Allelic recombination and *de novo* deletions in sperm in the human β-globin gene region

Kim Holloway* Victoria E. Lawson and Alec J. Jeffreys+

Department of Genetics,
University of Leicester,
Leicester LE1 7RH,
UK

*present address: Department of Biomedical Sciences, Cornell University, Ithaca, NY, 14853, USA; email jkh44@cornell.edu.

+ to whom correspondence should be addressed.
tel. +44 116 2523435
fax. +44 116 2523378
email ajj@le.ac.uk

ABSTRACT

Meiotic recombination is of fundamental importance in creating haplotype diversity in the human genome and has the potential to cause genomic rearrangements by ectopic recombination between repeat sequences and through other changes triggered by recombination-initiating events. However, the relationship between allelic recombination and genome instability in the human germline remains unclear. We have therefore analysed recombination and DNA instability in the δ-, β-globin gene region and its associated recombination hotspot. Sperm typing has for the first time accurately defined the hotspot and shown it to be the most active autosomal crossover hotspot yet described, although unusually inactive in non-exchange gene conversion. The hotspot just extends into a homology block shared by the δ- and β-globin genes, within which ectopic exchanges can generate Hb Lepore deletions. We developed a physical selection method for recovering and validating extremely rare *de novo* deletions in human DNA and used it to characterise the dynamics of these Hb Lepore deletions in sperm as well as other deletions not arising from ectopic exchanges between homologous DNA sequences. Surprisingly, both classes of deletion showed breakpoints that avoided the β-globin hotspot, establishing that it possesses remarkable fidelity and does not play a significant role in triggering these DNA rearrangements. This study also provides the first direct analysis of *de novo* deletion in the human germline and points to a possible deletion-controlling element in the β-globin gene separate from the crossover hotspot.

INTRODUCTION

The study of patterns of human DNA diversity is of fundamental importance in the search for optimal methods of association mapping, in the investigation of population history and in understanding the processes by which haplotypes are generated within populations. A key factor is the presence of meiotic recombination hotspots in the human genome, which serve to disrupt linkage disequilibrium (LD), creating haplotype blocks between hotspots which often extend for tens or even hundreds of kilobases (1-7).

Hotspots have in the past been inferred from studies of crossovers in families, both in the β-globin gene region (8) and elsewhere in the human genome (9-12). However, this approach is limited by the very small number of crossovers that can be identified per hotspot. A less direct but far more powerful approach is to infer patterns of historical crossover from contemporary patterns of haplotype diversity in human populations. Simple analyses of LD (1,2), as well as LD mapping approaches (13), have revealed putative hotspots as regions of localised breakdown in marker association. Coalescent approaches have been used to infer historical recombination rates from DNA diversity information and have shown that the bulk of historical crossovers across the human genome have been focussed within narrow hotspots spaced on average 50 kb apart (7,14,15).

The alternative approach to defining crossover hotspots is by direct analysis in sperm

(6,16). This approach circumvents potential problems of factors other than recombination, such as genetic drift, selection, admixture and migration, that can influence patterns of haplotype diversity and thus inferences of underlying population recombination rates (17,18). It also permits the analysis of contemporary crossover rates in individual men, allowing crossover rate polymorphisms to be identified which in turn can give clues about the evolution of hotspot activity. Analysis of single sperm has been used to study crossover rates and distributions around the β-globin gene (19) and elsewhere (20-23), but is unsuitable for screening the millions of sperm required for fully defining hotspot activity and morphology. The alternative approach is batch screening of sperm DNA (2,6) by allele-specific PCR to selectively amplify crossover molecules. This approach has been applied to the MHC class II region (2), to an interval on chromosome 1q42.3 (24) and to the pseudoautosomal *SHOX* gene (25), revealing crossover hotspots of widely varying intensities but of a constant 1-2 kb width that generally map to intervals of LD breakdown identified as putative hotspots by coalescent analysis. This correspondence between sperm hotspots and patterns of LD is however not perfect, with an example of an interval of LD breakdown that does not contain a sperm hotspot (26), as well as sperm hotspots that have left little if any mark on patterns of DNA diversity (24). The latter suggest that some hotspots may have arisen recently in humans and is consistent with major differences in LD landscape between chimpanzees and humans that imply rapid hotspot turnover during recent evolution (27,28).

Sperm crossover hotspots almost certainly mark sites of recombination initiation, as shown by their activity in gene conversion without exchange of flanking markers (29).

Some but not all hotspots show recombination resolution strongly biased towards conversion (29,30). Polymorphism between men in recombination initiation rates can be detected both by simple rate measurements and more powerfully using the crossover asymmetry test, in which reciprocal crossovers in sperm arise at the same rate but map to different locations within the hotspot (31). Such rate polymorphisms are common (30,31) and have been used to demonstrate that crossovers are also generally associated with conversion of markers near the site of exchange. In two cases, a single nucleotide polymorphism (SNP) within the hotspot has been identified that seems to influence crossover initiation rates directly (30,31). The recombination-suppressing SNP allele disrupts a putative crossover-promoting motif preferentially found in hotspots (15) and is strongly overtransmitted to recombinant progeny, creating a level of meiotic drive that can be sufficient to promote population fixation of recombination suppressing alleles (30,31). This raises the paradoxical issue of how hotspots can arise and persist in the face of such a deterministic drive to attenuation/extinction (32).

Crossover hotspots drive allelic exchanges at meiosis. What is not clear is whether they can also promote unequal crossover (ectopic exchange, non-allelic homologous recombination) between related DNA sequences or more generally drive genomic rearrangements such as deletions and duplications triggered by recombination-initiating double-strand DNA breaks. Such genome instability is of great importance given the prevalence of segmental duplications in the human genome (33,34), some of which are associated with copy number variation (35,36), and the likely role that ectopic recombination plays in their generation. Such structural variation can have pathological consequences (37); for example copy number variation in the α-globin gene cluster can cause α-thalassaemia (38) and exchanges between CMT1A-REP elements lying 1.5 Mb apart on chromosome 17 result in duplication or deletion of the 1.5 Mb region and cause Charcot-Marie-Tooth type 1A disease (CMT1A) and hereditary neuropathy with liability to pressure palsies (HNPP) respectively (39,40). Unequal exchanges between CMT1A-REP repeats cluster within a 1 kb hotspot, similar in width to hotspots of allelic exchange (40), and such clustering of unequal exchange points within dispersed repeats appears to be a common feature of such genomic disorders (37). However, it is not known whether hotspots for ectopic exchange such as CMT1A-REP are also hotspots for allelic crossover, nor whether hotspots for allelic recombination can trigger DNA rearrangements.

To address these issues, we focussed on the β-globin gene region on chromosome 11 and its associated recombination hotspot. This was the first hotspot to be provisionally identified, with breakdown of LD over a 9 kb interval immediately 5′ to the β-globin gene suggesting a localised hotspot 3-30 times more active than the genome average rate of crossover (41). Three crossovers in this region detected in families (one maternal and two paternal) all mapped to a 1.5 kb interval in the same region (8) but could not define the extent of the hotspot. Flow-sorted single-sperm typing identified an 11 kb interval with an 80-fold enhancement of recombination activity over the genome average rate and with a very high overall crossover frequency of $9x10^{-3}$, but did not refine the location or morphology of the hotspot (19). Most recently, coalescent analysis of genotype data has provided further evidence for a hotspot about 1.7-2.0 kb wide near the beginning of the β-globin gene, with a historical recombination frequency of $5.7x10^{-3}$ (42) and a peak activity estimated at 46 cM/Mb (27). This human hotspot is not discernible in the LD landscape of either rhesus macaques (42) or chimpanzees (27) and may therefore have evolved recently. The β-globin gene region is also prone to ectopic exchanges between a pair of homology blocks shared by the δ- and β-globin genes. These exchanges result in Hb Lepore deletions creating a δ-β fusion gene and anti-Lepore duplications generating an additional reciprocal β-δ fusion gene (43). Other deletions in this region not driven by ectopic exchanges between homologous DNA sequences have been identified in β-thalassaemia patients (44).

We have therefore characterised the LD landscape in the δ-, β-globin gene region and have used this as a guide to analyse allelic crossover and gene conversion activity in sperm. This was then combined with the development of new methods to characterise very rare *de novo* DNA deletions in sperm. The goal was to define the hotspot at very high resolution and to use this information to investigate the relationship between hotspot activity and the dynamics of DNA deletion.

RESULTS

SNP discovery and LD landscape in the δ-, β-globin gene region

Sperm recombination analysis requires a very high density of SNP markers. Twenty SNPs were identified in a 20 kb interval spanning the δ- and β-globin genes, 13 from previous literature (8,45,46) plus an additional 7 from dbSNP (47). This interval was also completely resequenced in 5 semen donors of north European origin, revealing an additional 22 novel SNPs, a surprisingly large number for such a well-studied region of the human genome. The LD landscape was investigated by genotyping all 42 SNPs in a panel of 50 unrelated semen donors of north European origin. Pairwise analysis of LD (2) between high frequency (MAF > 0.15) SNPs revealed a simple picture showing two blocks of intense LD separated by a 1.3 kb wide interval of relatively free marker association (Fig. 1A). This interval lies immediately 5′ of the β-globin gene and coincides with the putative hotspot identified from coalescent analysis (42), family crossovers (8) and single sperm typing (19).

Analysis of crossover molecules in sperm

By genotyping a selected set of 23 SNPs over an additional panel of 48 donors to maximise the number of informative donors for recombination analysis, we identified two men carrying suitable SNP heterozygosities for crossover recovery and mapping across a 13.5 kb interval spanning the putative hotspot. Nested repulsion-phase allele-specific long PCR was used to selectively amplify recombinant molecules directly from batches of sperm DNA (2,48). Each man was assayed for reciprocal (orientation A and B) crossovers (Fig. 1B); given the different haplotypes carried by these two men, this resulted in four different crossover assays.

Crossover molecules were detected in sperm but not in blood DNA (Fig. 1C), establishing their meiotic specificity. The crossover frequency was $1.4\text{x}10^{-3}$ per sperm in man 1 and $1.5\text{x}10^{-3}$ in man 2. There was no significant difference in crossover frequency between these two men ($P = 0.6$), nor between orientation A and B frequencies in each man ($P > 0.07$), consistent with reciprocal crossover. Mapping of crossover breakpoints by typing SNPs within the test interval showed that they all clustered into a narrow region located between the haplotype blocks, with a distribution that, as for other hotspots (2,24), appeared approximately normal with a width of 1.2 kb within which 95% of crossovers occur (Fig. 2A). Crossover distributions were similar for both men, with centres separated by only 46 bp and located about 1.0 kb upstream of the β-globin cap site. The hotspot is very active, with a peak activity of 200 cM/Mb in both man. In contrast, the 10 kb of flanking DNA is recombinationally inert, with no crossovers seen among 193,000 sperm, giving a mean crossover rate of <0.16 cM/Mb ($P > 0.95$) consistent with previous estimates of sperm crossover activity outside hotspots (2,24).

