

## ORIGINAL ARTICLE

# Fine mapping of meiotic NAHR-associated crossovers causing large *NF1* deletions

Morten Hillmer<sup>1</sup>, David Wagner<sup>1</sup>, Anna Summerer<sup>1</sup>, Michaela Daiber<sup>1</sup>, Victor-Felix Mautner<sup>2</sup>, Ludwine Messiaen<sup>3</sup>, David N. Cooper<sup>4</sup> and Hildegard Kehrer-Sawatzki<sup>1,\*</sup>

<sup>1</sup>Institute of Human Genetics, University of Ulm, 89081 Ulm, Germany, <sup>2</sup>Department of Neurology, University Hospital Hamburg Eppendorf, 20246 Hamburg, Germany, <sup>3</sup>Medical Genomics Laboratory, Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35242, USA and <sup>4</sup>Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK

\*To whom correspondence should be addressed at: Institute of Human Genetics, University of Ulm, Albert-Einstein-Allee 11, 89081 Ulm, Germany. Tel: +49 73150065421; Fax: +49 731 50065402; Email: hildegard.kehrer-sawatzki@uni-ulm.de

## Abstract

Large deletions encompassing the *NF1* gene and its flanking regions belong to the group of genomic disorders caused by copy number changes that are mediated by the local genomic architecture. Although nonallelic homologous recombination (NAHR) is known to be a major mutational mechanism underlying such genomic copy number changes, the sequence determinants of NAHR location and frequency are still poorly understood since few high-resolution mapping studies of NAHR hotspots have been performed to date. Here, we have characterized two NAHR hotspots, PRS1 and PRS2, separated by 20 kb and located within the low-copy repeats *NF1*-REPa and *NF1*-REPC, which flank the human *NF1* gene region. High-resolution mapping of the crossover sites identified in 78 type 1 *NF1* deletions mediated by NAHR indicated that PRS2 is a much stronger NAHR hotspot than PRS1 since 80% of these deletions exhibited crossovers within PRS2, whereas 20% had crossovers within PRS1. The identification of the most common strand exchange regions of these 78 deletions served to demarcate the cores of the PRS1 and PRS2 hotspots encompassing 1026 and 1976 bp, respectively. Several sequence features were identified that may influence hotspot intensity and direct the positional preference of NAHR to the hotspot cores. These features include regions of perfect sequence identity encompassing 700 bp at the hotspot core, the presence of PRDM9 binding sites perfectly matching the consensus motif for the most common PRDM9 variant, specific pre-existing patterns of histone modification and open chromatin conformations that are likely to facilitate PRDM9 binding.

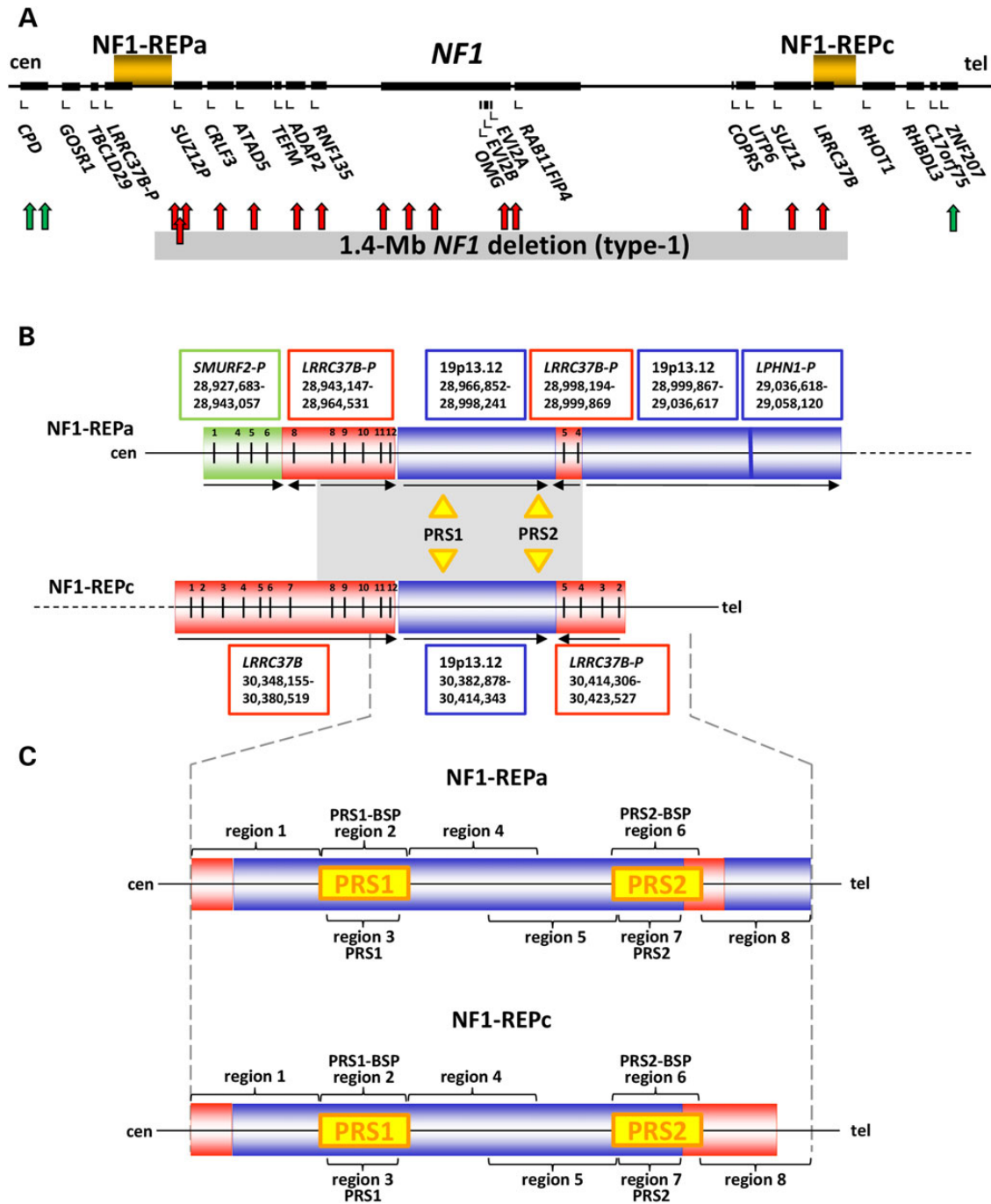
## Introduction

Neurofibromatosis type 1 (*NF1*; MIM 162200) is a common disease with an incidence at birth of ~1/3000 (1,2). Large deletions of the entire *NF1* gene and its flanking regions at 17q11.2 are observed in ~5% of all patients with neurofibromatosis type 1 (*NF1*) (3). Four types of large *NF1* deletion (type 1, type 2, type 3 and atypical) have been identified, which differ with respect to the extent of

the deleted region and the locations of the respective breakpoints (4). The majority of large *NF1* deletions (70–80%) span 1.4–1.9 Mb, as determined by multiplex ligation-dependent probe amplification (MLPA), a screening technique frequently used to detect them (5). It has been generally assumed that most of these 1.4–1.9 Mb deletions have breakpoints located within low-copy repeats, termed *NF1*-REPa and *NF1*-REPC. Large *NF1* deletions

Received: November 12, 2015. Revised and Accepted: November 19, 2015

© The Author 2015. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com



**Figure 1.** (A) Schema of the 17q11.2 region harbouring the *NF1* gene and its flanking genes. The size of the type 1 *NF1* deletion (1.4 Mb) is indicated by a grey bar. The vertical red and green arrows represent the binding sites of the MLPA probes included in the MLPA kit P122, version C2. For *NF1* deletions encompassing 1.4–1.9 Mb, a proportion of which are type 1, the target sequences for the probes shown in red are present in only one copy because they are located within the deletion region. By contrast, the target sequences for the MLPA probes shown in green are present in two copies since they are not located within the deletion region. (B) Schematic representation of the structure of *NF1-REPa* and *NF1-REPC*. The genomic positions indicated are according to hg19. (C) Positions of the genomic regions analysed to determine the degree of sequence homology between *NF1-REPa* and *NF1-REPC*. The length of the respective regions is indicated in Supplementary Material, Table S27.

spanning 1.4 Mb, with breakpoints located in *NF1-REPa* and *NF1-REPC* mediated by nonallelic homologous recombination (NAHR), have been designated ‘type 1 *NF1* deletions’. Using MLPA as the sole method of analysis, it is not immediately possible to distinguish between bona fide type 1 *NF1* deletions and atypical deletions of similar size that are not associated with breakpoints within the *NF1-REPs* and which are not mediated by NAHR. The

lack of resolution of MLPA is due to the wide spacing of the MLPA probes flanking the *NF1-REPs* (Fig. 1; Supplementary Material, Fig. S1). The application of commercially available microarrays to determine the precise extent of large *NF1* deletions would also be of limited utility owing to the lack of probes within the highly homologous *NF1-REPa* and *NF1-REPC* repeats, a prerequisite for accurate ascertainment of copy number changes.

In the study presented here, we systematically analysed a cohort of 68 NF1 patients from the University Hospital Hamburg Eppendorf, Germany (UKE cohort) who harboured 1.4–1.9 Mb NF1 deletions as initially determined by MLPA. The first goal of our analysis was to ascertain how many of these deletions were indeed bona fide 1.4 Mb type 1 NF1 deletions mediated by NAHR and possessing breakpoints located in homologous regions of NF1-REPa and NF1-REPC. Since the resolution of the MLPA assay would have been insufficient to be able to assign the deletion breakpoints unambiguously, we performed breakpoint-spanning polymerase chain reactions (PCRs) to identify the breakpoints located within the known NAHR hotspots PRS1 and PRS2 (Supplementary Material, Fig. S2) (6,7). This provided a measure of the relative proportions of the PRS1- versus PRS2-mediated deletions, thereby yielding an indication of the relative activity of these NAHR hotspots.

