Fine-scale analysis of mechanisms and controlling factors in a meiotic recombination hotspot in dogs (*canis familiaris*)

Dissertation

in fulfillment of the requirements for the degree
"*Doctor rerum naturalium*"
of the Faculty of Mathematics and Natural Sciences
at the Christian Albrecht University of Kiel

submitted by

Alina Marie Jeschke

Department of Evolutionary Genetics
Max Planck Institute for Evolutionary Biology

Plön, February, 2020

First examiner: Prof. Dr. Diethard Tautz

Second examiner: Prof. Dr. Eva Holtgrewe-Stukenbrock

Date of the oral examination: 20th May, 2020

Table of contents

Abstract

Meiotic recombination re-shuffles genomes from one generation to the next. In humans and most other mammals, meiotic recombination events are clustered in 1-2 kb wide recombination hotspots, whose locations are determined in *trans* by the protein PR-domain containing 9 (PRDM9). Mice lacking PRDM9 direct recombination to promoters and functional elements, resulting in meiotic defects. Dogs (*Canis familiaris*) lack a functional copy of PRDM9, yet linkage data showed that historical recombination events cluster in functional elements, suggesting that there may be a mechanism enabling controlled recombination at these locations, and in the absence of PRDM9. However nothing is known about the *de-novo* activity of dog recombination hotspots and the patters of recombination resolution in this PRDM9 deficient species. I investigated a dog recombination hotspot for *de-novo* recombination events using pooled sperm typing, and uncovered high crossover frequencies affecting up to 1 % of sperm. Frequencies can differ by one order of magnitude between dogs. Fine-scale analysis of crossover-breakpoints revealed wide distributions of breaks across up to 10 kb within the hotspot region. I further detect asymmetric breakpoint distributions between crossover orientations and crossover-associated transmission distortion, suggesting biased recombination-initiation or -repair. This work is an elaborate fine-scale dissection of a mammalian PRDM9-independent active recombination hotspot.

Zusammenfassung

Durch meiotische Rekombination entsehen neue Allel-Kombinationen in der nächten Generation. Meiotische Rekombination zentriert sich in Menschen und den meisten anderen Säugetieren in 1-2 kb breiten Regionen, die "Rekombinationshotspots" genannt werden und welche durch das Zinkfinger-Protein *PR-domain containing 9* (PRDM9) platziert werden. In Mäusen, die keine intakte Kopie des Gens tragen, findet meiotische Rekombinationsinitiation stattdessen in funktionalen Elementen und Genpromotern statt, was zu meiotischen Defekten führt. Hunde (*canis familiaris*) tragen eine mutierte, nicht funktionale Kopie des Prdm9-Gens, dennoch zeigen genetische Kopplungsdaten ancestrale Rekombinationsaktivität in funktionellen Elementen und Genpromotern. Dies weist auf einen alternativen Mechanismus hin, der die kontrollierte Rekombination an diesen Stellen in Abwesenheit von PRDM9 ermöglicht. Jedoch war zum Zeitpunkt der vorgelegten Arbeit noch nichts über die *de-novo* Rekombinations-Frequenz und Morphologie in diesen PRDM9-defizienten Hunde-Hotspots bekannt. Mittels Spermientypisierung wurde ein Hunde-Rekombinationshotspot analysiert und *de-novo* Rekombinationsfrequenzen von bis zu 1 % in Spermien gemessen. Die Frequenz unterschied sich zwischen den einzelnen Individuen bis zu 10-fach. Hochauflösungs-Untersuchungen der Crossover-Bruchpunkte zeigten eine sehr breite Verteilung von Bruchpunkten über eine Region von bis zu 10 kb in der Hotspot-Region, viel breiter als in PRDM9-regulierten Hotspots. Des weiteren konnten asymmetrische Bruchpunkte zwischen reziproken *Crossover*-Richtungen und nicht-mendelische Übertragung von Allelen in die nächste Generation gemessen werden, was darauf hinweist, dass Rekombinations-Initiation oder –Reperatur asymmetrisch sind. Die vorgelegte Arbeit ist die erste umfangreiche, hochauflösende Analyse eines PRDM9-unabhängigen Rekombinationshotspots in einem Säugetier.

Abbreviations

ASP	allele-specific primer
bp	base pair(s)
CCO	complex crossover
ChIP	chromatin-immuno-precipitation
cM	centimorgans
CO	crossover
dHJ	double Holliday junction
DNA	deoxyribonucleic acid
dNTPs	deoxy nucleoside triphosphates
DSB	double stranded DNA break
dsDNA	double stranded DNA
H3	histone 3
H3K36me3	histone 3 Lysine 36 trimethylation
H3K4me3	histone 3 Lysine 4 trimethylation
HJ	Holliday junction
HMW	high molecular weight
InDel	insertion/deletion
kb	kilo base pairs
LD	linkage disequilibrium
LD	linkage disequilibrium
Mb	mega base pairs
myr	million years
ng	nano gram(s)
PCR	polymerase chain reaction
pg	pico gram(s)
PRDM9	PR-domain containing 9
SDSA	synthesis-dependent strand annealing
SEIs	single-end-invasions
SNP	single nucleotide polymorphism
ssDNA	single stranded DNA
TD	transmission distortion
TF	transcription factor
TSS	transcription start sites
ZF	zinc finger
µl	microliter

1 Introduction

1.1 Meiosis

Sexually reproducing species form gametes during the process of meiosis. During two rounds of cell division the homologous chromosomes within a cell are first segregated into two haploid daughter cells, each of which undergo a second round of cell division where the sister chromatids are separated into the respective daughter cells. Further differentiation leads to the production of four sperm cells deriving from a single cell in males and one oocyte and three polar bodies deriving from a single cell in females.

1.1.1 Phases of Meiosis

Meiosis I, where segregation of sister chromatids takes place, can be sub-divided into several phases.

Prophase I

Prophase I in turn can be sub-divided into several stages.

In Leptotene, chromosomes start to condense to form visible strands, sister chromatids are closely associated and virtually indistinguishable and lateral elements of the synaptonemal complex assemble (Fawcett 1956).

During Zygotene, chromosomes line up to form homologous pairs. The telomeres cluster at the end of the nucleus ("Bouquet stage") and the synapsis is facilitated by central elements of the synaptonemal complex ("zipping"). At the end of zygotene, the chromosomes are lined up in pairs called bivalents or tetrads.

In pachytene, crossovers become visible between homologous chromosomes, an essential process that will be discussed in more detail later. The physical manifestations of crossovers are called chiasmata and these remain thoroughout the next stage of diplotene, where the synaptonemal complex is degraded and the homologues are connected only via chiasmata, allowing DNA transcription in the intervals between them. In mammalian female meiosis (oogenesis), the cells now

enter a resting stage until puberty, whereas male meiosis (spermatogenesis) directly proceeds to Metaphase I.

Metaphase I

The homologues move together along the metaphase plate and paired homologues align along the equatorial plane. Kinetochores pulling the chromosomes towards the poles create a tension antagonized by the chiasmata connecting homologous chromosomes. This tension is essential for correct bi-orientation of the chromosomes and lack thereof can lead to incorrect segregation, causing aneuploidies.

Anaphase I

Kinetochore microtubules shorten to pull the chromosomes towards the poles of the cell. While the cohesin connecting chromosome arms degrades, the centromeres are protected from this process, ensuring the connection of sister chromatids, while homologous chromosomes are pulled apart.

Telophase I

The first meiotic division ends when the chromosomes arrive at the poles. The spindle network disappears, new nuclear envelopes form and the chromosomes decondensate.

During Meiosis II the cell undergoes similar steps, yet here sister chromatids are separated into daughter cells, resulting in a total of four daughter cells deriving from one meiotic cell.

1.2 Meiotic Recombination

During meiosis the genetic content of an individuals offspring is determined. In addition to segregation of homologous chromosomes, there is also intense re-shuffling of genetic information facilitated by meiotic recombination between homologous chromosomes. Linkage between genetic loci is measured in centimorgans (cM), defined as the distance between loci for which the expected number of crossovers between chromosomes is 0.01. To translate the genetic

distance into physical distance, centimorgans per mega base pairs (cM/Mb) can be read as the number of meiotic homologous crossovers per 100 meioses within a physical distance of 1,000,000 base pairs.

Recombination Hotspots

In mammalian genomes, the sites where meiotic recombination takes place are not randomly distributed but cluster into 1-2 kb wide regions of intense recombination activity, called recombination hotspots (Jeffreys, Murray et al. 1998), (Jeffreys, Kauppi et al. 2001). The placement of these hotspots has great consequences to the organism and the submission of genetic material to the next generation. Mechanistically, a minimal number of meiotic crossovers are needed to ensure correct synapsis during meiosis (Murray and Szostak 1985), yet a controlling mechanism called "crossover interference" prevents the placement of additional crossovers close to an already existing crossover site (Sturtevant 1915, Muller 1916).

From an evolutionary perspective, recombination hotspots are intersections between linkage blocks in the genome. They allow for the genomic re-shuffling by creating new allelic combinations of the genes on either side of a hotspot. Meiotic recombination requires the controlled formation of a double stranded DNA break (DSB), a process likely linked to increased exposure to mutation (Arbeithuber, Betancourt et al. 2015) and often results in stretches of gene conversion, which is the non-reciprocal transfer of information from one homologue to the other, without exchange of flanking markers.

The control of hotspot site placement has multiple layers. Firstly, their position has a cytological function during meiotic segregation. Secondly, hotspot sites are exposed to increased mutation rates and have to be placed such that this does not have detrimental effects, for example by disrupting a gene. Thirdly, on a population genetic level, hotspots change the distribution of linkage blocks within a genome. It is unsurprising therefore, that the placement of meiotic recombination sites is tightly controlled.

1.2.1 Mechanism and controlling factors of mammalian recombination

The mechanism of meiotic recombination can roughly be distinguished into three parts. Firstly, meiotic recombination initiation and DNA double strand break formation; secondly, gap extension and recombination intermediate formation and finally, resolution of intermediates and double strand break repair.

1.2.2 Step one: Mammalian Recombination initiation

During zygonema of prophase I, chromosomes line up to form homologous pairs. For this process, reciprocal homologous recombination is essential in that it facilitates homology search and pairs homologous parts of the chromosomes (Cao, Alani et al. 1990).

In many taxa including most mammals, the zinc-finger protein PRDM9 defines the location of meiotic recombination initiation (Baudat, Buard et al. 2010, Berg, Neumann et al. 2010, Myers, Bowden et al. 2010, Parvanov, Petkov et al. 2010).

The zinc-finger domain of PRDM9 recognizes and binds specific DNA-sequences in the genome (Baudat, Buard et al. 2010, Grey, Barthes et al. 2011, Brick, Smagulova et al. 2012, Billings, Parvanov et al. 2013) and trimethylates Lysine 4 and Lysine 36 on the same histone 3 (H3) molecules (Baker, Walker et al. 2014, Powers, Parvanov et al. 2016) in close proximity to its own binding site. This signal recruits a number of other proteins, which then place a DSB in the resulting nucleosome depleted region (Baker, Walker et al. 2014).

The evolutionarily conserved Spo11 protein, a type II topoisomerase-like protein, forms DSBs, binding DNA as a dimer and creating a DSB in concert with several other proteins (Keeney, Giroux et al. 1997, Keeney 2008).

PRDM9

The key player in meiotic recombination initiation is PRDM9, a histone methyltransferase, which consists of three major domains (Baudat, Buard et al. 2010, Parvanov, Petkov et al. 2010). The widely conserved N-terminal part comprises a KRAB-domain (Krüppel-associated box domain) and a PR/SET-domain account for the methyltransferase activity (Hayashi, Yoshida et al. 2005). The C-terminal part is made up of a long zinc-finger array. The zinc-finger array is highly

polymorphic between species and even within the same species, differing in the zinc-fingers themselves, the numbers of zinc fingers and thus lengths of the array an the composition (Berg, Neumann et al. 2010, Myers, Bowden et al. 2010, Berg, Neumann et al. 2011).

The binding site of PRDM9 in the genome is defined by the respective sequence motif preferentially bound by three positions (-1, 3 and 6 in mice) of every zinc-finger of PRDM9 (Choo and Klug 1994, Choo and Klug 1994) which have evolved rapidly under positive selection over a long period of time (Oliver, Goodstadt et al. 2009).

In effect, even small changes in the zinc-finger array can activate a completely new set of recombination hotspots (Berg, Neumann et al. 2011).

PRDM9 places unique trimethylation marks

To draw DSB initiation away from other H3K4me3 marks, PRDM9 creates a unique signature by placing H3K4me3 as well as H3K36me3 marks on the same histones adjacent to its own binding sites. These double-methylation marks are highly enriched in recombination hotspots and PRDM9 is the only described mammalian methyltransferase that can place both these marks (Powers, Parvanov et al. 2016).

Structural arrangement of DNA and chromosome axis during recombination initiation

PRDM9's role in recombination initiation is not purely to define the sequence at which recombination occurs but it has a role in defining the structural arrangement of the condensed DNA during meiotic prophase I as well. During Leptonema chromosomal DNA condenses, forming a chromosomal axis (J.M. 1956). This axis is made up of the synaptonemal complex, a core structure around which the DNA is arranged in loops (Fawcett 1956, J.M. 1956, Møens and Pearlman 1988, Zickler and Kleckner 1999). Already in the 1970s it was established that meiotic recombination takes place in "recombination nodules" associated with the chromosome axis (Carpenter 1975). A recent study found evidence for a model in which PRDM9-positioned trimethylation marks are sufficient to recruit other players, possibly CxxC ZF proteins, which pull these PRDM9-associated DSB sites

towards the chromosomal axis. In the absence of PRDM9, CxxC ZF proteins direct methyltransferases to DNA sites including unmethylated CpG-islands. These sites are also marked with H3K4 trimethylated histones and mark sites of active transcription. In the absence of PRDM9, these sites could be directed to the chromosomal axis instead of PRDM9-binding sites and DSBs might instead be placed there(Brick, Smagulova et al. 2012, Diagouraga, Clement et al. 2018).

1.2.3 Step two: Gap extension and strand invasion

Spo11 remains covalently attached to the 5'-ends of the DSBs, which are then subject to 5'-3' resection. This process involves initiation by the MRX complex, containing Mre11 (Gravel, Chapman et al. 2008, Mimitou and Symington 2008, Zhu, Chung et al. 2008) creating entry sites for exonucleases like Exo1 and/or Sgs1-Dna2 which resect dsDNA (double stranded DNA) to ssDNA (single stranded DNA) in 5'-3' orientation (Zakharyevich, Ma et al. 2010, Garcia, Phelps et al. 2011, Cannavo and Cejka 2014, Mimitou, Yamada et al. 2017). The resulting ssDNA overhangs serve as templates for DMC1 (Bishop, Park et al. 1992, Yoshida, Kondoh et al. 1998) and Rad51, whereas Rad51 uses the sister-chromatid-associated end and DMC1 uses the homologous-chromosome-associated end. DSB repair during meiosis almost exclusively uses the homologue as a template for repair, thus the presence of DMC1 is used as a proxy for meiotic DSBs (reviewed in (Hong, Sung et al. 2013)).

DMC1 facilitates the homology search and strand exchange through a number of other proteins which will not be discussed here (reviewed for example in (Hunter 2007)), causing strand invasion of the DMC1-processed 3'-ssDNA-end of a homologous region of the template (homologous chromosome) duplex, forming a D-loop (Szostak, Orr-Weaver et al. 1983). This intermediate containing the un-cleaved template D-loop with one ssDNA-strand from the DSB annealed to it can give rise to several different downstream intermediates.

1.2.4 Step three: Recombination intermediates and final DSB repair

When one 3'-end of the DSB DNA has invaded one strand of the D-loop of the homologous chromosome, three strand ends remain free, one 3'-end from the

other strand and two 5'-ends, all of which have to be joined again and the missing sequence from the gap has to be repaired in order to re-form an intact DNA molecule. The most prominent intermediate to accomplish this is the double-Holliday-junction (dHJ) (Schwacha and Kleckner 1995, Holliday 2009), which can give rise to crossover- and non-crossover recombination and will be discussed below. An alternative mechanism of repair is explained by the synthesis-dependent strand annealing (SDSA) model, creating non-crossover recombinant molecules, which will not be discussed in detail here (reviewed in (Allers and Lichten 2001)).

Double Holliday-junction (dHJ) formation

The D-loop structure is solely a theoretical assumption of the stage that results when the D-loop is extended due to DNA-synthesis primed by the invading 3'-prime end (reviewed in (Hunter 2015)). The detectable intermediates are called single-end-invasions (SEIs), where one end of the DSB is still tethered to the axis, while the other has interacted with its homologue (Hunter and Kleckner 2001). DNA-synthesis then facilitates branch migration at the end of which the nascent strand is displaced and anneals to the second DSB-end. After another round of DNA-Synthesis a double-Holliday-junction (dHJ) is formed (Schwacha and Kleckner 1995, Lao, Oh et al. 2008).

dHJ-resolution

The final step of homologous recombination requires the resolution of intermediates to restore the two chromatids involved and fully repair the initiating DSBs. A dHJ has four potential cleavage sites representing the four strands involved in the dHJ. Each HJ can be repaired either by cleaving the crossed strands (non-crossover mode) or the non-crossed strands (crossover mode). Only when one of the two HJs is resolved in the non-crossover mode and the other in the crossover mode a full crossover occurs. Otherwise the chromatids will be restored without crossing over but containing stretches of gene conversion from gap repair.

1.2.5 Crossover Interference

Crossover are prevented from being spaced closely together by crossover interference (Sturtevant 1915, Muller 1916). There are several proposed models as to how crossover interference is facilitated. The first one was proposed by Nancy Kleckner (Kleckner, Zickler et al. 2004), and is based on the assumption that meiotic chromosomes are subject to mechanical tension or stress. The model assumes that common precursor sites (DSBs) can later either become a crossover or non-crossover. Once a site is designated to become a crossover due to increasing stress, the tension reduces to zero at this site, so the next site to be designated as a crossover will be located at a distance, where tension is still high.
Another model, advocated by Franklin Stahl, proposes that crossovers as well as non-crossovers derive from the same meiotic intermediates and that there is a fixed number of non-crossover events between any two crossover events (Foss, Lande et al. 1993).

Non-interfering crossovers

While the majority of crossovers are subject to crossover interference, a small number of crossovers seems to be non-interfering (class II) and can occur in close proximity to interfering crossovers (class I). These might arise from an altogether different recombination pathway called dissolution, where nicking of incomplete HJs could lead to crossovers without dHJs (Osman, Dixon et al. 2003).

1.2.6 Global regulation and chromatin status

Apart from the very fine-scale perspective of recombination intermediates and their regulators, recombination hotspots are as well controlled on a global scale across the genome. DSBs only occur in contact with the chromosome axis and not in the loops arranged around it, but loops can be recruited to the axis by specific complexes. Crossover interference limits crossovers that can be closely spaced. At the same time, there must be a mechanism in place to ensure that enough crossovers occur per chromosome, or chromosome arm, to ensure correct pairing (Murray and Szostak 1985).

All meiotic recombination hotspots are associated with epigenetic structural features. DSBs occur in close proximity to trimethylated histones (H3K4me3). In PRDM9-regulated hotpots, PRDM9 binds to a specific sequence motif, trimethylates the adjacent histones and causes remodeling of these histones to create a central nucleosome-depleted region where the DSBs occur (Brick, Smagulova et al. 2012, Baker, Walker et al. 2014). In PRDM9-independent recombination events, meiotic recombination DSBs are also placed near histones carrying H3K4-trimethylation. These are marks placed by other methyltransferases, for example at DNA-transcription start sites, creating accessible chromatin for Spo11 to place DSBs (Brick, Smagulova et al. 2012).

1.2.7 DSB repair of "default" initiation sites

The placement of meiotic DSBs in functional elements, independently of PRDM9, is called "default" initiation.

In PRDM9-carrying species, the disruption of functional elements and transcription by meiotic recombination DBSs can have a gradient of effects on meiosis depending on the underlying conditions. While one study finds that mice lacking functional PRDM9 have problems repairing breaks in functional elements, leading to severe meiotic defects and rendering them infertile (Brick, Smagulova et al. 2012), other data point towards a more diverse fertility phenotype (Davies, Hatton et al. 2016, Mihola, Pratto et al. 2019). First of all, even when functional PRDM9 is present, there seem to be default initiation sites concurrently, with their proportion increasing towards later stages of meiosis (Smagulova, Brick et al. 2016). There has also been a study with one human female lacking functional copies of PRDM9, who was fertile (Narasimhan, Hunt et al. 2016) and a very recent study found that snakes place meiotic DSBs in functional elements even in the presence of functional PRDM9 (Schield, Pasquesi et al. 2020). DSB repair in functional elements is not impaired in *Saccharomyces cerevisiae*, where meiotic recombination takes place in the nucleosome-depleted region adjacent to transcription start sites.

It appears that using default initiation sites is not altogether detrimental but less advantageous than using a specifically marked site for recombination rather than existing accessible chromatin.

From an evolutionary perspective this would be advantageous from two angles. Firstly, even mild degrees of asynapsis can cause severe phenotypes in the offspring and disrupting active transcription might have detrimental effects other than those related to synapsis. Secondly, a sequence-specific regulator of recombination landscapes poses an easy target for evolutionary forces to modulate linkage.

1.3 Meiotic Recombination in the context of evolution

Why sexual reproduction and correspondingly meiotic recombination have evolved to be ubiquitous in almost all higher organisms remains an unanswered biological enigma. Firstly, any sexually reproducing organism is subject to the cost of meiosis, referring to the fact that while any clonally reproducing organisms will contribute 100% of their offspring's genotype, a sexually reproducing organism confers only half its genetic content to its offspring (Williams 1975). Secondly, meiotic recombination breaks up linked genes. While it might be advantageous to create new allelic combinations in a phase of adaption, once a fitness peak is reached the breaking up of the "winning" combination of alleles should be a disadvantage (Otto 2009) and is often mildly deleterious. However, the benefits of sexual reproduction seem to outrun the disadvantages. Most newly acquired mutations are deleterious, and while they might accumulate in an asexually reproducing population (Muller's ratchet, (Muller 1964)), they can be purged more efficiently with recombination. Likewise, hitchhiking of mutations with a small deleterious effect, which are linked to a beneficial mutation that is selected for, can often be prevented by recombination.

Recombination hotspots are sites within the genome at which breaking up of linkage blocks occurs. Their placement has a most important implication for all the above-mentioned population genetic effects of meiotic recombination.

2 Thesis aim

PRDM9 can alter the sites at which recombination happens and draw recombination away from functional elements associated for example with transcription, thus protecting active sites from breaking up and genes from being disrupted. In the simplest eukaryotic cell, the yeast cell, recombination is drawn towards these functional sites, whereas in higher complexity genomes like those of mammals, PRDM9 directs recombination away from these sites.

Dogs and other canids have mammalian complex genomes, yet they have lost the regulator PRDM9. This thesis aims at uncovering some of the underlying mechanisms that characterize dog meiotic recombination in functional elements and in the absence of PRDM9.

3 Methods

3.1 Introduction

3.1.1 Adapting techniques for a non-model system

The analysis of recombinant molecules from germ cells has helped greatly in understanding the process of meiotic recombination ((Jeffreys, Murray et al. 1998, Jeffreys, Kauppi et al. 2001, Berg, Neumann et al. 2010, Cole, Baudat et al. 2014) and others). Most studies to date have focused on model organisms such as yeast and mice because these are available, reasonably easy to handle and it is possible to generate biological replicates. The analysis of human samples is possible in this case as well, because at least male germ cells (sperm) are easily accessible without harm to the donor.

The methods used to analyze meiotic processes, like sperm typing (Jeffreys, Holloway et al. 2004, Kauppi, May et al. 2009), ChIP-Sequencing of meiotic proteins (Pan and Keeney 2009, Smagulova, Gregoretti et al. 2011, Khil, Smagulova et al. 2012), immune-cytological analyses of germ line cells (Zickler and Kleckner 1999, Lipkin, Moens et al. 2002, Cole, Baudat et al. 2014) and whole genome linkage analysis (Auton, Rui Li et al. 2013, Pratto, Brick et al. 2014) are well-published and accepted in the field.

The analyses for the present thesis were performed in a non-model organism, the domestic dog, because it provides a unique natural knockout-system for PRDM9 in mammals (Oliver, Goodstadt et al. 2009, Munoz-Fuentes, Di Rienzo et al. 2011). The natural loss of Prdm9 in canid species must have lead to, or have been preceded by alternative mechanisms to fulfill the function PRDM9 carries out in other mammals. This unique natural loss thus cannot be mimicked by an artificial knockout in a model species. A crucial part of this thesis was to adapt the well-accepted published methods to a non-model system, the domestic dog.

3.1.2 Novel technologies becoming available

The advance in DNA-sequencing technologies has been rapid within the past years, providing many novel approaches, which had not been available before. In the analysis of meiotic recombinant molecules this holds especially true for third-generation sequencing because it provides long and, more importantly, phased sequences. Sperm typing has previously included radioactive hybridization using labeled oligotides to provide phased information on marker SNPs (Kauppi, May et al. 2009). Apart from safety-issues with using radioactivity, this method has a great caveat when using non-model organisms. It requires the previous knowledge of the present heterozygous sites and thus requires polymorphism-data as well as either pedigree data or else further phasing experiments.

Recently, recombinant molecules have been analyzed by Sanger-sequencing of single-molecule amplified PCR-products (Arbeithuber, Betancourt et al. 2015). With this method it is possible to analyze SNPs that were not collected on polymorphism data bases before and further to analyze *de-novo* mutations occurring during meiotic recombination. The method is limited to analyze regions within the range of one sequencing run however, in order to not loose phase-information, which can provide a maximum of ca. 1-2 kb when using overlapping primers. This is a sufficient range to pick up >95% of recombination events in PRDM9-dependent hotspots, which are generally 1-2 kb wide.

Third-generation sequencing can greatly increase the length of sequence covered, by providing very long phased reads (up to whole chromosome-reads in the case of the Oxford Nano-pore technology), however they have very high error rates of >10%, which need to be taken into account.

3.2 Aim

The work presented in this chapter aimed at developing an approach to analyze long PCR amplicons from sperm-typing PCRs accurately in a non-model organism (domestic dog) using Sanger- and Third-generation sequencing.

3.2.1 Optimization PCR buffer system

The commonly used buffer system for recombinant analysis was established in Alec J. Jeffreys' laboratory (Kauppi, May et al. 2009). It uses a combination of a highly efficient Taq-polymerase with a polymerase with proofreading activity (pfu) to ensure allele-specific and long-range amplification. It is reported to work for long amplicons (>10kb) with short allele-specific primers (ca. 18 bp) and with input ranging from very high to as low as single-molecule PCRs.

Trials to establish a long-range specific assay for the present analysis of recombinants in the domestic dog using this system yielded very poor results. Short amplicon allele-specific PCRs (<4 kb) yielded good results, however long amplicon PCRs lacked efficiency, with no PCR product visible on agarose gels, while the same buffer batch produced good results in control long amplicon PCRs on mouse DNA. It is possible, that this was due to the different quality of obtained high molecular weight (HMW) DNA from dogs. While even very highly concentrated sperm DNA obtained from mouse epididymis would always completely dissolve, the sperm DNA obtained from dogs was far more viscous and would remain heterogeneous, even in larger volumes of buffer or water. Re-precipitation of the dog DNA improved the purity (measured in a photo-spectrometer), yet viscosity did not change.

Consequenctly, a different buffer system was established, using a polymerase of high fidelity (TaKaRa LA Taq polymerase). Other polymerase-systems (Kapa HiFi, Phusion HiFi) were tested but they either lacked allele-specificity or efficiency in the long-range assays. It should be noted that the region to be amplified is of low complexity, with several repetitive elements, micro-satellites and long stretches of very low GC-content, all of which provide difficult sequences for polymerases to amplify.

The TaKaRa LA system yielded good results with intermediate length primers (between 18 and 22 bp) and with input ranging from very high DNA input to low input (5 to 10 molecules). Single-molecule PCRs, where the polymerase is presented with statistically just 0.5 DNA-molecules, were unsuccessful at least using dog DNA. Notably, this refers to the total input of DNA, including

recombinants as well as parental haplotypes. The polymerase proved to be single-molecule specific in the sense that the sole "correct" recombinant molecule could be amplified within a pool of parental haplotype DNA. False positive rates were tested by using somatic tissue controls, as well as homozygous template DNA for the respective opposite haplotype per allele-specific SNP. Over 400,000 molecules were screened in the somatic controls and ca. 3000 molecules of opposing haplotypes, none of which yielded a positive PCR reaction, corresponding to a combined effect of unspecific primer binding and template switching (for more detail see 3.2.5).

3.2.2 Optimizing long amplicon sequence analysis

Most prior analyses of recombinant molecules were performed using allele-specific radioactive hybridization on PCR products from allele-specific PCRs. In recent years, DNA-sequencing methods have become more readily available and less costly, opening the chance to determine crossover resolution points based on PCR-product sequencing.

In a typical mammalian recombination hotspot, over 95% of the crossover resolution takes place within 1000 bp up- and downstream of the hotspot center. Depending on the complexity of the DNA, a Sanger-sequencing run will result in up to 1000 bp of sequence, so that most mammalian hotspots can be sequenced fully, using overlapping sequencing-primers, starting from the hotspot center towards the up- and downstream boundaries.

Assuming that dog recombinant molecules would likewise contain most of the resolution points close to the center, it was first attempted to use the same method. The results were disappointing; although some haplotype switches could be detected most of the recombinant molecules appeared to show no switch in haplotype.

3.2.3 Single-recombinant Sanger-sequencing across 10.8 kb

Next, sequencing-primers across the whole amplified region were established, aiming at creating overlapping sequencing-reads.

A problem occurring with this method is, that in order to determine the switch from one haplotype to the other, the reads need to be phased, meaning that the sequences obtained from all primers must come from the same PCR reaction. The standard protocol used for in-house sequencing uses 1 µl of PCR product per sequencing reaction (equivalent to a ca. 800-1000 bp segment of the hotspot). All tested buffer systems yielded poor results when the total PCR reaction volume exceeded 15 µl. In order to define positive PCR reactions from the assay, at least 5 µl of the PCR product were needed to be loaded on a gel. From the repeated heating and cooling during a PCR reaction, the volume within a well of the PCR plate decreases by at least 1-2 µl, leaving a maximal amount ca. 8 µl to be used for sequencing, which was not enough to cover the whole hotspot region.

In order to overcome this problem, a new procedure was established. Low-melting agarose was used instead of using standard agarose, gels were photographed and positive PCR reactions were cut out separately. PCR products were recovered from the gel by melting the cut-out pieces (separately) and digesting the agarose enzymatically using agarase. This way enough PCR-product volume was generated to run multiple sequencing-runs from the same PCR reaction.

This method only worked on PCR reactions yielding very high amounts of DNA. A majority of the positive PCR reactions yielded intermediate or low amounts of DNA, which were still visible on the agarose gel, but did not supply enough DNA material to be sequenced. 96 recombinants with sufficient sequencing quality were obtained in total.

However, the very low complexity of the region lead to gaps in the assembly, because Sanger-sequencing is very error-prone when presented with templates that contain satellite DNA. “Polymerase slippage” created reads of heterogeneous signal that could not be analyzed and if two low-complexity stretches were too close together no sequencing primer could be found that would produce an unambiguous sequence to close the gap between them.

3.2.4 Long-read single recombinant analysis using third-generation sequencing

In order to close the gaps in the assembly, trials to establish an assay using long-read sequencing were conducted. The obvious advantages of these methods is that they can produce single sequencing reads of very long lengths (>> 10 kb) and they are not impeded by low-complexity regions, as was observed in Sanger-sequencing. The problem with third-generation sequencing is their very high per-base error rate. This rate has improved since their introduction but is still well above 10% for all methods available to date.

The solution to the very high error rate lies in higher coverage. In the allele-specific assays the input of sperm DNA is optimized such that in most reactions only one recombinant molecule is amongst many parental haplotype molecules. The PCR products after two-rounds of nested allele-specific PCR are all copies of the same single recombinant molecule from the original input pool. By sequencing multiple copies of this recombinant from the same PCR pool real haplotype switches, which occur at known heterozygous marker sites, can be distinguished from randomly distributed sequencing errors.

The sequencing error is not uniform but depends on the template DNA's complexity and GC content. The available sequencing data from Sanger-Sequencing of Crossovers as well as phased Haplotypes were used to establish a minimal coverage needed for the MinION-sequencing technology (described in paragraph 3.10).

PacBio SMRT bell library sequencing

A commercially available system from PacBio was used for the first attempt at third-generation sequencing for recombinant analysis, which would be analyzed by a collaborator. The challenge was to use multiplex adapters so that every read could later be assigned to a single PCR reaction.

There are several reasons why this attempt did not work out in the present case, although in retrospect some of them proved to be biologically true rather than errors in the methods.

Firstly, while the mean length for PacBio sequencing at the time was ca. 10 kb, the distribution of read lengths was problematic. Most of the reads acquired were shorter (ca. 3-4 kb) and some reads were very long (> 50 kb) and while these very long reads might be desirable in whole genome *de-novo* assemblies, in this case they most probably represented errors, during sequencing preparation (ligation of adapters), or possibly occurring during the PCR reaction by template switching, or unspecific amplification. The number of usable reads of the desired length (10.8 kb) was much lower than expected.

Another problem was, that the submission criteria for sequencing were highly impractical. The required total volume was 4 µl, the required amount of DNA was at least 1 ng. In order to concentrate such a high amount of DNA into such a small volume PCR-products had to re-precipitated, dried and eluted in a very small volume.

The third problem was that a collaborator based in the USA did the bioinformatics analyses, so any changes in the analysis procedure (for example discarding too short and too long reads) took more time compared to a pipeline where all laboratory steps and analyses are done in-house.

The results obtained from this method looked very different from the expectation. Far from finding a single switch point in each of the molecules, there was a very high number of reads that carried more than one haplotype switch and resolution points were not concentrated around the hotspot center. The same result was later supported by other sequencing analyses, but led to trying a different method at the time.

Oxford Nanopore sequencing using MinION

Using the MinION sequencer held many advantages over the PacBio technology and led to very good results after optimization. First of all, the whole pipeline, including sequencing and analysis, could be done in-house. Secondly, there is an optimized multiplex-adapter kit available which can be extended to a total of 24 adapters, which are designed to be unique even when allowing for some error within the barcode sequence, which greatly increases the recovery of labeled

reads after sequencing. Thirdly, the input volume is highly flexible because of a bead-based purification step, which made it possible to use the total PCR volume without any loss of material and yielded good results even for PCR reactions with low final yield. And finally, the method provided with a sufficient number of reads in the requested length of ca. 10.8 kb, with recovery of reads being solely dependent on the input (i.e. PCR reaction yield).

3.2.5 PCR errors in sequence-based recombinant analyses

There are different ways in how PCR results can be biased and not represent the underlying template DNA. Some of them, like template bias and PCR stochasticity, are not relevant to the present setup, because the system is set up so that every PCR reaction uses a single molecule as a template. The two types of relevant errors are "template switching", where the polymerase jumps from one template to another without releasing the first DNA strand, and "amplification errors", where the polymerase incorporates a false base pair.

Template switching

Template switching requires the physical distance of two DNA template molecules to be small, in order for the polymerase to switch between them and thus increases with increasing DNA concentration. Within the first cycles, DNA concentration is very low and it exponentially increases with every cycle. Template switching is very rare in early cycles and becomes more abundant only in later cycles, when its effect decreases. A study testing the effect of PCR errors found the fraction of molecules arising from template switching to be in the order of magnitude of 10^{-5} to 10^{-3} (Kebschull and Zador 2015).

Amplification errors

The manufacturers claim that the error rate of TaKaRa LA Taq is 6.5 times better than that of Taq polymerase. Taq polymerase has error rates of ca. 10^{-4} (Kebschull and Zador 2015), leaving an error rate of ca. $1.5 \cdot 10^{-5}$ for TaKaRa LA Taq. Due to a formula used in the above-mentioned study ((Kebschull and Zador 2015),Formula 10 and 11) this would result in a probability of 0.14 that a molecule contains at least one amplification error, considering a 10 kb PCR amplicon. However, by

using a pre-defined set of 78 markers, the number of sites at which an error is relevant reduces dramatically. The probability that an amplified molecule has at least one incorrectly amplified base at one of the marker sites is $1.2 \cdot 10^{-3}$. Thus, one incorrectly amplified marker was expected in every 833th molecule amplified. An amplification error will only be problematic if it occurs in one of the first cycles and is amplified to at least the same coverage as the correct molecule. In a perfect single-molecule PCR, after the 3rd cycle the copy number of the target should be eight copies. Any incorrectly amplified molecule occurring in later cycles would lack behind three cycles and is unlikely to reach the same abundance as the correct target (Kebschull and Zador 2015). If every 833rd molecule includes an amplification error and eight molecules are produced within the first three cycles, it would be expected that every 833rd reaction harbors an incorrectly amplified marker created within the first cycle, every 278th PCR reaction includes a molecule incorrectly amplified within the first two cycles and every 119th PCR reaction to include a molecule with an incorrectly amplified marker within the first 8 molecules amplified. Thus less than 1% of PCR reactions are assumed to be false positive due to amplification errors.

3.3 Final methods

3.4 Samples

Sperm samples were obtained from several different sources. Most samples came from breed dogs (supplied by Dr. Axel Wehrend, at the Veterinary Medicine Hospital in Giessen, Germany https://ebvs.eu/colleges/ECAR/members/prof-axel-wehrend) sperm from ejaculates was stored in insemination tubes in buffer at −196 °C and shipped in liquid nitrogen. Other samples were collected from epididymis of castration waste. Local veterinarians kept and stored whole testicles from castrations (at -20 °C). The remaining samples were obtained directly after castrations, where tissue could be snap-frozen within minutes of detachment (Supplementary information S1).

3.5 High molecular weight DNA (HMW DNA)

For sperm typing it is important to extract whole intact chromosomal DNA. This was done using a modified version of the published protocol by Francesca Cole and Maria Jasin (Cole and Jasin 2011). Obtained DNA was re-precipitated for purification (protocol supplied in supplementary information S2).

3.6 Multiple-displacement amplified DNA

For genotyping of the hotspot-flanking regions and primer testing whole genome amplified DNA (Qiagen Repli-g Single Cell Kit) was used.

3.7 Candidate hotspot

An appropriate candidate hotspot for historical recombination was identified by several criteria. A list of hotspots present in both available historical dog recombination data sets ((Axelsson, Webster et al. 2012, Auton, Rui Li et al. 2013)) was created and filtered further for hotspot heat. The heat per hotspot differed between data sets, and the 25 hotspots, which showed highest heat in both data sets, were chosen. Their positions were transferred to the most recent dog genome assembly (CanFam broad assembly 3.1). Those hotspots, which were close to assembly-gaps were discarded from the list and a panel of top ten hotspots, all of which had an annotated CpG-island was created. The hotspot in the present study was then picked because of the high marker density both within the hotspot region as well as in the hotspot flanking regions.

3.7.1 Genotyping of hotspot-flanking regions

To find informative heterozygous SNPs in the hotspot-flanking regions specifically for each individual, the flanking 3 kb both upstream and downstream of five candidate hotspots were amplified and PCR products were sequenced using Sanger-technology. Sequences were analyzed in Geneious program (Version 9 and Geneious Prime) using the "Heterozygotes"-plugin. Every heterozygous call was manually verified by checking the Sanger signal peak to eliminate false calls.

3.8 Recombination assays

The concept of allele-specific repulsion phase PCR was adapted from the method published by Liisa Kauppi, Cecila May and Alec Jeffreys (Kauppi, May et al. 2009).

3.8.1 Buffer system

In all PCR assays the TaKaRa LA Taq polymerase was used with the respective commercially available buffer system (PCR Buffer II (Magnesium-free) and dNTPs in the suggested concentration), however the amount of polymerase was reduced to a final concentration of 0.05 U/µl, because this resulted in higher specificity in the allele-specific assays with a reasonable efficiency. Shorter primers than suggested (18-22 bp instead of 20-25 bp) were used to ensure allele-specificity. The extension temperature was reduced to 68°C, and extension time to t=30seconds/kb+60seconds, improving specificity.

3.8.2 Allele-specific primers (ASPs)

To amplify recombinants, primers were designed that are specific to the haplotypes at heterozygous loci respectively. Primers differed at the most 3'-base targeting the heterozygous locus. Specificity of the primers was tested on individuals homozygous at the given locus and a universal complimentary primer. Temperature gradient PCRs for each candidate primer were conducted to find the optimal annealing temperature at which the primer would efficiently amplify the target allele but not amplify the opposite allele (Supplementary information S3, S14 and S15).

3.8.3 Phasing

From ASP testing a list of possible marker SNPs for the recombination assay were obtained and the panel of individuals screened for those that would be heterozygous at least at two of those markers both up- and downstream of the hotspot.

Since the two upstream and downstream SNPs were close to each other (within 1 kb) their phase could be inferred from the homozygous individuals of the panel. However, if there were doubts about the phase within upstream or downstream

blocks, an additional PCR was run with the outside allele-specific primer and a universal primer close by (within 2-3 kb) and sequenced across the inner allele-specific primer using Sanger-sequencing.

In order to know the phase across the hotspot, long PCRs on sperm DNA across the region, using all possible combinations for two of the informative SNPs (one upstream and one downstream) were conducted (Figure 1).

To avoid producing a PCR product that would serve as a template for the Crossover-assay and thereby posing a threat of contaminating the assay, the inner Forward ASP was tested with the outer Reverse ASP and vice versa. This leads to eight primer combinations to be tested of which only half of the reactions yield a positive result if the primers are efficient and specific (Figure 1).

The PCR bands from the phasing were cut out and used for both Sanger and long-read MinION sequencing to get extensive information on the phased haplotypes across the hotspot region (also see supplementary information S16).

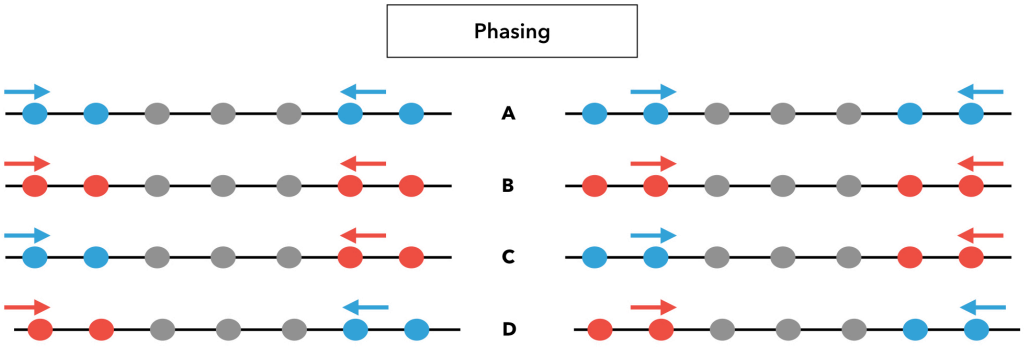


Figure 1 Phasing of Parental Haplotypes. Alleles per SNP are assigned arbitrarily to "blue" or "red" and every combination of Up- and Downstream SNPs is tested. Two sets of Haplotypes are possible (A and B or C and D). All individuals included in the present study had haplotype set A and B.

