MEIOTIC RECOMBINATION HOTSPOTS IN HUMANS -DYNAMICS AND CONTROLLING FACTORS

Thesis submitted for the degree of
DOCTOR OF PHILOSOPHY
at the University of Leicester

by

Linda Odenthal-Hesse, geb. Odenthal
Department of Genetics
University of Leicester

March 2012

The most beautiful experience we can have is the mysterious—the fundamental emotion which stands at the cradle of true art and true science.

Albert Einstein (1879-1955)

ABSTRACT

Meiotic gene conversion has a major impact on genome diversity. Both crossovers and non-exchange conversions cluster into distinct recombination-active regions that we call hotspots. Hotspot analysis in humans has focused on the description of crossover profiles and only few hotspots had been tested for crossover and non-exchange gene conversion. Therefore, very little was known about the frequency and distribution of non-exchange conversions and how well they correlate with crossing over.

Six extremely active recombination hotspots were analysed by using small pool PCR approaches on sperm DNA to detect both types of recombinant molecules. Non-exchange conversions were detectable and arose at high frequencies (0.01-0.47%) per sperm, in addition to crossovers (0.2-0.70%). Conversion tracts were short and their distribution defined a steep conversion gradient, centred at the peak of crossover activity and probably marking the zone of recombination initiation. It was also observed that non-exchange gene conversion and crossover frequencies were positively correlated, not just between men at the same hotspot but equally when compared across several hotspots. On average, non-exchange gene conversions spanning a marker close to the centre of the hotspot occurred at 50% of the crossover frequency.

Hotspot regulation in *cis* had been described previously to results in different initiation efficiencies between interacting haplotypes. Preferential initiation on a more active haplotype, in turn leads to the overtransmission of alleles from the less active haplotype. Additional hotspots that showed a phenomenon of biased gene conversion were described in this study, with crossovers and non-exchange gene conversions influenced to the same degree. More unusually, biased gene conversion specifically affecting non-exchange events was also observed at two hotspots. Here single SNP heterozygosities appear responsible for triggering the bias in *cis*. Crossovers were not affected, which may provide evidence for distinct crossover and non-crossover pathways operating at human hotspots.

Inter-individual differences in recombination frequencies between men at a given hotspot established PRDM9 as major trans-regulator of hotspot activity. PRDM9 regulation was characterised at two hotspots activated by specific sets of PRDM9 variants. At both hotspots a sequence motif, proposed to be the PRDM9 binding site in vitro, was not found to be responsible for hotspot activation as had been predicted previously. Curiously, better motif matches were not correlated with higher crossover frequencies, and PRDM9 can in fact activate hotspot without the proposed binding motif.

ACKNOWLEDGMENT

First and foremost I want to thank Alec, not only for accepting me as your PhD student, but also for all your support and encouragement and simply for the opportunity to see you work. Alec, you will continue to be a great inspiration. I am only hoping that I will be working in the lab for as long as you do and I feel very honoured to have been your last PhD student.

I also like to thank the Boehringer Ingelheim Fonds for their generous PhD fellowship and conference stipends. Thank you for being one of the few foundations in Germany that support basic biomedical research, with no obvious benefit to industry.

To all members of G18/G19, past and present: thank you for making me feel part of the family, and thank you all for guidance and encouragement you have given me along the way. A special thank you to Rita, firstly for making tremendous amounts of 11.1xbuffer, and secondly for sharing so much of your technical know-how. A thank you also goes to all those who gave me TMAC hybridisation-support during the bumpy period, and especially Nicky, but also Ingrid, Raheleh, Milly, Gabriel and Mahmut. Additionally, I want to especially thank Adam, Raheleh, Ingrid and Tim for becoming close friends in- and outside of the lab.

I would also like to thank my family, starting with my two lads. Firstly, thank you Michael for being happy to move countries at weeks' notice in order for me to start this PhD. Seondly to my son Maximilian for literally kicking me in the guts during my final year of research, so I had to stay focused on finishing. I am very proud to be your mommy, and I am looking forward to see all that you will become. A huge thank you also belongs to the rest of my family for their love and continuous support, starting with my siblings Nadja, Niklas and Valerie, my dad Alois and Opa Otto. But most importantly, I want to thank my mother for encouraging me to go far and beyond in following my dreams and to never give up.

Und dafür widme ich Dir, meiner Mama Sigrid, diese Doktorarbeit.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abbreviations

Common SI-units as well as IUPAC nomenclature were utilised; in addition the following abbreviations were used:

A	Adenine
AE	axial element
ASO	allele-specific oligonucleotide
ASP	allele-specific PCR using allele-specific primer
ATP	Adenosine tri-phosphate
Bp	base pairs
C	Cytosine
CEPH	Centre d´Etude du Polymorphisme Humain
CO	crossover, reciprocal recombinant
cM	centi Morgan
dHJ	double Holliday junction
D-loop	displacement loop
D. melanogaster	*Drosophila melanogaster*
DSB	double-stranded break
DSBR	double strand break repair model of recombination
E.coli	*Escherichia coli*
G	Guanine
H3K4	Histone 3 Lysine 4
hDNA	heteroduplex DNA
HGE	haploid genome equivalent
HJ	Holliday junction
JM	joint molecule
Kb	kilo bases
LD	linkage disequilibrium
LDU	linkage disequilibrium unit
LE	lateral element
MAF	minor allele frequency
Mb	mega base
MDA	multiple-displacement amplified
MHC	Major Histocompatibility Complex
MMR	mismatch repair
NCO	non-crossover, non-reciprocal recombinant
Pfu	*Pyrococcus furiosus* polymerase
PAR1	major pseudo-autosomal region
PAR2	minor pseudo-autosomal region
PCR	polymerase chain reaction
SC	synaptonemal complex
S. cerevisiae	*Saccharomyces cerevisiae*
SDSA	single nucleotide polymorphism
ss	single stranded
STR	short tandem repeat
S. pombe	*Saccharomyces pombe*
T	Thymine
Taq	*Thermus aquaticus* polymerase
TD	transmission distortion
TMAC	tetramethyl ammonium chloride
ZnF	Zinc finger

Chapter One Introduction

In 1865, a long time before anything was known about DNA, the monk Gregor Mendel first described the idea of "independent assortment of traits" in his famous experiments on heredity. Traits located on different chromosomes are assorted independently during meiotic recombination, exactly as Mendel predicted. But the assortment of traits is more complex for genes located on the same chromosome. Thomas Hunt Morgan, who studied heredity in flies, had observed that some traits, like wing mutations and eye colour, are sometimes, but not always assorted independently. To explain this, he described the concept of linkage for traits that are always inherited together on the same chromosome. In addition he postulated the idea of crossing-over, a mechanism to allow alleles to be re-shuffled before being inherited (Morgan 1911). As alleles are exchanged between chromosomes through crossover (CO), the alleles segregate in a 2:2 ratio in the offspring. In early studies of recombination using asci of fungi, where all four products of meiosis can be observed, another phenomenon of meiotic recombination was discovered. Some asci displayed a departure from the expected parental ratio; they had 3:1 non-mendelian segregation patterns (Fogel and Hurst 1963). These events were termed gene conversions (Holliday 1964).

1.1 Homologous Recombination

The generation of DNA breaks lies at the basis of all recombination. Homologous recombination involves the repair of these breaks using information from a homologous target. In mitotic cells breaks can be induced in the form of DNA damage through cell stress, ionising radiation, chemical reagents, de-amination and de-purination events, and DNA breaks are usually repaired via homologous recombination with the sister chromatid. The main goal of mitotic homologous recombination is to save the cells from going into apoptosis, by repairing the genotoxic lesions as faithfully as possible.

In contrast, meiotic recombination has the intriguing feature that it deliberately introduces DNA double-stranded breaks (DSBs) (Game *et al.* 1989; Sun *et al.* 1989; Cao *et al.* 1990), and then uses the homologous chromosome as the template for repair. This process is responsible for re-shuffling the raw material for evolution, variation generated by mutation. Meiotic recombination not only promotes genetic diversity, but also ensures the correct segregation of homologous chromosomes by providing physical connections

between chromosomes. In some cases repair is initiated between ectopic sites, referred to as ectopic recombination (or non-allelic, illegitimate). Ectopic recombination events can occur between dispersed copies of sequences with 2.2 kb of homology in yeast, and are often associated with crossing over of flanking sequences to create chromosome rearrangements (Lichten *et al.* 1987).

1.2 Meiotic recombination

Meiotic recombination is tightly linked to the process of meiosis. Most sexually reproducing organisms use meiosis to halve their chromosome number before forming germ cells, so that the genome stays the same size in offspring as in parents. Halving the chromosome number is achieved through two rounds of chromosome segregation after only one round of chromosome replication. During DNA replication chromatids stay attached through their centromeres to form a pair of sister chromatids. Chromosome association is partly achieved by the generation of physical links between the chromosomes, which are provided by cohesin and the synaptonemal complex. The first meiotic division separates maternal and paternal chromosomes, while the chromosome arms are separated during the second meiotic division, much in the same way as in a mitotic division.

For the correct segregation in the first meiotic division it is not only important that the chromosomes are paired with their correct partner, but they also have to engage in chromosomal exchanges to become physically connected as bi-valents. This is achieved during Prophase I, through reciprocal recombination (COs) between maternal and paternal chromosomes. Only then it is ensured that they orient correctly on the meiotic spindle. Without recombination random disjunction and aneuploidy can occur. At every meiosis, a higher number of breaks are produced, than are repaired as crossovers. The remaining breaks are repaired as non-reciprocal exchanges that can manifest as gene conversions.

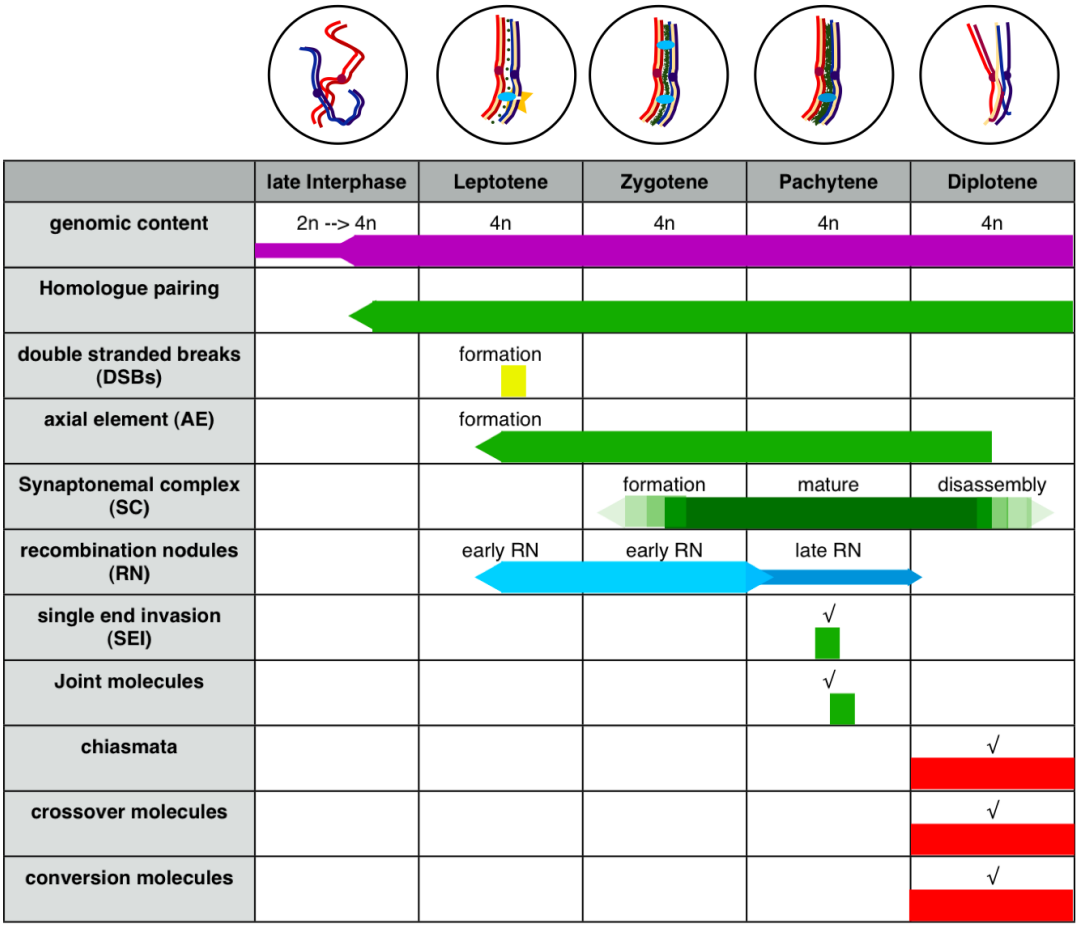
1.2.1 Cytological features of meiotic recombination

Recombination is a tightly regulated network of cytological as well as genetic interactions that are required to work together to ensure correct completion of meiosis. Cytological studies, observing structural features of the recombination process, are still used. Large cytological features, such as the synaptonemal complex, recombination nodules and chiasmata can be observed under the microscope, while proteins involved in recombina-

tion can be visualised through immuno-fluorescent staining, for example as Mlh1 foci that mark sites of crossover formation (Hassold *et al.* 2004).

The timing and occurrence of these cytological features is illustrated in Figure 1-1.

Figure 1-1 Stages of Prophase I



Homologous chromosomes are shown in blue and red. Chromosomes have replicated during interphase and then pair during leptotene and zygotene (Nag *et al.* 1995). During leptotene, axial elements (light green) are formed and DSBs are initiated. Early recombination nodules (teal) are associated with the axial elements. The SC (green) forms as early as late leptotene until mid-pachytene. DSBs occur before the SC is formed (leptotene), single-end invasion happens at the same time as the SC is forming and joint formation occurs when the SC is fully formed (pachytene). Crossover and non-crossover products appear around the same time at pachytene (represented by ticks) (Storlazzi *et al.* 1995).

1.2.1.1 The synaptonemal complex

The synapsis of homologous chromosome pairs is achieved through the scaffolding action of the Synaptonemal Complex (SC). The SC consists of an axial element (AE), lateral elements (LE) and chromatin loops. As shown in Figure 1-1, first the AEs are formed and

the condensing chromatin is then loaded on these protein cores. Only short stretches of DNA are bound to the AEs of the SC, most of the DNA is condensed into long loops, with variable loop sizes of tens or hundreds of kilo bases (kb) in length.

Chromatin loop sizes vary along the chromosome length, with smaller loops closer to the telomeres, and longer interstitial loops (Heng *et al.* 1996). This feature is shared between different species, and seems to be conserved. Although specific sequences have been proposed, which may be responsible for chromatin loading, the chromosomal position plays the dominant role in the type of chromatin packaging (Heng *et al.* 1996).

1.2.1.2 Recombination nodules

Recombination nodules are protein complexes that were originally identified as electron-dense ovoid structures in close association with the SC (Carpenter 1975). Early recombination nodules appear along the chromosome as early as zygotene (von Wettstein *et al.* 1984; Carpenter 1987), well before the appearance of double Holliday Junctions (dHJs) at mid-pachytene (Zickler and Kleckner 1999).

Early recombination nodules seem to correspond to DNA DSBs and they do not display interference. A subset of early nodules matures into late nodules, which are believed to correspond to the sites of crossovers, as they are in good agreement with the numbers and location of crossover formation, and they do display interference. Late nodules contain proteins specifically involved in crossover maturation but not gene conversion, namely Zip1, Zip2, Zip3, Msh4 and Msh5 (Ross-Macdonald and Roeder 1994; Novak *et al.* 2001; Fung *et al.* 2004).

1.2.1.3 Chiasmata

Chiasmata are the visible physical connection between homologous chromosomes, at the location where crossovers are formed (Carpenter 1994). Together with inter-sister connections provided by cohesins, this connection promotes a bi-orientation at metaphase I, which ensures the correct segregation of chromosomes during the first meiotic division (Moore and Orr-Weaver 1998; Petronczki *et al.* 2003). The need for at least one crossover per chromosome to ensure correct segregation is often termed an obligate chiasma.

1.2.1.4 Crossover interference

The presence of one chiasma at a given position decreases the probability that another chiasma occurs nearby. This property is known as interference resulting in wide spacing of crossovers along a chromosomal arm and ensuring a correct number of crossovers on each chromosome. Crossovers rarely occur within the same genetic interval (positive interference) and every chromosome, regardless of size, receives at least one crossover.

In yeast it has been shown that smaller chromosomes display less positive interference than larger ones (Kaback *et al.* 1999). As crossover interference is correlated with SC length, interference is stronger where shorter interstitial loops are bound to the axial element. As the ratio of DNA length to SC distance is smaller, the interval affected by interference is shorter and crossovers can be placed closer together. Synapsis initiation complexes containing Zip2 and Zip3 already show a pattern consistent with interference, even without synapsis or crossover formation (Fung *et al.* 2004).

1.2.2 Chronology of meiotic recombination

This section presents the steps of recombination, in a chronological order as much as possible, and the models of recombination as we understand them to date from studies in model organisms. The molecular events of the meiotic recombination machinery have been studied extensively in the budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) and the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) as well as the fungus *Ustilago maydis* (*U. maydis*). Fungal organisms are not only used because of their short generation time and relative ease for genetic manipulation, but lend themselves especially to the analysis of meiotic recombination, because all four products of a single meiosis can be recovered together, as tetrads. *Drosophila melanogaster* (*D. melanogaster*), the organism in which crossovers were first postulated, can be used for crosses between marker lines, and recombination events can be easily traced as phenotypes as there is no spontaneous meiotic recombination in males.

1.2.2.1 Double-Stranded DNA breaks and resection

Meiotic recombination is initiated by double-strand breaks (DSBs) (Sun *et al.* 1991) in most organisms. In yeast DSBs occur preferentially between genes, (Wu and Lichten 1994; Wu and Lichten 1995; Gerton *et al.* 2000) and often within the promoter regions

(Baudat and Nicolas 1997). The DNA DSBs are catalysed by the sporulation-specific 11 protein (Spo11), which remains covalently bound to the 5'-end of the break (Keeney and Kleckner 1995). Spo11 is removed together with a short oligonucleotide by the SAE2/COM1 complex and MRE11, XRS2 and Rad50 (Keeney *et al.* 1997). The Exo I protein is involved in the resection of the DSB in 3'-direction, generating single-stranded (ss) overhangs of approximately 600 nucleotides (Alani *et al.* 1990; Cao *et al.* 1990; Nairz and Klein 1997).

1.2.2.2 Strand invasion and junction molecule formation

To prevent the formation of secondary structures that would hinder recombination a nucleoprotein filament with the bacterial RecA homologues Dmc1 and Rad51 is formed around the single-stranded 3'-overhang in the budding yeast (Ogawa *et al.* 1993). The resulting nucleoproteins are active structures that search the homologous chromosome for a homologous target. Rad51 specifically fosters inter-allelic recombination (Bishop and Zickler 2004). Once the homologous target is annealed to, one strand of the homologous sequence is displaced at one end of the DSB first. This strand-exchange generates a Single-end invasion (SEIs) intermediate, which can mature into a double Holliday Junction (dHJ) after the subsequent engagement of the other end of the DSB with the homologous chromosome (Hunter and Kleckner 2001). After a joint molecule (JM) is formed by one of the free ends invading the homologous region in the donor duplex, it then acts as a primer for DNA synthesis. In meiosis in budding yeast (*S. cerevisiae*) inter-homologue recombination is highly favoured over inter-sister recombination as the latter does not join homologues and thus can lead to non-disjunction of chromosomes during meiosis; this is known as inter-homologue bias (Schwacha and Kleckner 1994; Schwacha and Kleckner 1997).

1.2.2.3 Mismatch repair

Invasion and annealing to homologous but not identical sequence can lead to the formation of single base mismatches and small insertion/deletion loops, which is referred to as heteroduplex DNA (hDNA). Heteroduplex DNA is detected and subsequently corrected by the mismatch repair machinery (MMR). Originally characterised in *Escherichia coli* mismatch repair involves recognition of a mismatched base, which is then cut out and repaired. Currently it is believed that mismatch repair preferentially occurs early during

recombination, before the HJs are resolved (Haber *et al.* 2004). But a mismatch that has escaped early mismatch repair can be repaired by late mismatch repair on the termini created by HJ resolution (Stahl and Foss 2010). Depending on which strand is used as template, repair can lead to either gene conversion or restoration of the original allele. When mismatch repair operating on heteroduplex DNA favours one allele over the other this results in biased mismatch repair, and consequently in biased gene conversion.

In yeast two independent pathways for mismatch repair have been characterised based on their dependence on Msh4. The pathway requiring Msh4 is known as the disjunction pathway; here the repair of mismatches is dependent on Mlh1, a homologue of the bacterial mismatch repair protein MutL. Mlh1-facilitated mismatch repair in the disjunction pathway restores 2:2 segregation of the marker (Stahl and Foss 2010). The Msh4-independent pathway is known as the pairing pathway; here Msh2 is responsible for mismatch repair and can lead to either gene conversion or restoration of the original marker (Stahl and Foss 2010).

1.2.2.4 Resolution

The resolution of dHJs occurs at the end of pachytene. Resolution can result in reciprocal exchanges or non-reciprocal exchanges. There are also distinct types of crossover resolution complexes, dependent on whether the resulting crossovers display interference or not. The Msh4/Msh5 resolvase-dependent type (class I) shows crossover interference, while the second type (class II) generated by the HJ resolvase MUS81/MMS4 exhibits no interference (Stahl and Foss 2010). The MMS4 resolvase is responsible for reducing crossing over on small chromosomes, which display less interference (de los Santos *et al.* 2003). Studies of homozygous-null mice revealed mus81 as a regulator of crossover frequency and placement in mammals (Holloway *et al.* 2008).

1.2.3 Models of recombination

Several models have been proposed to date to explain the generation of non-crossover recombinants (NCO) and crossover (CO) molecules based on the resolution plane of the recombination intermediate. It is now widely believed that the Double Strand Break Repair (DSBR) model of recombination mainly produces crossovers, while Synthesis-Dependent Strand Annealing (SDSA) predominantly produces non-exchange conversion.

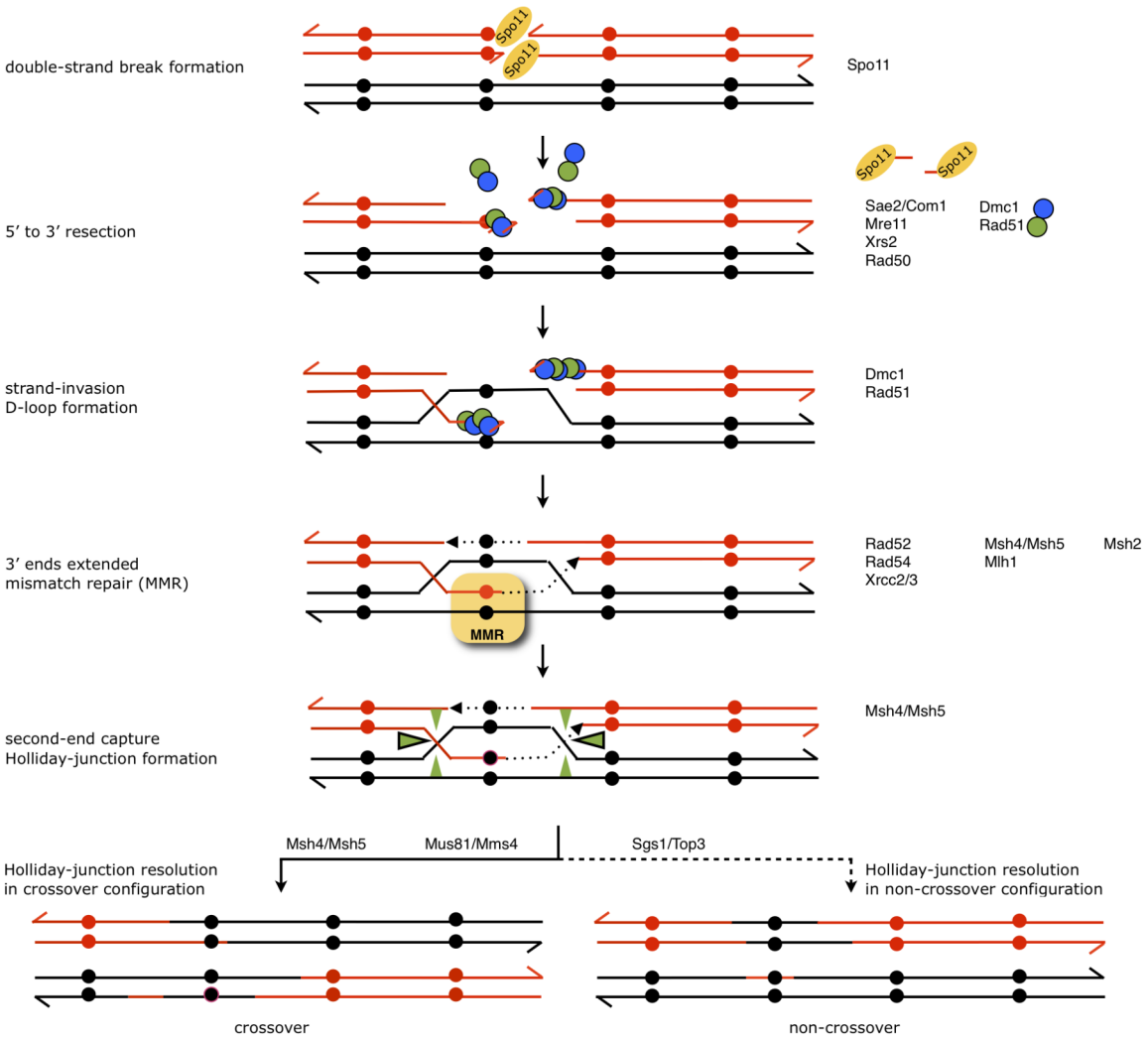
The DSBR and SDSA pathways are shown in Figure 1-2 and Figure 1-3, both modified from (Pâques and Haber 1999) and (Haber *et al.* 2004).

1.2.3.1 DSB Repair model of recombination (DSBR)

The DSBR model aims to explain the generation of crossover and non-crossover products as arising from the same initiating DSB, it was first developed by (Szostak *et al.* 1983) and later modified (Sun *et al.* 1991). The key structure in this model is a DNA joint molecule intermediate containing two HJs, which flank both sides of the break. Double HJs are generated through subsequent engagement of both strands with the homologous template, as shown in Figure 1-2. The existence of dHJs as intermediates has been verified (Schwacha and Kleckner 1995) and it has been shown that their formation is dependent on Msh5 (Börner *et al.* 2004).

In the original model it was proposed that the dHJ resection plane determines the respective end product of recombination such that resolution in the same plane creates a non-crossover molecule (cutting either both Watson-strands or both Crick-strands), while resolution in opposite plane generates crossover molecules (cutting one of the Watson and one of the Crick strands). Subsequent data inferred dHJ resolution leading only to crossovers (Allers and Lichten 2001a) as reviewed in (Bishop and Zickler 2004).

Figure 1-2 Double-strand break-repair (DSBR) model of recombination

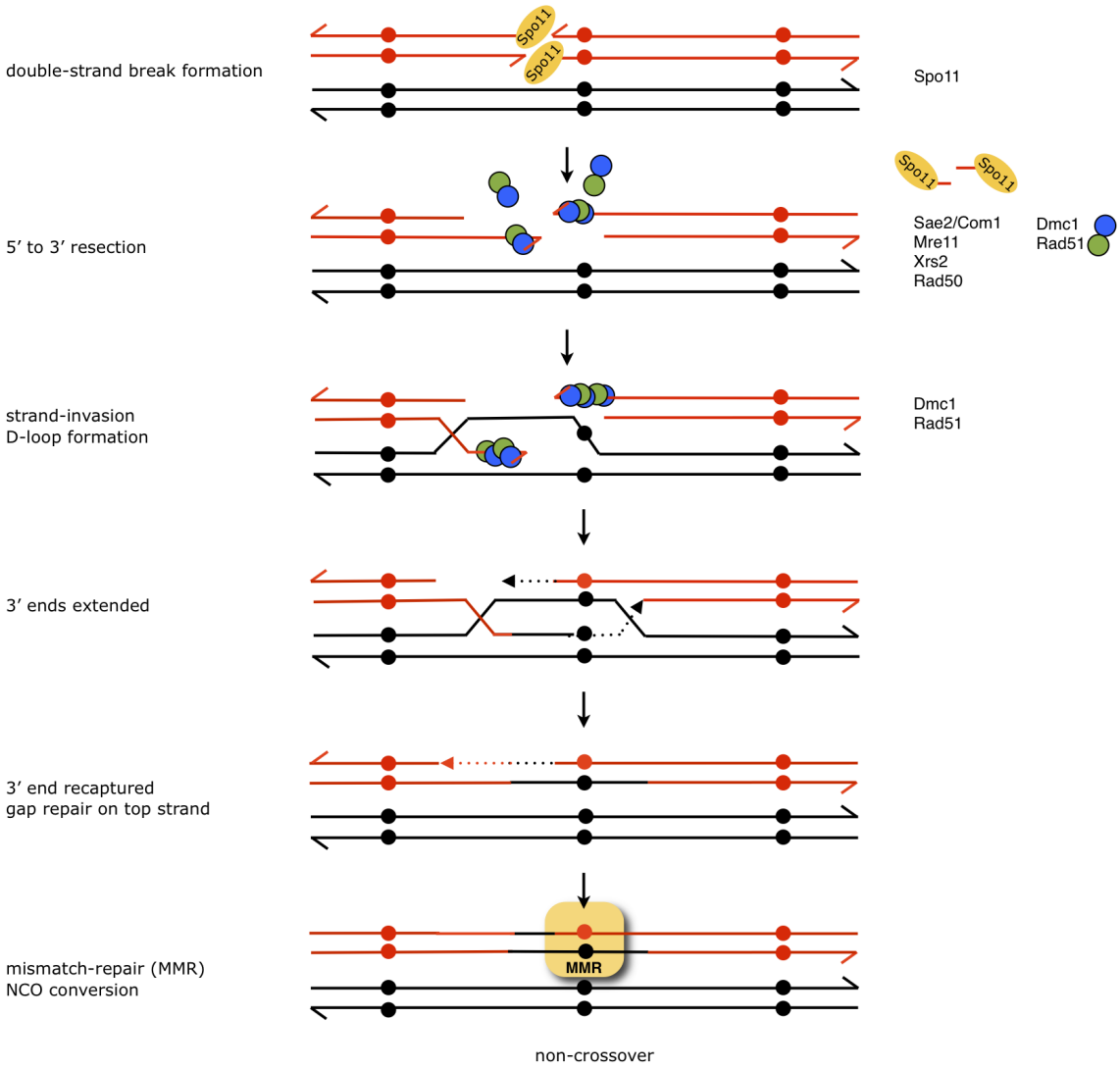


Model is presented in chronological order, from top to bottom. Proteins known to be involved in each step are listed on the right. For simplicity, not all four chromatids, but instead only the two chromatids directly involved in the break repair process are shown, indicated by lines. Chromosomal origin is differentiated by colour, with heterozygous markers represented by circles. After DNA double-strand break-formation on the red chromatid, ends are resected in 5′ to 3′ direction. Initial strand displacement on the homologous template leads to the formation of a D-loop. Subsequent DNA synthesis is used for gap-repair (indicated by hashed line), which leads to enlargement of the D-loop. Mismatch repair (MMR) at the location of heteroduplex DNA is indicated by yellow box. Once the 3′-end is recaptured a double HJ intermediate is formed. To release chromatids, the dHJ is resected (green triangles). Resolution either in crossover or non-crossover configuration is possible, depending on the resection plane.

1.2.3.2 Synthesis dependent Strand annealing (SDSA)

The SDSA model was first proposed by (Nassif *et al.* 1994) based on the finding that a high rate of complex re-arrangements, in which the breakpoints occurred at direct duplications, was observed in P-element induced gap repair in *D. melanogaster*. To account for this each of the termini would have had to invade a template for repair and serve as primers for DNA synthesis independently. In the SDSA model, both strands invade a homologous target independently and then act as primers for DNA synthesis, copying from the homologous chromosomal strand. After initial repair of the gap, they then re-anneal without formation of a dHJ. After re-annealing, full repair of the gap is then completed through copying from the sister chromatid as template. Several sub-categories of models based on the synthesis dependent strand annealing model have been proposed, some of which can generate both non-exchange conversion and crossovers as reviewed in: (Pâques and Haber 1999; McMahill *et al.* 2007; Stahl and Foss 2010).

Figure 1-3 Synthesis-dependent strand-annealing (SDSA) model of recombination



Model is presented in chronological order, from top to bottom. Proteins known to be involved in each step are listed on the right. For simplicity, not all four chromatids, but instead only the two chromatids directly involved in the break repair process are shown, indicated by lines. Chromosomal origin is differentiated by colour, with heterozygous markers represented by circles. After DNA double-strand break-formation on the red chromatid, ends are resected in 5´ to 3´ direction. After DNA synthesis (hashed lines) the 3´end is recaptured without formation of a junction intermediate, and DNA synthesis on the top strand completes gap repair.

1.2.3.3 D-loop nicking pathway in *S. pombe*

Fission yeast (*S. pombe*) recombination models vary substantially from the models proposed for recombination in budding yeast (*S. cerevisiae*), despite similarity in the initiation step, where the Spo11 orthologue Rec12 is essential for the formation of DSBs (Cervantes *et al.* 2000). Inter-sister recombination is preferred over inter-homologue recombination, and not double HJs but instead single HJs are the predominant joint molecule in meiosis (Cromie *et al.* 2006). A transient D-loop is nicked before it can form a double HJ (Cromie *et al.* 2006).

Mus81 is a part of the resolvase complex Mus81-Eme, which cuts a single HJ. In crosses involving fission yeast mutant for Mus81, which display normal numbers and kinetics of DSB formation, the number of crossovers is greatly reduced with little effect on the number of non-exchange conversion events (Cromie *et al.* 2006). The corresponding complex of Mus81/mms4 is not the major meiotic HJ resolvase in budding yeast, and mutations in mus81 only reduce the crossover frequency modestly (de los Santos *et al.* 2003).

1.2.4 The crossover / non-crossover decision

The decision whether to produce a crossover or non-crossover occurs early, before or at the appearance of SEIs (Börner *et al.* 2004). Non-crossover products are seen in the presence of mutations that prevent resolution of HJs in yeast (Allers and Lichten 2001b). Mutations in meiotic recombination proteins including Zip1, Zip2, Zip3, Mer3, Msh4 and Msh5 reduce the levels of crossover but maintain high frequencies of non-exchange conversion (Börner *et al.* 2004), pointing to the non-crossover mechanism as the main repair mechanism in the absence of crossover formation.

In contrast, when the number of DSBs is artificially reduced, crossovers tend to be maintained at the expense of non-crossovers at a given DSB location (Martini *et al.* 2006). This maintenance of crossovers points to a specific subset of DSBs "deciding" to enter a pathway of inter-homologue crossover differentiation, with the remaining DSBs resolved primarily without reciprocal exchange of chromosomal arms, as non-crossovers.

1.3 Recombination hotspots

In eukaryotic organisms studied at a sufficiently high resolution, the resolution of recombination events was not randomly distributed across the genome, but instead clustered into narrow regions highly active in recombination, so-called hotspots. A hotspot is defined as a locus or region with greater than average frequency of meiotic recombination, which is flanked by cold sequences that are suppressed for recombination. The definition of a region as a hotspot depends on the observer's frame of reference. Hotspots can be defined either by clustering of non-parental combinations of linked markers, by mapping of recombinant resolution points in the products of meiosis, or based on cytological features like chiasmata. Recombination in a hotspot begins with DSBs which are eventually repaired and resolved as crossover that results in the exchange of flanking markers, potentially as large as a whole chromosomal arm, with the remaining breaks repaired as non-crossover.

1.3.1 Yeast recombination hotspots

Yeast hotspots display a several kb wide distribution of crossover and non-exchange conversion, and exchanges are not always simple, some exchange events are complex, containing both crossover and gene conversion events, interrupted by unconverted markers (Borts and Haber 1989). Hotspots are significantly associated with regions of high G+C content and certain transcriptional profiles (Gerton *et al.* 2000). High resolution mapping of crossovers and non-exchange gene conversions in yeast has revealed that there are regions favouring crossover or non-crossovers, and that interference exists between both outcomes of meiosis, not just crossover (Mancera *et al.* 2008).

Hotspots can be classified depending on their activation or initiation signal. Nuclease sensitive hotspots that require transcription factors are called alpha-hotspots (Petes 2001). In contrast ß-hotspots are not dependent on transcription factors; instead CCGNN tracts stimulate recombination in *cis*, perhaps by producing an open chromatin domain as they exclude nucleosomes. Alpha and beta hotspots are mutually exclusive, but can simultaneously belong to the third subclass, gamma hotspots, which are located in regions of high GC content (Gerton *et al.* 2000).

1.3.2 Mammalian Recombination hotspots

The resolution sites of crossovers in mammals are spread over 500 -2000 bp and form what appears to be a normal distribution around a central point, defining the centre of the hotspot. This distribution of crossover resolution points probably reflects the outward migration of Holliday junction intermediates from a DSB site. A high number of DNA DSBs are formed at every meiosis, but it is estimated that only 10% of these breaks are repaired via an inter-homologue crossover in mammals (Baudat and de Massy 2007). Crossover frequencies at individual hotspots can vary substantially between hotspots and hotspots are not randomly distributed along chromosomes. They have also been found as doublets or clusters of several hotspots, for example in the MHC. The position of hotspots is sexually dimorphic in humans (Kong *et al.* 2002) and mice (Shifman *et al.* 2006). And the usage of individual hotspots is likely sex specific in humans (Kong *et al.* 2010) and mouse (Shiroishi *et al.* 1991).

1.4 Human meiotic recombination

Analysing meiotic recombination in humans poses unique challenges, as the classical geneticists' toolkit, from genetic manipulation to deliberate crossing of individuals, is not available to the human geneticist. Nevertheless there are powerful tools available to address meiotic recombination both genome-wide and in individuals. On the one hand genotype data can be used to estimate the position and intensity of recombination from a historical perspective by analysing haplotype diversity in a population. Secondly, fine-scale genetic experiments can directly measure recombination frequencies in individuals.

1.4.1 Population data

The amount of data available has increases in leaps and bounds in the last decade, with high-density SNP maps and large-scale population data now publicly available. In recent years, high resolution SNP maps were created from sequencing and alignment data of several human populations, generated by the International HapMap Project (phase 1 (The International HapMap Consortium 2005), phase 2 (The International HapMap Consortium 2007) and phase 3 (The International HapMap Consortium 2010) and the 1000 Genomes Project (1000 Genomes Project Consortium 2010). This large-scale human population data can be addressed statistically through linkage disequilibrium and coalescent analysis to infer historical recombination patterns.

1.4.1.1 Using the concept of Linkage and Linkage Disequilibrium

Two loci, which are not linked, segregate apart from each other with a probability of 50%, as described by Mendel's law of independent assortment. In contrast, loci that are inherited together much more often than this probability of 50% can be considered linked. If alleles are associated it means they are found on the same haplotype more often than expected by chance; this is described as Linkage Disequilibrium (LD). If alleles are strongly associated only very little, if any, historical recombination can have occurred between those two loci in the population being studied. The opposite is the case where all markers are in free association, which we call breakdown of linkage disequilibrium. The statistical analysis of the degree of LD is a powerful tool to infer historical recombination. But there are other factors influencing LD, like mutation and evolutionary processes such as admixture, genetic drift, population demography and selection.

1.4.1.1.1 Metric Linkage Disequilibrium maps

Locations of recombination hotspots can be inferred from LD when recombination events have accumulated in a population over many generations. The rate of LD-breakdown can be displayed in metric LD maps by cumulating LD-units (LDU) along a chromosome (Maniatis *et al.* 2002). Regions of strong marker association appear as plateaus while regions of LD breakdown appear as increments, as the lesser the marker association between two loci is, the more LD-units are cumulated, and the bigger the LDU step on the plot. Thus, a site of strong historical recombination (an LD hotspot) will appear as large increment in LDU, flanked by haplotype blocks that appear as plateaus. LDU steps are in good agreement with sites of meiotic recombination determined by sperm-typing (Zhang *et al.* 2002).

1.4.1.1.2 Coalescent analysis

Coalescent-based statistical analysis of LD data allows inferring the probabilities that haplotype boundaries represent historical hotspots. Haplotypes present in a population are taken to construct random genealogies, determining the ancestral state of these haplotypes. Individual lineages are associated to ancestor haplotypes at random as the model goes back in time, and whenever two lineages end up in the same ancestor, they coalesce. The model then implies models of evolutionary change, incorporating factors such as population bottlenecks, mutation, drift and in particular recombination onto these coales-

cent haplotypes, trying to recreate the haplotypes seen in the population today (Kingman 2000). Statistical analyses based on coalescent approaches are powerful in providing broad-scale, high-resolution recombination maps. However, they only describe sex-averaged, historical recombination within genetically heterogeneous populations. Historical crossover rates estimated by coalescent analysis correlate well with LD maps (McVean *et al.* 2004).

1.4.1.2 Genome wide historical recombination landscape

The International HapMap Project defined haplotype structure across the human genome on extensive genotype data of four human populations (The International HapMap Consortium 2005; The International HapMap Consortium 2007). These were initially generated with the intention to aid gene association studies, but also identified approximately 33,000 LD-hotspots, on average about one in every 100 kb. Some of the hotspots found by LD, have been validated at the individual level by sperm typing (as described in 1.4.3). The Hap-Map recombination rates are sex-averaged on autosomal chromosomes, whereas female specific recombination was inferred from X-chromosomal data. In a study by Myers *et al.* (2005) estimations of the recombination rates for three different human populations, based on coalescent analysis of HapMap SNP genotyping data were used to define a fine-scale genetic map of the human genome (Myers 2005). Essentially about 80% of historical recombination events were found to accumulate into 10-20% of the sequence (Myers 2005).

Human population studies have led to the observation that adjacent SNPs tend to form stretches of DNA with strong marker association, or haplotype blocks, that are 10-100 kb long. Within these blocks, most or all markers are in linkage disequilibrium, and the block boundaries are correlated among populations (Gabriel *et al.* 2002). Analysing SNP variation data from human populations by fitting a statistical model, based on the coalescent, to patterns of linkage disequilibrium, can identify LD recombination hotspots (Myers 2005). But as LD hotspots are inferred from genome wide, sex-averaged data, they cannot give information about sex differences or differences between individuals. Statistical analyses of genetic variation data can enable us to infer rates of historical recombination. Historical recombination maps of human chromosomes are measured in centi-Morgans and allow the genome-wide identification of LD hotspots.

1.4.2 Pedigree studies

Pedigree studies can give more information about individual and sex-differences in recombination than population data can, and avoid biases created by demographic history. Pedigree studies are particularly useful in large families and extended pedigrees, as they allow the tracing of a single allele, typically across at least three generations. Using molecular markers, segregation analysis is used to directly measure the recombination frequency between adjacent markers.

1.4.2.1 Segregation analysis in pedigrees

Earlier linkage maps, based on simple tandem repeat (STR) markers with resolution limited to 0.1 – 1 cM, revealed that meiotic recombination clusters into localised sites of recombination, for example in the MHC (Cullen *et al.* 1997). Due to the low resolution, these maps only had the power to identify megabase-scale regions with particularly high or low recombination rates, which were termed recombination "jungles" or "deserts" respectively (Yu *et al.* 2001). Later Kong *et al.* demonstrated that recombination rates are on average higher in females than in males, and that differences between females were common (Kong *et al.* 2002). But even large families have only a limited number of meioses to be followed, and the marker density is usually small.

1.4.2.2 High-resolution association studies in multi-generational pedigrees

Data from large multigenerational pedigrees can provide information of the distribution of recombination. Pedigree analyses have increased in resolution through the arrival of high-density single nucleotide polymorphism (SNP) typing techniques. A high-resolution genome wide linkage study of 728 human meioses has found that only 60% of mapped crossover events coincided with LD hotspots (Coop *et al.* 2008). A higher-density genome wide SNP data set from 15,257 parent-offspring pairs was used to generate a map with SNP markers regularly spaced at 10 kilo-base (kb) intervals, revealing that about 15% of hotspots in one sex are specific to that sex (Kong *et al.* 2010).

1.4.3 High-resolution studies in sperm

Similar to tetrad analysis in yeast, the highest resolution of recombination in a single individual can be achieved by direct analysis of gametes produced by meiosis. The resolution level is only limited by the number of informative markers present, and is often as high as one marker every 100 bp. The only obvious drawback of this method is that it can only address male recombination, given the limited availability of oocytes.

1.4.3.1 Single sperm studies

Single sperm studies require isolation of individual sperm cells and the amplification of their whole haploid genome using degenerate PCR primers (Li *et al.* 1988). Studies on crossover involving single sperm DNA molecules have produced a crossover map of the major histocompatibility complex (MHC) (Cullen *et al.* 2002). Sorted single sperm derived from a heterozygous individual were analysed for non-exchange conversion events by direct sequencing of exon 2 of the HLA-DPB1 in the MHC, which detected one potential non-exchange conversion event per 10,000 sperm (Zangenberg *et al.* 1995). Single sperm studies mostly type few hundreds of sperm and are restricted by the number of informative markers in the analysed region.

1.4.3.2 Batch sperm studies

Batch sperm typing can screen pools of large numbers of sperm DNA molecules for recombination-events. This approach is based on choosing a study interval, typically containing a region of LD-breakdown, followed by selective amplification of crossover molecules by repulsion-phase allele-specific PCR (Jeffreys *et al.* 1998a). With this method thousands of sperm molecules are analysed at once, and with a high resolution that is only limited by the number of markers present in the analysed region, as reviewed in (Kauppi *et al.* 2009). Batch sperm typing not only allowed the characterisation of general properties of human meiotic crossover hotspots but can also reveal polymorphisms in hotspot activity between individuals. Individual sperm crossover hotspots can be identified even in regions of strong marker association, which are undetectable by LD analysis (Jeffreys *et al.* 2005).

1.4.4 Human recombination hotspots

Studies in sperm have established that recombination events in humans cluster into narrow recombination hotspots 1-2 kb wide and spaced, on average, 50 kb apart. The recombination frequency seen at regular hotspots was typically in the region of 0.05%, ranging from 0.001-0.1%. At some locations, hotspots tend to cluster, for example in the class II region of the major histocompatibility complex (Jeffreys *et al.* 2001). Several examples of hotspot doublets have been found in other regions of the genome (Jeffreys *et al.* 2005; Webb *et al.* 2008).

In almost all crossover hotspots analysed to date crossovers were simple, and reciprocal crossover products arose at the same rate. Plotting exchange points across the analysed region displayed a symmetrical quasi-normal distribution. The uniform width across all hotspots clearly pointed to common processes operation within these hotspots. Despite the constant width, the intensity of recombination varies by orders of magnitude over different hotspots and also showed significant variation in crossover frequencies between men as reviewed in (Jeffreys *et al.* 2004; Kauppi *et al.* 2005; Holloway *et al.* 2006).

1.4.4.1 Recombination Superhotspots

The term "Superhotspot" is an arbitrary definition describing a sub-class of hotspots with crossover frequencies above 0.1%, first identified by Webb and colleagues. Candidate regions were identified from metric LD maps, showing a rapid and marked breakdown in LD flanked by LD-blocks, which implied high recombination frequencies. Coalescent analysis also indicated a high rate of historical recombination at the same location. The frequency and distribution of crossover breakpoints within these regions has been mapped in sperm, and verified Superhotspots for crossover (Webb *et al.* 2008; Berg *et al.* 2010), (Jeffreys and Neumann 2009).

Later additional Superhotspots were identified from SNP data, collected from nuclear families by Coop *et al.* (Coop *et al.* 2008), at intervals where familial crossovers clustered within minimal regions of overlap (Jeffreys and Neumann 2009; Berg *et al.* 2010) (Jeffreys, *unpublished data*). Direct sperm typing performed on seven of these candidate intervals, which had also shown high levels of historical LD in HapMap, confirmed Superhotspots with crossover frequencies as high as 3% in some men (Jeffreys and

Neumann 2009; Berg *et al.* 2010) (Jeffreys, *unpublished data*). The previously defined properties of hotspots, like their uniform width and quasi-normal distribution are also true for recombination Superhotspots, but crossover frequencies appear to show less variation between men (Webb *et al.* 2008).

1.5 Gene conversion

In the DSBR model, both non-crossover and crossover are believed to result from the same heteroduplex-containing recombination intermediate, generated from an initiating DSB. Resolution of recombination events in the non-crossover configuration can lead to gene conversion, 3:1 segregation of markers. But gene conversion is not limited to non-crossovers and can equally accompany crossovers. Gene conversion of a marker can also occur via gap repair, when a 3′-end has been degraded past the site of a marker near the DSB at the resection step, as reviewed in (Stahl and Foss 2010) or by mismatch repair of heteroduplex DNA.

1.5.1 Gene conversion in yeast

Traditionally gene conversion was often used as a marker for hotspot location, as 3:1 segregation of phenotypic markers in tetrads is sufficient to pinpoint recombination, even in the absence of additional flanking markers. The yeast *HIS4* and *ARG4* hotspots were initially characterised as gene conversion hotspots in this way. In these hotspots, gene conversion displays polarity, with high frequencies at one end of a gene declining towards the other end of the gene (Detloff *et al.* 1992; Nicolas and Petes 1994). Gene conversion is frequent in yeast, varying from 0.5% to more than 70% of tetrads containing a gene conversion event at a given locus. Genome-wide up to 1% of the genome is changed through conversion per meiosis, and gene conversion is often biased towards generating GC nucleotides (Mancera *et al.* 2008). Simple gene conversion tracts accompany around 90% of crossover molecules; the remaining 10% are associated with complex, discontinuous, conversion tract exchanges. The median non-exchange conversion tract is 1.8 kb and the median gene conversion tracts accompanying crossovers is 2 kb. The majority of gene conversions are believed to be produced by mismatch repair of heteroduplex DNA in yeast (Alani *et al.* 1994).

1.5.2 Gene conversion drives human genome instability and evolution

Gene conversion, like crossover, generates new combinations of alleles, which increases the genetic diversity of the human genome. Gene conversion leads to non-mendelian segregation of alleles in gametes. If copying from one haplotype to the other is equally likely in reciprocal orientations, this does not have any consequences on the population level. Overall the gamete pool would be unbiased. Biased transmission of one allele over the other has been observed at several recombination hotspots, and in some instances at least, it was caused by biased initiation frequency on one homologue. Initiation on an active haplotype leads to repair from the homologue and thus the transfer of markers is directional, from the uncut homologue to the cut, or active, homologue. In these cases there is an evolutionary advantage of the donor allele over the acceptor allele. The availability of large-scale human population genetic variation data and sperm typing have facilitated the progress towards understanding the mechanism of non-exchange conversion and its impact on the human genome. But the understanding of the mechanisms and regulating factors of meiotic gene conversion in humans is still very limited, and much work still needs to be done.

1.5.2.1 Gene conversion and LD

The impact that gene conversion has on the finer scale LD can have detrimental effects when misinterpreted in association studies. Fine-scale patterns of LD are shaped by crossover and by gene conversion (Ardlie *et al.* 2001; Ptak *et al.* 2004), but gene conversion cannot easily be inferred from LD patterns, though attempts have been made (Gay *et al.* 2007). As gene conversion tracts are short, only one or few markers are likely to be involved and the impact on LD will be very localised. When flanking markers are in LD, converted markers will not show LD with the flanking markers. With low marker density it is impossible to determine whether a base-substitution inferred from population data is the result of a gene conversion or a mutation event. In cases where high sequence similarity between one locus and a potential donor locus is present, substitution of one or two markers in a small tract is likely to be the result of a gene conversion event, rather than a single base substitution or a double-CO. Some pathogenic single base substitutions identified in the human genome to date might just as well have arisen by gene conversion. Incorporating gene conversion as well as crossover in LD studies significantly increases the fit of population data (Frisse *et al.* 2001).

However, accurate and reliable conversion to crossover ratios are not known, apart from a few small regions of the genome that have been subjected to sperm recombination analysis as summarised in (Holloway *et al.* 2006).

1.5.2.2 GC biased gene conversion

Gene conversion in the human genome appears to be biased towards generating more Guanine and Cytosine residues than Adenine and Thymine (Duret and Arndt 2008), gaining GC alleles a population advantage in highly recombining regions. This is known as GC biased gene conversion. Human "accelerated regions", containing functional elements that have significantly changed during human evolution, show GC bias (Katzman *et al.* 2010). Biased gene conversion favouring GC alleles appears to have had a major influence on the evolution of GC content in mammals, and may be responsible for the evolution of the isochores (Duret and Galtier 2009). In some cases GC-biased gene conversion can overcome purifying selection, and promote the fixation of deleterious amino acid changes in primates (Galtier and Duret 2007; Galtier *et al.* 2009). It is not yet fully understood whether it arises through biased recombination initiation on A or T alleles, or whether it results from biased mismatch repair, promoting GC alleles over AT-alleles.

1.5.2.3 Ectopic gene conversion and the evolution of gene families

It is important to allow genetic diversity to increase while equally ensuring relative genome stability. Gene conversion can influence both of these processes to a great degree. In contrast to allelic gene conversion, ectopic (or non-allelic) gene conversion occurs between sequences that are highly homologous, but not allelic. Ectopic gene conversion has been identified in several regions of high sequence homology, including genes in gene clusters, a gene and a pseudogene and palindromic arms on the Y chromosome.

1.5.2.3.1 Divergence of duplicated genes

Genomes expand in part through gene duplication events via unequal crossover. Duplicated genes then evolve through mutation in one of the gene copies that either render it non-functional or improve gene function.

Ectopic gene conversion can generate allelic diversity within highly similar genes within gene families. Several regions of the human genome have to remain highly polymorphic, to allow rapid evolution and adaptation to changing environments. One example is the HLA class II region of the Major Histocompatibility Complex (MHC). The highest degree of polymorphism is found in exon 2 of the *HLA-DRB1*, which encodes part of the antigen-recognition site of class II proteins (International Immunogenetics Project *HLA* database (Robinson *et al.* 2009)). Novel alleles in the HLA locus are likely to be continually generated by gene conversion (von Salomé and Kukkonen 2008). Evidence of inter-allelic gene conversion has been identified in the *HLA-DPB1* locus by single-sperm analysis (Zangenberg *et al.* 1995). Additional examples of ectopic gene conversion in the human genome fostering functional divergence come from the ß-globin gene family (Aguileta *et al.* 2004), immunoglobulin genes (Baltimore 1981; Darlow and Stott 2006), as well as the human rhodopsins (red and green pigments) (Kuma *et al.* 1988; Zhao *et al.* 1998) and ABO blood groups (Ogasawara *et al.* 2001).

1.5.2.3.2 Concerted evolution of duplicated genes

Not all gene copies necessarily diverge and, interestingly, gene conversion can also promote concerted evolution of genes within a gene family. The phenomenon of concerted evolution in duplicated genes is facilitated by the high sequence homology present between the gene copies. In some multi-gene families it has often been observed that paralogues in one species are much more closely related to each other, than they are to their respective orthologue in a closely related species. Against the principle of parsimony, it may appear that independent duplication events in each of the lineages have occurred. But, in fact, one gene duplication event in a common ancestor led to the duplication, but then differences in the paralogous sequences were repeatedly smoothed out by ectopic gene conversion. Examples of concerted evolution through biased gene conversion are found in the human Y chromosomal palindromes (Rozen *et al.* 2003) and between hotspots on the Y chromosome in humans (Hurles *et al.* 2004).

An example for maintenance of high homology between two genes comes from the human *CYP21A/ CYP21B* genes, located in the *HLA-MHC* on Chromosome 6. This region is a tandem-duplication, consisting of *CYP21A2* – a gene encoding the human steroid 21-hydroxylase, and the pseudogene *CYP21A1P* as well as their associated complement

factor genes *C4A* and *C4B*. High sequence homology between the *CYP21A2* gene and the pseudogene *CYP21A1P* leads to deletion and addition haplotypes of *CYP21+C4* units by unequal crossing over (Collier *et al.* 1989; White *et al.* 1994). Large-scale conversions proposed to involve the flanking *TNXB* gene have been discredited (Koppens *et al.* 2002). But small-scale gene conversion derived from the pseudogene contribute to more than 95% of mutations leading to CYP21A2 defects that are responsible for congenital adrenal hyperplasia (White *et al.* 1994; Tusié-Luna and White 1995).

1.5.2.4 Minisatellite instability

Minisatellites are highly unstable tandem repeated DNA loci in the human genome. Repeat instability is generated frequently, and gene conversion processes play a major role in driving this instability (Jeffreys *et al.* 1997), re-arrangements appear as gene-conversion-like transfers of blocks of repeats from one allele into the other, without exchange of flanking markers. A highly complex patchwork of DNA can be generated from rearranging various regions of a donor allele (Jeffreys *et al.* 1994; Jeffreys *et al.* 1995; Jeffreys and Neumann 1997; May *et al.* 2000). Conversions at some minisatellites are heavily polarised, being restricted to only one end of the repeat array (Armour *et al.* 1993; Jeffreys *et al.* 1995; May *et al.* 1996).

The first human meiotic crossover hotspot identified on the molecular level by sperm typing was located upstream of minisatellite MS32 (Jeffreys *et al.* 1998a), and the repeat instability at MS32 appears to have evolved as by-product of this localised recombination activity (Jeffreys *et al.* 1999).

1.5.2.5 Gene conversion is frequent within human recombination hotspots

Gene conversion within the very few human recombination hotspots analysed to date by sperm typing, displays a steep bidirectional gradient, clustering at the centre of crossover activity potentially marking the site of recombination initiating DSBs (Jeffreys and May 2004). Gene conversion tracts tend to be very short, typically involving less than 300 bp and sometimes as little as tens of base pairs (Jeffreys and May 2004; Jeffreys and Neumann 2005). Conversion tract detection thus crucially required heterozygous SNPs at the centre of the hotspot.

At the beginning of this work, four hotspots had been analysed for crossover and gene conversion. At the ß-globin hotspot, no gene conversion molecules could be detected from 12,000 tested sperm DNA molecules (Holloway *et al.* 2006), for the other hotspots variable frequencies of gene conversion were detected, as summarised in Table 1-1.

Table 1-1 highly variable rates of gene conversion to crossover between previously characterised hotspots

Hotspot name	*DNA3*	*DMB2*	*NID*1	*SHOX*
Non-exchange conversion frequency (%)	0.03	0.0004	0.0013	0.009
Non-exchange conversion to crossover ratio	2.7:1	1:1.3	1:4	1:3.3

Conversion frequencies and ratios of conversion to crossover are shown for the hotspots *SHOX, DNA*3 and *DMB*2 (Jeffreys and May 2004) as well as *NID*1 (Jeffreys and Neumann 2005)

1.6 Rapid evolution of the recombination landscape in the human genome

1.6.1 Inter-species comparison of fine-scale recombination rates

In studies comparing the LD-structure of the human genome with that of our nearest relative, the chimpanzee, only poor conservation of fine-scale recombination rates was observed (Ptak *et al.* 2005; Winckler 2005), suggesting a rapid turnover of recombination hotspots, despite the relatively small sequence divergence of 1.24% between humans and chimpanzees (Chen and Li 2001).

1.6.2 Hotspot are attenuated through biased gene conversion

If DSBs were introduced at equal rates on both chromosomes, and then processed and repaired by repair, one would expect haplotypes of both chromosomes to appear in crossover molecules at equal rates (Schwacha and Kleckner 1995), provided that no selection is acting on either allele. The initiation of recombination by DSBs on one chromosome leads to the copying of allelic sequence from the other chromosome. A higher frequency of initiation on one chromosome therefore leads to alleles from the less-initiating chromosome becoming over-transmitted into recombinant progeny. Biased initiation and subsequent gap repair also leads to an asymmetric distribution in gene conversion accompanying crossover, in a way that exchange points are not reciprocal in

location. This phenomenon is known as reciprocal crossover asymmetry. Hotspots can become attenuated through creating a meiotic drive against themselves, converting activating into suppressing alleles.