Although NAHR is an important mutational mechanism affecting many regions of the human genome, little is known as yet about the sequence features that influence NAHR position and frequency (8). Meiotic NAHR and allelic homologous recombination (AHR) are thought to be mechanistically similar processes (9,10) and both meiotic AHR- and NAHR-associated crossovers occur in hotspots of a few kilobases (kb) in length (6,11–17). PRDM9, a meiosis-specific methyltransferase, is an important regulator of meiotic AHR in many species including human since it directs the genome-wide positioning of AHR hotspots via sequence-specific DNA binding of its zinc finger array (18–23). PRDM9 is one of the most rapidly evolving genes in the human genome, and many allelic variants have been described (19,24–27). Different PRDM9 alleles have been shown to activate different sets of AHR hotspots and to influence hotspot strength differentially (18,19,28,29). A significant enrichment of the PRDM9 protein binding motif 5' CCNCCNTNNCCNC 3' has been observed in the vicinity of NAHR sites, consistent with both AHR and NAHR using identical recombination activating binding motifs (30). Thus, PRDM9 is likely to play an important role during NAHR as well as AHR.

In the study presented here, we investigated various sequence features of the NAHR hotspots PRS1 and PRS2, including the number of potential PRDM9 binding sites, which may serve to influence NAHR frequency. To maximize the number of PRS1- and PRS2-mediated deletions under study, we additionally screened a cohort of patients with suggested type 1 NF1 deletions initially ascertained in the Department of Genetics, University of Alabama at Birmingham, USA (UAB cohort). By combining patients from the UKE and UAB cohorts, we were able to analyse 50 PRS2- and 28 PRS1-mediated NF1 deletions. We sequenced across the sites of the NAHR-associated crossovers, the strand exchange regions (SERs) between NF1-REPa and NF1-REPC. The high-resolution mapping of the hotspots then enabled us to determine the extent of the PRS1 and PRS2 regions as well as the most commonly used SERs that demarcate the highly active cores of the NAHR hotspots. Our findings indicate that multiple factors are likely to influence the increased rate of NAHR observed within the hotspots PRS1 and PRS2.

## Results

### Proportion of PRS1- and PRS2-mediated NF1 deletions

We systematically investigated 68 unrelated NF1 patients (UKE cohort) with putative type 1 deletions according to MLPA to determine whether their deletions were indeed bona fide type 1

deletions mediated by NAHR (Supplementary Material, Tables S1–S4). Further, we determined the relative proportions of PRS1- versus PRS2-mediated deletions, which have not been previously ascertained. Breakpoint-spanning PCR and sequence analysis of the breakpoint-spanning PCR products (BSPs) indicated that 52 deletions (76.5%) were PRS2-mediated, whereas 13 deletions (19.1%) exhibited breakpoints within PRS1. Three deletions (4.4%) did not exhibit breakpoints within either PRS1 or PRS2. Our findings clearly indicate that the vast majority (95.6%) of large NF1 deletions classified as type 1 according to MLPA are indeed bona fide type 1 deletions exhibiting breakpoints located within the NAHR hotspots PRS1 and PRS2. The ratio of PRS2-mediated to PRS1-mediated deletions is of the order of 4:1.

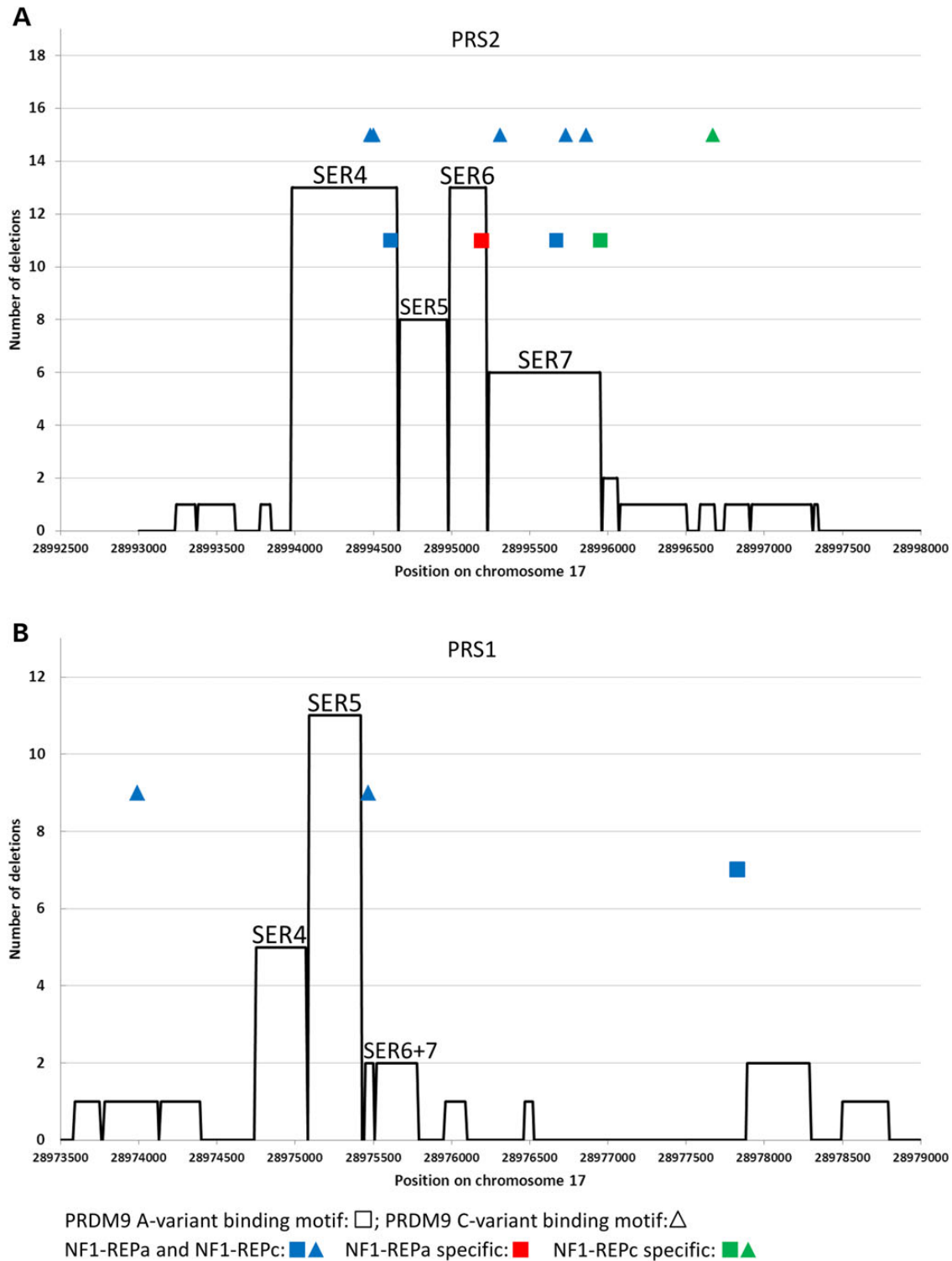
### Sequence analysis of the PRS2 hotspot

In our previous analysis (7), we determined the locations of the SERs between NF1-REPa and NF1-REPC in 20 of the type 1 NF1 deletions from the UKE cohort, which exhibited breakpoints located within the PRS2 hotspot. In order to identify the most common SERs within PRS2, and hence the hotspot core, it was necessary to analyse a much larger number of PRS2-positive deletions. Therefore we sequenced the BSPs of four additional PRS2-mediated deletions from the UKE cohort (Supplementary Material, Table S5) as well as 26 PRS2-mediated deletions from the UAB cohort (Supplementary Material, Table S6). These 26 deletions were initially identified by MLPA. Sequence analysis of BSPs performed during the course of this study indicated that they all harboured breakpoints within PRS2. In total, we analysed the SERs in 50 PRS2-mediated type 1 NF1 deletions (Supplementary Material, Table S7).

The SERs harbour the sites of NAHR-associated crossovers and are flanked by either paralogous sequence variants (PSVs) or single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) of  $\leq 1\%$  that serve to distinguish NF1-REPa from NF1-REPC. Within an SER, NF1-REPa and NF1-REPC exhibit absolute sequence identity. Consequently, higher resolution of the crossover locations within the SERs is not possible. The 50 PRS2-mediated deletions analysed exhibited crossovers within 13 different SERs ranging in size from 25 to 737 bp. According to the genomic locations of these SERs, PRS2 extends over 4096 bp (Supplementary Material, Fig. S2). Nine of the 13 SERs harboured the crossovers of only one or two deletions. By contrast, SERs 4–7 were recurrent since a total of 40 deletions (80%) had crossovers within these SERs (Supplementary Material, Table S8). SERs 4–7 span 1976 bp and represent the hotspot core (Fig. 2). We conclude that the deletion breakpoints in PRS2 are non-randomly distributed, clustering within specific SERs (Fig. 2). The most commonly encountered SERs (4 and 6) were detected in a total of 26 patients (52%) (Supplementary Material, Table S8).

