nature genetics

Conservation of hotspots for recombination in low-copy repeats associated with the *NF1* microdeletion

Thomas De Raedt[1], Matthew Stephens[2], Ine Heyns[1], Hilde Brems[1], Daisy Thijs[1], Ludwine Messiaen[3], Karen Stephens[4], Conxi Lazaro[5], Katharina Wimmer[6], Hildegard Kehrer-Sawatzki[7], Dominique Vidaud[8], Lan Kluwe[9], Peter Marynen[1] & Eric Legius[1]

Several large-scale studies of human genetic variation have provided insights into processes such as recombination that have shaped human diversity. However, regions such as low-copy repeats (LCRs) have proven difficult to characterize, hindering efforts to understand the processes operating in these regions. We present a detailed study of genetic variation and underlying recombination processes in two copies of an LCR (NF1REPa and NF1REPc) on chromosome 17 involved in the generation of *NF1* microdeletions and in a third copy (REP19) on chromosome 19 from which the others originated over 6.7 million years ago. We find evidence for shared hotspots of recombination among the LCRs. REP19 seems to contain hotspots in the same place as the nonallelic recombination hotspots in NF1REPa and NF1REPc. This apparent conservation of patterns of recombination hotspots in moderately diverged paralogous regions contrasts with recent evidence that these patterns are not conserved in less-diverged orthologous regions of chimpanzees.

Recombination in LCRs is particularly interesting because crossovers that occur between two nearby copies of an LCR (generally referred to as nonallelic homologous recombination; NAHR) result in deletion or duplication of the region between the copies and are often associated with genomic disorders such as Charcot-Marie-Tooth type 1A, hereditary neuropathy with liability to pressure palsies, Williams-Beuren syndrome and neurofibromatosis type 1 (NF1). These NAHR events have sometimes been observed to cluster in relatively short hotspots[1–3], but otherwise, little is known about recombination in LCRs, including the extent to which patterns of recombination are shared across different copies of the same LCR and whether hotspots for NAHR also correspond to hotspots for nonallelic homologous gene conversion (NAHGC) and/or allelic homologous recombination (AHR). In principle, one can obtain considerable insight into these questions through studying patterns of genetic variation in LCRs. However, because of high sequence identity, accurate genotyping of SNPs in LCRs is not straightforward, and when such SNPs are typed in high-throughput genotyping projects, the obtained genotypes are often inaccurate and show nonmendelian inheritance or lack of Hardy-Weinberg equilibrium. As a result, virtually no reliable data are available on population genetic variation in LCRs.

We therefore undertook a detailed sequencing-based study of the patterns of genetic variation in three copies of an LCR, of which two are associated with the *NF1* microdeletion. **Supplementary Figure 1** online gives a schematic overview of the *NF1* microdeletion region and its LCRs, NF1REPa and NF1REPc (on chromosome 17) and REP19 (on chromosome 19). Two types of recurrent *NF1* microdeletions have been described. Type I *NF1* microdeletions are typically 1.4 Mb in length and of meiotic origin[4]. In individuals with type I *NF1* microdeletions, the deletion breakpoints tend to cluster in two regions called PRS1 (ref. 5) and PRS2 (ref. 1) (**Supplementary Fig. 1**). Regions paralogous to PRS1 and PRS2 are also present in REP19 (**Supplementary Fig. 1**) but with about 22 kb of extra sequence inserted between them compared with NF1REPa and NF1REPc. Type II *NF1* microdeletions are less frequent, have a length of 1.2 Mb and are mitotic in origin; their breakpoints tend to cluster in a different segment of the LCRs[6,7]. **Supplementary Figure 1** gives an overview of the repeat content of the different LCRs and the location of the typed SNPs.

To better characterize NAHR in NF1REPa and NF1REPc, we accurately mapped the location of the breakpoints in 60 individuals with type I *NF1* microdeletions ascertained in an unbiased manner. The results (**Fig. 1**) confirmed the presence of two clusters of breakpoints: a larger cluster in PRS2 (3.4 kb, containing 40/60 breakpoints) and a smaller cluster in PRS1 (1.8 kb, containing 13/60 breakpoints). This suggests the presence of hotspots for the initiation of NAHR in PRS1 and PRS2, in either one or both of NF1REPa and NF1REPc.