Two recombination hotspots, *NID*1 (nidogen 1) (Jeffreys and Neumann 2005) as well as *DNA*2 in the MHC (Jeffreys and Neumann 2002) provide examples of this meiotic drive in the human genome. As observed for the hotspots in *NID*1 and *DNA*2, increased initiation on one chromosome through the active allele leads to over-transmission of the suppressed allele into recombinant sperm. This meiotic drive at hotspots *NID*1 and *DNA*2, favouring the SNP allele suppressing recombination, is strong enough to lead to attenuation of the hotspot over time. Asymmetry also seems to affect non-exchange conversion as found in the *NID*1 hotspot, further increasing the meiotic drive (Jeffreys and Neumann 2005).

The fact that suppressing alleles can drive attenuation of a hotspot may provide an explanation for why hotspots are ephemeral features that are not conserved between human and chimpanzee. The problem as to how a hotspot can persist, despite creating a meiotic drive against itself through biased gene conversion, was referred to as the "recombination hotspot paradox" (Boulton *et al.* 1997; Pineda-Krch and Redfield 2005).

1.6.3 Generation of new hotspots

Direct evidence for the appearance of human recombination hotspots has been identified at two hotspots, hotspot S (Jeffreys and Neumann 2009) and the MSTM1a hotspot (Neumann and Jeffreys 2006). Both hotspots were located in regions of strong marker association, as they had not yet left a mark on LD, and therefore appear to be young. Hotspot S is activated in *cis* through a single-base change, defining a haplotype estimated to have appeared around 70,000 years ago. Modelling of evolution showed that the observed population frequency of the founding mutation can arise through drift alone, despite being constantly removed through biased gene conversion. It also predicted that the hotspot is likely to go extinct with a median hotspot lifespan of 120,000 years. This example has provided one resolution for the hotspot paradox, showing that drift alone can create and extinguish hotspots with high recombination frequencies despite the major impact of meiotic drive (Jeffreys and Neumann 2009).

MSTM1a lies in proximity to hotspot MSTM1b, and is only active in crossover in three out of the 26 men tested for recombination at both hotspots. MSTM1b is active in all 26 men tested here, but displays up to 75-fold differences in crossover frequencies between men. In contrast, at MSTM1a the men active for crossover show similar frequencies, but display highly significant reciprocal crossover asymmetry. MSTM1a is active in crossover in men sharing the same haplotypes with men where MSTM1a is inactive, which ruled out a *cis*-activating factor for recombination at this hotspot (Neumann and Jeffreys 2006).

1.7 Regulation and control of Recombination

The observation that human meiotic recombination is not randomly distributed but instead clusters into recombination hotspots, indicates that recombination initiation must be highly regulated. Regulation by an element located at the same locus and on the same molecule of DNA is defined as *cis*-regulation, while regulation by an element located elsewhere defined as *trans*-regulation. Studies of recombination hotspots in this laboratory has shown differences in both, presence or absence, of given hotspots, as well as differences in the recombination frequency displayed by men at the same hotspot.

In two instances a *cis*-regulating SNP within the hotspot was found to influence hotspot heat. In another instance presence/absence polymorphisms were independent of local DNA sequence variation. Men with identical haplotypes showed completely different recombination patterns, which points to *trans*-regulation of hotspot location and activity.

1.7.1 Regulation in *cis*

Ever since the discovery of the chi-sequence in bacteria, a specific series of eight bases that stimulates recombination in the bacterial chromosome (Smith *et al.* 1984; Rosenberg and Hastings 1991; Myers and Stahl 1994), the question remained whether recombination initiation sequences exist for the human genome as well. This is true for at least some regions of the human genome, especially the somatic VJ-recombination, which uses specific sequences to join a variable gene segment and a joining segment to form an immunoglobulin gene. Several local predisposing factors have been described that have the potential to generate a target for the DSB machinery, including DNA sequence motifs and higher order chromatin structure.

1.7.1.1 Primary DNA sequence motifs

In concordance with a specific initiation sequence, sequence motifs have been found that are enriched within LD recombination hotspots in the human genome. Long terminal repeats of retrovirus-like retrotransposons and CT or GA rich repeats were over-represented, while TA and GC rich repeats were underrepresented in LD-hotspots. A core sequence motif (CCTCCCT) was identified, which has been associated with about 11% of all putative hotspots (Myers 2005). A SNP disrupting this motif was both necessary and sufficient to suppress recombination at hotspot *DNA*2 (Jeffreys and Neumann 2002). A degenerate form, the 13-mer CCNCCNTNNCCNC motif was found associated with 40% of LD hotspots (Myers 2005) and with 37% of hotspots analysed through sperm typing of individuals (Webb *et al.* 2008). Recently, an additional fully conserved 11-mer GGNGGNAGGGG motif was identified by comparing the association of recombination hotspots with genetic polymorphism (Zheng *et al.* 2010). The degenerate 13-mer motif of Myers *et al.* is partially complementary to the motif identified by Zheng *et al.* (2010).

1.7.1.2 Higher order chromatin structure

DNA is condensed into chromosomes, wound around histone complexes, creating nucleosomes. Nucleosome occupancy is generally stable during meiotic progression and crossover refractory zones within hotspots coincide with the location of nucleosomes in mouse (Getun *et al.* 2010).

It is easy to imagine that protein complexes involved in recombination would require an opening in the chromatin structure to access DNA. During transcription the chromatin has to be opened up to be accessible for transcription machinery. This is achieved through post-translational histone modifications, especially acetylation of histones. And indeed, post-translational histone modifications, such as acetylation, ubiquitination and methylation have also been implicated in the placement of recombination in yeast (Yamada *et al.* 2004; Yamashita *et al.* 2004; Mieczkowski *et al.* 2007). Especially histone-3 lysine 4-methylation histone-3 lysine-4 (H3K4) methylation stands out as being quantitatively correlated with the number of DSBs in yeast (Borde *et al.* 2009).

Methylation also seems to play an important role in mammalian recombination. High levels of H3K4 tri-methylation at mouse hotspots Psbm9 and Hlx1 correlate with high levels of recombination (Buard *et al.* 2009; Grey *et al.* 2009). And in the human Epigenome project dataset, germ-line methylation positively correlates with regional recombination rates, and is also increased in locations of known sperm recombination hotspots (Sigurdsson *et al.* 2009).

1.7.2 Regulation in *trans*

Studies on hotspot dynamics by Neumann and Jeffreys have identified that activity polymorphisms between men at a given hotspot could exist independently of local DNA sequence variation at hotspots MSTM1a and 1b (Neumann and Jeffreys 2006) pointing towards *trans*-activating factors. In a mouse recombination Superhotspot this had been observed as well (Shiroishi *et al.* 1991).

1.7.2.1 Genome wide recombination regulators

In humans an inversion polymorphism on chromosome 17q21.31 had been implicated to influence recombination, with females carrying a European specific haplotype having more children and higher recombination rates than non-carriers (Stefansson *et al.* 2005). By treating recombination rate as a quantitative phenotype and carrying out genome wide association studies on a high number of pedigreed individuals it was possible to detect association of markers with increased or decreased recombination rates. Analyses in the Icelandic population revealed that sequence variants in the *RNF212* gene were correlated with variation in recombination rates in males and females. Interestingly two single-nucleotide polymorphisms located within *RNF212*, are associated with opposite effects on males and females (Kong *et al.* 2008). A similar study on Caucasian samples identified three male-specific and three female-specific recombination regulators, one of which confirmed the *RNF212* as a genome-wide recombination regulator for males (Chowdhury *et al.* 2009).

1.7.2.1.1 *PRDM9* (Meisetz)

In mice, genome-wide recombination regulators have been proposed from using crosses between different mouse strains and observing varying recombination hotspot usage patterns either chromosome wide or at specific hotspots (Grey *et al.* 2009; Parvanov *et al.*

2009). In both studies, the overlapping candidate regions contained *Prdm9*, which stood out as a strong candidate gene for recombination regulation. *Prdm9* is expressed during early prophase in mouse males and females. A mouse knockout for *Prdm9* is blocked at the pachytene stage of meiosis I, and both sexes are sterile with males displaying azoospermia (Hayashi *et al.* 2005).

Prdm9 has three functional domains (Parvanov *et al.* 2010). Firstly an N-terminal KRAB domain can promote protein:protein interactions, and transcriptional repression, when bound to DNA by a tethered DNA binding domain. Secondly a central PR/SET domain (Meisetz) that provides histone methyl transferase activity, tri-methylating H3K4 (Hayashi *et al.* 2005) thus altering chromatin structure. Thirdly, a DNA binding domain is provided through the terminal C2H2 zinc-finger domain. The DNA binding properties of each zinc-finger are predicted to rely mainly on positions -1, 3 and 6 of the zinc-finger alpha-helix (Baudat *et al.* 2010). Amino-acids at these positions are likely to be involved in sequence specificity, and alterations that lead to coding of different amino acids should affect binding properties of a given zinc-finger (Baudat *et al.* 2010). Human reference sequence *PRDM9* has 13 zinc-fingers coded for by a minisatellite.

PRDM9 is the most diverged C2H2 containing zinc-finger protein between the human and chimpanzee lineages, these species sharing the DNA binding positions -1, 2, 3 and 6 on only the first zinc-finger (McVean and Myers 2010). DNA binding predictions of human *PRDM9* are consistent with five zinc-fingers binding the human 13-mer degenerate hotspot motif that also displays 3-bp periodicity (Myers *et al.* 2008). In-silico binding-prediction to the hotspot motif is consistent with motif degeneracy at positions 3, 6, 8, 9 and 12 and a lack of degeneracy on the other 8 positions, in addition there is an exact match on the opposite strand to an 8 bp region of the extended motif (Myers *et al.* 2010).

Comparison of the sequences of human *PRDM9* between individuals of the Centre d´Etude du Polymorphisme Humain (CEPH) panel revealed a large number of variant PRDM9 zinc finger alleles (Baudat *et al.* 2010). Zinc finger allele A, the most common allele in CEPH individuals with a population frequency of 90% was highly similar to the genome reference allele B (with 5% frequency in CEPH). Six additional variants were found, with the first five zinc fingers showing little variation to allele A and variability concentrated on the C-terminal zinc-fingers 8-11 (Baudat *et al.* 2010).

In a Hutterite population sample, only three *PRDM9* alleles were found, allele A and B and an additional allele I, not present in CEPH. Allele I was correlated with a shift in LD hotspot usage as AI individuals had a significantly lower likelihood of using predicted LD-hotspots than AA individuals, as shown by linkage analysis. Allele I is thus likely to shift recombination to other hotspots that have not left a mark on the LD landscape (Baudat *et al.* 2010). The A-variant of PRDM9 was found to have DNA binding properties, with a high affinity *in vitro* to DNA containing the 13-mer motif (Baudat *et al.* 2010). Conversely the I-variant had a high affinity to DNA containing a predicted binding motif for the I-allele, but did not bind the A-motif (Baudat *et al.* 2010).

Variation at the *PRDM9* locus has a strong effect on sperm hotspot activity in humans (Berg *et al.* 2010; Berg *et al.* 2011; Hinch *et al.* 2011). Berg *et al.* (2010) identified twenty-four additional PRDM9 variants, with a higher diversity of PRDM9 zinc-finger variants in men of African and Afro-Caribbean origin. Changes within the zinc-finger array can create variants that activate recombination at new hotspots, for example three men active at hotspot MSTM1a all carry a distinct PRDM9 variant responsible for initiating recombination at this hotspot (Berg *et al.* 2010). Even subtle changes between variants can create non-activating alleles that lead to suppression of recombination at a given hotspot (Berg *et al.* 2010).

PRDM9 variation not only influences hotspot activity, but also has a major influence on genome stability. Minisatellite instability is strongly influenced; likewise megabase-scale rearrangements at the CMT1A locus responsible for Charcot-Marie Tooth Disease type 1 A and Hereditary neuropathy with pressure palsies, are much more frequent in men with *PRDM9* A and B alleles, while non-A alleles show suppressed frequencies of rearrangements (Berg *et al.* 2010).

A second class of hotspots activated by the C-variant of PRDM9 were found to be enriched in African populations (Berg *et al.* 2011). Comparing linkage maps of African-Americans with those of European descent also identified significant fine-scale differences in the position of inferred recombination hotspots, and the use of different hotspots between populations is fully attributable to the PRDM9 variants that individuals carry (Hinch *et al.* 2011).

1.7.3 Disabling the "hotspot paradox"

The hotspot paradox now seems to have been fully resolved. Hotspots do not have to persist in the face of their destruction, but instead new locations are activated as recombination hotspots. This is achieved by generation of new PRDM9 zinc-finger arrays by rapid evolution of the *PRDM9* minisatellite (Berg *et al.* 2010). But, as cis-regulating factors had been described to strongly influence recombination activity within a given hotspot, there could also be secondary *cis*-acting factors that predispose chromosomal locations to become hotspots of recombination.

1.8 Aims and Objectives – Overview of this thesis

DNA within a chromosome is re-shuffled during meiosis through recombination events. Interactions between homologous allelic sequences result in conventional crossovers or alternatively in non-exchange conversion in which short patches of DNA are copied from one chromosome and pasted into the homologous chromosome, replacing allelic sequences. Recombination clusters into distinct recombination-active regions, that we call hotspots. Previously, hotspot analysis had focused on the description of crossover profiles because analysing non-exchange conversion efficiently without enrichment had not been possible. At the beginning of this work very little was known about the frequency and distribution of non-exchange conversion events across the human genome, and how well they correlate with crossover.

Equally, our knowledge about regulating factors of the human recombination machinery was very limited. Biased gene conversion had been found responsible for over-transmitting suppressing alleles, and therefore implicated in hotspot silencing. But we did not know how frequently biased gene conversion occurs, and whether the meiotic drive that it generates is equally strong when comparing different hotspots. Finally, it remained a paradox as to how hotspots persist in the human genome, if they are continuously eliminated by biased gene conversion? This work aims to gain insights into these very interesting questions.

Chapter 3 investigates gene conversion processes operating at human recombination hotspots, and the relationship between crossover and non-exchange conversions. Five crossover Superhotspots were chosen for analysis from published sources and then both types of recombinants analysed simultaneously in pools of sperm DNA in men carrying suitable SNP heterozygosities. Crossover frequencies and distribution were compared with previously established parameters and the frequency and distribution of non-exchange conversion determined. This Chapter will show for the first time whether conversion to crossover rates are variable between men and therefore give basic information on to whether crossovers and conversions are co-regulated or independently regulated.

Chapter 4 aims to characterise biased gene conversion by comparing recombination frequencies and distribution in both reciprocal orientations, allowing for the detection and quantification of disparity in reciprocal conversion rates. As for the first time a high number of non-exchange conversion molecules were detected, it was tested whether crossover and non-crossover are influenced by the same biases. Incidences of biased gene conversion within hotspots were then explored further, extending the analysis to more men to allow the identification of local DNA sequence variants that control the bias.

Chapter 5 investigates incidences of particularly low recombination frequencies, identified by comparing conversion and crossover rates between men at a given hotspot in Chapter 3. Two candidate factors were tested for the hotspot showing the most extreme variation in crossover frequencies between men.

Chapter 6 investigates a recombination Superhotspot, which is more active in African populations. This Superhotspot had shown marked reciprocal crossover asymmetry, indicative of biased gene conversion associated with crossover. Curiously the over-transmitted putative recombination suppressing allele strengthened a recombination motif. Analysis of non-exchange conversion at this hotspot was used to determine whether this bias also exists in non-crossovers, and what the contribution of non-exchange conversion is on the meiotic drive.

Chapter Two Materials and Methods

2.1 Materials

2.1.1 Consumables and Equipment

Consumables and equipment were purchased from Amersham Biosciences (Little Chalfont, UK), Bando Chemical Ltd. (Kobe, Japan), Barloworld Scientific (Stone, UK), BD Biosciences (Oxford, UK) Bio-Rad (Hemel Hempstead, UK), Clare Chemical Research (Dolores, USA), Cecil Instruments (Cambridge, UK), Corning Ltd. (Hemel Hempstead, UK,), Duran (Mainz, Germany), Eppendorf Scientific (Cambridge, UK), Flowgen (Nottingham, UK) Fujifilm (London, UK), GE Water & Process Technologies (Heverlee, Belgium), Genetic Research Instrumentation (GRI) (Braintree, UK), Hybaid (Teddington, UK), Millipore (Watford, UK), MJ Research (Cambridge, USA, Nalge Nuc International (Hereford, UK), New Bruinswick Scientific Co. (New Jersey, USA), Precisa (Milton Keynes, UK), Sanyo (Watford, UK), Sarstedt (Leicester, UK); Sartorius Ltd. (Epsom, UK), Scilabub Ltd. (Measham, UK), Starlab (Ahrensburg, Germany), Stuart Scientific (Stone, UK), Syngene (Cambridge, UK), Thermo Electron Oy (Vantaa, Finland), Thermo Hybaid (Franklin, USA), and UVP Life Sciences (Cambridge, UK).

2.1.2 Reagents

Reagents were purchased from ABgene (Surrey, UK), Ambion (Austin, TX), Amersham Biosciences (Little Chalfont, UK), Applied Biosystems (Warrington, UK), Cambrex Bio Science (Rockland, USA/ Karlskoga, Sweden), Edge Biosystems (Gateshead, UK), Fisher Scientific (Loughborough, UK), Geneflow Ltd (Fradley, UK), Invitrogen (Paisley, UK), KAPA Biosystems (Labtech International, Ringmer, UK), New England Biolabs (Hitchin, UK), Perkin Elmer (Cambridge, UK) Promega (Southampton, UK), Quiagen Ltd (Crawley, UK), Roche (Welwyn Garden City, UK), Sigma-Aldrich Company Ltd. (Gillingham, UK) and Stratagene (Amsterdam, The Netherlands).

2.1.3 Standard solutions

Sodium chloride-Sodium citrate (SSC) buffer as well as 20× and 10×Tris-borate-EDTA (TBE) buffer were prepared as described by (Sambrook and Russel 2000) and were supplied by the media kitchen, Department of Genetics, University of Leicester, UK.

The Standard 11.1×PCR buffer was prepared as outlined in Table 2-1 (Jeffreys *et al.* 1988) and provided by R. Neumann, Department of Genetics University of Leicester, UK.

Table 2-1 Constituents of 11.1x PCR buffer

Reagent	Stock concentration	Volume added (µl)	Final concentration
Tris-HCl (pH 8.8)	2 M	167	45 mM
Ammonium sulphate	1 M	83	11 mM
Magnesium chloride	1 M	33.5	4.5 mM
2-mercaptoethanol	100 %	3.6	6.7 mM
EDTA (pH 8.0)	10 mM	3.4	4.4 µM
dATP	100 mM	75	1 mM
dCTP	100 mM	75	1 mM
dGTP	100 mM	75	1 mM
dTTP	100 mM	75	1 mM
BSA	10 mg/ml	85	113 µg/ml

Roche Applied Science supplied all dNTPs and Ambion supplied non-acetylated ultra-pure bovine serum albumin (BSA). Aliquots of 100 µl PCR-buffer were stored at -20 °C.

Fragmented DNA for size standards, namely λ DNA (Invitrogen) digested with HindIII and ΦX174 DNA (ABgene) digested with HaeIII, were supplied by H. Roe (Department of Genetics, University of Leicester, UK) in concentrations of 200 ng/µl.

2.1.4 Enzymes

T4 Polynucleotide Kinase and Exonuclease I were supplied by New England Biolabs. *Taq*-Polymerase was obtained from KAPA Biosystems (Labtech International, Ringmer, UK) and cloned *Pfu*-Polymerase from Stratagene (Amsterdam, The Netherlands). Roche supplied Shrimp Alkaline Phosphatase.

2.1.5 Oligonucleotides

All primers for PCR and all oligonucleotides for ASO hybridisation were supplied by Sigma-Aldrich Ltd (Haverhill, UK) in lyophilised form. They were dissolved in ultra-pure water in a category II laminar flow hood, and stored at -20 °C in stock concentration of 100 µM. PCR primers were also stored in working concentrations of 10 µM at -20 °C.

2.1.6 Human DNA samples

Sperm, and in some cases blood, were collected from 200 men of north European descent with approval from the Leicestershire Health Authority Research Ethics Committee, and with informed consent. Most of the semen samples were from anonymous donors attending fertility clinics, which were provided by J. Blower (Leicester Royal Infirmary, Leicester UK). Additional blood and semen samples, donated by volunteers from the Department of Genetics, were anonymised before use. Blood and sperm DNA samples were prepared under high containment to minimise the risk of contamination, in a class II laminar flow hood using designated pipettes. All solutions used for DNA preparation were UV treated to disable the amplifiability of any potential contaminating DNA.

A.J. Jeffreys extracted sperm and blood DNA as described in (Kauppi *et al.* 2009). The North-European donor panel was built from those men showing highest sperm DNA yield. As positive controls and referencing samples, four donors were included twice, thus generating 100 samples from 96 North-European men. Additional sperm DNA from 174 Afro-Caribbean and Zimbabwean men were prepared, giving in total genomic sperm DNA from 270 donors available for analysis. In addition donor-panel master-plates containing whole-genome multiple displacement amplified (MDA) DNA were generated by R. Neumann using the GenomiPhi Amplification kit (GE Healthcare).

2.1.7 Computers and Scripts

Data and images were stored and processed using the software packages AutoAssembler, EndNote, Papers for Mac, Factura, Microsoft Excel, Microsoft PowerPoint and Microsoft Word, as well as Apple iWorks: Pages, Numbers and Keynote. Images were either transferred to .jpg via Mac Preview or using the Epson Perfection 1250 Photosmart scanner.

LDU plots were generated from un-phased diploid genotyping data (2.3.1.1.) using LD mapping software modified by Adam J. Webb from version 1.5 of LDMAP (Maniatis *et al.* 2002).

Maximum-likelihood Poisson-approximation, two-sample confidence-interval simulation, and least-squares best-fit normal distribution analysis were each determined using bespoke simulation programs written by A.J. Jeffreys in True Basic 4.1 in the Classic set-up of Mac OS9 (all are available from A.J. Jeffreys upon request). Basic statistics were calculated using software available at http://faculty.vassar.edu/lowry/VassarStats.html. T-test performed on Poisson means were calculated using software available at http://www.quantitativeskills.com/sisa/statistics/t-test.htm.

Plotting and statistical computing was performed using the R script for statistical computing version 2.7.0 (2008-04-22) (R Development Core Team 2008) on a MacOSX. SimpleR was used for linear regression analysis as well as Welch two-sample t-test (Verzani 2008). Data were plotted using simpleR (Verzani 2008), as well as the "lattice" graphics package version 0.9-12 (Sarkar 2010) and the "plotrix" package version 2.4 (Lemon *et al.* 2008).

2.2 Methods

Standard molecular biology procedures were performed according to descriptions in (Sambrook and Russel 2000). All reagents used for PCR or in conjunction with sperm DNA were prepared in a class II laminar flow hood using designated pipettes to minimise the risk of contamination with PCR products or environmental DNA as described in (Jeffreys *et al.* 1990).

2.2.1 Quantification of genomic sperm DNA

DNA was quantified using a NanoDrop 1000 spectrophotometer, measuring absorbance at 260 nm. DNA was then diluted to working stock concentrations of ~50 ng/µl and ~10 ng/µl. To verify DNA concentrations and simultaneously assess DNA quality, serial dilutions of genomic DNA with known concentration were electrophoresed alongside 20 ng aliquots of sperm DNA on a 0.8% agarose gel (Seakem LE agarose, Cambrex, Karl-skoga, Sweden) in 0.5×TBE, with 0.2 µg/ml ethidium bromide (Biorad, Hercules, CA, USA) in both the gel and the buffer.

2.2.2 Standard long-range Polymerase Chain Reaction (PCR)

PCR reactions were carried out in 1x PCR-buffer (Table 2-1) plus 12.5 mM Tris-base [tris(hydroxymethyl-)aminomethane] ultra-grade for molecular biology (Fluka Chemie, Buchs, Switzerland) as well as 0.2 µM each of forward and reverse primer, 0.025 U/µl Taq-Polymerase and 0.0033 U/µl Pfu-Polymerase. Before adding the respective human DNA sample 0.5 µg/ml salmon sperm DNA (Sigma-Aldrich, Gillingham, UK) were added as carrier DNA. This decreases the deprivation of trace human DNA samples in the wall-cavities of PCR plastic-tubes, and therefore increases amplifiability of the respective DNA samples.

PCR primers were designed from the human genome consensus sequence (build GRCh37). Universal primers were typically 18-20 bases long with 50-70% GC content, preferably with between one and three G/Cs at the last three nucleotides.
PCR conditions were optimised for non-skirted 96-well plates (ThermoFast®96, Thermo Scientific) either on a Peltier Thermo Cycler 225 Tetrad DNA engine (MJ Research), a PCR System 9700 thermal-cycler (GeneAmp), a GS4 Thermal Cycler (G-Storm) or a

Veriti PCR machine (Applied Biosystems). Typically, an initial denaturation at 96 °C for 1.5 min was followed by PCR cycles of denaturing at 96 °C for 20 sec; annealing, with annealing temperatures, typically 2-3 °C lower than the calculated primer melting temperature, for 30 sec; extension at 65 °C for 60 sec per kilo-base of DNA + 1 minute. Primary PCR, amplifying from sperm DNA directly, used 25 cycles of amplification, while secondary PCRs re-amplifying primary PCR product used 32 -35 cycles. After cycling, a final elongation step of 65° C for 5 min was added. Primer annealing temperatures, extension times and cycle numbers are given in Appendix III.

2.2.3 Gel Electrophoresis

PCR products were mixed with 1/5th multi-purpose loading dye (30% (v/v) glycerol in 0.5x TBE plus enough bromophenol blue to give adequate blue colour) and then loaded into a 0.8% agarose gel (Seakem LE agarose, Cambrex, Karlskoga, Sweden) in 0.5×TBE, with 0.2 µg/ml ethidium bromide (Biorad, Hercules, CA, USA) in both the gel and the buffer. It was then electrophoresed at 80 -130 V, depending on expected PCR product size. Presence of amplified DNA was then checked on a Syngene UV-transilluminator. Gel photographs were produced using the GeneSnap imaging system from Syngene (Fisher Scientific). Migration distances of DNA samples were compared with those of a standard ladder consisting of 30 ng λ DNA×HindIII and 20 ng ΦX174 DNA×HaeIII.

2.2.4 Transferring PCR products onto dot blots

Dot blotting is a generic method for transferring PCR products onto membranes, which were then used for oligonucleotide hybridisation. Firstly 200 µl denaturing solution (0.5 M NaOH, 2M NaCl, 25 mM EDTA and bromophenol blue) was added directly to secondary PCR products. A 96-well dotblot manifold (DHM-96, Scie-PLAS, Southam, UK) was assembled with three 13 cm×9.5 cm sheets of 3MM filter paper (Whatman) and one 12 cm×8.5 cm sheet of Hybond-NX membrane (GE Healthcare, formerly Amersham) pre-soaked in ddH_2O. Vacuum was applied and the denatured PCR product was directly pipetted onto the membrane. Typically 50 µl denatured PCR product containing at least 30 ng of DNA was applied to each well of the manifold.

The DNA was neutralised by rinsing with 120 µl 2×SSC (2×sodium chloride-sodium citrate buffer: 0.3 M NaCl, 30 mM sodium citrate-dihydrate pH 7.0). After disassembly of

the manifold, all membranes were dried at 80 °C for 5 min. Due to the high DNA concentration of the secondary PCR product; it was possible to generate four replica filters per PCR plate. DNA was then cross-linked to the membrane using a pre-set UV exposure of 0.07 joules/cm^2 on a DNA cross-linker (UVP CL-1000, UVP Life Sciences). Dotblots were stored dry, and away from UV and heat, and could still be re-used after several years (A. J. Jeffreys, *personal communication*).

2.2.5 Allele specific oligonucleotide (ASO) hybridisation

ASO hybridisation was used for genotyping and linkage phasing of internal markers (2.3.2.4) and the detection and mapping of non-exchange conversion and crossover molecules in the half-crossover assay (2.3.2.5). Membranes were hybridised with γ^{32}P-labelled allele-specific oligonucleotides (ASOs) and autoradiographed to detect bound labelled ASO.

2.2.5.1 Designing ASOs

ASOs were 18 nucleotides long, with the selector SNP site located at the 8th base from the 5´-end. ASOs were diluted in ultra-pure water to a concentration of 100 µM and stored at -20 °C.

2.2.5.2 Radioactive labelling of ASOs

Before labelling, 2 µl of 100 µM ASO was diluted into 140 µl ddH_2O to a working concentration of ~8 µg/ml. A master reaction mix, with kinase buffer (70 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM spermidine trichloride, 2 mM dithiothreitol) and 4 U/µl of T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) as well as 0.12-0.2 µCi of γ^{32}P-ATP per µl (Perkin Elmer, Waltham, MA) was set up on ice. Then 9 µl of this reaction mix were aliquotted into a 1.5 ml screw top Eppendorf tube and 1.2 µl ASO working solution was added and incubated for 2 h at 37 °C or overnight at room temperature (22 °C).

2.2.5.3 Solutions for ASO Hybridisation

To allow ASOs to be hybridised at the same temperature, irrespective of GC content, tetramethyl ammonium chloride (TMAC) was used in hybridisation and washing solutions as it alleviates differences in nucleotide annealing temperatures. In addition 5×Denhards solution was used as a blocking agent in the hybridisation buffer (50×Denhardt´s solution was prepared by dissolving 5 g ficoll^{400}, 5 g polyvinylpyrrolidone as well as 5 g BSA in 500 ml ddH_2O, it is stored at +4 °C).

2.2.5.4 Hybridisation

Dot blots were soaked in 3×SSC (3×sodium chloride-sodium citrate buffer: 0.45 M NaCl, 45 mM sodium citrate-dihydrate, pH 7.0). Small hybridisation bottles (~250 ml volume) were rinsed with 3×SSC prior to inserting the dotblots with the DNA side facing inwards. Membranes were the pre-hybridised by adding 2.5 ml hybridisation solution (3 M TMAC (Sigma, St. Louis, MO) 0.6% SDS, 10 mM sodium phosphate pH 6.8, 1 mM EDTA, 4 µg/ml yeast RNA, 5x Denhardt´s solution.) pre-warmed at 48.5 °C to the bottle and hybridising at 48.5 °C for 5 -10 min in a hybridisation oven (Thermo Shake ´n Stack).

The hybridisation solution was replaced with 2.5 ml fresh hybridisation solution and again incubated at 48.5 °C for 5 min. Before adding the labelled ASO to the hybridisation bottle, the labelling process was stopped by adding 20 µl of STOP-solution (25 mM diNa EDTA, 0.1% SDS, 10 µM ATP) to the labelling ASO. The blots were hybridised with the labelled ASO at 48.5 °C for 1 hour.

Filters were then washed by increasing the oven temperature to 54 °C. The hybridisation solution was discarded and 2 ml TMAC-wash-solution (3 M TMAC, 0.6% SDS, 10 mM sodium phosphate pH 6.8 and 1 mM EDTA) added and incubated for 1 to 5 min. The TMAC wash solution was exchanged for 2 ml fresh TMAC wash solution and incubated for 1 to 5 min. TMAC wash solution was exchanged twice for 3 ml TMAC wash solution, each time incubated for 10 – 20 min. After washing, the membranes were rinsed twice in the hybridisation bottle using 3×SSC at room temperature, and then transferred to a tray of 3×SSC for extensive washing to remove toxic TMAC.

2.2.5.5 Autoradiography

The membranes were blotted using Whatman 3MM paper to remove excess liquid, and then sealed in Saran plastic wrap to prevent drying out. They were auto-radiographed at -80 °C on X-ray film (Bio-Max MS film, Kodak, Rochester, NY, USA) using an intensifier screen for up to three days, depending on signal intensity.

2.2.5.6 Stripping and reusing of dot blots

To remove the radioactive probe, all membranes were pooled and washed in multiple changes of ~200 ml boiling 0.1% SDS. The probe removal was monitored with a Geiger counter, and when the count dropped down to below five counts per sec, the membranes were rinsed extensively in dH_2O at room temperature. The blots were then stored damp in Saran plastic wrap at +4 °C or re-used for hybridisation immediately. For long-term storage, blots were dried completely for 10 min at 80 °C and stored cold and dry.

2.3 Recombination assays

To directly recover recombinants from sperm DNA two types of recombination assays have been developed (Kauppi *et al.* 2009). The crossover assay relies on repulsion-phase allele-specific PCR using allele-specific primers (ASPs) flanking both sides of a hotspot. These ASPs are directed to nested selector sites, with forward and reverse primers directed to selector sites on different progenitor haplotypes, so that only crossover molecules will amplify. The crossover assay can efficiently score large pools of sperm DNA containing >1000 amplifiable haploid genomes for the occurrence of crossovers, detectable as pools returning PCR products. This assay can be used to determine crossover frequencies across a large number of men at a given hotspot, as described in (Kauppi *et al.* 2009). Gene conversion molecules derived from non-reciprocal recombination events are not detectable in the crossover assay, as gene conversion does not involve the exchange of flanking markers.

Alternatively it is possible to detect and characterise crossovers (CO) and non-exchange gene conversion molecules (CONs) simultaneously. Using allele-specific PCR, a hotspot region is selectively amplified from small pools of sperm DNA, containing between 15 and 45 amplifiable sperm DNA molecules, with ASPs flanking one side of the hotspot. Selective haplotype amplification is achieved by using nested ASPs, which flank only one

side of the hotspot. As only one set of ASPs are used, compared with two sets of ASPs in the full-crossover assay, this assay was therefore referred to as half-crossover assay.

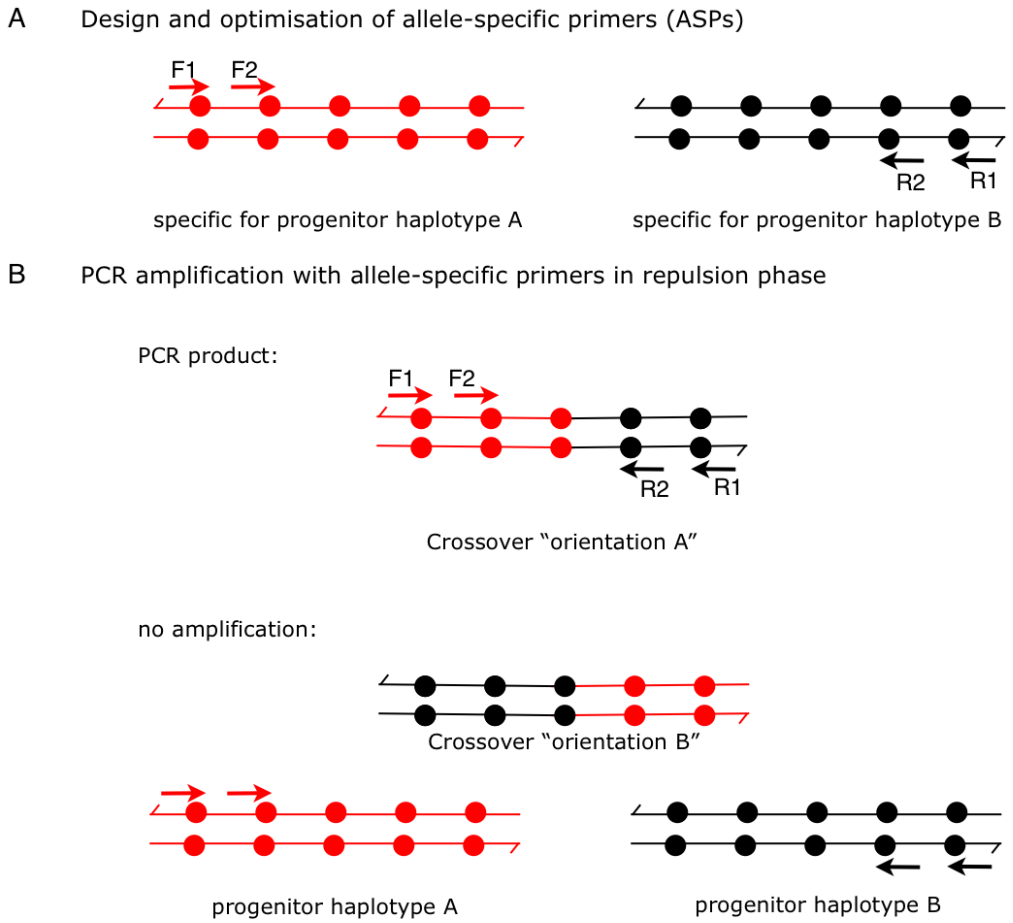
Half-crossover assays were performed in both reciprocal orientations, by alternatively amplifying each haplotype separately, and then probing for the presence of recombined markers by hybridisation with ASOs specific to alleles of the non-selected haplotype. The half-crossover assay not only allows the detection, but also the characterisation of recombination events at high resolution, only limited by the SNP density within a given hotspot interval.

Methodological aspects of both assays are presented in 2.3.1 and 2.3.2, with more detailed accounts of the specific requirements for the half-crossover and the full-crossover assay discussed in Chapter Three and Chapter Five, respectively.

2.3.1 Crossover assay

Crossover generation is reciprocal; therefore it is sufficient to perform a crossover assay on either reciprocal orientation to measure crossover frequency. Assay orientations are defined from the phasing of selector alleles on crossover molecules, an assay in orientation A would return crossovers with markers of haplotype A upstream, and markers of haplotype B downstream of the hotspot. The reciprocal assay uses forward primers specific to haplotype B and reverse primers specific to haplotype A, to detect crossovers in orientation B. The principle of crossover detection using repulsion-phase allele-specific PCR is illustrated in an example for a crossover orientation A in Figure 2-1.

Figure 2-1 Crossover assay



Each circle represents an informative SNP. Progenitor molecules do not amplify, as forward and reverse priming sites (arrows) were not located on the same molecule. Only recombinants in the chosen orientation (e.g. red to black) will amplify as priming sites have been recombined to the same molecule

2.3.1.1 SNP genotyping

Hotspot target regions were amplified in two to four short, partially overlapping, PCR targets by two successive rounds of nested PCR using universal primers. MDA DNA donor-panel master plates are used to seed reactions. Secondary PCR products were then transferred onto nylon-membranes using the standard dotblotting protocol for PCR plates (as described in 2.2.4). Typically 2-6 replicates of each amplicon were blotted, depending on the number of SNPs that needed to be typed.

2.3.1.2 Allele-specific primers

Using heterozygous SNPs as selector sites for allele-specific PCR amplification requires primers to be highly specific for the chosen SNP allele, and unique to only this locus in the genome. ASPs were designed with the target SNP nucleotide at the 3´-end; complementary to the respective SNP allele of the haplotype to be amplified. The genomic context of this SNP dictated the constitution of the primers, which were designed to be 15-19 bases long and preferably with at least 50% G/C content. Very G/C rich ASPs can be as short as 14 bps. The conditions for the ASPs were optimised to achieve highest specificity with highest yield of long PCR products. The efficiency of primers located in AT-rich sequences was boosted by increasing the G/C content of the primer through adding a synthetic CCCC 5´-extension to the primer.

2.3.1.3 Optimisation of primer pairs

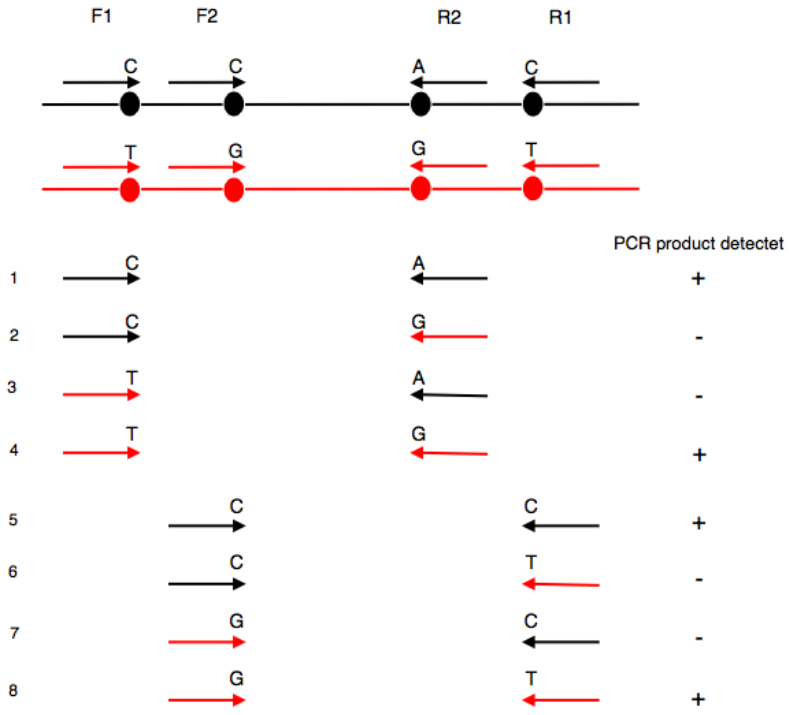
The selector sites for allele-specific primers have to be located outside of the hotspots, and can be located up to 13 kb apart, which is the maximum distance that can still be efficiently amplified at the single molecule level. Primer pairs used in repulsion-phase allele-specific PCR had to work well with each other, requiring high specificity and efficiency of both primers at the same annealing temperature. Each ASP was tested by amplifying 4-8 kb targets from 20 ng of MDA DNA, using the ASP plus a universal primer in opposite orientation. The optimal annealing temperature of each ASP was determined by annealing temperature titration (typically at 56, 58, 60, 62 and 65 °C). To test the specificity of ASPs, two PCR reactions were set up, one amplifying from individuals homozygous for one ASP allele, and the other from individuals homozygous for the alternative allele. To decrease any effects of variable quality of MDA DNA, three aliquots were pooled from men homozygous for each allele respectively, where available. Finally the universal primer was tested in combination with another universal primer at each annealing temperature to ensure that it worked efficiently at all annealing temperatures.

ASPs that were not specific and efficient, or assays in which forward and reverse primers would not work at the same temperature, required the re-design of ASPs. Low efficiency primers were improved by adding synthetic GC-rich 5´-extensions, and low specificity primers could often be improved by shortening by one or two bases. The final ASPs used are listed in Appendix I.

2.3.1.4 Determining linkage phase of selector SNP alleles

Selector SNP sites for forward primers F1 and F2 and selector SNP sites for reverse primers R1 and R2 are separated by the hotspot, and thus expected to be in free association, as indicated in Figure 2-2. For donors chosen for analysis the linkage phase of forward and reverse sets of alleles had to be determined by using allele-specific primer pairs in all possible combinations. If both, F1 and F2 as well as R1 and R2 are in strong LD respectively, then it is sufficient to only phase primer combination F1-R2 or F2-R1, as the remaining haplotypes can be deduced from LD. If only one pair of primers, either forward or reverse is in LD then both combinations F1-R2 and F2-R1 have to be analysed. If neither is in strong LD, then ASP combinations of nested alleles F1-R2, F2-R1 and F2-R2 have to be used. A hypothetical example of phasing PCRs using allele combinations of the F1-R2 and F2-R1 primers is illustrated in Figure 2-2.

Figure 2-2 Principle of linkage phasing, modified from (Kauppi *et al.* 2009)



DNA segments on homologous chromosomes are shown as black and red lines and circles represent heterozygous SNPs. Arrows represent ASPs, F1 and F2 are nested forward primers, R1 and R2 are nested reverse primers. All allele-specific primer combinations were tested (1-8) Reactions 1-4 contained primers specific to either F1 allele, as well as R2 alleles, and reactions 5-8 contained primers specific to either of the F2 and R1 alleles. Positive reactions that would be obtained in this hypothetical assay were indicated by + (plus), negative PCR reactions by – (minus)

PCR products are only generated when alleles that primers are specific to lay on the same haplotype, and thus primer phase can be deduced. Outside markers F1 and R1 were never used together, as this would generate PCR products that could easily contaminate primary PCRs, generating false positives in the subsequent crossover assays.

2.3.1.5 Amplification of crossover molecules from sperm DNA

Crossover molecules were selectively amplified from pools of sperm DNA using long PCR of 8-12 kb with forward and reverse primer in repulsion-phase. Diluted primary PCR products were re-amplified with nested ASPs. Typically four different pool sizes of 11 or 12 pools each were used per man. Pool sizes were chosen to contain 0.6, 1.2, 2.4 and 4.8 expected crossovers per pool, based on the crossover frequencies taken from published sources. Two men were analysed within the same using 96-well plate experimental set-up, with two wells per plate used as positive controls containing 0.1 ng and 1.0 ng DNA from a man with selector sites in coupling phase on the same haplotype.

2.3.1.5.1 Genomic sperm DNA input

As sperm cells are haploid, statistically only every second sperm genome would be of the haplotype analysed in the respective assay. In addition, not all DNA molecules in the PCR reaction will amplify, as primers are never fully efficient and DNA damage within the target sequence or priming sites might interfere with annealing of primers or PCR product extension. Extensive analysis of input molecule amplification efficiency for crossover assays has shown a consistent amplification efficiency of 50% (± 12%) (Jeffreys *et al.* 2000). Therefore sperm DNA input was calculated not per molecule, but per amplifiable sperm DNA molecule of the correct haplotype, which corresponds to 12 pg of genomic DNA (Jeffreys *et al.* 2000).

2.3.1.5.2 Primary PCR amplification from sperm DNA

Primary PCR was used to amplify a chosen locus directly from sperm DNA. Primary amplification was carried out in volume of 8µl reaction mixture. Different input concentrations were generated by serial dilution of genomic sperm DNA into the buffer mixture, containing carrier DNA and Polymerases, and then aliquotted directly into the PCR plate. After completion of the final elongation step, PCR plates were directly transferred on ice and diluted with ten times their volume of Diluent (5 mM Tris-HCl pH 7.5, 2 µg/ml

salmon sperm DNA). Diluted 1° product was used immediately to seed plates of secondary PCR, and then stored at -20 °C. Final PCR conditions used in crossover assays are collected in Appendix VII.

2.3.1.5.3 Secondary PCR amplification of primary PCR product

A nested set of ASPs was used for secondary PCR amplification to increase the PCR yield and haplotype specificity of each set of ASPs. It was set up immediately after completion of primary PCR. Secondary PCR reaction volume was 12.5 µl of reaction mixture plus 1.5 µl of diluted primary PCR product. PCR amplification was performed for 35 cycles.

2.3.1.6 Determining crossover frequencies

Aliquots of secondary PCR product were analysed for positive reactions by ethidium bromide staining after agarose gel electrophoresis (2.2.3). For gel electrophoresis, 3 µl of loading-dye was added directly to the 15 µl of secondary PCR product. Then 10 µl of this mixture was loaded into a 0.8% agarose gel. The mean number of crossovers per PCR reaction was then estimated from the number of negative reactions (Nr_{neg}) and the total number of reactions (Nr_{tot}) based on Poisson-approximation:

$$m = -ln(\frac{Nr_{neg}}{Nr_{tot}})$$

Crossover frequency can then be calculated by dividing the Poisson-corrected number of crossovers by the number of input molecules analysed per PCR (Nr_{mol}).

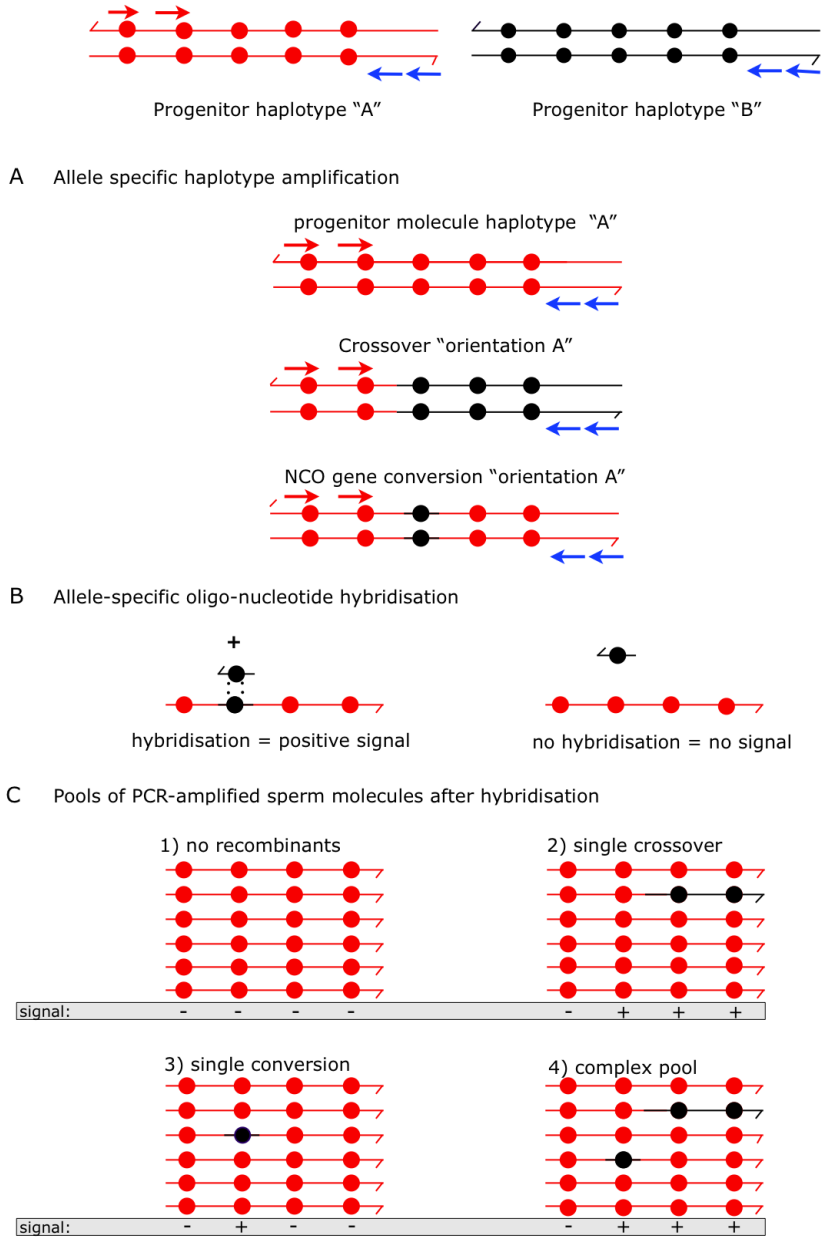
$$freq = \frac{m}{Nr_{mol}}$$

To get a more accurate approximation of the crossover frequency over multiple input pool sizes, a maximum-likelihood analysis was performed using software written by A.J. Jeffreys in True Basic 4.1, which combines data from all input pool sizes.

2.3.2 Half-crossover assay

The half-crossover assay was used to simultaneously detect crossover and non-exchange gene conversion molecules from small pools of genomic sperm DNA. Recombinant detection and breakpoint mapping was carried out by probing selected haplotypes with ASOs complementary for SNP alleles from the opposite, non-amplified, haplotype. A schematic representation of a half-crossover assay in orientation A is illustrated in Figure 2-3.

Figure 2-3 Half-crossover assay



(A) PCR products are generated from recombinant molecules in orientation A, as well as progenitor molecules of haplotype A (red) by selective amplification through two successive rounds of allele-specific PCR with two nested sets of primers, each using an ASP (red arrows) together with a universal primer (blue arrows).
(B) Membranes were hybridised with ASOs complementary to progenitor haplotype B (black) for all SNPs within the hotspot interval.
(C) Most pools do not contain recombinant molecules (1), while those pools that do contain recombinant molecules (2)-(4) return positive hybridisation signals. Crossovers are positive for a continuous string of markers (2) while non-exchange conversion events are positive at typically one or two markers, but not at flanking markers. (3) Pools can potentially contain more than one recombinant molecule (4) that cannot be resolved.

2.3.2.1 Primer design and optimisation

The half-crossover assay used ASPs specific to either a forward or the reverse set of selector sites in combination with universal primers located on the opposite side of the hotspot to selectively amplify one of the two haplotypes in a man carrying suitable SNP heterozygosities. Nested allele specific primers directed to SNP sites on one side, as well as nested universal primers located on the other side of the Superhotspot have to be designed. ASPs were designed (as described previously 2.3.1.2) and optimized to highest specificity with highest yield of long PCR products (as described in 2.3.1.3.). Annealing temperatures were determined in combination with universal primer through annealing temperature titration on sperm DNA on eight different temperatures across the temperature range where both ASPs worked highly specifically, typically in 0.5 °C steps. Universal primers that would not work efficiently at a temperature where the ASPs worked highly specifically, resulted in re-design of the universal primer.

2.3.2.2 Input pool sizes

As recombinant detection relies on ASO hybridisation, this creates a limitation on the pool size. ASO hybridisation has to be specific enough to distinguish a single molecule with recombinant markers from background binding to other non-recombinant molecules, and pools with up to 80 amplifiable molecules have been used successfully as described in (Kauppi *et al.* 2009). Secondly, ASO detection cannot distinguish one recombinant in one pool, from more than one recombinant present within the same pool. Therefore the pools size has to be chosen small enough to minimize the likelihood of multiple recombinants per pool. Pilot assays were set up with two pool-sizes based on known crossover frequencies and assuming a 1:1 ratio of crossover to detectable non-exchange conversion for the smaller pool size and a 1:2 ratio for the larger pool size. Subsequent assays then used optimized input pools sizes calculated from the observed non-exchange gene conversion and crossover frequencies.

2.3.2.3 Haplotype separation PCR

To ensure efficient haplotype separation, two nested sets of allele-specific primer (ASP) and universal primer were used. Re-amplifying primary PCR product also increased the PCR yield. Primary PCR uses the outside ASPs, designed as described previously (2.3.1.2) together with the outside universal primer at optimized PCR conditions for this primer pair.

Primary PCR amplification efficiency was checked in a subset of 12 PCR products by gel electrophoresis, by mixing 4 µl of undiluted primary PCR with 1 µl of 5x loading dye and electrophoresing the PCR products on a 0.8% agarose gel (as described in 2.2.3). To increase the PCR yield and the specificity of haplotype separation, the primary PCR product was re-amplified using the nested ASP together with a nested universal primer. Aliquots of 0.7 µl from a subset of 12 secondary PCR products were checked for uniform amplification yield and PCR efficiency by gel electrophoresis (see 2.2.3). Recombinant molecules carrying the selector marker were selectively amplified, together with molecules of one haplotype spanning the hotspot. Assays were also performed in reciprocal orientation, by amplifying the alternative haplotype.

2.3.2.4 Haplotyping via ASO hybridisation

The linkage phase of internal SNPs was determined by testing separated haplotypes by sequential ASO hybridisation at each of the heterozygous SNPs. Secondary PCR products of separated haplotypes generated at step 2.3.2 were denatured and dotblotted, with only one dot for each haplotype, as described in 2.2.4. For each SNP, both alleles are tested by subsequent hybridisation to the same blot as described previously (2.2.5).

2.3.2.5 Identifying recombinant DNA molecules via ASO hybridisation

To identify crossovers and non-exchange conversion molecules, separated haplotypes generated in step 2.3.2.3 were transferred onto membranes (as described in 2.2.4). Membranes were then probed for the presence of recombinants by hybridisation with γ^{32}P-labelled ASOs specific to the non-selected haplotype, using the TMAC method as described in 2.2.5. Comparing the acquired SNPs to parental haplotypes can detect whether the recombinant is a non-exchange gene conversion or crossover molecule.

As pools contain mainly progenitor molecules, the standard ASO-hybridisation protocol described in 2.2.5 was adapted. To decrease background binding to progenitor molecules, 1.12 µg of unlabelled competitor ASO, the ASO complementary to the progenitor haplotype, were added to the hybridisation solution, followed by hybridisation for 2-5 min prior to adding the labelled ASO. In addition the stringency of washing was adjusted to the specificity of the ASO if necessary, with more washes, washes at higher temperature, and/or a higher initial hybridisation temperature to increase ASO specificity.

A signal to noise ratio of 80:1, allowing recombinant detection in pools of DNA containing up to 80 molecules was readily achieved for the most specific ASOs as described in (Kauppi *et al.* 2009). To improve the occasional ASO which showed low specificity, and therefore caused background signal, the hybridisation temperature was raised and the stringency of washing increased through higher washing temperatures and longer washing time. In almost all cases, highly specific ASOs were generated for each SNP, rarely ASOs not specific enough to discriminate between the alleles could be observed, which indicated the necessity for ASO re-design to the opposite strand of DNA. In rare cases, when more than one SNP site was present within the ASO binding site, and ASOs containing ambiguity bases were unspecific, donor specific ASOs were designed for each of the tested individuals. Highly specific ASOs, as well as optimised hybridisation and washing temperatures are collected in Appendix IV.

2.3.3 SNP discovery

To ensure the highest possible SNP density for recombinant detection, and to identify polymorphisms not known from dbSNP, separated haplotypes of individuals were sequenced from PCR products generated at step 2.3.2.3. Sequencing targets were typically expected to be >600 bp long, therefore sequencing primers were designed to 3-4 targets to cover the hotspot widths of 1-2 kb. To aid assembly of sequences, targets were designed to overlap by ~100 bp. The hotspot centre was sequenced from both directions using forward primers proximal and reverse primers distal to the centre.

2.3.3.1 Purification of PCR product

This protocol simultaneously removes leftover primers as well as dNTPs. 7 µl of PCR product were mixed with 0.7 µl exonuclease I (20 U/µl, New England Biolabs) and 2.1 µl shrimp alkaline phosphatase (1 U/µl Roche), and incubated at 37 °C for 60 min, followed by heat inactivation of the enzymes at 80 °C for 15 min.

2.3.3.2 Sequencing reaction using Big Dye Terminators (Sanger sequencing)

Sequencing reactions of 20 µl were set up using 0.5×Big Dye Terminator Ready Reaction Mix v 3.1, 0.875×Big Dye Terminator Buffer, 20-30 ng/kb purified PCR product and 0.16 µM sequencing primer. The reaction was cycled for 10 sec denaturation at 96 °C followed by 5 sec annealing at 50 °C and sequence extension at 60 °C for four min per cycle for 25 cycles.

2.3.3.3 Cleanup of sequencing products

Completed sequencing reactions received 2 µl of 2.2% SDS and were then incubated for 5 min at 98 °C, followed by cooling at 25 °C for 10 min. The sequencing reaction was then purified using Performa® DTR-gel filtration Cartridges according to the manufacturer's protocol to remove dye terminators, dNTPs and other low molecular weight material from sequencing reactions by centrifugation. The cleaned sequencing products were submitted to the Protein and Nucleic Acid Laboratory (PNACL, University of Leicester) for nucleotide detection. Sequence traces were then supplied via the Universities CFS network.

2.3.3.4 Detection of polymorphisms

Sequence data were edited and assembled using Factura and Auto Assembler software on the Apple Macintosh computer in the classic (OX9) set-up. Sequence confidence ranges were identified by eye and were typically 600-900 bp in length. Possible SNPs were detected by the Auto Assembler software and identified by eye from the sequence traces. All SNPs identified through re-sequencing of single donors were genotyped across the semen donor panels by ASO hybridisation (described in 2.3.1.1), and a short tandem repeat was genotyped across a specific panel of men as described in 2.3.4.

2.3.4 Short Tandem Repeat (STR) genotyping

The STR target for PCR amplification was designed with at least 50 bp of non-repetitive sequence flanking the STR on both sides. PCR amplification of this target used a set of universal primers, with the forward primer labelled using a 5'-Hexachloro-Fluorescein (HEX) tag. Depending on STR repeat number, amplicon lengths of 300 – 400 bp were expected. The PCR reaction mixtures were set up with 1×PCR buffer (Table 2-1), 0.5 µg/ml salmon sperm DNA, 0.2 µM Forward primer (5'-HEX labelled), 0.2 µM Reverse primer, 0.025 U/µl KAPA Taq Polymerase and 1 ng PCR clean sperm DNA, in a volume of 10 µl. PCR conditions were 96 °C for 1 min 30 sec, followed by 33 cycles of denaturing at 96 °C for 30 sec, annealing at 57 °C for 30 sec, and extension at 65 °C for 1 min 30 sec. The PCR reaction was concluded with a final extension step of 65 °C for 5 min. The PCR products were then purified using the cleanup protocol as described in 2.3.3. Short Tandem Repeat length variation was then determined by capillary electrophoresis using a 3130xl Genetic Analyzer (Applied Biosystems).

Chapter Three Gene conversion frequency and distribution within human recombination Superhotspots

3.1 Introduction

In mammalian meiotic recombination, only a fraction of double-stranded breaks are repaired as crossovers (COs) (Baudat and de Massy 2007). The majority of breaks should be repaired as non-exchange gene conversion replacing the DNA sequence without exchange of flanking markers, also referred to as non-crossover (NCO) conversions. The distribution of crossovers was being studied extensively in selected regions of the human genome by single DNA molecule analysis of sperm. To date 40 recombination hotspots have been directly characterised with this method, showing varying crossover frequencies of between 0.0004% and 2.5% (Jeffreys *et al.* 2004; Berg *et al.* 2010). But only very little is known about the frequency and distribution of non-exchange conversions in the human genome and how well they correlate with crossover.