Analysis of separate A- and B-orientation crossovers in both men showed no evidence for reciprocal crossover asymmetry (Fig. 2B), in which A and B crossover distributions are centred at different locations (31). This suggests that both haplotypes in each man are equally efficient at initiating crossover (31). However, both men showed crossovers in one orientation somewhat more diffusely distributed than in the other orientation, though centred at the same location; for example, man 2 shows A crossovers more broadly spread than B crossovers. The result is that alleles at SNPs F0 and 3586 located closest to the centre of the hotspot are transmitted to crossover progeny with a ratio that is not significantly different from 50:50 (P = 0.4). In contrast, markers near the edge of the hotspot show distorted transmission, with transmission ratios of 46:54 for marker K20 to the left of centre, and 56:44 for markers F1 or F2 to the right. These distortions are of marginal significance (P = 0.01-0.1) and is seen for marker F2 in man 1 but not man 2. However, their detection in both men suggests that this unusual phenomenon, very different from the reciprocal crossover asymmetry described previously (30,31), may be real. If so, this indicates that biased gene conversion is occurring at the hotspot boundaries during crossover.

Gene conversion analysis at the β-globin hotspot

The alternative outcome of meiotic recombination is gene conversion without exchange of flanking markers. Previous analyses of four human crossover hotspots have shown that they are also active in gene conversion, particularly at the centre of the hotspot, but that conversion tracts are very short (mean length 60-300 bp) and require heterozygous SNPs very close to the hotspot centre for detection (29,30). We therefore chose man 2 for conversion analysis, given the presence of marker F0 approximately 20 bp from the centre of the hotspot (Fig. 2B). Conversions were assayed as described previously (29), using allele-specific PCR to amplify one haplotype from small pools of sperm DNA containing only 20-30 amplifiable molecules per pool, followed by allele-specific oligonucleotide (ASO) hybridisation to determine whether any pool contained markers from the other haplotype (Fig. 3A). This approach simultaneously detects crossovers and conversions.

Screening 9600 molecules in the orientation shown in Fig. 3B, plus 2400 molecules in the opposite orientation using allele-specific primers directed to the other haplotype, yielded 35 crossovers. The crossover frequency (2.9×10^{-3} per sperm) was significantly higher than the estimate of 1.5×10^{-3} in this man from crossover assays ($P < 0.001$). The difference in frequencies is however modest and most likely reflects subtle variation in the efficiency of PCR amplification between the two assays. There was no significant difference in crossover distribution across the hotspot as compared with data from conventional crossover assays (Fig. 2A) (Fisher exact test, $P > 0.05$). In contrast, no non-exchange gene conversion events were seen, even for marker F0 closest to the hotspot centre (Fig. 3B). The conversion rate must therefore be low ($<2.5 \times 10^{-4}$, $P > 0.95$) with a correspondingly low ratio of conversions to crossovers (<1:12).

Isolation of sperm deletion mutants in the δ-, β-globin gene region

The β-globin crossover hotspot extends into the beginning of the β-globin gene (Fig. 2A), into a region where the δ- and β-globin genes share a 570 bp region of homology and within which unequal exchanges can occur, leading to Hb Lepore deletions and anti-Lepore duplications. To determine whether recombination initiation events within the hotspot promote rearrangements such as Hb Lepore-type exchanges and other deletions (Fig. 4A), we used a size-enrichment technique originally developed for isolating minisatellite mutants (49) to recover deletion molecules directly from sperm DNA. The rate of such deletions was completely unknown, though was likely to be very low given the rarity of Hb Lepore in most populations. The target region chosen was located entirely within the interval analysed for sperm crossovers (Fig. 2A) and thus contains only the β-globin hotspot.

Large amounts of sperm DNA (800 μg DNA from two ejaculates) from a third man (man 3) were digested with *Eco*RV, to release the δ-, β-globin gene region on a 15.5 kb DNA fragment, then fractionated by agarose gel electrophoresis to recover size fractions that could contain Lepore-type deletions (8.1 kb molecules) and exclude progenitor DNA. Broken DNA molecules terminating within either of the homology blocks shared by δ- and β-globin genes could present a potential problem by promoting the formation of artefact Lepore deletions via strand annealing during PCR (Fig. 4Aiii). Since these molecules are all less than 6.3 kb long, the risk of such artefacts was minimised by restricting DNA fractions to the size range 6-14 kb (Fig. 4Bi).

The location of Lepore-sized deletions within fractionated DNA was monitored using a control 8.0 kb *Eco*RV genomic DNA fragment from the MHC matched in size to these deletions (Fig. 4Bii). PCR analysis of this control fragment showed a distribution across the size fractions as predicted from its size and indicated that a total of 8.0×10^7 amplifiable DNA molecules had been recovered. Similar analysis of the progenitor δ-, β-globin DNA fragment (Fig. 4Biii) showed major depletion in all except the largest DNA fraction. Only 5000 progenitor molecules remained in the three fractions that could contain Lepore-type deletions, indicating that 99.994% of progenitor molecules had been eliminated.

All size fractions except the largest were screened by nested long PCR to identify possible deletion mutants. The risk of jumping PCR artefacts was further minimised by using long extension times (15 min) and low extension temperatures (64^{o}) to maximise the efficiency of strand extension. Inputs of fractionated DNA were limited to levels known to be fully compatible with recovery of single molecules by PCR (not shown). All nine size fractions analysed were thus surveyed over a total of 860 PCR reactions. Each reaction showed progenitor PCR products plus a low level of Hb Lepore deletion artefact (Fig. 4C) that only appeared late during the nested PCR (not shown). Seven PCR reactions showed in addition a strong PCR signal from a putative deletion mutant. Four of these mutants (m1-m4) were the same size as Lepore deletions, while two were larger (m5, m7) and one smaller (m6). Similar analysis of 800 μg sperm DNA from a second man (man 4) yielded three mutants, none of which was identical in size to Lepore deletions.

Ten mutants were thus recovered from 1.6×10^{8} amplifiable molecules surveyed in the two men. The overall sperm deletion rate in this region, for deletions losing 3.5-9.4 kb DNA (larger or smaller deletions would have been excluded from the size fractions tested) was therefore extremely low, at 6.2×10^{-8} per sperm (95% CI $2.6\text{-}12.4 \times 10^{-8}$). Despite this exceedingly low rate, there is good evidence that these mutants are not PCR artefacts arising from progenitor molecules or broken DNA. First, all showed similar PCR signals far stronger than the very uniform low levels of Lepore deletion artefacts (Fig. 4C), especially at lower cycle numbers during nested PCR (not shown). In contrast, artefacts arising early during PCR should show a gradient of intensity depending on when the artefact arose. Second, the incidence of mutants across fractions did not correlate with the level of contaminating progenitor DNA (Fig. 4Biii). Third, each mutant was of a size appropriate for the size range of the fraction in which it was detected. Thus, the four Lepore-sized deletions from man 3 were all derived from the only three fractions that could contain Lepore-sized DNA molecules. If the ten mutants were in fact artefacts that could have arisen with equal likelihood in any size fraction, then the chance that all would be derived from correctly size-matched fractions is very low ($P = 0.00005$).

Characterisation of sperm deletion mutants

All mutants, as well as progenitor molecules of each haplotype from each man, were sequenced to define deletion breakpoints and to identify SNP heterozygosities flanking the deletion that could be used to test for exchange of markers accompanying deletion (Fig. 5A). The four Lepore-sized deletions m1-m4 were simple unequal exchanges between homologous sequences. All mapped to homology block 1, within one of the longer (54 bp) regions of sequence identity in an interval yet to be identified as an unequal exchange point in any Hb Lepore (Fig. 5C). One mutant was derived from one haplotype and the other three from the other haplotype, and none showed exchange of flanking DNA markers.

The non-Lepore sized mutants m5-m10 were all simple deletions within single-copy DNA, commencing in δ-globin homology block 1 or downstream of the δ-globin gene and terminating within or downstream of the β-globin gene (Fig. 5A, B). These mutants showed extremely limited homology of just 1-3 bp between 5′ and 3′ breakpoints (Fig. 5D), providing further evidence that they were not jumping PCR artefacts arising by annealing of incompletely extended DNA strands during PCR. Nor were there any significant stem-loop secondary structures between 5′ and 3′ breakpoint sequences that could promote deletion artefacts by Taq polymerase traversing the neck of a stem during PCR (not shown). None of the deletions in man 4 showed exchange of flanking markers; deletions in man 3 removed 3′ SNP markers and thus exchange could not be tested. None of these six mutants, nor any of the Lepore mutants, showed any other DNA sequence change over 2.4-4.0 kb of DNA sequenced around the site of deletion.

DISCUSSION

This work provides the first high-resolution definition of the β-globin crossover hotspot and shows that, contrary to previous claims (50), it can be identified and localised with some precision from genotype data by using simple pairwise LD analysis to characterise haplotype block structure (2), as well as by more sophisticated coalescent analyses (27,42). Direct analysis of sperm crossovers has shown that the hotspot is typical of other human recombination hotspots characterised by sperm typing (Table 1), with a width of 1.2 kb and flanked by recombinationally inert DNA. It is however more active than any other autosomal hotspot yet characterised. Its current activity in male meiosis (mean recombination frequency $1.5\text{x}10^{-3}$ in two men analysed) is similar to that previously estimated by single sperm typing (RF = $8.8\text{x}10^{-3}$, 95% CI $1.0\text{-}18\text{x}10^{-3}$ [19]) and is comparable to the historical recombination rate of $5.7\text{x}10^{-3}$ estimated from coalescent analysis (42). This hotspot is also likely to be active in female meiosis, given the identification of a maternal crossover that maps within the hotspot (8), and any difference between historical rates and current rates in sperm could be readily explained by greater activity in females. The hotspot shows similar activity in both men tested and no evidence for reciprocal crossover asymmetry (31) in either man, suggesting equal rates of recombination initiation on both haplotypes in each man. The four different haplotypes analysed therefore show no evidence of polymorphism in recombination rate as seen at some other crossover hotspots (24,30,31,51).

