A novel third type of recurrent NF1 microdeletion mediated by non-allelic homologous recombination between LRRC37B-containing low-copy repeats in 17q11.2

Kathrin Bengesser, David N. Cooper, Katharina Steinmann, Lan Kluwe, Nadia Chuzhanova, Katharina Wimmer, Marcos Tatagiba, Sigrid Tinschert, Victor Mautner, Hildegard Kehrer-Sawatzki

Human Mutation

A novel third type of recurrent NF1 microdeletion mediated by non-allelic homologous recombination between LRRC37B-containing low-copy repeats in 17q11.2

Journal:	*Human Mutation*
Manuscript ID:	humu-2010-0060.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	18-Mar-2010
Complete List of Authors:	Bengesser, Kathrin; University of Ulm, Human Genetics Cooper, David; Cardiff University, Institute of Medical Genetics, College of Medicine Steinmann, Katharina; University of Ulm, Human Genetics Kluwe, Lan; University Hospital Eppendorf, Hamburg, Department of Maxillofacial Surgery Chuzhanova, Nadia; University of Central Lancashire, School of Computing, Engineering and Physical Sciences Wimmer, Katharina; Medical University Innsbruck, Medical Genetics Section, Department of Medical Genetics, Molecular and Clinical Pharmacology Tatagiba, Marcos; Eberhard Karls University Tuebingen, Department of Neurosurgery Tinschert, Sigrid; Carl Gustav Carus Technical University, 7 Institute of Clinical Genetics Mautner, Victor; University Hospital Eppendorf, Maxillofacial Surgery Kehrer-Sawatzki, Hildegard; University of Ulm
Key Words:	Genomic disorders, NAHR, microdeletions, neurofibromatosis type-1, segmental duplications

A novel third type of recurrent *NF1* microdeletion mediated by non-allelic homologous recombination between *LRRC37B*-containing low-copy repeats in 17q11.2

Kathrin Bengesser[1], David N. Cooper[2], Katharina Steinmann[1], Lan Kluwe[3], Nadia A. Chuzhanova[4], Katharina Wimmer[5], Marcos Tatagiba[6], Sigrid Tinschert[7], Victor-Felix Mautner[3], Hildegard Kehrer-Sawatzki[1]

[1] Institute of Human Genetics, University of Ulm, Ulm, Germany

[2] Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, UK

[3] Department of Maxillofacial Surgery, University Hospital Eppendorf, Hamburg, Germany

[4] School of Computing, Engineering and Physical Sciences, University of Central Lancashire, Preston, UK

[5] Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University Innsbruck, Innsbruck, Austria

[6] Department of Neurosurgery, Eberhard-Karls-University, Tübingen, Germany

[7] Institute of Clinical Genetics, Medical Faculty Carl Gustav Carus, Technical University Dresden, Germany

Corresponding author:

Prof. Dr. Hildegard Kehrer-Sawatzki

Institute of Human Genetics

University of Ulm

Albert-Einstein-Allee 11

89081 Ulm, Germany

Email: hildegard.kehrer-sawatzki@uni-ulm.de

Phone: 0049 731 500 65421

ABSTRACT: Large microdeletions encompassing the neurofibromatosis type-1 (*NF1*) gene and its flanking regions at 17q11.2 belong to the group of genomic disorders caused by aberrant recombination between segmental duplications. The most common *NF1* microdeletions (type-1) span 1.4-Mb and have breakpoints located within NF1-REPs A and C, low-copy repeats (LCRs) containing *LRRC37*-core duplicons. We have identified a novel type of recurrent *NF1* deletion mediated by non-allelic homologous recombination (NAHR) between the highly homologous NF1-REPs B and C. The breakpoints of these ~1.0-Mb ('type 3') *NF1* deletions were characterized at the DNA sequence level in three unrelated patients. Recombination regions, spanning 275bp, 180bp and 109bp respectively, were identified within the *LRRC37B-P* paralogues of NF1-REPs B and C, and were found to contain sequences capable of non-B DNA formation. Both LCRs contain *LRRC37*-core duplicons, abundant and highly dynamic sequences in the human genome. NAHR between *LRRC37*-containing LCRs at 17q21.31 is known to have mediated the 970-kb polymorphic inversions of the *MAPT*-locus that occurred independently in different primate species, but also underlies the syndromes associated with recurrent 17q21.31 microdeletions and reciprocal microduplications. The novel *NF1* microdeletions reported here provide further evidence for the unusually high recombinogenic potential of *LRRC37*-containing LCRs in the human genome.

KEY WORDS: Genomic disorders, NAHR, microdeletions, neurofibromatosis type-1, NF1, non-allelic homologous recombination, 17q11.2

Introduction

The most frequently recurring mutations causing neurofibromatosis type-1 (NF1; MIM# 162200) are submicroscopic deletions in 17q11.2 that encompass the *NF1* gene and its flanking regions. Approximately 5% of all NF1 patients exhibit such *NF1* 'microdeletions' [Cnossen et al., 1997; Rasmussen et al., 1998; Kluwe et al., 2004]. So far, two distinct types of recurrent *NF1* microdeletion (type-1 and type-2) have been identified which differ both in terms of their size and their breakpoint positions. Type-1 deletions are the most common, encompassing 1.4 Mb and resulting in the loss of 14 genes including *NF1* [Figure 1; Dorschner et al., 2000; Jenne et al., 2001; Lopéz-Correa et al., 2001]. The breakpoints of these type-1 *NF1* deletions are located within low-copy repeats, termed NF1-REPs A and C, located 366 kb centromeric and 643 kb telomeric to the *NF1* gene, respectively. In common with many other types of low-copy repeat, NF1-REPs A and C are composed of different modules or blocks of paralogous sequences [Jenne et al., 2003; Forbes et al., 2004; De Raedt et al., 2004]. The largest sequence blocks shared by NF1-REPs A and C are 51 kb in length and exhibit 97.5% sequence identity [Forbes et al., 2004]. Recombination between the NF1-REPs constitutes the major mutational mechanism underlying type-1 *NF1* deletions. Since these recombination events occur between paralogous (non-allelic) sequences, this has been termed non-allelic homologous recombination (NAHR). NAHR is now recognized as being the most common mechanism underlying genomic disorders associated with microdeletions and their reciprocal microduplications [reviewed in Shaw and Lupski, 2004; Stankiewicz and Lupski, 2006; Gu et al., 2008]. The breakpoints of type-1 *NF1* deletions occur predominantly in two hotspot regions within NF1-REPs A and C viz. the paralogous recombination sites 1 and 2 (PRS1 and PRS2) which span 2.9 kb and 3.5 kb, respectively [Lopéz-Correa et al., 2001; Forbes et al., 2004; De Raedt et al., 2006; Steinmann et al., 2008]. A second type of recurrent gross *NF1* deletion ('type-2') has also been described which encompasses only 1.2

Mb. The breakpoints of type-2 deletions are located within the *SUZ12* gene (suppressor of zeste 12, MIM#606345) and its pseudogene, *SUZ12P,* located close to NF1-REPs A and C, respectively (Figure 1). Type-2 deletions lead to the loss of only 13 genes since the *LRRC37B* (leucine-rich repeat-containing 37B) gene is not included within the type-2 deletion interval (Figure 1). In contrast to type-1 deletions, type-2 *NF1* deletions are frequently associated with somatic mosaicism (i.e. where cells with the deletion co-exist with normal cells) indicating a postzygotic origin for the deletions. Type-2 *NF1* deletions are, like type-1 *NF1* deletions, predominantly caused by NAHR [Petek et al., 2003; Kehrer-Sawatzki et al., 2004; Spiegel et al., 2005; Steinmann et al., 2007; Roehl et al., 2010].

In addition to NF1-REPs A and C, a third low-copy repeat (termed NF1-REP B) has been identified within the *NF1* gene region, located ~44 kb centromeric to the *NF1* gene [Figure 1; Jenne et al., 2003; Forbes et al., 2004; De Raedt et al., 2004]. NF1-REP B spans 43 kb and constitutes a potential target for NAHR involving the *NF1* gene region since it contains sequences with high homology to NF1-REP C. The strategy of detecting a new genomic disorder by identifying highly homologous LCRs that could mediate NAHR, in a way that is predictable on the basis of their length, orientation and distance between them, is not new, having been pioneered by Sharp et al. [2006]. Adopting this approach, we investigated three NF1 patients whose unusually short (~1.0 Mb) deletions were of a size consistent with that expected if the deletion breakpoints were located within NF1-REPs B and C. Localization of the deletion breakpoints at the highest possible resolution indicated that the recombination events did indeed occur within paralogous sequence blocks of NF1-REPs B and C and that the causal mechanism underlying this novel type of recurrent *NF1* deletion (which we have termed 'type-3') was likely to be NAHR.

Patients, Materials and Methods

NF1 Patients Harbouring 'Type-3' Deletions

NF1 patients TOP and Z41/03 were identified as having a gross *NF1* deletion by FISH analysis in previous studies [Jenne et al., 2003; Kehrer-Sawatzki et al., 2004]. However, the breakpoints of the respective deletions were not characterized precisely. Patient 2176 was identified from among 800 NF1 patients, clinically examined in the Hamburg NF-Centre (Department of Maxillofacial Surgery, University Medical Centre, Hamburg-Eppendorf, Germany), using microsatellite markers as previously described [Kluwe et al., 2005].
The clinical phenotypes of the three 'type-3' NF1 patients are described below. All three deletions are thought to have occurred *de novo*, since none of the parents or their siblings evidenced any clinical features of NF1. This study was approved by the Local Institutional Review boards of the participating centres. Informed consent was obtained from all patients and their relatives.

Patient TOP

The clinical phenotype of this male patient has already been described [Jenne et al., 2003]. At the age of 11.5 years, he had 45 cutaneous neurofibromas, coarse facial appearance with a large nose. Mental retardation (IQ of 53) was noted in this patient as well as large hands and feet.

Patient 2176

This male patient was clinically investigated at the age of 36 years. He was of normal height (175 cm; 50th percentile) but microcephaly (head circumference: 55.5cm, 3rd percentile) and large hands and feet were noted. He had six café-au-lait-spots, axillary and inguinal freckling as well as Lisch nodules. Approximately 1000 cutaneous and 15 subcutaneous neurofibromas

but no externally visible plexiform neurofibromas were evident. Spinal neurofibromas caused spinal compression and spinal ataxia in this patient. He exhibited scoliosis (with a curve of 20°) and pectus excavatum. Hypertelorism, down-slanting palpebral fissures, teeth dysplasia, gingival hyperplasia and a broad neck were observed. Learning problems were also recorded. He attended secondary school and now works as a mechanic. He gives the impression of being cognitively impaired but formal IQ testing has not been performed.

Patient Z41/03

This female patient was first investigated at the age of 9 years. Delayed motor development, mental retardation, ataxia, hamartomas of the thalamus and facial dysmorphic features were noted. At the age of 11 years, she developed a malignant peripheral nerve sheath tumour. The patient was previously (and incorrectly) described as having a type-1 (1.4 Mb) *NF1* deletion [Kehrer-Sawatzki et al., 2004]. However, upon reinvestigation by MLPA (see below), it became apparent that she possessed a shorter ~1.0 Mb deletion.

FISH Analysis

Blood lymphocytes from all three patients were investigated by fluorescent *in situ* hybridisation (FISH) as previously described [Kehrer-Sawatzki et al., 2004]. In brief, FISH was performed using differentially labelled probes, BAC RP11-142O6 (spanning the proximal *NF1* gene region) and BAC RP11-55A13 (located at 17q24). A total of 30 metaphase spreads and at least 100 interphase nuclei were evaluated in order to exclude mosaicism (which would have been evidenced by the additional presence of normal cells lacking the deletion).

Deleted: such as

Deleted: ¶ ¶

Multiplex Ligation-Dependent Probe Amplification (MLPA) Analysis

DNA was isolated from peripheral blood samples of the patients using the QIAamp kit (Qiagen, Valencia, CA, USA). All three patients were investigated by MLPA using the SALSA P122 C1 MLPA assay (MRC-Holland, Amsterdam, Netherlands) according to the manufacturer's instructions and as described by Wimmer et al. [2005].

Analysis of Somatic Cell Hybrids Containing Only Deletion-bearing Chromosomes 17

Somatic human-mouse cell hybrids containing only the deletion-bearing chromosomes 17 derived from patients 2176 and TOP were generated as described [Kehrer-Sawatzki et al., 2004]. DNAs extracted from these hybrids were investigated by PCR in order to identify the extent of the deletions. In a first step, PCR was performed so as to determine whether or not the region-specific single copy sequences and microsatellite markers located in 17q11.2 were present within the deletion interval. This procedure helped to narrow down the locations of the deletion breakpoints as summarized in Supp. Table S1 (PCR primer sequences available on request). The characterization by PCR of somatic cell hybrids containing only the deletion-bearing chromosomes 17 offers the advantage that amplified sequences from the chromosome 17 containing the deletion can be analysed in the absence of co-amplified homologous sequences from the normal chromosome 17. In the second step of the analysis, PCR-fragments were amplified from different regions of the duplicated *LRRC37B* sequences at 17q11.2 (Supp.Table S2). To identify the origin of the PCR-products amplified from the hybrid DNAs (and hence from the deletion-bearing chromosomes 17), the PCR-products were sequenced and the resulting sequences were aligned by BLAT against the human reference sequence (http://genome.ucsc.edu/; human genome assembly 18, NCBI 36). The origin of the amplified PCR-fragments (and, in particular, from which of the three NF1-REPs they had been amplified) was determined by evaluation of the paralogous sequence variants (PSVs) that serve to distinguish NF1-REPs A, B and C. Breakpoint-spanning PCRs were performed with primers listed in Supp. Table S3 using the Expand Long Template PCR-system (Roche,

Deleted: . The primer sequences used to amplify these PCR-fragments are listed in

Deleted: on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA)

Penzberg, Germany). Sequence analysis of the deletion breakpoint-spanning fragments then revealed the locations of the recombination regions within which the breakpoints must have occurred. These recombination (deletion breakpoint) regions were identified by virtue of their being flanked by PSVs specific to either NF1-REP B or NF1-REP C (Supp. Figure S1).

