5 Application of array CGH and CGHPRO

5.1 Impact of segmental duplication on the generation of genome rearrangements

5.1.1 Molecular mechanisms underlying genome rearrangements

Genome rearrangements result from double-strand breaks (DSBs) that arise spontaneously during DNA replication or can be induced by ionizing radiation or chemicals (including anticancer drugs). DSBs are critical lesions which, if not repaired, may be lethal for the affected cell. So a number of cellular DNA repair mechanisms have evolved for the restoration of break sites. In eukaryotes, two major pathways have been identified that differ in their requirements of DNA homology. DSB repair by homologous recombination (HR) requires the presence of homologous sequences elsewhere in the genome (e.g. a homologous chromosome or a sister chromatid). In contrast, non-homologous end joining (NHEJ) fuses the two ends of a DSB through a process that is largely independent of terminal sequence homology and therefore can join ends with diverse chemical and physical characteristics. Both HR and NHEJ have been conserved during evolution, but vary in the contribution to overall DSBs repair in lower and higher eukaryotes. Generally speaking, while HR predominates in lower eukaryotes, DSB in mammals are primarily repaired by NHEJ. Furthermore, their relative contribution varies during development and depends also on the stage of the cell cycle: while NHEJ is active throughout the cell cycle, HR is limited to the late S and G2 phase. Although DSBs repair by either HR or NHEJ is normally efficient and precise, occasional errors can occur in the repair process and thus lead to genome rearrangements.

Regardless of the fact that chromosome rearrangements occur everywhere in the genome, they predominate in the intervals with a complex genomic architecture, such as segmental duplications and AT-rich palindromic repeats. This suggests that genome rearrangements are not random events, but rather result from chromosome instability that is due to the local genomic architecture (Shaw and Lupski, 2004).

5.1.1.1 Segmental duplication-mediated nonallelic homologous recombination

Segmental duplications are large, nearly identical copies of genomic DNA, which range in size from 1 to >200 kb and are present at two or more positions in the human genome. It has been estimated that 5% of the human genome are composed of such duplications, which are clustered in the pericentromeric transition zones, the subtelomers and several interspersed LCR hubs (Bailey et al., 2002; Bailey et al., 2001; Bailey et al., 2002; Cheung et al., 2003; Cheung et al., 2001; Eichler, 2001; Horvath et al., 2001). Many of the segmental duplication in the human genome appear to have arisen during primate speciation. It has been hypothesized that these duplications can drive adaptive evolution by generating new genes. This hypothesis is supported by a variety of studies which have shown DNA copy number changes between human and non-human primates (and Analysis ConsortiumThe Chimpanzee, 2005; Fortna et al., 2004; Fujiyama et al., 2002; Locke et al., 2003; Newman et al., 2005; Wilson et al., 2006; Yunis et al., 1980). Although segmental duplications may be important in an evolutionary sense, their existence poses a risk to the individual human genome, as their highly homologous sequences provide ample substrates for non-allelic homologous recombination (NAHR). As shown in Figure 14, segmental duplication-mediated NAHR can lead to deletions, duplications or inversions, depending on the orientation (direct/inverted) of the duplicated sequences and the involvement of interchromosomal, intrachromosomal or intrachromatid recombination (Stankiewicz and Lupski, 2002). In addition, NAHR between different chromosomes can also result in chromosomal reciprocal translocation. The probability of meiotic misalignment between duplicated sequences may depend on several factors—including length, sequence identity and orientation as well as the distance between duplications.

Recently, segmental duplication-mediated NAHR has been directly implicated in a growing list of recurrent genomic disorders. Similarly, there is increasing evidence that the duplication architecture of the genome may also mediate structural variation in the normal population. In the study of Tuzun et al (2005), more than half of the detected variant sites (163 of 297) map to regions with segmental duplications. The association was most pronounced for the

intrachromosmal segmental duplications where the degree of sequence identity exceeds 98% (Tuzun et al., 2005).

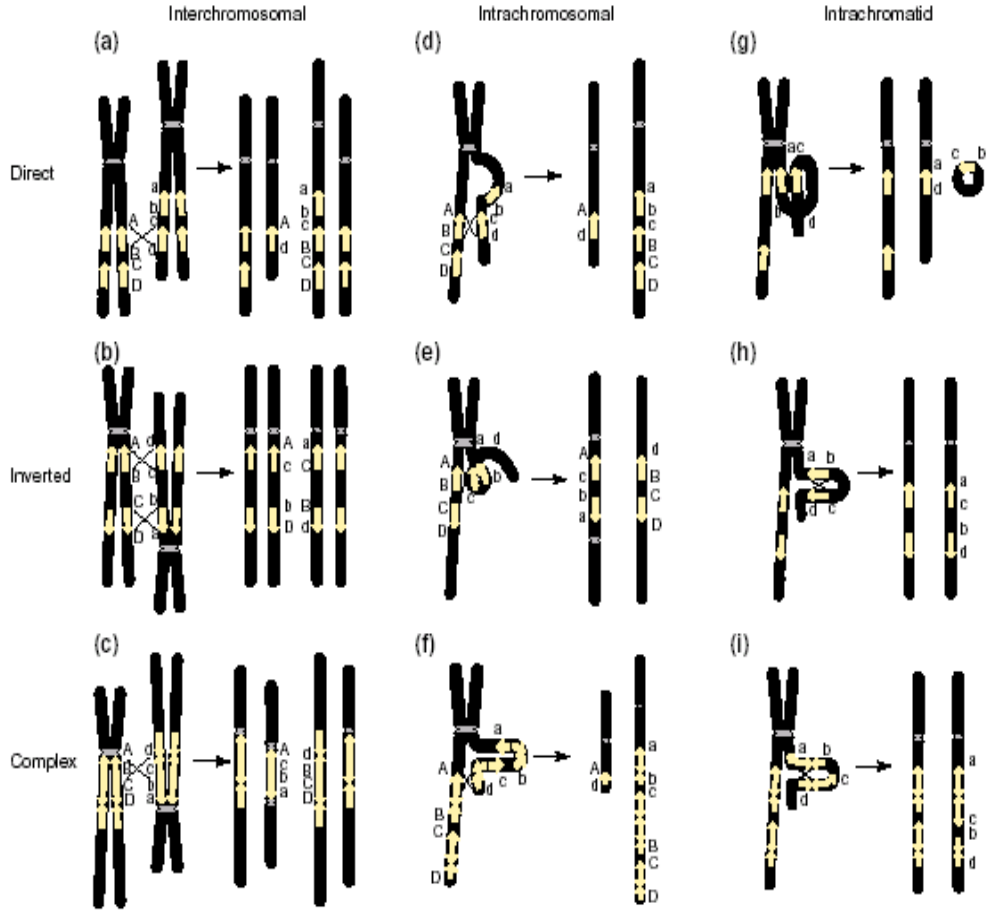


Figure 14: Mechanisms of genome rearrangements resulting from segmental duplications mediated NAHR (Stankiewicz and Lupski, 2002). Chromosomes are shown in black with the centromere depicted in gray line. Yellow arrows represented segmental duplications with specific orientation. All possible rearrangements mediated by segmental duplications are grouped horizontally by orientation and structure of segmental duplications (direct, inverted, complex), and vertically by the mechanisms (interchromosomal, intrachromosomal, intrachromatid).

