Crossovers are associated with mutagenesis and biased gene conversion in recombination hotspots

Barbara Arbeithuber[1], Andrea J. Betancourt[2], Thomas Ebner[3,4], Irene Tiemann-Boege[1*]

SI Appendix Content:

SI Materials and Methods

1. PCR conditions for crossover and non-recombinant collection

Allele-specific primers were designed for each SNP and used according to the required haplotype. Phosphorothioate bonds (indicated in lower case), protected the 3' primer ends from the 3'-5' exonuclease activity of the polymerase and increased the specificity of the assay. Red letters indicate additional bases in the primer sequences, not present in the genomic sequence, to adjust the annealing temperature of the primer.

Hotspot I (HSI)

Primer	SNP	Primer sequence	Product length
1st PCR forward	rs6517577	CTC AAT AGT CCA CAT GGA AAC tta (a/c)	4187 bp
1st PCR reverse	rs2299775	AGC AAT TCC CCT GGT TGt gt(t/c)	
2nd PCR forward	rs2244084	AGA ATC CAC CAT AGT GAG AGA Tagc (a/g)	3761 bp
2nd PCR reverse	rs2299774	AAA GCA GAT TGG CTC CTt gg(t/c)	

Cycling conditions:

1st PCR

94 °C	2 min	
94 °C	15 sec	5x
63 °C	15 sec	
72 °C	60 sec	
94 °C	15 sec	25x
63 °C	15 sec	
72 °C	90 sec	
72 °C	2 min	

2nd PCR

94 °C	2 min	
94 °C	15 sec	45x
56 °C	15 sec	
72 °C	60 sec	
82 °C	5 sec	
72 °C	2 min	
Melting curve (65 – 95 °C)		

Hotspot II (HSII)

Primer	SNP	Primer sequence	Product length
1st PCR forward	rs7201177	TAG GAC GTC TCT CTG ctt (c/g)	3566 bp
1st PCR reverse	rs12149730	GTA AGT GCT ATG TTC AGA ACa ga(t/c)	
2nd PCR forward	rs1861187	GCG ATT GAA ATA ATC AGG TTt ca(c/t)	3326 bp
2nd PCR reverse	rs4786855	GAA GTA GCA ATG AGA GAG AGA Aga a(t/g)	

Cycling conditions

1st PCR

94 °C	2 min	
94 °C	15 sec	5x
63 °C	15 sec	
72 °C	60 sec	
94 °C	15 sec	25x
63 °C	15 sec	
72 °C	90 sec	
72 °C	2 min	

2nd PCR

94 °C	2 min	
94 °C	15 sec	45x
56 °C	15 sec	
72 °C	60 sec	
82 °C	5 sec	
72 °C	2 min	
Melting curve (65 – 95 °C)		

2. Opposing effects of biased gene conversion and mutation

The expected GC content at equilibrium is estimated as 100% based on the formula $1/[1+\kappa(\exp(-2N_e b)]$ (1, 2), where b is the heterozygous selection coefficient favoring GC, N_e is the effective population size, and κ is the ratio of mutation rate to AT *vs.* rates to GC mutation rates (*i.e.*, the S>W mutation rate divided by the W>S mutation rate).

Transmission advantage (gBGC). The preferential transmission of GC alleles due to gBGC (expressed as b) can be obtained from the transmission bias (2, 3), and is calculated as $b = 2x-1$, where x is the fraction of over-transmition considering all gametes (crossovers and non-recombinants) and can be defined as $x = (1-c)0.5+c(p_{GC})$, with c being the crossover frequency estimated from the data (**SI Appendix, Fig. S4, Table S5**). If we assume that gBGC is restricted to male meioses, as suggested by (4), this advantage would be halved. The value for p_{GC} was calculated directly from the weighted odds-ratio, which is an estimate of the ratio of the odds of transmitting a GC allele and the odds of transmitting an AT allele, e.g, $wOR = [p_{GC}/(1- p_{GC})]/[p_{AT}/(1- p_{AT})]$. From this, we see that $p_{GC} = \sqrt{wOR} / (1+ \sqrt{wOR})$, which denotes the fraction GC alleles at polymorphic sites favored in crossovers.

Estimating N_e. To obtain an estimate of N_e specific to the local region, we used data from the 1000 genomes project from the 5kb region around HSI to calculate Watterson's θ, an estimate of $4N_e u$ (with u being the mutation rate). We used only segregating sites in Europe from sequences with genotypes with significant support (based on the genotype likelihoods given in the vcf files calculated by the 1000 genomes project). Using the observed theta of $1.36\text{x}10^{-3}$ and the corrected hotspot mutation rate for HSI (μHS) given in **Table 1** of $2.07\text{x}10^{-8}$, we obtain an N_e estimate of 16,425, very similar to the usual value of 20,000 (5). We note that while human demographic history includes dramatically changing population sizes [eg. (6)], our primary interest here is what our measurements would predict for equilibrium GC content— that is, whether AT-biased mutation or GC-biased gene conversion dominates patterns of sequence evolution in the long run.

Estimating κ. The mutation bias parameter, κ=μHS(S>W)/μHS(W>S), can be obtained from the data summarized in **Table 2.**

Taken together, we can use these parameter estimates to predict a GC content at equilibrium of 100%. If we consider that the observed rate of gBGC may only be valid for male meioses (see above), the predicted equilibrium GC content is still very high—99%. In general, this conclusion is quite robust to uncertainty in our estimates; as **SI Appendix, Fig. S3C** shows, most sites have GC alleles for values of κ and b within the 95% CI for our estimates. In fact, the equilibrium GC content dips below 50% only when the effect of biased gene conversion approaches neutrality (*i.e*, with $2N_e b$ close to 1, and $b \approx 1\text{x}10^{-5}$). The observed GC content is much lower, around 45% as described in the text. The reason is probably due to the short lifespan of recombination hotspots, though note that our analysis also ignored any effect of selection on base composition. **SI Appendix, Fig. S3A** shows that the equilibrium GC depends also on the intensity of the hotspot given as the recombination frequency, c. Assuming that the recombination frequency is reduced by a different percentage (0.2, 04, 0.6 and 0.8) from the previous step, once a very low recombination frequency is reached (~$3\text{x}10^{-6}$

equivalent to ~0.1cM/Mb a level below an active hotspot), the GC content is solely determined by the average human mutation rates μhAve (S>W) and μhAve(W>S), reaching an equilibrium GC of 31%.

3. Supporting References

1. Bulmer MG (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129(3):897-907.
2. Nagylaki T (1983) Evolution of a finite population under gene conversion. *Proc Natl Acad Sci U S A* 80(20):6278-6281.
3. Gutz H & Leslie JF (1976) Gene conversion: a hitherto overlooked parameter in population genetics. *Genetics* 83(4):861-866.
4. Duret L & Galtier N (2009) Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual review of genomics and human genetics* 10:285-311.
5. Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature reviews. Genetics* 10(3):195-205.
6. Schiffels S & Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. *Nat Genet.*
7. Pratto F, *et al.* (2014) Recombination initiation maps of individual human genomes. *Science* 346(6211):1256442.
8. Garwood F (1936) Fiducial Limits for the Poisson Distribution. *Biometrika* 28(3-4):437-442.
9. Patil VV & Kulkarni HV (2012) Comparison of Confidence Intervals For The Poisson Mean: Some New Aspects. *REVSTAT- Statistical Journal* 10(2):211-227.
10. Minton JA, Flanagan SE, & Ellard S (2011) Mutation surveyor: software for DNA sequence analysis. *Methods in molecular biology* 688:143-153.

