4   Results: Bioinformatic investigation of rearrangement breakpoints

4.1   Introduction

The methods used to obtain the sequence across the translocation junctions in 3 patients with *de novo* constitutional reciprocal translocations have been previously discussed in Chapter 3. This chapter discusses the effect the translocation may have on each patient's phenotype and the possible mechanisms underlying the rearrangements.

Translocation breakpoint positions were originally mapped in NCBI Build 35. The remapped breakpoint positions according to NCBI Build 36 of the human genome are detailed in Table 4.1.

Patient	Chromosome	Breakpoint position May 2004	Breakpoint position March 2006
t(2;7)(q37.3;p15.1)	2	236,548,579	236,431,318
	7	30,983,603	31,176,888
t(3;11)(q21;q12)	3	130,755,505	130,755,497
	11	61,032,619	61,032,619
t(7;13)(q31.3;q21.3)	7	121,245,446	121,438,731
	13	71,023,877	71,023,877

Table 4.1   ***Comparison of translocation breakpoint positions in the UCSC genome browser between May 2004 and March 2006 builds.***

The genome reference sequence from NCBI Build 36 was used for all bioinformatic analysis discussed in this chapter.

4.2   Breakpoints and phenotypes

Translocation breakpoints can affect a patient's phenotype by directly disrupting a gene and affecting it's function, by distancing a gene from its regulatory elements, or placing it under the control of other regulatory elements.

4.2.1 Patient phenotypes (as published in (Gribble et al. 2005))

4.2.1.1 Patient t(2;7)(q37.3;p15.1) phenotype

This patient was referred at the age of eight months with mild generalised developmental delay, a slightly beaked nose, and adducted thumbs. At the age of two years and eight months his development was clearly delayed: he was able to sit but unable to walk unaided and he had no intelligible words. He had a dysmorphic appearance with brachycephaly, blepharophimosis, medially flared eyebrows, a broad nasal tip, short philtrum, thin upper lip, and prominent lower jaw.

4.2.1.2 Patient t(3;11)(q21;q12) phenotype

This patient is one of phenotypically discordant, monochorionic, monoamniotic twins born at 29 weeks' gestation. She had a congenital duodenal obstruction requiring surgery, complex congenital heart disease, and facial dysmorphism. Her twin sister who carries the same apparently balanced t(3;11) is clinically normal.

4.2.1.3 Patient t(7;13)(q31.3;q21.3) phenotype

This patient was referred at the age of six years because of developmental delay and autistic features. By the age of seven years, he had epilepsy, learning difficulties, disordered speech and language, and an autism spectrum disorder.

4.2.2 Direct disruption of a gene by the translocation breakpoint

The underlying genetic cause of some diseases has been found by studying patients with translocations who present with the same characteristics as patients showing deletions or duplications in the same region. For example, the *rnex40* gene identified as a candidate for DiGeorge syndrome by the analysis of a t(2;22)(q14;q11.21) translocation and the *ProSAP2* gene as a candidate for terminal 22q13.3 deletion syndrome by the investigation into a t(12;22)(q24.1;q13.3) translocation (Budarf et al. 1995 Bonaglia et al. 2001).

4.2.2.1 Direct gene disruption in patient t(2;7)(q37.3;p15.1)

Analysis of the two breakpoints in NCBI Build 36 revealed that the chromosome 2 breakpoint directly disrupted Centaurin, Gamma 2 (*CENTG2*), whilst the chromosome 7 breakpoint did not directly disrupt a gene (Figure 4.1).
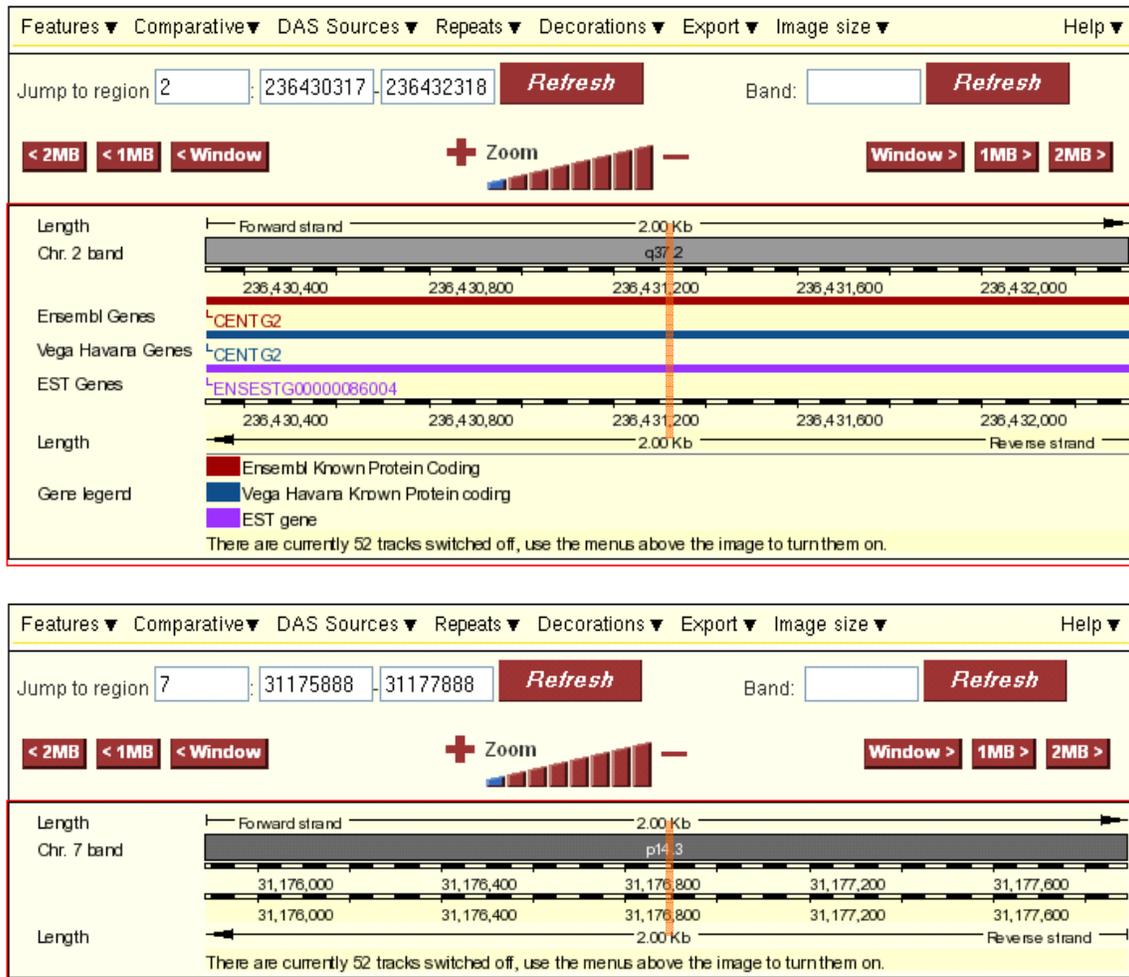
Figure 4.1 **NCBI Build 36 Ensembl downloads for the 2Kb region surrounding the chromosome 2 and chromosome 7 breakpoints for patient t(2;7)(q37.3;p15.1). The vertical orange bar marks the breakpoint position.**

CENTG2 (OMIM; 608651) is a gene belonging to a protein family involved in membrane trafficking and actin cytoskeleton dynamics – the ADP-ribosylation

GTPase-activating (ARF-GAP) family (Nie et al. 2002; Meurer et al. 2004). *CENTG2* has been considered to be a good candidate for developmental delay as it is expressed in the brain and nervous system, however the evidence for its involvement remains inconclusive (Wassink et al. 2005). Analysis of the chromosome 2 breakpoint in patient t(2;7)(q37.3;p15.1) showed that the breakpoint lay within the intron between exons 9 and 10 of *CENTG2*, resulting in a truncated protein.