5.1.1.2 Other genome architectural features

Segmental duplication-mediated NAHR cannot explain all cases of genome rearrangements. Other mechanisms such as NHEJ have been observed, particularly for rearrangements with scattered breakpoints (Roth and Wilson, 1986). Very often, complex genome architectural features are also involved. A systematic study of deletion junctions in the gene for Duchenne muscular dystrophy (DMD), revealed Alu and long tandem repeat (LTR) elements in 3 out

of 10 cases (Nobile et al., 2002). The sequence TTTAAA, which is known to bend the DNA molecule (Singh et al., 1997), was found at or near 3 of the junctions examined. In many translocation cases, AT-rich palindromes were found at the break points on the derivative chromosomes (Gotter et al., 2004; Kurahashi et al., 2003; Nimmakayalu et al., 2003). This suggests the possibility of secondary structures based on AT palindromes causing double strand breaks. In addition to that, centromeres, pericentromeric repeats and telomers are often implicated in non-recurrent breakpoints. Their involvement indicates that the chromatin structure can also play a role in genome rearrangements.

In this study, array CGH has been employed to study the impact of segmental duplications on the generation of both balanced and unbalanced genomic rearrangements. For this purpose, a set of 22 mentally retarded patients were examined, which has been pre-selected for the presence of chromosomal aberrations, and the results were compared with FISH mapping data from 41 mentally retarded patients with balanced translocations.

5.1.2 Copy number changes in 22 patients with mental retardation

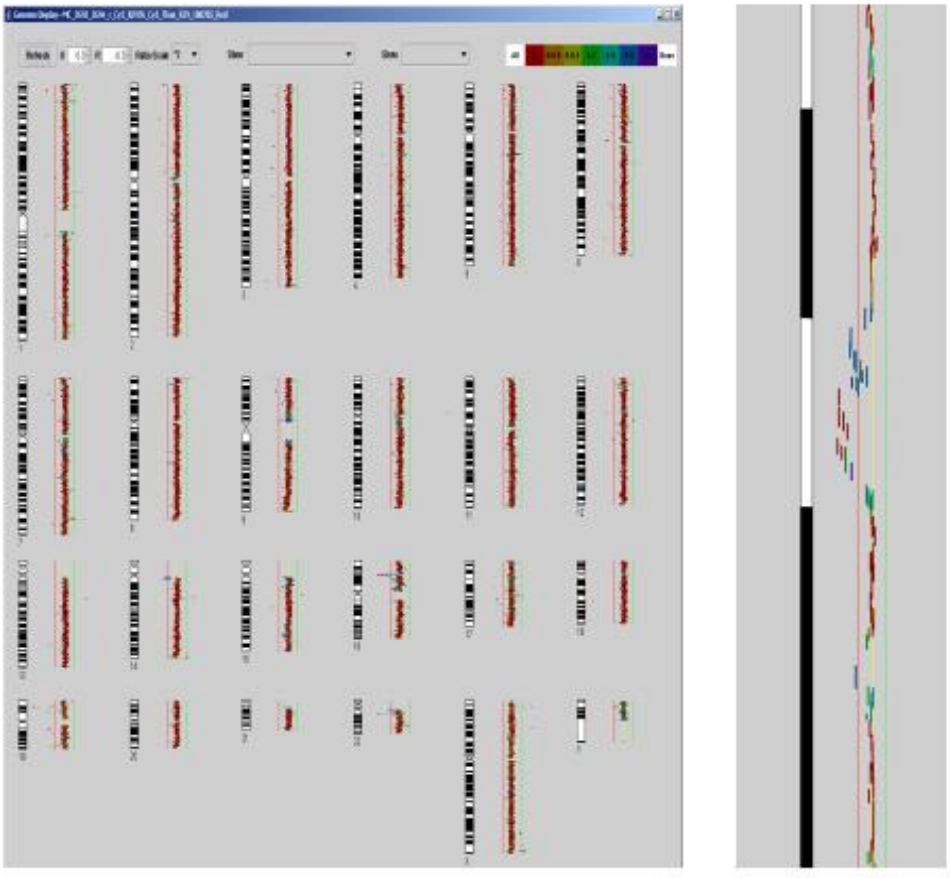
Array CGH was carried out for 22 patients with mental retardation. In all but four cases, the imbalances have been analysed and verified by HR-CGH (Kirchhoff, et al., 1999, Kirchhoff, et al., 2004). As controls, three patients with the known genomic disorders, Smith-Magenis Syndrome (SMS) (case16), Prader-Willi/ Angelman Syndrome (case15) and 22q11deletion syndrome (case7), were included.

For array CGH, a high resolution tiling path BAC array was used, comprising the human 32k Re-Array set (Krzywinski, et al., 2004; Osoegawa, et al., 2001; Ishkanian, et al., 2004), http://bacpac.chori.org/pHumanMinSet.html: (DNA kindly provided by Pieter de Jong), the 1Mb Sanger set (Fiegler, et al., 2003) (clones kindly provided by Nigel Carter, Wellcome Trust Sanger Center) and a set of 390 subtelomeric clones (assembled by members of the COST B19 initiative: Molecular Cytogenetics of solid tumors). Cases 5, 7, 8 and 16 were hybridised on a 14k array, which provided tiling path resolution only for chromosomes 4, 9, 10,

11, 16, 17, 21, 22, and X. All aberrations discussed here were detected with a sub-megabase tiling path BAC array.

Array CGH data were analyzed by CGHPRO. No background subtraction was applied. Raw data were normalised by “Subgrid LOWESS”. Copy number gains and losses were determined by a conservative log2 ratio threshold of 0.3 and -0.3, respectively. Aberrant ratios involving three or more neighbouring BAC clones were considered as genomic aberrations unless they coincided with a published polymorphism as shown in CGHPRO. In Figure 15, examples are shown of the genomic profiles from case 2 and case 4.

In total, 22 aberrations were identified in 22 patients. The size of aberrations ranged from 651 Kb to 14 Mb. Table 1 lists the aberrations found in each of the patients.



A B

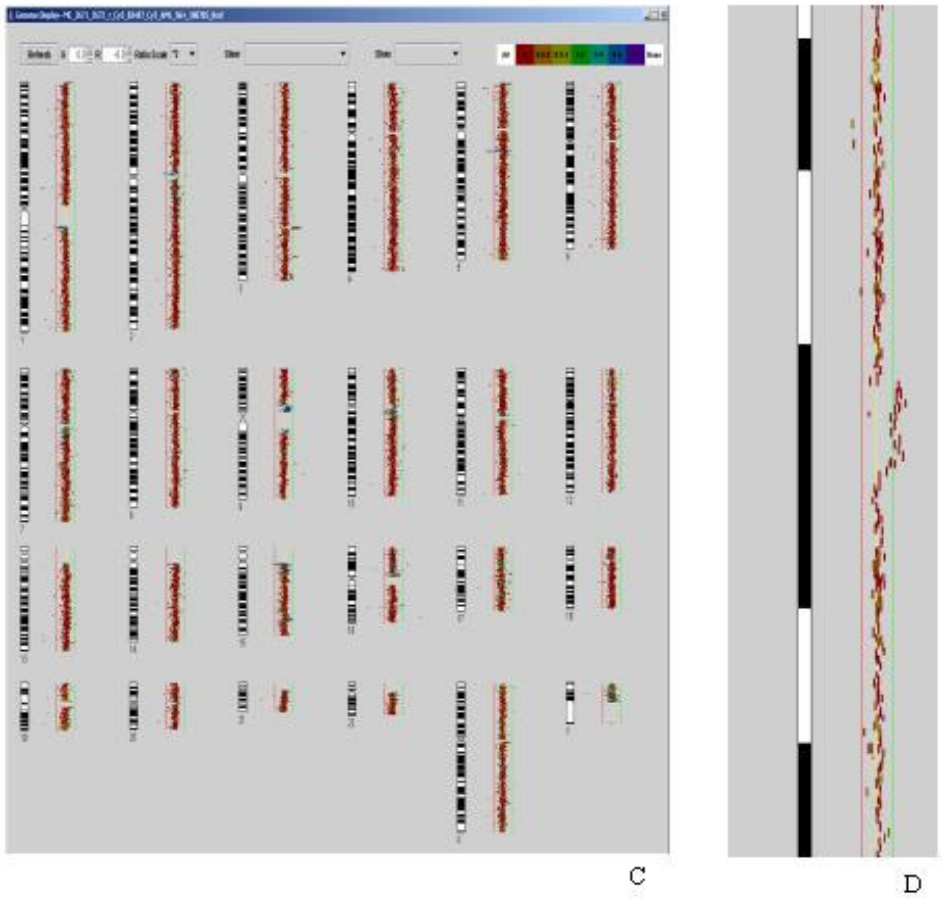


Figure 15: (A) Genome display of case 2. (B) Zoom-in view of the aberration in case 2 (C) Genome display of case 4. (B) Zoom-in view of the aberration in case 4.

