Control of Meiotic Recombination At A Human Crossover Hotspot

Mahmut Cerkez Ergoren
University of Leicester

Thesis submitted for the degree of Doctor of Philosophy

May 2013

Control of Meiotic Recombination At A Human Crossover Hotspot

ACKNOWLEDGEMENTS

First and foremost I would like to thank Alec for accepting me in his group, for his great help and helpful guidance, also importantly, for being very patient with me.

There are many people in the lab and the department, past and present, who were very helpful. To Rita, thank you for your great technical help and great support during my PhD; Celia, a special thanks for your priceless help, you are one in a million; Ingrid, for all your help, advice and surprise gifts, thank you; Nancy thank you very much for your incredible constant support and it was very nice dancing with you, love you; Raheleh, thanks for being a good and supportive friend in the lab; Lenny, thanks for helping me to adjust to the department; Rob, without you this thesis would not have been completed, thanks man; Pille thanks for the great statistical help during the Jubilee weekend.

To my amazing Leicester crew for making this city tolerable and enjoyable, you guys are very special to me. Thanks for being here and for your support; Tutku (ahbabım) THANK YOU; "my beeatches", Danielle and Angelica, without you Leicester and my life would not be so great and important, love you; Chiara, a perfect person always in the right place, thank you very much, ti amo; Housemate Aline and George, you guys are very important to me, miss you both.

To my specials, Irmak, Gunseli, Bahar and Mustafa; thank you for the last twelve years.

To my friends around the world, thank you for your support and being a call away. Andrea, "study bello!", "did you study today?", the first three chapters of this thesis were written with your great support, grazie mille bello, MK&DB; Davita and Keith, my tele-therapists, thanks for always being on the other side of the phone; Baldo, my little brother, ti voglio bene; Ryan, my special personal supporter, thanks for being there; Yeliz, how can I forget your great help, kuzen? And the others: Nasser (I do not know how to thanks, hayati!), Nick (Gaga), Enrico, Melek, Berya, Edge Gokhan, Argyris, Pierpaolo and Shanow.

Finally, and most importantly, to my family.

Annem ve Babam, bana inandı□ınız ve güvendi□iniz icin çok te□ekkür ederim, sizi çok seviyorum; Ahmet, sabrın icin çok te□ekkürler karde□im, Teyzem, canım, deste□in ve yardımların olmasa heralde bügün burada olamazdım. Son olarak, dede, bu tezi benimle her zaman gururlandı□ın ve övgüyle bahsetti□in icin, bir te□ekkür olarak sana ithaf ediyorum.

This thesis is dedicated to my grandfather Mahmut Cerkez…

ABBREVIATIONS

AE	axial element
ASO	allele-specific oligonucleotide
ASP	allele-specific primer
ASW	African ancestry in the south- western USA
CEPH	Centre d'Etude du Polymorphisme Humaine
CEU	the Centre d'Etude du Polymorphisme Humain in Utah
CHB	the Han Chinese of Beijing
CHD	Chinese in metropolitan Denver, Colorado, USA
CO	crossovers
dbSNP	database SNP
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DSB	double strand break
DSBR	double strand break repair
EN	early (recombination) nodule
GIH	Gujarati Indians in Houston, Texas, USA
JPT	the Japanese of Tokyo
LD	linkage disequilibrium
LDU	linkage disequilibrium unit
LE	lateral element
LINE	long interspersed nuclear element
LN	late (recombination) nodule
LR	likelihood ratio
LWK	Luhya in Webuye, Kenya
Mb, kb, bp	Mega-, kilo-, base pair
MDA	multiple displacement amplification
MHC	major histocompatibility complex
MKK	Maasai in Kinyawa, Kenya
MMR	mismatch repair
MXL	Mexican ancestry in Los Angeles, California, USA
NCO	no gene conversions
PAR	pseudoautosomal region
PCR	polymerase chain reaction
PRDM9	PR domain–containing 9, is a meiosis-specific histone
H3	methyltransferase
RN	recombination nodule
RF	recombination frequency
SEI	strand exchange intermediates
SINE	short interspersed nuclear element
SC	synaptonemal complex
SDSA	synthesis dependent strand annealing
SNP	single nucleotide polymorphism
TSI	Tuscans in Italy
WGA	whole genome amplification
YRI	Yoruba people of the Ibadan region in Nigeria
ZnF	zinc finger

Control of Meiotic Recombination At A Human Crossover Hotspot

Mahmut Cerkez Ergoren

ABSTRACT

Meiotic recombination plays a key role in reshuffling haplotypes in human populations, thus profoundly affecting evolution. However, our understanding of recombination dynamics is largely limited to descriptions of variation in populations and families. Higher resolution analysis (0.0001 cM or less) of *de novo* recombination events in human sperm DNA has revealed clustering into very narrow hotspots (1-2 kb) that generally coincide with abrupt breakdown of linkage disequilibrium (LD).

Myers *et al.* (2008) used population genetics approaches to survey HapMap Phase II genotype data across the whole genome to infer sites of historical recombination. Around 40% of ~30,000 hotspots so identified contain a 13-bp motif CCNCCNTNNCCNC which appears to be involved in hotspot specification and which may also drive modes of genome instability such as disease-causing genomic rearrangements, minisatellite mutation and common mitochondrial deletions. In addition, recent studies identified that the *trans*-regulator PRDM9 of meiotic recombination hotspots in humans and mice, contains a zinc finger array that can recognize the 13-bp sequence motif associated with hotspots, with binding to this motif possibly triggering hotspot activity via chromatin re-modelling in humans. Berg *et al*., (2010) reported that PRDM9 is a major global regulator of hotspots in humans, even hotspots lacking the sequence motif, and influences aspects of genome instability.

To examine the effect of the 13-bp motif on crossover frequencies and distribution, four candidate putative hotspots located on different chromosomes and identified from HapMap data were studied. These hotspots had the motif at their centre, and a SNP that disrupts the motif. LDU analysis confirmed the locations of the putative hotspots to a 1-2-kb interval. However, only the first candidate LD hotspot, 'Hotspot DA' could be analysed in detailed high-resolution analyses for understanding the influences of *cis*-regulation on hotspot activity. The other candidate LD hotspots were eliminated because of their disrupting SNP locations on the hotspot or because of a lack of suitable donors in our sperm donor panel.

Comparing the rates and distributions of sperm crossover events between donors heterozygous for the disrupting SNP showed that there is a huge asymmetry between the two alleles with the derived and motif-disrupting allele suppressing hotspot activity. This direct study to understand the effect of the disrupting allele on crossover initiation revealed that Hotspot DA is the first to show very strong direct *cis*-regulation for hotspot activity, and despite being influenced by the *trans*-factor PRDM9, it is not the major regulator for this hotspot.

CHAPTER 1: INTRODUCTION

Genetic variation between individuals is the product of two main processes: mutation and meiotic recombination. While recombination and independent assortment of chromosomes reshuffle DNA to create novel allelic combinations of the genetic material, it is mutation that drives novel variants into the genome. Sometimes these processes have deleterious effects, for instance by causing inherited disease. The other fundamental role of meiotic recombination is to ensure the proper segregation of homologous chromosomes during meiotic division. This chapter will introduce the basic features of meiosis and meiotic recombination in lower eukaryotes such as yeast, as well as in mice and humans, and discuss the current understanding of meiotic recombination hotspots in these three organisms.

1.1 OVERVIEW OF MEIOSIS

Meiosis consists of two sub-stages: meiosis I and meiosis II. In meiosis I there is pairing of homologous chromosomes followed by their subsequent segregation away from each other, and in meiosis II there is a mitosis-like division. Each sub-stage is divided further into prophase, metaphase, anaphase and telophase. The first stage of prophase of meiosis I is called leptotene, in which individual chromosomes begin to condense from long strands within the nucleus. However, the two sister chromatids are still so tightly bound that they are indistinguishable from one another. The zygotene stage occurs as the homologous chromosomes line up with each other and at this point the combined homologous chromosomes are said to be bivalent. They may also be referred to as a tetrad, in reference to the four sister chromatids. The two pairs of chromatids become “zipped” together, via the synaptonemal complex, in a process known as synapsis. The pachytene stage heralds crossing-over where non-sister chromatids of homologous chromosomes randomly exchange segments of genetic information over regions of homology. Exchange takes place at sites of recombination events, where the exchange of information between the non-sister chromatids can be either reciprocal or non-reciprocal.

During the diplotene stage, the synaptonemal complex degrades and homologous chromosomes begin to separate from one another. The chromosomes themselves begin to uncoil, allowing some transcription of DNA. However, the homologous chromosomes of each bivalent remain tightly bound at chiasmata (the regions where crossing-over has occurred). Chromosomes condense further during the diakinesis stage – this is the first point in meiosis where the four parts of the tetrads are actually visible under a light microscope. Sites of crossing-over entangle together and effectively overlap to make chiasmata clearly visible. Apart from this observation, the remainder of diakinesis closely resembles prometaphase of meiosis I: the nucleoli disappear, the nuclear membrane disintegrates into vesicles, and the meiotic spindle begins to form.

Following prophase I, homologous pairs enter metaphase I where the chromosomes line up along the centre of the cell and are independently and randomly assorted. In anaphase I, the cell elongates in preparation for division down the middle, and homologous chromosomes closely associated in synapsis exchange segments of DNA by crossing-over. The first division effectively ends when the centromeres arrive at the poles. Each daughter cell now has half the number of chromosomes, but each chromosome consists of a pair of chromatids. In prophase II, the nucleoli and the nuclear envelope begin to disappear again and chromatids shorten and thicken. Centrioles move to the polar regions and are arranged by spindle fibres. In metaphase II, the centromeres contain three kinetochores, organising fibres from the centrosomes on each side. This is followed by anaphase II, where the centromeres are cleaved, allowing the kinetochores to pull the sister chromatids apart. The sister chromatids by convention are now called sister chromosomes, and they are pulled towards opposing poles. The process ends with telophase II, which is similar to telophase I, marked by uncoiling and lengthening of the chromosomes together with the disappearance of the microtubules. Nuclear envelopes reform, and cleavage and cell wall formation eventually produces a total of four daughter cells, each with a haploid set of chromosomes. Meiosis is now complete (Figure 1-1).

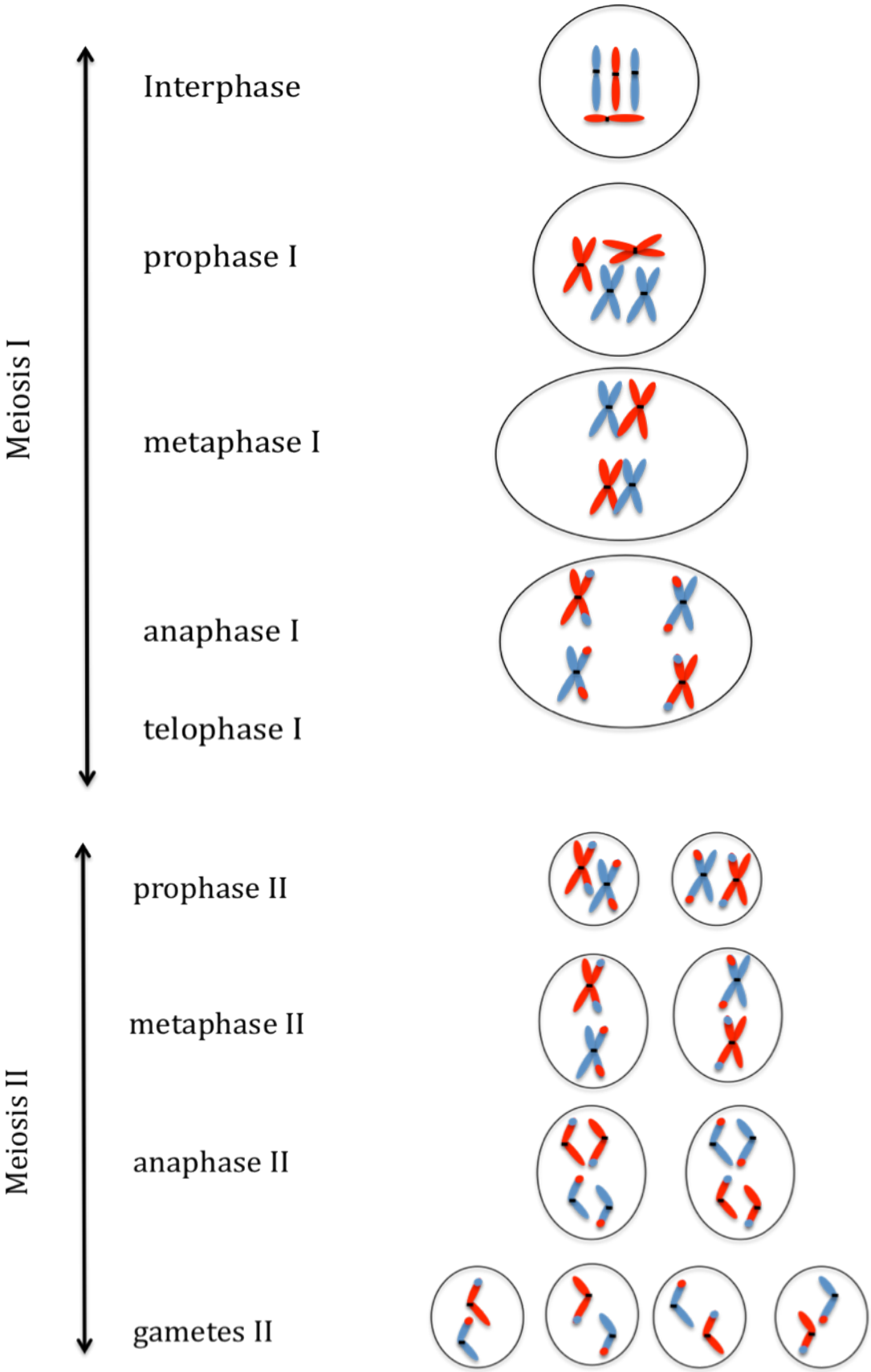


Figure 1-1 Overview of meiosis

1.1.1 Prophase I

Leptotene, zygotene, pachytene and diplotene/diakinesis are the phases of prophase I (Figure 1-2). During leptotene, homologous chromosomes physically line up with one another and axial elements (AE) form. Zygotene is the stage of initiation of synapsis between homologous chromosomes in the form of synaptonemal complexes (SC). Synapsis is then finished and the bivalent is stabilised by the time the cell reaches pachytene. In diplotene and diakinesis the SCs are disassembled and chiasmata are visible as the homologous chromosomes begin to separate. Metaphase I follows prophase I where the bivalent chromosomes are aligned on the metaphase plate. Until the transition to anaphase I, bivalents are held at metaphase I.

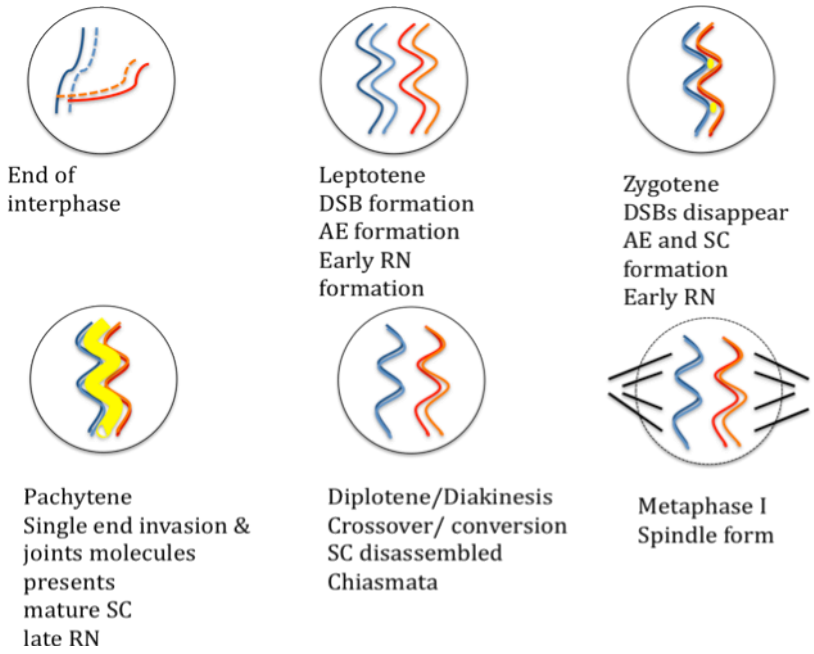


Figure 1-2 Interactions of one pair of homologous chromosomes (blue and red) during the stages of prophase. Replicated chromosomes during interphase are denoted by strands of different shades. Chromosomes condense during leptotene, and synapsis is initiated during zygotene and completed in pachytene. Chiasmata, can be seen during the diplotene/diakenesis stage. In metaphase I, breakdown of the nuclear envelope and formation of meiotic spindle are observed. DSB refers to double strand breaks: AE is axial element, RN is recombination nodules and SC is the synaptonemal complex (shown in yellow) (adapted from Page and Hawley, 2003; Roeder 1997).

1.2 MEIOTIC RECOMBINATION

Correct synapsis and proper segregation of homologous chromosomes occur during meiosis I by meiotic recombination. There are many proteins involved in the various steps of meiotic recombination: homologue recognition and pairing; the initiation of recombination and subsequent repair of homologous DNA strands; synapsis; and maintenance of the physical bonds between homologues.

1.2.1 Homologue Recognition and Pairing

Homologous pairing occurs prior to chromosome synapsis (Weiner and Kleckner, 1994; Nag *et al*., 1995) and implies that homologous chromosomes need to be brought together and aligned before recombination can take place, particularly in regions where the frequency of DSBs is reduced by DNA heterologies (Rocco and Nicolas, 1996; Xu and Kleckner, 1995). DSB formations can be affected by some heterologies, revealing the existence of an anti-initiation process sensing the presence of sequence non-homology between the homologous chromosomes (Rocco and Nicolar, 1996). There is evidence that the clustering of telomeres towards the nuclear envelope may have an effect on pairing, as this clustering occurs at the same time as pairing during meiosis. This telomere clustering might provide a mechanism for chromosome alignment that is conducive to homologous chromosomes finding one another. This assembly, termed the 'bouquet', is almost universal among sexually reproducing organisms, and has been extensively reviewed (Zickler and Kleckner, 1999; Bass, 2003).

1.2.2 Meiotic Recombination Is Initiated by Double Strand Breaks (DSB)

In *S. cerevisiae,* the formation of DSBs initiates recombination (Sun *et al*., 1989), and it is believed that DSBs are initiators of meiotic recombination in most organisms (Lin and Smith, 1994; Dalganov *et al*., 1996). There are many proteins involved in this multi-stage process. A non-random process of DSB formation happens in cluster sites of initiation (known as hotspots), which are usually located in promoters or in coding sequences (Lichten and Goldman, 1995, Bullard *et al*., 1996; Baudat and Nicolas, 1997). There is no evidence of any specific sequence motifs that control

DSB, but they are instead scattered across a 50-200-nucleotide tract within nuclease hypersensitive regions (Liu *et al*., 1995; Xu and Kleckner, 1995).

DSBs are formed by a topoisomerase-like transesterification which is catalysed by the Spo11 protein in *S. cerevisiae*, (Keeney *et al*., 1997) and by Spo11 orthologues in other organisms such as *Arabidopsis,* mouse and humans (Keeney *et al*., 1999; Shannon *et al*., 1999; Hartung and Puchta, 2000). It starts by breaking both strands of a DNA molecule and covalently linking the 5′ ends of the breaks to a tyrosine within itself (Keeney and Kleckner, 1995; Keeney *et al*., 1997). Even though Spo11 is an essential protein for initiation, there are many other proteins involved in this process (Johzuka and Ogawa, 1995; Roeder, 1997).

RAD50, MRE11 and *COM1/SAE2* are three genes fundamental for the resection process (Alani *et al*., 1990; McKee and Kleckner, 1997; Nairz and Klein, 1997). Rad50 and Mre11 proteins are required not only for DSB induction, but also for processing of the broken strands. Moreover, the exonuclease involved is thought to be a complex made up of Rad50 and Mre11 (Johzuka and Ogawa, 1995; Truijillo and Sung, 2001). After double-strand cleavage, the 5′- attached Spo11 protein is removed by the Com1/Sae2 protein along with the Rad50/Mre11/Xrs2 complex (Rattray *et al*., 2001). This prepares the chromosomes for strand invasion by leaving single tails with 3′– termini exposed by the Rad50/Mre11/Xrs2 complex (Cao *et al*., 1990; Nairz and Klein, 1997). From this, exonucleolytic resection is initiated.

Following resection, strand invasion of the 3' tail into the unbroken duplex of the homologous chromosome requires strand invasion of four yeast proteins: Rad51, Dmc1, Rad55 and Rad57 (Schwacha and Kleckner, 1997); which are homologues of the bacterial recA strand exchange enzyme (Shinohara *et al*., 1992). However, Rad51 is the catalyst for the invasion (Sung, 1994). The 3□-resected end invades the homologous chromosome to form structures known as Holliday junctions (Holliday, 1964).

In human cells, the Rad51 protein is required for both branch migration and Holliday junction resolution (Liu *et al*., 2004). With different cleavage orientation (vertical or horizontal) of a Holliday junction, the DSB repair system can cause either crossover

or non-crossover products (Storlazzi *et al*., 1995). However, a pathway without the involvement of a Holliday junction intermediate can also form non-crossover recombinants (Allers and Lichten, 2001).

1.2.2.1 Models of meiotic recombination

The first model is double-strand break repair (DSBR) (Szostak *et al*., 1983). In DSBR the D loop (Figure 1-3) captures the second 3′ single end and anneals to it via sequence complementarity. Ligation results in the generation of a double Holliday junction (Schwacha *et al*., 1995). Whether the double Holliday junction resolves into a crossover or noncross-over product depends on the direction in which the Holliday junction is resolved; if both Holliday junctions are cleaved in the same orientation then no crossover will occur, but if they are cut in opposite orientations then crossing over will occur. The DSBR model is reinforced by the visualisation of branched intermediates that correspond to Holliday junctions (Schwacha *et al*., 1997) (Figure 1-3).

A second model, called synthesis-dependent strand-annealing (SDSA), was first described in *D. melanogaster* while studying P-element induced gap repair (Nassif *et al*., 1994). This model has the same initial steps as the DSBR model, in that it invades the homologous chromosome to displace the 3' strand and anneal it to the original chromatid from which it came. The SDSA model can explain gene conversion because the conversion events are only expected on the recipient chromosome (Paques *et al*., 1999). The possible model for crossing over by the SDSA method has been suggested by Ferguson and Holloman (1996) and involves the formation of a D-loop and DNA synthesis, followed by re-annealing of the second 3' end to produce a single Holliday junction, and subsequent resolution. This model has since been adapted to incorporate a double Holliday junction, which can be cleaved to result in either a crossover or non-crossover (Paques *et al*., 1998) (Figure 1-4).

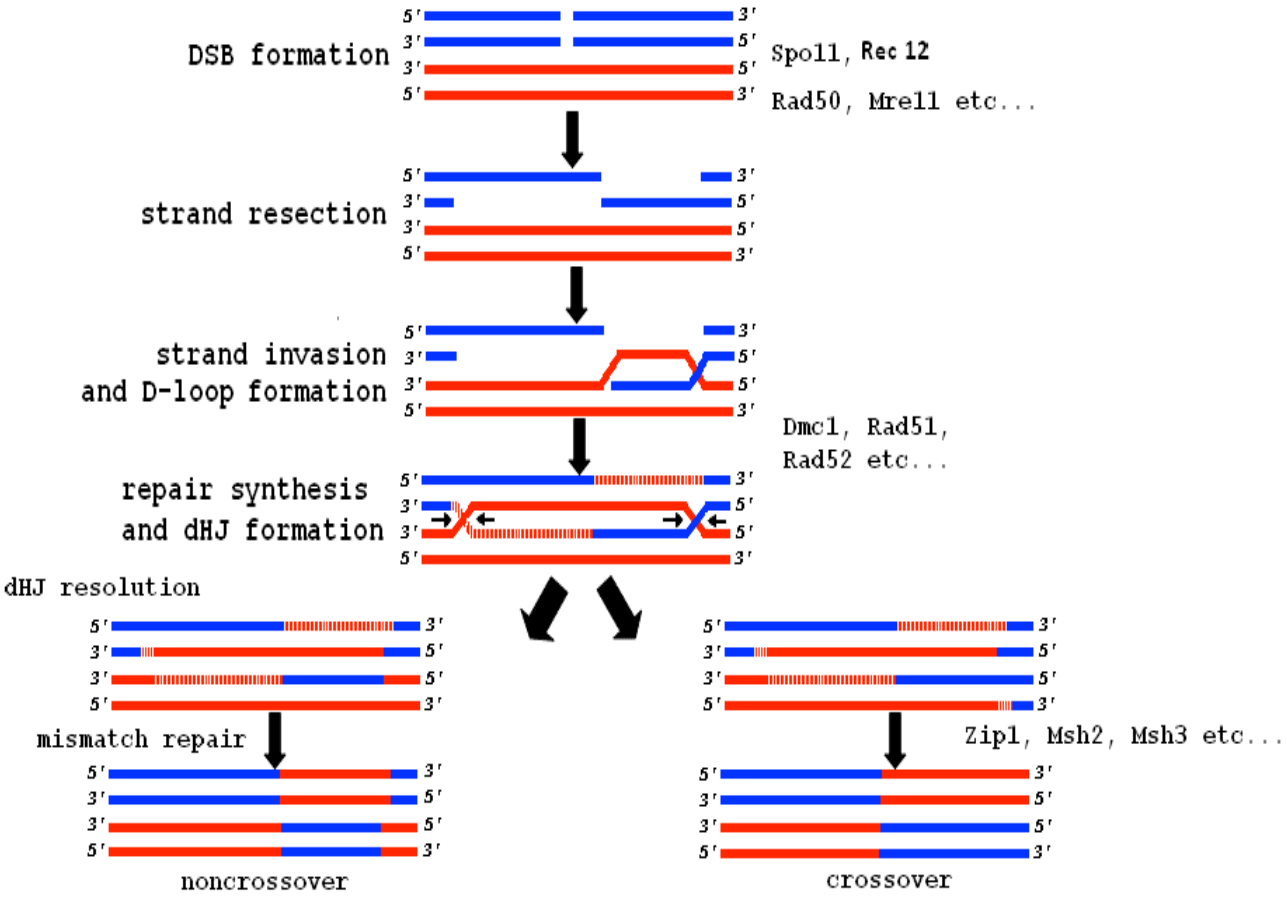


Figure 1-3 Double strand break repair model (DSBR). Diagrammatic representation of the DSBR model of recombination (adapted from Pagues and Haber, 1999; and Petes, 2001). DSB formation is followed by a 5' to 3' exonuclease resection of the ends, to leave a 3' overhang. This 3' overhang invades the intact homologue and produces a D loop, which can be extended by new DNA synthesis. In the DSBR model the second 3′ is captured and two Holliday Junctions (HJs) are formed; these can be resolved independently by cutting the crossed and non-crossed strands, resulting in crossover or non-crossover products.

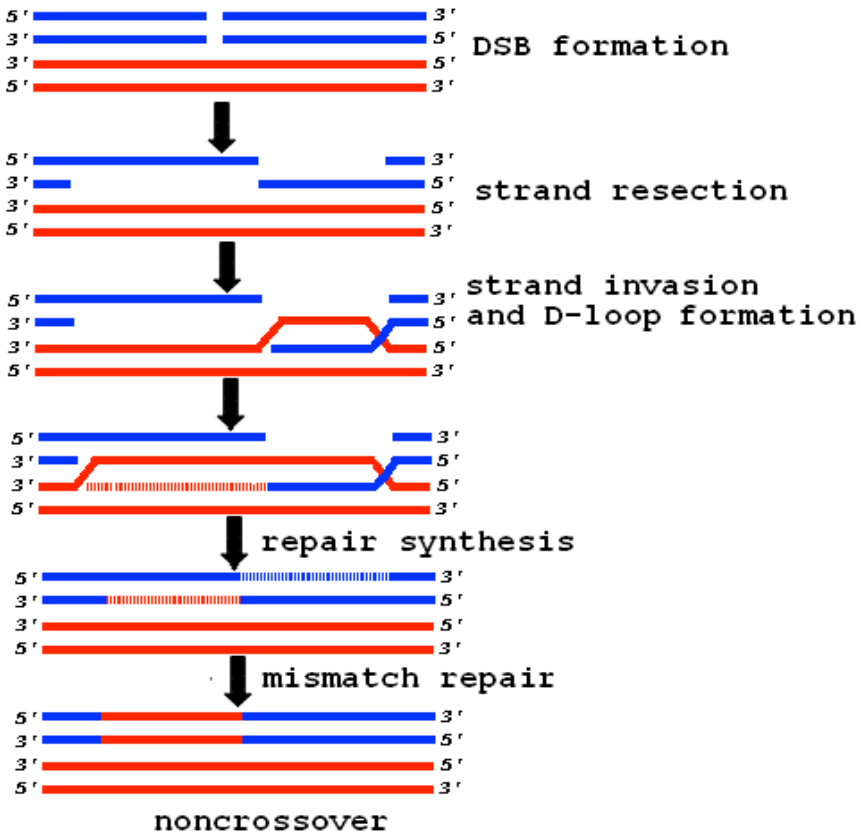


Figure 1-4 The synthesis dependent strand annealing (SDSA) model. Above is a diagrammatic representation of the SDSA model (Adapted from Pagues and Haber, 1999; Petes, 2001). DSB formation, resection, single strand invasion and extension occur as in the DSBR model. However, the extended 3' end is then expelled from the homologue and anneals to the original chromatid to produce a non-exchange gene conversion event. If the second 3' end invades the heteroduplex, it allows the formation of a double HJ, which can be resolved either with or without crossover. The SDSA model predicts crossover and conversions.

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

Despite similarity in the initiation step, where the Spo11 orthologue Rec12 is essential for the formation of DSBs; fission yeast (*Schizosaccharomyces pombe*) recombination models vary substantially from the models proposed for recombination in budding yeast (*Saccharomyces cerevisiae*) (Cervantes *et al.,* 2000). Inter-sister recombination is preferred over-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).

Single HJ is cut by Mus81 that is a part of resolvase complex Mus81-Eme. In crosses involving fission yeast mutant for Mus81, which display kinetics and normal numbers

of DSB formation, the number of crossovers is greatly reduce with little effect on the number of non-exchange conversion events (Cromie *et al.,* 2006). In budding yeast, the corresponding complex of Mus81/mms4 is not the major meiotic HJ resolvase, and the crossover frequency is reduced modestly by mutations in mus81 (de los Santos *et al.,* 2003).

1.2.2.3 The crossover/ non-crossover decision

Producing a crossover or non-crossover decision occurs before or at the appearance of strand exchange intermediates (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, 2001). The level of crossover is reduced by meiotic recombination proteins including Zip1, Zip2, Zip3, Mer3, Msh4 and Msh5, but these proteins maintain high frequencies of non-exchange conversion (Börner *et al.,* 2004). This points to the non-crossover mechanism as the main repair mechanism in the absence of crossover formation.

In contrast, crossovers tend to be maintained at the expense of non-crossovers at a given DSB location when the number of DSBs is artificially reduced (Martini *et al.,* 2006). This points to a specific subset 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.

1.2.3 The Synaptonemal Complex (SC) links homologous chromosomes

The SC is a meiosis-specific, chromatin-associated protein structure which is evolutionary conserved and found in many eukaryotes including *S.cerevisiae, D. melanogaster* and humans (Zickler and Olson, 1975; McKim *et al*., 1998; Moses *et al*., 1975). During leptotene, zygotene and pachytene, the SC, a tripartite ribbon-like structure, forms a physical link between homologous chromosomes.

During prophase I, an axial element (AE) develops between two sister chromatids of a single chromosome, and the SC is initially formed from AE. After formation of the SC, the AE is renamed as the lateral element (LE). During zygotene, synapsis begins

where transverse filaments form the SC structure and connect the AEs of two homologous chromosomes. Chromatin loops anchor the LE. The entire space encompassing the lateral and central elements is known as the central space (Zickler and Kleckner, 1999). Fewer crossovers are seen within larger loops (Roeder, 1995), with DSBs preferentially occurring in the middle of these chromatin loops (Blat *et al*., 2002)

The Zip1 protein of *S. cerevisiae* is required for SC production and subsequent synapsis, and is also involved in chromosome interference (Sym *et al*., 1993; Tung and Roeder, 1998; Sym and Roeder, 1994). Also, the CP1/Syn1 complex has been identified and characterised as protein components of the central element of the SC in mammals (Dobson *et al*., 1994; Meuwissen *et al*., 1997). Early studies using a *zip1* mutant in *S. cerevesiae* showed that the Zip1 protein is an essential structural element of the SC (Sym and Roeder, 1994). Moreover, chromosome synapsis is not necessarily required for recombination, as in some yeast mutants recombination can occur without the presence of a mature SC and hence synapsis. In *zip1* mutants, gene conversion occurs even in the absence of the SC (Roeder, 1995). The Zip1 protein was found to promote centromere coupling, to favour homologue pairing, and to serve as sites of synapsis initiation (Tsubouchi and Roeder, 2003). Recently, studies showed that the formation of the SC is a critical process in meiosis, although a positive correlation between crossovers and sites of synaptic initiation has been seen in some organisms (Henderson and Keeney, 2005).

1.2.4 Recombination Nodules mark the locations of recombination events

Recombination nodules (RNs) are electron-dense structures found within the SC and are visible with electron microscopy. Early RNs (ENs) form between the late leptotene or early zygotene and mid-pachytene phases, and are thought to mark the locations of recombination events. A second type of RN is the late RN (LN). LNs form in pachytene and sometimes in the diplotene phases. LNs are less frequent in most organisms (Zickler and Kleckner, 1999). Dmc1 and Rad51 are two RecA-like proteins thought essential in strand invasion and Holliday junction formation during DSBs (Anderson *et al*., 1997). Components of LNs are thought to be mismatch repair

proteins. Msh4, Nad, and Mlh1 and linked to the stabilisation or resolution of Holliday junctions (Ross-Macdonald and Roeder, 1994).

1.2.5 Two homologous non-sister chromatids exchange genetic material during crossover at chiasmata

Chiasmata are the physical structures left by the process of recombination and serve to covalently attach a sister chromatid from one chromosome to a sister chromatid of the homologous chromosome, to ensure correct disjunction of recombinant chromosomes at meiosis I. For correct disjunction, at least one chiasma is required per chromosome. Chiasmata can be visualised due to the accumulation of Mlh1 in mice (Baker *et al*., 1996) and humans (Tease *et al*., 2002), and counts of crossovers made. The cleavage of Rec8 protein at metaphase/anaphase dissolves chiasmata, when homologous chromosomes are separated to opposite poles of the cell (Lee and Orr-Weaver, 2001).

1.2.6 The numbers and distribution of crossovers are controlled by interference

At least one crossover must occur in each chromosome for prevention of non-disjunction, and therefore the size of the chromosome is important. With this knock-on effect, smaller chromosomes have proportionally more crossovers per unit length than larger chromosomes (Kaback *et al*., 1992). Moreover, if a large chromosome is split into two smaller chromosomes, the number of crossovers per unit length increases (Kaback *et al*., 1992).

Yeast mutants of zip1, zip2, and msh4 do not exhibit interference, possibly preventing the correct formation of the SC (Sym and Roeder, 1994; Novak *et al*., 2001). Organisms that lack an SC, such as *S.pombe* and *Aspergillus nidulans* also do not exhibit interference (Egel-Mitani *et al*., 1982; Kohli and Bahler, 1994). However, some organisms, yeast (Stahl *et al*. 2004), humans (Housworth and Stahl, 2003) and *Arabidopsis thaliana* (Copenhaver *et al*., 2002), are thought to generate a few crossovers that lack interference, compared with the total number of crossovers generated. Other organisms, such as *Drosophila* and *Neurospora*, are thought to only generate crossovers that exhibit interference (Copenhaver *et al*., 2002).

1.2.7 Mismatch repair proteins involved in recombination

During meiotic recombination, the correlation between heteroduplex DNA formation and the frequency of meiotic gene conversion indicates that mismatched base pairs are necessarily repaired by gene conversion after heteroduplex DNA is formed (Nag *et al*., 1995). Mismatch repair (MMR) was first characterised in bacteria. MutS, MutL and MutH are three proteins in the family, and their orthologues are components of the MMR system in yeast and mammals. In yeast, MutS and MutL are orthologues, whilst in mammals they are Msh1-6 and Mlh1-3, Pms1 and Pms 2.

Mismatches are recognised by a heterodimer made up of the MutS homologues Msh2 and either Msh3 or Msh6 (Marsischky *et al*., 1996), leading to recruitment of other heterodimers; Mlh1-Pms1 for the excision and synthesis of DNA (Prolla *et al*., 1994; Svetlanov and Cohen, 2004) and Mlh1-Mlh3. Mlh1 and Mlh3 are essential for proper recombination as mutants defective in the *mlh1* and *mlh3* genes show decreased numbers or complete absence of crossovers in yeast (Hunter and Borts, 1997; Wang *et al*., 1999). In mammals, MLH-1 and MLH-3 proteins are considered as markers for sites of recombination (Svetlanov and Cohen, 2004).

1.3 RECOMBINATION HOTSPOTS IN EXPERIMENTAL ORGANISMS

In most eukaryotic organisms, recombination has been observed in regions of either high or low activity (Lichten and Goldman 1995; Wahls 1998). Hotspots are generally relatively narrow regions of elevated recombination, whereas cold spots tend to be recombinationally inert (Petes, 2001). As a result, hotspots have been identified in model organisms, such as yeast, maize, mice etc.

1.3.1 Recombination Hotspots In Yeast

Most of the understanding behind the recombination processes involved in meiosis has come from studies of lower eukaryotes, in particular the yeast *S.cerevisiae*, in which recombination analysis is greatly helped by the ability to recover all four products from a single meiotic event.

At least ten meiotic recombination hotspots have been well characterised in yeast (Petes, 2001) including ARG4 (Sun *et al.,* 1991; Sun *et al.,* 1989), HIS4 (Fan *et al.,* 1995) and HIS4-LEU2 (Xu and Kleckner, 1995) in *S.cerevisiae* and ade6-M26 in *S.pombe* (Steiner *et al*., 2002). *S.cerevesiae* and *S.pombe* separated from their common ancestor about 420 to 330 million years ago (Sipiczki, 2000). High levels of DSBs are associated with many regions of 50-100 bp containing recombination hotspots (Xu and Kleckner, 1995), although there is no evidence of consensus sequences among these hotspots. Hotspots in yeast are also associated with the promoter regions of genes, with DSB sites preferentially occurring at intergenic regions rather than within genes (Wu and Lichten, 1994; Gerton *et al*., 2000).

Petes (2001) characterised hotspots in yeast into three classes; □-, □- and □-hotspots. □-hotspot activity depends on trans-acting transcription factors, but not on transcription itself – variation in □-hotspot activity may reflect the varying ability of different transcription factors to efficiently recruit the Spo11 endonuclease that initiates recombination by introducing a DSB. An example is the HIS4 hotspot, which is found within the promoter region of the *HIS4* gene, but studies using transcription factor mutants (Bas1, Bas2 and Rap1) show that hotspot activity is controlled by transcription factor binding (White *et al.,* 1993; Kirkpatrick *et al.,* 1999). Another example is *ade6-M26* hotspot activity in *S.pombe*, which is controlled by the binding of the heteromeric transcription factor Atf1/Pcr1 (Mts1/Mts2) (Kon *et al.,* 1997). □-hotspots such as HIS4LEU2 associate with only nuclease-sensitive sequences, and all DSB sites tend to fall in regions of DNaseI or micrococcal nuclease sensitivity (Wu and Lichten, 1994; Wu and Lichten, 1995; Fan and Petes, 1996). A sequence of 12 tandemly repeating CCGNN motifs stimulates recombination and transcription (Kirkpatrick *et al*., 1999). However, there is no evidence that transcription factors bind to these tracts, which supports the idea that □-hotspots are controlled by an alternative mechanism to □-hotspots (Petes, 2001). A third category is the □-hotspots, which have been observed in regions with high GC content (Gerton *et al*., 2000).

1.3.2 Maize Recombination Hotspots

In maize, there is a disproportionate frequency of crossing over in dista1 chromosomal regions (Dooner *et al.,* 1997). A sharper resolution of the variability of meiotic recombination in maize has been attained in studies of intragenic recombination at various loci. Dooner *et al*., (1997) studied the *bronze* (Bz) gene locus as a recombinational hotspot in maize and meiotic recombination products between heteroallelic pairs of *bz* mutations with both the presence and absence of heterologies have been studied. These studies have revealed an extreme non-randomness in the distribution of recombination events, and although genes comprise a small fraction of the maize genome, they behave in general as recombination hotspots (Dooner, 1986).

1.3.3 Mouse Recombination Hotspots

The majority of hotspots characterised in the mouse have been in the major histocompatibility complex (MHC) region. Pedigree analysis in mice is more practical than parallel experiments in humans, as their breeding can be controlled and the number of offspring is greater than in humans.

The *Pmsb9 (Lmp2)* hotspot was previously identified using pedigree analysis (Shiroishi *et al*., 1995). This recombination hotspot is located close to the *Pmsb9* gene and has been demonstrated to be a site of recombination initiation (Baudat and de Massy, 2007). Baudat and de Massy*,* (2007) have analysed this hotspot in both male and female germlines and compared the level of recombination in different hybrid mice. *Trans*-acting elements were observed to increase the recombination activity at Pmsb9 by 2,000-fold, while *cis*-acting elements were shown to repress the initiation of recombination. In addition, subtle variations in the frequency and distribution of recombination events were described between different strains and sexes. These findings suggest that most of the regulation observed acts at the level of initiation, and provide the first analysis of the control of the activity of a meiotic recombination hotspot in the mouse genome to reveal the interactions of elements located both in and outside the hotspot (Baudat and de Massy, 2007). A recombination hotspot has

also been characterised in detail at the Eβ gene in mice. It was initially analysed using inbred mouse pedigree methods, and subsequently by sperm typing (Yauk *et al*., 2003). At least seven recombination hotspots are known to exist in the mouse MHC, however not all of these have been refined by sperm typing.

1.4 MEASURING RECOMBINATION RATES IN HUMANS

In yeast, recombination can be studied efficiently by analysis of the four spores (tetrad analysis) resulting from a single meiotic event. Based on the knowledge developed from lower eukaryotes, meiotic recombination studies in humans have been analysed using more indirect approaches. Ratios of genetic to physical distance predict recombination frequencies, and hotspot locations in the human genome can also be estimated by these methods.

1.4.1 Pedigree Analysis

Pedigree analysis has been efficiently used to study recombination hotspots, providing us with an approximate indication of where recombination hotspots may lie. It is difficult to obtain detailed analyses of recombination events in the human genome by traditional pedigree analysis because of the given low frequency of meiotic recombination events per unit physical distance (~1 cM/Mb) (Kong *et al*., 2002). Polymorphic markers, for instance short tandem repeats in nuclear families with two to three generations, can be used for mapping (Kong *et al*., 2002; Kong *et al*., 2004). Recombination within the dystrophin gene has been shown to be approximately 4 times the expected value based on its length. Two hotspots have been identified from studies of minisatellite markers in CEPH families (Oudet *et al*., 1992). Furthermore, pedigree studies have characterised hotspots in the MHC class II region: one in the TAP2 gene (Cullen *et al*., 1995), one in the *DNA-RING3* region of the MHC, and one in the *DQB3-DQB1* region (Cullen *et al*., 1997). Of the 30 recombinant sequences in the *TAP2, DNA-RING3* and *DQB3-DQB1* regions, 21 were maternally derived, indicating a higher rate of recombination in females than in males (Cullen *et al*., 1997). One of the advantages of pedigree analysis is that both maternal and paternal meioses can be scored for crossovers.

Pedigree analysis of the phosphoglucomutase 1 (*PGM1*) gene, in Centre d'Etude du Polymorphisme Humain (CEPH) families have shown a region of increased recombination over a total of 58 kb, supported by additional linkage disequilibrium analysis which has marked the existence of two putative recombination hotspots in the region. Family data contributed to the discovery of two hotspots in the *PGM1* gene region, by studying patterns of inheritance of markers in CEPH families (Yip *et al.,* 1999). A recent study of fine scale patterns of recombination with pedigree data used single nucleotide polymorphism data (SNP) in 725 Hutterites (364 male and 364 female gametes), whereas previous studies were done by microsatellite data, and so this was able to develop a higher resolution linkage map using a very small amount of offspring. Also, this study reveals that overall recombination hotspot usage is similar in males and females. Moreover, extensive and heritable variation among both males and females in the proportion of crossovers occurs in this population (Coop *et al*., 2008).

1.4.2 Linkage Disequilibrium

Linkage disequilibrium (LD) is a term used for the non-random association of alleles at two or more loci. It describes a situation in which some combinations of alleles or genetic markers occur more or less frequently in a population than would be expected from a random formation of haplotypes from alleles based on their frequencies. LD can estimate the quantity and localisation of historical recombination in a region of interest.

LD estimations can help determine the location of recombination hotspots by estimating the association between alleles, for instance, if strong LD appears between alleles it can be inferred that little or no recombination has occurred historically between the markers. If weak LD appears between two markers, historical recombination can be inferred. LD can be used in association mapping by examining associations between an allele of interest and other alleles. There are many statistical measures of LD, with the most commonly used being D' (Lewontin, 1964).

D' values depend on the value of D. D can be calculated using the equation $D = x_{11} - p_1q_1$, where x_{11} is the observed frequency of haplotype AB, and p and q are the

expected frequency of alleles A and loci B. D is simple to calculate but its main disadvantage is its dependency on the frequency of the alleles inspected. There can be no D observed if any locus has an allele frequency 0 or 1, and it is maximal when frequencies are at 0.5. Lewontin (1964) suggested normalising D by dividing it with the theoretical maximum for the observed allele frequencies. Thus, $D' = D/D_{max}$, with D_{max} being the maximum level of D possible given the observed haplotype frequencies. (D_{max}=min [$p_1 q_1$, $p_2 q_2$] when $D < 0$, or, D_{max} = min [$p_1 q_2$, $p_2 q_1$] when $D > 0$) (Lewontin, 1964). When $|D'|=1$, this means the markers are said to be in complete LD, and there is no evidence for historical recombination between the two markers. If $D'=0$, it suggests there is free association between the two markers and therefore it is likely that extensive historical recombination has occurred.

Measured LD values can be used to create an LD plot (Figure 1-5) for the target region. These LD plots can then be used to determine the location of putative recombination hotspots within a region of interest prior to high-resolution sperm typing.

Previous studies using LD analysis have led to the characterisation of a number of putative human meiotic recombination hotspots inferred from regions of local LD breakdown. This approach includes hotspots in the □-globin gene cluster (Chakravarti *et al*., 1984), near the insulin gene (Chakravarti *et al*., 1984) and at the *PGMI* gene (Yip *et al*., 1999). Nevertheless, patterns of LD can be influenced not only by recombination rate but also by recurrent mutation (Jeffreys *et al*., 2001), natural selection, genetic drift and admixture (Ardlie *et al*., 2002).

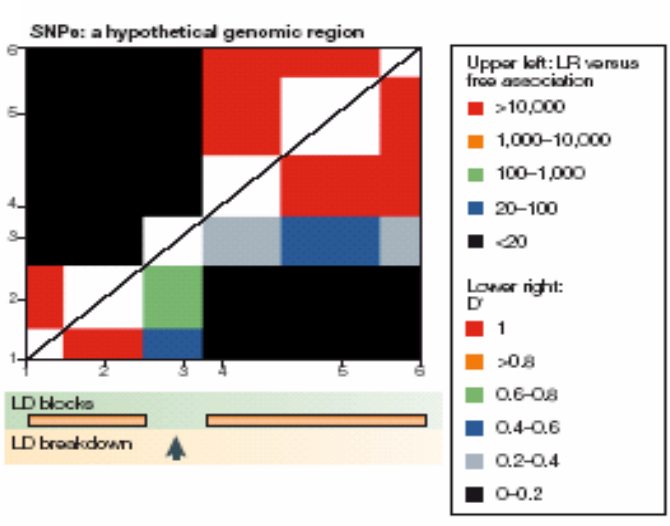


Figure 1-5 Graphical representation of a hypothetical LD plot: Linkage disequilibrium (LD) in a hypothetical genomic region. LD is calculated for each pair of single nucleotide polymorphisms (SNPs) from a panel of genotyped individuals (shown as vertical bars on the axis of plot) and is plotted as a rectangle centred on each SNP. The statistical importance of these values is shown above the diagonal and $|D'|$ values are shown below the diagonal. They are represented by a colour code. A region of LD breakdown is apparent between markers 2 and 4 indicated by the arrow. However, the likelihood ratio (LR) shows that the odds of equilibrium are high and therefore does not statistically support the D' value. (Diagram taken from Kauppi *et al.*, (2004))

1.4.2.1 Linkage disequilibrium unit (LDU) maps

Allowing LD to be presented as a metric map called a Linkage Disequilibrium Unit (LDU) map is similar to a centimorgan map, but with additive distances. Describing the relationship between distance and LD (Collins and Morton, 1998) as in the Malecot model (Malecot, 1948) can be adapted to pairwise LD data. Time (in generations) and recombination frequency control the decay of LD.

In LDU maps, each marker is given a position on the resulting map. The most efficient metric is *rho* (Morton *et al*., 2001), which is equivalent to $\square D' \square_{max}$, (Collins and Morton, 1998). However, any pairwise metric can be used while constructing the LDU map. Showing large-scale correlation with pedigree-based linkage maps allows LDU maps to be used in increasing the resolution of pedigree maps (Zhang *et al*., 2002; Tapper *et al*., 2003). Comparing LDU maps with coalescent-based maps (described below) shows that they are different from each other. LDU maps reflect

the underlying LD structure, whereas coalescent maps reconstruct the underlying recombination rate with a particular demographic model.

In this thesis, LDU maps are used prevalently. In Chapter 4, these maps are used to confirm putative LD hotspots in a donor panel. Previously, 493,408 markers have been used to produce an LDU map of the genome (Tapper *et al*., 2005). Recently, genomewide LD maps have been created using over 2 million SNPs and the LDMAP algorithm (http://cedar.genetics.soton.ac.uk/ pub/PROGRAMS/LDMAP). These LD maps were used to determine LD hotspots which have been confirmed by high-resolution sperm crossover analysis (Webb *et al*., 2008). In this thesis LD maps play a significant role for determining the putative hotspots.

1.4.3 Coalescent Analysis

Coalescent simulations are used to predict or infer historical recombination events from contemporary population genotype data, and for estimating recombination rates and locating hotspots (Fearnhead *et al*., 2004). Thus, they differ from LD maps in that complete haplotypes are related to a single common haplotype under a coalescent model. When using this approach for estimating recombination rates and predicting the location of hotspots, results are generally confirmed by direct estimated sperm typing (Schneider *et al*., 2002; McVean *et al*., 2004) or population-based surveys (Smith *et al*., 1998; Wall *et al*., 2003).

$P= 4N_e r$ is the formula for calculation of the population scale recombination rate (P) using coalescent approaches. External estimates of the effective population size (N_e) can infer the per-generation recombination fraction (r).

However, using LD data, recombination hotspots in the NID/TM7HSF1 (206 kb) region on chromosome 1q42.3 were discovered by high-resolution sperm typing (Jeffreys *et al*., 2005). Comparisons of coalescent analysis and sperm typing techniques in this region shows that there is a significant difference between sperm crossover frequencies and historical recombination rates. The reason could be that the hotspots may have evolved very recently.

1.5 MEASURING RECOMBINATION RATES IN HUMANS

Studying recombination in humans is more difficult than in lower eukaryotes. For example, the recombination rate in yeast is on average 1 cM per 3 kb (Petes, 2001), while the human sex-averaged rate is ~1.1 cM/Mb (Kong *et al*., 2002).

The huge number of sperm available from males (typically 10^8 sperm per ejaculate) provides the possibility of high-resolution analysis and recombination studies from individual men (Li *et al*., 1998). Sperm typing can be used instead of pedigree analysis to study recombination events in detail. There are two important methods of studying recombination in sperm DNA molecules. The first is to pick out and type individual sperm (Li *et al*., 1988), and the second relies on the amplification of single recombinant molecules from pools or batches of sperm DNA of up to 50,000 molecules in size (Jeffreys *et al*., 2000).

As with every method, sperm typing has drawbacks; firstly, sperm typing can only supply information on males and estimates of the male recombination frequency. Female recombination rates can only be estimated from family data, as obtaining eggs is far more difficult and would not provide the necessary number of gametes required for each study.

1.5.1 Single Sperm Typing Technique

Single sperm typing relies on the ability to isolate individual sperm. Each individual sperm can then be analysed by whole-genome amplification, followed by the polymerase chain reaction (PCR) to amplify markers of interest. Single sperm typing has led to the identification of putative recombination hotspots in a 208-kb-long DNA segment near the Huntington disease locus (Hubert *et al*., 1994) and allowed recombination frequency analysis within the pseudoautosomal region PAR1 (Lien *et al*., 2000). This technique is suitable for studying large genomic regions, however the resolution of single sperm typing studies has been restricted by low marker density and especially by the limited number of sperm that can be typed, typically in the region of a few hundred per study. Even the largest study to date, within the MHC, yielded only 325 recombinants from 20,031 sperm typed (Cullen *et al*., 2002)

1.5.2 Recombination Detection Using High-Resolution Sperm Typing (Batch Sperm Typing)

An alternative method for the detection of recombination events in DNA molecules has been developed to allow batches of sperm DNA rather than single sperm to be used. This method is termed High Resolution Sperm Typing and is capable of producing a high resolution of analysis of sperm molecules (0.0001 cM or less). Linkage Disequilibrium (LD) patterns are analysed by comparing identified single nucleotide polymorphism (SNP) markers and a target region for the analysis can be established where a breakdown in LD occurs; this indicates historical recombination and the presence of a putative hotspot. Initially, a target region spanning the LD breakdown in suitable donors is selected and amplified in a round of allele-specific PCR directed to alleles in repulsion phase at the outermost SNP sites, followed by a second round of allele-specific PCR directed to sites internal to the primary PCR sites. If a crossover product can be picked up after secondary PCR, a tertiary PCR can be performed using internal universal primers and the PCR product dot blotted to type SNPs by Allele-Specific Oligos (ASOs). The point of exchange is then determined for each crossover (Figure 1-6).

Both these methods were first developed to search and characterise the relationship between meiotic crossover and tandem repeat DNA instability in the germline (Jeffreys *et al*., 1998a). With the high-resolution sperm typing method not only can hotspots be determined, but gene conversions can also be obtained (Guillon and de Massy, 2002; Jeffreys and May, 2004). Indeed, the first human crossover hotspot (MS32) to be characterised at the molecular level was identified through the detection and analysis of crossovers at the MS32 minisatellite (Jeffreys *et al.,* 1998b) on chromosome 1. Approximately 40 allelic recombination hotspots have been identified in humans (Jeffreys *et al.,* 1998a; Jeffreys *et al.,* 2001; Jeffreys and Neumann, 2002; May *et al.,* 2002; Kauppi *et al.,* 2004; Jeffreys *et al.,* 2005; Holloway *et al.,* 2006; Webb *et al.,* 2008).

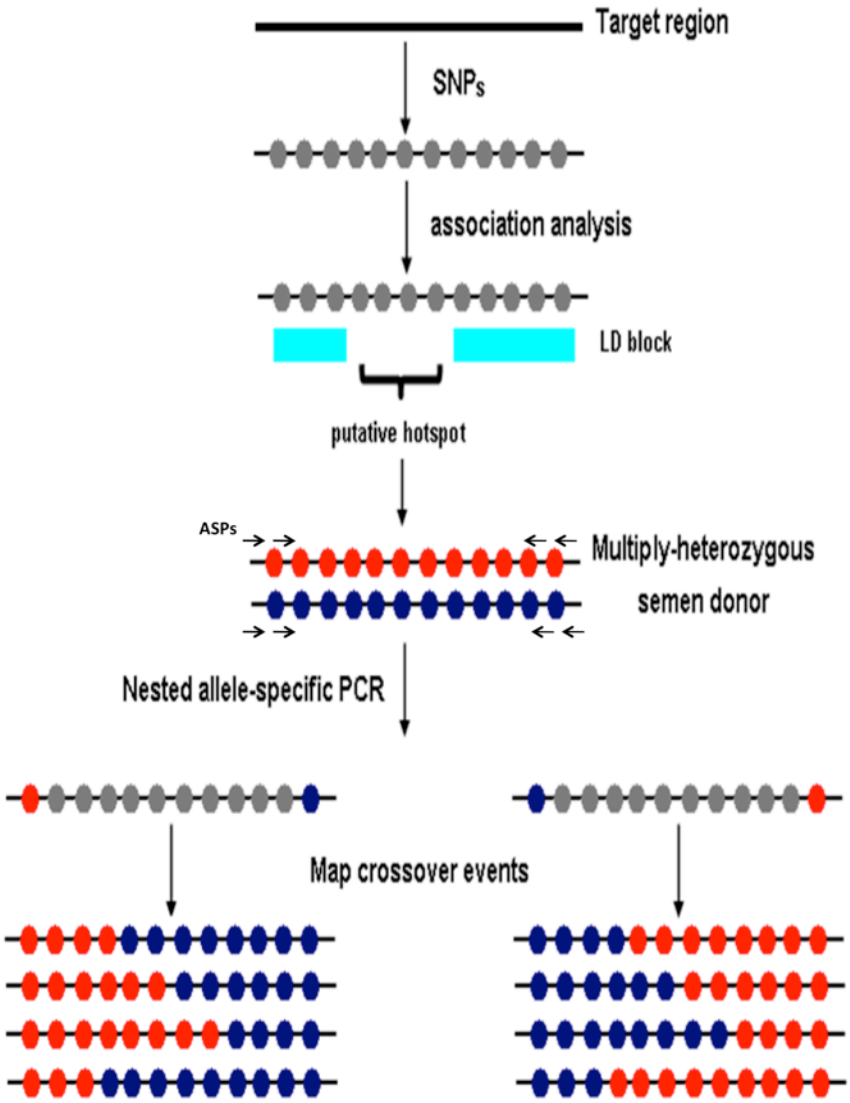


Figure 1-6 The high-resolution sperm typing approach for recovering crossover molecules directly from human sperm DNA. SNPs are identified in the chosen target region, followed by genotyping in a panel of semen donors. LD mapping allows the identification of LD blocks and can give clues about historical recombination events. Such regions of LD breakdown are putative recombination hotspots. The second stage starts with identifying a suitable semen donor. Two rounds of repulsion phase allele-specific PCR (allele-specific primers are shown by black arrowheads) and tertiary polymerase chain reaction (PCR) are used to selectively amplify recombinant molecules from sperm DNA; cross-over breakpoints are mapped by typing internal SNPs. Reciprocal crossovers *(red-blue, blue-red)* can be recovered separately using appropriate primer combinations.

1.6 HUMAN RECOMBINATION HOTSPOTS

Direct analysis of the fine-scale distribution of recombination events is not feasible using pedigree analysis but is possible using single molecule PCR methods to recover recombinant DNA molecules directly from sperm DNA (Jeffreys *et al.*, 2000; Jeffreys *et al.,* 2001).

The first human recombination hotspot associated with minisatellite MS32 on chromosome 1 was characterised using sperm typing and LD analysis (Jeffreys *et al.,* 1998b), and it was the discovery of this hotspot that provided the first clear indication that human meiotic crossovers generally concentrate into very narrow (1-2-kb) hotspots interspersed with recombinationally inert DNA (Jeffreys *et al.,* 2000). Other examples near minisatellite MS32 are the *NID1* recombination hotspot, which is located 58 kb upstream from minisatellite MS32. Several hotspots within the human major histocompatibility complex (MHC) class II region (Jeffreys *et al.*, 2001), and two loci within the *SHOX* gene in the Xp/Yp pseudoautosomal region of the short arms of the sex chromosomes have been characterised (May *et al.,* 2002). A near-continuous 103-kb region on chromosome 21 (Tiemann-Boege *et al.,* 2006), a recombination hotspot at β–globin (Holloway *et al.,* 2006), and 21 LD hotspots have been identified using sperm techniques (Webb *et al.,* 2008; Jeffreys and Neumann, 2009). Table 1-1 presents the hotspots in the MHC region, the MS32 region, *SHOX* and β–globin.

Region	Hotspot	Chr	95% width activity (kb)	Activity cM	Peak activity (cM/Mb)	Location	References
MHC	DPA1	6	1.7	0.028	27	intergenic	(Kauppi *et al.*, 2005)
	DNA1	6	1.9	0.0005	0.4	*HLA-DOA* promoter	(Jeffreys *et al.*, 2001)
	DNA2	6	1.3	0.0037	8	Intergenic, Alu	(Jeffreys *et al.*, 2001)
	DNA3	6	1.2	0.13	100	Intergenic, Alu	(Jeffreys *et al.*, 2001)
	DMB1	6	1.8	0.0027	5	exon/intron boundary	(Jeffreys *et al.*, 2001)
	DMB2	6	1.2	0.026	45	Intergenic	(Jeffreys *et al.*, 2001)
	TAP2	6	1	0.0058	8	*TAP2* intron	(Jeffreys *et al.*, 2000; Jeffreys *et al.*, 2001)
MS32	NID3	1	2	0.096	70	intergenic, Alu	(Jeffreys *et al.*, 2005)
	NID2a	1	1.4	0.0085	10	Intron	(Jeffreys *et al.*, 2005)
	NID2b	1	1.1	0.003	4	Intron	(Jeffreys *et al.*, 2005)
	NID1	1	1.5	0.05	70	Intron, Alu	(Jeffreys *et al.*, 2005)
	MS32	1	1.5	0.039	40	Intergenic, in RTLV-LTR	(Jeffreys *et al.* 1998a; Jeffreys *et al.*, 2005)
	MSTM1a	1	1.6	0.001	9	Intergenic	(Jeffreys *et al.*, 2005; Neumann and Jeffreys, 2006)
	MSTM1b	1	2.1	0.009	16	Intergenic	(Jeffreys *et al.*, 2005; Neumann and Jeffreys, 2006)
	MSTM2	1	1.3	0.0007	0.9	Intergenic	(Jeffreys *et al.*, 2005)
PAR1	*SHOX*	X/Y	2	0.3	300	Exon	(May *et al.*, 2002)
β-globin	β-globin	11	1.2	0.15	200	Intergenic	(Holloway *et al.*, 2006)

Table 1-1 Human recombination hotspots analysed by high resolution sperm typing.

1.6.1 Hotspots Within The Major Histocompatibility Complex (MHC) Region

Seven hotspots called *DPA1, DNA1, DNA2, DNA3, DMB1, DMB2,* and *TAP2* were identified in the class II major histocompatibility complex (MHC) (Kauppi *et al.,* 2005; Jeffreys *et al.,* 2000; Jeffreys *et al.,* 2001). Family studies gave the first clue of a putative hotspot in the *TAP2* gene (Cullen *et al.,* 1995; Cullen *et al.,* 1997), using a sperm typing technique to identify the 1-kb hotspot (Jeffreys *et al.,* 2000). The activity of the crossover hotspot in female meiosis was shown by pedigree studies (Cullen *et al.,* 1995). There are two more regions of LD breakdown occurring near the *TAP2* gene: the *DNA* locus and the *DMB* gene. Five crossover hotspots, three at the *DNA* locus *(DNA1, DNA2, DNA3)* and two in the *DMB* locus *(DMB1, DMB2)* have been characterised using crossover assay analyses at the region of *DNA* and *DMB* LD breakdown, with all showing 1-2 kb width. Referring to the observations of the pattern of LD blocks and associated hotspots in the MHC region, the other haplotype patterns in the human genome are traced (Jeffreys *et al.,* 2001).

1.6.2 The MS32 Region

Sperm typing around the GC-rich minisatellite MS32 on chromosome 1q42.3 revealed the first human crossover hotspot to be defined at high resolution (Jeffreys *et al.,* 1998a). This hotspot is centred 200 bp upstream of the minisatellite, with a peak activity of 40 cM/Mb and a width of 1.5 kb within which 95% of the crossovers occur (Jeffreys *et al.,* 1998a). It appears that this hotspot is responsible for driving repeat instability at MS32 (Jeffreys *et al.,* 1998a). Surprisingly, SNP markers around this active hotspot are in strong LD. Analysis of 206 kb surrounding MS32 revealed a block-like pattern of LD, as seen in the MHC, with five intervals of LD breakdown near the *NID* gene and upstream of *TM7SF1*, that might signal the presence of crossover hotspots (Jeffreys *et al.,* 2005). Sperm typing showed that all were genuine hotspots of 1-2 kb width, though with activities that correlated poorly with historical activity as estimated by coalescent analysis. Two of these hotspots proved to be doublets, with each showing an additional hotspot in a region of intense LD. This indicates that regions of strong marker association are not necessarily recombinationally inactive and might point to these hotspots being too young to have

left a mark on haplotype diversity. Also, one of these hotspots displays an on/off polymorphism, with only some men showing hotspot activity (Neumann and Jeffreys 2006). Surprisingly, active and inactive men can share the same hotspot haplotypes, suggesting that distal or *trans*-acting factors are involved in regulating hotspot activity.

1.6.3 Recombination Hotspot in the β-globin Region

Pedigree studies (Smith *et al.,* 1998), LD analyses (Antonarakis *et al.,* 1982; Chakravarti *et al.,* 1984; Wall *et al.,* 2003), coalescent analyses (Wall *et al.,* 2003), single sperm typing analyses (Schneider *et al.,* 2002), and high resolution sperm typing analyses (Holloway *et al.,* 2006) are all techniques used with β-globin recombination hotspots.

1.6.4 *SHOX* Hotspot In The Xp/Yp Pseudoautosomal Region

It is known that the most of human Y chromosome does not undergo recombination. Male meiosis is restricted to the terminal pseudoautosomal regions PAR1 and PAR2. The first region, PAR1, spans 2.6 Mb of the Xp/Yp terminus (Brown, 1988; Petit *et al.,* 1988). An obligate recombination event occurs, creating a male-specific recombination hot domain with a recombination rate ~20 fold higher than the genome average (Lien *et al.,* 2000). The 320-kb PAR2 is located at the end of Xq/Yq; its overall recombination rate is lower than that PAR1, though still six–fold higher than the genome average (Freije *et al.,* 1992; Li and Hamer, 1995). In females, recombination can take place along the entire X chromosome, but the recombination rate is lower at a genome average 1.5 cM/Mb (Henke *et al.,* 1993). The *SHOX* gene, in the PAR1 region, is located ~ 500 kb from the Xp/Yp telomere (Rao *et al.,* 1997), showing high male recombination in LD patterns, which can be assayed by sperm typing studies. In the *SHOX* gene, as with the MHC region, crossovers were not randomly distributed but instead crossover assays within a 9.9-kb sub-region revealed a recombination hotspot with a peak activity between 190- 300 cM/MB in a 2-kb-wide hotspot (May *et al.,* 2002). Previous studies revealed that recombination in this

region is static, and recombination does not occur along the PAR1 (Baird *et al.,* 1995).

1.7 HOTSPOT EVOLUTION AND GENE CONVERSION IN HUMANS

Evolution of Hotspots

Comparisons of DNA diversity in humans and chimpanzees have revealed remarkably divergent LD landscapes at orthologous loci despite ~99% identity at the DNA sequence level (Wall *et al.,* 2003; Ptak *et al.,* 2005; Winckler *et al.,* 2005). This information suggests that the turnover of recombination hotspots occurred within the last six million years and to an extent disproportionate to that of sequence divergence. Rapid hotspot turnover also increases the possibility of hotspot polymorphism within human populations, increasing from newly evolved hotspots. Extreme examples show a region of LD breakdown that is in fact recombinationally inert in sperm (Kauppi *et al.,* 2005) and a recombination hotspot that is yet to have an impact on the LD landscape (Jeffreys *et al.,* 2005). There is also evidence of hotspots that do not result in a discernible breakdown of LD, yet appear to have recombination rates similar to those leaving a crisp breakdown in LD, which is not predicted by the coalescent model (Jeffreys *et al.,* 2005). These 'invisible' hotspots may be young hotspots that have not yet made an impression on LD patterns.

In contrast to most other hotposts, in the MS32, *DNA2* and *NID1* hotspots both haplotypes show different exchange frequencies; asymmetric locations of exchange points were identified in the same men (Jeffreys *et al*., 1998a; Jeffreys and Neumann, 2002; Jeffreys and Neumann, 2005). Crossover asymmetry is caused by a heterozygous SNP at the centre of the hotspot; with one allele initiating the crossover and promoting biased gene conversion with over transmission of the recombination allele. As estimated in the DSB repair model, recombinational meiotic drive of a suppressing allele can eradicate the hotspots, resulting in the"hotspot paradox", as fortified by these findings (Boulton *et al.,* 1997; Pineda-Krch and Redfield, 2005).

Gene Conversion

As in yeast, gene conversion at human crossover hotspots has been observed (Jeffreys and May, 2004; Jeffreys and Neumann, 2005). Human gene conversions have been examined at the *DNA3, DMB2, SHOX, NID1, S1* and *S2* hotspots (Jeffreys *et al.,* 2001; May *et al.,* 2002; Jeffreys *et al.,* 2005; Jeffreys and Neumann 2009). These studies indicated that the same recombination can trigger both crossovers and gene conversions, although gene conversion events show more variation between hotspots compared to crossovers (Jeffreys and May, 2004; Holloway *et al.,* 2006). Moreover, according to the short tract of conversion events, difficulties in detecting them from LD data and the affects of a few SNPs is thus highly localised within recombination hotspots. A lack of gene conversions in LD regions may have serious consequences for association studies. Thus it is essential for LD to be characterised at a high resolution to identify these conversions. Crossover rate and marker density restrict the high-resolution comparisons of crossover and conversion at recombination hotspots. More research is essential in this area, because little is known about the distribution of conversion events.

1.8 THE HAPMAP PROJECT

In October 2002, the international HapMap project (a haplotype map of the human genome; www.hapmap.org) was established with the aim of producing a freely available single nucleotide polymorphism (SNP) haplotype map of the human genome to facilitate tagSNP selection. A tagSNP is a representative SNP in a region of the genome with high linkage disequilibrium. Such SNPs are useful in genome-wide association studies in which hundreds of thousands of SNPs across the entire genome are genotyped. For this reason, the International HapMap Project aimed to use tag SNPs to discover genes responsible for certain disorders. In identifying such SNPs, the project was successful in aiding researchers to identify genes that underlie the susceptibility of the carrier to common diseases. Initially, the project analysed DNA from 269 individuals from four populations, namely the Yoruba people of the Ibadan region in Nigeria (YRI), the Japanese of Tokyo (JPT), the Han Chinese of

Beijing (CHB), and the Centre d'Etude du Polymorphisme Humain in Utah (CEPH), USA. CEPH residents have ancestry from northern and western Europe (CEU).

October 2005 heralded the completion of the first phase of the project, yielding data on more than one million validated SNPs with an average density of 1 per 5 kb (The International HapMap Consortium, 2005). This was followed in July 2006 by HapMap with a further ~4.6 million SNPs showing a density of 1 per kb being released in phase II of the project data (International HapMap Consortium 2005). These markers provide the resources to produce a highly detailed LD map of the human genome. Recently, in HapMap Phase III, 7 more populations of African ancestry have been completed: in the south-western USA (ASW); Chinese in metropolitan Denver, Colorado, USA (CHD); Gujarati Indians in Houston, Texas, USA (GIH); Luhya in Webuye, Kenya (LWK); Maasai in Kinyawa, Kenya (MKK); Mexican ancestry in Los Angeles, California, USA (MXL); and Tuscans in Italy (Toscani in Italia, TSI) are included (International HapMap 3 Consortium, 2010). As a result, 1.6 million common SNPs in 1,184 reference individuals from 11 global populations have been genotyped.

Furthermore, 10 million common DNA variants have been identified by The Human Genome Project (International Human Genome Sequencing Consortium, 2001) the SNP Consortium (The International SNP Map Working Group, 2001) and the International HapMap Project (The International HapMap Consortium, 2007). This information along with linkage disequilibrium patterns allows an understanding of genome-wide association studies. In this thesis, the international HapMap Phase II data is used for identifying the SNPs for regions of interest.

1.8.1 Confirming Human Hotspots from LD Data by High-Resolution Sperm Crossover Analysis

Using coalescent analyses genomewide LD landscapes were determined by the International HapMap Project (The International HapMap Consortium, 2005, The International HapMap Consortium, 2007), revealing kilobase level resolution of hotspots. Additionally, fewer then 20 human meiotic hotspots were identified in sperm surveys that equal 0.6 Mb of human DNA to date, and they have shown good harmony between the location of LD hotspots and sperm hotspots (Webb *et al*., 2008). According to previous studies, and to clarify whether LD landscape can accurately predict and locate hotspots or correctly estimate their historical activity, hotspots which show the most extreme LD breakdown in HapMap genotypes have been analysed by sperm crossover assays, and the repertoire of human crossover hotspots has been doubled with hotspots A, B, C1, C2, D, E, F, G1, G2, H, J1, J2, K, L, M, N, P, Q, R, S1, S2 (Webb *et al*., 2008, Jeffreys and Neumann 2009). These studies revealed that comparisons of LD patterns with crossover profiles show that population diversity data are excellent predictors of hotspots and can generally locate them with accuracy but are poor at predicting hotspot intensity (Webb *et al*., 2008).

1.9 DNA SEQUENCE MOTIFS ASSOCIATED WITH RECOMBINATION HOTSPOTS

Previous studies have shown that many sequences in the human genome are associated with enhanced recombination through their close proximity to recombination hotspots. Only two factors stand out as being strong and consistent correlators of recombination rate: GC content and the location of the motif most strongly associated with hotspots (Myers *et al.* 2006). Sequence motifs promoting crossover activity have been previously found in several organisms, for example a single eukaryotic hotspot, M26 in *S.pombe* (Fox *et al.,* 2000). Studies in *S.cerevisiae*, have also shown that there is a strong association between recombination rate and sequence motifs (Myers *et al.,* 2006).

Recent studies have shown that fine-scale coalescent analysis show peaks that are highly related in rate, most likely due to recombination hotspots, with about 80% of

all recombination events occurring within one-fifth of the genome sequence (Myers *et al.,* 2005). HapMap Phase II data (22,699 autosomal and 608 chromosome X hotspots mapped to within 5 kb) revealed in excess of 30,000 hotspots in the human genome with an average density of around 1 per 50-100 kb (Myers *et al.*, 2005; Myers *et al.,* 2008). To investigate particular repeat sequences and simple sequence motifs associated with hotspots, Myers *et al.,* (2005) compared hotspot regions with regions of the same size and SNP density which showed no evidence of recombination activity; they named these matched regions "coldspots". Hotspots and coldspots show different frequencies of particular sequences (Myers *et al.,* 2005). DNA sequences which can move around to different positions within the genome of a single cell, the long terminal repeats of the THE1A and THE1B retrovirus-like transposons, along with CT-rich and GA-rich repeats were over-represented in hotspots. GC-rich repeats, certain L1 elements and $(TA)_n$ repeats, were significantly under-represented (Myers *et al.,* 2005).

The strongest correlation of all was seen with a seven-nucleotide oligomer CCTCCCT that aligns to the positions of 261 to 267 in the THE1B consensus (Myers *et al.,* 2005). THE1B in hotspots occurs more frequently in hotspots when compared to coldspots. Chromosomes active at the hotspot contain the motif CCTCCCT, with the "suppressor" mutation being a change from T to C in its third position. Suppressor alleles disrupt the hotspot activity of motifs. At most, this motif accounts for 11% of the 25,000 hotspots examined (Myers *et al.,* 2005). The latest studies show a 13bp hotspot motif CCNCCNTNNCCNC, a 6 bp extension of CCTCCCT motifs that determines the location of at least 40% of ~30,000 human crossover hotspots (Myers *et al.,* 2008). These motifs are found in hypervariable minisatellites, clusters in the breakpoint regions of disease-causing non allelic homologous recombination hotspots, common mitochondrial deletion hotspots, and it is assumed that it drives genome instability. In addition, a nine nucleotide long motif CCCCACCCC is enriched amongst HapMap hotspots. The hotspots *DNA2* and *NID1* (Jeffreys *et al.,* 2002, Jeffreys *et al.,* 2005), characterised by sperm typing and crossover rate polymorphism data contain the motifs CCTCCCT and CCCCACCCC at their centre. These polymorphisms may be determined by recombination frequency (RF) measurements in sperm by analysis of differences in crossover distributions in the reciprocal products of crossing-over (Jeffreys *et al.,* 2002). SNPs are located in one of the

hotspot motifs, with the lower activity allele being the derived state and disrupting the motif (Myers *et al.,* 2005). Consequently, hotspot activity can be influenced by local sequence determinants. Jeffreys *et al.,* (2002) have shown that the recombination-suppressing allele is over-transmitted to recombinant progeny due to biased repair processes operating during recombination. A recent study found DNA sequence motifs associated with recombination in G-quadruplex DNA, and the co-occurrence of potential G4 DNA within 7 short sequence elements (SES) that are associated with recombinogenic regions (Mani *et al.,* 2009). This finding is interesting because most organisms' DNA is in the double stranded formation (B-DNA) (Mani *et al.,* 2009). In this thesis the argument for association of a 13-bp DNA sequence motif with recombination hotspots will be discussed.

1.10 *TRANS*-REGULATOR EFFECTS ON RECOMBINATION HOTSPOTS

Little is known about the location of the elements associated with hotspot activity, despite the determination of the recombination landscape. The direct molecular detection of DSBs allows mapping of initiation sites in *S. cerevisiae* and *S. pombe* (Kon *et al*., 1997; Keeney, 2008). Previous studies in yeast and mammals show that some DNA sequences determine the location of hotspots (Kon *et al.,* 1997; Shiroishi *et al.,* 1991; Neumann and Jeffreys, 2006). For instance, the loci *Dsbc1* and *Rcr1* are *trans*-acting loci that control the activation of specific hotspots. These loci are located in overlapping 5.4 Mb and 6.3 Mb regions on mouse chromosome 17 (Parvanov *et al.,* 2009; Grey *et al.,* 2009). Additionally, the trimethylation of lysine 4 of histone H3 (H3K4me3) is known as a common chromatin feature that also defines yeast and mouse initiation sites (Borde *et al.,* 2009; Buard *et al.,* 2009). Moreover, *in silico* studies show that LD-based hotspots were found highly associated with a 13-bp motif (Myers *et al.,* 2008), and sperm-typing analysis illustrates associations with variation of hotspot activity in *cis* (Jeffreys and Neumann, 2002; Jeffreys and Neumann, 2005).

PRDM9, PR domain–containing 9, is a meiosis-specific histone H3 methyltransferase with a C-terminal tandem-repeat C2H2 zinc finger (ZnF) domain encoded by a minisatellite (Hayashi *et al.,* 2005). It is uniquely expressed during early meiosis in both males and females (Parvanov *et al.,* 2010). Recent studies revealed that, depending on the carried *PRDM9* alleles, the recombination profiles and crossover

hotspot activity showed variation in mouse subspecies hybrids (Parvanov *et al.,* 2010). In humans, according to computational analyses (Myers *et al.,* 2010), PRDM9 has been identified as a human ZnF protein that recognizes the 13-bp CCNCCNTNNCCNC motif. This motif might also serve as a binding site for the ZnF protein (Myers *et al.,* 2010; Myers *et al.,* 2008). Nevertheless, an allele of *PRDM9* is the most common allele that binds *in vitro* (Baudat *et al.,* 2010), but the *PRDM9* allele 'I' cannot bind to this motif (Baudat *et al.,* 2010).

Previous studies in the Jeffreys laboratory have investigated the influence of variation in the PRDM9 ZnF array (Berg *et al*., 2010). The results demonstrate that crossover activity at individual human recombination hotspots and genome instability both at minisatellites and at pathological non-allelic homologous recombination (NAHR) rearrangements, are all influenced by PRDM9 variation. Sixteen different forms of PRDM9 containing between 8 and 18 zinc fingers have been found in 74 African and 156 European semen donors (Berg *et al.,* 2010) (Table 1.2). Ten active hotspots (including 5 hotspots with the sequence motif) were examined, and the results showed that all 10 hotspots showed activity dependent on PRDM9, and were generally activated by the common allele 'A'. This study implies that PRDM9 operates across a much longer track than the 13-bp sequence motif (Berg *et al.,* 2010). The most recent study shows that the activation of PRDM9 variants that are common in Africans but rare in Europeans reveals second-class hotspots, and these hotspots act differently even in the same populations (Berg *et al.,* 2011).

Table 1.2 Activating and non-activating *PRDM9* alleles. Allele structures are coded as in Berg *et al.*, 2010, with predicted DNA binding sequences and best motif matches shown. (This figure is adapted from Berg *et al.*, 2010)

Transgenic mouse studies showed that the hotspot activity (as inferred by histone H3 lysine 4 trimethylation (H3K4me3) levels, and chromosome-wide distribution of crossovers) is changed by a sole modification of PRDM9 zinc fingers (Grey *et al.,* 2011). Also, this study demonstrated with an *in vitro* assay that the three tested hotspots are activated when a *PRDM9* variant binds specifically to DNA sequences located at the hotspot centre (Grey *et al.,* 2011). Thus, mutations in *cis* located at the hotspot centre and associated with a decrease of hotspot activity can affect PRDM9 binding. In this study, PRDM9 is presented as a master regulator of hotspot localisation through the DNA binding specificity of its zinc finger array, and that binding of PRDM9 at hotspots promotes local H3K4me3 enrichment (Grey *et al.,* 2011) (Figure 1.7).

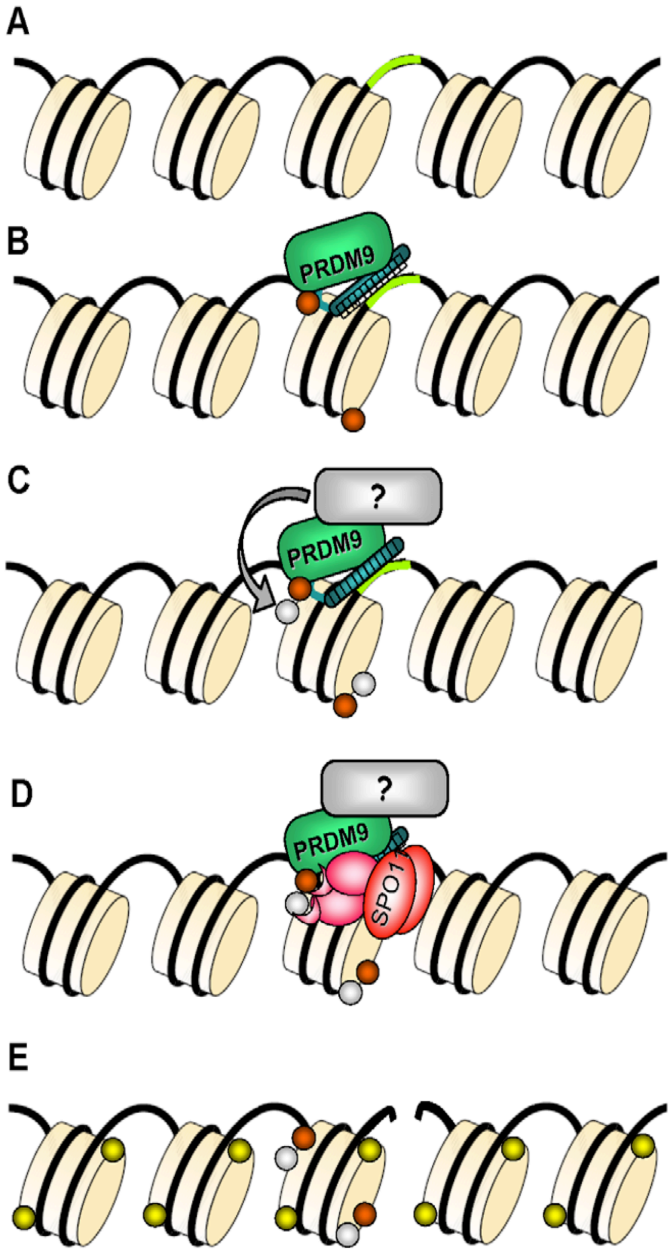


Figure 1.7 Model for hotspot specification by PRDM9. a) The DNA and several nucleosomes are shown. A DNA sequence motif recognised by PRDM9 is represented in light green. b) PRDM9 binds to its target DNA motif through the zinc finger array and catalyses H3K4me3 (orange). c) A protein partner of PRDM9 may catalyse another post-translational histone modification (grey), allowing for the formation of a hotspot-specific signature. d) PRDM9, a partner, or other component of the chromatin may recruit the recombination initiation complex containing SPO11 or may create a favorable chromatin environment allowing access of SPO11 to the DNA. e) A DSB is formed by SPO11 and triggers the phosphorylation of histone H2Ax (yellow) in the surrounding nucleosomes. Then, homologous recombination is repaired by the DSB and gives a crossover or non-crossover. (This figure is adapted from Grey *et al.*, 2011).

Single nucleotide polymorphisms (SNPs) from whole-genome resequencing and genotyping studies have been used to estimate recombination rates across the mouse genome (Brunschwig *et al*., 2012). This analysis identified 47,068 historical hotspots - an average of over 2477 per chromosome. The positions of historical hotspots were identified by simulation in inbred mouse strains (Brunschwig *et al.,* 2012), and considering the predicted binding sequences for different alleles of the PRDM9 protein enriched the location of recombination hotspots. Recombination rates were on average lower near transcription start sites. This study suggests that characterising and studying the dynamics of historical hotspots can be done in inbred strains, and also strengthens previous findings on mouse recombination hotspots, and specifically the impact of sequence variants in *Prdm9* (Brunschwig *et al.,* 2012).

Finally, the *SPRY3* hotspot in the PAR2 region was studied by Sarbajna *et al.,* (2012), and this revealed that PRDM9 strongly modulates both crossovers (COs) and non-exchange gene conversions (non-crossovers, NOs) in *trans*. Additionally, a SNP located at the hotspot centre that appears to influence recombination initiation, causes biased gene conversion in *cis*. This study showed that recombination initiation frequencies vary between haplotypes but are not mediated by a *cis*-effect. Also, through subsequent processing, they have the potential to significantly intensify the meiotic drive of hotspot-suppressing alleles (Sarbajna *et al.,* 2012).

1.11 WORK IN THIS THESIS

Recombination is an important evolutionary force in shaping the human genome and creating human diversity. In humans, at least one recombination event per chromosome is required for proper segregation of chromosomes, and in cases of non-disjunction (when chromosomes fail to separate normally into the daughter cells), genetic disorders such as Down's syndrome (three copies of chromosome 21) may result. The recombination event may be resolved as a crossover, where information is reciprocally exchanged between chromosomes, or a non-exchange gene conversion event. Recombination events have mostly been studied in yeast and mice. The reciprocal exchange that takes place during meiotic division can occur nearly anywhere along a chromosome. However, meiotic recombination occurs more

frequently in some regions of the genome called hotspots, which arise in 1-2 kb intervals across the human genome (Jeffreys *et al.,* 2000). Hotspot analysis has focussed on the description of crossover profiles, but at the beginning of this work little was known about the factors that influence the localisation and frequency of human crossover hotspots.

Furthermore, our knowledge about regulating factors of the human recombination machinery is very limited. Recombination hotspots could be influenced by single base changes known as single nucleotide polymorphisms (SNPs). The International HapMap Project, established in 2002, (www.hapmap.org) is a multinational effort to identify and catalogue genetic similarities and differences in human beings. This project has helped and will continue to help the discovery of uncharacterised inherited genetic variation in human disease by establishing a SNP map of the human genome. Approximately 4.6 million SNPs have been characterised in four different populations in this project so far. This data can also help us to identify putative localised crossover hotspots. Sperm typing has been used most successfully to identify hotspots in humans. The first characterised hotspot using sperm typing was the MS32 hotspot on chromosome 1 (Jeffreys *et al.,* 2000). Biased gene conversion had been found responsible for over-transmitting alleles, and therefore implicated in hotspot silencing. Could hotspots be affected by a DNA sequence motif? DNA sequence motif composition is thought to play an important role in recombination, and many sequence components have an effect on recombination rates. The aim of this thesis is to identify hotspots with the DNA sequence motif and which include a motif-disrupting SNP to test whether the disrupting allele influences the crossover frequency and distribution. Or are there any other regulatory factors (*cis*- or *trans*-) that influence hotspot activity? This work aims to gain insights into these interesting questions.

Chapter 3 investigates four previously identified LD hotspots by Donnelly and his colleagues from low-resolution HapMap Phase II data. These LD hotspots based on having a 13-bp CCNCCNTNNCCNC motif with a disrupting SNP and a good historical activity. To confirm the presence of these LD hotspots and whether the disrupting SNP within the motif located at the centre of the hotspots, high-resolution genotyping techniques used for genotype out European semen donor panel. This

chapter shows and prove by LDU mapping that the motif disrupting SNP within the 13-bp motif is located at the centre of only one LD hotspot.

Chapter 4 includes the recovery of recombinant DNA molecules, the analysis of recombination rate and distribution, and determination of the hotspot centre of Hotspot DA. Crossover frequencies and distributions were compared with previously established parameters. Also, this chapter characterises biased gene conversion by comparing recombination frequencies and distributions in both reciprocal orientations. Interestingly, the results show the highest observed transmission distortion ratio. Furthermore, it was tested whether crossovers and non-crossovers are influenced by the same biases. For better understanding of the effect of *cis*-regulation on the studied hotspot, the analysis was expanded to more men to allow the identification of DNA sequence motifs that carry the disrupting SNP to control hotspot activation. This chapter shows for the first time an example of strong *cis*-regulatory control on a human crossover hotspot.

Chapter 5 investigates the important question remaining, as to the effect of the *trans*-regulator factor PRDM9 on the activity of Hotspot DA. Therefore, in this chapter, more men with variant PRDM9 genotypes were analysed to explore possible influences of PRDM9 status on crossover frequencies. This chapter shows for the first time whether *cis*-regulation is more effective than the *trans*-regulator factor PRDM9.

CHAPTER 2: MATERIAL AND METHODS

2.1 MATERIALS

2.1.1 Suppliers

ABgene (Epsom, UK), Applied Biosystems (Warrington, UK), Bio-Rad (Hemel Hempstead, UK), Cecil Instruments (Cambridge, UK), Clare Chemical Research (Delores, USA), Eppendorf Scientific (Hamburg, Germany), Fisher Scientific (Loughborough, UK), Fisons (Beverley, USA), Flowgen (Ashby-de-la-Zouch, UK), FMC Bioproducts (Rockland, USA), GE Healthcare (Little Chalfont, UK), Hybaid (Teddington, UK), Invitrogen (Paisley, UK), MJ Research (Waltham, USA), New Brunswick Scientific Co. (New Jersey, USA), PerkinElmer (Cambridge, UK), Qiagen Ltd. (Crawley, UK), Serva (Heilderberg, Germany), Sigma-Aldrich (Poole, UK), Sysngene (Cambridge, UK), Thermo Scientific (Pittsburg, USA), USB Corporation (Cleveland, USA), UVP Life Sciences (Cambridge, UK).

2.1.2 Chemical Reagents

2.1.2.1 Enzymes

T4 polynucleotide kinase and Exonuclease I was supplied by New England Biolabs and Fermentas. *Pfu* polymerase was obtained from Gibco-BRL and *Taq* polymerase was obtained from Abgene and Kapa *Taq* from Kaba biosystems. Shrimp Alkaline Phosphatase (SAP) was obtained from Roche.

2.1.2.2 Molecular Weight Markers

λ DNA digested with *Hin*dIII and ϕX174 RF digested with *Hae*III was supplied by ABgene.

2.1.2.3 Oligonucleotides

Three suppliers have been used for synthesing oligonucleotides; The Protein and Nucleic Acid Chemistry Laboratary (PNACL), University of Leicester, UK; Invitrogen (Paisley, UK); Sigma-Genosys (Haverhill, UK).

2.1.2.4 Human DNA

Semen, and in some cases blood were collected from 200 men of North European decent 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, University of Leicester, UK, were anonymised before use. Blood and sperm DNA samples were prepared under high containment to minimize 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 (Department of Genetics, University of Leicester, UK) 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 (donated by A.D Nakomo and S.B. Kanoyangwa, Forensic Science Laboratory, Causeway, Zimbabwe) men were prepared, giving in total genomic sperm DNA from 270 donors available for analysis. Donor-panel master-plates containing whole-genome multiple displacement amplified (MDA) DNA were generated by R. Neumann (Department of Genetics, University of Leicester, UK) using the GenomiPhi Amplification kit (GE Healthcare). All studies were granted Local Ethical Committee approval.

2.1.2.5 Standard solutions

20x sodium chloride-sodium citrate (SSC) buffer and 10x Tris-borate/EDTA (TBE) electrophoresis buffer, were as described by Sambrook *et al.*, (1989), and were supplied by the kitchen, Department of Genetics, University of Leicester, UK.

2.1.2.6 Preparing 11.1 X PCR buffer

Typically 10 ng of DNA was amplified using an MJ tetrad PCR machine PTC 250 (supplied by GRI, Braintree, Essex, UK) and unless otherwise stated, was in a 10 μl total reaction volume. All reagents and plasticware were PCR clean, all pipettes dedicated to PCR only and all reactions were carried out in a category II laminar flow hood to limit contamination.

11.1 x PCR buffer (Jeffreys *et al.,* 1990) was prepared by Rita Neumann (Department of Genetics, University of Leicester, UK) as indicated below. dNTPs and BSA were supplied by GE Healthcare.