The effect of NAHGC will be to 'insert' a small segment of one LCR into another LCR. Thus, frequent NAHGC would tend to lead to

[1]Department of Human Genetics, Catholic University Leuven, Leuven, Belgium. [2]Department of Statistics, University of Washington, Seattle, Washington, USA. [3]Laboratory Medical Genetics, Department of Genetics, University of Alabama, Birmingham, Alabama, USA. [4]Department of Medicine, University of Washington, Seattle, Washington, USA. [5]Institut Catala d'Oncologia–Instituto de Investigación Biomédica de Bellvitge, L'Hospitalet de Llobregat, Barcelona, Spain. [6]Department of Human Genetics, Medical University of Vienna, Vienna, Austria. [7]Department of Human Genetics, University of Ulm, Ulm, Germany. [8]Laboratoire de Génétique Moléculaire, Institut National de la Santé et de la Recherche Médicale U745, Université Paris 5, Paris, France. [9]Laboratory for Tumor Biology and Development and Malformation, University Hospital Eppendorf, Hamburg, Germany. Correspondence should be addressed to E.L. (Eric.Legius@uz.kuleuven.ac.be).

Received 6 July; accepted 16 October; published online 19 November 2006; doi:10.1038/ng1920

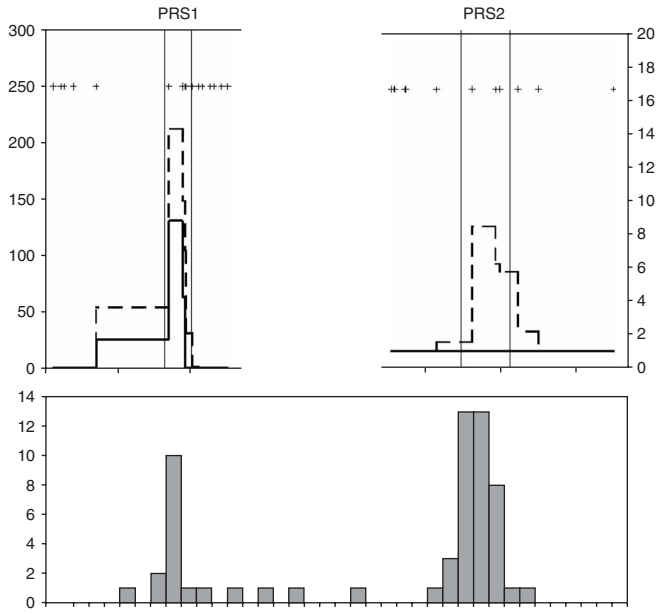

Figure 1 Comparison of hotspots of recombination inferred from haplotype and microdeletion data. Top: estimated rates of crossover across PRS1 and PRS2 in REP19 as calculated by PHASE. Lines show the posterior mean (dashed line) and median (full line) crossover rates from PHASE when both the location and intensity of any hotspot are estimated from the population haplotype data. The position of the SNPs (REP19) is represented by +. Vertical lines show locations of the NAHR breakpoint hotspots identified from the data in the bottom figure. Bottom: distribution of the NAHR breakpoints of 60 individuals with *NF1* microdeletions. The upper and lower panels are aligned: the estimated locations of AHR hotspots from the haplotype data on chromosome 19 agree closely with the locations of the NAHR breakpoint hotspots in the paralogous regions on chromosome 17. Note the very different scales of the *y* axes in the upper plot.

