UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

School of Medicine

DETAILED MOLECULAR STUDIES OF CHROMOSOME REARRANGEMENTS IN MAN

by

Júlia da Conceição Pereira Baptista

Thesis for the degree of Doctor of Philosophy

February 2008

UNIVERSITY OF SOUTHAMPTON
ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES
SCHOOL OF MEDICINE
Doctor of Philosophy

DETAILED MOLECULAR STUDIES OF CHROMOSOME REARRANGEMENTS IN MAN

by Júlia da Conceição Pereira Baptista

Apparently balanced chromosome rearrangements (ABCRs), mainly reciprocal translocations and inversions, are common in our species and are present both in patients with clinical abnormalities and in phenotypically normal individuals. The four main features that are thought to explain clinical abnormalities in patients with ABCRs are: (i) breakpoint-mediated gene disruption, (ii) breakpoint-associated genomic imbalances, (iii) additional chromosomal complexity and (iv) genomic imbalances unrelated to the breakpoints. The work presented in this thesis represents one of the first systematic studies to ascertain the occurrence of the above four features in both phenotypically normal and abnormal carriers of ABCRs.

Molecular cytogenetic analyses by FISH and/or array painting and by array CGH were applied in the characterisation of ABCRs in 31 phenotypically normal individuals (control cohort) and in 16 phenotypically abnormal patients (patient cohort). The occurrence of the above four features was assessed in both cohorts and the results were compared in an attempt to determine if the ABCRs in these two groups are molecularly distinct.

Genomic imbalances both at the breakpoints and unrelated to the breakpoints and additional chromosomal complexity were present in 25% of the cases in the patient cohort, but in none of those in the control cohort. Surprisingly, breakpoint-mediated gene disruption was equally frequent in both cohorts. However, there was a difference in the type of genes involved, with those of the patient cohort being more commonly involved in development and function of the nervous system. These observations suggest that there are molecular differences in the two groups of carriers. Furthermore, among the four features analysed, genomic imbalances both at the breakpoints and elsewhere in the genome appear to be the main cause of phenotypic abnormalities in carriers of ABCRs; nevertheless disease candidate genes have also been identified and future studies will assess their contribution to the abnormal phenotypes.

In summary, the application of molecular techniques is an invaluable approach to characterise genomic regions involved in chromosome rearrangements and other genomic regions unrelated to the breakpoints, thus aiding in the understanding of the contribution of these genomic regions to human disease and to human variation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DECLARATION OF AUTHORSHIP

I, Júlia da Conceição Pereira Baptista,

declare that the thesis entitled:

Detailed molecular studies of chromosome rearrangements in man and the work presented in it are my own. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- no part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:

1) Baptista J *et al.* (2008) Breakpoint mapping and array CGH in translocations: comparison of a phenotypically normal and an abnormal cohort. Am J Hum Genet. (in press)

2) Gribble SM *et al.* (2007) Ultra-high resolution array painting facilitates breakpoint sequencing. J.Med.Genet. 44:51-58

3) Baptista J *et al.* (2005) Molecular cytogenetic analyses of breakpoints in apparently balanced reciprocal translocations carried by phenotypically normal individuals. Eur.J.Hum.Genet. 13:1205-1212

CONTRIBUTION OF THE AUTHOR TO THE WORK PRESENTED IN THIS THESIS

The work presented in this thesis resulted from collaboration between the Wessex Regional Genetics Laboratory (WRGL) and Dr Nigel Carter's research team at the Wellcome Trust Sanger Institute (WTSI).

Chapter 2: Dr. Catherine Mercer, a Clinical Geneticist from the Wessex Clinical Genetics Service, visited and carried out clinical examinations on the majority of the study subjects. Detailed clinical descriptions were provided by Dr. Catherine Mercer and the referring consultants.

Chapter 3: The lymphoblastoid cell lines established at the WRGL were prepared by the author. PHA stimulated cultures for FISH studies were prepared either by the author or by colleagues in the Blood laboratory (WRGL). The extraction of genomic DNA was carried out by colleagues in the DNA laboratory (WRGL). The preparation of probe DNA for FISH, including bacterial growth, DNA extraction and labelling were all carried out by the author, with the exception of 6 rearrangements which were mapped by Viv Maloney from the National Genetics Reference Laboratory (Wessex). Additional probes were provided by colleagues in the Probe laboratory (WRGL).

Chapter 4: Elena Prigmore (WTSI) performed array CGH analysis on 13 phenotypically normal individuals and the high resolution array painting experiments in 5 of these cases. Array CGH in the remaining 18 individuals were performed by the author at the WRGL, using the microarray slides provided by the WTSI.

Chapter 5: Bivariate karyotype analysis and flow sorting of derivative chromosomes were performed by Bee Ng (WTSI). The array painting and the array CGH experiments with Sanger WGTP arrays were performed by the author during two laboratory visits to the WTSI. The work carried out during these visits was undertaken under the supervision of Susan Gribble. Array CGH with the Agilent 244K oligonucleotide array was performed by Shuwen Huang (WRGL). Microsatellite analysis for parent of origin studies were performed by Victoria Bryant and N. Simon Thomas (WRGL).

ACKNOWLEDGEMENTS

I am truly indebted to my supervisor, Dr John Crolla for his help, guidance and support throughout this project. I also thank Professor Nick Cross for providing the infrastructure which makes this work possible.

Thanks to all patients and their families for taking part in this study. They made this work possible and more importantly, they were the reason for this project in the first place.

Thanks to Catherine Mercer for her dedication and for being willing to travel many miles to visit patients and collect invaluable clinical data.

Thanks to Barbara O'Prey for a great job in contacting the patients.

Thanks to all colleagues at the Wessex Regional Genetics Laboratory who have contributed to this project in many different ways. I have been very fortunate to be surrounded by experts who gave me their time, advice and the training I needed in the past few years. Thanks to Sarah Beal and to Viv Maloney for teaching me all about FISH. Thanks to Andrew Fisher for training me in cell culture techniques and to Jo Score for training me in PCR techniques. I also thank Derek Richardson, Annette Cockwell and N Simon Thomas for all their help and useful suggestions.

Thanks to all my collaborators in the research group led by Dr Nigel Carter at the Wellcome Trust Sanger Institute, especially Susan Gribble, and Elena Prigmore for all their input in this project.

I also acknowledge the Wellcome Trust for funding part of this project.

Thanks to the Archives Team at the Welcome Trust for providing the many FISH probes used in this study and also to Professor Rocchi at the University of Bari for additional clones.

Thanks to my family and friends for being there for me. Special thanks to Francesco, for everything!

Very special thanks to Professor Pat Jacobs, whose enthusiasm, energy and passion for science are an inspiration. Her attempts to shape me into a scientist and her support during this project, which included reading many drafts of this thesis, are greatly appreciated. I really wouldn't have made it without Pat and being her student has been a great privilege.

ABBREVIATIONS

ABCR	Apparently balanced chromosome rearrangement
AP	Array painting
Array CGH	Array comparative genomic hybridisation
BAC	Bacterial artificial chromosome
bps	Basepairs
Chr	Chromosome
CNV	Copy number variation
Cy3	Cyanine-3 dye
Cy5	Cyanine-5 dye
Der	Derivative
DNA	Deoxyribonucleic acid
DOP PCR	Degenerated oligonucleotide primed polymerase chain reaction
DGV	Database of Genomic Variants
FISH	Fluorescence in situ hybridisation
GO	Gene ontology
ISCN	International System for Human Cytogenetic Nomenclature
Kb	Kilobasepairs
LR-PCR	Long range polymerase chain reaction
Mb	Megabasepairs
mat	maternal
MIM	Mendelian Inheritance in Man
min	minute
mM	millimolar
NCBI	National Centre for Biotechnology Information
ng	nanogram
OFC	Occipitofrontal head circumference
PAC	P1 artificial chromosome
pat	paternal
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
s	second
SSC	Saline/Sodium Citrate
STOIC	Salisbury treasury of interesting chromosomes
UCSC	University of California Santa Cruz
URL	Uniform Resource Locator
µL	microliter
µM	micromolar
WGTP	Whole Genome Tiling Path
WRGL	Wessex Regional Genetics Laboratory
WTSI	Wellcome Trust Sanger Institute
YAC	Yeast artificial chromosome

1 INTRODUCTION

1.1 Modern human cytogenetics: the beginning

Modern human cytogenetics began in 1956 after Tjio and Levan established the correct diploid number of chromosomes in man as 46 and not 48 as supposed for the previous three decades. In the same year, Ford and Hamerton (1956) independently confirmed this finding. Three years later, the first chromosome abnormalities were described showing that a deviation from the diploid number of 46, i.e. aneuploidy, could, in some instances, explain clinical abnormalities; Jacobs and Strong (1959) reported an additional chromosome X in a patient with Klinefelter syndrome, an extra chromosome 21 was identified in patients with Down syndrome (Jacobs *et al.*, 1959; Lejeune *et al.*, 1959) and a 45,X karyotype was reported in a case of Turner syndrome (Ford *et al.*, 1959). These findings were rapidly followed by others, including reports of trisomy 13 (Patau *et al.*, 1960), trisomy 18 (Edwards *et al.*, 1960) and the first deletion syndrome, Cri du chat (Lejeune *et al.*, 1963). The role of chromosome abnormalities in cancer was also recognised after an abnormal chromosome was reported to cause chronic myeloid leukaemia (Nowell and Hungerford, 1960). This aberrant chromosome was referred to as the Philadelphia chromosome and was later shown to be a derivative of a translocation between chromosomes 9 and 22 (Rowley, 1973). Taken together, these pioneering studies showed that there was a causal relationship between chromosome complement and phenotypic abnormalities, thus establishing the clinical applications of human cytogenetics.

Numerical abnormalities were the first class of chromosome abnormality described, but were rapidly followed by the documentation of large structural rearrangements, including Robertsonian translocations and deletions (review in Hamerton, 1971). A more detailed characterisation of structural rearrangements was possible with the introduction of chromosome banding techniques. Caspersson and colleagues developed the first banding method, Q-banding. This technique used quinacrine mustard or quinacrine dihydrochloride to obtain a fluorescence banding pattern unique to each pair of homologues, thus enabling the identification of each individual chromosome (Caspersson *et al.*,

1970). Other banding techniques were developed soon afterwards, of which G-banding became the most prominent (Seabright, 1971).

1.2 The frequency of chromosome abnormalities in man

Following the realisation that a chromosome abnormality could represent an observation of great clinical importance, studies were established to determine the frequency of chromosome abnormalities in the human species. This was attempted by studies in three main sample groups: spontaneous abortions (fetal death between 6-8 and 24 weeks), perinatal deaths (fetal and neonatal death between 24 weeks and up to one week following term) and newborn babies. These studies suggested that at least 50% of spontaneous abortions have a chromosome abnormality, of which numerical abnormalities account for the great majority, while structural abnormalities are seen in only 2% of samples (reviewed by Hassold, 1986). Among perinatal deaths, the overall frequency of chromosome abnormalities is 5%, of which 0.8% are structural rearrangements (Machin and Crolla, 1974). By comparison, cytogenetic studies of series of consecutive newborn babies showed the frequency of chromosome abnormalities to be much lower in this population, with an overall frequency of 0.6% of which 0.26% are structural rearrangements (Hook and Hamerton, 1977). Many of these early studies were carried out using block staining; hence the figures presented for structural abnormalities must have been underestimated. In this context, using moderate levels of banding applied to a series of cytogenetic prenatal diagnoses, the frequency of human liveborn chromosome abnormalities was estimated to be 0.92%. Over half of the abnormalities were structural chromosome rearrangements and of these 0.52% were balanced and 0.06% unbalanced (Jacobs *et al.*, 1992).

An important biological observation from these studies was that chromosome abnormalities represent an important cause of fetal loss as there is a 50-fold enrichment for chromosome abnormalities in spontaneous abortions when compared to livebirths. Furthermore, there was relatively little variation in the

frequency of structural rearrangements among the three populations (which varied between 2 and 0.6%). Thus, the differences in the frequency of chromosome abnormalities among the three types of populations is due to a higher incidence of numerical abnormalities in spontaneous abortions and stillbirths by comparison with livebirths, a finding that reflects the incompatibility of many gross abnormalities with normal fetal development. Moreover, these surveys have shown that structural chromosome rearrangements both balanced and unbalanced, which include translocations, inversions, insertions, ring chromosomes, and complex rearrangements are relatively common in man, being detectable in 0.58% of liveborns using moderate levels of banding (Jacobs *et al.*, 1992). They are of particular interest because, unlike numerical abnormalities, they represent a class of chromosome abnormality that is commonly inherited.

1.3 Apparently balanced structural chromosome rearrangements (ABCRs)

Apparently balanced chromosome rearrangements, i.e. without microscopically visible gain or loss of DNA are divided in three main classes: reciprocal translocations, Robertsonian translocations and inversions. Balanced reciprocal translocations occur in approximately 1 in 500 people (Jacobs *et al.*, 1992) and are interchromosomal rearrangements characterised by two breaks and exchange of DNA segments between non homologous chromosomes. Robertsonian translocations are a particular type of translocation limited to the acrocentric chromosomes and are present in 1 in 1000 people (Jacobs *et al.*, 1992). Inversions are intrachromosomal rearrangements characterised by two breaks on the same chromosome and inversion of the interstitial DNA segment and are estimated to have a frequency of 1 in 1000 (Jacobs *et al.*, 1992). Robertsonian translocations will not be discussed here and henceforth the term apparently balanced chromosome rearrangement (ABCR) will be used to refer collectively to reciprocal translocations and inversions.

ABCRs segregate in many families without phenotypic effect. Some however, are diagnosed when associated with reproductive difficulties, including recurrent miscarriages, stillbirths and/or offspring with congenital abnormalities due to segregation of an unbalanced form of the rearrangement (Jacobs *et al.*, 1975). In a study by Youings *et al.* (2004) based on all samples sent to a regional diagnostic cytogenetics laboratory, 29% of the apparently balanced translocations had been ascertained because of reproductive difficulties, 27% because of an abnormal phenotype and 25% at prenatal diagnosis. In contrast, inversions had been ascertained prenatally in 38% of the cases, through an abnormal phenotype in 31% and only 17% due to reproductive difficulties.

The probability of an ABCR being diagnosed is higher for individuals with an abnormal phenotype or reproductive difficulties, who are more likely to be clinically investigated than individuals with a normal phenotype. This ascertainment bias is well recognised. Hence, when a *de novo* ABCR is diagnosed post-natally in association with an abnormal phenotype it is difficult to determine whether the rearrangement is the cause of the clinical problems or whether it is a co-incidental finding. The proportion of *de novo* cases is estimated to be 23.5% for reciprocal translocations and 10% for inversions (Jacobs *et al.*, 1992); and for these, a combined risk for clinical abnormalities is quoted as 6.7% according to estimates derived from prenatally ascertained cases (Warburton, 1991).

1.4 Chromosome rearrangements and the identification of disease genes

One of the mechanisms by which an ABCR is associated with clinical abnormalities is the disruption of a disease gene. The gene for Duchenne muscular dystrophy (DMD [MIM 310200]) was one of the first to be mapped following the characterisation of apparently balanced translocations. The observation that affected females with X;autosomal translocations shared a common breakpoint on the short arm of chromosome X, led to the hypothesis

that this breakpoint had split the *DMD* gene (Greenstein *et al.*, 1977; Jacobs *et al.*, 1981). This gene was one of the first to be isolated by positional cloning (Monaco *et al.*, 1986). Similarly, the study of other chromosome rearrangements with common breakpoints in patients with similar phenotypes has lead to the identification of disease loci thus accelerating the process of gene isolation and mapping by positional cloning (Collins, 1992). The study of chromosome rearrangements has enabled the identification of several genes for Mendelian disorders (reviewed by Tommerup, 1993) and continues to be a powerful approach to identify additional genes implicated in Mendelian or in complex disease (reviewed by Bugge *et al.*, 2000; Bache *et al.*, 2006).

1.5 New developments for the study of ABCRs

It has long been thought that besides gene disruption by a breakpoint, other mechanisms might be the underlying cause of an association between an ABCR and clinical abnormalities. These mechanisms include the rearrangement not being truly balanced, a position effect of the breakpoints (see section 1.6.2) or a coincidental rather than causal association (Jacobs, 1974). A detailed investigation of these potential scenarios often requires higher resolution than that afforded by conventional cytogenetics alone, particularly for the exclusion of submicroscopic imbalances.

1.5.1 Molecular cytogenetics

The development of fluorescence in situ hybridisation (FISH) has permitted the study of chromosome rearrangements at a resolution significantly higher than that of conventional cytogenetics. FISH is based on the hybridisation of a labelled DNA or RNA probe to patient genomic DNA and it was developed from the observation that radioactively labelled ribosomal RNA probes hybridised to acrocentric chromosomes (Gall and Pardue, 1969; John *et al.*, 1969). Radioactive isotopic labels were used initially, but were later replaced

with fluorochromes, rendering the technique safer and easier to use (Langer-Safer *et al.*, 1982; van Prooijen-Knegt *et al.*, 1982; Landegent *et al.*, 1985). Fluorochromes can be used to label the DNA directly, alternatively, the DNA can be labelled indirectly with haptens (digoxigenin and biotin are widely used) which are detected by antibodies (anti-digoxigenin to detect digoxigenin and avidin for detection of biotin) conjugated to fluorochromes. The availability of these straightforward systems, together with a range of fluorochromes of different colours which enable the testing of more than one probe simultaneously and an increasing accessibility to probes, greatly expanded the use of FISH.

1.5.2 The Human Genome Project: resources for molecular cytogenetics

The hierarchical shotgun sequencing approach adopted by the International Human Genome Sequencing Consortium consisted in initially fragmenting the genome in large segments which were subsequently cloned into various types of vectors for downstream applications, i.e. generation of physical maps followed by sequencing (Lander *et al.*, 2001). Thus, efficient cloning vectors were needed to ensure the success of this enterprise (reviewed by Zhao and Stodolsky, 2004). Cosmids were one of the first types of vectors available and are suitable to clone inserts of up to 45 kb (Collins and Hohn, 1978). However, the resulting clones can present instability, with insert loss during replication. Yeast artificial chromosomes (YACs) (Burke *et al.*, 1987) are another type of commonly used vector. One of the main advantages of YACs is that inserts of up to 1 Mb can be cloned, thus reducing the number of clones required to represent a whole genome. However, YAC clones were difficult to construct and also tended to undergo rearrangements leading to chimaerism. The limitations of these early systems lead to the development of more efficient vectors and the introduction of P1-derived artificial chromosomes (PACs), bacterial artificial chromosomes (BACs) and fosmids. The average insert size cloned into a PAC is 120 kb and the resulting clones proved easy to purify and presented low or non existent chimaerism or insert instability (Ioannou *et al.*,

1994). These same advantageous characteristics were shared by BACs, which accept an insert of an average size of 170 kb (Shizuya *et al.*, 1992) and fosmids, with an average insert size of 40 kb (Kim *et al.*, 1992). As a result of these developments, libraries of BAC, PACs and fosmids covering the entire human genome have been created. These individual clones can be utilised as DNA probes for FISH studies. A summary of the characteristics of the different cloning vectors is given in table 1.I.

Table 1.I. Vectors for molecular cloning

Name	Type	Average insert size
cosmid	plasmid-phage hybrid	35 kb
YAC	artificial chromosome, based on Yeast	500 kb
PAC	artificial chromosome, based on bacteriophage P1	120 kb
BAC	artificial chromosome, based on *E.coli* F factor	170 kb
fosmid	plasmid-phage hybrid based on *E.coli* F factor	40 kb

Besides the creation of large genomic libraries for molecular cytogenetic studies, the major contribution of the Human Genome Project has been the assembly of the sequence of the human genome and the mapping and identification of human genes. These invaluable data for the study of chromosome rearrangements has been deposited in publicly available databases and can be accessed via human genome browsers. Ensembl, hosted by the Wellcome Trust Sanger Institute, and the University of California Santa Cruz (UCSC) browser, are two of the main browsers available (URLs in Appendix 1). These browsers integrate sequence data from the reference human genome and provide annotation for known and predicted genes (see section 3.6) alongside numerous other features, e.g. expression data, comparative genomics and human variation and repeat elements.

1.6 Mechanisms for the basis of clinical abnormalities in ABCRs

1.6.1 Gene disruption by one or both breakpoints

As discussed in section 1.4, the identification of ABCRs in patients with clinical abnormalities presents the breakpoint regions as obvious locations for candidate disease genes. For example, Mansouri *et al.* (2005) characterised the breakpoints in a t(X;15)(q21.2;p12) associated with profound mental retardation, seizures, and normal stature. The patient had been previously reported (Gustavson *et al.*, 1984) and was thought to have Prader-Willi syndrome [MIM 176270], but molecular studies showed that the chromosome 15 breakpoint mapped 4.5 Mb centromeric to the Prader-Willi syndrome region. The Xq breakpoint was found to map to Xq13.3 and not Xq21.2 as assigned cytogenetically and it disrupted the *ZDHHC15* (zinc finger DHHC domain-containing protein 15) gene. Following expression analysis studies, the authors showed that no transcripts of *ZDHHC15* were present in the patient's lymphoblasts. In addition, the normal X was shown to be inactivated in all the cells analysed. This study suggested that haploinsufficiency of *ZDHHC15* might be a cause for nonsyndromic X-linked mental retardation. Furthermore, this study is a very good example of the importance of a well delineated clinical phenotype for the interpretation of cytogenetic findings. In another report, a t(1;8)(p34.3;q21.12)*de novo* was characterised in an attempt to identify a gene for congenital bilateral isolated blepharoptosis which had been previously suggested, by linkage studies, to map to 1p32-34.1 [MIM 178300]. Interestingly, the authors found that the 1p breakpoint was mapped ~13 Mb telomeric to the reported candidate locus, but the 8q breakpoint caused disruption of *ZFH-4* (zinc finger homeodomain 4 protein), which was considered a good candidate gene for congenital bilateral isolated blepharoptosis (McMullan *et al.*, 2002). Similarly, FitzPatrick *et al.* (2003) analysed translocations with breakpoints in 2q32-33 in an attempt to identify a gene for cleft palate. This genomic location had been previously implicated in cleft development from studies of deletions and translocations. The authors

found that the *SATB2* (SATB homeobox 2) gene was interrupted by the 2q translocation breakpoint in one of the patients analysed, while in a second patient, the 2q breakpoint mapped 130 kb 3' to the *SATB2* polyadenylation signal. Studies of the pattern of expression of *SATB2* in mouse embryos provided further evidence of a role of the gene in palate development. Other reports of identification of disease candidate genes by the study of breakpoint regions include for example, a report of *SNX3* (sorting nexin 3) as a candidate for microcephaly, microphthalmia, ectrodactyly, prognathism phenotype [MIM 601349] (Vervoort *et al.*, 2002a), *NPAS3* (neuronal PAS domain protein 3) as a candidate gene for schizophrenia (Kamnasaran *et al.*, 2003a; Pickard *et al.*, 2005) and *NTNG1* (netrin G1) as a candidate gene for Rett syndrome [MIM 312750] (Borg *et al.*, 2005).

Following the identification of a good disease candidate gene, further evidence for a role in disease is given by the detection of mutations in other affected patients. For example, Zemni *et al.* (2000) reported the disruption of the *TSPAN7* (tetraspanin 7) gene at Xp11.4 in a patient with mental retardation and a balanced t(X;2). Subsequent mutational analysis identified two families with *TSPAN7* mutations and X-linked mental retardation. Similarly, Johnson *et al.* (2006) confirmed the involvement of *CHD7* (chromodomain helicase DNA binding protein 7) in CHARGE (Coloboma, heart anomaly, choanal atresia, retardation, genital and ear abnormalities) syndrome [MIM 214800] when the gene was found to be disrupted by the 8q11 breakpoint of a t(8;13)(q11.2;q22)*de novo* in a monozygotic twin pair. This finding was supported by a previous report of microdeletions and point mutations involving *CHD7* in unrelated CHARGE patients (Vissers *et al.*, 2004).

Gene interruption can also occur at both breakpoints of a translocation making it difficult to interpret the contribution of each gene to the phenotype, especially when fusion transcripts are formed (Nothwang *et al.*, 2001; Ramocki *et al.*, 2003; Yue *et al.*, 2005; Kim *et al.* 2005; Mansouri *et al.*, 2006). These difficulties are well illustrated by a report of a patient with mental retardation and a t(X;9)(p11.23;q34.3)*de novo* (Kleefstra *et al.*, 2004). *ZNF81* on Xp11.23 and *EHMT1* (euchromatic histone-lysine N-methyltransferase 1) on 9q34.3

were both disrupted. *ZNF81* (zinc finger protein 81) was thought to be a good candidate gene for mental retardation and subsequent mutational analyses identified a missense mutation in this gene in an unrelated affected family. However, a re-evaluation of the patient's phenotype (Kleefstra *et al.*, 2005) suggested a similarity with the 9q34 subtelomeric deletion syndrome and the authors suggested that haploinsufficiency of *EHMT1* was the cause of the patient's abnormalities. However, it could well be that both genes contribute to the phenotype. This report is another example of the value of a detailed clinical description in the interpretation of molecular cytogenetic findings.

1.6.1.1 The unmasking of a recessive mutation

For cases with breakpoints within genes, an abnormal phenotype can arise due to gene haploinsufficiency caused by the inactivation of a gene by a breakpoint. Another possible scenario is that of the inactivation of a normal allele by a breakpoint leading to the unmasking of a recessive mutation in a heterozygous carrier (Buhler, 1983). The gene for Alstrom syndrome [MIM 203800], *ALMS1* (Alstrom syndrome 1), was identified by just such a mechanism. Hearn *et al.* (2002) reported the analysis of an apparently balanced t(2;11)(p13;q21)mat in an affected patient. *ALMS1* was disrupted by the 2p13 breakpoint in the maternally inherited chromosome and a frameshift mutation was present in the paternal chromosome. Further evidence of the role of *ALMS1* in Alstrom syndrome was provided when mutations in the gene were detected in unrelated affected families.

1.6.2 Gene modulation by position effect

Phenotypic abnormalities in patients with an ABCR can arise by position effect when the breakpoints do not disrupt a gene directly. A position effect (PE) can be defined by gene function alteration despite the fact that the transcription unit of the gene remains intact. The mechanism was first described in *Drosophila*,

where translocation of active genes into heterochromatic domains led to gene inactivation by what was termed position effect variegation (reviewed by Karpen, 1994). A similar, but not identical PE mechanism for gene inactivation has also been reported for a number of genes implicated in human disease, e.g. patients with aniridia [MIM 106210] in whom the 11p13 translocation breakpoints mapped 100 and 125 kb downstream of the responsible gene *PAX6* (paired box 6) (Fantes *et al.*, 1995). While most PE breakpoints occur within 100-200 kb of the target gene, Velagaleti *et al.* (2005) reported two patients with translocation breakpoints mapping 900 kb upstream and 1.3 Mb downstream of the *SOX9* (sex determining region Y-box 9, campomelic dysplasia, autosomal sex-reversal) gene in patients with campomelic dysplasia [MIM 114290]. Therefore, PE can be caused by breakpoints mapped as distant as 1 Mb both downstream and upstream of the responsible gene. It is thought that PE is due to chromosome rearrangements altering transcriptional control by affecting interactions between the promoter or transcription unit and cis regulators or by alterations of regulation of chromatin structure (reviewed by Kleinjan and van Heyningen, 2005).

1.6.3 Cryptic imbalances in ABCRs

Kumar *et al.* (1998) reported one of the first systematic studies of ABCRs using FISH, which aimed to determine if cryptic imbalances near the breakpoints represented a main cause of phenotypic abnormalities in clinically abnormal carriers of ABCRs. The authors detected cryptic deletions of 4-6 Mb near the breakpoints in 2/3 patients. Similarly, Wirth *et al.* (1999) detected cryptic deletions of 3-5 Mb in one or both breakpoint regions in 2/6 patients with mental retardation and an apparently balanced *de novo* translocation. These findings have been confirmed by other reports (Borg *et al.*, 2002; Cox *et al.*, 2003; Astbury *et al.*, 2004) showing that a proportion of ABCRS are in fact unbalanced.

Imbalances not directly at but near the breakpoints have also been described; Fantes *et al.* (2003) reported the characterisation of a *de novo* t(3;11)(q26.3;p11.2) in a patient with anophthalmia. The authors identified a deletion mapping ~600 kb centromeric to the 3q breakpoint, which included a good candidate gene for anophthalmia [MIM 206900], *SOX2* (sex determining region Y-box 2). The identification of *de novo* mutations in 4 of 35 affected patients provided further evidence of a role of *SOX2* in anophthalmia.

1.6.4 Additional chromosomal complexity in ABCRs

Besides imbalances, molecular studies can reveal if an ABCR is associated with cryptic chromosomal complexity. The exclusion of a cryptic rearrangement has important genetic counselling implications as the risk of phenotypic abnormalities is increased when additional complexity is present. This is because a higher number of breakpoints will increase the probability of gene disruption or the likelihood of a genomic imbalance (Madan *et al.*, 1997, reviewed by Patsalis, 2007).

Patsalis *et al.* (2004) reported one of the first systematic studies of ABCRs by FISH that aimed to assess the occurrence of cryptic chromosomal complexity in ABCRs. The authors analysed 20 apparently balanced translocations, including both familial and *de novo* translocations, the majority of which were associated with recurrent miscarriages (n=16) and the others with mental retardation and/or congenital malformations. Three cases were found to be complex rearrangements, of which two were patients with clinical abnormalities and the third case was an individual with an apparently normal phenotype which had been referred because of recurrent miscarriages.

In another report, Kamnasaran *et al.* (2003b) characterised an apparently balanced t(4;14)(q25;q13) in a patient with mild craniofacial and acallosal central nervous system midline defects. The 14q breakpoint was mapped within the *MIPOL1* (mirror-image polydactyly 1) gene, while two breakpoints were

identified on 4q, suggesting the presence of an additional rearrangement (potentially an inversion). In addition, the authors identified a microdeletion (~50 kb) of the *PITX2* (paired-like homeodomain 2) gene and suggested that this was likely to be the cause of the phenotypic abnormalities. This report is of particular interest as it shows that an apparently balanced translocation can be associated with breakpoint-mediated gene disruption, cryptic imbalances and additional complexity, all features that are likely to explain abnormal phenotypes and which also add complexity to data interpretation.

1.6.5 Uniparental disomy

Uniparental disomy (UPD), that is the inheritance of both homologues from the same parent (Engel, 1980), has also been suggested as a mechanism by which an ABCR can be associated with an abnormal phenotype. This is because carriers of ABCRs are at higher risk of meiotic nondisjunction, which in turn can lead to UPD. UPD can cause clinical abnormalities if the chromosomes involved have regions that are subject to imprinting (i.e. expression dependent on parental origin). It is estimated that 1% of human genes are subject to imprinting and in a recent effort to provide a map of imprinted genes in the human genome, several new candidate genes have been identified (Luedi *et al.*, 2007). Alternatively, UPD might lead to homozygosity of a recessive mutation and cause clinical abnormalities by this mechanism. Nevertheless, a study of apparently balanced translocations in 65 families detected UPD in only one case suggesting that UPD is likely to be a rare event among apparently balanced translocations (James *et al.*, 1994).

1.7 DNA microarrays: applications to the study of ABCRs

Recently, the application of DNA probes to microarrays has emerged as a powerful technology in genetics studies. Microarray based comparative genomic hybridisation (array CGH) enables the detection of copy number

changes by competitively hybridising differentially labelled test and reference DNA to arrays of spotted and mapped clones. Thus, the technique allows the rapid screening of the whole genome at a resolution determined by the density of the markers spotted onto the array. This technique evolved from comparative genomic hybridisation (Kallioniemi *et al.*, 1992) with the replacement of metaphase chromosomes with spotted clones as the hybridisation target (Solinas-Toldo *et al.*, 1997; Pinkel *et al.*, 1998). Array CGH was used at first in the study of tumour samples (reviewed by Albertson and Pinkel, 2003). More recently, this technique has been widely applied to the characterisation of patients with a normal karyotype and congenital abnormalities or mental retardation (Vissers *et al.*, 2003; Shaw-Smith *et al.*, 2004; deVries *et al.*, 2005, Rosenberg *et al.*, 2006; Menten *et al.*, 2006; Friedman *et al.*, 2006) and it is also thought to have a potential application in prenatal diagnosis (Rickman *et al.*, 2005).

Gribble *et al.* (2005) reported one of the first systematic studies of ABCRs in 10 patients with phenotypic abnormalities using a combination of FISH, array painting and array CGH. This study showed that 3 of the ABCRs were actually complex rearrangements with deletions, inversions and insertions at or near the breakpoints. Three other patients had imbalances in a chromosomal region independent of the translocation; which would have remained undetected if array CGH analysis had not been used. The finding that such a significant proportion of ABCRs had imbalances independent of the breakpoints was particularly surprising and more importantly, it showed that other genomic regions independent of the ABCRs might contribute to the phenotypic abnormalities and that array CGH will be invaluable to assess these imbalances. Similarly, Ciccone *et al.* (2005) analysed 4 ABCRs by array CGH and FISH and confirmed previous findings that many ABCRs might have an unsuspected level of complexity. More recently, the potential of array CGH for the study of ABCRs was further validated by a report of De Gregori *et al.* (2007). The authors analysed a series of 47 apparently balanced reciprocal translocations and 18 complex rearrangements both in patients with clinical abnormalities, in phenotypically normal individuals and in a group of prenatal samples of unknown phenotype. Their findings showed that cryptic deletions both at the

breakpoints and in regions independent of the breakpoints are often present in patients with phenotypic abnormalities and ABCRs.

1.8 Analysis of normal individuals: the link between molecular findings and clinical abnormalities

The finding of a *de novo* mutation in association with clinical abnormalities is often regarded as the explanation for the phenotype. However, only the systematic characterisation of normal individuals will provide the necessary reference data to aid in establishing whether a mutation is likely to be causal or coincidental to an abnormal phenotype.

Evidence that cytogenetically visible imbalances can segregate without phenotypic effect is available from studies of directly transmitted unbalanced chromosome abnormalities (UBCA) and euchromatic variants (EV). Barber (2005) has reviewed 200 families and found that 18% of the UBCAs had no phenotypic effect and in 23% an affected proband had inherited the UBCA from a normal carrier. For EV, 54% of the probands were normal and 43% were abnormal, but all had unaffected family members with the same EV.

Recently, a systematic assessment of human submicroscopic variation has been facilitated by the application of array CGH to the analysis of normal individuals (reviewed by Feuk *et al.*, 2006). Large regions (~100 kb and greater) of copy number variation (CNV) have been identified across the genome, showing that human genetic variation is considerable (Iafrate *et al.*, 2004; Sebat *et al.*, 2004; Sharp *et al.*, 2005, Redon *et al.*, 2006). In addition to deletions and duplications, submicroscopic insertions and inversions contribute greatly to structural variation (Tuzun *et al.*, 2005; Feuk *et al.*, 2005). These discoveries emphasize that the finding of imbalances, insertions or inversions in patients with clinical abnormalities might reflect normal variation rather than pathogenic mutation.

1.9 The study of ABCRs in phenotypically normal individuals

There are very few studies of ABCRs in individuals with no phenotypic abnormalities, with the exception of the t(11;22)(q23;q11), which has been extensively studied as it is the only common recurrent non-Robertsonian constitutional translocation in man. The lack of studies of ABCRs in normal individuals is mainly because these do not obviously offer the prospect of identifying disease causing genes. However, ABCRs might help to identify genes associated with late onset disease; for example, Foster *et al.* (2007) have characterised the breakpoints in a constitutional t(3;6)(q22;q16.1) in an attempt to identify genes for renal cell carcinoma [MIM 144700].

In one of the first systematic efforts to characterise ABCRs in phenotypically normal individuals, Baptista *et al.* (2005) analysed 13 ABCRs using FISH and array CGH. The authors aimed to determine if ABCRs in this group of individuals share the molecular characteristics seen in phenotypically abnormal individuals, i.e. breakpoint-mediated gene disruption, cryptic imbalances or additional chromosomal complexity. These analyses showed that, at the level of resolution of the techniques applied, the ABCRs had no genomic imbalances at the breakpoints or elsewhere or additional complexity. However, gene disruption was found in at least 3/13 patients. Due to lack of a detailed clinical history on these normal individuals, it was not possible to evaluate the significance of the gene disruptions observed. Since the individuals analysed were developmentally and intellectually apparently normal, it is reasonable to assume that the genes disrupted were not dosage sensitive or developmentally regulated. However, these could have a role in late onset disorders. These observations prompted the need to expand studies of ABCRs to additional normal individuals from whom full clinical details were available.

1.10 Aims of this thesis

A review of the literature, as presented in the preceding sections, shows that the study of apparently balanced rearrangements is a valuable resource for knowledge advancement both in the field of medical genetics and in the understanding of normal variation. An ABCR associated with a clinical abnormality offers an opportunity to understand the role of the genomic regions involved, leads to an understanding of the underlying chromosomal pathogenesis, helps to resolve the contribution of imbalances and gene disruption in human development and disease, and as a result, new genes can be discovered and the role of known genes can be more fully assessed. This potential has led to an almost exclusive focus on the analysis of cases associated with phenotypic abnormalities and virtually all the literature available is on single case reports or small series of patients, mainly selected because of potentially interesting clinical features. Hence, a gap remains in the understanding of the characteristics of apparently balanced rearrangements in the general population. The work presented in this thesis will attempt to address this aspect by studying unselected ABCRs in both a control cohort, i.e. carriers of apparently balanced rearrangements with no clinical abnormalities (chapter 4) and a patient cohort, i.e. patients with clinical abnormalities (chapter 5). The availability of a detailed clinical history in both cohorts will aid in the interpretation of the findings and in the establishment of genotype-phenotype correlations. Molecular analyses will be undertaken with up to date technologies available for the study of chromosome rearrangements, i.e. FISH, array painting and array CGH.

The main aim of these analyses is to attempt to answer the following question: are apparently balanced rearrangements molecularly distinct in the patient cohort by comparison to the control cohort? Both cohorts will be evaluated with respect to the four main features known to cause phenotypic abnormalities in carriers of apparently balanced rearrangements: (i) breakpoint-mediated gene disruption, (ii) cryptic breakpoint-associated imbalances, (iii) additional chromosomal complexity and (iv) genomic imbalances unrelated to the

breakpoints. Additionally, the results of both cohorts will be compared (chapter 6) to determine whether there is a distinction in the phenotypically normal and abnormal carriers and to help to assess the contribution of each of the above four features to the genesis of phenotypic abnormalities.

2 STUDY POPULATIONS

2.1 Overview

The Salisbury Treasury of Interesting Chromosome (STOIC) is a collection documenting all cytogenetic abnormalities diagnosed at the Wessex Regional Genetics Laboratory (WRGL) since 1967. Part of this collection was used to identify subjects with apparently balanced chromosome rearrangements. In an initial phase of the current project, thirteen anonymised samples of individuals with no known phenotypic abnormalities were investigated. These cases were selected solely by the fact that they presented with a normal phenotype and a lymphoblastoid cell line had been set up and stored at the WRGL. The results of this preliminary study prompted the need to recruit additional normal individuals from whom a detailed clinical history was available. Thus, after ethical approval was granted by a Multi-Centre Research Ethics Committee (reference MREC 03/8/095), additional subjects (n=42) were selected from STOIC. They were then contacted via a letter sent to the consultant who referred them for cytogenetic analysis. Eighteen of these individuals provided consent and were therefore included in the study. Simultaneously, a second group of patients (n=35), all of whom were ascertained because of phenotypic abnormalities, were selected and contacted in a similar manner. Eleven of these patients consented into the study. In addition, five patients with phenotypic abnormalities were referred by Consultants from other regional genetic units.

2.2 The control cohort: 31 phenotypically unaffected individuals

This study included thirty one apparently balanced reciprocal translocations in phenotypically normal individuals. Nineteen translocations were *de novo* events, seven were maternally inherited and the remaining five were paternally inherited. The majority of individuals (fourteen cases) were ascertained through a relative presenting with phenotypic abnormalities and nine others were ascertained because of recurrent miscarriages. The remaining eight individuals were ascertained following the detection of an abnormal karyotype in a prenatal

sample (four tests were performed because of advanced maternal age, three due to a family history of Down syndrome and one because of high serum screen risk). Table 2.I lists the karyotypes and details of the mode of ascertainment of these phenotypically normal individuals. Detailed clinical descriptions of these individuals are available for the eighteen individuals recruited in the later stages of the project, but not for the thirteen cases recruited early (available clinical details are listed in Appendix 2).

2.3 The patient cohort: 16 phenotypically affected patients

Sixteen patients presenting with an apparently balanced rearrangement and phenotypic abnormalities were analysed, including fourteen reciprocal translocations, one insertional translocation and one pericentric inversion. Fifteen rearrangements had occurred *de novo* and one was of unknown origin (the maternal karyotype was normal, but a paternal sample was unavailable). Developmental delay was the major reason for ascertainment and was seen in nine cases. Three other patients were referred because of premature ovarian failure (one with primary amenorrhea and two with secondary amenorrhea), two patients had severe oligospermia, one had dysmorphic features and the remaining patient was ascertained because of isolated truncus arteriosus. Karyotypic details and the mode of ascertainment of the phenotypically abnormal patients are given in table 2.II. Patients were subjected to a detailed clinical examination by a Clinical Geneticist and these descriptions are given in Appendix 3.

Table 2.I. Karyotypes and mode of ascertainment in the control cohort

Case[1]	Karyotype	Ascertainment
1.1	46,XY,t(2;14)(p21;q13)*de novo*	Parent of a 46,XX,t(2;14)(p21;q13) miscarriage
1.2	46,XX,t(3;9)(p26.2;p22.3)*de novo*	Parent of a 46,XY,add(3)(p25.3) phenotypically abnormal child
1.3	46,X,t(X;7)(?q27;q22)*de novo*	Parent of a 46,X,Xq+ phenotypically abnormal child
1.4	46,XX,t(10;18)(q24.3;q12.2)mat	Relative of 46,XY,t(10;18)(q24.3;q12.2) ascertained prenatally because of high serum screen risk
2.1	46,XX,t(4;16)(q35.1;p13.13)*de novo*	Parent of a phenotypically abnormal child presenting a rea(4)
2.2	46,XX,t(1;13)(q32.3;q32.3)*de novo*	Parent of a 46,XY,13q+ phenotypically abnormal child
2.3	46,XY,t(2;18)(q35;q21.3)*de novo*	Recurrent miscarriages
2.4	46,XX,t(2;9)(q21.3;p13)*de novo*	Parent of a 46,XX,t(2;9)(q21.3;p13) amniocentesis because of AMA
2.5	46,XX,t(7;17)(q36.1;q25.1)*de novo*	Recurrent miscarriages
2.6	46,XY,t(8;15)(p11.2;q24)*de novo*	Parent of a 46,XY,t(8;15)(p11.2;q24) amniocentesis because of AMA
2.7	46,XY,t(1;13)(p22;q32)*de novo*	Parent of a 46,XY,t(1;13)(p22;q32) phenotypically abnormal child
2.8	46,XY,t(11;21)(p15.4;p12)*de novo*	Parent of a phenotypically abnormal child presenting a paternal dup(11)
2.9	46,XX,t(2;7)(p23.3;p22.3)*de novo*	Prenatal for AMA
2.10	46,XX,t(11;17)(p13;p13.1)mat	Parent of a 46,XY,t(11;17)(p13;p13.1) phenotypically abnormal child
2.11	46,XX,t(7;16)(p15;q22)mat	Recurrent miscarriages
2.12	46,XY,t(8;16)(q22.1;q13)pat	Sibling of a 46,XX,t(8;16)(q22.1;q13) girl ascertained because of delayed puberty
2.13	46,XX,t(16;18)(q24;q21.1)mat	Recurrent miscarriages
2.14	46,XX,t(5;18)(p13;q11)pat	Relative of 46,XX,t(5;18)(p13;q11)pat amniocentesis because of FHD
2.15	46,XY,t(1;11)(q42.3;q21)pat	Parent of 46,XX, t(1;11)(q42.3;q21) amniocentesis because of FHD
2.16	46,XX,t(3;10)(p23;q21.2)pat	Relative of a 46,XY,t(3;10)(p23;q21.2) phenotypically abnormal child
3.1	46,XX,t(4;6)(q27;p25)*de novo*	Recurrent miscarriages

Table 2.I. Karyotypes and mode of ascertainment in the control cohort (Cont.)

Case[1]	Karyotype	Ascertainment
3.2	46,XX,t(8;12)(p23.1;p13.1)*de novo*	Parent of a 46,XY, add(8)(p23.1) phenotypically abnormal child
3.3	46,XX,t(6;9)(q22.2;p22.3)*de novo*	Recurrent miscarriages
3.4	46,XY,t(2;4)(p23;p12)*de novo*	Recurrent miscarriages
3.5	46,X,t(X;22)(p11.23;q13.1)*de novo*	Parent of a 46,X,t(X;22)(p11.23;q13.1) amniocentesis because of FHD
3.6	46,XX,t(11;15)(q23;q22)*de novo*	Parent of a 46,XY,t(11;15)(q23;q22) amniocentesis because of AMA
3.7	46,XX,t(2;6)(q32.2;p23)*de novo*	Recurrent miscarriages
3.8	46,XX,t(6;22)(p21.3;q13)pat	Parent of a 46,XX,del(15)(q11q12) phenotypically abnormal child
3.9	46,XX,t(1;19)(q42.13;p13.2)mat	Recurrent miscarriages
3.10	46,XX,t(9;20)(p24.1;p11.2?3)mat	Parent of a 46,XY,der(9)t(9;20)(p24.1;p11.2?3) phenotypically abnormal child
3.11	46,XY,t(2;3)(p23.1;q29)mat	Relative of a 46,XX,der(3)t(2;3)(p23.1;q29) phenotypically abnormal child

[1]Prefix 1 refers to cases with no genes at the breakpoints. Prefix 2 indicates cases with obligatory breakpoint-mediated gene disruption and prefix 3 refers to cases with potential breakpoint-mediated gene disruption (see section 4.2).