4.2.2.2 Direct gene disruption in patient t(3;11)(q21;q12)

Analysis of the breakpoints for patient t(3;11)(q21;q12) showed that neither of the translocation breakpoints directly disrupted a gene (Figure 4.2).
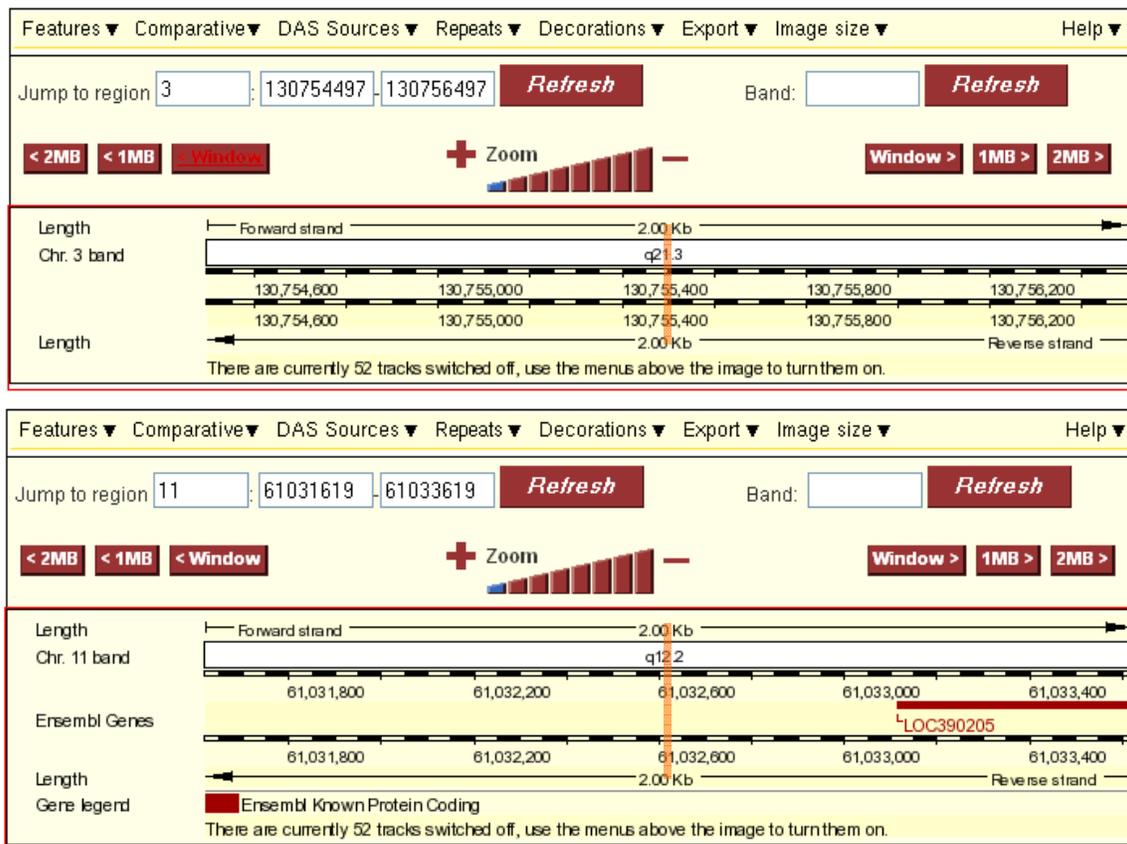
Figure 4.2 ***NCBI Build 36 Ensembl downloads for the 2Kb region surrounding the chromosome 3 and chromosome 11 breakpoints for patient t(3;11)(q21;q12). The vertical orange bar marks the breakpoint position.***

121

4.2.2.3 Direct gene disruption in patient t(7;13)(q31.3;q21.3)

Analysis of the derivative chromosome sequences showed that the chromosome 7 breakpoint directly disrupted the Protein-Tyrosine Phosphatase, Receptor-Type, Zeta-1 gene (*PTPRZ1*) (Figure 4.3) and the chromosome 13 breakpoint directly disrupted the Dachshund, Drosophila, Homolog of, 1 gene (*DACH1*) (Figure 4.4).
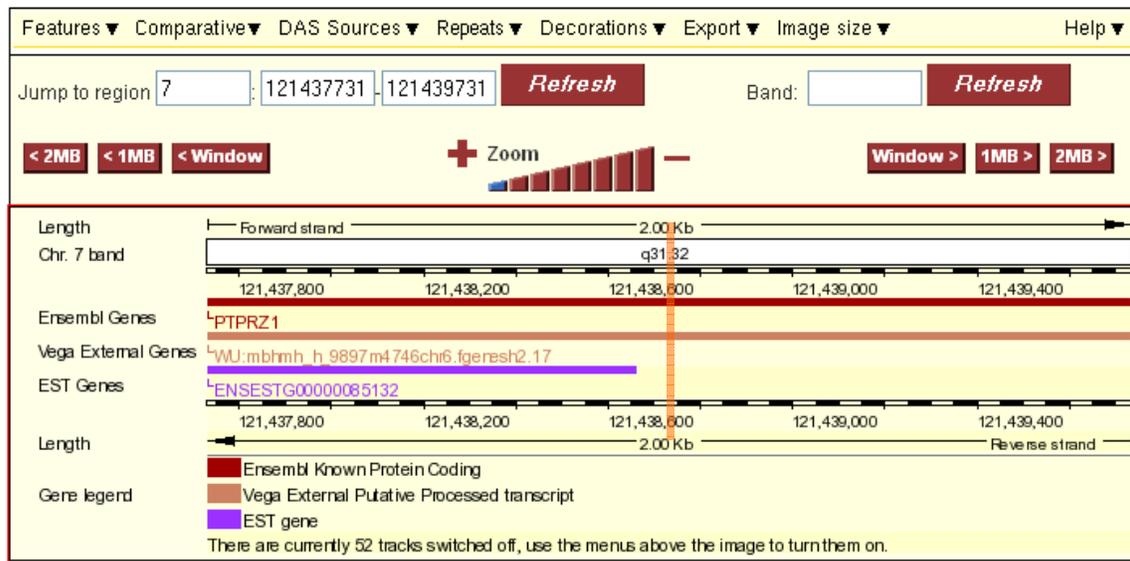
Figure 4.3   NCBI Build 36 Ensembl download for the 2Kb region surrounding the chromosome 7 breakpoint for patient t(7;13)(q31.3;q21.3). The vertical orange bar marks the breakpoint position.

PTPRZ1 (OMIM; 176891) is a member of the Protein Tyrosine Phosphatase (PTP) family and is expressed only in the central nervous system (Levy et al. 1993). There is a high level of expression in the embryonic CNS suggesting that it plays a key role in its development, and it is also expressed in the adult brain at sites of mitotic activity.

PTPRZ1 has been studied in direct relation to its links with autism however the study did not provide any evidence that it is causal to the autism spectrum

disorder (Bonora et al. 2005). This research was supported by a further study in the Japanese population (Marui et al. 2005).
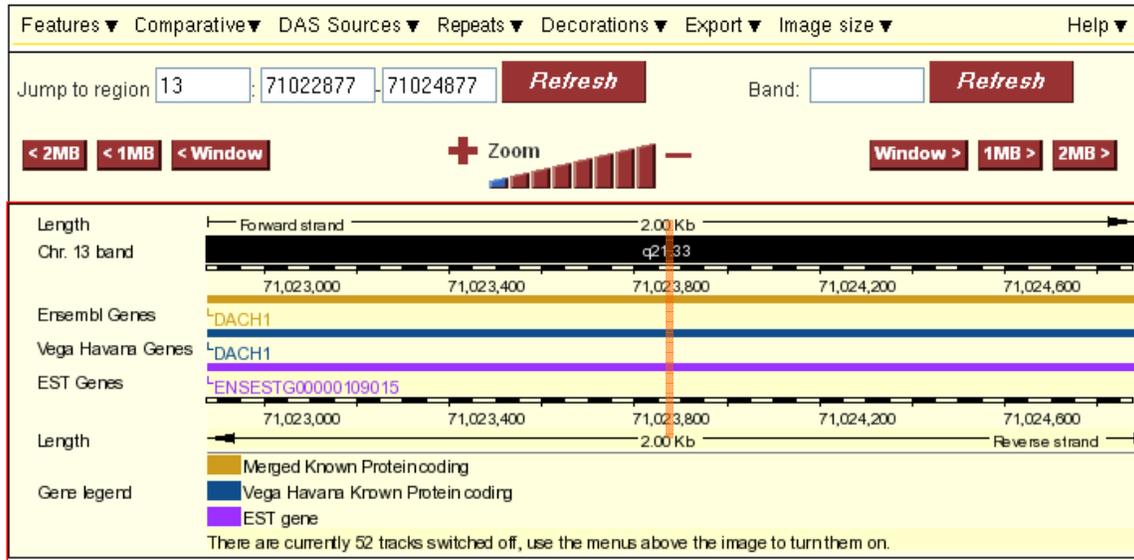
Figure 4.4 *NCBI Build 36 Ensembl download for the 2Kb region surrounding the chromosome 13 breakpoint for patient t(7;13)(q31.3;q21.3). The vertical orange bar marks the breakpoint position.*

DACH1 (OMIM; 603803) is the human homologue of the Drosophila 'dac' gene involved in leg and eye development (Hammond et al. 1998). As well as playing a key role in development, *DACH1* has been shown to inhibit transforming growth factor-ß (TGF- ß) induced apoptosis (Wu et al. 2003).