At the beginning of this work five hotspots had been analysed for both crossovers and non-exchange conversions. These hotspots were located in the Major Histocompatibility Complex (MHC) (Jeffreys and May 2004), the major pseudo-autosomal region on the sex chromosomes (PAR1) (Jeffreys and May 2004) and the ß-globin gene region (Holloway *et al.* 2006). Only one hotspot that was characterised for non-exchange conversion was located in a region with average gene density (1q42.3). However, this intragenic hotspot, *NID*1, is associated with a minisatellite (Jeffreys *et al.* 1998b).

Non-exchange conversion molecules can be successfully detected in sperm DNA using DNA enrichment by allele-specific hybridisation (DEASH). This method addresses one internal SNP at a time with biotinyl-labelled ASOs (Jeffreys and May 2003) and is therefore labour-intensive. DEASH was applied to hotspots *DNA3*, *DMB*2, *SHOX* and *NID*1, and aided the detection of non-exchange conversions at frequencies of 0.0004 % to 0.03 % (Jeffreys and May 2004; Jeffreys and Neumann 2005).

The ability to efficiently detect non-exchange conversion events without enrichment would facilitate the characterization of non-exchange conversion across several hotspots. A PCR based approach, the half-crossover assay, could be used for that purpose. The half-crossover assay relies on allele-specific amplification of one haplotype at a time, followed

by probing for the presence of recombined markers using allele-specific oligonucleotide hybridisation specific to alleles of the non-selected haplotype. An attempt to simultaneously detect crossover and non-exchange conversion molecules with this method at the ß-globin hotspot, had however failed to return any non-exchange conversion events, though detected a crossover frequency of 0.03%. To use the half-crossover assay efficiently appears to require hotspots with much higher recombination frequencies than the ß-globin hotspot. Sperm recombination hotspots very active for crossover should also be very active for non-exchange conversion, as the same initiating lesions appear to generate crossover and non-crossover products (Jeffreys and May 2004).

A key study by A.J. Webb and colleagues detected sixteen hotspots with crossover frequencies above 0.1%, which were defined as Superhotspots (Webb *et al.* 2008). Subsequently five additional putative Superhotspots were identified by A.J. Jeffreys based on familial linkage data collected by (Coop *et al.* 2008), which were then validated as highly active sperm crossover hotspots (Jeffreys and Neumann 2009; Berg *et al.* 2010). These Superhotspots provided an excellent resource for the characterisation of the distribution and dynamics of non-exchange gene conversion events without the need for enrichment.

3.2 This work

The most suitable hotspots for crossover and non-exchange conversion analysis were chosen from a panel of Superhotspots. Half-crossover assays were designed and optimised for five Superhotspots that would allow the detection of crossovers and non-exchange conversion molecules. The crossover frequencies and breakpoint distributions obtained in these half-crossover assays were compared with published data to validate the method used to detect non-exchange gene conversions. Detected non-exchange conversion molecules were mapped across the hotspot interval. This gave limited information on conversion tract lengths that were obtained by comparing minimal and maximal tract length estimates. To determine for the first time how well non-exchange conversion frequencies correlate with crossover frequencies, they were compared between hotspots.

3.3 Hotspot analysis via the Half-crossover assay

The half-crossover assay is a PCR based approach using allele-specific amplification of a 5-13 kb wide Superhotspot target interval. For selective haplotype amplification, target intervals not only have to include the hotspot but also sufficient flanking sequence outside of the hotspot to allow suitable heterozygous selector sites. Using suitable markers as selector sites located in a LD block flanking either side of the hotspot, haplotypes were separately amplified by two rounds of nested allele-specific PCR from small pools of 15-45 amplifiable sperm DNA molecules. Separated haplotypes were then tested for the presence of markers from the opposite haplotype, to detect recombinants (2.3.2).

Conversion tracts detected in human sperm DNA at hotspots *DNA3* and *SHOX* clustered at the centre of the crossover distribution and were very short (Jeffreys and May 2004). As non-crossover not spanning an informative marker would not be detectable as non-exchange conversion event, their detection thus crucially relies on the presence of marker(s) very close to the hotspot. Additional markers are also needed to allow clear distinction of crossover events from non-exchange conversion events.

Recombinant detection in the half-crossover assay is based on ASO hybridisation to amplified pools of sperm DNA that contain mostly progenitor molecules. The detection of non-exchange gene conversion events involving a single marker hence relies on the successful distinction of single-base changes against a background of progenitor molecules. As smaller pool sizes allow a better signal-to-noise ratio, it followed that hotspots with higher recombination frequencies would facilitate clear distinction of non-exchange gene conversion molecules from progenitor background noise.

3.3.1 Selection of hotspots for analysis

Candidate hotspots for half-crossover assays were selected from a panel of Superhotspots identified by (Webb *et al.* 2008; Berg *et al.* 2010). Only Superhotspots with crossover frequencies above 0.1% were chosen. In addition, Hotspots that lacked markers within 150 bp of the centre were excluded at this stage. Ten hotspots fulfilled these initial criteria. A high SNP density across the hotspot, and especially close to the hotspot centre, was desirable to be able to detect double-site events. The five hotspots with the highest crossover frequencies, and the highest SNP densities within the hotspot interval and close to the hotspot centre, were analysed in the timeframe of this work, four from (Webb *et al.* 2008) and one from (Berg *et al.* 2010); these are summarised in Table 3-1.

Superhotspot K was especially suitable for non-exchange conversion analysis, since two SNPs directly flank the hotspot centre, which are located only 23 bp apart. These two SNPs should allow the detection of even very short gene conversion tracts on both sides of the hotspot centre. The SNP proximal to the hotspot centre is directly adjacent to an extended degenerate motif associated with hotspots (CACCNNTCCCTCC) (Myers *et al.* 2006).

A second Superhotspot contains a sequence (CCACCCTGACCCC) complementary to 11 of 13 bases of this motif (Myers *et al.* 2006). This motif was located 47 bp from the centre of Superhotspot F (Webb *et al.* 2008). The other hotspots did not contain an obvious hotspot recombination initiation motif (Webb *et al.* 2008; Berg *et al.* 2010).

Table 3-1 Superhotspots chosen for analysis

Superhotspot	E	F	H	K	T
chromosome	8q	12p	3p	8q	3p
centre location (GRCh37)	137,524,198 likely intergenic (hypothetical protein LOC51059)	5,749,255 intronic (ANO2)	14,817,625 intergenic	94,302,762 intergenic	2,366,642 intronic (CNTN4)
95% width (kb)	1.4	1.4	1.2	1.4	1.5
Peak crossover frequency (cM/Mb)	550	1100	180	260	1200
mean cross-over frequency per sperm (%)	0.49	0.97	0.13	0.22	1.10
Central SNP(s) and distance to hotspot centre	rs28541869 10 bp rs2255417 49 bp	rs10492181 108 bp	rs3899614 102 bp	rs1374632 4 bp rs1374633 20 bp	rs9854419 10 bp

Superhotspots E, F, H, K were identified by (Webb *et al.* 2008), Superhotspot T was identified by (Berg *et al.* 2010) based on familial crossover clustering data from (Coop *et al.* 2008). Sperm typing of three men per hotspot determined mean crossover frequencies. Hotspot centre and 95% hotspot widths, the interval within which 95% of crossovers exchange points occur, were both determined from a least-squares best fit normal distribution of crossover exchange points pooled from all three men. Genomic locations are based on Genome Reference Consortium Human genome build 37, patch release 2 (GRCh37.p2).

Two of the selected Superhotspots were located within genes (Table 3-1). Superhotspot F is located intronic of anoctamin 2 (*ANO2*), coding for a *trans*-membrane protein with eight membrane-spanning domains that acts as a calcium-activated chloride channel in photoreceptors (Stöhr *et al.* 2009). Superhotspot T is located in an intron of contactin-4 (*CNTN4*), which codes for Contactin-4, a member of the immunoglobulin superfamily (Yoshihara *et al.* 1995). Contactin-4 is a CPI anchored neuronal membrane protein (Hansford *et al.* 2003), responsible for cell-surface interactions in neuronal cell development. *CNTN4* is a candidate gene for spinocerebellar ataxia type 16 (Miura *et al.* 2006), and deletion of this gene may be responsible for the autism spectrum disorder associated with 3p deletion syndrome (Cottrell *et al.* 2011).

3.3.2 Identifying men most suitable for analysis

The most suitable donors for analysis were identified from un-phased diploid genotyping data of 94 North-European semen donors (the European donor panel), which had been generated by (Webb *et al.* 2008). Men only qualified for analysis if they were heterozygous at the most central SNP for which the hotspot was originally chosen in addition to two selector SNPs used for haplotype separation PCR. In addition donors also had to have heterozygous SNP sites across the hotspot interval plus sufficient heterozygous SNPs outside the hotspot interval to differentiate crossover and non-exchange gene conversion molecules.

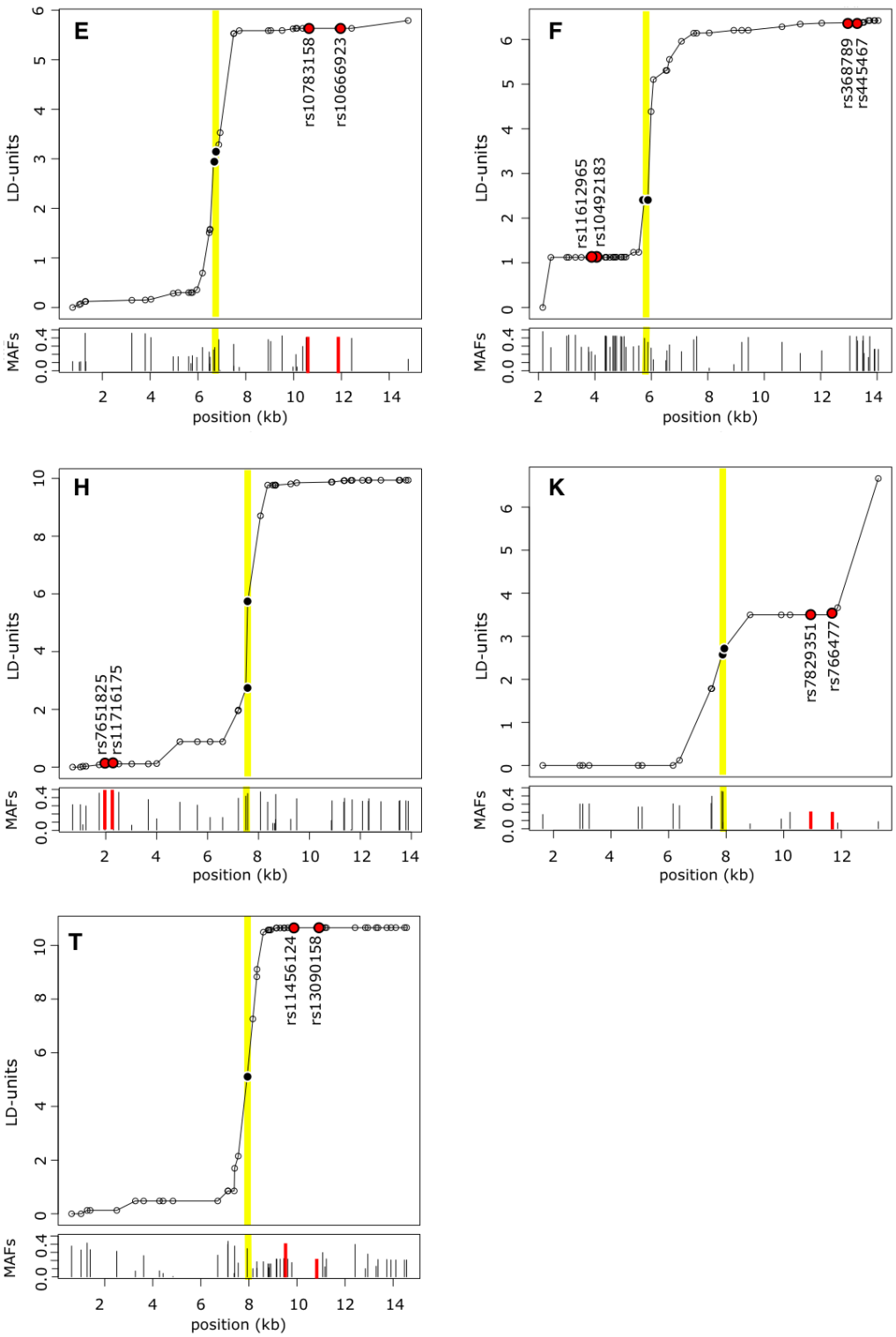
Typically 5-10 men from the donor panel fulfilled these stringent criteria. As blood-controls were available for only very few men, it was not possible to test whether non-exchange conversions would arise in blood DNA. But the authenticity of sperm recombinants has been established by testing large samples of blood DNA molecules for recombinants, and neither crossover (Jeffreys *et al.* 2001) nor non-exchange conversion molecules (Jeffreys and May 2004) have ever been found in blood DNA, consistent with these being products of meiotic events.

Two men were typically analysed per hotspot, to compare recombination frequencies at a given hotspot. The donors with the highest density of informative SNPs across the hotspot interval were chosen.

3.3.3 Half-crossover assay design at each of the selected Superhotspots

Half-crossover assays were developed for each of the five Superhotspots chosen for analysis. To illustrate the location of hotspot centres determined previously (Webb *et al.* 2008; Berg *et al.* 2010) as well as the location of selector sites used for half-crossover assays, metric LD plots for each target interval are shown in Figure 3-1. Breakdown of association, and therefore inferred historical recombination is displayed by an increase in metric LD. The markers located on the largest LD step (yellow lines in Figure 3-1), are equally the locations of central marker(s) within the crossover hotspot, as determined by sperm typing (black filled circles in Figure 3-1) (Webb *et al.* 2008; Berg *et al.* 2010).

Figure 3-1 Metric LD maps across each of the chosen Superhotspot target intervals



Metric LD plots in cumulative linkage disequilibrium units (LDU) are displayed for each Superhotspot, with their initial in the top left corner within each plot.
Upper panels: Circles represent polymorphic SNPs, at their target interval position in kilo bases (kb). Central SNPs (black) taken from (Webb *et al.* 2008; Berg *et al.* 2010), SNPs used as selector sites (red) are labelled with their rs numbers. Hotspot centres are highlighted in yellow. Lower panels: Ticks represent minor allele frequencies (MAFs) - the population frequency of the less common allele within the European donor panel.

As shown in Figure 3-1, allele-specific primers for half-crossover assays were designed to selector sites located within LD blocks outside of the hotspot. Where possible, ASPs used by Webb *et al.* (2008) were used, as they had already been proven efficient and specific for crossover isolation (Webb *et al.* 2008). Universal primers were placed outside of the hotspot interval, on the other side of the hotspot in respect to the allele-specific primers, and PCR targets of 5 - 13 kb emerged. Each hotspot was analysed using reciprocal assays, to detect recombinants generated in opposite orientation of recombination. PCR conditions for all half-crossover assays are collected in Appendix (III).

SNPs were given indicator names, referring to their location in kb within the target interval, in addition to their rs numbers. For example SNP rs10783158 T/C at Superhotspot E was located 10.6 kb into the target interval, and is therefore referred to as E10.6 T/C.

3.3.3.1.1 Superhotspot E

Allele-specific PCR amplification of the target interval at Superhotspot E used selector sites E10.6 T/C (rs10783158) and E12.4 +/– (rs10666923) downstream of the hotspot. These have been previously used by (Webb *et al.* 2008). The haplotype with SNP alleles E 10.6 T and E12.4 + was defined as progenitor haplotype A, the other haplotype carrying E10.6 C and E12.4 –, defined as progenitor haplotype B. The most suitable candidate d17 was analysed in the time frame of this work and compared with data gathered by A.J. Jeffreys on d56, analysed in a half-crossover assay using a set of allele-specific primers upstream of the hotspot (A. J. Jeffreys, *unpublished data*).

3.3.3.1.2 Superhotspot F

Two assays were developed to analyse Superhotspot F. One used a set of ASPs proximal, and the other a set of ASPs distal to the hotspot, as illustrated in Figure 3-1. The distal selector SNPs F13.0 T/G (rs 445467) and F13.3 C/T (rs 368789) were used for allele-specific amplification of the haplotypes in the first assay. The haplotype with SNP alleles F10.3 T and F13.3 C was defined as progenitor haplotype A, the other haplotype carrying F10.3 G and F13.3 T defined as progenitor haplotype B. Two men qualified for analysis using these allele-specific primers. One man, d23, was analysed for the study presented in this Chapter. The second man, d28, had been previously analysed for non-exchange gene

conversions by A. J. Jeffreys. This data had indicated a curious incidence of unidirectional biased non-exchange conversion (A. J. Jeffreys, *unpublished data*) and was therefore not included in this survey, but will be discussed in Chapter 4.

A third man, d11, was analysed using the second half-crossover assay. It was developed using a set of allele-specific primers specific to SNPs located on the other side of Super-hotspot F, proximal selector sites F3.8a G/A (rs11612965) and F3.9 T/C (rs10492183). The haplotype carrying alleles F3.8a G and F3.9 T was defined as haplotype A, the other haplotype carrying alleles F3.8a A and F3.9 C defined as haplotype B.

All three men were heterozygous at two central SNPs F6.0 A/T (rs386440) and F6.1 A/G (rs10492181).

3.3.3.1.3 Superhotspot H

The half-crossover assay for Superhotspot H used proximal selector SNPs H2.0 G/A (rs7651825) and H2.5 G/A (rs11716175) for allele-specific amplification. The haplotype with SNP alleles H2.0 A and H2.5 A was defined as progenitor haplotype A, while the haplotype carrying H2.0 G and H2.5 G was defined as haplotype B. Five men could be analysed, however only the two men with the highest SNP density across the hotspot interval (d52 and d25) were analysed in the time-frame of this work.

3.3.3.1.4 Superhotspot K

Distal selector sites were used for allele specific amplification of Superhotspot K, namely K10.9 A/G (rs7829351) and K11.7 C/T (rs766477). SNP alleles G at K10.9 and T at K11.7 are located on haplotype A, while alleles A at K10.9 and C at K11.7 defined haplotype B. Four men were heterozygous at two SNPs directly flanking the hotspot centre and two men (d28 and d90) were analysed in the initial assay presented in this Chapter.

3.3.3.1.5 Superhotspot T

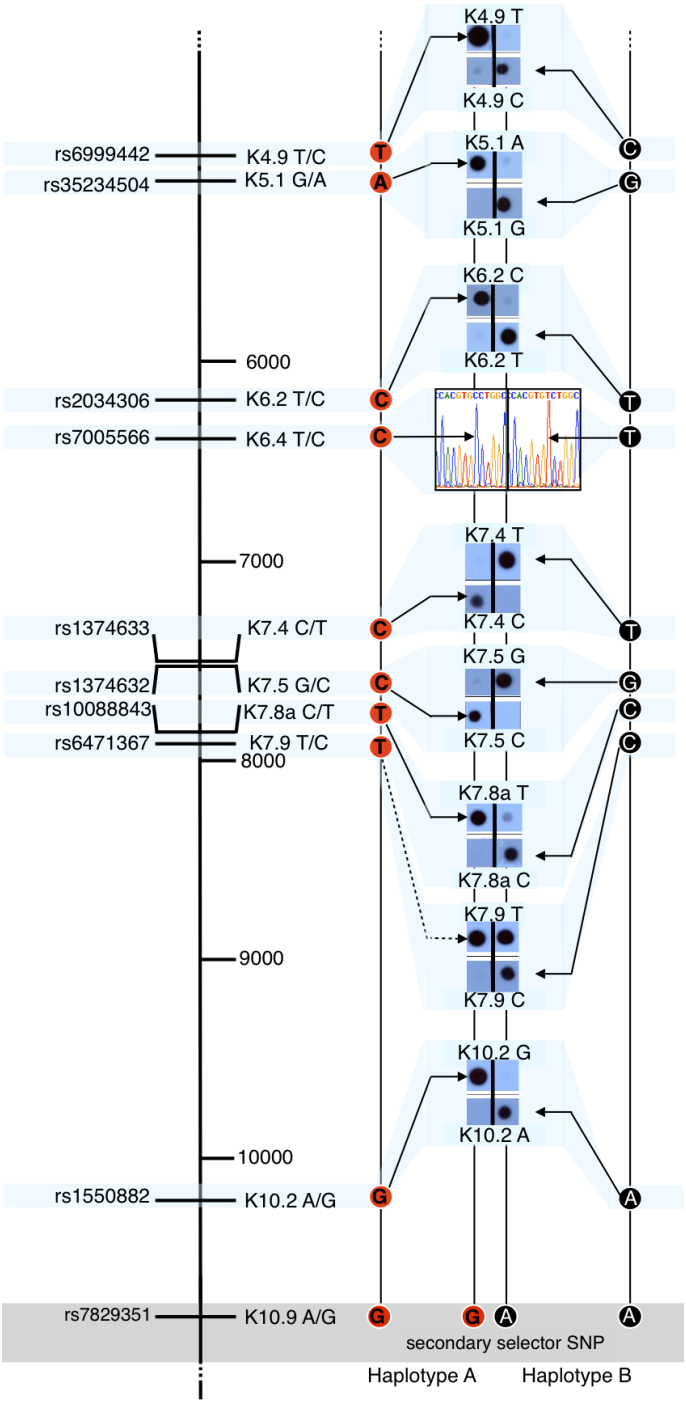
The target interval at Superhotspot T was amplified using allele-specific primers distal to the hotspot, markers T9.6 +/- (rs 11456124) and T10.8 T/C (rs13090158). Haplotype A was defined as the haplotype carrying the T9.6 + and T10.8 C allele, while haplotype B

was defined as carrying the T9.6 - and the T10.8 T allele. Two men were analysed using the same sets of ASPs, d25 and d60, but to allow more markers proximal of the hotspot a different universal primer was used to analyse d60.

3.3.4 Phasing of internal SNPs

Genotyping data had established which SNPs were heterozygous in each candidate donor. To identify which allele was present on which haplotype, internal phase was determined experimentally. An example of how internal phase was determined is shown in Figure 3-2 for man d90, heterozygous at eight SNPs within the Superhotspot K target interval. Haplotypes for all other men analysed at each of the Superhotspots are collected in Appendix V.

Figure 3-2: Phasing of internal markers; an example for d90 at Superhotspot K



Left: Horizontal ticks represent local coordinates within the target interval (in bp). The location of SNPs within the target interval are shaded in blue, and were characterised by their rs numbers as well as indicator name, referring to their location within the target interval (in kb).
Right: PCR products of separated haplotypes were dot-blotted next to each other. The presence of alleles was then tested by ASO hybridisation. The tested allele is written either above or below the image to aid readability. When an ASO hybridised to a pool, it became visible as black circle in the autoradiograph negative image. Haplotypes were deduced from the positive signals (arrows), and build accordingly. Alleles associated with specific haplotypes are arbitrarily displayed in red for haplotype A or black for haplotype B. SNP K6.4T/C was identified through haplotype re-sequencing of the hotspot centre.

As shown in Figure 3-2, most ASOs were specific, meaning that they hybridised to either haplotype A or B and therefore allowed unambiguous phasing of internal markers, though rarely background binding of the occasional ASO (K7.8a T) or unspecific ASOs (K7.9 T) were observed (Figure 3-2). The centre of the hotspot was re-sequenced on each of the separated haplotypes, identifying and simultaneously determining the phase of the additional SNP K6.4 T/C (rs7005566).

3.3.5 Recombinant detection and criteria for scoring crossovers and non-exchange conversions

To test for the presence of recombined markers, PCR pools that were selectively amplified using primers specific to one haplotype were hybridised with ASOs specific to the non-selected haplotype. PCR pools positive for markers of the non-selected haplotype indicated the presence of at least one recombinant molecule within this pool as illustrated in Figure 3-3. In this example pool sizes of 25 or 45 amplifiable sperm DNA molecules of the correct haplotype were used, which are also referred to as haploid genome equivalents (HGEs).

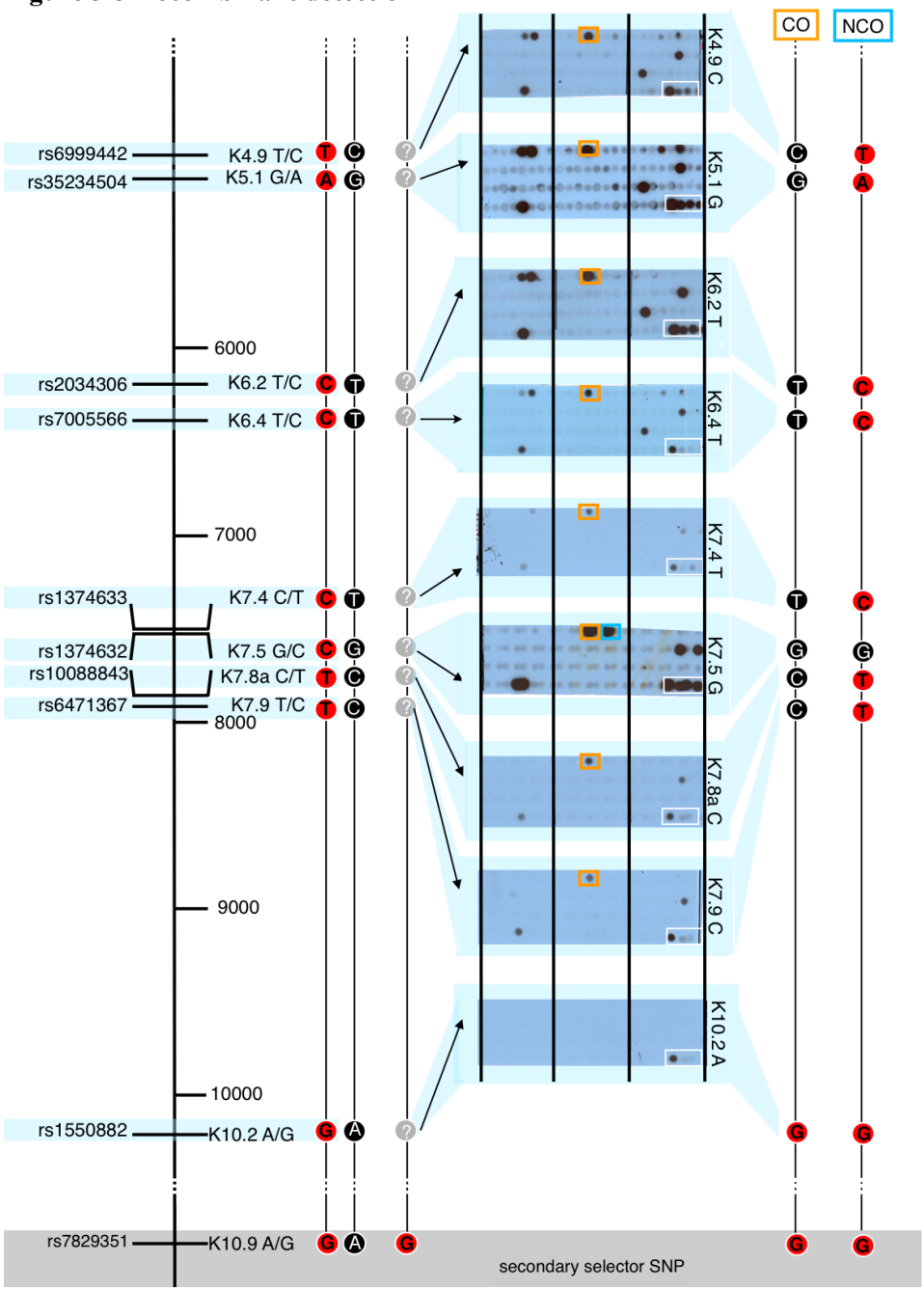
Signal strengths were compared to positive controls containing 10%, 3%, 1%, and 0.3% of PCR product of the opposite haplotype. Mapping exchanged SNPs allowed non-exchange conversions and crossovers to be distinguished. Assuming that PCR amplification of all molecules was uniform, a recombinant molecule present in a pool containing PCR products of 24-44 progenitor molecules should return a hybridisation signal of 2-4%.

Both types of recombinant molecules were successfully detected in all assays just as shown for d90 in Figure 3-3. Signals of between 1% and 10% were generally observed for crossover molecules.

3.3.5.1 Crossover detection

As ASOs will hybridise to more than one SNP, crossover detection is a self-validating process. The crossover breakpoint cannot be determined precisely, but is located within a breakpoint interval - between the last marker positive for recombinant alleles and the first progenitor marker following the stretch of positive markers (between marker K7.9 C and marker K10.2 A in the example illustrated in Figure 3-3).

Figure 3-3 Recombinant detection



Left: Nine heterozygous SNPs are located within the target interval (vertical lines). They are identified by their rs number as well as indicator name. Their location into the target interval (in bp) is indicated on the ruler (as in Figure 3-2).
Right: Assay in orientation A. Dot-blot membranes contained 92 pools of secondary PCR products amplified with ASPs specific to haplotype A (red). Four additional pools were included as hybridisation-signal controls (boxed in white), these contain (from left to right) 10%, 3%, 1%, and 0.3% of PCR product amplified with ASPs specific to haplotype B. Membranes were hybridised with ASOs specific to haplotype B, written to the right of the autoradiograph. Autoradiograph images after hybridisation with each SNP within the target interval are shown. Orange and blue boxes frame hybridisation signals to a crossover molecule as well as a non-exchange gene conversion molecule, respectively; the inferred recombinant structure is illustrated on the right.

3.3.5.2 Criteria for scoring non-exchange conversions

Pools containing non-exchange conversion molecules should by definition have positive hybridisation signals to SNPs within the hotspot interval, but not at flanking markers. Typically, hybridisation signals to one or two adjacent SNPs were observed (see example in Figure 3-3). Pools with positive signals to two adjacent SNPs were classified as double site conversions, as the signals validate each other. However, single site exchanges could potentially be the result of an early PCR misincorporation in any of the progenitor molecules also present in a PCR pool. To limit the scoring of such events as non-exchange gene conversions, only pools with a comparable signal-strength to a crossover molecule on the same membrane, typically of 1-10%, were therefore scored as single site non-exchange gene conversion (blue box in Figure 3-3). However, even with a lower threshold of 1% a single site event could still be mis-scored as non-exchange conversion, even though it arose from a PCR misincorporation. It is therefore necessary to estimate how common such detectable misincorporations would be.

PCR misincorporations resulting in a signal strength of ≥1%, in a pool size containing 40 amplifiable molecules, would require that the changed allele was present in ~40% of amplified molecules from the single molecule that gave rise to this misincorporation. The expected rate of misincorporations at this level was estimated using extensive data on misincorporations rates gathered by Neumann and Jeffreys (2010). They observed misincorporation-levels of at least 40% in 1 base per 20,000 bp sequenced DNA of single molecules amplified using the same PCR conditions as those used here (Neumann *et al.* 2010) (A.J. Jeffreys *unpublished data*). Most of these changes were transitions and must have occurred during the first few cycles of PCR. Any later and the level of misincorporations would be much lower. Single-site exchanges giving a hybridisation signal above 1%, but resulting from PCR misincorporation will therefore occur within this survey. However they will be rare. At this level the chance of a detectable misincorporation at the chosen SNP site corresponds to 0.005% in a pool containing 40 molecules. Caution was therefore applied when single-site non-exchange conversions were observed only once or twice per survey over multiple plates.

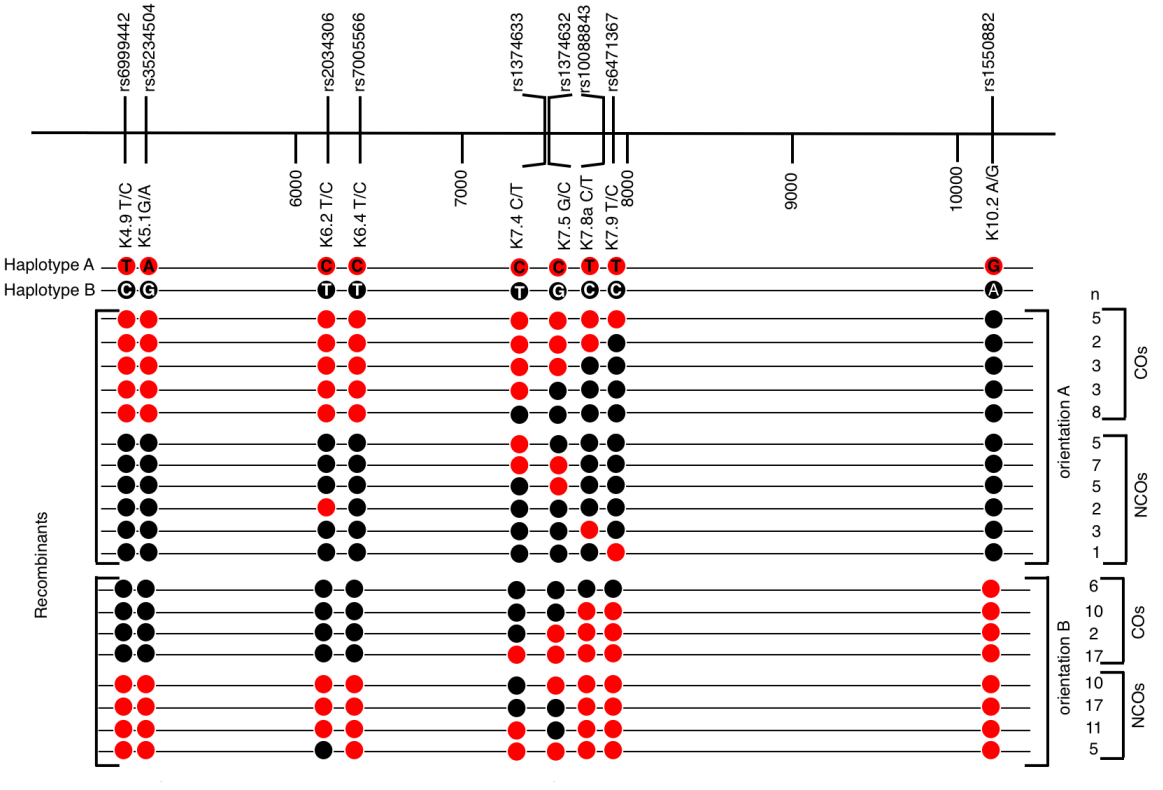
An additional PCR error could be introduced by progenitor "bleed-through", which can occur when allele-specific primers bind to the wrong haplotype. However, these are

detectable as pools entirely positive for markers from the opposite haplotype, typically with hybridisation signals above the maximum expected 10%. Two such pools were detected within the given dataset and were subsequently excluded.

3.3.6 COs and non-exchange gene conversions were detected at each Superhotspot

Each man and each hotspot was analysed using reciprocal assays. In both assays recombinants were scored using the criteria outlined in 3.3.5. Both crossovers, as well as non-exchange gene conversions, were successfully isolated in each of the half-crossover assays. A schematic representation of recombinant molecules detected in reciprocal assays for d90 at Superhotspot K is shown in Figure 3-4.

Figure 3-4 Recombinant molecules detected at hotspot K in sperm DNA from man d90



Top panel: The relative location of SNP markers is shown, with allele names at the bottom, and rs-numbers at the top. Vertical ticks represent local position of each marker in the target interval, in bp. Selector sites for this half-crossover assay lay to the right of the shown interval.
Bottom panel: The types of recombinants detected in this assay are shown, with SNPs originating from haplotype A and B indicated by red and black respectively. In each reciprocal assay 368 pools containing 40 amplifiable molecules of the selected haplotype were screened, 14720 molecules in total. Numbers of pools containing each type of recombinant are shown on the right.

All crossovers detected in both reciprocal assays for d90 at Superhotspot K were simple, meaning that crossover breakpoints mapped to a single interval between heterozygous SNPs (Figure 3-4). Similar illustrations of recombinants detected for each man analysed at each Superhotspot are collected in Appendix V, and most crossover molecules detected in all assays were simple. However 2.8% of crossovers were complex molecules that displayed more than one breakpoint. Most half-crossover assays yielded between 20-56 crossovers per man, with two notable exceptions. Very few crossovers were detected in comparable datasets for man d23 at Superhotspot F and man d60 at Superhotspot T, with assays yielding only seven and two crossover molecules, respectively (Appendix V).

In contrast to crossovers, only non-crossovers that resulted in gene conversion were detectable, as only non-crossover tracts that span at least one marker can result in conversion of this marker. Nevertheless non-exchange conversion molecules were detected at every hotspot. They involved at most only two SNPs and clustered near the hotspot centre, as for example shown in Figure 3-4. Most non-exchange conversions were single site events, but at hotspots that had two markers located directly at the centre, double-site conversions were also observed. The men for whom only very few crossovers were detected, equally returned only very few non-exchange conversion molecules, one in d23 and seven in man d60. For all other men between 15 and 66 pools were positive for non-exchange conversions.

3.3.7 Determining recombination frequencies

To correct for the possibility of several recombinant molecules present in the same pool masking one another, recombination frequencies were determined by Poisson-approximation. Poisson approximation on data from several pool sizes uses a maximum-likelihood approach that not only determined the most likely number of recombinants in a given pool size but also simultaneously tested whether the occurrence of recombinants was proportional to pool sizes. Minor deviation from proportionality was only occasionally observed, and was always statistically insignificant.

As crossover molecules present within the same pool can mask non-exchange conversion molecules, non-exchange conversion frequencies were determined after excluding pools that contained a crossover molecule spanning the respective SNP or SNPs. Overall

conversion frequencies were determined by treating each single-site and each multiple site exchange as a single non-exchange gene conversion event, respectively. Additionally, the conversion frequency at a given SNP was also determined. Here multiple-site conversions were counted for each of the SNPs they involved in addition to single site conversions involving the respective SNP.

Crossover numbers within an interval were at most corrected by a factor of 1.15, and numbers of non-exchange gene conversions at a given marker were, at most, corrected by a factor of 1.25.

3.4 Recombination Superhotspots are active in crossover and non-crossover

3.4.1 Crossover frequency and distribution

Crossover frequencies and the distribution of crossover breakpoints have already been established for each Superhotspot using full-crossover assays (Webb *et al.* 2008). These parameters were also determined for each of the men analysed using the half-crossover assays, and then compared with the previous results.

3.4.1.1 Crossover frequencies

Crossover frequencies for each man were determined from the Poisson-corrected number of crossover molecules. And since crossovers are reciprocal large-scale events involving several markers, and all are detectable, numbers of crossover molecules detected in opposite orientations should not be significantly different from each other. Differences between numbers of crossovers detected in reciprocal orientation would therefore point to differences between reciprocal assays that affect recombinant detection. Numbers of crossovers detected in reciprocal assays are summarised in Table 3-2.

Table 3-2 Comparing numbers of crossovers detected in reciprocal assays

hotspot	man	Number of HGEs per orientation	A crossover		B crossover		A/B difference
			Poisson no	Frequency (%)	Poisson no	Frequency (%)	
E	d17	7360	13	0.18	21	0.28	*P*=0.170
E	d56	2290	15	0.73	15	0.75	*P*=1
F	d11	8280	28	0.34	19	0.22	*P*=0.189
F	d23	4600	2	0.04	5	0.11	*P*=0.257
H	d25	12420	15	0.12	18	0.14	*P*=0.612
H	d52	11040	18	0.16	27	0.25	*P*=0.180
K	d28	6980	9	0.13	11	0.16	*P*=0.655
K	d90	14720	37	0.25	22	0.15	*P*=0.051
T	d25	3496	17	0.48	15	0.42	*P*=0.724
T	d60	4600	2	0.04	-	-	*P*=0.157

P-values were determined using a t-test for Poisson means (Poisson-corrected numbers of crossovers) comparing crossover frequencies averaged from reciprocal half-crossover assays.

In most cases no significant variation between numbers of crossovers detected in reciprocal orientations was observed (Table 3-2). The only exception was marginally significant variation between reciprocal crossover frequencies in man d90 at Superhotspot K. This man was analysed with the same assay as d28, therefore crossover data was pooled from both men, which revealed that frequencies detected in reciprocal assays were not significantly different from each other overall (t-test comparing two Poisson means, p=0.144). Recombination frequencies obtained in reciprocal assays can therefore be compared at all hotspots. Data from reciprocal assays was therefore subsequently averaged to determine overall crossover frequencies.

To test whether the half-crossover assays gave reliable results. These were then compared to frequencies obtained previously in full-crossover assays at the same hotspots, which are summarised in Table 3-3.

Table 3-3 Overview of crossover frequencies at five recombination Superhotspots

Superhotspot name	Crossover frequencies per sperm (%) in crossover assay	man	Crossover frequency per sperm (%)	Deviation from published frequency
E	0.45 0.61 0.41	d17	0.23	Yes ***(*P*=0.00001)
		d56[1]	[1]0.74	[1]No (*P*=0.367)
F	0.81 1.00 1.09	d11	0.28	Yes ***(P<0.0001)
		d23	0.08	Yes ***(P<0.0001)
H	0.13 0.14 0.13	d25	0.13	No (*P*=1)
		d52	0.20	Yes *(*P*=0.015)
K	0.19 0.21 0.26	d28	0.15	No (*P*=0.115)
		d90	0.20	No (*P*=0.585)
T	1.28 1.33 0.74	d25	0.46	Yes ***(*P*=0.00001)
		d60	0.02	Yes ***(P<0.00001)

Analysed men indicated by donor number. Crossover frequencies in % per sperm were based on Poisson-corrected numbers, [1]Additional data from a half-crossover assay on d56 at Superhotspot E performed by A.J. Jeffreys (*unpublished data*). P-values are two-tailed exact binomial probabilities.

Overall crossover frequencies were much more variable between men compared to the observations of (Webb *et al.* 2008) (Table 3-3) with only Superhotspot K and one of the men analysed at Superhotspot H as well as Superhotspot E showing good agreement.

On the one hand, a moderate but significant two- to fourfold reduction in crossover frequencies was observed for d17 at Superhotspot E, for d11 at Superhotspot F and for d25 at Superhotspot T. On the other hand, d52 at Superhotspot H displayed a small, but significantly higher crossover frequency than observed by (Webb *et al.* 2008).

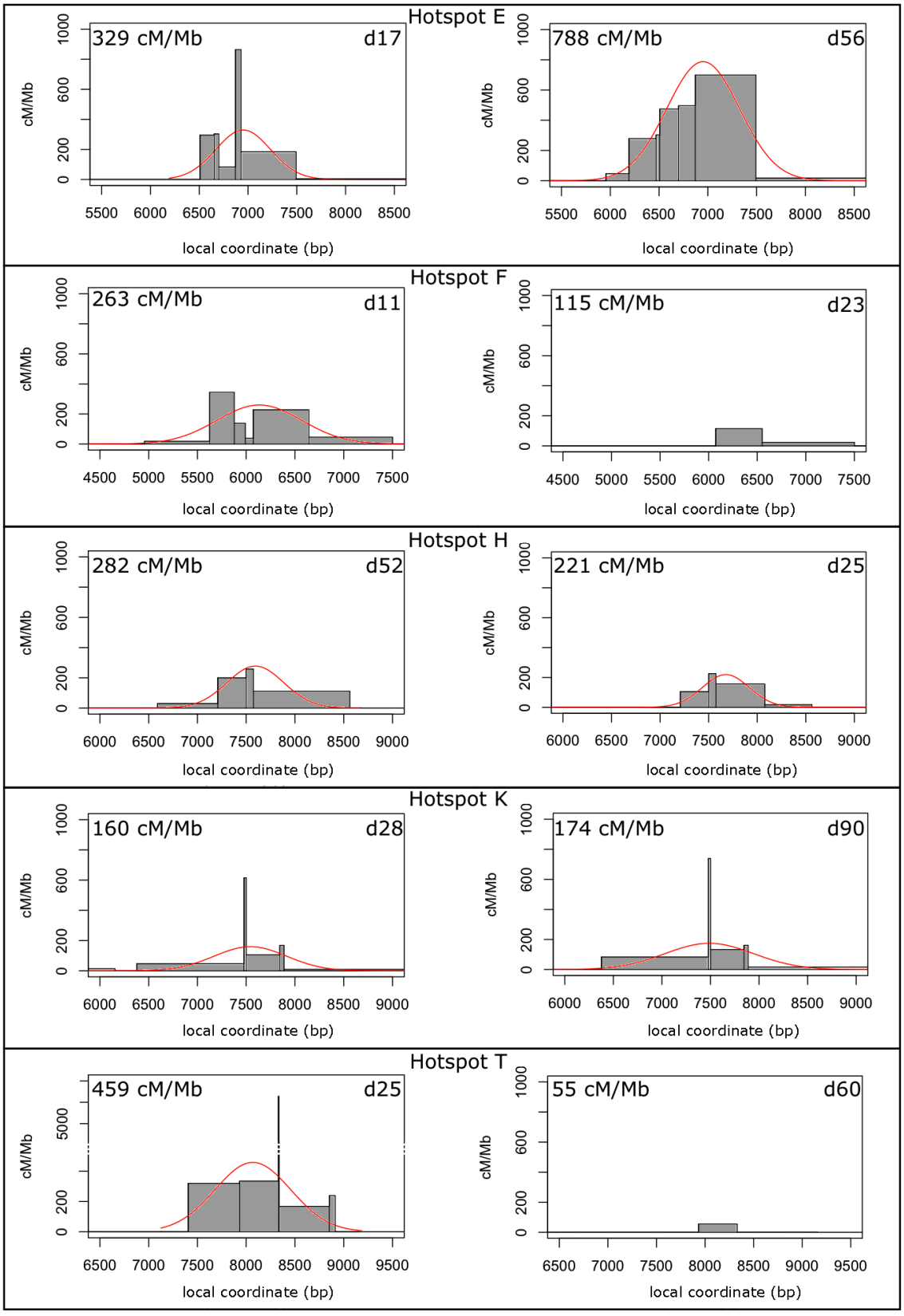
Instances of much more marked reduction of crossover frequencies were also observed, d23 at Superhotspot F showed a more than 12-fold reduced crossover frequency compared to the previously determined crossover frequency, additionally d60 at Superhotspot T showed a >50-fold reduction of crossover frequency compared to those obtained previously (Table 3-3). This phenomenon will be investigated further in Chapter Five.

3.4.1.2 Crossover distribution

The morphology of a crossover hotspot is defined by the distribution of crossover resolution points, which are harboured by crossover breakpoint intervals. All sperm crossover hotspots analysed to date had crossover breakpoint distributions consistent with a normal-distribution model, as summarised in (Jeffreys *et al.* 2004). Consistent with this model, hotspot peak recombination frequency and width can be estimated from a least-squares best-fit analysis of a normal-distribution of breakpoints (Jeffreys *et al.* 2001).

Hotspot morphologies were determined for all of the men for whom ample numbers of recombinants were detected, excluding d23 at Superhotspot F and d60 at Superhotspot T. Firstly the Poisson-corrected number of crossovers exchanging within each interval was determined. To do so, complex crossover molecules were assigned to breakpoint intervals proportionally i.e. a complex crossover molecule with two breakpoints would be counted as 0.5 crossover molecule for each of the breakpoint intervals. Crossover breakpoint distributions were displayed as block histograms, whose heights represent the recombination activity in cM/Mb, and whose widths represent the width of the breakpoint interval. Distributions of crossover breakpoints were then simulated using a bespoke program written by A. J. Jeffreys that determined the best-fit normal distribution using the least-squares approach. Crossover distributions and their corresponding simulated normal distributions are collected in Figure 3-5 for all of the analysed men.

Figure 3-5 Distribution of crossover resolution points



Crossover distributions are shown for each Superhotspot with their name shown on top of the panel. Breakpoint distributions are shown for each donor separately, with donor number as well as peak recombination frequency in cM/Mb shown within each graph. Simulated least-squares best-fit normal distributions are shown as red lines.

As summarised in Figure 3-5, crossover breakpoint distributions obtained in the half-crossover assay display a typical pattern of the peak crossover frequency at the centre and bi-directional decay away from the centre, and centre locations correspond well to the Superhotspot parameters established previously (Webb *et al.* 2008) (A. J. Jeffreys, *unpublished data*). Superhotspots also display typical widths of between 1.0 kb and 1.8 kb, defined by 95% widths of the least-squares best-fit normal distributions.

However, a few crossover breakpoints occurred within very small intervals between closely spaced SNPs at the centres of hotspots E, T and K, which resulted in steep, almost needle-like appearance, in cM/Mb. The observed number of breaks was tested for significant difference from the expected number of breaks within the same interval in the simulated normal distribution as summarised in Table 3-4.

Table 3-4 crossover breakpoints occurring within a given interval

hotspot	Donor	HGEs analysed	interval (bp)	Observed	Expected	binomial probability
E	17	14720	57	7	2.71	*P*=0.042
K	28	13960	24	2	0.73	*P*=0.333
	90	29440	24	5	1.27	*P*=0.019
T	25	6992	8	3	0.17	*P*=0.0014

Observed raw (not Poisson-corrected) numbers of crossover breakpoints within a given interval and expected number of crossovers breaking within the same interval in the least-squares best-fit normal distribution were compared. P-values are Poisson-approximated two-tailed binomial probabilities.

Significantly more crossover breakpoints than were expected from the simulated crossover breakpoint normal distribution, occurred within a short interval at the centre of the crossover hotspot observed for d17 at Superhotspot E as well as for d90 at Superhotspot K and d25 at Superhotspot T. Possible reasons for this observation will be discussed (3.6.1.2).

3.4.2 Non-exchange conversion frequency and distribution

3.4.2.1 Non-exchange conversion frequency

Non-exchange conversions were detected in addition to crossovers. Conversion frequencies were determined by Poisson-approximation on non-exchange conversion molecules detected in both reciprocal half-crossover assays (summarised in Table 3-5.)

Table 3-5 non-exchange conversion frequencies at five recombination Superhotspots

Superhotspot name	man	Crossover frequency per sperm (%)	non-exchange conversion frequency per sperm (%)	Combined recombination frequency per sperm (%)
E	d17	0.25	0.18	0.43
	d56[1]	0.701	0.471	1.16
F	d11	0.28	0.13	0.41
	d23	0.07	0.01	0.08
H	d25	0.13	0.07	0.20
	d52	0.21	0.07	0.28
K	d28	0.14	0.24	0.37
	d90	0.21	0.28	0.49
T	d25	0.46	0.23	0.68
	d60	0.02	0.01	0.03

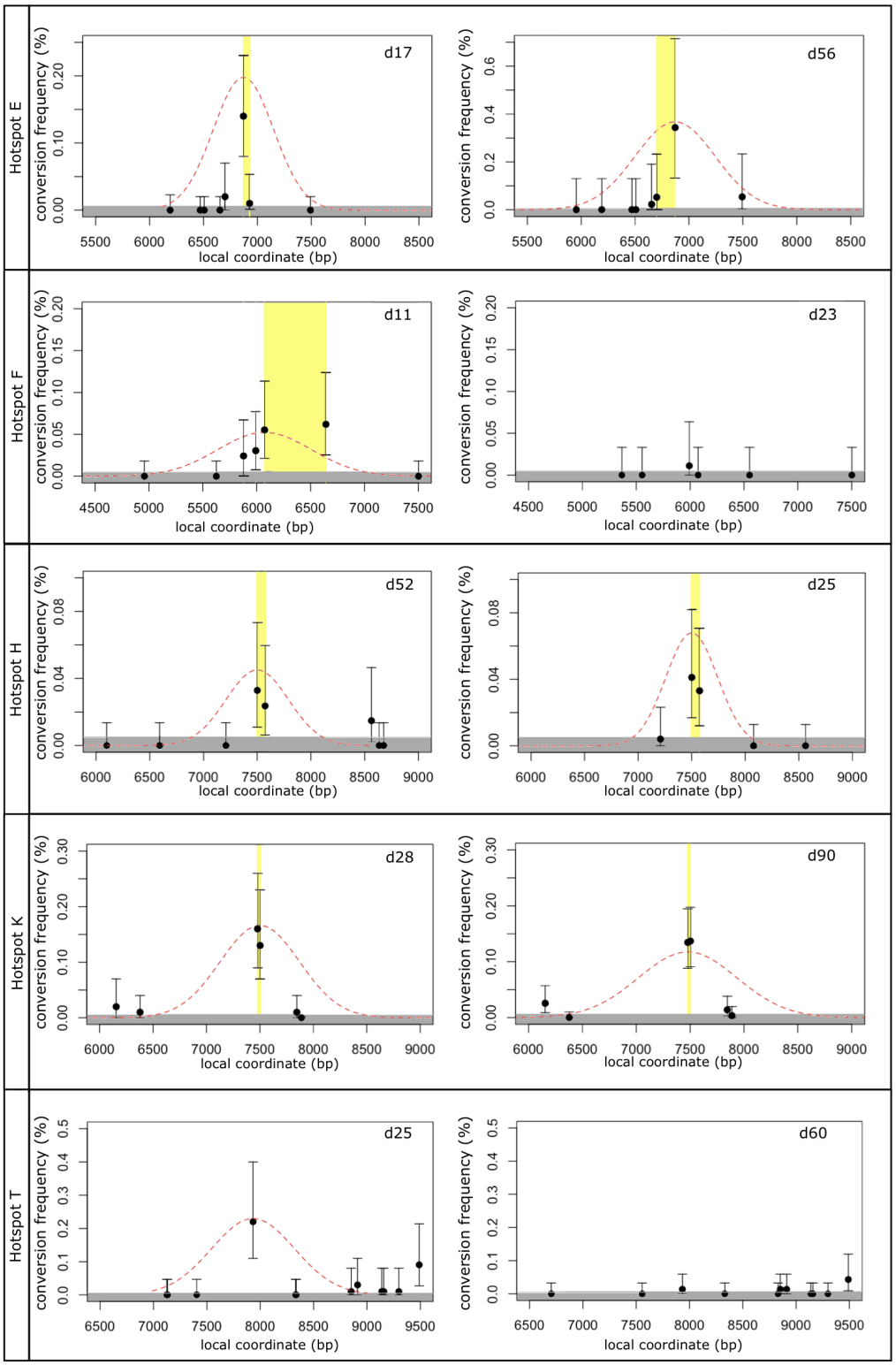
Detected crossover frequencies and detectable non-exchange conversion frequencies in % per sperm are based on Poisson-corrected numbers.[1]Additional data from a half-crossover assay on d56 at Superhotspot E performed by A.J. Jeffreys was included (A.J. Jeffreys, *unpublished data*).

Most men displayed high frequencies of detectable non-exchange conversions in addition to crossovers, occurring at 0.07 – 0.47% per sperm. Those men with lower crossover frequencies also displayed low non-exchange conversion frequencies.

3.4.2.2 Non-exchange conversion distribution

To display non-exchange gene conversion profiles at each Superhotspot, conversion frequencies per marker were plotted across each of the hotspot intervals. These profiles display the conversion frequency per SNP. As described previously these were established by not only counting single-site events occurring per SNP but also adding multiple-site events for each of the involved SNPs (3.3.7). Conversion profiles for all of the men analysed at each Superhotspot are collected in Figure 3-6.

Figure 3-6 non-exchange conversion profiles



Non-exchange gene conversion frequencies per sperm, and confidence intervals were determined by Poisson-approximation for each marker. The co-conversion zone is highlighted in yellow. Least-squares best-fit normal distribution of crossovers at each hotspot are shown by hashed red line, peak location was corrected to pass through the location of the most central SNP. Expected background rate of single-base changes indistinguishable from early PCR misincorporation is shaded in grey (see section 3.3.5).

Non-exchange conversion profiles displayed a distinct peak of conversion with the central SNP(s) converted at the highest frequencies and markers converted less frequently the further away they are from the centre (Figure 3-6). One notable exception is d11 at Superhotspot F, who shows the highest conversion frequency at a marker not located at the centre of the hotspot, as defined by the peak of the crossover distribution.

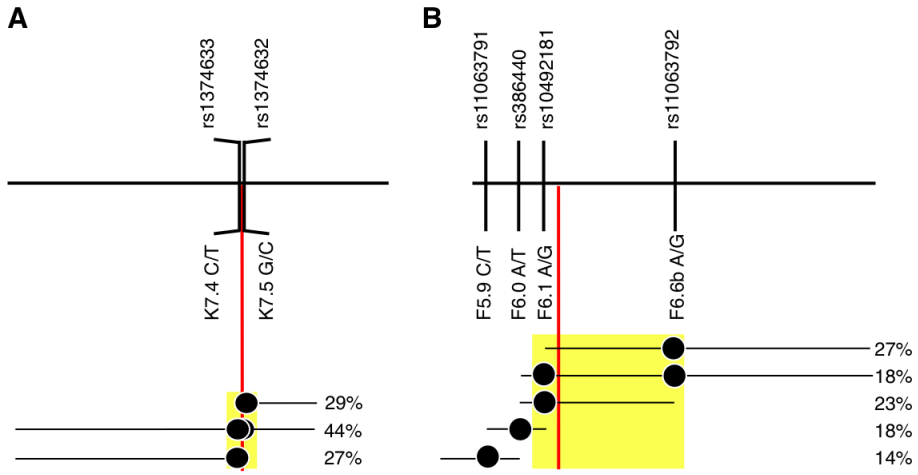
At most hotspots the peak non-exchange gene conversion frequency was detected mainly at the centre of the crossover distribution, and the distribution of non-exchange gene conversion events appeared to be narrower than the crossover resolution interval.

Markers outside of the crossover resolution interval were rarely exchanged (Figure 3-6). Most are single-site transitional exchanges occurring at a frequency consistent with the expected PCR misincorporation rate of 0.005% (the grey zone in Figure 3-6). Two transitional SNPs, namely K6.2 at Superhotspot K as well as T9.5 located 1.5 kb away from the centre of Superhotspot T, were changed at a frequency significantly elevated above this threshold in both men analysed at a given hotspot (Poisson-approximated binomial probability, p=0.00083, and p<0.000001, respectively). Possible explanations for this observation will be discussed.

3.4.2.2.1 Co-conversion tracts

At the centres of the hotspots, and where two suitable markers were located closely together, co-conversion tracts were observed. These co-conversion zones always incorporated markers located at the centre of the crossover breakpoint distribution. Man d11 at Superhotspot F had non-exchange co-conversion tracts that never included the two markers located closest to hotspot centre, but instead a marker off-centre. Secondly, a wider distribution of non-exchange gene conversions was observed compared to the other hotspots with co-conversion tracts always incorporating only those markers closest to the centre. To illustrate this curious pattern in more detail non-exchange co-conversions detected at Superhotspot F are compared with those detected at Superhotspot K in Figure 3-7.

Figure 3-7 co-conversion zones at Superhotspot K and Superhotspot F



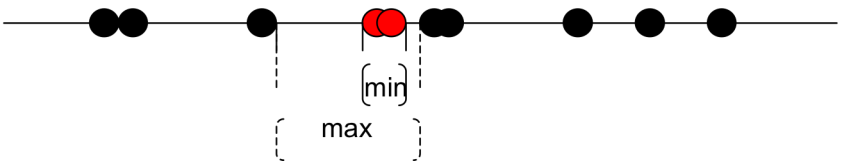
Gene conversion molecules involving the most central SNPs (A) for donor 90 at Superhotspot K and (B) for d11 at Superhotspot F. Circles represent SNPs, with their indicator name and rs numbers shown above. Horizontal lines represent sequence distance to flanking markers (maximum conversion tract lengths). The co-conversion zone is underlined in yellow and the location of the hotspot centre, defined as the peak of the crossover breakpoint distribution (3.6.1.2) indicated by vertical red line.

At most hotspots, Co-conversion tracts involved the markers that were located closest to the hotspot centre, as for example illustrated for d90 at Superhotspot K (Figure 3-7). In contrast at Superhotspot F, the markers located at the closest distance to the hotspot centre, markers F6.0 and F6.1, were not involved in co-conversion tracts with each other, but instead marker F6.6b, located off-centre was found in all co-conversion tracts (Figure 3-7). Despite the off-centre location of F6.6b, co-conversion tracts nevertheless spanned the hotspot centre.

3.4.2.2.2 Non-exchange conversion tract length

The length of non-crossover tracts cannot be measured, but limited information can be gathered by estimating minimal and maximal conversion tract lengths, as illustrated in Figure 3-8.