'shared' SNPs that are polymorphic in both LCRs. We therefore assessed relative rates of NAHGC in three regions (900 bp in PRS1, 1,500 bp in PRS2 and 800 bp in a region between PRS1 and PRS2) by sequencing these regions in both NF1REPa and NF1REPc (**Supplementary Table 1** online) and determining how many of the identified SNPs were 'shared' (that is, polymorphic in both NF1REPa and NF1REPc). In total, we identified 28 SNPs in NF1REPa and 26 in NF1REPc in the three specified regions. Of these, 12 were shared, providing strong evidence for NAHGC (the probability of 12 or more SNPs being shared if one draws 28 and 26 SNPs independently at random from 3,200 bp is $P = 2.2 \times 10^{-18}$). The region sequenced in PRS2 contained a higher proportion of shared SNPs (9/22) than the region sequenced in PRS1 (2/9), which contained a higher proportion than the intervening region (1/11; **Table 1**). Although only the difference between PRS2 and the intervening region approached statistical significance ($P = 0.066$), the results are consistent with the possibility that rates of NAHGC vary in a similar way to rates of NAHR; that is, the highest rate is found in PRS2, an intermediate rate in PRS1 and a lower rate in the intervening region, just as hotspots for allelic homologous gene conversion (AHGC) have been found to coincide with hotspots for AHR in the MHC and PAR1 regions[8]. This would be expected if NAHGC and NAHR were two alternative resolutions of a Holliday junction formed after a double-strand break, and if the hotspots for NAHR were due to an increased rate of double-strand break formation. Sequencing the same three regions in REP19, we did not find any SNPs shared with either NF1REPa or NF1REPc, and thus, as expected, there is no evidence for any ongoing genetic exchange between NF1REPa or NF1REPc and REP19.

In order to assess patterns of AHR, we first identified and typed SNPs in each of the three LCRs in 43 trios (parents plus child). We typed a total of 20 SNPs in the 33-kb region surrounding PRS1 and PRS2 in NF1REPa and NF1REPc. In REP19, where an additional 22 kb is found between PRS1 and PRS2, we typed 16 SNPs surrounding PRS1 and 12 SNPs surrounding PRS2 (**Supplementary Table 2** online; LD patterns are shown in **Supplementary Fig. 2** online). We then used a statistical approach[9–11] based on the observed haplotype data to assess the evidence for AHR hotspots at PRS1 and PRS2 in each of the LCRs. Specifically, as in ref. 11, we prespecified PRS1 and PRS2 as locations for potential hotspots and used the statistical method to estimate the background population-scaled recombination rate (ρ) and the relative intensity of recombination in the hotspot segments (λ). Thus, $\lambda = 1$ corresponds to no hotspot; we will refer to λ between 1 and 10 as a 'weak' hotspot and $\lambda > 10$ as a 'strong' hotspot. To summarize the strength of the evidence for either a strong or weak hotspot, we use the Bayes factor (BF), which is the probability of obtaining the data under the model that a weak or strong hotspot is present divided by the probability of obtaining the data under the model that no hotspot is present. Thus, a BF of >1 is evidence for a hotspot, and a BF of 10 means that the data are ten times more likely under a hotspot model than under a model without a hotspot, which might be considered strong evidence for a hotspot. The results are summarized in **Table 2**. One explanation for the different results for NF1REPa and NF1REPc (BFs ranging from 1.1–3.2) as compared with REP19 (BFs of 10.1 and 14,835) is that the polymorphism data for the former are comparatively uninformative about the presence or absence of a hotspot. This is suggested by results shown in **Supplementary Figure 3** online, which demonstrates little difference between the prior and posterior distribution of λ. In contrast, the polymorphism data in REP19 are substantially more informative and provide strong evidence for a hotspot of some kind in both PRS1 and PRS2 (BF $>$ 10,000 (strong hotspot) in PRS1, and BF $=$ 10.1 (weak hotspot) in PRS2). Further, when we repeated these analyses for the REP19 data without prespecifying the location of the hotspot and allowing the statistical method to estimate the location, the estimated hotspot locations corresponded very closely to the regions paralogous to PRS1 and PRS2 (**Fig. 1**). (The estimated positions of the AHR and NAHR hotspots are, respectively, 6973–9059 and 7682–9525 for PRS1 and 5567–7316 and 4394–7564 for PRS2.)

Table 1 Number of SNPs across the different regions of the LCRs

	PRS1 (900 bp)	PRS2 (1,500 bp)	Region between hotspots (800 bp)	Total (3,200 bp)
NF1REPa only	4	5	7	16
NF1REPc only	3	8	3	14
Shared NF1REPa/c	2	9	1	12
REP19 only	6	11	2	19
Shared 17/19	0	0	0	0

'NF1REPa only': total number of nucleotide positions polymorphic only in NF1REPa and not at the paralogous position in NF1REPc. 'NF1REPc only': total number of nucleotide positions polymorphic only in NF1REPc and not at the paralogous position in NF1REPa. 'Shared NF1REPa/c': number of nucleotide positions polymorphic in NF1REPa and at the paralogous position in NF1REPc. 'REP19 only': total number of nucleotide positions polymorphic only in REP19 and not at the paralogous position in NF1REPa and/or NF1REPc. 'Shared 17/19': total number of nucleotide positions polymorphic in REP19 and at the paralogous position in NF1REPa or NF1REPc.