Component	**Concentration of Stock Solution**	**Volume (arbitary units)**	**Final Concentration (in PCR)**
Tris/HCl pH 8.8	2M	167	45 mM
Ammonium Sulphate	1M	83	11 mM
$MgCl_2$	1M	33.5	4.5 mM
2-mercaptethanol	100%	3.6	6.7 mM
EDTA pH 8.0	10mM	3.4	4.4 μM
dATP	100mM	75	1 mM
dCTP	100mM	75	1 mM
dGTP	100mM	75	1 mM
dTTP	100mM	75	1 mM
BSA	10mg/ml	85	113 μg/ml
Total Volume		**676**	

For amplicon sizes of less than 2 kb, no Tris base or *Pfu* were added. For long-PCR these reagents help amplification. *Pfu* is a proofreading 3' to 5' polymerase, which removes base mismatches that can cause *Taq* polymerase to stall during amplification, resulting in shorter products or PCR jumping. Tris base was added to raise the pH of the reaction, thus reducing the risk of template depurination at high temperatures and allowing denaturation at lower temperatures.

In reactions where the input of template DNA was below 5 ng carrier herring sperm DNA was added to the PCR reaction at 1μg/ml to coat the inside of the PCR tubes and prevent the target DNA being sequestered as DNA binds to certain types of plasticware.

Cycling conditions for PCR varied according to amplicon length. Typically the reaction conditions were as follows: denaturing for 2 min at 94^{o}C, followed by 25-32 cycles of 94^{o}C for 20 sec, 54-66 ^{o}C for 30 sec, and 66^{o}C for 4-15 min and finally a soak at 15^{o}C. Small numbers of PCRs, typically less than 50, were carried out in 200 μl PCR tubes, with larger numbers of PCRs carried out in PCR plates and sealed with self-adhesive PCR film (ABgene).

2.1.2.7 Whole Genome Amplification by Multiple Displacement Amplification (MDA)

Whole genome amplification was carried out as described by Dean *et al.,* 2002. Reagents for a 50 μl reaction were as follows Table 2-1:

Reagent	Stock concentration	Final concentration	Volume Added (μl)
Tris-HCl pH 7.5	2 M	37 mM	0.925
Potassium chloride	2 M	50 mM	1.25
Magnesium chloride	1 M	10 mM	0.5
Ammonium sulphate	1 M	5 mM	0.25
dATP	100 mM	1 mM	0.5
dCTP	100 mM	1 mM	0.5
dGTP	100 mM	1 mM	0.5
dTTP	100 mM	1 mM	0.5
Thiophosphate modified random hexamer primer	1m M	50 μM	2.5
Yeast pyrophosphatase	200 u/ml	1 u/ml	0.25
ϕ29DNA polymerase	unknown	800 u/ml	0.5

Table 2-1 Reagents used in whole genome amplification reaction.

The reagents were added to 40 ng of good quality PCR clean DNA and the reaction volume made up to 50 μl with PCR clean water. After 30^{o}C for 18 hours incubations and then 65^{0}C 3 minutes to stop the reaction, samples are ready to use. MDA DNAs were prepared by A.J. Jeffreys (Department of Genetics, University of Leicester, UK).

2.1.3 Computers

Data and images were stored and processed using the software packages AutoAssembler, Papers for Mac, Factura, Chroma, Microsoft Office and Cricket

Graph. Images either tranfered to .jpg via Mac Preview or using the Epson Perfection 1250 Photosmart scanner.

LDU plots were generated from un-phased diploid genotyping data using LD mapping software modified by Adam J. Webb (University of Leicester, UK) from version 1.5 of LDMAP (Maniatis *et al.,* 2002). Historical recombination rates (coalescent analysis) were estimated using data with the LDhat program version 2.0, available from www.stats.ox.ac.uk/~mcvean/LDhat/ (McVean *et al.*, 2004).

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 (Department of Genetics, University of Leicester, UK) 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 available at http://www.quantitativeskills.com/sisa/statistics/t-test.htm. Again, Mann Whitney U test performed on Poisson means were calculated using available at http://vassarstats.net/utest.html

Plotting and this thesis was produced on a MacBook Pro and PC running Microsoft Windows XP. Word processor used was Microsoft Word (2007) and 2008 (Mac). Citation manager was Paper2 (Mac). Graphs were produced in Microsoft Excel 2007 (PC), Microsoft Power Point 2008 (Mac), Cricket (Mac). Internet searches were performed using Safari. DNA sequences were analysed on Autoassembler and Factura (Mac) and Chrome (PC).

2.2 METHODS

2.2.1 DNA Extraction from Semen

DNA extraction was carried out in a category II laminar flow hood. Frozen semen samples were slow thawed on ice for 1 hour and an aliquot of 200 µl removed and transferred to a screw-top Eppendorf tube. 1 ml of 1 x SSC was added to the tube, the

contents mixed and then centrifuged at 13,000 rpm for 2 min. The pellet was resuspended in 1ml of 1 x SSC and SDS added to 0.2% (w/v) to lyse non-sperm cells. The contents was mixed gently by flicking and centrifuged at 13,000rpm for 2 min. The supernatant was removed, the pellet re-suspended in 1ml of 1 x SSC and SDS added to 0.2% (w/v). The contenets were centrifuged for 2 min, the supernatant removed and the pellet re-suspended in 800 μl of 1 x SSC. The transparent pellet was resuspended in 450 μl 0.2 x SSC. Sperm heads were lysed by the addition of SDS to a final concentration of 1% and 2-mercaptoethanol to a final concentration of 1M and the sample incubated at room temperature for 5 min. Proteinase K was added to a concentration of 200 μg/ml and this was incubated for 45 mins-1 hour at 37^{o}C, mixing occasionally. Proteins were removed by addition of 300 μl phenol/chloroform (with gentle mixing for 5-10 min to allow emulsification, and then centrifuged for 2 min at 13,000 rpm. The organic layer was re-extracted with 1 x SSC and 1% (w/v) SDS. DNA was ethanol precipitated with 2 volumes of 100% (v/v) ethanol and gentle swirling. After centrifuging, the supernatant was removed and the pellet washed with 1ml of 80% (v/v) ethanol, and dried. The pellet was dissolved in distilled water and 0.1 volumes of sodium acetate (pH7.0) were added, followed by 2 volumes of 100% (v/v) ethanol and centrifuged for 1 minute. The supernatant was removed and the pellet air-dried and dissolved in 5mM Tris-HCl pH 7.5.

2.2.2 DNA Extraction from Blood

Venous blood samples (delivered into equal volumes of 1x SSC and stored at -80^{o}C) were thawed at 37^{o}C and 500 μl was transferred to a 1.5 ml screw-topped eppendorf tube. 800 μl of 1x SSC was added gently mixed and then centrifuged for 2 min. The supernatant was removed and the cell pellet washed twice with 1x SSC and centrifuging. The pellet was resuspended in 300 μl 0.2x SSC and cells were lysed by adding 30 μl 10% (w/v) SDS and incubated at room temperature for 5 min. Proteinase K was added to a final concentration of 200 μg/ml and was incubated at 37^{o}C for 1 hour, with occasional mixing. Trace proteins were removed by addition of phenol/chloroform, with gentle mixing to allow emulsification and then centrifuged for 3 min. The organic layer was re-extracted twice with 1x SSC and 0.2% (w/v) SDS. The DNA was ethanol precipitated, dried and diluted as for the sperm DNA.

2.2.3 Measuring DNA Concentration

DNA concentrations were estimated by visual comparisons of signal intensity against DNA of known concentration following agarose gel electrophoresis in the presence of ethidium bromide. The DNA concentration was then estimated by measuring optical density at a wavelength of 260nm (OD_{260}) at a dilution of 1 in 200, or if very concentrated 1 in 1000, using a Cecil Instruments CE 202 Ultraviolet Specrophotometer at wavelength. Two aliquots of each sample were measured and the average taken. This was then used to calculate the concentration of DNA by using the equation (dilution factor x OD_{260}) / 0.02 with 0.02 being the weight given to double-stranded DNA measured at 260nm. Oligonucleotide concentrations were also estimated in this way, however 0.024 was substituted for 0.02.

2.2.4 Gel Electrophoresis

SeaKem LE agarose (Cambrex) was used to make gels at a concentration of 0.8-1.2% (w/v) and sizes 20 cm x 20 cm. PCR products were run using a horizontal submarine format with 0.5xTBE (44.5 mM Tris-borate (pH 8.3), 1 mM EDTA) buffer containing 0.5 μg/ml ethidium bromide. ¼ volume of loading dye (30% glycerol, 0.5 x TBE, bromophenol blue to colour) was added to each sample prior to loading between 1-5 μl into the well. The samples were electrophoresed along side markers of known size. DNA bands was visualised using a GeneGenius system (Sysgene). Photographic records were acquired using the GeneGenius (Syngene) and printed using a Sony digital graphics printer UP-D895 (Syngene). Electrophoresis tanks were manufactured in-house, and power supplies were provided by Bio-Rad.

2.2.5 PCR Amplification

Generally, polymerase chain reaction (PCR) DNA amplification was carried in 10 μl total reaction volumes in 200 μl tubes or 96-well plates sealed with self-adhesive film on an a MJ Tetrad PCR machine PTC 250 (supplied by GRI, Braintree, Essex, UK) and Veriti Thermal Cycler (Applied Biosystems, Warrington, UK). All PCR reactions were carried out in a category II laminar flow hood to limit contamination;

moreover all reagents and plastic ware were PCR clean and all pipettes dedicated to PCR clean. To amplify 5-10 ng DNA template, 0.9 μl of 11.1 x PCR buffer (described above), 12.5 mM Tris base, 0.2 μM of each primer, 0.05 U/μl *Taq* polymerase, 0.005 U/μl *Pfu* polymerase were mixed for per reaction. For accurate *Pfu* activity, which is a 3′ to 5′ proofreading exonuclease that removes mismatches, is helped by Tris base that raises the pH. These possible mismatches may cause *Taq* to stall during primer extension.

2.2.6 PCR Cleanup by Exonuclease I and Shrimp Alkaline Phosphatase (SAP) Purification

Amplified DNA (PCRs) that produces a single clean PCR product can be treated with Exonuclease 1 (20U/μl) and SAP (1 U/μL) before resequencing or reamplification. Excess primers and single-stranded DNA can be digested by Exonuclease 1 while excess dNTPs are removed by SAP. 1 μl of 20U/μl Exonuclease 1 and 3 μl of 1 U/μl SAP were added into 10μl PCR product. An MJ Tetrad PCR machine was used to hold tubes with the following conditions: 37^{o}C 1 hour; 80^{o}C 15 minutes to inactivate enzymes.

2.2.7 Automated DNA Sequencing

ABI PRISM BigDye^{TM} Terminator Cycle sequencing reaction Kit, version 3.1 was used for an automatic sequencing. 20-30 ng per kilobase of template DNA, 4 μl Big Dye mix and 3.2 pmol of sequencing primer were prepared in 10 μl reaction. An MJ Tetrad PCR machine was used to cycle tubes 25 times with the following conditions: 96^{o}C 10 sec, 50^{o}C 5 sec, 60^{o}C 4 min. The second step is adding 10 μl H_20 and 2 μl Sodium Dodecyl Sulphate (SDS) (10% w/v). After that tubes were held at 98^{o}C for 5 minutes followed by 25^{o}C for 10 minutes. The Qiagen DyeEx 2 sequencing cleanup kit (QIagen) was used to clean up the sequencing following the manufacturer's protocol. Sequencing was carried out on an Applied Biosystems 3730 sequencer at the Protein and Nucleic Acid Chemistry Laboratory, University of Leicester. Sequences were analysed by eye using Factura (Mac) and assembled using AutoAssembler (Mac) and Chroma (PC).

2.2.8 Dot Blots

3-100 ng per kilobase DNA was available per reaction after PCR amplification; ¼-loading dye (30% [v/v] glycerol, 0.5x TBE, bromophenol blue to colour) was added to 20 μl of PCR product followed by at least 5x volume of denaturing mix (0.5 M NaOH, 2M NaCl, 25mM EDTA) was added to provide at least 30 μl per filter. A vacuum was applied to the assembled 96 well dotblot manifold harbouring one sheet of Hybond^{TM}-N^{fp} (Amersham) nylon membrane (pre-soaked in 2x SSC [100 ml distilled water, 17.6 g of sodium chloride and 8.9 g of tri-sodium citrate, pH 7.0]), and two sheets of Whatman 3mm chromatography paper to act as a backing. 30μl of the denatured DNA was loaded into each well for each replica filter. Each well was then washed with 150 μl 2x SSC to neutralise the sample. The filters were dried at 80˚C for 10 min and the DNA covalently linked to the membrane by 7x10^4 J/cm^2 of UV light in an RPN 2500 ultraviolet crosslinker (GE Healthcare).

2.2.9 Dot Blot Hybridization

18mer Allele-specific oligonucleotides (ASOs) were designed and ordered from Sigma, with the SNP site as position 8 from the 5' end of the oligo. 8 ng of each Allele-Specific Oligonucleotide was labelled in a 10 □l reaction containing in 1μl 10 x Kinase mix (700mM Tris-HCl pH 7.5, 100mM $MgCl_2$, 50mM spermidine trichloride, 20mM dithiothreitol) 0.35 U/μl of T4 polynucleotide kinase, 7.8 μl water and 0.12 μl 10mCi/ml γ-^{32}P-ATP) for 45 min-5 hr at 37˚C. Following incubation 20 μl of kinase stop solution (25 mM diNa EDTA, 0.1% SDS, 10 μM ATP) was added to the labelling reaction. Dot blot filters were soaked in 3x SSC and pre-hybridised in 2ml hybridisation solution [3M TMAC (tetramethylammonium chloride, Sigma), 0.6% (w/v) SDS, 1 mM diNa EDTA (pH 8.0), 10 mM sodium phosphate (pH 6.8), 0.1% (w/v) Ficoll 400, 0.1% polyvinylpyrrolidone, 0.1% (w/v) BSA, 0.4 μg 10mg/ml yeast RNA] at 48.5˚C for 5 min (10 min if the signal was likely to be low due to low DNA yield) in a mini hybridisation oven. The hybridisation solution was discarded and 2.5ml of fresh hybridisation solution was added. This was hybridised for 5 min at 48.5˚C. Prior to adding the [γ-^{32}P]-labelled oligo and unlabelled 160-320 ng

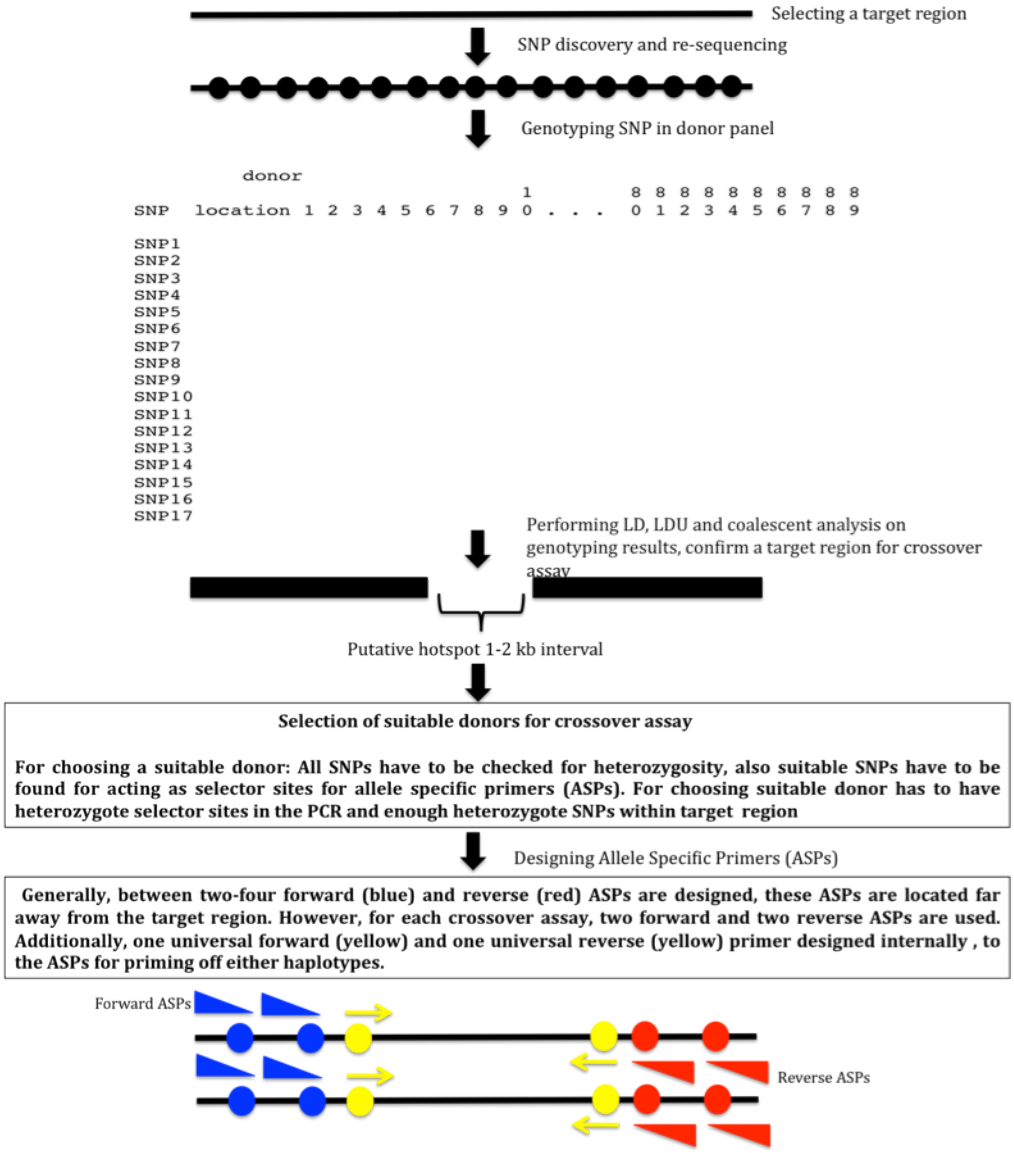
competitor ASO (e.g. opposite allele) and hybridising for 1-2 hr at 48.5˚C. The hybridisation solution was discarded, and the filters washed in TMAC wash solution (3M TMAC, 0.6% [w/v] SDS, 1 mM EDTA (pH 8.0), 10 mM sodium phosphate, (pH 6.8)) for 2x 10 min and 1x 15 min, using a fresh aliquot of 2.5ml of wash solution each time. The filters were then rinsed in 3x SSC at room temperature, blot dried and wrapped in Saran wrap before exposure to Fuji RX100 X-ray film at -80˚C with an intensifier screen from 3 hours to overnight, depending on the strength of the signal. Screens were scanned using TypHoon 9400 variable mode imager (GE Healthcare). Images analysed using imageQuant TL v2005 (GE Healthcare).

2.2.10 Probe Removal from Membrane

Probes on membrane were removed by washing in boiling 0.1% (w/v) SDS. The washing was continued (4-5 changes) and monitored by a Geiger counter until probe removal was complete. Membranes were rinsed in 2x SSC and stored damp at 4^{o}C.

2.2.11 Crossover (Recombination) Assay

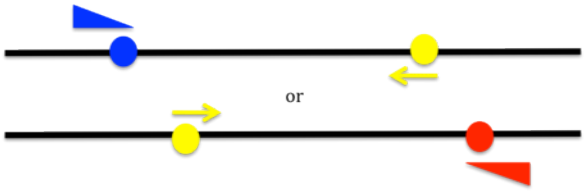
Design and execution of recombination assays was an extremely complicated and time-consuming process, with many stages. A simplified version of the steps involved in designing and performing the assay is shown in figure 2.1. Each stage is described in more detail below.



(Figure 2.1 continues on the next page)

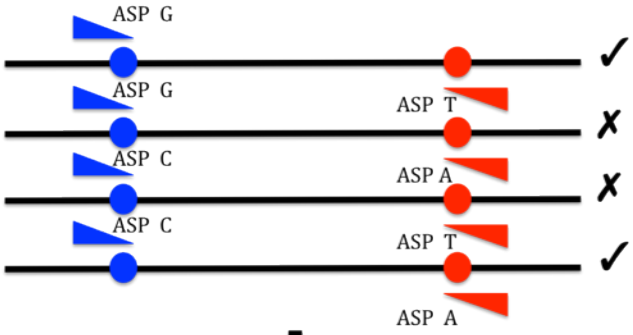
Optimisation of ASPs

PCR performed using a forward ASP in conjuction with reverse universal primer, or a forward universal primer in conjuction with a reverse ASP on donor homozygous for the allele in the primer and those homozygous for the opposite allele



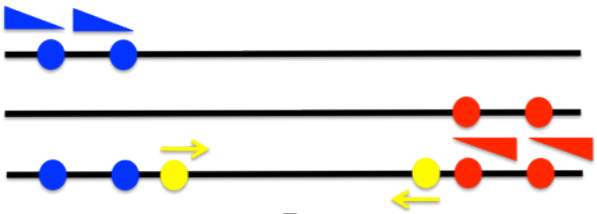
Establishing linkage phase

Amplify DNA from chosen suitable donor using a forward ASP (i.e SNP G/C is used for designing ASP G/C) in conjunction with a reverse ASP (i.e SNP A/T is used for designing ASP T/C), use all four combinations of primers in separate reaction. Those ASPs directed to alleles in coupling phase will produce a product (✓), those in repulsion phase will not (✗). The outermost forward primer should never be used in conjunction with the outer reverse ASP, to eliminate any chance of introducing false recombinants to future experiments.



Pilot recombination assay

Nested repulsion phase allele-specific PCRs performed on pools of sperm DNA and blood DNA to selectively amplify recombinant sperm molecules only. Pool sizes typically range from 300-9600 molecules



Large scale recombination assay and mapping of recombinant products

The large scale assay is performed in pool sizes depending on the optimum pool size determined from the pilot assay that maximises the number of recombinants recovered, while limiting the amount of mixed reactions. Mixed reactions occur when more than one recombinant molecule is present in a single PCR reaction. These reactions are impossible to identify from gel electrophoresis of the products and can only be seen when the internal SNPs are mapped. The mixed reactions show up as a mix of alleles at a SNP site. Typing of internal SNPs is done in the presence of progenitor haplotype PCRs.

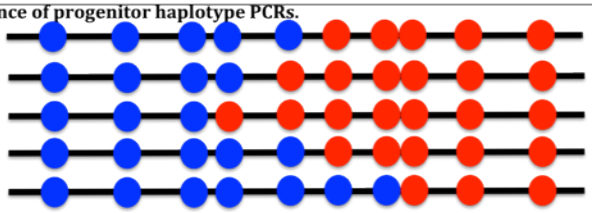


Figure 2.1 Steps involved in design and implementation of a recombination assay. The figure shows a flow chart of steps that need to be taken when designing and carrying out a recombination assay. Re-sequencing and SNP discovery is performed on a subset of the DNA donors. All SNPs genotyped by ASO hybridisation in every donor and LD analysis performed on the genotype data. LD analysis allows identification of a target region for the crossover assay. By examining the genotype data, it is necessary to select a semen donor on which to perform the recombination assay. The criteria for this are for the individual to be heterozygous at a number of SNP sites within and flanking the target region identified by the LD analysis and they are heterozygous at potential primer sites. Allele-specific primers (APSs) are designed at suitable SNP sites, away from the target region. Usually two forward and two reverse allele-specific primers are designed and used in nested primary and secondary PCRs, as well as one forward and one reverse universal primer (a primer that can amplify both haplotypes). ASPs are optimised using donors homozygous for each SNP allele. Each donor is amplified using a forward ASP in conjunction with a reverse universal primer at several different temperatures and the optimum temperature determined. Reverse ASPs are optimised with the forward universal primer. Linkage phase of the donor selected needs to be established in order to perform the recombination assay. DNA from this donor is amplified using forward ASPs in conjunction with reverse ASPs. The four possible combinations of forward and reverse ASPs are used in separate reactions. If the reaction produces a PCR product, the alleles in the ASPs are said to be in coupling phase. If the PCR is unsuccessful the alleles are said to be in recombinant phase. A pilot recombination assay is performed, using primary ASPs that are in recombinant phase and nested secondary primers that are also in recombinant phase on differing pool sizes of semen DNA. The pilot assay is used to determine the optimum pool size of sperm DNA molecules. Blood DNA is used as a control. A large scale assay is performed using the pool size(s) determined from the pilot assay. Recombinant products are amplified in a tertiary reaction using the two universal primers and the products dot blotted and hybridised with ASOs to map each crossover to a SNP interval.

2.2.11.1 Allele Specific Primers (ASPs)

Allele specific primers (ASPs) 17-19 bases in length, depending on the GC content of the sequence, were designed with the differential alleles at the 3` end of the oligo for amplifying DNAs. Each primer was optimised, in conjunction with a primer that will prime from both haplotypes (universal), using a donor homozygous for the allele and a donor homozygous for the opposite allele. Optimisation was initially performed at 56°C, 59°C, 62°C and 65°C for the annealing step of the PCR. Specificity usually improved with increasing temperature, while efficiency normally decreased. The primers used as forward primers in the PCR had to be both efficient and specific at

the same conditions as the reverse primers, to ensure they would be both efficient and specific at the same reaction conditions. Allele specific PCRs were set up as in section 2.2.2. Inputs and conditions are given in chapters 3, 4, 5 and 6.

2.2.11.2 Primary PCR

Allele Specific PCRs were set up as in section 2.2.2. Inputs and conditions are given in chapters 3, 4, 5 and 6.

2.2.11.3 Secondary PCR

Immediately following primary PCR was halted by addition of 40 µl of dilution buffer (10 mM Tris-HCl pH 7.5, 2 µg/ml salmon DNA). Arbitrarily, samples can be stored at 4°C before use in secondary PCR. 0.6 µl of diluted primary PCR product can be added in 9.4 µl per reaction for secondary PCR set up as in section 2.2.9.3. Conditions are given in Chapters 3, 4, 5 and 6.

2.2.11.4 Tertiary PCR

Once positive secondary PCR results had been confirmed, 0.5 µl from the secondary PCR was added in to a tertiary PCR. Universal tertiary primers were situated internally to the secondary primers. Conditions are given in Chapters 4, 5 and 6.

2.2.11.4 Mapping crossovers to SNP intervals

Tertiary PCR products were denatured, dotblotted and then sequentially hybridised for all informative SNPs in the target region (see section 2.2.8 and 2.2.9), first with one allele then with the opposite allele. Progenitor haplotype PCRs generated during phasing experiments were hybridised in parallel to establish the progenitor haplotypes (Figure 2.2). The same theoretical results have been used in this section to explain the procedure.

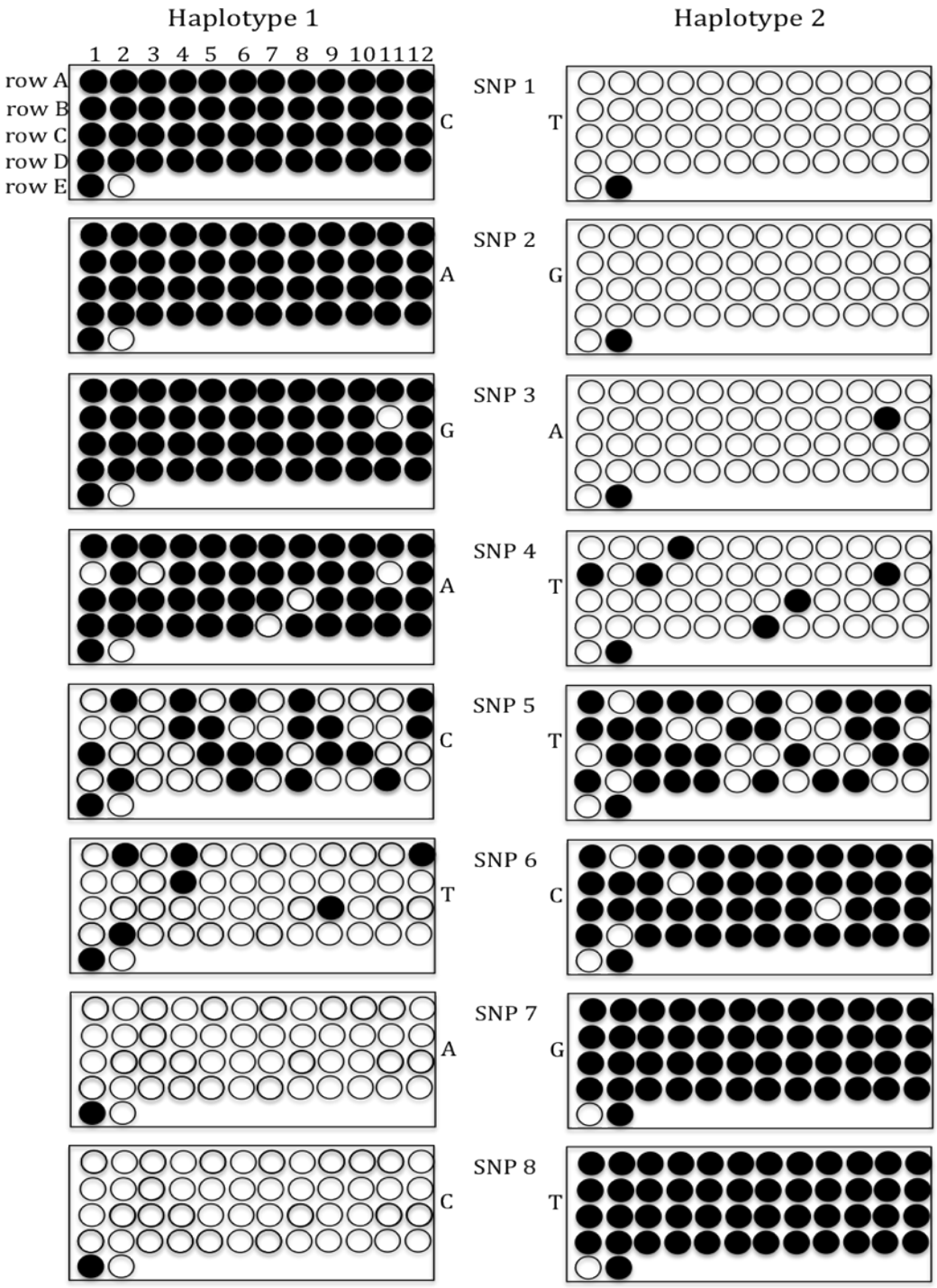


Figure 2.2 Typing of hypothetical informative SNPs in a recombination assay interval by dot blot hybridisation. Each rectangle represents a dot blot hybridisation for one allele of the SNP. The dots are arranged in rows A-E, top to bottom, and columns 1-12 left to right. The SNPs are numbered 1-8, with the allele hybridised shown next to each blot. Rows A-D represent 48 x reactions performed. Black circles show PCR product to which the probe allele has bound, white circles show PCR product to which the probe has not bound. SNPs arranged sequentially 5` to 3`. Row E 1 represents haplotype 1 progenitor PCR (green haplotype), row E 2 represents haplotype 2 progenitors PCR (red haplotype).

Each haplotype was assigned a colour, either red or green. In this theoretical case haplotype 1 was green and haplotype 2 was red. Therefore this assay was detecting crossovers from green to red haplotypes. Every SNP was typed for each reaction, and the genotype of that SNP put into the corresponding column and colour coded for the progenitor haplotype it belonged to. The results of this can be seen in Figure 2.3.

Reaction	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8
Progenitor 1	C	A	G	A	C	T	A	G
Progenitor 2	T	G	A	T	T	C	G	T
A1	C	A	G	A	T	C	G	T
A2	C	A	G	A	C	T	G	T
A3	C	A	G	A	T	C	G	T
A4	C	A	G	M	M	M	G	T
A5	C	A	G	A	T	C	G	T
A6	C	A	G	A	C	C	G	T
A7	C	A	G	A	T	C	G	T
A8	C	A	G	A	C	C	G	T
A9	C	A	G	A	T	C	G	T
A10	C	A	G	A	T	C	G	T
A11	C	A	G	A	M	M	G	T
A12	C	A	G	A	C	T	G	T
B1	C	A	G	T	T	C	G	T
B2	C	A	G	A	T	C	G	T
B3	C	A	G	T	T	C	G	T
B4	C	A	G	A	C	T	G	T
B5	C	A	G	A	C	C	G	T
B6	C	A	G	A	T	C	G	T
B7	C	A	G	A	T	C	G	T
B8	C	A	G	A	C	C	G	T
B9	C	A	G	A	C	C	G	T
B10	C	A	G	A	T	C	G	T
B11	C	A	A	T	T	C	G	T
B12	C	A	G	A	C	C	G	T
C1	C	A	G	A	C	C	G	T
C2	C	A	G	A	T	C	G	T
C3	C	A	G	A	T	C	G	T
C4	C	A	G	A	T	C	G	T
C5	C	A	G	A	M	C	G	T
C6	C	A	G	A	C	C	G	T
C7	C	A	G	A	C	C	G	T
C8	C	A	G	T	T	C	G	T
C9	C	A	G	A	C	T	G	T
C10	C	A	G	A	C	C	G	T
C11	C	A	G	A	T	C	G	T
C12	C	A	G	A	T	C	G	T

D1	C	A	G	A	T	C	G	T
D2	C	A	G	A	C	T	G	T
D3	C	A	G	A	M	C	G	T
D4	C	A	G	A	T	C	G	T
D5	C	A	G	A	T	C	G	T
D6	C	A	G	A	C	C	G	T
D7	C	A	G	T	T	C	G	T
D8	C	A	G	A	C	C	G	T
D9	C	A	G	A	T	C	G	T
D10	C	A	G	A	T	C	G	T
D11	C	A	G	A	C	C	G	T
D12	C	A	G	A	T	C	G	TD

Crossovers sorted and mixtures resolved

Reaction	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8
Progenitor 1	C	A	G	A	C	T	A	G
Progenitor 2	T	G	A	T	T	C	G	T
B11	C	A	A	T	T	C	G	T
A4a	C	A	G	T	T	C	G	T
B1	C	A	G	T	T	C	G	T
B3	C	A	G	T	T	C	G	T
C8	C	A	G	T	T	C	G	T
D7	C	A	G	T	T	C	G	T
A1	C	A	G	A	T	C	G	T
A3	C	A	G	A	T	C	G	T
A5	C	A	G	A	T	C	G	T
A7	C	A	G	A	T	C	G	T
A9	C	A	G	A	T	C	G	T
A10	C	A	G	A	T	C	G	T
A11a	C	A	G	A	T	C	G	T
B2	C	A	G	A	T	C	G	T
B6	C	A	G	A	T	C	G	T
B7	C	A	G	A	T	C	G	T
B10	C	A	G	A	T	C	G	T
C2	C	A	G	A	T	C	G	T
C3	C	A	G	A	T	C	G	T
C4	C	A	G	A	T	C	G	T
C5a	C	A	G	A	T	C	G	T
C11	C	A	G	A	T	C	G	T
C12	C	A	G	A	T	C	G	T
D1	C	A	G	A	T	C	G	T
D3a	C	A	G	A	T	C	G	T
D4	C	A	G	A	T	C	G	T
D5	C	A	G	A	T	C	G	T
D9	C	A	G	A	T	C	G	T
D10	C	A	G	A	T	C	G	T
D12	C	A	G	A	T	C	G	T
A6	C	A	G	A	C	C	G	T

A8	C	A	G	A	C	C	G	T
B5	C	A	G	A	C	C	G	T
B8	C	A	G	A	C	C	G	T
B9	C	A	G	A	C	C	G	T
B12	C	A	G	A	C	C	G	T
C1	C	A	G	A	C	C	G	T
C5b	C	A	G	A	C	C	G	T
C6	C	A	G	A	C	C	G	T
C7	C	A	G	A	C	C	G	T
C10	C	A	G	A	C	C	G	T
D3b	C	A	G	A	C	C	G	T
D6	C	A	G	A	C	C	G	T
D8	C	A	G	A	C	C	G	T
D11	C	A	G	A	C	C	G	T
A2	C	A	G	A	C	T	G	T
A4b	C	A	G	A	C	T	G	T
A11b	C	A	G	A	C	T	G	T
A12	C	A	G	A	C	T	G	T
B4	C	A	G	A	C	T	G	T
C9	C	A	G	A	C	T	G	T
D2	C	A	G	A	C	T	G	T

crossovers	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8
No. positive	0	1	5	24	15	7	0	
No. negative	80	79	75	55	63	73	80	
No.unscorable	0	0	0	1	2	0	0	
No.of Poisson corrected crossovers	0	1	5	28	17	7	0	

Figure 2.3 Hypothetical crossover mapping results. Progenitor haplotypes were colour coded either red or blue. Each SNP was typed and assigned to a haplotype. SNPs belonging to red haplotype were coloured red and those belonging to the blue haplotype were coloured blue. SNPs that showed mixed results for a particular reaction, ie. both alleles at a SNP site showed positive hybridisation, are coded as M (mixed). A crossover has occurred between the marker SNPs that change from the green to red haplotype. In the case of mixed sites the crossovers needed to be resolved, so that one was counted before the mixed sites and one after. If more than one mixture was present, as in the case of reaction A4, still only 2 crossovers could definitely be accounted for, one between SNP3 and SNP 4 and one between SNP6 and SNP7. Any crossovers occuring between SNP4 and SNP6 would be masked by the other crossovers. This was corrected when estimating crossover rates (see below). The resolved mixed crossovers, as in the case of A4, were called A4a and A4b. Once all of the mixtures were resolved and the crossovers sorted into order, the number of crossovers in each SNP interval was counted. The numbers of negative, positive and unscorable reactions for each SNP interval are shown at the bottom of the table with unscorable refering to any interval that were between two mixed sites. These sites, such as interval SNP4-SNP5 in sample A4, cannot be counted as either negative or positive for a crossover occurring as any crossover would be masked by the crossovers in both the preceding

and following SNP intervals. These numbers were used to estimate the Poisson corrected numbers of crossovers.

Once the mixtures have been resolved and sorted, maximum likelihood methods were used to estimate mean frequency of crossovers in each interval, taking into account the number of input molecules. Where more than one mixed site was observed, the intervening intervals were excluded from the analysis, because a crossover in these intervals could not be discounted. For each interval, the Poisson estimate of recombination frequency was multiplied by the number of input molecules screened in the interval, to yield the Poisson corrected number of crossovers. Once the Poisson corrected numbers of crossovers was known, it could be used to calculate the recombination activity for each SNP interval in cM/Mb. This was done by initially calculating the recombination activity in cM by dividing the number of Poisson corrected crossovers by the total number of molecules screened and multiplying the answer by 100. The activity in cM was converted to cM/Mb by dividing the recombination fraction in cM by the length of the SNP interval in base pairs and multiplying by 1 million.

CHAPTER 3:
HIGH-RESOLUTION LINKAGE DISEQUILIBRIUM ANALYSIS OF PUTATIVE HOTSPOTS CONTAINING A MOTIF-DISRUPTING SNP

3.1 INTRODUCTION

Donnelly and colleagues have identified more than 30,000 putative hotspots from localised linkage disequilibrium (LD) breakdown, and carried out an exhaustive search of short (5- to 9-mer) motifs for enrichment in those historical hot spots (Myers *et al.,* 2005). They identified two motifs, CCTCCCT and CCCCACCCC, which were strongly overrepresented in a small fraction (~10%) of human hot spots (Myers *et al.,* 2005). The first of these motifs, CCTCCCT, is found in THE1A/B retro-transposons and is over-represented (five to sixfold) in THE1A/B elements within the hotspots (Myers *et al.*, 2005). Previous polymorphic hotspot studies have reported the role of these motifs in hotspot activity where single nucleotide variants disrupting the centrally located motifs CCTCCCT and CCCCACCCC reduce crossover activity in *cis* at hotspots DNA2 (Jeffreys and Neumann, 2002) and NID1 (Jeffreys and Neumann, 2005) respectively. Additionally, Myers et al. (2008) used the power of HapMap Phase 2 data (22,699 autosomal and 608 chromosome X hotspots mapped to 5 kb resolution), to identify classes of repeat elements that are over-represented in hotspots. According to low-resolution HapMap Phase II data, Donnelly and colleagues drew up a shortlist of the four best candidate LD hotspots based on having a motif with a disrupting SNP and a good historical activity (personal communication with A.J Jeffreys, University of Leicester, UK). Subsequently, they searched for motifs that are independently associated with hotspots on multiple repeat-family backgrounds. This approach revealed a common 13-bp motif CCNCCNTNNCCNC, which is a 6-bp extension of the previous CCTCCCT sequence motif (Myers *et al.,* 2008). Additionally, this study showed how well the presence of the motif predicts hotspots; for example the presence of CCTCCCTNNCCAC in a THE1A background resulted in the detection of a hotspot 73% of the time, whereas in unique DNA it led to detection of hotspots only 10% of the time. The sequence context of the repeat element in THE1A leads to the presence of the most

recombinogenic nucleotides outside of the motif. Additionally, the DNA3 hotspot in the HLA class II region (Jeffreys *et al.*, 2001) and the MS32 hotspot (Jeffreys *et al.*, 1998b) contained the 13-mer motif within a few base pairs of the estimated centre.

The study of Myers *et al.* (2005) implicates the 13-mer motif in allelic crossover activity during meiosis. This raises the question of whether the nature of the motif offers any clues for understanding the molecular basis for recombination hotspots (Myers *et al.*, 2008). The four best candidate putative hotspots, their motif-disrupting SNPs, the 13-bp motif and their locations, are all listed in Table 3.1 for further analysis. The aim of this thesis is to identify hotspots with the 13-bp motif and which include a motif-disrupting SNP to test whether the disrupting allele influences the crossover frequency and distribution. In this chapter, four identified LD hotspots from low-resolution HapMap Phase II data were genotyped by high-resolution genotyping techniques in our European semen donor panel and tested to see if the motif-disrupting SNP lay within the motif at the centre of the hotspot by linkage disequilibrium (LD) and linkage disequilibrium unit (LDU) mapping.

LD Hotspots	SNP	Motif	Location
DA	rs7036542	MCNCCNTNNCCNC	Chr 9 p22.9
DB	rs6035457	CCWCCNTNNCCNC	Chr 20 p19.70
DC	rs6578087	CMNCCNTNNCCNC	Chr 8 q24.3
DD	rs1982437	YCNCCNTNNCCNC	Chr 11 q21

Table 3.1 Candidate SNPs and the locations of these SNPs within the motif. The table shows the four candidate SNPs in hotspots identified by Donnelly (personal communication with A.J Jeffreys, University of Leicester, UK) and their chromosomal location. Also, the table shows the single base changes (labelled in red) inside the motif (M: C, A; W: T, A; Y: C, T).

3.2 FOUR SELECTED LD HOTSPOTS

The first LD hotspot to be studied was Hotspot DA. This hotspot contains an intergenic SNP (rs7036542), which is located on the short arm of chromosome 9 in its sub terminal region (Figure 3.1a), and within a well-localised putative hotspot within a THE1B element. This SNP changes the 13-bp motif from CCNCCNTNNCCNC to ACNCCNTNNCCNC. The ancestral SNP allele is C (or complementary G). In addition, downstream of the CCTCCCT motif was a sequence, CCTCCCTAGCCT that is overrepresented in the hotspot THEIB elements. This suggests that the presence of this downstream sequence is also helpful in promoting hotspots. Interestingly, previous evidence from genome-wide statistical comparisons shows that ACNCCNTNNCCNC is associated with a lowered hotspot activity (Myers *et al.,* 2005). The ancestral SNP allele is C, and the allele frequency is 0.58/0.34 (C/A) in CEU population.

The second LD hotspot, Hotspot DB, has an intergenic SNP (rs6035457) that is located on the short arm of chromosome 20 (Figure 3.1b). This SNP is outside of any repeat elements, and changes the sequence CCTCCNTNNCCNC to CCACCNTNNCCNC. The analysis of Donnelly and colleagues (personal communication with A.J Jeffreys, University of Leicester, UK) did not give an especially precise localisation for this hotspot, but the hotspot did localise exactly to the SNP location based on HapMap Phase II LD patterns. The downstream sequence CCTCCT..CCAC is found 3.5 times more frequently than expected by chance in narrow hotspots. The ancestral SNP allele is T, and the allele frequency is 0.36/0.64 (T/A) in CEU population.

Hotspot DC is the third LD hotspot, and SNP rs6578087 is a very well localised intronic SNP that locates to the long arm of chromosome 8 (Figure 3.1c). This SNP is in a LINE/L1 element and changes CCNCCNTNNCCNC to CANCCNTNNCCNC. This change shows strong genome-wide evidence of less activity (personal communication with A.J Jeffreys, University of Leicester, UK). The downstream sequence CCTCCCTGACCCC is two times enhanced in well-localised hotspots. The alelle frequency in the CEU population is 0.633/0.367 (C/A, C is the ancestral allele)

The last LD hotspot, Hotspot DD, has an intergenic SNP (rs1982437) on the long arm of chromosome 11 (Figure 3.1d). This SNP is in a SINE/MIR retrotransposon element and changes CCNCCNTNNCCNC to TCNCCNTNNCCNC. The downstream sequence CCTCCCTGACCTC is 2.1 times enhanced in well-localised hotspots. The allele frequency of the SNP in the CEU population is 0.701/0.299 (C/T).

3.2.1 SNP Discovery and Annotation

For each putative hotspot containing a SNP within a CCNCCNTNNCCNC motif, a 15 kb interval of DNA sequence centring on the target SNP (the hotspot) was downloaded from ENSEMBL (www.ensembl.org). The sequence was annotated to include all SNPs and repeat sequences (LINEs, Alus etc) using information gained from Phase II HapMap data, dbSNP data and Repeat Masker (Appendix I). This was carried out so that Alu elements, with their high genomic copy number, could be taken into consideration while designing PCR primers and Allele Specific Oligos (ASOs). In summary, 88, 77, 70 and 69 SNPs were characterised using HapMap Phase II and dbSNP data at 15 kb intervals for LD Hotspots DA, DB, DC and DD respectively. The sequences are shown in Appendix 3.1.

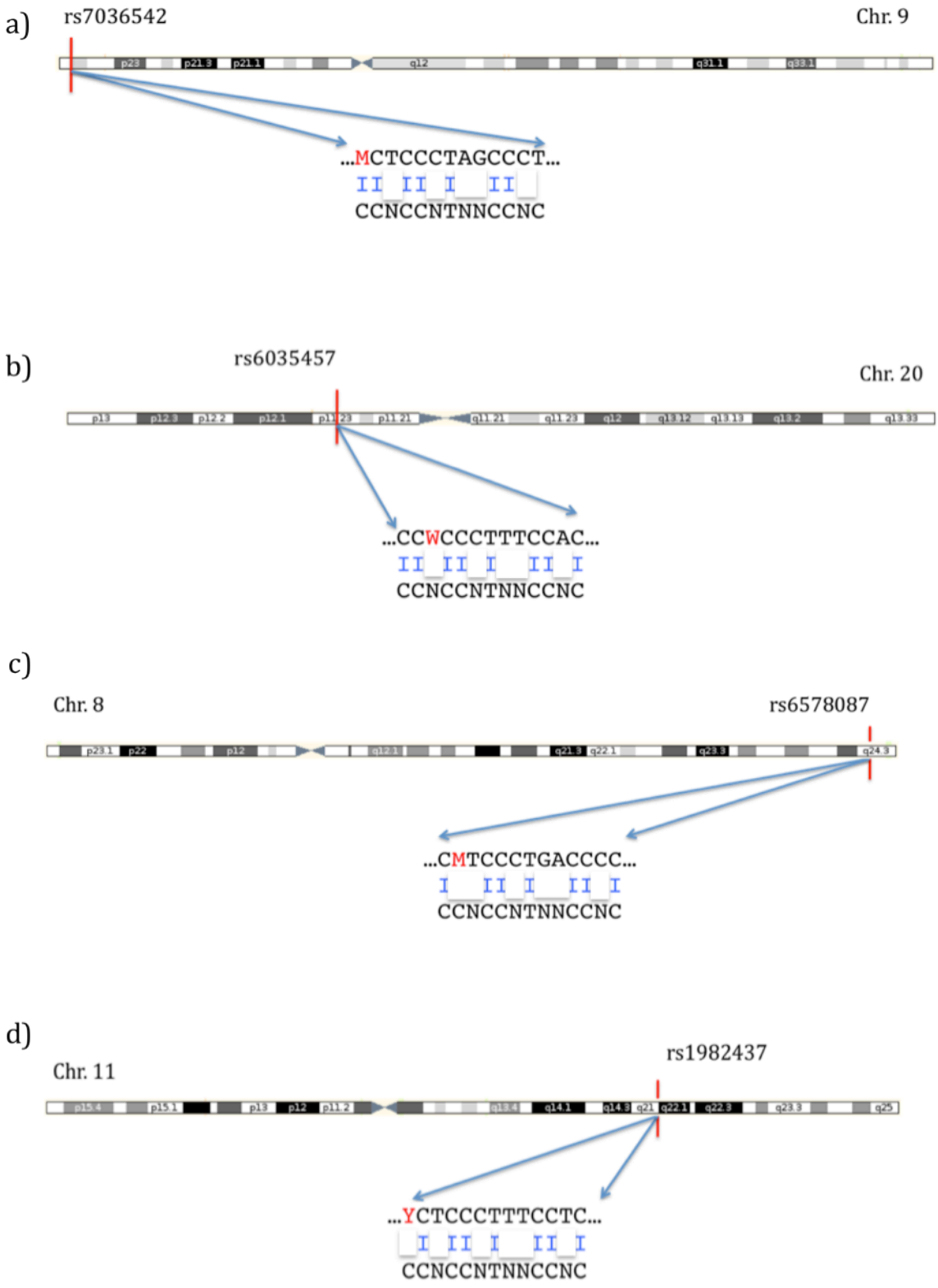


Figure 3.1 shows the location of disrupting SNPs on the chromosomes for each target. Additionally, the figure shows the location of the SNPs (red) in the sequence and the matches within the 13-bp sequence motif. (W: A/T, M: C/A, Y: C/T) (Adapted from www.ensembl.org)

3.2.2. Typing SNPs In Semen Donors by Allele Specific Oligo (ASO) Hybridisation

All identified SNPs were genotyped using Allele Specific Oligo (ASO) hybridisation (see Materials and Methods) (Figure 3.2) and these genotypes were used for LD and LDU mapping, and coalescent analysis to test for the presence of historical hotspots. The 15-kb interval of DNA sequence that centred on each LD hotspot was divided into 3 overlapping 4-6 kb amplicons. These amplicons were amplified using nested PCR on whole-genome amplified (MDA) sperm DNA; MDA DNAs were used to save genomic DNA.

Initially, sperm DNA from 92 semen donors was whole-genome amplified (PCR conditions and primers are shown in Appendix II). The PCR products used for SNP genotyping were used in ASO hybridization (all ASOs are shown in Appendix III). Depending on the number of SNPs in each amplicon, 2-6 replicates of each amplicon for donors 1-96 (there are three donors who are duplicated in the panel: donor 19 and donor 49, donor 30 and donor 33, donor 31 and donor 86 and donor 50 and donor 95) were produced, with a single membrane being hybridised sequentially, first with one allele ASO and then with the opposite allele ASO. The resulting genotypes were normally unambiguous, although if the PCR products were weak then it was necessary to assume an unknown result.

In total, 39 SNPs were typed in 92 donors for Hotspot DA. A further 49 SNPs were either non-validated by HapMap or did not have a frequency in the CEU population and were therefore discarded. The data from all 38 SNPs typed is shown in Table 3.2 (donors 1-50) and Table 3.3 (donors 50-96). For the other LD hotspots, a total of 30, 31 and 23 SNPs were typed in 93 donors out of 77, 70 and 69 annotated SNPs for Hotspots DB, Hotspot DC and Hotspot DD respectively. Tables 3.4 - 3.9 show genotyped SNPs in our donor panel for Hotspots DB, DC and DD. These genotypes were used for LD, LDU and coalescent analysis.

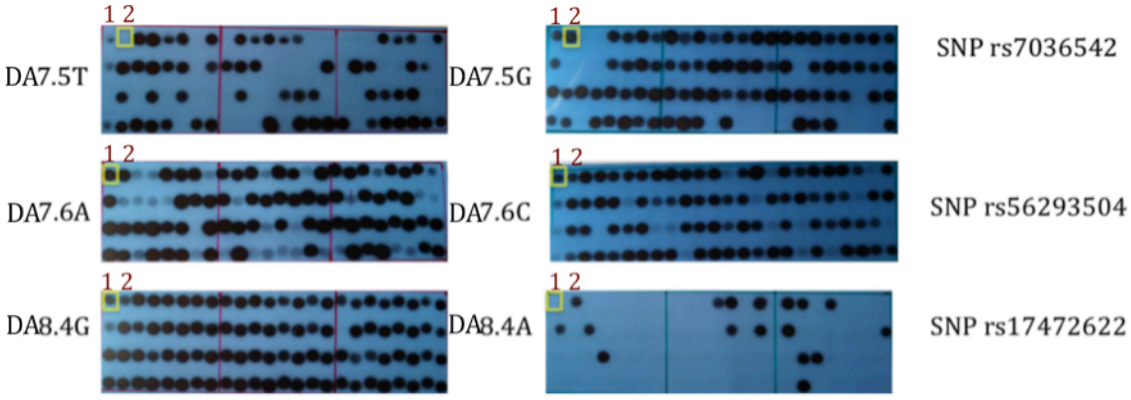


Figure 3.2 An example of SNP genotyping from ASO hybridisation for Hotspot DA markers. PCR products from 92 men were loaded onto membranes. Blots are probed with radiolabelled ASOs, one membrane with one allele and the other membrane with the other allele. Genotypes can thus be determined. For example, donor 2 for SNP rs7036542 (DA7.5T/G) is homozygous (G/G), donor 1 for SNP rs56293504 (DA7.6A/C) is heterozygous (A/C) and donor 1 for SNP rs17472622 (DA8.4G/A) is homozygote (G/G). (Donors 1 and 2 are coloured in red)

SNP	location	donor 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
DA2.4-/+	2368	-	-	-	H	-	-	-	-	-	-	H	-	-	-	H	H	-	-	-	-	-	-	-	H	-	-	-	-	-	H	-	H	-	-	-	-	-	-	-	-	H	+	H	-	-	-	-	?	-
DA2.6G/C	2593	H	G	G	G	G	H	H	H	C	G	G	H	H	G	G	H	H	H	C	C	H	H	H	G	C	H	C	C	H	H	H	H	H	H	C	H	C	G	H	G	G	H	G	G	H	C	G
DA3.5T/G	3489	T	T	T	H	H	T	T	T	T	T	H	T	T	T	H	H	T	T	T	T	T	T	H	T	T	T	T	T	H	T	H	T	T	T	T	T	T	T	H	H	H	G	H	T	T	T	T	T	T
DA3.7T/G	3667	H	T	T	T	T	H	H	H	G	T	T	H	H	H	T	H	H	H	G	G	H	H	H	T	G	G	G	G	H	H	H	H	H	H	G	H	G	T	H	T	T	H	T	H	H	T	T	H	G	T
DA4.0C/T	4005	C	C	C	H	C	C	C	C	C	C	H	C	C	C	H	H	C	C	C	C	C	C	C	H	C	C	C	C	C	H	C	H	C	C	C	C	C	C	C	C	H	T	H	C	C	C	C	?	C
DA4.0aG/A	4044	H	G	G	G	G	H	H	H	A	G	G	H	H	G	G	H	H	H	A	A	H	H	H	G	A	H	A	A	H	H	H	H	H	H	A	H	A	G	H	G	G	H	G	H	H	G	G	H	A	G
DA4.8A/C	4777	A	A	A	H	H	A	A	A	A	A	H	A	A	A	H	H	A	A	A	H	A	A	H	A	A	A	A	A	H	A	H	A	A	A	A	A	A	A	H	H	H	C	H	A	H	A	A	A	A
DA5.0T/C	5112	H	T	T	T	T	H	H	H	H	T	T	H	H	T	T	H	H	H	C	C	H	H	H	T	C	C	C	C	H	H	H	H	H	H	C	T	C	T	H	T	T	H	T	T	H	C	T
DA5.3A/T	5368	A	A	A	H	A	A	A	A	A	A	H	A	A	A	H	H	A	A	A	A	A	A	H	A	A	A	A	A	H	A	H	A	A	A	A	A	A	A	H	H	H	T	H	A	A	A	A	A	A
DA6.0A/G	6057	A	A	A	H	H	A	A	A	A	A	H	A	A	A	H	H	A	A	A	A	A	A	H	A	A	A	A	A	H	A	H	A	A	A	A	A	A	A	H	H	H	G	H	A	A	A	A	A	A
DA6.1C/T	6139	H	T	H	C	C	C	C	C	C	C	H	C	H	T	H	C	H	H	C	C	C	H	C	H	C	C	C	C	C	C	C	H	C	H	C	C	C	C	T	H	H	C	C	C	C	H	C	T	H	C	H
DA6.1aA/G	6171	H	A	A	A	A	A	H	H	G	A	A	H	H	A	A	H	H	H	G	G	H	H	H	A	G	G	G	G	H	H	A	H	H	H	G	H	G	A	H	A	A	H	A	H	A	A	A	H	G	A
DA6.3T/C	6335	H	T	T	H	H	T	H	H	C	T	H	H	T	T	H	C	H	H	H	T	H	H	H	T	H	C	H	C	H	H	H	H	H	H	C	H	H	T	H	H	H	C	C	C	T	T	T	H	T
DA6.3aA/T	6352	A	A	A	H	H	A	H	?	T	A	H	H	A	A	H	T	H	H	H	A	H	H	H	A	?	T	H	T	H	H	H	H	H	H	T	H	H	A	H	H	H	T	T	T	A	A	A	A	H	A
DA7.0A/G	7025	H	G	H	A	A	H	A	A	A	A	H	A	H	G	H	A	H	H	A	A	A	H	A	H	A	A	A	A	A	A	H	H	A	H	H	A	A	A	G	H	H	A	A	A	A	G	A	G	H	A	H
DA7.0a-/+	7049	H	+	H	H	H	H	H	H	+	-	+	H	H	+	+	+	+	+	H	-	H	+	H	H	H	+	H	+	H	+	H	H	+	+	H	+	+	H	H	+	H	+	H	+	+	+	+	-	+	H	H	H
DA7.0bT/G	7140	H	T	T	T	H	T	G	H	G	T	H	G	H	T	H	G	T	H	H	G	H	H	H	T	G	G	G	G	G	H	T	H	H	H	G	H	G	T	H	H	H	H	T	H	T	H	T	H	H	T
DA7.4A/C	7481	H	A	A	H	H	H	H	H	A	H	A	A	H	A	A	H	A	H	A	H	H	C	H	H	H	H	H	A	H	H	C	H	A	A	H	A	H	A	H	A	H	H	C	A	H	A	A	A	A	H	A
DA7.5T/G	7501	H	T	H	G	G	H	H	G	G	G	H	G	H	T	H	G	H	H	G	G	G	H	G	H	G	G	G	G	G	G	H	H	G	H	H	G	H	T	H	H	G	G	G	T	T	H	T	T	G	T
DA7.5aT/C	7542	C	C	C	C	C	C	H	H	C	C	C	C	C	C	C	C	C	H	C	H	C	C	C	H	C	C	H	H	H	C	H	C	C	H	C	C	C	C	C	T	C	C	C	C	C	H	H	C
DA7.6A/C	7631	H	C	H	H	H	H	A	C	H	H	H	A	H	C	C	H	H	H	H	A	H	H	H	C	A	H	H	C	A	H	H	H	H	H	H	A	C	H	C	C	C	C	H	C	H	C	H	H	C
DA8.2G/A	8229	G	G	G	G	G	H	G	G	G	G	G	H	G	G	H	G	H	G	H	G	G	G	G	H	G	H	H	H	G	G	H	G	H	G	G	G	G	G	G	G	H	G	H	G	G	G	H	H	G	H	G
DA8.4G/A	8413	G	G	G	G	G	G	H	H	H	G	G	G	H	H	G	G	G	G	H	H	A	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G
DA8.5C/G	8590	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DA9.5C/T	9580	H	H	H	C	C	H	C	C	C	C	C	C	C	H	C	C	H	H	C	C	C	H	C	H	C	C	C	C	C	C	H	C	C	C	C	C	C	T	C	C	C	C	C	H	T	C	H	H	C	T
DA10.0G/A	10126	G	G	H	A	G	G	H	H	H	A	G	H	G	G	G	A	G	H	G	G	G	H	G	G	G	G	H	G	H	H	H	G	A	H	G	H	A	H	G	A	G	G	G	A	H	G	G	G	G	G
DA10.2C/T	10203	C	H	C	C	C	H	H	H	H	C	H	H	T	H	H	C	H	C	T	H	T	C	C	H	C	H	H	H	C	C	H	C	C	C	T	C	H	C	C	C	H	H	C	T	C	H	C	C	T	C
DA10.6G/A	10563	G	G	H	A	G	G	H	H	H	A	G	H	G	G	G	A	G	H	G	G	G	H	G	G	G	G	H	G	H	H	H	G	A	H	G	H	A	H	G	A	G	G	G	A	H	G	G	G	G	G
DA10.6aA/C	10604	C	C	C	C	A	C	C	C	C	C	H	C	C	C	H	C	C	C	C	H	C	C	A	C	A	H	C	H	H	H	C	C	C	C	H	C	H	C	C	C	A	H	C	C	C	H	C	H	C	C
DA10.6bC/A	10614	C	H	C	C	C	H	H	H	H	C	H	H	A	H	H	C	H	C	A	H	A	C	C	H	C	H	H	H	C	C	H	C	C	A	C	C	C	C	C	A	C	H	C	C	C	H	H	C	A	C
DA11.5C/T	11543	C	C	H	T	C	C	H	H	H	T	C	H	C	C	C	T	C	H	C	C	C	H	C	C	C	C	H	C	H	H	C	T	H	C	H	T	H	C	T	C	C	C	T	H	C	C	C	C	C	C
DA12.0C/T	12063	H	H	H	C	C	H	C	C	C	C	C	C	C	H	C	C	H	H	C	C	C	H	C	H	C	C	C	C	C	C	H	C	C	C	C	C	H	T	C	C	C	C	C	H	T	C	H	T	C	T
DA13.4A/G	13364	A	A	H	G	G	A	H	?	H	G	H	H	A	A	H	G	A	H	A	H	A	H	G	A	G	H	H	H	G	G	A	G	G	A	G	G	H	A	G	A	G	H	G	H	A	H	A	A	A	A
DA13.4a-/+	13384	+	H	H	-	-	H	-	-	-	-	-	-	-	H	-	-	H	H	-	-	-	H	-	H	-	-	-	-	-	-	+	-	-	-	-	-	H	+	-	H	H	-	+
DA13.4bC/T	13388	T	H	H	C	C	H	C	C	C	C	C	C	C	H	C	C	H	H	C	C	C	H	C	H	C	C	C	C	C	C	H	C	C	C	C	C	C	T	C	C	C	C	C	H	T	C	H	H	C	T
DA13.5T/C	13559	T	T	H	C	C	T	H	H	H	C	H	H	T	T	H	C	T	H	T	H	T	H	C	T	C	H	H	H	C	C	T	C	C	T	C	C	C	T	C	T	C	H	C	H	T	H	T	H	T	T
DA13.8aG/A	13798	H	H	A	A	A	H	H	H	H	A	H	H	G	H	H	A	H	A	G	H	G	A	A	H	A	H	H	H	A	A	H	A	A	H	A	A	G	A	A	H	A	A	G	A	H	A	A	A	H	H	H	G	A
DA13.8bC/T	13858	C	H	C	C	C	H	C	C	C	C	H	H	H	C	H	C	H	C	H	C	C	C	C	H	C	H	H	H	C	C	H	C	C	H	C	C	H	C	C	C	C	C	T	C	H	C	C	C	H	H	C	H	C
DA14.4T/C	14479	T	T	T	T	C	T	T	T	T	T	H	T	T	T	H	T	T	T	T	H	T	T	C	T	C	H	T	H	H	H	H	T	T	H	T	H	T	H	T	T	T	T	T	?	H	T	T	T	H	T	T	T	T

Table 3.2 SNP genotypes from semen donors 1-50 for the 39 SNPs for Hotspot DA. SNP names are listed down the left and donor number across the top. Each genotype is depicted as either a homozygote (A, C, G, T, +, -) or heterozygote (H). Motif disrupting SNP DA7.5T/G is highlighted in red.

SNP	location	donor 5 1	5 2	5 3	5 4	5 5	5 6	5 7	5 8	5 9	6 0	6 1	6 2	6 3	6 4	6 5	6 6	6 7	6 8	6 9	7 0	7 1	7 2	7 3	7 4	7 5	7 6	7 7	7 8	7 9	8 0	8 1	8 2	8 3	8 4	8 5	8 6	8 7	8 8	8 9	9 0	9 1	9 2	9 3	9 4	9 5	9 6
DA2.4-/+	2368	-	-	-	-	H	H	H	-	-	-	-	-	-	-	-	-	-	H	-	-	-	-	-	-	H	H	-	+	-	-	H	-	H	-	-	H	-	-	H	-	-	-	-	-	-	-
DA2.6G/C	2593	C	H	H	H	H	H	H	H	H	G	G	G	C	H	H	C	H	G	G	H	C	H	H	G	G	H	H	H	G	G	H	H	H	G	G	H	H	C	H	H	G	G	G	C	C	G	C
DA3.5T/G	3489	T	T	T	T	H	H	H	T	T	T	T	T	T	T	T	T	H	T	T	T	T	T	T	H	H	T	G	H	T	H	T	H	T	T	H	T	T	H	T	T	T	T	T	T	T	T
DA3.7T/G	3667	G	H	H	H	H	H	H	H	H	T	T	T	G	H	H	G	H	T	T	H	G	H	H	T	T	H	H	T	T	H	H	H	T	T	H	H	G	H	H	T	T	T	G	G	T	G
DA4.0C/T	4005	C	C	C	C	H	H	H	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	H	H	C	T	C	C	H	C	H	C	C	H	C	C	H	C	C	C	C	C	C	C
DA4.0aG/A	4044	A	H	H	H	H	H	H	H	H	G	G	G	A	H	H	A	H	G	G	H	A	H	H	G	G	H	H	G	G	H	H	H	G	G	H	H	A	H	H	G	G	G	A	A	G	A
DA4.8A/C	4777	A	A	A	A	H	H	H	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	H	H	A	C	H	A	H	A	H	A	A	H	A	A	A	H	A	A	A	A	A	A	A
DA5.0T/C	5112	H	H	H	H	H	H	H	H	H	T	T	T	C	H	H	C	H	T	T	H	C	H	H	T	T	H	H	H	T	T	H	H	H	C	H	H	T	T	T	H	H	T	H	T	H	H
DA5.3A/T	5368	A	A	A	A	H	H	H	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	H	H	A	T	H	A	H	A	H	A	A	H	A	A	H	A	A	A	A	A	A	A	A
DA6.0A/G	6057	A	A	A	A	H	H	H	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	H	H	A	G	H	A	H	A	H	A	A	H	A	A	H	A	A	A	A	A	A	A	A
DA6.1C/T	6139	C	H	H	H	C	C	C	C	H	H	T	T	C	C	H	C	C	H	T	C	C	C	H	C	C	C	H	C	H	C	C	H	C	C	H	C	C	H	C	T	T	T	C	C	H	C
DA6.1aA/G	6171	H	H	H	H	H	H	H	A	H	A	A	A	G	A	H	G	H	A	A	H	G	H	H	A	A	H	H	A	A	H	H	H	A	H	H	A	G	H	H	A	A	A	G	G	A	H
DA6.3T/C	6335	H	H	H	H	H	C	C	T	H	T	T	T	H	T	H	H	T	H	T	H	H	T	T	T	H	C	H	C	H	T	H	H	H	H	H	H	C	H	C	T	T	T	H	C	T	H
DA6.3aA/T	6352	H	H	H	H	H	T	H	A	H	A	A	A	H	A	H	H	A	H	A	H	H	A	A	A	H	T	H	T	H	A	H	H	H	H	H	H	T	H	T	A	A	A	H	T	A	H
DA7.0A/G	7025	H	H	H	H	A	A	A	H	H	H	G	G	A	H	H	A	A	H	G	A	A	A	H	A	A	A	H	A	H	A	H	A	A	H	H	A	H	A	G	G	G	G	A	A	H	H
DA7.0a-/+	7049	+	+	+	+	H	+	+	H	+	H	+	+	H	H	+	H	-	+	+	H	H	-	H	-	H	+	+	+	+	-	H	+	H	H	+	+	+	+	+	+	+	H	+	H	+
DA7.0bT/G	7140	H	H	H	H	H	G	H	T	H	T	T	T	G	T	H	G	H	T	T	G	G	H	H	T	T	H	H	T	H	G	H	H	T	H	H	T	G	H	H	T	T	T	G	G	T	H
DA7.4A/C	7481	H	A	A	A	H	C	A	A	A	A	A	A	H	A	A	H	H	A	A	H	A	A	H	A	H	H	A	C	H	H	C	A	H	H	A	A	H	A	A	A	A	A	H	A	A	A
DA7.5T/G	7501	H	H	H	H	H	G	G	H	H	H	T	T	H	H	H	G	G	H	T	G	H	G	H	H	H	G	H	G	H	G	G	H	T	G	H	H	G	H	G	T	T	T	G	H	T	H
DA7.5aT/C	7542	C	C	C	C	H	C	C	C	C	C	C	C	H	C	C	C	C	H	C	H	H	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	H	H	C	C	C	C	H	C	H
DA7.6A/C	7631	H	C	H	C	H	A	H	C	H	C	C	C	C	A	H	C	H	H	H	C	H	A	C	H	A	H	A	H	C	C	A	A	H	H	A	H	H	H	H	A	C	C	C	A	A	C	H
DA8.2G/A	8229	G	H	G	G	G	G	H	G	G	H	H	G	H	G	H	G	H	G	G	H	G	H	G	H	G	G	G	H	H	H	H	G	H	G	G	H	G	H	G	H	G	G	H	G	G	G	G
DA8.4G/A	8413	G	G	G	H	G	G	G	H	H	G	G	G	G	G	G	G	G	G	G	G	G	H	G	H	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G
DA8.5C/G	8590	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H
DA9.5C/T	9580	H	C	C	C	C	C	C	H	H	C	C	H	C	H	H	H	C	H	H	C	C	C	C	C	C	C	C	H	C	C	H	H	C	H	H	C	H	C	H	H	C	C	C	T	H
DA10.0G/A	10126	H	G	H	G	G	H	G	G	H	H	G	G	H	H	G	G	G	G	G	G	H	G	G	H	G	G	G	G	G	A	G	G	H	G	H	G	A	G	G	G	G	H	H	H	G	H
DA10.2C/T	10203	C	T	H	T	H	C	T	H	C	H	T	H	H	C	H	C	H	C	H	H	C	T	H	H	C	T	C	C	H	C	H	C	C	H	H	H	H	H	C	C	C	C	C	C	C	C
DA10.6G/A	10563	H	G	H	G	G	H	G	G	H	?	G	G	H	H	G	G	G	G	G	G	H	?	G	H	G	G	G	G	G	A	G	G	H	G	H	G	A	G	G	G	G	H	H	H	G	H
DA10.6aA/C	10604	C	C	C	C	H	H	C	C	C	C	C	C	C	C	C	H	H	H	C	H	H	C	H	C	A	H	C	H	C	C	C	H	C	H	C	C	C	H	H	C	C	C	C	H	C	C
DA10.6bC/A	10614	C	A	H	A	H	C	A	H	C	?	A	H	H	C	H	C	H	C	H	H	C	?	H	H	C	H	A	H	H	C	A	C	C	H	C	H	C	C	H	H	H	H	H	C	C	C
DA11.5C/T	11543	H	C	H	C	C	H	C	C	H	H	C	C	H	H	C	C	C	C	C	C	H	C	C	H	C	C	C	C	C	T	C	C	H	C	H	C	T	C	C	C	C	H	H	H	C	H
DA12.0C/T	12063	H	C	C	C	H	C	C	H	H	C	C	H	C	H	H	H	C	H	H	C	H	C	C	C	C	C	C	C	H	C	C	H	H	H	H	H	C	H	C	H	H	C	C	C	T	H
DA13.4A/G	13364	H	A	H	A	A	G	A	A	H	H	A	A	H	H	A	A	H	H	A	H	H	A	H	H	G	H	A	H	A	G	A	H	H	H	H	A	G	H	H	A	A	H	H	G	A	H
DA13.4a-/+	13384	H	-	-	-	-	-	-	-	H	H	-	-	H	-	H	H	H	-	H	H	-	-	-	-	-	-	-	-	-	H	-	-	H	H	-	H	-	H	-	H	H	-	-	-	+	H
DA13.4bC/T	13388	H	C	C	C	C	C	C	H	H	C	C	H	C	H	H	H	C	H	H	C	C	C	C	C	C	C	C	C	H	C	C	H	H	C	H	H	C	H	C	H	H	C	C	C	T	H
DA13.5T/C	13559	H	T	H	T	H	C	T	T	H	?	T	T	H	H	T	H	H	H	T	H	C	T	H	H	C	H	T	H	T	C	T	H	H	H	H	T	C	H	H	T	T	H	H	C	T	H
DA13.8aG/A	13798	A	G	H	G	G	A	G	H	A	H	G	H	H	A	H	H	H	A	H	H	H	G	H	H	A	H	G	H	H	A	G	A	A	G	A	H	A	A	H	H	H	H	H	A	A	A
DA13.8bC/T	13858	C	T	H	H	H	C	H	C	C	?	T	H	H	C	H	C	H	C	H	H	C	?	H	C	C	H	T	H	H	C	H	C	C	H	C	H	C	C	H	H	H	H	C	C	C	C
DA14.4T/C	14479	T	T	T	T	T	H	T	T	T	T	T	T	T	H	T	T	H	H	T	H	T	T	H	T	C	H	T	H	H	T	T	T	H	T	T	T	T	T	H	H	T	T	T	T	H	T	T

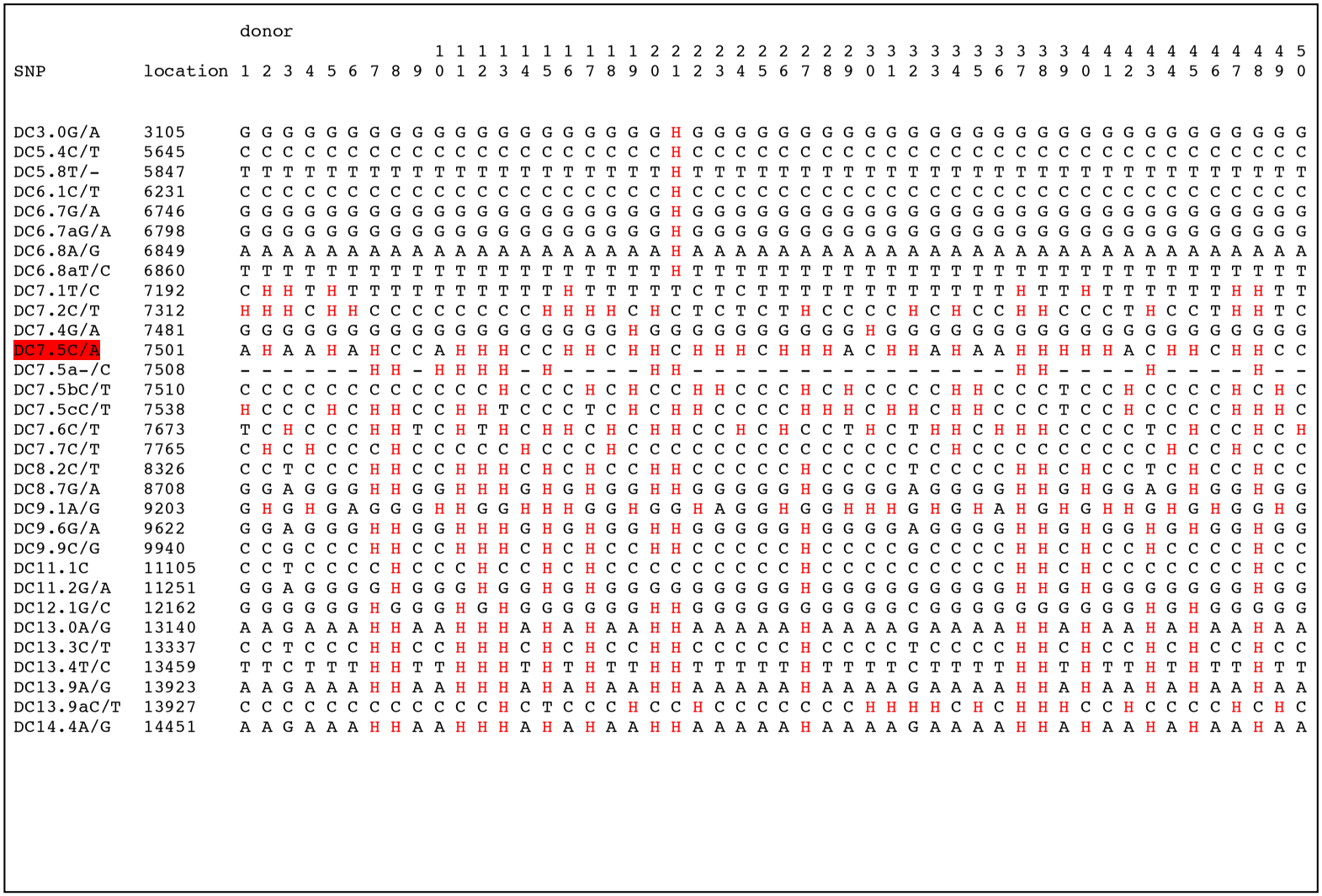
Table 3.3 SNP genotypes from semen donors 50-96 for the 39 SNPs for Hotspot DA. SNP names are listed down the left and donor number across the top. Each genotype is depicted as either a homozygote (A, C, G, T, +, -) or heterozygote (H). Motif disrupting SNP DA7.5T/G is highlighted in red.

SNP	location	donor 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
DB2.2T/C	2286	T	T	T	T	T	T	T	T	T	T	T	T	T	H	T	H	T	T	T	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	H	T	T	T	T	T	T	H	T	T	T	T	T	T	T	T
DB2.3G/C	2390	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	H	H	H	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DB2.4G/A	2448	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	H	H	H	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DB2.5T/C	2612	C	H	C	C	H	C	C	C	T	H	H	C	H	T	H	C	H	C	C	C	H	H	H	C	T	H	T	C	C	C	C	C	C	C	C	H	C	C	H	T	C	C	H	C	C	H	H	H	C	H
DB2.7T/C	2704	T	T	T	T	T	T	T	T	H	T	T	T	H	H	T	T	T	T	T	T	T	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
DB2.7aC/G	2747	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	H	H	H	C	H	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C
DB3.7A/G	3777	A	A	A	A	A	A	A	A	H	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
DB3.9G/C	3927	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	H	H	H	G	H	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G
DB4.1C/A	4152	C	C	C	C	C	C	C	C	C	H	C	C	C	H	H	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H
DB4.1bA/G	4191	A	A	A	A	A	A	A	A	H	A	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A	H	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H
DB4.2A/G	4222	G	G	G	G	H	G	G	G	H	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G
DB4.3C/G	4406	H	H	G	G	H	G	G	G	C	H	H	G	H	C	H	G	H	H	G	G	H	H	H	G	C	H	C	G	G	G	G	G	G	G	G	H	G	H	H	C	G	G	H	G	G	H	H	H	G	H
DB4.3aA/G	4413	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	H	H	H	A	H	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A
DB4.5A/G	4517	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	H	H	H	A	H	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A
DB5.4G/A	5448	H	H	A	A	H	A	A	A	G	A	H	H	H	H	H	A	H	H	A	A	H	A	H	A	H	A	H	H	G	A	A	A	A	A	A	A	H	A	H	A	G	A	A	H	H	H	H	H	A	A
DB6.6A/G	6689	G	H	G	G	H	G	G	G	A	G	H	G	H	A	H	G	H	H	G	G	G	G	G	G	G	H	A	G	G	G	G	G	G	G	G	H	G	H	G	H	G	G	H	H	H	G	H	H	G	H
DB7.0T/C	7058	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	H	C	C	C	C	C	C
DB7.0aC/A	7077	H	C	H	C	C	H	A	H	A	H	H	H	H	A	H	A	A	H	A	A	H	H	C	H	C	H	A	H	H	H	H	?	H	H	A	H	C	A	A	H	C	A	C	A	H	H	H	A	A	H
DB7.1G/T	7155	G	G	H	G	G	H	T	H	G	H	G	G	G	G	G	G	H	H	T	T	G	H	G	H	G	H	G	G	G	H	G	G	H	H	T	H	G	G	T	H	G	H	G	H	G	H	G	T	G
DB7.2T/C	7301	T	T	T	T	T	T	T	T	T	T	T	H	T	H	T	C	T	T	T	T	H	T	T	T	H	T	H	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	H	T	T	T	H	T	T	H
DB7.5T/A	7501	A	H	H	H	T	H	A	H	A	T	T	A	H	H	H	A	A	H	H	A	H	T	T	T	T	H	A	A	H	T	H	A	T	A	H	H	T	A	H	A	H	A	H	H	H	A	A	H	A
DB7.6C/T	7712	H	H	C	T	H	T	C	H	C	T	H	H	T	C	H	C	C	H	H	C	H	T	H	T	C	C	C	H	H	T	C	C	T	H	C	H	H	C	H	C	T	C	H	C	H	H	C	C	H	H
DB7.8T/C	7889	H	H	H	T	T	H	H	H	C	H	T	H	T	C	T	C	C	H	H	C	H	T	T	H	T	H	C	H	H	H	C	T	H	H	C	H	T	C	H	C	T	H	T	H	T	H	C	H	H	H
DB7.9T/A	7921	H	H	H	H	T	H	H	H	A	T	T	H	T	A	T	A	A	H	H	A	H	T	T	H	H	H	A	H	H	T	H	T	T	H	H	T	H	A	H	A	T	A	T	A	H	H	A	A	H	H
DB9.0C/T	9048	H	C	H	C	C	C	H	C	H	C	C	C	C	H	C	C	H	C	H	H	C	C	C	C	C	C	T	C	C	C	T	T	C	H	C	C	H	T	H	C	C	C	C	H	H	H	C	C	H	C
DB9.2G/A	9267	H	G	H	G	G	G	H	G	H	G	G	G	G	H	G	H	H	G	H	H	G	G	G	G	G	G	A	G	G	G	A	H	G	H	G	G	H	A	H	G	G	G	G	H	H	H	G	G	H	G
DB10.5A/G	10502	G	H	H	H	A	A	H	H	H	G	A	A	H	H	G	H	G	G	A	H	G	H	A	A	H	H	H	H	G	G	H	A	G	H	A	H	H	G	A	H	G	A	H	G	H	H	H	G	A	G
DB12.3T/A	12339	T	H	T	T	T	T	T	T	H	H	T	T	T	H	T	H	T	T	H	T	T	T	H	T	H	T	T	T	T	T	T	T	T	A	T	T	T	T	H	T	H	T	H	H	T	H	A	T	T	T
DB13.5G/A	13520	H	H	H	G	G	G	H	H	A	G	G	G	H	H	H	G	A	G	H	A	G	G	G	H	G	H	A	G	G	G	A	H	G	H	A	G	H	A	H	H	G	H	G	A	A	H	H	A	H	G
DB14.3G/C	14400	G	H	G	G	G	G	G	H	H	G	G	G	H	G	H	G	H	G	G	G	H	G	G	G	H	G	H	G	G	G	G	G	G	G	G	G	H	G	G	G	G	H	G	H	G	H	H	G	H	G	G	G

Table 3.4 SNP genotypes from semen donors 1-50 for the 30 SNPs for Hotspot DB. SNP names are listed down the left and donor number across the top. Each genotype is depicted as either a homozygote (A, C, G, T, +, -) or heterozygote (H). Motif disrupting SNP DB7.5T/A is highlighted in red.

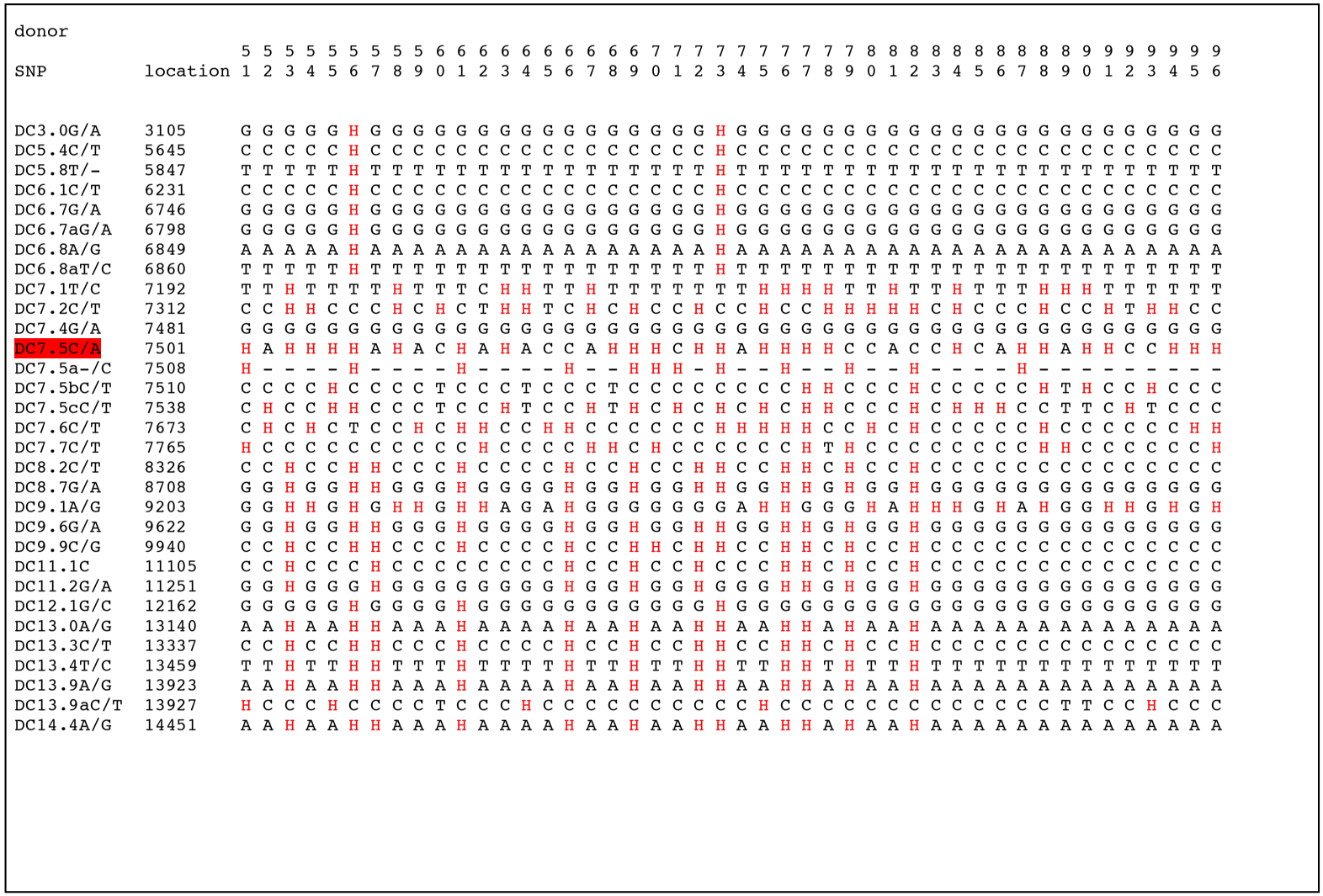
donor		5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9
SNP	location	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
DB2.2T/C	2286	T	T	T	T	T	H	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	H	H	T	T	H	T	T	T	T	T	T	T	T	T	T	H	T	T
DB2.3G/C	2390	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	H	H	G	G	G	G	G	G	H	G	H	G	G	G	G	H	G	G	H
DB2.4G/A	2448	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	H	H	G	G	G	G	G	G	H	G	H	G	G	G	G	H	G	G	H
DB2.5T/C	2612	C	C	C	H	C	H	H	H	H	H	C	C	H	C	C	C	H	H	C	H	C	C	C	H	C	T	T	H	C	H	H	C	C	C	H	C	C	C	C	T	C	H	T	H	H
DB2.7T/C	2704	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	T
DB2.7aC/G	2747	C	C	C	C	C	C	H	C	H	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	H	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	H
DB3.7A/G	3777	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A
DB3.9G/C	3927	G	G	G	G	G	G	H	G	H	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	H	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	H
DB4.1C/A	4152	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	H	C	C
DB4.1bA/G	4191	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	H	A	A
DB4.2A/G	4222	G	G	G	G	G	G	G	G	H	G	G	H	G	G	G	H	G	G	G	G	G	G	G	H	G	G	G	H	G	G	H	G	G	G	G	G	G	G	G	G	A	G	G	H	G	G
DB4.3C/G	4406	G	G	G	H	G	H	H	H	H	H	G	G	H	G	G	G	H	H	G	H	G	G	G	H	H	C	C	H	G	H	H	G	G	G	H	G	G	G	G	H	C	G	H	C	H	H
DB4.3aA/G	4413	A	A	A	A	A	A	H	A	H	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	H
DB4.5A/G	4517	A	A	A	A	A	A	H	A	H	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	H
DB5.4G/A	5448	A	A	A	H	A	H	H	H	A	H	A	A	H	A	A	A	A	H	A	A	A	H	A	A	A	H	H	H	G	H	A	H	H	A	A	H	A	A	A	A	H	G	A	H	G	A	H
DB6.6A/G	6689	G	G	G	H	G	H	G	H	G	H	G	G	H	G	G	H	H	G	G	G	G	G	G	H	G	H	H	H	H	H	H	G	G	G	H	G	G	G	G	G	A	G	G	A	H	G
DB7.0T/C	7058	C	C	C	C	C	C	H	C	C	C	C	C	C	H	C	C	C	H	C	C	C	C	C	C	C	C	C	C	H	C	C	H	C	C	C	H	C	C	C	C	C	T	C	C	T	C	C
DB7.0aC/A	7077	H	H	C	H	A	H	H	H	A	H	A	H	H	C	C	H	C	A	H	H	H	A	C	H	A	A	H	C	A	H	C	C	H	A	A	H	H	H	H	A	C	H	A	H	H	H
DB7.1G/T	7155	G	H	G	G	H	H	G	G	T	G	H	H	H	G	G	G	G	T	H	H	H	T	G	G	T	H	G	G	G	G	G	G	H	H	H	G	H	H	H	T	G	H	G	G	G	H
DB7.2T/C	7301	T	T	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	H	T	T	T	T	H	T	T	T	T	T	T	T	T	H	T
DB7.5T/A	7501	H	A	T	H	A	H	H	A	H	H	H	H	H	H	T	H	H	H	H	A	H	H	H	A	A	H	H	T	A	A	T	H	H	H	A	H	H	H	H	A	T	A	T	H	A	T
DB7.6C/T	7712	H	H	T	H	C	C	C	H	T	H	H	H	H	T	T	H	H	H	H	T	H	T	T	H	T	H	H	H	C	H	H	T	H	C	C	C	H	H	H	H	C	H	H	C	H	H
DB7.8T/C	7889	H	H	T	H	C	C	H	H	H	H	H	H	H	T	T	H	H	H	H	H	H	H	T	H	H	H	H	T	C	H	T	T	H	H	H	C	H	H	H	H	C	H	T	H	H	T
DB7.9T/A	7921	H	H	T	T	A	A	H	H	T	H	H	H	T	T	T	H	H	H	H	T	H	T	T	H	T	H	H	T	A	H	T	T	H	A	A	H	H	H	H	T	A	H	T	H	H	T
DB9.0C/T	9048	H	H	C	C	T	C	C	H	C	H	H	C	C	C	C	H	C	H	C	C	C	C	C	C	C	C	H	H	C	H	H	C	C	C	H	C	C	H	H	C	C	C	H	C	C
DB9.2G/A	9267	H	H	G	G	A	G	G	H	G	H	H	G	G	G	G	H	G	H	G	G	G	G	G	G	G	H	H	G	H	H	G	G	G	G	H	A	H	G	H	H	G	G	G	H	G	G
DB10.5A/G	10502	H	G	A	H	G	G	H	G	A	A	H	H	A	A	A	H	G	H	H	H	?	A	A	G	H	H	H	A	G	H	A	H	H	G	G	G	H	H	H	A	G	A	A	A	H	A
DB12.3T/A	12339	T	T	T	H	T	A	T	T	T	T	T	H	T	T	T	T	H	T	H	T	H	T	T	H	T	T	T	T	T	T	T	T	H	H	H	T	T	H	T	T	A	T	T	T	T	T
DB13.5G/A	13520	H	H	G	H	A	A	G	H	G	H	H	H	G	G	G	H	H	H	H	G	H	G	G	H	G	H	H	G	H	H	G	G	H	H	A	A	H	G	H	H	A	?	G	G	G	G
DB14.3G/C	14400	G	G	G	H	G	C	G	G	G	G	G	H	G	G	G	G	H	G	H	G	H	G	G	H	G	G	G	G	G	G	G	G	H	H	H	G	G	H	G	G	C	H	G	G	C	G

Table 3.5 SNP genotypes from semen donors 51-96 for the 30 SNPs for Hotspot DB. SNP names are listed down the left and donor number across the top. Each genotype is depicted as either a homozygote (A, C, G, T, +, -) or heterozygote (H). Motif disrupting SNP DB7.5T/A is highlighted in red.