### Sequence analysis of the PRS1 hotspot

Type 1 NF1 deletion breakpoints located within PRS1 have not been extensively characterized as yet. We sequenced the BSPs of the 13 PRS1-mediated type 1 deletions identified among the 68 patients from the UKE cohort (Supplementary Material, Table S9). To maximize the number of PRS1-mediated deletions under study, we also included 15 PRS1-mediated type 1 deletions from the UAB cohort (Supplementary Material, Table S10). Taken together, we were able to analyse 28 PRS1-mediated type 1 NF1 deletions and detected 11 different SERs within 5202 bp



**Figure 2.** Localization of the SERs within PRS2 (A) and PRS1 (B). Predicted PRDM9 A-variant binding sites (A-motifs) are indicated by rectangles, whereas the predicted binding sites of the less common PRDM9 C-variant (C-motifs) are indicated by triangles. PRDM9 binding sites shown in blue are located within NF1-REPA and NF1-REPC, whereas those marked in green are specific to NF1-REPC. PRDM9 binding sites marked in blue and green are invariant since no SNPs are present within these sequences that match perfectly with the binding site consensus sequence motif 5' CCNCCNTNNCCNC 3' for the PRDM9 A-variant (31) or with the binding site consensus motif 5' CCNCCNTNNCCNTNCC 3' for the PRDM9 C-variant (32,33). The A-motif indicated in red is polymorphic; a perfect match to the 13 bp A-motif consensus is only evident if the C-allele of SNP rs114479203 is present. The frequency of the C-allele of SNP rs114479203 was found to be 0.199 in Europeans (34).

(Supplementary Material, Tables S11 and S12). The recurrent SERs 4–7 harboured the crossovers of 20 deletions (71%) and represent the 1026 bp core of PRS1. Thus, clustering of crossovers

within a hotspot core is apparent in PRS1 as well as in PRS2, indicating preferred sites of double-strand breaks (DSBs) and/or double Holliday junction (dHj) resolution within both hotspots.

## Gene conversion within PRS1 and PRS2

Sequence analysis of the BSPs also indicated sites of putative gene conversion within PRS1 and PRS2 (Supplementary Material, Tables S13–S17). To determine whether these gene conversion events were associated with the deletion-causing NAHR or whether they instead resulted from nonallelic homologous recombination without crossover (NAHGC) occurring in a previous generation, we analysed the non-recombinant (wild-type) PRS1 and PRS2 sequences in the parents of the patients. NAHR-associated gene conversion was observed in 4 of the 28 PRS1-mediated type 1 NF1 deletions (14%) and 5 of the 50 PRS2-mediated deletions (10%) (Supplementary Material, Table S17). The rate of gene conversion events within PRS2 observed in our present study was higher than noted previously (7). This discrepancy may be explained in terms of the analysis of a larger number of patients and improved variation data provided by the 1000 Genomes Project, which helped us to distinguish between SNPs with a higher allele frequency and PSVs. In most instances, the gene conversion events affected only single PSVs and did not impair the ascertainment of the SERs.

The majority of the gene conversion events detected here favoured GC alleles, which was suggestive of GC-biased gene conversion. Within PRS2, we observed a GC content of 58.7%, which was markedly higher than its flanking regions and the local genomic average (49.6% GC within 40.4 kb). By contrast, the GC content within PRS1 was 52.8%, which was only moderately elevated when compared with the local average GC content (Supplementary Material, Fig. S3).

## Sequence features of PRS1 and PRS2

### PRDM9 binding sites

As part of our attempt to explore the sequence features of the NAHR hotspots PRS1 and PRS2 that could account for the higher incidence of NAHR events within PRS2 when compared with PRS1, we screened both hotspots for the presence of putative PRDM9 binding sites. The PRDM9 gene is highly polymorphic in humans, with the PRDM9 A-allele being the most common PRDM9 allele in Europeans (18,24,35–38) (Supplementary Material, Table S18). The A-allele was also the most common PRDM9 allele in the healthy parents of 24 of the patients with type 1 NF1 deletions (Supplementary Material, Tables S19 and S20). The parents did not harbour the deletion in their somatic cells; rather, the deletion occurred *de novo* in their germlines by meiotic NAHR and was subsequently transmitted to their offspring. Hence, it is very likely that the protein encoded by the PRDM9 A-allele initiated NAHR in PRS1 and PRS2. Three non-variant and one polymorphic PRDM9 protein binding sites, perfectly matching the PRDM9 A-variant consensus motif (5' CCNCCNTNNCCNC 3'), are located within the core of PRS2 (Fig. 2; Supplementary Material, Fig. S4 and Table S21). By contrast, PRDM9 binding sites perfectly matching the consensus A-motif are absent from the core of PRS1 encompassing SERs 4–7 (Supplementary Material, Table S22). Only binding sites exhibiting seven of eight matches to the consensus motif were identified in the PRS1 core (Supplementary Material, Table S23). Several PRDM9 C-variant binding sites are predicted to be located within PRS1 and PRS2 (Fig. 2), but PRDM9 C-alleles are much less frequent than A-alleles in Europeans and the transmitting parents (Supplementary Material, Tables S18 and S19).

### Distance between the SERs and the PRDM9 binding sites

We next calculated the distance between each NAHR-associated SER located within PRS2 and the nearest predicted PRDM9 A-variant binding site. The observed distances ranged from 111 to

1367 bp (Supplementary Material, Table S24). This analysis was necessarily confined to PRDM9 A-variant binding sites since the parents of patients with PRS2-mediated deletions were invariably carriers of the PRDM9 A-allele (Supplementary Material, Table S19), which encodes the PRDM9 protein A-variant. This variant specifically activates recombination events by binding to the consensus sequence A-motif (19,24).

Although we were able to investigate only five transmitting parents of patients with PRS1-mediated deletions, the PRDM9 A-allele was the most common allele in the deletion-transmitting parents of these patients as well as in the general population (Supplementary Material, Table S19). However, in contrast to PRS2, only one PRDM9 A-variant binding site perfectly matching the A-variant consensus motif is located in PRS1, some 2.3–3 kb distant from the hotspot core (Fig. 2; Supplementary Material, Fig. S4 and Table S25).

### Chromatin and histone modifications

To investigate whether chromatin conformation and histone modifications could serve to distinguish PRS1 and PRS2 from their flanking regions, we analysed ENCODE data for the regions in question (39). Type 1 NF1 deletions with breakpoints in PRS1 or PRS2 are mostly of meiotic origin occurring in the maternal germline (40). Hence, ideally, the chromatin state and histone modifications within PRS1 and PRS2 should be analysed in oocytes. However, germ cell lines were not included in the ENCODE project. Since multipotent adult germline stem cells exhibit embryonic stem cell-like features (41), we opted to analyse the ENCODE data obtained from the embryonic stem cell line H1-hESC as a proxy for oocytes. In the centromeric portion of PRS1 and in the core of PRS2, open chromatin states suggestive of promoter-like features were identified, which were lacking from their flanking regions (Fig. 3; Supplementary Material, Figs S5 and S6). These observations may at first seem surprising since PRS1 and PRS2 are not located upstream of any functional gene in 17q11.2. However, the evolutionary origin of PRS1 and PRS2 within the NF1-REPs may be informative in this context. During primate genome evolution, after the separation of the orang-utan from the human lineage ~14 Mya, a 32 kb sequence segment harbouring the ancestral PRS1 and PRS2 regions, located upstream of the LPHN1 gene at 19p13.2, was duplicated and transposed to 17q11.2 forming part of one of the NF1-REPs (42). PRS1 and PRS2 are located within this 32 kb duplicon paralogous to 19p13.12 (Fig. 1). The 19p13.12 PRS2 paralog may constitute an alternative promoter for the LPHN1 gene (Supplementary Material, Fig. S7) and hence may have retained its promoter-like open chromatin conformation after the duplicative transposition from 19p13.12 to 17q11.2. By contrast, the PRS1 region at 19p13.12 does not exhibit promoter-like features (Supplementary Material, Fig. S7), although it has acquired an open chromatin state characteristic of a weak promoter after its duplicative transposition to 17q11.2 (Fig. 3; Supplementary Material, Fig. S5).

PRS1 and PRS2 clearly differ from their flanking regions with regard to their open chromatin/ promoter-like features. Further, they also exhibit a very specific histone modification pattern which is lacking in their flanking regions. In both PRS1 and PRS2, histone H3 lysine 4 dimethylation (H3K4me2) was overrepresented. In PRS2, H3K4me3, histone H2 acetylated at lysine 9 (H3K9ac) and histone H3 acetylated at lysine 27 (H3K27ac) were enriched when compared with the flanking regions (Fig. 3; Supplementary Material, Fig. S6 and Table S26).

Both the open chromatin state within PRS1 and PRS2 and the specific histone modifications in the vicinity were identified in



**Figure 3.** Chromatin conformation and histone modifications within PRS1 and PRS2 as well as their flanking regions in the embryonic stem cell line H1-hESC according to the ENCODE data (<https://genome.ucsc.edu/>) (39). The PRS1 and PRS2 regions within NF1-REPa (A) and NF1-REPC (B) are highlighted by dashed lines. Track 1 indicates the chromatin state segmentation (Chrom-HMM) based on ChIP-seq data using a hidden Markov model (HMM). The coloured bars in Track 1 highlight predicted functional elements and chromatin states: dark red, active promoter; light red, weak promoter; green, weakly transcribed; orange, strong enhancer; yellow, weak enhancer; blue, insulator). Tracks 2–15 indicate histone modifications analysed by ChIP-seq. For every histone modification type, signal (s) and peak (p) tracks are indicated. The peak tracks show regions of statistically significant histone modification enrichment, whereas the signal tracks display signal graphs of histone modifications. Track 16 indicates regions of open chromatin and/or transcription factor binding sites determined by combined data obtained by DNaseI hypersensitivity (HS) assay, Formaldehyde-assisted isolation of regulatory elements (FAIRE) and chromatin immunoprecipitation (ChIP) for regulatory factors. Regions exhibiting features of open chromatin conformation in Track 16 are indicated by rectangles according to the following colour code: blue, high significance; green, low significance, regions identified by DNaseI HS assays but not by FAIRE as peaks; red, low significance, regions identified by FAIRE as peaks but not identified by DNaseI HS; black, low significance, regions identified as peaks by both DNaseI HS and FAIRE; and pink, regions identified by ChIP-seq as peaks. In Tracks 17 and 18, CpG methylation patterns are indicated. Short unmethylated regions observed only in PRS2 are indicated as blue bars.

the H1-hESC cell line, as well as in other cell lines for which ENCODE data are available (Supplementary Material, Figs S8–S10).