AMA Advanced Maternal Age

FHD Family History of Down syndrome

Table 2.II. Karyotypes and mode of ascertainment in the patient cohort

Case	Karyotype	Ascertainment
16	46,XX,t(10;22)(q24.3;q13.31)*de novo*	DD, epilepsy
20	46,XX,t(2;5)(q33;q12)*de novo*	DD, mild mental retardation
43	46,XY,t(4;17)(q35.1;q25.1)*de novo*	Truncus arteriosus, NAA
44	46,XY,inv(6)(p24q16.2)*de novo*	DD
45	46,X,t(X;19)(q21;p13.11)*de novo*	Premature ovarian failure, NAA
48	46,XY,t(4;6)(q33;q22.2)*de novo*	DD, dysmorphic features
49	46,XX,t(2;10)(q33;q21.2)*de novo*	DD, cleft palate, behavioural abnormalities
50	46,X,t(X;8)(q22.1;q24.13)nk	Premature ovarian failure, NAA
51	47,XX,t(4;20)(p15.2;p11.23)*de novo*,+mar mat	DD, autistic spectrum disorder
52	46,XX inv ins (11;4)(q22.2;q13.2q21.3)*de novo*	DD, LD, short stature, scoliosis
53	46,XX,t(4;8)(q21.1;p12)*de novo*	DD, regressive skills
54	46,XY,t(14;15)(q23;q26.3)*de novo*	dysmorphic features, coarctation of the aorta
55	46,XY,t(19;20)(q13.43;q11.1)*de novo*	Severe oligospermia, NAA
56	46,XY,t(6;21)(q16.2;q11.2)*de novo*	Severe oligospermia, NAA
57	46,XY,t(2;5)(p23;q11.2)*de novo*,t(18;22)(q11.2;p13)*de novo*	DD, short stature, macrocephaly, epilepsy
58	46,XX,t(6;18)(p23;q22)*de novo*	Premature ovarian failure, NAA

DD Developmental Delay
LD Learning Difficulties
NAA No Additional Abnormalities
nk parental origin not known

3 MATERIALS AND METHODS

3.1 Overview

This chapter describes the methods used in the study of the chromosome rearrangements, first for the control cohort (results in chapter 4) and then for the patient cohort (results in chapter 5).

3.2 Preparation of specimens for analysis of both cohorts

Peripheral blood samples provided from each subject and both parents (whenever available) were used in the preparation of phytohaemagglutinin (PHA) stimulated lymphocyte cultures by standard methods (Webber and Garson, 1983), which were used in the preparation of chromosomes for FISH studies. Lymphoblastoid cell lines (LCLs) were also prepared from each study subject by Epstein-Barr virus transformation according to standard methods (adapted from Walls and Crawford, 1987). LCLs were established either at the Wessex Regional Genetics Laboratory (WRGL) (n=31) or at the European Cell and Culture Collection (n=15) and all LCLs were deposited in the cell bank at the WRGL (cell line identification numbers are listed in Appendix 4). LCLs were mainly used for the preparation of chromosomes for FISH studies and array painting analysis. Genomic DNA for array CGH and PCR (including microsatellite analysis) was extracted directly from blood samples and/or from LCLs by a standard salt extraction method (adapted from Miller *et al.*, 1988). For case 58, genomic DNA for array CGH was extracted from a cell suspension in 3:1 methanol:acetic acid (prepared from PHA stimulated cultures) using the Versagene Blood kit according to the supplier's instructions (Gentra Systems, USA).

3.3 Molecular studies of the control cohort

3.3.1 Breakpoint mapping by fluorescence in situ hybridisation

The breakpoint regions were initially mapped to intervals of ~5 Mb by cytogenetic analysis at 550-700 bands as part of the cytogenetics diagnostic service at the WRGL. However, this represents a large interval particularly for determining the gene content of the breakpoints as many human genes have sizes well below this level of resolution, with the mean gene size estimated as 27 kb, but with numerous genes being >100 kb (Lander *et al.*, 2001). Thus, in order to refine the breakpoints, molecular cytogenetic techniques were utilised, i.e. FISH with BAC/PAC clones (average size of ~170 kb). The breakpoints in case 2.9 were mapped both by FISH and by array painting (see section 3.4.1) with the Sanger 1 Mb array as described by Fiegler *et al.* (2003a).

The FISH strategy consisted in selecting BAC/PAC clones mapped both to the bands of the cytogenetically determined breakpoints and to the immediately adjacent bands (to take into account the limits of light microscopy). Probes were applied in rounds of FISH experiments, by "walking" along the chromosome until the breakpoint was mapped within a single BAC/PAC clone, which was "split" by the breakpoint and subsequently would show FISH signals on both translocation derivatives. At this stage, the gene content of the breakpoints was determined with the Ensembl and the UCSC human genome browsers (see Appendix 1 for URLs and section 3.6 for browsers' assembly details). Further refining of the breakpoints, by applying smaller FISH probes derived from fosmid clones (resolution of ~40 kb) was undertaken for regions found to map within genes in an attempt to obtain accurate estimates of the likelihood of breakpoint-mediated gene disruption.

3.3.1.1 Methodology

BAC/PAC clones in the regions of interest were selected mainly from the Ensembl browser, but also from the UCSC browser if no probes were available in Ensembl. BAC/PAC clones from the Ensembl clone sets were grown from glycerol stocks stored at the WRGL, which were kindly provided by the Wellcome Trust Sanger Institute (WTSI). Additional probes were kindly supplied by Prof. Mariano Rocchi from the Department of Genetics and Microbiology, University of Bari, or acquired from the Children's Hospital Oakland Research Institute or Invitrogen. All fosmid clones were selected from the UCSC browser and were kindly provided by the Archives Team at the WTSI. Details of all probes used in this study can be obtained from the probe providers' web resources (URLs listed in Appendix 1).

FISH analysis was undertaken according to the method described by Pinkel *et al.* (1988), with slight modifications. Briefly, the DNA was propagated by bacterial growth and extracted by alkaline lysis as previously described (Sambrook *et al.*, 1989). DNA was indirectly labelled with digoxigenin-11-dUTP (Roche, Germany) or biotin-16-dUTP (Roche, Germany) by nick translation according to a protocol provided by the WTSI (URL in Appendix 1). Labelled probe DNA was co-precipitated with carrier DNA (Sigma, USA), Cot-1 DNA (Roche, Germany) and a centromeric probe (provided by the Probe Laboratory, WRGL) to enable the correct identification of the chromosomes of interest. Probe DNA was resuspended in complete hybridisation mixture (50% deionised formamide, 2xSSC, 10% dextran sulphate) and hybridised to metaphase chromosomes. Target and probe DNA were co-denatured on a hot plate at 72°C for 5 min and hybridised overnight at 37°C. Post-hybridisation washes were as follows: once in 2xSSC at room temperature for 2 min; twice in 1xSSC/50% formamide at 42°C for 5 min and once in 4xSSC/Tween 20, at room temperature for 2 min. Biotin-labelled probes were detected with Avidin-FITC (Avidin-fluorescein isothiocyanate) (Vector Laboratories, USA) and digoxigenin-labelled probes with sheep anti-Dig-TRITC (Anti-digoxigenin-tetramethyl-rhodamine isothiocyanate) (Vector Laboratories, USA). Slides were mounted in 4',6-diamino-2-phenylindole (DAPI) Anti-Fade and were

analysed with a Carl Zeiss Axiophot epifluorescence microscope. Images were captured using a cooled charged-coupled device camera initially with MacProbe software v4.3 (Applied Imaging, USA), which was replaced in later analyses with Isis software v5.0.9 (MetaSystems, Germany).

3.3.2 High resolution breakpoint mapping by array painting

Cases 2.9, 2.10, 2.12, 2.13 and 3.8 were randomly selected for further studies. The main aim of these analyses was to apply and evaluate the potential of array painting (section 3.4.1) applied to high resolution microarrays for the rapid mapping of translocation breakpoints and sequence analysis as described by Gribble *et al.* (2007). Briefly, these analyses involved the construction of custom designed arrays in which the targets on the array were initially fosmid clones mapped within the breakpoint region previously defined by BAC/PAC FISH. In a second experiment, a custom designed array with oligonucleotide probes was used. Custom arrays were designed and printed at the WTSI microarray facility (Cambridge, UK). Case 2.10 was also analysed with commercial arrays (NimbleGen Systems Inc, USA) as described by Gribble *et al.* (2007). Due to time constraints, sequence analysis was only completed for case 2.10 (Gribble *et al.*, 2007), but in the remaining cases the breakpoints were refined to genomic intervals between 1 and 20 kb (results in section 4.3).

3.3.3 Whole genome analysis by 1 Mb array CGH

Array CGH experiments were performed using the Sanger 1Mb array platform, which contains ~3,500 BAC/PAC clones and has a resolution of 1 clone/Mb (Fiegler *et al.*, 2003b). Experiments were performed as described by Fiegler *et al.* (2003b), with slight modifications. Briefly, test and reference genomic DNA were labelled in Cy3 or Cy5 dyes with the BioPrime labelling kit (Invitrogen, UK), using a modified nucleotide mix (0.5 mM dCTP, 2 mM dATP, 2 mM dGTP, 2 mM dTTP in TE buffer). In a final reaction volume of 150 µL, 150 ng

of DNA and 60 μL 2.5x random primers solution were denatured at 100°C for 10 min. After adding 15 μL 10x dNTPmix, 1.5 μL 1mM Cy3 or Cy5 labelled dCTP (NEN Life Science Products, USA) and 3 μL of Klenow fragment, samples were incubated at 37°C overnight. The reactions were stopped by adding 15 μL of stop buffer and unincorporated nucleotides were removed with Microcon YM-30 columns (Millipore, UK) according to the manufacturer's instructions. Probe DNA (160 μL of each test and reference) was co-precipitated in ethanol with 135 μL human Cot-1 DNA and 65 μL 3M NaAc pH 5.2 and ressuspended in 120 μL of hybridisation buffer (50 % formamide, 10 % dextran sulphate, 0.1 % Tween 20, 2x SSC, 10 mM Tris pH 7.4) before hybridised to the arrays.

The samples tested were received over a long period of time and as a consequence the methods of analysis were slightly modified between experiments. For the 13 cases analysed in the earlier stages of this study (cases 1.4, 2.9 to 2.16 and 3.8 to 3.11), array CGH experiments were carried out at the site of the WTSI. Samples were tested in duplicate using dye reversal. The reference DNA consisted of a pool of 20 individuals and sex mismatch between test and reference DNA was not always performed. Hybridisation and post-hybridisation washes were performed manually as described by Fiegler *et al.* (2003b). Slides were scanned on an Axon 4000B scanner (Axon Instruments, Burlingame, CA) and data was analysed with GenePix Pro 3.0 software (Axon Instruments) according to the method in Fiegler *et al.* (2003b). Clones with $\log_2$ fluorescence ratios Cy3/Cy5 of ±0.5 were called as copy number changes.

The remaining 18 individuals were screened at the WRGL after the implementation of a microarray facility at this site. No dye reversal experiments were performed and the hybridisations were carried out against a sex mismatched reference DNA consisting of multiple donors (Promega Corporation, UK). Hybridisation and post-hybridisation washes were performed with a Tecan hybridisation station HS 4800 (Tecan, Austria) according to the program given in Appendix 5. Slides were scanned in a DNA Microarray Scanner (Agilent Technologies, USA) and data analysis was undertaken with BlueFuse software v3.4 (BlueGnome Limited, UK). Data was

normalised by a block median method and data points of low quality were excluded using the following criteria: quality <1 or confidence <0.5 or fluorescence amplitude of Cy3 or Cy5 inferior to 100. Regions of copy number change were called manually, for a $\log_2$ ratio threshold of ±0.5.

3.4 Molecular studies of the patient cohort

3.4.1 Breakpoint mapping by array painting/FISH

Breakpoint mapping studies in cases 44 and 58 were undertaken by FISH using the strategy described above in section 3.3.1. For the remaining 14 patients, the breakpoints were mapped by array painting (AP) with the Sanger Whole-genome tiling path (WGTP) array, which contains >30,000 overlapping BAC/PAC clones (Fiegler *et al.*, 2006). AP was first described by Fiegler *et al.* (2003a) and involves the separation of both derivatives of a translocation by flow sorting. The derivatives are amplified and one derivative chromosome is labelled in Cy3 and the other in Cy5 dyes. The derivatives are combined and competitively hybridised to an array of spotted clones. The subsequent analysis of the fluorescence ratios of Cy3/Cy5 for each clone spotted on the array allows the positioning of the arrayed clones to one derivative or the other as fluorescence ratios of a given spotted clone would either be high or low depending on which derivative contains the sequences represented in that clone. Furthermore, a transition of high to low ratios or vice versa reflects a translocation of sequences from one derivative to the other. The breakpoint spanning clones are expected to hybridise to both derivatives giving intermediate fluorescence ratios and once these are identified, the results can be rapidly confirmed by targeted FISH.

3.4.1.1 Methodology

AP was performed as described by Fiegler *et al.* (2003a). Briefly, chromosomes were prepared from LCLs and labelled with Hoechst 33258 dye and Chromomycin A3 dyes (Carter, 1994). The Hoechst dye has an affinity for AT bases whereas Chromomycin binds preferably to GC and consequently, the different base composition and DNA content of human chromosomes enables their separation on analysis of a bivariate flow karyotype (Langlois *et al.*, 1982). Bivariate karyotypes were analysed and the derivative chromosomes were selected for flow sorting with a MoFlo sorter (DAKO, USA). For cases 20, 45, 49 and 57, the derivative chromosomes were not unequivocally identified on the bivariate karyotypes and reverse chromosome painting (Carter *et al.*, 1992) was used. This consists in fluorescently labelling candidate chromosomes and hybridising these to normal metaphases as in a standard FISH experiment. By analysis of the resulting painting pattern it is possible to determine the identity of the chromosomes tested. Following the identification of the correct derivatives, these were amplified by degenerated oligonucleotide primed polymerase chain reaction (DOP PCR) and GenomePlex Complete Whole Genome Amplification (Sigma, USA). For DOP PCR, aliquots of 500 chromosomes in 33 µL of sterile water were amplified with the 6MW primer as described by Carter (1994). For the GenomePlex Complete Whole Genome Amplification method, aliquots of 25,000 chromosomes in 25 µL of sterile water were amplified according to the manufacturer's instructions (Sigma, USA). Following amplification, the derivatives were labelled by random priming (as in section 3.3.3) and combined.

Hybridisation and post-hybridisation washes were performed with a Tecan hybridisation station HS 4800 (Tecan, Austria) using the program given in Appendix 5. Slides were scanned in a DNA Microarray Scanner (Agilent Technologies, USA). Data was analysed with BlueFuse software v3.4 (BlueGnome Limited, UK), after normalisation by a global median method and exclusion of low quality data points (confidence <0.1 or fluorescence amplitude of Cy3 or Cy5 inferior to 200). Finally, AP results were validated by FISH (see section 3.3.1.1).

3.4.1.2 High resolution array painting with the 244K array

Higher resolution AP was undertaken in case 49 as it was of interest to map the breakpoint regions further (results in section 5.2.1.1.). AP was performed with the Agilent 244K array platform, which contains ~236,000 oligonucleotides, each 60-mer in length, with a resolution of 6.4 kb. DOP PCR amplified derivatives were labelled by random priming and hybridised to the array according to the manufacturer's instructions (Agilent Technologies, USA). The microarray slide was scanned with an Agilent microarray scanner and data analysis was carried out with the Agilent CGHAnalytics (v3.4) microarray software using the z-score algorithm.

3.4.2 Breakpoint mapping by PCR

In cases 49 and 51, the breakpoints were mapped to intervals containing candidate disease-causative genes (results in sections 5.2.1 and 5.2.2). In order to further refine the breakpoint regions, a PCR-based approach was developed as described by McMullan *et al.* (2002). This consisted of designing primer pairs to amplify products from a chromosome of interest. By testing a given primer pair in both derivatives of a translocation, the presence of a product from one derivative or the other would determine the position of the breakpoint in relation to the primers tested. Primers were designed to "walk" towards the breakpoint until the breakpoint containing region was progressively refined.

3.4.2.1 Methodology

The sequence of the breakpoint containing intervals as defined by AP with the 244K Agilent array (case 49) or by fosmid FISH (case 51) was obtained from the UCSC browser. Sequences were repeat-masked and imported to the Primer3 program v.0.4.0 (URL in Appendix 1). Primers were designed to amplify products of 100-300 bps and were spaced by ~2 kb whenever the

characteristics of the sequence, i.e. absence of repeat elements, made this possible. Primers were synthesised and provided by Sigma Genosys (Sigma, USA) and are listed in Appendix 6. Primers were initially tested on normal genomic DNA to ensure these were working. PCR experiments were carried out by standard methods with the HotStar Taq DNA polymerase (Qiagen, UK) which was used to amplify derivative DNA alongside a positive control (patient's genomic DNA) and a negative control (no template). PCR products were visualised in a 1% agarose gel alongside the 1 kb plus DNA ladder (Invitrogen, UK).

3.4.3 Long range PCR and sequence analysis

For case 49, who had a translocation between chromosomes 2 and 10, sequence analysis of the breakpoint junctions was attempted. The strategy utilised to amplify the chromosome 2 breakpoint junction consisted of designing a left primer mapped to chromosome 2 and a right primer mapped to chromosome 10. Conversely, for amplification of the chromosome 10 breakpoint junction, left primers were designed from chromosome 10 and right primers from chromosome 2. This study design ensures that a PCR product will only be obtained from the translocated chromosomes when using patient's genomic DNA as PCR template. Since the distance between the left and right primers could not be predicted, a long range PCR protocol, which enables the amplification of products of 10-30 kb, was utilised. Primers for long range PCR were designed with the Primer3 software and are listed in Appendix 6. Several combinations of left and right primers were tested by long range PCR with the Expand Long Template Kit (Roche, Germany) according to the supplier's protocol. PCR products were visualised in a 1% agarose gel alongside the 1 kb plus DNA ladder (Invitrogen, UK).

Following amplification of breakpoint junction products, these were treated with ExoSap nuclease I (BioLabs, UK) for removal of long range PCR primers. Sequencing reactions were set up with the Big Dye Terminator v3.1 (Applied

Biosystems, USA) by mixing: 6.5 μL of nuclease-free water, 1 μL Big Dye v3.1, 1 μL Big Dye Buffer, 0.5 μl of the forward or reverse primer (10 μM) and 1 μL of the product to be sequenced. Reactions were run in a thermocycler under the following conditions: Step 1: 96°C, 30s; Step 2: 50°C, 15s; Step 3: 60°C, 2 min; Step 4: Go to step 1, 24 times; Step 5: 15°C, hold. PCR products were cleaned up with the Montage Seq96 kit (Millipore, UK) according to the manufacturer's instructions and analysed in an ABI 3130 GeneAnalyzer (Applied Biosystems, USA). The resulting sequence was positioned on the human genome by using the Human BLAT search tool from the UCSC browser.

3.4.4 Whole genome analysis by array CGH

Array CGH analysis in cases 44 and 58 was undertaken with the Agilent 244K array (Agilent technologies) following the manufacturer's instructions (Agilent Technologies). Patient DNA was hybridised against a sex mismatched reference DNA from multiple donors (Promega Corporation, UK) and data analysis was undertaken using the Agilent CGHAnalytics (v3.4) microarray software. Regions of copy number change were called when including at least 3 consecutive oligonucleotides showing $\log_2$ fluorescence ratios Cy3/Cy5 of ±0.5.

For the remaining patients, array CGH was performed with the Sanger WGTP BAC array using the method described by Fiegler *et al.* (2003b) with slight modifications as detailed in section 3.3.3. All experiments were performed in duplicate using dye reversal. The reference DNA used consisted in genomic DNA from a single male individual. Slides were scanned with a DNA Microarray Scanner (Agilent Technologies, USA) and the replicate data from dye reversal experiments were combined with a custom Perl script developed by the WTSI. The format of data output was an Excel spreadsheet designed by the WTSI. Regions of copy number change were analysed with the CNVfinder algorithm described by Fiegler *et al.* (2006). Regions of copy number change

were called when including two consecutive clones with $\log_2$ fluorescence ratios Cy3/Cy5 of ±0.5.

3.5 Validation and parental origin of array CGH imbalances

Array CGH imbalances not documented in the Database of Genomic Variants (DGV) and including at least two BAC/PAC clones were validated by FISH (section 3.3.1.1) prior to parent of origin studies by microsatellite analysis and/or FISH. This type of imbalance was detected in cases 2.11 (see section 4.6) and 20, 52, 53 and 57 (see section 5.5.1). Microsatellite analysis involved initially the selection of microsatellite markers (listed in Appendix 7) from the Ensembl or UCSC browsers. Sequence information of primers to amplify the markers selected is available within these genome browsers. Primers were synthesised by Thermo Electron (Thermo Fisher Scientific, Germany), and the left primer was fluorescently labelled. Genomic DNA from the proband and available family members (where appropriate) was amplified by standard PCR methods and amplified products were analysed on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, USA).

3.6 Human genome browsers and data analysis

Since the beginning of this project, different human genome assemblies based on the NCBI builds 34, 35 and 36.1 have been made available. As a consequence of the databases' updating, some of the initial mapping information gathered on FISH probes, genes or primers has been modified. However, the results presented in the next chapters have been reviewed and reflect the most recent human genome assembly at the time of writing: NCBI build 36.1 accessed in December 2007.

Most BAC/PAC clones were selected from the Ensembl clone sets but a minority of BAC/PAC clones and all fosmids used in this study were selected

from the UCSC browser. As a result, both browsers were used for the analysis of the results, showing that there are gene mapping and clone mapping discrepancies between the browsers. This is because although both browsers display data based on the same version of the human reference genome sequence, different algorithms are utilised to analyse the sequence and to assign gene start and end positions. These dissimilarities will be clearly marked in the results (chapters 4 and 5) as and where they occur.

3.7 Overall summary of the methods

Since a range of methods were utilised for the characterisation of the different rearrangements, a summary of techniques is provided in tables 3.I (for the control cohort) and 3.II (for the patient cohort).

Table 3.I. Methods used in the study of the ABCRs in the control cohort.

	Breakpoint mapping			Array CGH
Case	**FISH**	**1 Mb AP**	**HR-AP**	**Sanger 1 Mb**
1.1	+	-	-	+
1.2	+	-	-	+
1.3	+	-	-	+
1.4	+	-	-	+
2.1	+	-	-	+
2.2	+	-	-	+
2.3	+	-	-	+
2.4	+	-	-	+
2.5	+	-	-	+
2.6	+	-	-	+
2.7	+	-	-	+
2.8	+	-	-	+
2.9	+	+	+	+
2.10	+	-	+	+
2.11	+	-	-	+
2.12	+	-	+	+
2.13	+	-	+	+
2.14	+	-	-	+
2.15	+	-	-	+
2.16	+	-	-	+
3.1	+	-	-	+
3.2	+	-	-	+
3.3	+	-	-	+
3.4	+	-	-	+
3.5	+	-	-	+
3.6	+	-	-	+
3.7	+	-	-	+
3.8	+	-	+	+
3.9	+	-	-	+
3.10	+	-	-	+
3.11	+	-	-	+

HR High resolution
AP Array painting

Table 3.II. Methods used in the study of the ABCRs in the patient cohort.

	Breakpoint mapping				Array CGH	
Case	**FISH**	**WGTP-AP**	**244K-AP**	**PCR**	**WGTP**	**Agilent 244K**
16	+	+	-	-	+	-
20	+	+	-	-	+	-
43	+	+	-	-	+	-
44	+	-	-	-	-	+
45	+	+	-	-	+	-
48	+	+	-	-	+	-
49	+	+	+	+	+	-
50	+	+	-	-	+	-
51	+	+	-	+	+	-
52	+	+	-	-	+	-
53	+	+	-	-	+	-
54	+	+	-	-	+	-
55	+	+	-	-	+	-
56	+	+	-	-	+	-
57	+	+	-	-	+	-
58	+	-	-	-	-	+

WGTP Whole Genome Tiling Path
AP Array Painting

4 MOLECULAR CYTOGENETIC ANALYSES IN A CONTROL COHORT OF 31 PHENOTYPICALLY NORMAL INDIVIDUALS

4.1 Overview

This chapter presents the results of the analysis of ABCRs in 31 phenotypically normal individuals (clinical descriptions in Appendix 2). As stated in chapter 1, the translocations in this control cohort were analysed with the aim of determining the frequency of the four main features known to be associated with clinical abnormalities in carriers of ABCRs. Firstly, the presence of breakpoint-mediated gene disruption was ascertained by consulting the human reference genome data available in human genome browsers. Secondly, the presence of cryptic deletions and duplications at the breakpoint regions was evaluated by examining whether FISH probes mapped near the breakpoints were present in two copies as expected in normal individuals. Thirdly, the breakpoint regions were screened for additional chromosomal complexity. Chromosomal complexity refers to the presence of breakpoints in addition to the two expected from a simple reciprocal translocation or inversion suggesting the presence of additional chromosomal rearrangements such as deletions, duplications, insertions or inversions near the breakpoint regions. Finally, array CGH analysis was utilised to characterise the whole genome of each individual to determine if genomic imbalances unrelated to the breakpoints were present. Ultimately, the results of these analyses were compared to similar studies of ABCRs in 16 phenotypically abnormal patients (see chapters 5 and 6).

4.2 Breakpoint mapping results

Breakpoint mapping studies were undertaken by FISH as described in chapter 3 (section 3.3.1). Results for each study subject are described below and summarised in tables 4.III, 4.IV and 4.V. The identity of the clones spanning or flanking a breakpoint region is given together with the cytogenetic band to which the breakpoints were mapped. In most cases the FISH breakpoint was not mapped to the same cytogenetic band as the cytogenetically determined breakpoint (details in Table 4.II) and these revisions of the breakpoints are shown using an adaptation of the ISCN (2005) nomenclature. Each

translocation was given an identification number, in which the first digit is 1, 2 or 3, reflecting the clustering of the translocations in three groups based on the gene content of the breakpoint regions. Briefly, translocations in group 1 have both breakpoints mapped to regions containing no known genes. Translocations in group 2 have obligatory breakpoint-mediated gene disruption by at least one breakpoint. Finally, translocations in group 3 show potential gene disruption by at least one of the breakpoints. The likelihood of gene disruption was given by the ratio of the length of a gene to the length of the breakpoint-containing interval.

The FISH mapping results for cases 3.8 and 3.11 are presented in detail because, unlike the remaining 29 translocations, their results suggest that they might not be straight-forward reciprocal translocations. Case 3.8 showed inconsistent FISH results that could be due to additional chromosomal complexity, whereas case 3.11 showed FISH results consistent with a non reciprocal translocation.

4.2.1 Translocations in group 1

Group 1 included 4 apparently balanced reciprocal translocations, for which no known genes mapped to either of the breakpoint regions.

Case 1.1: 46,XY,t(2;14)(p21;q13)*de novo*.ish t(2;14)(p21;q21.1)

This male individual was referred at 29 years of age for routine parent of origin analysis because his partner had a miscarriage of a fetus with an apparently balanced translocation which was shown to be paternally inherited. The 2p21 breakpoint was mapped to an 183 kb region within BAC RP11-391M15 and the 14q21.1 breakpoint to an 169 kb region within BAC RP11-254B15. No known genes map to either breakpoint.

Case 1.2: **46,XX,t(3;9)(p26.2;p22.3)*de novo*.ish t(3;9)(p26.3;p23)**

This female individual was referred at 32 years of age because her developmentally delayed child was found to have a 46,XY,add(3)(p25.3) karyotype. The 3p26.3 breakpoint was mapped to a 377 kb region proximal to BAC RP11-128A12 and distal to BAC RP11-86C13 and the 9p23 breakpoint to a 461 kb region proximal to BAC RP11-165C19 and distal to BAC RP11-58B8. No known genes map to either breakpoint.

Case 1.3: **46,X,t(X;7)(?q27;q22)*de novo*.ish t(X;7)(q28;q31.1)**

This female individual was referred at 42 years of age, after her mentally retarded child was found to have a 46,X,Xq+ karyotype. The Xq28 breakpoint was mapped to an 181 kb interval within BAC RP11-49C9 and the 7q31.1 breakpoint to an 197 kb region spanned by BAC RP11-540M6. No known genes map to either breakpoint.

Case 1.4: **46,XX,t(10;18)(q24.3;q12.2)mat.ish t(10;18)(q25.1;q12.3)**

This female individual, previously described as case 12 by Baptista *et al.* (2005), was referred for routine testing of relatives of carriers of chromosome abnormalities. The proband of the family was referred for prenatal testing because of increased serum screen risk and found to have an apparently balanced translocation. A cytogenetically identical translocation was found in case 1.4. The 10q25.1 breakpoint mapped to an 166 kb interval within BAC RP11-163F15 and the 18q12.3 to an 164 kb region within BAC RP11-142I20. No known genes map to either breakpoint.

4.2.2 Translocations in group 2

Group 2 included 16 apparently balanced reciprocal translocations, for which breakpoint-mediated gene disruption by at least one of the breakpoint regions is obligatory.

Case 2.1: 46,XX,t(4;16)(q35.1;p13.13)*de novo*.ish t(4;16)(q35.2;p13.2)

This female individual was referred at 32 years of age for parental studies because her child had multiple congenital abnormalities and a 46,XY,-4,+rea(4)(4pter->4q35.1::?) karyotype. The 4q35.2 breakpoint was mapped to a 41 kb region within fosmid WI2-3578O1. This fosmid is fully contained within the *FAT* (FAT tumour suppressor homolog 1-*Drosophila*) gene, thus *FAT* must be disrupted by the breakpoint. The 16p13.2 breakpoint was mapped to an 80 kb region proximal to fosmid WI2-1047H17 and distal to fosmid WI2-3013O19. The gene *ABAT* (4-aminobutyrate aminotransferase) is fully contained within this region and is therefore disrupted by the breakpoint.

Case 2.2: 46,XX,t(1;13)(q32.3;q32.3)*de novo*.ish t(1;13)(q32.2;q32.1)

This female individual was ascertained at 24 years of age, after she miscarried a fetus with congenital abnormalities and a 46,XY,13q+ karyotype. The 1q32.2 breakpoint was mapped to a 43 kb region within fosmid WI2-1778P11. The *PLXNA2* (plexin A2) gene is fully contained within this fosmid, and must be disrupted by the breakpoint. The 13q32.1 breakpoint was mapped to an 83 kb region distal to fosmid WI2-2792L8 and proximal to fosmid WI2-1163M19. This region is completely occupied by the *DNAJC3* (DnaJ -Hsp40-homolog, subfamily C, member 3) gene, which must be disrupted by the breakpoint.

Case 2.3: 46,XY,t(2;18)(q35;q21.3)*de novo*.ish t(2;18)(q34;q21.31)

This male individual was referred at 31 years of age because of recurrent miscarriages in his partner. The 2q34 breakpoint mapped to an 178 kb within BAC RP11-221D24. This region is fully contained within the *SPAG16* (sperm associated antigen 16) gene which must be disrupted by the breakpoint. The 18q21.31 breakpoint was mapped to a 77 kb region distal to fosmid WI2-1250J9 and proximal to BAC RP11-380K20. Part of this genomic region (20%) contains the *ST8SIA3* (ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 3) gene.

Case 2.4: 46,XX,t(2;9)(q21.3;p13)*de novo*.ish t(2;9)(q22.1;p13.3)

This female individual was referred at 40 years of age for prenatal testing because of advanced maternal age. Her fetus inherited her apparently balanced

translocation. The 2q22.1 breakpoint was mapped to a 92 kb region distal to fosmid WI2-642B17 and proximal to fosmid WI2-1298I22. A portion of this region (75%) contains the *SPOPL* (speckle-type POZ protein-like) gene. The 9p13.3 breakpoint was mapped to a 36 kb region within fosmid WI2-1672N4. This region is fully contained within the *UNC13B* (unc-13 homolog B -*C. elegans*) gene, which must be disrupted by the breakpoint.

Case 2.5: 46,XX,t(7;17)(q36.1;q25.1)*de novo*.ish t(7;17)(q33;q23.2)

This female individual was referred for cytogenetic analysis, at 23 years of age, because of recurrent miscarriages. The 7q33 breakpoint was mapped to a 218 kb region within BAC RP11-322P10. This clone is fully contained within the *EXOC4* (exocyst complex component 4) gene, which must be disrupted by the breakpoint. The 17q23.2 breakpoint was mapped to an 86 kb region distal to fosmid WI2-691B9 and proximal to BAC RP11-435J17. A part of this region (80%) contains the genes *SUPT4H1* (suppressor of Ty 4 homolog 1, *S. cerevisiae*) and *RNF43* (ring finger protein 43).

Case 2.6: 46,XY,t(8;15)(p11.2;q24)*de novo*.ish t(8;15)(p11.21;q24.1)

This male individual, aged 36, was referred after his partner had an amniocentesis because of advanced maternal age. The fetus had inherited the paternal apparently balanced translocation. The 8p11.21 breakpoint was mapped to an 158 kb region proximal to fosmid WI2-3053K15 and distal to BAC RP11-317M6. Part of this region (80%) contains *SLC20A2* (solute carrier family 20 -phosphate transporter-member 2) and *C8orf40* (chromosome 8 open reading frame 40). The 15q24.1 breakpoint was mapped to an 128 kb region distal to BAC RP11-519P13 and proximal to fosmid WI2-1957A19. This region is fully occupied by the *NPTN* (neuroplastin) gene and the predicted gene *LOC283677* (unknown function) therefore, one of these genes must be disrupted by the breakpoint.

Case 2.7: 46,XY,t(1;13)(p22;q32)*de novo*.ish t(1;13)(p21.2;q31.3)

This male individual, aged 37, was referred for parent of origin studies. His developmentally delayed child was found to have inherited the paternal apparently balanced translocation. The 1p21.2 breakpoint was mapped to an

159 kb region within BAC RP11-176N14, which contains no known genes. The 13q31.3 breakpoint mapped to an 168 kb region within BAC RP11-95C14. This BAC is fully contained within the *GPC5* (glypican 5) gene and as a result *GPC5* must be disrupted by the breakpoint.

Case 2.8: 46,XY,t(11;21)(p15.4;p12)*de novo*.ish t(11;21)(p15.4;p12)

This male individual was ascertained at 28 years of age because his child had some characteristics of Beckwith-Wiedemann syndrome and a paternal dup(11). This was later shown to be an unbalanced translocation due to malsegregation of a paternal apparently balanced translocation. The 11p15.4 breakpoint was mapped to a 39 kb interval within fosmid WI2-2397B2. According to the UCSC human genome browser, this fosmid is fully contained within the *ST5* (suppression of tumorigenicity 5) gene and consequently the breakpoint must disrupt *ST5*. However, in the Ensembl human genome browser, this gene occupies only 70% of the breakpoint spanning fosmid. The 21p12 breakpoint mapped to the satellite stalk.

Case 2.9: 46,XX,t(2;7)(p23.3;p22.3)*de novo*.ish t(2;7)(p22.2;p21.3)

This female individual, previously described as case 1 by Baptista *et al.* (2005), was ascertained antenatally because of advanced maternal age. The 2p22.2 breakpoint mapped to an 81 kb interval within BAC RP11-288C18. Part of this region (80%) contains the *STRN* (striatin) gene. The 7p21.3 breakpoint was mapped to an 83 kb interval within PAC RP5-887P4, which contains no known genes. Further mapping of the breakpoints by high resolution array painting (section 4.3) showed that *STRN* was disrupted by the 2p22.2 breakpoint, while the 7p21.3 disrupted the *PHF14* (PHD finger protein 14) gene. However, according to the FISH analysis, the 7p21.3 breakpoint mapped 90 kb centromeric to *PHF14*. The mapping discordance between FISH and high resolution array painting might be explained by mapping alterations between different releases of the reference human genome sequence or to probe mapping errors.

Case 2.10: 46,XX,t(11;17)(p13;p13.1)mat.ish t(11;17)(p12;p12)

This female individual, previously reported as case 2 by Baptista *et al.* (2005), was referred at 33 years of age for parental origin studies. Her child had developmental delay and the maternal apparently balanced translocation. The 11p12 breakpoint was mapped to an 175 kb region within BAC RP11-2G8, which contains no known genes. The 17p12 breakpoint was mapped to a 66 kb region proximal to fosmid WI2-648I13 and distal to fosmid WI2-1659O13, which is fully contained within the *QRICH2* (glutamine rich 2) gene.

Case 2.11: 46,XX,t(7;16)(p15;q22)mat.ish t(7;16)(p15.3;q22.1)

This female individual, previously reported as case 4 by Baptista *et al.* (2005), was referred at 28 years of age because of recurrent miscarriages. The 7p15.3 breakpoint was mapped to a 204 kb region within BAC RP11-842N18, which contains no known genes. The 16q22.1 breakpoint was mapped to an 87 kb interval distal to BAC RP11-419C5 and proximal to fosmid WI2-2601F178. The *WWP2* (WW domain containing E3 ubiquitin protein ligase 2) gene is fully contained within this region, and must therefore be disrupted by the breakpoint.

Case 2.12: 46,XY,t(8;16)(q22.1;q13)pat.ish t(8;16)(q22.2;q21)

This male individual, previously reported as case 5 by Baptista *et al.* (2005), was referred at 9 years of age for routine analysis of relatives of carriers of chromosome abnormalities. The proband of the family was his sibling and presented with puberty delay and a paternally inherited apparently balanced translocation. The 8q22.2 breakpoint was mapped to a 42 kb interval within fosmid WI2-1991H16. Part of this (96%) contains the *STK3* (serine/threonine kinase 3- STE20 homolog, yeast) and *OSR2* (odd-skipped related 2,Drosophila) genes, according to the UCSC browser, whereas in Ensembl, 18% of this region contains *OSR2*. High resolution array painting (section 4.3) refined further this breakpoint region to within *STK3*, which must be disrupted according to the UCSC browser. However, in the Ensembl browser no known genes map to this region and *STK3* is located ~100 kb centromeric to the breakpoint. The 16q21 breakpoint was mapped to a 208 kb region within BAC RP11-486F10, which contains no known genes.

Case 2.13: 46,XX,t(16;18)(q24;q21.1)mat.ish t(16;18)(q24.1;q21.1)

This female individual, previously reported as case 7 by Baptista *et al.* (2005), was referred at 24 years of age because of recurrent miscarriages. The 16q24.1 breakpoint mapped to a 68 kb region distal to fosmid WI2-1243C9 and proximal to fosmid WI2-2443E17. The *ATP2C2* (ATPase, Ca++ transporting, type 2C, member 2) gene is fully contained within this region and must therefore by split by the breakpoint. The 18q21.1 breakpoint was mapped to a 59 kb region distal to fosmid WI2-2841H18 and proximal to fosmid WI2-1586M9. Part of this region (90%) contains the *HDHD2* (haloacid dehalogenase-like hydrolase domain containing 2) and the *IER3IP1* (immediate early response 3 interacting protein 1) genes. Further mapping by high resolution array painting (section 4.3) refined the breakpoint to within *HDHD2*.

Case 2.14: 46,XX,t(5;18)(p13;q11)pat.ish t(5;18)(p13.2;q11.1)

This female individual, previously reported as case 9 by Baptista *et al.* (2005), was referred at 30 years of age for routine testing of relatives of carriers of chromosome abnormalities. The proband of the family was ascertained at prenatal testing because of a family history of Down syndrome. The 5p13.2 breakpoint was mapped to a 39 kb region within fosmid WI2-1419B1. This region is fully contained within the *EGFLAM* (EGF-like, fibronectin type III and laminin G domains) gene, which must therefore be disrupted by the breakpoint. The 18q11.1 breakpoint was mapped to a 43 kb region within fosmid WI2-2791H16. Part of this region (25%) contains the *ROCK1* (Rho-associated protein kinase 1) gene.

Case 2.15: 46,XY,t(1;11)(q42.3;q21)pat.ish t(1;11)(q43;q22.1)

This male individual, previously described as case 10 by Baptista *et al.* (2005), was referred at 29 years of age for parent of origin studies. Prenatal testing because of family history of Down syndrome showed his fetus to have the paternally inherited apparently balanced translocation. The 1q43 breakpoint was mapped to a 39 kb region within fosmid WI2-1874N21. This fosmid is fully contained within the *RYR2* (ryanodine receptor 2, cardiac) gene, which must be interrupted by the breakpoint. The 11q22.1 breakpoint was mapped to an 175 kb interval within BAC RP11-99C10. In the UCSC browser, the

CNTN5 (contactin 5) gene is mapped to part (~50%) of this breakpoint-containing region, but not in the Ensembl browser where *CNTN5* is located ~700 kb distal to the breakpoint region.

Case 2.16: 46,XX,t(3;10)(p23;q21.2)pat.ish t(3;10)(p24.2;q22.1)

This female individual, previously described as case 11 by Baptista *et al.* (2005), was referred at 33 years of age for routine testing of relatives of carriers of chromosome abnormalities. The proband of the family had mental retardation and an apparently balanced translocation. The 3p24.2 breakpoint was mapped to a 38 kb interval within fosmid WI2-3033M15, which contains no known genes. The 10q22.1 breakpoint was mapped to a 44 kb region within fosmid WI2- 2217B17. This fosmid is fully contained within the *COL13A1* (alpha 1 type XIII collagen isoform 1) gene, which must be disrupted by the breakpoint.

4.2.3 Translocations in group 3

Group 3 included 11 apparently balanced reciprocal translocations, for which breakpoint-mediated gene disruption by at least one of the breakpoint regions was a possibility.

Case 3.1: 46,XX,t(4;6)(q27;p25)*de novo*.ish t(4;6)(q28.1;p24.3)

This female individual was referred for cytogenetic analysis at 25 years of age because of recurrent miscarriages. The 4q28.1 breakpoint was mapped to an 180 kb region distal to BAC RP11-396E2 and proximal to fosmid WI2- 887N9. Part of this region (15%) contains the *ANKRD50* (ankyrin repeat domain 50) gene. The 6p24.3 breakpoint was mapped to an 85 kb region proximal to fosmid WI2-737K2 and distal to fosmid WI2-3350C15. Part of this interval (45%) contains the *CAGE1* (cancer antigen 1) and *RIOK1* (RIO kinase 1 -yeast) genes.

Case 3.2: 46,XX,t(8;12)(p23.1;p13.1)*de novo*.ish t(8;12)(p23.1;p13.31)

This female individual was ascertained at 31 years of age because she had a developmentally delayed child with a 46,XY, add(8)(p23.1) karyotype. The 8p23.1 breakpoint mapped to a 334 kb region proximal to BAC RP11-5E15 and distal to BAC RP11-151L20. Part of this region (10%) contained six genes: *DEFB1* (defensin, beta 1), *DEFA6* (defensin, alpha 6, Paneth cell-specific), *DEFA4* (defensin, alpha 4, corticostatin), *DEFA1* (defensin, alpha 1), *DEFA3* (defensin, alpha 3, neutrophil-specific) and *DEFA5* (defensin, alpha 5, Paneth cell-specific). The 12p13.31 breakpoint was mapped to a 323 kb region proximal to BAC RP11-69M1 and distal to BAC RP11-266K4. Part of this interval (25 or 76% in UCSC or Ensembl, respectively) contains six genes: *FOXJ2* (Forkhead box protein J2 -Fork head homologous X), *C3AR* (C3a anaphylatoxin chemotactic receptor), *NECAP1* (adaptin ear-binding coat-associated protein 1), *CLEC4A* (C-type lectin domain family 4 member A), *ZNF705A* (zinc finger protein 705A) and *FAM90A1* (family with sequence similarity 90, member A1). Further refining of these breakpoint regions was hindered by the presence of sequences of shared homology between chromosomes 8 and 12, with caused the FISH probes mapped to the vicinity of the breakpoints to cross hybridise to both chromosomes.

Case 3.3: 46,XX,t(6;9)(q22.2;p22.3)*de novo*.ish t(6;9)(q22.31;p22.3)

This female individual was referred at 32 years of age because of recurrent miscarriages. The 6q22.31 breakpoint was mapped to an 87 kb region distal to fosmid WI2-743A11 and proximal to fosmid WI2-960K23. Part of this region (40 or 65% in UCS or Ensembl, respectively) contains the *ASF1* (ASF1 anti-silencing function 1 homolog A -*S. cerevisiae*) and *MCM9* (minichromosome maintenance complex component 9) genes. The 9p22.3 breakpoint was mapped to a 278 kb region proximal to BAC RP11-54D18 and distal to BAC RP11-271D19. Part of this interval (10%) contains the *C9orf52* (chromosome 9 open reading frame 52) gene.