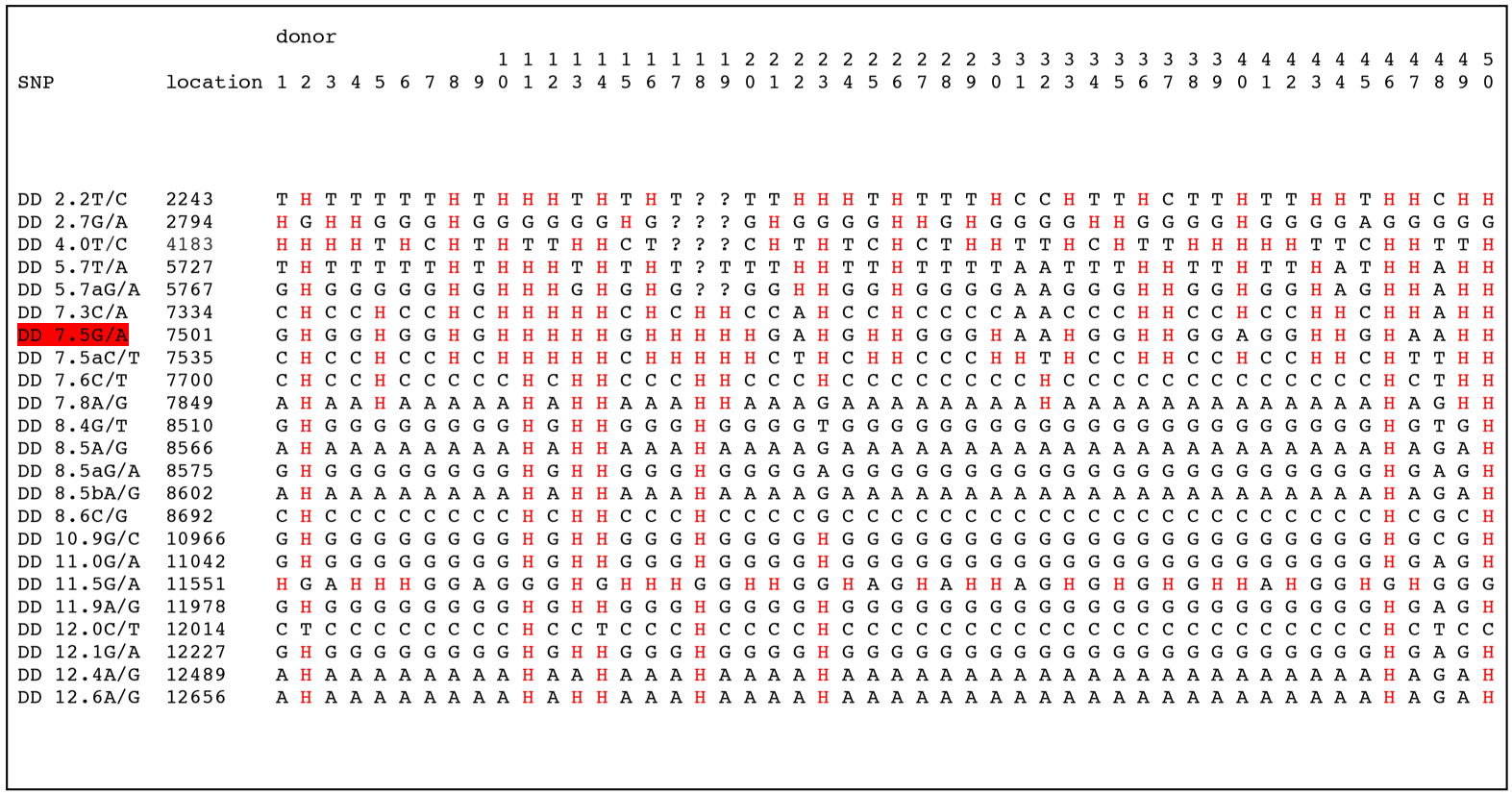
SNP	location	donor 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
DC3.0G/A	3105	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DC5.4C/T	5645	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DC5.8T/-	5847	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
DC6.1C/T	6231	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DC6.7G/A	6746	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DC6.7aG/A	6798	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DC6.8A/G	6849	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
DC6.8aT/C	6860	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
DC7.1T/C	7192	C	H	H	T	H	T	T	T	T	T	T	T	T	T	H	T	T	T	T	T	C	T	C	T	T	T	T	T	T	T	T	T	T	H	T	T	H	T	T	T	T	T	H	H	T	T
DC7.2C/T	7312	H	H	H	C	H	H	C	C	C	C	C	C	C	H	H	H	H	C	H	C	T	C	T	C	T	H	C	C	C	C	H	C	H	C	C	H	H	C	C	C	T	H	C	C	T	H	H	T	C
DC7.4G/A	7481	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DC7.5C/A	7501	A	H	A	A	H	A	H	C	C	A	H	H	H	C	C	H	H	C	H	H	C	H	H	H	C	H	H	H	A	C	H	H	A	H	A	A	H	H	H	H	H	A	C	H	H	C	H	H	C	C
DC7.5a-/C	7508	-	-	-	-	-	-	H	H	-	H	H	H	H	-	H	-	-	-	-	H	H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	-	-	-	-	H	-	-	-	-	H	-	-
DC7.5bC/T	7510	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	H	C	C	H	H	C	C	C	H	C	H	C	C	C	C	C	H	H	C	C	C	C	T	C	C	H	C	C	C	C	H	C	H	C
DC7.5cC/T	7538	H	C	C	C	H	C	H	H	C	C	H	H	T	C	C	C	T	C	H	C	H	H	C	C	C	H	H	H	C	H	H	C	H	H	C	C	C	T	C	C	H	C	C	C	C	H	H	H	C
DC7.6C/T	7673	T	C	H	C	C	C	H	H	T	C	H	T	H	C	H	H	C	H	C	H	H	C	C	H	C	H	C	C	T	H	C	T	H	H	C	H	H	H	C	C	C	C	T	C	H	C	C	H	C	H
DC7.7C/T	7765	C	H	C	H	C	C	C	H	C	C	C	C	C	H	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	H	C	C	H	C	C	C
DC8.2C/T	8326	C	C	T	C	C	C	H	H	C	C	H	H	H	C	H	C	H	C	C	H	H	C	C	C	C	C	H	C	C	C	C	T	C	C	C	H	H	C	H	C	C	T	C	H	C	C	H	C	C
DC8.7G/A	8708	G	G	A	G	G	G	H	H	G	G	H	H	H	G	H	G	H	G	G	H	H	G	G	G	G	G	H	G	G	G	G	A	G	G	G	H	H	G	H	G	G	A	G	H	G	G	H	G	G
DC9.1A/G	9203	G	H	G	H	G	A	G	G	G	H	H	G	G	H	H	H	G	G	H	G	G	H	A	G	G	H	G	G	H	H	H	G	H	G	H	A	H	G	H	G	H	H	G	H	G	G	H	G	H	G
DC9.6G/A	9622	G	G	A	G	G	G	H	H	G	G	H	H	H	G	H	G	H	G	G	H	H	G	G	G	G	G	H	G	G	G	G	A	G	G	G	H	H	G	H	G	G	H	G	H	G	G	H	G	G
DC9.9C/G	9940	C	C	G	C	C	C	H	H	C	C	H	H	H	C	H	C	H	C	C	H	H	C	C	C	C	C	H	C	C	C	C	G	C	C	C	H	H	C	H	C	C	H	C	C	C	H	C	C
DC11.1C	11105	C	C	T	C	C	C	C	H	C	C	C	H	C	C	H	C	H	C	H	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	H	H	C	H	C	C	C	C	C	H	C	C
DC11.2G/A	11251	G	G	A	G	G	G	G	H	G	G	G	H	G	G	H	G	H	G	H	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	H	H	G	H	G	G	G	G	G	H	G	G
DC12.1G/C	12162	G	G	G	G	G	G	H	G	G	G	H	G	H	G	G	G	G	G	G	H	H	G	G	G	G	G	G	G	G	G	G	C	G	G	G	G	G	G	G	H	G	H	G	G	G	G	G	G	G	G
DC13.0A/G	13140	A	A	G	A	A	A	A	H	H	A	A	H	H	H	A	H	A	H	A	A	H	A	A	A	A	A	A	G	A	A	A	A	H	H	A	H	H	A	A	H	A	H	H	A	A	H	A	A
DC13.3C/T	13337	C	C	T	C	C	C	C	H	H	C	C	H	H	H	C	H	C	H	C	C	H	H	C	C	C	C	C	T	C	C	C	C	C	H	H	C	H	C	C	H	C	H	C	C	H	C	C	H	C	C
DC13.4T/C	13459	T	T	C	T	T	T	T	H	H	T	T	H	H	H	T	H	T	H	T	T	H	H	T	T	T	T	T	H	T	T	T	T	C	T	T	T	H	H	T	H	T	T	H	T	H	T	T	H	T	T
DC13.9A/G	13923	A	A	G	A	A	A	A	H	H	A	A	H	H	H	A	H	A	H	A	A	H	H	A	A	A	A	A	H	A	A	A	A	G	A	A	A	A	H	H	A	H	A	A	H	A	H	A	A	H	A
DC13.9aC/T	13927	C	C	C	C	C	C	C	C	C	C	C	C	H	C	T	C	C	C	H	C	C	H	C	C	C	C	C	C	C	C	H	H	H	H	C	H	C	H	H	H	C	C	H	C	C	C	C	H	C	C
DC14.4A/G	14451	A	A	G	A	A	A	H	H	A	A	H	H	H	A	H	A	H	A	A	H	H	A	A	A	A	A	H	A	A	A	A	G	A	A	A	A	H	H	A	H	A	A	H	A	H	A	A	H	A	A

Table 3.6 SNP genotypes from semen donors 1-50 for the 31 SNPs for Hotspot DC. SNP names are listed down the left and donor number across the top. Each genotype is depicted as either a homozygote (A, C, G, T, +, -) or heterozygote (H). Motif disrupting SNP DC7.5C/A is highlighted in red.



donor		5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9
SNP	location	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
DC3.0G/A	3105	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DC5.4C/T	5645	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DC5.8T/-	5847	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
DC6.1C/T	6231	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DC6.7G/A	6746	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DC6.7aG/A	6798	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DC6.8A/G	6849	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
DC6.8aT/C	6860	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
DC7.1T/C	7192	T	T	H	T	T	T	T	H	T	T	T	C	H	H	T	T	H	T	T	T	T	T	T	T	H	H	H	H	T	T	H	T	T	H	T	T	H	T	T	T	T	T	T	T
DC7.2C/T	7312	C	C	H	H	C	C	C	H	C	H	C	T	H	H	T	C	H	C	H	C	C	H	C	C	H	C	C	H	H	H	H	H	C	H	C	C	C	H	C	C	H	T	H	H	C	C
DC7.4G/A	7481	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DC7.5C/A	7501	H	A	H	H	H	H	A	H	A	C	H	A	H	A	C	C	A	H	H	H	C	H	H	A	H	H	H	H	C	C	A	C	C	H	C	A	H	H	A	H	H	C	C	H	H	H
DC7.5a-/C	7508	H	-	-	-	-	H	-	-	-	-	-	H	-	-	-	-	H	-	-	H	H	H	-	H	-	-	H	-	-	H	-	-	H	-	-	-	-	H	-	-	-	-	-	-
DC7.5bC/T	7510	C	C	C	C	H	C	C	C	C	T	C	C	C	T	C	C	C	T	C	C	C	C	C	C	C	C	C	H	H	C	C	C	H	C	C	C	C	C	H	T	H	C	C	H	C	C	C
DC7.5cC/T	7538	C	H	C	C	H	H	C	C	C	T	C	C	H	T	C	C	H	T	H	C	H	C	H	C	H	C	H	H	C	C	C	H	C	H	H	H	C	C	T	T	C	H	T	C	C	C
DC7.6C/T	7673	C	H	C	H	C	T	C	C	H	C	H	H	C	C	H	H	C	C	C	C	C	C	C	H	H	H	H	H	C	C	H	C	H	C	C	C	C	C	H	C	C	C	C	C	H	H
DC7.7C/T	7765	H	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	H	H	C	H	C	C	C	C	C	C	H	T	H	C	C	C	C	C	C	C	H	H	C	C	C	C	C	C	H
DC8.2C/T	8326	C	C	H	C	C	H	H	C	C	C	H	C	C	C	C	H	C	C	H	C	C	H	H	C	C	H	H	C	H	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C
DC8.7G/A	8708	G	G	H	G	G	H	H	G	G	G	H	G	G	G	G	H	G	G	H	G	G	H	H	G	G	H	H	G	H	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G
DC9.1A/G	9203	G	G	H	H	G	H	G	H	H	G	H	H	A	G	A	H	G	G	G	G	G	G	G	A	H	H	G	G	G	H	A	H	H	H	G	H	A	H	G	G	H	H	G	H	G	H
DC9.6G/A	9622	G	G	H	G	G	H	H	G	G	G	H	G	G	G	G	H	G	G	H	G	G	H	H	G	G	H	H	G	H	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G
DC9.9C/G	9940	C	C	H	C	C	H	H	C	C	C	H	C	C	C	C	H	C	C	H	H	C	H	H	C	C	H	H	C	H	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DC11.1C	11105	C	C	H	C	C	C	H	C	C	C	C	C	C	C	C	H	C	C	H	C	C	H	C	C	C	H	H	C	H	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DC11.2G/A	11251	G	G	H	G	G	G	H	G	G	G	G	G	G	G	G	H	G	G	H	G	G	H	G	G	G	H	H	G	H	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DC12.1G/C	12162	G	G	G	G	G	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
DC13.0A/G	13140	A	A	H	A	A	H	H	A	A	A	H	A	A	A	A	H	A	A	H	A	A	H	H	A	A	H	H	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A
DC13.3C/T	13337	C	C	H	C	C	H	H	C	C	C	H	C	C	C	C	H	C	C	H	C	C	H	H	C	C	H	H	C	H	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DC13.4T/C	13459	T	T	H	T	T	H	H	T	T	T	H	T	T	T	T	H	T	T	H	T	T	H	H	T	T	H	H	T	H	T	T	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T
DC13.9A/G	13923	A	A	H	A	A	H	H	A	A	A	H	A	A	A	A	H	A	A	H	A	A	H	H	A	A	H	H	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A
DC13.9aC/T	13927	H	C	C	C	H	C	C	C	C	T	C	C	C	H	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	T	T	C	C	H	C	C	C
DC14.4A/G	14451	A	A	H	A	A	H	H	A	A	A	H	A	A	A	A	H	A	A	H	A	A	H	H	A	A	H	H	A	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A

Table 3.7 SNP genotypes from semen donors 51-96 for the 31 SNPs for Hotspot DC. SNP names are listed down the left and donor number across the top. Each genotype is depicted as either a homozygote (A, C, G, T, +, -) or heterozygote (H). Motif disrupting SNP DC7.5C/A is highlighted in red.



SNP	location	donor 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
DD 2.2T/C	2243	T	H	T	T	T	T	T	H	T	H	H	H	T	H	T	H	T	?	?	T	T	H	H	H	T	H	T	T	T	H	C	C	H	T	T	H	C	T	T	H	T	T	H	H	T	H	H	C	H	H
DD 2.7G/A	2794	H	G	H	H	G	G	G	H	G	G	G	G	G	G	H	G	H	?	?	?	G	H	G	G	G	G	G	H	H	G	H	H	G	G	G	G	H	H	G	G	G	G	H	G	G	G	G	A	G	G	G	G	G
DD 4.0T/C	4183	H	H	H	H	H	T	H	C	H	T	H	H	T	T	H	H	C	T	?	?	?	C	H	T	H	H	T	C	H	C	T	H	H	T	T	H	C	H	H	T	T	H	H	H	H	H	H	T	T	C	H	H	T	T	H
DD 5.7T/A	5727	T	H	T	T	T	T	T	H	T	H	H	H	T	H	T	H	T	?	T	T	T	H	H	H	T	T	H	T	T	T	T	A	A	T	T	T	H	H	T	T	H	T	T	H	A	T	H	H	A	H	H
DD 5.7aG/A	5767	G	H	G	G	G	G	G	H	G	H	H	H	G	H	G	H	G	?	?	G	G	H	H	G	G	H	G	G	G	G	A	A	G	G	G	H	H	G	G	H	G	G	H	A	G	H	H	A	H	H
DD 7.3C/A	7334	C	H	C	C	H	C	C	H	C	H	H	H	H	H	C	H	C	H	H	C	C	A	H	C	C	H	C	C	C	C	A	A	C	C	C	H	H	C	C	H	C	C	H	H	C	H	H	A	H	H
DD 7.5G/A	7501	G	H	G	G	H	G	G	H	G	H	H	H	H	H	G	H	H	H	H	H	G	A	H	G	H	H	G	G	G	H	A	A	H	G	G	H	H	G	G	A	G	G	H	H	G	H	A	A	H	H
DD 7.5aC/T	7535	C	H	C	C	H	C	C	H	C	H	H	H	H	H	C	H	H	H	H	H	C	T	H	C	H	H	C	C	C	C	H	H	T	H	C	C	H	H	C	C	H	C	C	H	H	C	H	T	T	H	H
DD 7.6C/T	7700	C	H	C	C	H	C	C	C	C	C	H	C	H	H	C	C	C	H	H	C	C	C	H	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	T	H	H
DD 7.8A/G	7849	A	H	A	A	H	A	A	A	A	A	H	A	H	H	A	A	A	H	H	A	A	A	G	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	G	H	H
DD 8.4G/T	8510	G	H	G	G	G	G	G	G	G	G	H	G	H	H	G	G	G	H	G	G	G	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	T	G	H
DD 8.5A/G	8566	A	H	A	A	A	A	A	A	A	A	H	A	H	H	A	A	A	H	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	G	A	H
DD 8.5aG/A	8575	G	H	G	G	G	G	G	G	G	G	H	G	H	H	G	G	G	H	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	A	G	H
DD 8.5bA/G	8602	A	H	A	A	A	A	A	A	A	A	H	A	H	H	A	A	A	H	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	G	A	H
DD 8.6C/G	8692	C	H	C	C	C	C	C	C	C	C	H	C	H	H	C	C	C	H	C	C	C	C	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	G	C	H
DD 10.9G/C	10966	G	H	G	G	G	G	G	G	G	G	H	G	H	H	G	G	G	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	C	G	H
DD 11.0G/A	11042	G	H	G	G	G	G	G	G	G	G	H	G	H	H	G	G	G	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	A	G	H
DD 11.5G/A	11551	H	G	A	H	H	H	G	G	A	G	G	G	H	G	H	H	H	G	G	H	H	G	G	H	A	G	H	A	H	H	H	A	G	H	G	H	G	H	H	A	H	G	G	H	G	H	G	G	G	G	G
DD 11.9A/G	11978	G	H	G	G	G	G	G	G	G	G	H	G	H	H	G	G	G	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	A	G	H
DD 12.0C/T	12014	C	T	C	C	C	C	C	C	C	C	H	C	C	T	C	C	C	H	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	T	C	C
DD 12.1G/A	12227	G	H	G	G	G	G	G	G	G	G	H	G	H	H	G	G	G	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	A	G	H
DD 12.4A/G	12489	A	H	A	A	A	A	A	A	A	A	H	A	A	H	A	A	A	H	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	G	A	H
DD 12.6A/G	12656	A	H	A	A	A	A	A	A	A	A	H	A	H	H	A	A	A	H	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	G	A	H

Table 3.8 SNP genotypes from semen donors 1-50 for the 23 SNPs for Hotspot DD. SNP names are listed down the left and donor number across the top. Each genotype is depicted as either a homozygote (A, C, G, T, +, -) or heterozygote (H). Motif disrupting SNP DD7.5G/A is highlighted in red.

donor		5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	8	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9
SNP	location	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
DD 2.2T/C	2243	T	T	H	H	H	T	T	H	T	H	T	T	H	T	T	C	T	T	H	T	T	T	T	T	H	H	T	H	T	H	T	T	C	H	H	C	T	T	T	T	T	T	H	T	H	C
DD 2.7G/A	2794	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	H	G	H	G	G	G	G	G	G	G	H	G	H	H	G	G	G	G	G	G
DD 4.0T/C	4183	H	C	H	T	T	H	H	H	T	T	C	T	H	H	C	T	H	T	T	T	T	T	H	C	C	T	H	H	T	H	T	H	T	T	T	H	T	T	H	T	C	C	H	C	H	T	H	T
DD 5.7T/A	5727	T	T	H	H	H	T	T	H	T	H	T	T	H	T	T	A	T	T	H	T	T	T	T	T	T	H	H	T	H	T	H	T	T	T	A	H	H	A	T	T	T	T	T	T	T	H	A
DD 5.7aG/A	5767	G	G	H	H	H	G	G	H	G	H	G	G	H	G	G	A	G	G	H	G	G	G	G	G	G	H	H	G	H	G	H	G	G	A	H	H	A	G	G	G	G	G	G	G	H	A
DD 7.3C/A	7334	C	C	H	H	H	C	C	H	C	H	C	C	H	C	C	H	C	C	H	H	C	C	C	C	H	H	C	H	H	H	C	C	A	H	H	A	C	C	C	C	C	C	C	C	H	A
DD 7.5G/A	7501	G	H	H	H	H	G	H	H	G	H	H	G	H	G	G	H	G	G	H	H	G	G	G	G	H	A	G	H	H	H	G	G	A	H	H	A	G	H	G	G	G	G	G	G	H	A
DD 7.5aC/T	7535	C	H	H	H	H	C	H	H	C	H	H	C	H	C	C	H	C	C	H	H	C	C	H	C	H	T	H	H	H	H	C	C	T	H	C	H	C	H	C	C	C	C	C	C	H	T
DD 7.6C/T	7700	C	C	C	C	H	C	C	C	C	H	C	C	C	C	C	H	C	C	C	H	C	C	C	C	H	C	C	H	H	H	C	C	T	C	C	C	C	C	C	C	C	C	C	C	H	C
DD 7.8A/G	7849	A	A	A	H	H	A	A	A	A	H	A	A	A	A	A	H	A	A	A	H	A	A	A	A	H	A	A	H	H	H	A	H	G	H	A	A	A	A	A	A	A	A	A	A	H	H
DD 8.4G/T	8510	G	G	G	H	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	H	G	G	T	H	G	G	G	G	G	G	G	G	G	G	H	G
DD 8.5A/G	8566	A	A	A	H	H	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	H	A	A	G	H	A	A	A	A	A	A	A	A	A	A	H	A
DD 8.5aG/A	8575	G	G	G	H	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	H	G	G	A	H	G	G	G	G	G	G	G	G	G	G	H	G
DD 8.5bA/G	8602	A	A	A	H	H	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	H	A	A	G	H	A	A	A	A	A	A	A	A	A	A	H	A
DD 8.6C/G	8692	C	C	C	H	H	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	H	C	C	G	H	C	C	C	C	C	C	C	C	C	C	H	C
DD 10.9G/C	10966	G	G	G	H	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	H	G	G	C	H	G	G	G	G	G	G	G	G	G	G	H	G
DD 11.0G/A	11042	G	G	G	H	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	H	G	G	A	H	G	G	G	G	G	G	G	G	G	G	H	G
DD 11.5G/A	11551	H	A	G	G	H	G	A	H	A	H	H	G	G	G	H	G	A	H	G	G	A	H	H	H	H	H	H	H	H	G	G	A	G	G	G	A	H	A	H	G	H	H	G	H	G	G
DD 11.9A/G	11978	G	G	G	H	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	H	G	G	A	H	G	G	G	G	G	G	G	G	G	G	H	G
DD 12.0C/T	12014	C	C	C	T	H	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	T	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DD 12.1G/A	12227	G	G	G	H	H	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	H	G	G	A	H	G	G	G	G	G	G	G	G	G	G	H	G
DD 12.4A/G	12489	A	A	A	H	H	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	H	A	A	G	H	A	A	A	A	A	A	A	A	A	A	H	A
DD 12.6A/G	12656	A	A	A	H	H	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	H	A	A	G	H	A	A	A	A	A	A	A	A	A	A	H	A

Table 3.8 SNP genotypes from semen donors 51-96 for the 23 SNPs for Hotspot DD. SNP names are listed down the left and donor number across the top. Each genotype is depicted as either a homozygote (A, C, G, T, +, -) or heterozygote (H). Motif disrupting SNP DD7.5G/A is highlighted in red.

3.2.3 Linkage Disequilibrium (LD) and Linkage Disequilibrium Unit (LDU) Mapping, and Coalescent Analysis

As part of the high-resolution study of meiotic recombination, LD analysis is used to define regions that potentially contain a recombination hotspot. Previously, Donnelly and colleagues (Myers *et al.,* 2005; Myers *et al.,* 2008) screened this target region DA on low-resolution Phase II HapMap data, and it was one of the best candidate LD hotspots based on having a motif with a disrupting SNP and due to its historical activity, as inferred by its strong LD breakdown. Genotyped SNPs in our donor panel were used to determine haplotype frequencies across the 15-kb target region, via a maximum likelihood approach (software written using TrueBASIC 4.1 by A. Jeffreys). Firstly, allele frequencies for each SNP were calculated and the most likely haplotypes for each pairwise comparison of SNPs were calculated from the genotype data. The analysis program calculates the most likely haplotypes from all the genotype data in a similar manner. Therefore, the D' value was calculated for the genotype data from the target region of DA. This identified the approximate position of a change from a region of free association to strong LD. In our analysis pairwise D' values were assigned a colour, with red representing high D' and black representing low D' (explained in detail in Chapter 1). For example, the D' value is red (= 1) and the corresponding odds of linkage equilibrium are also red, with a value of <0.0001 indicating that the statistical significance of the association between the two markers is high. Conversely, if it is blue then this indicates that the value of D' between these two sites is between 0.4-0.6. This would suggest substantial LD. However, the statistical significance (LR>0.05) shows that the odds of equilibrium are high and therefore it does not statistically support the D' value. Figure 3.3 shows higher-resolution LD mapping for the target region DA. LD breakdown was reproducible in these European donors and the Hotspot DA was narrow and flanked by regions of strong LD. LD plots showed evidence for localised LD breakdown but with some SNP pairs showing strong LD across a putative hotspot.

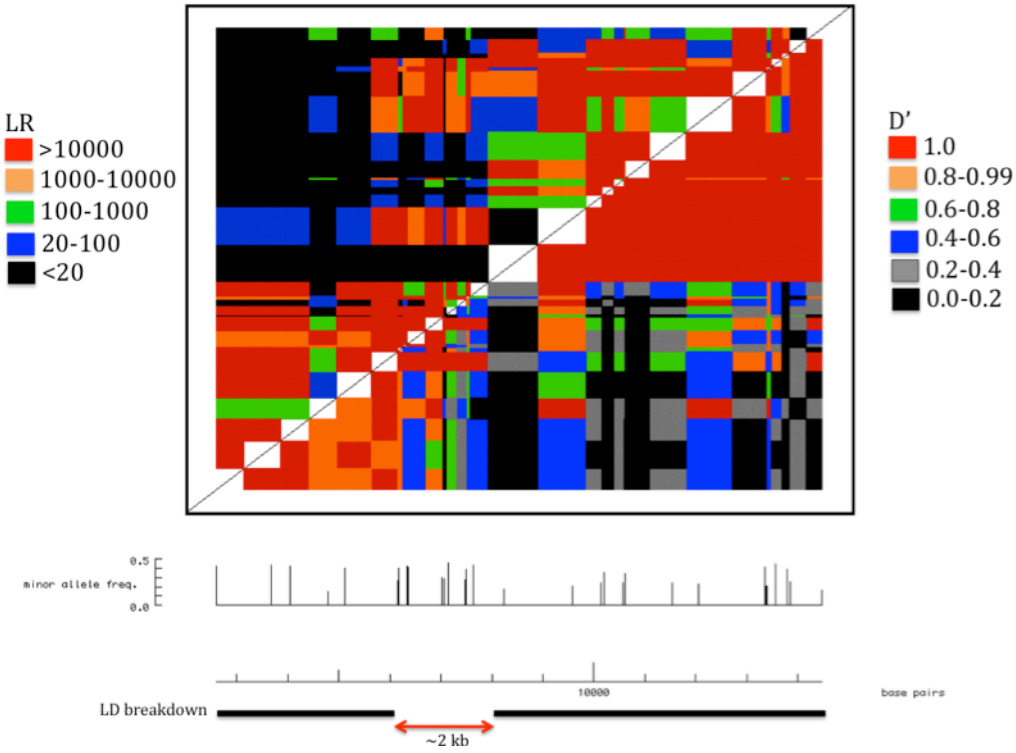


Figure 3.3 Linkage disequilibrium breakdown across the target region DA. All analyses were based on genotyped SNP typed data for Hotspot DA in our donor panel. Maximum likelihood haplotype frequencies for each pair of SNPs were determined and used to estimate |D'| levels of LD (lower right), plus the associated likelihood ratio (LR, odds of LD) versus free association (upper left), and are colour-coded as indicated. Only SNPs with minor allele frequencies (MAFs) ≥ 0.15 were included in the analysis. The locations of the SNPs are shown below and to the right of the plot, with positions centred on the middle of DA at co-ordinate 0. The LD block is shown in black below the plot, and the position and the approximate width of the LD hotspot DA is indicated by the red arrow.

Linkage Disequilibrium Unit (LDU) mapping:

Before testing the importance of the disrupting SNPs within the motif and carrying out crossover assays to estimate the recombination rate, all the genotyped targets (DA, DB, DC, and DD) were LDU mapped to confirm the presence of the putative hotspots, and compared with HapMap CEU data (Figure 3.4). In this anlaysis, the disrupting SNPs had to be within the motif localised at the centre of the hotspots. All putative hotspots show a big similarity with HapMap Phase II CEU data and with having 1-2-kb hotspot intervals. This reflected the normal concentration of human meiotic recombination hotspots (Jeffreys *et al.,* 2000). LD hotspot target DA was an excellent candidate, with strong and narrow LDU steps and a central SNP with a

disrupting allele affecting a key motif base. LD hotspot target DB showed the same features as target DA, but the disrupting SNP base change (T/A) did not affect a key base in the 13-bp moitf. LD hotspot target DC clearly showed a very narrow LDU step, but a key SNP disrupts the motif at the boundary of the LD hotspot and not in the centre. Therefore, this target was a poor candidate for a crossover assay. This information was not given in the results of Donnelly and colleagues, presumably because this LD hotspot was mapped at lower resolution using Phase II HapMap Data (Myers *et al.,* 2005). LD hotspot target DD was not a clearly localised LD hotspot, with LD weakly decaying over an interval of at least 4 kb. This LD hotspot was thus a very poor candidate and was rejected as a LD hotspot.

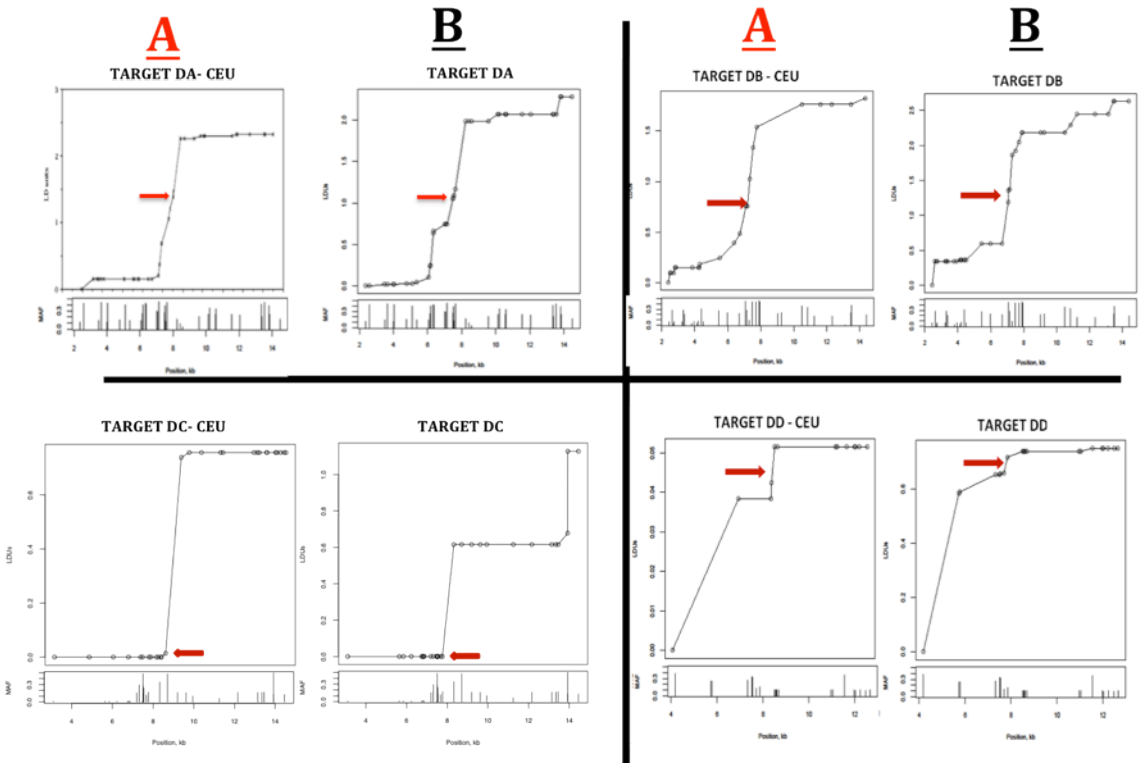


Figure 3.4 LDU maps of targets DA, DB, DC and DD. In the figure, LDU (B) maps confirmed the putative hotspots with 1-2 kb intervals, apart from target DD; also, each putative hotspot was compared with the LDU of CEU data from HapMap (A) (minor allele frequency > 0.3). The disrupting SNP within the motif is shown with red arrows. In targets DA and DB, the disrupting SNPs are localised close to the centre; therefore, both putative hotspots are eligible for crossover assays for testing the recombination rate of the hotspot. However, the disrupting SNP is localised very far away at target DC. Target DD showed only a very weak LDU step, and thus targets DC and DD were eliminated from this study.

Coalescent analyses

Comparing LDU maps with coalescent-based maps shows how different they are. LDU maps reflect the underlying LD structure, whilst coalescent maps reconstruct the underlying recombination rate with a particular demographic model (Fearnhead *et al*., 2004). Coalescent analyses were used for estimating recombination rates and predicting the location of hotspots as generally confirmed by direct estimates from sperm typing (Schneider *et al*., 2002; McVean *et al*., 2004). According to low-resolution analysis from HapMap Phase II data, these targets showed good historical activity. Coalescent analyses were carried out on genotyped SNPs for each target LD hotspot in our donor panel (Figure 3.5). Coalescent analysis of Hotspot DA suggested a highest peak of activity of ~50 cM/Mb crossover rate in all four LD hotspots, with an approximate historical recombination frequency (RF) of $\sim 5\times10^{-4}$. This RF is typical for crossover hotspots analysed by sperm typing. Hotspots analysed by sperm typing vary widely in crossover activity, with RF values ranging from 5×10^{-6} to 3×10^{-3} for currently characterised hotspots. This range of nearly three orders of magnitude almost certainly reflects widely differing rates of recombination initiation. The LD hotspots DB and DC showed a similar peak activity of ~32 cM/Mb and a crossover rate of $\sim 3.2\times10^{-4}$. These results defined hotspots with >2 kb resolution. This width is more than previously observed hotspot intervals (Jeffreys *et al.,* 2000), and moreover these hotspots do not show typical coalescent mapping for a normal hotspot. The lesser peak activity (~14 cM/Mb) for the LD Hotspot DD was reported along with a very narrow hotspot width (< 0.4 kb). Therefore, according to coalescent analysis, we confirmed the LDU results for all four LD hotspots.

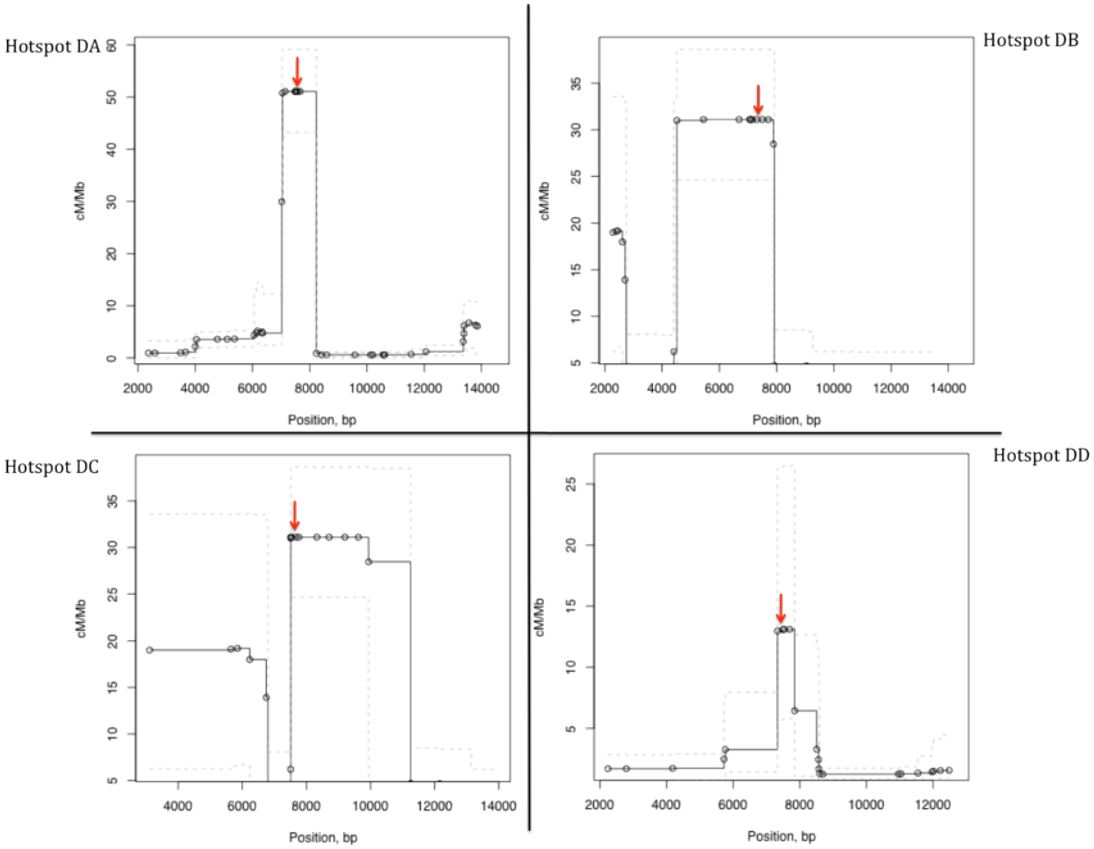


Figure 3.5 Coalescent analyses of four LD hotspots (DA, DB, DC and DD). Coalescent mapping gives clues about historical recombination and recombination frequency. For example, Hotspot DA showed a highest peak activity of ~ 50 cM/Mb crossover rate in all four LD hotspots (RF ~ 5 × 10^{-4}.) The width of Hotspot DA was ~1.4 kb. Despite not having any good coalescent mapping, Hotspots DB and DC suggested a similar peak activity of ~32 cM/Mb with a very broad hotspot width of >2 kb. The lesser peak activity (~14 cM/Mb) suggested for Hotspot DD came with a very narrow hotspot width (<0.4 kb). Therefore, LDU maps were confirmed by coalescent analyses for these LD hotspots. Also, red arrows show the location of the disrupting SNP within the motif in each LD hotspot.

3.4 DISCUSSION

According to our knowledge, 40% of human recombination hotspots have a 13-bp CCNCCNTNNCCNC sequence motif that determines activity in the hotspots (Myers *et al.,* 2008). The work of previous studies (Jeffreys *et al.,* 2002; Jeffreys *et al.,* 2005) has shown that CCTCCCT and CCCCACCCC motifs are present at the centre of hotspots *DNA2* and *NIDI* (Jeffreys and Neumann, 2002; Jeffreys *et al.,* 2005). Furthermore, SNPs in both the two motifs disrupt hotspot activity, with the non-reference SNP base acting as a recombination-suppressing allele (Myers *et al.,* 2005). The four best candidate SNPs within the 13-bp CCNCCNTNNCCNC motif theoretically locate to the middle of four hotspots and have been analysed to confirm the presence of the putative hotspot in our donor panel by LDU mapping. Confirming these hotspots with LDU mapping is crucial because it establishes the existence of the hotspot in our donor panel and the location of disrupting SNPs within the motif, and also the size of the hotspot can be estimated. Therefore, further analysis with a reciprocal crossover assay (Jeffreys *et al.,* 2002) can be carried out according to LDU mapping, genotyping of the target regions, and the availability of genomic DNA in our stocks.

Thirty-nine SNPs have been genotyped using ASO hybridisation for the putative hotspot DA. LDU mapping confirmed 1-2 kb intervals (Jeffreys *et al.,* 2000) with the disrupting SNP in the middle, and estimated the historical recombination frequency to be $\sim 5\times10^{-4}$. This RF shows that the hotspot is an ordinary hotspot, but the exact RF should be deduced and more information from the crossover assay should be collected. Following the genotyping of 30 and 23 SNPs in the regions of the second and fourth putative hotspots DB and DD respectively, LDU maps confirmed the location of the disrupting SNP to be close to the centre of both hotspots. However, for the 31 SNPs genotyped for target DC, the LDU map showed the disrupting SNP to be outside of the centre. Thus, target DC cannot be studied for this project because it does not fit into the objectives of this thesis.

In summary, given our knowledge about putative hotspot DA, good heterozygosity of donors in our panel and the availability of genomic DNA, this is the best candidate for further crossover assay analysis. Chapter 4 describes its detailed study. Putative

hotspot DB is also quite promising for the future study of crossover activity. Even though the disrupting SNP is localised at the centre of the putative hotspot DD, the heterozygosity of the donors does not allow us to conduct further analyses. Given the location of the disrupting SNP in target DC and the lack of good donors in our donor panel for target DD, SNPs in these putative hotspots were deemed unsuitable for study in this thesis.

CHAPTER 4: FREQUENCY AND DISTRIBUTION OF RECOMBINATION EVENTS AT HOTSPOT DA

4.1 INTRODUCTION

Intergenic SNP rs703642, which is located on the short arm of chromosome 9, disrupts the 13-bp CCNCCNTNNCCNC motif at the first base of the motif sequence (SNP rs703642: CCNCCNTNNCCNC → TCNCCNTNNCCNC). Previously, LD plots from HapMap phase II genotype data (Stephens and Donnelly, 2003) have confirmed the presence of the hotspot in four different populations, Utah residents with ancestry from northern and western Europe (CEU), Yoruba in Ibadan, Nigeria (YRI), Japanese in Tokyo, Japan (JPT) and Han Chinese in Beijing, China (CHB). Additionally, Donnelly and colleagues drew up a shortlist of the four best candidate LD hotspots based on having a motif with disrupting SNP accompanied by high historical crossover activity based on low-resolution HapMap Phase II data (Myers *et al.,* 2005) (Chapter 3). In Chapter 3, the 15-kb interval region of Hotspot DA was genotyped using ASO hybridization for a 92-donor panel, testing for the presence of the hotspot. LD and LDU map analysis confirmed the presence of the hotspot (1-2-kb interval) (Figure 3.3 and 3.4).

This Chapter discusses the optimisation of the allele-specific PCRs, identification of appropriate semen donors for the assay, recovery of recombinant DNA molecules, the analysis of recombination rate and distribution, and determination of the hotspot centre of Hotspot DA.

4.2 THE CROSSOVER ASSAY STRATEGY FOR HOTSPOT DA

The recombination rates and distributions of crossovers at recombination hotspots were measured using repulsion-phase allele-specific PCR of recombinant molecules from genomic sperm DNA. The strategy behind this method was that selective PCR recovers recombinant molecules from batches of genomic DNA, if necessary allowing millions of molecules to be rapidly screened. In Chapter 2, the reagents and methods of recovering crossovers and the estimation of recombination frequency are discussed.

4.3 SELECTION OF SEMEN DONORS FROM THE DONOR PANEL

It was necessary to have blood controls, as well as the sperm reactions while recovering crossover molecules, in order to test for meiotic specificity of putative crossovers and to check for production of PCR artefacts during the reaction. As only a small proportion of the semen donor panel also donated blood samples, it was necessary to optimise the assay on one of these donors first.

4.4 OPTIMISATION FOR RECOMBINATION ASSAY

When performing crossover assays, it helps to keep the amplified fragment as short as possible, while still including multiple informative heterozygous SNPs. Targets of recombination assays are generally ~4 - 7 kb long (Jeffreys *et al.,* 1998a; Jeffreys *et al.,* 2000; Jeffreys *et al.,* 2001; May *et al.,* 2002; Jeffreys and Neumann, 2005; Kauppi *et al.,* 2009). Heterozygous SNPs were used as selector sites for allele-specific amplification, and heterozygous SNPs in the intervening region acted as markers to map crossover breakpoints. Possible allele-specific PCR selector sites were selected based on their proximity to the target region, and the requirement for them to be present in the heterozygous state in a number of semen donors. However, for testing the effect of the disrupting SNP on the hotspot initiation activity, also both homozygote donors for motif disrupting SNPs were analysed (Jeffreys and Neumann, 2002; Jeffreys and Neumann, 2005; Neumann and Jeffreys 2006). Two possible forward and two reverse allele-specific primers were designed, and the annealing temperature of each primer optimised on DNA from semen donors homozygous for one or other allele at the locus. These optimisation experiments were necessary to achieve efficient and highly selective amplification of crossover molecules.

4.4.1 Recombination Assay for Hotspot DA

4.4.1.1 Optimisation of forward Allele-Specific Primers (ASPs)

The three SNPs chosen for the forward primer selector sites were DA2.6, DA3.7 and DA4.0a (Appendix II), as most of the semen donors that possessed these SNPs in the heterozygous state also had informative SNPs within the target region. There are several selection criteria for any SNP site to be used in recovery of crossovers; the SNP site must be within the LD block well away from the putative hotspot, sites must be within single-copy DNA (important for primary selector sites, but not absolutely necessary for secondary selector sites) and they must have a relatively high GC content (~60%).

In the optimisation experiments, MDA DNAs from donor 50 and donor 51 were used, as these donors were homozygous for all three SNPs; donor 50 had the alleles G, T and G for DA2.6, DA3.7 and DA4.0a, respectively and donor 51 had C, G and A. These donors were used for the entirety of the forward primer optimisation.

Universal primers (those which prime from either haplotype) were designed at internal positions to the allele-specific primers, within the target region. One forward and one reverse primer were designed; these primers were 20-24 bases in length, and roughly 65% GC rich. They were designed for three reasons: to ensure amplification was possible across the entire target region, to use in conjunction with allele-specific primers during optimisation, and to use in the recombination assay to the increase yield of the crossover PCR products. Optimisation PCRs for each of the allele-specific primers were performed on semen donors' MDA DNA, homozygous either for the allele being tested (+) or for the opposite allele (-), using the forward universal primer (DA5.9aR) at temperatures shown in Appendix II. The PCR products were electrophoresed on an agarose gel (Figure 4-1).

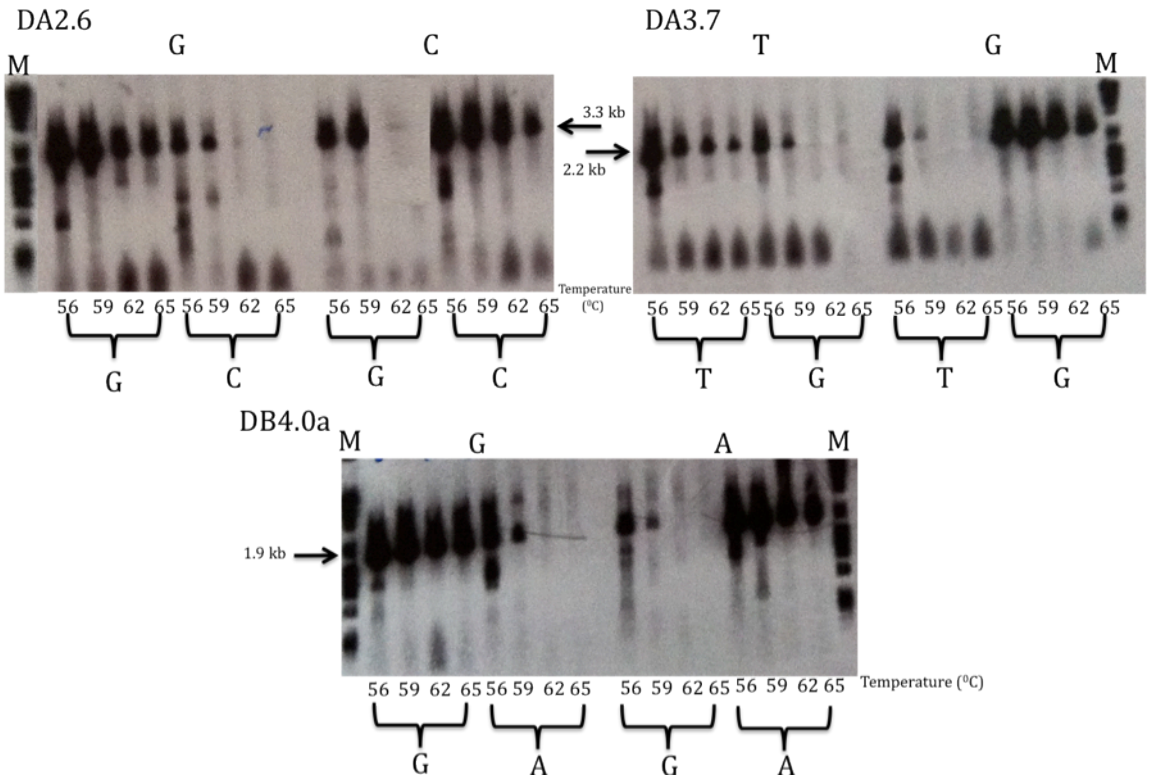


Figure 4-1 Optimisation of forward allele-specific primers on ethidium-bromide-stained agarose gels (presented as negatives). Three forward ASP pairs were optimised: DA2.6FG/C, DA3.FT/G and DA4.0aFG/A from donors homozygous for either the allele being tested or the opposite allele amplified using MDA DNA. DA5.9aR was used as a reverse universal primer. PCRs using each primer were performed at four different annealing temperatures 56°C, 59°C, 62°C and 65°C. DNA inputs were 10ng per 10µl PCR. (M: marker λ*Hin*dIII and ϕ*XHae*III).

All primers worked, and showed similar efficiency and specificity. All primers at 62°C showed specificity. It was preferred to amplify the all-forward ASPs at an annealing temperature greater than 59°C. However, the optimum temperature for the PCRs could not be decided until the reverse primers were optimised.

4.4.1.2 Optimisation of reverse Allele-Specific Primers (ASPs)

The four SNPs chosen for the reverse primer selector sites were DA10.6RG/A, DA10.6aRA/C, DA10.6bRC/A and DA12.0RC/T (Appendix II). In the optimisation experiments, MDA DNA from donor 4, donor 5, donor 34 and donor 38 were used, as these donors were homozygous for all four SNPs. Donor 4 had the alleles A, C, C and C for DA10.6, DA10.6a, DA10.6b and DA12.0, respectively, donor 5 had G, A, C

and C, donor 34 had G, C, A and C and donor 38 had G, C, C, T. Optimisation of primers directed to 3′ SNPs was more difficult than for the forward primers. ASP DA10.6RG/A had to be redesigned, and was named DA10.6RG2/A2. DA9.9aF has been designed as a forward universal primer (PCR profile and primers are shown in Appendix II). Figure 4-2 shows the electrophoresed PCR products at different annealing temperatures.

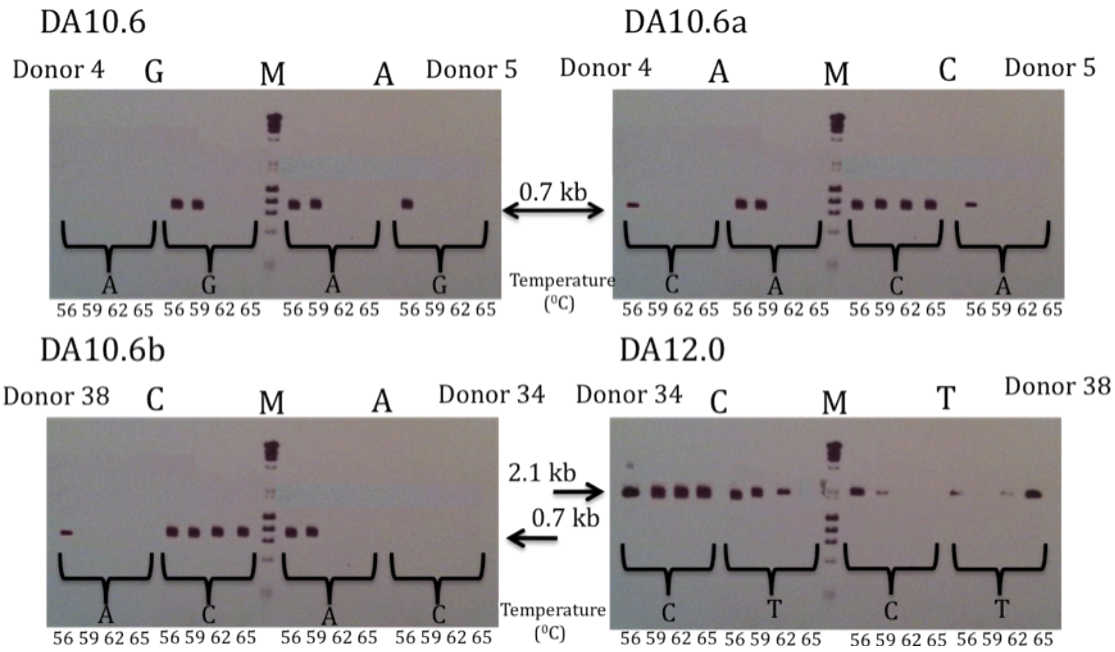


Figure 4-1 Optimisation of reverse allele-specific primers on ethidium-bromide-stained agarose gels. Four reverse ASPs were optimised: DA10.6RG2/A2, DA10.6aRA/C, DA10.6bRC/A and DA12.0RC/T, from donors homozygous for either the allele being tested or the opposite allele amplified. DA9.9aF was used for universal forward primer. PCRs using each primer were performed at four different annealing temperatures 56°C, 59°C, 62°C and 65°C. DNA inputs were 10ng per 10µl PCR/ reaction. (M: marker λ*Hin*dIII and ϕ*XHae*III).

As with all forward allele-specific primers, all reverse allele-specific primers worked, but with varying degrees of efficiency and specificity.

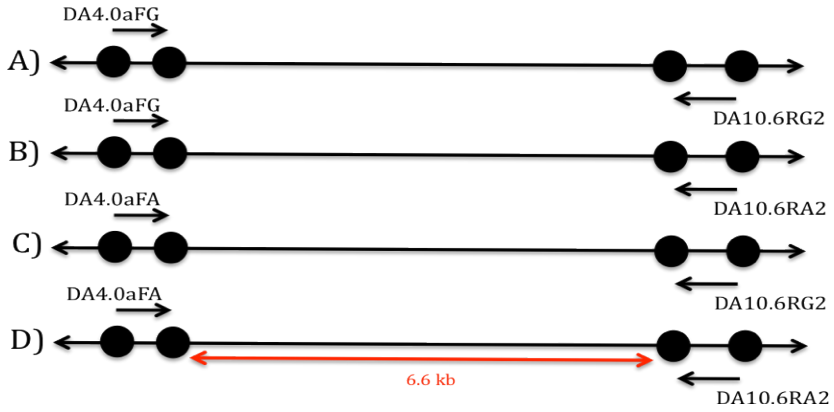
4.4.1.3 Linkage phasing of donors

Donors were selected on the basis of having multiple heterozygous SNP sites both flanking and within the target hotspot region, which were identified by genotyping analysis and their locations in the LDU map. Donors required heterozygous sites at

two forward ASP sites and two reverse ASP sites. In order to characterise the distribution of crossovers within the hotspot, it was necessary to choose donors with as many informative SNP sites as possible within the interval of LD breakdown. Therefore, primarily 13 donors have been selected from the donor panel. Linkage phase of each donor was determined by PCR amplifications using different combinations of secondary forward and reverse ASP selector sites (Figure 4.3) (PCR conditions and primers are shown in Appendix II).

After linkage phasing, the availability of sperm and blood DNA in the available stock was the other elimination criteria for donor selection. Donor 7 was heterozygous for all of the primer sites and had 11 informative sites (including disrupting SNP within the motif) within the region. Compared with other 12 donors, donor 7 was the best option when these criteria were considered. Also, more importantly, blood DNA from this donor was available, and this donor's sperm DNA was used to optimise the recombination assay. Linkage phasing results for donor 7 and the other donors (those having the same secondary ASP selector site genotypes) are given in Figure 4.3b.

a)



b)

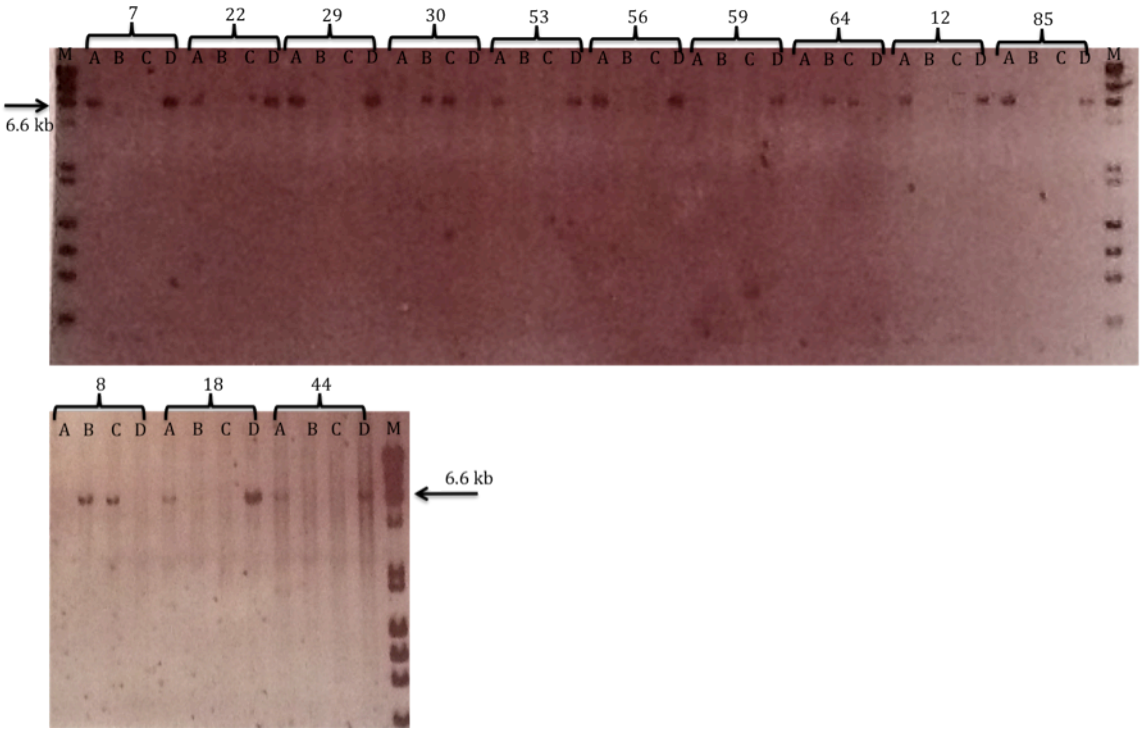


Figure 4.3 Designing linkage phasing assay for suitable donors for crossover assay. a) Shows four different combinations of internal allele-specific primers (ASPs). Internal allele-specific primers used for designing the linkage-phasing assay. The phasing of two forward ASPs and two reverse ASPs were done using linkage mapping. b) PCR results of 13 selected donors. Electrophoresis results give us information on the phasing for each donor; for example, donor 7 showed positive PCR reaction on A and D. Thus, the haplotype phasing of donor 7 is G – G, A-A). (M: marker λ*Hin*dIII and ϕ*XHae*III).

4.4.1.4 Performing the recombination assay on donor 7

The reciprocal crossover asymmetry test is a powerful rate-independent test that examines the distributions of exchange points to test whether reciprocal exchanges are distributed differently across a hotspot, and can be used to detect even minor variation in crossover activity. The reciprocal crossover asymmetry test also allows recombination proficiency at the level of individual haplotypes to be analysed (Jeffreys and Neumann, 2002). A pilot experiment using a crossover assay strategy consisting of two rounds of repulsion-phase allele-specific PCR was performed on sperm DNA from donor 7. Each reaction had a number of amplifiable molecules varying from 300 to 9600; using DA2.6FG and DA10.6bRC in the primary PCR and DA4.0aFG and DA10.6A2 in the secondary PCR (PCR conditions and primers are shown in Appendix II). Additionally, blood DNA of donor 7 was used for negative controls with the number of molecules in each reaction varying from 4800 to 9600.

In detail, in the pilot assay 12 x 300 molecules, 12 x 600 molecules, 12 x 1200 molecules, 12 x 2400 molecules, 12 x 4800 molecules and 12 x 9600 molecules from sperm DNA (a total of 2.27×10^5 amplifiable sperm DNA molecules for each haplotype); and 12 x 4800 molecules and 12 x 9600 molecules blood DNA controls (a total of 1.7×10^5 amplifiable blood DNA molecules for each haplotype) were used. A total of 22 positive reactions were seen. The numbers of amplifiable crossover molecules are lower than the true numbers of molecules, since some molecules are non-amplifiable due to nicking or other damage; numbers were calculated using the Poisson distribution programme written by Alec J. Jeffreys, University of Leicester, UK. Poisson-correction was used for each analysed men for estimating the number of crossover molecules from the proportion of PCR that were negative; maximum-likehood software This programme uses a mathematical model to determine the probability of a negative reaction (P_0) using the equation, $P_0=e^{-m}$, where m is the mean number of amplifiable molecules per PCR. In the pilot assay, the recombination frequency (RF) was found to be 1.15×10^{-4} (95% CI: 0.6×10^{-4} - 2.24×10^{-4}) The recombination frequency (RF) is given as the number of crossovers per input molecule (Figure 4.4).

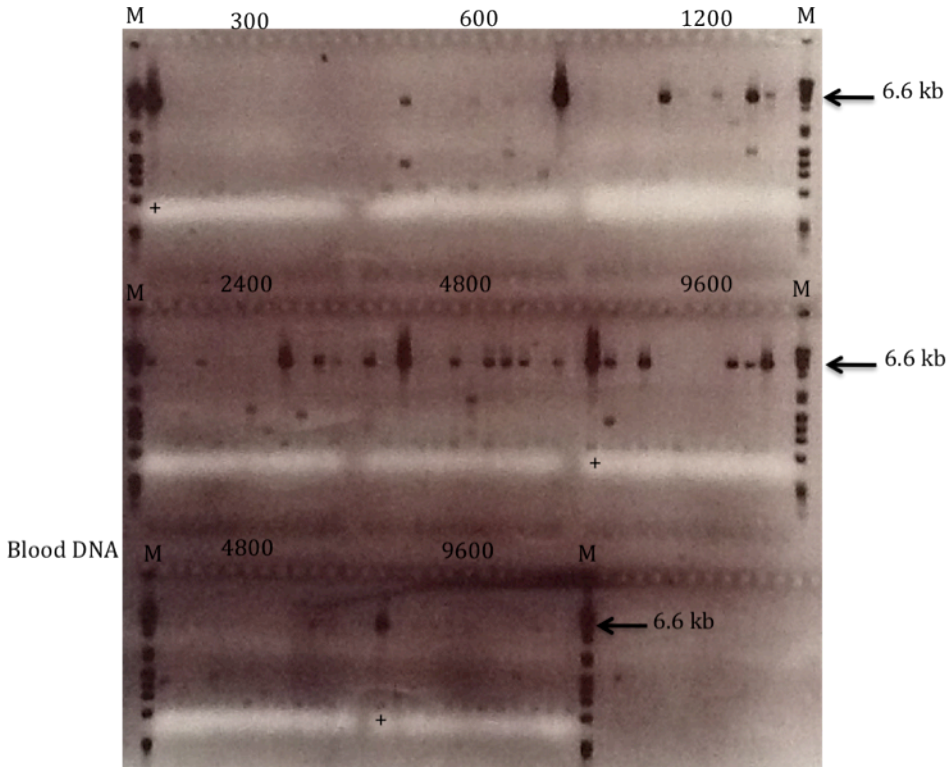


Figure 4.4 Pilot assay results for donor 7. This pilot assay has genomic DNA pool sizes covering a wide range of inputs (300, 600, 1200, 2400, 4800, and 9600 molecules per reaction). Blood DNA was used as a negative control in pool sizes of 4800 and 9600. Positive controls (donor 8) including sperm DNA from a man with primers in coupling phase is indicated by (+) with respectively, 0.24 ng sperm DNA input. (M: marker λ*Hin*dIII and ϕ*XHae*III).

The large-scale assay is performed in pool sizes depending upon on the optimum pool size determined from the pilot assay that maximises the number of recombinants recovered, while limiting the amount of mixed reactions. Mixed reactions occur when more than one recombinant molecule is present in a single PCR. The recombination assay was based on the two repulsion-phase PCRs and tertiary PCR using universal primers. However, when amplified with universal primers (DAXOF + DA10.4R) (Appendix II) these products formed a visible band when electrophoresed on an agarose gel and stained with ethidium bromide. Four pool sizes were used for the recombination assays for both orientation A and B (Figure 4.5), including; 24 x 4200, 24 x 5830, 24 x 8400 and 24 x 11670 molecules per pool (a total of 7.22 x 10^5 amplifiable molecules for each haplotype), and amplified to select for recombinants (PCR profile and primers are shown in Appendix II). Universal primers were used for

tertiary PCR and total of 60 positive reactions for orientation A, and 73 positive reactions for orientation B were observed. Therefore, in total approximately 723,000 molecules were used for both orientations A and B.

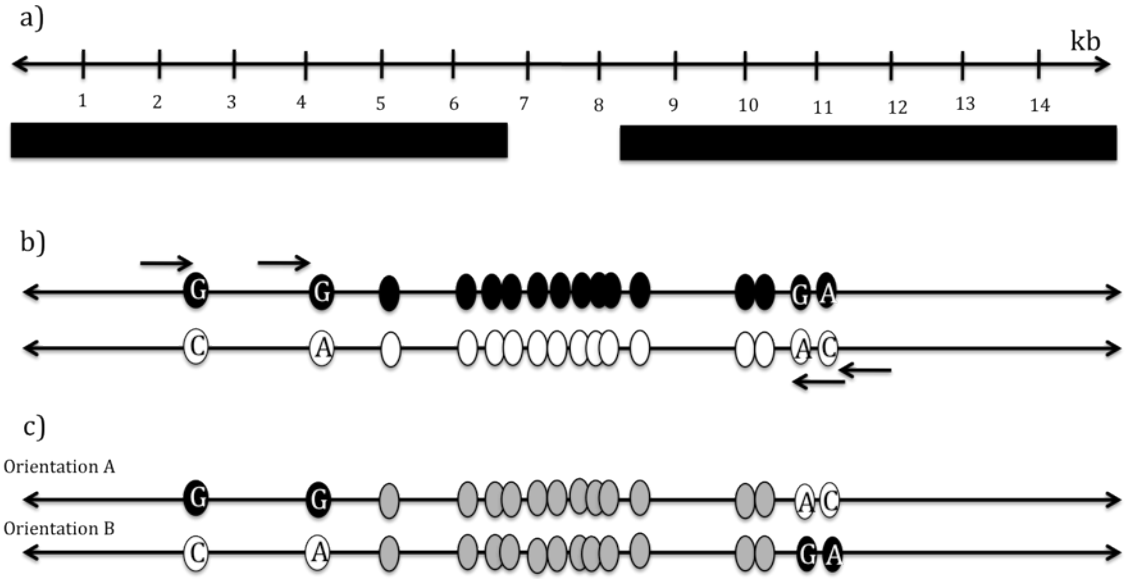


Figure 4.5 Illustrates the design of a recombination assay for both orientations for donor 7 according to linkage phasing and genotyping. a) 15 intervals are shown with an approximately 1.5-kb region of LDU breakdown. b) Two forward and two reverse allele-specific primers for each haplotype are shown with heterozygous selector sites. c) Both orientation A and B have been designed for crossover assay.

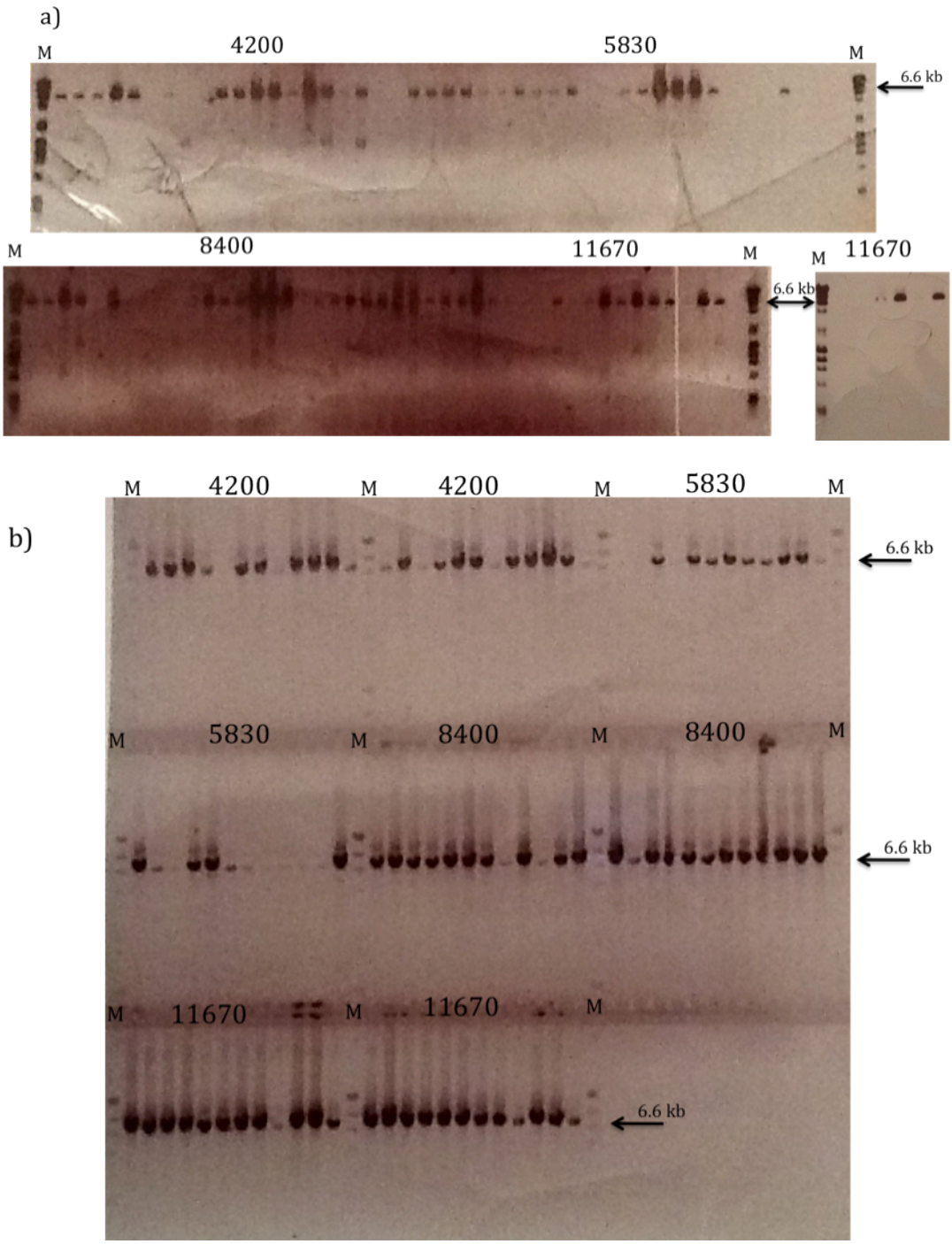


Figure 4.6 Secondary crossover assay PCR results of both orientation A (a) and B (b). a) Shows the result of the orientation A crossover assay. Sperm DNAs were amplified in different concentrations consisting of 4200 molecules, 5830 molecules, 8400 molecules and 11670 molecules respectively. A total of 60 positive reactions were observed. b) Shows result of orientation B crossover assay. Sperm DNAs were amplified in different concentrations as orientation A. A total of 73 positive reactions were observed. (M: marker λ*Hin*dIII and ϕ*XHae*III).

4.4.1.5 Mapping recombinant PCR products

The tertiary PCR products from recombination assays for donor 7 (both orientations A and B) were then dot blotted and hybridised sequentially with first one allele for a given SNP and then the opposite allele (Figure 4.7b). All informative SNPs within the amplified target were typed (PCR conditions and primers are shown in Appendix II). Hybridising PCR products with allele-specific oligo (ASO) probes is necessary, so that the recombination rate, distribution and the centre of the hotspots can be calculated by crossover mapping by ASO hybridisation. The results of hybridisations for crossover recombinants for donor 7 are shown in figure 4.7. In orientation A, recombination frequency was calculated as 1.1×10^{-4} (95% CI: $0.6 \times 10^{-4} – 2.4 \times 10^{-4}$) or 1 crossover in 9100 molecules and for orientation B 1.5×10^{-4} (95% CI: $0.7 \times 10^{-4} – 2.5 \times 10^{-4}$) or 1 crossover in 6700 molecules was used to combine data across all input pool sizes.

Figure 4.7 Mapping of crossover products in donor 7. a) Schematic representation of layout of dotblots. PCR products from orientation A (red) and from orientation B (yellow) are shown. Progenitor sites determine the haplotypes for SNP alleles. Progenitors for orientation A (red square) and B (yellow square) are also shown. b) Constituent haplotypes in donor 7 are shown on either side, with informative SNPs indicated. PCR products from crossover assay were sequentially hybridised with one allele of an informative SNP, followed by the opposite allele. The progenitor haplotypes were hybridised at the same time, and can be seen with each blot. The motif-disrupting SNP is DA7.5G/T.

4.5 DISTRIBUTION OF CROSSOVERS WITHIN HOTSPOTS DA

4.5.1 Calculating Crossover Activity

The crossover activity is calculated by as (inter-marker crossovers / total number of molecules) / inter-marker distance x 100 (cM) x 1000000 (Mb), and is given in cM/Mb (table 4.1). Each PCR was examined to determine in which SNP interval the crossover breakpoint lies, the numbers of total crossovers for each interval were determined, and the numbers of reactions positive for crossover were subjected to Poisson correction. This accounts for any crossovers that have occurred in mixed reactions (sometimes one crossover reaction might contain two molecules) that cannot be identified by SNP typing (see Materials and Methods). Table 4.1 shows a summary of crossover data obtained for donor 7, in both orientations, and the numbers of crossovers typed for each.

Donor 7

SNP	location	distance (bp)	No. A COs	No. B COs	cM/Mb, A COs	cM/Mb, B COs
DA5.O	5112		0	0		
DA6.1a	6171	1058	0	0	0.0	0.0
DA6.3	6335	163	0	0	0.0	0.0
DA6.3a	6352	16	0	0	0.0	0.0
DA7.0a	7049	696	15.6	0	3.2	0.0
DA7.4	7481	431	38	3.2	12.5	0.5
DA7.5	7501	19	3.2	2.2	23.8	7.8
DA7.5a	7542	40	10.8	18.8	38.2	31.8
DA8.4	8413	870	10	230.7	1.6	18.0
DA10.0	10126	1712	0	1.5	0.0	0.1
DA10.2	10203	76	0	0.5	0.0	0.4
DA10.6	10563	359	0	0	0.0	0.0
		Poisson corrected total COs	76.6	259.8		

Table 4.1 Distribution of crossovers in the Hotspot DA. The name of each informative SNP site in the target interval for each donor, their locations and their distance (bp) to next marker, number of crossovers (No. A COs) for orientation A, number of crossovers for orientation B (No. B COs), calculated recombination activity in cM/Mb for orientation A (A COs) and orientation B (B COs) (separately) and Poisson corrected number of crossovers for both orientation A and B are also shown. Number of crossovers and estimated recombination activity (cM/Mb) are given between two markers, for example 15.6 crossovers observed in orientation A between SNP DA7.0A and DA7.4. The numbers of crossovers are different in both orientations. This means that mixed PCRs were observed in some crossovers, containing more than one molecule.

When the peak activity was plotted against the size of SNP intervals, the hotspot was clear, and the width was estimated at around 1.2 kb (Figure 4.8).

The centre point of the hotspot for donor 7 was estimated at position 7601 and the width within which 95% of crossovers took place was estimated at approximately 1.6 kb. The same program was used to generate values for a best-fit normal distribution curve, and for a best fit cumulative frequency curve. The graphs of cumulative frequency values are shown in Figure 4.9.

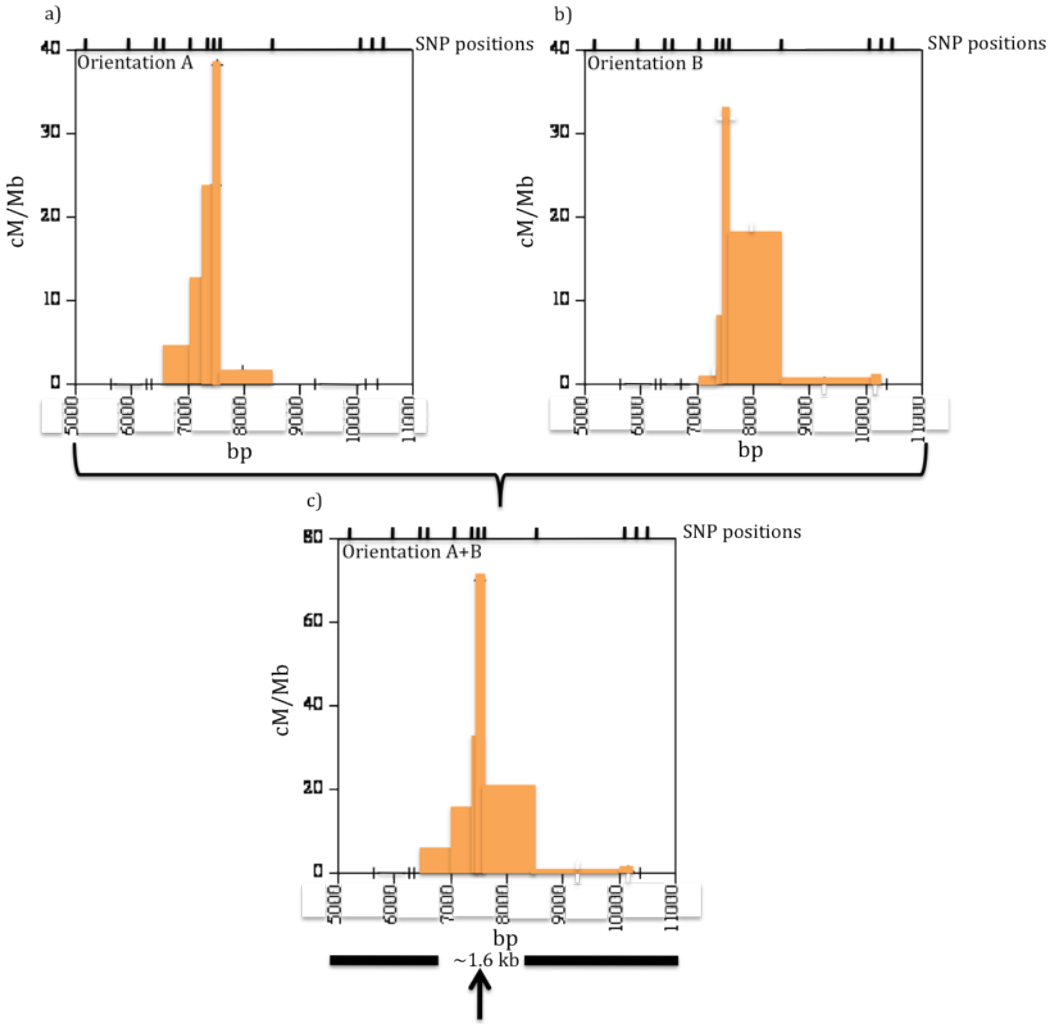


Figure 4.8 Crossover distributions in donor 7 in hotspot DA. Crossovers were scored for each SNP interval, and the peak activity calculated in cM/Mb. This value was plotted against the distance, with each column in the graph representing the width of the SNP interval. Informative SNP positions are shown as short lines above each graph. a) orientation A. b) orientation B. c) orientation A+B. The black arrow indicates the centre point of the hotspot for donor 7.

4.5.2 Distribution of Crossovers and the Centre of the Hotspot

The position of the centre point of the hotspot was estimated by the programme that written by Alec J. Jeffreys, University of Leicester, UK. The centre point for donor 7 was estimated at position 7601, and the width within which 95% of crossovers took place was estimated at approximately 1.6 kb. Analysis of cumulative crossover frequency distributions across the target interval (Figure 4.9) showed that hotspot DA was active in donor 7. Also, the cumulative frequencies of orientations A, B and combined A+B crossovers for donor 7 are shown in Figure 4.8. The A+B distributions define a typical narrow crossover hotspot ~1.4 kb wide. The separate A and B crossovers show major asymmetry in donors 7, with A and B cumulative distributions shifted by 330 bp. Thus, one allele of the motif-disrupting SNP suppressed the hotspot activity.

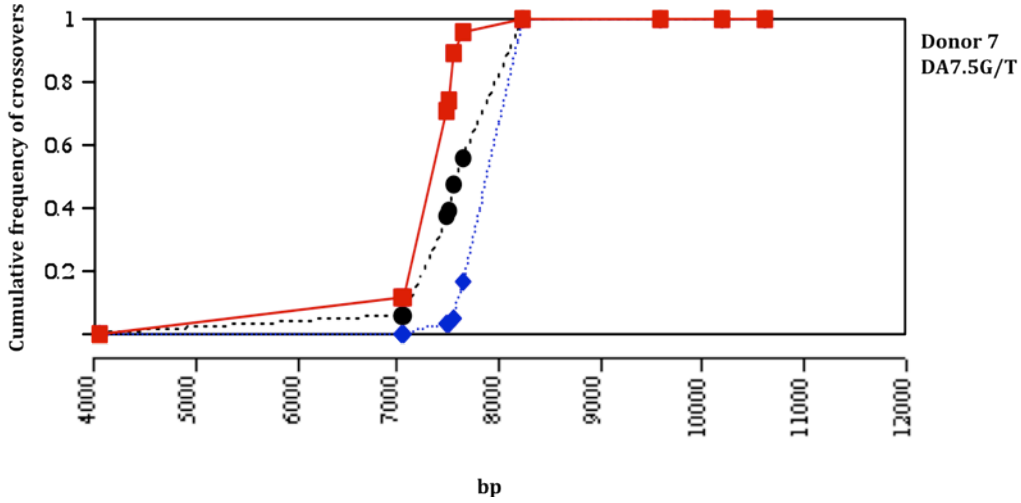


Figure 4.9 This figure explains how recombination breakpoint asymmetry around the hotspot arises as a consequence of donor/recipient asymmetry. The cumulative frequency values for orientation A (red), and B (blue) for each donor are shown. The fit curve for orientations A+B (black) are shown for each donor. The curves for both donors show a similar shape. Cumulative distributions shifted by 330 bp for A and B crossovers for donor 7.

4.5.3 Biased Gene Conversion in Crossover Progeny

Markers within the hotspot showed massive over-transmission (Figure 4.10) with central markers (DA7.5G/T and DA7.5aC/T) showing 90:10 transmission ratios into crossover progeny in donor 7. This biased gene conversion is maximal for the motif-disrupting SNP located just 90 bp from the hotspot centre. Donor 7 over-transmitted the disrupting allele T into crossovers, consistent with motif disruption down-regulating crossover initiation.

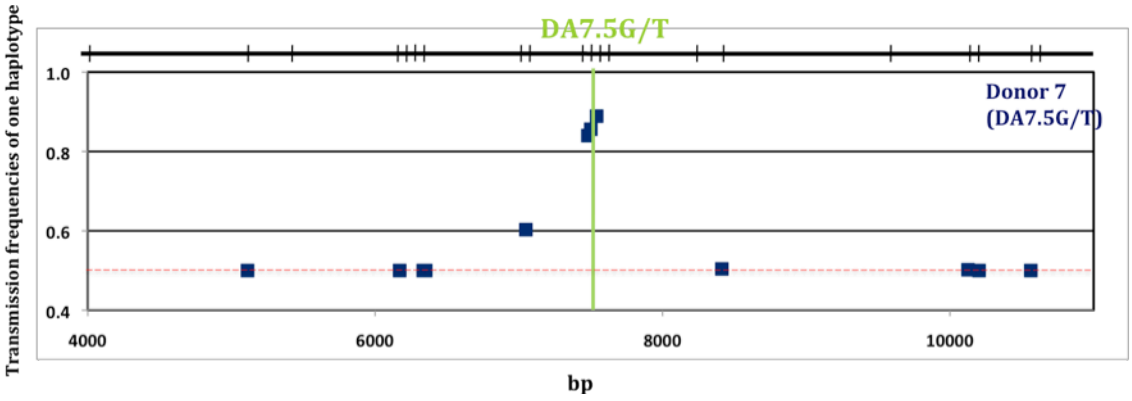


Figure 4.10 Transmission of alleles from one haplotype into crossover progeny in donor 7. Massive over-transmission was observed for the markers within the hotspot. Central markers (DA7.5G/T and DA7.5aC/T) showed 90:10 transmission ratios into crossover progeny in donor 7. This biased gene conversion is maximal for the motif-disrupting SNP located just 90 bp from the hotspot centre (the location of the motif-disrupting SNP DA7.5G/T is shown by the green line).

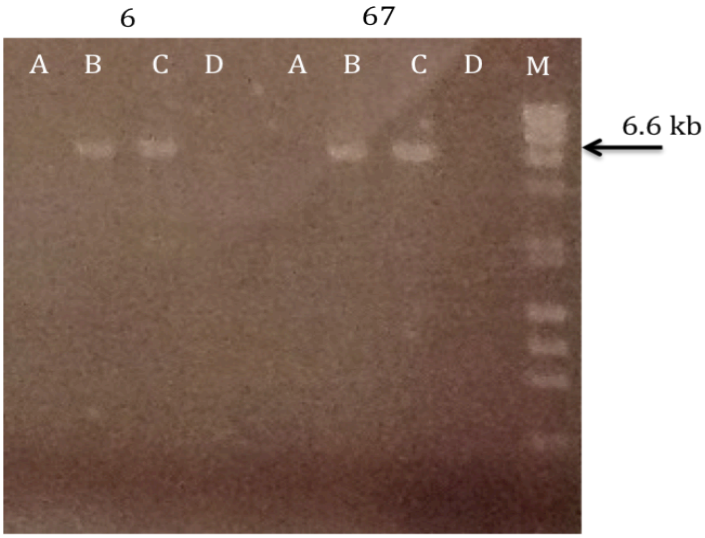
4.6 RECOMBINATION ASSAY IN MORE INFORMATIVE DONORS

For further information about the role of the disrupting SNP within the motif at the centre of the hotspot, more donors who were heterozygous or homozygous for the disrupting SNP were assayed at Hotspot DA.

4.6.1 Recombination Assay Performed on Sperm DNA of Donor 6

Four more donors were selected for the recombination assay for Hotspot DA. As with donor 7, donor 6 was heterozygous (G/T) for the disrupting SNP 7.5 within the motif. In addition to this donor, two homozygous donors (donor 33 and donor 67) for the G-

allele and one homozygous donor (donor 44) for the T-allele at the disrupting SNP DA7.5 were selected. Linkage phasing was required for donors 6 and 67. Four different combinations of forward allele-specific primers DA4.0aFG/A and reverse allele-specific primers DA10.6bC/A were used for linkage phasing (PCR conditions and primers are shown in Appendix II) (Figure 4.11).



A) DA4.0aFG + DA10.6bRC
B) DA4.0aFG + DA10.6bRA
C) DA4.0aFA + DA10.6bRC
D) DA4.0aFA + DA10.6bRA

Figure 4.11 Linkage phasing electrophoresis results for donors 6 and 67. Four combinations of allele-specific primers are shown above. Electrophoresis results give us information on the phasing for each donor; for example, donor 6 showed positive PCR reaction on B and C. Thus, the haplotype phasing of donor 6 is G – A, A-C). (M: marker λ*Hin*dIII and ϕ*XHae*III).

The first informative heterozygous marker within the interval was SNP DA7.0. There was 300 bp between the second forward selector site DA4.0a and first SNP heterozygote marker DA7.0. Possible positive crossovers would be missed out. Therefore, ASP DA4.0a was included as an informative heterozygote marker and previously optimized ASPs DA3.7FG1/T1 were selected as a second forward allele-specific primer. There was strong linkage between SNP DA3.7 and SNP 4.0a, thus linkage-phasing results considered while designing the crossover assay. Sperm DNA from donor 6 was amplified for both orientations A and B, that consisted of 24 x 1320 molecules, 24 x 1890 molecules, 24 x 2450 molecules and 24 x 3400 molecules, using

allele-specific primers, DA2.6FG and DA12.0RT for primary PCR and DA3.7FT1 and DA10.6bRC for secondary PCR for orientation A. And, for orientation B; DA2.6FC and DA12.0RC for primary PCR and DA3.7FG1 and 10.6bRA allele-specific primers were used. 66 positive molecules from orientation A and 60 positive molecules from orientation B have been detected in a total of 217,000 for both orientations. According to the secondary PCR gel results, the estimated recombination rates for orientation A and B, respectively were 8.0×10^{-4} (95% CI: 5.0×10^{-4} – 13.0×10^{-4}) and 6.0×10^{-4} (95% CI: 4.0×10^{-4} – 13.0×10^{-4}) recombinants per molecule, or 1 crossover in 1200 molecules for orientation A and 1 crossover in 1700 molecules for orientation B (these recombination frequencies were calculated and Poisson-corrected after crossover mapping) (Figure 4.12). The PCR products were then sequentially hybridised, first one allele then the opposite allele, for all informative SNPs in the target interval.

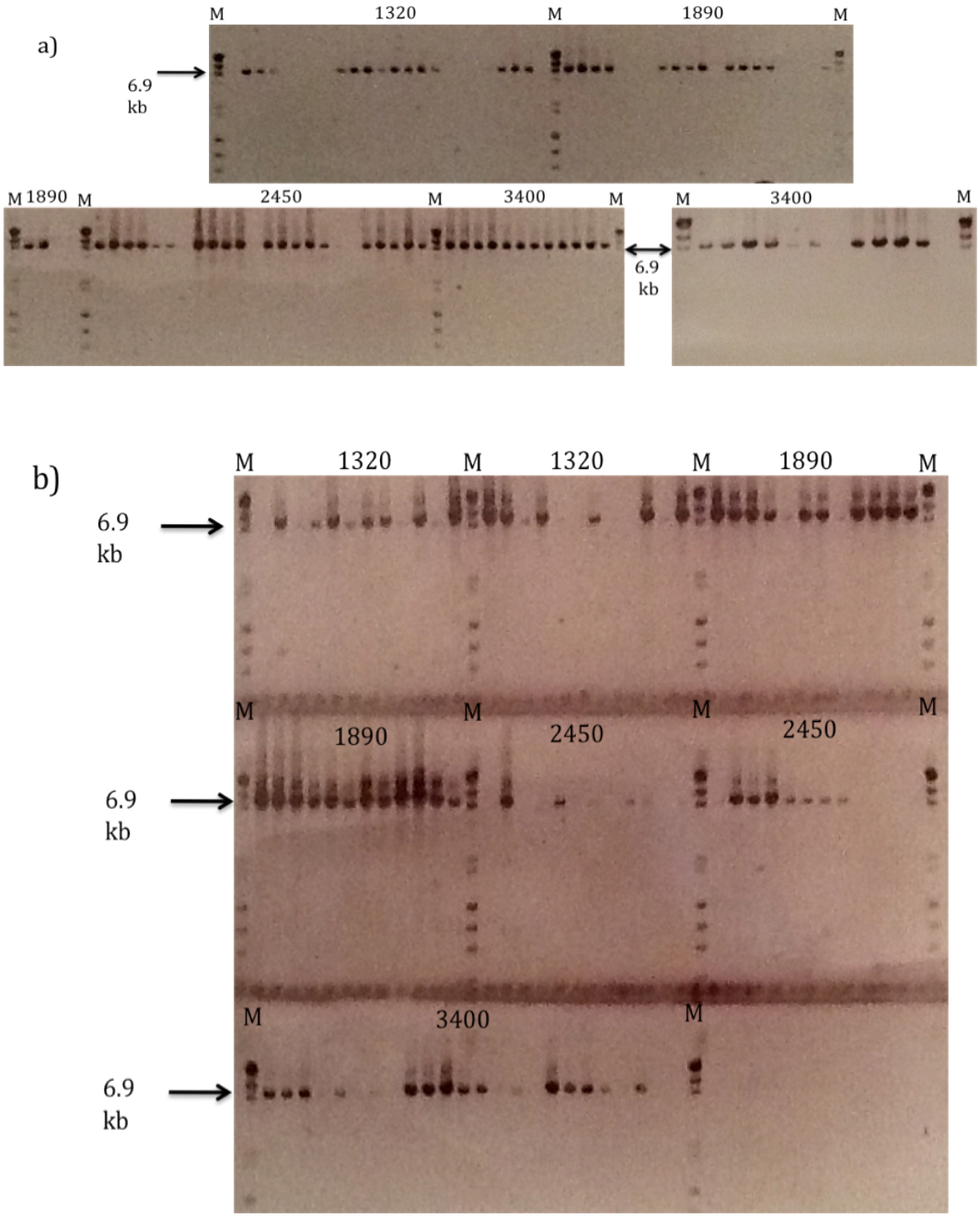


Figure 4.12 Recombination assay carried out on donor 6. Recombinant molecules were amplified in repulsion phase using allele-specific primers. Reactions were carried out on sperm DNA in pools of 24 x 1320 molecules, 24 x 1890 molecules, 24 x 2450 molecules and 24 x 3400 molecules for both orientations. Respectively a) shows the gel results from orientation A, and b) shows the gel results of orientation B. (M: marker λ*Hin*dIII and ϕ*XHae*III).