3.8.4 Crossover assays

The concept of small pool sperm typing crossover recombination assay is to recover the very small proportion of recombinant molecules at a given recombination site from the bulk of parental haplotype molecules. The proportion of crossover recombinant molecules can differ between recombination hotspots and between individuals, yet have been reported to be well below 10% for mammalian hotspots (for example (Jeffreys, Holloway et al. 2004, Webb, Berg et

al. 2008, Cole, Keeney et al. 2010, Berg, Neumann et al. 2011)). To recover those, the assay uses two rounds of nested PCRs, where the forward primers are allele-specific to one haplotype, the reverse primers to the other. Batches of these PCRs are performed on increasingly diluted pools of high molecular weight sperm DNA. By assessing the overall input number of haploid sperm genomes and assuming that the DNA is distributed into the multiple PCR reaction wells following Poisson distribution, it is later possible to infer the underlying number of recombinant molecules within the input pool and thus the *de-novo* crossover recombination frequency (Figure 2).

Pilot assays covering a broad range of input molecules were performed first to define an optimal input amount. Assays were then performed in six pools per individual and direction, giving a total of 144 independent reactions per individual. Samples were run on 1% low melting agarose gel electrophoreses and visualized using SybrSafe and a UV-light imager with blue-screen conversion. Visible bands were cut out from the gel for sequencing (also see supplementary information S17).

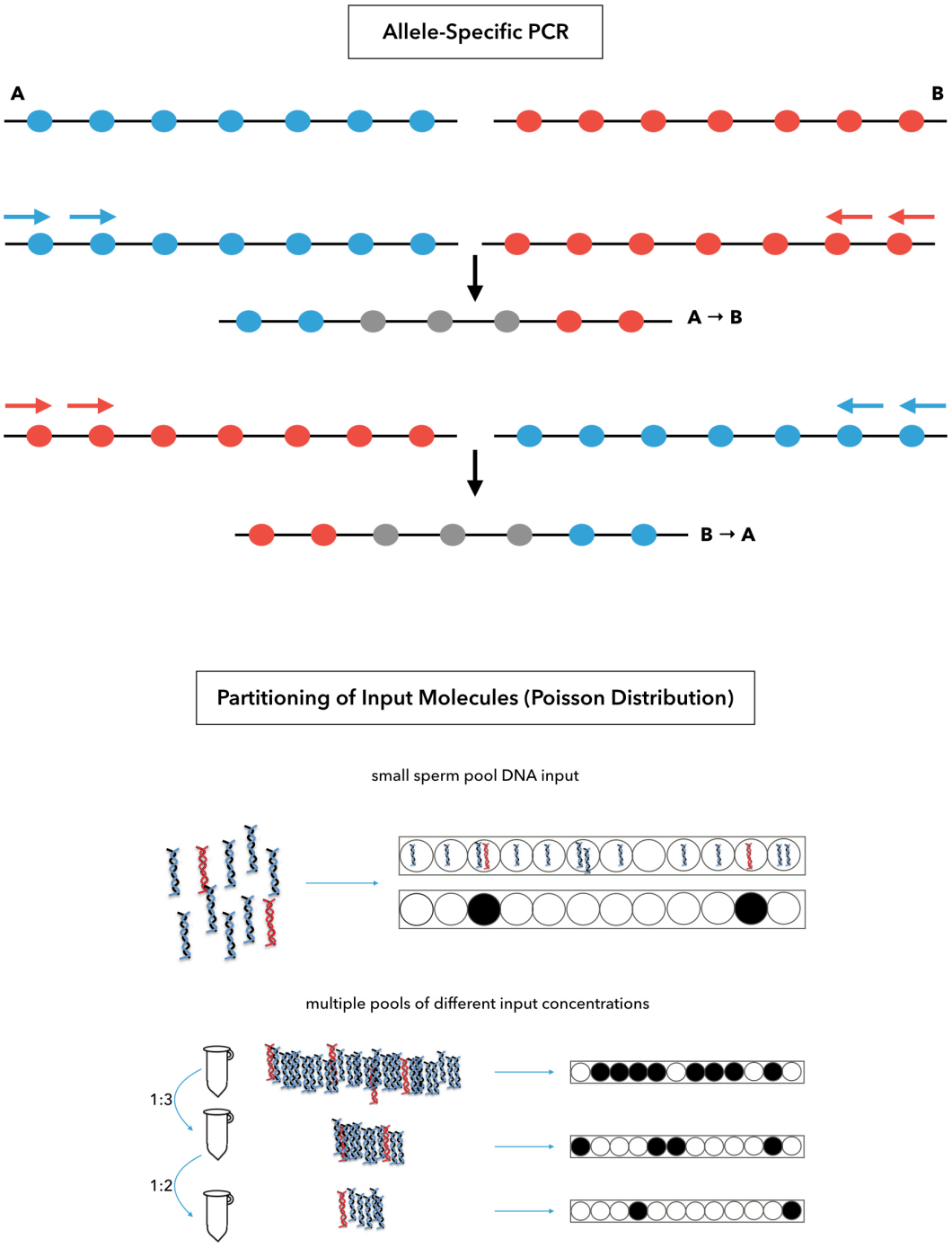


Figure 2 Principle of small-pool sperm typing. In an allele-specific PCR primers are used such that only molecules resulting from a crossover can be amplified. Partitioning (multiple pool sizes and biological replicates) allows statistical inference of underlying crossover frequency.

Determining *de-novo* crossover frequencies

Frequencies were determined using the published software "hotspot" (Odenthal-Hesse, Dutheil et al. 2016), which finds a log-likelihood best fit estimate of the

Poisson-distribution based frequency calculation as described in the paper. The program takes as input the number of positive and negative PCR reactions from crossover assays, as well as the respective pool-sizes (i.e. numbers of amplifiable molecules per reaction).

Amplifiable molecules

To determine the number of input molecules per pool, an aliquot of the HMW DNA to be used in each recombination assay was taken. The aliquot was always large enough to ensure that 3 µl were left over for quantification. The quantification was done by vortexing the left over aliquot thoroughly then using a nanodrop spectrometer to quantify. This ensured that the quantification was as accurate as possible for it is very difficult to quantify HMW DNA, due to its viscosity.

Molecule input was then calculated as follows: One haploid dog sperm genome has a weight of 3 pg. Because only one molecule can be amplified per reciprocal crossover directions per PCR reaction, one amplifiable recombinant has a weight of 6 pg. It has to be taken into account that PCR is not 100% efficient and that some molecules may be broken or otherwise un-amplifiable within the template interval. Extensive experimental data concluded that, on average, only every second correct target molecule can be amplified (Cole and Jasin 2011), giving a weight of 12 pg per amplifiable molecule. It is possible to determine the exact amplification efficiency (Cole and Jasin 2011), however for this project it proved impractical for two reasons. Firstly, the amplification efficiency is determined per DNA batch, which in the case of dog sperm DNA was far less homogeneous than the results expected for mouse DNA. Secondly, the TaKaRa LA Taq can amplify DNA from very low input, yet failed to yield PCR product in single-molecule reactions, which are needed for amplification efficiency tests. Therefore an empirical value was used (0.5) for PCR efficiency and quantified DNA-aliquots used per PCR reactions rather than DNA batches.

Controls

Somatic controls

For one individual (d37) somatic tissue was accessible. Somatic recombination is very rare hence somatic tissue can be used as a negative control to test for the frequency of false positive PCR reactions (parental bleed-through). A crossover assay on somatic DNA was run using very high input molecule numbers (total number of 401544 molecules screened, 200772 per reciprocal orientation) and no positive reactions were found. Assuming that the next tested PCR reaction would yield a positive result can then give an upper limit of the fraction of expected false positive results, which was $1.068 \cdot 10^{-5}$ for d37 and is assumed to be the same for all other individuals.

Negative and positive controls

With every crossover assay two negative control reactions and two positive control reactions were included. For the controls the same master mix was used, but sperm DNA from an individual that is homozygous at all four discriminating SNPs was added. The individuals used for the positive controls carried the homozygous alleles serving as template for each of the ASPs, the negative controls carried the opposite alleles. Negative controls never yielded a positive PCR reaction, whereas all positive control reaction showed a band well visible on the agarose gel.

False positive detection

The newly established buffer system using TaKaRa LA *Taq* yields good results in the sperm-typing assay, albeit slightly increased false positive rates compared to the published *Taq/Pfu*-assay. By sequencing the PCR outcome false positives could be detected. A reaction was classified as false positive if no haplotype switch could be detected between the two markers closest to the forward- and reverse inner allele-specific primer respectively, and those reactions were removed from the resolution point analysis. For the determination of *de-novo* crossover frequencies it was not possible to use the same method, because PCR products with very little final DNA yield did not produce the required coverage for third-generation sequencing.

To ensure that the increased false positive rate did not affect the *de-novo* crossover frequencies significantly, the frequency based on the visible bands on an agarose gel (as in the published sperm-typing method) were measured and then re-analysed by considering only those positive reactions that were supported by a positive sequencing result. This method underestimates the crossover-frequency because a false positive reaction might mask a real crossover recombinant, whereas a negative PCR reaction might still include a crossover recombinant that was not detected by the polymerase. The method does however ensure that no false positive reactions are considered. The corrected dataset results in slightly lower crossover-frequencies, as expected, but the means where not significantly different from each other (Paired t-test on log-likelihood best-fit values, P=0.1, Supplementary information S6).

Interestingly, the somatic controls did not result in a positive PCR reaction, but false positives were only seen in the crossover assays in sperm and their rate was not uniform between individuals. One possible explanation for this could be, that the false positive rate is dependent on the DNA quality or the number of input molecules, which differs by the nature of the experiment. However, the false positives were found in all pool sizes, making this explanation less likely. The false positive rate was highest in some but not all B→A orientation assays. Given the detected wide distribution of crossover resolution points in some individuals, it is also possible, that the molecules scored as "false positive" are actually crossovers that switch very early or very late. A false positive is defined as a molecule that shows no haplotype switch between the first and last marker tested. These are located 506 bp downstream of the inner forward primer and 234 bp upstream of the inner reverse primer respectively. Any haplotype switches between primer and first marker were determined as false positive molecules, because it is impossible to distinguish them from parental bleedthrough, where one of the primers incorrectly binds the opposite haplotype to which it should be specific.

3.9 Sanger sequencing

All Sanger-sequencing was performed in-house and using the same primers as were used for the PCR or else alternative sequencing primers (Supplementary Information S3). Raw files were mapped to the reference genome (CanFam 3.1) in Geneious Software (Version 9) and heterozygotes were called using the Geneious Heterozygotes plug-in and manual verification.

3.10 MinION-sequencing

PCR products were prepared for long-read sequencing with the Oxford Nanopore 1D barcoding genomic DNA kit (with EXP-NBD104, EXP-NBD114 and SQK-LSK109) and adapters specific for every independent PCR reaction were added. Sequencing every independent PCR reaction to high depth (50-3000x) resulted in a set of reads that represents PCR products all amplified from the same single recombinant sperm DNA molecule. The high coverage can thus compensate for the relatively high sequencing error of the MinION system. The minimal coverage was established by comparing sequencing-results with the sequences obtained from Sanger-sequencing (for crossovers as well as phased haplotypes). The main errors induced by MinION-sequencing are short insertion-deletion polymorphisms, which are not as relevant in the present set-up, because a pre-defined set of marker SNPs was used. An advantage of using higher coverage is further, that PCR errors, especially template switching, happening in later PCR cycles and are thus present in low copy numbers, have little impact on the overall sequencing results. From manual comparison of sequences from Sanger- and MinION sequencing a cutoff at 50x coverage was drawn for the MinION-sequencing pipeline. In some cases, sequencing-results with lower coverage were used, in which cases the raw read data in were manually analyzed in Geneious to ensure that the marker sites where unambiguously identifiable as one or the other haplotype.

Basecalling and demultiplexing

For efficiency reasons the guppy basecaller was used (https://community.nanoporetech.com/protocols/Guppy-protocol-

preRev/v/gpb_2003_v1_revk_14dec2018/linux-guppy) and run on parallel cores on the in-house computer cluster. The inbuilt-quality filter was applied and only high-quality reads were passed using the following standard command (Code in Supplementary information S4).

Guppy barcoder was used to de-multiplex and sort reads based on the PCR-reaction specific adapters. The commercially available adapter-set does not require custom configuration so the inbuilt configuration file was used (Code in Supplementary information S4).

3.10.1 Run-to-Run contamination

Two sets of 12 adapters per set are commercially available and usage was alternated with every run on the MinION to reduce run-to-run contamination. Quantification was done by comparing the number of reads found for the adapters used in the previous run to the number of reads for the adapters used in the given run. Run-to-run contamination from the previous runs was ca. 10%, thus by alternating adapter sets contamination rate of as low as 1% is assumed.

3.10.2 *De-novo* assembly

CANU software (https://github.com/marbl/canu/releases/tag/v1.6) was used for *de-novo* assembly of reads per PCR-reaction. The advantage of *de-novo* assembly in comparison to mapping the reads to the reference is, that an unbiased consensus sequence is obtained, including even larger insertions or deletions.

Only reads longer than 4 kb and shorter than 12 kb were considered for the assemblies to further reduce errors due to, for example, short incorrectly amplified PCR fragments, duplications when attaching adapters or late template switching (Code in Supplementary information S4).

3.11 Recombinant sequence analysis

The best assembly (based on length and coverage) was loaded into the Geneious Prime software (Version 2019.1.3) and mapped to the reference together with both parental haplotype sequences. All informative SNPs across the region were

screened manually in each recombinant molecule to find any haplotype switches across the whole amplified region, including complex crossover switches.

3.11.1 Switch-point analysis

To define the recombination frequency for every informative interval of the hotspot region (hotspot morphology), Poisson-corrected crossover numbers were used as input for the hotspot program for every informative interval. These were calculated by multiplying the recombination rate per interval in Morgans with the number of molecules screened per respective individual. These intervals can be different for every parental genotype since the heterozygous sites differ for every dog. By correcting for the number of molecules screened in every individual and taking the length of the informative interval into account, a corrected recombination frequency per base pair can be calculated. The corresponding values for historical recombination rate were extracted from the available linkage maps (Auton, Rui Li et al. 2013).

3.11.2 Non-linear regression and statistical calculation of hotspot center and width

The statistical center and width of a hotspot can be calculated from the cumulative fractions of crossover resolution points (Cole, Keeney et al. 2010). Resolution points in mammals follow normal distribution (Gaussian density function) (Kauppi, Jeffreys et al. 2004) and the cumulative fractions can be regressed using the Gaussian cumulative distribution function. Two standard deviations of this cumulative Gaussian regression cover 95% of the data and represent the hotspot width. The statistical midpoints of these curves correspond to the statistical hotspot center. All regression analyses were done in GraphPad Prism8 software (Version 8.3.1) with default settings unless stated otherwise.

The midpoints of the Gaussian density function are overall consistent with the midpoints from the cumulative distribution function. The midpoints from the cumulative crossover (CO) fractions are reported because this minimizes the effect of different CO numbers obtained per individual, especially minimizing the effect of needle-like peaks in cM/Mb in comparison to the Gaussian density function.

These needle-like peaks occur when a crossover is found in a small interval with closely spaced informative SNPs and from a PCR of large pool-size (high number of input molecules). Especially in individuals with low number of sequenced crossovers these peaks can skew the resolution point distribution and thus cumulative fractions of Poisson-corrected CO numbers give a more accurate depiction of the distribution.

Considering the cumulative fraction distribution, it might be questioned whether a single cumulative Gaussian function best describes the present data. Using standard error as a measure for goodness of fit, the values for each individual are between ca. 0.03 for d26 and d56 and close to 0.1 for d07. The standard error represents the mean distance a point has from the curve in units of the y-axis (in this case, normalized fractions between 0 and 1).

In a data set following normal distribution, the residuals (individual distance of any point to the regression curve) should follow Gaussian distribution as well. In the present data set this seems only to be true for individual d26 but not for any other individual nor for the combined data set (Supplementary Information S7-S12). It is possible that the relatively low number of crossovers per individual (compared to published data sets in mice or humans) is responsible for the rather lower statistical power of the regression.

Given the fact that the distribution for individual d26 meets the criteria for normal distribution (for nPoisson=41.3) whereas it does not for other individuals with similar overall number of crossovers (for example d06, nPoisson=30.2) it is also possible that the underlying shape of the distribution of crossovers varies between individuals.

The distribution of raw data points looks in some cases reminiscent of published data sets from double-hotspots (Webb, Berg et al. 2008). Here, a steep increase in recombination is followed by a plateau and then succeeded by a second steep increase. To account for this, the authors of that study used a set of two separate cumulative Gaussian regression curves to describe both peaks separately. Re-analyzing of the cumulative crossover fractions of the candidates from this study (d06, d07, d30) to see whether they are better described by two successive

Gaussian distributions did not improve best-fit values but the double-hotspot fit was worse than the single-hotspot fit (data not shown).

Overall, the present results show that hotspot usage, hotspot center and possibly the underlying shape of crossover resolution points all vary between individuals and might generally be less well described by Gaussian distribution functions than observed in mice or humans and that data inferred from these regressions should be treated merely as indications.

3.11.3 Transmission distortion (TD)

To analyze transmission frequencies per informative marker only those recombinant molecules that were obtained from individuals heterozygous at the site in question were considered. Of these, the number of molecules switched to the strong/purine allele in the weak/pyrimidine to strong/purine orientation to the number of molecules switched to the weak/pyrimidine allele in the opposite (strong/purine to weak/pyrimidine) orientation were compared, using the following formulas

$$\%\ S\ transmission = \frac{Fraction\ (S\ in\ W\ to\ S) + (1 - Fraction\ (W\ in\ W\ to\ S))}{2} x100$$

$$\%\ W\ transmission = 100\% - \%\ S\ transmission$$

In the case of unbiased transmission, the frequencies should be equal to or close to 50:50. To assess whether any observed transmission ratio is significantly different from the expected 50:50 transmission, two-tailed Fisher's exact test on the raw, uncorrected recombinant molecule counts per SNP were used (Table 1).

3.11.4 Motif search

To identify binding sites for regulating factors at SNPs showing significant transmission distortion, a publically available R package called JASPAR TFBSTools was used (R version 3.5.2 (2018-12-20), JASPAR2018_1.1.1, TFBSTools_1.20.0). JASPAR is a non-profit collection of transcription factor binding affinities (represented as position weight matrices or frequency matrices) that are confirmed by experimental data such as ChIP-Sequencing (as opposed to purely

bioinformatics predictions). The R package uses the site in question as an input and outputs any transcription factor, which might bind to it, with a given cutoff of sequence similarity. For every TD-SNP the motif search was conducted with each alternative allele, including 10 bp up- and downstream of the SNP. The following specifications for the motif search were used:

TF-library: JASPAR 2018
Species-group: vertebrates
Cutoff: specific search: 90% similarity, broader search: 85% similarity
strand: both

The results for each search are attached in the supplementary information (S13). For every candidate factor a tissue-specificity search was conducted, using the human protein atlas to assess whether they are testis-expressed and if so, whether they were found to be enriched in seminiferous ducts, where meiosis takes place.
All functional information was obtained from the uniprot data base (www.uniprot.org) for human, mouse and dog version if available.

4 *De-novo* recombination frequency and hotspot morphology

4.1 Introduction

In most mammals, PRDM9 regulates meiotic recombination hotspots in two ways. First of all, the zinc-finger array of the protein binds specifically to a sequence motif on one of the two present chromosomes, and then the protein places two different trimethylation marks on adjacent histones. These marks are recognized by other proteins and initiate the formation of a double-stranded DNA break. These DNA breaks are limited to a narrow zone of less than 200 bp (Baker, Walker et al. 2014). By placing trimethylation marks on only the two histones closest to its binding site, PRDM9 also limits the migration of crossover-intermediates to a close range of ca. 1 kb up- and downstream of the initiation site, thereby creating a limited range of ca. 2 kb in which recombination takes place (Baker, Walker et al. 2014).

A study in mice showed that without functional PRDM9, DSBs are instead placed in regions with existing H3K4me3 marks, such as promoters and transcription start sites (Brick, Smagulova et al. 2012). While break formation is not impaired in those cases, the repair of DSBs seems to be delayed or overall impossible, leading to a reduced number of crossovers. Apart from the evolutionary disadvantages of reproduction without recombination, reducing the number of crossovers also poses a more imminent threat on the outcome of meiosis. To create germ cells of uniform genomic content, chromosomes have to be correctly lined up in prophase I, and then segregate to opposite poles of the cell. At least one crossover per chromosome, or per chromosome arm in larger chromosomes, is needed to ensure this (Murray and Szostak 1985). The reduction of crossover numbers below this critical point can lead to asynapsis of chromosomes and cause chromosomal mis-segregation leading to aneuploidies.

Interestingly, dogs and other canids lack functional PRDM9 (Oliver, Goodstadt et al. 2009, Munoz-Fuentes, Di Rienzo et al. 2011).

4.1.1 Canids carry a pseudogenized version of PRDM9

The functional domains of PRMD9 are highly conserved across diverse metazoan taxa; however, the *canis familiaris* Prdm9 ortholog has acquired multiple disruptive mutations including several stop-codons (Oliver, Goodstadt et al. 2009, Munoz-Fuentes, Di Rienzo et al. 2011). Re-sequencing of Prdm9 orthologs in species across the canid species tree, including several species from the wolf-like clade to which the domestic dog belongs, the red fox as a representative of the red-fox-like clade and the bush dog as a representative of the South American clade, suggest that Prdm9 was pseudogenized across the whole canid group (Lindblad-Toh, Wade et al. 2005, Munoz-Fuentes, Di Rienzo et al. 2011, Axelsson, Webster et al. 2012). Not all the mutations are shared between species of the canid group (Munoz-Fuentes, Di Rienzo et al. 2011). This suggests that there has not been a single early disruption event, but rather the function of PRDM9 became redundant, allowing for multiple mutations to accumulate. The closest relative proven to still carry functional PRDM9 is the Panda and the Prdm9 sequence of the grey fox and the island fox, the most distant species of the canid tree to the domestic dog, have not yet been analyzed. Consequently, Prdm9 was lost in the canid group between 10 myr and ~49 myr ago (Wayne, Geffen et al. 1997, Lindblad-Toh, Wade et al. 2005, Eizirik, Murphy et al. 2010, Axelsson, Webster et al. 2012) and thus well before the domestication of the dog.

4.1.2 Historical Recombination in dogs

There are two historical recombination maps available for dogs to date, one an LD map calculated using SNP-array data sets from nearly 500 samples across 30 breeds (Axelsson, Webster et al. 2012). The second study used whole genome sequences from another 50 samples of village dogs. This second study has a more fine-scale resolution due to using whole genome sequences as well as using less inbred individuals with higher marker density (Auton, Rui Li et al. 2013). The overlap between both studies is significant, confirming that both studies detect

recombination but at different resolutions (Berglund, Quilez et al. 2014). Both data sets confirm that dog recombination clusters in hotspots, although elevation from the background seems to be less intense than in other mammals, and overall, hotspots seem to be more evenly spaced across the genome. These effects could also be bioinformatics artifacts because of the lower marker SNP density and higher effective population size in dogs (Auton, Rui Li et al. 2013).

The recombination map reveals a very strong association of dog hotspots with CpG-rich motifs with a relative enrichment of 61% for the strongest motif found (CGCCGCG). In line with this observation, the authors find highly elevated rates of recombination around transcription start sites (TSS), which seems to be due to the nature of CpG-islands serving as promoter regions. There is a close correlation of dog hotspots with TSS carrying a CpG-island that lessens in TSS without CpG-islands, suggesting an important role of the presence of a CpG-islands for dog recombination hotspots (Auton, Rui Li et al. 2013). A further study revealed that intergenic CpG islands are depleted for hotspots, but genic and promoter-associated CpGs are enriched in hotspots, suggesting a specific enrichment of dog hotspots with promoter-associated CpG-islands (Berglund, Quilez et al. 2014). A study in knockout-mice lacking functional PRDM9 likewise observed clustering of recombination hotspots in functional elements, showing that in mammalian PRDM9-deficient recombination, initiating DSBs are placed at existing trimethylation marks (H3K4me3) which can be found in promoter regions. While the PRDM9-knockout mice studied showed a range of meiotic defective phenotypes, some as severe as complete sterility, dogs seem to be able to complete recombination in functional elements without severe meiotic defects (Hayashi, Yoshida et al. 2005, Brick, Smagulova et al. 2012, Mihola, Pratto et al. 2019).

4.1.3 Meiotic double strand break initiation in dogs

Data from DMC1-ChIP-sequencing (Galina Petukhova, unpublished) confirm strong enrichment of dog initiation hotspots in functional elements and a generally broader initiation zone of hotspots. These initiation maps also suggest a more

even distribution of recombination across the genome compared to other mammals. First of all, the area around strong peaks of initiation also show slightly elevated recombination compared to the background and the conductors of the study also observe large regions of elevated background activity. These "high-background" regions still possess hotspots in the classical sense, but larger adjacent regions show a more uniform elevation of recombination compared to the genome average (Kevin Brick and Galina Pethukova, *personal communication*).

4.2 Aim of the experiment

The first part of this thesis aimed at testing sites of high levels of historical recombination ("LD-hotspots") in dogs, for de-novo recombination events in dog sperm DNA. This approach is used to determine if they are still active to date, but also of the temporal stability of this hotspot. A historical hotspot, that is still active to date reveals some stability across long evolutionary time scales. The second aim of this chapter is to determine the crossover resolution distribution spatially across the hotspot and flanking interval.

4.3 Results

4.3.1 *De-novo* recombination frequency of a historical hotspot in dogs

Screening a total of 1.96 million sperm across seven parental haplotypes revealed that the analyzed historical hotspot is active *de-novo* in sperm at an overall crossover frequency of 0.148 cM (Figure 3). This frequency is in accordance with the expected rate from historical linkage data (expected rate 0.091 cM, from Haldane's map function on linkage data from Adam Auton). To test the robustness of the data set, it can be sub-divided in several ways. First of all, the crossover frequencies between reciprocal crossover orientations (Haplotype A to B and B to A) should not be significantly different from each other as shown in Figure 4, right panel. The tendency visible in Figure 4A, that the frequencies of the B to A orientation are slightly higher in most cases, might be explained by the design of the assay. The two markers discriminating haplotypes downstream of the hotspot are both A/G-SNPs. Due to the haplotype phase present in the tested individuals,

the reverse primer for the B to A orientation assay had to end on a thymine. Thymine has a higher affinity to bind non-complementary bases than the other nucleobases, making these primers slightly less specific.

In contrast to the overall *de-novo* recombination frequency matching well with historical recombination, frequencies between individuals showed variation. This was unexpected with regard to the current hypothesis that PRDM9-deficient species show stable recombination hotspots, as observed for example in birds (Singhal, Leffler et al. 2015) (Figure 4). Individual d26 and d37 had a 10-fold increased crossover frequency compared to the expectation from LD and individual d07 had a 2-fold reduced frequency compared to the expectation. The difference between individuals is in the order of one magnitude, which cannot be explained by the expected variation in the assay (95%-confidence intervals) and is much greater than the variation between reciprocal crossover orientations.

While direct quantification of the amplification efficiency per DNA batch was difficult in dogs (see 3.8.4) the DNA quality is unlikely to cause the observed frequency differences. Firstly, Quantification of input molecules was performed per aliquot instead of per batch to account for heterogeneity of the extracted DNA; secondly, phasing PCRs (3.8.3) yielded reproducible positive results for every sample, showing that the region is generally amplifiable with the used primers in every individual.

4.3.2 Crossover resolution points

104 recombinant molecules were sequenced across the whole 10.8 kb amplified region. Unambiguous recombinant sequences were generated by using long-read sequencing at high coverage of the amplified PCR product. Because PCR products were assumed to be generated from single-molecules, the high coverage of at least 50x per molecule ensured accurate sequences even considering the relatively high error rate of MinION-sequencing. The *de-novo* assembled sequences per recombinant were mapped to the reference genome (CanFam 3.1) together with the parental haplotypes, getting detailed information on haplotype switches at informative sites within the hotspot region. The marker density varied,

the shortest intervals being several base pairs long, the longer ones several hundred base pairs and informative SNPs and insertions or deletions (InDels) sometimes differed between individuals.

To account for this, Poisson corrected crossover numbers were used and the recombination frequency in cM/Mb was calculated per single base pair (Figure 3, upper panel).

The bulk of resolution points across all individuals was consistent with the historical distribution from the LD map, with the midpoint at chr27:16,038,200 and a maximum rate of 44.6 cM/Mb. The distribution from historical recombination data suggested a hotspot width of ca. 2500-3000 bp (statistical width from Gaussian density curve on recombination data from Auton(Auton, Rui Li et al. 2013), curve not shown). The statistical width of the *de-novo* recombination hotspots was inferred to be 7216 bp (Figure 3, lower panel) and was thus even wider than expected from historical recombination and much wider than PRDM9-dependent hotspots described in mouse or humans (for example (Webb, Berg et al. 2008, Cole, Keeney et al. 2010)). The Gaussian cumulative distribution regression does not provide a very accurate fit to the data. It is likely that the distribution of resolution points detected here is represented by either a different, more heavy-tailed distribution or that the underlying data represent a cluster of normally distributed resolution points which overlap and can not be represented by a single non-linear regression.

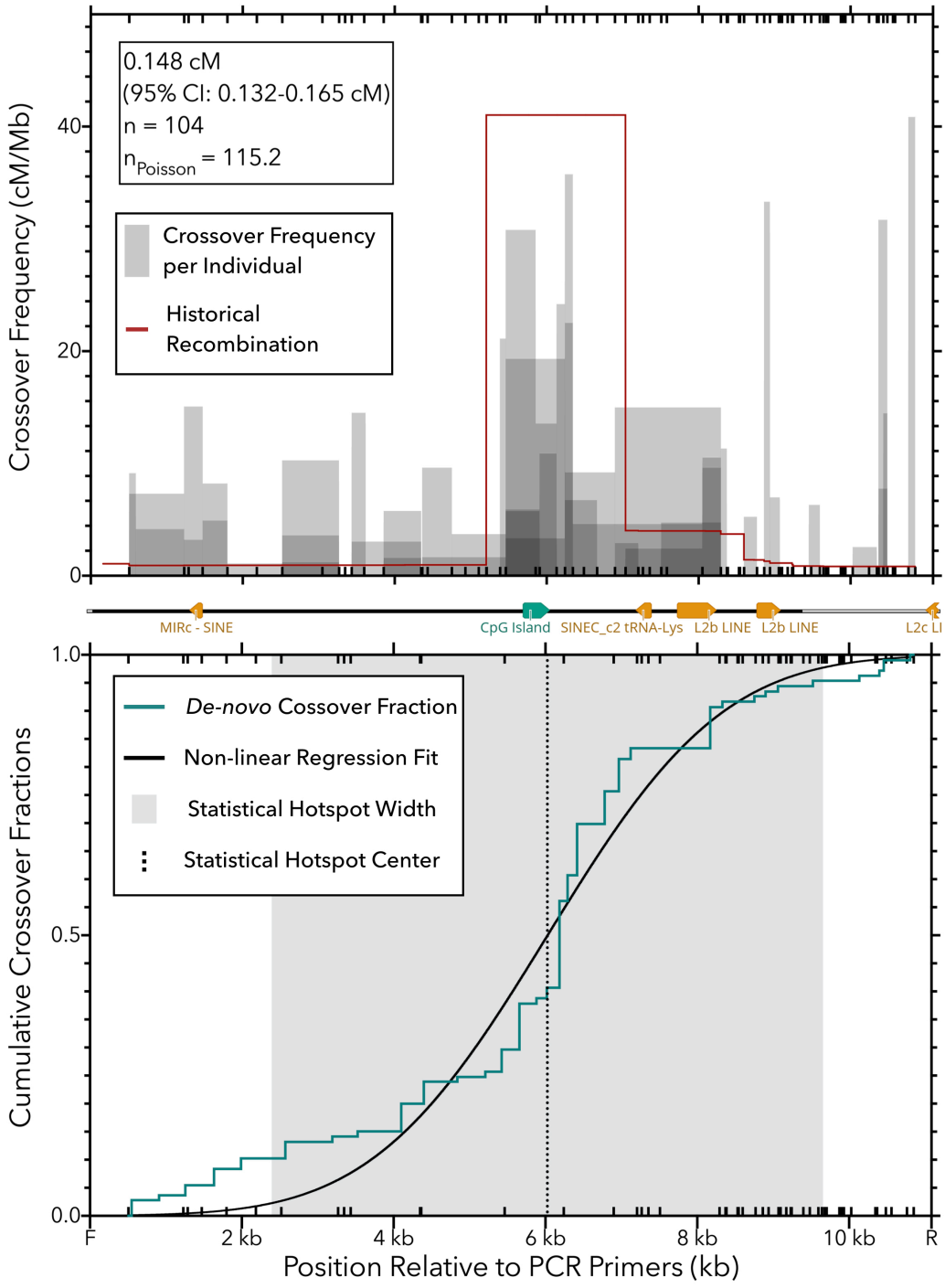


Figure 3 Crossover resolution points. "F" and "R" refer to the inner allele-specific primers and inward ticks demark all markers informative for at least one individual. Annotations are shown between panels. Upper panel: resolution per individual, overlapping distributions in darker shades of grey and red line indicating the LD-peak. Lower panel: Cumulative crossover fractions per interval. Green line represents haplotype switches in the data set. Smooth black line represents the non-linear regression fit of the cumulative Gaussian function. Dotted line indicates statistical midpoint from non-linear regression and grey shaded area indicates 95%-width of the non-linear regression (hotspot width).

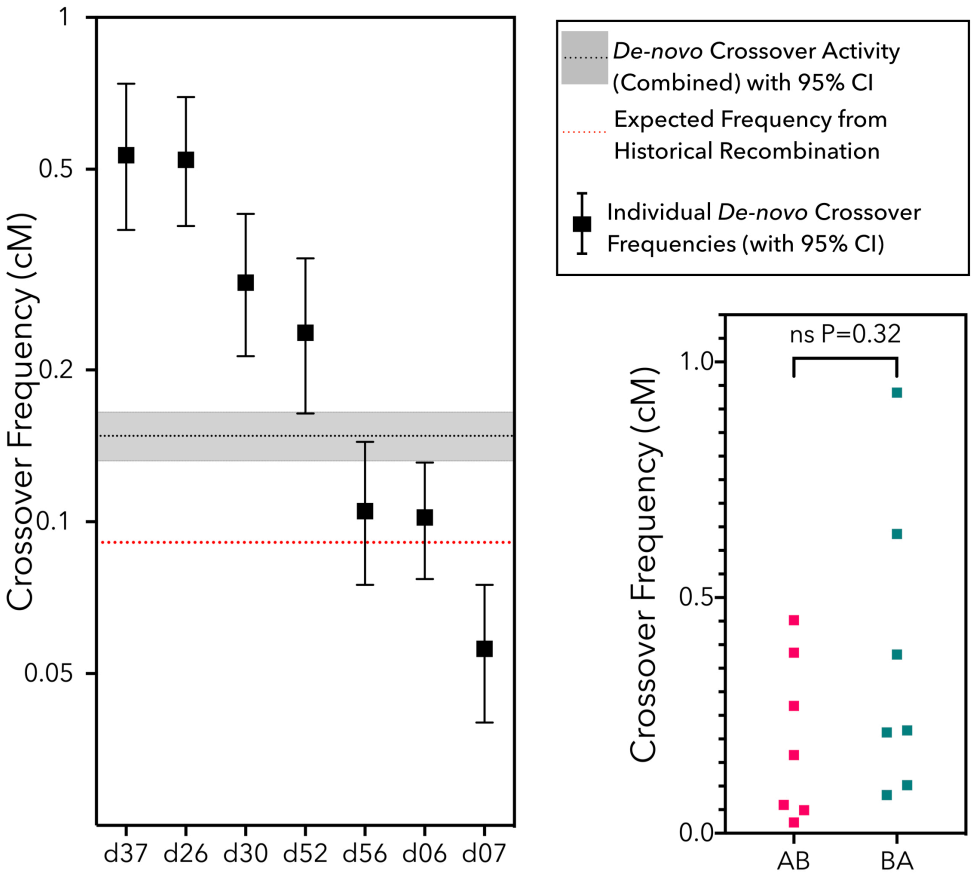


Figure 4 Individual crossover frequencies. Left panel: overall frequency per individual; error bars represent 95%-confidence intervals. Red line indicates expected rate from historical linkage data (0.091 cM). Right panel: Frequencies grouped by reciprocal orientations; Mann-Whitney U test on mean frequencies in cM, P=0.32.

4.4 Analysis of frequency and morphology per individual

The differences in *de-novo* crossover frequencies between individuals suggest that the dog recombination hotspot may be stable across time, but not across individuals as originally expected. Consequently, individual crossover resolution sites were analyzed to further test for morphological differences between individuals.

Figure 5 shows individual crossover resolution points. For this analysis, individuals were excluded if less than five crossovers could be sequenced (d52, d37), leaving five individuals to compare.

The variation in *de-novo* crossover frequency described above was also reflected in the crossover resolution points. The hotspot widths and midpoints varied

between individuals. Midpoint location differded by up to 2000 bp between individuals and the inferred widths range from 6.5 kb to 12.1 kb. Again it should be noted that Gaussian regression does not represent a very good fit of the data and the information inferred should be treated as indicative but possibly not conclusive.

Figures 5-9. The overall number of crossovers sequenced per individual is shown in the top right part of each upper panel. All x-axes demark the spatial position across the tested region on Chromosome 27, with inward ticks representing informative SNPs for the respective individuals. Top panel: Resolution points in bar plots, where bar widths represents the maximum interval in which a haplotype switch occurred. The right edge of each bar represents the tested SNP and corresponds to the SNPs tested for the cumulative crossover fractions in the bottom panel. Bar heights represent recombination activity in cM/Mb. Bottom panels shows the cumulative breakpoint fractions. This was calculated using Poisson-corrected Crossover-numbers per Interval, standardizing to the overall Poisson corrected crossover number for that individual. Hotspot width is represented in gray shaded areas and statistical midpoints are shown as dotted lines (inferred from cumulative non-linear regression).

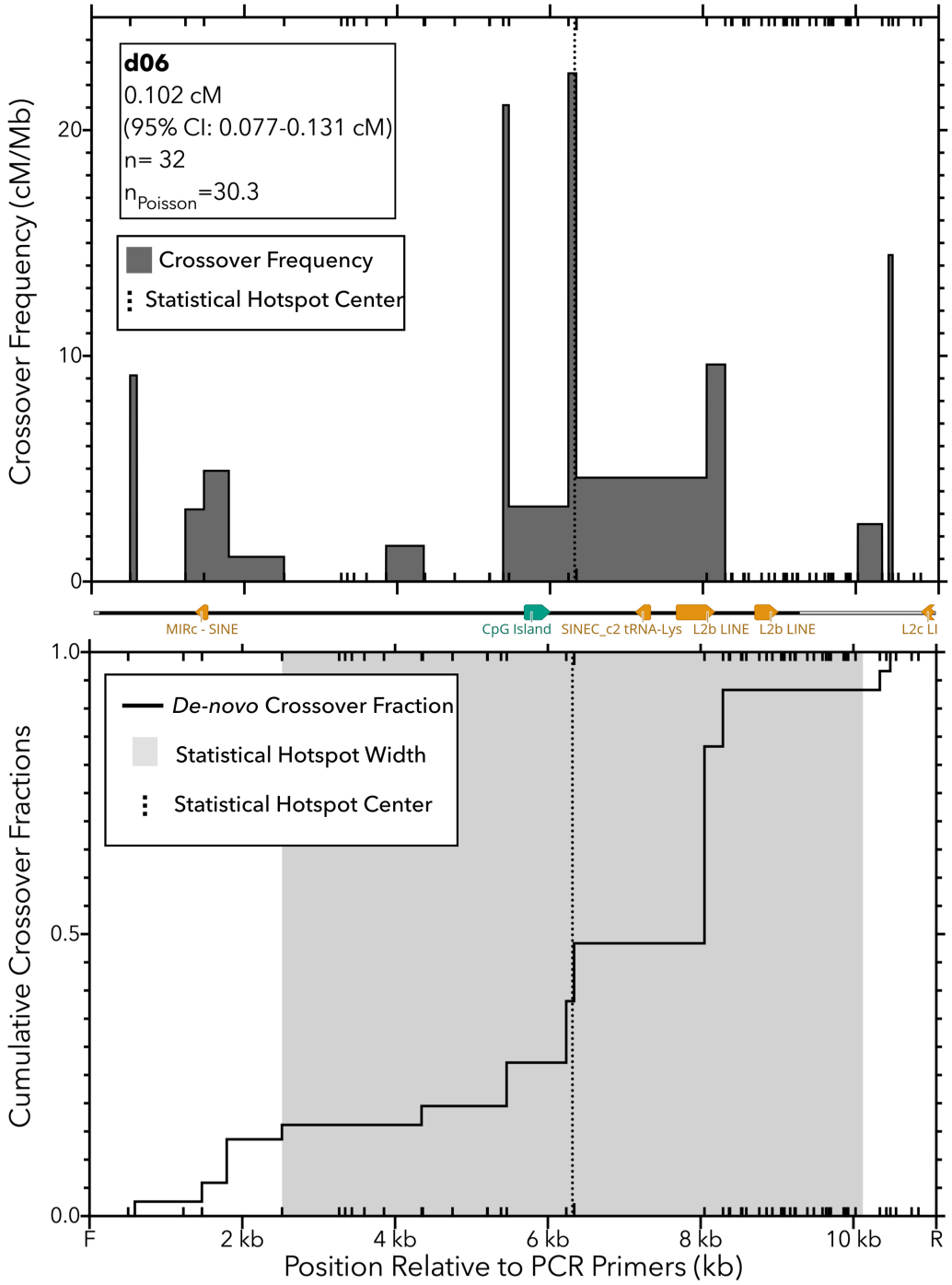


Figure 5 Individual crossover resolution d06.