Table 1: Array CGH results of 25 patients

case No.	chromosome band	gain/loss	start (kb)	end (kb)	size (kb)
1	16p11.2	loss	29573	30225	652
2	16p13.11	loss	14805	16425	1620
3	2q13	loss	111155	112776	1621
4	3q24	gain	145246	146988	1742
5	17q12	loss	31438	33481	2043
6	17q23.2-17q23.3	loss	55339	57686	2347
7	**22q11.21**	**loss**	**17202**	**20050**	**2848**
8	10q22.2-10q22.3	loss	75118	78473	3355
9	1p34.2-1p34.1	gain	40749	44190	3441
10	10q24.31-10q25.1	loss	102682	106175	3493
11	11q14.1-1q14.2	loss	82029	85900	3871
12	1q25.2	loss	176351	180687	4336
13	8q12.1-8q12.3	loss	60271	64625	4354
14	3q27.1-3q27.3	loss	184933	189324	4391
15	**15q11.2-15q13.1**	**loss**	**21265**	**26123**	**4858**
16	**17p12**	**loss**	**15545**	**20629**	**5084**
17	1p32.1-1p31.3	loss	59040	64557	5517
18	7p22.3-7p21.3	loss	1611	7158	5547
19	1p36.13-1p36.12	loss	17035	23035	6000
20	10q11.21-10q11.23	loss	45432	51611	6179
21	5q14.3-5q21.1	loss	90411	98322	7911
22	7p15.2-7p14.2	loss	26202	35318	9116
23	2q24.1-2q24.3	loss	154773	164127	9354
24	2p25.2-2p24.1	loss	6922	19766	12844
25	1q23.3-1q25.2	loss	160035	174241	14206

*Three patients with previously known genomic disorders are shown in bold. The 25 cases are sorted by aberration size.

5.1.3 Overlap of breakpoints in unbalanced aberrations with segmental duplication and CNPs

To estimate the content of segmental duplications and copy number polymorphisms (CNPs) around the breakpoints of the above 25 cases, a 400kb breakpoint interval, including 200kb proximal and 200kb distal of each breakpoint, was searched against the Segmental Duplication Database (http://humanparalogy.gs.washington.edu/) and the Database of Genomic Variants (http://projects.tcag.ca/variation/, version Dec 13, 2005) , respectively. Breakpoints were defined as the midpoint between end and start of the two neighbouring clones with alternate states, i.e. the one clone with normal and the other one with an aberrant ratio.

When segmental duplications were found to flank both breakpoints, the respective entries in the Segmental Duplication Database were checked for homology and degree of sequence similarity. The same procedure was also applied to the imbalances of an independent cohort of mentally retarded patients published recently (de Vries et al., 2005).

The segmental duplication content and DNA copy number polymorphisms (CNPs) found in the vicinity of breakpoint are shown in Table 2.

Table 2: Segmental duplication content and DNA copy number polymorphisms in 25 patients with unbalanced aberrations

Case No.	LCR content** upper breakpoint	CNP*** upper breakpoint	LCR content** lower breakpoint	CNP*** lower breakpoint	Size of homologous sequence(kb)	Sequence identity
1	3.088	-	2.898	-	146	0.996
2	5.162	-	4.37	-	310	0.957
3	1.218	+	1.615	+	44	0.995
4	0.0	-	0.0	+	0	0.0
5	2.299	+	3.041	+	200	0.987

6	1.454	+	1.769	-	48	0.987
7	**3.256**	**+**	**3.319**	**+**	**332**	**0.979**
8	1.835	-	0.0	+	0	0.0
9	0.0040	-	0.04	-	0	0.0
10	0.0	-	0.0	-	0	0.0
11	0.015	-	0.0	+	0	0.0
12	0.139	-	0.0080	-	0	0.0
13	0.0	-	0.0	-	0	0.0
14	0.0	-	0.0	-	0	0.0
15	**2.631**	**+**	**3.227**	**+**	**72**	**0.980**
16	**0.963**	**+**	**1.708**	**-**	**112**	**0.986**
17	0.0080	-	0.0	-	0	0.0
18	0.076	-	0.041	+	0	0.0
19	0.978	+	0.0	-	0	0.0
20	2.522	-	2.897	+	189	0.951
21	0.0	-	0.0	-	0	0.0
22	0.0080	-	0.0	+	0	0.0
23	0.0	-	0.0	+	0	0.0
24	0.0	-	0.0	-	0	0.0
25	0.0	-	0.0	-	0	0.0

*Three patients with previously known genomic disorders are shown in bold. The 25 cases are sorted by aberration size. Calculation is based on a 400 kb interval centered around the breakpoint.
**LCR (Low Copy Repeats, same as segmental duplications) content is calculated using the formula (∑Length of Duplication * Copy Number)/ Length of Clone
*** CNP: DNA Copy Number Polymorphism

The three images displayed in Figure 16a-c represent a case where both breakpoints are flanked by segmental duplications (Figure 16a), a case with a segmental duplication present at only one breakpoint (Figure 16b) and an imbalance without segmental duplication in the breakpoints regions (Figure 16c).

Figure 16d shows the classification and distribution of all 36,000 BAC clones with respect to their segmental duplication content, as described in the Section 4.10.1.

Chromosomal imbalances found in this study can be grouped into four classes:

1. Proximal and distal breakpoints are enriched for segmental duplications with high sequence similarity (6/22; 27,3%)
2. Proximal and distal breakpoints are enriched for segmental duplications, but with low sequence similarity (3/22; 13,6%)
3. Only one breakpoint lies within a segmental duplication (5/22; 22,7%)
4. No segmental duplication lies in the vicinity of both breakpoints (8/22; 36,4%).

When group 1 was compared with the rest groups, a significant difference in aberration size was observed (p=0.018878; Wilcoxon rank sum test).

It is noteworthy that in each of the three patients with known genomic disorders, which are due to non-allelic homologous recombination, homologous segmental duplications were found to flank the respective breakpoints.

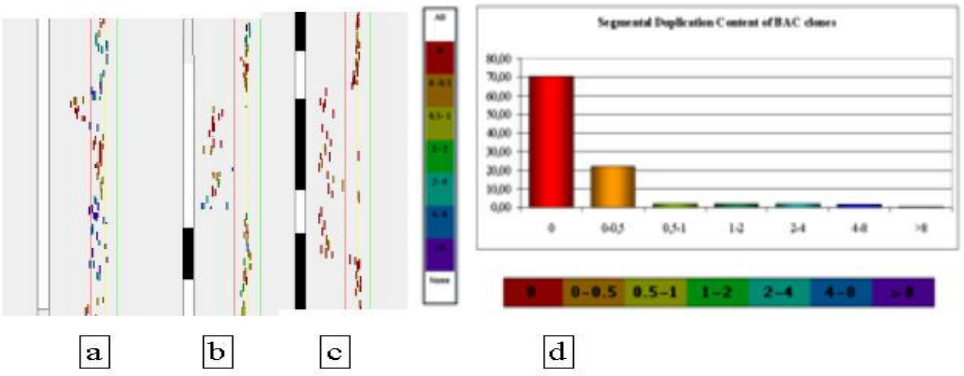


Figure 16: a-c Three example with two/one/no breakpoint covered by segmental duplication enriched clones. d Distribution of BAC clones with regard to their contents of segmental duplication (see text for detail).