Microsatellite Marker Analysis to Identify the Parental Origin of the Deletion

To determine the parental origin of the deletion identified in patient 2176, we analysed DNA isolated from somatic cell hybrids containing either the deletion-bearing chromosome 17 or the normal chromosome 17 from the patient, as well as DNA extracted from blood samples of the patient and his parents by PCR of polymorphic markers on chromosome 17 as previously described [Steinmann et al., 2008].

Deleted: .

Deleted: For this purpose, we used 6FAM-labelled primers and investigated the respective marker alleles by capillary electrophoresis on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA)

Deleted: (

Deleted:)

Bioinformatic Analysis

Sequences spanning the breakpoint (recombination) regions of all three deletions were screened for the presence of 32 DNA sequence motifs (and their complements) of length ≥5 nucleotides (nt) which have been reported to be associated with DNA breakage and repair [Abeysinghe et al., 2003; Ball et al., 2005; Chuzhanova et al., 2009]. In addition, both the recombination regions and 100-200 bp segments flanking them on either side, were analysed by complexity analysis [Gusev et al., 1999] to identify direct, inverted and symmetric repeats ≥6 nt separated by no more than 20 bp. These repeats are capable of non-B DNA formation, in particular slipped, cruciform and triplex structures [Wells, 2007].

For each of the above searches, the statistical significance of our findings was assessed by comparison with 1000 control datasets using z-score statistics [as described in Chuzhanova et al., 2009]. Two regions of chromosome 17, flanking the *NF1* gene but not including any of the known recombination hotspots, were used to generate matching control datasets: [coordinates 26,109,477 to 26,359,000 (5' to the *NF1* gene, between but not including NF1-

REPs A and B) and 26,889,356 to 27,203,003 (3' to the *NF1* gene up to, but not including, NF1-REP C); human genome assembly hg18 (NCBI build 36)].

Results

MLPA and FISH Analysis

The three patients (TOP, Z41/03 and 2176) were selected for analysis on the basis that they (i) did not fit into the category of either type-1 or type-2 deletions and (ii) were all determined to be ~1.0 Mb in size as measured by MLPA. Whilst the MLPA probe located within the *RNF135* gene (MIM#611358) was not deleted, the MLPA probe within *NF1* exon 1 revealed a copy number loss (Figure 1; Supp. Table S4). We therefore surmised that the centromeric deletion breakpoint might reside within NF1-REP B. The telomeric deletion breakpoints were found, in all three patients, to be located between the MLPA probes *LRRC37B* and *ZNF207*, suggesting that they were located within NF1-REP C (Figure 1; Supp. Table S4). All three deletions were estimated to encompass ~1 Mb, implying the loss of 9 genes including *NF1*.

These findings indicated that patient Z41/03 must have been erroneously reported, on the basis of previously performed FISH analysis, as having a 1.4 Mb type-1 *NF1* deletion [Kehrer-Sawatzki et al. 2004]. Clearly, FISH has a much lower resolution than MLPA and must have been insufficiently precise on this occasion to determine the exact extent of the deletion. Mosaicism with normal cells was not observed in the three patients by means of interphase FISH.

Deleted: as determined

Approximation of the Deletion Breakpoint Regions in Patients 2176 and TOP

To this end, DNA from somatic cell hybrids, harbouring only the deletion-bearing chromosomes 17 from patients TOP and 2176, were used to narrow down the deletion breakpoint regions by PCR of both single copy sequences and microsatellite markers. Taken

together with the MLPA results, this analysis suggested that the centromeric deletion breakpoints would map between genomic positions 26,339,386 and 26,429,386 in both patients, whereas the telomeric deletion breakpoint regions were assigned to between nucleotide positions 27,372,703 and 27,717,868 (Supp. Table S1). These results strongly implied that the deletion breakpoints were located within NF1-REPs B and C (Figure 1 and Supp. Figure S2).

Deleted: s

Reinvestigation of the Structure of the NF1-REPs

In order to assign the breakpoint regions to the paralogous subunits of the NF1-REPs, it was necessary to re-analyse the structure of the NF1-REPs in 17q11.2 according to the base-pair numbering of the human genome assembly (hg 18, NCBI 36, Ensembl version 54.36p). We therefore determined the positions of the different paralogous sequence blocks or modules that constitute the NF1-REPs based upon previous investigations of BAC/PAC contigs of the corresponding region [Jenne et al., 2003; Forbes et al., 2004; De Raedt et al., 2004]. NF1-REP A spans 128 kb and is located 366 kb centromeric to the *NF1* gene, whereas NF1-REP C is 75 kb in length and lies 643 kb telomeric to the *NF1* gene (Supp. Figure S2). NF1-REP B, located 44 kb centromeric to *NF1,* is the shortest of the three LCRs, spanning only ~43 kb. Our analysis confirmed previous findings that NF1-REPs A and C are paralogues with homology to sequences located at 19p13.12 [Forbes et al., 2004; De Raedt et al., 2004]. Indeed, NF1-REP A contains a duplicated pseudogene fragment of the latrophilin-1 precursor gene (*LPHN1*) gene located at 19p13.12 [Forbes et al., 2004]. By contrast, no sequences with homology to 19p13.12 are present within NF1-REP B. Although the functional *LRRC37B* gene is located exclusively in NF1-REP C, pseudogene fragments of this gene are present within all three NF1-REPs [Supp. Figure S2; Jenne et al., 2003; Forbes et al., 2004; De Raedt et al., 2004].

Deleted: (Figure 2)

Sequence alignments of the *LRRC37B* gene against the human reference sequence (hg18, NCBI 36) indicated that in addition to the various *LRRC37B* pseudogene fragments within 17q11.2, several other genes located telomeric to the *NF1* gene region also possess strong sequence similarity to *LRRC37B* (Supp. Table S5) viz. the genes *LRRC37A*, *LRRC37A2*, *LRRC37A3* and multiple pseudogene fragments thereof. The members of the *LRRC37* multigene family are located in clusters at 17q11.2 (25-27 Mb), 17q21.31 (40-42 Mb) and 17q24.1 (60-63 Mb) [Supp. Table S5; Zody et al., 2006a].

In addition to the *LRRC37B* sequences, NF1-REPs A and B also contain pseudogene fragments of the *SMURF2* (SMAD-specific E3 ubiquitin protein ligase 2, 17q24.1, MIM#605532) gene, which are absent from NF1-REP C. Thus, the paralogous sequence blocks of the NF1-REPs include *LRRC37B* sequences, *SMURF2* pseudogenes and sequences with homology to 19p13.12, respectively, together constituting the different modules that form the characteristic structures of each NF1-REP [Jenne et al., 2003; Forbes et al., 2004; De Raedt et al., 2004]. Assuming that non-allelic homologous recombination (NAHR) was responsible for the three ~1.0 Mb *NF1* deletions under investigation, only the *LRRC37* sequences within NF1-REPs B and C, as direct repeats, would have had the potential to mediate NAHR events. We therefore designed a series of different PCR experiments, with primers located within the *LRRC37B* sequences, in order to identify the recombination regions more precisely.

Deleted: ;

Deleted: (Figure 2)

Identification of the Breakpoints in Patient TOP

To identify the deletion breakpoints, PCR experiments were performed using primers located within the *LRRC37B* sequences and DNA from a hybrid cell line containing only the deletion-bearing chromosome 17 from patient TOP as a template. Sequence analysis of the resulting PCR-products, alignments against the human reference sequence, and investigation of the paralogous sequence variants (PSVs) that distinguish the paralogous *LRRC37B*

Deleted: using

sequences, together served to identify the regions not included within the bounds of the deletion. For example, PCR-fragment B7/8 was derived exclusively from NF1-REP A. We therefore inferred that the paralogous fragment B7/8, located within NF1-REP C, had been deleted (Figure 2A). Adopting this approach, we analysed a set of 11 PCR-fragments (listed in Table 1); the results then enabled us to assign the centromeric deletion breakpoint to intron 3 of the *LRRC37B* pseudogene within NF1-REP B and the telomeric breakpoint to the homologous region within the *LRRC37B* pseudogene of NF1-REP C (Figure 2A).

In the next step of our analysis, we cloned a 1.6 kb breakpoint-spanning PCR-fragment amplified with primers B9 and Int3Cr, and sequenced it completely (Supp. Table S3). This enabled the breakpoints of the deletion in patient TOP to be assigned to 275 bp regions between positions 26,391,423 and 26,391,697 in NF1-REP B and between positions 27,440,992 and 27,441,266 in NF1-REP C. Consequently, the deletion of patient TOP must span a total of 1,049,843 bp (~1.0 Mb). Since the 275 bp-spanning deletion breakpoint regions lacked any PSVs or other distinguishing sequence differences between NF1-REPs B and C, it was not possible to narrow down the deletion breakpoints any further (Supp. Figure S1). The recombination-associated breakpoints must have occurred within these identical 275 bp segments located within introns 3 of the *LRRC37B* pseudogene copies present in both NF1-REPs B and C (Figure 3). Thus, non-allelic homologous recombination (NAHR) gave rise to the deletion in patient TOP.

Identification of the Breakpoints in Patient 2176

The mapping of the PCR-fragments listed in Table 1 indicated that the centromeric deletion breakpoint in this patient was located within NF1-REP B, between PCR-fragments Int2A and Ex1 (Figure 2B; Table 1). Sequence analysis of PCR-fragment Int1B revealed that this PCR-fragment was breakpoint-spanning. To verify this by means of an independent PCR assay, we performed a breakpoint-spanning PCR on genomic DNA from patient 2176, using primer

Deleted: indicate the origin of the amplified sequences and hence
Deleted: Thus, f
Deleted: , amplified from hybrid DNA containing only the deletion-bearing chromosome 17 from patient TOP,
Deleted: 3
Deleted: 3
Deleted: with primers Int3Df and Int3Dr
Deleted: s 2 and
Deleted: e 275 bp sequences were absolutely identical between both LCRs, strongly supporting
Deleted: caused
Deleted: as the mechanism underlying the
Deleted: 3

Int2C located in NF1-REP B and primer Int2r located in NF1-REP C (Supp. Table S3). Sequence analysis of this fragment indicated that the deletion breakpoints were located between positions 26,397,050 and 26,397,229 in NF1-REP B and between positions 27,446,691 and 27,446,870 in NF1-REP C. Thus, the deletion of patient 2176 spans 1,049,820 bp (~1.0 Mb). The 180 bp deletion breakpoint regions within *LRRC37B-P* introns 1 of NF1-REPs B and C were found to be completely identical (Supp. Figure S1). Hence, the deletion of patient 2176 was also mediated by NAHR.

Microsatellite marker analysis indicated that the germline deletion of patient 2176 occurred on the paternally inherited chromosome (Supp. Figure S3).

Identification of the Deletion Breakpoint Regions in Patient Z41/03

Unfortunately, we were unable to establish a somatic cell hybrid line from patient Z41/03 which would have facilitated the mapping of the deletion breakpoints. Working under the assumption that the deletion breakpoint of patient Z41/03 was located close to the breakpoint region of patient 2176, we performed the Int2C/Int2r breakpoint-spanning PCR on lymphocyte-derived DNA from patient Z41/03. This PCR was designed to amplify deletion breakpoint-spanning fragments specifically; wild-type PCR fragments that would have been derived either from NF1-REPs B or C were not co-amplified. Sequence analysis of this 2.3 kb fragment with nested primers (Supp. Table S3) indicated that the deletion breakpoints in patient Z41/03 were located between positions 26,397,231 and 26,397,339 in NF1-REP B and between positions 27,446,872 and 27,446,980 in NF1-REP C. A more precise determination of the breakpoints within these 109 bp-spanning recombination regions was not possible owing to the sequence identity between the regions of strand exchange within NF1-REPs B and C (Supp. Figure S1). Interestingly, the deletion breakpoint regions of patients 2176 and Z41/03 lie immediately adjacent to one another, separated by only 2 bp within introns 1 of the *LRRC37B-P* sequences (Figure 3).

Deleted: . The primer sequences and their respective positions are listed in
Deleted: s
Deleted: The 2.3 kb breakpoint-spanning PCR-fragment generated was sequenced using nested primers Int1Af and Int1Ar. Alignments against the human reference sequence
Deleted: we may surmise that
Deleted: in this individual
Deleted: Nevertheless, w
Deleted: identified in
Deleted: genomic
Deleted: A 2.3 kb breakpoint-spanning fragment was amplified from lymphocyte-derived DNA of this patient. The Int2C/Int2r
Deleted: (as confirmed by sequence analysis of the PCR product)
Deleted: .
Deleted: The Int2C/Int2r PCR-fragment was then fully sequenced using
Deleted: Int1Af and Int1Ar
Deleted: . The evaluation of PSVs
Deleted: of NF1-REPs B and C
Deleted: 4

Comparative Analysis of the Deletion Breakpoint Regions

The deletion breakpoint regions in all three patients were localized to regions flanked by PSVs that were uniquely characteristic for either NF1-REP B or NF1-REP C (Supp. Figure S1). The NAHR-mediated DNA breaks that gave rise to the ~1.0 Mb *NF1* deletions must therefore have occurred within these deletion breakpoint regions. The deletion breakpoint regions were consequently assigned as recombination regions (RRs) and are identical between NF1-REPs B and C (Supp. Figure S4). In patients Z41/03 and 2176, the juxtaposed RRs are located within introns 1 of the *LRRC37B-P* of NF1-REP B and C whereas in patient TOP, the RR lies within introns 3 of the same *LRRC37B* pseudogene, ~6 kb distant from the RRs of patients Z41/01 and 2176 (Figure 3). None of the RRs of the ~1.0 Mb deletions characterized in this study colocalized with the recurrent breakpoints of the common 1.4 Mb *NF1* deletions which are located within either the PRS2 or PRS1 breakpoint hotspot regions. The distances between the breakpoint regions of the ~1.0 Mb deletions and the type-1 deletion hotspots PRS1 and PRS2, given in Table 2, are minimally >7.9 kb. Moreover, both PRS1 and PRS2 are located within the paralogous sequence blocks with homology to sequences from 19p13.12 (Supp. Figure S2). By contrast, the telomeric breakpoints of the three deletions investigated here lie telomeric to PRS1 and PRS2 within NF1-REP C.