Analysis of both breakpoints in relation to the gene structures showed that the chromosome 7 breakpoint disrupted *PTPRZ1* within Exon 12, and the chromosome 13 breakpoint disrupted *DACH1* within Intron 7-8.

4.2.3 Translocation breakpoints and position effect

Positional cloning studies using balanced translocations have shown that in approximately 10% of cases, the breakpoints fall outside of the candidate causative gene (Crisponi et al. 2004). The furthest documented distance of a translocation breakpoint exerting an effect on a gene is 1.3Mb away in a patient with Campomelic Dysplasia (Velagaleti et al. 2005). Tools for the prioritisation of candidate genes in relation to a given phenotype have been developed to aid in genotype to phenotype correlation. Endeavour is a web-based system which has previously identified a novel gene in a 2Mb region involved in craniofacial development which is deleted in some patients with DiGeorge like birth defects (Aerts et al. 2006). The DECIPHER database will also prioritise genes in a region according to published data linking the gene to known characteristics using precise clinical diagnoses.

4.2.3.1 Position effect in patient t(2;7)(q37.3;p15.1)
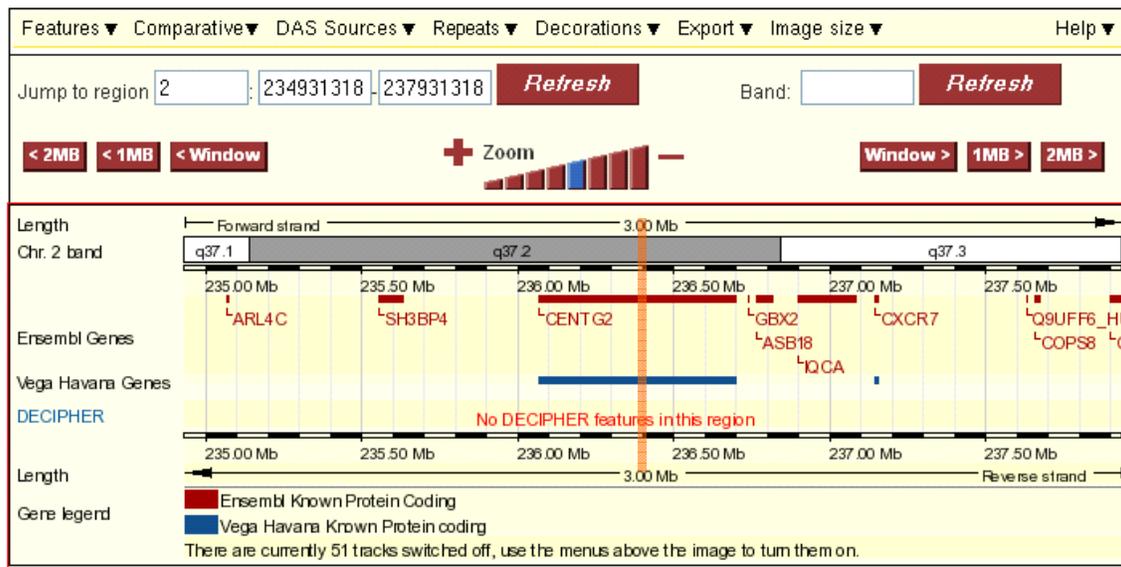
Figure 4.5 ***NCBI Build 36 Ensembl download for the 3Mb region surrounding the chromosome 2 breakpoint for patient t(2;7)(q37.3;p15.1). The vertical orange bar marks the breakpoint position.***

The 3Mb region surrounding the chromosome 2 breakpoint encompasses a total of 10 genes (Figure 4.5). Although Ensembl shows that there are no DECIPHER features there is a common 2q37.3 deletion syndrome in this region. Patients have a range of characteristics, but generally present with varying degrees of facial dysmorphism and mental retardation (Giardino et al. 2001; Aldred et al. 2004; Lukusa et al. 2004) similar to those exhibited by patient t(2;7)(q37.3:p15.1). The exact genetic basis of each characteristic has yet to be determined.

Analysis of the 3Mb region around the chromosome 7 breakpoint identified 24 genes within the potential range for a position effect (Figure 4.6).
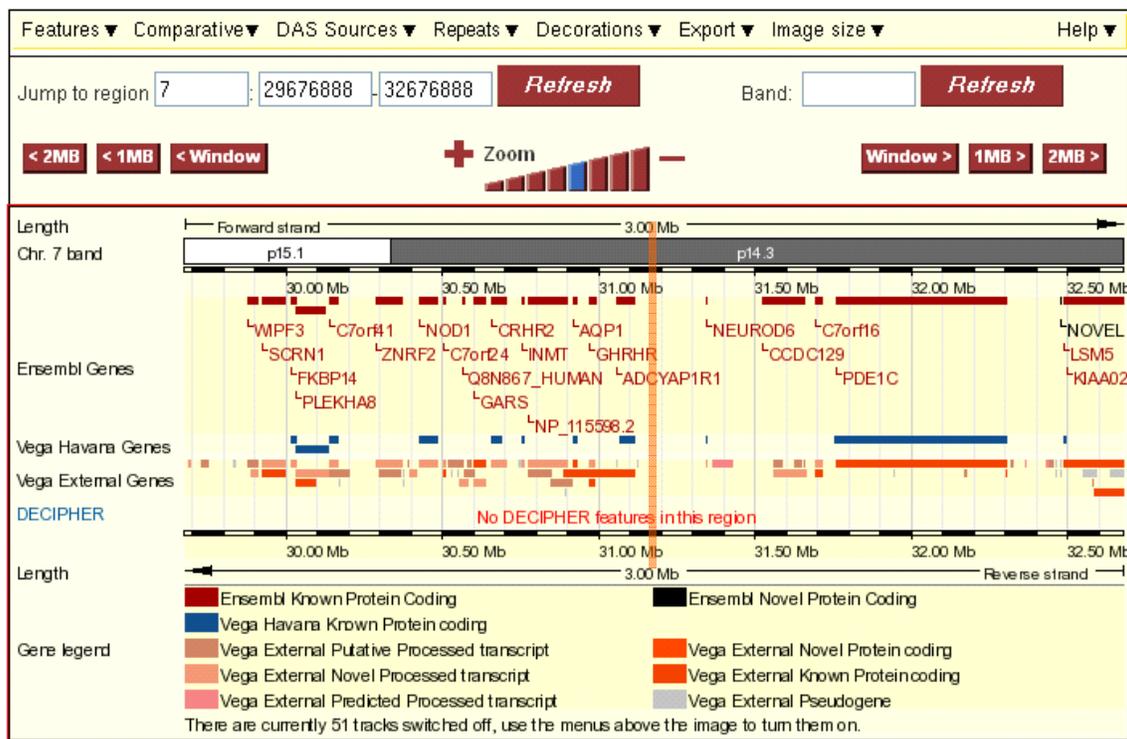
Figure 4.6 *NCBI Build 36 Ensembl download for the 3Mb region surrounding the chromosome 7 breakpoint for patient t(2;7)(q37.3;p15.1). The vertical orange bar marks the breakpoint position.*

In order to prioritise the genes according to their relevance to the patients phenotype, the DECIPHER database was used. The precise clinical characteristics were entered along with the 3Mb region around the translocation breakpoints for both chromosomes. A strong candidate gene for the patient's phenotype around the chromosome 2 breakpoint is the Collagen Type VI alpha 3 (*COL6A3*) gene. Mutations in *COL6A3* (OMIM; 120250) have been linked to general myopathy of varying severity (Pan et al. 1998; Demir et al. 2002). Analysis of the chromosome 7 breakpoint region using the DECIPHER prioritisation tool identified Aquaporin 1 (*AQP1*) as a candidate gene. *AQP1* (OMIM; 107776) encodes a membrane protein involved in water transport (Smith et al. 1993).

