"Molecular characterization of a pericentric inversion of chromosome 3 in a 3-generation family with short stature"

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

Naturwissenschaftlichen Fakultät I
Biowissenschaften

der Martin-Luther-Universität Halle-Wittenberg

von

Frau Dutta Usha Rani

geb. am: 17.9.72 in Chirala (India)

Gutachter

1. Prof. Ingo Hansmann

2. Prof. G. Reuter

3. Prof. Broinitz

Halle (Saale), den 04.03.2008

Contents

Abbreviations

BAC	Bacterial Artificial Chromosome
BAAD	Biotinylated anti-Avidin D
BCIP	5-Bromo-4-chloro-3-indolyl-phosphate
bp	Base pair
BSA	Bovine serum albumin
cDNA	Complementary deoxyribonucleic acid
dATP	Deoxyadenosine triphosphate
DAPI	$4^{'}$, 6-Diamidino-2-phenylindole
dCTP	Deoxycytidine triphosphate
ddNTPs	Dideoxynucleotide triphosphates
dGTP	Deoxyguanosine triphosphate
ddH$_2$O	Double distilled water
DIG	Digoxigenin
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
dUTP	Deoxyuridine triphosphate
EDTA	Ethylenediamino-tetraacetic acid
EST	Expressed sequence tags
Fig	Figure
FITC	Fluorescein Avidin D
FISH	Fluorescence *in situ* hybridization
HCl	Hydrochloric acid
hr	hour
IMAGE	International molecular analysis of genomes and their expression (cDNA clones)

Inv	Inversion
IPTG	Isopropyl β-D-thiogalactopyranoside
Kb	Kilo-base pair
LB	Luria Bertani
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
M	Molar
Mb	Megabase
min	Minute
µg	Microgram
ml	Milliliter
mM	Millimolar
mRNA	Messenger ribonucleic acid
NBT	4-Nitro blue tetrazolium chloride
nm	Nanometer
OD_{260}	Optical density at 260 nm
OD_{280}	Optical density at 280 nm
OD_{560}	Optical density at 560 nm
Oligo	Oligonucleotide
O/N	over night
PBS	Phosphored buffered saline
PCR	Polymerase chain reaction
PI	Propidium iodide
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions per minute
RT	room temperature
SDS	Sodium dodecyl sulphate
sec	seconds
SINE	Short interspersed nuclear element
SSC	Saline sodium citrate

Tab	Table
Taq	*Thermus aquaticus*
TAE	Tris Acetate EDTA buffer
TBE	Tris Borate EDTA buffer
TE	Tris EDTA buffer
Tris	Tris (hydroxymethyl) amino methane
Tween 20	Polyethylenglycol-sorbitan-monolaurate
U	units (enzymatic)
Vol	Volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
YAC	Yeast artificial chromosome

1   Introduction

1.1   Structural abnormalities

Chromosomal abnormalities and gene mutations are some of the major causes of human sufferings. Structural changes in chromosomes can be classified according to cytological types and their effect on the phenotype. The main cytological types are translocations, deletions, inversions, insertions, isochromosomes, dicentric chromosomes and ring chromosomes (Fig. 1). Structural rearrangements alter the genomic architecture and may result in human disease traits. These alterations can be termed as Genomic disorders. Genomic disorders can lead to the complete loss or gain of a gene(s) sensitive to a dosage effect, or can disrupt the structural integrity of a gene (Lupski, 1998). Genome alterations can occur through many mechanisms, one of which is the nonallelic homologous recombination during meiosis between region specific, low-copy repeated sequences (Weinstock et al., 1998). Recombination between direct repeats leads to deletions and duplications, while between inverted repeats results in inversion of the intervening genomic sequences.

1.1.1   Chromosomal rearrangements as tools for disease gene identification

Balanced chromosomal rearrangements, which truncate, delete or otherwise inactivate genes, are powerful tools for mapping novel disease genes (Tommerup et al., 1993; Collins, 1995; Bugge et al., 2000). The cytogenetic and molecular analyses of such rearrangements constitute an efficient strategy for mapping and cloning these disease genes. There are many examples of human diseases in which analyses of chromosomal anomalies have led to the discovery of diseases and further identification of genes involved in the etiology of the disease. A few among those are X-linked ectodermal dysplasia (Srivastava et al., 1996), Aniridia (Gessler et al., 1989) and holoprosencephaly (Belloni et al., 1996).
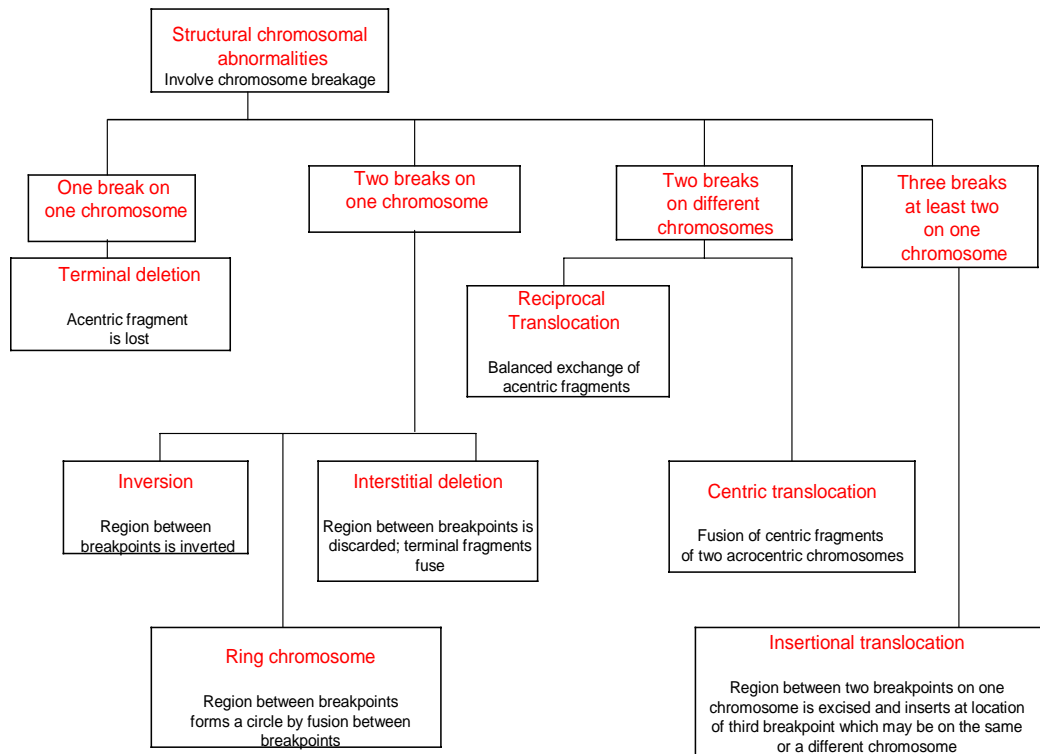
Fig. 1: Major types of structural chromosomal abnormalities (adapted from Strachan and Read, 1996).

In majority of the cases a single gene is expected to be disrupted either in its structure or expression, thus leading to a disease phenotype. An autosomal dominant disorder can be caused either by a gain-of-function mutation on one of the alleles or by the loss-of-function on one homologue (haploinsufficiency). Chromosomal rearrangements associated with an autosomal dominant disorder are usually expected to cause haploinsufficiency, since the disruption of the coding region of a gene may lead to a functional-null-allele (Vortkamp et al., 1991; Fahsold et al., 1995). In most of the cases chromosomal aberrations occur sporadically (*de novo*) and do not segregate within a family. Such cases are simple in interpreting the correlation of the patient phenotype with the chromosomal aberration. In rare cases chromosomal aberrations segregate within a family with a specific clinical phenotype. Such rare families with chromosomal rearrangements segregating with an "autosomal dominant" trait are highly significant for identifying the disease causing gene(s). In most such cases one of the chromosomal

breakpoints map within or close to the respective gene thus disturbing its gene activity. If a gene is located in the breakpoint region that is disrupted then it can be considered as a candidate gene for the disease observed in the patient. Mutation screening of this candidate gene can be done in patients presenting with the same disorder and having a normal karyotype. The identification of mutations in this gene will thus establish a causal relation between the mutation and that particular disease. For example Tsukamoto et al. (1992) cloned and characterized an inversion breakpoint on chromosome 2q35 with Waardenburg syndrome type I. The HuP2 gene is found to be disrupted by the inversion and thus suggested that this gene is a candidate for the Waardenburg syndrome type I. Hence, the identification of disease causing genes is a prerequisite for proper diagnosis and genetic counselling of patients and their families as well as for studying the molecular pathogenesis leading to the disease.

1.1.2   Mechanism of inversion

Two breaks in a single chromosome can lead to an inversion, or an interstitial deletion or else a ring chromosome. An inversion results when the chromosomal segment between the two breakpoints takes a 180 degree turn prior to resealing the breakpoints (Fig. 2). Usually, this is a balanced rearrangement with no loss of chromosomal material. The inverted segment may include the centromere (pericentric inversion) or may be confined to one chromosome arm (paracentric inversion).
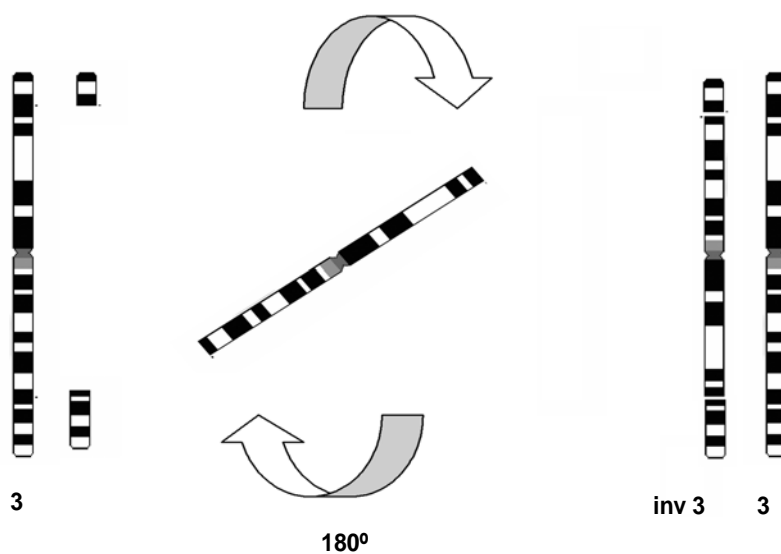
Fig. 2: Schematic representation showing the mechanism of inversion.

1.2   The Human Genome Project

The Human Genome Project (HGP) is an international project conceived in 1986, which officially began in 1990 and was completed in 2003. The complete nucleotide sequence of the human genome was one of the several goals of the HGP. The major goal then was to acquire the fundamental information concerning our genetic make-up. This is crucial for basic scientific understanding of human genetics and the role of various genes in health and disease. The detail interaction with research focussing on mapping disease genes is shown in Figure 3. Knowledge of the human genome sequence enables us to understand how the genetic information determines the development, the structure and the function of the human body. It helps in exploring how variations in our human genome can cause disease, how they affect our interaction with our environment, and more importantly how to develop new and effective ways to improve human health (Bentley, 2000).
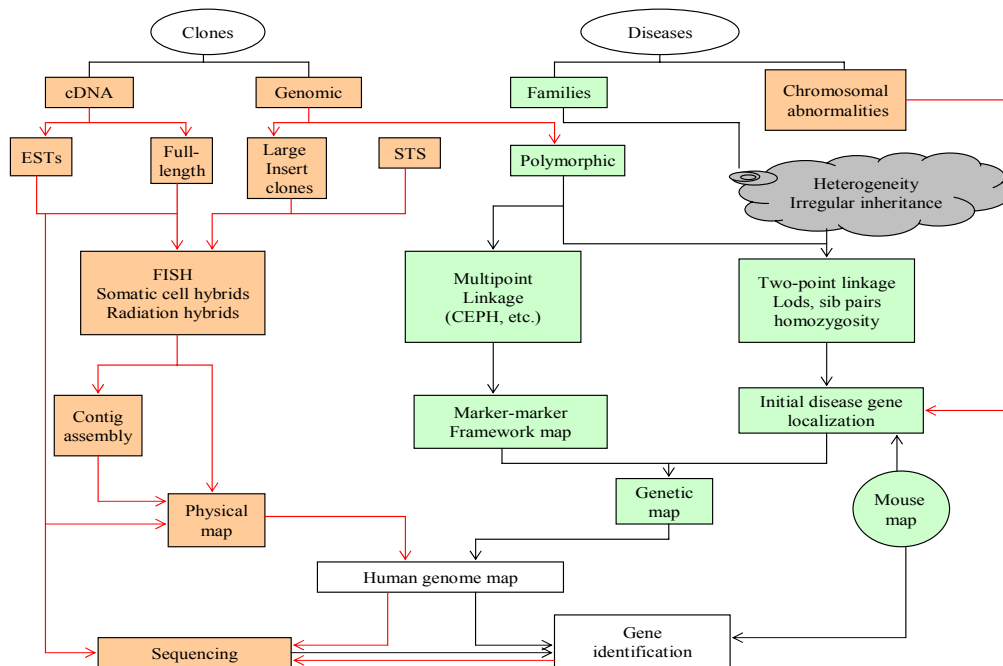
Fig. 3: Major scientific strategies and approaches used in the Human Genome Project (adapted from Strachan and Read, 1996)

Physical map of the Human genome

A physical map is not only a scaffold for genomic sequencing but also offers access to any genomic region, which is very important for gene cloning. Gene maps have been constructed in humans and different mammalian species for two reasons: first as a resource for locating the genetic determinants of heritable characters, behaviour and phenotypes; and second as a template for resolving and interpreting pattern of evolving genome organization in their ancestry. A new dimension has been introduced into biomedical research by HGP and other related programmes for various organisms. The main goal is to determine the sequence of three billion nucleotide pairs in the DNA of the human genome and to find all the genes therein.

A variety of different physical maps of the human genome like Cytogenetic, Chromosome breakpoint maps, Restriction map, Clone contig map, Sequence-tagged site (STS) map, Expressed sequence tag (EST) map and DNA sequence maps, were constructed and further are used to map the human genome.

The first physical map was obtained when cytogenetic banding technique made way for subchromosomal identification (ISCN, 1995). Though the resolution was low, it has been very useful as a framework for assigning the locations of human DNA sequences by chromosome *in situ* hybridization techniques. Some rare-cutter restriction maps have been achieved for a few human chromosomes; and the first restriction map using *NotI* restriction enzyme on chromosome 21 was published by Ichikawa et al. (1993). With the advent of the YAC system (Burke et al., 1987) it became possible to clone large inserts up to 2 Mb (Anderson, 1993) and the first generation physical map of the human genome was constructed with CEPH YAC library by Cohen et al. (1993). But a major disadvantage of the YAC libraries is its high frequency of chimeric clone formation (Green et al., 1991; Bray-Ward et al., 1996). This problem was overcome to some extent by verifying the physical location of the individual YAC by FISH and some were established in terms of cytogenetic bands on some chromosomes. Since most of the YAC clones were positive for STS markers (Hudson et al., 1995) with defined genetic linkage distances (Dib et al., 1996; Broman et al., 1998) corresponding to their physical locations, the data facilitated integration of cytogenetic, genetic and physical maps of the human genome (Bray-Ward et al., 1996; Cox et al., 1996). This integration has been particularly useful in the maximum coverage of the genome (Crollius et al., 1996; Bouffard et al.,

1997). However in some gene rich regions, higher resolution contig maps were constructed using overlapping cosmid, bacteriophage lambda, P, P1-artificial chromosome (PAC) and Bacterial artificial chromosome (BAC) clones. The development of cloning systems like PAC (Ioannou et al., 1994) and BAC (Shizuya et al., 1992) proved to be quite important in the success of the whole-genome map as they were more stable than YAC clones. Ultimately, high resolution maps based on PAC and BAC clones provided suitable framework for sequencing and much better coverage of the genome. Though initial efforts to construct clone based regional and even chromosomal physical maps of the human genome using cosmid libraries derived from isolated human chromosomes met with limited success (Doggett et al., 1995), a reasonable physical map of the human genome based on clones was published in 2001 by McPherson et al. The physical map of the overlapping clones was constructed by checking all BACs from different libraries for the restriction fragments pattern, and thus established a fingerprint for each BAC (Soderlund et al., 1997; Sulston et al., 1989). This helped in distinguishing different BACs, and in turn assessing the degree of overlaps. Hence, the clone based map has been vital for the accurate assembly of the human genome sequence without any gaps (International Human Genome Mapping Consortium (IHGMC), 2001). Other types of map like the EST map required cDNA sequencing and also mapping cDNAs back to other physical maps (Adams et al., 1995). Subsequently, subchromosomal localization of human cDNA clones by *FISH* has also been possible (Korenberg et al., 1995). A complete contig map of a chromosome would therefore comprise the entire DNA without any gaps. Finally, obtaining the physical map with complete nucleotide sequence of the genome providing the highest possible resolution up to a single base pair is the desired ultimate goal.

The advances of the HGP and the completion of the total genome sequences of many species enable us to view the gene information of the entire genome. As a result the mechanisms for some of the genetic diseases are best understood at a genomic level.

1.2.1 Identification of human disease genes

The resources provided by the HGP are shaping the strategies used for disease gene identification (Ballabio, 1993). The identification of genes involved in human diseases is important to understand the pathophysiology of the disease, further it often provides new insights into normal human development and biology. Basically, diseases can be identified by four main approaches. They are positional cloning, functional cloning, positional candidate approach and cloning of a candidate gene. A general strategy of functional cloning requires prior knowledge of the function of the gene. For example the gene for phenylketonuria was identified by purifying the enzyme from liver and raising antibodies which immunoprecipitated polysomes containing phenylalanine-hydroxylase mRNA (Robson et al., 1982). Positional cloning, a term coined by Collins in 1992, requires the knowledge of its physical location in the genome and usually with little information about the function. This requires laborious methods of chromosome walking and the identification of expressed sequences. The first gene isolated by positional cloning was chronic granulomatous disease gene in 1986 by Royer-Pokora et al. The genes for many important diseases like Duchenne muscular dystrophy, Cystic fibrosis, Huntington's disease, Adult polycystic kidney disease, Colorectal cancer, and Breast cancer were also isolated by positional cloning (Ballabio, 1993). The candidate gene approach relies on the availability of the information from previously isolated genes (Collins in 1992). The positional candidate approach involves the combination of mapping the genes to the correct chromosomal subregion and then checking for candidates in that region. For example Marfan syndrome was mapped to chromosome 15q using the information available from both positional and candidate approach (Kainulainen et al., 1990; Magenis et al., 1991; Dietz et al., 1991).

1.3   Subject of the present study

In this work, a three generation familial pericentric inversion of chromosome 3 with short stature is studied. The proband had been referred to the Department of Human Genetics and Medical Biology, Martin-Luther University, Halle-Wittenberg for chromosomal analysis of short stature. She was born in 1986 and was 146 cm tall in 2002. Apart from short stature, no other dysmorphic features were identified in either clinical or biochemical point of view. Initial cytogenetic analysis revealed a karyotype of 46, XX inv (3)(p23; q25q26)(Fig. 4A). To establish the origin of this pericentric inversion, chromosomal analyses of the parents and the sibling were done. The karyotypes of both the father and the brother were normal with 46,XY but the karyotype of the mother showed the same inversion as that of the proband. In order to further elucidate the inversion, the chromosomes of the maternal grandmother and her other daughter were also analyzed. The maternal aunt of the proband showed a normal karyotype, whereas the maternal grandmother carried the same inversion. The pedigree of the family, along with the year of birth, heights and the initial karyotypes in three generations, is shown in Figure 4B.
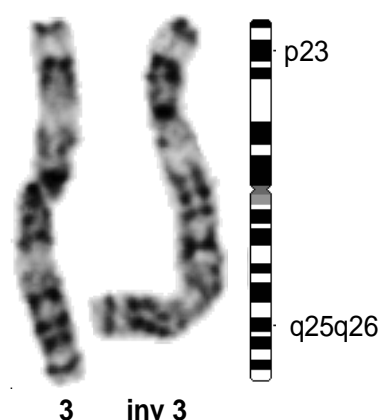
Fig. 4A: The GTG banded partial karyogram of the proband showing the pericentric inversion of chromosome 3 and the ideogram showing the cytogenetically mapped inversion breakpoint regions.
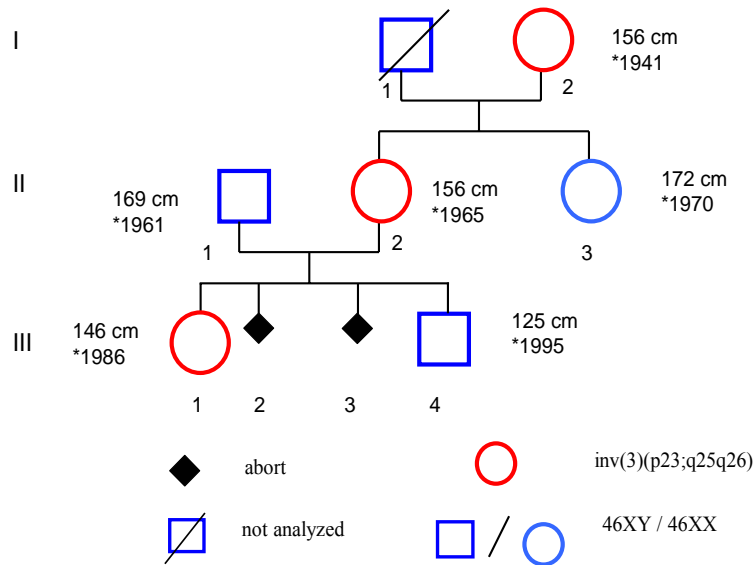
Fig. 4B: The pedigree showing the familial pericentric inversion. The heights and the years of birth are also shown.

1.4   Possible causes of short stature

The causes of short stature are heterogeneous. Many genes and environmental influences are involved in the individual shaping of body size. So a systematic classification of the various syndromes involving short stature is not exactly interpreted till date (Enders, 1992).

Hormonal involvement

Growth hormone deficiency may occur by itself or in combination with one or more other pituitary hormone deficiencies. Growth hormone (GH) deficiency is a rare cause of short stature. The gene encoding growth hormone is GH1, which shows a variety of defects, including deletions, frameshifts, splice sites and nonsense mutations, and leads to either decreased GH expression or action (Wagner et al., 1998). A novel homozygous 5' splice site mutation in the Growth Hormone-Releasing hormone Receptor (GHRHR) gene as cause of dwarfism with familial isolated growth hormone deficiency (IGHD) was

described by Salvatori et al. (1999). In pituitary dwarfism, the Pit-1 gene encodes the POU-domain transcription factor Pit-1 which is important for the development of the anterior pituitary gland and the expression of the GH, prolactin (PRL) and Thyroid stimulating hormone (TSH) genes. R271W mutation of the Pit-1 gene causes hypoplasia of the pituitary gland and deficiencies of those hormones (Aarskog et al., 1997). Many mutations have been identified in the transcription factors which regulate the pituitary development like Pit1 (POU1F1) (Pfaeffle et al., 1992), PROP1 (Wu et al., 1998), HESX1 (Dattani et al., 1998) or LHX3 (Netchine et al., 2000) and LHX4 (Machinis et al., 2001).

Major syndromes involved in short stature

About 74 major syndromes describe short stature as one of the characteristic features (OMIM). However, only a few syndromes are described here. SHORT syndrome associated with mild intrauterine growth retardation, postnatal growth deficiency and delayed bone age was first reported in 1975 by Gorlin et al. and Sensenbrenner et al. Achondroplasia (ACH) is the common genetic form of dwarfism. ACH candidate region includes the gene encoding fibroblast growth factor receptor 3 (FGFR3). Point mutation in the FGFR3 gene in ACH heterozygotes and homozygotes is responsible for ACH (Shiang et al., 1994). The most common cytogenetic cause of short stature in females is Turner syndrome (Ullrich, 1930; Turner, 1938). It is a common developmental disorder in females which is also characterized by short stature. A pseudoautosomal location for a dosage-sensitive locus involved in stature has been suggested based on the analyses with deletions of a specific segment of the short arm pseudoautosomal region (PAR1); hemizygosity for this putative locus probably contributes to the short stature in Turner individuals. Individuals with deletions of portions of short arm of PAR1 were short (Ogata et al., 1992). Two allelic variants of Turner syndrome Leri-Weill Dyschondrosteosis and Langer mesomelic dysplasia were also reported (Belin et al., 1998; Shears et al., 1998). Dyschondrosteosis is an autosomal dominant form of mesomelic dysplasia where short stature due to shortening of the forelegs, madelung deformity of the forearm with bowing of the radius and dorsal dislocation of the distal ulna is frequently observed.

Identification of genes influencing short stature

On sex chromosomes

The genetic causes of growth abnormalities were well reviewed by Mendez et al. in 1985. The identification of genes that influence the complex trait height is more complex. However candidate gene and positional cloning approaches, using linkage analysis on a whole genome scale, helped in isolating some genes on both sex and autosomal chromosomes. Many chromosomal aberrations in patients with short stature were reported. The Short stature HOmeobox-containing gene (SHOX) on the short arm of the X and Y chromosome is an important determining factor of stature phenotype (Musebeck et al., 2000). The SHOX gene was mapped to the Xp22.3 (Ellison et al., 1997) and the haploinsufficiency of this gene leads to growth failure and short stature (Rao et al., 1997). Interestingly, SHOX mutations have also been described as causative for the Leri-Weill syndrome, a mesomelic short stature syndrome (Belin et al., 1998). In addition, homozygous SHOX mutations have been shown to cause Langer-type mesomelic dwarfism (Belin et al., 1998; Shears et al., 1998). A homeodomain identical to SHOX called PHOG was found to be a candidate for involvement in the short stature of Turner syndrome (Ellison et al., 1997).

On autosomal chromosomes

Autosomal aberrations with short stature also have been reported. The SHOX HOmologous gene on chromosome Three (SHOT) a SHOX-related human gene has been identified and localized on chromosome 3q25-26 (Blaschke et al; Baere et al., 1998). This SHOT is closely related to SHOX on sex chromosomes and also has a murine counterpart OG12X (Clement-Jones et al., 2000). Some of the known growth related genes identified on other autosomes are Cathepsin K on 1p13, FGFR3 on 4p16, Oestrogen receptor on 6q25, GHRH receptor on 7p14, ROR2 on 9q22, LHX3 on 9q32-34, Col1A1, GH1 on 17q22-23 (Pfaeffle, 2006), and Growth hormone-releasing factor (GHRF) or Somatocrinin on 20p12 (Rao et al., 1991).

1.5   Insight into some common chromosomal rearrangements involving 3q26

Many different structural anomalies involving 3q26 region are well reported in the literature. The structural abnormalities taking the form of intra- and inter-chromosomal

rearrangements either between chromosomes 3, for example t(3;3)(q21;q26) or between 3q21 or 3q26 and other chromosomes were reported by Pintado et al., 1985. But majority of the rearrangements with 3q26 are involved with cancer. For example translocations or inversions of chromosome 3 with breakpoints involving band 3q26 were specifically associated with megakaryoblastic acute phase or abnormal megakaryocytopoiesis. The occurrence in myelodysplastic syndromes and myeloid leukemias of acquired abnormalities of chromosome 3 is well documented by Belloma et al. (1992). Acute leukaemia with abnormal thrombopoiesis is also associated with pericentric inversion or homologous translocations of chromosome 3 involving bands 3q21 and 3q26.2; these were described in a few patients with acute non lymphocytic leukemias (ANLL) (Bitter et al., 1985) and dysmyelopoietic syndromes (DMPS) (Carroll et al., 1986) or accelerated phases of chronic granulocytic leukaemia (CGL) (Carbonell et al., 1982). This association suggested the presence of important thrombopoiesis gene on chromosome 3q (Bitter et al., 1985). Ecotropic virus integration-1 (EVI1) gene located at 3q26 has been reported in individuals with AML involving translocation or inversions of chromosome 3 long arm (Levy et al., 1994). Philadelphia chromosome as a secondary abnormality in inv(3)(q21q26) acute myeloid leukaemia (Han and Theil, 2006) and t(3;17)(q26;q22) as an additional change in Philadelphia positive chronic myelogenous Leukemia in acceleration were also observed (Mugneret et al., 1992).

1.6   Chromosome 3 and short stature association

A few cases associated with short stature and chromosome 3 abnormalities are also known in the literature. For example, a case of Turner syndrome with a familial pericentric inversion inv (3)(p25q21)pat. This case showed primarily short stature with minor form of Turner syndrome like monosomy and isochromosome X and a familial pericentric inversion on chromosome 3 (Stine et al., 1982). Also, a female child with short stature showing ring chromosome 3 abnormalities along with hypoplastic thumb and coloboma of iris (Barajas et al., 2001) was described. Two cases of trisomy 21(pter-q22.1) with no major features of Down syndrome but with moderate mental retardation and short stature were also reported. Interestingly, the extra chromosome 21 contained the distal part of chromosome 3p in one case and 14q in the other (Kondo et al., 2006). In another case a 11 year old boy with short stature and learning difficulties showed a

chromosome 3p23 break with ring formation and translocation of displaced 3p23-->pter segment to 6pter (Yip et al., 1996). A case of distal segment of chromosome 3q deletion with primary ovarian failure and short stature as one of the major feature was reported by Nguyen et al. (2005). Finally, a partial monosomy 3q in a boy with short stature, developmental delay and mild dysmorphic features was also reported (Brueton et al., 1989).

1.7   Aim of the study

As a primary aim, an inversion chromosome 3 segregating in a three generation family with short stature was studied by positional cloning:

- Initially, YAC (Yeast Artificial Chromosome) clones should be selected by *in silico* analysis using the human genome database and should be used to delineate the inversion breakpoint by FISH (Fluorescence *in situ* hybridization).

- Further, BAC (Bacterial Artificial Chromosome) clones should be selected and breakpoints spanning clones on both 3p and 3q regions should be identified and characterization of those clones by different strategies and techniques should be used to narrow down the breakpoint.

- Data mining using the DNA sequences of the clones anchoring the breakpoints should be carried out to find out known or putative gene(s) located in or close to the critical regions which might be a candidate gene for this phenotype.

- By molecular methods the junction fragments of the breakpoint regions should be characterized.

- Subsequent *in silico* analysis using the public databases should give information about the gene contents and the sequence features, like repetitive elements and segmental duplications at the inversion breakpoint region. The characterization of the DNA motifs at the breakpoint regions should also be performed to understand the mechanism underlying this familial inversion.

The detailed analysis of the inversion breakpoints might help to provide a better understanding of the reasons causing short stature, such as the forces driving chromosomal rearrangements and structural types promoting breakage, but also may provide insights in human pathology of short stature.

2   Materials and Methods

2.1   Materials

The common materials required for this study are listed below along with their companies. All the solutions and buffers were prepared with Millipore filtered water if not described otherwise.

2.1.1   Reagents

Acetic acid	Merck
Adenine	Merck
Agarose	Roth
Agarose, low melt	Roth
Agarose Pulsed-Field certified	BioRad
Ammonium acetate	Merck
Ammonium chloride	Sigma
Ampicillin	Boehringer
Ampuwa	Fresenius Kabi
Anti-Digoxigenin-AP Fab fragments	Roche
L-Arginine-Hcl	Serva
Bacto-Agar	Fluka
Bacto-Tryptone	Roth
Bacto-Yeast-Extract	Roth
BAAD	Vector
BCIP	Roche
Biotin-16-dUTP	Roche
Blocking reagent	Roche
Boric acid	Calbiochem
BSA	Serva

Calcium chloride	Merck
Chloramphenicol	Boehringer
Chloroform	Sigma
Citric acid monohydrate	Chemapol
Colcimid	Calbiochem
Cot 1-human DNA	Roth
DAPI	Sigma
Dextran sulphate	Sigma
Digoxigenin-11-dUTP	Roche
N, N-Dimethyl formamide	Roth
Dithiothreitol	Merck
dNTP	Invitrogen
ECF substrate	Amersham
EDTA	Roth
Ethanol	Sigma-Aldrich
Ethidium bromide	Roth
FITC	Vector
Fixogum rubber cement	Marabu
Formaldehyde	Sigma
Formamide	Fluka
Giemsa stain	Sigma
D-Glucose	Roth
Glycerol	Roth
Glycogen	Roche
Guanidine hydrochloride	Roth
L-Histidine	Serva

Hydrochloric acid	Sigma
L-Isoleucine	Serva
Isopropanol	Fluka
IPTG	AppliChem
Kanamycin	Boehringer
L-Leucine	Serva
Lithium chloride	Sigma
L-Lysine-HCl	Serva
Maleic acid	Roth
Magnesium chloride	Merck
Methanol	Roth
L-Methionine	Serva
NBT	Roche
PBS	Biochrom AG
Pepsin	Sigma
Phenol	Roth
L-Phenylalanine	Serva
Phosphoric acid	Merck
Phytohaemaglutinine	Biochrom AG
Potassium acetate	Roth
Potassium chloride	Merck
Potassium *di* hydrogen phosphate	Merck
Potassium hydrogen carbonate	Merck
Propidium iodide	Sigma

Rapid Hyb	Amersham
Salmon sperm DNA	Sigma
SDS	Roth
Sodium acetate	Roth
tri-Sodium citrate-*di* hydrate	Roth
di-Sodium hydrogen phosphate	Merck
Sodium hydroxide	Roth
Sodium chloride	Roth
D-Sorbitol	Roth
Streptavidin AP conjugate	Roche
TAE (10X)	Roth
TBE (10X)	Roth
Tris-HCl	Roth
Triton-X-100	Serva
Tween®-20	Roth
L-Tyrosine	Serva
L-Valine	Serva
Vectashield	Vector
X-gal	Roth
Yeast Nitrogen Base	Invitrogen

2.1.2   Enzymes

Agarase	Roche
Alkaline phosphatase (Shrimp)	Roche
DNase I	Serva

DNA polymerase-I	Roche
Klenow polymerase	Roche
Lyticase	Sigma
Pepsin	Sigma
Proteinase K	Sigma
Restriction endonucleases	Fermentas, NEB
RNase A	Sigma
Taq DNA polymerase	Fermentas
Trypsin	Serva
T4 DNA ligase	Promega
Long PCR Enzyme Mix	Fermentas

2.1.3   Antibiotic stocks

Antibiotic	stock	working concentration
Ampicillin	50mg/ml in ddH$_2$O	50-100µg/ml
Chloramphenicol	34mg/ml in Ethanol	100µg/ml
Kanamycin	30mg/ml in ddH$_2$O	30µg/ml

2.1.4   Buffers and stock solutions

All the solutions and buffers were made according to the protocols adapted from Sambrook et al. (1989). All chemicals were dissolved in either double distilled water or Ampuwa (pyrogen free double distilled water) and autoclaved or filter sterilized.

Amino acid stock (10X)	200 µg/ml Adenine
	200 µg/ml L-Arginine-HCl
	200 µg/ml L-Histidine
	200 µg/ml L-Isoleucine
	600 µg/ml L-Leucine
	200 µg/ml L-Lysine HCl
	200 µg/ml L-Methionine

500 µg/ml L- Phenylalanine

300 µg/ml L-Tyrosine*

1.5 µg/ml L-Valine

*L-Tyrosine was dissolved separately
in a drop of 1M NaOH and added to
the stock solution.

Blocking solution (*FISH*)	5 % BSA
	2 X SSC
	Filtered in Whatman
	folded filter and again
	filter sterilized.
Blocking solution (DOT-BLOT)	10 % Blocking reagent
	in Maleic acid buffer
Cell Resuspension buffer	1 M Sorbitol
	0.1 M Tris-Hcl
	pH 7.6
DAPI stock solution	1 mg/ml in 10 X PBS
DAPI working concentration	2 µl in 1ml 10 X PBS
Denaturation solution	1.5 M Sodium chloride
	0.5 M Sodium hydroxide
Denaturation solution (*FISH*)	70 % Formamide
	2 X SSC
	pH 7.0

Dig DNA labeling mix (10X)   1 mM dATP

                                                    1 mM dCTP

                                                    1 mM dGTP

                                                    0.65 mM dTTP

                                                    0.35 mM Dig-dUTP

DNase I   1 mg/ml in Ampuwa

                                                    aliquots stored at -20ºC

dNTP-mix   10 mM dATP

                                                    10 mM dCTP

                                                    10 mM dGTP

                                                    10 mM dTTP

dNTP-mix (Nick Translation)   1 mM dATP

                                                    1 mM dCTP

                                                    1 mM dGTP

                                                    0.65 mM dTTP

Buffer-I   25 mM Tris-HCl

                                                    50 mM Sodium chloride

                                                    0.05 % Triton-X-100

                                                    pH 7.5

Buffer-II   1 % Blocking solution

                                                    in Buffer-I

Buffer-III   25 mM Tris-HCl

                                                    50 mM Sodium chloride

                                                    pH 9.5

Hypotonic solution	0.4 % Sodium citrate
	0.592 % Potassium chloride
	1:1 ratio and stored at 37ºC
Lyticase buffer	1 M Sorbitol
	1 mM EDTA
	10 mM Tris-HCl
Maleic acid buffer	0.1 M Maleic acid
	0.15 M Sodium chloride
	pH 7.5
Master-mix	20 % Dextran sulphate
	2 X SSC
Neutralization solution	0.5 M Tris-HCl
	1.5 M Sodium chloride
Nick Translation buffer	500 mM Tris-HCl
	100 mM Magnesium chloride
	10 mM DTT
	500 µg/ ml BSA
PBS (10X)	1.36 M Sodium chloride
	0.026 M Potassium chloride
	0.15 M *di*Sodium hydrogen phosphate *di*hydrate
	0.017 M potassium *di*hydrogen *ortho* phosphate
	pH 7.4 adjusted with HCl

PI Stock solution   1 mg/ml in 10 X PBS

PI working concentration   1 µl in 100 µl 10 X PBS

Post fixative solution   2.5 % Formaldehyde

                                                           0.02 M Magnesium chloride

                                                           made up with 1 X PBS

                                                           Stored at 4ºC

Proteinase-K Reaction buffer   75 mM Sodium chloride

                                                           1 mM EDTA

                                                           1 % SDS

Salmon sperm DNA   10 mg/ml

                                                           in TE (pH 7.5)

Solution I   50 mM Glucose

                                                           25 mM Tris-HCl

                                                           10 mM EDTA

                                                           100 µg/ml RNase A

                                                           pH 8.0

Solution II   0.1 M Sodium hydroxide

                                                           1 % SDS

Solution III   3 M Potassium acetate

                                                           pH 4.8

SSC (20X)   3 M Sodium chloride

                                                           0.3 M Sodium Citrate

                                                           pH 7.0

Tris buffer	10 mM Tris-HCl
	pH 8.5

TE buffer	10 mM Tris-HCl
	1 mM EDTA
	pH 7.5

Sterilization

The buffers and solutions were autoclaved for 20 minutes at 121ºC at 2000 hPa and heat labile solutions were filter sterilized with a 0.2 μm sterile filter.

2.1.5   Media

LB	1 % Bacto-Tryptone
	0.5 % Bacto-Yeast extract
	1 % Sodium chloride
	pH 7.0 with 10 M Sodium hydroxide
	Autoclaved

LB-Agar	LB medium
	1.5 % agar
	Autoclaved and approximately 25 ml poured in each Petri plate.

LB-X-gal	LB agar medium
	Autoclaved and cooled till 50ºC
	Ampicillin (100 μg/ml)
	X-gal (80 μg/ml)
	IPTG (0.5 mM)

SOC medium	2 % Bacto Tryptone
	0.5 % Yeast extract
	0.05 % Sodium chloride
	1 M Potassium chloride
	Autoclaved
	20 mM Glucose
	filter sterilized

YAC medium	0.067% Yeast nitrogen base
	2 % Glucose
	1 X amino acid stock

| YAC agar | YAC medium |
| | 1.5 % agar |

Commercial mediums

| RPMI 1640   + GlutaMax^{TM} | Gibco |
| + 25mM HEPES | |

| Amnio Max^{TM}-C100 | Gibco |

| MEM | Gibco |

2.1.6   DNA Molecular weight standards

Gene Ruler^{TM} 1Kb DNA ladder	Fermentas
Gene Ruler^{TM} 1Kb DNA ladder plus	Fermentas
Gene Ruler^{TM} 100bp DNA ladder	Fermentas
Lambda-pUC mix marker, 4	Fermentas

Lambda-pUC mix marker, 8   Fermentas

Lambda mix marker, 19   Fermentas

Mass Ruler^TM DNA ladder mix   Fermentas

Lambda DNA/*Eco*130I (*Sty*I)/*Mlu*I marker, 17   Fermentas

Yeast chromosome PFG marker   NEB

Mid range marker II   NEB

Low range marker   NEB

2.1.7   Vectors

pGEM 5Zf(+)   Promega

pBluescript SK   Promega

pUC 19   Promega

pGEM®-T   Promega
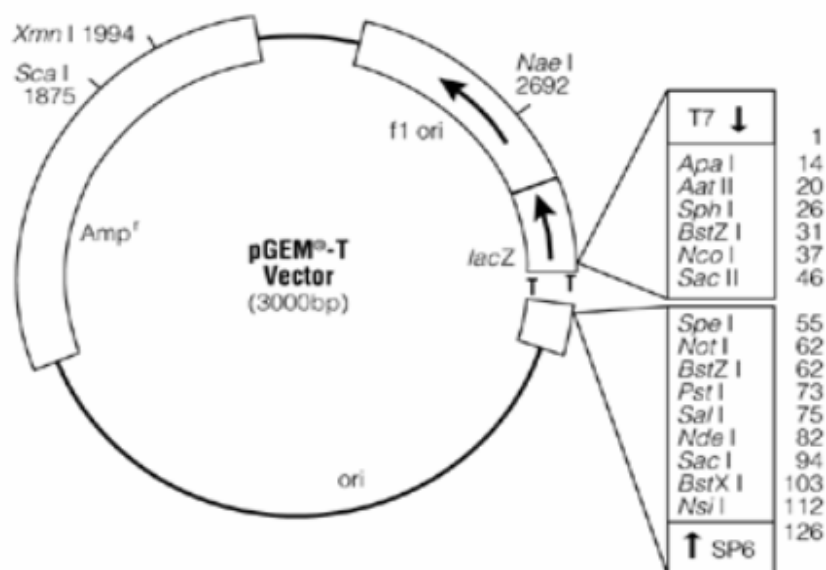
Fig. 5: Schematic representation of cloning vector pGEM-T (adapted from Promega Technical Manual # TM042).