Deleted: 2

Deleted: 4

Deleted: (Figures 2 and 4)

To investigate more closely why the NAHR events responsible for the 'type-3' deletions reported here occurred within the observed deletion breakpoint regions, we screened the RRs for the presence of recombination-promoting motifs potentially involved in DNA breakage and repair [Love et al., 1995; Giese et al., 1997; Chuzhanova et al., 2009]. Two recombination-associated motifs, namely the topoisomerase cleavage site, YCCTT, and the deletion hotspot consensus sequence, TGRRKM, were found to be overrepresented within the RRs at the 1% level. None of the other motifs known to be associated with site-specific cleavage/recombination and gene rearrangement [Abeysinghe et al. 2003; Ball et al. 2005;

Cullen et al. 2002; Myers et al. 2008] were found to be overrepresented within the RRs (Supp. Table S6). In order to identify other sequence features that could have been responsible for local DNA breakage within the RRs, we screened the RRs on their own (and the RRs plus their flanking regions) for the presence of repeats capable of non-B DNA structure formation. Simple repetitive DNA sequences capable of adopting non-B DNA conformations are known to be highly mutagenic and predispose to DNA breakage as evidenced both by *in vitro* studies and empirical biochemical data on non-B DNA conformations [Bacolla et al. 2004, 2008; Bacolla and Wells, 2004; Wang and Vasquez, 2006]. Both direct and inverted repeats capable of non-B DNA structure formation were found within the RRs of patients 2176 and TOP (Supp. Figure S5). However, only direct repeats were found to be overrepresented ($p<0.05$) within the RRs. In addition, long (10 bp) polypyrimidine tracts, also implicated in non-B DNA structure formation, were overrepresented ($p<0.01$) within the RRs of all patients (Supp. Figure S5).

Discussion

In this study, we have identified a novel type of recurrent *NF1* microdeletion ('type-3') that spans ~1.0 Mb and is mediated by non-allelic homologous recombination (NAHR) between NF1-REPs B and C at 17q11.2 (Figure 1). Whilst NAHR between NF1-REPs A and C is well known to cause the common 1.4 Mb deletions observed in 70-80% of all patients with large *NF1* deletions [Lopéz-Correa et al., 2001; Forbes et al., 2004; Kehrer-Sawatzki et al., 2004; De Raedt et al., 2006], NAHR between NF1-REPs B and C would appear to be much less common, probably accounting for <5% of all *NF1* microdeletions. To our knowledge, only two patients have previously been reported to possess breakpoints located within NF1-REPs B and C: patient UWA 113-1 [Dorschner et al., 2000] and patient TOP [Jenne et al., 2003] reanalysed here. Since these deletions were originally mapped by FISH, unequivocal confirmation that the breakpoints were located within the NF1-REPs was initially lacking.

In this study, we reinvestigated patient TOP and demonstrated at the highest possible resolution that the deletion in this patient was mediated by NAHR between NF1-REPs B and C. Further, we identified two additional NF1 patients whose similarly sized deletions originated via the same mechanism. The deletion breakpoints of these two patients, Z41/03 and 2176, were mapped to within intron 1 of the *LRRC37B* pseudogenes in NF1-REPs B and C. The breakpoints occurred within regions of absolute sequence identity between NF1-REPs B and C; hence, recombination regions were assigned that spanned 109 bp and 180 bp, respectively (Supp. Figure S1). The deletion breakpoint regions identified in patient TOP were however located within intron 3 of the *LRRC37B-P* sequences of NF1-REPs B and C, ~6 kb distant from the breakpoint regions in patients 2176 and Z41/03 (Figure 3). These results indicate that the paralogous *LRRC37B* pseudogene sequences in NF1-REPs B and C, which constitute highly homologous repeats that lie in the same orientation, exhibit an increased propensity to undergo NAHR giving rise to the ~1.0 Mb *NF1* deletions. It should be noted that the breakpoint regions, in which the homologous recombination events underlying the deletions in patients Z41/03 and 2176 must have occurred, are immediately adjacent to one another, separated by only 2 bp (Figure 3B). We developed a breakpoint-spanning PCR that permitted the specific amplification of the deletion breakpoints from genomic DNA of patients Z41/03 and 2176. This PCR-based approach could be used in future studies to screen other NF1 patients for ~1.0 Mb *NF1* deletions in order to investigate (i) the question of breakpoint recurrence and (ii) whether *LRRC37B-P* intron 1 constitutes a specific hotspot for NAHR within this genomic region.

Interphase FISH analysis of peripheral blood lymphocytes of the three patients with type-3 *NF1* deletions was not indicative of mosaicism with normal cells. Thus, we surmise that the deletions must have occurred in the parental germline. Mosaicism is indeed an issue in *NF1* deletion diagnostics because it occurs much more frequently than previously assumed [reviewed in Kehrer-Sawatzki and Cooper, 2008].

Deleted: 4

Deleted: 4

Deleted: did

Deleted: e

Deleted: it is most likely to assume

Formatted

Deleted: is

Microsatellite marker analysis indicated that the deletion of patient 2176 occurred on the paternally inherited chromosome. By contrast, the majority of type-1 *NF1* deletions are mediated by NAHR during maternal meiosis I [López-Correa et al., 2000]. Further studies are needed to determine whether or not there is a difference in the parental origin of type-1 and type-3 *NF1* deletions.

Patients with large deletions in the *NF1* gene region have been reported as a group to exhibit a more severe clinical phenotype than the general NF1 population. The deletion-associated phenotype includes mental retardation, high tumour burden and dysmorphic facial features [Venturin et al., 2004; Mensink et al., 2006; Mautner et al., 2010]. Since mental retardation was observed in all three patients with type-3 *NF1* deletions described here, we speculate that a modifying gene, located either within the *NF1* gene (e.g. the oligodendrocyte-myelin glycoprotein gene, *OMG*, MIM#164345) or telomeric to it (Figure 1), could have contributed to this phenotype. Genotype-phenotype analysis of further patients with type-3 *NF1* deletions would be necessary to substantiate this tentative conclusion.

Deleted: T
Deleted: o
Deleted: necessary
Deleted: as a group
Deleted: as
Deleted:
Deleted: i
Deleted:
Deleted: may surmise
Deleted: contributing to this phenotype is
Deleted: in
Deleted: direction
Deleted: /
Deleted: are
Deleted: observation

The recombination events underlying the ~1.0 Mb deletions appear to have been caused by aberrant pairing of the directly oriented *LRRC37B* pseudogene copies located within NF1-REPs B and C. It is however unclear what factors could have influenced the local DNA breakage that ultimately gave rise to the recombination-associated strand exchange. Recently, Chuzhanova et al. [2009] showed that motifs associated with recombinational activity and sequences with the potential to form non-B DNA structures are both overrepresented in DNA sequence tracts involved in gene conversion causing human inherited disease. Both gene conversion and the non-allelic strand exchange underlying the ~1.0 Mb deletions reported here are known to be associated with homologous recombination and hence could in principle be influenced by such sequence features. Within the recombination regions in introns 1 and 3 of the *LRRC37B-P* paralogues in NF1-REPs B and C, an overrepresentation of two recombination-associated motifs was observed. Also overrepresented were direct repeats and

polypyrimidine tracts (10 bp) both of which have been implicated in non-B DNA structure formation (Supp. Figure S5). Our analysis therefore suggests that sequences capable of adopting non-B DNA conformations may have combined with recombination-inducing motifs so as to promote NAHR, thereby resulting in the ~1.0 Mb 'type-3' deletions identified in NF1 patients TOP, 2176 and Z41/03.

The NF1-REPs giving rise to either the common type-1 *NF1* deletions exhibit a modular structure as do the NF1-REPs mediating the much rarer ~1.0 Mb 'type-3' *NF1* deletions investigated here. The different component modules comprise distinct paralogous sequence blocks in either direct or inverted orientation (Supp. Figure S2). One of these modules contains *LRRC37B* sequences. Although the single copy functional *LRRC37B* gene maps to within NF1-REP C, additional pseudogene fragments of *LRRC37B* are present within all three NF1-REPs [Figure 1; Jenne et al., 2003; Forbes et al., 2004; De Raedt et al., 2004]. *LRRC37B* is a member of the *LRRC37* multigene family which appears to have arisen by multiple intrachromosomal duplicative transposition events during primate evolution. Multiple copies of *LRRC37* sequences are present on macaque chromosome 16, homologous to human chromosome 17 [Zody et al., 2008]. Further, BLAT (http://genome.ucsc.edu) sequence alignment of human *LRRC37* sequences against the marmoset genome (Callithrix_jacchus-3.2, Feb 2009), indicated that this New World monkey also possesses multiple copies of this gene (data not shown) suggesting that the expansion of this gene family began very early during primate evolution.

Deleted: 2

LRRC37 sequences are among the most abundant segmental duplications on chromosome 17, representing ~20% of all intrachromosomal duplications on this chromosome [Zody et al., 2006a]. Intrachromosomal LCRs with multiple paralogues of high abundance (located on different human chromosomes) are sometimes known to contain LCR-specific core duplicons which represent some of the most dynamic regions in the human genome (Johnson et al., 2006; Popesco et al., 2006; Zody et al., 2006a, 2006b, 2008; Jiang et al., 2007; Marques-

Bonet and Eichler, 2009). Such a core duplicon of ~4 kb is present in most *LRRC37*-containing LCRs despite their different overall lengths [Zody et al., 2006a]. LCRs containing *LRRC37* core duplicons are located exclusively within 17q and are found within three main clusters at 25-27 Mb, 40-42 Mb and 60-63 Mb [Supp. Table 5; Zody et al., 2006a, 2008]. Phylogenetic reconstruction has indicated that the *LRRC37* duplicons located at 40-42 and 60-63 Mb are more closely related to one another than to those within the NF1-REPs at 25-27 Mb [De Raedt et al., 2004; Zody et al., 2006a]. The ancestral source of the *LRRC37* core duplicons is likely to be the *LRRC37A* gene located at 41.7 Mb because it is the only copy of *LRRC37* with an orthologue in the mouse genome [*Lrrc37a*; XM_137868; Zody et al., 2006a]. Remarkably, *LRRC37* sequences were found in 9 of the 19 syntenic breakpoint regions that were identified when comparing the structure of human chromosome 17 with murine chromosome 11, suggesting genomic instability of the corresponding genomic regions [Zody et al., 2006a]. The high recombinogenic potential of *LRRC37*-containing LCRs on human chromosome 17 is also reflected by the fact that, in addition to mediating the recurrent 'type-3' *NF1* deletions at 17q11.2, they also give rise to other genomic disorders such as the microdeletions and their reciprocal microduplications at 17q21.31 associated with various specific syndromes [Sharp et al., 2006; Koolen et al., 2006; Shaw-Smith et al., 2006; Tan et al., 2009; Kirchhoff et al., 2007; Grisart et al., 2009]. These rearrangements are all mediated by large LCRs that contain the *LRRC37A* gene (and/or sequences highly homologous to it).

Large LCRs containing the *LRRC37A* gene also flank the *MAPT* (microtubule-associated protein tau) gene; NAHR between these LCRs not only mediates the disease-associated microdeletions/duplications within 17q21.31, but has also been responsible for generating the ~970 kb polymorphic inversion of the *MAPT* region [Stefansson et al., 2005; Gijselinck et al. 2006; Cruts et al., 2005; Zody et al. 2008; Antonacci et al., 2009]. The *MAPT* locus and its flanking regions occur in humans as two distinct inversion haplotypes, H1 and H2, which display no recombination between them over a distance of 1.5 Mb [Zody et al. 2008]. These

haplotypes differ in terms of their functional impact: whereas the H1 haplotype has been associated with neurological disorders including Alzheimer disease and amyotrophic lateral sclerosis [Myers et al., 2005; Sundar et al., 2007], the H2 haplotype appears to predispose to recurrent microdeletions associated with the 17q21.31 microdeletion syndrome [Sharp et al., 2006; Koolen et al., 2006; Shaw-Smith et al., 2006]. Importantly, the inversion of the *MAPT* region is not only polymorphic in humans but also in other extant primate species such as chimpanzee, bonobo and orangutan [Zody et al., 2008]. Such polymorphism is strongly suggestive of the likely independent recurrence of this inversion in primate genomes, a conclusion supported by evidence that the inversion has occurred independently in humans and chimpanzees [Zody et al., 2008]. Recurrent NAHR events leading to near identical inversions in different primate species are indicative of the high degree of genomic instability driven by the *LRRC37* core duplicons. The presence of *LRRC37*-containing LCRs at multiple evolutionary breakpoints as well as at a variety of recurring disease-associated rearrangements (including the *NF1* microdeletions reported here) provides powerful evidence to support the contention that *LRRC37* sequences are intimately involved in the process of DNA breakage and non-allelic homologous recombination.

Acknowledgment

This work was supported by the Deutsche Forschungsgemeinschaft grant no. DFG KE724/7-1. The authors wish to thank Helene Spöri for her technical assistance.

References

Abeysinghe SS, Chuzhanova N, Krawczak M, Ball EV, Cooper DN. 2006. Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. Hum Mutat 22:229-244.