4.2.3.2 Position effect in patient t(3;11)(q21;q12)

The 3Mb region surrounding the chromosome 3 breakpoint contained 43 genes and the chromosome 11 breakpoint, 91 genes. The phenotypes for patient t(3;11)(q21;q12) were ill defined, so no useful data was obtained from the DECIPHER prioritisation tool and no other cases of rearrangement were identified from the database.
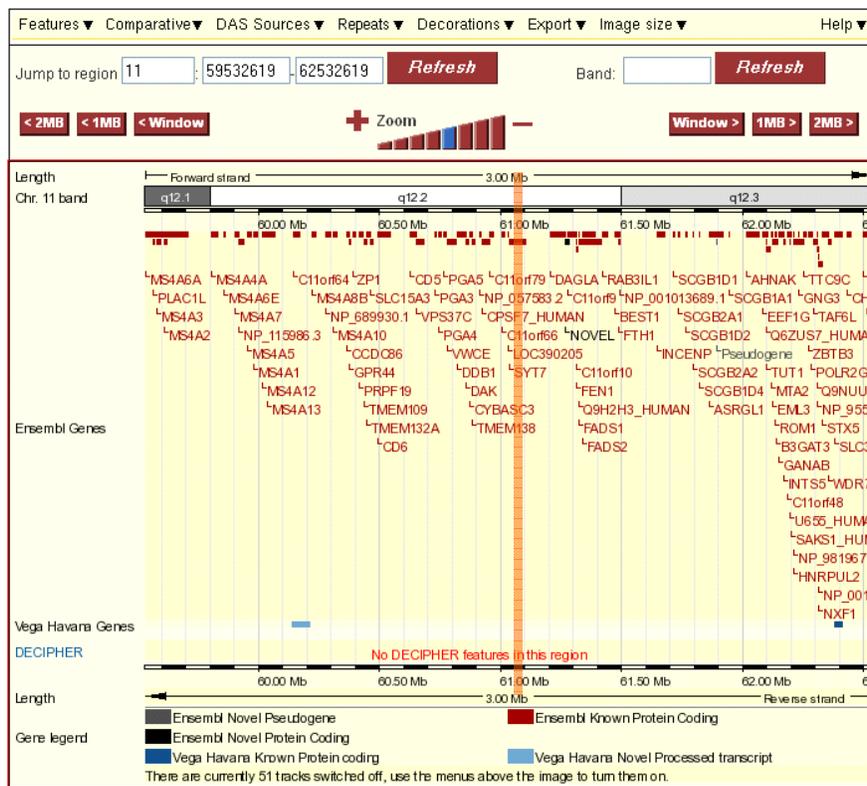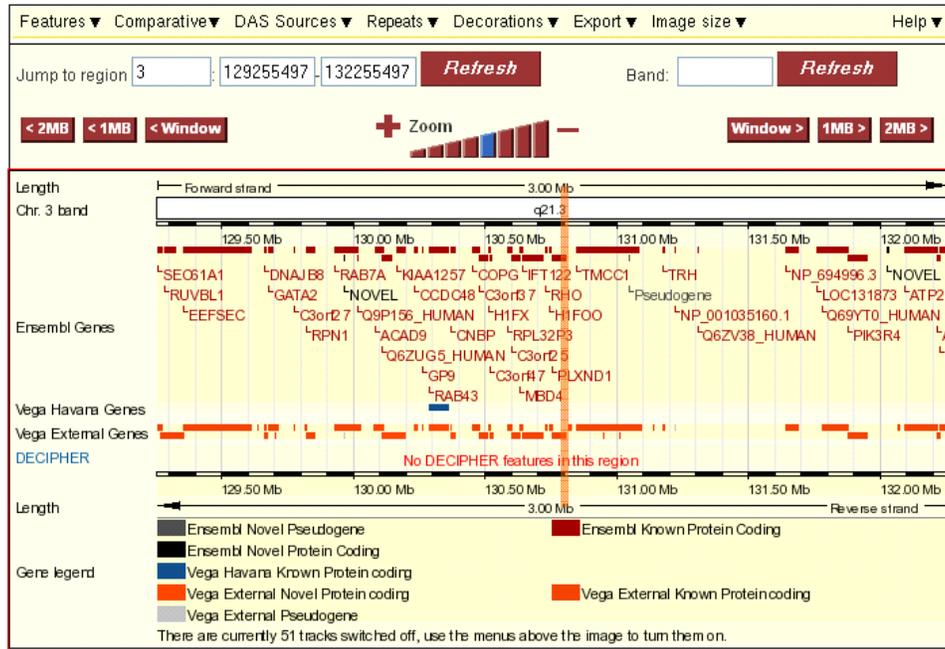
Figure 4.7 ***NCBI Build 36 Ensembl download for the 3Mb region surrounding*** <u>***A***</u> ***the chromosome 3 breakpoint and*** <u>***B***</u> ***the chromosome 11 breakpoint for patient t(3;11)(q21;q12). The vertical orange bar marks the breakpoint position.***

4.2.3.3 Position effect in patient t(7;13)(q31.3;q21.3)
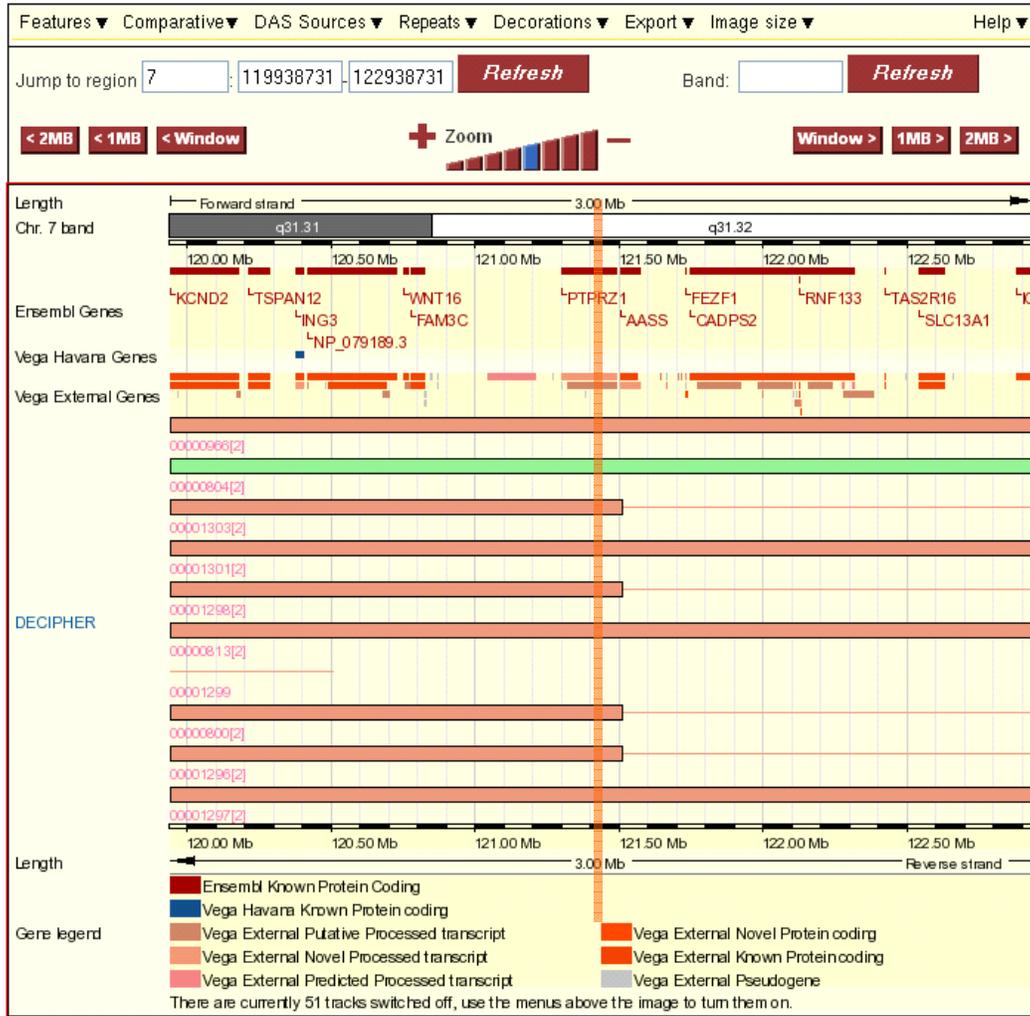
Figure 4.8 ***NCBI Build 36 Ensembl download for the 3Mb region surrounding the chromosome 7 breakpoint for patient t(7;13)(q31.3;q21.3). The vertical orange bar marks the breakpoint position.***

The Ensembl download shows that multiple patients have been entered into the DECIPHER database around the chromosome 7 breakpoint region (Figure 4.8). These patients exhibit characteristics similar to patient t(7;13)(q31.3;q21.3) including developmental delay and autistic features, however so far, no genes have been definitively identified as being causative to the phenotypes.