Case 3.4: 46,XY,t(2;4)(p23;p12)*de novo*.ish t(2;4)(p23.3;p12)

This male individual was referred at 30 years of age because his partner had recurrent miscarriages. The 2p23.3 breakpoint was mapped to a 39 kb region

within fosmid WI2-3158P17. Part of this fosmid (20%) contains the *FAM59B* (family with sequence similarity 59, member B) gene. The 4p12 breakpoint was mapped to an 87 kb segment proximal to fosmid WI2-2624O16 and distal to fosmid WI2-894F19. Part of this region (55%) contains the *FRYL* (FRY-like) gene.

Case 3.5: 46,X,t(X;22)(p11.23;q13.1)*de novo*.ish t(X;22)(p11.1;q11.21)

This female individual was referred at 28 years of age for an amniocentesis because of family history of Down syndrome. Her fetus had the maternally inherited apparently balanced translocation. The Xp11.1 breakpoint was mapped to a 43 kb region within fosmid WI2-3044A1. Part of this region (10%) contains the *SPIN2B* (spindlin family, member 2B) and *SPIN2A* (spindlin family, member 2A) genes. The 22q11.21 breakpoint was mapped to an 80 kb region distal to BAC b562F10 and proximal to fosmid WI2-1379B7. Part of this genomic region (35 or 75% in UCSC or Ensembl, respectively) contains the *USP41* (ubiquitin specific peptidase 41) and *ZNF74* (zinc finger protein 74) genes.

Case 3.6: 46,XX,t(11;15)(q23;q22)*de novo*.ish t(11;15)(q23.3;q21.2)

This female individual was referred at 37 years of age for an amniocentesis because of advanced maternal age. Her fetus had the maternally inherited apparently balanced translocation. The 11q23.3 breakpoint was mapped to an 162 kb region within BAC RP11-56K19, which contains no known genes. The 15q21.2 breakpoint was mapped to an 84 kb region distal to fosmid WI2-426O12 and proximal to fosmid WI2-2248H19. Part of this region (75%) contains the *HDC* (histidine decarboxylase) and *GABPB2* (GA binding protein transcription factor, beta subunit 2) genes.

Case 3.7: 46,XX,t(2;6)(q32.2;p23)*de novo*.ish t(2;6)(q33.1;p25.1)

This female individual was referred for cytogenetic analysis at 22 years of age because of recurrent miscarriages. The 2q33.1 breakpoint was mapped to a 39 kb interval within fosmid WI2-2910N4. Part of this fosmid (95%) contains the *FLJ38973* (predicted gene of unknown function) and *C2orf60* (chromosome 2

open reading frame 60) genes. The 6p25.1 breakpoint was mapped to an 103 kb region within PAC RP3-470L22, which contains no known genes.

Case 3.8: 46,XX,t(6;22)(p21.3;q13)pat

This case is presented in detail in section 4.2.4.

Case 3.9: 46,XX,t(1;19)(q42.13;p13.2)mat.ish t(1;19)(q42.12;p13.12)

This female individual, previously described as case 6 by Baptista *et al.* (2005), was referred at 26 years of age because of recurrent miscarriages. The 1q42.12 breakpoint was mapped to an 126 kb region distal to BAC RP11-757H9 and proximal to fosmid WI2-2745B10. Part of this region (20%) contains the *LBR* (lamin B receptor) gene. The 19p13.12 breakpoint was mapped to a 40 kb interval within fosmid WI2-1278H6. Part of this fosmid (60%) contains the *RLN3* (relaxin 3) and *IL27RA* (interleukin 27 receptor, alpha) genes.

Case 3.10: 46,XX,t(9;20)(p24.1;p11.2?3)mat.ish t(9;20)(p24.1;p11.21)

This female individual, previously described as case 8 by Baptista *et al.* (2005), was referred at 19 years of age for parent of origin studies after her clinically abnormal child was found to have a 46,XY,der(9)t(9;20)(p24.1;p11.2?3) karyotype. The 9p24.1 breakpoint was mapped to an 174 kb region within BAC RP11-472F14. Part of this BAC (85%) contains the *UHRF2* (Np95-like ring finger protein isoform a) and *GLDC* (glycine dehydrogenase -decarboxylating) genes. The 20p11.21 breakpoint was mapped to a 41 kb region defined by fosmid WI2-831H20. Part of this fosmid (15%) contains the *CD93* (CD93 molecule) gene.

Case 3.11: 46,XY,t(2;3)(p23.1;q29)mat

This case is presented in detail in section 4.2.5.

4.2.4 FISH results for case 3.8: 46,XX,t(6;22)(p21.3;q13)pat.ish t(6;22)(p21.1;q12.3)

This female individual, previously described as case 3 by Baptista *et al.* (2005), was referred for cytogenetic analysis at 31 years of age for parental origin studies after an abnormal karyotype was found in her child. The child had developmental delay and a clinical diagnosis of Angelman syndrome was confirmed by demonstrating that she had a 46,XX,del(15)(q11.2q12)*de novo* chromosome complement. When case 3.8 was analysed, an abnormal karyotype unrelated to that of her child was identified; an apparently balanced translocation, t(6;22)(p21.3;q13) inherited from her clinically normal father. The 6p21.1 breakpoint mapped to an 179 kb interval defined by BAC RP11-249P15, which contains no known genes (see figure 4.1). The 22q12.3 breakpoint mapped to within a 97 kb region defined by PAC RP1-127L4. Part of this region (25%) contained *NP_001010859.1* (predicted gene) and *RFPL2* (ret finger protein-like 2) (figure 4.2).

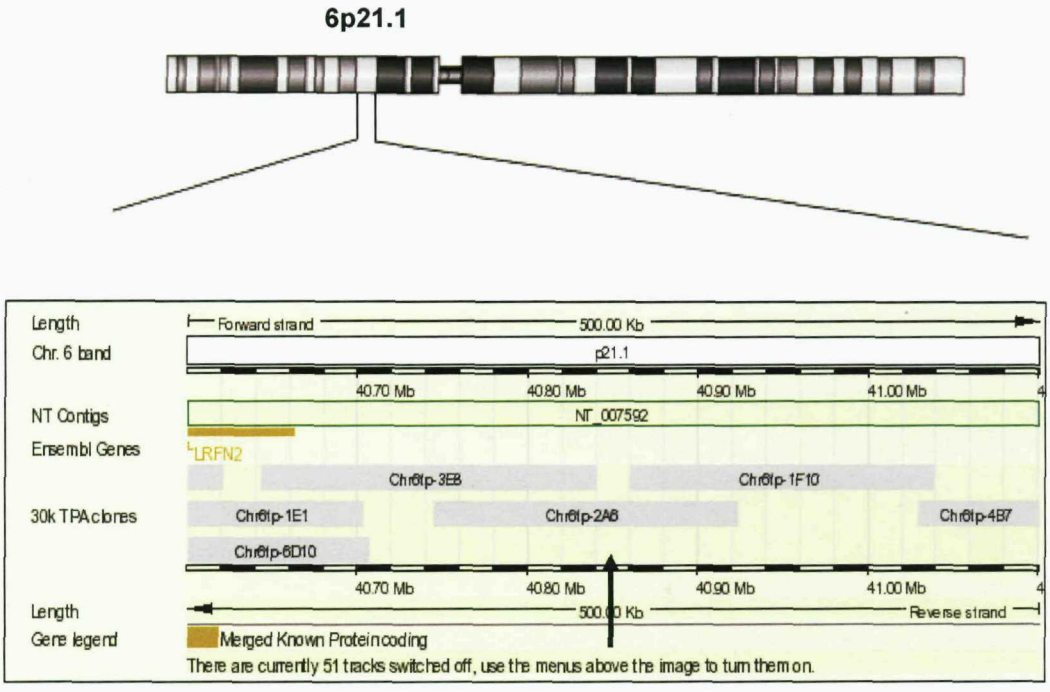

Fig.4.1. Screen shot from the Ensembl browser showing the 6p21.1 breakpoint region in case 3.8.

BAC RP11-249P15, indicated by the arrow, was split by the breakpoint. Note that no known genes co-localise to this BAC, and the nearest gene, LRFN2 (leucine rich repeat and fibronectin type III domain containing 2) is ~110 kb telomeric to the breakpoint spanning clone.

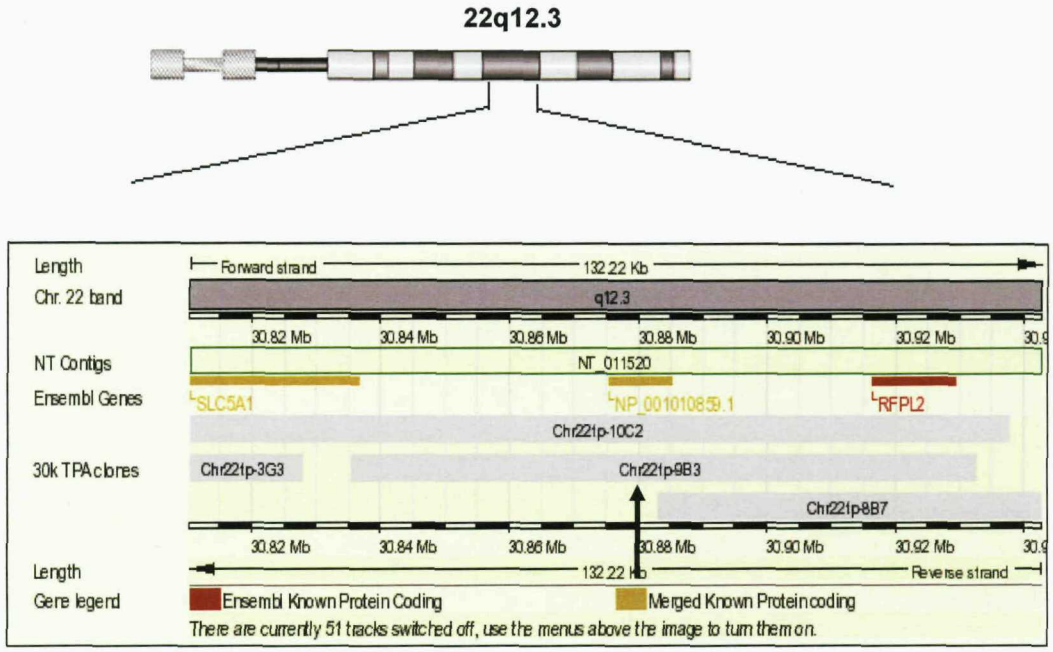

Fig.4.2. Screen shot from the Ensembl browser showing the 22q12.3 breakpoint region in case 3.8.

PAC RP1-127L4 (Chr22tp-9B3), indicated by the arrow, was split by the breakpoint. Note that this PAC contains the genes *NP_001010859.1* and *RFPL2*.

The chromosome 22 breakpoint region was further refined by fosmid mapping. FISH results for three informative fosmid clones (WI2-1419L13, WI2-2818H11 and WI2-2776M6) are summarised in figure 4.3, which shows that the 22q12.3 breakpoint mapped within fosmid WI2-2818H11. Part of this fosmid (25%) contains the predicted and uncharacterised gene *NP_001010859.1* (also known as *LOC150297*). However, the FISH results obtained are inconsistent with the order assigned to the fosmids tested; fosmid WI2-1419L13 is mapped centromeric to the breakpoint, therefore, it would be expected that this probe would hybridise to the normal chromosome 22 and to the derivative 22. Instead, the pattern of hybridisation obtained (on chromosome 22 and on the derivative 6) indicates that WI2-1419L13 is mapped telomeric to the breakpoint. These results could be explained by an error in the identity or in the position assigned to the fosmids. Alternatively, a small inversion in this region could be present, but due to the close physical proximity of the fosmids it is not possible to test this hypothesis by FISH as the hybridisation signals merge together and cannot be discerned. In an attempt to gather more information on this region, a literature search was made for previously reported inversion polymorphisms, but none were found.

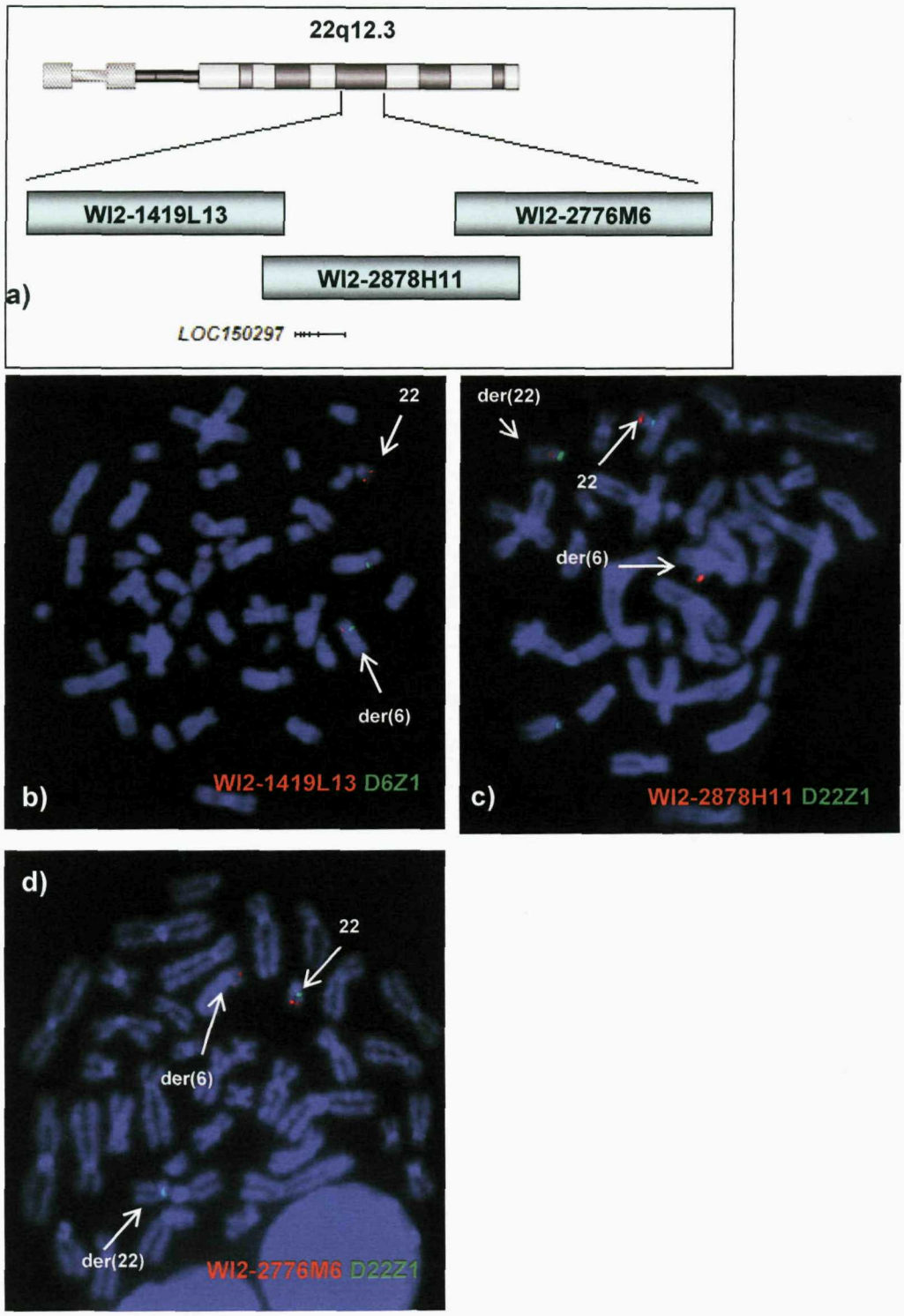

Fig.4.3. Fosmid FISH results for case 3.8.

a) Diagram adapted from the UCSC browser showing the relative position of the fosmid clones analysed and the *LOC150297* gene. Note that the same hybridisation pattern was observed for fosmids WI2-1419L3 (b) and WI2-2776M6 (d), which map centromeric and telomeric to the breakpoint, respectively. WI2-2878H11 was split by the breakpoint (c).

4.2.5 FISH results for case 3.11: 46,XY,t(2;3)(p23.1;q29)mat.ish t(2;3)(p23.3;)

This male individual, previously reported as case 13 by Baptista *et al.* (2005), was referred at 5 years of age for routine analysis of relatives of carriers of chromosome abnormalities. The proband of the family was his clinically abnormal sibling who had an unbalanced karyotype 46,XX,der(3)t(2;3)(p23.1;q29). The 2p23.3 breakpoint was mapped to an 192 kb region within BAC RP11-236C20. Part of this region (60%) contains the genes: *C2orf16* (chromosome 2 open reading frame 16), *ZNF512* (zinc finger protein 512), *CCDC121* (coiled-coil domain containing 121), *XAB1* (XPA binding protein 1, GTPase), *SUPT7L* (suppressor of Ty 7 -*S. cerevisiae*-like) and *SLC4A1AP* (solute carrier family 4 -anion exchanger-member 1, adaptor protein). On the derivative 3, no breakpoint region was identified. Therefore, although 2p23.2→2pter material was translocated to distal 3q, no reciprocal chromosome 3 product was identified on the derivative 2, suggesting that the translocation was balanced and non-reciprocal instead of reciprocal as determined by conventional cytogenetics.

The finding that no translocated chromosome 3 material was present on the der(2) suggested that either the 2p telomere had re-formed or an additional chromosome was involved in this rearrangement. FISH analysis with the subtelomere probes of all human chromosomes excluded the involvement of any other chromosome, suggesting that healing of the 2p telomere is likely. FISH results with some of the most informative probes that characterise this translocation are shown in figure 4.4.

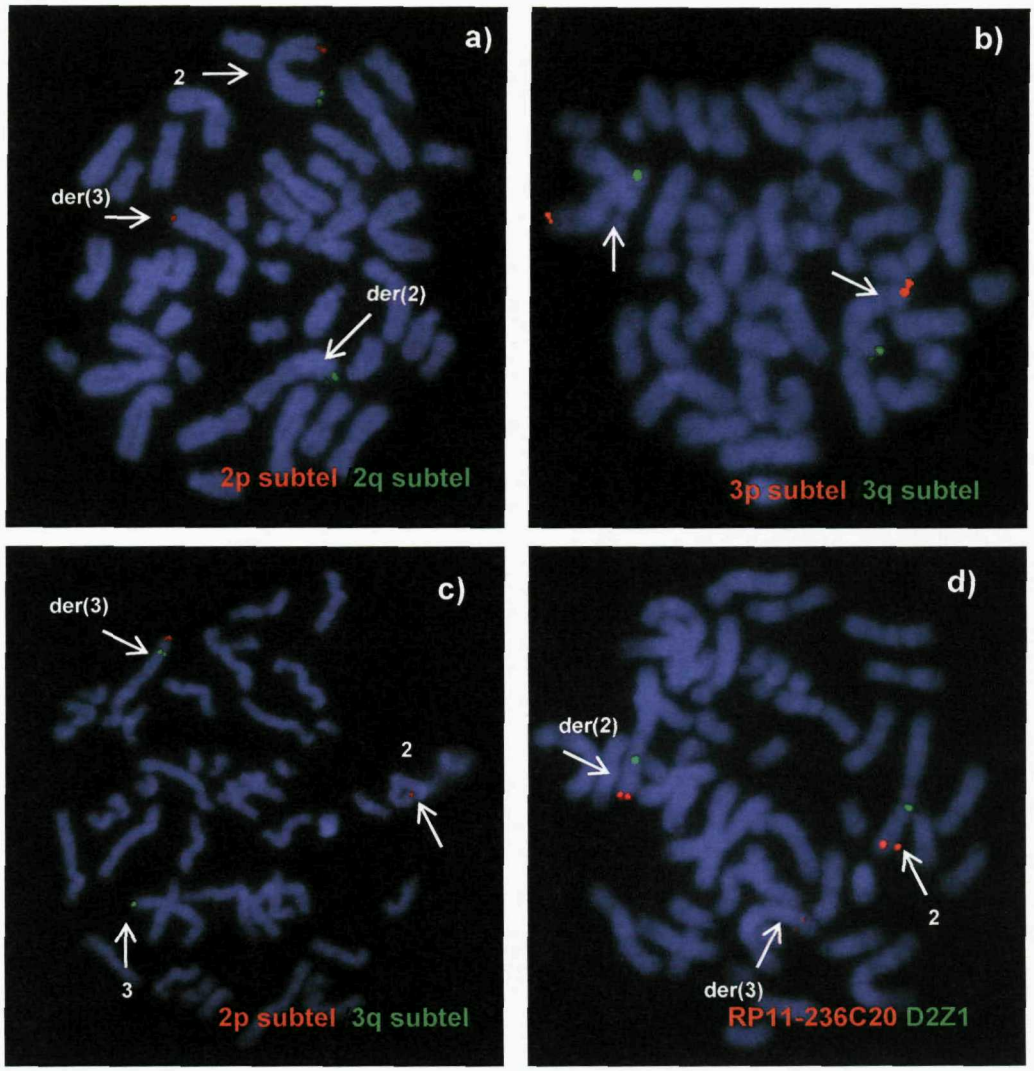

Fig.4.4. FISH images for case 3.11.

a) the der(2) retained the 2q subtelomeric region, whereas the 2p subtelomere is translocated to 3q, confirming that the translocation involved 2p. b) both chromosomes 3 and der(3) retained their subtelomeres p and q, confirming the absence of a terminal breakpoint on chromosome 3. c) the 2p subtelomere is translocated to a region distal to the 3q subtelomere, suggesting that the translocation is non-reciprocal. d) RP11-236C20 is split by the 2p23.3 breakpoint.

4.3 High resolution breakpoint mapping in a selection of individuals

Following the FISH breakpoint mapping studies described above, high resolution analyses were undertaken in a group of five randomly selected cases (2.9, 2.10, 2.12, 2.13 and 3.8). The studies in these individuals involved the refining of the breakpoints by array painting technology using custom designed arrays as detailed in section 3.3.2. A summary of the results is given in table 4.I. Furthermore, case 2.10 was analysed further at the sequence level by Gribble *et al.* (2007).

Table 4.I. Summary of the high resolution array painting results

Case	BCI genomic position (bps)	BCI size (bps)
2.9	chr2:36,974,205-36,984,164	9,959
2.9	chr7:11,125,470-11,131,360	5,890
2.10	chr11:37,799,191-37,800,372	1,181
2.10	chr17:12,739,251-12,759,331	20,080
2.12	chr8:100,008,198-100,013,461	5,263
2.12	chr16:59,138,187-59,146,748	8,561
2.13	chr16:82,997,304-83,005,803	8,499
2.13	chr18:42,902,560-42,915,062	12,502
3.8	chr6:40,838,172-40,840,190	2,018

BCI Breakpoint containing interval

4.4 Summary of the breakpoint mapping results

As stated in section 4.2, many breakpoint regions were found to map to different cytogenetic bands to those assigned by conventional cytogenetics. A comparison of the results using conventional cytogenetics and molecular cytogenetics is given in Table 4.II, which shows that only 6/60 breakpoints mapped to the same location by both techniques, whereas the remaining were re-assigned to a sub-band of the same band or to a different band altogether. Furthermore, the majority of the breakpoints were re-assigned to a band adjacent to the cytogenetically determined breakpoint and only 5/60 were re-assigned to 2 or more bands away.

The details of the breakpoint regions are summarised in Tables 4.III (group 1), 4.IV (group 2) and 4.V (group 3). These include the genomic position (in bps) of the breakpoint-containing intervals (BCI). Additionally, the genes annotated to the BCI are given, together with their likelihood of disruption (estimated as described in section 4.2). For the cases with gene mapping discrepancies between the Ensembl and UCSC browsers, both estimates are given. These discrepancies were particularly relevant for cases 2.8 and 2.12, for which the use of one browser or the other would place the translocation in a different group. Thus, case 2.8 could either be associated with obligatory gene disruption as in group 2 (according to UCSC) or with only potential gene disruption as in group 3 (according to Ensembl). Similarly, case 2.12 could either be associated with breakpoint-mediated gene disruption as in group 2 (according to UCSC) or with no gene disruption as in group 1 (according to Ensembl). The choice of placing both cases in group 2, according to the mapping data from the UCSC browser, was simply because this was the browser used to select FISH probes for these cases.

Table 4.II. Comparison between the cytogenetically and the molecularly assigned breakpoints

Case	Cytogenetic breakpoint	Molecular breakpoint
1.1	**2p21**	**2p21**
1.1	14q13	14q21.1
1.2	3p26.2	3p26.3
1.2	9p22.3	9p23
1.3	X?q27	Xq28
1.3	7q22	7q31.1
1.4	10q24.3	10q25.1
1.4	18q12.2	18q12.3
2.1	4q35.1	4q35.2
2.1	16p13.13	16p13.2
2.2	1q32.3	1q32.2
2.2	13q32.3	13q32.1
2.3	2q35	2q34
2.3	18q21.3	18q21.31
2.4	2q21.3	2q22.1
2.4	9p13	9p13.3
2.5	7q36.1	7q33
2.5	17q25.1	17q23.2
2.6	8p11.2	8p11.21
2.6	15q24	15q24.1
2.7	1p22	1p21.2
2.7	13q32	13q31.3
2.8	**11p15.4**	**11p15.4**
2.8	21p12	not determined
2.9	2p23.3	2p22.2
2.9	7p22.3	7p21.3
2.10	11p13	11p12
2.10	17p13.1	17p12
2.11	7p15	7p15.3
2.11	16q22	16q22.1
2.12	8q22.1	8q22.2
2.12	16q13	16q21
2.13	16q24	16q24.1
2.13	**18q21.1**	**18q21.1**
2.14	5p13	5p13.2
2.14	18q11	18q11.1
2.15	1q42.3	1q43

Table 4.II. Comparison between the cytogenetically and the molecularly assigned breakpoints (Cont.).

Case	Cytogenetic breakpoint	Molecular breakpoint
2.15	11q21	11q22.1
2.16	3p23	3p24.2
2.16	10q21.2	10q22.1
3.1	4q27	4q28.1
3.1	6p25	6p24.3
3.2	**8p23.1**	**8p23.1**
3.2	12p13.1	12p13.31
3.3	6q22.2	6q22.31
3.3	**9p22.3**	**9p22.3**
3.4	2p23	2p23.3
3.4	**4p12**	**4p12**
3.5	Xp11.23	Xp11.1
3.5	22q13.1	22q11.21
3.6	11q23	11q23.3
3.6	15q22	15q21.2
3.7	2q32.2	2q33.1
3.7	6p23	6p25.1
3.8	6p21.3	6p21.1
3.8	22q13	22q12.3
3.9	1q42.13	1q42.12
3.9	19p13.2	19p13.12
3.10	**9p24.1**	**9p24.1**
3.10	20p11.2?3	20p11.21
3.11	2p23.1	2p23.3
3.11	3q29	none detected

Bold type indicates identical results by both techniques

Table 4.III. Breakpoint mapping results for translocations in group 1

Case	BCI genomic position	BCI size	Genes in BCI	LOGD
1.1	chr2:42,991,855-43,174,584	182,729	----	----
1.1	chr14:40,212,036-40,381,064	169,028	----	----
1.2	chr3:581,645-958,276	376,631	----	----
1.2	chr9:11,303,383-11,764,217	460,834	----	----
1.3	chrX:148,868,102-149,049,276	181,174	----	----
1.3	chr7:108,572,249-108,768,779	196,530	----	----
1.4	chr10:110,235,641-110,401,130	165,489	----	----
1.4	chr18:37,199,353-37,363,565	164,212	----	----

BCI Breakpoint containing interval
LOGD Likelihood of gene disruption

Table 4.IV. Breakpoint mapping results for translocations in group 2

Case	BCI genomic position	BCI size	Genes in BCI	LOGD
2.1	chr4:187,776,924-187,818,201	41,277	*FAT*	100
2.1	chr16:8,687,646-8,767,351	79,705	*ABAT*	100
2.2	chr1:206,420,427-206,463,044	42,617	*PLXNA2*	100
2.2	chr13:95,145,761-95,229,198	83,437	*DNAJC3*	100
2.3	chr2:214,463,710-214,641,904	178,194	*SPAG16*	100
2.3	chr18:53,150,966-53,228,407	77,441	*ST8SIA3*	20
2.4	chr2:138,955,571-139,047,575	92,004	*SPOPL*	75
2.4	chr9:35,316,944-35,352,859	35,915	*UNC13B*	100
2.5	chr7:132,668,430-132,886,439	218,009	*EXOC4*	100
2.5	chr17:53,768,215-53,854,109	85,894	*SUPT4H1, RNF43*	80
2.6	chr8:42,400,367-42,558,693	158,326	*SLC20A2, C8orf40*	80
2.6	chr15:71,549,843-71,677,866	128,023	*NPTN, LOC283677*	100
2.7	chr1:101,729,637-101,888,180	158,543	----	----
2.7	chr13:91,284,473-91,451,991	167,518	*GPC5*	100
2.8	chr11:8,789,238-8,828,410	39,172	*ST5*	70-100
2.8	chr21: satellite stalk	nd	nd	nd
2.9	chr2:36,974,205-36,984,164	9,959	*STRN*	100
2.9	chr7:11,125,470-11,131,360	5,890	*PHF14*	100
2.10	chr11:37,799,191-37,800,372	1,181	----	----
2.10	chr17:12,739,251-12,759,331	20,080	*QRICH2*	100
2.11	chr7:24,352,315-24,556,482	204,167	----	----
2.11	chr16:68,412,623-68,499,123	86,500	*WWP2*	100
2.12	chr8:100,008,198-100,013,461	5,263	*STK3*	0-100
2.12	chr16:59,138,187-59,146,748	8,561	----	----
2.13	chr16:82,997,304-83,005,803	8,499	*ATP2C2*	100
2.13	chr18:42,902,560-42,915,062	12,502	*HDHD2*	100
2.14	chr5:38,450,546-38,489,171	38,625	*EGFLAM*	100
2.14	chr18:16,935,079-16,977,780	42,701	*ROCK1*	25
2.15	chr1:235,591,808-235,631,043	39,235	*RYR2*	100
2.15	chr11:98,312,650-98,487,722	175,072	*CNTN5*	0-50
2.16	chr3:25,187,406-25,225,486	38,080	----	----
2.16	chr10:71,325,230-71,369,369	44,139	*COL13A1*	100

BCI Breakpoint containing interval
LOGD Likelihood of gene disruption
nd not determined

Table 4.V. Breakpoint mapping results for translocations in group 3

Case	BCI genomic position	BCI size	Genes in BCI	LOGD
3.1	chr4:125,649,232-125,829,198	179,966	*ANKRD50*	15
3.1	chr6:7,324,263-7,409,496	85,233	*CAGE,RIOK1*	45
3.2	chr8:6,719,131-7,053,466	334,335	*DEFB1,DEFA6 ,DEFA4,DEFA 1,DEFA3,DEF A5*	10
3.2	chr12:7,987,953-8,310,582	322,629	*FOXJ2, C3AR, NECAP1, CLEC4A, ZNF705A, FAM90A1*	25-76
3.3	chr6:119,215,818-119,302,998	87,180	*ASF1A, MCM9*	40-65
3.3	chr9:14,905,963-15,184,142	278,179	*C9orf52*	10
3.4	chr2:26,218,578-26,257,187	38,609	*FAM59B*	20
3.4	chr4:48,429,074-48,516,206	87,132	*FRYL*	55
3.5	chrX:57,142,101-57,184,980	42,879	*SPIN2B, SPIN2A*	10
3.5	chr22:19,027,823-19,108,026	80,203	*USP41, ZNF74*	35-75
3.6	chr11:115,668,958-115,830,806	161,848	----	----
3.6	chr15:48,315,253-48,399,500	84,247	*HDC, GABPB2*	75
3.7	chr2:200,490,832-200,529,742	38,910	*FLJ38973, C2orf60*	95
3.7	chr6:6,602,851-6,705,437	102,586	----	----
3.8	chr6:40,838,172-40,840,190	2,018	----	----
3.8	chr22:30,875,582-30,913,878	38,296	*LOC150297*	25
3.9	chr1:223,565,344-223,691,038	125,694	*LBR*	20
3.9	chr19:13,999,383-14,039,028	39,645	*RLN3, IL27RA*	60
3.10	chr9:6,427,961-6,601,726	173,765	*UHRF2, GLDC*	85
3.10	chr20:23,000,436-23,041,115	40,679	*CD93*	15
3.11	chr2:27,620,439-27,812,447	192,008	*C2orf16, ZNF512, CCDC121, XAB1, SUPT7L, SLC4A1AP*	60

BCI Breakpoint containing interval
LOGD Likelihood of gene disruption

4.5 Further characterisation of the breakpoint regions

The breakpoint mapping studies undertaken here had the primary aim of determining the gene content of the breakpoint regions. However, it was also of interest to characterise the breakpoint regions further by determining the types of genes and the genomic context of the breakpoints.

4.5.1 The types of genes at the breakpoints

In an attempt to obtain additional information on the genes mapped to the breakpoints, several databases were consulted, including the GO database, OMIM, the Human Gene Mutation database, UniProt and GeneCards (URLs in Appendix 1). This analysis showed that the genes mapped to the breakpoint regions were involved in a variety of biological processes (full details are given in Appendix 8), of which the most common were transcription, signal transduction and cell adhesion (Table 4.VI). Many of the genes mapped to the breakpoints have unknown function, but some are known to cause Mendelian diseases (listed in Table 4.VII) or have been associated with susceptibility to complex or late onset disorders (Table 4.VIII).

Table 4.VI. Most common GO terms for genes mapped to the breakpoints in the control cohort

GO term	Gene
transcription regulation of transcription transcription factor activity	*ASF1A*, *FOXJ2*, *FRYL*, *GABPB2*, *SUPT4H1*, *SUPT7L*, *ZNF512*, *ZNF74*
signal transduction Rho protein signal transduction cell surface receptor linked signal transduction	*C3AR*, *IL27RA*, *PLXNA2*, *QRICH2*, *ROCK1*, *RYR2*, *STK3*
cell-cell adhesion cell adhesion homophilic cell adhesion	*CD93*, *CLEC4A*, *CNTN5*, *COL13A1*, *FAT*, *NPTN*

Table 4.VII. Genes associated with Mendelian disorders

Case	Gene	Disorder	Inheritance
2.1	***ABAT***	GABA-AT deficiency [MIM 137150]	AR
2.15	***RYR2***	familial arrhythmogenic right ventricular dysplasia 2 [MIM 600996]	AD
2.15	***RYR2***	stress-induced polymorphic ventricular tachycardia [MIM 604772]	AD
2.15	***RYR2***	familial polymorphic ventricular tachycardia [MIM 192605]	AD
3.9	*LBR*	Pelger-Huet anomaly (PHA) [MIM 169400]	AD
3.9	*LBR*	hydrops-ectopic calcification-moth-eaten skeletal dysplasia [MIM 215140]	AR
3.10	*GLDC*	Glycine encephalopathy [MIM 605899]	AR

Bold type indicates genes obligatory disrupted by a breakpoint, whereas normal type indicates potentially disrupted genes (Tables 4.IV and 4.V).
AR autosomal recessive
AD autosomal dominant

Table 4.VIII. Genes associated with complex disorders or disease susceptibility

Type of disease	Gene
Malignancy	*C3AR, CLEC4A, DEFA1, DEFA3, DEFA4, DNAJC3 , FAT, FRYL, GPC5, HDC, ROCK1 , SLC20A2 , ST5*
Inflammation	*C3AR, CLEC4A, DEFA1, DEFA3, DEFA5, DEFA6, DEFB1, EXOC4*
Psychiatric disease	*NPTN, PLXNA2, ZNF74*

4.5.2 Genomic context of the breakpoints

4.5.2.1 Cytogenetic bands

It was of interest to assess the distribution of the breakpoint regions to cytogenetic bands of characteristic and specialised genomic architecture such as pericentromeric and centromeric regions, satellite stalks and telomeres. In addition, the GDB database (URL in Appendix 1) was utilised to ascertain whether the breakpoint regions mapped to cytogenetic bands previously reported to contain fragile sites, which are chromosome regions that form cytogenetically visible gaps or breaks when cultured under conditions that hamper DNA replication or repair (reviewed by Debacker and Kooy, 2007) and that have been suggested to co-localise to breakpoints in ABCRs (Warburton, 1991). The results of this analysis are summarised in Table 4.IX, which shows that 5 breakpoints mapped to pericentromeric regions, 1 to within satellite stalks and 2 in telomeric bands, while fragile sites have been reported in 12 of the breakpoints regions.

Table 4.IX. Cytogenetic bands of interest in the control cohort

Case	Cytogenetic band	Type of region[1]
1.3	Xq28	telomeric, FRAXE
2.1	4q35.2	telomeric
2.1	17p12	FRA17A
2.2	13q32.1	FRA13D
2.6	8p11.21	pericentromeric
2.7	1p21.2	FRA1E
2.8	21p12	satellite stalk
2.11	16q22.1	FRA16C
2.14	5p13.2	FRA5A
2.16	3p24.2	FRA3A
2.16	10q22.1	FRA10D
3.4	4p12	pericentromeric
3.5	Xp11.1	pericentromeric
3.5	22q11.21	pericentromeric
3.6	11q23.3	FRA11G
3.7	2q33.1	FRA2I
3.7	6p25.1	FRA6B
3.9	1q42.12	FRA1H
3.10	20p11.21	pericentromeric

[1]Fragile sites have a prefix FRA and were mapped according to the GDB database

4.5.2.2 Segmental duplications

Segmental duplications (SDs) are repeat regions, typically with a size between 1 and 400 kb, which share a high degree (>90%) of sequence identity (reviewed by Eichler, 2001). These regions represent 5% of the human genome and are known to predispose and mediate genomic rearrangements, including deletions, duplications, inversions and translocations, thus playing a role in human disease and variability (reviewed by Lupski and Stankiewicz, 2005). SDs are involved in the generation of recurrent rearrangements and hence, it was of interest to determine whether SDs might mediate the rearrangements analysed in this study, which are essentially non recurrent. The UCSC genome browser was

used to analyse the SD content of the breakpoint regions. The results of this analysis are summarised in table 4.X, which shows that SDs mapped to 15/60 (25%) of the breakpoint regions.

4.5.2.3 Normal copy number variation

Copy number variation (CNV) refers to a copy number gain or loss of > 1 kb in size, for which the frequency or clinical significance might or might not be known (Scherer *et al.*, 2007). Studies of normal populations suggest that normal CNV accounts for at least 12% of the human genome (Redon *et al.*, 2006) and might have an important role in genetic diversity and susceptibility to disease (reviewed by Freeman *et al.*, 2006). Thus, it was of interest to determine whether the breakpoints described in the present study were preferentially mapped to regions previously reported to contain normal CNV. The Database of Genomic Variants was used for this purpose and the results are summarised in Table 4.X, which shows that 30/60 (50%) of the breakpoint regions co-localised to known sites of normal CNV.

4.5.2.4 Potentially disease-causative copy number changes

The characterisation of the breakpoints in phenotypically normal individuals is useful for the identification of genes whose disruption is compatible with normal development. The information collected from this type of study can prove useful for the understanding of mutations detected in phenotypically abnormal patients. In this context, it was of interest to determine the proportion of the breakpoints in the phenotypically normal individuals of the present study that mapped within potentially pathogenic regions of copy number change previously detected in phenotypically abnormal patients. This analysis was carried out by using the DECIPHER database (URL in Appendix 1), a web-

based resource containing data on genomic imbalances identified by array CGH in phenotypically abnormal patients. The results of this analysis are given in Table 4.X, which shows that 17/60 (28.3%) of the breakpoints map to regions of copy number change in clinically abnormal patients. Furthermore, 2 breakpoints mapped to within regions of known syndromes: the 2q32.2 deletion syndrome and the 22q11 deletion/duplication syndrome regions, thus providing additional insight into these regions.

Table 4.X. Genomic context of the breakpoint regions in the control cohort

Case	Breakpoint region	SDs	Normal CNV	DECIPHER data
1.1	chr2:42,991,855-43,174,584	-	-	-
1.1	chr14:40,212,036-40,381,064	-	-	+
1.2	chr3:581,645-958,276	+	+	-
1.2	chr9:11,303,383-11,764,217	-	+	+
1.3	chrX:148,868,102-149,049,276	-	+	+
1.3	chr7:108,572,249-108,768,779	-	-	+
1.4	chr10:110,235,641-110,401,130	-	-	-
1.4	chr18:37,199,353-37,363,565	-	-	-
2.1	chr4:187,776,924-187,818,201	-	+	+
2.1	chr16:8,687,646-8,767,351	-	-	-
2.2	chr1:206,420,427-206,463,044	-	+	-
2.2	chr13:95,145,761-95,229,198	-	-	+
2.3	chr2:214,463,710-214,641,904	-	+	+
2.3	chr18:53,150,966-53,228,407	-	-	-
2.4	chr2:138,955,571-139,047,575	+	-	-
2.4	chr9:35,316,944-35,352,859	-	+	-
2.5	chr7:132,668,430-132,886,439	-	-	+
2.5	chr17:53,768,215-53,854,109	+	+	+
2.6	chr8:42,400,367-42,558,693	-	-	-
2.6	chr15:71,549,843-71,677,866	-	-	-
2.7	chr1:101,729,637-101,888,180	-	+	-
2.7	chr13:91,284,473-91,451,991	+	-	+
2.8	chr11:8,789,238-8,828,410	-	+	-
2.8	chr21: satellite stalk	nd	nd	nd
2.9	chr2:36,974,205-36,984,164	+	-	+
2.9	chr7:11,125,470-11,131,360	-	-	-
2.10	chr11:37,799,191-37,800,372	-	-	-
2.10	chr17:12,739,251-12,759,331	-	-	-
2.11	chr7:24,352,315-24,556,482	+	+	-
2.11	chr16:68,412,623-68,499,123	-	+	-
2.12	chr8:100,008,198-100,013,461	-	-	-

Table 4.X. Genomic context of the breakpoint regions in the control cohort (Cont.)

Case	Breakpoint region	SDs	Normal CNV	DECIPHER data
2.12	chr16:59,138,187-59,146,748	-	-	-
2.13	chr16:82,969,935-83,038,274	-	+	-
2.13	chr18:42,902,560-42,915,062	-	+	-
2.14	chr5:38,450,546-38,489,171	-	-	-
2.14	chr18:16,935,079-16,977,780	-	-	-
2.15	chr1:235,591,808-235,631,043	-	-	-
2.15	chr11:98,312,650-98,487,722	+	-	+
2.16	chr3:25,187,406-25,225,486	-	+	-
2.16	chr10:71,325,230-71,369,369	-	+	-
3.1	chr4:125,649,232-125,829,198	-	-	-
3.1	chr6:7,324,263-7,409,496	-	+	-
3.2	chr8:6,719,131-7,053,466	+	+	-
3.2	chr12:7,987,953-8,310,582	+	+	-
3.3	chr6:119,215,818-119,302,998	-	+	-
3.3	chr9:14,905,963-15,184,142	+	-	-
3.4	chr2:26,218,578-26,257,187	-	+	-
3.4	chr4:48,429,074-48,516,206	-	-	-
3.5	chrX:57,142,101-57,184,980	+	+	-
3.5	chr22:19,027,823-19,108,026	+	+	22q11 del/dup*
3.6	chr11:115,668,958-115,830,806	-	-	+
3.6	chr15:48,315,253-48,399,500	-	+	+
3.7	chr2:200,490,832-200,529,742	-	-	2q32.2 del†
3.7	chr6:6,675,920-6,706,437	-	-	-
3.8	chr6:40,838,172-40,840,190	-	-	-
3.8	chr22:30,875,582-30,913,878	+	+	-
3.9	chr1:223,565,344-223,691,038	-	+	+
3.9	chr19:13,999,383-14,039,028	-	+	-
3.10	chr9:6,427,961-6,601,726	-	+	+
3.10	chr20:23,000,436-23,041,115	-	+	-
3.11	chr2:27,620,439-27,812,447	+	+	-

*22q11 deletion/duplication syndrome region
†2q32.2 deletion syndrome region

4.6 Array CGH results

Array CGH imbalances were detected in all 31 phenotypically normal individuals and a summary of the number of regions of imbalance in each individual is given in Table 4.XI. In total, 73 different genomic regions defined by 80 BAC/PAC clones were involved. These regions were compared to sites of normal CNV documented in the Database of Genomic Variants, which showed that many regions constituted know CNV, whereas 21 were novel (listed in Table 4.XII). In an attempt to characterise these regions further, their content in SDs, known genes and disease loci (according to the OMIM database, URL in Appendix 1) was determined and is given in table 4.XII. This table shows that SDs were present in 4/21 (19%), known genes in 16/21 (76%) and disease loci in 8/21 (38%) of these regions.

Overall, on analysis of the array CGH profiles, all individuals had apparently normal profiles with sites of CNV being mainly changes of single clones. Case 2.11, who had an apparently balanced translocation between 7p and 16q was an exception as its array CH profile showed a deletion of >2 Mb, including 2-4 clones on 3p21.31 (figure 4.5). This observation was made prior to the publication of most studies of normal CNV and it was unknown whether it represented normal CNV, but since the region involved was considerably large, it was of interest to characterise it further. FISH analysis on metaphases prepared from lymphoblastoid cell lines showed that this was a mosaic deletion, present in 70-90% of the cells analysed. Subsequent analysis of microsatellite markers showed that the deletion was present in the patient's transformed lymphoblasts but not in genomic DNA (figure 4.6). This result is consistent with the deletion being a somatic mutation that occurred *in vitro*.

