



HAL
open science

Telomere Healing Following DNA polymerase Arrest Induced Breakages is likely the Main Mechanism Generating Chromosome 4p Terminal Deletions

Femke Hannes, Jeroen van Houdt, Oliver W. Quarrell, Martin Poot, Ron Hochstenbach, Jean-Pierre Fryns, Joris Robert Vermeesch

► **To cite this version:**

Femke Hannes, Jeroen van Houdt, Oliver W. Quarrell, Martin Poot, Ron Hochstenbach, et al.. Telomere Healing Following DNA polymerase Arrest Induced Breakages is likely the Main Mechanism Generating Chromosome 4p Terminal Deletions. *Human Mutation*, 2010, 31 (12), pp.1343. 10.1002/humu.21368 . hal-00593433

HAL Id: hal-00593433

<https://hal.science/hal-00593433>

Submitted on 16 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Telomere Healing Following DNA polymerase Arrest
Induced Breakages is likely the Main Mechanism Generating
Chromosome 4p Terminal Deletions**

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0301.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	01-Sep-2010
Complete List of Authors:	Hannes, Femke; Catholic University of Leuven, Center for Human Genetics Van Houdt, Jeroen; Catholic University of Leuven, Center for Human Genetics Quarrell, Oliver; Sheffield children's Hospital Poot, Martin; University Medical Center Utrecht Hochstenbach, Ron; University Medical Center Utrecht Fryns, Jean-Pierre; University Hospitals Leuven, Centre of Human Genetics Vermeesch, Joris; University of Leuven, Center for Human Genetics
Key Words:	terminal deletions, telomere healing, telomerase, DNA polymerase arrest sites, Wolf-Hirschhorn syndrome, Chromosome 4p

SCHOLARONE™
Manuscripts

1
2
3 **Telomere Healing Following DNA polymerase Arrest Induced Breakages is**
4 **likely the Main Mechanism Generating Chromosome 4p Terminal Deletions**
5
6
7
8
9

10
11 Femke Hannes,¹ Jeroen Van Houdt,¹ Oliver W. Quarrell,² Martin Poot,³ Ron Hochstenbach,³
12
13 Jean-Pierre Fryns,¹ and Joris R. Vermeesch^{1*}
14
15

16
17
18 ¹*Centre for Human Genetics, University Hospital, Catholic University of Leuven, Belgium;*

19
20 ²*Sheffield Children's Hospital, Sheffield, UK;* ³*University Medical Center Utrecht, The*
21
22 *Netherlands*
23
24

25
26
27
28
29
30 *Correspondence to: Joris R. Vermeesch, Center for Human Genetics, Herestraat 49,
31
32 3000 Leuven, Belgium.
33

34 E-mail: Joris.Vermeesch@med.kuleuven.be
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT:

Constitutional developmental disorders are frequently caused by terminal chromosomal deletions. The mechanisms and/or architectural features that might underlie those chromosome breakages remain largely unexplored. Since telomeres are the vital DNA protein complexes stabilizing linear chromosomes against chromosome degradation, fusion and incomplete replication, those terminal deleted chromosomes acquired new telomeres either by telomere healing or by telomere capture. To unravel the mechanisms leading to chromosomal breakage and healing, we sequenced nine chromosome 4p terminal deletion boundaries. A computational analysis of the breakpoint flanking region, including 12 previously published pure terminal breakage sites, was performed in order to identify architectural features that might be involved in this process. All terminal 4p truncations were likely stabilized by telomerase mediated telomere healing. In the majority of breakpoints multiple genetic elements have a potential to induce secondary structures and an enrichment in replication stalling site motifs were identified. These findings suggest DNA replication stalling induced chromosome breakage during early development is the first mechanistic step leading towards terminal deletion syndromes.

KEY-WORDS: terminal deletions, telomere healing, telomerase, DNA polymerase arrest sites, Wolf-Hirschhorn syndrome, Chromosome 4p

Introduction

Terminal chromosomal deletions are the most common class of subtelomeric abnormalities and are often associated with mental retardation and multiple congenital anomalies [Ballif et al., 2007]. The most frequent terminal deletion syndromes include the 1p36 deletion syndrome (MIM 607872), the 4p terminal deletion leading to Wolf-Hirschhorn syndrome (MIM 194190), the 5p terminal deletions causing Cri-du-chat syndrome (MIM 123450), the 16p terminal deletion leading to alpha thalassaemia (MIM 141750), 9q34 deletion syndrome (MIM 610253) and the 22q terminal deletion syndrome (MIM 606232). Despite their clinical importance, it remains largely unknown when during development those terminal deletions arise, what the mechanisms are causing the chromosomal breakages leading up to those terminal deletion syndromes or which mechanisms rescue the broken chromosomes.

Recent studies show that many apparent simple terminal deletion breakpoints have proved difficult to sequence [Rooms et al., 2007; Yatsenko et al., 2009; D'Angelo et al., 2009]. In a recent study on chromosome 9q deletions, only 2 out of 14 apparently pure terminal deletions could be sequenced [Yatsenko et al., 2009]. The inability to analyze the breakpoints may relate to the complexity of the genomic sequence at the breakpoint such as inverted and tandem duplications or intervening sequence of unknown or ectopic origin [Flint et al., 1994; Varley et al., 2000; Ballif et al., 2003; Ballif et al., 2007; Rooms et al., 2007; Gajecka et al., 2007]. Another reason for the inability to identify precisely the genomic location of the telomeric breakpoint is its location within repetitive sequences [D'Angelo et al., 2009]. Overall, relatively few breakpoints have been analyzed at the sequence level and systematic studies mapping larger series of broken chromosomes are nearly non-existent. Moreover, the finding of mainly pure terminal deletions at chromosome 16pter versus more complex rearrangements at 1p, 22q and 9qter might imply that different chromosomes

1
2
3 predispose to different mechanisms leading to the chromosome breakage and subsequent
4
5 rescue.
6
7

8 Apparently simple 4p terminal deletions are the most frequent cause of WHS (55%-75%)
9
10 [Lurie et al., 1980; Shannon et al., 2001; South et al., 2008; Zollino et al., 2008]. Thus far, no
11
12 chromosome 4 terminal deletions breakpoints have been mapped at sequence level. Here, we
13
14 sequence systematically 9 consecutive terminal deletions of the short arm of chromosome 4
15
16 and explore the genomic architecture surrounding the breakpoints. In contrast to studies at
17
18 other chromosomes, we were able to map and sequence all 9 telomere junction sites. Our data
19
20 suggest that all telomeres were added *de novo* to the chromosome specific sequence.
21
22 Furthermore, analysis of the breakage sites provide evidence that specific sequence motifs
23
24 potentially induce chromosomal breaks leading up to terminal deletion syndromes.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Materials and Methods

Human subjects

Patients with developmental anomalies and mental retardation were ascertained over a period of five years of screening. Blood samples were collected after informed consent. Seven patients were seen in Leuven (BE), one in Sheffield (UK) and one in Utrecht (NE). Except for the latter patient were all patients diagnosed with Wolf-Hirschhorn syndrome. The phenotype and genotype of the latter patient have been described elsewhere [Engbers et al., 2009]. All patients, but one, had normal high resolution G-banded chromosome analysis and were ascertained through either abnormal Fluorescent in situ hybridization (FISH) results using a commercially available WHS probe (Cytocell) [Van Buggenhout et al., 2004] and/or 1 Mb BAC array following the protocol as described elsewhere [Menten et al., 2006]. Unbalanced 4p translocations as well as interstitial 4p deletion carriers were excluded.

Ultra-high resolution oligonucleotide microarray

A tiling path oligonucleotide microarray containing 385.000 probes (Roche NimbleGen systems) covering the region between 1.0 Mb to 11.8 Mb on chromosome 4p was custom designed. Both olfactory receptor gene clusters, located respectively around 3.9 Mb and 8.8 Mb, were avoided from the design. The average probe resolution was 30 bp. In addition, for one patient with the largest deletion, a HD2 whole genome oligonucleotide microarray with 2.1 million probes was used (Roche NimbleGen systems). For this array the average probe resolution was 1.5 kb.