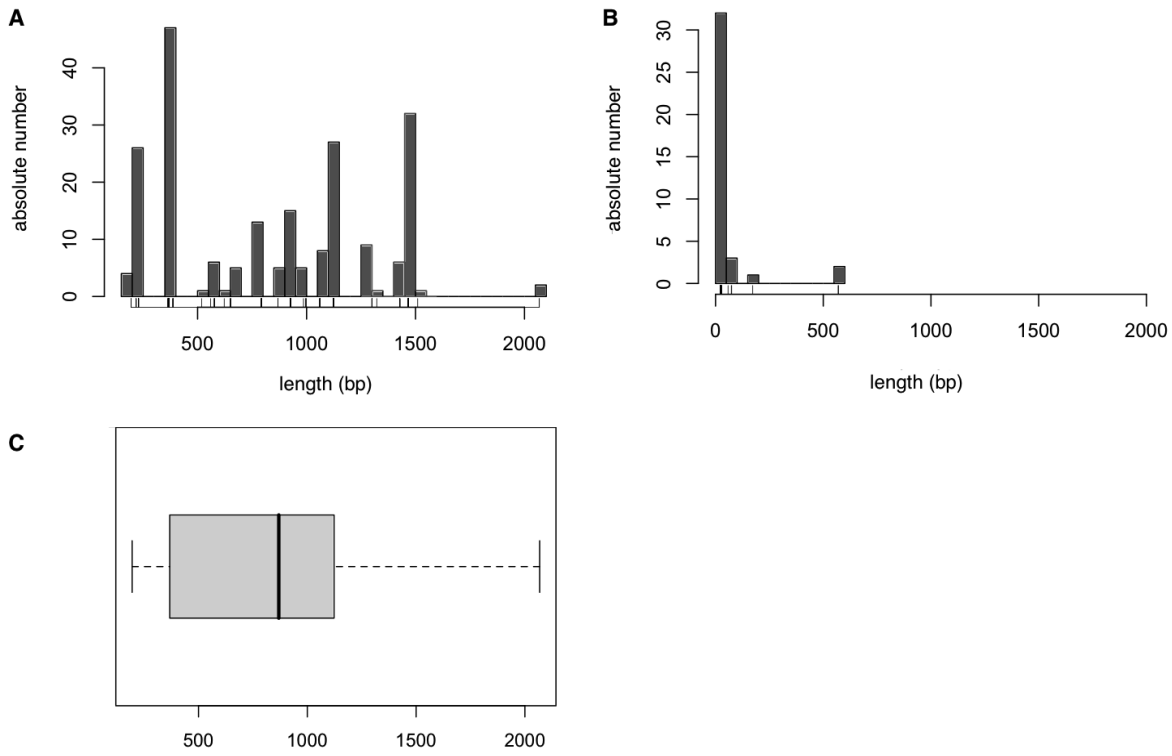
Figure 3-8 Estimating minimal and maximal non-exchange conversion tract lengths



Maximal conversion tract lengths are defined as sequence distance between flanking, non-converted markers (hashed vertical lines). Minimal conversion tract lengths of multiple-site conversions encompass the interval from the first to the last converted marker within a given non-exchange gene conversion tract (solid vertical line).

For each conversion event the maximal conversion tract length was defined as the distance between the non-converted markers flanking the converted marker. Single site conversions always have a minimum tract length of 1 bp. For co-conversions, which included more than one marker, the minimal conversion tract length was also determined - as the distance between the first and last markers of the converted tract, including the markers itself (Figure 3-8). The detection of non-exchange gene conversions and the accuracy of minimum and maximum conversion tract length estimates increased with higher polymorphism density. Estimations of maximum and minimum tract lengths are collected in Figure 3-9.

Figure 3-9 Conversion tract lengths estimations



Histograms were generated based on the counts of maximum and minimum tract lengths that were collected in lengths interval bins, the hashes in the bottom ruler within each histogram show their estimated lengths. (A) Absolute number of maximum non-exchange conversion tracts binned by length into 50 bp size intervals; (B) Absolute numbers of minimum co-conversion tracts (>1 bp) observed at the hotspots, collected in 10 bp bin sizes. (C) box-plot of the maximum conversion tract lengths data set

The maximum conversion tract length was dependent on the SNP density not only at the hotspot centre, but also away from the centre. The shortest maximum tract observed was 195 bp long. The mean maximum tract length was 833 bp ± 32 bp, which is in good agreement with the median maximum tract length of 869 bp. The longest maximum tract length observed was 2068 bp. This tract belonged to a conversion detected at Superhot-spot E from d56, where there were no additional markers between marker E7.5 (rs1371745) and marker E8.9 (rs2581558), located outside of the hotspot.

The maximum tract length for many single-site conversions was very long as the marker density was limited, and markers outside of the centre interval were often located far away from the conversionally active central SNP that converted at peak frequency. Hotspots with a high SNP density in the centre, but with additional markers located far away from the centre gave long maximum conversion tract lengths. For example the double site

conversions involving the most central SNPs at Superhotspot K have a maximum tract lengths of 1467 bp and in contrast a minimum tract length of merely 24 bp.

The longest minimum tract length for a double-site conversion was 569 bp, which was observed at Superhotspot F for d11, who had shown a very unusual distribution of non-exchange conversion events. Here co-conversion tracts involved marker F6.6b (rs406430) located off-centre, but within the 95% width of the crossover breakpoint distribution. Overall, the minimum tract lengths of co-conversion were short, with a mean minimum co-conversion tract length of 60±20 bp.

3.5 The relationship between non-exchange conversion and crossover

Non-exchange conversion molecules were detected at all hotspots and appeared to arise at frequencies of the same magnitude as crossovers. The men with very low crossover frequencies also displayed low non-exchange conversion frequencies. Crossover and non-exchange conversion frequencies were compared between men and between hotspots by comparing CON:CO ratios, which are collected in Table 3-6. Two different approaches were used to determine CON:CO ratios. Firstly the overall non-exchange conversion: crossover ratio was determined from all events detected within a given hotspot. Secondly, to decrease the effect of greater non-exchange conversion detection ability at hotspots with two or more central SNPs peak conversion frequencies were determined. Conversions to crossover ratios were not determined for the men with low numbers of both crossover and non-exchange conversion, as conversion events arose at a frequency indistinguishable from PCR misincorporations. Conversion to crossover ratios determined from both the peak frequency and the overall non-exchange conversion frequency are collected in Table 3-6.

Table 3-6 Conversion to crossover ratios at five recombination Superhotspots

Superhot-spot name	man	Crossover frequency (%)	Detectable non-exchange conversion frequency (%)	CON peak transmission (%)	CON:CO ratio	Peak CON:CO ratio
E	d17	0.25	0.18	0.15	0.72	0.60
	d56[1]	0.70[1]	0.47[1]	0.36[1]	0.67	0.51
F	d11	0.28	0.13	0.05	0.46	0.17
	d23	0.07	0.01	N/A	N/A	N/A
H	d25	0.13	0.07	0.04	0.54	0.31
	d52	0.21	0.07	0.03	0.33	0.14
K	d28	0.14	0.24	0.07	1.71	0.50
	d90	0.21	0.28	0.14	1.33	0.67
T	d25	0.46	0.23	0.16	0.50	0.34
	d60	0.02	0.01	N/A	N/A	N/A

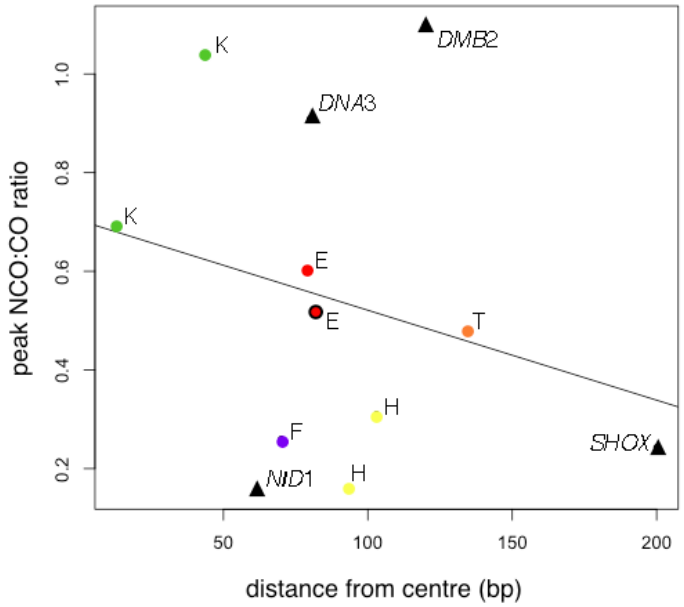
Detected crossover frequencies and the detectable non-exchange conversion frequency per sperm were based on Poisson-corrected numbers. [1]analysis based on a half-crossover assay performed by A.J. Jeffreys (*unpublished data*). N/A = not applicable as frequencies were below expected PCR misincorporation rate.

Across all men that had shown high crossover frequencies, conversions arose at 33% -171 % of this frequency. When compared between all men across all hotspots non-exchange conversion to crossover ratios were significantly different (Fisher-exact test on 8x2 contingency table, 7d.f P=0.00005, after Bonferroni correction for the 8 men). However, their difference appeared greater between hotspots rather than between men at a given hotspot. CON:CO ratios did not vary significantly between men at those hotspots where two men could be compared (Fisher-exact two-tailed, Superhotspot E: p=0.284, Superhot-spot H: p=0.251 and Superhotspot K: p=0.384).

In order to compare non-exchange conversion to crossover ratios across hotspots, conver-sion frequencies at the SNP converting at peak frequency were used. As it was based on a single marker only, differences in the ability to detect non-exchange conversion were minimised. If non-crossover forms at the centre of the hotspot then the key variable might be the distance of the central SNP from the hotspot centre. To test this relationship, peak conversion to crossover ratios (Table 3-6) were correlated with the distance of this SNP to the centre of the crossover breakpoint distribution.

The correlation analysis summarised in Figure 3-10 was performed including previously published data on non-exchange conversion to crossover ratios from hotspots *DNA3*, *DMB*2, *SHOX* (Jeffreys and May 2004) and *NID*1 (Jeffreys and Neumann 2005).

Figure 3-10 correlating peak non-crossover frequencies with the distance of the converting marker to the hotspot centre

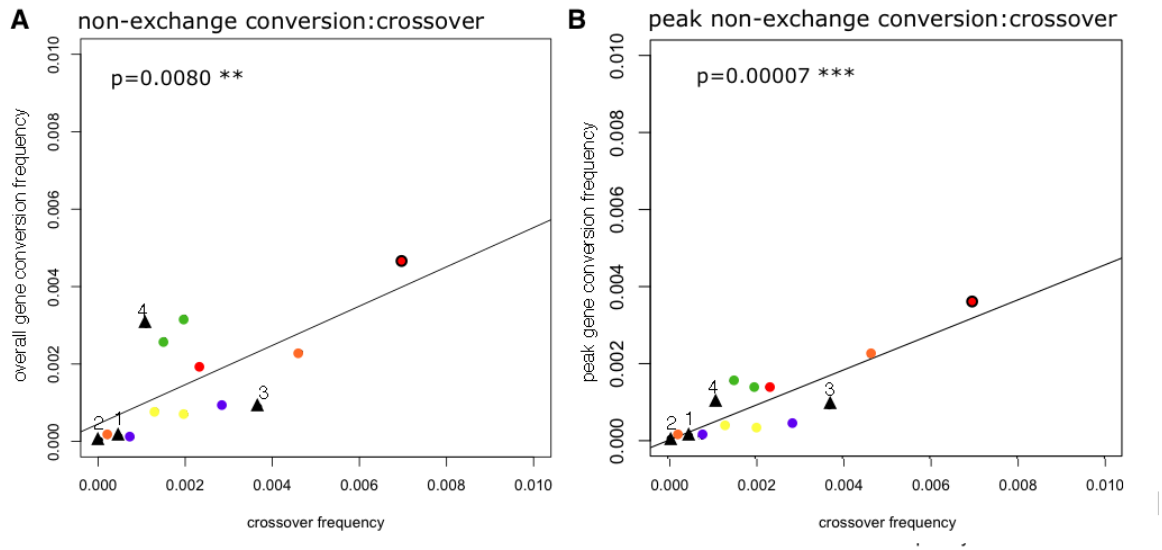


Poisson-corrected crossover and conversion frequencies of all hotspot analysed in this study, as well as published data were tested for the correlation between pCON/Crossover ratios and the distance of the most central SNP to the hotspot centre. Linear regression line plotted with the peak conversion to crossover ratio as a function of the distance to the centre of the hotspot. Hotspots were differentiated by colour: Superhotspot E red, Superhotspot F blue, Superhotspot H yellow, Superhotspot K green, Superhotspot T orange, published data from hotspots *DNA3*, *DMB*2, *SHOX* (Jeffreys and May 2004) and *NID*1 (Jeffreys and Neumann 2005) indicated by triangles. Data included from a half-crossover assay performed by A.J. Jeffreys circled in black.

A weak indication for negative correlation was observed between the peak non-exchange conversion to crossover frequency at a given SNP and the distance of this SNP to the hotspot centre (R= -0.26, coefficient of determination R^2=0.07). This observation is not significant (F-statistics, *P*=0.419, 10 d.f.) and the observed variation therefore cannot be explained as a function of the distance from the centre of the hotspot.

To analyse whether conversion and crossover frequencies are correlated, frequencies obtained here as well as previous estimates from published sources were both used and correlated using regression analysis (Figure 3-11).

Figure 3-11 non-exchange conversion and crossover frequencies are positively correlated



Poisson-corrected crossover and conversion frequencies of all hotspot analysed in this study, as well as published data from hotspots [1]*DNA3*, [2]*DMB*2, [3]*SHOX* (Jeffreys and May 2004) and [4]*NID*1 (Jeffreys and Neumann 2005), indicated by triangles. Hotspots were differentiated by colour as in Figure 3-10. (A) Least-squares regression line plotted from the observed overall gene conversion frequency as a function of the observed crossover frequency. (B) Least-squares regression line plotted from the observed peak non-exchange conversion frequency as a function of the observed crossover frequency

Crossover and non-exchange conversion frequencies were positively correlated (linear regression, correlation coefficient R=0.64, coefficient of determination R^2=0.41, F-Statistics, *P*=0.008, 12 d.f.). An even stronger direct correlation exists between peak non-exchange conversion and crossover frequencies (correlation coefficient R=0.85, coefficient of determination R^2=0.73, F-statistics *P*=0.00007, 12.d.f.), suggesting that the non-exchange conversion frequency is proportional to crossover frequency at a given hotspot. Across all hotspots the detectable non-exchange conversion peak frequency is on average observed at (54±10)% of the overall crossover frequency.

3.6 Summary and Discussion

3.6.1 Crossover frequencies are variable between men

Half-crossover assays were designed to analyse five recombination Superhotspots, and two men were analysed at each hotspot for crossovers and non-exchange conversions. The crossover frequencies obtained in these assays were not all comparable with previous estimates. Previous analyses had suggested that men show a fairly uniform crossover frequency at Superhotspots (Webb *et al.* 2008). The present data show that statistical variation can exist similar to the variation between men observed at less active hotspots (Jeffreys *et al.* 2004).

Only few recombinants were detected for man d23 at Superhotspot F and man d60 at Superhotspot T, while many pools were positive for recombinant molecules for a second man analysed at each hotspot (see Appendix V). Minor variation in inter-individual recombination frequencies can result from variation in DNA quality and from subtle differences in primer-amplification efficiencies that affect recombinant recovery. For both of the men, the DNA quality was good as it showed robust amplifiability. Additionally, the strength of secondary PCR product was both uniform as well as comparable with those of the men active for recombination at either hotspot, which argues against lower PCR efficiency affecting recombinant recovery and detection.

The observed frequencies, albeit lower than expected, were well above previous estimates of recombination-cold sequences (Jeffreys *et al.* 2001; Jeffreys and Neumann 2005; Neumann and Jeffreys 2006) and elevated above the genome average rate of 0.89 cM/Mb in male meiosis (Gyapay *et al.* 1994).

The strong degree of variation between men was intriguing and implies differences in hotspot regulation between men. *Cis*-acting elements have previously been shown to influence the efficiency of initiation, leading to variation in crossover frequencies between men and to reciprocal crossover asymmetry in men carrying haplotypes of different activity. Strong variation, even as extreme as complete suppression, has been observed at regular human recombination hotspots (Neumann and Jeffreys 2006).

Possible *cis*-influences on hotspot activity will be investigated in Chapter Four. Additionally, a *trans*-acting factor that might influence recombination frequencies is examined in Chapter Five.

3.6.1.1 Authenticity of conversion events and noise of the assay

About 82% of detectable non-exchange conversion molecules were single-site events, and are therefore indistinguishable from base-mutations or PCR misincorporation. However, de-novo base substitution will not have a significant effect, as this rate is in the region of approximately 10^{-8} per base per generation in the male germline (1000 Genomes Project Consortium 2010) while gene conversions were detected at far higher frequencies of 0.01% to 0.47%.

The assay itself relies on faithful PCR amplification. Incorrect estimates of gene conversion through the inability to discriminate PCR errors from genuine events could lead to an overestimation of the detectable amount of gene conversion. The expected level of PCR misincorporation that is indistinguishable from true events was estimated to be 0.005%. This estimate was based on the observed frequency of base-changes in sequenced single-molecules amplified using the same PCR conditions that were used here (Neumann *et al.* 2010) (3.3.5.). The misincorporation data from Neumann *et al.* (2010) were for any base switch, while mis-scoring a non-exchange conversion requires precisely the "correct" misincorporation. Therefore the true level of PCR misincorporations is likely to be much lower. However given the transitional bias this estimate should be a good approximation for any transitional SNP, but will be much too high for a transversional SNP or an insertion/deletion polymorphism.

Only a few markers were involved in single-site events outside of the hotspot interval, and those that occurred at a frequency below 0.005% were therefore viewed as PCR-artefacts. At Superhotspot T, several markers located more than 1.2 kb away from the hotspot centre were involved in single-site exchanges. These may represent true non-exchange conversion events, as firstly a transversional SNP conversion was observed at marker T9.2 and secondly conversions at SNP T9.5 were significantly elevated above the expected PCR misincorporation rate. This observation may therefore point to another recombination active region within the test interval. Closely spaced hotspot doublets do

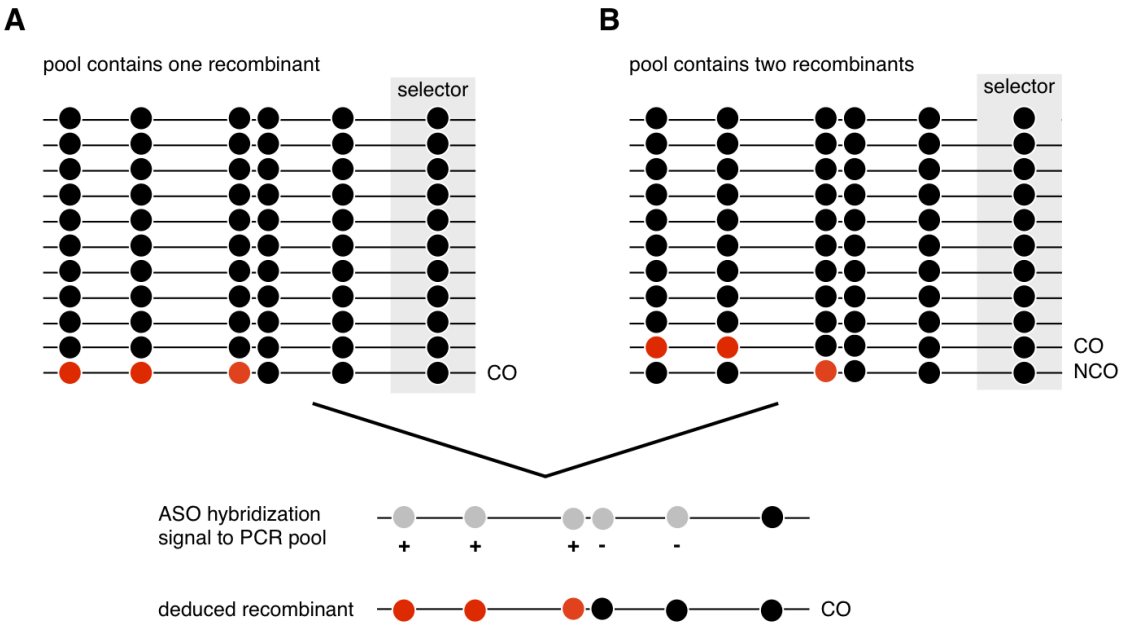
exist (Jeffreys *et al.* 2005; Webb *et al.* 2008) and can be analysed for crossover, though unfortunately it was impossible here, as re-sequencing of LD-blocks upstream of the hotspot failed to identify suitable selector sites needed for allele-specific PCR.

3.6.1.2 Recombination hotspot morphology

Non-exchange conversion profiles were established at all hotspots and their peak centred at the centre of the crossover distribution, defining steep bi-directional gradients. The non-exchange conversion distributions observed here therefore corresponds to those observed previously at human hotspots (Jeffreys and May 2004). Crossover-breakpoints were found within 1-2 kb intervals, and the distribution of crossover breakpoints was similar to those observed previously using full crossover assays. Curiously at some hotspots needle-like peaks of recombination activity were observed that appear significant. One explanation for these "spikes" may be mis-mapping of crossover breakpoints to very short intervals. Crossover profiles generated previously in full-crossover assays at the same hotspots did not detect such "spikes", and it therefore appears that the half-crossover assay may be more prone to mis-mapping than a regular full-crossover assay. This may be due to the fact that in the half-crossover assays PCR pools can potentially contain both, a crossover and non-exchange conversion molecule. With the half-crossover assay, a pool that contained a crossover molecule or a pool that contained both, a crossover and a non-exchange conversion molecule would have been indistinguishable. Breakpoints could be mis-mapped between two very closely spaced markers if this non-exchange conversion occurred at one of the markers flanking the short interval, as illustrated in Figure 3-12.

As illustrated in Figure 3-12, mis-scoring a pool to contain only a crossover, when in fact it also contains a non-exchange conversion molecule, could have lead to mis-mapping of crossover breakpoints to a very narrow different interval. Consistent with this theory, crossover-"spikes" were observed exactly at the same location as the peak non-exchange conversion activity at Superhotspots E and K.

Figure 3-12 Complex pools



Circles represent polymorphic SNP with progenitor origin indicated by colour (red and black). Selector sites are located to the right (grey box). Two types of pools could explain the observed breakpoint distribution. (A) Crossover molecule with a breakpoint within the very short interval between two closely spaced markers (B) crossover molecule with a breakpoint located within a wider interval, plus a non-exchange conversion molecule involving one the adjacent SNPs that flanked the short interval.

The frequency of non-exchange conversions involving the peak marker, as well as the frequency of crossovers mapping to the previous interval (as illustrated in Figure 3-12) have been determined. Therefore it can be estimated how likely it is for both types of recombinants to be present within one and the same pool. This likelihood was 0.07% at Superhotspot E and 0.24%, at Superhotspot K, respectively.

At Superhotspot E we would therefore predict that 0.4 pools contain both types of recombinants. Adding this frequency to the expected number of breakpoints in the given interval, shows that the combined number was not significantly different from the observed number of breaks (Poisson-approximated binomial probability P=0.079). The same was true for Superhotspot K; we would expect more than 1.7 pools containing both types of recombinants. The number of expected breaks within the peak interval and the number of pools expected to contain non-exchange conversions combined give a frequency that was not significant different from the observed frequency of breaks within this small interval (Poisson-approximated binomial probability P=0.373). Curiously at Superhotspot T, the peak was observed at a transitional marker that did not show any non-

exchange conversion and it remains to be determined whether crossovers were mis-mapped to this interval and why.

Crossover breakpoints mark the sites of crossover resolution, not initiation and can therefore not be used to locate the recombination initiation site. Non-exchange conversion tracts were found within a somewhat narrower interval than crossover breakpoints and were always located well within the 95% width of the crossover breakpoint distribution. Co-conversion tracts always involved markers flanking both sides of the hotspot centre. This is consistent with a hotspot initiation model where recombination is initiated at a narrow zone within the hotspot centre (Jeffreys and Neumann 2002).

3.6.1.3 Conversion tract lengths

Conversion tract estimations based on observable marker exchanges were used to gain limited information about tract lengths. Maximum conversion tracts were by definition overestimated, and could potentially involve only a couple of bases on each side of the converted marker(s). Nevertheless, conversion tracts appeared to be short, and the minimal and maximal estimates of 60 and 833 bp, respectively, fit well with previous estimates of conversion tract lengths at human recombination hotspots (Jeffreys and May 2004).

3.6.1.4 Crossover and non-exchange conversion frequencies were positively correlated

A positive correlation was observed between non-exchange conversion and crossover frequencies, with hotspots highly active in crossover also showing high non-exchange conversion frequency. It can therefore be concluded that the higher the number of initiation events is, the higher both the crossover as well as non-exchange conversion frequency. This observation argues against a simple model were the numbers of initiating events was constant at all hotspots, and the junction was resolved either as non-crossover or crossover. If that were true, hotspots with a high frequency of crossovers would have little non-exchange conversion, as DSBs would be preferentially repaired as crossovers and *vice versa*. In this case differences between conversion and crossover ratios would dependent only on the ability to detect non-crossover recombinants as non-exchange conversions. Contrary to this model, non-exchange conversion and crossover frequencies

were instead positively and fairly tightly correlated, with peak non-exchange conversions occurring at about (54±10)% of crossovers. If the non-exchange conversion to crossover ratio were dependent on the orientation of Holliday-junction resolution, as described in the DSBR model of recombination, this constant CON:CO could be the result of a given likelihood of resolution in either direction. However this observation does not imply that non-crossovers arise at a lower frequency than crossovers. Non-crossover recombination does not necessarily result in detectable gene conversion, as it might not involve conversion of an informative marker. Consistent with this observation the non-exchange conversion frequency was observed to be roughly twice as high in hotspots, where two SNPs were located close to the hotspot centre, compared to those observed at one SNP.

Chapter Four Biased gene conversion at human recombination hotspots

4.1 Introduction

Crossover generation is fully reciprocal, and not only leads to equal numbers of reciprocal crossovers but also to a 50:50 ratio of alleles transmitted to crossover progeny at any given marker. Biased gene conversion associated with crossover leads to breakpoints in reciprocal orientation not mapping to the same interval. As a consequence of non-reciprocal tracts accompanying crossovers, alleles are now found at a non 50:50 ratio, which is referred to as transmission distortion (TD).

Reciprocal crossover asymmetry and TD has been observed at two human recombination hotspots, *DNA*2 and *NID*1. Heterozygosity at a single SNP was necessary and sufficient to induce biased gene conversion associated with crossover. Men homozygous for an activating allele had high crossover frequencies while men homozygous for suppressing alleles had reduced crossover frequencies; therefore this phenomenon likely resulted from biased initiation (Jeffreys and Neumann 2002; Jeffreys and Neumann 2005). Biased initiation leads to biased gene conversion as a break initiated on one homologue is repaired with alleles from the uncut homologue. Hence, frequent initiation on an active homologue leads to overtransmission of alleles from the suppressed homologue continually enriching recombinant progeny for suppressing alleles. This creates a meiotic drive that, if strong enough, promotes the attenuation or extinction of the hotspot over time. At Hotspot *NID*1 non-exchange gene conversions were also analysed (Jeffreys and Neumann 2005), and biased transmission of the suppressed allele into non-exchange gene conversion further increased the meiotic drive towards fixation of the suppressing allele (Jeffreys and Neumann 2005).

Variation in recombination frequencies between men was observed in Chapter 3. These could be the result of differences in initiation efficiency between haplotypes. By testing whether hotspots exhibit reciprocal crossover asymmetry, active and inactive haplotypes should be identifiable. Additionally, as biased gene conversion can have a strong effect on the dynamics of hotspot turnover, it would be interesting to know whether crossover and gene conversion are always influenced by the same biases, as initially observed for Hotspot *NID*1, or whether additional biases in non-crossover exist.

4.2 This work

This Chapter aims to detect incidences of biased gene conversion within the Superhot-spots analysed using reciprocal half-crossover assays in Chapter 3. In this Chapter, recombinant molecules detected in reciprocal assays were tested for disparity in reciprocal conversion rates by comparing the frequency and distribution of recombinant molecules generated by reciprocal recombination events. This allowed biased gene conversion to be detected and quantified. As for the first time a high number of non-exchange conversion molecules was also detected, this Chapter not only addresses gene conversion associated with crossover but also in non-exchange conversions, and whether they were influenced by the same biases.

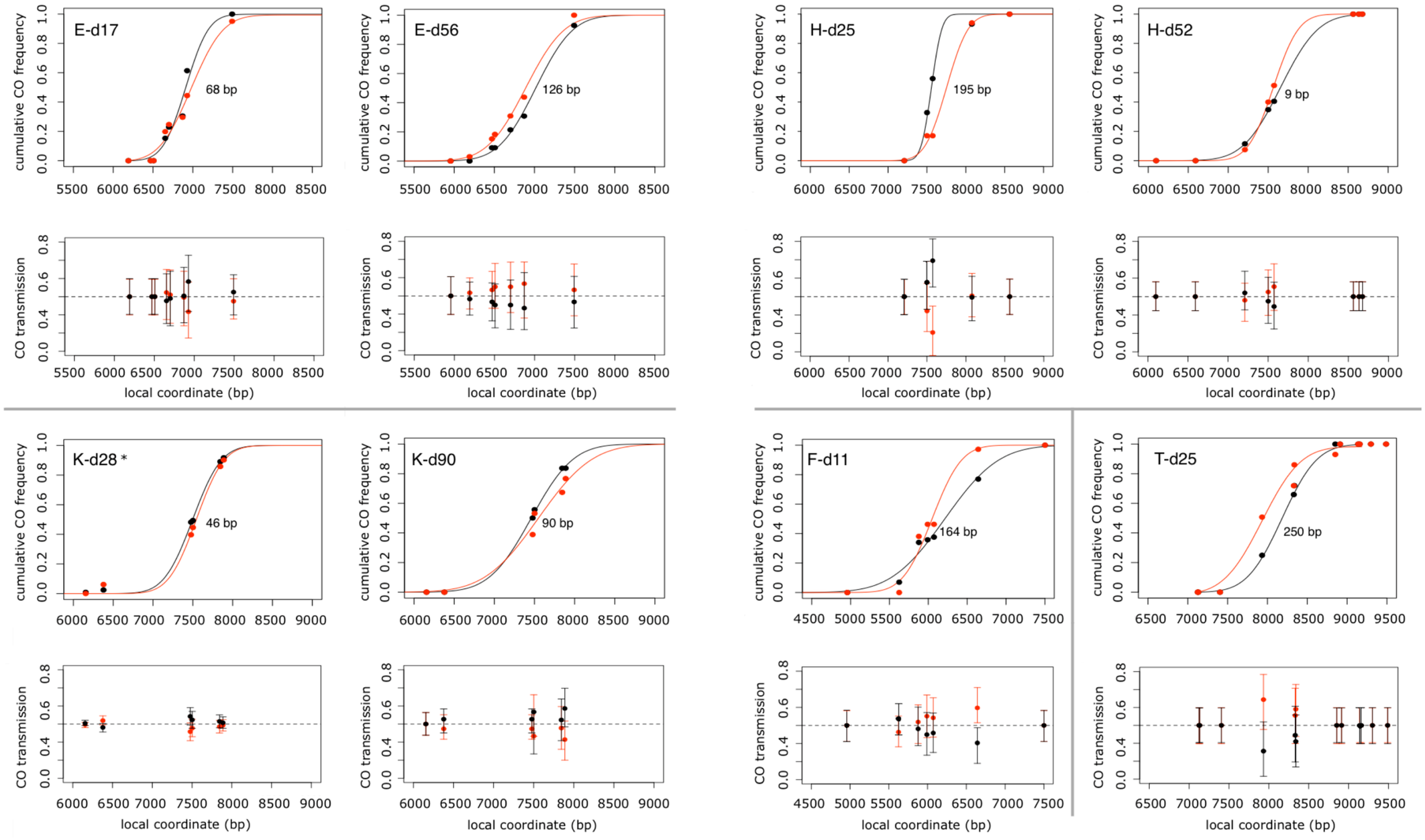
4.3 Testing for disparity in reciprocal recombination rates

In humans reciprocal products of the same recombination event cannot be detected, as they were separated prior to spermatogenesis. Additionally, reciprocal half-crossover assays were not performed on the same pools of sperm DNA, and further used different sets of primers. So before comparing recombinants detected in reciprocal orientations, it first had to be established whether differences between reciprocal assays have signifi-cantly affected recombinant detection. Numbers of crossovers detected in reciprocal assays had been compared previously and significant deviation between reciprocal half-crossover assays was not observed (3.4.1.1,Table 3-2). As reciprocal assays at all hotspots gave comparable results, it therefore followed that reciprocal recombination products can be meaningfully compared and tested for biased gene conversion into crossovers, as well as non-exchange conversions.

4.3.1 Biased gene conversion associated with crossover

Cumulative distributions of crossovers for each reciprocal orientation were generated and compared to test for reciprocal crossover asymmetry. The locations of hotspot centres, defined by the centres of crossover breakpoint distributions, were then compared between orientations. To identify biased gene conversion associated with crossover, alleles trans-mitted in each orientation of recombination were compared. Transmission frequencies were determined from normalized proportions of crossovers, as numbers detected in reciprocal assays were not equal. Reciprocal cumulative crossover distributions and marker transmission frequencies at all hotspots and men tested are shown in Figure 4-1.

Figure 4-1 Testing for biased gene conversion associated with crossover



Top panel: To visualise reciprocal CO asymmetry, observed cumulative CO breakpoint distributions were plotted for each orientation separately, orientation A (red) and orientation B (black). Red or black lines show least-squares best-fit cumulative CO distributions, for each orientation, respectively. Observed cumulative frequencies are indicated by red or black dots. Shifts between centres of these best-fit distributions were displayed as distance in bp next to the curve.
Bottom panel: To detect biased allele transmission into COs, normalised allele-transmission frequencies were compared at each marker. Red dots indicate normalised transmission frequency in orientation A, black dots in orientation B, respectively. 95% confidence intervals were determined using a bespoke simulation programme written by A. J. Jeffreys. *Data from the analysis of d28 at superhotspot K performed Webb *et al.* (2008), was included in this graph.

In most cases no evidence for biased gene conversion accompanying crossover was observed, with no significant asymmetry in reciprocal crossover distributions and no significant departure from a 50:50 transmission ratio of allele into crossover. However, there are three exceptions.

For man d11 at Superhotspot F the peak locations of reciprocal breakpoint distributions were shifted by 164 bp. The central markers, F6.0 and F6.1 showed no indication for deviation from the expected 50:50 ratio (two-tailed Fisher-exacts P=0.655 and P=0.361, respectively). Curiously, at off-centre marker F6.6b, an excess of G-alleles was transmitted into crossovers, with a transmission ratio of 60:40. This ratio is only marginally significantly distorted from 50:50 (two-tailed Fisher-exact, P=0.067). However, two additional men, heterozygous at marker F6.6b that had been analysed for crossover transmission by Webb *et al.* also showed significant TD at this marker (Webb *et al.* 2008) equally overtransmitting the G-allele, and the overall degree of TD is highly significant across the three men tested (combined p=0.0002, 6.d.f).

In man d25 at Superhotspot H, the breakpoints of crossovers in orientation A mapped to different intervals than the breakpoints of crossovers in orientation B, with a 195 bp shift between peak locations. Significant TD was observed at the central marker H7.6 (rs3899614) with the G allele transmitted into 69 % of crossover molecules. Transmission ratios of G-allele and A-allele were marginally significantly different from 50:50 in this dataset (two-tailed Fisher-exact, p=0.061). But the same man had also been analysed for crossover by Webb *et al.*, and pooling crossovers detected in both surveys reveals a significant overtransmission of the G-allele over the A-allele in d25 (two-tailed Fisher-exact, p=0.031). The second man analysed at Superhotspot H, d52, showed no indication for reciprocal crossover asymmetry, instead the G-allele was found in only 47% of crossovers, a proportion not significantly different from 50% (two-tailed Fisher-exact, P=0.761).

In man d25 at Superhotspot T, recombinant breakpoints in orientation A and orientation B mapped to either side of the central marker T7.9 (rs9854419), with a 250 bp shift in peak locations. Transmission into crossovers at T7.9 indicated modest TD with the T7.9 G allele found in 63% of crossover molecules, but this was not significant (two-tailed Fisher-exact, P=0.106). Data from A.J. Jeffreys on crossover distributions had shown

partial A/B displacement with indication of TD at the same marker in two other men heterozygous at T7.9 (A. J. Jeffreys, *unpublished data*). The transmission frequencies can be compared between all men analysed with the different assays (two-tailed Fisher-exact on 2x3 contingency table, P=0.752, after Bonferroni correction for the 3 men) and showed a combined transmission ratio of 57:43 in favour of the G-allele a weak but nevertheless significant deviation from 50:50 (two-tailed exact binomial, p=0.0013).

4.3.2 Testing for biased non-exchange conversion

Reciprocal crossover asymmetry and transmission distortion in crossover molecules was found at Superhotspots F and H, with indication of reciprocal crossover asymmetry at Superhotspot T. In a study by Jeffreys and Neumann, at hotspot *NID*1, a similar degree of transmission bias was also seen in non-exchange conversion molecules in addition to biased TD in crossovers (Jeffreys and Neumann 2005). To test whether crossovers and non-exchange conversions were influenced by the same biases in the hotspots analysed here, numbers of non-exchange conversions in reciprocal orientation were compared (Table 4-1). This was done for all of the hotspots in this survey to also determine whether biases specifically in non-exchange conversions could be identified.

Table 4-1 comparing non-exchange conversion frequencies between orientations

hotspot	man	Sperm molecules analysed per orientation	A non-exchange conversions		B non-exchange conversions		A/B difference
			Poisson no	Frequency	Poisson no	Frequency	
E	d17	7360	16.5	0.22	8.1	0.11	*P*=0.090
E	[1]d56	[1]2290	[1]6.9	0.30	[1]14.6	0.65	*P*=0.097
F	d11	8280	9.0	0.11	13.0	0.16	*P*=0.394
H	d25	12420	11.1	0.09	5.0	0.04	*P*=0.129
H	d52	11040	8.1	0.07	7.1	0.06	*P*=0.798
K	d28	6980	17.9	0.28	15.7	0.24	*P*=0.704
K	d90	14720	25.9	0.18	50.0	0.32	*P*=0.006
T	d25	3496	10.3	0.29	5.1	0.14	*P*=0.185

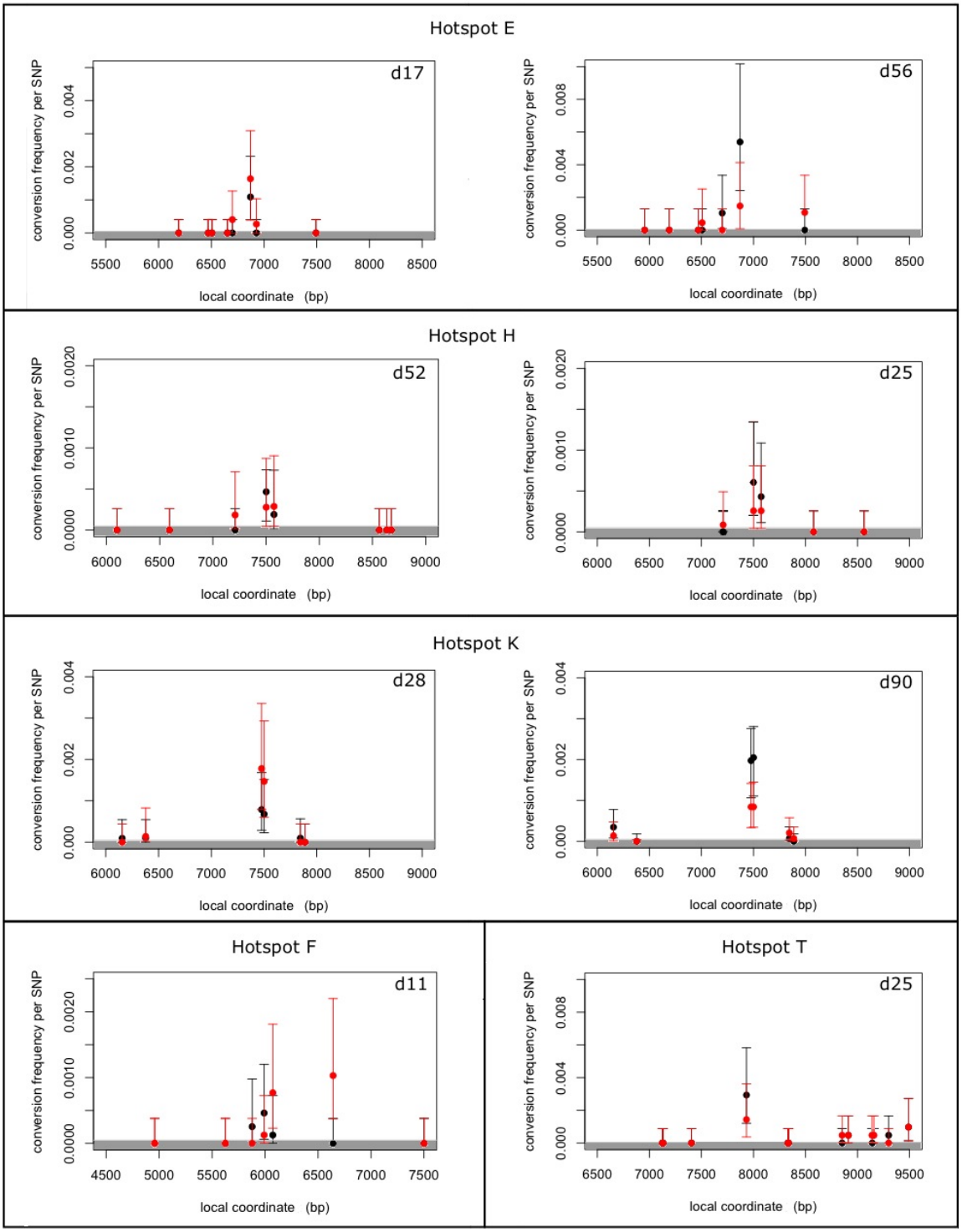
P-values are two-tailed probabilities of a t-test comparing Poisson means. [1]Data from Alec J. Jeffreys (Webb *et al.* 2008) was included in this table.

Numbers of non-exchange conversion detected in reciprocal orientations were in most cases not significantly different from each other, as summarised in Table 4-1. Interest-

ingly, significant variation between numbers generated in reciprocal orientation cannot be detected for Superhotspots H, F and T, which had shown indication of biased gene conversion accompanying crossover (Figure 4-1). However, a significant difference was observed in man d90, analysed at Superhotspot K, pointing to a bias in non-exchange conversion. Curiously, the same man had shown no indication for reciprocal crossover asymmetry and biased gene conversion associated with crossover (Figure 4-1).

To analyse non-exchange conversion events in more detail, conversion profiles were generated for each orientation of recombination separately. Conversion profiles enable the detection of transmission biases by allowing comparison of transmission frequencies at each marker across the hotspot interval (Figure 4-2).

Figure 4-2 non-exchange conversion frequency profiles separated by orientation



To detect biased non-exchange conversion, reciprocal allele transmission frequencies were compared between both orientations of recombination. Distributions of non-exchange conversions in orientation A (red) and orientation B (black) are shown, with 95% confidence intervals determined by Poisson-approximation. Expected false-positive rates, the expected frequency of single-site exchanges generated through PCR misincorporation, are shaded in grey (as in Figure 3-6).

As shown in Figure 4-2, non-exchange conversion profiles in reciprocal orientations show the same width and distribution. However, non-exchange conversion profiles at Superhot-spot F vary between orientations with the peak non-exchange conversion activity located at off-centre marker F6.6b in orientation A (red), though at marker F6.0 in orientation B (black).

The Superhotspots that had shown significant biased gene conversion accompanying crossover, Superhotspots F and H displayed an excess of the same alleles in non-exchange conversions (Figure 4-2). Significant variations between non-exchange conversions in reciprocal orientations were only observed at marker F6.6b at Superhotspot F.

Superhotspot F had shown crossover-associated gene conversion of the A→G allele at off-centre marker F6.6b. In non-exchange conversions, transmission in the same direction was 4-times more frequent that conversion of the G→A-allele, but this ratio was not significantly different from 50:50 (two-tailed exact binomial, P=0.109). As the degree of transmission was comparable between non-exchange conversions and crossovers (two-tailed Fisher-exact, P=0.291) transmission frequencies were combined, which revealed that 60% of recombinants involving marker F6.6b obtained the G-allele, a significant distortion from 50:50 transmission ratio (two-tailed exact binomial, p=0.024).

At Superhotspot H, even though transmission in non-exchange conversion was not significantly different from 50:50 (two-tailed exact binomial, P=0.726) transmission frequencies of alleles were comparable between non-exchange conversion and crossover recombinants (P=1, two-tailed Fisher-exact). Combining both types of recombinant progeny revealed a 63:38 ratio in favour of the G-allele at marker H7.6, a transmission ratio significantly different from 50:50 (two-tailed exact binomial p=0.0074).

At Superhotspot T, the G allele at central SNP T7.9 was found in 63 % of crossover molecules, a ratio that was marginally significantly different from 50:50. Non-exchange conversions also show an excess of G-alleles, with 67% of non-exchange conversions having acquired the G-allele, though this data on its own is not significant (two-tailed exact binomial, P=0.302).

In addition, significant variation in transmission frequencies between reciprocal orientations was also observed at Superhotspots E and K, which had shown no indication for reciprocal crossover asymmetry (Figure 4-1). At Superhotspot E, man d56 shows an excess of G-alleles at marker E6.8, however this excess is not significantly different from 50:50 transmission (two-tailed exact binomial, P=0.092). Additionally, man d17 at the same Superhotspot equally shows an excess of allele G at marker E6.8 A/G, this excess is again not significantly distorted from 50:50 (two-tailed exact binomial P=0.503). TD at marker E6.8 in non-exchange conversions can be compared between both men (Fisher-exact, two tailed, P=0.455). Crossovers did not display any evidence for reciprocal crossover asymmetry, and transmission ratios in crossover and non-exchange conversions are comparable for d17 and d56 (two-tailed Fisher exact tests, P=0.571 and P=0.197, respectively).

Man d90 at Superhotspot K had a significantly higher non-exchange conversion frequency in orientation B (as shown in Table 4-1). This difference in numbers is reflected in the significantly higher transmission frequencies at the central SNPs K7.4 T/C (rs1374633 A/G) and K7.5 C/G (rs1374632 G/C). More than 69% of non-exchange conversion molecules obtained the C-allele at K7.4, a proportion significantly different from 50% (two-tailed exact binomial, P=0.013). Equally, marker K7.5 C/G was affected, here the C–allele was found in 70% of non-crossover progeny, again a significant difference from 50:50 transmission (two-tailed exact binomial, P=0.010). Reciprocal crossover asymmetry and biased transmission into crossover was not observed (as shown in Figure 4-1) and transmission frequencies were significantly different between crossover and non-exchange conversion molecules (two-tailed Fisher-exact, p=0.036).

4.4 A more detailed analysis of transmission bias differences between crossovers and non-exchange conversions

Biased transmission that only affects non-exchange conversions had never previously been observed. This novel observation is most curious and requires further analysis.

4.4.1 Comparing haplotypes and transmission profiles between men

To test for local sequence variants that could control the bias purely into non-exchange conversion in man d90, interacting haplotypes analysed at Superhotspot K were compared

between men. Interestingly the second man d28 also exhibits an excess of alleles, however with opposite directionality (Figure 4-2). The two men had nearly identical haplotypes, differing only in the phase of SNP K7.4 resulting in the C-and T-allele being located on opposite haplotypes. Curiously, therefore it is again the C-allele at K7.4 that is found in excess in non-crossover progeny. The observed rate of T→C compared to C→T transmission at K7.4 (rs1374633) is therefore elevated in both men, and the degree of elevation was comparable (two-tailed Fisher-exact, P=0.565) as summarized in Table 4-2.

Table 4-2 Transmission of alleles at marker K7.4

Non-exchange conversion		K7.4 T→ C		K7.4 C→ T	
man	haploid genome equivalents (HGEs)	CONs	CON freq	CONs	CON freq
d28	6980	12	1.7×10-3	8	1.1×10-3
d90	14720	27	1.8×10-3	12	0.8×10-3
Sum	21700	39	1.8×10-3	20	0.8×10-3

Crossover		K7.4 T→ C		K7.4 C→ T	
man	haploid genome equivalents (HGEs)	COs	Crossover freq	COs	Crossover freq
d28*	72880	124	1.7×10-3	154	2.1×10-3
d90	14720	26	1.8×10-3	30	2.0×10-3
Sum	87600	150	1.7×10-3	184	2.1×10-3

Absolute number of times a given allele was found in non-exchange conversion molecules as well as crossover molecules was counted, *included crossover data from d28, generated by Webb *et al.* (Webb *et al.* 2008).

Combining data from both men indicated that the derived C-allele at rs1374633 was significantly over-transmitted into non-exchange conversion molecules (two-tailed exact binomial, p=0.018), but not into crossover progeny in the two men analysed (two-tailed exact binomial, p=0.071) (Table 4-2). This bias was significantly different between crossovers and non-exchange conversions (two-tailed Fisher-exact, P=0.003), which is a novel observation. If biases purely in non-exchange conversion were found more often, they would indicate a degree of independence in the generation of non-crossovers and crossovers in humans.

If this observation is true, then not the haplotype *per se*, but the phase of K7.4 alleles determined the direction of the bias. This phenomenon was most interesting and required further analysis. Could the SNP itself trigger a bias into non-exchange conversion with the adjacent SNP co-converting? To test whether the central K7.4 SNP marker itself could cause the distortion, then heterozygosity at this marker should be necessary and sufficient to trigger this bias, and the bias should disappear in men homozygous at K7.4.

4.4.2 Biased transmission in non-crossovers was not restricted to Superhotspot K

Interestingly a strikingly similar phenomenon was identified in an additional study. Unpublished data from A.J. Jeffreys had identified what appears to be biased gene conversion affecting non-crossovers to a higher degree than crossovers at Superhotspot F. A half-crossover assay on d28 had shown unidirectional non-exchange conversion involving the two central SNPs F6.0A/T (rs386440) and F6.1A/G (rs10492181) that often co-converted. The data indicated at least a 81:19 ratio in favour of conversions in one direction over the other, a significant deviation from the expected 50:50 ratio under parity. A bias in the same direction was also seen in crossovers but was far less pronounced, with the central F6.1A/G (rs10492181) showing only insignificant TD of 55:45 in crossover progeny (A. J. Jeffreys, *unpublished data*).

4.4.3 Extending the analysis to donors with specific haplotypes

To investigate whether heterozygosity at the central SNP marker itself can cause TD purely into non-exchange conversion, the datasets were increased. More men at both hotspots were analysed using the half-crossover assay to detect crossovers and non-exchange conversions (3.3). To be able to efficiently detect non-exchange conversion bias, if present, it was also necessary to increase the numbers of molecules analysed. Donors had to have suitable selector sites outside of the hotspot to allow haplotype differentiation, plus additional heterozygosities to distinguish crossover from non-exchange conversion molecules as described previously (Chapter 3, 3.3.2), though not necessarily the same high marker density across the hotspots. To include as many candidate donors as possible meeting these criteria, an additional set of ASPs (shown in Appendix I) was optimised for both hotspots.

4.4.3.1 Superhotspot K

At Superhotspot K it was tested whether biased transmission at marker K7.5 would disappear in men homozygous for the candidate *cis*-regulatory SNP K7.4. Two men, homozygous for either the C or the T allele at K7.4 and heterozygous at K7.5 were included in the analysis. These men could still be scored for non-exchange conversions as the K7.5 SNP was previously found to have converted at a frequency of > 0.1% per sperm.

In addition, a man heterozygous at both markers was included. Together with the men that had already been analysed, this survey thus allowed the comparison of three men heterozygous at K7.4 and two men homozygous at K7.4 each with an alternative K7.4 allele to test the effect of homozygosity at K7.4.

Man d21, homozygous for allele T at marker K7.4, was analysed using the same assay that had been used for d28 and d90 (Chapter 3, 3.3.3.1.4). The other two men, d35, heterozygous at K7.4, and d67, homozygous for allele C at K7.4, were not heterozygous at the selector sites used in this assay. Therefore an additional assay was set up that used selector sites K2.9 T/C (rs4735211) and K3.2 C/G (rs10090492) instead. Three 96-well plates were tested for each reciprocal assay with each well containing 40 amplifiable sperm DNA molecules, also referred to as haploid genome equivalents (HGE). This resulted in 11040 HGEs analysed in each orientation of recombination. PCR conditions of both assays are collected in Appendix III. Recombinant breakpoint mapping was carried out as described in Chapter 3 (3.3.5).

4.4.3.2 Superhotspot F

Simultaneously A.J. Jeffreys and I. L. Berg performed an assay at Superhotspot F for eight additional men (*unpublished data*). Two 96-well plates with inputs of 25 or 40 HGEs per pool were tested, resulting in 5980 HGEs screened per orientation for each man. Including the men analysed previously, this allowed the comparison of data of ten men at Superhotspot F.

4.4.4 Relative frequencies of crossovers and non-exchange conversions

Crossover and non-exchange conversion recombinants were detected for all of the additional men analysed (collected in Appendix V). At Superhotspot K, only recombinants involving marker K7.5 could be compared between all men, as men homozygous for the candidate regulatory SNP K7.4 were included into analysis. In contrast, men analysed at Superhotspot F can all be compared at marker F6.1, the marker that showed the strongest distortion initially. Poisson-corrected numbers of crossovers as well as non-exchange conversions spanning marker K7.5 at Superhotspot K or alternatively spanning marker F6.1 at Superhotspot F, are collected in Table 4-3 and Table 4-4, respectively.

Table 4-3 crossovers and non-exchange conversions spanning K7.5 at Superhotspot K

man	half-crossover assay	haploid genome equivalents (A+B)	Number of crossovers	Number of CONs at K7.5 C/G	Crossover frequency (%)	CON frequency (%)	CON:CO
d35	new	22080	13	11	0.06	0.048	0.85
d28	original	13800	21	19	0.152	0.138	0.9
d90	original	29440	58	46	0.197	0.156	0.79
d67	new	22080	54	52	0.245	0.236	0.96
d21	original	22080	58	35	0.263	0.158	0.6
				mean:	0.183	0.147	0.82

Men were sorted by ascending crossover frequency, determined from Poisson-corrected numbers of crossovers. In addition non-exchange conversions spanning marker K7.5 are shown, as they can be compared between the five men.

Hotspot centres and widths were indistinguishable between half-crossover assays and full-crossover assays performed by (Webb *et al.* 2008), but again a larger variation between crossover frequencies was observed (as summarised in Table 4-3). Men d21, d28 and d90 were analysed using a different assay than men d35 and d67. But data generated using the two half-crossover assays was comparable (two-tailed t-test, P=0.567), and equally numbers of crossovers obtained in reciprocal orientations were comparable in both the original assay and the new assay (t-test for Poisson means, p=0.930 and p=0.211).

Frequencies of crossovers and non-exchange conversions were both variable between men (4 to 5-fold, respectively). Non-exchange conversion and crossover frequencies were positively correlated as observed in Chapter 3 (linear regression, P=0.016). As observed

previously at Superhotspot K, CON:CO ratios did not vary significantly between men (P=0.587, 2×5 contingency table, 4.d.f.) with the non-exchange conversion frequencies at K7.5 arising on average at 82±6% of the crossover frequency.

A.J Jeffreys and I.L. Berg performed a similar analysis for Superhotspot F. Here non-exchange conversion frequencies were compared at SNP F6.1 (Table 4-4). The mean crossover frequency obtained here was comparable with the mean frequency obtained by (Webb *et al.* 2008), but again a higher degree of variation was observed. The lowest crossover frequency was found in d35, who had equally shown the lowest recombination frequency at Superhotspot K.

Table 4-4 crossovers and non-exchange conversions spanning F6.1 at Superhotspot F

man	haploid genome equivalents (A+B)	Number of crossovers	Number of CONs at F6.1A/G	Crossover frequency (%)	CON frequency (%)	CON:CO
*d35	12218	25	6	0.203	0.052	0.25
[2]d11	16560	47	9	0.282	0.057	0.20
*d77	11960	65	5	0.542	0.045	0.08
*d87	11935	73	4	0.615	0.037	0.06
*d55	11960	75	10	0.630	0.082	0.13
[1]d28	16660	167	13	1.005	0.078	0.08
*d20	12180	133	7	1.089	0.055	0.05
*d51	12016	136	18	1.133	0.146	0.13
*d31	11785	151	18	1.281	0.155	0.12
*d2	11960	155	11	1.298	0.092	0.07
			mean:	0.808	0.080	0.118

Poisson-corrected numbers of crossovers as well as non-exchange conversions involving marker F6.1A/G (rs10492181) were shown with men sorted by ascending crossover frequency. Data from [1]earlier assay by A.J. Jeffreys, [2] from assay described in Chapter 3.* data generated by A.J. Jeffreys and I. L. Berg.

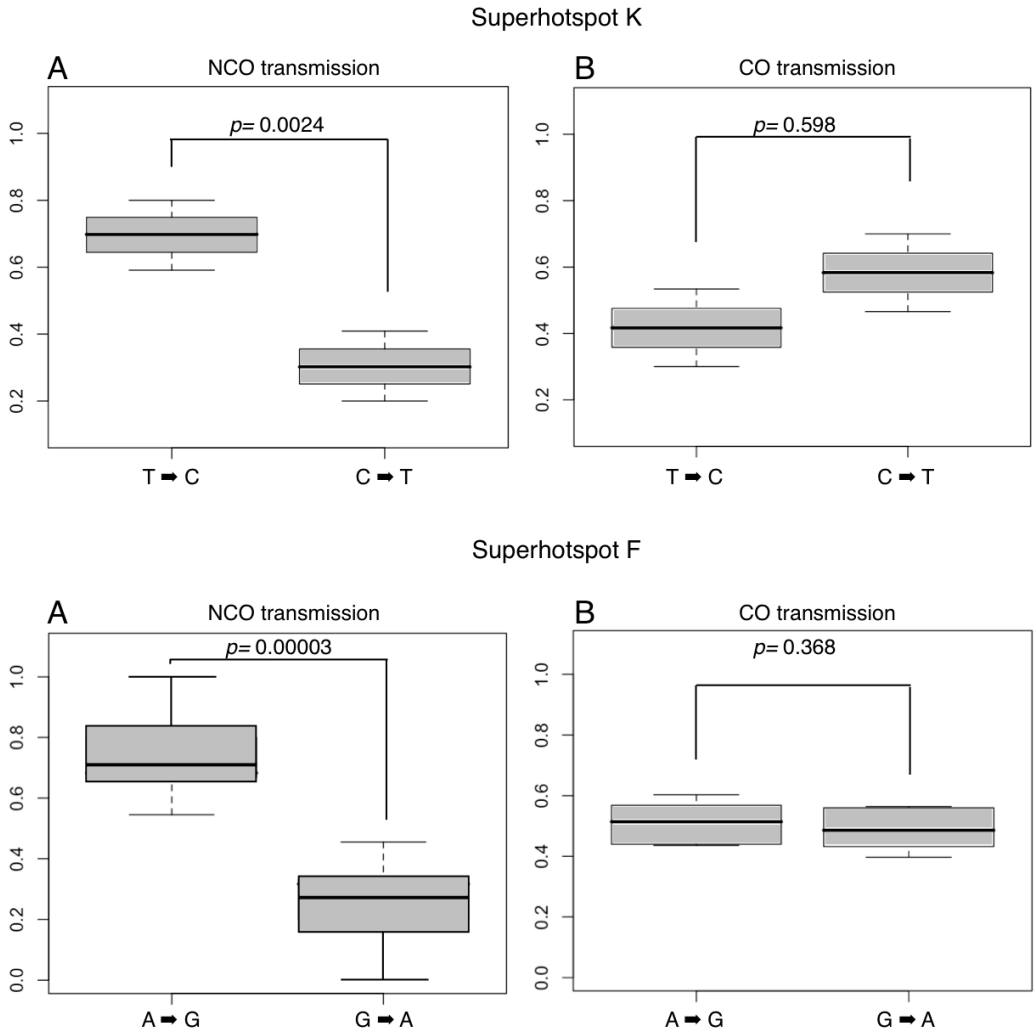
Again crossover as well as non-exchange conversion frequencies varied between men (by 4-and 6-fold respectively), and frequencies of crossover and non-exchange conversion were positively correlated (linear regression, P=0.030). Variation in CON:CO ratios between men was not significant (2×10 contingency table, 9 d.f. P=0.057) as seen before.