For the chromosome 13 breakpoint region (Figure 4.9), no patients have been entered in to DECIPHER. The phenotype for this patient has not been defined enough for DECIPHER to prioritise genes, however both the chromosome 7 and 13 breakpoints directly disrupt candidate genes as discussed in Section 4.2.2.3.
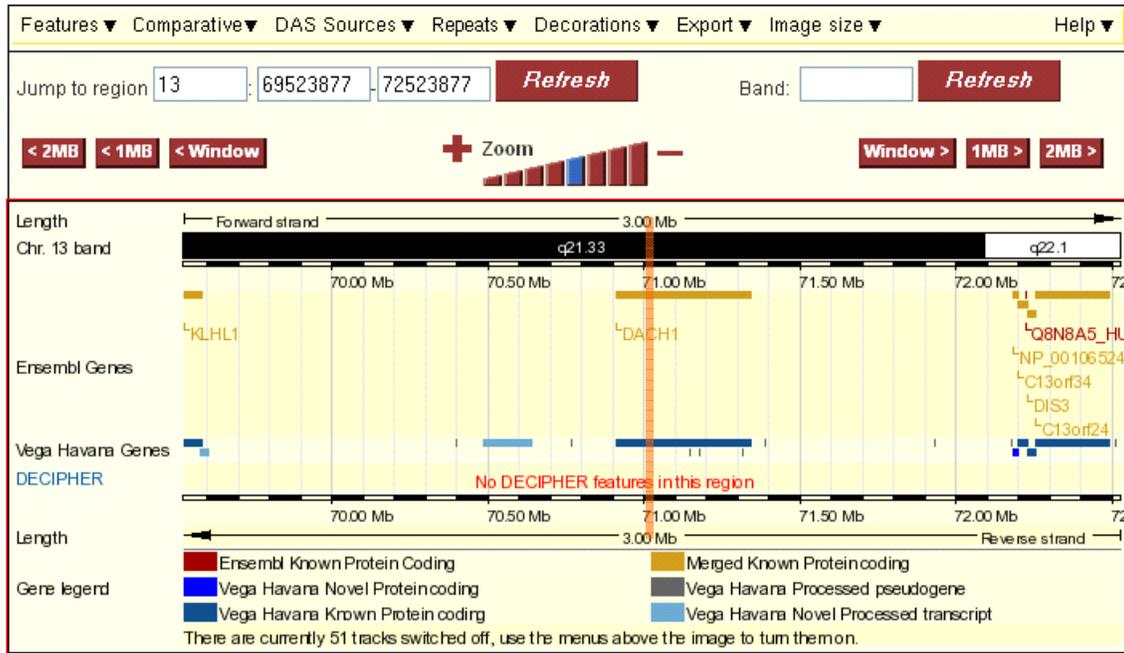
Figure 4.9 ***NCBI Build 36 Ensembl download for the 3Mb region surrounding the chromosome 13 breakpoint for patient t(7;13)(q31.3;q21.3). The vertical orange bar marks the breakpoint position.***

4.2.4 Duplication in patient t(2;7)(q37.3;p15.1) and phenotype

The region of chromosome 3 duplicated in patient t(2;7)(q37.3;p15.1) lies between two Alu repeats at 1.7Mb and 3.6Mb and was found to be approximately 1.9Mb in size (Figure 4.10).   Whilst the precise breakpoints could not be determined at the basepair level, the distal breakpoint was mapped to a 260bp region from 1,756,993 to 1,757,252bp and the proximal breakpoint to a 258bp

region from 3,614,132 to 3,614,389bp as discussed in Chapter 3 (there is no variation in the genomic positions for chromosome 3 between NCBI Builds 35 and 36).
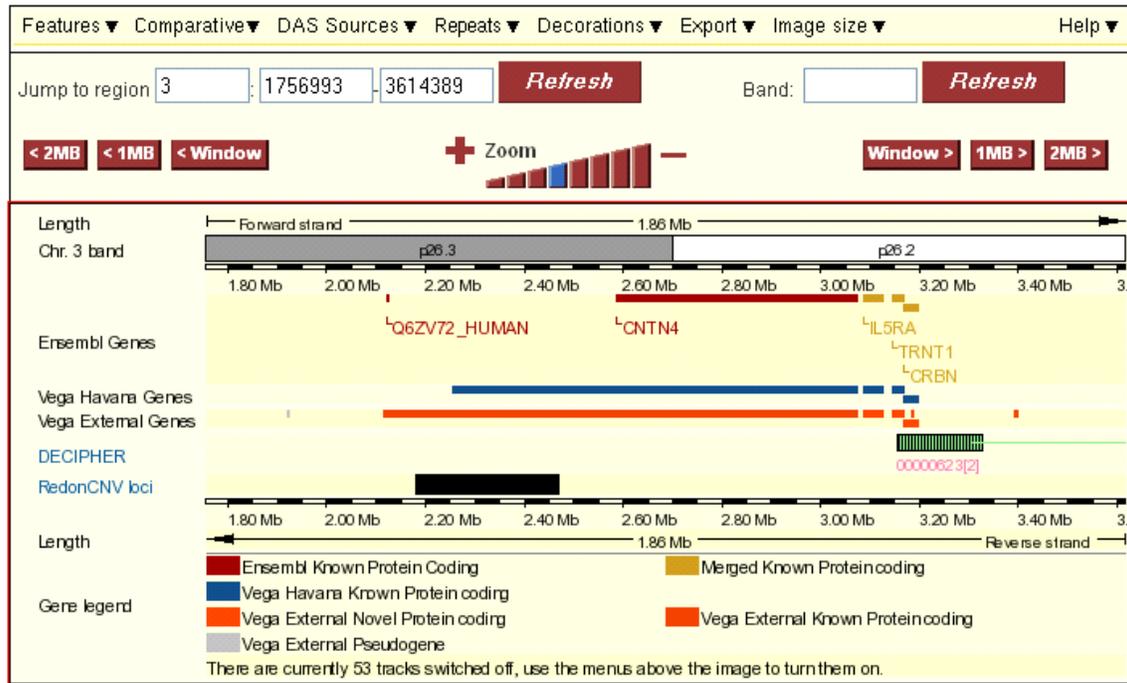
Figure 4.10   Ensembl download showing the duplicated region at 3p26.2-3p26.3 for patient t(2;7)(q37.3;p15.1).

This region contains a known CNV locus at approximately 2.3Mb along the chromosome. 2 out of 270 phenotypically normal individuals used to generate this CNV data showed amplification at this locus. A literature search revealed reports of several aberrations in this region of the genome; the most common being a 3p deletion syndrome characterised by developmental delay, growth retardation and dysmorphic features (Neri et al. 1984; Moncla et al. 1995; Lukusa et al. 1999; Wahlstrom et al. 1999; Cargile et al. 2002; Fernandez et al. 2004; Lalli et al. 2007). The DECIPHER track for this region reports a patient with a 1.4Mb duplication of dup(3)(p26.1;p26.2) presenting with an absent uterus and

developmental delay. There was only one published report of a duplication case covering the same region; the patient showed a two clone duplication (by array CGH on a 1Mb resolution microarray) and presented with dysmorphic features and learning disabilities (Shaw-Smith et al. 2004). However, as was the case for patient t(2;7)(q37.3;p151), the duplication was inherited from a phenotypically normal father. It is unlikely that the duplication plays a major part in the proband's phenotype – it is more likely that the translocation breakpoints are responsible for the phenotype as discussed in 4.2.2.1 and 4.2.3.1.

4.3   Breakpoints and genome architecture

Investigations into recurrent rearrangements have shown that genomic structures may pre-dispose the sequence to rearrangement. For example; Sotos syndrome is mediated by low copy repeats (LCRs) (Kurotaki et al. 2005; Visser et al. 2005), the duplications associated with Charcot-Marie-Tooth disease type 1A and the deletions associated with hereditary neuropathy with liability to pressure palsies are both mediated by LCRs on chromosome 17p12 (Shaw et al. 2004) and the recurrent t(11;22) translocation is mediated by AT-rich palindromic repeats (Edelmann et al. 2001; Kurahashi and Emanuel 2001). However, the mechanisms underlying *de novo* constitutional translocations have yet to be elucidated.