Table 2 Background recombination rates and λ values in the PRS1 and PRS2 regions of the LCRs

		NF1REPa	NF1REPc	REP19
PRS1	P(λ > 1)[a]	0.707	0.564	1.00
	P(λ > 10)[a]	0.393	0.296	0.989
	BF(weak)[b]	3.2	1.8	327[d]
	BF(strong)[c]	2.0	1.0	14,835[d]
	Median λ	5.47	1.78	152
PRS2	P(λ > 1)	0.719	0.532	0.865
	P(λ > 10)	0.429	0.327	0.412
	BF(weak)[b]	3.1	1.3	10.1[d]
	BF(strong)[c]	2.2	1.1	4.6
	Median λ	7.23	1.66	7.99
Background recombination rate (90% confidence interval)		1.31×10^{-4} (2.3×10^{-5} to 3.3×10^{-4})	6.41×10^{-5} (1.0×10^{-5} to 1.7×10^{-4})	PRS1: 1.48×10^{-4} (3.8×10^{-5} to 5.9×10^{-4}) PRS2: 5.09×10^{-4} (1.9×10^{-4} to 1.1×10^{-3})

[a]Proportion of the 50,000 iterations with λ > 1 or λ > 10. [b]BF comparing the model 'weak hotspot' versus the model 'no hotspot'. [c]BF comparing the model 'strong hotspot' versus the model 'no hotspot'. [d]BF values >10: strong indication in favor of a hotspot.

The high informativeness of the population data in REP19 compared with NF1REPa and NF1REPc could be due to several factors, including the greater number of SNPs available (28 in REP19 versus 20 in NF1REPa and NF1REPc) and the higher average minor allele frequency (average minor allele frequency of 29.8% in REP19 compared with 18.0% in NF1REPa and 14.7% in NF1REPc)[12,13]. It is also possible that differences in the background recombination rates[14] in the different regions could contribute to different power to detect hotspots, as the estimated background recombination rate (ρ) differs in the three LCRs. The estimate of ρ is about two times higher in NF1REPa (1.31×10^{-4}) compared with NF1REPc (6.41×10^{-5}). The background recombination rate for REP19 PRS1 is 1.48×10^{-4} (comparable to NF1REPa), whereas the background in REP19 PRS2 is higher (5.09×10^{-4}). It is possible that, due to the formation of deleterious microdeletions, a selective pressure against recombination exists in the LCRs of chromosome 17, reducing the historical recombination signature in the region.

Although it is notable that hotspots for AHR in REP19 identified from population data seem to occur in the same locations as the PRS1 and PRS2 hotspots for NAHR identified from individuals with *NF1* microdeletions, it is also worth noting that the estimated relative intensities of the hotspots identified from the two approaches differ appreciably. Intensities of the NAHR hotspots, computed by comparing the frequency of deletion breakpoints occurring inside each hotspot versus outside the hotspots, are 28.7 for PRS1 and 46.7 for PRS2. (Note that these estimates are likely biased slightly upward because the locations of the breakpoint hotspots were based on the observed NAHR events.) In contrast, estimated intensities of these hotspots from the REP19 population data are 152 and 8, respectively. In addition, in REP19, the population data suggest a lower average rate of AHR in the region surrounding PRS2 than in the region surrounding PRS1 (**Supplementary Fig. 3**), whereas the microdeletion data suggest the reverse for NAHR in chromosome 17. Although some of these discrepancies could be due to imprecision of the estimated intensities and deviations from modeling assumptions, other factors could also have a role. For example, the NAHR data measure the current recombination rate (microdeletions being formed), whereas population genetic data reflect the average historical recombination rate, and differences can exist between current and historical recombination rates[15]. Other factors that could influence the relative intensity of the recombination hotspots include the fact that REP19 contains 22 kb of additional sequence inserted between PRS1 and PRS2, or that differences in methylation pattern could result in a different DNA folding in each LCR (epigenetic effects).