It was necessary to confirm whether the recombination activity seen at the hotspot DA was fully reciprocal, as well as crossovers occurring from one haplotype to the other haplotype (orientation B), and *vice versa*.

4.6.2 Recombination Assay Performed on Other Semen Donors

The best way to understand the effect of disrupting SNP within the motif of the Hotspot DA is to undertake a crossover asymmetry assay. If the disrupting SNP is present in both the active and suppressed allele (i.e. G/T) for Hotspot DA, mapping of crossover assays after dot blot hybridization should show crossover asymmetry. However, to confirm the contribution of the disrupting SNP to hotspot activation/inactivation, donors that are homozygous for the G and T alleles for the motif disrupting SNP were picked. Since previous studies (Jeffreys and Neumann, 2002; Jeffreys and Neumann, 2005; Jeffreys and Neumann, 2009) showed that homozygote alleles for cis-regulator SNPs, for the hotspot activation, did not show crossover asymmetry, thus, recombination assay was performed on homozygote donors 33, 67, 44 and 55 for the motif disrupting SNP.

Donor 33 was homozygous for the disrupting G-allele at SNP DA7.5 within the motif, this donor had 10 nicely spread informative SNP sites in the target interval. Both recombinant phases were amplified, using primary PCR primers DA2.6FG and DA10.6aRA and secondary primers DA4.0aFG and DA10.6RG2 for orientation A; and primary PCR primers DA2.6FC and DA10.6aRC and secondary primers DA4.0aFA and DA10.6RA2 for orientation B. Input pool sizes were 24 x 373 molecules, 24 x 561 molecules, 24 x 831 molecules and 24 x 1246 molecules for both orientations. All secondary products were re-amplified with tertiary universal primers. The rate of orientation A crossovers calculated from the number of putative crossover positive PCRs and the number of negative PCRs is approximately 1.3×10^{-3} (95% CI: $0.7 \times 10^{-3} – 2.4 \times 10^{-3}$)crossovers per molecule, or 1 crossover in 770 molecules. The rate of orientation B crossovers, also calculated from the number of putative crossover positive and negative PCRs is 1.0×10^{-3} (95% CI: $0.5 \times 10^{-3} – 2.2 \times 10^{-3}$) crossover per molecule or 1 crossover in 980 molecules.

As well as donor 33, second donor 67 was homozygote for allele G for disrupting SNP DA7.5. Donor 67, was also performed with 7 informative SNP markers in the target interval. For orientation A, DA2.6FG and DA10.6bRC as primary ASP primers and DA4.0aFG and DA10.6aRA as secondary ASP primers were used. Additionally, ASP primers DA2.6FC and DA10.6bRA for primary PCRs and ASP primers

DA4.0aDA and DA10.6aRC for secondary PCR were used for orientation B. 24 x 473 molecules, 24 x 710 molecules, 24 x 1050 molecules and 24 x 1575 molecules were the pool sizes for both orientations. Recombination frequencies were calculated for orientation A and B respectively, 1,4 x 10^{-3} (95% CI: 0.7 x 10^{-3} – 2.7 x 10^{-3}) and 1.6 x 10^{-3} (95% CI: 0.8 x 10^{-3} – 3.0 x 10^{-3}) or 1 crossover in 675 molecules for orientation A and 1 crossover in 620 molecules for B.

Donor 44 was homozygous for allele T for disrupting SNP. He was the best possible donor to perform for this assay among all European donor panels. Donor 44 had 9 informative SNP sites in the target interval. Both recombinant phases were amplified, one using primary PCR primers DA2.6FG and DA12.0RC and secondary primers DA4.0aFG and DA10.6RA2 for orientation A and primary PCR primers DA2.6FC and DA10.20RT and secondary primers DA4.0aFA and DA10.6RG2 for orientation B. Input pool size was 48 x 5200 molecules for both orientation A and B. There were no recombinant molecules observed in both orientations. In this manner, donor 44, that homozygote for allele T for disrupting SNP turned off the hotspot activity at the Hotspot DA. And, the highest recombination rates were estimated from donors 33 and 67 (G/G homozygous). Thus, G allele is an active allele for disrupting SNP DA7.5G/T and initiates the hotspot activity. Allele T is a suppressed allele and turns off the hotspot activity. Figure 4.13 shows active and suppressed haplotypes for hotspot DA. All PCR products resulting from the tertiary PCRs were applied to eight replica membranes and sequentially hybridised with both alleles of all informative SNPs (PCR conditions and all primers are shown in Appendix II)

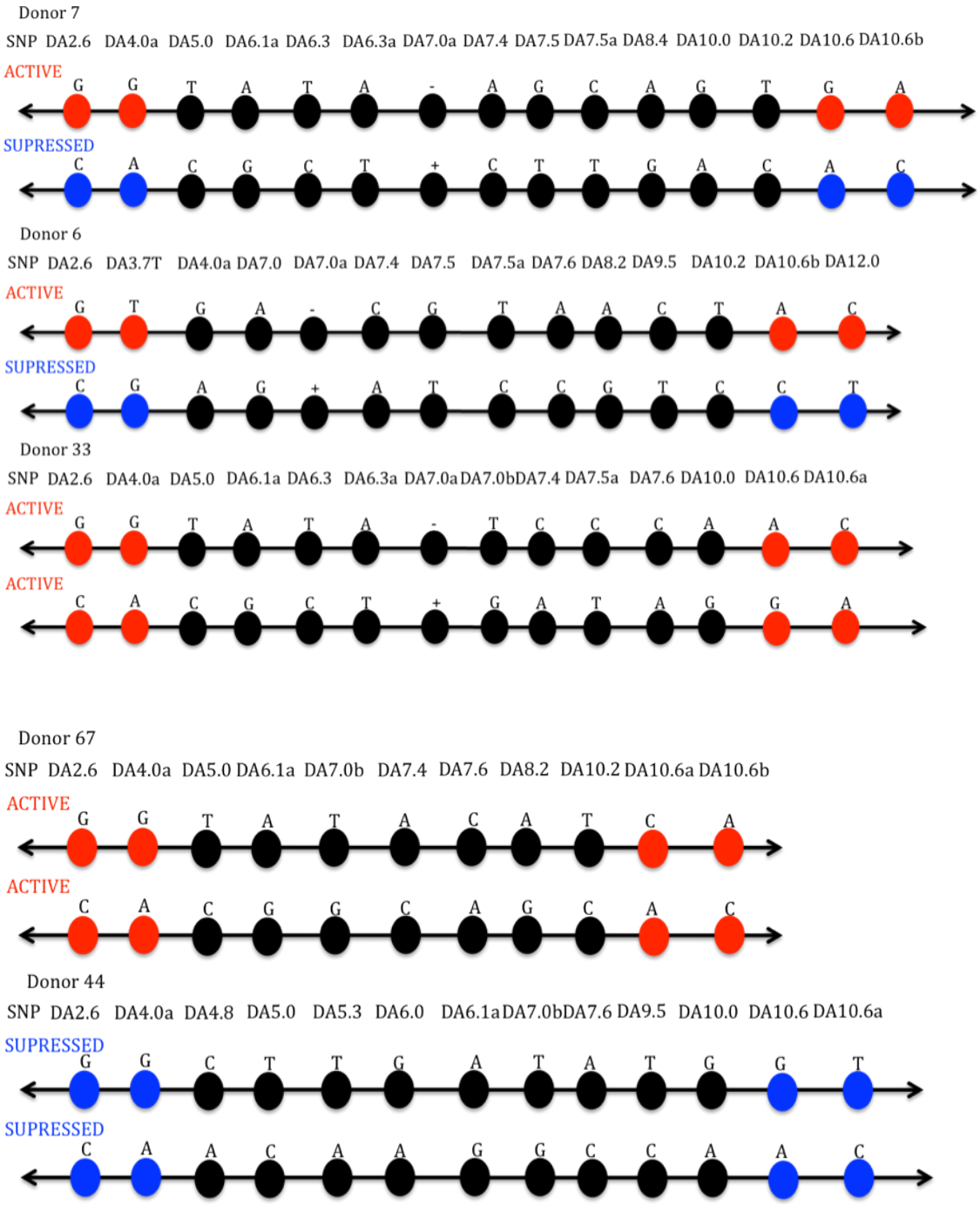


Figure 4.13 All haplotypes of four donors. According to reciprocal crossover assay; Donors 7 and 6, heterozygous (G/T) for the motif disrupting SNP had one active and one suppressed haplotype. The donors (33 and 67) were homozygous (G/G) for the motif disrupting SNP had two active allele, and additionally donor 44 that homozygous (T/T) for the motif disrupting SNP had two suppressed haplotypes.

4.6.3 Calculating Crossover Activity

Table 4.2 shows a summary of recombination assays performed on donors 6, 33 and 67, in both orientations, and the numbers of crossovers typed for each orientation. The recombination assay data is not shown for donor 44, because there was no single positive reaction in the assays.

Donor 6

SNP	location	distance (bp)	No. A XOs	No. B XOs	cM/Mb, AXOs	cM/Mb, BXOs
DA4.0a	4044		0	0		
DA7.0	7025	2980	21.3	0	3.1	0.0
DA7.0a	7049	23	0	0	0.0	0.0
DA7.4	7481	431	108.4	4.4	109.3	4.4
DA7.5	7501	19	6.3	0	144.1	0.0
DA7.5a	7542	40	27.3	2.2	296.6	23.9
DA7.6	7631	88	12.9	15.5	63.7	76.5
DA8.2	8229	597	7.4	114	5.4	83.0
DA9.5	9580	1350	0	0	0.0	0.0
DA10.2	10203	622	0	0	0.0	0.0
DA10.6b	10614	410	1	0	1.1	0.0
		Poisson corrected total COs	184.6	136.1		

Donor 67

SNP	location	distance (bp)	No. A XOs	No. B XOs	cM/Mb, AXOs	cM/Mb, BXOs
DA 5.0	5112		0.0	0.0		
DA 6.1a	6171	1058	0.0	0.0	0.0	0.0
DA 7.0b	7140	968	5.4	15	6.1	16.9
DA 7.4	7481	340	20.3	33.4	65.4	107.8
DA 7.6	7631	149	20.5	43	150.6	316
DA 8.2	8229	597	53	45.3	97.2	83
DA 10.2	10203	1973	0.0	0.0	0.0	0.0
DA 10.6a	10604	400	0.0	0.0	0.0	0.0
		Poisson corrected total COs	99.2	136.7		

Donor 33

SNP	location	distance (bp)	No. A XOs	No. B XOs	cM/Mb, AXOs	cM/Mb, BXOs
DA6.1a	6171		0.0	0.0		
DA6.3	6335	163	0.0	0.0	0.0	0.0
DA6.3a	6352	16	0.0	0.0	0.0	0.0
DA7.0a	7049	966	7	5	10.2	10
DA7.0b	7140	90	0.0	0.0	0.0	0.0
DA7.4	7481	340	24	17	97.6	69
DA7.5a	7542	60	19	13	438	300
DA7.6	7631	88	29	26	456	409
DA10.0	10126	2494	15	12	8.4	7.0
DA10.6	10563	436	0.0	0.0	0.0	0.0
DA10.6a	10604	40	0.0	0.0	0.0	0.0
		Poisson corrected total COs	94	73		

Table 4.2 Distribution of crossovers in the Hotspot DA. The name of each informative SNP site in the target interval for each donor, their locations and their distance (bp) to next marker, number of crossovers for orientation A, number of crossover for orientation B, calculated recombination activity in cM/Mb for orientation A and orientation B (separately) and Poisson corrected number of crossovers for both orientation A and B are also shown above. The number of crossover showed differences between orientation A and B in some donors. This means, mixture PCR reactions were observed in some crossovers and these mixture reactions carried more then one molecules.

4.6.4 Distribution of Crossover and The Centre of The Hotspot

Table 4.3 shows the summary of the centre point of the hotspot for four donors. The centre of the hotspot was calculated by the programme written by Alec Jeffreys, University of Leicester, UK.

Donors	Centre of the Hotspot (bp)	95% CI, width (kb)
7	7601	1.6
6	7583	1.3
67	7597	1.0
33	7532	1.0

Table 4.3 The summary of the centre of the hotspot for four donors. All four donors showed very similar centres; the small differences could result from sampling differences.

Analysis of cumulative crossover frequency distributions across the target interval (Figure 4.14) showed that hotspot DA was active in all four donors, but not in donor 44, who is homozygous (T/T) for the motif-disrupting SNP. Moreover, these donors (7. 6. 67 and 33) showed varying crossover frequencies. Also, the cumulative frequencies of orientations A, B and combined A+B crossovers for each donor are shown in Figure 4.14. The A+B distributions are very similar in all four donors and define a typical narrow crossover hotspot 1.4 kb wide.

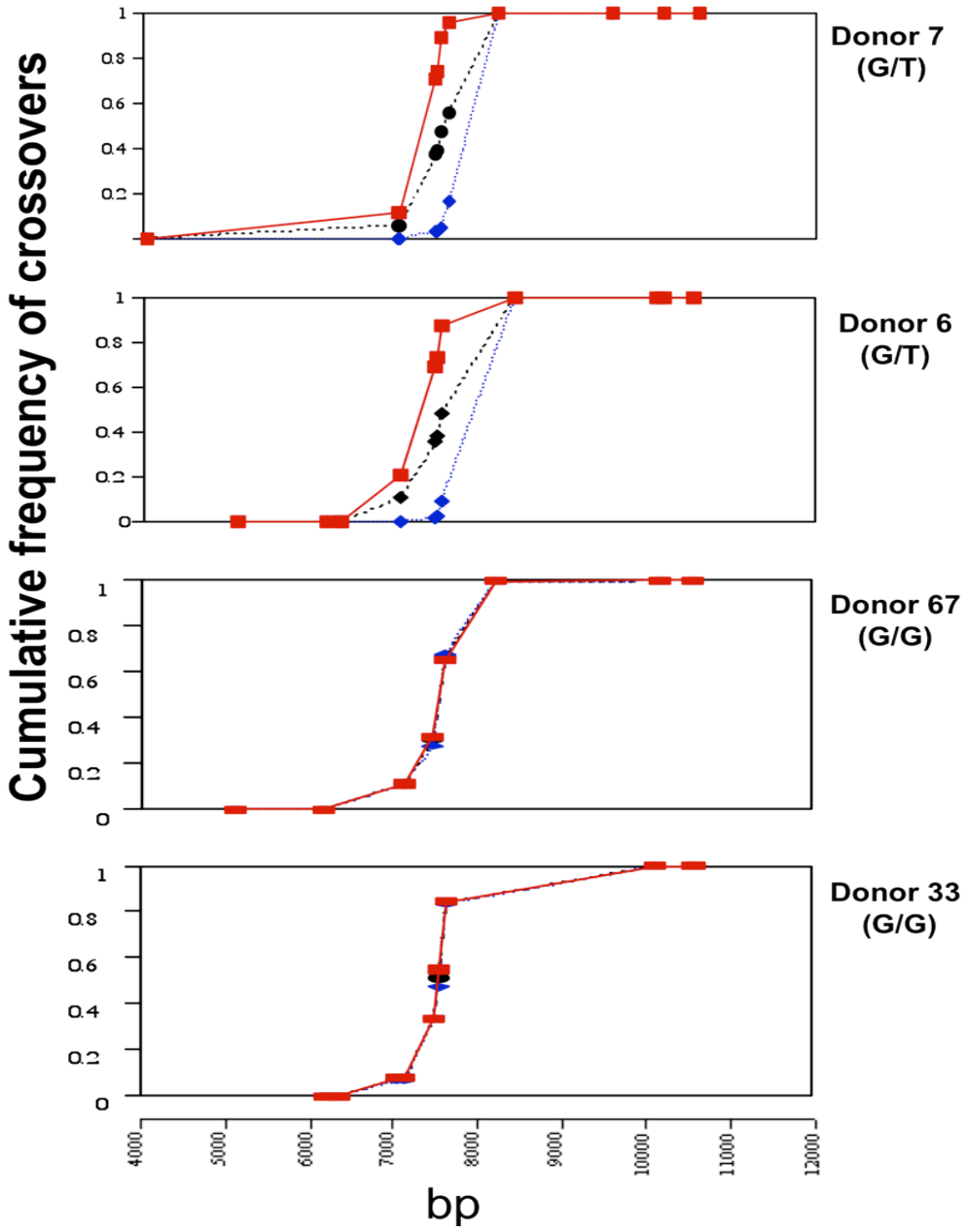


Figure 4.14 Cumulative frequency of crossovers for Hotspot DA. The cumulative frequency values for orientation A (red), and B (blue) for each donor are shown. The best fit curve for orientations A+B (black) is shown for each donor. The curves for all four donors show a similar shape. Cumulative distributions shifted by 330-450 bp for A and B crossovers for donors 7 and 6. The crossover asymmetry was not observed at donors 67 and 33.

The separate A and B crossovers show major asymmetry in the heterozygous donors 7 and 6, with A and B cumulative distributions shifted by 330-450 bp. This asymmetry is not present in the homozygous donors 67 and 33, and therefore appears to result from heterozygosity for an intact and disrupted 13-bp sequence motif, consistent with a major reduction in recombination initiation efficiency triggered by motif disruption.

4.6.5 Biased Gene Conversion In Crossover Progeny

Markers within the hotspot showed massive over-transmission (figure 4.15) with central markers (DA7.5G/T and DA7.5aC/T) showing 90:10 transmission ratios into crossover progeny in donor 7 and donor 6. A two-tailed Fisher exact test was performed to check if the numbers were significantly different from the 50:50 (Mendelian transmission) ratio and is showed statistically extremely significant (P = 0.0001). This biased gene conversion is maximal for the motif-disrupting SNP located just 90 bp from the hotspot centre. Donor 7 and donor 6 both over-transmit the disrupting T-allele into crossovers, consistent with motif disruption down-regulating crossover initiation. Transmission distortion in donor 33 is much less marked, with central marker (DA7.5aC/T) showing at most a 53:47 transmission ratio (P = 0.7773, statistically not significant), donor 67 showed 51:49 (P = 1.000) transmission ratio at the central markers (DA7.4A/C and DA7.6C/A).

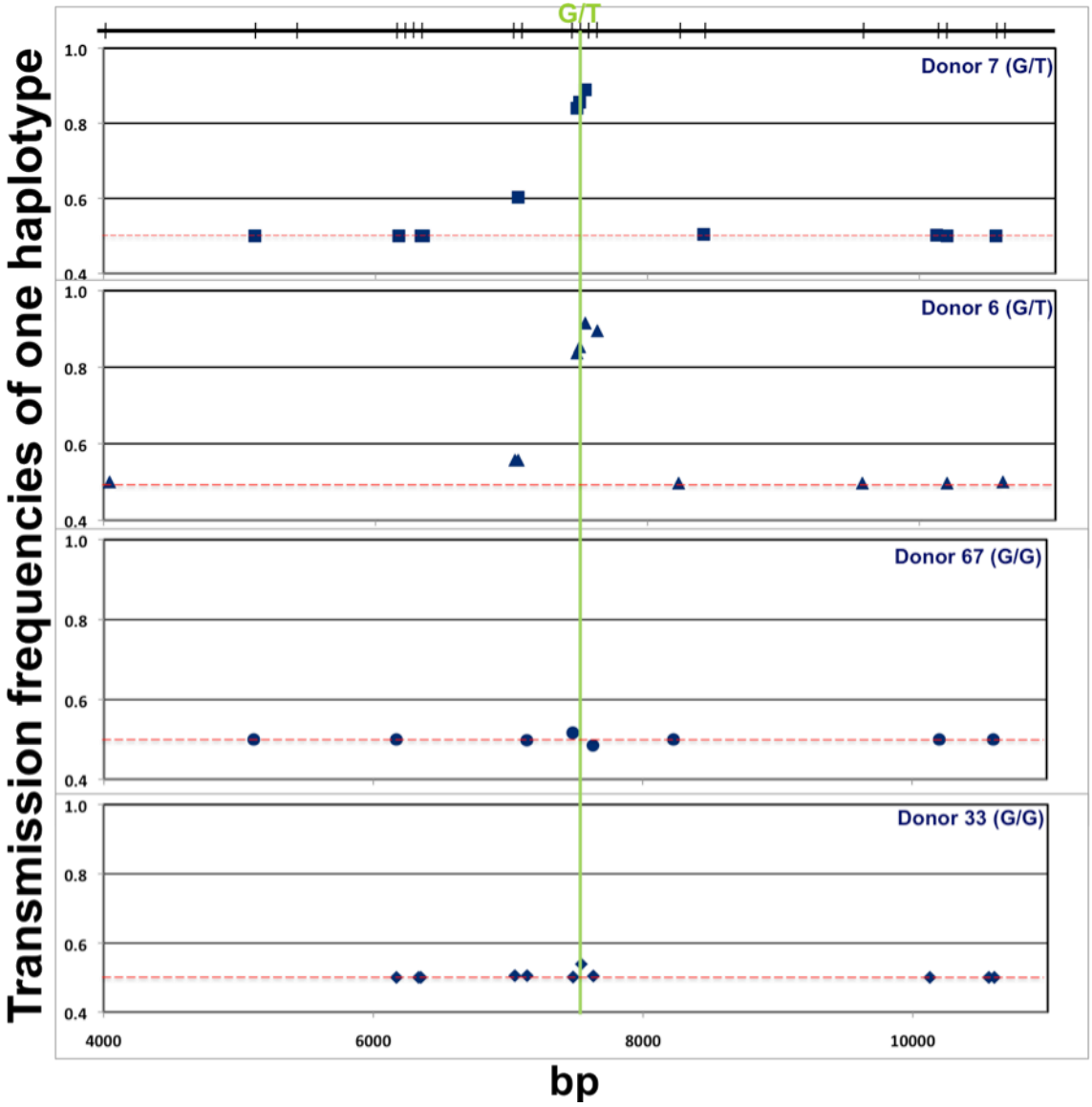


Figure 4.15 Transmission of alleles from one haplotype into crossover progeny. Central Markers (DA7.5G/T and DA7.5aC/T) showed 90:10 transmission ratios into crossover progeny in donor 7 and donor 6. This biased gene conversion is maximal for the motif-disrupting SNP located just 90 bp from the hotspot centre. However, transmission distortion in donor 33 and donor 67 is much less marked, with central markers showing at most respectively 53:47 and 51:49 transmission ratios. The location of the disrupting SNP DA7.5G/T is shown by the green line.

4.7 GENE CONVERSION AT THE HOTSPOT DA

A non-exchange gene conversion assay for Hotspot DA was carried out by Alec Jeffreys, University of Leicester, UK, while this thesis was being written up.

4.7.1 Strategy of Recovering Crossovers and Non-crossovers

The strategy for detecting crossovers (COs) and non-exchange gene conversions (NCOs) is summarised in Figure 4.16. Donor 6 and donor 55 were the best candidates for this assay, since both tested donors were heterozygous for SNPs DA7.4 and DA7.5. Also, these SNPs are located close to the hotspot centre and should show the highest frequency of conversion compared with markers further from the hotspot centre. However, given the relatively low crossover frequency at hotspot DA, it is not possible to detect conversions at just one of these sites using the half-crossover assay approach.

SNPs DA7.4 and DA7.5 are separated by only 19 bp and should therefore co-convert in the majority of NCO events that span at least one of these sites. The strategy was therefore to identify and map DA7.4/DA7.5 co-conversion events in pools of sperm DNA using nested allele-specific PCR. This approach will miss conversion tracts that include only one of these SNPs, as well as tracts that do not include either site, so NCO frequencies will be under-estimated; this is however true for all NCO assays, since markerless tracts will always go undetected.

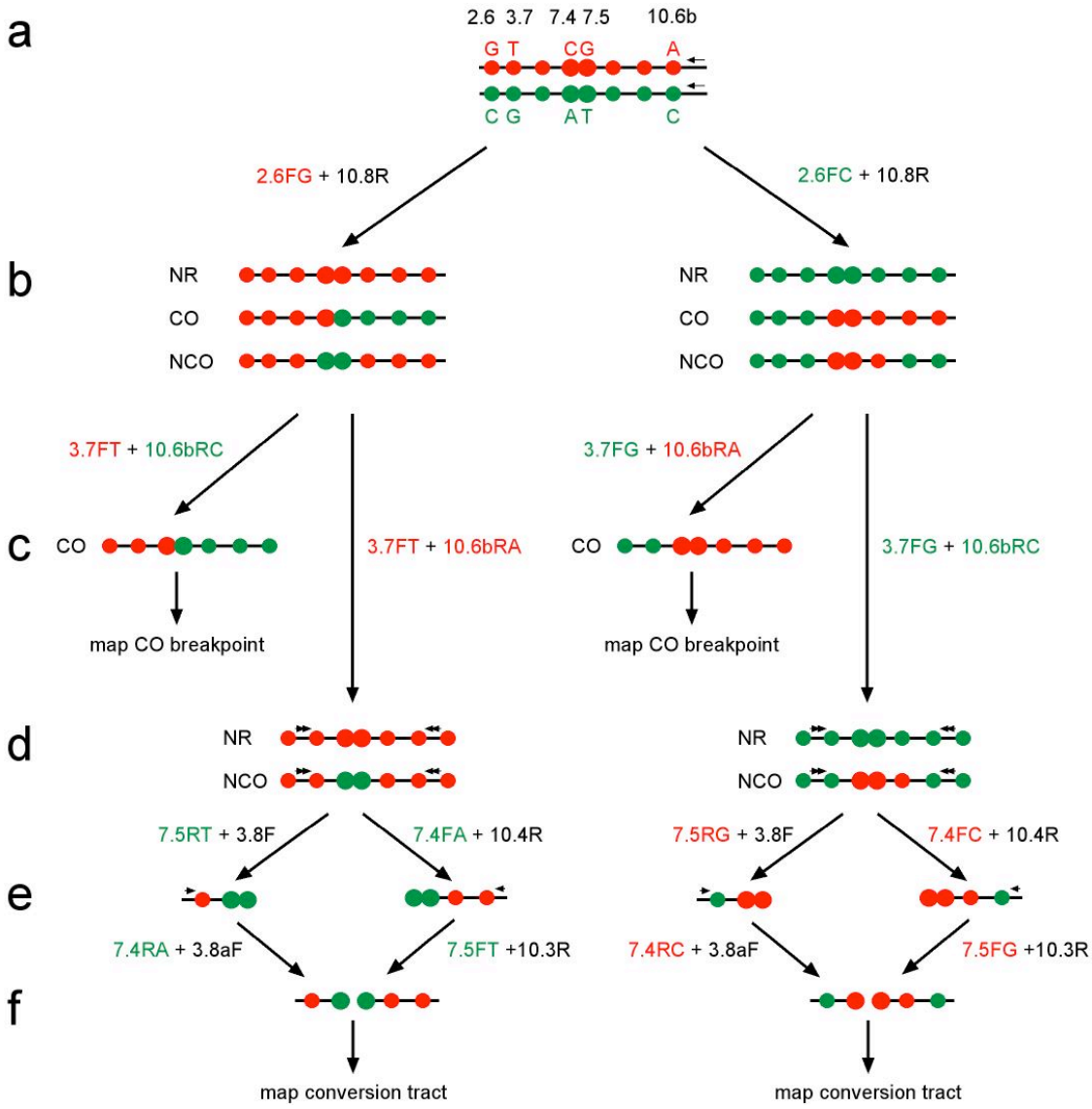


Figure 4.16 The strategy for detecting crossovers (COs) and non-exchange gene conversions (NCOs) at Hotspot DA. a) Progenitor haplotypes in the two donors (6 and 55) tested. Both showed the same linkage phase for SNPs DA2.6, DA3.7 and DA10.6b used to separate COs and NCOs, and also for SNPs DA7.4 and DA7.5 used to detect NCO events spanning these sites. Universal (non-allele-specific primers) also used in this analysis are shown as black arrows. b) Two separate primary PCRs were used on multiple pools of sperm DNA containing DNA inputs ranging from 150 to 1500 amplifiable molecules of each haplotype, and also on blood DNA from donor 6, across two 96-well plates per man per orientation. 63,000 sperm molecules were scored in each orientation for donor 6, and 156,000 molecules for donor 55. These primary PCRs used one or other allele-specific primer directed to SNP DA2.6 in combination with a downstream universal primer. PCR products will be enriched for one or other type of non-recombinant progenitor haplotype molecules (NR) plus different types of any COs and NCOs present, depending on the SNP DA2.6 primer used. All further analyses were carried out on these primary PCR products. c) Re-amplification with nested allele-specific primers in repulsion

phase, as in a regular crossover assay, selectively amplified CO molecules. The specificity of ASPs used for SNP 10.6b was good enough for CO molecules to be detected and mapped directly by ASO hybridisation to these secondary PCR products, exactly as for regular CO mapping. d) Re-amplification of the same primary PCRs with nested allele-specific primers in coupling phase resulted in NR and NCO molecules being amplified and CO molecules being eliminated. e,f) NCOs with switches at SNPs DA7.4 and DA7.5 were selectively amplified from these coupling-phase secondary PCRs by nested PCR using DA7.4 plus DA7.5 ASPs in combination with universal primers as indicated. Forward ASPs selectively amplified the 3′ end of NCO molecules, while reverse ASPs amplified the 5′ end of the same molecules. Thus each NCO was recovered in two pieces. Conversion tracts were then mapped by ASO hybridisation of these two pieces. This approach gave good specificity and yielded NCOs, detectable by agarose gel electrophoresis, from sperm DNA but not from blood DNA (no events seen in 63,000 blood molecules screened from donor 6), providing good evidence that these NCOs are genuine and of meiotic origin. Note that COs and NCOs were analysed in the same primary PCRs (b) and were, as expected, distributed across PCRs independently of each other. (Figure drawn by Alec jeffreys, University of Leicester, UK).

4.7.2 Crossover and Non-crossover Frequencies

The crossover frequency for donor 6 was higher than previously measured (0.152% vs. 0.070%), reflecting maybe variation in DNA quality and/or quantity, or primer efficiency. However the crossover frequency for donor 55 showed avarage (0.062%) frequency for Hospot DA (Table 4.4).

NCO frequencies were low (Table 4.4), at about 10% of the CO frequency in both donors. Previously, studies showed a wide range of NCO: CO ratios, with typical values of ~1:3 at other hotspots. The NCO frequency at hotspot DA seems unusually low, though this might well reflect missing NCOs since only those that had co-converted at SNP DA7.4 and DA7.5 were scored. There is no significant difference between the two donors in the ratio of NCOs to COs (Fisher exact test, $P = 0.14$).

Donors	Sperm molecules screened	Number of COs	CO frequencies, %	Number of NCOs	NCO frequencies, %	NCOs:COs
6	126,000	192	0.152	15	0.0119	1: 12.8
55	312,000	193	0.062	26	0.0083	1: 7.4

Table 4.4 Poisson-corrected numbers of recombination eventx for donors 6 and 55 are shown above.

4.7.3 Reciprocal Crossover and Conversion Asymmetry

The centre point of the hotspot for donor 6 was estimated at position 7564 and the width within which 95% of crossovers took place was estimated at approximately 0.8 kb (previously the centre of donor 6 was estimated at position 7583 and the width was 1.3 kb). The sampling differences caused only minor differences in centre point location (19 bp) and width (0.5 kb) . However, the centre point for donor 55 was at position 7523 and the width 1.5 kb. Marker DA5.7G/T has transmitted the T allele and showed 80:20 for donor 6 and 86:14 for donor 55 statistically extremely significant (respectively, $P = 0.0001$, two-tailed Fisher exact test and $P = 0.0001$, two-tailed Fisher exact test) transmission ratios into crossover progeny. Marker DA7.5aC/T showed the highest transmission ratio (91:9) ($P = 0.0001$, two-tailed Fisher exact test) for donor 6 in both crossover assays. While donor 55 seems to have a broader distribution of crossover breakpoints, the basic features are much as observed in terms of hotspot centre, the strong over-transmission of allele DA7.5T (Figure 4.17). These transmission ratios confirmed previous results (Figure 4.15).

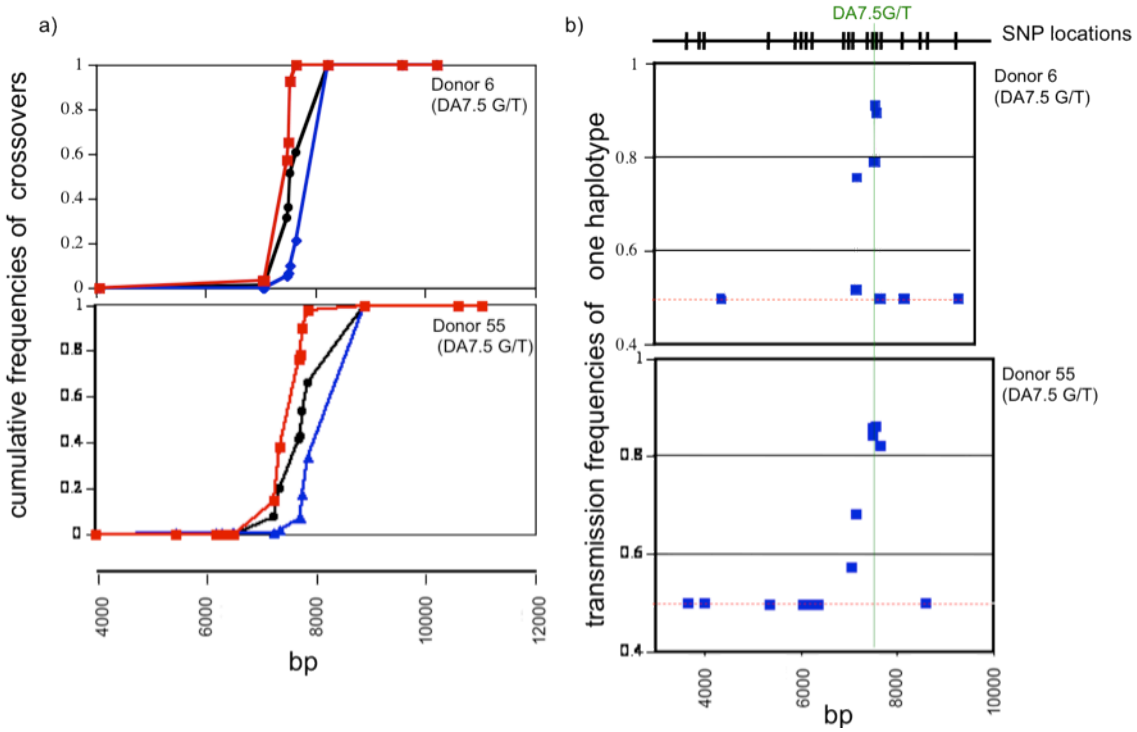


Figure 4.17 Cumulative frequencies of crossovers (a), and transmission of alleles from one haplotype into crossover progeny (b) in donors 6 and 55. a) The cumulative frequency values for orientation A (red), and B (blue) for donors are shown. The best fit curve for orientations A+B (black) are shown for each donor. The curves for both donors show a similar shape. Cumulative distributions are shifted by 380 bp for donor 6 and 495 bp for donor 55 for A and B crossovers. b) Transmission of alleles from one haplotype into crossover progeny in those donors normalised to equal frequencies of A and B orientations. Massive over-transmission was observed the markers within the hotspot. Central Marker DA7.5G/T showed 80:20 (donor6) and 86:14 (donor 55) over-transmission ratios into crossover progeny (the location of disrupting SNP DA7.5G/T is shown by a green line).

Conversion asymmetry

The non-crossovers showed complete bias, with all 41 NCOs (34 events before Poisson correction) resulting in the transfer of DA7.5T to the DA7.5G haplotype, and with no instances of G>T transfers. This extreme bias is highly significant (P = 1.2×10^{-10}) and is in the same direction as the bias seen in crossovers. Table 4.5 compares the transmission distortion seen in COs and NCOs:

	d6	d55	d6+d55
COs with 7.5G *	34	26	60
COs with 7.5T *	130	156	286
Observed NCOs with 7.5G	0	0	0
Observed NCOs with 7.5T	14	20	34
P bias in COs and NCOs same	0.075	0.083	0.0049
Poisson-corrected NCOs with 7.5G	0	0	0
Poisson-corrected NCOs with 7.5T	15	26	41
P bias in COs and NCOs same	0.078	0.051	0.0010

Table 4.5 Comparison of the transmission distortion seen in crossovers (COs) and non-crossovers (NCOs) in both donors 6 and 55. The table respectively shows the number of COs with alleles 7.5G and 7.5T, the number of observed NCOs with alleles 7.5G and 7.5T and P values of bias in COs and NCOs at the direction as crossovers, the number of Poisson-corrected NCOs with haplotypes 7.5G and 7.5T and P values (after Poisson correction) of bias in COs and NCOs for each donors 6 and 55, 6+55. (*Numbers normalised to equal numbers of A and B crossovers.)

For both donors 6 and 55, the strength of distortion is greater in NCOs than in COs. For each donor, this disparity is not significant (Fisher exact tests on numbers of 7.5G- and T-containing COs vs. numbers of G- and T-containing NCOs). However, it is significant for data from both donors combined, even on raw counts of numbers of NCOs before Poisson correction. It therefore appears that NCOs show a significantly stronger bias towards acquiring 7.5T compared with COs.

4.7.4 Conversion Tract Lengths

40-48% of NCOs show co-conversion of one or more flanking markers, most particularly at SNP DA7.5a that lies just 41 bp away from SNP DA7.5. There are examples of quite long conversion tracts incorporating 5-6 SNPs in each donor, with minimum tract lengths of 607 and 583 bp in donor 6 and donor 55 respectively, though they could be much longer since there are not sufficient markers to narrow the location (e.g. the tract with minimum length 583 bp could be as long as 2237 bp).

The conversion data were selected for NCOs showing conversion at SNPs DA7.4 plus DA7.5, and could not be used to determine an overall conversion profile across the hotspot. SNP DA7.5a is of particular interest, as it seems to map closer to the centre

of the hotspot and might show a higher NCO activity. Nevertheless, the data can be used to determine the likelihood that a flanking SNP is co-converted in those NCO molecules known to have converted at DA7.4 plus DA7.5. These data, with 95% confidence intervals, are shown below for both donors; the locations of SNPs DA7.4 and DA7.5 are shown by dotted lines (Figure 4.18).

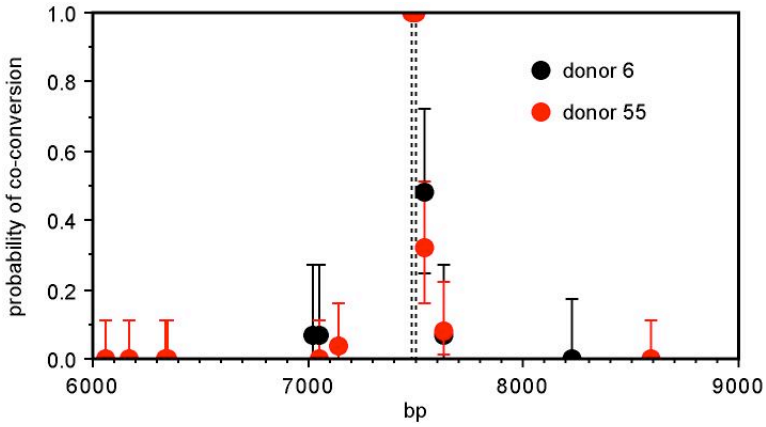


Figure 4.18 Probability of co-conversions. (This graph was drawn by Alec Jeffreys, University of Leicester, UK)

While the sample sizes are small, there is clear evidence for a very steep gradient, with the likelihood of co-conversion declining rapidly as one moves away from the selected sites DA7.4 and DA7.5. This fits with previous studies of NCO distributions in human hotspots (Sarbajna *et al.,* 2012).

4.7 DISCUSSION

A 15-kb target interval around Hotspot DA which was centred on the motif-disrupting SNP 7.5 within the 13-bp hotspot sequence motif was assayed for recombination activity in a total of six men. The crossover rates showed Hotspot DA to be a regular hotspot with an average crossover rate among hotspots assayed on autosomal chromosomes. Hotspot DA provides strong evidence that the motif is likely involved in promoting the initiation of recombination. Five assayed donors (6, 7, 33, 55 and 67) showed very similar hotspot locations for orientation A + B crossovers combined. This centre is displaced 3′ to DA7.5 by about 50-70 bp. The numbers of crossovers mapping 5′ and 3′ of DA7.5 were normalised to equal numbers of A and B crossovers, and a Fisher test was performed to check if the numbers were significantly

different from the 50:50 ratio predicted if DA.7.5 is located exactly at the centre. Donors 6, 7, 33 and 67 showed statistically significant crossovers mapping of DA7.5 was; the only exception was donor 55 where the shift was of borderline significance ($P = 0.054$, two-tailed Fisher exact test). Thus, DA7.5 and the motif were not located at the centre of hotspot DA. Sperm crossovers in men heterozygous for a motif-disrupting variant show the greatest transmission distortion ratio (~ 90-10) ever seen in a human crossover hotspot (Jeffreys and Neumann, 2002; Neumann and Jeffreys, 2006).

The transmission frequency results showed that the asymmetries in three heterozygous donors (donors 6, 7 and 55) for the motif disrupting SNP DA7.5G/T were very similar in terms of the transmission of DA7.5T to crossover progeny (90:10 transmission ratio) and in terms of the displacement of orientation A versus orientation B crossover distributions. Donors (33 and 67) that carried the active G-allele showed the highest recombination frequency respectively of 0.115% and 0.155% among analysed 6 donors. Donors 6 and 55 are both heterozygous for the disrupting SNP and showed similar recombination frequencies (0.07% and 0.06%). However another heterozygous for the disrupting SNP, donor 7 showed markedly different crossover frequencies (0.013%) than donors 6 and 55. This shows that there might be other affects (*cis*- or *trans*- factors) for this donor for the hotspot activity.

The direction of biased gene conversion indicates that the chromosome carrying the disrupting allele was suppressed for crossover initiation. This was supported by donor 44, who is homozygous for suppressing allele for the motif disrupting SNP (T/T). There were no single recombinant molecules observed in the crossover assay. The strength of biased conversion in favour of the suppressing allele, combined with the high activity of the hotspot, was sufficient to ensure its eventual population fixation, highlighting the necessarily ephemeral nature of recombination hotspots.

Gene conversion assays carried out on two donors (6 and 55) showed biased non-crossover frequencies to be low, at about 10% of the crossover frequency in both cases. However to date, observed relative rates of NCO: CO showed major variation between human hotspots (Holloway *et al.,* 2006). However, those studied hotspots showed a wide range of NCO: CO ratios were typically ~1:3 (Holloway *et al.,* 2006;

Sarbajna *et al.,* 2012). The non-crossover frequency at hotspot DA seems unusually low, though this might well reflect missing non-crossovers since we only scored those that had co-converted at SNP 7.4 and 7.5. There was no significant difference between the two donors in the ratio of non-crossovers to crossovers (Fisher exact test, $P = 0.14$). Nevertheless, the data can be used to determine the likelihood that a flanking SNP is co-converted in those non-crossover molecules known to have converted at 7.4 plus 7.5. While the sample sizes are small, there was clear evidence for a very steep gradient, with the likelihood of co-conversion declining rapidly as one moves away from the selected sites 7.4 and 7.5. This fits with previous studies of non-crossover distributions in human hotspots.

This suggests perhaps two different pathways for generating NCOs and COs, as seen in yeast and mice. The NCO pathway (maybe the SDSA model, (Petes, 2001)) would require that SNP 7.5 is always incorporated into a G/T heteroduplex in G/T heterozygotes and is always repaired in favour of the T-allele during e.g. early mismatch repair. The CO pathway would have to be different, either in terms of a lower likelihood that a G/T heteroduplex is generated, or in terms of a lower visibility of the heteroduplex to mismatch repair or a lower strength of bias in favour of T during repair. Also, a similar story is seen at a PAR2 hotspot (Sarbajna *et al.,* 2012). Gene conversion that favors the transmission of GC-alleles over AT-alleles, was not observed at Hotspot DA.

To sum up, Hotspot DA is the unique example to date of strongly direct *cis*-regulation for hotspot on/off polymorphism in which a disrupting SNP is located close to its centre. This is confirmed by our reciprocal crossover asymmetry and non-exchange bias gene conversion assays.

CHAPTER 5: EFFECT OF TRANS-REGULATOR FACTOR PRDM9 ON HOTSPOT DA

5.1 INTRODUCTION

In Chapter 4, variation in the crossover frequencies between men heterozygous for the disrupting SNP (DA7.5G/T) within the motif was observed using crossover assays. In particular, a low recombination frequency was observed for one of the heterozygous men. This result was unexpected, because of the high sequence similarity between all heterozygous men showing higher and lower crossover frequencies. This suggests that there might be other factors that affect the hotspot activity besides the *cis*-regulator SNP (DA7.5G/T). The question thus remained: are there any other elements within Hotspot DA that could influence hotspot initiation? Or, based on recent studies (Baudat *et al.*, 2010; Myers *et al.,* 2010; Berg *et al.,* 2010) are there any *trans*-regulators that may be responsible instead?

The *trans*-regulator factor PRDM9 has been identified as a regulator of hotspot usage in mice and humans (Baudat *et al.*, 2010; Myers *et al.,* 2010; Berg *et al.,* 2010). The protein encoded by the *PRDM9* gene consists of an N-terminal KRAB domain and a SET domain encoding a histone methyl-transferase followed by a tandem repeat zinc finger domain. The PRDM9 variant A has 13 C2H2 zinc-fingers (ZnF), each encoded by an 84 bp repeat of a minisatellite (Hayashi *et al.*, 2005), and is the most common variant in Europeans and Africans (Baudat *et al.*, 2010, Berg *et al.*, 2010). At positions -1, 2, 3 and 6 of the α-helix domain, every ZnF has four predicted DNA contact residues (Baudat *et al.*, 2010). Based on these residues, the last six ZnFs of the variant PRDM9 allele A are predicted to be responsible for binding to eight non-degenerate bases of the 13-mer CCNCCNTNNCCNC hotspot motif (Baudat *et al.,* 2010, Myers *et al.,* 2010, Berg *et al.,* 2010). This hotspot motif can be found in about 40% of human recombination hotspots (Myers *et al.,* 2008; Myers *et al.,* 2010).

The first direct evidence for *PRDM9* regulation in humans came from Hutterite individuals in a single pedigree carrying the variant PRDM9 allele I, which encodes a

ZnF array that cannot bind to this motif (Baudat *et al*., 2010). This study used linkage analysis to show a genome-wide shift in hotspots activity (Baudat *et al.,* 2010).

PRDM9 alleles have been genotyped in our semen donors by A. Jeffreys, R. Neumann and I. Berg (University of Leicester, Leicester/ UK) as described in (Berg *et al.,* 2010). PRDM9 variability was limited in Europeans, therefore ZnF alleles in a panel of 74 African semen donors and 156 European donors based on the number of ZnF repeats (8-18 ZnFs) were sequenced in our laboratory (Berg *et al.*, 2010). All reported PRDM9 alleles (A-E), except variant PRDM9 allele I, that were seen in the Hutterite population (Baudat *et al.,* 2010), were also present within our European semen donor panel. In addition, 24 new *PRDM9* alleles (L1-L24) was identified (Berg *et al.,* 2010). *PRDM9* variability was much higher in the African donor panel. Additionally, the classification of alleles according to the predicted ability of the resulting protein variants to recognise the hotspot motif showed that nearly half of these variants should result in impaired binding relative to the common A allele as shown in Figure 5.1 (Berg *et al.,* 2010).

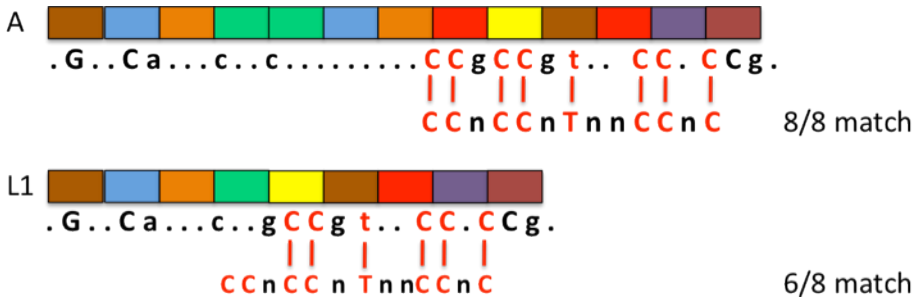


Figure 5.1 PRDM9 variants A and L1, with examples of tandem repeats encoding the ZnF array with variant repeat units coloured differently (modified from Berg *et al*., 2010). The predicted binding sequence is shown above, with dots indicating weakly predicted bases and uppercase letters indicating the most strongly predicted bases, and aligned with the hotspot motif CCNCCNTNNCCNC (Myers *et al.*, 2008). The binding sequence for the variant PRDM9 allele A matches all eight specified bases in the motif, while allele L1 matches at best only six of the eight bases.

Recent studies indicated that the *trans*-regulator PRDM9 has a strong effect on the activity of human sperm hotspots (Berg *et al.,* 2010, Berg *et al.,* 2011) and mouse hotspots (Baudat *et al.,* 2010, Parvanov *et al.,* 2010, Grey *et al.,* 2011, Brunschwig *et al.,* 2012), and even at hotspots lacking the sequence motif (Berg *et al.,* 2010). The changes within the ZnF can create hotspot non-activating or enhancing variants and

can even trigger the appearance of a new hotspot. Therefore, PRDM9 is a major global regulator of hotspots in mammals (Berg *et al.,* 2010, Berg *et al.,* 2011, Grey *et al.,* 2011, Brunschwig *et al.,* 2012). In chapter 4, *cis*-regulation was studied for hotspot DA; this hotpsot has a disrupting SNP (DA7.5G/T) within its motif sequence close to the centre of the hotspot, and strong *cis* influences were observed. According to recent studies, the question remained: what is the effect of the *trans*-regulator PRDM9 on the activity of Hotspot DA? Therefore, in this chapter, more men with variant PRDM9 genotypes were analysed to explore possible influences of PRDM9 status on crossover frequencies.

5.2 DESIGNING THE CROSSOVER ASSAYS

To test whether the *trans*-regulator PRDM9 has any influence on crossover frequencies at Hotspot DA, a more efficient crossover assay was used to analyse crossover frequencies. The crossover assay relies on nested repulsion-phase allele-specific PCR to selectively amplify crossovers from sperm DNA. The principle of the crossover assay has been explained in detail in the Materials and Methods chapter, but in summary, two sets of heterozygous selector sites located within LD blocks flanking both sides of the hotspot are identified and optimised for primers to allow for efficient crossover detection in as many men as possible. The crossover frequency results of multiple men would give clues about the factors that influence hotspot activation in *cis* and in *trans*. To allow as many PRDM9 variants as possible to be tested, 74 Afro-Caribbean and Zimbabwean (African) and 156 European men were investigated.

5.2.1 SNP Discovery and Annotation for the African Panel

As explained in Chapter 3, a 15-kb interval of DNA sequence spanning Hotspot DA was downloaded from ENSEMBL (www.ensembl.org). The sequence was annotated to include all SNPs (including both European and African) and repeat sequences (LINEs, Alus etc) (Appendix I) using information gained from Phase II HapMap data, dbSNP data and Repeat Masker. All identified 38 SNPs were genotyped using Allele Specific Oligo (ASO) hybridisation (see Materials and Methods for more details) (Table 5.1). 74 African semen donors were amplified on DA amplicons (PCR conditions and primers are shown in Appendix II).

Donors 134–239:

SNPs	Location	134	146	150	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	211	232	235	236	237	238	239
DA2.6G/C	2593	G	G	H	G	G	G	G	G	G	G	G	G	G	H	H	G	G	G	G	H	G	H	H	G	G	H	G	G
DA3.5T/G	3489	T	T	H	T	T	T	T	T	H	H	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
DA3.7T/G	3667	T	T	H	T	T	T	T	T	T	T	T	T	T	H	H	T	T	T	T	H	T	T	H	T	H	T	H	T
DA4.0aG/A	4044	G	G	H	G	G	G	G	G	G	G	G	G	G	H	H	G	G	G	G	H	G	H	H	G	H	G	H	G
DA4.8A/C	4777	A	A	H	A	A	H	H	A	C	C	A	A	H	A	A	A	H	A	A	A	H	A	A	H	A	H	A	A
DA5.OT/C	5112	T	T	H	T	T	T	T	T	T	T	T	T	T	H	H	T	T	T	T	H	T	H	H	T	T	H	T	T
DA5.3A/T	5368	A	A	H	A	A	H	H	A	T	T	A	A	H	A	A	A	H	A	A	A	H	A	A	H	A	H	A	A
DA6.0A/G	6057	A	A	H	A	A	H	H	A	G	G	A	A	H	A	A	A	H	A	A	A	H	A	A	H	A	H	A	A
DA6.1C/T	6139	H	C	C	C	H	C	H	T	C	C	H	T	C	H	H	H	C	T	H	H	H	H	C	H	H	H	H	C
DA6.1aA/G	6171	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	A	A	A	A	A	A	H	H	A	A	A	A	A
DA6.3T/C	6335	T	T	H	T	T	H	H	T	C	C	T	T	H	T	T	T	H	T	T	T	H	T	T	H	T	H	T	T
DA6.3aA/T	6352	A	A	H	A	A	H	H	A	T	T	A	A	H	A	A	A	H	A	A	A	H	A	A	H	A	H	A	A
DA7.0A/G	7025	H	A	H	A	H	A	G	G	A	A	H	G	A	H	H	H	A	G	H	G	H	H	A	H	G	G	G	A
DA7.0a-/+	7049	H	-	+	-	H	-	H	+	H	H	H	H	-	H	-	H	-	+	-	H	-	H	-	H	H	+	H	-
DA7.0bT/G	7140	H	T	T	T	T	H	G	G	T	G	H	G	T	H	H	H	H	H	T	T	H	H	H	G	G	G	G	T
DA7.4A/C	7481	H	H	A	A	C	H	H	C	A	H	C	H	H	H	A	H	A	H	A	A	A	H	H	A	C	H	C	A
DA7.5T/G	750	H	G	H	T	T	T	T	H	H	H	H	H	T	H	T	T	G	H	T	T	T	H	H	G	H	H	H	T
DA7.5aT/C	7542	C	C	H	T	H	H	T	H	H	C	H	C	T	H	H	T	C	C	C	C	C	C	H	C	H	C	H	T
DA7.5cG/T	7573	G	G	G	G	H	G	G	G	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	G	H	G	H	G
DA7.6A/C	7631	A	H	H	A	H	A	A	A	A	A	A	C	A	H	A	A	H	H	A	H	H	H	H	C	H	H	H	A
DA7.6aG/C	7661	G	G	G	G	H	G	G	G	G	H	G	H	G	G	G	G	G	G	G	G	G	G	G	G	G	H	G	G
DA8.2G/A	8229	H	G	G	H	G	H	H	H	G	H	A	G	H	G	H	A	G	G	G	H	G	G	H	G	H	G	H	G
DA8.5C/G	8590	G	G	G	G	G	G	G	G	G	G	G	G	H	H	H	G	G	G	H	G	G	G	H	G	G	G	G	G
DA8.6C/T	8659	C	C	C	H	C	H	H	H	C	H	H	C	H	C	H	T	C	C	C	C	C	C	C	C	H	C	H	C
DA9.5C/T	9580	C	C	H	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DA10.0G/A	10126	H	A	H	G	H	G	G	H	H	G	G	H	G	G	G	G	G	H	G	G	G	G	G	G	G	H	G	G
DA10.2C/T	10203	H	C	C	H	C	H	T	H	C	H	T	H	H	C	H	T	T	H	H	H	H	H	H	H	H	H	H	C
DA10.6G/A	10563	H	A	H	G	H	G	G	H	H	G	G	H	G	G	G	G	G	H	G	G	G	G	G	G	G	H	G	G
DA10.6aA/C	10604	C	C	C	C	C	C	C	C	C	C	C	C	H	H	H	C	C	C	H	C	C	C	H	C	C	C	C	C
DA10.6bC/A	10614	H	C	C	H	C	H	H	H	C	H	H	C	H	C	C	A	H	H	C	C	C	H	H	C	H	C	H	C
DA11.5C/T	11543	H	T	H	H	H	C	C	H	H	C	C	H	C	C	C	C	C	H	C	C	C	C	C	H	C	H	C	T
DA12.0C/T	12063	C	C	H	C	C	C	C	H	C	C	C	C	H	T	H	C	C	C	H	H	C	C	H	C	C	C	C	T
DA13.4A/G	13364	H	G	H	H	H	H	H	H	G	H	A	A	A	A	A	H	A	A	A	A	A	H	A	H	G	H	H	A
DA13.4a-/+	13384	-	-	H	-	-	-	-	H	-	-	-	+	H	H	-	-	-	H	-	H	H	-	-	-	-	-	H	-
DA13.5T/C	13559	H	C	H	H	H	H	H	H	C	H	T	T	H	H	H	H	T	T	H	T	T	H	H	H	C	H	H	T
DA13.8aG/A	13798	H	A	A	H	H	H	H	A	A	H	G	A	H	H	G	H	G	H	G	H	H	H	G	H	A	H	A	G
DA13.8bC/T	13858	H	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	H	C	C	C	H	H	C	C	C	C	C
DA14.2C/T	14280	C	C	C	C	C	C	H	H	H	C	C	C	C	C	C	H	C	C	C	C	C	C	C	C	T	C	H	C

Donors 240–265:

SNPs	Location	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265
DA2.6G/C	2593	G	G	H	H	G	G	H	H	G	G	H	G	H	H	G	G	H	G	G	G	G	G	G	G	G	H
DA3.5T/G	3489	T	T	T	T	T	T	T	T	T	H	T	T	H	T	T	H	T	T	T	T	T	T	H	T	T	T
DA3.7T/G	3667	T	T	H	H	T	T	H	H	T	T	G	H	H	H	T	T	H	T	T	T	T	T	T	T	T	G
DA4.0aG/A	4044	G	G	H	H	G	G	H	H	G	G	A	H	H	H	G	G	H	G	G	G	G	G	G	G	G	A
DA4.8A/C	4777	A	A	A	A	C	A	A	A	A	H	A	A	H	A	C	H	A	A	A	C	A	H	H	H	A	A
DA5.OT/C	5112	T	T	H	H	T	T	H	H	T	T	H	T	H	H	T	T	H	T	T	T	T	T	T	T	T	H
DA5.3A/T	5368	A	A	A	A	T	A	A	A	A	H	A	A	H	A	T	H	A	A	A	T	A	H	H	H	A	A
DA6.0A/G	6057	A	A	A	A	G	A	A	A	A	H	A	A	H	A	G	H	A	A	A	G	A	H	H	H	A	A
DA6.1C/T	6139	T	T	H	H	C	T	H	H	H	C	C	H	C	C	H	H	H	T	T	C	T	H	H	H	H	C
DA6.1aA/G	6171	A	A	G	A	A	A	H	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
DA6.3T/C	6335	T	T	T	T	C	T	T	T	T	H	T	T	H	T	C	H	T	T	T	C	T	H	H	H	T	T
DA6.3aA/T	6352	A	A	A	A	T	A	A	A	A	H	A	A	H	A	T	H	A	A	A	T	A	H	H	H	A	A
DA7.0A/G	7025	G	G	H	G	H	G	H	H	A	A	H	G	G	H	H	H	G	G	G	H	G	A	H	G	H	H
DA7.0a-/+	7049	H	+	H	H	-	+	H	H	-	H	-	H	+	H	H	+	H	+	H	-	+	-	H	-	-	+
DA7.0bT/G	7140	G	G	G	G	H	G	G	G	H	T	H	G	T	H	G	G	H	G	G	G	G	H	H	H	T	T
DA7.4A/C	7481	H	C	C	H	A	C	C	C	A	A	C	C	A	H	A	A	A	H	H	H	C	A	A	A	H	H
DA7.5T/G	750	T	T	G	T	T	T	H	H	T	H	T	H	T	T	T	H	T	H	T	T	T	T	H	T	H	H
DA7.5aT/C	7542	H	T	C	H	T	T	H	H	T	C	C	H	H	C	H	H	C	H	H	H	T	H	C	H	H	C
DA7.5cG/T	7573	G	G	G	G	G	G	G	H	G	G	G	H	G	H	G	G	G	G	G	G	G	G	G	G	G	G
DA7.6A/C	7631	H	A	A	A	A	A	A	A	H	A	C	H	H	H	A	A	C	A	H	H	H	A	H	A	H	H
DA7.6aG/C	7661	H	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	H	G	H	H	H	G	H	G	G	G
DA8.2G/A	8229	H	A	G	A	G	A	H	H	G	G	G	H	G	G	A	H	H	H	H	G	A	H	G	H	G	H
DA8.5C/G	8590	G	G	G	G	G	G	H	H	G	H	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G
DA8.6C/T	8659	H	T	C	T	C	T	H	H	C	C	C	H	C	C	T	C	C	H	H	H	T	H	C	H	C	H
DA9.5C/T	9580	C	C	C	C	C	C	C	C	C	C	C	C	T	H	C	C	C	C	C	C	C	C	C	C	C	C
DA10.0G/A	10126	G	G	H	G	G	G	G	G	G	H	G	G	G	H	G	G	G	H	G	G	G	G	H	G	H	H
DA10.2C/T	10203	T	T	H	T	H	T	H	H	H	C	C	H	C	C	T	H	H	H	T	H	T	H	H	T	C	T
DA10.6G/A	10563	G	G	H	G	G	G	G	G	G	H	G	G	G	H	G	G	G	H	G	G	G	G	H	G	H	H
DA10.6aA/C	10604	C	C	C	C	C	C	H	H	C	H	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C
DA10.6bC/A	10614	H	A	C	A	C	A	H	H	C	C	C	H	C	C	A	A	C	H	H	C	H	H	C	H	C	H
DA11.5C/T	11543	C	C	H	C	C	C	C	C	T	T	C	C	C	H	C	C	C	H	C	C	C	C	H	C	T	H
DA12.0C/T	12063	C	C	C	C	C	C	H	H	T	C	H	C	H	H	C	C	H	C	C	C	C	C	C	C	H	C
DA13.4A/G	13364	H	H	A	A	A	H	A	A	A	G	H	H	H	H	A	G	H	G	A	A	A	A	G	A	A	G
DA13.4a-/+	13384	H	-	H	H	H	H	H	H	H	-	H	H	H	H	H	-	H	-	+	H	+	-	-	-	H	-
DA13.5T/C	13559	H	H	T	T	T	H	H	H	T	C	H	H	H	H	T	C	H	C	T	T	T	H	C	T	T	C
DA13.8aG/A	13798	A	H	H	H	H	A	H	H	G	H	A	A	A	A	G	A	A	A	A	G	A	G	A	G	H	A
DA13.8bC/T	13858	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DA14.2C/T	14280	H	H	C	C	C	H	C	C	C	T	H	H	H	C	C	C	H	H	C	C	C	C	H	C	C	T

Donors 266–290:

SNPs	Location	266	267	268	269	270	271	272	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290
DA2.6G/C	2593	G	C	G	H	C	G	G	G	G	G	G	C	G	H	H	H	G	G	G	H	G	G	G	C
DA3.5T/G	3489	H	T	T	H	T	T	T	H	H	T	T	T	T	T	T	T	H	T	H	T	H	T	T	T
DA3.7T/G	3667	T	H	H	H	G	T	T	T	T	T	T	G	T	H	H	G	T	T	T	H	T	T	T	G
DA4.0aG/A	4044	G	H	H	H	A	G	G	G	G	G	G	A	G	H	G	A	G	G	G	H	G	G	G	A
DA4.8A/C	4777	H	H	A	H	A	A	A	H	H	A	A	A	A	A	A	A	H	A	H	A	H	C	A	A
DA5.OT/C	5112	T	T	T	H	T	T	T	T	T	T	T	T	T	H	H	H	T	T	T	H	T	T	T	C
DA5.3A/T	5368	H	H	A	H	A	A	A	H	H	A	A	A	A	A	A	A	H	A	H	A	H	T	A	A
DA6.0A/G	6057	H	H	A	H	A	A	A	H	H	A	A	A	A	A	A	A	H	A	H	A	H	G	A	A
DA6.1C/T	6139	H	C	T	C	C	T	H	H	H	H	T	C	T	C	H	C	H	T	H	C	H	C	H	C
DA6.1aA/G	6171	A	H	A	H	A	A	A	A	A	A	A	H	A	H	H	G	A	A	A	A	A	A	A	G
DA6.3T/C	6335	H	H	T	H	T	T	T	H	H	T	T	T	T	T	T	T	H	T	H	T	H	C	T	T
DA6.3aA/T	6352	H	H	A	H	A	A	A	H	H	A	A	A	A	A	A	A	H	A	H	A	H	T	A	A
DA7.0A/G	7025	H	H	G	A	A	G	H	H	H	H	G	H	G	H	G	A	H	G	H	H	H	A	H	A
DA7.0a-/+	7049	H	H	+	H	-	+	H	+	+	-	+	-	H	-	+	-	H	+	H	-	+	-	H	-
DA7.0bT/G	7140	H	H	G	G	T	G	H	G	G	H	G	T	H	G	H	H	G	G	G	H	G	T	G	G
DA7.4A/C	7481	A	C	C	H	H	H	C	H	H	C	C	A	H	H	C	H	H	H	H	C	A	A	C	H
DA7.5T/G	750	H	G	H	H	G	H	T	T	H	G	T	T	T	G	H	H	H	G	H	H	H	G	H	G
DA7.5aT/C	7542	C	T	T	H	C	H	T	T	H	C	T	H	H	C	H	H	C	C	C	C	H	C	H	H
DA7.5cG/T	7573	G	G	H	H	G	G	G	G	G	T	G	G	G	G	G	T	G	G	G	H	G	G	H	H
DA7.6A/C	7631	H	C	H	A	H	H	A	A	H	A	A	A	A	H	A	H	H	C	H	C	A	A	H	H
DA7.6aG/C	7661	H	G	G	G	G	G	G	G	H	G	G	G	G	G	G	H	H	G	H	G	G	G	G	G
DA8.2G/A	8229	G	H	A	H	G	H	H	A	A	G	A	G	A	G	H	G	G	G	G	G	H	G	H	G
DA8.5C/G	8590	G	G	G	H	G	G	G	G	G	G	G	G	G	G	G	H	G	G	G	G	G	G	G	H
DA8.6C/T	8659	C	H	T	C	C	H	H	H	T	C	T	C	T	C	H	C	C	C	C	C	C	C	H	C
DA9.5C/T	9580	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T
DA10.0G/A	10126	H	G	G	G	H	G	G	G	G	A	G	G	G	G	G	G	G	H	G	G	G	G	H	G
DA10.2C/T	10203	H	H	T	H	H	H	H	T	T	C	T	C	T	H	H	C	T	H	T	C	H	C	H	C
DA10.6G/A	10563	H	G	G	G	H	G	G	G	G	A	G	G	G	G	G	G	G	H	G	G	G	G	H	G
DA10.6aA/C	10604	C	C	C	H	C	C	C	C	C	C	C	C	C	C	C	H	C	C	C	C	C	C	C	H
DA10.6bC/A	10614	C	C	A	H	C	H	H	A	C	C	A	C	A	C	H	C	C	C	C	C	A	A	H	C
DA11.5C/T	11543	H	C	C	C	H	C	C	C	C	T	C	C	C	C	C	C	C	H	C	C	C	C	H	C
DA12.0C/T	12063	C	C	C	H	C	C	C	C	C	C	C	C	C	H	C	H	C	C	C	H	C	C	C	T
DA13.4A/G	13364	G	H	A	H	H	A	A	H	A	H	A	G	A	A	G	A	A	A	A	H	G	H	H	A
DA13.4a-/+	13384	-	-	+	-	-	H	H	H	H	-	-	-	H	H	-	H	H	H	H	H	-	H	H	H
DA13.5T/C	13559	C	H	T	C	H	T	T	H	T	H	T	C	T	T	C	H	T	T	T	H	C	C	H	H
DA13.8aG/A	13798	A	H	A	H	A	A	H	A	G	H	G	A	H	H	A	H	H	H	H	A	A	A	A	H
DA13.8bC/T	13858	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
DA14.2C/T	14280	H	C	C	C	C	C	C	C	C	C	C	H	C	C	T	C	C	C	C	H	C	T	C	C

Table 5.1 SNP genotypes from the African semen donor panel. 38 SNPs for Hotspot DA were genotyped. SNP names and their locations and donor number are shown. Each genotype is depicted as either a homozygote (A, C, G, T, +, -) or heterozygote (H). The motif disrupting SNP DA7.5T/G is higlighted in red.

5.2.2 LDU Mapping for African Panel

Using this genotype data, metric LD analysis (LDU) was performed (Figure 5.2) to see if the African population showed a historical hotspot as well as the Europeans (Chapter 3).

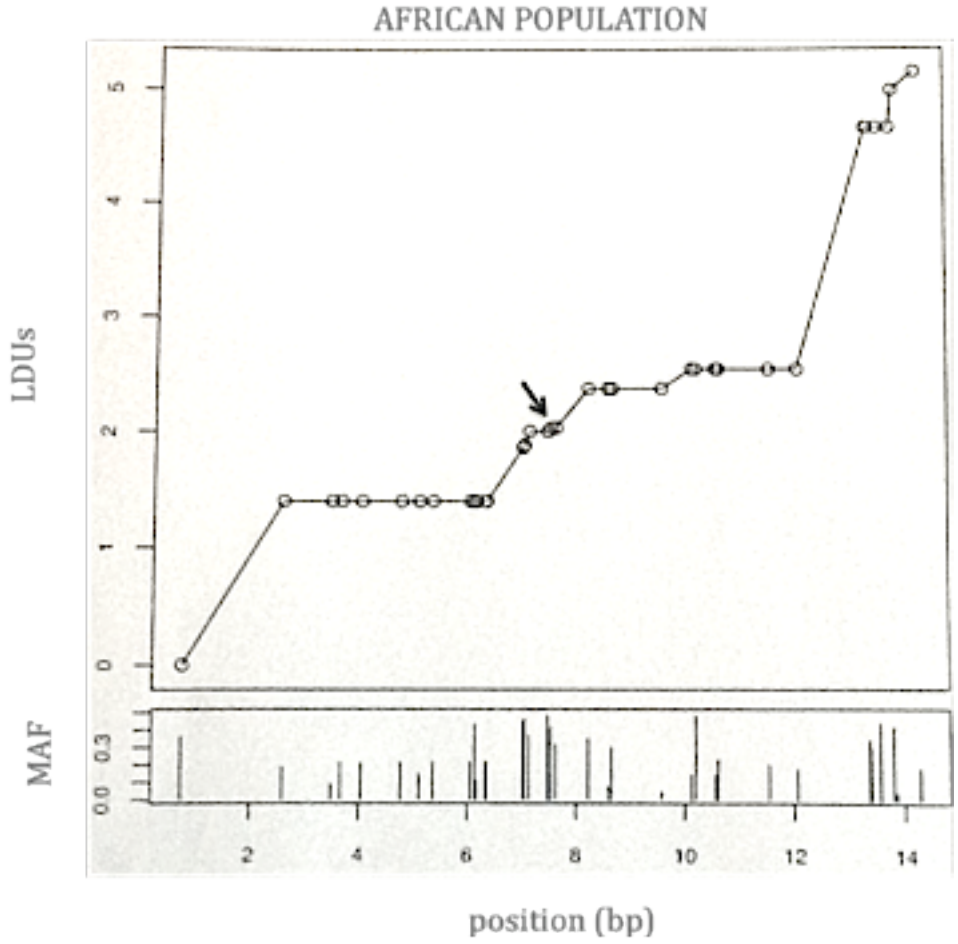


Figure 5.2 LDU map as established from the genotyped African panel. The length of ticks in the bottom panel represents the minor allele frequencies (MAFs) of the SNPs within the population tested. The location of the disrupting SNP DA7.5G/T within the 13-bp motif is shown with a black arrow.

The LDU map across Hotspot DA showed an increase of LD across the target interval in the African population, implying that historical recombination has taken place. The LDU step of the African population was reduced two-fold compared to the European population (see Figure 3.4). However, historical recombination has not necessarily influenced LD to a similar extent in both populations.

5.2.3 Allele-Specific Primers (ASPs) and Optimisation of Selector Sites

It was important to analyse as many men as possible. According to genotype LDU maps of both European and African populations, several sites became apparent as potential selector sites, located outside of Hotspot DA. Additional to previously designed and tested ASPs (see Chapter 4), 16 forward ASPs and 12 reverse ASPs were designed. Universal reverse primer DA10.3R was used for optimising forward ASPs at four different annealing temperatures respectively: 56°C, 59°C, 62°C and 65°C (Figure 5.3a). ASP DA4.8FA did not work with the chosen annealing temperatures. Additionally, ASPs DA5.3FA and DA7.0FA did not show any specificity. ASPs DA4.8FC, DA5.0RC, DA5.3FA, DA5.3FT, DA7.0bFG and DA7.4FC and DA5.0RC, DA6.1FC and DA7.0bFT showed optimal efficiency and specificity (respectively, 56-65+ °C and 59-65+ °C). Moreover, ASPs DA7.0FG and DA7.4FA showed efficiency and specificity between 56°C and 62°C, and ASPs DA6.0FG and DA6.1FT showed between 62°C and 65°C. Finally, ASP DA5.0FT showed specificity at 65°C.

DA5.2F and DA9.9aF were used as forward universal primers for the optimisation of reverse allele-specific primers. The optimised temperatures for ASPs DA10.0RG/A and DA10.2RC/T were 50°C, 53°C, 56°C and 59°C, however for ASPs DA13.4R-/+, DA13.4RA/G, DA13.5RT1/C1 and DA13.8aRA/G were 56°C, 59°C, 62°C and 65°C (Figure 5.3b). ASP DA10.0RA did not show any specificity at these temperatures. ASPs DA13.5RC1 and DA13.8aFG showed very good efficiency and specificity at the temperatures between 56-65°C. ASPs DA10.2RC, DA10.2RT and DA13.4RA showed their specificity at 59°C. ASP DA10.0RG worked efficiently between 50 °C and 59 °C. The other ASPs DA10.2RC, DA13.4R-/+, DA13.4RG, DA13.5RT1 and DA13.8aFA showed specificity at annealing temperatures of 50-59°C, 56°C, 62-65°C, 56-59°C and 56-62°C respectively.

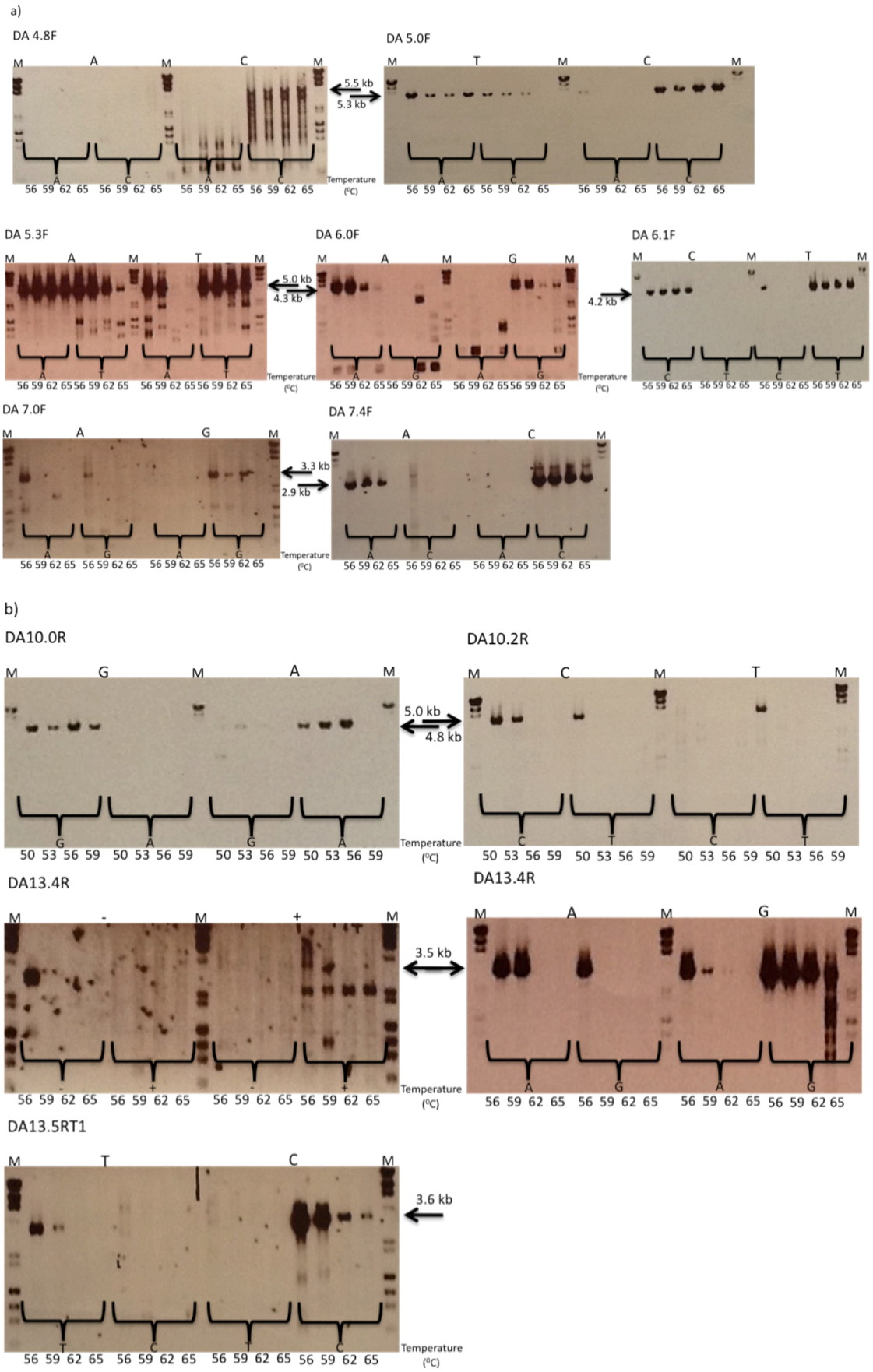

a)
DA 4.8F
DA 5.0F
5.5 kb
5.3 kb
DA 5.3F
DA 6.0F
DA 6.1F
5.0 kb
4.3 kb
4.2 kb
DA 7.0F
DA 7.4F
3.3 kb
2.9 kb
b)
DA10.0R
DA10.2R
5.0 kb
4.8 kb
DA13.4R
DA13.4R
3.5 kb
DA13.5RT1
3.6 kb
Temperature (°C)

Figure 5.3 Optimisations of forward and reverse allele-specific primers. a) Eight forward ASP pairs were optimised: DA4.8FA/C, DA5.0FT/C, DA5.3FA/T, DA6.0FA/G, DA6.1FC/T, DA7.0FA/T and DA7.4FA/C (DA7.0bFT/G is not shown in the figure) from donors homozygous for either the allele being tested or the opposite allele amplified from MDA DNA. DA10.3R was used as a reverse universal primer. PCRs using each primer were performed at four different annealing temperatures 56°C, 59 °C, 62 °C and 65 °C. b) Six reverse ASP pairs were optimised: DA10.0RG/A, DA10.2RC/T with optimisation temperatures 50 °C, 53 °C, 56 °C and 59 °C respectively. ASPs DA13.4R-/+, DA13.4RA/G, DA13.5RT1/C1 and DA13.8aRA/G, which are not shown on the figure, annealing temperatures were 56 °C, 59 °C, 62 °C and 65 °C respectively. Universal forward primer DA5.2F was used for DA10.RG/A and DA10.2RC/T, and universal primer DA9.9aF was used for the rest of the reverse ASPs. DNA inputs were 10ng MDA DNA (previously measured and stored by Rita Neumann and Alec Jeffreys, University of Leicester, UK) per 10μl PCR/ reaction. (M: marker λ*Hind*III and ϕ*XHae*III).

5.2.4 Phasing of Selector Alleles

Analysis of population LD had revealed that forward and reverse sets of alleles were in strong LD in Europeans and Africans. But haplotypes needed to be determined. To solve this problem, the designing of linkage phasing are illustrated in Figure 5.4a. The linkage phase of forward and reverse selector sites, as well as between two forward and two reverse alleles, had to be determined experimentally in each donor selected for crossover assay. In Figure 5.3b the linkage phasing results for d232 are shown with the two haplotypes of d232 as determined by linkage phasing. 28 donors were selected for crossover assay.

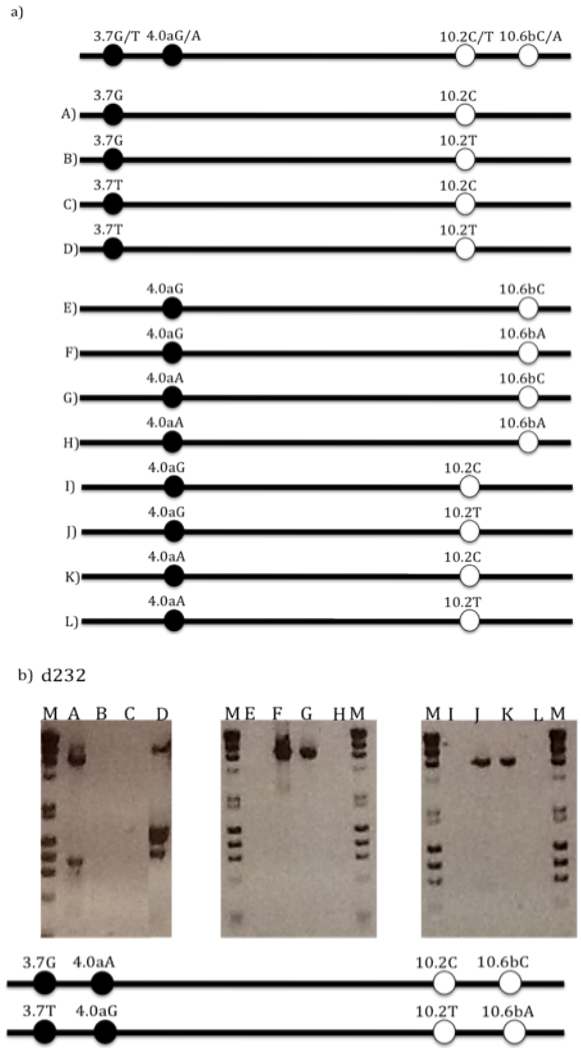


Figure 5.4 Phasing of selector alleles. Linkage is represented by the line, with SNP alleles represented by circles, and the allele indicated above each circle. a) Strategy for phasing and determining the haplotypes. The forward selector sites 3.7 and 4.0a and two reverse selector site 10.2 and 10.6b were selected for donor 232. 12 possible reactions were designed for determining the haplotypes. b) Agarose gel results for donor 232. Positive reactions determined the two haplotypes of donor 232. (M: marker λ*Hind*III and ϕ*XHae*III).

5.2.5 The Crossover Assays

Crossover assays were set up with forward and reverse primers in repulsion-phase (see Chapters 2 and 4). Primary PCR amplification of pools of sperm DNA used outside ASPs and secondary amplification of the primary PCR product was performed using inside nested ASPs to improve specificity and PCR yield. Crossover assays were carried out on 28 men, with each assay testing only one orientation of recombination. Four different pool sizes of sperm DNA molecules were set up for

each man. According to a previous estimate of crossover frequency for men homozygous for the disrupting SNP DA7.5G/T and assuming the crossover frequency is ~0.13 % for this hotspot, the pool sizes were adjusted so that they contained 0.6, 1.2, 2.4 and 4.8 crossover molecules, respectively (Table 5.2).

A universal PCR master-mix was made up for two plates to minimise variation in amplification results. Two men were analysed on the same PCR plate (in total ~1 μg sperm DNA was used for each donor), using twelve aliquots for each pool size, except for the smallest pool size. For both of the men, the first and second samples were replaced with genomic DNA from a man with selector SNPs in coupling phase in amounts of 0.1 ng and 1 ng respectively, to function as a positive control.

Donor	Population	1st pool size	2nd pool size	3rd pool size	4th pool size	Primary PCR ASPs	Secondary PCR ASPs
8	British	482	964	1929	3858	DA2.6FC + DA10.6bRC	DA3.7FG1 + DA10.6RG2
12	French	381	762	1525	3092	DA2.6FG + DA10.6bRC	DA3.7FG1 +DA10.6RA2
17	British	473	946	1892	3784	DA2.6FG + DA12.0RC	DA3.7FT1 + DA10.6bRA
26	British	309	618	1236	2472	DA2.6FC + DA10.6bRC	DA3.7FG1 + DA10.6aRA
31	British	411	849	1672	3345	DA2.6FG + DA12.0RT	DA3.7FT + DA10.6bRC
73	British	469	938	1875	3750	DA2.6FG + DA10.6bRC	DA3.7FT1 + DA10.6aRA
178	Afro-Caribbean	440	880	1760	3520	DA4.8FC + DA13.5RC1	DA5.3FT + DA13.4RaR1-
180	Afro-Caribbean	369	739	1479	2958	DA2.6FG + DA12.0RT	DA3.7FT1 + DA10.6aRA
181	Afro-Caribbean	440	880	1760	3520	DA6.1FT + DA13.8aRG	DA7.0FG + DA13.5RT1
184	Afro-Caribbean	461	922	1845	3691	DA6.1FT + DA13.8aRA	DA7.0FG + DA13.5RC1
185	Afro-Caribbean	514	1029	2047	4117	DA2.6FG + DA12.0RT	DA3.7FT1 + DA10.2C1
186	African	461	922	1845	3691	DA4.8FC + DA13.8aRG	DA5.3FT + DA13.4aR1-
211	British	464	927	1854	3708	DA2.6FG + DA10.6bRC	DA4.0aFG + DA10.2RC
232	British	477	954	1907	3820	DA2.6FC + DA12.0RC	DA3.7FG1 + DA10.6bRA
236	Zimbabwean	311	622	1244	2487	DA3.7FT +DA10.6bRC	DA4.0aFG + DA10.2RC
238	Zimbabwean	369	738	1476	2953	DA3.7FT + DA10.6bRC	DA4.0aFG + DA10.2RC
243	Zimbabwean	461	922	1845	3691	DA6.1FC + DA13.8aRA	DA7.4FA + DA13.4aR1+
244	Zimbabwean	460	920	1840	3680	DA7.0FG + DA13.4R1-	DA7.0bFT + DA10.2RC
247	Zimbabwean	413	827	1654	3300	DA2.6FG + DA10.6bRC	DA3.7FG1 + DA10.6aRA
251	Zimbabwean	312	624	1247	2495	DA3.7FT +DA10.6bRC	DA4.0aFG + DA10.2RC
253	Zimbabwean	470	941	1882	3764	DA2.6FG + DA12.0RT	DA3.7FT1 + DA10.6RG2
256	Zimbabwean	369	739	1479	2958	DA2.6FG + DA12.0RC	DA3.7FT1 + DA10.2C1
259	Zimbabwean	462	924	1849	3699	DA7.0FG + DA13.4R1-	DA7.4FC + DA10.2RC
261	Zimbabwean	440	880	1760	3520	DA4.8FC + DA13.5RC1	DA5.3FT + DA10.6bRA
269	Zimbabwean	366	732	1464	2927	DA2.6FG + DA12.0RT	DA3.7FT1 + DA10.6aRA
272	Zimbabwean	465	930	1860	3720	DA6.1FC + DA13.8aRA	DA13.4aR+ + DA7.0bFG
279	Zimbabwean	462	925	1850	3701	DA7.0bFT + DA13.8aRA	DA7.4FA + DA13.4aR+
280	Zimbabwean	369	739	1479	2958	DA2.6FG + DA12.0RT	DA3.FT1 + DA10.2T1

Table 5.2 Crossover assay designs for all possible donors. The Table illustrates respectively, donors, the population to which he belongs, the pool sizes for recombination assays, and primary and secondary PCR primers. Only one orientation was used for each donor, with orientation A coloured in blue and B marked in green. All PCR profiles and primers are shown in Appendix II.

5.2.6 Determining Crossover Frequencies

Crossover frequencies were calculated from the pooled number of positive and negative results as seen from PCR products and detected by ethidium bromide staining after gel-electrophoresis. Positive PCRs indicated the presence of at least one crossover molecule in the original pool of sperm DNA. In Figure 5.5 an example of a recombinant molecule detection is shown. Pools containing fewer molecules were less likely to be positive for crossovers, while pools containing more molecules were more likely to be positive for recombinants. Poisson-correction was used for each of the analysed men to estimate the number of crossover molecules from the proportion of PCRs that were negative. A maximum-likehood software (written by Alec Jeffreys, University of Leicester, Leicester, UK) was used to combine data across all input pool sizes.

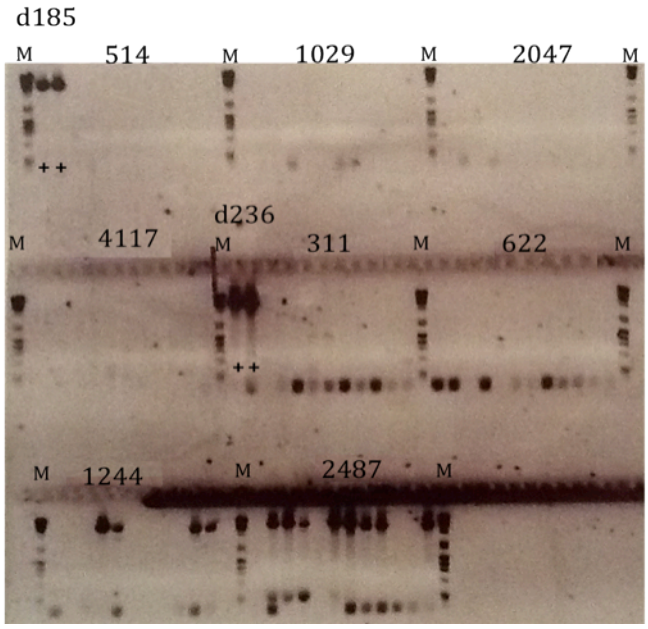


Figure 5.5 Crossover detection after repulsion-phase allele-specific PCR. Secondary PCR products of a crossover assay for d185 (homozygous (T/T) for disrupting SNP DA7.5) and d236 (heterozygous (G/T) for disrupting SNP DA7.5) men were analysed by agarose gel electrophoresis followed by ethidium bromide staining. For all four of the analysed pool sizes, the number of amplifiable molecules of each progenitor haplotype in the respective pools is shown for each man. Positive recombinants appeared as black bands. Positive controls, including sperm DNA from a man with primers in coupling phase, is indicated by (+) with respectively, 0.1 ng and 1 ng sperm DNA input. There were no recombinant molecules observed for donor 185. The reason might lay with either *cis*-regulation or *trans*-regulation or both (Berg *et al.,* 2010). (M: marker λ*Hind*III and ϕ*XHae*III).

A crossover assay was carried out for donors 185 (homozygous (T/T)) for disrupting SNP DA7.5) and 236 (heterozygous (G/T)) for disrupting SNP DA7.5) in Figure 5.5. There were no crossover molecules observed for donor 185, and moreover the recombination frequency of donor 236 (RF: ~0.016%) was lower than the previously estimated recombination rate for this hotspot. To understand the variation of the RFs and their regulatory effects on Hotspot DA, crossover assays for 28 donors were carried out as in Table 5.3.

Donor	Population	DA7.5G/T	RF %	lower 95% CI	upper 95% CI
8	British	G	0.196	0.012	0.336
12	French	G	0.203	0.011	0.346
17	British	H	0.032	0.017	0.06
26	British	G	0.361	0.204	0.625
31	British	H	0.045	0.024	0.077
73	British	H	0.02	0.001	0.039
178	Afro-Caribbean	T	0.000	0.000	0.000
180	Afro-Caribbean	T	0.008	0.002	0.019
181	Afro-Caribbean	T	0.000	0.000	0.000
184	Afro-Caribbean	T	0.000	0.000	0.000
185	Afro-Caribbean	T	0.000	0.000	0.000
186	African	T	0.000	0.000	0.000
211	British	H	0.000	0.000	0.000
232	British	H	0.026	0.013	0.046
236	Zimbabwean	H	0.016	0.008	0.043
238	Zimbabwean	H	0.09	0.053	0.146
243	Zimbabwean	T	0.000	0.000	0.000
244	Zimbabwean	T	0.000	0.000	0.000
247	Zimbabwean	H	0.063	0.035	0.104
251	Zimbabwean	H	0.036	0.017	0.07
253	Zimbabwean	T	0.000	0.000	0.000
256	Zimbabwean	T	0.006	0.001	0.017
259	Zimbabwean	T	0.000	0.000	0.000
261	Zimbabwean	T	0.000	0.000	0.000
269	Zimbabwean	H	0.014	0.008	0.022
272	Zimbabwean	T	0.007	0.000	0.010
279	Zimbabwean	T	0.000	0.000	0.000
280	Zimbabwean	G	0.042	0.023	0.071

Table 5.3 Crossover frequencies as estimated for assayed donors. 28 men were assayed for crossovers. The Table shows respectively donors, the populations of the donors to which they belong, the genotype of each man for the motif disrupting SNP7.5G/T, and recombination frequencies with lower and upper 95% confidence intervals (CI), as determined by a Poisson-correction. For example, donor 8 (G/G, homozygous for the active allele of the disrupting SNP DA7.5) showed a higher recombination rate than donor 17 (G/T (H), heterozygous for the active allele for disrupting SNP DA7.5), and moreover, donor 17 showed a higher recombination frequency than donor 178 (T/T, homozygote for the suppressed allele for disrupting SNP DA7.5)

As shown in Table 5.3, different variations of crossover frequency were seen between men. It is clear that *cis*-regulation has a major effect on Hotspot DA activity. For example, donors who are homozygous for active allele (G/G) for the disrupting SNP DA7.5 showed the highest recombination rates, and as a contrast, those homozygote for the suppressing allele (T/T) for disrupting SNP DA7.5 showed the lowest frequencies with most cases showing no crossovers. As expected, the recombination frequencies of heterozygous (G/T) donors are located in the middle. However, donors sharing the same genotype at the disrupting SNP showed substantial differences. For example, donors 26 and 280 are homozygous (G/G) for disrupting allele DA7.5 with RFs of 0.361 and 0.042 respectively, or the RFs of donors 238 and 236 heterozygous (G/T) for disrupting SNP7.5 are 0.090 and 0.016 respectively. According to recent studies, Hotspot DA has a 13-bp sequence motif that is recognised and regulated by PRDM9 (Myers *et al.,* 2010, Baudat *et al.,* 2010, Berg *et al.,* 2010).