This work confirms that the recombination hotspot is centred in single-copy DNA 1 kb upstream of the β-globin gene, in a region containing the promoter and a replication origin (52). The association with a promoter is unusual for human hotspots (Table 1) (2) and might suggest that the β-globin hotspot is an example of an α-hotspot as described in yeast (53). As noted previously (8), this region contains several motifs including a Chi sequence, a *Pur* binding element, and $(TG)_n$ and $(ATTTT)_n$ repeats, all of which have been implicated in promoting recombination in viruses, *S. cerevisiae* and humans (54-58). The hotspot also contains two copies of the sequence CCTCCCT strongly over-represented within human crossover hotspots (15). However, these are both located 800-900 bp away from the centre of the hotspot, at the beginning of the β-globin gene. It remains unclear whether any of these sequence motifs are directly involved in hotspot activity.

While the β-globin crossover hotspot is typical in morphology compared with other human hotspots, it does show two unusual features. First is an unusually low level of gene conversion activity without exchange compared with the frequency of crossover. This is unlikely to be due to conversion tracts being missed because of lack of markers. Most conversion activity occurs near the centre of a hotspot (29), and marker F0 lies as close if not closer to the centre than the central markers tested at other hotspots assayed for conversion activity (Table 1). There is therefore real and considerable variation between hotspots in the choice between resolving recombination initiation events as non-exchanges or exchanges, with the observed ratios (uncorrected for missed marker-less conversion tracts) varying from 2.7:1 at hotspot *DNA*3 to <1:12 at the β-globin hotspot, a range of >30-fold. What controls this variation remains wholly unclear, but it does provide further evidence that human crossovers and conversions are generated by separate pathways, as seen in yeast (59).

The second unusual feature of the β-globin hotspot is the different spread of crossover breakpoints in reciprocal crossovers seen in both men tested (Fig. 2). This phenomenon has not been seen before in a human hotspot and its cause is unclear. One possibility is provided by the double-strand break (DSB) repair model for recombination in which recombination is initiated by a DSB which is then resected to generate single-strand ends that invade the homologous chromosome (60). Mismatch recognition could remove mismatches on the invading strands, leading to replacement of central markers on the initiating chromosome by alleles from the non-initiating chromosome (61). If both chromosomes initiate at the same rate, then the transmission of these central markers to crossover progeny will be restored to 50:50. However, incomplete removal of mismatches on the invading strand and/or subsequent branch migration at ensuing Holliday junctions could create mismatches in heteroduplex DNA nearer the edges of the hotspot. Biased mismatch repair operating independently on these 5′ and 3′ mismatches could lead to the observed non-Mendelian transmission of markers near the edge of the hotspot.

The β-globin hotspot only just extends into the beginning of the β-globin gene and into homology block 1 (HB1) shared by the δ- and β-globin genes in which Hb Lepores are generated by unequal exchange. One allelic crossover, in man 1, maps to a region terminating in the 5′ region of HB1 (Fig. 2A) within the Hb Senegalese exchange interval (Fig. 5C). This suggests a crossover rate within HB1 of very roughly 5×10^{-6} per sperm. Likewise, the best-fit normal distributions for crossovers (Fig. 2A) suggest a crossover rate in HB1 of very approximately 1.5×10^{-6}. In contrast, the frequency of Hb Lepore deletions in sperm (4 mutants seen in man 3, none in man 4) is extremely low, at 2.5×10^{-8} per sperm (95% CI $0.5\text{-}6.9\times10^{-8}$), roughly 100-fold lower than the estimated allelic crossover rate. Furthermore, all Lepore exchanges map to the 3′ side of HB1, away from the hotspot, with none in the second longest region of sequence identity between δ and β HB1 within which the Hb Senegalese exchange maps and which should be the most likely target for any unequal exchanges driven by the crossover hotspot (62). There is therefore no evidence that the β-globin crossover hotspot drives ectopic recombinational exchanges. This contrasts sharply with ectopic exchanges in yeast, in which ectopic recombination frequencies between repeats on the same chromosome can approach allelic exchange frequencies (63). This difference may however be due to the substantial (8.6%) sequence divergence between δ and β HB1s resulting in mismatches arising during ectopic exchange that could lead to these events being aborted (64).

The few Lepore exchanges seen in sperm are clearly the products of recombination between homologous sequences within HB1, but it is unclear whether they arise by aberrant meiotic recombination. None of the mutants recovered showed exchange of flanking DNA markers (Fig. 5A) and they could instead have arisen premeiotically by unequal mitotic recombination between sister chromatids or by intramolecular recombination. Given the extremely low frequency of these deletions in sperm, analysis of somatic instability in blood DNA will be extremely difficult, and the recovery of reciprocal anti-Lepore products of unequal exchange will be impossible using current approaches.

Interestingly, this mutation survey, while initially designed for Lepore-type deletions, also yielded non-Lepore deletions from both men. Again, their frequency in sperm was extremely low, at $3.8x10^{-8}$ (95% CI $1.1\text{-}8.8x10^{-8}$). All six deletions removed part or all of the β-globin gene and spanned the crossover hotspot. None showed a deletion breakpoint terminating within the hotspot, as might be anticipated for a deletion triggered by a DSB arising within the hotspot. There is therefore no evidence that the β-globin hotspot promotes this class of deletion. The mechanism of deletion is unclear but appears not to involve homologous recombination given the lack of exchange of flanking markers and the absence of significant DNA sequence homology shared by 5′ and 3′ deletion breakpoints. Non-homologous end joining (NHEJ) is a plausible mechanism, though none of the mutants showed abnormal "orphan" sequences at the deletion junction as can sometime arise during NHEJ and which have been seen in some β-globin gene deletions ascertained in patients (65).

These non-Lepore sperm deletions all involve loss of 4.0-8.0 kb of DNA, similar in length to the 7.4 kb of DNA lost in Hb Lepore deletions. However, computer simulations based on the possible range of deletion sizes detectable in the enriched DNA (3.5-9.4 kb) showed that this apparent clustering of deletion sizes is not significant ($P = 0.11$). 5′ breakpoints also appear to be randomly distributed (Fig. 5B) (given the observed deletion sizes, the chance that six randomly located breakpoints would map within an 3.7 kb interval is 0.26). In contrast, 3′ breakpoints are strongly and significantly ($P = 0.0023$) clustered into a 1.3 kb interval across the β-globin gene, with three mapping within just 120 bp of each other at the end of the gene. This suggests the possible existence of a controlling element in or very near this gene that promotes such deletions, though current data cannot further localise this putative regulator. The existence of this regulator might also predict clustering of 5′ breakpoints for deletions extending 3′ of the β-globin gene. Such deletions would have been excluded from our survey and merit further investigation.

This work describes, to our knowledge, the first direct detection and quantification of spontaneous deletions in human germline DNA. The deletions recovered are all pathogenic and would cause β-thalassaemia. Similar deletions have been detected in thalassaemic and HPFH individuals (44), with deletion breakpoints showing extremely limited junctional homology and with some evidence for association between breakpoints and transcription units (65). Interestingly, the frequency of these sperm deletions is comparable to the rate of base mutation in the human germline. Thus, the frequency with which a given base is lost in a deletion is about $2.0\text{x}10^{-8}$ per sperm (deletion rate of $3.8\text{x}10^{-8}$, with on average 53% of the 12 kb target lost per deletion). This frequency is similar to the mean rate of $3.9\text{x}10^{-8}$ for *de novo* base substitution estimated for the human male germline from human/chimpanzee divergence (66). This suggests that, if the dynamics of deletion in the β-globin gene region are more generally applicable in the human genome, then deletion might be of comparable significance to base substitution in driving mutations into the human genome. The challenge now is to investigate more generally the dynamics of deletion and to test directly whether deletion processes can also generate reciprocal duplications that could be a major source of segmental duplications and duplicate genes within the genome.

MATERIALS AND METHODS

DNA samples

We collected, with approval from the Leicestershire Health Authority Research Ethics Committee, semen and blood samples with informed consent from 200 men of north European descent, including volunteers and men attending fertility clinics, and selected 98 men showing good sperm DNA yields for further analysis. Sperm and blood DNAs were extracted as described previously (67). Sperm DNAs were whole genome amplified by multiple displacement amplification (MDA) (68) prior to routine genotyping.

PCR amplification, SNP discovery and genotyping

DNA was amplified using the PCR buffer described previously (69) supplemented with 12 mM Tris base, 0.2 μM of each primer, 0.03 U/μl *Taq* polymerase and 0.003 U/μl *Pfu* polymerase. PCR reactions were carried out in 0.2 ml PCR tubes or 96 well plates in an MJ Research PTC-225 Tetrad DNA engine or an Applied Biosystems GeneAmp PCR System 9700 thermal cycler, using primers designed from the consensus β-globin sequence (GenBank accession number NG00007). The NCBI SNP database (47) and previous literature (8,45,46) were scanned for SNPs over the β-globin gene region. Short (1-6 kb) targets from this region were PCR amplified from MDA-amplified genomic DNA and genotyped by allele-specific oligonucleotide (ASO) hybridisation to dotblots of PCR products using the tetramethylammonium chloride (TMAC) method as described previously (48). SNPs were also discovered through resequencing as described previously (2) and genotyped as above. The linkage phase of alleles was established by allele-specific PCR directed to a heterozygous SNP inside the target region in conjunction with a universal primer outside the recombination assay interval, followed by typing PCR products by ASO hybridisation.

Linkage disequilibrium analysis

Pairwise LD analysis was carried out on unphased diploid genotype data and plotted as described previously (2). Briefly, maximum-likelihood haplotype frequencies estimated from pairwise comparison of diploid genotypes were used to determine the |D′| measure of complete association. Observed allele frequencies at each SNP were then used to predict the haplotype frequencies expected if pairs of SNPs were in linkage equilibrium and these were used in turn to estimate the likelihood ratio (LR) in favour of significant association.