When applying the same procedure to the independent dataset from de Vries et.al (de Vries et al., 2005), their aberrations can also be grouped into the same four classes, although the percentages are different. Table 3 lists the chromosome aberrations found in their study and Table 4 summarizes the segmental

duplication content as well as the DNA copy number polymorphisms (CNPs) found in the respective breakpoint regions.

Table 3: Genomic imbalances in patients with mental retardation, as determined by array CGH (from de Vries et al., 2005)

Patient	Type of Submicroscopic Aberration and Chromosome Band	Gain or Loss	Start (Mb)	End (Mb)	Length (Mb)
1	1p34.3-1p34.2	Loss	39.22	43.15	3.93
2	2q23.1-2q23.2	Loss	149.17	150.09	.92
3	3q27.1-3q29	Loss	184.43	196.8	12.37
4	5q35.1	Gain	170.52	171.76	1.24
5	9q31.1	Loss	99.74	102.58	2.85
6	9q33.1	Loss	115.3	115.84	.54
7	11q14.1-11q14.2	Loss	78.12	85.61	7.49
8	12q24.21-12q24.23	Gain	114.91	117.21	2.3
9	17p13.2-17p13.1	Gain	4.27	7.16	2.89
9	17p13.1	Gain	7.67	9.10	1.43
9	17p12	Gain	12.65	15.54	2.88
9	17p11.2	Gain	18.55	20.03	1.48
10	22q11.21	Loss	17.1	19.75	2.66
11	1q21.1	Gain	143.25	145.38	2.12
12	3p14.1	Loss	67.59	68.15	.56
13	7q11.21	Loss	64.23	64.58	0.35
14	9p24.3	Gain	.21	.45	.23
15	15q24.1-15q24.2	Loss	72.21	73.86	1.65

Table 4: Segmental duplication content and DNA copy number polymorphisms in patients with unbalanced aberrations* (de Vries et al., 2005)

patient	LCR content** upper breakpoint	CNPs*** upper breakpoint	LCR content** lower breakpoint	CNPs*** lower breakpoint	size of homologous sequence(kb)	sequence identity
1	0.0080	-	0.029	-	0	0
2	0	-	0.01	-	0	0
3	0.013	-	0.613	+	0	0
4	0.038	-	0	-	0	0
5	0.012	-	0	-	0	0
6	0	-	0	-	0	0
7	0	+	0.0040	-	0	0
8	0	-	0.0040	-	0	0
9	0	+	0	-	0	0
9	0	+	0	+	0	0
9	0	+	0.963	+	0	0
9	1.684	-	1.118	-	264	0.992
10	3.07	+	2.261	+	283	0.969
11	0.786	-	4.808	-	0	0
12	0.011	-	0	+	0	0
13	2.624	+	5.824	+	172	0.994
14	1.805	+	0	+	0	0
15	0.9	+	1.484	-	54	0.934

*Calculation is based on a 400 kb interval centered around the breakpoint.
**LCR (Low Copy Repeats, same as segmental duplications) content is calculated using the formula (∑Length of Duplication * Copy Number)/ Length of Clone
*** CNP: DNA Copy Number Polymorphism

5.1.4 Overlap of breakpoints in balanced translocation with segmental duplication and CNPs

Compared with unbalanced aberrations, less is known of the impact of segmental duplications on constitutive balanced translocations. Chromosomes 11, 17 and 22, containing regions with AT-rich palindromes embedded in segmental duplications seem to be more frequently involved and it appears that the affected duplications preferentially fuse with the telomeric regions of their translocation partner chromosome (Edelmann et al., 2001; Gotter et al., 2004; Kurahashi et al., 2004; Kurahashi et al., 2003; Kurahashi et al., 2000; Kurahashi et al., 2000; Shaikh et al., 2001; Spiteri et al., 2003; Stankiewicz et al., 2003)

To investigate the effect of segmental duplications on balanced translocations, we analysed the content of segmental duplications and CNPs around breakpoints in 41 balanced translocations, where the breakpoints had been mapped to single BACs by FISH. The procedure was the same as the analysis for chromosome imbalances except that the position of breakpoints in balanced translocations was defined as the midpoint of the respective breakpoint-flanking clone. Table 5 summarized the results.

Table 5: Segmental duplication content and DNA copy number polymorphisms in 41 mentally retarded patients with balanced translocation*

Case No.	LCR content** breakpoint1	CNP*** breakpoint 1	LCR content** breakpoint 2	CNP*** breakpoint 2	Size of homologous sequence(kb)	Sequence identity
1	0.009	-	0.0	-	0	0
2	0.000	+	0.0	-	0	0
3	0.000	+	0.0030	+	0	0
4	0.000	-	0.0	+	0	0
5	0.004	-	0.072	-	0	0
6	0.000	-	0.0	-	0	0
7	0.566	-	0.0	-	0	0

8	0.000	-	0.0	-	0	0
9	0.000	-	0.0	-	0	0
10	0.000	-	0.0	-	0	0
11	0.000	-	0.0	-	0	0
12	0.000	+	0.0	-	0	0
13	0.008	-	0.358	-	0	0
14	0.000	-	0.0	-	0	0
15	0.000	-	0.01	-	0	0
16	0.000	-	0.15	-	0	0
17	0.000	-	0.0	-	0	0
18	0.009	-	0.0	-	0	0
19	0.000	-	0.0	-	0	0
20	0.047	+	0.0	+	0	0
21	1.503	-	3.812	+	0	0
22	0.042	+	1.361	+	0	0
23	0.000	-	0.0	-	0	0
24	0.000	-	0.0070	-	0	0
25	0.000	-	0.0040	-	0	0
26	0.008	+	0.019	-	0	0
27	0.000	+	0.012	+	0	0
28	0.000	-	0.0	-	0	0
29	0.016	+	0.0070	-	0	0
30	0.0	-	0.087	-	0	0
31	6.382	-	0.0	-	0	0
32	0.000	-	0.0	-	0	0
33	1.987	+	0.000	+	0	0

34	0.000	-	0.032	-	0	0
35	0.007	+	0.0040	-	0	0
36	0.000	-	0.0030	+	0	0
37	0.000	-	0.0040	-	0	0
38	0.000	-	0.0030	-	0	0
39	0.0090	-	0.0	-	0	0
40	0.0080	-	0.358	-	0	0
41	0.133	+	0.0	-	0	0

*Calculation is based on a 400 kb interval centered around the breakpoint.
**LCR (Low Copy Repeats, same as segmental duplications) content is calculated using the formula (∑Length of Duplication * Copy Number)/ Length of Clone
*** CNP: DNA Copy Number Polymorphism

For the constitutive translocations we have encountered the following distribution:

- Proximal and distal breakpoints are enriched for segmental duplications, but with low sequence similarity: 8/41; 20%.
- Only one breakpoint lies within a segmental duplication: 19/41; 46%).
- No segmental duplications in the vicinity of both breakpoints: 14/41; 34%.

5.1.5 Segmental duplication and CNPs in balanced and unbalanced rearrangements

Contrary to our findings in unbalanced rearrangements, both breakpoints in patients with constitutive balanced translocations were never found to be flanked by highly similar segmental duplications. Figure 17 shows segmental duplication frequency in the breakpoint regions of 41 balanced translations and 22 unbalanced aberrations.