#### Recombination rates within PRS1 and PRS2

During primate genome evolution, sequences at 19p13.12 were duplicated and transposed to 17q11.2 giving rise to the NF1-REPs including the PRS1 and PRS2 regions (42). A previous study suggested that the progenitor sequences of PRS1 and PRS2 at 19p13.12 harbour hotspots of AHR (6). We evaluated the HapMap Phase II data and observed a 2- to 3-fold increased recombination rate in PRS1 and PRS2 of NF1-REPC as well as in their paralogous regions at 19p13.12 when compared with the local average recombination rate estimated from the recombination rates observed within 100 kb flanking PRS1 and PRS2 on both sides (Supplementary Material, Fig. S11). Although the number of informative SNPs within the NF1-REPs is low, our data concur with the observations of De Raedt *et al.* (6) in that they support the view that the NAHR hotspots PRS1 and PRS2 overlap with regions of pre-existing AHR hotspots.

#### Sequence homology between the NF1-REPs

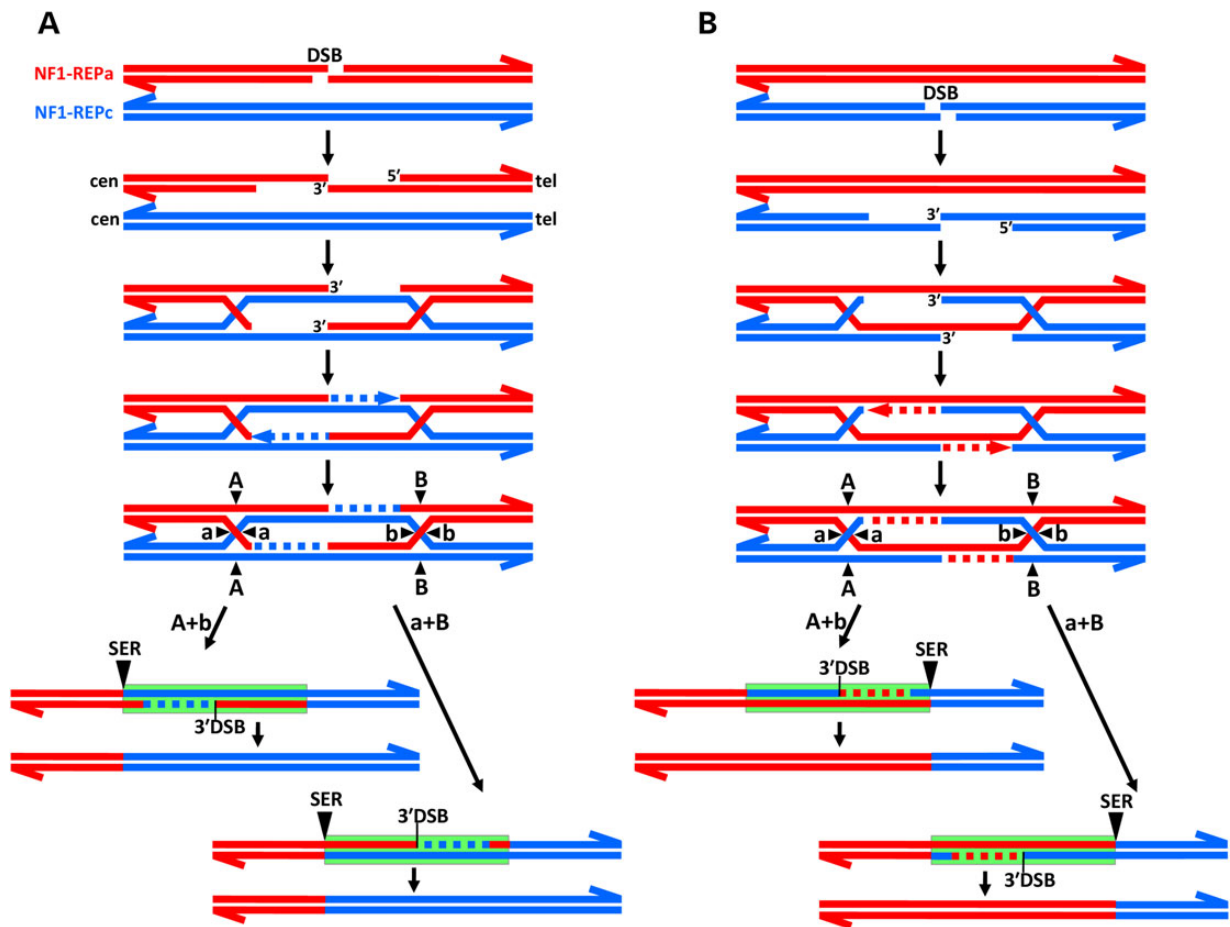
To investigate whether there are DNA sequence differences between NF1-REPa and NF1-REPC that might influence the location of NAHR-associated strand exchanges, we performed BLAST sequence alignments. In this context, regions exhibiting 100% sequence homology (perfect identity) between the NF1-REPs are particularly interesting since they may influence the recombination efficiency (43). Within PRS1 and PRS2, the average sequence homology between NF1-REPa and NF1-REPC was 98% (Supplementary Material, Tables S27 and S28). We did not observe any differences between PRS1 and PRS2 in terms of the number of

regions spanning 200–300 bp that exhibited perfect sequence identity between the NF1-REPs, although we did identify a 700 bp region within PRS2 (but not in PRS1) that exhibited perfect sequence identity between NF1-REPa and NF1-REPC (Supplementary Material, Fig. S12).

Sequence alignments of 10 kb regions located centromeric to PRS1 and PRS2, respectively, did not indicate striking differences with regard to the number of segments showing perfect sequence identity between NF1-REPa and NF1-REPC (Supplementary Material, Figs S13A and S14A). However, sequence alignments of regions located telomeric to PRS1 and PRS2 revealed a considerable number of differences. A high degree of sequence homology between NF1-REPa and NF1-REPC was observed within 10 kb flanking PRS1 in a telomeric direction (Supplementary Material, Fig. S13B), but the sequence homology between NF1-REPa and NF1-REPC ends abruptly 2.5 kb telomeric to PRS2 (Supplementary Material, Fig. S14B). This interruption in sequence homology could contribute to the positional preference of dHj resolution within PRS2.

## Discussion

The systematic analysis of 68 NF1 deletions from the UKE cohort, which were initially identified by MLPA and suggested to be type 1, indicated that 65 (96%) were indeed bona fide type 1 NF1 deletions. By definition, type 1 NF1 deletions encompass 1.4 Mb and are mediated by NAHR with crossovers occurring within the two NAHR hotspots, PRS1 and PRS2. Thus, despite the wide spacing of the MLPA probes employed (Supplementary Material, Fig. S1), MLPA proved to be an effective technique for the prediction of type 1 NF1 deletions. However, formal



**Figure 4.** NAHR between NF1-REPa and NF1-REPC according to the DNA DSB repair model involving a dHj. A DSB is introduced at the NAHR hotspot either in NF1-REPa (A) or in NF1-REPC (B). Subsequently, 5' to 3' resection occurs on one strand leaving 3' single-stranded overhangs that invade the unbroken paralog, base pairing in the process with the complementary strand and displacing the other strand to form a D-loop. The overhangs at the 3' end remain unprocessed and demarcate the site of the original DSB close to the PRDM9 binding site. New DNA synthesis is denoted by dotted lines with the unbroken paralogous strand serving as the template. The dHj is completed after the ligation of the DNA ends and can be resolved either by crossover (as shown here) or by noncrossover (not shown). Two types of crossover resolution are possible, distinguished by the sites of cleavage, either  $A + b$  or  $a + B$ . Both types of resolution yield a deletion (as shown here) and a duplication chromatid (not shown). Heteroduplex regions are corrected using the unbroken DNA strand as template. The SER representing the NAHR-associated crossover is marked by a black triangle. The 3' end of the recombination-initiating DSB (3'DSB) is located close to the PRDM9 binding site. The green rectangles indicate regions of PRDM9-modified nucleosomes. These nucleosomes contain histone H3 trimethylated at lysine K4 (H3K4me3) by PRDM9 and serve to limit the branch migration of the dHj (48). The green rectangle indicates the extent of the genomic region involved in forming the dHj.

confirmation of a bona fide type 1 NF1 deletion is only possible by means of breakpoint-spanning PCRs with primers that distinguish between the highly homologous NF1-REPa and NF1-REPC. Therefore, we performed breakpoint-spanning PCRs in order to confirm that the NF1 deletions were indeed mediated by NAHR and exhibited breakpoints located within PRS1 or PRS2. Mapping of type 1 NF1 deletion breakpoints using commercially available array CGH platforms is much less efficient than the breakpoint-spanning PCR approach due to the lack of robust copy-number-sensitive array probes located close to the NAHR hotspots. The analysis of the BSPs performed as part of this study indicated that PRS2 is a 4-fold stronger NAHR hotspot than PRS1. Of the 65 bona fide type 1 NF1 deletions of the UKE cohort, 52 (80%) exhibited crossovers within PRS2, whereas only 13 (20%) were within PRS1. This difference in hotspot strength between PRS1 and PRS2 has been observed in previous studies (6,44–47) but has not been systematically analysed before in terms of the precise proportions of PRS1- versus PRS2-mediated deletions.