Table 4.XI. Number of regions of CNV in the control cohort

Case	Regions of loss	Regions of gain	Total
1.1	2	2	4
1.2	5	1	6
1.3	3	1	4
1.4	3	3	6
2.1	2	1	3
2.2	5	1	6
2.3	1	3	4
2.4	10	0	10
2.5	5	1	6
2.6	7	3	10
2.7	1	2	3
2.8	2	1	3
2.9	4	31	35
2.1	3	4	7
2.11	5	3	8
2.12	6	1	7
2.13	5	1	6
2.14	2	3	5
2.15	5	0	5
2.16	4	3	7
3.1	1	1	2
3.2	8	1	9
3.3	4	1	5
3.4	6	3	9
3.5	2	1	3
3.6	5	1	6
3.7	4	0	4
3.8	0	4	4
3.9	3	4	7
3.1	5	4	9
3.11	3	0	3

Table 4.XII. Characteristics of the novel regions of CNV identified in the control cohort

Region[1]	Chr	SDs	Genes	Disease /Syndrome region
1	2p12	-	+	
2	3q13.33	-	+	
3	3q26.1	-	+	Postanesthetic apnea
4	3q26.33	-	-	
5	4q13.1	-	+	
6	5p15.2	-	+	Cri du chat
7	5p14.1	+	-	
8	5p14.1	-	-	
9	5q23.2	-	+	Congenital contractural arachnodactyly
10	5q23.3	-	+	Susceptibility to rheumatoid arthritis, systemic primary carnitine deficiency
11	5q32	-	+	
12	5q33.3	-	-	
13	6q22.31	-	-	
14	7q21.11	-	+	
15	7q31.2	-	+	
16	7q32.1	-	+	Renal cell carcinoma papillary 2
17	8q21.13	+	+	
18	11q23.3	-	+	Glycogen storage disease
19	19p13.2	-	+	Susceptibility to cerebral malaria, Landsteiner-Wiener blood group
20	19p13.13	+	+	Episodic ataxia 2, familial hemiplegic migraine, spinocerebellar ataxia-6, idiopathic generalized epilepsy
21	19q13.32	+	+	

[1]Bold type indicates regions seen in more than one individual, whereas normal type indicates regions seen in a single individual.

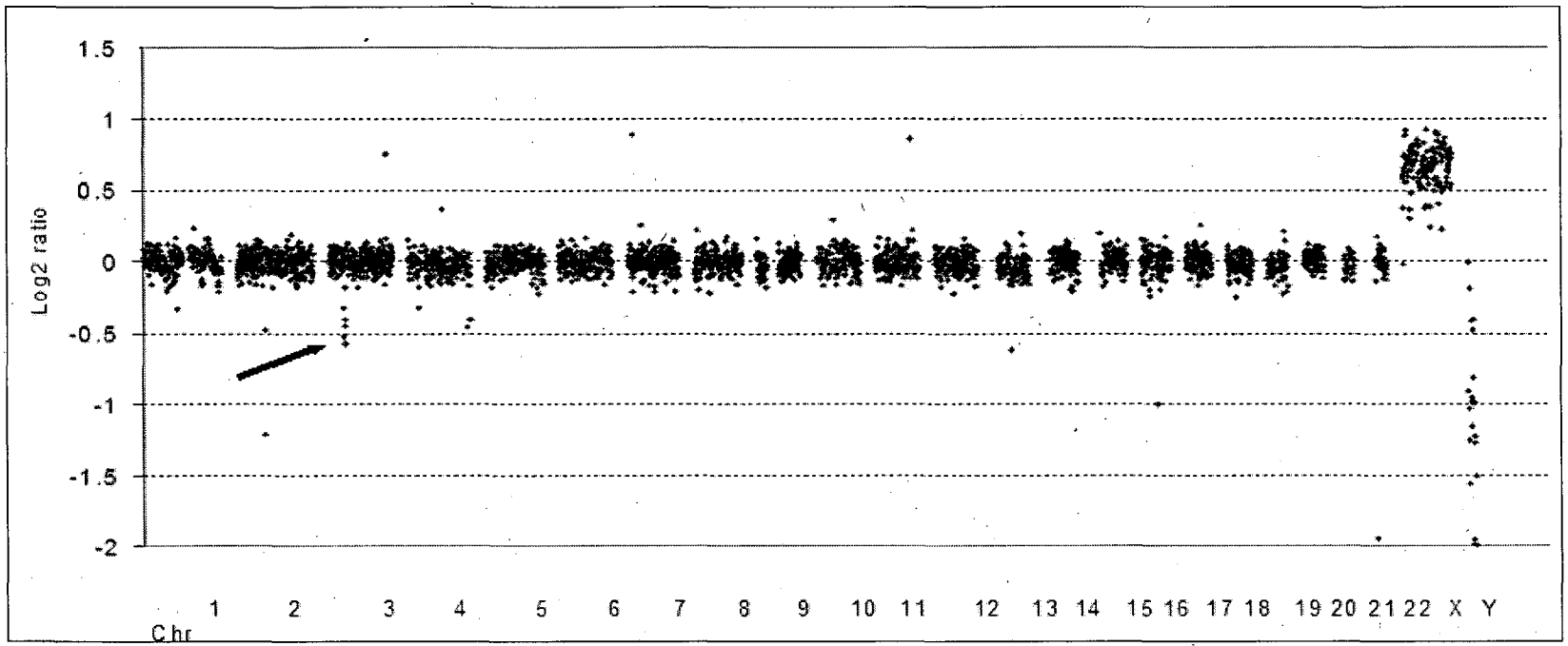

Fig.4.5. Array CGH results for case 2.11.

Genomic clones are shown in the x-axis and are grouped by chromosome 1 to Y. Log_2 fluorescence ratios of test/reference DNA are given in the y-axis. The array CGH profile in case 2.11 shows the gain of sequences from chromosome X and loss of those from chromosome Y, consistent with the sex mismatch test/reference. The arrow indicates the 3p21.31 somatic deletion.

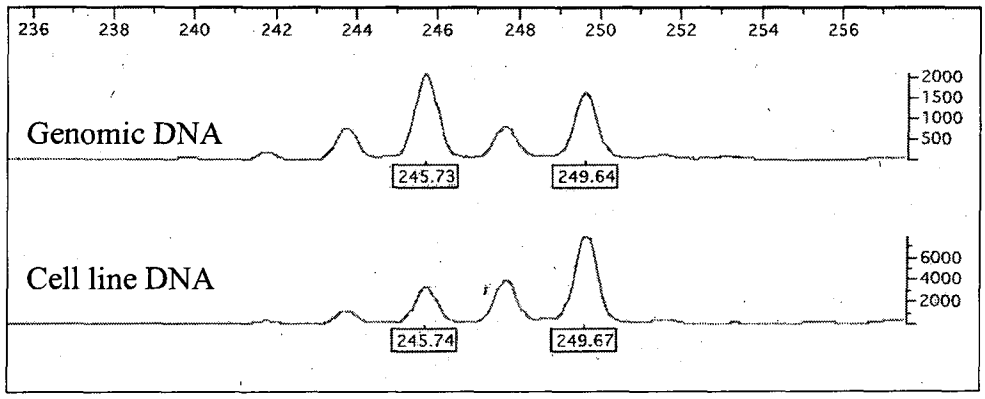

Fig.4.6. Analysis of marker D3S3629 in DNA extracted from blood and from a lymphoblastoid cell line from case 2.11.

Two alleles were detected in each sample but, in the cell line DNA the amount of the 245 allele is significantly reduced. This result is consistent with the mosaic somatic deletion detected by FISH in the cell line.

4.7 Overall summary of the results

The use of molecular cytogenetic techniques for the analysis of the ABCRs in a control cohort of 31 phenotypically normal individuals enabled the re-assignment of the majority of the breakpoints, which were refined to small genomic intervals, with 60% of the breakpoints mapped to intervals <100 kb, 28.3% between 100 and 200 kb and 11.7% to intervals >200 kb.

An analysis of the types of regions involved in breakpoints showed that many breakpoints mapped to cytogenetic bands with no remarkable features, with only 8 (13.1%) of the breakpoints in pericentromeric regions, satellite stalks or telomeric bands and 12 (19.7%) in bands previously reported to contain fragile sites. Further analysis of the genomic context of the breakpoints showed that 25% mapped to regions containing segmental duplications and 50% to regions previously reported to contain normal CNV. In addition, 28.3% mapped to regions that are deleted or duplicated in patients with clinical abnormalities according to the DECIPHER database, including two syndrome regions: the

2q32.2 deletion ad the 22qdeletion/duplication. The four features that are the focus of the present study are discussed below:

(i) Breakpoint-mediated gene disruption

The most remarkable finding in the present study was that a high proportion of the breakpoints in this control cohort are associated with gene disruption. As many as 16 translocations were definitely associated with breakpoint-mediated gene disruption and in a further 11, gene disruption by one of the breakpoints was a possibility, while only 4 had both breakpoint in regions containing no known genes.

In total, the breakpoint regions spanned >60 genes, of which 18 were definitely disrupted (Tables 4.IV and 4.V). Surprisingly, as many as 25% of the genes mapped to the breakpoint regions were implicated in important biological processes as transcription and signal transduction. Furthermore, several genes have been implicated in complex disorders (Table 4.VIII), mainly in malignancy, inflammation and psychiatric disorders. This finding is likely to be biased by the fact that the above disorders are the object of many published studies. Among these genes, *FAT* was of particular interest. This gene is largely uncharacterised in humans, but the *Drosophila* homolog codes for a tumour suppressor protein involved in cell proliferation during development. This gene was found to be disrupted in case 2.1, who had a prolactinoma diagnosed at 43 years age. Further studies would be of interest to ascertain whether disruption of *FAT* is causally associated with development of this type of tumor.

Genes implicated in Mendelian disorders were identified in the breakpoint regions of 4 individuals (Table 4.VII). *ABAT* was disrupted by the breakpoint in case 2.1. Mutations in this gene have been reported to cause GABA-AT deficiency [MIM 137150], an autosomal recessive disorder (Gibson *et al.*, 1985) characterised by severe psychomotor retardation, hypotonia, hyperreflexia (Jaeken *et al.*, 1984) and lethargy, refractory seizures, and EEG abnormalities (Medina-Kauwe *et al.*, 1999). The absence of these clinical features in case 2.1 can be explained by the autosomal recessive nature of

GABA-AT deficiency. The other genes *GLDC*, *LBR* and *RYR2* mapped to the breakpoints in patients from whom a detailed clinical history was not available and thus, genotype-phenotype correlations cannot be made in these cases.

(ii) Breakpoint-associated imbalances

Imbalances at the breakpoints were not identified in any of the individuals analysed.

(iii) Additional chromosomal complexity

Additional chromosomal complexity was not detected in 30/31 translocations, but could not be excluded in the remaining case. For case 3.8, the finding of FISH results inconsistent with the probes tested for mapping of the 22q12.3 breakpoint suggested that this breakpoint could be associated with a cryptic inversion or that there was an error in the order or identity of the probes tested (discussed in section 4.2.4). Since this region of 22q12.3 is characterised by a high content in segmental duplications, it is likely that sequencing errors exist in this region or that an inversion has been mediated by this genomic architecture. A search for previously reported inversions involving this region failed to retrieve any results, but this might simply reflect the little data available on submicroscopic inversions in normal individuals, which have been surveyed in only a small number of studies (Tuzun *et al.*, 2005; Feuk *et al.*, 2005).

(iv) Genomic imbalances unrelated to the breakpoints

Genomic imbalances unrelated to the breakpoints were identified, the majority of which being previously reported normal CNV from studies of normal populations. Novel regions of normal CNV were also identified (Table 4.XII), contributing to the growing list of these genomic sites; however, further validation studies will be needed before these regions are added to databases of normal CNV. These novel sites of CNV were found to overlap to genes that might contribute to human disease and diversity as reported in most surveys of CNV of normal populations (Iafrate *et al.*, 2004; Sebat *et al.*, 2004; Redon *et al.*, 2006). In addition a genomic imbalance unrelated to the breakpoint was found to be a somatic mutation that arose during transformation of human lymphoblasts by Epstein Barr virus (EBV). This result and previous reports on

the acquisition of *in vitro* abnormalities in EBV-transformed cells (Manolov *et al.*, 1981; Steel *et al.*, 1985) illustrates the need to carry out validation studies of any genomic imbalance detected when using this type of material.

4.8 Conclusions

Apparently balanced translocations in phenotypically normal individuals are essentially balanced and simple. However, they are often associated with breakpoint-mediated gene disruption, a feature that might help to elucidate the role of the genes mapped to these regions, will help in the delimitation of critical regions for microdeletion/microduplication syndromes and overall, aid in the interpretation of genotype-phenotype correlations in patients with clinical abnormalities.

5 MOLECULAR CYTOGENETIC ANALYSES IN A PATIENT COHORT OF 16 PHENOTYPICALLY ABNORMAL CASES

5.1 Overview

This chapter presents the results of the analysis of ABCRs in 16 phenotypically abnormal patients (clinical descriptions in Appendix 3). As discussed in chapter 4, the aim of these analyses was to determine the frequency of the four features thought to be associated with clinical abnormalities in this type of patient, namely: i) breakpoint-mediated gene disruption; ii) cryptic imbalances at the breakpoints; iii) additional chromosomal complexity; and iv) genomic imbalances unrelated to the breakpoints.

5.2 Breakpoint mapping results

The results of the breakpoint mapping studies by array painting and/or FISH are summarised below. Cases 49, 51 and 52 are presented in detail. This is because cases 49 and 51 provided significant evidence of a role for breakpoint-mediated gene disruption to the abnormal phenotypes, while case 52 presented all the four aforementioned features associated with abnormal phenotypes. The results presented below include the size of the breakpoint region, the identity of the genomic clones that defined this region and their chromosomal band location, which often differed from the cytogenetically determined breakpoint (details listed in Table 5.IX) as shown using an adaptation of the ISCN (2005) nomenclature.

Case 16: 46,XX,t(10;22)(q24.3;q13.31)*de novo*.ish t(10;22)(q24.31;q13.31)
This female patient was first referred at 8 years of age because of developmental delay, temporal lobe epilepsy and mild hearing impairment. The 10q24.31 breakpoint was mapped to a 39 kb genomic region within fosmid WI2-455C17. Part of this fosmid (40%) contains the *PAX2* (paired box gene 2) gene. The 22q13.31 breakpoint was mapped to a 41 kb region within cosmid cN75B3. This cosmid is fully contained within the predicted gene *KIAA1644* (unknown function) and as a result, this gene must be disrupted by the breakpoint.

Case 20: 46,XX,t(2;5)(q33;q12)*de novo*.ish t(2;5)(q33.2;q12.1)

This female patient was ascertained at 13 years of age because of mild mental retardation. The 2q33.2 breakpoint mapped to an 80 kb region distal to fosmid WI2-2222L4 and proximal to fosmid WI2-1034P1. Part of this region (80%) contains the *ABI2* (abl interactor 2) gene. The 5q12.1 breakpoint was mapped to a region distal to BAC RP11-2O17. FISH studies with probes mapped in the vicinity of this BAC showed a breakpoint-associated deletion, which was further characterised by array CGH (results given in section 5.5.1).

Case 43: 46,XY,t(4;17)(q35.1;q25.1)*de novo*.ish t(4;17)(q34.3;q24.2)

This male patient was referred antenatally for cytogenetic analysis because of an isolated truncus arteriosus. The 4q34.3 breakpoint was mapped to an 155 kb region within BAC RP11-352B19, which contained no known genes. The 17q24.2 breakpoint was mapped to a 75 kb region distal to fosmid WI2-653E19 and proximal to fosmid WI2-410L2. Part of this region (60%) contains the *PSMD12* (proteasome, prosome, macropain, 26S subunit, non-ATPase, 12) and *PITPNC1* (phosphatidylinositol transfer protein, cytoplasmic 1) genes.

Case 44: 46,XY,inv(6)(p24q16.2)*de novo*.ish inv(6)(p24.3q16.2)

This male patient was referred at 2 years of age because of mild global developmental delay, learning difficulties, dysmorphic features and a specific speech impairment. The 6p24.3 breakpoint was mapped to a 40 kb interval within fosmid WI2-506N5. The 6q16.2 breakpoint was mapped to an 175 kb region within BAC RP11-390H11. No known genes mapped to either of these breakpoint regions.

Case 45: 46,X,t(X;19)(q21;p13.11)*de novo*.ish t(X;19)(q21.1;q11.?)

This female patient was referred at 29 years of age because of premature ovarian failure (secondary amenorrhea). The Xq21.1 breakpoint mapped to an 163 kb region within BAC RP11-2L17, which contains no known genes. The chromosome 19 breakpoint mapped to proximal 19q. The whole short arm and the centromere of chromosome 19 were translocated to chromosome X. As a result, the derivative X was dicentric, with both X and 19 centromeres. FISH

analysis with an alpha satellite probe showed that the derivative 19 had retained some alpha satellite sequences.

Case 48: 46,XY,t(4;6)(q33;q22.2)*de novo*.ish t(4;6)(q33-34.1;q21)

This male patient was referred at 15 years of age because of developmental delay and dysmorphic features. The 4q33-34.1 breakpoint was mapped to a 203 kb region, within BAC RP11-344G13, which contains no known genes. The 6q21 breakpoint mapped to a 93 kb region distal to fosmid WI2-2439J5 and proximal to fosmid WI2-1412A5. Part of this region (20%) contains the *AMD1* (adenosylmethionine decarboxylase 1) gene.

Case 49: 46,XX,t(2;10)(q33;q21.2)*de novo*

Breakpoint mapping results are presented in detail in section 5.2.1.

Case 50: 46,X,t(X;8)(q22.1;q24.13).ish t(X;8)(q23;q24.3)

This female patient was referred at 29 years of age because of secondary amenorrhoea and was found to have an apparently balanced t(X;8)(q22.1;q24.13). The maternal karyotype was normal, but a paternal sample was not available for parent of origin analysis. The Xq23 breakpoint mapped to an 191 kb region within BAC RP11-107O11. Part of this region (3%) contains the *AGTR2* (angiotensin II receptor, type 2) gene. The 8q24.3 breakpoint was mapped to a 209 kb region, distal to fosmid WI2-1885B12 and proximal to BAC RP11-370K2, which contains no known genes.

Case 51: 47,XX,t(4;20)(p15.2;p11.23)*de novo*,+mar[23]mat /46,XX,t(4;20)(p15.2;p11.23)[7]

Breakpoint mapping results are shown in detail in section 5.2.2.

Case 52: 46,XX inv ins (11;4)(q22.2;q13.2q21.3).ish

Breakpoint mapping results are shown in detail in section 5.2.3.

Case 53: 46,XX,t(4;8)(q21.1;p12)*de novo*.ish t(4;8)(q13.3;p12)

This female patient was referred at 1 year of age because of developmental delay and an unusual pattern of progression/regression in skills. The 4q13.3

breakpoint was mapped distal to BAC RP11-88J6 and was associated with a deletion, which was characterised further by array CGH (presented in section 5.5.1). The 8p12 breakpoint was mapped to an 163 kb region within BAC RP11-274F14. This clone is fully contained within the *UNC5D* (unc-5 homolog D, *C. elegans*) gene which must be disrupted by the breakpoint.

Case 54: 46,XY,t(14;15)(q23;q26.3)*de novo*.ish t(14;15)(q22.3;q26.2)
This male patient was first referred for cytogenetic analysis at 3.5 months of age because of coarctation of the aorta and dysmorphic features. The 14q22.3 breakpoint was mapped to an 106 kb region distal to fosmid WI2-815L3 and proximal to fosmid WI2-3837K2. According to Ensembl, part of this region (2%) contains the predicted gene *Q6NVV1_HUMAN* (no description). However, in the UCSC browser this region contains no genes. The 15q26.2 breakpoint was mapped to a 36 kb region within fosmid WI2-1929I14. Part of this clone (40%) contains the *NR2F2* (nuclear receptor subfamily 2, group F, member 2) gene.

Case 55: 46,XY,t(19;20)(q13.43;q11.1)*de novo*.ish t(19;20)(q13.43;q11.1)
This male patient was referred at 35 years of age because of severe oligospermia, but had no other clinical abnormalities. The 19q13.43 breakpoint was mapped to a 44 kb region within fosmid WI2-452J8. Part of this clone (60%) contains the *ZNF606* (zinc finger protein 606) gene. The 20q11.1 breakpoint was mapped to an 165 kb region within PAC RP4-610C12. According to the UCSC browser, no genes map to this clone. In Ensembl, a portion of the clone (5%) contains the predicted gene of unknown function *Q6ZS48_HUMAN*.

Case 56: 46,XY,t(6;21)(q16.2;q11.2)*de novo*.ish t(6;21)(q16.1;q11.1)
This male patient was referred at 30 years of age because of severe oligoteratospermia. The 6q16.1 breakpoint mapped to an 82 kb region distal to fosmid WI2-2659I10 and proximal to fosmid WI2-2846G24. Part of this interval (15%) contains the *C6orf167* (chromosome 6 open reading frame 167) gene. The 21q11.2 breakpoint was mapped to within centromeric sequences with plasmid L1.26.

Case 57: 46,XY,t(2;5)(p23;q11.2)*de novo*,t(18;22)(q11.2;p13)*de novo*.ish t(2;5)(p23.3;q11.2),t(18;22)(q11.2;p13)

This male patient was referred at 34 years of age because of epilepsy, short stature and macrocephaly. He had two independent apparently balanced reciprocal translocations and his karyotype was: 46,XY,t(2;5)(p23;q11.2)*de novo*,t(18;22)(q11.2;p13)*de novo*. In the t(2;5)(p23.3;q11.2), the 2p23.3 breakpoint was mapped to a 42 kb region within fosmid WI2-1700C17. Part of this clone (50%) contains the *NCOA1* (nuclear receptor coactivator 1) gene. The 5q11.2 breakpoint was mapped to a 368 kb region, distal to BAC RP11-352H23 and proximal to BAC RP11-313I12, which contains no known genes. In the t(18;22)(q11.2;p13), the 18q11.2 breakpoint was mapped to an 174 kb region within BAC RP11-467C13. This region is fully contained within the *ZNF521* (zinc finger protein 521) gene which must be disrupted by the breakpoint. The chromosome 22 breakpoint was mapped near the satellite stalk.

Case 58: 46,XX,t(6;18)(p23;q22)*de novo*.ish t(6;18)(p24.3;q21.1)

This female patient was referred for cytogenetic analysis because of primary amenorrhoea. The 6p24.3 breakpoint was mapped to a 99 kb region within PAC RP1-90O12, which contains no known genes. The 18q21.1 breakpoint was mapped to an 166 kb interval within BAC RP11-346H17. This clone is fully contained within the *DCC* (deleted colorectal carcinoma) gene, which must therefore be disrupted by the breakpoint.

5.2.1 Breakpoint mapping in case 49: 46,XX,t(2;10)(q33;q21.2)*de novo*.ish t(2;10)(q33.1;q21.3)

This female patient was referred at 29 years of age because of severe developmental delay and cleft palate. The results of breakpoint mapping studies by array painting are shown in figures 5.1 and 5.2 for clones mapped to chromosomes 2 and 10, respectively.

Analysis of figure 5.1 shows that clones mapped from 2pter to 2q33.1 have high values for the fluorescence ratios Cy3/Cy5, whereas clones mapped from 2q33.1 to 2qter have low values. This is consistent with translocation of 2q33.1 to 2qter sequences from the derivative 2 (labelled in Cy3) to the derivative 10 (labelled in Cy5). Conversely, the array painting profile for clones mapped to chromosome 10 (figure 5.2) shows a transition from low to high values of Cy3/Cy5 for clones mapped from 10q21.3 to 10qter.

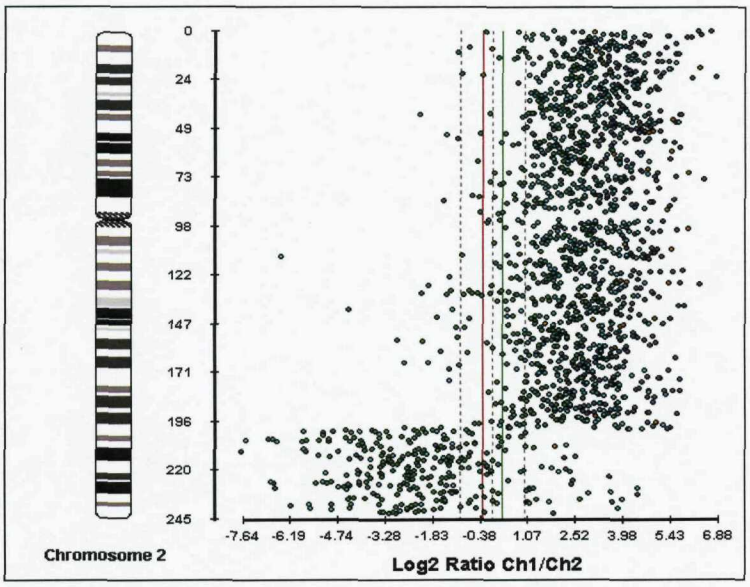

Fig.5.1. Array painting results for mapping of the 2q33.1 breakpoint in case 49.

This image is a screen shot from the BlueFuse software. The $\log_2$ fluorescence ratio Cy3/Cy5 is given in the x-axis, whereas the genomic position of clones from 2pter to 2qter is in the y-axis. Note the ratio transition from high to low values at 2q33.1.

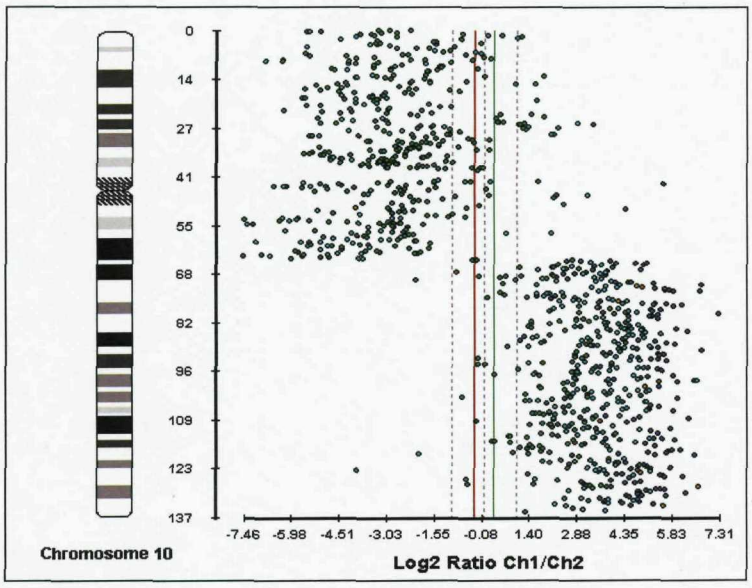

Fig.5.2. Array painting results for mapping of the 10q21.3 breakpoint in case 49.

This image is a screen shot from the BlueFuse software. Log2 fluorescence ratios Cy3/Cy5 are given in the x-axis. Genomic positions of clones from 10pter to 10qter are given in the y-axis. Note the ratio transition from low to high values at 10q21.3.

Following the initial analysis of the array painting profiles given above, a detailed examination of the values of fluorescence intensities was undertaken. The fluorescence values for the clones mapped near the breakpoint regions are presented in Table 5.I, which shows that a transition in ratio values occurred between BACs RP11-486F17 and RP11-530J6 on chromosome 2. On chromosome 10, the transition occurred between BACs RP11-436D10 and RP11-132E18. These clones were then applied in FISH experiments and the results are summarised in Table 5.II and illustrated in figures 5.3 and 5.4.

Table 5.I. Fluorescence ratios for clones mapped near the 2q33.1 and 10q23.1 breakpoints in case 49

Chr	Clone name	Amplitude Cy3	Amplitude Cy5	Log_2 ratio Cy3/Cy5
2	RP11-505H14	1389.198	195.59	2.828
2	RP11-486F17	894.728	349.683	1.355
2	RP11-530J6	159.478	592.672	-1.894
10	RP11-436D10	214.223	765.539	-1.837
10	RP11-132E18	736.449	567.297	0.376
10	RP11-144G16	709.365	338.424	1.068

Table 5.II. FISH results with BAC clones mapped near the breakpoints on 2q33.1 and 10q21.3

Chr	Clone name	2	der(2)	der(10)	Breakpoint position[1]
2	RP11-505H14	+	+	-	telomeric
2	RP11-486F17	+	+	+	within
2	RP11-530J6	+	-	+	centromeric
		10	**der(10)**	**der(2)**	
10	RP11-436D10	+	+	-	telomeric
10	RP11-132E18	+	+	+	within
10	RP11-144G16	+	-	+	centromeric

[1] Position of the breakpoint region in relation to the FISH clone tested.

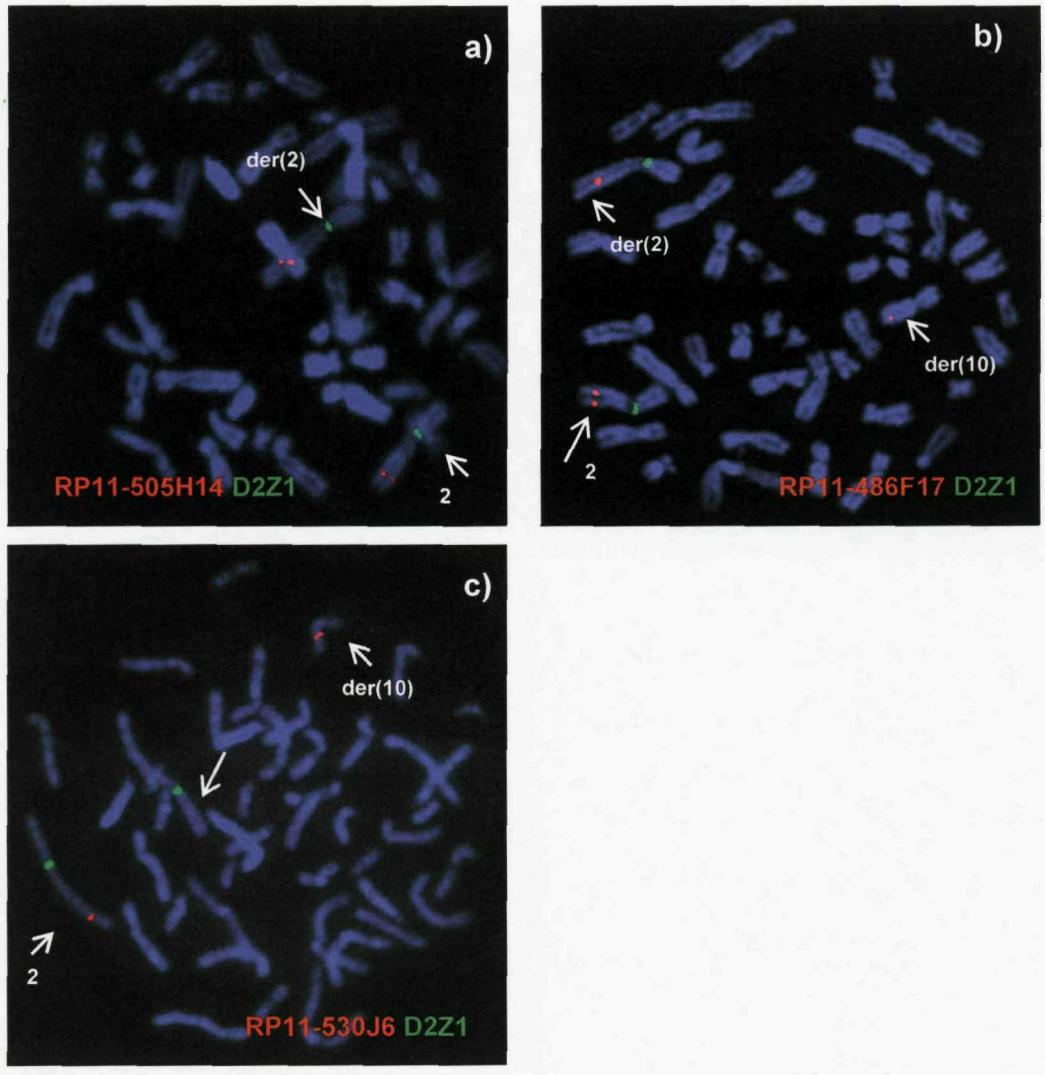

Fig.5.3. FISH results at the 2q33.1 breakpoint in case 49.

a) The breakpoint is mapped telomeric to BAC RP11-505H14. b) BAC RP11-486F17 is split by the breakpoint and shows hybridisation signals on the normal chromosome 2, and on both derivatives 2 and 10. c) the breakpoint lays centromeric to BAC RP11-530J6.

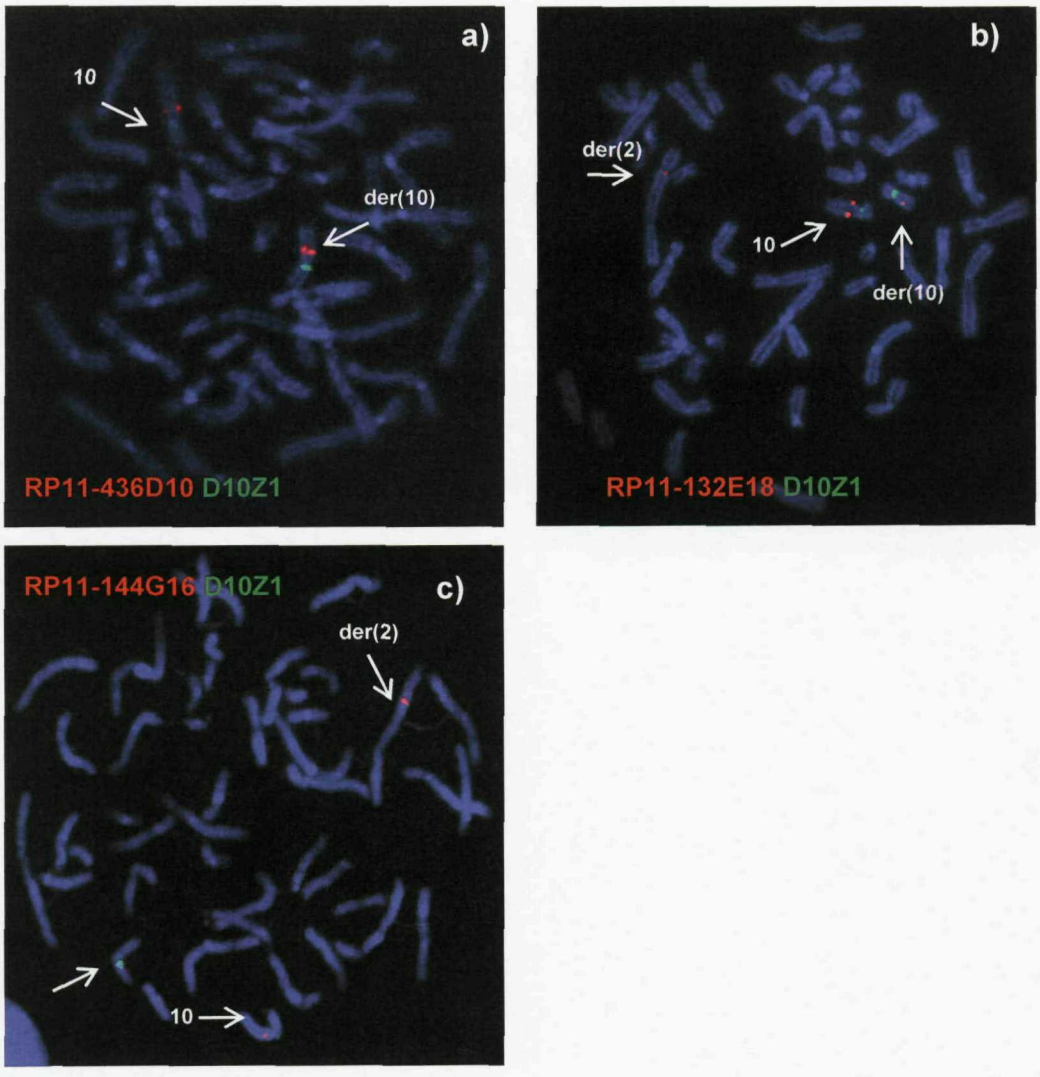

Fig.5.4. FISH results at the 10q23.1 breakpoint in case 49.

a) The breakpoint is mapped telomeric to BAC RP11-436D10. b) BAC RP11-132E18 is split by the breakpoint. c) The breakpoint is located centromeric to BAC RP11-144G16.

In summary, BAC RP11-486F17 was split by the 2q33.1 breakpoint and BAC RP11-132E18 by the 10q21.3 breakpoint. The regions defined by these clones were visualised in the Ensembl and USCS human genome browsers in a search for known genes. No known genes mapped to the 10q23.1 breakpoint region. However, on 2q33.1, part of the breakpoint region (80%) co-localised to the *SATB2* (SATB homeobox 2) gene. Thus, it was of interest to further refine this breakpoint region and obtain a more accurate estimate of the likelihood of disruption of *SATB2* by the breakpoint. This was done by selecting fosmid clones in the region defined by BAC RP11-486F17. Initially, three fosmid clones (WI2-1418M13, WI2-1367G19 and WI2-648C8) were selected for FISH. This analysis mapped the breakpoint to the 37 kb region defined by fosmid WI2-1367G19 (see figure 5.5). This clone was fully contained within *SATB2*; therefore the gene must be disrupted by the breakpoint.

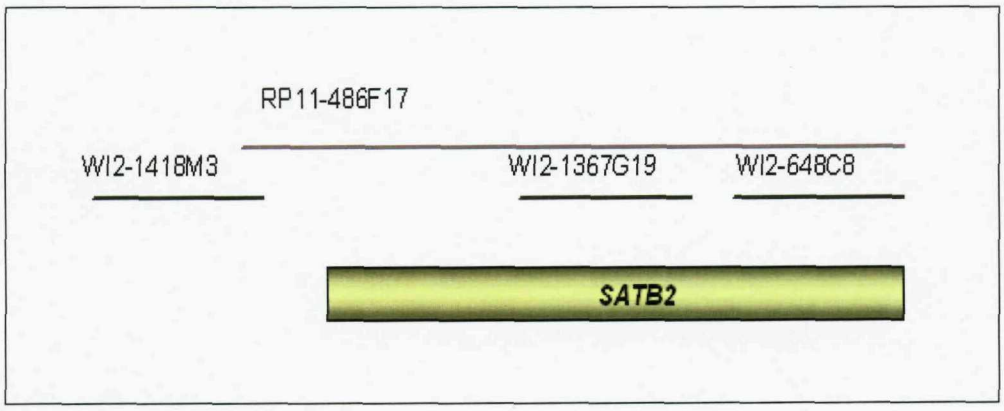

Fig.5.5. Assembly of fosmid clones mapped to 2q33.1 in case 49.
This figure shows the relative order of the fosmid clones (with prefix WI2) mapped within the breakpoint spanning BAC RP11-486F17 in the 2q breakpoint region of case 49. The position of the *SATB2* gene is also given. Fosmid WI2-1367G19 was split by the breakpoint.

5.2.1.1 High resolution mapping and sequence analysis of the breakpoints in case 49

The finding that the *SATB2* gene was interrupted in case 49 suggested that this is the cause of the cleft palate seen in this patient, as *SATB2* has been reported to cause this phenotype in patients with translocation breakpoints in the vicinity of the gene (Brewer *et al.*, 1999; FitzPatrick *et al.*, 2003). Further characterisation of this breakpoint was undertaken with the aim of excluding the presence of microrearrangements in *SATB2*. Firstly, the breakpoints were refined by high resolution array painting using the Agilent 244K oligonucleotide array. This analysis refined the 2q33.1 breakpoint to a 15 kb interval distal to probe A_14_P134396 and proximal to A_16_P15995021. The 10q23.1 breakpoint was refined to a 25 kb region distal to probe A_16_P18952001 and proximal to A_16_P18952049. The array painting profiles for chromosomes 2 and 10 are given in figures 5.6 and 5.7, respectively.

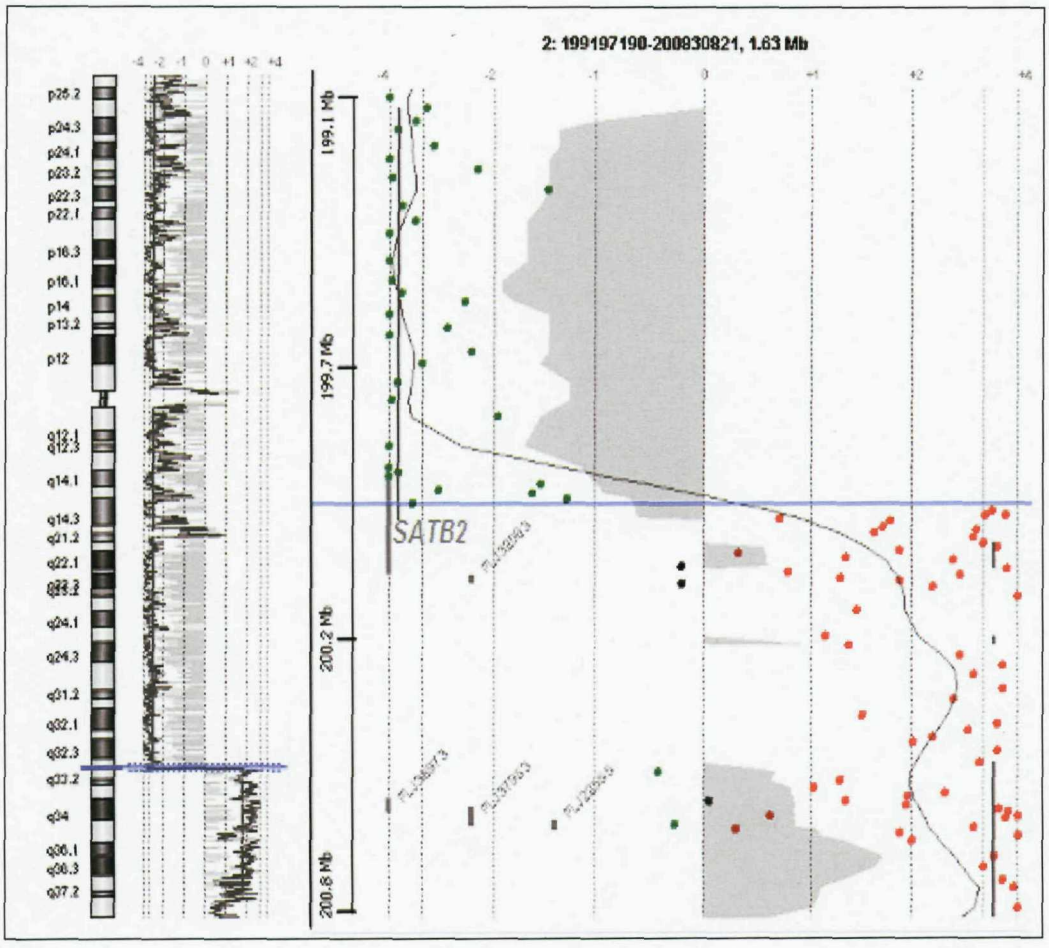

Fig.5.6. High resolution array painting results at the 2q33.1 breakpoint in case 49.

This is a screen shot from the Agilent software. The results for the whole chromosome 2 are shown on the left side alongside an idiogram. Fluorescence ratios of Cy5/Cy3 are given in the x-axis, for probes mapped from 2pter to 2qter. A zoomed-in view of the breakpoint region is given on the right hand side of the figure. Note the ratio transition, from negative (probes shown in green) to positive values (probes shown in red), indicating the breakpoint region within the sequence of *SATB2*.

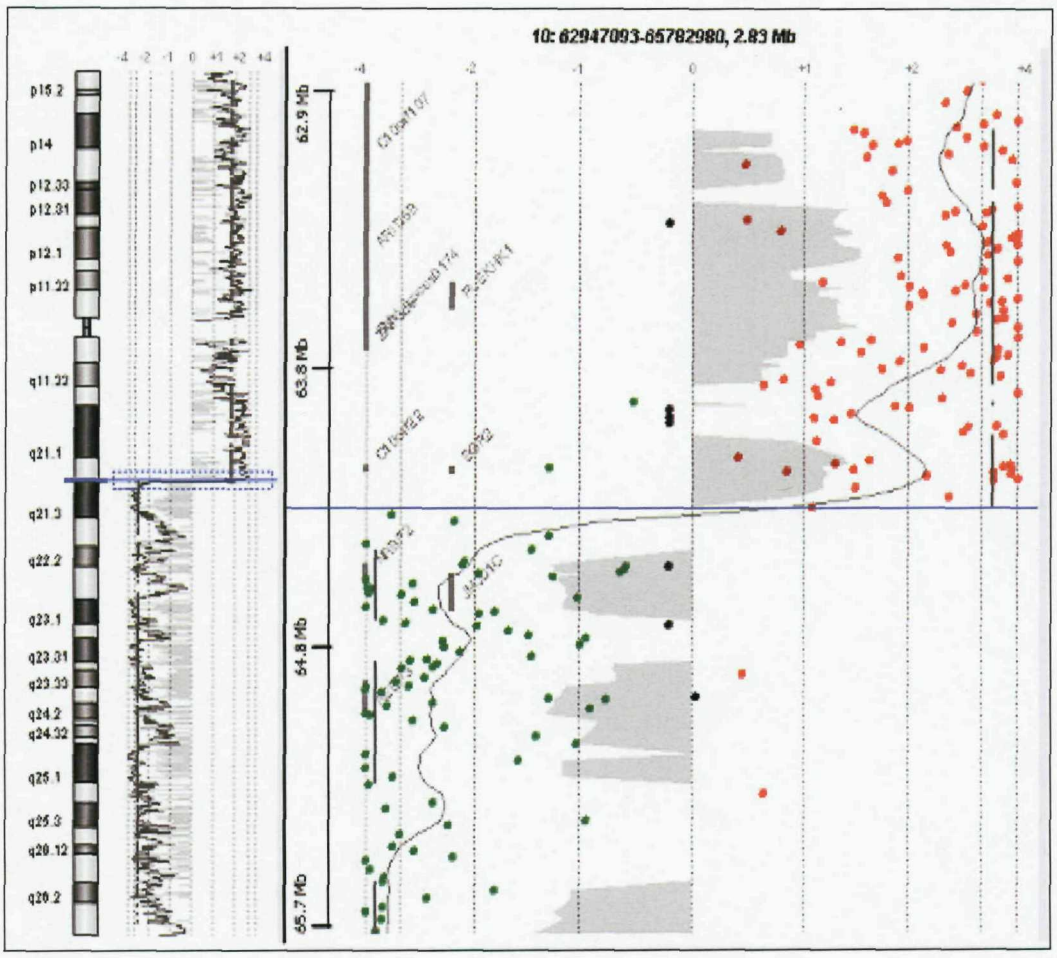

Fig.5.7. High resolution array painting results at the 10q23.1 breakpoint in case 49.