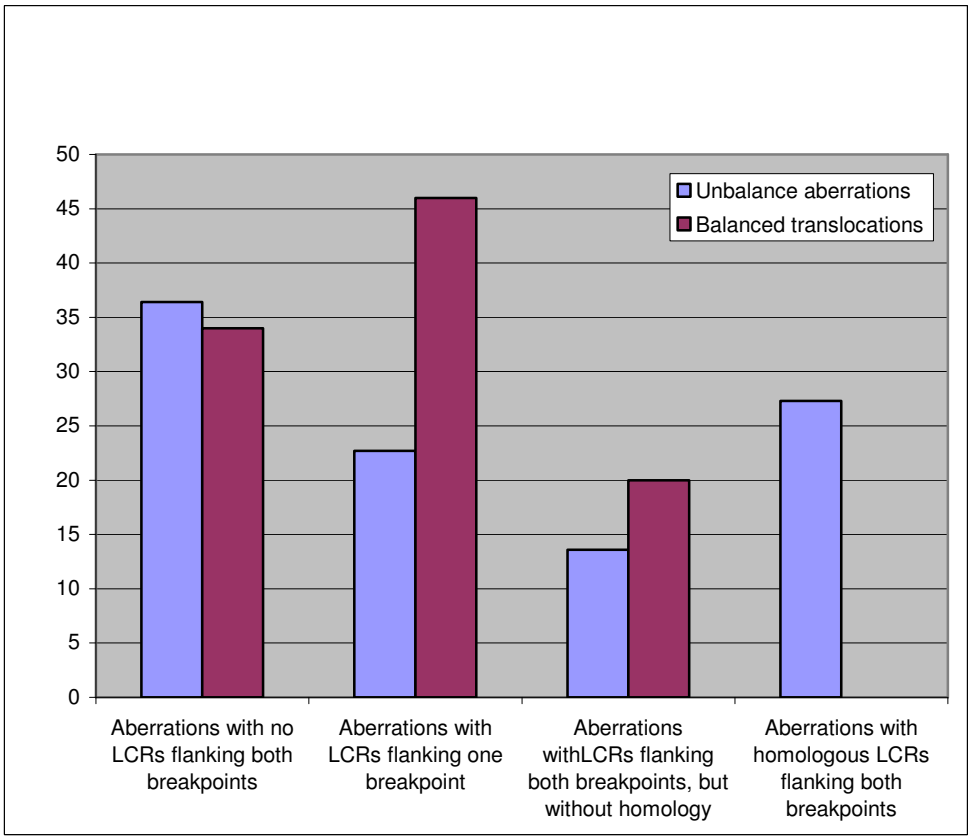


Figure17: Comparison of segmental duplication frequency and homology in breakpoint regions** of balanced and unbalanced aberrations. Data are based on a 400kb interval centered on the breakpoint. *LCR (low copy repeat) is same as segmental duplication. **Defined as 400kb interval centered around the respective breakpoint.

Figure 18 shows the frequency of DNA copy number polymorphisms at the breakpoint regions of 41 balanced translocations and 22 unbalanced aberrations.

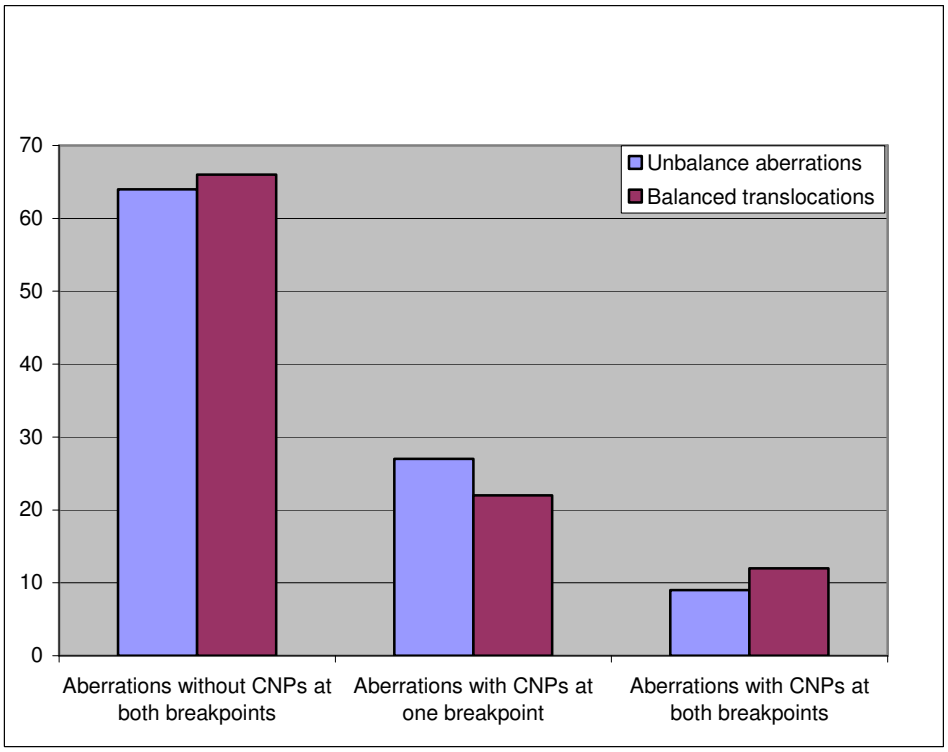


Figure 18: Comparison of CNP frequency in breakpoint regions** of balanced and unbalanced aberrations. Data are based on a 400kb interval centered on the breakpoint. *CNP: copy number polymorphism. ** Defined as 400kb interval centered around the respective breakpoint.

In order to avoid any loss of information due to the way of breakpoint definition, the whole procedure was repeated with breakpoint regions defined as 200kb and 1 Mb interval centered around the respective breakpoint, respectively. In Supplemental Material, Table S1 to Table S5 show the results obtained for the 200Kb interval, whereas Table S6 to Table S10 shows the results for the 1 Mb interval. Although the results differed between the intervals with different length, the underlying trends could still be observed. In this regard, interval size has no significant bearing on the conclusions.

5.1.6 Segmental duplication could mediate non-allelic homologous recombination (NAHR)

In this study, I have found a conspicuous clustering of segmental duplications at or near the borders of sub-microscopic deletions and duplication that were identified in mental retarded patients. In 41% of these imbalances, both breakpoint regions carried segmental duplications and in two third of these cases,

sequence comparisons identified NAHR as the most likely cause of rearrangement. Several reasons may account for this unexpected observation. For instance, it could be due to an ascertainment bias: deletions or duplications arising from NAHR were associated with a more pronounced phenotype than those due to non-homologous end joining (NHEJ), which may affect the dosage in fewer genes. Alternatively, this observation may reflect certain flexibility in the choice of the most appropriate DSB repair pathway, and the presence of homologous sequences in the immediate vicinity of the damage can shift the decision in favour of NAHR. Therefore it seems plausible that DSB repair by NAHR is favoured in regions, where segmental duplications are clustered. The similarity of duplicated sequences, which is a prerequisite for NAHR, may be maintained by gene conversion. The reciprocal relationship of aberration size and frequency of NAHR perfectly matches this idea. Finally, if the distances between segmental duplications are small, there will be a low probability for meiotic crossover in the intervening fragments. This increases the relative frequency of NAHR, since such connections can serve as anchors to prevent chromosome slippage (Inoue and Lupski, 2002).

5.1.7 NAHR involving segmental duplications cannot explain all unbalanced rearrangements

NAHR between segmental duplications cannot be the sole mechanism underlying unbalanced rearrangements; after all, only nine out of 22 patients had segmental duplications in both breakpoint regions. Moreover, in some of these cases, the replicated sequences found near the deletion/duplication borders were not similar, and in other patients, segmental duplications were only observed at one of the two breakpoints.