Sperm crossover detection and mapping

Blood and sperm DNAs were prepared and subsequently manipulated under conditions designed to minimise the risk of contamination (67). Allele-specific primers (ASPs) 15-18 nt long were designed for appropriate heterozygous SNP sites in the 5′ and 3′ LD blocks. These ASPs were optimised by PCR on genomic DNA from individuals homozygous for the correct or incorrect allele, to identify primers that showed good efficiency and excellent allele specificity and to determine optimal annealing temperatures. The final ASPs used in crossover assays were: M5/TF, 5′ CTC CCA AGT AGC TGG CAT 3′; M5/CF, 5′ CTC CCA AGT AGC TGG CAC 3′; M7/AF, 5′ CCT CGG CCT CTG AAT GTA 3′; M7/GF, 5′ CCT CGG CCT CTG AAA GTG 3′ ; F19/GR, 5′ CAG GAC AGT CAA ACC 3′ ; F19/TR, 5′ CAG GAC AGT CAA ACA 3′ ; M11/AR, 5′ GGG TGG GCC TAT GAT 3′ and M11/GR, 5′ GGG TGG GCC TAT GAC 3′. Crossover molecules were selectively amplified from multiple batches of sperm DNA, each containing on average one amplifiable crossover molecule (total of 50–2000 amplifiable progenitor molecules of each haplotype per reaction), by long PCR using ASPs in repulsion phase directed to selector SNP sites 13.6 kb apart. ASPs used were M5/TF or M5/CF in combination with M11/AR or M11/GR, and DNA was amplified for 23-26 cycles at 96^{o} for 20 sec, 62-65^{o} for 30 sec, 66^{o} for 15 min. Primary PCR products were immediately digested with S1 nuclease to remove any single-stranded DNA and PCR artefacts (51), then re-amplified as above with nested secondary ASPs M7/AF or M7/GF plus F19/GR or F19/TR. Secondary products were barely detectable by ethidium bromide staining after agarose gel electrophoresis, so all secondary PCRs were re-amplified with nested non-allele-specific tertiary primers 60.4F (5′ CAT GTA ACC AGA TCT CCC AAT GTG 3′) and 72.1R (5′ CCT CAG AAA AGG ATT CAA GTA GAG GC 3′) as above to identify positive PCR reactions. Positive tertiary PCR products were dotblotted and internal crossover points mapped by ASO hybridisation. All crossover analyses included blood DNA and negative controls containing no DNA. Poisson analysis of limiting dilutions of sperm DNA was used to estimate the number of amplifiable input molecules, and established that one amplifiable molecule of each haplotype was present per 12 pg DNA. Crossover data were Poisson-corrected for PCR reactions containing more than one crossover molecule as described previously (51).

Analysis of gene conversions

Single haplotypes were amplified from genomic DNA with the nested reverse ASPs as above in conjunction with nested universal forward PCR primers 60.4F and 60.7F (5′ GTG GTA GTG ATT CAC ACA GC 3′), using pools of sperm DNA each containing 20-30 amplifiable DNA molecules of each haplotype. PCR products were typed using ASOs directed to alleles on the non-amplified haplotype to identify pools containing crossovers or conversions. These ASO hybridisations included a control series of PCR products from mixtures of the non-selected and selected haplotypes, in ratios of 1:10, 1:50, 1:100 and 1:500, to provide controls for hybridisation signal intensity of pools containing recombinant PCR products (29).

Size fractionation of sperm DNA to enrich for deletion mutants

A total of 800 μg sperm DNA from each of two individuals, purified from 1-2 ejaculates per man, was digested to completion with *Eco*RV (New England BioLabs), ethanol precipitated and dissolved in 5 mM Tris-HCl, pH 7.5. A 200 μg aliquot of DNA was loaded in 1 ml 0.5xTBE (44mM Tris-borate pH 8.3, 1mM EDTA), 5% v/v glycerol, 400 μg/ml ethidium bromide plus bromophenol blue into a 10 cm x 0.3 cm slot in a 40 cm long, 1.5 cm deep 0.8% SeaKem HGT agarose gel in 0.5xTBE, 0.5 μg/ml ethidium bromide and electrophoresed at 60 V for 20 min. The current was then reversed for 5 min to allow any DNA overloaded at the gel interface to return into free solution, then electrophoresis continued for an additional 20 min to allow all DNA to enter the gel. This procedure was repeated three times until all 800 μg DNA had been loaded without overloading the gel. The gel was then electrophoresed in the dark at 60 V for 4 days until a 6.6 kb λ DNA x *Hin*dIII marker had migrated 30 cm. DNA markers were visualised using a Dark Reader transilluminator (Clare Chemical Research) and gel slices containing genomic DNA collected over the size range 6-14 kb to include any mutant DNA molecules similar in size to Hb Lepore deletions (8.1 kb) and to exclude progenitor β-globin DNA molecules (15.5 kb). The top and bottom 2 mm of each gel slice was excised to remove any aberrantly migrating DNA molecules, and remaining genomic DNA was recovered by electroelution onto dialysis membrane. Each DNA fraction was ethanol precipitated and dissolved in 100 μl 5 mM Tris-HCl, pH 7.5.

Characterisation of size fractions of sperm DNA

Aliquots of each size fraction of *Eco*RV-digested sperm DNA were analysed by agarose gel electrophoresis, both individually and pooled, against a dilution series of the initial *Eco*RV digest to estimate overall yield (50-70%) and the size distribution of each fraction. The distribution across the fractions of *Eco*RV DNA fragments corresponding in size to Hb Lepore mutants (8.1 kb) was determined using a control 8.035 kb *Eco*RV DNA fragment from the class II region of the MHC. Aliquots of each fraction, together with a dilution series of the initial *Eco*RV digest, were PCR amplified using MHC primers R46.5F (5′ GGC AGG TAT CTG ATA CAG AGC 3′) and R51.9R (5′ GAC AAA GTT TCC CCT GTT GC 3′) and yields of the 5.4 kb PCR product derived from within this MHC *Eco*RV fragment compared to estimate recovery in each size fraction. The total number of amplifiable MHC molecules was estimated by Poisson analysis of multiple aliquots of DNA pooled from all fractions, diluted to the single molecule level and amplified by PCR. The total yield of amplifiable DNA molecules ($1.0\text{x}10^8$ for each semen donor) corresponded to a 54% yield from 800 μg DNA. This yield was similar to that estimated from bulk DNA recovery and established that little DNA damage had occurred during DNA fractionation. The residual level of contaminating progenitor β-globin DNA molecules in each fraction was similarly estimated by amplifying a 12.2 kb δ-, β-globin DNA interval located within the 15.5 kb *Eco*RV globin DNA fragment using primers B33.3F (5′ TAA ACA TGT AAC CAG ATC TCC C 3′) and B45.5R (5′ TGC AGA GCC AGA AGC ACC 3′).

Screening fractionated sperm DNA for mutant molecules

Multiple aliquots of each DNA fraction, each containing at most 0.25 μg DNA and, depending on the fraction, 2-50 molecules of progenitor globin DNA, were PCR amplified in 20 μl reactions using primers B33.2F (5′ ACA AAT CCT CTC AAT GCA ATC C 3′) and B45.6R (5′ CAG AAT CTA GCA TCT ACC TAC C 3′), using one cycle of 96° for 1.5 min followed by 26 cycles of 96° for 20 sec, 61° for 30 sec, 64° for 15 min. These primary PCRs were diluted with 180 μl H_2O and 0.7 μl used to seed a 10 μl secondary PCR containing nested primers B33.3F and B45.5R (details above) followed by amplification under the same conditions but for 17 cycles. Secondary PCR products were analysed by agarose gel electrophoresis to identify reactions containing a deletion mutant as well as progenitor. 80% of each fraction was tested for mutants. The efficiency of these PCR primers at amplifying single DNA molecules was established by Poisson analysis of multiple aliquots of diluted sperm DNA (4.2-13.9 pg DNA per PCR) amplified as above and tested for the 12.2 kb progenitor PCR product. The number of amplifiable DNA molecules per haploid genome (3 pg DNA) was 0.72 (95% CI 0.50-1.03), indicating 72% efficiency of amplification of progenitor DNA molecules.

Analysis of mutants

Mutants were re-amplified using nested PCR primers B33.4F (5′ TAA CCA GAT CTC CCA ATG TG 3′) and B45.4R (5′ CCA TAA GGG ACA TGA TAA GGG 3′) and deletion breakpoints roughly located by restriction mapping. All mutants were sequenced over a 2.4-4.0 kb interval spanning the breakpoint and including all known SNP heterozygosities. To control for possible sequencing errors resulting from gel fractionation and single molecule PCR, single progenitor molecules 12.2 kb long were amplified as above from extreme dilutions of the pooled fractions and two molecules of each haplotype were sequenced. No errors were seen over 20.3 kb of DNA sequenced.

ACKNOWLEDGEMENTS

We thank J. Blower and volunteers for providing semen and blood samples, and M.T. Slingsby, C.A. May and others for helpful discussions. This work was supported by grants to AJJ from the Medical Research Council, the Biotechnology and Biological Sciences Research Council, the Royal Society and the Louis-Jeantet Foundation.

Conflict of Interest statement. None declared.

REFERENCES

1. Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J. and Lander, E.S. (2001) High-resolution haplotype structure in the human genome. *Nat. Genet.,* **29,** 229-232.
2. 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.
3. Goldstein, D.B. (2001) Islands of linkage disequilibrium. *Nat. Genet.,* **29,** 109-111.
4. Ardlie, K.G., Kruglyak, L. and Seielstad, M. (2002) Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.,* **3,** 299-309.
5. Dawson, E., Abecasis, G.R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D.M., Pabial, J., Dibling, T., Tinsley, E., Kirby, S. *et al.* (2002) A first-generation linkage disequilibrium map of human chromosome 22. *Nature,* **418,** 544-548.
6. Kauppi, L., Jeffreys, A.J. and Keeney, S. (2004) Where the crossovers are: recombination distributions in mammals. *Nat. Rev. Genet.,* **5,** 413-424.
7. Altshuler, D., Brooks, L.D., Chakravarti, A., Collins, F.S., Daly, M.J. and Donnelly, P.; International HapMap Consortium (2005) A haplotype map of the human genome. *Nature,* **437,** 1299-1320.
8. Smith, R.A., Ho, P.J., Clegg, J.B., Kidd, J.R. and Thein, S.L. (1998) Recombination breakpoints in the human β-globin gene cluster. *Blood,* **92,** 4415-4421.
9. Oudet, C., Hanauer, A., Clemens, P., Caskey, T. and Mandel, J.L. (1992) Two hot spots of recombination in the DMD gene correlate with the deletion prone regions. *Hum. Mol. Genet.,* **1,** 599-603.

10. Cullen, M., Erlich, H., Klitz, W. and Carrington, M. (1995) Molecular mapping of a recombination hotspot located in the second intron of the human *TAP2* locus. *Am. J. Hum. Genet.,* **56,** 1350-1358.

11. Cullen, M., Noble, J., Erlich, H., Thorpe, K., Beck, S., Klitz, W., Trowsdale, J. and Carrington, M. (1997) Characterization of recombination in the HLA class II region. *Am. J. Hum. Genet.,* **60,** 397-407.