This is a screen shot from the Agilent software. The results for the whole chromosome 10 are shown on the left side alongside an idiogram of chromosome. Fluorescence ratios of Cy5/Cy3 are given in the x-axis, for probes mapped from 10pter to 10qter. A zoomed-in view of the breakpoint region is given on the right hand side of the figure. Note the ratio transition, from positive (probes in red) to negative (probes in green) values, which indicates the breakpoint region.

Following the refining of the breakpoints by high resolution array painting, a PCR based approach was applied (see section 3.4.2). Briefly, for mapping of the chromosome 2 breakpoint, primer pairs mapped to chromosome 2 were used to amplify products from both derivatives 2 and 10. The amplification of a product from the derivative 2 would indicate that the breakpoint mapped telomeric to the primer pair tested, whereas the presence of a product from the derivative 10 would indicate that the breakpoint was centromeric to the primers. This same approach was utilised to map the chromosome 10 breakpoint. The results of these analyses are given in Table 5.III, which shows that the 2q33.1 breakpoint mapped centromeric to primer P2_chr2 and the 10q23.1 breakpoint was centromeric to primer pair P19_chr10. Additionally, Table 5.III shows that several primer pairs failed to amplify a product from either derivative chromosome. Identical results were seen in replicate experiments. However, these same primers amplified a product from the patient's genomic DNA. The lack of a PCR product when the derivative DNA is used as a PCR template could be due to a deletion in the derivatives or it could be an experimental artefact caused by sub-optimal quality of the derivative DNA due to degradation or incomplete amplification by DOP PCR.

Table 5.III. Mapping of the breakpoints in case 49 by a PCR approach

Chr	Primer	Primer genomic start position (bps)	Primer position[1]	PCR result
2	P2_chr2	199,899,026	2.2	der(10)
2	P3_chr2	199,900,621	3.8	der(10)
2	P4_chr2	199,901,839	5.0	der(10)
2	P5_chr2	199,902,688	5.9	der(10)
2	P6_chr2	199,904,568	7.8	np
2	P7_chr2	199,906,717	9.9	np
2	P8_chr2	199,908,777	12.0	np
2	P9_chr2	199,909,138	12.3	der(10)
2	P10_chr2	199,910,209	13.4	der(10)
2	P11_chr2	199,911,389	14.6	der(10)
10	P12_chr10	64,366,013	0.98	np
10	P13_chr10	64,368,939	3.9	np
10	P14_chr10	64,370,891	5.8	np
10	P15_chr10	64,372,916	7.8	np
10	P16_chr10	64,372,951	8.0	np
10	P17_chr10	64,375,013	9.9	np
10	P18_chr10	64,375,834	10.8	np
10	P19_chr10	64,385,923	20.8	der(2)

[1]Primer position from the start of the breakpoint region as defined by the 244K-AP experiment

AP Array painting

np no product

Finally, long range PCR was undertaken in an attempt to analyse the chromosome 2 breakpoint region at the sequence level. The use of a long range PCR strategy (described in chapter 3) aimed to amplify the breakpoint junction fragment on the derivative 2, by using a left primer mapped to chromosome 2 and a right primer mapped to chromosome 10. Following the testing of several primer combinations (primers listed in Appendix 6), a product of ~10 kb was obtained when the left primer P2_chr2 was tested with the right primer P16_chr10 (figure 5.8). This result indicated that primer P2_chr2 mapped to the derivative 2, a finding that was inconsistent with the previous observation of

a product from this primer on the derivative 10, but not on the derivative 2 (Table 5.III). The reason for this inconsistency is unclear. So, to exclude the possibility of this junction fragment product being a PCR artefact, the product was sequenced from both ends. The results confirmed that the 5' end mapped to chromosome 2 and the 3' end to chromosome 10, in the expected orientation. The sequence of this product was further determined by using additional primers from both ends, but the sequence across the breakpoint was not obtained due to results inconsistent with the location of the right primers tested. The results of this analysis with right primers from chromosome 2 are given in Table 5.IV and illustrated in figure 5.9. These results suggest the presence of a microrearrangement, including an inversion in this region, but future studies will be needed to elucidate its exact nature.

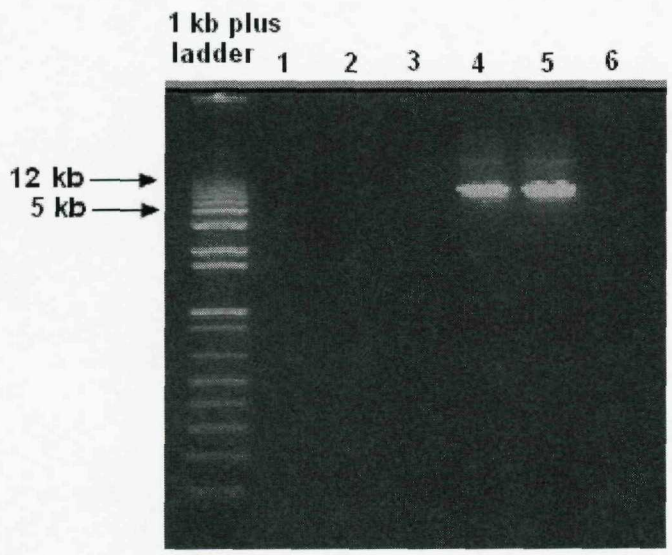

Fig.5.8. Long range PCR amplification of a breakpoint junction fragment at 2q33.1 in case 49.

This figure shows that a chr 2 breakpoint junction fragment from the genomic DNA of case 49 was obtained when the left primer P2_chr2 was mixed with the right primer P16_chr10 (lanes 4 and 5). This primer combination did not amplify a product from normal DNA (lane 6), showing the specificity of the primers for the derivative 2. No product was obtained when using the primer combination: P2_chr2 (left primer) /P15_chr10 (right primer) on case 49's genomic DNA (lanes 1-2) or in a normal individual (lane 3).

Table 5.IV. Sequence analysis of the long range PCR product mapped to the breakpoint junction on the derivative 2

Primer	Primer genomic start position (bps)	Genomic position of sequence product	Sequence product ID from fig. 5.9
P2_chr2	199,899,026	199,899,169-199,899,580	A
P41_chr2	199,899,254	199,899,345-199,899,824	B
P42_chr2	199,899,600	199,899,620-199,900,185	C
P43_chr2	199,900,089	199,900,110-199,900,700	D
P3_chr2	199,900,621	199,900,544-199,901,191	E
P4_chr2	199,901,839	199,901,862-199,902,412	F
P5_chr2	199,902,688	199,902,732-199,903,277	G
P6_chr2	199,904,568	np	np
P7_chr2	199,906,717	199,899,053-199,899,618	A-B
P8_chr2	199,908,777	199,901,523-199,901,946	F
P9_chr2	199,909,138	np	np
P10_chr2	199,910,209	np	np
P11_chr2	199,911,389	np	np

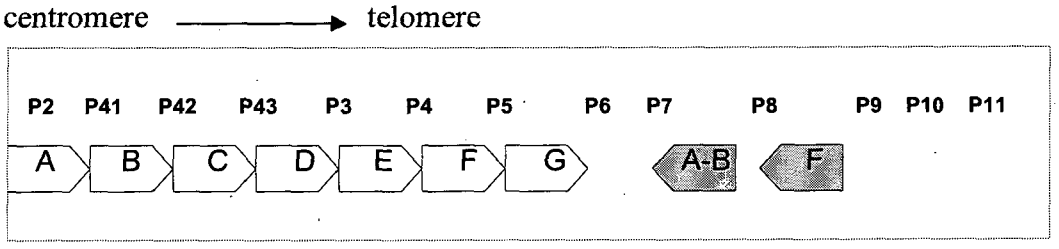

Fig.5.9. Summary of the 2q33.1 breakpoint region in case 49 according to the results of sequence analysis.

Sequence products are labelled A to G and the primers used to amplify these are given according to the data from table 5.IV. Note that primers P6, P9, P10 and P11 failed to amplify a product, whereas P7 and P8 amplified products mapped centromeric to their position, a result that is inconsistent with the primer assembly.

5.2.2 Breakpoint mapping in case 51: 47,XX,t(4;20)(p15.2;p11.23)*de novo*,+mar[23]mat /46,XX,t(4;20)(p15.2;p11.23)[7].ish t(4;20)(p15.32;p12.2)

This female patient was referred at 6 years of age because of developmental delay, autistic spectrum disorder and attention deficit hyperactive disorder (ADHD). She had an apparently balanced t(4;20)(p15.2;p11.23)*de novo* in all of her cells, and a marker derived from chromosome 22 in 70% of the cells analysed. This marker chromosome, which was mainly composed of heterochromatin, was also present in her phenotypically normal mother and maternal grandmother, thus it is most likely that this marker is coincidental to the clinical abnormalities in case 51. The translocation breakpoints were mapped by array painting. The results of this analysis are given in figures 5.10 and 5.11, for chromosomes 4 and 20, respectively.

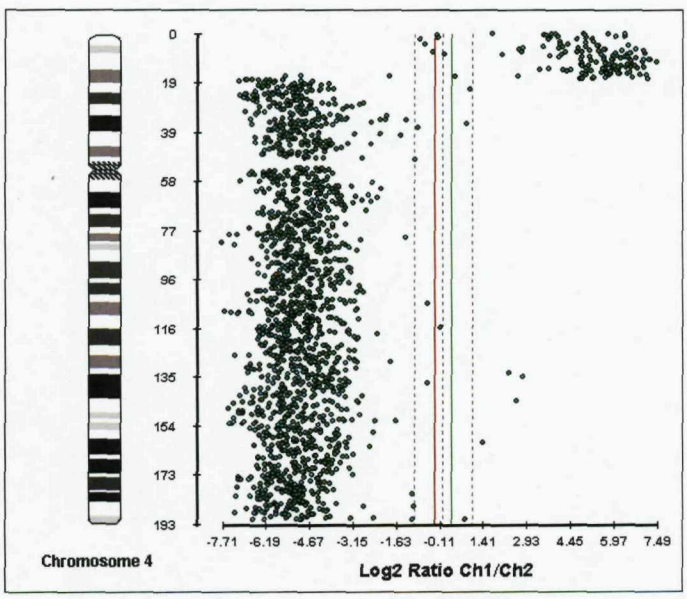

Fig.5.10. Array painting results for mapping of the 4p15.32 breakpoint in case 51.

Log2 fluorescence ratios Cy3/Cy5 are given in the x-axis. Genomic positions of clones from 4pter to 4qter are given in the y-axis. Note the ratio transition from high to low values at 4p15.32.

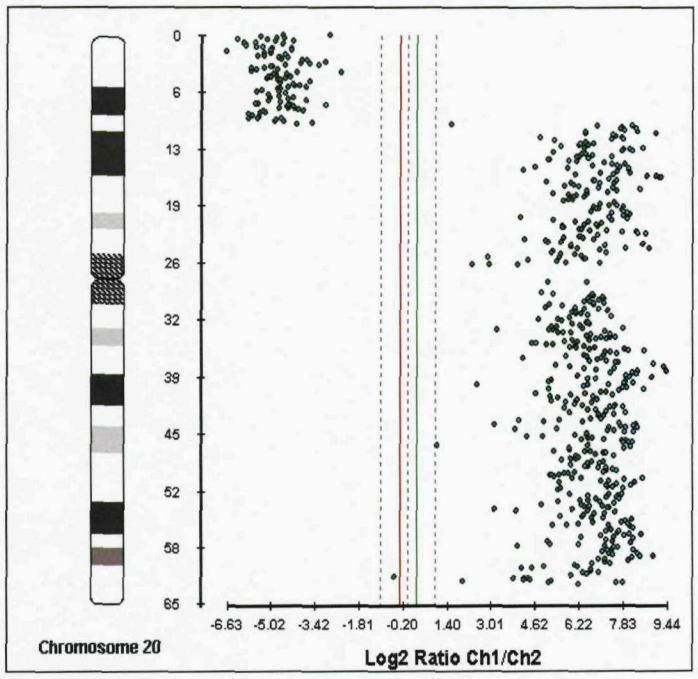

Fig.5.11. Array painting results for the mapping of the 20p12.2 breakpoint in case 51.

Log2 fluorescence ratios Cy3/Cy5 are given in the x-axis. Genomic positions of clones from 20pter to 20qter are given in the y-axis. Note the ratio transition from low to high values at 20p12.2.

Analysis of figure 5.10 shows that a fluorescence ratio Cy3/Cy5 transition from high to low values occurred for clones mapped from 4p15.32 to 4qter. This is consistent with translocation of 4p15.32 to 4pter sequences from the derivative 4 (labelled in Cy5) to the derivative 20 (labelled in Cy3). Clones mapped to chromosome 20 (figure 5.11) showed a transition from low to high fluorescence ratios of Cy3/Cy5 for clones mapped from 20p12.2 to 2qter. The fluorescence ratios for clones mapped to the regions of transition is given in Table 5.V.

Table 5.V. Fluorescence ratios for clones mapped near the 4p15.32 and 20p12.2 breakpoints in case 51

Chromosome	Clone name	Amplitude Cy3	Amplitude Cy5	Log_2 ratio Cy3/Cy5
4p15.32	RP11-141E5	6860.644	243.791	4.815
4p15.32	RP11-210E17	84.635	9754.941	-6.849
20p12.2	RP11-416N4	289	4762.931	-4.043
20p12.2	RP5-1068F16	12934.56	4273.785	1.598
20p12.2	RP5-931K24	16452.848	132.908	6.952

FISH analysis with the clones in the regions of fluorescence ratio transition mapped the 4p15.32 breakpoint to a 327 kb region, proximal to BAC RP11-141E5 and distal to BAC RP11-210E17. No known genes mapped to this region. The 20p12.2 breakpoint mapped to a 140 kb region within PAC RP5-1068F16. Part of this clone (60%) contained the *SNAP25* (synaptosomal-associated protein, 25kDa) gene. To obtain a better estimate of the likelihood of disruption of *SNAP25*, this breakpoint region was further mapped with fosmid clones. The breakpoint mapped to a 42 kb region defined by fosmid WI2-938I7 and the likelihood of disruption of *SNAP25* was estimated as 85%. *SNAP25* has been implicated in ADHD (Feng *et al.*, 2005) and since case 51 presented with this disorder, it was of interest to refine this breakpoint region further. These analyses were undertaken by amplification of sequences from both derivatives using the PCR based approach described above for case 49 (see sections 5.2.1.1 and 3.4.2). Briefly, the breakpoint would map telomeric to primers amplifying products from the derivative 4 and centromeric to primers amplifying products from the derivative 20. Table 5.VI summarises the results obtained.

Table 5.VI. Mapping of the 20p12.2 breakpoint in case 51 by a PCR approach

Primer	Primer genomic start position (bps)	Primer position[1](kb)	PCR Result
P33_chr20	10,212,934	3.9	np
P34_chr20	10,219,303	10	der(4)
P35_chr20	10,221,614	12.6	der(4)
P36_chr20	10,225,476	16.5	der(4)
P37_chr20	10,226,453	18	np
P38_chr20	10,227,931	18.9	np
P39_chr20	10,234,611	25.6	der(20)
P40_chr20	10,234,842	26	der(20)

[1]Primer position from the start of the breakpoint region defined by fosmid WI2-938I7
AP Array Painting
np no product

Table 5.VI shows that the 20p12.2 breakpoint mapped between primers P36_chr20 and P39_chr20, in a 9.4 kb interval fully contained within the *SNAP25* gene. Primers mapped between the aforementioned breakpoint flanking primers failed to amplify a product from any of the derivatives, whilst amplifying a product when genomic DNA was used as a PCR template. As discussed for case 49 (section 5.2.1.1), this result could be due to a deletion in the derivative DNA, or to a technical artefact generated by the use of flow sorted and amplified derivative DNA. Futures studies will be needed to clarify these results.

5.2.3 Breakpoint mapping in case 52: 46,XX inv ins (11;4)(q22.2;q13.2q21.3)*de novo*.ish inv ins (11;4)(q22.1;q13.1q21.23)

This female patient was referred for cytogenetic analysis at 12 years of age because of learning difficulties. She also presented with laryngomalacia with vocal cord palsy (requiring tracheostomy), short stature and scoliosis. She had an insertional translocation with two breakpoints on 4q and insertion of the

interstitial material into 11q as defined by the karyotype: 46,XX inv ins (11;4)(q22.2;q13.2q21.3)*de novo*. The results of the array painting analysis for mapping of the chromosome 4 breakpoints are shown in figure 5.12.

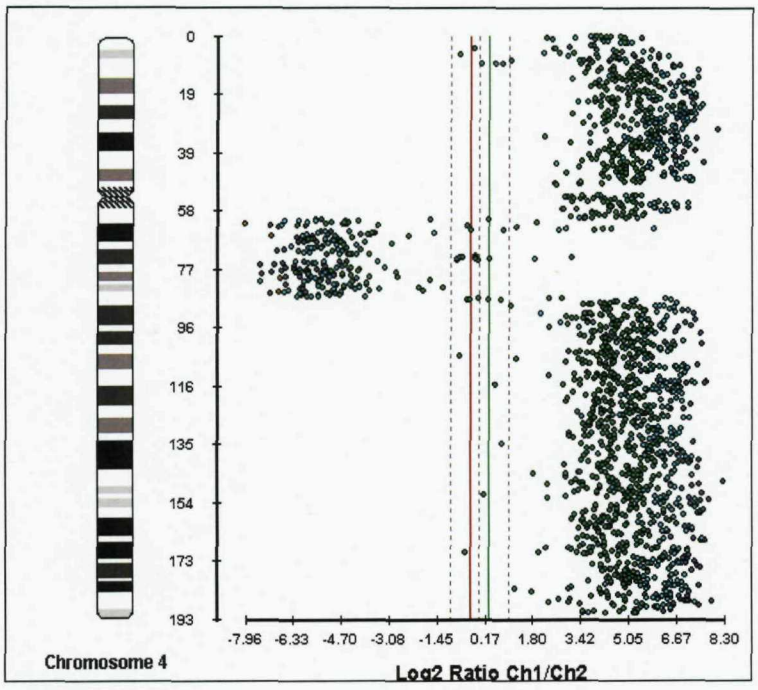

Fig.5.12. Array painting results for mapping of the chromosome 4 breakpoints in case 52.

Log_2 fluorescence ratios Cy3/Cy5 are given in the x-axis. Genomic positions of clones from 4pter to 4qter are given in the y-axis. Note the ratio transition for clones mapped between 4q13.1 and 4q21.23.

Figure 5.12 shows that clones mapped between 4q13.1 and 4q21.23 show a transition of fluorescence ratios Cy3/Cy5 from high to low values. This is consistent with insertion of these sequences from the derivative 4 (labelled in Cy3) to the derivative 11 (labelled in Cy5). Analysis of the fluorescence ratios for clones mapped near the 4q13.1 breakpoint showed a ratio transition between BACs RP11-467I16 and RP11-357L18 (Table 5.VII).

Table 5.VII. Fluorescence ratios for clones mapped near the 4q13.1 breakpoint in case 52

Chr	Clone name	Amplitude Cy3	Amplitude Cy5	Log_2 ratio Cy3/Cy5
4q13.1	RP11-467I16	3357.805	1819.686	0.884
4q13.1	RP11-357L18	2002.261	2351.783	-0.232
4q13.1	RP11-352A20	86.281	3167.113	-5.198

FISH analysis confirmed that RP11-467I16 and RP11-357L18 were spanning the 4q13.1 breakpoint, which mapped to a 184 kb region with no known genes. A similar analysis for clones near the 4q21.23 breakpoint mapped this breakpoint region to a 201 kb interval within BAC RP11-350M17. This clone is fully contained within the *MAPK10* (mitogen-activated protein kinase 10) gene which must be disrupted by the breakpoint. Furthermore, FISH probes mapped in the vicinity of the 4q21.23 breakpoint gave results inconsistent with their location suggesting the presence of an additional rearrangement. A diagram of the BAC clones tested and their assembly is given in figure 5.13 and FISH results are summarised in Table 5.VIII.

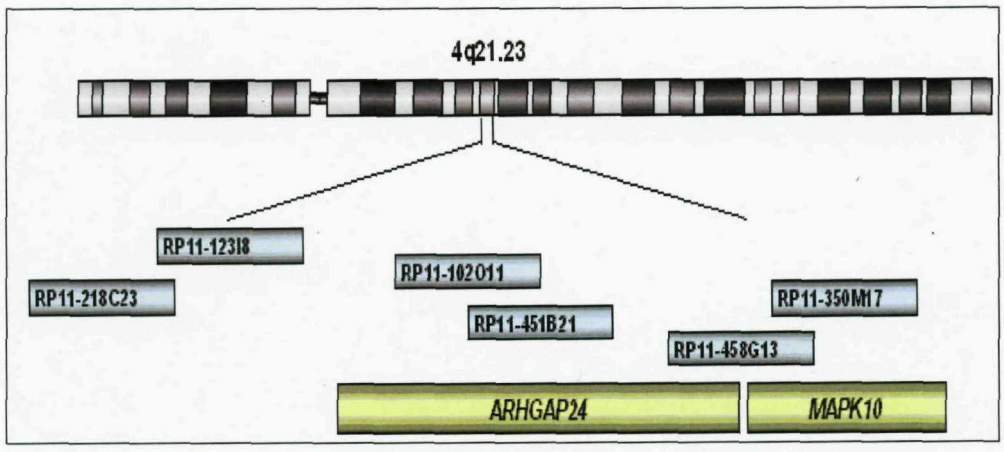

Fig.5.13. Assembly of BAC clones mapped to the 4q21.23 breakpoint in case 52.

The diagram shows the BAC clone assembly and the gene content of this region (*ARHGAP24* and *MAPK10*).

Table 5.VIII. FISH results for clones mapped near the 4q21.23 breakpoint in case 52

Chr	Clone name	4	der(4)	der(11)	Breakpoint position[1]
4q21.23	RP11-218C23	+	-	+	telomeric
4q21.23	RP11-123I8	+	-	-	
4q21.23	RP11-102O11	+	+	+	within
4q21.23	RP11-451B21	+	+	-	centromeric
4q21.23-21.3	RP11-458G13	+	+	+	within
4q21.23-21.3	RP11-350M17	+	+	+	within

[1]Position of the breakpoint region in relation to the FISH clone tested.

Analysis of Table 5.VIII shows BAC RP11-123I8 is deleted. Furthermore, BAC RP11-451B21 is flanked by two breakpoint regions defined by BACs RP11-102O11 and RP11-458G13. This result could be explained by an error in the assembly in this region or by the presence of an inversion or other microrearrangement. However, due to the physical proximity of these clones, FISH analysis could not clarify these results. Future studies will be needed to define this region further.

The chromosome 11 breakpoint could not be mapped by array painting. This is because no sequences of this chromosome were translocated to the derivative 4 and as a result no fluorescence ratio transition occurred for these clones. Hence, breakpoint mapping was carried out by FISH, which showed that the 11q22.1 breakpoint mapped to a 316 kb region between BACs RP11-49M9 and RP11-199N10. Part of this region (40%) contains the *CNTN5* (contactin 5) gene.

5.3 Summary of the breakpoint mapping results

Table 5.IX summarises and compares the breakpoint locations determined by cytogenetic analysis and by FISH and shows that 8/34 of the breakpoints were assigned to be same cytogenetic band by both techniques, whereas the

remaining were reassigned either to a sub-band of the same band or to a different band altogether. Moreover, the majority of the breakpoints were re-assigned to a band adjacent to the cytogenetically determined breakpoint and only 2/34 were re-assigned to 2 or more bands away.

The details of the breakpoint regions are summarised in Table 5.X, which lists the genomic positions of the breakpoint-containing intervals (BCI [in bps]), the genes annotated within these and their likelihood of disruption according to both Ensembl and UCSC (estimated as described in section 4.2). Analysis of these results shows that only 2 patients had rearrangements in which both breakpoints mapped to regions with no known genes, 7 had breakpoint-mediated gene disruption by at least one breakpoint and the remaining 7 were potentially associated with breakpoint-mediated gene disruption. Furthermore, for the cases with no genes in the breakpoint regions, the nearest gene was mapped to <1.2 Mb to the breakpoint region, in all but one case (the chromosome 4 breakpoint in case 43). This interval is within the maximum range of 1.3 Mb for position effects (Velagaleti *et al.*, 2005), suggesting that candidate genes for the clinical abnormalities might well be present outside the breakpoint regions.

Table 5.IX. Comparison between the cytogenetically and the molecularly assigned breakpoints

Case	Cytogenetic breakpoint	Molecular breakpoint
16	10q24.3	10q24.31
16	**22q13.31**	**22q13.31**
20	2q33	2q33.2
20	**5q12**	**5q12**
43	4q35.1	4q34.3
43	17q25.1	17q 24.2
44	6p24	6p24.3
44	**6q16.2**	**6q16.2**
45	Xq21	Xq21.1
45	19p13.11	19q11
48	**4q33**	**4q33-34.1**
48	6q22.2	6q21
49	2q33	2q33.1
49	10q21.2	10q21.3
50	Xq22.1	Xq23
50	8q24.13	8q24.3
51	4p15.2	4p15.32
51	20p11.23	20p12.2
52	4q13.2	4q13.1
52	4q21.3	4q21.23
52	11q22.2	11q22.1
53	4q21.1	4q13.3
53	**8p12**	**8p12**
54	14q23	14q22.3
54	15q26.3	15q26.2
55	**19q13.43**	**19q13.43**
55	20q11.1	20q11
56	6q16.2	6q16.1
56	21q11.2	21q11.1
57	2p23	2p23.3
57	**5q11.2**	**5q11.2**
57	**18q11.2**	**18q11.2**
57	22p13	nd
58	6p23	6p24.3
58	18q22	18q21.1

Bold type indicates identical results by both techniques

Table 5.X. Breakpoint mapping results for the patient cohort

Case	BCI genomic position	BCI size	Genes in BCI	LOGD
16	chr10:102,470,424-102,509,479	39,055	*PAX2*	40
16	chr22:42,975,154-43,015,629	40,475	*KIAA1644*	100
20	chr2:203,884,002-203,963,572	79,570	*ABI2*	80
20	chr5:60,897,709-63,578,254	2,680,545	(6 genes)*	nd
43	chr4:181,557,356-181,712,629	155,273	----	----
43	chr17:62,750,206-62,824,693	74,487	*PSMD12, PITPNC1*	60
44	chr6:10,438,104-10,477,782	39,678	----	----
44	chr6:99,103,218-99,277,964	174,746	----	----
45	chrX:78,833,496-78,996,054	162,558	----	----
45	chr19: near centromere	nd	nd	nd
48	chr4:172,112,115-172,314,669	202,554	----	----
48	chr6:111,251,547-111,344,802	93,255	*AMD1*	20
49	chr2: 199,896,744-199,911,974	15,230	*SATB2*	100
49	chr10: 64,365,036-64,389,795	24,759	----	----
50	chrX:115,042,293-115,233,378	191,085	*AGTR2*	3
50	chr8:142,582,086-142,790,558	208,472	----	----
51	chr4:17,806,804-18,134,168	327,364	----	----
51	chr20:10,225,476-10,234,860	9,384	*SNAP25*	100
52	chr4:64,164,472-64,348,036	183,564	----	----
52	chr4:87,196,675-87,397,572	200,897	*MAPK10*	100
52	chr11:99,613,159-99,929,275	316,116	*CNTN5*	40
53	chr4:74,827,036- 77,105,689	2,278,653	(21 genes)*	nd
53	chr8:35,542,879-35,705,808	162,929	*UNC5D*	100
54	chr14:55,276,675-55,382,178	105,503	*Q6NVV1_H.*	0-2
54	chr15:94,659,356-94,695,383	36,027	*NR2F2*	40
55	chr19:63,177,283-63,220,873	43,590	*ZNF606*	60
55	chr20:28,033,231-28,197,751	164,520	*Q6ZS48_H.*	0-5
56	chr6:97,826,527-97,908,737	82,210	*C6orf167*	15
56	chr21:near centromere	nd	nd	nd
57	chr2:24,640,005-24,682,052	42,047	*NCOA1*	50
57	chr5:57,413,898-57,782,197	368,299	----	----
57	chr18:20,893,727-21,067,881	174,154	*ZNF521*	100
57	chr22:satellite stalk	nd	nd	nd
58	chr6:9,634,150-9,733,184	99,034	----	----
58	chr18:48,220,604-48,386,185	165,581	*DCC*	100

*Genes deleted due to breakpoint-associated deletions (see Table 5.XVI).
LOGD Likelihood of gene disruption
nd not determined

5.4 Further characterisation of the breakpoint regions

As explained in section 4.5 it was relevant to characterise the breakpoint regions further by determining the types of genes and also the genomic context of these regions, mainly with respect to the type of cytogenetic band involved in breakpoints and their content in segmental duplications, normal CNV and pathogenic imbalances.

5.4.1 The types of genes at the breakpoints

In silico analysis (as described in section 4.5.1) showed that the genes mapped to the breakpoint regions were implicated in various biological processes (see Appendix 9), of which the most common were related to signal transduction, the nervous system and transcription (Table 5.XI). Genes associated with Mendelian diseases or with susceptibility to complex disorders were also identified (Table 5.XII) and are discussed in the context of genotype-phenotype correlations in section 5.6.

Table 5.XI. Most common GO terms for genes mapped to the breakpoints in the patient cohort

GO term	Gene
signal transduction	*CXCL6*, *HTR1A*, *MAPK10*, *NCOA1*, *NR2F2*, *PITPNC1*, *UNC5D*
nervous system development, neuron migration, neurotransmitter secretion/uptake, axon guidance, axonogenesis	*CXCL1*, *DCC*, *KIF2A*, *NR2F2*, *PAX2*, *SNAP25*
transcription, DNA-dependent regulation of transcription, DNA-dependent positive regulation of transcription	*NCOA1*, *NR2F2*, *PAX2*, *SATB2*, *ZNF606*

Table 5.XII. Genes associated with known disorders

Case	Gene	Disorder
16	*PAX2*	Renal-coloboma syndrome [MIM 120330]
49	*SATB2*	Isolated cleft palate [MIM 119540]
50	*AGTR2*	X-linked mental retardation [MIM 300034]
51	*SNAP25*	Attention Deficit-Hyperactivity disorder [MIM 143465]
52	*MAPK10*	Epileptic encephalopathy Lennox-Gastaut type [MIM 606369]

5.4.2 Genomic context of the breakpoints

5.4.2.1 Cytogenetic bands

Table 5.XIII lists the type of cytogenetic bands involved in the breakpoints, and shows that 5 breakpoints mapped to pericentromeric bands, 1 to within a satellite region and 1 in a telomeric band, while fragile sites mapped to 5 of the breakpoint regions.

Table 5.XIII. Cytogenetic bands of interest in the patient cohort

Case	Cytogenetic band	Type of region[1]
20	2q33.2	FRA2I
45	19q11	pericentromeric
49	2q33.1	FRA2I
49	10q21.3	FRA10C
50	8q24.3	telomeric, FRA8D
51	20p12.2	FRA20B
55	20q11	pericentromeric
56	21q11.1	pericentromeric
57	5q11.2	pericentromeric
57	18q11.2	pericentromeric
57	22p13	satellite region

[1]Fragile sites have a prefix FRA and were mapped according to the GDB database

5.4.2.2 Segmental duplications

Segmental duplications were found to map to 11/32 (34%) of the breakpoint regions as shown in table 5.XIV.

5.4.2.3 Normal copy number variation

Sites of normal CNV were mapped to 17/32 (53%) of the breakpoints (Table 5.XIV).

5.4.2.4 Potentially disease-causative copy number changes

Six breakpoints (~19%) were mapped to potentially pathogenic regions of deletion or duplication identified in patients with clinical abnormalities according to the DECIPHER database (Table 5.XIV). Most of these regions await further characterisation, but these might well represent novel syndrome regions. The only defined syndrome among these was the 2q32.2 deletion syndrome region and two breakpoints (case 20 and 49) were found to map within this region.

Table 5.XIV. Genomic context of the breakpoint regions in the patient cohort

Case	Breakpoint region	SDs	normal CNV	DECIPHER data
16	10q24.31	-	+	-
16	22q13.31	+	+	+
20	2q33.2	+	-	2q32.2 del*
20	5q12	+	+	-
43	4q34.3	-	+	-
43	17q 24.2	-	+	-
44	6p24.3	+	-	-
44	6q16.2	-	-	-
45	Xq21.1	-	+	-
45	19q11	nd	nd	----
48	4q33-34.1	-	-	-
48	6q21	+	-	-
49	2q33.1	-	-	2q32.2 del*
49	10q21.3	-	-	-
50	Xq23	+	+	-
50	8q24.3	+	+	+
51	4p15.32	-	-	-
51	20p12.2	-	-	-
52	4q13.1	-	+	-
52	4q21.23	-	+	+
52	11q22.1	-	-	+
53	4q13.3	+	+	-
53	8p12	-	-	-
54	14q22.3	+	-	-
54	15q26.2	-	-	-
55	19q13.43	-	+	-
55	20q11	+	+	-
56	6q16.1	-	-	-
56	21q11.1	nd	nd	----
57	2p23.3	-	-	-
57	5q11.2	+	+	-
57	18q11.2	-	+	-
57	22p13	nd	nd	----
58	6p24.3	-	+	-
58	18q21.1	-	+	-

*2q32.2 deletion syndrome region
nd not determined

5.5 Array CGH results

Array CGH analysis with the Sanger 30K WGTP BAC array detected a large number (between 28 and 86) of regions of copy number change in all genomes analysed. These regions of copy number change were identified by the CNVfinder method described by Fiegler *et al.* (2006). Cases 44 and 58 were excluded from this analysis as the 244K Agilent oligonucleotide platform and not the 30K BAC array was applied in these cases (see chapter 3, Table 3.II). For the 244K oligonucleotide array, the CNVfinder method was not applicable due to the different array design, and so the search for regions of copy number change was done manually. The calling criteria included regions spanned by a minimum of 3 oligonucleotides with a $\log_2$ fluorescence ratio test/reference DNA of ±0.5. Case 44 showed a normal array CGH profile. The results for case 58 also suggested a normal result, but interpretation was hampered by experimental noise. This was due to sub-optimal quality of the DNA sample, which was extracted from archival fixed cells in 3:1 methanol:acetic acid instead of fresh blood (section 3.2).

A summary of the number of regions of copy number change detected is provided in Table 5.XV. In total, 188 genomic regions, spanned by 721 BAC clones were found to have copy number changes. These regions were compared to data from surveys of copy number variation in normal populations documented in the DGV (references are available within the database). The changes detected in this study which have been previously reported in normal populations were assumed to constitute normal CNV, while unreported changes, including at least two BAC clones were singled out for parental origin studies. Changes present in a clinically normal parent were considered to be normal variation, whereas *de novo* imbalances were considered to be candidate regions for a role in the phenotypic abnormalities present in the patients.

Table 5.XV. Number of regions of copy number change in the patient cohort

Case	Regions of loss	Regions of gain	Total regions of CNV
16	39	28	67
20	21	35	56
43	31	39	70
44	nd	nd	nd
45	33	39	72
48	26	34	60
49	13	16	29
50	24	41	65
51	14	23	37
52	35	28	63
53	34	51	85
54	7	21	28
55	28	30	58
56	18	27	45
57	28	31	59
58	nd	nd	nd

nd not determined

5.5.1 Candidate disease-causing imbalances

Candidate imbalances to explain the abnormal phenotypes were identified in cases 20, 52, 53 and 57. The DECIPHER database was used to determine whether these regions have been previously reported in patients with clinical abnormalities, but no records were retrieved. The gene content of these deletions was determined and a systematic analysis was undertaken with the GO database, OMIM, the Human Gene Mutation database, UniProt and GeneCards (URLs in Appendix 1), in an attempt to identify good candidate genes for the clinical abnormalities of the affected patients (see Table 5.XVI).

Case 20: 46,XX,t(2;5)(q33;q12)*de novo*

The array CGH results of case 20 are given in figure 5.14. A 2.5 Mb deletion was detected at the 5q translocation breakpoint, between 61,052,997 and 63,598,644 bps. This was shown, by FISH analysis, to have occurred *de novo.* Parental origin studies by microsatellite analysis determined that this deletion originated on the paternal chromosome.

Case 52: 46,XX inv ins (11;4)(q22.2;q13.2q21.3)*de novo*

The array CGH results of case 52 are given in figure 5.15. Two deletions were detected on 4q. The most proximal deletion, with a size of 1.8 Mb, mapped ~8.4 Mb distal to the 4q13.1 translocation breakpoint and ~12 Mb proximal to the 4q21.23 breakpoint, between 72,636,161 and 74,421,392 bps. The second deletion, with a size of ~170 kb mapped to the 4q21.23 breakpoint between 86,384,696 and 86,558,8920 bps. Both deletions were shown, by FISH analysis, to have occurred *de novo*. Parental origin studies by microsatellite analysis determined that both deletions originated on the paternal chromosome.

Case 53: 46,XX,t(4;8)(q21.1;p12)*de novo*

The array CGH results of case 53 are given in figure 5.16. A 2.1 Mb deletion was detected at the 4q translocation breakpoint between 74,932,950 and 77,060,297 bps and a 1.1 Mb deletion at 10p14 between 8,423,513 and 9,583,339 bps. Both deletions were shown, by microsatellite analysis, to have occurred *de novo*, with their origin on the paternal chromosome.

Case 57: 46,XY,t(2;5)(p23;q11.2)*de novo*, t(18;22)(q11.2;p13)*de novo*

The array CGH results of case 57 are given in figure 5.17. A 2 Mb deletion was detected on 4q32.1, between 156,373,857 and 158,388,330 bps. This was shown, by microsatellite analysis, to have occurred *de novo,* with origin on the paternal chromosome.

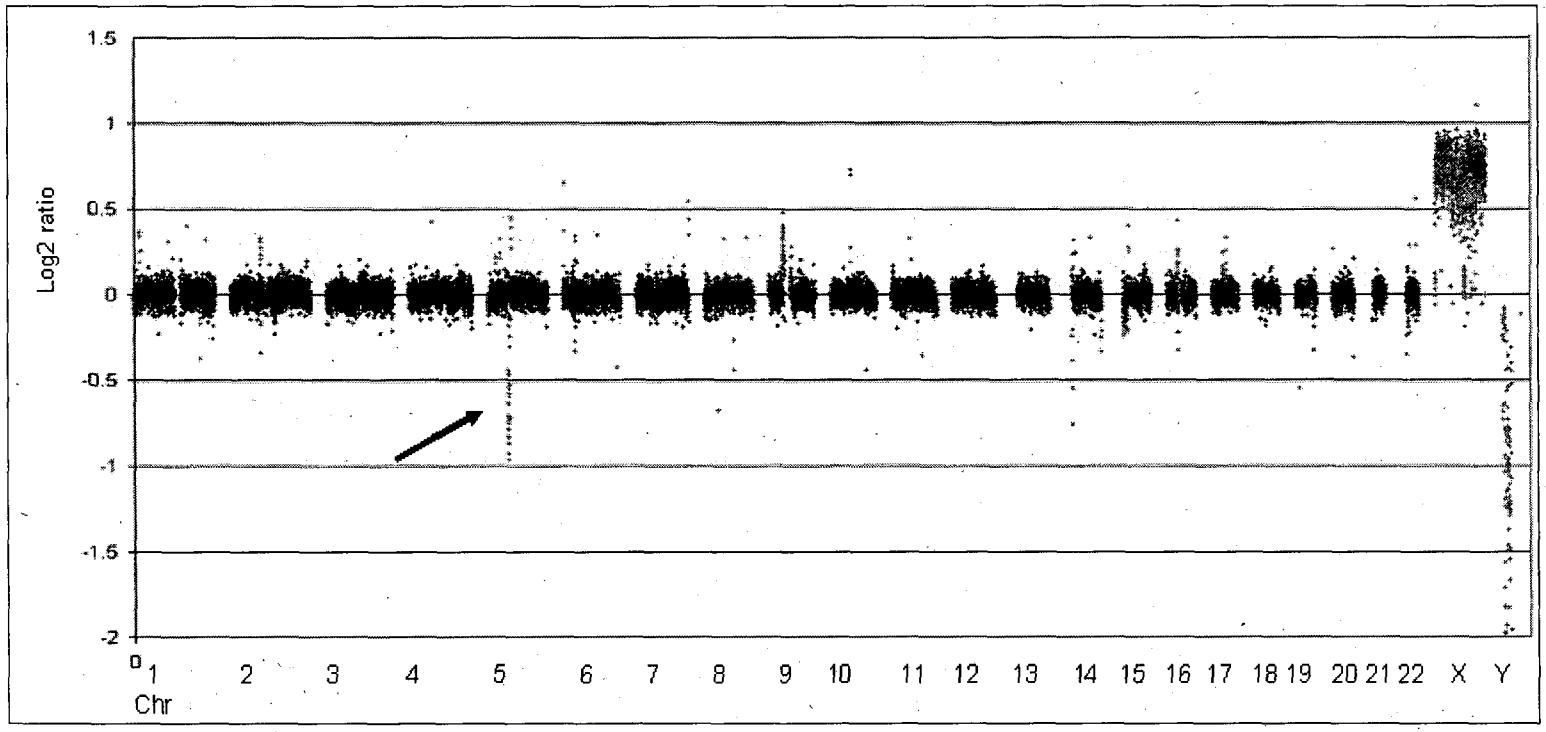

Fig.5.14. Array CGH results for case 20.

Genomic clones are shown in the x-axis and are grouped by chromosome 1 to Y. Log2 fluorescence ratios of test/reference DNA are given in the y-axis. The array CGH profile in case 20 shows the gain of sequences from chromosome X and loss of those from chromosome Y, consistent with the sex mismatch test/reference. Copy number changes are present throughout the genome. The arrow indicates the breakpoint-associated deletion at 5q.

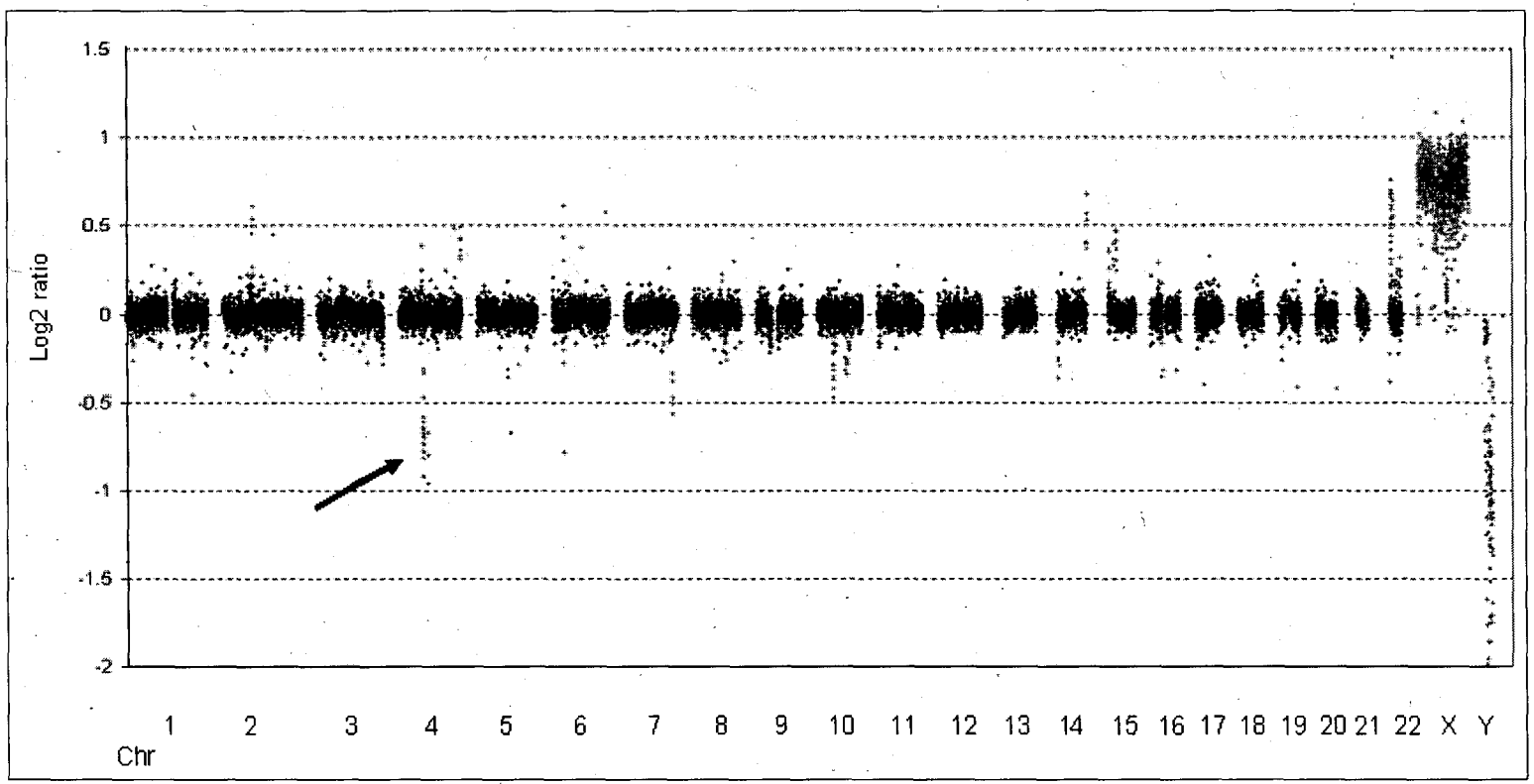

Fig.5.15. Array CGH results for case 52.