In summary, the mapping of *NF1* microdeletion breakpoints (**Fig. 1**) and the presence of NAHGC give a clear view that sites of NAHR in (at least one of) NF1REPa and NF1REPc are not randomly distributed and are located in two distinct hotspots. Population genetic data in both NF1REPa and NF1REPc are not very informative for a hotspot of AHR but do not argue strongly against the presence of hotspots in either NF1REPa or NF1REPc. Population genetic data in the paralogous regions of REP19 provide strong evidence for hotspots of AHR. The initial duplication of the LCR from chromosome 19 to the *NF1* region took place before the human-gorilla split, about 6.7 million years ago[16]. Conservation of AHR hotspots in REP19 with NAHR hotspots in NF1REPa and NF1REPc in sequences that have been evolving separately for such a long time is notable, as hotspots for recombination in humans do not seem to be highly conserved in chimpanzee[11,17]. Others[18] have investigated the fine-scale recombination rate of three 500-kb regions both in humans and chimpanzees. Eighteen hotspots for recombination were found in humans, and of these, only one may correspond to a site of historical recombination in chimpanzees. Although it could be that we have simply found hotspots that happen to be conserved by chance, it is also possible that patterns of recombination in orthologous regions between humans and chimpanzees are less evolutionarily stable than patterns of recombination in paralogous regions. This could be due, for example, to differences in the recombination machinery or in epigenetic factors in the two species[17,18]. This potential role in hotspot evolution for distal regulators of recombination or epigenetic factors has been suggested before when looking at the presence or absence polymorphism of a human recombination hotspot[19]. Examining whether the AHR hotspots in PRS1 and PRS2 are present in the chimpanzee orthologs could help distinguish among these competing explanations.

METHODS

Samples. We collected DNA samples extracted from blood leukocytes from 43 nuclear families (trios). In addition, genomic DNA from 70 individuals was available for full sequencing of three regions in each LCR investigated. All individuals were of Flemish origin and gave their informed consent, and all experiments were approved by the ethical commission of the medical faculty of the Catholic University of Leuven.

Specific PCR amplification of sequences from the three different copies of the LCRs. Using the Expand Long Template PCR system (Roche), we amplified each copy of the LCR in fragments of a maximum of 5 kb. Primers were chosen to end on paralogous sequence variants (PSVs) in such a way that only one copy of the LCR would be amplified. Genotypes from some SNPs showed problems with mendelian inheritance and Hardy-Weinberg equilibrium. Primer positions of the relevant PCR fragments were sequenced in all individuals, and the PSV sites used for the LCR-specific amplification proved to be polymorphic and hence not LCR specific. In these cases, new primers were designed using other PSVs in the vicinity that were not polymorphic in

our set of 43 families. With the new primer sets, there were no longer problems with mendelian inheritance or Hardy-Weinberg Equilibrium. A complete list of primers is available in **Supplementary Table 3** online.

Detection of type I *NF1* microdeletion breakpoints. We amplified, nonspecifically, fragments from both NF1REPa and NF1REPc simultaneously during the same PCR reaction and sequenced PSVs at several sites throughout the LCR. In an individual with a microdeletion, one typically amplifies three different fragments, one derived from the normal NF1REPa, one derived from the normal NF1REPc and one derived from the hybrid NF1REPa/NF1REPc on the chromosome 17 with the microdeletion. By simply scoring the relative copy number of both nucleotides of the PSV, one can determine if the breakpoint is located centromeric of the PSV investigated (NF1REPc-specific nucleotide has a higher relative intensity on sequencing readout) or telomeric of the PSV (NF1REPa-specific nucleotide has a higher relative intensity) (**Supplementary Fig. 4** online). In this way, the interval where the breakpoint is located can be narrowed to an interval between two PSVs.

SNP discovery. In NF1REPa and NF1REPc, 31 kb of the 33 kb region investigated in the three LCRs were fully sequenced, and about 22 kb were sequenced in REP19 (BigDye Terminator Sequencing Kit v3.1, Applied Biosystems) in ten random individuals for SNP discovery. Regions left unsequenced in NF1REPa and NF1REPc had a high repetitive content. All SNPs typed in the 43 families were submitted to the dbSNP database. **Supplementary Table 2** gives an overview of all the SNPs typed in the different LCRs together with the minor allele frequency. In addition, three smaller regions in the different LCRs were fully sequenced in 70 unrelated individuals (900 bp in PRS1, 1,500 bp in PRS2 and 800 bp in a region between PRS1 and PRS2). The number of SNPs shared between the different copies of the LCRs was counted. **Supplementary Table 1** gives an overview of the total number of SNPs and the number of SNPs shared between LCRs.