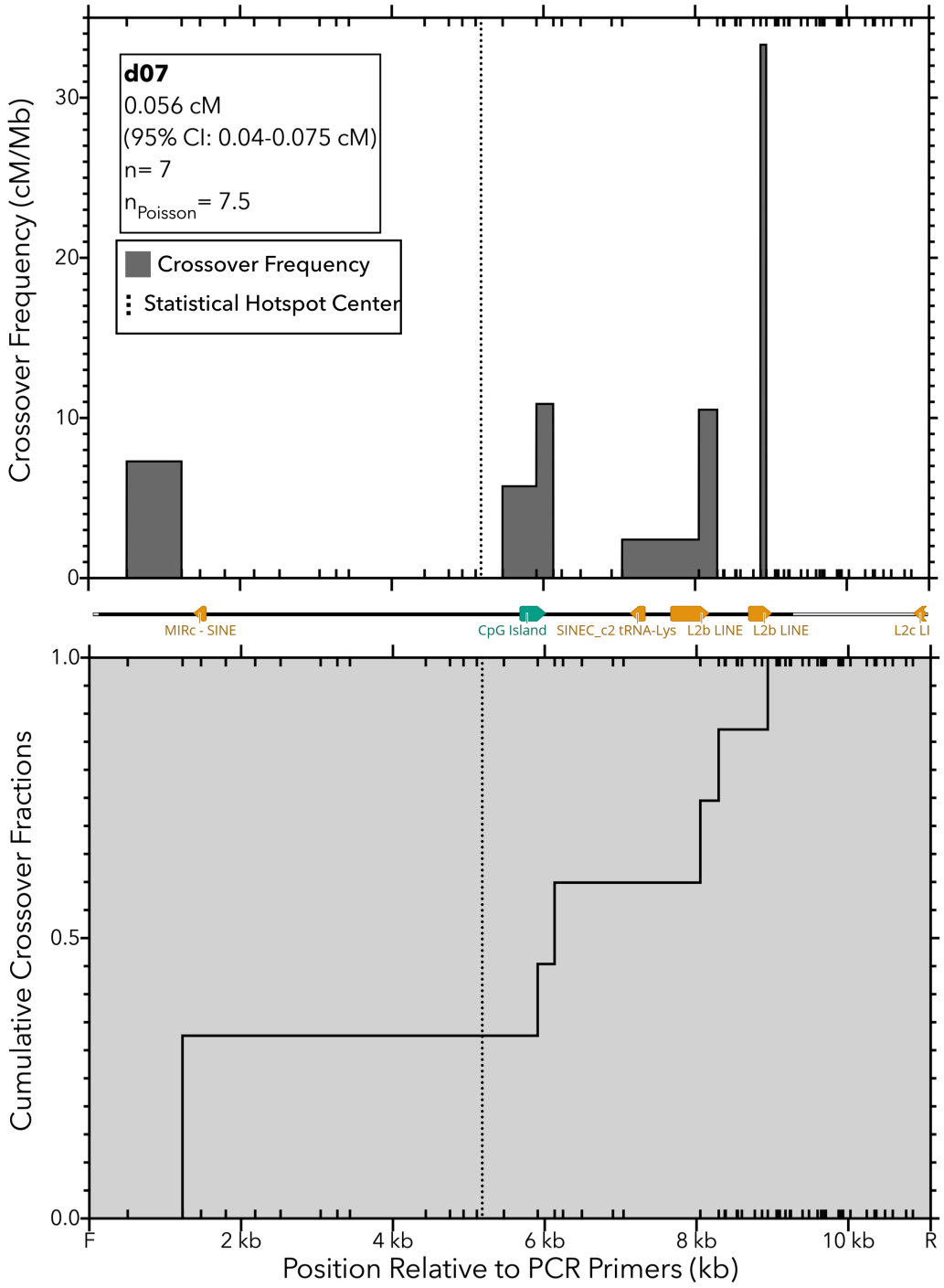


Figure 6 Individual crossover resolution d07.

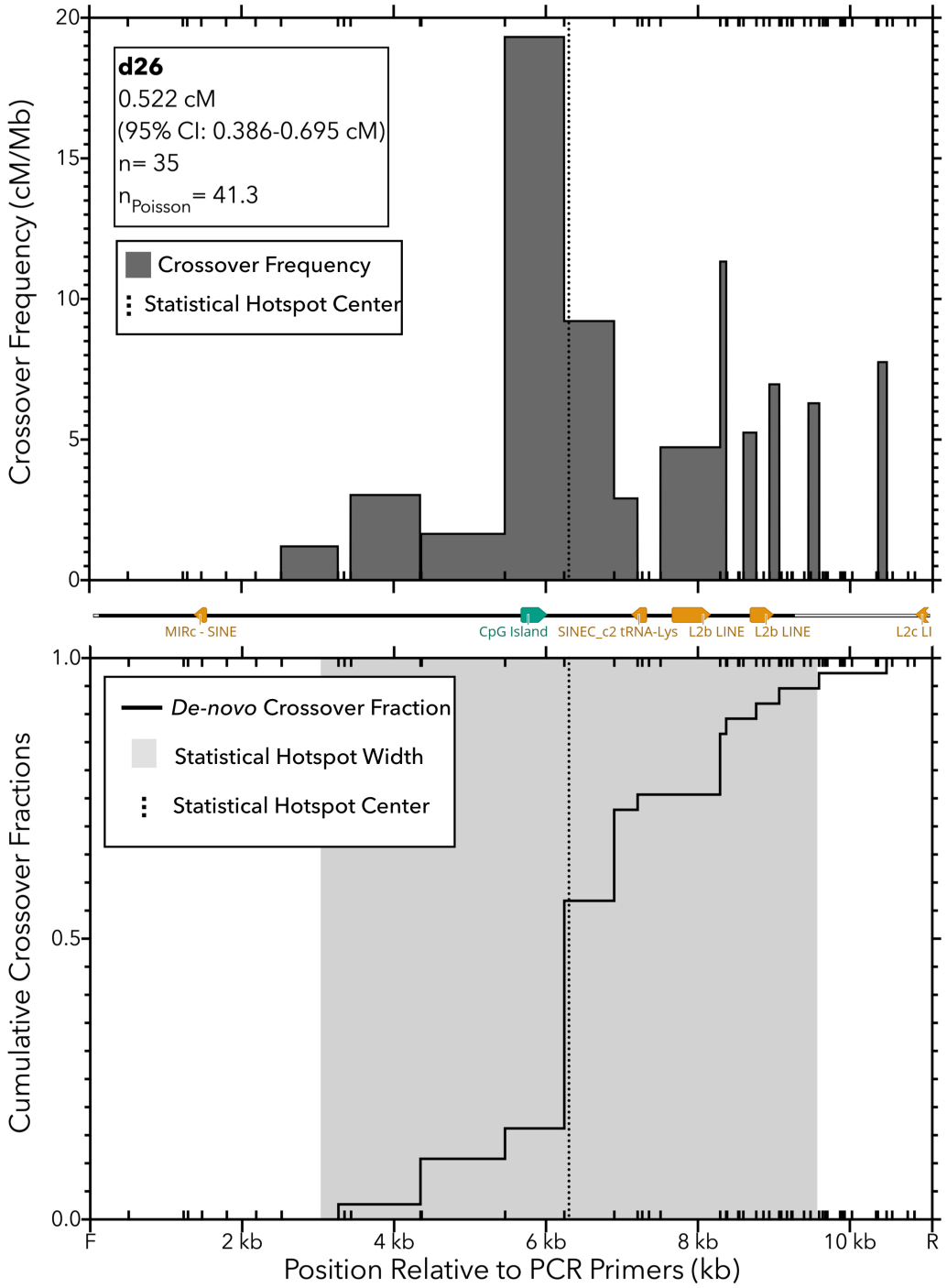


Figure 7 Individual crossover resolution d26.

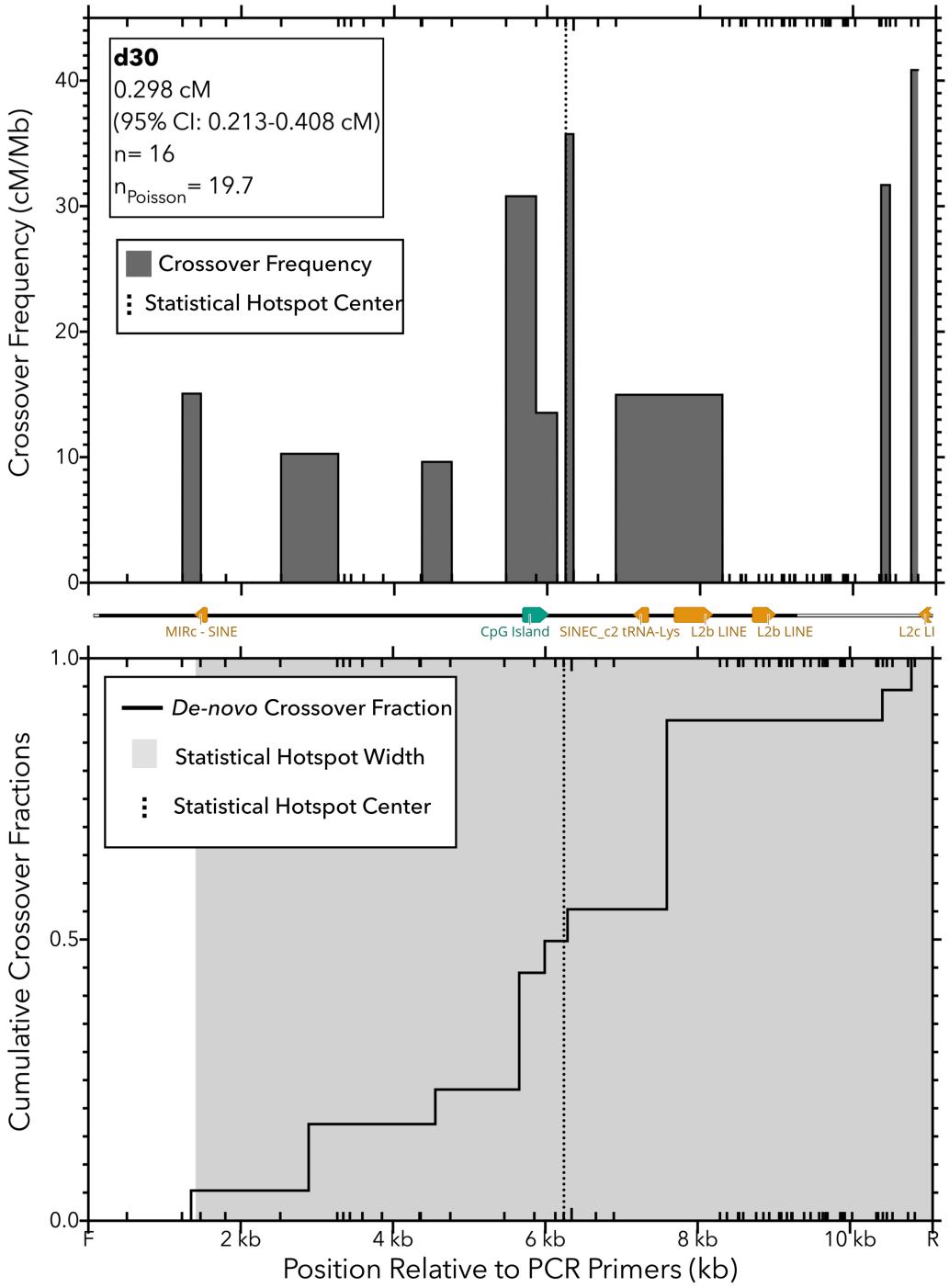


Figure 8 Individual crossover resolution d30.

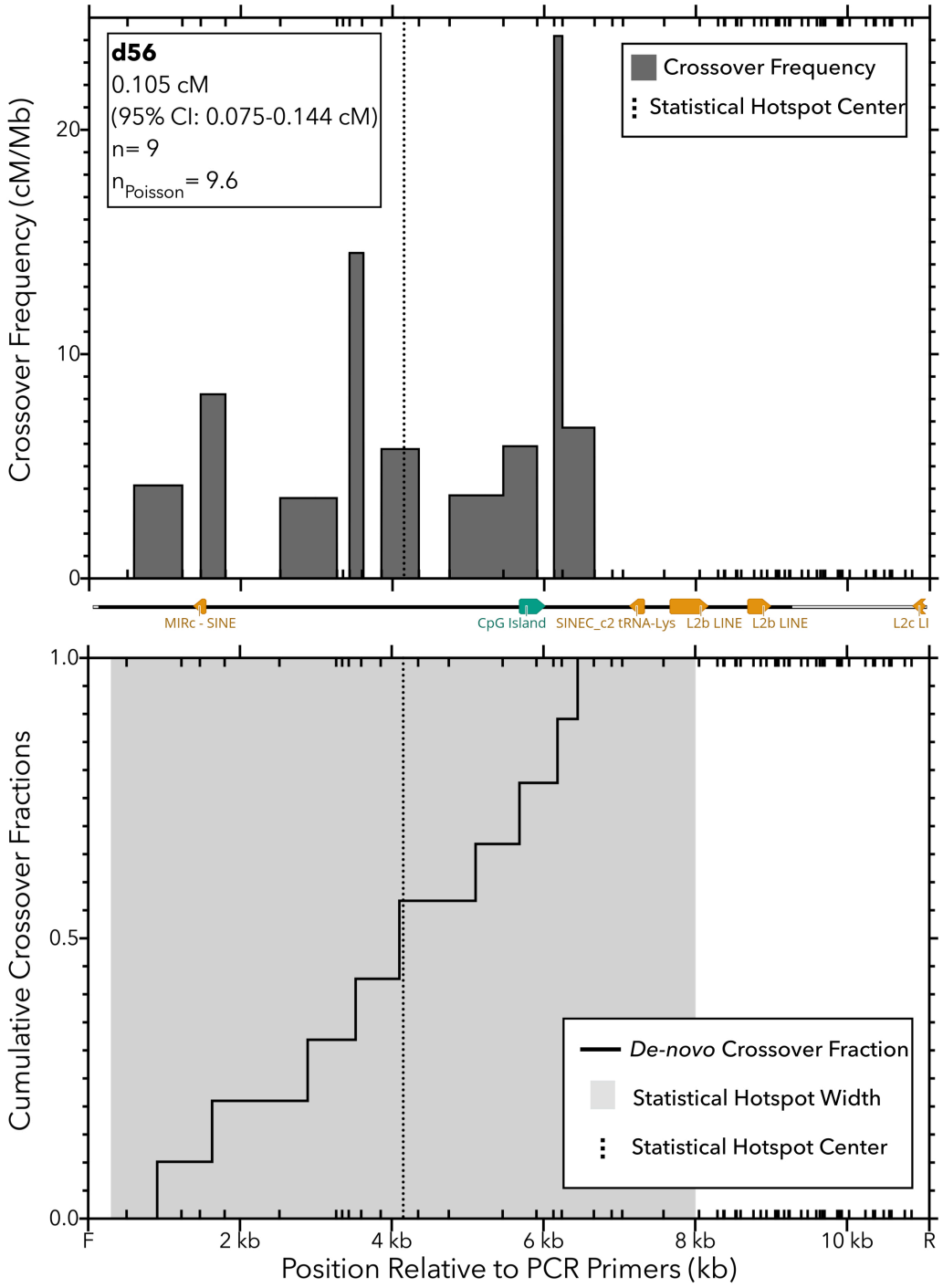


Figure 9 Individual crossover resolution d56.

4.4.1 Genomic context

PRDM9 directs recombination hotspots away from functional elements possibly conferring a protective role against potential mutagenic effects of recombination in functional elements (Brick, Smagulova et al. 2012).

It was shown before that PRDM9-deficient recombination uses existing trimethylation marks (H3K4me3) in gene promoter regions. While PRDM9-deficient recombination in yeast is located at the upstream nucleosome depleted regions of promoters (Wu and Lichten 1994, Pan, Sasaki et al. 2011), mice lacking PRDM9 seem to place DSBs just downstream of the TSS (Brick, Smagulova et al. 2012). Auton *et al.* found that hotpots tended to cluster around TSS as well, however it appears that this effect is correlated with CpG-islands (Berglund, Quilez et al. 2014). The DMC1-ChIP-sequencing data from dogs are, unfortunately, not published yet, so analysis of fine-scale placement of DSBs is not yet possible.

Exactly how meiotic recombination is facilitated in "default" sites remains elusive but the strong correlation with functional elements suggests the genetic architecture to be essential. The available annotations for the hotspot region from USCS genome browser (Figure 10, source: http://genome.ucsc.edu/cgi-bin/hgTracks?db=canFam3&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr27%3A16032181%2D16043266&hgsid=798167103_KZhEta9vxHC0DLdXOetE0TpZlTmu) reveal a CpG-island placed centrally in the hotspot, as well as several transposable elements (MIRc, SINEC_c2_tRNA-Lys, 2x L2b LINE, L2c LINE). A predicted gene spanning the region is the PKP2 gene (from mouse and human) exon2 and exon3. The only available expression data for dogs is a spliced EST consistent with PKP2 exons 2 and 3, with the gene ending with exon 3 just before the CpG-island, while mouse and human genes continue several 10s of kb downstream. The re-sequencing data of the tested individuals confirms the location and placement of all these elements in every dog tested.

To test if the observed amount of repetitive elements is representative of the dog genome it was split into bins of 10 kb length, which is roughly the length of the hotspot region in the present assay, and the occurrences of repetitive elements annotated in the UCSC genome browser were counted. The histogram (Figure 11

and Supplementary Information S39) shows that ca. 60 % of 10 kb bins contained one or no repetitive elements, but only 3.1 % of the bins have five or more repetitive elements within 10kb.

The statistical midpoint of the hotspot is co-located with the CpG-island. It should be noted that the boundaries of the CpG-island are defined computationally. The surrounding regions include further CpGs and are distinctly GC-rich across ca. 300 bp around the CpG-island.

The exons shown in Figure 10 correspond to the annotated dog ESTs. Annotations from other mammals, including human and mouse, are very similar at Exon 2 (upstream exon) but Exon3 continues downstream, spanning the CpG-island region. No crossover resolution was detected in the annotated dog exons, however haplotype switches did occur just downstream of the dog exon, within the annotated mammalian exon3 (as seen in Figure 3).

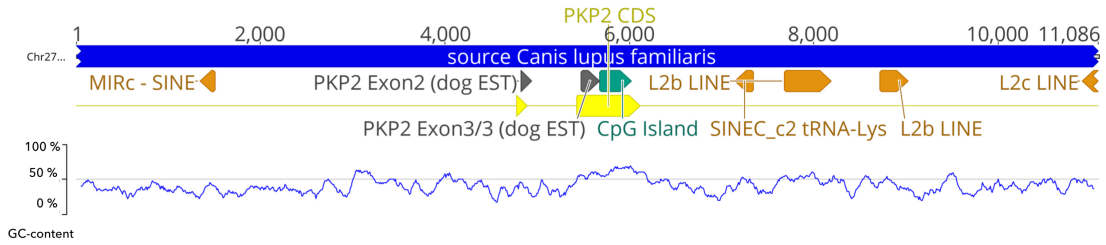


Figure 10 Available annotations for the Hotspot region from USCS genome browser. Orange: Repetitive elements, green: CpG-island, grey: dog ESTs, yellow: mammalian exons, blue: source. Positions refer to base counts starting from the outer allele-specific primer.

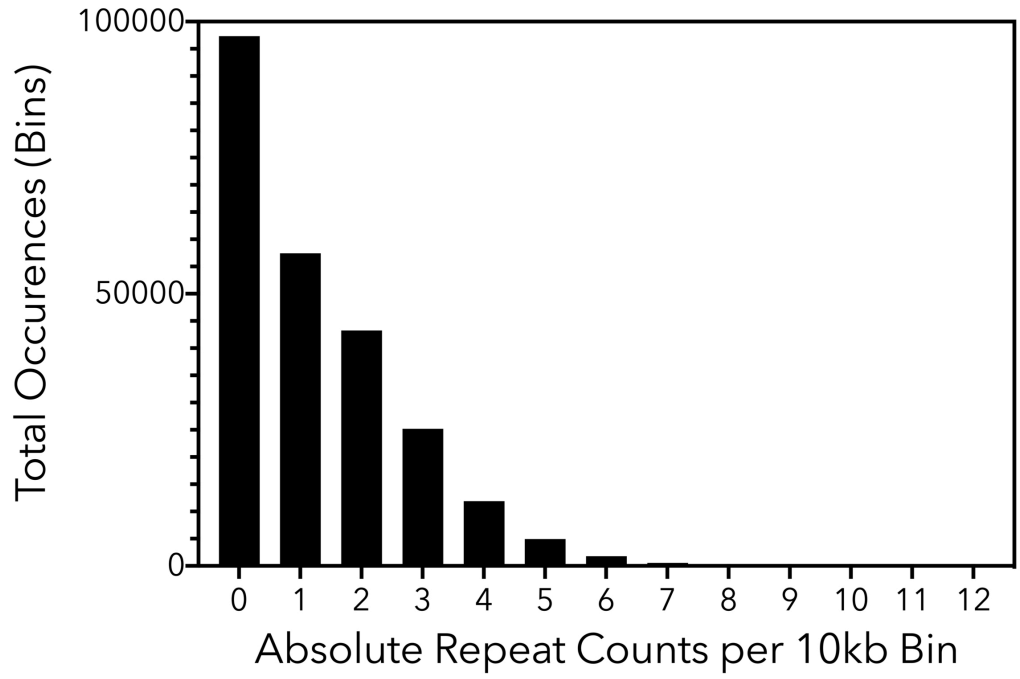


Figure 11 Count histogram for repetitive elements within the dog genome in 10 kb windows. Dog genome (CanFam 3.1) was split into 10 kb windows and number of annotated elements found per window reported.

4.5 Discussion

The first question to be answered about dog recombination hotspots was whether the peaks observed in LD maps represented sites of still active hotspots or if these were merely marks of past recombination. The present data show that the historical hotspot is still active *de-novo* and at an overall frequency consistent with the expectation from the LD data. These finding are completely novel, as there has to date been no assessment on the outcome of recombination in any canine species on a fine-scale level. That historical hotspots should be active *de-novo* is not surprising, given the current model of PRDM9-deficient recombination. In a stable recombination landscape it would be expected that the same set of hotspots is used across generations, between individuals of the same species and even between different species along the canine tree.

Contradictory to this theory, variation in the hotpot heat even between dogs (i.e within species) was detected. The differences in recombination frequency, which are in the order of one magnitude, suggest that there is some regulation of

hotspot heat, which cannot be attributed to the general chromatin state at the hotspot region but must lie in differences between the individuals.

One possible model could be, that dog recombination is always linked to active transcription and that the differences observed are due to different expression patterns between individuals. These might not necessarily be linked to meiotic recombination. The gene found in the analyzed hotspot encodes PKP2 (plakophilin 2). This gene has been linked with atopic dermatitis in dogs and with a specific cardiomyopathy in humans (RefSeq, accession number: NM_001005242.3). In humans, it is mainly expressed in heart muscle tissue but expression in testis was detected as well, albeit at much lower levels.

If hotspot regulation indeed depends on transcription activity, the question remains how it can be advantageous to disrupt an active promoter rather than directing recombination away from transcription. Recombination sites demark those positions where existing allele combinations can be broken up and new combinations are created. New allele combinations might be beneficial if a deleterious allele was purged, but to break up a functional gene and disrupt its expression would be predicted to hold more disadvantages than beneficial effects. Recombination activity has also been linked with increased *de-novo* mutation rate (Arbeithuber, Betancourt et al. 2015), posing an additional threat to promoters in which recombination takes place, since most new mutations are deleterious.

It is possible that canids use an alternative trans-acting regulating factor to PRDM9. There are other methyltransferases with similar properties to PRDM9 (especially other proteins from the PRDM-family), however none of them can place the characteristic double-mark (H3K4me3 and H3K36me3) that is unique to PRDM9-mediated recombination sites (Powers, Parvanov et al. 2016). It is possible that there is a regulating factor, determining hotspot heat and possibly crossover resolution in dogs, nonetheless keeping recombination sites around CpG-islands and not directing them away from functional elements. A very recent study found meiotic recombination to be clustered in functional elements in snakes, despite possessing functional PRDM9 (Schield, Pasquesi et al. 2020).

Another result is the extensive width of the hotspot, with 95%-width of ca. 7 kb and crossover resolution observed across the whole region analyzed. The detection of slightly broader distributions of breakpoints than in mice or humans was expected because of the distinctive role of PRDM9 in determining hotspot width.

Unpublished DMC1-ChIP-sequencing data from dogs (Galina Petukhova, *personal communication*) show wider zones of DSB initiation compared to mouse and human data, probably due to larger nucleosome-depleted areas. The present data set detects DSB initiation widths in hotspot of several thousands of base pairs, which might be an explanation for the non-Gaussian distribution of resolution points (further discussed in 5.8).

Interestingly, additionally to the "classical" hotspot structure, where a steep increase in recombination activity is surrounded by an area of low recombination levels, they also detect hotspots which are located in areas of generally increases recombination activity ("high background hotspots").

The extensive hotspot width as well as the rather poor fit of a Gaussian distribution to the crossover resolution points might suggest that recombination in dogs is organized altogether differently than in other mammals. The detection limits for crossover resolution points in this study are defined by the PCR-primers used for the sperm-typing assay. It is likely that if an even broader interval around the historical hotspot center would have been analyzed, additional resolution points even outside this already broad interval could have been detected. Together with the unpublished data from Galina Petukhovas group, the present data suggest that dog recombination is not so much clustered in recombination hotspots but rather in recombination "zones" or "areas" of the genome, albeit with individual differences in the heat or usage. Whether any given hotspot is characterized by a steep increase in recombination in a low recombination background, or located within a "high background" area might be defined by the chromatin structure at that site.

The limiting factors determining hotspot width in dogs remain elusive. Possibly, repetitive elements, which are highly enriched in dog hotspots, could play a role in

limiting accessible chromatin, as they are suggested to perform this role in transcription (reviewed in (Elbarbary, Lucas et al. 2016)).

The promoters associated with the CpG-islands clustered in dog recombination hotspots might not be active in the same meiotic stages that recombination occurs at. There is no data available to link active promoters in early prophase I to active recombination hotspots, especially since LD hotspots might be a general predictor for sites of recombination but not for individual hotspot usage.

There have been indications that dog recombination might follow a nested regulation model, where the genome is clustered into areas generally accessible to the DSB machinery. These regions could be defined by overall chromatin modification, chromatin states or the location on the DNA-loops relative to the chromosomal axis versus the axis of the chromosome. On the fine-scale level, other factors could modify individual placement and frequency of recombination (Berglund, Quilez et al. 2014).

5 Aspects of fine-scale hotspot morphology

5.1 Introduction

The results described in Chapter 4 show that PRDM9-deficient canine recombination does not impair general usage of a hotspot, however the differences in hotspot heat as well as resolution point distributions between individuals suggest that there is variation in fine-scale regulation of the hotspot. The analyses in this chapter will focus on resolution points from reciprocal crossover orientations. The double strand break repair model (DSBR model) states that recombination occurs when one homologue (initiation homologue) experiences a DSB, which is extended into a gap, this gap is then repaired using the other homologue (template homologue)(Szostak, Orr-Weaver et al. 1983, Sun, Treco et al. 1991). Depending on the intermediates formed and the way these intermediates are resolved this might lead to crossover- or non-crossover recombination, which can be associated with gene conversion, where the genetic material of the template homologue replaces genetic material from the initiation homologue in a non-mendelian fashion. If gene conversion happens equally often in reciprocal recombination directions, the overall transmitted allele frequencies into the next generation does not change. However when biases occur, for example during DSB initiation or mismatch repair, the transmission frequencies can differ from 50:50, causing transmission distortion (TD). TD can play a major role on the evolution of genomes by influencing allele frequencies from one generation to the next.

5.1.1 How does TD arise?

Biased DSB initiation

In PRDM9-carrying mammals, DSB initiation biases are attributed to PRDM9-binding (Baudat, Buard et al. 2010, Berg, Neumann et al. 2010, Myers, Bowden et al. 2010, Berg, Neumann et al. 2011). In a given hotspot, when there is variation

within binding motifs, there is often one haplotype that is preferentially bound by PRDM9 and thus becomes the initiation homologue. Curiously, PRDM9 initiates a DSB in close proximity to its own binding site, which is then repaired by the template homologue, thus disrupting the original binding site and replacing it with the sequence from the template homologue, to which it has lower binding affinity (Myers, Bowden et al. 2010). On a population genetic level this leads to fast erosion of highly used binding sites and ultimately this might lead to complete loss of binding sites. However, this has not led to a more uniform distribution of recombination in PRDM9-carrying organisms, but recombination hotspots continue to exist. This phenomenon has been called "the hotspot paradox".

PRDM9 zinc fingers experience accelerated evolution

The structure of PRDM9, containing an N-terminal domain that is highly conserved and a highly variable C-terminal zinc-finger domain creates a possible counter-measure to the effects of hotspot erosion. The zinc-finger domain of PRDM9 experiences accelerated evolution, especially at those positions important for DNA-binding (Oliver, Goodstadt et al. 2009). Even a single SNP within the DNA-binding zinc-fingers might lead to a different DNA-recognition motif for that allele, thus activating a completely different set of hotspots genome-wide. Hotspot usage between closely related species show very little overlap (Myers, Bowden et al. 2010) and even within species, hotspot usage might differ significantly due to the present PRDM9-alleles (shown for example in (Berg, Neumann et al. 2011)).
The erosion and evolution of new alleles is reminiscent of an "arms-race" scenario, where PRDM9 can evolve to bind new sites before hotspot erosion is complete.

Mismatch repair bias

Mismatch repair biases are independent of DSB initiation. In two human hotspots it was shown that TD can affect non-crossover recombinants while crossover recombinants do not reveal TD within the same individuals (Odenthal-Hesse, Berg et al. 2014). The bias cannot be attributed to DSB initiation but likely occurs during mismatch repair. The full mechanism to explain this remains elusive but it is possible that mismatches are handled differently in different crossover-generating

pathways (resulting in interfering or non-interfering crossovers), or else that mismatches influence the decision as to how recombination intermediates are resolved (Stahl and Foss 2010, Haber 2014, Odenthal-Hesse, Berg et al. 2014).

5.1.2 Stable recombination landscapes in PRDM9-deficient species

In organisms lacking PRDM9, recombination placement is organized in a different way. Some species, like honey bees (*Apis mellifera)* do not have hotspots, instead recombination is associated with genomic features like GC-content or microsatellite abundance (Ross, DeFelice et al. 2015).

Dogs lack PRMD9, yet they show a clustering of recombination in hotspots. These are associated with structural features of open chromatin (CpG-islands, H3K4me3) (Axelsson, Webster et al. 2012, Auton, Rui Li et al. 2013, Berglund, Quilez et al. 2014). Birds likewise lack PRDM9 yet show recombination clustered in hotspots, however their placement does not appear to be sequence-specific but rather dependent on structural features, such as chromatin modification and is stable across generations (Singhal, Leffler et al. 2015).

The hypothesis and current opinion in the field is, that eukaryotic PRDM9-dependent recombination is dependent on structural, rather than DNA-sequence signatures and thus hotspot usage between individuals and even between species of the canid tree (for example grey wolf and red fox) should be stable.

5.2 Aim of the experiment

The analyses in this chapter aim at investigating the stability of a canid recombination hotspot over shorter evolutionary timescales, which can also be projected into the future. This can be tested by investigating signs of hotspot erosion, which can be uncovered as transmission distortion and asymmetric crossover resolution distributions between reciprocal orientations.

5.3 Results

5.3.1 Assessing crossover associated transmission distortion (TD)

To detect significant distortions in the transmission of alleles from spermatocytes to sperm it is necessary to consider a large enough number of recombinant

molecules from reciprocal crossover orientations. Moreover, for obvious reasons, it is only possible to detect differences in transmission at heterozygous sites. Every marker across the hotspot region was analyzed independently, including only those individuals that were heterozygous at the respective site.

For consistency reasons transmission of the strong allele (S; G or C) over the weak allele (W; A or T) is shown in transitions and the purine allele (G or A) in transversions.

The transmission frequencies per informative site are shown in Figure 12 and Table 1. The numbers refer to actual observed numbers, as opposed to Poisson corrected numbers and statistical significance was tested separately for every SNP using two-tailed Fisher's exact test. Notably, only crossover-associated TD could be detected in the present assay. Non-crossover gene conversion might also be subject to transmission distortion but could not be picked up in the present assay.

5.3.2 Five SNPs show significant transmission distortion (TD)

Of the 78 markers, five markers showed significant TD ($P<0.05$). Two more SNPs showed TD that was marginally significant, with $P<0.1$. These were spread across the whole hotspot region (Figure 13).

Due to the overall lower number of sequenced recombinant molecules, compared to data from mouse or human sperm, it is possible that there are more markers subject to TD that this assay or the present panel of parental genotypes cannot detect. For every SNP it was assessed whether the TD is in favor of the strong allele (S; G or C) in transitions or the purine allele (R; G or A) in transversions and whether the dominant substitution creates or disrupts a CpG. The sites of these SNPs (including 10 bp up- and downstream) were subjected to a binding site search to find potential trans-acting factors responsible for the TD (Table 2).

Table 1 Transmission per marker. Position refers to chromosome27 (CanFam 3.1), S-SW refers to number of recombinants holding a strong allele in the strong to weak crossover orientation. W-WS refers to recombinants holding a weak allele in the weak to strong orientation, etc. Purine alleles are sorted as "strong" alleles in transversions and pyrimidine alleles as "weak". Statistical significance (P) refers to two-tailed Fisher's exact test on raw molecule count numbers (before Poisson correction).

	Pos Chr27	S-SW	W-WS	W-SW	W-SW	P	% S
Int1	16032687	0	55	0	52	1.000	50
Int3	16033410	2	53	2	50	1.000	50
Int4	16033468	0	15	0	20	1.000	50
Int5	16033653	2	53	4	48	0.429	48
Int6	16033979	4	48	5	50	1.000	49
Int7	16034366	1	23	1	20	1.000	50
Int8	16034701	5	47	6	49	1.000	49
Int10	16035299	1	3	4	2	0.523	29
Int11	16035449	8	47	7	45	1.000	51
Int12	16035528	6	40	7	44	1.000	50
Int13	16035615	8	47	7	45	1.000	51
Int14	16035799	5	26	13	18	0.049	37
Int15	16036036	5	26	13	18	0.049	37
Int16	16036529	10	36	8	43	0.797	53
Int17	16036544	10	36	8	34	0.797	51
Int19	16036933	11	26	15	20	0.327	43
Int22	16037387	5	13	10	8	0.176	36
Int23	16037392	8	8	5	11	0.473	59
Int25	16037568	7	11	6	12	1.000	53
Int26	16037643	16	39	20	30	0.304	45
Int28	16038089	10	3	7	4	0.659	57
Int29	16038312	11	8	13	7	0.748	46
Int30	16038422	24	31	30	21	0.126	42
Int31	16038526	16	8	8	16	0.042	67
Int33	16038845	10	5	13	2	0.390	40
Int34	16039080	29	7	27	12	0.298	56
Int35	16039219	12	1	9	2	0.576	55
Int36	16039387	10	5	16	4	0.451	43
Int37	16039448	11	4	16	4	0.700	47
Int39	16039689	10	5	16	4	0.451	43
Int40	16039763	7	0	7	0	1.000	50
Int41	16040231	25	2	18	11	0.010	65
Int42	16040474	51	4	45	7	0.351	53
Int43	16040553	41	6	40	4	0.741	48

Table 1 Transmission per marker (continued). Position refers to chromosome27 (CanFam 3.1), S-SW refers to number of recombinants holding a strong allele in the strong to weak crossover orientation. W-WS refers to recombinants holding a weak allele in the weak to strong orientation, etc. Purine alleles are sorted as "strong" alleles in transversions and pyrimidine alleles as "weak". Statistical significance (P) refers to two-tailed Fisher's exact test on raw molecule count numbers (before Poisson correction).

	Pos Chr27	S-SW	W-WS	W-SW	W-SW	P	% S
Int44	16040565	46	6	49	6	1.000	50
Int45	16040712	46	6	52	3	0.311	47
Int46	16040725	52	3	46	6	0.311	53
Int47	16040779	46	6	52	3	0.311	47
Int49	16040949	53	2	46	6	0.154	54
Int50	16041044	46	6	53	2	0.154	46
Int51	16041120	54	1	46	6	0.056	55
Int52	16041236	21	8	27	0	0.005	36
Int53	16041251	54	1	47	5	0.106	54
Int54	16041278	47	5	54	1	0.106	46
Int55	16041353	47	5	54	1	0.106	46
Int56	16041412	47	5	54	1	0.106	46
Int57	16041424	47	5	54	1	0.106	46
Int58	16041578	46	6	54	1	0.056	45
Int59	16041633	54	1	47	5	0.106	54
Int60	16041776	48	4	54	1	0.197	47
Int62	16041823	48	4	54	1	0.197	47
Int63	16041858	48	4	54	1	0.197	47
Int64	16041885	54	1	48	4	0.197	53
Int65	16042053	48	4	54	1	0.197	47
Int66	16042084	20	0	15	1	0.444	53
Int67	16042085	54	1	48	4	0.197	53
Int68	16042112	48	4	54	1	0.197	47
Int69	16042209	48	4	54	1	0.197	47
Int70	16042424	10	3	11	0	0.222	38
Int71	16042527	49	3	54	1	0.354	48
Int72	16042550	34	2	38	1	0.604	49
Int73	16042612	16	0	15	1	1.000	53
Int74	16042665	52	0	54	1	1.000	51
Int75	16042739	27	1	23	0	1.000	48
Int76	16042767	10	3	11	0	0.223	38
Int77	16042943	54	1	52	0	1.000	49
Int78	16043032	55	0	52	0	1.000	50

De-novo crossover frequency clustering based on underlying genotypes

The clustering of *de-novo* crossover frequencies based on the underlying genotype has been observed in mouse and humans and has been attributed solely to the binding affinities of PRDM9-variants to the different sequence motifs (Berg, Neumann et al. 2010, Berg, Neumann et al. 2011). A hotspot activated by a certain PRMD9-allele will be used at high frequencies in an individual carrying the respective underlying PRDM9-binding site at both homologues ("homozygous hot

genotype"). An individual homozygous for a disrupted binding site will show drastically reduced recombination frequencies at this hotspot ("homozygous cold genotype"), because PRDM9 cannot bind. And an individual heterozygous for one activating and one disrupted binding site will show intermediate recombination frequency ("heterozygous genotype") and characteristic resolution point shifts resulting in over-transmission of the "cold allele" (Berg, Neumann et al. 2010, Cole, Keeney et al. 2010).

For every TD-SNP, individuals were assigned to the respective groups (Homozygous for the "weak" allele, homozygous for the "strong" allele or heterozygous) and log-likelihood best-fit values for *de-novo* crossover frequencies were calculated for every group (Figure 14, Table 2). With seven individuals tested in total, it is difficult to test whether any differences observed are significant, thus 95%-confidence intervals of the log-likelihood best-fit value are plotted as error bars as an indicator of significance.

Genomic background of TD-SNPs

SNP 1 and SNP 2 are 257 bp apart and likely co-converted, both show over-transmission of the weak allele. SNP 2.C is located in a CpG and surrounded by a low-complexity area (surrounding 100 bp have a GC-content of ca. 30% and adjacent sequence shows signatures of a microsatellite repeat, see Figure 13). The CpG lies outside the annotated CpG-island and might be methylated. It is possible that in this case the single-base conversion does not occur after meiotic DSB-repair but shows AT-biased repair as observed in mutation-repair mechanisms after spontaneous deamination of the methylated cytosine.

SNP 3 is placed centrally in the hotspot region, just downstream of the CpG-island and of PKP2-Exon3 (Figure 13). The *de-novo* crossover frequencies showed an almost 5-fold difference between individuals homozygous for the G-allele and the individual homozygous for the T-allele, whereas the heterozygous individuals showed intermediate frequencies.

SNP 4 is placed within an annotated L2b LINE element and within a CpG outside the CpG-island. It is possible that the substitution has an effect on LINE-element activity.

The distortion in transmission at SNP 5 does not come from an overall skewed resolution point distribution at this site, but is due the presence of complex crossovers in two individuals, showing single-base conversion of the strong to weak allele at this site, thus two out of five TD-SNPs show crossover associated over-transmission of the weak allele, two of the strong allele and one shows no distortion in single-switch crossovers but over-transmission of the weak allele only in complex crossovers.

Table 2 TD-SNPs (transmission significantly different from 50:50). Position refers to CanFam 3.1 assembly, TD shows over-transmitted:under-transmitted base, P was determined using two-tailed Fisher's exact test on raw (not Poisson-corrected) recombinant molecule counts. N total refers to the number of recombinant molecules assessed for the respective SNP (i.e. molecules recovered from individuals heterozygous at the respective SNP). N S->W refers to the number of recombinants recovered for the strong->weak orientation. Log-likelihood best fit values for *de-novo* crossover frequencies for respective groups of individuals (homozygous for the strong allele, heterozygous or homozygous for the weak allele) are shown in bold numbers, 95%-confidence intervals are reported in brackets. "Na" refers to groups where no individual from the panel possessed the respective genotype (for example there was no individual homozygous for C at SNP 1 within the tested panel).

SNP	Position	TD	Ratio	P	N total	N S->W	Freq HomS	Freq Het	Freq HomW	Placed in element?	Disrupts/ creates CpG?
SNP 1	chr27:16,035,779	T:C	63:37	0.05	62	31	na	**15.53** cM/Mb (95%-CI: 13.53-17.76 cM/Mb)	**11.52** cM/Mb (95%-CI: 9.49-13.88)	no	no
SNP 2	chr27:16,036,036	T:C	63:37	0.05	62	31	na	**15.53** cM/Mb (95%-CI: 13.53-17.76 cM/Mb)	**11.52** cM/Mb (95%-CI: 9.49-13.88)	no	yes
SNP 3	chr27:16,038,526	G:T	67:33	0.04	48	24	**23.42** cM/Mb (95%-CI: 19.98-27.31 cM/Mb)	**12.97** cM/Mb (95%-CI: 10.58-15.76 cM/Mb)	**5.02** cM/Mb (95%-CI: 3.6-6.79 cM/Mb)	no	no
SNP 4	chr27:16,040,231	G:T	65:35	0.01	56	27	**35.78** cM/Mb (95%-CI: 28.71-44.22 cM/Mb)	**10.70** cM/Mb (95%-CI: 9.33-12.22 cM/Mb)	na	yes	yes
SNP 5	chr27:16,041,236	T:C	64:36	0.01	56	29	na	**10.70** cM/Mb (95%-CI: 9.33-12.22 cM/Mb)	**35.78** cM/Mb (95%-CI: 28.71-44.22 cM/Mb)	no (-33 bp)	no

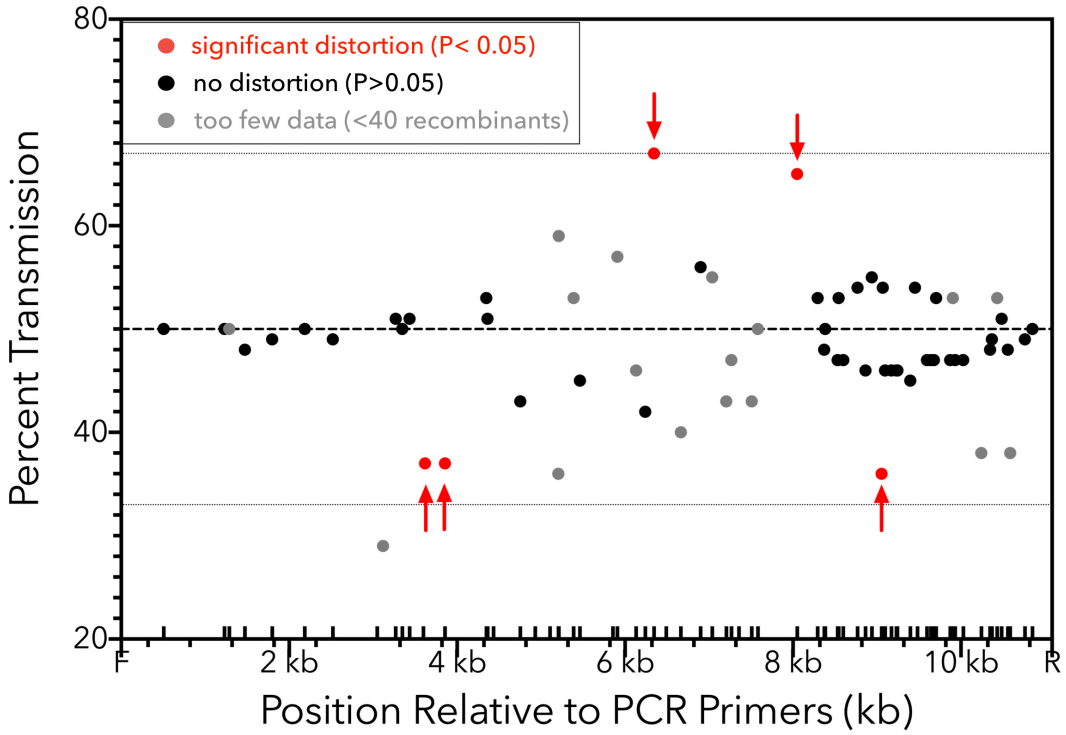


Figure 12 Transmission of strong (G/C) or purine (G/A) alleles. Every heterozygous marker present in at least one of the individuals is represented by a dot. Grey dots represent SNPs with less than 40 recombinant molecules sequenced. SNPs with significantly different (P<0.050) transmission from 50:50 are marked in red and with red arrows. Dotted line represents equal transmission (50:50) of both alleles. Any point above this line is indicative of strong over weak over-transmission, whereas points below this line represent SNPs where the weak (or pyrimidine) allele is over-transmitted. Grey lines represent 33% and 67% transmission respectively.