To identify the sequence features characteristic of both hotspots, as well as those that might be responsible for the higher frequency of NAHR-associated crossovers within PRS2 when compared with PRS1, we included in our analysis 15 type 1 NF1 deletions identified in the UAB cohort of NF1 patients. In total, we assessed 28 type 1 NF1 deletions with breakpoints in PRS1 and 50 PRS2-mediated deletions. We sequenced the respective BSPs and determined the SERs between NF1-REPa and NF1-REPC within both NAHR hotspots. The SER of an NAHR-mediated deletion indicates the location of the dHj resolution and hence the region of crossover (Fig. 4). Within PRS2, 13 SERs were identified, encompassing 4096 bp, thereby indicating the extent of the PRS2 hotspot. A clustering of crossovers within specific SERs was observed since 40 (80%) of the 50 PRS2-mediated deletions exhibited crossovers located within SERs 4–7. These SERs encompass 1976 bp and represent the hotspot core (Supplementary Material, Table S8; Fig. 2). Within PRS1, 11 SERs were identified encompassing 5202 bp. Consequently, the PRS1 hotspot is ~1 kb longer than that of PRS2. This notwithstanding, a clustering of crossovers was

also detected in PRS1. Twenty (80%) of the 28 PRS1-mediated deletions exhibited crossovers in SERs 4–7 located within 1026 bp (Supplementary Material, Table S12; Fig. 2). We therefore conclude that a strong positional preference must exist for NAHR to be initiated within both hotspots which are separated from each other by 20 kb (Fig. 1). Although NAHR-associated gene conversion was observed in 4 of the 28 PRS1-mediated type 1 NF1 deletions (14%) and 5 of the 50 PRS2-mediated deletions (10%) (Supplementary Material, Table S17), it is unlikely that gene conversion will have hampered the determination of the SERs. Most of the gene conversion events detected here favoured GC alleles, an observation consistent with previous reports of GC-biased gene conversion being associated with recombination events in a variety of different species (29,49–54). Within PRS2, and to a lesser extent also in PRS1, we observed a higher GC content when compared with their flanking regions, most likely resulting from GC-biased gene conversion within these hotspots. This finding concurs with those of other NAHR hotspots in the human genome (30).

It has been suggested that the synopsis of genomic regions exhibiting extensive sequence homology is a prerequisite for the pairing and crossover underlying NAHR (10,55). Hence, both the frequency and location of NAHR within the NF1-REPs are likely to be influenced by the degree of sequence homology (56). In view of the fact that NF1-REPa and NF1-REPC exhibit on average 98% sequence homology within 46.4 kb (Supplementary Material, Table S28), the strong positional preference for NAHR to be initiated within PRS1 and PRS2 is remarkable, with both hotspots exhibiting narrow cores encompassing only 1026 and 1976 bp, respectively. The NF1-REPs comprise several duplicate sequences with different lengths, orientations and evolutionary origins (Supplementary Material, Fig. S1). Analysis of the sequence homology between the NF1-REPs indicated a continuous segment of 700 bp exhibiting perfect sequence identity between them; this 700 bp segment was present exclusively in PRS2, and neither in PRS1 nor in the flanking regions of either hotspot (Supplementary Material, Figs S12–S14). This 700 bp region of sequence identity overlaps with SER 7, located within the core of PRS2, and may well facilitate NAHR-associated strand invasion and dHj formation leading to the high number of crossovers within PRS2.

The sequence homology between NF1-REPa and NF1-REPC ends abruptly 2.5 kb telomeric to PRS2 (Supplementary Material, Fig. S14). This sharp drop in sequence homology may also contribute to frequent crossovers within PRS2. By contrast, the 10 kb regions flanking PRS1 on both sides exhibit continuous high sequence homology (Supplementary Material, Figs S13 and S14). Our findings with PRS2 are reminiscent of the NAHR hotspots within the Smith–Magenis syndrome-REPs at 17p11.2, which were found to be located close to a discontinuity of sequence homology, suggesting that such boundaries might influence the positioning of dHj resolution (55).

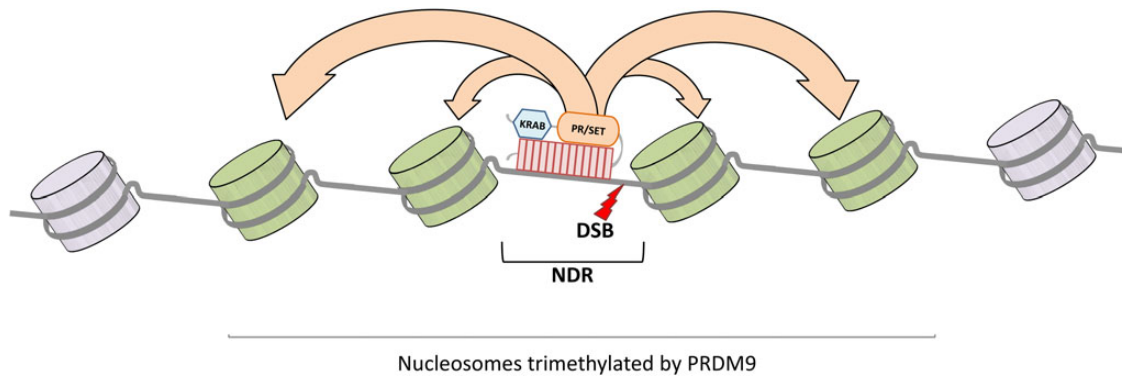
Although sequence homology promotes the occurrence of NAHR events, other DNA sequence features are also likely to contribute to the observed positional preference of NAHR-associated crossovers. The previous results (6) taken together with our own analysis of recombination rates using the HapMap Phase II data (Supplementary Material, Fig. S11) imply that the NAHR hotspots PRS1 and PRS2 emerged on a background of pre-existing hotspots for AHR originally present at paralogous positions in the progenitor sequences of the NF1-REPs located at 19p13.12. AHR hotspots have been shown to exhibit certain specific histone modifications in a variety of different species (Supplementary Material, Table S26) (57–63). Remarkably, AHR hotspots occupy 'bivalent chromatin regions' that harbour both active (H3K4me3) and

repressive (H3K27me3) marks. We observed a specific pattern of histone modification at the PRS1 and PRS2 hotspots that was absent from their flanking regions. In PRS2, and to a lesser extent also in PRS1, H3K4me2 and H3K4me3 were significantly enriched (Fig. 3, Supplementary Material, Table S26). In PRS2, but not in PRS1, H3K9ac and H3K27ac were also significantly enriched. The specific histone modification patterns and the open chromatin organization within PRS1 and PRS2 may facilitate NAHR. This pattern of histone modification and open chromatin was more pronounced in PRS2 than in PRS1, which may also contribute to PRS2 being a stronger NAHR hotspot when compared with PRS1 (Fig. 3; Supplementary Material, Figs S5 and S6).

In the present study, we assessed various other NAHR hotspots identified in previous studies (12,15,17,64,65) in order to ascertain their histone modification patterns and chromatin conformations using the ENCODE data. However, the specific histone modification patterns and open chromatin conformation we observed in association with the NAHR hotspots PRS1 and PRS2 have not been consistently observed features at these NAHR hotspots (data not shown). By contrast, a recent genome-wide study of human deletion breakpoints suggested that NAHR-associated breakpoints tend to occur in open chromatin and exhibit a high density of specific histone modifications such as H3K4me2 and H3K79me2 (66). This suggests firstly that a hierarchy of different features may be involved in determining the location and intensity of NAHR hotspots, and secondly that the pre-existing histone modification pattern or open chromatin state is not necessarily relevant in all cases.

In addition to its open chromatin conformation and a specific histone modification pattern, PRS2 is also distinguishable from PRS1 by the number and location of predicted PRDM9 binding sites (Fig. 2; Supplementary Material, Fig. S4). As in the general European population, the PRDM9 A-allele was found to be the most common PRDM9 allele in the healthy parents of 24 patients with type 1 NF1 deletions (Supplementary Material, Table S19). Three non-polymorphic binding sites perfectly matching the 13-mer consensus binding motif for PRDM9 A-variants were predicted to be located within the core of PRS2 encompassing SERs 4–7. By contrast, PRDM9 A-variant binding sites perfectly matching the consensus binding motif were absent from the core of PRS1, and only PRDM9 A-variant binding sites with weaker matches to the consensus motif were detected (Supplementary Material, Table S23). Sperm-typing studies have shown that single-point mutations in the 13 bp consensus motif for the PRDM9 A-variant can completely abolish hotspot activity (67–69). Consequently, the absence of PRDM9 A-variant binding sites perfectly matching the consensus binding motif close to the core of PRS1 may well contribute to a lower NAHR-initiation rate within PRS1 when compared with PRS2. However, PRDM9 activity at recombination hotspots is likely to be regulated in a complex manner by *cis*-acting sequence motifs as well as by the local chromatin environment and *trans*-acting factors (70,71). This may also be concluded from the identification of human AHR hotspots that are dependent upon PRDM9 activity yet do not contain the predicted PRDM9 binding motif (19,24). This notwithstanding, *in vitro* studies have clearly indicated that the PRDM9 A-variant protein exhibits high binding affinity to the A-variant consensus binding motif (13-mer) but only low affinity to the 13-mer consensus motif if single nucleotides at conserved positions are mutated (18). Furthermore, SNPs within the consensus PRDM9 binding sites have been shown to influence AHR hotspot activity in humans (24). These findings lend support to our hypothesis that the sequences perfectly matching the consensus motif for





**Figure 5.** Model of PRDM9 activation of recombination hotspots according to Baker et al. (48). PRDM9 binds to the DNA within the hotspot core via its zinc finger domain and modifies surrounding nucleosomes by trimethylation at histone H3. PRDM9-modified nucleosomes are remodelled to create a central NDR. Subsequently, a DNA DSB is introduced within the NDR, spanning ~120 bp, by meiosis-specific SPO11. The location of the meiotic crossover (e.g. the branch migration) is constrained by PRDM9-dependent chromatin modification and hence is restricted to the nucleosomes modified by PRDM9.

the PRDM9 A-variant within PRS2 are important for PRDM9 binding and are associated with the higher NAHR activity in PRS2 when compared with PRS1 that lacks a binding site perfectly matching the consensus motif in its core.