2.1.8 Genomic Clones

All the clones were obtained from RZPD (Ressourcenzentrum Primaer Datenbank, Berlin) as agar stabs. The YAC vector pYAC4 (Fig. 6) and the BAC vector pBACe3.6 (Fig.7) are also shown.

2.1.8.1 YAC clones

Clone ID

Region: 3q	Accession number	
CEPHy904E01891D1	891E01	Library No: 904
CEPHy904E05855D1	855E05	Source: Lymphoblastoid cell lines
CEPHy904F10938D1	938F10	Host: S. *Cerevisiae* AB1380
CEPHy904G12806D1	806G12	Vector: pYAC4
CEPHy904H12943D1	943H12	Resistance: Ampicillin
CEPHy904G06807D1	807G06	
CEPHy904A03750D3	750A03	
CEPHy904B09845D3	845B09	
CEPHy904B11956D3	956B11	
CEPHy904C09858D3	858C09	
CEPHy904F11783D3	783F11	
CEPHy904G07889D3	889G07	
CEPHy904H01954D3	954H01	
CEPHy904H03886D3	886H03	

Region: 3p		
CEPHy904A11791D1	791A11	
CEPHy904B04925D1	925B04	
CEPHy904G05803D1	803G05	
CEPHy904H08787D1	787H08	
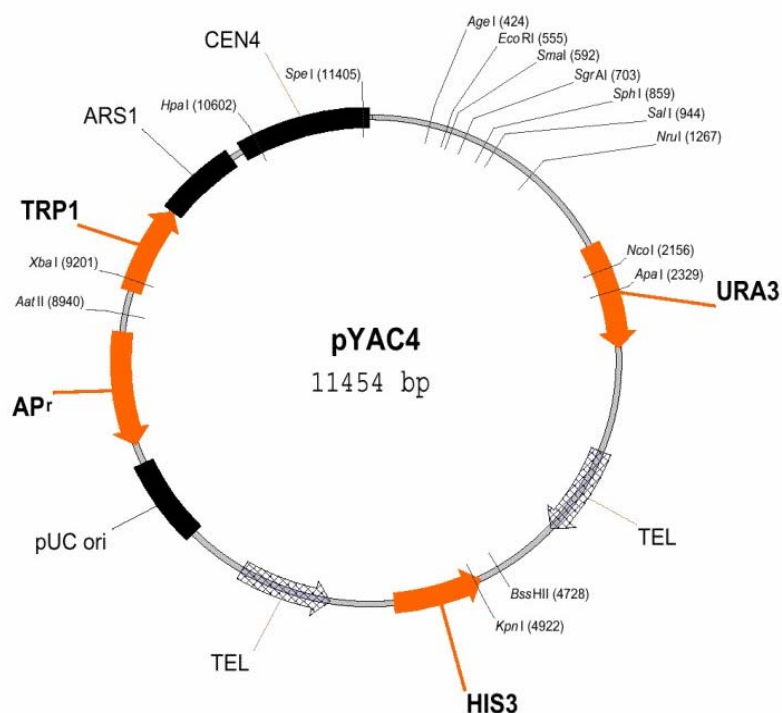
Fig. 6: Schematic representation of pYAC4 (adapted from http://www.rzpd.de/info/vectors/pYAC4-_pic.shtml).

2.1.8.2 BAC clones

Library: 753, Source: Blood, Host: E.*coli* DH 10B, Vector: pBACe3.6, Resistance: Chloramphenicol.

Clone ID	region	Standard clone
RPCIB753K2211	3q	RP11-11K22
RPCIB753L1412	3q	RP11-12L14
RPCIB753A20451	3p	RP11-451A20
RPCIB753K0491	3p	RP11-91K04
RPCIB753I20451Q8	3q	RP11-451I20
RPCIB753K1026Q8	3q	RP11-26K10
RPCIB753O7636Q8	3q	RP11-636O7
RPCIB753B20358Q8	3q	RP11-358B20
RPCIB753G08775Q8	3q	RP11-775G08

RPCIB753L09480Q8	3q	RP11-480L09
RPCIB753M05280Q8	3q	RP11-280M05
RPCIB753A05592	3p	RP11-592A05
RPCIB753I17130	3p	RP11-130I17
RPCIB753E16669Q8	3p	RP11-669E16
RPCIB753G20666Q8	3p	RP11-666G20

Library No: 737, Source: Blood, Host: E.*coli* DH 10B, Vector: pBACe3.6, Resistance: Chloramphenicol

RZPDB737A022128D6	3q	RP11-452B04
RZPDB737B012078D6	3q	RP11-265C03
RZPDB737D052006D6	3q	RP11-12N13
CTD-2007B5	3p	CTD-2007H5
CTD-2265H12	3p	CTD-2265H12
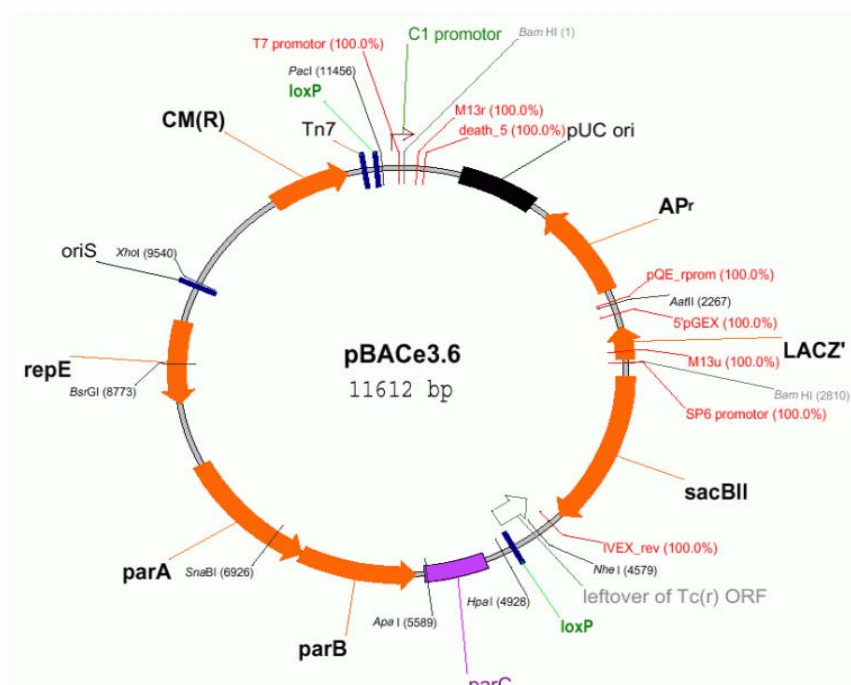
Fig. 7: Schematic representation of pBACe3.6 (adapted from http://www.rzpd.de/info/vectors/pBACe3.6-_pic.shtml).

2.1.8.3   cDNA clones
Clone ID

Region: 3p	**vector**
IMAGp998H1417075Q2	pExpress-1
IMAGp998B0410432Q1	pCMV-SPORT6
IMAGp998I0413343Q1	pCMV-SPORT6
IMAGp998K099654Q1	pCMV-SPORT6
IMAGp998M0514122Q1	pCMV-SPORT6
IMAGp958H192314Q2	pOTB7
IMAGp958B062469Q2	pOTB7
Region: 3q	
IMAGp998F1113301Q1	pCMV-SPORT6
IMAGp998B0415915Q1	pExpress-1
IMAGp998O074892Q1	pCMV-SPORT6

2.1.9   Bacterial stocks

Escherichia coli JM109	Promega
Escherichia coli DH5α	Invitrogen

2.1.10   Oligonucleotides
All the designed primers were delivered desalted by Invitrogen.

2.1.10.1 Primers for *Alu* PCR
YAC Vector primers

Ura1	$^{5`}$CGATATAGGCGCCAGCAACCGCACCTGTGG$^{3`}$
Trp 1	$^{5`}$CGCACCCGTTCTCGGAGCACTGTCCGACC$^{3`}$

***Alu* primers**

Alu-14	$^{5`}$TGCACTCCAGCCTGGGCAACA$^{3`}$
Alu	$^{5`}$GGATTACAGGCGTGAGCCAC$^{3`}$
Alu-3	$^{5`}$GATCGCGCCACTGCACTCC$^{3`}$
Alu-13	$^{5`}$GTGGCTCACGCCTGTAATCCC$^{3`}$
Alu-153	$^{5`}$GGGATTACAGGCGTGAGCCAC$^{3`}$

Alu-154 $5`$TGCACTCCAGCCTGGGCAACA$3`$

Alu-451 $5`$GTGAGCCGAGATCGCGCCACTGCACT$3`$

2.1.10.2 Primers for Long range PCR

d1 $5`$CTGTGTCTTAGGGTCTACTTTGGTCAGAAT$3`$

d2 $5`$CAGTGTTGAAATAGTTATCACCATGAGGAC$3`$

d3 $5`$CTCTTCCAATCTTTGAGCTATTCTCTGAAG$3`$

d4 $5`$CAGTTTTTCATACAGCTTACTCCCCTGACTT$3`$

d5 $5`$GATAGGCTACACACCTGACCTAAATACCAC$3`$

d6 $5`$AGTCTCAAGAGTTTCTTGGCTTGTATATGG$3`$

d7 $5`$GTTTTGTGCCCTATTAAGGAGAGAAGTAAC$3`$

d8 $5`$TATTCACATCCAAGATGGATATAGCAGAGG$3`$

d9 $5`$AGAAGCTGCAAGAAGATAAGAAGCCTCCTG$3`$

d10 $5`$TTCTTCAGGCTAGGTCTTATTTTGTCTGGG$3`$

d11 $5`$TAGAAATTTGTATGGCCAATCCTTCACATG$3`$

d12 $5`$CTTTGTGGCTAGTCAGTTGGGGAAATAGTC$3`$

d13 $5`$AGATTTACAGCTATTTGTTGCCCTGACATC$3`$

d14 $5`$GTGCTGTCAAGTTTCAAGAAGATTGTATTC$3`$

2.1.10.3 Primers for breakpoint spanning PCR

U1 $5`$CTGATGATTAAAGGGATGAAGAC$3`$

U2 $5`$ACCTGGTTGTTGGAGCTTATC$3`$

U3 $5`$TTGACAAGCATGTAATCCAGG$3`$

U4 $5`$AGTTGGGATTGGAGCTATGG$3`$

U5 $5`$TCATACAGCTACTACGTGGTACC$3`$

U6 $5`$GACGGAGTCTGGATCTGTTACC$3`$

U7 $5`$AAGTGTGCTTTGGGACTGG$3`$

U8 $5`$GGTCTGGATGGTCGCGATCTC$3`$

U9 $5`$TTATCCCCATGCCCAAAGC$3`$

U10 $5`$AAGCTACCAGTGGCAGATGC$3`$

U11 $5`$ACCAAGTTGTTAAACCTGACC$3`$

U12 $5`$ATTAACCCAAGACCTTAGCG$3`$

U13   5'CAACAACAGAGGGCATAGTG3'

U14   5'TGTGGCAGGTTGTATTTCC3'

U15   5'TGGTTCTTACACTGTGTTCAG3'

U16   5'TTAAGTTGAGCCATAAATGC3'

U17   5'AGTTTCCCTTGGATGTCAC3'

U18   5'TGGAAGATTGGAGAGCTATC3'

U19   5'CATTAACAAGCACGCTACTG3'

U20   5'ACCTATGCTAACCACTCACAG3'

U21   5'TGCACTTCTTCAATAGATTCC3'

U22   5'CCACTTGCCACCTACTCATAC3'

U23   5'GAAATGAGCAAAGGTGAACC3'

U24   5'CTCGGTTCGTCTAAAGCTGC3'

U25   5'CATGGATTGGCAATTATAAGTAG3'

U26   5'AGCTGCTGTCCAGTACAAGAG3'

U27   5'TGCCAGCTTCAAATTGACTG3'

U28   5'GTCTATGTCCGCTAAGGTCTTG3'

U29   5'CAATGCTCAGTGGTAGGCC3'

U30   5'GCTTGCTTCTCATTGATCTG3'

U31   5'GACCATTAGCACCTCCTCTTC3'

2.1.10.4 Primers for RT-PCR

Ut1   5'GTCCCATCAAGTATTGTGAG3'

Ut2   5'ATTCACCTGGACCGTTCTG3'

Ut3   5'CGAAGTGATAGTTTCAGCATG3'

Ut4   5'CATAACTCAACTCATTGGC3'

Ut5   5'TGGAGACTTTCAGCTTGC3'

Ut6   5'GACATTGTCTGGGAGCAGC3'

Ut7   5'AGTAGATCCTGAAGGCGTG3'

Ut8   5'CCCACTTGTGAGCTTGATAG3'

Ut9   5'AGCATTGCCAAACTCTAAGACC3'

Ut10   5'GTTTCCAGCCTGCACCTTG3'

Ut11   5'GCCTCTCTGTATAATGCTGTC3'

Ut12 $^{5`}$TGGTAGAGGTGCCGAATGAG$^{3`}$

Ut13 $^{5`}$GGACTCAGAAGGAAAACAGC$^{3`}$

Ut14 $^{5`}$GTCTCCAGGATGTACAGTTG$^{3`}$

Ut15 $^{5`}$ATCTTCCCACTGCATTAGCTG$^{3`}$

2.1.10.5 Primers for cycle sequencing PCR

M-13 forward $^{5`}$GTTTTCCCAGTCACGAC$^{3`}$

M-13 reverse $^{5`}$CAGGAAACAGCTATGAC$^{3`}$

2.1.11   Kits

Big Dye$^{®}$ Terminator v.1.1 cycle sequencing kit   Applied Biosystems

Nucleobond$^{®}$ AX plasmid isolation kit   Macherey-Nagel

pGEM$^{®}$–T Vector system II   Promega

Qiagen$^{®}$ plasmid mini kit   Qiagen

Qiaprep$^{®}$ Spin miniprep kit   Qiagen

Wizard$^{®}$ DNA Clean-up system   Promega

2.1.12   Glassware and Disposable ware

Cover slips   Roth

Micro centrifuge tubes   Sarstedt

Folded filters   Schleicher & Schuell

Glass slides   Menzel-Glaser

Glassware   Schott

Hybond^{TM}- N   Amersham

Montáge$^{®}$ PCR filter units   Millipore

Parafilm   Alfresa pharma corporation

Petriplates   Sarstedt

Pipette tips   Sarstedt

Plasticware   Sarstedt

Sterile filters 0.20μm   Heinemann Labortechnik

Sterile filter units   Millipore

Filter papers   Whatmann

2.1.13   Special Instruments

CHEF-DR$^{®}$III System   Biorad

Smart card reader EasyjeeT   EquiBio

Phosphoimager, Strom 860   Molecular Dynamics

Pressure control station   Stratagene

Microscope and attachments

AXioplan 2   Carl Zeiss

AXioskop   Carl Zeiss

UV lamp HBO 100W

IMAC-CCDS30, Video, Camera module

Objectives

Plan-NEOFLUAR 100X/1.30, oil immersion

Ph1, Plan-NEOFLUAR 10X/0.30

Plan-Apochromat 63X/1.40, oil immersion

2.2 Methods

All the methods used in this study are refered and the methods which are not mentioned otherwise are adapted from Sambrook et al. (1989).

2.2.1 Cell culture and Cytological preparations

2.2.1.1 Lymphocyte cultures
- 10 ml of Amniomax C-100 medium/ MEM medium / RPMI 1640 medium was taken in T25 flask with 200 μl of Phytohaemaglutinine (PHA) (1.2 mg/5ml) and 1 ml of heparinized whole blood
- The flasks were incubated at 37ºC for 72 hrs

2.2.1.2 Metaphase chromosome preparation
- 200 µl of colcemid (10 µg/ml) was added 1 ½ hr before the end of the culture time and incubated at 37ºC
- Cultures were transferred to 15 ml tubes and centrifuged at 1000 rpm for 10 min
- Supernatant was discarded, pellet was broken and then resuspended in 10 ml of hypotonic solution and incubated at 37ºC for 20 min
- Tubes were prefixed with a few drops of fixative (Methanol: Acetic acid 3:1) and centrifuged at 1000 rpm for 10 min
- Pellet was resuspended in 10 ml of ice cold fixative and then washed thrice
- 500 µl of the fixative was mixed with the pellet to make a cell suspension
- 2-3 drops of cell suspension were dropped on precleaned slides from a distance with an angle of approximately 45º, and then air dried and was finally stored at 4ºC or -20ºC

2.2.1.3 GTG banding (G-banding using trypsin and Giemsa)
- Slides were treated with trypsin (0.04g/100 ml of PBS) for 10-20 sec
- They were then washed in ddH$_2$O
- Finally, these were stained with Giemsa (0.4%)

2.2.2 DNA Isolation

2.2.2.1 YAC DNA isolation in agarose blocks

(The Fifth Wellcome Summer School, London, 1990)

- 100 ml O/N grown cultures were transferred to 50 ml tubes and centrifuged at 2000 rpm at 4ºC for 15 min

- Pellet was resuspended in 25 ml 0.05 M EDTA pH-8.0 and centrifuged

- Pellet was dissolved in 250 µl of Cell Resuspension buffer, 15 µl of Lyticase (5000 U/ml) and transferred to an micro centrifuge tube

- Equal volume of 2 % low melt agarose was cooled to 50ºC and added to the cell suspension, mixed and loaded into plug caps

- Plug caps were cooled on ice and the blocks transferred into 50 ml tubes

- 2.5 ml of Lyticase buffer and 85 µl of Lyticase were added and incubated at 37ºC for 2 hrs

- The blocks were washed with ddH$_2$O

- 2.5 ml Proteinase K-Reaction buffer, 100 µl Proteinase-K (600 U/ml) were added and incubated at 50ºC for O/N

- Blocks were washed with 1 X TE and stored in 20 ml of 0.05M EDTA at 4ºC

2.2.2.2 YAC DNA isolation from Agarose blocks

(Modified protocol from Boehringer)

- Two agarose blocks were weighed and 0.04 vol of 25X Agarase buffer was added and incubated at 65ºC for 15 min

- These were reincubated at 45ºC with 5-6 units of Agarase enzyme for 2 hrs

- Equal volumes of phenol and chloroform were added and centrifuged for 10 min

- To the upper watery phase 0.1 vol 3M Sodium acetate and 2.5 vol ethanol were added and then precipitated at -20ºC for 30 min

- Thew mix was micro centrifuged at 15,000 rpm at 4ºC for 20 min

- Pellet was washed with 70 % ethanol for 5 min

- It was dried and dissolved in 20 µl TE

2.2.2.3 BAC DNA Isolation

2.2.2.3.1 BAC DNA preparation by Nucleobond kit

(Birnboim and Doly, 1979)

This method was followed as per the instructions provided by the user manual, PT 3167-1, (PR2Y247 published on 12th Nov 2002). The basic principle lies with the lysis of bacterial cells by buffers based on the NaOH/ SDS lysis method. The plasmid DNA is bound to the resin in the column, later washed and then precipitated. The yield of the DNA is based on the amount of the culture and the column used.

Column	Expected DNA
Mini (AX-20 column)	20 μg
Midi (AX-100 column)	100μg

- 5 ml/ 100 ml of O/N grown cultures were centrifuged at 4000 rpm for 15 min at 4ºC for mini/ midi preparations respectively

- Pellet was suspended in 0.4/ 4 ml of S1 buffer+ RNase A

- 0.4/ 4 ml of S2 buffer was added, mixed gently and incubated at RT for 5 min

- 0.4/ 4 ml of S3 buffer was added and kept on ice for 15/ 25 min respectively

- Lysate was filtered in a folded filter

- AX-20/ AX-100 columns were equilibrated with 1/ 2.5 ml of N2 buffer

- Cleared lysate was added to the column and allowed to flow through

- Columns were washed twice with 1.5/ 5 ml of N3 buffer

- 1/ 5 ml prewarmed N5 buffer (50ºC) was added to the column to elute DNA

- 0.75/ 3.5ml Isopropanol was added to the elute and precipitated at -20ºC for 30 min

- It was micro centrifuged at 15,000 rpm at 4ºC for 20 min and pellet was washed with 70% ethanol for 5 min

- Pellet was dried and dissolved in 10/ 25 μl of TE

2.2.2.3.2 Mini plasmid DNA preparation by Qiagen plasmid mini kit

Qiagen plasmid purification protocols are based on a modified alkaline lysis procedure, followed by binding of plasmid DNA to QIAGEN anion exchange resin under appropriate low salt and pH conditions. RNA, proteins, dyes and low molecular weight impurities are removed by a medium salt wash. Plasmid DNA is

eluted in a high salt buffer and then concentrated and desalted by isopropanol precipitation.

- 5 ml of O/N grown cultures were centrifuged and pellet suspended in 0.3 ml P1 buffer and 0.3 ml P2 buffer was added, mixed gently and incubated at RT for 5 min

- 0.3 ml ice cold P3 buffer was added and then incubated on ice for 5 min

- Micro centrifuged at 13,000 rpm for 10 min and supernatant removed

- A QIAGEN tip 20 was equilibrated by adding 1ml of QBT buffer and the column was allowed to empty by gravitational flow

- The removed supernatant was added to the column and allowed to flow down

- Column was washed with 4 X 1 ml QC buffer

- DNA was precipitated with 0.7 vol of isopropanol at RT

- Micro centrifuged at 15,000 rpm for 20 min at 4°C and washed with 70% ethanol for 5 min

- Pellet was dried and dissolved in 10-20 µl of ddH$_2$O

2.2.2.3.3 Mini plasmid DNA preparation by Qiagen spin columns

- 5 ml O/N grown culture was centrifuged and the pellet was broken by adding 250 µl of P1 buffer, 250 µl of P2 buffer and tubes inverted gently

- 350 µl of N3 buffer was added and centrifuged for 10 min at maximum speed on a desktop centrifuge

- Supernatant was then added to the QIAprep column and micro centrifuged for 30-60sec

- Flow through was discarded and the spin column washed with 0.75ml of PE buffer

- Micro centrifuged for an additional min to remove residual wash buffer

- Column was placed on a clean micro centrifuge tube

- 50 µl of buffer EB or ddH$_2$O was added to the column, allowed to stand for a minute and micro centrifuged

2.2.2.3.4 Mini plasmid DNA preparation by alkaline lysis method

 (Adapted from Sambrook et al., 1989)

Alkaline lysis depends on a unique property of the plasmid DNA to rapidly anneal following denaturation. This allows the plasmid DNA to be separated from the

bacterial chromosomal DNA. This method was used to screen large recombinant colonies. The DNA thus obtained was successfully used for restriction digestion.

- 3 ml of the O/N grown cultures were micro centrifuged for 2 min
- The supernatant was discarded and pellet dissolved in 250 µl of ice cold Solution I
- 250 µl of Solution II was added and then allowed to incubate at RT for 5 min
- Then 350 µl of Solution III was added and incubated on ice for 5 min
- Micro centrifuged for 10 min and supernatant was transferred to a fresh tube
- 700 µl of isopropanol was added and precipitated for 30 min at -20ºC
- Micro centrifuged at 15000 rpm for 20 min at 4ºC
- Pellet was washed with 70 % ethanol for 5 min
- The pellet was dried and then dissolved in 30 µl ddH$_2$O

2.2.3 Photometric quantification of nucleic acid concentration

The nucleic acids (DNA and RNA) concentration and purity can be determined by Spectrophotometric analysis. The maximum absorbance for nucleic acid is measured at a wavelength of 260 nm. The concentration can be calculated by using the formula.

The concentration (µg/µl) = OD$_{260}$ X dilution factor X weight per OD

1 OD$_{260}$ corresponds to: 50 µg/µl double strand DNA

40 µg/µl single strand DNA or RNA

20 µg/µl oligonucleotides

The purity of the sample was evaluated by measuring the maximum absorbance of proteins at 280 nm.

The ratio of absorbance at (OD$_{260}$/ OD$_{280}$) should be within 1.8 to 2.

2.2.4 Enzymatic modifications of DNA

2.2.4.1 DNA digestion with restriction enzymes

Restriction endonucleases recognize specific sequences in the DNA and digest them. 1µg of plasmid DNA was digested in 1 unit of the enzyme with 1X concentration buffer at the specified temperature of the enzyme for 1½ - 2 hrs. The genomic DNA was digested in 5 units of the enzyme for 4 hrs.

Reaction Assay	1 µg DNA
	1U Restriction enzyme
	2 µl 10X buffer
	made up to 20µl with Ampuwa

Double Restriction digestions were performed by choosing compatible buffers for the selected enzymes. When necessary, restriction was terminated by heat inactivation at 65ºC for 20 min or by freezing the sample at - 20ºC. The restricted digested products were checked in an agarose gel electrophoresis (2.2.5.1).

2.2.4.2 DNA Ligation

DNA Ligases catalyze the formation of a phosphodiester bond between adjacent nucleotide with the hydrolysis of ATP to AMP and inorganic phosphate. T4 DNA Ligase was used in ligating the vector and the insert.

Reaction Assay	50 ng Vector DNA
	300 ng insert fragment
	heated at 65ºC for 5 min
	2 µl T4 DNA Ligase buffer (10X)
	1 µl T4 DNA Ligase (1U)
	made up to 20 µl with ampuwa

The ligation reaction was carried out at 16ºC for 2 ½ hrs.

2.2.4.3 DNA Labeling

2.2.4.3.1 DNA Labeling by Nick Translation

(Langer et al., 1981)

This is a non radioactive method of labeling oligonucleotides involving a hapten such as Biotin (Vitamin H), based on the ability of the DNase I to introduce random nicks in double stranded DNA at low enzyme concentration. After the DNA is nicked, the $5' \rightarrow 3'$ exonuclease activity of DNA polymerase I extend the nicks to gaps; then the polymerase replaces the excised nucleotides with labeled

ones (biotin labeled). The reaction produces optimal labeling after 120 min. The size of the probe after the reaction should be about 200 to 500 bp.

Reaction Assay for YACs	15 μg DNA
	20 μl nick translation buffer
	5.6 μl Biotin-16-dUTP (1mM)
	4.8 μl dNTP mix (1mM)
	6 μl DNase I*
	*1:1000 dilution from 1mg/ml stock
	4 μl DNA polymerase I (5U/μl)
	made up to 200 μl with Ampuwa
Reaction Assay for BACs	20-300 ng DNA
	4 μl nick translation buffer
	1.5 μl Biotin-16-dUTP (1mM)
	6 μl dNTP mix (1mM)
	1 μl DNase I
	0.8 μl DNA polymerase I (5U/μl)
	made up to 40 μl with Ampuwa

The reaction was incubated at 14ºC for 2 hrs.

2.2.4.3.2 DNA Labeling by Random priming

(Feinberg and Vogelstein, 1983)

This method is also a non radioactive labeling of oligonucleotides involving a hapten such as Digoxigenin (a steroid from *Digitalis purpurea*, foxglove), it is based on the random hybridization of a mixture of all possible hexanucleotides to the single stranded DNA that needs to be labeled. The complementary strand is synthesized from the 3' OH termini of the hexanucleotide primers using the 5'→3' polymerase activity of Klenow polymerase. The reaction produces optimal labeling at 37ºC for O/N. The size of the probe after the reaction should be about 200-1000 bp.

Reaction Assay	50 ng DNA
	10 μl ddH$_2$O
	10 min denatured at 95ºC
	5 min on ice
	2 μl Hexanucleotide (10X)
	2 μl Dig DNA labeling mix(10X)
	1 μl Klenow-Polymerase(2U)
	made up to 20 μl

The reaction was incubated at 37ºC for O/N.

2.2.4.3.3 Detection of labeled DNA by DOT-BLOT hybridization

The nick translated and random primed DNA was detected by DOT-BLOT.

Biotin has affinity to Streptavidin alkaline phosphate conjugate and Digoxigenin to anti Digoxigenin.

- 1 μl of the labeled DNA by Nick translation/ Random priming was serially diluted (1:1, 1:10, 1:100 and 1:1000) and spotted on a nylon membrane and allowed to dry

- Denatured at 120 mJoules for 1 min in a UV Stratalinker

- Incubated with 1:5000 Streptavidin alkaline phosphate conjugate and Buffer II for 20 min - Biotin labeling

- Incubated with 1:10,000 antidigoxigenin and buffer II for 20 min - Dig labeling

- 2 X 5 min washes in buffer I to remove excess probe

- 3 min wash in buffer III

- Color was detected by BCIP and NBT in buffer III (1:200) after incubating in the dark for 1 hr

- The labeling efficiency was visualized by the appearance of blue color dots on the membrane

2.2.5 DNA separation by Gel Electrophoresis

2.2.5.1 Agarose Gel electrophoresis

This method is widely used to separate DNA strands by size, and to estimate the size of the separated strand by comparing with the known fragments (DNA ladder).This is achieved by pulling negatively charged DNA molecules through an agarose matrix with an electric field. The concentration of the agarose depends on the size of the DNA fragments to be separated (Tab. 1).

- The agarose gels were prepared by heating agarose in 1X TAE or 0.5X TBE

- Gels were run at constant voltage between 30-90 volts. The loading buffers 6X dye a mixture of Xylene cyanol and Bromophenol blue was mixed with the sample. This dye runs about 5000 bp and 300 bp respectively (for 1% gels). The run can be estimated by these loading dye movements

- To determine the size of the DNA fragments, DNA standard ladders were used and DNA mass ladder was used to estimate the quantity of DNA sample

- The gels were stained in 0.001 % of Ethidium bromide for 30 min and viewed under UV light (366nm) and photographed.

Tab. 1: Optimal agarose concentrations for resolving DNA fragments (from Roche Molecular biochemical lab FAQS)

Agarose concentration %(w/v)	DNA size (kb)
0.4	2.5- 30
0.8	1- 15
1.0	0.5- 10
1.25	0.25- 5
1.5	0.25- 5
2.0	0.1- 2.5

2.2.5.2 Pulsed Field Gel Electrophoresis (PFGE)

(Schwartz and Cantor, 1984; Carle et al., 1986)

PFGE is used to separate very large DNA fragments up to several Mb. The principle behind it is applying the discontinuous electric field so that the DNA molecules are intermittently forced to change their conformation and direction of migration during their passage through the gel. The separation depends on many

factors; the pulsed time, temperature, running buffer and type and concentration of agarose.

- ▪ The YAC agar blocks were loaded in a 1 % low melting agarose gel and the wells were sealed with agarose

- ▪ The running buffer was 0.5X TAE

Conditions	Angle: 120ºC
	Gradient: 6V/cm
	Temperature: 14ºC
	Run time: 30 hrs
	Initial switch time: 18.75 s
	Final switch time: 2 min 58.46 s

The gels were stained with 0.001% Ethidium bromide and photographed.

2.2.6 DNA extraction from gels

The desired DNA fragments were excised from the gel after staining with 0.001% Ethidium bromide.

2.2.6.1 DNA extraction by centrifugation

The excised gel fragment was placed on a Whatman-1 filter paper in a 1.5 ml micro centrifuge tube.

- ▪ A small hole was made at the bottom of this micro centrifuge tube and was placed above another micro centrifuge tube

- ▪ Micro centrifuged at 12,000 rpm for 1 min

- ▪ The solution collected in the bottom tube contains the DNA

- ▪ DNA was concentrated by precipitating with 0.1 vol 3 M Sodium acetate, 2.5 vol ice cold ethanol for 30 min at -20ºC

- ▪ Micro centrifuged at 15,000 rpm for 20 min at 4ºC

- ▪ Pellet was washed with 70% ethanol for 5 min, dried and dissolved in 5-10 µl of TE or ddH$_2$O

2.2.6.2   DNA extraction from low melting point agarose (LMP)

The DNA extraction from LMP agarose and agarose gels was done using Promega Wizard® DNA Clean-Up system. The instructions were followed as in Technical Bulletin No 141, 10th/01.

- To the gel piece 1 ml of Wizard® DNA Clean-Up Resin was added and allowed to dissolve

- A mini column with a 1 ml syringe above it was placed on a micro centrifuge tube

- The resin was forced through the syringe

- Column was washed with 2 ml of 80 % isopropanol

- Micro centrifuged at 13,000 rpm for 2 min to dry the resin

- Column transferred on to a fresh tube and 50 µl of ddH$_2$O or TE buffer was added and then allowed to stand for a min

- Micro centrifuged for 20 sec and then DNA was collected

2.2.7   Southern blot hybridization

(Modified from Southern, 1975)

In this procedure, the target DNA is digested with one or more restriction endonucleases, size fractionated by agarose gel electrophoresis, denatured and transferred to a nitrocellulose or nylon membrane for hybridization.

2.2.7.1   Preparing the gel for transfer

- The PFGE gel was stained with Ethidium bromide and photographed with a scale

- Gel was destained for 15 min in ddH$_2$O

- Depurinated with 0.25 N HCl for 10 min

- The gel was washed with ddH$_2$O and denatured for 60 min with Denaturation solution while shaking moderately

- Washed with ddH$_2$O and neutralized for 15 min with Neutralization solution

2.2.7.2   Transfer of DNA onto nylon membrane

The DNA from the gel was transferred to immobilized membranes for hybridization.

2.2.7.2.1 Capillary transfer

- A nylon membrane and a Whatman-3 paper with correct dimensions of the gel were cut
- A glass plate with a Whatman-3 paper as supporting membrane was placed on a tray with 20X SSC as transfer buffer
- Gel was placed on the glass plate
- Nylon membrane was wetted in 2X SSC and placed on top of the gel
- Whatman-3 paper was placed above it with a stack of paper towels and weight
- This causes the DNA to move from the gel onto the membrane by capillary action
- Transfer was allowed to last for 16-20 hrs
- Membrane was washed with 2X SSC and then dried
- The DNA was cross linked to the membrane by exposing it to UV light 120 mjoules for 1 min
- Membrane was then stored in a plastic bag till further use

2.2.7.2.2 Pressure transfer

- The pressure transfer was performed using a pressure station at 35 mm pressure
- Whatman-3 paper was placed on a plate provided by the apparatus and nylon membrane above it and then the gel
- A sponge carrying the transfer buffer 20X SSC was placed above the set up
- Pressure was applied evenly on the gel by using suction
- Transfer was allowed to last O/N and membrane was washed in 2X SSC and cross linked with UV light and stored

2.2.8 Molecular hybridization of labeled DNA

2.2.8.1 Probe preparation

An isolated DNA molecule with a specific sequence that pairs with the appropriate sequence was used as a probe. These probes were labeled by Random priming (2.2.4.3.2) and used for Southern hybridization.

- To the labeled DNA 2.5 µl of 4 M lithium chloride, 1 µl glycogen (20 mg/ml), 2.5 vol ethanol was added and precipitated for 30 min at - 20ºC
- Micro centrifuged at 15,000 rpm at 4ºC for 20 min

- Pellet was washed in 70 % ethanol for 5 min, air dried and then dissolved in 50 µl of ddH$_2$O

- Probe was denatured at 95ºC for 10 min and kept on ice immediately for 5 min

2.2.8.2 Hybridization

- Nylon membrane was wetted in 2X SSC and rolled in an Hybridization tube with 10 ml of Rapid Hyb buffer

- Pre hybridized for 30 min at 62ºC in a hybridization oven

- Half the quantity of the buffer was discarded and probe added to 3-4 ml of the buffer and hybridized O/N at 62ºC

2.2.8.3 Membrane washes

- The hybridization solution was decanted and the excess probe washed in wash buffer(0.5X SSC/ 0.1 % SDS) at RT

- 3 X 10 min with wash buffer pre warmed at 65ºC

2.2.8.4 Detection of Digoxigenin labeled DNA by Enzyme Immuno assay

The YACs were detected by antibody conjugate for digoxigenin known as Anti-Digoxigenin-AP Fab fragments.

- The membrane was washed for 2 min in Buffer I

- Incubated in Buffer II for 15 min

- 30 min in 1: 10,000 Anti-Digoxigenin-AP Fab fragments and Buffer II

- 2 X 5 min washes in Buffer I

- Washed for 3 min in Buffer III

- 500 µl of ECF substrate was added on the membrane, sealed in a plastic bag and incubated for 4 hrs for color development

- The hybridization signals were detected with STORM

2.2.9 Subcloning

Subcloning is a basic procedure required to move inserts from one vector to another. This involves releasing and purifying the target insert from a parent vector, ligating this insert into a destination vector and transforming this ligation reaction into competent bacterial cells. Finally, the transformed cells are screened for the target vector to confirm successful transfer.

2.2.9.1 Dephosphorylation of the linearized vector

Removing the 5' phosphates of the linearized vector prevents self ligation of the vector.

<div style="border:1px solid black;">

Streamlined restriction digestion and dephosphorylation assay:

1 µg pBluescript SK+

2 µl Buffer A

1 µl Sac I

1 µl alkaline phosphatase (shrimp)

made up to 20 µl with ddH$_2$O

</div>

- The digest was performed at 37ºC for 15 min
- Heat inactivated at 65ºC for 20 min
- Purified it by using Promega Wizard® DNA Clean-Up system (2.2.6.2)

2.2.9.2 Insert DNA isolation

The BAC DNA was digested with selected restriction endonucleases for the isolation of the inserts.

- The restriction digestion was performed (2.2.4.1)
- The desired insert fragments were isolated from LMP agarose gels (2.2.6.2)

2.2.9.3 Competent cell preparation

Chemical competent cell preparation

- DH5α and JM 109 cells were grown till the OD$_{560}$ reached between 0.5-0.6
- Centrifuged at 4000 rpm for 10 min at 4ºC
- Pellet was dissolved in 50 mM Calcium chloride and incubated on ice for 15 min
- Centrifuged and dissolved pellet in glycerol and calcium chloride
- Aliquots stored at - 80ºC

Electro competent cell preparation

- The pellet was suspended in ice cold ddH_2O and centrifuged

- Washes repeated thrice with ice cold ddH_2O

- Pellet was resuspended in 10 % ice cold glycerol and centrifuged

- Pellet was then dissolved in 3 ml of 10 % ice cold glycerol

- Aliquots were finally stored at - 80ºC

2.2.9.4   Transformation

Electroporation

It is a method of exposing cells to high voltages in order to relax the selective permeability of the plasma membranes.

- 1 µl of the ligation DNA reaction mix (2.2.4.2) was added to 80 µl of electro competent cells

- Electroporated at 2500 volts

- Cells were transferred to 200 µl of SOC medium

- Allowed to grow at 37ºC for 1 hr by shaking at 200 rpm

- 100 µl cells were plated on LB- X-gal plates and incubated at 37ºC for O/N

Chemical Transformation

- 1 µl ligation DNA reaction mix was added to 100 µl of JM 109, and then incubated on ice for 30 min

- Heat shock was given at 42ºC for 90 sec

- Cold shock was then given on ice for 5 min

- 900 µl of SOC medium was added and incubated at 37ºC for 1 hr by shaking at 200 rpm

- 100 µl cells were plated on LB- X-gal plates and incubated at 37ºC for O/N

2.2.9.5   Blue/White Screening for Recombinants

The recombinants were screened by Blue/White selection. The *E.coli* cells JM 109 and DH5α were transformed for lac ZΔ. This mutation deletes a portion of the β-galactosidase gene with ω-fragment. The plasmid vector supplies this deleted portion or α-fragment. Once inside the bacterium both α and ω fragments are produced and appear as blue colonies on X-gal plates. Disruption of the reading frame due to the insert produces a non functional α fragment and hence produce white colonies. These are differentiated as white and blue colonies, hence termed blue white selection.

- The selected white colonies were cultured and mini preparation was performed
- The presence of the inserts was determined by restriction digestion analysis

2.2.10   DNA Amplification (PCR)

(Saiki et al., 1988)

Polymerase chain reaction (PCR) is an enzymatic method for exponential amplification of specific DNA *in vitro*. The three main steps involved are template denaturation, primer annealing and primer elongation. The heat stable *Taq* polymerase can withstand a temperature of 95ºC for a minimum of 1 hr.

2.2.10.1 Standard PCR

The standard PCR was performed in a 20 or 50 µl assay. Note that its annealing temperature depends on the melting temperatures of the primers used.