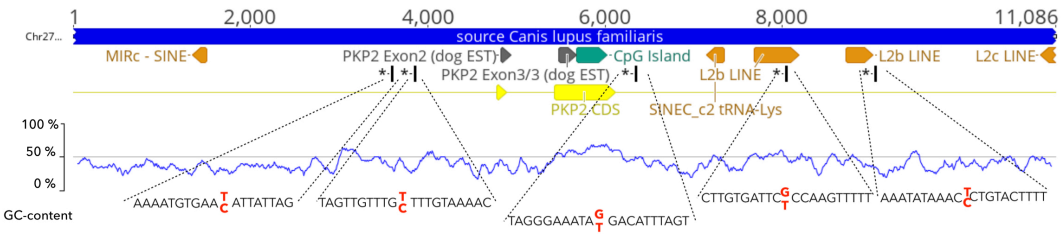


Figure 13 Genomic background and surrounding sequence of significantly distorted (from 50:50 transmission) SNPs (TD-SNPs). Annotations include LINE and SINE elements, CpG island, mammalian exons, dog ESTs and in black the positions of the TD-SNPs. Base pair annotations refer to distance from outer allele-specific forward primer. SINEC elements are canid specific elements. The local GC-content in percent is shown below in blue line, 50% is indicated by dotted line.

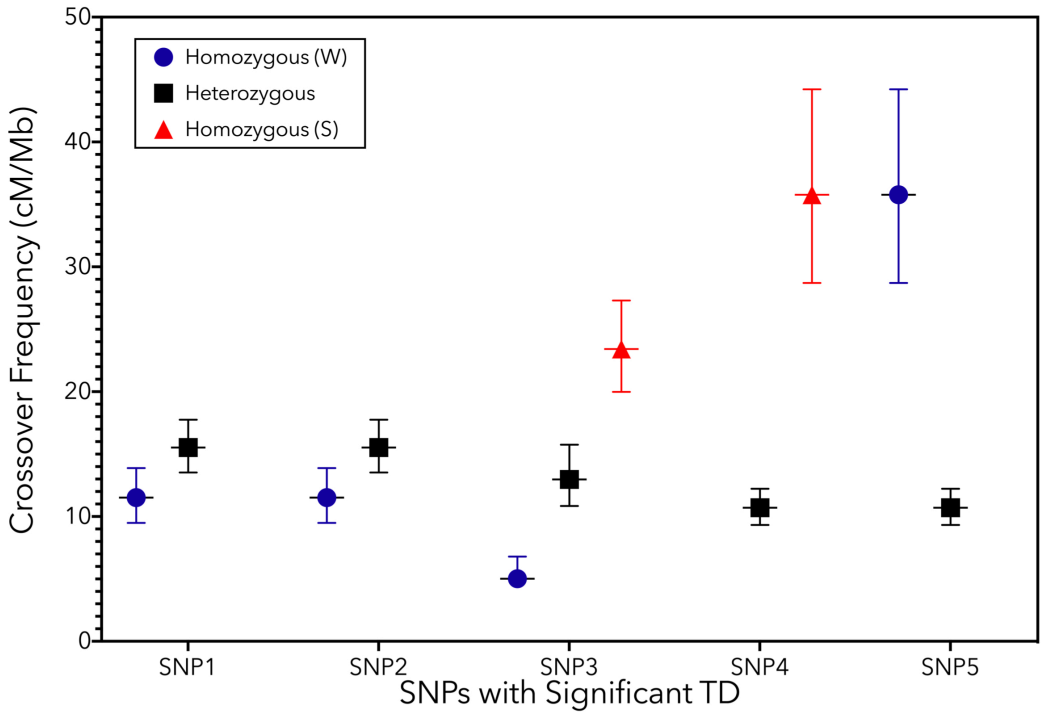


Figure 14 *De-novo* crossover frequencies per genotype group at TD-SNPs. For every TD-SNP, log-likelihood best-fit values for groups pooled based on their underlying genotype: Homozygous for either weak (blue dot) or strong (red triangle) allele or heterozygous (black square). Error bars represent 95%-confidence intervals.

5.4 Shifts in resolution point distributions per TD-SNP

Crossover-associated transmission distortion can occur when the location of crossover intermediate resolution is different between reciprocal crossover orientations. In one orientation, most crossovers are resolved in a more upstream position, in the other orientation more downstream, leading to an over-transmission of one or more markers in the area subject to the shift.

Resolution points were analyzed separately per reciprocal orientation. It is important to note that the haplotypes and reciprocal orientations do not refer to the initially used PCR haplotypes (A and B) but to the haplotypes at the respective TD-SNP. For every TD-SNP separately the transmission per marker was assessed and tested using two-tailed Fisher's exact test. Complex crossovers were included, and the final switch was assigned to the first stretch of at least two consecutive kilobases from the opposite haplotype.

The five TD-SNPs can be grouped into three underlying haplotypes. SNP 1 and SNP 2 are heterozygous in the same individuals, all of which have the same phase between both SNPs so the crossover resolution map is identical for both SNPs. The centrally placed SNP 3 is heterozygous in two of the tested individuals and placed ca. 2.5 kb downstream from SNP 2. The two remaining SNPs (SNP 4 and SNP 5) were heterozygous in five individuals with identical phase between the SNPs so their resolution point maps are identical.

The three resolution point maps are shown in Figure 15. As expected, the Gaussian regression fit was poor as described in previous chapters, so statistics inferred should be treated merely as indication but might not represent the underlying distribution. Differences in resolution maps were characterized using two-tailed Fisher's exact test on Poisson corrected crossover numbers, rounded to the next integer. For every SNP separately, it was asked whether the number of crossovers already switched to the other haplotype is different from each other in reciprocal crossover orientations. Here, only those crossovers obtained from individuals that are heterozygous at the query SNP were taken into account.

The reciprocal distributions for both the two upstream as well as the two downstream SNPs show no significant differences in resolution. The resolution distributions for the centrally placed SNP 3 are shifted between reciprocal orientations, with significant differences in haplotype switches between reciprocal orientations (significantly different SNPs are indicated by asterisks, see also Supplementary Information S27-S29). The midpoints from cumulative Gaussian regression are shifted by 1.3 kb, while the overall width of the distribution is the same in both orientations (7688 bp and 7956 bp respectively). The T→C orientation shows overall earlier haplotype switches than the C→T orientation, facilitating the C:T over-transmission at SNP 3.

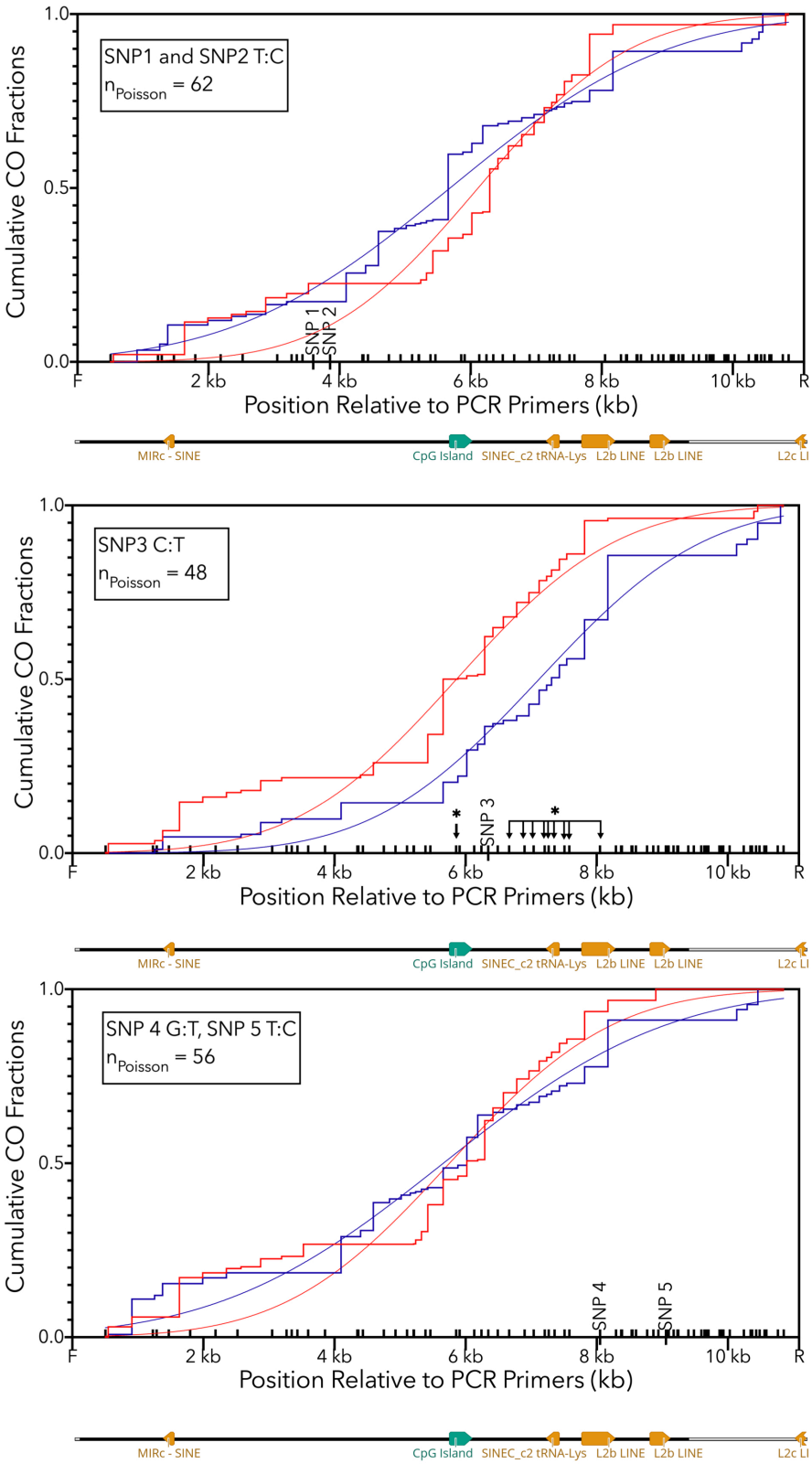


Figure 15 Crossover resolution maps per reciprocal crossover orientation grouped by haplotypes at TD-SNPs. Cumulative fractions of Poisson corrected crossover orientations are shown in blue (strong to weak) and red (weak to strong) lines respectively; only recombinants obtained from individuals heterozygous at the respective SNPs were taken into account. Smoothed curves represent Gaussian non-linear regression fits. Underlying TD-SNPs are indicated and any significant transmission difference in reciprocal crossover resolution ("crossover resolution shift") are indicated by asterisks.

5.5 Candidate influencing factors

The results described in the paragraphs above suggest that there might be factors influencing fine-scale DBS initiation or crossover resolution, independent of PRDM9, at least in the analyzed dog recombination hotspot. SNP 3, with significant TD as a result of a significant shift in reciprocal crossover resolution maps and a 5-fold difference of hotspot heat of individuals carrying the SNP 3.C over SNP 3.T allele, and intermediate frequencies for the heterozygotes, is a strong candidate for a marker influenced by a regulating trans-acting factor.

The JASPAR motif search (described in 3.11.4) revealed a large number of transcription factors binding sites within the hotspot region. Given that the hotspot is located within a promoter region, this was expected, as promoters naturally contain binding sites for many factors regulating transcription.

The majority of the factors found bind DNA via recognition sequences of low to moderate specificity. These include mainly Helix-turn-helix proteins, homeodomain proteins and basic leucine zipper proteins, predicted to bind this site. Interestingly, at a cutoff of 80% similarity, most of the TFs found to bind one allele were not found to bind the other allele, confirming that the single SNP is sufficient to potentially disrupt binding sites. The motif search revealed several candidates with different binding affinities for the C- and T-allele at the most promising candidate regulatory SNP 3.

RORC

The nuclear receptor ROR (retinoic acid receptor-related orphan receptor)-gamma binds DNA as a monomer to ROR response elements via a C4-zinc finger domain and is a key regulator in multiple cellular processes. It is a regulator of circadian expression of clock genes and involved in the regulation of rhythmic expression in metabolic pathways. It is expressed in the male reproductive organs in humans, including testis, if at moderate to low levels. Tissue specificity for meiotic or supporting tissue data is not available. The over-transmitted SNP3.G allele

provides higher binding affinity for the protein, whereas the T-allele disrupts the motif.

HLTF

Helicase-like transcription factor (HLTF) has helicase and E3 ubiquitin ligase activity and possesses nucleosome-remodeling activity in humans, which may be required for transcriptional activation or repression. It is also involved in post-replication repair of damaged DNA, although exact mechanisms remain elusive. In humans, it is expressed in many tissues but with high relative expression in testis and specifically in seminiferous ducts containing meiotic cells.

The motif search revealed binding sites both at SNP 3 and SNP 4. In both cases the under-transmitted weak-allele (T) is bound with slightly higher affinity than the strong allele (90.8 % similarity versus 91.1% similarity and 93.9% versus 94.1% similarity). These binding affinities are based on any vertebrate binding motifs however and could be different in dogs.

NOTO and NKX2-8

Homeobox protein notochord (NOTO) is a transcription-regulator acting during notochord development. It is expressed in brain but shows even higher expression in testis. Homeobox protein Nkx-2.8 is expressed at very high levels in testis and to some extent in brain tissue. There is no functional data available to date. Both SNP 1 and SNP 3 provide a binding site for NOTO, but with similar binding affinities for both alleles. NKX2-8 likewise binds both SNP 3 and SNP 4 with similar binding affinities for both alleles.

5.6 CG-biased gene conversion

Canine recombination is associated with GC-rich regions but it is unclear whether the recombination machinery is drawn towards GC-rich regions, or whether regions in which recombination takes place become GC-rich over time as a consequence of GC-biased gene conversion. Figure 12 is a representation of transmission frequencies per marker within the tested hotspot region. Transmission is always shown for the strong allele (G or C) in transitions and for the purine allele (G or A) in transversions. Of the five SNPs exhibiting significant TD, all

of which are transitions, two show over-transmission of the strong allele, and three of the weak allele. It is possible that some of these TD effects are due to co-conversion, occurring when the overall resolution distribution is skewed. The central SNP 3(G:T) is the only marker exhibiting a significant skew in resolution point distributions between reciprocal orientations as well as significant differences between hotspot usages segregating with parental haplotypes. This SNP shows strong over weak biased gene conversion, which could be an index of biased gene conversion. Overall the present data neither confirm nor reject systematic GC-biased gene conversion in the analyzed dog hotspot.

5.7 Complex Crossovers

A complex crossover (CCO) describes a recombinant molecule that contains more than one switch from one haplotype to the other. This could be a gene conversion stretch placed up- or downstream from the crossover resolution point or a double crossover. In PRDM9-dependent meiosis, CCOs are rare, with the vast majority of recombinant molecules showing just one single haplotype switch or one short non-crossover gene conversion stretch (for example examined in (Arbeithuber, Betancourt et al. 2015)).

Of the 104 recombinant molecules sequenced in the present data set, 25 were complex (24%). CCOs were found in most individuals tested, although the ratio of complex to normal crossovers differed between individuals, with individual d06 showing 17 out of 32 CCO, whereas d26 showed only 4 out of 35 CCO. The distance between conversion and full haplotype switch varied between several 10s of base pairs and over 1000 base pairs. CCOs occurred in high-input as well as low-input pool PCRs and converted bases differed between molecules, although some conversions occurred repeatedly.

5.8 Discussion

Contradictory to the current hypothesis, that the canid recombination landscape should be stable both on a broad- as well as fine-scale level the results from this thesis provided evidence for inter-individual variation both in hotspot usage as well as fine-scale crossover resolution. The presence of significant TD and a shift

between reciprocal crossover orientations are indicative of a regulating mechanism that is sequence-based, such as a trans-acting factor.

The motif search for transcription factors binding those sites specifically, at which TD was observed, revealed binding sites for many different transcription factors. For the two upstream sites these are mainly forkhead box proteins, a group of genes shown to play a role in diverse processes during development, metabolism, cell-cycle control and others. Compared to C_2H_2-zinc finger proteins such as PRDM9, their DNA-binding recognition sequence is far less specific. The central SNP displays significant TD, significant shift in reciprocal crossover resolution maps and a ca. 5-fold difference of hotspot usage of individuals carrying the SNP 3.C over SNP 3.T and intermediate frequencies for the heterozygotes, however the motif-search for the underlying sequence did not reveal a strong candidate to influence crossover outcome. Factors showing different binding affinities to the different alleles were either reported to have low expression in meiotic cells or else show only moderate binding affinity to either of the alleles. Those factors known to be expressed in testis specifically showed no pronounced difference in binding affinities for either allele if looking at the SNPs separately. However, two of them appeared as candidates for two of the five SNPs, suggesting that the differences in resolution point distribution and hotspot usage may also be due to an additive effect of several factors in dog and not attributable to a single factor like PRDM9 in other mammals.

It is also possible that the motif binding affinities for the dog orthologs of transcription factors are different than the ones reported, which are based on the mouse or human ortholog. The binding affinities of the transcription factors discussed here could be different in dogs and there might be other factors that the search misses for that reason.

Dog hotspots are significantly clustered around transcription start sites containing CpG-islands. Previous data suggests that dog recombination uses un-methylated CpG-islands as structural hallmarks to place meiotic DSBs (Berglund, Quilez et al. 2014). Some of the substitutions showing significant TD create or disrupt a CpG,

which might influence hotspot usage if there is a causal link between CpG-island and hotspot locations.

Two of the SNPs showing TD are placed in or close to annotated L2b-LINE elements. While LINE-activity is thought to be high in dogs compared to humans (Wang and Kirkness 2005), this refers mainly to L1 and not L2 elements. Both L2b LINE elements annotated in the analyzed hotspot region show high sequence divergence, indicating that they might no longer be active.

The Hotspot region further contains a canine specific tRNA-derived SINE in anti-sense orientation (SINEC-c2 placed centrally within the hotspot, just downstream of the CpG-island). Loci bi-morphic for SINE insertion are highly prevalent in canids and have played an important role in the dynamics of genome evolution (reviewed in (Malakowski 2000)). SINEs can also affect adjacent genes, influencing transcription, mRNA splicing and protein expression (Wang and Kirkness 2005). Individual d26 has two heterozygous SNPs within the centrally placed SINE element (chr27:16,039,387-A/G and chr27:16,039,448-C/G). There are two more annotated SINEC-c elements both just upstream and just downstream of the analyzed PCR amplicon. Generally, the number of annotated repetitive elements is far above the median in this hotspot.

It would be highly interesting to further analyse the role of repetitive elements, especially SINEC elements, on canid recombination, using a larger panel of individuals showing heterozygous SINEC sites.

Dog hotspots are wider than PRDM9-dependent hotspots, which is supported by the resolution point analysis in the present study, as well as by an unpublished data set by Galina Petukhova (*personal comm.)* which shows wider peaks in DMC1-ChIP-sequencing (as a proxy for DSB initiation sites). The data from this thesis suggest that rather than having one distinct hotspot peak, dogs use a broader area in which recombination can take place. One explanation for wide recombination resolution peaks could be extended Holliday-junction migration. While PRDM9-dependent hotspots have boundaries defined by the trimethylation marks placed by PRDM9 (Baker, Walker et al. 2014), default sites might lack these

boundaries and cause Holliday-junctions to migrate further and resolve into a crossover or non-crossover much further away from the DSB site than observed in PRMD9-dependent hotspots (Stahl and Foss 2010).

Another explanation could be, that the placement of initiating DSBs is less concentrated. The high number of complex crossovers suggests that there are many recombinant molecules, which have experienced more than one DSB. While the assay used for the present work does not allow for the detection of double-crossovers, the advanced sequencing approach does detect recombinant molecules in which one DSB could have been repaired by a non-crossover event, the other one by a crossover event.

Data from other groups also support this theory. In the DMC1-ChIP-sequencing data, many of the hotspots detected were located in regions of "high-background" recombination, areas in which recombination was generally elevated. The historical recombination data from Adam Auton (Auton, Rui Li et al. 2013) also point towards wider and less intense peaks of recombination activity. The crossover resolution maps (4.3.2) detect recombination resolution almost across the whole 10.8 kb analyzed. It is likely that if a larger interval was analyzed, more breakpoints could be found further away from the center. The distribution of resolution maps does not follow normal distribution like in PRDM9-dependent hotspots. It is possible that multiple DSB initiation sites lie within the region, with overlapping resolution point distributions.

Taken together, the results from this chapter show that contradictory to the present hypothesis, dog hotspots are likely subject to fine-scale regulation and not fully stable between individuals. Potential factors include methylation status of CpGs, trans-acting factors and repetitive element involvement and it will be highly interesting to investigate these further.

6 General Discussion and Perspectives

The present analysis of meiotic recombinants in dogs on a fine scale has revealed unexpected variation in hotspot usage as well as crossover resolution sites between individuals, even though the mammalian recombination regulator PRDM9 is absent in dogs.

The extensive width of the dog hotspot could be explained by a) the more widespread placement of recombination initiating DSBs and b) increased dHJ migration compared to PRDM9-dependent hotspots; or a combination of both. Historical dog recombination hotspots have shown to be associated with genomic features, like histone modification (H3K4me3) and CpG-islands and might be linked to active DNA-transcription sites. In PRDM9-functional mammals, recombination can be detected in similar sites as well but at much lower levels. They likely occur in later meiotic phases, when PRDM9 is no longer expressed, suggesting that these are the "default"-recombination sites that are not used when PRDM9 is present and directs recombination away from functional elements. In dogs, recombination initiation seems to be drawn to "default" sites. The present study is the first to reveal that default-DSBs can resolve into recombinant molecules and show that recombination in default sites does not necessarily lead to incomplete meiosis, as has been suggested from a study in PRDM9-knockout-mice (Brick, Smagulova et al. 2012).

Intriguingly, neither hotspot usage nor crossover resolution is stable between individuals at this dog default recombination site. The crossover-associated transmission distortion observed at a central SNP within the hotspot suggests a fine-scale regulating mechanism that might affect either DSB placement (initiation bias) or crossover resolution after a common intermediate (mismatch repair bias). *De-novo* crossover frequencies segregate in correlation to the underlying haplotypes at this central SNP, showing 5-fold differences between crossover frequencies of respective homozygous genotypes and intermediate frequencies for the heterozygous genotypes. Yet, the over-transmitted allele is correlated with

higher recombination frequencies than the under-transmitted allele, which is contrary to the observation made with PRDM9-associated TD. With PRDM9 disrupting its own binding site, the lower-affinity allele ("cold" allele) is over-transmitted, causing meiotic drive against hotspot activating alleles. The data from this thesis suggest a hotspot-stabilizing meiotic drive in favor of the activating allele. The mechanism responsible for this effect remains elusive. It is possible that there is an alternative regulating factor to PRDM9 in dogs, which can cause increased DSB initiation on the unbound allele. It is also possible that the bias arises when recombination intermediates have already formed, due to biased mismatch repair. This has been shown to occur in non-crossover recombination in humans (Odenthal-Hesse, Berg et al. 2014). Gap repair of the resected ends must always result in repair using the other, unbroken homologue and thus cause over-transmission of the unbroken allele. In contrast, increased migration of recombination intermediates, as was proposed as an alternative recombination mechanism by Stahl et al (Stahl and Foss 2010), could cause mismatches, which might be subject to bias towards maintaining the template allele. Resolution points were detected across the whole amplified hotspot region, whereas the dog DSB initiation map suggests initiation zones of ca. 4-5 kb (unpublished data, *personal comm*. Galina Petukhova). Taken together, the data agree with wide zones of DSB initiation and suggest that recombination intermediate migration might further contribute to the extensive width of crossover resolution maps.

This work has proven that dog recombination hotspots are still active *de-novo* and are subject to fine-scale inter-individual variation. Future work could include genome-wide analysis of chromatin status, DNA-methylation, protein expression and recombination initiation to resolve the relationship between genomic features, transcription and meiotic recombination in PRDM9-deficient recombination. It would also be interesting to fine-scale analyze recombinants from more LD hotspots, as well as hotspots from the unpublished DSB initiation data set, to create a link between initiation and resolution in dog recombination hotspots. The analysis of non-crossover recombinants could give insights into recombination

intermediate resolution and possible non-crossover associated transmission distortion.

7 Literature

Studies revealing facts that are now assumed to be common knowledge where not explicitly cited. These were defined as facts published in textbooks. More specifically, general biological facts and information about meiosis were drawn from "Alberts et al., Molecular Biology of the Cell" (Alberts, Johnson et al. 2008) and recombination-specific data can be found in "Haber, Genome Stability" (Haber 2014).

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter (2008). Molecular Biology of the Cell. New York, Garland Science.

Allers, T. and M. Lichten (2001). "Intermediates of Yeast Meiotic RecombinationContain Heteroduplex DNA." Cell **8**(1): 225-231.

Arbeithuber, B., A. J. Betancourt, T. Ebner and I. Tiemann-Boege (2015). "Crossovers are associated with mutation and biased gene conversion at recombination hotspots." Proc Natl Acad Sci U S A **112**(7): 2109-2114.

Auton, A., Y. Rui Li, J. Kidd, K. Oliveira, J. Nadel, J. K. Holloway, J. J. Hayward, P. E. Cohen, J. M. Greally, J. Wang, C. D. Bustamante and A. R. Boyko (2013). "Genetic recombination is targeted towards gene promoter regions in dogs." PLoS Genet **9**(12): e1003984.

Axelsson, E., M. T. Webster, A. Ratnakumar, L. Consortium, C. P. Ponting and K. Lindblad-Toh (2012). "Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome." Genome Res **22**(1): 51-63.

Baker, C. L., M. Walker, S. Kajita, P. M. Petkov and K. Paigen (2014). "PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration." Genome Research **24**(5): 724-732.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop and B. de Massy (2010). "PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice." Science **327**(5967): 836-840.

Berg, I. L., R. Neumann, K. W. Lam, S. Sarbajna, L. Odenthal-Hesse, C. A. May and A. J. Jeffreys (2010). "PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans." Nat Genet **42**(10): 859-863.

Berg, I. L., R. Neumann, S. Sarbajna, L. Odenthal-Hesse, N. J. Butler and A. J. Jeffreys (2011). "Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations." Proc Natl Acad Sci U S A **108**(30): 12378-12383.

Berglund, J., J. Quilez, P. F. Arndt and M. T. Webster (2014). "Germline methylation patterns determine the distribution of recombination events in the dog genome." Genome Biol Evol **7**(2): 522-530.

Billings, T., E. D. Parvanov, C. L. Baker, M. Walker, K. Paigen and P. M. Petkov (2013). "DNA binding specificities of the long zinc-finger recombination protein PRDM9." Genome Biol **14**(4): R35.

Bishop, D. K., D. Park, L. Xu and N. Kleckner (1992). "DMC1: A meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression." Cell **69**(3): 439-456.

Brick, K., F. Smagulova, P. Khil, R. D. Camerini-Otero and G. V. Petukhova (2012). "Genetic recombination is directed away from functional genomic elements in mice." Nature **485**(7400): 642-645.

Cannavo, E. and P. Cejka (2014). "Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks." Nature **514**(7520): 122-125.

Cao, L., E. Alani and N. Kleckner (1990). "A Pathway for Generation and Processing of Double-Strand Breaks during Meiotic Recombination in S. cerevisiae." Cell **61**(6): 1089-1101.

Carpenter, T. C. (1975). "Electron microscopy of meiosis in Drosophila melanogaster females: II: The recombination nodule-a recombination-associated structure at pachytene?" Proc Natl Acad Sci U S A **72**(8): 3186-3189.

Choo, Y. and A. Klug (1994). "Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions." Proc Natl Acad Sci U S A **91**(23): 11168-11172.

Choo, Y. and A. Klug (1994). "Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage." Proc Natl Acad Sci U S A **91**(23): 11163-11167.

Cole, F., F. Baudat, C. Grey, S. Keeney, B. de Massy and M. Jasin (2014). "Mouse tetrad analysis provides insights into recombination mechanisms and hotspot evolutionary dynamics." Nature Genetics **46**(10): 1072-1080.

Cole, F. and M. Jasin (2011). "Isolation of meiotic recombinants from mouse sperm." Methods Mol Biol **745**: 251-282.

Cole, F., S. Keeney and M. Jasin (2010). "Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis." Mol Cell **39**(5): 700-710.

Davies, B., E. Hatton, N. Altemose, J. G. Hussin, F. Pratto, G. Zhang, A. G. Hinch, D. Moralli, D. Biggs, R. Diaz, C. Preece, R. Li, E. Bitoun, K. Brick, C. M. Green, R. D. C. Amerini-Otero, S. R. Myers and P. Donnelly (2016). "Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice." Nature **530**(7589): 171-+.

Diagouraga, B., J. A. J. Clement, L. Duret, J. Kadlec, B. de Massy and F. Baudat (2018). "PRDM9 Methyltransferase Activity Is Essential for Meiotic DNA Double-Strand Break Formation at Its Binding Sites." Mol Cell **69**(5): 853-865 e856.

Eizirik, E., W. J. Murphy, K. P. Koepfli, W. E. Johnson, J. W. Dragoo, R. K. Wayne and S. J. O'Brien (2010). "Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences." Mol Phylogenet Evol **56**(1): 49-63.

Elbarbary, R. A., B. A. Lucas and L. E. Maquat (2016). "Retrotransposons as regulators of gene expression." Science **351**(6274): aac7247.

Fawcett, D. W. (1956). "The fine structure of chromosomes in the meiotic prophase of vertebrate spermatocytes." J Biophys Biochem Cytol **2**(4): 403-406.

Foss, H. M., R. Lande, F. W. Stahl and C. M. Steinberg (1993). "Chiasma Interference as a Function of Genetic Distance." Genetics **133**(3): 681-691.

Garcia, V., S. E. Phelps, S. Gray and M. J. Neale (2011). "Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1." Nature **479**(7372): 241-244.

Gravel, S., J. R. Chapman, C. Magill and S. P. Jackson (2008). "DNA helicases Sgs1 and BLM promote DNA double-strand break resection." Genes Dev **22**(20): 2767-2772.

Grey, C., P. Barthes, G. Chauveau-Le Friec, F. Langa, F. Baudat and B. de Massy (2011). "Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination." PLoS Biol **9**(10): e1001176.

Haber, J. E. (2014). Genome Stability DNA repair and recombination. New York, Garland Science (Taylor & Francis Group).

Hayashi, K., K. Yoshida and Y. Matsui (2005). "A histone H3 methyltransferase controls epigenetic events required for meiotic prophase." Nature **438**(7066): 374-378.

Holliday, R. (2009). "A mechanism for gene conversion in fungi." Genetical Research **5**(2): 282-304.

Hong, S., Y. Sung, M. Yu, M. Lee, N. Kleckner and K. P. Kim (2013). "The logic and mechanism of homologous recombination partner choice." Mol Cell **51**(4): 440-453.

Hunter, N. (2007). Meiotic Recombination. Molecular Genetics of Recombination. A. Aguilera and R. Rothstein. Berlin, Heidelberg, Springer. **17**.

Hunter, N. (2015). "Meiotic Recombination: The Essence of Heredity." Cold Spring Harb Perspect Biol **7**(12).

Hunter, N. and N. Kleckner (2001). "The Single-End Invasion: An Asymmetric Intermediate at the Double-Strand Break to Double-Holliday Junction Transition of Meiotic Recombination." Cell **106**: 59-70.

J.M., M. (1956). "Chromosomal Structures in Crayfish Spermatocytes." J Biophys Biochem Cytol **2**(2): 215-218.

Jeffreys, A. J., J. K. Holloway, L. Kauppi, C. A. May, R. Neumann, M. T. Slingsby and A. J. Webb (2004). "Meiotic recombination hot spots and human DNA diversity." Philos Trans R Soc Lond B Biol Sci **359**(1441): 141-152.

Jeffreys, A. J., L. Kauppi and R. Neumann (2001). "Intensely punctate meiotic recombination in the class IIregion of the major histocompatibility complex." Nat Genet(29): 217-222.

Jeffreys, A. J., J. Murray and R. Neumann (1998). "High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot." Mol Cell **2**(2): 267-273.

Jeffreys, A. J., J. Murray and R. Neumann (1998). "High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot." Mol Cell **2**(2): 267-273.

Kauppi, L., A. J. Jeffreys and S. Keeney (2004). "Where the crossovers are: recombination distributions in mammals." Nat Rev Genet **5**(6): 413-424.

Kauppi, L., C. A. May and A. J. Jeffreys (2009). Analysis of Meiotic Recombination Products from Human Sperm. Meiosis. S. Keeney, Humana Press, Springer. **1:** 323-355.

Kebschull, J. M. and A. M. Zador (2015). "Sources of PCR-induced distortions in high-throughput sequencing data sets." Nucleic Acids Res **43**(21): e143.

Keeney, S. (2008). "Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis." Genome Dyn Stab **2**: 81-123.

Keeney, S., C. N. Giroux and N. Kleckner (1997). "Meiosis-Specific DNA Double-Strand BreaksAre Catalyzed by Spo11, a Memberof a Widely Conserved Protein Family." Cell **88**(3): 375-384.

Khil, P. P., F. Smagulova, K. M. Brick, R. D. Camerini-Otero and G. V. Petukhova (2012). "Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA." Genome Res **22**(5): 957-965.

Kleckner, N., D. Zickler, G. H. Jones, J. Dekker, R. Padmore, J. Henle and J. Hutchinson (2004). "A mechanical basis for chromosome function." Proc Natl Acad Sci U S A **101**(34): 12592-12597.

Lao, J. P., S. D. Oh, M. Shinohara, A. Shinohara and N. Hunter (2008). "Rad52 promotes postinvasion steps of meiotic double-strand-break repair." Mol Cell **29**(4): 517-524.

Lindblad-Toh, K., C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe, M. Kamal, M. Clamp, J. L. Chang, E. J. Kulbokas, M. C. Zody, E. Mauceli, X. H. Xie, M. Breen, R. K. Wayne, E. A. Ostrander, C. P. Ponting, F. Galibert, D. R. Smith, P. J. deJong, E. Kirkness, P. Alvarez, T. Biagi, W. Brockman, J. Butler, C. W. Chin, A. Cook, J. Cuff, M. J. Daly, D. DeCaprio, S. Gnerre, M. Grabherr, M. Kellis, M. Kleber, C. Bardeleben, L. Goodstadt, A. Heger, C. Hitte, L. Kim, K. P. Koepfli, H. G. Parker, J. P. Pollinger, S. M. J. Searle, N. B. Sutter, R. Thomas, C. Webber, E. S. Lander and B. I. G. S. Plat (2005). "Genome sequence, comparative analysis and haplotype structure of the domestic dog." Nature **438**(7069): 803-819.

Lipkin, S. M., P. B. Moens, V. Wang, M. Lenzi, D. Shanmugarajah, A. Gilgeous, J. Thomas, J. Cheng, J. W. Touchman, E. D. Green, P. Schwartzberg, F. S. Collins and P. E. Cohen (2002). "Meiotic arrest and aneuploidy in MLH3-deficient mice." Nat Genet **31**(4): 385-390.

Malakowski, W. (2000). "Genomic scrap yard: how genomes utilize all that junk." Gene **259**: 61-67.

Mihola, O., F. Pratto, K. Brick, E. Linhartova, T. Kobets, P. Flachs, C. L. Baker, R. Sedlacek, K. Paigen, P. M. Petkov, R. D. Camerini-Otero and Z. Trachtulec (2019). "Histone methyltransferase PRDM9 is not essential for meiosis in male mice." Genome Res **29**(7): 1078-1086.

Mimitou, E. P. and L. S. Symington (2008). "Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing." Nature **455**(7214): 770-774.

Mimitou, E. P., S. Yamada and S. Keeney (2017). "A global view of meiotic double-strand break end resection." Science **355**(6320): 40-45.

Møens, P. B. and R. E. Pearlman (1988). "Chromatin Organization at Meiosis." BioEssays **9**(5): 151-152.

Muller, H. J. (1916). "The mechanism of crossing-over." Am.Nat. **50**: 193-221ff.

Muller, H. J. (1964). "The Relation of Recombination to Mutational Advance." Mutation Research **1**(1): 2-9.

Munoz-Fuentes, V., A. Di Rienzo and C. Vila (2011). "Prdm9, a major determinant of meiotic recombination hotspots, is not functional in dogs and their wild relatives, wolves and coyotes." PLoS One **6**(11): e25498.

Murray, A. W. and J. W. Szostak (1985). "Chromosome segregation in mitosis and meiosis." Annu Rev Cell Biol **1**: 289-315.

Myers, S., R. Bowden, A. Tumian, R. E. Bontrop, C. Freeman, T. S. MacFie, G. McVean and P. Donnelly (2010). "Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination." Science **327**(5967): 876-879.

Narasimhan, V. M., K. A. Hunt, D. Mason, C. L. Baker, K. J. Karczewski, M. R. Barnes, A. H. Barnett, C. Bates, S. Bellary, N. A. Bockett, K. Giorda, C. J. Griffiths, H. Hemingway, Z. Jia, M. A. Kelly, H. A. Khawaja, M. Lek, S. McCarthy, R. McEachan, A. O'Donnell-Luria, K. Paigen, C. A. Parisinos, E. Sheridan, L. Southgate, L. Tee, M. Thomas, Y. Xue, M. Schnall-Levin, P. M. Petkov, C. Tyler-Smith, E. R. Maher, R. C. Trembath, D. G. MacArthur, J. Wright, R. Durbin and D. A. van Heel (2016). "Health and population effects of rare gene knockouts in adult humans with related parents." Science **352**(6284): 474-477.

Odenthal-Hesse, L., I. L. Berg, A. Veselis, A. J. Jeffreys and C. A. May (2014). "Transmission distortion affecting human noncrossover but not crossover recombination: a hidden source of meiotic drive." PLoS Genet **10**(2): e1004106.

Odenthal-Hesse, L., J. Y. Dutheil, F. Klotzl and B. Haubold (2016). "hotspot: software to support sperm-typing for investigating recombination hotspots." Bioinformatics.

Oliver, P. L., L. Goodstadt, J. J. Bayes, Z. Birtle, K. C. Roach, N. Phadnis, S. A. Beatson, G. Lunter, H. S. Malik and C. P. Ponting (2009). "Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa." PLoS Genet **5**(12): e1000753.

Osman, F., J. Dixon, C. L. Doe and M. C. Whitby (2003). "Generating Crossovers by Resolutionof Nicked Holliday Junctions:A Role for Mus81-Eme1 in Meiosis." Moelcular Cell **12**(3): 761-774.

Otto, S. P. (2009). "The evolutionary enigma of sex." Am Nat **174 Suppl 1**: S1-S14.

Pan, J. and S. Keeney (2009). Detection of SPO11-Oligonucleotide Complexes from Mouse Testes. Meiosis. S. Keeney, Humana Press. **1:** 197-207.

Pan, J., M. Sasaki, R. Kniewel, H. Murakami, H. G. Blitzblau, S. E. Tischfield, X. Zhu, M. J. Neale, M. Jasin, N. D. Socci, A. Hochwagen and S. Keeney (2011). "A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation." Cell **144**(5): 719-731.

Parvanov, E. D., P. M. Petkov and K. Paigen (2010). "Prdm9 Controls Activation of Mammalian Recombination Hotspots." Science **327**(5967): 835-835.

Powers, N. R., E. D. Parvanov, C. L. Baker, M. Walker, P. M. Petkov and K. Paigen (2016). "The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo." PLoS Genet **12**(6): e1006146.

Pratto, F., K. Brick, P. Khil, F. Smagulova, G. V. Petukhova and R. D. Camerini-Otero (2014). "Recombination initiation maps of individual human genomes." Science **346**(6211): 2564421-2564429.

Ross, C. R., D. S. DeFelice, G. J. Hunt, K. E. Ihle, G. V. Amdam and O. Rueppell (2015). "Genomic correlates of recombination rate and its variability across eight recombination maps in the western honey bee (Apis mellifera L.)." BMC Genomics **16**: 107.

Schield, D. R., G. I. M. Pasquesi, B. W. Perry, R. H. Adams, Z. L. Nikolakis, A. K. Westfall, R. W. Orton, J. M. Meik, S. P. Mackessy and T. A. Castoe (2020). "Snake recombination landscapes are concentrated in functional regions despite PRDM9." Mol Biol Evol.

Schwacha, A. and N. Kleckner (1995). "Identification of double holliday junctions as intermediates in meiotic recombination." Cell **83**(5): 783-791.

Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Ranay, C. N. Balakrishnan, S. C. Griffith, G. McVean and M. Przeworski (2015). "Stable recombination hotspots in birds." Science **350**(6263): 928-932.

Smagulova, F., K. Brick, Y. Pu, R. D. Camerini-Otero and G. V. Petukhova (2016). "The evolutionary turnover of recombination hot spots contributes to speciation in mice." Genes Dev **30**(3): 266-280.

Smagulova, F., I. V. Gregoretti, K. Brick, P. Khil, R. D. Camerini-Otero and G. V. Petukhova (2011). "Genome-wide analysis reveals novel molecular features of mouse recombination hotspots." Nature **472**(7343): 375-378.