In Figure 5.6 the association between crossover frequencies and their disrupting SNP (DA7.5) status for the analysed men is shown for Hotspot DA. The classification for their disrupting SNP (DA7.5) status was men homozygous for active allele (G/G), suppressed allele (T/T) and heterozygous (G/T). The median frequencies were also determined for each group and compared between groups. Men homozygous for the active allele (G/G) showed the highest crossover frequencies, and men homozygous for the suppressed allele (T/T) showed no crossover activity. Therefore, *cis*-regulator SNP DA7.5 has a major impact on Hotspot DA activity. Despite strong *cis* regulation on Hotspot DA, variation between crossover frequencies in the groups also indicates the effect of *trans* regulation.

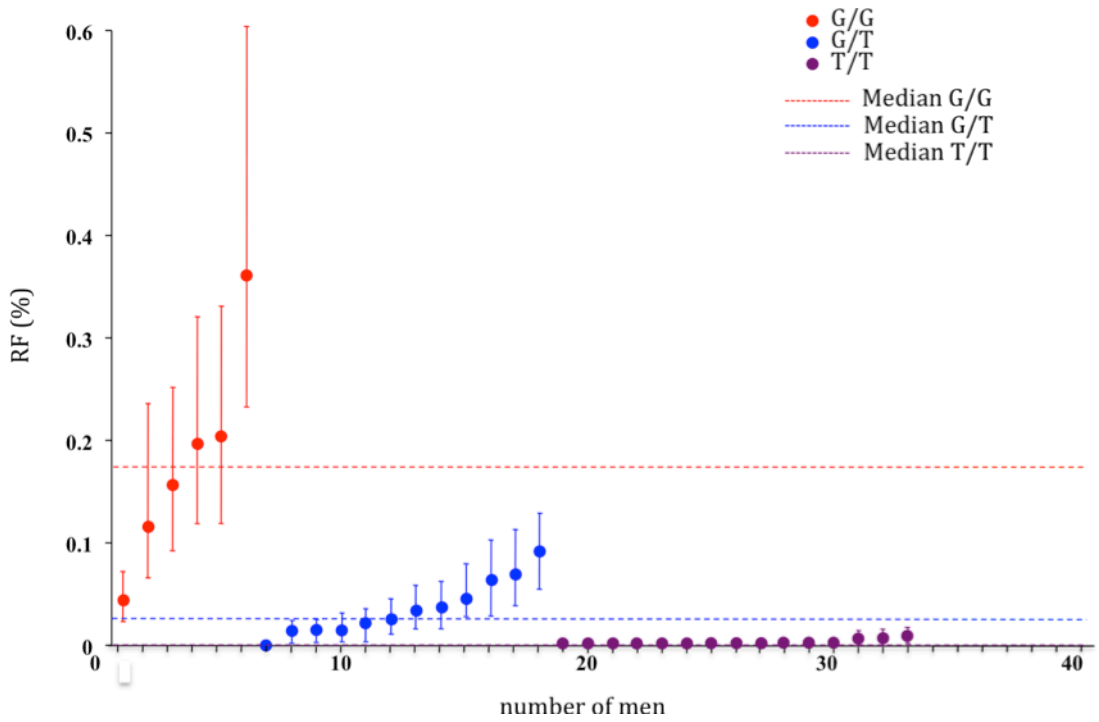


Figure 5.6 Variation in crossover activity between men at Hotspot DA, indicating their status for the disrupting SNP DA7.5. Individuals were grouped to their status as men homozygous for the active allele (G/G, shown in red) and suppressed allele (T/T, shown in purple), and heterozygous (G/T, shown in blue) for the disrupting SNP DA7.5. 95% Confidence intervals (CI) were estimated using a Poisson correction. Dashed lines indicate the median crossover frequencies observed within each group.

5.2.7 Tertiary PCR Carried Out Men That Homozygous for Suppressed Allele (T/T) for The Motif Disrupting SNP DA7.5

After repulsion-phase allele-specific PCR and according to secondary gel results, crossover frequencies were estimated. Donors that were homozygous for the suppressed allele (T/T) for the disrupting SNP and that have PRDM9 A-allele variants did not show any positive crossovers. Therefore, tertiary PCR was carried out for 15 men that were homozygous for the suppressed allele (T/T) for the 13-bp motif-disrupting SNP DA7.5, in order to determine if those negative crossovers were real. The secondary PCR agarose gel results were confirmed by the tertiary PCR agarose gel results, giving all negative crossovers for men homozygous (T/T) for SNP DA7.5. Thus, the suppressing allele T turned off crossover initiation in the studied men for Hotspot DA (Figure 5.7).

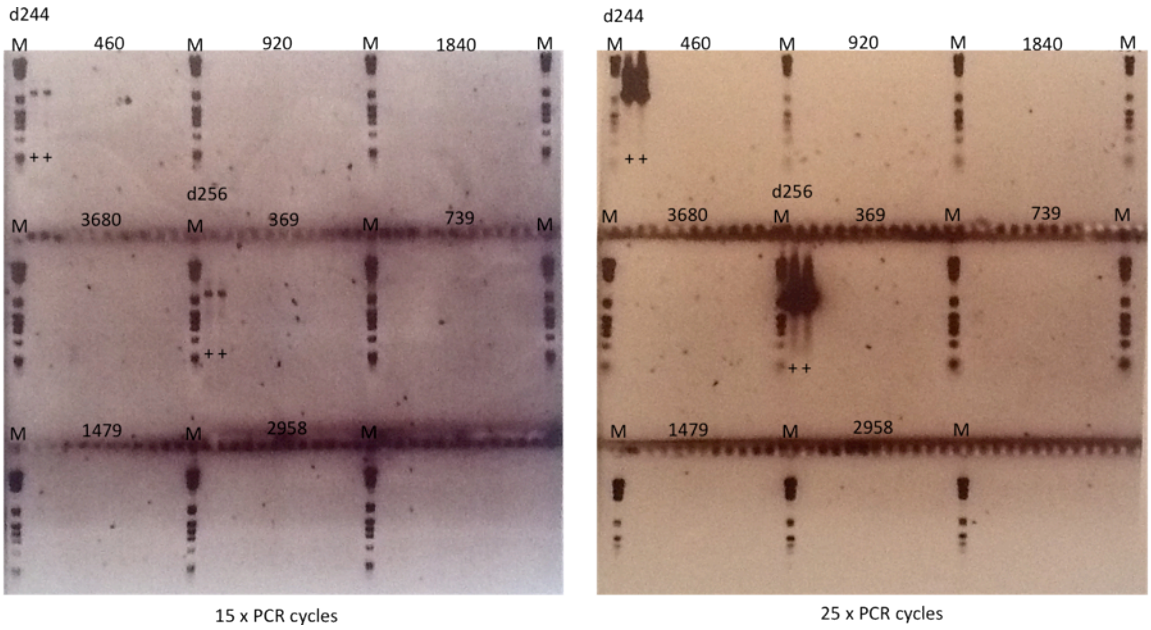


Figure 5.7 Two different PCR cycles of tertiary PCRs ethidium-bromide-stained agarose gel electrophoresis of donors 244 and 256 are shown above. All four analysed pool sizes are shown for each man. Positive controls, including sperm DNA men with primers in coupling phase, are indicated by (+) with respectively, 0.1 ng and 1 ng sperm DNA input. DA7.4F and DA10.0R were used as universal primers. There were no recombinant molecules observed in the donors. (M: marker λ*HindIII* and ϕ*XHaeIII*).

5.2.8 Effect of *Trans*-Regulator PRDM9 on Hotspot DA

Table 5.4 shows the PRDM9 status of crossovers assayed in a total of 33 men (28 new donors and 5 previously analysed men (donors 6, 7, 33, 44 and 67, as studied in Chapter 4)). The men that did not carry an A-allele of *PRDM9* displayed the lowest crossover frequencies, and of the men who did carry an A-allele of *PRDM9*, those who were also homozygous for the suppressed allele (T/T) for the disrupting SNP DA7.5 within the motif showed the lowest recombination frequencies. In contrast, donors homozygous for the active allele for the disrupting SNP DA7.5 and who carried the *PRDM9* A allele showed the highest recombination rate. Moreover, the man (donor 236) who is heterozygous for the disrupting SNP DA 7.5 and carrying a non-A allele for *PRDM9* had a recombination rate of % 0.016 , whilst the other men (four men homozygous (T/T) and one man heterozygous (G/T) for the disrupting allele SNP DA7.5) who carried a non-A allele for *PRDM9* did not show any recombination.

donor	population	PRDM9 allele 1	PRDM9 allele 2	DA7.5G/T	RF %	lower 95% CI	upper 95% CI
44	British	A	A	T	0.000	0.000	0.000
256	Zimbabwean	A	B	T	0.006	0.001	0.017
272	Zimbabwean	A	B	T	0.007	0.000	0.010
73	British	A	A	H	0.020	0.001	0.039
31	British	A	A	H	0.045	0.024	0.077
247	Zimbabwean	A	A	H	0.063	0.035	0.104
6	British/Indian	A	A	H	0.069	0.004	0.135
238	Zimbabwean	A	A	H	0.090	0.053	0.146
67	British	A	A	G	0.155	0.007	0.302
8	British	A	A	G	0.196	0.012	0.336
12	British	A	B	G	0.203	0.011	0.346
26	British	A	A	G	0.361	0.204	0.625
261	Zimbabwean	L21	A	T	0.000	0.000	0.000
181	Afro-Caribbean	A	L11	T	0.000	0.000	0.000
243	Zimbabwean	B	L15	T	0.000	0.000	0.000
256	Zimbabwean	A	L12	T	0.000	0.000	0.000
244	Zimbabwean	L22	A	T	0.000	0.000	0.000
178	Afro-Caribbean	C	A	T	0.000	0.000	0.000
184	Afro-Caribbean	A	L7	T	0.000	0.000	0.000
180	Afro-Caribbean	L4	A	T	0.008	0.002	0.019
7	British	A	E	H	0.013	0.007	0.025
269	Zimbabwean	A	L11	H	0.014	0.008	0.022
232	British	A	E	H	0.026	0.013	0.046
17	British	A	L20	H	0.032	0.017	0.060
251	Zimbabwean	A	L16	H	0.036	0.017	0.070
280	Zimbabwean	L4	A	G	0.042	0.023	0.071
33	British	A	L2	G	0.115	0.060	0.224
186	African	L6	L13	T	0.000	0.000	0.000
253	Zimbabwean	C	C	T	0.000	0.000	0.000
185	Afro-Caribbean	C	L12	T	0.000	0.000	0.000
279	Zimbabwean	L6	C	T	0.000	0.000	0.000
211	British	L20	E	H	0.000	0.000	0.000
236	Zimbabwean	L4	L22	H	0.016	0.008	0.043

Table 5.4 Sperm crossover frequencies and their relationship with *PRDM9* status. Men were ranked with recombination frequencies (RFs), and 95% confidence intervals were determined by a Poisson-correction. *PRDM9* variants and SNP information for each man for the motif disrupting SNP7.5G/T are also shown above. The *PRDM9* variant allele B acts like allele A (Berg *et al.*, 2010). The other variant alleles of *PRDM9* are indicated in blue.

To test the association status with crossover frequencies further, *PRDM9* 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).

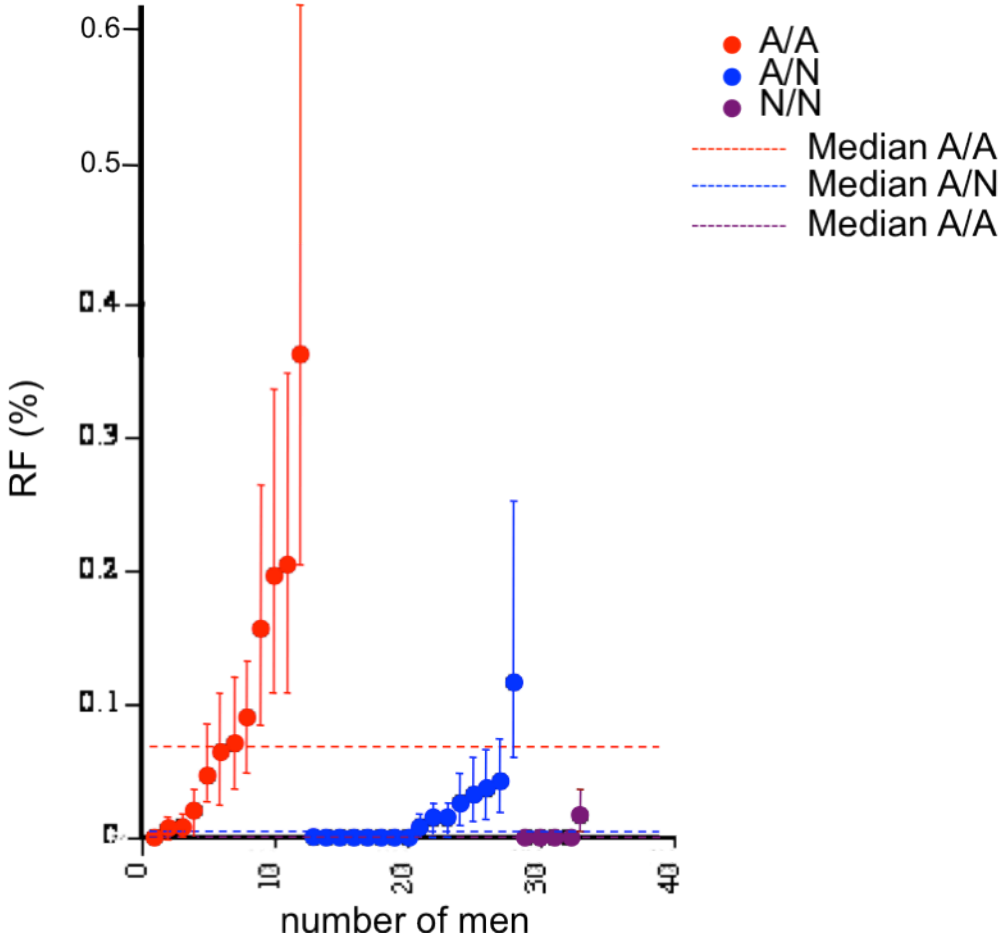


Figure 5.8 Variation in crossover activity between men at Hotspot DA. Individuals were grouped by their status into men carrying two A-alleles (A/A shown in red), one A-allele (A/N shown in blue) or non-A alleles (N/N shown in purple). 95% Confidence intervals (CI) were estimated using a Poisson-correction. Dashed lines indicate the median crossover frequencies observed within each group.

The difference between crossover frequencies of men homozygous A/A compared to the frequency of men without A alleles is significant (one-tailed t-test, *P*= 0.035; one-tailed Mann Whitney test *P*=0.0045), however when men who carry the suppressed allele T for the disrupting motif were excluded from the group, *PRDM9* status became insignificant (*P*= 0.0741). This was probably due to the small number of men with N/N genotypes for *PRDM9*. The higher crossover frequencies of men with an A/A *PRDM9* status compared to men with an A/N status is of marginal significance (one-tailed Mann Whitney Test, *P*=0.0029; one-tailed t-test, *P*=0.0034). However, when the men homozygous for the suppressed allele (T/T) for the disrupting SNP DA7.5 were excluded from the group, the frequencies observed in the A/A men were significantly higher than those in the A/N men (one-tailed t-test, *P*=0.0224). Additionally, the higher crossover frequencies of men with A/A *PRDM9* status compared to men with A/N and N/N statuses are extremely significant (one-tailed Mann Whitney test, *P*=0.0006; one-tailed t-test *P*=0.0007). Moreover, when the men carrying the suppressed allele (T/T) were excluded, the compared frequencies between A/A versus A/N and N/N stayed significant (one-tailed t-test, *P*= 0.0081). These findings established the *PRDM9* variant allele A as an activating variant, while *PRDM9* variants L2 and L4 were found to be non-activating in men carrying the active G allele at the disrupting SNP DA7.5.

5.2.9 Re-Sequencing of Donors

14 crossover-assayed donors were re-sequenced for Hotspot DA to search for any other potential *cis*-regulators of hotspot activity. Separate haplotypes were generated using allele-specific primers and universal primers (DA7.3F and DA7.6R) to amplify the approximate 1.5-kb interval centred around the motif disrupting SNP (PCR conditions and primers are shown in Appendix II). In Table 5.5 the re-sequencing results are summarised (for full re-sequencing data, see Appendix IV). Two haplotypes of 14 men were re-sequenced and four new SNPs (7.5b, 7.5c, 7.6a and 7.7) close to the hotspot centre were identified. These SNPs were genotyped in both our European and African panels. SNPs 7.5b and 7.7 were donor-specific SNPs. SNP 7.5c was observed in only four donors in the European panel, but in the African panel this SNP is normally distributed. SNP 7.6a was also picked up in both panels. Table

5.5 shows a list of the active and suppressed haplotypes for Hotspot DA. Active haplotypes were identified as both the haplotypes found in each active man where crossovers have been mapped and there was no asymmetry, and also as the haplotypes in each man showing strong asymmetry that corresponds to the inactive haplotypes. Inactive haplotypes were identified in men carrying a *PRDM9* A variant (in practice, these are all T/T men), and in active men where mapping showed strong asymmetry. Comparing the two sets of haplotypes showed that only the G allele for SNP DA7.5, and no other variant, was found within the active haplotype group. To sum up, the re-sequencing results demonstrated that the *cis* effect is mediated by the motif disrupting SNP DA7.5 and no other sequences had an effect.

	SNP	DA7.0	DA7.0a	DA7.0b	DA7.4	DA7.5	DA7.5a	DA7.5b	DA7.5c	DA7.6	DA7.6a	DA7.7	DA8.2
	locations	7025	7049	7140	7481	7501	7542	7551	7573	7631	7661	7768	8229
Donors													
Suppressed haplotypes													
6		G	+	T	A	T	C	C	G	C	G	T	G
7		A	+	G	C	T	T	C	G	A	G	G	G
44		A	+	T	A	T	C	C	G	C	G	G	G
44		A	+	G	A	T	C	C	G	A	G	G	G
73		A	-	T	A	T	C	C	G	A	G	G	G
180		G	-	T	A	T	C	C	G	A	G	G	A
180		A	-	G	A	T	T	C	G	A	G	G	G
185		G	+	T	A	T	C	C	G	C	G	G	?
185		G	-	T	A	T	C	C	G	A	G	G	?
211		G	+	T	A	T	C	C	G	C	G	G	G
232		A	-	T	A	T	T	C	G	A	G	G	G
238		G	+	G	C	T	T	C	G	A	G	G	A
251		G	+	G	C	T	T	C	G	A	G	G	A
253		G	+	G	C	T	C	C	T	A	G	G	G
253		A	-	T	A	T	C	C	G	C	G	G	G
Active Haplotypes													
6		A	-	T	C	G	T	C	G	A	G	G	A
7		A	-	G	A	G	C	C	G	A	G	G	A
12		A	+	G	A	G	C	C	G	A	C	G	G
12		A	-	G	A	G	C	C	G	A	G	G	A
67		A	-	T	A	G	C	C	G	C	C	G	G
67		A	-	G	C	G	C	T	G	A	G	G	G
73		G	+	G	C	G	C	C	G	C	C	G	G
211		A	-	G	C	G	C	C	G	A	G	G	G
232		A	-	G	C	G	C	C	G	C	G	G	A
238		G	-	G	C	G	C	C	T	C	G	G	G
251		G	-	G	C	G	C	C	T	C	G	G	G
280		G	-	G	A	G	C	C	G	C	G	G	G
280		A	-	G	C	G	C	C	G	A	G	G	G

Table 5.5 Re-sequencing results. The active and suppressed haplotypes of 14 men are shown above. The locations and names of the SNPs are shown in the table. Newly identified SNPs 7.5b, 7.5c, 7.6a and 7.7 were typed and are coloured in pink. Donors that were heterozygous for each SNP are shown in red and blue letters. The motif disrupting SNP 7.5 is coloured in yellow. According to the active and suppressed haplotypes, the motif disrupting SNP DA7.5 is the only *cis* regulator that can effect hotspot activity for Hotspot DA.

5.3 DISCUSSION

Cis-regulator SNP DA7.5 is the main factor that influences hotspot activation on Hotspot DA, however the *trans*-regulator factor *PRDM9* also has an important impact. Both regulators showed a good correlation with each other. Hotspot DA was inactive in men that had the A-allele of *PRDM9* and were also homozygous for the suppressed allele (T/T) for the motif disrupting SNP. The crossover frequencies were higher in men that had the A-allele of *PRDM9* and who were also homozygous for the active allele (G/G) for the disrupting motif SNP. Moreover, men that had one A-allele for *PRDM9* showed variation in their crossover frequencies according to both their carried *PRDM9* alleles, and their disrupting SNP genotypes. Berg *et al.,* (2010) studied 5 hotspots that lack the 13-bp hotspot motif proposed to be the *in vitro* binding site for the protein encoded by the *PRDM9* A-allele. Additionally, *PRDM9* regulates hotspots with or without the hotspot motif to the same extent (Berg *et al.,* 2010), but in Hotspot DA the *cis*-regulatory effect is very significant.

Twelve men who were homozygous for the suppressed allele (T/T) of the disrupting SNP DA7.5 and who had different *PRDM9* non-activating variants (C, E, L4, L6, L7, L11, L12, L13, L15, L21, L22) were studied to see whether any of the *PRDM9* variants can rescue the suppression caused by the suppressed T SNP allele acting in *cis.* The *PRDM9* variant L4 generally binds five of the exact 8 bases (Berg *et al.,* 2010) of the 13-bp motif, but one of its exact bases is the disrupting SNP, so L4 (donor 236) variants only bind to four exact bases of the motif and so rescued only a few crossovers (~1-10 possitive crossover reactions in 100,000 molecules). Interesting results come with the men carrying A and L4 variants (donors 180 and 280). These men had different disrupting SNP genotypes (donor 180 (T/T) and donor 280 (G/G)), and it was the men homozygous for the active allele (G/G) of the disrupting SNP who showed higher recombination frequencies. However, d280 and d236 were both homozygous for the suppressed allele (T/T) of the disrupting SNP, and showed very low hotspot activity when compared to the other men who were homozygous for the suppressed allele. Previously, the L20 variant had been observed to be non-activating at Super-hotspots F and U (Berg *et al.,* 2010). Conversely, L20

has been shown to activate the hotspot MSTM1b. The PRDM9 variant L20 therefore has entirely different activation profiles between different super-hotspots (Berg *et al.,* 2010). Even though it is not that strong, perhaps the *PRDM9* L4 variant has an activation effect on Hotspot DA. Recently, a hotspot activated by C variants of *PRDM9* was studied by Berg *et al.,* 2011. The C variant of *PRDM9* totally inactivates Hotspot DA.

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/A alleles show variation according to their disrupting SNP genotype, and when considering that the *cis*-effect of the variation was as large as 60-fold between donors 26 (G/G) and d256 (T/T). However, the same disrupting SNP genotypes showed a very small, ~3-fold difference between donors d26 (G/G) and donor 67 (G/G), and 4.5-fold between donors 238 (G/T) and donor 73 (G/T).

However, a comparison of recombination rates between A/A men and A/N men that were heterozygous for the motif-disrupting SNP is significant (one-tailed t-test, P= 0.0153). Thus, the *trans*-regulator *PRDM9* has a significant effect on hotspot initiation in Hotspot DA.

To conclude, both the *trans*-regulator PRDM9 and the *cis*-regulatory disrupting SNP DA7.5 within the 13-bp motif close to the centre of the hotspot, have major influences on hotspot initiation in Hotspot DA. Hotspot DA is the only human crossover hotspot in which the disrupting SNP within the motif has a very strong *cis* effect for the hotspot turn on/off polymorphism, and this *cis*-regulation of the disrupting SNP was confirmed by re-sequencing.

CHAPTER 6: DISCUSSION

There are three main techniques used to investigate recombination frequency in humans: cytological analysis, pedigree or sperm analysis to observe the segregation of markers, and population genetics approaches. The simplest method for measuring recombination is pedigree analysis, but this can only give a low resolution due to the combination of the low numbers of informative meioses observed and the low average recombination rate per unit of physical distance along DNA in humans. An alternative method, sperm analysis, allows male recombination rates to be estimated directly, but assays are restricted to short intervals (typically up to 10 kb) and are practically very challenging to design and optimise. In contrast, population LD analyses allow recombination to be estimated indirectly by inferring historical recombination events to explain haplotype structure. About 33,000 hotspots were estimated to be present in the human genome from historical LD block boundaries (Myers *et al.,* 2005), corresponding to about one in every 50 to 100 kb of sequence. To date, only ~40 hotspots have been directly characterised in individuals by sperm typing (Jeffreys *et al.,* 1998a; Jeffreys *et al.,* 2001; Jeffreys and Neumann, 2002; May *et al.,* 2002; Kauppi *et al.,* 2004; Jeffreys *et al.,* 2005; Holloway *et al.,* 2006; Webb *et al.,* 2008).

In addition to this ~40% of human recombination hotspots have a 13-bp CCNCCNTNNCCNC sequence motif that determines hotspot activity (Myers *et al*. 2008). Previous studies have reported the role of these motifs in hotspot activity where single nucleotide variants disrupting the centrally located motifs CCTCCCT and CCCCACCCC reduce crossover initiation activity in *cis* at hotspots DNA2 (Jeffreys and Neumann, 2002) and NID1 (Jeffreys and Neumann, 2005) respectively.

6.1 HIGH-RESOLUTION LINKAGE DISEQUILIBRIUM ANALYSIS OF PUTATIVE HOTSPOTS CONTAINING A MITF-DISTRUPTING SNP

Using low-resolution HapMap Phase II data, Donnelly and colleagues drew up a shortlist of the four best candidate LD hotspots (DA, DB, DC and DD) based on

having the 13-bp CCNCCNTNNCCNC motif with a disrupting SNP and good historical activity (Donnelly, personal communication). This raised the question of whether the nature of the motif offers any clues for understanding the molecular basis of recombination hotspots. In Chapter 3 these four LD hotspots were genotyped using high-resolution genotyping techniques in our European semen donor panel and tested to see if the motif-disrupting SNP lay at the centre of the hotspot as defined by LDU mapping. Confirming these hotspots with LDU mapping is crucial because it confirms the existence of the LD hotspot in our donor panel, the location of the motif with the disrupting SNPs, and it also allows the width of the hotspot to be estimated. LDU mapping confirmed all four putative LD hotspots (DA, DB, DC and DD) to have a width of 1-2 kb which is characteristic of hotspots (Jeffreys *et al.* 2000). According to the LDU map, the motif disrupting SNP was at the centre of Hotspot DA, and this hotspot had an estimated historical recombination frequency of $\sim 5 \times 10^{-4}$. Hotspot DA was thus a very promising candidate for further analysis. The second hotspot, Hotspot DB, had the disrupting SNP close to the centre of the hotspot, but crossover assays designed for Hotspot DB failed to detect any crossovers. This could be explained by various reasons. Firstly, the recombination frequency of Hotspot DB is very low, and therefore one explanation is that the assay could not pick up any positive crossovers in our European panel. Here, a higher DNA input would help give the answer, but unfortunately there was only a limited DNA stock for these selected men. The second explanation is that Hotspot DB does not exist anymore in the population, but we did not have enough time to explore this further.

For Hotspot DC the LDU map showed the disrupting SNP to be outside of the centre and so this hotspot did not fit with our hypothesis and was eliminated. With Hotspot DD the disrupting SNP was localised close to its centre, but the heterozygosity of the donors did not allow us to conduct further analyses. For these reasons Hotspots DB, DC and DD were not studied for this thesis, and instead further analysis was carried out on only Hotspot DA.

6.2 DETERMINING THE RECOMBINATION RATE AND HOTSPOT POLYMORPHISM IN HOTSPOT DA

Previously, major variation in crossover frequencies had been observed (Neumann and Jeffreys, 2006; Jeffreys and Neumann, 2005), and some of these variations were associated with specific hotspot sequence variants that influence the efficiency of crossover initiation between chromosomes (Jeffreys and Neumann, 2002; Jeffreys and Neumann, 2005; Jeffreys and Neumann, 2009). This manifests as reciprocal crossover asymmetry in heterozygotes, as higher initiation on one haplotype can lead to biased gene conversion tracts accompanying the crossover, with markers acquired from the less active homologue (Jeffreys and Neumann, 2002). In a previously characterised hotspot, both crossovers and non-crossovers indicated comparable biased transmission in both crossover and non-crossover exchange conversion products (Jeffreys and Neumann, 2005). Therefore, crossovers and non-crossovers were in general influenced by the same factors.

A 15-kb target interval around Hotspot DA that was centered on the motif-disrupting SNP DA7.5 within the 13-bp CCNCCNTNNCCNC hotspot sequence motif was assayed for recombination activity in six men. The crossover frequency showed Hotspot DA to be a regular hotspot with an average crossover rate (~8×10^{-4}) among hotspots assayed on autosomes.

Another factor potentially at play is meiotic drive. This is observed in sperm when a recombination suppressing haplotype is over-transmitted to progeny, both at crossovers and at gene conversions, without exchange of flanking markers. It suggests that changes within the hotspot could become attenuated in activity during evolution (Jeffreys and Neumann, 2002; Jeffreys and Neumann, 2005). Other factors besides the sequence within the hotspot of identical haplotypes could regulate the activity of a presence/absence hotspot polymorphism (Neumann and Jeffreys, 2006). Therefore, Hotspot DA provides evidence that the motif is likely to be involved in promoting the initiation of recombination. Five assayed donors showed very similar hotspot locations for orientation A and B crossovers. The centre of the hotspot is displaced 3′

to DA7.5 by ~80 bp. The numbers of crossovers mapping to the 5′ and 3′ ends of DA7.5 were normalised to equal numbers of A and B crossovers, and a Fisher test was performed to check if the numbers were significantly different from the predicted 50:50 Mendelian transmission ratio if DA.7.5 is located exactly at the centre. Thus it was concluded that DA7.5 and the motif were not located at the centre of Hotspot DA. Sperm crossovers in men heterozygous for a motif-disrupting variant show the greatest transmission distortion ratio (~90-10, T: G) ever seen in human crossover progeny (Jeffreys and Neumann, 2002; Neumann and Jeffreys, 2006).

In humans (Jeffreys and Neumann, 2002; Neumann and Jeffreys, 2005), mice (Boulton *et al.,* 1997) and fungi (Nicolas *et al*., 1989; Nicolas and Petes, 1994), different recombination initiation rates result in over-transmission of alleles from the suppressed haplotype into recombinant progeny (Neumann and Jeffreys 2006). The highest transmission distortion had been seen in hotspot MSTM1a (on average 72:28 versus 50:50 for Mendelian transmission) for markers nearest the centre of the hotspot, which is similar in intensity to that seen at the other mammalian hotspots showing distortion (Jeffreys and Neumann, 2002; Neumann and Jeffreys, 2005; Yauk *et al.,* 2003). These meiotic drives in favour of alleles from the suppressed haplotype allow the identifcation of the putative active haplotype in each of the analysed men (Neumann and Jeffreys, 2006). The transmission frequency results showed that the asymmetries in the three donors heterozygous for the disrupting SNP DA7.5G/T were very similar in terms of the transmission frequency of DA7.5G to crossover progeny (10:90, G:T transmission ratio) and in terms of the displacement of orientation A versus orientation B crossover distributions. Donors that are homozygous for the active G-allele showed the highest recombination rate among the other donors. The direction of biased gene conversion indicates that the chromosome carrying the disrupting allele was suppressed for crossover initiation. This is supported by donor 44, who is homozygous for the disrupting allele (T/T) and shows no recombination (RF = 0%). *Cis*-regulation has previously been seen and has reduced the hotspot activation at studied hotspots (Jeffreys and Neumann, 2002; Neumann and Jeffreys, 2005; Neumann and Jeffreys, 2006; Jeffreys and Neumann, 2009; Berg *et al*., 2010, Sarbajna *et al*., 2012), but the *cis*-regulation of Hotspot DA has provided the first strong evidence that *cis* factors have a direct influence on the turn on/off activity of

hotspots.

6.3 THE RELATIONSHIP BETWEEN CROSSOVERS AND NON-CROSSOVERS

Understanding the relationship between crossovers (COs) and non-crossovers (non-exchange gene conversions, NCOs) has been facilitated by examination of the large number of non-exchange events at several hotspots. This revealed that crossover and non-crossover frequencies were positively correlated (Berg *et al.*, 2011; Sarbajna *et al.,* 2012).

Hotspots with high recombination rates also tend to show high non-crossover frequencies. This correlation was seen either between men at a given hotspot or when compared between hotspots (Jeffreys and May, 2004, Jeffreys and Neumann, 2005) at a fairly stable CO:NCO ratio of 2:1. Previously, the CO:NCO ratio was estimated at 4:1 in Europeans and 1:1 in Africans by population genetics approaches (Ptak *et al.,* 2004). In these estimates the influence of gene conversion on LD, which can have detrimental effects in association studies, was ignored. The inflation of historical recombination on LD cannot be explained by crossover-based recombination rates alone (Ptak *et al*., 2004). Another study by Ardlie *et al*. (2002) reported non-exchange conversion at a ratio of 3:1-10:1 in favour of non-exchange conversions, with 6:1 being the best estimate (Ardlie *et al.,* 2002). This showed incomplete LD that was not readily explained by the expected historical recombination rate based on crossovers alone. It is important to remember that COs are always detected by repulsion-phase PCR, but NCOs can only be detected when they occur in a region including an informative SNP. Therefore, many NCOs are missed, so the quoted ratio may be misleading.

A wide range of CO: NCO ratios have typical values of ~3:1 between men. In this work, the directly observed CO: NCO ratios varied between two analysed donors for Hotspot DA, but at the lower end of what would be expected (10:1). The non-crossover frequency at Hotspot DA seem unusually low, although this might well reflect missing non-crossovers since only those that had co-converted at SNPs 7.4 and 7.5 were scored. There was no significant difference between the two donors in the

ratio of non-crossovers to crossovers (Fisher exact test, $P = 0.14$). Nevertheless, the data can be used to determine the likelihood that a flanking SNP is co-converted in those non-crossover molecules known to have converted at both SNP7.4 and SNP7.5. While the sample sizes are small, there was clear evidence for a very steep gradient, with the likelihood of co-conversion declining rapidly as one moves away from the selected sites SNPs 7.4 and 7.5. This fits with previous studies of non-crossover distributions in human hotspots. In contrast, CO: NCO ratios showed a significant and very strong variation between 14 men at hotspot *SPRY3* (Sarbajna *et al.,* 2012). Hotspot *SPRY3* is located in the minor human pseudoautosomal region (PAR2) and may behave differently to Hotspot DA because either PAR1 or PAR2 must engage in a crossover event at any given meiosis to prevent non-disjunction. Therefore, factors influencing the CO/NCO 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 not yet been detected at autosomal hotspots.

When reciprocal transmission rates of markers in crossovers and non-crossovers were compared for Hotspot DA, the non-crossovers showed complete bias resulting in the transfer of DA7.5T to the DA7.5G haplotype, and with no instances of G>T transfers. This extreme bias is highly significant ($P = 1.2\times10^{-10}$) and is in the same direction as the bias seen in crossovers. For both donors 6 and 55 the strength of the distortion is greater in NCOs than in COs. For each donor, this disparity is not significant (Fisher exact tests on numbers of 7.5G- and T-containing COs versus numbers of G- and T-containing NCOs). However, it is significant for data from both donors combined, even on raw counts of numbers of NCOs before Poisson correction. It therefore appears that NCOs show a significantly stronger bias towards acquiring 7.5T compared with COs.

6.4 THE PATHWAYS FOR GENERATING CROSSOVERS AND NON-CROSSOVERS

Hotspot DA is an example of a hotspot with both types of biased gene conversion operating within the same hotspot. Biased gene conversion into crossovers and also non-crossovers was observed at SNP DA7.5, located ~80 bp away from the centre of

Hotspot DA. Significantly distorted transmission ratios were observed for all men that were heterozygous at this marker.

Mechanistically, human recombination hotspots can be explained for all observations of recombination events by the single canonical double-strand repair model of recombination (DSBR).

Biased gene conversions in COs and NCOs are influenced to the same degree, if COs and NCOs are generated as alternative products of the same double Holliday Junction (dHJ) molecule. Heteroduplex DNA within a dHJ intermediate is subjected to mismatch repair (MMR), with repair being directional from the invaded to the invading CO. This was seen as transmission distortion in favour of CG at Hotspot *DNA2* (Jeffreys and Neumann, 2002) and *SPRY3* (Sarbajna *et al.,* 2012), as well as AT alleles at Super-hotspot S2 (Jeffreys and Neumann, 2009) and *NID1* (Jeffreys and Neumann, 2005). This observation is consistent with early mismatch repair being biased towards repairing from the invaded duplex to the invading strand.

The canonical DSBR model has the potential to explain both biases within the same intermediates. In the DSBR model of recombination, both COs and NCOs are produced from the same recombination intermediate and a dHJ by resolution in different planes. Therefore, the intermediates destined to become crossover and non-crossover molecules are perhaps sensed and processed differently by the MMR process, resulting in gene conversion with different degrees of bias. This may be manifested by differences in MMR recognising the heteroduplex and/ or differences in the strength of the bias in the lower of T during repair. Alternatively, the extent of resection of the Spo11-induced DSB may differ between molecules destined to become COs and NCOs such that there may be a difference in the likelihood that a G/T heteroduplex is generated at SNP DA7.5.

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, 2001). Evidence for two pathways of non-crossover generation in the human genome comes from the observations of Sarbajna *et al.* (2012) at the minor

pseudoautosomal hotspot SPRY3. They observed at least a proportion of non-exchange conversions being generated via a second, perhaps SDSA pathway (Sarbajna *et al.,* 2012). This SDSA pathway could be applicable to Hotspot DA. This pathway would require that SNP 7.5 is always incorporated into a G/T heteroduplex in G/T heterozygotes, and is always repaired in favour of the T-allele during, for example, early mismatch repair. Also, meiotic drive in favour of T will lead to hotspot extinction, and not just modest down-regulation as seen at most hotspots.

6.5 PRDM9 REGULATION ON HOTSPOT DA

At the beginning of this work, knowledge about the factors regulating the human recombination machinery was limited. Early findings of initiation biases that resulted in reciprocal crossover asymmetry pointed to *cis*-acting factors regulating hotspot activity. High variation in crossover frequencies between men had also been observed independent of specific sequence variation at two closely linked hotspots (Neumann and Jeffreys, 2006). These findings prompted the investigation of a *trans*-regulatory factor, which has recently been identified as PRDM9 (Baudat *et al.,* 2010, Myers *et al.,* 2010, Berg *et al.,* 2010).

The identification of PRDM9 as the major *trans*-regulatory factor for specifying and regulating hotspot activity (Berg *et al.,* 2010, Berg *et al.,* 2011), has dramatically increased our understanding as to what controls human crossover hotspots.

Variant *PRDM9* allele A is the most common allele in the European population, and it binds to a 13-bp motif in human LD hotspots identified using European HapMap data (Myers *et al.* 2010). Berg et al. (2010) provided 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). Moreover, additional data addressing PRDM9 regulation and the relationship with the hotspot motif came from the investigation of recombination hotspots tuned to *PRDM9* Ct variants, which are more common in Africans (Berg *et al.* 2011). However, the LD Hotspot DA was shown to be activated by the *PRDM9* A-variant, and only in men that are homozygous for the active allele (G/G) and heterozygous (G/T) for the disrupting SNP DA7.5 within the motif. Additionally, the men homozygous for the suppressed allele (T/T) of

the disrupting SNP DA7.5 showed either no or extremely low recombination rates in the presence of the *PRDM9* allele A. Men homozygous for the active allele (G/G) of the disrupting SNP and who carried a non-A *PRDM9* variant need to be studied to prove the direct influence of the T allele on the hotspot turn on/off status, but unfortunately, there were no suitable donors in our panel.

6.5.1 The Protein Encoded by *PRDM9* Binds to a 13-bp Motif

The data presented in this thesis is the only evidence for human crossover hotspot regulation by a very strong *cis*-regulatory disrupting SNP. However, the *trans*-regulator *PRDM9* has a secondary effect. Sperm crossover activity was highly dependent up on *cis*-regulation in Hotspot DA. However, men who carried different *PRDM9* variants showed variable crossover frequencies depending on their disrupting SNP genotypes and *PRDM9* status. In contrast, at some Super-hotspots independent of the presence of the hotspot motif (Berg *et al*., 2010), sperm crossover activity was highly dependent on specific *PRDM9* variants. Moreover, a motif-strengthening SNP at Hotspot 5A was associated with a suppressed haplotype (Berg *et al.,* 2011), and it appears that the presence of a motif is not sufficient for recombination either, as the motif is common and can be seen in recombination cold sequence outside of a hotspot (Berg *et al*., 2010). However, a motif-disrupting SNP was found to be associated with the suppressed haplotype at hotspot NID1 (Jeffreys and Neumann, 2005; Myers *et al.,* 2005; Myers *et al.,* 2008), which is opposite to Hotspot 5A, where a better-matched motif associated with the suppressed haplotype (Berg *et al*., 2011).

Men homozygous for the motif-suppressing allele T at SNP DA7.5 showed no recombination or extremely low recombination frequency. In contrast men homozygous (G/G) and heterozygous (G/T) for SNP DA7.5 showed regular hotspot crossover frequencies depending on their *PRDM9* status. T/T men completely turned off their recombination activity regardless of the presence of the variant *PRMD9* A allele. Therefore, *cis*-regulation is a major factor that controls the hotspot activity in Hotspot DA.

Subtle changes in ZnF arrays between *PRDM9 v*ariants can have a strong effect on recombination activity. The *PRDM9* variant L20 has been shown to activate the

hotspot MSTM1b, and show an entirely different activation profile between Super-hotspots (Berg *et al.,* 2010). Also, the C variant of *PRDM9* showed hotspot activation (Berg *et al.,* 2011). 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). 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 (Berg *et al.,* 2010). In Hotspot DA, non-activating *PRDM9* variant L4 binds five of the exact 8 bases of the 13-bp motif, including the SNP DA7.5. Therefore, in SNP DA7.5 T/T men, the *PRDM9* L4 variant is predicted to bind to 4 bases for the 13-bp motifs and activate hotspots with extremely low recombination frequency.

Finally, all active and suppressed haplotypes were determined for Hotspot DA, and the disrupting SNP DA7.5 within the 13-bp motif sequence was found to be the only *cis*-regulator and major factor for hotspot activation in this hotspot.

6.6 FINAL REMARKS

Hotspot generation and constant biased gene conversion extinction contribute to the regulation of recombination in the human genome, and this appears to be extremely dynamic. New *PRDM9* variants might create young recombination hotspots, which are then silenced over time through mutation and biased gene conversion. Hotspots would not have to persist given that they are silenced by biased gene conversion, but instead new locations could be activated as recombination hotspots by the generation of new PRDM9 ZnF arrays (Berg *et al.,* 2010).

Intensive biased gene conversion, both in to crossovers and non-crossovers, has been found at Hotspot DA. Biased gene conversion that influences crossover and non-crossover activated hotspot activity correlates with *PRDM9* allele A. In Hotspot DA, the lifetime of the hotspot mostly depends on the *cis*-regulatory disrupting SNP DA7.5, and on the *trans*-regulatory factor PRDM9. Recent studies showed that a neutral system of recombination landscape evolution could only be achieved if PRDM9 could evolve at an equal rate or more rapidly than the time it takes for a motif to be completely depleted (Ponting, 2011).

6.7 FUTURE WORK

Apart from PRDM9, no other protein is known to have a profound influence on recombination activity in humans. The factors that influence PRDM9 binding *in vivo* are still unknown. It also 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. The other important issue is whether determination of PRDM9 specifically depends on local hotspot-specific chromatin complexes. Grey *et al*. (2011) enriched for hotspots in mouse for H3K4 tri-methylation patterns, and it still remains to be understood where these arise through direct interactions with PRDM9.

Positive selection, insertion, deletion and gene conversion affect human PRDM9 ZnFs (Ponting, 2011). Finally, the influence of PRDM9 ZnF variants on minisatellite instability is intriguing, given that these variants themselves are encoded by a minisatellite (Berg *et al.,* 2010). If *PRDM9* influences its own evolution by generating alleles that affect meiotic instability at the *PRDM9* minisatellite and promote the generation of new alleles, a whole raft of new *PRDM9* alleles could be generated whenever an interaction allele appears in a population, implying a potential chaotic mode of hotspot evolution (Berg *et al*., 2010). So, a number of questions in the future will need to be answered. Firstly, how does the ZnF array specify hotspots and what are the rules governing PRDM9/DNA interactions? There is a poor correlation between hotspot sequence and *in silico* predicted binding motifs, in part due to problems with *in silico* predictions. Evidence came from the study of Grey and his colleagues (2011) that mouse *Prdm9* variants bind to their cognate hotspot centres *in vitro* on naked DNA, despite a lack of obviously strong motifs. This study had a limited dataset, but is really important and needs testing in humans by examining interactions between known *PRDM9* variants and known hotspots. A key question is whether *in vitro* binding can identify hotspots, and if so, what exactly does PRDM9 bind to or whether chromatin imposes an additional layer of rules controlling PRDM9/hotspot interactions. On the other hand, how does PRDM9 activate a hotspot? Does it active either indirectly via histone modifications, or directly by interacting with components of the recombination machinery? How does *PRDM9* itself evolve and how does its rate of evolution compare with the rate of hotspot loss from the genome? All evidence so far says that *PRDM9* evolves rapidly, and

considerably faster than the hotspot loss, so that PRDM9 actually influences its own instability (personal communication with Alec J. Jeffreys).

APPENDIX I: DNA SEQUENCES OF TARGET HOTSPOTS

A 15-kb interval of DNA sequences for Target DA, DB, DC and DD, centred on the target SNP, was downloaded from ENSEMBL. SNPs have been characterised from Hapmap and dbSNP data. (SNPs in red are validated and genotyped SNPs, polymorphic in Europeans and Africans, SNPs in cyan are putative but not-genotyped SNPs, and SNPs in scratch-through cyan are SNPs genotyped but fixed in Europeans.) 13-bp sequence motif (black) is shown with matches to sequence.

TARGET DA

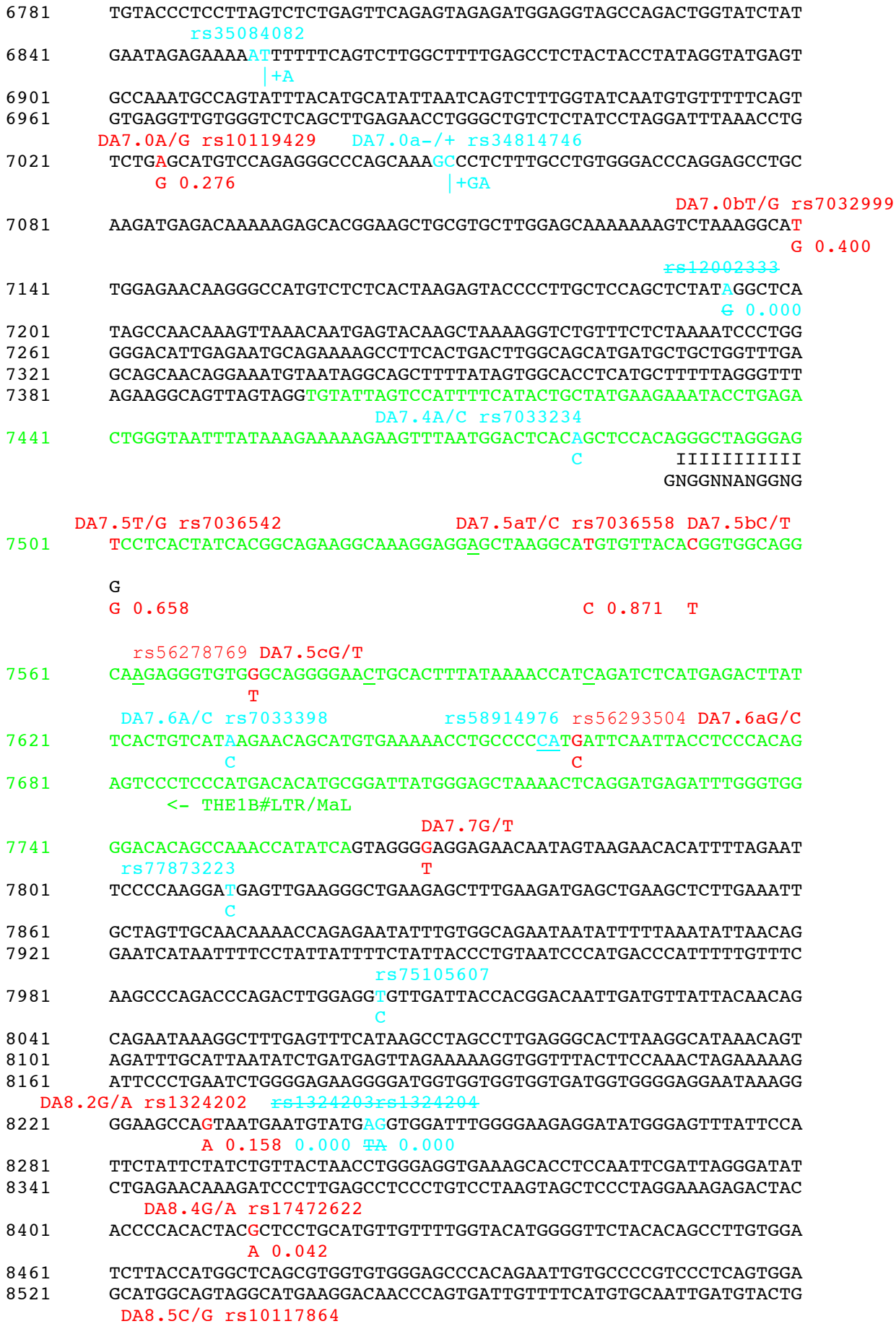
```
                          rs16938276
1       CAGGGGCACCAGTTACATTGTATTCTTTGTGAACGGCCATTCCTGGAGTTTTGTAATGCT
                          C 0.062
61      GTGGCCCTTTTTTCAAGCAAGGCCTTATGAGGACTTCAGGTCTCTAGGAAAAATCTGGTA
                                        rs10811785
121     GAAGGATATGGGGCAGCACCAGTAGAGGAAAGTGGCTTGAGGGGCGAGCCTCTGGAAGCT
                                        A 0.250
181     CTCTTGGCTCTGCTGAACTTTGGTGTGTTGGGTAAAAGGTCTGAAGCTTGCTTCATGGCC
                                  rs16938278
241     TGGTGAGAACTCACAGAGCTGCCAAGAAGGGCCTGGCATCAGAAATAAGCAACTACCAAG
                                  C 0.062
301     GACAGGAAAGTCAACAACAAATGCCAGCCATGAGCATCAGATGATAAACTTCATGCCACG
                                                          rs35119837
361     AGGCCAGGTGAAATTGTCAGAATTGGCAAATGCCAATCCTGAGATCAGATCGACACCCCA
                                                rs7853650       _+G
421     TGCCAGGTTGGCAGAGAGGGGCAGTACCCAGGCCAGCTTTCTCTCACCACCTCAACCCCC
                                                      G 0.000
481     TTCCTATGTCTCCACTCCCCCAGTTATCAGAAATACCCTCAAGCTTTCACCTCCACCCAG
                                  rs11999827                rs34504242
541     AAGCCATCTTGAAAAGCAGGAAGAAGACGGGGGTAAACAGCTCTGTTCAAAGAGTTTAAG
                                  T 0.049                          _+T
601     TTAAAATTATATGAATGTTTGTTCCCAAAAAAGCATTTGAATATGTAGATACTATTAATT
661     TCCAAATTCTGCCCTTATTCTCACCCCATCTCCCTACTATCTAGCAAGATTGGGAGGGAG
721     ACTGGCAGTTAATAAGGAAGACTGGATCAGTTAAAGACTATAAAGAAGCTTGTATTTTGT
781     TTTGTTTTGTTTTTGCCAGCTTAGTGAAGTATGCATTTGCAATCCTGTTCTACCATGTTG
841     GTGCATAGCTCTTAGTCCTTTAGACAAACAGCTCTTAAACAATTCTACTGTGATTTGTCT
901     GAGATTCCTAGAAATCTCCTACCTTCCAAGTAAGGAAGGGATTAATGGCTGAGTACACAC
961     ACTCTGGGATGGAGTTAAAGCATCAGACTCTTGGGGCTGGAGCCATCTGGGAGACCATCT
1021    CATTAATGAAAGGTTTTATTTTCAGACAGAAAACTGCAACCCAAGAAGCTTCTACTTCTT
1081    GTCTTCTGGAACTCTATCTGGGGTCTCTGAGTTGCTGAAAAGAAGCCTGACTACCCACCT
                                  MLT1H1#LTR/Ma ->
1141    ACACATGCTAGAGAACATCTAGACACTCCATTCCCACCAATGTGTTAAACATTGAGTGAG
1201    CTGTCTTAGACCATACTTACTCATTCACCAGATGAATTCCACCTAGTGACTTCAGTTTAT
1261    ATCACGTGAAGAAGAAAAATTGCCAACCTGAGCCCTTTCCGAATTCCTGATCCAAAATCA
1321    TTAGATAAAATAAAATGCCAGTTGTTTTAACTGGTTACAGTTTGGGATGGTCTGCTATCC
1381    AGCCATAGGCAACTAGTGTAGTAGAAAGAAAGTGCAATAGAGTGTTAGAAAAGAAGAAAT
1441    TAATGCTAATTTGGGAGATCTCTGGGATAGGGGGAGGTAGGGTATAAGTATAGCTTTAGA
1501    AATGGCCAGACCTTGGGCAAGTGGTGATGGTGAGAGGCATGGAGAGGAGAGTAAGAAAGC

1561    TGTGGAAGTGGAAGCACATAGAGAATGTTCTGAGCTTGCCCAGAGTGACTGGAGTGAGGC
1621    TGGCATGAATGGGAGCAGCAACAGGGAGAGATTTTTGAAAAAAGTCACTGGTTAATGCCA
1681    AGACTGTGCATGACATTAAGGATGCTTGTTACTACCAAAAAGGAAGTTTTAAAGTGAAAG
```

```
                rs16938284
1741     CCTTGTCTCCGGTTATTGCTCCAGTGTCAGAAGCTGTCTATTGAAATATGTCACAGTGAT
                 T 0.000
                                                          DA1.9-/+ rs34951391
1801     AAACAACTAGGTTCATGTGTTCAGCCTGAACCCTAATGTGAGTTTTCTTTCTCCAGTGGA
                                                                    |+G
1861     AGTACTAAACTGTGGCTTACACTAAATCTATATGACCTTTGGGTGGTTCCCTGGTAAGGA
1921     GTTAAACCTAAAACCTACATCCTCAAAAGTCCCAGAAAGAAATTCTGATTTACCCTTTAT
         THE1D#LTR/MaL ->
1981     CAGTCTATTCTCACACTGCTATAAAGATACAATCTGAGACTGGGTAATTTCTAAACTAAA
         rs7030775
2041     CAAAAGAGTTTTAATTGATTCACAGTTCTGCATGGCTGGAGAGGCCTCAGGAAACTTACA
          G
2101     ATCATGGCAGAAGGCAAAGGGGAAGCAAGGCACATCTTACACGGCAGCAGGAGAGACCGA
2161     GCAAGCACGAGGAAGTGCCACACTTTTAAACCATCAAGATCTCACGAGAACTCCCTCACT
2221     GTCATGAGAACAGCATGGAAGAAACCATCCCCATGATCCAATCACCTCCCACCAGGTCTC
2281     TCCCTCAACATGTGGGGATTACAATTCAAGATGAGAGTTGGGTGGGGACATAGAGCCAAA
                                  DA2.4-/+ rs34397119
2341     CCATATCAGAACCCTTGGTTAATCTCCCTGCAACATAGGGATGGCCATTCCTAGAGTTGT
                                       |+C
2401     GCAATTTGTGCAATTCTGAAATTCCAGCTATGTGTGTAATTCTGAAATTCCTGTATTAAT
2461     GGCCCTTCCTTTTCAGGTTTCAGAATCCCCTTATTAAATCGTCACCATTCTGCTCATTTA
2521     TTTGTTTAACAAATGTTTCTTGATGGCCTACTCTATATGCCAGGCACAGCTTTAGATACT
               DA2.6G/C rs9777039
2581     TCCCAACATTGAGTAAGATAGACAAGCCTGTGGTTCTGGGAGGAACACCATTGGAGGAAA
                         C 0.342
2641     CCAATGGTAAACAAGTAAACAAGCAAGCACACAGGATAATTGCAGGAGGTATAAGAAAAG
2701     AAACTAGCACATGTAACACAAAACCAGCACATGTGGAAAGAGAGTGTGGAATAGTGAGTG
                            rs17406823
2761     AAAGGGGGACAGTCAGAACAGATCAGAGAGTCAGGCAGTGTTGGCCATGGAAAGTCTGTG
                            G 0.042
2821     GGGCCTGGGCAGAAGCTGGTTTTCATTTCAAGCGTAGAGGGAAGCAGTAGGAGAGAGCAT
2881     GAGTAGGTTTATTTTTAAGCAGATCTTTGGCTGATGAAATGAATTGTCAATATTAGTTAT
2941     GAAGTTACTGGGGATTCAACCAGAGGTTGGCACACTTGTTTTCTGTAAAAGGCCAGAGAG
3001     TAAATATTTTAGGCTTTGGGGCAATTTAGTCTCTGTTGCAACTACTCAATTCTGCCTCAC
3061     AAAAGCAGCCACACACAAGATGTGTAACTTTAGTTACAAAAACAGGTGGTGGACTAGATT
                       <- MER58A#DNA/ME
3121     TGTCCCAAGGGCTGTAGTTTGGAGACCACTATATTGGATAATGGGGTGTGGGAAAAGTAG
3181     GCAGTTGAGTGCACCAAGCAGAGTGGGGATGTATGGTTAGTGGGAGCATGGAGAATTCTA
3241     AGAGGCAAGGCCACAGGCTCTGAGAGAAAAGGCAGGTGGAGGAGAAGGATGTCAGTGACA
3301     GATTCTGCATTTTCTAATTTCGCAAGGGCTGACCCCCATGAGCACGTGTGTGCATGTCTG
3361     CGTTCAGCCTGGTGGCAACATGGACCAAAGGCTCTGGTGTACTGGGGTGGCTGGTTTTGT
3421     GGGCCATGAAACCTAACAATATGGGAGTGCTCTTTAAGAAAGTGAGTATAAAATTACAAA
DA3.5T/G rs12554219 rs7850124                                  rs7850134
3481     TACAAAATTGAGGACAGGGCCTTGGAAGAAACGTAGGCAAGTGACAGGCCGTGAAGCGTG
           0.102 G    A     <- MER69A#DNA/Ti                    C 0.000
3541     AGGCTTATTAGCTCCAGGGTAAATTTACCTCTGCTTGTGTATACCTACTCTCTCCCCACA
3601     AGCTCACAGCAACAATCTGAGCCTCATTAAGTCAGACTGTTTCAAATAAAACCTCTTTTA
       DA3.7T/G rs10965506
3661     GTGACATCTGGTGCATGATCGATACTTCGAAACACTCAGAATCCTAAACAAGGAAAAGGG
               G 0.342
3721     CACCTGCTCTGGGTCCTGCTGTGTCTGATTTGTCATCACCAGGGTGCAGGTGCTCAGGCC
3781     TCGCCCCACCCCGCTGTCCTTAGCAACACAAACAGCAGTAACTGGAAGCCACACTGCAGC
          rs7869518
3841     TTGTGTGGCTGATGTTTGCTTTATGTCCTCATACTTCTATTGCTGGAAATGACACTGGAA
             G
3901     CCCAGAAAGAATGTATGCTGGTGCTTTAGGTCATGAAATGGGAGAAATTTTGTGGATTGT
                                       DA4.0C/T rs7851589  rs10811799
3961     GTTTATTCTATCTTATCCTGAAAAATTTTGTCTGTGAGACAGCCCCTCCTCTGGATTCCA
                                                0.078 T     G
                         DA4.0aG/A rs10965510
4021     CAGATCTGTACGGGGAAAATACTGTAATAAGCACCTGCAGTGGGAGCTCAGTCTCCAGTG
                   rs4740673     A 0.331
```

```
4081      CATCTAGGCATGAGTGGAACTGCCCTGTGGAAGGTTGGAGGAGAGGGTGTGCTTGGTAGC
                          T 0.000
4141      TTAGGAGAAAGAGAGAGAGGTAAGGGCAAGGGGGAGAGAGGGATGGAGCCAGTGGATGGT
4201      GAAAGGCCTTGCCCTCTGCAGTGTGAGCTCCCTTCAAATCTCTGCTCTGTCCTTTCCCTA
4261      TTATGCAAATTGTCAGTTAAATTTATGAAGTCAAGGAAGGAAGGAATGAGGGACATAAAG
4321      TGCTTGATGCACATGATTTTCTTTAGAAACAATAGAAGTAGCTGAATTCTTCAGCAAGCA
4381      TGTGGAGGAGGATGTTGGAAGTGGGTGAGTGGTAAGCAAAATGGATAAGGAGGATGGAAG
4441      CCATGGAAACTCGTGGGAATTGTGCGGGAGGCTTAGGTTGCCACAGCTCTGGATTTAAGG
4501      TTGGTTCCAAATTTAACCACATCTATGAAATACAGAGGGGCCCTGCCTGAGAGCTCTGGT
4561      GGTGGCAGATGGCACGTGCCTGGGTCGCACAGCAGCCAGCAGGCACAGCCTGTTGATGAG
4621      CACATGTGAGACATCCCTTTAGGCCTCCAGAAACATTGGGCAATTTAGCAAGGGGCTGAT
4681      TTTCTGTGTGAGTAGATGGGTCTATGGAAATCAATGAGATGTGGGTATAAAAGTCCTGTG
                                    DA4.8A/C rs7020468
4741      ACCTCATTGGAGACCACAGGGGAGAGAGGCCTGGGGACATTTAGGCTGCCATGGGTTGCA
                                                C
4801      TATCTACCTACATCACCTGATGTAAAGTTACTGAGACCACAAACCTTGTCTTATTCATTG
                                                     L2b#LINE/L2 ->
4861      TCATAACTCTTCCTGTCCCCCAATCACCTTTCCCAACTTACCCAGTGCCTAGCATAACCC
4921      CTGGCATATATGGTATTCACTCAGTGAATATATGTGCAATGAATAAATGAATAAAGAATG
                                  DA5.OT/C rs13292475
4981      CATGAATAAAACACTCTTATGCAGGAGTCTGTACCCTGGCCCATGCTATCCTTCAGGCTT
                                             C 0.625
5041      GATCAACTCTAGCTTCTTCCTGGACTTTTGGTCTCTATCTGCCATGAGAGAATTAATGGT
5101      GCTATCACAGCATCATTCAGGCTGACACTTAGAATTATAGAATGGCAGAGCTGGAAAGGA
5161      CCCTAGAGCAGAGTCCAATGCTTTTGATGAAATCTGAGCCCTTAAAAGAAAAGCCAAGTG
5221      ACTTCCATTTATCTGAAATCAAAACAAGCAAGAAATGCGAATGACTAGAATATTTTTTAT
5281      ACACCCTATTTATAAGAATACCTTATATATGTGAAGGTTTTTATGGGATATCAAGTTCTC
                               DA5.3A/T rs7021249
5341      TCACTTACATTATCTCACTGGATTTTCATAGTATATTTTTAAGTAGTCATTGATCTCCGT
             rs7035418                 T 0.093
5401      TTTAGAAAGGGTGAGTGACTTAACCCAGATCATTCAGCTGGCTAATGACAGCCGGTGTTC
                    T 0.000
5461      AAAGTGGTTGCTTTAGTTCTTTCTCACACAGCTACAAAGAACTACCTGAGACGGGTAATT
5521      AATGAAGAAAAGAGGTTTAATTGGCTCACAGTTTTACAGGTTGTACAGGAGGCATTGCTA
5581      GGGAGGCGTCAGGAAACTTACAATCATAGTGGAAGGTGAAGTAGAAGCAAGCACGTCTTC
5641      ACATGGCTGGCAGGAGAGAGAGAGGGAGGGAGGGAGGGAGGGAAGAGAGAGAGAGACC
5701      AGAGAGAGAGAGAGAGAAGGGGGAGGTGAGACACACTTAAACAACCAGACCTCATGAGGG
5761      AACTCTATCATGAGACAGCACTAGGGAATGGTGCTAAACCATTAGAAACCACCCCCGTGA
5821      TCCAGTCACCTCCCACCAGGCCCCACCTCCAACACTGGGAATTACAATTCAGCATAAGCT
                         <- MSTA#LTR/MaLR
5881      TTGGGTGGGGACACAGAGCAAAACCTTATCAGTGGTCTACCGATTCCAGGTCCAATGTGA
5941      TATCCACCATCTGGCAGGAAGCTTACCAGTTCATGGATAACAAATAGGCAAACTGATCTC
                                                DA6.0A/G rs12684262
6001      AAGCATCTCCTTGGTAAGAAACACAGCTTCTAGGTGCTCTCAGTCTCCTATTTCTTAGAT
                                                                  G 0.102
6061      AACTCCAGAGCTATATGGCAAAATTCTAGATAAAAGACACATGCTGTTCCACTAATTCTG

                        DA6.1C/T rs10965522          DA6.1aA/G rs13294267
6121      CCAAGCTGTGAGAGAGTCCGAGTTTCCCATTGCCCCTGTAACAAATTACCACAAACTTAA
                               T 0.267                              G
6181      TGTTTTGGATGTCTTATGGTTCTGGAAGTCACATGTCGGAAATGAGTTGTACAAGGCTAA
6241      AATCGAAGTGTCAGCAGGGTATATTTCTTCTGCAGCTCTCAAAGGGGAATCTGTTCCTTT
                                  DA6.3T/C rs13298799    DA6.3aA/T rs13294613
6301      CTCCTTTCAGCTTCTAGAAGTTGCTGGCATTCCTTGGCTTGTAGCCCTCTCACTCCAATC
                                                C 0.250          T 0.255
6361      TCTGCTTCTGTCCTCACATAGCTTTCTCCCCTGACTCTCTGACTCCTCCTGCATTCCCTT
6421      TATAAGGACTATTGTGATTACATTAGGCCCAGAATAATCCAGGATAACCTCCCCATTTCA
6481      AGATCCTTGATTTAACCATCTTGACAAAGACCCTTTCGCCATGTAAGGTTAAATTCACAG
                                                <- MLT1D#LTR/MaL
6541      GGTCTGGAATTAGAGCATGAACATATTTAGGGGATATTATTCAGCCTACCACAGGGAAGA
6601      ATTTTACTTTTATAATGTGTCAGGTTCAGCAGTTTTCAAAGAAATATTTAAAAATTATAC
6661      TTGCCAAAGGGTTTATCTGAGAAGAGAATGGGAAGATTATGGTATAAAAGCCACTGTACT
6721      CCCAACAATGAGGAAGAAGCTATCTTCCTCAATTTCTCTAGGGCATGGACTGTCCTCCCA
```

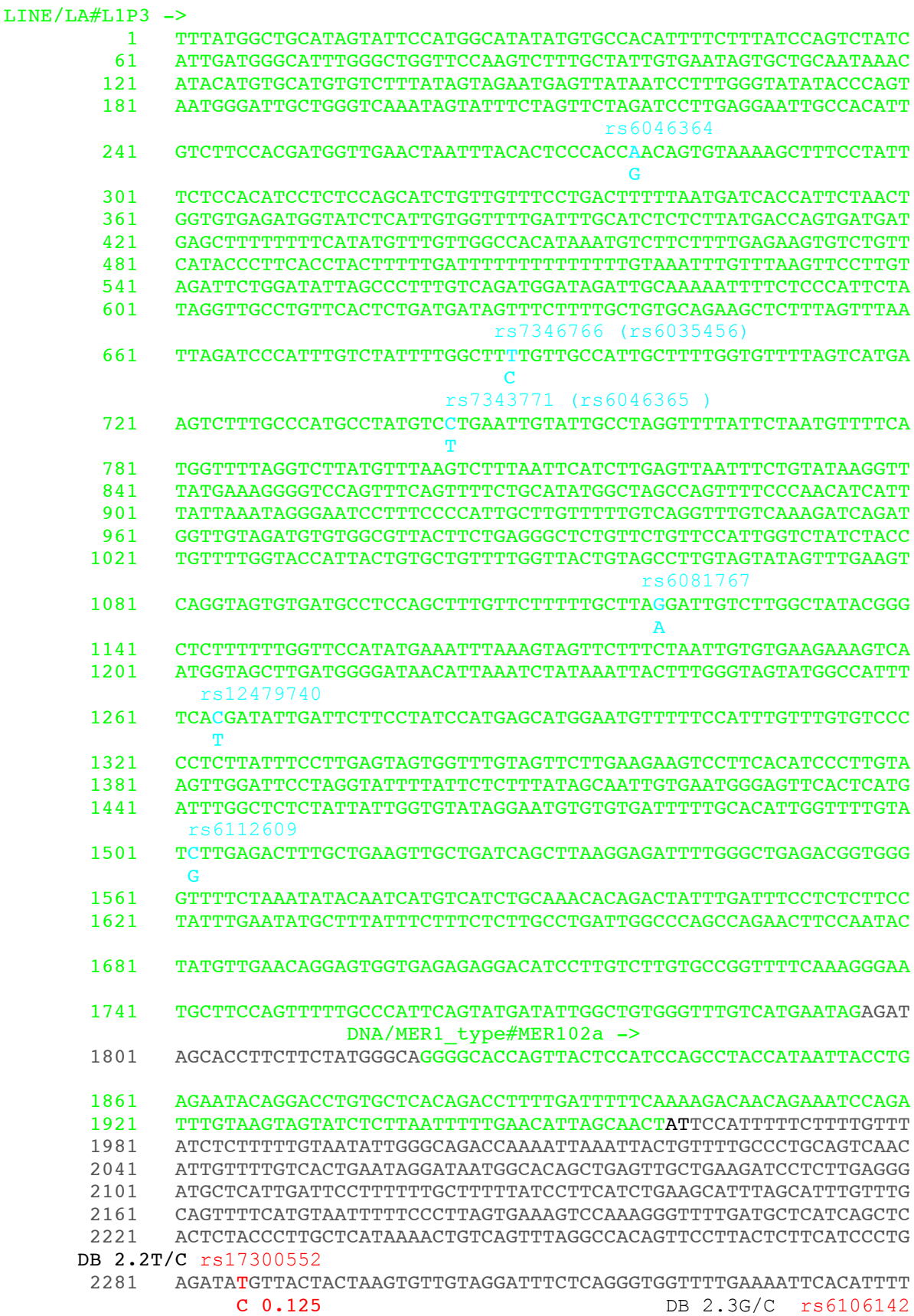

6781 TGTACCCTCCTTAGTCTCTGAGTTCAGAGTAGAGATGGAGGTAGCCAGACTGGTATCTAT
rs35084082
6841 GAATAGAGAAAAATTTTTTCAGTCTTGGCTTTTGAGCCTCTACTACCTATAGGTATGAGT
+A
6901 GCCAAATGCCAGTATTTACATGCATATTAATCAGTCTTTGGTATCAATGTGTTTTTCAGT
6961 GTGAGGTTGTGGGTCTCAGCTTGAGAACCTGGGCTGTCTCTATCCTAGGATTTAAACCTG
DA7.0A/G rs10119429 DA7.0a-/+ rs34814746
7021 TCTGAGCATGTCCAGAGGGCCCAGCAAAGCCTCTTTGCCTGTGGGACCCAGGAGCCTGC
G 0.276 +GA
DA7.0bT/G rs7032999
7081 AAGATGAGACAAAAAGAGCACGGAAGCTGCGTGCTTGGAGCAAAAAAAGTCTAAAGGCAT
G 0.400
rs12002333
7141 TGGAGAACAAGGGCCATGTCTCTCACTAAGAGTACCCCTTGCTCCAGCTCTATAGGCTCA
G 0.000
7201 TAGCCAACAAAGTTAAACAATGAGTACAAGCTAAAAGGTCTGTTTCTCTAAAATCCCTGG
7261 GGGACATTGAGAATGCAGAAAAGCCTTCACTGACTTGGCAGCATGATGCTGCTGGTTTGA
7321 GCAGCAACAGGAAATGTAATAGGCAGCTTTTATAGTGGCACCTCATGCTTTTTAGGGTTT
7381 AGAAGGCAGTTAGTAGGTGTATTAGTCCATTTTCATACTGCTATGAAGAAATACCTGAGA
DA7.4A/C rs7033234
7441 CTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGGCTAGGGAG
C IIIIIIIIIII
GNGGNNANGGNG
DA7.5T/G rs7036542 DA7.5aT/C rs7036558 DA7.5bC/T
7501 TCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCATGTGTTACACGGTGGCAGG
G
G 0.658 C 0.871 T
rs56278769 DA7.5cG/T
7561 CAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGAGACTTAT
T
DA7.6A/C rs7033398 rs58914976 rs56293504 DA7.6aG/C
7621 TCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACCTCCCACAG
C C
7681 AGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATTTGGGTGG
<- THE1B#LTR/MaL
DA7.7G/T
7741 GGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACATTTTAGAAT
rs77873223 T
7801 TCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCTTGAAATT
C
7861 GCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATATTAACAG
7921 GAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTTTTGTTTC
rs75105607
7981 AAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATTACAACAG
C
8041 CAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCATAAACAGT
8101 AGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTAGAAAAAG
8161 ATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGTGGGGAGGAATAAAGG
DA8.2G/A rs1324202 rs1324203rs1324204
8221 GGAAGCCAGTAATGAATGTATGAGGTGGATTTGGGGAAGAGGATATGGGAGTTTATTCCA
A 0.158 0.000 TA 0.000
8281 TTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTCGATTAGGGATAT
8341 CTGAGAACAAAGATCCCTTGAGCCTCCCTGTCCTAAGTAGCTCCCTAGGAAAGAGACTAC
DA8.4G/A rs17472622
8401 ACCCCACACTACGCTCCTGCATGTTGTTTTGGTACATGGGGTTCTACACAGCCTTGTGGA
A 0.042
8461 TCTTACCATGGCTCAGCGTGGTGTGGGAGCCCACAGAATTGTGCCCCGTCCCTCAGTGGA
8521 GCATGGCAGTAGGCATGAAGGACAACCCAGTGATTGTTTTCATGTGCAATTGATGTACTG
DA8.5C/G rs10117864

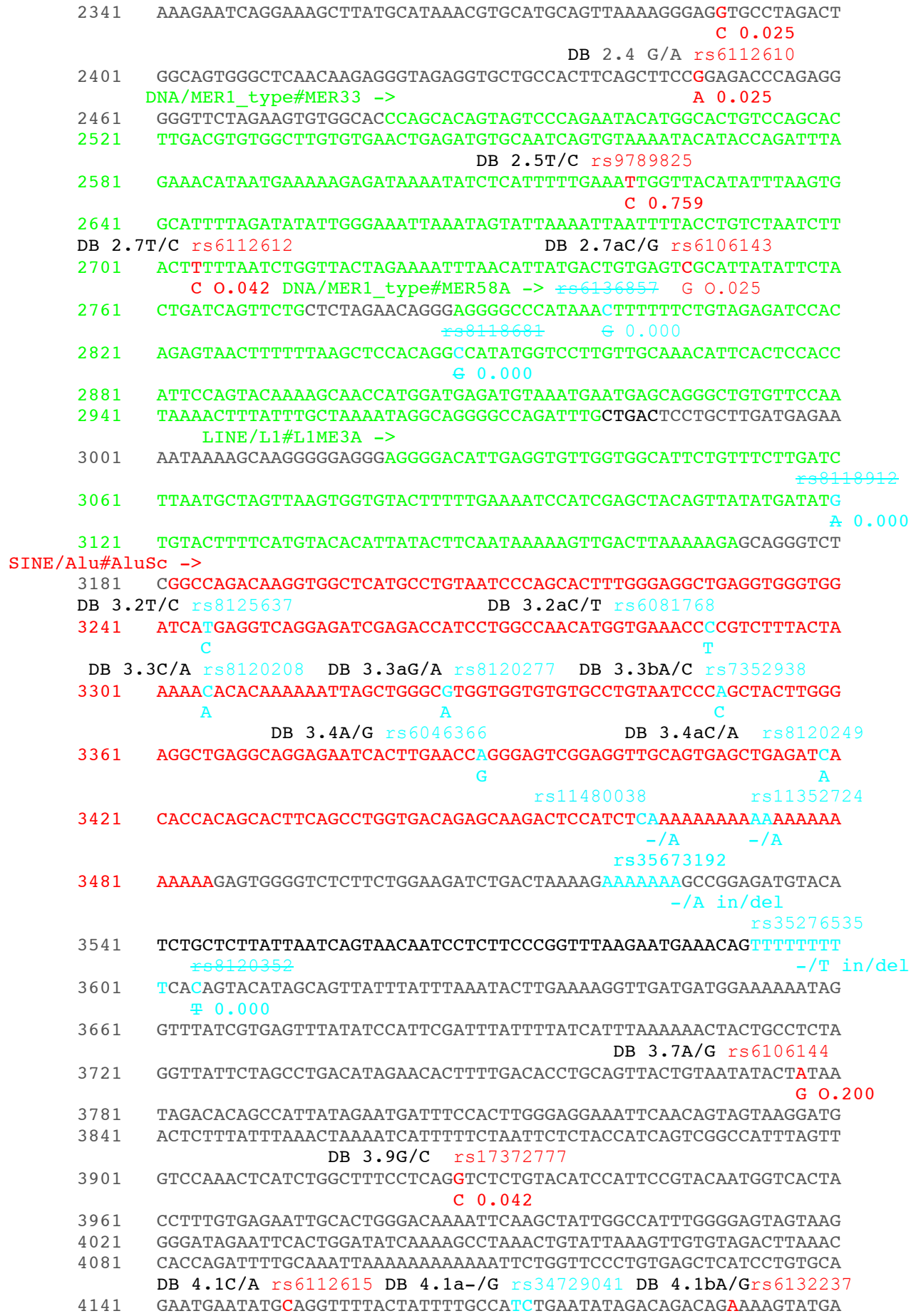
```
8581     TTGGAAATACGAATGTCTCAAAGGATGTTGATCTGGAGAGACGATGACATAGAGTGCCTC
                  G 0.975  DA8.6 rs10117889
8641     CTTCTCACAATGATACCACTTAAACCCCCTCCCCAGACTTACACACAATCCCAGGGAAAA
                         T O.OOO
                           0.308 (AFRICAN)
8701     GGTAAATCCAAAATAGAATGTGATTCAGTCTATGCTATTTTGGCATAAAGGGAAATCAAA
8761     ATAGAAATCATATTGTTCAGTTGCAGAAAAATAAATACAATCGTACTTTGTTGACATCTG
             rs16931305                                      rs13288842
8821     AGTGAGGGACCACACTTTGCATTCTCTACATTCTGTAGTTGCATTTGCAGGTGACCTGA
              C O.OOO                          rs13290069     T O.OOO
8881     AACTCCCTGGTGTATAGCAAGGTACCATTTTGATGAGGCAAGAATTTCAATCCTTCACGT
                                                    T 0.000
8941     ATGGTAATGAGATATTTATCTGAAATGTGCTTCTATTTGATCAGCAGCATCTCAGTGTTG
9001     CTCTTGCATATGAAGTTACTCTCAGTGAGGGTCTGAGCAGATTACATAAAATTTAATTTT
9061     TATTTACCTAATATATTGCCCAGGACTTGGGGGAATGTGATCACAGCAAATTCTATGCTG
9121     TAATGATTTTGTTAATCTGAATATTTGGAACAGCTTTGGGATATATTTAAGTCATTTGTG
               rs10123974
9181     ATGTTTGATGTCTTCAAAAATTGATCAGCAGTAGAGAATATTTAGAGAAAATAGATTATT
                   T 0.000
   DA9.2+/- rs34794098
9241     TTTTGAATGTGTTATGCAAAGTTTTGAACATATCCCAAATAAGAGAGAACAGTATAATGT
          -
9301     TCCCCCATGTTCCCATTCCCCAGCTTCAACTATTAATATTCTGCCAATCTGAAAATACAG
9361     AATTTTAAAATGGCAAATGCTTCATTTGAAAAGTGAGGAAACAGAGGGGAGGGGTACAAG
9421     GGATCTTTCCAGTCACACATCAGTGACCTTTCTGCTCTGAAGTGAATAAAATTATGCACT
9481     AATCAATATTTAATAGCCATCATATTTCATAGTTCAGCCATTCTGAATATATGGGAGATT
                                        DA9.5C/T rs17472664
9541     GCCCCCAAATCTTGATTGAGCCTGAATCCTCTGCAATTTCTGAAGAACTCCCAAAGTGAA
                                                  T 0.188
9601     GGTGGTTAAGCACAGGGTTGCCAAACACTGGGACTAAGCTTAGGCAGTGAGGCACCTAGG
9661     ACACAACATTTAAGGGGCCCTGAGTCTCAGGTTCACCCATACACAGGGCCAGCACCTCAT
9721     TCACTCCATCCCATCCCTTGGGCTACCACATGCTCATTCCAATTTCTGACTCATATTAAC
9781     CCCAGGATCCCCAAACTAGCTGCTCAACCGCCTTGACTCTTGGGATCACACACTTTTGTG
9841     AGAAGAGGTGGTCGACAAGGCACGACTGACTCCTCCTGCACTCACAAATTCACTTGATGT
9901     GAAGTGACTCCTGGATGCCAGAAATGAGGCTAGATGCCTGGTGTGAAGATAAGTCAATGG
9961     CCTTGACTTTGAGGACATTGCAGTCTGCTAGGGATTCTCTCTAAATGGTTTATTTTTCCA
10021    TTTGCATTTGAAGCAGTAAACTCTAAACCAGTATTAATTCCAGGGAGCACATGGACTATG
                                        DA10.0G/A rs10965543
10081    GCTGAATGGAAACTATTTATTCCTGGCAATTTAGAGGATGCTGGAGGGGGTAGTAGTAGT
                                                    A 0.239
10141    TGACAATACATTTTTCCTATAGTTGTTTATACTTGATTCTCACAGCAATTCTATGAGATG
       DA10.2C/T rs7026913
10201    GGCATTATTTCATTCTTTTAGAATTGACAAAATGTTGGCTTAGCAGTATTGAGTAATTTG
           T 0.358
10261    CAAAAGATCCAGCAACAAATGTTTCCCATTGGGATCTGAACGCTAGCCTAGTGCCTCAAA
10321    TTAAATGTCTTATGACATGATAACTACAACAGTCAAAAGGTCCCCACTAGAAATGACAAA
10381    GGAATATTGCAAAACTGGGGACTTGGAAGGAACATGTTACAAAGCAGCTACCAAGGTGGT
10441    GAGAGAGGCCCTCGGGATGGCACAGACACGCCCAAATTCTCATCCTCACCTGCAGTTCTT
             rs9987957
10501    CCCTCGCCATCCTCAGATTATGATCCATAAAGCCTCTCTGAAGAGGCAGCCGACAGCCCT
             A 0.000
  DA10.6G/A rs1981006                DA10.6aA/C rs2376221 DA10.6bC/A rs9987682
10561    GTGCTTGCTAACAACCTGCGTATCCTCTGAATAGTATCTCCATACTCTCCTGACAATGGG
           A 0.229                                    C     A 0.438
10621    GAGTTCCTGTTACTGTCTGCTGTTAGAGGCAGGAAGCCACAGAAGAGAATGGCTTCTTCC
10681    TAGCTTGAATGCTTCAGTTCCATAATGGAGCTGACTTCCAGGCACTGATGGGGCCTGGGC
10741    TAAGAAGCAGTGTGGAGTGGGGACTCTTTCCAGGAGGCAAAGGATGCTGGGATTTCATAT
10801    TTGAAAGTGAATTCAATTCATCCCAGCGTCACTGATGACTAGTAAAAAACCAATTTTGGA
       DA10.8-/+ rs35200569
10861    GTCTACTGGGTTGGAAGGGTGGGACAGTCCTCACACTCTTGTGGAGCCCTGAGTTTGGTG
           |+T
10921    TCAGAGATCTCTGAATGAGATCAGACATAAAGCCATAGGGGTATGCTAGCAGACTCAGAG
10981    GAAGGCTATCATACAGCTCTAACCTGCATCCACATTGTTTGTGGTGCCCTGTGAATTAAA
```


11041 AATACATAGAAACTCTGTGTCTTTAAAATAGCACAGATATGTGTATACTGTTAACTTTAA
11101 AAATAGCAATATTTACTGCATTTTGTTACCTTCTTTTGACATTTCTTCATTTTAATCTTC
11161 AGAAGCATCAAAATGTGAAATTGGTCTTTTATGACATCTGAAGGTCTTGCCTAGCCTTAA
11221 TTTTTTAGGAGGCCTAGTCCCAAGAAAAGAGTATTGTTTCATAAGCCACTGATGTGTTCA
11281 GGGTCTAGATTTTATTTAACGTCTTCAAACCACATCAAGAAATTACTTAATAAGCTGTGT
11341 CTACCACCGCAGAGTGACTTTGAAGAGTTTGAGGCAGTTGCCTGAGTTATTATAGTCCTA
11401 AAAGCTGTTCATCTCAAATATGCAGGTAGTAGACACAAGGATGACTTCACCTTTATCTGG
11461 GTAATTGTATCAATTGCCCTGTAAATCACAGTAAATCAATGTATTATATAGGTGTGTATG
DA11.5C/T rs10811811 AluSx ->
11521 TATAACACTATAAAATATATAACATGGCCTGGCTCAGTGGCTCACACCTGTAATCCTAGT
rs2039352 T 0.229
11581 ACTTTGGGAGTCTGAGGCGGGAAATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCA
C
11641 ACATGGTGAAACCCCATCTCTACTAAAAATACAAAAAAATTAGCTGGGCGTGGTGGCGGA
11701 TGCCTGTAATCCCATCTACTCAAGAGGCTGAGGCAGGAGAATCGCTTGAACCTGAGAGGC
11761 GGAGGTTGCCGTGAGCCTAGATAAAGCCACCACACTCCAGCCTGGGCAACAGAGTGAGAC
rs35199905
11821 TCTGTCTCAAAAAAAAAAAAAAAAATTATATATATATATGTATATAAAACATTGAATAAA
--
11881 ATATTTTTATACTTTAGTACTCATGTTCACAATCAGTTCTTAACGTTTAAAAAATAATGC
Rs9987427
11941 ATTGTAGGTTATATAAACATTTCTCCTACTTTCGTTTACATCCTGTTTTCTAGTCAACGT
A 0.000
12001 CTTCCTTGTTTGGTTGTTTAACTTTGGAACATGTGCTTTGTTGGTTCAGTTCAGGATTCA
DA12.0C/T rs13287382
12061 CGCGCCCAGTACCTAACCTAGTCATGCAGTGATGTGTCACAGTTTCTACATCTAGGTCCC
T
12121 TGTGATGGTCCTTGGTGCTGTTCACCAGCTATTTCCAGGCATTTCCTGTTTGCCTTGTGT
12181 TGGGGAGGCCATGTGACTTTCTCTGGCTGAATTTTAAGTGGAAGCACATGTTCCTAGGCC
12241 AGACCATTTAACTCCTGACATAATACCTCTGGAGCTTTCTTTGCCATTTCTGGGGTAACC
12301 AGGAATCTTGCAGACTATGGCTGCTCCATCAGCTTAAGTGTTGGAGTTAAGATGATCTGG
12361 AACATAAAAGAGAAATAAACTTTTGTTGTATTAAGCATCTGAGATCCTTTTTTTCCCCAT
12421 AACATAGCCTACTCCATCTTGACCAGTAGAGAATGTACTACATTTCTTTTAATTTTTTAA
rs10965555
12481 AAAACATTTTTATTTTGTGGATACATAGTAGGTGTACGTATTTAGGGGGTACATGCGATA
C
12541 TTTTGGCACAGGCATGCAATGTGAAATAATGGTATTATGGAAAATGGGTATCCATCCCCT
rs2039353 rs2039354
12601 CAAGCATTTATCCTTTGTGTTACAAACAATCCAATTATACACTTTTAGTTGTTTTAAAAT
G G
12661 GTGCAATAACATTATTATTGTTATTATAGTCTCTCTGTTGTGCTATCAAATAGTAGGCCT
12721 TATTTATTCATTCTAGCTATTTTTTTTTGTACCCATTAGCCATCCCCACCTTCCCAGCAA
12781 GGCTCCCTTCCCCAATACATTTCCCAGCCTGTGGTAACCATCCTTCTACTCTCTATCTCC
12841 ATGGGTTCAATTGTTTAATTTTCAGATCCCACAAATAAGTGAGAACATGCAATATCTGTT
12901 TTTCTGTGCCTGGCTTATTTCACTTAACATGATGATCTCCAGTTCCATTCGTGTTGTTGC
12961 AAATGTTCCAAATGACTGAATCTCATTCTTTTTTTATGGCTGAATAGTAATCCATCGTGT
13021 ACAAGGACCACAGTTTTTTAATCCATTCATTTGTTGATGGACACTTAGGTTGCTTCCAAA
13081 TCTTGGCTATTGTGAACAGACCTGCAAGAAACAAGGGACTGAAGACATCTCTTCAGTATA
13141 CTGATTTTCTTTCTTTTGGGTATACACCCAGCAGTAAAATTGCTGGATTGTATGGTAGCT
13201 CTATTTTTTATTTGCATTTCTCTGATAATCATTGATATTGAGCATGTTTTCATATGCTTG
13261 TTTAGCATTTGTATGTCTTATTTTGAGAAATGTCTATTCAAATCTTTTGCTCATTTATTA
rs12003624 DA13.4A/G rs10811818
13321 ATCAGATTAGATTTTTTCCTGTAGTGTTGTTTGAGCTCTTTGTATATTCTAGTTATGAAT
G G 0.500
DA13.4a-/+ rs34370591 DA13.4bC/T rs34775873
13381 CCCCTATCAGATGGGTAGTTTGCAAATATTTTCTCTTATTCTGTGGGTTGTCTCTTCATT
+T| T
13441 TTGTTGATTGTTTCCTTTGCTGTGCAGAAGCTTTTTAACTTGATGTGATCCCATTTGTCT
DA13.5T/C rs7859510rs10965560
13501 ATTTTTGCTTTTGTTGCTTGTGCTTGTGGGGTATTACTCAAGAAATCATTGCCCAGATTG
0.550 CA
13561 TTGTCTTGGAGAGTTTCCCCAATGTTTTCTTGTAGTAGTTTTATAGTTTGAGGTCTTAGT
13621 TTTAAGTCTTTAATCCATTTTAATTTGATTTTAGCATATATCAAGAGACAGGGGTCTAGT