Standard PCR assay:	100 ng Template DNA
	5 µl PCR buffer (10X)
	3 µl Magnesium chloride (25mM)
	1 µl Primer 1 (10 pmol/µl)
	1 µl Primer 2 (10 pmol/µl)
	1 µl dNTPs (10 mM)
	0.5 µl *Taq* polymerase (5U/µl)
	made up to 50 µl with ddH$_2$O

Standard PCR Program:

Segment	Temperature	Time	cycles
Initial denaturation	95ºC	2 min	
Denaturation	95ºC	30 sec	
Annealing	50-70ºC	30 sec	30-35
Elongation	72ºC	4 min	

The PCR products were analyzed in an agarose gel electrophoresis.

2.2.10.2 *Alu* PCR

(Nelson et al., 1989)

Alu PCR is an exceptional type of PCR reaction in which one *Alu* primer and one vector primer is used, corresponding to a sequence which is chosen to be close to the end of the *Alu* repeat consensus sequence. This is a convenient method to end sequence YAC inserts. Here, YAC vector primers and *Alu* specific primers were used.

PCR Reaction Assay	
	5 µl YAC DNA (100 ng)
	2 µl Taq buffer (10X)
	1 µl vector primer (5 pmol/µl)
	1 µl alu primer (5 pmol/µl)
	5 µl dNTPs (1mM)
	0.75 U Taq polymerase (5 U/µl)
	made up to 20 µl with ddH$_2$O

Program:

Segment	Temperature	Time	cycles
Initial denaturation	95ºC	2 min	
Denaturation	95ºC	30 sec	
Annealing	55ºC	30 sec	35
Elongation	72ºC	3 min	

Insert end fragments were purified from the gel and sequenced.

2.2.10.3 Long range PCR

The long range PCR helps in amplifying fragments greater than 8 kb. Fermentas long PCR Enzyme mix was used, which contains a unique blend of high processive *Taq* DNA polymerase and second thermostable polymerase that exhibits 3' →5' exonuclease (proof reading) activity.

PCR Reaction Assay	3 µl DNA template (12ng)
	5 µl long PCR buffer (10X)
	1.5 µl Primer I (10 pmol/µl)
	1.5 µl Primer II (10 pmol/µl)
	2.5 µl dNTPs (10mM)
	2.5 U Enzyme mix (5 U/µl)
	made up to 50 µl with ddH$_2$O

Program:			
Segment	**Temperature**	**Time**	**cycles**
Initial denaturation	94ºC	2 min	1
Denaturation	95ºC	15 sec	
Annealing	60ºC	30 sec	20
Elongation	68ºC	9 min	
Denaturation	95ºC	15 sec	
Annealing	60ºC	30 sec	20
Elongation	68ºC	9 min	
		+10 sec/cycle	
Final elongation	68ºC	9 min	1

2.2.10.4 Colony PCR

To screen a large amount of recombinant colonies for the desired insert, single colony PCR was performed. This is a fast method to screen a large number of colonies.

- A single white colony was picked up with an yellow tip and dipped in 20 µl of ddH$_2$O

- Incubated at 95ºC for 10 min

- The cultures were then centrifuged for a minute at 13,000 rpm

- 3 µl was taken from this and used as a template for PCR

- Standard PCR conditions were used for the amplification of the cloned colonies with M-13 forward and reverse primers

2.2.10.5 RT (Reverse Transcription) PCR

This PCR is extremely useful in amplification of cDNA sequences corresponding to both abundant and rare transcripts, thereby providing an important source of DNA for mutation screening.

- 1µl of cDNA was used as a template from human testis and brain

- The cDNA was amplified with exon specific primer pairs by standard PCR assay conditions

2.2.11 Purification of PCR products

The PCR products should be purified to eliminate contaminants like oligonucleotides, enzymes and the presence of salts before cloning or sequencing. Montage® PCR purification units were used.

- 300 µl of 10 mM Tris pH 8.0 and 6 µl of the PCR product were added to the PCR unit

- Micro centrifuged at 2000 rpm for 30 min

- 20 µl of TE was added to the unit

- PCR unit was inverted in a clean centrifuge tube

- Micro centrifuged at 13,000 rpm for 20 sec

- DNA collected was further used.

2.2.12 Cloning of PCR products

pGEM®-T Vector systems (Promega, Technical manual Part# TM042)

The pGEM®-T Vector system (Fig. 5) is a convenient system for cloning of PCR products. The vectors are made by cutting pGEM Vectors with *Eco*RV and adding

a 3' terminal thyamine to both ends. These 3' -T overhangs at the insertion site greatly improve the efficiency of ligation of a PCR product into the plasmids by preventing re-circularization of the vector by providing a compatible overhang for PCR products generated by certain thermo stable polymerases. Recombinant *Taq* provides 3' A overhangs for the PCR products.

2.2.12.1 Ligation and Transformation of PCR products

Reaction assay			
	Standard	**+ve control**	**-ve control**
Rapid ligation buffer(2X)	5 µl	5 µl	5 µl
pGEM®-T Vector	1 µl	1 µl	1 µl
PCR product	5 µl	-	-
Control insert DNA (4 ng/µl)	-	2 µl	-
T4 DNA ligase (3 U/µl)	1 µl	1 µl	1 µl
Made upto 15 µl with ddH₂O			

- Ligated at 4ºC for O/N
- 2 µl of the ligated DNA was chemically transformed into 50 µl of DH5α competent cells (2.2.9.4.)
- Screening of the recombinants was done by colony PCR (2.2.10.4) and mini preparation (2.2.2.3.4)

2.2.13   DNA Sequencing

(Sanger et al., 1977)

The automated non radioactive DNA sequencing follows the enzymatic method with the principle of ddNTP incorporation. During DNA synthesis there is no bond formation between 3' position and the next nucleotide as the ddNTPs lacks the 3' OH group. Thus, the synthesis of new chain is terminated at this site.

2.2.13.1 Thermal Cycle Sequencing

The products formed by this method were used for sequencing. Only one primer is used with ddNTPs in the reaction mixture. This generates a series of different chain-terminated strands, each dependent on the position of the particular nucleotide base where the chain is being terminated. Advantage of this method is

that dsDNA can be used as starting material. The Big Dye®Terminator v.1.1 Cycle sequencing kit was used.

Sequence Reaction Assay:	10-50 ng of DNA
	2 µl of sequencing buffer (5X)
	0.2-0.5 µl Primer (5 p mol/µl)
	3 µl Premix
	made up to 20 µl ddH$_2$O

Program:

Segment	Temperature	Time	cycles
Initial denaturation	96ºC	1 min	
Denaturation	96ºC	10 sec	
Annealing	50ºC	5 sec	25
Elongation	60ºC	4 min	

- The PCR products obtained by thermal sequencing were precipitated by adding 0.1 vol 3 M Sodium acetate, 0.025 vol 0.5M EDTA pH 8.0, 0.025 vol glycogen and 2.5 vol of ice cold ethanol at - 20ºC for 30 min

- Micro centrifuged at 15,000 rpm at 4ºC for 20 min

- Pellet was washed in 70% ethanol for 5 min, dried and then stored at -20ºC

- The samples were denatured and separated by gel electrophoresis

- The products were analysed on an ABI Prism® Genetic analyzer at Zentrum fuer Medizinische Grundlagenforschung(ZMG)

2.2.14   Fluorescence *in situ* hybridization (FISH) on metaphase chromosomes

(Pinkel et al., 1986; Trask et al., 1988)

FISH is a technique to directly visualize DNA sequences on metaphase chromosomes and interphase cells. Indirect FISH uses hapten-labeled DNA probes in combination with immunocytochemistry. Both probe and target DNA were denatured and allowed to reanneal during hybridization. After post hybridization washing and immunocytochemical detection of the *in situ* bound hapten molecules, a specific fluorescence signal is produced at the hybridization site which can be viewed through the fluorescence microscope.

2.2.14.1   Single hybridization FISH

2.2.14.1.1   Chromosomal *in situ* suppression (CISS)

(Lichter et al., 1988)

This is a form of Competition hybridization before the main hybridization; the probe is mixed with a large excess of unlabelled total genomic DNA and denatured there by saturating the repetitive elements in the probe, so that they do not mask the signals generated by the unique sequences.

- The biotin labeled DNA (2.2.4.3.1) was precipitated with 0.1 volume of 3M Sodium acetate, 2.5 volume of ice cold ethanol, 20 μl/ 9 μl of Cot 1 Human DNA for YACs/BACs and 5 μl/ 6μl of Salmon sperm DNA for YACs/BACs at - 20ºC for 30 min

- Micro centrifuged at 15,000 rpm at 4ºC for 20 min

- Pellet was washed in 70% ethanol for 5 min and then dried at 37ºC

- 5 μl of cold formamide was added and then incubated at 37ºC for 30 min

- 5 μl of master mix was added, mixed and then denatured at 95ºC for 5 min

- The probe was incubated on ice for 5 min and then used for hybridization

2.2.14.1.2 Pepsin pretreatment of slides

Pepsin treatments help for a better FISH in three ways. First, to remove extraneous RNA and proteins which bind to the probe and detection reagents, increasing background. Second, to enable access of the probes to the DNA-permeabilizing the target material and lastly, to fix the preparations so that chromosomes and nuclei are not lost from the slide during the procedure.

- The slides were incubated in 0.005% of pepsin in 0.01N HCl at 37ºC for 20 min

- Washed in 1X PBS at RT for 5 min

- Fixed for 5 min in Post fixative solution at RT

- Washed again in 1X PBS at RT for 5 min

- Air dried and dehydrated in 70 %, 80 %, and 90 % and twice in 99 % ethanol series at RT for 1 min each and again air dried

2.2.14.1.3 Denaturation of chromosomes

To allow hybridization of the labeled probe DNA, the DNA target on the slide must be denatured so as to make it single-stranded. Using high temperatures in formamide is a frequently employed technique to denature chromosomes.

- Slides were denatured in Denaturation solution for 3 min and then immediately dehydrated in ice cold alcohol series (70 %, 80 %, 90 %, 2 X 99 %) for 1 min each

- Air dried

2.2.14.1.4 Hybridization

- 10 µl of the prepared probe was hybridized on to the slide and covered with a 22 X 22 mm cover slip and sealed with rubber gum

- Slides were hybridized in a humid chamber and incubated O/N at 37ºC

2.2.14.1.5 Post Hybridization washes

The post hybridization washes removes the hybridization mixture and unbound probe. The three variables to be considered are the composition of the solutions, washing temperature and the washing time.

- Slides were washed thrice in 50 % formamide in 2X SSC, pH 7 at 47ºC for BACs and 51ºC for YACs

- 5 X 2 min in 2X SSC washes

- 3 X 5 min in 0.1X SSC washes at 66ºC

- 2 X 5 min in 4X SSC, 0.1% Tween 20 washes

2.2.14.1.6 Detection and amplification

The biotin present in the probe has an affinity to the primary antibody FITC and a secondary antibody bound to biotin- BAAD was attached to it. The signals were enhanced by adding the primary antibody again to the primary and secondary antibody complex.

- 500 µl of blocking solution was added on the slides and incubated at RT for 10 min

- 100 µl primary antibody FITC was added and incubated at 37ºC for 30 min

- 5 X 5 min 4X SSC, 0.1% Tween 20 washes

- 100 µl of second antibody BAAD was added and incubated at 37ºC for 30 min
- 5 X 5 min 4 X SSC, 0.1% Tween 20 washes
- 100 µl of primary antibody FITC was added again and incubated at 37ºC for 30 min
- 5 X 5 min 4 X SSC, 0.1% Tween 20 washes
- Rinsed in ddH$_2$O and air dried

2.2.14.1.7 Counter staining the chromosomes

Counter staining reagents such as DAPI (blue) and/or Propidium iodide (PI, orange-red) had been used along with the antifade solution, which prevents the fading of signals when viewed.

- 27 µl of Vectashield antifade mounting medium, 4µl of DAPI and 4µl of PI were added, covered with a 22 X 40 mm cover slip and sealed with nail polish
- Slides stored at 4ºC

2.2.14.1.8 Microscope analysis and Digital imaging

Images were taken with a Zeiss Axioplan 2 fluorescence microscope equipped with a CCD (charged coupled device) camera. Switching between different fluorochromes was achieved by changing filter sets. The spectral properties of fluorophores are shown in Table 2.

Tab. 2: Spectral properties of fluorophores used in FISH analysis (adapted from Non radioactive in situ Hybridization Application manual, 2nd edition, Boehringer Mannheim).

Fluorophores	color	Excitation max. (nm)	Emission max. (nm)
Fluorescein	Green	494	523
CY3	Red	552	565
AMCA	Blue	399	446

2.2.14.2 Double hybridization FISH

This method involves the hybridization of two different DNAs labeled with biotin on a single slide.

- Two BACs or a YAC and a BAC were biotin labeled separately and 5 μl of both the probes were mixed and hybridized on a slide

- All the other steps were similar as in (2.2.14.1)

2.2.14.3 FISH with Restriction fragments

The DNA is fragmented to the desired size with selected restriction enzymes. These fragments were then directly labeled with biotin and used as probes for FISH. This method is convenient for using bigger fragments.

2.2.14.4 Whole chromosome paint

(Licher et al., 1988; Pinkel et al., 1988)

The whole chromosome paint probes are a mixture of DNA sequences specific to the chromosome that are directly labeled with any one fluorophore. Usually, unlabeled blocking DNA is included in the mixture to suppress the more highly repetitive sequences that are common to many different chromosomes. When hybridized and visualized, these probes appear to "paint" the entire chromosome. This allows chromosome enumeration, as well as identifying some chromosomal structural rearrangements and numerical changes, in metaphase cells. In this method commercially available whole chromosome paint from Star FISH was used to paint the chromosome 3. The procedure is followed as per the instructions provided by the manufacturer.

2.2.14.4.1 WCP Probe preparation and Hybridization

- 5 μl of the paint probe and 5 μl of hybridization buffer was aliquoted in an micro centrifuge tube and incubated at 37ºC for 10 min

- Denatured at 72ºC for 5 min and re incubated at 37ºC for 1 hour

- Probe was then hybridized at 37ºC for O/N

2.2.14.4.2 WCP Post hybridization washes

- Slides were washed 2 X 5 min in 50% Formamide in 2X SSC, pH 7 at 42ºC

- 2 X 5 min in 0.1X SSC at 65ºC

- 1 X 5 min in 4X SSC, 0.1% Tween 20 washes

- Detection similar to (2.2.14.1.6)

- 3 X 5 min 4X SSC, 0.1% Tween 20 washes

- Slides were mounted and visualized

2.2.15 *In silico* sequence analysis

All *in silico* information required in this study are taken from the following databases. Most of them can be easily accessed from the public domains.

2.2.15.1 Databases

- Databank from National Center for Biotechnology Information (NCBI)

  http://www.ncbi.nlm.nih.gov

- Databank from Wellcome Trust Sanger Institute (ensembl)

  http://www.ensembl.org

- The UCSC Genome Browser

  http://genome.cse.ucsc.edu

- The Human Genome Segmental Duplication Database

  http://projects.tcag.ca/humandup/

  http://humanparalogy.gs.washington.edu

2.2.15.2 Software for sequence analysis

- BLAST

  http://www.ncbi.nlm.nih.gov/BLAST/

  The BLAST was used to search sequence homologies in the database with in the whole human genome or other species.

- Repeat Masker

  http://repeatmasker.org/cgi-bin/WEBRepeatMasker

- GrailEXP (also to analyze repeats)

  http://compbio.ornl.gov/grailexp/

- ORF Finder (Open reading Frame)

  http://www.ncbi.nlm.nih.gov/projects/gorf/orfig.cgi

2.2.15.3 Software for primer designing

- Primer3 Output

  http://frodo.wi.mit.edu/cgi-bin/primer3/primer3-www.cgi

- GENERUNNER (version 3.01, Hastings software)

2.2.15.4 Software for Imaging from Metasystems

- Ikaros for metaphase chromosomes

  ISIS (*FISH* Imaging System)

2.2.15.5 Software for gel documentation

- E.A.S.Y win from Herolabs

3   Results

The proband as well as her mother and grand mother with an inversion chromosome 3 and short stature were studied in this project to characterize the breakpoints. Cytogenetic analysis had revealed a familial pericentric inversion 3, being heterozygous in the proband, her mother and maternal grand mother. Initially, the karyotype was assigned as inv(3)(p23;q25q26). In order to characterize the breakpoints FISH experiments were performed using YACs and BACs.

3.1   WCP (Whole chromosome paint) FISH

To rule out the involvement of other chromosomes in this rearrangement WCP FISH was performed with paint 3. Though small aberrations may not be detected it helped to confirm that the rearrangement in this study is only involved with chromosome 3. Both the normal chromosome and the inverted chromosomes were painted without any gaps in the sequences except the centromeric regions suggesting no gross involvement of other chromosomes (Fig. 8).
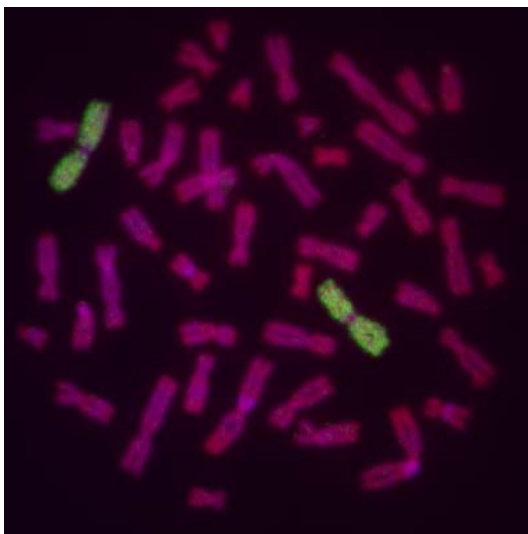

Fig. 8: Whole chromosome paint FISH on lymphocyte metaphase spreads of the proband.

The probe was labelled by FITC (green signals). The signals were seen on both the normal and inverted chromosome 3.

3.2   *In-silico* selection of YACs in the inversion breakpoint regions

All the genomic sequence data in this study was retracted from NCBI. Initially five q specific YAC clones were selected by *in silico* analysis using Map Viewer of the human genome database (http://www.ncbi.nlm.nih.gov/geneview). The selection was based on the tentative location of the breakpoint regions assigned by conventional cytogenetics. These 5 YACs (Tab. 3) covered a distance of 11 Mb around the 3q25 region.

3.3   YAC insert size determination by PFGE

All the YACs obtained were initially determined for the insert size by PFGE. The PFGE gel was run along with a yeast chromosome marker to differentiate the extra band present in the clone. The gel was photographed with a scale (Fig. 9A). Southern hybridization was performed with the PFGE gel and labelled pBR322 DNA was used as a probe to detect the YAC inserts (Fig. 9B). The bands detected by the Southern hybridization were measured and compared with the graduated gel photograph. The sizes were then determined with the band position of PFGE yeast chromosome marker. Four YACs showed a band corresponding to a size between 1900 and 1640 kb. YAC clone 943H12 showed an insert size of 1100 kb. First positive control showed a size around 350 kb and the second positive control between 1900 and 1640 kb (Tab. 3).
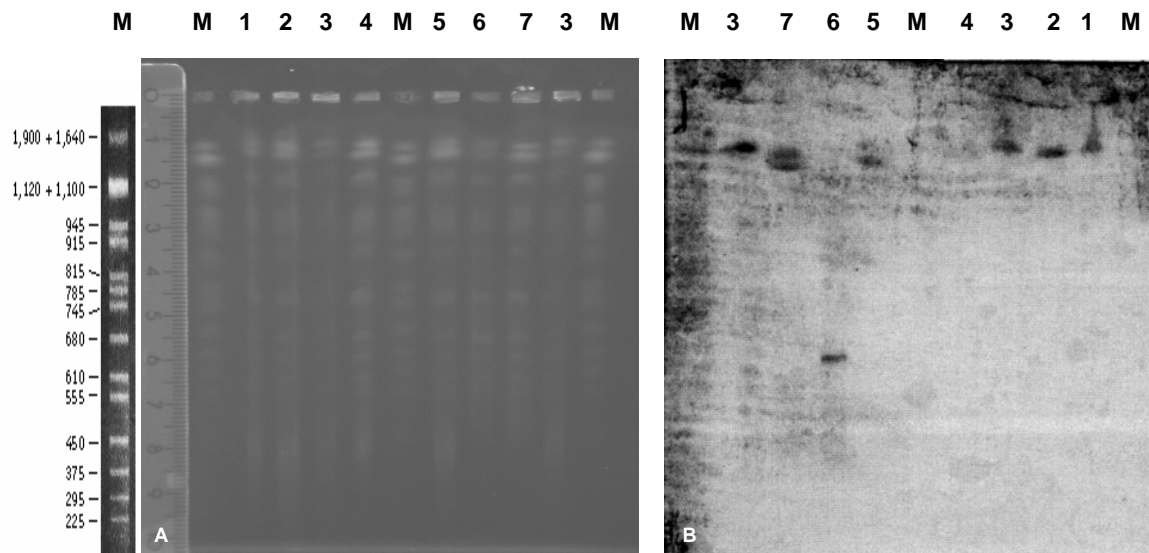

Fig. 9: PFGE gel and the blot. M: PFGE marker. **A:** PFGE gel Lane 1 is YAC clone 891E01, 2 is 855E05, 3 is 938F10, 4 is 806G12, 5 is 943H12, 6 and 7 are positive controls. **B:** The Southern hybridized blot showing the sizes of the inserts.

Tab. 3: The insert sizes of five YACs determined by PFGE.

The details of the five YACs showing the respective STS markers, position on the chromosome, its band position and the size of the inserts by PFGE were also shown in comparison with the PFGE results. NB refers to "no band". NA refers to "not available".

3q specific YACs	STS markers	Start position on chromosome (bp)	Band position	Insert size (RZPD kb)	Insert Size (PFGE kb)
CEPHy904E01891D1 (891E01)	D3S3692	160079350	3q25.33	1350	1900+ 1640
CEPHy904E05855D1 (855E05)	D3S3419 WI-3118	159635409 158896633	3q25.32	1430	1900+ 1640
CEPHy904F10938D1 (938F10)	D3S3360 WI-7070	158896633 158643463	3q25.32	1590	1900+ 1640
CEPHy904G12806D1 (806G12)	WI-7070 WI-3118	158643463 158896633	3q25.32	1580	NB
CEPHy904H12943D1 (943H12)	TO3514 IB398	158348000 158643463	3q25.32	NA	1120+1100

3.4   Double hybridization FISH with YACs and BACs

The human chromosome 3 is metacentric and the inversion studied in this case is also equidistant from the telomeres. To overcome the problem of arm identity double hybridization FISH was standardized. Four terminal BACs were selected two on either side of the chromosome 3 as reference signal. The BACs RP11-11K22 and RP11-12L14 were 3q specific and confined to 3q28 and 3q27-28 region and the BACs RP11-451A20 and RP11-91K04 were 3p terminal BACs, confined to 3p25 and 3p25-3p26 regions. Most of the double hybridizations were performed by using p specific terminal BACs. The possible occurrence of the FISH signals on both normal and inverted chromosome using the terminal p and q BACs with the respective q or p YACs/BACs are depicted in Figure 10.
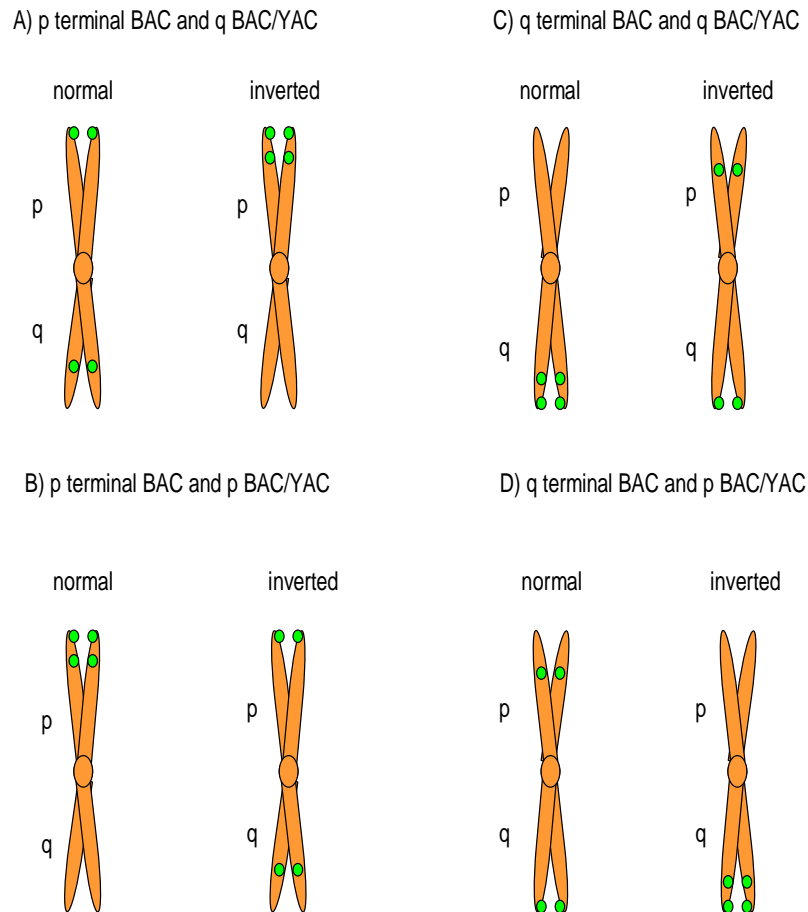
Fig. 10: Pictorial representation of the double hybridization FISH signals with different combination of YAC and BACs.

The chromosomes are shown in orange colours and the FISH signals with BAC/YACs are shown as green circles. **A:** The combination of p terminal BAC and q BAC/YAC showing terminal signals on p arm and signals on q arm of normal chromosome and terminal signals on p arm and signals on p arm on inverted chromosome. **B:** p terminal BAC and p BAC/YAC showing terminal signals on p arm and also signals on p arm on normal chromosome and the terminal signals on p arm and signals on q arm of inverted chromosome. **C:** q terminal BAC and q BAC/YAC showing terminal signals on q arm and signals on q arm on normal chromosome and terminal signals on q arm and signals on p arm of inverted chromosome. **D:** q terminal BAC and p BAC/YAC showing terminal signals on q arm and signals on p arm of normal chromosome and terminal signals on q arm and signals on q arm of inverted chromosome.

Initially, two YACs were selected for mapping 3q segments, one the proximal YAC (807G06) and other the most distal YAC (886H03). The BAC clone RP11-91K04 was taken as a reference on p arm. Double hybridization FISH results showed that the FISH signals of the proximal YAC for 3q arm were inverted and found on the p arm and that of the distal YAC signals were not inverted and remained on the q arm of the inverted chromosome (Fig. 11). This confirmed that the critical region of the 3q breakpoint was between these two regions and thus the double hybridizations were standardized for further FISH experiments.
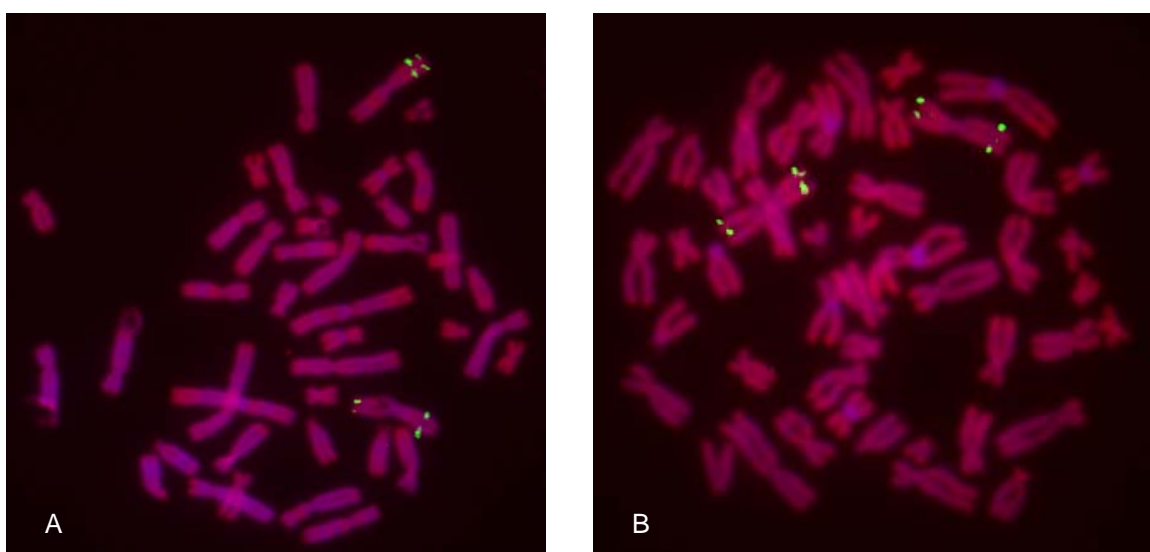

Fig. 11: Double hybridization FISH with proximal and distal YACs of 3q and terminal BAC for 3p.

The BAC RP11-91K04 signals are seen on the p terminal regions on both the metaphases. **A:** YAC clone 807G06 showing the signals on the p arm of the inverted chromosome and on q arm of the normal chromosome. **B:** Distal YAC clone 886H03 showing signals on the q arm on inverted as well as on normal chromosome.

3.5   3q inversion breakpoint region- Delineation by FISH with YACs

The 5 YACs mentioned in Table 3 were used as probes for FISH analysis. Signals were obtained on the metaphase chromosomes of the proband slides. All the YACs showed signals on chromosome 3q regions but YAC clone 943H12 showed chimerism, and the signals were found on chromosomes 1, 2, 3 and 16 (data not shown).

Further knowing the position of the signals with the above five YACs, 9 more YACs were obtained spanning about 86 Mb towards the distal end. The details of all the YACs are shown in (Tab. 4). FISH with YAC clone 889G07 (1610 Kb) on proband metaphase chromosomes showed split signals (Fig. 12) on both p and q region. Split signals indicate that the target sequence carries the inversion breakpoint region. Thus, the breakpoint was localized to chromosomal band region 3q26.1.
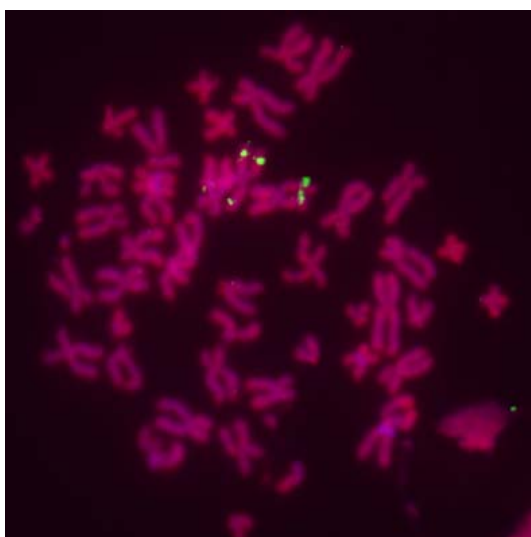

Fig. 12: FISH with YAC clone 889G07 on lymphocyte metaphase spreads of the proband.

The probe is detected by FITC (green signals). The split signals are seen on both the arms of the inverted chromosome 3. The chromosomes were counterstained with PI (red) and DAPI (blue) and the FISH signals and the DAPI banding pattern was merged for figure configuration. Only a pair of signals was seen on q region on normal chromosome.

Tab. 4: The details of 9 distal YACs for 3q region.

The YAC clone which showed a split signal on the metaphase chromosomes of the proband is shown in red.

3q specific YACs	STS markers	Start position on chromosome (bp)	Band position	Insert size RZPD (kb)
CEPHy904G06807D1 (807G06)	WI-9752 WI-3176	160784534 160753011	3q25.33	1020
CEPHy904H01954D3 (954H01)	WI-6917 WI-9638	161700650 162285467	3q26.1	870
CEPHy904B09845D3 (845B09)	D3S3523	163812224	3q26.1	710
CEPHy904G07889D3 (889G07)	D3S1258 D3S3712	167408075 168452294	3q26.1	1610
CEPHy904F11783D3 (783F11)	D3S3622 D3S3682	169346588 168621288	3q26.2-1	900
CEPHy904C09858D3 (858C09)	D3S3622 WI-9940	169346588 169896441	3q26.2	2200
CEPHy904A03750D3 (750A03)	WI-6894 WI-3218	170515576 173753346	3q26.31	1280
CEPHy904B11956D3 (956B11)	D3S1564 WI-3847	171671509 171857078	3q26.2	1560
CEPHy904H03886D3 (886H03)	D3S3511 WI-1231	178878254 178659317	3q26.32	1090

.

3.6   YAC end sequencing by *Alu*-PCR

YAC end sequencing and an alternative approach to identify the overlapping BAC clones was achieved by *Alu*-PCR using *Alu* primers and YAC-vector primers with YAC clone 889G07 as a template.

Alu PCR and *Alu*-Vector PCR

The ability to isolate the YAC insert or a portion of it is an important feature in analyzing the clone and confirming it to a specific chromosomal band position. The amplification of the fragments from the YAC clone 889G07 (1610 kb) insert was carried out using 7 Alu primers (Alu, Alu3, Alu13, Alu14, Alu153, Alu154, and Alu454) and 2 YAC vector primers. Two primers adjacent to the insert of the pYAC4 sequences Ura1 and Trp1 were designed and used as YAC-Vector primers. For the amplification of the insert both the primers were used together. By using same *Alu* primer inter-*Alu* fragments were generated and with one *Alu* primer and YAC vector primer *Alu*-Vector products were generated. The amplified PCR products (Fig. 13) were analyzed by comparing the inter-*Alu* and *Alu*-Vector products. The details of the inter-*Alu* and *Alu*-Vector products combinations are shown in Tab. 5. The amplified *Alu*-Vector products were isolated from the gel and sequenced. Out of the 8 fragments formed 6 could be sequenced. Two of the eight fragments, a 2,690 bp Ura-Alu3 fragment and a 300 bp Trp-Alu451 product did not show any sequencing results.

Tab. 5: The details of inter-*Alu* and *Alu*-Vector combinations along with the amplified product size.

NP refers to "no product formed".

Inter-*Alu* combinations	*Alu*-Vector combinations	Amplified products/ size(bp)
Alu3-Alu3	Trp-Alu3	NP
	Ura-Alu3	881
		2690
Alu154-Alu154	Trp-Alu154	6000
	Ura-Alu154	NP
Alu13-Alu13	Trp-Alu13	350
		3280
	Ura-Alu13	NP
Alu14-Alu14	Trp-Alu14	NP
	Ura-Alu14	400
Alu451-Alu451	Trp-Alu451	300
	Ura-Alu451	900
Alu153-Alu153	Trp-Alu153	NP
	Ura-Alu153	NP

3.6.1   BLAST search of the six *Alu*-Vector sequences

Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) from the NCBI database was used to check our results for the homologous sequences in the human genome. Each sequence was blasted separately and analyzed. Both ends of the YAC insert were sequenced by the Trp and Ura primers combination products. So further, with the 350 bp Trp-Alu13 and 400 bp Ura-Alu14 sequences two BAC clones were identified. RP11-450H5 at the Trp end of the YAC insert and RP11-190F16 at the Ura end. Both these clones were present on the 3q26.1 region.
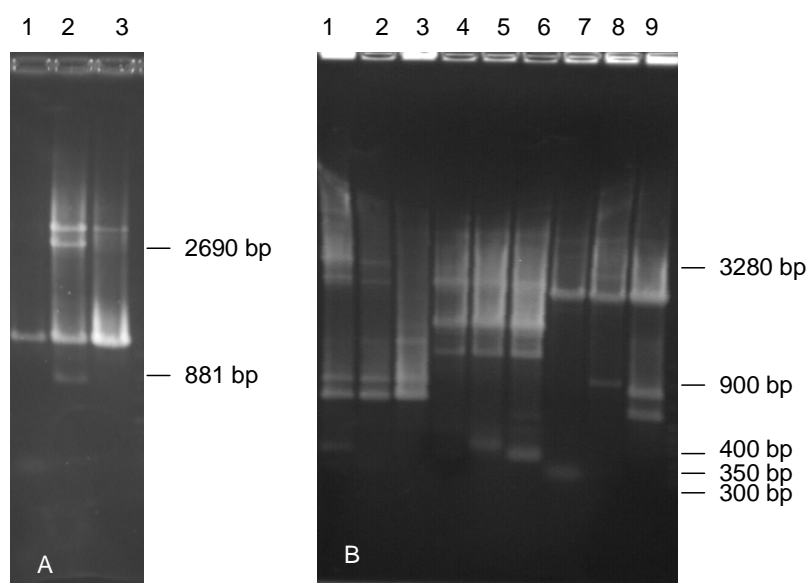
Fig. 13: The gel showing the inter-Alu and Alu-Vector products.

A: Lane 1: Trp-Alu3, Lane 2: Ura-Alu3, Lane 3: Alu3-Alu3 **B**: Lane 1: Trp-Alu13, Lane 2: Ura-Alu13, Lane 3: Alu13-Alu13, Lane 4: Trp-Alu14, Lane 5: Ura-Alu14, Lane 6: Alu14-Alu14, Lane 7: Trp- Alu451, Lane 8: Ura-Alu451 and Lane 9: Alu451-Alu451.

3.7   Initial selection of BACs

The 3q breakpoint region was defined to a 1610 kb region with the YAC clone 889G07 by FISH and sequencing the ends (see 3.6). In order to further narrow down the breakpoint region, the whole 1610 kb of the YAC was considered and randomly five BACs were selected from the database. The BACs selected and their band positions are shown in Tab. 6. The double hybridization FISH experiments were performed with all the five BACs as probes along with 3p terminal BACs RP11-91K04 and RP11-451A20. The double hybridizations helped in identifying the presence of the signals on p or q arm. FISH with four BACs RP11-452B04, RP11-265C03, RP11-451I20 and RP11-26K10 showed signals on the q arm of the inverted chromosome i.e., non-inverted signals, but the BAC RP11-636O7 showed signals on the p arm of the inverted chromosome i.e., inverted signals (Fig. 15). Thus, the breakpoint was found to be anchored between these

two BACs RP11-636O7 and RP11-451I20. The distance narrowed down was ~411kb, the distance between these two BACs (Fig. 14).

Tab. 6: The status of the five random BACs covering the sequence of the YAC clones 889G07.

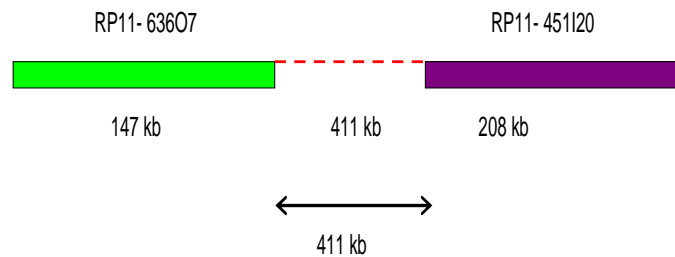| BAC clones	Region	Signal position	Status
RP11-636O7	3q26.1	p arm	inverted
RP11-451I20	3q26.1	q arm	non-inverted
RP11-265C03	3q26.1	q arm	non-inverted
RP11-452B04	3q26.1	q arm	non-inverted
RP11-26K10	3q26.1	q arm	non-inverted

Fig. 14: **The schematic representation of the two adjacent BACs which showed inverted (green) and non-inverted (violet) signals with the breakpoint within a distance of ~411 kb.**

The BACs are shown as blocks with the names above and the insert sizes below. The gap is shown as a red interrupted line and the line with double arrowheads shows the total narrowed down breakpoint anchoring region.
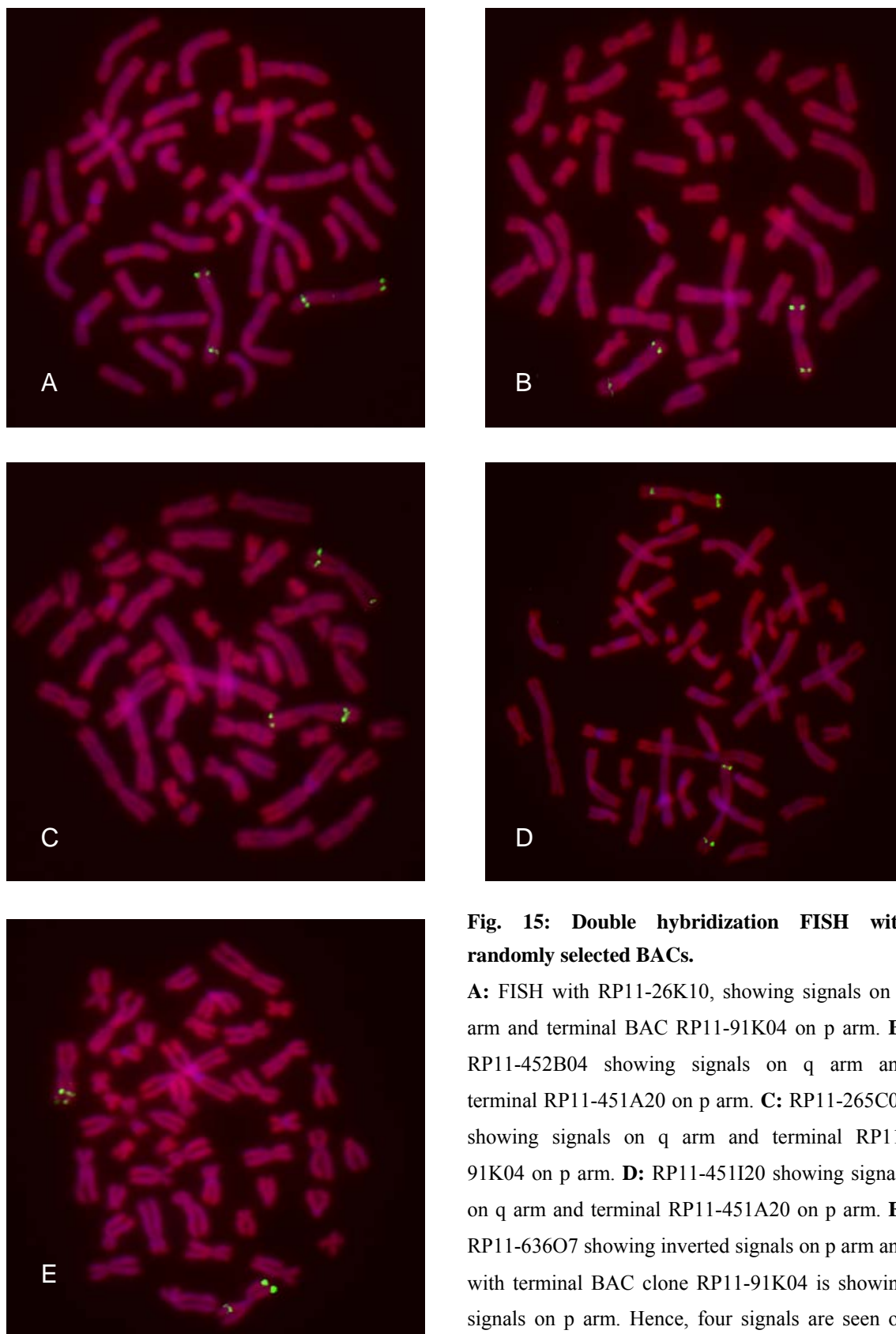
Fig. 15: Double hybridization FISH with randomly selected BACs.