12. Rana, N.A., Ebenezer, N.D., Webster, A.R., Linares, A.R., Whitehouse, D.B., Povey, S. and Hardcastle, A.J. (2004) Recombination hotspots and block structure of linkage disequilibrium in the human genome exemplified by detailed analysis of *PGM1* on 1p31. *Hum. Mol. Genet*., **13**, 3089-3102.

13. Tapper. W., Collins. A., Gibson. J., Maniatis. N., Ennis. S. and Morton N.E. (2005) A map of the human genome in linkage disequilibrium units. *Proc. Natl. Acad. Sci. USA*, **102**, 11835-11839.

14. 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.

15. Myers, S., Bottolo, L., Freeman, C., McVean, G. and Donnelly, P. (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science*, **310**, 321-324.

16. Arnheim, N., Calabrese, P. and Nordborg, M. (2003) Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved. *Am. J. Hum. Genet.*, **73**, 5-16.

17. Wang, N., Akey, J.M., Zhang, K., Chakraborty, R. and Jin, L. (2002) Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am. J. Hum. Genet,* **71,** 1227-1234.
18. Zhang, K., Akey, J.M., Wang, N., Xiong, M., Chakraborty, R. and Jin, L. (2003) Randomly distributed crossovers may generate block-like patterns of linkage disequilibrium: an act of genetic drift. *Hum. Genet.,* **113,** 51-59.
19. Schneider, J.A., Peto, T.E., Boone, R.A., Boyce, A.J. and Clegg, J.B. (2002) Direct measurement of the male recombination fraction in the human β-globin hot spot. *Hum. Mol. Genet.,* **11,** 207-215.
20. Hubert, R., Stanton, V.P., Aburatani, H., Warren, J., Li, H., Housman, D.E. and Arnheim, N. (1992) Sperm typing allows accurate measurement of the recombination fraction between D3S2 and D3S3 on the short arm of human chromosome 3. *Genomics,* **12,** 683-687.
21. Hubert, R., MacDonald, M., Gusella, J. and Arnheim, N. (1994) High resolution localization of recombination hot spots using sperm typing. *Nat. Genet.,* **7,** 420-424.
22. Lien, S., Szyda, J., Schechinger, B., Rappold, G. and Arnheim, N. (2000) Evidence for heterogeneity in recombination in the human pseudoautosomal region: high resolution analysis by sperm typing and radiation-hybrid mapping. *Am. J. Hum. Genet.*, **66**, 557-566.
23. Cullen, M., Perfetto, S.P., Klitz, W., Nelson, G. and Carrington, M. (2002) High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am. J. Hum. Genet.*, **71**, 759-776.

24. Jeffreys, A.J., Neumann, R., Panayi, M., Myers, S. and Donnelly, P. (2005) Human recombination hot spots hidden in regions of strong marker association. *Nat. Genet.*, **37**, 601-606.

25. May, C.A., Shone, A.C., Kalaydjieva, L., Sajantila, A. and Jeffreys, A.J. (2002) Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene *SHOX*. *Nat. Genet.,* **31,** 272-275.

26. Kauppi, L., Stumpf, M.P. and Jeffreys, A.J. (2005) Localized breakdown in linkage disequilibrium does not always predict sperm crossover hot spots in the human MHC class II region. *Genomics,* **86,** 13-24.

27. Winckler, W., Myers, S.R., Richter, D.J., Onofrio, R.C., McDonald, G.J., Bontrop, R.E., McVean, G.A., Gabriel, S.B., Reich, D., Donnelly, P. *et al*. (2005) Comparison of fine-scale recombination rates in humans and chimpanzees. *Science*, **308**, 107-111.

28. Ptak, S.E., Hinds, D.A., Koehler, K., Nickel, B., Patil, N., Ballinger, D.G., Przeworski, M., Frazer, K.A. and Paabo, S. (2005) Fine-scale recombination patterns differ between chimpanzees and humans. *Nat. Genet.*, **37**, 429-434.

29. Jeffreys, A.J. and May, C.A. (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nat. Genet.,* **36,** 151-156.

30. 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.

31. Jeffreys, A.J. and Neumann, R. (2002) Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat. Genet.,* **31,** 267-271.

32. Boulton, A., Myers, R.S. and Redfield, R.J. (1997) The hotspot conversion paradox and the evolution of meiotic recombination. *Proc. Natl. Acad. Sci. USA*, **94**, 8058-8063.
33. Samonte, R.V. and Eichler, E.E. (2002) Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.*, **3**, 65-72.
34. Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. and Eichler, E.E. (2002) Recent segmental duplications in the human genome. *Science*, **297**, 1003-1007.
35. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M. *et al*. (2004) Large-scale copy number polymorphism in the human genome. *Science*, **305**, 525-528.
36. Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Segraves, R. *et al*. (2005) Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78-88.
37. Shaw, C.J. and Lupski, J.R. (2004) Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum. Mol. Genet.*, **13**, R57-64.
38. Higgs, D.R. (1993) Alpha-thalassaemia. *Baillieres Clin. Haematol.*, **6**, 117-150.
39. Pentao, L., Wise, C.A., Chinault, A.C., Patel, P.I. and Lupski, J.R. (1992) Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nat. Genet.*, **2**, 292-300.

40. Reiter, L.T., Murakami, T., Koeuth, T., Pentao, L., Muzny, D.M., Gibbs, R.A. and Lupski, J.R. (1996) A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element. *Nat. Genet.*, **12**, 288-297.

41. Chakravarti, A., Buetow, K.H., Antonarakis, S.E., Waber, P.G., Boehm, C.D. and Kazazian, H.H. (1984) Nonuniform recombination within the human β-globin gene cluster. *Am. J. Hum. Genet.,* **36,** 1239-1258.

42. Wall, J.D., Frisse, L.A., Hudson, R.R. and Di Rienzo, A. (2003) Comparative linkage-disequilibrium analysis of the β-globin hotspot in primates. *Am. J. Hum. Genet.,* **73,** 1330-1340.

43. Efremov, G.D. (1978) Hemoglobins Lepore and anti-Lepore. *Hemoglobin,* **2,** 197-233.

44. Hardison, R.C., Chui, D.H., Giardine, B., Riemer, C., Patrinos, G.P., Anagnou, N., Miller, W. and Wajcman, H. (2002) HbVar: A relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. *Hum. Mutat.*, **19**, 225-233.

45. Fullerton, S.M., Bond, J., Schneider, J.A., Hamilton, B., Harding, R.M., Boyce, A.J. and Clegg, J.B. (2000) Polymorphism and divergence in the β-globin replication origin initiation region. *Mol. Biol. Evol.,* **17,** 179-188.

46. Fullerton, S.M., Harding, R.M., Boyce, A.J. and Clegg, J.B. (1994) Molecular and population genetic analysis of allelic sequence diversity at the human β-globin locus. *Proc. Natl. Acad. Sci. USA,* **91,** 1805-1809.

47. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29,** 308-311.
48. Jeffreys, A.J., Ritchie, A. and Neumann, R. (2000) High resolution analysis of haplotype diversity and meiotic crossover in the human *TAP2* recombination hotspot. *Hum. Mol. Genet.,* **9,** 725-733.
49. Jeffreys, A.J. and Neumann, R. (1997) Somatic mutation processes at a human minisatellite. *Hum. Mol. Genet.*, **6**, 129-136.
50. Fearnhead, P., Harding, R.M., Schneider, J.A., Myers, S. and Donnelly, P. (2004) Application of coalescent methods to reveal fine-scale rate variation and recombination hotspots. *Genetics,* **167,** 2067-2081.
51. Jeffreys, A.J., Murray, J. and Neumann, R. (1998) High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell*, **2**, 267-273.
52. Aladjem, M.I., Groudine, M., Brody, L.L., Dieken, E.S., Fournier, R.E., Wahl, G.M. and Epner, E.M. (1995) Participation of the human β-globin locus control region in initiation of DNA replication. *Science,* **270,** 815-819.
53. Petes, T.D. (2001) Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.,* **2,** 360-369.
54. Eggleston, A.K. and West, S.C. (1997) Recombination initiation: easy as A, B, C, D… chi? *Curr. Biol.,* **7,** R745-749.
55. Stringer, J.R. (1985) Recombination between poly[d(GT).d(CA)] sequences in simian virus 40-infected cultured cells. *Mol. Cell. Biol.,* **5,** 1247-1259.

56. Treco, D. and Arnheim, N. (1986) The evolutionarily conserved repetitive sequence d(TG.AC)n promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Mol. Cell. Biol.,* **6,** 3934-3947.

57. Majewski, J. and Ott, J. (2000) GT repeats are associated with recombination on human chromosome 22. *Genome Res.,* **10,** 1108-1114.

58. Wahls, W.P. and Moore, P.D. (1990) Homologous recombination enhancement conferred by the Z-DNA motif $d(TG)_{30}$ is abrogated by simian virus 40 T antigen binding to adjacent DNA sequences. *Mol. Cell. Biol.,* **10,** 794-800.

59. Paques, F. and Haber, J.E. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev*., **63**, 349-404.

60. Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. and Stahl, F.W. (1983) The double-strand-break repair model for recombination. *Cell*, **33**, 25-35.

61. Alani, E., Reenan, R.A. and Kolodner, R.D. (1994) Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae*. *Genetics***,** **137**, 19-39.

62. Zertal-Zidani, S., Ducrocq, R., Weil-Olivier, C., Elion, J. and Krishnamoorthy, R. (2001) A novel δβ fusion gene expresses hemoglobin A (HbA) not Hb Lepore: Senegalese $\delta^0\beta^+$ thalassemia. *Blood*, **98**, 1261-1263.

63. Goldman, A.S.H. and Lichten, M. (1996) The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. *Genetics*, **144**, 43-55.

64. Borts, R.H. and Haber, J.E. (1987) Meiotic recombination in yeast: alteration by multiple heterozygosities. *Science*, **237**, 1459-1465.

65. Henthorn, P.S., Smithies, O. and Mager, D.L. (1990) Molecular analysis of deletions in the human β-globin gene cluster: deletion junctions and locations of breakpoints. *Genomics*, **6**, 226-237.

66. Nachman, M.W. and Crowell, S.L. (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics*, **156**, 297-304.

67. Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L. and Armour, J.A. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.,* **6,** 136-145.

68. Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J. *et al.* (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA,* **99,** 5261-5266.

69. Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, J. (1988) Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res.,* **16,** 10953-10971.

70. Metzenberg, A.B., Wurzer, G., Huisman, T.H. and Smithies, O. (1991) Homology requirements for unequal crossing over in humans. *Genetics*, **128**, 143-161.