Supporting Figures

Figure S1

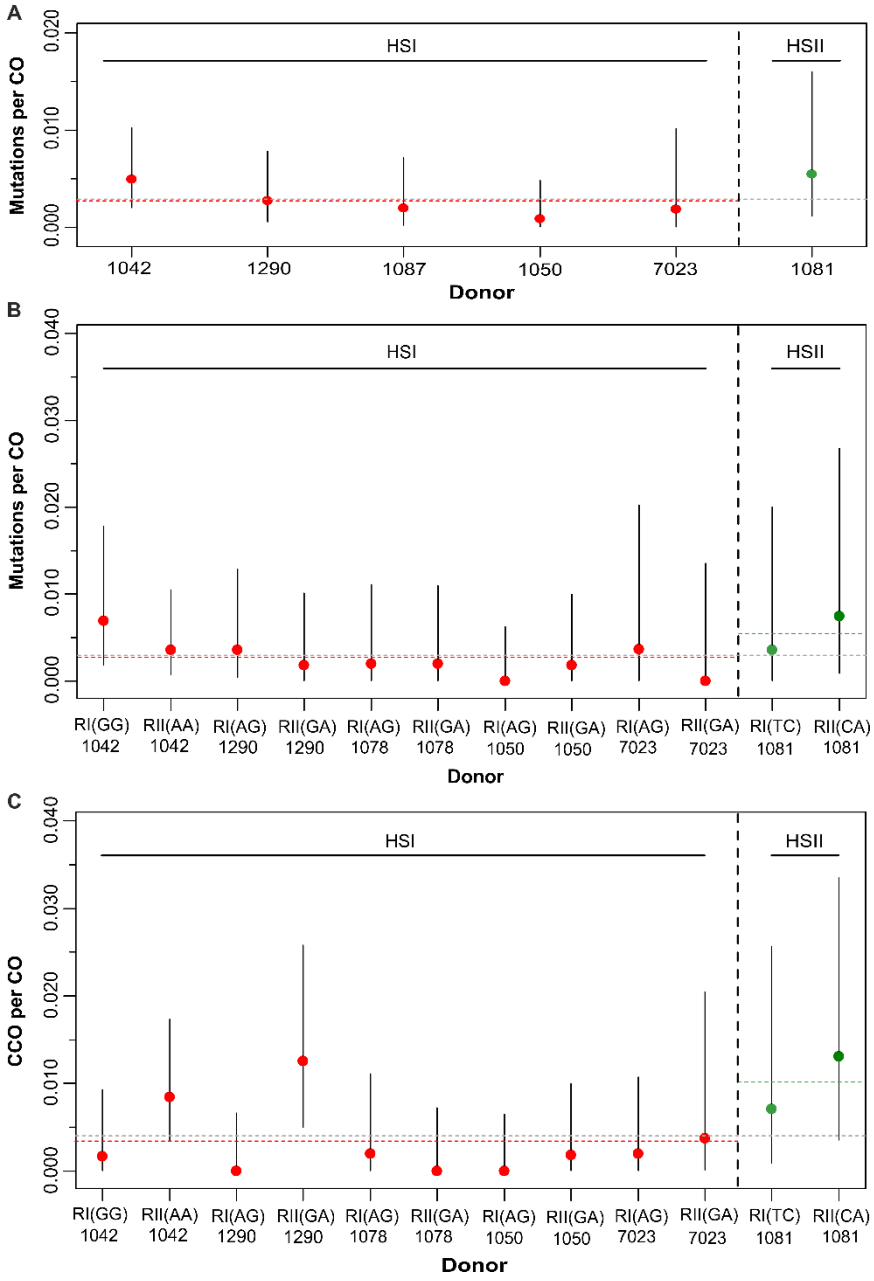


Figure S1. Analysis of differences in mutations and CCOs between donors and hotspots.
(A+B) Differences of mutation frequencies between donors and reciprocals. The number of mutations showed no hetereogeneity among donors (exact multinomial test, $p = 0.630$), or among the reciprocals treated individually (exact multinomial test, $p = 0.3593$), and were also statistically indistinguishable between hotspots (Fisher's exact test, $p = 0.215$). The dotted grey line denotes the average mutation frequency for HSI and HSII, the dotted red line for HSI, and the dotted green line for HSII. **(C) Distribution of CCO frequencies per crossover (CO) measured in six different donors for both types of CO (RI and RII).** The dashed grey line denotes the average CCO frequency per CO at 0.41% (0.26-0.60%; Poisson CI), the dashed red and green line show the average CCO frequency per CO at 0.35% (0.21-0.54; Poisson CI) and 1.02% (0.37-2.22; Poisson CI) for HSI and HSII, respectively. Donors 1042 and 1290 show larger differences in CCO frequencies per CO between reciprocals, but none are statistically significant (Fisher's exact test $p = 0.748$ for donor 1042, $p = 0.092$ for donor 1290, and $p = 1$ for all others after Bonferonni multiple-testing correction).

Figure S2

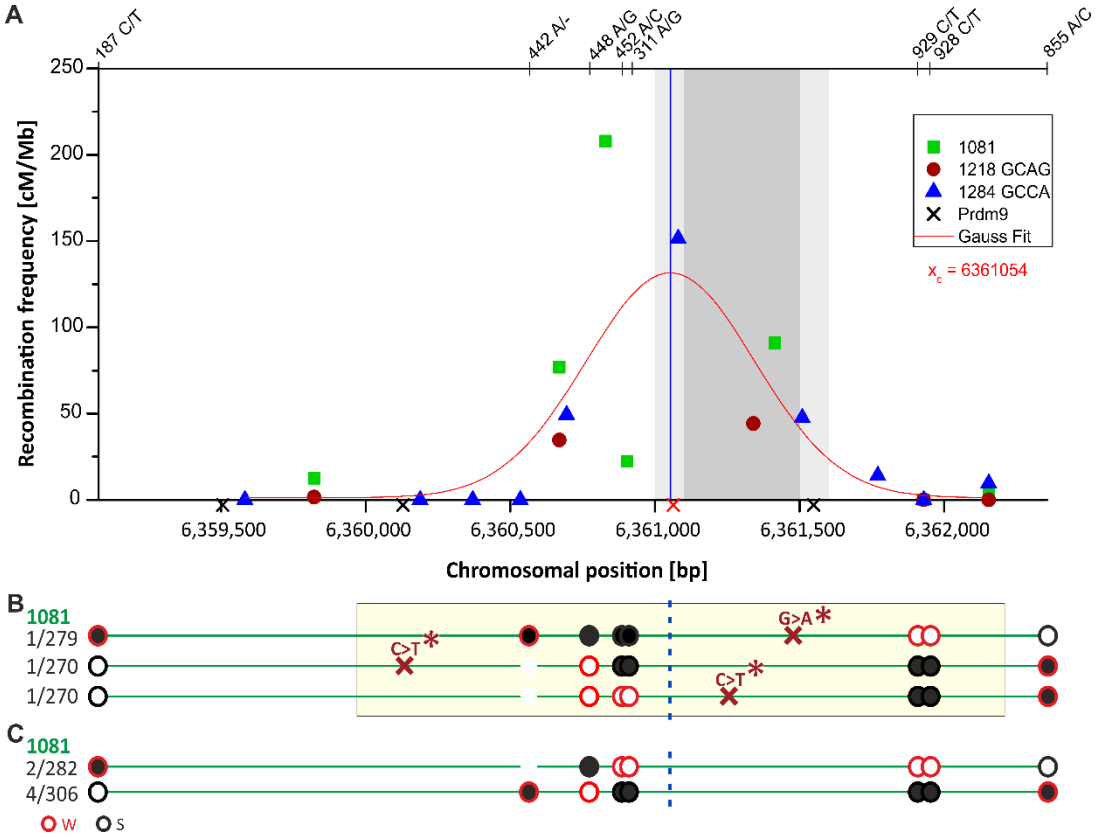


Figure S2. Crossover distribution, mutations, and CCOs in HSII.

(**A**) **CO distribution based on both reciprocal crossovers (RI+RII)**. Data comes from donor 1081 and one reciprocal of donors 1218 and 1284 (each mark represents a different donor). A best-fit normal distribution (Gaussian function) shows the hotspot center at its maximum at chr16:6,361,054 (vertical line). The region harboring the DSBs with high probability (7) is marked by the grey shaded area. Motifs for PRDM9 allele A are shown as crosses (red without mismatch, black with one mismatch) on the x-axis. (**B**) **Distribution of mutations.** The mutations identified on different haplotypes for donor 1081 are shown as red crosses (CpG sites are denoted with an asterisk). The yellow shaded area denotes the sequenced region. Aligned with the crossover distribution are black and white circles representing heterozygous SNPs with a red and black rim denoting the type of SNP (AT-Weak and GC-Strong, respectively; no rim is an InDel), whereas grey shaded circles represent homozygous polymorphisms. The vertical dotted line is the estimated hotspot center. (**C**) **CCOs identified in the same donor as above from 588 collected crossovers**. The haplotype of each CCO is shown with circles representing SNPs. The frequency of each CCO per crossover haplotype is shown to the left under the donor-ID.