The labeling, hybridization, washing and scanning were performed according the manufacturer's instructions (Roche NimbleGen systems). Data extraction and calculation was done via NimbleScan (Roche NimbleGen systems). The log₂ ratios were calculated following

1
2
3 the SegMNT algorithm. A 10x average window was used for breakpoint determination and
4
5 the visualization was done via SignalMap V1.9 (Roche NimbleGen systems).
6
7
8
9

10 **Quantitative PCR (qPCR)**

11 qPCR primers were designed with Universal ProbeLibrary (Roche Applied Science).
12
13 Primers were chosen free from any single nucleotide polymorphisms or repeats using
14
15 repeatmasker provided by the UCSC browser (<http://genome.ucsc.edu/>). qPCR was performed
16
17 using the lightcycler 480 instrument (Roche Applied Science) working in a total volume of 15
18
19 μ l including 7.5 μ l of mastermix, 2.5 μ l of primermix (2.5 μ M) and 50 ng template DNA. The
20
21 following amplification conditions were used: 95°C for 5 min, and 40 cycles at 95°C for 10 s
22
23 and 60°C for 20 s. After the amplification protocol, a melting curve was obtained at 95°C for
24
25 30 s, 60°C for 30 s and 95°C for a continuous mode (5 acquisitions per °C) and finally cooling
26
27 down to 40°C for 30 s. Next, data was analyzed with Excel (Microsoft) according to the
28
29 comparative ddCt method (Sequence Detection System bulletin 2 [Applied Biosystems]).
30
31
32
33
34
35
36
37
38

39 **Telomere anchored PCR and breakpoint sequencing**

40
41 A unique primer, located in the normal copy region, was chosen per breakpoint in
42
43 combination with a telomere specific primer
44
45 (5'TATGGATCCCTAACCCTGACCCTAACCC3') to amplify across the junction. A total
46
47 volume of 50 μ l contained platinum Taq polymerase (Invitrogen), 1.5 mM MgCl₂, 200 μ M of
48
49 dNTPs, 0.25 μ M forward and reverse primer (Eurogentec) and 100 ng of template DNA. The
50
51 PCR program is defined as follows: 95°C-5 min, 35 cycles of (95°C-30s, 58°C-30s, 72°C-
52
53 1.30 min) with a final extension of 72°C-10 min. After amplification, the PCR products were
54
55 analyzed on a 1% agarose gel and sequenced on an ABI 3130 automated capillary DNA
56
57 sequencer using the BigDye Terminator v. 3.1 Cycle Sequencing Kit (Applied Biosystems).
58
59
60

***In silico* analysis of breakpoint flanking sequences**

The Blat tool provided by the UCSC browser (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>) [Kent, 2002] was used to map the obtained sequences to the human genome for determining the exact location of the breakpoint. Furthermore, we collected from the literature, 14 additional pure terminal breakpoints that have been mapped at the sequence level including seven breakpoints at chromosome 16p [Flint et al., 1994; Horsley et al., 2001], one breakpoint cluster at chromosome 22q containing three terminal deletion breakpoints [Wong et al., 1997; Bonaglia et al., 2006], two at chromosome 9q34 [Yatsenko et al., 2009] and two single breakpoints at respectively chromosomes 7q [Varley et al., 2000] and 1p36 [Ballif et al., 2003]. To exclude a bias during our statistical analysis when using all three terminal breakpoints at chromosome 22q, only one breakpoint flanking region [Wong et al., 1997] was taken into account. In total, 12 terminal breakpoints were used to complete our analysis (Supplementary table 1).

All breakpoint flanking regions for further *in silico* analyses were defined by 150 bp including 75 bp flanking both sides of the breakpoint. In order to assess whether the breakpoint regions (4p and others) were enriched for certain genomic features, we randomly selected 500 sequences of 150 bp from the human genome (build 18), distributed over the different chromosomes and avoiding gaps and centromeric regions. Different online tools were used (online resources) to analyze these 150 bp random human sequences and the breakpoint sequences surrounding each junction (including 75 bp at both sides of the exact breakpoint).

RepeatMasker (<http://www.repeatmasker.org/>) (Smit, AFA, Hubley, R & Green, P.

RepeatMasker Open-3.0) was used to identify the highly repetitive sequences like long-interspersed nuclear elements (LINEs), Short-interspersed tandem nuclear elements (SINEs)

1
2
3 long terminal repeats (LTRs) and simple tandem repeats (STRs). Non-B structural elements
4 including left-handed DNA and tetraplex structures were detected by respectively ZHUNT
5
6 (<http://bioinfo.cgrb.oregonstate.edu/zDNA/>) [Champ et al., 2004] and QGRS
7
8 (<http://bioinformatics.ramapo.edu/QGRS/analyze.php>) [Kikin et al., 2006]. REPuter
9
10 (<http://bibiserv.techfak.uni-bielefeld.de/reputer/submission.html>) [Kurtz et al., 2001] allows
11
12 identifying different repeat types, we recorded direct, mirror, complementary and inverted
13
14 repeats. The presence of known motifs was analyzed with Fuzznuc (<http://inn->
15
16 temp.weizmann.ac.il/cgi-bin/emboss/fuzznuc) (EMBOSS package, [Rice et al., 2000]). The
17
18 motifs presented by Abeysinghe and co-workers were used for further analysis [Abeysinghe et
19
20 al., 2003]. Melina (<http://melina2.hgc.jp/public/index.html>) [Poluliakh et al., 2003; Okumura
21
22 et al., 2007] was used for identifying novel common motifs and Weblogo
23
24 (<http://weblogo.berkeley.edu/logo.cgi>) [Crooks et al., 2004] was used to generate sequence
25
26 logo's which are the graphical representation of a nucleotide multiple sequence alignment.
27
28 Weblogo enabled us to screen for the presence of common nucleotides at certain positions
29
30 from the exact breakpoint.
31
32
33
34
35
36
37

38 For all features we recorded the number of observations in the breakpoint sequences (4p
39
40 and others from literature) and the random human sequences (Table 1). In both sets of
41
42 sequences we counted the number of sequences with at least one hit for a given feature (Table
43
44 1). We performed a Fisher exact test in order to determine whether the occurrence of a feature
45
46 was significantly different for the breakpoint sequences and the random human sequences.
47
48
49
50
51
52

53 Results

54 Sequence analysis of terminal breakpoints

55
56
57
58 DNA of 8 patients were hybridized on specific tiling path oligonucleotide microarrays of
59
60 chromosome 4p which enabled further fine mapping of the breakpoints to an approximately

1
2
3 200-1500 bp interval (Fig. 1A). The breakpoint in one larger terminal deletion, mapped at
4
5 approximately 15 Mb according the 1 Mb BAC array, was further characterized using a whole
6
7 genome HD2 NimbleGen array. This breakpoint is mapped to an approximately 12 kb
8
9 interval. Since terminal deletions are often reported to be associated with submicroscopic
10
11 terminal duplications [Zuffardi et al., 2009] the probes flanking the contiguous stretch of
12
13 probes with reduced intensity ratios due to the terminal deletion were inspected for increased
14
15 intensity ratios. No increased intensities were identified near the breakpoints thus excluding
16
17 the presence of inverted deletion duplications. Nevertheless, in breakpoint n° 6, 7 and 8, a
18
19 group of probes around 4 Mb, more specifically between 3,854,200 bp and 4,322,200 bp,
20
21 showed an apparently normal copy number within a larger terminal deleted region (Fig. 1A).
22
23 The reason for this elevation in \log_2 values is most likely the presence of repetitive regions
24
25 flanking the olfactory receptor gene clusters which have multiple copies scattered within the
26
27 genome. Hence, a deletion of a few copies will not be detected and will appear as normal copy
28
29 number value.
30
31
32
33
34

35
36 To validate the results of the high resolution arrays and further fine map the location of
37
38 the breakpoint, a series of qPCRs were performed in the purported breakpoint region.
39
40 Quantitative PCR using different primer sets showed a contiguous stretch present as a single
41
42 allele while a flanking set of primer pairs showed the presence of two alleles and thus no
43
44 additional complex rearrangements flanking the breakpoint could be registered. The region in-
45
46 between those primer pairs with altered copy number contains the terminal breakpoint. Once
47
48 the breakpoint region was refined down to 200-800 bp, a breakpoint spanning PCR was
49
50 designed. To this end, telomere anchored PCR was performed using a telomere specific
51
52 primer combined with the reverse primer from the first primer pair with a diploid copy
53
54 number state. A positive control without a 4p deletion was performed in parallel to distinguish
55
56 the correct PCR product from the other bands as seen in the positive control. All amplicons
57
58
59
60

1
2
3 were smaller than 600 bp and subsequently sequenced (Fig. 1B). In all sequenced products, 4
5 to 27 TTAGGG hexamers were obtained (Fig. 1B). The chromosome specific sequences
6 flanking the telomeric repeats were aligned with the human genome reference sequence and
7 revealed the immediate proximity of the telomeric repeats to the chromosome specific
8 sequence in all fragments (Fig. 1C). *De novo* telomere addition to a DNA fragment is thought
9 to be mediated by telomerase. Telomerase mediated addition of telomeric repeats onto a DNA
10 fragment is stimulated by the presence of a telomeric repeat sequence to which the RNA
11 template of telomerase can bind [Muller et al., 1991; Melek and Shippen, 1996; Bottius et al.,
12 1998]. Hence, we determined whether such telomeric repeats are located at or near the
13 telomeric breakpoint. In 7/9 junctions we observed, in frame with the TTAGGG repeat,
14 microhomology of 2 to 5 bp with the RNA component of telomerase (Fig. 1C and Table 2). In
15 addition to terminal breakages at chromosome 4p16, microhomology in frame with the
16 telomere sequence was also detected in 9/12 breakpoint flanking sequences previously
17 reported.