Antonacci F, Kidd JM, Marques-Bonet T, Ventura M, Siswara P, Jiang Z, Eichler EE. 2009. Characterization of six human disease-associated inversion polymorphisms. Hum Mol Genet 18:2555-2566.

Bacolla A, Jaworski A, Larson JE, Jakupciak JP, Chuzhanova N, Abeysinghe SS, O'Connell CD, Cooper DN, Wells RD. 2004. Breakpoints of gross deletions coincide with non-B DNA conformations. Proc Natl Acad Sci USA 101:14162-14167.

Bacolla A, Wells RD. 2004. Non-B DNA conformations, genomic rearrangements, and human disease. J Biol Chem 279:47411-47414.

Bacolla A, Larson JE, Collins JR, Li J, Milosavljevic A, Stenson PD, Cooper DN, Wells RD. 2008. Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. Genome Res 18:1545-1553.

Ball EV, Stenson PD, Krawczak M, Cooper DN, Chuzhanova NA. 2005. Micro-deletions and micro-insertions causing human genetic disease: common mechanisms of mutagenesis and the role of local DNA sequence complexity. Hum Mutat 26:205-213.

Chuzhanova N, Chen JM, Bacolla A, Patrinos GP, Férec C, Wells RD, Cooper DN. 2009. Gene conversion causing human inherited disease: evidence for involvement of non-B-DNA-forming sequences and recombination-promoting motifs in DNA breakage and repair. Hum Mutat 30:1189-1198.

Cnossen MH, van der Est MN, Breuning MH, van Asperen CJ, Breslau-Siderius EJ, van der Ploeg AT, de Goede-Bolder A, van den Ouweland AM, Halley DJ, Niermeijer MF. 1997. Deletions spanning the neurofibromatosis type 1 gene: implications for genotype-phenotype correlations in neurofibromatosis type 1? Hum Mutat 9:458-464.

Cruts M, Rademakers R, Gijselinck I, van der Zee J, Dermaut B, de Pooter T, de Rijk P, Del-Favero J, van Broeckhoven C. 2005. Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region. Hum Mol Genet 14:1753-1762.

Cullen M, Perfetto SP, Klitz W, Nelson G, Carrington M. 2002. High-resolution patterns of meiotic recombination across the human major histocompatibility complex. Am J Hum Genet 71:759–776.

De Raedt T, Brems H, Lopez-Correa C, Vermeesch JR, Marynen P, Legius E. 2004. Genomic organization and evolution of the NF1 microdeletion region. Genomics 84:346-360.

De Raedt T, Stephens M, Heyns I, Brems H, Thijs D, Messiaen L, Stephens K, Lazaro C, Wimmer K, Kehrer-Sawatzki H, Vidaud D, Kluwe L, Marynen P, Legius E. 2006. Conservation of hotspots for recombination in low-copy repeats associated with the NF1 microdeletion. Nat Genet 38:1419-1423.

Dorschner MO, Sybert VP, Weaver M, Pletcher BA, Stephens K. 2000. NF1 microdeletion breakpoints are clustered at flanking repetitive sequences. Hum Mol Genet 9:35-46.

Forbes SH, Dorschner MO, Le R, Stephens K. 2004. Genomic context of paralogous recombination hotspots mediating recurrent *NF1* region microdeletion. Genes Chrom Cancer 41:12-25.

Giese K, Pagel J, Grosschedl R. 1997. Functional analysis of DNA bending and unwinding by the high mobility group domain of LEF-1. Proc Natl Acad Sci USA 94:12845-12850.

Gijselinck I, Bogaerts V, Rademakers R, van der Zee J, Van Broeckhoven C, Cruts M. 2006. Visualization of *MAPT* inversion on stretched chromosomes of tau-negative frontotemporal dementia patients. Hum Mutat 27:1057-1059.

Grisart B, Willatt L, Destrée A, Fryns JP, Rack K, de Ravel T, Rosenfeld J, Vermeesch JR, Verellen-Dumoulin C, Sandford R. 2009. 17q21.31 microduplication patients are characterised by behavioural problems and poor social interaction. J Med Genet 46:524-530.

Gu W, Zhang F, Lupski JR. 2008. Mechanisms for human genomic rearrangements. Pathogenetics 1: 4.

Gusev VD, Nemytikova LA, Chuzhanova NA. 1999. On the complexity measures of genetic sequences. Bioinformatics 15:994-999.

Jenne DE, Tinschert S, Reimann H, Lasinger W, Thiel G, Hameister H, Kehrer-Sawatzki H. 2001. Molecular characterization and gene content of breakpoint boundaries in patients with neurofibromatosis type 1 with 17q11.2 microdeletions. Am J Hum Genet 69:516-527.

Jenne DE, Tinschert S, Dorschner MO, Hameister H, Stephens K, Kehrer-Sawatzki H. 2003. Complete physical map and gene content of the human *NF1* tumor suppressor region in human and mouse. Genes Chrom Cancer 37:111-120.

Jiang Z, Tang H, Ventura M, Cardone MF, Marques-Bonet T, She X, Pevzner PA, Eichler EE. 2007. Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. Nat Genet 39:1361-1368.

Kehrer-Sawatzki H, Kluwe L, Sandig C, Kohn M, Wimmer K, Krammer U, Peyrl A, Jenne DE, Hansmann I, Mautner VF. 2004. High frequency of mosaicism among patients with neurofibromatosis type 1 (NF1) with microdeletions caused by somatic recombination of the *JJAZ1* gene. Am J Hum Genet 75:410-423.

Kehrer-Sawatzki H, Cooper DN. 2008. Mosaicism in sporadic neurofibromatosis type 1: variations on a theme common to other hereditary cancer syndromes? J Med Genet 45:622-631.

Kirchhoff M, Bisgaard AM, Duno M, Hansen FJ, Schwartz M. 2007. A 17q21.31 microduplication, reciprocal to the newly described 17q21.31 microdeletion, in a girl with severe psychomotor developmental delay and dysmorphic craniofacial features. Eur J Med Genet 50:256-263.

Kluwe L, Siebert R, Gesk S, Friedrich RE, Tinschert S, Kehrer-Sawatzki H, Mautner V-F. 2004. Screening of 500 unselected neurofibromatosis 1 patients for deletions of the *NF1* gene. Hum Mutat 23:111-116.

Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R, Kooy RF, Reyniers E, Romano C, Fichera M, Schinzel A, Baumer A, Anderlid BM, Schoumans J, Knoers NV, van Kessel AG, Sistermans EA, Veltman JA, Brunner HG, de Vries BB. 2006. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nat Genet 38:999-1001.

Lazaro C, Gaona A, Ainsworth P, Tenconi R, Vidaud D, Kruyer H, Ars E, Volpini V, Estivill X. 1996. Sex differences in mutational rate and mutational mechanism in the *NF1* gene in neurofibromatosis type 1 patients. Hum Genet 98:696-699.

López-Correa C, Brems H, Lazaro C, Marynen P, Legius E. 2000. Unequal meiotic crossover: a frequent cause of NF1 microdeletions. Am J Hum Genet 66:1969-1974.

López-Correa C, Dorschner M, Brems H, Lazaro C, Clementi M, Upadhyaya M, Dooijes D, Moog U, Kehrer-Sawatzki H, Rutkowski JL, Fryns JP, Marynen P, Stephens K, Legius E. 2001. Recombination hotspot in NF1 microdeletion patients. Hum Mol Genet 10:1387-1392.

Love JJ, Li X, Case DA, Giese K, Grosschedl R, Wright PE. 1995. Structural basis for DNA bending by the architectural transcription factor LEF-1. Nature 376:791-795.

Marques-Bonet T, Eichler EE. 2009. The evolution of human segmental duplications and the core duplicon hypothesis. Cold Spring Harb Symp Quant Biol. 2009 Aug 28.[Epub ahead of print]

Mautner V-F, Kluwe L, Friedrich RE, Roehl AC, Bammert S, Högel J, Spöri H, Cooper DN, Kehrer-Sawatzki H. 2010. Clinical characterization of 29 neurofibromatosis type-1 patients with molecularly ascertained 1.4 Mb type-1 *NF1* deletions**.** J Med Genet, in press

Mensink KA, Ketterling RP, Flynn HC, Knudson RA, Lindor NM, Heese BA, Spinner RJ, Babovic-Vuksanovic D. 2006. Connective tissue dysplasia in five new patients with NF1 microdeletions: further expansion of phenotype and review of the literature. J Med Genet 43:e8.

Myers AJ, Kaleem M, Marlowe L, Pittman AM, Lees AJ, Fung HC, Duckworth J, Leung D, Gibson A, Morris CM, de Silva R, Hardy J. 2005. The H1c haplotype at the *MAPT* locus is associated with Alzheimer's disease. Hum Mol Genet 14:2399-2404.

Myers S, Freeman C, Auton A, Donnelly P, McVean G. 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. Nat Genet 40:1124-1129.

Petek E, Jenne DE, Smolle J, Binder B, Lasinger W, Windpassinger C, Wagner K, Kroisel PM, Kehrer-Sawatzki H. 2003. Mitotic recombination mediated by the *JJAZF1* (KIAA0160) gene causing somatic mosaicism and a new type of constitutional NF1 microdeletion in two children of a mosaic female with only few manifestations. J Med Genet 40:520-525.

Popesco MC, Maclaren EJ, Hopkins J, Dumas L, Cox M, Meltesen L, McGavran L, Wyckoff GJ, Sikela JM. 2006. Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. Science 313:1304-1307.

Rasmussen SA, Colman SD, Ho VT, Abernathy CR, Arn PH, Weiss L, Schwartz C, Saul RA, Wallace MR. 1998. Constitutional and mosaic large *NF1* gene deletions in neurofibromatosis type 1. J Med Genet 35:68-71.

Roehl AC, Cooper DN, Kluwe L, Helbrich A, Wimmer K, Högel J, Mautner V-F, Kehrer-Sawatzki H. 2010. Extended runs of homozygosity at 17q11.2: An association with type-2 *NF1* deletions? Hum Mutat 31:325-334.

Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC, Fitzpatrick CA, Segraves R, Richmond TA, Guiver C, Albertson DG, Pinkel D, Eis PS, Schwartz S, Knight SJ, Eichler EE. 2006. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet 38:1038-1042.

Shaw CJ, Lupski JR. 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet 13:R57– R64.

Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, Curley R, Cumming S, Dunn C, Kalaitzopoulos D, Porter K, Prigmore E, Krepischi-Santos AC, Varela MC, Koiffmann CP, Lees AJ, Rosenberg C, Firth HV, de Silva R, Carter NP. 2006. Microdeletion encompassing *MAPT* at chromosome 17q21.3 is associated with developmental delay and learning disability. Nat Genet 38:1032-1037.

Deleted: In press [doi10.1002/humu.21191]

Spiegel M, Oexle K, Horn D, Windt E, Buske A, Albrecht B, Prott EC, Seemanová E, Seidel J, Rosenbaum T, Jenne D, Kehrer-Sawatzki H, Tinschert S. 2005. Childhood overgrowth in patients with common NF1 microdeletions. Eur J Hum Genet 13:883-888.

Stankiewicz P, Lupski JR. 2006. The genomic basis of disease, mechanisms and assays for genomic disorders. Genome Dyn 1:1-16.

Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, Desnica N, Hicks A, Gylfason A, Gudbjartsson DF, Jonsdottir GM, Sainz J, Agnarsson K, Birgisdottir B, Ghosh S, Olafsdottir A, Cazier JB, Kristjansson K, Frigge ML, Thorgeirsson TE, Gulcher JR, Kong A, Stefansson K. 2005. A common inversion under selection in Europeans. Nat Genet 37:129-137.

Steinmann K, Cooper DN, Kluwe L, Chuzhanova NA, Senger C, Serra E, Lazaro C, Gilaberte M, Wimmer K, Mautner VF, Kehrer-Sawatzki H. 2007. Type 2 *NF1* deletions are highly unusual by virtue of the absence of nonallelic homologous recombination hotspots and an apparent preference for female mitotic recombination. Am J Hum Genet 81:1201-1220.

Steinmann K, Kluwe L, Cooper DN, Brems H, De Raedt T, Legius E, Mautner VF, Kehrer-Sawatzki H. 2008. Copy number variations in the *NF1* gene region are infrequent and do not predispose to recurrent type-1 deletions. Eur J Hum Genet 16:572-580.

Sundar PD, Yu CE, Sieh W, Steinbart E, Garruto RM, Oyanagi K, Craig UK, Bird TD, Wijsman EM, Galasko DR, Schellenberg GD. 2007. Two sites in the *MAPT* region confer genetic risk for Guam ALS/PDC and dementia. Hum Mol Genet 16:295-306.

Tan TY, Aftimos S, Worgan L, Susman R, Wilson M, Ghedia S, Kirk EP, Love D, Ronan A, Darmanian A, Slavotinek A, Hogue J, Moeschler JB, Ozmore J, Widmer R, Bruno D, Savarirayan R, Peters G. 2009. Phenotypic expansion and further characterisation of the 17q21.31 microdeletion syndrome. J Med Genet 46:480-489.

Upadhyaya M, Roberts SH, Maynard J, Sorour E, Thompson PW, Vaughan M, Wilkie AO, Hughes HE. 1996. A cytogenetic deletion, del(17)(q11.22q21.1), in a patient with sporadic neurofibromatosis type 1 (NF1) associated with dysmorphism and developmental delay. J Med Genet 33:148-152.

Venturin M, Guarnieri P, Natacci F, Stabile M, Tenconi R, Clementi M, Hernandez C, Thompson P, Upadhyaya M, Larizza L, Riva P. 2004. Mental retardation and cardiovascular malformations in NF1 microdeleted patients point to candidate genes in 17q11.2. J Med Genet 41:35-41.

Wells RD. 2007. Non-B DNA conformations, mutagenesis, and diseases. Trends Biochem Sci 32:271–278.