At Superhotspot F, non-exchange conversions arise on average at 12% of the crossover frequency, which is much lower than at Superhotspot K, where non-exchange conversions arise on average at 82% of the crossover frequency.

4.4.4.1 Heterozygosity at the candidate markers results in TD in non-exchange conversion but not crossover recombinants

Crossover and non-exchange conversions were tested for distortion from 50:50 transmission ratios of alleles at markers K7.4 and F6.1, respectively. Comparison of allele transmissions between crossovers and non-exchange conversions for Superhotspots K and F are shown in Figure 4-3 (including data generated by A.J. Jeffreys and I.L Berg).

Figure 4-3 Allele transmission into recombinant progeny



Transmission frequencies of particular alleles observed in (A) non-exchange conversions or (B) crossovers (based on Poisson-corrected numbers) separated by Superhotspot.

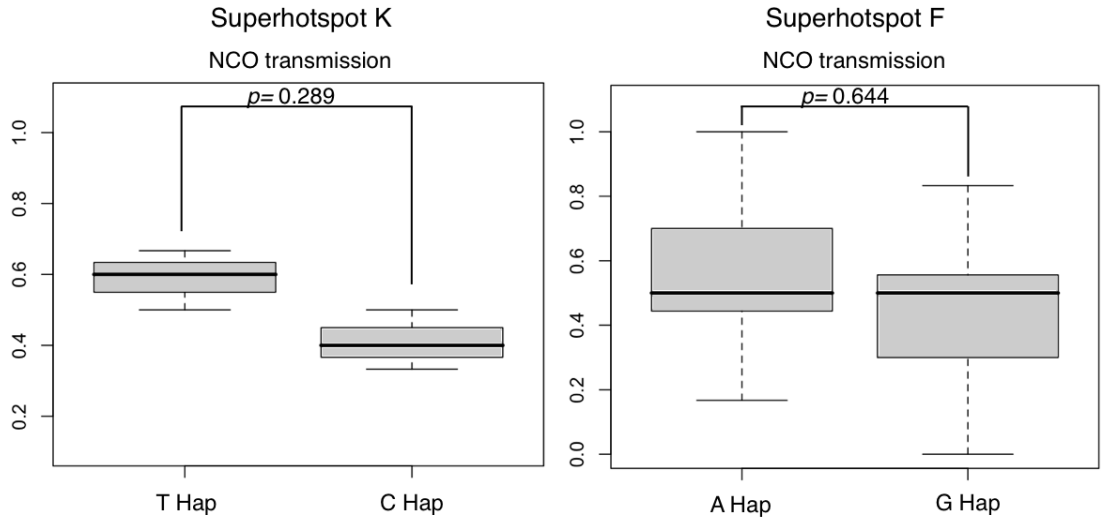
For Superhotspot K three of the analysed men were heterozygous at marker K7.4, the candidate regulating SNP. Transmission ratios of K7.4 C and T-alleles were indistinguishable between these men (d28, d35, d90) in both crossovers and non-exchange conversions (Fisher-exact on 2x3 contingency table, P=0.564 and P=1, respectively after Bonferroni-correction for the 3 men). Crossovers showed no indication for TD, with 46.7% of crossovers containing K7.4C, which was not significantly different from 50:50 (exact binomial, p=0.598). Each donor was also tested independently, and none showed significant departure from 50:50 ratio of transmission at marker K7.4 in crossover. In contrast 68% of non-exchange conversions acquired the C-allele at K7.4 and only 32% the T-allele, a significant deviation from 50:50 (two-tailed exact binomial, p=0.0024) (Figure 4-3). Transmission frequencies in non-exchange conversions, spanning marker K7.4, and crossovers were significantly different from each other (two-tailed Fisher-exact, p=0.007).

The same phenomenon was observed in the men analysed at Superhotspot F by A.J. Jeffreys and I.L Berg. Crossovers at Superhotspot F equally did not show evidence for TD, overall 51.5% of crossovers contained F6.1 G, a proportion not significantly different from 50% (two-tailed exact-binomial, p=0.368). Each donor was also tested independently, and none showed significant departure from 50:50 ratio of transmission at marker F6.1 (combined P=0.240). As opposed to the balanced transmission in crossovers, non-exchange conversions involving marker F6.1 were strongly biased. The F6.1G allele was found in 71% of non-exchange conversions spanning marker F6.1, a ratio significantly different from 50:50, (two-tailed exact binomial, p=0.00003). Every man had an excess of non-exchange conversions spanning F6.1 G compared to non-exchange conversions with F6.1 A, and this was seen despite the small numbers of non-exchange conversions detected. Man d28 had initially shown this phenomenon as unidirectional transmission of alleles at F6.1 A→G but not F6.1 G→A in non-exchange conversions (A.J. Jeffreys, *unpublished data*). This transmission ratio remained the most extreme TD seen, but it was still compatible with the overall level of TD in all men (P=0.230, after Bonferroni correction for the ten men). Even if this man was excluded, the increased frequency of G-alleles over A-alleles remained highly significant (two-tailed exact binomial, p=0.00279). Again, the difference in transmission ratios between non-exchange conversions spanning the tested marker compared to crossovers was highly significant (two-tailed Fisher-exact, p=0.0002).

4.4.4.2 Only non-exchange conversions spanning the candidate markers show TD

At both Superhotspots, TD was concentrated at a single SNP heterozygosity, and additional non-exchange conversions not spanning either of these SNPs did not show significantly biased transmission distortion, as illustrated in Figure 4-4.

Figure 4-4 Allele transmission in non-exchange conversions not spanning K7.4 or F6.1.



Observed transmission frequencies in non-exchange conversion molecules at markers not spanning either marker K7.4 or F6.1 are shown differentiated by progenitor haplotype.

Non-exchange conversions that did not span marker K7.4 at Superhotspot K showed a transmission ratio of 60:40, which was not significantly different from 50:50 (exact binomial, p=0.289). The same was observed at Superhotspot F, non-exchange conversions not spanning F6.1 did not show any bias with only 53% involving transfer of markers from the F6.1 G haplotype (two-tailed exact binomial, p= 0.644).

The level of TD seen in non-exchange conversions, which spanned marker F6.1, was significantly different to the transmission seen in non-exchange conversions not spanning F6.1 (two-tailed Fisher-exact, p=0.018). Curiously, non-exchange conversions spanning marker K7.4 at Superhotspot K showed comparable transmission frequencies to those non-exchange conversions not spanning marker K7.4 (two-tailed Fisher-exact, P=0.231).

Men homozygous at K7.4 were tested for biased non-exchange conversion at marker K7.5. Both men showed no significant TD at marker K7.5. Man d21 who was homozygous for the T-allele showed a 57:43 transmission ratio in favour of the C-allele, not

significantly different from 50:50 (binomial probability, P=0.585). In d67, homozygous for the K7.4 C-allele, the K7.5 C-allele was found in 61% of non-exchange conversions, but this was equally not significantly different from 50:50 (two-tailed exact binomial, P=0.153). The transmission ratios at K7.5 were comparable between both men (two-tailed Fisher-exact, P=0.814).

4.5 Summary and Discussion

The analysis of allele transmission frequencies in crossover as well as detectable non-crossover recombinants has detected incidences of biased gene conversion. These biases could be grouped into two categories. Firstly biased gene conversion affecting non-crossovers and crossovers to the same degree, and secondly, a bias that was significantly different between crossovers and non-crossovers.

4.5.1 Biased gene conversion affecting crossovers and non-crossovers

Weak evidence for reciprocal crossover asymmetry and TD was observed at two hotspots. Although an excess of alleles could not be significantly established for non-exchange conversions within the given datasets, the transmission ratios into crossover and non-exchange conversion were comparable. Significant biased transmission of markers into both crossovers and non-exchange conversions combined was found to occur at hotspots F and H, with weak indication of biased transmission at Superhotspot T. In these hotspots reciprocal crossover products mapped to somewhat different intervals with A/B displacement of under 200 bp (187±12 bp) and only weak transmission distortion. Modest TD has been observed at Superhotspots before, including at Superhotspots E and F (Webb *et al.* 2008). The strong transmission distortion reported for human hotspots by Jeffreys and Neumann (2002, 2005 and 2006) resulted from crossovers in reciprocal orientation mapping to intervals displaced by about 400 bp (397±39 bp) (Jeffreys and Neumann 2002; Jeffreys and Neumann 2005; Neumann and Jeffreys 2006).

At Superhotspot F, off-centre marker F6.6b displayed TD into crossover. This marker is located 472 bp away from the mean centre of the crossover hotspot but nevertheless shows significantly distorted transmission ratios. All men tested for transmission ratios at F6.6b show an overtransmission of the G-allele in crossovers progeny, pointing to it being a true phenomenon despite the off-centre location. This includes d11 as well as two men

analysed via a full-crossover assay by A.J. Jeffreys (Webb *et al.* 2008). This was the first observation of significant TD at a marker not located directly at the centre of the hotspot, but it is highly significant compared across all men, and at d11 who was tested for non-exchange conversion the G-allele is equally overtransmitted in non-exchange conversions.

Intriguingly, the central marker F6.1 did not show any indication for TD in crossover in any of the men but was instead involved in biased non-exchange conversion with TD purely into non-crossover recombinants (4.5.2) (This will be discussed further in the Final Discussion (7.3.2).)

4.5.1.1 Reciprocal crossover asymmetry and reduced crossover frequencies

As observed initially in Chapter 3 (Table 3-3) the frequency of crossovers at hotspot T was significantly lower in d25 than expected from crossover analysis on three men tested by A.J Jeffreys (Berg *et al.* 2010). Man d25 had also shown a lower crossover frequency than d52 analysed using the same half-crossover assay at Superhotspot H (Table 3-3). A comparable decrease in crossover frequencies was also observed for d11 at Superhotspot F as well as d17 at Superhotspot E. All of these men had shown mild 2-4 fold reduced frequencies. Differences in recombination frequencies of this magnitude may be explained by subtle variation in the efficiency of PCR amplification between assays, however d25 at Hotspot H and d11 at Hotspot F showed indication for reciprocal crossover asymmetry.

From previous work on *cis*-regulating factors it was known that TD likely occurs through biased initiation on a more active haplotype, leading to over-transmission of alleles from the less active haplotype (Jeffreys and Neumann 2002; Jeffreys and Neumann 2005). Men homozygous for the activating SNP allele had high crossover frequencies, in contrast men homozygous for the suppressing SNP allele had reduced crossover frequencies (Jeffreys and Neumann 2002). Men heterozygous for the suppressing and the activating allele displayed reciprocal crossover asymmetry (Jeffreys and Neumann 2005). Though because they had only one activating allele, their crossover frequencies were reduced compared to men homozygous for the activating allele. Therefore a reduction of recombination frequencies would be consistent with observations of reciprocal crossover asymmetry through biased initiation.

Men for whom indication of reciprocal crossover asymmetry was observed displayed a somewhat reduced crossover frequency, consistent with active and inactive haplotypes operating at the given hotspots.

However, a clear association of specific SNP heterozygosities with active and inactive haplotypes is not possible at all of the hotspots.

Discrepancy between what would be the active and suppressed haplotype and crossover frequencies were observed at Superhotspot H. Two men were analysed at Superhotspot H, namely d25 and d52. Crossovers detected in sperm DNA from d25 displayed reciprocal crossover asymmetry with significant TD observed at SNP H7.6. Here the C-allele was significantly overtransmitted into both types of recombinants. In contrast to d25, the C-allele at SNP H7.6 was found in only 47% of crossovers of d52, a proportion very close to 50:50 (two-tailed Fischer-exact, P=0.647). One of the men analysed by Webb et al, is equally heterozygous at H7.6, and had shown transmission ratios of 51:49 in favour of the C-allele (Webb *et al.* 2008), a deviation not significantly different from 50:50 (two-tailed Fisher-exact, P=0.674). It appears that the haplotype carrying the C-allele at H7.6 is an inactive haplotype in d25, but not in d52 or the man analysed by (Webb *et al.* 2008).

Man d52 had a higher than expected crossovers frequency, compared to d25 and the data by (Webb *et al.* 2008). This is most curious. The central markers that show non-exchange conversion are heterozygous in all three men, but men differ between haplotypes away from the centre, thus making it impossible to narrow down potential other *cis*-DNA sequence differences that could account for the disparity. Jeffreys and Neumann had observed biases where men would show opposite TD, with no clear association of haplotypes to activating and suppressing alleles (Jeffreys and Neumann 2009).

4.5.1.2 Men displaying reduced frequencies

At Superhotspot F significant reciprocal crossover asymmetry was observed at marker F6.6b in d11 and all men heterozygous at the same SNP analysed by A.J. Jeffreys (Webb *et al.* 2008). Man d23 showed a marked reduction of crossover frequencies at this hotspot, however he is equally heterozygous at F6.6b and therefore not homozygous for the F6.6b G allele marking what would be the inactive haplotype.

Similarly, one of the men (d25) analysed at Superhotspot T had shown a 2-fold reduction while the second man, d60, had shown a 50-fold reduced crossover frequency, compared to the frequencies seen by initially by A.J. Jeffreys (Berg *et al.* 2010). Significant reciprocal crossover asymmetry was not observed at this hotspot. Therefore a simple model of *cis*-inactivating SNPs cannot explain the strong degree of variation that was observed between crossover frequencies at Superhotspot F or Superhotspot T. Neumann *et al.* (2006) have also observed significant variation in crossover profiles and in crossover frequencies between men that was independent of local sequence variation at hotspot MSTM1b (Neumann and Jeffreys 2006). The phenomenon of reduced frequencies will be investigated further in Chapter 5.

Six of the 19 Superhotspots from Webb *et al.* had displayed TD in crossovers, with two displaying allelic ratios as strong as 80:20 in favour of the suppressing allele (Webb *et al.* 2008), (A.J Jeffreys, *unpublished data*). An approach targeting only these hotspots that had already shown indications of reciprocal crossover asymmetry, and then analysing fewer hotspots but at higher depth would probably have returned more significant results that could clarify how pronounced the biases that affect both crossover and non-crossover recombinants are in several men. But by instead using an unbiased approach, without pre-selecting hotspots it was possible to identify a previously unknown phenomenon – transmission distortion purely into non-exchange conversions. This is the first direct evidence in man that crossover and non-crossovers might to some extent be regulated independently.

4.5.2 Biased gene conversion specifically affecting non-crossovers

In contrast to the biased initiation phenomenon described earlier, biased transmission of markers specifically into non-crossovers was observed at two recombination Superhotspots. Non-exchange conversions showed significant TD at the candidate markers at both hotspots, while crossovers show no indication of reciprocal crossover asymmetry. This observation may indicate separate pathways generating non-crossover and crossovers in humans, with crossovers mainly produced by the DSBR pathway and non-crossovers produced preferentially by an SDSA pathway, consistent with observations in yeast recombination (Allers and Lichten 2001a), which will be discussed further in Chapter Seven.

Theoretically the observed TD in non-exchange conversions could be either caused by biased initiation or biased heteroduplex repair. It is unlikely that bias specifically affecting non-crossovers results from initiation bias. In biases affecting crossovers and non-crossovers simultaneously, biased initiation resulted from one haplotype being less active or inactive. Hence men were also reduced in their recombination frequencies. At Superhotspot K, men homozygous for either over-transmitted or under-transmitted alleles have comparably high recombination frequencies, as high as men who were heterozygous. The lowest recombination frequencies were observed in d35, who was analysed at both Superhotspots and was heterozygous at each of the candidate regulatory SNPs. Transmission distortion therefore unlikely results from initiation bias but instead by biased mismatch repair of heteroduplex DNA in recombination intermediates destined to become non-crossover recombinants. For biases in mismatch repair to occur, repair would have to be strongly biased in favour of correction to one allele over the other and the mismatch machinery would have to be able to detect whether recombination intermediates are destined to become crossover or non-crossover. This will be discussed further in the Final discussion (Chapter Seven, 7.4)

At Superhotspot F, conversions spanning marker SNP F6.1 and conversions of additional markers were significantly different to each other. Here adjacent marker F6.0 did not display biased transmission on its own. In contrast at Superhotspot K, even though the significant TD purely into non-exchange conversion was only observed at K7.4 heterozygote men, TD observed at adjacent marker SNP K7.5 were not significantly different to the TD seen only at K7.4.

The discovery of biased gene conversion specifically affecting non-crossover also means that previous estimates based on biased gene conversion associated with crossover may be underestimating the true extent and impact of meiotic drive. Hotspots that appear to be churning out balanced recombinant progeny, when analysed only for crossovers, may in fact harbour biased gene conversion generated by non-crossover recombination.

Chapter Five Analysis of *cis*- and *trans*-acting factors that influence recombination activity at Superhotspot T

5.1 Introduction

An unexpectedly high variation in crossover frequencies between men analysed via the half-crossover assay had been observed in Chapter 3. Particularly low recombination frequencies were observed for one of the men analysed at each Superhotspots F and T, which had excluded them from the study of non-exchange conversion profiles.

High sequence similarity between men showing higher and lower crossover frequencies appeared to exclude a *cis*-acting SNP as regulator of hotspot activity at both Superhotspots. The question remained whether there are any other elements within the Superhotspots that could influence initiation? Or is there a candidate *trans*-regulator that may be responsible instead? Two candidate factors for recombination suppression were identified.

5.1.1 Candidate factors for recombination suppression

5.1.1.1 Could an STR influence recombination in *cis* at Superhotspot T?

At Superhotspot T, a short tandem repeat (STR) is located only 47 bp away from the centre of the hotspot. Recombination hotspots are enriched for STRs in yeast (Bagshaw *et al.* 2008). Repeat DNA is also associated with recombination hotspots in humans. Myers *et al.* (2005) observed a significant correlation between LD hotspots and repeat elements such as THE1A and THE1Bs and found that CT-and GA-rich repeats are highly over-represented in LD hotspots while TA and GC rich repeats were under-represented (Myers 2005). It was therefore investigated whether this STR with its 2 bp TG -periodicity influences recombination at Superhotspot T.

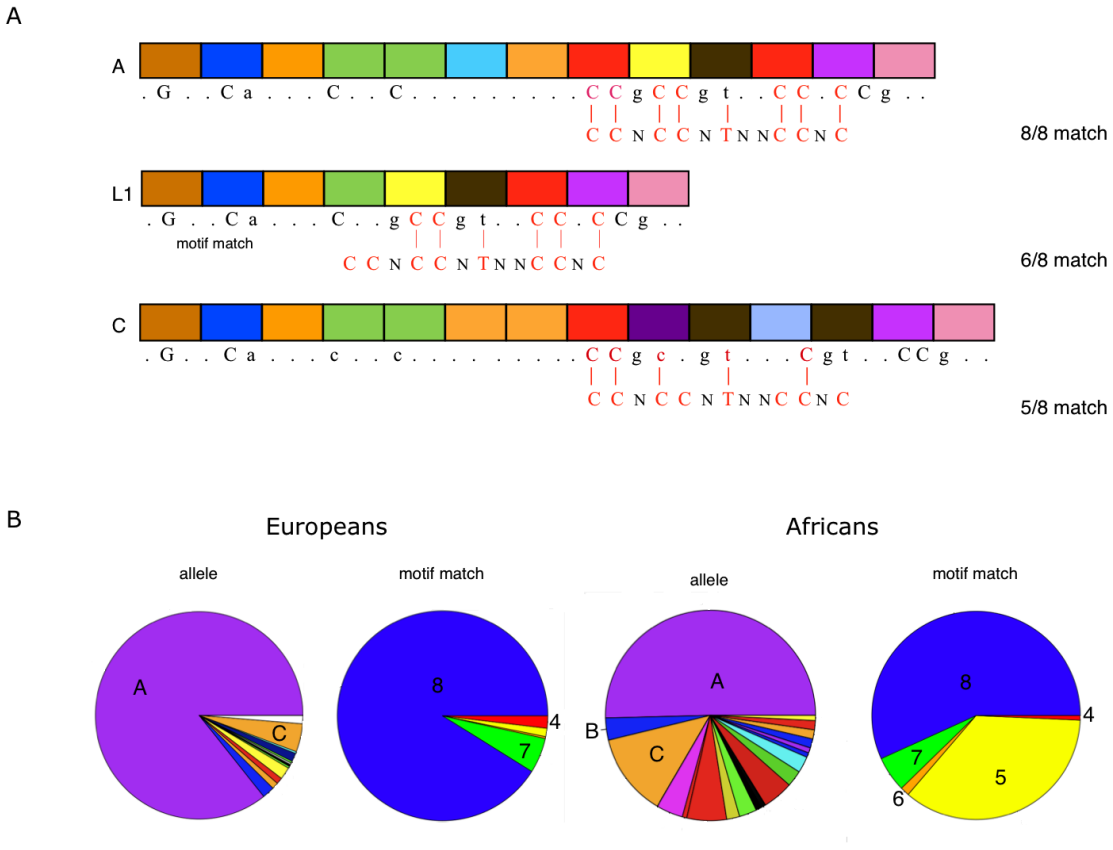
5.1.1.2 *Trans*-regulation by *PRDM9*

A second candidate for regulation was the *trans*-regulating factor *PRDM9*. At the time when the high variation in crossover frequencies had been observed between men at Superhotspots F and T, *PRDM9* had just been identified as regulator of hotspot usage in mice and humans (Baudat *et al.* 2010). The protein encoded by *PRDM9* consists of an N-

terminal KRAB domain and a SET-histone methyl-transferase followed by a zinc-finger domain. The most common variant in Europeans, variant A, has 13 C2H2 zinc-fingers (ZnF) each encoded by an 84 bp repeat of a minisatellite (Hayashi *et al.* 2005). Every ZnF has four predicted key DNA contact residues at positions -1, 2, 3 and 6 of the α-helix domain (Baudat *et al.* 2010). Binding predictions based on these residues had implicated the last six ZnFs of the PRDM9 A-variant as responsible for binding to the eight non-degenerate bases of the 13-mer CCNCCNTNNCCNC hotspot motif. This hotspot motif, henceforth referred to as the A-motif, can be found in about 40% of recombination hotspots (Myers *et al.* 2010). The first direct evidence for *PRDM9* regulation in humans came from a study where pedigreed Hutterite individuals, who carry a PRDM9 variant I, show a genome wide shift in LD-hotspot usage (Baudat *et al.* 2010).

A. Jeffreys, R. Neumann and I. Berg had carried out genotyping of *PRDM9* alleles as described in (Berg *et al.* 2010). *PRDM9* alleles were classified for men in our semen donor panels based on the number of ZnF repeats they encoded, and also on the predicted recognition capability of these ZnFs with respect to the A-motif. All reported alleles, except allele I seen in Hutterites (Baudat *et al.* 2010), were also present within the European semen donor panel. In addition 24 new *PRDM9* alleles were identified (Berg *et al.* 2010). The variability of *PRDM9* was low in European men, with only 20 men of the original donor panel of 96 men carrying variant alleles in addition to allele A, and most of these variants had predicted binding properties that were identical to the most common PRDM9 A-variant (Figure 5-1). In contrast *PRDM9* variability was much higher in an African donor panel. The lengths of ZnF arrays encoded by *PRDM9* alleles found in the African population ranged from eight to eighteen repeats and many had predicted recognition capability for as little as five out of the eight non-degenerate sites of the A-motif (Berg *et al.* 2010) as illustrated in Figure 5-1.

Figure 5-1 *PRDM9* ZnF diversity



(A) PRDM9 variants A, L1 and C, with variant structure indicated by coloured boxes (modified from (Berg *et al.* 2010; Berg *et al.* 2011)). The predicted DNA binding sequence is shown below, with dots indicating weakly predicted bases and uppercase letters indicating the most strongly predicted bases. DNA binding sequences are aligned with the A-motif. Vertical lines indicate binding predictions that match the A-motif of *PRDM9*. The binding sequence for variant A matches all eight specified bases in the motif, while variant L1 matches at best only six of the eight bases and variant C at best only five of the eight bases.
(B) *PRDM9* ZnF diversity within the European and African donor panels. Alleles were classified both by structure and by the strengths of the predicted match with the A-motif (from (Berg *et al.* 2010))

Comparison of *PRDM9* genotypes of the men analysed at both hotspots revealed that the men with lower recombination frequencies carried alleles in addition to allele A. Man d23, who had displayed low recombination frequencies at Superhotspot F was heterozygous for *PRDM9* alleles A and L1. Allele L1 has only 9 ZnFs comprising the first four and last five of the 13 ZnFs of allele A, and only a predicted 7/8 match to the hotspot motif (Figure 5-1 A). Man d60, who had shown lower recombination frequency at Superhotspot T, was heterozygous for *PRDM9* alleles A and C. Variant C has 14 ZnF binding domains, and was predicted to bind the A-motif at only 5 out of the 8 non-degenerate positions (Figure 5-1 A).

Analysis of the suppression data in light of these findings made *PRDM9* a candidate for regulation, at least for Superhotspot F. This Superhotspot had the A-motif located within 47 bp of the hotspot centre, which made it an ideal candidate to test the influence of PRDM9 on recombination at the individual level. Extensive crossover analysis of 32 men, carried out by A. J. Jeffreys, showed a significant regulation by *PRDM9* at Superhotspot F and other hotspots (Berg *et al.* 2010). Hotspots containing a hotspot-activating motif showed high sperm crossover activities in men homozygous for the motif-recognising PRDM9 A-variant. Conversely, men lacking motif-recognising PRDM9 variants were suppressed in recombination frequency. Consistent with an additive model of PRDM9 regulation, it was observed that men with one A-allele displayed about half of the cross-over frequency seen in A/A men, and men with no A-alleles were usually fully suppressed in recombination (Berg *et al.* 2010).

At Superhotspot T, the three men analysed initially by A.J. Jeffreys (Berg *et al.* 2010) as well as d25, analysed for Chapter 3 displayed crossover frequencies of 0.45% - 1.33% and were homozygous for the A-allele of *PRDM9*. However, Superhotspot T does not contain an apparent A-motif. Could the reduced crossover activity, observed for the *PRDM9* A/C heterozygous man d60 at Superhotspot T, be due to PRDM9 regulation as observed at Superhotspot F, but independent of the motif? If PRDM9 regulates Superhotspot T, then it should be activated by the A-variant, even though the A-motif is absent.

5.2 This work

Two candidate factors for crossover suppression had been identified for Superhotspot T, Firstly a STR as a potential *cis*-acting factor, and secondly PRDM9 as potential *trans*-acting factor. Multiple men have to be analysed to explore possible influences of STR and *PRDM9* status on crossover frequencies. The half-crossover assay is very laborious as every SNP has to be addressed individually, in addition it can only test very few men at a time as it is dependent on high marker density within the hotspot to detect crossover and non-exchange conversion molecules. A more efficient method of crossover analysis is the crossover assay, which relies on nested repulsion-phase allele-specific PCR to selectively amplify crossovers from sperm DNA. A larger number of men could be tested with the crossover assay and crossover frequencies of these men, in addition to the men analysed previously, were then correlated with genotypes of the potential regulating factors. Some of the work described in this chapter has been published (Berg *et al.* 2010).

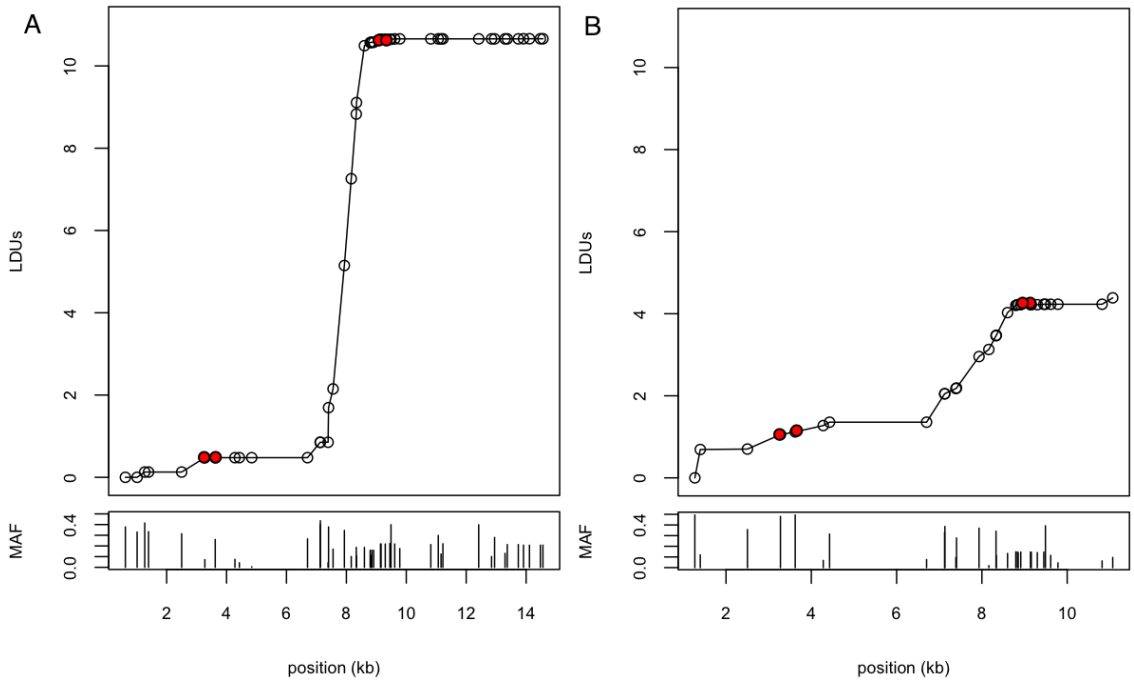
5.3 Analysis of crossover frequencies using a full-crossover assay

To test whether any of the candidate factors influence crossover frequencies at Superhotspot T, a more efficient conventional full-crossover assay was used. Two sets of heterozygous selector sites located within LD blocks flanking both sides of the hotspot have to be identified and optimised to allow efficient crossover detection in as many men as possible. Comparing crossover frequencies across many men should smooth out factors such as variable DNA quality and possible *cis*-acting variants, in addition to the tested variables. To allow many PRDM9 variants to be tested the donor panel was extended to include 74 men of Afro-Caribbean and Zimbabwean descent, who displayed a higher diversity of *PRDM9* alleles compared to European men (Berg *et al.* 2010).

5.3.1 Choosing selector sites to analyse multiple men

To identify multiple men heterozygous at selector sites, genotyping data were established and compared. As Superhotspot T had initially been identified by high-resolution mapping of crossovers in Hutterites, a population of European descent (Coop *et al.* 2008) and then validated as a sperm crossover hotspot in the European donor panel (Berg *et al.* 2010) it was not known if Superhotspot T was active in Africans. So by doing metric LD analysis of Europeans and Africans separately it was tested if both populations show a historical hotspot. Metric LD mapping on un-phased diploid genotyping data, including all HapMap and dbSNP markers plus SNPs found through re-sequencing in the given interval, determined the rate of LD breakdown across Superhotspot T. The resulting LD maps for both populations are displayed in Figure 5-2.

Figure 5-2 Metric LD maps in LD-units across Superhotspot T



Metric LD maps established from diploid un-phased genotyping data from (A) 156 Europeans and (B) 74 Africans from the African and European donor panels respectively. Lengths of ticks in the bottom panels represent minor allele frequencies (MAFs) of the SNPs within the populations tested.

The LDU-profiles across hotspot T in Figure 5-2 illustrated that both populations had an increase of LD across the target interval, implying that historical recombination has taken place. The LDU step of the African population was 3-fold reduced compared to the European population (Figure 5-2), which indicated that even though historical recombination has likely occurred in both populations; it has not necessarily influenced LD to a similar extent. The marker T7.9 (rs9854419), established as the most central SNP within Superhotspot T in Chapter 3 (3.4.1.2), is located in the middle of the greatest LDU step in both the European and African populations.

Several sites became apparent as potential selector sites, located outside of the hotspot determined by (Berg *et al.* 2010). The largest number of donors were heterozygous at markers T3.3 A/G and 3.6 T/C as well as T9.1 T/C and T9.3 A/G, therefore these markers were chosen as selector sites for crossover analysis (marked by red dots in Figure 5-2). These markers lie within LD blocks on each side of Superhotspot T, though there was a small increment in LD between markers T3.3 and T3.6 in Africans.

Markers T9.1 and T9.3 are in absolute LD with two haplotypes (TA and CG) present at haplotype frequencies of 78.3% and 21.7% respectively in Europeans as well as 85.4% and 14.6% respectively in Africans. Markers T3.3 and T3.6 are in complete LD in the European population, with 3 haplotypes (AT, GC and AC) present at haplotype frequencies of 73.4%, 11.4% and 15.2%, respectively; however in Africans, all four haplotypes were observed.

5.3.2 Men amenable to crossover detection

Twelve men were amenable to crossover detection using selector sites T3.3 and T3.6 for forward, as well as selector sites T9.1 and T9.3 for reverse allele specific primers (ASPs) (summarised in Table 5-1).

Table 5-1 Genotypes of men selected for crossover analysis at Superhotspot T

		donor	4	24	43	80	173	177	235	267	268	269	278	281
		PRDM9 Genotype												
		Allele1	A	A	L20	A	L7	A	L4	C	A	L11	A	A
		Allele 2	A	A	A	A	A	A	L21	L14	A	A	A	A
SNP														
rs number	Name	location	genotype											
17013514	T3.3A/G	3276	H	H	H	H	H	H	H	H	H	H	H	H
6808240	T3.6T/C	3622	H	H	H	H	H	H	H	H	H	H	H	H
1976359	T4.3T/C	4280	C	H	C	C	C	C	C	C	?	C	C	C
17013524	T4.4T/G	4428	H	T	H	T	H	H	T	?	T	T	H	T
66595506	T6.6T/G	6703	T	H	T	T	T	T	T	T	H	T	T	T
9830036	T7.1C/T	7125	T	H	T	T	H	T	T	T	H	T	T	H
11714875	T7.1aT/G	7132	H	H	H	G	T	H	H	H	T	G	H	T
11926996	T7.4G/A	7390	H	G	H	G	G	H	G	G	G	G	H	G
3856841	T7.4aG/A	7405	H	H	A	G	A	A	A	A	A	H	A	H
9854419	T7.9A/G	7933	A	G	H	G	H	H	?	H	H	A	G	A
1827451	T8.3G/A	8331	G	G	H	G	G	H	G	G	G	H	A	G
17013574	T8.3aT/A	8338	H	T	H	T	A	T	T	H	T	H	T	T
35068617	T8.8aC/T	8831	H	T	H	T	H	H	H	H	H	H	H	H
12636785	T8.8T/C	8852	H	T	H	T	H	H	H	H	H	H	H	H
12636800	T8.9T/C	8912	H	T	H	T	H	H	H	H	H	H	H	H
12636837	T9.1T/C	9139	H	H	H	H	H	H	H	H	H	H	H	H
12636283	T9.2A/C	9158	H	H	H	H	H	H	H	H	H	H	H	H
66906407	T9.3A/G	9299	H	H	H	H	H	H	H	H	H	H	H	H
1032784	T9.4G/A	9452	G	H	G	H	G	G	G	G	G	G	G	H
1502567	T9.5T/C	9490	H	T	H	T	H	H	H	H	H	T	H	T
11456124	T9.6-/+	9614	H	H	H	H	H	H	H	H	H	H	H	H

Donor numbers and *PRDM9* status are shown on the top. SNP names and rs-numbers are displayed in the left, followed by SNP location in bp within the target interval, SNPs located within the 95% width of the hotspot are highlighted in yellow. Genotype of each marker is shown, with homozygous alleles indicated by the respective allele, heterozygous SNPs by "H" and inconclusive genotyping results indicated by "?".

As summarised in Table 5-1, seven men homozygous for allele A of *PRDM9* as well as five men with *PRDM9* alleles in addition to allele A could be analysed using the chosen selector sites.

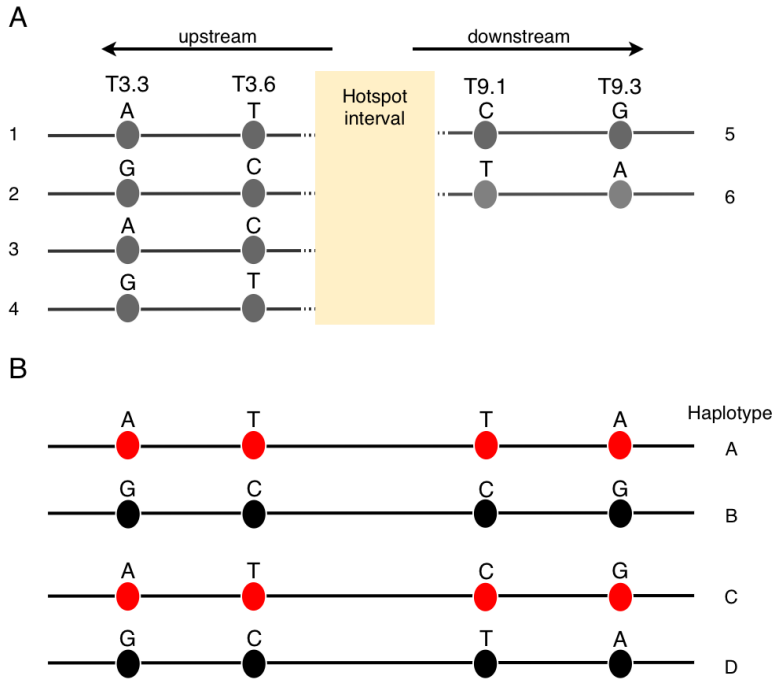
5.3.3 Primer optimisation and phasing of selector sites

Forward ASPs were developed for selector sites T3.3 A/G as well as T3.6 T/C and reverse ASPs were designed specific to selector sites T9.1 T/C and T9.3 A/G. Primers were tested on multiple-displacement amplified (MDA) DNA of men homozygous for either the desired allele (positive control) or for the alternative allele (negative control) at a range of annealing temperatures. This was used to determine at which temperatures PCR product was generated efficiently, but only from PCR reactions that contained DNA carrying the desired allele. PCR reactions with DNA carrying the alternative allele must have remained negative for PCR product at the same temperatures. ASPs designed for all selector sites were found to be highly specific and worked efficiently at a range of temperatures. Primer sequences and temperature ranges are collected in Appendix I.

5.3.4 Phasing of selector alleles

Analysis of population LD had revealed that only the reverse set of alleles was in strong LD, with the forward set of alleles in full LD in Europeans, but not Africans. This LD pattern left four possible upstream and two possible downstream haplotypes, which are illustrated in Figure 5-3 A. The linkage phase of forward and reverse set of alleles, as well as between alleles at markers T3.3 and T3.6 had to be determined experimentally in each donor selected for crossover analysis (Table 5-1). To test which of the possible eight composite haplotypes were present a phasing assay using T3.3FA and T3.3FG together with T9.1RT and T9.1CR as well as T3.6FT and T3.6FC together with T9.3RA and T9.3RG was carried out. Four composite haplotypes were detected and are displayed in Figure 5-3 B.

Figure 5-3 Phasing of selector alleles identified from LD



(A) Four upstream (1-4) and two downstream (5-6) haplotypes were found in the semen donor panels. Linkage is represented by line, with SNP alleles represent by circles, and the allele indicated above each circle. (B) Four composite haplotypes (A-D) were found in the tested men. Haplotypes are arbitrarily labelled to red and black.

As shown in Figure 5-3, eight haplotypes were present in the donor panel; with only four composite haplotypes present in the men amenable to crossover detection (Figure 5-3, B). Men d43, d173, d278 and d281 had haplotypes A and B and the remaining eight men, d4, d24, d80, d177, d235, d267, d268 and d269, had haplotypes C and D.

5.3.5 Crossover assays

Crossover assays were set up with forward and reverse primers in repulsion-phase. Primary PCR amplification of pools of sperm DNA used the outside primers and secondary amplification of the primary PCR product was performed using inside nested primers, which improved specificity and PCR yield. Two separate crossover assays were used to analyse men, depending on the haplotype they carried, with each assay testing only one orientation of recombination as summarised in Table 5-2.

Table 5-2 Crossover assays used to test men with different haplotypes

Assay	haplotypes	PCR	Primer combination	men analysed
1	A and B	primary	T3.3FA & T9.3RG	d43, d173, d278 and d281
		secondary	T3.6FT & T9.1RC	
2	C and D	primary	T3.3FG & T9.3RG	d4, d24, d80, d177, d235, d267, d268 and d269
		secondary	T3.6FC & T9.1RC	

As summarised in Table 5-2, the first assay analysed men carrying haplotypes A and B. and the second crossover assay analysed men with haplotypes C and D. PCR conditions of both crossover assays are collected in Appendix VII.

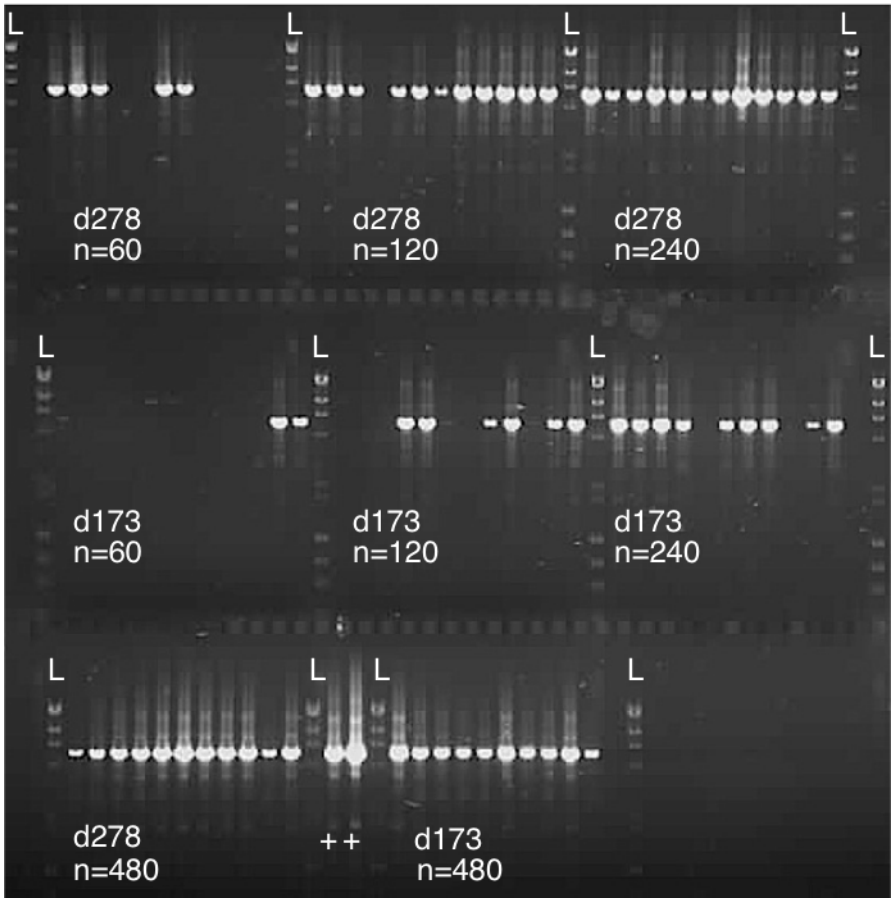
Four different pool sizes of 60, 120, 240 and 480 amplifiable sperm DNA molecules were set up for each man, chosen such that, at the mean crossover frequency of 1% seen previously at this hotspot, they will contain 0.6, 1.2, 2.4 and 4.8 crossover molecules, respectively.

To minimise variation in amplification results, a universal PCR master-mix was made up for two plates. Two men were analysed on the same PCR plate, using twelve aliquots for each pool size, except for the largest pool size. For both of the men, the 12^{th} pool was replaced with DNA from a man with the selector SNPs in coupling phase in concentrations of 0.1 ng or 1 ng, respectively, to function as positive control. These amounts were chosen so that eight and eighty molecules of the correct haplotype would amplify at an expected 50% single molecule efficiency. In reality, numbers of molecules will fluctuate between pools, but with this input the chance of adding no amplifiable molecules to a single positive control was minimised.

5.3.6 Determining crossover frequencies

Crossover frequencies were estimated from the number of pools positive for PCR product, detected by ethidium bromide staining after gel-electrophoresis. Positive PCR reactions indicated the presence of at least one crossover molecule in the original pool of sperm DNA. An example of crossover detection for two men, d173 and d278, is shown in Figure 5-4.

Figure 5-4 crossover detection after repulsion-phase allele specific PCR



Secondary PCR products of a crossover assay for men d278 and d173 were analysed by agarose gel electrophoresis followed by staining with ethidium bromide. All four analysed pool sizes are shown for each man, with n=numbers of amplifiable molecules of each progenitor haplotype in the respective pool. Pools are separated by wells with a size standard (L) equalling 30 ng λ DNA×HindIII and 20 ng ΦX174 DNA×HaeIII. Pools that contained crossovers amplified with primers in repulsion-phase, and appeared as white bands after ethidium bromide staining. Positive controls including sperm DNA from a man with primers in coupling phase are indicated by (+) with 0.1ng sperm DNA input on the left and 1 ng on the right.

Pools containing fewer molecules were less often positive for PCR product, while pools containing more molecules were more often positive for crossovers (Figure 5-4). For each of the analysed men, crossover frequencies were estimated by Poisson-approximation on the proportion of PCR reactions that were negative; maximum-likelihood software (written by A. J. Jeffreys) was used to combine data across all input pool sizes.

For men with low crossover frequencies the analysis was repeated at larger pool sizes. For two men, d235 and d267, follow-up assays in which the number of molecules was 10-fold increased did not yield any pools positive for crossover molecules either. Positive controls had consistently shown the expected strengths, and there was no indication of low sperm

DNA quality; each sperm DNA sample had shown good amplifiability in the phasing PCRs that used the same PCR conditions of the final crossover assays as well as the same primers, albeit in different combinations. The absence of PCR product from both assays was therefore trusted not to be a result of consistent PCR failure or low DNA quality, but instead a representation of fully suppressed crossover frequencies in the analysed men.

In contrast, all PCR pools were positive for PCR product in two subsequent assays for d80, including fewer molecules input. This man could be displaying a substantially higher crossover frequency than any of the other men, with a minimal crossover frequency of at least 4.28% per sperm. This value is compatible with the upper confidence level of 4.72 % crossover per HGE observed for d268. However, d80 was excluded from the dataset as this result could have also arisen from mis-genotyping at one of the selector sites.

5.4 Results

Crossover frequencies varied substantially between analysed men, from less than 0.03% to 2.53%, the highest crossover frequency observed at a human hotspot at the time. To analyse the effect of *PRDM9* status and STR allele length, these frequencies as well as crossover frequencies obtained in a full crossover assay by A.J. Jeffreys and frequencies obtained using the half-crossover assay in Chapter 3 were included in the analysis, and are summarised in Table 5-3.

Table 5-3 crossover frequencies obtained in different assays

Assay	donor	population	Crossover frequency per sperm		
			mean	lower CI	upper CI
Crossover assay 1	43	British	1.42	1.07	1.69
	173	Afro-Caribbean	0.52	0.21	0.59
	278	Zimbabwean	1.51	0.84	2.68
	281	Zimbabwean	1.86	1.05	3.22
Crossover assay 2	4	British	1.59	0.89	2.79
	24	British	0.36	0.80	2.44
	177	Afro-Caribbean	1.13	0.29	0.70
	235	Zimbabwean	0.00	0.00	0.05
	267	Zimbabwean	0.00	0.00	0.03
	268	Zimbabwean	2.53	1.34	4.79
	269	Zimbabwean	0.43	0.24	0.72
Full-crossover assay (A. J. Jeffreys)	2	French	1.33	0.64	1.07
	24	British	0.36	0.80	2.44
Half-crossover assay	25	British	0.46	0.002	0.08
	60	British	0.02	0.30	0.85

Men are grouped by the assay with which they were analysed. Mean crossover frequency as well as lower and upper 95% confidence intervals (CI) were determined by Poisson approximation.

As shown in Table 5-3 crossover frequencies obtained in crossover assay 1 and 2, using different combinations of allele-specific primers, were comparable (two-tailed t-test, P=0.397). Equally, crossover frequencies of the two men analysed with the half-crossover assay in Chapter 3, and crossover frequencies of two men analysed by A.J. Jeffreys (Berg *et al.* 2010) fall within the variation observed using these crossover assays (two-tailed t-tests, P=0.770 and P=0.220, respectively).

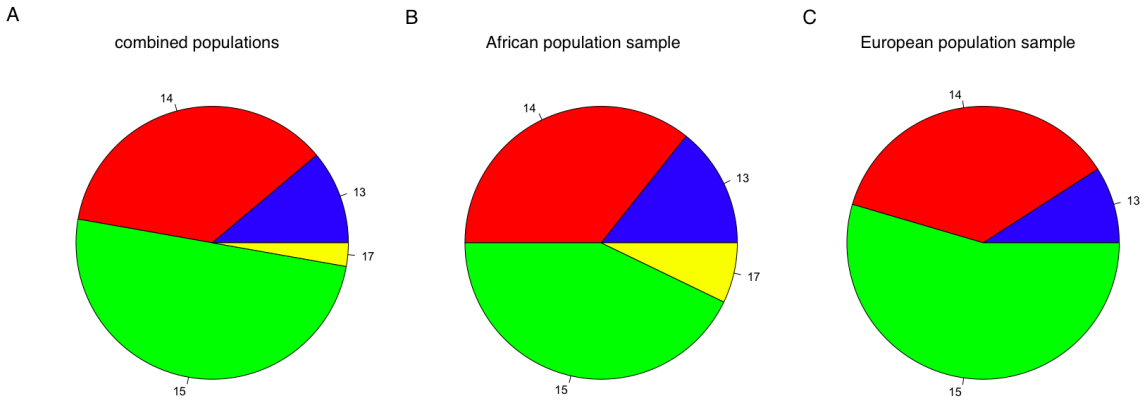
Men from African and Afro-Caribbean descent were not biased towards low crossover frequencies, as might have been predicted from LDU profiles (Figure 5-2). Instead they displayed similar degrees of variability in crossover frequencies, and both the highest and lowest crossover frequencies were observed in men of Zimbabwean descent.

5.4.1 STR variability and crossover frequencies

To test whether the STR has an influence on the crossover frequency, it was typed in all 15 men for whom crossover frequencies have been obtained (Table 5-5). A target spanning the STR was amplified from sperm DNA, using a hexachlorofluorescein (HEX)-

tagged primer. Fluorescent signal was then detected via capillary electrophoresis and STR lengths were determined using GeneMapper (Applied Biosystems) correlating the peaks of fluorescence with the expected PCR product lengths. STR alleles lengths were then classified by the number of pure TG repeats present in the 3′-region of the STR. The STR varied in length between 13 and 17 repeats. The longest allele displayed an anomalous size increment of 1 instead of 2 bps, but was made up of at least 17 repeats. The African population sample displayed higher variability in repeat lengths compared to the European population sample (Figure 5-5).

Figure 5-5 STR repeat length variability



STR lengths, classified based on a 3´- pure TG repeat region of the STR, differentiated by colour: TG_{13}, blue; TG_{14}, red; TG_{15}, green and $TG_{17?}$,yellow. Allele frequencies within the respective population sample: (A) all tested men, (B) the African population sample or (C) the European population sample.

Of the men originally analysed at Superhotspot T, the man with the higher crossover activity, d25, carried STR alleles with 13 di-nucleotide-repeats TG_{13} and 14-dinucleotide repeats TG_{14}. The man with lower crossover activity (d60), was homozygous for an even longer 15-dinucleotide STR allele TG_{15}, which is the human consensus allele. The hypothesis therefore was that longer STR alleles might function as suppressors of recombination if the STR indeed acts as *cis*-regulator of recombination. STR lengths and crossover frequencies for all men are summarised in Table 5-4.

Table 5-4 Sperm crossover frequencies and their relationship with STR-length alleles

donor	population	STR alleles		Crossover frequency per sperm		
				mean	lower CI	upper CI
268	Zimbabwean	14	14	2.535	1.340	4.792
281	Zimbabwean	14	15	1.863	1.051	3.224
4	British	15	15	1.591	0.889	2.786
278	Zimbabwean	14	15	1.514	0.842	2.676
43	British	13	15	1.418	0.804	2.444
2	French	14	15	1.354	1.067	1.694
83	British	14	15	0.836	0.641	1.073
177	Afro-Caribbean	14	14	0.518	0.299	0.848
173	Afro-Caribbean	15	17	0.500	0.342	0.710
25	British	13	14	0.464	0.288	0.699
269	Zimbabwean	13	15	0.435	0.245	0.722
24	British	14	15	0.362	0.207	0.585
60	British	15	15	0.022	0.002	0.084
235	Zimbabwean	15	15	0.000	0.000	0.005
267	Zimbabwean	13	15	0.000	0.000	0.028

Men were ranked in descending order of crossover activity, mean crossover frequency as well as lower and upper 95% confidence intervals (CI) were determined by Poisson approximation. Cells are shaded to visualize variant STR-length alleles.

The 15-dinucleotide repeat, which was the candidate allele for recombination suppression, was found in d60 as well as d235, both of which show very low crossover frequencies. However this allele combination is equally found in d4 with a high crossover frequency of 1.59 % per sperm (Table 5-5). The longest alleles, 15 plus 17, were observed in d173 a man from Afro-Caribbean origin who showed a recombination frequency of 0.52 % that is close to the observed mean crossover frequency. There was no specific allele length associated with high or low recombination frequencies and it thus appears that the STR has no influence on the recombination frequency within this hotspot.

5.4.2 *PRDM9* status and crossover frequency

As Superhotspot T does not have a binding motif, the expectation was that it would not be regulated by *PRDM9*. However the two men that do not carry an A-allele of *PRDM9* displayed the lowest crossover frequencies as shown in Table 5-5.

Table 5-5 Sperm crossover frequencies and their relationship with *PRDM9* status

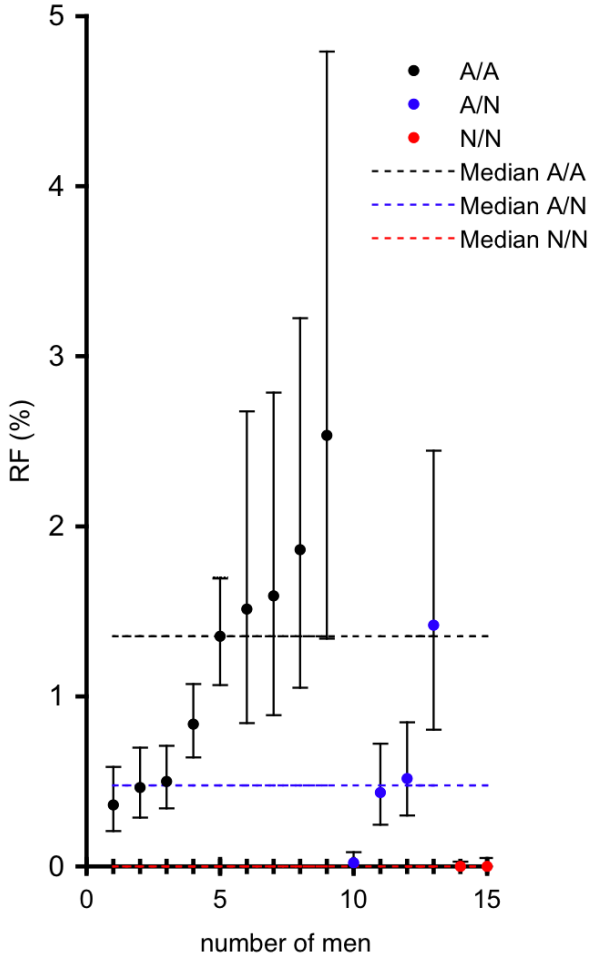
donor	population	PRDM9 variants		Crossover frequency per sperm		
				mean	lower CI	upper CI
268	Zimbabwean	A	A	2.535	1.340	4.792
281	Zimbabwean	A	A	1.863	1.051	3.224
4	British	A	A	1.591	0.889	2.786
278	Zimbabwean	A	A	1.514	0.842	2.676
43	British	A	L20	1.418	0.804	2.444
2	French	A	A	1.354	1.067	1.694
83	British	A	A	0.836	0.641	1.073
177	Afro-Caribbean	A	A	0.518	0.299	0.848
173	Afro-Caribbean	A	L7	0.500	0.342	0.710
25	British	A	A	0.464	0.288	0.699
269	Zimbabwean	A	L11	0.435	0.245	0.722
24	British	A	A	0.362	0.207	0.585
60	British	A	C	0.022	0.002	0.084
235	Zimbabwean	L4	L21	0.000	0.000	0.005
267	Zimbabwean	C	L14	0.000	0.000	0.028

Men were ranked in descending order of crossover activity with mean crossover frequency and 95% confidence intervals determined by Poisson approximation. Cells are shaded to visualize PRDM9 variants; orange cells indicate allelic variants other than allele A.

Crossover frequencies of men carrying *PRDM9* alleles other than allele A cluster at the bottom of the table (Table 5-5). The low ranking of d235 and d267, the two men without A-alleles among the 15 tested men is significant (two-tailed exact binomial, P=0.010). It therefore follows that crossover frequencies at Superhotspot T are influenced by *PRDM9* status with the tested N-variants not activating the hotspot and consistent with the A-variant being activating.

To test the association of *PRDM9* status with crossover frequencies further, allele classification was firstly simplified to A or N (non-A). With this classification, individuals could either be carrying A/A alleles, A/N alleles or N/N alleles. Individuals were then grouped according to their *PRDM9* status, and median crossover frequencies were determined for each group and compared between groups (Figure 5-6).

Figure 5-6 Variation in crossover activity between men at Superhotspot T modified from (Berg *et al.* 2010)



Individuals were grouped according to their *PRDM9* status into men carrying two A-alleles (A/A shown in black), one A-allele (A/N shown in blue) or no A-alleles (N/N shown in red). Confidence intervals have been estimated using Poisson-approximation. Hashed lines indicate the median crossover frequencies observed within each group.

As shown in Figure 5-6, median crossover frequencies are proportional to the number of A-alleles that individuals carry, with A/N men displaying 35% and N/N men displaying <2% of the median crossover frequency observed in A/A men. The difference between mean crossover frequencies of men homozygous A/A compared to the mean frequency of men without A-alleles is significant (two-tailed t-test, P=0.035). Likewise the mean frequency observed in men with at least one A-allele compared to men without A-alleles is significantly higher (one-tailed t-test, P=0.026). The higher mean crossover frequencies of men with A/A *PRDM9* status compared to men with A/N status is of marginal significance (one-tailed t-test, P=0.041). The mean frequency observed in A/N men is not significantly higher than the mean frequency observed for N/N men (one-tailed t-test, P=0.126).