Figure S3

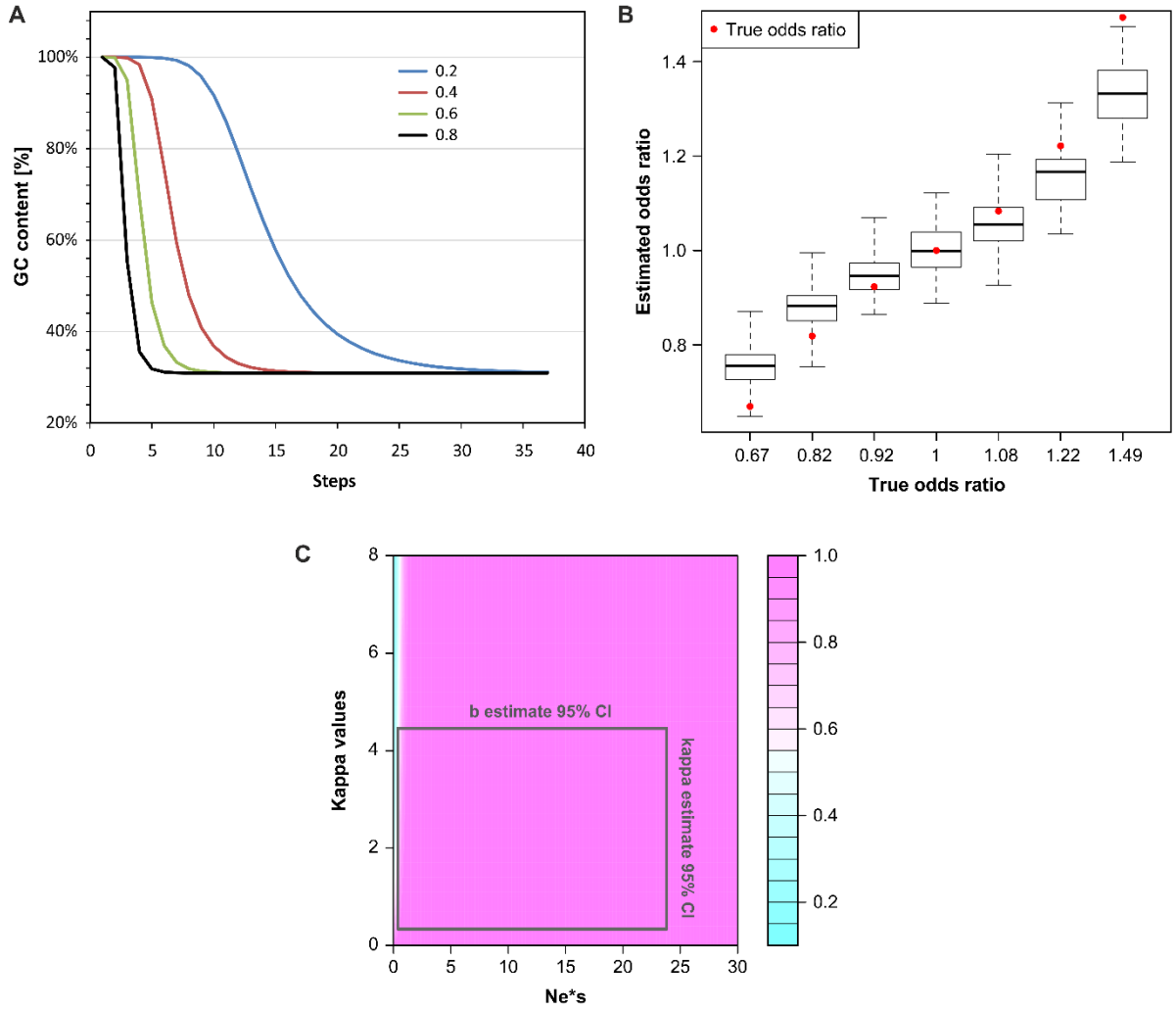


Figure S3. Analysis of gBGC and equilibrium GC content.

(A) Effect of the crossover frequency on equilibrium GC content. Equilibrium GC content was estimated by the Li-Bulmer equation (1) assuming a gBGC of 52.3% and a corrected mutation frequency μCO_{total} of 8.8×10^{-7}; $\mu CO_{S>W=}1.71 \times 10^{-6}$; $\mu CO_{W>S}=1.55 \times 10^{-7}$ estimated from the data of HSI. The crossover frequency starts at 1×10^{-2}, equivalent to 532 cM/Mb and is reduced by 20, 40, 60 or 80% from the previous step. If the crossover rate is low enough (~3×10^{-6} equivalent to 0.1cM/Mb), then the equilibrium GC is only influenced by the effect of genome wide average mutation rates reaching 31%, and the contribution of CO associated gBGC and mutagenesis is neglegible. **(B) Simulations testing the estimation procedure for the odds-ratio from the Cochrane Mantel Haenszel (CMH) used to calculate the gBGC.** As this analysis suffers from some non-independence (COs that stop distal to one SNP are also those that stop proximal to the next SNP), we evaluated its performance with simulations. We simulated CO and gene conversion under a range of biases (with 45-55% transmission of GC alelles) to obtain simulated data sets of a size corresponding to the fewest recombinants recovered per donor, and analyzed the simulated data in the same way as the real data. Recombinants from 5 donors were simulated, corresponding roughly to the minimum data set sizes for HSI (n = 601, 562, 503, 571, 275). For each recombinant, a breakpoint was chosen depending on its distance from the DSB; the locations of these breakpoints were exponentially distributed with scale parameter = 100. GC alleles were favored in the simulations according to the true input odds-ratio (x-axis and red points); odds-ratios were estimated from the simulated data as described for the real data from the CMH test (boxplots). For each odds-ratio, 100 data sets were simulated. The whiskers on the boxplots extend to the extreme simulated data points, and the red dots indicate the odds ratios used in the simulations. In addition, we also performed 50,000 simulations under the null hypothesis of equal

transmission (*i.e*, with a true odds ratio of 1, and find that we reject the null model with the CMH test less than 3% of the time, suggesting this analysis is slightly conservative. **(C) Equilibrium GC content with varying kappa and *b* values.** Equilibrium GC content was calculated for a range of kappa and $N_e b$ values. The grey rectangle indicates values within the 95% CI limits of our estimates. For kappa, confidence intervals were calculated as ±1.96 s.e, with s.e calculated from the number of mutational events in crossovers, adjusted for the non-crossover rate, using $\sqrt{((1/\mu_{SW})+(1/\mu_{WS})-(1/\text{GC-sites})-(1/\text{AT-sites}))}$. Confidence limits for *b* were calculated from the 95% CI of the odds-ratio estimates from the Cochrane-Mantel Haenszel test done on HSI data, *i.e* 1.04, 1.40. The lower (upper) limit assumes a transmission ratio consistent with this lower (upper) bound, gene conversion occurring only during male (both male and female) meioses, and is calculated using the lower (upper) bound CI for the donor with the lowest (highest) crossing over rate. The colors in the heatmap indicate the percent GC content expected at equilibrium given the values of *b* and kappa, calculated using the Li-Bulmer equation.

Figure S4

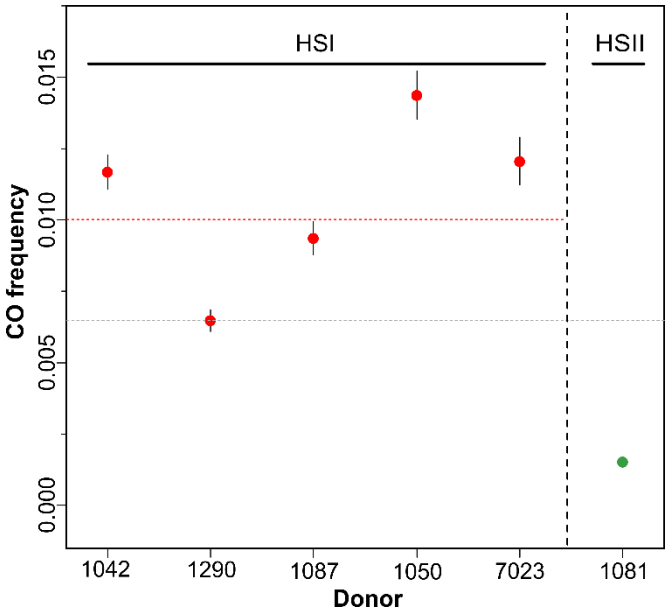


Figure S4. Estimation of crossover frequency between donors.

Individual donor crossover frequencies. Crossovers were measured in HSI or HSII in a 3761 bp or 3326 bp region, respectively; in a total of 6061 sequenced samples (25 complex crossover sequences were not included). Poisson confidence intervals (CI) of crossover frequencies were calculated according to Garwood 1938 (8), following (9), with lower and upper bounds of $\chi^2_{2x,\ 0.025}/2$ and $\chi^2_{2x+1,0.975}/2$, respectively, where x is the observed number of crossovers. CI for rates were determined by dividing these limits by the number of total amplifiable meiosis. The dashed red or grey lines show the average crossover (CO) frequency for HSI or for both HSI + HSII, respectively.