Stahl, F. W. and H. M. Foss (2010). "A two-pathway analysis of meiotic crossing over and gene conversion in Saccharomyces cerevisiae." Genetics **186**(2): 515-536.

Sturtevant, A. H. (1915). "The behavior of the chromosomes as studied through linkage." Zeitschrift für induktive Abstammungs- und Vererbungslehre volume **13**: 234-287.

Sun, H., D. Treco and J. W. Szostak (1991). "Extensive 3'-Overhanging, Single-Stranded DNA Associated with the Meiosis-Specific Double-Strand Breaks at the ARG4 Recombination Initiation Site." Cell **64**(6): 1155-1161.

Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein and F. W. Stahl (1983). "The double-strand-break repair model for recombination." Cell **33**(1): 25-35.

Wang, W. and E. F. Kirkness (2005). "Short interspersed elements (SINEs) are a major source of canine genomic diversity." Genome Res **15**(12): 1798-1808.

Wayne, R. K., E. Geffen, D. J. Girman, K. P. Koepfli, L. M. Lau and C. R. Marshall (1997). "Molecular systematics of the canidae." Syst Biol **46**(4): 622-653.

Webb, A. J., I. L. Berg and A. Jeffreys (2008). "Sperm cross-over activity in regions of the human genome showing extreme breakdown of marker association." Proc Natl Acad Sci U S A **105**(30): 10471-10476.

Williams, G. C. (1975). "Sex and Evolution." Monogr Popul Biol **8**: 3-200.

Wu, T.-C. and M. Lichten (1994). "Meiosis-Induced Double-Strand Break Sites Determined by Yeast Chromatin Structure." Science **263**(5146): 515-518.

Yoshida, K., G. Kondoh, Y. Matsuda, T. Habu, Y. Nishimune and T. Morita (1998). "The Mouse RecA-like Gene Dmc1 Is Required for Homologous Chromosome Synapsis during Meiosis." Mol Cell **1**(5): 707-718.

Zakharyevich, K., Y. Ma, S. Tang, P. Y. Hwang, S. Boiteux and N. Hunter (2010). "Temporally and biochemically distinct activities of Exo1 during meiosis: double-strand break resection and resolution of double Holliday junctions." Mol Cell **40**(6): 1001-1015.

Zhu, Z., W. H. Chung, E. Y. Shim, S. E. Lee and G. Ira (2008). "Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends." Cell **134**(6): 981-994.

Zickler, D. and N. Kleckner (1999). "Meiotic Chromosomes: Integrating Structure and Function." Annu Rev Genet **33**: 603-754.

Acknowledgements

I would first and foremost like to thank my supervisor and mentor Linda Odenthal-Hesse for giving me the chance to work on this project but more importantly for giving me the freedom to develop it myself. I am grateful for all the trust you put in me, for the way you let me make my own mistakes and learn from them, for all the confidence you gave me and for the support both in professional as well as private matters. I would further like to thank Diethard Tautz, for having me in his department and giving me advice when I needed it and at the same time trusted me to carry out my project independently.

I thank my family, first of all my son, for keeping me grounded and showing me what is most important in life; my soon-to-be husband Jonathan Hartmann, for supporting me in every possible way, celebrating the good times and sticking with me through the less enjoyable periods. My parents and parents-in-law, my sister and siblings-in-law for all the conversations, for babysitting, the interest in my work and for the way you were so proud of me.

I am so thankful for all the technical support I got, most of all from Nicole Thomsen and also Olga Eitel, Malte Dittmann, Christina Nimke, Tjorben Nawroth and Marvin Suhr, who have made sure that I had a great set of data to come back to after my maternity leave.

I would like to thank Ute Barz, who in a way has become my most important colleague in my final year. Thank you for taking such good care of my son. I have never had to worry about him when he was in your care and this above all has made it possible for me to finish this thesis with a clear head.

Thank you to the people in our group, who have become friends: Ela Iwaskiewicz, Khawla Abu Alia, Nicole Thomsen and especially Elena Damm. Thank you for countless lunches, walks, coffees, advice and motivation. I have loved coming to Plön every day because of you.

I am immensely grateful for all the support from the administration, especially Petra Salenz and Katharina Schrader, who have made it so much easier for me to

come back to work after having a baby. I would also like to thank the sequencing-team for the countless sequencing runs and for never losing your patience with me, even if it took me ten attempts to get the submission file right.

I would also like to mention Lutz Becks, thank you for your supervision during my master thesis. It was the work in your group and your supervision that made me want to continue in science and start a PhD in the first place.

Finally, I would like to thank the Christiane-Nüsslein-Volhard-Foundation for supporting my work financially, which has helped me greatly in dealing with all the "ordinary life stuff" that is taking up so much capacity in every parent's mind.

I am grateful to have such a wonderful network of people around me, who have not only made it possible for me to finish my PhD but who have made it a great and worthwhile time.

I hereby declare that:

i. Apart from my supervisor's guidance, the content and design of this dissertation is the product of my own work;

ii. This thesis has not already been submitted either partially or wholly as part of doctoral degree to another examination body, and no other materials are published or submitted for publication than indicated in the thesis;

iii. The preparation of the thesis has been subjected to the Rules of Good Scientific Practice of the German Research Foundation;

iv. no academic degree has ever been withdrawn from me.

Alina Jeschke

Supplementary

S1. Initial panel of dogs analyzed for hotspot flanking markers. Dogs for which ASPs were tested are marked with an X in column "Hotspot 27". Individuals d04 and d18 had Phase C/D, where the crossover assay could not be established.

Identifier	HMW DNA conc (ng/ul)	MDA?	Breed Info	somatic tissue	Hotspot 27
CanFam001	4.62	yes	mix (Dachshund/rhodesian	no	
CanFam002	0.43	no	mix (German Shepherd)	no	
CanFam003	32.00	yes	mix	no	
CanFam004	42.87	yes	Bullmastiff	no	X
CanFam005	116.40	yes	Boxer	no	
CanFam006	567.37	yes	mix	no	X
CanFam007	342.59	yes	Weimaraner	no	X
CanFam008	165.31	yes	Bullterrier	no	
CanFam009	8.09	yes	Bullterrier	no	
CanFam010	-2.03	no	Mix	no	
CanFam011	165.96	yes	Jack Russel Terrier	no	
CanFam012	-2.79	yes	Chihauhau	no	
CanFam013	11.30	yes	Rhodesian Rhidgeback	no	
CanFam014	61.03	yes	Rhodesian Rhidgeback	no	
CanFam015	74.76	yes	American Akita	no	
CanFam016	36.79	yes	German Shepheard	no	
CanFam017	137.39	yes	Boxer	no	
CanFam018	19.85	yes	Bullmastiff	no	X
CanFam019	7.97	yes	German Shepheard	no	
CanFam020	14.85	yes	Mini Australian Shepheard	no	
CanFam021	100.09	yes	Beagle	no	
CanFam022	133.68	yes	mix	no	
CanFam023	68.13	yes	Golden Retriever	no	
CanFam024	23.53	yes	mix	no	
CanFam025	32.61	yes	mix	no	
CanFam026	41.69	yes	mix	no	X
CanFam027	3.00	yes	Chihauhau	tunica albuginea	
CanFam028	2.00	yes	Husky-Akita-Inu	tunica albuginea	

S1 (continued). Initial panel of dogs analyzed for hotspot flanking markers. Dogs for which ASPs were tested are marked with an X in column “Hotspot 27”. Individuals d04 and d18 had Phase C/D, where the crossover assay could not be established.

Identifier	HMW DNA conc (ng/ul)	MDA?	Breed Info	somatic tissue	Hotspot 27
CanFam029	3.00	yes	Zwergspitz	tunica albuginea	
CanFam030	70.00	yes	Malteser	no	X
CanFam031	0.00	yes	no information	tunica albuginea	
CanFam032	0.00	no	no information	tunica albuginea	
CanFam033	12.00	yes	no information	tunica albuginea	
CanFam034	14.00	yes	no information	tunica albuginea	
CanFam035	0.00	yes	no information	tunica albuginea	
CanFam036	2.00	yes	no information	tunica albuginea	
CanFam037	48.00	yes	no information	tunica albuginea	X
CanFam038	not extracted	no	Husky-Australien Shepeard	tunica albuginea	
CanFam039	not extracted	no	Labrador mix	tunica-albuginea/fur	
CanFam040	0.00	yes	Border Collie	no	
CanFam041	70.00	yes	French bulldoge	no	
CanFam042	not extracted	yes	Mini Australian Shepheard	tunica albuginea	
CanFam043	not extracted	no	no information	no	
CanFam044	48.00	yes	no information	no	
CanFam045	no DNA	no	Golden Retriever	no	
CanFam046	no sample	no			
CanFam047	no sample	no			
CanFam048	no sample	no			
CanFam049	no sample	no			
CanFam050	13.00	yes	Chow-Chow	no	
CanFam051	not extracted	no	Australian Shepheard	no	
CanFam052	25.00	yes	Grosspudel	no	X
CanFam053	70.00	yes	mix	no	
CanFam054	45.00	yes	Caucasian Ovcharka	no	
CanFam055	70.00	yes	Labrador	no	
CanFam056	45.00	yes	Bobtail	no	X

S2. DNA extraction protocol

DNA extraction from dog sperm samples

Time (per 4 samples): plan 2 to 2.5 hours in the morning, incubation (2h), 2 to 2.5 hours in the afternoon. Incubation overnight plus ~1h on the next day.
Process one DNA sample at a time in the clean-bench, using designated pipettes. In between DNA extractions decontaminate the clean bench using Ethanol, then H_2O_2 solution or DNA-Away and finally at least 45 minutes of UV.

Equipment and materials
Clean bench
Vortexer
Microcentrifuge
Set of designated pipettes
Low retention tubes (siliconized) → screwtop tubes
Thermo-shaker (rotating heating block)
NanoDrop 3300 Fluorospectrometer

PCR clean buffers:
(Autoclaved and treated to 5050 µJoules x 100 in 254 nm UV crosslinker)
0.2X SSC – adjust pH to 7.0 just prior to use to avoid pellet re-suspension problems
1X SSC/0.2% SDS, adjust pH to 7.0-7.2 before use
10% SDS, pH 7.2
20 µl of 20 mg/ml proteinase K – (40 µg) → from Qiagen kit
3M Sodium acetate pH 5.2
5mM Tris-HCl pH 7.5
TE-Buffer

Reagents (purchased as Ultrapure):
ß-mercaptoethanol
5 ml aliquot of Phenol/Chloroform /Isoamyl alcohol 25:24:1,
saturated with 225 µl 10mM Tris-HCl pH 8.0 in a dark bottle
Ethanol (EtOH),
Absolute (at -20°C)
70% at room temperature

Other Reagents
0.8% Agarose gel
FastRuler High Range Ladder
SybrGreen
SybrSafe stain

S2 (continued). DNA extraction protocol

Work in Category II laminar flow hood for all times tubes are opened:

1. Place 10%SDS and 1%SSC/0.2%SDS bottles in water bath and heat (~55°C).
2. Thaw an aliquot of dog sperm (~500 µl) and transfer the solution to a low-retention microcentrifuge tube
3. Pellet the aliquot in a microcentrifuge at 5,000 rpm for 5 min.
4. Carefully remove the supernatant with a 1 ml pipet tip. (Remove most supernatant, Spin again briefly, draw off residual freezing extender).
5. Re-suspend the pellet in 1 ml of 1X SSC/0.2% SDS, by thorough vortexing for 10 sec. Somatic cells (such as the egg yolk present in the sperm buffer) will lyse, but sperm are resistant to this lower concentration of SDS. Mix briefly by inverting and spin at 5,000 rpm for 5 min.
6. Remove the supernatant with a 1 ml pipet tip and discard. Briefly spin the cells, and remove the residual 1X SSC/0.2% SDS. Repeat steps 5-6 twice.
7. Re-suspend in 1 ml of 0.2X SSC. Pellet the cells at 5,000 rpm for 3 min and remove the supernatant with a 1 ml pipet tip as before.
8. Repeat this step (7) once and then re-suspend in 960 µl of 0.2X SSC.
9. Add 20 µl of freshly prepared 20 mg/ml proteinase K, and 100 µl of 10% SDS. **In the fume hood:** Add 120 µl of β-mercaptoethanol (this reduces disulfide bonds that inter- and intra-molecularly connect protamines, cysteine- and arginine-rich proteins that highly compact sperm DNA, replacing histones.) Invert 10 times to mix and incubate at 37°C for 2 h on a low rotating heat block.

While you wait:

Place a Falcon tube with 5ml of 100% Ethanol in the –20◦C freezer.

Put 10% SDS in waterbath to homogenize.

Label 12 screwtop tubes per sample.

10. Split the resulting slurry into four tubes of ~320 µl each. Add an equivalent volume of phenol:chloroform:isoamyl alcohol (25:24:1) to extract the protein. Mix well and spin at 12,000 rpm (full speed) for 5 min.
11. For the subsequent phenol extractions: Use 1 ml wide tips . Transfer the aqueous layers (top) into clean siliconized tubes. Leave a

significant portion of the aqueous phases at the interphase to prevent protein contamination.

12. Re-extract the phenol layers from Step 10 by adding 200 µl of 1X SSC and 20 µl of 10% SDS to each tube. Mix well by inverting tubes vigorously and spin 12,000 rpm (full speed) for 5 min. Combine the aqueous phases from the re-extractions to the previous samples.
13. Add 500 µl of phenol:chloroform:isoamyl alcohol (25:24:1) to each of the tubes with the combined aqueous phases, mix well and spin at 12,000 rpm (full speed) for 5 min. Remove the aqueous phase to a fresh tube.
14. Precipitate the DNA by adding 1 ml of –20°C 100% ethanol to each tube. Mix well and spin at 12,000 rpm (full speed) for 10 min.
15. Decant supernatant and pellet DNA with a twist at full speed for 5 minutes, discard last of the supernatant.
16. Wash the pellet by adding 1 ml of 70% ethanol. Repeat the centrifugation step and decant the supernatant. Briefly spin and remove the last of the supernatant with a pipet tip.
17. Immediately re-suspend the pellet in 75 or 150 µl TE-buffer, depending on the size of the pellet (pipette around the sides of the tubes). Incubate at 55°C for 1 h on a slow rotating heating block. Transfer to 4°C overnight.

18. Pool the four aliquots together in a freezer safe tube (screwtop).
19. Add 30 µl 3 M Na acetate (pH 5.2) and 900 µl of –20°C 100% ethanol. Mix well and spin as in Step 14. Decant supernatant and wash the pellet with 70% ethanol (as in steps 16+17). Air-dry the pellet in the PCR laminar flow hood, but do not overdry the pellet as this will hamper re-suspension.
20. Re-suspend in 100 µl of 5 mM Tris pH 7.5. Incubate at 55°C for 1 h on a slow rotating heating block. Transfer to 4°C overnight.
21. Quick spin, take aliquot and quantify with a nanodrop spectrometer.
22. Run dilution series of (.01, .05 and 0.1 µg) on a 0.8% agarose gel at 120 V for 2 hours and compare with high quality DNA of defined concentration.
23. Can be stored indefinitely at -20°C

Protocol amended from: (Cole and Jasin 2011)

Cole, F. and M. Jasin (2011). "Isolation of meiotic recombinants from mouse sperm." Methods Mol Biol **745**: 251-282.

S2 (continued). DNA extraction protocol

Re-precipitation

- add 1/10th of volume Sodium actetate (pH 5.2, 3 M) and 3 volumes ice cold absolute ethanol
- spin 15 minutes at max speed (14,000 rpm)
- add 3 volumes 70% ethanol
- spin 10 minutes at max speed
- decant ethanol and fully dry pellet (ca. 45-90 min)
- elute in 5 mM Tris buffer (100 µl or 300 µl) over night at 4°C
- store at -20°C

S3. List of primers

Name	**usage**	**Sequence**
Clf27A_30.0_UF	Re-sequencing PCR	GAGCACCTTCTACCCACACC
Clf27A_30.2_UF	Sanger Seq	TGCAGGACTGTGCTAGAAGC
Clf27A_32.0B_CT_F1	Outer Forward A/B	AGAGTCCTGTTGGGTGGT
Clf27A_32.0B_TG_F1	Outer Forward A/B	AGAGTCTTGTTGGGTGGG
Clf27A_32.1_A_F2	Inner Forward A/B	ACCACCACCAATTTTACA
Clf27A_32.1_G_F2	Inner Forward A/B	ACCACCACCAATTTTACG
Clf27A_32.2_R	Re-sequencing PCR	ATTTGGCCTTTCGGCACAAC
Clf27A_32.2UF	Sanger Seq	CAGAGGTTGTGCCGAAAG
Clf27A_34.9_UF	Sanger Seq	GATGATGAGCTAGAGGGGGC
Clf27A_34.9UR	Primer testing	CTGCCCCCTCTAGCTCAT
Clf27A_38.7_UR	Sanger Seq	CCTCCAGCTCTCCACCTTTT
Clf27A_40.6_UF	Sanger Seq	GAGCATTAGAATGATCTTCCC
Clf27A_41.0UF	Primer testing and Re-sequencing PCR	TGCCCCAGATAGATGCAC
Clf27A_42.2_UF	Sanger Seq	TGGAACGGTCTATAGGAAGGTC
Clf27A_43.2A4R	Inner Reverse A/B	ACACTGCATGCCTCTTT
Clf27A_43.2G3R	Inner Reverse A/B	CACTGCATGCCTCTTC
Clf27A_43.4A4R	Outer Reverse A/B	GGTCTGCTACAGCTGTT
Clf27A_43.4G3R	Outer Reverse A/B	GTCTGCTACAGCTGTC
Clf_27A_59_F	Sanger Seq	AGGGAAGATACCTGCCTCAT
Clf_27A_88_F	Sanger Seq	GCCTCCTACTCAGTGTTTG
Clf_27A_119_F	Sanger Seq	TCTAGCAGCAGTGATTGGAG
Clf_27A_137_F	Sanger Seq	ACCTGGAGGTCATGCTTATC
Clf_27A_143_F	Sanger Seq	aggtagctcagatactgcct
Clf27A_43.4_UR	Re-sequencing PCR	TCAACCCAGTGCTACTTCACT

S4. Code for bioinformatics analyses (Basecalling, demultiplexing and de-novo assembly).

Basecalling

```
guppy_basecaller --input path/to/runfiles/Run#/fast5 --
save_path /output/path/Run#/basecall --flowcell FLO-MIN106 --
kit SQK-LSK109 --qscore_filtering --num_callers 16 --
cpu_threads_per_caller 2
```

Demultiplexing

```
guppy_barcoder -t 12 --input /path/to/file/Run#/basecall/pass
--save_path /output/path/Run#/adapters --config
configuration.cfg
```

Read filtering

```
for file  in …/Run#/adapters/barcode#/*.fastq; do
cat $file >> …/Run#/adapters/barcode#.fastq; done
```

```
awk -F"\t" 'BEGIN {FS = "\t" ; OFS = "\n"} {header = $0 ;
getline seq ; getline qheader ; getline qseq ; if (length(seq)
>= 4000 && length(seq) <= 12000) {print header, seq, qheader,
qseq}}' < barcode#.fastq > filtered-barcode#.fastq
```

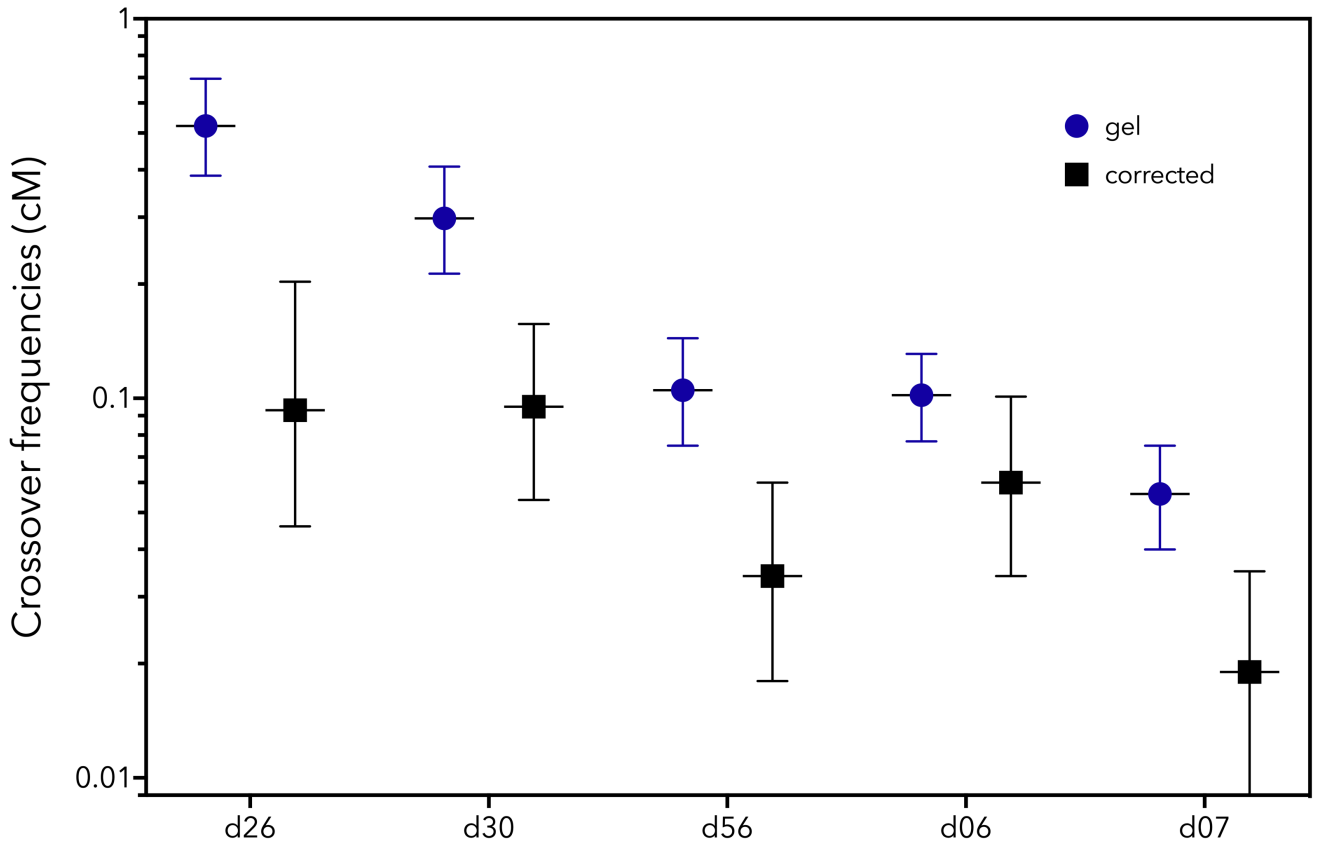
***De-novo* assembly**

```
…/canu -p Run#-Ad# -d …/Run#/adapters/Run#-# genomeSize=10k
useGrid=false maxThreads=12 maxMemory=256g -nanopore-raw
…/Run#/adapters/filtered-barcode#.fastq
```

S5. Haplotypes per individual across hotspot region. Haplotypes were obtained by sequencing phasing PCR products using MinION-long read sequencing and validated by Sanger-sequencing.

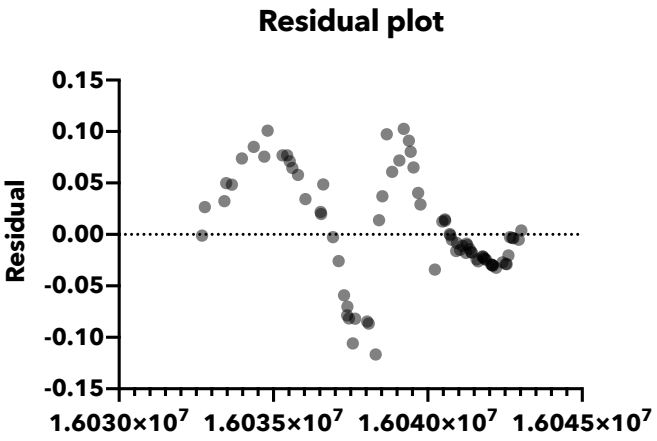
					d06		d07		d26		d30		d37		d52		d56	
from ASP	chr27:160x	Int	Reference	Variation	A	B	A	B	A	B	A	B	A	B	A	B	A	B
3,992	32,687	1	C	A	A	C	A	C	A	C	A	C	A	C	A	C	A	C
4,080	32,775	2	TTT	DEL	8 T	10 T	8 T	10 T	8 T	10 T	8 T	10 T	8 T	10 T	8 T	10 T	7T	10T
4,715	33,410	3	C	T	T	C	T	C	T	C	T	C	T	C	T	C	T	C
4,773	33,468	4	A	G	A	A	A	A	G	A	A	A	A	A	A	A	A	A
4,958	33,653	5	G	C	C	G	C	G	C	G	C	G	C	G	C	G	C	G
5,284	33,979	6	DEL	IN(24BP)	IN	DEL	IN	DEL	IN	DEL	IN	DEL	IN	DEL	IN	DEL	IN	DEL
5,671	34,366	7	G	T	G	G	T	G	T	G	G	G	G	G	G	G	G	G
6,006	34,701	8	T	C	C	T	C	T	C	T	C	T	C	T	C	T	C	T
6,115	34,810	9	23A	DEL	16 A	16 A	16A	20A	16 A	16 A	16 A	16A	16A	16A	16A	16A	16A	16A
6,604	35,299	10	G	T	T	T	T	G	T	T	T	T	T	T	T	T	T	T
6,754	35,449	11	G	T	T	G	T	G	T	G	T	G	T	G	T	G	T	G
6,833	35,528	12	G	A	G	A	G	G	G	A	G	A	G	A	G	A	G	A
6,920	35,615	13	TT	DEL	DEL	TT	DEL	TT	DEL	TT	DEL	TT	DEL	TT	DEL	TT	DEL	TT
7,104	35,799	14	T	C	C	T	T	T	T	T	C	T	C	T	C	T	C	T
7,341	36,036	15	T	C	C	T	T	T	T	T	C	T	C	T	C	T	C	T
7,834	36,529	16	G	A	G	A	G	G	G	A	G	A	G	A	G	A	G	A
7,849	36,544	17	T	C	T	C	T	T	T	C	T	C	T	C	T	C	T	C
7,919	36,614	18	C	A	C	C	A	C	C	C	C	C	C	C	C	C	C	C
8,238	36,933	19	DEL	AG	AG	DEL	AG	DEL	DEL	DEL	AG	DEL	AG	DEL	AG	DEL	AG	DEL
8,416	37,111	20	G	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A
8,592	37,287	21	T	A	T	T	A	T	T	T	T	T	T	T	T	T	T	T
8,692	37,387	22	G	T	G	T	G	G	G	G	G	G	T	G	T	T	T	T
8,697	37,392	23	C	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C
8,734	37,449	24	30 T	DEL	27 T	23 T	21 T	19 T	18 T	20 T	20 T	18 T	23 T	20 T	23 T	24 T	23T	23T
8,873	37,568	25	A	T	A	T	A	A	A	A	A	A	T	A	T	T	T	T
8,948	37,643	26	C	A	C	A	A	C	A	C	C	A	C	C	A	A	C	A
9,341	38,036	27	T	C	C	C	T	T	C	C	C	T	C	C	C	C	C	C
9,394	38,089	28	A	C	A	A	C	A	A	A	A	A	C	A	C	A	C	A
9,617	38,312	29	C	T	T	T	T	C	C	C	T	C	T	C	T	T	T	C
9,727	38,422	30	G	A	G	A	A	G	A	G	A	G	A	G	A	A	A	G
9,831	38,526	31	T	G	T	G	T	T	G	G	G	T	G	G	G	G	G	G
9,976	38,671	32	A	C	A	A	C	A	A	A	A	A	A	A	A	A	A	A
10,150	38,845	33	T	G	G	G	G	G	G	G	G	T	T	G	T	G	T	G
10,385	39,080	34	G	A	A	A	G	A	G	A	G	A	G	A	G	A	G	A
10,524	39,219	35	C	T	T	T	C	T	T	T	T	T	C	T	C	T	C	T
10,692	39,387	36	A	G	A	A	A	A	G	A	A	A	A	A	A	A	A	A
10,753	39,448	37	C	G	C	C	C	C	G	C	C	C	C	C	C	C	C	C
10,842	39,537	38	6T	3T	6T	6T	6T	6T	3T	6T	6T	6T	6T	6T	6T	6T	6T	6T
10,994	39,689	39	T	C	T	T	T	T	C	T	T	T	T	T	T	T	T	T
11,068	39,763	40	A	G	G	G	G	G	G	G	G	G	A	G	A	G	A	G
11,536	40,231	41	T	G	T	G	T	G	G	G	G	G	T	G	T	G	T	G
11,779	40,474	42	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G
11,858	40,553	43	A	G	A	G	A	G	A	G	A	A	A	G	A	G	A	G
11,870	40,565	44	TA	DEL	TA	DEL	TA	DEL	TA	DEL	TA	DEL	TA	DEL	TA	DEL	TA	DEL
12,017	40,712	45	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
12,030	40,725	46	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C
12,084	40,779	47	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
12,226	40,921	48	9C	6C	6C	6C	6C	6C	6C	6C	6C	6C	6C	6C	6C	6C	6C	6C
12,254	40,949	49	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G
12,349	41,044	50	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A
12,425	41,120	51	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C
12,541	41,236	52	C	T	C	T	C	T	T	T	T	T	C	T	C	T	C	T
12,556	41,251	53	DEL	TCTA	DEL	TCTA	DEL	TCTA	DEL	TCTA	DEL	TCTA	DEL	TCTA	DEL	TCTA	DEL	TCTA
12,583	41,278	54	IN (*)	DEL	IN(*)	DEL	IN(*)	DEL	IN(*)	DEL	IN(*)	DEL	IN	DEL	IN(*)	DEL	IN	DEL
12,658	41,353	55	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A
12,717	41,412	56	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A
12,729	41,424	57	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T
12,883	41,578	58	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A
12,938	41,633	59	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A
13,081	41,776	60	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
13,085	41,780	61	11T	9T	9T	11T	11T	11T	11T	11T	11T	11T	11T	11T	11T	11T	9T	10T
13,128	41,823	62	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A
13,163	41,858	63	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A
13,190	41,885	64	1T	9T	1T	9T	1T	9T	1T	9T	1T	9T	1T	9T	1T	9T	1T	6T
13,358	42,053	65	12G+IN	DEL	6G+IN	6G+DEL	6G+IN	6G+DEL	6G+A+IN	6G+DEL	6G+A+IN	6G+DEL	6G+IN	6G+DEL	6G+IN	6G+DEL	6G+IN	6G+DEL
13,389	42,084	66	DEL	G	DEL	DEL	DEL	DEL	DEL	G	DEL	DEL	DEL	DEL	DEL	G	A	A
13,390	42,085	67	TT	AAC	TT	AAC	TT	AAC	TT	AAC	TT	AAC	TT	AAC	TT	AAC	TT	AAC
13,417	42,112	68	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
13,514	42,209	69	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
13,729	42,424	70	G	A	A	A	G	A	A	A	A	A	G	A	G	A	G	A
13,832	42,527	71	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
13,855	42,550	72	G	A	A	A	G	A	G	A	G	A	G	A	G	A	G	A
13,917	42,612	73	C	T	T	C	C	C	C	C	C	C	C	C	C	C	C	C
13,970	42,665	74	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T
14,044	42,739	75	C	T	C	C	C	C	T	C	T	C	C	C	C	C	C	C
13,072	41,767	76	C	T	T	T	C	T	T	T	T	T	C	T	C	T	C	T
14,248	42,943	77	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C
14,337	43,032	78	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C
14,571	43,266		ASP R															
14,713	43,408																	
15,654	44,349																	
15,693	44,388																	
15,718	44,413																	
16,787	45,482																	
16,933	45,628																	
16,969	45,664																	

S6. Individual crossover frequencies obtained from gel analysis (blue dots), compared to MinION-sequencing validated positive crossovers (black squared). Error bars represent 95%-confidence intervals



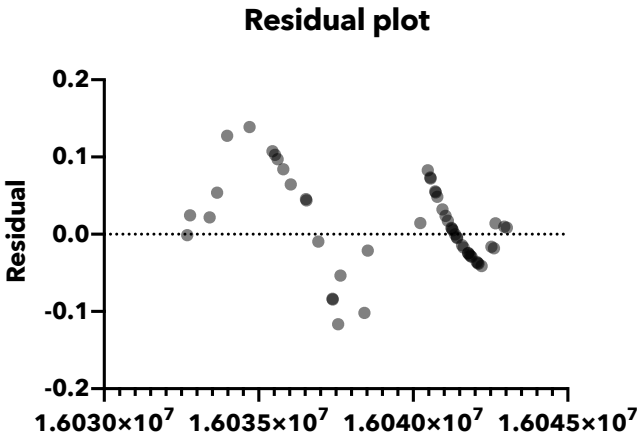
S7. Regression analysis (combined all samples). Table represents statistics inferred by Prism8 software, the plot depicts residuals-vs.-x plots.

	sum
Cumulative Gaussian -- Fraction	
Best-fit values	
Mean	16038206
SD	1816
95% CI (profile likelihood)	
Mean	16038104 to 16038306
SD	1653 to 1982
Goodness of Fit	
Degrees of Freedom	76
R squared	0.9805
Sum of Squares	0.1993
Sy.x	0.05121
Constraints	
SD	SD > 0
Number of points	
# of X values	78
# Y values analyzed	78



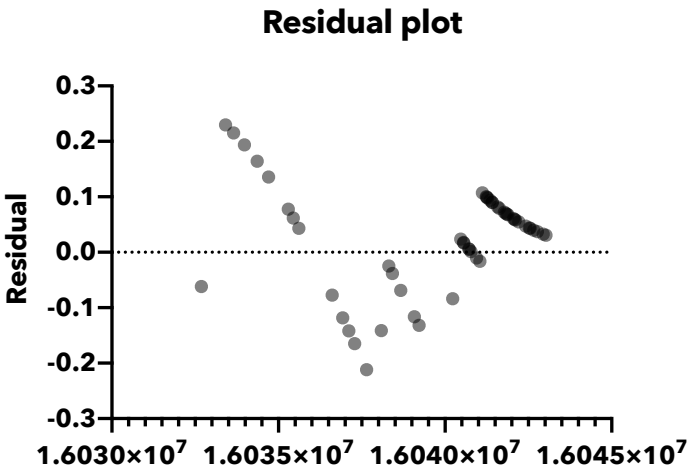
S8. Regression analysis (d06). Table represents statistics inferred by Prism8 software, the plot depicts residuals-vs.-x plots.

	combined
Cumulative Gaussian -- Fraction	
Best-fit values	
Mean	16038503
SD	1901
95% CI (profile likelihood)	
Mean	16038326 to 16038680
SD	1695 to 2111
Goodness of Fit	
Degrees of Freedom	50
R squared	0.9775
Sum of Squares	0.1668
Sy.x	0.05775
Constraints	
SD	SD > 0
Number of points	
# of X values	52
# Y values analyzed	52



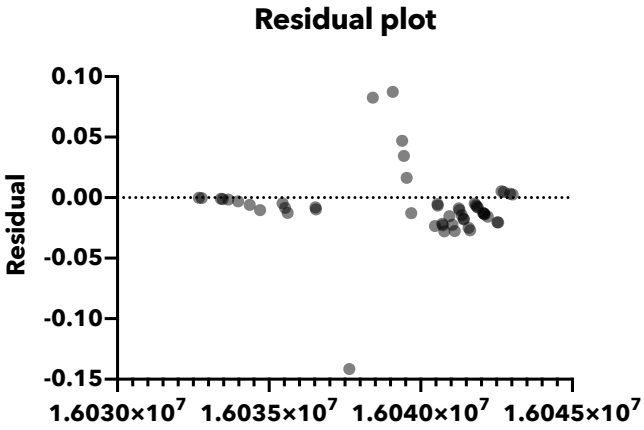
S9. Regression analysis (d07). Table represents statistics inferred by Prism8 software, the plot depicts residuals-vs.-x plots.

	combined
Cumulative Gaussian -- Fraction	
Best-fit values	
Mean	16037358
SD	3029
95% CI (profile likelihood)	
Mean	16036922 to 16037768
SD	2571 to 3534
Goodness of Fit	
Degrees of Freedom	51
R squared	0.8944
Sum of Squares	0.4985
Sy.x	0.09886
Constraints	
SD	SD > 0
Number of points	
# of X values	53
# Y values analyzed	53



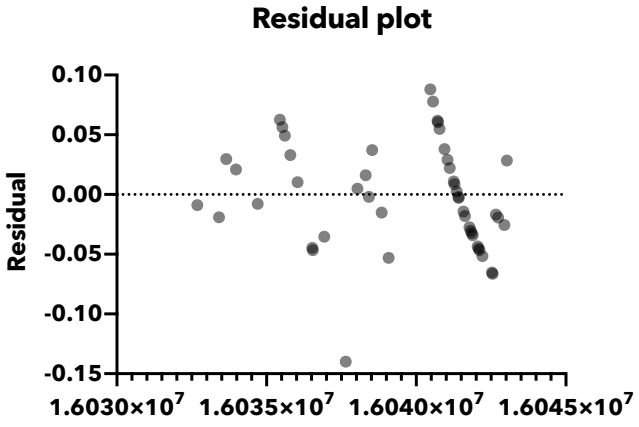
S10. Regression analysis (d26). Table represents statistics inferred by Prism8 software, the plot depicts residuals-vs.-x plots.

	combined
Cumulative Gaussian -- Fraction	
Best-fit values	
Mean	16038483
SD	1635
95% CI (profile likelihood)	
Mean	16038381 to 16038582
SD	1517 to 1754
Goodness of Fit	
Degrees of Freedom	49
R squared	0.9943
Sum of Squares	0.04714
Sy.x	0.03102
Constraints	
SD	SD > 0
Number of points	
# of X values	51
# Y values analyzed	51



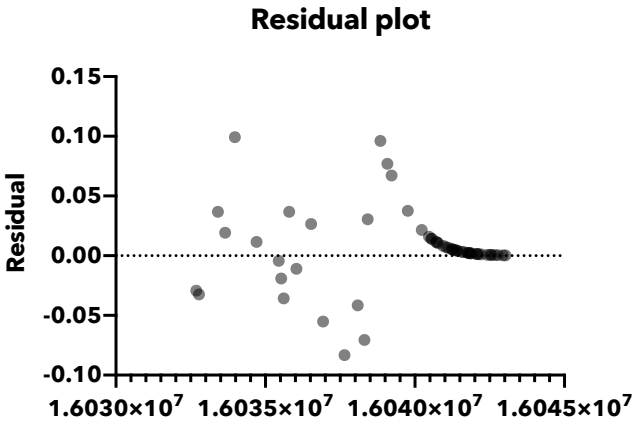
S11. Regression analysis (d30). Table represents statistics inferred by Prism8 software, the plot depicts residuals-vs.-x plots.

	combined
Cumulative Gaussian -- Fraction	
Best-fit values	
Mean	16038425
SD	2418
95% CI (profile likelihood)	
Mean	16038282 to 16038566
SD	2244 to 2598
Goodness of Fit	
Degrees of Freedom	47
R squared	0.9838
Sum of Squares	0.09344
Sy.x	0.04459
Constraints	
SD	SD > 0
Number of points	
# of X values	49
# Y values analyzed	49



S12. Regression analysis (d56). Table represents statistics inferred by Prism8 software, the plot depicts residuals-vs.-x plots.

	combined
Cumulative Gaussian -- Fraction	
Best-fit values	
Mean	16036333
SD	1927
95% CI (profile likelihood)	
Mean	16036229 to 16036437
SD	1791 to 2067
Goodness of Fit	
Degrees of Freedom	53
R squared	0.9896
Sum of Squares	0.05801
Sy.x	0.03308
Constraints	
SD	SD > 0
Number of points	
# of X values	55
# Y values analyzed	55



S13. Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

	seqnames	source	feature	start	end	absScore	relScore	strand	ID	TF	class	siteSeqs
2	SNPupstream1C	TFBS	TFBS	6	12	6.199879	0.9114242	+	MA0042.2	FOXI1	Fork head / winged helix factors	GTGAACA
3	SNPupstream1C	TFBS	TFBS	6	12	5.979245	0.9324868	+	MA0033.2	FOXL1	Fork head / winged helix factors	GTGAACA
4	SNPupstream1C	TFBS	TFBS	6	13	7.623073	0.9114842	+	MA0157.2	FOXO3	Fork head / winged helix factors	GTGAACAT
9	SNPupstream1C	TFBS	TFBS	11	20	8.034653	0.9261313	-	MA0707.1	MNX1	Homeo domain factors	GCTAATAATG
10	SNPupstream1C	TFBS	TFBS	11	20	5.420905	0.9114544	-	MA0710.1	NOTO	Homeo domain factors	GCTAATAATG
11	SNPupstream1C	TFBS	TFBS	9	16	4.303342	0.9101038	+	MA0132.2	PDX1	Homeo domain factors	AACATTAT
14	SNPupstream1C	TFBS	TFBS	3	13	9.695565	0.9004773	+	MA0845.1	FOXB1	Fork head / winged helix factors	AATGTGAACAT
16	SNPupstream1C	TFBS	TFBS	6	12	3.411538	0.9060533	+	MA0849.1	FOXO6	Fork head / winged helix factors	GTGAACA
17	SNPupstream1C	TFBS	TFBS	9	16	5.928750	0.9268935	+	MA0875.1	BARX1	Homeo domain factors	AACATTAT
19	SNPupstream1C	TFBS	TFBS	9	16	4.745514	0.9064299	+	MA0876.1	BSX	Homeo domain factors	AACATTAT
20	SNPupstream1C	TFBS	TFBS	11	20	6.002568	0.9101469	-	MA0877.1	Barhl1	Homeo domain factors	GCTAATAATG
21	SNPupstream1C	TFBS	TFBS	11	20	6.881417	0.9318304	+	MA0886.1	EMX2	Homeo domain factors	CATTATTAGC
22	SNPupstream1C	TFBS	TFBS	11	20	7.690088	0.9229463	+	MA0887.1	EVX1	Homeo domain factors	CATTATTAGC
23	SNPupstream1C	TFBS	TFBS	11	20	7.543310	0.9256725	+	MA0888.1	EVX2	Homeo domain factors	CATTATTAGC
24	SNPupstream1C	TFBS	TFBS	11	20	6.915413	0.9096295	+	MA0892.1	GSX1	Homeo domain factors	CATTATTAGC
25	SNPupstream1C	TFBS	TFBS	11	20	6.042584	0.9004688	+	MA0900.1	HOXA2	Homeo domain factors	CATTATTAGC
26	SNPupstream1C	TFBS	TFBS	8	17	5.281033	0.9041953	+	MA0903.1	HOXB3	Homeo domain factors	GAACATTATT
27	SNPupstream1C	TFBS	TFBS	11	20	6.476499	0.9226395	+	MA0903.1	HOXB3	Homeo domain factors	CATTATTAGC