However, this does not strictly rule out NAHR in these cases, as it is possible and even likely that many of segmental duplications are not represented in the human reference sequence because they were overlooked during sequence assembly (Cheung et al., 2003; She et al., 2004). Also, homologous segments that are necessary for NAHR can actually be fairly small, as illustrated by recombination

events observed between Alu or Mer5B elements (Deininger and Batzer, 1999; Shaw and Lupski, 2005). Given the limited resolution of this analysis, such small repetitive elements may have been missed. However, even the presence of segmental duplications at one of the breakpoints may be enough to predispose for unbalanced rearrangements. In a recent report on Xq22.2 duplications in Pelizaeus-Merzbacher syndrome (Woodward et al., 2005), a one-sided distribution of segmental duplications at the chromosomal breakpoints has been explained by a mechanism involving homologous and non-homologous recombination (Richardson and Jasin, 2000). The resolution of the BAC array is not high enough to decide whether or not this mechanism accounts for some of the duplications in this study, but it cannot explain the respective deletions since this process is characterised by the generation of duplicated sequence. In two other studies, one with a deletion in a patient with Pelizaeus-Merzbacher like phenotype (Inoue et al., 2002) and the other dealing with Smith-Magenis syndrome (Shaw and Lupski, 2005), segmental duplications were also confined to one of the breakpoints. Detailed sequence analysis provided clear evidence for the involvement of NHEJ in these cases.

5.1.8 NAHR involving segmental duplications cannot explain balanced rearrangements

In patients with balanced translocations, no evidence was found for a major role of NAHR involving segmental duplications. Eight patients displayed low copy repeats in both breakpoint-spanning BAC clones, but the sequence similarity was only minimal. Inter-chromosomal duplication events are thought to have occurred more frequently at early stages of genome evolution, which explain why on average, the similarity between duplicated sequences on different chromosomes is lower than between duplicated sequences on the same chromosome (Zhang et al., 2005). Therefore, the chance to find a highly similar sequence on a heterologous chromosome is relatively low. Moreover, NAHR between duplicated sequences on different chromosomes may be restrained by the larger spatial distance in the nucleus.

Almost half of balanced translocations (19/41) displayed clustering of segmental duplication at one of the breakpoints, similar to the finding in patients with unbalanced rearrangements Thus, segmental duplication seem to be involved, but it is unlikely that NAHR is an important cause of balanced chromosomal rearrangements.

5.1.9 Clusters of segmental duplications may cause chromosomal instability

Given the conspicuous clustering of segmental duplications in breakpoint regions and the various arguments against a major role of NAHR, it is tempting to speculate that segmental duplications decrease the stability of DNA. This instability does not seems to be dependent on recombination events and may not only be caused by the potential of segmental duplications to form secondary structures at replication forks, but may also involve other e.g. epigenetic, mechanisms that are not yet understood.

In summary, we have demostrated an accumulation of segmental duplications at chromosomal breakpoints in mentally retarded patients, and it is possible that their presence predisposes chromosomes to rearrangement. However, it is also possible that segmental duplications are not always the cause of chromosomal instability, e.g. both may be secondary to other genetic or epigenetic factors, which are hitherto unknown (Bailey et al., 2004; Eichler and Sankoff, 2003).

5.2 Mapping of balanced translocation breakpoints using array CGH and CGHPRO

Disease-associated balanced chromosomal rearrangements form a unique resource for bridging genotypes and phenotypes (Bugge et al., 2000) and fine-mapping of chromosomal breakpoints in these patients has led to the identification of many disease genes. The traditional methods of mapping breakpoints in balanced translocations consist of FISH and Southern Blot analysis. Typically FISH mapping and ending up with the identification of a breakpoint-spanning clone requires several rounds of hybridisation starting with widely spaced clones. The whole process is rather labour-intensive and time-consuming, which has limited the large-scale application of this strategy for identifying disease genes.

As introduced in Section 3.2, Array CGH has been mainly used to detect copy number changes. In order to study balanced translocation, a novel technique termed ‘array painting’ has been developed (Fiegler et al., 2003; Veltman et al., 2003), which combines array CGH and chromosome sorting. As array CGH is the high-resolution variant of CGH, array painting is essentially an advanced variant of reverse chromosome painting (Carter et al., 1992). By replacing metaphase chromosomes with DNA sequences spotted on microarrays as hybridization targets, array painting greatly improves the resolution that can be attained by conventional reverse chromosome painting. As shown in Figure 19, array painting involves two steps. First, the two derivative chromosomes are isolated by preparative flow-sorting. The sorted chromosomes are then amplified, differentially labeled, and hybridized onto DNA microarrays. Fluorescence intensities are quantified with scanning device. Plotting the signal intensity ratios can reveal the DNA segment containing the respective junction fragment.

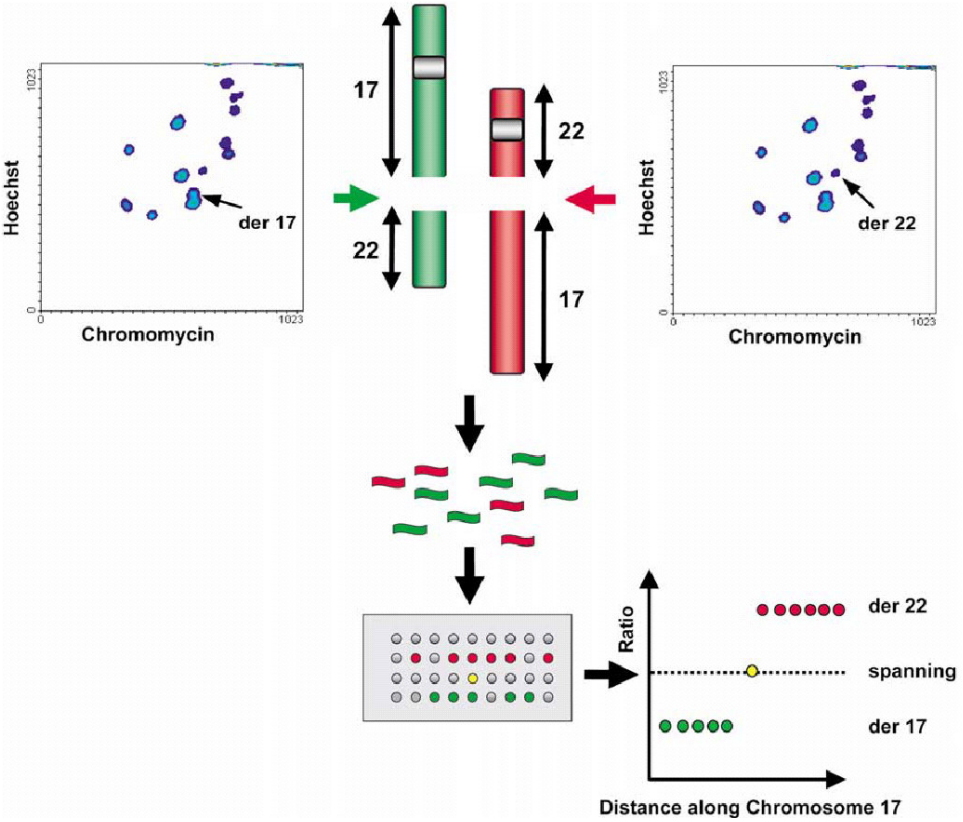


Figure 19: Principle of array painting (Gribble et al., 2004). First, the 2 derivative chromosomes, der 17 and der 22 are flow sorted. Then the sorted chromosomes are amplified, differentially labelled and hybridised onto DNA microarray. Fluorescence intensities are quantified with scanning device. Plotting the signal intensity ratios can reveal the DNA segment containing the respective junction fragment.