Figure S5

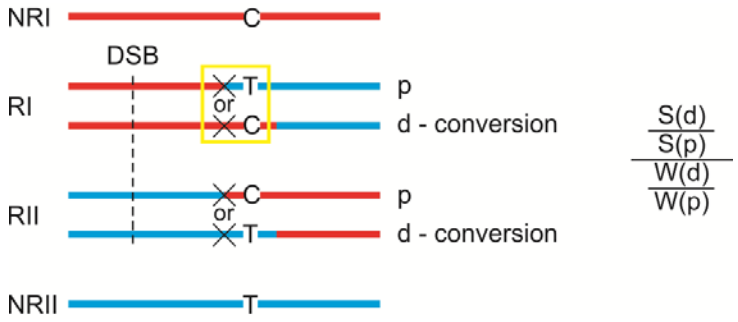
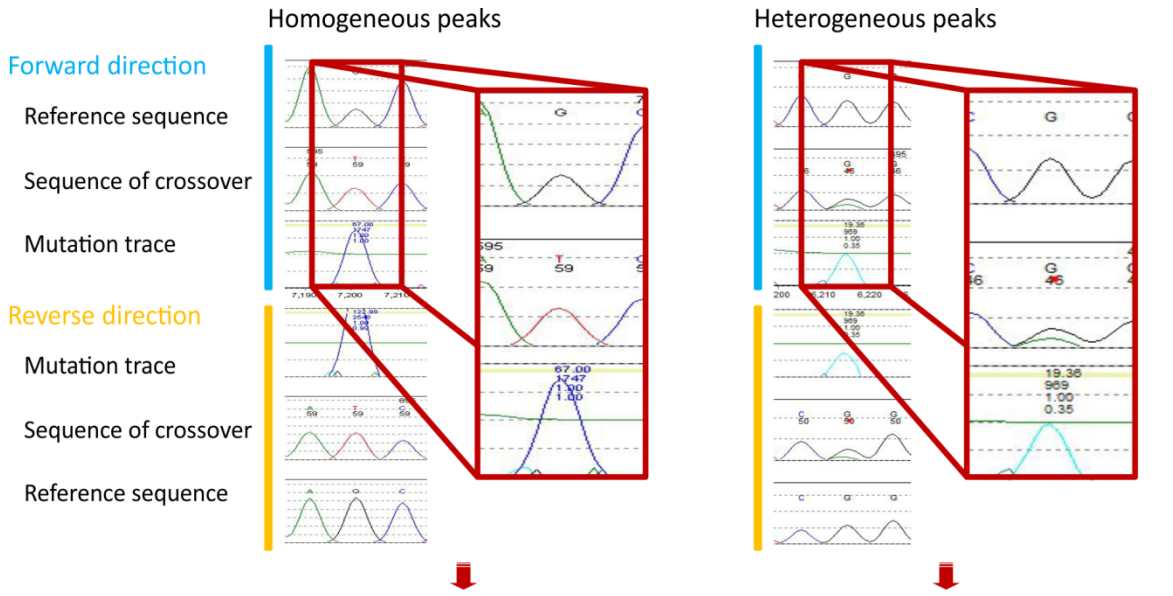


Figure S5. Rationale for the test of an effect of strong (S) vs. weak (W) alleles on the distribution of crossover reciprocals.

Recombination occurs between a red haplotype and blue haplotype. The DSB point is shown as a dotted line; from this point, crossovers can end at any point up- or downstream from the DSB (the figure shows a downstream crossover breakpoint marked with the cross). With equal transmission, the ratio of proximal (p) and distal (d) alleles recovered should be the same for both reciprocal crossovers. But if heteroduplexes preferentially resolve as a strong allele (eg. C) via a conversion event (yellow box), there will be more crossovers that seem to end distal from the C polymorphism in the RI crossover type, than crossovers that seem to end proximal to the C in the RII crossover type.

Figure S6



Dir.	Mutation	Drop	Overlap	SnRatio	Qual. (0-100)
F	40055G>T	1.00	1.00	4.94	59
R	40055G>T	0.99	1.00	7.21	59

Dir.	Mutation	Drop	Overlap	SnRatio	Qual. (0-100)
F	39953G>GA	0.35	1.00	3.00	46
R	39953G>GA	0.35	1.00	3.00	50

Figure S6. Sequence analysis.

Sequence analysis was performed using the Mutation Surveyor software (10). An example of a homogeneous and heterogeneous chromatogram peak is shown for both forward and reverse direction. Based on a number of parameters, such as the fraction of drop, overlap, signal to noise ratio and quality scores, the Mutation Surveyor package categorizes positions with alternate nucleotides as homogenous or heterogenous mutations. The alternate trace is compared to the reference sequence, which is a consensus chromatogram of all the sequencing reads.

Supporting Tables

Table S1. Mutations in crossovers and non-recombinant controls.

Donor	HS	Recipr.	#CO or #NR	# mut	mut/CO or mut/NR (%)	Effective sequenced sites			mut type	Position (hg19)	2nd PCR repeats	drop forward	drop reverse	Sequence context	Distance to HS center (bp)	Distance to CO	CO region size (bp)	Rel. to CO
						Total (Mb)	CpG	S (G/C)										
1042	I	RI (GG)	577	4	0.693	1.3271	34620	619121	C → T	41277650	4x	1.00/0.99/1.00/1.00	0.99	CTGCAAT	-860	~ - 1380 bp	154	x ⌊
									G → A	41278433	1x	0.87	0.98	CCTGCCC	-77	at CO	1084	⌊
									G → A	41278855	2x	0.96	0.99/1.00	CACGCGC	345	at CO	1084	⌊
									G → A	41279487	2x	1.00/1.00	1.00/1.00	GCCGAGG	977	~ 1000 bp	1084	⌊ x
		RII (AA)	836	3	0.359	1.9228	50160	897028	C → T	41279039	2x	1.00/1.00	1.00	GCACGGA	529	~ 800 bp	1084	⌊ x
									G → A	41279077	2x	1.00/0.99	1.00	GCTGTCA	567	~ 800 bp	1084	⌊ x
									C → T	41279315	2x	0.99/0.99	0.90	ACCCAGA	805	~ 900 bp	1084	⌊ x
1290	I	RI (AG)	562	2	0.356	1.2926	33720	603588	C → T	41277923	3x	0.98/1.00/0.99	0.98/0.95/0.97	AGCCGAG	-587	at CO	1084	⌊
									C → T	41278531	3x	0.95/1.00/1.00	0.99/0.94/0.93	CTCCGTC	21	at CO	1084	⌊
		RII (GA)	553	1	0.181	1.2719	33180	593369	C → T	41278329	3x	1.00/0.99/0.96	0.86/0.98/0.98	TGGCAAA	-181	at CO	1084	⌊
1087	I	RI (AG)	504	1	0.198	1.1592	30240	541296	T → C	41278834	3x	1.00/0.97/0.97	1.00/1.00/1.00	TTTTATG	324	~850 bp	544	⌊ x
		RII (GA)	510	1	0.196	1.1730	31620	548250	G → A	41279231	3x	1.00/0.93/0.96	0.99/0.94/0.96	TTCGGAC	721	~900 bp	39	⌊ x
1050	I	RI (AG)	595	0	0.000	1.3685	35700	637840	-	-	-	-	-	-	-	-	-	-
		RII (GA)	562	1	0.178	1.2926	33720	602464	C → T	41279582	4x	1.00/0.90/0.97/0.99	0.97	GGGCGTG	1072	at CO	2462	⌊
7023	I	RI (AG)	276	1	0.362	0.6348	16008	295596	G → A	41277901	3x	1.00/1.00/1.00	1.00/1.00/1.00	CCAGGAG	-609	~ - 550 bp	392	x ⌊
		RII (AG)	272	0	0.000	0.6256	17408	292672	-	-	-	-	-	-	-	-	-	-
1081	II	RI (TC)	279	1	0.358	0.5859	16182	269793	G → A	6361480	1x	0.98	0.90	TCCGCTC	426	at CO	986	⌊
		RII (CA)	270	2	0.741	0.5670	15120	260550	C → T	6360138	1x	1.00	0.92	CCTCGGC	-916	~- 650 bp	113	x ⌊
									C → T	6361259	1x	0.98	0.90	CACCGTA	205	at CO	986	⌊
		TOTAL	**5796**	**17**	**0.293**	**13.221**	**347678**	**6161567**										
1042	I	NRI (GA)	551	1	0.181	1.2673	33060	591223	G → A	41277790	1x	0.99	1.00	CACGGTG	-	-	-	-
		NRII (AG)	558	0	0.000	1.2834	33480	598734	-	-	-	-	-	-	-	-	-	-
1290	I	NRI (AA)	278	0	0.000	0.6394	16680	298572	-	-	-	-	-	-	-	-	-	-
		NRII (GG)	282	0	0.000	0.6486	16920	302586	-	-	-	-	-	-	-	-	-	-
1087	I	NRI (AA)	283	0	0.000	0.6509	16980	303942	-	-	-	-	-	-	-	-	-	-
		NRII (GG)	286	0	0.000	0.6578	17732	307450	-	-	-	-	-	-	-	-	-	-
1050	I	NRI (AA)	868	1	0.115	1.9964	52080	930496	C → T	41278301	1x	0.91	1.00	TATCTCA	-	-	-	-
		NRII (GG)	N/A	N/A	N/A	N/A	N/A	N/A	-	-	-	-	-	-	-	-	-	-
1081	II	NRI (TA)	286	1	0.350	0.6006	16588	276562	G → A	6361109	1x	1.00	1.00	GCCGACA	-	-	-	-