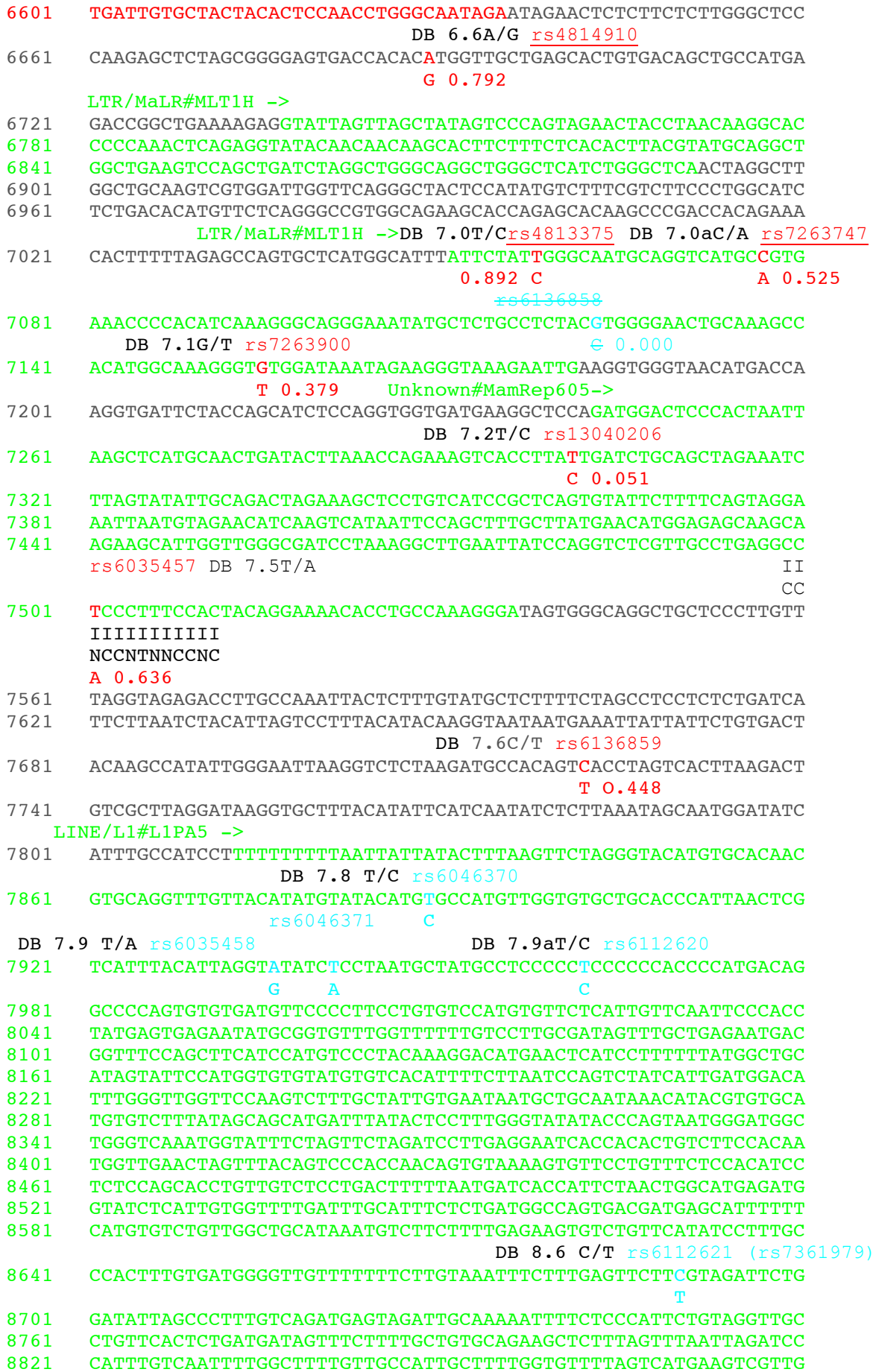

13681 TTTATTCTTCTAGGTATGGATATCTGGTTTTCTCAGCACCATTAATTGAAGAGGCTGTCT
DA13.8-/+ rs35857441
DA13.8aG/A rs7869886
13741 TTTCCCCAGTGTGTGTTCTTGGCACCTTTGTCAAAAATGTTTGTTGTAAGTGTTTGGGTT
+C
A 0.617
DA13.8bC/T rs10965562
13801 TGTTTATGGATTTTCTATTGTGTTCCATTGGTCTATGTGTCTATTTTTATGCCACTACCA
rs34158483
T 0.284
13861 TTCTGTTTTGGTTACTATAGCTCAGTAGTATAATTTGAAGTCAGGTAATGTGATTCTTTC
C
13921 AATTTTCTTCTTTTTGCTTAGGATAGCTTTGGCTATTCTGGGACTTACGTGGTTTCATAT
13981 AAATTTTAGGATAGGTTTTTTTCTATGTCTGTGAATAATGTCATTGGTATTTTGATAGGG
14041 ATTGCATCAAATCTGTAGATTGCTTTGGGAAGTGTGGATGTTTTAACAATATTTATTATT
14101 CCAATCCATGAACATGAAATATTTTTCTATTTTTTACTGTTCTCTTCAATTTCTTTTATT
14161 AGTATTTTATAGTTTTCATTATAGAGATCTTTCACTTCTTTGGTTAATTTCTAGGTATTT
rs10965564
14221 AATTTTATGTGTGGCTCTTGAAAATAGGATTACATTTTTTATTTATTTTTCAGACTGTTC
T 0.000
14281 CCTGTTGACATATGAAAATGGTATTGATTTTTGAATGCTGAGTTTATGTCCTGCAACTTT
14341 ACTGAATTTGTTTATCAGTTCTAACAGTTTTTTGGTGGAGTATATAGGTTTTTCCAAATA
Rs10811819
14401 TAGGATCATACCATCTGCAAACAAGGATAATTTGACTTCTTCCTTTCCAATTTGGATACG
DA14.4T/C rs11531720
C 0.000
14461 CTTTATACCTTTCTCTTGTCTGATTGCTCTAGCTAGGACTTCCAGTACTATTTTGAATAA
C
14521 TAGTGGTGAAAGTGGGCATTCTTACTGTGTTTCAGATCTTAGAGGAAAAGCTTTCAGTTT
14581 TTCCCCCATTCAGCTGTGGGCCTGTTGTATATGGCCTTTATTATGTTGAATATGTTCCTT
14641 CTATATACAGTTTTTTGAGGGATTTTATCATGAAGGGATGTTGAATTTTATCAAATGCTC
14701 TTCAGTATCAATTAAAATGATCATATGGATTTTGTCCTTCATTCCTTTGATATGATGTAT
rs12377063
rs34900306
14761 CACATTGATTGATTTACATATGTTGAACCATCCATGCATCCCTGGGATAAGTCCCATTTG
T
A
14821 GTTATGATGAATAATCTTTTTAATGTGTTGTTGAGTTTGGTTTGCTAGTATTTTCTGAGG
14881 TTTTGCATCAATATTTATCAGTGATATTGGCCTATAGTTGTTTTTGCTTTTGTTGTTGTT
<- L1
14941 GTTTGATGTGTCCTTTGTCTGGTTTGGGTATCATGGTAATACTGACCTCACTGAGTTTGG

TARGET DB


LINE/LA#L1P3 ->
rs6046364
rs7346766 (rs6035456)
rs7343771 (rs6046365)
rs6081767
rs12479740
rs6112609
DNA/MER1_type#MER102a ->
DB 2.2T/C rs17300552
C 0.125
DB 2.3G/C rs6106142


2341 AAAGAATCAGGAAAGCTTATGCATAAACGTGCATGCAGTTAAAAGGGAGGTGCCTAGACT
C 0.025
DB 2.4 G/A rs6112610
2401 GGCAGTGGGCTCAACAAGAGGGTAGAGGTGCTGCCACTTCAGCTTCCGGAGACCCAGAGG
DNA/MER1_type#MER33 ->
A 0.025
2461 GGGTTCTAGAAGTGTGGCACCCAGCACAGTAGTCCCAGAATACATGGCACTGTCCAGCAC
2521 TTGACGTGTGGCTTGTGTGAACTGAGATGTGCAATCAGTGTAAAATACATACCAGATTTA
DB 2.5T/C rs9789825
2581 GAAACATAATGAAAAAGAGATAAAATATCTCATTTTTGAAATTGGTTACATATTTAAGTG
C 0.759
2641 GCATTTTAGATATATTGGGAAATTAAATAGTATTAAAATTAATTTTACCTGTCTAATCTT
DB 2.7T/C rs6112612
DB 2.7aC/G rs6106143
2701 ACTTTTTAATCTGGTTACTAGAAAATTTAACATTATGACTGTGAGTCGCATTATATTCTA
C O.042 DNA/MER1_type#MER58A -> rs6136857 G O.025
2761 CTGATCAGTTCTGCTCTAGAACAGGGAGGGGCCCATAAACTTTTTTCTGTAGAGATCCAC
rs8118681 G 0.000
2821 AGAGTAACTTTTTTAAGCTCCACAGGCCATATGGTCCTTGTTGCAAACATTCACTCCACC
G 0.000
2881 ATTCCAGTACAAAAGCAACCATGGATGAGATGTAAATGAATGAGCAGGGCTGTGTTCCAA
2941 TAAAACTTTATTTGCTAAAATAGGCAGGGGCCAGATTTGCTGACTCCTGCTTGATGAGAA
LINE/L1#L1ME3A ->
3001 AATAAAAGCAAGGGGGAGGGAGGGGACATTGAGGTGTTGGTGGCATTCTGTTTCTTGATC
rs8118912
3061 TTAATGCTAGTTAAGTGGTGTACTTTTTGAAAATCCATCGAGCTACAGTTATATGATATG
A 0.000
3121 TGTACTTTTCATGTACACATTATACTTCAATAAAAAGTTGACTTAAAAAGAGCAGGGTCT
SINE/Alu#AluSc ->
3181 CGGCCAGACAAGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGTGG
DB 3.2T/C rs8125637
DB 3.2aC/T rs6081768
3241 ATCATGAGGTCAGGAGATCGAGACCATCCTGGCCAACATGGTGAAACCCCGTCTTTACTA
C
T
DB 3.3C/A rs8120208 DB 3.3aG/A rs8120277 DB 3.3bA/C rs7352938
3301 AAAACACACAAAAAATTAGCTGGGCGTGGTGGTGTGTGCCTGTAATCCCAGCTACTTGGG
A
A
C
DB 3.4A/G rs6046366
DB 3.4aC/A rs8120249
3361 AGGCTGAGGCAGGAGAATCACTTGAACCAGGGAGTCGGAGGTTGCAGTGAGCTGAGATCA
G
A
rs11480038
rs11352724
3421 CACCACAGCACTTCAGCCTGGTGACAGAGCAAGACTCCATCTCAAAAAAAAAAAAAAAA
-/A
-/A
rs35673192
3481 AAAAAGAGTGGGGTCTCTTCTGGAAGATCTGACTAAAAGAAAAAAAGCCGGAGATGTACA
-/A in/del
rs35276535
3541 TCTGCTCTTATTAATCAGTAACAATCCTCTTCCCGGTTTAAGAATGAAACAGTTTTTTTT
rs8120352
-/T in/del
3601 TCACAGTACATAGCAGTTATTTATTTAAATACTTGAAAAGGTTGATGATGGAAAAAATAG
T 0.000
3661 GTTTATCGTGAGTTTATATCCATTCGATTTATTTTATCATTTAAAAAACTACTGCCTCTA
DB 3.7A/G rs6106144
3721 GGTTATTCTAGCCTGACATAGAACACTTTTGACACCTGCAGTTACTGTAATATACTATAA
G O.200
3781 TAGACACAGCCATTATAGAATGATTTCCACTTGGGAGGAAATTCAACAGTAGTAAGGATG
3841 ACTCTTTATTTAAACTAAAATCATTTTTCTAATTCTCTACCATCAGTCGGCCATTTAGTT
DB 3.9G/C rs17372777
3901 GTCCAAACTCATCTGGCTTTCCTCAGGTCTCTGTACATCCATTCCGTACAATGGTCACTA
C 0.042
3961 CCTTTGTGAGAATTGCACTGGGACAAAATTCAAGCTATTGGCCATTTGGGGAGTAGTAAG
4021 GGGATAGAATTCACTGGATATCAAAAGCCTAAACTGTATTAAAGTTGTGTAGACTTAAAC
4081 CACCAGATTTTGCAAATTAAAAAAAAAAAATTCTGGTTCCCTGTGAGCTCATCCTGTGCA
DB 4.1C/A rs6112615 DB 4.1a-/G rs34729041 DB 4.1bA/Grs6132237
4141 GAATGAATATGCAGGTTTTACTATTTTGCCATCTGAATATAGACAGACAGAAAAGTATGA

```
                A 0.075              -/G                G 0.061
                 DB 4.2A/G rs6081769
4201   CTTTTCCTTCTAGACCTCGATACTGACCAACTATCAAGACACTTGCTCAGGCCTGCTCAG
                           G 0.058
4261   ACAGAAATGAGAAAGCGTGGAAACCGGGGCCGAGGTTAGAGAACAGGACAGCCATGGACA
4321   CAGGCTCCCAAGGTCAGGGCATGGAAGTGCCAGACCCAGCAGGGTCCTGGTAGCTGCGAG
          DB 4.3C/G rs3950057   DB 4.3aA/G rs34634068
4381   AGCATCGTTCTACCCCGACTCTGCTCCTGCTCAGCAGCAAAGTTGCAGAGAGAGAGAGCT
                               G      G 0.028
4441   TCCAGAAGTGGAAAGATTTTTAAATAAGTCTTTTAACTTTTTGCCTGTGGATACAGCTGC
          DB 4.5A/G rs34181998
4501   GTTGACACAAAGCCACATTTGAGGCATGATGTCTCGGAAGCCTGGCACCACGGACTGTTT
                     G 0.028
4561   TTCTACTGGGACCCATAATTGCTGCCTTTTAAGAGGTGTCCTGTGTTCAGTTTTATCAAT
4621   GTGGTGCTTGAACCGGACTTTCAGCTGACTCCCAAAAGGAAATAGTTTTAAAAGGAAGAA
4681   TTTTAAGTCTGCCAAGAGCCAGATAATGGCATAGCACAGAGAAAAGCCAGGGTTTTGTGG
4741   TTTGTATATTAAAATACAAGTCTCTTAATGCCTCAAAACAGCAGCCTCCAGAAAATGGGA
4801   AGTTTAATAGATGAGGCTAGAGAAAGCAGTTTTAAAGATCGAAATTTTCCCTCTTTGTTC
4861   CTTGTGATTTACCCGAAAGGAACAAGTTCCTTTCTCCTGAACCTCCTCCTATGTCTATTT
4921   TTGGTTTTATACAAGTGACTGTTCTCATTAAAGTAAATCTCAAAGTGCTCTGAAAATGTC
4981   CATGCTGGGCTGCCGGCCTCACGCCTGAGGAGCTTTTGGCCTGGACACTCTGCAAAAATG
5041   TGTCCTGAGGACAAAGGCCATTAACATTTGGTTAATCTAACATCATAAATTACTTTTCCC
5101   ATAAATTACTATTAACAGACCTCCCTCCATACTTTTTCTGAGTGCTGTCTTGCCGTTTTT
5161   AACCTCCCTTACATTTCCACCCTTTTGCCCAACACAGAATCACATGAGAGCCCCTGGAGG
                                                        rs6112617
5221   ACACGGGGGAGGGGTGGGTTCAGCTGTTGTCCTAGCACAGTCTGCAGCTCCTCCGTGCAG
 DB 5.2G/A rs34367386                  rs6112618                A 0.000
5281   GCTGTTGTTTTCCCTGACAGTAATCATCTGTCAGGAGAAATGTTACGTATCCCATCATCA
            A                          A 0.000
 DB 5.3G/A rs17300678
5341   GTGTGGACTCTGGCTGGACTCTAGACAATGCACAGAACCATCAGCCATGGTCCTCACAAA
          A 0.042
                      rs6081770   DB 5.4G/A rs941736 DB 5.4aG/A rs6112619
5401   CCACTTGGCCACTATTGTCTTCTTACCTTCTGTTTGTAGCAGTCACTGCAATAGGTTGGG
                       G 0.000                0.490 A      A
5461   CCAGGGGACAGCAGAGGAGGTTTGGGGATGCCTGAATTCAACTGGAATGCCACCTACACT
5521   TGTATTATCCAACACATGCCCCTGGTGCAGTGTATTTGGTTTTAAAACCAGTTTTAAACG
                    SINE/Alu#AluSq ->
5581   GATGCACATTGTGGTTAGAAATGGCAGACCAGGCTGGGTGCAGTGGCTCATGCCTGTAAT
5641   CCCAGCACTTTGGAAGGCCGAGGCTGGTGGATCATTTGGGGTCAGGAGTTCAAGACCAGC
5701   CTGACCAACATGGTGAAAACCTGTCTCCTAAAAATACAAACAAAAACCAAAAATTAGCCT
5761   GGCATGTTGGCACATGCCTGTAATCTCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCT
5821   TGAACCTGGGATGCAGAAATTGCAGTAAGCCAAGATCGTGCCACTGCACTCCAGCCTGGG
          Simple_repeat#(GGAA)n ->
5881   CAACAAAAGCAAGACTCTGTCAAAAAAAGAAAGGGAGAAAAGGAAGGAAGGAAGGAAGGA
                                  DB 5.9A rs114757 rs3966681 (-/A/AGAAAGA/G)
                                       DB 5.9aGAAG/-rs35906767 (rs11467912)
5941   AGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAGAAAGAAAGGAAGGAAGGAA
        DB 6.0T/A rs6046369    SINE/Alu#AluSq ->   -/del -/del   - in/del
6001   GGAAGGAAGGAAGGTAGGAAGGAAGGAAAGAGAAGAGAAGAGAAAGAAAAAGGGCGAATT
                     A
6061   TGCCTAACCTGCAGTCTGACACTGATCTCTAGGCATGCTGGGTGCTGGGAACTGGGGTGA
6121   TGGAGGGGGTCAGTGGCAAAGAGAAAACCTTGGGGCAGGTGAGATGATGCTGGGGAAGGG
6181   ACCCCAGTGGTGTGGATGTGACACTAGGCCATAGAGGACTCAATCAATAAGAAAGGAAGG
6241   ATTTGCAGTGTCAGGGGTTCCAGACACAGGTGAATCTTGAAGCCAGGTACAGATGAAGGA
                            rs16981220
6301   GATTGGGAAAAGTTTCCAGCCACTGGGTAGGCCAACCCAGGTGCCAGGGACAAAGAACCT
SINE/Alu#AluJo ->            T 0.017 (0.100 AFR)
6361   TGGTCTTGGGGCTGGGCATGATGGCTCAAGCCTGTAATCCCAGCACTTAAGGAGGCTGAG
6421   GCAGGAAGATTGCTTGAGGCCAGAAGTTCTTGAGATCAGCCTGTGCAACATAACGAGACT
6481   TCATCTCTACAAAAACAAAAAATCAGCTGGGCTTGGTGGTGTGTGCCGGTGGTCCCAGCT
6541   ACTCAGAGGCTAAGGTGGGAAGATCGCTTAAGCCTAGGAGGTCGAGAATGCAGTGAGTTA
```


6601 TGATTGTGCTACTACACTCCAACCTGGGCAATAGAATAGAACTCTCTTCTCTTGGGCTCC
DB 6.6A/G rs4814910
6661 CAAGAGCTCTAGCGGGGAGTGACCACACATGGTTGCTGAGCACTGTGACAGCTGCCATGA
G 0.792
LTR/MaLR#MLT1H ->
6721 GACCGGCTGAAAAGAGGTATTAGTTAGCTATAGTCCCAGTAGAACTACCTAACAAGGCAC
6781 CCCCAAACTCAGAGGTATACAACAACAAGCACTTCTTTCTCACACTTACGTATGCAGGCT
6841 GGCTGAAGTCCAGCTGATCTAGGCTGGGCAGGCTGGGCTCATCTGGGCTCAACTAGGCTT
6901 GGCTGCAAGTCGTGGATTGGTTCAGGGCTACTCCATATGTCTTTCGTCTTCCCTGGCATC
6961 TCTGACACATGTTCTCAGGGCCGTGGCAGAAGCACCAGAGCACAAGCCCGACCACAGAAA
LTR/MaLR#MLT1H ->DB 7.0T/Crs4813375 DB 7.0aC/A rs7263747
7021 CACTTTTTAGAGCCAGTGCTCATGGCATTTATTCTATTGGGCAATGCAGGTCATGCCGTG
0.892 C A 0.525
rs6136858
7081 AAACCCCACATCAAAGGGCAGGGAAATATGCTCTGCCTCTACGTGGGGAACTGCAAAGCC
DB 7.1G/T rs7263900 C 0.000
7141 ACATGGCAAAGGGTGTGGATAAATAGAAGGGTAAAGAATTGAAGGTGGGTAACATGACCA
T 0.379 Unknown#MamRep605->
7201 AGGTGATTCTACCAGCATCTCCAGGTGGTGATGAAGGCTCCAGATGGACTCCCACTAATT
DB 7.2T/C rs13040206
7261 AAGCTCATGCAACTGATACTTAAACCAGAAAGTCACCTTATTGATCTGCAGCTAGAAATC
C 0.051
7321 TTAGTATATTGCAGACTAGAAAGCTCCTGTCATCCGCTCAGTGTATTCTTTTCAGTAGGA
7381 AATTAATGTAGAACATCAAGTCATAATTCCAGCTTTGCTTATGAACATGGAGAGCAAGCA
7441 AGAAGCATTGGTTGGGCGATCCTAAAGGCTTGAATTATCCAGGTCTCGTTGCCTGAGGCC
rs6035457 DB 7.5T/A II
CC
7501 TCCCTTTCCACTACAGGAAAACACCTGCCAAAGGGATAGTGGGCAGGCTGCTCCCTTGTT
IIIIIIIIIII
NCCNTNNCCNC
A 0.636
7561 TAGGTAGAGACCTTGCCAAATTACTCTTTGTATGCTCTTTTCTAGCCTCCTCTCTGATCA
7621 TTCTTAATCTACATTAGTCCTTTACATACAAGGTAATAATGAAATTATTATTCTGTGACT
DB 7.6C/T rs6136859
7681 ACAAGCCATATTGGGAATTAAGGTCTCTAAGATGCCACAGTCACCTAGTCACTTAAGACT
T 0.448
7741 GTCGCTTAGGATAAGGTGCTTTACATATTCATCAATATCTCTTAAATAGCAATGGATATC
LINE/L1#L1PA5 ->
7801 ATTTGCCATCCTTTTTTTTTTAATTATTATACTTTAAGTTCTAGGGTACATGTGCACAAC
DB 7.8 T/C rs6046370
7861 GTGCAGGTTTGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACTCG
rs6046371 C
DB 7.9 T/A rs6035458 DB 7.9aT/C rs6112620
7921 TCATTTACATTAGGTATATCTCCTAATGCTATGCCTCCCCCTCCCCCCACCCCATGACAG
G A C
7981 GCCCCAGTGTGTGATGTTCCCCTTCCTGTGTCCATGTGTTCTCATTGTTCAATTCCCACC
8041 TATGAGTGAGAATATGCGGTGTTTGGTTTTTTGTCCTTGCGATAGTTTGCTGAGAATGAC
8101 GGTTTCCAGCTTCATCCATGTCCCTACAAAGGACATGAACTCATCCTTTTTTATGGCTGC
8161 ATAGTATTCCATGGTGTGTATGTGTCACATTTTCTTAATCCAGTCTATCATTGATGGACA
8221 TTTGGGTTGGTTCCAAGTCTTTGCTATTGTGAATAATGCTGCAATAAACATACGTGTGCA
8281 TGTGTCTTTATAGCAGCATGATTTATACTCCTTTGGGTATATACCCAGTAATGGGATGGC
8341 TGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATCACCACACTGTCTTCCACAA
8401 TGGTTGAACTAGTTTACAGTCCCACCAACAGTGTAAAAGTGTTCCTGTTTCTCCACATCC
8461 TCTCCAGCACCTGTTGTCTCCTGACTTTTTAATGATCACCATTCTAACTGGCATGAGATG
8521 GTATCTCATTGTGGTTTTGATTTGCATTTCTCTGATGGCCAGTGACGATGAGCATTTTTT
8581 CATGTGTCTGTTGGCTGCATAAATGTCTTCTTTTGAGAAGTGTCTGTTCATATCCTTTGC
DB 8.6 C/T rs6112621 (rs7361979)
8641 CCACTTTGTGATGGGGTTGTTTTTTTCTTGTAAATTTCTTTGAGTTCTTCGTAGATTCTG
T
8701 GATATTAGCCCTTTGTCAGATGAGTAGATTGCAAAAATTTTCTCCCATTCTGTAGGTTGC
8761 CTGTTCACTCTGATGATAGTTTCTTTTGCTGTGCAGAAGCTCTTTAGTTTAATTAGATCC
8821 CATTTGTCAATTTTGGCTTTTGTTGCCATTGCTTTTGGTGTTTTAGTCATGAAGTCGTTG

```
 8881   CCCATGCCTATGTCCTGAATGGTATTGCCTAGGTTTTCTTCTAGGGTTTTGATGGTTTTA
 8941   GGTCTAACCTTTAAGTCTTTAATCCATCTTGAATTAATTTTTGTATAAGGTGTAAGGAAG
                                          DB 9.0C/T rs6075582 (rs7352902)
 9001   GGATCCAGTTTCAGCTTTCTACATATGGCTAGCCAGTTTTCCCAGCACCATTTATTAAAT
                                                        T
 9061   AGGGAATCCTTTCCCCATTTCTTGTTTTTGTCAGGTTTGTCAAAGATCAGATGGTTGTAG

 9121   ATGTGTGGTATTACTTCTGAGGGCTCTGTTCTGTTCCATTGGTCTATATCTCTGTTTTGG

 9181   TACCAGTACCATGCTGTTTTGGTTACTGTAGCCTTGTAGTATAGTTTGAAGTCAGGTAGT
                  DB 9.2G/A   rs6081771 (rs7352972)
 9241   GTGATGCCTCCAGCTTTGTTCTTTTGGCTTAGGATTGCCTTGGCAATGCGGGCCCTTTTT
                                    A
 9301   TGGTTCCATATGAACTTTAAAGTAGTTTTTCCCAATTCTGTGAAGAAAGTCATTGGTAGC
 9361   TTGATGGGGATGGCATTGAATCTATAAATTTCCTTGGGCAGTATGGCCATTTTCATGATA
 9421   TTGATTCTTCCTATCTATGAGCATGGAATGTTCTTCCATTTGTTTGTGTCCTCTTTTGTT
 9481   TCGTTGAGCAGTGGTTTGTAGTTCTCCTTGAAGAGGTCCTTCACATCCCTTATAAATTGG
 9541   ATTCCTAGGTATTTTGTTCTCTTTGAAGCAATAGTGAATGGTAGTTCACTCATGATTTGG
 9601   CTCTCTGTTTGTCTGTTATTGGTGTATAAGAATGCTTGTGATTTTTGCACATTGATTTTG
                                                 DB 9.6C/T rs7269719
 9661   TATCCTGAGACTTTGCTGAAGTTGCTGATCAGCTTAAGGAGATTTTGGGCTGAGACGATG
                                                          T
 9721   GGGTTTTCTAAATATATAATCATGTCATCTGCAAACAGGGACAATTTGACTTCCTCTTTT
 9781   CCTAATTGAATACCCTTTATTTCTTTCTCCTGCCTGATTGCCCTGGCCAGAACTTCCAAG
                                                   DB 9.8C/- rs35363632
 9841   GCTATGTTGAATAGGAGTGGTGAGAGAGGGCATCCCTGTCTTGTGCCAGTTTTCAAAGGG
                                                             -/C
 9901   AATGCTTCCAGTTTTTGCCCATTCAGTATGATATTGGCTGTGGGTTTGTCATAAATAGCT
 9961   CTTATTATTTTGAGATACGTCCCATCAATACCTAATTTATTGAGAGTTTTTGGCATGAAG
10021   GGCTGTTGAATTTCGTCAAAGGCCTTTTCTGCATCTATTGAGATAATCATGTGGTTTTTG
10081   TCTTTGGTTCTGTTTATATGCTGAATTACATTTATTGATTTGCGTATGTTGAACCAGCCT
10141   TGCATCCCAGGGATGAAGCCCACTTGATCTTGGTGGATAAGCTTTTTGATGTGCTGCTGG
10201   ATTCGGTTTGCCAGTATGTTATTGAGGACTTTTGCATCAATGTTCATCAGGGATATTTGT
10261   CTAAAATTCACTTTTTTTGTTGTGTCTCTGTCAGGCTTTGGTATCAGGATGATGCTGGCC
10321   TCATAAAATGAATTAGGGAGGATTCCCTCTTTTTCTATTGATTGGAATAGTTTCAGAAGG
10381   AATGGTGCCAGCTCCTCCTCATACCTCTGGTAGAATTCAGTTGTGAATCCGTCTGGTCCT
                      DNA/Tc2#Kanga1a ->
10441   GGACTTTTTCTGGTTGGTACATTTGCCATCCTTTGATGGCAACTTACAACCCCTTCTGAC
  rs6035459 DB 10.5A/G
10501   CATGTAGTCATCAAGTGGTTGTCATTTCCTGAATGTGTACAAGTTTGCAACTGGACATCC
         G 0.578
10561   TGAAGTTTCCATCAACCAACCATTTAAGGAGCATTTAAGGAAGGAATATGAGTCTTTCAC
10621   TGACCTCTTCTGTAAAATCAAGAAAACACCAACAAGAGACTGGCAAAACTGACACCGGAG
                                   SINE/Alu#AluSg ->
10681   GCTTGGGGGAAAAAAATCCCAGAGACAAAAATGAAACATCTTGGCCAGGCATGGTGGCTC
10741   ACGCTTGTAATCCTAGCACTTTGGGAGGCCAAGGCGGGTGGATCACAAGGTCAGGAGTTC
                                                   DB 10.8T/C rs6081772
10801   GAGACCAGCCTGGCCAACATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCAG
                                                               C
10861   GGTATGGTGGCAGGCACCTGTGATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCT
10921   TGAACCTGAGAGGCGGAGGTTGCAGTGAGCAGAGATCGCACCATTGCGCTCCAGCCTGCG
                                     DNA/Tc2#Kanga1a ->
10981   TGACACAGCAAGACTCCTTCTAGAACAAAAAAAAAGAAGAAAAATCTTTTAAGAGACGCT
                              rs11907801   DB 11.0A/G rs6112622
11041   ACATCATGAATGTCCTTGAAAGATCATACCCTGTATATGAAAGAACTTTAGGGATATCTT
                                     T 0.00  G
11101   AATTAATTAATTTTGCTTTTATTTTCCTTGTAATGCGCATGGGAAAAACTGTGCATGACG
11161   ACATTCTGTGTCTAAATATGATTTTAGGGTTTTAAGAGTTATTTCAATAAGATTAAAGTA
                    SINE/Alu#AluJo -> rs28722182 DB 11.2G/A
11221   AATTCTAAGAGATAAGAAAATAAATTCAGCCTGGTCAACATAGTGACAACCCATCTCTAC
                                     A
11281   AAAATAAATTTAAAAAATAATTTTAATTAGCAGTAGTTGGGCATGGTGACATTTGCCTGT
```


11341 AGTCCCAGCTACTCCAGAGGCTGAGGTGGGAGGATCGCTTGAGCCCAGGAGTTCAAGACT
11401 ACAATGAGTCACGATCATGTCACTGTACTCCAGGCTGGGAGAGAGAGTGAGACTTTGTCT
Simple_repeat#(GGAA)n ->
11461 CAGAAAAAAAAAAAAGAGAGAGGGAGAGAAAGGGAGGAAGGAAGGAAAGGAAGGGAGGGA
11521 GGGAGGATGGAAGAAAGGAAGGAAGGAAGGAAAGAAGGAAGGAAGGAAGGGAGGGAGGGA
DNA/Tc2#Kanga1a ->
rs35771019
11581 GAGAGATATTGCCATGGTTTAATTGGCAATGTCTGTTTCATTAGATGATACATAAAATAA
rs6081773
rs6132238
I-/A
11641 TGATGTCCTTTACTATTGATGGAGGCTCAAATGCTATTAAATTTTGTATTTCTCCACTGG
G 0.569
G 0.000
11701 AGCATTCCAGCGACTGGTCATCAGAAATGGAACTAGCACCATGAGCGAGGTAGCTTTGGC
11761 TGTGAACTGATTCCAAATCTTGCAACGTGGCCAGTAGTAAAGAAACAAGATGAGAGGCCA
11821 CTTTTTATATGGCTTTTTAAAGCCTTTATATATACATACTGGAATATCTTAGCCAGTATC
LINE/L1#L1ME3 ->
11881 CCCCCTCATTTTTTAAAACCTTAATTTTATTGAGATATAATTTGTATACAGCAAAGTGCA
rs6136860
11941 AAGTCTTAAGTGTACAGCCCAGTGAACTTTAGGCAGAGACCCCGGTAGCCACCATCCAGA
T 0.000
12001 TCAAGGTCTAGAACATTTTCATTATCCCGGATGGTTTCCTGTCCCTTTCCTAGTCAATGC
12061 CTCCCTAAACAAGAATACCCATTATTCTGTTTTCTATTAATAGCAATGTAGATTAGTTTT
12121 GCCTGTTATTGATCTCCATATATATGGAATCAAAATGTATACACCTTTGTGACTGGCCTC
12181 TTTTGCTCAACATAATGTTTTGAGCAAATGCATCCACGTTGCTTCATATATCAGTAATTT
12241 GTTCCTTTTTGTTCCTGGGTACTTATCCATTGTATGAATAATACAATGGATTCATTGTAT
DB 12.3T/A rs7268810
12301 TATTTTCATATTGATGGATATTTAAATTGTTTCTAGTCTGGACTATAATGAATAAAGCTG
A 0.237
12361 CAATGAACATTTCTGTACGTGTATCTTTTTGTAGACACATACATCCATTTATTTTGGATG
12421 TATAACTAAGAGTGGAATTACTGGATTATAGAGCAAGCATAGATATTAGTAGATTATTCC
12481 AGCTTTCTAAGGTAGTTGGACCAATTTTTATTCCCACCAACACCATATGAGAGTTCCAGT
12541 TGCTAACACTTGGTATGATCAGTCTTTTAAATTTTACTTATTCCAGAGGGTGCATGGGAG
SINE/Alu#FLAM_A ->
12601 TGTCTCAATGTGGCTTTAATCTGGATTTTCCTCACTGGATATGGTGGTACACACCTGGAA
12661 TCCCAGCTACTTGGGAGACTGAGGTGGGAAGATTGCTTGAGCCCAGGAATTTGAAGCCAG
Simple_repeat#(TAA)n ->
LINE/L1L1M ->
12721 TCTGGGCAACAAAGTGAGACCCTGTCTCTAATAATAATAATAATAATAATTTGGATTTTC
12781 CTGATGAGTAATAATGTTGAACAGTCTTTTGAATGTTTTAGGTTTGAATATTCTCTGTAG
12841 TGAAGGCTCCATTTAAGCCTTTTGTTCTTTTGAAAATGTGTTTTGTCTTCCTTATATATT
12901 CTAGCCAGAAGTTATTTGTCTTACATGTGCATTTCAAATATCTTCTCCTGCTCTATGTCT
12961 TATTCATTCACTTTTTAATGATGTGTTTTGTTGAACAAAAGCATTTAACTTTAATGAGGT
rs6035460
13021 CCAACTCATCCCTTTTTATGGTCTGTCTTTTTTTAAAAAAAAATATTAAGATGGCTTGAG
A 0.444
DB 13.0C/A rs34800086
13081 TTTTTTATGTCCTGTTTAAGACATTTTTGCCTACTCCTTGTTTTTCTGTCTTACAACCCC
A
13141 TGTTTCTTTTGCCAGTTTAATTGCCTGCCCTTTTCTTGAGTGGGGAGAATAGGAATCAGG
13201 TCAGGCTACATCTGAGTGAGGAAGGAAGGGTGACAGAGTTAAATTAGGAAAGCGAGGCCA
13261 ACAAATTAGGGGAATAGATTGAAGCAGGAAGAGGCCCAACACTGTCTGATTTTTGGTCTA
LINE/CR1#L3 ->
13321 AGGGACCTGAAAGCCTTTGAGTCTGGGTAACCTGATAACTGACTTCAGTGCTGTCCAGCC
13381 CAAATATCTGCATTGCGTCTCTTAGAGTGTTTGGAGAAGCTTGTCAAATTTTATGAAATC
DB 13.4C/- rs34759328
DB 13.4aA/- rs35133530
13441 AACCTCTTCCGTGCTCACAGCCCTGATCTATTGCCTAGAAACATGAAAAAGTTAATGAAG
-/C
-/A
DB 13.5G/A rs6081774
13501 TTCGTGTTTTAAGGACATCGTGCTAATGTCTAGAGACCATCATTTTCCTTCTAAGTGCTC
A 0.438
13561 AAAAACCATCTTGATAATCCCTTCTTGACCATATCCAGGAAATGAAACCTCCATTTGGAA
13621 TAATATTTACTTATTCCTAATCCTCTGACTCCTTCCCCATTCTCCATAAGTTTTCCTTAC
13681 TTACTAAAGATGTCAATAATTCCATGATCTCATCTGTCTTCCTACAATCAAGGGTAAATG
13741 TTTGGGTCTGCTTTTCATAAAGAGACCCTGGCAGGGCCTGAACTGAGCTGACCATGTAGT
13801 GGAGACCAGAACAAAGCAGGAGTTGGCAGGGGGCTCGGGACATGCAGGGAGAAGGCGGGA

13861 CTGCTGCAAACACGGTGGAGGACAGAATACATTTCTTAAAGTAGCTTCAGATTAGGTAGC
13921 AGCAACTGAGTTGAGATCCCTGGGCTCAGAAGATGGACTCTGGAGTCCATAGGGCTTTCT
13981 TGGGAAGATAAACAGCTCTACAGGGAGAAAATACTTCCATGCCAGGCCCTTCCAGGGGGT
14041 CATACTTAGAATGTTGGGACTACATTGATGAATCAGAGTCTCCTTAGAGGCACTAATTGG
14101 CAGGAACAAGATGGGGAGGAGCCATCCAACTTCCAGGGATGAGCTGGATGAGGGCAGGCA
SINE/MIR#MIRb ->
14161 GGATGTGGCTGCAGAGACGCCCCTGGGCGTCCACCCTCAGGCACAGGGAAAGGGAGTATG
14221 GTGTAATAGAAAGAGAGGACTGCTGGGATTCACACATTCCTGGTTCTAATCCTGACTCTA
14281 CCACTTATTAATTGTGTAACATTAGGCAAGAATAGTGCTTCCATCCGAAGATTGACTAGA
DB 14.3G/C rs6081775
14341 GGGTCAAATAGGATACTGTGCATAAAATGCCAATTTCCCCATTTTCCTCCCATTTCTTTG
LINE/L2#L2 ->
C 0.237
14401 GCTCTTCCTTCTGGTTGCTCTACCCAACAGACAGCTTTCAATTCTTATCTTTCTTGCCCA
LINE/L2#L2 ->
14461 CTTTGAAGCACTGGCACCATTAAATAGTCACAAGGTCTTGAAATATAACTTTCCCTTGAC
14521 TTCCATGAAGATCCCATCTATTCCATGACAGTAGCTACCATGTGCATGCTGACAACTCCC
14581 ACGTGACTATCTCCGCACATAACTCCTTCCTAAGAGCCAGATCCATTGATATACCTGCTT
14641 CCTGAACTTTCCACATGGATGTCCTACTGACAGCTCAAACTCAACAATTCCAAATTCAGC
14701 TCATTATCTTTTCCATAAATCTTTCCCTTCCTCACCTGCGTTCTCTATATGGAGAATCCA
LTR/ERVL#LTR16C ->
14761 GTCATCAAGCCCAAAACTTGGGAGTCATGGATTCAAGTGCAAGTGGTTTATTTGGGAGAT
14821 GATCCCCAGAAGCATCACTGGGGGATCTTGGAAGTGAAACAAGGAAGGGAAGGAAGCCAG
rs6046374
14881 TATAGAGTGTTTTATCAAGCCTGGCACCCCTGTGGACAACTGGAACTGAATCCCACTGAA
C 0.000
14941 AAACTTGGAGATGCTGTAGAGAGCACAAGCTTCAGAGTCACTCATCCCAAGGTGCAAAGG
15001 A

TARGET DC

```
(AT) rich/ Low complexity ->
       1 GCCTGGCAACAGTCCGAGACTCCATCTCAAAAAAATAAAATTAAAATTAAAAAGATTTTA
      61 AAAGACCCTTCATTCAGAGTTCAGAAAGCAACACCAGATCCTCACAGCCAGACAGCTCTC
                                                  rs4736153
     121 CCGTCTCTGAACTCAGGGTGTGGAACACTCGGCAGCCTTCGCAGCGTGCCTTCCGTCCCT
                                                  A
     181 CACTTGCTTAACGTATGCCAAGAAGCTGCTCTGACCAGGCCTCACTGCATACTGGGAAAG
     241 GAGAGAGCACACGGCAGCTACGTGGGAGACTCAGGGAAGGATGAGCTCGAGCTCTGCCAC
     301 CACAGGACATGGAGGGAGGGGACACTCAAGGGCTTGCAGGAGGGAGGGGCCGAGTGGGTT
     361 GGCAGAAAAAGAGAGACGGACATTCCAGGTGGAGGGAAGAGCACAGATGAAGGGCATTTC
                                 (GA)n/ simple repeat ->
     421 ACAAAGGCCAGGGTACGGCATGCAAGGGCCCAGCAACACGAAGAGGCAGACAGAGAGAGA
     481 GAGAGAGAGAGAGAGAGAGGGAGACTGGGACCAGATTACATAGAGACTTAGACACCAAAG
             rs28678828
     541 TGAGGAGCCCATTTGTTTTTGAGAAAACTGTAAGACACACCAGAAGATGGATTCAAGAGG
             A
     601 GCCAGGCTAGTGACAGCGAAAGAGTTAACTAGGCTCCCGTGAAGGCTCAGATAAAACGTG
     661 TTTGGCATGTGTACCAAGGCAAAGAGGGAGAGAAGATGGAGGGATATTTAAGAGAAAGAA
     721 CAGATGAGCCTTGGGGAAGAACGGATGGAACTCCAGCGGTGAGAGAACCATCTGCAACAG
     781 CCCCAGGACTCTTGCCGGCACTTCCCAGCAGATGATAATGGCTTTTGCCACCAAAGGAAA
     841 CACAGAAAGGTGAGCCCGTCTGGAGAGAATGACAATCCATCAGAGTTTGAACAAGCTGAA
     901 CTGGAGGTGTTTTCGGGGCTGCCAGGTGAAAAAGTCAGGAAGCCAGCTAAAAACATTGAC
                                 rs7002644
     961 ATGGAATTCAAGACCGGAAATAGACATTTCAGAGTCATTACCATCATGTTAGAGCCTGTG
                                 G 0.033
    1021 AAATCATTTAGGAAAAATATAGTGTATGGTAAGTATTAAGGACAAAACTGTTAGGAACAT
    1081 CCATAGCTGGGGATGCAGCAGTGAAGCAGAGTCGGGAACAGTCAGTGTGGCACGGCACGG
              rs6982187                                   rs34989198
    1141 GCCAGGGGAGGGAAGCTCCCCACCTACAGGAGGACTGGCCAGTCACTGGCGTCTCCCAAG
              A                                           /T
    1201 GATGAGTAAGTGGCAACCACGGAGGGGGAGGGAGGTTAAAAGCCTGATTTCAATGGCGGA
    1261 AAAGTACAAGAAGGATGAAAAAGTGAAGGGGAAAGAGTAAGGAGTAGCCAGAGAGAAACC
    1321 CATGGTTTAGAAGAGGATTTGGGCATATCGGCTGAGGGACTGGTACAAAACTGCTCAAAG
    1381 CTTAGTGCGCAGAACCCCCCAGAGAATGGGTCAGAAACACAGATCGCAGCCCAGCGCCCA
    1441 CAGAAAGGGATGTCGAAGAGCTGGGTGGGCCGAGGAATCTGCGTTTTAGCAAAGACACTT
    1501 TGATGCAGGTGGTCCATGGGCCGTGACCTACAAAAACAAGGGCAAGTTCAAAAGACCAAA
    1561 GAGAAAGGATACCTGATCCGGAACACAAGTCAGACGATTCACCTCTGATGAGAGGAGGAG
    1621 CCCATCTTCCTCTACAACGAAACAGATGGGAAAAACAAGTGTATCCCAGAGGAAAGAGGG
    1681 GAAGCAGAGAAGGTTTCTGCCTGATGGCCTCTGTGGATCTTTGAAAAGAGACCATGTGCT
    1741 GAGAGGAAGGGATACAAGAGTGGGGAAAAGCCTTAAAAAAAGTGGCATGTTGACAACAGC
    1801 TGAAACTAGGCACACACTTGCCCCTGTGTTTTGATGAATGAATGTTTAGTGGGAGAGCCC
    1861 TATGATACAGATAAGCATATTTTCCCCATGGTGTTCAAATTCATGGGATAAAATGGGCAC
    1921 TACTCACCAAAGTCAACCTGTCCCCTTCTCAAGATACAAAAGCAGCAGATGACTTAAAGG
    1981 ACCAGGTTTGCCGCCCACTCTCTATTACCCTCCTCTCTGACAAAGACAGGAAGATTCCTT
    2041 ACATGAACTGCTATCATATCCCCATTTTTGAAAAAAGCAAGTCTCCATTTGTCCACCAGA
    2101 ATTGTGACATTCACTTCCCAGCAAACCAGAAGTTTATAGGTACCACAAAGATGGACCTAG
                                DNA/MER1_type# MER1B ->
    2161 AAATACTGTTCTGTTTGTTCCCTAAGGCGGGGGCCCGCAACCCCTGGTGGCCTGTTAGGA
                                            DC 2.2G/T rs4736154
    2221 ACTGAGCCGCACAGCAGGAGGTGGGTGGCAGCCTAGCGAGCTTTGCAGCCTGAGCTCTGC
                                                T 0.042
    2281 CTCCCGTCAGCTCAGCAGCGGCCTCAGGTTCTCACAGGAGTGCAAACTCTATTGTGACCT
    2341 GTGCACACCAAGGATCTAGGTTGCACGCTCCTTATGAGAATCTAATGCCTGATGATCTGA
    2401 GGTGGAAGAGTTTCAACCCAAAACCATCCCCTTGCCACCATGGTCCGCAGAAAAACTGTC
    2461 TTCCACAAAACTGGTCCTTGGTGCCAAAGAGGTTGGGGACCGCTGCCCAAAGGGAAGGAA
    2521 GGAGTGGATAACAGCAGGACAAGCTCCGCAGCAGGCAGGGCCCCTGGTGCTGCTGTGTGA
    2581 GGCCTTAGACAAATCATTCCACCTTCCTGGGACTCAGGTTCCTGTTCTGTCAACCAAATT
                          rs10110957
    2641 GATTAGCTGACATCTACCGTCTCGTCTCATTCTGTGATACTAGGGATCACTATTTCCCAT
                          A 0.000
    2701 ACTGAAACTGAAATGCCACAGTTAATTATCATGTCTAACAGAAACTGTCAGCAGGGAAAG
                                      AT_rich/Low_complexity ->
    2761 TTGACATTTCTCTTACAAAACCAAAAGCCAAAAAGTACACAGAGAAAGAAAACCAAATTA
    2821 AATAATTAATTAATTGAAAAGGAAACCAAAGCAGGTGATGCTTTCATCCCACCGTCAAGA
    2881 AACGGCAACCCAAAGGCAGTTTTCATGTGACACATCAATCCCAATAAAATCAAAACACTA
```


2941 ATTTATCAATCTTCCCGACAGCTCCCTGTGCTCAAACATGTCCCTCAGACAAAGAGGGCA
3001 CGTGCAGCAGCCTGGCTCGATCTTCTGGATAATTCCTTCATGAAAACCTTCTTGGATAAA
DC 3.0G/A rs7823441
3061 ATAATTTTTCAATCTCATCAGATGACAAAGCTGTCTGTTACGCAGAAGTATTATTACTAA
rs35034476 A 0.033
3121 GGTGTTTCCAATTCCTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTTTCGCTCTGA
/-
3181 CACCAGGCTGGAGTGCAGTGGCGCAATCTCGGCTCACTGCAATCTCTGCTTTCTGGGTTC
3241 AAGCGATTATCCTGCCTCAGCCTCCCGAGTACCTGGGATTAAGGTGCACGCCACCACACC
SINE/Alu# AluSc ->
3301 AAGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGATGGTCTC
rs4736156
3361 GATCTCCTGACCCCATGATTCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCAC
A
3421 GAGCCACCGCGCCTGGCCTCTTCTTTTAAAAAACACTTTATTGTGAAATACAACACTGAT
LINE/L1# L1ME3E ->
3481 TGACAAAACCACCAAAAATAAAGGTATCACTTAATGAATTATTATGAAGCAAACACCCTA
3541 GAAAACACGACACAGGTCCAGAAGCAGAATTTTACCAGCCCCCTAGAAGCCCTCGATTTG
3601 CCCATCCCAATGCTGACGTCCTCCCTCGCCCACAGGGCCCACCACCCAGACTTTGTGCTA
3661 GTCACTGCCTTGCTTCTCTTTGCCATTTTACCACGTAAGCATGCAGCAATGAAACTACAG
LINE/L1# L1MC5 -> rs34120094 DC 3.7 TA/-
3721 TTTAGCTGTGCCTGTTTTTAAATGCGTCATTTAACTCTCTCTTAAACTACAGATAGCCTT
(TG)n/ Simple_repeat ->/- LINE/L1# L1MB3 ->
rs35366571
3781 TTCATCTCTTTGTTTTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTCTGTAAAATACAC
/-
3841 ATAATTTAAAGTTTACCACTTTAACCATTTTTAAATGTACGGGGTCAGCAGCATTAAGCA
LINE/L1# L1MC5 ->
3901 CATTCACACTGTTGTACAATCATCACCATCATCTATCTCCAGAACTTCATGCCTTTTTCC
3961 CCCTGATAATTTACTTGTTGAGAGAACCAATTCATGGACTGCAGGGTTTCCTGGAGTCTG
4021 GACTGCCAACTGTGTCCCCGTGGTGTAGTTTAACAAGTTTTTCTGCCTTCTGAAGTCCTG
4081 TAAATTAGCAGTTAGAGCCAGAGCGCTGATCAGCCTCAGGTGTGATTCTGGGACACTGCT
4141 AATTCACAGAAGGTGCTGTGTTCTCCCAGCAGGAGGAACATAATGTGTGGTTGTCTCTTA
4201 TTTTGTTATGTGGATGCTTACTGTCTCGACTCATTAATCTATTAGGGATGGCAAAATGAT
4261 GATATCCGAATTCTATCCTCCTTATCTCCATTGTTAGCCCAAATACTTCTATAAAGAGAG
4321 ACTCCCCTCATCAACTGGTATTAGATCGTCCAATCGCGCAGTCTACATGAAAAGGCAGAA
4381 TAAATACTCATTGCTTACCATTTTCAAAATGATAAATCACTTCCCCATCATTTTCCAAGG
4441 GCGAACAATTCTCTCAAGTAGCATTGAGAACTCATGGACTGAAACACAGTTCGCATCTTT
4501 ACTGATGCTCAAATTGCCAGTGGTCACTTCTTCAGGCGGAGCCCTTCAACACAACCTCAA
DC 4.5 A/G rs6991134
4561 TGCTCTCTAATCACTCCTTTGCTGCAAGTTTGACAAGATGGTCCAAGATCACCTTCCTGC
G 0.033
4621 TCTGTTTCCTGCTCCACATCTAGAACTGGTTATTTTTCTCTGAGTGAGAAAGTGTGTTTC
4681 AAGACCAGGAGGTGGGCATTAGGAATGCTCACTGCTGCTGGACTGGTCCGTTTCTAGACT
4741 CGGACACTGTACAAGGCTAGGAGATTGATAGCTTTTCAAATAGTTCATGATTTCTTGCTA
4801 TTTTTTATTCAAATTCTAAACTACAAGGTTTTTACTTAACTTCATATCTGACATCAGTTA
4861 CTTCCTGTCTTCCATATGAGAACCCTAGTTCAAAACTAGATTCAGATGATAGAATGAGAT
4921 TATCACAGAATTAATCATTTACTTTACCCTACACTGATACCTAGTTTCAGAATATGAGTA
SINE/Alu# AluSx ->
4981 CTAACACTATCTCTACCAAAACAATTACAGAGAACAGTTTAAAAATTACTTTTGGGGCCG
rs9644445
5041 GGCATGGTGGCTCACACTTGTAATCGCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCA
A
5101 CCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACGTGGTGAAACCCCGTCTCTCCTAAA
5161 AATACAAAAATTACCCAGGTGTGGTGGCATACGCATGTAGTCCCAGCTACTTGGGAGGCT
5221 GAGGCAGGAGAATCACTTGAACCCGGGAAGAGGAGGTGGCAGTGAGCTGAGATCGTGCCA
5281 CTGCACTCTAGCCCGGGCGATAGAGCCAGACTTTGTCTCAAAAAAAAAAAAATTACTTTTG
5341 CATATGACTTCCCATGCTCCCCCCATGTTTTAAATTATTGTGATGGATACTGTAAGAGCA
5401 GATTACCATGACATACTATCTTCTCCCTTTGAATGCTCATTTGGTCTTAGCTATACAAGC
DC 5.4 C/T rs7014855
5461 AACGATATGTTTAATGTTCACCACAAGTCCTTATGACAACGTCTCTCTAGTTACTTTGGC
LINE/L1# L1MC4 -> T 0.019
5521 TGTGCTTCCCTAGTAGAATCCTCAAGAAATCATGAGAATTTAATTCTTCCAGTCTTGCGT
5581 GTTGACAACAATTTGTCTAAGCCCTGTATACCTAAAAGTTAGTGTGGCTGTGCCTACAAA
DC 5.6 T/C rs28514735
5641 TCTCTGGCTTAAATTTTCATTACATGAGTAGCTAGACATAGTCTAGATACATGCTATTCT
C
5701 ATTGGCTTCTGGAATACTGCTGCTGTCAAAAAGTCCAATGATAATCTGATTTTCTTTATT
LINE/L1# L1MC4 ->
5761 TCATAAATCACTTGCTATCTGGGTCTAAAAGCCCAAAAGACTTTAAGTCCAGTAATATAA


DC 5.8 T/- rs10709306
5821 CTAGAATATGTCCCGCTCATTCTAAGTCAAGATTCTCATGTATACAGTGTACCCATTCAT
/-
5881 GCAACAGTTTCAAGGGTTTTTTTTTTTAATTACAGAAAGTTTTCTGTGAATTATAGTTTT
5941 TAACATGTTGTGGTTCCATTGTTTCCGTTTCCTTCTCTGGTGTCTCATAGCAGATATATG
6001 TATGTATATATGTATGCATGCATGCATGTACGTTAGATCTTCTTTGCCTAGCTTCTAAGT
6061 GTGTCACTTTCTCTTTAATCATCTTTCTCTCGATTCTTTTTTCTTCTCTCTTTTCATCTT
6121 CTGCTTCCCTTAATGTATCTGGCAGGTATATATAACTCACTAGAAACTACAATGCTGGAA
DC 6.1C/T rs6981937
6181 AAAAAAATTTATTCTATTACAAATTCTTTCCTGAGTTATATCCATCTCATCTCTGAGTTT
T 0.033
6241 TTCTCATTCAGATTTATGTTATTCTTTTGTACTTCTTCTTCATGTCTTTTAGCTCATTAT
DC 6.3 T/A rs7005691
6301 GGGTTTTTGTAATTTTTATCAGCTTCTGAGCACGTTCTTCTATGTTATCTCGCTGCTTAT
A 0.034
6361 GTTTTCACTAGAATCACTCTGCATAGAGTTTGGCCATGATCCTTTTCTGTTGACCATCTT
6421 TACATGAAACTAGTTTTCATGAATGTTTAGGAGAAATGATTCAGGATAGCTCTTCTAATT
6481 CTAGTAAATCTACACACTGTATTCTGTGTTTGGGGGTGGTACTCAAGAATATAAGAAATT
6541 ACTTTCTAGACCACTTAACTCTGTTCTCCTCCCCTAGTTTCAGCCAGATCTTCTCTTTCC
6601 TTGCTTCTGCTATCCCACCTCTGTTCCCAGCAGCTTCTCCACAGGTGAGGCTTTGTTCTA
6661 GAGGAATTATTTGCACTCATCTGCTACTGGAACATACAAACTCCTCCACATTTCAGCTGC
DC 6.7 G/A rs4422756
6721 TGTTCTCAACTTGGCCCACTGAGCTGCCCAATGAGTACCTGCTGGCTGTGTGGGGATTCC
A 0.017
DC 6.7a G/A rs4422757
6781 TGGGTGTGTGTGTTTCCGGCTTCCTCTTGTGCAGACGGTGAAACCGTGTTGTTATTAGTG
A
DC 6.8 A/G rs4307318 rs4409383 DC 6.8a T/C
6841 GTTTGTCCACATCTACTTATATTTTGGGGTCCGTGGGGATACTCTGGTAACTGGTTTTGC
G C 0.021 DC 6.8 F
6901 TATATATTTTTGCCTATGGGTATTTGGTTTTTCTATTCTAGTTGTTCAGTCTGTTTGACT
6961 GAATAGTTCATATATTCAAAACCTAAGACAGTTTGCCACCATCTTCCAGGACTTCTAATT
7021 TTAAACATAAAATTCCAAGAAATCAGAGATTTCAAGTATTATAAGGATCATAGGAAAAGA
7081 ATATACTATTTTGTGGGTGAAAAGTTGACTTCTCTTACCCTAGGCTAGTTATATGAAGAA
DC 7.1 T/C rs7010692
7141 GATAGATATCTGCCAAGAATTTGAAGATAAGGTAAATAATTTGCTTTCTGATGTTTCATT
C 0.008
7201 TAGAAACAATTTCTAAAATTCACTTAGAAATCAAAATAATACATCCAAACAGTGCTCTAC
LINE/L1# L1MC4 -> DC 7.2 C/T rs6578086
7261 AAAATCTTTTTCTTGTTACTATTTTGAAAAATATCGAGCCTACAGAAAAGTCACAAGAAA
LINE/L1# L1PA16 ->rs33992585 T 0.425
7321 TGAAGATCTATATCTCTTTATCTAAAGCCAACTTTTTTTTGTAATTTCAACTTTTATTTT
/-
7381 AGATTTAGCGGGTACGTGTTCAGGTTTGTTACATGAGTATACTGTGTGATGCTGAGGTTT
DC 7.4 G/A rs4736023
7441 GGGAAATGAGTGATTCTGTCATCCGGACAGTGAGCGTGCTGCCCAACAACTGGTTTTTCC
DC 7.5 C/A rs12682161 DC 7.5b C/T A 0.150 I
rs6578087 rs11382806 DC 7.5a -/C rs4736024 DC 7.5c C/T C
7501 CTCCCTGACCCTCACTCCCTGCCTTCTCGAGTAGCCCCCGTGTCTGTTGCTGCCACCTT
IIIIIIIIIIII
CNCCNTNNCCNC
A 0.367/CT T 0.292
7561 TATGTCTGTGAGGACTCAATGCATACCTCCCACTTATATGTGAGAATATGCTGTATTTGC
DC 7.6 C/T rs4736155
7621 TTTTCTATTCCTGCATTCATTGGTTTAGGATAATGGCTTCCAGATGCATCCACGTTGCTG
T
7681 CAAAGGCCATGATTTCATTCTGTTCTATGGCTGCATAGTCTTCCAGGGTGTATATGTACC
DC 7.7 C/T rs10110769
7741 ACATTTTCTTTATCCAATTCACTACTGATGGCCACTTAGGTCGATTCCACATATTTGGTA
LINE/L1# L1MC4 -> T O.110
7801 TTGTAAATAGTTAAAGACACTTCTTTTTAATAATTTGACATTTTTACTTTATTTCTCTCT
7861 AGATTCACATTCACACTGTATTTTTTTTCCTTCACCCTATTAAAGTAAGTGCAGGCAAAT
rs35511380
rs34710276 rs10611493
rs11988522
SINE/Alu# AluY -> rs11988521
7921 CATCTTTCACTCTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCGCTCTGTCACC
DC 7.9 R C A /-
C
/-

```
7981 CAGGCTGGAGGGCAGTGGCGCGATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACG
8041 CCATTCTCCCACCTCAGCCTCCCAAGTAGCTGGGACTACAGGCACCCGCCACCACGCCCG
8101 GCTAATTTTTGTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTAGCCAGGATGG
8161 TCTCAATCTCCTGACCTCGTGATCCGCCCACCTCGGCCTCCCAAAGTGCTGGGATTACAG
  rs4736157   LINE/L1# L1MC4 ->
8221 GCGTGAGCCACCGTGCCTGGCCTGATCTTTCACTCTTAAATGCTTCTGCATTTATTTCCT
       A                                    DC8.2 C/T rs4236870
8281 AAGAACAAGAATTTTCTGCTACAATTTCAATTTCATTATCATACTCGATATATTTGTCAT
                                                      T 0.125
8341 TGATACAATAATATTATAGTATTCAATTTCTTCAATTGTCCCAAAACAAAGTCTTCTTTT
8401 TTAAATCCAGGATACAATTAAGGATCATGCACGGCATTTGATTGTTACAGCTCTTCAATC
8461 TCCTTTAATCTAGAACATTCTTCCCACTTGCGTTTATAAAATGGGAATTACAACATTTAC
8521 TTTATAGAGTTTTTACAAAGATTAAATAGAAATAATGCCAGAAAAAAAGTGGCTAGCCCA
8581 GTCTTTATACACAACATGTGACAATGCTCAGTAAACATTAATTATCTTCCAGCTATCTAT
8641 GGCCGACACTGAATCCAGGATCCCGGTTTGTGCAACAGCTATCCAGGTTGCTGTGAATTT
  DC8.7 G/A rs6992126
8701 TTTAAGCGTCATTTAACTTCAGTAACCATAATTAATGCGAGTCCATGCCTGAGAATCAAA
           A 0.117
8761 TTGGATCCACACGTGGTCCCCCATTAGATGTCTCTGAGAATGTTAGACTGTGAGCTACAG
8821 GGGGTCGTCTCCCAGTTCCCAGAGGCCACAAATGAGCCTCTGCCACTGAGTTCACAGCCA
8881 CTTCACATAAGGGCGCCGTGATTGTCTCGTGCTTCCTCAGACTGAACTGCACGTAACAAG
8941 GCAGCCTCCATGGACCAAGCTCTGGGTGGTTCTTCGTGCCTACCTTTTCCTCCATAAAAC
9001 TTGACCTGGTTTCCTGGTTTTTCTAAGGTAAAACTTACCTTAGAGATTTGCTGTCTCTTT
9061 CTGCCACATACTATGTAAACAACCAGCCAGAGCTTCTATTAACTGTTATTAATAAAGGGG
9121 GCATCTCATGAGTAACCACTACGTGCCATGACCAGCAATCAGGCCAGCCTCACATTTACA
                DC9.1 A/G rs7017431
9181 AACCACCTGGTGACCATAATTCAAGGCCTGATTTGGTTTTGTTTTAAAAGATAGAAGAAG
                           G 0.642
9241 GTATTCCCAGGCTACAAGTCCCAAATTGCAATTTTGCTCAATTTTTCTCTCTGACCTTAG
9301 AAGGGCAGAGAGACTGCCCTAGTCCCCTCTATAACAAAAGGCAGAAGCACAGTCATAACA
9361 GGACAAAATAGGTCTGCACTGAACATGCAAATTCAAACAACCAAGCACAAAAAGCCAAGC
9421 CTTCCACATGGAAGCCTTCCACACTGTTCCTCAGGAATAAGACGACCCTGGCAGGCCATG
      rs34882077
9481 GGAGAAACCACGGTCACCATTCCCTCACTGCTAGCCTGGAGTCCACGTGCCAAGACATAG
      /-
9541 CCCTGGGGTAATATCCAGTCCCAAGTGTTCCAGGGGCAGGTGCTGGGCACCATTTATTCT
                DC9.6 G/A  rs4736158
9601 GCCAAACCCCCTTCTTGGGAAGAAGAAGTCAGAACCTATGTAACCCTCTGTCTACTGATT
                         A 0.125
9661 AAAAAGAAAAAATGCAGTTCAGCGGGATAAAGCCGTTTACACACCAACCAAAAAAATCCG
9721 TGTCATCAGACATTTTTAAGTCTTGGTATAATTTGAAATCCTCTCAACTTTTTTCATTTT
9781 CTAATGTATCCAACCCTAAATCTGTTTTTCTACATTTGTTAATAATATAAAAAGAATATT
9841 TTACCAAATACCAGCACTATACAAGAAGCTCATGCTGACTATTTGAACTGCAGAGGGTGC
             LTR/MaLR# MLT1A ->  DC9.9 C/G rs4736025
9901 AGACAGCAAACTCTATGGAGGGCGTCTGTTAGGGACTGAGCTGCGTCCCCCTTCCCCAGC
                                          G 0.125
9961 TGCATCTACTGAAGGCCTATCCCCCAGTGGGACAGCTTTAGGAGGTGGGGCCTTTGGGAG
10021 GTAATTAGGGTTAGATGAAGCAATGGAGATGGAGCCCCCTCTATGAGATGAGTGCCCTTG
10081 TAAGAAGGGGTGGACACCAGAGCTCCCTCTCTCCTCCACGTGGAGACGCAGCACGAAGGT
10141 GGCCATCTGCAAGCCAGGAAGAGGGCCCTCCCTCACCAGAAAATCACCACGGCTGACAGT
                                                 rs35260714
10201 CTGATCTCAGACTCCCAGTCTCCAGAGCTATGAGGAAATAAATCTTGGTTGTTTAAGCTG
                           rs11782549             /T
10261 TTCTGTCTGTGGTATTTTGTTAGGGGAGCCTGAGTTGACTGAAAGACTCCACACTGCCTC
                           A     SINE/Alu# AluSx ->
10321 CAAATGAATGCTTCATAATTTGGCATAATTAATTTATGAAATCTTTTTTTTTTTTTTGAG
10381 ATGGAGTCTCACTGTGTTGCCCAGGCTGGAGTGCAGTGGCGTGATCTCGACTCACTGCAA
10441 CCTCCGCCTCCCAGGTTCAAGCGATTCTCCTGCCCTCAGCCTCCTGAGTAGCTGGGATTA
10501 CAGGCATGCGCCACCATGCCCGACGAATTTTTGTATTGTTAATAGAGACAGGATTTCACC
10561 ATGTTGTCCAGGCTGGTCTGAAAGTCCTGGCCTCAGGTGATCCGCCCGCCCTGGCCTCCC
10621 AGAGTGCTGAGATTATAGGTGTGAGCCATCACACCCAGCCAATCTATGAAATCTTATCTC
10681 TAATCAGTCCTAATGTTACAACAAATATAAATTTCCATTTCTTCCACACTCCACCTCCTT
                                  rs35936231
10741 ATAGATATCATCAGACTTTGAAGGAAAACTTTATCATCACAGCTAATTAGATGTCTAAAG
                                  /C
10801 AATCACAGCAGTTATACTTGAAAAAACGCAGTAATTCCTACATTCCACTAATTTAGAATG
10861 ACCTGCTGGTAAATGAGTACAAAATAATAAAGAGTGAGAATTAAATCATTTAGATAAAAA
10921 GTACTTTTATTACTTTCAGACACATACAGTAACACACTCTGTTAACAAAGTGCCCAGGAG
10981 ATGTCTCAGGCAGCTTGAGTTAATGAATGACAAGTGAGTTACAAACAGCAAACAGCAGGC
                      rs11986788
```

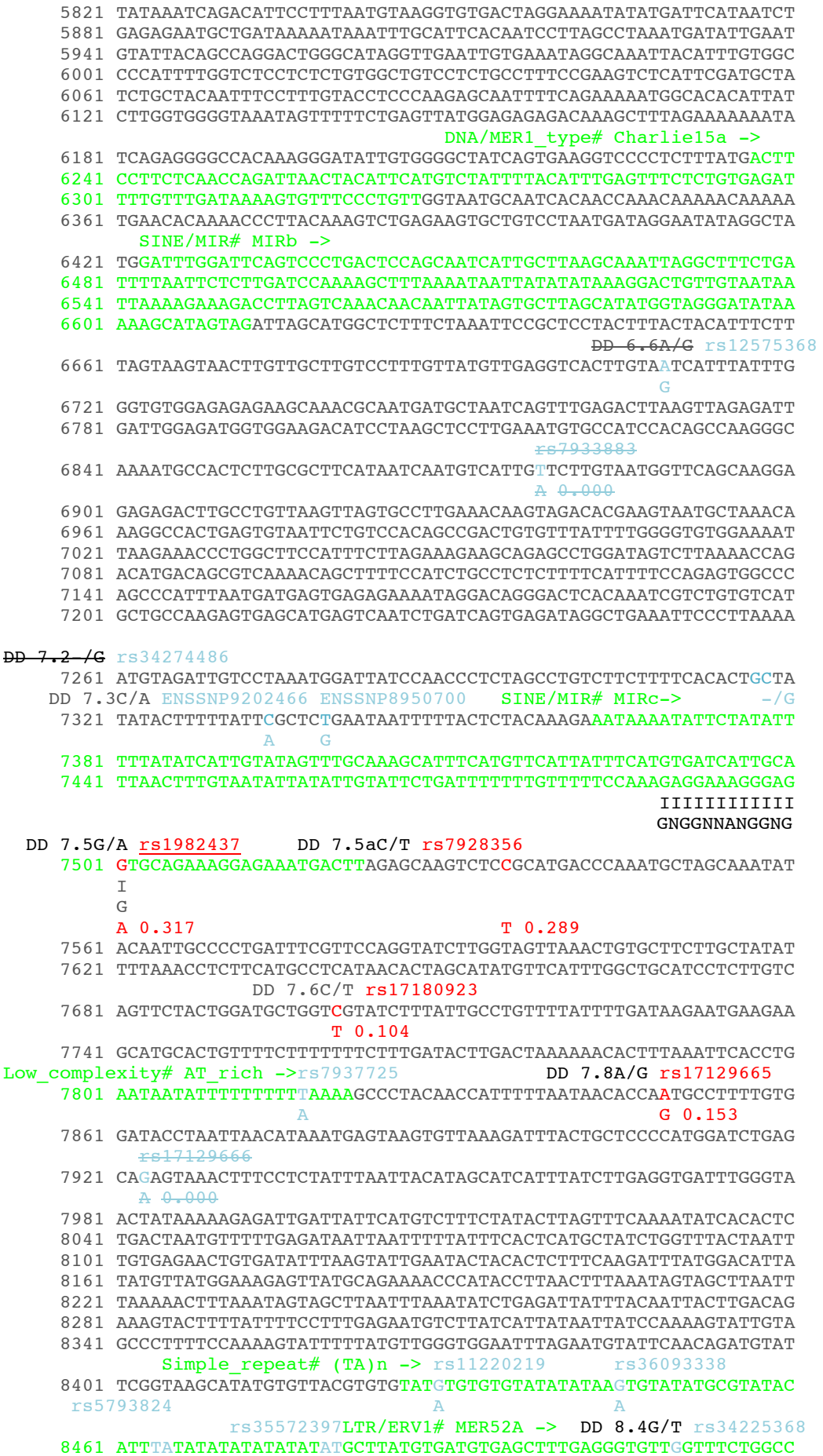

11041 ATCAAGGGATGGGGGTTGCTACACAGTATCTAGGACCCTGACGTGCTAATTAAAATTAGG
A 0.000
DC11.1 C/T rs11166964
11101 AAAGCGCTTACCATCACTGAGATCCTGCAGGAGATTCTCCTGGCAGGAGAAATCCAGCTT
T 0.085
11161 CACTTTGATGTTGATTGTGAACGTGGCCACCTTCCCAGGCTGCAAAGGGAACTGGGCAAG
DC11.2 G/A rs11166965
11221 GGTTTCCTCTAGCTTCCAGCTCAAGAAGTCGCCATACAATTTTTCTGTAATAATAACAAT
A 0.096
11281 GACAGTGATGAGGATGCTGTAACTGAGGCAGTATGGGACTTACCAATCCCCTAGAAAATT
11341 CTGATGCATTCTCAAAAGCCAAGGAAAAAGAAGAGCTTTTAAAATGTGTAAGATCAAAGT
11401 GACAAGAACATTCCAGAACCAAGAAGTGTTGCAAGGCAATCACCCCTGTTTATGAAGGAG
11461 GCACAGGGAGGAGGAAGAACTGCTCATCTCTAGAGGCACCGAGCTCAAGCCAAGCCATCT
rs11166966 rs12334515
11521 TGACGCCTTAGTGAGCAACTCCTTCTCATTGACCAAATAAATGCCGTTTTGCTCTTATAA
T A
11581 ACTAAATCAGAGTTAAACTAAATTTCTCCACAATGAATATTTTCTGTATTATATCACCTT
11641 AAAGAGAAACAGATCCAAGTGTAGTAGGGAAAGTGTATATTTAATATGAGACACCTCTTC
SINE/Alu# AluSx ->rs7821968
11701 ACAAAAAGAGAATAAAAAATCAAAACCCAAGCAGGACGCTGTGGCTCACACTTATAATCC
G
11761 CAGCACTTTGAAAGGCTGAGGCAGGGTGAATCACTTGAGATCAGGAGTTCCAGTCCAGCC
11821 TGGACAACATGGTGAAACCCCATCTCTACAAAAAAATACAAAATTTAGCCAGTGTGGCAG
11881 CGCGCCTCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGC
rs4736159 rs12543314
11941 AGGCAGAGGTTGCAATGAGCCAAGATCATGCCACTGCACTCCAGCCTGGGCAACAGAGCA
rs7841395 T T
12001 AGACTCCATCTCAAAAGAAAAAAAAAACATCAAATCAAAGAATTAAGGACGGCACCTTGG
A
12061 GAAGGGCCTCCGGGGAGAGGTGTTTCCCTTGCTTTCTGGCAAATACACCATGGTCTTTAA
SINE/MIR# MIRc -> DC12.1 G/C rs7841647
12121 TAAAAAGTCTTGAATGTTAGAGTCACATGGCCTTGACTTTGGTGCCTTTTTGACACTGAC
C
12181 TAACTCGAGTGCCATGGAGAATAATGATCCCTTGGAGTCAAGGTTTCTTCGTGTGTAAAA
12241 TGGGGATAACAATTACTGCACCACAGGGTTATTTTACAGATTCAATAAAGTACTATCTGT
12301 AAAAGCACCAAGAAGACCACAGCACTGTAAGAACCTCAGGCAACACGGGAGGAAGGGGGA
12361 CAGGAAGGGGACATGGGGAGGCCACTGCTGGTGCCCGCTGAAGTAGCTGCGTCTCTCCAC
LTR/ERVL# LTR53 ->
12421 GGCTCCAGTTCCTAAAACCCAGGCACAGTATGATATGCTGCCTTGACATCCGGCAAAACT
12481 GGGAGGGCCCCAAACGGCCTCACCACAAGTCCCCTCCTGACTCTGCTCCTGCAGATACAG
12541 GGTCCCTCAGCCAGGCAGGTTTCCAGGTGCGGAAAGCAGGTGCAGCCCTGTGTATCCTGG
12601 GCAGTGGGTGTCAGTTCCCTGCCAGCCTGGAAGGGACTGAAATAAGCCAGTCACATCCTC
rs6996327rs7013680
12661 CCTGGGGAGCCGGGGGCAGCCCTCCCTCTCCATACCCCAAAGCCCGCCTCCCCCAGCCCC
GC
12721 TGGGCAGCTGCTCTCTCTGTTCCTGAGTGCAGCCCACATGGCCCTACGTGGCATGGTATC
rs35027015
12781 CTCCTTCCCCAGCTGTGAGTGTCTGTGACCAATGAACTGCTGTCAGGTGCACATGCCCTG
/C
12841 TGCACATGCTAGGTCTCTGGCTGTCCCCACATGGTAGAACAAGAAACCCTCCTGACCAAC
12901 ACAGGGGACAGGAAGCCATGAGCACGGGGCCCTCCTCCTCCCAGGAAGGGCCCAGCTTCC
12961 AGGCTCTGCAAAGGTCCTCCTCCTGGTGCCCCTCCAGCCTAGGAGGCAGGAGCGACTTGC
rs36112566
13021 TGTTTTCCCACTATCCCCTTTCTGGGAACCAGTTTTTTATGCTGAACGTCTTCGGTCTGA
T
DC13.0 A/G rs4736160
13081 ATAAGGAGACGGTTTTAGAGAGATGAAGTAACTTATCCATGGTCACCAACTATGGAATCA
rs4736161 G 0.121
13141 GTCTAGAATCCTAGATCCAAATGATTCCAAAGTCACATTCACTTCCTGCATGAAGACATT
G
13201 TCATGGTGAAGTTCAAATGTAAAGTCCTGGTCGGAAGACCCAGAAGTGAACCTCACAGGA
13261 GCTCAGGGGAAGACGGAGAGCTTGACCAAAATCCCCATGAAAGAAAAAAGATACGTGGTG
DC13.3 C/T rs4736026 rs7005324
13321 TGCTAACTCCAGAACACGAGGACAAACTGAGCTGCAAAATAAGCAGACATATTCCTGGCC
T 0.125rs6980517 C
13381 GGAGGTGTGCTCGACCATGGCACGGTCCTGCAGAACTGGCCAAGCTCCCTCTGGAGCACG
DC 13.4 T/C rs7005477 A 0.000
13441 TCTGACAGACGGCGTTTATGTCTCCATGTTTCCTTAAAGAAGGATACTAACACAGAGAGA
C 0.125

```
13501 CAGATCCAGGGGACAGCGTCCTTGAAAAGTGTTTAGAAACCACATCACATTCATTTATTT
13561 ACTTAGTCATTCACAGAACACCTCAAATGTGCCAGACACTCCACGTGTTACCTCACTTAA
13621 TCCTCCCAACAACCAGTGAAATCATGGAAGGAGGTTAGCATATGTAACAGGTCATTCAGC
13681 TGTCAGTGGAAGACTTGGCATATTATACATGTTTGATAAATATTAGCTGCCCTTCAGCAT
13741 CACCCAGTCCACCACCTTGGATGGTGTTCAGATGCTTTGCAAACTGATACTATCACAGTA
13801 TTTTAGAATGATCTGGGAAAAGACAGATGGCTGGGAAAAGGATGATGGAATCGTTGAGCA
13861 GGCTGGTTCCTGGAAAGTCAACTTTAACGACAGTCTGAGGAATGAATCCCTACATTCAGA
rs4078541 DC13.9A/G
            DC13.9a C/T rs4475456    SINE/Alu# AluSx ->
13921 TGATTTCGAGAAGCTGTGGACCGGGCGCAGTGACTCACACCTGTAATCCCAGCACTTTGG
  0.125 G   T 0.062
13981 GAGGCCAAGGCGGGCAGATCACCTGAGGCCAGGCGTTCGAGGCCAGCCTGGGCAACAGGG
14041 TGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCGCACACCTGTA
14101 GTCCTGGCTATTCAGGAGCCCAAGGCATGAGAGTCACTTGACCCAGGGAGGTGGAGGTTG
14161 CAGTGAGTCAAGGTCGCGCCACTACACTCCAGCCTGGGCAACAAAGCGAGACTCCGTCCC
14221 CAAAACAGAAGAAGCTGTGGCTGTTGGATGAGATTCTTATAGGAATTTAGTGAAAGGGAA
14281 ATGGAGAATTGGAGGAAAGTATTATCCAAAACTCAAAGTGGGAGATGGAATGGGTCAACA
14341 AGAATCACCTAAAGTCAAAGAAAGTGAATCATATCCAATTATAGCCACTCTAATTACAGG
                                                  DC14.4 A/G rs11778176
14401 ATAAAAATCTAAACACTTCAGTCAAGCTTTGGGGGCAAAGGGTACAAAACACTGGTCGCT
                                                           G 0.125
14461 TTGGTATTTCATGATGACAACAGGCTACTGAATTCTCTGTCTGGAAACCCGCTTCAACAC
14521 ACAACATTTGGCCCTTGGCTGCTCAGGGGAATGTGTGGATACTCACATATATGATTTATC
14581 CTTTAGACGGTAAGTCACTCAGTCAGAGAAAGAACAAAGATCTGTCTCCAAATGGGAAAA
14641 GAACTTCAAAGCCATCTTACAGGTGAGAAAGCAGAAGGAAAAGGAGAGTCTGACACAAAA
14701 TCACATCATCAATTGGGTCGCAGCTTTCTTCCCACTCCTTCAGAAGGGTTCTGACGACAT
14761 CTTGGCTCATTAGTAAGGAAGGTAACTGACTTTCTTCCTTCCCACCTGCTCTCTCAAGCC
                                                        rs34455732
14821 ATGCCAGTTCTCTGCCTACTCGATGTGACCTCCTCCACGCAGGACCCCCGGGCAGGCCCA
                                                        /-
14881 CCTGCTTCCGCTTGGTTGCCATTCTATGTGTATTCTGAGAATCTACCCTCCCACAAGACG
                                                     rs11778339
14941 CTGGAGTTCCTTTCCACTGTGCTCTTCTGTAGGTGCCGACACATCTGTGGGCAGCTGCAG
                                                     G 0.112
15001 A
```

TARGET DD


LTR/ERVL# MER76-int ->
LTR/ERVL# MER76 ->
ENSSNP10484456
rs11219891
LTR/ERVL# ERVL-E-int ->
rs11219899
rs7925506
LTR/ERVL# ERVL-E-int ->
DD 2.2T/C rs11219948
C 0.242
DD 2.3G/A rs35721278
Simple_repeat# +(TTA)n->rs34703815
LTR/ERVL# LTR16A ->
DD 2.7G/A rs35272776


rs35547878
2941 GCCAATGACTAGTTGATGTAAAACTACAAAAAGCTAAGCCCACTTGCCCTACTGGAGAAT
T
3001 ACCTTTAAAGGATACCTCAGTTCCTGATCCCATTGTTGGGTTTGGCTGGAGCCTTTACTG
3061 CGAATTCATCTTTTCCTTTCACTCACTCATAGATATTAATAACTAGAGTATCTCCCAACA
3121 AGTTTCTTACACACAAATCTTACCTCAGTATCTTATCTCAGTATTTTCAATAGCAGCCTA
DD 3.1A/G ENSSNP10484457
3181 ACAATATTTCAAAAATTCTTTAACGGCAAATTTTATGCTTAGGAATCCATGTTATAGAAA
G
3241 TACTTGAACATGTTTCCAATAGTTGGTTGGTCCAAGACATTGGTCATCCTTAGTCATAGG
3301 GTTGAGACTCTCCAAAGTAGTGTCTATTCTTCAAACATCAATAAAACACAGTGGAAATGC
3361 TCATCTCTAATCTGCCAAACCAATGCATTATGACGAGAATTCTCTGCTTTTCATCTTAAA
LINE/L1# L1ME4a ->
rs35786280
3421 CTTATTAATAACTAGTCCAATGTTTATGGCTAAACATGTTTATTTTAGCACTGATTTTAA
-/T
3481 TTGGGTAAAATTTGGAATATTTAAATGTCCATCATTTGGGGAGGACTTAAATTATGGTAT
3541 ATTCATAAAACTGAATGCTACTTAGCCATTTAAAAATGAGATCTTAAAACACTAATTTGA
3601 AAAAAATATCTAAGGTCCATTATTAAACGAATAAAGGAAAATATAGAAAATATGTTTAAT
3661 AGAATCCATTTGTTTAACCTCTCTGATTTTGTAGGTTTATTTTTATTTATACTACACAGG
3721 CAGATGCAATTTGAGAGTATCAAGCATATAAATAACATCTGTTAAGTTTTCAGAGATTGT
3781 TTAAAAAGTCATGTATAAAGCAACCAGAAATCAATCATTAGACTTTTTATTAGCCACACT
3841 GTGCATTAGACTACTATGGAACCATGTACTCAAAATCTTGAGAAAAAATTGTATTAACAC
3901 ATAATTTTATACCTGGACAGTTTTTAATGACTGAAGTTTTTGCTATATTACAATAAATTA
Simple_repeat# (TA)n ->
3961 TGCCTGCAATATTACAGAATATATATGTAAGAGAATATATATATTATGTATATAGAATAT
rs36013891 rs10594066
4021 ATATATATTATGAGTGTATATATATATAATGTGTGTGTGTATATATATATATATATATAT
G -/TATATATA
4081 ATATAAAGAAAGAAAAAAACCTGTGATATATATAGAATGAAAAAAACCTGTAGATTATTA
LINE/L1# L1ME4a -> DD 4.0T/C rs34428483
4141 TATATATACCTAGACGTTATTGTGCTGATTAAATAGCAGCTATATTGGAGTCGTAGGACT
C
4201 ATGGGTGATTATGTTTTCTTCTCAGTATTTTTTTCGTCTGATATTTTCCAATCCTCTACA
ENSSNP10484460 DNA/MER1_type# MER5A->
4261 AAGCATATTCAGAAAAAAAAACACACATAGAAAACACTTAAGGTATGTTTCTTCATGGT
G
4321 CACCAGACTACTTGTTTCTGAATCAGTTAAGAGAAATAGACTTGTTGAAACTCTATACCA
ENSSNP10484461
4381 GGCCCATTGAGTTAGAAACTTTAGATGTTGGGCTTGGGATCCTGCACTTTAAATTAGCGC
T
4441 CCTTTGGTGGTACTTATATACTTTAAAGTTTAAGATGAAATGCAGAGAATTCTCATTACA
4501 ATGCATTGGTTTGGCAGACTAGAGGTGAGTATTTCCATTGATGTTTGAAGAATAGACACT
4561 ACTTTGCAGAGTCTCAAACCTACGACCAAGGAAGACCAATGTCATCTTTCAAATTGTTCC
4621 TGGGACCCCAACATTAAATACGCTTTTCTTTTTAATTGCATTTCTATACAGTGTATGTTC
4681 TGAACCCAGCTACAGAAAAAGAAACTTGACTTAGTCCACATTTCATAGAAGATTTCTTTT
4741 TGTAGCTGTTATCTACTAATTATTTGTCTCTTTTTTTTTTGTTTTGGTCTACTCAATAAGC
4801 CATTTCCACTTGAAGATTTAGTGCAGTGAACAATTTTTTCATAAAATGGTTTGTACTTAG
4861 GAAAATAAAGAACAAAGGAGTAAACCCTACTAAAGAGTCTTCTTGCTAGTTCAGCTCGTG
4921 ACACAAAGAATAGTTGAAAAATTAAACACACCATAGAGTTAGTATTCAGAAATATCACAT
4981 CTTAACATTTCTGAGTATTAAATGAAAGTAATGTTGGTTAAAGCAAAGGCACGTAATAAA
5041 TGTTTGTTGATTATCAAACTGAATGACTTTGCCAAATAAAATGACAAGTGTATTTGCATA
5101 TAACTATTTTATAACAAGGGTAAATGATTTCACAGATCCTTTTAATGAATAGTATTTCAT
5161 TCTTAACAGTTCTTCAACTGTCAGCAAAAGATTGCTCTCAACATTTTTTTCAGTCTGTTT
5221 TTTTTTAAAGAAACTTTTCTCACTCAATTTTGTCAACTCCTATGTCATATTTGCTTAGCC
5281 TGTATTTTTGGATATGAATTAAAACCACCAATAATTTAAAAGTCCTTCTGCCCTTGGCTA
rs1452936
5341 GCAAGAAACAAAAAATGGCAAATGAGAAGGTGAAGCTACAAGTATGTGCACTTTGGTTCC
T 0.000
5401 CAAGTTACTCTGGTCCATTGCAGCTGCTGTGTTCCATTGTTAGGCTGTGCCCACCTGGGA
rs7947133
5461 CAGAAGACAGAAATTAAAATAAGTTCAATTTTAAAAAAATGCATATCTTAGGAAAATGTT
G 0.000
5521 TAATATTATTTGCTCTTAAGAGAAGTTTTATTTATGAGCTATTAATTTCCCATTGTGAAT
5581 AAGTTGTAATATAAATTCAGGAACCCAACTTTTCAGTCAGTTAAAACTTTAAAACCTAAT
5641 GAGTTGAAACAATTGATTCTGCTCTTTCTCCTAGAGGGGGTGGAAAAAAAAATCAGGATC
DD 5.7T/A rs1869267
5701 TTACGAAGAAAATTTATTTTGTTTAATTAAAGGCACATTAATTAAGATGCTGTTTAATTC
DD 5.7aG/A rs1869268
A
5761 AATAATGGTGCCTTTAAAATGCATTTTTCATCTTTATTCTTTAAGGACAAATAAAGCAGT
A 0.217


5821 TATAAATCAGACATTCCTTTAATGTAAGGTGTGACTAGGAAAATATATGATTCATAATCT
5881 GAGAGAATGCTGATAAAAATAAATTTGCATTCACAATCCTTAGCCTAAATGATATTGAAT
5941 GTATTACAGCCAGGACTGGGCATAGGTTGAATTGTGAAATAGGCAAATTACATTTGTGGC
6001 CCCATTTTGGTCTCCTCTCTGTGGCTGTCCTCTGCCTTTCCGAAGTCTCATTCGATGCTA
6061 TCTGCTACAATTTCCTTTGTACCTCCCAAGAGCAATTTTCAGAAAAATGGCACACATTAT
6121 CTTGGTGGGGTAAATAGTTTTTCTGAGTTATGGAGAGAGACAAAGCTTTAGAAAAAAATA
DNA/MER1_type# Charlie15a ->
6181 TCAGAGGGGCCACAAAGGGATATTGTGGGGCTATCAGTGAAGGTCCCCTCTTTATGACTT
6241 CCTTCTCAACCAGATTAACTACATTCATGTCTATTTTACATTTGAGTTTCTCTGTGAGAT
6301 TTTGTTTGATAAAAGTGTTTCCCTGTTGGTAATGCAATCACAACCAAACAAAAACAAAAA
6361 TGAACACAAAACCCTTACAAAGTCTGAGAAGTGCTGTCCTAATGATAGGAATATAGGCTA
SINE/MIR# MIRb ->
6421 TGGATTTGGATTCAGTCCCTGACTCCAGCAATCATTGCTTAAGCAAATTAGGCTTTCTGA
6481 TTTTAATTCTCTTGATCCAAAAGCTTTAAAATAATTATATATAAAGGACTGTTGTAATAA
6541 TTAAAAGAAAGACCTTAGTCAAACAACAATTATAGTGCTTAGCATATGGTAGGGATATAA
6601 AAAGCATAGTAGATTAGCATGGCTCTTTCTAAATTCCGCTCCTACTTTACTACATTTCTT
DD 6.6A/G rs12575368
6661 TAGTAAGTAACTTGTTGCTTGTCCTTTGTTATGTTGAGGTCACTTGTAATCATTTATTTG
G
6721 GGTGTGGAGAGAGAAGCAAACGCAATGATGCTAATCAGTTTGAGACTTAAGTTAGAGATT
6781 GATTGGAGATGGTGGAAGACATCCTAAGCTCCTTGAAATGTGCCATCCACAGCCAAGGGC
rs7933883
6841 AAAATGCCACTCTTGCGCTTCATAATCAATGTCATTGTTCTTGTAATGGTTCAGCAAGGA
A 0.000
6901 GAGAGACTTGCCTGTTAAGTTAGTGCCTTGAAACAAGTAGACACGAAGTAATGCTAAACA
6961 AAGGCCACTGAGTGTAATTCTGTCCACAGCCGACTGTGTTTATTTTGGGGTGTGGAAAAT
7021 TAAGAAACCCTGGCTTCCATTTCTTAGAAAGAAGCAGAGCCTGGATAGTCTTAAAACCAG
7081 ACATGACAGCGTCAAAACAGCTTTTCCATCTGCCTCTCTTTTCATTTTCCAGAGTGGCCC
7141 AGCCCATTTAATGATGAGTGAGAGAAAATAGGACAGGGACTCACAAATCGTCTGTGTCAT
7201 GCTGCCAAGAGTGAGCATGAGTCAATCTGATCAGTGAGATAGGCTGAAATTCCCTTAAAA
DD 7.2 /G rs34274486
7261 ATGTAGATTGTCCTAAATGGATTATCCAACCCTCTAGCCTGTCTTCTTTTCACACTGCTA
DD 7.3C/A ENSSNP9202466 ENSSNP8950700 SINE/MIR# MIRc-> -/G
7321 TATACTTTTTATTCGCTCTGAATAATTTTTACTCTACAAAGAAATAAAATATTCTATATT
A G
7381 TTTATATCATTGTATAGTTTGCAAAGCATTTCATGTTCATTATTTCATGTGATCATTGCA
7441 TTAACTTTGTAATATTATATTGTATTCTGATTTTTTTGTTTTTCCAAAGAGGAAAGGGAG
IIIIIIIIIIII
GNGGNNANGGNG
DD 7.5G/A rs1982437 DD 7.5aC/T rs7928356
7501 GTGCAGAAAGGAGAAATGACTTAGAGCAAGTCTCCGCATGACCCAAATGCTAGCAAATAT
I
G
A 0.317 T 0.289
7561 ACAATTGCCCCTGATTTCGTTCCAGGTATCTTGGTAGTTAAACTGTGCTTCTTGCTATAT
7621 TTTAAACCTCTTCATGCCTCATAACACTAGCATATGTTCATTTGGCTGCATCCTCTTGTC
DD 7.6C/T rs17180923
7681 AGTTCTACTGGATGCTGGTCGTATCTTTATTGCCTGTTTTATTTTGATAAGAATGAAGAA
T 0.104
7741 GCATGCACTGTTTTCTTTTTTTCTTTGATACTTGACTAAAAAACACTTTAAATTCACCTG
Low_complexity# AT_rich ->rs7937725 DD 7.8A/G rs17129665
7801 AATAATATTTTTTTTTTAAAAGCCCTACAACCATTTTTAATAACACCAATGCCTTTTGTG
A G 0.153
7861 GATACCTAATTAACATAAATGAGTAAGTGTTAAAGATTTACTGCTCCCCATGGATCTGAG
rs17129666
7921 CAGAGTAAACTTTCCTCTATTTAATTACATAGCATCATTTATCTTGAGGTGATTTGGGTA
A 0.000
7981 ACTATAAAAAGAGATTGATTATTCATGTCTTTCTATACTTAGTTTCAAAATATCACACTC
8041 TGACTAATGTTTTTGAGATAATTAATTTTTATTTCACTCATGCTATCTGGTTTACTAATT
8101 TGTGAGAACTGTGATATTTAAGTATTGAATACTACACTCTTTCAAGATTTATGGACATTA
8161 TATGTTATGGAAAGAGTTATGCAGAAAACCCATACCTTAACTTTAAATAGTAGCTTAATT
8221 TAAAAACTTTAAATAGTAGCTTAATTTAAATATCTGAGATTATTTACAATTACTTGACAG
8281 AAAGTACTTTTATTTTCCTTTGAGAATGTCTTATCATTATAATTATCCAAAAGTATTGTA
8341 GCCCTTTTCCAAAAGTATTTTTATGTTGGGTGGAATTTAGAATGTATTCAACAGATGTAT
Simple_repeat# (TA)n -> rs11220219 rs36093338
8401 TCGGTAAGCATATGTGTTACGTGTGTATGTGTGTGTATATATAAGTGTATATGCGTATAC
rs5793824 A A
rs35572397LTR/ERV1# MER52A -> DD 8.4G/T rs34225368
8461 ATTTATATATATATATATATGCTTATGTGATGTGAGCTTTGAGGGTGTTGGTTTCTGGCC