4.3.1   Analysis of sequence at breakpoints

The sequence data available for constitutional non-recurrent reciprocal translocation breakpoints in the literature is summarised in Table 4.2. The majority of these translocations are accompanied by the loss, addition or duplication of a small number of bases, believed to be a characteristic of non-homologous end joining as discussed in the Introduction.

Translocation	insertions at bpt	deletions at bpt	duplications at bpt	Reference
t(X;21)(p21;p12)	3bp on derX	71-72bp from chrX 16-23bp from chr21		Bodrug et al., 1987
t(X;2)(p21;q37)	2bp on der2	1-3bp from chr2 0-2bp from chrX		Bodrug et al., 1991
t(X;4)(p21;q35)	3bp on derX	2-3bp from chrX 7-8bp from chr4		
t(X;1)(p21;p34)	2-5bp on derX	4-7bp from chrX		Cockburn, 1991
t(X;4)(p21.2;q31.22)		5Kb from chrX 4-6bp from chr4		Giacalone and Francke, 1992
t(4;22)(q12;q12.2)				Arai, Ikeuchi and Nakamura, 1994
t(2;22)(q14;q11.21)		5-10bp from chr2 1-6bp from chr22		Budarf et al., 1995
t(X;5)(p21;q31.1)	4bp on derX 6bp on der5	3bp from chrX		van Bakel et al., 1995
t(X;9)(p21.1;p34.3)	7bp on derX 3bp on der9	1bp from chrX 55bp from chr9		Toriello et al., 1996
t(21;22)(p12;q11)	1bp on der21 or der22			Holmes et al., 1997
t(X;8)(p22.13;q22.1)				Ishikawa_Brush et al., 1997
t(17;22)(q11.2;q11.2)	109bp on der17	14bp from chr17		Kehrer-Sawatzki et al., 1997
t(6;7)(q16.2;p15.3)		518bp from chr7		Krebs et al., 1997
t(8;17)(p11.2;p13.3)			3bp of chr8	Kurahashi et al., 1998
t(1;10)(p22;q21)				Roberts, Chernova and Cowell, 1998
t(2;19)(q11.2;q13.3)		1bp from chr2 2bp from chr19		Yoshiura et al., 1998
t(6;12)(q16.2;q21.2)	15bp on der12	6bp from chr12	5bp of chr6	Ikegawa et al., 1999
t(1;6)(p22.1;q16.2)		1bp from chr6		Holder, Butte and Zinn, 2000
t(1;8)(q21.1;q22.1)			6bp of chr8	Matsumoto et al., 2000
t(1;11)q42.1;q14.3)	2bp on der11	4bp from chr11		Millar et al., 2000
t(12;22)(q24.1;q13.3)			5bp of chr22	Bonaglia et al., 2001
t(1;19)(q21.3;q13.2)				Nothwang et al., 2001
t(9;11)(p24;q23)	41bp on der9	2bp from chr11		Willett-Brozick et al., 2001
t(7;16)(q11.23;q13)				Duba et al., 2002
t(1;8)(p34.3;q21.12)	5bp on der1 12bp on der8	1bp from chr1 10bp from chr8		McMullan et al., 2002
t(2;8)(q31;p21)	5bp on der2 13bp on der8	7bp from chr8		Spitz et al., 2002
t(2;8)(q31;p21)	13bp on der8	2bp from chr8		Sugawara et al., 2002
t(6;13)(q21;q12)			2bp of chr6	Vervoort et al., 2002
t(7;22)(p13;q11.2)	13bp on der7 46bp on der22	75bp from chr7 4bp from chr22		Hill et al., 2003
t(6;11)(q14.2;q25)		8bp from chr6 9bp from chr11	16bp of chr11	Jeffries et al., 2003
t(4;22)(q35.1;q11.2)		1bp from chr4 ~168bp from chr22		Nimmakayalu et al., 2003
t(X;7)(p11.3;q11.21)				Shoichet et al., 2003
t(1;7)(q41;p21)		3-6bp from chr1 3-6bp from chr7		David et al., 2003
t(1;22)(p21.2;q11)				Gotter et al., 2004
t(3;8)(p14.2;q24.2)		5Kb from chr3		Rodriguez-Perales et al., 2004
t(2;6)(q24.3;q22.31)			18bp of chr6	Bocciardi et al., 2005
t(4;17)(q28.3;q24.3)		10-13bp from chr4 0-3bp from chr17		Velagaleti et al., 2005
t(1;7)(p22;q32)	2bp on der1 3bp on der7			Borg et al., 2005
t(4;15)(q27;q11.2)	1bp on der4			Schule et al., 2005
t(4;15)(q22.3;q21.3)	7bp on der4 37bp on der15	1bp from chr4 13bp from chr15		Klar et al., 2005
t(9;11)(q33.1;p15.3)				Tagariello et al., 2006
t(6;17)(p21.31;q11.2)		2bp from chr6	7bp of chr17	Mansouri et al., 2006
t(5;14)(q21;q32)	2bp on der14	1bp from chr5 1bp from chr14		Haider et al., 2006
t(17;22)(q21.1;q12.1)	4bp on der17			Gribble et al., 2007
t(2;7)(q37.1;q36.3)		11bp from chr2 1bp from chr7		
t(11;17)(p13;p13.1)	3bp on der11	2bp from chr11		
t(2;7)(q37.1;q21.3)				Bocciardi et al., 2007

Table 4.2   Summary of sequence data at the breakpoints of published reciprocal constitutional translocations.

4.3.1.1 Patient t(2;7)(q37.3;p15.1)

Alignment of the sequence across the two breakpoint regions revealed a 7bp deletion of chromosome 2 material, a 4bp duplication of chromosome 7 sequence and a 19bp insertion of unknown origin at the derivative 7 junction (Figure 4.11). An example of mitochondrial DNA insertion at the breakpoints has been observed in a familial t(9;11)(p24;q23) translocation (Willett-Brozick et al. 2001) however this is not the origin of the 19bp insertion for patient t(2;7)(q37.3;p15.1).

```
t(2;7)(q37.3;p15.1)
2r/c;   aaatgaaaatcccccaagatcacccaggaagcagccctgggtataagcctcacactcccaacat
chr2;   actgtctgcccatgttgggagtgtgaggcttatacccagggctgcttcctgggtgatcttgggg
der2;   actgtctgcccatgttgggagtgtgaggcttataTTAAGGCTTTAAATATCTGAATGTAGCCAT
der7;   cagcccaggaagcagtttaatctattaatctttgTTAATAGATTAAACAAAAACTGAGCGTAAA
chr7;   GGTACCTCATGGCTACATTCAGATATTTAAAGCCTTAATAGATTAAACAAAAACTGAGCGTAAA
7r/c;   ACCCTGAGTTTACGCTCAGTTTTTGTTTAATCTATTAAGGCTTTAAATATCTGAATGTAGCCAT
```

Figure 4.11   *Sequence alignment of translocation junctions against human reference sequence for patient t(2;7)(q37.3;p15.1).*

4.3.1.2 Patient t(3;11)(q21;q12)

In this patient there is a 1-3bp deletion of chromosome 3 sequence, and a 8-10bp deletion of chromosome 11 sequence (Figure 4.12). The precise number of basepairs involved cannot be determined due to the 2 bp homology (GC) at the breakpoints of both chromosomes.

```
t(3;11)(q21;q12)
chr3    acaccctgatcctgagttcactcctcggccccgccccatccaagcgccccgtgcggctggcgtt
der3;   acaccctgatcctgagttcactcctcggcccGGGCTGGATGGGCAGGTAGGGGGCGGGCTCCGG
der11;  CGGGGCGGGGAATCTCTCGGCTTGTGCTTGCcccatccaagcgccccgtgcggctggcgttctc
chr11;  CGGGGCGGGGAATCTCTCGGCTTGTGCTTGCTCCGCGGTGGGCTGGATGGGCAGGTAGGGGGCG
```

Figure 4.12   *Sequence alignment of translocation junctions against human reference sequence for patient t(3;11)(q21;q12).*