NRII (CC)	280	0	0.000	0.5880	15680	270200	-	-	-	-	-	-	-	-	-	-	-	-
TOTAL	**3672**	**3**	**0.082**	**8.3324**	**219200**	**3879765**												

Mutations in crossover (CO) products (both reciprocals: RI and RII), and in single non-recombinants (NRI and NRII) assayed using the same experimental conditions as for crossovers, were analyzed in six Caucasian donors (aged 27-40 years). The number of sequenced single COs (#CO) and single NRs (#NR), the total amount of nucleotides sequenced (Mb), the effective sequenced sites classified as CpG or Strong (G/C), the number of *de novo* mutations identified (#mut), and the position of the mutation in the hg19 genome assembly is given for the different hotspots (HSI and HSII), located on chromosome 21 and 16, respectively. For most of the identified mutations, the 2nd PCR for CO collection was repeated multiple times and verified by sequencing again (confirming the mutation in all cases). Mutations were called by assessing the dropping factor (drop) of the chromatogram peak from the forward and reverse sequencing reads using the Mutation Surveyor software. Colored letters show the mutated nucleotide in its sequence context with green denoting a CpG site. The hotspot center was calculated according to a best-fit normal distribution (Gaussian function) of the crossover distribution (for HSI, chr21:41278510, and HSII, chr16:6361054). Symbols in the last column show the location of the mutation relative to the CO; xլ denotes the mutation is located upstream of the CO, լ is within the CO region, and լx is downstream of the CO. There was no evidence of heterogeneity in the mutation frequency among donors or reciprocals **(SI Appendix, Fig. S1)**.

Table S2. CpG methylation in sperm and testis.

Sample	CpG site											Average methylation
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	
1042	0.95	0.95	0.87	0.94	0.84	0.81	1.00	0.90	0.79	0.80	0.87	
1050	0.96	0.92	0.90	0.99	0.72	0.88	0.99	-	0.85	0.80	0.83	**88%**
TOTAL	**0.96**	**0.93**	**0.89**	**0.96**	**0.78**	**0.85**	**1.00**	**0.90**	**0.82**	**0.80**	**0.85**	
Testis	0.95	0.95	0.88	0.94	0.74	0.80	0.87	0.81	0.71	0.70	0.83	**83%**

CpG methylation levels were analyzed using bisulfite sequencing for 11 CpG sites of HSI lying within three regions, 218 bp, 156 bp, and 392 bp in size, distributed over the hotspot. The center of these regions is 359 bp, -21 bp, and 85 bp from the hotspot center, respectively. Sperm DNA of two donors (1042 and 1050) and DNA from one testis biopsy of a different Caucasian donor were analyzed. Percent methylation was estimated with the Mutation Surveyor software via the analysis of the dropping factor of the chromatograms obtained from amplicons after bisulfite treatments. Sperm DNA showed an average methylation level of 88% summed over all analyzed sites, the mean methylation level in testis DNA is 83%.

Table S3. Transmission bias of haplotypes.

Donor	SNP	RI	RII	SNP_RI	SNP_RII	nRI	nRII	%GC RI	%GC RII	Distance to HS center (bp)	chi-square residuals RI vs RII	chi-square residuals strong vs. weak
1042	rs2244084	Aag*tcgaG	**Ggt*caagA**	A	G	92213	152560					
1042	rs2299762	**G** ag*tcgaG	A **gt*caagA**	a	g	1	3	63%	38%	-810	-0.80	1.29
1042	rs2299765	**Gg** g*tcgaG	Aa **t*caagA**	g	t	16	9	75%	25%	-665	3.72*	3.48*
1042	rs2244287	**Ggt** *tcgaG	Aag ***caagA**	t	c	536	760	63%	38%	419	-2.44	-2.49
1042	rs968582	**Ggt*c** cgaG	Aag*t **aagA**	c	a	13	4	75%	25%	445	**5.93***	**5.68***
1042	rs2244297	**Ggt*ca** gaG	Aag*tc **agA**	g	a	29	40	63%	38%	599	-0.01	-0.30
1042	rs2299767	**Ggt*caa** aG	Aag*tcg **gA**	a	g	4	6	50%	50%	1073	-0.17	0.08
1042	rs2299774	**Ggt*caag** G	Aag*tcga **A**	G	A	2	5	63%	38%	2254	-0.86	-0.94
											X^2 = 50.6, p = 0.0005	X^2 = 47.3, p = 0.0005
1290	rs2244084	Gggtt*tcgaG	**Aaacg*caatA**	G	A	157935	187564					
1290	760	**A** ggtt*tcgaG	G **aacg*caatA**	g	a	3	1	50%	40%	-817	1.97	1.99
1290	rs2299762	**Aa** gtt*tcgaG	Gg **acg*caatA**	g	a	0	0	40%	50%	-810	-	-
1290	rs2299763	**Aaa** tt*tcgaG	Ggg **cg*caatA**	t	c	2	4	30%	60%	-721	-1.03	1.47
1290	rs2299765	**Aaac** t *tcgaG	Gggt **g*caatA**	t	g	1	1	40%	50%	-665	-0.02	0.03
1290	rs2244287	**Aaacg** *tcgaG	Gggtt ***caatA**	t	c	531	515	50%	40%	419	1.27	-0.36
1290	rs968582	**Aaacg*c** cgaG	Gggtt*t **aatA**	c	a	7	5	60%	30%	445	0.86	0.88
1290	rs2244297	**Aaacg*ca** gaG	Gggtt*tc **atA**	g	a	15	16	50%	40%	599	-0.32	-0.28
1290	rs2299768	**Aaacg*caa** aG	Gggtt*tcg **tA**	a	t	1	6	40%	50%	1177	-2.08	-
1290	rs2299774	**Aaacg*caat** G	Gggtt*tcga **A**	G	A	2	5	40%	50%	2254	-1.37	-1.36
											X^2 = 12.0, p = 0.0975	X^2 = 8.80, p = 0.182
1087	rs2244084	Ggaca*aG	**AagtctA**	G	A	106990	109886					
1087	rs2299762	**A** gaca*aG	G **agtc*tA**	g	a	7	3	43%	43%	-810	2.36	2.19
1087	rs2244188	**Aa** aca*aG	Gg **gtc*tA**	a	g	154	153	29%	57%	-266	0.30	-2.85
1087	rs2244189	**Aag** ca*aG	Gga **tc*tA**	c	t	14	13	43%	43%	-208	0.33	0.11
1087	rs62236567	**Aagt** a*aG	Ggac **c*tA**	a	c	4	14	29%	57%	-169	**-2.68**	**4.67***
1087	rs2299768	**Aagtc** *aG	Ggaca ***tA**	a	t	323	325	43%	43%	1177	0.23	-
1087	rs2299774	**Aagtc*t** G	Ggaca*a **A**	G	A	1	2	43%	43%	2254	-0.69	-0.76
											X^2 = 13.2, p = 0.0310	X^2 = 28.1, p = 0.0005
1050	rs2244084	Ggttc*G	**Aacgt*A**	G	A	87508	70268					
1050	rs2299762	**A** gttc*G	G **acgt*A**	g	a	5	2	50%	50%	-810	2.08	2.04
1050	rs2299763	**Aa** ttc*G	Gg **cgt*A**	t	c	5	9	33%	67%	-721	-1.39	1.80
1050	rs2299765	**Aac** tc*G	Ggt **gt*A**	t	g	3	4	50%	50%	-665	-0.53	0.58
1050	rs2244189	**Aacg** c*G	Ggtt **t*A**	c	t	180	142	67%	33%	-208	**3.41***	3.14*
1050	rs2299774	**Aacgt** *G	Ggttc ***A**	G	A	378	404	50%	50%	2254	-3.10*	**-3.81***
											X^2 = 17.9, p = 0.0045	X^2 = 18.9, p = 0.003
7023	rs2244084	Ggct*atcgaG	**Aatc*gcaagA**	G	A	81309	49958					
7023	rs2299762	**A** gct*atcgaG	G **atc*gcaagA**	g	a	2	0	50%	50%		-	-