-/TA
-/AT
T
DD 8.5A/G rs35030714DD 8.5aG/Ars35560250
8521 AGAAACCTCTGTGGTCAACAGTGCCTTTGCCTGACTTTTGCTGGGACCCACTGGGCTCTT
DD 8.5bA/G rs35539924
G
A
8581 TCTGCCTGCTTTGCCAGGCAGACTGCACTTAGCTCATGCTATTGGCCTGGGTCCGTGCTT
rs34124130
G
DD 8.6C/G rs34376453
8641 GCCAAGGGCAAATAGAGTGGTAAGGGGTGTGTGAGTGAGCAAGCATGGGGTCTGGCCACT
G
G
8701 GTGCACAGCCAGGCACACTGGCTGCAGTGGGGCAGGCAGCTCCAGGCACTGGTACAGGTG
8761 CCGGCTCCCTGTGAGGCTATAGCTGCACAAGGTGTTACAACGAGCAGCTTCCATGGCTGG
8821 CACTGGGAAATGCAGTGACACCTGGAAGCTTGGAGACACTAGGAACCACAGAGCCCCAAA
8881 GAGTGTGTCACAGCTCTGGCTTGGGGAGCTCCTAGGTCTGGGTTCCCCAAAAGGCCACAG
rs35377089
8941 CTCTTCTTTCCTTCTCTCTTTTCTTGTCATCCGCAACATGGTGAGCAAGGGTCATGTTTG
rs35860137 T
9001 AGCCCTGTTTGTGTTACAGCTCTTTCAGCGCCACCATTCAGCAGGTCCTGAGTTCTTTAC
A
9061 CAGTGTCCAGGAAAAATGAGGTATGTGGACAAGTGGAGAGTGAGCAAGGTGAAGAGAAGC
9121 TTTAGTGAGTGACATAACAGCTCAGAGGAGACCCACAGTGGGTCACTACTCTCTGCAGGC
9181 AGGTTGTCCCAATGAGTGTTCAGCTCTCAGCAGAGAGGAGACCCACAGTGGGTAGCTCCT
rs34680366
9241 TTCTTTGGGTAGCTCTTCTCCACTACTGGTAGTCCTGATATCTGCTCAGCTCTCAGCAGA
G
9301 GGGGAAACCTTGGCATAAGTAGCTCCTGTTCACAGGCAGGTCATCCTGTCATCTCTCTGA
9361 GTCTGGCTGAGTCTGGGGGTTTTATGGGCTTCAGAGGGGAGGAAGTGCATGCTGATTGGT
rs35445893 rs2084911
9421 TCCTGGGTGGCCATGGGCAGCCCTGGAAAAGGCATCATAAGTTCTCATTCTGGTCCATGG
C C
9481 AACTGACAGCCCAGCCCCCAGGCTTCAGGCCACCCCTGGCTTGAAGGTGGGGCTTCACCA
9541 GGGACCCACCCCTTTCTATCCTGGAGCCTGTCTGCCTCCTGCCCTCATCCATGGCACTGA
9601 GGCTGTTCTTGCCAAGGGGTGCCTGCAGGTCATCTGTGAGCTTCTCCGAGCATGCCCCGG
9661 GCCTCCCTCCCATCCTCGTCAGCACTCAAAGTCCAGAGTCAGCACTGTGTTGAGCGTGCA
9721 CACACCTGGCTGGGTTGTGGCAGCACCCAGGCTTGGCCTCAATTTTACTCTGAGATTGGA
9781 GCAAGCACCAGGATCGGGGAGAGGCCAGGCAGCAGGAGCAGGCATATGCATACCTGTGGG
9841 GGAAGGAGGGCCTTCTCTGGCCCCTGAGAGTGCAGAGATGTTTGGGTCTGCAGCTGCAGC
9901 TTGGGCAGCTGCAGCTGTAGGTGGGAGGGTGGAGGTCCTGCTTACTCCCAGTCTCCAAGA
9961 GCAAAGGGGTGCCCAGGTCCAGAGCAGTGGCTGGGAAGCTGCAGCTGCACATGGGGAGTA
10021 TGAGGTTCCCACCCTGCCAGTTCAGAAGGGTGCAGGGCTTCTGCCTATTCCTGGCTCCCA
10081 CTGGCTCCATGGAGTGCACAACCCTGGCTGCACCTCCCCCATGGTAGCCAGCATCATGGC
Simple_repeat# (TA)n ->
10141 AGCGGCCACTCCAGATGGGCCACCACTGCCATCACTTACACACACACATACATATATGTG
10201 TATATATATGATTATGTAGAAAGAGTGCTACTCTACCAATAACCTAACATTATCTGTCTT
10261 TCTCTTTTCCTAGCATAATAAATGTGTTTGTGTGAATTTTTCATATTCCTCAGTTGTTTC
SINE/MIR# MIRb ->
10321 CCTGTGTGACTAATAGCAAGGTTTGAGGAGACATATTGCTCATCTGAGGGTTGATTTGGG
10381 CCTAAGTTTAAATCTCAGTTCTAACACTCTCAATAGCTATAGAATTCAGCTAACTTCTCA
10441 GATTTTAGGCTTCTCTATCTGTGAAATAATCCTTTATAAGTTTTTTTAAGGATTAAATGA
rs35116437
10501 GATAATATATGTAATGTATGTAGTACAACGTCTGCCACAAATAGAAATTCAATAGATTTC
A
10561 TTTTTCTCTTTAATGAATCCTTTCTATTTTAAGGAGTTAACAACTATCACTGAAGAGGAA
10621 AAAGCTAAATAATGGCTAATCTAGTAGTTCAAATATGCTCTTGAATTCCTAGTACCTCTG
rs35714742
10681 CATTTATTAAATAAGCTAGTCAACCAGAGTTCCTCAAATAATCAACTTTCATTAGAGATC
-/AG
10741 AAGATTTTTGTTTTTCACTGAATCCACTAATTTTTCATCATTGAATAAAGCAACTGGTAT
10801 TTTCTGTAGTGGAAACACATATGACTTCCATTTGAAACCATTTAAACTGAATACAGAGAT
rs1452937
10861 CCAGGAAACAGACTGCAGTAGATCAGAGAGAATCCAGAACATTATTTTTTTAAACATATG
DD 10.9G/C rs1452938 A 0.000
10921 AAAGATACAGGTTTGAATATCAGAAGCACCTGGCTTCCCTTTATAGATAGATGCCAACCA
C 0.102
10981 AATTAGGAATTCTTAGAAGAGAAGAGTTACATCATTAAATAAAGGGTTTGGAGAAGAAAC
DD 11.0G/A rs1452939
rs35234923
11041 TGCCTGGATTTTCAATGTTCATCAACTATGGTTGTGCATATTTTACCAGGTTGCATTTTA
A 0.042
A
11101 TTTTCCTTATTCTTAAAAATATCTATGTAGTTTCTAAAAACATTAGGCTCCATTTGTGCC
11161 AAGGTTAAAGTTTCTTCTCTTCGCAAGAATTCTAATGGAAAAAGCTCCATGTAGTACATT
11221 CATTAGAGCAAATAAAAATAGCAAAGTGCTGCATTCCATTATGATTTCTGCATTTTCATT


11281 CTGTGAAATATTTAAACTAAGTACTAACATAATTTTTGTATTTGATAATGAGCTGATGTT
11341 TGTTAAATTCTATTTGTTTTTTCTTTGGGGAAGCAATGCTGAAGAACTCCTAGGCAGTGG
11401 GCTTTGGCAAGGACCTTTATTACATTTTGTTACTGCTGTATACAGCCTTTTGACATAAAT
11461 TATCTTTTCAAGAAAAGTAATTGATTTTTTGAGTAAGGAATGATAAAGCAACTACTAAAT
DD 11.5G/A rs1452940
11521 CTGATTTTTGTGCCATGCAAAAAGTAAAGCGCTTCTCCACTGGTCACACCTTCGCATTTA
A 0.568
11581 CCCACCCCAAGTCTAAATTTAATCCTGCTTGTGTGTACACCTATTACCTCTAGGTATTTT
11641 TACTCATAGCATTTGATGCATCGTGTTATCATGATTTACCTACCCTCTGTTATGATTATA
11701 ATATTAATAGTATCCCCCTCATAGAAGTAACTTGTGTTATTTATATTTTTGACGAAGAAC
11761 TAAACACAGTAGCTAGTGCCTAGAAGATGCTGCTGTACACAGGTTTATTGATTCAATGAA
11821 TGTTTTGCAACTCAACAAAAACTCTTCCCTAAGAGGGAAAATACAATGGTACTGGCTGCT
11881 TTACTCTTATGCCATCCCCTAGCTTTAGTGTTTCAATTTTCACCAATCAATCCCTCCTAG
DD 11.9A/G rs1901729 rs12805681
11941 TTAATTATCTTGAGTTGAAAACATAACTAGAATAAAAATCTCCTGTCACTTACAATCAAT
DD 12.0C/T rs1901730 rs1901731 G 0.042 C 0.000
12001 ATTTACATGTATTCTCTGACTTCATAATTAATATTTAGATTTCTGTTTATGGACTTAGGC
T 0.095 G 0.000
Simple_repeat# (TAGA)n ->
12061 CAAAACTGTCTTCATAACTTTATTATAGATATATATGTAGATAGATAGATACAGACAGAT
12121 AAACAGATAGTAGATAGGTAGATGTTATCATTAAAGAAAGTATGATGTTAGTAATTTGTT
DD 12.1G/A rs1901732
12181 TATAATTTTGCTTCAATAATTACTTTGTCTTAACATTCTAGAATGAGGTTTTCTTAAATC
A 0.085
12241 TGGTATTTTATGAAGGTAAATGGTTGAATTTGTGGTTCATTCACTATGAGCTATGACAAA
LTR/ERVL# LTR80A ->
12301 GGGTCCAATTTTCAAATAATGCTCGTACTCTCAGCTCTTACCTGATGTTAAGGATACAAG
12361 CAACCAAATGGTGTAGGCACAGCAGAGAGAGGTCTTGGATATCCCATATGGCTTCCTAGG
rs34690804
12421 TCCTTCTTTCCTCATGAGACACATTTTGTGGCCTACAGTAACAGAAAAGGTTTGTAATCT
DD 12.4A/G rs17266200 T
12481 AGGTATGCATGTCATCTTACAGTGTGAAGGCATTTAGATGATGAAAGGTCTGTAGTTTAT
G 0.042
12541 GCCTAGAATATTCAACAGTTTTCATGACATTTTCCAGGGGTATGTGCTTCATAATTTCAT
DD 12.6A/G rs17129669
12601 AAATTCTAGTTTTATTTCCCATAAGAAATCCTAAACCAAGGGCATCCTTTTAAGAACATT
G 0.100
12661 ACACAGATAAGCATTCTTCTGCTAGCAAACATTATTAGTCTGTTACCAAAAGGTCCGTTA
12721 TTAGCAAGTTTATAAAGAAATTCATCTGGAACCCTTTCCTTCTTCCCTCTGGGATTGTGT
12781 TTCCATTTTTGGTACACTAGCAGATCTGGGGCAGTTCCAGTAACCCTTTTCTCCTATGTC
12841 CACCTTCTGACATCAGTTAATAAATATCCATCACATATTTTTCAGGTTTTAACATTTTTT
12901 TATCCTACAAAAATTTTTTTGTGTTACTTAATGAAAAATGTAAGTACACTAACAGTCTAG
12961 TGGCCACCAGAATTTGAAGCATACAGTAGGGATAGAAAGCACCTTGTAAAGCATATTAAT
13021 GACGTACATATCTCTTCAGGGATCTTATATTCTTTAATCAGCTCAGTTACTCTTTTTGCT
Low_complexity# AT_rich ->
13081 CCATCCTCAGTTATTTGAAAACGTTAATGTGTGTTGGTGGAGAGGCTTAATTTAGTTTTA
rs1349700 LINE/L1# L1PA16 ->
13141 ATTTTTTTTACTTATTTTTAAGTTTCATTTTGTAAACATTATAATAGACTTATGACCTTT
G
13201 ATTTTAATTCCAACTTTTATCTTGGATTTGGGGTATACATATGCAGGTTTGTAACATGAG
13261 TATATTGCATGATGCTGAGGTGTGGAATATGGATGATCCCATCAGCCAGGTAGTAAGCAT
13321 TGTACCCAAGGGGTAGTTTTTCAACCCGTTTCCTTTCCTTCCTTTCTAGCCTTCTAGTAG
ENSSNP3991572
13381 TCCTCAGTGCTTATTGTTCCCATCATTATGTCCGTGAGAACCCCAGATTTAGCTCCTACT
G
13441 TATAAGTGAGAATGTGTGGTGTCTGGTTTTCTGTTCCTTCATTAATTTGCTTGGGATAAT
13501 GGCCTCCAGCTACATCCATGTTGCTGCAAAGGACATGATTTATTTCTTTATATGGCCATG
13561 TGGTATTCCGTGGTGTTTATGTACCACATTTTCTTTTTCCAGTCCATAATTTATGAACAC
rs1349701 rs1349702
13621 CTGGGTTGATTCCATGTATTTGCTATTGTGAATAATGCTAAGATAAACATATGTGTGCAT
A A
13681 GTGTCTTTTTGGTAGAACAATTTATTTTCCTTTGGGTATATATCCAGTAATGAGATTGCT
13741 GGGTCAAATGGTAATTGTGTTTTAAGTTCTTTGAGAAATCTCTGAACTACTTTCCACAGT
rs12785351
13801 GGCTGAACTAATTTGCATTCCCACAAACTGTGTAAGTGTTCCCTTTCCCCTGCAGCCTTG
T 0.000
13861 CCAGCATCTGATATTTTTTGACTTTTTAATAATAGCCATTCTGTCTGGTGTGAGACAGTA
13921 TCTCACTGTGGTTTTGATTTGCATTTCCCTAATGGCTAGTGATATGGAGCATTTCTTCAT
13981 ATGTTTATCAGCTAGTTGAATGTCTTCTTTTGAGAAGTGTCTGTTCATGTCCTTTGCCCA
14041 TTTTTTAAAATGGGGTTGTTTTTTGCTTGTTGATTTGTTTAAGTTCTTTGTAGATTCTGG

```
                          rs34071053
14101 ATATTAGACCTTTGTCAGATGCATAGTTCGTGAATGTTTTCTCCTGTTCCGTGGGTTATC
                            -/C
14161 TGTTTACTCTGTTGATAATTTCATTTGCTGTGCAGAAGTCCCACTTGTCAATTTTTGTTT
14221 TTGTTGCAATTGCTTATGGGGATTTAGCCATAAAATCTTTGCTGAAGCTGATGTTCAAAG
14281 GGTATTTCCTAGGTTTTCTTCTAGGATTTCTATAATTTGAGGTCTTACATTTAAATTTTT
14341 AGTCTATCTTGGTTAATATTTTGGGAGGTATATTGTGAGAGGTAGGGCTCCAGTTTCTTT
14401 TTTTTTTTTTTTTTTGCATATGGATATCTAGTTTTCCCAACATAATTTATTGAAGAAACTG
     LINE/L1# L1MA5 ->
14461 GAGCACATTATTCAGCAATCCCACTACTGGGTATATATGCAAAGGAAAGGAAATAAGTAT
                                                    rs1597877
14521 GTTGAAGAGATATCTGCACTATTATGTTTATTGCAGCATTATTCACAATAACCAAAGTAT
                                                      G
14581 GAAATTAACCTATGTCTATCAACTGATGAATGCATAAGAATTTGTGGTGTATTTACACAA
14641 TGAAATATGATTTAGCCATTACAGTGAATGAAATTCTGTCATTTGTGACAACATGAATGT
                                                     rs11220489
14701 GCTTGGAGAATATTTTGCTAAGTAAAATAAGCCAGGCACAGAAAGACAAATACTGCATGT
                                                      T
14761 TCTCACTCATATGTGGAAGCTAAATGCATTGATTTCATCGAAGTAGAGAGTAGAATTGTG
14821 GTAACTACAGGTTAGGAAGGGTAGGGGGTAATGAGGATAGAGAGATGTTGGCTAATGGAT
14881 ACAAAATTACAGCTAGATAGGAGGAATAAGTTCTGGTGTTTTATACCATGTAGCATGCCT
14941 ATAGTTAACAATACTTTATATTTTCAAATAGCTAGAAAAGAGGATTTTGAATGTTTCTAA
15001 T
```

APPENDIX II: PCR PROFILES AND PRIMERS

Appendix II.I PCR Cycles

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
3.2.2, 5.2.1	4.3	DA1.6F + DA5.9R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 60°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA 1st Amplicon 1° PCR
3.2.2, 5.2.1	4.1	DA1.8F + DA5.9aR	1 x (96°C, 1.5'), 34 x (96°C, 20"; 58°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA 1st Amplicon 2° PCR
3.2.2, 5.2.1	5.3	DA5.1F + DA10.4R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 58°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA 2nd Amplicon 1° PCR
3.2.2, 5.2.1	5.1	DA5.2F + DA10.3R	1 x (96°C, 1.5'), 34 x (96°C, 20"; 60°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA 2nd Amplicon 2° PCR
3.2.2, 5.2.1	4.7	DA9.9F + DA14.6R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 58°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA 3rd Amplicon 1° PCR
3.2.2, 5.2.1	4.6	DA9.9aF + da14.5R	1 x (96°C, 1.5'), 34 x (96°C, 20"; 58°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA 3rd Amplicon 2° PCR

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
3.2.2	3.5	DB2.0F + DB5.5R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 64°C, 30"; 65°C, 5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DB 1st Amplicon 1° PCR
3.2.2	3.3	DB2.1F + DB5.4R	1 x (96°C, 1.5'), 34 x (96°C, 20"; 58°C, 30"; 65°C, 4.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DB 1st Amplicon 2° PCR
3.2.2	5.9	DB4.6F + DB10.5R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 64°C, 30"; 65°C, 7'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DB 2nd Amplicon 1° PCR
3.2.2	5.0	DB5.1F + DB10.1R	1 x (96°C, 1.5'), 34 x (96°C, 20"; 60°C, 30"; 65°C, 5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DB 2nd Amplicon 2° PCR
3.2.2	4.6	DB9.9F + DB14.5R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 62°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DB 3rd Amplicon 1° PCR
3.2.2	4.5	DB9.9aF + DB14.4R	1 x (96°C, 1.5'), 34 x (96°C, 20"; 64°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DB 3rd Amplicon 2° PCR

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
3.2.2	5.2	DC1.4F + DC6.6R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 62°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DC 1st Amplicon 1° PCR
3.2.2	4.5	DC1.8F + DC6.3R	1 x (96°C, 1.5'), 34 x (96°C, 20"; 62°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DC 1st Amplicon 2° PCR
3.2.2	4.2	DC5.7F + DC9.9R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 62°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DC 2nd Amplicon 1° PCR
3.2.2	3.5	DC6.1F + DC9.6F	1 x (96°C, 1.5'), 34 x (96°C, 20"; 62°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DC 2nd Amplicon 2° PCR
3.2.2	5.4	DC9.2F + DC14.6R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 62°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DC 3rd Amplicon 1° PCR
3.2.2	4.9	DC9.5F + DC14.4R	1 x (96°C, 1.5'), 34 x (96°C, 20"; 62°C, 30"; 65°C, 5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DC 3rd Amplicon 2° PCR

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
3.2.2	4.4	DD1.9F + DD6.3R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 63°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DD 1st Amplicon 1° PCR
3.2.2	4.2	DD2.2 + DD6.0R	1 x (96°C, 1.5'), 34 x (96°C, 20"; 58°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DD 1st Amplicon 2° PCR
3.2.2	6.0	DD5.4F + DD11.4R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 61°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DD 2nd Amplicon 1° PCR
3.2.2	5.6	DD5.5F + DD11.1F	1 x (96°C, 1.5'), 34 x (96°C, 20"; 63°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DD 2nd Amplicon 2° PCR
3.2.2	4.5	DD10.3F + DD14.8R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 63°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DD 3rd Amplicon 1° PCR
3.2.2	3.7	DD10.5F + DD14.2R	1 x (96°C, 1.5'), 34 x (96°C, 20"; 60°C, 30"; 65°C, 5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DD 3rd Amplicon 2° PCR

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
4.4.1.1	3.3	DA2.6FG + DA5.9aR	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
4.4.1.1	3.3	DA2.6FC + DA5.9aR	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
4.4.1.1	2.2	DA3.7FT + DA5.9aR	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
4.4.1.1	2.2	DA3.7FG + DA5.9aR	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
4.4.1.1	1.9	DA4.0aFG + DA5.9aR	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
4.4.1.1	1.9	DA4.0aFA + DA5.9aR	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
4.4.1.2	0.6	DA9.9aF + DA10.6RG2	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
4.4.1.2	0.6	DA9.9aF + DA10.6RA2	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
4.4.1.2	0.7	DA9.9aF + DA10.6aRA	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
4.4.1.2	0.7	DA9.9aF + DA10.6aRC	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
4.4.1.2	0.7	DA9.9aF + DA10.6bRC	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
4.4.1.2	0.7	DA9.9aF + DA10.6bRA	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
4.4.1.2	2.1	DA9.9aF + DA12.0RC	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
4.4.1.2	2.1	DA9.9aF + DA12.0RT	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
4.4.1.3, 4.6.1	6.6	DA4.0aFG + DA10.6RG2	1 x (96°C, 1.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 59°C, 30"; 65°C, 10'), 23 x (96°C, 20"; 58°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing
4.4.1.3, 4.6.1	6.6	DA4.0aFG + DA10.6RA2	1 x (96°C, 1.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 59°C, 30"; 65°C, 10'), 23 x (96°C, 20"; 58°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing
4.4.1.3, 4.6.1	6.6	DA4.0aFA + DA10.6RG2	1 x (96°C, 1.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 59°C, 30"; 65°C, 10'), 23 x (96°C, 20"; 58°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing
4.4.1.3, 4.6.1	6.6	DA4.0aFA + DA10.6RA2	1 x (96°C, 1.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 59°C, 30"; 65°C, 10'), 23 x (96°C, 20"; 58°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
4.4.1.4	8	DA2.6FG + DA10.6bRC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 64°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 63°C, 30"; 65°C, 10'), 16 x (96°C, 20"; 62°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 7 orientation A (1° PCR)
4.4.1.4	6.6	DA4.0aFG + DA10.6RA2	1 x (96°C, 1.5'), 5 x (96°C, 20"; 61°C, 30"; 65°C, 8.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 8.5'), 24 x (96°C, 20"; 59°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 7 orientation A (2° PCR)
4.4.1.4	8	DA2.6FC + DA10.6bRA	1 x (96°C, 1.5'), 5 x (96°C, 20"; 58°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 57°C, 30"; 65°C, 10'), 16 x (96°C, 20"; 56°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 7 orientation B (1° PCR)
4.4.1.4	6.6	DA4.0aFA + DA10.6RG2	1 x (96°C, 1.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 8.5'), 5 x (96°C, 20"; 59°C, 30"; 65°C, 8.5'), 24 x (96°C, 20"; 58°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 7 orientation B (2° PCR)
4.4.1.4, 4.4.1.5	6.4	DAXOF + DA10.4R	1 x (96°C, 1.5'), 30 x (96°C, 20"; 64°C, 30"; 65°C, 8'), 1 X (65°C, 7')	Hotspot DA, tertiary PCR for donor 7 orientations A and B

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
4.6.1	9.4	DA2.6FG + DA12.0RT	1 x (96°C, 1.5'), 5 x (96°C, 20"; 64°C, 30"; 65°C, 11.5'), 5 x (96°C, 20"; 63°C, 30"; 65°C, 11.5'), 16 x (96°C, 20"; 62°C, 30"; 65°C, 11.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 6 orientation A (1° PCR)
4.6.1	6.9	DA3.7FT1 + DA10.6bRC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 61°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 10'), 24 x (96°C, 20"; 59°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 6 orientation A (2° PCR)
4.6.1	9.4	DA2.6FC + DA12.0RC	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 11.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 6 orientation B (1° PCR)
4.6.1	6.9	DA3.7FG1 + 10.6bRA	1 x (96°C, 1.5'), 5 x (96°C, 20"; 57°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 10'), 24 x (96°C, 20"; 55°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 6 orientation B (2° PCR)
4.6.1	6.6	DA3.8XO +DA10.4R	1 x (96°C, 1.5'), 30 x (96°C, 20"; 64°C, 30"; 65°C, 9.5'), 1 X (65°C, 7')	Hotspot DA, tertiary PCR for donor 6 orientations A and B

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
4.6.2	8	DA2.6FG + DA10.6aRA	1 x (96°C, 1.5'), 5 x (96°C, 20"; 61°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 10'), 16 x (96°C, 20"; 59°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 33 orientation A (1° PCR)
4.6.2	6.6	DA4.0aFG + DA10.6RG2	1 x (96°C, 1.5'), 5 x (96°C, 20"; 63°C, 30"; 65°C, .8.5'), 5 x (96°C, 20"; 62°C, 30"; 65°C, 8.5'), 24 x (96°C, 20"; 61°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 33 orientation A (2° PCR)
4.6.2	8	DA2.6FC + DA10.6aRC	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 33 orientation B (1° PCR)
4.6.2	6.6	DA4.0aFA + DA10.6RA2	1 x (96°C, 1.5'), 5 x (96°C, 20"; 62°C, 30"; 65°C, .8.5'), 5 x (96°C, 20"; 61°C, 30"; 65°C, 8.5'), 24 x (96°C, 20"; 60°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 60°C, 5')	Hotspot DA, pilot and CO assay for donor 33 orientation B (2° PCR)
4.6.2	6.4	DAXO +DA10.4R	1 x (96°C, 1.5'), 30 x (96°C, 20"; 64°C, 30"; 65°C, 8'), 1 X (65°C, 7')	Hotspot DA, tertiary PCR for donor 33 orientations A and B

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
4.6.2	8	DA2.6FG + DA10.6bRC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 64°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 63°C, 30"; 65°C, 10'), 16 x (96°C, 20"; 62°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 67 orientation A (1° PCR)
4.6.2	6.6	DA4.0aFG + DA10.6aRA	1 x (96°C, 1.5'), 5 x (96°C, 20"; 62°C, 30"; 65°C, 8.5'), 5 x (96°C, 20"; 61°C, 30"; 65°C, 8.5'), 24 x (96°C, 20"; 60°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 67 orientation A (2° PCR)
4.6.2	8	DA2.6FC+ DA10.6bRA	1 x (96°C, 1.5'), 5 x (96°C, 20"; 61°C, 30"; 65°C, 10'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 10'), 16 x (96°C, 20"; 59°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 67 orientation B (1° PCR)
4.6.2	6.6	DA4.0aDA + DA10.6aRC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 61°C, 30"; 65°C, 8.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 8.5'), 24 x (96°C, 20"; 59°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 67 orientation B (2° PCR)
4.6.2	6.4	DAXO +DA10.4R	1 x (96°C, 1.5'), 30 x (96°C, 20"; 64°C, 30"; 65°C, 8'), 1 X (65°C, 7')	Hotspot DA, tertiary PCR for donor 67 orientations A and B

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
4.6.2	9.4	DA2.6FG + DA12.0RC	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 11.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 44 orientation A (1° PCR)
4.6.2	6.6	DA4.0aFG + DA10.6RA2	1 x (96°C, 1.5'), 5 x (96°C, 20"; 62°C, 30"; 65°C, 8.5'), 5 x (96°C, 20"; 61°C, 30"; 65°C, 8.5'), 24 x (96°C, 20"; 60°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 44 orientation A (2° PCR)
4.6.2	9.4	DA2.6FC + DA10.20RT	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 11.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 44 orientation B (1° PCR)
4.6.2	6.6	DA4.0aFA + DA10.6RG2	1 x (96°C, 1.5'), 5 x (96°C, 20"; 62°C, 30"; 65°C, 8.5'), 5 x (96°C, 20"; 61°C, 30"; 65°C, 8.5'), 24 x (96°C, 20"; 60°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 44 orientation B (2° PCR)
4.6.2	6.4	DAXO +DA10.4R	1 x (96°C, 1.5'), 30 x (96°C, 20"; 64°C, 30"; 65°C, 8'), 1 X (65°C, 7')	Hotspot DA, tertiary PCR for donor 44 orientations A and B

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
5.2.3	5.5	DA4.8FA + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	5.5	DA4.8FC + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	5.3	DA5.0FT + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	5.3	DA5.0FC + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	5.0	DA5.3FA + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	5.0	DA5.3FT+ DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	4.3	DA6.0FA + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 5.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	4.3	DA6.0FG + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 5.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	4.2	DA6.1FC + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 5.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	4.2	DA6.1FT + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 5.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	3.3	DA7.0FA + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 4'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	3.3	DA7.0FG+ DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 4'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	3.3	DA7.0bFT+ DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 4'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	3.3	DA7.0bFG + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 4'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	2.9	DA7.4FA + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 4'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation
5.2.3	2.9	DA7.4FC + DA10.3R	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 4'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 5' ASP optimisation

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
5.2.3	4.8	DA5.2F + DA10.0RG	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 50°, 53°, 56°, 59°)	Hotspot DA, 3' ASP optimisation
5.2.3	4.8	DA5.2F + DA10.0RA	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 50°, 53°, 56°, 59°)	Hotspot DA, 3' ASP optimisation
5.2.3	5.0	DA5.2F + DA10.2RC	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 50°, 53°, 56°, 59°)	Hotspot DA, 3' ASP optimisation
5.2.3	5.0	DA5.2F + DA10.2RT	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 50°, 53°, 56°, 59°)	Hotspot DA, 3' ASP optimisation
5.2.3	3.5	DA9.9aF + DA13.4R1-	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 4.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
5.2.3	3.5	DA9.9aF + DA13.4R1+	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, 3' ASP optimisation
5.2.3	3.5	DA9.9aF + DA13.4RA	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 4.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
5.2.3	3.5	DA9.9aF + DA13.4RG	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 4.5'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
5.2.3	3.6	DA9.9aF + DA13.5RT1	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, 3' ASP optimisation
5.2.3	3.6	DA9.9aF + DA13.5R1C	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
5.2.3	3.9	DA9.9aF + DA13.8aFG	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation
5.2.3	3.9	DA9.9aF + DA13.8aFA	1 x (96°C, 1.5'), 33 x (96°C, 20"; A°C, 30"; 65°C, 6'), 1 X (60°C, 30"; 65°C, 3') (A: 56°, 59°, 62°, 65°)	Hotspot DA, 3' ASP optimisation

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
5.2.4	6.5	DA3.7FG + DA10.2C	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23
5.2.4	6.5	DA3.7FG + DA10.2T	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23
5.2.4	6.5	DA3.7FT + DA10.2C	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23
5.2.4	6.5	DA3.7FT + DA10.2T	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23
5.2.4	6.6	DA4.0aFG + DA10.6bRC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 59°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 58°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23

5.2.4	6.6	DA4.0aFG + DA10.6bRA	1 x (96°C, 1.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 59°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 58°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23
5.2.4	6.6	DA4.0aFT + DA10.6bRC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 59^{0}C, 30"; 65°C, 8'), 23 x (96°C, 20"; 58°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23
5.2.4	6.6	DA4.0aFT + DA10.6bRA	1 x (96°C, 1.5'), 5 x (96°C, 20"; 60°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 59°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 58°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23
5.2.4	6.2	DA4.0aFG + DA10.2RC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23
5.2.4	6.2	DA4.0aFG + DA10.2RT	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65^{0}C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23
5.2.4	6.2	DA4.0aFT + DA10.2RC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23
5.2.4	6.2	DA4.0aFT + DA10.2RT	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	solving linkage phasing for d23

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
5.2.5	8	DA2.6FG + DA10.6bRC	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 8 orientation B and for donor 12 and for donors 73, 211, 247 orientation A (1° PCR)
5.2.5	6.9	DA3.7FG1 + DA10.6RG2	1 x (96°C, 1.5'), 33 x (96°C, 20"; 59°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 8 orientation B (2° PCR)
5.2.5	6.9	DA3.7FG1 +DA10.6RA2	1 x (96^{0}C, 1.5'), 33 x (96^{0}C, 20"; 58^{0}C, 30"; 65^{0}C, 8.5'), 1 X (60^{0}C, 30"; 65^{0}C, 5')	Hotspot DA, pilot and CO assay for donor 12 orientation A (2° PCR)
5.2.5	9.4	DA2.6FG + DA12.0RC	1 x (96^{0}C, 1.5'), 26 x (96^{0}C, 20"; 63^{0}C, 30"; 65^{0}C, 11.5'), 1 X (60^{0}C, 30"; 65^{0}C, 5')	Hotspot DA, pilot and CO assay for donors 17, 256 orientation A (1° PCR)
5.2.5	6.9	DA3.7FT1 + DA10.6bRA	1 x (96°C, 1.5'), 33 x (96°C, 20"; 58°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 17 orientation A (2° PCR)
5.2.5	8.0	DA2.6FC + DA10.6bRC	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 26 orientation B (1° PCR)
5.2.5	6.9	DA3.7FG1 + DA10.6aRA	1 x (96°C, 1.5'), 5 x (96°C, 20"; 59°C, 30"; 65°C, 8.5'), 5 x (96°C, 20"; 58°C, 30"; 65°C, 8.5'), 23 x (96°C, 20"; 57°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65^{0}C, 5')	Hotspot DA, pilot and CO assay for donor 26 orientation B (2° PCR)
5.2.5	9.4	DA2.6FG + DA12.0RT	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 11.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 31, 185 orientation A (1° PCR)
5.2.5	6.9	DA3.7FT +DA10.6bRC	1 x (96°C, 1.5'), 33 x (96°C, 20"; 63°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 31 orientation A (2° PCR)
5.2.5	6.9	DA3.7FT1 + DA10.6aRA	1 x (96°C, 1.5'), 33 x (96°C, 20"; 58°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 73, 180, 269 orientation A (2° PCR)
5.2.5	8.7	DA4.8FC + DA13.5RC1	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 10.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 178 orientation A and for donor 261 orientation B (1° PCR)
5.2.5	8.1	DA5.3FT + DA13.4aR1-	1 x (96°C, 1.5'), 33 x (96°C, 20"; 56°C, 30"; 65°C, 10'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 178 orientation A and for donor 186 orientation B (2° PCR)
5.2.5	9.4	DA2.6FG + DA12.0RT	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 11.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 180, 253, 269, 280 orientation A (1° PCR)
5.2.5	7.7	DA6.1FT + DA13.8aRA	1 x (96°C, 1.5'), 26 x (96°C, 20"; 62°C, 30"; 65°C, 9.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 181, 184 orientation B (1° PCR)
5.2.5	6.5	DA7.0FG + DA13.5RC1	1 x (96°C, 1.5'), 33 x (96°C, 20"; 60°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 181, 184 orientation B (2° PCR)

5.2.5	6.5	DA3.7FT1 + DA10.2C	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 185, 253, 256, 280 orientation A (2° PCR)
5.2.5	8.7	DA4.8FC + DA13.8aRG	1 x (96°C, 1.5'), 26 x (96°C, 20"; 59°C, 30"; 65°C, 11'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 186 orientation A (1° PCR)
5.2.5	6.2	DA4.0aFG + DA10.2RC	1 x (96°C, 1.5'), 33 x (96°C, 20"; 57°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 211, 236, 251 orientation A (2° PCR)
5.2.5	9.4	DA2.6FC + DA12.0RC	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 11.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 232 orientation B (1° PCR)
5.2.5	6.9	DA3.7FG1 + DA10.6bRA	1 x (96°C, 1.5'), 33 x (96°C, 20"; 58°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 232 orientation B (2° PCR)
5.2.5	6.9	DA3.7FT +DA10.6bRC	1 x (96°C, 1.5'), 26 x (96°C, 20"; 63°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 236, 238, 251 orientation A (1° PCR)
5.2.5	6.2	DA4.0aFG + DA10.2RC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 8'), 5 x (96°C, 20"; 55°C, 30"; 65°C, 8'), 23 x (96°C, 20"; 54°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 238 orientation A (2° PCR)
5.2.5	7.7	DA6.1FC + DA13.8aRA	1 x (96°C, 1.5'), 26 x (96°C, 20"; 60°C, 30"; 65°C, 9.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 243, 272 orientation B (1° PCR)
5.2.5	6.9	DA7.4FA + DA13.4aR1+	1 x (96°C, 1.5'), 33 x (96°C, 20"; 60°C, 30"; 65°C, 7.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 243 orientation A (2° PCR)
5.2.5	6.4	DA7.0FG + DA13.4R1-	1 x (96°C, 1.5'), 26 x (96°C, 20"; 59°C, 30"; 65°C, 8'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 244, 259 orientation B (1° PCR)
5.2.5	3.2	DA7.0bFT + DA10.2RC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 57°C, 30"; 65°C, 4'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 4'), 23 x (96°C, 20"; 56°C, 30"; 65°C, 4'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 244 orientation B (2° PCR)
5.2.5	6.9	DA3.7FG1 + DA10.6aRA	1 x (96°C, 1.5'), 26 x (96°C, 20"; 58°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 247 orientation A (2° PCR)
5.2.5	2.8	DA7.4FC + DA10.2RC	1 x (96°C, 1.5'), 5 x (96°C, 20"; 57°C, 30"; 65°C, 4'), 5 x (96°C, 20"; 56°C, 30"; 65°C, 4'), 23 x (96°C, 20"; 56°C, 30"; 65°C, 4'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donors 259 orientation B (2° PCR)
5.2.5	5.3	DA5.3FT + DA10.6bRA	1 x (96°C, 1.5'), 33 x (96°C, 20"; 59°C, 30"; 65°C, 7'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 261 orientation B (2° PCR)
5.2.5	6.4	DA13.4aR+ + DA7.0bFG	1 x (96°C, 1.5'), 33 x (96°C, 20"; 60°C, 30"; 65°C, 7.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 272 orientation A (2° PCR)
5.2.5	6.8	DA7.0bRT + DA13.8aRA	1 x (96°C, 1.5'), 26 x (96°C, 20"; 60°C, 30"; 65°C, 8.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 279 orientation A (1° PCR)
5.2.5	6.0	DA7.4FA + DA13.4aR+	1 x (96°C, 1.5'), 33 x (96°C, 20"; 60°C, 30"; 65°C, 7.5'), 1 X (60°C, 30"; 65°C, 5')	Hotspot DA, pilot and CO assay for donor 279 orientation A (2° PCR)

Section	Target Molecule Size (kb)	Primers used	Cycling conditions	Remarks
5.2.7	2.6	DA7.4F +DA10.0R	1 x (96°C, 1.5'), 15 x (96°C, 20"; 58°C, 30"; 65°C, 3.5'), 1 X (65°C, 7')	Hotspot DA, tertiary PCR for donor homozygous for supressing SNP DA7.5
5.2.7	2.6	DA7.4F +DA10.0R	1 x (96°C, 1.5'), 25 x (96°C, 20"; 58°C, 30"; 65°C, 3.5'), 1 X (65°C, 7')	Hotspot DA, tertiary PCR for donor homozygous for supressing SNP DA7.5
5.2.9	3.3	DA5.2F + DA8.6R	1 x (96°C, 1.5'), 33 x (96°C, 20"; 62°C, 30"; 65°C, 4.5'), 1 X (65°C, 7')	Hotspot DA, tertiary PCR for sequencing
5.2.9		DA7.3F	1 x (96°C, 1.5'), 15 x (96°C, 10"; 50°C, 5"; 60°C, 4'), 1 X (65°C, 7')	Hotspot DA, sequencing using forward primer
5.2.9		DA7.6R	1 x (96°C, 1.5'), 15 x (96°C, 10"; 50°C, 5"; 60°C, 4'), 1 X (65°C, 7')	Hotspot DA, sequencing using reverse primer
5.2.9			98°C, 5'; 25°C, 10'	Hotspot DA, Big Dye V3.1 clean up after sequencing cycling

Appendix II.II Primers

II.II.I Universal Primers

Section	Universal Primer	Sequence (5' -> 3')
3.2.2, 5.2.1	DA1.6F	AGAATGTTCTGAGCTTGCCC
3.2.2, 5.2.1	DA1.8F	CATGTGTTCAGCCTGAACCC
3.2.2, 5.2.1	DA5.9R	GAACTGGTAAGCTTCCTGCC
3.2.2, 5.2.1	DA5.9aR	GGTGGATATCACATTGGACC
3.2.2, 5.2.1	DA5.1F	GGACTTTTGGTCTCTATCTGCC
3.2.2, 5.2.1, 5.2.3	DA5.2F	AGGACCCTAGAGCAGAGTCC
3.2.2, 5.2.1, 5.2.3	DA10.3R	TAGGCTAGCGTTCAGATCCC
3.2.2, 4.4.1.4, 4.4.1.5, 4.6.1, 4.6.2, 5.2.1	DA10.4R	ACATGTTCCTTCCAAGTCCC
3.2.2, 5.2.1	DA9.9F	GACAAGGCACGACTGACTCC
3.2.2, 5.2.1, 5.2.3	DA9.9aF	CAGAAATGAGGCTAGATGCC
3.2.2, 5.2.1	DA14.5R	CTGAAACACAGTAAGAATGCCC
3.2.2, 5.2.1	DA14.6R	AGGCCATATACAACAGGCCC

Section	Universal Primer	Sequence (5' -> 3')
3.2.2	DC1.4F	CGAGGAATCTGCGTTTTAGC
3.2.2	DC1.8F	GAATGTTTAGTGGGAGAGCC
3.2.2	DC6.6R	GAACAGAGGTGGGATAGCAG
3.2.2	DC6.3R	GATGGTCAACAGAAAAGGATC
3.2.2	DC5.7F	GGCTTCTGGAATACTGCTGC
3.2.2	DC6.1F	CCCTTAATGTATCTGGCAGGTA
3.2.2	DC9.6R	GGTTGGTGTGTAAACGGC
3.2.2	DC9.9R	TAGAGTTTGCTGTCTGCACCC
3.2.2	DC9.2F	GGCTACAAGTCCCAAATTGC
3.2.2	DC9.5F	GGGTAATATCCAGTCCCAAGT
3.2.2	DC14.4R	ATGTTGTGTGTTGAAGCGGG
3.2.2	DC14.6R	GCTTTCTCACCTGTAAGATGGC

Section	Universal Primer	Sequence (5' -> 3')
3.2.2	DB2.5FT	ccccccAATATCTCATTTTTGAAAT
3.2.2	DB2.5FC	ccccccAATATCTCATTTTTGAAAC
3.2.2	DB4.3FC	GTTCTACCCCGACTCTGCTC
3.2.2	DB4.3FG	GTTCTACCCCGACTCTGCTG
3.2.2	DB10.5RA	ACCACTTGATGACTACAT
3.2.2	DB10.5RG	CCACTTGATGACTACAC
3.2.2	DB12.3RT	cccccGCTTTATTCATTATAGTCCA
3.2.2	DB12.3RA	cccccGCTTTATTCATTATAGTCCT
3.2.2	DB13.5RG	cccccGTCTCTAGACATTAGCAC
3.2.2	DB13.5RA	cccccGTCTCTAGACATTAGCAT
3.2.2	DB14.3RG	CAACCAGAAGGAAGAGCC
3.2.2	DB14.3RC	CAACCAGAAGGAAGAGCG

Section	Universal Primer	Sequence (5' -> 3')
3.2.2	DD 1.9F	GCCTGTAGGACTGTAGCCA
3.2.2	DD 2.2F	CCCTAAGATTATGGGTGC
3.2.2	DD 6.0R	GAGAGGAGACCAAAATGGG
3.2.2	DD 6.3R	CACTTCTCAGACTTTGTAAGGG
3.2.2	DD 5.4F	GGTTCCCAAGTTACTCTGG
3.2.2	DD 5.5F	CAGGAACCCAACTTTTCAGTC
3.2.2	DD 11.1R	CCTTGGCACAAATGGAGCC
3.2.2	DD 11.4R	CTTGCCAAAGCCCACTGC
3.2.2	DD 10.3F	GTGACTAATAGCAAGGTTTGAG
3.2.2	DD 10.5F	CTATCACTGAAGAGGAAAAAGC
3.2.2	DD 14.2R	GCTTCAGCAAAGATTTTATGGC
3.2.2	DD 14.8R	CATCTCTCTATCCTCATTACCC

Section	Universal Primer	Sequence (5' -> 3')
4.4.1.4, 4.4.1.5, 4.6.2	DA4.0X0F	CTGCAGTGGGAGCTCAGTCTCC
4.6.1	DA3.8XO	CCTGCTGTGTCTGATTTGTCATc
5.2.7	DA7.4F	GCTCCACAGGGCTAGGGAG
5.2.7	DA10.0R	CAGCATCCTCTAAATTGCAGG
5.2.9	DA7.3F	GGCAGTTAGTAGGTGTATTAG
5.2.9	DA7.6R	CATCCTGAGTTTTAGCTCCC

II.II.II Allele-Specific Primers (ASPs)

Section	Allele Specific Primer (ASP)	Sequence (5' -> 3)
4.4.1.1, 4.4.1.4, 4.6.1, 4.6.2, 5.2.5	DA2.6FG	cccccGATACTTCCCAACATTGAG
4.4.1.1, 4.4.1.4, 4.6.1, 4.6.2, 5.2.5	DA2.6FC	cccccGATACTTCCCAACATTGAC
4.4.1.1, 5.2.4, 5.2.5	DA3.7FT	cccccAAACCTCTTTTAGTGACAT
4.4.1.1, 5.2.4, 5.2.5	DA3.7FG	cccccAAACCTCTTTTAGTGACAG
4.4.1.1, 4.4.1.3, 4.6.1, 4.4.1.4, 4.6.2, 5.2.4, 5.2.5	DA4.0aFG	cccCTGTACGGGGAAAATACTG
4.4.1.1, 4.4.1.3, 4.6.1, 4.4.1.4, 4.6.2, 5.2.4	DA4.0aFA	ccccCTGTACGGGGAAAATACTA
4.4.1.2, 4.4.1.3, 4.6.1, 4.4.1.4, 4.6.2, 5.2.5	DA10.6G2	CAGGTTGTTAGCAAGC
4.4.1.2, 4.4.1.3, 4.6.1, 4.4.1.4, 4.6.2, 5.2.5	DA10.6A2	GCAGGTTGTTAGCAAGT
4.4.1.2, 4.6.2, 5.2.5	DA10.6aRA	CCCCATTKTCAGGAGAGT
4.4.1.2, 4.6.2	DA10.6aRC	CCCCATTKTCAGGAGAGG
4.4.1.2, 4.4.1.4, 4.6.1, 4.6.2, 5.2.4, 5.2.5	DA10.6bRC	TAACAGGAACTCCCCATTG
4.4.1.2, 4.4.1.4, 4.6.1, 4.6.2, 5.2.4, 5.2.5	DA10.6bRA	TAACAGGAACTCCCCATTT
4.4.1.2, 4.6.1, 4.6.2, 5.2.5	DA12.0RC	TAGGTTAGGTACTGGGCG
4.4.1.2, 4.6.1, 4.6.2, 5.2.5	DA12.0RT	TAGGTTAGGTACTGGGCA
4.6.1, 5.2.5	DA3.7FT1	ccAAACCTCTTTTAGTGACAT
4.6.1, 5.2.5	DA3.7FG1	ccAAACCTCTTTTAGTGACAG
5.2.3	DA4.8FA	GGAGAGAGGCCTGGGGA
5.2.3, 5.2.5	DA4.8FC	GGAGAGAGGCCTGGGGC
5.2.3	DA5.0FT	CTTATGCAGGAGTCTGT
5.2.3	DA5.0FC	CTTATGCAGGAGTCTGC
5.2.3	DA5.3FA	CCCCATTATCTCACTGGATTTTCA
5.2.3, 5.2.5	DA5.3FT	CCCCATTATCTCACTGGATTTTCT

Section	Universal Primer	Sequence (5' - 3')
5.2.3	DA6.0FA	GCTCTCAGTCTCCTATTTCTTA
5.2.3	DA6.0FG	GCTCTCAGTCTCCTATTTCTTG
5.2.3, 5.2.5	DA6.1FC	CCAAGCTGTGAGAGAGTCC
5.2.3, 5.2.5	DA6.1FT	CCAAGCTGTGAGAGAGTCT
5.2.3	DA7.0FA	cGGATTTAAACCTGTCTGA
5.2.3, 5.2.5	DA7.0FG	cGGATTTAAACCTGTCTGG
5.2.3, 5.2.5	DA7.0bFT	ccGCAAAAAAAGTCTAAAGGCAT
5.2.3, 5.2.5	DA7.0bFG	ccGCAAAAAAAGTCTAAAGGCAG
5.2.3, 5.2.5	DA7.4FA	ccGAAGTTTAATGGACTCACA
5.2.3	DA7.4FC	ccGAAGTTTAATGGACTCACC
5.2.3	DA10.0RG	GTCAACTACTACTACCCCC
5.2.3	DA10.0RA	GTCAACTACTACTACCCCT
5.2.3, 5.2.4, 5.2.5	DA10.2RC	ccccccCTAAAAGAATGAAATAATG
5.2.3, 5.2.4	DA10.2RT	ccccccCTAAAAGAATGAAATAATA
5.2.3, 5.2.5	DA13.4aR1-	CAAACTACCCATCTGATA
5.2.3, 5.2.5	DA13.4aR1+	CAAACTACCCATCTGATAA
5.2.3	DA13.4RA	CGGGGATTCATAACTAGAATAT
5.2.3	DA13.4RG	CGGGGATTCATAACTAGAATAC
5.2.3	DA13.5RT1	GAAACTCTCCAAGACAAYA
5.2.3, 5.2.5	DA13.5RC1	GAAACTCTCCAAGACAAYG
5.2.3, 5.2.5	DA13.8aRG	ccccGAAAATCCATAAACAAAC
5.2.3, 5.2.5	DA13.8aRA	ccccGAAAATCCATAAACAAAA

APPENDIX III: ALLELE-SPECIFIC OLIGOS (ASOs)

ASO	Sequence (5' - 3')	ASO	Sequence (5' - 3')
DA1.9-	TCCAGTGGAAGTACTAAA	DA7.6A	CTGTCATAAGAACAGCAT
DA1.9+	TCCAGTGGGAAGTACTAA	DA7.6C	CTGTCATCAGAACAGCAT
DA2.4-	AATCTCCCTGCAACATAG	DA8.2G	GAAGCCAGTAATGAATGT
DA2.4+	AATCTCCCCTGCAACATA	DA8.2A	GAAGCCAATAATGAATGT
DA2.6G	ACATTGAGTAAGATAGAC	DA8.4G	ACACTACGCTCCTGCATG
DA2.6C	ACATTGACTAAGATAGAC	DA8.4A	ACACTACACTCCTGCATG
DA3.5T	ACAAAATTGAGRACAGGG	DA8.5C	GGAAATACGAATGTCTCA
DA3.5G	ACAAAATGGAGRACAGGG	DA8.5G	GGAAATAGGAATGTCTCA
DA3.7T	AGTGACATCTGGTGCATG	DA9.2+	AGATTATTTTTTGAATGT
DA3.7G	AGTGACAGCTGGTGCATG	DA9.2-	TAGATTATTTTTGAATGT
DA4.0C	GACAGCCCCTCCTSTGGA	DA9.5C	GCAATTTCTGAAGAACTC
DA4.0T	GACAGCCTCTCCTSTGGA	DA9.5T	GCAATTTTTGAAGAACTC
DA4.0aG	AAATACTGTAATAAGCAC	DA10.0G	TGCTGGAGGGGGTAGTAG
DA4.0aA	AAATACTATAATAAGCAC	DA10.0A	TGCTGGAAGGGGTAGTAG
DA4.8A	CCTGGGGACATTTAGGCT	DA10.2C	AGATGGGCATTATTTCAT
DA4.8C	CCTGGGGCCATTTAGGCT	DA10.2T	AGATGGGTATTATTTCAT
DA5.OT	GAGTCTGTACCCTGGCCC	DA10.6G	GCCCTGTGCTTGCTAACA
DA5.OC	GAGTCTGCACCCTGGCCC	DA10.6A	GCCCTGTACTTGCTAACA
DA5.3A	GATTTTCATAGTATATTT	DA10.6aA	ATCTCCATACTCTCCTGA
DA5.3T	GATTTTCTTAGTATATTT	DA10.6aC	ATCTCCATCCTCTCCTGA
DA6.0A	ATTTCTTAGATAACTCCA	DA10.6bC	CTCCTGACAATGGGGAGT
DA6.0G	ATTTCTTGGATAACTCCA	DA10.6bA	CTCCTGAAAATGGGGAGT
DA6.1C	GAGAGTCCGAGTTTCCCA	DA10.8-	TGGAGTCTACTGGGTTGG
DA6.1T	GAGAGTCTGAGTTTCCCA	DA10.8+	TGGAGTCTTACTGGGTTG
DA6.1aA	AATTACCACAAACTTAAT	DA11.5C	TATATAACATGGCCTGGC
DA6.1aG	AATTACCGCAAACTTAAT	DA11.5T	TATATAATATGGCCTGGC
DA6.3T	CATTCCTTGGCTTGTAGC	DA12.0C	ATTCACGCGCCCAGTACC
DA6.3C	CATTCCTCGGCTTGTAGC	DA12.0T	ATTCACGTGCCCAGTACC
DA6.3aA	CCCTCTCACTCCAATCTC	DA13.4A	TSTTTGTATATTCTAGTT
DA6.3aT	CCCTCTCTCTCCAATCTC	DA13.4G	TSTTTGTGTATTCTAGTT
DA7.0A	CTGTCTGAGCATGTCCAG	DA13.4a-	GAATCCCCTATYAGATGG
DA7.0G	CTGTCTGGGCATGTCCAG	DA13.4a+	GAATCCCCTTATYAGATG
DA7.0a-	CAGCAAAGCCCTCTTTGC	DA13.4bC	CCCYTATCAGATGGGTAG
DA7.0a+	CAGCAAAGGACCCTCTTT	DA13.4bT	CCCYTATTAGATGGGTAG
DA7.0bT	AAAGGCATTGGAGAACAA	DA13.5T	CCCAGATTRTTGTCTTGG
DA7.0bG	AAAGGCAGTGGAGAACAA	DA13.5C	CCCAGATCRTTGTCTTGG
DA7.4A	GACTCACAGCTCCACAGG	DA13.8-	TTTTCCCCAGTGTGTGTT

DA7.4C	GACTCACCGCTCCACAGG	DA13.8+	TTTTCCCCCAGTGTGTGT
DA7.5T	TAGGGAGTCCTCACTATC	DA13.8aG	TGTTTGGGTTTGTTTATG
DA7.5G	TAGGGAGGCCTCACTATC	DA13.8aA	TGTTTGGATTTGTTTATG
DA7.5aT	TAAGGCATGTGTTACACG	DA13.8bC	GCCACTACCATTCTGTTT
DA7.5aC	TAAGGCACGTGTTACACG	DA13.8bT	GCCACTATCATTCTGTTT
		DA14.4T	TCTCTTGTCTGATTGCTC
		DA14.4C	TCTCTTGCCTGATTGCTC

ASO	Sequence (5' - 3')	ASO	Sequence (5' - 3')
DB 2.2T	TGAGATATGTTACTACTA	DB 5.9A	GAGAAAGAAAGGAAGGAA
DB 2.2C	TGAGATACGTTACTACTA	DB 5.9-	GAGAAAGAAGGAAGGAAG
DB 2.3G	AAGGGAGGTGCCTAGACT	DB 5.9aGA	AGGAAGGAAGGAAGGAAG
DB 2.3C	AAGGGAGCTGCCTAGACT	DB 5.9a-	AGGAAGGAAGGAAGGAAG
DB 2.4G	AGCTTCCGGAGACCCAGA	DB 6.0T	AGGAAGGTAGGAAGGAAG
DB 2.4A	AGCTTCCAGAGACCCAGA	DB 6.0A	AGGAAGGAAGGAAGGAAG
DB 2.5T	TTTGAAATTGGTTACATA	DB 6.6A	ACCACACATGGTTGCTGA
DB 2.5C	TTTGAAACTGGTTACATA	DB 6.6G	ACCACACGTGGTTGCTGA
DB 2.7T	TCTTACTTTTTAATCTGG	DB 7.0T	ATTCTATTGGGCAATGCA
DB 2.7C	TCTTACTCTTTAATCTGG	DB 7.0C	ATTCTATCGGGCAATGCA
DB 2.7aC	TGTGAGTCGCATTATATT	DB 7.0aC	GTCATGCCGTGAAACCCC
DB 2.7aG	TGTGAGTGGCATTATATT	DB 7.0aA	GTCATGCAGTGAAACCCC
DB 3.2T	TGGATCATGAGGTCAGGA	DB 7.1G	AAAGGGTGTGGATAAATA
DB 3.2C	TGGATCACGAGGTCAGGA	DB 7.1T	AAAGGGTTTGGATAAATA
DB 3.2aC	TGAAACCCCGTCTTTACT	DB 7.2T	CACCTTATTGATCTGCAG
DB 3.2aT	TGAAACCTCGTCTTTACT	DB 7.2C	CACCTTACTGATCTGCAG
DB 3.3C	CTAAAAACACACAAAAAA	DB 7.5T	TGAGGCCTCCCTTTCCAC
DB 3.3A	CTAAAAAAACACAAAAAA	DB 7.5A	TGAGGCCACCCTTTCCAC
DB 3.3aG	GCTGGGCGTGGTGGTGTG	DB 7.6C	CCACAGTCACCTAGTCAC
DB 3.3aA	GCTGGGCATGGTGGTGTG	DB 7.6T	CCACAGTTACCTAGTCAC
DB 3.3bA	TAATCCCAGCTACTTGGG	DB 7.8T	ATACATGTGCCATGTTGG
DB 3.3bC	TAATCCCCGCTACTTGGG	DB 7.8C	ATACATGCGCCATGTTGG
DB 3.4A	TTGAACCAGGGAGTCGGA	DB 7.9T	GTATATCTCCTAATGCTA
DB 3.4G	TTGAACCGGGGAGTCGGA	DB 7.9A	GTATATCACCTAATGCTA
DB 3.4aC	CTGAGATCACACCACAGC	DB 8.1C	AGTTCTTCGTAGATTCTG
DB 3.4aA	CTGAGATAACACCACAGC	DB 8.1T	AGTTCTTTGTAGATTCTG
DB 3.7A	ATATACTATAATAGACAC	DB 9.0C	CCCAGCACCATTTATTAA
DB 3.7G	ATATACTGTAATAGACAC	DB 9.0T	CCCAGCATCATTTATTAA
DB 3.9G	TCCTCAGGTCTCTGTACA	DB 9.2G	TCTTTTGGCTTAGGATTG
DB 3.9C	TCCTCAGCTCTCTGTACA	DB 9.2A	TCTTTTGACTTAGGATTG
DB 4.1C	GAATATGCAGGTTTTACT	DB 9.6C	TTTTGGGCTGAGACGATG
DB 4.1A	GAATATGAAGGTTTTACT	DB 9.6T	TTTTGGGTTGAGACGATG
DB 4.1a-	TTTGCCATCTGAATATAG	DB 9.8C	CAGTTTTCAAAGGGAATG
DB 4.1aG	TTTGCCATGCTGAATATA	DB 9.8-	CAGTTTTAAAGGGAATGC
DB 4.1bA	CAGACAGAAAAGTATGAC	DB 10.5A	TCTGACCATGTAGTCATC
DB 4.1bG	CAGACAGGAAAGTATGAC	DB 10.5G	TCTGACCGTGTAGTCATC
DB 4.2A	CCTCGATACTGACCAACT	DB 10.8T	ACAAAAATTAGCAGGGTA
DB 4.2G	CCTCGATGCTGACCAACT	DB 10.8C	ACAAAAACTAGCAGGGTA

DB 4.3C	CTCTGCTCCTGCTCRGCA	DB 11.0A	CCTGTATATGAAAGAACT
DB 4.3G	CTCTGCTGCTGCTCRGCA	DB 11.0G	CCTGTATGTGAAAGAACT
DB 4.3aA	SCTGCTCAGCAGCAAAGT	DB 11.2G	AAATTCAGCCTGGTCAAC
DB 4.3aG	SCTGCTCGGCAGCAAAGT	DB 11.2A	AAATTCAACCTGGTCAAC
DB 4.5A	AAGCCACATTTGAGGCAT	DB 12.3T	TCTAGTCTGGACTATAAT
DB 4.5G	AAGCCACGTTTGAGGCAT	DB 12.3A	TCTAGTCAGGACTATAAT
DB 5.2G	GCTGTTGTTTTCCCTGAC	DB 13.0C	TTACAACCCCTGTTTCTT
DB 5.2A	GCTGTTATTTTCCCTGAC	DB 13.0A	TTACAACACCTGTTTCTT
DB 5.3G	TCAGTGTGGACTCTGGCT	DB 13.4C	CACAGCCCTGATCTATTG
DB 5.3A	TCAGTGTAGACTCTGGCT	DB 13.4-	CACAGCCTGATCTATTGC
DB 5.4G	AGTCACTGCAATARGTTG	DB 13.4aA	ATGAAAAAGTTAATGAAG
DB 5.4A	AGTCACTACAATARGTTG	DB 13.4a-	ATGAAAAGTTAATGAAGT
DB 5.4aG	TRCAATAGGTTGGGCCAG	DB 13.5G	GGACATCGTGCTAATGTC
DB 5.4aA	TRCAATAAGTTGGGCCAG	DB 13.5A	GGACATCATGCTAATGTC
		DB 14.3G	TTTCTTTGGCTCTTCCTT
		DB 14.3C	TTTCTTTCGCTCTTCCTT

ASO	Sequence (5' - 3')	Allele Specific Oligo	Sequence (5' - 3')
DC2.2G	CGAGCTTTGCAGCCTGAG	DC7.5bC1	CCCTGACCCCCTCACTCC
DC2.2T	CGAGCTTTTCAGCCTGAG	DC7.5bT2	CCCTGACCTCCTCACTCC
DC3.0G	GTTACGCAGAAGTATTAT	DC7.5cC	AGTAGCCCCCGTGTCTGT
DC3.0A	GTTACGCAAAAGTATTAT	DC7.5cT	AGTAGCCTCCGTGTCTGT
DC3.7TA	CTGTTTTTAAATGCGTCAT	DC7.6C	GCATCCACGTTGCTGCA
DC3.7-	CTGTTTTAATGCGTCAT	DC7.6T	GCATCCATGTTGCTGCA
DC4.5A	TCTAATCACTCCTTTGCT	DC7.7C	TTCACTACTGATGGCCAC
DC4.5G	TCTAATCGCTCCTTTGCT	DC7.7T	TTCACTATTGATGGCCAC
DC5.4C	ATGACAACGTCTCTCTAG	DC8.2C	TCATACTCGATATATTTG
DC5.4T	ATGACAATGTCTCTCTAG	DC8.2T	TCATACTTGATATATTTG
DC5.6T	AAATCTCTGGCTTAAATT	DC8.7G	TTTAAGCGTCATTTAAC
DC5.6C	AAATCTCCGGCTTAAATT	DC8.7A	TTTAAGCATCATTTAAC
DC5.8T	TTCTAAGTCAAGATTCTC	DC9.1A	CATAATTCAAGGCCTGATT
DC5.8-	TTCTAAGCAAGATTCTCA	DC9.1G	CATAATTCGAGGCCTGATT
DC6.1C	ATCTCATCTCTGAGTTTT	DC9.6G	TTGGGAAGAAGAAGTCAG
DC6.1T	ATCTCATTTCTGAGTTTT	DC9.6A	TTGGGAAAAAGAAGTCAG
DC6.3T	TCGCTGCTTATGTTTTCAC	DC9.9C	GACTGAGCTGCGTCCCCC
DC6.3A	TCGCTGCATATGTTTTCAC	DC9.9G	GACTGAGGTGCGTCCCCC
DC6.7G	CTGAGCTGCCCAATGAGT	DC11.1C	AGGAAAGCGCTTACCATC
DC6.7A	CTGAGCTACCCAATGAGT	DC11.1T	AGGAAAGTGCTTACCATC
DC6.7aG	GTGTTTCCGGCTTCCTCTTG	DC11.2G	AGAAGTCGCCATACAAT
DC6.7aA	GTGTTTCCAGCTTCCTCTTG	DC11.2A	AGAAGTCACCATACAAT

DC6.8A	GTTTGTCCACATCTACTT	DC12.1G	GACTTTGGTGCCTTTTTG
DC6.8G	GTTTGTCCGCATCTACTT	DC12.1C	GACTTTGCTGCCTTTTTG
DC6.8aT	CTACTTATATTTTGGGGTC	DC13.0A	TGGAATCAGTCTAGAAKC
DC6.8aC	CTACTTACATTTTGGGGTC	DC13.0G	TGGAATCGGTCTAGAAKC
DC7.1T	CTTTCTGATGTTTCATT	DC13.3C	CAGAACACGAGGACAAAC
DC7.1C	CTTTCTGACGTTTCATT	DC13.3T	CAGAACATGAGGACAAAC
DC7.2C	GAAAAGTCACAAGAAATG	DC13.4T	GCGTTTATGTCTCCATGT
DC7.2T	GAAAAGTTACAAGAAATG	DC13.4C	GCGTTTACGTCTCCATGT
DC7.4G	GCGTGCTGCCCAACAACTG	DC13.9A	TCAGATGATTTYGAGAAG
DC7.4A	GCGTGCTACCCAACAACTG	DC13.9G	TCAGATGGTTTYGAGAAG
DC7.5C	GTTTTTCCCTCCCTGAC	DC13.9aC	ATGRTTTCGAGAAGCTGTG
DC7.5A	GTTTTTCCTTCCCTGAC	DC13.9aT	ATGRTTTTGAGAAGCTGTG
DC7.5a-	TCCCTGACYCCTCACTCC	DC14.4A	ACAAAACACTGGTCGCTT
DC7.5aC	TCCCTGACCYCCTCACTC	DC14.4G	ACAAAACGCTGGTCGCTT
DC7.5bC	CCCTGACCCCTCACTCCC		
DC7.5bT	CCCTGACTCCTCACTCCC		

ASO	Sequence (5' - 3')	ASO	Sequence (5' - 3')
DD 2.2T	GTGCCATTTTTCAGGATG	DD 8.5aG	CCACTGGGCTCTTTCTGC
DD 2.2C	GTGCCATCTTTCAGGATG	DD 8.5aA	CCACTGGACTCTTTCTGC
DD 2.3G	CAAAGGGGTACATAAAAT	DD 8.5bA	CAGGCAGACTGCACTTAG
DD 2.3A	CAAAGGGATACATAAAAT	DD 8.5bG	CAGGCAGGCTGCACTTAG
DD 2.7G	TGTTTTGGTTCACCGGCC	DD 8.6C	CATGGGGTCTGGCCACTG
DD 2.7A	TGTTTTGATTCACCGGCC	DD 8.6G	CATGGGGTGTGGCCACTG
DD 4.0T	GCAGCTATATTGGAGTCG	DD 8.9C	GTCATCCGCAACATGGTG
DD 4.0C	GCAGCTACATTGGAGTCG	DD 8.9T	GTCATCTGCAACATGGTG
DD 5.7T	TGTTTAATTAAAGGCACA	DD 9.2C	CAGCTCTCAGCAGAGGG
DD 5.7A	TGTTTAAATAAAGGCACA	DD 9.2G	CAGCTCTGAGCAGAGGG
DD 5.7aG	CAATAATGGTGCCTTTAA	DD 10.9G	CCTTTATAGATAGATGCC
DD 5.7aA	CAATAATAGTGCCTTTAA	DD 10.9C	CCTTTATACATAGATGCC
DD 6.6A	ACTTGTAATCATTTATTTG	DD 11.0G	GAAACTGCCTGGATTTTC
DD 6.6G	ACTTGTAGTCATTTATTTG	DD 11.0A	GAAACTACCTGGATTTTC
DD 7.2-	CACACTGCTATATACTTT	DD 11.5G	GTAAAGCGCTTCTCCACTG
DD 7.2G	CACACTGGCTATATACTT	DD 11.5A	GTAAAGCACTTCTCCACTG
DD 7.3C	CTTTTTATTCGCTCKGA	DD 11.9A	GAATAAAAATCTCCTGTCAC

DD 7.3A	CTTTTTATTAGCTCKGA	DD 11.9G	GAATAAAAGTCTCCTGTCAC
DD 7.5G	AAGGGAGGTGCAGAAAGG	DD 12.0C	CATGTATTCTCTGACTTC
DD 7.5A	AAGGGAGATGCAGAAAGG	DD 12.0T	CATGTATTTTCTGACTTC
DD 7.5aC	CAAGTCTCCGCATGACCCA	DD 12.1G	GAATGAGGTTTTCTTAAATC
DD 7.5aT	CAAGTCTCTGCATGACCCA	DD 12.1A	GAATGAAGTTTTCTTAAATC
DD 7.6C	GCTGGTCGTATCTTTATTG	DD 12.4A	GGTATGCATGTCATCTTAC
DD 7.6T	GCTGGTTGTATCTTTATTG	DD 12.4G	GGTATGCGTGTCATCTTAC
DD 7.8A	AACACCAATGCCTTTTGTG	DD 12.6A	CTTTTAAGAACATTACACAG
DD 7.8G	AACACCAGTGCCTTTTGTG	DD 12.6G	CTTTTAAGAGCATTACACAG
DD 8.4G	GGGTGTTGGTTTCTGGCC	DD 13.6G	CTGGGTTGATTCCATGTA
DD 8.4T	GGGTGTTTGTTTCTGGCC	DD 13.6A	CTGGGTTAATTCCATGTA
DD 8.5A	GCTGGGACCCACTGGRC	DD 13.6aG	CATATGTGTGCATGTGTC
DD 8.5G	GCTGGGGCCCACTGGRC	DD 13.6aA	CATATGTATGCATGTGTC

APPENDIX IV RE-SEQUENCING RESULTS

The reference sequence of Hotspot DA is represented as a “Refseq”.

Donor 6

5′ → 3′ sequence for 1st haplotype

```
DONOR  6      ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGG  60
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq 7433   ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 6   61  CTAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCATGTGTTACACG  120
              |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||
Refseq 7493   CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 6   121 GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 6   18  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
              |||||||||||||||| |||||||||||||||||||||||| ||||||||||
Refseq 7613   AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 6 241   TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7673   TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 6   301 TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Refseq  7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 6   361 TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 6   421 TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 6   481 ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 6   541 TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 6   601 TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 6   661 TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 6   721 AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGGNGGGGGA  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |||| |
Refseq  8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT-GGGG-A  8210

DONOR 6   781 GGAATAAAGGGGAAGCCAATAATGAATGTATGTAGTGGATTTGGGGAAGAGGATATGGGA  840
              |||||||||||||||||| |||||||||||||  |||||||||||||||||||||||||
Refseq  8211  GGAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGA  8270

DONOR 6   841 GTTTATTCCATTCTATTCTATCTGNTACNNNNCTGGNNGGTGNAAGCACCTCCA-TTCGA  899
              ||||||||||||||||||||||||| |||    ||||  |||| |||||||||||| |||||
Refseq  8271  GTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTCGA  8330

DONOR 6   900 TTAGGGATATCTGANAACAANN-TCCCTTGANCCTCCCTGTCCTAAGTAACTCC-TAGGA  957
              |||||||||||||| |||||   |||||||| |||||||||||||||| |||| |||||
Refseq  8331  TTAGGGATATCTGAGAACAAAGATCCCTTGAGCCTCCCTGTCCTAAGTAGCTCCCTAGGA  8390
```

```
DONOR 6  958 AAGAGANTAC  967
             ||||||  |||
Refseq  8391 AAGAGACTAC  8400
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 6   1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq 7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 6  61  CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
             |||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq 7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 6 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 6 181  AGACTTATTCACTGTCATCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
             |||||||||||||||||| |||||||||||||||||||||||||| ||||||||||
Refseq 7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 6 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 6 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGTAGGAGAACAATAGTAAGAACACAT  360
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Refseq 7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 6 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 6 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 6 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 6 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 6 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 6 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 6 721  AGAAAAAGATTCCCTGAATCtggggagaaggggatggtggtggtggtganggnnggggna  780
             |||||||||||||||||||||||||||||||||||||||||||||||| ||  |||| |
Refseq 8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT-GGGG-A  8210

DONOR 6 781  ggaataaaggggaagCCAGTAATGAATGTATGAGGTGGATTTGGGGAAGAGGATATGGGA  840
             |||||||||||||||||| ||||||||||||  ||||||||||||||||||||||||||
Refseq 8211  GGAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGA  8270

DONOR 6 841  GTTTATTCNNTTCTNTTCTATCTGTTACTAACCTGGGANGNGAAAGNACCTCCNATTCGA  900
             ||||||||  |||| |||||||||||||||||||| | ||||| ||||||  ||||||
Refseq 8271  GTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTCGA  8330

DONOR 6 901  TTAGGGNATATCTGANAACNA-GANCCCTTGNGNCTNCCTGTCCTA  945
             ||||||  ||||||||  |||  |  ||  ||||||  |  ||  |||||||||
Refseq 8331  TTAGGG-ATATCTGAGAACAAAGATCCCTTGAGCCTCCCTGTCCTA  8375
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 6   1  CTGTGGGAGGT-NNTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  59
             ||||||||||    ||||| ||||||||||||||||||||||||||||| ||||||||||
Refseq 7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 6  60  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 6 120  CCTGCCACCGTGTAACACATGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  179
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq 7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 6 180  CTCCCTAGCCCTGTGGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  239
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq 7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 6 240  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 6 300  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 6 360  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 6 420  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 6 480  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  539
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 6 540  ATGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 6 600  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAGA  659
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Refseq 7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAGA  7021

DONOR 6 660  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7020  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  6961

DONOR 6 720  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 6960  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  6901

DONOR 6 780  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 6900  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  6841

DONOR 6 840  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANANACTNANNAGGGTACA  899
             |||||||||||||||||||||||||||||||||||||||| | ||| |  ||||||||
Refseq 6840  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACA  6781

DONOR 6 900  TGGGANGACAGTCCATGCCCTANAGAAATTGANGAANATAGCTTCTTCCTCATTGTTGGG  959
             ||||| ||||||||||||||| ||||||||| ||| ||||||||||||||||||||||||
Refseq 6780  TGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGG  6721

DONOR 6 960  AGTACAGTNN-TTTTATACCATAATCNTCCCNTTNNCTTCTCA  1001
             ||||||||   ||||||||||||||| |||| ||  |||||||
Refseq 6720  AGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTCA  6678
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 6    1  CTGTGGGANGNAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACAGTGA  60
              |||||||| | |||||||| |||||||||||||||||||||||||||| ||||||||||
Sbjct  7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 6 61   ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 7620   ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 6 121  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  180
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 6 181  CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 6 241  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 6 301  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 6 361  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 6 421  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 6 481  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
             ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 6 541  ATGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 6 601  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGTCCTTTGCTGGGCCCTCTGGACATGCCCA  660
             ||||||||||||||||||||||||||||||||  |||||||||||||||||||||||| ||
Sbjct  7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCA  7023

DONOR 6 661  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7022  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  6963

DONOR 6 721  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6962  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  6903

DONOR 6 781  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6902  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  6843

DONOR 6 841  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTNNN-AGGGTA  899
             |||||||||||||||||||||||||||||||||||||||||||| |||||    ||||||
Sbjct  6842  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTA  6783

DONOR 6 900  CATGGGAGGANAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  959
             |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6782  CATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  6723

DONOR 6 960  GGAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCT  999
             |||||||||||||||||||||||||||||||||||||||
Sbjct  6722  GGAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCT  6683
```

Donor 7

5′ → 3′ sequence for 1st haplotype

```
DONOR 7   1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGG  60
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 7  61  CTAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
             |||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 7 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 7 181  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
             |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Sbjct  7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 7 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 7 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
             |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Sbjct  7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 7 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 7 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 7 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 7 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 7 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 7 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 7 721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGNGGGGGAG  780
             |||||||||||||||||||||||||||||||||||||||||||||||||| |||| ||
Sbjct  8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGTGGGG-AG  8211

DONOR 7 781  GAATAAAGGGGAAGCCAGTAATGAATGTATGANGTGGATTTGGGGAAGANGATATGGGAG  840
             |||||||||||||||| ||||||||||||  ||||||||||||||| ||||||||||
Sbjct  8212  GAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGAG  8271

DONOR 7 841  TTTATTCCATTCTATTCTATCTGTTnnnnnnnCTGGGAGGTGAAANCACCTCCNATTTCG  900
             |||||||||||||||||||||||||       |||||||||||||| ||||||| ||| ||
Sbjct  8272  TTTATTCCATTCTATTCTATCTGTTACTAAC-CTGGGAGGTGAAAGCACCTCCAATT-CG  8329

DONOR 7 901  ATTANGGANNTCTGAGANCAAAGATCCCTTGAGCCTCCCTGTCC  944
             |||| |||  ||||||| ||||||||||||||||||||||||||
Sbjct  8330  ATTAGGGATATCTGAGAACAAAGATCCCTTGAGCCTCCCTGTCC  8373
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 7   1  ACCTGAGANTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGG  60
             ||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq 7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 7  61  CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCATGTGTTACACG  120
             ||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq 7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 7 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 7 181  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
             |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||
Refseq 7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 7 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 7 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Refseq 7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 7 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 7 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 7 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 7 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 7 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 7 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 7 721  AGAAAAAGATTCCCTGAATCtggggagaaggggatggtggtggtggtgatgggtgggggga  780
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |
Refseq 8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGG-TGGGG-A  8210

DONOR 7 781  ngaataaaggggaagCCAGTAATGAATGTATGAGGTGNATTTGGGGAAGAGGATATGGGA  840
              |||||||||||||||| ||||||||||||  ||| ||||||||||||||||||||||
Refseq 8211  GGAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGA  8270

DONOR 7 841  GTTTATTCNNTTCTATTCTATCTGTTACTAACCTGGGANGTGAAAGCACCTCCAATTCGA  900
             ||||||||  |||||||||||||||||||||||||| ||||||||||||||||||||||
Refseq 8271  GTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTCGA  8330

DONOR 7 901  TNANGGATNTCTGANAACANANATCCCNTGANCCTCCCTGTNCTA  945
             | | |||| ||||| |||| | ||||| ||| |||||||||| |||
Refseq 8331  TTAGGGATATCTGAGAACAAAGATCCCTTGAGCCTCCCTGTCCTA  8375
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 7   1  CTGTGGGAGGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  60
             |||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Refseq 7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 7  6   ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 7 121  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  180
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq 7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 7 181  CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq 7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 7 241  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 7 301  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 7 361  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 7 421  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 7 481  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 7 541  CTGCCTTTAGACtttttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCT  600
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7140  MTGCCTTTAGAC-TTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCT  7082

DONOR 7 601  TGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAG  660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Refseq 7081  TGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAG  7022

DONOR 7 661  ACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCA  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7021  ACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCA  6962

DONOR 7 721  CACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGG  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 6961  CACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGG  6902

DONOR 7 781  CACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATT  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 6901  CACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATT  6842

DONOR 7 841  CATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTNNN-AGGGTAC  899
             |||||||||||||||||||||||||||||||||||||||||||||||||    |||||||
Refseq 6841  CATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTAC  6782

DONOR 7 900  ATGGNANGACAGTCCATGCCCTANAGAAATTGANGAAGANAGCTTCTTCCTCATTGTTGG  959
             |||| | |||||||||||||||| ||||||||| ||||| ||||||||||||||||||||
Refseq 6781  ATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGG  6722

DONOR 7 960  GANTACAGTGNNTTT-ATACCATAN-CTTCC-ATTCT  993
             || |||||||  ||| ||||||||  ||||| |||||
Refseq 6721  GAGTACAGTGGCTTTTATACCATAATCTTCCCATTCT  6685
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 7   1  CTGTGGGAGGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  60
             |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Refseq 7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 7 61   ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 7 121  CCTGCCACCGTGTAACACATGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  180
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq 7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 7 181  CTCCCTAGCCCTGTGGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq 7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 7 241  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 7 301  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 7 361  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 7 421  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 7 481  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 7 541  CTGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 7 601  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGTCCTTTGCTGGGCCCTCTGGACATGCTCA  660
             |||||||||||||||||||||||||||||  |||||||||||||||||||||||||| ||
Refseq 7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCA  7023

DONOR 7 661  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7022  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  6963

DONOR 7 721  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 6962  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  6903

DONOR 7 781  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 6902  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  6843

DONOR 7 841  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAANGAGGGTA  900
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Refseq 6842  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTA  6783

DONOR 7 901  CATGGGAGNACAGTCCATGCCCTANAGAAATTGAGGAANANAGCTTCTTCCTCATTGTTG  960
             |||||||| ||||||||||||||| |||||||||||||| | ||||||||||||||||||
Refseq 6782  CATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  6723

DONOR 7 961  GGAGTACAGNNGNTTTTATACCATNATCTTCCCATTCTCTTC  1002
             |||||||||  | |||||||||| ||||||||||||||||
Refseq 6722  GGAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTC  6681
```

Donor 12

5′ → 3′ sequence for 1st haplotype

```
DONOR 12   1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGG  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGG  7492

DONOR 12  61  CTAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
              ||||||||  ||||||||||||||||||||||||||||||||||||  ||||||||||
Refseq  7493  CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCATGTGTTACACG  7552

DONOR 12 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 12 181  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATCATTCAATTACC  240
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Refseq  7613  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  7672

DONOR 12 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq 7673   TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 12 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  7792

DONOR 12 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 12 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 12 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 12 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 12 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 12 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 12 721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGANNGNNGGGGGA  780
              ||||||||||||||||||||||||||||||||||||||||||||||||  |  |||| |
Refseq  8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT-GGGG-A  8210

DONOR 12 781  GGAATAAAGGGGAAGCCAATAATGAATGTATGTAG-TGGANTTGGGGAAGAGGATATGGG  839
              ||||||||||||||||||  ||||||||||||  ||  ||||  ||||||||||||||||||||
Refseq  8211  GGAATAAAGGGGAAGCCAGTAATGAATGTATG-AGGTGGATTTGGGGAAGAGGATATGGG  8269

DONOR 12 840  AGTTTATNCCATTCTATTCTATCTGTT  866
              |||||||  ||||||||||||||||||
Refseq  8270  AGTTTATTCCATTCTATTCTATCTGTT  8296
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 12   1  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGGC  60
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq  7434  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGGC  7493

DONOR 12  61  TAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACGG  120
              ||||||| ||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq  7494  TAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACGG  7553

DONOR 12 121  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7554  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  7613

DONOR 12 181  GACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACCT  240
              |||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Refseq  7614  GACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACCT  7673

DONOR 12 241  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7674  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  7733

DONOR 12 301  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACATT  360
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Refseq  7734  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACATT  7793

DONOR 12 361  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7794  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  7853

DONOR 12 421  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7854  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  7913

DONOR 12 481  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7914  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  7973

DONOR 12 541  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7974  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  8033

DONOR 12 601  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8034  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  8093

DONOR 12 661  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8094  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  8153

DONOR 12 721  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGANNGNNGGGGGAG  780
              ||||||||||||||||||||||||||||||||||||||||||||||||  |  |||| ||
Refseq  8154  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT-GGGG-AG  8211

DONOR 12 781  GAATAAAGGGGAANGCCAGTAATGAATGTATGANGTGGATTTGGGGAAGAGGATATGGGA  840
              |||||||||||||| |||| ||||||||||||||  |||||||||||||||||||||||
Refseq  8212  GAATAAAGGGGAA-GCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGA  8270

DONOR 12 841  GTTTATTCCATTCCNATTCTATCNGNTACTAACCNGGGAGG  881
              |||||||||||||  |||||||| | |||||||| ||||||
Refseq  8271  GTTTATTCCATTCT-ATTCTATCTGTTACTAACCTGGGAGG  8310
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 12    1  TGTGGGAGGTAATTGAATGATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGAA  60
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq  7679  TGTGGGAGGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGAA  7620

DONOR 12   61  TAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTGC  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7619  TAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTGC  7560

DONOR 12 121  CTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGCC  180
               |||||||||||||||| |||||||||||||||||||||||||||||||||||||| |
Refseq  7559  CTGCCACCGTGTAACACATGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGAC  7500

DONOR 12 181  TCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAGT  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7499  TCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAGT  7440

DONOR 12 241  CTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCTA  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7439  CTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCTA  7380

DONOR 12 301  AACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGCT  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7379  AACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGCT  7320

DONOR 12 361  CAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCCC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7319  CAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCCC  7260

DONOR 12 421  CAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTAT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7259  CAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTAT  7200

DONOR 12 481  GAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCAC  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7199  GAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCAA  7140

DONOR 12 541  TGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTTG  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7139  TGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTTG  7080

DONOR 12 601  CAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAGAC  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7079  CAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAGAC  7020

DONOR 12 661  AGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCACA  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7019  AGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCACA  6960

DONOR 12 721  CTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGCA  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6959  CTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGCA  6900

DONOR 12 781  CTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTCA  840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6899  CTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTCA  6840

DONOR 12 841  TAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTANN-AGGGTACAT  899
               ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Refseq  6839  TAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACAT  6780

DONOR 12 900  GGGAGGACAGTCCATGCCCTANAGAAATTGANGAANATAGCTTCTTCCTCATTGTTGGGA  959
               |||||||||||||||||||| ||||||||| ||| ||||||||||||||||||||||||
Refseq  6779  GGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGGA  6720

DONOR 12 960  GTACAGTGGNTTTTATACCATAATCTTCCCATTCTCTTCTCnnanaanCC-TTTNGCAAG  1018
               ||||||||| |||||||||||||||||||||||||||||||  | || || ||| |||||
Refseq  6719  GTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTCAGATAAACCCTTTGGCAAG  6660

DONOR 121019  TATNANTTTTAANTANTTCNTTGAAAN-TGCTGA  1051
               ||| | |||||| || ||| ||||||  ||||||
Refseq  6659  TATAATTTTTAAATATTTCTTTGAAAACTGCTGA  6626
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 12   1  CTGTGGGAGGT-NNTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  59
              ||||||||||||   ||||| |||||||||||||||||||||||||||| ||||||||||
Refseq  7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 12  60  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 12 120  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  179
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq  7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 12 180  CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  239
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq  7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 12 240  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 12 300  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 12 360  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 12 420  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 12 480  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  539
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 12 540  CTGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 12 600  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGTCCTTTGCTGGGCCCTCTGGACATGCTCA  659
              |||||||||||||||||||||||||||||  |||||||||||||||||||||||||| ||
Refseq  7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCA  7023

DONOR 12 660  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7022  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  6963

DONOR 12 720  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6962  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  6903

DONOR 12 780  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6902  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  6843

DONOR 12 840  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTNANNAGGGTA  899
              |||||||||||||||||||||||||||||||||||||||||||| ||||| |  ||||||
Refseq  6842  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTA  6783

DONOR 12 900  CATGGGAGGACAGTCCATGCCCTANAGAAATTGAGnaananaGCTTCTTCCTCATTGNTG  959
              |||||||||||||||||||||||| |||||||||| || | |||||||||||||||| ||
Refseq  6782  CATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  6723

DONOR 12 960  GGAGTACAGTNNNTTT-ATACCATAATCTTCCCATTCTCTTCTC  1002
              ||||||||||   ||| |||||||||||||||||||||||||
Refseq  6722  GGAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTC  6679
```

Donor 44

5′ → 3′ sequence for 1st haplotype

```
DONOR 44   1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGG  60
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq  7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 44  61  CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
              |||||||| |||||||||||||||||||||||||||||||||| ||||||||||
Refseq  7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 44 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 44 181  AGACTTATTCACTGTCATCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
              |||||||||||||||||| |||||||||||||||||||||||| ||||||||||
Refseq  7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 44 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 44 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Refseq  7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 44 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 44 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 44 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 44 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 44 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 44 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 44 721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGTNGGGGGA  780
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Refseq  8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT-GGGG-A  8210

DONOR 44 781  GGAAT  785
              |||||
Refseq  8211  GGAAT  8215
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 44   1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGG  60
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq  7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 44  61  CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
              |||||||| ||||||||||||||||||||||||||||||||||| ||||||||||
Refseq  7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 44 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 44 181  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
              |||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Refseq  7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 44 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 44 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Refseq  7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 44 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 44 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 44 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 44 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 44 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 44 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 44 721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGGTGGGGAG  780
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Refseq  8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGATGG-TGGGGAG  8211

DONOR 44 781  GAATAAAGGGGAAGCCA  797
              |||||||||||||||||
Refseq  8212  GAATAAAGGGGAAGCCA  8228
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 44    1  CTGTGGGAGGNAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACAGTGA  60
               |||||||||| ||||||||  ||||||||||||||||||||||||||||| ||||||||||
Refseq  7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 44   61  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 44 121  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  180
               |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Refseq  7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 44 181  CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
               |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Refseq  7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 44 241  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 44 301  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 44 361  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 44 421  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 44 481  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
               ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 44 541  ATGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 44 601  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGTCCTTTGCTGGTCCCTCTGGACATGCTCA  660
               |||||||||||||||||||||||||||||||  |||||||||  ||||||||||||||| ||
Refseq  7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCA  7023

DONOR 44 661  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7022  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  6963

DONOR 44 721  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6962  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  6903

DONOR 44 781  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6902  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  6843

DONOR 44 841  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANANACTNANNAGGGTA  900
               |||||||||||||||||||||||||||||||||||||||||||| | ||| |  ||||||
Refseq  6842  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTA  6783

DONOR 44 901  CATGGGAGNACAGTCCATGCCCTANAGAAATTGAGGAAGANAGCTTCTTCCTCATTGTTG  960
               ||||||||  ||||||||||||||  ||||||||||||||  ||||||||||||||||||
Refseq  6782  CATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  6723

DONOR 44 961  GGAGTACAGTGN-TTTTATACCATAATCTTCCCATTCNNCTNNTCNNANAA-CCCTTTGG  1018
               ||||||||||||  |||||||||||||||||||||||||  ||  ||  | || ||||||||
Refseq  6722  GGAGTACAGTGGCTTTTATACCATAATCTTCCCATTCT-CTTCTCAGATAAACCCTTTGG  6664

DONOR 441019  CAAGNATANTTTT-AANNATTTCTTTGAAAACTG  1051
               ||||  |||  ||||  ||  ||||||||||||||||
Refseq  6663  CAAGTATAATTTTTAAATATTTCTTTGAAAACTG  6630
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 44    1  CTGTGGGANGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  60
               |||||||| |||||||||| |||||||||||||||||||||||||||| ||||||||||
Refseq  7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 44   61  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 44 121  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  180
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq  7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 44 181  CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq  7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 44 241  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 44 301  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 44 361  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 44 421  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 44 481  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 44 541  CTGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 44 601  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGTCCTTTGCTGGGCCCTCTGGACATGCTCA  660
               |||||||||||||||||||||||||||||||  |||||||||||||||||||||||| ||
Refseq  7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCA  7023

DONOR 44661   GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7022  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  6963

DONOR 44 721  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6962  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  6903

DONOR 44 781  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6902  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  6843

DONOR 44 841  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGANACTAANNAGGGTA  900
               |||||||||||||||||||||||||||||||||||||||||||||| |||||  ||||||
Refseq  6842  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTA  6783

DONOR 44 901  CATGGGAGGACAGTCCATGCCCTANAGAAATTGAGGAANATAGCTTCTTCCTCATTGNNG  960
               |||||||||||||||||||||||| |||||||||||| |||||||||||||||||  |
Refseq  6782  CATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  6723

DONOR 44 961  GGAGTACAGNGGNTTTTATACCATAN-CTTCCCATTCTCTTCT  1002
               ||||||||| || ||||||||||||  |||||||||||||||||
Refseq  6722  GGAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCT  6680
```