A: FISH with RP11-26K10, showing signals on q arm and terminal BAC RP11-91K04 on p arm. **B:** RP11-452B04 showing signals on q arm and terminal RP11-451A20 on p arm. **C:** RP11-265C03 showing signals on q arm and terminal RP11-91K04 on p arm. **D:** RP11-451I20 showing signals on q arm and terminal RP11-451A20 on p arm. **E:** RP11-636O7 showing inverted signals on p arm and with terminal BAC clone RP11-91K04 is showing signals on p arm. Hence, four signals are seen on the inverted chromosome of the patient.

3.8   BAC contig formation

The 3q breakpoint region was thus narrowed down to a region of ~411 kb. The next step was to analyze this ~411 kb region, so a series of 5 overlapping BAC clones (RP11-12N13, RP11-480L9, RP11-358B20, RP11-280M5, RP11-775G8) were selected as a BAC contig (Fig. 18). FISH with BAC clone RP11-12N13 showed a split signal refining the breakpoint region to ~174kb (Fig. 16A). Double hybridization FISH with the proximal adjacent BAC clone RP11-480L9 and terminal clone RP11-91K04 showed the signals on p region (inverted) (Fig. 16B). The proximal adjacent BAC clone RP11-480L9 overlapped a region of 95 kb with RP11-12N13. Thus, combining the overlapping region and FISH results the breakpoint region was now narrowed down to ~79 kb (Fig. 17). The detailed physical map of the 3q breakpoint region covering a distance of 1.61 Mb is shown in Fig. 18.
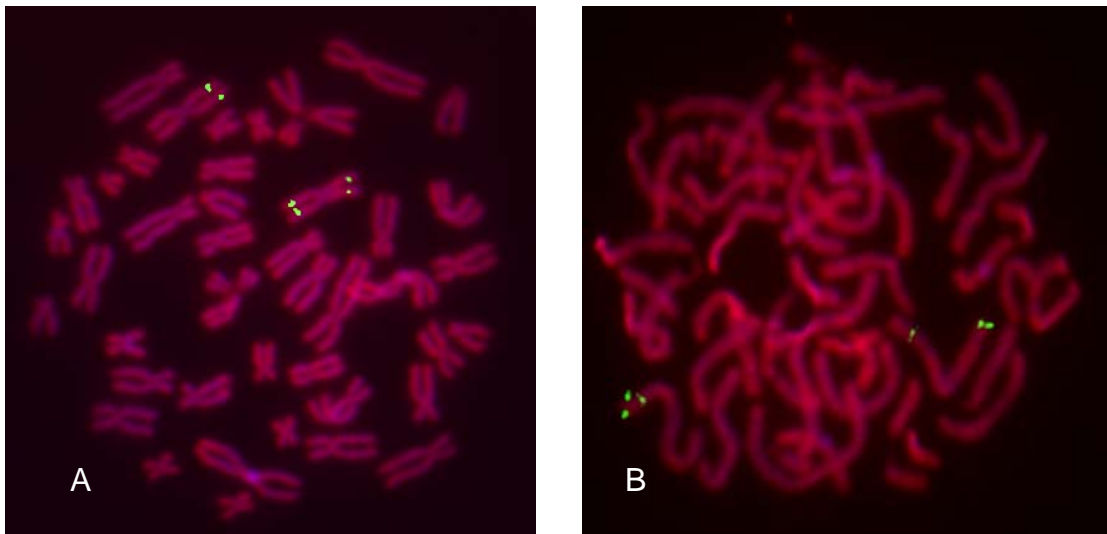

Fig. 16: FISH with BACs of a contig.

A: FISH only with RP11-12N13 showing split signals on both the arms of the inverted chromosome. **B:** Double hybridization FISH with proximal adjacent BAC clone RP11-480L9 and terminal RP11-91K04 showing the signals on the p arm of the inverted chromosome and on q arm of the normal chromosome.
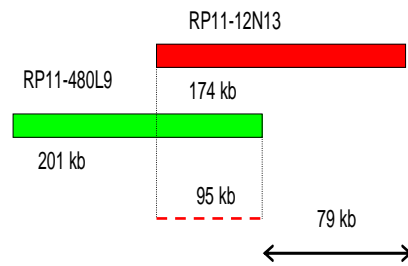
Fig. 17: Partial schematic representation of the two BAC clones RP11-12N13 and RP11-480L09 from the BAC contig.

The BACs are shown with the name above it and the size below. The overlapping distance of the two BACs is shown by a red interrupted line covering a distance of 95 kb. The black double arrow head line represents the region narrowed down or the breakpoint anchoring region.
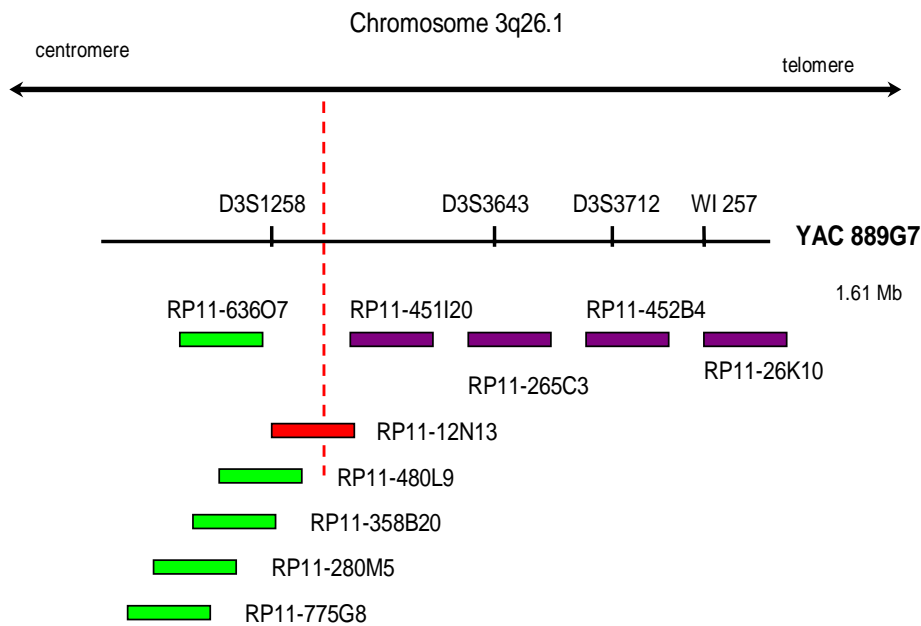
Fig. 18: A detailed 1.61 Mb physical map of the 3q26.1 inversion breakpoint region of the proband.
The YAC clone 889G07 carrying 4 STS markers (indicated by vertical lines). BAC clone RP11-636O7 showed signals on p arm (inverted) and BAC clone RP11-451I20 showed signals on q arm (non-inverted). The interval between these two clones was further tested with 5 additional BAC clones in order to narrow down the breakpoint region even further. The position of the breakpoint is shown in vertical red line. The BAC clone indicated in red encompasses the breakpoint. All the BAC clones in green showed signals on p arm and the BAC clones in violet showed signals on q arm.

3.9   Two precise approaches

To further narrow down the breakpoint region two different approaches were followed: subcloning, use of Restriction enzymes and Long range PCR. Subcloning was used to move inserts from one vector to another and restriction enzymes were selected to overcome the difficulties in cloning large fragments. To amplify larger genomic DNA products Long range PCR was used. Both these methods were used with the breakpoint spanning BACs.

3.9.1 Subcloning

In order to define the breakpoint region more precisely, a few fragments of the breakpoint anchoring BAC RP11-12N13 were subcloned. The 100 kb distal region of the BAC insert was targeted to be analyzed in detail. To obtain the required fragments *SacI* restriction enzyme was selected (Fig. 19). Five fragments 1.2 kb, 2.7 kb, 2.8 kb, 4.5 kb and 6.9 kb were cloned into a pBluescript-SK plasmid vector (Fig. 20). The positive clones were confirmed by sequencing. Two clones of insert sizes 6.9 kb and 4.5 kb were used as probes for FISH experiments. Double hybridization FISH with p terminal BAC clone RP11-91K04 and both 6.9 kb (Fig. 21) and 4.5 kb clones showed signals on 3p region (inverted) indicating that the breakpoint was distal to this region (Fig. 19). A 17 kb restricted fragment adjacent and distal to the 4.5 kb region was gel isolated and used as a probe for FISH (data not shown). FISH with this 17 kb fragment also showed the signals on p arm (inverted) suggesting the presence of the breakpoint distal to this region. The region remained to be analyzed was approximately 53 kb.
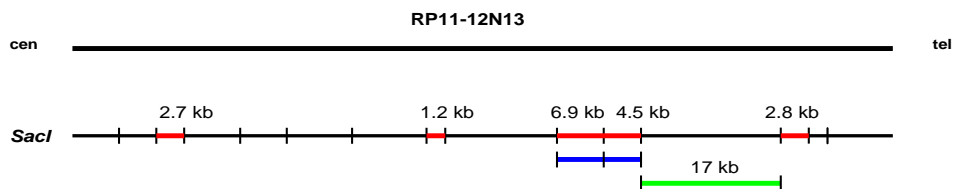
Fig. 19: Schematic representation of the subcloned inserts of BAC RP11-12N13. The upper line represents the insert of BAC RP11-12N13. The vertical lines on the lower line represent *SacI* restriction sites of RP11-12N13. The fragments shown in red were cloned and the fragments shown in blue were used as probes for FISH experiments. The 17 kb fragment shown in green is restricted and directly labeled to be used as a probe for FISH (the fragments are not drawn to scale).
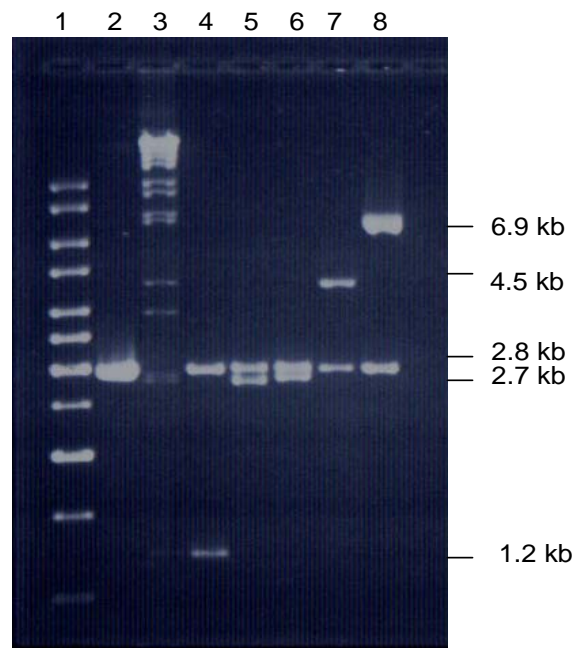
Fig. 20: Gel picture showing the recombinants of subcloning.

Lane 1: 1 kb marker, Lane 2: pBluescript SK vector of 3 kb size, Lane 3: BAC 12N13 restricted with *SacI*, Lane 4: 1.2 kb insert clone Lane 5: 2.7 kb insert clone, Lane 6: 2.8 kb insert clone, Lane 7: 4.5 kb insert clone, Lane 8: 6.9 kb insert clone. The respective insert sizes are shown by horizontal lines.
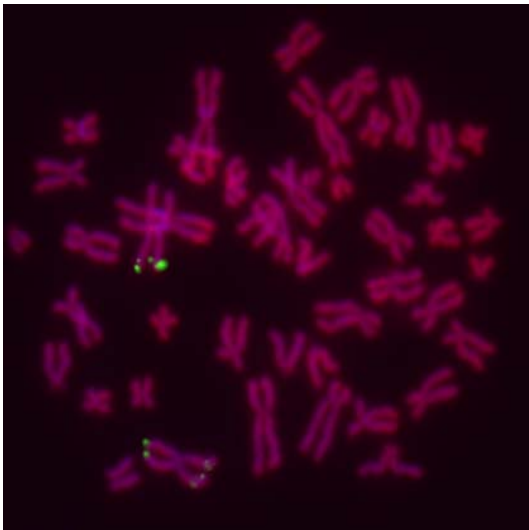
Fig. 21: Double hybridization FISH with the 6.9 kb subcloned fragment and 3p terminal BAC clone RP11-91K04. The normal chromosome is seen with the terminal signals on 3p arm and the signals on 3q arm but the inverted chromosome is seen with the terminal signals on the 3p arm as well as the inverted signals on the 3p arm in close proximity.

3.9.2   Long range PCR

The FISH analyses with BAC RP11-12N13 showed a split signal thus narrowed down the region to approximately 95 kb. An alternative approach to further narrow down the breakpoint region was carried out by Long range PCR. The whole breakpoint anchoring region of RP11-12N13 was considered and 6 primers in both forward and reverse directions were designed. The forward and reverse primers in the first half of the 30 kb region were labeled as d3 and d4, the middle 30 kb region as d5 and d6 and the last 39.5 kb region as d1 and d2. Three different sizes of products were formed, a 10.7 kb product with d3 and d4, a 10.8 kb product with d5 and d6 and a 13.7 kb product with d1 and d2 (Fig. 22).

3.9.2.1 FISH with Long range PCR products

The products formed by the long range PCR were labeled and used as probes for FISH experiments. Double hybridizations were performed with p terminal BAC clone RP11-91K04. FISH with the 10.7 kb PCR product showed that the signals were on the p arm (inverted) and the 10.8 kb PCR product also showed inverted signals on p arm. This indicated that the breakpoint was distal to these two regions. FISH with 13.7 kb PCR product showed a split signal, which indicated that the breakpoint was anchored with in this 13.7 kb region (Fig. 23). Thus, the breakpoint was narrowed down to 13.7 kb region. Single hybridization FISH with the 13.7 kb product showed split signals on both the arms of the inverted chromosome of the proband.
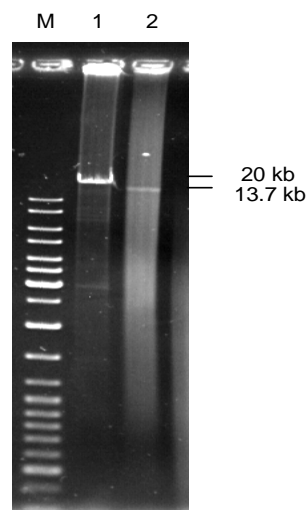
Fig. 22: The long range PCR gel

M is the 1 kb marker. Lane 1 a 20 kb λDNA control fragment and Lane 2 is the 13.7 kb amplified fragment.
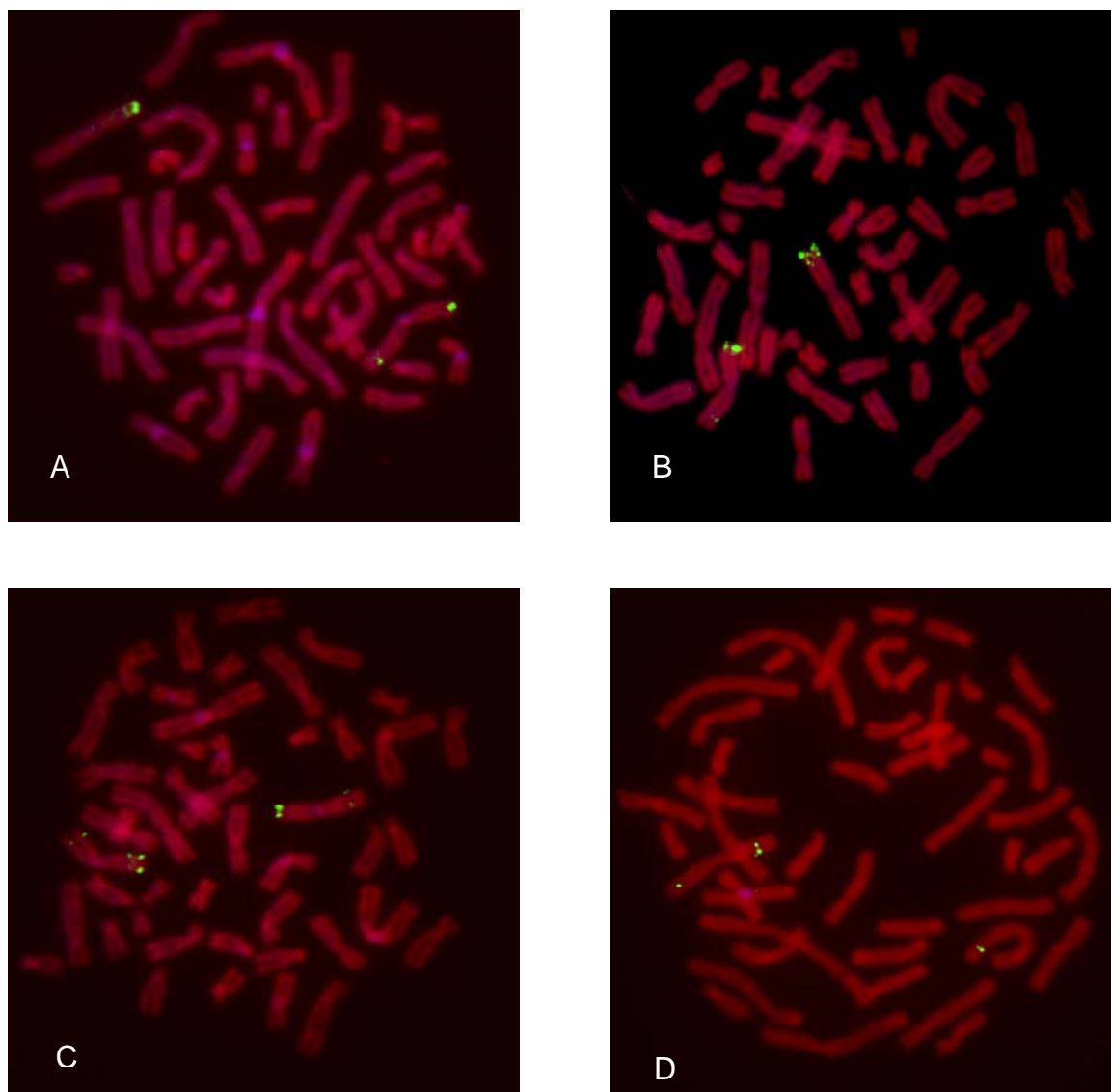
Fig. 23: FISH with Long range PCR products.

A: Double hybridization FISH with 10.7 kb product showing the inverted signals on p arm on the inverted chromosome and normal signals on q arm on normal chromosome. **B:** Double hybridization FISH with 10.8 kb PCR product showing inverted signals on p arm of the inverted chromosome and normal signals on the q arm of the normal chromosome. **C:** Double hybridization FISH with 13.7 kb PCR product showing a split signal on both the arms. **D:** Single hybridization FISH with 13.7 kb product showing the split signal on both the arm of the inverted chromosome and normal signals on q arm on the normal chromosome.

FISH with restricted fragments

As an alternative approach for subcloning the restriction fragments were used. The 3q region was narrowed down to a 13.7 kb region with the long range PCR product by FISH (see 3.9.2), the breakpoint was thus positioned from 144 kb to 157.7 kb of the BAC clone RP11-12N13. Further, two restriction enzymes *Sca*I and *Kpn*I were selected and fragments overlapping this 13.7 kb region were identified by using custom digest of NEBcutter (http://tools.neb.com/NEBcutter2). A 14.6 kb *Sca*I restricted fragment with its position from 138.6 kb-153.2 kb and a 7.4 kb *Kpn*I restricted fragment with its position from 140.3 kb-147.7 kb were selected (Fig. 24). These two fragments were gel eluted, biotin-labeled and used as probes for FISH experiments. Double hybridization FISH with p terminal BAC clone RP11-91K04 and the restricted fragments was performed and with both the fragments split signals were observed on both arms of the inverted chromosome of the proband (Fig. 25). By overlapping all these three fragment regions a minimum overlapping region was taken into consideration to deduce the approximate region of the breakpoint. With these results the breakpoint region was thus narrowed down to an approximate 3.7 kb region.
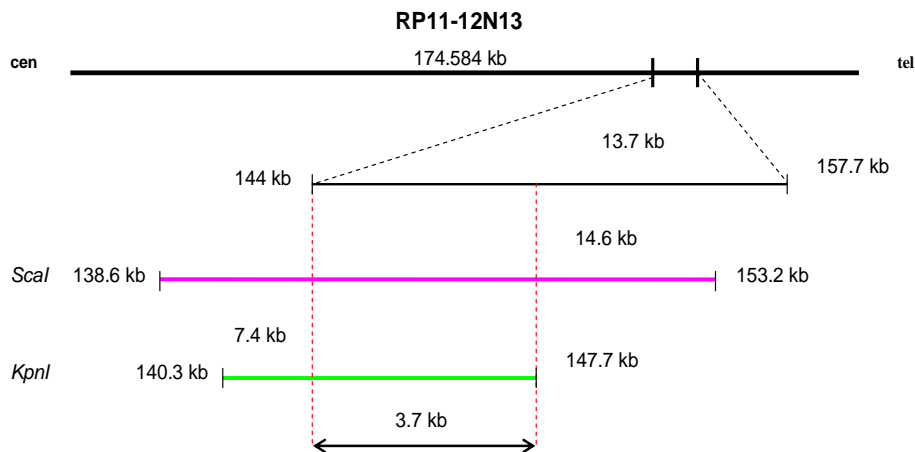
Fig. 24: Schematic representation of the restriction fragments selected on 3q region.

The first line represents the BAC clone RP11-12N13 and the second line is a magnified 13.7 kb Long range PCR product of the BAC RP11-12N13. The third line in pink is the *Sca*I fragment, the fourth green line is *Kpn*I fragment and the fifth line is the 3.7 kb region anchoring the breakpoint. The sizes are given above each line and the start position and end position of the fragments are shown on either side of the lines. The red interrupted vertical lines are representing the final narrowed down region.
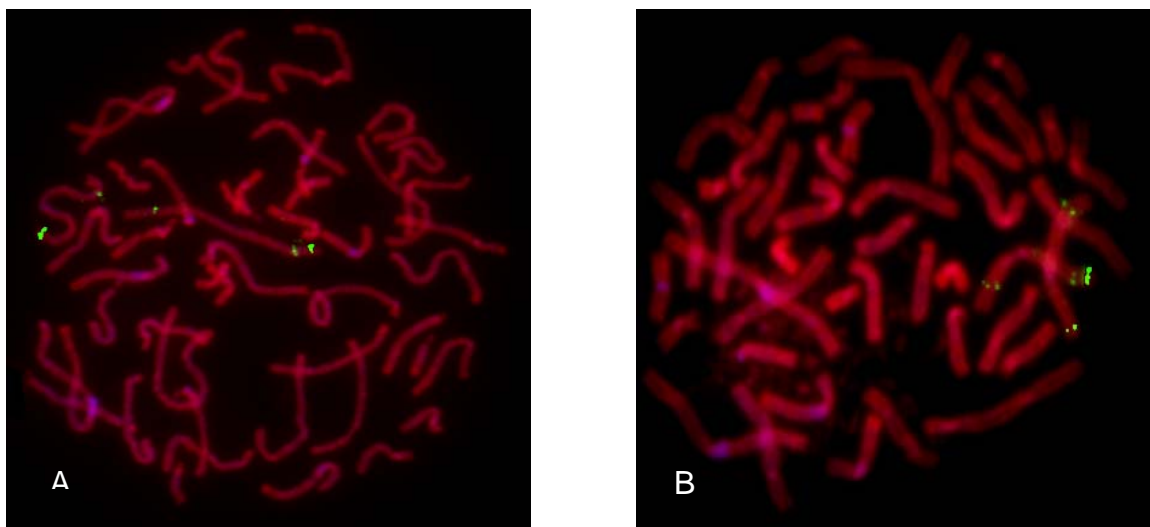
Fig. 25: FISH with restricted fragments.

A: Double hybridization FISH with 14.6 kb *ScaI* restricted fragment and terminal BAC RP11-91K04 showing terminal signals and split signals on the inverted chromosome and terminal signals and normal signals on q arm of the normal chromosome. **B:** FISH with 7.4 kb *KpnI* fragment showing p terminal signals and split signals on both the arms of the inverted chromosome and p terminal signals with signals on q arm of normal chromosome.

3.10   3p inversion breakpoint region- Delineation by FISH with YACs

The approach for delineation of the 3p region is almost similar to 3q region i.e., initially working with YACs and than with BACs. Four YACs were selected by *in silico* analysis and the details of the YACs were given in the Tab. 7. These 4 YACs were used as probes for FISH experiments. FISH with YAC clones 791A11 and 925B04 showed no signals on both metaphases and interphase cells. PFGE results confirmed the absence of inserts in these two clones. FISH with 803G05 showed signals on chromosome 3p and FISH with YAC clone 787H08 showed split signals on both the arms of the inverted chromosome of the proband (Fig. 26). Thus the breakpoint was assigned to chromosomal band region 3p24.1-24.2 and thus the region was narrowed down to ~1090 kb.

Tab. 7: The details of four 3p specific YACs with band position and insert sizes, and the YAC which showed split signals is in red.

p specific YACs	STS markers	Start position on chromosome (bp)	Band position	RZPD insert size (kb)
CEPHY904A11791D1 (791A11)	D3S1293	21902053	3p24.3	1170 (no signals)
CEPHY904H08787D1 (787H08)	D3S2335	26424557	3p24.2-24.1	1090
CEPHY904B04925D1 (925B04)	D3S2369	36472099	3p22.3	1440 (no signals)
CEPHY904G05803D1 (803G05)	WI-6058	38142018	3p22.2	2200
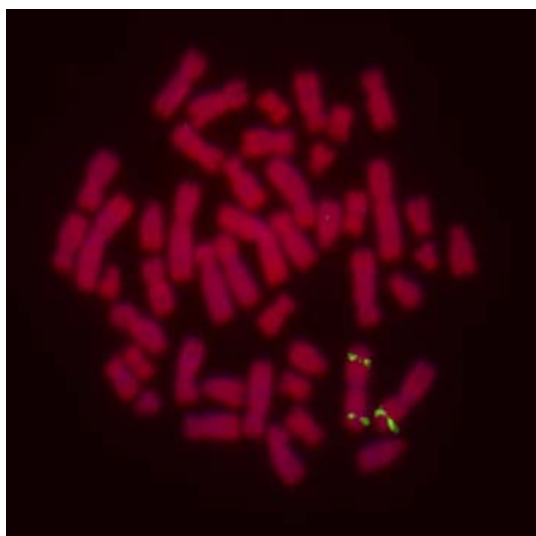
Fig. 26: FISH with YAC clone 787H08 on lymphocyte metaphase spreads of the proband.

The split signals are seen on both arms of the inverted chromosome 3. Signals only on the p arm are seen on the normal chromosome 3.

3.11   BAC selection

Now that the p region was defined to ~1090 kb, several BACs were selected and split signals with two superimposing BACs were obtained by our lab. These two BACs were RP11-666G20 and CTD-2007B5. Further characterization was carried on using the BAC clone CTD-2007B5. The detailed physical map from the breakpoint spanning YAC to the BAC clone CTD-2007B5 is shown in Figure 27.
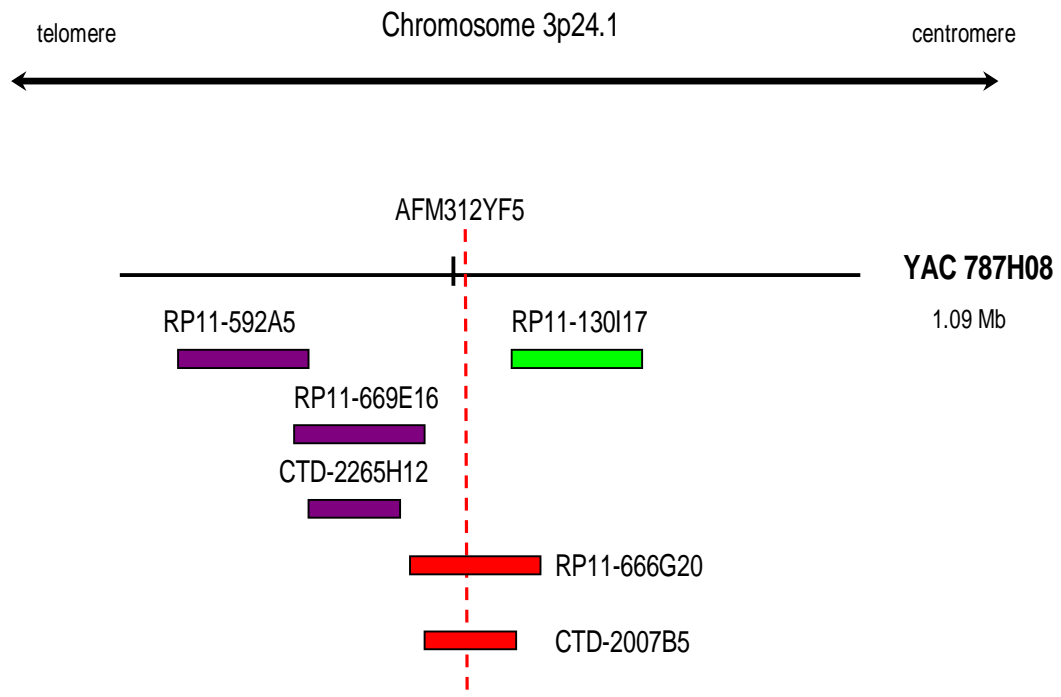
Fig. 27: A detailed 1.09 Mb physical map of the 3p24.1 inversion breakpoint region of the proband.
The YAC clone 787H08 is indicated by a horizontal line and its STS marker is indicated by vertical line. BAC clone RP11-592A5 showed a signal on p arm (non-inverted) and BAC clone RP11-130I17 showed signals on q arm (inverted). The interval between these two clones was further tested with 4 additional BAC clones in order to narrow down the breakpoint region even further. The BACs with violet colours showed signals on p arm, the BAC with green colour showed signals on q arm and the BACs in red colour showed signals on both the arms. The BAC clones indicated in red encompasses the breakpoint and the position of the breakpoint is shown as a vertical red interrupted line.

3.12   FISH with restriction fragments

The beakpoint was thus localized to 3p24.1 region. To further narrow down the breakpoint region, FISH with selected restriction fragments was performed. Initially, three restriction enzymes *Pac*I, *Sac*I and *Bam*HI were selected (Tab. 8) from the NEB cutter covering the 44.7 kb region of the CTD-2007B5 clone. The desired size anchored as the first fragment was selected for the easy isolation from the gel. Double hybridizations were performed using the p terminal BAC clone RP11-91K04 as reference. FISH with 11.8 kb *Pac*I fragment and 23.8 kb *Sac*I fragment showed weak signals on 3p region but FISH with the 26.4 kb *Bam*HI fragment showed split signals (Fig. 29). This suggested that the breakpoint was anchored within this 26.4 kb region. Additional five restriction enzymes *Acc*I, *Eco*RV, *Xba*I, *Hinc*II and *Kpn*I were selected covering this 26.4 kb *Bam*HI region (Tab. 8).

Tab. 8: The selected restriction enzymes with their coordinates on CTD-2007B5 and the sizes of the fragments.

Restriction enzymes	Coordinates	Length (bp)
*Pac*I	45410 - 57216	11807
*Sac*I	52834 - 76585	23752
*Bam*HI	71122 - 97555	26434
*Acc*I	73565 - 89266	15702
*Eco*RV	80632 - 92333	11702
	73660 - 80631	6972
*Xba*I	82673 - 97555	14883
*Hinc*II	85052 - 96626	11575
*Kpn*I	88967 - 97555	8589

FISH with 15.7 kb *Acc*I fragment, 11.7 kb *Eco*RV fragment, 14.9 kb *Xba*I fragment and 11.6 kb *Hinc*II fragment showed split signals (Fig. 29). Since all the fragments were overlapping each other the end position of *Acc*I fragment and the start position of the *Hinc*II fragments were subtracted and thus the breakpoint was defined to ~4 kb region. Further, confirmation was done by taking fragments from either side of the breakpoint and checking their position. FISH with an 8.6 kb *Kpn*I fragment on the other side showed signals to be present on the q region only (Fig. 30). This also confirmed that the breakpoint was anchored with in this 4 kb region. The detailed map of the restriction fragments is shown in Fig. 28.
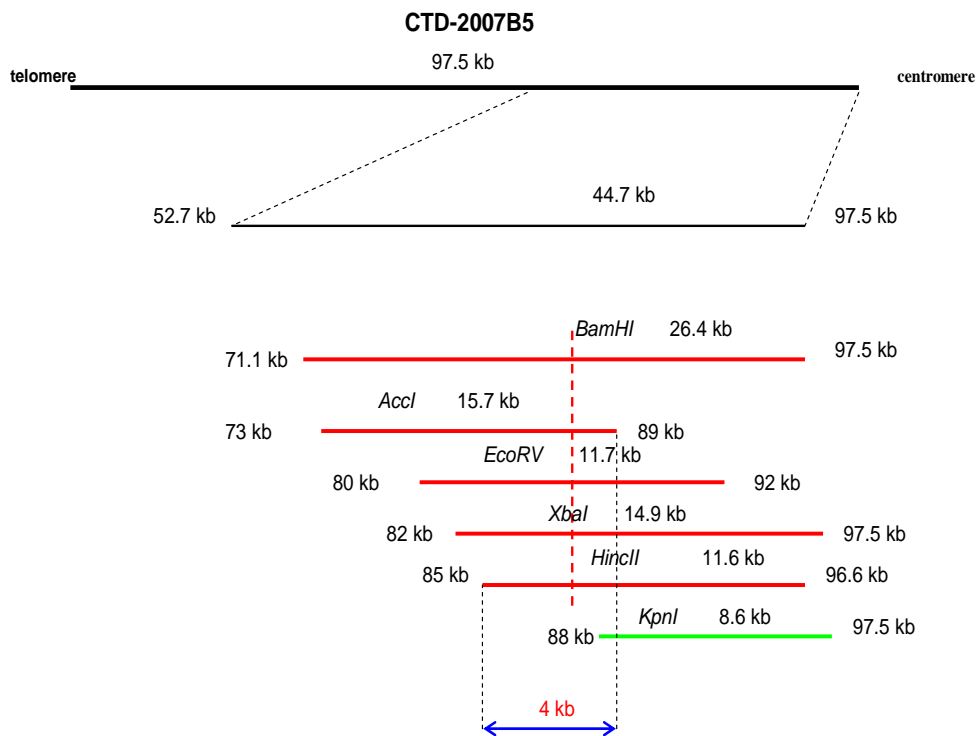
Fig. 28: Schematic representation of the detailed physical map of the restriction fragments.

The first line represents the BAC clone CTD-2007B5 and the second line is a magnified region anchoring the breakpoint. The sizes are given above each line and the start position and end position of the fragments are shown on either side of the lines. The third line in red is the 26.4 kb *Bam*HI fragment which gave a split signal. The other four fragments shown in red *Acc*I (15.7 kb*), Eco*RV (11.7 kb), *Xba*I (14.9 kb) and *Hinc*II (11.6 kb) were selected overlapping the *Bam*HI region. All these fragments anchored the breakpoint and the vertical interrupted line in red encompasses the breakpoint region. The 8.6 kb *Kpn*I fragment in green showed signals on the q region. The vertical black interrupted lines shows the overlapping region and the arrowhead blue line shows the final narrowed down region of ~4 kb.
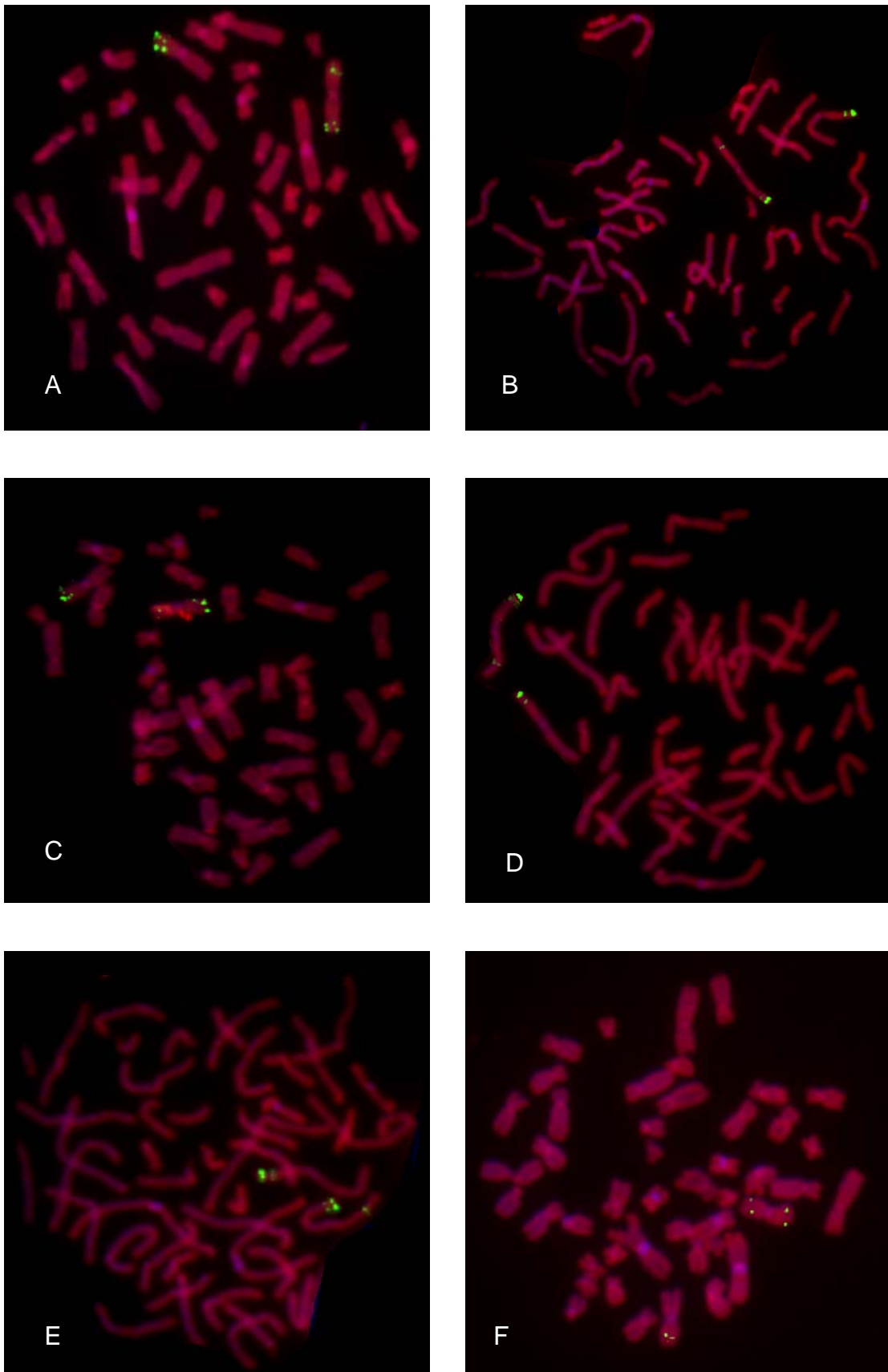
Fig. 29: (legend on the following page).

 (See figure on the previous page)

Fig. 29: FISH with restriction fragments on the metaphase chromosomes of the proband.