7023	711	**Aa** ct*atcgaG	Gg **tc*gcaagA**	c	t	36	11	40%	60%	-589	3.35*	2.14
7023	rs2244189	**Aat** t*atcgaG	Ggc **c*gcaagA**	t	c	129	88	30%	70%	-208	-3.65*	0.70
7023	rs2299766	**Aatc** *atcgaG	Ggct ***gcaagA**	a	g	188	109	40%	60%	184	-1.79	-1.27
7023	rs2244287	**Aatc*g** tcgaG	Ggct*a **caagA**	t	c	108	48	50%	50%	419	1.89	-2.84
7023	rs968582	**Aatc*gc** cgaG	Ggct*at **aagA**	c	a	9	5	60%	40%	445	-0.18	-0.79
7023	rs2244297	**Aatc*gca** gaG	Ggct*atc **agA**	g	a	26	7	50%	50%	599	3.51*	2.47
7023	rs2299767	**Aatc*gcaa** aG	Ggct*atcg **gA**	a	g	5	1	40%	60%	1073	2.24	-1.23
7023	rs2299774	**Aatc*gcaag** G	Ggct*atcga **A**	G	A	16	2	50%	50%	2254	**6.26***	**5.30***
											X^2 = 80.5, p = 0.0005	X^2 = 47.6, p = 0.0005
1081	rs1861187	C_aaa*ttC	**Tagcg*ccA**	C	T	396479	379875					
1081	rs35094442	**T** _aaa*ttC	C **agcg*ccA**	del	ins	17	55	13%	75%	-487	**-5.26***	-
1081	rs12102448	**Ta** aaa*ttC	C_ **gcg*ccA**	a	g	36	26	13%	75%	-280	2.53	-1.37
1081	rs12102452	**Tag** aa*ttC	C_a **cg*ccA**	a	c	44	47	25%	63%	-167	0.07	1.07
1081	rs199937311	**Tagc** a*ttC	C_aa **g*ccA**	a	g	1	2	38%	50%	-132	-0.63	1.12
1081	rs12445929	**Tagcg** *ttC	C_aaa ***ccA**	t	c	180	168	50%	38%	854	2.92	0.17
1081	rs8060928	**Tagcg*c** tC	C_aaa*t **cA**	t	c	0	0	63%	25%		-	-
1081	rs4786855	**Tagcg*cc** C	C_aaa*tt **A**	C	A	2	4	63%	13%	1302	-0.89	-0.67
											X^2 = 33.5, p = 0.0005	X^2 = 4.27, p = 0.352
											X^2 = 47.6, p = 0.0005	**X^2 = 207.6, p = 0.0005**

Crossover haplotypes measured in the six donors for both reciprocals. Asterisks within the haplotype denote the hotspot center and the underlined position corresponds to the reported SNP. The respective numbers of haplotypes (nRI or nRII) are given for the six assayed donors. The first row for each donor denotes the non-recombinant haplotype and the number of amplifiable meioses. The GC content is estimated as the proportion of GC (S) alleles of the heterozygous alleles of that haplotype. The HS center was calculated according to a best-fit normal distribution (Gaussian function) of the crossover distribution (for HSI, at chr21:41278510, and HSII, at chr16:6361054). For each donor, we used the chi-square test to examine the data for transmission biases. The null hypothesis predicts equal transmission for both alleles; we therefore calculated the expected number of haplotypes with RI or strong alleles based on the transmission rate of the RII or weak allele haplotype and vice versa (*i.e*, expected_RI = nRII*totalRI/totalRII; where total denotes the sum of all crossovers collected per reciprocal for that donor). We tested for deviation from expected values, for each donor and overall, using chi-square tests; *p* values were obtained by simulations under the null hypothesis of equal transmission (2000 iterations), as there were small expected count numbers for some entries. The standardized Pearson residual chi-square values are given for each site, with values above zero indicating an excess of the haplotype containing the RI, and those below indicating excess RII; we considered cells with absolute chi-square residual values larger than 3 (marked with an asterisk) to have significantly unequal transmission (with $p < 0.003$). Haplotypes that have the strongest evidence of heterogenity are marked in bold.

Table S4. Complex crossovers (CCO).

Donor	HS	Recipr.	# COs	Mb	# CCOs	CCO/CO (%)	95% Poisson CI (%)		Samples with CCO type	Possible HTs	Converted SNP	Type	Distance to HS Center	Distance to CO (bp)	cM/ Mb	CO region size (bp)	Difference RI and RII (p-value)
							lower	upper									
1042	I	RI (GG)	602	1.38	1	0.166	0.004	0.926	1	**C-G**-g-t-t-c-g-g-**G-G**	rs2299767	A → G	1073	~ 1197	222.8	1084	0.748
		RII (AA)	834	1.92	7	0.839	0.337	1.729	2	**A-A**-g-g-c-a-a-g-**A-A**	rs2299765	T → G	-665	~ 470	1.3	651	
										A-A-g-g-c-a-a-g-**A-A**	rs2299762	A → G	-810	~ - 687	190.9	1084	
									3	**A-A**-a-g-c-c-a-g-**A-A**	rs968582	T → C	450	~ 568	190.9	1084	
										A-A-a-g-c-c-a-g-**A-A**	rs2244287	A → C	424	~ - 103	70.7	154	
									1	**A-A**-a-g-c-c-g-a-**A-A**	rs2244287	T → C	419	~ - 1248	1.2	1181	
									1	**A-A**-a-g-t-a-g-g-**A-A**	rs2244297	A → G	599	~ 167	41.9	26	
										A-A-a-g-t-a-g-g-**A-A**	rs968582	C → A	445	~ - 3913	3.4	474	
		total	**1436**	**3.3**	**8**	**0.557**	**0.241**	**1.098**									
1290	I	RI (AG)	562	1.29	0	0.000	0.000	0.656	-	-	-	-		-	-	-	0.092
		RII (GA)	560	1.29	7	1.250	0.503	2.575	4	**C-G**-g-g-t-t-c-c-a-t-**A-A**	rs968582	A → C	450	~ 568	129.8	1084	
										C-G-g-g-t-t-c-c-a-t-**A-A**	rs2244287	T → C	424	~ - 103	28.4	154	
									2	**C-G**-g-g-c-t-c-a-a-t-**A-A**	rs2299765	G → T	-665	~101	12.3	89	
										C-G-g-g-c-t-c-a-a-t-**A-A**	rs2299763	T → C	-721	~ -598	129.8	1084	
									1	**C-G**-g-g-t-t-t-a-g-t-**A-A**	rs2244297	A → G	599	~167	52.6	26	
										C-G-g-g-t-t-t-a-g-t-**A-A**	rs968582	C → A	445	~ -443	2.8	578	
		total	**1122**	**2.58**	**7**	**0.624**	**0.251**	**1.285**									
1087	I	RI (AG)	504	1.16	1	0.198	0.005	1.105	1	**A-A**-a-g-c-c-a-**G-G**	rs62236567	A → C	-169	~ 68	107.1	58	1
										A-A-a-g-c-c-a-**G-G**	rs2244189	T → C	-208	~ - 712	106.5	1346	
		RII (GA)	510	1.17	0	0.000	0.000	0.723	-	-	-	-		-	-	-	
		total	**1014**	**2.33**	**1**	**0.099**	**0.002**	**0.549**									
1050	I	RI (AG)	571	1.31	0	0.000	0.000	0.646	-	-	-	-		-	-	-	1
		RII (GA)	562	1.29	1	0.178	0.005	0.991	1	**C-G**-g-c-t-t-**A-A**	rs2299765	G → T	-665	~ 100	59.7	89	
										C-G-g-c-t-t-**A-A**	rs2299763	T → C	-721	~ - 285	183.6	457	
		total	**1133**	**2.61**	**1**	**0.088**	**0.002**	**0.492**									
7023	I	RI (AG)	520	1.2	1	0.192	0.005	1.071	1	**A-A**-a-c-c-a-t-c-g-a-**G-G**	rs2244189	T → C T	-208	~ 487	74.7	231	1
										A-A-a-c-c-a-t-c-g-a-**G-G**	711	→ C	-579	~ - 567	256.8	392	
		RII (GA)	272	0.63	1	0.368	0.009	2.048	1	**C-G**-g-c-t-a-c-c-a-g-**A-A**	rs968582	A → C	445	~ 144	180.9	235	
										C-G-g-c-t-a-c-c-a-g-**A-A**	rs2244287	T → C	419	~ - 103	40.3	154	
		total	**792**	**1.82**	**2**	**0.253**	**0.031**	**0.912**									
total HSI			**5497**	**12.6**	**19**	**0.346**	**0.208**	**0.540**									