71. Fioretti, G., De Angioletti, M., Masciangelo, F., Lacerra, G., Scarallo, A., de Bonis, C., Pagano, L., Guarino, E., De Rosa, L., Salvati, F. *et al*. (1992) Origin heterogeneity of Hb Lepore-Boston gene in Italy. *Am. J. Hum. Genet.*, **50**, 781-786.

72. Chan, V., Au, P., Yip, B. and Chan, T.K. (2004) A new cross-over region for hemoglobin-Lepore-Hollandia. *Haematologica*, **89**, 610-611.

FIGURE LEGENDS

Figure 1. Detecting crossovers in the β-globin gene region. (**A)** Patterns of pairwise LD in a 20 kb region spanning the δ- and β-globin genes, determined from genotypes of 50 UK semen donors of north European descent and plotted as in (2) with colour-coded estimates of |D′| shown below the diagonal and the likelihood ratio (LR) in favour of significant LD above. The locations of SNPs are shown below and to the right of the plot. LD blocks indicated below the plot identify the target region for the crossover assay. The regions of putative hotspot activity previously identified by LD breakdown (41) and by single sperm typing (19) are shown below in green and blue respectively, and the interval chosen for sperm crossover analysis in grey. The kb scale relates to base position in Genbank accession NG000007. (**B)** Strategy for detecting sperm crossover molecules. All SNPs are shown, together with informative SNPs for each man analysed for crossovers and allele-specific primers (black, white triangles) used in nested allele-specific PCR to recover crossover molecules arising from exchanges (X) within the test interval. Reciprocal (A-, B-type) crossovers are defined arbitrarily by the 5′ selector sites used for crossover recovery. Note the different haplotypes present in the two men analysed. (**C**) Examples of crossover molecules detected in sperm DNA. PCRs seeded with the indicated numbers of amplifiable molecules of each haplotype were amplified by two rounds of nested allele-specific PCR, then reamplified using universal primers. These tertiary PCR products were analysed by agarose gel electrophoresis and staining with ethidium bromide. No crossover molecules were seen in blood DNA and in DNA-free control PCRs.

Figure 2. Sperm crossover distribution at the β-globin hotspot. (**A**) Distribution of combined orientation A+B crossovers in man 1 and man 2 recovered from 85,000 and 108,000 amplifiable molecules of each haplotype (1.02 and 1.30 μg sperm DNA) respectively. Informative SNPs are shown as ticks above each plot, and the numbers of crossovers mapping to each interval (shown above the histogram) were used to estimate local crossover activity in cM/Mb. Least-squares best-fit normal distributions (2) for each man are shown in red. (**B**) Cumulative frequency distributions for separate orientation A and B crossovers in each man, together with the best-fit cumulative normal distribution for combined A+B crossovers. Informative markers in each man are shown above the plot. (**C**) Transmission frequencies of markers from the white haplotype in each man (Fig. 1B) to crossover progeny, normalised to equal numbers of A and B crossovers and with 95% confidence intervals (31). Black, man 1; red, man 2.

Figure 3. Assaying crossovers and conversions without exchange at the β-globin hotspot. (**A**) Detection strategy. Multiple small pools of sperm DNA are amplified using nested ASPs 3′ to the hotspot (grey arrows) plus universal (not allele-specific) primers 5′ to the hotspot (black arrows). PCR products will be derived from non-recombinant molecules of haplotype 2 plus the occasional crossover or conversion molecule. ASO typing of alleles from the non-amplified grey haplotype can identify the presence of these recombinant DNA molecules. (**B**) Examples of recombinant detection in sperm DNA from man 2. Each PCR was seeded with 20 amplifiable molecules of each haplotype. The controls contain a 1:10 mixture of PCR products from haplotype 1 and 2. The recombinants

detected were all crossovers (breakpoints marked with crosses); no conversions without flanking marker exchange were seen in this assay.

Figure 4. Recovery of deletion mutants by size fractionation of sperm DNA. (**A**) The target region analysed, showing the location of δ- and β-globin genes and the recombination hotspot, plus *Eco*RV sites and nested universal PCR primers (arrows) used to recover mutant DNA molecules from fractionated DNA. Sperm DNA will include 15.5 kb *Eco*RV progenitor DNA molecules (12.2 kb after PCR amplification) plus in theory 8.1 kb *Eco*RV Hb Lepore deletions (4.8 kb after PCR) arising from unequal exchange between homology blocks shared by δ- and β-globin genes (**i**) and other deletions triggered for example by the hotspot (**ii**). Fractionated DNA will also contain broken DNA molecules (**iii**) which, if terminating within homology blocks, could yield artefact Lepore deletions by annealing during PCR. Such broken molecules will lie in the size range 1.8-6.3 kb. (**B**) Analysis of 800 μg sperm DNA digested with *Eco*RV and fractionated by agarose gel electrophoresis. (**i**) Size range of DNA fragments in each of the 10 fractions, with sizes of progenitor and Lepore deletion molecules indicated. (**ii**) Number of amplifiable 8.0 kb *Eco*RV MHC genomic DNA molecules in each fraction, matched in size to Hb Lepore deletion mutants. (**iii**) Number of intact progenitor δ-, β-globin DNA molecules. The fractions within which the seven mutants (m1-m7) were recovered are indicated. Fraction 10 was not screened due to significant contamination with progenitor DNA. (**C**) Examples of screening fractions 5 and 7 for deletion molecules using nested PCR. Each PCR contained DNA from 700,000 sperm with 5 or 20 remaining progenitor molecules respectively. Two Hb Lepore mutants were detected

(m1, m2) plus Lepore-sized artefacts generated by jumping PCR from the progenitor but at a low and uniform level. M, λ DNA x *Hin*dIII.

Figure 5. Characterisation of deletion mutants recovered from sperm DNA. (**A**) Location of deletion breakpoints in mutants m1-m10, together with flanking SNP information on progenitor haplotypes and each mutant. The region sequenced in each mutant is indicated in black. Homology blocks HB1 and HB2 shared by δ- and β-globin genes are shaded. Lepore-type deletions are marked with an asterisk. (**B**) Morphology of the β-globin crossover hotspot estimated from Fig. 2, together with the location of 5′ and 3′ deletion breakpoints (grey and black arrows respectively). (**C**) Location within HB1 of deletion breakpoints in the Lepore-type sperm mutants, plus the 5′ breakpoint in mutant m5 and 3′ breakpoint in mutant m10 . Sequence differences between the δ- and β-globin genes in HB1 are shown by vertical ticks and the lengths of the longest blocks of sequence identity indicated. The locations of characterised Lepore-type exchanges identified in human populations (62,70-72) are shown below. (**D**) Sequences around the 5′ breakpoint (top) and 3′ breakpoint (bottom) compared with the deletion mutant (middle) for each of the six non-Lepore deletions, with sequence matches indicated by lines and with sequences shared by 5′ and 3′ breakpoints marked in bold.

Table 1. Comparison of human meiotic recombination hotspots characterised by sperm typing.

	centre-point location	width (kb)[a]	peak activity (cM/Mb)	sperm crossover frequency (x 10^{-5})[b]	distance closest marker to centre (bp)	sperm conversion frequency (x 10^{-5})[c]	Ratio conversion: crossover[d]
β-globin	promoter/origin of replication	1.2	200	290	20	<25	<1:12
_DPA_1	HERV repeat	1.7	27	30	-	-	-
_DNA_1	promoter	1.9	0.5	0.5	-	-	-
_DNA_2	intergenic *Alu*	1.3	3.7	2.2	-	-	-
_DNA_3	intergenic *Alu*	1.2	130	110	80	300	2.7:1
_DMB_1	intragenic intron/exon?	1.8	3.1	27	-	-	-
_DMB_2	intergenic	1.2	28	5	120	4	1:1.3
_TAP_2	intron	1.0	5.8	7.3	-	-	-
_NID_1	intragenic – *Alu* *NID* intron 4	1.5	70	50	70	13	1:4
_NID_2a	intragenic – *NID* intron 12	1.4	10	8.5	-	-	-
_NID_2b	intragenic - *NID* intron 12	1.1	4	3.0	-	-	-
_NID_3	intergenic *Alu*	2.0	70	96	-	-	-
MS32	RTLV-LTR	1.5	43	39	-	-	-
MSTM1a	intergenic, single copy	1.6	9	8.8	-	-	-
MSTM1b	intergenic, single copy	2.1	16	15	-	-	-
MSTM2	intergenic, single copy	1.3	0.9	0.7	-	-	-
SHOX	intragenic exon?	2.2	370	370	200	90	1:3.3

Data are from this work and from refs. 2, 24-26, 29, 30, 48 and 51. Hotspots *DPA*1–*TAP2* are located within the MHC Class II region, *NID*1–MSTM2 in chromosome 1q42.3 and *SHOX* in the pseudoautosomal pairing region PAR1.

a, width of hotspot within which 95% of crossovers occur, estimated from best-fit normal distributions (2).

b, mean frequency over men tested. Crossover frequencies are taken from gene conversion assays where available.

c, frequency of non-exchange conversion events involving the marker closest to the centre of the hotspot.

d, ratio not corrected for undetectable marker-less conversion tracts.