The terminal p specific BAC clone RP11-91K04 was a reference for all the double hybridizations. **A:** Double hybridization FISH with 26.4 kb *Bam*HI fragment showing split signals on both the arm of the inverted chromosomes and normal signals on p arm on the normal chromosome. **B:** Double hybridization FISH with a 15.7 kb *Acc*I fragment showing split signals on both the arms. **C:** Double hybridization FISH with a 11.6 kb *Hinc*II fragment showing split signals on both the arms of the inverted chromosome of the proband. **D:** Double hybridization FISH with a 14.9 kb *Xba*I fragment showing split signals on both the arms of the inverted chromosome. **E:** Double hybridization FISH with a 11.7 kb *Eco*RV fragment showing split signals on both the arms of the inverted chromosome. **F:** Single hybridization FISH with a 11.7 kb *Eco*RV fragment showing split signals on both the arms of the inverted chromosome.
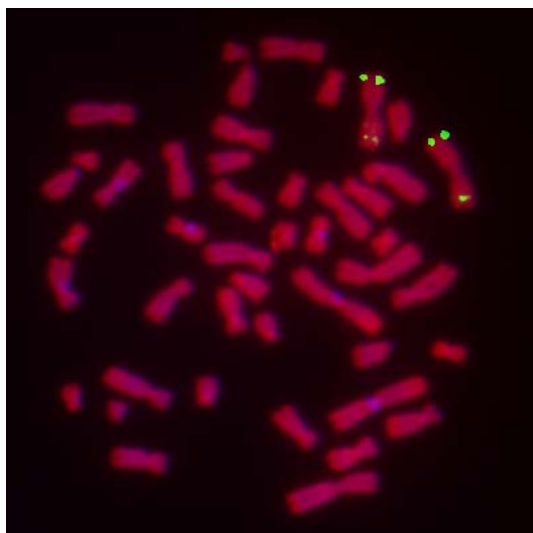
Fig. 30: FISH with an 8.5 kb *Kpn*I fragment. The p terminal signals are seen on both the chromosomes and the q signals of the 8.5 kb *Kpn*I fragment are also seen on the q arm of both normal and inverted chromosome.

3.13 Redefining the inversion breakpoints

All the FISH analyses were performed on the proband, her mother and grand mother and the three subjects showed the same results for all the experiments carried out. The initial karyotype was cytogenetically assigned as inv(3)(p23;q25q26) but after the delineation of the breakpoints the karyotype was redefined as inv(3)(p24.1;q26.1) (Fig. 31).
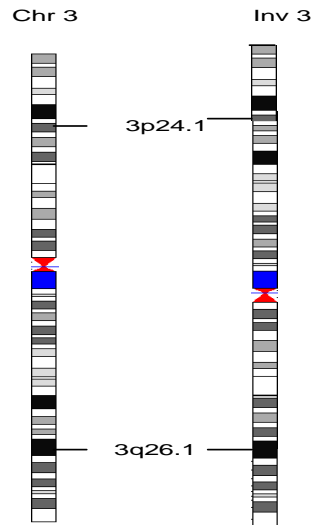
Fig. 31: Ideogram showing the refined inversion breakpoint regions.

3.14   Identification and characterization of both inversion breakpoints on genomic DNA

The breakpoint regions of the proband were also characterized at the genomic level. To characterize the junction fragments of the breakpoints at the molecular level, PCR was performed. Approximately a 6 kb breakpoint anchoring region on the 3p BAC CTD-2007B5 and a 5.7 kb region on the 3q RP11-12N13 were selected to design several primers. Breakpoint spanning PCR was performed with more than 17 different possible primer combinations in order to amplify the genomic DNA of the proband and control. Initially, a junction product of ~ 500 bp was formed with U5/U6 primer combination (Fig. 32 and Supplementary Fig. 3). This PCR product was cloned and sequenced. The sequenced U5/U6 product from one end showed 81% homology to the 3q breakpoint region ~252 bp (Fig. 33) and the rest of the sequences showed homology to an IMAGE clone on chromosome 1 open reading frame. To find out any possible insertions or translocations from chromosome 1 several primers were designed extending the non homologous region but the data showed no involvement of chromosome 1 in this inversion. To check for any possible error in the genomic sequence in this region a 4 kb control DNA product was amplified and sequenced. The control sequence showed 100% homology to the database sequence. In order to find the junction fragment several

combination of primers were used and several different products were formed (Tab. 9). The region was also extended on both the sides of the breakpoint to design new primers and long range PCR was performed which again yielded many different products. All these products were cloned and sequenced and blasted for homologous sequences and all the sequences showed homology to different chromosomes (Tab. 9). The 3q region showed 85 % homology to a BAC clone on 1q44 region. It also showed 86% Alu warning. This region also showed homology to chromosome 21. On further analysis it was revealed that this breakpoint region is associated with strong repetitive region which is detailed in 3.18.
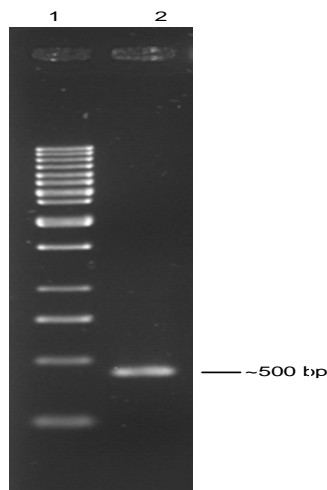
Fig. 32: Gel picture showing the ~ 500 bp U5/U6 products designed to amplify the junction fragment at 3q26.1.

Lane 1: 1 kb marker

Lane 2: ~ 500 bp PCR product

```
Chr 3      GTCTGGATCTGTTACCCAGGCTGGAGTGCAGTGGTATGATCTCCGCTCACTGCAACCTCT
           ||||||||||||||||| ||||||||||||||||   ||||| |||||||||||  |
U5/U6      GTCTGGATCTGTTACCG-GGCTGGAGTGCAGTGGTGCAATCTCGGCTCACTGCAATGTTC

Chr 3      GCCTCCCAGGTTCAAGTGATTCTCCTGCCTCAGCTTCCAGAGTATCTGGGATTACAGGCT
            |||||| ||||||||| |||||||||||||||||||| ||| ||| | |||||| |||||||
U5/U6      ACCTCCCCGGTTCAAGGGATTCTCCTGCCTCAGCCTCCCGAGCAGCTGGGACTACAGGCG

Chr 3      CCCGCCACCACATCCAGCTAATTTTTGTGTTTTTAGTAGAGACGGGGTTTCGCCATGTTG
            | |||||| || |||||||||||| |||| ||||||||| ||||||||| ||||||||
U5/U6      GGTGACACCATGCCCGGCTAATTTTTGTATTTTCAGTAGAGATGGGGTTTCACCATGTTG

Chr 3      GCCAGGCTGATCTCAAACTCCTGTCCTCCAGTGATCCACCCACCTTGGCCTCCCAAAGTG
            | | || || || | | ||| || |||| ||||||||| |||||||| | |||||||||||
U5/u6      GTCTGGATGGTCGCGATCTCTTGGCCTC--GTGATCCGCCCACCTCTGTCTCCCAAAGTG

Chr 3      CTGGAATT
           |||| |||
U5/U6      CTGGGATT
```

Fig. 33: Schematic representation of the homologous sequences of U5/U6 PCR product of genomic DNA with the chromosome 3q region.

Tab. 9: The PCR products formed with the different primer combinations of the breakpoint spanning region.

The products with their sizes and the homology of more than 85% to other chromosomes are also shown.

PCR Product (primer combination)	Fragment size (bp)	Chromosome homology
U5/U6	~500	1
U6/U7	~500	1
U7/U8	~150	1
U6/U10	~500	5
U3/U10	~500	9
U8/U10	~300	7
U8/U9	~500	20
	~1,000	20
U5/U8	~350	1
U6/U9	~800	16
	~900	5
	~1,200	22
	~2,500	21
U17/U20	~1,500	14
d1/U10	~500	9
U23/U19	~1,000	10
U29/U21	~2,500	14
U22/U26	~750	15
U21/U29	~5,500	2
d9/d12	~3,500	12
U20/d10	~1,300	16
U13/U31	~900	4

3.15   Confirmation and characterization of both inversion breakpoints

Due to the difficulties in breakpoint spanning PCR, further confirmation was carried on by FISH using PCR amplified products. In order to further confirm and refine the breakpoints on q and p regions, several primers were designed on both the ends of the breakpoint regions. An approximate 6 kb was taken on either side of the breakpoint to design the primers. BAC clone RP11-12N13 was taken as a template for amplifying the q specific primers and a product of 5,353 bp was obtained. On the other side CTD-2007B5 was taken as a template and amplified with p specific primers. Four products of sizes 2,933 bp, 3000 bp, 4500 bp and 5000 bp were obtained. All these PCR amplified products were biotin labeled and used as probes for FISH experiments. Double hybridizations and single hybridizations were performed and all the products showed a split signal confirming the presence of the breakpoint. As there were 4 products which showed split signals on the p region, the smallest size product i.e., 2,933 bp was taken into consideration for further analysis (Fig. 34). The 5,353 bp 3q PCR product showed split signals on the q region (Fig. 34). Both these 2,933 bp 3p and 5,353 bp 3q PCR products were also sequenced. The FISH results with the other three PCR products of 3000 bp, 4500 bp and 5000 bp on p region are shown in the Figure 35.

Fig. 34: FISH with PCR amplified products on the metaphase chromosomes of the proband.

A: Double hybridization FISH with a p terminal BAC RP11-91K04 and the 3q 5,353 bp PCR product showing terminal p signals and normal signals on q arm and the inverted chromosome is seen with terminal p signals and split signals on both the arm. **B:** Single hybridization FISH with the 5,353 bp PCR product showing split signals on both the arms. **C:** Double hybridization FISH with a p terminal BAC RP11-91K04 and the 3p 2,933 bp PCR product showing terminal p signals and signals on the same arm whereas the inverted chromosome is seen with the p terminal signals and split signals on both the arms of the chromosome. **D:** Single hybridization FISH with the 2,933 bp PCR product showing a split signals on both the arms of the inverted chromosome.
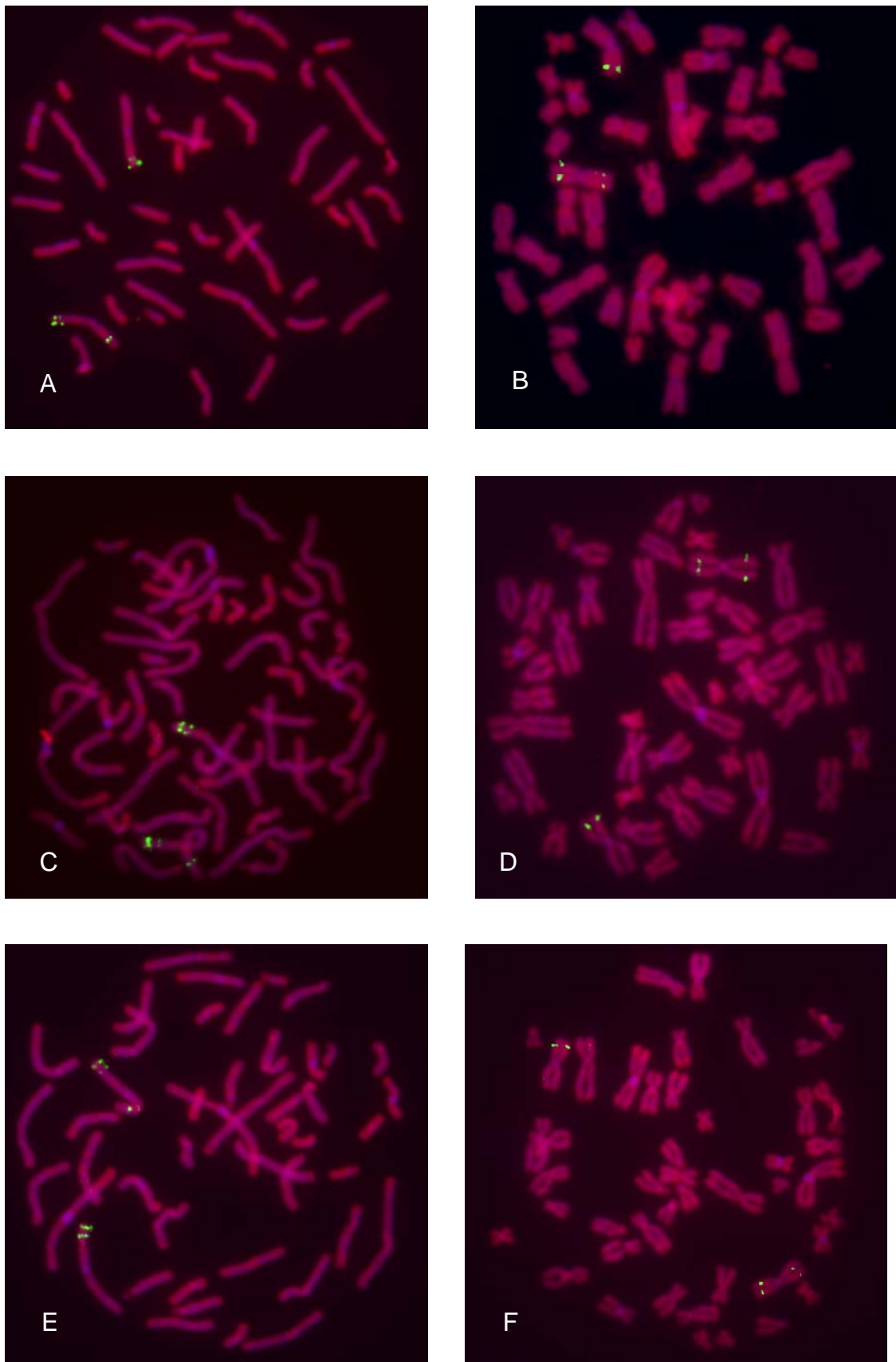
Fig. 35: (Legend on the following page).
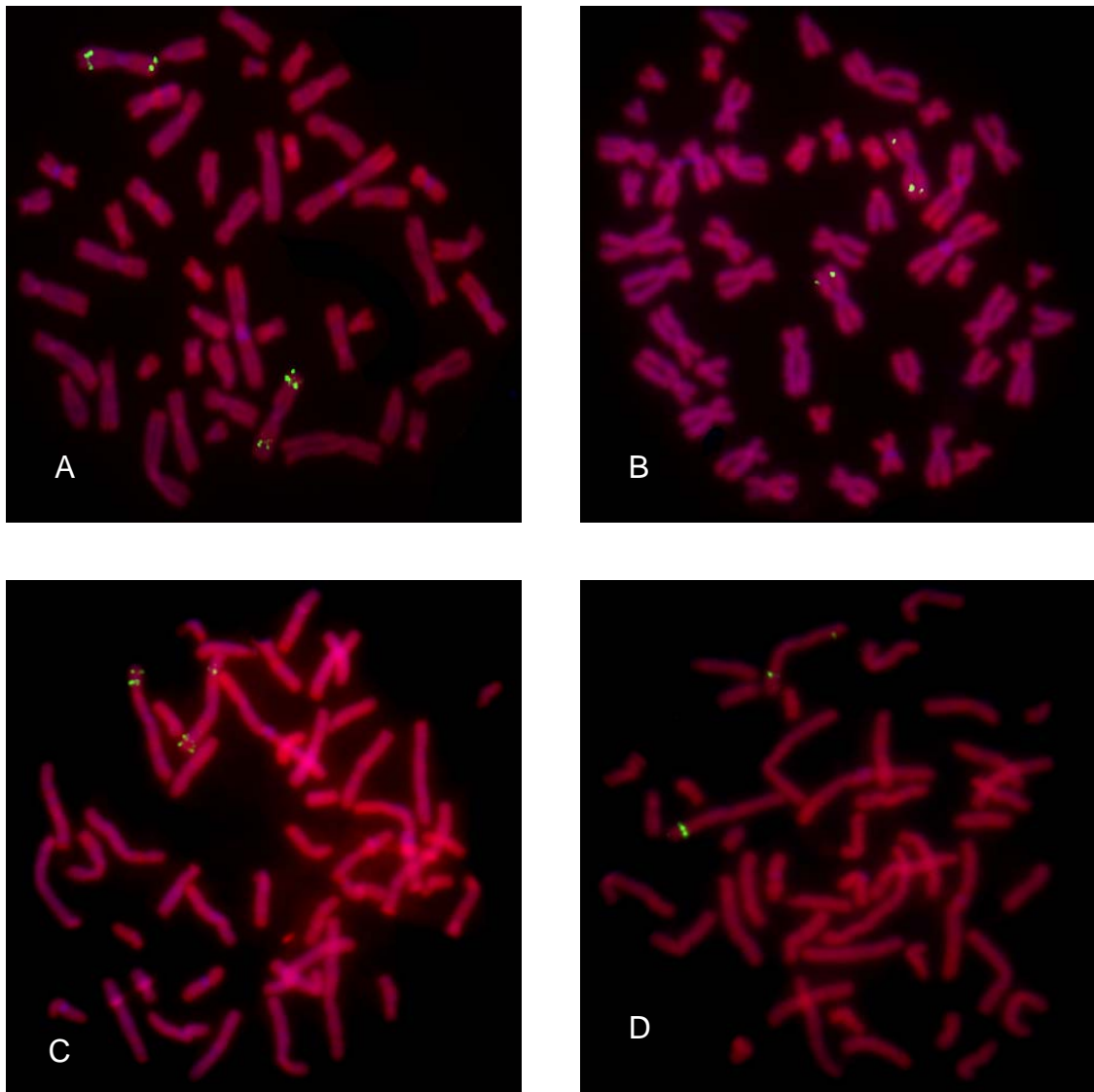
(Figure on previous page).

Fig. 35: FISH with 3p PCR amplified products on the metaphase chromosomes of the proband.

A: Double hybridization FISH with a p terminal BAC RP11-91K04 and the 4 kb PCR product showing terminal p signals and normal signals on p arm and the inverted chromosome is seen with terminal p signals and split signals on both the arm. **B:** Single hybridization FISH with the 4 kb PCR product showing split signals on both the arms. **C:** Double hybridization with terminal BAC RP11-91K04 and 4.5 kb PCR product showing terminal signals on p arm and product signals on p arm of normal chromosome and terminal p signals and split signals on inverted chromosome. **D:** Single hybridization FISH with 4.5 kb PCR product showing split signals on inverted chromosome. **E:** Double hybridization with terminal BAC RP11-91K04 and 5 kb PCR product showing terminal signals on p arm and product signals on p arm of normal chromosome and terminal p signals and split signals on inverted chromosome. **F:** Single hybridization FISH with 5 kb PCR product showing split signals on inverted chromosome.

3.16   Database searches for the identification of putative novel genes

As described in 3.8 the 3q breakpoint region was localized within the sequence of 167,343,069-167,517,652 bp on genomic region of chromosome 3 within the BAC clone RP11-12N13 (174,584 bp). BLAST analysis with the genomic sequence of this clone revealed 162 homologies to EST (Expressed sequence tag) sequences. All the sequences were analyzed *in silico* and they formed consensus sequences of 4,649 bp. The consensus sequences spanned from 86,381-91,030 bp of the BAC clone RP11-12N13. Two speculated genes sleyjabu and slajabu were found to be anchored within this BAC clone from NCBI Aceview (www.ncbi.nlm.nih.gov/IEB/Research/Acembly). According to Aceview sleyjabu has shown one exon of 828 bp. This sleyjabu exon spanned from 109,990-110,950 bp of the BAC RP11-12N13. The other putative gene slajabu, according to Aceview showed 4 exons; Exon 1-389 bp, exon 2-94 bp, exon 3-125 bp and exon 4-303 bp. The last two exons of slajabu were anchored in this clone and former two exons were found in two proximal BAC clones; exon 1 in RP11-480L9 and exon 2 in RP11-233K8. Exon 3 was found spanning from 48,960- 49,085 bp and exon 4 was found spanning from 50,625-50,929 bp of RP11-12N13 (Fig. 36).

Subsequently, as the breakpoint was narrowed down to a 13.7 kb region with the long range PCR product as described in 3.9, the sequence of 13.7 kb was taken to search the database for ESTs. This 13.7 kb region spanned from 167,487,077-167,500,077 bp of

RP11-12N13 and there were no ESTs found within this breakpoint region. Further to check for any overlapping ESTs, search was gradually increased to 100 kb on either side of the breakpoint. Many ESTs were identified but approximately 10 ESTs were selected for further analysis.
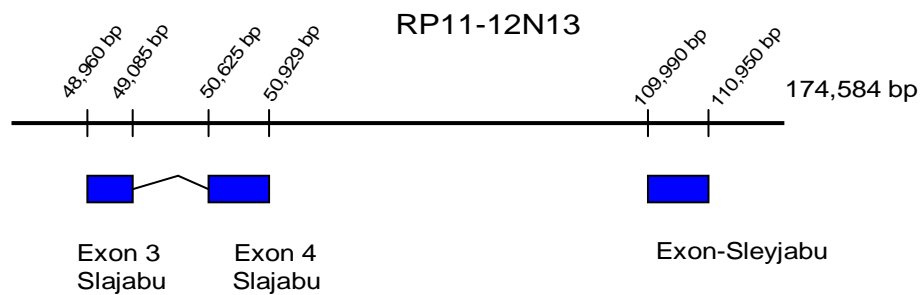
Fig. 36: The schematic representation showing the position of the exons of the putative genes sleyjabu and slajabu on 3q BAC RP11-12N13.

The horizontal line represents the BAC clone RP11-12N13 with 174,584 bp. The vertical small lines show the position of the exons and consensus sequences formed and the blue boxes represent the exons and the consensus sequences.

Similarly, as described in 3.12 the p region was localized within the sequence of (26792920- 2689047 bp) on the clone CTD-2007B5 (total insert sequences 97,555 bp). The *Bam*HI 26.4 kb fragment was localized at 26864040- 26890474 bp on the clone. BLAST analysis with this region showed 428 EST homologies. The EST search was further continued with a much narrowed down 11.7 kb *Eco*RV fragment. This fragment was localized within the sequence of 26873552- 26885252 bp on the clone. EST search was extended to approximately 80 kb on either side of the breakpoint region. Many ESTs were identified but no EST was found within the breakpoint region. Approximately 10 ESTs from either side of the breakpoint regions were selected for further analysis.

3.17 cDNA characterization and *in silico* Gene identification

The ESTs selected on either side of the breakpoint regions were ordered as cDNA clones. Seven p specific cDNA IMAGE clones: H1417075Q2, B0410432Q1, I0413343Q1, K099654Q1, M0514122Q1, H192314Q2 and B062469Q2 and three q specific cDNA clones: F1113301Q1, B0415915Q1 and O074892Q1 were procured and characterized. All the cDNAs were initially checked for the insert size by restriction digestion. All the inserts were then sequenced from either side by using M-13 forward and reverse primers. The positions of the clones were checked by aligning the sequences with the genomic clones.

To find out for any existing genes homology based searches were performed. The sequence similarities of the genomic sequence with the sequences of the cDNAs were compared, using BLASTN. Some cDNA sequences showed homology to the genomic sequence but failed to show any exon-intron pattern.

3.18   Inversion breakpoints are associated with repetitive elements

The frequency and classes of repeated elements in the breakpoint regions were analyzed by RepeatMasker (http://www.repeatmasker.org). The 3q breakpoint region showed 26.4% repeats and the 3p breakpoint region showed a higher percentage of repetitive elements of 27.5%. Out of the different classes of repeats, SINEs, LINEs, LTR elements, simple repeats and low complex repeats were found. However, the repeat composition of the two breakpoint regions varied considerably (Tab. 10). The average human genome shows 2.2% MIR, 20.4% LINEs, 8.3 % LTRs and 2.8% DNA transposons (The International Human Genome Sequencing Consortium, 2001). The LTR elements in the p region and DNA transposons on the q region were much higher than those on the average human genome. All the types of repeats and the classes identified are shown in the schematic maps. The detailed search was carried out with both the breakpoint regions and also further analyzed in USCS browser.

Tab. 10A: The 2933 bp breakpoint region on 3p showing different composition of the repeats.

```
==================================================

file name:          3p
sequences:           1
total length:      2933 bp  (2933 bp excl N/X-runs)
GC level:          39.82 %
bases masked:       806 bp ( 27.48 %)
==================================================
             number of      length   percentage
             elements*     occupied  of sequence
--------------------------------------------------
SINEs:              0             0 bp    0.00 %
     ALUs           0             0 bp    0.00 %
     MIRs           0             0 bp    0.00 %

LINEs:              1            74 bp    2.52 %
     LINE1          0             0 bp    0.00 %
     LINE2          1            74 bp    2.52 %
     L3/CR1         0             0 bp    0.00 %

LTR elements:       4           680 bp   23.18 %
     MaLRs          0             0 bp    0.00 %
     ERVL           2           226 bp    7.71 %
     ERV_classI     2           454 bp   15.48 %
     ERV_classII    0             0 bp    0.00 %

DNA elements:       0             0 bp    0.00 %
     MER1_type      0             0 bp    0.00 %
     MER2_type      0             0 bp    0.00 %

Unclassified:       1            52 bp    1.77 %

Total interspersed repeats:     806 bp   27.48 %
Small RNA:          0             0 bp    0.00 %

Satellites:         0             0 bp    0.00 %
Simple repeats:     0             0 bp    0.00 %
Low complexity:     0             0 bp    0.00 %
==========================================
```

Tab. 10B: The 5354 bp breakpoint region on 3q showing different composition of the repeats.

```
===================================================
sequences:              3q
total length:        5354 bp  (5354 bp excl N/X-runs)
GC level:            35.28 %
bases masked:        1414 bp ( 26.41 %)
===================================================
                number of      length    percentage
                elements*    occupied    of sequence
---------------------------------------------------
SINEs:                 0           0 bp     0.00 %
      ALUs             0           0 bp     0.00 %
      MIRs             0           0 bp     0.00 %

LINEs:                 0           0 bp     0.00 %
      LINE1            0           0 bp     0.00 %
      LINE2            0           0 bp     0.00 %
      L3/CR1           0           0 bp     0.00 %

LTR elements:          2        1037 bp    19.37 %
      MaLRs            0           0 bp     0.00 %
      ERVL             1         445 bp     8.31 %
      ERV_classI       1         592 bp    11.06 %
      ERV_classII      0           0 bp     0.00 %

DNA elements:          1         200 bp     3.74 %
      MER1_type        1         200 bp     3.74 %
      MER2_type        0           0 bp     0.00 %

Unclassified:          0           0 bp     0.00 %

Total interspersed repeats:     1237 bp    23.10 %


Small RNA:             0           0 bp     0.00 %

Satellites:            0           0 bp     0.00 %
Simple repeats:        2          81 bp     1.51 %
Low complexity:        2          96 bp     1.79 %
===================================================

* most repeats fragmented by insertions or deletions
  have been counted as one element
```

3.18.1   Types of repeats identified in the 3p breakpoint region

The 2933 bp breakpoint anchoring sequence was masked with the repeat masker and five different regions were found to be repeat rich. There were 4 types of repeats identified; LTR33A, MER67C, L2 and MER67D. In mammals, almost all transposable elements fall into one of the four types, of which three transpose through RNA intermediates and one directly as DNA. These are long interspersed elements (LINEs), short interspersed elements (SINEs), LTR retrotransposons and DNA transposons (The International Human Genome sequencing Consortium, 2001). All the four types of repeats identified in this region were transposable elements (Fig. 37 and Supplementary Fig. 1). The details of their position in the genomic sequence were mentioned in detail in Table 11.

Tab. 11: Types of repetitive elements identified on the 2,933 bp breakpoint region on 3p.

The families, classes, start position and end position on the chromosome and the sizes are also shown.

Name	Family	Class	Start position	End position	Size(bp)
LTR33A	ERVL	LTR	26879845	26880146	302
LTR33A	ERVL	LTR	26880746	26880794	49
MER67C	ERV1	LTR	26880974	26881351	378
L2	L2	LINE	26881376	26881449	74
MER67D	ERV1	LTR	26881458	26881533	76

Fig. 37: Schematic map of the types of repeats identified in the 3p breakpoint region.

The first line represents the BAC clone CTD-2007B5. The two black vertical lines on the BAC show the 2,933 bp breakpoint anchoring region. The second line represents the zoomed up region of 2933 bp. The vertical lines show the position of the repeats and the boxes below with different colors show the different types of repeats and their names below the boxes.
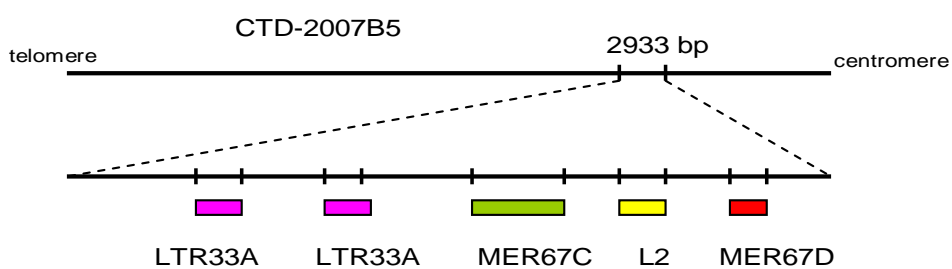
3.18.2   Types of repeats identified in the 3q breakpoint region

The 5,354 bp breakpoint anchoring sequence of RP11-12N13 was masked with the repeat masker and seven different repeat regions were found. Six different types of repeats were identified: LTR16C, MER20, MTL2B1, LTR1B, simple repeats and low complexity elements. Two types of transposable elements, two simple repeats, two low complexity and simple tandem repeats were observed in this region. The details of their position in the sequence were mentioned in Table 12 and the repeat map shows the details of the position and type of repeats (Fig. 38). The positions of the repeats are also shown in the sequence (Supplementary Fig. 2).

Tab. 12: Types of repetitive elements identified on 5,354 bp breakpoint region on 3q.

The families, classes, start position and end position on the chromosome and the sizes of the repeats were shown in detail.

Name	Family	Class	Start position	End Position	Size(bp)
LTR16C	ERVL	LTR	167487545	167487740	196
MER20	MER_type	DNA	167488669	167488868	200
(TATG)n	Simple repeat	Simple repeat	167488869	167488907	39
AT_rich	Low complexity	Low complexity	167490634	167490656	23
MLT2B1	ERVL	LTR	167490935	167491379	445
(CA)n	Simple repeat	Simple repeat	167491395	167491436	42
AT_rich	Low complexity	Low complexity	167491550	167491622	73
LTR1B	ERV1	LTR	167492070	167492873	804
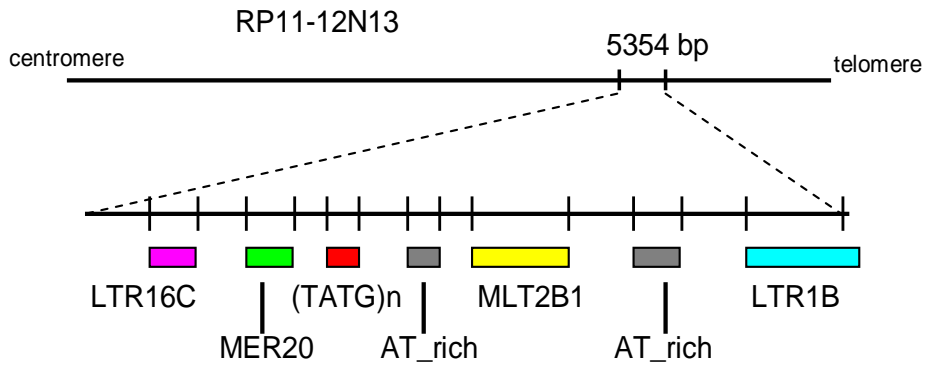
Fig. 38: Schematic map of the types of repeats identified in the 3q breakpoint region.

The first line represents the BAC clone RP11-12N13, the second line represents the zoomed out region of 5,354 bp. The boxes in different colors show the different types of repeats and the names of the repeats are listed below the boxes.

3.19   *In silico* analysis of inversion breakpoint regions for segmental duplications

The inversion breakpoint regions were studied for the presence of paralogous segments using the Human Genome Segmental Duplication Database. This database contains the updated data regarding recent segmental duplications in the finished human genome (build36; March 2006; hg18). It focuses on genomic duplications ≥1 kb and ≥90% identity. The inversion breakpoint regions were anchored within a distance from 26 - 27 Mb on p region and 167-168 Mb on q region of chromosome 3. There were no segmental duplications detected in the vicinity of 1 Mb region anchoring these breakpoints (Fig. 39). The chromosome 3 is relatively devoid of segmental duplications, having just 1.7% of its

bases composed of duplicated sequence compared to the whole-genome average of 5.3%. This is the lowest percentage for any chromosome in the genome (Muzny et al., 2006).
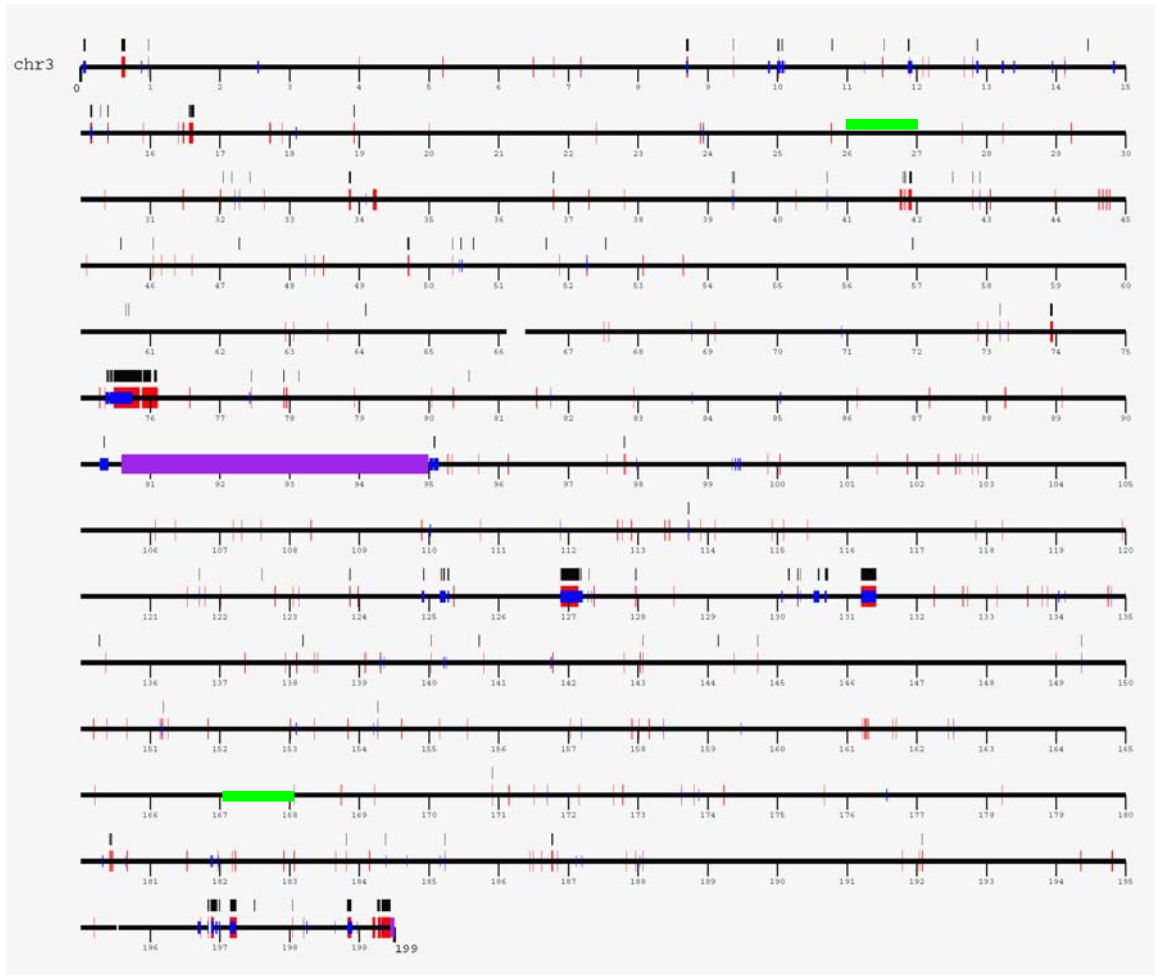
Fig. 39: Chromosomal view of segmental duplications in chromosome 3.

The black color thick horizontal lines represent the genomic sequence of chromosome 3 from 0 - 199 Mb. Gaps are depicted as line discontinuous on the chromosomal sequences. The large-insert clones that tested positive with the whole-genome shotgun detection were shown as black vertical lines above the genomic sequence. The whole genome assembly duplications are shown in red and blue vertical lines for inter and intra chromosomes respectively. The green color highlighted regions from 26 - 27 Mb on p region and from 167-168 Mb on the q region are the regions anchoring the inversion breakpoints which are devoid of segmental duplications.

3.20   Gene content of the two breakpoint regions

The National center for biotechnology information (NCBI) database was used to determine the gene content of the two inversion breakpoint regions. There were no known validated gene(s) disrupted by the inversion, but many bona fide genes and predicted genes in an approximate vicinity of 1 Mb from the distal and proximal breakpoint regions were identified. In the 3p breakpoint region towards the distal end six putatively transcribed sequences LRRC3B, LOC729108, VENTXP4, LOC645101, LOC645065 and LOC246135 were identified. The gene LRRC3B maps ~ 151 kb adjacent to the distal breakpoint region. Two known genes SLC4A7 and EOMES and three putative transcripts NEK10, LOC643634, and MGC6157 were identified proximal to the breakpoint. NEK10 was the nearest putative gene about ~426 kb away from the proximal breakpoint (Tab. 13). In addition to the NCBI database, AceView showed six hypothetical genes skeyvarbo, kluswerby, deefybo, dorfybo, klarswerby and fawfybo towards the proximal end of the breakpoint region (Fig. 40).

The region surrounding the 3q proximal breakpoint showed two genes BCHE and SLITRK3. The locus of BCHE gene was ~449 kb proximal to the breakpoint region. Towards the distal breakpoint region two genes FLJ23049, SERPINI2 and two putative transcripts LOC389173 and LOC131055 were identified (Tab. 13). SERPINI2 is found to be ~1 Mb away from the distal breakpoint region. In addition to these genes Aceview showed three hypothetically predicted genes bujubu, jyjybu and kleejybu towards the distal end. Towards the proximal end of the breakpoint 6 predicted genes sleyjabu, slajabu, shyjabu, transposase_22.556, skersweyby and rarjabu were identified (Fig. 40). The hypothetical gene sleyjabu was 34.7 kb to the proximal end of the breakpoint region and was within the breakpoint spanning BAC RP11-12N13 as mentioned in 3.16. The last two exons of slajabu were also found on the insert of this BAC clone. Interestingly, a Genscan predicted gene was identified spanning the 3q breakpoint region, which is detailed in 3.21.

Tab. 13: Gene contents of the two inversion breakpoints (according to NCBI database).

Breakpoint	Gene	Description	Status*
3p24.1	LRRC3B	leucin rich repeat containing 3B	E
	LOC729108	hypothetical protein LOC729108	M
	VENTXP4	VENT homeobox (Xenopus laevis) pseudogene 4	I
	LOC645101	similar to high-mobility group box 3	P
	LOC645065	similar to ribulose-5-phosphate-3-epimerase	P
	LOC246135	TBP-associated factor 9-like pseudogene	I
	NEK10 or FLJ32685	NIMA (never in mitosis gene a)-related kinase 10	E
	SLC4A7	solute carrier family 4, sodium bicarbonate cotransporter, member 7	C
	EOMES	eomesodermin homolog (Xenopus laevis)	C
	MGC6157	hypothetical protein MGC61571	A
	LOC643634	hypothetical LOC643634	P
3q26.1	BCHE	butyrylcholinesterase	C
	SLITRK3	SLIT and NTRK-like family, member 3	C
	LOC389173	similar to phosphoserine aminitransferase isoform 1	P
	LOC131055	similar to peptidylprolyl isomerase A isoform 1	M
	FLJ23049	hypothetical protein FLJ23049	C
	SERPINI2	serpin peptidase inhibitor, clade I (pancpin), member 2	C

*Status: The gene type and reference sequence status is depicted as a single alphabet. C: confirmed gene model based on alignment of mRNA or mRNA plus ESTs to the genomic sequence. Protein coding reviewed or validated; E: protein coding and provisional; P: pseudo gene model; M: protein coding model; A: protein coding and predicted; I: pseudo gene and provisional.
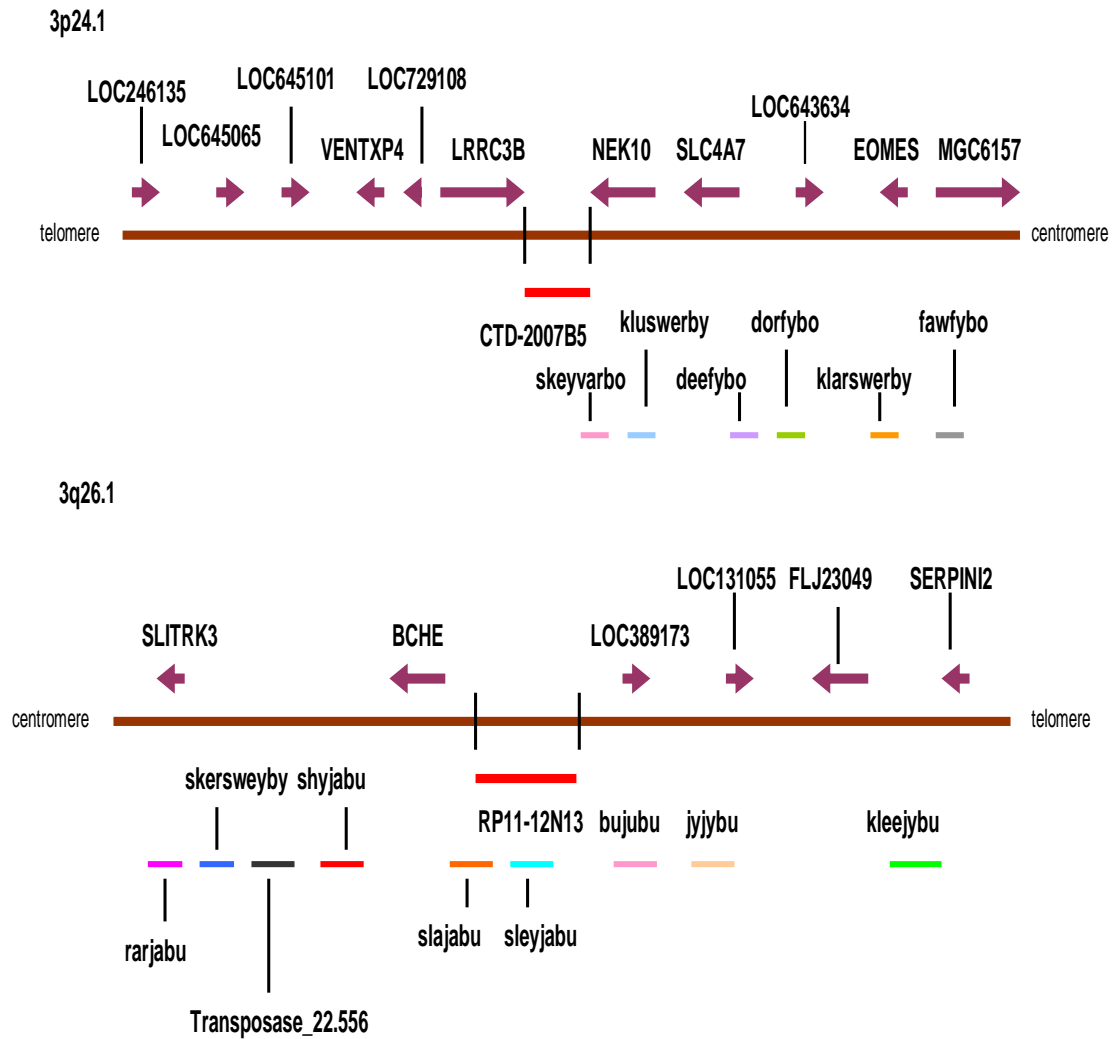
Fig. 40: Genomic structure of the two breakpoint regions.