Wang G, Vasquez KM. 2006. Non-B DNA structure-induced genetic instability. Mutat Res 598:103–119.

Zody MC, Garber M, Adams DJ, Sharpe T, Harrow J, Lupski JR, Nicholson C, Searle SM, Wilming L, Young SK, Abouelleil A, Allen NR, Bi W, Bloom T, Borowsky ML, Bugalter BE, Butler J, Chang JL, Chen CK, Cook A, Corum B, Cuomo CA, de Jong PJ, DeCaprio D, Dewar K, FitzGerald M, Gilbert J, Gibson R, Gnerre S, Goldstein S, Grafham DV, Grocock

R, Hafez N, Hagopian DS, Hart E, Norman CH, Humphray S, Jaffe DB, Jones M, Kamal M, Khodiyar VK, LaButti K, Laird G, Lehoczky J, Liu X, Lokyitsang T, Loveland J, Lui A, Macdonald P, Major JE, Matthews L, Mauceli E, McCarroll SA, Mihalev AH, Mudge J, Nguyen C, Nicol R, O'Leary SB, Osoegawa K, Schwartz DC, Shaw-Smith C, Stankiewicz P, Steward C, Swarbreck D, Venkataraman V, Whittaker CA, Yang X, Zimmer AR, Bradley A, Hubbard T, Birren BW, Rogers J, Lander ES, Nusbaum C. 2006a. DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. Nature 440:1045-1049.

Zody MC, Garber M, Sharpe T, Young SK, Rowen L, O'Neill K, Whittaker CA, Kamal M, Chang JL, Cuomo CA, Dewar K, FitzGerald MG, Kodira CD, Madan A, Qin S, Yang X, Abbasi N, Abouelleil A, Arachchi HM, Baradarani L, Birditt B, Bloom S, Bloom T, Borowsky ML, Burke J, Butler J, Cook A, DeArellano K, DeCaprio D, Dorris L 3rd, Dors M, Eichler EE, Engels R, Fahey J, Fleetwood P, Friedman C, Gearin G, Hall JL, Hensley G, Johnson E, Jones C, Kamat A, Kaur A, Locke DP, Madan A, Munson G, Jaffe DB, Lui A, Macdonald P, Mauceli E, Naylor JW, Nesbitt R, Nicol R, O'Leary SB, Ratcliffe A, Rounsley S, She X, Sneddon KM, Stewart S, Sougnez C, Stone SM, Topham K, Vincent D, Wang S, Zimmer AR, Birren BW, Hood L, Lander ES, Nusbaum C. 2006b. Analysis of the DNA sequence and duplication history of human chromosome 15. Nature 440:671-675.

Zody MC, Jiang Z, Fung HC, Antonacci F, Hillier LW, Cardone MF, Graves TA, Kidd JM, Cheng Z, Abouelleil A, Chen L, Wallis J, Glasscock J, Wilson RK, Reily AD, Duckworth J, Ventura M, Hardy J, Warren WC, Eichler EE. 2008. Evolutionary toggling of the *MAPT* 17q21.31 inversion region. Nat Genet 40:1076-1083.

Legends to figures

Figure 1: (A) Schema of the *NF1* gene region. **A:** The relative genomic positions of NF1-REPs A, B and C are indicated as well as the positions of functional genes and the *SUZ12* pseudogene (*SUZ12P*) located in this region of 17q11.2. The vertical arrows indicate the locations of the MLPA probes used in this analysis. **B:** The extent of the common 1.4 Mb type-1 *NF1* deletions, the type-2 deletions with breakpoints located within *SUZ12/SUZ12P* sequences and the extent of the ~1.0 Mb ('type-3') deletions observed in patients TOP, 2176 and Z41/03, are given in relation to the map of the extended *NF1* gene region shown above.

Figure 2: Positions of the paralogous *LRRC37B* sequences within the NF1-REPs and mapping of the deletion breakpoint regions in patients TOP and 2176 by PCR. The numbering of the *LRRC37B* exons (and pseudogene exons) is indicated together with the designation of the different paralogous sequences amplified by PCR (B7/7, B5/6, Int5A etc). These paralogous sequences are distinguishable by reference to paralogous sequence variants (PSVs). To identify the deletion breakpoint regions, PCR was performed with DNAs from somatic cell hybrids of patients TOP and 2176 containing only the chromosomes 17 harbouring the deletions of the respective patients. Sequence analysis and evaluation of PSVs indicated the NF1-REP origin of the PCR products amplified. The PCR-fragments given in bold were located within the deleted regions. **A:** The breakpoints of the deletion in patient TOP were assigned to the region within NF1-REPs B and C, respectively, between the fragments B9/10 and Int3C. **B:** The deletion breakpoints in patient 2176 were assigned to the region between PCR-fragments Int2A and Ex1 in NF1-REPs B and C, respectively. Sequence analysis of PCR-fragment Int1B indicated that it is breakpoint-spanning.

Deleted: Figure 2: Schema of the modular sub-structure of NF1-REPs A, B and C comprising different paralogous sequence blocks at 17q11.2. The location and transcriptional location of the *NF1* gene is given. Pseudogene fragments of the *LRRC37B* gene (marked in red) are present in all three NF1-REPs. The functional *LRRC37B* gene, containing 12 exons, lies within NF1-REP C. Paralogous sequences with homology to 19p13.12, marked in blue, comprise portions of NF1-REPs A and C. NF1-REP A also contains a pseudogene fragment encompassing exon 1 and partial intron 1 of the latrophilin-1 precursor gene (*LPHN1*) from 19p13.12. NF1-REPs A and B also include pseudogene fragments of the *SMURF2* (SMAD-specific E3 ubiquitin-protein ligase 2) gene, marked in green. The extent of the respective paralogous sequence blocks are given according to the base-pair numbering of chromosome 17 (hg18, NCBI 36, Ensembl database version 54.36p). The horizontal arrows below the paralogous sequence blocks indicate their transcriptional orientation or their orientation in relation to the ancestral sequence. The positions of the paralogous recombination sites 1 and 2 (PRS1, PRS2) within NF1-REPs A and C are also indicated.

Deleted: 3

Deleted: ue-specific

Deleted: red

Deleted: , whereas those given in black were not deleted.

Deleted: (marked in bold)

Figure 3: **A**: *LRRC37B* sequences with the NF1-REPs and localization of the deletion breakpoint regions in patients TOP, 2176 and Z41/03 within NF1-REPs B and C. The base-pair numbering of chromosome 17 is in accordance with hg18, NCBI 36, Ensembl database version 54.36p. The arrows represent the transcriptional orientation of the *LRRC37B* gene and its pseudogenes; the corresponding exon numbers are also indicated. In patient TOP, the breakpoints were assigned to *LRRC37B-P* intron 3. In patients 2176 and Z41/03, the breakpoints were identified within *LRRC37B-P* intron 1. **B:** Schematic diagram of the positions of the deletion breakpoint regions of patient TOP within intron 1, and patients 2176 and Z41/03 within intron 3. The latter breakpoint regions are directly adjacent to one another, separated by only 2 bp. The numbers indicate the positions of the breakpoint regions within NF1-REP B and those in parentheses within NF1-REP C.

Deleted: 4

Deleted: ¶
¶
Figure 5: Repeats capable of non-B DNA structure formation within the recombination regions (from patients Z41/03, 2176 and TOP) harbouring the deletion breakpoints and their flanking regions. The recombination regions are highlighted in grey, whereas the centromeric flanking regions are marked in yellow and the telomeric flanking sequences are highlighted in turquoise. Sequences capable of non-B DNA structure formation include long inexact polypyrimidine tracts, direct repeats, inverted repeats and symmetric elements. Long inexact polypyrimidine tracts, composed of at least one continuous tract ≥10bp, are underlined. Direct repeats are highlighted in pink. Inverted repeats and symmetric elements are shown in blue and red letters, respectively.

Table 1: Identification of the deletion breakpoint regions in NF1 patients 2176 and TOP by PCR direct sequencing from different regions of the functional *LRRC37B* gene and its *LRRC37B* pseudogene fragments located within the NF1-REPs A, B, and C.

PCR product	Genomic position of the respective fragment in				Origin of the PCR product amplified from the deletion-bearing chromosome 17 in patient[a]	
	NF-REP A *LRRC37B-P*	NF1-REP B *LRRC37B-P*	NF1-REP C *LRRC37B*	NF1-REP C *LRRC37B-P*	2176	TOP
B7/8	25,985,319-25,985,730	—	27,401,316-27,401,727	—	NF1-REP A	NF1-REP A
B5/6	26,019,880-26,020,251	—	—	27,435,982-27,436,354	NF1-REP A	NF1-REP A
Int5A	26,022,532-26,022,910	26,389,046-26,389,425	27,382,883-27,383,261	27,438,621-27,438,999	NF1-REPs A and B	NF1-REPs A and B
Int3B	—	26,390,554-26,390,883	27,381,425-27,381,754	27,440,128-27,440,457	NF1-REP B	NF1-REP B
B9/10	—	26,390,646-26,390,973	27,381,335-27,381,646	27,440,236-27,440,543	NF1-REP B	NF1-REP B
Int3C	—	26,391,580-26,392,274	27,380,035-27,380,729	27,441,149-27,441,846	NF1-REP B	NF1-REP C (*LRRC37B-P*)
Int3A	—	26,393,049-26,393,427	27,378,888-27,379,266	27,442,615-27,442,993	NF1-REP B	NF1-REP C (*LRRC37B-P*)
Int2A	—	26,396,250-26,396,488	27,375,780-27,376,018	27,445,890-27,446,128	NF1-REP B	NF1-REP C (*LRRC37B-P*)
Int1B	—	26,397,173-26,397,698	27,374,570-27,375,095	27,446,814-27,447,339	NF1-REPs B and C (breakpoint-spanning)	NF1-REP C (*LRRC37B-P*)
Ex1	—	26,401,114-26,401,719	27,372,713-27,373,324	—	deleted	deleted
B3/4	—	—	27,354,170-27,354,473	27,447,960-27,448,268	NF1-REP C (*LRRC37B-P*)	NF1-REP C (*LRRC37B-P*)

a: PCRs were performed with somatic cell hybrids containing the deletion-bearing chromosomes 17 from the respective patients.

—: Sequences paralogous to the PCR fragment were absent from the respective NF1-REPs.

Table 2: Distance between the centromeric boundary of PRS2 spanning 2.9 kb (PRS1, 3.5 kb) within NF1-REP C and the beginning of the deletion breakpoint regions of the patients with ~1.0 Mb deletions ('type-3') in NF1-REP C

Distance (bp) between the centromeric boundary of PRS2 (PRS1) within NF1-REP C and the beginning of the deletion breakpoint region of patient		
TOP	2176	Z41/03
7,962 (27,170)	13,661 (32,869)	13,842 (33,050)

The PRS1 region is located between 27,413,822 and 27,416,683 whereas PRS2 is located between 27,433,020 and 27,436,542. The telomeric deletion breakpoint region of patient TOP lies between 27,440,992 and 27,441,266, of patient 2176 between 27,446,691-27,446,870 and of patient Z41/03 between 27,446,872-27,446,980 according to the base-pair numbering of chromosome 17 (hg18, NCBI build 36.1).

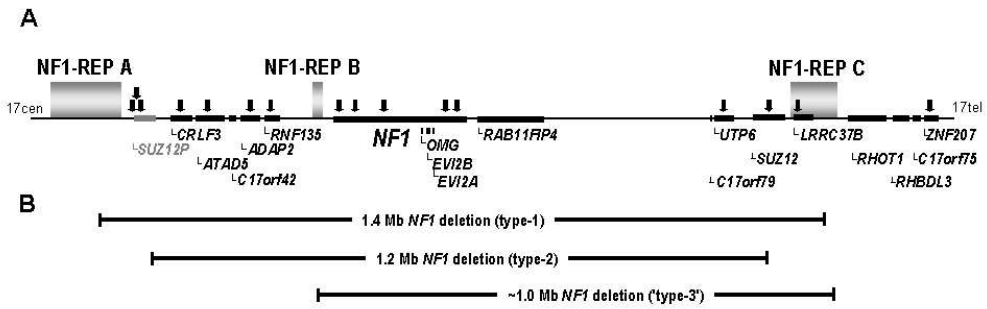


Figure 1

81x60mm (300 x 300 DPI)

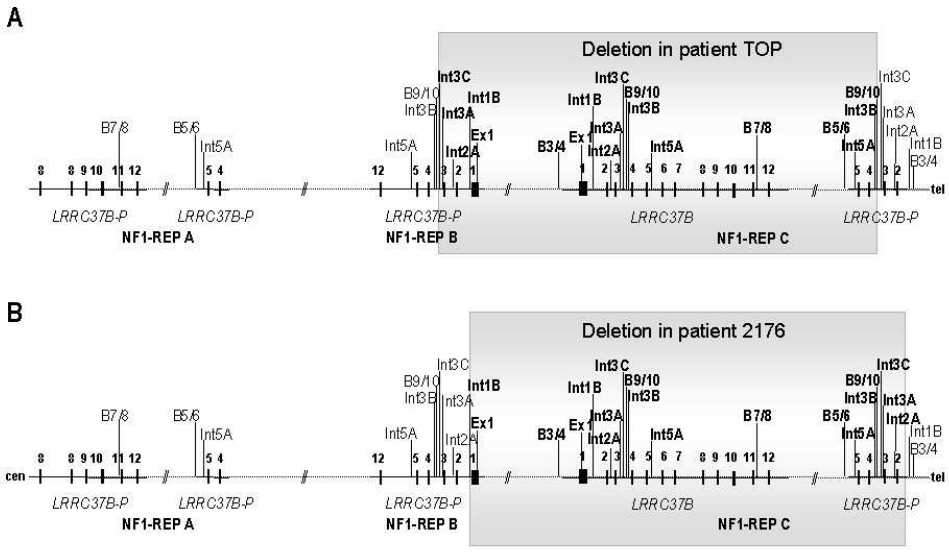


Figure 2

81x60mm (300 x 300 DPI)

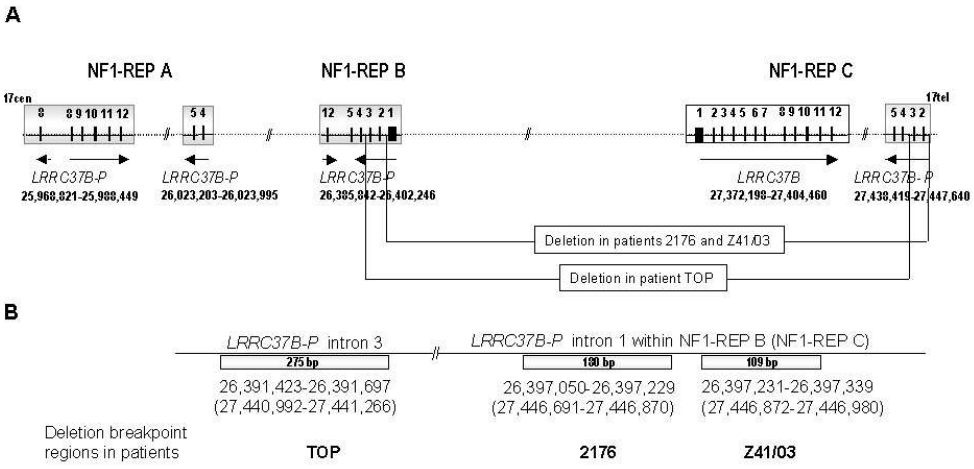


Figure 3

81x60mm (300 x 300 DPI)

Supplementary Table S1: Identification of the deletion breakpoint regions in patients TOP and 2176 by PCR analysis of microsatellite markers and single copy sequences in 17q11.2 using somatic cell hybrids containing only the deleted chromosomes 17 of the respective patients. The precise positions of the MLPA probes and the corresponding MLPA results obtained using genomic DNA of both patients are given.