Fig. 1

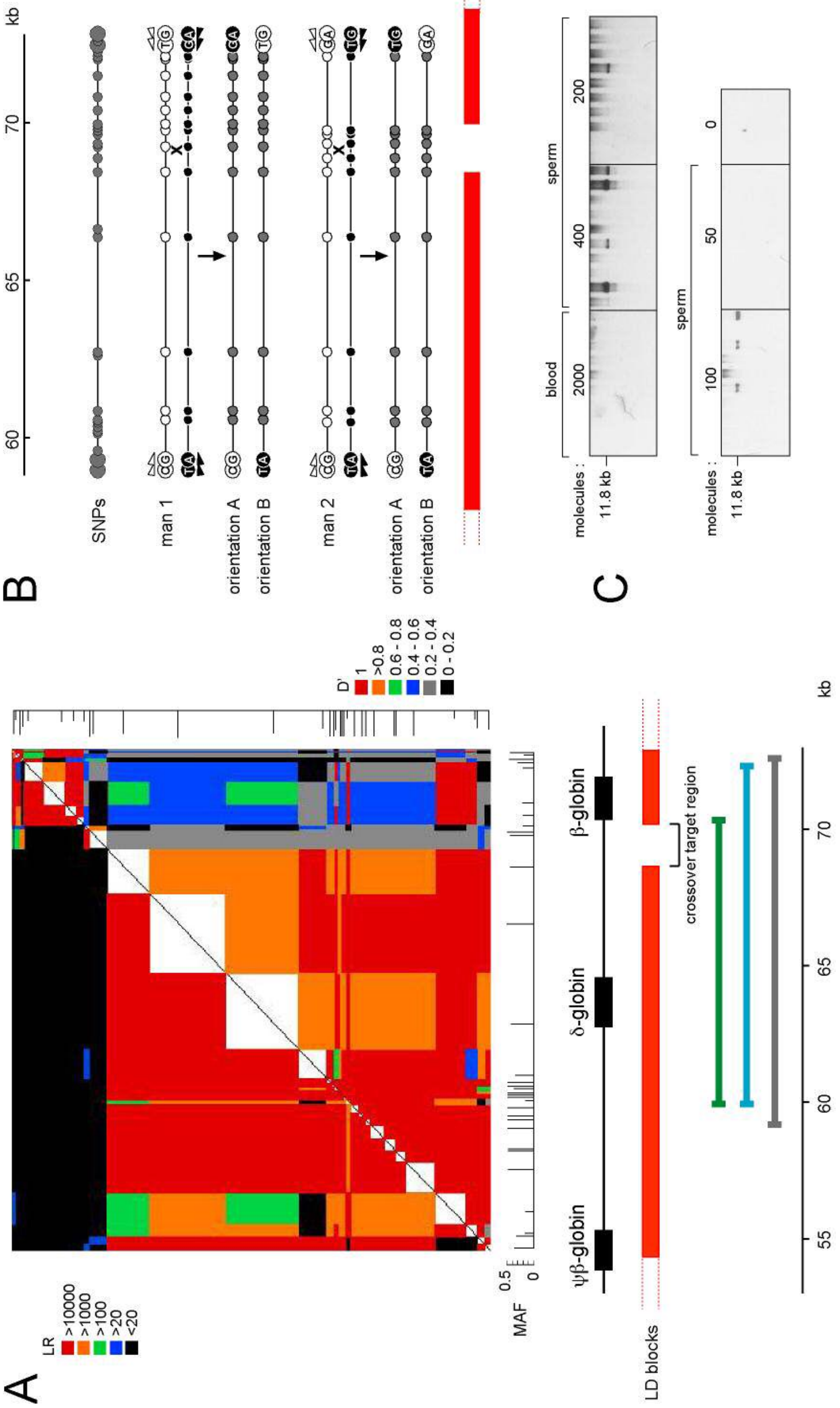

A
LR
>10000
>1000
>100
>20
<20
MAF
0.5
0
D'
1
>0.8
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0 - 0.2
ψβ-globin
δ-globin
β-globin
LD blocks
crossover target region
55
60
65
70
kb
B
kb
60
65
70
SNPs
man 1
orientation A
orientation B
man 2
orientation A
orientation B
CG
TA
TG
GA
X
C
molecules :
11.8 kb
blood
sperm
2000
400
200
100
50
0

Fig. 2

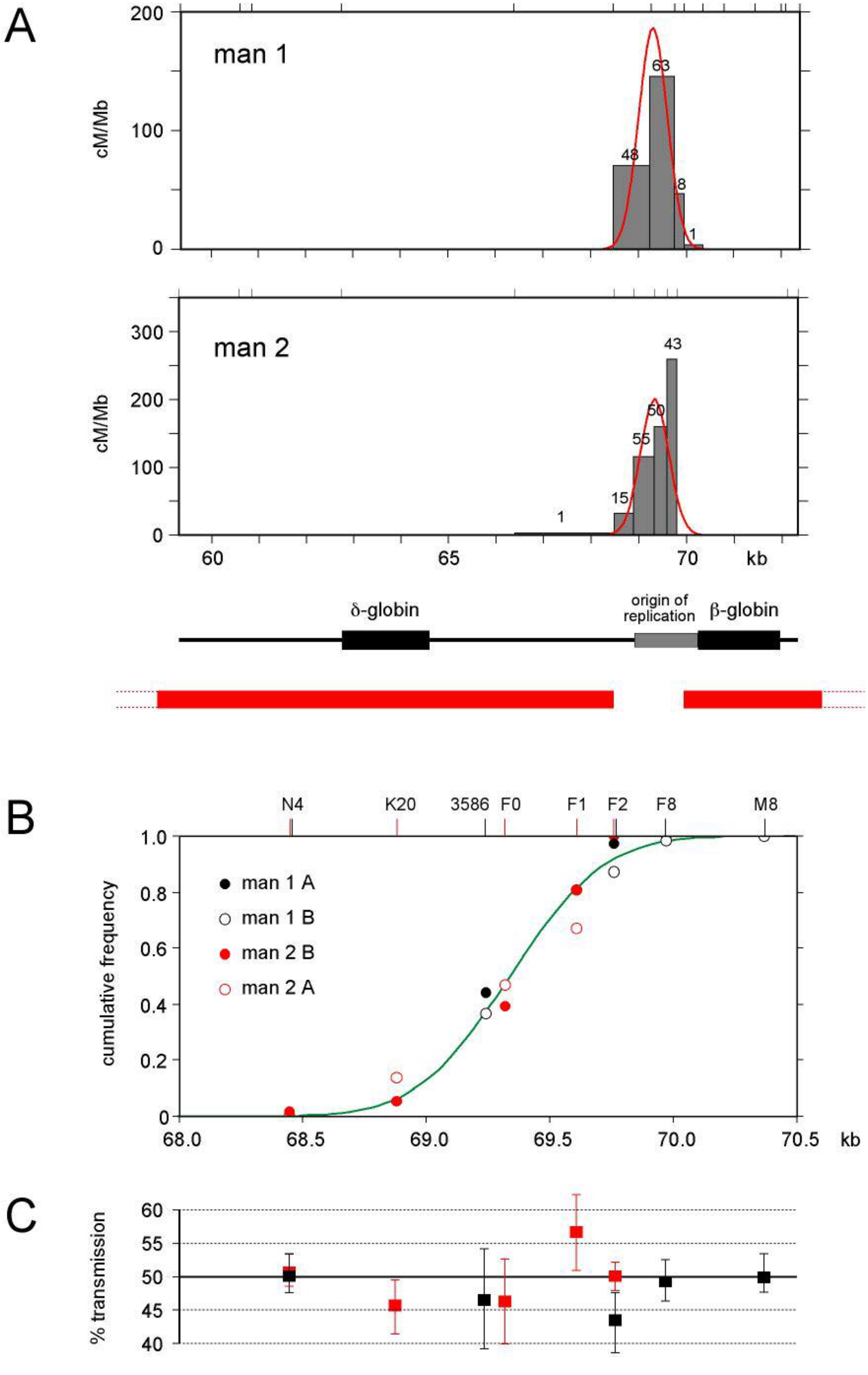

A
man 1
cM/Mb
200
100
0
63
48
8
1
man 2
cM/Mb
300
200
100
0
43
50
55
15
1
60
65
70
kb
δ-globin
origin of replication
β-globin
B
N4
K20
3586
F0
F1
F2
F8
M8
cumulative frequency
1.0
0.8
0.6
0.4
0.2
0
● man 1 A
○ man 1 B
● man 2 B
○ man 2 A
68.0
68.5
69.0
69.5
70.0
70.5
kb
C
% transmission
60
55
50
45
40

Fig. 3

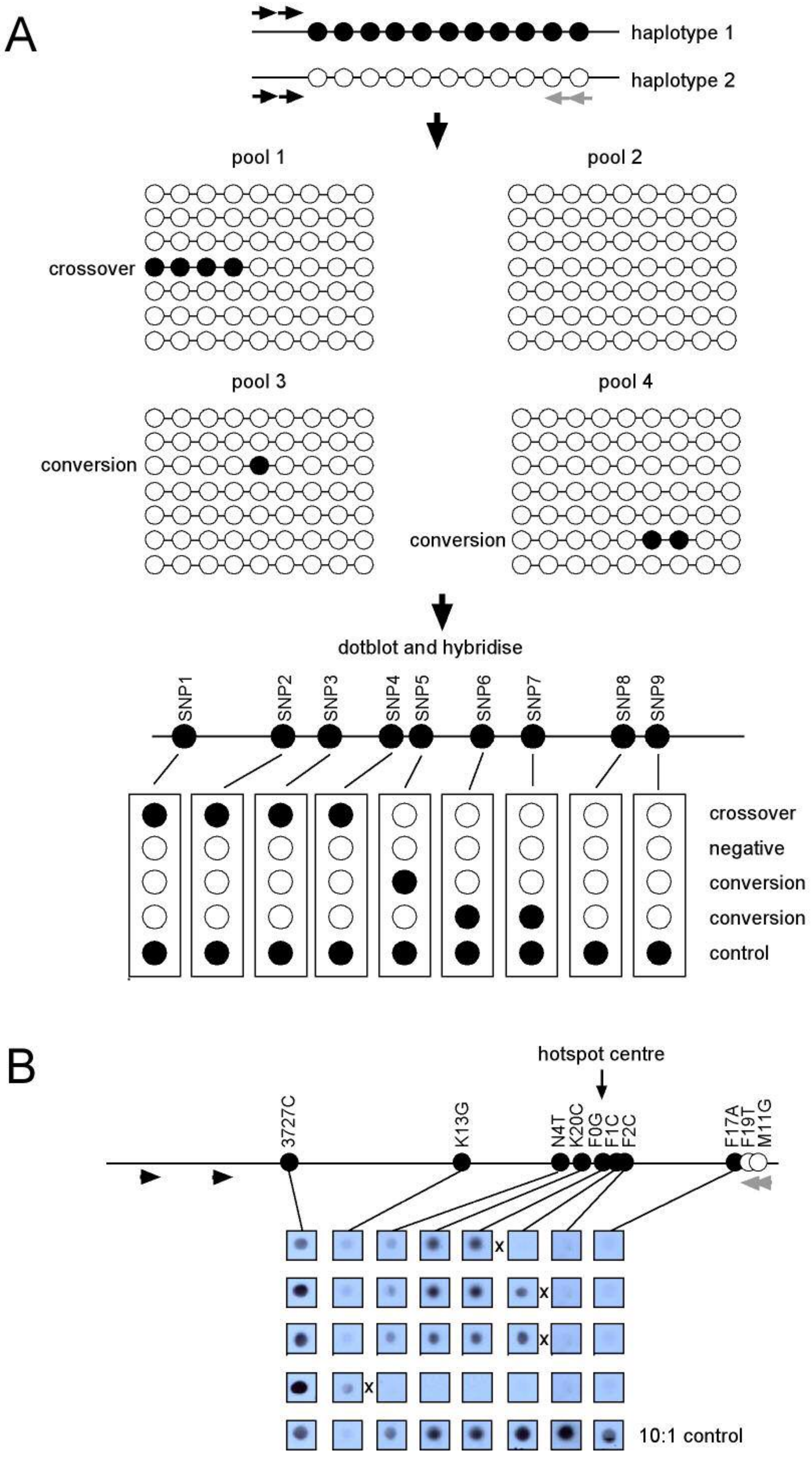

A
haplotype 1
haplotype 2
pool 1
pool 2
crossover
pool 3
pool 4
conversion
conversion
dotblot and hybridise
SNP1
SNP2
SNP3
SNP4
SNP5
SNP6
SNP7
SNP8
SNP9
crossover
negative
conversion
conversion
control
B
hotspot centre
3727C
K13G
N4T
K20C
F0G
F1C
F2C
F17A
F19T
M11G
x
x
x
x
10:1 control