We have streamlined this approach even further. Below, I describe the successful application of this novel protocol to fine map the breakpoint in balanced translocation t(1;13). This protocol combining high-resolution array CGH analysis of flow-sorted chromosomes, the generation of DNA subarrays covering the respective breakpoint-spanning BACs, and long range PCR across the breakpoints and eventually sequencing of the junction fragments.

5.2.1 Cytogenetic investigation of the translocation

Chromosome analysis (GTG banding) in the proband had identified a balanced translocation between the short arm of chromosome 1 and the long arm of chromosome 13 (Figure 20). The karyotype was 46,XY,t(1;13)(p22;q14).

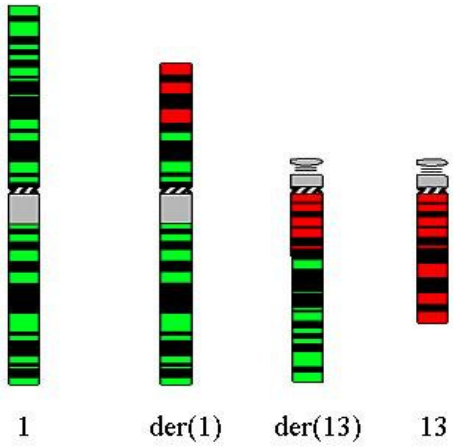


Figure 20: Partial karyotype of the patient with balanced translocation, showing the normal and derivative chromosomes 1 (green) and 13 (red).

5.2.2 Array CGH investigation of the translocation

Co-hybridization of differentially labeled DNA from the flow-sorted derivative chromosomes (der(1) and der(13), respectively) on the BAC array clearly showed the breakpoint regions on both derivative chromosomes (Figure 21 A B C). With the Zoom-in function of CGHPRO, even the breakpoint-spanning BAC clones, RP11-764P11 on chromosome 1 and RP11-339I10 on chromosome 13, could be identified.

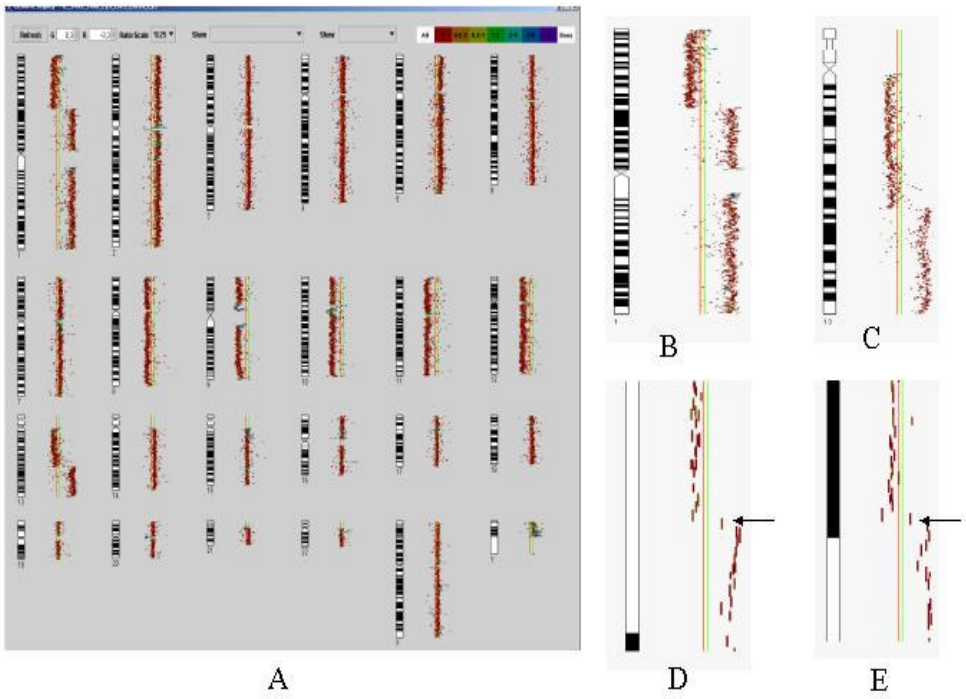


Figure 21: A) Results of the hybridisation of differentially labeled DNA from derivative chromosomes (der(1) and der(13)) on a whole-genome 36k array CGH

chip. B and C) Chromosome display of chromosomes 1 (B) and 13 (C), clearly indicating the translocated regions. D and E) Zoon-in view at the breakpoints reveals breakpoint-spanning BAC clones (arrows): RP11-764P11 on chromosome 1 and RP11-339I10 on chromosome 13.

5.2.3 Confirmation of the chromosome breakpoints by FISH analysis

To confirm the results, FISH experiments were performed. As shown in Figure 22 A, hybridisation with BAC clone RP11-764P11 gave rise to signals on both derivative chromosomes [(der(1) and der (13)], as well as a hybridisation signal on the normal, non-rearranged chromosome 1, thus confirming RP11-764P11 as the breakpoint-spanning clone on chromosome 1. Similarly, on chromosome 13, BAC clone RP11-339I10 was shown to span the breakpoints on both derivative chromosomes (Figure 22B).

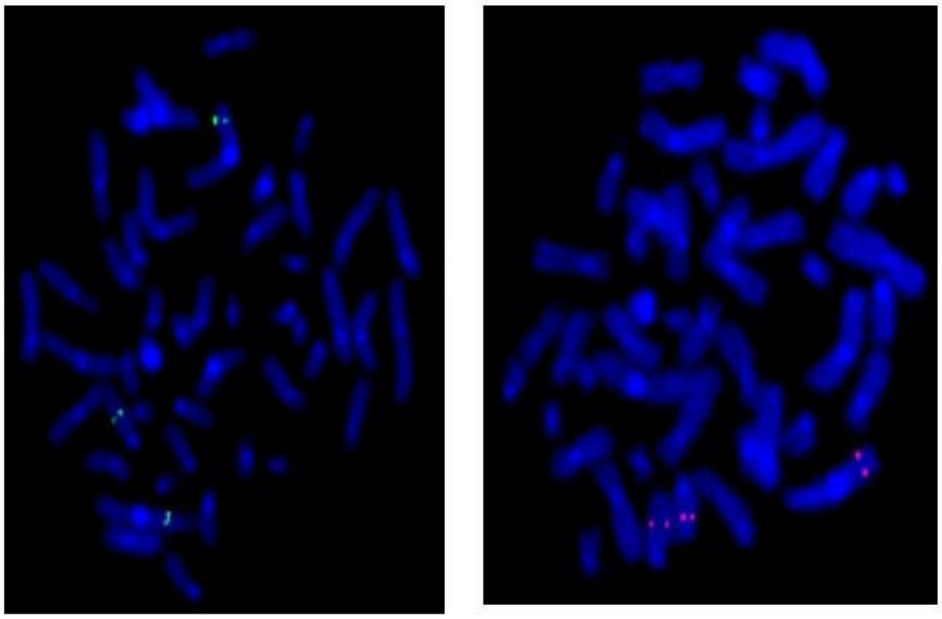


Figure 22: a) FISH result with the breakpoint-spanning BAC RP11-764P11 (green signals, and b) BAC RP11-339I10 (red signals). The three paired hybridisation signals correspond to the breakpoints of the derivative chromosomes and the respective normal counterpart (i.e., chr.1 or 13)

5.2.4 PCR fragment subarray

To further narrow down the breakpoint intervals, a so-called ‘sub-array’ was spotted with PCR products amplified from short segments distributing evenly along the two breakpoint-flanking clones, RP11-764P11 and RP11-339I10.