Designation of markers, PCR products and MLPA probes	Position on chromosome 17[a]	Extent of the deletion according to PCR and MLPA analysis
D17S975	25,124,113-25,124,371	not deleted
D17S1532	25,708,122-25,708,291	not deleted
RNF135 (MLPA)	26,335,816-26,335,835	not deleted
hks332/hks331	26,339,038-26,339,386	not deleted
hks329/hks330	26,429,386-26,429,775	deleted
NF1 Exon 1 (MLPA)	26,445,726-26,445,745	deleted
D17S1307	26,497,479-26,497,687	deleted
D17S1849	26,594,230-26,594,552	deleted
NF1.PCR3	26,614,084-26,614,274	deleted
D17S2237	26,641,851-26,642,254	deleted
D17S1166	26,673,142-26,673,342	deleted
D17S1972	26,725,380-26,725,585	deleted
D17S1800	26,960,942-26,961,211	deleted
D17S1666	27,203,022-27,203,239	deleted
hks405/hks403	27,347,427-27,347,984	deleted
hks17s4/hks17s2	27,350,506-27,351,290	deleted
hks103/hks104	27,369,524-27,369,916	deleted
hks68/hks69	27,386,045-27,386,567	deleted
hks109/hkstg1	27,359,461-27,359,798	deleted
LRRC37B probe 3787-L3296 (MLPA)	27,372,684-27,372,703	deleted
ZNF207 probe 9637-L9949 (MLPA)	27,717,868-27,717,887	not deleted
D17S1416E	27,576,410-27,576,504	not deleted
D17S1927	27,675,519-27,675,669	not deleted

a: According to the sequence of chromosome 17 in hg18 (March 2006), NCBI build 36.1.

Supplementary Table S2: Primers used for the PCRs to amplify *LRRC37B* sequences in order to identify the deletion breakpoint regions in patients 2176 and TOP

PCR primer designation	**Primer sequences 5' → 3'**
Ex1f	CCTAGTTCAGCTTCCTCGCCTC
Ex1r	TCCACATCTGCAGGTTTGACTGTAA
B3	CCCCCATTTTTAAAGAGAAGAA
B4	TTGCAGTGATTTCATACCAACA
B7	CCAACGTCACTTTCCCACTT
B8	CACCTTACACACTGGAGAAGGA
B9	GAGCTCCTAACCCTGCCTGT
B10	GGGGTAACTTCCTTTCAAATGC
Int5Af	CGACCTGCTAATTTTTGTATCG
Int5Ar	TCCCAGAACTTTAGGAGCTGAG
Int3Bf	AAGTTATCAGTGTTGACTGGTATGAA
Int3Br	GGTGTATAAATGTGAGGCTATTCG
Int3Af	TAGAAGAAGATGGTATGTAAACCCTTA
Int3Ar	TTTCATTTTCCAGAATTCTCAGTG
Int2Af	AGGGGAAATATCTTTTCTCAGAGG
Int2Ar	TTTTATTTATCCTGTTCATTCTTTGC
Int1Bf	TGTAGCTTTCCCTGATCTCCA
Int1Br	ACAGCAGCCTGTCACTCTCC
Int3Cf	TGTCCGATGCTACACAATGG
Int3Cr	CTGCAGAGGCCATGAAGTAA
B5	CAGAGTTGACCCGAAAGAGC
B6	TGGCCTCTGCTCTTCCTTTA

Supplementary Table S3: Primers used to amplify breakpoint-spanning PCR fragments from somatic cell hybrids containing only the deletion-bearing chromosome 17 from patient TOP (primers B9 and Int3Cr) and from genomic DNA of patient 2176 (primers Int2C and Int2r). Primers Int1Af and Int1Ar were used as nested primers to fully sequence PCR-fragment Int2C/Int2r. Primers Int3Df and Int3Dr were used to fully sequence PCR-fragment B9/Int3Cr. Primers Int2C and Int2r were also used to amplify a breakpoint-spanning fragment from genomic DNA of patient Z41/03, which was fully sequenced using primers Int1Af and Int1Ar.

Primer	Sequence 5'→ 3'	Position[a] in	
		NF1-REP B	NF1-REP C
Int2C	AAATTGCCTCTTAGAGAGAGCTTC	26,396,165-26,396,188	—
Int2r	TTGCAGTGATTTCATACCAACA	—	27,448,247-27,448,268
Int1Af	CCCTATGCCAGGTAACACCA	26,396,821-26,396,840	27,446,461-27,446,480
Int1Ar	CTTTGTGGAAGCCAGTTTCAG	—	27,447,725-27,447,745
B9	GAGCCCCTAACCCTGCCTGT	26,390,662-26,390,681	27,440,236-27,440,255
Int3Cr	CTGCAGAGGCCATGAAGTAA	26,392,255-26,392,274	27,441,827-27,441,846
Int3Df	TTGTCTGGCTGATGTCTCCA	26,391,150-26,391,169	27,440,720-27,440,739
Int3Dr	TTTAAACATGGTGGGCAATG	26,391,806-26,391,825	27,441,375-27,441,394

[a]: according to the base-pair numbering of chromosome 17 of hg18, NCBI Build 36.1.
—: no homology to sequences within this NF1-REP.

Supplementary Table S4: Results of the MLPA analysis in NF1 patients 2176, Z41/03 and TOP

MLPA probe	Localisation of probe	Probe position on chromosome 17 [a]	MLPA results of patients		
			2176[b]	Z41/03[c]	TOP[b]
TRAF4 probe 9176-L9350	17q11.2 *TRAF4* Exon 2	24,098,420-24,098,439	1.04	0.98	0.95
TRAF4 probe 8620-L8632	17q11.2 *TRAF4* Exon 4	24,099,181-24,099,200	0.97	1.02	0.96
SSH2 probe 9635-L9920	17q11.2 *SSH2* Exon 14	24,987,708-24,987,727	0.98	0.97	1.02
SSH2 probe 9634-L9919	17q11.2 *SSH2* Exon 4	25,046,623-25,046,642	0.99	0.99	1.04
BLMH probe 9627-L9912	17q11.2 *BLMH* Exon 9	25,623,740-25,623,759	0.99	1.03	0.98
BLMH probe 9626-L9911	17q11.2 *BLMH* Exon 2	25,642,606-25,642,625	0.98	1.01	1.05
CPD probe 9628-L9913	17q11.2 *CPD* Exon 11	25,795,038-25,795,057	1.01	0.99	0.97
CPD probe 9629-L9914	17q11.2 *CPD* Exon 20	25,813,548-25,813,567	1.01	1	0.98
SUZ12P probe 11798-L12590	17q11.2 *SUZ12P* proximal	26,082,519-26,082,538	0.97	-	0.96
SUZ12P probe 11800-L12591	17q11.2 *SUZ12P* Exon 1	26,082,967-26,082,986	1.09	-	0.99
SUZ12P probe 11801-L12592	17q11.2 *SUZ12P* Exon 6	26,109,273-26,109,293	0.96	-	1.09
ATAD5 probe 3781-L3290	17q11.2 *ATAD5*	26,186,172-26,186,191	0.96	0.97	1.02
CRLF3 probe 3780-L3289	17q11.2 *CRLF3*	26,148,508-26,148,527	0.99	1	1.01
CENTA2 probe 3782-L3291	17q11.2 *CENTA2*	26,278,001-26,278,020	1	1	0.98
RNF135 probe 3783-L3292	17q11.2 *RNF135*	26,335,816-26,335,835	0.99	1	0.98
NF1 probe 2491-L1922	17q11.2 *NF1* Exon 1	26,445,726-26,445,745	0.56	0.55	0.48
NF1 probe 2507-L1938	17q11.2 *NF1* Exon 12B	26,576,330-26,576,349	0.55	0.5	0.51
NF1 probe 2512-L1943	17q11.2 *NF1* Exon 23	26,600,151-26,600,170	0.56	0.54	0.52
NF1 probe 2525-L1956	17q11.2 *NF1* Exon 40	26,700,280-26,700,299	0.55	0.54	0.51
NF1 probe 5220-L3309	17q11.2 *NF1* Exon 48	26,711,704-26,711,723	0.54	0.53	0.51
UTP6 probe 3785-L3294	17q11.2 *UTP6*	27,226,463-27,226,482	0.52	0.54	0.52
SUZ12 probe 3786-L3295	17q11.2 *SUZ12*	27,339,525-27,339,544	0.54	0.54	0.54
LRRC37B probe 3787-L3296	17q11.2 *LRRC37B*	27,372,684-27,372,703	0.55	0.51	0.56
ZNF207 probe 9637-L9949	17q11.2 *ZNF207*	27,717,868-27,717,887	0.99	0.98	1.01
PSMD11 probe 9632-L9917	17q11.2 *PSMD11* Exon 2	27,798,094-27,798,113	0.95	1.03	0.96
PSMD11 probe 9633-L9918	17q11.2 *PSMD11* Exon 6	27,820,186-27,820,205	0.98	0.97	0.92
MYO1D probe 9631-L9916	17q11.2 *MYO1D* Exon 7	28,118,825-28,118,844	1.05	1.06	0.99
MYO1D probe 9630-L9915	17q11.2 *MYO1D* Exon 2	28,131,767-28,131,786	1.03	1.03	0.9
Reference probe 2876-L2343	1p32		1.02	1	0.91
Reference probe 4683-L4061	1p36		1.01	0.98	1.02
Reference probe 1542-L0985	5q22		1.04	101	1
Reference probe 0797-L0463	5q31		0.97	1	0.97
Reference probe 0662-L0158	6p21		1.03	1.02	1.02
Reference probe 3205-L2566	6q26		1	0.98	0.96
Reference probe 3532-L2898	7q36		0.98	0.98	0.95
Reference probe 3007-L2447	9q34		0.95	0.99	1.01
Reference probe 1232-L0780	10p14		1.01	1,01	1.05
Reference probe 1463-L0928	17p12		0.99	0.97	0.91
Reference probe 1325-L7456	17p13.3		0.96	1.01	1.03

[a]: Position according to the sequence numbering of 17 (hg18, NCBI Build 36.1, Ensembl version 44.36f)
[b]: MLPA analysis performed with P122-C1 MLPA probemix Lot 0608
[c]: MLPA analysis performed with P122-B MLPA probemix Lot 1007. The *SUZ12P* probes 11798-L12590, 11800-L12591 and 11801-L12592 were not included in this version of the MLPA-kit.

Supplementary Table S5: BLAT results of all *LRRC37B* exons against the human reference sequence (hg18; NCBI 36.1) in order to identify sequences with high similarity to these exons. In addition to the *LRRC37B* pseudogene exons in the NF1-REPs A, B, and C, sequence similarites to other members of the *LRRC37* gene family were observed.