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

	seqnames	source	feature	start	end	absScore	relScore	strand	ID	TF	class	siteSeqs
1	SNPupstream1T	TFBS	TFBS	10	16	6.801457	0.9085944	-	MA0063.1	Nkx2-5	Homeo domain factors	ATAATAT
3	SNPupstream1T	TFBS	TFBS	10	20	8.885573	0.9125470	+	MA0601.1	Arid3b	ARID domain factors	ATATTATTAGC
4	SNPupstream1T	TFBS	TFBS	6	12	4.162951	0.9132815	+	MA0033.2	FOXL1	Fork head / winged helix factors	GTGAATA
9	SNPupstream1T	TFBS	TFBS	11	20	9.309310	0.9451462	-	MA0707.1	MNX1	Homeo domain factors	GCTAATAATA
10	SNPupstream1T	TFBS	TFBS	11	20	4.991191	0.9065551	-	MA0710.1	NOTO	Homeo domain factors	GCTAATAATA
11	SNPupstream1T	TFBS	TFBS	12	19	5.183630	0.9205843	+	MA0132.2	PDX1	Homeo domain factors	ATTATTAG
12	SNPupstream1T	TFBS	TFBS	9	20	8.434206	0.9019948	-	MA0052.3	MEF2A	MADS box factors	GCTAATAATATT
14	SNPupstream1T	TFBS	TFBS	3	13	10.722472	0.9133417	+	MA0845.1	FOXB1	Fork head / winged helix factors	AATGTGAATAT
17	SNPupstream1T	TFBS	TFBS	11	20	5.723804	0.9053822	-	MA0877.1	Barhl1	Homeo domain factors	GCTAATAATA
18	SNPupstream1T	TFBS	TFBS	11	19	7.652903	0.9012105	-	MA0878.1	CDX1	Homeo domain factors	CTAATAATA
19	SNPupstream1T	TFBS	TFBS	11	20	6.286404	0.9258167	+	MA0886.1	EMX2	Homeo domain factors	TATTATTAGC
20	SNPupstream1T	TFBS	TFBS	11	20	6.889824	0.9094898	+	MA0887.1	EVX1	Homeo domain factors	TATTATTAGC
21	SNPupstream1T	TFBS	TFBS	11	20	6.813044	0.9143956	+	MA0888.1	EVX2	Homeo domain factors	TATTATTAGC
22	SNPupstream1T	TFBS	TFBS	11	20	6.584303	0.9243027	+	MA0903.1	HOXB3	Homeo domain factors	TATTATTAGC

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

	seqnames	source	feature	start	end	absScore	relScore	strand	ID	TF	class	siteSeqs
1	SNPupstream2C	TFBS	TFBS	2	13	7.003959	0.8679500	+	MA0041.1	Foxd3	Fork head / winged helix factors	AGTTGTTTGCTT
2	SNPupstream2C	TFBS	TFBS	12	20	8.345006	0.9051205	-	MA0084.1	SRY	High-mobility group (HMG) domain factors	TTTTACAAA
3	SNPupstream2C	TFBS	TFBS	5	16	13.389973	0.9473507	+	MA0047.2	Foxa2	Fork head / winged helix factors	TGTTTGCTTTGT
4	SNPupstream2C	TFBS	TFBS	14	20	8.001974	0.8904714	-	MA0152.1	NFATC2	Rel homology region (RHR) factors	TTTTACA
5	SNPupstream2C	TFBS	TFBS	8	21	2.402744	0.8600447	-	MA0523.1	TCF7L2	High-mobility group (HMG) domain factors	GTTTTACAAAGCAA
6	SNPupstream2C	TFBS	TFBS	1	15	15.187672	0.9769421	+	MA0148.3	FOXA1	Fork head / winged helix factors	TAGTTGTTTGCTTTG
7	SNPupstream2C	TFBS	TFBS	3	13	3.352037	0.8611555	-	MA0593.1	FOXP2	Fork head / winged helix factors	AAGCAAACAAC
8	SNPupstream2C	TFBS	TFBS	4	11	4.581801	0.8707457	-	MA0613.1	FOXG1	Fork head / winged helix factors	GCAAACAA
9	SNPupstream2C	TFBS	TFBS	4	11	8.607966	0.9401418	-	MA0614.1	Foxj2	Fork head / winged helix factors	GCAAACAA
10	SNPupstream2C	TFBS	TFBS	5	11	6.894339	0.9200633	-	MA0042.2	FOXI1	Fork head / winged helix factors	GCAAACA
11	SNPupstream2C	TFBS	TFBS	1	7	0.814988	0.8778806	-	MA0033.2	FOXL1	Fork head / winged helix factors	ACAACTA
12	SNPupstream2C	TFBS	TFBS	5	11	8.842435	0.9627618	-	MA0033.2	FOXL1	Fork head / winged helix factors	GCAAACA
13	SNPupstream2C	TFBS	TFBS	4	11	3.482348	0.8608392	-	MA0157.2	FOXO3	Fork head / winged helix factors	GCAAACAA
14	SNPupstream2C	TFBS	TFBS	12	21	2.918885	0.8684199	+	MA0466.2	CEBPB	Basic leucine zipper factors (bZIP)	TTTGTAAAAC
15	SNPupstream2C	TFBS	TFBS	12	21	2.115208	0.8619573	+	MA0836.1	CEBPD	Basic leucine zipper factors (bZIP)	TTTGTAAAAC
16	SNPupstream2C	TFBS	TFBS	12	21	4.394396	0.8618829	+	MA0837.1	CEBPE	Basic leucine zipper factors (bZIP)	TTTGTAAAAC
17	SNPupstream2C	TFBS	TFBS	4	14	8.360484	0.8837525	-	MA0845.1	FOXB1	Fork head / winged helix factors	AAAGCAAACAA
18	SNPupstream2C	TFBS	TFBS	4	14	10.725144	0.9268870	-				

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
MA0032.2  FOXC1         Fork head / winged helix factors
AAAGCAAACAA
19 SNPupstream2C   TFBS    TFBS     3  14 12.499471 0.9315696      -
MA0846.1  FOXC2         Fork head / winged helix factors
AAAGCAAACAAC
20 SNPupstream2C   TFBS    TFBS     2   8  1.762082 0.8703907      -
MA0847.1  FOXD2         Fork head / winged helix factors
AACAACT
21 SNPupstream2C   TFBS    TFBS     5  11  9.513351 0.9705395      -
MA0847.1  FOXD2         Fork head / winged helix factors
GCAAACA
22 SNPupstream2C   TFBS    TFBS     5  11  6.525744 0.9080208      -
MA0848.1  FOXO4         Fork head / winged helix factors
GCAAACA
23 SNPupstream2C   TFBS    TFBS     5  11 -1.758387 0.8571429      -
MA0849.1  FOXO6         Fork head / winged helix factors
GCAAACA
24 SNPupstream2C   TFBS    TFBS     1   7  2.573295 0.8689965      -
MA0850.1  FOXP3         Fork head / winged helix factors
ACAACTA
25 SNPupstream2C   TFBS    TFBS     5  11  9.686235 0.9710642      -
MA0850.1  FOXP3         Fork head / winged helix factors
GCAAACA
26 SNPupstream2C   TFBS    TFBS    11  20  5.125400 0.8548346      +
MA0895.1 HMBOX1          Homeo domain factors
CTTTGTAAAA
27 SNPupstream2C   TFBS    TFBS    11  21  6.963717 0.8679373      +
MA0907.1 HOXC13          Homeo domain factors
CTTTGTAAAAC
28 SNPupstream2C   TFBS    TFBS    12  21  3.753105 0.8630769      +
MA0909.1 HOXD13          Homeo domain factors
TTTGTAAAAC
29 SNPupstream2C   TFBS    TFBS     2  13  9.246677 0.8759914      -
MA0481.2  FOXP1         Fork head / winged helix factors
AAGCAAACAACT
30 SNPupstream2C   TFBS    TFBS     9  19 10.811773 0.9337272      -
MA0442.2  SOX10 High-mobility group (HMG) domain factors
TTTACAAAGCA
31 SNPupstream2C   TFBS    TFBS    10  19  7.501956 0.8861792      +
MA1152.1  SOX15 High-mobility group (HMG) domain factors
GCTTTGTAAA
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

	seqnames	source	feature	start	end	absScore	relScore	strand	ID	TF	class	siteSeqs
2	SNPupstream2C	TFBS	TFBS	5	16	13.389973	0.9473507	+	MA0047.2	Foxa2	Fork head / winged helix factors	TGTTTGCTTTGT
3	SNPupstream2C	TFBS	TFBS	1	15	15.187672	0.9769421	+	MA0148.3	FOXA1	Fork head / winged helix factors	TAGTTGTTTGCTTTG
4	SNPupstream2C	TFBS	TFBS	4	11	8.607966	0.9401418	-	MA0614.1	Foxj2	Fork head / winged helix factors	GCAAACAA
5	SNPupstream2C	TFBS	TFBS	5	11	6.894339	0.9200633	-	MA0042.2	FOXI1	Fork head / winged helix factors	GCAAACA
6	SNPupstream2C	TFBS	TFBS	5	11	8.842435	0.9627618	-	MA0033.2	FOXL1	Fork head / winged helix factors	GCAAACA
7	SNPupstream2C	TFBS	TFBS	4	14	10.725144	0.9268870	-	MA0032.2	FOXC1	Fork head / winged helix factors	AAAGCAAACAA
8	SNPupstream2C	TFBS	TFBS	3	14	12.499471	0.9315696	-	MA0846.1	FOXC2	Fork head / winged helix factors	AAAGCAAACAAC
9	SNPupstream2C	TFBS	TFBS	5	11	9.513351	0.9705395	-	MA0847.1	FOXD2	Fork head / winged helix factors	GCAAACA
10	SNPupstream2C	TFBS	TFBS	5	11	6.525744	0.9080208	-	MA0848.1	FOXO4	Fork head / winged helix factors	GCAAACA
11	SNPupstream2C	TFBS	TFBS	5	11	9.686235	0.9710642	-	MA0850.1	FOXP3	Fork head / winged helix factors	GCAAACA
12	SNPupstream2C	TFBS	TFBS	9	19	10.811773	0.9337272	-	MA0442.2	SOX10	High-mobility group (HMG) domain factors	TTTACAAAGCA

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

	seqnames	source	feature	start	end	absScore	relScore	strand	ID	TF	class	siteSeqs
1	SNPupstream2T	TFBS	TFBS	2	12	10.5312196	0.8902438	+	MA0040.1	Foxq1	Fork head / winged helix factors	AGTTGTTTGTT
2	SNPupstream2T	TFBS	TFBS	2	13	9.2518860	0.8953416	+	MA0041.1	Foxd3	Fork head / winged helix factors	AGTTGTTTGTTT
3	SNPupstream2T	TFBS	TFBS	7	15	9.6920310	0.9340298	-	MA0084.1	SRY	High-mobility group (HMG) domain factors	CAAAACAAA
4	SNPupstream2T	TFBS	TFBS	12	20	8.3450060	0.9051205	-	MA0084.1	SRY	High-mobility group (HMG) domain factors	TTTTACAAA
5	SNPupstream2T	TFBS	TFBS	5	16	10.2311457	0.9085172	+	MA0047.2	Foxa2	Fork head / winged helix factors	TGTTTGTTTTGT
6	SNPupstream2T	TFBS	TFBS	14	20	8.0019740	0.8904714	-	MA0152.1	NFATC2	Rel homology region (RHR) factors	TTTTACA
7	SNPupstream2T	TFBS	TFBS	5	14	3.2247862	0.8600872	+	MA0515.1	Sox6	High-mobility group (HMG) domain factors	TGTTTGTTTT
8	SNPupstream2T	TFBS	TFBS	5	14	0.5432618	0.8653781	+	MA0514.1	Sox3	High-mobility group (HMG) domain factors	TGTTTGTTTT
9	SNPupstream2T	TFBS	TFBS	9	19	-4.7042584	0.8546955	-	MA0102.3	CEBPA	Basic leucine zipper factors (bZIP)	TTTACAAAACA
10	SNPupstream2T	TFBS	TFBS	1	15	11.3644108	0.9522460	+	MA0148.3	FOXA1	Fork head / winged helix factors	TAGTTGTTTGTTTTG
11	SNPupstream2T	TFBS	TFBS	4	11	3.8459152	0.8609385	-	MA0613.1	FOXG1	Fork head / winged helix factors	ACAAACAA
12	SNPupstream2T	TFBS	TFBS	4	11	8.3852163	0.9375760	-	MA0614.1	Foxj2	Fork head / winged helix factors	ACAAACAA
13	SNPupstream2T	TFBS	TFBS	7	16	2.8521495	0.8830776	-	MA0635.1	BARHL2	Homeo domain factors	ACAAAACAAA
14	SNPupstream2T	TFBS	TFBS	5	11	3.8288704	0.8819285	-	MA0042.2	FOXI1	Fork head / winged helix factors	ACAAACA
15	SNPupstream2T	TFBS	TFBS	1	7	0.8149880	0.8778806	-	MA0033.2	FOXL1	Fork head / winged helix factors	ACAACTA
16	SNPupstream2T	TFBS	TFBS	5	11	8.4930715	0.9590676	-	MA0033.2	FOXL1	Fork head / winged helix factors	ACAAACA
17	SNPupstream2T	TFBS	TFBS	9	20	4.0177810	0.8639307	+	MA0043.2	HLF	Basic leucine zipper factors (bZIP)	TGTTTTGTAAAA

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
18 SNPupstream2T   TFBS    TFBS     9  20  6.7181587 0.8943173
- MA0043.2    HLF      Basic leucine zipper factors (bZIP)
TTTTACAAAACA
19 SNPupstream2T   TFBS    TFBS    12  21  2.9188846 0.8684199
+ MA0466.2  CEBPB      Basic leucine zipper factors (bZIP)
TTTGTAAAAC
20 SNPupstream2T   TFBS    TFBS    10  19  6.1659529 0.9012068
- MA0466.2  CEBPB      Basic leucine zipper factors (bZIP)
TTTACAAAAC
21 SNPupstream2T   TFBS    TFBS    12  21  2.1152082 0.8619573
+ MA0836.1  CEBPD      Basic leucine zipper factors (bZIP)
TTTGTAAAAC
22 SNPupstream2T   TFBS    TFBS    10  19  5.5295248 0.8951202
- MA0836.1  CEBPD      Basic leucine zipper factors (bZIP)
TTTACAAAAC
23 SNPupstream2T   TFBS    TFBS    12  21  4.3943964 0.8618829
+ MA0837.1  CEBPE      Basic leucine zipper factors (bZIP)
TTTGTAAAAC
24 SNPupstream2T   TFBS    TFBS    10  19  6.9409181 0.8934304
- MA0837.1  CEBPE      Basic leucine zipper factors (bZIP)
TTTACAAAAC
25 SNPupstream2T   TFBS    TFBS     4  14  5.7993638 0.8516687
- MA0845.1  FOXB1         Fork head / winged helix factors
AAAACAAACAA
26 SNPupstream2T   TFBS    TFBS     4  14  9.2367253 0.9080203
- MA0032.2  FOXC1         Fork head / winged helix factors
AAAACAAACAA
27 SNPupstream2T   TFBS    TFBS     3  14 11.3703540 0.9149176
- MA0846.1  FOXC2         Fork head / winged helix factors
AAAACAAACAAC
28 SNPupstream2T   TFBS    TFBS     2   8  1.7620820 0.8703907
- MA0847.1  FOXD2         Fork head / winged helix factors
AACAACT
29 SNPupstream2T   TFBS    TFBS     5  11  8.4540375 0.9568528
- MA0847.1  FOXD2         Fork head / winged helix factors
ACAAACA
30 SNPupstream2T   TFBS    TFBS     9  15  4.3894210 0.9043367
- MA0847.1  FOXD2         Fork head / winged helix factors
CAAAACA
31 SNPupstream2T   TFBS    TFBS     5  11  4.8377778 0.8832991
- MA0848.1  FOXO4         Fork head / winged helix factors
ACAAACA
32 SNPupstream2T   TFBS    TFBS     1   7  2.5732951 0.8689965
- MA0850.1  FOXP3         Fork head / winged helix factors
ACAACTA
33 SNPupstream2T   TFBS    TFBS     5  11  8.7994459 0.9583392
- MA0850.1  FOXP3         Fork head / winged helix factors
ACAAACA
34 SNPupstream2T   TFBS    TFBS    11  21  7.0651866 0.8691869
+ MA0907.1 HOXC13                    Homeo domain factors
TTTTGTAAAAC
35 SNPupstream2T   TFBS    TFBS    12  21  3.7531046 0.8630769
+ MA0909.1 HOXD13                    Homeo domain factors
TTTGTAAAAC
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
36 SNPupstream2T   TFBS    TFBS     4  14  7.2069950 0.8682272
- MA0442.2  SOX10 High-mobility group (HMG) domain factors
AAAACAAACAA
37 SNPupstream2T   TFBS    TFBS     7  18  5.5617251 0.8637327
- MA1125.1 ZNF384                 C2H2 zinc finger factors
TTACAAAACAAA
38 SNPupstream2T   TFBS    TFBS     5  14  7.5905238 0.8878052
+ MA1152.1  SOX15 High-mobility group (HMG) domain factors
TGTTTGTTTT
39 SNPupstream2T   TFBS    TFBS    10  19  7.7839981 0.8913569
+ MA1152.1  SOX15 High-mobility group (HMG) domain factors
GTTTTGTAAA
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
1  SNPupstream2T   TFBS    TFBS     7  15  9.692031 0.9340298      -
MA0084.1   SRY High-mobility group (HMG) domain factors
CAAAACAAA
3  SNPupstream2T   TFBS    TFBS     5  16 10.231146 0.9085172      +
MA0047.2 Foxa2         Fork head / winged helix factors
TGTTTGTTTTGT
4  SNPupstream2T   TFBS    TFBS     1  15 11.364411 0.9522460      +
MA0148.3 FOXA1         Fork head / winged helix factors
TAGTTGTTTGTTTTG
5  SNPupstream2T   TFBS    TFBS     4  11  8.385216 0.9375760      -
MA0614.1 Foxj2         Fork head / winged helix factors
ACAAACAA
6  SNPupstream2T   TFBS    TFBS     5  11  8.493072 0.9590676      -
MA0033.2 FOXL1         Fork head / winged helix factors
ACAAACA
7  SNPupstream2T   TFBS    TFBS    10  19  6.165953 0.9012068      -
MA0466.2 CEBPB      Basic leucine zipper factors (bZIP)
TTTACAAAAC
8  SNPupstream2T   TFBS    TFBS     4  14  9.236725 0.9080203      -
MA0032.2 FOXC1         Fork head / winged helix factors
AAAACAAACAA
9  SNPupstream2T   TFBS    TFBS     3  14 11.370354 0.9149176      -
MA0846.1 FOXC2         Fork head / winged helix factors
AAAACAAACAAC
10 SNPupstream2T   TFBS    TFBS     5  11  8.454037 0.9568528      -
MA0847.1 FOXD2         Fork head / winged helix factors
ACAAACA
11 SNPupstream2T   TFBS    TFBS     9  15  4.389421 0.9043367      -
MA0847.1 FOXD2         Fork head / winged helix factors
CAAAACA
12 SNPupstream2T   TFBS    TFBS     5  11  8.799446 0.9583392      -
MA0850.1 FOXP3         Fork head / winged helix factors
ACAAACA
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

	seqnames	source	feature	start	end	absScore	relScore	strand	ID	TF	class	siteSeqs
1	SNPcenterG	TFBS	TFBS	1	7	5.9108420	0.8843503	+	MA0081.1	SPIB	Tryptophan cluster factors	TAGGGAA
2	SNPcenterG	TFBS	TFBS	11	20	5.6190021	0.9082201	+	MA0109.1	HLTF	Tryptophan cluster factors	GGACATTTAG
3	SNPcenterG	TFBS	TFBS	3	9	7.2282853	0.8711784	-	MA0152.1	NFATC2	Rel homology region (RHR) factors	ATTTCCC
4	SNPcenterG	TFBS	TFBS	1	10	8.2169874	0.9227841	-	MA0606.1	NFAT5	Rel homology region (RHR) factors	TATTTCCCTA
5	SNPcenterG	TFBS	TFBS	1	10	7.3347522	0.8689360	-	MA0624.1	NFATC1	Rel homology region (RHR) factors	TATTTCCCTA
6	SNPcenterG	TFBS	TFBS	1	10	7.2931324	0.8640039	-	MA0625.1	NFATC3	Rel homology region (RHR) factors	TATTTCCCTA
7	SNPcenterG	TFBS	TFBS	12	21	9.2179187	0.9594117	-	MA0635.1	BARHL2	Homeo domain factors	ACTAAATGTC
8	SNPcenterG	TFBS	TFBS	4	10	3.4448867	0.8771517	+	MA0042.2	FOXI1	Fork head / winged helix factors	GGAAATA
9	SNPcenterG	TFBS	TFBS	13	20	2.9160074	0.8794234	+	MA0654.1	ISX	Homeo domain factors	ACATTTAG
10	SNPcenterG	TFBS	TFBS	12	21	2.5914979	0.8560545	-	MA0669.1	NEUROG2	Basic helix-loop-helix factors (bHLH)	ACTAAATGTC
11	SNPcenterG	TFBS	TFBS	3	11	-0.3588502	0.8546796	-	MA0673.1	NKX2-8	Homeo domain factors	CTATTTCCC
12	SNPcenterG	TFBS	TFBS	13	20	4.1439521	0.8840883	+	MA0704.1	Lhx4	Homeo domain factors	ACATTTAG
13	SNPcenterG	TFBS	TFBS	11	20	1.7283023	0.8693543	+	MA0710.1	NOTO	Homeo domain factors	GGACATTTAG
14	SNPcenterG	TFBS	TFBS	11	20	0.3311182	0.8534248	-	MA0710.1	NOTO	Homeo domain factors	CTAAATGTCC
15	SNPcenterG	TFBS	TFBS	12	21	4.0743552	0.8522554	+	MA0718.1	RAX	Homeo domain factors	GACATTTAGT
16	SNPcenterG	TFBS	TFBS	12	20	1.1080435	0.8681647	+	MA0724.1	VENTX	Homeo domain factors	GACATTTAG
17	SNPcenterG	TFBS	TFBS	10	16	4.7120213	0.9302700	+	MA0498.2	MEIS1	Homeo domain factors	AGGACAT
18	SNPcenterG	TFBS	TFBS	1	11	5.3854083	0.8592022	+	MA0032.2	FOXC1	Fork head / winged helix factors	TAGGGAAATAG
19	SNPcenterG	TFBS	TFBS	14	20	2.7326506	0.8829308	-	MA0847.1	FOXD2	Fork head / winged helix factors	CTAAATG
20	SNPcenterG	TFBS	TFBS	4	10	4.8691780	0.8837590	+	MA0848.1	FOXO4	Fork head / winged helix factors	GGAAATA
21	SNPcenterG	TFBS	TFBS	4	10	1.3836697	0.8868685	+	MA0849.1	FOXO6	Fork head / winged helix factors	GGAAATA
22	SNPcenterG	TFBS	TFBS	13	20	2.3287053	0.8704122	+	MA0876.1	BSX	Homeo domain factors	ACATTTAG
23	SNPcenterG	TFBS	TFBS	4	13	3.7725970	0.8720316	+	MA0877.1	Barhl1	Homeo domain factors	GGAAATAGGA
24	SNPcenterG	TFBS	TFBS	12	21	8.3688602	0.9505923	-	MA0877.1	Barhl1	Homeo domain factors	ACTAAATGTC
25	SNPcenterG	TFBS	TFBS	13	20	4.2207139	0.8677986	+	MA0882.1	DLX6	Homeo domain factors	ACATTTAG
26	SNPcenterG	TFBS	TFBS	5	12	3.5884996	0.8557950	-	MA0882.1	DLX6	Homeo domain factors	CCTATTTC

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
27 SNPcenterG   TFBS    TFBS    13  20  4.0847629 0.8624876      +
MA0885.1    Dlx2                   Homeo domain factors     ACATTTAG
28 SNPcenterG   TFBS    TFBS     3  12  5.5098562 0.8760777      -
MA0891.1    GSC2                   Homeo domain factors   CCTATTTCCC
29 SNPcenterG   TFBS    TFBS    11  20  3.1629898 0.8715173      +
MA0903.1   HOXB3                   Homeo domain factors   GGACATTTAG
30 SNPcenterG   TFBS    TFBS    12  21  3.3210061 0.8739553      +
MA0903.1   HOXB3                   Homeo domain factors   GACATTTAGT
31 SNPcenterG   TFBS    TFBS     9  19  8.8643746 0.8647662      -
MA1110.1   NR1H4 Nuclear receptors with C4 zinc fingers  TAAATGTCCTA
32 SNPcenterG   TFBS    TFBS    10  19  5.4390374 0.8566363      -
MA1112.1   NR4A1 Nuclear receptors with C4 zinc fingers   TAAATGTCCT
33 SNPcenterG   TFBS    TFBS     6  16  9.4746721 0.8874584      +
MA1150.1    RORB Nuclear receptors with C4 zinc fingers  AAATAGGACAT
34 SNPcenterG   TFBS    TFBS     4  15  9.9282326 0.9043176      +
MA1151.1    RORC Nuclear receptors with C4 zinc fingers GGAAATAGGACA
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

	seqnames	source	feature	start	end	absScore	relScore	strand	ID	TF	class	siteSeqs
1	SNPcenterG	TFBS	TFBS	11	20	5.619002	0.9082201	+	MA0109.1	HLTF	Tryptophan cluster factors	GGACATTTAG
4	SNPcenterG	TFBS	TFBS	10	16	4.712021	0.9302700	+	MA0498.2	MEIS1	Homeo domain factors	AGGACAT
6	SNPcenterG	TFBS	TFBS	4	15	9.928233	0.9043176	+	MA1151.1	RORC	Nuclear receptors with C4 zinc fingers	GGAAATAGGACA

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
1  SNPcenterT   TFBS    TFBS    11  18  4.7587254 0.8618823      +
MA0067.1         Pax2                     Paired box factors
TGACATTT
2  SNPcenterT   TFBS    TFBS     8  15  5.5791547 0.8838277      -
MA0067.1         Pax2                     Paired box factors
TGTCATAT
3  SNPcenterT   TFBS    TFBS     1   7  5.9108420 0.8843503      +
MA0081.1         SPIB             Tryptophan cluster factors
TAGGGAA
4  SNPcenterT   TFBS    TFBS     9  14  8.6864048 0.9770921      +
MA0089.1 MAFG::NFE2L1    Basic leucine zipper factors (bZIP)
TATGAC
5  SNPcenterT   TFBS    TFBS    11  20  5.7326589 0.9107424      +
MA0109.1         HLTF             Tryptophan cluster factors
TGACATTTAG
6  SNPcenterT   TFBS    TFBS     3   9  7.2282853 0.8711784      -
MA0152.1       NFATC2      Rel homology region (RHR) factors
ATTTCCC
7  SNPcenterT   TFBS    TFBS     2  16  6.2532464 0.8680125      -
MA0258.2         ESR2 Nuclear receptors with C4 zinc fingers
ATGTCATATTTCCCT
8  SNPcenterT   TFBS    TFBS     5  15 -0.1556113 0.8630717      +
MA0093.2         USF1  Basic helix-loop-helix factors (bHLH)
GAAATATGACA
9  SNPcenterT   TFBS    TFBS    10  17  3.5752552 0.8761767      +
MA0604.1         Atf1    Basic leucine zipper factors (bZIP)
ATGACATT
10 SNPcenterT   TFBS    TFBS     9  16  8.3589207 0.9266418      +
MA0605.1         Atf3    Basic leucine zipper factors (bZIP)
TATGACAT
11 SNPcenterT   TFBS    TFBS     1  10  8.2169874 0.9227841      -
MA0606.1        NFAT5      Rel homology region (RHR) factors
TATTTCCCTA
12 SNPcenterT   TFBS    TFBS     5  14  5.4861568 0.8539820      -
MA0623.1      Neurog1  Basic helix-loop-helix factors (bHLH)
GTCATATTTC
13 SNPcenterT   TFBS    TFBS     1  10  7.3347522 0.8689360      -
MA0624.1       NFATC1      Rel homology region (RHR) factors
TATTTCCCTA
14 SNPcenterT   TFBS    TFBS     1  10  7.2931324 0.8640039      -
MA0625.1       NFATC3      Rel homology region (RHR) factors
TATTTCCCTA
15 SNPcenterT   TFBS    TFBS    12  21  9.2179187 0.9594117      -
MA0635.1       BARHL2                   Homeo domain factors
ACTAAATGTC
16 SNPcenterT   TFBS    TFBS    10  19  7.1798829 0.8974080      -
MA0643.1        Esrrg Nuclear receptors with C4 zinc fingers
TAAATGTCAT
17 SNPcenterT   TFBS    TFBS     4  10  3.4448867 0.8771517      +
MA0042.2        FOXI1       Fork head / winged helix factors
GGAAATA
18 SNPcenterT   TFBS    TFBS    13  20  2.9160074 0.8794234      +
MA0654.1          ISX                    Homeo domain factors
ACATTTAG
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
19 SNPcenterT   TFBS    TFBS    12  21  2.5914979 0.8560545     -
MA0669.1      NEUROG2  Basic helix-loop-helix factors (bHLH)
ACTAAATGTC
20 SNPcenterT   TFBS    TFBS    10  18  5.3332832 0.8689673     -
MA0676.1       Nr2e1 Nuclear receptors with C4 zinc fingers
AAATGTCAT
21 SNPcenterT   TFBS    TFBS    13  20  4.1439521 0.8840883     +
MA0704.1        Lhx4                   Homeo domain factors
ACATTTAG
22 SNPcenterT   TFBS    TFBS    11  20  0.8545695 0.8593927     +
MA0710.1        NOTO                   Homeo domain factors
TGACATTTAG
23 SNPcenterT   TFBS    TFBS    12  21  4.0743552 0.8522554     +
MA0718.1          RAX                  Homeo domain factors
GACATTTAGT
24 SNPcenterT   TFBS    TFBS    12  20  1.1080435 0.8681647     +
MA0724.1       VENTX                   Homeo domain factors
GACATTTAG
25 SNPcenterT   TFBS    TFBS    10  16  9.3153945 0.9869590     +
MA0498.2       MEIS1                   Homeo domain factors
ATGACAT
26 SNPcenterT   TFBS    TFBS    10  17  4.5310106 0.8614151     +
MA0774.1       MEIS2                   Homeo domain factors
ATGACATT
27 SNPcenterT   TFBS    TFBS    10  17  8.0900520 0.9263910     +
MA0775.1       MEIS3                   Homeo domain factors
ATGACATT
28 SNPcenterT   TFBS    TFBS     5  14  5.9021424 0.8796796     +
MA0461.2       Atoh1  Basic helix-loop-helix factors (bHLH)
GAAATATGAC
29 SNPcenterT   TFBS    TFBS     5  14  4.3310084 0.8587230     -
MA0461.2       Atoh1  Basic helix-loop-helix factors (bHLH)
GTCATATTTC
30 SNPcenterT   TFBS    TFBS     5  14  3.7075025 0.8612276     +
MA0827.1       OLIG3  Basic helix-loop-helix factors (bHLH)
GAAATATGAC
31 SNPcenterT   TFBS    TFBS     1  11  8.1532177 0.8811560     +
MA0845.1       FOXB1       Fork head / winged helix factors
TAGGGAAATAT
32 SNPcenterT   TFBS    TFBS     1  11  8.7976416 0.9024546     +
MA0032.2       FOXC1       Fork head / winged helix factors
TAGGGAAATAT
33 SNPcenterT   TFBS    TFBS    14  20  2.7326506 0.8829308     -
MA0847.1       FOXD2       Fork head / winged helix factors
CTAAATG
34 SNPcenterT   TFBS    TFBS     4  10  4.8691780 0.8837590     +
MA0848.1       FOXO4       Fork head / winged helix factors
GGAAATA
35 SNPcenterT   TFBS    TFBS     4  10  1.3836697 0.8868685     +
MA0849.1       FOXO6       Fork head / winged helix factors
GGAAATA
36 SNPcenterT   TFBS    TFBS    13  20  2.3287053 0.8704122     +
MA0876.1          BSX                  Homeo domain factors
ACATTTAG
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
37 SNPcenterT   TFBS    TFBS    12  21  8.3688602 0.9505923     -
MA0877.1       Barhl1                   Homeo domain factors
ACTAAATGTC
38 SNPcenterT   TFBS    TFBS    13  20  4.2207139 0.8677986     +
MA0882.1        DLX6                    Homeo domain factors
ACATTTAG
39 SNPcenterT   TFBS    TFBS    13  20  4.0847629 0.8624876     +
MA0885.1        Dlx2                    Homeo domain factors
ACATTTAG
40 SNPcenterT   TFBS    TFBS     3  12  4.2735290 0.8550839     -
MA0891.1        GSC2                    Homeo domain factors
CATATTTCCC
41 SNPcenterT   TFBS    TFBS    11  20  2.9518615 0.8682599     +
MA0903.1       HOXB3                    Homeo domain factors
TGACATTTAG
42 SNPcenterT   TFBS    TFBS    12  21  3.3210061 0.8739553     +
MA0903.1       HOXB3                    Homeo domain factors
GACATTTAGT
43 SNPcenterT   TFBS    TFBS     5  14  3.2550152 0.8563563     -
MA0905.1       HOXC10                   Homeo domain factors
GTCATATTTC
44 SNPcenterT   TFBS    TFBS     9  19  9.2125634 0.8815318     -
MA1111.1       NR2F2 Nuclear receptors with C4 zinc fingers
TAAATGTCATA
45 SNPcenterT   TFBS    TFBS    10  19  9.3742130 0.9140343     -
MA1112.1       NR4A1 Nuclear receptors with C4 zinc fingers
TAAATGTCAT
46 SNPcenterT   TFBS    TFBS     1  11  8.5272510 0.8510041     -
MA1117.1        RELB      Rel homology region (RHR) factors
ATATTTCCCTA
47 SNPcenterT   TFBS    TFBS     5  15  9.1856683 0.8544290     +
MA1118.1        SIX1                    Homeo domain factors
GAAATATGACA
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

	seqnames	source	feature	start	end	absScore	relScore	strand	ID	TF	class	siteSeqs
1	SNPcenterT	TFBS	TFBS	9	14	8.686405	0.9770921	+	MA0089.1	MAFG::NFE2L1	Basic leucine zipper factors (bZIP)	TATGAC
2	SNPcenterT	TFBS	TFBS	11	20	5.732659	0.9107424	+	MA0109.1	HLTF	Tryptophan cluster factors	TGACATTTAG
3	SNPcenterT	TFBS	TFBS	9	16	8.358921	0.9266418	+	MA0605.1	Atf3	Basic leucine zipper factors (bZIP)	TATGACAT
6	SNPcenterT	TFBS	TFBS	10	16	9.315395	0.9869590	+	MA0498.2	MEIS1	Homeo domain factors	ATGACAT
7	SNPcenterT	TFBS	TFBS	10	17	8.090052	0.9263910	+	MA0775.1	MEIS3	Homeo domain factors	ATGACATT
8	SNPcenterT	TFBS	TFBS	1	11	8.797642	0.9024546	+	MA0032.2	FOXC1	Fork head / winged helix factors	TAGGGAAATAT
10	SNPcenterT	TFBS	TFBS	10	19	9.374213	0.9140343	-	MA1112.1	NR4A1	Nuclear receptors with C4 zinc fingers	TAAATGTCAT

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
        seqnames source feature start end   absScore  relScore
strand       ID     TF                                  class
siteSeqs
1  SNPdownstream1G   TFBS    TFBS     2  11  9.4318248 0.9161284
- MA0038.1   Gfi1              C2H2 zinc finger factors  CGAATCACAA
2  SNPdownstream1G   TFBS    TFBS    10  19  7.0064785 0.9390115
- MA0109.1   HLTF            Tryptophan cluster factors  AAACTTGGCG
3  SNPdownstream1G   TFBS    TFBS     3  13 -7.1009601 0.8510757
+ MA0491.1   JUND   Basic leucine zipper factors (bZIP) TGTGATTCGCC
4  SNPdownstream1G   TFBS    TFBS     9  17  4.7300082 0.8623536
+ MA0597.1  THAP1    C2CH THAP-type zinc finger factors   TCGCCAAGT
5  SNPdownstream1G   TFBS    TFBS     3  12  3.3886576 0.8656042
- MA0642.1    EN2                  Homeo domain factors  GCGAATCACA
6  SNPdownstream1G   TFBS    TFBS     3  12  1.0064547 0.8582022
- MA0658.1   LHX6                  Homeo domain factors  GCGAATCACA
7  SNPdownstream1G   TFBS    TFBS     8  16  8.3936130 0.9558531
+ MA0671.1   NFIX  SMAD/NF-1 DNA-binding domain factors   TTCGCCAAG
8  SNPdownstream1G   TFBS    TFBS    11  20  1.6264739 0.8563105
- MA0672.1 NKX2-3                  Homeo domain factors  AAAACTTGGC
9  SNPdownstream1G   TFBS    TFBS    11  19  0.6543582 0.8658960
- MA0673.1 NKX2-8                  Homeo domain factors   AAACTTGGC
10 SNPdownstream1G   TFBS    TFBS     9  17  1.7189299 0.8513962
+ MA0738.1   HIC2              C2H2 zinc finger factors   TCGCCAAGT
11 SNPdownstream1G   TFBS    TFBS    11  19  3.8860726 0.8699724
+ MA0745.1  SNAI2              C2H2 zinc finger factors   GCCAAGTTT
12 SNPdownstream1G   TFBS    TFBS     5  12  4.1739951 0.8756955
+ MA0766.1  GATA5     Other C4 zinc finger-type factors    TGATTCGC
13 SNPdownstream1G   TFBS    TFBS     9  15  1.0526573 0.8852061
+ MA0498.2  MEIS1                  Homeo domain factors     TCGCCAA
14 SNPdownstream1G   TFBS    TFBS     2   8  3.5559289 0.9160331
- MA0498.2  MEIS1                  Homeo domain factors     ATCACAA
15 SNPdownstream1G   TFBS    TFBS     9  15 -0.3255371 0.8682341
- MA0498.2  MEIS1                  Homeo domain factors     TTGGCGA
16 SNPdownstream1G   TFBS    TFBS    15  21  4.4925171 0.9056688
- MA0847.1  FOXD2      Fork head / winged helix factors     AAAAACT
17 SNPdownstream1G   TFBS    TFBS     3  12  4.1985494 0.8696500
- MA0900.1  HOXA2                  Homeo domain factors  GCGAATCACA
18 SNPdownstream1G   TFBS    TFBS     3  12  1.8424635 0.8511437
- MA0903.1  HOXB3                  Homeo domain factors  GCGAATCACA
19 SNPdownstream1G   TFBS    TFBS     8  18  9.1159023 0.8816916
- MA0161.2   NFIC  SMAD/NF-1 DNA-binding domain factors AACTTGGCGAA
20 SNPdownstream1G   TFBS    TFBS     3  13  7.6632877 0.8546925
- MA0060.3   NFYA                        Other alpha GGCGAATCACA
21 SNPdownstream1G   TFBS    TFBS     1   8  8.8032129 0.9202858
+ MA0831.2   TFE3 Basic helix-loop-helix factors (bHLH)    CTTGTGAT
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
  seqnames source feature start end absScore  relScore strand
ID    TF                                  class   siteSeqs
1 SNPdownstream1G   TFBS    TFBS     2  11 9.431825 0.9161284      -
MA0038.1  Gfi1              C2H2 zinc finger factors CGAATCACAA
2 SNPdownstream1G   TFBS    TFBS    10  19 7.006478 0.9390115      -
MA0109.1  HLTF            Tryptophan cluster factors AAACTTGGCG
3 SNPdownstream1G   TFBS    TFBS     8  16 8.393613 0.9558531      +
MA0671.1  NFIX  SMAD/NF-1 DNA-binding domain factors  TTCGCCAAG
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

	seqnames	source	feature	start	end	absScore	relScore	strand	ID	TF	class	siteSeqs
1	SNPdownstream1T	TFBS	TFBS	2	11	8.297175	0.8928164	-	MA0038.1	Gfi1	C2H2 zinc finger factors	AGAATCACAA
2	SNPdownstream1T	TFBS	TFBS	10	19	7.104659	0.9411904	-	MA0109.1	HLTF	Tryptophan cluster factors	AAACTTGGAG
3	SNPdownstream1T	TFBS	TFBS	8	14	8.060391	0.8919281	+	MA0152.1	NFATC2	Rel homology region (RHR) factors	TTCTCCA
4	SNPdownstream1T	TFBS	TFBS	3	13	-7.100960	0.8510757	+	MA0491.1	JUND	Basic leucine zipper factors (bZIP)	TGTGATTCTCC
5	SNPdownstream1T	TFBS	TFBS	8	16	4.699542	0.8892788	+	MA0671.1	NFIX	SMAD/NF-1 DNA-binding domain factors	TTCTCCAAG
6	SNPdownstream1T	TFBS	TFBS	11	20	1.829414	0.8586132	-	MA0672.1	NKX2-3	Homeo domain factors	AAAACTTGGA
7	SNPdownstream1T	TFBS	TFBS	7	15	3.524546	0.8976697	+	MA0673.1	NKX2-8	Homeo domain factors	ATTCTCCAA
8	SNPdownstream1T	TFBS	TFBS	11	19	1.355871	0.8736620	-	MA0673.1	NKX2-8	Homeo domain factors	AAACTTGGA
9	SNPdownstream1T	TFBS	TFBS	11	19	3.188505	0.8597160	+	MA0745.1	SNAI2	C2H2 zinc finger factors	TCCAAGTTT
10	SNPdownstream1T	TFBS	TFBS	2	8	3.555929	0.9160331	-	MA0498.2	MEIS1	Homeo domain factors	ATCACAA
11	SNPdownstream1T	TFBS	TFBS	15	21	4.492517	0.9056688	-	MA0847.1	FOXD2	Fork head / winged helix factors	AAAAACT
12	SNPdownstream1T	TFBS	TFBS	3	12	3.090097	0.8511248	-	MA0900.1	HOXA2	Homeo domain factors	GAGAATCACA
13	SNPdownstream1T	TFBS	TFBS	1	8	8.803213	0.9202858	+	MA0831.2	TFE3	Basic helix-loop-helix factors (bHLH)	CTTGTGAT