Genomic clones are shown in the x-axis and are grouped by chromosome 1 to Y. Log_2 fluorescence ratios of test/reference DNA are given in the y-axis. The array CGH profile in case 52 shows the gain of sequences from chromosome X and loss of those from chromosome Y, consistent with the sex mismatch test/reference. Copy number changes are present throughout the genome. The arrow indicates the breakpoint-associated deletions at 4q.

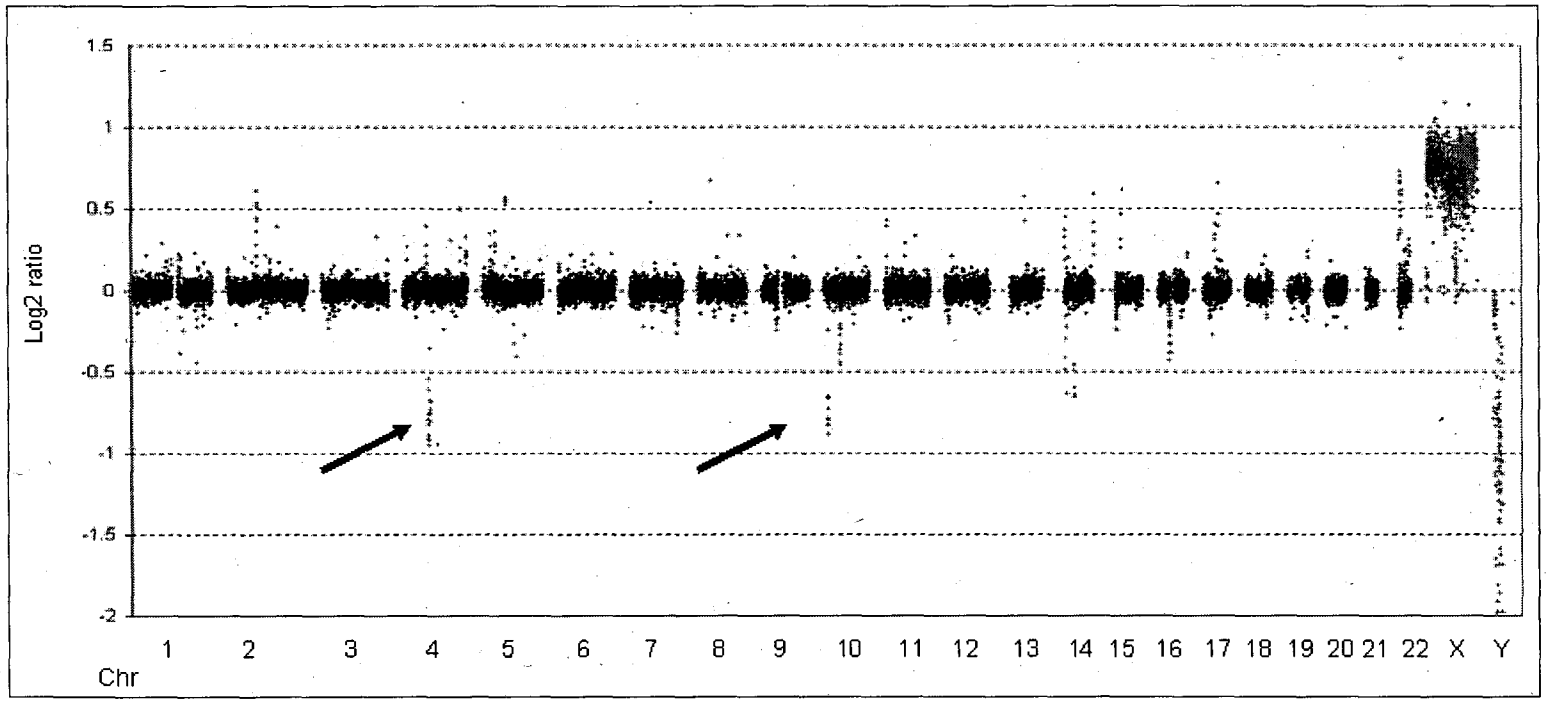

Fig.5.16. Array CGH results for case 53.

Genomic clones are shown in the x-axis and are grouped by chromosome 1 to Y. Log2 fluorescence ratios of test/reference DNA are given in the y-axis. The array CGH profile in case 53 shows the gain of sequences from chromosome X and loss of those from chromosome Y, consistent with the sex mismatch test/reference. Copy number changes are present throughout the genome. The arrows indicate the breakpoint-associated deletion at 4q and the breakpoint unrelated deletion on 10p.

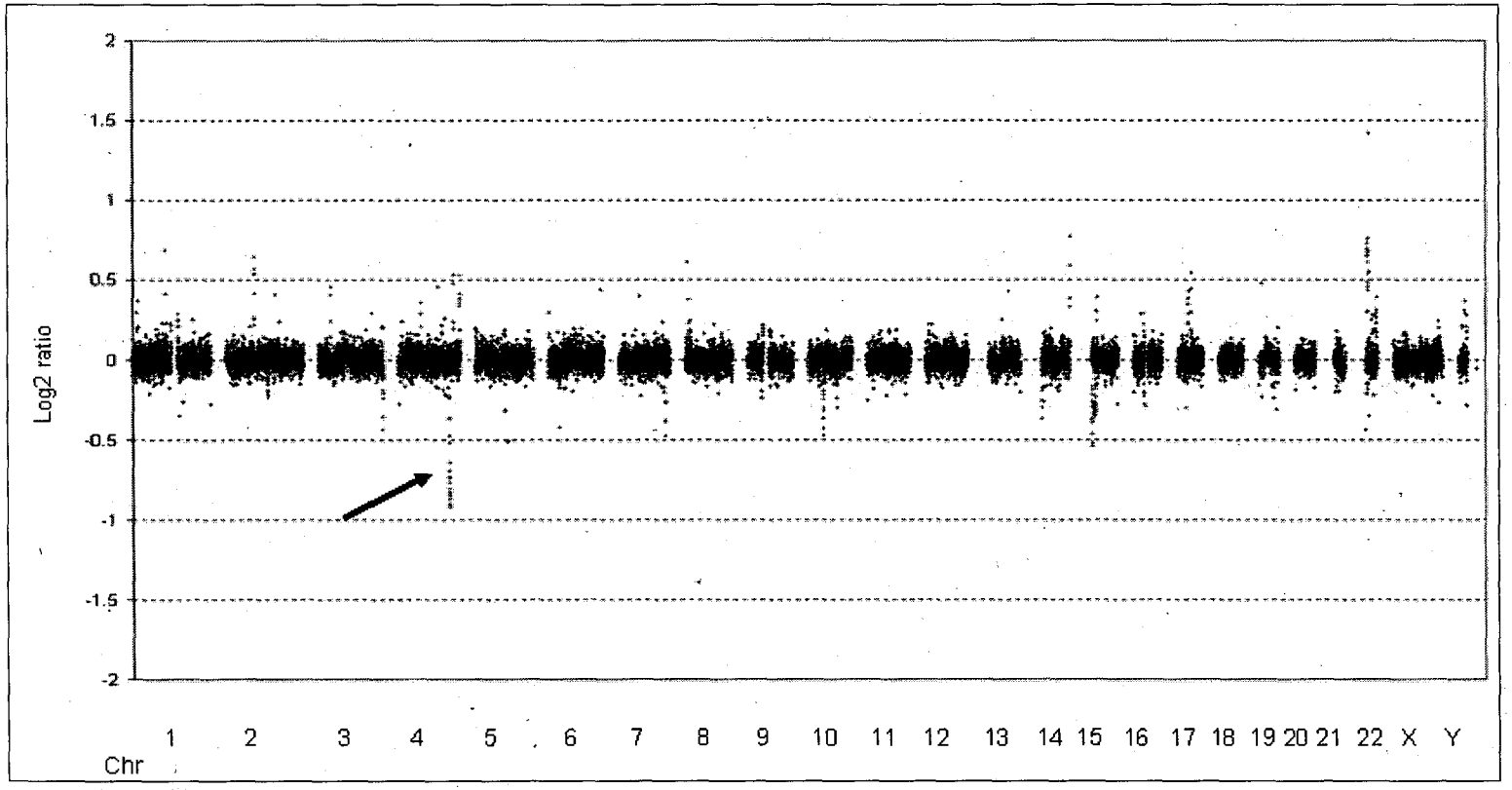

Fig.5.17. Array CGH results for case 57.

Genomic clones are shown in the x-axis and are grouped by chromosome 1 to Y. Log2 fluorescence ratios of test/reference DNA are given in the y-axis. The array CGH profile in case 57 shows the 1:1 ratio for clones mapped to chromosomes X and Y, consistent with the use of reference DNA of the same gender. Copy number changes are present throughout the genome. The arrow indicates the breakpoint unrelated deletion on 4q.

Table 5.XVI. Genes within the candidate disease-causing deletions in the patient cohort

Case	Chr	Gene	Disease association
20	5q12.1	***KIF2A***	
20	5q12.1	*DIMT1L*	
20	5q12.1	*IPO11*	
20	5q12.1	*ISCA1L*	
20	5q12.1	***HTR1A***	Tourette syndrome [MIM 137580]
20	5q12.1	*RNF180*	
52	4q13.3	*SLC4A4*	Proximal renal tubular acidosis with ocular abnormalities [MIM 604278]
52	4q13.3	*GC*	susceptibility to Graves disease [MIM 275000]
52	4q13.3	***NPFFR2***	
52	4q13.3	*ADAMTS3*	
52	4q13.3	*COX18*	
52	4q13.3	*ANKRD17*	
53	4q13.3	*PF4V1*	
53	4q13.3	***CXCL1***	
53	4q13.3	*PF4*	
53	4q13.3	*PPBP*	
53	4q13.3	***CXCL5***	
53	4q13.3	***CXCL3***	
53	4q13.3	***PPBPL2***	
53	4q13.3	***CXCL2***	
53	4q13.3	*MTHFD2L*	
53	4q13.3	***EPGN***	
53	4q13.3	***EREG***	
53	4q13.3	***AREG***	Psoriasis-like phenotype
53	4q13.3	*BTC*	
53	4q13.3	*PARM1*	
53	4q13.3	*RCHY1*	
53	4q13.3	*THAP6*	
53	4q13.3	***CDKL2***	
53	4q13.3	***G3BP2***	
53	4q13.3	*USO1*	
53	4q13.3	*PPEF2*	
53	4q13.3	*ASAHL*	
53	10p14	*LOC389936*	

Table 5.XVI. Genes within the candidate disease-causing deletions in the patient cohort (Cont.)

Case	Chr	Gene	Disease association
57	4q32.1	*MAP9*	
57	4q32.1	***GUCY1A3***	
57	4q32.1	***GUCY1B3***	
57	4q32.1	*ACCN5*	
57	4q32.1	***TDO2***	possible role in behavioural disorders
57	4q32.1	*CTSO*	
57	4q32.1	***PDGFC***	
57	4q32.1	***GLRB***	Hyperekplexia [MIM 149400]
57	4q32.1	***GRIA2***	

Genes in bold type have been implicated in biological processes of the nervous system, signal transduction or transcription and are therefore, good candidates for further investigations.

5.6 Overall summary of the results and genotype-phenotype correlations

The work presented in this chapter shows that molecular cytogenetic analyses of the breakpoints in ABCRs in 16 phenotypically abnormal patients enabled the re-assignment of the majority of the cytogenetic breakpoints, with 44% mapped to a genomic interval <100 kb, 31% between 100 and 200 kb and 19% mapped to >200 kb.

An analysis of the types of regions involved in breakpoints showed that many breakpoints mapped to cytogenetic bands with no remarkable features, with 7 (20%) of the breakpoints in pericentromeric regions, satellite stalks or telomeric bands and 5 (14.3%) in bands previously reported to contain fragile sites. Further analysis of the genomic context of the breakpoints showed that 34% mapped to regions containing segmental duplications and 53% to regions previously reported to contain normal CNV. In addition, 19% mapped to regions that are deleted or duplicated in patients with clinical abnormalities according to the DECIPHER database.

Table 5.XVII summarises the occurrence of the four features associated with phenotypic abnormalities investigated in this chapter. The results show that the four features were present in this unselected group of patients. Breakpoint-mediated gene disruption was the most frequent feature. This was obligatory in 7 patients and potential in a further 7 cases. However, even the breakpoint regions with no known genes mapped close enough to a gene (<1.2 Mb in all but one case) for position effects to have a role in the clinical abnormalities. Disease-candidate genes which are good candidates for further characterisation were identified among the breakpoint regions. Genomic imbalances at the breakpoints were detected in 3 patients. Genomic imbalances unrelated to the breakpoints were also seen in 3 patients. Finally, additional chromosomal complexity was detected in only one case, which was originally reported as a complex rearrangement with 3 breakpoints, but was shown to have a minimum of 6 breakpoints. With exception of cases 44 and 45,

all patients potentially presented with one of the four features. These results and genotype-phenotype correlations are discussed below.

Table 5.XVII. Overall summary of the results in the patient cohort

Case	Breakpoint-mediated gene disruption	Breakpoint-associated imbalances	Additional chromosomal complexity	Imbalances unrelated to breakpoints
16	+	-	-	-
20	+/-	+	-	-
43	+/-	-	-	-
44	-	-	-	-
45	-	-	-	-
48	+/-	-	-	-
49	+	-	-	-
50	+/-	-	-	-
51	+	-	-	-
52	+	+	+	+
53	+	+	-	+
54	+/-	-	-	-
55	+/-	-	-	-
56	+/-	-	-	-
57	+	-	-	+
58	+	-	-	-

\+ detected
\- not detected
+/- potential

Case 16: 46,XX,t(10;22)(q24.3;q13.31)*de novo*.ish t(10;22)(q24.31;q13.31)

Molecular studies of this translocation did not detect imbalances at the breakpoints or elsewhere nor chromosomal complexity. However, gene disruption remains as a possible mechanism to explain the abnormalities in this patient. *PAX2* (on 10q) was potentially disrupted (likelihood of 40%). *PAX2* encodes a probable transcription factor and it is implicated in development of the urogenital tract, the eyes, and the central nervous system. Defects in this gene cause renal-coloboma syndrome [MIM

120330] and isolated renal hypoplasia, neither of which is present in case 16. However, the patient does present with mild hearing impairment and a role for *PAX2* in ear development has been suggested with basis on its expression pattern (Sanyanusin *et al.*, 1995). *KIAA1644* (on 22q) was definitely disrupted. This is an uncharacterised gene and its possible functions remain elusive. Further studies will be needed to establish the contribution of *PAX2* and *KIAA1644* for the clinical abnormalities in case 16, which included intellectual impairment and psychiatric disease.

Case 20: 46,XX,t(2;5)(q33;q12)*de novo*.ish t(2;5)(q33.2;q12.1)

Molecular studies of this translocation did not detect chromosomal complexity or imbalances unrelated to the translocation. However, breakpoint-mediated gene disruption was a possibility. The *ABI2* (on 2q) was potentially disrupted (likelihood of 80%). This gene is thought to be involved in cell growth, transformation and in cytoskeletal reorganization. Interestingly, this breakpoint region mapped to the 2q32.2 deletion syndrome, which is characterised by severe mental retardation, facial dysmorphism, short stature and a cleft palate or high palate (van Buggenhout *et al.*, 2005). Case 20 had mild mental retardation but none of the other features and this mapping information could prove valuable to refine the critical region of the 8 Mb interval that defines the 2q32.2 syndrome. Furthermore, this patient had a deletion at the 5q breakpoint, and it is likely that haploinsufficiency of genes mapped to this region could contribute to the patient's intellectual impairment. Amongst the deleted genes, *KIF2A* (kinesin heavy chain member 2A) and *HTR1A* (5-hydroxytryptamine, serotonin, receptor 1A) have been implicated in processes of the nervous system and are good candidates for further investigations.

Case 43: 46,XY,t(4;17)(q35.1;q25.1)*de novo*.ish t(4;17)(q34.3;q24.2)

Molecular studies of this translocation did not detect imbalances at the breakpoints or elsewhere nor chromosomal complexity. However, gene disruption remains as a possible mechanism to explain the abnormalities in this patient. The genes *PSMD12* (on 17q) and *PITPNC1* (on 17q) were potentially disrupted (combined likelihood of

60%). *PSMD12* encodes a regulatory subunit of the proteasome 26S, a complex responsible for degradation of ubiquinated proteins (reviewed by Coux *et al.*, 1996). This gene has not been extensively characterised in man, but in yeast, the disruption of its possible counterpart, *NAS5*, resulted in lethality (Saito *et al.*, 1997). *PITPNC1* encodes a member of the phosphatidylinositol transfer protein family. These proteins are ubiquitously expressed and have a role in the monomeric transport of lipids from one membrane compartment to another (reviewed by Allen-Baume *et al.*, 2002). Furthermore, the *Drosophila* homologue to *PITPNC1* has been implicated in retinal and olfactory neurosensory signalling and null mutants have light-induced retinal degeneration (Woodard *et al.*, 1992). Neither *PSMD12* nor *PITPNC1* appear to be obvious candidates for isolated truncus arteriosus, and further studies are needed to elucidate their putative role in this malformation.

Case 44: 46,XY,inv(6)(p24q16.2)*de novo*.ish inv(6)(p24.3q16.2)

Molecular studies of this inversion did not detect breakpoint-mediated gene disruption, imbalances at the breakpoints or elsewhere nor chromosomal complexity. In an attempt to further evaluate whether this inversion was coincidental or causative to the phenotypic abnormalities, a search was made for genes mapped in the vicinity of the breakpoint regions. This analysis aimed to identify candidate genes that could cause the phenotype by a position effect mechanism (see section 1.6.2). The 6p24.3 breakpoint was found to map <30 kb of the 3' end of the *TFAP2A* (transcription factor AP-2 alpha, activating enhancer binding protein 2 alpha) gene. *TFAP2A* codes for a transcription factor, its expression is developmentally regulated and it is implicated in craniofacial development. The 6q16.2 breakpoint mapped <120 kb 5' of the *POU3F2* (POU class 3 homeobox 2) gene. *POU3F2* is a transcription factor of the Class III POU genes, a gene class that is expressed predominantly in the central nervous system. Both the *TFAP2A* and the *POU3F2* genes map within the maximum range (~1.3 Mb) described for position effects (Velagaleti *et al.*, 2005) and both appear to be very good candidates for further studies of their putative contribution to the phenotype in case 44, which is mainly of developmental delay, dysmorphic features and a speech disorder.

Case 45: 46,X,t(X;19)(q21;p13.11)*de novo*.ish t(X;19)(q21.1;q11.?)

Molecular studies of this translocation did not detect breakpoint-mediated gene disruption, imbalances at the breakpoints or elsewhere or chromosomal complexity. However, it seems likely that the phenotype, premature ovarian failure (POF), is due to the abnormal karyotype. This is because approximately half of carriers of X;autosome translocations have been reported to present gonadal dysfunction (Schmidt and Du Sart, 1992). Furthermore, breakpoints in affected females seem to cluster in two regions critical for normal ovarian function: Xq13-Xq21 and Xq23-Xq27 (Madan, 1983; Therman *et al.*, 1990, Sala *et al.*, 1997, Bione *et al.*, 1998). It is of interest that the breakpoint in case 45 maps to one of these critical regions: Xq21.1. However, the mechanism to explain the association of POF with breakpoints in this critical region remains largely unclear and speculative. Breakpoint-mediated gene disruption has been suggested as a possible mechanism (Sala *et al.*, 1997). Schlessinger *et al.* (2002) suggested that X breakpoints could lead to structural alterations that ultimately would cause aberrations in pairing or X inactivation during folliculogenesis. These aberrations would ultimately lead to apoptosis at meiotic checkpoints. More recently, another hypothetical mechanism has been proposed by Rizzolio *et al.* (2007). The authors characterised the breakpoints in 4 X;autosome balanced translocations and found that while the chromosome X breakpoints were mapped to regions with no ovary-expressed candidate genes, this class of genes was present near the autosomal breakpoints in all 4 cases. The authors proposed that POF might be caused by an ovary-specific position effect on autosomal genes translocated to the derivative X. In case 45, both breakpoints mapped to regions containing no known genes, suggesting that the models proposed by Sala *et al.* (1997) and by Rizzolio *et al.* (2007) are unlikely to explain POF in this patient, but the model proposed by Schlessinger *et al.* (2002) is possible.

Case 48: 46,XY,t(4;6)(q33;q22.2)*de novo*.ish t(4;6)(q33-34.1;q21)

Molecular studies of this translocation did not detect imbalances at the breakpoints or elsewhere or chromosomal complexity. However, gene disruption remained as a

possible mechanism to explain the abnormalities in this patient, since the *AMD1* gene (on 6q) was potentially disrupted (likelihood of 20%). *AMD1* codes for an intermediate enzyme involved in polyamine biosynthesis. Polyamines are critical for cellular proliferation and might be involved in brain development and cognitive function (Cason *et al.*, 2003). Thus, further studies would be of interest to evaluate the putative role of *AMD1* in the developmental delay present in this patient.

Case 49: 46,XX,t(2;10)(q33;q21.2)*de novo*.ish t(2;10)(q33.1;q21.3)

Molecular studies of this translocation did not detect imbalances at the breakpoints or elsewhere or chromosomal complexity. However, gene disruption remained as a possible mechanism to explain the abnormalities in case 49. *SATB2* (on 2q) was definitely disrupted and haploinsufficiency of this gene has been reported to cause cleft palate in two patients with cleft palate, facial dysmorphism, mild learning disability and balanced translocations involving 2q (FitzPatrick *et al.*, 2003). Case 49 presented with cleft palate, dysmorphic features and developmental delay and thus haploinsufficiency of *SATB2* is most likely causative to this phenotype.

Case 50: 46,X,t(X;8)(q22.1;q24.13).ish t(X;8)(q23;q24.3)

Molecular studies of this translocation did not detect imbalances at the breakpoints or elsewhere nor chromosomal complexity. Breakpoint-mediated gene disruption was very unlikely in this case. As discussed for case 45, the premature ovarian failure present in this patient is likely to be due to the X;autosome translocation. Furthermore, the Xq breakpoint mapped to Xq23, which is within the critical region for normal ovarian function (Madan, 1983; Therman *et al.*, 1990).

Interestingly, the *AGTR2* (angiotensin II receptor, type 2) gene mapped in the vicinity of the Xq breakpoint. The likelihood of breakpoint-mediated disruption of this gene was only 3%. Moreover, the FISH results indicated that it was highly unlikely that *AGTR2* was interrupted. *AGTR2* might have a role in cell morphogenesis, growth and development, but its function remains elusive. Vervoort *et al.* (2002b) implicated *AGTR2* in X-linked mental retardation (XLMR) following

breakpoint mapping studies in a patient with a t(X;7)(q24;q22) and moderately severe mental retardation. The authors showed that although the Xq breakpoint was mapped >150 kb upstream of *AGTR2*, the expression of this gene was abolished. It was suggested that lack of *AGTR2* was due to a position effect of the breakpoint. Furthermore, skewed X inactivation was shown, with the normal X inactivated in all the cells analysed. Further evidence of a role of *AGTR2* in XLMR was given by the identification of point mutations in 8/590 unrelated affected males. In case 50, the breakpoint was mapped 50-100 kb upstream of *AGTR2*, but there was no evidence of mental retardation in this patient. X inactivation studies were not undertaken in this case, but the most likely X inactivation pattern would be that seen in the majority of X; autosome carriers with POF and no other abnormalities, which is inactivation of the normal X (Mattei *et al.*, 1982). Future studies would be of interest to further elucidate the expression pattern of *AGTR2* in this patient and its reported role in cognitive development.

Case 51: 47,XX,t(4;20)(p15.2;p11.23)*de novo*,+mar[23]mat /46,XX,t(4;20)(p15.2;p11.23)[7].ish t(4;20)(p15.32;p12.2)

Molecular studies of this translocation did not detect imbalances at the breakpoints or elsewhere nor chromosomal complexity. However, gene disruption remained as a possible mechanism to explain the abnormalities in case 51. *SNAP25* (on 20p) was definitely disrupted. This gene has been reported as a susceptibility *locus* for attention deficit hyperactivity disorder (ADHD) after polymorphisms were detected in affected children (Feng *et al.*, 2005). *SNAP25* is thought to have a role in neurotransmitter release and synaptic function. These characteristics suggest that *SNAP25* is a good candidate for further studies of its role in the phenotype of case 51, which includes ADHD, developmental delay and autistic spectrum disorder.

Case 52: 46,XX inv ins (11;4)(q22.2;q13.2q21.3)*de novo*.ish inv ins (11;4)(q22.1;q13.1q21.23)

Molecular studies of this translocation detected breakpoint-mediated gene disruption, imbalances at the breakpoints, imbalances elsewhere in the genome and

additional chromosomal complexity. *MAPK10* (on 4q) was definitely disrupted; this gene encodes the mitogen-activated protein kinase 10 of the MAPK family of serine/threonine kinases. These proteins have been extensively characterised and are known to have an essential role in signal transduction by modulating gene transcription in the nucleus in response to cellular environment changes. The MAPK signalling pathway is involved in cellular processes such as proliferation, differentiation, migration and apoptosis, and also in diseases such as chronic inflammation and cancer (reviewed by Turjanski *et al.*, 2007). *MAPK10* has also been implicated in pharmacoresistant epileptic encephalopathy of the Lennox-Gastaut type (Shoichet *et al.* 2006) apparently by disrupting normal signal transduction in neuronal cells. In addition, this patient had potential disruption (likelihood of 40%) of *CNTN5* (on 11q). This gene encodes an adhesion molecule of the immunoglobulin superfamily. This group of proteins has a role in axon guidance during development and in plasticity and maintenance of synaptic connections in the adult brain (Kamei *et al.*, 2000). The involvement of both *MAPK10* and *CNTN5* in synaptic connections suggests that both genes could be implicated in the developmental delay and learning difficulties seen in case 52. Moreover, this patient had a number of deleted genes, which might well contribute to the abnormal phenotype. Amongst these, the *NPFFR2* (neuropeptide FF receptor 2) has been implicated in signal transduction and in neuropeptide receptor activity and thus it is of particular interest.

Case 53: 46,XX,t(4;8)(q21.1;p12)*de novo*.ish t(4;8)(q13.3;p12)

Molecular studies of this translocation detected breakpoint-mediated gene disruption, imbalances at the breakpoints and elsewhere, but no chromosomal complexity. *UNC5D* (on 8q) was disrupted. *UNC5D* has not been extensively characterised in humans and its function is largely unknown, but it is thought to encode a netrin receptor. Netrins are involved in neuronal migration and in axon guidance during the development of the nervous system (reviewed by Cirulli and Yebra, 2007). Further studies of *UNC5D* would be of interest to establish its putative contribution to the intellectual impairment in this patient. Moreover, this

patient had a number of deleted genes (listed in Table 5.XVI) which could be contributing to the clinical abnormalities.

Case 54: 46,XY,t(14;15)(q23;q26.3)*de novo*.ish t(14;15)(q22.3;q26.2)

Molecular studies of this translocation did not detect imbalances at the breakpoints or elsewhere nor chromosomal complexity. However, gene disruption remained as a possible mechanism to explain the abnormalities in case 54. *NR2F2* (on 15q) was potentially disrupted (likelihood of 40%). Studies of *NR2F2* in mice showed that null mutants had angiogenesis, heart defects and died early in embryogenesis (Petit *et al.*, 2007). These studies suggest that *NR2F2* might be a good candidate for cardiac abnormalities and further studies would be of interest to evaluate its putative contribution to the coarctation of the aorta present in this patient.

Case 55: 46,XY,t(19;20)(q13.43;q11.1)*de novo*.ish t(19;20)(q13.43;q11.1)

Molecular studies of this translocation did not detect imbalances at the breakpoints or elsewhere nor chromosomal complexity. However, gene disruption remained as a possible mechanism to explain the oligospermia in this patient. The 19q breakpoint mapped within a large cluster of zinc finger proteins of the Kruppel-associated box domain family, and *ZNF606* was potentially disrupted (likelihood of 60%). The Kruppel-associated box domain family represents the largest single family of transcriptional regulators in mammals (reviewed by Urrutia, 2003) but, it is unknown whether they might have a role in oligospermia. Additional studies in this patient excluded the presence of common causes of male infertility such as microdeletions of the *AZF* locus (Vogt *et al.*, 1996) and the presence of the 5T allele in the *CFTR* gene (Costes *et al.*, 1995). These results together with the observation that there is a higher frequency of chromosome rearrangements reported in infertile males by comparison to the general population: 4.6% *versus* 0.5% (van Assche *et al.*, 1996) suggest that the translocation could be causative to the oligospermia in this patient. The underlying mechanism for oligospermia due to chromosome rearrangements is largely unknown, but it has been proposed that incompletely paired chromosomes might have an effect (Burgoyne and Baker, 1984). Animal

studies support this model by showing that meiosis progression is regulated by checkpoints and synaptic errors trigger apoptosis among spermatocytes (Odorisio *et al.*, 1998).

Case 56: 46,XY,t(6;21)(q16.2;q11.2)*de novo*.ish t(6;21)(q16.1;q11.1)

Molecular studies of this translocation did not detect imbalances at the breakpoints or elsewhere nor chromosomal complexity. *C6orf167* (on 6q) was potentially disrupted (likelihood of 15%), but the function of this gene remains elusive. Further analysis of the causes of the infertility in this patient included testing for the presence of the 5T allele (Costes *et al.*, 1995), which gave normal results. Microdeletions of the Y chromosome were not excluded in this case. As discussed above for case 55, the presence of a reciprocal translocation might contribute to the oligospermia in this patient (van Assche *et al.*, 1996).

Case 57: 46,XY,t(2;5)(p23;q11.2)*de novo*,t(18;22)(q11.2;p13)*de novo*.ish t(2;5)(p23.3;q11.2),t(18;22)(q11.2;p13)

Molecular studies of these translocations detected no imbalances at the breakpoints or chromosomal complexity. Breakpoint-mediated gene disruption was detected. *ZNF521* (on 18q) was disrupted, but its function is largely unknown. *NCOA1* (on 2p) was potentially disrupted (likelihood of 50%). This gene codes for a coactivator molecule that binds nuclear receptors and stimulates transcription in a hormone-dependent mode. *NCOA1* might also be implicated in remodelling of chromatin and recruitment of general transcription factors. Further studies of *ZNF521* and *NCOA1* would be necessary to establish their contribution to the developmental delay and epilepsy in case 57. Moreover, this patient had a deletion independent of the breakpoints and haploinsufficiency for genes within the deletion might well contribute to the clinical abnormalities (epilepsy, short stature and microcephaly).

Case 58: 46,XX,t(6;18)(p23;q22)*de novo*.ish t(6;18)(p24.3;q21.1)

Molecular studies of this translocation did not detect imbalances at the breakpoints or elsewhere nor chromosomal complexity. However, gene disruption remained as a

possible mechanism to explain the primary amenorrhea in case 58. *DCC* (on 18q) was disrupted; this gene has been implicated in colorectal carcinomas and is often deleted in these tumours (Vogelstein *et al.*, 1988). A role for *DCC* in the development of the nervous system has also been suggested (Colon-Ramos *et al.*, 2007). This is because *DCC* presents binding properties to netrins which in turn are molecules involved in axon growth and guidance in pathfinding during development of the nervous system (reviewed by Cirulli and Yebra, 2007). However, there is no evidence that *DCC* might cause primary amenorrhea and available data suggests that this gene is not expressed in ovary. A review of the literature showed that there are very few reports of autosomal translocations in women with primary amenorrhea and for these it can be assumed that either the translocation is coincidental to the phenotype or the breakpoints affect autosomal genes implicated in normal gonadal function. Further studies of *DCC* are necessary to evaluate the effect of its disruption in case 58.

5.7 Conclusions

Apparently balanced rearrangements in patients with clinical abnormalities can be associated with breakpoint-mediated gene disruption, breakpoint-associated imbalances, additional chromosomal complexity and genomic imbalances unrelated to the breakpoints. The results presented here suggest that genomic imbalances both at the breakpoints and independent of the breakpoints are the most important cause of clinical abnormalities in this type of patient. Furthermore, the molecular characterisation of the breakpoint regions was invaluable for the identification of disease causative genes and of candidate genes for further studies.

6 COMPARISON OF THE CONTROL AND THE PATIENT COHORTS

6.1 Overview

This chapter presents a comparison of the results between the control cohort (chapter 4) and the patient cohort (chapter 5), focusing on the four features analysed in the present study. The first three features were related to the breakpoints and concerned the presence of gene disruption, imbalances or complexity at these regions, whereas the last feature was the presence of genomic imbalances unrelated to the breakpoints. With respect to the features related to the breakpoints, it is possible to carry out direct comparisons as both cohorts were analysed at the same level of resolution, i.e. typically to the BAC/PAC or fosmid level. However, for the survey of genomic imbalances unrelated to the breakpoints a direct comparison of the results is not applicable due to the use of different platforms for array CGH. The use of a lower resolution analysis (1 Mb array) in the control cohort has the possible limitation that smaller genomic imbalances might have gone undetected.

The results reported here will be compared to those of other similar studies of apparently balanced rearrangements. There are several reports in the literature of this type of study, but the great majority of these are single case reports of patients with remarkable clinical features and only three studies report on series of patients comparable to those of the present study. FitzPatrick *et al.* (2006) reported, in abstract, the analyses of the breakpoints in apparently balanced translocations in 46 phenotypically abnormal patients and in 30 phenotypically normal individuals. This is the largest study of this type that included both patient and control cohorts. However, the analyses were focused on the characteristics of the breakpoint regions and whole genome analyses by array CGH were not undertaken. In another study, Gribble *et al.* (2005) characterised the apparently balanced reciprocal translocations of 10 clinically abnormal patients by 1 Mb array CGH, and carried out breakpoint mapping studies by array painting and FISH. This study provided information both on the characteristics of the breakpoint regions and on the genomes as a whole, but did not include a control cohort. Lastly, De Gregori *et al.* (2007) characterised apparently balanced rearrangements and complex rearrangements by array CGH.

The patients with apparently balanced reciprocal translocations were sub-divided in two groups. One was represented by 27 clinically abnormal patients and the other included 14 prenatal samples. There were no follow-up studies on 12 of these prenatal samples (pregnancies were in progress), one other patient had normal development at 3 months of age and the remaining patient had delayed psychomotor development at 6 months of age. Thus, these two groups cannot be defined as control and patient cohorts as reported in the present study. The authors used array CGH at 100 kb resolution, but did not carry out breakpoint mapping studies in any of the carriers of apparently balanced reciprocal translocations. Data from these three studies will be discussed in the context of the findings reported here and, where appropriate, an estimate of the frequency of the four features analysed will be generated by combining data from these studies with the data from the present study.

6.2 The breakpoint regions

6.2.1 Gene content at the breakpoint regions

In the present study, the proportion of breakpoints mapped to regions containing no known genes was 29.3% in the control cohort *versus* 39.3% in the patient cohort. Breakpoints definitely associated with gene disruption accounted for 31% in the control cohort *versus* 25% in the patient cohort. Finally, breakpoints potentially associated with gene disruption accounted for 39.7% in the control cohort *versus* 35.7% in the patient cohort. These results suggest that there is no major difference in the pattern of distribution of breakpoint regions between the two cohorts. Similar observations have been reported by FitzPatrick and colleagues (2006).

The proportion of the human genome with annotated genes has been estimated as 38.5% (Sakharkar *et al.*, 2004) or 34.8% (Ensembl database version 46.36h). Based on these estimates, the observation that 31% (control cohort) and 25% (patient cohort) of the breakpoints map within genes suggests a random distribution of

breakpoints. Nonetheless, 70.7% (control cohort) and 60.7% (patient cohort) of the breakpoints mapped in the vicinity (<200 kb) of a gene, and for the breakpoints refined to regions that contained no known genes, all but one mapped within 1.2 Mb of a gene. This clearly shows that most breakpoints in the present study map near genes, although not necessarily within them. Furthermore, since the majority of the breakpoints map in the vicinity of a gene position effects may be exerted by the breakpoints.

6.2.2 The types of genes mapped to the breakpoint regions

Many breakpoint regions, in both cohorts, contained genes with gene ontology (GO) descriptions that included important biological processes such as transcription and signal transduction. However, there were differences between the two cohorts and the most remarkable was related to genes implicated in biological processes of the nervous system. Genes with this category of GO terms were present in the breakpoint regions of 2/31 (6.5%) individuals in the control cohort *versus* 6/16 (37.5%) in the patient cohort. In the control cohort, the genes involved were *ABAT* (disrupted in case 2.1) and *NPTN* (potentially disrupted in case 2.6). The exact GO terms for these genes were: neurotransmitter catabolic process and positive regulation of long-term neuronal synaptic plasticity, respectively. The *ABAT* gene is associated with GABA-AT deficiency, and the autosomal recessive nature of this disorder is likely to explain the normal phenotype in case 2.1 (discussed in section 4.7). *NPTN* remains largely uncharacterised, but has been proposed as a candidate gene for susceptibility to schizophrenia (Saito *et al.*, 2007). In the patient cohort, the genes involved in processes linked to the nervous system were *PAX2* (potentially disrupted in case 16), *KIF2A* (deleted in case 20), *SNAP25* (disrupted in case 51), *CXCL1* (deleted in case 53), *NR2F2* (potentially disrupted in case 54) and *DCC* (disrupted in case 58). In all cases, with exception of case 58, the patients had some level of cognitive impairment. The exact gene ontologies represented by these genes were: nervous system development, neuron migration, neurotransmitter

secretion/uptake, axon guidance and axogenesis. The observation that genes linked to processes of the nervous system are overrepresented in the patient cohort by comparison to the control cohort indicates, rather unsurprisingly, that breakpoints within genes with a role in nervous system are likely to be an important cause of cognitive impairment in patients with ABCRs, thus supporting the utility of molecular studies of the type described here to identify candidate genes.

6.2.3 The genomic context of the breakpoint regions

Cytogenetic bands

An analysis of the types of cytogenetic bands involved in breakpoints showed that centromeric/pericentromeric regions, satellite stalks and telomeric bands accounted for 8/61 (13.1%) of the breakpoints in the control cohort versus 7/35 (20%) in the patient cohort. Furthermore, 12 (19.7%) of the breakpoints in the control cohort versus 5/35 (14.3%) in the patient cohort mapped to bands reported to contain fragile sites. These results suggest that the distribution of the breakpoints within bands of specialised genome architecture is similar in both cohorts and a large sample size would be needed to ascertain whether there is a biased distribution of the breakpoints.

Segmental duplications (SDs)

Segmental duplications (see section 4.5.2.2) mapped to 25% of the breakpoint regions in the control cohort *versus* 34% in the patient cohort. This suggests that SDs are relatively common at the breakpoint regions, but not necessarily important at mediating this class of rearrangements as in the majority of the cases SDs were present at only one of the rearrangement breakpoints.

For the control cohort, only two cases had both breakpoints mapped to regions with SDs: cases 3.2 and 3.5. For case 3.2, the breakpoints mapped to 8p23.1 and 12p13.31; the 8p23.1 breakpoint was localised within the defensin gene cluster, a

well characterised genomic region which is genetically unstable and is often involved in chromosome rearrangements (Floridia *et al.*, 1996; Giglio *et al.*, 2001; Hollox *et al.*, 2003). The 12p13.31 breakpoint mapped to a region of high sequence homology to 8p23.1 as described by Taudien *et al.* (2004). For case 3.5, the breakpoints mapped to Xp11.1 and 22q11.21, both of which are pericentromeric bands, a type of region notable for enrichment in SDs and for genetic instability (Eichler, 1998). The search for SDs at the breakpoints in the patient cohort also identified two cases which had both breakpoints mapped to regions neighbouring SDs: cases 20 and 50. The breakpoint regions in case 20 mapped to 2q33.2 and 5q12.1 and those in case 50 to Xq23 and 8q24.3. Taken together, these observations suggest that SDs might have predisposed to the breakpoints in only a minority, 4/47 (~8.5%), of the cases analysed. Comparable observations have been reported in a study of 4 apparently balanced translocations at the sequence level, which had no SDs or any degree of shared sequence homology at the breakpoints (Gribble *et al.*, 2007) suggesting that SDs are not common mediators of breakpoint regions in non recurrent rearrangements. It has been previously suggested that SDs are important mediators of recurrent rearrangements, but are not so crucial in the formation of non recurrent rearrangements, which are likely to be formed by non homologous end joining (NHEJ), a mechanism that does not require homology between the regions involved (Lupski and Stankiewicz, 2005). Recently, Lee *et al.* (2007) proposed a novel mechanism for the generation of non recurrent rearrangements, Fork Stalling and Template Switching (FoSTeS), which is replication-based and like NHEJ does not require homology between the breakpoint regions.

Normal copy number variation (CNV)

Regions of known normal CNV mapped to 50% of the immediate breakpoint regions in the control cohort versus 53% in the patient cohort. These results show that there is a similarly high distribution of the breakpoints in both cohorts to within regions of known CNV. This observation suggests that the genomic locations prone to CNV might also be prone to rearrangement breakpoints of the type analysed here, perhaps due to the underlying unstable genomic architecture of these regions (Iafrate *et al.*,

2004). The analysis of further breakpoints would be of interest to determine whether there is a bias for breakpoints in regions of normal CNV.

Potentially disease-causative copy number changes

Regions of deletion or duplication thought to cause clinical abnormalities in affected patients and recorded in the DECIPHER database were found to co-localise to 17 breakpoints in the control cohort and 6 in the patient cohort. These included regions of unknown significance and also regions of known syndromes: the 2q32.2 deletion syndrome region, for which breakpoints were identified in one normal individual and in two phenotypically abnormal patients; and the 22q11 deletion/duplication syndrome region, which co-localised to one of the breakpoints in a phenotypically normal individual. These results suggest that the detailed characterisation of breakpoint regions in carriers of ABCRs, particularly in normal individuals might prove valuable in gathering information that can be used to establish the regions of clinical relevance in patients with clinical abnormalities.

6.3 Cryptic imbalances at the breakpoints

Breakpoint-associated cryptic imbalances were not detected in the control cohort. In contrast, 3/16 (18.75%) cases in the patient cohort had *de novo* candidate disease-causing deletions. Of these, two deletions were over 2 Mb and mapped to regions containing known genes. The remaining deletion was 170 kb in size and mapped to a region with no known genes. Similarly, FitzPatrick *et al.* (2006) reported no breakpoint-associated imbalances in their control cohort, whereas 6/46 (13%) of the patient cohort had this type of imbalance, of which 3 mapped directly at the breakpoints and the others between 6 to 30 Mb from the breakpoints. In the study by Gribble *et al.* (2005), *de novo* breakpoint-associated imbalances were detected in 2/10 patients. These deletions were estimated to have genomic sizes of 5.5 and 6.2 Mb. In addition, De Gregori *et al.* (2007) reported breakpoint-associated deletions in 7/27 (26%) of the phenotypically abnormal patients and in none of the 14 prenatal

samples. The deletions identified spanned regions between 1.1 and 7.8 Mb. Taken together, these results show that breakpoint-associated deletions are present in 13 to 26% of phenotypically abnormal carriers of ABCRs, but have not been detected in any of the phenotypically normal carriers of ABCRs.