The first brown line represents the gene details in 3p24.1 breakpoint region and the second brown line represents the gene details from 3q26.1 breakpoint region. The brown lines represent the genomic sequences. The red line below shows the BAC, which anchored the breakpoint and the name written below it. Gene sequences are indicated by arrows in the 5'-3' direction of the sense strand above the genomic sequence line. Each gene is indicated by NCBI name. The colored lines below the main line are the predicted genes from Aceview genome database and the names of each gene are written either above are below.

3.21   Identification of a putative gene in 3q26.1 region

The UCSC genome browser showed a Genscan predicted gene NT_005612.1229 spanning the 5,354 bp 3q breakpoint region. This gene covered a genomic distance of 92,125 bp from 167,425,251-167,517,375 bp on the human chromosome 3q. It showed a predicted mRNA sequence of 483 bp and a predicted protein of 160 amino acids. In order to further study this gene BLAST analysis was performed. Three exons were found on the BAC RP11-12N13. The first exon was 164 bp positioned from 82183-82346 bp, second exon was 142 bp positioned from 122,937-123,080 bp and the third exon was 171 bp positioned from 174,131-174,302 bp on the BAC clone RP11-12N13 (Fig. 41). The breakpoint (144,286-149,639 bp) was found to be interrupting the second intron of this putative gene.

3.21.1   Identification of nine novel exons of the putative gene

Four sequences were analyzed in order to achieve the nine novel exons. The first sequence was the mRNA of the putative gene NT_005612.1229. Blast analysis of this mRNA showed homology to a second sequence of predicted mRNA XR 015783.1 (LOC732249) which was 803 bp. The first and the second sequence showed homology to a 164 bp sequence which was the first exon of the putative gene. The remaining 639 bp unknown sequences when blasted showing homology to a third predicted mRNA XR_017393.1 (LOC389173) of 689 bp. This showed homology to phosphoserine aminotransferase isoform 1 (LOC389173) mRNA and to a genomic sequence on BAC RP11-450H5 (AC07046.11) on chromosome 3q26.1. This indicated the possibility of extension of this putative gene still further. A homology of 647 bp was formed with the third sequence on the BAC RP11-450H5 from position 131,466-132,104 bp and the initial 42 bp sequences showed homology proximally on the same BAC from 131,149-131,190 bp. Thus two new exons of 45 bp and 647 bp were identified. All these exons covered a distance of 467,054 kb on the genomic DNA. In the process of analyzing this extended genomic region a fourth sequence of the Genscan predicted gene NT_005612.1230 was identified. This was adjacent and distal to the first Genscan gene covering a genomic size of 431,711 bp. Its mRNA showed 1,377 bp and protein was predicted to be 458 amino acids. It showed 8 exons of sizes 62 bp, 106 bp, 177 bp, 565

bp, 104 bp, 68 bp, 145 bp and 128 bp. The 565 bp exon showed homology to the identified 647 bp exon. Hence a total of 12 exons were formed as a part of a single gene structure (Fig. 41) and their sizes and position on chromosome 3 are shown in detail in Tab. 14.
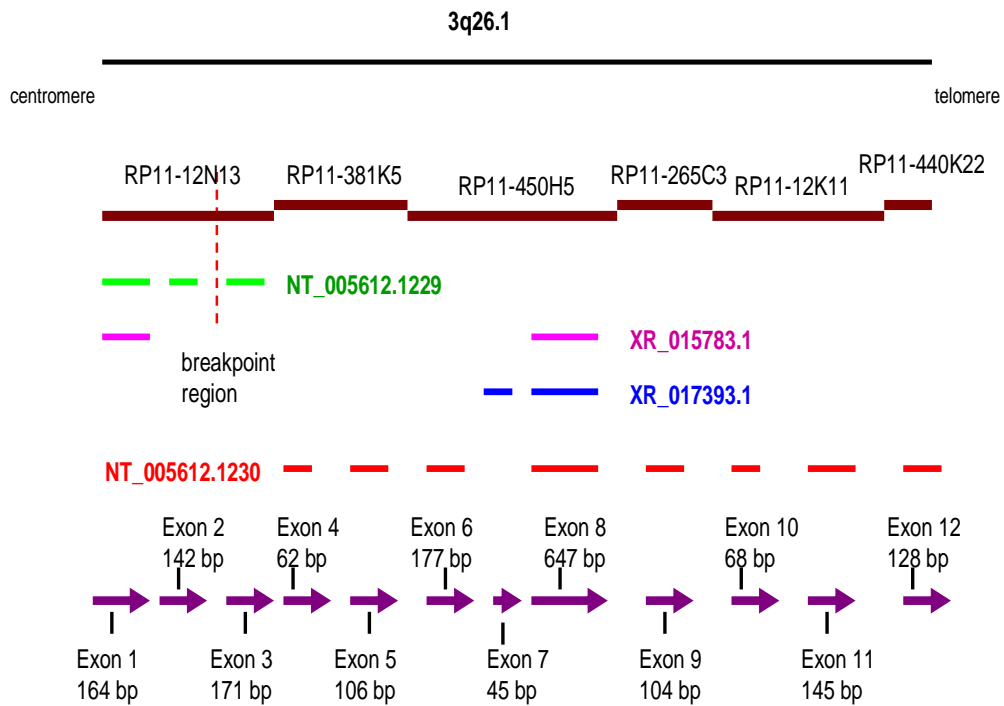
Fig. 41: Schematic representation showing the formation of 12 exons and their position on different BACs.

The first black line represents the 3q26.1 region. The BACs anchoring the different exons are shown below as brown lines with their names above each line. The three green color lines represent exons of mRNA NT_005612.1229 the two pink color lines represent the exons of mRNA XR_015783.1, the two blue color lines represent exons of XR_017393.1 mRNA sequences and the eight red color lines represent the exons of mRNA NT_005612.1230. The exons with their orientation are shown as violet colored arrows with the sizes mentioned either above or below the arrows. The breakpoint region is shown as a red vertical interrupted line.

Tab. 14: The sizes of all the 12 putative exons with their position on the genomic clones.

BAC clone	Exon	Size (bp)	Start position on Chromosome (bp)	End position on Chromosome (bp)
RP11-12N13	Exon 1	164	167,425,252	167,425,416
	Exon 2	142	167,466,008	167,466,150
	Exon 3	171	167,517,200	167,517,372
RP11-381K05	Exon 4	62	167,726,115	167,726,177
	Exon 5	106	167,746,921	167,747,027
RP11-450H05	Exon 6	177	167,871,406	167,871,583
	Exon 7	45	167,891,350	167,891,392
	Exon 8	647	167,891,659	167,892,306
RP11-265C03	Exon 9	104	168,011,519	168,011,623
RP11-12K11	Exon 10	68	168,129,034	168,129,102
	Exon 11	145	168,132,261	168,132,406
RP11-440K22	Exon 12	128	168,157,698	168,157,826

3.21.2   Gene structure of the novel putative gene

The structure of the novel putative gene covers a genomic size of 732,573 bp and is described in Fig. 42. It has 12 exons with sizes of 164 bp, 142 bp, 171 bp, 62 bp, 106 bp, 177 bp, 45 bp, 647 bp, 104 bp, 68 bp, 145 bp and 128 bp. The 12 exons are separated by 11 introns and almost all the introns follow the presence of GT at the 5` splice site and AG at the 3` splice site, giving rise to the so-called GT-AG rule exception being the intron 3 which follows the AT-AC splicing mechanism (Wu et al., 1999). The details of the slice donor and acceptor sites with the consensus sequences and the presumptive branch sites are shown in the Tab. 15. Though, intron 9 formed the GT-AG rule there was no presumptive branch site identified. The promoter sequence TATAAA was found 28 bp upstream from exon 1 but the polyadenylation site was not identified. As the cis-element is a TATA box the associated trans-acting factor could be TFIID. The putative mRNA

sequence is 1959 bp which shows a single ORF (Open Reading Frame) in the NCBI ORF Finder. This ORF predicts a polypeptide sequence of 652 amino acids with a calculated molecular weight of 169.99 kilo Daltons (Fig. 43). The 3q inversion breakpoint was found disrupting the second intron of this identified putative gene.
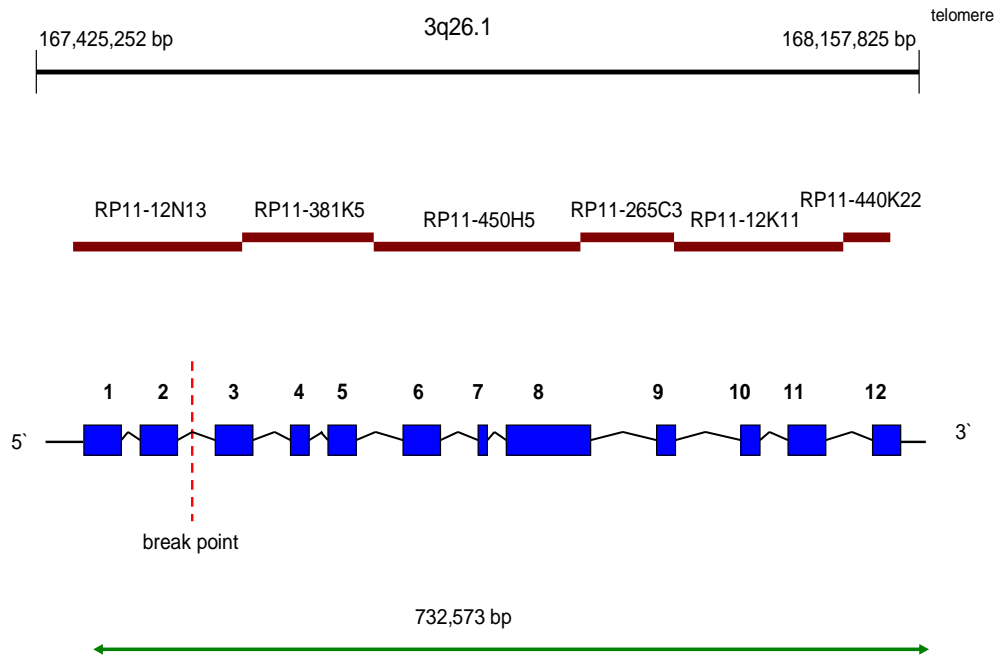
Fig. 42: Schematic representation of the exon-intron structure of the novel putative gene on 3q26.1 breakpoint region. The first black line represents the genomic sequences on chromosome 3. The positions of the sequences are indicated by black vertical lines. The brown lines show the BACs anchoring the exons. The blue boxes indicate the exons and the connecting zigzag black lines indicate the introns. The orientation of the gene is 5'-3' from centromere to telomere. The vertical red interrupted line indicates the breakpoint region and the bottom green arrowhead line shows the size of the putative gene.

Tab.15: The splice sites and the sizes of the 11 introns.

The sequences around the 5' splice site, 3' splice site and presumptive branch site of the introns are shown. Intron sequences are shown in lowercase letters, flanked by two exon nucleotides on each side, in uppercase letters. The consensus positions at the 5'-splice-site, 3'-splice-site and presumptive branch site sequences are shown in boldface for nearly invariant nucleotides. The distance refers to the number of nucleotides from the branch nucleotide to the 3' splice junction. NA refers to "not available".

Intron no.	5' splice site	Presumptive Branch site	3' splice site	Distance (bp)	Size(bp)
1	AG**gta**ggta	**ctttaat**	tg**tttag**TC	21	40,590
2	CT**gtgag**gc	**ctttaat**	aa**atag**CA	9	51,050
3	CT**at**tttga	a**cctt**tg	ggcca**ac**AT	25	208,743
4	TG**gtaag**a	**cttcaat**	**cctctag**TA	92	20,744
5	CT**gtaag**t	**ttctaat**	**ccctag**GT	36	124,379
6	GC**gtt**atag	**tgttaac**	ta**aag**TA	44	19,764
7	AA**gta**tat	**atctaac**	aa**gtag**CA	19	267
8	TG**aacat**	**caccgat**	ctg**gcag**GC	68	119,213
9	GG**gtaa**agt	NA	**ctgctag**AA	NA	117,411
10	AG**gtaag**ga	**tactga**g	**ccctcag**GT	34	3,159
11	AG**gta**ttgt	**tattaat**	**ccctccag**TT	43	25,292

```
  1 atgccctttggaacccctaaggggccccagatgtcaataatacca
    M   P   F   G   T   P   K   G   P   Q   M   S   I   I   P
 46 tatggtcccatcaagtattgtgagaagttctgctttgccactctc
    Y   G   P   I   K   Y   C   E   K   F   C   F   A   T   L
 91 ggattttcttttt ccattcacctggaccgttctgactcttctgta
    G   F   S   F   S   I   H   L   D   R   S   D   S   S   V
136 gattctgcgcctgtttgcttgaatgtcagtcaaactaatggtgga
    D   S   A   P   V   C   L   N   V   S   Q   T   N   G   G
181 gtactaggtttaccagcagttttaaacttctgtggaaacagctat
    V   L   G   L   P   A   V   L   N   F   C   G   N   S   Y
226 tattcttcagaagccaatgagttgagttatgattatgacatgctg
    Y   S   S   E   A   N   E   L   S   Y   D   Y   D   M   L
271 aaactatcacttcgttataaagccattatagagtctcaaatgatc
    K   L   S   L   R   Y   K   A   I   I   E   S   Q   M   I
316 atagaggatcagatgaaagaagtagcttcacggaacaaccactgg
```

```
          I   E   D   Q   M   K   E   V   A   S   R   N   N   H   W
     361  agtgggacaagagtaaaagagaacattcaaaagccattgcaagct
          S   G   T   R   V   K   E   N   I   Q   K   P   L   Q   A
     406  gaaagtctccagtggattggctcctgggaaattcaccgccttggg
          E   S   L   Q   W   I   G   S   W   E   I   H   R   L   G
     451  gcagaaggaagaataggaaaatttactatgagtcacaaatgcaca
          A   E   G   R   I   G   K   F   T   M   S   H   K   C   T
     496  ctatatatggttgtggctagttttgaaaatgcattttgtccatgt
          L   Y   M   V   V   A   S   F   E   N   A   F   C   P   C
     541  attcctgctgagcttagacattgtctgggagcagcccaaagagaa
          I   P   A   E   L   R   H   C   L   G   A   A   Q   R   E
     586  tatgacctcagtgttgcagtagatcctgaaggcgtgaaagctgga
          Y   D   L   S   V   A   V   D   P   E   G   V   K   A   G
     631  ggctattgtctgactgtcttctggtcacttgaagcttctggccta
          G   Y   C   L   T   V   F   W   S   L   E   A   S   G   L
     676  tcaagctcacaagtgggctcgcagactggcatctggctctttgct
          S   S   S   Q   V   G   S   Q   T   G   I   W   L   F   A
     721  tatgtttt cacctatcaagaatctatttacttactttctctgcat
          Y   V   F   T   Y   Q   E   S   I   Y   L   L   S   L   H
     766  ctcgtcaaaggaatccgtggtcttagagtttggcaatgctcccat
          L   V   K   G   I   R   G   L   R   V   W   Q   C   S   H
     811  gttaggtggcagtatatggtattaaaaaccgaccttgctaatgta
          V   R   W   Q   Y   M   V   L   K   T   D   L   A   N   V
     856  aacatacataaacagcaaaggcattttgctggtgcccagaagaat
          N   I   H   K   Q   Q   R   H   F   A   G   A   Q   K   N
     901  gttggctctgctggggtcaccgtggagattgtccatgatgacctg
          V   G   S   A   G   V   T   V   E   I   V   H   D   D   L
     946  ctgggatttgccctcctagagtgcccctcagtcctggaatacaag
          L   G   F   A   L   L   E   C   P   S   V   L   E   Y   K
     991  gtgcaggctggaaacagctccttgtacaacatgcctccatgtttc
          V   Q   A   G   N   S   S   L   Y   N   M   P   P   C   F
    1036  agcatctacgtcagggggcttggtcctggagtcgattaaaaacaat
          S   I   Y   V   R   G   L   V   L   E   S   I   K   N   N
    1081  ggaggtgttgcggccatggagaagcgtagctccatcaaatctcaa
          G   G   V   A   A   M   E   K   R   S   S   I   K   S   Q
    1126  atgatttatgagattattgataattctcaaggattctacatatgt
          M   I   Y   E   I   I   D   N   S   Q   G   F   Y   I   C
    1171  ccagtggggggccccaaaatagaagcaagatgaatattccattctgc
          P   V   G   P   Q   N   R   S   K   M   N   I   P   F   C
    1216  attggcaatgccaaaggagatgatgctttagaaaaaagatttctt
          I   G   N   A   K   G   D   D   A   L   E   K   R   F   L
    1261  gataaagctcttgaactcaatatgttgtccttgaaagggcataag
          D   K   A   L   E   L   N   M   L   S   L   K   G   H   K
    1306  tctgtgggaggcatctgggcctctctgtataatgctgtcacgatt
          S   V   G   G   I   W   A   S   L   Y   N   A   V   T   I
    1351  gaagatgttcagaagctagccgccttcatgaaaaaattttggaga
          E   D   V   Q   K   L   A   A   F   M   K   K   F   W   R
    1396  tgcatcagctacaaacacatcctaaccaggatatactctgttctt
          C   I   S   Y   K   H   I   L   T   R   I   Y   S   V   L
    1441  gaacaacatacaaagtttaaagtaaattggggatggctagaaaaa
          E   Q   H   T   K   F   K   V   N   W   G   W   L   E   K
    1486  gttaacaaacacagtatttttctcaaatggcaccatgtggaaaaa
          V   N   K   H   S   I   F   L   K   W   H   H   V   E   K
    1531  tttaaacaagaaaaaactgtggtagaggtgccgaatgaggaaaag
          F   K   Q   E   K   T   V   V   E   V   P   N   E   E   K
    1576  gagtgtgtggtggaagaagttctccactgttgtgcggtgaaggaa
          E   C   V   V   E   E   V   L   H   C   C   A   V   K   E
    1621  gagactcacaggactcagaaggaaaacagcaataagaaaagaaaa
          E   T   H   R   T   Q   K   E   N   S   N   K   K   R   K
```

```
1666 tttgagcagggaagaagtaaggtgtggatgaaacgaggcttcacc
     F   E   Q   G   R   S   K   V   W   M   K   R   G   F   T
1711 cctgagtctgcaaccttagacaagtttctccacaatgctactctg
     P   E   S   A   T   L   D   K   F   L   H   N   A   T   L
1756 agatcaagatctcaactgtacatcctggagacaacacaacaacac
     R   S   R   S   Q   L   Y   I   L   E   T   T   Q   Q   H
1801 aagggaataaaatcttatacaactgtcagagttggcatgaatcta
     K   G   I   K   S   Y   T   T   V   R   V   G   M   N   L
1846 cgttttgatctccccaacaaagacgtgtgccagagacctcagaac
     R   F   D   L   P   N   K   D   V   C   Q   R   P   Q   N
1891 gcaagctgtggaccaaagcagctgtgggcactggcagctaatgca
     A   S   C   G   P   K   Q   L   W   A   L   A   A   N   A
1936 gtgggaagattatggaaattctga 1959
     V   G   R   L   W   K   F   *
```

Fig. 43: The Nucleotide and amino acid sequence of the putative mRNA.

The deduced amino acid sequence is shown below the nucleotide sequence.

3.21.3 Characterization of the putative gene

In order to characterize the putative gene expression RT-PCR of the cDNA libraries from human testis and brain were analyzed. Several exon specific primers were designed and amplifications were performed on human testis and brain cDNAs. No PCR products were amplified. Due to this observation and the fact that no EST sequence corresponded to this region further characterization of this putative gene was not proceeded.

4   Discussion

4.1   Genomic structure of the inversion breakpoint region

Disease genes have been identified and isolated by different approaches and the technique of positional cloning has become a familiar component of modern human genetics research (Collins, 1995). One method of identifying disease genes is to clone the breakpoint regions of the chromosomal rearrangement associated with the phenotype and generate a detailed physical map (Tadin et al., 2004). In most cases one of the chromosomal breakpoints maps within or close to the respective gene thus disturbing its gene activity. An example is the inversion event disrupting the factor VIII gene causing severe Haemophilia A (Lakich et al., 1993). The disease associated balanced chromosomal rearrangements, which truncate, delete, or otherwise inactivate specific genes, have been instrumental for positional cloning of many disease genes (Tommerup, 1993; Collins, 1995). So far, over 100 disease loci have been identified based solely on their chromosomal position. An example is Duchenne muscular dystrophy (Monaco et al., 1986). Here, a familial pericentric inversion of chromosome 3 with short-stature was characterized by positional cloning strategy. By FISH the breakpoint regions were localized on 3p24.1 and 3q26.1, thus refining the previous karyotype as 46,XX, inv(3)(p24.1q26.1). Evidently, the 3q26.1 breakpoint region is involved in many inter and intra chromosomal rearrangements as detailed in 1.5. It is also well documented in cancers both as recurrent and non-recurrent rearrangements. A cancer fusion cluster Ecotropic virus integration-1 (EVI1) has been reported in individuals with AML involving translocations or inversions of chromosome 3 long arm (Levy et al., 1994). Recently, Muzny et al. (2006) mapped a large number of cancer lesions to chromosome 3. They also compared those cancer cases with structural chromosomal aberrations and assigned them to individual chromosomal band positions (Fig. 44). In this study the 3q inversion breakpoint mapped to the 3q26 cluster region. Incidentaly, this is the second largest hot spot region for chromosomal rearrangements on chromosome 3.
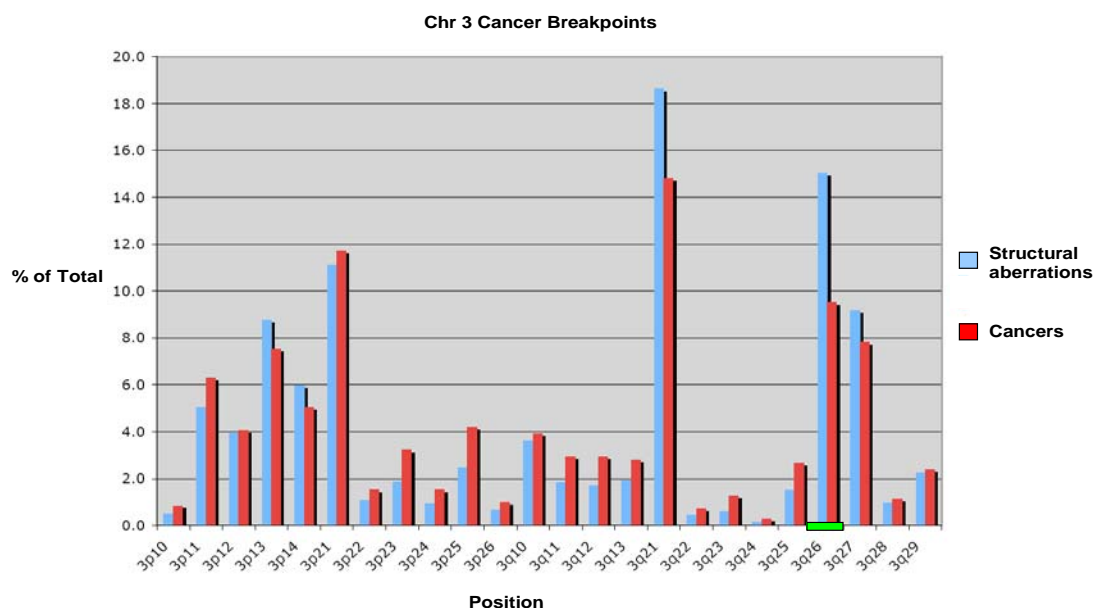
Fig. 44: Percentage of chromosome 3 total chromosomal abnormalities and cancer breakpoints versus chromosome band position.

The green block shows the 3q inversion breakpoint region (adapted from Muzny et al., 2006).

4.1.1 The inversion breakpoint and paralogous nature

Another hotspot for the occurrence of non-allelic homologous recombinations or unequal crossing over leading to genomic mutation are segmental duplications or duplicons which are the segments of DNA with near-identical sequences in the human genome (Cheung et al., 2003). A variety of mutational events including small-scale (single-base pair changes, microsatellite slippage, insertion/deletions) and large-scale events (retrotransposition, chromosomal rearrangements, segmental duplications) are thought to have molded the human genome. Segmental duplications and low-copy repeats (LCRs) play a role in several genomic disorders forming paralogous recombination events giving rise to contiguous gene syndromes like Smith-Magenis syndrome on 17p11.2 and Charcot-Marie-Tooth syndrome 1A (Reiter et al., 1997). A paralogous nature describes two distinct nonallelic genomic segments, with highly similar primary DNA sequence, which is derived from a duplication event. Segmental duplications are rich in pericentromic and subtelomeric regions, pericentromic duplications are rarely found in chromosomes 3, 4

and 19 (Bailey et al., 2002; She et al., 2004). Previous studies have suggested that recent segmental duplications, which are often involved in chromosome rearrangements underlying genomic disease, account for about 5% of the human genome. In general, chromosome 3 is relatively devoid of segmental duplications having a low percentage of 1.7% in the genome (Muzny et al., 2006). In this study the inversion breakpoint regions were localized to 3p24.1 and 3q26.1 and the chromosomal view of these regions showed no segmental duplications. Further, in the process of identifying the junction fragments by PCR several products were amplified showing sequence homology to other chromosomes. Here two PCR products homologous to chromosome 1 and 21 are discussed. A 500 bp PCR amplified product showed 100% similarity to an IMAGE clone 5266681 and a BAC clone RP11-1086F11 on chromosome 1q44. This BAC clone contained a part of the gene for a novel protein (FLJ32001), a novel gene and a ribosomal protein L7a (RPL7A) pseudogene. To check why the homology with chromosome 1 occurred, this region was further extensively analyzed for ancient duplications between chromosome 3 and 1 (from http://wolfe.gen.tcd.ie/dup/human5.28) (Fig. 46). A paralogous cluster of 11.9 Mb of chromosome 1p was identified with an approximately 40 Mb of 3q region. The identified cluster is one of the 20 large paralogous blocks in the human genome. The ENSP00000295830 or BCHE gene on chromosome 3 which is an adjacent gene to the 3q breakpoint region showed homology to another gene RPL22 on chromosome 1p36.1 (Fig.45). These results confirmed the ancient paralogous relation with chromosome 1. Interestingly, the 3q inversion breakpoint region was anchored within this paralogous block. Although the paralogous nature of both the chromosomes is discussed, the significant similarity of 3q genomic sequence in this study was found to be on 1q44 region (Fig. 46). Our findings suggest that within this paralogous block there are segments which do not map to 1p region alone.

With the first PCR product mapped to the paralogous block on 1q region the second 2.5 kb PCR amplified junction product showed homology to chromosome 21 which could be explained in terms of the evolutionary aspect. The genome evolution is marked by chromosomal synteny conservation, and the association of human chromosome 3 and 21 is the largest conserved syntenic block known for mammals and may be considered as ancestral for all placental mammals (Mueller et al., 2000). Collectively, the evolutionary aspect and the 3q inversion breakpoint region showing paralogous nature to a large block

of chromosome 1, supports the fact that the duplication architecture of the mammalian genome plays a fundamental role in shaping the process of evolution, genomic diversity, and phenotype.
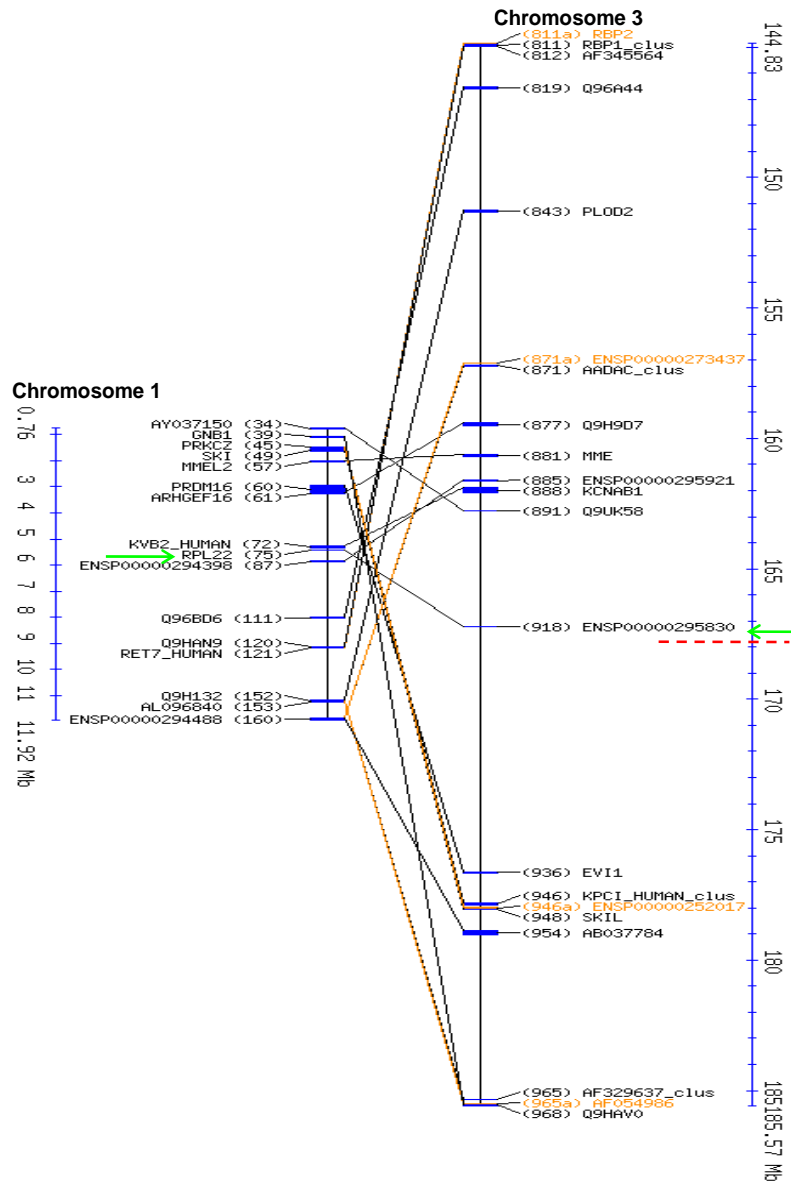

Fig. 45: Graphical view of the duplicated genes on paralogous block of chromosome 3.

The paralogous block of chromosome 3 with chromosome 1 is compared. The blue lines shows the similarities of the paralogous block and the blue vertical lines show the gene similarity. The red vertical interrupted line shows the 3q breakpoint region and the green arrows shows the similarities of the gene adjacent to the 3q breakpoint region (adapted from http://wolfe.gen.tcd.ie/dup/human5.28).
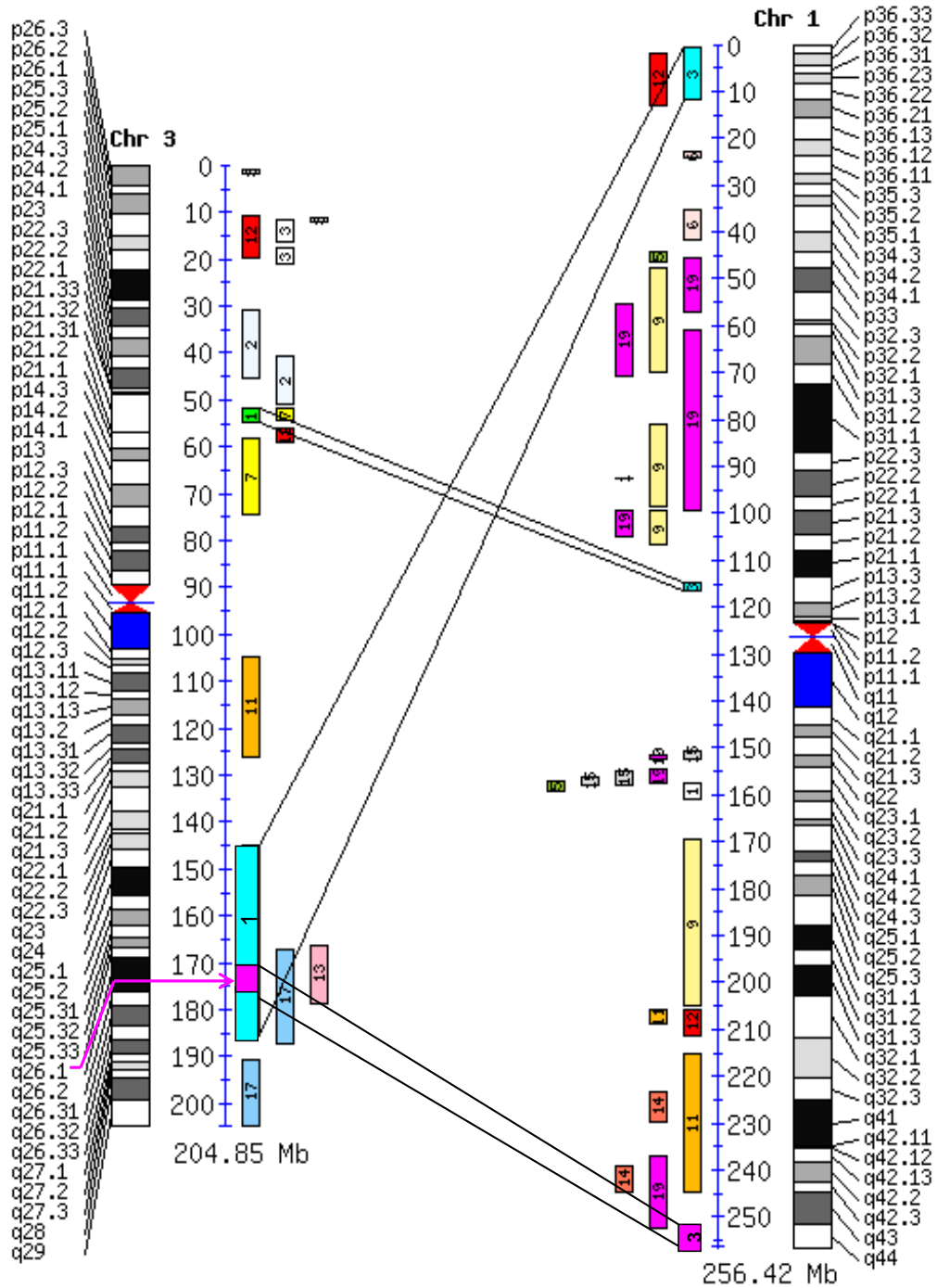
Fig. 46: Ancient duplications mapped between chromosome 3 and 1.

Colored boxes denote paralogous clusters of six or more genes by chromosome. The cluster at position 144 Mb to 185 Mb on chromosome 3 is shared with a cluster on chromosome 1 at 0.76 Mb to 11.9 Mb. The pink arrow shows the 3q breakpoint region and the homology on 1q is shown as a pink color box (adapted from http://wolfe.gen.tcd.ie/dup/human5.28/chrom_plot).

4.1.2   Association of repetitive elements with the inversion breakpoints

Analogous to the paralogous segments, the repeats represent an extraordinary trove of information about biological roles such as gene expression, genome structural stabilization and recombination (Murakami et al., 2004). They also constitute a rich palaeontological record holding clues about evolutionary events and forces which provide assays for studying processes of mutation and selection (IHGSC, 2001). Most human repeats in the genome are derived from transposable elements (Smit, 1999; Prak et al., 2000) and 45% belong to this class with long interspersed repeats (LINEs), short interspersed repeats (SINEs), long terminal repeat (LTR) retrotransposons and DNA transposons. SINEs, LINEs and LTRs are propagated by reverse transcription of an RNA intermediate but DNA transposons propagate through a "cut-and-paste" mechanism of DNA sequences. The contribution of the transposons to the human genome is that a few hundred genes, for e.g. use transcriptional terminators donated by LTR retrotransposons. Other genes employ regulatory elements derived from repeat elements (Brosius, 1999). However, analyses of the sequences surrounding some breakpoints showed the presence of multiple repetitive elements which may have mediated the rearrangement event. For example a pericentric inversion 8 breakpoint was characterized and the sequences surrounding the breakpoints showed highly repetitive LTR region, which mediated the inversion (Graw et al., 2000). In many other cases the inversions took place in regions enriched with *Alu* and LTR elements suggesting that repetitive elements themselves are recombinogenic and that they contributed to evolutionary rearrangements (Kehrer-Sawatzki et al., 2002). Further in some cases such as deletion of α-globin gene, leading to α-thalassemia (Ottolenghi et al., 1974; Taylor et al., 1974), recombination was shown between homologous segments of two members of a gene family. In other cases, such as in LDL-receptor (Hobbs et al., 1986), α-globin (Nichollos et al., 1987), β-globin (Vanin et al., 1983; Henthorn et al., 1986) rearrangements have been shown to be mediated by *Alu* repetitive elements. It has been proposed that *Alu* elements function as recombinatorial hotspots (Calabretta et al., 1982; Barsh et al., 1983). In addition, other repetitive elements, such as LINE-1 (L1), have also been shown to participate in chromosomal rearrangements (Toriello et al., 1996). In a case of a constitutional translocation t(8;17), resulting in isolated lissencephaly the region immediately surrounding the 17p13.3 breakpoint had five *Alu* repetitive elements and the regions

surrounding the 8p11.2 breakpoint contained three L1 sequences, leading to speculation that the repetitive elements themselves contribute to instability and recombination (Kurahashi et al., 1998). Further, Tigger-1 DNA transposon mediated recombination has been proposed as the mechanism in Charot-Marie-Tooth type 1A (CMT1A) by Reiter et al. (1997). Thereby, most recurrent chromosomal breakpoints occur in repetitive sequences that function as recombination hotspots (Kurahashi et al., 2003). In this study a 2.9 kb region around the 3p breakpoint and a 5.3 kb region around the 3q breakpoint region were analyzed. Both the breakpoints were surrounded by repetitive regions. The 3p region contained 27.5% of repeats with LTR33A, MER67C, L2 and MER67D and the 3q region contained 26.4% of repeats with LTR16C, MER20, MLT2B1, LTR1B, (TATG)n simple repeats and low complexity AT rich repeats. The percentage of LTR elements on the p region and the DNA transposons on the q region were higher than that of the average human genome. Though the repeat elements on both the regions were of different families they formed the main class of LTR elements. Apparently, this inversion event might have taken place between these LTR classes of elements thus supporting the hypothesis that repetitive elements are themselves recombinogenic. In summary, several chromosomal rearrangements have been characterized, leading to the speculation of the molecular mechanism behind such rearrangements.