Fig. 4

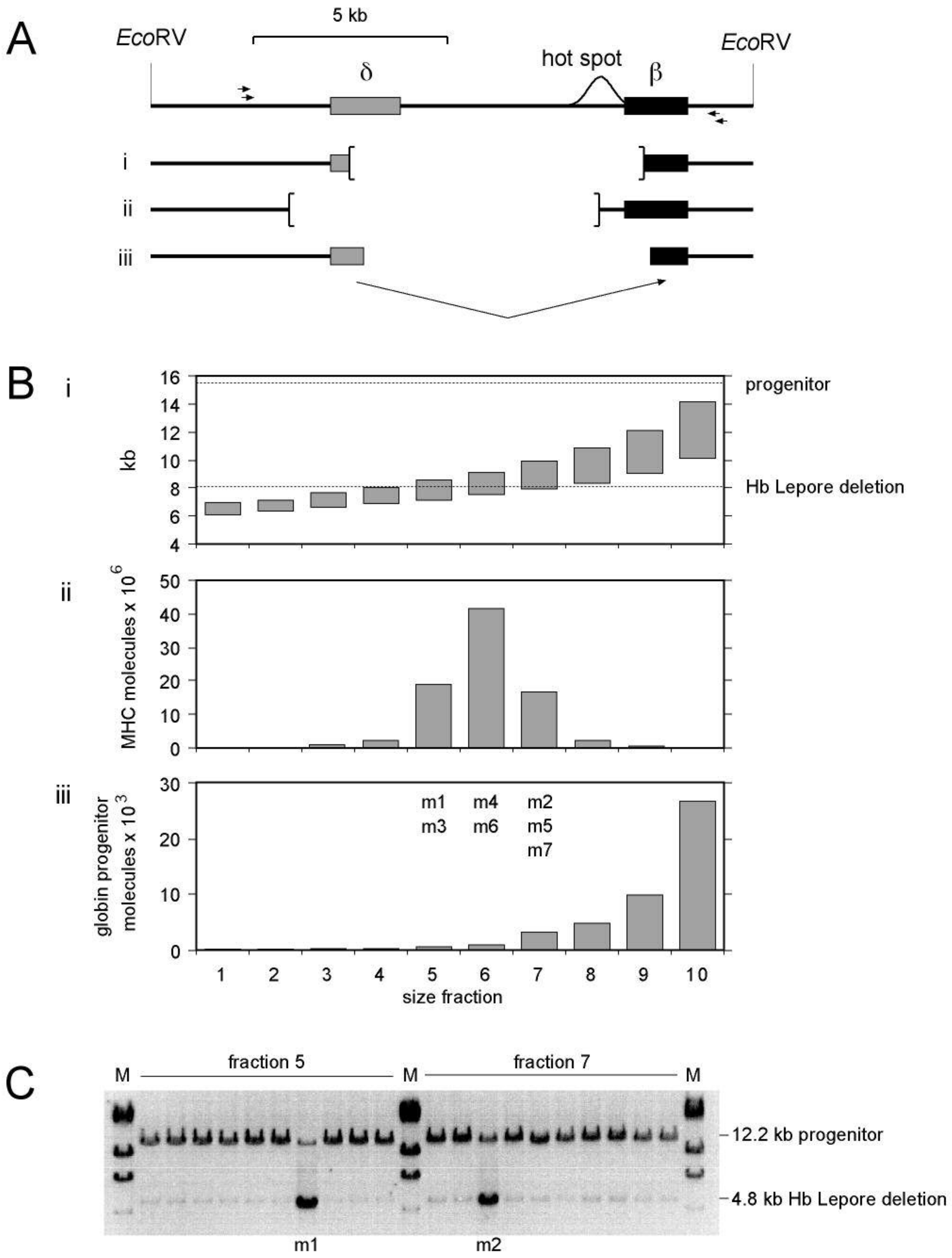
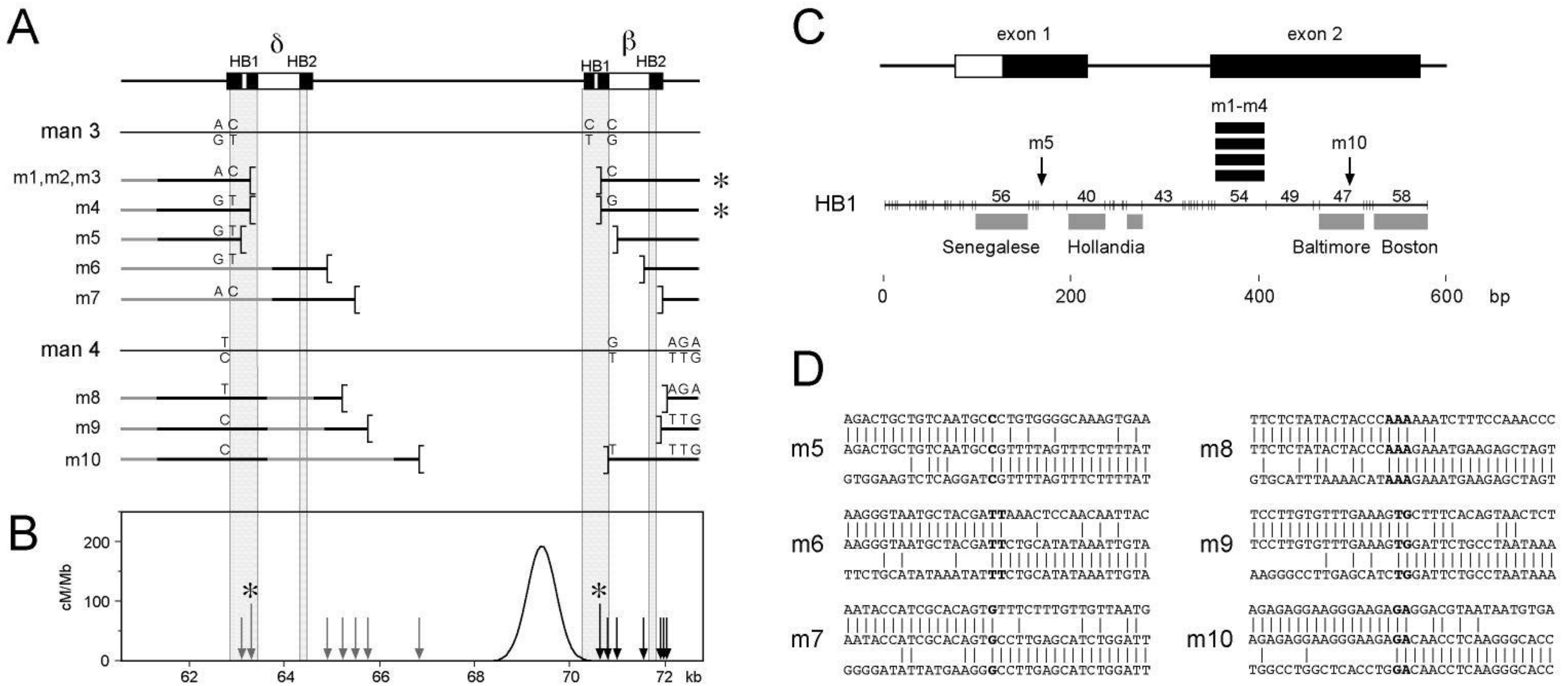

A
5 kb
EcoRV
EcoRV
hot spot
δ
β
i
ii
iii
B
i
16
14
12
10
8
6
4
kb
progenitor
Hb Lepore deletion
ii
50
40
30
20
10
0
MHC molecules x 10^6
iii
30
20
10
0
globin progenitor molecules x 10^3
m1
m3
m4
m6
m2
m5
m7
1
2
3
4
5
6
7
8
9
10
size fraction
C
M
fraction 5
M
fraction 7
M
12.2 kb progenitor
4.8 kb Hb Lepore deletion
m1
m2

Fig. 5


A
δ
β
HB1
HB2
man 3
m1,m2,m3
m4
m5
m6
m7
man 4
m8
m9
m10
A C
G T
C C
T G
T
C
G
T
AGA
TTG
*
B
cM/Mb
200
100
0
62
64
66
68
70
72
kb
C
exon 1
exon 2
m1-m4
m5
m10
HB1
56
40
43
54
49
47
58
Senegalese
Hollandia
Baltimore
Boston
0
200
400
600
bp
D
m5
AGACTGCTGTCAATGCCCTGTGGGGCAAAGTGAA
AGACTGCTGTCAATGCCGTTTTAGTTTCTTTTAT
GTGGAAGTCTCAGGATCGTTTTAGTTTCTTTTAT
m6
AAGGGTAATGCTACGATTAAACTCCAACAATTAC
AAGGGTAATGCTACGATTCTGCATATAAATTGTA
TTCTGCATATAAATATTTCTGCATATAAATTGTA
m7
AATACCATCGCACAGTGTTTCTTTGTTGTTAATG
AATACCATCGCACAGTGCCTTGAGCATCTGGATT
GGGGATATTATGAAGGGCCTTGAGCATCTGGATT
m8
TTCTCTATACTACCCAAAAAATCTTTCCAAACCC
TTCTCTATACTACCCAAAGAAATGAAGAGCTAGT
GTGCATTTAAAACATAAAGAAATGAAGAGCTAGT
m9
TCCTTGTGTTTGAAAGTGCTTTCACAGTAACTCT
TCCTTGTGTTTGAAAGTGGATTCTGCCTAATAAA
AAGGGCCTTGAGCATCTGGATTCTGCCTAATAAA
m10
AGAGAGGAAGGGAAGAGAGGACGTAATAATGTGA
AGAGAGGAAGGGAAGAGACAACCTCAAGGGCACC
TGGCCTGGCTCACCTGGACAACCTCAAGGGCACC