4.3.1.3 Patient t(7;13)(q31.3;q21.3)

In this patient there is a 2-3bp deletion of chromosome 7 sequence and a 4-5bp duplication of chromosome 13 sequence (Figure 4.13). The exact number cannot be determined due to the T at the breakpoints being present on both chromosomes.

```
t(7;13)(q31.3;q21.3)
chr7;   tagtgattcggccttgcatgctacgcctgtatttcccagtgtcgatgtgtcatttgaatccatc
der7;   tagtgattcggccttgcatgctacgcctgGCTTTGTATGAAAATGGTCCATAAGATTGGATGCT
der13;  GTGATACATAATGATTTTCTTTCAAAAGGGCTTTttcccagtgtcgatgtgtcatttgaatcca
chr13;  GTGATACATAATGATTTTCTTTCAAAAGGGCTTTGTATGAAAATGGTCCATAAGATTGGATGCT
```

Figure 4.13   ***Sequence alignment of translocation junctions against human reference sequence for patient t(7;13)(q31.3;q21.3).***

4.3.1.4 Summary of basepair sequence analysis at the breakpoints

Analysis of the derivative chromosome sequence for patients t(2;7)(q37.3;p15.1), t(3;11)(q21;q12) and t(7;13)(q31.3;q21.3) revealed that a small number of bases were deleted (1-10bp), duplicated (4-5bp) and/or inserted (19bp) in each case. This is comparable to the numbers seen in the published cases where deletions of 1-518bp, duplications of 1-18bp and insertions of 2-109bp were observed. These small basepair changes are indicative of non-homologous end joining as a mechanism for the rearrangements.

4.3.2 Analysis of repeat structures and recombination motifs at breakpoints

As described in the Introduction, certain repeat structures and recombination motifs have been associated with chromosomal rearrangements. Recurrent chromosomal rearrangements are frequently associated with repeat structures within the genome, whilst the mechanisms behind non-recurrent rearrangements remain undetermined. Analysis of the sequence surrounding the translocation breakpoints in the 3 patients studied in this thesis, and a comparison with non-recurrent translocations in the literature may help to elucidate a mechanism.

134

Repeat structures present at or around the breakpoints of published non-recurrent constitutional translocations are summarised in Table 4.3.

Translocation	Repeats in proximity	Reference
t(X;21)(p21;p12)		Bodrug et al., 1987
t(X;2)(p21;q37)	ALU 50bp from chr 2bp	Bodrug et al., 1991
t(X;4)(p21;q35)		
t(X;1)(p21;p34)		Cockburn, 1991
t(X;4)(p21.2;q31.22)		Giacalone and Francke, 1992
t(4;22)(q12;q12.2)		Arai, Ikeuchi and Nakamura, 1994
t(2;22)(q14;q11.21)	Chi like octamer close to chr22 bpt	Budarf et al., 1995
t(X;5)(p21;p31.1)		van Bakel et al., 1995
t(X;9)(p21.1;q34.3)	LINE elements on chrX	Toriello et al., 1996
t(21;22)(p12;q11)		Holmes et al., 1997
t(X;8)(p22.13;q22.1)		Ishikawa_Brush et al., 1997
t(17;22)(q11.2;q11.2)	AT rich repeats on both chrs	Kehrer-Sawatzki et al., 1997
t(6;7)(q16.2;p15.3)	chr7 breakpoint within ALU repeat	Krebs et al., 1997
t(8;17)(p11.2;p13.3)	5xALU repeats distal to chr17 breakpoint 3xL1distal to chr8 breakpoint	Kurahashi et al., 1998
t(1;10)(p22;q21)		Roberts, Chernova and Cowell, 1998
t(2;19)(q11.2;q13.3)	ALU repeat on chr19	Yoshiura et al., 1998
t(6;12)(q16.2;q21.2)		Ikegawa et al., 1999
t(1;6)(p22.1;q16.2)		Holder, Butte and Zinn, 2000
t(1;8)(q21.1;q22.1)		Matsumoto et al., 2000
t(1;11)q42.1;q14.3)		Millar et al., 2000
t(12;22)(q24.1;q13.3)		Bonaglia et al., 2001
t(1;19)(q21.3;q13.2)	chr1 bpt within AluSp chr19 bpt within AluY	Nothwang et al., 2001
t(9;11)(p24;q23)	chr9 bpt within L1 repeat chr11 bpt within ALU repeat	Willett-Brozick et al., 2001
t(7;16)(q11.23;q13)		Duba et al., 2002
t(1;8)(p34.3;q21.12)		McMullan et al., 2002
t(2;8)(q31;p21)		Spitz et al., 2002
t(2;8)(q31;p21)		Sugawara et al., 2002
t(6;13)(q21;q12)		Vervoort et al., 2002
t(7;22)(p13;q11.2)	chr22 bpt within Immunogolbulin Lambda light chain locus	Hill et al., 2003
t(6;11)(q14.2;q25)	chr6 bpt within LINE L1 element	Jeffries et al., 2003
t(4;22)(q35.1;q11.2)	554bp palindrome on chr 4 PATTR on chr 22	Nimmakayalu et al., 2003
t(X;7)(p11.3;q11.21)	LINE repeats at both bpts (L1ME on chrX, L1 on chr7)	Shoichet et al., 2003
t(1;7)(q41;p21)	chr1 bpt 200bp proximal to L1 element chr7 bpt within L2 element	David et al., 2003
t(1;22)(p21.2;q11)	AT rich region on chr1 LCR22 on chr22 Palindromic repeats on both chrs	Gotter et al., 2004
t(3;8)(p14.2;q24.2)	chr3 bpt in AT rich region	Rodriguez-Perales et al., 2004
t(2;6)(q24.3;q22.31)	SINE and LINE elements lie close to both breakpoints	Bocciardi et al., 2005
t(4;17)(q28.3;q24.3)	chr4 bpt within mariner-transposon like element (HSMAR2)	Velagaleti et al., 2005
t(1;7)(p22;q32)		Borg et al., 2005
t(4;15)(q27;q11.2)	chr4 bpt within LTR1B chr15 bpt surrounded by SINE, AluY and LINE;L1M4	Schule et al., 2005
t(4;15)(q22.3;q21.3)	polypurine and polypyrimidine tracts around both breakpoints	Klar et al., 2005
t(9;11)(q33.1;p15.3)		Tagariello et al., 2006
t(6;17)(p21.31;q11.2)	both breakpoints lie within polypyrimidine tracts	Mansouri et al., 2006
t(5;14)(q21;q32)		Haider et al., 2006
t(17;22)(q21.1;q12.1)	chr22 bpt within SINE (MIR)	Gribble et al., 2007
t(2;7)(q37.1;q36.3)	chr2 bpt 2bp away from LINE (L2) chr7 bpt within LINE (L1)	
t(11;17)(p13;p13.1)	both bpts within LINE (L1) repeats	
t(2;7)(q37.1;q21.3)		Bocciardi et al., 2007

Table 4.3 *Summary of sequence motifs found at or around the breakpoints of published reciprocal constitutional translocations.*

135

A selection of repeat structures were searched for in the 2Kb regions around the translocation breakpoints in the patients with translocations t(2;7)(q37.3;p15.1), t(3;11)(q21;q12) and t(7;13)(q31.3;q21.3) which were mapped and sequenced in Chapter 3. These structures have been discussed in detail in the Introduction, but briefly include recombination motifs such as Chi sequences, Topoisomerase I and II sites, repetitive DNA sequences such as mini satellite sequences, purine/pyrimidine tracts and AT rich sites, SINE elements including Alu repeats and MIR repeats, LINE elements, long terminal repeats, segmental duplications and DNA motifs such as palindromic sequences. 2Kb regions surrounding the translocation breakpoints for each chromosome were exported from Ensembl (http://www.ensembl.org/Homo_sapiens/index.html) and analysed for these structures. A full list of websites used to investigate the presence of sequence motifs is summarised in Appendix A1. Any structures found to be present are detailed in Sections 4.3.2.1 to 4.3.2.3.