It is of interest that most of the genomic imbalances reported are relatively large (>1 Mb) and well within the limits of resolution of the methodologies used to ascertain these. This is well illustrated in the study of De Gregori *et al.*(2007), who used the highest level of resolution amongst the aforementioned studies, but even with a resolution of 100 kb, the imbalances detected spanned regions between 1.1 and 7.8 Mb. The likely explanation for this observation is the proportional correlation between size of imbalance and number of genes affected. Furthermore, the probability of affecting a gene required for normal development also increases with the size of the imbalance. As a result, larger imbalances are more likely to be disease-causative suggesting that this explains their presence in a proportion of phenotypically affected patients, but not in any of the 61 phenotypically normal individuals described by FitzPatrick *et al.* (ASHG, 2006) and the present study. In an attempt to explore further whether most imbalances reported in patients with clinically abnormal phenotypes are large, the literature was reviewed. Since there were not many studies of series of patients with ABSCRs, data was also reviewed from array CGH studies of patients with clinical abnormalities, but normal karyotypes (Vissers *et al.*, 2003; Shaw-Smith *et al.*, 2004; de Vries *et al.*, 2005; Gribble *et al.*, 2005; Schoumans *et al.*, 2005; Tyson *et al.*, 2005; Friedman *et al.*, 2006; Menten *et al.*, 2006; Rosenberg *et al.*, 2006; De Gregori *et al.*, 2007). Different methodologies were used in these studies, thus limiting the power of data comparison, nevertheless, as a general observation it was noted that ~75% of all the imbalances reported were larger than 1 Mb and 60% were larger than 2 Mb. As discussed above, this is not due only to limits of the methods used as shown further by the study of Friedman *et al.* (2006) who applied the Affimetrix GeneChip 100K platform (resolution of 30 kb) to the study of 100 children with idiopathic mental retardation and both their unaffected parents. The candidate disease-causing

imbalances detected included 8 regions of copy number loss and 2 of copy number gain. These ranged in size between 178 kb and 11.1 Mb, which is well above the 30 kb resolution of the array, showing that large genomic imbalances are an important occurrence amongst phenotypically abnormal patients. Large imbalances can also be found in individuals with normal phenotypes. Studies of cytogenetically visible imbalances, including directly transmitted unbalanced chromosome abnormalities and euchromatic variants have shown that many large imbalances (average size of 10 Mb) are compatible with normal phenotypes (reviewed by Barber, 2005). Hence, in these cases it is reasonable to assume that no dosage sensitive genes with a role in normal development are affected.

It is of interest that although no breakpoint-associated genomic imbalances have been detected in the control cohort, when analysed at the sequence level, translocations of phenotypically normal individuals have been reported to have cryptic imbalances involving a small number of nucleotides. Gajecka *et al.* (2006) reported imbalances of 4-5 bps in 2/3 clinically normal parents of children with unbalanced karyotypes due to malsegregation of the apparently balanced parental translocations. In another report, which included case 2.10 of the present study, cryptic imbalances at the nucleotide level were also observed (Gribble *et al.*, 2007). Case 2.10 had a t(11;17), which was balanced by FISH studies, but by sequence analysis had deletion of a single nucleotide and insertion of three nucleotides at the 11p breakpoint whereas the 17q breakpoint was balanced. Due to a lack of literature on sequence analysis of rearrangements in phenotypically normal carriers of ABCRs, it is not yet possible to evaluate whether imbalances at the nucleotide level are a common finding. However, amongst phenotypically abnormal carriers this seems to be the case (review within McMullan *et al.*, 2002; Gribble *et al.*, 2007). McMullan *et al.* (2002) proposed that these deletions or insertions of nucleotides might be a result of complex mechanisms for translocation formation rather than simply chromosome breakage and re-joining. In this context, it could be proposed that cryptic imbalances at the nucleotide level are a usual by-product of complex mechanisms of translocation formation, whereas the cryptic imbalances detectable at

the submicroscopic level by array CGH are a more unusual by-product, which due to their larger size are more likely to involve regions with a role in phenotypic abnormalities. This could explain the presence of nucleotide imbalances but not submicroscopic imbalances in the control cohort.

6.4 Additional chromosomal complexity

For the control cohort, additional chromosomal complexity was not detected in 30/31 translocations. The remaining case, case 3.8, had a breakpoint within band 22q12.3, in a region of complex genomic architecture due to segmental duplications, which could not be fully characterised by FISH studies (discussed in chapter 4). In the patient cohort, additional chromosomal complexity was present in a single case, case 52, which was initially reported cytogenetically as a complex rearrangement with 3 breakpoints and was shown by molecular cytogenetic studies to have additional unsuspected complexity (discussed in chapter 5). Taken together, these observations suggest that additional chromosomal complexity is not significant amongst the chromosome rearrangements analysed in the present study. In other studies, unsuspected complexity was more significant and was detected in 3/10 (Gribble *et al.*, 2005) and in 5/27 (De Gregori *et al.*, 2007) of the patients. By combining the data of the present study with those reported by Gribble *et al.* (2005) and De Gregori *et al.* (2007), the prevalence of unsuspected chromosomal complexity in phenotypically abnormal carriers of ABCRs is estimated to be 9/53 (~17%).

6.5 Genomic imbalances unrelated to the breakpoints

Genomic imbalances unrelated to the breakpoints were detected in both the control and the patient cohort. However, the direct comparison of these two data sets is not appropriate mainly because different array CGH platforms (1 Mb BAC array,

WGTP BAC array and 244K oligo array, see chapter 3), with different levels of resolution and different types of probes were utilised. Thus, it is expected that smaller imbalances might have gone undetected when using the lower resolution platform. In addition, regions of normal copy number variation are likely to be underestimated when using oligonucleotide arrays as these are typically designed away from repeat regions, whereas regions of CNV are biased towards repeats (Sharp *et al.*, 2005). These same observations have been reported by Aradhya *et al.* (2007) after analysing a cohort of 20 clinically affected patients in both a 1Mb BAC array and a 44K oligonucleotide array. Another reason that precludes the direct comparison of the study cohorts was the use of different sets of reference DNA, with some experiments using a single individual and others using a pool of multiple donors. The choice of reference DNA is likely to influence the number of regions of CNV detected, with the use of multiple donors being likely to reduce the number of regions of CNV detected. As a consequence, this study cannot answer questions that require a direct comparison of the cohorts such as those dealing with the relative distribution, size and number of CNVs. Nonetheless, the two sets of results must be treated differently because any CNV found in the control cohort, by definition, must be associated with normal development (although might be associated with late onset disorders), while those found in the patient cohort might either represent normal CNV or a pathogenic change with a role in clinical abnormalities.

In the control cohort, the great majority of the regions of CNV identified had been previously reported in studies of normal populations. However, 21 novel regions of CNV, each consisting in a single BAC clone, were detected; 7 were present in more than one individual, whereas the remaining 14 were present in only a single person (see chapter 4, Table 4.XII). Many (76.2%) of these novel regions mapped to known genes, which can contribute to human variation and disease as previously reported in surveys of CNV (Iafrate *et al.*, 2004; Sebat *et al.*, 2004). Redon *et al.* (2006) have discussed the finding that although a high number of genes are found to map to regions of CNV, overall, CNVs are actually biased away from genes.

For the patient cohort, sites of CNV included both regions previously reported from studies of normal populations and novel regions. The latter were assumed to be candidate disease-causing changes if not present in the patient's parents and were identified in 3/16 (18.75%) of the patients. All regions of imbalance detected were deletions with sizes between 1.1 and 2 Mb, and it is therefore likely that haploinsufficiency of genes within these regions have a role in clinical abnormalities (discussed in section 5.5.1). Previously, Gribble *et al.* (2005) have reported breakpoint unrelated imbalances in 3/10 of the patients analysed, including one *de novo* deletion of 2.2-3.4 Mb, one deletion of ~6 Mb and unknown origin and one paternally inherited duplication of 1.1-2.9 Mb. Thus, the proportion of *de novo* imbalances in that study was between 10 and 20%. Similarly, De Gregori *et al.* (2007) reported deletions independent of the breakpoints in 4/27 (14.8%) phenotypically abnormal carriers of reciprocal translocations. The parental origin was determined in only one patient who had three *de novo* deletions, all of paternal origin, of sizes between 1.7 and 4.6 Mb. The other 3 patients had *de novo* deletions between 0.6 and 8.4 Mb of unknown origin. Taken together, the results of Gribble *et al.* (2005), De Gregori *et al.* (2007) and those of the present study indicate that in phenotypically abnormal carriers of apparently balanced rearrangements, the prevalence of candidate disease-causing imbalances independent of the breakpoints is 9/53 (16.98%). The finding of such a significant proportion of genomic imbalances unrelated to the breakpoints might be explained by some degree of genomic instability in these patients. According to this model, ABCRs in association with clinical abnormalities would be a marker for unstable genomes and this instability would be the underlying cause of the cytogenetically visible rearrangement and any submicroscopic imbalances. An argument in favour of the genomic instability concept is the fact that some patients with ABCRs have multiple regions of genomic imbalance independent of the breakpoints as reported in the present study and also by De Gregori *et al.* (2007). However, to test whether cytogenetically visible chromosome rearrangements are markers of genomic instability, it is necessary to compare the rate of array CGH imbalances independent of the breakpoints in patients with clinical abnormalities and abnormal karyotypes to

those of a similar group of patients but with normal karyotypes. Data on this last group of patients has been gathered in the last few years in several studies showing that the rate of imbalances likely to be disease-causing varies between 5 and 16.8% (reviewed by Stankiewicz and Beaudet, 2007). These studies used various array CGH platforms, with different levels of resolution and this, together with different criteria for patient inclusion is likely to contribute to the different estimates provided. In an attempt to obtain some degree of uniformity between these estimates, the different study designs were reviewed. It was noted that some of the studies used a targeted array instead of a whole genome array. Targeted arrays were designed to cover either regions of known clinical significance, regions of segmental duplications or to tile a single chromosome and thus the rate of abnormality derived from these studies is biased because of incomplete coverage of the genome. The results of the nine studies that have used whole genome arrays are summarised in table 6.I., which shows that the rate of genomic imbalances varied between 10 and 16.8%, the average estimate being 13.25%. This result shows that the rate of genomic imbalances in patients with normal karyotypes (13.25%) is lower, but apparently not very different to that of patients with abnormal karyotypes (16.98%). However, these estimates are imperfect as the former was based on studies of >600 patients, whereas the latter was based on 53 cases and many more cases will need to be analysed to evaluate further the hypothesis that ABCRs are markers for genomic instability.

Table 6.I. Rate of array CGH imbalances from studies of clinically abnormal patients with normal karyotypes

Study	No cases	Resolution	No abnormal cases	Ab.R (%)
Vissers *et al.*, 2003	20	1 Mb	3	15.0
Shaw-Smith *et al.*, 2004	50	1 Mb	7	14.0
Rosenberg *et al.*, 2006	81	1 Mb	13	16.0
Schoumans *et al.*, 2005	41	1 Mb	4	9.8
Menten *et al.*, 2006	140	1 Mb	19	13.6
Krepischi-Santos *et al.*, 2006	95	1 Mb	16	16.8
Tyson *et al.*, 2005	22	1 and 3 Mb	3	13.6
de Vries *et al.*, 2005	100	Tiling	10	10.0
Friedman *et al.*, 2006	100	30 kb	11	11.0
Combined data	649	----	86	13.25

Ab.R Abnormality rate

6.6 The origin of genomic imbalances

Candidate disease-causing deletions both at the breakpoints and unrelated to the breakpoints were identified in 4/16 (25%) of the patients reported in the present study. Gribble *et al.* (2005) and De Gregori *et al.* (2007) reported higher rates of 4/10 (40%) and 11/27 (40.7%). Taken together, these results show that deletions detected by array CGH are a significant finding amongst carriers of ABCRs. The parental origin of these deletions was determined in all 4 patients analysed in the present study, on 3 of the cases reported by Gribble *et al.* (2005), on 5 of those reported by De Gregori *et al.* (2007) and on 2 of those reported by Ciccone *et al.* (2005), which showed that 13/14 deletions had arisen *de novo* with their origin on the paternal chromosome. Taken together, these results show that deletions both at the breakpoints and independent of the breakpoints are predominantly of paternal origin, suggesting that spermatogenesis is very much more prone to generate this type of deletion than oogenesis. However, the numbers presented here are too small

to determine whether this paternal bias is a widespread finding and a larger systematic analysis of the parental origin of ABCRs is required to address this question.

Interstitial deletions unrelated to the breakpoints might have occurred by events independent of those that gave rise to the translocation. The observation that these were predominantly paternal in origin is comparable to the findings reported by Thomas *et al.* (2006). The authors determined the parental origin of 44 interstitial deletions and found that 84% were of paternal origin; similarly, other unbalanced rearrangements, including terminal deletions, rings, duplications and unbalanced translocations were found to be mainly paternal, but the extent of paternal bias varied among the different types of abnormalities.

For breakpoint-associated imbalances, it seems reasonable to suppose that these are part of the same events that lead to the translocation. However, despite translocations being a common type of chromosome rearrangement in man there are very few data on their mechanisms of origin, probably because this type of study is technically very challenging and since the great majority of translocations are non recurrent, analysis would need to be done on a case by case basis. Data from studies of the only common non-Robertsonian recurrent translocation, t(11;22)(q23;q11), show that this translocation certainly arises during male meiosis but no data are available on female meiosis (Kurahashi and Emanuel, 2001). By contrast, studies of Robertsonian translocations show that these arise during meiosis, predominantly, but not exclusively, during oogenesis (Page and Shaffer, 1997; Bandyopadhyay *et al.*, 2002). Taken together, these results show that each type of rearrangement is likely to have different underlying mechanisms.

However, it can be speculated that the paternal excess of array CGH deletions is likely to reflect a biological phenomenon as it has long been known that the differences between male and female germ cell development lead to vulnerability to different classes of mutations. Thus, structural rearrangements of all types, with the

exception of Robertsonian translocations, are mainly paternal (reviewed by Chandley, 1991) and deletions detected by array CGH in ABCRs might be no exception.

6.7 Summary

The main aim of the present study was to attempt to answer the following question: are apparently balanced rearrangements molecularly distinct in patients with phenotypic abnormalities by comparison to a control cohort? The findings presented in the preceding chapters suggest that this is the case. The ABCRs in patients with clinical abnormalities are more likely to have breakpoint-associated imbalances, additional chromosomal complexity or genomic imbalances independent of the breakpoints which might account for phenotypic abnormalities. Surprisingly, breakpoint-mediated gene disruption was present at the same frequency in patients with clinical abnormalities and in clinically normal individuals. Both cohorts had potential or obligatory disruption of genes involved in transcription and in signal transduction, but genes with a role in the nervous system are more frequent among the breakpoint regions of patients with clinical abnormalities.

The analysis of the phenotypically normal individuals showed that many genes mapped to the breakpoint regions, most of unknown function but some have been associated with Mendelian disorders, with complex disorders or with microdeletion or microduplication syndromes. Thus, this study has provided data that will prove valuable for future analysis of gene function, to define critical regions for known syndromes and to establish the boundaries between normal variation and pathogenic mutation. It would be of particular interest to review the data presented here in the future and to re-visit each individual in an attempt to determine if there are associations between genes at the breakpoints and late onset disease.

The analysis of the phenotypically abnormal patients enabled the identification of the potential link i.e. gene disruption, genomic imbalances or chromosomal complexity, between the ABCRs and the clinical abnormalities in most of the patients. Genomic imbalances both at the breakpoints and independent of the breakpoints appear to be the most significant cause of clinical abnormalities and were present in 25% of the patients. Breakpoint-mediated gene disruption is almost certainly the cause of phenotypic abnormalities in two patients; one with disruption of the *SATB2* gene and one with disruption of *SNAP25*. In addition, several other candidate genes for phenotypic abnormalities were identified and it would be of interest to undertake further characterisation of these genes to elucidate their role.

In summary, this study provided further insight into apparently balanced chromosome rearrangements both in phenotypically normal individuals and in abnormal patients and will contribute to elucidate the role of the genomic regions involved in these rearrangements, as well as any other genomic regions found to contain copy number changes, to both human disease and to human variation.

APPENDICES

Appendix 1: URLs

1) Children's Hospital Oakland Research Institute (BACPAC Resources): http://bacpac.chori.org/order.php

2) Database of Genomic Variants: http://projects.tcag.ca/variation/

3) DECIPHER (DatabasE of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources): http://www.sanger.ac.uk/PostGenomics/decipher/

4) Department of Genetics and Microbiology, University of Bari (Prof. Rocchi): http://www.biologia.uniba.it/rmc/

5) Ensembl Human Genome Browser: http://www.ensembl.org/Homo_sapiens/

6) GDB Human Genome Database: http://www.gdb.org/

7) GeneCards: http://www.genecards.org/

8) Gene Ontology: http://www.geneontology.org/

9) Human Gene Mutation Database: http://www.hgmd.cf.ac.uk/

10) Invitrogen: http://clones.invitrogen.com

11) Nick translation labelling protocol: http://www.sanger.ac.uk/HGP/methods/cytogenetics/

12) OMIM: http://www.ncbi.nlm.nih.gov/Omim

13) Primer3 program: http://primer3.sourceforge.net/

14) UniProt: http://www.ebi.uniprot.org/

15) University of California Santa Cruz Human Genome browser: http://genome.ucsc.edu/

Appendix 2: Clinical details of the control cohort

Case 1.1: 46,XY,t(2;14)(p21;q13)de novo.ish t(2;14)(p21;q21.1)

Age at assessment	53 years
Birth history	No data
Medical history	Intermittent bilateral hip and knee pain from age 6 years Tonsillectomy in childhood Left hip replacement aged 52 years Right hip replacement pending Hypercholesterolaemia not requiring treatment Allergic to aspirin Teeth all lost secondary to trauma, e.g. rugby, ?'soft' secondary dentition
Developmental history	No history of developmental delay Academic achievement consistent for family
Vision and hearing	No hearing difficulties Myopic since teenage years
Height	173 cm
Centile	25^{th}-50^{th}
Est for parental heights	179 cm
OFC	57 cm
Centile	25^{th}-50^{th}
General examination	Faded strawberry naevus, right thigh
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	None
Family history	None

Case 1.2: 46,XX,t(3;9)(p26.2;p22.3)de novo.ish t(3;9)(p26.3;p23)

Age at assessment	35 years
Birth history	Term
Medical history	Perthe's disease, aged 8 years

	Road traffic accident, aged 14 – sustained metatarsal fracture Age 20 - onset of back pain and sciatica Femoral osteotomy aged 25 years, leg length discrepancy of 2 cm diagnosed subsequently Cluster of fainting episodes in late 20s, no causative diagnosis made
Developmental history	No history of developmental delay Academic achievement consistent for family
Vision and hearing	No hearing difficulties No visual difficulties
Height Centile Est for parental heights	162.5 cm 25^{th}-50^{th} 169 cm
OFC Centile General examination	56 cm 50^{th}-75^{th} Unremarkable
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	None
Family history	None

Case 1.3: 46,X,t(X;7)(?q27;q22)de novo.ish t(X;7)(q28;q31.1)

Age at assessment Birth history	62 years No data
Medical history	No health problems in childhood Cervical polyps Age 38, episode of collapse, secondary to anaemia. No overt loss and no cause found. No further episodes. Arthritis in small joints
Developmental history	No history of developmental delay Academic achievement consistent for family
Vision and hearing	No hearing difficulties

	Amblyopia and reduced vision in left eye
Height	160 cm
Centile	25^{th}-50^{th}
Est for parental heights	163 cm
OFC	57 cm
Centile	75^{th}-91^{st}
General examination	Unremarkable
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	None
Family history	None

Case 2.1: 46,XX,t(4;16)(q35.1;p13.13)de novo.ish t(4;16)(q35.2;p13.2)

Age at assessment	46 years
Birth history	No data
Medical history	No health problems in childhood Episode of epigastric pain aged 45 years, no diagnosis made but splenic cyst found incidentally Prolactinoma diagnosed at 43 years, requiring medical treatment with Cabergoline
Developmental history	No history of developmental delay Academic achievement consistent for family
Vision and hearing	No hearing difficulties No visual difficulties. Normal visual fields.
Height	160 cm
Centile	25^{th}-50^{th}
Est for parental heights	No data
OFC	55 cm
Centile	25^{th}-50^{th}
General examination	Unremarkable
Neurological examination	Normal tone, power and co-ordination in all four limbs

Dysmorphic features	None
Family history	None

Case 2.2: 46,XX,t(1;13)(q32.3;q32.3)de novo.ish t(1;13)(q32.2;q32.1)

Age at assessment	35 years
Birth history	Term
Medical history	Nasal polyps removed aged 12 years Unilateral oophrectomy due to ovarian cyst, age 21 years Recurrent knee dislocations, arthroscopy aged 22 years x1 miscarriage, x1 termination of pregnancy for multiple fetal anomalies, further x6 miscarriages
Developmental history	No history of developmental delay Developed speech aged 3 years Academic achievement consistent for family
Vision and hearing	No hearing difficulties Borderline hypermetropia
Height	150 cm
Centile	0.4^{th}-2^{nd}
Est for parental heights	158 cm
OFC	51.5 cm
Centile	$<0.4^{th}$
General examination	Dental crowding, gum hypertrophy and gum inflammation
Neurological examination	Normal tone and power in all four limbs. Mild intention tremor in both hands. Mildly impaired balance on heel-toe walking.
Dysmorphic features	High arched palate
Family history	Daughter with same translocation, developmentally normal. Single palmar crease.

Case 2.3: 46,XY,t(2;18)(q35;q21.3)de novo.ish t(2;18)(q34;q21.31)

Age at assessment	40 years
Birth history	No data
Medical history	Tonsillectomy aged 6 years No other significant history in childhood Single episode of chest pain during coitus. No cause found; diagnosed as musculo- skeletal
Developmental history	No history of developmental delay Slow learning to read, educated in special needs class until age 10 years. Progress then rapid, completed university education
Vision and hearing	No hearing difficulties No visual difficulties
Height	170 cm
Centile	9^{th}-25^{th}
Est for parental heights	167.5 cm
OFC	57 cm
Centile	25^{th}-50^{th}
General examination	Single large naevus on right thigh No other skin markings
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	None
Family history	None

Case 2.4: 46,XX,t(2;9)(q21.3;p13)de novo.ish t(2;9)(q22.1;p13.3)

Age at assessment	55 years
Birth history	No data
Medical history	Recurrent tonsillitis Hepatitis A aged 25 years Cervical intraepithelia neoplasia, treated by colposcopy and ablation x3 miscarriages Pneumonia aged 44 years

Developmental history	No history of developmental delay Slow learning to read, possibly dyslexic Eventual academic achievement consistent for family
Vision and hearing	No hearing difficulties Myopic
Height	162.5 cm
Centile	25^{th}-50^{th}
Est for parental heights	165 cm
OFC	56 cm
Centile	50^{th}-75^{th}
General examination	Two small naevi on left upper arm
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	None
Family history	Daughter with same balanced translocation has normal phenotype

Case 2.5: 46,XX,t(7;17)(q36.1;q25.1)de novo.ish t(7;17)(q33;q23.2)

Age at assessment	29 years
Birth history	No data
Medical history	Molar pregnancy aged 23 years, presented at 6 weeks gestation. x3 miscarriages
Developmental history	No history of developmental delay Walked at less than 1 year
Vision and hearing	No hearing difficulties No visual difficulties
Height	155 cm
Centile	9^{th}
Est for parental heights	159 cm
OFC	55 cm
Centile	25^{th}-50^{th}

General examination	Unremarkable
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	Small ears, right ear over folded
Family history	Over folded ear found in two other family members

Case 2.6: 46,XY,t(8;15)(p11.2;q24)de novo.ish t(8;15)(p11.21;q24.1)

Age at assessment	59 years
Birth history	Term
Medical history	Limited past medical history No hospital admissions other than to treat numerous fractures, all due to horse riding accidents. Single episode of 'blackout' aged 21 years. Depression aged 58 years Partner x11 miscarriages
Developmental history	No data Unable to read, poor writing skills, but academic achievement consistent for family
Vision and hearing	No hearing difficulties Hypermetropia, recent onset. Previously no difficulties.
Height	162.5 cm
Centile	0.4^{th}-2^{nd}
Est for parental heights	154 cm
OFC	55 cm
Centile	9^{th}
General examination	Unremarkable
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	None
Family history	None

Case 2.7: 46,XY,t(1;13)(p22;q32)de novo.ish t(1;13)(p21.2;q31.3)

Age at assessment	58 years
Birth history	Term
Medical history	Ear polyps diagnosed aged 5 years. Road traffic accident aged 19 causing multiple fractures Low sperm count; conceived after 10 years Deep vein thrombosis, secondary to trauma aged 40 years.
Developmental history	No history of developmental delay Mild dyslexia
Vision and hearing	Mild hearing impairment secondary to loud noise exposure No visual difficulties
Height	188 cm
Centile	91^{st}-98^{th}
Est for parental heights	No data
OFC	59 cm
Centile	75^{th}-91^{st}
General examination	Unremarkable
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	No dysmorphic facial features Feet - low arches
Family history	Son with same translocation, with developmental delay and epilepsy

Case 2.8: 46,XY,t(11;21)(p15.4;p12)de novo.ish t(11;21)(p15.4;p12)

Age at assessment	33 years
Birth history	Born at 39/40 weeks, 3.5 Kg
Medical history	No childhood medical problems Left knee arthroscopy and trimming of cartilage Right knee recurrent bursitis

Developmental history	No history of developmental delay Academic achievement consistent for family
Vision and hearing	No hearing difficulties Right amblyopia, under treated, subsequent poor vision in right eye
Height	185 cm
Centile	75^{th}-91^{st}
Est for parental heights	174 cm
OFC	60.5 cm
Centile	91^{st}-98^{th}
General examination	Raised body mass index
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	High arched palate No dysmorphic facial features Feet- low arches
Family history	Two sons with Beckwith-Wiedemann syndrome, due to segregation of unbalanced product of the paternal translocation

Case 3.1: 46,XX,t(4;6)(q27;p25)de novo.ish t(4;6)(q28;p24.3)

Age at assessment	41 years
Birth history	No data
Medical history	Asthma diagnosed at 5 years. Mild asthma only as an adult Childhood eczema x7 first trimester miscarriages, originally thought due to cervical incompetence Breast Cancer, aged 37 years Metastatic disease diagnosed aged 40 years
Developmental history	No history of developmental delay Academic achievement consistent with family
Vision and hearing	No hearing difficulties No visual difficulties

Height	152.5 cm
Centile	2^{nd}-9^{th}
Est for parental heights	168 cm
OFC	54 cm
Centile	9^{th}-25^{th}
General examination	Cachetic
Neurological examination	Bradykinesia, consistent with terminal disease No focal neurological signs
Dysmorphic features	None
Family history	None

Case 3.2: 46,XX,t(8;12)(p23.1;p13.1)de novo.ish t(8;12)(p23.1;p13.31)

Age at assessment	35 years
Birth history	Term
Medical history	No medical problems as a child Breast reduction surgery, requiring repeat procedures x2 miscarriages
Developmental history	No history of developmental delay Academic achievement consistent with family
Vision and hearing	No hearing difficulties No visual difficulties
Height	162.5 cm
Centile	25^{th}-50^{th}
Est for parental heights	No data
OFC	56 cm
Centile	50^{th}-75^{th}
General examination	Single café au lait patch on left aspect of neck
Neurological examination	High arched palate Normal tone, power and co-ordination in all four limbs
Dysmorphic features	None

Family history	Son with unbalanced translocation, multiple medical problems and global delay

Case 3.3: 46,XX,t(6;9)(q22.2;p22.3)de novo.ish t(6;9)(q22.31;p22.3)

Age at assessment	36 years
Birth history	No data
Medical history	Recurrent chest infections as a child Mild asthma Heart murmur diagnosed in childhood; resolved spontaneously x3 miscarriages
Developmental history	No history of developmental delay Speech and language mildly delayed Academic achievement consistent for family
Vision and hearing	No hearing difficulties No visual difficulties
Height	162.5 cm
Centile	25^{th}-50^{th}
Est for parental heights	164 cm
OFC	57 cm
Centile	75^{th}-91^{st}
General examination	Area of freckling on right forearm, 2 cmx3 cm
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	None
Family history	Daughter with developmental delay and epilepsy, normal karyotype, diagnosis: cerebral palsy

Case 3.4: 46,XY,t(2;4)(p23;p12)de novo.ish t(2;4)(p23.3;p12)

Age at assessment	38 years
Birth history	Term
Medical history	Right orchidopexy aged 9 years

	Renal stone aged 11 years, no underlying cause found. Ruptured left sided pre-patellar bursa aged 35 years.
Developmental history	No history of developmental delay Academic achievement consistent for family
Vision and hearing	No hearing difficulties Right strabismus in infancy, corrected at 2 years
Height	178 cm
Centile	50^{th}-75^{th}
Est for parental heights	176.5 cm
OFC	58 cm
Centile	50^{th}-75^{th}
General examination	Unremarkable
Neurological examination	Normal tone and power Mild dysdiadodyskinesis; lower limb co-ordination normal
Dysmorphic features	Ears - prominent crus
Family history	None

Case 3.5: 46,X,t(X;22)(p11.23;q13.1)de novo.ish t(X;22)(p11.1;q11.21)

Age at assessment	48 years
Birth history	No data
Medical history	Rheumatic fever aged 4 years Tonsillectomy aged 5 years Borderline hypertension
Developmental history	Walked aged 1 year Talked 'early' Academic achievement consistent with family
Vision and hearing	No hearing difficulties No visual difficulties
Height	162.5 cm
Centile	25^{th}-50^{th}
Est for parental heights	168 cm

OFC	57 cm
Centile	75^{th}-91^{st}
General examination	Unremarkable
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	Feet - large for height, flat arches
Family history	Daughter with unbalanced product of maternal translocation. Mild Asperger's syndrome, memory/concentration difficulties

Case 3.6: 46,XX,t(11;15)(q23;q22)de novo.ish t(11;15)(q23.3;q21.2)

Age at assessment	57 years
Birth history	No data
Medical history	No medical problems during childhood x1 miscarriage Umbilical hernia, repaired aged 55 years
Developmental history	No history of developmental delay Academic achievement consistent for family
Vision and hearing	Mild hearing loss, age related Myopia, prescription -5, required glasses since 8 years
Height	157.5 cm
Centile	9^{th}-25^{th}
Est for parental heights	160 cm
OFC	56.5 cm
Centile	75^{th}-91^{st}
General examination	Unremarkable
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	None
Family history	None

Case 3.7: 46,XX,t(2;6)(q32.2;p23)de novo.ish t(2;6)(q33.1;p25.1)

Age at assessment	41 years
Birth history	No data
Medical history	Tonsillectomy aged 13 years 34 years: severe road traffic accident, involving fractured femur and wrist. Residual leg length discrepancy. Blackout' during admission, ?embolic episode Following dental work, has been commented that teeth are very dense x4 miscarriages Breast cyst
Developmental history	No history of developmental delay Academic achievement consistent for family
Vision and hearing	No hearing difficulties No visual difficulties
Height	160 cm
Centile	25^{th}-50^{th}
Est for parental heights	161.5 cm
OFC	55.2 cm
Centile	25^{th}-50^{th}
General examination	Small strawberry naevus on breast
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	Ears - large upper part of pinnae, small ear lobes, normally rotated Feet - flat arches
Family history	Son has normal phenotype and same balanced translocation

Appendix 3: Clinical details of the patient cohort

Case 16: 46,XX,t(10;22)(q24.3;q13.31)*de novo*.ish t(10;22)(q24.31;q13.31)

Age at assessment	34 years
Birth history	Term. Born in good condition
Medical history	Absence seizures from 4 years. Age 6 years grand mal seizures, diagnosed as temporal lobe epilepsy. Psychiatric history, beginning at 11 years. Obsessional behaviour and anxiety requiring prolonged psychiatric admission. Obesity, treated with gastric by-pass.
Developmental history	Sat at 6 months. Walked at 18 months. Speech development normal. No concerns until age 11 years. Adult IQ 74
Vision and hearing	Mild hearing impairment. No visual difficulties
Height	162.5 cm
Centile	25-50th
Est for parental heights	158 cm
OFC	58 cm
Centile	98th Centile
General examination	Flat affect with monotonal voice, high body mass index
Neurological examination	Bradykinesia. Normal tone and power in all four limbs, poor co-ordination; to heel-toe walk. Normal eye movements.
Dysmorphic features	Downslanting palpebral fissures.
Family history	Brother with temporal lobe epilepsy and normal karyotype

Case 20: 46,XX,t(2;5)(q33;q12)*de novo*.ish t(2;5)(q33.2;q12.1)

Age at assessment	32 years
Birth history	38/40. Birth weight 2.18kg. 2nd Centile

Medical history	Quiet and hypotonic from birth Ventilated for 1/7, required nasogastric feeds for 4/7. Remained hypotonic until ambulant. Menorrhagia. Migraines throughout teenage years.
Developmental history	Sat at 6 months. Walked at 14 months. Speech development slow, main difficulty with understanding. Attended mainstream school with support
Vision and hearing	No hearing difficulties. Amblyopia
Height	157.5 cm
Centile	9^{th}-25^{th}
Est for parental heights	158 cm
OFC	55 cm
Centile	25^{th}-50^{th}
General examination	Slow, slightly slurred speech. Quiet, shy
Neurological examination	Normal tone and power in all four limbs. Poor co-ordination. Unable to heel-to-toe walk. Poor upper limb co-ordination.
Dysmorphic features	Coarse facial features. Dark, curly hair. Single palmar crease on left hand
Family history	Sister with anorexia nervosa

Case 43: 46,XY,t(4;17)(q35.1;q25.1)*de novo*.ish t(4;17)(q34.3;q24.2)

Age at assessment	3 months
Medical history	Isolated truncus arteriosus
Developmental history	Normal development

Case 44: 46,XY,inv(6)(p24q16.2)*de novo*.ish inv(6)(p24.3q16.2)

Age at assessment	2 years
Birth history	no data

Developmental history	mild global developmental delay, very specific speech and language problems, with a disordered rather than delayed speech (? possible palatal dysfunction)
Neurological examination	Learning problems
Dysmorphic features	Heterochromia of one iris, slightly coarse facies
Family history	no data

Case 45: 46,X,t(X;19)(q21;p13.11)*de novo*.ish t(X;19)(q21.1;q11.?)

Age at assessment	32 years
Birth history	No data
Medical history	No medical problems as a child. Menarche aged 12 years. Menses initially erratic and then regular. Amenorrhoea from age 31 years. No hot flushes or other menopausal symptoms. Gonadotrophin levels high consistent with premature menopause
Developmental history	No delay known in early milestones. Attended mainstream school
Height	157 cm
Centile	9^{th}-25^{th}
Est for parental heights	No data
General examination	Unremarkable
Neurological examination	Neurology grossly intact
Dysmorphic features	None
Family history	No data

Case 48: 46,XY,t(4;6)(q33;q22.2)*de novo*.ish t(4;6)(q33-34.1;q21)

Age at assessment	15 years
Developmental history	Severe developmental delay
Dysmorphic features	Present

Case 49: 46,XX,t(2;10)(q33;q21.2)*de novo*.ish t(2;10)(q33.1;q21.3)

Age at assessment	21 years
Medical history	Isolated cleft palate repaired at 6 months. Two operations for a scoliosis
Developmental history	No speech, but good communication with British sign language
General examination	Very sociable but with significant behavioural abnormalities
Neurological examination	No data
Dysmorphic features	Long thin fingers with hypoplastic distal phalanges. Long face with a pointed chin and a high anterior hairline with a cowlick. Coarse features, including a long nose.

Case 50: 46,X,t(X;8)(q22.1;q24.13).ish t(X;8)(q23;q24.3)

Age at assessment	30 years
Birth history	No data
Medical history	Menarche aged 14 years. Menses at one year intervals. Aged 19 years took combined oral contraceptive pill for 9 years; on discontinuation, no resumption of menstruation, associated with high gonadotrophins
Developmental history	No delay known in early milestones
Vision and hearing	No hearing difficulties No visual difficulties
Height	166 cm
Centile	50^{th}-75^{th}
Est for parental heights	159 cm
OFC	56 cm
Centile	50^{th}-75^{th}
General examination	Slight build, long fingers
Neurological examination	Neurology grossly normal

Dysmorphic features	No dysmorphic features
Family history	Mother had very irregular menses, menarche at 15 years, menopause at 50 years, normal karyotype

Case 51: 47,XX,t(4;20)(p15.2;p11.23)*de novo*,+mar[23]mat /46,XX,t(4;20)(p15.2;p11.23)[7].ish t(4;20)(p15.32;p12.2)

Age at assessment	6 years
Birth history	Term. 2.86kg. Feeding difficulties, wouldn't breastfeed
Medical history	Autistic Spectrum Disorder and Attention Deficit Hyperactivity Disorder aged 6 years Sleeps poorly, requiring melatonin Shows mannerisms, repetitive behaviours and cself injurious behaviour No seizures
Developmental history	Rolled at 1 year. Sat at approximately 1 year Walked at 19 months No language until 4 years Severe language delay Attends special needs school
Vision and hearing	No hearing difficulties No visual difficulties
Height	131.6 cm
Centile	75^{th}-91^{st}
Est for parental heights	Tall for family
OFC	50.5 cm
Centile	2^{nd}
General examination	Tall stature 1 café au lait patch on left back
Neurological examination	Neurology grossly normal
Dysmorphic features	Large, upslanting eyes Well defined philtrum Mild fifth finger clinodactyly
Family history	Brother with 47XYY in 90% of cells and marker chromosome 22

Case 52: 46,XX inv ins (11;4)(q22.2;q13.2q21.3)*de novo*.ish inv ins (11;4)(q22.1;q13.1q21.23)

Age at assessment	13 years
Birth history	No data
Medical history	Respiratory problems and vocal cord palsy which required tracheostomy. Scoliosis (60%) required surgery. Learning difficulties Normal growth rate. Short stature.
Developmental history	Delayed speech
Centile	2nd
General examination	Asymmetry more marked in the feet
Dysmorphic features	Slight facial dysmorphism

Case 53: 46,XX,t(4;8)(q21.1;p12)*de novo*.ish t(4;8)(q13.3;p12)

Age at assessment	7 years
Birth history	41/40. 3.83kg. Born in good condition
Medical history	Failed to fix with eyes at 3/12 8/12 hypotonia identified. Episode of loss of use of left upper and lower limbs, aged 3 years. CT scan normal, recovery made but slight weakness remained. Seizure at 8/12 ?febrile fit. Initial MRI brain - delayed myelination. Repeat scan aged 6 years, complete myelination
Developmental history	Smiled at 6 weeks. Rolled at 6 months. Sat at 18 months. Able to walk with support. No speech, understands limited number of words
Vision and hearing	Hearing 'excellent' ?hyperacusis Left strabismus
OFC	52 cm
Centile	25th centile

General examination	Left strabismus, quiet
Neurological examination	Hypotonic, spontaneous movements of all four limbs. Deep tendon reflexes present, plantars downgoing
Dysmorphic features	No dysmorphic facial features Slender fingers

Case 54: 46,XY,t(14;15)(q23;q26.3)*de novo*.ish t(14;15)(q22.3;q26.2)

Age at assessment	11years
Birth history	Term. 3.8kg. Born in good condition
Medical history	Co-arctation of aorta diagnosed at 8 weeks. Associated with cardiomegaly, pneumothorax and pneumonia x2 further pneumothoraces in first year. Repeated chest infections. Undescended testes, orchidopexy at 3 years. Behavioural difficulties, on autistic spectrum
Developmental history	Obsessional tendencies but very sociable. Hypercholesterolaemia, diet controlled. Smiled at 10 weeks. Sat at 9 months. Walked at 17 months. Speech delay. Attends mainstream school, mild learning difficulties
Vision and hearing	No hearing difficulties Left strabismus, corrected at 5 years
Height	No data
Centile	No data
Est for parental heights	Tall for family
OFC	57 cm
Centile	98^{th}-99.6^{th}
General examination	Tall stature Left strabismus
Neurological examination	Normal tone and power in all four limbs Mild ataxia demonstrated by reduced ability to heel-toe walk
Dysmorphic features	Ears - large ear lobes, flat pinnae Hands - brachydactyly, mild joint hypermobility Feet - wide with short toes, low arches

	Pectus excavatum
Family history	Father has hypercholesterolaemia

Case 55: 46,XY,t(19;20)(q13.43;q11.1)*de novo*.ish t(19;20)(q13.43;q11.1)

Age at assessment	42 years
Birth history	No data
Medical history	No childhood medical problems Allergies, including hayfever Gastro-oesophageal reflux, diet controlled Gum disease and poor dentition Low sperm count Partner had 'multiple' miscarriages
Developmental history	No known delay in early milestones Attended mainstream school Academic achievement consistent with family
Vision and hearing	No hearing difficulties No visual difficulties
Height	175 cm
Centile	25^{th}-50^{th}
Est for parental heights	176 cm
OFC	60.5 cm
Centile	91^{st}-98^{th} (Brother also has a large head)
General examination	Obese
Neurological examination	Normal, tone power and co-ordination in all four limbs. High palate
Dysmorphic features	Hyperteloric Long, slim fingers
Family history	No data

Case 56: 46,XY,t(6;21)(q16.2;q11.2)*de novo*.ish t(6;21)(q16.1;q11.1)

Age at assessment	39 years

Birth history	No data
Medical history	No childhood medical problems Infertility, with very low spem count, mainly immotile
Developmental history	No known delay in early milestones Academic achievement consistent with family Difficulties in visuo-spatial assessment, For example unable to drive
Vision and hearing	No hearing difficulties. No visual difficulties
Height Centile Est for parental heights	180 cm 50^{th}-75^{th} 178 cm
OFC Centile General examination	57.5 cm 50^{th}-75^{th}
Neurological examination	Normal tone, power and co-ordination in all four limbs
Dysmorphic features	No dysmorphic facial features. Feet- very flat arches. Pectus carinatum
Family history	No data

Case 57: 46,XY,t(2;5)(p23;q11.2)*de novo*,t(18;22)(q11.2;p13)*de novo*.ish t(2;5)(p23.3;q11.2),t(18;22)(q11.2;p13)

Age at assessment	43 years
Birth history	41/40. Birth weight 2.55kg. 2^{nd} centile. Quiet on delivery, required nasogastric feeding for 3 weeks
Medical history	Epilepsy diagnosed at 2 years, grand mal seizures. Had a seizure every few months during childhood. No episodes of status epilepticus. Hypertension, diagnosed at 40 years. Hypercholesterolaemia. Joint laxity, knee dislocation requiring surgery
Developmental history	Sat at around 8 months. Walked at 16 months. Speech development normal. Attended mainstream

	school. Paid unskilled employment, lives independently. Academic achievement less than unaffected family members
Vision and hearing	No hearing difficulties. No visual difficulties
Height	160 cm
Centile	0.4^{th}-2^{nd} centile
Est for parental heights	171.5 cm
OFC	No data
Centile	No data
General examination	Speech slow but not dysarthric. Normal palate. Dental crowding
Neurological examination	Normal tone and power in all four limbs. Poor co-ordination of upper and lower limbs. Unsteady gait, unable to heel-toe walk. Normal eye movements
Dysmorphic features	No dysmorphic facial features
Family history	Brother with normal karyotype and intrauterine growth retardation, stellate cataract strabismus, short stature, learning difficulties and depression. No facial dysmorphic features

Case 58: 46,XX,t(6;18)(p23;q22)*de novo*.ish t(6;18)(p24.3;q21.1)

Age at assessment	26-27 years
General examination	Primary amenorrhea. Streak left ovary and primordial folicules on the right ovary
Dysmorphic features	none reported

Appendix 4: Cell line identification numbers

CASE	WRGL[1]	ECACC[2]
1.1	LN2/9C63 & LN3/1A24	-
1.2	LN2/9C57 & LN3/1A17	-
1.3	LN2/9C61 & LN3/1A27	
1.4	LN2/8B65/66 & LN2/8C26/27	DD0946
2.1	LN2/9C60 & LN3/1A23	
2.2	LN2/9C47/48 & LN3/1A10	-
2.3	LN2/9C55 & LN3/1A19	-
2.4	LN2/9C54 & LN3/1A18	
2.5	LN2/9D60 & LN3/1B56	-
2.6	LN2/9D58 & LN3/1B54	-
2.7	LN2/6B97/98 & LN2/6C74/75	DD0665
2.8	LN2/9C49/50 & LN3/1A11	-
2.9	LN2/7C21/22 & LN2/6A72	DD0348
2.10	LN2/6D89/90 & LN2/7A64/65	DD0505
2.11	LN2/6D83/84 & LN2/7A58/59	DD1407
2.12	LN2/6D87/88 & LN2/7A62/63	DD0845
2.13	LN2/7B43/44/53/54	DD1521
2.14	LN2/8B59/60 & LN2/8B95/96	DD0182
2.15	LN2/8B61/62 & LN2/8B97/98	DD0282
2.16	LN2/8B63/64 & LN2/8B99/100	DD1277
3.1	LN2/9D27/29	-
3.2	LN2/9C56 & LN3/1A16	-
3.3	LN2/9C62 & LN3/1A28	
3.4	LN2/9D56 & LN3/1B52	
3.5	LN2/9D74 & LN3/1B58	
3.6	LN2/9C66 & LN3/1A39	
3.7	LN2/9D24 & LN3/1B20/21	
3.8	LN2/6D85/86 & LN2/7A60/61	DD1319
3.9	LN2/6D81/82 & LN2/7A56/57	DD2058
3.10	LN1/3D19/20 & LN1/3B03/04	-
3.11	LN2/7A90/91 & LN2/7B41/42	DD3206
16	LN2/8A18/ 19 /24 /25	DD1713
20	LN1/2E/70/71	DD0762
43	LN2/9C/53 LN3/1A/20	-
44	LN1/3A/45/46 LN2/9C/70	DD3560
45	LN2/9C/59 LN3/1A/26	-

CASE	WRGL[1]	ECACC[2]
48	LN2/9C/78/81	-
49	LN2/9C/97/98 LN3/1A/88	-
50	LN2/9C/58/69 LN3/1A/25	-
51	LN2/9D/13 LN3/1B/11/12	-
52	LN2/9D/19/20 LN3/1B/15	-
53	LN2/9D/25 LN3/1B/22/23	
54	LN2/9D/32/33 LN3/1B/33	
55	LN2/9D/53 LN3/1B/46	-
56	LN2/9D/54 LN3/1B/47	-
57	LN2/9D/57 LN3/1B/53	-
58	failed	-

1WRGL Wessex Regional Genetics Laboratory
[2]ECACC European Cell and Culture Collection

Appendix 5: Hybridisation program for the Tecan hybridisation station HS 4800

The hybridisation station was connected to a liquid unit in which the channels were defined as follows:

channel 1: 1xPBS/0.05% Tween 20
channel 3: 0.1xSSC
channel 5: dH_2O

Step 1.	Wash: Temperature.: 25°C, Channel 1, Time: 30 s
Step 2.	Probe injection: Temperature.: 37°C
Step 3.	Hybridisation: Temperature.: 37°C, Medium agitation frequency, Time: 45 min
Step 4.	Probe injection: Temperature.: 37°C
Step 5.	Hybridisation: Temperature.:37°C, Medium agitation frequency, Time: 21 h.
Step 6.	Wash: Temperature.: 37°C, Channel: 1, Runs: 15, Wash time: 30 s, Soak time: 30 s
Step 7.	Wash: Temperature.: 54°C, Channel: 3, Runs: 5, Wash time: 1 min, Soak time: 2 min
Step 8.	Wash: Temperature.: 25°C, Channel: 1, Runs: 10, Wash time: 30 s, Soak time: 30 s
Step 9.	Wash: Temperature.:25°C, Channel: 5, Runs: 1, Wash time: 30 s
Step 10.	Slide drying: Temperature.:25°C, Time: 2 min 30 s. Final Manifold cleaning: Yes, Channel 5.