4.2   Nonhomologous recombination as a molecular mechanism for inversion 3 rearrangement

The characterization of a disease associated with a chromosomal rearrangement helps in the identification of novel genes. However, the majority of the genomic rearrangements are not random events and analysis of each non-recurrent rearrangement will help to enlighten the mechanistic properties of the recombination. The presence of region-specific low-copy repeats (LCRs) can stimulate non-allelic homologous recombination (NAHR) that results in chromosome rearrangements (Stankiewicz and Lupski, 2002). LCRs play a role in formation of constitutional, evolutionary and somatic chromosome aberrations. However, NAHR is a major mechanism for human disease and the other mechanism is the Nonhomologous end joining (NHEJ) (Fig. 48) (Roth et al., 1986). In some cases analysis of the products of the recombination at the junctions of the

rearrangements revealed both homologous recombination and nonhomologous end joining as causative mechanisms (Shaw and Lupski, 2004). In general, mispairing between inverted repeats on homologous chromosomes results in an inversion when crossover involves both repeats (Fig. 47) (Shaffer et al., 2000).
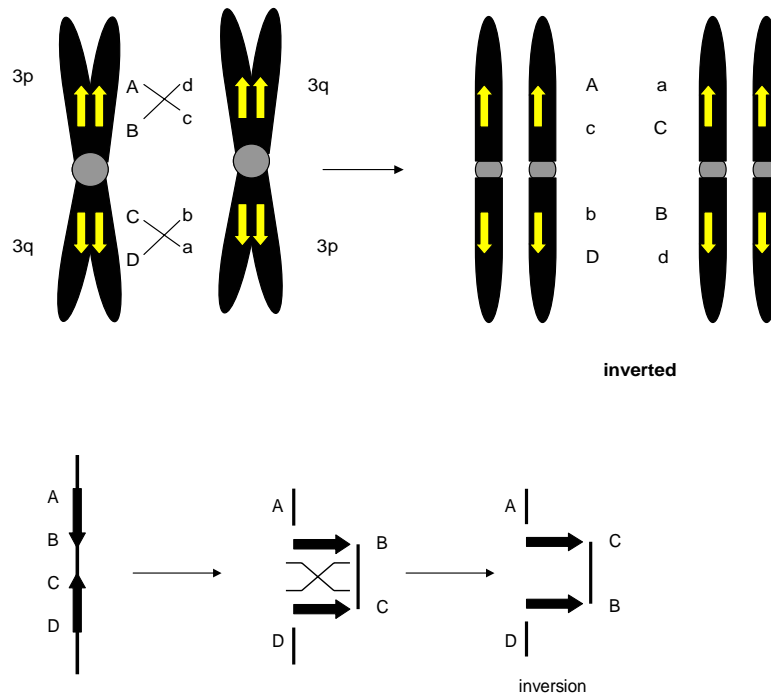

Fig. 47: The pictorial and diagrammatic representation of low-copy repeat/nonallelic homologous recombination (LCR/NAHR)-based mechanisms for pericentric inversion.

The chromosome 3 is shown in black; both normal and inverted, with the centromere in grey and yellow arrows depict LCRs. The figure depicts LCRs in the inverted orientation. The interchromosomal alignment of inverted repeats is shown here (adapted from Stankiewicz and Lupski, 2002).

Various types of recombinogenic motifs such as DNA polymerase pause site, immunoglobulin heavy chain class switch repeat, heptamer recombination signal sequences, translin binding sites etc. were shown to promote nonhomologous recombination (Abeysinghe et al., 2003). For example in many of the inversions analyzed at the nucleotide level, short common sequences like chi-like sequences, sequences homologous to topoII sites and sequences homologous to V(D)J heptamers and

nonamers, Alu repeats etc. were identified which participated in the formation of the secondary structures and facilitated the recombination process (Todd et al., 2001). In the present study, analyses of the DNA sequences at the inversion breakpoint region revealed the presence of repetitive elements. Primarily, this suggests that the repetitive elements are involved in the mechanism of inversion. The repeat elements identified on both the breakpoint regions are different and nonhomologous. Eventually, this suggests that the inversion had occurred by nonhomologous (illegitimate) recombination. How this non-homologous recombination might have occurred can be addressed by considering some of the possible mechanisms. It is of interest to note that in the 3q26.1 breakpoint region two AT-rich low complexity repeat regions were identified. The palindromic nature of the sequence was analyzed in the DNA-palindromic finder (http://www.alagu-molbio.net/palin.html). A mechanism with Palindromic AT-Rich Repeats (PATRR) is described here. PATRR are highly unstable because such sequences can replicate, leading to cleavage of the DNA and loss of the hairpin, resulting in stabilization of the locus (Gordenin et al., 1993). Further such PATRR form a cruciform structure that induces the genomic instability leading to the rearrangements resulting in illegitimate recombination events (Kurahashi et al., 2003). For example in normal chromosomes 11, 17 and 22 these events were already reported (Fig. 49) (Edelman et al., 2001). The mechanism of this PATRR-mediated double strand breaks is unknown but it was hypothesized that a small disruption of symmetry in the palindrome could subsequently stabilize the locus (Akguen et al., 1997) and two models were explained. In one model, replication slippage could occur, generating two-sided palindrome deletions spanning the central axis of symmetry or tip of the hairpin. This is analogous to the mechanism described for bacteria and yeast. In the second model, a single-stranded nick in the tip of the hairpin would result in a double strand break, and this might serve as an initiating event for illegitimate recombination. Recently, Kato et al. (2006) demonstrated a mechanism that genetic variations play an important role in PATRR mediated translocations. Further, it has been proposed that the repetitive sequence elements such as direct repeats, inverted repeats, symmetric elements, and inversions of inverted repeats may be involved in the formation of secondary structure intermediates between single-stranded DNA ends that recombine during rearrangements (Chuzhanova et al., 2003). Interestingly, it was also shown by Gotter et al. (2004) that a putative secondary structure sequence surrounding the

translocation breakpoint with low copy repeats revealed the presence of palindromes, forming stem-loop structures that did not share large sequence homology. From this it was suggested that illegitimate recombination might occur between regions having similar characteristics but not necessary sequence homology (Gotter et al., 2004). In summary, though the exact breakpoint region was not known at the nucleotide level due to difficulty in PCR amplification, perhaps because of its repeat nature (Vishwanathan et al., 1999) and as 3q26.1 breakpoint region showed two AT rich regions the probability of forming cruciform structures inducing the genomic stability leading to the rearrangement cannot be ruled out. The fact that all the repeats on both the breakpoint regions belong to different families showing no homology explain clearly the illegitimate recombination. Thus, PATRR mediated non-homologous recombination can be speculated as one of the molecular mechanism for this inversion, by suggesting the second model of illegitimate recombination.
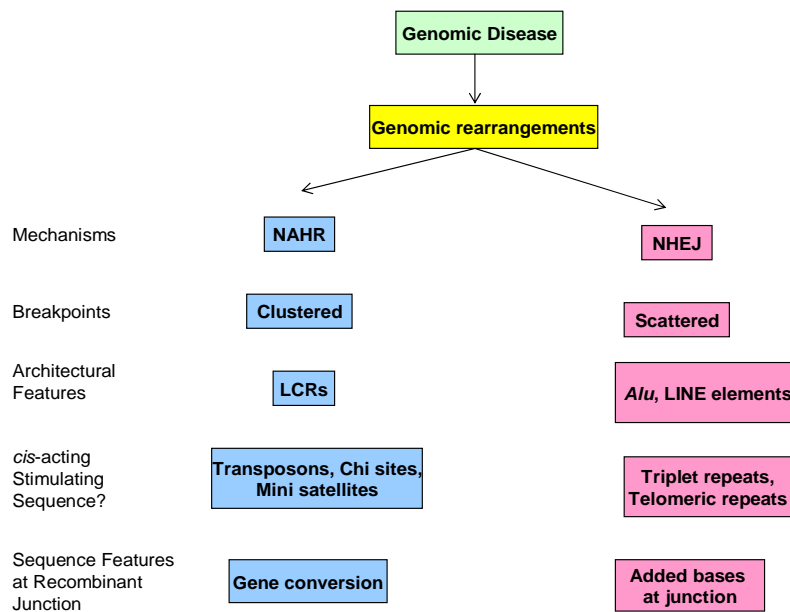

Fig. 48: Mechanisms of genomic rearrangements.

Two primary recombination mechanisms, NAHR (blue) and NHEJ (red), are shown. Features associated with NAHR or NHEJ are shown in blue and red, respectively (adapted from Shaw and Lupski, 2004).

Alternatively, the inversion could also be the result of incorrect repair of two simultaneous double-strand breaks (DSBs) by the nonhomologous end-joining (NHEJ) repair pathway (Richardson and Jasin, 2000). The abundant repetitive elements may be responsible for the NHEJ (Shaw and Lupski, 2004). Furthermore, the familial autosomal dominant trait can be ascertained, as the genomic disorders can also manifest Mendelian traits.
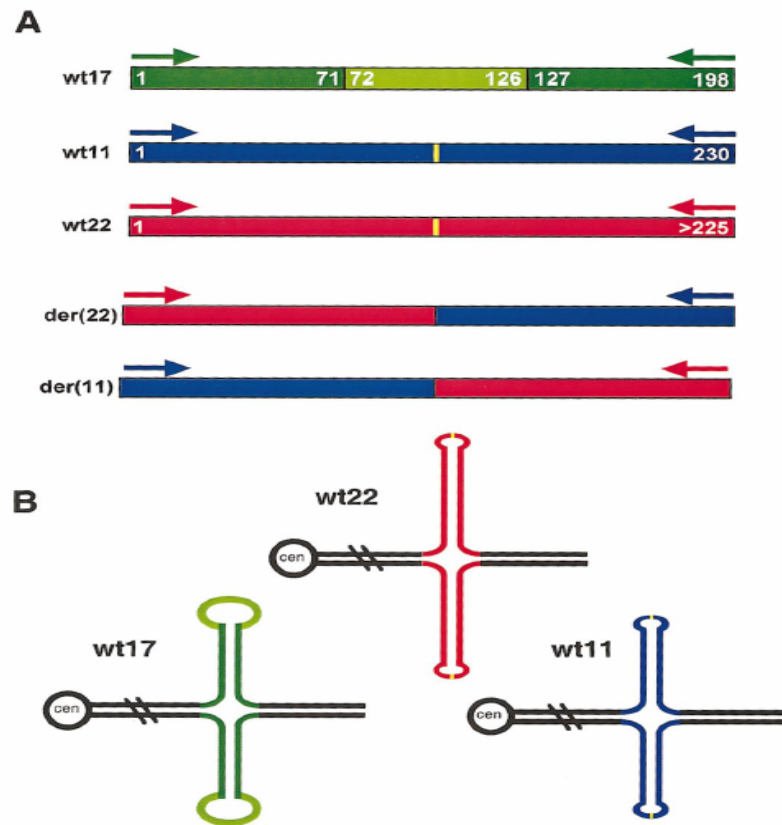

Fig. 49: Schematic representation of a PATRR mediated recombination.

A. The chromosomes 17, 11, and 22 with relative orientation of palindromic sequence in the vicinity of the chromosome breakpoint junction are depicted by the arrows. The wild-type chromosome 17 has a palindromic sequence of 71 bp (dark green) on either side of a unique, nonpalindromic sequence of 55 bp (light green). The wildtype chromosomes 11 and 22 have a palindrome structure of 230 bp (blue) and >225 base pairs (red), respectively.The yellow bar denotes the center of symmetry of the palindrome and the position of the deletions in the carriers. The derivative chromosomes represent the recombinant chromosomes, which are no longer palindromic. **B**: The palindromic sequences in the normal chromosomes 17, 11, and 22 can theoretically form hairpins or cruciform structures as indicated. The hairpin structures on chromosomes 11 and 22 can extend the entire length of the palindrome, and the loop is assumed because of steric hindrance. In contrast, there is a stem-loop structure hypothesized to form on chromosome 17, and the light green interval depicts the region that lacks base pairing (adapted from Edelmann et al., 2001).

4.3   Disease phenotype of the inversion patient

A large number of chromosomal rearrangements were shown to be associated with a disease phenotype as a result of interrupting or modifying the expression of gene(s) localized in or close to the breakpoints (Bugge, 2000). In order to identify the disease genes, breakpoints should be characterized and if any gene is disrupted by the rearrangement then it is considered as a candidate gene for the disease. Post human genome sequencing, the large scale strategies to establish the genotype-phenotype relationships also gained greater importance as it helps in identifying the unknown genes. The expression of a gene requires the presence of an intact transcription unit and also the correct function of regulatory elements. As little is known about the gene regulation in humans, one can study the rearrangements that result in diseases where chromosomal breakpoints do not disrupt the causative gene but instead map outside the intact gene (Velagaleti et al., 2005). In a number of diseases like Aniridia, X-Linked deafness, and Rieger syndrome the chromosomal breakpoints have been found to map outside disease genes (Fantes et al., 1995; de Kok et al., 1995; Flomen et al., 1998). Furthermore, mutations in the *POU3F4* (POU domain, class 3, transcription factor 4) gene were shown to cause X-linked deafness type 3 (DFN3) (de Kok et al., 1995). Evidently, five DFN3 patients were characterized which showed a microdeletion of an 8 kb DNA fragment located 900 kb upstream *POU3F4*. The molecular cause in these cases is not known, disruption of another DFN3 gene or of sequences that are involved in transcriptional regulation of the *POU3F4* gene was discussed (de Kok et al., 1996). In addition, chromosomal rearrangements that are located outside the transcription and promoter regions of a gene may cause a disease phenotype by a position effect (Kleinjan and van Heyningen, 1998). A position effect is defined as a deleterious change in the level of gene expression brought about by a change in the position of the gene relative to its normal chromosomal environment, but not associated with an intragenic mutation or deletion. Kleinjan and van Heyningen (1998) reviewed position effects of chromosomal rearrangements and proposed a few possible mechanisms that may cause such effects: (1) separation of the gene from its enhancer or promoter region, (2) juxtaposition with an enhancer element from another gene, (3) removal of the long-range insulator or boundary element, (4) competition with another enhancer, and (5) position-effect variegation that is the insertion of the gene in a new heterochromatin environment. For example, Lettice et

al. (2003) demonstrated that point mutations in a regulatory element located ~1 Mb from the target gene *Shh* which were localized in the intron of the LMBR1 gene, are capable to cause congenital abnormalities and to possess the capacity to modify gene activity. An evidence for position effects on *SOX9* associated with chromosome breakpoints mapping ~1.3 Mb downstream and ~0.9 Mb upstream of *SOX9* was also provided making it the longest-range position effect found in the field of human genetics (Velagaleti et al., 2005). These examples collectively show the position effect, though the gene is up to 1.3 Mb away from the breakpoint region. In this study the 3q inversion breakpoint region is of interest as the breakpoint is anchored between two bona fide genes where as the 3p breakpoint region had no validated genes in the near vicinity (Fig. 50). In addition, a bona fide gene SHOX2, the S̲HOX H̲O̲mologous gene on chromosome T̲hree (SHOT); a SHOX-related human gene was identified and localized on chromosome 3q25-26 (Blaschke et al; Baere et al., 1998). Though this gene is involved in short-stature, the mutation with this gene was already excluded in this inversion phenotype (Peisker, 2003). Further SHOX2 gene was found to be located ~8 Mb proximal to the 3q breakpoint. The first gene is SERPINI2 (serine protease inhibitor, clade I, member 2) which is 1 Mb distal to the 3q breakpoint region and the protein encoded by this gene plays a central role in the regulation of a wide variety of physiological processes, including coagulation, fibrinolysis, development, malignancy and inflammation. The gene product reportedly had a role in growth-control, possibly growth-suppressing pathway and, when impaired, may be involved in pancreatic carcinogenesis. A second gene BCHE for butyrylcholinesterase was found to be 449 kb away from the proximal breakpoint on 3p24.1, due to the inversion event. In summary, the 3q breakpoint occurred between the two genes BCHE and SERPINI2 and due to the inversion the BCHE gene is located on 3p region. The SERPINI2 gene is located distal to the breakpoint and has a role in growth control and especially growth suppressing pathway. Due to which it can be speculated that the breakpoint in 3q26.1 of the inversion described in this work might have caused down regulation of SERPINI2 gene which is located almost 1 Mb away from the breakpoint showing position effect. Alternatively, the inversion rearrangement might also have affected some other gene located far from the breakpoint thus causing the 'unique' phenotype of short-stature in this family.
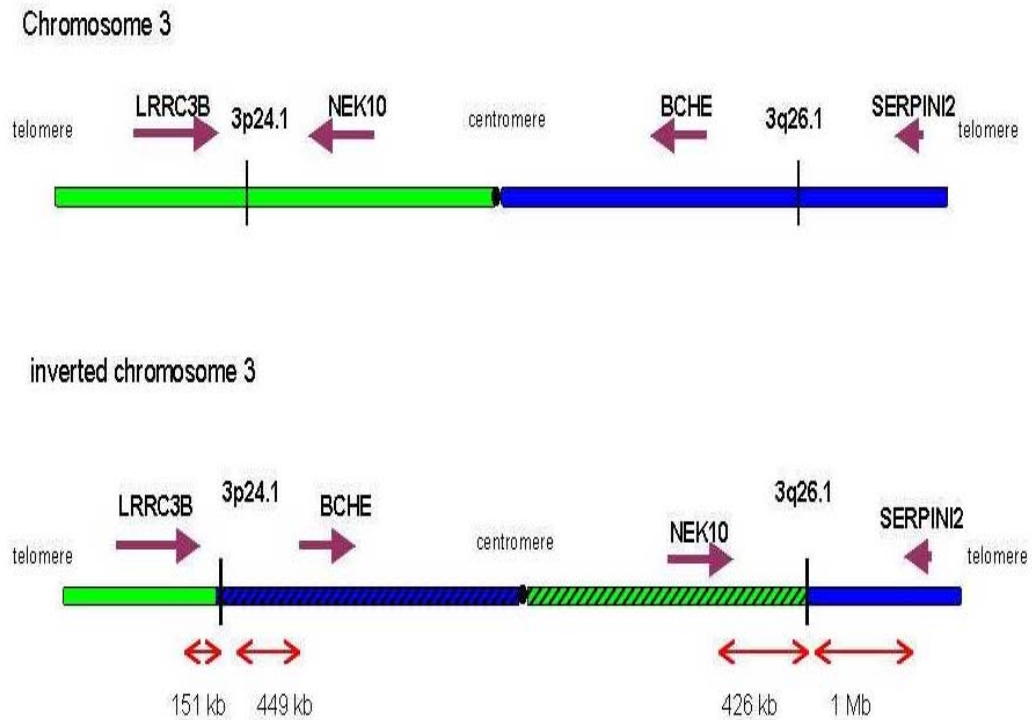
Fig. 50: The schematic representation of the gene content of inversion chromosome 3.

The upper green line represents the 3p arm and the blue line 3q arm and the breakpoints defined as vertical black lines. Gene sequences are indicated by arrows in the 5'-3' direction of the sense strand above the genomic sequence line with names above them. The dashed lines represent the inverted region along with the gene content. The red arrow headed lines below show the distance of the nearest genes to the breakpoint region.

4.4   Analysis of the putative gene

A frequent and important goal in genome research is to identify a particular gene with its structure and expression and relate these to normal and abnormal function. The availability of both genomic and cDNA clones provides the opportunity of mapping exon/intron boundaries. Once the cDNA sequence had been obtained identification of the

ORF in one of the three possible translational reading frames is essential. The 5` and 3` untranslated regions and the introns cannot be ruled out by the mutation screening approaches used routinely. For example the congenital cataract facial dysmorphism neuropathy (CCFDN) is caused by a single nucleotide substitution in intron 6 (IVS6+389C>T) of CTDP1 (Varon et al., 2003). In this study the inversion breakpoints did not disrupt the genomic structure of any bona fide genes. In the 3q26.1 breakpoint region a putative gene was identified (Fig. 44). The gene was initially predicted by Genscan with 3 exons but in this study we had possibly extended the gene and found 9 more novel exons to the existing putative gene. This putative gene structure showed 12 exons and RNA splicing followed the standard GT-AG rule except the intron 3 which followed the AT-AC splicing mechanism (Wu et al., 1999). No transcription units were identified in the two tissues studied by RT-PCR. Usually, a gene is represented by multiple ESTs that correspond to different portions of a transcript or various alternatively spliced transcripts (Schuler et al., 1996). In addition, database searches revealed no EST sequences matching to the identified putative exons in the 3q26.1 breakpoint region. Probably, the putative gene could be expressed in such part of the tissues where there is less expression and thus making it a difficult task for the construction of cDNA libraries. The predicted protein showed homology to an enzyme phosphoserine aminotransferase (PSAT1) which is involved in the phosphorylated pathway of L-serine biosynthesis. PSAT1 gene has 9 exons covering approximately 56 kb on 9q21.31. L-serine serves as a building block for protein synthesis and can be modified in different metabolic pathways for the generation of several essential compounds, including glycine, D-serine, cysteine, serine phospholipids, sphingomyelins and cerebrosides. It is also required for the synthesis of nucleotide precursors, such as purines and thymine, which are directly linked to cellular replication (Snell, 1986). The significant increase of L-serine also confers a growth advantage to tumour cells by coupling with nucleotide biosynthesis, suggesting an involvement in neoplasia (Snell and Weber, 1986; Snell et al., 1988). Baek et al. (2003) had characterized two forms of PSAT cDNA (HsPSATα and HsPSATβ). Neither the polyadenylated additional signal nor the poly(A) tail was detected in the 3`-untranslated region of both the forms. The PSAT1 gene also showed homology to PSAT1 of mouse on chromosome 19qA region. In summary, the putative gene identified in the 3q26.1 region has a 652 amino acids protein sequence which showed homology to a PSAT1 gene on

chromosome 9q21.31. The possible functions of the predicted gene in this study were known using the URL (http://dragon.bio.purdue.edu/pfp). This putative gene predicted a function in biological process like L-serine biosynthesis, Pyridoxine biosynthesis as well as in transaminase activity and phosphoserine transaminase activity in membranes. If this putative gene encoding the 652 amino acids protein can be validated, subsequent functional analysis could provide insights into the inversion breakpoint and its association with the phenotype.

4.5   Outlook of the project

It is important to characterize the chromosomal rearrangements as it leads to the identification of new genes. In this study the inversion breakpoint regions were localized on 3p24.1 and 3q26.1. These regions were further narrowed down to a 2.9 kb region on 3p and a 5.3 kb region on 3q. To further enlighten on the molecular mechanism of the inversion event a detailed analysis of the junction fragments should be carried out. In order to know the role of the identified putative gene on 3q26.1 region and its possible correlation with short stature phenotype it may be analyzed further in two ways. First, the transcripts should be analyzed for expression in different tissues or developmental stages. Second, it could also be very useful to analyze synthenic regions in mouse chromosomes to find out if there is a homologous gene.

5   Summary

In this work, three family members with a phenotype of short-stature carrying a familial inversion inv(3)(p23;q25q26) were studied. In general, balanced chromosomal rearrangements associated with a Mendelian disorder are powerful tools for mapping novel disease gene(s). Molecular analysis of the breakpoint regions may help to identify interrupted gene(s) responsible for the phenotype of the patient. By mutation analysis of the gene(s) located in the breakpoint regions in patients with the same phenotype and a normal karyotype, the disease causing gene can be identified. Thus, the gene identified functions as a candidate gene for that disease phenotype. Initially, breakpoint spanning YAC and BAC clones were identified by Fluorescence *in situ* hybridization on both 3p and 3q regions. The karyotype was thus refined to inv(3)(p24.1q26.1).

FISH with PCR amplified products confirmed the localization of the breakpoint with in a 2,934 bp fragment on 3p24.1 region and 5,353 bp fragment on 3q26.1 region. Analysis of the breakpoint regions on both the sides revealed repetitive elements. The 3p region showed LTR33A, MER67C, L2 and MER67D repeat elements and the 3q region showed LTR16C, MER20, MLT2B1, LTR1B, (TATG)n simple repeats and low complexity AT rich repeats. The repeat elements on both the sides showed no homology to each other but majority of them belonged to the same class of LTR elements. Analysis with some of the PCR junction fragments revealed a new paralogous block on 1q region as a part of the existing paralogous cluster. Thus support the homology of 3q breakpoint region with chromosome 1 and confirm the correlation with an ancient paralogous cluster.

Neither of the inversion breakpoint regions was found to disrupt a validated gene(s); however a novel putative gene was identified in the 3q26.1 region by *in silico* analysis. This putative gene showed 12 exons, with a 1,952 bp long mRNA showing a single ORF with 652 amino acids. The inversion rearrangement on the 3q26.1 region was found to disrupt the second intron of this identified putative gene. The function of this putative gene is predicted to be similar to the PSAT1 gene on chromosome 9. No transcript could be identified in the cDNA of the human brain and testis so far. Further expression analysis with different tissues could shed light on the nature of this putative gene.

6   References

Aarskog, D., Eiken, H.G., Bjerknes, R. and Myking, O.L. (1997): Pituitary dwarfism in the R271W Pit-1 gene mutation.

European Journal of Paediatrics 156: 829-834.

Abeysinghe, S.S., Chu, N., Krauzak, M., Ball, E.V. and Cooper, D.N. (2003): Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination associated motifs.

Human Mutation 22: 229-244.

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirknesse, E.F. and et al. (1995): Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence.

Nature 377: 3S-174S.

Akguen, E., Zahn, J., Baumes, S., Brown, G., Liang, F., Romanienko, P.J., Lewis, S. and Jasin, M. (1997): Palindrome resolution and recombination in the mammalian germ line.

Molecular Cell Biology 17: 5559-5570.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990): Basic local alignment search tool.

Journal of Molecular Biology 215: 403-410.

Anderson, C. (1993): Genome shortcut leads to problems.

Science 259: 1684-1687.

Baek, J.Y., Jun, D.Y., Taub, D. and Kim, Y.H. (2003): Characterization of human phosphoserine aminotransferase involved in the phosphorylated pathway of L-serine biosynthesis.

The Biochemical Journal 373: 191-200.

Baere, E.D, Speleman, F., Van, R.N., De Paepe, A. and Messiaen, L. (1998): Assignment of SHOX2 (alias OG12X and SHOT) to human chromosome bands 3q25→q26.1 by in situ hybridization.

Cytogenetics and Cell Genetics 82: 228-229.

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. and Eichler, E.E. (2002): Recent segmental duplications in the human genome.

Science 297: 1003-1007.

Ballabio, A. (1993): The rise and fall of positional cloning?

Nature Genetics 3: 277-279.

Barajas-Barajas, L.O., Velarde-Felix, S., Elizarraras-Rivas, J., Hernandez-Zaragoza, G. and Vazquez-Herrera, J.A. (2001): De novo ring chromosome 3 in a girl with hypoplastic thumb and coloboma of iris.

Genetic Counseling 12: 151-156.

Barsh, G.S., Seeburg, P.H. and Gelinas, R.E. (1983): The human growth hormone gene family: structure and evolution of the chromosome locus.

Nucleic Acids Research 11: 3939–3958.

Belin, V., Cusin, V., Viot, G., Girlich, D., Toutain, A., Moncla, A., Vekemans, M., Le Merrer, M., Munnich, A. and Cormier-Daire, V. (1998): SHOX mutations in dyschondrosteosis (Leri-Weill syndrome).

Nature Genetics 19: 67-69.

Belloma, M.J., Parlier, V., Muhlematter, D., Grob, J.P. and Beris, P. (1992): Three new cases of chromosome 3 rearrangements in bands q21 and q26 with abnormal thrombopoiesis bring further evidence to the existence of a 3q21.3q26 syndrome.

Cancer Genetics and Cytogenetics 59(2): 138-160.

Belloni, E., Muenke, M., Roessler, E., Traverso, G., Siegel-Bartelt, J, Frumkin, A., Mitchell, H.F., Donis-Keller, H., Helms, C., Hing, A.V., Heng, H.H.Q., Koop, B., Martindale, D., Rommens, J.M., Tsui, L-C. and Scherer, S.W. (1996): Identification of Sonic Hedgehog as a candidate gene responsible for holoprosencephaly.

Nature Genetics 14: 353-356.

Bentley, D.R. (2000): Decoding the human genome sequence.

Human Molecular Genetics 9: 2353-2358.

Birnboim, H.C. and Doly, J. (1979): A rapid alkaline extraction procedure for screening recombinant plasmid DNA.

Nucleic Acid Research 7: 1513-1523.

Bitter, M.A., Neilly, M.E., LeBeau, M.M., Pearson, M.G. and Rowley, J.D. (1985): Rearrangements of chromosome 3 involving bands 3q21 and 3q26 are associated with normal or elevated platelets counts in acute non-lymphocytic leukemis.

Blood 66: 1362-1370.

Blaschke, R.J., Monaghan, A.P., Schiller, S., Schechinger, B., Rao, E., Padilla-Nash, H., Ried, T. and Rappold, G.A. (1998): SHOT, a SHOX-related homeobox gene, is implicated in craniofacial, brain, heart, and limb development.

Proc. Natl. Acad. Sci. USA 95: 2406-2411.

Bouffard, G.G., Idol, J.R., Branden, V.V., Iyer, L.M., Cummingham, A.F., Weintraub, L.A., Touchman, J.W., Mohr-Tidwell, R.M., Peluss, D.C., Fulton, R.S., Veltzen, M.S., Weissenbach, J., Magness, C.L. and Green, E.D. (1997): A physical map of human chromosome 7: an integrated YAC contig map with average STS spacing of 79 Kb.

Genome Research 7: 673-692.

Bray-Ward, P., Menninger, J., Lieman, J., Desai, T., Mokady, N., Banks, A. and Ward, D.C. (1996): Integration of the cytogenetic, genetic, and physical maps of the human genome by FISH mapping of CEPH YAC clones.

Genomics 32: 1-14.

Broman, K., Murray, J.C., Sheffield, V.C., White, R.L. and Weber, J.L. (1998): Comprehensive human genetic maps: individual and sex-specific variation in recombination.

American Journal of Human Genetics 63: 861-869.

Brosius, J. (1999): Genomes were forged by massive bombardments with reteroelements and retrosequences.

Genetica 107: 209-238.

Brueton, L.A., Barber, J.C., Huson, S.M. and Winter, R.M. (1989). Partial monosomy 3q in a boy with short stature, developmental delay, and mild dysmorphic features.

Journal of Medical Genetics 26: 729-730.

Bugge, M., Brunn-Petersen, G., Brondum-Nielsen, K., Friedrich, U., Hansen, J., Jensen, G., Jensen, P.K.A., Kristoffersson, U., Lundsteen, C., Niebuhr, E., Rasmussen, K.R., Rasmussen, K. and Tommerup, N. (2000): Disease associated balanced chromosome rearrangements: a resource for large scale genotype-phenotype delineation in man.

Journal of Medical Genetics 37: 858-865.

Burke, D.T., Carle, G.F. and Olson, M.V. (1987): Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors.

Science 236: 806-812.

Calabretta, B., Robberson, D.L., Barrera-Saldana, H.A., Lambrou, T.P. and Saunders, G.F. (1982): Genome instability in a region of human DNA enriched in Alu repeat sequences.

Nature 296: 219–225.

Carbonell, F., Hoelze, D., Thiel, E. and Bartl, R. (1982): Ph1-positive CML associated with megakaryocytic hyperplasia and thrombocythemia and an abnormality of chromosome no 3.

Cancer Genetics and Cytogenetics 6: 153-161.

Carle, G.F., Frank, M. and Olson, M.V. (1986): Electrophoretic separations of large DNA molecules by periodic inversion of the electric field.

Science 232: 65-68.

Carroll, A.J., Poon, M-C., Robinson, N.C. and Crist, W.M. (1986): Sideroblastic anemia associated with thrombocytosis and a chromosome 3 abnormality.

Cancer Genetics and Cytogenetics 22: 183-187.

Cheung, J., Estivill, X., Khaja, R., MacDonald, J.R., Lau, K., Tsui, L-C., Scherer, S.W. (2003): Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence.

Genome Biology 4: R25.

Chuzhanova, N., Abeysinghe, S.S., Krawczak, M. and Cooper, D.N. (2003): Translocation and gross deletion breakpoints in human inherited disease and cancer II: Potential involvement of repetitive sequence elements in secondary structure formation between DNA ends.

Human Mutation 22: 245-251.

Clement-Jones, M., Schiller, S., Rao, E., Blaschke, R.J., Zuniga, A., Zeller, R., Robson, S.C., Binder, G., Glass, I., Strachan, T., Lindsay, S. and Rappold, G.A. (2000): The short stature homeobox gene SHOX is involved in skeletal abnormalities in Turner syndrome.

Human Molecular Genetics 9: 695-702.

Cohen, D., Chumakov, I. and Weissenbach, J. (1993): A first-generation physical map of the human genome.

Nature 366: 698-701.

Collins, F.S. (1992): Positional cloning: Let's not call it reverse anymore.

Nature Genetics 1: 3-6.

Collins, F.S. (1995): Positional cloning moves from perditional to traditional.

Nature Genetics 9: 347-350.

Cox, S.A., Attwood, J., Bryant, S.P., Bains, R., Povey, S., Rebello, M., Kapsetaki, M., Moschonas, N.K., Grzeschik, K.H., Otto, M., Dixon, M., Sudworth, H.E., Kooy, R.F., Wright, A., Teague, P., Terrenato, L., Vergnaud, G., Monfouilloux, S., Weissenbach, J., Alibert, O., Dib, C., Faure, S., Bakker, E., Pearson, N.M., Spurr, N.K. and et al. (1996): European Gene Mapping Project (EUROGEM): breakpoint panels for human chromosomes based on the CEPH reference families. Centre d'Etude du Polymorphisme Humain.

Annals of Human Genetics 60 (Pt 6): 447-486.

Crollius, H.R., Ross, M.T., Grigoriev, A., Knights, C.J., Holloway, E., et al. (1996): An integrated YAC map of the human X chromosome.

Genome Research 6: 943-955.

Dattani, M.T., Martinez-Barbera, J.P., Thomas, P.Q., Brichman, J.M., Gupta, R., Martensson, I.L., Toresson, H., Fox, M., Wales, J.H., Hindmarsh, P.C., Krauss, S., Beddington, R.S. and Robinson, I.C. (1998): Mutations on the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse.

Nature Genetics 19: 125-133.

de Kok, Y.J., van der Maarel, S.M., Bitner-Glindzicz, M., Huber, I., Monaco, A.P., Malcolm, S., Pembrey, M.E., Ropers, H.H. and Cremers, F.P. (1995): Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4.

Science 267: 685-688.

de Kok, Y.J., Vossenaar, E.R., Cremers, C.W., Dahl, N., Laporte, J., Hu, L.J., Lacombe, D., Fischel-Ghodsian, N., Friedman, R.A., Parnes, L.S., Thorpe, P., Bitner-Glindzicz, M., Pander, H.J., Heilbronner, H., Graveline, J., den Dunnen, J.T., Brunner, H.G., Ropers, H.H. and Cremers, F.P. (1996): Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4.

Human Molecular Genetics 5: 1229-1235.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Kazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. and Weissenbach, J. (1996): A comprehensive genetic map of the human genome based on 5,264 microsatellites.

Nature 380: 152-154.

Dietz, H.C., Culting, R.C., Pyeritz, R.E., Maslen, C.L., Sakai, L.Y., Corson, G.M., Puffenberger, E.G., Hamosh, A., Nanthakumar, E.J., Curristin, S.M., Stelten, G., Meyers, D.A. and Francomano, C.A. (1991): Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene.

Nature 353: 337-339.

Doggett, N.A., Goodwin, L.A., Tesmer, J.G., Meincke, L.J., Bruce, D.C., Clark, L.M., et al. (1995): An integrated physical map of human chromosome 16.

Nature 377: 335-365.

Edelmann, L., Spiteri, E., Koren, K., Pulijaal, V., Bialer, M.G., Shanske, A., Goldberg, R. and Morrow, B.E. (2001): AT-Rich Palindromes Mediate the Constitutional t(11;22) Translocation.

American Journal of Human Genetics 68: 1-13.

Ellison, J.W., Wardak, Z., Young, M.F., Gehron Robey, P., Laig-Webster, M. and Chiong, W. (1997): PHOG, a candidate gene for involvement in the short stature of Turner syndrome.

 Human Molecular Genetics 6: 1341-1347.

Enders, H. (1992): Chromosomal and genetic forms of growth failure.

Baillieres Clinical Endocrinology and Metabolism 6: 621-643.

Fahsold, R., Habash, T., Trautmann, U., Haustein, A. and Pfeiffer, R.A. (1995): Familial reciprocal translocation t(17;19)(q11.2;q13.2) associated with neurofibromatosis type 1, including one patient with non-Hodgkin lymphoma an additional t(14;20) in B lymphocytes.

Human Genetics 96: 65-69.

Fantes, J., Redeker, B., Breen, M., Boyle, S., Brown, J., Fletcher, J., Jones, S., Bickmore, W., Fukushima, Y., Mannens, M. and et al. (1995) Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype.

Human Molecular Genetics 4: 415-422.

Feinberg, A.P. and Vogelstein, B. (1983): A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity.

Analytical Biochemistry 137: 266-267.

Flomen, R.H., Vatcheva, R., Gorman, P.A., Baptista, P.R., Groet, J., Barisic, I., Ligutic, I. and Nizetic, D. (1998): Construction and analysis of a sequence-ready map in 4q25: Rieger syndrome can be caused by haploinsufficiency of RIEG, but also by chromosome breaks approximately 90 kb upstream of this gene.

Genomics 47: 409-413.

Gessler, M., Simola, K.O. and Bruns, G.A. (1989): Cloning of breakpoints of a chromosome translocation identifies the AN2 locus.

Science 244: 1575-1578.

Gordenin, D.A., Lovachey, K.S., Degtyareya, N.P., Malkova, A.L., Perkins, E. and Resnick, M.A. (1993): Inverted DNA repeats: a source of eukaryotic genomic instability.

Molecular Cell Biology 13: 5315-5322.

Gorlin, R.J. and et al. (1975): Rieger anomaly and growth retardation (the SHORT syndrome).

Birth defects original article series 11: 464.

Gotter, A.L., Shaikh, T.H., Budarf, M.L., Rhodes, C.H. and Emanuel, B.S. (2004): Ampalindrome mediated mechanism distinguishes translocation involving LCR-B of chromosome 22q11.2.

Human Molecular Genetics 13:103-115.

Graw, S.L., Sample, T., Bleskan, J., Sujansky, E. and Patterson, D. (2000): Cloning, Sequencing, and Analysis of Inv8 Chromosome Breakpoints Associated with Recombinant 8 Syndrome.

American Journal of Human Genetics 66: 1138-1144.

Green, E.D., Riethman, H.C., Dutchik, J.E. and Olson, M.V. (1991): Detection and characterization of chimeric yeast artificial-chromosome clones.

Genomics 11: 658-669.

Han, J-Y. and Theil, K.S. (2006): The Philadelphia chromosome as a secondary abnormality in inv(3) (q21q26) acute myeloid leukemia at diagnosis: confirmation of p190 BCR-ABL mRNA by real-time quantitative polymerase chain reaction.

Cancer Genetics and Cytogenetics 165: 70-74.

Henthorn, P.S., Mager, D.L., Huisman, T.H. and Smithies, O. (1986): A gene deletion ending within a complex array of repeated sequences 30 to the human beta-globin gene cluster.

Proc. Natl. Acad. Sci. USA 83: 5194–5198.

Hobbs, H.H., Brown, M.S., Goldstein, J.L. and Russell, D.W. (1986): Deletion of exon encoding cysteine-rich repeat of low density lipoprotein receptor alters its binding specificity in a subject with familial hypercholesterolemia.

The Journal of Biological Chemistry 261:13114-13120.

Hudson, T.J., Stein, L.D., Gerety, S.S., Ma, J., Castle, A.B., Silva, J., Slonim, D.K., Baptista, R., Kruglyak, L., Xu, S.H. and et al. (1995): An STS-based map of the human genome.

Science 270: 1945-1954.

Ichikawa, H., Hosoda, F., Arai, Y., Shimizu, K., Ohira, M. and Ohki, M. (1993): A *Not*I restriction map of the entire long arm of human chromosome 21.

Nature Genetics 4: 361-365.

IHGSC (International Human Genome sequencing Consortium) (2001): Initial sequencing and analysis of the human genome.

Nature 409: 860-921.

Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A. and de Jong, P.J. (1994): A new bacteriophage P1-derived vector for the propagation of large human DNA fragments.

Nature Genetics 6: 84-89.

ISCN (An International System for Human Cytogenetic Nomenclature) (1995):

Karger publishers.

Kainulainen, K., Pulkkinen, L., Savolainen, A., Kaitila, I. and Peltonen, L. (1990): Location on chromosome 15 of the gene defect causing Marfan syndrome.

The New England Journal of Medicine 323: 935-939.

Kato, T., Inagaki, H., Yamada, K., Kogo, H., Ohye, T., Kowa, H., Nagaoka, K., Taniguchi, M., Emanuel, B.S. and Kurahashi, H. (2006): Genetic Variation Affects de Novo Translocation Frequency.

Science 311: 971.

Kehrer-Sawatzki, H., Schreiner, B., Taenzer, S., Platzer, M., Mueller, S. and Hameister H. (2002): Molecular characterization of the Pericentric Inversion That Causes Differences between Chimpanzee Chromosome 19 and Human Chromosome 17.

American Journal of Human Genetics 71: 375-388.

Kleinjan, D.J. and van Heyningen, V. (1998): Position effect in human genetic disease.

Human Molecular Genetics 7: 1611-1618.

Kondo, Y., Mizuno, S., Ohara, K., Nakamura, T., Yamada, K., Yamamori, S., Hayakawa, C., Ishii, T., Yamada, Y. and Wakamatsu, N. (2006): Two cases of partial trisomy 21 (pter-pq22.1) without the major features of Down syndrome.

American Journal of Medical Genetics 140: 227-232.

Korenberg, J.R., Chen, X.N., Adamn, M.D. and Venter, C. (1995): Toward a cDNA Map of the human genome.

Genomics 29: 364-370.