According to the DNA DSB repair model of meiotic recombination, subsequent to the DSB, single-stranded DNA ends invade the homologous sequences (Fig. 4). The DSB is then repaired by the reciprocal exchange of DNA, forming a dHj associated with branch migration. Consequently, the location of the crossover is different from the site of the original DSB (Fig. 4) (72). PRDM9 directs the recombination-initiating SPO11-mediated DSB close to its binding site (21,28,73). In the mouse, PRDM9 catalyses the trimethylation of histone H3 at lysine 4 and reorganizes nucleosomes within recombination hotspots in a symmetrical pattern around a central nucleosome-depleted region (NDR) of ~120 bp. Within the NDR, the DSB occurs close to the PRDM9 binding site (Fig. 5) (48,74). Further, branch migration of the dHj is limited to the region of nucleosomes trimethylated by PRDM9 encompassing 1000–2500 bp (48). Accordingly, the extent of the dHj is determined by the number of nucleosomes with PRDM9-modified H3K4me3 and should encompass 1 up to 2.5 kb.

If we apply this model of PRDM9 activity to NAHR within PRS2 causing type 1 NF1 deletions, the SERs determined in the current work represent the regions of dHj resolution or crossover, whilst the DSB occurred close to the predicted PRDM9 binding sites. In view of the fact that a dHj is a symmetrical structure, the distance between the SER and the PRDM9 binding site corresponds to half the length of the dHj (Fig. 4). We observed distances between the SERs and the predicted PRDM9 A-variant binding sites within PRS2 of up to 1367 bp (Supplementary Material, Table S24). Since these regions represent only half the length of the dHj as shown in Figure 4, it follows that the entire H3K4me3 track length mediated by PRDM9, and hence the extent of the dHjs underlying the NAHR events causing PRS2-mediated deletions, should encompass 2–3 kb. This estimate of the H3K4me3 track length mediated by PRDM9 is thus similar to that observed at mouse AHR hotspots (48).

We conclude that the high frequency of NAHR-associated crossovers within PRS2 is potentiated by optimal conditions for PRDM9 binding and effective reorganization of nucleosomes leading to stable dHj formation and branch migration. The sequence features responsible for providing these optimal conditions for PRDM9 action within PRS2 are likely to include open chromatin conformation as well as a specific histone modification pattern preceding PRDM9-mediated nucleosome reorganization. By contrast, within PRS1, the conditions for

PRDM9 binding and nucleosome reorganization are probably suboptimal owing to the lack of PRDM9 binding sites perfectly matching the consensus motif leading to a reduced number of crossovers within PRS1 when compared with PRS2. These conclusions are supported by recent studies showing that PRDM9 not only activates AHR hotspots but is also responsible for quantitative modulation of the recombination rate in mice (75).

In summary, our findings indicate that a number of different chromosomal features are likely to play a role in influencing the positional preference of NAHR within the NF1-REPs and the NAHR hotspots. These features include the presence of several 100 bp segments of uninterrupted sequence identity, the number of PRDM9 binding sites perfectly matching the consensus motif of the most frequent PRDM9 variants, specific patterns of histone modification and an open chromatin conformation that facilitates PRDM9 binding. We postulate that the activity of other NAHR hotspots elsewhere in the human genome may be influenced by similar mechanisms.

## Materials and Methods

### Patients

The 68 NF1 patients investigated here to determine the frequency of bona fide type 1 NF1 deletions were referred by the NF1 clinic at the Department of Neurology, University Hospital Hamburg Eppendorf, Germany (UKE cohort). They were initially analysed by microsatellite markers located within the NF1 gene region and subsequently by MLPA (P122 NF1 area probemix, version C2, MRC Holland, The Netherlands) using blood-derived genomic DNA. When a type 1 NF1 deletion was tentatively identified by MLPA, the patient DNA was further investigated by breakpoint-spanning PCRs in order to confirm that the breakpoints were indeed located within the PRS1 and PRS2 NAHR hotspots. Twenty of these 68 type 1 NF1 deletions had already been previously studied (7), but the remaining 48 deletions had not previously been analysed systematically.

In addition to these 68 patients, we analysed 41 patients with putative type 1 NF1 deletions as suggested by initial MLPA screening performed at the Medical Genomics Laboratory, Department of Genetics, University of Alabama at Birmingham, USA (UAB cohort). In the present study, breakpoint-spanning PCRs were performed using blood-derived genomic DNA from the patients from the UAB cohort in order to identify deletions with breakpoints located within the NAHR hotspots PRS1 and PRS2.

All patients fulfilled the diagnostic criteria for NF1 and gave their informed consent for the genetic analysis to be performed. This study was approved by the Institutional Review Boards of the participating institutions.

### Breakpoint-spanning PCRs, determination of SERs and gene conversion

We performed breakpoint-spanning PCRs with primers previously described or newly established for the purposes of this study (Supplementary Material, Table S1). The PCRs were performed using the Expand Long Range dNTPack (Roche) and 400 ng genomic DNA derived from blood of the patients. The breakpoint-spanning PCRs performed to detect SERs within PRS2 have been established previously (7), but those performed to identify breakpoints located within PRS1 were newly designed. The breakpoint-spanning PCR products (BSPs) were sequenced to determine the SERs that were identified by means of PSVs (Supplementary Material, Table S2). PSVs are non-polymorphic sequence differences between NF1-REPa and NF1-REPC. In order to identify the SERs, we also used SNPs with a MAF of  $\leq 1\%$  (Supplementary Material, Tables S3 and S4).

PSVs were distinguished from SNPs by cross-comparison with all variants located within these regions as identified by the 1000 Genomes Project (<http://www.1000genomes.org/home>) ([ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/ALL.chr17.phase1\\_release\\_v3.20101123.snps\\_indels\\_svsvs.genotypes.vcf.gz](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/ALL.chr17.phase1_release_v3.20101123.snps_indels_svsvs.genotypes.vcf.gz)) (34). Discrimination of PSVs from SNPs with a MAF of  $\geq 1\%$  is important in order to determine the SERs, and ascertain the gene conversion patterns, since polymorphisms with high MAFs are unsuitable for determining the origin of the respective nucleotides within BSPs.

Putative gene conversion events detected in the BSPs were further investigated by analysing the nucleotide patterns of the wild-type PRS1 or PRS2 sequences in the parents of the patients. Type 1 NF1 deletions arise in the germlines of otherwise healthy parents by meiotic NAHR affecting single germ cells. The parental origin of the deletions was determined by microsatellite marker analysis. Non-recombinant wild-type sequences from NF1-REPa and NF1-REPC encompassing the PRS2 or PRS1 regions were PCR-amplified from genomic blood-derived DNA of the healthy parents using the Expand Long Range dNTPack (Roche) and primers listed in Supplementary Material, Tables S13 and S14. Sequence analysis of the PCR fragments encompassing the wild-type PRS2 or PRS1 regions amplified from the transmitting parents enabled us to ascertain whether the gene conversion patterns detected in the breakpoint-flanking sequences of their children represented variant haplotypes present already in the parents prior to the NAHR events or whether these gene conversion patterns were associated with the deletion-causing NAHR. The phase of the haplotypes of the wild-type PRS1 and PRS2 regions in the parents was determined by cloning of the PCR products and sequence analysis of single clones.

### PRDM9 binding site prediction

Binding sites for the PRDM9 protein A- and C-variants were predicted within PRS1, PRS2 and 10 kb regions flanking these hotspots on either side. The extent and genomic positions of the analysed regions are indicated in Supplementary Material, Table S27 and Figure 1. To predict putative PRDM9 binding sites, we screened the reference sequence of the human genome (hg19) using the JavaScript software DNA Pattern Find, Sequence Manipulation Suites (Version 2) (<http://www.bioinformatics.org/>

[sms2/dna\\_pattern.html](#)) (76). Predicted binding sites for the PRDM9 A-variant exhibited a perfect match to the consensus motif 5' CCNCCNTNNCCNC 3' (31), whereas predicted binding sites for the PRDM9 C-variant showed 100% sequence identity to the consensus motif 5' CCNCCNTNNCCNTNNC 3' (32,33). The PRS1 region was also screened for the presence of PRDM9 A-variant binding sites exhibiting only seven of eight matches to the consensus binding motif 5' CCNCCNTNNCCNC 3'.

### Genotyping of PRDM9 alleles

To determine the genotype of PRDM9 in the parents of patients with type 1 NF1 deletions, we performed PCR using genomic blood-derived DNA and primers HsPrdm9-F3 (5' TGTAAGGAAT GACTGCCCTGA 3') and HsPrdm9-R1 (5' ATGTCCCCGAA CACTTACAGAA 3') (18). The PCR was performed with the Expand Long Range dNTPack (Roche). The primers used to fully sequence the PCR fragments of the PRDM9 alleles were PN1.2F (5' TGAATC CAGGGAACACAGGC 3') and PN2.4R (5' GCAAGTGTGTGGTGAC CACA 3') (19).