18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36 The enzymatic part of telomerase also known as TERT contains mainly 3 functional
37 domains: the active site located at the C-terminus, the TERC binding site is located centrally,
38 and the N-terminus interacts with the ssDNA end. This N-terminal anchor site will determine
39 the efficiency by which telomerase can conduct his function. Recently, Sealy and colleagues
40 showed a ssDNA length and sequence dependency of the telomerase enzyme within a short
41 template sequence. Moreover the interaction of the N-terminus preferentially interacts with
42 oligonucleotides of minimal 13 nt in length and a G-rich character at certain positions in the
43 oligonucleotide [Sealey et al.,]. In order to characterize such a sequence specificity flanking
44 the breakpoint at the junctions that could define the interaction with telomerase, a 16 bp
45 fragment flanking each breakpoint was analyzed via Weblogo. No common nucleotides at
46 certain positions flanking the breakpoint could be detected on chromosome 4p16 or others.

In silico analysis of the breakpoint

Low copy repeats and highly repetitive elements

The mechanism(s) generating terminal chromosomal deletions remain(s) largely unknown. If the genomic architecture influences the generation of chromosomal breaks leading up to terminal deletions, specific sequences or DNA structures might be present near the deletion breakpoint both at chromosome 4p16 and others. To investigate this possibility, *in silico* analysis was conducted on the 150 bp breakpoint junction interval defined by 75 bp on either side of the breakpoint. The breakpoint flanking sequences were first analyzed for the presence of low copy repeats (LCRs) and highly repetitive elements. LCRs are mediators of non-allelic homologous recombination [Stankiewicz and Lupski, 2002], and could be involved in chromosome instability leading up to terminal deletions. No LCRs were detected in the nine 4p deletion breakpoint sequences, whereas six of the nine contained highly repetitive elements. In 3 breakpoint flanking sequences we observed long-interspersed nuclear elements (LINEs), in two other fragments short-interspersed nuclear elements (SINEs) were present and one sequence contained a simple tandem repeat (STR)(See Table 2). In light of the 12 previously reported breakpoints, in total 4 breakpoint flanking sequences contained repetitive elements including one breakpoint flanking sequence contained a SINE element, the other two contained STRs and one breakpoint flanking sequence contained a LCR.

Non-B-conformations genetic elements

Certain genetic elements in the genome can adopt non-B conformations such as triplexes, tetraplexes, cruciforms, slipped hairpin structures and left-handed Z-DNA. Those conformations can hamper DNA replication and as a consequence may result in a break [Bacolla et al., 2006; Wells, 2007]. To investigate the potential role of non-B structural

1
2
3 elements residing in the sequence surrounding the terminal breakpoint, different applications
4 such as REPuter were used to investigate the presence of direct, inverted, complementary,
5 mirror repeats and ZHUNT was used to search for left handed Z-DNA conformations. In total,
6
7
8
9
10 6/9 (67%) breakpoint flanking sequences on chromosome 4p16 contained one or more non-B
11 structural elements. Moreover, direct repeats can potentially form slipped hairpin structures
12 which in turn may trigger the rearrangement. Multiple direct repeats of minimal 8 bp were
13 found within one breakpoint flanking sequence. Mirror repeats of minimal 8 bp, potentially
14 leading to triplex structures, were detected in 2 breakpoint flanking sequences and
15 complementary repeats were identified in 3 breakpoint flanking sequences. Finally, 1
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
breakpoint flanking sequence contained an inverted repeat potentially adopting cruciform
conformations and is proposed to be responsible for genome rearrangements and/or gross
deletions [Wells, 2007]. 12 other previously reported simple terminal deletions were analyzed
in parallel with those on chromosome 4p. We observed direct, inverted and mirror repeats
from at least 8 bp in respectively 4, 3 and 2 breakpoint flanking sequences. To investigate
whether the finding of a certain element is significant or accidental, similar *in silico* analysis
was performed for randomly chosen sequences extracted from the human genome and
subsequently compared to the combined number at terminal breakpoints both at chromosome
4p and others. An overrepresentation of individual structural elements such as mirror repeats
($p=0.106$) and alternating purine/pyrimidine sequences ($p=0.118$) leading to respectively
triplex structures and left-handed Z-DNA were identified (Table 1 and Fig. 2). Although those
elements were often seen in breakpoint flanking sequences, their p-values did not reach
statistical significance ($p<0.05$) upon frequency determination.

Furthermore, the local flipping of small regions of right-handed B-DNA to left-handed Z-
DNA in the sequence requires negative super coiling making the sequence prone to breakage.
This B to Z transition is strongly favored in alternating purine/pyrimidine sequences and can

1
2
3 be *in silico* determined. ZHUNT analysis predicted at chromosome 4p in one breakpoint
4 flanking sequence n° 4 a 15 bp sequence (TGGTGCGTGCCCGTA) 21 bp away from the
5
6
7 breakpoint inducing a left-handed structure. Equally, within the group of 12 previously
8
9 reported breakpoint flanking sequences, one contained a 15 bp sequence
10
11 (AAGTGTGCACGCTCG) 56 bp proximal to the breakpoint (Fig. 2). Finally, oligo(G)_n tracts
12
13 were identified which may form tetraplex structures potentially inducing breakage. Among
14
15 nine breakpoint flanking sequences on chromosome 4p, 4 different oligo(G)_n tracts of 15 to 29
16
17 bp were identified in three breakpoint flanking sequences, whereas in the 12 other breakpoint
18
19 flanking sequences 7 contained at least one oligo(G)_n repeat with a size varying between 11
20
21 and 26 bp (Table 1 and Fig. 2).
22
23
24
25
26
27
28

29 ***Known sequence motifs***

30
31 Besides the complex architectural features and repeats inducing altered secondary
32 structures, certain sequence motifs are also known to predispose to breakage.
33
34 Overrepresentation of certain sequence motifs are found in the vicinity of the converted region
35
36 which can promote recombination not by their sequence per se but because their potential to
37
38 form non-B conformations leading to a break [Chuzhanova et al., 2009]. We evaluated the
39
40 presence of known motifs within the breakpoint flanking sequences on both the 5' to 3' and
41
42 the 3' to 5' strands. In total, 17/40 motifs were obtained from the nine breakpoint flanking
43
44 sequences. Moreover, 3 of those motifs such as the deletion hotspot consensus (TGRRKM),
45
46 the DNA polymerase arrest site (WGGAG) and the Murine parvovirus recombination hotspot
47
48 (CTWTTY) were found to be frequent (> 50%), however not necessarily simultaneously
49
50 present. Amongst the nine breakpoint flanking sequences, 5 contained such a deletion hotspot
51
52 recombination hotspot motif (Table 1). All of the three motifs are scattered throughout the
53
54 150 bp. Whereas among the 12 previously reported breakpoint flanking sequences, in 9/12
55
56
57
58
59
60

1
2
3 one or more DNA polymerase arrest site(s), in 7/12 Ig heavy chain class switch repeats 5 and
4
5 in 4/12 a Ig heavy chain class switch repeats 3 were found (Table 1). Comparison using 500
6
7 sequences generated at random, not derived from the human genome sequence, showed
8
9 statistically significant p-values for the DNA polymerase arrest site ($p=0.026$) and Ig heavy
10
11 chain class switch repeat 5 ($p=0,01$).
12
13

14
15 However the same analysis with 500 randomly generated sequences from the human
16
17 genome, showed that these particular motifs such as the DNA polymerase arrest site
18
19 ($p=0.119$) and Ig heavy chain class switch repeats 5 and 3 ($p=0.102$ and 0.08) were
20
21 overrepresented, although not statistically significant, in the group of in total 21 terminal
22
23 breakpoints including 9 on chromosome 4p and 12 other terminal breakpoints previously
24
25 reported.
26
27
28
29
30