Kurahashi, H., Shaikh, T., Takata, M., Toda, T. and Emanuel, B.S. (2003): The Constitutional t(17;22): Another Translocation Mediated by Palindromic AT-Rich Repeats.

American Journal of Human Genetics 72: 733-738.

Kurahashi, H., Sakamoto, M., Ono, J., Honda, A., Okada, S. and Nakamura, Y. (1998): Molecular cloning of the chromosomal breakpoint in the LIS1 gene of a patient with isolated lissencephaly and balanced t(8;17).

Human Genetics 103:189–192.

Lakich, D., Kazazian, H.H.Jr, Antonarakis, S.E. and Gitschier, J. (1993): Inversions disrupting the factor VIII gene are a common cause of severe haemoplilia A.

Nature Genetics 5: 236-241.

Langer, P.R., Waldrop, A.K. and Ward, D.C. (1981): Enzymatic synthesis of biotin labeled polynucleotides: novel nucleic acid affinity probes.

Proc. Natl. Acad. Sci. USA 78: 6633-6637.

Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E. and de Graaff, E. (2003): A long range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly.

Human Molecular Genetics 12: 1725-1735.

Levy, B.R., Parganas, E., Morishita, K., Fichelson, S., Jame, L., Oscier, D., Gisselbrecht, S., Ihle, J.N. and Buckle, V.J. (1994): DNA Rearrangements Proximal to the EVI1 Locus Associated With the 3q21q26 Syndrome.

Blood 83: 1348- 1354.

Lichter, P., Cremer, T., Borden, J., Manuelidis, L. and Ward, D.C. (1988): Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries.

Human Genetics 80: 224-34.

Lupski, J.R. (1998): Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits.

Trends in Genetics 14: 417-422.

Machinis, K., Pantel, J., Netchine, I., Leger, J., Camand, O.J., Sobrier, M.L., Dastot-Le Moal, F., Duquesnoy, P., Abitbol, M., Czernichnow, P. and Amselem, S. (2001): Syndromic short stature in patients with germline mutation in the LIM homeobox LHX4.

American Journal of Human Genetics 69: 961-968.

Magenis, R.E., Maslen, C.L., Smith, L., Allen, L. and Sakai, L.Y. (1991): Localization of the fibrillin (FBN) gene to chromosome 15, band q21.1.

Genomics 11: 346-351.

McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R. and et al., (2001): A physical map of the human genome.

Nature 409: 934-41.

Mendez, H., Opitz, J.M. and Reynolds, J.F. (1985): Introduction to the study of pre-and postnatal growth in humans: A review.

American Journal of Medical Genetics 20: 63-85.

Monaco, A.P., Neve, R.L., Collelti-feener, C., Bertelson, C.J., Kurnit, D.M. and Kunkel, L.M. (1986): Isolation of candidate cDNAs for portions of Duchenne muscular dystrophy gene.

Nature 323(6089): 646-650.

Mueller, S., Stanyon, R., Finelli, P., Archidiacona, N. and Weinberg, J. (2000): Molecular cytogenetic dissection of human chromosomes 3 and 21 evolution.

Proc. Natl. Acad. Sci. USA 97: 206-211.

Mugneret, F., Solary, E., Favre, B., Caillot, D., Sidaner, I. and Guy, H. (1992): New case of t(3;17)(q26;q22) as an Additional Change in a Philadelphia- Positive Chronic Myelogenous Leukemia in Acceleration.

Cancer Genetics and Cytogenetics 60: 90-92.

Murakami, H., Sugaya, N., Sato, M., Imaizumi, A., Aburatani, S. and Horimoto, K. (2004): Detection of Inter-Spread Repeat Sequence in Genomic DNA Sequence.

Genome Informatics 15: 170-179.

Musebeck, J., Mohnike, K., Beye, P., Tonnies, H., Neitzel, H., Schnabel, D., Gruters, A., Wieacker, P.F. and Stumm, M. (2001): Short stature homeobox-containing gene deletion screening by fluorescence in situ hybridization in patients with short stature.

European Journal of Pediatrics 160: 561-565.

Muzny, D.M., Scherer, S.E., Kaul, R., Wang, J., Yu, Jun. and et al., (2006): The DNA sequence, annotation and analysis of human chromosome 3.

Nature 440: 1194-1198.

Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H. and Caskey, C.T. (1989): Alu polymerase chain reaction: a method for rapid isolation of human specific sequences from complex DNA sources.

Proc. Natl. Acad. Sci. USA 86: 6686-6690.

Netchine, I., Sobrier, M.L., Krude, H., Scanabel, D., Maghnie, M., Marcos, E., Duriez, B., Cacheux, V., Moers, A., Goossens, M., Gruters, A. and Amsalem, S. (2000): Mutations in LHX3 result in a new syndrome revealed by combined pituitary hormone deficiency.

Nature Genetics 25: 182-186.

Nguyen, T., McDonnell, C.M. and Zacharin, M.R. (2005): Primary ovarian failure and deletions of the long arm of chromosome 3.

Journal of Pediatric Endocrinology and Metabolism 18: 1013-1017.

Nicholls, R.D., Fischel-Ghodsian, N. and Higgs, D.R. (1987): Recombination at the human alpha-globin gene cluster: sequence features and topological constraints.

Cell 49:369–378.

Ogata, T., Goodfellow, P., Petiti, C., Aya, M. and Matsuo, N. (1992): Short stature in a girl with a terminal Xp deletion distal to DXYS15: localization of a growth gene(s) in the pseudoautosomal region.

Journal of Medical Genetics 29: 455-459.

Ottolenghi, S., Lanyon, W.G., Paul, J., Williamson, R., Weatherall, D.J., Clegg, J.B., Pritchard, J. and et al (1974): The severe form of alpha thalassaemia is caused by a haemoglobin gene deletion.

Nature 251:389–391.

Prak, E.T.L. and Haig, H.K. Jr (2000): Mobile elements and the human genome.

Nature Reviews Genetics 1: 134-144.

Peisker, K. (2003): Molekulare Charakterisierung einer familiaeren, mit Kleinwuchs einhergehenden Inversion des Chromosoms 3.

Diplomarbeit, FB Medizin, Martin-Luther Universitaet, Halle-Wittenberg.

Pfaeffle, R. (2006): Genetics of growth in the normal child.

European Journal of Endocrinology 155: S27-S33.

Pfaeffle, R.W., DiMattia, G.E., Parks, J.S., Browm, M.R., Wit, J.M., Jansen, M., Van der Nat, H., Van den Brande, J.L., Rosenfeld, M.G. and Ingraham, H.A. (1992): Mutation of the POU-specific domain of Pit-1 and hypopituitarism without pituitary hypoplasia.

Science 257: 1118-1121.

Pinkel, D., Straume, T. and Gray, J.W. (1986): Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization.

Proc. Natl. Acad. Sci. USA 83: 2934-2938.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J, Segraves, R., Lucas, J. and Gray, J.W. (1998): Fluorescence *in situ* hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4.

Proc. Natl. Acad. Sci. USA 85: 9138-9142.

Pintado, T., Ferro, M.T., San-Roman, C., Mayayo, M. and Iarana, J.G. (1985): Clinical correlations of the 3q21;3q26 cytogenetic anomaly. A leukemic or myelodysplastic syndrome with preserved or increased platelet production and lack of response to cytotoxic drug theraphy.

Cancer 55: 534-541.

Rao, E., Weiss, B., Fukami, M., Rump, A., Niesler, B., Mertz, A., Muroya, K., Binder, G., Kirsch, S., Winkelmann, M., Nordsiek, G., Heinrich, U., Breuning, M.H., Ranke, M.B., Rosenthal, A., Ogata, T. and Rappold, G.A. (1997): Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome.

Nature Genetics 16: 54-63.

Rao, V.V., Loffler, C., Schnittger, S. and Hansmann, I. (1991): The gene for human growth hormone-releasing factor (GHRF) maps to or near chromosome 20p12.

Cytogenetics Cell Genetics 57: 39-40.

Reiter, L.T., Murakami, T., Koeuth, T., Gibbs, R.A. and Lupski, J.R. (1997): The human COX10 gene is disrupted during homologous recombination between the 24 kb proximal and distal CMT1A-REPs.

Human Molecular Genetics 6:1595-1603.

Richardson, C. and Jasin, M. (2000): Frequent chromosomal translocations induced by DNA double-strand breaks.

Nature 405: 697-700.

Robson, K.J.H., Chandra, T., MacGillivray, R.T.A. and Woo, S.L.C. (1982): Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA.

Proc. Natl. Acad. Sci. USA 79: 4701-4705.

Roth, D.B. and Wilson, J.H. (1986): Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction.
Molecular Cell Biology 6: 4295–4304.

Royer-Pokora, B., Kunkel, L.M., monaco, A.P., Goff, S.C., Newburger, P.E., Baehner, R.L., Cole, F.S., Curnutte, J.T. and Orkin, S.H. (1986): Cloning the gene for an inherited human disorder-chronic granulomatous disease-on the basis of its chromosomal location.

Nature 322: 32-38.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988): Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.

Science 239: 487-491.

Salvatori, R., Hayashida, C.Y., Aguiar-Oliveira, M.H., Philips III, J.A., Douza, A.H.O., Gondo, R.G., Maheshwari, H.G., Baumann, G. and Levine, M.A. (1999): Familial Dwarfism due to a Novel Mutation of the Growth Hormone-Releasing Hormone Receptor Gene.

The Journal of Clinical Endocrinology and Metabolism 84: 917-923.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989): Molecular cloning. A laboratory manual. Cold Spring Harbor, New York, Cold Spring Laboratory Press.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977): DNA sequencing with chain-terminating inhibitors.

Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B.B., Butler, A., Castle, A.B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P.J., Dibling, T.,Drouot, N., Dunham, I., Duprat, S., East, C., Hudson, T.J. and et al. (1996): A gene map of the human genome.

Science 274: 540-546.

Schwartz, D.C. and Cantor, C.R. (1984): Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis.

Cell 37(1): 67-75.

Sensenbrenner, J.A., Hussels, I.E. and Levis, L.S. (1975): CC- A low birthweight syndrome, Rieger syndrome.

Birth Defects Original Article Series 11: 423-426.

Shaffer, L.G. and Lupski, J.R. (2000): Molecular mechanisms for constitutional chromosomal rearrangements in humans.

Annual Review of Genetics 34: 297-329.

Shaw, C.J. and Lupski, J.R. (2004): Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease.

Human Molecular Genetics 13: R57-R64.

She, X., Horvath, J.E., Jiang, Z., Liu, G., Furey, T.S., Christ, L., Clark, R., Graves, T., Gulden, C.L., Alkan, C., Bailey, J.A., Sahinalp, C., Rocchi, M., Haussler, D., Wilson, R.K., Miller, W., Schwartz, S. And Eichler, E.E. (2004): The structure and evolution of centromeric transition regions within the human genome.

Nature 430: 857-864.

Shears, D.J., vassal, H.J., Goodman, F.R., Palmer, R.W., Reardon, W., Superti-Furga, A., Scambler, P.J. and Winter, R.M. (1998): Mutation and deletion of the pseudoautosomal gene *SHOX* cause Leri-Weill dyschondrosteosis.

Nature Genetics 19: 70-73.

Shiang, R., Thompson, L.M., Zhu, Y.Z., Church, D.M., Fielder, T.J., Bocian, M., Winokur, S.T. and Wasmuth, J.J. (1994): Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia.

Cell 78: 335-342.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992): 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. USA 89: 8794-8797.

Smit, AF. (1999): Interspersed repeats and other mementos of transposable elements in mammalian genomes.

Current Opinion in Genetics and Development 9: 657-663.

Snell, K. (1986): The duality of pathways for serine biosynthesis is a fallacy.

Trends in Biochemical Sciences 11: 241-243.

Snell, K., Natsumeda, Y., Eble, J.N., Glover, J.L. and Weber, G. (1988): Enzymatic imbalance in serine metabolism in human colon carcinoma and rat sarcoma.

British Journal of Cancer 57: 87-90.

Snell, K. and Weber, G. (1986): Enzymatic imbalances in serine metabolism in rat hepatomas.

The Biochemical Journal 233: 617-620.

Soderlund, C., Longden, I. and Mott, R. (1997): FPC: a system for building contigs from restriction fingerprinted clones.

Computer Applications in the Biosciences 13: 523-535.

Southern, E.M. (1975): Detection of specific sequences among DNA fragments separated by gel electrophoresis.

Journal of Molecular Biology 98: 503-517.

Srivastava, A.K., Montonen, O., Saarialho-Kere, U., Chen, E., Baybayan, P., Pispa, J., Limon, J., Schlessinger, D. and Kere, J. (1996): Fine mapping of the EDA gene: a translocation breakpoint is associated with a CpG island that is transcribed.

American Journal of Human Genetics 58: 126-132.

Stankiewicz, P. and Lupski, J. (2002): Genome architecture, rearrangements and genomic disorders.

Trends in Genetics 18: 74-82.

Stine, S.B., Clark, C.E., Telfer, M.A., Casey, P.A. and Cowell, H.R. (1982): Ullrich-Turner syndrome (45,X/46,Xi[Xq]) in a child with a familial inversion of chromosome 3.

American Journal of Medical Genetics 12: 57-62.

Strachan, T. and Read, P.A. (1996): Human Molecular Genetics.

Bios scientific publishers

Sulston, J., Mallett, F., Durbin, R. and Horsnell, T. (1989): Image analysis of restriction enzyme fingerprint autoradiograms.

Computer Applications in the Biosciences 5: 101-106.

Tadin, S.M., Warburton, D., Baumeister, F.A.M., Fischer, S.G., Yonan, J., Gilliam, T.C. and Christiano, A.M. (2004): Cloning of the breakpoints of a de novo inversion of chromosome 8, inv(8)(p11.2q23.1) in a patient with Ambras syndrome.

Cytogentic and Genome Research 107: 68-76.

Taylor, J.M., Dozy, A., Kan, Y.W., Varmus, H.E., Lie-Injo, L.E., Ganesan, J. and Todd, D. (1974): Genetic lesion in homozygous alpha thalassaemia (hydrops fetalis).

Nature 251: 392–393.

The Fifth Wellcome Summer School (1990): DNA related methods in human genetics: YAC cloning in genome analysis. The Trustees of the Wellcome Trust.

Todd, R., Bia, B., Johnson, E., Jones, C. and Cotter, F. (2001): Molecular characterization of a myelodysplasia-associated chromosome 7 inversion.

Bristish Journal of Haematology 113: 143-152.

Tommerup, N. (1993): Mendelian cytogenetics. Chromosome rearrangements associated with Mendelian disorers.

Journal of Medical Genetics 30: 713-727.

Toriello, H.V., Glover, T.W., Takahara, K., Byers, P.H., Miller, D.E., Higgins, J.V. and Greenspan, D.S. (1996): A translocation interrupts the COL5A1 gene in a patient with Ehlers-Danlos syndrome and hypomelanosis of Ito.

Nature Genetics 13: 361–365.

Trask, B., van den Engh, G., Pinkel, D., Mullikin, J., Waldman, F., van Dekken, H. and Gray, J. (1988): Fluorescence *in situ* hybridization to interphase cell nuclei in suspension allows flow cytometric analysis of chromosome content and microscopic analysis of nuclear organization.

Human Genetics 78: 251-259.

Tsukamoto, K., Tohma, T., Ohta, T., Yamakawa, K., Fukushima, Y., Nakamura, Y. and Niikawa, N. (1992): Cloning and characterization of the inversion breakpoint at chromosome 2q35 in a patient with Waardenburg syndrome type I.

Human Molecular Genetics 1: 315-317.

Turner, H.H. (1938): A syndrome of infantilism, congenital webbed neck, and cubitus valgus.

Endocrinology 23: 566-567.

Ullrich, O. (1930): Über typische Kombinationsbilder multipler Abartungen.

Kinderheilkunde 49: 271-276.

Vanin, E.F., Henthorn, P.S., Kioussis, D., Grosveld, F. and Smithies, O. (1983): Unexpected relationships between four large deletions in the human beta-globin gene cluster.

Cell 35: 701–709.

Varon, R., Gooding, R., Steglich, C., Marns, L., Tang, H., Angelicheva, D., Yong, K.K., Ambrugger, P., Reinhold, A., Morar, B., Baas, F., Kwa, M., Tournev, I., Guerguelcheva, V., Kremensky, I., Lochmuller, H., Mullner-Eidenbock, A., Merlini, L., Neumann, L., Burger, J., Walter, M., Swoboda, K., Thomas, P.K., von Moers, A., Risch, N. and Kalaydjieva, L. (2003): Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome.

Nature Genetics 35: 185-189.

Velagaleti, G.V.N., Bien-Willer, G.A., Northuo, J.K., Lockhart, L.H., Hawkins, J.C., Jalal, S.M., Withers, M. and Lupski, J.R. (2005): Position Effects Due to Chromosome Breakpoints that Map ~900 kb Upstream and ~1.3 Mb Downstream of *SOX*9 in Two Patients with Campomelic Dysplasia.

American Journal of Human Genetics 76: 652-662.

Vishwanathan, V.K., Krcmarik, K. and Cianciotto, N.P. (1999): Template secondary structure promotes jumping during PCR amplification.

Biotechniques 27: 508-511.

Vortkamp, A., Gessler, M. and Grzeschik, K.H. (1991): GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families.

Nature 352: 539-540.

Wagner, J.K., Eble, A., Hindmarsh, P.C. and Mullis, P.E. (1998): Prevalence of human GH-1 alterations in patients with isolated growth hormone deficiency.

Pediatric Research 43: 105-110.

Weinstock, G.M. and Lupski, J.R. (1998): in Bacterial Genomes: Physical structure and Analysis (de Bruijn, F.J., Lupski, J. R. and Weinstock, G .M. eds), pp 112- 118, Chapman & Hall.

Wu, Q. and Krainer, A.R. (1999): AT-AC Pre-mRNA Splicing Mechanisms and Conservation of Minor Introns in Voltage-Gated Ion Channel Genes.

Molecular and Cellular Biology 19: 3225-3236.

Wu, W., Cogan, J.D., Pfaffle, R.W., Dasen, J.S., Frisch, H., O'Connell, S.M., Flynnm S.E., Brown, M.R., Mullis, P.E., Parks, J.S., Phillips, J.A.III. and Rosenfeld, M.G.(1998): Mutations in PROP1 cause familial combined pituitary hormone deficiency.

Nature Genetics 18: 147-149.

Yip, M.Y., MacKenzie, H., Kovacic, A. and McIntosh, A. (1996): Chromosome 3p23 break with ring formation and translocation of displaced 3p23-->pter segment to 6pter.

Journal of Medical Genetics 33(9): 789-792.

7   Appendix

Supplementary Figure 1: Nucleotide sequence of the 2,933 bp breakpoint anchoring region.

The highlighted sequences are the repeat regions. The two pink color highlighted sequences are the LTR33A elements, green color highlighted sequences are the MER67C repeat region, yellow color highlighted sequences are the L2 repeat region and the red color highlighted sequences are the MER67D region.

```
CTGATGATTAAAGGGATGAAGACATATGGAAAAATAATGAATATAAAGGG       50
AATGTTCCTGAGTAGCAGGCCAGAATGCTCTAGGAAATGTTAGAAAAATG       100
AGAGTCAGAGAGGTAATCATTGGTTATGCCCATCCAAGCCCTTTACTAAG       150
AGCCAAATGCTAACCACTTTAAATAAACTTCTTTGAATATCATTTTTTCC       200
TTAGTCCTCCACAAATAGTAAGAACTCTTTCACATTAATAATTACTTGCA       250
GCTTCCAGTAATTAAGTTCTAATCAATTTGTTTATTAAAGAGCATTTAAT       300
ATATTCATAATGTTATGGGGGAAAGGTAAGTATAGGATTGCAATGAAGAA       350
TTGCTTAGAAAAAAGTAACTAATAATAGTATATCCATAAACACAAAATT       400
CAAACTATTTACTCACTTAAGCCTATCTCAGTTGCCTTAGGTATTTGTGA       450
CATATCAAACAACAAATGCTGTTATATCATAAGTAATAAGATTTGTTTCT       500
TTGTAGTTTCCTTTAAGCAATAGAATGCTAGCAGTTCAATGCTTCTTAAA       550
GTCTTGCTGTAACTCAACAAATTTTGATTTTCCCTTCTTTATCTGACAAA       600
GAGAAGGAAACCTCTGAGGGATTACAGATGACCTCATCTTTTTGCTTTTC       650
AAGTCTCTGCTGCAAACCTAGCTGTGAAGGAAATATGAATTCAGCTAACA       700
TCATGTCACCTCCTTTATTCTTAGAGGCTGAAACAGTAGAGAAATTTCCT       750
CTGAGGGAGCTCCCAGACAAAATAAGACCTAGCCTGAAGAATAAAGGTGA       800
TATTGGATTATATTTTGGCTATCCAGAGAAGGTCTGTGCCAGTCAGAATC       850
CAATCAGATTTTAAACAGAGAATCTAATTCATTGAATTGGTGGCACAGGT       900
CATAGAACAGCCGGTGAAAACAAGAGACTGTAAGGCCATTAGGGATTAGG       950
ATAATCAGGAAGCTGCTTCCACTGTTAGGAAAGAAGGGAGCAGTGGTAAC       1000
ATCAGCCCCTAGTGTTGTGACATCCAAGGGAAACTGGAGCCGTGGTGGGC       1050
TTGCCTCATGTAGGGGGTGAAGACTGCAGCAGAGAGTAAGGCACAGGATT       1100
GCCCTGGCTTCTCCCTTCTTCCCATTCTCTAGTTTCTGGGAAACATCACC       1150
TAGGGGCTCAGCAGCAGAAGGGCAAGGAACAGATCTCAGGGCAAACAGGA       1200
CAGGATTGGCAAAAAGGCTTAGAAGCAGAATTGTTTCATTATAATGTACT       1250
TGAAGCTGCACCTACAAGAATTGAGCTGCAAGAGATGCAGATTTAACTGT       1300
```

```
TGGTCCAATGAGGTACAACCATCTCACAAGCCCAAGCTAACCTCTTTCCA        1350

GGAAGAGGAGCAGCTATAGGCCTGCAGTGGCAACATATGTGAGAATTGTT        1400

TGTGAATTCATGTTTGACACTTGCAGCATCCCCCTCCCAACAAAGATCCC        1450

AAAAATATTTGAGACTATGAAGAAGTAGAGAAGAGGAGGTGCTAATGGTC        1500

CCCCAATATTCAAGGAGACTAAATCTATGAAACTTCAATTATTACCATAA        1550

ATAGGACCTAGCATTACCTCTGGACTAATGCTACGTGGAAGAGTTCTAAT        1600

TACTCATACATGTTAGATAACTAAAAACATATACATCAACAGTGATTATG        1650

TCTATATTTTCTGAAATTTCTACAATGAGTTTATATTACTTTTATTACTT        1700

TCTAAGGTTTTTAAAAATAAGGTACCATAATAGAT TTTGCTTAAAATACT        1750

TGGTGCTTTCTGTTTTCCTGGTTGGACTTTGACT CACATGGACATTTCTC        1800

GGTAGCTAGCATCTGATTATCCATCTCAAGGCAGTTCAACAAGGACCCAT        1850

CTCCCAGAGCACTGCTTGCTGTGAGCATATTTAATTGGACTCCCTTGTCA        1900

AAACTAGGCATTCAGCTTTGACTCATGAGCTTCTCTGAAAGCATTTAACC        1950

AAAGGAAAACCTC TGACCAAGGTAGCCTTAATGTTCCCCACAGCTTGACT        2000

ACACTTTAGACAGATTTCTTCCTGACTATAGGCCCCTGACCTCCTTTTCC        2050

TTAGAGCATTTACTTTACAAAACTTGTAATTGTAAATTCTTCCTCTGTCT        2100

CCTTTGAAATAAATCTCCTTAGAAACCTCACGAGTTTTAAGACCCAATAA        2150

ATGTCTTTCTCAAGGACCTGGGAGCTGTCTCTTTGAAATGTAAACATCAG        2200

AGGAGATAGCTCTCCAATCTTCCAGTTTCTGTGGAAGGGTAGGAGCCTAA        2250

CTTCAGTGGGCTCCTTGTTCCAAGCTGTAAAAGTACTTCTTGTCATGAAG        2300

ACACAGAAAGTTTACTTTTCCTCTGGGTAGAAGCCAGTTAG AAATCACGT        2350

GCGTCCAATGATTTC TCCCACCCCAGCTCCTGAAAACCCTCCAGTGAGAA        2400

ATGATTGTGAACTCATGTTTGATACCTGCAATATCTCCC TTCCAGGT TCT        2450

CTCACCCCTATTGTGATAGTCTTGAACAAAGTCTTCCTGTTTAACTTTTT        2500

CCAGTGCAATTCTTACTTTGACA CCTCCCAGTTCCCTCCTGAGATTATCC        2550

TATATTTATGAATCAGCAGTGATTTTTTAGCAAGGGGAAATAGGATTCTC        2600

AGTTGGCTCCCTCAACTCTCCGGAGTGCCGAGGCAGATCAATGAGAAGCA        2650

AGCTCAAAAGGGTCTTGCAAGTTGCACTTTGAAAACAAGTTGCCTTTACT        2700

ACTAAAAGGCCTGCAATTAAAAAAAAAATGTGTTGGGTTCCATTCTCTTC        2750

TTTCCCCTCTCTCCTCTTGATTCGCTTCTCTTTCTACCTTCTTCCTTCCT        2800

ACACAACACCCCCTGACTCCTACACACAGCCATGCATACACCTCTCTACC        2850

TATGTCATTTACAATAAGTAGGTACCCTTCTTGTCGATTTAGAGAAGACG        2900

CTCCAGGCCTGGGATAAGCTCCAACAACCAGGT                        2933
```

Supplementary Figure 2: Nucleotide sequence of the 5,354 bp breakpoint anchoring region.

The repeat sequences are highlighted with different colors. The pink color highlighted sequences represents the LTR16C elements, green are MER20, red are (TATG)n, the two grey colors highlighted sequences are the AT rich regions, yellow color highlighted sequences are the MLT2B1 repeat region, light blue color highlighted sequences are the LTR1B elements. The dark blue color highlighted sequences are the CA repeat region.

```
CTCGGTTCGTCTAAAGCTGCAGCAGACATTTTCTGTACATGTATCACTCT          50
ATGTCTGTGCCTGATGGTTTTTCCTTGCCTAATATCACAGGAGAATGCTC          100
AGCTTATGGGGAAGGTATTCTGGAAAATAAAAGTATTTGATTGGATAGAT          150
AGTTCACATTTTCTGTTTGTTAGGGTGTAAATATTAAAGCATGGTTTATA          200
TAGTTCCCTAGAGAATTTCCAATAAAATGAATCCTAATTTCTCATATCAG          250
TAGCGTGCTTGTTAATGCACCCCTTATTATATTTTCCTCTTTCCTTGCCT          300
TAATTGCCCACTGTTTTAGTTATATTACCTATTTTTACTTAACTCTGTGT          350
CTTAGGGTCTACTTTGGTCAGAATCCAAACTAAAACATACTTTGTGAAAA          400
GTGCATACGCTAGGGCCTGGCCCTGGCATTCTTCACCACGGTCAGTTTCC          450
CAACTTACTTAGCTCTCCTTTGACATTATCAGTGTTTATCTGCTCTGTGT          500
TCTATTTCTGGACCTCTAGAAGTGACTGTCTACTAATCCTCCTTCCCAAA          550
AAAATTACGCTAAATTGCTGTTTTCTTAATATTTTTACATCTTTTCTCTG          600
GGATATTTTTGCTTAAATAGATAATCATGAAATGAAAAGCCATAGGCCAA          650
AAAAAAAAAAAAAAATGAATCCCTATGATATTTTAGATACTTTGTGGCTAG          700
TCAGTTGGGGAAATAGTCTTAAATATGAATTCTTTCAAAATGTCTCTAAA          750
TAATCACAGGTAGGTAATGAACTATCCTAATTAACTTTATTTACATCAGT          800
ATGTCTTAAAATAAATGAGTCATTAATAGCATAATGACATGGTGCCAACA          850
ATTCCAAATAAGTTTTATGTGATTCTCACATTGTATTTTGTATTTTCTTA          900
AAGCATTTTGCACTAAAATGAAAGAGAATGAATTTTTGAATTTGCTTTAC          950
TCTTAGCAGATATGTATTTTACATGGTGCAACTTCACTTTCATAATTGGA          1000
ATAAATCAAACTTCATGGTAATGAAAATATATCTTGAACTTGACATAATC          1050
CTCCAAGCATCATTTCAATTTATTATGGATGCTTAAGACAAATGTAATTA          1100
TCATTATTTTTATGTAAAAAATTATCAAAAAGGTTTGAGATTTCTCTTTC          1150
CTTTTGCATGGATTGGCAATTATAAGTAGTTAATTATAAGAAATATTTGG          1200
ATAGATTTCCTTCAAAAAACTGTATGCAAATAAATAAAATGACATTTAAA          1250
AATATTATTACAGTCTTATAGTGCATAGTTCTTTAGATCCATTAGTCTGT          1300
AAGCATCAGCAGTTAGCACTCCTGATACTTTGGACTAGATAATTATTTGT          1350
TGTGGGAGGTTGTGCTGTGCATTGTAGGATGTTTGGCAGCATCCTTAGTC          1400
```

TCTACCTACTGGATGCCAGTAGCACCTCCTCACCAGTGATGACAGTCAAA 1450

AATATCTCCAAACATTTGCAAAATATTCCTTGAGGGATAAAATTATTTTG 1500

GATTGAGAGCCACTG ATGTATGTGTATATATATGTATGCATGTATGTATG 1550

TATG AACATATAAGTGTATATATATTCATCTGTGAGTGGTTAGCATAGGT 1600

CCCTGAACATGTTTACTACATGCTTAAGGAGGAAGGGAAGAAAAGATTTA 1650

CTAAAATTTACTATGATGCATGTTTCTGTTTTTACATTTAAATATTCAAA 1700

TTCAATAATTGACAAGCATGTAATCCAGGCAATCTCTCAACAAAATAAAA 1750

ATAGGCAGCAATGACTATATACACAGCCAGAGGTACTTGATCATGAACAT 1800

TTTTTACTTAAGATTTACAGCTATTTGTTGCCCTGACATCAATCTTTTCT 1850

TGTATCCCAGTCTAGTTGATCAAATGAAGGTCCTGCCCATCCTGGGCCTG 1900

GAATCTGCATCATCTATCAGAACCCTGAGTGCTTGAGGCTATCTTTTGTT 1950

TTCTAAATCTCACATCTTTTAGGGAAACAATTATTTCTCTATCTTTGGCT 2000

AGTTCCCTCACACTCTGCTGCCATGGCCAGCCATCAAAATATAAATTCCA 2050

ATGCAACTTTCTGTTTCTAGTTGACTCTTTTAGTTAAAAATAATTCATTA 2100

CATGATGTTATCAACTCATGGTAATTAAGTTACATGATGAAGATCCGTGA 2150

AGGCAATATTTTTTTCAAAACCATTTAAATGGGTTAACAATACTGTTAGT 2200

ATCAATCTAAACAGTAGGTATATGAAGCAAATAGCAGAAACTAAATTCCT 2250

TTAGTGACTCAAAAATTCCATATTTATTCCCCAATACAGTCATGGGAAGT 2300

CAAGTATCCTCTCTCTGGGAATGATGCATCTGGTAGAGAATGCAATAAAA 2350

AGGATTTGTTGAAGACTTCAGCTTGTCAGCTACTGCACTCCCAACTTGTC 2400

ATTGTTCAAGTAGAAAGTCAAAGAGAAGTCTCAAGATATCACTTCATCAA 2450

GATAGATGAACACCACAATAATGAGAAGTATGAGAAGATGTTTAAAGCTA 2500

CTTATGGAGTGTCTCCTCTGTTCCCATTACTAGAAGCAGGGCTCATTGGA 2550

AACAGGATTTTTGTTTGCTTTTACAATCACTTTTTTATAAATTTCTTTTA 2600

CCTACTTATTAGATGTGTAGATCTACTGCCTTTGTATTATCTTTTTTAGT 2650

TACATAAAGTTTAATCATAATCATCCTAGCAAAATAAGTTAGGGGAGGTG 2700

GAAGCTGCTGTCCAGTACAAGAGTTTTGGTTCTGCTCATTTAGCACAGTA 2750

CACGATATCATTAACTAGAGATACTCATCTTCTGTTAGATCCCAAGAGAA 2800

GGTAATCCAAAACAAATGTAGGAAATCTCTACCCATGCCATAGCTCCAAT 2850

CCCAACTTTTGTCTTTTTCCATTAAATCAACTGAAGTATACAGTAGTAAC 2900

AATTTAGAACAAAAGAATATTGCTTTCAATCTTAGAGGGCTCTCTTTAA 2950

CTTTGCAGACAAAGACACTACCTAAATGAAATCTGAAATAAAACGCATAG 3000

TGCCACATGCATTTTCCATAGAAAAAATAATGAATGAAAGTTAGAACTTG 3050

GTGCTAACAGCATATCTGTTTGGCTACAACATGGATTAAGTTTGTTTCTG 3100

TGGTACTATTTTGAAACCAACCACCTCATATCTCATTGTTTCATATTTTA 3150

TTATGGTAAATTAGCTTTATTAACTATTTTCTAAGTTTGAAATAACTGGC 3200

```
AGAATACACAAACACTTAGAATACAATCTTCTTGAAACTTGACAGCACAT          3250
AGCTCCTTTTTCTCTCTTTCTGTATCTTTCATTATATATTTATATTTTAT          3300
AATGATAAATGTGTATATGCAAGCAGCATGTAATACTAAATCAGATAAAA          3350
AATATTGGCATCAGATTAGATATTCTATAGTGTTGAAACTATATAAAAT           3400
ATATTTTATGGCAGGTAGTTACAATATTATTAATACACTTTGGTGATAAA          3450
AAGGAAAAAAAAAGCTTTGCACAATATCATCTAAGCAACTGGTACCACAT          3500
TGTTAAAAATGTGCTCAGAACACTTAATAAGAGTATATGCATTATATGAA          3550
ATGTGAAATACGGTGATAAGCCCATGCTATATGTGATGGATAATTTTATG          3600
TGTCAACCTGACTGGGCTAAAGTTTTTAGCCAAAATATTTGGTGAAATAG          3650
CATTTCTAGATATGTTTGTGTGGGTGATTTTGTATGAGATTAACATTTGA          3700
ATCAGTAGATTAAGCAAAGCAGATTGACCTCCCCAGTGTAGGTGGACCTC          3750
ATCCAATCAGTGGAAGACCTGAATAGAATAAAAGGCTGAGTAAGAAAGAA          3800
TTATTTTTCCCTGCCTGACCATCTTCAATCTTGGATATTGATCTTCTTTT          3850
GTCTTCAGTCTCAAACTTGGACTAGAAACTTACACCATTACCGCCCATGG          3900
TTCACAGGCCTTTGGACTTGGACTGGAAATATATCATATACTCTCCTGGG          3950
TCTGCAGCTTGTCCATTGCAGATCATGGAACTTCTCTGTCTCCATAATCA          4000
CATAAGCCAGTTCCTTCTAGGAAATCATGTATGCATAATTTACACACACA          4050
CACACACACACACACACACACACACACACACACCTCATTAGATATAATAT          4100
AACATATATTTTCATATGAATATGTGTTTATTTAAATATAATACTGTATA          4150
TAATAGTATAATTGTGTCTGTTTCTGGAGAACCCAGAATAATACACTATT          4200
TAACTTAAATTTTATATTACTTTTTGTACAATTTCTAAAAAATAAATTTA          4250
AAAGTATTTCATATTTTTACTAGCAGTTTTACTGATTTGCAGTTTATCAT          4300
TCTATGTAAAATTCAATATTAACTCAATTATTTTTTAGTCATTATCTAGA          4350
GATCAATTTCAAAAAGCCTCTCTTCTCCTCTATTAAATAAAAGCAATGCT          4400
AATGATATTCTGCCAGCTTCAAATTGACTGATTTATCACATCATTACACA          4450
GCCATCCTGCATCTGCACTTTGTTTGTGGAATACTATAGCTTGACAATAA          4500
CAAATCACTATGCCCTCTGTTGTTGAGAATGTTAGAAGAAAGGCAACTCA          4550
CACATGCAAAACCAAATGGTGCTGGAATGTTTAGATGCATTTGGCTTCTG          4600
CATTAGCTATTCATCAGCCACACTTGTAATCATTTAGAATATTTTCATCT          4650
CGAAGATTTCAAATTCTGTCTCCTCACTCGTTCCTCTGTGTCTACTCTGC          4700
TTTGATTCTGTGACAGTGAAACTGGAGAGCTCCCTTATCTCCCTAGCAGG          4750
ATGTACTATAGGGGTGTGGCTCACTCCTTTGGTCACCCCACTGCCCTGAG          4800
GCTCAAACCTCTAGGGGGAGCATGCAGAGGGGCAGGTGCAGAGGCCCTGC          4850
AGAGTGCTTTTGGGCTCCAGCTCCACAGCAGCATCTACAAGTGGGTTTCT          4900
ATGACTCCTGAAGCCCAAGCGGGCATGTGTTACAGTGTGTTATTTCAGCT          4950
TTGCTGTCTACAGACGGCTTGTGTTAATTAGCTCAGTAGATCTTCTGCCT          5000
```

```
TATCACAAGGGAAGGGGGATCAGTGTGACAGCCTGAGTTCTTGCCCAGTG        5050

TACTGGAAGAATCGGATCACACTTGGATTTGAAGGATGAGTACAAGGTTT        5100

TGTTGAGTGGTGGAAGTCGCTCTCAGCTAGATGGAGATGGATGGGGAGCT        5150

GGAAGCAGGAGATGGAGTGGGAAGGTGGTCTTCCCCTGGAATCCAGATGC        5200

CCACTGGTTAGACTCTTCTCCGCCCCTGGCTGAATTCCCTTCTGTGTGTC        5250

CAGACATCTTTCTCTTCTCTCTTTCTCTGTCATGCCACCCTGCCCCTCTG        5300

TTCCTCTGGATGTTCAGCTGCTTGTGTCTATGTCCGCTAAGGTCTTGGGTTAAT    5354
```

Supplementary Figure 3: The inversion chromosome 3 ideogram showing the U5 and U6 primers designed for the junction PCR.
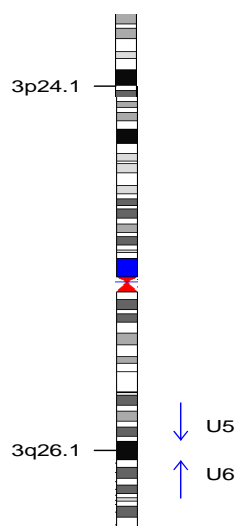
Acknowledgements

These years of my study have been a wonderful experience which has helped me to hone my intellectual and practical skills. I take this opportunity to thank all those people who have helped me through these years.

I am deeply indebted to my supervisor Prof. Dr. Ingo Hansmann for his constant help, encouragement and support both professionally and personally. I would like to thank Prof. Dr. Gunter Reuter for accepting to be my official supervisor.

I thank Dr. Dietmar Schlote for his constant guidance, his excellent supervision and support at every step in my work. I would like to thank my colleague Falko Matthes for all his help.

I thank Dr. Christiane Höfers for all her help bestowed on me. My special thanks to Ms. Petra Demuth, Ms. Sylvia Weiβflog, Ms. Uta Grasemann and Ms. Ruth Enders for all their love, care and constant support during my good and bad times making my stay at the Institute a memorable one.

I extend my thanks to Dr. Seyed E Hasnain, former director and Dr. Gowri Shankar director, CDFD for their support. I would like to specially thank Dr. Nalini Mallikarjuna, Scientist at ICRISAT for helping me in enhancing my scientific career.

Finally, I take this opportunity to thank my father, mother, brother for their persistent encouragement, love, support and strong faith in my abilities which had helped me to reach this stage. It is beyond words to express my gratitude to my husband who stood as a pillar of support in developing my career. Lastly, I would like to say thanks as well as sorry to my daughter Amora who was deprived of her motherly affection during the period of my study.

I am grateful to DAAD for providing me the fellowship and also Faculty of Biology, Martin-Luther University for providing me PhD registration.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Halle (Saale), den 15. August 2007   Usha Dutta

Curriculum vitae

17.09.1972	Born in Chirala, India
1977- 1987	Schooling at BHEL Higher Secondary School BHEL, Hyderabad, Andhra-Pradesh, India
1987-1989	Board of Intermediate Education St. Ann's College for Women, Osmania University Hyderabad, Andra-Pradesh, India
1989-1992	Bachelor of Science St. Ann's College for Women, Osmania University Hyderabad, Andra-Pradesh, India
1992- 1994	P.G Diploma in methods in Biotechnology Madurai Kamraj University, Madurai, Tamil Nadu, India
1995- 1998	Employee in Cellular & Molecular Biology Division of International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Asia Center, Patancheru, Andhra Pradesh, India
1997- 1999	Master of Science Annamalai University, Chidambaram, Tamil Nadu, India
1998- till date	Employee in Center for DNA Finger printing and Diagnostics (CDFD), Hyderabad, India
2004- 2007	PhD student Fellowship from DAAD Institute of Human Genetics and Medical Biology Martin-Luther-University Halle-Wittenberg, Germany