LRRC37B exons	Characterization of the sequences identified by BLAT						
	Position on chromosome 17		Strand	Exon size (bp)	Homo-logy	Designation of the gene/ mRNA[a]	Chromosomal position
Exon 1	27372156	27373957	+	1802	100.0%	*LRRC37B*	NF1-REP C
	26398309	26402281	-	3973	94.9%	*LRRC37B-P*	NF1-REP B
	60321229	60323911	-	2683	90.0%	*LRRC37A3*	17q24.1
	40980876	40983562	-	2687	90.1%	AK124512	17q21.31
	41945317	41948003	+	2687	90.0%	*LRRC37A2*	17q21.31
	39355670	39360370	-	4701	89.7%	–	17q21.31
	42449443	42452165	+	2723	89.7%	BC006271, AK303650	17q21.32
	41728199	41730885	+	2687	90.6%	*LRRC37A*	17q21.31
	63657005	63660287	-	3283	88.4%	CR595591	17q24.2
	31258281	31268175	-	9895	90.3%	–	17q12
	34439148	34441568	+	2421	93.8%	AK125814, BC111725	17q12
Exon 2	27375843	27375914	+	72	100.0%	*LRRC37B*	NF1-REP C
	27445994	27446065	-	72	98.7%	*LRRC37B-P*	NF1-REP C
	26396354	26396425	-	72	98.7%	*LRRC37B-P*	NF1-REP B
	60319109	60319180	-	72	97.3%	*LRRC37A3*	17q24.1
	41950044	41950115	+	72	97.3%	*LRRC37A2*	17q21.31
	40978754	40978825	-	72	95.9%	AK124512	17q21.31
	41732937	41733008	+	72	95.9%	*LRRC37A*	17q21.31
	42454215	42454286	+	72	94.5%	BC006271, AK303650	17q21.32
	63655188	63655253	-	66	95.5%	CR595591	17q24.2
	34443949	34444020	+	72	91.7%	AK125814, BC111725	17q12
Exon 3	27378901	27378972	+	72	100.0%	*LRRC37B*	NF1-REP C
	26393343	26393414	-	72	98.7%	*LRRC37B-P*	NF1-REP B
	27442909	27442980	-	72	97.3%	*LRRC37B-P*	NF1-REP C
	34449247	34449319	+	73	93.1%	AK125814, BC111725	17q12
	60316261	60316332	-	72	91.7%	*LRRC37A3*	17q24.1
	40975913	40975984	-	72	91.7%	AK124512	17q21.31
	41735781	41735852	+	72	91.7%	*LRRC37A*	17q21.31
	41952894	41952965	+	72	91.7%	*LRRC37A2*	17q21.31
	42457045	42457116	+	72	91.7%	BC006271, BC006271	17q21.32
Exon 4	27381788	27381859	+	72	100.0%	*LRRC37B*	NF1-REP C
	27440023	27440094	-	72	98.7%	*LRRC37B-P*	NF1-REP C
	26390450	26390520	-	71	97.2%	*LRRC37B-P*	NF1-REP B
	26023935	26023995	-	61	98.4%	*LRRC37B-P*	NF1-REP A
	63653262	63653333	-	72	90.3%	CR595591, AK130509	17q24.2
	34452255	34452322	+	68	91.2%	AK125814, BC111725	17q12
	60313403	60313466	-	64	92.2%	*LRRC37A3*	17q24.1
	40973049	40973112	-	64	92.2%	AK124512	17q21.31
	41738651	41738714	+	64	92.2%	*LRRC37A*	17q21.31
	41955766	41955829	+	64	92.2%	*LRRC37A2*	17q21.31
	42459922	42459985	+	64	92.2%	BC006271, AK303650	17q21.32
Exon 5	27382510	27382590	+	81	100.0%	*LRRC37B*	NF1-REP C
	27439292	27439372	-	81	98.8%	*LRRC37B-P*	NF1-REP C
	26023203	26023283	-	81	98.8%	*LRRC37B-P*	NF1-REP A
	26389718	26389798	-	81	98.8%	*LRRC37B-P*	NF1-REP B
	34452988	34453058	+	71	91.6%	AK125814, BC111725	17q12
	63652528	63652598	-	71	90.2%	CR595591	17q24.2
	60312672	60312736	-	65	93.9%	*LRRC37A3*	17q24.1
	40972318	40972382	-	65	93.9%	AK124512	17q21.31
	41739381	41739445	+	65	93.9%	*LRRC37A*	17q21.31

	41956496	41956560	+	65	93.9%	*LRRC37A2*	17q21.31
	42460655	42460719	+	65	92.4%	BC006271	17q21.32
Exon 6	27386042	27386113	+	72	100.0%	*LRRC37B*	NF1-REP C
	60295675	60295746	-	72	91.7%	*LRRC37A3*	17q24.1
	41680569	41680640	+	72	90.3%	–	17q21.31
	41755453	41755524	+	72	90.3%	*LRRC37A*	17q21.31
	41972858	41972929	+	72	90.3%	*LRRC37A2*	17q21.31
	63649020	63649091	-	72	88.9%	CR595591	17q24.2
Exon 7	27386697	27386771	+	75	100.0%	*LRRC37B*	NF1-REP C
	60295042	60295116	-	75	92.0%	*LRRC37A3*	17q24.1
	41756084	41756158	+	75	92.0%	*LRRC37A*	17q21.31
	41973489	41973563	+	75	92.0%	*LRRC37A2*	17q21.31
	63648356	63648430	-	75	90.7%	CR595591	17q24.2
	41030913	41030987	-	75	90.7%	–	17q21.31
	41681228	41681302	+	75	90.2%	–	17q21.31
Exon 8	27396832	27396950	+	119	100.0%	*LRRC37B*	NF1-REP C
	25968821	25968935	-	115	94.0%	*LRRC37B-P*	NF1-REP A
	25980824	25980938	+	115	94.0%	*LRRC37B-P*	NF1-REP A
	63637940	63638058	-	119	96.7%	CR595591	17q24.2
	40954456	40954574	-	119	95.8%	BT007332, AK124512	17q21.31
	41761522	41761640	+	119	95.8%	*LRRC37A*	17q21.31
	41978937	41979055	+	119	95.8%	*LRRC37A2*	17q21.31
	42479511	42479629	+	119	95.8%	BC006271	17q21.32
	60289480	60289598	-	119	95.0%	*LRRC37A3*	17q24.1
Exon 9	27398893	27399033	+	141	100.0%	*LRRC37B*	NF1-REP C
	25982899	25983038	+	140	98.6%	*LRRC37B-P*	NF1-REP A
	60287414	60287553	-	140	92.9%	*LRRC37A3*	17q24.1
	40948389	40948528	-	140	92.2%	BT007332, BC088374, AK124512, AK000982	17q21.31
	40952382	40952521	-	140	92.2%	BT007332, AK124512, BC088374, BC065911, BC064489	17q21.31
	42481567	42481706	+	140	91.5%	BC006271, BC064489, BC065911, BC088374	17q21.32
	63635878	63636016	-	139	91.4%	CR595591, BC053686, BC018754, BC067079, BC073783, BC094861	17q24.2
	41763577	41763716	+	140	90.8%	*LRRC37A*	17q21.31
	41980994	41981133	+	140	90.8%	*LRRC37A2*	17q21.31
Exon 10	27400234	27400526	+	293	100.0%	*LRRC37B*	NF1-REP C
	25984242	25984534	+	293	99.4%	*LRRC37B-P*	NF1-REP A
	41982230	41982518	+	289	85.4%	*LRRC37A2*	17q21.31
	41764813	41765101	+	289	85.4%	*LRRC37A*	17q21.31
	42482790	42483071	+	282	85.8%	BC006271, BC065911, BC088374	17q21.32
	60286038	60286317	-	280	84.1%	*LRRC37A3*	17q24.1
	63634213	63634479	-	267	84.9%	CR595591, AL137380	17q24.2
Exon 11	27401135	27401239	+	105	100.0%	*LRRC37B*	NF1-REP C
	25985138	25985242	+	105	97.2%	*LRRC37B-P*	NF1-REP A
	63633516	63633620	-	105	88.6%	BC045718, CR609563, CR595591	17q24.2
	40950340	40950397	-	58	88.0%	BT007332, AK124512, AK000982	17q21.31
Exon 12	27404398	27404460	+	63	100.0%	*LRRC37B*	NF1-REP C
	25988387	25988449	+	63	100.0%	*LRRC37B-P*	NF1-REP A
	26385842	26385954	+	112		*LRRC37B-P*	NF1-REP B
	63630646	63630671	-	26	92.4%	BC045718, CR609563, CR595591	17q24.2

a: The GenBank accession numbers of the respective mRNAs are indicated.

Supplementary Table S6: Motifs identified in the RRs of NF1 patients TOP, 2176 and Z41/03

Motif [a]	Consensus sequence [b]	Motifs (number) in the breakpoint regions of patient		
		TOP	2176	Z41/03
Vaccinia topoisomerase I consensus cleavage site	YCCTT	TCCTT (1) CCCTT (3)	AAGGA*(1)	CCCTT (1)
Immunoglobulin heavy chain class switch repeat	GAGCT, TGGGG	AGCTC* (1) CCCCA* (1)		
Long polypurine/ polypyrimidine tract	R_{10}	yes		yes
Human Fra(X) breakpoint cluster[c]	CGGCGG			
Murine parvovirus recombination hotspot	CTWTTY	CTATTC (1)		CTATTC (1), CTTTTT (1)
Deletion hotspot consensus sequence	TGRRKM	TGGGTC (1) GCTCCA*(1) TACTCA*(1) TCCCCA*(1) GCTTCA* (1)	GACCCA*(1) TATCCA*(1) TCTCCA*(1)	TCTTCA*(1)
DNA polymerase arrest site	WGGAG	CTCCA* (1) CTCCT* (1)		CTCCT*(1)
DNA polymerase β frameshift hotspot	ACCCWR	CTGGGT* (1)	ACCCAA (1)	
Hamster deletion hotspot sequence	TGGAG	CTCCA* (1)	CTCCA*(1)	
Hamster and human *APRT* deletion hotspot	TTCTTC			TTCTTC (1)
Super hotspot motifs [d]	CCAAR, CCCAG, GGGACA, AGCTG, GGAGAA, SAAGT	CTTGG* (1) CTGGG* (1) TGTCCC* (1) CAGCT* (2) CAAGT (1) GAAGT (1) GGAGAA	CCAAG(1) CTTGG*(1)	
Chinese hamster scaffold attachment site 2 [e]	TTWTWTTWTT		AATAAAATAA*(2)	
TCF-1/LEF-1 [f]	WWCAAAG	TTCAAAG (2)		

a: As summarized by Abeysinghe et al., [2003] except where otherwise specified.
b: R=A/G, Y=C/T, W=A/T, S=G/C, M=A/C, K=G/T, N= any base. The complementary sequence has also been investigated, but the corresponding sequence motifs are not listed.
c: Motif listed in Chuzhanova et al., [2009]
d: Motifs listed in Ball et al., [2006]
e: Motifs listed in Cullen et al., [2002]
f: Motif listed in Love et al., [1995] and Giese et al., [1997]
* Sequences complementary to the corresponding motif.

(A): Deletion breakpoint region and flanking segments in patient TOP

```
ACAGTCATCA TATAAACTCT CCAGAAGTGG CTTGCAAAGA CCAGCACCTC 26391253
CTGGGAAATT ATTGAGGATG CAAATTCTTG GTCCCACTCT AGACCAACTG 26391303
AATCCAATAA CTATGAGGGC GGAGTCCAGA ACTGAGTTCT AACGTGCCCT 26391353
CTCAATGATT GTGATGCAGA TCCATCTTAC AGCACTGCTG TGATAACATG 26391403-26391422
                    T (PSV specfic for NF1-REP C)
GTTCCCCATG ATGGCTCCGC AGCTTGGCAT TCAAAGCCCT AGAAGATCTG
GCTCCAATTT CCTACTCTCC CTCCTTGTAA TCTACAGGTA CTCATGGCAT
TCCTCGAATA CTTTCCTGTG TTTTGCCCTT CTCCCATTTC CCTTTTGCAA
GTTTAGAATA TTGTCCCCAA GATGTCTGTC CGATGCTACA CAATGGCCCT
TCAAAGTCCT ATTCAAATGG GATTGTTCTA GTAACAGCTT CCTGGGTCCA
                                                    T (PSV specfic for NF1-REP B)
AACTGAAGTC ATTTTCCCCT CTTCTATGCT TCAGAAACAA TTTACCCCTC 27441267-27441272
CTAGTGCCCA CATCCATTTC TTCTTCTTAA TGTAGTTATT GTTCATCCCA 27441322
TTTTTTCTGA GCATAAACTC CTTGAAAGCA TGGACTAGGT CTTGCTCATC 27441372
TGCATTGCCC ACCATGTTTA AAACTGACAC ATGGAAATAA AGTAAATTAA 27441422
```