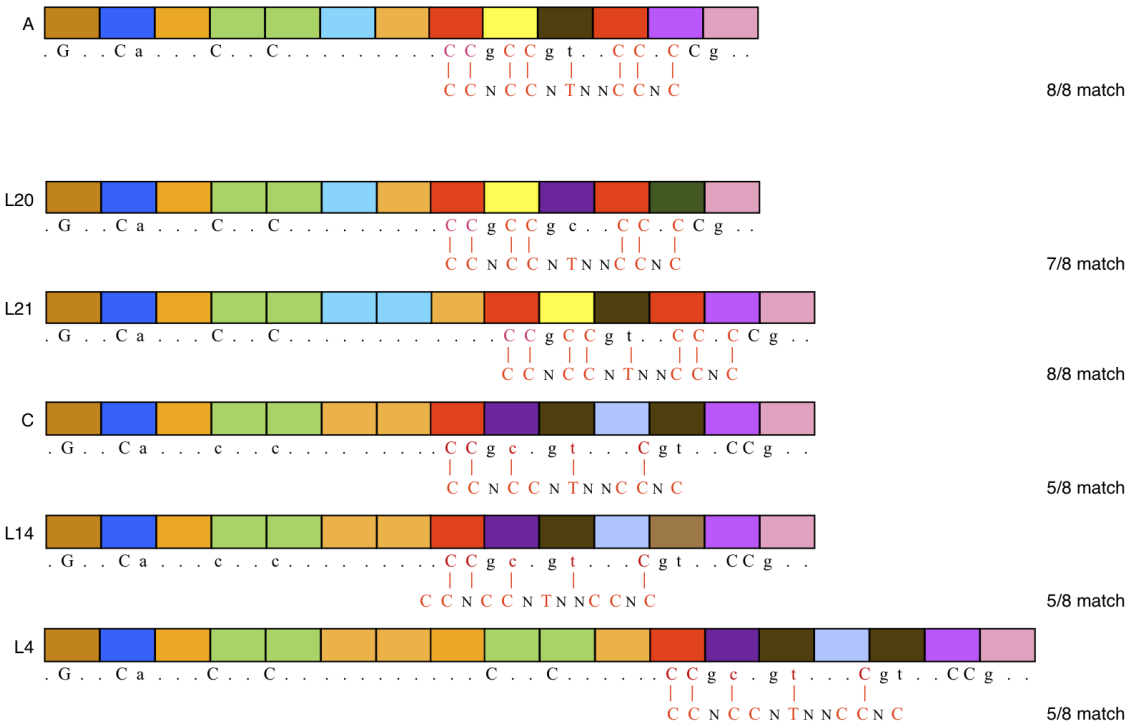
5.5 Summary and Discussion

The association between *PRDM9* genotype and suppressed recombination frequency was highly significant, and it can therefore be concluded that PRDM9 does regulate hotspot activity at Superhotspot T. Superhotspot T was inactive in men that do not have the A-allele of *PRDM9*, and the median crossover frequencies between men of different *PRDM9* status increased with the number of A-alleles that they carry. The observation of increasing median crossover frequencies between men of different *PRDM9* status is consistent with an additive model of hotspot activation proposed by (Berg *et al.* 2010). This implies that the A-variant of PRDM9 is activating Superhotspot T, despite the lack of an apparent hotspot motif proposed to be the *in-vitro* binding site for the protein encoded by the *PRDM9* A-allele. It has now been shown that *PRDM9* regulates hotspots with and without the hotspot motif to the same extent (Berg *et al.* 2010). In contrast, the STR lengths do not obviously influence crossover activity at Superhotspot T.

5.5.1 Activating and non-activating alleles

Allele A of *PRDM9* is associated with men that have high recombination frequencies, and therefore Superhotspot T is activated by the A-variant. Conversely, N/N men fail to activate Superhotspot T, and therefore all variants encoded by these N-alleles, namely C, L4, L14 and L21 must be non-activating. The ZnF structures, as well as binding predictions to the putative A-motif, are illustrated in Figure 5-7.

Figure 5-7 Activating and non-activating PRDM9 variants at Superhotspot T



PRDM9 variants A, L20, L21, C, L14 and L4, with ZnF structure indicated by coloured boxes and predicted DNA binding sequence shown below as in Figure 5-1 (modified from (Berg *et al.* 2010)). DNA binding sequences are aligned with the A-motif, and the binding sequence of variants A and L21 each match all eight specified bases of the A-motif, while binding predictions for variants L20, C, L14 and L4 matches at best only five of the eight bases.

As illustrated in Figure 5-7, non-activating alleles vary in the number and order of ZnFs, as well as in their binding predictions to the putative A-motif. Interestingly, allele L21 is very similar to allele A, with a duplicated ZnF in the middle of the array, but with the same binding prediction of all eight specific bases of the hotspot motif.

All variants identified as non-activating at Superhotspot T, were also found to be non-activating at additional hotspots (Berg *et al.* 2010). All except L21 were also testable at Hotspots that do contain the hotspot motif, and were found to be equally inactivating.

5.5.2 Significant variation of crossover frequencies between men with the same *PRDM9* status

Despite a significant association of *PRDM9* status with crossover frequencies, a large degree of variation can be observed between men with the same *PRDM9* status. Men who carry A/N alleles show variation as large as 65-fold between d60 and d43. In fact the

crossover frequency observed in d60 is comparable to the frequency observed for N/N men, despite carrying an A-allele. In contrast, man d43 showed a recombination frequency not significantly different to that of A/A homozygous men, and higher than that of A/N men, despite being heterozygous for A/L20. This suggested that the L20 variant might have activated Superhotspot T. The ability of some non-A PRDM9 variants to activate A-regulated hotspots has been observed by Berg *et al.* (2010) at additional hotspots, with the L20 variant activating hotspot MSTM1b (Berg *et al.* 2010).

Significant variation can also be observed in the A/A men, with frequencies varying 7-fold. Although differences in primer efficiencies between assays and man tested could have occurred, they do not explain the strong variation seen. The most extreme variation seen within the A/A group was observed between d24 and d268, but these men were tested using the same assay. PCR artefacts do not provide a comprehensive explanation for the strong variation seen. Theoretically, false-positives could have been generated from false annealing of partial extension products of homologous targets, referred to as haplotype switching (Dieffenbach and Dveksler 1993) and from progenitor "bleed through", which results from unspecific primer binding. For two men in this analysis, all pools remained negative for PCR product, and therefore neither showed jumping artefacts nor bleed-through, providing good negative controls. In addition, men tested on blood and sperm DNA by Berg *et al.* (2010) in additional crossover assays never showed any blood exchanges, as expected for products of meiosis and proving that PCR jumping artefacts are rare (Berg *et al.* 2010).

In conclusion, variation in frequencies cannot be explained solely by differences in primer efficiencies or the presence of false-positives generated by PCR artefacts, which suggests other regulatory elements influencing recombination at Superhotspot T.

5.5.3 *PRDM9* regulation provides an explanation for most of the variation observed between men analysed at the same hotspot

Superhotspot T is regulated by *PRDM9* and specifically activated by the A-variant of PRDM9. Berg *et al.* have shown that the A-variant of PRDM9 likely activates all hotspots discussed in previous chapters, and this activation has been tested and established at all hotspots examined in this thesis, except for Superhotspot H (Berg *et al.* 2010). Addition-

ally not only the major variation tested in this chapter but also almost all differences in recombination frequencies between donors analysed at any of the Superhotspots in Chapter 3 might be attributable to PRDM9 regulation.

The previously unexplained 2-fold reduction of crossover frequencies observed in man d17 at Superhotspot E might result from differences in amplifiability between assays, but it could also be the result of a non-activating L20 variant, as d17 was heterozygous for PRDM9 variants A and L20. The L20-variant had been observed to be non-activating at Superhotspots F and U (Berg *et al.* 2010). Conversely L20 has been shown to activate hotspot MSTM1b. PRDM9 variant L20 therefore has entirely different activation profiles between Superhotspots, which has also been observed for other PRDM9 variants (Berg *et al.* 2010).

All Superhotspots had been identified in individuals of North-European descent, but the presence and activity of recombination Superhotspot T was not population specific, though still dependent on activation by the European-enhanced A-allele. The presence of a smaller LD-step in the African population might therefore be in part explained by the fact that the frequency of allele A was just 50% within the African donor panel compared to 86% in the European donor panel.

5.5.4 The extremely low recombination frequency of d60 at Superhotspot T and d23 at Superhotspot F is not fully explained.

Intriguingly, the low crossover frequencies of the two men that have initiated the analysis of *PRDM9* effect on crossover frequencies at both hotspots cannot be explained by the findings in this Chapter. The presence of one activating A-variant should be necessary and sufficient for hotspot activation, as was observed at this and all of the other nine hotspots tested by (Berg *et al.* 2010). Man d60 is heterozygous for PRDM9 variants A and C and should therefore be active at Superhotspot T, at a reduced frequency in the range of the median frequency (0.46%), but certainly higher than the observed frequency of 0.02%.

Similarly, d23, the man who had shown extremely low frequencies at Superhotspot F, should have been active as he is heterozygous for alleles A/L1. Sperm crossover frequencies for this man were also determined by A. J Jeffreys using a full-crossover assay, but

again showed these low crossover frequencies. In addition, this man displayed equally low crossover activities at additional hotspots PAR2 and Q (Berg *et al.* 2010).

Low DNA quality was ruled out as cause for the low numbers of detectable recombinants, as the DNA sample had shown good amplifiability in phasing assays and sequencing PCRs, which pointed to the low crossover frequency being genuine in these men.

One explanation for inactivity might be that their *PRDM9* A-alleles are in fact null-alleles, mutants leading to non-functional proteins. Both of the men are donors who have provided their semen samples at fertility clinics. Interestingly, mouse knockouts for *PRDM9* displayed azoospermia in males, and were therefore infertile (Hayashi *et al.* 2005). If the association of *PRDM9* with infertility is true for humans, as has been suggested (Miyamoto *et al.* 2008; Irie *et al.* 2009), the donor panel may well be enriched for men with *PRDM9* null-alleles. However, the sperm DNA-yield of d23 and d60 is high, found at the 96th and 81st percentile in the entire donor panel, respectively. Even though sperm DNA yield may not fully account for the sperm count in these men, it therefore appears unlikely that they have a mutated *PRDM9* allele that resulted in low sperm counts.

Interestingly, even though *PRDM9* A/C heterozygous man d60 did not display the expected crossover frequency at this A-regulated hotspot, he did display an average crossover activity at a Hotspot activated by the C-variant of *PRDM9* (Berg *et al.* 2011). A hotspot activated by C variants of PRDM9 is discussed in the following Chapter.

Chapter Six A recombination hotspot activated by PRDM9 variant C variants displays biased gene conversion

Most of the data presented in this Chapter has been published (Berg *et al.* 2011).

6.1 Introduction

In Chapter Five, PRDM9 variant A was described as a regulator of a hotspot without the A-motif. Men with PRDM9 variants, different from the A-variant of PRDM9 showed lower crossover frequencies all tested hotspots, regardless of the presence of the A-motif (Berg *et al.* 2010). All of these hotspots are activated by the A-variant of PRDM9, and were initially identified from LD breakdown of crossover clustering in European populations. The A-allele is the predominant allele not only in the European population, but also the Indian population, both of which display a very low diversity of other PRDM9 variants (Berg *et al.* 2011).

Recently a second class of hotspots, that appear to have been more active in Africans than in Europeans, has been identified from population diversity data (Berg *et al.* 2011). To find these hotspots, Berg *et al.* screened 26 Mb of the human genome concentrating to within 6Mb of the telomeres to identify regions that showed a large LDU step in African populations, but not in the European and Indian populations. Coalescent analysis confirmed high historical recombination. Using this combined approach twenty LD-hotspots were short-listed as possibly having enhanced activity in Africans (Berg *et al.* 2011).

Recombinant breakpoint mapping of sperm crossovers at six of these target intervals confirmed sperm hotspots within all of them, and all were regulated by *PRDM9*. Even though these hotspots had shown consistently stronger effect on LD in Africans, the A-variant of PRDM9 nevertheless activated one of the six tested intervals.

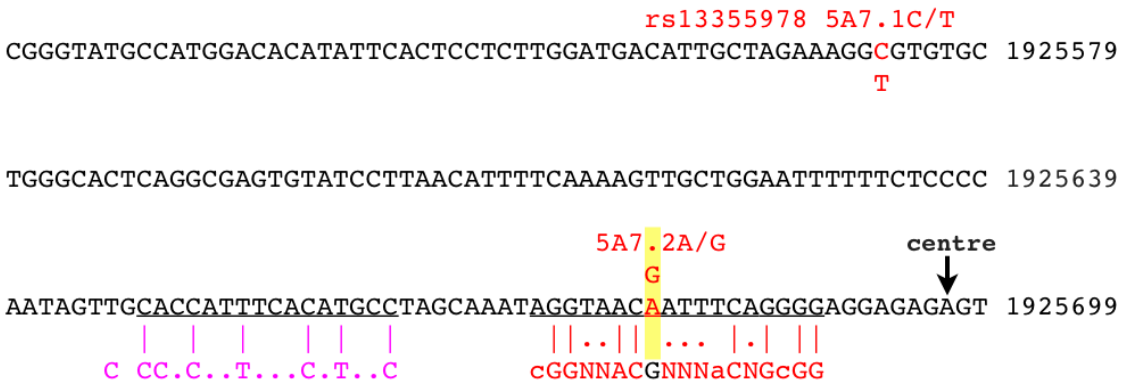
The remaining five hotspots were activated either by the C-variant of PRDM9 or by variants that had the same motif binding prediction as the C-variant, the latter referred to as C-type (Ct) variants (Berg *et al.* 2011). Not all of the six tested intervals contained a single hotspot; the five Ct regulated ones did, but the A-regulated interval contained more than one hotspot.

PRDM9 C and Ct-variants are predicted to bind (CCNCNNTNNNCNTNNC), also referred to as the C-motif (Kong *et al.* 2010). This motif was found in four of the five C-activated hotspots, with varying degrees of degeneracy (Berg *et al.* 2011). All of the Hotspots activated by C or Ct PRDM9 variants, with or without a C-motif located near the centre of the hotspot, showed high recombination frequencies of between 0.1% and 4% per sperm. The latter was the highest crossover frequency observed at a human Hotspot to date. The activity of the five Ct-regulated hotspots is extremely low in men lacking Ct variants. As Ct variants are common in Africans but rare in Europeans and Indians, this explained the difference in LD steps between those populations (Berg *et al.* 2011)**.**

One of the African-enhanced Ct-regulated hotspots, hotspot 5A, stood out in several ways. A single crossover hotspot with a consistent hotspot width of 1.5 kb was defined for all men for whom recombinant breakpoints had been mapped. But the central location of 7195 bp into the 15 kb target interval, centred at the LD hotspot, was the mean sperm hotspot centre for only three of the four men analysed. However the hotspot centre was shifted by about 320 bp upstream in the fourth man, who was heterozygous for the Ct variant L6 and variant A of PRDM9. The hotspot shift seen in this man most likely indicates that a new hotspot is activated near 5A by Ct variant L6.

For the three men sharing the common hotspot, crossover breakpoint analysis showed strong reciprocal crossover asymmetry, indicating biased gene conversion associated with crossover at the central SNP termed 5A7.2 G/A. There is a second marker 5A7.1C/T (rs13355978) located 75 bp upstream of 5A7.2 (Figure 6-1), and reciprocal crossover asymmetry was only observed in men that were heterozygous at 5A7.2. Five men homozygous at 5A7.2 did not show reciprocal crossover asymmetry at 5A7.1 and any other central SNPs. Heterozygosity at 5A7.2 therefore appeared to be required to trigger asymmetry. Here biased gene conversion associated with crossover results in over-transmission of the derived G-allele. This conversion of A→G results in strengthening of the C-motif as illustrated in Figure 6-1.

Figure 6-1 Matches to the predicted PRDM9 C-motif spanning SNP 5A7.2 (modified from (Berg *et al.* 2011))



The consensus sequence across the centre of Hotspot 5A (based on Genome Reference Consortium Human genome build 37, patch release 2 (GRCh37.p2)) is shown with the location of central SNPs, 5A7.1 and 5A7.2, and predicted PRDM9 C-motifs. One predicted C-motif (purple) and a better matched motif in inverted orientation (red) are present within the hotspot, vertical lines indicate matches between motifs and sequence. Transition of the A to the G-allele at 5A7.2 (highlighted in yellow) generates a better-matched inverted motif.

As shown in Figure 6-1, the active haplotype carrying the A-allele at 5A7.2 has only five out of seven base matches to the putative C-motif, while the suppressed haplotype has six of the non-degenerate sites matching this motif. Curiously, the sixth match is being provided exactly by the suppressing G-allele at 5A7.2 (Berg *et al.* 2011). This unexpected association of a better-matched motif with suppressed recombination activity was most intriguing.

As hotspot 5A had been initially analysed using a full-crossover assay by Berg *et al.* (2011), there was no information on the occurrence, distribution and frequency of non-exchange conversion.

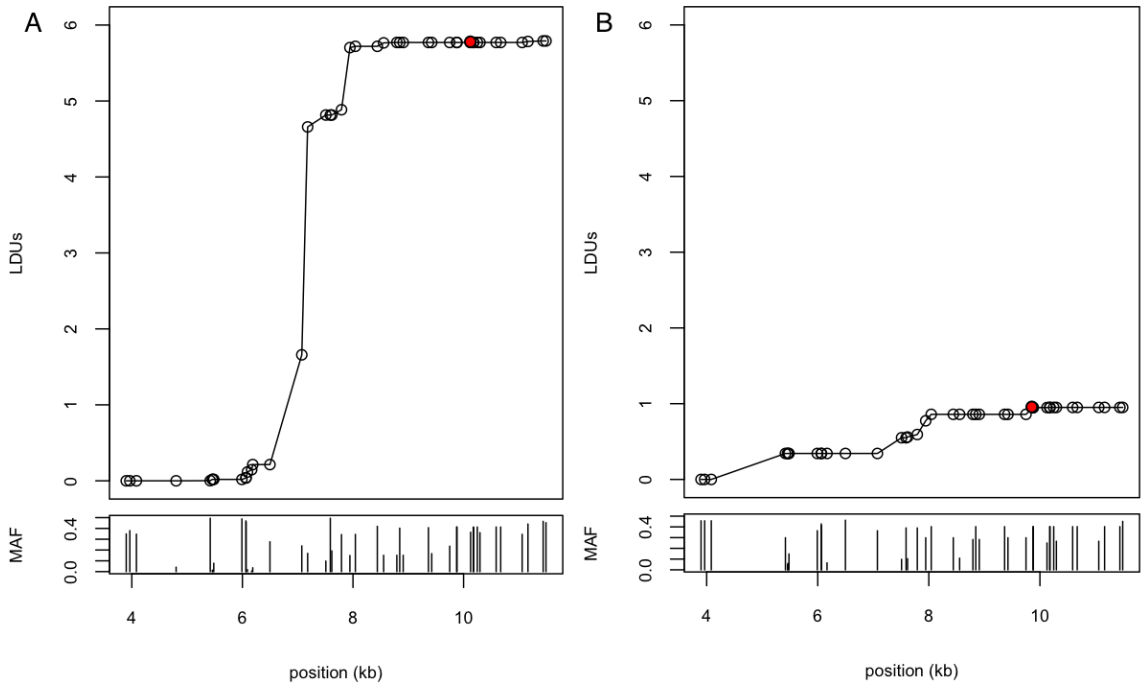
6.2 This work

Hotspot 5A was analysed using the half-crossover assay as described in Chapter 3 to test whether non-exchange conversions, if present, would show the same bias towards strengthening the C-motif. To test whether biased non-exchange conversion was present at hotspot 5A, transmission frequencies of alleles into both crossovers and non-exchange conversions was tested and compared as described in Chapter 4.

6.2.1 Target interval LDU profiles of the African and European populations samples

Hotspot 5A is located on chromosome 5p, between the genes encoding Iroquois-class homeodomain proteins IRX-4 and IRX-2. The target interval of hotspot 5A was initially identified as a region with strong LD breakdown in the four African HapMap populations, and weaker breakdown in the European and Indian HapMap populations (Berg *et al.* 2011). To test whether Hotspot 5A was equally detectable in LD from the semen donors available for analysis, metric LD maps were generated from un-phased diploid genotyping data from two population samples. The European population sample consisted of 103 North-European semen donors and the African population sample was compiled from 56 Zimbabwean and 18 Afro-Caribbean men.

Figure 6-2 LDU profiles of the African and European population samples across a region containing hotspot 5A



Metric profiles in LD-units generated from un-phased diploid genotyping data of (A) the African and (B) European donor populations, respectively. Circles represent polymorphic SNP markers, and lengths of corresponding ticks in bottom panel represent minor allele frequencies (MAFs) of these markers. The red dot indicated the location of reverse selector sites used for non-exchange conversion and crossover detection; these were located outside of the interval showing LD-breakdown.

A steep increment of six LD-units was observed in Africans, as shown in Figure 6-2. The markers located at the steepest point were markers 5A7.1 and 5A7.2, corresponding to the markers closest to the mapped mean centre of hotspot 5A as determined by (Berg *et al.* 2011). The location of the LD hotspot in the African semen donors thus corresponds to the exact location of crossover hotspot 5A. In contrast, little detectable historical recombination has occurred at the same location in the European population, displaying only a small increment of less than one LD-unit, which is a more than 5-fold reduced step-size compared to the African population. The information contained within the semen donor panels available for analysis therefore recapitulates the difference in historical recombination observed in the HapMap population samples analysed by (Berg *et al.* 2011).

6.2.2 Half-crossover assay design

To detect crossover as well as non-exchange conversion molecules reciprocal half-crossover assays were designed as described in Chapter 3. The target interval for a half-crossover assay was selected for hotspot 5A, as described for the A-regulated recombination hotspots in Chapter 3. The reverse selector sites used by Berg *et al.* (2011) were chosen as selector sites for the half-crossover assay, as they had shown high allele-specificity and PCR efficiency; the marker locations with respect to the hotspot are shown in Figure 6-2 (these two selector sites overlap in the graph, and a single red filled circle indicates the location of both markers). Both orientations of recombination were analysed (using PCR conditions summarised in Appendix VIII). Hotspot 5A qualified for the half-crossover assay not only in *PRDM9* Ct/Ct homozygous but also in *PRDM9* Ct/N men, as these still showed a crossover frequency >0.2% (Berg *et al.* 2011). Hotspot 5A has two markers located directly at the centre of the hotspot (Berg *et al.* 2011). Men in the donor panels were either only heterozygous at marker 5A7.1 or marker 5A7.2, however the small distances to the mean hotspot centre, of 93 bp or 18 bp respectively, provided the necessary marker density for non-exchange conversion detection using either marker.

Two men qualified for analysis via the half-crossover assay. Both men had varying haplotypes across the hotspot interval, with each having one of the heterozygous markers located in close proximity to the hotspot centre. Both men had been analysed previously using a full-crossover assay at hotspot 5A (Berg *et al.* 2011).

The men also had different PRDM9 variants. The first man (d253) was homozygous for *PRDM9* C-alleles and had a crossover frequency of 0.54% in the original full-crossover assay. The second man (d243) was heterozygous for Ct allele L15 and allele B of *PRDM9*. Despite only having one activating Ct variant this man had a crossover frequency of 0.39% (Berg *et al.* 2011), which should be sufficiently high to efficiently detect non-exchange conversions using the half-crossover assay.

6.3 Results

6.3.1 Frequent non-exchange conversions in addition to crossovers at hotspot 5A

For both men, crossover and non-exchange conversion molecules were detected in the half-crossover assay (illustrated in Appendix IX). Numbers of crossovers detected in reciprocal assays were comparable in d253 and d243 (t-test of Poisson means, P=0.458 and P=0.383, respectively).

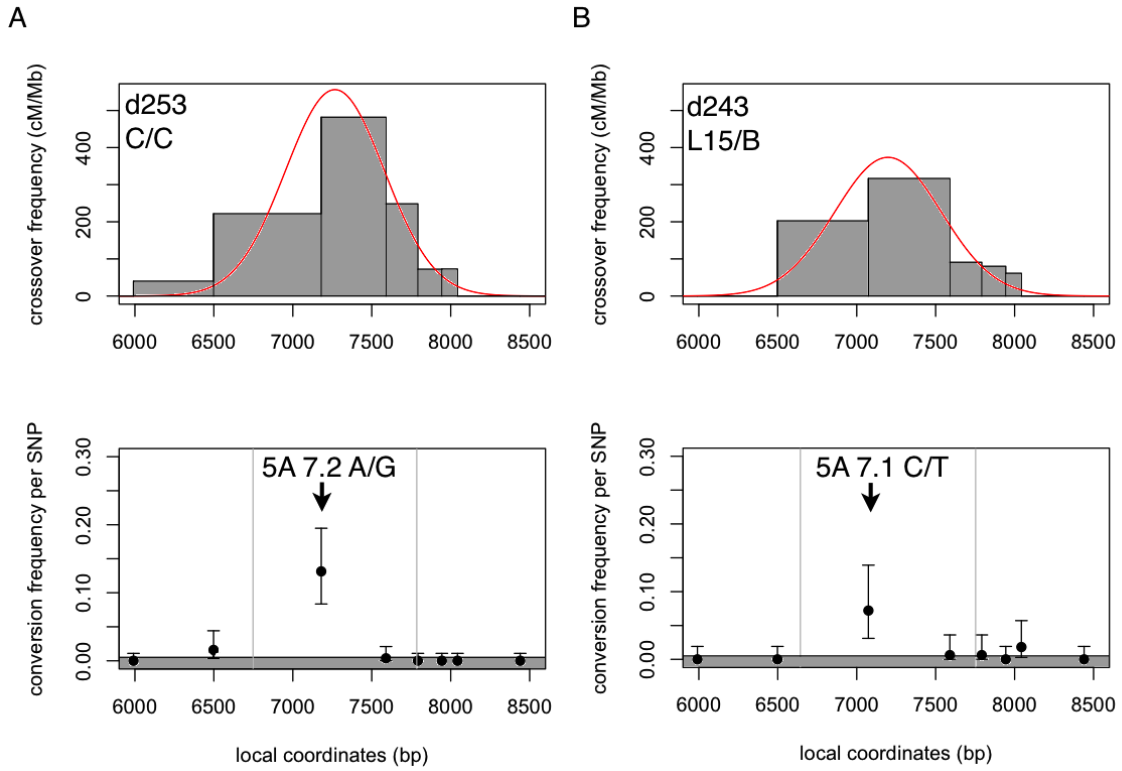
For the *PRDM9* C/C homozygous man d253, 28000 sperm DNA molecules were analysed, and crossovers were detected at a frequency of 0.44% per sperm. This crossover frequency was not significantly different to the crossover frequency of 0.54% detected by (Berg *et al.* 2011) using a full-crossover assay (Pearson's χ^2, P=0.108). In addition, man d253 showed an overall non-exchange conversion frequency of 0.15%, with a peak transmission frequency of 0.13% at marker 5A7.2 A/G, which has not been assigned an rs-number (as of 03/03/2012).

For the *PRDM9* A/L15 heterozygous man d243, 16560 sperm DNA molecules were analysed. Crossovers were detected in the half-crossover assay at a frequency of 0.31% per sperm, which was not significantly different to the frequency obtained previously, using a full-crossover assay (Pearson's χ^2, P=0.193) (Berg *et al.* 2011). Non-exchange conversions were also detected at an overall frequency of 0.10%, with a peak frequency of 0.07% at marker 5A7.1 C/T (rs13355978).

6.3.2 Morphology of recombination hotspot 5A

The distribution of crossover breakpoints and the distribution of gene conversion events, estimated as described in Chapter 3, are shown in Figure 6-3.

Figure 6-3 Distribution of recombinants at hotspot 5A for two men



(A) Man d253 is homozygous for PRDM9 variant C, and (B) man d243 is heterozygous for PRDM9 variants L15 and B. Top panels display the reciprocal crossover distribution and frequency in cM/Mb as histograms, as well as least-squares best-fit normal distributions (red line), both averaged over reciprocal half-crossover assays. Bottom panels display non-exchange conversion frequencies per SNP averaged over reciprocal assays with 95% confidence intervals based on Poisson-approximation, the grey zone indicated the expected rate of PCR misincorporation (as described in 3.3.5).

A typical crossover breakpoint distribution was observed, consistent with a normal distribution of crossover resolution points (Figure 6-3). This distribution is consistent with the data gathered in full-crossover assays, testing the same men, and therefore half-crossover assays used here give reliable results when compared to full-crossover assays as observed previously (3.4.1). Non-exchange conversions clustered at the most central marker as illustrated in Figure 6-3, comparable to the distribution of non-exchange conversion molecules at A-regulated hotspots analysed in Chapter 3.

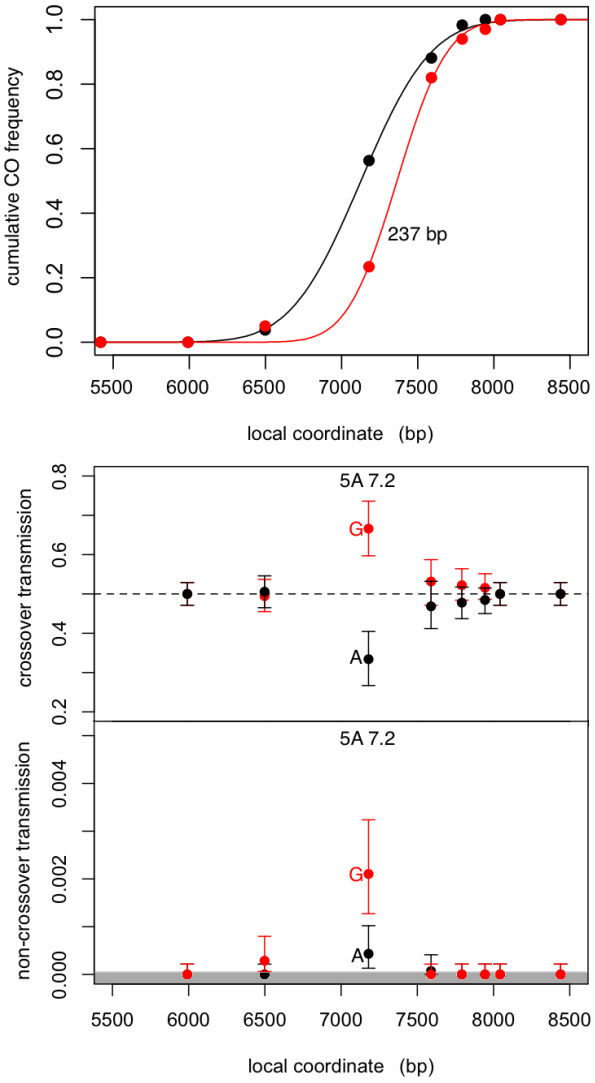
One transitional SNP 5A8.0 located outside of the crossover breakpoint interval, displayed a marginally significant elevation above the expected rate of PCR misincorporation (Poisson-approximated binomial probability, P=0.041).

6.3.3 Biased gene conversion in non-crossover as well as crossover molecules

Two men were tested that not only had different *PRDM9* genotypes but also different haplotypes in respect to the central SNP. The man homozygous for PRDM9 variant C, man d253, is heterozygous for SNP 5A7.2, located directly within a C-hotspot motif. The other man, d243 is homozygous at this SNP and instead heterozygous at adjacent SNP 5A7.1. Previously this SNP did not appear to be affected by TD in crossover across five men tested (Berg *et al.* 2011). To detect incidence of for biased gene conversion, in crossover and non-crossover molecules, they were tested for disparity in reciprocal recombination rates as described in Chapter 4 (4.3).

Cumulative crossover frequency plots that allowed the visualisation of reciprocal crossover asymmetry, as well as graphs displaying transmission frequencies in both crossover and non-exchange conversion are displayed in Figure 6-4 for man for d253, and in Figure 6-5 for man d243, respectively.

Figure 6-4 biased gene conversion and reciprocal crossover asymmetry in sperm recombinants from man d253



Top panel shows cumulative crossover frequencies of orientation A (red) and orientation B (black). Least-squares best-fit cumulative crossover distributions for both orientations are shown by black and red line, respectively. Middle panel shows transmission frequencies per SNP allele into crossover molecules, based on normalised proportions of crossovers. Lower panel shows transmission frequencies of alleles found in non-exchange conversions.

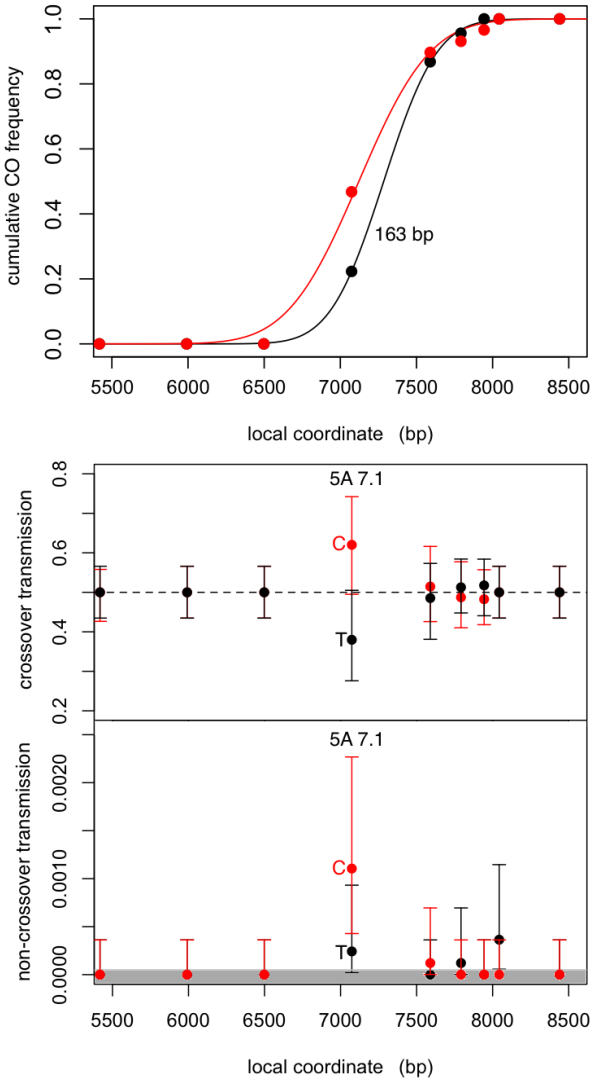
Man d253 showed reciprocal crossover asymmetry focused at marker 5A7.2 (Figure 6-4), with 67% of crossovers receiving the G-allele. As mentioned previously, it is the G-allele at 5A7.2 that generated a better-matched PRDM9 C-motif (Figure 6-1). Berg *et al.* had observed marker transmission frequencies of (69:31) in favour of the G-allele for d253, which were not significantly different from those observed here (two-tailed Fisher-exact, P=0.102). In the dataset generated by (Berg *et al.* 2011), reciprocal crossover asymmetry was not restricted to marker 5A7.2, and a second marker 5A7.6a (+/−) was also involved, with the "−" allele over-transmitted into crossover. In the data generated in the half-

crossover assay the "–" allele was transmitted into 53% of crossover molecules, but this distortion was not significant. In addition, a single non-exchange conversion molecule involved this marker and conversion of the + allele into – allele was observed.

As shown in Figure 6-4, non-exchange conversions also showed a peak of activity focused at the 5A7.2 A→G substitution at the centre of the hotspot and, as with crossovers, were strongly biased towards generating more G-alleles. For man d253, significant biased gene conversion in non-crossover was only observed at marker 5A7.2, with 83% of conversion events having acquired the G allele. Even though transmission frequencies are higher in non-crossovers than in crossovers, this difference is not statistically significant (two-tailed Fisher-exact, P=0.091).

Previous observations showed that men homozygous at 5A7.2 but heterozygous at marker 5A7.1 show no indication for reciprocal crossover asymmetry. Man d243, was the only man analysed using a full-crossover assay by A. J. Jeffreys, for which weak TD at marker 5A7.1 was found. For this man, 58% of crossovers had obtained the C-allele, a proportion that is significantly different from 50:50 (two-tailed Fisher-exact, P=0.036). However, as transmission ratios were not significantly different between this man and three other men, transmission was balanced when averaged across all men. Therefore it was concluded that the data did not show significant TD at marker 5A7.1 overall (Berg *et al.* 2011). This man was now also analysed with the half-crossover assay. As shown in Figure 6-5, the data generated in reciprocal half-crossover assays for d243, also indicated reciprocal crossover asymmetry and transmission distortion in crossovers, and additionally also biased transmission in non-exchange conversions.

Figure 6-5 biased gene conversion and reciprocal crossover asymmetry in sperm recombinants from man d243



Top panel shows cumulative crossover frequencies of orientation A (red) and orientation B (black). Least-squares best-fit cumulative crossover distributions for each orientation are also shown in black and red. Middle panel shows normalised proportions of crossovers acquiring either allele. Lower panel shows transmission frequencies of alleles found in non-exchange conversion obtained in reciprocal assays.

The data generated in the half-crossover assay, 62.0% of crossovers had acquired the C-allele at marker 5A7.1 (Figure 6-5), a transmission ratio that is not significantly different from 50:50 (two-tailed Fisher-exact, P=0.137). The full-crossover data and the half-crossover data have indistinguishable allele-transmission ratios (two-tailed Fisher-exact, P= 0.872), and the combined dataset clearly show a significant deviation from 50:50 (two-tailed exact binomial, p=0.011) with an overall transmission ratio of 59:41 in favour of the C-allele. It also follows that transmission ratios of C and T alleles vary significantly

between d243 and the three additional men analysed for TD by A.J. Jeffreys in a full-crossover assay (χ^2 on 2x4 contingency table, P=0.015, after Bonferroni-correction for the 4 men). Non-exchange conversions also showed peak transmission at marker 5A7.1, with 82% of detected non-exchange conversion molecules resulting from conversion of the T-allele into the C-allele, but this data is not significant on its own (two-tailed exact binomial, P=0.109). The bias into crossover molecules was not significantly different from the bias into non-exchange conversion molecules (two-tailed Fisher-exact, P=0.220), and the overall transmission ratio in both types of recombinants is 60:40 in favour of the G allele, a highly significant deviation from 50:50 transmission (two-tailed exact binomial, P=0.0038).

6.4 Summary and Discussion

The reciprocal crossover asymmetry, which had already been observed in crossovers in men that were heterozygous at SNP 5A7.2 (Berg *et al.* 2011) was also present into non-exchange conversions of d253. Originally men heterozygous at SNP 5A7.2 showed strong reciprocal crossover asymmetry, in contrast four men homozygous A/A at 5A7.2 but heterozygous at nearby SNP 5A7.1 had displayed averaged transmission ratios close to 50:50 in the initial crossover analysis. Weak TD of 58:42 had only been observed in one of the men, d243 (P=0.036). Pooling transmission frequencies across all four men heterozygous at 5A7.1 had relinquished significance, pointing to SNP 5A7.1 not showing reciprocal crossover asymmetry overall (Berg *et al.* 2011).

Surprisingly, reciprocal crossover asymmetry with corresponding bias in non-exchange conversion was observed at both men in this survey. Each man shows biased gene conversion associated with crossover and biased gene conversion in non-crossover, not only at a different marker in each man but also to varying degrees. Biased gene conversion was found associated with crossover as well as non-crossover recombinants for both men.

6.4.1 Non-crossover transmission increases the A→G substitution frequency in man d253

The first man, d253 showed strong TD into crossover and non-exchange conversion at marker 5A7.2, the marker located within the inverted PRDM9 C-motif. The strong biased gene conversion, not only in crossovers but also in non-exchange conversion leads to

gametic ratios of 50.085:49.915 in favour of the G-allele. Gametic ratios of 50.009:49.991, observed at Hotspot *NID*1 by Jeffreys and Neumann (2005) were sufficiently strong to lead to the eventual fixation of the over-transmitted allele (Jeffreys and Neumann 2005). The gametic ratio observed at 5A7.2 should therefore virtually guarantee fixation of the G-allele, and thus equally the eventual fixation of the better matched inverted C-motif located within hotspot 5A (Berg *et al.* 2011). However, generation of a better-matched motif is associated with hotspot inactivation.

As discussed previously, a cut homologue is repaired from the uncut homologue, resulting in the over-transmission of alleles from the haplotype that had not been cut i.e. the suppressed haplotype. Consequently, the motif generating allele should nevertheless be the recombination-suppressing allele. This directionality is opposed to previous observations at hotspots *DNA*2 (Jeffreys and Neumann 2002) and *NID*1 (Jeffreys and Neumann 2005), were over-transmitted alleles disrupted the hotspot-motif (Myers *et al.* 2008).

6.4.2 Biased gene conversion is not restricted to marker 5A7.2

A weak but significant TD was observed at marker 5A7.1 in man d243 in crossovers and non-exchange conversions. Reciprocal crossovers mapped to intervals shifted by less than 200 bp. Despite the weak degree of TD, biased transmission is nonetheless observed in crossover and non-exchange conversions and both overtransmitted the ancestral C-allele at 5A7.1. Since the hotspot is active in 0.3% of meioses in man d243, 50.027% of sperm will have acquired the C-allele compared to only about 49.973% of sperm acquiring the T-allele. However, as TD was not observed in all men at marker 5A7.1, it remains to be understood whether strong meiotic drive observed in one man, will actually affect variant fixation on the population level.

This milder form of reciprocal crossover asymmetry is comparable to that observed in Chapter 4 at Superhotspots H and F, again suggesting weak TD and weak reciprocal crossover asymmetry being possible and not affecting all men with the same heterozygosity. The inconsistencies of allele transmission between men may suggest biased conversion not being controlled by 5A7.1 itself. So how can men with identical haplotypes not all show reciprocal crossover asymmetry? Biases initiation may be the result of interactions between *cis*-and *trans*-acting factors at a given hotspot.

Man d243, the only man that had shown reciprocal crossover asymmetry at marker 5A7.1, has PRDM9 variants L16/B, of which only L16 is a Ct-PRDM9 variant. The other men analysed at hotspot 5A had activating PRDM9 variants C or L6 (Berg *et al.* 2011). As subtle changes within PRDM9 ZnF had shown different binding properties across several hotspots, it can be speculated that PRDM9 variants, despite activating the same hotspot, may be influenced by different SNP markers in *cis*.

Chapter Seven Summary and final Discussion

The human genome was estimated to contain about 33,000 hotspots, defined by historical LD block boundaries (Myers 2005), about one in every 50 to 100 kb. Only about 40 hotspots have been directly characterised in individuals by sperm typing. Previously hotspot analysis had focused on the description of crossover profiles using repulsion phase allele-specific PCR, a method in which non-exchange conversion products are inaccessible. At the beginning of this work, non-exchange conversions had only been detected at four human recombination hotspots *DNA*3, *DMB*2, *SHOX* and *NID*1 by using DNA enrichment by allele-specific hybridisation (DEASH). Hotspots *DNA*3 and *DMB*2 were located in the Major Histocompatibility Complex, *SHOX* in the major pseudoautosomal region and *NID*1 adjacent to a minisatellite. At these hotspots, non-exchange conversions clustered at the centre of the crossover distribution. However, overall very little was known about the processes that regulate recombination hotspots and especially whether crossovers and non-exchange conversions were generated by the same processes.

7.1 Fine-scale analysis of non-exchange conversions in addition to crossover at a range of human hotspots

To elucidate the relationship between crossover and non-crossover, as well as the frequency and distribution of non-exchange conversion events, half-crossover assays were performed at six recombination hotspots to detect both types of recombinants simultaneously from pools of sperm DNA. As a half-crossover assay used by Holloway *et al.* to analyse the ß-globin hotspot had not returned any non-exchange conversion molecules, it was inferred that their recombination frequency was too low for efficient detection via this method (Holloway *et al.* 2006). Efficient detection would therefore require hotspots with much higher recombination frequencies. The detection of non-exchange conversion events with the half-crossover assay relied on the successful distinction of base changes using ASO hybridisation to amplified pools of sperm DNA, which contained mostly progenitor molecules. As smaller pool sizes allowed for a better signal-to-noise ratio, it follows that hotspots with higher recombination frequencies would facilitate cleaner half-crossover assays. Therefore Superhotspots with crossover frequencies of at least 0.1% per sperm were selected for analysis. Secondly, the detection of conversion events crucially relies on the presence of marker(s) very close to the centre of the hotspot and hotspots with a suitable marker density were chosen for analysis.

Half-crossover assays performed at six Superhotspots, with a marker located in close proximity of the centre, allowed the detection of high numbers of non-exchange conversion events in addition to crossovers. With this work the number of hotspots analysed for both types of recombinants has been increased from four to ten hotspots, and these data have also allowed us to address the relationship between crossovers and non-exchange conversions. Five of the six hotspots in this survey were subterminal, in fact most of the "Superhotspots" found to date were located within 6Mb of the telomeres (Webb *et al.* 2008; Jeffreys and Neumann 2009; Berg *et al.* 2010; Berg *et al.* 2011).

7.1.1 Non-exchange conversion frequency and distribution at recombination hotspots

Non-exchange conversion molecules were detected at every hotspot, and occurred at frequencies of 0.01% to 0.47% per haploid genome equivalent (HGE). Half-crossover assays at all hotspots gave crossover data consistent with data gathered at the same hotspots using repulsion-phase crossover assays, followed by crossover breakpoint mapping. Reciprocal crossover products arose with the same frequency and plotting crossover exchange points across the analysed region displayed a symmetrical quasi-normal distribution as observed previously for these hotspots (Webb *et al.* 2008; Berg *et al.* 2011) and all crossover hotspots analysed to date (as reviewed in (Jeffreys *et al.* 2004; Kauppi *et al.* 2004; Arnheim *et al.* 2007). Conversion tracts clustered at the centre of crossover activity, and non-exchange conversion profiles displayed steep bidirectional gradients at all of the hotspots.

The 1-2 kb width of the crossover distribution as well as the distribution of non-exchange conversions clustering at the centre of this crossover distribution have also been observed at mouse hotspots (Baudat and de Massy 2007; Cole *et al.* 2010), suggesting highly similar processes operating at mouse and human hotspots.

Conversion tracts were found to be short, with mean minimum and maximum co-conversion tracts being 60 and 833 bp long, respectively. Considering that tract lengths estimations are highly dependent on the marker density within a hotspot interval, these estimates are consistent with previous observations at human hotspots of a probable range between 55-290 bp (Jeffreys and May 2004). Overall most tracts were found to be shorter

than 1 kb, in contrast to meiotic gene conversion tracts in yeast, which are on average 1.6 kb long (Borts and Haber 1989; Mancera *et al.* 2008). Thus far, long conversion tracts have not been observed in these or other human recombination hotspots. The co-localisation of crossover resolution sites and short non-exchange conversion tracts observed at all hotspots substantiates that this zone probably marks a short zone of recombination initiation, originally proposed to be about 400bp wide in human hotspots (Jeffreys and May 2004).

7.1.2 The relationship between crossovers and non-exchange conversions

The detection of a large number of non-exchange conversion events at several hotspots allowed the investigation of the relationship between crossovers and non-exchange conversions, revealing that crossover and non-exchange conversion frequencies were positively and fairly tightly correlated.

If crossover and non-exchange conversion arise by the same initiation lesion, and if the number of DSBs would be a constant at each hotspot, then hotspots very active in crossover would show very little gene conversion, and *vice versa*. Instead hotspots with high crossover frequencies also tend to show high non-exchange conversion frequencies.
Interestingly, this correlation was not only seen between men at a given hotspot, but also when compared between hotspots, including all hotspots that had been analysed previously for non-exchange conversions (Jeffreys and May 2004; Jeffreys and Neumann 2005). This observation argues against recombination being initiated at a constant frequency at a given hotspot, but instead suggests varying initiation frequencies between hotspots with a fairly stable ratio of CON:CO with an average CON:CO ratio of 1:2. Directly observed CON:CO ratios varied between hotspots, from 1.7:1 observed at Superhotspot K to 0.3:1 at Superhotspot H (Chapter 3, 3.5).

Several attempts have been made to predict CON:CO ratios in the human genome, as ignoring the influence of gene conversion on LD can have detrimental effects in association studies. Ptak *et al.* had estimated a CON:CO ratio of 1:4 in Europeans and 1:1 in Africans (Ptak *et al.* 2004) to explain the inflation of historical recombination in LD that could not be explained by crossover based recombination rates alone. This estimate is consistent with the observed level of gene conversion at Hotspots analysed here (3.6.1.4).

However other estimates by Ardlie *et al.* invoked non-exchange conversion at a ratio of 3:1 -10:1 in favour of non-exchange conversion, with 6:1 being the best estimate (Ardlie *et al.* 2001). These estimated stemmed from the observation that a significant proportion of genotyped locus pairs, separated by short distances (~124bp), showed incomplete LD that was not readily explained by the expected historical recombination rate based on crossover alone.

In this work CON:CO ratios were fairly constant compared between men at a given hotspot. Ten men analysed at Superhotspot F did not display significant variation in their CON:CO ratios and neither did five men analysed at Superhotspot K (4.4.4.1), indicating stable CON:CO ratios between men at a given hotspot (Chapter 4). This data would suggest a fixed rate of crossovers and non-exchange conversions at a given recombination hotspot. In contrast Sarbajna *et al.* (2012) observed significant and very strong variation in CON:CO ratios between 14 men at hotspot *SPRY3* (Sarbajna *et al.* 2012). Hotspot *SPRY3* is located in the minor human pseudo-autosomal region PAR2, and may be behaving differently to the hotspots tested here, this may be because either PAR1 or PAR2 must engage in a crossover event at any given meiosis to prevent non-disjunction. Therefore factors influencing the crossover/non-crossover decision may be functioning differently between pseudo-autosomal and autosomal hotspots. Alternatively, it is equally possible that differences between men at a given hotspot have just not yet been detected in autosomal hotspots.

7.2 *PRDM9* regulation of human recombination hotspots

At the beginning of this work, our knowledge about regulating factors of the human recombination machinery was limited. Findings of initiation biases that resulted in reciprocal crossover asymmetry pointed to *cis*-acting factors regulating hotspot activity.

Crossover frequencies obtained during half-crossover analysis were found to be much more variable than initially observed at Superhotspots by (Webb *et al.* 2008) and not all could be explained by *cis*-acting elements influencing initiation efficiency. High variation in crossover frequencies between men had also been observed independent of local sequence variation at two other closely linked hotspots (Neumann and Jeffreys 2006). These data suggested the involvement of a *trans*-regulating factor, which has recently been identified as PRDM9 (Buard *et al.* 2009; Berg *et al.* 2010).

The identification of *PRDM9* as the major *trans*-regulating factor responsible for specifying and regulating hotspot activity, in part through work presented in this thesis, has increased our understanding as to what controls recombination hotspots in humans (Chapter 4) (Chapter 6) (Buard *et al.* 2009; Berg *et al.* 2010; Berg *et al.* 2011; Hinch *et al.* 2011).

Binding predictions of PRDM9 variant A, encoded by the most common allele in the European population, were consistent with binding to a motif enriched in human LD hotspots identified from European HapMap data (Myers *et al.* 2010). However this work provides evidence for PRDM9 affecting crossover activities in sperm independently of a hotspot motif that had been identified to be the PRDM9 binding site *in vitro* (Baudat *et al.* 2010). Superhotspot T, the hotspot showing the most extreme variation in crossover frequencies between men in this survey, was activated by the PRDM9 A-variant, though it did not contain an obvious A-motif (Chapter 5) (Berg *et al.* 2010). Additional data addressing PRDM9 regulation and the relationship with the hotspot motif came from work presented in Chapter 6. This Chapter investigated a recombination hotspot tuned to PRDM9 Ct variants, which are more common in Africans. At this hotspot, termed 5A, strong reciprocal crossover asymmetry had been observed, indicative of biased gene conversion associated with crossover (Berg *et al.* 2011). As discussed previously, preferred initiation on one of the haplotypes lead to overtransmission of markers from the other haplotype, allowing the less active haplotype to be identified. Curiously the overtransmitted, and therefore putative recombination suppressing allele characterised a stronger predicted binding motif for the PRDM9 C-variant. Analysis of non-exchange conversion at this hotspot showed that biased gene conversion exists not only in crossovers but also in non-exchange conversions leading to even stronger over-transmission of the motif-strengthening allele (Chapter 6) (Berg *et al.* 2011). The meiotic drive is sufficiently strong to lead to eventual fixation of the better-matched C-motif, but simultaneously to attenuation of the hotspot.

The data presented in this thesis is not the only evidence for hotspot regulation by PRMD9, but independent of proposed PRDM9 binding motifs. An association of PRDM9 status and recombination frequencies was observed at all hotspots analysed for the effect of PRDM9, and this irrespective of the presence of a hotspot motif (Berg *et al.* 2010; Berg *et al.* 2011). Hotspots were tuned to specific PRDM9 variants with the A-allele activating

a subset of hotspots. Conversely the same variant was not activating at recombination hotspots tuned to the C-variant of PRDM9 (Berg *et al.* 2011). Hotspots were activated by a distinct set of Ct variants, with some hotspots tuned to a narrow spectrum of variants, while others are more widely tuned (Berg *et al.* 2011).

Overall non-exchange conversion frequencies and distributions at hotspots activated by A- or C-variants of PRDM9 showed highly similar distribution, suggesting that the same processes operated at hotspots activated by different PRDM9 variants. Individuals with differing *PRDM9* genotype, even individuals within the same population, can use different sets of hotspots. Individual differences in PRDM9 genotype can explain at least 80% of the heritable variation between genome wide hotspot usage (Grey *et al.* 2009; Fledel-Alon *et al.* 2011; Hinch *et al.* 2011). However, as the A-allele is the most common allele in the European population, most hotspots identified from European HapMap data were activated by the A-allele. Conversely, the C-variant is the second most common allele in the African population and regions with intense LD breakdown in the African, but not the European HapMap populations led to the identification of hotspots tuned to C-variants of PRDM9. Overall, variation between predominant alleles of a given population can affect the recombination landscape between populations (Berg *et al.* 2011; Hinch *et al.* 2011).

7.2.1.1 Does the protein encoded by *PRDM9* bind a specific sequence motif?

Data presented in this thesis argues against a simple relationship between PRDM9 variation and hotspot motifs, as postulated by (Myers *et al.* 2010). Sperm crossover activity was highly dependent on specific PRDM9 variants, but independent of the presence of the hotspot motif as observed for Superhotspot T (Chapter 5) as well as at four additional recombination Superhotspots analysed by Berg *et al.* (2010). Secondly, a motif-strengthening SNP at hotspot 5A was associated with the suppressed haplotype (Chapter 6). It appears that the presence of a motif is not sufficient for recombination either, as a motif is often located in recombination cold sequence outside of a hotspot (Berg *et al.* 2010). Though a motif-disrupting SNP was found to be associated with the suppressed haplotype at hotspot *NID*1 (Jeffreys and Neumann 2005; Myers *et al.* 2008), opposite to the observation at Hotspot 5A, where a better-matched motif associated with the suppressed haplotype (Chapter 6) (Berg *et al.* 2011).

One might speculate that binding of the protein encoded by *PRDM9* would allow a certain degree of degeneracy that could explain regulation of hotspots without the motif. However, subtle changes in ZnF arrays between PRDM9 variants can have a strong effect on the recombination activity at a given hotspot. For example the non-activating PRDM9 variant L21 at Superhotspot T has almost identical ZnFs to the activating variant A, with the only difference being a duplication of the fifth ZnF. Binding predictions based on the last six ZnF are identical (Chapter Five, Figure 5-7). Additionally, men with variants that have identical binding predictions to the hotspot motif can display variation in crossover frequencies as strong as activation/non-activation (Berg *et al.* 2010). And hence, PRDM9 binding to a specific recognition site would be the most ready explanation for how subtle changes between alleles created non-activating variants incapable of recombination (Chapter 5) (Berg *et al.* 2010).

The observation of hotspots activated by PRDM9 without the presence of the motif, while at the same time being affected by subtle changes in the ZnF array, creates a conundrum as to how PRDM9 can be highly specific but at the same time highly permissive to allow regulation without a specific binding motif.

One solution to this paradox may be that PRDM9 does bind to a specific sequence, however not the sequence predicted with the current methods. Binding predictions of PRDM9 were based on theoretical algorithms established on short ZnF arrays (Persikov *et al.* 2009), and it is therefore unclear whether these predictions can be extrapolated to a protein with 13 or more ZnFs. And as the methods only evaluate whether a particular ZnF protein can theoretically bind a fragment of DNA (Persikov *et al.* 2009), they hence may not accurately reflect binding *in vivo*. Additionally, more than the last five ZnFs may be involved in DNA binding (Berg *et al.* 2011). These five ZnF had predicted affinity to the 13-mer hotspot motif, and were therefore initially implicated to be the key binding residues. However, as hotspots are not dependent on the presence of this sequence motif, this initial association does not hold true. To explore this question further requires understanding *PRDM9* binding *in vivo*.

7.2.2 Does *PRDM9* regulate crossovers and non-crossovers to the same degree?

Hotspots with major variation in crossover frequencies between men showed that recombination suppression consistently affected both crossover and non-crossover events (Chapter 3). Additionally, CON:CO ratios were generally stable between men at the same hotspot, even across men with different *PRMD9* alleles. Although *PRDM9* regulation of non-crossovers could not be established with work in this thesis alone, it has now been shown that *PRDM9* affects crossovers and non-crossovers to the same degree (Berg *et al.* 2011; Sarbajna *et al.* 2012). PRDM9 is believed to influence recombination prior to DSB formation, which is consistent with models in which the crossover/non-crossover decision occurs after recombination initiation, at the level of intermediate formation (Börner *et al.* 2004).

7.3 Are crossover and non-exchange conversion influenced by the same biases?

Major variation in crossover frequencies had been observed previously (Jeffreys and Neumann 2005; Neumann and Jeffreys 2006), and some of this variation was associated with specific hotspot sequence variants influencing the efficiency of initiation between chromosomes (Jeffreys and Neumann 2002; Jeffreys and Neumann 2005; Jeffreys and Neumann 2009). This phenomenon manifests as reciprocal crossover asymmetry in heterozygotes, as higher initiation on one haplotype would lead to biased gene conversion tracts accompanying crossovers, with markers acquired from the less active homologue (Jeffreys and Neumann 2002). One hotspot that was characterised for crossovers and non-exchange conversions had indicated comparable biased transmission in both crossover and non-crossover products (Jeffreys and Neumann 2005). The question therefore remained as to whether crossover and non-crossover were in general influenced by the same biases, and whether additional biases could be identified.

7.3.1 Initiation biases affecting crossovers and non-crossovers

Reciprocal transmission rates of markers in crossovers and non-exchange conversions were compared, detecting significant biases at four of the six hotspots analysed in this thesis (Chapters 4 and 6), suggesting that biases are actually fairly common.

Three of these four hotspots showed incidences of the previously described reciprocal crossover asymmetry. The strongest reciprocal crossover asymmetry and transmission distortion were observed for d253 at Superhotspot 5A. Here reciprocal crossovers mapped to intervals displaced by more than 200 bp. Transmission ratios at the most central SNP 5A7.2 A/G were in excess of 68% for the G-allele, with mild TD also observed at an additional marker. Comparing haplotypes across men associated the 5A7.2 A-allele with the active haplotype in all men heterozygous at this SNP. The TD observed at marker 5A7.2 within Hotspot 5A was weaker as those observed at hotspots *DNA*2 and *NID*1 (Jeffreys and Neumann 2002; Jeffreys and Neumann 2005). However, these three hotspots had clearly identifiable active and inactive haplotypes, and also displayed weaker transmission distortion at other markers located further from the centre.

At Hotspot 5A, an indication of weaker TD was also observed in one man homozygous at 5A7.2 but heterozygous at a nearby SNP, 5A7.1. This observation was curious, as other men with the same haplotype in regard to these two SNPs show no indication for reciprocal crossover asymmetry. As d243 was also analysed using the half-crossover assay biased gene conversion accompanying crossovers as well as biased non-exchange conversion were observed at 5A7.1 (Chapter 6).

At Superhotspots H and F, additional incidences of TD were also observed, with transmission ratios of 68:32 and 60:40, respectively (Chapter 4). Transmission ratios in non-exchange conversions were comparable to those in crossovers (Chapter 4, Chapter 6), as observed for Hotspot 5A, albeit the bias into non-exchange conversions often being insignificant. Combining transmission ratios of both types of recombinants strengthened the transmission distortion (Chapter 4). TD of 68:32, as observed at 5A and H, reflects 2.1-fold increased initiation efficiency on the active homologue compared to the suppressed homologue (Jeffreys and Neumann 2002). For Superhotspot F, the modest TD at SNP F6.1 reflects a 1.45-fold more efficient recombination initiation on the haplotype carrying the A-allele when taking allele-transmission in crossovers and non-exchange conversions into account. However these mild differences in crossover frequencies fall within the variation in primer efficiencies between assays and are therefore not readily detectable.

At Superhotspot F, all men that were heterozygous at F6.6b showed significant TD at this marker, both in this study as well as in the crossover survey by (Webb *et al.* 2008). This observation is consistent with F6.6b being necessary and sufficient to generate biased transmission, and is therefore likely to be the specific driver SNP marking active and inactive haplotypes at Superhotspot F. In contrast at Superhotspot H, significant TD was only observed at SNP H7.6 in man d25 but not man d52 (Chapter 4). Similarly at Hotspot 5A, d243 was the only man that displayed TD at marker 5A7.1. Neither SNP H7.6 nor 5A7.1 may therefore control the reciprocal crossover asymmetry itself. This phenomenon has been observed before. Modest distortion inconsistent in direction, and not favouring a particular hotspot allele has been seen at other Superhotspots (Webb *et al.* 2008; Jeffreys and Neumann 2009). As markers predicting active and inactive haplotypes were not clearly identifiable, the question remains what could be responsible for subtle difference in initiation efficiency between haplotypes. In Chapter 6, it was proposed that these differences might arise from interactions of *cis*- and *trans*-acting factors at a given hotspot.