SNP typing. A total of 68 SNPs (20 in NF1REPa and NF1REPc, 16 in the PRS1 region of REP19 and 12 in the PRS2 region of REP19) were typed in the 43 nuclear families and were submitted to the dbSNP database (**Supplementary Table 2**). The SNPs were typed by direct sequencing (BigDye Terminator Sequencing Kit v3.1, Applied Biosystems) or by SNaPshot analysis (Applied Biosystems) on the LCR-specific amplified PCR products. Because of the high sequence identity between the different copies of the LCR and the inherent problems of specific amplification of the LCRs, extensive precautions were taken to avoid mistyping. None of the SNPs in this report showed problems with mendelian inheritance, and all genotypes are in Hardy-Weinberg equilibrium. SNPs with a frequency lower than 4% were not included in the data analysis.

Haplotype analysis. For each copy of the LCR, the 172 parental haplotypes were estimated from the genotype data of the 43 trios using PHASE v2.1.1 (refs. 9,10,20–22). These estimated haplotypes were assumed to be known without error for subsequent analyses. We then used PHASE to assess the evidence for the presence of AHR hotspots in the three LCRs (NF1REPa, NF1REPc and REP19) at the location of the NAHR hotspots (PRS1 and PRS2) in NF1REPa and NF1REPc. The boundaries of PRS1 and PRS2 were chosen in such a way that they encompassed 90% of the breakpoints located in the PRS1 and PRS2 area, respectively. Based on the haplotype data, PHASE estimates the overall background recombination rate (ρ) and the relative intensity of any hotspots (λ). For example, $\lambda = 1$ corresponds to no hotspot, whereas $\lambda = 10$ corresponds to a hotspot where recombination occurs at a rate 10 times higher than in the surrounding sequence. The program makes a number of iterations (for example 10,000), and each iteration produces a sampled value of ρ and λ. We assumed the following prior distribution on λ: with probability 0.5, $\lambda = 1$ (that is, no hotspot); otherwise a hotspot is present ($\lambda > 1$) and $\log10(\lambda)$ is uniform in the range 0–3. The number of iterations that provide a sampled value of $\lambda > 1$ indicates the strength of the evidence for a hotspot. For our analysis, we used five different seeds, each time making 10,000 iterations, to provide a total of 50,000 samples. The numbers shown in **Table 2** reflect the proportion of iterations where $\lambda > 1$ or $\lambda > 10$, with $1 < \lambda < 10$ being considered a weak hotspot, and $\lambda > 10$ being considered a strong hotspot. For the prior on λ we assumed (see above) the prior probability of no hotspot is 0.5, the prior probability of a weak hotspot λ is 1/6 and the prior probability of a strong hotspot is 1/3. The evidence for a weak or strong hotspot is measured using the Bayes factor (BF). The BF is the probability of obtaining the data under the model that a weak or strong hotspot is present divided by the probability of obtaining the data under the model that no hotspot is present. A BF of 10 indicates that the data are ten times more likely to be obtained if there is a hotspot than if there is no hotspot. The BF is calculated as follows: $BF(\text{weak}) = (Pr(\text{weak hotspot}) \times 0.5)/(Pr(\lambda = 1) \times 1/6)$, and $BF(\text{strong}) = (Pr(\text{strong hotspot}) \times 0.5)/(Pr(\lambda = 1) \times 1/3)$, where $Pr(x)$ is the percentage of iterations observed under condition x, and 0.5, 1/6 and 1/3 are the prior probabilities of finding 'no hotspot', a 'weak hotspot' or a 'strong hotspot', respectively.

For the REP19PRS1 and REP19PRS2 regions, a second analysis with PHASE was performed, assuming that one AHR hotspot was present in each region without specifying the exact location of the AHR hotspot.

Statistical analysis. The number of observed shared SNPs between the different copies of the LCR was compared with the expected number by the binomial distribution. The *a priori* probability was calculated as the number of SNPs per bp present in the copy of the LCR to which the other LCR was compared. The α-level was Bonferroni corrected.