Donor 67

5′ → 3′ sequence for 1st haplotype

```
DONOR 67   1  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGGC  60
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq  7434  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGGC  7493

DONOR 67  61  TAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACGG  120
              ||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq  7494  TAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACGG  7553

DONOR 67 121  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7554  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  7613

DONOR 67 181  GACTTATTCACTGTCATCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACCT  240
              ||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Refseq  7614  GACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACCT  7673

DONOR 67 241  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7674  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  7733

DONOR 67 301  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACATT  360
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Refseq  7734  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACATT  7793

DONOR 67 361  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7794  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  7853

DONOR 67 421  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7854  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  7913

DONOR 67 481  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7914  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  7973

DONOR 67 541  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7974  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  8033

DONOR 67 601  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8034  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  8093

DONOR 67 661  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8094  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  8153

DONOR 67 721  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGGNNGGGGAG  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Refseq  8154  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT--GGGGAG  8211

DONOR 67 781  GAATAAAGGGGAAGCCAATAATGAATGNNNGTAGTGGATTTGGGGGAAGAGGATATGGGA  840
              |||||||||||||||| |||||||||   |  |||||||||||| ||||||||||||||
Refseq  8212  GAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGG-AAGAGGATATGGGA  8270

DONOR 67 841  GTTTATTCCATTNCNATTCNATCTGTTACTNACCTGGGAGGTGA  884
              |||||||||||| | |||| |||||||||| ||||||||||||
Refseq  8271  GTTTATTCCATT-CTATTCTATCTGTTACTAACCTGGGAGGTGA  8313
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 67   1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGG  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq  7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 67  61  CTAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACATG  120
              |||||||| |||||||||||||||||||||||||||||||||| |||||||| |
Refseq  7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 67 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 67 181  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
              |||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Refseq  7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 67 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 67 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Refseq  7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 67 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 67 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 67 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 67 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 67 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 67 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 67 721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGGTGGGGAG  780
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Refseq  8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGG-TGGGGAG  8211

DONOR 67 781  GAATAAAGGGGAAGCCAGTAATGAATGTATGAGGTGGATTTGGGGAAGAGGATATGGGAG  840
              |||||||||||||||| |||||||||||||  ||||||||||||||||||||||||||||
Refseq  8212  GAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGAG  8271

DONOR 67 841  TTTATTCCATTCNATTCNATCNGNTACTAACNNGGGAGGTGAAAGCACTTCAAATT  896
              |||||||||||| |||| ||| | |||||||  ||||||||||||||| || ||||
Refseq  8272  TTTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATT  8327
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 67  1    CTGTGGGAGGNAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACAGTGA  60
               |||||||||| |||||||| |||||||||||||||||||||||| ||||||||||
Refseq 7680    CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 67  61   ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7620   ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 67 121   CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  180
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq  7560   CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 67 181   CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq  7500   CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 67 241   TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7440   TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 67 301   AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7380   AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 67 361   TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7320   TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 67 421   CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7260   CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 67 481   TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
               |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7200   TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 67 541   ATGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7140   MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 67 601   GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAGA  660
               |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Refseq  7080   GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAGA  7021

DONOR 67 661   CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7020   CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  6961

DONOR 67 721   ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6960   ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  6901

DONOR 67 781   ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6900   ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  6841

DONOR 67 841   ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTAANNAGGGTACA  900
               |||||||||||||||||||||||||||||||||||||||||| |||||||  ||||||||
Refseq  6840   ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACA  6781

DONOR 67 901   TGGGANGACAGTCCATGCCCTANAGAAATTGAGGAANANAGCTTCTTCCTCATTGTTGGG  960
               ||||| ||||||||||||||| |||||||||||| | ||||||||||||||||||||||
Refseq  6780   TGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGG  6721

DONOR 67 961   AGTACAGNGGNTTTTNATACCATAATCNTCCCATTCTCTTCTCNNATAANCCCTTTGGNA  1020
               ||||||| || |||| |||||||||| |||||||||||||  |||| |||||||| |
Refseq  6720   AGTACAGTGGCTTTT-ATACCATAATCTTCCCATTCTCTTCTCAGATAAACCCTTTGGCA  6662

DONOR 67 1021 AGTA  1024
               ||||
Refseq  6661   AGTA  6658
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 67   1  CTGTGGGAGG-NNNTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  59
              ||||||||||    ||||| ||||||||||||||||||||||||||||| ||||||||||
Refseq  7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 67  60  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 67 120  CCTGCCACCATGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  179
              ||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq  7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 67 180  CTCCCTAGCCCTGTGGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  239
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq  7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 67 240  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 67 300  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 67 360  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 67 420  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 67 480  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  539
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 67 540  CTGCCTTTAGACtttttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCT  599
               |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7140  MTGCCTTTAGAC-TTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCT  7082

DONOR 67 600  TGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAG  659
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Refseq  7081  TGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAG  7022

DONOR 67 660  ACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCA  719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7021  ACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCA  6962

DONOR 67 720  CACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGG  779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6961  CACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGG  6902

DONOR 67 780  CACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATT  839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6901  CACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATT  6842

DONOR 67 840  CATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTNNGGAGGGTAC  899
              |||||||||||||||||||||||||||||||||||||||||| |||||  |||||||||
Refseq  6841  CATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTAC  6782

DONOR 67 900  ATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGANAGCTTCTTCCTCATTGTTGG  959
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Refseq  6781  ATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGG  6722

DONOR 67 960  GNAGTACANNNN-TTTTATACCATNATCTTCCCATTCNCTTCTCAGANAAACCCTTTGGN  1018
              | ||||||     |||||||||||| |||||||||||| ||||||||| ||||||||||||
Refseq  6721  G-AGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTCAGATAAACCCTTTGGC  6663

DONOR 67 1019 AAGTATANTTT--AANTATTTCNTTGAAA-CTGCTGAACCTGAC  1059
              ||||||| |||  || |||||| |||||| ||||||||||||||
Refseq  6662  AAGTATAATTTTTAAATATTTCTTTGAAAACTGCTGAACCTGAC  6619
```

Donor 73

5′ → 3′ sequence for 1st haplotype

```
DONOR 73 1    ACCTGAGANTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGG  60
              |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq  7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 73  61  CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
              |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||
Refseq  7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 73 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 73 181  AGACTTATTCACTGTCATCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
              |||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Refseq  7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 73 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 73 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Refseq  7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 73 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 73 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 73 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 73 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 73 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 73 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 73 721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGANGNNNNGNGGA  780
              |||||||||||||||||||||||||||||||||||||||||||||||| |    | |||
Refseq  8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGTG-G-GGA  8210

DONOR 73 781  GGAATAAAGGGGAAGCCAGTAATGAATGTATGAGGTGGATTTGGGGGAAGAGGATATGGG  840
              |||||||||||||||| |||||||||||  |||||||||||| ||||||||||||||
Refseq  8211  GGAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGG-AAGAGGATATGGG  8269

DONOR 73 841  AGTTTATTCCATTCTATTCTATCTGTTACNAACCNTGGGAGGTGA  885
              |||||||||||||||||||||||||||||| |||| ||||||||||
Refseq  8270  AGTTTATTCCATTCTATTCTATCTGTTACTAACC-TGGGAGGTGA  8313
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 73  1    ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGG  60
               |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq  7433   ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 73  61   CTAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
               |||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq  7493   CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 73 121   GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7553   GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 73 181   AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATCATTCAATTACC  240
               |||||||||||||||| |||||||||||||||||||||||| ||||||||||
Refseq  7613   AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 73 241   TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7673   TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 73 301   TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
               |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Refseq  7733   TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 73 361   TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7793   TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 73 421   TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7853   TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 73 481   AATAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7913   ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 73 541   TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7973   TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 73 601   TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8033   TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 73 661   TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8093   TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 73 721   AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGANNNNNGGGGGA  780
               ||||||||||||||||||||||||||||||||||||||||||||||||     |||| |
Refseq  8153   AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT-GGGG-A  8210

DONOR 73 781   GGAATAAAGGGGAAGCCAGTAATGAATGTATGAGGTGGATTTGGGGAAGAGGATATGGGA  840
               |||||||||||||||||| ||||||||||||  ||||||||||||||||||||||||||
Refseq  8211   GGAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGA  8270

DONOR 73 841   GTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGNGGTGAAAGCACCTCCAATTCNA  900
               |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||| |
Refseq  8271   GTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTCGA  8330

DONOR 73 901   TTANGGATATCTGANAACAAAGATCCCTTGANCCNTCCNTGTCCTANGTAGCTCCCTA  958
               ||| |||||||||| |||||||||||||||| || ||| |||||||| ||||||||||
Refseq  8331   TTAGGGATATCTGAGAACAAAGATCCCTTGAGCC-TCCCTGTCCTAAGTAGCTCCCTA  8387
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 73   1  CTGTGGGAGG-NNNTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACAGTGA  59
              ||||||||||    ||||| ||||||||||||||||||||||||||||| ||||||||||
Refseq  7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 73  60  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 73 120  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  179
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq  7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 73 180  CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  239
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq  7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 73 240  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 73 300  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 73 360  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 73 420  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 73 480  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  539
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 73 540  ATGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  599
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 73 600  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGTCCTTTGCTGGGCCCTCTGGACATGCCCA  659
              |||||||||||||||||||||||||||||  |||||||||||||||||||||||||| ||
Refseq  7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCA  7023

DONOR 73 660  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7022  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  6963

DONOR 73 720  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6962  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  6903

DONOR 73 780  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6902  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  6843

DONOR 73 840  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTNANNAGGGTA  899
              |||||||||||||||||||||||||||||||||||||||||||| ||||| |  ||||||
Refseq  6842  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTA  6783

DONOR 73 900  CATGGGAGGACAGTCCATGCCCTANAGAAATTGAGNAANATAGCTTCTTCCTCATTGTTG  959
              |||||||||||||||||||||||| |||||||||| || ||||||||||||||||||||
Refseq  6782  CATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  6723

DONOR 73 960  GGAGTACAGTGGNTTTTATACCATAATCTTCCCATTCTCTTCTC  1003
              |||||||||||| |||||||||||||||||||||||||||||
Refseq  6722  GGAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTC  6679
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 73   1  CTGTGGGAGGT-NNTGAATGATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  59
              ||||||||||    ||||| ||||||||||||||||||||||||||| ||||||||||
Refseq  7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 73  60  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 73 120  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  179
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq  7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 73 180  CTCCCTAGCCCTGTGGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  239
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq  7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 73 240  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 73 300  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 73 360  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 73 420  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 73 480  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  539
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 73 540  CTGCCTTTAGACtttttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCT  599
               |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7140  MTGCCTTTAGAC-TTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCT  7082

DONOR 73 600  TGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAG  659
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Refseq  7081  TGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAG  7022

DONOR 73 660  ACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCA  719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7021  ACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCA  6962

DONOR 73 720  CACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGG  779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6961  CACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGG  6902

DONOR 73 780  CACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATT  839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  6901  CACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATT  6842

DONOR 73 840  CATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTNNN-AGGGTAC  898
              |||||||||||||||||||||||||||||||||||||||||| |||||    |||||||
Refseq  6841  CATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTAC  6782

DONOR 73 899  ATGGGAGGACAGTCCATGCCCTANAGAAATTGAGNAANATAGCTTCTTCCTCATTGNTGG  958
              ||||||||||||||||||||||| |||||||||| || |||||||||||||||||| |||
Refseq  6781  ATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGG  6722

DONOR 73 959  GAGTACAGTGN-TTTTATACCANNANCTTCCCATTCNCTTCTCNGANAA-CCCTTTNGCA  1016
              |||||||||| |||||||||| | |||||||||| |||||| || || |||||| |||
Refseq  6721  GAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTCAGATAAACCCTTTGGCA  6662

DONOR 73 1017 AGTA  1020
              ||||
Refseq  6661  AGTA  6658
```

Donor 180

5′ → 3′ sequence for 1st haplotype

```
DONOR 180   1  CCTGANACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGGC  60
               |||||  ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7434  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGGC  7493

DONOR 180  61  TAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACGG  120
               |||||||  |||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7494  TAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACGG  7553

DONOR 180 121  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7554  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  7613

DONOR 180 181  GACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACCT  240
               ||||||||||||||||  |||||||||||||||||||||||||||| ||||||||||||
Refseq   7614  GACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACCT  7673

DONOR 180 241  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7674  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  7733

DONOR 180 301  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACATT  360
               ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Refseq   7734  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACATT  7793

DONOR 180 361  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7794  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  7853

DONOR 180 421  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7854  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  7913

DONOR 180 481  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7914  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  7973

DONOR 180 541  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7974  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  8033

DONOR 180 601  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8034  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  8093

DONOR 180 661  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8094  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  8153

DONOR 180 721  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGGTGGGGAGG  780
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Refseq   8154  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGG-TGGGGAGG  8212

DONOR 180 781  AATAAAGGGGAAGCCAATAATGAATGTATGTAGTGGATTTGGGGAAGAGGATATGGGAGT  840
               ||||||||||||||||  ||||||||||||  |||||||||||||||||||||||||||||
Refseq   8213  AATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGAGT  8272

DONOR 180 841  TTATTCCATTCTATTCTATCTGNTNNTNNCCTGGGAGGTnnannnnCCTCCAATTCGATT  900
               ||||||||||||||||||||||  |  |  ||||||||||  |    |||||||||||||
Refseq   8273  TTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTCGATT  8332

DONOR 180 901  NAGGGATATCTGanannaaaGATCCCTTGAGCCNTCCCTGTCC  943
               ||||||||||||  |  ||||||||||||||||  |||||||||
Refseq   8333  -AGGGATATCTGAGAACAAAGATCCCTTGAGCC-TCCCTGTCC  8373
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 180  1  CCTGAGANTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGGC  60
              ||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq  7434  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGGC  7493

DONOR 180 61  TAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCATGTGTTACACGG  120
              ||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq  7494  TAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACGG  7553

DONOR 180 121  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7554  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  7613

DONOR 180 181  GACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACCT  240
               |||||||||||||||| |||||||||||||||||||||||||| ||||||||||||
Refseq   7614  GACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACCT  7673

DONOR 180 241  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7674  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  7733

DONOR 180 301  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACATT  360
               |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Refseq   7734  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACATT  7793

DONOR 180 361  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7794  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  7853

DONOR 180 421  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7854  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  7913

DONOR 180 481  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7914  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  7973

DONOR 180 541  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7974  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  8033

DONOR 180 601  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8034  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  8093

DONOR 180 661  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8094  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  8153

DONOR 180 721  GAAAAAGATTCCCTGAATCtggggagaaggggatggtggtggtggtgatgggtgggggag  780
               ||||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||
Refseq   8154  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGG-TGGGG-AG  8211

DONOR 180 781  gaataaaggggaAGCCAGTAATGAATGTATGAGGTGGATTTGGGGAAGAGGATATGGGAG  840
               |||||||||||||||| ||||||||||||  ||||||||||||||||||||||||||
Refseq   8212  GAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGAG  8271

DONOR 180 841  TTTATTCCCATTCTATTCTATCTGTNNNNANCCTGGGAGGTGAAANCACCTCCAATTC  898
               |||||||| ||||||||||||||||    | |||||||||||||| ||||||||||||
Refseq   8272  TTTATTCC-ATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTC  8328
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 180   3  CTGTGGGAGGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  62
               |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Refseq   7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 180  63  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  122
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 180 123  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  182
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq   7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 180 183  CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  242
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq   7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 180 243  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  302
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 180 303  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  362
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 180 363  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  422
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 180 423  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  482
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 180 483  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  542
               |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 180 543  ATGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  602
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 180 603  GCAGGCTCCTGGGTCCCANNGGCAAAGAGGGCTTTGCTGGGCCCCTCTGGACATGCCCAG  662
               ||||||||||||||||||  |||||||||||||||||||||| |||||||||||| |||
Refseq   7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCC-TCTGGACATGCYCAG  7022

DONOR 180 663  ACAGGTTTAAATCC  676
               |||||||||||||
Refseq   7021  ACAGGTTTAAATCC  7008
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 180   1  TGTGGGANGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGAA  60
               ||||||| |||||||||| |||||||||||||||||||||||||||| ||||||||||
Refseq   7679  TGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGAA  7620

DONOR 180  61  TAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTGC  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7619  TAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTGC  7560

DONOR 180 121  CTGCCACCGTGTAACACATGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGAC  180
               |||||||||||||||| |||||||||||||||||||||||||||||||||||||||| |
Refseq   7559  CTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGMC  7500

DONOR 180 181  TCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAGT  240
               |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Refseq   7499  TCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAGT  7440

DONOR 180 241  CTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCTA  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7439  CTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCTA  7380

DONOR 180 301  AACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGCT  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7379  AACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGCT  7320

DONOR 180 361  CAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCCC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7319  CAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCCC  7260

DONOR 180 421  CAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTAT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7259  CAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTAT  7200

DONOR 180 481  GAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCAC  540
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7199  GAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCAM  7140

DONOR 180 541  TGCCTTTAGACtttttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
               |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7139  TGCCTTTAGAC-TTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 180 601  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAGA  660
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Refseq   7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAGA  7021

DONOR 180 661  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7020  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  6961

DONOR 180 721  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6960  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  6901

DONOR 180 781  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6900  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  6841

DONOR 180 841  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANANACTNNNGAGGGTACA  900
               |||||||||||||||||||||||||||||||||||||||||| | |||   |||||||||
Refseq   6840  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACA  6781

DONOR 180 901  TGGGGAGGACAGTCCATGCCCTANAGAAATTGAGGAANANAGCTTCTTCCTCATTGTTGN  960
               |||| |||||||||||||||| |||||||||||||| | ||||||||||||||||||
Refseq   6780  TGGG-AGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGG  6722

DONOR 180 961   -AGTACAGTNNNTTT-ATACCATAATCTTCCCATTC  994
               ||||||||   ||| ||||||||||||||||||||
Refseq   6721  GAGTACAGTGGCTTTTATACCATAATCTTCCCATTC  6686
```

Donor 185

5′ → 3′ sequence for 1st haplotype

```
DONOR 185  16  TTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGGCTAGGGAGTCCTCACTA  75
               ||||||||||||||||||||||||||||||| |||||||||||||||||| ||||||||
Refseq   7450  TTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGGCTAGGGAGKCCTCACTA  7509

DONOR 185  76  TCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACGGTGGCAGGCAAGAGGGT  135
               |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Refseq   7510  TCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACGGTGGCAGGCAAGAGGGT  7569

DONOR 185 136  GTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGAGACTTATTCACTGTCA  195
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7570  GTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGAGACTTATTCACTGTCA  7629

DONOR 185 196  TCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACCTCCCACAGAGTCCCTCC  255
               | |||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Refseq   7630  TMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACCTCCCACAGAGTCCCTCC  7689

DONOR 185 256  CATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATTTGGGTGGGGACACAGC  315
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7690  CATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATTTGGGTGGGGACACAGC  7749

DONOR 185 316  CAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACATTTTAGAATTCCCCAAGG  375
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq   7750  CAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACATTTTAGAATTCCCCAAGG  7809

DONOR 185 376  ATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCTTGAAATTGCTAGTTGC  435
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7810  ATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCTTGAAATTGCTAGTTGC  7869

DONOR 185 436  AACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATATTAACAGGAATCATAA  495
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7870  AACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATATTAACAGGAATCATAA  7929

DONOR 185 496  TTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTTTTGTTTCAAGCCCAGA  555
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7930  TTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTTTTGTTTCAAGCCCAGA  7989

DONOR 185 556  CCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATTACAACAGCAGAATAAA  615
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7990  CCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATTACAACAGCAGAATAAA  8049

DONOR 185 616  GGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCATAAACAGTAGATTTGCA  675
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8050  GGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCATAAACAGTAGATTTGCA  8109

DONOR 185 676  TTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTAGAAAAAGATTCCCTGA  735
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8110  TTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTAGAAAAAGATTCCCTGA  8169

DONOR 185 736  ATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGG  770
               |||||||||||||||||||||||||||||||||
Refseq   8170  ATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGG  8204
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 185    1  ACCTGAGANTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGG  60
                ||||||||  ||||||||||||||||||||||||||||||||||||  ||||||||||
Refseq    7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 185   61  CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
                ||||||||  ||||||||||||||||||||||||||||||||||||  ||||||||||
Refseq    7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 185  121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 185  181  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
                ||||||||||||||||||  ||||||||||||||||||||||||||||||  ||||||||||
Refseq    7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 185  241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 185  301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
                ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Refseq    7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 185  361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 185  421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 185  481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 185  541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 185  601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 185  661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 185  721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT  773
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT  8205
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 185  1    CTGTGGGAGGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACAGTGA  60
                |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Refseq  7680    CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 185  61   ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7620   ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 185 121   CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  180
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq   7560   CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 185 181   CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq   7500   CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 185 241   TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7440   TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 185 301   AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7380   AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 185 361   TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7320   TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 185 421   CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7260   CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 185 481   TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
                |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7200   TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 185 541   ATGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7140   MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 185 601   GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGTCCTTTGCTGGGCCCTCTGGACATGCCCA  660
                |||||||||||||||||||||||||||||||  |||||||||||||||||||||||| ||
Refseq   7080   GCAGGCTCCTGGGTCCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCA  7023

DONOR 185 661   GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCANGCTGAGACCCACAACCTC  720
                |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Refseq   7022   GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  6963

DONOR 185 721   ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6962   ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  6903

DONOR 185 781   GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6902   GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  6843

DONOR 185 841   TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCNNAGACTAANGAGGGTA  900
                |||||||||||||||||||||||||||||||||||||||||  ||||||| |||||||
Refseq   6842   TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTA  6783

DONOR 185 901   CATGGGAGGACAGTCCATGCCCTananaaaTTGAGGAAGANAGCTTCTTCCTCATTGTTG  960
                |||||||||||||||||||||||| | |||||||||||||| ||||||||||||||||||
Refseq   6782   CATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  6723

DONOR 185 961   GGNGTACAGTNN-TTT-ATACCATAAT  985
                || |||||||   ||| ||||||||||
Refseq   6722   GGAGTACAGTGGCTTTTATACCATAAT  6696
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 185    1  CTGTGGGAGGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  60
                ||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Refseq    7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 185   61  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 185  121  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  180
                ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq    7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 185  181  CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
                ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq    7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 185  241  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 185  301  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 185 361  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 185 421  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 185 481  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
                |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 185 541  ATGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 185 601  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCCCAGA  660
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Refseq    7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAGA  7021

DONOR 185 661  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7020  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  6961

DONOR 185 721  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    6960  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  6901

DONOR 185 781  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    6900  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  6841

DONOR 185 841  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTNANNAGGGTACA  900
                |||||||||||||||||||||||||||||||||||||||||||||||| |  ||||||||
Refseq    6840  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACA  6781

DONOR 185 901  TGGGANGACAGTCCATGCCCTANAGAAATTGAGNAANATAGCTTCTTCCTCATTGNTGGG  960
                ||||| ||||||||||||||| |||||||||| || |||||||||||||||||| ||||
Refseq    6780  TGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGG  6721

DONOR 185 961  AGTACAGTGN-TTTTATACCATAN-CTTCCCATTCNCTTCTC  1000
                |||||||||  ||||||||||||  |||||||||| ||||||
Refseq    6720  AGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTC  6679
```

Donor 211

5′ → 3′ sequence for 1st haplotype

```
DONOR 211   1  ACCTGAGANTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGG  60
               |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||
Refseq   7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 211  61  CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
               |||||||| |||||||||||||||||||||||||||||||||| ||||||||||
Refseq   7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 211 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 211 181  AGACTTATTCACTGTCATCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
               |||||||||||||||||| |||||||||||||||||||||||| ||||||||||
Refseq   7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 211 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 211 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
               |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Refseq   7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 211 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 211 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 211 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 211 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 211 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 211 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 211 721  AGAAAAAGATTCCCTGAATCtggggagaaggggatggtggtggtggtgatgggtgggggga  780
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |
Refseq   8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGG-TGGGG-A  8210

DONOR 211 781  ggaataaaggggaAGCCAGTAATGAATGTATGAGGTGNATTTGGGGAAGAGGATATGGGA  840
               |||||||||||||||| ||||||||||||  ||| ||||||||||||||||||||||
Refseq   8211  GGAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGA  8270

DONOR 211 841  GTTTATTCCATNCTATTCTATCTGNTACTAACCTGGNNGN-GAAAGCNCCTCCAATTCGA  899
               |||||||||||| |||||||||||| |||||||||||  |  |||||| ||||||||||||
Refseq   8271  GTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTCGA  8330

DONOR 211 900  TTAGGNNTATCTGANANCAAAGATCCCTTGAGCCTCCCTGTC-TAA  944
               |||||  ||||||| | |||||||||||||||||||||||| |||
Refseq   8331  TTAGGGATATCTGAGAACAAAGATCCCTTGAGCCTCCCTGTCCTAA  8376
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 211    1  TGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGGCTAGGGAGG  60
                ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Refseq    7442  TGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGGCTAGGGAGK  7501

DONOR 211   61  CCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACGGTGGCAGGC  120
                |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Refseq    7502  CCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACGGTGGCAGGC  7561

DONOR 211  121  AAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGAGACTTATT  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7562  AAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGAGACTTATT  7621

DONOR 211  181  CACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACCTCCCACAGA  240
                ||||||||| |||||||||||||||||||||||||||| ||||||||||||||||||||
Refseq    7622  CACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACCTCCCACAGA  7681

DONOR 211  241  GTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATTTGGGTGGG  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7682  GTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATTTGGGTGGG  7741

DONOR 211  301  GACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACATTTTAGAATT  360
                |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Refseq     7742 GACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACATTTTAGAATT  7801

DONOR 211  361  CCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCTTGAAATTG  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7802  CCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCTTGAAATTG  7861

DONOR 211  421  CTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATATTAACAGG  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7862    CTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATATTAACAGG  7921

DONOR 211  481  AATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTTTTGTTTCA  540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7922  AATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTTTTGTTTCA  7981

DONOR 211  541  AGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATTACAACAGC  600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7982  AGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATTACAACAGC  8041

DONOR 211  601  AGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCATAAACAGTA  660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    8042  AGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCATAAACAGTA  8101

DONOR 211  661  GATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTAGAAAAAGA  720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    8102  GATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTAGAAAAAGA  8161

DONOR 211  721  TTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGGTNNGGGGAGGAATAAA  780
                |||||||||||||||||||||||||||||||||||||||||||| |  ||||||||||||
Refseq    8162  TTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGG-T--GGGGAGGAATAAA  8218

DONOR 211  781  GGGGAANCCNNTAATGAATGTATGAGGTGGATTTGGGGAAGAGGATATGGGAGTTNNTTC  840
                |||||| ||  ||||||||||||  |||||||||||||||||||||||||||||  |||
Refseq    8219  GGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGAGTTTATTC  8278

DONOR 211  841  CATTCTATTCNATCTGNTACTAACNNGGGANGTGAAA  877
                |||||||||| ||||| |||||||  |||| ||||||
Refseq    8279  CATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAA  8315
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 211  16  TGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACAGTGAATAAGTCTCATGAG  75
               |||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||
Refseq   7666  TGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGAATAAGTCTCATGAG  7607

DONOR 211  76  ATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTGCCTGCCACCGTGTA  135
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7606  ATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTGCCTGCCACCGTGTA  7547

DONOR 211 136  ACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGACTCCCTAGCCCTGT  195
               |||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||
Refseq   7546  ACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGMCTCCCTAGCCCTGT  7487

DONOR 211 196  GGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAGTCTCAGGTATTTCT  255
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7486  GGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAGTCTCAGGTATTTCT  7427

DONOR 211 256  TCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCTAAACCCTAAAAAGC  315
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7426  TCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCTAAACCCTAAAAAGC  7367

DONOR 211 316  ATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGCTCAAACCAGCAGCA  375
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7366  ATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGCTCAAACCAGCAGCA  7307

DONOR 211 376  TCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCCCCAGGGATTTTAGA  435
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7306  TCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCCCCAGGGATTTTAGA  7247

DONOR 211 436  GAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTATGAGCCTATAGAGC  495
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Refseq   7246  GAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTATGAGCCYATAGAGC  7187

DONOR 211 496  TGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCAATGCCTTTAGACtt  555
               |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7186  TGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCAMTGCCTTTAGACTT  7127

DONOR 211 556  ttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTTGCAGGCTCCTGGGT  615
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7126  TTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTTGCAGGCTCCTGGGT  7067

DONOR 211 616  CCCACAGGCAAAGAGGGTCCTTTGCTGGGCCCTCTGGACATGCCCAGACAGGTTTAAATC  675
               ||||||||||||||||  |||||||||||||||||||||| ||||||||||||||||||
Refseq   7066  CCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCAGACAGGTTTAAATC  7009

DONOR 211 676  CTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCACACTGAAAAACAC  735
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7008  CTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCACACTGAAAAACAC  6949

DONOR 211 736  ATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGCACTCATACCTAT  795
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6948  ATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGCACTCATACCTAT  6889

DONOR 211 796  AGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTCATAGATACCAGT  855
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6888  AGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTCATAGATACCAGT  6829

DONOR 211 856  CTGGCTACCTCCATCTCTACTCTGAACTCANANACTNNN-AGGGTACATGGGANGANAGT  914
               |||||||||||||||||||||||||||||| | |||    |||||||||||||| || |||
Refseq   6828  CTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACATGGGAGGACAGT  6769

DONOR 211 915  CCATGCCCTANAGAAATTGAGNAAGATAGCTTCTTCCTCATTGTNGGGAGTACAGTGGNT  974
               |||||||||| |||||||||| |||||||||||||||||||||| |||||||||||||| |
Refseq   6768  CCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGGAGTACAGTGGCT  6709

DONOR 211 975  TTTATACCATAN-CTTCCCATTCTCTTCTCA  1004
               ||||||||||  ||||||||||||||||||
Refseq   6708  TTTATACCATAATCTTCCCATTCTCTTCTCA  6678
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 211    1  CTGTGGGAGGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  60
                ||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Refseq    7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 211   61  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 211  121  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  180
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq    7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 211  181  CTCCCTAGCCCTGTGGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq    7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 211  241  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 211  301  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 211  361  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 211  421  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 211  481  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
                |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 211  541  CTGCCTTTAGACtttttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCT  600
                 |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7140  MTGCCTTTAGAC-TTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCT  7082

DONOR 211  601  TGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAG  660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Refseq    7081  TGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAG  7022

DONOR 211  661  ACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCA  720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7021  ACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCA  6962

DONOR 211  721  CACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGG  780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    6961  CACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGG  6902

DONOR 211  781  CACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATT  840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    6901  CACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATT  6842

DONOR 211  841  CATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTNANNAGGGTAC  900
                |||||||||||||||||||||||||||||||||||||||||||| ||||| |  |||||||
Refseq    6841  CATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTAC  6782

DONOR 211  901  ATGGGAGGACAGTCCATGCCCTANAGAAATTGAGGAAGANAGCTTCNTCCTCATTGTTGG  960
                |||||||||||||||||||||||| ||||||||||||||| |||||| ||||||||||||||
Refseq    6781  ATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGG  6722

DONOR 211  961  GAGTACAGTNNNTTT-ATACCANAN-CTTCCCATTCT  995
                |||||||||   ||| |||||| |  ||||||||||||
Refseq   6721   GAGTACAGTGGCTTTTATACCATAATCTTCCCATTCT  6685
```

Donor 232

5′ → 3′ sequence for 1st haplotype

```
DONOR 232   1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGG  60
               |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq   7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 232  61  CTAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
               |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||
Refseq   7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 232 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 232 181  AGACTTATTCACTGTCATCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
               |||||||||||||||| |||||||||||||||||||||||||| ||||||||||
Refseq   7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 232 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 232 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
               |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Refseq   7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 232 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 232 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 232 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 232 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 232 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 232 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 232 721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGNNGGGGGA  780
               |||||||||||||||||||||||||||||||||||||||||||||||||||  |||| |
Refseq   8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT-GGGG-A  8210

DONOR 232 781  GGAATAAAGGGGAAGCCAATAATGAATGTATGTAGTGGATTTGGGGAAGAGGATATGGGA  840
               |||||||||||||||||| ||||||||||||  ||||||||||||||||||||||||||
Refseq   8211  GGAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGA  8270

DONOR 232 841  GTTTATTCCATTCTAATTCTATCTGTTACNNNCNTGGGAGGTGAAAGCNCCNNCNNTTCG  900
               ||||||||||||||| ||||||||||||    | ||||||||||||| ||  |  ||||
Refseq   8271  GTTTATTCCATTCTA-TTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTCG  8329

DONOR 232 901  ATTAGGGATATCTGANANCAA-GNATCCCTTGAGCNTCCCTGTCC  944
               |||||||||||||||| | ||| | |||||||||||| |||||||||
Refseq   8330  ATTAGGGATATCTGAGAACAAAG-ATCCCTTGAGCCTCCCTGTCC  8373
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 232   1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGG  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq   7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 232  61  CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCATGTGTTACACG  120
               |||||||| ||||||||||||||||||||||||||||||||||| ||||||||||
Refseq  7493   CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 232 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7553   GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 232 181  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
               |||||||||||||||||| |||||||||||||||||||||||||| ||||||||||||
Refseq   7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 232 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 232 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
               |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Refseq   7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 232 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 232 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 232 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 232 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 232 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 232 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 232 721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGANNGGNNGGGGN  780
               |||||||||||||||||||||||||||||||||||||||||||||||||  ||  ||||
Refseq   8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGAT-GGT-GGGG-  8209

DONOR 232 781  AGGAANTAAAGGGGAAGCCAGTAATGAATGTATGAGGTGGATTTGGGGAAGAGGATATGG  840
               ||||| |||||||||||||| |||||||||||  |||||||||||||||||||||||||
Refseq   8210  AGGAA-TAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGG  8268

DONOR 232 841  GGAGTTTATTCN-TTCTATTCTATCTGTTACTAACCTGGNANGTGAA  886
               | |||||||||  |||||||||||||||||||||||||| | |||||
Refseq   8269  G-AGTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAA  8314
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 232  1  TGTGGGAGGTANTTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACAGTGA  60
              |||||||||||| |||||| |||||||||||||||||||||||||||||| ||||||||||
Refseq  7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 232 61  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 232 121 CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  180
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq   7560 CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 232 181 CTCCCTAGCCCTGTGGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq   7500 CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 232 241 TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7440 TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 232 301 AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7380 AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 232 361 TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7320 TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 232 421 CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7260 CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 232 481 TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7200 TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 232 541 ATGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7140 MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 232 601 GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAGA  660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Refseq   7080 GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAGA  7021

DONOR 232 661 CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7020 CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  6961

DONOR 232 721 ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6960 ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  6901

DONOR 232 781 ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6900 ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  6841

DONOR 232 841 ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTNANNAGGGTACA  900
              |||||||||||||||||||||||||||||||||||||||||||| ||||| |  ||||||||
Refseq   6840 ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACA  6781

DONOR 232 901 TGGGAGGACAGTCCATGCCCTANAGAAATTGAGNAAGANAGCTTCTTCCTCATTGTTGGG  960
              |||||||||||||||||||| |||||||||| |||| ||||||||||||||||||||||
Refseq   6780 TGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGG  6721

DONOR 232 961 AGTACAGTGGNTTTTATACCATAATCTTCCCATTCTCTTCT  1001
              |||||||||| ||||||||||||||||||||||||||||
Refseq   6720 AGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCT  6680
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 232   1  CTGTGGGAGGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  60
               |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Refseq   7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 232  61  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 232 121  CCTGCCACCGTGTAACACATGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  180
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq   7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 232 181  CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq   7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 232 241  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 232 301  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 232 361  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 232 421  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 232 481  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
               |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 232 541  CTGCCTTTAGACtttttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCT  600
                ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7140  MTGCCTTTAGAC-TTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCT  7082

DONOR 232 601  TGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAG  660
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Refseq   7081  TGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAG  7022

DONOR 232 661  ACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCA  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7021  ACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCA  6962

DONOR 232 721  CACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGG  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6961  CACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGG  6902

DONOR 232 781  CACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATT  840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6901  CACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATT  6842

DONOR 232 841  CATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTANNGAGGGTAC  900
               |||||||||||||||||||||||||||||||||||||||||| ||||||  ||||||||
Refseq   6841  CATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTAC  6782

DONOR 232 901  ATGGGANGACAGTCCATGCCCTAGAGAAATTGAGnaananaGCTTCTTCCTCATTGTTGG  960
               |||||| |||||||||||||||||||||||| || | ||||||||||||||||||||
Refseq   6781  ATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGG  6722

DONOR 232 961  GAGTACAGTG-CTTTTATACCANNANCTTCCCATTCNCTTCT  1001
               |||||||||| |||||||||||  | |||||||||| |||||
Refseq   6721  GAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCT  6680
```

Donor 238

5′ → 3′ sequence for 1st haplotype

```
DONOR 238   1  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGGC  60
               |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7434  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGGC  7493

DONOR 238  61  TAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACGG  120
               ||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7494  TAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACGG  7553

DONOR 238 121  TGGCAGGCAAGAGGGTGTGTGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  180
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq   7554  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  7613

DONOR 238 181  GACTTATTCACTGTCATCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACCT  240
               ||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||±
Refseq   7614  GACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACCT  7673

DONOR 238 241  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7674  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  7733

DONOR 238 301  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACATT  360
               |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Refseq   7734  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACATT  7793

DONOR 238 361  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7794  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  7853

DONOR 238 421  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7854  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  7913

DONOR 238 481  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7914  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  7973

DONOR 238 541  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7974  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  8033

DONOR 238 601  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8034  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  8093

DONOR 238 661  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8094  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  8153

DONOR 238 721  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGnnnnnnnG-GGGGAGG  779
               |||||||||||||||||||||||||||||||||||||||||||||||       | |||||||
Refseq   8154  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATG-GTGGGGAGG  8212

DONOR 238 780  AATAAAGGGGAAGCCAGTAATGAATGTATGANGTGGATTTGGGGAAGAGGATATGGGAGT  839
               |||||||||||||||| |||||||||||||  ||||||||||||||||||||||||||||
Refseq   8213  AATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGAGT  8272

DONOR 238 840  TTATTCCATTCTNATTCTNTCTGTTACTAACCNNGGGAGGTGAAAGCNCCNTCCAATTCN  899
               |||||||||||| ||||| |||||||||||||  |||||||||||||| || ||||||||
Refseq   8273  TTATTCCATTCT-ATTCTATCTGTTACTAACCT-GGGAGGTGAAAGCACC-TCCAATTCG  8329

DONOR 238 900  NNTNNGGATATCTGAGAACAA-GNTCCCTTGAGCCTTCCCTGTCCTA-GTAGCTCCCTAG  957
                 |  ||||||||||||||| | |||||||||||| |||||||||| ||||||||||||
Refseq   8330  ATTAGGGATATCTGAGAACAAAGATCCCTTGAGCCT-CCCTGTCCTAAGTAGCTCCCTAG  8388

DONOR 238 958  GAAN-AGACTACACCCC  973
               |||  ||||||||||||
Refseq   8389  GAAAGAGACTACACCCC  8405
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 238   1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGG  60
               |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq   7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 238  61 CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCATGTGTTACACG  120
               |||||||| |||||||||||||||||||||||||||||||||||| |||||||||||
Refseq   7493 CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 238 121 GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7553 GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 238 181 AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
               |||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Refseq   7613 AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 238 241 TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7673 TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 238 301 TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
               |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Refseq   7733 TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 238 361 TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7793 TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 238 421 TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7853 TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 238 481 ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7913 ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 238 541 TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7973 TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 238 601 TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 238 661 TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8093 TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 238 721 AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGT  764
               |||||||||||||||||||||||||||||||||||||||||
Refseq   8153 AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGT  8196
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 238   1  CTGTGGGAGGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACNGTGA  60
               |||||||||||||||||| |||||||||||||||||||||||||||| ||||| ||||
Refseq   7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 238  61  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCACACACCCTCTTG  120
               |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 238 121  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  180
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq   7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 238 181  CTCCCTAGCCCTGTGGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
               |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Refseq   7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 238 241  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 238 301  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 238 361  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 238 421  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 238 481  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
               |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 238 541  CTGCCTTTAGACttttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq  7140   MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 238 601  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCCCAGA  660
               |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Refseq   7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAGA  7021

DONOR 238 661  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7020  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  6961

DONOR 238 721  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6960  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  6901

DONOR 238 781  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6900  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  6841

DONOR 238 841  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTANN-AGGGTACA  899
               |||||||||||||||||||||||||||||||||||||||||| ||||||   ||||||||
Refseq   6840  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACA  6781

DONOR 238 900  TGGGAGGANAGTCCATGCCCTANAGAANT-Gnnnan-anaGCTTCTTCCTCATTGNTGGG  957
               |||||||| |||||||||||||| |||| | |   |  | ||||||||||||||||| ||||
Refseq   6780  TGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGG  6721

DONOR 238 958  ANTACAGTGN-TTTTATACCATNATCTTCCCNTTNNCTTCTCAGA  1001
               | |||||||  |||||||||| |||||||| ||  |||||||||
Refseq   6720  AGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTCAGA  6676
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 238  14  TGAATCNTGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGAATAAGTCTCATGAG  73
               |||||  |||||||||||||||||||||||||| ||||||||||||||||||||||||
Refseq   7666  TGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGAATAAGTCTCATGAG  7607

DONOR 238  74  ATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTGCCTGCCACCGTGTA  133
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7606  ATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTGCCTGCCACCGTGTA  7547

DONOR 238 134  ACACATGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGACTCCCTAGCCCTGT  193
               ||||  |||||||||||||||||||||||||||||||||||| ||||||||||||||
Refseq   7546  ACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGMCTCCCTAGCCCTGT  7487

DONOR 238 194  GGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAGTCTCAGGTATTTCT  253
               |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7486  GGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAGTCTCAGGTATTTCT  7427

DONOR 238 254  TCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCTAAACCCTAAAAAGC  313
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7426  TCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCTAAACCCTAAAAAGC  7367

DONOR 238 314  ATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGCTCAAACCAGCAGCA  373
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7366  ATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGCTCAAACCAGCAGCA  7307

DONOR 238 374  TCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCCCCAGGGATTTTAGA  433
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7306  TCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCCCCAGGGATTTTAGA  7247

DONOR 238 434  GAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTATGAGCCTATAGAGC  493
               |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Refseq   7246  GAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTATGAGCCYATAGAGC  7187

DONOR 238 494  TGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCACTGCCTTTAGACtt  553
               |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7186  TGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCAMTGCCTTTAGACTT  7127

DONOR 238 554  tttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTTGCAGGCTCCTGGGT  613
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7126  TTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTTGCAGGCTCCTGGGT  7067

DONOR 238 614  CCCACAGGCAAAGAGGGTCCTTTGCTGGGCCCTCTGGACATGCCCAGACAGGTTTAAATC  673
               ||||||||||||||||  |||||||||||||||||||||||| |||||||||||||||||
Refseq   7066  CCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCAGACAGGTTTAAATC  7009

DONOR 238 674  CTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCACACTGAAAAACAC  733
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7008  CTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCACACTGAAAAACAC  6949

DONOR 238 734  ATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGCACTCATACCTAT  793
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6948  ATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGCACTCATACCTAT  6889

DONOR 238 794  AGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTCATAGATACCAGT  853
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6888  AGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTCATAGATACCAGT  6829

DONOR 238 854  CTGGCTACCTCCATCTCTACTCTGAACTCANAGACTAAGNAGGGTACATGGGAGGACAGT  913
               |||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||||
Refseq   6828  CTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACATGGGAGGACAGT  6769

DONOR 238 914  CCATGCCCTANAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGGAGTACAGTGGNT  973
               |||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |
Refseq   6768  CCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGGAGTACAGTGGCT  6709

DONOR 238 974  TTTATACCATNATCTTCCCATTCTCTTCTC  1003
               |||||||||| ||||||||||||||||||
Refseq   6708  TTTATACCATAATCTTCCCATTCTCTTCTC  6679
```

Donor 251

5′ → 3′ sequence for 1st haplotype

```
DONOR 251    1 ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGG  60
               |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq    7433 ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 251   61 CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCATGTGTTACACG  120
               |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||
Refseq    7493 CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 251  121 GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7553 GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 251  181 AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
               |||||||||||||||| |||||||||||||||||||||||| ||||||||||
Refseq    7613 AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 251  241 TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7673 TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 251  301 TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
               |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Refseq    7733 TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 251  361 TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7793 TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 251  421 TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7853 TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 251  481 ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7913 ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 251  541 TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7973 TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 251  601 TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    8033 TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 251  661 TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    8093 TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 251  721 AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGANGGNNGGGGGA  780
               |||||||||||||||||||||||||||||||||||||||||||||||||| ||  |||| |
Refseq    8153 AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT-GGGG-A  8210

DONOR 251  781 GGAATAAAGGGGACGCCAATAATGAATGTATGTAGTGGATTTTGGGGAAGAGGATATGGG  840
               |||||||||||||| |||| ||||||||||||  |||||||| ||||||||||||||||||
Refseq    8211 GGAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTT-GGGGAAGAGGATATGGG  8269

DONOR 251  841 AGTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGANGTGAAAGCACCTCCAATTCG  900
               |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Refseq    8270 AGTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTCG  8329

DONOR 251  901 ATTANGGATATCTGANAACAAAGATCCCTTGAGCCTCCCTGTCCTAAGTAGCTCCCTAGG  960
               |||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Refseq  8330  ATTAGGGATATCTGAGAACAAAGATCCCTTGAGCCTCCCTGTCCTAAGTAGCTCCCTAGG  8389

DONOR 251  961 AAAGAGACTACNCCCCNCANNACGCTCCTGCATGTTGTTT  1000
               |||||||||| |||| ||  || ||||||||||||||||
Refseq    8390 AAAGAGACTACACCCCACACTACRCTCCTGCATGTTGTTT  8429
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 251    1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGG  60
                ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq    7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 251   61  CTAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
                |||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq    7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 251  121  GTGGCAGGCAAGAGGGTGTGTGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
                |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Refseq    7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 251  181  AGACTTATTCACTGTCATCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
                |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Refseq    7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 251  241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 251  301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
                |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Refseq    7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 251  361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 251  421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 251  481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 251  541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 251  601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 251  661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 251  721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGNNNGNNNNGGGG  780
                ||||||||||||||||||||||||||||||||||||||||||||||||   |    ||||
Refseq    8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT---GGGG  8209

DONOR 251  781  NNAGGAATTAAANGGGGAAGCCAGTAATGAATGTATGAGGTGGATTTGGGGAAGAGGATA  840
                  |||||| ||| |||||||||| ||||||||||||  |||||||||||||||||||||
Refseq    8210  --AGGAAT-AAA-GGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATA  8265

DONOR 251  841  TGGGAGTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGANGTGAAAGCACCTCCNA  900
                |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |
Refseq    8266  TGGGAGTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAA  8325

DONOR 251  901  TTCNATTAGGGNTATCTGANAACAAAGATCCCTTGAGCCTCCCTGT  946
                ||| ||||||| ||||||| ||||||||||||||||||||||||||
Refseq    8326  TTCGATTAGGGATATCTGAGAACAAAGATCCCTTGAGCCTCCCTGT  8371
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 251    1  CTGTGGGAGGTNNTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGA  59
                ||||||||||||   ||||| |||||||||||||||||||||||||||| ||||||||||
Refseq    7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 251   60  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  119
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 251  120  CCTGCCACCGTGTAACACATGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  179
                |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Refseq    7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 251  180  CTCCCTAGCCCTGTGGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  239
                |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Refseq    7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 251  240  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  299
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 251  300  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  359
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 251  360  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  419
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 251  420  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  479
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 251  480  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  539
                |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 251  540  CTGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  599
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 251  600  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGTCCTTTGCTGGGCCCTCTGGACATGCCCA  659
                ||||||||||||||||||||||||||||||  |||||||||||||||||||||||| ||
Refseq    7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCA  7023

DONOR 251  660  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  719
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7022  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  6963

DONOR 251  720  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  779
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    6962  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  6903

DONOR 251  780  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  839
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    6902  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  6843

DONOR 251  840  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANAGACTNNN-AGGGTA  898
                |||||||||||||||||||||||||||||||||||||||||||| |||||    ||||||
Refseq    6842  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTA  6783

DONOR 251  899  CATGGGAGGACAGTCCATGCCCTANAGAAATTGAGGAANATAGCTTCTTCCTCATTGTTG  958
                |||||||||||||||||||||||| |||||||||||| |||||||||||||||||||||
Refseq    6782  CATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  6723

DONOR 251  959  GGAGTACAGTGGCTTTTATACCANNATCTTCCCATTCTCTTCTC  1002
                ||||||||||||||||||||||  ||||||||||||||||||
Refseq    6722  GGAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTC  6781
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 251    1  CTGTGGGAGGT-NNTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACAGTGA  59
                ||||||||||||   ||||| |||||||||||||||||||||||||||| ||||||||||
Refseq    7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 251   60  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCACACACCCTCTTG  119
                |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq    7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 251  120  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  179
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq    7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 251  180  CTCCCTAGCCCTGTGGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  239
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq    7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 251  240  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  299
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 251  300  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  359
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 251  360  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  419
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 251  420  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  479
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 251  480  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  539
                |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 251  540  CTGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  599
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 251  600  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCCCAGA  659
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Refseq    7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAGA  7021

DONOR 251  660  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  719
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    7020  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  6961

DONOR 251  720  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  779
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    6960  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  6901

DONOR 251780   ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  839
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    6900  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  6841

DONOR 251  840  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCANANACTNANNAGGGTACA  899
                |||||||||||||||||||||||||||||||||||||||| | ||| |  ||||||||
Refseq    6840  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACA  6781

DONOR 251  900  TGGGAGGACAGTCCATGCCCTANAGAAATTGAGGAANATAGCTTCTTCCTCATTGTTGG-  958
                |||||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||
Refseq    6780  TGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGG  6721

DONOR 251  959  AGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTCAGA  1003
                |||||||||||||||||||||||||||||||||||||||||||||
Refseq    6720  AGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTCAGA  6676
```

Donor 253

5′ → 3′ sequence for 1st haplotype

```
DONOR 253   1  ACCTGAGANTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGG  60
               |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 253  61  CTAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
               |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||
Refseq   7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 253 121  GTGGCAGGCAAGAGGGTGTGTGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
               |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Refseq   7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 253 181  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
               |||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||
Refseq   7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 253 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 253 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Refseq   7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 253 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 253 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 253 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 253 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 253 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 253 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 253 721  AGAAAAAGATTCCCTGAATCtggggagaaggggatggtggtggtggtgatgggtgggggga  780
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |
Refseq   8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGG-TGGGG-A  8210

DONOR 253 781  ggaataaaggggaAGCCNGTNATGAATGTATGANGTGGATTTGGGGAAGAGGATATGGGA  840
               ||||||||||||||||  | ||||||||||  ||||||||||||||||||||||||||
Refseq   8211  GGAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGA  8270

DONOR 253 841  GTTTNTTCCATTCTATTCTATCTGTTACTAACCTGGGANGTGAAANCACCTCCNATTCGA  900
               |||| |||||||||||||||||||||||||||||| |||||| |||||||| ||||||
Refseq   8271  GTTTATTCCATTCTATTCTATCTGTTACTAACCTGGGAGGTGAAAGCACCTCCAATTCGA  8330

DONOR 253 901  TTAGGGNNATCTGAGAACNAAGATCCCTTGAG  932
               ||||||  |||||||||| ||||||||||||
Refseq   8331  TTAGGGATATCTGAGAACAAAGATCCCTTGAG  8362
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 253   1  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGGC  60
               ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7434  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGGC  7493

DONOR 253  61  TAGGGAGTCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACGG  120
               ||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7494  TAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACGG  7553

DONOR 253 121  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7554  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  7613

DONOR 253 181  GACTTATTCACTGTCATCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACCT  240
               |||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Refseq   7614  GACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACCT  7673

DONOR 253 241  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7674  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  7733

DONOR 253 301  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACATT  360
               |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Refseq   7734  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACATT  7793

DONOR 253 361  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7794  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  7853

DONOR 253 421  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7854  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  7913

DONOR 253 481  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7914  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  7973

DONOR 253 541  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7974  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  8033

DONOR 253 601  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8034  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  8093

DONOR 253 661  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8094  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  8153

DONOR 253 721  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGNNGGGGAGG  780
               |||||||||||||||||||||||||||||||||||||||||||||||||  |||||||
Refseq   8154  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGATGGT-GGGGAGG  8212

DONOR 253 781  AATAAAGGGGAAGCCAGTAATGAATGTATGANGTGGATTTGGGGAAGANGATATGGGAGT  840
               ||||||||||||||| ||||||||||||  |||||||||||||||| ||||||||||||
Refseq   8213  AATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGAGT  8272

DONOR 253 841  TTATTCCNTTCTATTCTATCTGTTACTAACCTGGGANGGNGAAANNNCCTCCNATTCNAN  900
               ||||||| |||||||||||||||||||||||| || ||||   ||||| |||| |
Refseq   8273  TTATTCCATTCTATTCTATCTGTTACTAACCTGGGA-GGTGAAAGCACCTCCAATTCGAT  8331

DONOR 253 901  TAGGGANANCCTGANAACNAAGAANCCCTTGANCCTCCCTGTC-TAN-TANCTCCCTNNG  958
               |||||| | | ||| ||| ||||  |||||||| |||||||||| ||  || ||||||  |
Refseq   8332  TAGGGATATC-TGAGAACAAAGAT-CCCTTGAGCCTCCCTGTCCTAAGTAGCTCCCTAGG  8389

DONOR 253 959  AAAGANACTA  968
               ||||| ||||
Refseq   8390  AAAGAGACTA  8399
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 253   2  TCTGTGGGAGG-NNNTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTG  60
               ||||||||||||    ||||| |||||||||||||||||||||||||||||| |||||||||
Refseq   7681  TCTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTG  7622

DONOR 253  61  AATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCACACACCCTCTT  120
               |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq   7621  AATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTT  7562

DONOR 253 121  GCCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGG  180
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq   7561  GCCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGG  7502

DONOR 253 181  ACTCCCTAGCCCTGTGGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCA  240
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Refseq   7501  MCTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCA  7442

DONOR 253 241  GTCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTC  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq    441  GTCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTC  7382

DONOR 253 301  TAAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTG  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7381  TAAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTG  7322

DONOR 253 361  CTCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7321  CTCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCC  7262

DONOR 253 421  CCCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7261  CCCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCT  7202

DONOR 253 481  ATGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCC  540
               ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7201  ATGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCC  7142

DONOR 253 541  ACTGCCTTTAGACtttttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATC  600
               | |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Refseq   7141  AMTGCCTTTAGAC-TTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATC  7083

DONOR 253 601  TTGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCA  660
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Refseq   7082  TTGCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCA  7023

DONOR 253 661  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7022  GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  6963

DONOR 253 721  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6962  ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  6903

DONOR 253 781  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6902  GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  6843

DONOR 253 841  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGANACTNNN-AGGGTA  899
               |||||||||||||||||||||||||||||||||||||||||||||| |||    ||||||
Refseq   6842  TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTA  6783

DONOR 253 900  CATGGGANGACAGTCCATGCCCTANAGAAATTGAGGAANANNGCTTCTTCCTCATTGTTG  959
               ||||||| ||||||||||||||| |||||||||||||  |  ||||||||||||||||||
Refseq   6782  CATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  6723

DONOR 253 960  GGANTACNGTNNNTTT-ATACCATAN-CNTCCCNTNCNCTTCTCA  1002
               ||| ||| ||   ||| ||||||||  | |||| | | |||||||
Refseq   6722  GGAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCTCA  6678
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 253   1  CTGTGGGANGTAATTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACAGTGA  60
               ||||||||| |||||||||| |||||||||||||||||||||||||||| ||||||||||
Refseq   7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 253  61 ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7620 ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 253 121 CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGA  180
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq   7560 CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 253 181 CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  240
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq   7500 CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 253 241 TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7440 TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 253 301 AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7380 AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 253 361 TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7320 TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 253 421 CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7260 CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 253 481 TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  540
               |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7200 TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 253 541 ATGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7140 MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 253 601 GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGTCCTTTGCTGGGCCCTCTGGACATGCCCA  660
               |||||||||||||||||||||||||||||||  |||||||||||||||||||||||| ||
Refseq   7080 GCAGGCTCCTGGGTCCCACAGGCAAAGAGGG--CTTTGCTGGGCCCTCTGGACATGCYCA  7023

DONOR 253 661 GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7022 GACAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTC  6963

DONOR 253 721 ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6962 ACACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTG  6903

DONOR 253 781 GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6902 GCACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTAT  6843

DONOR 253 841 TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTNANNAGGGTA  900
               |||||||||||||||||||||||||||||||||||||||||||||||| |  ||||||
Refseq   6842 TCATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTA  6783

DONOR 253 901 CATGGGAGGACAGTCCATGCCCTANAGAAATTGAGGAANATAGCTTCTTCCTCATTGTTG  960
               |||||||||||||||||||||||| |||||||||||| ||||||||||||||||||||||
Refseq   6782 CATGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTG  6723

DONOR 253 961 GGAGTACAGNGGNTTTTATACCATAATCTTCCCATTCTCTTC  1002
               ||||||||| || |||||||||||||||||||||||||||
Refseq   6722 GGAGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTC  6681
```

Donor 280

5′ → 3′ sequence for 1st haplotype

```
DONOR 280   1  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACAGCTCCACAGGGC  60
               |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7434  CCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGGC  7493

DONOR 280  61  TAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACGG  120
               ||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Refseq   7494  TAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACGG  7553

DONOR 280 121  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7554  TGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATGA  7613

DONOR 280 181  GACTTATTCACTGTCATCAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACCT  240
               |||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Refseq   7614  GACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACCT  7673

DONOR 280 241  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7674  CCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGATT  7733

DONOR 280 301  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACATT  360
               |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Refseq   7734  TGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACATT  7793

DONOR 280 361  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7794  TTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTCT  7853

DONOR 280 421  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7854  TGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAATA  7913

DONOR 280 481  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7914  TTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATTT  7973

DONOR 280 541  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7974  TTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTATT  8033

DONOR 280 601  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8034  ACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCAT  8093

DONOR 280 661  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8094  AAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACTA  8153

DONOR 280 721  GAAAAAGATTCCCTGAATCtggggagaaggggatggtggtggtggtgatgggtgggggag  780
               |||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||
Refseq   8154  GAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGG-TGGGG-AG  8211

DONOR 280 781  gaataaaggggaAGCCAGTAATGAATGTATGAGGTGGATTTGGGGAAGAGGATATGGGAG  840
               |||||||||||||||| ||||||||||||  |||||||||||||||||||||||||||
Refseq   8212  GAATAAAGGGGAAGCCARTAATGAATGTATGWRGTGGATTTGGGGAAGAGGATATGGGAG  8271

DONOR 280 841  TTTATTCCATTCNANTNCNATNNGTTNCTANNCTGGGAGGTGAAAGCACCTC  892
               |||||||||||| | | | ||  ||| |||  ||||||||||||||||||||
Refseq   8272  TTTATTCCATTCTATT-CTATCTGTTACTAACCTGGGAGGTGAAAGCACCTC  8322
```

5′ → 3′ sequence for 2nd haplotype

```
DONOR 280   1  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACCGCTCCACAGGG  60
               |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq   7433  ACCTGAGACTGGGTAATTTATAAAGAAAAAGAAGTTTAATGGACTCACMGCTCCACAGGG  7492

DONOR 280  61  CTAGGGAGGCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCACGTGTTACACG  120
               |||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||
Refseq   7493  CTAGGGAGKCCTCACTATCACGGCAGAAGGCAAAGGAGGAGCTAAGGCAYGTGTTACACG  7552

DONOR 280 121  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7553  GTGGCAGGCAAGAGGGTGTGGGCAGGGGAACTGCACTTTATAAAACCATCAGATCTCATG  7612

DONOR 280 181  AGACTTATTCACTGTCATAAGAACAGCATGTGAAAAACCTGCCCCCATGATTCAATTACC  240
               |||||||||||||||||| |||||||||||||||||||||||||| ||||||||||
Refseq   7613  AGACTTATTCACTGTCATMAGAACAGCATGTGAAAAACCTGCCCCCATSATTCAATTACC  7672

DONOR 280 241  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7673  TCCCACAGAGTCCCTCCCATGACACATGCGGATTATGGGAGCTAAAACTCAGGATGAGAT  7732

DONOR 280 301  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGGAGGAGAACAATAGTAAGAACACAT  360
               |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Refseq   7733  TTGGGTGGGGACACAGCCAAACCATATCAGTAGGGKAGGAGAACAATAGTAAGAACACAT  7792

DONOR 280 361  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7793  TTTAGAATTCCCCAAGGATGAGTTGAAGGGCTGAAGAGCTTTGAAGATGAGCTGAAGCTC  7852

DONOR 280 421  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7853  TTGAAATTGCTAGTTGCAACAAAACCAGAGAATATTTGTGGCAGAATAATATTTTTAAAT  7912

DONOR 280 481  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7913  ATTAACAGGAATCATAATTTTCCTATTATTTTCTATTACCCTGTAATCCCATGACCCATT  7972

DONOR 280 541  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7973  TTTGTTTCAAGCCCAGACCCAGACTTGGAGGTGTTGATTACCACGGACAATTGATGTTAT  8032

DONOR 280 601  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8033  TACAACAGCAGAATAAAGGCTTTGAGTTTCATAAGCCTAGCCTTGAGGGCACTTAAGGCA  8092

DONOR 280 661  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   8093  TAAACAGTAGATTTGCATTAATATCTGATGAGTTAGAAAAAGGTGGTTTACTTCCAAACT  8152

DONOR 280 721  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGNNGGGGGA  780
               ||||||||||||||||||||||||||||||||||||||||||||||  |||| |
Refseq   8153  AGAAAAAGATTCCCTGAATCTGGGGAGAAGGGGATGGTGGTGGTGGTGATGGT-GGGG-A  8210

DONOR 280 781  GGAATAAAGGGGAAGCCAGTAATGNANTGTATGAGGTGGATTTGGGGAAGAGGATATGGG  840
               |||||||||||||||||| ||||| | ||||||  |||||||||||||||||||||||
Refseq   8211  GGAATAAAGGGGAAGCCARTAATGAA-TGTATGWRGTGGATTTGGGGAAGAGGATATGGG  8269
```

3′ → 5′ sequence for 1st haplotype

```
DONOR 280   1  CTGTGGGAGG-NNNTGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTGATGACAGTGA  59
               ||||||||||    ||||| |||||||||||||||||||||||||||| ||||||||||
Refseq   7680  CTGTGGGAGGTAATTGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGA  7621

DONOR 280  60  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  119
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7620  ATAAGTCTCATGAGATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTG  7561

DONOR 280 120  CCTGCCACCGTGTAACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGC  179
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq   7560  CCTGCCACCGTGTAACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGM  7501

DONOR 280 180  CTCCCTAGCCCTGTGGAGCTGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  239
               |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Refseq   7500  CTCCCTAGCCCTGTGGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAG  7441

DONOR 280 240  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  299
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7440  TCTCAGGTATTTCTTCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCT  7381

DONOR 280 300  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  359
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7380  AAACCCTAAAAAGCATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGC  7321

DONOR 280 360  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  419
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7320  TCAAACCAGCAGCATCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCC  7261

DONOR 280 420  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  479
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7260  CCAGGGATTTTAGAGAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTA  7201

DONOR 280 480  TGAGCCTATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  539
               |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7200  TGAGCCYATAGAGCTGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCA  7141

DONOR 280 540  CTGCCTTTAGACtttttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  599
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7140  MTGCCTTTAGACTTTTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTT  7081

DONOR 280 600  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCCCAGA  659
               |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Refseq   7080  GCAGGCTCCTGGGTCCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAGA  7021

DONOR 280 660  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  719
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7020  CAGGTTTAAATCCTAGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCAC  6961

DONOR 280 720  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  779
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6960  ACTGAAAAACACATTGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGC  6901

DONOR 280 780  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  839
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6900  ACTCATACCTATAGGTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTC  6841

DONOR 280 840  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTNANNAGGGTACA  899
               |||||||||||||||||||||||||||||||||||||||||||||||| |  ||||||||
Refseq   6840  ATAGATACCAGTCTGGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACA  6781

DONOR 280 900  TGGGAGGACAGTCCATGCCCTananaaaTTGAGGAANATAGCTTCTTCCTCATTGTTGGG  959
               |||||||||||||||||||||| | |||||||||||| ||||||||||||||||||||||
Refseq   6780  TGGGAGGACAGTCCATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGG  6721

DONOR 280 960  AGTACAGTNNNTTT-ATACCATAATCTTCCCATTCTCTTCT  999
               ||||||||   ||| ||||||||||||||||||||||||||
Refseq   6720  AGTACAGTGGCTTTTATACCATAATCTTCCCATTCTCTTCT  6680
```

3′ → 5′ sequence for 2nd haplotype

```
DONOR 280  14  TGAATCATGGGGGCAGGTTTTTCACATGCTGTTCTTATGACAGTGAATAAGTCTCATGAG  73
               |||||.|||||||||||||||||||||||||||.|||||||||||||||||||||||||
Refseq   7666  TGAATSATGGGGGCAGGTTTTTCACATGCTGTTCTKATGACAGTGAATAAGTCTCATGAG  7607

DONOR 280  74  ATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTGCCTGCCACCGTGTA  133
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7606  ATCTGATGGTTTTATAAAGTGCAGTTCCCCTGCCCACACCCTCTTGCCTGCCACCGTGTA  7547

DONOR 280  134 ACACGTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGCCTCCCTAGCCCTGT  193
               ||||.|||||||||||||||||||||||||||||||||||||.|||||||||||||||
Refseq   7546  ACACRTGCCTTAGCTCCTCCTTTGCCTTCTGCCGTGATAGTGAGGMCTCCCTAGCCCTGT  7487

DONOR 280 194  GGAGCGGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAGTCTCAGGTATTTCT  253
               |||||.|||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7486  GGAGCKGTGAGTCCATTAAACTTCTTTTTCTTTATAAATTACCCAGTCTCAGGTATTTCT  7427

DONOR 280 254  TCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCTAAACCCTAAAAAGC  313
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7426  TCATAGCAGTATGAAAATGGACTAATACACCTACTAACTGCCTTCTAAACCCTAAAAAGC  7367

DONOR 280 314  ATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGCTCAAACCAGCAGCA  373
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7366  ATGAGGTGCCACTATAAAAGCTGCCTATTACATTTCCTGTTGCTGCTCAAACCAGCAGCA  7307

DONOR 280 374  TCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCCCCAGGGATTTTAGA  433
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7306  TCATGCTGCCAAGTCAGTGAAGGCTTTTCTGCATTCTCAATGTCCCCCAGGGATTTTAGA  7247

DONOR 280 434  GAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTATGAGCCTATAGAGC  493
               |||||||||||||||||||||||||||||||||||||||||||||||||||.|||||||
Refseq   7246  GAAACAGACCTTTTAGCTTGTACTCATTGTTTAACTTTGTTGGCTATGAGCCYATAGAGC  7187

DONOR 280 494  TGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCACTGCCTTTAGACtt  553
               |||||||||||||||||||||||||||||||||||||||||||||||||||.||||||||
Refseq   7186  TGGAGCAAGGGGTACTCTTAGTGAGAGACATGGCCCTTGTTCTCCAMTGCCTTTAGACTT  7127

DONOR 280 554  ttttGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTTGCAGGCTCCTGGGT  613
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7126  TTTTTGCTCCAAGCACGCAGCTTCCGTGCTCTTTTTGTCTCATCTTGCAGGCTCCTGGGT  7067

DONOR 280 614  CCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCTCAGACAGGTTTAAATCCT  673
               |||||||||||||||||||||||||||||||||||.|||||||||||||||||||||||
Refseq   7066  CCCACAGGCAAAGAGGGCTTTGCTGGGCCCTCTGGACATGCYCAGACAGGTTTAAATCCT  7007

DONOR 280 674  AGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCACACTGAAAAACACAT  733
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   7006  AGGATAGAGACAGCCCAGGTTCTCAAGCTGAGACCCACAACCTCACACTGAAAAACACAT  6947

DONOR 280 734  TGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGCACTCATACCTATAG  793
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6946  TGATACCAAAGACTGATTAATATGCATGTAAATACTGGCATTTGGCACTCATACCTATAG  6887

DONOR 280 794  GTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTCATAGATACCAGTCT  853
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Refseq   6886  GTAGTAGAGGCTCAAAAGCCAAGACTGAAAAAATTTTTCTCTATTCATAGATACCAGTCT  6827

DONOR 280 854  GGCTACCTCCATCTCTACTCTGAACTCANAGACTAANGAGGGTACATGGGAGGACAGTCC  913
               |||||||||||||||||||||||||||||.|||||||.|||||||||||||||||||||||
Refseq   6826  GGCTACCTCCATCTCTACTCTGAACTCAGAGACTAAGGAGGGTACATGGGAGGACAGTCC  6767

DONOR 280 914  ATGCCCTANAGAAATTGANGAANANAGCTTCTTCCTCATTGTTGGGAGTACAGTGGCTTT  973
               ||||||||.|||||||||.|||.|||||||||||||||||||||||||||||||||||||
Refseq   6766  ATGCCCTAGAGAAATTGAGGAAGATAGCTTCTTCCTCATTGTTGGGAGTACAGTGGCTTT  6707

DONOR 280 974  TATACCATAATCTTCCCATTCTCTTCT  1000
               |||||||||||||||||||||||||||
Refseq   6706  TATACCATAATCTTCCCATTCTCTTCT  6680
```

REFERENCES

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. Cell 61:419-36

Allers T, Lichten M (2001) Differential timing and control of noncrossover and crossover recombination during meiosis. Cell 106:47-57

Anderson LK, Offenberg HH, Verkuijlen WM, Heyting C (1997) RecA-like proteins are components of early meiotic nodules in lily. Proc Natl Sci Acad USA 94:6868-73

Antonarakis SE, Boehm CD, Giardina PJ, Kazazian HH, Jr. (1982) Nonrandom association of polymorphic restriction sites in the beta-globin gene cluster. Proc Natl Acad Sci USA 79:137-41

Ardlie KG, Kruglyak L, Seielstad M (2002) patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299-309

Baird DM, Jeffreys AJ, Royle NJ (1995) Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. EMBO J 14:5433-43

Baker SM, Plug AW, Prolla TA, Bronner CE, Harris AC, Yao X, Christie DM, Monell C, Arnheim N, Bradley A, Ashley T, Liskay RM (1996) Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over. Nat Genet 13:336-42

Bass HW (2003) Telomere dynamics unique to meiotic prophase: formation and significance of the bouquet. Cell Mol Life Sci 60:2319-24

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. Science 327:836-40

Baudat F, de Massy B (2007) Cis- and trans-acting elements regulate the mouse Psmb9 meiotic recombination hotspot. PLoS Genet 3:e100

Baudat F, Nicolas A (1997) Clustering of meiotic double-strand breaks on yeast chromosome III. Proc Natl Acad Sci USA 94:5213-8

Berg IL, Neumann R, Lam KW, Sarbajna S, Odenthal-Hesse L, May CA, Jeffreys AJ (2010) PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. Nat Genet 42:859-63

Berg IL, Neumann R, Sarbajna S, Odenthal-Hesse L, Butler NJ, Jeffreys AJ (2011) Variants of the protein PRDM9 differentially regulate a set human

meiotic recombination hotspots highly active in African populations. Proc Natl Acad Sci USA 108:12378-83

Blat Y, Protacio RU, Hunter N, Kleckner N (2002) Physical and functional interactions among basic chromosome organizational features govern early steps of meiotic chiasma formation. Cell 111:791-802

Borde V, Robine N, Lin W, Bonfils S, Geli V, Nicolas A (2009) Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. EMBO J 28:99-111

Boulton A, Myers RS, Redfield RJ (1997) The hotspot conversion paradox and the evolution of meiotic recombination. Proc Natl Acad Sci USA 94:8058-63

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

Brown TC, Jiricny J (1988) Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. Cell 54:705-11

Brunschwig H, Levi L, Ben-David E, Williams RW (2012) Fine-scale map of recombination rates and hotsots in the mouse genome. Genetics 10.1534

Buard J, Barthes P, Grey C, de Massy B (2009) Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. EMBO J 28:2616-24

Bullard SA, Kim S, Galbraith AM, Malone RE (1996) Double strand breaks at the HIS2 recombination hot spot in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 93:13054-9

Cao L, Alani E, Kleckner N (1990) A pathway for generation and processing of double-strand breaks during meiotic recombination in S. cerevisiae. Cell 61:1089-101

Cervantes MD, Farah JA, Smith GR (2000) Meiotic DNA breaks associated with recombination in S.pombe. Mol Cell 5:883-3

Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH (1984) Nonuniform recombination within the human beta-globin gene cluster. Am J Hum Genet 36:1239-58.

Cromie GA, Hyppa RW, Taylor AF, Zakharyevich K, Hunter N, Smith GR (2006) Single Holliday junctions are intermediates of meiotic recombination. Cell 127:1167-78

Collins A, Morton NE (1998) Mapping a disease locus by allelic association. Proc Natl Acad Sci USA 95:1741-1745

Copenhaver GP, Housworth EA, Stahl FW (2002) Crossover interference in Arabidopsis. Genetics 160:1631-9

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 Science 319:1395-8

Cullen M, Erlich H, Klitz W, Carrington M (1995) Molecular Mapping of a Recombination Hotspot Located in the Second Intron of the Human TAP2 locus. Am J Hum Genet 56:1350-1358

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. Am J Hum Genet 60:397-407

Cullen M, Perfectto SP, Klitz W, Nelson G, Carrington M (2002) High-resolution patterns of meiotic recombination across the human major histocompatibility complex. Am J Hum Genet 71:759-76

Dalganov GM, Maser RS, Novikov A, Tosto L, Chong S, Bressan DA, Petrini JH (1996) Human Rad50 is physically associated with human Mre11: identification of a concerved multiprotein complex implicated in recombination DNA repair. Mol Cell Biol 16:4832-41

de los Santos T, Hunter N, Lee C, Larkin B, Loidi 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. Genetics 164:81-94

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS (2002) Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci USA. 99:5261-6

Dobson MJ, Pearlman RE, Karaiskakis A, Spyropoulos B, Moens PB (1994) Synaptonemal complex proteins: occurrence, epitope mapping and chromosome disjunction. J Cell Sci 107:2749-60

Dooner HK (1986) Genetic Fine Structure of the BRONZE Locus in Maize. Genetics 113:1021-1036.

Dooner HK, Martínez-Férez IM (1997) Germinal excisions of the maize transposon activator do not stimulate meiotic recombination or homology-dependent repair at the bz locus. Genetics 147:1923-32.

Egel-Mitani M, Olson LW, Egel R (1982) Meiosis in Aspergillus nidulans: another example for lacking synaptonemal complexes in the absence of crossover interference. Hereditas 97:179-87

Fan Q, Xu F, Petes TD (1995) Meiosis-specific double-strand DNA breaks at the HIS4 recombination hot spot in the yeast Saccharomyces cerevisiae: control in cis and trans. Mol Cell Biol 15:1679-88

Fan QQ, Petes TD (1996) Relationship between nuclease-hypersensitive sites and meiotic recombination hot spot activity at the HIS4 locus of Saccharomyces cerevisiae. Mol Cell Biol 16:2037-43.

Ferguson DO, Holloman WK (1996) Recombinational repair of gaps in DNA is asymmetric in Ustilago maydis and can be explained by a migrating D-loop model. Proc Natl Acad Sci USA 93:5419-24

Fearnhead P, Harding RM, Schneider JA, Myers S, Donnelly P (2004) Application of coalescent methods to reveal fine-scale rate variation and recombination hotspots. Genetics 167:2067-81

Fox ME, Yamada T, Ohta K, Smith GR (2000) A family of cAMP-response-element-related DNA sequences with meiotic recombination hotspot activity in Schizosaccharomyces pombe. Genetics 156:59-68

Freije D, Helms C, Watson MS, Donis-Keller H (1992) Identification of a second pseudoautosomal region near the Xq and Yq telomeres. Science 258:1784-7

Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD (2000) Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci USA 97:11383-90

Grey C, Barthes P, Chauveau-Le Friec G, Baudat F, de Massy B (2011) Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. PLoS Biol 9:e1001176

Grey C, Baudat F, de Massy B (2009) Genome-wide control of the disribution of meiotic recombination. PLoS Biol 7:e35

Guillon H, de Massy B (2002) An initiation site for meiotic crossing-over and gene conversion in the mouse. Nat Genet 32:296-9

Hartung F, Puchta H (2000) Molecular characterisation of two paralogous SPO11 homologues in Arabidopsis thaliana. Nucleic Acids Res 28:1548-54

Hayashi K, Yoshida K, Matsul Y (2005) A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. Nature 438:374-8

Henderson KA, Keeney S (2005) Synaptonemal complex formation: where does it start? Bioessays 27:995-8

Henke A, Fisher C, Rappoid GA (1993) Genetics map of the human pseudoautosomal region reveals a high rate of recombination in female meiosis at the Xp telomere. Genomics 18:478-85

Holliday R (1964) The Induction of Mitotic Recombination by Mitomycin C in Ustilago and Saccharomyces. Genetics 50:323-35

Holloway K, Lawson VE, Jeffreys AJ (2006) Allelic recombination and de novo deletions in sperm in the human beta-globin gene region. Hum Mol Genet 15:1099-111

Housworth EA, Stahl FW (2003) Crossover interference in humans. Am J Hum Genet 73:188-97

Hubert R, MacDonald M, Gusella J, Arnheim N (1994) High resolution localization of recombination hot spots using sperm typing. Nat Genet 7:420:424

Hunter N, Borts RH (1997) Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis. Genes Dev 11:1573-82

International HapMap Consortium (2005) A haplotype map of human genome. Nature 437:299-320

The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851-61

International HapMap 3 Consortium (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467:52-8

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921

International SNP Map Working Group (2001) A map of huma genome sequence variation containing 1.42 million single nucleotide polymorpisms. Nature 409:928-33

Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 29:217-22

Jeffreys AJ, May CA (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. Nat Genet 36:151-6

Jeffreys AJ, Murray J, Neumann R (1998a) High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. Mol Cell 2:267-73

Jeffreys AJ, Neil DL, Neumann R (1998b) Repeat instability at human minisatellites arising from meiotic recombination. EMBO J 17:4147-57

Jeffreys AJ, Neumann R (2002) Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. Nat Genet 31:267-71

Jeffreys AJ, Neumann R (2005) Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. Hum Mol Genet 14:2277-87

Jeffreys AJ, Neumann R (2009) The rise and fall of a human recombination hot spot. Nat Genet 41:625-9

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

Jeffreys AJ, Neumann R, Wilson V (1990) Repat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell 60:473-85

Jeffreys AJ, Ritchie A, Neumann R (2000) High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. Hum Mol Genet 9:725-33

Johzuka K, Ogawa H (1995) Interaction of Mre11 and Rad50: two proteins required for DNA repair and meiosis-specific double-strand break formation in Saccharomyces cerevisiae. Genetics 139:1521-32

Kaback DB, Guacci V, Barber D, Mahon JW (1992) Chromosome size-dependent control of meiotic recombination. Science 256:228-32

Kauppi L, Jeffreys AJ, Keeney S (2004) Where the crossovers are: recombination distributions in mammals. Nat Rev Genet 5:413-24

Kauppi L, May CA, Jeffreys AJ (2009) Analysis of meiotic recombination products from human sperm. Methods Mol Biol 557:323-55

Kauppi L, Stumpf MP, Jeffreys AJ (2005) Localized breakdown in linkage disequilibrium does not always predict sperm crossover hot spots in the human MHC class II region. Genomics 86:13-24

Keeney S (2008) Spo11 and the formation of DNA double-strand breaks in meiosis. Genome Dyn Stab 2:81-123

Keeney S, Baudat F, Angeles M, Zhou ZH, Copeland NG, Jenkins NA, Manova K, Jasin M (1999) A mouse homolog of the Saccharomyces cerevisiae meiotic recombination DNA transesterase Spo11p. Genomics 61:170-82

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. Cell 88:375-84

Keeney S, Kleckner N (1995) Covalent protein-DNA complexes t the 5' strand termini of meiosis-specific double-strand breaks in yeast. Proc Natl Acad Sci USA 92:11274-8

Kirkpatrick DT, Fan Q, Petes TD (1999) Maximal stimulation of meiotic recombination by a yeast transcription factor requires the transcription activation domain and a DNA-binding domain. Genetics 152:101-15

Kohli J, Bahler J (1994) Homologous recombination in fission yeast: absence of crossover interference and synaptonemal complex. Experientia 50:295-306

Kon N, Krawchuk MD, Warren BG, Wahls WP (1997) Transcription factor Mts1/Mts2 (Atf1/Pcr1) activates the M26 meiotic recombination hotspot in Schizosaccharomyces pombe. Proc Natl Acad Sci USA 94:13765-70

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241-7.

Kong A, Barnard J, Gudbjartsson DF, Thorleifsson G, Jonsdottir G, Sigurdardottir S, Richardsson B, Jonsdottir J, Thorgeirsson T, Frigge ML, Lamb NE, Sherman S, Gulcher JR, Stefansson K (2004) Recombination rate and reproductive success in humans. Nat Genet 36:1203-6

Lee JY, Orr-Weaver TL (2001) The molecular basis of sister-chromatid cohesion. Annu Rev Cell Dev Biol 17:753-77

Lewontin RC (1964) The Interaction of Selection and Linkage. Ii. Optimum Models. Genetics 50:757-782.

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. Nature 335:414-7

Li L, Hamer DH (1995) Recombination and allelic association in the Xq/Yq homology region. Hum Mol Genet 4:2013-6

Lichten M, Goldman AS (1995) Meiotic recombination hotspots. Annu Rev Genet 29:423-44

Lien S, Szyda J, Schechinger B, Rappold G, Arnheim N (2000) Evidence for heterogeneity in recombination in the human pseudoautosomal region: high resolution analysis by sperm typing and radiation-hybrid mapping. Am J Hum Genet. 66(2):557-66

Lin Y, Smith GR (1994) Transient, meiosis-induced expression of the rec6 and rec12 genes of Schizosaccharomyces pombe. Genetics 136:769-79

Liu Y, Masson JY, Shah R, O'Regan P, West SC (2004) RAD51C is required for Holliday junction processing in mammalian cells. Science 303:243-6

Liu J, Wu TC, Linchten M (1995) The location and structure of double-strand DNA breaks induced during yeast meiosis: evidence for a covalently linked DNA-protein intermediate. EMBO J 14:4599-608

Malécot G (1948) Les mathématiques de l'hérédité. Masson et Cie, Paris. [The mathematics of heredity (1968 English translation by D.M. Yermanos). W.H. Freeman and Co., San Francisco.]

Mani P, Yadav VK, Das SK, Chowdhury S (2009) Genome-wide analyses of recombination prone regions predict role of DNA structural motif in recombination. PLoS One 4:e4399

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. Proc Natl Acad Sci USA 99:228-33

Martini E, Diaz RL, Hunter N, Keeney S (2006) Crossover homeostasis in yeast meiosis. Cell 126:285-95

Marsischky GT, Filosi N, Kane MF, Kolodner R (1996) Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. Genes Dev 10:407-20

May CA, Shone AC, Kalaydjieva L, Sajantila A, Jeffreys AJ (2002) Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene SHOX. Nat Genet 31:272-5.

McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P (2004) The fine-scale structure of recombination rate variation in the human genome. Science 304:581-4

McKee AH, Kleckner N (1997) Mutations in Saccharomyces cerevisiae that block meiotic prophase chromosome metabolism and confer cell cycle arrest at pachytene identify two meiosis-specific genes SAE1 and SAE3. Genetics 146:817-34

McKim KS, Green-Marroquin BL, Sekelsky JJ, Chin G, Steinberg C, Khodosh R, Hawley RS (1998) Meiotic synapsis in the absence of recombination. Science 279:876-8

Meuwissen RL, Meerts I, Hoovers JM, Leschot NJ, Heyting C (1997) Human synaptonemal complex protein 1 (SCP1): isolation and characterization of the cDNA and chromosomal localization of the gene. Genomics 39:377-84

Morton NE, Zhang W, Taillon-Miller W, Ennis S, Kwok –Y, Collins A (2001) The optimal measure of allelic association. Proc Natl Acad Sci USA 98:5217-5221

Moses MJ, Counce SJ, Paulson DF (1975) Synaptonemal complex complement of man in spreads of spermatocytes, with details of the sex chromosome pair. Science 187:363-5

Myers S, Bottolo Leonardo, Freeman C, McVean G, Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome Science 310: 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. Science 327:876

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. Nat Genet 40:1124-9

Myers S, Spencer CC, Auton A, Bottolo L, Freeman C, Donnelly P, McVean G (2006) The distribution and causes of meiotic recombination in the human genome. Biochem Soc Trans 34:526-30

Nairz K, Klein F (1997) mre11S—a yeast mutation that blocks double-strand-break processing and permits nonhomologous synapsis in meiosis. Genes Dev 11: 2272-90

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. Mol Cell Biol 14:1613-25

Nag DK, Scherthan H, Rockmill B, Bhargava J, Roeder GS (1995) Heterodublex DNA formation and homolog pairing in yeast meiotic mutants. Genetics 141:75-86

Neumann R, Jeffreys AJ (2006) Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation. Hum Mol Genet. 15:1401-11.

Novak JE, Ross-Macdonald PB, Roeder GS (2001) The budding yeast Msh4 protein functions in chromosome synapsis and the regulation of crossover distribution. Genetics 158:1013-25

Page SL, Hawley RS (2003) Chromosome choreography: the meiotic ballet. Science 301:785-9

Oudet C, Hanauer A, Clemens P, Caskey T, Mandel JL (1992) Two hot spots of recombination in the DMD gene correlate with the deletion prone regions. Hum Mol Genet 1:599-603

Page SL, Hawley RS (2003) Chromosome choreography: the meiotic ballet. Science 301:785-9

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

Paques F, Leung WY, Haber JE (1998) Expansions and contractions in a tandem repeat induced by double-strand break repair. Mol Cell Biol 18:2045-54

Parvanov ED, Ng SH, Petkov PM, Paigen K (2009) Trans-regulation of mouse meiotic recombination hotspots by Rcr1. PLoS Biol 7:e36

Parvanov ED, Petkov PM, Paigen Kenneth (2010) Prdm9 controls activation of mammalian recombination hotspots. Science 327:835

Petes TD (2001) Meiotic recombination hot spots and cold spots. Nat Rev Genet 2:360-9

Petit C, Levilliers J, Weissenbach J (1988) Physical mapping of the human pseudo-autosomal region; comparison with genetics linkage map. EMBO J 7:2369-76

Pineda-Krch M, Redfield RJ (2005) Persistence and loss of meiotic recombination hotspots. Genetics 169:2319-33

Ponting CP (2011) What are the genomic drivers of the rapid evolution of PRDM9? Trends Genet 27:165-71

Prolla TA, Pang Q, Alani E, Kolodner RD, Liskay RM (1994) MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. Science 265:1091-3

Ptak SE, Hinds DA, Koehler K, Nickel B, Patil, Ballinger DG, Przeworski M, Frazer KA, Paabo S (2005) Fine-scale recombination pattern differ between chimpanzees and humans. Nat Genet 37:429-434

Ptak SE, Voelpel K, Przeworski M (2004) insights into recombination from patterns of linkage disequilibrium in humans. Genetics 167:387-97

Rao E, Weiss B, Fukami M, Mertz A, Meder J, Ogata T, Heinrich U, Garcia-Heras J, Schiebel K, Rappoid GA (1997) Hum Genet 100:236-9

Rattray AJ, McGill CB, Shafer BK, Strahern JN (2001) Fidelity of meiotic doubsle-strand-breaks repair in Saccharomyces cerevisiae: a role for SAE/COM1. Genetics 158:109-22

Rocco V, Nicolas A (1996) Sensing of DNA non-homology lowers the initiation of meiotic recombination in yeast. Genes Cells 1:645-61

Roeder GS (1995) Sex and the single cell: meiosis in yeast. Proc Natl Acad Sci USA 92:10450-6

Roeder GS (1997) Meiotic chromosomes: it takes two to tango. Genes Dev 11:2600-21

Ross-Macdonald P, Roeder GS (1994) Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. Cell 79:1069-80

Sarbajna S, Denniff M, Jeffreys AJ, Neumann R, Soler Artigas M, Veselis A, May CA (2012) A major recombination hotspot in the XqYq preudoautosomal region gives new sights into processing of human gene conversion events. Hum Mol Genet 21:2029-38

Schneider JA, Peto TE, Boone RA, Boyce AJ, Clegg JB (2002) Direct measurement of the male recombination fraction in the human beta- globin hot spot. Hum Mol Genet 11:207-15.

Schwacha A, Kleckner N (1995) Identification of double Holliday junctions as intermediates in meiotic recombination. Cell 83:783-91

Schwacha A, Kleckner N (1997) Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. Cell 90:1123-35

Shannon M, Richardson L, Christian A, Handel MA, Thelen MP (1999) Differential gene expression of mammalian SPO11/TOP6A homologs during meiosis. FEBS Lett 462:329-34

Shinohara A, Ogawa H, Ogawa T (1992) Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. Cell 69:457-70

Shiroishi T, Koide T, Yoshino M, Sagai T, Moriwaki K (1995) Hotspots of homologous recombination in mouse meiosis. Adv Biophys 31:119-32

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. EMBO J 10:68106

Sipiczki M (2000) Where does fission yeast sit on the ree of life? Genome Biol 10.1186

Smith RA, Ho PJ, Clegg JB, Kidd JR, Thein SL (1998) Recombination breakpoints in the human beta-globin gene cluster. Blood 92:4415-21.

Steiner WW, Schreckhise RW, Smith GR (2002) Meiotic DNA breaks at the S. pombe recombination hot spot M26. Mol Cell 9:847-55

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162-9

Storiazzi A, Xu L, Cao L, Kleckner N (1995) Crossover and noncrossover recombination during meiosis: timing and pathway relationships. Proc Natl Acad Sci USA 92:8512-6

Sun H, Treco D, Schultes NP, Szostak JW (1989) Double-strand breaks at an initiation site for meiotic gene conversion. Nature 338:87-90

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. Cell 64:1155-61

Sung P (1994) Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. Science 265:1241-3

Svetlanov A, Cohen PE (2004) Mismatch repair proteins, meiosis, and mice: understanding the complexities of mammalian meiosis. Exp Cell Res 296:71-9

Sym M, Engebrecht JA, Roeder GS (1993) ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. Cell 72:365-78

Sym M, Roeder GS (1994) Crossover interference is abolished in the absence of a synaptonemal complex protein. Cell 79:283-92

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33:25-35

Tapper W, Collins A, Gibson J, Maniatis N, Ennis S, Morton NE (2005) A map of the human genome in linkage disequilibrium units. Proc Natl Acad Sci USA 102:11835-9

Tapper W, Maniatis N, Morton NE, Collins A (2003) A metric linkage disequilibrium map of human chromosome. Ann Hum Genet 67:487-94

Tease C, Hartshorne GM, Hulten MA (2002) Patterns of meiotic recombination in human fetal oocytes. Am J Hum Genet 70:1469-79

Tiemann-Boege I, Calabrese P, Cochran DM, Sokol R, Arnheim N (2006) High-resolution recombination patterns in a region of human chromosome 21 measured by sperm typing. PLoS Genet 2-e70

Trujillo KM, Sung P (2001) DNA structure-specific nuclease activities in the Saccharomyces cerisiae Rad50*Mre11 complex. J Biol Chem 276:35458-64

Tsubouchi H, Roeder GC (2003) The importance of genetics recombination for fidelity of chromosome pairing in meiosis. Dev Cell 5:915-25

Tung KS, Roeder GS (1998) Meiotic chromosome morphology and behavior in zip1 mutants of Saccharomyces cerevisiae. Genetics 149:817-32

Wahls WP (1998) Meiotic recombination hotspots: shaping the genome and insights into hypervariable minisatellite DNA change. Curr Top Dev Biol 37:37-75

Wall JD, Frisse LA, Hudson RR, Di Rienzo A (2003) Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates. Am J Hum Genet 73:1330-40

Wang TF, Kleckner N, Hunter N (1999) Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. Proc Natl Acad Sci USA 96:13914-9

Webb AJ, Berg IL, Jeffreys A (2008) Sperm cross-over activity in regions of the human genome showing extreme breakdowns of marker association. Proc Natl Acad Sci USA 105-10471-6

Weiner BM, Kleckner N (1994) Chromosome pairing via multiple interstitial interactions before and during meiosis in yeast. Cell 77:977-91

Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P, Altshuler D (2005) Comparison of fine-scale recombination rates in humans and chimpanzees. Science 308:107-11

White MA, Dominska M, Petes TD (1993) Transcription factors are required for the meiotic recombination hotspot at the HIS4 locus in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 90:6621-5

Wu TC, Lichten M (1994) Meiosis-induced double-strand-break sites determined by yeast chromatin structure. Science 263:515-8

Wu TC, Lichten M (1995) Factors that affect the location and frequency of meiosis-induced double-strand breaks in Saccharomyces cerevisiae. Genetics 140:55-66

Xu L, Kleckner N (1995) Sequence non-specific double-strand breaks and interhomolog interactions prior to double-strand break formation at a meiotic recombination hot spot in yeast. EMBO J 14:5115-28

Yauk CL, Bois PR, Jeffreys AJ (2003) High-resolution sperm typing of meiotic recombination in the mouse MHC Ebeta gene. EMBO J 22:1389-97

Yip SP, Lovegrove JU, Rana NA, Hopkinson DA, Whitehouse DB (1999) Mapping recombination hotspots in human phosphoglucomutase (PGM1). Hum Mol Genet 8:1699-706.

Zhang W, Collins A, Maniatis N, Tapper W, Morton NE (2002) Properties of linkage disequilibrium (LD) maps. Proc Natl Acad Sci USA 99:17004-7

Zickler D, Kleckner N (1999) Meiotic chromosomes: integrating structure and function. Annu Rev Genet 33:603-754

Zickler D, Olson LW (1975) The synaptonemal complex and the spindle plaque during meiosis in yeast. Chromosoma 50:1-23