31 *Novel sequence motifs*

32
33
34 Alternatively, the identification of novel common motifs was performed using Melina.
35
36 Among the breakpoint flanking sequences at chromosome 4p, one particular motif of 5 bp
37
38 (KGGMA) was commonly found in all breakpoint flanking sequences. This novel motif was
39
40 tested in the 12 previously reported breakpoint junctions. In 9/12 breakpoint flanking
41
42 sequences, 19 hits were obtained in both sense and antisense direction. Investigation of this
43
44 novel motif at 500 randomly generated sequences of the human genome, revealed no
45
46 statistical overrepresentation or an enrichment within the breakpoint flanking sequences.
47
48
49
50
51
52

53 **Discussion**

54
55 To unravel the mechanisms leading up to terminal deletion syndromes, we mapped and
56
57 sequenced the breakpoints of nine terminal 4p deletions. We demonstrate that all broken
58
59 chromosomes contain pure TTAGGG repeats flanking chromosome specific sequence.
60

1
2
3 Similar findings were previously reported for 12 pure terminal breakpoints at other
4
5 chromosome arms. Microhomology with the TTAGGG repeat was found in 16/21 (76.2%)
6
7 terminal breakpoints. Furthermore, we analyzed the breakpoint flanking region for the
8
9 presence of repeats, potential genomic architectural elements and sequence motifs that might
10
11 ultimately lead to chromosome breakage. Herein, we demonstrate an enrichment of DNA
12
13 polymerase arrest sites and non-B structural elements that might impede the progression of
14
15 DNA polymerase during DNA replication. Hence, our data suggest that DNA polymerase
16
17 pausing may lead to chromosomal breaks and those breaks are subsequently healed by
18
19 telomerase.
20
21
22
23
24
25
26

27 **Healing of broken chromosomes**

28
29 Broken chromosomes can acquire a telomere sequence either by capturing the telomere or
30
31 by telomerase mediated telomere healing. There are two lines of evidence that imply that all
32
33 chromosome 4p telomeres analyzed were healed by telomerase [Ballif et al., 2007]. *In vitro*
34
35 studies have demonstrated that telomerase preferably adds a segment of the telomeric
36
37 hexamer repeat unit onto a DNA fragment at its 3' end. A microhomology in frame with the
38
39 telomere repeat of 2-5 bp with the telomerase template in 7/9 (77.7%) breakpoints was
40
41 demonstrated. This result is concordant with previous published terminal breakpoints at other
42
43 chromosomes in which this type of microhomology is detected at 9/12 (75%) breakpoints
44
45 [Flint et al., 1994; Wong et al., 1997; Varley et al., 2000; Yatsenko et al., 2009]. When a
46
47 telomere is captured from another chromosome, usually not only the telomere but also
48
49 subtelomeric sequences are captured [Meltzer et al., 1993]. No such subtelomeric sequences
50
51 were detected. In addition, telomeric repeats flanking the unique euchromatic sequence have
52
53 been replicated during multiple cell divisions and thus acquire at the most proximal location
54
55 many permutations of the hexameric TTAGGG sequence such as e.g. TTGGG, TGAGGG
56
57
58
59
60

1
2
3 and TAGGG. Hence, when a telomere sequence is captured from another chromosome, those
4
5 variants of the TTAGGG sequence would be present. On the other hand, when the telomere is
6
7 acquired via telomerase only exact TTAGGG hexamers will be present. In all our junctions
8
9 fragments exact TTAGGG repeats were flanking the breakpoint junction. Hence, all 4p
10
11 broken chromosomes are likely stabilized via telomerase promoted repair and thus suggest a
12
13 telomere healing event.
14
15

16
17 Indirectly, we can deduce when during development breaks arise. First, it is known that
18
19 telomerase activity is low in mature oocytes and spermatozoa, but highly expressed from the
20
21 blastocyst stage onwards [Wright et al., 1996; Bekaert et al., 2004; Liu et al., 2007]. If a
22
23 breakage occurs in the oocyte, spermatocyte or in the zygote other mechanisms of DSB repair
24
25 will be activated rather than repair via telomerase leading possibly to other types of
26
27 rearrangements [Ballif et al., 2004]. Otherwise, when a break occurs in early development,
28
29 telomerase is present and will interfere with the DSB repair processes such as non allelic
30
31 homologous recombination (NAHR) and non homologous end joining (NHEJ). Moreover, it
32
33 has recently been shown that such a *de novo* addition via telomerase is negatively regulated by
34
35 DNA damage signaling preventing aberrant healing of broken DNA ends by telomerase. This
36
37 shows that healing via telomerase is subordinate to both non allelic homologous
38
39 recombination (NAHR) and non homologous end joining (NHEJ) [Makovets and Blackburn,
40
41 2009]. Hence, telomeric deletions are often stabilized as derivative chromosomes rather than
42
43 pure deletions [South et al., 2008]. Secondly, our results show an extensive microhomology at
44
45 the terminal breakpoints, suggesting an important role in telomerase dependent repair. Finally,
46
47 it was recently shown that genomes of early cleavage stage embryos are prone to chromosome
48
49 instability and that a high number of blastomeres carry chromosomal deletions [Vanneste et
50
51 al., 2009]. Since chromosome instability is high in early cleavage stage embryos and since
52
53 telomerase is most active during early development, it seems likely that terminal deletions
54
55
56
57
58
59
60

1
2
3 arise during early human development. This implies that the origin of generation and
4
5 stabilization occurs during mitosis rather than meiosis and thus, it seems likely that the broken
6
7 4p chromosomes are generated and stabilized during early embryonic development.
8
9

10 Besides that approximately 76% of the terminal deletion breakpoints contained
11
12 microhomology with the RNA template of telomerase, a small proportion of the breakpoints
13
14 did not contain microhomology. It is known that binding and elongation of the telomerase
15
16 enzyme involves base pairing with the telomerase RNA template and the substrate [Morin,
17
18 1991; Harrington and Greider, 1991]. In vitro studies have recently shown that binding of
19
20 telomerase to a substrate also depends on the G-rich character at certain positions in the
21
22 substrate [Sealey et al., 2010]. According to our data, no obvious G-rich position nearby the
23
24 breakpoint could be defined in a 16 bp sequence fragment flanking the terminal breakpoints
25
26 without any microhomology. This suggests that *in vivo* telomerase can elongate broken
27
28 chromosomes without having a complement to the RNA template nor a certain enrichment for
29
30 a nucleotide near the healing site.
31
32
33
34
35
36
37
38

39 **Mechanism of chromosomal breakage**

40
41 Sequences such as repeats, non-B structural elements or certain motifs can predispose to
42
43 DNA breakage [Vilenchik and Knudson, 2003; Abeysinghe et al., 2003; Wang and Vasquez,
44
45 2006; Chuzhanova et al., 2009; Branzei and Foiani, 2010]. Here, the enrichment of particular
46
47 motifs was observed at terminal breakpoints. A DNA polymerase arrest site was identified to
48
49 impede the progress of DNA polymerase alpha at phiX174 DNA resulting in stalling of the
50
51 replication fork [[Weaver and DePamphilis, 1982]. Replication arrest actually stimulates
52
53 slippage and can trigger recombination and thus the finding of a replication arrest motif can be
54
55 a key factor in genome instability [Stary and Sarasin, 1992; Hyrien, 2000; Michel et al.,
56
57 2001]. In E.coli a mechanistic link between a replication arrest site and homologous
58
59
60

1
2
3 recombination has been established [Seigneur et al., 1998]. In about 50% of the breakpoint
4 flanking sequences such an arrest site was present. Likewise, an overrepresentation of DNA
5 polymerase arrest motifs ($p=0.01$) are found at sequences flanking the gene conversion region
6 [Chuzhanova et al., 2009]. Furthermore, both Ig heavy chain switch class repeats 3 and 5 were
7 enriched at our terminal breakpoints. These repeats mediate a process of nonhomologous
8 recombination that introduces variation in the constant region of an immunoglobulin heavy
9 chain gene. Those switch repeats have been noted at translocation breakpoints associated with
10 a number of hematological malignancies [Fenton et al., 2002]. Moreover, repeat 5 was also
11 shown to be significantly overrepresented within close proximity at translocation breakpoints
12 ($p=0.05$) [Abeysinghe et al., 2003].
13
14
15
16
17
18
19
20
21
22
23
24
25
26