Statistical analysis software and URLs. PHASE can be found at http://www.stat.washington.edu/stephens/software.html. All input files for PHASE and all derived haplotypes are available in **Supplementary Table 4** online.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The authors thank T. de Ravel for critically reading the manuscript. T.D. is supported by the Emmanuel Vanderschueren Fonds. M.S. is supported by US National Institutes of Health grant 1RO1HG/LM02585-01. E.L. is a part-time clinical researcher of the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO). This work is also supported by research grants from the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (G.0096.02, E.L.; G.0507.04, P.M.); the Interuniversity Attraction Poles (IAP) granted by the Federal Office for Scientific, Technical and Cultural Affairs, Belgium (2002-2006; P5/25) (P.M. and E.L.); by a Concerted Action Grant from the Catholic University of Leuven and by the Belgian Federation against Cancer (SCIE2003-33 to E.L.).

AUTHOR CONTRIBUTIONS

This study was coordinated by T.D., P.M. and E.L.; the manuscript was written by T.D., M.S. and E.L.; breakpoint detection was performed by T.D., I.H., H.B., L.M., K.S., C.L., K.W., H.K., D.V. and L.K.; SNP detection and typing was performed by T.D., I.H., H.B. and D.T. and computational analysis was performed by T.D. and M.S.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Lopez-Correa, C. *et al.* Recombination hotspot in NF1 microdeletion patients. *Hum. Mol. Genet.* **10**, 1387–1392 (2001).
2. Repping, S. *et al.* Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am. J. Hum. Genet.* **71**, 906–922 (2002).
3. Reiter, L.T. *et al.* Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. *Am. J. Hum. Genet.* **62**, 1023–1033 (1998).
4. Lopez Correa, C., Brems, H., Lazaro, C., Marynen, P. & Legius, E. Unequal meiotic crossover: a frequent cause of NF1 microdeletions. *Am. J. Hum. Genet.* **66**, 1969–1974 (2000).
5. Forbes, S.H., Dorschner, M.O., Le, R. & Stephens, K. Genomic context of paralogous recombination hotspots mediating recurrent NF1 region microdeletion. *Genes Chromosom. Cancer* **41**, 12–25 (2004).
6. Petek, E. *et al.* 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 (2003).
7. Kehrer-Sawatzki, H. *et al.* 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 (2004).

8. Jeffreys, A.J. & May, C.A. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nat. Genet.* **36**, 151–156 (2004).
9. Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233 (2003).
10. Crawford, D.C. *et al.* Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.* **36**, 700–706 (2004).
11. Ptak, S.E. *et al.* Absence of the TAP2 human recombination hotspot in chimpanzees. *PLoS Biol.* **2**, e155 (2004).
12. Fearnhead, P., Harding, R.M., Schneider, J.A., Myers, S. & Donnelly, P. Application of coalescent methods to reveal fine-scale rate variation and recombination hotspots. *Genetics* **167**, 2067–2081 (2004).
13. Fearnhead, P. & Smith, N.G. A novel method with improved power to detect recombination hotspots from polymorphism data reveals multiple hotspots in human genes. *Am. J. Hum. Genet.* **77**, 781–794 (2005).
14. Hudson, R.R. Two-locus sampling distributions and their application. *Genetics* **159**, 1805–1817 (2001).
15. Jeffreys, A.J., Neumann, R., Panayi, M., Myers, S. & Donnelly, P. Human recombination hot spots hidden in regions of strong marker association. *Nat. Genet.* **37**, 601–606 (2005).
16. De Raedt, T. *et al.* Genomic organization and evolution of the NF1 microdeletion region. *Genomics* **84**, 346–360 (2004).
17. Ptak, S.E. *et al.* Fine-scale recombination patterns differ between chimpanzees and humans. *Nat. Genet.* **37**, 429–434 (2005).
18. Winckler, W. *et al.* Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* **308**, 107–111 (2005).
19. Neumann, R. & Jeffreys, A.J. Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation. *Hum. Mol. Genet.* **15**, 1401–1411 (2006).
20. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
21. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**, 1162–1169 (2003).
22. Marchini, J. *et al.* A comparison of phasing algorithms for trios and unrelated individuals. *Am. J. Hum. Genet.* **78**, 437–450 (2006).