### Analysis of sequence homology

Sequence homology between NF1-REPa and NF1-REPC was determined by BLAST sequence alignments ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE\\_TYPE=BlastSearch&BLAST\\_SPEC=blast2seq&LINK\\_LOC=align2seq](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=align2seq)). The chromosomal positions of the regions analysed are listed in Supplementary Material, Table S27 and Figure 1. The location and extent of the regions exhibiting sequence identity were graphically displayed using the mVISTA tool (<http://genome.lbl.gov/vista/mvista/submit.shtml>).

### Recombination rates

The recombination rates in PRS1, PRS2 and their flanking regions were extracted from the Phase II HapMap data set compiling information from over 3.1 million SNPs genotyped in 270 individuals of African, Asian and European origin (77). Fine-scale recombination rates were determined from genetic variation patterns (78) and were downloaded from the HapMap website (<http://hapmap.ncbi.nlm.nih.gov/downloads/recombination/>), and converted to GRCh37 (hg19).

### GC content

The GC content of the genomic regions encompassing nucleotide positions 28 962 945–29 003 359 (according to hg19), including the PRS1 and PRS2 hotspots as well as 10 kb centromeric to PRS1 and 4 kb telomeric to PRS2, was analysed using the Visual Gene Developer 1.7 (<http://visualgenedeveloper.net/index.html>).

### Supplementary Material

Supplementary material is available at HMG online.

*Conflict of Interest statement.* None declared.

### Funding

This work was supported by the Deutsche Forschungsgemeinschaft (KE 724/12-1 given to H.K.-S.).

### References

1. Friedman, J.M. and Riccardi, V.M. (1999) Clinical and epidemiologic features. In Friedman, J.M., Gutmann, D.H., MacCollin, M. and Riccardi, V.M. (eds), *Neurofibromatosis: Phenotype,*

- Natural History and Pathogenesis*, 3rd edn. Johns Hopkins University Press, Baltimore, USA, pp. 29–86.
- Lammert, M., Friedman, J.M., Kluwe, L. and Mautner, V.F. (2005) Prevalence of neurofibromatosis 1 in German children at elementary school enrolment. *Arch. Dermatol.*, **141**, 71–74.
  - Kluwe, L., Siebert, R., Gesk, S., Friedrich, R.E., Tinschert, S., Kehrer-Sawatzki, H. and Mautner, V.F. (2004) Screening 500 unselected neurofibromatosis 1 patients for deletions of the NF1 gene. *Hum. Mutat.*, **23**, 111–116.
  - Kehrer-Sawatzki, H. and Cooper, D.N. (2012) NF1 microdeletions and their underlying mutational mechanisms, Chap. 14. In Upadhyaya, M. and Cooper, D.N. (eds), *Neurofibromatosis Type 1: Molecular and Cellular Biology*, Springer, Heidelberg, Germany, Vol. 1, pp. 187–211.
  - Wimmer, K., Yao, S., Claes, K., Kehrer-Sawatzki, H., Tinschert, S., De Raedt, T., Legius, E., Callens, T., Beiglböck, H., Maertens, O. et al. (2006) Spectrum of single- and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. *Genes Chromosomes Cancer*, **45**, 265–276.
  - De Raedt, T., Stephens, M., Heyns, I., Brems, H., Thijs, D., Messiaen, L., Stephens, K., Lazaro, C., Wimmer, K., Kehrer-Sawatzki, H. et al. (2006) Conservation of hotspots for recombination in low-copy repeats associated with the NF1 microdeletion. *Nat. Genet.*, **38**, 1419–1423.
  - Bengesser, K., Vogt, J., Mussotter, T., Mautner, V.F., Messiaen, L., Cooper, D.N. and Kehrer-Sawatzki, H. (2014) Analysis of crossover breakpoints yields new insights into the nature of the gene conversion events associated with large NF1 deletions mediated by nonallelic homologous recombination. *Hum. Mutat.*, **35**, 215–226.
  - Liu, P., Carvalho, C.M., Hastings, P.J. and Lupski, J.R. (2012) Mechanisms for recurrent and complex human genomic rearrangements. *Curr. Opin. Genet. Dev.*, **22**, 211–220.
  - Lupski, J.R. (2004) Hotspots of homologous recombination in the human genome: not all homologous sequences are equal. *Genome Biol.*, **5**, 242.
  - Liu, P., Lacia, M., Zhang, F., Withers, M., Hastings, P.J. and Lupski, J.R. (2011) Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over. *Am. J. Hum. Genet.*, **89**, 580–588.
  - Baudat, F., Imai, Y. and de Massy, B. (2013) Meiotic recombination in mammals: Localization and regulation. *Nat. Rev. Genet.*, **14**, 794–806.
  - Coutton, C., Abada, F., Karaouzene, T., Sanlaville, D., Satre, V., Lunardi, J., Jouk, P.S., Arnoult, C., Thierry-Mieg, N. and Ray, P.F. (2013) Fine characterisation of a recombination hotspot at the DPY19L2 locus and resolution of the paradoxical excess of duplications over deletions in the general population. *PLoS Genet.*, **9**, e1003363.
  - Elinati, E., Kuentz, P., Redin, C., Jaber, S., Vanden Meerschaut, F., Makarian, J., Koscinski, I., Nasr-Esfahani, M.H., Demiroglu, A., Gurgan, T. et al. (2012) Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots. *Hum. Mol. Genet.*, **21**, 3695–3702.
  - Lindsay, S.J., Khajavi, M., Lupski, J.R. and Hurles, M.E. (2006) A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. *Am. J. Hum. Genet.*, **79**, 890–902.
  - MacArthur, J.A., Spector, T.D., Lindsay, S.J., Mangino, M., Gill, R., Small, K.S. and Hurles, M.E. (2014) The rate of nonallelic homologous recombination in males is highly variable, correlated between monozygotic twins and independent of age. *PLoS Genet.*, **10**, e1004195.
  - Paigen, K. and Petkov, P. (2010) Mammalian recombination hot spots: Properties, control and evolution. *Nat. Rev. Genet.*, **11**, 221–233.
  - Turner, D.J., Miretti, M., Rajan, D., Fiegler, H., Carter, N.P., Blayney, M.L., Beck, S. and Hurles, M.E. (2008) Germline rates of *de novo* meiotic deletions and duplications causing several genomic disorders. *Nat. Genet.*, **40**, 90–95.
  - Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G. and de Massy, B. (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*, **327**, 836–840.
  - Berg, I.L., Neumann, R., Lam, K.W., Sarbajna, S., Odenthal-Hesse, L., May, C.A. and Jeffreys, A.J. (2010) PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat. Genet.*, **42**, 859–863.
  - Billings, T., Parvanov, E.D., Baker, C.L., Walker, M., Paigen, K. and Petkov, P.M. (2013) DNA binding specificities of the long zinc-finger recombination protein PRDM9. *Genome Biol.*, **14**, R35.
  - Grey, C., Barthès, P., Chauveau-Le Fric, G., Langa, F., Baudat, F. and de Massy, B. (2011) Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol.*, **9**, e1001176.
  - Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G. and Donnelly, P. (2010) Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*, **327**, 876–879.
  - Parvanov, E.D., Petkov, P.M. and Paigen, K. (2010) Prdm9 controls activation of mammalian recombination hotspots. *Science*, **327**, 835.
  - Berg, I.L., Neumann, R., Sarbajna, S., Odenthal-Hesse, L., Butler, N.J. and Jeffreys, A.J. (2011) Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. *Proc. Natl. Acad. Sci. USA*, **108**, 12378–12383.
  - Buard, J., Rivals, E., Dunoyer de Segonzac, D., Garres, C., Caminade, P., de Massy, B. and Boursot, P. (2014) Diversity of Prdm9 zinc finger array in wild mice unravels new facets of the evolutionary turnover of this coding minisatellite. *PLoS One*, **9**, e85021.
  - Jeffreys, A.J., Cotton, V.E., Neumann, R. and Lam, K.W. (2013) Recombination regulator PRDM9 influences the instability of its own coding sequence in humans. *Proc. Natl. Acad. Sci. USA*, **110**, 600–605.
  - Oliver, P.L., Goodstadt, L., Bayes, J.J., Birtle, Z., Roach, K.C., Phadnis, N., Beatson, S.A., Lunter, G., Malik, H.S. and Ponting, C.P. (2009) Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. *PLoS Genet.*, **5**, e1000753.
  - Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R.D. and Petukhova, G.V. (2012) Genetic recombination is directed away from functional genomic elements in mice. *Nature*, **485**, 642–645.
  - Pratto, F., Brick, K., Khil, P., Smagulova, F., Petukhova, G.V. and Camerini-Otero, R.D. (2014) DNA recombination. Recombination initiation maps of individual human genomes. *Science*, **346**, 1256442.
  - Dittwald, P., Gambin, T., Szafranski, P., Li, J., Amato, S., Divon, M.Y., Rodríguez Rojas, L.X., Elton, L.E., Scott, D.A., Schaaf, C.P. et al. (2013) NAHR-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and Mendelizing traits. *Genome Res.*, **23**, 1395–1409.