A customized Perl script, which is part of the CGHPRO package, was used to design the primers for the amplicons spotted on the subarray. The script first divides the BAC clones into evenly distributed intervals of 2 kb. Then, by using the Primer3 software (Rozen and Skaletsky, 2000), it designs primers for the generation of PCR probes within each of these intervals (average probe size: 500-800bp). To confirm that the amplicons are specific for the target region, the script searches the whole human genome for the presence of the amplicon sequences by using BLAST with the default parameter. Amplicons with more than one match in the genome are excluded from PCR amplification and spotting.

In total, 175 PCR products covering the breakpoint spanning BAC clones, RP11-764P11 and RP11-339I10, were generated and spotted on a subarray (for further technical details, see Section 8.1.2.6). Co-hybridisation of labeled DNA from the flow-sorted chromosomes on the subarray further narrowed down the two breakpoints, as shown in Figure 23.

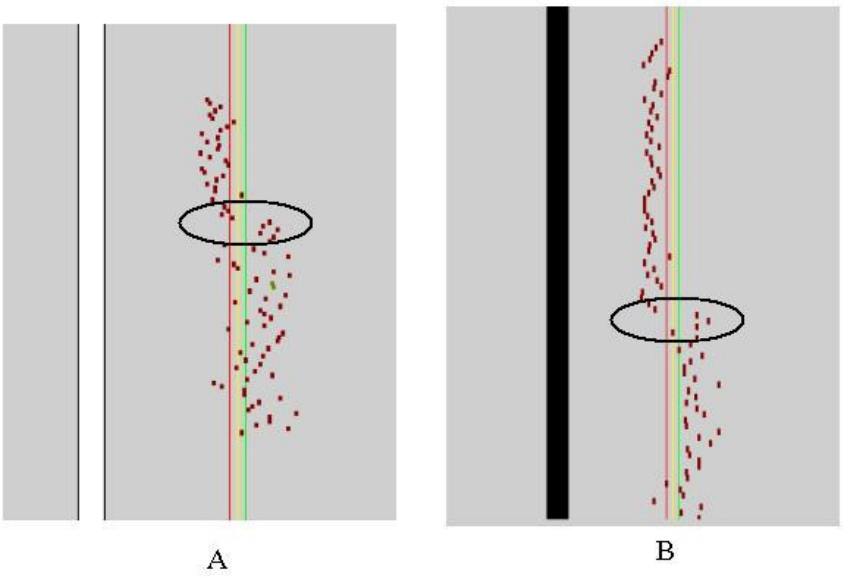


Figure 23:. Flow-sorted DNA from derivative chromosomes 1 (A) and 13 (B) was hybridised to a high resolution subarray spotted with multiple fragments of approximately 800 bp in size for fine mapping of the translocation breakpoint. The breakpoints (encircled) are thus mapped to a region of about 2-4 kb.

5.2.5 Long-range PCR and sequencing

The junction fragments from the derivative chromosomes were amplified by long-range PCR with primer pair 3 and 8 (more details, see Section 8.1.2.7). Sequencing of the specific PCR products and subsequent BLAST search against the human genome reference sequence mapped the chromosome 1 breakpoint to 67776107 bp, the chromosome 13 breakpoint to 71824031 bp (May 2004 UCSC Genome Browser, NCBI build 35). The sequencing chromatograph of the 2 amplified junction fragments and the corresponding genomic sequence from normal chromosome 1 and chromosome 13 are shown in Figure 24.

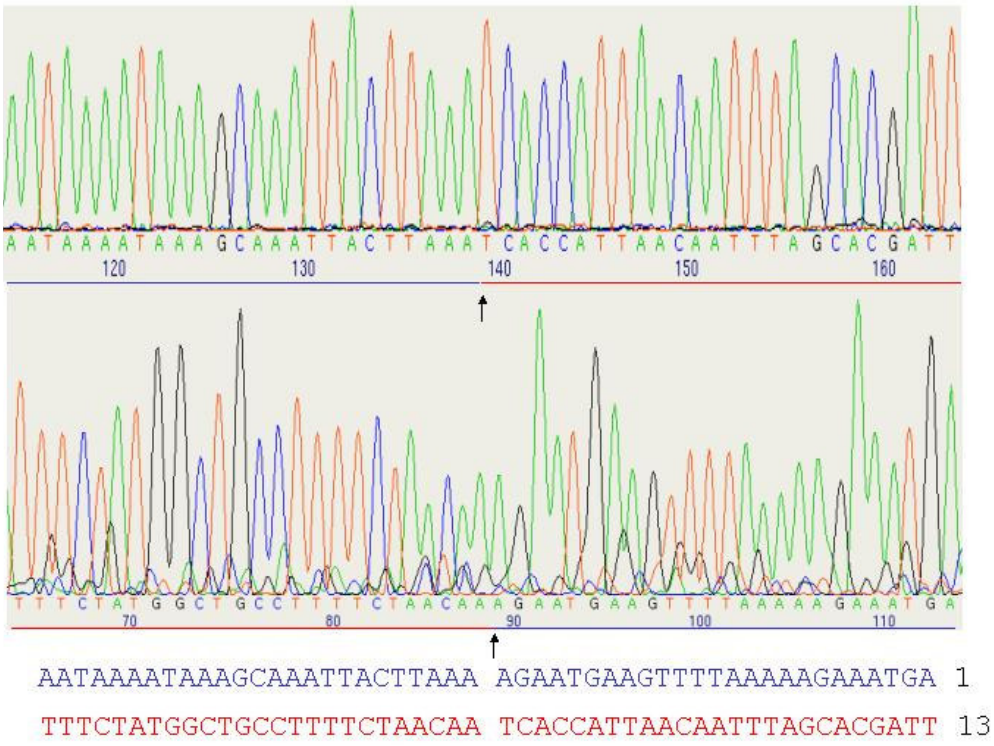


Figure 24: Sequencing chromatograph of the 2 amplified junction fragments and the genomic sequence from normal chromosome1 and chromosome 13. Arrows marked the exact breakpoints.

Genomic sequences around the translocation breakpoints were searched for specific features that may promote chromosome instability. In the junction fragments, microhomology and loss of one adenosine was observed, two characteristic features usually associated with NHEJ. On chromosome 1, a Mer-1 type element, MER5C, was found to span the breakpoint, while the short interspersed element (SINE) ALUJo was found to reside about 60 bps away from

the breakpoint on chromosome 13. It is noteworthy that a Mer1-type element has been implicated in two related rearrangements (Abeysinghe et al., 2003) deposited in Gross Rearrangement Breakpoint Database [GRaBD; http://www.uwcm.ac.uk/uwcm/mg/grabd/grabd.html].

5.2.6 Advantage of array painting in mapping balanced translocation breakpoints

The strategy for chromosome breakpoint analysis described here is less laborious, less time-consuming and thus less expensive than previously employed methods such as single clone FISH experiments and Southern blotting. The costs of manual labour and reagents of the traditional method easily outweigh the considerable costs of high-resolution array CGH and chromosome flow-sorting experiments. Even though the application of this protocol is limited to the analysis of derivative chromosomes that can be separated by flow sorting, its implementation promises to pave the way for large-scale breakpoint mapping and gene finding in patients with disease-associated balanced translocations.

Hitherto, chromosome breakpoint analysis has been nearly exclusively restricted to de-novo rearrangements associated with congenital or early-onset disorders. However, such cases represent only a small percentage of carriers of balanced chromosome rearrangements. Balanced chromosome rearrangements were also shown to be associated with complex and late-onset disorders, and the identification of genes affected by the breakpoints could elucidate candidate genes for complex disorders such as schizophrenia, dyslexia, Tourette's syndrome and psoriasis. Therefore, the availability of fast and cost-efficient methods of breakpoint analysis should also contribute to the identification of genetic factors in the etiology of late onset and complex disorders.