In summary, about half of the human crossover hotspots analysed at sufficient resolution to date displayed some form of biased gene conversion accompanying crossover (Jeffreys and Neumann 2002; Jeffreys and Neumann 2005; Neumann and Jeffreys 2006; Webb *et al.* 2008; Jeffreys and Neumann 2009; Sarbajna *et al.* 2012), and hotspots analysed for non-exchange conversion in addition to crossover showed comparable degrees of TD in both types of recombinants at the same marker (Chapter 4) (Chapter 6) (Neumann and Jeffreys 2006; Sarbajna *et al.* 2012). The results obtained in this work thus recapitulate the frequency of biases observed at other hotspots, and indicated that incidences of subtler biases in recombination initiation may be fairly common. These data also point to non-crossovers being generally influenced by the initiation biases affecting crossovers.

7.3.2 Biases affecting only non-crossovers

Curiously, in addition to biases that affect both crossovers and non-crossovers, a bias specifically into non-crossovers was observed. This is the first observation that points to some component of crossover and non-crossover generation being independently regulated. Biased gene conversion specifically affecting non-crossovers was observed at two Superhotspots and it appeared that they were influenced in *cis*- by a single SNP heterozy-

gosity, both necessary and sufficient for controlling the bias, and determining the direction of allele-transmission. But alleles did not appear to be "activating" or "suppressing", as men homozygous for different alleles did not show observable differences in recombination frequencies. Therefore there is no support indicating that this phenomenon arises from initiation bias.

Gene conversions are believed to mainly arise through mismatch repair of heteroduplexes during recombination. Biased mismatch repair from the invaded duplex to the invading strand could readily explain biased gene conversion affecting both crossover and non-crossover (Jeffreys and Neumann 2002). For crossovers and non-exchange conversions to be affected differently, intermediates destined to become either crossover or non-crossover have to be sensed and processed differently, and secondly the process of mismatch repair would have to be biased towards converting particular alleles.

There have been proposals suggesting that the mismatch-repair process itself may be biased, as the human genome has been enriched with G and C-alleles particularly at the locations of recombination hotspots (Duret and Galtier 2009; Ratnakumar *et al.* 2010; Katzman *et al.* 2011; Marsolier-Kergoat 2011). Interestingly at hotspots K and F the bias into non-crossover always involved the over-transmission of G/C alleles, consistent with GC biased mismatch repair operating at human hotspots. But more hotspots need to be tested to see if this observation can be generalised.

7.4 Separate pathways for non-crossover generation?

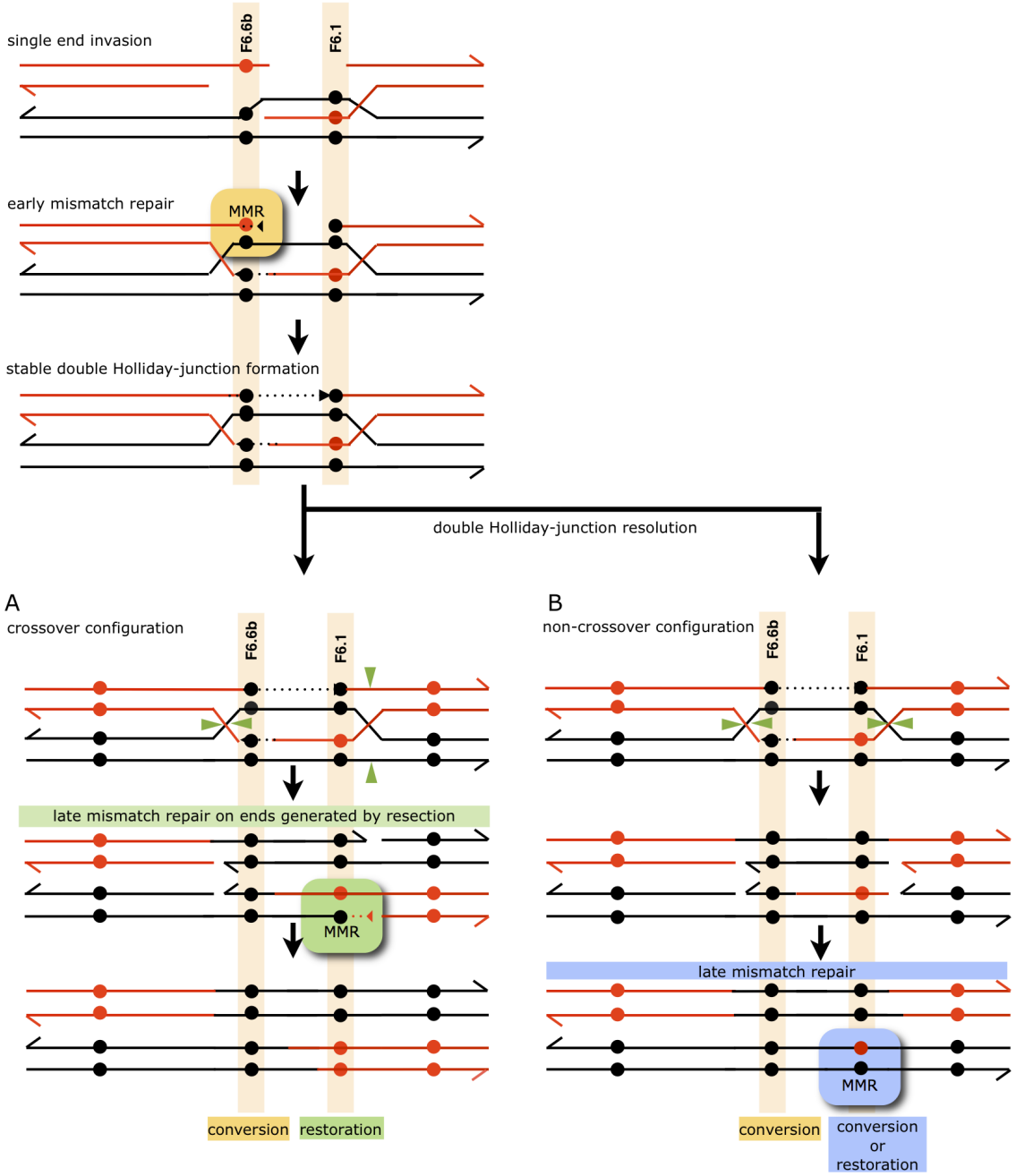
Intriguingly, Superhotspot F was an example of both types of biased gene conversion operating within the same hotspot. Firstly, biased gene conversion into crossovers and also non-crossovers was observed at marker F6.6b, located 472 bp away from the mean centre of the crossover hotspot. Significantly distorted transmission ratios were observed for all men that were heterozygous at this marker, not only in this study (4.3.2) but also in the data of Webb *et al.*, (Webb *et al.* 2008). Secondly, biased gene conversion specifically into non-crossover was observed at the most central marker F6.1 with no indication for biased TD in crossovers. The degree of TD at marker F6.1 was significantly different between crossovers and non-exchange conversions (Chapter 4).

To date all observations of recombination events at human recombination hotspots were explainable by the simple canonical double-strand repair model of recombination (DSBR). If crossovers and non-crossovers are generated as alternative products of the same dHJ molecule this readily provides an explanation for biased gene conversion in crossover and non-crossover being influenced to the same degree. Heteroduplex DNA within a dHJ intermediate is subjected to MMR, with repair being directional from the invaded to the invading strand changing the allele of the open DNA end. Biased gene conversion accompanying crossover has shown TD in favour of CG at Hotspots *DNA*2 (Jeffreys and Neumann 2002) and *SPRY*3 (Sarbajna *et al.* 2012) as well as AT alleles at Superhotspot S2 (Jeffreys and Neumann 2009) and *NID*1 (Jeffreys and Neumann 2005). This observation is consistent with early mismatch repair being biased towards repairing from the invaded duplex to the invading strand and not towards particular alleles.

But how can non-exchange conversion molecules that contain the most central marker be biased only in non-crossover, while recombinant molecules that contain the off-centre marker are biased in both crossover and non-crossover molecules? One model based on the canonical DSBR model with the potential to explain how these two biases arise within the same intermediate is illustrated in Figure 7-1.

In the DSBR model of recombination, crossover and non-crossovers are produced from the same recombination intermediate by resolution in different resection planes. It could therefore be possible that intermediates destined to become crossover and non-crossover are sensed and processed differently by MMR processes that result in gene conversion. Late mismatch repair operating on the ends generated by this resection process may provide an explanation for the bias observed only in non-exchange conversions. Resolution in non-crossover configuration requires cleavage of internal strands, potentially creating molecules containing marker F6.1 spanning hDNA that cannot simply be repaired by late mismatch repair (as illustrated in Figure 7-1). The necessity to repair these mismatches may then require excision repair, which might be biased toward generating alleles of a particular type. This could explain biased gene conversion observed purely in non-crossover at F6.1 within the same pathway that also generated biased gene conversion in crossover and non-crossover at marker F6.6b.

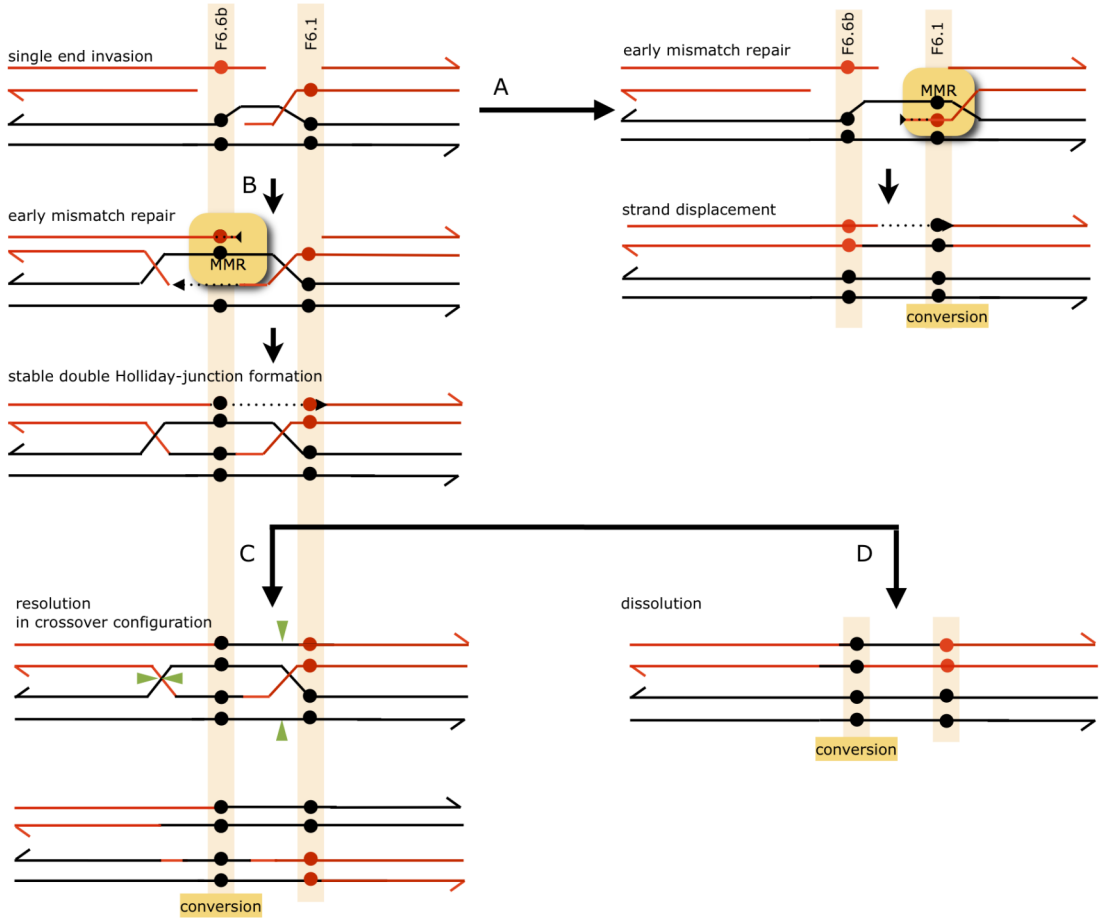
Figure 7-1 Proposed recombination model



Model is presented in chronological order, from top to bottom after DNA double-strand break-formation on the red chromatid has already taken place and ends have been resected in 5′ to 3′ direction. For simplicity only the two chromatids directly involved in the break repair process are shown, indicated by lines. Chromosomal origin is differentiated by colour, with heterozygous markers represented by circles. Mismatch repair (MMR) at the location of heteroduplex DNA is indicated by orange box. In this model, a mismatch at F6.1 is not repaired by early mismatch repair as it is located such that a long tract would have to be chewed back, jeopardising stable strand exchange. Once the 3′-end is recaptured a double HJ intermediate is formed with this base still mismatched. To release chromatids, the dHJ is resected (green triangles). Resolution either in crossover or non-crossover configuration is possible, depending on the resection plane. (A) A nick created by resolution in crossover configuration allows for restoration at F6.1. (B) Resolution in non-crossover configuration does not allow F6.1 to be repaired by nick-directed mismatch repair. To repair the mismatch invokes creation of a nick (possibly by excision repair).

It is possible to explain both conversion accompanying crossover as well as non-exchange conversion bias within the framework of the canonical DSBR pathway. However it is also tempting to speculate that crossovers and a proportion of non-crossovers are generated as alternative resolution pathways of the same initiation complex, with additional non-crossovers generated in a separate pathway, as illustrated in Figure 7-2.

Figure 7-2 Proposed two-pathway model of meiotic recombination in humans



Model is presented in chronological order, from top to bottom after DNA double-strand break-formation on the red chromatid has already taken place and ends have been resected in 5′ to 3′ direction. For simplicity only the two chromatids directly involved in the break repair process are shown, indicated by lines as in Figure 7-1. The free end anneals to the complementary strand in the invaded duplex, initiating DNA synthesis (hashed lines). (A) Heteroduplex DNA contains marker F6.1. A mismatch at F6.1 has to be repaired, delaying D-loop formation, the 3′end is then recaptured without formation of a junction intermediate, and DNA synthesis on the top strand completes gap repair leading only to non-exchange gene conversions. (B) Alternatively, F6.1 is not contained in hDNA and does not require early MMR, therefore a full double Holliday junction intermediate can be formed (C) the stable dHJ intermediate is then processed down the canonical DSBR-pathway (as shown in Figure 7-1) that can lead to both crossover (and potentially also non-exchange conversion molecules) that show conversion of F6.6b. (D) Dissolution of the dHJ can also generate non-exchange conversions at marker F6.6b.

Two pathways that generate non-crossover recombinants have been proposed in *S. cerevisiae* recombination, with the DSBR pathway mainly responsible for crossover formation and the SDSA pathway responsible for non-crossover generation (Allers and Lichten 2001a). The crossover/non-crossover decision likely occurs before stable strand exchange (Hunter and Kleckner 2001; Bishop and Zickler 2004; Börner *et al.* 2004). Non-crossover intermediates containing hDNA are believed to be present at the same time as dHJ form (Allers and Lichten 2001a). A bifurcating pathway from the point of single end invasion molecule formation has been proposed by Börner *et al.,* where single-end invasion molecules are formed and lead to a displacement loop in the invaded homologue (Börner *et al.* 2004). Either a dHJ is formed after engagement with the remaining free end (Figure 7-2 B), or alternatively a transient D-loop with DNA synthesis from the invaded homologue dissociates and forms non-crossovers much as proposed in the mitotic SDSA pathway (Börner *et al.* 2004) as illustrated in Figure 7-2 A. If separate pathways generate crossover and non-crossover, the most central markers at Superhotspot F would intriguingly have to be more likely involved in intermediates destined for non-crossover, while the off-centre marker would be equally likely contained in intermediates destined for non-crossover or crossover. Shorter hDNA tracts at the centre of the hotspot, especially if engaged in MMR at several markers may result in dissolution of dHJ intermediates, and therefore in non-crossovers. While longer tracts may facilitate stable strand exchange, and therefore full dHJ formation and crossover formation (as illustrated in Figure 7-2 C).

Evidence for two pathways of non-crossover generation in the human genome comes from observation of Sarbajna *et al.* (2012) at minor pseudo-autosomal hotspot *SPRY*3. They observed high levels of variation between CON:CO ratios, that were most readily explained by at least a proportion of non-exchange conversions being generated via a second, perhaps the SDSA, pathway (Sarbajna *et al.* 2012).

7.5 Conclusion

7.5.1 *Cis*- and *trans*- regulating factors facilitate the turnover of the human recombination landscape.

The regulation of recombination in the human genome appears to be extremely dynamic, with hotspot generation and extinction through biased gene conversion going on constantly. PRDM9 activation of hotspots and biased gene conversion might form an evolutionary model in which new recombination hotspots were always re-introduced by the generation of new PRDM9 variants, and then silenced over time through mutation and biased gene conversion. This model has the ability to fully resolve the hotspot paradox. Hotspots would not have to persist despite being silenced by biased gene conversion, but instead new locations were activated as recombination hotspots by the generation of new PRDM9 zinc-finger arrays (Berg *et al.* 2010).

Significant biased gene conversion either into crossovers and non-exchange conversions has been found at three hotspots, and biases specifically into non-crossover were found at two hotspots in this survey. As one of the hotspot simultaneously shows both types of biases, biases have therefore been identified at five out of six of the analyzed hotspots. Biased gene conversion that influences crossovers and non-crossovers was observed at Hotspots activated by PRDM9 variants A and C, often with meiotic drive strong enough to lead to eventual fixation of the suppressing allele (Chapter 6), (Jeffreys and Neumann 2005; Berg *et al.* 2011). It appears that crossover and non-crossover were not necessarily influenced by the same biases and that biases were very common, at least at human recombination Superhotspots. Previous studies based only on biased gene conversion associated with crossovers underestimated the degree of meiotic drive generated within recombination hotspots. At all hotspots where non-exchange conversions were also tested, biased gene conversion into non-crossovers occurred to the same degree as crossover, resulting in a much stronger effect of meiotic drive within recombination hotspots. This argues for a rapid elimination of hotspots by biased gene conversion, sufficient to explain the negligible overlap observed between recombination patterns of humans and chimpanzees (Ptak *et al.* 2005; Winckler 2005).

If PRDM9 binding sites are constantly depleted by biased gene conversion, then at a critical point most alleles would have been converted to suppressing alleles, which could result in insufficient numbers of DSBs on which recombination can act. At this point, crossover homeostasis would fail, which would lead to non-disjunction. This could have deleterious effects resulting in fertility-selection. There is evidence for positive selection on the ZnF residues involved in DNA binding, consistent with these being adaptations to rapidly evolving DNA sequences to which they bind. Pairs of human ZnF share total sequence identity at 19 of 28 codons and differences virtually always occur at the three amino acids that contact the major groove of double-stranded DNA, exceeding the neutral substitution rate by 10-fold (Ponting 2011).

A neutral system of recombination landscape evolution can only be achieved if *PRDM9* should evolve at equal rate, or more rapidly, than the time it takes for a motif to be completely depleted (Ponting 2011). It has been proposed that hotspots are created and destroyed rapidly, but at nearly equal rates following Red Queen dynamics (Ubeda and Wilkins 2011).

7.6 Future work

The question remains as to what exactly influences *PRDM9* binding *in vivo*. Firstly, it remains to be understood how a large ZnF array behaves *in vivo*, and whether more than the initially proposed five ZnFs are involved in DNA binding. Secondly, it is also important to determine whether PRDM9 regulation is fully recapitulated by *in-vitro* binding and if PRMD9 specificity is depended on local hotspot specific chromatin-complexes. Hotspots in mouse are enriched for H3K4 tri-methylation patterns, (Grey *et al.* 2011) and it remains to be understood whether these arise through direct interactions with PRDM9, as has been suggested by (Grey *et al.* 2011).

In addition the evolution of *PRDM9* itself is most intriguing, especially of those regions that code for amino-acids responsible for DNA binding. PRDM9 is the most diverged C2H2-containing ZnF protein between human and chimpanzee (Myers *et al.* 2010). A minisatellite codes for the PRMD9 C2H2 ZnF domain, responsible for DNA binding, and it is this domain that was the most variable between men (Berg *et al.* 2010). Minisatellites are very instable loci and known to mutate mostly via gene-conversion-like transfer of blocks of repeats, generating hetero-allelic repeat arrays (Jeffreys *et al.* 1995). However,

somatic recombination processes like replication slippage as well as mis-repair may also foster additional minisatellite instability but at lower frequencies.

Human PRDM9 ZnFs exhibit traces of multiple events of positive selection, insertion, deletion and gene conversion (Ponting 2011). Additionally, large variation in the number of PRDM9 ZnFs, ranging from 8-18 ZnFs in the C-terminal array, were observed in humans (Berg *et al.* 2010). Together these observations suggest that the PRDM9 minisatellite may be evolving by classical meiotic minisatellite processes. Additionally, direct evidence for meiotic processes driving PRDM9 repeat instability in the germline has been observed by A.J. Jeffreys (*personal communication*). If the PRDM9 minisatellite evolves by the classic meiotic minisatellite evolution process, then it might be driven by a flanking hotspot. Thus it would be interesting to analyse the *PRDM9* minisatellite for evidence of non-exchange conversion.

Intriguingly, PRDM9 itself has been shown to influence minisatellite instability (Berg *et al.* 2010). If PRDM9 also influences repeat instability of its own minisatellite, the evolution of new *PRDM9* alleles would be dependent on PRDM9 itself. This concept of a "perpetual mobile" of recombination landscape evolution is most fascinating. However, a new PRDM9 allele with the ability to influence meiotic instability at the *PRDM9* minisatellite would lead to bouts of explosive increase of allele diversity at the *PRDM9* locus. This would be followed by gradual conversion of recombination active alleles through mutation and biased gene conversion, which would eventually equally silence the PRDM9 hotspot (if it existed).

Appendix

I. Allele specific Primers

Appendix 7-I Allele specific primers at Hotspots E, F, H and K

SNP	allele	name	sequence 5'-->3'	specific annealing temperatures (°C)
Hotspot E				
rs10783158	C	E10.6RC	AGCCTGGTCAACCTGG	59-65+
"	T	E10.6RT	CAGCCTGGTCAACCTGA	62-65+
rs10666923	–	E12.4R–	ACCCAGTTTACTACACAGT	56-59
"	+	E12.4R+2	ATTGACCCAGTTTACTACACATAGT	56-65+
Hotspot F				
rs11612965	G	F3.8aFG2	GYACCAACCCAGTCACG	62-65+
"	A	F3.8aFA2	GCACCAACCCAGTCACA	59-65+
rs10492183	T	F3.9FT	ccccCAATGCTACTCTTCTGGAT	59-65+
"	C	F3.9FC	ccccCAATGCTACTCTTCTGGAC	65+
rs445467	T	F13.0RT2	CCAGCATGGGCAAGACA	56-65+
"	G	F13.0RG2	CAGCATGGGCAAGACC	59-65+
rs368789	C	F13.3RC	CGTTTGTGGATAGACCAGG	62-65+
"	T	F13.3RT	CGTTTGTGGATAGACCAGA	(59)-62-65+
Hotspot H				
rs7651825	A	H2.0FA	TCCCAGAGCTACCTCAAA	(55)-59
"	G	H2.0FG	TCCCAGAGCTACCTCAAG	59-65+
rs11716175	A	H2.5FA	GCTCCAGAAGGACCCTCA	62-65
"	G	H2.5FG	CTCCAGAAGGACCCTCG	59-65
Hotspot K				
rs4735211	C	K2.9FC	CCAGGTCCCCTGTCTC	(55)-59-65
"	T	K2.9FT2	GTTCCAGGTCCCCTGTCTT	(55-59)-62-65
rs10090492	C	K3.2FC2	ccccGAAAAACAGCACATCAGAC	55-65
"	G	K3.2FG2	ccccGAAAAACAGCACATCAGAG	(55)-59-65
rs7829351	A	K10.9RA	cccCCCCCCTTAATTTTTAAACT	59-65
"	G	K10.9RG	cccCCCCCCTTAATTTTTAAACC	(55)-59-65
rs766477	C	K11.7RC	TGAAGGTAGAATTGGCCG	55-59-(62)
"	T	K11.7RT	GTGAAGGTAGAATTGGCCA	55-65

Bases used to achieve allele-specificity are indicated in red, lowercase bases (blue) are synthetic poly-Cytosine tails, ambiguity bases are indicated in green.

Allele specific primers continued…

Appendix 7-II Allele specific primers at Hotspots T and 5A

SNP	allele	name	sequence 5'-->3'	specific annealing temperatures (°C)
Hotspot T				
rs17013514	A	T3.3FA	TCAGCCTCAATGGTAACA	(58)-62-(65)
"	G	T3.3FG	TCAGCCTCAATGGTAACG	(58)-62-65
rs6808240	T	T3.6FT	CCCAGGTCATTTACTACAT	54-(58)
"	C	T3.6FC	CCCAGGTCATTTACTACAC	(54)-65
rs12636837	T	T9.1RT	CCTTCTGGGAACCATGTA	(54)-58
"	C	T9.1RC	CCTTCTGGGAACCATGTG	58-62(-65)
rs66906407	A	T9.3RA	ACTTTACCACCTGGCCAT	(58)- 65
"	G	T9.3RG	ACTTTACCACCTGGCCAC	(54-58) 62
rs11456124	minus(-)	T9.6R-	ccccTGCAATTTTCATGTACTC	54-58-(62)
"	plus(+)	T9.6R+	ccccGCAATTTTCATGTATCTC	54-58-(62)
rs13090158	T	T10.8RT	ccccCTTTGTTTTCTGTCTTCTA	54-58
	C	T10.8RC	ccccCTTTGTTTTCTGTCTTCTG	54-62
Hotspot 5A				
rs10067185	C	5A9.9RC	TCTCACAGCACAGGTTGG	54-58
"	G	5A9.9RG	TCTCACAGCAGAGGTTGC	54-58
rs4304112	C	5A9.9aRC	CTTGTGATCTCACAGCAG	58-62(-65)
"	G	5A9.9aRG	CTTGTGATCTCACAGCAC	58-62(-65)

II. Universal Primers

Universal primer	sequence 5'-->3'	used in 1/2 CO assay	used for Sequencing	genotyping/ sequencing target amplification
Hotspot E				
E4.9F	GAGTATGAATGGTGTCTCCC	√		
E5.6F	TAACCATGCACTGAGCTCCC	√		
E 5.6aF	CTGAAGAAGCAATGAGGAACGAG			√
E 9.1 R	GAACATGCAGCAGGATACAG			√
E 6.1 F	GTGCCTTGCACACATAGGAT		√	
E 6.7 F	TTCAGCGAGCTAGAAGCCTT		√	
E 7.5 F	ACCTGTCTCCGTGGTTCTAA		√	
E 7.3 R	GGTGCATTCCCAGGAGACA		√	
E 8.5 R	ACTCCTCTCCTCATTCTGCT		√	
Hotspot F				
F3.5F	CAACTCCAGACCTACTGACC	√		
F4.0F	ATTGATCCAAGGTCTGGAGG	√		
F9.5R	AGTACATATATCCTGGGGCC	√		
F9.7R	ATCCAAATGGCCAATGAGCC	√		
F 5.1 F	AAGTAACCTCTGCTGATCCC			√
F 8.7 R	GACAGCAGGAAAGAGCCAC			√
F 5.1a F	AACCTCTGCTGATCCCCAGC		√	
F 5.5 F	GTGGTTGGGAAAGAGACAGC			√
F 6.0 F	CAGATGAAATAGCACTGGGGC			√
F 6.3 R	CCTAGAACAAGGTGTGGAGC		√	
F 7.0 R	GCTCAGAATGATGGTTTCCAGC		√	
Hotspot H				
H6.1F	ACAGGCCCAGCGAATGTCC	√	√	√
H6.7F	GAATCCTTTCTTGTTCAGGCTGA		√	
H7.3F	AAGGAGACTGAAGCCCAGA		√	
H7.9R	CAGTTTGTGGTACACCGT		√	
H8.6R	CAGACATCCTGCTGGTTCT		√	
H10.3R	ATCTCTACTCCTACCCCTAC	√		√
H12.3R	ATACACGGCAGGCTTTGCCC	√		
Hotspot K				
K4.4F	CACATGTCAGCACACATTGCCC	√		
K4.5F	GCGAGGAGTCCTAAACTCCC	√	√	√
K6.5F	ACCAAAATCTTCAGCTGCCC		√	
K6.7R	CAGCCACCTGAGGCCCC		√	
K7.2F	GTTCCTTCTTTCCACTGCTT		√	
K10.8R	CTTGGCTGAAGCCTCTCCC	√	√	√
K11.0R	TTGCCTCTCAGCCTGCTGCA	√		
Hotspot T				
T 0.4 F	TTTCTCTAGCTTCCTCACCC	√		√
T 0.4 aF	GTTACTTCTGTGCTGTGCCC	√	√	
T 6.5 R	ATTCCCTAGGACCCAGACCC	√		√
T 6.5 aR	ACCTAATCCCTCCAGGCCTC	√		
T1.1F	GGCCCAACATACATTGGCTC	√		
T1.3F	GAGGACAGTCCTGGCTGTG	√	√	
T 3.8 R	CTCTTTCCTCTCTTAGGCATCCC		√	
T 5.7R	AAGTAAGATAAAGGCCAGGC		√	
T6.2F	CCTATCCAGCATGTGAATCC			√
T 6.4 F	CCTTAATCTTCCCACAACCC		√	
T6.5 aR	ACCTAATCCCTCCAGGCCTC		√	
T 6.9 F	GTTGGAGACCTTCTGTGTATG		√	
T 7.4 F	GTTAGAGAATGCCAAGAAGAG		√	
T 8.3 R	GACTCTGCCTTAAGTATGCT		√	
T 8.2 F	CTACAGTAATATCAAATCAC		√	
T 9.1 R	CCCTTCTGGGAACCATGT		√	
T 9.6 R	GCTTCTCTTACATAAACGTCTCC		√	√
Hotspot 5A				
5A3.6F	AGCTGACATGGCCAAGTTCC	√		
5A3.7F	CCTGGCGAATAGCTCTTTCC	√		

III. PCR conditions in half-crossover assays

Hotspot	ASPs located	orientation	PCR type	Primer combination	cycling conditions	PCR machine
E	reverse -->	A	1°	E4.9F & E12.4R+2	96 °C, 1.5 min; (96 °C, 20 s; 58 °C, 30 s; 65 °C, 9 min) x 25; 65 °C, 5 min; hold at 10 °C	Veriti
	reverse -->		2°	E5.6F & E10.6RT	94 °C, 1.5 min; (94 °C, 20 s; 62 °C, 30 s; 65 °C, 6min) x 35; 65 °C, 5 min; hold at 10 °C	Veriti
	reverse -->	B	1°	E4.9F & E12.4R-	96 °C, 1.5 min; (96 °C, 20 s; 58 °C, 30 s; 65 °C, 9 min) x 25; 65 °C, 5 min; hold at 10 °C	Veriti
	reverse -->		2°	E5.6F & E10.6RC	94 °C, 1.5 min; (94 °C, 20 s; 62 °C, 30 s; 65 °C, 6 min) x 35; 65 °C, 5 min; hold at 10 °C	Veriti
F	reverse -->	A	1°	F3.5F & F13.3RC	96 °C, 1.5 min; (96 °C, 20 s; 63 °C, 30 s; 65 °C, 11 min) x 25; 65 °C, 5 min; hold at 10 °C	G-Storm
	reverse -->		2°	F4.0F & 13.0RT2	96 °C, 1.5 min; (96 °C, 20 s; 63 °C, 30 s; 65 °C, 10 min) x 35; 65 °C, 5 min; hold at 10 °C	G-Storm
	reverse -->	B	1°	F3.5F & F13.3RT	96 °C, 1.5 min; (96 °C, 20 s; 63 °C, 30 s; 65 °C, 11 min) x 25; 65 °C, 5 min; hold at 10 °C	G-Storm
	reverse -->		2°	F4.0F & 13.0RG2	96 °C, 1.5 min; (96 °C, 20 s; 63 °C, 30 s; 65 °C, 10 min) x 35; 65 °C, 5 min; hold at 10 °C	G-Storm
	<-- forward	A	1°	F3.8aFG2 & F9.7R	96 °C, 1.5 min; (96 °C, 20 s; 64 °C, 30 s; 65 °C, 9 min) x 25; 65 °C, 5 min; hold at 10 °C	9700
	<-- forward		2°	F3.8FT & F9.5R	96 °C, 1.5 min; (96 °C, 20 s; 60 °C, 30 s; 65 °C, 7 min) x 35; 65 °C, 5 min; hold at 10 °C	9700
	<-- forward	B	1°	F3.8aFA2 & F9.7R	94 °C, 1.5 min; (94 °C, 20 s; 62 °C, 30 s; 65 °C, 9 min) x 25; 65 °C, 5 min; hold at 10 °C	9700
	<-- forward		2°	F3.8FC & F9.5R	96 °C, 1.5 min; {(96 °C, 20 s; 65 °C, 30 s; 67 °C, 7 min) x 10; (96 °C, 20 s; 64 °C, 30 s; 65 °C,7 min) x 25}; 65 °C, 5 min; hold at 10 °C	9700
H	<-- forward	A	1°	H2.0FA & H12.3R	94 °C, 1.5 min; (94 °C, 20 s; 61 °C, 30 s; 65 °C, 11 min) x 25; 65 °C, 5 min; hold at 10 °C	Veriti
	<-- forward		2°	H2.5FA & H10.3R	94 °C, 1.5 min; (94 °C, 20 s; 64 °C, 30 s; 65 °C, 9 min) x 35; 65 °C, 5 min; hold at 10 °C	Veriti
	<-- forward	B	1°	H2.0FG & H12.3R	94 °C, 1.5 min; (94 °C, 20 s; 62 °C, 30 s; 65 °C, 11 min) x 25; 65 °C, 5 min; hold at 10 °C	Veriti
	<-- forward		2°	H2.5FG & H10.3R	94 °C, 1.5 min; (94 °C, 20 s; 61 °C, 30 s; 65 °C, 9 min) x 35; 65 °C, 5 min; hold at 10 °C	Veriti
K	reverse -->	A	1°	K4.4F & K 11.7RT	96 °C, 1.5 min; {(96 °C, 20 s; 65 °C, 30 s; 65 °C, 9 min) x 10; (96 °C, 20 s; 64 °C, 30 s; 65 °C, 9 min) x 15}; 65 °C, 5 min; hold at 10 °C	Tetrad
	reverse -->		2°	K4.5F & K10.9RG	96 °C, 1.5 min; (96 °C, 20 s; 62 °C, 30 s; 65 °C, 8 min) x 35; 65 °C, 5 min; hold at 10 °C	Tetrad
	reverse -->	B	1°	K4.4F & K 11.7RC	96 °C, 1.5 min; (96 °C, 20 s; 61 °C, 30 s; 65 °C, 9 min) x 25; 65 °C, 5 min; hold at 10 °C	Tetrad
	reverse -->		2°	K4.5F & K10.9RA	96 °C, 1.5 min; (96 °C, 20 s; 62 °C, 30 s; 65 °C, 8 min) x 35; 65 °C, 5 min; hold at 10 °C	Tetrad
	<-- forward	A	1°	K2.9FC & K11.0R	96 °C, 1.5 min; (96 °C, 20 s; 60 °C, 30 s; 65 °C, 9 min) x 28; 65 °C, 5 min; hold at 10 °C	9700
	<-- forward		2°	K3.2FC2 & K10.8R	96 °C, 1.5 min; {(96 °C, 20 s; 61 °C, 30 s; 67 °C, 7 min) x 10; (96 °C, 20 s; 60 °C, 30 s; 65 °C, 7 min) x 25}; 65 °C, 5 min; hold at 10 °C	9700
	<-- forward	B	1°	K2.9FT2 & K11.0R	96 °C, 1.5 min; (96 °C, 20 s; 60 °C, 30 s; 65 °C, 9 min) x 25; 65 °C, 5 min; hold at 10 °C	9700
	<-- forward		2°	K3.2FG2 & K10.8R	96 °C, 1.5 min; {(96 °C, 20 s; 61 °C, 30 s; 67 °C, 7 min) x 10; (96 °C, 20 s; 60 °C, 30 s; 65 °C,7 min) x 25}; 65 °C, 5 min; hold at 10 °C	9700
T	reverse -->	A	1°	T1.1F & T10.8RC	94 °C, 1.5 min; (94 °C, 20 s; 62 °C, 30 s; 65 °C, 11 min) x 25; 65 °C, 5 min; hold at 10 °C	Veriti
	reverse -->		2°	T1.3F & T 9.6+	94 °C, 1.5 min; (94 °C, 20 s; 58 °C, 30 s; 65 °C, 9.5 min) x 35; 65 °C, 5 min; hold at 10 °C	Veriti
	reverse -->	B	1°	T1.1F & T10.8RT	94 °C, 1.5 min; (94 °C, 20 s; 58 °C, 30 s; 65 °C, 11 min) x 25; 65 °C, 5 min; hold at 10 °C	Veriti
	reverse -->		2°	T1.3F & T 9.6-	94 °C, 1.5 min; (94 °C, 20 s; 58 °C, 30 s; 65 °C, 9.5 min) x 35; 65 °C, 5 min; hold at 10 °C	Veriti
5A	reverse -->	A	1°	5A3.6F & 5A9.9aRG	96 °C, 1.5 min; (96 °C, 20 s; 55 °C, 30 s; 65 °C, 7.5 min) x 25; 65 °C, 5 min; hold at 10 °C	Veriti
	reverse -->		2°	5A3.7F & 5A9.9RC	96 °C, 1.5 min; (96 °C, 20 s; 58 °C, 30 s; 65 °C, 7.5 min) x 32; 65 °C, 5 min; hold at 10 °C	Veriti
	reverse -->	B	1°	5A3.6F & 5A9.9aRC	96 °C, 1.5 min; (96 °C, 20 s; 55 °C, 30 s; 65 °C, 7.5 min) x 25; 65 °C, 5 min; hold at 10 °C	Veriti
	reverse -->		2°	5A3.7F & 5A9.9RG	96 °C, 1.5 min; (96 °C, 20 s; 55 °C, 30 s; 65 °C, 7.5 min) x 32; 65 °C, 5 min; hold at 10 °C	Veriti

IV. Allele specific Oligonucleotides

Appendix 7-III ASOs used for hybridisation at hotspots E, F and H

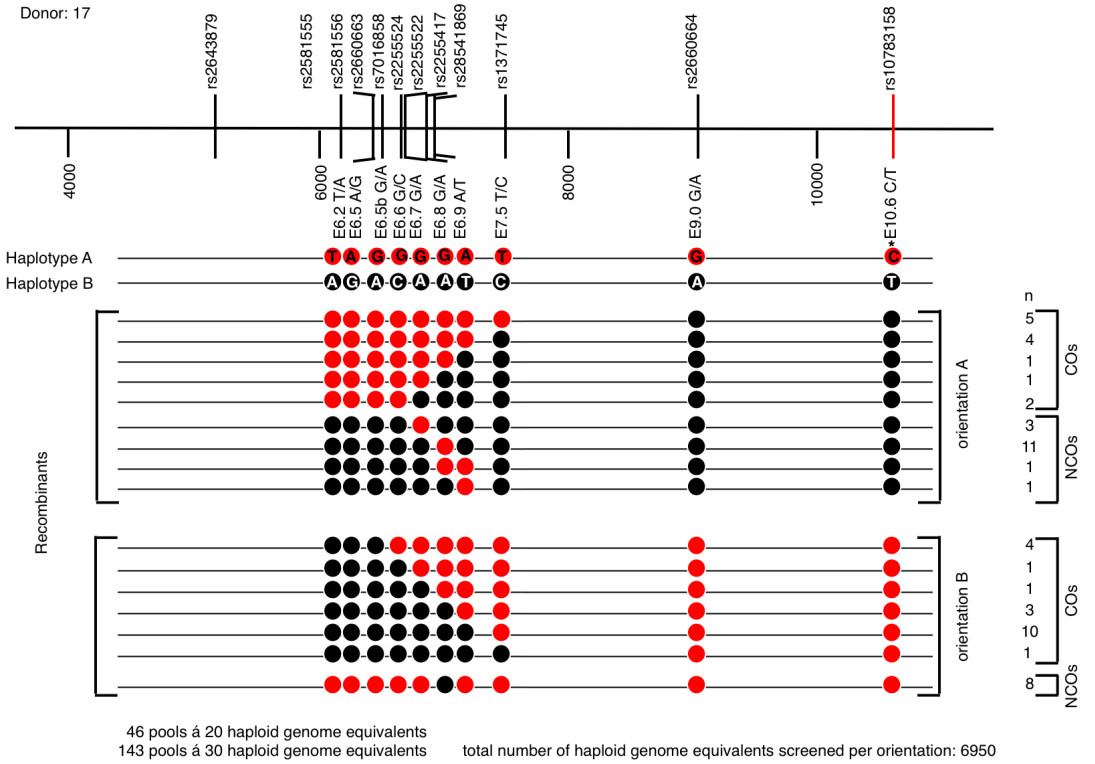
Hotspot	rs number	location	Allele	Name of probe	Sequence 5'-->3'	Hybridisation temperature (°C)	Washing temperature (°C)	ASO quality
E	2581556	6191	A	E6.2A2	GACATTTTGTTGCAGAGA	48.5	54	***
			T	E6.2T2	GACATTTAGTTGCAGAGA	48.5	54	***
E	2660663	6470	G	E6.5G	GTATATATTGTATATGTA	48.5	54	***
			A	E6.5A	GTATATATTATATATGTA	48.5	54	***
E	7016858	6507	G	E6.5bG2	GTAGTATCCATAAAATAT	48.5	54	***
			A	E6.5bA2	GTAGTATCTATAAAATAT	48.5	54	***
E	2255524	6653	G	E6.6G	ACCAAAAGGTACCAATGC	50	56	***
			C	E6.6C	ACCAAAACGTACCAATGC	48.5	56	***
E	2255524	6700	A	E6.7A	TGTGACGATGGTGCAAGG	48.5	54	***
			G	E6.7G	TGTGACGGTGGTGCAAGG	48.5	54	***
E	2255417	6871	G	E6.8G2	CCTCTCTCCGCCTTCTAT	50	56	***
			A	E6.8A2	CCTCTCTCTGCCTTCTAT	48.5	54	***
E	28541869	6928	A	E6.9A2	CTCCAGCTGTTGCTAGTG	48.5	54	***
			T	E6.9T2	CTCCAGCAGTTGCTAGTG	48.5	54	***
E	1371745	7493	C	E7.5C2	AATGACTGCTGTTGGAGT	48.5	56	***
			T	E7.5T2	AATGACTACTGTTGGAGT	50	56	***
E	2660664	9043	A	E9.0A	GCCCTGGATATTTTGGTT	48.5	54	***
			G	E9.0G	GCCCTGGGTATTTTGGTT	48.5	54	***
F	11063788	4958	T	F5.0T	GAGAAAACTCCACAGAAA	48.5	54	***
			G	F5.0G	GAGAAAACGCCACAGAAA	48.5	54	***
F	10849308	5366	T	F5.4T2	CGTGCTGAAGGTACTACT	48.5	54	***
			C	F5.4C2	CGTGCTGGAGGTACTACT	48.5	54	***
F	365211	5554	T	F5.5T2	CCACTCCACCTGTCCCC	48.5	54	***
			A	F5.5A2	CCACTCCTCCTGTCCCC	48.5	54	***
F		5625	C	F5.6aC	CCTCAGGCGATGGGAAGC	48.5	54	***
			A	F5.6aA	CCTCAGGAGATGGGAAGC	48.5	54	***
F	393861	5757	G	F5.8G	GATGAGAGTCCAGGCTGT	48.5	54	***
			A	F5.8A	GATGAGAATCCAGGCTGT	48.5	54	***
F	11063791	5878	T	F5.9T	TCGCCTCTGACAATTCCC	48.5	54	***
			C	F5.9C	TCGCCTCCGACAATTCCC	48.5	54	***
F	386440	5991	A	F6.0A2	CTTGTGTTAGAAACCAGC	48.5	54	***
			T	F6.0T2	CTTGTGTAAGAAACCAGC	48.5	54	***
F	10492181	6073	A	F6.1A2	TGCCAATCTACCGAGTCA	48.5	54	****
			G	F6.1G2	TGCCAATCCACCGAGTCA	48.5	54	****
F	384401	6551	G	F6.6G2	GGGATGGTACCTCATTGT	49°	56	**
			A	F6.6A2	GGGATGGTATCTCATTGT	49°	56	**
F	11063792	6642	A	F6.6bA	CTTTTACACTGTTGGTGG	48.5	54	***
			G	F6.6bG	CTTTTACGCTGTTGGTGG	48.5	54	***
F	374793	7067	A	F7.1A	TGGACACAGGGCGAGGAA	48.5	54	***
			G	F7.1G	TGGACACGGGGCGAGGAA	48.5	54	***
F	417672	7501	T	F7.5T2	GAATGGGGAGAGCCCC	48.5	52	***
			C	F7.5C2	GAATGGGGGGAGCCCC	48.5	54	***
F	406430	10630	A	F10.6A	TCGGTTGATAGCAAACTC	48.5	54	***
			G	F10.6G	TCGGTTGGTAGCAAACTC	48.5	54	***
H	829182	5600	G	H5.6G	TGAAATGGCTTGCGATAT	48.5	56	***
			A	H5.6A	TGAAATGACTTGCGATAT	48.5	56	***
H	829183	6099	C	H6.1C2	CTAAGGTGCATTTGATTC	48.5	54	***
			T	H6.1T2	CTAAGGTACATTTGATTC	50	56	**
H	829184	6592	T	H6.6T	TCACCTCTCAGGTTATCA	50	56	**
			C	H6.6C	TCACCTCCCAGGTTATCA	48.5	56	**
H	864361	7209	G	H7.2aG	TTTGTTGGTCTTGTTACA	48.5	54	****
			T	H7.2aT	TTTGTTGTTCTTGTTACA	very low specificity		
H	864361	7209	G	H7.2aG2	AACAAGACCAACAAAAAA	very low specificity		
			T	H7.2aT2	AACAAGAACAACAAAAAA	48.5	54	***
H	853334	7501	G	H7.5G2	ATGTTATCCAGTATTTCT	48.5	56	***
			A	H7.5A2	ATGTTATTCAGTATTTCT	48.5	56	***
H	3899614	7574	G	H7.6G2	TGTGTTTCAGCCCCTGAA	48.5	56	***
			A	H7.6A2	TGTGTTTTAGCCCCTGAA	48.5	56	****
H	829185	8078	C	H8.1C2	GACATTGGACATCTTGTC	48.5	56	***
			T	H8.1T2	GACATTGAACATCTTGTC	48.5	54	***
H	829186	8563	A	H8.6A2	TTGGACCTAGGGGTTCTT	48.5	56	***
			G	H8.6G2	TTGGACCCAGGGGTTCTT	50	56	****
H	829187	8634	A	H8.6aA2	TGGGTGGTTTCCGTCGGA	50	56	**
			G	H8.6aG2	TGGGTGGCTTCCGTCGGA	50	56	**
H	829188	8677	T	H8.7aT2	GGAGCCAGCAGTTCATGT	50	56	**
			C	H8.7aC2	GGAGCCAGCGGTTCATGT	50	56	***
H	13072398	9499	A	H9.5A2	AGATTTATTGTCACTCC	50	56	**
			G	H9.5G2	AGATTTACTGTCACTCC	50	56	**

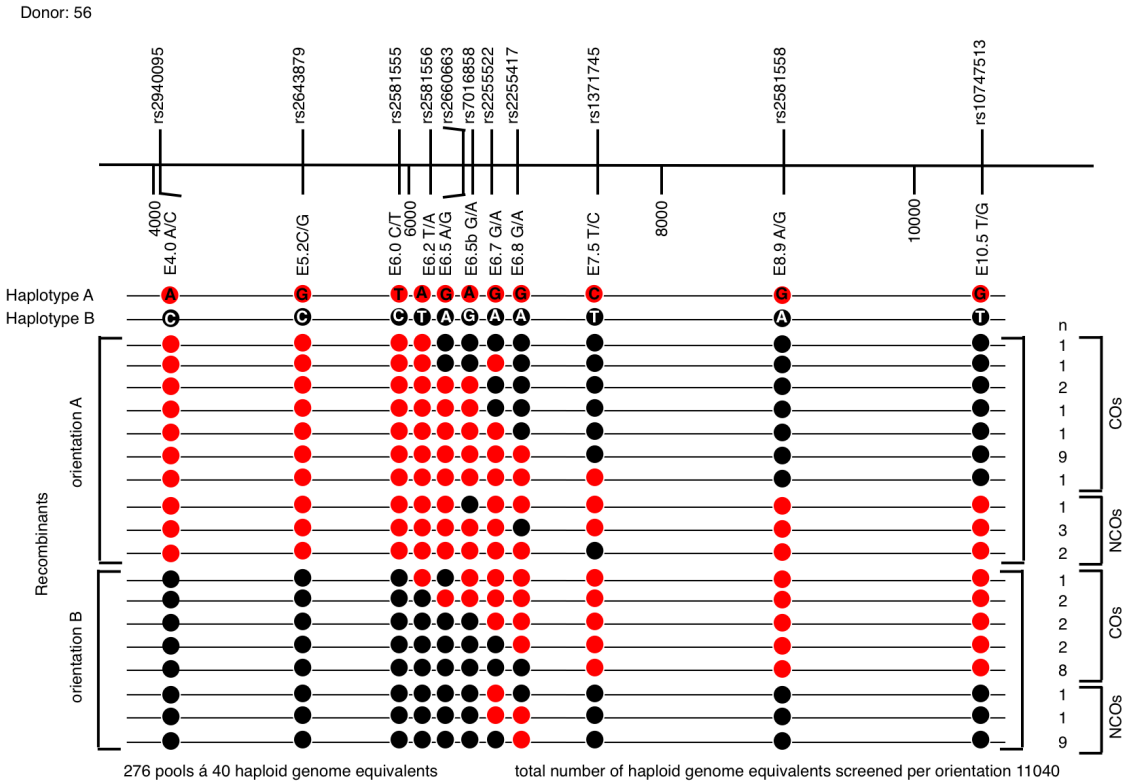
Appendix 7-IV ASOs used for hybridisation at hotspots K, T and 5A

Hotspot	rs number	location	Allele	Name of probe	Sequence 5'-->3'	Hybridisation temperature	Washing temperature	ASO quality
K	6999442	4944	T	K4.9T	AGTAATGTATTGAACCAT	48.5	54	**
			C	K4.9C	AGTAATGCATTGAACCAT	48.5	54	***
K	35234504	5078	A	K5.1A	TCTGGCTTATGGGTCACC	48.5	56	***
			G	K5.1G	TCTGGCTCATGGGTCACC	48.5	56	****
K	2034306	6155	C	K6.2C	AAGACATCTTGGACTTCT	48.5	54	***
			T	K6.2T	AAGACATTTTGGACTTCT	48.5	54	****
K	7005566	6377	C	K6.4C2	GTGCCAGACAGCTGGTAA	48.5	54	***
			T	K6.4T2	GTGCCAGGCACGTGGTAA	48.5	54	***
K	1374633	7477	C	K7.4C	GAGAGGGCACCCTGCCTA	48.5	54	***
			T	K7.4T	GAGAGGGTACCCTGCCTA	48.5	56	****
K	1374632	7501	C	K7.5C	AGGGAGGCCATCACAGAG	48.5	56	***
			G	K7.5G	AGGGAGGGCATCACAGAG	48.5	56	***
K	10088843	7845	T	K7.8aT	ATTCACCTGCCTTGGCCT	48.5	56	***
			C	K7.8aC	ATTCACCCGCCTTGGCCT	48.5	56	**
K	6471367	7891	T	K7.9gT	CGGGCACAGTGGCTCACA	48.5	56	**
			C	K7.9gC	CGGGCACGGTGGCTCACA	48.5	56	***
K	1550883	9913	G	K9.9G	TTCATGCGTCTCAGTCAC	48.5	56	**
			A	K9.9A	TTCATGCATCTCAGTCAC	48.5	56	**
K	1550882	10217	G	K10.2G	TGGGTCAGTTTCACAGAT	48.5	54	***
			A	K10.2A	TGGGTCAATTTCACAGAT	48.5	54	***
T	17013501	2502	C	T2.5C2	GACAAAAGAGGCCTGTGG	48.5	54	***
			T	T2.5T2	GACAAAAAAGGCCTGTGG	48.5	54	***
T	66595506	6703	C	T6.6T	CTCTCCTTCTTCTGTCTC	48.5	54	***
			T	T6.6G	CTCTCCTGCTTCTGTCTC	48.5	54	***
T	9830036	7125	C	T7.1T	AATTGAGCTTATAGCATG	48.5	54	**
			T	T7.1C	AATTGAGCTTACAGCATG	48.5	54	***
T	11714875	7132	C	T7.1aG	AGCATGGAGCCGTGTGCA	48.5	54	***
			T	T7.1aT	AGCATGTAGCCGTGTGCA	48.5	54	**
T	3856841	7405	C	T7.4aG2	TAACAGGCGGATTTATGT	48.5	54	***
			T	T7.4aA2	TAACAGGTGGATTTATGT	48.5	54	***
T	1171581	7556	C	T7.6aT	CAAATSTTAAGTCATGGT	48.5	54	***
			T	T7.6aG	CAAATSTGAAGTCATGGT	48.5	54	***
T	9854419	7933	C	T7.9A	CTTCTACAGTGGTTGACG	48.5	54	***
			T	T7.9G	CTTCTACGGTGGTTGACG	48.5	54	***
T	1827451	8331	C	T8.3G	TATATTACTCTTGTTGCA	48.5	54	**
			T	T8.3A	TATATTACTCTTATTGCA	48.5	54	***
T	1701357	8338	C	T8.3aA2	GTTATTCTGTGCAAYAAG	48.5	54	***
			T	T8.3aT2	GTTATTCAGTGCAAYAAG	48.5	54	***
T	35068617	8831	C	T8.8aT	CAGGGGTCAAGACTTCAA	48.5	54	***
			T	T8.8aC	CAGGGGTTAAGACTTCAA	48.5	54	***
T	12636785	8852	C	T8.8C	ATCTTTTCAGCGACACAA	48.5	56	***
			T	T8.8T	ATCTTTTTAGCGACACAA	48.5	54	**
T	12636800	8912	C	T8.9T	GAATACATGCATTCTCAC	49.5	56	**
			T	T8.9C	GAATACACGCATTCTCAC	48.5	54	***
T	12636837	9139	C	T9.1T	ATTGCCTTACATGGTTCC	48.5	54	***
			T	T9.1C	ATTGCCTCACATGGTTCC	48.5	54	***
T	12636283	9158	C	T9.2C	AGAAGGGCAAAATGAGTG	48.5	54	***
			T	T9.2A	AGAAGGGAAAAATGAGTG	48.5	54	***
T	66906407	9299	C	T9.3A	CAGTAGAATGGCCAGGTG	48.5	54	***
			T	T9.3G	CAGTAGAGTGGCCAGGTG	48.5	54	***
T	1502567	9490	C	T9.5C	TGCAATACGTGTTTACAT	48.5	54	***
			T	T9.5T	TGCAATATGTGTTTACAT	48.5	54	***
5A	4975767	5418	A	5A5.4A	TTTAAACATGCAGATTTC	48.5	54	***
			T	5A5.4T	TTTAAACTTGCAGATTTC	48.5	54	***
5A	6893465	5991	G	5A6.0A	TGTATTTAGACGGCTAAT	48.5	54	***
			A	5A6.0G	TGTATTTGGACGGCTAAT	48.5	54	***
5A	rs6872163	6498	A	5A6.5A	TTTTGACAAATGCACACC	48.5	54	***
			G	5A6.5G	TTTTGACGAATGCACACC	48.5	54	***
5A	13355978	7074	T	5A7.1T2	AGCACACACCTTTCTAGC	48.5	54	***
			C	5A7.1C2	AGCACACGCCTTTCTAGC	48.5	54	***
5A	no rs# assigned yet	7180	A	5A7.2A	AGGTAACAATTTCAGGGG	48.5	54	***
			G	5A7.2G	AGGTAACGATTTCAGGGG	48.5	54	***
5A	34992917	7590	plus	5A7.6a+	GTTACACATGTTTCCTTG	48.5	54	***
			minus	5A7.6a-	GTTACACGTTTCCTTGTG	48.5	54	***
5A	34829806	7792	plus	5A7.8+	AATAAACATAAGTTTTAC	48.5	54	***
			minus	5A7.8-	AATAAACAAGTTTTACAT	48.5	54	***
5A	2129470	7944	A	5A7.9A2	GAAGACGTGAAAAAGCTG	48.5	54	***
			C	5A7.9T2	GAAGACGGGAAAAAGCTG	48.5	54	***
5A	2129469	8043	A	5A8.0A	GTGGCCAATATCTCTATT	48.5	54	***
			G	5A8.0G	GTGGCCAGTATCTCTATT	48.5	54	***
5A	6865173	8440	G	5A8.4G	CCATGGCGTGGGCTAGGA	48.5	54	***
			A	5A8.4A	CCATGGCATGGGCTAGGA	48.5	54	***

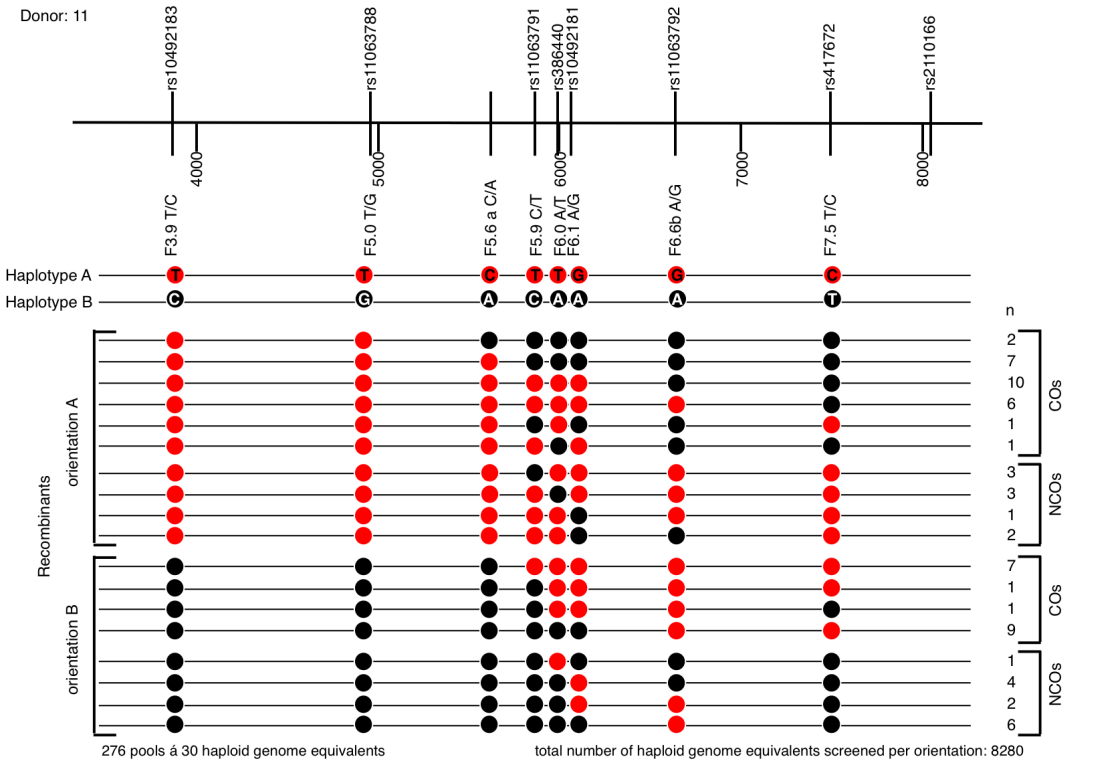
V. Detected recombinants

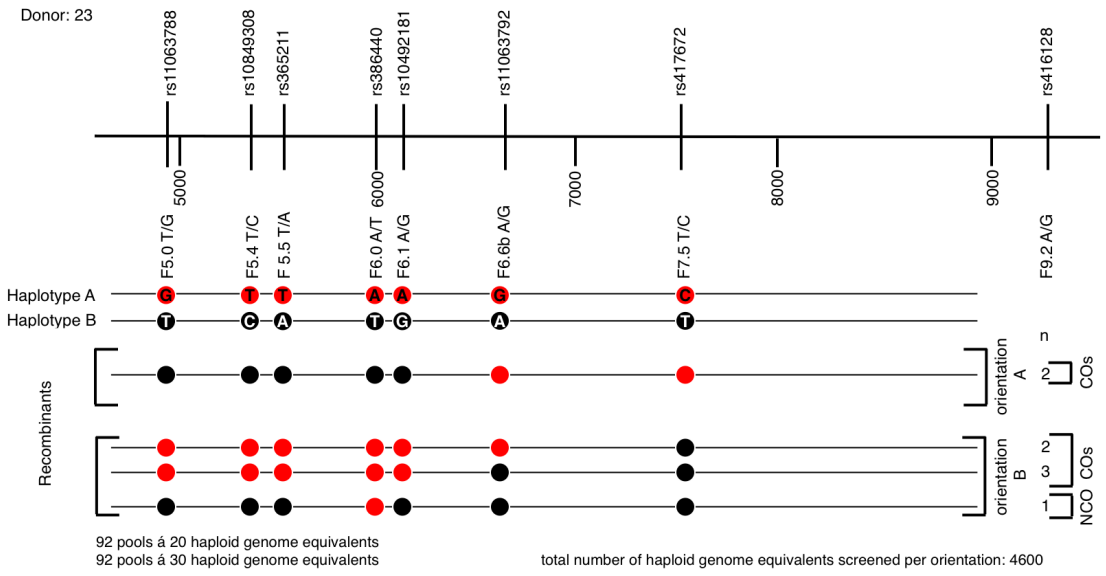
Appendix 7-V Superhotspot E recombinants:



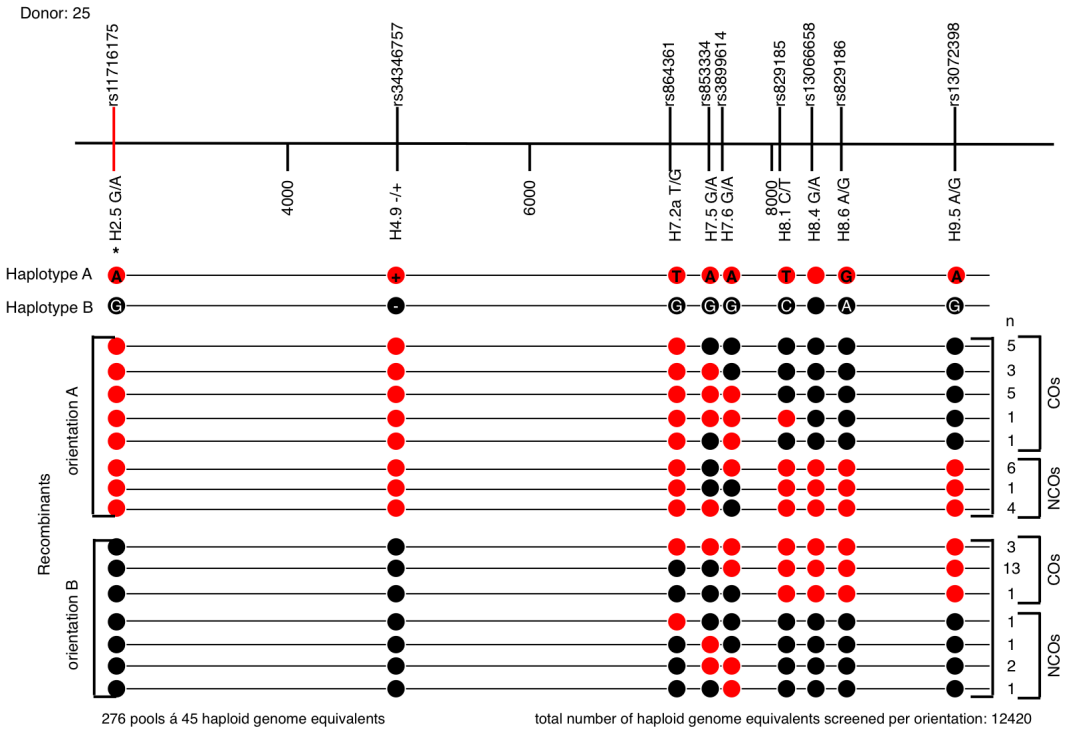


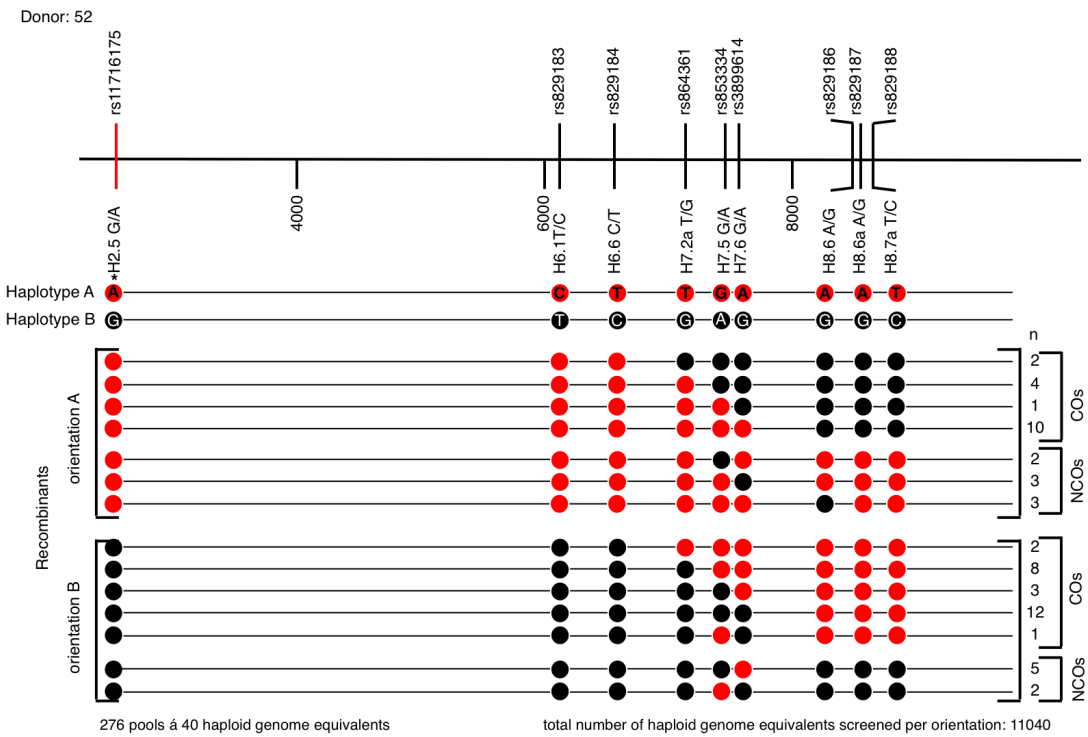
Appendix 7-VI Superhotspot F recombinants



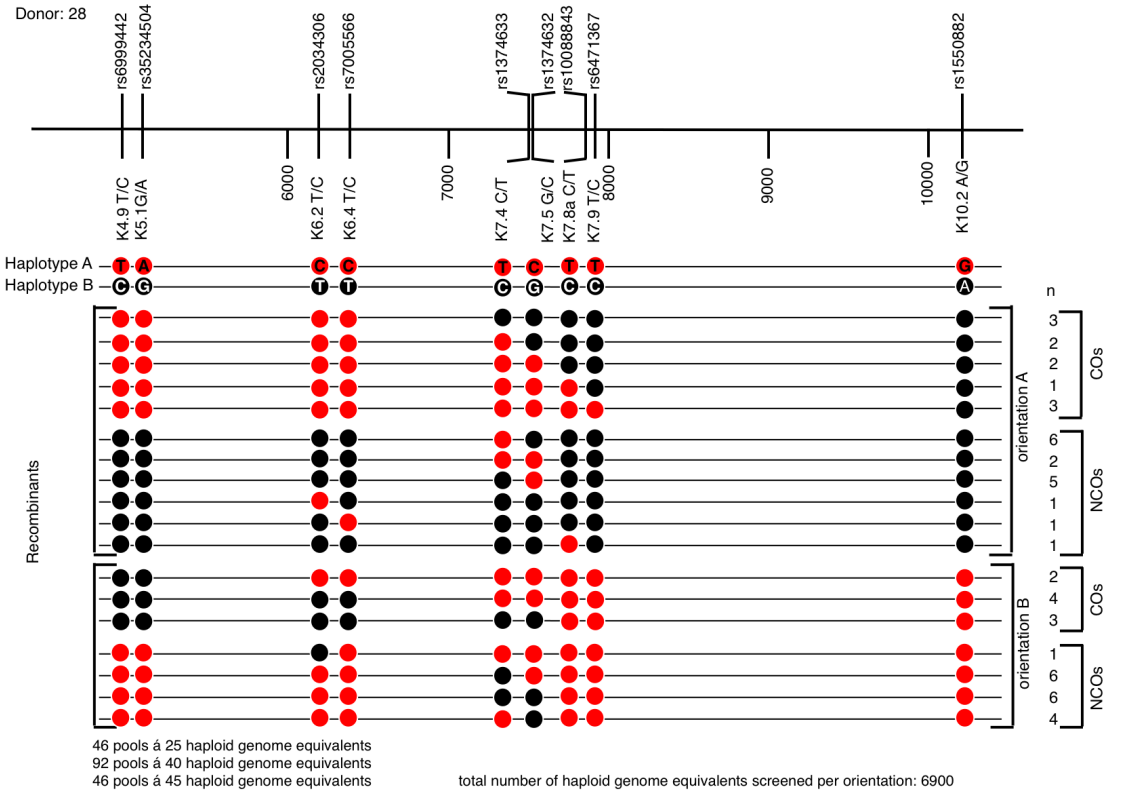


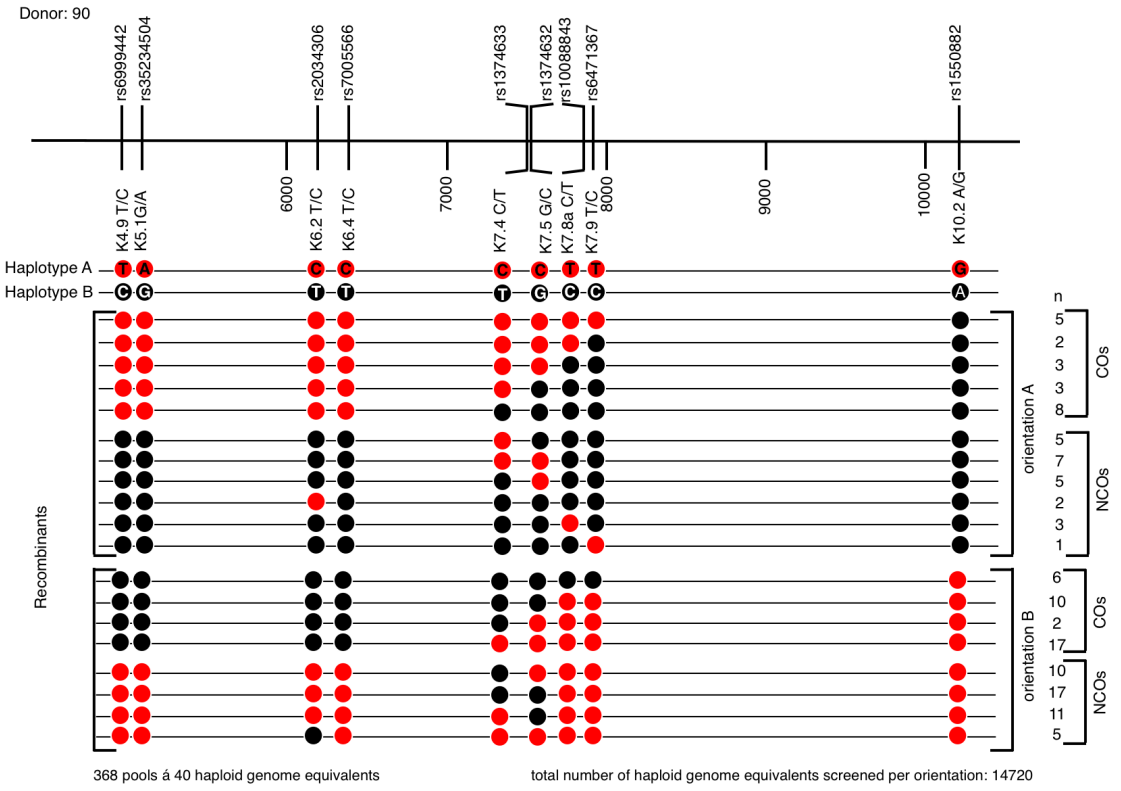
Appendix 7-VII Superhotspot H recombinants



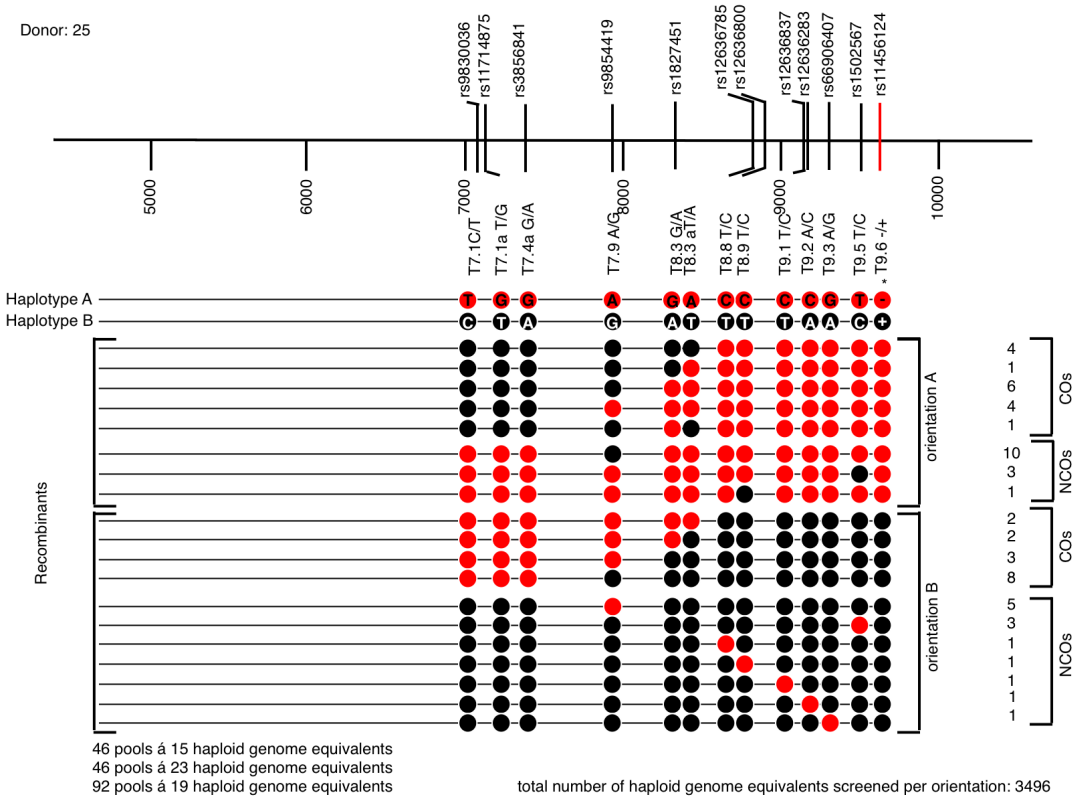


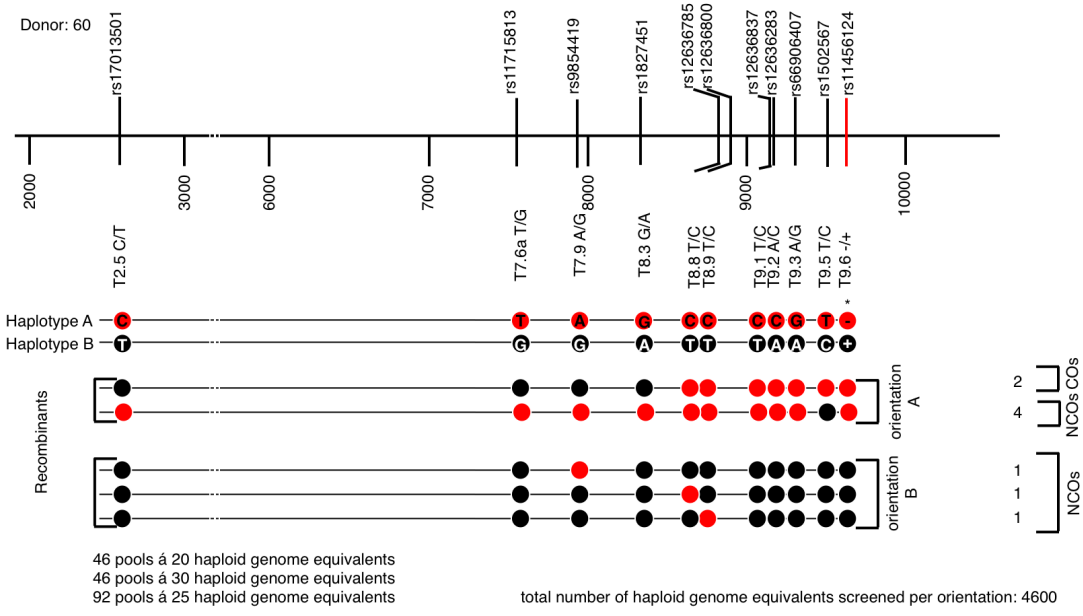
Appendix 7-VIII Superhotspot K recombinants





Appendix 7-IX Superhotspot T recombinants





VI. Transmission frequencies of alleles

Appendix 7-X Superhotspot K, transmission frequencies at marker K7.4

man	K7.4C/T phasing	CONs with K7.4C	CONs with K7.4T	CONs with K7.4C (%)	additional CONs with markers from K7.4C haplotype	additional CONs with markers from K7.4T haplotype	additional CONs with markers from K7.4C haplotype (%)	A+B cross overs with K7.4C	A+B cross overs with K7.4T	A+B cross overs with K7.4C (%)
d35	3	8	2	80.0	4	2	66.7	5	7	41.7
d28	1	13	9	59.1	8	8	50.0	6	14	30.0
d90	2	30	13	69.8	21	14	60.0	31	27	53.4
total		51	24	68.0	33	24	57.9	42	48	46.7

Data gathered by L. Odenthal-Hesse

Appendix 7-XI Superhotspot F, transmission frequencies at marker F6.1

man	F6.1G/A phasing	CONs with F6.1G	CONs with F6.1A	CONs with F6.1G (%)	additional CONs with markers from F6.1G haplotype	additional CONs with markers from F6.1A haplotype	additional CONs with markers from F6.1G haplotype (%)	A+B cross overs with F6.1G	A+B cross overs with F6.1A	A+B cross overs with F6.1G (%)
d28	–	13	0	100.0	4	0	100.0	74	94	44.0
*d11	1	6	3	66.7	8	8	50.0	26	18	59.1
d35	1	5	1	83.3	3	1	75.0	12	13	48.0
d77	1	4	1	80.0	3	3	50.0	32	33	49.2
d20	1	4	2	66.7	4	5	44.4	58	75	43.6
d51	2	10	7	58.8	7	3	70.0	77	60	56.2
d31	1	12	6	66.7	1	5	16.7	81	70	53.6
d2	2	6	5	54.5	4	6	40.0	94	62	60.3
d55	1	7	3	70.0	6	4	60.0	33	42	44.0
d87	1	3	1	75.0	0	0	-	42	32	56.8
total		70	29	70.7	40	35	53.3	529	499	51.5

data gathered by A.J. Jeffreys and I. L. Berg, *data gathered by LOH

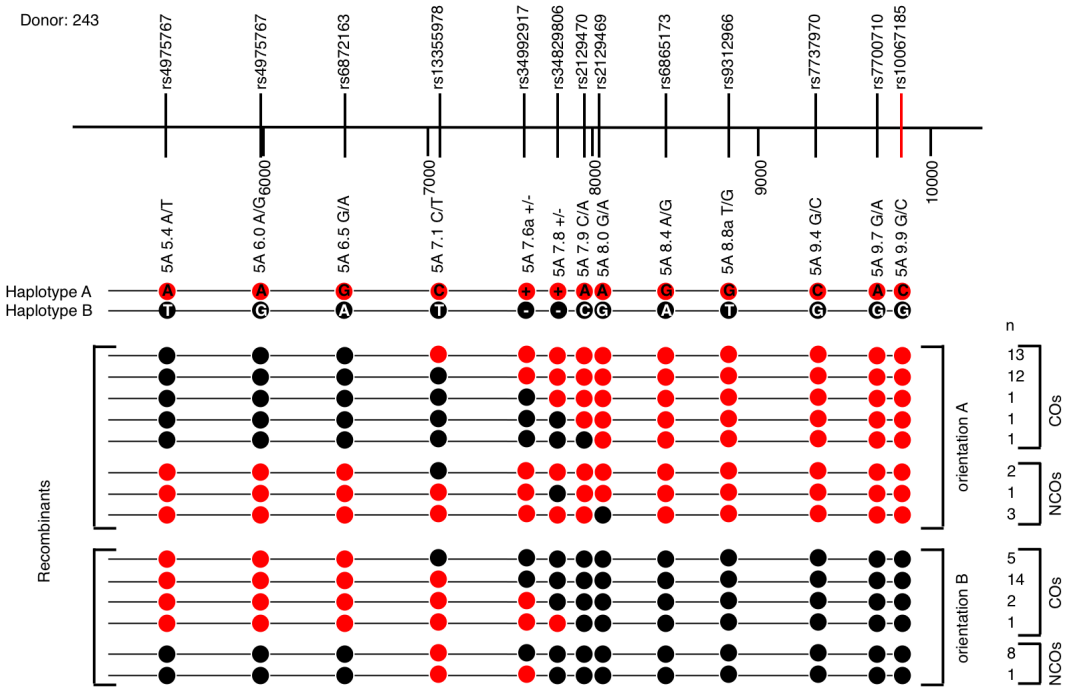
VII. PCR conditions in crossover assay at Superhotspot T

hotspot	Assay	amplification type	Primer combi-nation	cycling conditions	PCR machine
T	1	1°	T3.3FG & T 9.3RG	96°C, 1.5min; (96°C, 20s; 58°C, 30s; 65°C, 7 min)×25; 65°C, 5min; hold at 10°C	9700
		2°	T3.6FC & T9.1RC	96°C, 1.5min (96°C, 20s; 57°C, 30s; 65°C, 6.5 min)×35; 65°C, 5min; hold at 10°C	9700
	2	1°	T3.3FA & T9.3RG	96°C, 1.5min (96°C, 20s; 58°C, 30s; 65°C,7 min)×25; 65°C, 5min; hold at 10°C	9700
		2°	T3.6FT & T9.1RC	96°C, 1.5min (96°C, 20s; 58°C, 30s; 65°C, 6.5 min)×32; 65°C, 5min; hold at 10°C	9700

VIII. PCR conditions in half-crossover assay at Hotspot 5A

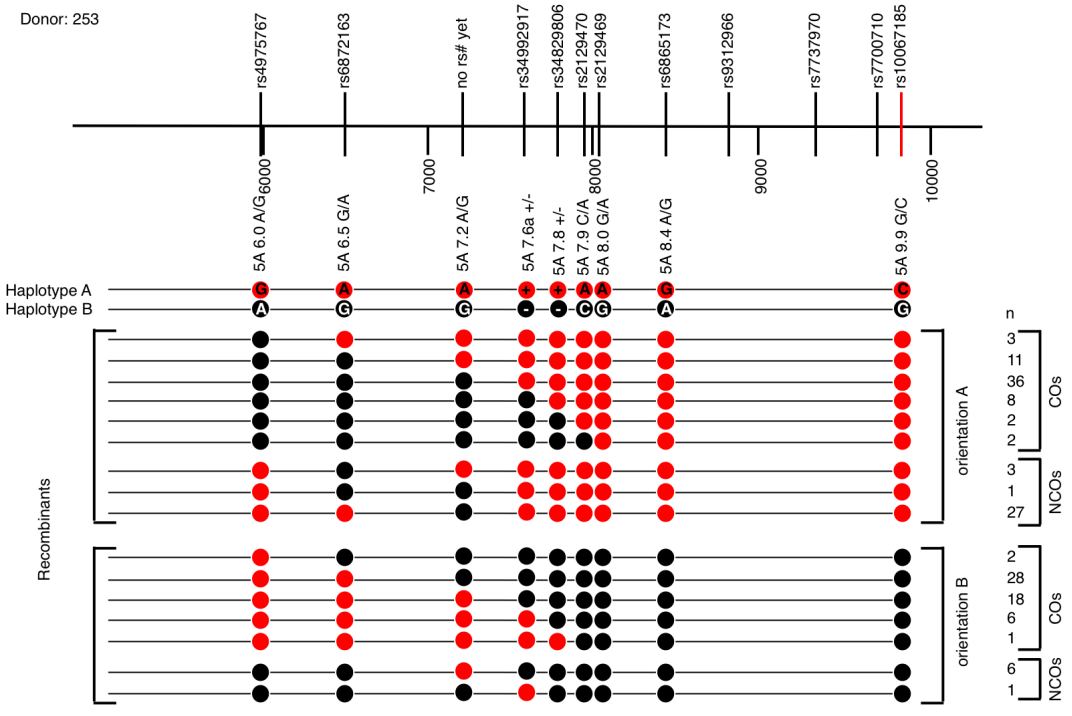
hotspot	orientation	ASPs located	amplification type	Primer combination	cycling conditions	PCR ma-chine
5A	A	-->	1°	5A3.6F& 5A9.9aRG	96°C, 1.5min (96°C,20s; 55°C, 30s; 65°C 7.5min)×25; 65°C 5 min	Veriti
	A	-->	2°	5A3.7F& 5A9.9RC	96°C, 1.5min (96°C,20s; 58°C, 30s; 65°C 7.5min)×32; 65°C 5 min; hold at 10°C	Veriti
	B	-->	1°	5A3.6F& 5A9.9aRC	96°C, 1.5min (96°C,20s; 55°C, 30s; 65°C 7.5min)×25; 65°C 5 min	Veriti
	B	-->	2°	5A3.7F& 5A9.9RG	96°C, 1.5min (96°C,20s; 55°C, 30s; 65°C 7.5min)×25; 65°C 5 min; hold at 10°C	Veriti

IX. Hotspot 5A recombinants



276 pools á 30 haploid genome equivalents

total number of haploid genome equivalents screened per orientation:



276 pools á 30 haploid genome equivalents
184 pools á 37.5 haploid genome equivalents

total number of haploid genome equivalents screened per orientation:

References

1000 Genomes Project Consortium, Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome variation from population-scale sequencing. In *Nature*, Vol 467, pp. 1061-1073.

Aguileta G, Bielawski JP, Yang Z. 2004. Gene conversion and functional divergence in the beta-globin gene family. In *J Mol Evol*, Vol 59, pp. 177-189.

Alani E, Padmore R, Kleckner N. 1990. Analysis of wild-type and rad50 mutants of yeast suggests an intimate relationship between meiotic chromosome synapsis and recombination. In *Cell*, Vol 61, pp. 419-436.

Alani E, Reenan RA, Kolodner RD. 1994. Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae*. In *Genetics*, Vol 137, pp. 19-39.

Allers T, Lichten M. 2001a. Differential timing and control of noncrossover and crossover recombination during meiosis. In *Cell*, Vol 106, pp. 47-57.

Allers T, Lichten M. 2001b. Intermediates of yeast meiotic recombination contain heteroduplex DNA. In *Molecular Cell*, Vol 8, pp. 225-231.

Ardlie K, Liu-Cordero SN, Eberle MA, Daly M, Barrett J, Winchester E, Lander ES, Kruglyak L. 2001. Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. In *Am J Hum Genet*, Vol 69, pp. 582-589.

Armour JA, Harris PC, Jeffreys AJ. 1993. Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. In *Hum Mol Genet*, Vol 2, pp. 1137-1145.

Arnheim N, Calabrese P, Tiemann-Boege I. 2007. Mammalian meiotic recombination hot spots. In *Annu Rev Genet*, Vol 41, pp. 369-399.

Bagshaw ATM, Pitt JPW, Gemmell NJ. 2008. High frequency of microsatellites in *S. cerevisiae* meiotic recombination hotspots. In *BMC Genomics*, Vol 9, p. 49.

Baltimore D. 1981. Gene conversion: some implications for immunoglobulin genes. In *Cell*, Vol 24, pp. 592-594.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. 2010. PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. In *Science*, Vol 327, pp. 836-840.

Baudat F, de Massy B. 2007. Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. In *Chromosome Res*, Vol 15, pp. 565-577.

Baudat F, Nicolas A. 1997. Clustering of meiotic double-strand breaks on yeast chromosome III. In *Proc Natl Acad Sci USA*, Vol 94, pp. 5213-5218.

Berg IL, Neumann R, Lam K-WG, Sarbajna S, Odenthal-Hesse L, May CA, Jeffreys AJ. 2010. PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. In *Nat Genet*, Vol 42, pp. 859–863.

Berg IL, Neumann R, Sarbajna S, Odenthal-Hesse L, Butler NJ, Jeffreys AJ. 2011. Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. In *Proc Natl Acad Sci USA*, p. 10.1073/pnas.1109531108.

Bishop DK, Zickler D. 2004. Early decision; meiotic crossover interference prior to stable strand exchange and synapsis. In *Cell*, Vol 117, pp. 9-15.

Borde V, Robine N, Lin W, Bonfils S, Géli V, Nicolas A. 2009. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. In *EMBO J*, Vol 28, pp. 99-111.

Börner GV, Kleckner N, Hunter N. 2004. Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. In *Cell*, Vol 117, pp. 29-45.

Borts RH, Haber JE. 1989. Length and distribution of meiotic gene conversion tracts and crossovers in *Saccharomyces cerevisiae*. In *Genetics*, Vol 123, pp. 69-80.

Boulton A, Myers RS, Redfield RJ. 1997. The hotspot conversion paradox and the evolution of meiotic recombination. In *Proc Natl Acad Sci USA*, Vol 94, pp. 8058-8063.

Buard J, Barthès P, Grey C, de Massy B. 2009. Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. In *EMBO J*, Vol 28, pp. 2616-2624.

Cao L, Alani E, Kleckner N. 1990. A pathway for generation and processing of double-strand breaks during meiotic recombination in S. cerevisiae. In *Cell*, Vol 61, pp. 1089-1101.

Carpenter AT. 1975. Electron microscopy of meiosis in *Drosophila melanogaster* females: II. The recombination nodule--a recombination-associated structure at pachytene? In *Proc Natl Acad Sci USA*, Vol 72, pp. 3186-3189.

Carpenter AT. 1987. Gene conversion, recombination nodules, and the initiation of meiotic synapsis. In *Bioessays*, Vol 6, pp. 232-236.

Carpenter AT. 1994. Chiasma function. In *Cell*, Vol 77, pp. 957-962.

Cervantes MD, Farah JA, Smith GR. 2000. Meiotic DNA breaks associated with recombination in S. pombe. In *Molecular Cell*, Vol 5, pp. 883-888.

Chen FC, Li WH. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. In *Am J Hum Genet*, Vol 68, pp. 444-456.

Chowdhury R, Bois PRJ, Feingold E, Sherman SL, Cheung VG. 2009. Genetic analysis of variation in human meiotic recombination. In *PLoS Genetics*, Vol 5, p. e1000648.

Cole F, Keeney S, Jasin M. 2010. Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis. In *Molecular Cell*, Vol 39, pp. 700-710.

Collier S, Sinnott PJ, Dyer PA, Price DA, Harris R, Strachan T. 1989. Pulsed field gel electrophoresis identifies a high degree of variability in the number of tandem 21-hydroxylase and complement C4 gene repeats in 21-hydroxylase deficiency haplotypes. In *EMBO J*, Vol 8, pp. 1393-1402.

Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. 2008. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. In *Science*, Vol 319, pp. 1395-1398.

Cottrell CE, Bir N, Varga E, Alvarez CE, Bouyain S, Zernzach R, Thrush DL, Evans J, Trimarchi M, Butter EM *et al.* 2011. Contactin 4 as an autism susceptibility locus. In *Autism research : official journal of the International Society for Autism Research*, Vol 4, pp. 189–199.

Cromie GA, Hyppa RW, Taylor AF, Zakharyevich K, Hunter N, Smith GR. 2006. Single Holliday junctions are intermediates of meiotic recombination. In *Cell*, Vol 127, pp. 1167-1178.

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

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

Darlow JM, Stott DI. 2006. Gene conversion in human rearranged immunoglobulin genes. In *Immunogenetics*, Vol 58, pp. 511-522.

de los Santos T, Hunter N, Lee C, Larkin B, Loidl J, Hollingsworth NM. 2003. The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. In *Genetics*, Vol 164, pp. 81-94.

Detloff P, White MA, Petes TD. 1992. Analysis of a gene conversion gradient at the HIS4 locus in *Saccharomyces cerevisiae*. In *Genetics*, Vol 132, pp. 113-123.

Dieffenbach CW, Dveksler GS. 1993. Setting up a PCR laboratory. In *PCR Methods Appl*, Vol 3, pp. S2-7.

Duret L, Arndt PF. 2008. The impact of recombination on nucleotide substitutions in the human genome. In *PLoS Genet*, Vol 4, p. e1000071.

Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. In *Annual review of genomics and human genetics*, Vol 10, pp. 285-311.

Fledel-Alon A, Leffler EM, Guan Y, Stephens M, Coop G, Przeworski M. 2011. Variation in human recombination rates and its genetic determinants. In *PLoS ONE*, Vol 6, p. e20321.

Fogel S, Hurst DD. 1963. Coincidence relations between gene conversion and mitotic recombination in Saccharomyces. In *Genetics*, Vol 48, pp. 321-328.

Frisse L, Hudson RR, Bartoszewicz A, Wall JD, Donfack J, Di Rienzo A. 2001. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. In *Am J Hum Genet*, Vol 69, pp. 831-843.

Fung JC, Rockmill B, Odell M, Roeder GS. 2004. Imposition of crossover interference through the nonrandom distribution of synapsis initiation complexes. In *Cell*, Vol 116, pp. 795-802.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, Defelice M, Lochner A, Faggart M *et al.* 2002. The structure of haplotype blocks in the human genome. In *Science*, Vol 296, pp. 2225-2229.

Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. In *Trends Genet*, Vol 23, pp. 273-277.

Galtier N, Duret L, Glémin S, Ranwez V. 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. In *Trends Genet*, Vol 25, pp. 1-5.

Game JC, Sitney KC, Cook VE, Mortimer RK. 1989. Use of a ring chromosome and pulsed-field gels to study interhomolog recombination, double-strand DNA breaks and sister-chromatid exchange in yeast. In *Genetics*, Vol 123, pp. 695-713.

Gay J, Myers S, McVean G. 2007. Estimating meiotic gene conversion rates from population genetic data. In *Genetics*, Vol 177, pp. 881-894.

Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. In *Proc Natl Acad Sci USA*, Vol 97, pp. 11383-11390.

Getun IV, Wu ZK, Khalil AM, Bois PRJ. 2010. Nucleosome occupancy landscape and dynamics at mouse recombination hotspots. In *EMBO Rep*, Vol 11, pp. 555-560.

Grey C, Barthès P, Chauveau-Le Friec G, Langa F, Baudat F, de Massy B. 2011. Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. In *Plos Biol*, Vol 9, p. e1001176.

Grey C, Baudat F, de Massy B. 2009. Genome-wide control of the distribution of meiotic recombination. In *Plos Biol*, Vol 7, p. e35.

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J. 1994. The 1993-94 Généthon human genetic linkage map. In *Nature Genetics*, Vol 7, pp. 246-339.

Haber JE, Ira G, Malkova A, Sugawara N. 2004. Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. In *Philos Trans R Soc Lond, B, Biol Sci*, Vol 359, pp. 79-86.

Hansford LM, Smith SA, Haber M, Norris MD, Cheung B, Marshall GM. 2003. Cloning and characterization of the human neural cell adhesion molecule, CNTN4 (alias BIG-2). In *Cytogenet Genome Res*, Vol 101, pp. 17-23.

Hassold T, Judis L, Chan ER, Schwartz S, Seftel A, Lynn A. 2004. Cytological studies of meiotic recombination in human males. In *Cytogenet Genome Res*, Vol 107, pp. 249-255.

Hayashi K, Yoshida K, Matsui Y. 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. In *Nature*, Vol 438, pp. 374-378.

Heng HH, Chamberlain JW, Shi XM, Spyropoulos B, Tsui LC, Moens PB. 1996. Regulation of meiotic chromatin loop size by chromosomal position. In *Proc Natl Acad Sci USA*, Vol 93, pp. 2795-2800.

Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, Chen GK, Wang K, Buxbaum SG, Akylbekova EL *et al.* 2011. The landscape of recombination in African Americans. In *Nature*, Vol 476, pp. 170-175.

Holliday R. 1964. A mechanism for gene conversion in fungi. In *Genet Res*, Vol 89, pp. 282-304.

Holloway JK, Booth J, Edelmann W, McGowan CH, Cohen PE. 2008. MUS81 generates a subset of MLH1-MLH3-independent crossovers in mammalian meiosis. In *PLoS Genetics*, Vol 4, p. e1000186.

Holloway K, Lawson VE, Jeffreys AJ. 2006. Allelic recombination and de novo deletions in sperm in the human beta-globin gene region. In *Hum Mol Genet*, Vol 15, pp. 1099-1111.

Hunter N, Kleckner N. 2001. The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. In *Cell*, Vol 106, pp. 59-70.

Hurles ME, Willey D, Matthews L, Hussain SS. 2004. Origins of chromosomal rearrangement hotspots in the human genome: evidence from the AZFa deletion hotspots. In *Genome Biol*, Vol 5, p. R55.

Irie S, Tsujimura A, Miyagawa Y, Ueda T, Matsuoka Y, Matsui Y, Okuyama A, Nishimune Y, Tanaka H. 2009. Single-nucleotide polymorphisms of the PRDM9 (MEISETZ) gene in patients with nonobstructive azoospermia. In *J Androl*, Vol 30, pp. 426-431.

Jeffreys A, Neumann R. 2009. The rise and fall of a human recombination hot spot. In *Nat Genet*, p. 10.1038/ng.1346.

Jeffreys AJ, Allen MJ, Armour JA, Collick A, Dubrova Y, Fretwell N, Guram T, Jobling M, May CA, Neil DL. 1995. Mutation processes at human minisatellites. In *Electrophoresis*, Vol 16, pp. 1577-1585.

Jeffreys AJ, Barber R, Bois P, Buard J, Dubrova YE, Grant G, Hollies CR, May CA, Neumann R, Panayi M *et al.* 1999. Human minisatellites, repeat DNA instability and meiotic recombination. In *Electrophoresis*, Vol 20, pp. 1665-1675.

Jeffreys AJ, Bois P, Buard J, Collick A, Dubrova Y, Hollies CR, May CA, Murray J, Neil DL, Neumann R *et al.* 1997. Spontaneous and induced minisatellite instability. In *Electrophoresis*, Vol 18, pp. 1501-1511.

Jeffreys AJ, Holloway JK, Kauppi L, May CA, Neumann R, Slingsby MT, Webb AJ. 2004. Meiotic recombination hot spots and human DNA diversity. In *Philos Trans R Soc Lond, B, Biol Sci*, Vol 359, pp. 141-152.

Jeffreys AJ, Kauppi L, Neumann R. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. In *Nat Genet*, Vol 29, pp. 217-222.

Jeffreys AJ, May CA. 2003. DNA enrichment by allele-specific hybridization (DEASH): a novel method for haplotyping and for detecting low-frequency base substitutional variants and recombinant DNA molecules. In *Genome Res*, Vol 13, pp. 2316-2324.

Jeffreys AJ, May CA. 2004. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. In *Nat Genet*, Vol 36, pp. 151-156.

Jeffreys AJ, Murray J, Neumann R. 1998a. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. In *Molecular Cell*, Vol 2, pp. 267-273.

Jeffreys AJ, Neil DL, Neumann R. 1998b. Repeat instability at human minisatellites arising from meiotic recombination. In *EMBO J*, Vol 17, pp. 4147-4157.

Jeffreys AJ, Neumann R. 1997. Somatic mutation processes at a human minisatellite. In *Hum Mol Genet*, Vol 6, pp. 129-132; 134-126.

Jeffreys AJ, Neumann R. 2002. Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. In *Nat Genet*, Vol 31, pp. 267-271.

Jeffreys AJ, Neumann R. 2005. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. In *Hum Mol Genet*, Vol 14, pp. 2277-2287.

Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P. 2005. Human recombination hot spots hidden in regions of strong marker association. In *Nat Genet*, Vol 37, pp. 601-606.

Jeffreys AJ, Neumann R, Wilson V. 1990. Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. In *Cell*, Vol 60, pp. 473-485.

Jeffreys AJ, Ritchie A, Neumann R. 2000. High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. In *Hum Mol Genet*, Vol 9, pp. 725-733.

Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JA. 1994. Complex gene conversion events in germline mutation at human minisatellites. In *Nat Genet*, Vol 6, pp. 136-145.

Jeffreys AJ, Wilson V, Neumann R, Keyte J. 1988. Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. In *Nucleic Acids Res*, Vol 16, pp. 10953-10971.

Kaback DB, Barber D, Mahon J, Lamb J, You J. 1999. Chromosome size-dependent control of meiotic reciprocal recombination in *Saccharomyces cerevisiae*: the role of crossover interference. In *Genetics*, Vol 152, pp. 1475-1486.

Katzman S, Capra JA, Haussler D, Pollard KS. 2011. Ongoing GC-biased evolution is widespread in the human genome and enriched near recombination hotspots. In *Genome Biol Evol*, p. 10.1093/gbe/evr1058.

Katzman S, Kern AD, Pollard KS, Salama SR, Haussler D. 2010. GC-biased evolution near human accelerated regions. In *PLoS Genetics*, Vol 6, p. e1000960.

Kauppi L, Jeffreys AJ, Keeney S. 2004. Where the crossovers are: recombination distributions in mammals. In *Nat Rev Genet*, Vol 5, pp. 413-424.

Kauppi L, May CA, Jeffreys AJ. 2009. Analysis of meiotic recombination products from human sperm. In *Methods Mol Biol*, Vol 557, pp. 323-355.

Kauppi L, Stumpf MPH, Jeffreys AJ. 2005. Localized breakdown in linkage disequilibrium does not always predict sperm crossover hot spots in the human MHC class II region. In *Genomics*, Vol 86, pp. 13-24.

Keeney S, Giroux CN, Kleckner N. 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. In *Cell*, Vol 88, pp. 375-384.

Keeney S, Kleckner N. 1995. Covalent protein-DNA complexes at the 5' strand termini of meiosis-specific double-strand breaks in yeast. In *Proc Natl Acad Sci USA*, Vol 92, pp. 11274-11278.

Kingman JF. 2000. Origins of the coalescent. 1974-1982. In *Genetics*, Vol 156, pp. 1461-1463.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G *et al.* 2002. A high-resolution recombination map of the human genome. In *Nat Genet*, Vol 31, pp. 241-247.

Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB, Jonasdottir A, Gylfason A, Kristinsson KT *et al.* 2010. Fine-scale recombination rate differences between sexes, populations and individuals. In *Nature*, Vol 467, pp. 1099-1103.

Kong A, Thorleifsson G, Stefansson H, Masson G, Helgason A, Gudbjartsson DF, Jonsdottir GM, Gudjonsson SA, Sverrisson S, Thorlacius T *et al.* 2008. Sequence variants in the RNF212 gene associate with genome-wide recombination rate. In *Science*, Vol 319, pp. 1398-1401.

Koppens PFJ, Hoogenboezem T, Degenhart HJ. 2002. Carriership of a defective tenascin-X gene in steroid 21-hydroxylase deficiency patients: TNXB -TNXA hybrids in apparent large-scale gene conversions. In *Hum Mol Genet*, Vol 11, pp. 2581-2590.

Kuma K, Hayashida H, Miyata T. 1988. Recent gene conversion between genes encoding human red and green visual pigments. In *Jpn J Genet*, Vol 63, pp. 367-371.

Lemon J, Bolker B, Oom S, Klein E, Rowlingson B, Wickham H, Tyagi A, Eterradossi O, Grothendieck G, Toews M *et al.* 2008. plotrix: Various plotting functions. R package version 2.5.

Li HH, Gyllensten UB, Cui XF, Saiki RK, Erlich HA, Arnheim N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. In *Nature*, Vol 335, pp. 414-417.

Lichten M, Borts RH, Haber JE. 1987. Meiotic gene conversion and crossing over between dispersed homologous sequences occurs frequently in *Saccharomyces cerevisiae*. In *Genetics*, Vol 115, pp. 233-246.

Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. In *Nature*, Vol 454, pp. 479-485.

Maniatis N, Collins A, Xu CF, McCarthy LC, Hewett DR, Tapper W, Ennis S, Ke X, Morton NE. 2002. The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. In *Proc Natl Acad Sci USA*, Vol 99, pp. 2228-2233.

Marsolier-Kergoat M-C. 2011. A simple model for the influence of meiotic conversion tracts on GC content. In *PLoS ONE*, Vol 6, p. e16109.

Martini E, Diaz RL, Hunter N, Keeney S. 2006. Crossover homeostasis in yeast meiosis. In *Cell*, Vol 126, pp. 285-295.

May CA, Jeffreys AJ, Armour JA. 1996. Mutation rate heterogeneity and the generation of allele diversity at the human minisatellite MS205 (D16S309). In *Hum Mol Genet*, Vol 5, pp. 1823-1833.

May CA, Tamaki K, Neumann R, Wilson G, Zagars G, Pollack A, Dubrova YE, Jeffreys AJ, Meistrich ML. 2000. Minisatellite mutation frequency in human sperm following radiotherapy. In *Mutat Res*, Vol 453, pp. 67-75.

McMahill MS, Sham CW, Bishop DK. 2007. Synthesis-dependent strand annealing in meiosis. In *Plos Biol*, Vol 5, p. e299.

McVean G, Myers S. 2010. PRDM9 marks the spot. In *Nat Genet*, Vol 42, pp. 821-822.

McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. In *Science*, Vol 304, pp. 581-584.

Mieczkowski PA, Dominska M, Buck MJ, Lieb JD, Petes TD. 2007. Loss of a histone deacetylase dramatically alters the genomic distribution of Spo11p-catalyzed DNA breaks in *Saccharomyces cerevisiae*. In *Proc Natl Acad Sci USA*, Vol 104, pp. 3955-3960.

Miura S, Shibata H, Furuya H, Ohyagi Y, Osoegawa M, Miyoshi Y, Matsunaga H, Shibata A, Matsumoto N, Iwaki A *et al.* 2006. The contactin 4 gene locus at 3p26 is a candidate gene of SCA16. In *Neurology*, Vol 67, pp. 1236-1241.

Miyamoto T, Koh E, Sakugawa N, Sato H, Hayashi H, Namiki M, Sengoku K. 2008. Two single nucleotide polymorphisms in PRDM9 (MEISETZ) gene may be a genetic risk factor for Japanese patients with azoospermia by meiotic arrest. In *J Assist Reprod Genet*, Vol 25, pp. 553-557.

Moore DP, Orr-Weaver TL. 1998. Chromosome segregation during meiosis: building an unambivalent bivalent. In *Curr Top Dev Biol*, Vol 37, pp. 263-299.

Morgan TH. 1911. Random Segregation versus Coupling in Mendelian Inheritance. In *Science*, Vol 34, p. 384.

Myers RS, Stahl FW. 1994. Chi and the RecBC D enzyme of Escherichia coli. In *Annu Rev Genet*, Vol 28, pp. 49-70.

Myers S. 2005. A Fine-Scale Map of Recombination Rates and Hotspots Across the Human Genome. In *Science*, Vol 310, pp. 321-324.

Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, Macfie TS, McVean G, Donnelly P. 2010. Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination. In *Science*, Vol 327, pp. 876-879.

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

Myers S, Spencer CCA, Auton A, Bottolo L, Freeman C, Donnelly P, McVean G. 2006. The distribution and causes of meiotic recombination in the human genome. In *Biochem Soc Trans*, Vol 34, pp. 526-530.

Nag DK, Scherthan H, Rockmill B, Bhargava J, Roeder GS. 1995. Heteroduplex DNA formation and homolog pairing in yeast meiotic mutants. In *Genetics*, Vol 141, pp. 75-86.

Nairz K, Klein F. 1997. mre11S--a yeast mutation that blocks double-strand-break processing and permits nonhomologous synapsis in meiosis. In *Genes Dev*, Vol 11, pp. 2272-2290.

Nassif N, Penney J, Pal S, Engels WR, Gloor GB. 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. In *Molecular and Cellular Biology*, Vol 14, pp. 1613-1625.

Neumann R, Jeffreys AJ. 2006. Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation. In *Hum Mol Genet*, Vol 15, pp. 1401-1411.

Neumann R, Lawson VE, Jeffreys AJ. 2010. Dynamics and processes of copy number instability in human gamma-globin genes. In *Proc Natl Acad Sci USA*, Vol 107, pp. 8304-8309.

Nicolas A, Petes TD. 1994. Polarity of meiotic gene conversion in fungi: contrasting views. In *Experientia*, Vol 50, pp. 242-252.

Novak JE, Ross-Macdonald PB, Roeder GS. 2001. The budding yeast Msh4 protein functions in chromosome synapsis and the regulation of crossover distribution. In *Genetics*, Vol 158, pp. 1013-1025.

Ogasawara K, Yabe R, Uchikawa M, Nakata K, Watanabe J, Takahashi Y, Tokunaga K. 2001. Recombination and gene conversion-like events may contribute to ABO gene diversity causing various phenotypes. In *Immunogenetics*, Vol 53, pp. 190-199.

Ogawa T, Yu X, Shinohara A, Egelman EH. 1993. Similarity of the yeast RAD51 filament to the bacterial RecA filament. In *Science*, Vol 259, pp. 1896-1899.

Pâques F, Haber JE. 1999. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. In *Microbiol Mol Biol Rev*, Vol 63, pp. 349-404.

Parvanov ED, Ng SHS, Petkov PM, Paigen K. 2009. Trans-regulation of mouse meiotic recombination hotspots by Rcr1. In *Plos Biol*, Vol 7, p. e36.

Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots. In *Science*, Vol 327, p. 835.

Persikov AV, Osada R, Singh M. 2009. Predicting DNA recognition by Cys2His2 zinc finger proteins. In *Bioinformatics*, Vol 25, pp. 22-29.

Petes TD. 2001. Meiotic recombination hot spots and cold spots. In *Nat Rev Genet*, Vol 2, pp. 360-369.

Petronczki M, Siomos MF, Nasmyth K. 2003. Un ménage à quatre: the molecular biology of chromosome segregation in meiosis. In *Cell*, Vol 112, pp. 423-440.

Pineda-Krch M, Redfield RJ. 2005. Persistence and loss of meiotic recombination hotspots. In *Genetics*, Vol 169, pp. 2319-2333.

Ponting CP. 2011. What are the genomic drivers of the rapid evolution of PRDM9? In *Trends Genet*, Vol 27, pp. 165-171.

Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Pääbo S. 2005. Fine-scale recombination patterns differ between chimpanzees and humans. In *Nat Genet*, Vol 37, pp. 429-434.

Ptak SE, Voelpel K, Przeworski M. 2004. Insights into recombination from patterns of linkage disequilibrium in humans. In *Genetics*, Vol 167, pp. 387-397.

R Development Core Team. 2008. R: A language and environment for statistical computing. In *R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0*.

Ratnakumar A, Mousset S, Glémin S, Berglund J, Galtier N, Duret L, Webster MT. 2010. Detecting positive selection within genomes: the problem of biased gene conversion. In *Philos Trans R Soc Lond, B, Biol Sci*, Vol 365, pp. 2571-2580.

Robinson J, Waller MJ, Fail SC, McWilliam H, Lopez R, Parham P, Marsh SGE. 2009. The IMGT/HLA database. In *Nucleic Acids Res*, Vol 37, pp. D1013-1017.

Rosenberg SM, Hastings PJ. 1991. The split-end model for homologous recombination at double-strand breaks and at Chi. In *Biochimie*, Vol 73, pp. 385-397.

Ross-Macdonald P, Roeder GS. 1994. Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. In *Cell*, Vol 79, pp. 1069-1080.

Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. In *Nature*, Vol 423, pp. 873-876.

Sambrook J, Russel D. 2000. Molecular cloning: a laboratory manual. In *CSH Press*, Vol 3rd.

Sarbajna S, Denniff M, Jeffreys AJ, Neumann R, Artigas MS, Veselis A, May CA. 2012. A major recombination hotspot in the XqYq pseudoautosomal region gives new insight into processing of human gene conversion events. In *Hum Mol Genet*, Vol 21, pp. 2029–2038.

Sarkar D. 2010. lattice: Lattice Graphics. R package version 0.17-15. In *The R-Project*, Vol 0.18-8.

Schwacha A, Kleckner N. 1994. Identification of joint molecules that form frequently between homologs but rarely between sister chromatids during yeast meiosis. In *Cell*, Vol 76, pp. 51-63.

Schwacha A, Kleckner N. 1995. Identification of double Holliday junctions as intermediates in meiotic recombination. In *Cell*, Vol 83, pp. 783-791.

Schwacha A, Kleckner N. 1997. Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. In *Cell*, Vol 90, pp. 1123-1135.

Shifman S, Bell JT, Copley RR, Taylor MS, Williams RW, Mott R, Flint J. 2006. A high-resolution single nucleotide polymorphism genetic map of the mouse genome. In *Plos Biol*, Vol 4, p. e395.

Shiroishi T, Sagai T, Hanzawa N, Gotoh H, Moriwaki K. 1991. Genetic control of sex-dependent meiotic recombination in the major histocompatibility complex of the mouse. In *EMBO J*, Vol 10, pp. 681-686.

Sigurdsson MI, Smith AV, Bjornsson HT, Jonsson JJ. 2009. HapMap methylation-associated SNPs, markers of germline DNA methylation, positively correlate with regional levels of human meiotic recombination. In *Genome Res*, Vol 19, pp. 581-589.

Smith GR, Amundsen SK, Chaudhury AM, Cheng KC, Ponticelli AS, Roberts CM, Schultz DW, Taylor AF. 1984. Roles of RecBC enzyme and chi sites in homologous recombination. In *Cold Spring Harb Symp Quant Biol*, Vol 49, pp. 485-495.

Stahl FW, Foss HM. 2010. A Two-pathway Analysis of Meiotic Crossing Over and Gene Conversion in *Saccharomyces cerevisiae*. In *Genetics*, Vol 186, pp. 515–536.

Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG *et al.* 2005. A common inversion under selection in Europeans. In *Nat Genet*, Vol 37, pp. 129-137.

Stöhr H, Heisig JB, Benz PM, Schöberl S, Milenkovic VM, Strauss O, Aartsen WM, Wijnholds J, Weber BHF, Schulz HL. 2009. TMEM16B, a novel protein with calcium-dependent chloride channel activity, associates with a presynaptic protein complex in photoreceptor terminals. In *J Neurosci*, Vol 29, pp. 6809-6818.

Storlazzi A, Xu L, Cao L, Kleckner N. 1995. Crossover and noncrossover recombination during meiosis: timing and pathway relationships. In *Proc Natl Acad Sci USA*, Vol 92, pp. 8512-8516.

Sun H, Treco D, Schultes NP, Szostak JW. 1989. Double-strand breaks at an initiation site for meiotic gene conversion. In *Nature*, Vol 338, pp. 87-90.

Sun H, Treco D, Szostak JW. 1991. Extensive 3'-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the ARG4 recombination initiation site. In *Cell*, Vol 64, pp. 1155-1161.

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. 1983. The double-strand-break repair model for recombination. In *Cell*, Vol 33, pp. 25-35.

The International HapMap Consortium: Altshuler D Brooks LD Chakravarti A Collins FS Daly MJ Donnelly P Sekine A Kent A Smith AV *et al.* 2005. A haplotype map of the human genome. In *Nature*, Vol 437, pp. 1299-1320.

The International HapMap Consortium: Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE *et al.* 2010. Integrating common and rare genetic variation in diverse human populations. In *Nature*, Vol 467, pp. 52-58.

The International HapMap Consortium: Frazer KA Ballinger DG Cox DR Hinds DA Stuve LL Gibbs RA Belmont JW Boudreau A Hardenbol P *et al.* 2007. A second generation human haplotype map of over 3.1 million SNPs. In *Nature*, Vol 449, pp. 851-861.

Tusié-Luna MT, White PC. 1995. Gene conversions and unequal crossovers between CYP21 (steroid 21-hydroxylase gene) and CYP21P involve different mechanisms. In *Proc Natl Acad Sci USA*, Vol 92, pp. 10796-10800.

Ubeda F, Wilkins JF. 2011. The Red Queen theory of recombination hotspots. In *Journal of evolutionary biology*, Vol 24, pp. 541–553.

Verzani J. 2008. Using R in Introductory Statistics Courses with the pmg Graphical User Interface. In *Journal of Statistics Education*, Vol 16.

von Salomé J, Kukkonen JP. 2008. Sequence features of HLA-DRB1 locus define putative basis for gene conversion and point mutations. In *BMC Genomics*, Vol 9, p. 228.

von Wettstein D, Rasmussen SW, Holm PB. 1984. The synaptonemal complex in genetic segregation. In *Annu Rev Genet*, Vol 18, pp. 331-413.

Webb AJ, Berg IL, Jeffreys A. 2008. Sperm cross-over activity in regions of the human genome showing extreme breakdown of marker association. In *Proc Natl Acad Sci USA*, Vol 105, pp. 10471-10476.

White PC, Tusie-Luna MT, New MI, Speiser PW. 1994. Mutations in steroid 21-hydroxylase (CYP21). In *Hum Mutat*, Vol 3, pp. 373-378.

Winckler W. 2005. Comparison of Fine-Scale Recombination Rates in Humans and Chimpanzees. In *Science*, Vol 308, pp. 107-111.

Wu TC, Lichten M. 1994. Meiosis-induced double-strand break sites determined by yeast chromatin structure. In *Science*, Vol 263, pp. 515-518.

Wu TC, Lichten M. 1995. Factors that affect the location and frequency of meiosis-induced double-strand breaks in *Saccharomyces cerevisiae*. In *Genetics*, Vol 140, pp. 55-66.

Yamada T, Mizuno K-I, Hirota K, Kon N, Wahls WP, Hartsuiker E, Murofushi H, Shibata T, Ohta K. 2004. Roles of histone acetylation and chromatin remodeling factor in a meiotic recombination hotspot. In *EMBO J*, Vol 23, pp. 1792-1803.

Yamashita K, Shinohara M, Shinohara A. 2004. Rad6-Bre1-mediated histone H2B ubiquitylation modulates the formation of double-strand breaks during meiosis. In *Proc Natl Acad Sci USA*, Vol 101, pp. 11380-11385.

Yoshihara Y, Kawasaki M, Tamada A, Nagata S, Kagamiyama H, Mori K. 1995. Overlapping and differential expression of BIG-2, BIG-1, TAG-1, and F3: four members of an axon-associated cell adhesion molecule subgroup of the immunoglobulin superfamily. In *J Neurobiol*, Vol 28, pp. 51-69.

Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW *et al.* 2001. Comparison of human genetic and sequence-based physical maps. In *Nature*, Vol 409, pp. 951-953.

Zangenberg G, Huang MM, Arnheim N, Erlich H. 1995. New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. In *Nat Genet*, Vol 10, pp. 407-414.

Zhang W, Collins A, Maniatis N, Tapper W, Morton NE. 2002. Properties of linkage disequilibrium (LD) maps. In *Proc Natl Acad Sci USA*, Vol 99, pp. 17004-17007.

Zhao Z, Hewett-Emmett D, Li WH. 1998. Frequent gene conversion between human red and green opsin genes. In *J Mol Evol*, Vol 46, pp. 494-496.

Zheng J, Khil PP, Camerini-Otero RD, Przytycka TM. 2010. Detecting sequence polymorphisms associated with meiotic recombination hotspots in the human genome. In *Genome Biol*, Vol 11, p. R103.

Zickler D, Kleckner N. 1999. Meiotic chromosomes: integrating structure and function. In *Annu Rev Genet*, Vol 33, pp. 603-754.