31. Myers, S., Freeman, C., Auton, A., Donnelly, P. and McVean, G. (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.*, **40**, 1124–1129.
32. Hinch, A.G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C.D., Chen, G.K., Wang, K., Buxbaum, S.G., Akylbekova, E.L. et al. (2011) The landscape of recombination in African Americans. *Nature*, **476**, 170–175.
33. Kong, A., Thorleifsson, G., Gudbjartsson, D.F., Masson, G., Sigurdsson, A., Jonasdottir, A., Walters, G.B., Jonasdottir, A., Gylfason, A., Kristinsson, K. et al. (2010) Fine-scale recombination rate differences between sexes, populations and individuals. *Nature*, **467**, 1099–1103.
34. 1000 Genomes Project Consortium Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
35. Alemany-Schmidt, A., Navarro-Palou, M., Voltes-Cobo, A., Rosell, J., Heine-Suñer, D., Picornell, A. and Oliver-Bonet, M. (2014) Geographical genetic variability: a factor to consider when assessing clinical implications of PRDM9. *Mol. Genet. Genomic Med.*, **2**, 201–203.
36. Borel, C., Cheung, F., Stewart, H., Koolen, D.A., Phillips, C., Thomas, N.S., Jacobs, P.A., Eliez, S. and Sharp, A.J. (2012) Evaluation of PRDM9 variation as a risk factor for recurrent genomic disorders and chromosomal non-disjunction. *Hum. Genet.*, **131**, 1519–1524.
37. Hussin, J., Sinnett, D., Casals, F., Idaghdour, Y., Bruat, V., Sallour, V., Healy, J., Grenier, J.C., de Malliard, T., Busche, S. et al. (2013) Rare allelic forms of PRDM9 associated with childhood leukemogenesis. *Genome Res.*, **23**, 419–430.
38. Woodward, E.L., Olsson, M.L., Johansson, B. and Paulsson, K. (2014) Allelic variants of PRDM9 associated with high hyperdiploid childhood acute lymphoblastic leukaemia. *Br. J. Haematol.*, **166**, 947–949.
39. ENCODE Project Consortium Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Fritze, S., Harrow, J. et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
40. López Correa, C., Brems, H., Lázaro, C., Marynen, P. and Legius, E. (2000) Unequal meiotic crossover: a frequent cause of NF1 microdeletions. *Am. J. Hum. Genet.*, **66**, 1969–1974.
41. Dihazi, H., Dihazi, G.H., Nolte, J., Meyer, S., Jahn, O., Müller, G. A. and Engel, W. (2009) Multipotent adult germline stem cells and embryonic stem cells: comparative proteomic approach. *J. Proteome Res.*, **8**, 5497–5510.
42. De Raedt, T., Brems, H., Lopez-Correa, C., Vermeesch, J.R., Marynen, P. and Legius, E. (2004) Genomic organization and evolution of the NF1 microdeletion region. *Genomics*, **84**, 346–360.
43. Waldman, A.S. (2008) Ensuring the fidelity of recombination in mammalian chromosomes. *Bioessays*, **30**, 1163–1171.
44. Dorschner, M.O., Sybert, V.P., Weaver, M., Pletcher, B.A. and Stephens, K. (2000) NF1 microdeletion breakpoints are clustered at flanking repetitive sequences. *Hum. Mol. Genet.*, **9**, 35–46.
45. Jenne, D.E., Tinschert, S., Reimann, H., Lasinger, W., Thiel, G., Hameister, H. and 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.
46. Forbes, S.H., Dorschner, M.O., Le, R. and Stephens, K. (2004) Genomic context of paralogous recombination hotspots mediating recurrent NF1 region microdeletion. *Genes Chromosomes Cancer*, **41**, 12–15.
47. López-Correa, C., Dorschner, M., Brems, H., Lázaro, C., Clementi, M., Upadhyaya, M., Dooijes, D., Moog, U., Kehrer-Sawatzki, H., Rutkowski, J.L. et al. (2001) Recombination hotspot in NF1 microdeletion patients. *Hum. Mol. Genet.*, **10**, 1387–1392.
48. Baker, C.L., Walker, M., Kajita, S., Petkov, P.M. and Paigen, K. (2014) PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Res.*, **24**, 724–732.
49. Arbeithuber, B., Betancourt, A.J., Ebner, T. and Tiemann-Boege, I. (2015) Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *Proc. Natl. Acad. Sci. USA*, **112**, 2109–2114.
50. Capra, J.A. and Pollard, K.S. (2011) Substitution patterns are GC-biased in divergent sequences across the metazoans. *Genome Biol. Evol.*, **3**, 516–527.
51. Duret, L. and Arndt, P.F. (2008) The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.*, **4**, e1000071.
52. Lesecque, Y., Mouchiroud, D. and Duret, L. (2013) GC-biased gene conversion in yeast is specifically associated with crossovers: molecular mechanisms and evolutionary significance. *Mol. Biol. Evol.*, **30**, 1409–1419.
53. Odenthal-Hesse, L., Berg, I.L., Veselis, A., Jeffreys, A.J. and May, C.A. (2014) Transmission distortion affecting human noncrossover but not crossover recombination: a hidden source of meiotic drive. *PLoS Genet.*, **10**, e1004106.
54. Williams, A.L., Genovese, G., Dyer, T., Altemose, N., Truax, K., Jun, G., Patterson, N., Myers, S.R., Curran, J.E., Duggirala, R. et al. (2015) Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. *Elife*, **25**, 4.
55. Bi, W., Park, S.S., Shaw, C.J., Withers, M.A., Patel, P.I. and Lupski, J.R. (2003) Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2. *Am. J. Hum. Genet.*, **73**, 1302–1315.
56. Peng, Z., Zhou, W., Fu, W., Du, R., Jin, L. and Zhang, F. (2015) Correlation between frequency of non-allelic homologous recombination and homology properties: evidence from homology-mediated CNV mutations in the human genome. *Hum. Mol. Genet.*, **24**, 1225–1233.
57. Borde, V., Robine, N., Lin, W., Bonfils, S., Géli, V. and Nicolas, A. (2009) Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J.*, **28**, 99–111.
58. Buard, J., Barthès, P., Grey, C. and de Massy, B. (2009) Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. *EMBO J.*, **28**, 2616–2624.
59. Smagulova, F., Gregoret, I.V., Brick, K., Khil, P., Camerini-Otero, R.D. and Petukhova, G.V. (2011) Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature*, **472**, 375–378.
60. Szekvolgyi, L., Ohta, K. and Nicolas, A. (2015) Initiation of meiotic homologous recombination: flexibility, impact of histone modifications, and chromatin remodeling. *Cold Spring Harb Perspect. Biol.*, **7**, a016527.
61. Yamada, S., Ohta, K. and Yamada, T. (2013) Acetylated histone H3K9 is associated with meiotic recombination hotspots, and plays a role in recombination redundantly with other factors including the H3K4 methylase Set1 in fission yeast. *Nucleic Acids Res.*, **41**, 3504–3517.
62. Zeng, J. and Yi, S.V. (2014) Specific modifications of histone tails, but not DNA methylation, mirror the temporal variation of mammalian recombination hotspots. *Genome Biol. Evol.*, **6**, 2918–2929.

63. Zhang, L., Ma, H. and Pugh, B.F. (2011) Stable and dynamic nucleosome states during a meiotic developmental process. *Genome Res.*, **21**, 875–884.
64. Szafranski, P., Schaaf, C.P., Person, R.E., Gibson, I.B., Xia, Z., Mahadevan, S., Wiszniewska, J., Bacino, C.A., Lalani, S., Potocki, L. et al. (2010) Structures and molecular mechanisms for common 15q13.3 microduplications involving CHRNA7: benign or pathological? *Hum. Mutat.*, **31**, 840–850.
65. Visser, R., Shimokawa, O., Harada, N., Kinoshita, A., Ohta, T., Niikawa, N. and Matsumoto, N. (2005) Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion. *Am. J. Hum. Genet.*, **76**, 52–67.
66. Abyzov, A., Li, S., Kim, D.R., Mohiyuddin, M., Stutz, A.M., Parrish, N.F., Mu, X.J., Clark, W., Chen, K., Hurles, M. et al. (2015) Analysis of deletion breakpoints from 1,092 humans reveals details of mutation mechanisms. *Nat. Commun.*, **6**, 7256.
67. Jeffreys, A.J., Kauppi, L. and Neumann, R. (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.*, **29**, 217–222.
68. Jeffreys, A.J. and Neumann, R. (2002) Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat. Genet.*, **31**, 267–271.
69. Jeffreys, A.J. and Neumann, R. (2005) Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum. Mol. Genet.*, **14**, 2277–2287.
70. Séguirel, L., Leffler, E.M. and Przeworski, M. (2011) The case of the fickle fingers: how the PRDM9 zinc finger protein specifies meiotic recombination hotspots in humans. *PLoS Biol.*, **9**, e1001211.
71. Walker, M., Billings, T., Baker, C.L., Powers, N., Tian, H., Saxl, R.L., Choi, K., Hibbs, M.A., Carter, G.W., Handel, M.A. et al. (2015) Affinity-seq detects genome-wide PRDM9 binding sites and reveals the impact of prior chromatin modifications on mammalian recombination hotspot usage. *Epigenetics Chromatin*, **8**, 31.
72. Neale, M.J. and Keeney, S. (2006) Clarifying the mechanics of DNA strand exchange in meiotic recombination. *Nature*, **442**, 153–158.
73. Keeney, S. (2008) Spo11 and the formation of DNA double-strand breaks in meiosis. *Genome Dyn. Stab.*, **2**, 81–123.
74. de Massy, B. (2013) Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes. *Annu. Rev. Genet.*, **47**, 563–599.
75. Baker, C.L., Petkova, P., Walker, M., Flachs, P., Mihola, O., Trachtulec, Z., Petkov, P.M. and Paigen, K. (2015) Multimer formation explains allelic suppression of PRDM9 recombination hotspots. *PLoS Genet.*, **11**, e1005512.
76. Stothard, P. (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, **28**, 1102–1104.
77. International HapMap Consortium Frazer, K.A., Ballinger, D. G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J. W., Boudreau, A., Hardenbol, P. et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
78. McVean, G.A., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R. and Donnelly, P. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science*, **304**, 581–584.