27 Besides the specific motifs, repetitive sequences inducing altered DNA conformations are
28 frequent within the surroundings of the rearrangement breakpoints [Bacolla et al., 2006]. Such
29 elements can elevate the rate of single-strand lesions, and hence contribute to their conversion
30 to double-strand breaks [Vilenchik and Knudson, 2003; Patel et al., 2004]. Within the
31 terminal breakpoint flanking sequences of both chromosome 4p and others, direct, mirror,
32 inverted and complementary repeats were found even as puridine/pyrimidine and oligo(G)_n
33 tracts. Interestingly, the fragments were likely to induce more left-handed Z-DNA ($p=0.118$)
34 and triplex ($p=0.106$) structures than what we would expect from the random genome. Other
35 studies with similar findings conclude that both structures are significantly associated with the
36 site of DNA breakage at translocation junctions [Abeysinghe et al., 2003; Bacolla et al.,
37 2006].
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 In addition, the presence of repetitive elements including high and low copy repeats
54 might play a role in stimulating the formation of non-recurrent breakpoints [Gu et al., 2008].
55 No low copy repeats were identified at or near the breakpoints whereas highly repetitive
56 elements were abundant. 6/9 breakpoints are embedded within a repetitive element most often
57
58
59
60

1
2
3 involving interspersed sequences (Table 2). In addition, 4/12 pure terminal breakpoints
4
5 extracted from the literature are within segmental duplications, simple tandem repeats or
6
7 SINEs (Table 1). Overall, the repetitive content of the breakpoint flanking sequences is
8
9 heterogeneous and the fragments do not *per se* contain LCR nor repetitive elements. Hence, in
10
11 our dataset, 50% of the breakpoints associate with LCR or highly repetitive sequences which
12
13 are comparable with that of the human genome and thus there is not strong evidence of the
14
15 contribution of those elements in the generation of breakages.
16
17
18

19
20 The relatively high frequency of DNA polymerase arrest sites at the breakpoint flanking
21
22 sequences suggests one plausible mechanism of DNA breakage following DNA replication
23
24 stalling. The presence of other structures rendering the DNA prone to DNA breakage might be
25
26 witnesses from other breakage mechanisms. It is well known that dicentric chromosomes can
27
28 be formed (either during meiosis or during mitosis). Following mitosis, the dicentric
29
30 chromosome would break leading to two daughter cells having either a deleted 4p or an inv
31
32 dup del (4p) and it can be expected that the breaks would occur at the “weakest” loci [Zuffardi
33
34 et al., 2009]. No inv del dupl (4p) were detected during our study. Another mechanism that
35
36 might induce breakages is homologous recombination intermediates. Larger groups of
37
38 terminal deletions need to be sequenced to determine the involvement of specific motifs in the
39
40 breakage process and to discriminate whether the occurrence of different motifs might be
41
42 relics of different mechanisms.
43
44
45
46
47
48
49

50 **Acknowledgments**

51
52 We thank all families who participated in the study. This work was made possible in part by
53
54 grants GOA/2006/12, and the SymbioSys Center of Excellence (Research Council,
55
56 K.U.Leuven, EF/05/007) to J.R.V.
57
58
59
60

References

- 1
2
3
4
5
6
7
8
9
10 Abeyasinghe SS, Chuzhanova N, Krawczak M, Ball EV, Cooper DN. 2003. Translocation and
11 gross deletion breakpoints in human inherited disease and cancer I: Nucleotide
12 composition and recombination-associated motifs. *Hum Mutat* 22:229-244.
13
14
15
16
17 Bacolla A, Wojciechowska M, Kosmider B, Larson JE, Wells RD. 2006. The involvement of
18 non-B DNA structures in gross chromosomal rearrangements. *DNA Repair (Amst)* 5:1161-
19 1170.
20
21
22
23
24 Ballif BC, Sulpizio SG, Lloyd RM, Minier SL, Theisen A, Bejjani BA, Shaffer LG. 2007. The
25 clinical utility of enhanced subtelomeric coverage in array CGH. *Am J Med Genet A*
26 143A:1850-1857.
27
28
29
30
31 Ballif BC, Wakui K, Gajecka M, Shaffer LG. 2004. Translocation breakpoint mapping and
32 sequence analysis in three monosomy 1p36 subjects with der(1)t(1;1)(p36;q44) suggest
33 mechanisms for telomere capture in stabilizing de novo terminal rearrangements. *Hum*
34 *Genet* 114:198-206.
35
36
37
38
39
40
41 Ballif BC, Yu W, Shaw CA, Kashork CD, Shaffer LG. 2003. Monosomy 1p36 breakpoint
42 junctions suggest pre-meiotic breakage-fusion-bridge cycles are involved in generating
43 terminal deletions. *Hum Mol Genet* 12:2153-2165.
44
45
46
47
48 Bekaert S, Derradji H, Baatout S. 2004. Telomere biology in mammalian germ cells and
49 during development. *Dev Biol* 274:15-30.
50
51
52
53 Bonaglia MC, Giorda R, Mani E, Aceti G, Anderlid BM, Baroncini A, Pramparo T, Zuffardi
54 O. 2006. Identification of a recurrent breakpoint within the SHANK3 gene in the 22q13.3
55 deletion syndrome. *J Med Genet* 43:822-828.
56
57
58
59
60

- 1
2
3 Bottius E, Bakhsis N, Scherf A. 1998. Plasmodium falciparum telomerase: de novo telomere
4 addition to telomeric and nontelomeric sequences and role in chromosome healing. Mol
5 Cell Biol 18:919-925.
6
7
8
9
10 Branzei D, Foiani M. 2010. Maintaining genome stability at the replication fork. Nat Rev Mol
11 Cell Biol 11:208-219.
12
13
14
15 Champ PC, Maurice S, Vargason JM, Camp T, Ho PS. 2004. Distributions of Z-DNA and
16 nuclear factor I in human chromosome 22: a model for coupled transcriptional regulation.
17 Nucleic Acids Res 32:6501-6510.
18
19
20
21
22 Chuzhanova N, Chen JM, Bacolla A, Patrinos GP, Ferec C, Wells RD, Cooper DN. 2009.
23 Gene conversion causing human inherited disease: evidence for involvement of non-B-
24 DNA-forming sequences and recombination-promoting motifs in DNA breakage and
25 repair. Hum Mutat 30:1189-1198.
26
27
28
29
30
31
32 Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator.
33 Genome Res 14:1188-1190.
34
35
36
37 D'Angelo CS, Gajecka M, Kim CA, Gentles AJ, Glotzbach CD, Shaffer LG, Koiffmann CP.
38 2009. Further delineation of nonhomologous-based recombination and evidence for
39 subtelomeric segmental duplications in 1p36 rearrangements. Hum Genet 125:551-563.
40
41
42
43
44 Engbers H, van der Smagt JJ, van 't SR, Vermeesch JR, Hochstenbach R, Poot M. 2009.
45 Wolf-Hirschhorn syndrome facial dysmorphic features in a patient with a terminal 4p16.3
46 deletion telomeric to the WHSCR and WHSCR 2 regions. Eur J Hum Genet 17:129-132.
47
48
49
50
51 Fenton JA, Vaandrager JW, Aarts WM, Bende RJ, Heering K, van Dijk M, Morgan G, van
52 Noesel CJ, Schuurin E, Kluin PM. 2002. Follicular lymphoma with a novel t(14;18)
53 breakpoint involving the immunoglobulin heavy chain switch mu region indicates an origin
54 from germinal center B cells. Blood 99:716-718.
55
56
57
58
59
60

- 1
2
3 Flint J, Craddock CF, Villegas A, Bentley DP, Williams HJ, Galanello R, Cao A, Wood WG,
4
5 Ayyub H, Higgs DR. 1994. Healing of broken human chromosomes by the addition of
6
7 telomeric repeats. *Am J Hum Genet* 55:505-512.
8
9
10 Gajecka M, Mackay KL, Shaffer LG. 2007. Monosomy 1p36 deletion syndrome. *Am J Med*
11
12 *Genet C Semin Med Genet* 145C:346-356.
13
14
15 Gu W, Zhang F, Lupski JR. 2008. Mechanisms for human genomic rearrangements.
16
17 *Pathogenetics* 1:4.
18
19
20 Harrington LA, Greider CW. 1991. Telomerase primer specificity and chromosome healing.
21
22 *Nature* 353:451-454.
23
24
25 Horsley SW, Daniels RJ, Anguita E, Raynham HA, Peden JF, Villegas A, Vickers MA, Green
26
27 S, Wayne JS, Chui DH, Ayyub H, MacCarthy AB, Buckle VJ, Gibbons RJ, Kearney L,
28
29 Higgs DR. 2001. Monosomy for the most telomeric, gene-rich region of the short arm of
30
31 human chromosome 16 causes minimal phenotypic effects. *Eur J Hum Genet* 9:217-225.
32
33
34 Hyrien O. 2000. Mechanisms and consequences of replication fork arrest. *Biochimie* 82:5-17.
35
36
37 Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome Res* 12:656-664.
38
39
40 Kikin O, D'Antonio L, Bagga PS. 2006. QGRS Mapper: a web-based server for predicting G-
41
42 quadruplexes in nucleotide sequences. *Nucleic Acids Res* 34:676-682.
43
44
45 Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001.
46
47 REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids*
48
49 *Res* 29:4633-4642.
50
51
52 Liu L, Bailey SM, Okuka M, Munoz P, Li C, Zhou L, Wu C, Czerwiec E, Sandler L, Seyfang
53
54 A, Blasco MA, Keefe DL. 2007. Telomere lengthening early in development. *Nat Cell Biol*
55
56 9:1436-1441.
57
58
59 Lurie IW, Lazjuk GI, Ussova YI, Presman EB, Gurevich DB. 1980. The Wolf-Hirschhorn
60
syndrome. I. Genetics. *Clin Genet* 17:375-384.