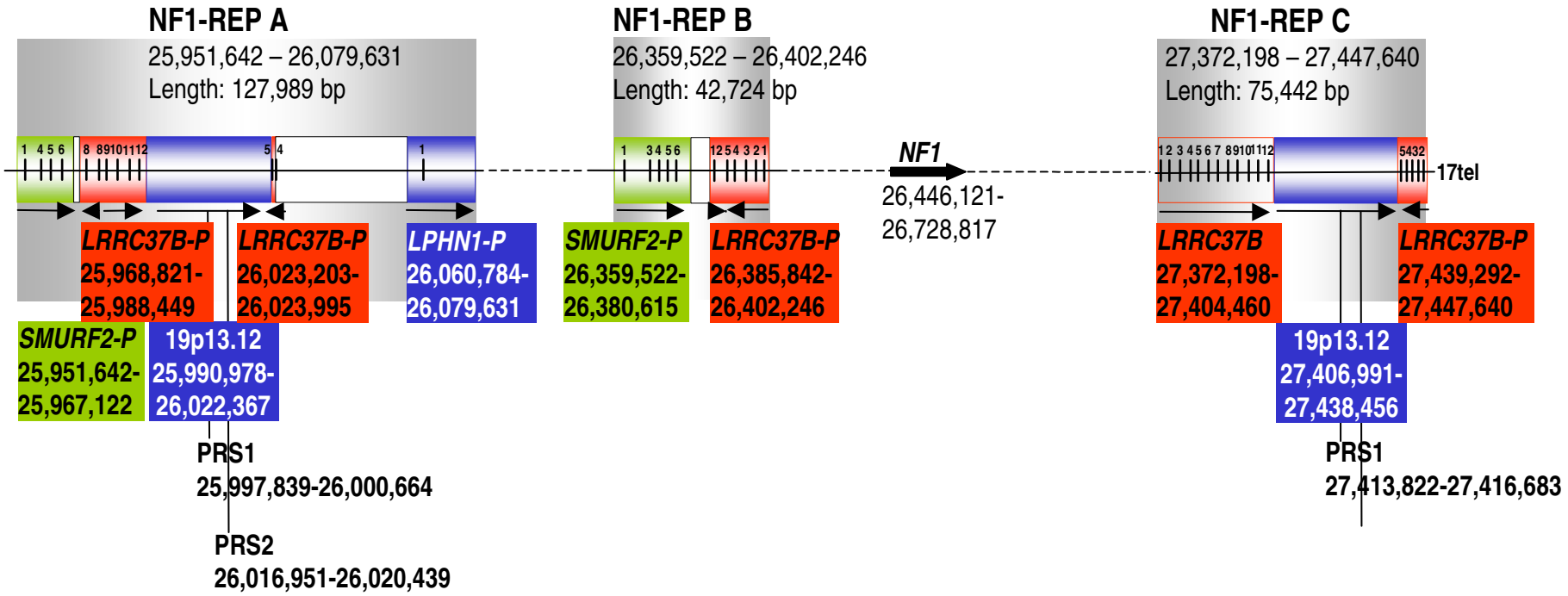
(B) Deletion breakpoint region and flanking segments in patient Z41/03

```
TTTTTCTTTC CACTCTTATT ACAACTGATC TGATTTAGCC CCTTCACACT  26396921
TTACTCCTAC ATTTTGCTAG ACCTTTCTAT TTTGTCTCCC TGAAATAAGC  26396971
TTTTCCTTTT GGGCACTCTA CATATGGATT CTAAAATAAT CCTTTGCATG  26397021
TCTATACTTG AACATAAAGC AAAAAAAACA AAATAAAATA AAATAATCTT  26397071
CCTGTTTTAA TCATTTAATC TCTCTTGCTT GGAAACTTTC AATGGCTTTC  26397121
CATACCTCAT TGTGTACTCT TTAAGACCCA AGTAAACTTT TAGCTTATCC  26397171
ATGTAGCTTT CCCTGATCTC CACACCTCAA GGACCACAGA TTCTGGTAAA  26397221-26397230
          C (PSV specfic for NF1-REP C)
TTCTAGCAAC GATGGTATGC ATTATCTTTC AGGAATTAAT TATATACACC
GCCCTTTTTA TGCCTATTCT CTTTCTTCCC TCCTATCATC TTCATGTTCT
                    T (PSV specfic for NF1-REP B)
GCGGACAATA GCATACTTGG TCCTGGGTTT GTCACAAAAT GAGTGCTACA  27446981-27447011
TATAAAAAGC CAGGGACAAA GAGGAGTTCT GGACATAACT AAAGGACAGA  27447061
GTAATCATGT TAACAGCCAA CACTCCTACA TATGCTAGGC ACTGATGAAG  27447111
```

(C) Deletion breakpoint region and flanking segments in patient 2176

```
AAAATGGGCA TAAGCTTATC TTTATCGATA AATAAGTAGA AAGAAACCCT  26396824
ATGCCAGGTA ACACCAAAGC TTTGCCCTCT GTTGAAACAT CCTAGGCTTT  26396874
TTCTTTCCAC TCTTATTACA ACTGATCTGA TTTAGCCCCT TCACACTTTA  26396924
CTCCTACATT TTGCTAGACC TTTCTATTTT GTCTCCCTGA AATAAGCTTT  26396974
TCCTTTTGGG CACTCTACAT ATGGATTCTA AAATAATCCT TTGCATGTCT  26397024-26397049
                          A (PSV specfic for NF1-REP C)
ATACTTGAAC ATAAAGCAAA AAAAA.CAAA ATAAAATAAA ATAATCTTCC
TGTTTTAATC ATTTAATCTC TCTTGCTTGG AAACTTTCAA TGGCTTTCCA
TACCTCATTG TGTACTCTTT AAGACCCAAG TAAACTTTTA GCTTATCCAT
GTAGCTTTCC CTGATCTCCA CACCTCAAGG ACCACAGATT CTGGTAAATT
      A (PSV specfic for NF1-REP B)
CTAGCACCGA TGGTATGCAT TATCTTTCAG GAATTAATTA TATACACCGC  27446971-27446914
CCTTTTTATG CCTATTCTCT TTCTTCCCTC CTATCATCTT CATGTTCTGC  27446964
GGACAATAGC ATACTTGGTC CTGGGTTTGT CACAAAATGA GTGCTACATA  27447014
```

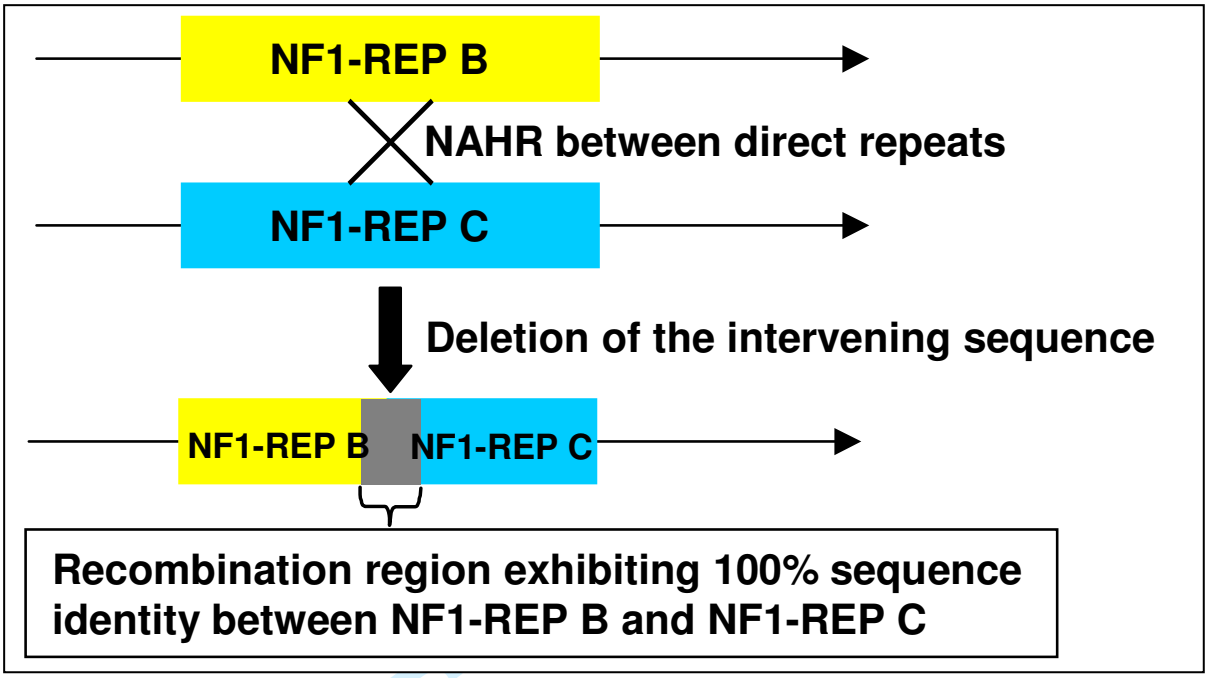
Supp. Figure S1: The deletion breakpoint regions (grey) in each patient are the regions within which the recombination events underlying the deletions must have occurred. They were identified by the inspection of paralogous sequence variants (PSVs) which are indicated in bold and underlined. The proximal flanking sequences (yellow) are homologous to NF1-REP B, whereas the distal flanking sequences (turquoise) are homologous to NF1-REP C. Those PSVs that demarcate the deletion breakpoint regions (recombination regions) that are identical between NF1-REP B and C are marked in red.



Supp. Figure S2: Schema of the modular sub-structure of NF1-REPs A, B and C comprising different paralogous sequence blocks at 17q11.2. The location and transcriptional location of the *NF1* gene is given. Pseudogene fragments of the *LRRC37B* gene (marked in red) are present in all three NF1-REPs. The functional *LRRC37B* gene, containing 12 exons, lies within NF1-REP C. Paralogous sequences with homology to 19p13.12, marked in blue, comprise portions of NF1-REPs A and C. NF1-REP A also contains a pseudogene fragment encompassing exon 1 and partial intron 1 of the latrophilin-1 precursor gene (*LPHN1*) from 19p13.12. NF1-REPs A and B also include pseudogene fragments of the *SMURF2* (SMAD-specific E3 ubiquitin-protein ligase 2) gene, marked in green. The extent of the respective paralogous sequence blocks are given according to the base-pair numbering of chromosome 17 (hg18, NCBI 36, Ensembl database version 54.36p). The horizontal arrows below the paralogous sequence blocks indicate their transcriptional orientation or their orientation in relation to the ancestral sequence. The positions of the paralogous recombination sites 1 and 2 (PRS1, PRS2) within NF1-REPs A and C are also indicated.

Position of marker on chromosome 17	Marker	Patient 2176, blood		Hybrid with normal chr. 17	Hybrid with del(17)	Father of patient 2176, blood		Mother of 2176, blood	
24481596 - 24481738	D17S1873	122	122	122	122	122	122	122	120
24885942 - 24886214	D17S1841	264	272	264	272	272	272	264	272
25124113 - 25124371	D17S975	258	262	258	262	262	254	258	254
25708122 –25708291	D17S1532	169	169	169	169	169	169	169	172
26594230 - 26594552	D17S1849	207	del	207	del	207	207	207	207
26497479 - 26497687	D17S1307	210	del	210	del	210	210	210	206
26641851 - 26642254	D17S2237	400	del	400	del	400	400	400	396
26665007 - 26665309	IVs27TG24.8	272	del	272	del	272	272	272	280
26668420 - 26668626	IVs27AC28.4	207	del	207	del	207	207	207	207
26673142 - 26673342	D17S1166	194	del	194	del	194	194	194	202
26960942 - 26961211	D17S1800	274	del	274	del	276	280	274	278
28038006 - 28038197	D17S1880	190	174	190	174	174	182	190	174
29584279 - 29584562	D17S1293	276	280	276	280	280	272	276	284
30890423 –30890752	D17S907	286	312	286	312	312	286	286	328
32909950 - 32910102	D17S1788	160	160	160	160	160	156	160	166

Supp. Figure S3: Analysis of microsatellite markers located on chromosome 17 using DNA from blood samples of patient 2176 and his parents, as well as DNA isolated from human/mouse somatic cell hybrids containing either the deletion-bearing chromosome 17 of patient 2176 or the normal chromosome of this patient.



Supplementary Figure S4: Schema of the non-allelic homologous recombination between NF1-REP B and NF1-REP C causing the type-3 *NF1* deletions.

Patient Z41/03

TTGTGTACTCTTTAAGACCCAAGTAAACTTTTAGCTTATCCATGTAGCTT
TCCCTGATCTCCACACCTCAAGGACCACAGATTCTGGTAAATTCTAGCAA
CGATGGTATGCATTATCTTTCAGGAATTAATTATATACACCGCCCTTTTT
ATGCCTATTCTCTTTCTTCCCTCCTATCATCTTCATGTTCTGCGGACAAT
AGCATACTTGGTCCTGGGTTTGTCACAAAATGAGTGCTACATATAAAAAG
CCAGGGACAAAGAGGAGTTCTGGACATAACTAAAGGACAGAGTAATCATG
TTAACAGCCAACACTCCTACATATGCTAGGCACTGATGAAGTGTTCCATA
TATTCACTC

Patient 2176

TTTTGTCTCCCTGAAATAAGCTTTTCCTTTTGGGCACTCTACATATGGAT
TCTAAAATAATCCTTTGCATGTCTATACTTGAACATAAAGCAAAAAAAA.
CAAAATAAAATAAAATAATCTTCCTGTTTTAATCATTTAATCTCTCTTGC
TTGGAAACTTTCAATGGCTTTCCATACCTCATTGTGTACTCTTTAAGACC
CAAGTAAACTTTTAGCTTATCCATGTAGCTTTCCCTGATCTCCACACCTC
AAGGACCACAGATTCTGGTAAATTCTAGCACCGATGGTATGCATTATCTT
TCAGGAATTAATTATATACACCGCCCTTTTTATGCCTATTCTCTTTCTTC
CCTCCTATCATCTTCATGTTCTGCGGACAA

Patient TOP

CGGAGTCCAGAACTGAGTTCTAACGTGCCCTCTCAATGATTGTGATGCA
GATCCATCTTACAGCACTGCTGTGATAACATGGTTCCCCATGATGGCTCC
GCAGCTTGGCATTCAAAGCCCTAGAAGATCTGGCTCCAATTTCCTACTCT
CCCTCCTTGTAATCTACAGGTACTCATGGCATTCCTCGAATACTTTCCTG
TGTTTTGCCCTTCTCCCATTTCCCTTTTGCAAGTTTAGAATATTGTCCCC
AAGATGTCTGTCCGATGCTACACAATGGCCCTTCAAAGTCCTATTCAAAT
GGGATTGTTCTAGTAACAGCTTCCTGGGTCCAAACTGAAGTCATTTTCCC
CTCTTCTATGCTTCAGAAACAATTTACCCCTCCTAGTGCCCACATCCATT
TCTTCTTCTTAATGTAGTTATTGTTCATCCCATTTTTTCTGAGCATAAAC
TCCTTGAAAGCATGGACTAGGTCTTG

Supp. Figure S5: Repeats capable of non-B DNA structure formation within the recombination regions (from patients Z41/03, 2176 and TOP) harbouring the deletion breakpoints and their flanking regions. The recombination regions are highlighted in grey, whereas the centromeric flanking regions are marked in yellow and the telomeric flanking sequences are highlighted in turquoise. Sequences capable of non-B DNA structure formation include long inexact polypyrimidine tracts, direct repeats, inverted repeats and symmetric elements. Long inexact polypyrimidine tracts, composed of at least one continuous tract ≥10bp, are underlined. Direct repeats are highlighted in pink. Inverted repeats and symmetric elements are shown in blue and red letters, respectively.