4.3.2.1 Analysis of repeat structures around t(2;7)(q37.3;p15.1) breakpoints

A MER1B repeat was found 526bp proximal to the chromosome 2 breakpoint. The chromosome 7 breakpoint was found to be more repetitive. Two SINE elements were found; an AluSx element 375bp distal and a MIR repeat 526bp proximal to the breakpoint. A total of 4 LINE elements were found; 3 L3 elements 2bp distal and 114bp and 302bp proximal and an L2 element 8bp proximal to the breakpoint. Also found were 6.5 copies of a simple 4bp tandem repeat (AAAG) 375 bp proximal to the breakpoint. A Chi motif was observed 895 bp proximal to the chromosome 2 breakpoint with a Translin motif (ATGCAG) observed 570bp distal to the breakpoint. Translin motifs (ATGCAG) were discovered 546bp and 894 bp proximal to the chromosome 7 breakpoint. DNA bind motifs (TTTAAA) were observed 381bp proximal and 594bp distal to the chromosome 2 breakpoint and 826 and 8bp distal to the chromosome 7 breakpoint.

4.3.2.2 Analysis of repeat structures around t(3;11)(q21;q12) breakpoints

Two LINE elements were found proximal to the chromosome 3 breakpoint; an L2 repeat 900bp from the breakpoint and an L1ME4a repeat 20bp from the breakpoint. In addition, a MIRb SINE element was found 776bp distal to the breakpoint. The chromosome 11 breakpoint was found to lie distal to 2 SINE elements; a MIRb 906bp proximal and a MIR 614bp proximal to the breakpoint. In addition, 128bp of a simple repeat (CGG) was found 292bp distal to the chromosome 11 breakpoint with a further 3.8 copies of a 5bp (GCCCC) repeat and 2 copies of a 9bp repeat (GAGCTGCGC) both distal to the breakpoint by 74 and 621bp respectively. The chromosome 3 sequence was noted to contain a chi motif 279bp proximal, a translin motif (ATGCAG) 544bp distal and 2 translin motifs (GCCC[A/T][G/C][G/C][A/T]) 525bp and 945bp distal to the breakpoint. The chromosome 11 breakpoint was 880bp distal, 563bp distal and 512bp distal to 3 Chi motifs and 616bp distal, 606bp distal, 554bp distal and 794bp proximal to 4 Translin motifs (GCCC[A/T][G/C][G/C][A/T]). The chromosome 11 sequence was noted to contain palindromic sequence from 61,032,591 to 61,032,600bp with 61,032,698 to 61,032,707bp or 61,032,693 to 61,032,702bp. Both of these regions of palindromic sequence surround the breakpoint on chromosome 11. DNA bind motifs were observed 292 and 368bp distal to the chromosome 3 breakpoint and 697bp proximal to the chromosome 11 breakpoint

4.3.2.3 Analysis of repeat structures around t(7;13)(q31.3;q21.3) breakpoints

The chromosome 13 breakpoint was found to lie 130bp distal to an L1MA4 LINE repeat, 46bp distal to a 28bp AT rich low complexity repeat and 154bp proximal to a simple 16bp tandem repeat present in 2 copies. The chromosome 7 breakpoint was 114bp and 550bp proximal to 2 Translin motifs (ATGCAG). In addition, 3 immunoglobulin heptamers were observed; GATAGTG 610bp distal and CACAGTC 810bp distal and 825bp proximal to the chromosome 7 breakpoint. However, the immunoglobulin nonamer motif was not observed in close proximity to the heptamer motifs. The chromosome 13 breakpoint was

128bp proximal to a single Translin motif (ATGCAG). DNA bend motifs were discovered 898bp and 784bp proximal to the chromosome 7 breakpoint and 733, 609, 500 and 94bp proximal and 958bp distal to the chromosome 13 breakpoint.

4.3.2.4 Summary of sequence motifs found at translocation breakpoints

Whilst recombinogenic motifs were found within 1Kb of the breakpoint in all 6 regions studied, none were found to cross the exact breakpoint locations. In addition, repeat structures were observed within 1Kb of the breakpoints in all 6 regions studied, however the same repeat structures were not observed on both donor chromosomes for each patient. Whilst a comparison of the data generated from the 3 patients studied did not appear to implicate any particular repeat structure as being definitively involved in the formation of non-recurrent translocations. However, a study of 75 breakpoint sequences and 5000 control sequences has revealed a significant enrichment of MIR repeats (Kalaitzopoulos 2006).

4.4   Mechanisms underlying genomic rearrangements

Whilst recurrent translocations such as the t(11;22) are believed to be caused by illegitimate homologous recombination, the mechanism behind non-recurrent constitutional translocations has yet to be defined.

4.4.1   Translocations in patients t(2;7)(q37.3;p15.1), t(3;11)(q21;q12) and t(7;13)(q31.3;q21.3)

Analysis of the repeat structures and recombination sites at or around the breakpoints in the 3 patients and their comparison to the structures observed at the breakpoints in published translocation cases did not highlight any structure involved in all cases (as discussed in section 4.3.2.4). No regions of homology were observed at the sites of translocation between the donor chromosomes. In addition, a few bases (1-19bp) were seen to be deleted, duplicated or inserted in the derivative chromosomes of the 3 patients which were comparable to the 1-

518bp observed in the literature. These observations have led to the conclusion that the reciprocal translocations observed in the 3 patients have arisen due to non-homologous end joining.

4.4.2   Duplication in patient t(2;7)(q37.3;p15.1)

Sequence analysis of the dup(3)(p26.3p26.3) breakpoints showed that both breakpoints fell within Alu repeats at 1.7Mb and 3.6Mb along chromosome 3. Due to the high degree of homology (97%) within the Alu elements the distal breakpoint was only mapped within a 260bp region from 1,756,993 to 1,757,252bp and the proximal breakpoint was mapped within a 258bp region from 3,614,132 to 3,614,389bp. The homology observed between both breakpoint regions is indicative of a rearrangement mediated by non-allelic homologous recombination (Figure 4.14).
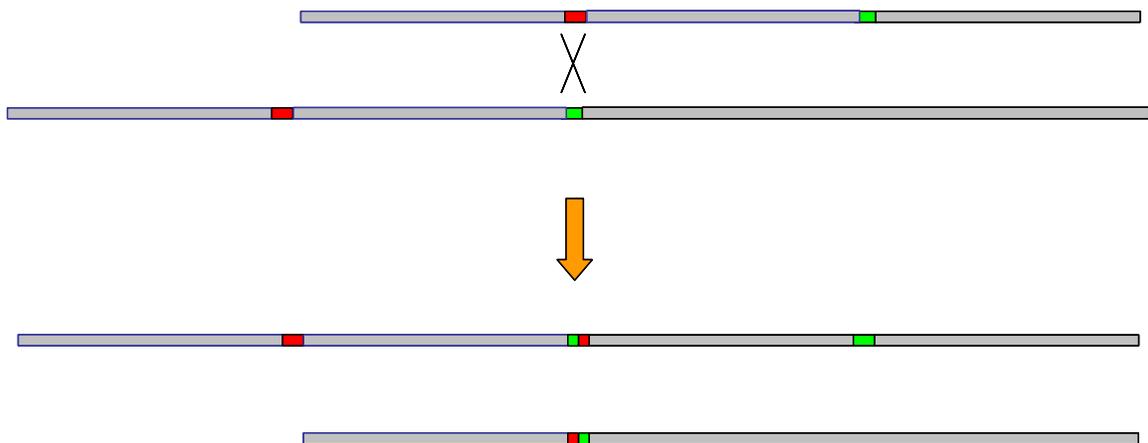
Figure 4.14   Schematic of the non-allelic homologous recombination mechanism proposed for the dup(3)(p26.3p26.3) observed in patient t(2;7)(q37.3;p15.1). Alignment of homologous segments of DNA (in this case direct Alu repeats with 97% homology coloured red and green) can lead to unequal crossing over resulting in a duplication of material on one chromosome and the reciprocal deletion on the other.

4.5 Conclusions

Analysis of the sequence across the translocation breakpoints in all 3 patients revealed that 3 of the 6 breakpoints directly disrupted a gene; CENTG2 on chromosome 2 for patient t(2;7)(q37.3;p15.1) and PTPRZ1 on chromosome 7 and DACH1 on chromosome 13 for patient t(7;13)(q31.1;q21.3), providing candidate genes for the 2 of the 3 patient phenotypes. In addition, analysis of the genes in the surrounding areas susceptible to position effect identified candidate genes for the patient analysed.

Investigation into the genomic architecture revealed that the dup(3)(p26.3p26.3) in patient t(2;7)(q37.3;p15.1) was likely to have resulted from non allelic homologous recombination mediated by 2 Alu repeats with 97% homology. This was in direct contrast to the proposed mechanism of non homologous end joining for the translocations where no repeat structures/recombination motifs were found to be common to all the breakpoints.