- 1
2
3 Makovets S, Blackburn EH. 2009. DNA damage signalling prevents deleterious telomere
4 addition at DNA breaks. *Nat Cell Biol* 11:1383-1386.
5
6
7
8 Melek M, Shippen DE. 1996. Chromosome healing: spontaneous and programmed de novo
9 telomere formation by telomerase. *Bioessays* 18:301-308.
10
11
12 Meltzer PS, Guan XY, Trent JM. 1993. Telomere capture stabilizes chromosome breakage.
13 *Nat Genet* 4:252-255.
14
15
16
17 Menten B, Maas N, Thienpont B, Buysse K, Vandesompele J, Melotte C, de Ravel T, Van
18 Vooren S, Balikova I, Backx L, Janssens S, De Paepe A, De Moor B, Moreau Y, Marynen
19 P, Fryns JP, Mortier G, Devriendt K, Speleman F, Vermeesch JR. 2006. Emerging patterns
20 of cryptic chromosomal imbalance in patients with idiopathic mental retardation and
21 multiple congenital anomalies: a new series of 140 patients and review of published
22 reports. *J Med Genet* 43:625-633.
23
24
25
26
27
28
29
30
31 Michel B, Flores MJ, Viguera E, Grompone G, Seigneur M, Bidnenko V. 2001. Rescue of
32 arrested replication forks by homologous recombination. *Proc Natl Acad Sci U S A*
33 98:8181-8188.
34
35
36
37
38
39 Morin GB. 1991. Recognition of a chromosome truncation site associated with alpha-
40 thalassaemia by human telomerase. *Nature* 353:454-456.
41
42
43
44 Muller F, Wicky C, Spicher A, Tobler H. 1991. New telomere formation after
45 developmentally regulated chromosomal breakage during the process of chromatin
46 diminution in *Ascaris lumbricoides*. *Cell* 67:815-822.
47
48
49
50
51 Okumura T, Makiguchi H, Makita Y, Yamashita R, Nakai K. 2007. Melina II: a web tool for
52 comparisons among several predictive algorithms to find potential motifs from promoter
53 regions. *Nucleic Acids Res* 35:227-231.
54
55
56
57
58 Patel HP, Lu L, Blaszak RT, Bissler JJ. 2004. PKD1 intron 21: triplex DNA formation and
59 effect on replication. *Nucleic Acids Res* 32:1460-1468.
60

- 1
2
3 Poluliakh N, Takagi T, Nakai K. 2003. Melina: motif extraction from promoter regions of
4
5 potentially co-regulated genes. *Bioinformatics* 19:423-424.
6
7
8 Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open
9
10 Software Suite. *Trends Genet* 16:276-277.
11
12
13 Rooms L, Reyniers E, Kooy RF. 2007. Diverse chromosome breakage mechanisms underlie
14
15 subtelomeric rearrangements, a common cause of mental retardation. *Hum Mutat* 28:177-
16
17 182.
18
19
20 Sealey DC, Zheng L, Taboski MA, Cruickshank J, Ikura M, Harrington LA. The N-terminus
21
22 of hTERT contains a DNA-binding domain and is required for telomerase activity and
23
24 cellular immortalization. *Nucleic Acids Res* 38:2019-35.
25
26
27 Sealey DC, Zheng L, Taboski MA, Cruickshank J, Ikura M, Harrington LA. 2010. The N-
28
29 terminus of hTERT contains a DNA-binding domain and is required for telomerase activity
30
31 and cellular immortalization. *Nucleic Acids Res* 38:2019-2035.
32
33
34 Seigneur M, Bidnenko V, Ehrlich SD, Michel B. 1998. RuvAB acts at arrested replication
35
36 forks. *Cell* 95:419-430.
37
38
39 Shannon NL, Maltby EL, Rigby AS, Quarrell OW. 2001. An epidemiological study of Wolf-
40
41 Hirschhorn syndrome: life expectancy and cause of mortality. *J Med Genet* 38:674-679.
42
43
44 South ST, Whitby H, Battaglia A, Carey JC, Brothman AR. 2008. Comprehensive analysis of
45
46 Wolf-Hirschhorn syndrome using array CGH indicates a high prevalence of translocations.
47
48 *Eur J Hum Genet* 16:45-52.
49
50
51 Stankiewicz P, Lupski JR. 2002. Genome architecture, rearrangements and genomic
52
53 disorders. *Trends Genet* 18:74-82.
54
55
56 Stary A, Sarasin A. 1992. Molecular analysis of DNA junctions produced by illegitimate
57
58 recombination in human cells. *Nucleic Acids Res* 20:4269-4274.
59
60

- 1
2
3 Van Buggenhout G, Melotte C, Dutta B, Froyen G, Van Hummelen P, Marynen P, Matthijs
4
5 G, de Ravel T, Devriendt K, Fryns JP, Vermeesch JR. 2004. Mild Wolf-Hirschhorn
6
7 syndrome: micro-array CGH analysis of atypical 4p16.3 deletions enables refinement of
8
9 the genotype-phenotype map. *J Med Genet* 41:691-698.
- 10
11
12 Vanneste E, Voet T, Le Caignec C, Ampe M, Konings P, Melotte C, Debrock S, Amyere M,
13
14 Vikkula M, Schuit F, Fryns JP, Verbeke G, D'Hooghe T, Moreau Y, Vermeesch JR. 2009.
15
16 Chromosome instability is common in human cleavage-stage embryos. *Nat Med* 15:577-
17
18 583.
- 19
20
21
22 Varley H, Di S, Scherer SW, Royle NJ. 2000. Characterization of terminal deletions at 7q32
23
24 and 22q13.3 healed by De novo telomere addition. *Am J Hum Genet* 67:610-622.
- 25
26
27 Vilenchik MM, Knudson AG. 2003. Endogenous DNA double-strand breaks: production,
28
29 fidelity of repair, and induction of cancer. *Proc Natl Acad Sci U S A* 100:12871-12876.
- 30
31
32 Wang G, Vasquez KM. 2006. Non-B DNA structure-induced genetic instability. *Mutat Res*
33
34 598:103-119.
- 35
36
37 Weaver DT, DePamphilis ML. 1982. Specific sequences in native DNA that arrest synthesis
38
39 by DNA polymerase alpha. *J Biol Chem* 257:2075-2086.
- 40
41
42 Wells RD. 2007. Non-B DNA conformations, mutagenesis and disease. *Trends Biochem Sci*
43
44 32:271-8.
- 45
46
47 Wong AC, Ning Y, Flint J, Clark K, Dumanski JP, Ledbetter DH, McDermid HE. 1997.
48
49 Molecular characterization of a 130-kb terminal microdeletion at 22q in a child with mild
50
51 mental retardation. *Am J Hum Genet* 60:113-20.
- 52
53
54 Wright WE, Piatyszek MA, Rainey WE, Byrd W, Shay JW. 1996. Telomerase activity in
55
56 human germline and embryonic tissues and cells. *Dev Genet* 18:173-179.
- 57
58
59
60

1
2
3 Yatsenko SA, Brundage EK, Roney EK, Cheung SW, Chinault AC, Lupski JR. 2009.

4
5 Molecular mechanisms for subtelomeric rearrangements associated with the 9q34.3
6
7
8 microdeletion syndrome. *Hum Mol Genet* 18:1924-1936.