The pre-hybridisation probe added in step 2 was prepared by precipitating 100 µL of 10 mg/mL Herring sperm in ethanol and ressuspending it in 120 µL of hybridisation buffer (50% formamide, 5% dextran sulphate, 0.1%Tween 20, 2xSSC, 10mM Tris pH7.4).

Appendix 6: Primers for breakpoint mapping studies using PCR based approaches

Primer name	Target	Chr	Left primer (5' to 3')	Right primer (5' to 3')
P1_chr2	case 49	2	AATTGAATGATTTCCCATCAGC	GCTGCCAGAGGTTCTAAGAAAA
P2_chr2	case 49	2	TGTGCACTAAATTGCTGACTCC	GAGAATACGGTCATGGCATTG
P3_chr2	case 49	2	CCCTTGAGTTAGGAGCTTTGAA	ATGCAGTGTCCAGATCAGAATG
P4_chr2	case 49	2	GAGGGTCTTCTTCCTTACGC	AAGGCAAGCACATTGACTTC
P5_chr2	case 49	2	AGTCAGGGTTTTCATGTGCTCT	AATCCAACAAGTGTGCAGAATG
P6_chr2	case 49	2	GGTGGGCACAAGAAGAGTTTTA	TAGGTGGGGTGATTTGTGGTAT
P7_chr2	case 49	2	TTGTCACCTGCAACAAATTAGG	AACTGCCACAAACAGAAACAAA
P8_chr2	case 49	2	TTTCAGGCACCCTAAGTCTG	TGTGGCTTGGTCAAAGGTAT
P9_chr2	case 49	2	GGCTATGCACAATCCAAGTC	GTAAATTTTTGGGGCAGGTT
P10_chr2	case 49	2	GATGCAAAATGACCACATGA	AATCCTTCATGGAACAATGC
P11_chr2	case 49	2	TCACTAGAGGTAAAAGTCCTCATTG	TCTTTTCCTTTCATCCCTGA
P12_chr10	case 49	10	GGAAATCTGGGGAATGTAGGA	TGCAATGAAGGCTTAACAATC
P13_chr10	case 49	10	AAGCAACAGTCATGGGCTCT	ATGGTGTGACAAGCCACAAA
P14_chr10	case 49	10	TTCCCTAAGAAATACTAAGCACTACA	GCACACACACAAAATGAGACC
P15_chr10	case 49	10	GCATCTGAACCATGTCAGGA	CAGCAGCCAAACAGAAAACA
P16_chr10	case 49	10	GGCAGGAGCCTTTTATCTTC	TGCCTCAGTCCTAGAGGTTG
P17_chr10	case 49	10	TGGGAATAGGACATGGGTCT	GGAGCTTATTGCAGGGTGAA
P18_chr10	case 49	10	GTTTCCAAACCACTCCAATG	TTCCATGTGTGAATGAGTGC
P19_chr10	case 49	10	CAGCCCCAATACTTGGAAAA	CCTGGCATGTCCTGTTTGTA
P20_chr4	case 51	4	ATAAGATCTGTGCCATGCCCTAA	TTTTGAGTCTCTGGTCTTGGGAG

Primer name	Target	Chr	Left primer (5' to 3')	Right primer (5' to 3')
P21_chr4	case 51	4	ACTACTGGGCAAGCTGAATG	GCAGAGGAAACCCTAGGAAG
P22_chr4	case 51	4	CTTGTGTGGGAGAGAGCAGT	TGTCTGTCCATCCTGAACCT
P23_chr4	case 51	4	Caaaatccctcatcatttgc	aacagtggacctctcagcag
P24_chr4	case 51	4	gggaaatccaggttctgtct	gtcaagatttgtccccacag
P25_chr4	case 51	4	ggggaggctagtaaaagtgc	GCTCATCCATCTCATCTGCT
P26_chr4	case 51	4	ctccaacatgaatcccaaag	AGGCCATTTCAGGGAATAAC
P27_chr4	case 51	4	TTCATTGCTACTGGGGTGTTGTC	TAATCTGTGTGCTTACATTTTAGGG
P28_chr4	case 51	4	CGTACCTCCTCTACACCCTCC	GCTCAGGTAACATTCGTGCA
P29_chr4	case 51	4	AGCTGGGATTTGAACTCTTCC	ACCAGGATTGTCAAGTAGCCA
P30_chr4	case 51	4	GTACCCAAGAGAAAGCCC	GGGATGTAAAATTGGGATG
P31_chr4	case 51	4	CCAAATAATTCAGGATAATCTTGCC	TGTGAGAGCAGAAGAGACCAACT
P33_chr20	case 51	20	CCCCAATTTCACTGATGAAG	TGCTTTAGGCAGAATGAGGA
P34_chr20	case 51	20	TTTTTTTGGCCAAAATATTAACTC	TTTGTTTTTCCCTTTTTAATATCAA
P35_chr20	case 51	20	AATGCTGTGGCCTTTTCATA	GTCCATCCCTTCCTCAATG
P36_chr20	case 51	20	ATTCGCTTGGTTCGATTCTT	ACAGTGTATGCTGCATGAGG
P37_chr20	case 51	20	ACCCAGCTCCTCTTACCCTC	GCAATAGTTACGACTGTGGGC
P38_chr20	case 51	20	CCGAGAAAATGAAATGGATG	GAATCCAAGGTCGTGTTTTG
P39_chr20	case 51	20	TAGAGCCATAGCCGTCATTC	CACTTCCCAGCATCTTTGTT
P40_chr20	case 51	20	TTAAGTGTGCCCACCCGTG	CAAAAGACACAAAGTATTGTTGGG
P41_chr2	case 49	2	TGCCATGACCGTATTCTCAA	
P42_chr2	case 49	2	AAGTATGCCTGGGCTCTTCT	
P43_chr2	case 49	2	GCCCATTTTTCAGCTTTGTA	
P44_chr2	case 49	2		TACAAGGCCCCACAGTTAGG

Primer name	Target	Chr	Left primer (5' to 3')	Right primer (5' to 3')
P45_chr10	case 49	10	atttcccatcagcgctaaga	
P46_chr10	case 49	10	ctgacatgggaaaatgagca	
P47_chr10	case 49	10		Ctgtagtgcagggttgtcca
P48_chr10	case 49	10		AACAAATGGCCTAGAACCAA
P49_chr10	case 49	10		Tgctcattttcccatgtcag
P50_chr10	case 49	10		tatgcaggatggtgtgacag

Appendix 7: Microsatellite markers for parent of origin studies

Case	Chr	Markers[1]
2.11	3p21.31	D3S3640
2.11	3p21.31	D3S643
2.11	3p21.31	D3S3629
2.11	3p21.31	D3S1359
20	5q12.1-q12.3	D5S1474
20	5q12.1-q12.3	D5S76
20	5q12.1-q12.3	D5S1718
20	5q12.1-q12.3	D5S427
20	5q12.1-q12.3	D5S1956
20	5q12.1-q12.3	D5S1359
52	4q13.3	D4S2641
52	4q13.3	D4S2389
52	4q21.23	D4S1534
52	4q21.23	D4S2691
53	4q13.3-q21.1	D4S3249
53	4q13.3-q21.1	D4S2958
53	4q13.3-q21.1	D4S1558
53	10-p14	D10S1649
53	10-p14	D10S465
57	4q32.1	D4S3016
57	4q32.1	D4S1556
57	4q32.1	D4S1498

[1]Primer details available within the Ensembl and UCSC genome browsers

Appendix 8: GO terms for genes mapped to the breakpoint regions in the control cohort

Gene	Case	GO term	Category
ASF1A	3.3	transcription	Transcription
FOXJ2	3.2	transcription	
FRYL	3.4	regulation of transcription, DNA-dependent	
GABPB2	3.6	transcription	
SUPT4H1	2.5	transcription factor activity	
SUPT7L	3.11	transcription	
ZNF512	3.11	regulation of transcription, DNA-dependent	
ZNF74	3.5	regulation of transcription, DNA-dependent	
C3AR	3.2	signal transduction	Signal transduction
IL27RA	3.9	cell surface receptor linked signal transduction	
PLXNA2	2.2	signal transduction	
QRICH2	2.1	signal transduction	
ROCK1	2.14	Rho protein signal transduction	
RYR2	2.15	signal transduction	
STK3	2.12	signal transduction	
CD93	3.1	cell-cell adhesion	Cell adhesion
CLEC4A	3.2	cell adhesion	
CNTN5	2.15	cell adhesion	
COL13A1	2.16	cell-cell adhesion	
FAT	2.1	cell adhesion	
NPTN	2.6	homophilic cell adhesion	
ATP2C2	2.13	calcium ion transport	Transport
COL13A1	2.16	phosphate transport	
EXOC4	2.5	protein transport	
NECAP1	3.2	protein transport	
RYR2	2.15	calcium ion transport	
SLC20A2	2.6	transport	
ROCK1	2.14	protein amino acid phosphorylation	Protein modification
ST8SIA3	2.3	protein amino acid glycosylation	
STK3	2.12	protein amino acid phosphorylation	
UHRF2	3.1	protein modification process	
WWP2	2.11	protein modification process	

Gene	Case	GO term	Category
CD93	3.1	macrophage activation	Immunity
DEFA1	3.2	immune response	
DNAJC3	2.2	defense response	
IL27RA	3.9	defense response to Gram-positive bacterium	
SPOPL	2.4	ubiquitin cycle	Ubiquitination
UHRF2	3.1	ubiquitin-dependent protein catabolic process	
USP41	3.5	ubiquitin-dependent protein catabolic process	
WWP2	2.11	ubiquitin cycle	
ROCK1	2.14	apoptosis	Apoptosis
SPIN2B	3.5	apoptosis	
STK3	2.12	apoptosis	
UNC13B	2.4	induction of apoptosis	
COL13A1	2.16	cell differentiation	Cell proliferation/ differentiation
SPAG16	2.3	cell projection organization and biogenesis	
UHRF2	3.1	cell proliferation	
COL13A1	2.16	multicellular organismal development	Multicellular development
PLXNA2	2.2	multicellular organismal development	
ZNF74	3.5	multicellular organismal development	
SPIN2B	3.5	cell cycle	Cell cycle
UHRF2	3.1	regulation of cell cycle	
ABAT	2.1	neurotransmitter catabolic process	Nervous system
NPTN	2.6	positive regulation of long-term neuronal synaptic plasticity	

Appendix 9: GO terms for genes mapped to the breakpoint regions in the patient cohort

Gene	Case	GO term	Category
CXCL6	53	signal transduction	Signal transduction
HTR1A	20	signal transduction	
MAPK10	52	signal transduction	
NCOA1	57	signal transduction	
NR2F2	54	signal transduction	
PITPNC1	43	signal transduction	
UNC5D	53	signal transduction	
CXCL1	53	nervous system development	Nervous system
DCC	58	axon guidance	
KIF2A	20	nervous system development	
NR2F2	54	neuron migration	
PAX2	16	axonogenesis	
SNAP25	51	neurotransmitter secretion/uptake	
NCOA1	57	positive regulation of transcription, DNA-dependent	Transcription
NR2F2	54	transcription	
PAX2	16	transcription	
SATB2	49	regulation of transcription, DNA-dependent	
ZNF606	55	transcription	
CXCL1	53	negative regulation of cell proliferation	Cell proliferation/ differentiation
HTR1A	20	positive regulation of cell proliferation	
IL8	53	negative regulation of cell proliferation	
KIF2A	20	cell differentiation	
PAX2	16	cell differentiation	
AGTR2	50	G-protein coupled receptor protein signalling pathway	G-protein receptor
CXCL1	53	G-protein coupled receptor protein signalling pathway	
HTR1A	20	G-protein coupled receptor protein signalling pathway	
IL8	53	G-protein coupled receptor protein signaling pathway	
DCC	58	multicellular organismal development	Multicellular development
KIF2A	20	multicellular organismal development	
PAX2	16	multicellular organismal development	
UNC5D	53	multicellular organismal development	
AGTR2	50	apoptosis	Apoptosis
DCC	58	apoptosis	
UNC5D	53	apoptosis	
AGTR2	50	behaviour	Behaviour
HTR1A	20	behaviour	

Gene	Case	GO term	Category
DCC	58	negative regulation of cell cycle	Cell cycle
IL8	53	cell cycle arrest	
CXCL6	53	cell-cell signalling	Cell-cell signalling
IL8	53	cell-cell signalling	
ABI2	20	cell migration	Cell migration

REFERENCES

Albertson DG, Pinkel D. Genomic microarrays in human genetic disease and cancer. Hum Mol Genet 2003; 12 Spec No 2:R145-52. Epub@2003 Aug 5.:R145-R152.

Allen-Baume V, Segui B, Cockcroft S. Current thoughts on the phosphatidylinositol transfer protein family. FEBS Lett 2002; 531(1):74-80.

Aradhya S, Manning MA, Splendore A, Cherry AM. Whole-genome array-CGH identifies novel contiguous gene deletions and duplications associated with developmental delay, mental retardation, and dysmorphic features. Am J Med Genet A 2007; 143(13):1431-1441.

Astbury C, Christ LA, Aughton DJ, Cassidy SB, Kumar A, Eichler EE *et al.* Detection of deletions in *de novo* "balanced" chromosome rearrangements: further evidence for their role in phenotypic abnormalities. Genet Med 2004; 6:81-89.

Bache I, Hjorth M, Bugge M, Holstebroe S, Hilden J, Schmidt L *et al.* 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 2006; 14(4):410-417.

Bandyopadhyay R, Heller A, Knox-DuBois C, McCaskill C, Berend SA, Page SL *et al.* Parental origin and timing of *de novo* Robertsonian translocation formation. Am J Hum Genet 2002; 71(6):1456-1462.

Baptista J, Prigmore E, Gribble SM, Jacobs PA, Carter NP, Crolla JA. Molecular cytogenetic analyses of breakpoints in apparently balanced reciprocal translocations carried by phenotypically normal individuals. Eur J Hum Genet 2005; 13(11):1205-1212.

Barber JC. Directly transmitted unbalanced chromosome abnormalities and euchromatic variants. J Med Genet 2005; 42(8):609-629.

Bione S, Sala C, Manzini C, Arrigo G, Zuffardi O, Banfi S *et al.* A human homologue of the Drosophila melanogaster diaphanous gene is disrupted in a patient with premature ovarian failure: evidence for conserved function in oogenesis and implications for human sterility. Am J Hum Genet 1998; 62(3):533-541.

Borg I, Squire M, Menzel C, Stout K, Morgan D, Willatt L *et al.* A cryptic deletion of 2q35 including part of the PAX3 gene detected by breakpoint mapping in a child with autism and a *de novo* 2;8 translocation. J Med Genet 2002; 39(6):391-399.

Borg I, Freude K, Kubart S, Hoffmann K, Menzel C, Laccone F *et al.* Disruption of Netrin G1 by a balanced chromosome translocation in a girl with Rett syndrome. Eur J Hum Genet 2005; 13(8):921-927.

Brewer CM, Leek JP, Green AJ, Holloway S, Bonthron DT, Markham AF, et al. A locus for isolated cleft palate, located on human chromosome 2q32. Am J Hum Genet 1999 Aug;65(2):387-96.

Bugge M, Bruun-Petersen G, Brondum-Nielsen K, Friedrich U, Hansen J, Jensen G *et al.* Disease associated balanced chromosome rearrangements: a resource for large scale genotype-phenotype delineation in man. J Med Genet 2000; 37(11):858-865.

Buhler EM. Unmasking of heterozygosity by inherited balanced translocations. Implications for prenatal diagnosis and gene mapping. Ann Genet 1983; 26(3):133-137.

Burgoyne PS, Baker TG. Meiotic pairing and gametogenic failure. Symp Soc Exp Biol 1984;38:349-62.:349-62.

Burke DT, Carle GF, Olson MV. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 1987; 236(4803):806-812.

Carter NP, Ferguson-Smith MA, Perryman MT, Telenius H, Pelmear AH, Leversha MA *et al.* Reverse chromosome painting: a method for the rapid analysis of aberrant chromosomes in clinical cytogenetics. J Med Genet 1992; 29(5):299-307.

Carter NP. Bivariate chromosome analysis using a commercial flow cytometer. Methods Mol Biol 1994; 29:187-204.:187-204.

Cason AL, Ikeguchi Y, Skinner C, Wood TC, Holden KR, Lubs HA *et al.* X-linked spermine synthase gene (SMS) defect: the first polyamine deficiency syndrome. Eur J Hum Genet 2003; 11(12):937-944.

Caspersson T, Zech L, Johansson C. Differential binding of alkylating fluorochromes in human chromosomes. Exp Cell Res 1970; 60(3):315-319.

Chandley AC. On the parental origin of *de novo* mutation in man. J Med Genet 1991 Apr;28(4):217-23.

Ciccone R, Giorda R, Gregato G, Guerrini R, Giglio S, Carrozzo R *et al.* Reciprocal translocations: a trap for cytogenetists? Hum Genet 2005; 117(6):571-582.

Cirulli V, Yebra M. Netrins: beyond the brain. Nat Rev Mol Cell Biol 2007; 8(4):296-306.

Collins FS. Positional cloning: let's not call it reverse anymore. Nat Genet 1992; 1(1):3-6.

Collins J, Hohn B. Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. Proc Natl Acad Sci U S A 1978; 75(9):4242-4246.

Colon-Ramos DA, Margeta MA, Shen K. Glia promote local synaptogenesis through UNC-6 (netrin) signaling in C. elegans. Science 2007 Oct 5;318(5847):103-6.

Costes B, Girodon E, Ghanem N, Flori E, Jardin A, Soufir JC *et al.* Frequent occurrence of the CFTR intron 8 (TG)n 5T allele in men with congenital bilateral absence of the vas deferens. Eur J Hum Genet 1995; 3(5):285-293.

Coux O, Tanaka K, Goldberg AL. Structure and functions of the 20S and 26S proteasomes. Annu Rev Biochem 1996; 65:801-847.

Cox JJ, Holden ST, Dee S, Burbridge JI, Raymond FL. Identification of a 650 kb duplication at the X chromosome breakpoint in a patient with 46,X,t(X;8)(q28;q12) and non-syndromic mental retardation. J Med Genet 2003; 40(3):169-174.

de Gregori M, Ciccone R, Magini P, Pramparo T, Gimelli S, Messa J *et al.* Cryptic deletions are a common finding in "balanced" reciprocal and complex chromosome rearrangements: a study of 59 patients. J Med Genet 2007; 44(12):750-762.

de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM *et al.* Diagnostic genome profiling in mental retardation. Am J Hum Genet 2005; 77(4):606-616.

Debacker K, Kooy RF. Fragile sites and human disease. Hum Mol Genet 2007; 16 Spec No. 2:R150-R158.

Edwards JH, Harnden DG, Cameron AH, Crosse VM, Wolff OH. A new trisomic syndrome. Lancet 1960; 1:787-790.

Eichler EE. Masquerading repeats: paralogous pitfalls of the human genome. Genome Res 1998; 8(8):758-762.

Eichler EE. Recent duplication, domain accretion and the dynamic mutation of the human genome. Trends Genet 2001; 17(11):661-669.

Engel E. A new genetic concept: uniparental disomy and its potential effect, isodisomy. Am J Med Genet 1980; 6(2):137-143.

Fantes J, Redeker B, Breen M, Boyle S, Brown J, Fletcher J *et al.* Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Hum Mol Genet 1995; 4(3):415-422.

Fantes J, Ragge NK, Lynch SA, McGill NI, Collin JR, Howard-Peebles PN *et al.* Mutations in SOX2 cause anophthalmia. Nat Genet 2003; 33(4):461-463.

Feng Y, Crosbie J, Wigg K, Pathare T, Ickowicz A, Schachar R *et al.* The SNAP25 gene as a susceptibility gene contributing to attention-deficit hyperactivity disorder. Mol Psychiatry 2005; 10(11):998-1005, 973.

Feuk L, MacDonald JR, Tang T, Carson AR, Li M, Rao G *et al.* Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. PLoS Genet 2005; 1(4):e56.

Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 2006; 7(2):85-97.

Fiegler H, Gribble SM, Burford DC, Carr P, Prigmore E, Porter KM *et al.* Array painting: a method for the rapid analysis of aberrant chromosomes using DNA microarrays. J Med Genet 2003a; 40(9):664-670.

Fiegler H, Carr P, Douglas EJ, Burford DC, Hunt S, Scott CE *et al.* DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. Genes Chromosomes Cancer 2003b; 36(4):361-374.

Fiegler H, Redon R, Andrews D, Scott C, Andrews R, Carder C *et al.* Accurate and reliable high-throughput detection of copy number variation in the human genome. Genome Res 2006; 16(12):1566-1574.

FitzPatrick DR, Carr IM, McLaren L, Leek JP, Wightman P, Williamson K *et al.* Identification of SATB2 as the cleft palate gene on 2q32-q33. Hum Mol Genet 2003; 12(19):2491-2501.

FitzPatrick DR, Fantes JA, Boland E, Ramsay J, Donnai D, Clayton-Smith J, *et al.* Molecular cytogenetic analysis of 149 breakpoints in 76 cases with apparently balanced chromosome rearrangements. [Abstract]. American Society of Human Genetics Annual Meeting 2006.

Floridia G, Piantanida M, Minelli A, Dellavecchia C, Bonaglia C, Rossi E *et al.* The same molecular mechanism at the maternal meiosis I produces mono- and dicentric 8p duplications. Am J Hum Genet 1996; 58(4):785-796.

Ford CE, Hamerton JL. The chromosomes of man. Nature 1956; 178(4541):1020-1023.

Ford CE, Jones KW, Polani PE, De Almeida JC, Briggs JH. A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). Lancet 1959; 1(7075):711-713.

Foster RE, Abdulrahman M, Morris MR, Prigmore E, Gribble S, Ng B *et al.* Characterization of a 3;6 translocation associated with renal cell carcinoma. Genes Chromosomes Cancer 2007; 46(4):311-317.

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM *et al.* Copy number variation: new insights in genome diversity. Genome Res 2006; 16(8):949-961.

Friedman JM, Baross A, Delaney AD, Ally A, Arbour L, Armstrong L *et al.* Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. Am J Hum Genet 2006; 79(3):500-513.

Gajecka M, Glotzbach CD, Jarmuz M, Ballif BC, Shaffer LG. Identification of cryptic imbalance in phenotypically normal and abnormal translocation carriers. Eur J Hum Genet 2006; 14(12):1255-1262.

Gall JG, Pardue ML. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proc Natl Acad Sci U S A 1969; 63(2):378-383.

Gibson KM, Sweetman L, Nyhan WL, Jansen I. Demonstration of 4-aminobutyric acid aminotransferase deficiency in lymphocytes and lymphoblasts. J Inherit Metab Dis 1985; 8(4):204-208.

Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T *et al.* Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am J Hum Genet 2001; 68(4):874-883.

Greenstein RM, Reardon MP, Chan TS. An X-autosome translocation in a girl with Duchenne muscular dystrophy (DMD): evidence for DMD gene localization. [Abstract] Pediat. Res. 1977;11: 457

Gribble SM, Prigmore E, Burford DC, Porter KM, Ng BL, Douglas EJ *et al.* The complex nature of constitutional *de novo* apparently balanced translocations in patients presenting with abnormal phenotypes. J Med Genet 2005; 42(1):8-16.

Gribble SM, Kalaitzopoulos D, Burford DC, Prigmore E, Selzer RR, Ng BL *et al.* Ultra-high resolution array painting facilitates breakpoint sequencing. J Med Genet 2007; 44(1):51-58.

Gustavson K.-H, Anneren G, Jagell, S.Prader-Willi syndrome in a child with a balanced (X;15) *de novo* translocation. Clin. Genet. 1984;26: 245-247.

Hamerton JL. Human cytogenetics.New York. Academic Press, 1971;pp 232-289.

Hassold, T. Chromosome abnormalities in human reproductive wastage. Trends Genet. 1986;2: 105-110.

Hearn T, Renforth GL, Spalluto C, Hanley NA, Piper K, Brickwood S *et al.* Mutation of ALMS1, a large gene with a tandem repeat encoding 47 amino acids, causes Alstrom syndrome. Nat Genet 2002; 31(1):79-83.

Hollox EJ, Armour JA, Barber JC. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. Am J Hum Genet 2003; 73(3):591-600.

Hook EB, Hamerton JL.The frequency of chromosome abnormalities detected in consecutive newborn studies-differences between studies-results by sex and by severity of phenotypic involvement. In: Hook EB, Porter IH, eds. Population cytogenetics. New York State Department of Health, Birth Defects Institute, Symposium (1975), Albany, New York, October 1975. New York: Academic Press, 1977:63.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y *et al.* Detection of large-scale variation in the human genome. Nat Genet 2004; 36(9):949-951.

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C *et al.* A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat Genet 1994; 6(1):84-89.

Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex-determining mechanism. Nature 1959; 183(4657):302-303.

Jacobs PA, Baikie AG, Court BrownWM, Strong JA. The somatic chromosomes in mongolism. Lancet 1959; 1(7075):710.

Jacobs PA. Correlation between euploid structural chromosome rearrangements and mental subnormality in humans. Nature 1974; 249(453):164-165.

Jacobs PA, Frackiewicz A, Law P, Hilditch CJ, Morton NE. The effect of structural aberrations of the chromosomes on reproductive fitness in man. II. Results. Clin Genet 1975; 8(3):169-178.

Jacobs PA, Hunt PA, Mayer M, Bart RD. Duchenne muscular dystrophy (DMD) in a female with an X/autosome translocation: further evidence that the DMD locus is at Xp21. Am J Hum Genet 1981; 33(4):513-518.

Jacobs PA, Browne C, Gregson N, Joyce C, White H. Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. J Med Genet 1992; 29:103-108.

Jaeken J, Casaer P, de Cock P, Corbeel L, Eeckels R, Eggermont E *et al.* Gamma-aminobutyric acid-transaminase deficiency: a newly recognized inborn error of neurotransmitter metabolism. Neuropediatrics 1984; 15(3):165-169.

James RS, Temple IK, Patch C, Thompson EM, Hassold T, Jacobs PA. A systematic search for uniparental disomy in carriers of chromosome translocations. Eur J Hum Genet 1994; 2(2):83-95.

John HA, Birnstiel ML, Jones KW. RNA-DNA hybrids at the cytological level. Nature 1969; 223(206):582-587.

Johnson D, Morrison N, Grant L, Turner T, Fantes J, Connor JM *et al.* Confirmation of CHD7 as a cause of CHARGE association identified by mapping a balanced chromosome translocation in affected monozygotic twins. J Med Genet 2006; 43(3):280-284.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F *et al.* Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 1992; 258(5083):818-821.

Kamei Y, Takeda Y, Teramoto K, Tsutsumi O, Taketani Y, Watanabe K. Human NB-2 of the contactin subgroup molecules: chromosomal localization of

the gene (CNTN5) and distinct expression pattern from other subgroup members. Genomics 2000; 69(1):113-119.

Kamnasaran D, Muir WJ, Ferguson-Smith MA, Cox DW. Disruption of the neuronal PAS3 gene in a family affected with schizophrenia. J Med Genet 2003a; 40(5):325-332.

Kamnasaran D, O'Brien PC, Zackai EH, Muenke M, Ferguson-Smith MA, Cox DW. Rearrangement in the PITX2 and MIPOL1 genes in a patient with a t(4;14) chromosome. Eur J Hum Genet 2003b; 11(4):315-324.

Karpen GH. Position-effect variegation and the new biology of heterochromatin. Curr Opin Genet Dev 1994; 4(2):281-291.

Kim UJ, Shizuya H, de Jong PJ, Birren B, Simon MI. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Res 1992; 20(5):1083-1085.

Kim HG, Herrick SR, Lemyre E, Kishikawa S, Salisz JA, Seminara S *et al.* Hypogonadotropic hypogonadism and cleft lip and palate caused by a balanced translocation producing haploinsufficiency for FGFR1. J Med Genet 2005; 42(8):666-672.

Kleefstra T, Yntema HG, Oudakker AR, Banning MJ, Kalscheuer VM, Chelly J *et al.* Zinc finger 81 (ZNF81) mutations associated with X-linked mental retardation. J Med Genet 2004; 41(5):394-399.

Kleefstra T, Smidt M, Banning MJ, Oudakker AR, Van Esch H, de Brouwer AP *et al.* Disruption of the gene Euchromatin Histone Methyl Transferase1 (Eu-HMTase1) is associated with the 9q34 subtelomeric deletion syndrome. J Med Genet 2005; 42(4):299-306.

Kleinjan DA, van H, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 2005; 76(1):8-32.

Krepischi-Santos AC, Vianna-Morgante AM, Jehee FS, Passos-Bueno MR, Knijnenburg J, Szuhai K *et al.* Whole-genome array-CGH screening in undiagnosed syndromic patients: old syndromes revisited and new alterations. Cytogenet Genome Res 2006; 115(3-4):254-261.

Kumar A, Becker LA, Depinet TW, Haren JM, Kurtz CL, Robin NH *et al.* Molecular characterization and delineation of subtle deletions in *de novo* "balanced" chromosomal rearrangements. Hum Genet 1998; 103(2):173-178.

Kurahashi H, Emanuel BS. Unexpectedly high rate of *de novo* constitutional t(11;22) translocations in sperm from normal males. Nat Genet 2001; 29(2):139-140.

Landegent JE, Jansen in dW, van Ommen GJ, Baas F, de Vijlder JJ, van Duijn P *et al.* Chromosomal localization of a unique gene by non-autoradiographic in situ hybridization. Nature 1985; 317(6033):175-177.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J *et al.* Initial sequencing and analysis of the human genome. Nature 2001; 409(6822):860-921.

Langer-Safer PR, Levine M, Ward DC. Immunological method for mapping genes on Drosophila polytene chromosomes. Proc Natl Acad Sci U S A 1982; 79(14):4381-4385.

Langlois RG, Yu LC, Gray JW, Carrano AV. Quantitative karyotyping of human chromosomes by dual beam flow cytometry. Proc Natl Acad Sci U S A 1982; 79(24):7876-7880.

Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 2007; 131(7):1235-1247.

Lejeune J, Gautier M, Turpin R. [Study of somatic chromosomes from 9 mongoloid children.]. C R Hebd Seances Acad Sci 1959; 248(11):1721-1722.

Lejeune J, Lafourcade J, Berger R, Vialatte J, Boeswillwald M, Seringe P *et al.* [3 Cases of partial deletion of the short arm of a 5 chromosome]. C R Hebd Seances Acad Sci 1963; 257:3098-3102.

Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, Hartemink AJ. Computational and experimental identification of novel human imprinted genes. Genome Res 2007; 17(12):1723-1730.

Lupski JR, Stankiewicz P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet 2005; 1(6):e49.

Machin GA, Crolla JA. Chromosome constitution of 500 infants dying during the perinatal period. With an appendix concerning other genetic disorders among these infants. Humangenetik 1974; 23(3):183-198.

Madan K. Balanced structural changes involving the human X: effect on sexual phenotype. Hum Genet 1983; 63(3):216-221.

Madan K, Nieuwint AW, van Bever Y. Recombination in a balanced complex translocation of a mother leading to a balanced reciprocal translocation in the child. Review of 60 cases of balanced complex translocations. Hum Genet 1997; 99(6):806-815.

Manolov G, Manolova Y, Klein G, Levan A, Kieler J. Chromosome #14 markers in two Epstein-Barr virus (EBV)-transformed lymphoblastoid cell lines of normal origin differ from the Burkitt lymphoma (BL)-associated 14q+ marker. Cancer Genet Cytogenet 1981; 4(2):179-184.

Mansouri MR, Marklund L, Gustavsson P, Davey E, Carlsson B, Larsson C *et al.* Loss of ZDHHC15 expression in a woman with a balanced translocation t(X;15)(q13.3;cen) and severe mental retardation. Eur J Hum Genet 2005; 13(8):970-977.

Mansouri MR, Carlsson B, Davey E, Nordenskjold A, Wester T, Anneren G *et al.* Molecular genetic analysis of a *de novo* balanced translocation t(6;17)(p21.31;q11.2) associated with hypospadias and anorectal malformation. Hum Genet 2006; 119(1-2):162-168.

Mattei MG, Mattei JF, Ayme S, Giraud F. X-autosome translocations: cytogenetic characteristics and their consequences. Hum Genet 1982; 61(4):295-309.

McMullan TW, Crolla JA, Gregory SG, Carter NP, Cooper RA, Howell GR *et al.* A candidate gene for congenital bilateral isolated ptosis identified by molecular analysis of a *de novo* balanced translocation. Hum Genet 2002; 110(3):244-250.

Medina-Kauwe LK, Tobin AJ, De Meirleir L, Jaeken J, Jakobs C, Nyhan WL *et al.* 4-Aminobutyrate aminotransferase (GABA-transaminase) deficiency. J Inherit Metab Dis 1999; 22(4):414-427.

Menten B, Maas N, Thienpont B, Buysse K, Vandesompele J, Melotte C *et al.* Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports. J Med Genet 2006; 43(8):625-633.

Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988; 16(3):1215.

Monaco AP, Neve RL, Colletti-Feener C, Bertelson CJ, Kurnit DM, Kunkel LM. Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature 1986; 323(6089):646-650.

Nothwang HG, Kim HG, Aoki J, Geisterfer M, Kubart S, Wegner RD *et al.* Functional hemizygosity of PAFAH1B3 due to a PAFAH1B3-CLK2 fusion gene in a female with mental retardation, ataxia and atrophy of the brain. Hum Mol Genet 2001; 10(8):797-806.

Nowell P, Hungerford D. A minute chromosome in human chronic granulocytic leukemia [abstract]. Science. 1960;132:1497.

Odorisio T, Rodriguez TA, Evans EP, Clarke AR, Burgoyne PS. The meiotic checkpoint monitoring synapsis eliminates spermatocytes via p53-independent apoptosis. Nat Genet 1998; 18(3):257-261.

Page SL, Shaffer LG. Nonhomologous Robertsonian translocations form predominantly during female meiosis. Nat Genet 1997; 15(3):231-232.

Patau K, Smith DW, Therman E, Inhorn SL, Wagner HP. Multiple congenital anomaly caused by an extra autosome. Lancet 1960; 1:790-3.:790-793.

Patsalis PC, Evangelidou P, Charalambous S, Sismani C. Fluorescence in situ hybridization characterization of apparently balanced translocation reveals

cryptic complex chromosomal rearrangements with unexpected level of complexity. Eur J Hum Genet 2004; 12(8):647-653.

Patsalis PC. Complex chromosomal rearrangements. Genet Couns 2007; 18(1):57-69.

Petit FG, Jamin SP, Kurihara I, Behringer RR, DeMayo FJ, Tsai MJ *et al.* Deletion of the orphan nuclear receptor COUP-TFII in uterus leads to placental deficiency. Proc Natl Acad Sci U S A 2007; 104(15):6293-6298.

Pickard BS, Malloy MP, Porteous DJ, Blackwood DH, Muir WJ. Disruption of a brain transcription factor, NPAS3, is associated with schizophrenia and learning disability. Am J Med Genet B Neuropsychiatr Genet 2005; 136(1):26-32.

Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J *et al.* Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc Natl Acad Sci U S A 1988; 85(23):9138-9142.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998; 20(2):207-211.

Ramocki MB, Dowling J, Grinberg I, Kimonis VE, Cardoso C, Gross A *et al.* Reciprocal fusion transcripts of two novel Zn-finger genes in a female with absence of the corpus callosum, ocular colobomas and a balanced translocation between chromosomes 2p24 and 9q32. Eur J Hum Genet 2003; 11(7):527-534.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD *et al.* Global variation in copy number in the human genome. Nature 2006; 444(7118):444-454.

Rickman L, Fiegler H, Carter NP, Bobrow M. Prenatal diagnosis by array-CGH. Eur J Med Genet 2005; 48(3):232-240.

Rizzolio F, Sala C, Alboresi S, Bione S, Gilli S, Goegan M *et al.* Epigenetic control of the critical region for premature ovarian failure on autosomal genes translocated to the X chromosome: a hypothesis. Hum Genet 2007; 121(3-4):441-450.

Rosenberg C, Knijnenburg J, Bakker E, Vianna-Morgante AM, Sloos W, Otto PA *et al.* Array-CGH detection of micro rearrangements in mentally retarded individuals: clinical significance of imbalances present both in affected children and normal parents. J Med Genet 2006; 43(2):180-186.

Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature 1973; 243(5405):290-293.

Saito A, Watanabe TK, Shimada Y, Fujiwara T, Slaughter CA, DeMartino GN *et al.* cDNA cloning and functional analysis of p44.5 and p55, two regulatory subunits of the 26S proteasome. Gene 1997; 203(2):241-250.

Saito A, Fujikura-Ouchi Y, Kuramasu A, Shimoda K, Akiyama K, Matsuoka H, *et al.* Association study of putative promoter polymorphisms in the neuroplastin gene and schizophrenia. Neurosci Lett 2007 Jan 16;411(3):168-73.

Sakharkar MK, Chow VT, Kangueane P. Distributions of exons and introns in the human genome. In Silico Biol 2004; 4(4):387-393.

Sala C, Arrigo G, Torri G, Martinazzi F, Riva P, Larizza L *et al.* Eleven X chromosome breakpoints associated with premature ovarian failure (POF) map to a 15-Mb YAC contig spanning Xq21. Genomics 1997; 40(1):123-131.

Sambrook, J., Fritsch, E.F., and Maniatis, T.; Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.

Sanyanusin P, Schimmenti LA, McNoe LA, Ward TA, Pierpont ME, Sullivan MJ, *et al.* Mutation of the PAX2 gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. Nat Genet 1995 Apr;9(4):358-64.

Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP *et al.* Challenges and standards in integrating surveys of structural variation. Nat Genet 2007; 39(7 Suppl):S7-15.

Schlessinger D, Herrera L, Crisponi L, Mumm S, Percesepe A, Pellegrini M *et al.* Genes and translocations involved in POF. Am J Med Genet 2002; 111(3):328-333.

Schmidt M, Du SD. Functional disomies of the X chromosome influence the cell selection and hence the X inactivation pattern in females with balanced X-autosome translocations: a review of 122 cases. Am J Med Genet 1992; 42(2):161-169.

Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM, Nordenskjold M. Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH). J Med Genet 2005; 42(9):699-705.

Seabright M. A rapid banding technique for human chromosomes. Lancet 1971; 2(7731):971-972.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P *et al.* Large-scale copy number polymorphism in the human genome. Science 2004; 305(5683):525-528.

Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU *et al.* Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 2005; 77(1):78-88.

Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H *et al.* Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. J Med Genet 2004; 41(4):241-248.

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc Natl Acad Sci U S A 1992; 89(18):8794-8797.

Shoichet SA, Duprez L, Hagens O, Waetzig V, Menzel C, Herdegen T *et al.* Truncation of the CNS-expressed JNK3 in a patient with a severe developmental epileptic encephalopathy. Hum Genet 2006; 118(5):559-567.

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H *et al.* Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer 1997; 20(4):399-407.

Stankiewicz P, Beaud*et al.* Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. Curr Opin Genet Dev 2007; 17(3):182-192.

Steel CM, Morten JE, Foster E. The cytogenetics of human B lymphoid malignancy: studies in Burkitt's lymphoma and Epstein-Barr virus-transformed lymphoblastoid cell lines. IARC Sci Publ 1985;(60):265-292.

Taudien S, Galgoczy P, Huse K, Reichwald K, Schilhabel M, Szafranski K *et al.* Polymorphic segmental duplications at 8p23.1 challenge the determination of individual defensin gene repertoires and the assembly of a contiguous human reference sequence. BMC Genomics 2004; 5(1):92.

Therman E, Laxova R, Susman B. The critical region on the human Xq. Hum Genet 1990; 85(5):455-461.

Thomas NS, Durkie M, Van ZB, Sanford R, Potts G, Youings S *et al.* Parental and chromosomal origin of unbalanced *de novo* structural chromosome abnormalities in man. Hum Genet 2006; 119(4):444-450.

Tjio JH, Levan A. The chromosome number in man. Hereditas 1956; 42:1-6.

Tommerup N. Mendelian cytogenetics. Chromosome rearrangements associated with mendelian disorders. J Med Genet 1993; 30(9):713-727.

Turjanski AG, Vaque JP, Gutkind JS. MAP kinases and the control of nuclear events. Oncogene 2007; 26(22):3240-3253.

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM *et al.* Fine-scale structural variation of the human genome. Nat Genet 2005; 37(7):727-732.

Tyson C, Harvard C, Locker R, Friedman JM, Langlois S, Lewis ME *et al.* Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH. Am J Med Genet A 2005; 139(3):173-185.

Urrutia R. KRAB-containing zinc-finger repressor proteins. Genome Biol 2003; 4(10):231.

van Assche E, Bonduelle M, Tournaye H, Joris H, Verheyen G, Devroey P *et al.* Cytogenetics of infertile men. Hum Reprod 1996; 11 Suppl 4:1-24.

van Buggenhout G, Ravenswaaij-Arts C, Mc MN, Thoelen R, Vogels A, Smeets D *et al.* The del(2)(q32.2q33) deletion syndrome defined by clinical and molecular characterization of four patients. Eur J Med Genet 2005; 48(3):276-289.

van Prooijen-Knegt AC, Van Hoek JF, Bauman JG, van Duijn P, Wool IG, Van der PM. In situ hybridization of DNA sequences in human metaphase chromosomes visualized by an indirect fluorescent immunocytochemical procedure. Exp Cell Res 1982; 141(2):397-407.

Velagaleti GV, Bien-Willner GA, Northup JK, Lockhart LH, Hawkins JC, Jalal SM *et al.* Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. Am J Hum Genet 2005; 76(4):652-662.

Vervoort VS, Viljoen D, Smart R, Suthers G, DuPont BR, Abbott A *et al.* Sorting nexin 3 (SNX3) is disrupted in a patient with a translocation t(6;13)(q21;q12) and microcephaly, microphthalmia, ectrodactyly, prognathism (MMEP) phenotype. J Med Genet 2002a; 39(12):893-899.

Vervoort VS, Beachem MA, Edwards PS, Ladd S, Miller KE, de M, X *et al.* AGTR2 mutations in X-linked mental retardation. Science 2002b; 296(5577):2401-2403.

Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO *et al.* Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. Am J Hum Genet 2003; 73(6):1261-1270.

Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM *et al.* Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. Nat Genet 2004; 36(9):955-957.

Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, *et al.* Genetic alterations during colorectal-tumor development. N Engl J Med 1988 Sep 1;319(9):525-32.

Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F *et al.* Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Hum Mol Genet 1996; 5(7):933-943.

Walls EV, Crawford DH. Generation of B lymphoblastoid cell lines using Epstein-Barr virus. In Lymphocytes: a Practical Approach. Edited by Klauss GGB. Oxford, Washington DC: IRL Press; 1987: 149 -162.

Warburton D. *De novo* balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. Am J Hum Genet 1991; 49(5):995-1013.

Webber LM, Garson OM. Fluorodeoxyuridine synchronization of bone marrow cultures. Cancer Genet Cytogenet 1983; 8(2):123-132.

Wirth J, Nothwang HG, vanderMaarel S, Menzel C, Borck G, Lopezpajares I *et al.* 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 1999; 36:271-279.

Woodard C, Alcorta E, Carlson J. The rdgB gene of Drosophila: a link between vision and olfaction. J Neurogenet 1992; 8(1):17-31.

Youings S, Ellis K, Ennis S, Barber J, Jacobs P. A study of reciprocal translocations and inversions detected by light microscopy with special reference to origin, segregation, and recurrent abnormalities. Am J Med Genet A 2004; 126(1):46-60.

Yue Y, Grossmann B, Holder SE, Haaf T. *De novo* t(7;10)(q33;q23) translocation and closely juxtaposed microdeletion in a patient with macrocephaly and developmental delay. Hum Genet 2005; 117(1):1-8.

Zemni R, Bienvenu T, Vinet MC, Sefiani A, Carrie A, Billuart P *et al.* A new gene involved in X-linked mental retardation identified by analysis of an X;2 balanced translocation. Nat Genet 2000; 24(2):167-170.

Zhao S, Stodolsky M (Editors). Bacterial artificial chromosomes. Totowa New Jersey, Humana Press, 2004.