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
          seqnames source feature start end absScore  relScore strand
ID    TF                                    class   siteSeqs
1 SNPdownstream1T   TFBS    TFBS    10  19 7.104659 0.9411904      -
MA0109.1  HLTF            Tryptophan cluster factors AAACTTGGAG
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
seqnames source feature start end   absScore  relScore strand
ID     TF                                  class        siteSeqs
1  SNPdownstream2C   TFBS    TFBS     1  15  7.3002853 0.8720509
+ MA0108.2    TBP                 TATA-binding proteins
AAATATAAACCCTGT
2  SNPdownstream2C   TFBS    TFBS     6  20  9.3054439 0.9056326
- MA0484.1  HNF4G Nuclear receptors with C4 zinc fingers
AAAGTACAGGGTTTA
3  SNPdownstream2C   TFBS    TFBS     1  15 -0.3483109 0.8765883
- MA0148.3  FOXA1       Fork head / winged helix factors
ACAGGGTTTATATTT
4  SNPdownstream2C   TFBS    TFBS     5  11  1.5259880 0.8532803
+ MA0042.2  FOXI1       Fork head / winged helix factors
ATAAACC
5  SNPdownstream2C   TFBS    TFBS     5  11  5.5521593 0.9279708
+ MA0033.2  FOXL1       Fork head / winged helix factors
ATAAACC
6  SNPdownstream2C   TFBS    TFBS     5  12  5.7520749 0.8768641
+ MA0682.1  Pitx1                   Homeo domain factors
ATAAACCC
7  SNPdownstream2C   TFBS    TFBS     5  12  4.2671897 0.9011044
+ MA0711.1   OTX1                   Homeo domain factors
ATAAACCC
8  SNPdownstream2C   TFBS    TFBS     5  12  4.7574877 0.8852428
+ MA0712.1   OTX2                   Homeo domain factors
ATAAACCC
9  SNPdownstream2C   TFBS    TFBS     4  12  6.3652281 0.8751235
+ MA0714.1  PITX3                   Homeo domain factors
TATAAACCC
10 SNPdownstream2C   TFBS    TFBS     1   7  0.6104832 0.8555117
+ MA0847.1  FOXD2       Fork head / winged helix factors
AAATATA
11 SNPdownstream2C   TFBS    TFBS     4  10  0.5500784 0.8547313
- MA0847.1  FOXD2       Fork head / winged helix factors
GTTTATA
12 SNPdownstream2C   TFBS    TFBS     4  10  1.8272052 0.8582904
- MA0850.1  FOXP3       Fork head / winged helix factors
GTTTATA
13 SNPdownstream2C   TFBS    TFBS    15  21  2.8981875 0.8736586
- MA0850.1  FOXP3       Fork head / winged helix factors
AAAAGTA
14 SNPdownstream2C   TFBS    TFBS     4  13  2.8171528 0.8557009
+ MA0877.1 Barhl1                   Homeo domain factors
TATAAACCCT
15 SNPdownstream2C   TFBS    TFBS     6  13  8.3715331 0.9038473
- MA0693.2    VDR Nuclear receptors with C4 zinc fingers
AGGGTTTA
16 SNPdownstream2C   TFBS    TFBS    13  20  6.5603957 0.8561533
- MA0693.2    VDR Nuclear receptors with C4 zinc fingers
AAAGTACA
```

S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
  seqnames source feature start end absScore  relScore strand
ID    TF                                    class        siteSeqs
1 SNPdownstream2C   TFBS    TFBS     6  20 9.305444 0.9056326      -
MA0484.1 HNF4G Nuclear receptors with C4 zinc fingers
AAAGTACAGGGTTTA
2 SNPdownstream2C   TFBS    TFBS     5  11 5.552159 0.9279708      +
MA0033.2 FOXL1       Fork head / winged helix factors
ATAAACC
3 SNPdownstream2C   TFBS    TFBS     5  12 4.267190 0.9011044      +
MA0711.1  OTX1                  Homeo domain factors
ATAAACCC
4 SNPdownstream2C   TFBS    TFBS     6  13 8.371533 0.9038473      -
MA0693.2   VDR Nuclear receptors with C4 zinc fingers
AGGGTTTA
```

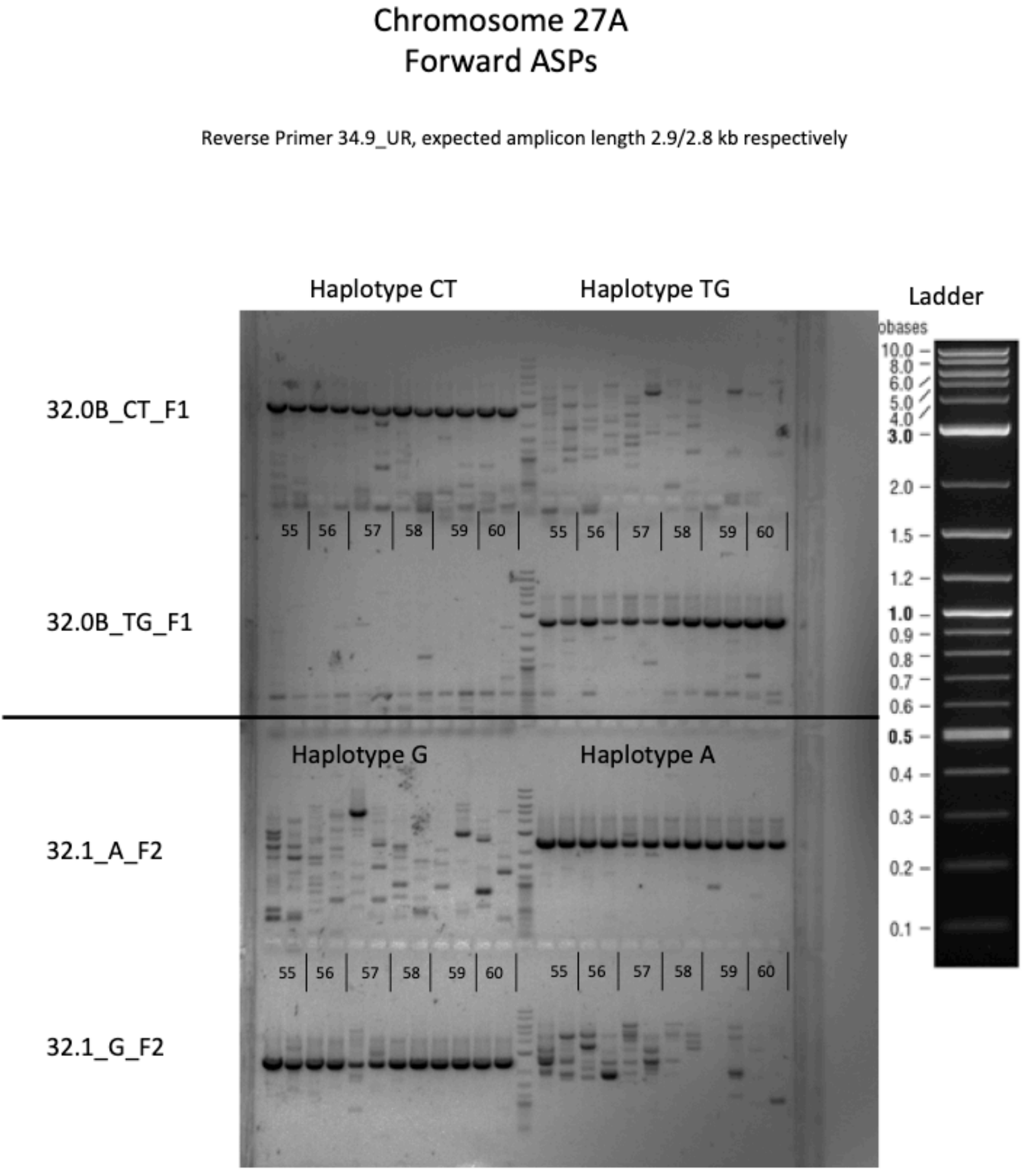
S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
seqnames source feature start end   absScore  relScore strand
ID     TF                                      class        siteSeqs
1  SNPdownstream2T   TFBS    TFBS     6  20  4.9014282 0.8585228
- MA0504.1  NR2C2   Nuclear receptors with C4 zinc fingers
AAAGTACAGAGTTTA
2  SNPdownstream2T   TFBS    TFBS     6  20 12.4788502 0.9347750
- MA0484.1  HNF4G   Nuclear receptors with C4 zinc fingers
AAAGTACAGAGTTTA
3  SNPdownstream2T   TFBS    TFBS     7  20  1.5493804 0.8527661
- MA0523.1 TCF7L2 High-mobility group (HMG) domain factors
AAAGTACAGAGTTT
4  SNPdownstream2T   TFBS    TFBS     1  15  2.1762790 0.8928958
- MA0148.3  FOXA1         Fork head / winged helix factors
ACAGAGTTTATATTT
5  SNPdownstream2T   TFBS    TFBS     4  13  6.5349852 0.9272398
+ MA0635.1 BARHL2                     Homeo domain factors
TATAAACTCT
6  SNPdownstream2T   TFBS    TFBS     5  11  3.8962750 0.9104617
+ MA0033.2  FOXL1         Fork head / winged helix factors
ATAAACT
7  SNPdownstream2T   TFBS    TFBS     5  12  1.4660498 0.8670023
+ MA0711.1   OTX1                     Homeo domain factors
ATAAACTC
8  SNPdownstream2T   TFBS    TFBS     5  12  3.0092813 0.8590024
+ MA0712.1   OTX2                     Homeo domain factors
ATAAACTC
9  SNPdownstream2T   TFBS    TFBS     1   7  0.6104832 0.8555117
+ MA0847.1  FOXD2         Fork head / winged helix factors
AAATATA
10 SNPdownstream2T   TFBS    TFBS     5  11  7.3293939 0.9423221
+ MA0847.1  FOXD2         Fork head / winged helix factors
ATAAACT
11 SNPdownstream2T   TFBS    TFBS     4  10  0.5500784 0.8547313
- MA0847.1  FOXD2         Fork head / winged helix factors
GTTTATA
12 SNPdownstream2T   TFBS    TFBS    11  17  1.2221221 0.8634143
- MA0847.1  FOXD2         Fork head / winged helix factors
GTACAGA
13 SNPdownstream2T   TFBS    TFBS     5  11  4.0542012 0.8718229
+ MA0848.1  FOXO4         Fork head / winged helix factors
ATAAACT
14 SNPdownstream2T   TFBS    TFBS    11  17  0.3575004 0.8771604
- MA0849.1  FOXO6         Fork head / winged helix factors
GTACAGA
15 SNPdownstream2T   TFBS    TFBS     4  10  1.8272052 0.8582904
- MA0850.1  FOXP3         Fork head / winged helix factors
GTTTATA
16 SNPdownstream2T   TFBS    TFBS    15  21  2.8981875 0.8736586
- MA0850.1  FOXP3         Fork head / winged helix factors
AAAAGTA
17 SNPdownstream2T   TFBS    TFBS     4  13  3.8231303 0.8728953
+ MA0877.1 Barhl1                     Homeo domain factors
TATAAACTCT
18 SNPdownstream2T   TFBS    TFBS     6  13  8.8951925 0.9176373
```

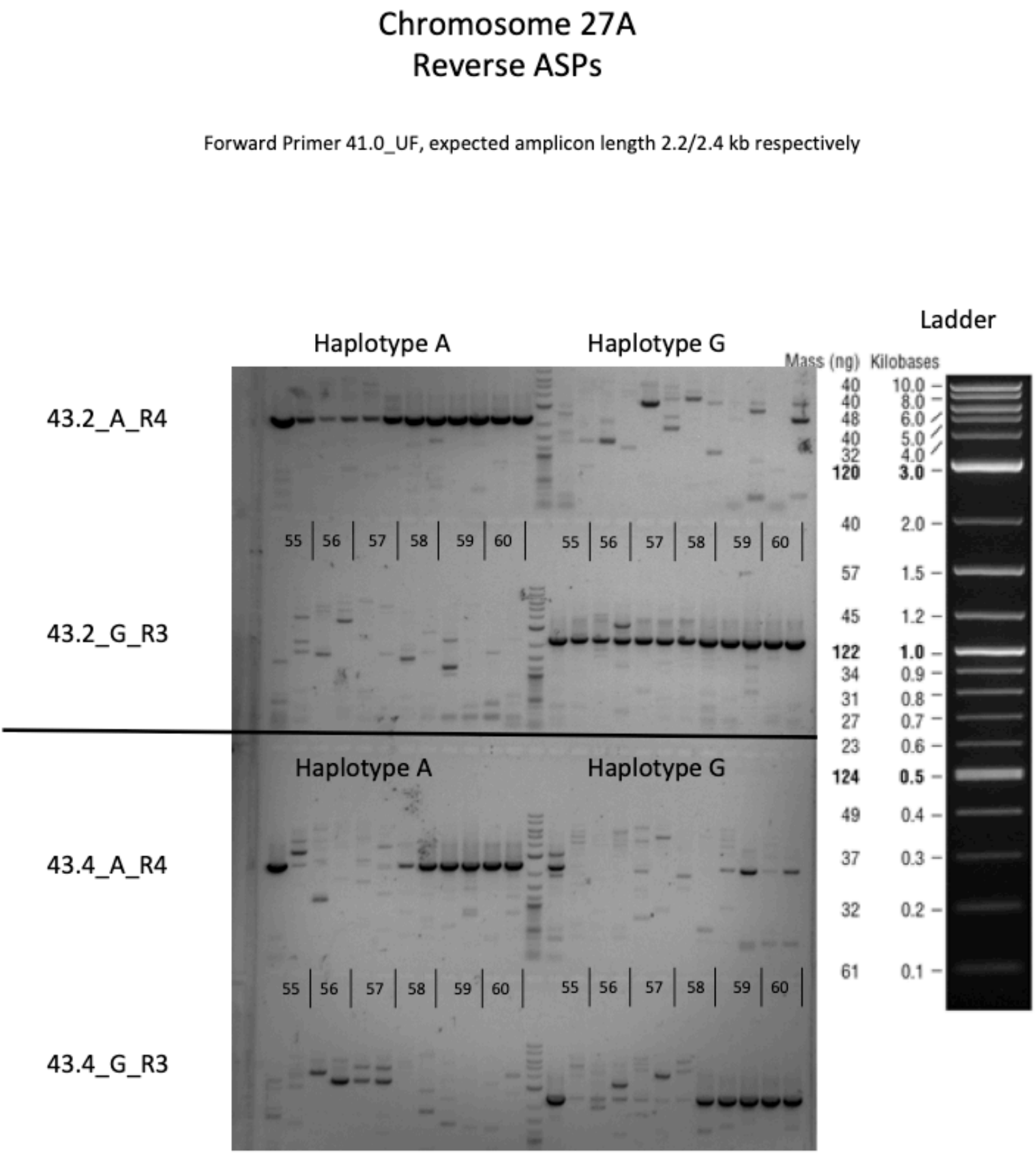
S13 (continued). Results from JASPAR motif search. SNPupstream1 and 2 refer to TD-SNPs 1 and 2. SNPcenter refers to TD-SNP 3 and SNPdownstream1 and 2 refer TD-SNPs 4 and 5.