9
10 Zollino M, Murdolo M, Marangi G, Pecile V, Galasso C, Mazzanti L, Neri G. 2008. On the
11
12 nosology and pathogenesis of Wolf-Hirschhorn syndrome: genotype-phenotype correlation
13
14 analysis of 80 patients and literature review. *Am J Med Genet C Semin Med Genet*
15
16
17 148C:257-269.

18
19 Zuffardi O, Bonaglia M, Ciccone R, Giorda R. 2009. Inverted duplications deletions:
20
21
22 underdiagnosed rearrangements?? *Clin Genet* 75:505-513.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Legend

Figure 1. Overview of 9 terminal breakpoints at chromosome 4p16. A) Results of a targeted oligonucleotide microarray (Nimblegen) starting from 1.0 Mb to 11.8 Mb. Y-axis represents the log₂ values of the intensity ratios and the X-axis, the position of the oligonucleotides on chromosome 4p. Breakpoint n°9 is analyzed with a HD2 oligo array (Nimblegen) of which only a part of chromosome 4p is visualized. B) The junction sequence. The red arrow indicates the start of the telomere repeat sequence. The green arrow refers to a TTAGGG repeat. C) Global alignment of the junctions and indication of the microhomology with TTAGGG sequence in yellow. Upper line refers to the reference sequence, the second line is patient derived sequence at the breakpoint, and third line is the telomere sequence.

Figure 2. Overview of all the elements potentially leading towards a non-B conformation or specific motifs that hamper DNA replication or facilitate recombination. Both groups of terminal deletions are represented including aberrations at chromosome 4p (left) and 12 other chromosomal terminal deletions previously reported in the literature (supplementary table 1) (right). All sequences have the same orientation starting from the telomere towards the centromere.

Table 1. Overview of the non-B structural elements and motifs tested at chromosome 4p and others

		4p-			Other chr			Combined			Human sequences			Random sequences		p-value
		n° seq	all hits	%	n° seq	all hits	%	n° seq	all hits	%	n° seq	all hits	%	n° seq	all hits	COM/ HUMAN
		n=9	n=9		n=12	n=12		n=21	n=21		n=500	n=500		n=500	n=500	
MICROHOMOLOGY		7	-	77.8	9	-	75.0	16	-	76.2	-					
NON B-STRUCTURES	zDNA	1	1	11.1	1	1	8.3	2	2	9.5	12	14			0.12	
	Oligo(G)n repeats	3	4	33.3	7	14	58.3	10	18	47.6	181	220	36.2		0.50	
	Direct repeats	1	3	11.1	4	19	33.3	5	22	23.8	6	6			1.00	
	Mirror repeat	2	3	22.2	2	4	16.7	4	7	19.0	0	0			0.11	
	Complementary repeats	3	3	33.3	0	0		3	3	14.3	2	3			1.00	
	Inverted repeats	1	1	11.1	3	3	25.0	4	4	19.0	2	2			0.66	
MOTIFS		n=9			n=12			n=21			n=500	n=500		n=500	n=500	
DNA polymerase arrest site	WGGAG	6	10	66.7	9	20	75.0	15	30	71.4	266	450	53.2	231	322	0.12
deletion hotspot consensus	TGRRKM	5	11	55.6	9	18	75.0	14	29	66.7	374	688	74.8	341	581	0.44
Murine parvovirus recombination hotspot	CTWTTY	5	5	55.6	3	3	25.0	8	8	38.1	204	298	40.8	127	140	1.00
Ig Heavy chain class switch repeat 5	TGAGC	4	5	44.4	7	10	58.3	11	15	52.4	171	208	34.2	127	144	0.10
DNA polymerase a/b frameshift hotspot 2	TGGNGT	3	3	33.3	4	5	33.3	7	8	33.3	137	168	27.4	138	154	0.62
Ig Heavy chain class switch repeat 3	GGGGT	3	3	33.3	4	16	33.3	7	19	33.3	90	106	18	145	165	0.09
Ig Heavy chain class switch repeat 4	TGGGG	3	5	33.3	6	11	50.0	9	16	42.9	135	179	27	138	156	0.13
translin binding site 2	GCCCWSSW	2	2	22.2	2	2	16.7	4	4	19.0	56	67	11.2	28	30	0.29
Ig Heavy chain class switch repeat 1	GAGCT	2	2	22.2	3	7	25.0	5	9	23.8	124	160	24.8	120	157	1.00

Ig Heavy chain class switch repeat 2	GGGCT	2	2	22.2	4	7	33.3	6	9	28.6	130	171	26	123	138	0.80
consensus scaffold attachment region 3	WADAWAYAWW	2	2	22.2	0	0	0.0	2	2	9.5	105	151	21	8	8	0.27
consensus scaffold attachment region 4	TWWTDTTWWW	2	2	22.2	1	11	8.3	3	13	14.3	113	289	22.6	11	14	0.59
DNA polymerase a frameshift hotspot 2	CTGGCG	1	1	11.1	0	0	0.0	1	1	4.8	6	6	1.2	32	34	0.25
DNA polymerase b frameshift hotspot 1	ACCCWR	1	1	11.1	6	9	50.0	7	10	33.3	149	168	29.8	125	142	0.81
DNA polymerase a/b frameshift hotspot 1	ACCCCA	1	1	11.1	1	1	8.3	2	2	9.5	40	43	8	39	42	0.68
Human minisatellites conserved seq/X-like element	GCWGGWGG	1	1	11.1	1	1	8.3	2	2	9.5	18	19	3.6	9	9	0.19
Vaccinia Topoisomerase I consensus	YCCTT	1	1	11.1	7	10	58.3	8	11	38.1	246	364	49.2	244	336	0.38
DNA polymerase a frameshift hotspot 1	TCCCC	0	0	0	3	4	25.0	3	4	14.3	30	32	6	41	43	0.14
translin binding site 1	ATGCAG	0	0	0	1	2	8.3	1	2	4.8	51	52	10.2	38	38	0.71
X-element E.coli	GCTGGTGG	0	0	0	0	0	0	0	0	0	4	4	0.8	2	2	
Ade6-M26	ATGACGT	0	0	0	0	0	0	0	0	0	1	1	0.2	13	14	
ARS consensus S cerevisiae	WTTTATRTTTW	0	0	0	0	0	0	0	0	0	1	1	0.2	1	1	
ARS consensus S Pombe	WRTTATTTAW	0	0	0	0	0	0	0	0	0	9	19	1.8	2	2	
consensus scaffold attachment region 1	AATAAYAAA	0	0	0	0	0	0	0	0	0	8	17	1.6	0	0	
consensus scaffold attachment region 2	TTWTWTTWTT	0	0	0	1	8	8.3	1	8	4.8	40	144	8	0	0	1.00
Drosophila topoisomerase 2 consensus	GTNWAYATTVATNNR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Heptamer recombination signal	CACAGTG	0	0	0	2	2	16.7	2	2	9.5	19	20	3.8	7	7	0.21

Human hypervariable minisatellites seq 1	GGAGGTGGGCAGGARG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Human hypervariable minisatellites seq 2	AGAGGTGGGCAGGTGG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Human minisatellites core sequence	GGGCAGGARG	0	0	0	0	0	0	0	0	0	2	2	0.4	0	0		
Human replication origin consensus	WAWTTDDWWWDHWGWHM AWTT	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
LT-IS motif	TGGAAATCCCC	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Mariner transposon-like element	GAAAATGAAGCTATTTACCCA GGA	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Murine MHC recombination hotspot	GAGRCAGR	0	0	0	2	2	16.7	2	2	9.5	56	67	11.2	9	9	1.00	
Nonamer recombination signal	ACAAAAACC	0	0	0	0	0	0	0	0	0	2	2	0.4	0	0		
Pur binding site	GGNNGAGGGAGARRRR	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Recombination hotspot	CCNCCNTNNCCNC	0	0	0	1	3	8.3	1	3	4.8	4	4	0.8	3	3	0.19	
Retrotransposon	TCATACACCACGCAGGGGTA GAGGACT	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Vertebrate topoisomerase II consensus	RNYNNXNNGYNGKTNYNY	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
XY32 homopurine-pyrimidine H-palindrome motif	AAGGGAGAARGGGTATAGGG RAAGAGGGAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0		

Table 2. Overview of terminal breakpoints at chromosome 4p

N°	Genbank ID	Breakpoint (bp)	Repeat at breakpoint	Micro-homology with TTAGGG (bp)
1	HQ167740	1.691.685	No	2
2	HQ167741	2.161.732	LINE	2
3	HQ167742	2.168.865	LINE	0
4	HQ167743	2.424.502	SINE	1
5	HQ167744	2.615.988	No	3
6	HQ167745	4.887.950	LINE	3
7	HQ167746	5.581.455	STR	2
8	HQ167747	5.981.420	No	5
9	HQ167748	16.696.986	SINE	3
			6/9	7/9

STR: simple tandem repeats, LINE: long interspersed nuclear elements, SINE: short interspersed nuclear elements.