1081	II	RI (TC)	282	0.59	2	0.709	0.086	2.562	2	**C-T**-_-g-a-a-t-t-**C-A** **C-T**-_-g-a-a-t-t-**C-A**	rs12102448 rs35094442	A → G A → _	-280 -487	~ 952 ~ - 264	2.4 82.5	1490 113	1
		RII (CA)	306	0.64	4	1.307	0.356	3.374	4	**G-C**-a-a-c-g-c-c-**A-G** **G-C**-a-a-c-g-c-c-**A-G**	rs12102448 rs35094442	G → A _ → A	-280 -487	~ 952 ~ - 264	8.1 91.8	1490 113	
total HSII			**588**	**1.23**	**6**	**1.020**	**0.374**	**2.221**									
TOTAL HSI + HSII			**6085**	**13.9**	**25**	**0.411**	**0.266**	**0.606**									

CCOs in both reciprocals (RI and RII) were analyzed in five different donors for HSI and in one donor for HSII. For most CCOs, there are two different possibilities for the location of the conversion, so both possible haplotypes (HTs) are shown. The change in color (black-red) indicates the location of the COs, green letters show the converted SNP. The hotspot center was calculated according to a best-fit normal distribution (Gaussian function) of the crossover distribution (for HSI, near chr21:41278510, and HSII, near chr16: 6361054). Differences in CCO frequency between RI and RII were tested for significance using the Fisher's exact test with Bonferonni multiple-testing correction.

Table S5. Crossover frequencies in HSI and HSII.

Donor	Age	HS	#CO	Meioses	Correction factors	Amplifiable meioses	bp	cM/Mb	CO_freq (x 10^{-3})	CI for CO_freq (x 10^{-3})	
										upper	lower
1042	40	I	1428	1,178,300	0.208	244,773	3761	620.5	11.67	12.29	11.07
1290	35	I	1115	1,348,000	0.256	345,499	3761	343.2	6.45	6.84	6.08
1087	34	I	1014	913,800	0.237	216,876	3761	497.3	9.35	9.94	8.78
1050	37	I	1132	760,050	0.208	157,777	3761	763.1	14.35	15.21	13.53
7023	29	I	790	593,300	0.221	131,267	3761	640.1	12.04	12.91	11.21
HSI total		**I**	**5479**	**4,793,450**	**0.221**	**1,096,193**	**3761**	**531.6**	**10.00**	**10.26**	**9.73**
1081	27	II	582	1,851,600	0.419	776,355	3326	90.2	1.50	1.63	1.38
Total			**6061**	**6,645,050**		**1,872,548**		**310.9**	**6.47**	**6.64**	**6.31**

Estimates are based on the number of amplifiable meiosis, which is the number of measured sperm genomes multiplied by correction factors derived from the non-recombinant controls (Materials and Methods). For donor 7023, the number of amplifiable meiosis was determined using the average correction factor of the 4 other donors in HSI. The crossover frequency is measured as the number of crossovers (#CO) measured per number of amplifiable meiosis per length of the hotspot, expressed in centiMorgans per megabase (cM/Mb) or crossovers per amplifiable meiosis (CO_freq). Total numbers are expressed as the sum of crossovers or meiosis, total cM/Mb as averages of total cM/Mb calculated per hotspot, and total CO_freq as twice the ratio of total CO per total amplifiable meiosis (accounting for the fact that only one of the reciprocals was measured per reaction).

Table S6. Primers and annealing temperatures used for genotyping.

SNP	Forward primer		Reverse primer		T_M [°C]	Polymerase
	Primer name	Primer sequence	Primer name	Primer sequence		
rs6517577 A/C	F-6517577	CTC AAT AGT CCA CATGGA AAC tta(a/c)	OR-6517577	TGA CAT TTC TGA CACACG TT	62	Phusion
rs2244084 A/G	F-2244084	AGAATCCACCATAGTGAGAGATagc(a/g)	OR-2244084	CCCATGTGCCTCTGGTATTC	68	Phusion
rs2299762 A/G	OF-2299762	GCA AGG AAC ACC TCG GAT AA	R-2299762	TTA CAG ACA TGA TCC Accg(t/c)	60	Phusion
rs2244188 A/G	OF2244188	CCTCTTGACCAGGGTCTTGT	R2244188	GCTAAGATGTAGCCCATTaac(t/c)	64	One*Taq*
rs2244189 C/T	OF-2244189	GGGCTACATCTTAGCCAAACC	R-2244189	CCAGAGGCTAGTTAACTAAACTGatg(g/a)	66	One*Taq*
rs2299766 A/G	F-2299766	CCGC TAC ATT ATT CTCAAT GAatt(a/g)	OR-2299766	TGAAACATTTGAAACCTGGAATA	59	One*Taq*
rs2244287 C/T	OF-2244287	CCGCTTGAAAACACTTTTGC	R-2244287	CTGCTTCTGAAAAACTGcct(g/a)	66	One*Taq*
rs968582 A/C	F-968582	CAG TTT TTC AGA AGC AAA Accc(a/c)	OR-968582	GAGGACAATTCAGCCCACTC	57	Phusion
rs2244297 A/G	OF-2244297	GTACATCTGGGATTACAAAAGCA	R-2244297	GCTTGAGAGGGAGATCTACtct(t/c)	62	One*Taq*
rs2299767 A/G	F-2299767	GGGAATACAAAAATTATCTGggc(a/g)	OR-2299767	AGT TTT GGC TGG GAA AGT CC	60	One*Taq*
rs2299774 A/G	OF-2299774	AGGTCTCAGAGGAGAGGCTAA	R-2299774	AAA GCA GAT TGG CTCCTtgg(t/c)	68	Phusion
rs2299775 A/G	OF-2299775	GCA GGA TCA GCT GCTTAA AA	R-2299775	AGC AAT TCC CCT GGTTGtgt(t/c)	68	Phusion
rs7201177 C/G	F-7201177	TAG GAC GTC TCT CTG ctt(c/g)	OR-7201177	CT GGG TAT AGG GTG AGA GGA	63	One*Taq*
rs1861187 C/T	F-1861187	GCG ATT GAA ATA ATC AGG TCtca(c/t)	OR-1861187	GAA TTC AAA ACA GGC GAA CG	63	One*Taq*
rs4786855 A/C	OF-4786855	CCA GGA AGA ACC AGC ATT TC	R-4786855	GAA GTA GCA ATG AGA GAG AGA Agaa(t/g)	63	One*Taq*
rs12149730 A/G	OF-12149730	AAG TGT GCC TTG CAA ATT CC	R-12149730	GTA AGT GCT ATG TTC AGA ACaga(t/c)	63	One*Taq*

Allele-specific primers (phosphorothioate bonds are indicated in lower case), outer primers (OF = outer forward, OR = outer reverse), and the annealing temperatures used are listed. Two alternative versions of each allele-specific primer were used, one for each allele (letters in brackets).

Table S7. Sequencing primers.

Name	HS	Primer sequence
HSInt15-Reg1-fwd	I	CTTCTGATATTGATCCAGATG
HSInt15-Reg2-fwd	I	CTGGTGAACTCAGGATTGTC
HSInt15-Reg3-fwd	I	CAAGCAGGAGATATTCCAGG
HSInt15-Reg1-rev	I	GCTAAGATGTAGCCCATTAAC
HSInt15-Reg2-rev	I	GAGGACAATTCAGCCCACTC
HSInt15-Reg3-2rev	I	TGTCTGCTCACCTCAATCTCC
HSInt15-Reg3-3rev	I	CTCCACCTAATCATTGCTCT
HSII-Reg1-fwd	II	GAGGAGCTGGGAATATAGGTG
HSII-Reg1-rev	II	GCACCTGTTCTTCATAGCTTC
HSII-Reg2-fwd	II	AACAGAATCCCAGACATAGG
HSII-Reg3-fwd	II	GCAAAAGGAGATGATGTTGG
HSII-Reg3-rev	II	TTTGAATGGATTTCTGTTGC

Sequences of primers used for forward (fwd) and reverse (rev) Sanger sequencing of the three analyzed regions of HSI and HSII are shown in the table below. When a mutation was detected in a read in one direction (the first three primers listed for each HS), sequencing was repeated in the opposite direction

Table S8. Primers for CpG methylation analysis.

Name	Primer sequence	Product length	CpG
F-Region1	TGG TTT AGT TTG AGA TTT AGG	218 bp	#1 + #2
R-Region1	ACC TTT AAA AAC CTA CCC C		
F-Region2	gGA AGG AAG AAA AGG ATG AAA GG	156 bp	#3 + #4
R-Region2	AAC CTC TTC ATA TTT CAC CTA CCC		
F-Region3	ccg GGA GTT TTA TTA TGT TGG TTA GG	392 bp	#5 - #11
R-Region3	ggc AAA AAT CAA CCT TAC AAC CC		

Primers used in the amplification of bisulfite converted DNA for the methylation analysis of 11 CpG sites lying in Region 1 (41278760-41278977), Region 2 (41278412-41278566) and Region 3 (41279164-41279549) (GRCh37/hg19). Red letters indicate additional bases in the primer sequences in order to increase the annealing temperature.