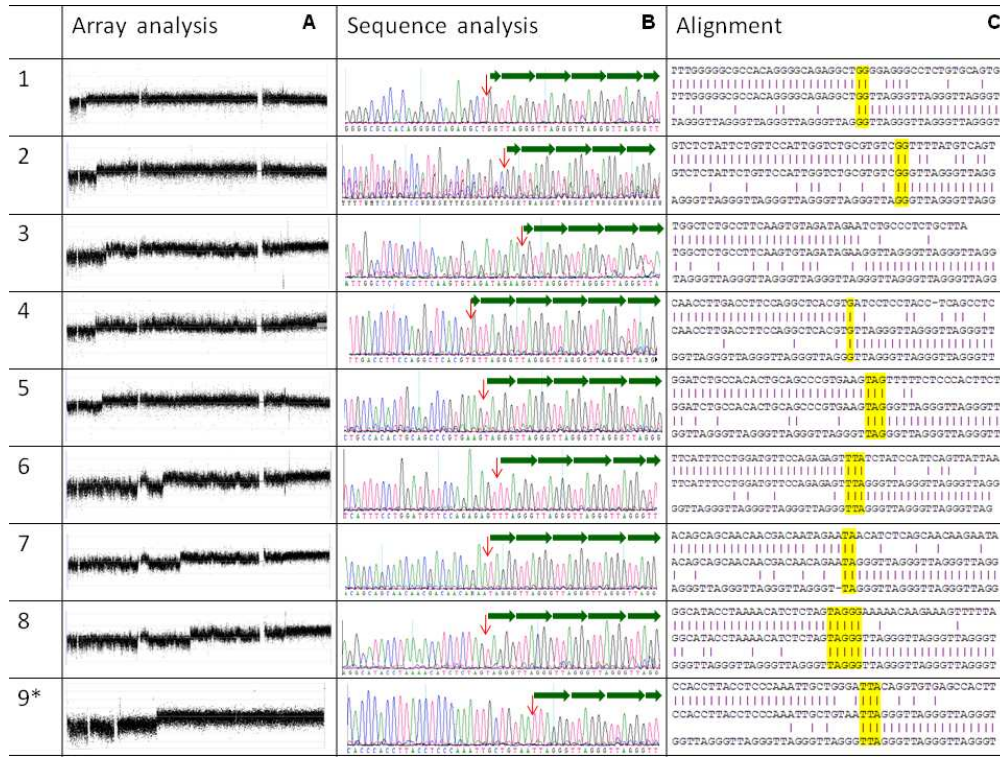


Figure 1
254x190mm (96 x 96 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

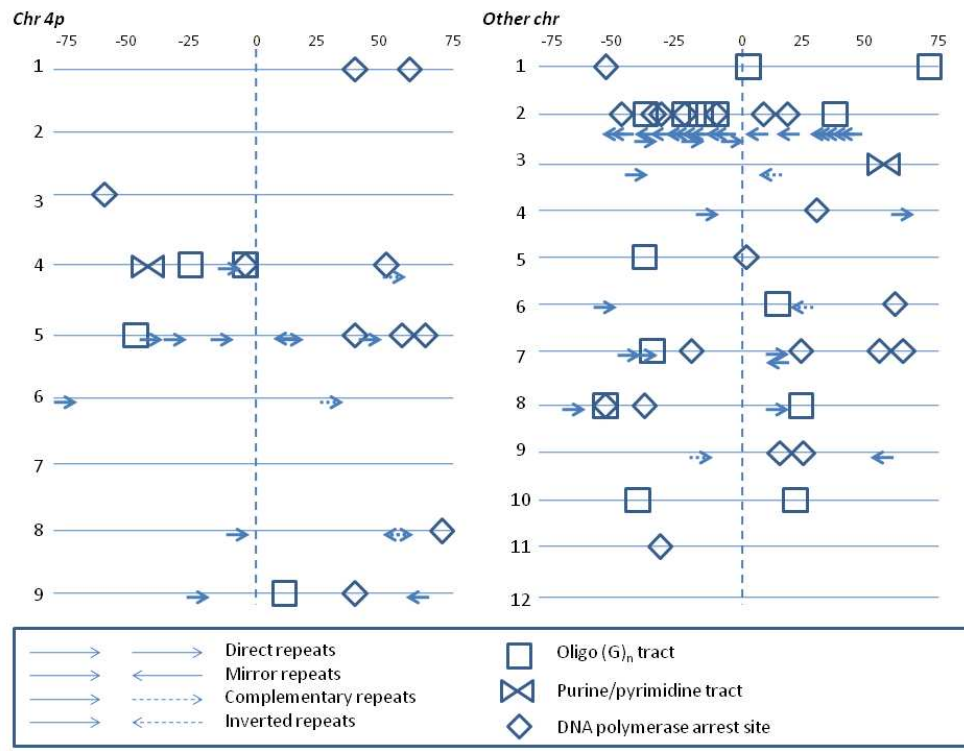


Figure 2
254x190mm (96 x 96 DPI)

Supplementary table 1

N°	Chromosome	Breakpoint (bp)	Repeat at breakpoint	Micro-homology with TTAGGG (bp)	Reference	Patient initials
1	7q32	131.449.009	No	1	[5]	
2	22q13.3	49.469.068	STR	0	[6]	
3	16p	1.886.763	No	4	[3]	BO
4	16p	147.681	No	3	[3]	CMO
5	16p	144.446	No	3	[3]	IDF
6	16p	142.819	SD	4	[3]	TAT
7	16p	112.851	STR	3	[3]	TI
8	1p36	4.788.530	No	2	[1]	
9	9q34	137.329.926	SINE	5	[7]	P40
10	9q34	139.129.925	No	2	[7]	P3
11	16p	210.918	No	1	[4]	HW
12	16p	212.156	No	5	[4]	BR
*13	22q13.3	49.469.074	STR	(0)	[2]	Subj 1
*14	22q13.3	49.469.086	STR	(3)	[2]	Subj 2

*Not included in the analysis, because of similar locus as described previously by Wong and colleagues [6].

References in the table

- [1] Ballif B. C., Yu W., Shaw C. A., Kashork C. D. and Shaffer L. G. Monosomy 1p36 breakpoint junctions suggest pre-meiotic breakage-fusion-bridge cycles are involved in generating terminal deletions. *Hum Mol Genet.* 2003;12(17):2153-65
- [2] Bonaglia M. C., Giorda R., Mani E., Aceti G., Anderlid B. M., Baroncini A., Pramparo T. and Zuffardi O. Identification of a recurrent breakpoint within the SHANK3 gene in the 22q13.3 deletion syndrome. *J Med Genet.* 2006;43(10):822-8
- [3] Flint J., Craddock C. F., Villegas A., Bentley D. P., Williams H. J., Galanello R., Cao A., Wood W. G., Ayyub H. and Higgs D. R. Healing of broken human chromosomes by the addition of telomeric repeats. *Am J Hum Genet.* 1994;55(3):505-12
- [4] Horsley S. W., Daniels R. J., Anguita E., Raynham H. A., Peden J. F., Villegas A., Vickers M. A., Green S., Wayne J. S., Chui D. H., Ayyub H., MacCarthy A. B., Buckle V. J., Gibbons R. J., Kearney L. and Higgs D. R. Monosomy for the most telomeric, gene-rich region of the short arm of human chromosome 16 causes minimal phenotypic effects. *Eur J Hum Genet.* 2001;9(3):217-25
- [5] Varley H., Di S., Scherer S. W. and Royle N. J. Characterization of terminal deletions at 7q32 and 22q13.3 healed by De novo telomere addition. *Am J Hum Genet.* 2000;67(3):610-22
- [6] Wong A. C., Ning Y., Flint J., Clark K., Dumanski J. P., Ledbetter D. H. and McDermid H. E. Molecular characterization of a 130-kb terminal microdeletion at 22q in a child with mild mental retardation. *Am J Hum Genet.* 1997;60(1):113-20

[7] Yatsenko S. A., Brundage E. K., Roney E. K., Cheung S. W., Chinault A. C. and Lupski J. R. Molecular mechanisms for subtelomeric rearrangements associated with the 9q34.3 microdeletion syndrome. Hum Mol Genet. 2009;18(11):1924-36

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60