Table S9. gBGC analysis.

Donor	HS	SNP	U/D from DSB	RI(p)	nRI (p)	S/W (p)	RI(d)	nRI (d)	S/W (d)	RII(p)	nRII (p)	S/W (p)	RII(d)	nRII (d)	S/W (d)	S(p)	S(d)	W(p)	W_d
1042	I	rs2299762	U	Ggg*tcgaG	16	S	Gag*tcgaG	1	W	Aat*caagA	9	W	Agt*caagA	3	S	16	3	9	1
1042	I	rs2299765	U	Ggt*tcgaG	536	W	Ggg*tcgaG	16	S	Aag*caagA	760	S	Aat*caagA	9	W	760	16	536	9
1042	I	rs2244287	D	Ggt*tcgaG	536	W	Ggt*ccgaG	13	S	Aag*caagA	760	S	Aag*taagA	4	W	760	13	536	4
1042	I	rs968582	D	Ggt*ccgaG	13	S	Ggt*cagaG	29	W	Aag*taagA	4	W	Aag*tcagA	40	S	13	40	4	29
1042	I	rs2244297	D	Ggt*cagaG	29	S	Ggt*caaaG	4	W	Aag*tcagA	40	W	Aag*tcggA	6	S	29	6	40	4
1042	I	rs2299767	D	Ggt*caaaG	4	W	Ggt*caagG	2	S	Aag*tcggA	6	S	Aag*tcgaA	5	W	6	2	4	5
1042	I	rs2299774	D	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	<NA	NA	NA	N/A	N/A	N/A	N/A
1050	I	rs2299762	U	Aattc*G	5	W	Agttc*G	5	S	Ggcgt*A	9	S	Gacgt*A	2	W	9	5	5	2
1050	I	rs2299763	U	Aactc*G	3	S	Aattc*G	5	W	Ggtgt*A	4	W	Ggcgt*A	9	S	3	9	4	5
1050	I	rs2299765	U	Aacgc*G	180	S	Aactc*G	3	W	Ggttt*A	142	W	Ggtgt*A	4	S	180	4	142	3
1050	I	rs2244189	U	Aacgt*G	378	W	Aacgc*G	180	S	Ggttc*A	404	S	Ggttt*A	142	W	404	180	378	142
1050	I	rs2299774	D	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1087	I	rs2299762	U	Aaaca*aG	154	W	Agaca*aG	7	S	Gggtc*tA	153	S	Gagtc*tA	3	W	153	7	154	3
1087	I	rs2244188	U	Aagca*aG	14	S	Aaaca*aG	154	W	Ggatc*tA	13	W	Gggtc*tA	153	S	14	153	13	154
1087	I	rs2244189	U	Aagta*aG	4	W	Aagca*aG	14	S	Ggacc*tA	14	S	Ggatc*tA	13	W	14	14	4	13
1087	I	rs62236567	U	Aagtc*aG	323	S	Aagta*aG	4	W	Ggaca*tA	325	W	Ggacc*tA	14	S	323	14	325	4
1087	I	rs2299768	D	Aagtc*aG	323	W	Aagtc*tG	1	W	Ggaca*tA	325	W	Ggaca*aA	2	W	N/A	N/A	N/A	N/A
1087	I	rs2299774	D	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
7023	I	rs2299762	U	Aact*atcgaG	36	W	Agct*atcgaG	2	S	Ggtc*gcaagA	11	S	Gatc*gcaagA	0	W	11	2	36	0
7023	I	711	U	Aatt*atcgaG	129	W	Aact*atcgaG	36	S	Ggcc*gcaagA	88	S	Ggtc*gcaagA	11	W	88	36	129	11
7023	I	rs2244189	U	Aatc*atcgaG	188	S	Aatt*atcgaG	129	W	Ggct*gcaagA	109	W	Ggcc*gcaagA	88	S	188	88	109	129
7023	I	rs2299766	D	Aatc*atcgaG	188	W	Aatc*g tcgaG	108	S	Ggct*gcaagA	109	S	Ggct*acaagA	48	W	109	108	188	48
7023	I	rs2244287	D	Aatc*gtcgaG	108	W	Aatc*gccgaG	9	S	Ggct*acaagA	48	S	Ggct*ataagA	5	W	48	9	108	5
7023	I	rs968582	D	Aatc*gccgaG	9	S	Aatc*gcagaG	26	W	Ggct*ataagA	5	W	Ggct*atcagA	7	S	9	7	5	26
7023	I	rs2244297	D	Aatc*gcagaG	26	S	Aatc*gcaaaG	5	W	Ggct*atcagA	7	W	Ggct*atcggA	1	S	26	1	7	5
7023	I	rs2299767	D	Aatc*gcaaaG	5	W	Aatc*gcaagG	16	S	Ggct*atcggA	1	S	Ggct*atcgaA	2	W	1	16	5	2
7023	I	rs2299774	D	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1290	I	760	U	Aagtt*tcgaG	0	W	Aggtt*tcgaG	3	S	Ggacg*caatA	0	S	Gaacg*caatA	1	W	0	3	0	1
1290	I	rs2299762	U	Aaatt*tcgaG	2	W	Aagtt*tcgaG	0	S	Gggcg*caatA	4	S	Ggacg*caatA	0	W	4	0	2	0
1290	I	rs2299763	U	Aaact*tcgaG	1	S	Aaatt*tcgaG	2	W	Gggtg*caatA	1	W	Gggcg*caatA	4	S	1	4	1	1
1290	I	rs2299765	U	Aaacg*tcgaG	531	S	Aaact*tcgaG	1	W	Gggtt*caatA	515	W	Gggtg*caatA	1	S	531	1	515	
1290	I	rs2244287	D	Aaacg*tcgaG	531	W	Aaacg*ccgaG	7	S	Gggtt*caatA	515	S	Gggtt*taatA	5	W	515	7	531	5
1290	I	rs968582	D	Aaacg*ccgaG	7	S	Aaacg*cagaG	15	W	Gggtt*taatA	5	W	Gggtt*caatA	16	W	7	15	5	16
1290	I	rs2244297	D	Aaacg*cagaG	15	S	Aaacg*caaaG	1	W	Gggtt*tcatA	16	W	Gggtt*tcgtA	6	S	15	6	16	1
1290	I	rs2299768	D	Aaacg*caaaG	1	W	Aaacg*caatG	2	W	Gggtt*tcgtA	6	W	Gggtt*tcgaA	5	W	N/A	N/A	N/A	N/A
1290	I	rs2299774	D	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

1081	II	rs35094442	U	Taaaa*ttC	36	W	T_aaa*ttC	17	N/A	C_gcg*ccA	26	N/A	Cagcg*ccA	55	W	N/A	N/A	N/A	N/A
1081	II	rs12102448	U	Tagaa*ttC	44	S	Taaaa*ttC	36	W	C_acg*ccA	47	W	C_gcg*ccA	26	S	44	26	47	36
1081	II	rs12102452	U	Tagca*ttC	1	S	Tagaa*ttC	44	W	C_aag*ccA	2	W	C_acg*ccA	47	S	1	47	2	44
1081	II	rs199937311	U	Tagcg*ttC	180	S	Tagca*ttC	1	W	C_aaa*ccA	168	W	C_aag*ccA	2	S	180	2	168	1
1081	II	rs12445929	D	Tagcg*ttC	180	W	Tagcg*ctC	0	S	C_aaa*ccA	168	S	C_aaa*tcA	0	W	168	0	180	0
1081	II	rs8060928	D	Tagcg*ctC	0	W	Tagcg*ccC	2	S	C_aaa*tcA	0	S	C_aaa*ttA	4	W	0	2	0	4
1081	II	rs4786855	D	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Number of recombinants recovered at each segregating site for strong (S) and weak (W) alleles of the SNP of interest (underlined) either upstrem (U) or downstream (D) the DSB center. Under the null hypothesis of 1:1 segregation, the ratio of the number of SNPs proximal (p) and distal (d) from the crossover of a strong allele should equal the ratio of the number of SNPs before and after a weak allele. Hotspots are HS: I or II; RI and RII allele indicate the allele occurring on the (arbitrarily defined) recombinant I or recombinant II haplotype; S indicates whether the RI or RII haplotypes contains either a G or C allele at this site, and W indicates whether it is A or T allele.