```
– MA0693.2    VDR   Nuclear receptors with C4 zinc fingers
AGAGTTTA
19 SNPdownstream2T   TFBS    TFBS    13  20  6.5603957 0.8561533
– MA0693.2    VDR   Nuclear receptors with C4 zinc fingers
AAAGTACA
```

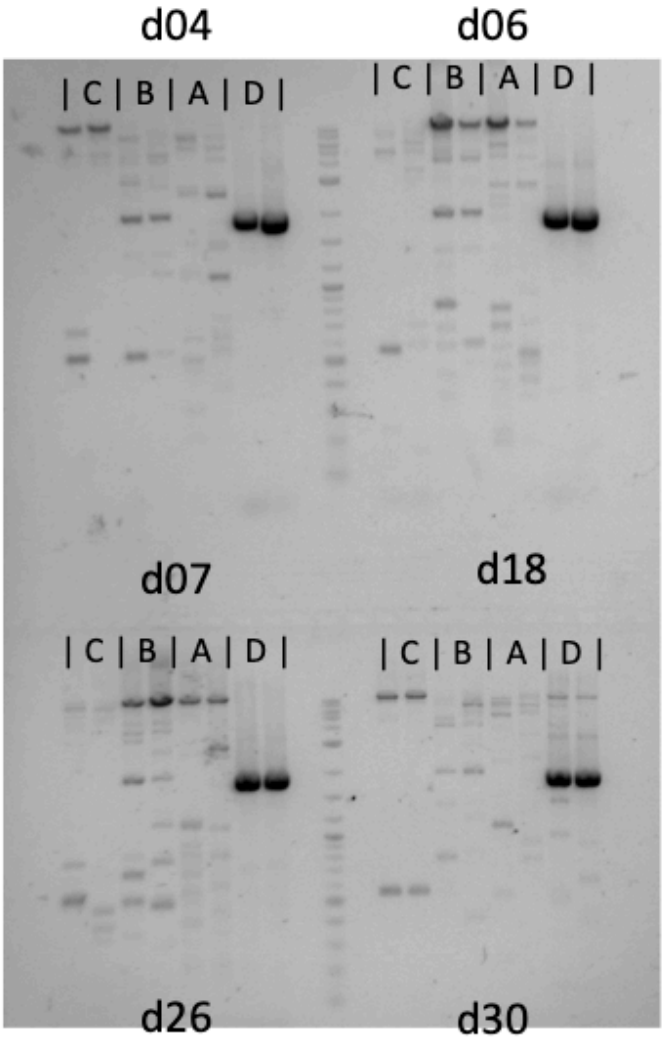
S14. Photo of agarose gel from ASP testing PCR. Each reaction was carried out twice (adjacent lanes on the gel). Annealing temperatures in °C are indicated above lanes. ASP tests foward

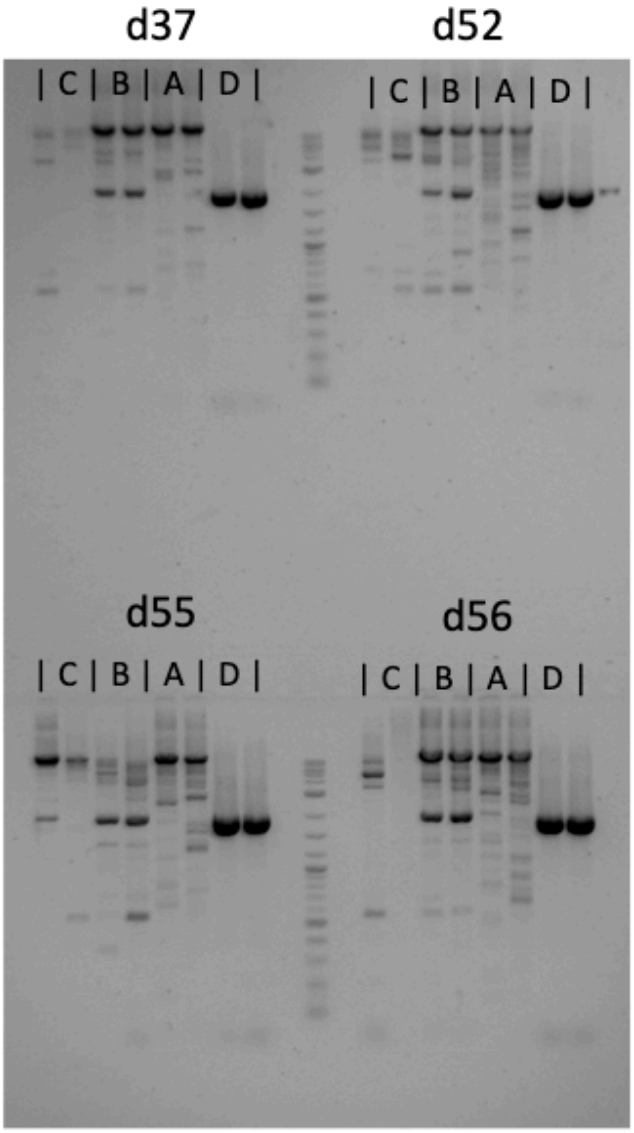


S15. Photo of agarose gel from ASP testing PCR. Each reaction was carried out twice (adjacent lanes on the gel). Annealing temperatures in °C are indicated above lanes. ASP tests reverse

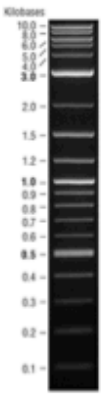


S16. Photos of agarose gels from Phasing PCRs. Each reaction was carried out twice (adjacent lanes on the gel). Respective Haplotypes as described in paragraph 3.8.3 are shown above lanes. Expected fragment lengths were ca. 10.5 kb

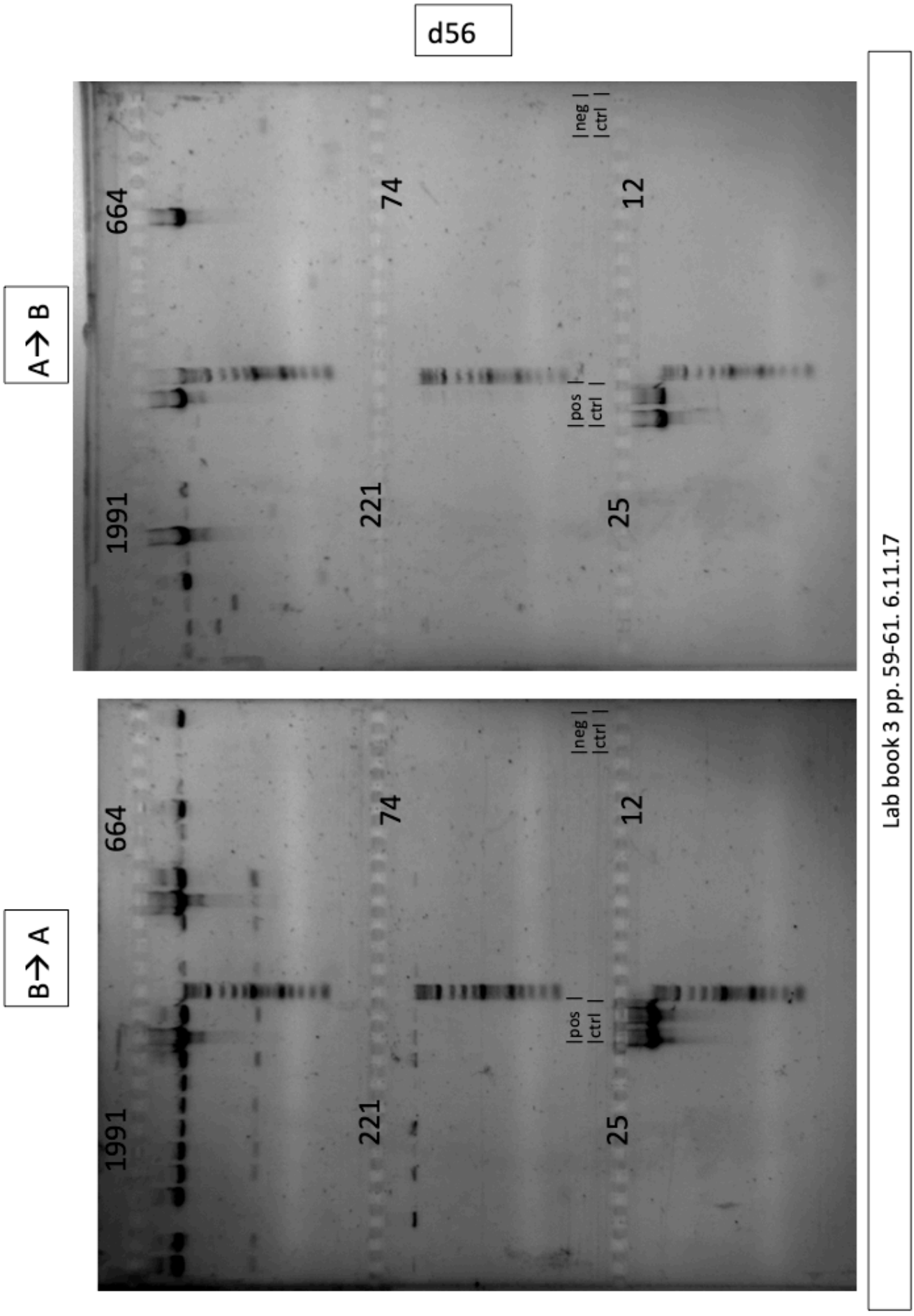




d55 was excluded from further analysis due to inconsistent phasing → probably one of the markers was incorrectly sequenced heterozygous while it is really homozygous
assays for C/D haplotypes could not be established



S17. Exemplary photo of agarose gel from crossover assay PCRs. Reciprocal orientations are indicated above pictures. Input molecules per per pool (amplifiable molecules) are shown above each strip of 12 reactions.



S18. Poisson corrected crossover numbers per individual and interval. Position refers to Chromosome 27 (CanFam3.1).

		sum	d06	d07	d26	d30	d37	d52	d56	sum-rounded
Int1	**1.6032687e+007**	0.00000	0.00000	0.00000	0.00000	0.0000	0.00000	0.000	0.00000	0
Int2	**1.6032775e+007**	3.23266	0.77527	2.45739	0.00000	0.0000	0.00000	0.000	0.00000	3
Int3	**1.6033410e+007**	4.20506	0.77527	2.45739	0.00000	0.0000	0.00000	0.000	0.97240	4
Int4	**1.6033468e+007**	6.27840	1.78601	2.45739	0.00000	1.0626	0.00000	0.000	0.97240	6
Int5	**1.6033653e+007**	6.27840	1.78601	2.45739	0.00000	1.0626	0.00000	0.000	0.97240	6
Int6	**1.6033979e+007**	9.64801	4.11182	2.45739	0.00000	1.0626	0.00000	0.000	2.01620	10
Int7	**1.6034366e+007**	11.80928	4.88709	2.45739	0.00000	1.0626	0.00000	1.386	2.01620	12
Int8	**1.6034701e+007**	11.80928	4.88709	2.45739	0.00000	1.0626	0.00000	1.386	2.01620	12
Int9	**1.6034810e+007**	15.18068	4.88709	2.45739	0.00000	3.3902	0.00000	1.386	3.06000	15
Int10	**1.6035299e+007**	15.18068	4.88709	2.45739	0.00000	3.3902	0.00000	1.386	3.06000	15
Int11	**1.6035449e+007**	16.29614	4.88709	2.45739	1.11546	3.3902	0.00000	1.386	3.06000	16
Int12	**1.6035528e+007**	16.29614	4.88709	2.45739	1.11546	3.3902	0.00000	1.386	3.06000	16
Int13	**1.6035615e+007**	16.29614	4.88709	2.45739	1.11546	3.3902	0.00000	1.386	3.06000	16
Int14	**1.6035799e+007**	17.33994	4.88709	2.45739	1.11546	3.3902	0.00000	1.386	4.10380	17
Int15	**1.6036036e+007**	17.33994	4.88709	2.45739	1.11546	3.3902	0.00000	1.386	4.10380	17
Int16	**1.6036529e+007**	23.03411	5.89783	2.45739	4.46184	3.3902	0.00000	1.386	5.44085	23
Int17	**1.6036544e+007**	23.03411	5.89783	2.45739	4.46184	3.3902	0.00000	1.386	5.44085	23
Int18	**1.6036614e+007**	27.54451	5.89783	2.45739	4.46184	4.6046	3.29600	1.386	5.44085	28
Int19	**1.6036933e+007**	27.54451	5.89783	2.45739	4.46184	4.6046	3.29600	1.386	5.44085	28
Int20	**1.6037111e+007**	28.51691	5.89783	2.45739	4.46184	4.6046	3.29600	1.386	6.41325	29
Int21	**1.6037287e+007**	28.51691	5.89783	2.45739	4.46184	4.6046	3.29600	1.386	6.41325	29
Int22	**1.6037387e+007**	28.51691	5.89783	2.45739	4.46184	4.6046	3.29600	1.386	6.41325	29
Int23	**1.6037392e+007**	29.59995	5.89783	2.45739	4.46184	4.6046	4.37904	1.386	6.41325	30
Int24	**1.6037449e+007**	29.59995	5.89783	2.45739	4.46184	4.6046	4.37904	1.386	6.41325	30
Int25	**1.6037568e+007**	29.59995	5.89783	2.45739	4.46184	4.6046	4.37904	1.386	6.41325	30
Int26	**1.6037643e+007**	34.15668	8.22364	2.45739	6.69276	4.6046	4.37904	1.386	6.41325	34
Int27	**1.6038036e+007**	43.56339	11.52576	3.41958	6.69276	8.7032	4.37904	1.386	7.45705	44
Int28	**1.6038089e+007**	44.67659	11.52576	3.41958	6.69276	9.8164	4.37904	1.386	7.45705	45
Int29	**1.6038312e+007**	46.85611	11.52576	4.51606	6.69276	9.8164	5.46208	1.386	7.45705	47
Int30	**1.6038422e+007**	64.68196	11.52576	4.51606	23.42466	9.8164	5.46208	1.386	8.55100	65
Int31	**1.6038526e+007**	69.94004	14.62684	4.51606	23.42466	10.9296	5.46208	1.386	9.59480	70
Int32	**1.6038671e+007**	80.48949	25.17629	4.51606	23.42466	10.9296	5.46208	1.386	9.59480	80
Int33	**1.6038845e+007**	80.48949	25.17629	4.51606	23.42466	10.9296	5.46208	1.386	9.59480	80
Int34	**1.6039080e+007**	87.18225	25.17629	4.51606	30.11742	10.9296	5.46208	1.386	9.59480	87
Int35	**1.6039219e+007**	93.81085	25.17629	4.51606	30.11742	17.5582	5.46208	1.386	9.59480	94
Int36	**1.6039387e+007**	96.02279	25.17629	5.61254	31.23288	17.5582	5.46208	1.386	9.59480	96
Int37	**1.6039448e+007**	96.02279	25.17629	5.61254	31.23288	17.5582	5.46208	1.386	9.59480	96
Int38	**1.6039537e+007**	96.02279	25.17629	5.61254	31.23288	17.5582	5.46208	1.386	9.59480	96
Int39	**1.6039689e+007**	96.02279	25.17629	5.61254	31.23288	17.5582	5.46208	1.386	9.59480	96

S18 (continued). Poisson corrected crossover numbers per individual and interval. Position refers to Chromosome27 (CanFam3.1).

		sum	d06	d07	d26	d30	d37	d52	d56	sum-rounded
Int40	**1.6039763e+007**	96.02279	25.17629	5.61254	31.23288	17.5582	5.46208	1.386	9.59480	96
Int41	**1.6040231e+007**	96.02279	25.17629	5.61254	31.23288	17.5582	5.46208	1.386	9.59480	96
Int42	**1.6040474e+007**	104.47904	28.20851	6.57473	35.69472	17.5582	5.46208	1.386	9.59480	104
Int43	**1.6040553e+007**	105.59450	28.20851	6.57473	36.81018	17.5582	5.46208	1.386	9.59480	106
Int44	**1.6040565e+007**	105.59450	28.20851	6.57473	36.81018	17.5582	5.46208	1.386	9.59480	106
Int45	**1.6040712e+007**	105.59450	28.20851	6.57473	36.81018	17.5582	5.46208	1.386	9.59480	106
Int46	**1.6040725e+007**	105.59450	28.20851	6.57473	36.81018	17.5582	5.46208	1.386	9.59480	106
Int47	**1.6040779e+007**	105.59450	28.20851	6.57473	36.81018	17.5582	5.46208	1.386	9.59480	106
Int48	**1.6040921e+007**	105.59450	28.20851	6.57473	36.81018	17.5582	5.46208	1.386	9.59480	106
Int49	**1.6040949e+007**	106.70996	28.20851	6.57473	37.92564	17.5582	5.46208	1.386	9.59480	107
Int50	**1.6041044e+007**	106.70996	28.20851	6.57473	37.92564	17.5582	5.46208	1.386	9.59480	107
Int51	**1.6041120e+007**	107.67215	28.20851	7.53692	37.92564	17.5582	5.46208	1.386	9.59480	108
Int52	**1.6041236e+007**	107.67215	28.20851	7.53692	37.92564	17.5582	5.46208	1.386	9.59480	108
Int53	**1.6041251e+007**	108.78761	28.20851	7.53692	39.04110	17.5582	5.46208	1.386	9.59480	109
Int54	**1.6041278e+007**	108.78761	28.20851	7.53692	39.04110	17.5582	5.46208	1.386	9.59480	109
Int55	**1.6041353e+007**	108.78761	28.20851	7.53692	39.04110	17.5582	5.46208	1.386	9.59480	109
Int56	**1.6041412e+007**	108.78761	28.20851	7.53692	39.04110	17.5582	5.46208	1.386	9.59480	109
Int57	**1.6041424e+007**	108.78761	28.20851	7.53692	39.04110	17.5582	5.46208	1.386	9.59480	109
Int58	**1.6041578e+007**	108.78761	28.20851	7.53692	39.04110	17.5582	5.46208	1.386	9.59480	109
Int59	**1.6041633e+007**	108.78761	28.20851	7.53692	39.04110	17.5582	5.46208	1.386	9.59480	109
Int60	**1.6041776e+007**	109.90307	28.20851	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	110
Int61	**1.6041780e+007**	109.90307	28.20851	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	110
Int62	**1.6041823e+007**	109.90307	28.20851	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	110
Int63	**1.6041858e+007**	109.90307	28.20851	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	110
Int64	**1.6041885e+007**	109.90307	28.20851	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	110
Int65	**1.6042053e+007**	109.90307	28.20851	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	110
Int66	**1.6042084e+007**	109.90307	28.20851	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	110
Int67	**1.6042085e+007**	109.90307	28.20851	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	110
Int68	**1.6042112e+007**	109.90307	28.20851	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	110
Int69	**1.6042209e+007**	109.90307	28.20851	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	110
Int70	**1.6042424e+007**	110.91381	29.21925	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	111
Int71	**1.6042527e+007**	110.91381	29.21925	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	111
Int72	**1.6042550e+007**	110.91381	29.21925	7.53692	40.15656	17.5582	5.46208	1.386	9.59480	111
Int73	**1.6042612e+007**	111.97641	29.21925	7.53692	40.15656	18.6208	5.46208	1.386	9.59480	112
Int74	**1.6042665e+007**	114.10261	30.22999	7.53692	41.27202	18.6208	5.46208	1.386	9.59480	114
Int75	**1.6042739e+007**	114.10261	30.22999	7.53692	41.27202	18.6208	5.46208	1.386	9.59480	114
Int76	**1.6042767e+007**	114.10261	30.22999	7.53692	41.27202	18.6208	5.46208	1.386	9.59480	114
Int77	**1.6042943e+007**	114.10261	30.22999	7.53692	41.27202	18.6208	5.46208	1.386	9.59480	114
Int78	**1.6043032e+007**	115.21581	30.22999	7.53692	41.27202	19.7340	5.46208	1.386	9.59480	115

S19. Individual cumulative crossover fractions (normalized to total Poisson corrected crossover numbers per individual) per interval. Position refers to Chromosome27 (CanFam3.1).

	End	sum	d06	d07	d26	d30	d37	d52	d56
Int1	**16032687**	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0	0.000000000
Int2	**16032775**	0.028057434	0.025645725	0.326046979	0.000000000	0.000000000	0.000000000	0	0.000000000
Int3	**16033410**	0.036497248	0.025645725	0.326046979	0.000000000	0.000000000	0.000000000	0	0.101346563
Int4	**16033468**	0.054492521	0.059080734	0.326046979	0.000000000	0.053846154	0.000000000	0	0.101346563
Int5	**16033653**	0.054492521	0.059080734	0.326046979	0.000000000	0.053846154	0.000000000	0	0.101346563
Int6	**16033979**	0.083738595	0.136017908	0.326046979	0.000000000	0.053846154	0.000000000	0	0.210134656
Int7	**16034366**	0.102497044	0.161663633	0.326046979	0.000000000	0.053846154	0.000000000	1	0.210134656
Int8	**16034701**	0.102497044	0.161663633	0.326046979	0.000000000	0.053846154	0.000000000	1	0.210134656
Int9	**16034810**	0.131758654	0.161663633	0.326046979	0.000000000	0.171794872	0.000000000	1	0.318922750
Int10	**16035299**	0.131758654	0.161663633	0.326046979	0.000000000	0.171794872	0.000000000	1	0.318922750
Int11	**16035449**	0.141440137	0.161663633	0.326046979	0.027027027	0.171794872	0.000000000	1	0.318922750
Int12	**16035528**	0.141440137	0.161663633	0.326046979	0.027027027	0.171794872	0.000000000	1	0.318922750
Int13	**16035615**	0.141440137	0.161663633	0.326046979	0.027027027	0.171794872	0.000000000	1	0.318922750
Int14	**16035799**	0.150499658	0.161663633	0.326046979	0.027027027	0.171794872	0.000000000	1	0.427710843
Int15	**16036036**	0.150499658	0.161663633	0.326046979	0.027027027	0.171794872	0.000000000	1	0.427710843
Int16	**16036529**	0.199921434	0.195098642	0.326046979	0.108108108	0.171794872	0.000000000	1	0.567062367
Int17	**16036544**	0.199921434	0.195098642	0.326046979	0.108108108	0.171794872	0.000000000	1	0.567062367
Int18	**16036614**	0.239068840	0.195098642	0.326046979	0.108108108	0.233333333	0.603433124	1	0.567062367
Int19	**16036933**	0.239068840	0.195098642	0.326046979	0.108108108	0.233333333	0.603433124	1	0.567062367
Int20	**16037111**	0.247508654	0.195098642	0.326046979	0.108108108	0.233333333	0.603433124	1	0.668408930
Int21	**16037287**	0.247508654	0.195098642	0.326046979	0.108108108	0.233333333	0.603433124	1	0.668408930
Int22	**16037387**	0.247508654	0.195098642	0.326046979	0.108108108	0.233333333	0.603433124	1	0.668408930
Int23	**16037392**	0.256908752	0.195098642	0.326046979	0.108108108	0.233333333	0.801716562	1	0.668408930
Int24	**16037449**	0.256908752	0.195098642	0.326046979	0.108108108	0.233333333	0.801716562	1	0.668408930
Int25	**16037568**	0.256908752	0.195098642	0.326046979	0.108108108	0.233333333	0.801716562	1	0.668408930
Int26	**16037643**	0.296458273	0.272035816	0.326046979	0.162162162	0.233333333	0.801716562	1	0.668408930
Int27	**16038036**	0.378102536	0.381269064	0.453710534	0.162162162	0.441025641	0.801716562	1	0.777197023
Int28	**16038089**	0.387764405	0.381269064	0.453710534	0.162162162	0.497435897	0.801716562	1	0.777197023
Int29	**16038312**	0.406681253	0.381269064	0.599191712	0.162162162	0.497435897	1.000000000	1	0.777197023
Int30	**16038422**	0.561398301	0.381269064	0.599191712	0.567567568	0.497435897	1.000000000	1	0.891211906
Int31	**16038526**	0.607035094	0.483851963	0.599191712	0.567567568	0.553846154	1.000000000	1	1.000000000
Int32	**16038671**	0.698597614	0.832824953	0.599191712	0.567567568	0.553846154	1.000000000	1	1.000000000
Int33	**16038845**	0.698597614	0.832824953	0.599191712	0.567567568	0.553846154	1.000000000	1	1.000000000
Int34	**16039080**	0.756686517	0.832824953	0.599191712	0.729729730	0.553846154	1.000000000	1	1.000000000
Int35	**16039219**	0.814218552	0.832824953	0.599191712	0.729729730	0.889743590	1.000000000	1	1.000000000
Int36	**16039387**	0.833416785	0.832824953	0.744672890	0.756756757	0.889743590	1.000000000	1	1.000000000
Int37	**16039448**	0.833416785	0.832824953	0.744672890	0.756756757	0.889743590	1.000000000	1	1.000000000
Int38	**16039537**	0.833416785	0.832824953	0.744672890	0.756756757	0.889743590	1.000000000	1	1.000000000
Int39	**16039689**	0.833416785	0.832824953	0.744672890	0.756756757	0.889743590	1.000000000	1	1.000000000

S19 (continued). Individual cumulative crossover fractions (normalized to total Poisson corrected crossover numbers per individual) per interval. Position refers to Chromosome27 (CanFam3.1).

	End	sum	d06	d07	d26	d30	d37	d52	d56
Int40	**16039763**	0.833416785	0.832824953	0.744672890	0.756756757	0.889743590	1.000000000	1	1.000000000
Int41	**16040231**	0.833416785	0.832824953	0.744672890	0.756756757	0.889743590	1.000000000	1	1.000000000
Int42	**16040474**	0.906811661	0.933129981	0.872336445	0.864864865	0.889743590	1.000000000	1	1.000000000
Int43	**16040553**	0.916493144	0.933129981	0.872336445	0.891891892	0.889743590	1.000000000	1	1.000000000
Int44	**16040565**	0.916493144	0.933129981	0.872336445	0.891891892	0.889743590	1.000000000	1	1.000000000
Int45	**16040712**	0.916493144	0.933129981	0.872336445	0.891891892	0.889743590	1.000000000	1	1.000000000
Int46	**16040725**	0.916493144	0.933129981	0.872336445	0.891891892	0.889743590	1.000000000	1	1.000000000
Int47	**16040779**	0.916493144	0.933129981	0.872336445	0.891891892	0.889743590	1.000000000	1	1.000000000
Int48	**16040921**	0.916493144	0.933129981	0.872336445	0.891891892	0.889743590	1.000000000	1	1.000000000
Int49	**16040949**	0.926174628	0.933129981	0.872336445	0.918918919	0.889743590	1.000000000	1	1.000000000
Int50	**16041044**	0.926174628	0.933129981	0.872336445	0.918918919	0.889743590	1.000000000	1	1.000000000
Int51	**16041120**	0.934525826	0.933129981	1.000000000	0.918918919	0.889743590	1.000000000	1	1.000000000
Int52	**16041236**	0.934525826	0.933129981	1.000000000	0.918918919	0.889743590	1.000000000	1	1.000000000
Int53	**16041251**	0.944207310	0.933129981	1.000000000	0.945945946	0.889743590	1.000000000	1	1.000000000
Int54	**16041278**	0.944207310	0.933129981	1.000000000	0.945945946	0.889743590	1.000000000	1	1.000000000
Int55	**16041353**	0.944207310	0.933129981	1.000000000	0.945945946	0.889743590	1.000000000	1	1.000000000
Int56	**16041412**	0.944207310	0.933129981	1.000000000	0.945945946	0.889743590	1.000000000	1	1.000000000
Int57	**16041424**	0.944207310	0.933129981	1.000000000	0.945945946	0.889743590	1.000000000	1	1.000000000
Int58	**16041578**	0.944207310	0.933129981	1.000000000	0.945945946	0.889743590	1.000000000	1	1.000000000
Int59	**16041633**	0.944207310	0.933129981	1.000000000	0.945945946	0.889743590	1.000000000	1	1.000000000
Int60	**16041776**	0.953888794	0.933129981	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int61	**16041780**	0.953888794	0.933129981	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int62	**16041823**	0.953888794	0.933129981	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int63	**16041858**	0.953888794	0.933129981	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int64	**16041885**	0.953888794	0.933129981	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int65	**16042053**	0.953888794	0.933129981	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int66	**16042084**	0.953888794	0.933129981	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int67	**16042085**	0.953888794	0.933129981	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int68	**16042112**	0.953888794	0.933129981	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int69	**16042209**	0.953888794	0.933129981	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int70	**16042424**	0.962661374	0.966564991	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int71	**16042527**	0.962661374	0.966564991	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int72	**16042550**	0.962661374	0.966564991	1.000000000	0.972972973	0.889743590	1.000000000	1	1.000000000
Int73	**16042612**	0.971884067	0.966564991	1.000000000	0.972972973	0.943589744	1.000000000	1	1.000000000
Int74	**16042665**	0.990338132	1.000000000	1.000000000	1.000000000	0.943589744	1.000000000	1	1.000000000
Int75	**16042739**	0.990338132	1.000000000	1.000000000	1.000000000	0.943589744	1.000000000	1	1.000000000
Int76	**16042767**	0.990338132	1.000000000	1.000000000	1.000000000	0.943589744	1.000000000	1	1.000000000
Int77	**16042943**	0.990338132	1.000000000	1.000000000	1.000000000	0.943589744	1.000000000	1	1.000000000
Int78	**16043032**	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1	1.000000000

S20. De-novo crossover frequencies per individual and per reciprocal orientation. Upper/ lower limit correspond to 95% confidence intervals. Position refers to Chromosome27 (CanFam3.1).

	combined			AB			BA		
	Mean	**Upper Limit**	**Lower Limit**	**Mean**	**Upper Limit**	**Lower Limit**	**Mean**	**Upper Limit**	**Lower Limit**
Hotspot	0.148	0.165	0.132	0.115	0.135	0.097	0.186	0.215	0.160
d37	0.533	0.739	0.379	0.452	0.716	0.276	0.635	1.001	0.390
d26	0.522	0.695	0.386	0.383	0.557	0.255	0.935	1.493	0.567
d30	0.298	0.408	0.213	0.166	0.262	0.102	0.379	0.562	0.248
d52	0.237	0.333	0.164	0.270	0.428	0.165	0.218	0.346	0.133
d56	0.105	0.144	0.075	0.049	0.026	0.084	0.214	0.325	0.137
d06	0.102	0.131	0.077	0.060	0.090	0.038	0.102	0.145	0.069
d07	0.056	0.075	0.040	0.023	0.044	0.010	0.081	0.113	0.056

S21. d06 - Individual cumulative crossover fractions per interval and per reciprocal crossover orientation. Position refers to Chromosome27 (CanFam3.1).

	Position	combined	AB	BA
Int1	**1.6032687e+007**	0.00000	0.00000	0.00000
Int2	**1.6032775e+007**	0.02565	0.04167	0.00000
Int3	**1.6033410e+007**	0.02565	0.04167	0.00000
Int5	**1.6033653e+007**	0.05908	0.04167	0.08696
Int6	**1.6033979e+007**	0.13602	0.16667	0.08696
Int8	**1.6034701e+007**	0.16166	0.20833	0.08696
Int11	**1.6035449e+007**	0.16166	0.20833	0.08696
Int12	**1.6035528e+007**	0.16166	0.20833	0.08696
Int13	**1.6035615e+007**	0.16166	0.20833	0.08696
Int14	**1.6035799e+007**	0.16166	0.20833	0.08696
Int15	**1.6036036e+007**	0.16166	0.20833	0.08696
Int16	**1.6036529e+007**	0.19510	0.20833	0.17391
Int17	**1.6036544e+007**	0.19510	0.20833	0.17391
Int19	**1.6036933e+007**	0.19510	0.20833	0.17391
Int22	**1.6037387e+007**	0.19510	0.20833	0.17391
Int23	**1.6037392e+007**	0.19510	0.20833	0.17391
Int25	**1.6037568e+007**	0.19510	0.20833	0.17391
Int26	**1.6037643e+007**	0.27204	0.33333	0.17391
Int30	**1.6038422e+007**	0.38127	0.37500	0.39130
Int31	**1.6038526e+007**	0.48385	0.54167	0.39130
Int41	**1.6040231e+007**	0.83282	1.00000	0.56522
Int42	**1.6040474e+007**	0.93313	1.00000	0.82609
Int43	**1.6040553e+007**	0.93313	1.00000	0.82609
Int44	**1.6040565e+007**	0.93313	1.00000	0.82609
Int45	**1.6040712e+007**	0.93313	1.00000	0.82609
Int46	**1.6040725e+007**	0.93313	1.00000	0.82609
Int47	**1.6040779e+007**	0.93313	1.00000	0.82609
Int49	**1.6040949e+007**	0.93313	1.00000	0.82609
Int50	**1.6041044e+007**	0.93313	1.00000	0.82609
Int51	**1.6041120e+007**	0.93313	1.00000	0.82609
Int52	**1.6041236e+007**	0.93313	1.00000	0.82609
Int53	**1.6041251e+007**	0.93313	1.00000	0.82609
Int54	**1.6041278e+007**	0.93313	1.00000	0.82609
Int55	**1.6041353e+007**	0.93313	1.00000	0.82609
Int56	**1.6041412e+007**	0.93313	1.00000	0.82609
Int57	**1.6041424e+007**	0.93313	1.00000	0.82609
Int58	**1.6041578e+007**	0.93313	1.00000	0.82609
Int59	**1.6041633e+007**	0.93313	1.00000	0.82609
Int60	**1.6041776e+007**	0.93313	1.00000	0.82609
Int61	**1.6041780e+007**	0.93313	1.00000	0.82609
Int62	**1.6041823e+007**	0.93313	1.00000	0.82609
Int63	**1.6041858e+007**	0.93313	1.00000	0.82609
Int64	**1.6041885e+007**	0.93313	1.00000	0.82609
Int65	**1.6042053e+007**	0.93313	1.00000	0.82609
Int67	**1.6042085e+007**	0.93313	1.00000	0.82609
Int68	**1.6042112e+007**	0.93313	1.00000	0.82609
Int69	**1.6042209e+007**	0.93313	1.00000	0.82609
Int71	**1.6042527e+007**	0.96656	1.00000	0.91304
Int73	**1.6042612e+007**	0.96656	1.00000	0.91304
Int74	**1.6042665e+007**	1.00000	1.00000	1.00000
Int77	**1.6042943e+007**	1.00000	1.00000	1.00000
Int78	**1.6043032e+007**	1.00000	1.00000	1.00000

S22. d07 - Individual cumulative crossover fractions per interval and per reciprocal crossover orientation. Position refers to Chromosome27 (CanFam3.1).

		combined	AB	BA
Int1	**16032687**	0.000	0.00	0.000
Int3	**16033410**	0.326	0.25	0.405
Int5	**16033653**	0.326	0.25	0.405
Int6	**16033979**	0.326	0.25	0.405
Int7	**16034366**	0.326	0.25	0.405
Int8	**16034701**	0.326	0.25	0.405
Int10	**16035299**	0.326	0.25	0.405
Int11	**16035449**	0.326	0.25	0.405
Int13	**16035615**	0.326	0.25	0.405
Int18	**16036614**	0.326	0.25	0.405
Int19	**16036933**	0.326	0.25	0.405
Int20	**16037111**	0.326	0.25	0.405
Int21	**16037287**	0.326	0.25	0.405
Int26	**16037643**	0.326	0.25	0.405
Int28	**16038089**	0.454	0.50	0.405
Int29	**16038312**	0.599	0.50	0.703
Int30	**16038422**	0.599	0.50	0.703
Int32	**16038671**	0.599	0.50	0.703
Int34	**16039080**	0.599	0.50	0.703
Int35	**16039219**	0.599	0.50	0.703
Int41	**16040231**	0.745	0.50	1.000
Int42	**16040474**	0.872	0.75	1.000
Int43	**16040553**	0.872	0.75	1.000
Int44	**16040565**	0.872	0.75	1.000
Int45	**16040712**	0.872	0.75	1.000
Int46	**16040725**	0.872	0.75	1.000
Int47	**16040779**	0.872	0.75	1.000
Int49	**16040949**	0.872	0.75	1.000
Int50	**16041044**	0.872	0.75	1.000
Int51	**16041120**	1.000	1.00	1.000
Int52	**16041236**	1.000	1.00	1.000
Int53	**16041251**	1.000	1.00	1.000
Int54	**16041278**	1.000	1.00	1.000
Int55	**16041353**	1.000	1.00	1.000
Int56	**16041412**	1.000	1.00	1.000
Int57	**16041424**	1.000	1.00	1.000
Int58	**16041578**	1.000	1.00	1.000
Int59	**16041633**	1.000	1.00	1.000
Int60	**16041776**	1.000	1.00	1.000
Int62	**16041823**	1.000	1.00	1.000
Int63	**16041858**	1.000	1.00	1.000
Int64	**16041885**	1.000	1.00	1.000
Int65	**16042053**	1.000	1.00	1.000
Int67	**16042085**	1.000	1.00	1.000
Int68	**16042112**	1.000	1.00	1.000
Int69	**16042209**	1.000	1.00	1.000
Int70	**16042424**	1.000	1.00	1.000
Int71	**16042527**	1.000	1.00	1.000
Int72	**16042550**	1.000	1.00	1.000
Int74	**16042665**	1.000	1.00	1.000
Int76	**16042767**	1.000	1.00	1.000
Int77	**16042943**	1.000	1.00	1.000
Int78	**16043032**	1.000	1.00	1.000

S23. d26 - Individual cumulative crossover fractions per interval and per reciprocal crossover orientation. Position refers to Chromosome27 (CanFam3.1).

	Position	combined	AB	BA
Int1	**16032687**	0.00000	0.00000	0.00000
Int2	**16032775**	0.00000	0.00000	0.00000
Int3	**16033410**	0.00000	0.00000	0.00000
Int4	**16033468**	0.00000	0.00000	0.00000
Int5	**16033653**	0.00000	0.00000	0.00000
Int6	**16033979**	0.00000	0.00000	0.00000
Int7	**16034366**	0.00000	0.00000	0.00000
Int8	**16034701**	0.00000	0.00000	0.00000
Int11	**16035449**	0.02703	0.00000	0.06383
Int12	**16035528**	0.02703	0.00000	0.06383
Int13	**16035615**	0.02703	0.00000	0.06383
Int16	**16036529**	0.10811	0.03226	0.19149
Int17	**16036544**	0.10811	0.03226	0.19149
Int26	**16037643**	0.16216	0.03226	0.31915
Int30	**16038422**	0.56757	0.48387	0.68085
Int34	**16039080**	0.72973	0.80645	0.68085
Int36	**16039387**	0.75676	0.83871	0.68085
Int37	**16039448**	0.75676	0.83871	0.68085
Int38	**16039537**	0.75676	0.83871	0.68085
Int39	**16039689**	0.75676	0.83871	0.68085
Int42	**16040474**	0.86486	0.96774	0.74468
Int43	**16040553**	0.89189	0.96774	0.80851
Int44	**16040565**	0.89189	0.96774	0.80851
Int45	**16040712**	0.89189	0.96774	0.80851
Int46	**16040725**	0.89189	0.96774	0.80851
Int47	**16040779**	0.89189	0.96774	0.80851
Int49	**16040949**	0.91892	1.00000	0.80851
Int50	**16041044**	0.91892	1.00000	0.80851
Int51	**16041120**	0.91892	1.00000	0.80851
Int53	**16041251**	0.94595	1.00000	0.87234
Int54	**16041278**	0.94595	1.00000	0.87234
Int55	**16041353**	0.94595	1.00000	0.87234
Int56	**16041412**	0.94595	1.00000	0.87234
Int57	**16041424**	0.94595	1.00000	0.87234
Int58	**16041578**	0.94595	1.00000	0.87234
Int59	**16041633**	0.94595	1.00000	0.87234
Int60	**16041776**	0.97297	1.00000	0.93617
Int62	**16041823**	0.97297	1.00000	0.93617
Int63	**16041858**	0.97297	1.00000	0.93617
Int64	**16041885**	0.97297	1.00000	0.93617
Int65	**16042053**	0.97297	1.00000	0.93617
Int66	**16042084**	0.97297	1.00000	0.93617
Int67	**16042085**	0.97297	1.00000	0.93617
Int68	**16042112**	0.97297	1.00000	0.93617
Int69	**16042209**	0.97297	1.00000	0.93617
Int71	**16042527**	0.97297	1.00000	0.93617
Int72	**16042550**	0.97297	1.00000	0.93617
Int74	**16042665**	1.00000	1.00000	1.00000
Int75	**16042739**	1.00000	1.00000	1.00000
Int77	**16042943**	1.00000	1.00000	1.00000
Int78	**16043032**	1.00000	1.00000	1.00000

S24. d30 - Individual cumulative crossover fractions per interval and per reciprocal crossover orientation. Position refers to Chromosome27 (CanFam3.1).

	Position	combined	AB	BA
Int1	**16032687**	0.00000	0.00000	0.00000
Int3	**16033410**	0.00000	0.00000	0.00000
Int5	**16033653**	0.05385	0.00000	0.10938
Int6	**16033979**	0.05385	0.00000	0.10938
Int8	**16034701**	0.05385	0.00000	0.10938
Int11	**16035449**	0.17179	0.11111	0.23438
Int12	**16035528**	0.17179	0.11111	0.23438
Int13	**16035615**	0.17179	0.11111	0.23438
Int14	**16035799**	0.17179	0.11111	0.23438
Int15	**16036036**	0.17179	0.11111	0.23438
Int16	**16036529**	0.17179	0.11111	0.23438
Int17	**16036544**	0.17179	0.11111	0.23438
Int19	**16036933**	0.23333	0.11111	0.35938
Int26	**16037643**	0.23333	0.11111	0.35938
Int27	**16038036**	0.44103	0.11111	0.78125
Int29	**16038312**	0.49744	0.22222	0.78125
Int30	**16038422**	0.49744	0.22222	0.78125
Int31	**16038526**	0.55385	0.33333	0.78125
Int33	**16038845**	0.55385	0.33333	0.78125
Int34	**16039080**	0.55385	0.33333	0.78125
Int42	**16040474**	0.88974	0.88889	0.89063
Int44	**16040565**	0.88974	0.88889	0.89063
Int45	**16040712**	0.88974	0.88889	0.89063
Int46	**16040725**	0.88974	0.88889	0.89063
Int47	**16040779**	0.88974	0.88889	0.89063
Int49	**16040949**	0.88974	0.88889	0.89063
Int50	**16041044**	0.88974	0.88889	0.89063
Int51	**16041120**	0.88974	0.88889	0.89063
Int53	**16041251**	0.88974	0.88889	0.89063
Int54	**16041278**	0.88974	0.88889	0.89063
Int55	**16041353**	0.88974	0.88889	0.89063
Int56	**16041412**	0.88974	0.88889	0.89063
Int57	**16041424**	0.88974	0.88889	0.89063
Int58	**16041578**	0.88974	0.88889	0.89063
Int59	**16041633**	0.88974	0.88889	0.89063
Int60	**16041776**	0.88974	0.88889	0.89063
Int62	**16041823**	0.88974	0.88889	0.89063
Int63	**16041858**	0.88974	0.88889	0.89063
Int64	**16041885**	0.88974	0.88889	0.89063
Int65	**16042053**	0.88974	0.88889	0.89063
Int67	**16042085**	0.88974	0.88889	0.89063
Int68	**16042112**	0.88974	0.88889	0.89063
Int69	**16042209**	0.88974	0.88889	0.89063
Int71	**16042527**	0.88974	0.88889	0.89063
Int72	**16042550**	0.88974	0.88889	0.89063
Int74	**16042665**	0.94359	0.88889	1.00000
Int75	**16042739**	0.94359	0.88889	1.00000
Int77	**16042943**	0.94359	0.88889	1.00000
Int78	**16043032**	1.00000	1.00000	1.00000

S25. d56 - cIndividual cumulative crossover fractions per interval and per reciprocal crossover orientation. Position refers to Chromosome27 (CanFam3.1).

	Pos	combined	AB	BA
Int1	**16032687**	0.00000	0.0	0.00000
Int2	**16032775**	0.00000	0.0	0.00000
Int3	**16033410**	0.10135	0.0	0.22222
Int5	**16033653**	0.10135	0.0	0.22222
Int6	**16033979**	0.21013	0.2	0.22222
Int8	**16034701**	0.21013	0.2	0.22222
Int11	**16035449**	0.31892	0.4	0.22222
Int12	**16035528**	0.31892	0.4	0.22222
Int13	**16035615**	0.31892	0.4	0.22222
Int14	**16035799**	0.42771	0.6	0.22222
Int15	**16036036**	0.42771	0.6	0.22222
Int16	**16036529**	0.56706	0.6	0.52778
Int19	**16036933**	0.56706	0.6	0.52778
Int26	**16037643**	0.66841	0.6	0.75000
Int28	**16038089**	0.77720	0.8	0.75000
Int29	**16038312**	0.77720	0.8	0.75000
Int30	**16038422**	0.89121	0.8	1.00000
Int33	**16038845**	1.00000	1.0	1.00000
Int34	**16039080**	1.00000	1.0	1.00000
Int35	**16039219**	1.00000	1.0	1.00000
Int40	**16039763**	1.00000	1.0	1.00000
Int41	**16040231**	1.00000	1.0	1.00000
Int42	**16040474**	1.00000	1.0	1.00000
Int43	**16040553**	1.00000	1.0	1.00000
Int44	**16040565**	1.00000	1.0	1.00000
Int45	**16040712**	1.00000	1.0	1.00000
Int46	**16040725**	1.00000	1.0	1.00000
Int47	**16040779**	1.00000	1.0	1.00000
Int49	**16040949**	1.00000	1.0	1.00000
Int50	**16041044**	1.00000	1.0	1.00000
Int51	**16041120**	1.00000	1.0	1.00000
Int52	**16041236**	1.00000	1.0	1.00000
Int53	**16041251**	1.00000	1.0	1.00000
Int54	**16041278**	1.00000	1.0	1.00000
Int55	**16041353**	1.00000	1.0	1.00000
Int56	**16041412**	1.00000	1.0	1.00000
Int57	**16041424**	1.00000	1.0	1.00000
Int58	**16041578**	1.00000	1.0	1.00000
Int59	**16041633**	1.00000	1.0	1.00000
Int60	**16041776**	1.00000	1.0	1.00000
Int61	**16041780**	1.00000	1.0	1.00000
Int62	**16041823**	1.00000	1.0	1.00000
Int63	**16041858**	1.00000	1.0	1.00000
Int64	**16041885**	1.00000	1.0	1.00000
Int65	**16042053**	1.00000	1.0	1.00000
Int67	**16042085**	1.00000	1.0	1.00000
Int68	**16042112**	1.00000	1.0	1.00000
Int69	**16042209**	1.00000	1.0	1.00000
Int70	**16042424**	1.00000	1.0	1.00000
Int71	**16042527**	1.00000	1.0	1.00000
Int72	**16042550**	1.00000	1.0	1.00000
Int74	**16042665**	1.00000	1.0	1.00000
Int76	**16042767**	1.00000	1.0	1.00000
Int77	**16042943**	1.00000	1.0	1.00000
Int78	**16043032**	1.00000	1.0	1.00000

S26. De-novo crossover frequencies per group (Homozygous S, Homozygous W and Heterozygous) per TD-SNP. Upper/lower limit correspond to 95% confidence intervals. Position refers to Chromosome27 (CanFam3.1).

	Homozygous W			Heterozygous			Homozygous S		
	Mean	Upper Limit	Lower Limit	Mean	Upper Limit	Lower Limit	Mean	Upper Limit	Lower Limit
SNP1	11.52	13.88	9.49	15.53	17.76	13.53			
SNP2	11.52	13.88	9.49	15.53	17.76	13.53			
SNP3	5.02	6.79	5.02	12.97	15.76	10.85	23.42	27.31	19.98
SNP4				10.70	12.22	9.33	35.78	44.22	28.71
SNP5	35.78	44.22	28.71	10.70	12.22	9.33			

S27. Cumulative crossover fractions grouped by TD-SNPs - SNP 1&2

		d06AB	d06BA	d30AB	d30BA	d37-AB	d37-BA	d52-BA	d56-AB	d56-BA	sumTC	sumCT
Int1	**1.6032687e+007**	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000
Int2	**1.6032775e+007**	0.041383161	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.021371041	0.000000000
Int3	**1.6033410e+007**	0.041383161	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.222659414	0.021371041	0.034042839
Int4	**1.6033468e+007**	0.041383161	0.020750194	0.000000000	0.026077932	0.000000000	0.000000000	0.000000000	0.000000000	0.222659414	0.021371041	0.051324000
Int5	**1.6033653e+007**	0.041383161	0.086936158	0.000000000	0.109257541	0.000000000	0.000000000	0.000000000	0.000000000	0.222659414	0.021371041	0.106444944
Int6	**1.6033979e+007**	0.166559456	0.086936158	0.000000000	0.109257541	0.000000000	0.000000000	0.000000000	0.199964046	0.222659414	0.115025726	0.106444944
Int7	**1.6034366e+007**	0.188962326	0.086936158	0.000000000	0.109257541	0.000000000	0.000000000	0.536011080	0.199964046	0.222659414	0.126594988	0.119564347
Int8	**1.6034701e+007**	0.208354992	0.086936158	0.000000000	0.109257541	0.000000000	0.000000000	1.000000000	0.199964046	0.222659414	0.136609725	0.130920935
Int9	**1.6034810e+007**	0.208354992	0.086936158	0.016222276	0.127503706	0.000000000	0.000000000	1.000000000	0.229153552	0.222659414	0.145361071	0.137117781
Int10	**1.6035299e+007**	0.208354992	0.086936158	0.088999275	0.209360356	0.000000000	0.000000000	1.000000000	0.360104639	0.222659414	0.184621696	0.164918307
Int11	**1.6035449e+007**	0.208354992	0.086936158	0.111323507	0.234469757	0.000000000	0.000000000	1.000000000	0.400273684	0.222659414	0.196664832	0.173446076
Int12	**1.6035528e+007**	0.208354992	0.086936158	0.111323507	0.234469757	0.000000000	0.000000000	1.000000000	0.400273684	0.222659414	0.196664832	0.173446076
Int13	**1.6035615e+007**	0.208354992	0.086936158	0.111323507	0.234469757	0.000000000	0.000000000	1.000000000	0.400273684	0.222659414	0.196664832	0.173446076
Int14	**1.6035799e+007**	0.208354992	0.086936158	0.111323507	0.234469757	0.000000000	0.000000000	1.000000000	0.599766914	0.222659414	0.225607823	0.173446076
Int15	**1.6036036e+007**	0.208354992	0.086936158	0.111323507	0.234469757	0.000000000	0.000000000	1.000000000	0.599766914	0.222659414	0.225607823	0.173446076
Int16	**1.6036529e+007**	0.208354992	0.174053371	0.111323507	0.234469757	0.000000000	0.000000000	1.000000000	0.599766914	0.528677519	0.225607823	0.255602861
Int17	**1.6036544e+007**	0.208354992	0.174053371	0.111323507	0.234469757	0.000000000	0.000000000	1.000000000	0.599766914	0.528677519	0.225607823	0.255602861
Int18	**1.6036614e+007**	0.208354992	0.174053371	0.111323507	0.256973790	0.000000000	0.179948586	1.000000000	0.599766914	0.528677519	0.225607823	0.277104242
Int19	**1.6036933e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.000000000	1.000000000	1.000000000	0.599766914	0.528677519	0.225607823	0.375089109
Int20	**1.6037111e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.000000000	1.000000000	1.000000000	0.599766914	0.584508448	0.225607823	0.383625210
Int21	**1.6037287e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.000000000	1.000000000	1.000000000	0.599766914	0.639712063	0.225607823	0.392065400
Int22	**1.6037387e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.000000000	1.000000000	1.000000000	0.599766914	0.671077753	0.225607823	0.396860963
Int23	**1.6037392e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.013805004	1.000000000	1.000000000	0.599766914	0.672646038	0.226437336	0.397100741
Int24	**1.6037449e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.171182053	1.000000000	1.000000000	0.599766914	0.690524481	0.235893783	0.399834212
Int25	**1.6037568e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.499741156	1.000000000	1.000000000	0.599766914	0.727849653	0.255636190	0.405540931
Int26	**1.6037643e+007**	0.332263155	0.174053371	0.111323507	0.359527883	0.499741156	1.000000000	1.000000000	0.599766914	0.751373921	0.319624690	0.409137603
Int27	**1.6038036e+007**	0.353350841	0.284010136	0.111323507	0.781610073	0.499741156	1.000000000	1.000000000	0.776113516	0.751373921	0.356099586	0.597128862
Int28	**1.6038089e+007**	0.356194728	0.298838911	0.132652149	0.781610073	0.499741156	1.000000000	1.000000000	0.799895628	0.751373921	0.366956717	0.603149257
Int29	**1.6038312e+007**	0.368160515	0.361231681	0.222393416	0.781610073	1.000000000	1.000000000	1.000000000	0.799895628	0.751373921	0.428180487	0.628480350
Int30	**1.6038422e+007**	0.374062920	0.392008383	0.222393416	0.781610073	1.000000000	1.000000000	1.000000000	0.799895628	1.000000000	0.431228600	0.678988443
Int31	**1.6038526e+007**	0.539882200	0.392008383	0.332804110	0.781610073	1.000000000	1.000000000	1.000000000	0.849093866	1.000000000	0.554738126	0.678988443
Int32	**1.6038671e+007**	0.579012453	0.406847326	0.332804110	0.781610073	1.000000000	1.000000000	1.000000000	0.917687563	1.000000000	0.584897473	0.685012966
Int33	**1.6038845e+007**	0.625968756	0.424654058	0.332804110	0.781610073	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.621088689	0.692242393
Int34	**1.6039080e+007**	0.689386752	0.448703380	0.332804110	0.781610073	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.653838932	0.702006274
Int35	**1.6039219e+007**	0.726897822	0.462928298	0.388340507	0.792541468	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.688672286	0.711494076
Int36	**1.6039387e+007**	0.772234943	0.480121004	0.455463634	0.805753513	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.730773031	0.722961347
Int37	**1.6039448e+007**	0.788696635	0.486363594	0.479835721	0.810550744	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.746059611	0.727125059
Int38	**1.6039537e+007**	0.812714514	0.495471635	0.515394997	0.817549982	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.768362982	0.733199982
Int39	**1.6039689e+007**	0.853733814	0.511026941	0.576125445	0.829503737	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.806454132	0.743575133

S27 (continued). Cumulative crossover fractions grouped by TD-SNPs - SNP 1&2

		d06AB	d06BA	d30AB	d30BA	d37-AB	d37-BA	d52-BA	d56-AB	d56-BA	sumTC	sumCT
Int40	1.6039763e+007	0.873703736	0.518599919	0.605691584	0.835323328	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.824998508	0.748626192
Int41	1.6040231e+007	1.000000000	0.566493887	0.792677437	0.872128311	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.942279154	0.780570734
Int42	1.6040474e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int43	1.6040553e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int44	1.6040565e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int45	1.6040712e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int46	1.6040725e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int47	1.6040779e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int48	1.6040921e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int49	1.6040949e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int50	1.6041044e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int51	1.6041120e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int52	1.6041236e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int53	1.6041251e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int54	1.6041278e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int55	1.6041353e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int56	1.6041412e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int57	1.6041424e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int58	1.6041578e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int59	1.6041633e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int60	1.6041776e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int61	1.6041780e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int62	1.6041823e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int63	1.6041858e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int64	1.6041885e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int65	1.6042053e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int66	1.6042084e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int67	1.6042085e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int68	1.6042112e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int69	1.6042209e+007	1.000000000	0.827302362	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.892947743
Int70	1.6042424e+007	1.000000000	0.886136822	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.916834184
Int71	1.6042527e+007	1.000000000	0.914322634	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.928277456
Int72	1.6042550e+007	1.000000000	0.914322634	0.889766246	0.891238591	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.928277456
Int73	1.6042612e+007	1.000000000	0.914322634	0.889766246	0.949875177	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	0.948191879
Int74	1.6042665e+007	1.000000000	1.000000000	0.889766246	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	1.000000000
Int75	1.6042739e+007	1.000000000	1.000000000	0.889766246	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	1.000000000
Int76	1.6042767e+007	1.000000000	1.000000000	0.889766246	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	1.000000000
Int77	1.6042943e+007	1.000000000	1.000000000	0.889766246	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.969309729	1.000000000
Int78	1.6043032e+007	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000

S28. Cumulative crossover fractions grouped by TD-SNPs - SNP 3

		d06AB	d06BA	d30AB	d30BA	sumTC	sumCT
Int1	**1.6032687e+007**	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000
Int2	**1.6032775e+007**	0.041383161	0.000000000	0.000000000	0.000000000	0.027174043	0.000000000
Int3	**1.6033410e+007**	0.041383161	0.000000000	0.000000000	0.000000000	0.027174043	0.000000000
Int4	**1.6033468e+007**	0.041383161	0.020750194	0.000000000	0.026077932	0.036128033	0.011141144
Int5	**1.6033653e+007**	0.041383161	0.086936158	0.000000000	0.109257541	0.064688173	0.046677553
Int6	**1.6033979e+007**	0.166559456	0.086936158	0.000000000	0.109257541	0.146884554	0.046677553
Int7	**1.6034366e+007**	0.188962326	0.086936158	0.000000000	0.109257541	0.161595285	0.046677553
Int8	**1.6034701e+007**	0.208354992	0.086936158	0.000000000	0.109257541	0.174329381	0.046677553
Int9	**1.6034810e+007**	0.208354992	0.086936158	0.016222276	0.127503706	0.180594294	0.054189804
Int10	**1.6035299e+007**	0.208354992	0.086936158	0.088999275	0.209360356	0.208700188	0.087891551
Int11	**1.6035449e+007**	0.208354992	0.086936158	0.111323507	0.234469757	0.217321629	0.098229511
Int12	**1.6035528e+007**	0.208354992	0.086936158	0.111323507	0.234469757	0.217321629	0.098229511
Int13	**1.6035615e+007**	0.208354992	0.086936158	0.111323507	0.234469757	0.217321629	0.098229511
Int14	**1.6035799e+007**	0.208354992	0.086936158	0.111323507	0.234469757	0.217321629	0.098229511
Int15	**1.6036036e+007**	0.208354992	0.086936158	0.111323507	0.234469757	0.217321629	0.098229511
Int16	**1.6036529e+007**	0.208354992	0.174053371	0.111323507	0.234469757	0.217321629	0.145004275
Int17	**1.6036544e+007**	0.208354992	0.174053371	0.111323507	0.234469757	0.217321629	0.145004275
Int18	**1.6036614e+007**	0.208354992	0.174053371	0.111323507	0.256973790	0.225048502	0.145004275
Int19	**1.6036933e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.260260969	0.145004275
Int20	**1.6037111e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.260260969	0.145004275
Int21	**1.6037287e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.260260969	0.145004275
Int22	**1.6037387e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.260260969	0.145004275
Int23	**1.6037392e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.260260969	0.145004275
Int24	**1.6037449e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.260260969	0.145004275
Int25	**1.6037568e+007**	0.208354992	0.174053371	0.111323507	0.359527883	0.260260969	0.145004275
Int26	**1.6037643e+007**	0.332263155	0.174053371	0.111323507	0.359527883	0.341624637	0.145004275
Int27	**1.6038036e+007**	0.353350841	0.284010136	0.111323507	0.781610073	0.500395816	0.204041997
Int28	**1.6038089e+007**	0.356194728	0.298838911	0.132652149	0.781610073	0.502263240	0.221880746
Int29	**1.6038312e+007**	0.368160515	0.361231681	0.222393416	0.781610073	0.510120513	0.296938123
Int30	**1.6038422e+007**	0.374062920	0.392008383	0.222393416	0.781610073	0.513996298	0.313462676
Int31	**1.6038526e+007**	0.539882200	0.392008383	0.332804110	0.781610073	0.622880689	0.364591926
Int32	**1.6038671e+007**	0.579012453	0.406847326	0.332804110	0.781610073	0.648575371	0.372559216
Int33	**1.6038845e+007**	0.625968756	0.424654058	0.332804110	0.781610073	0.679408990	0.382119963
Int34	**1.6039080e+007**	0.689386752	0.448703380	0.332804110	0.781610073	0.721052096	0.395032467
Int35	**1.6039219e+007**	0.726897822	0.462928298	0.388340507	0.792541468	0.749436900	0.428388004
Int36	**1.6039387e+007**	0.772234943	0.480121004	0.455463634	0.805753513	0.783743713	0.468702610
Int37	**1.6039448e+007**	0.788696635	0.486363594	0.479835721	0.810550744	0.796200354	0.483340651
Int38	**1.6039537e+007**	0.812714514	0.495471635	0.515394997	0.817549982	0.814374797	0.504697793
Int39	**1.6039689e+007**	0.853733814	0.511026941	0.576125445	0.829503737	0.845414294	0.541172913

S28 (continued). Cumulative crossover fractions grouped by TD-SNPs - SNP 3

		d06AB	d06BA	d30AB	d30BA	sumTC	sumCT
Int40	**1.6039763e+007**	0.873703736	0.518599919	0.605691584	0.835323328	0.860525629	0.558930536
Int41	**1.6040231e+007**	1.000000000	0.566493887	0.792677437	0.872128311	0.956094609	0.671235509
Int42	**1.6040474e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int43	**1.6040553e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int44	**1.6040565e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int45	**1.6040712e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int46	**1.6040725e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int47	**1.6040779e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int48	**1.6040921e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int49	**1.6040949e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int50	**1.6041044e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int51	**1.6041120e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int52	**1.6041236e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int53	**1.6041251e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int54	**1.6041278e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int55	**1.6041353e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int56	**1.6041412e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int57	**1.6041424e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int58	**1.6041578e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int59	**1.6041633e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int60	**1.6041776e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int61	**1.6041780e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int62	**1.6041823e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int63	**1.6041858e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int64	**1.6041885e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int65	**1.6042053e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int66	**1.6042084e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int67	**1.6042085e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int68	**1.6042112e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int69	**1.6042209e+007**	1.000000000	0.827302362	0.889766246	0.891238591	0.962656220	0.856228287
Int70	**1.6042424e+007**	1.000000000	0.886136822	0.889766246	0.891238591	0.962656220	0.887817545
Int71	**1.6042527e+007**	1.000000000	0.914322634	0.889766246	0.891238591	0.962656220	0.902951003
Int72	**1.6042550e+007**	1.000000000	0.914322634	0.889766246	0.891238591	0.962656220	0.902951003
Int73	**1.6042612e+007**	1.000000000	0.914322634	0.889766246	0.949875177	0.982789388	0.902951003
Int74	**1.6042665e+007**	1.000000000	1.000000000	0.889766246	1.000000000	1.000000000	0.948952688
Int75	**1.6042739e+007**	1.000000000	1.000000000	0.889766246	1.000000000	1.000000000	0.948952688
Int76	**1.6042767e+007**	1.000000000	1.000000000	0.889766246	1.000000000	1.000000000	0.948952688
Int77	**1.6042943e+007**	1.000000000	1.000000000	0.889766246	1.000000000	1.000000000	0.948952688
Int78	**1.6043032e+007**	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000

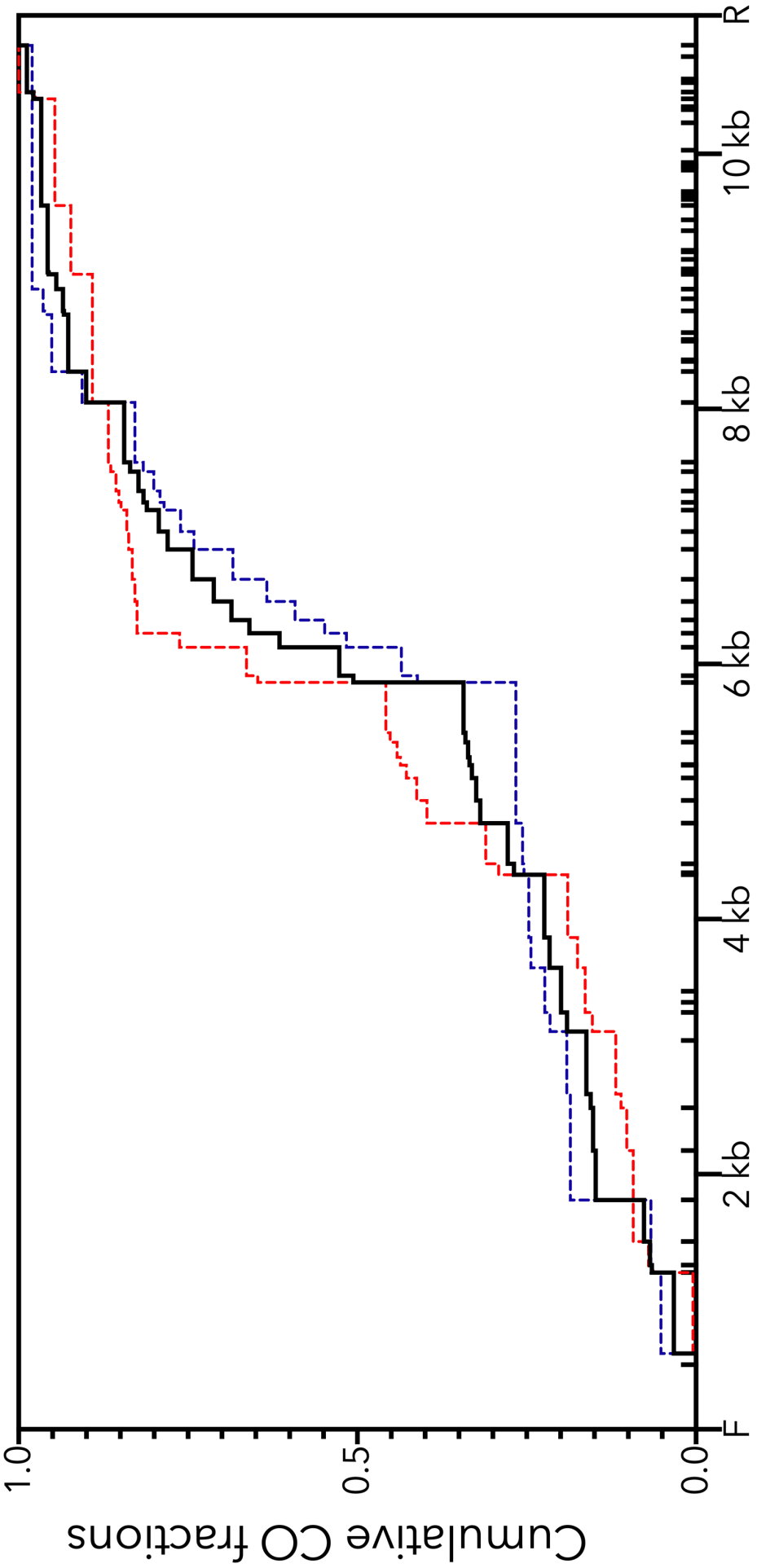
S29. Cumulative crossover fractions grouped by TD-SNPs - SNP 4&5

		d06AB	d06BA	d07AB	d07BA	d37-AB	d37-BA	d52-BA	d56-AB	d56-BA	sumTG	sumGT
Int1	**1.6032687e+007**	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000
Int2	**1.6032775e+007**	0.041383161	0.000000000	0.030544838	0.049383483	0.000000000	0.000000000	0.000000000	0.000000000	0.000000000	0.029736247	0.008072382
Int3	**1.6033410e+007**	0.041383161	0.000000000	0.250953615	0.405730211	0.000000000	0.000000000	0.000000000	0.000000000	0.222659414	0.058139657	0.109446045
Int4	**1.6033468e+007**	0.041383161	0.020750194	0.250953615	0.405730211	0.000000000	0.000000000	0.000000000	0.000000000	0.222659414	0.058139657	0.120117803
Int5	**1.6033653e+007**	0.041383161	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	0.000000000	0.000000000	0.222659414	0.058139657	0.154157030
Int6	**1.6033979e+007**	0.166559456	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	0.000000000	0.199964046	0.222659414	0.171203563	0.154157030
Int7	**1.6034366e+007**	0.188962326	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	0.053601108	0.199964046	0.222659414	0.185170467	0.170776156
Int8	**1.6034701e+007**	0.208354992	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	1.000000000	0.199964046	0.222659414	0.197260681	0.185162221
Int9	**1.6034810e+007**	0.208354992	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	1.000000000	0.229153552	0.222659414	0.202373218	0.185162221
Int10	**1.6035299e+007**	0.208354992	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	1.000000000	0.360104639	0.222659414	0.225309279	0.185162221
Int11	**1.6035449e+007**	0.208354992	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	1.000000000	0.400273684	0.222659414	0.232344880	0.185162221
Int12	**1.6035528e+007**	0.208354992	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	1.000000000	0.400273684	0.222659414	0.232344880	0.185162221
Int13	**1.6035615e+007**	0.208354992	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	1.000000000	0.400273684	0.222659414	0.232344880	0.185162221
Int14	**1.6035799e+007**	0.208354992	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	1.000000000	0.599766914	0.222659414	0.267286086	0.185162221
Int15	**1.6036036e+007**	0.208354992	0.086936158	0.250953615	0.405730211	0.000000000	0.000000000	1.000000000	0.599766914	0.222659414	0.267286086	0.185162221
Int16	**1.6036529e+007**	0.208354992	0.174053371	0.250953615	0.405730211	0.000000000	0.000000000	1.000000000	0.599766914	0.528677519	0.267286086	0.289235096
Int17	**1.6036544e+007**	0.208354992	0.174053371	0.250953615	0.405730211	0.000000000	0.000000000	1.000000000	0.599766914	0.528677519	0.267286086	0.289235096
Int18	**1.6036614e+007**	0.208354992	0.174053371	0.250953615	0.405730211	0.000000000	0.179948586	1.000000000	0.599766914	0.528677519	0.267286086	0.306790429
Int19	**1.6036933e+007**	0.208354992	0.174053371	0.250953615	0.405730211	0.000000000	1.000000000	1.000000000	0.599766914	0.528677519	0.267286086	0.386792592
Int20	**1.6037111e+007**	0.208354992	0.174053371	0.250953615	0.405730211	0.000000000	1.000000000	1.000000000	0.599766914	0.584508448	0.267286086	0.397605778
Int21	**1.6037287e+007**	0.208354992	0.174053371	0.250953615	0.405730211	0.000000000	1.000000000	1.000000000	0.599766914	0.639712063	0.267286086	0.408297468
Int22	**1.6037387e+007**	0.208354992	0.174053371	0.250953615	0.405730211	0.000000000	1.000000000	1.000000000	0.599766914	0.671077753	0.267286086	0.414372291
Int23	**1.6037392e+007**	0.208354992	0.174053371	0.250953615	0.405730211	0.013805004	1.000000000	1.000000000	0.599766914	0.672646038	0.268287509	0.414676033
Int24	**1.6037449e+007**	0.208354992	0.174053371	0.250953615	0.405730211	0.171182053	1.000000000	1.000000000	0.599766914	0.690524481	0.279703733	0.418138682
Int25	**1.6037568e+007**	0.208354992	0.174053371	0.250953615	0.405730211	0.499741156	1.000000000	1.000000000	0.599766914	0.727849653	0.303537603	0.425367722
Int26	**1.6037643e+007**	0.332263155	0.174053371	0.250953615	0.405730211	0.499741156	1.000000000	1.000000000	0.599766914	0.751373921	0.380787232	0.429923840
Int27	**1.6038036e+007**	0.353350841	0.284010136	0.471895854	0.405730211	0.499741156	1.000000000	1.000000000	0.776113516	0.751373921	0.453293428	0.486474248
Int28	**1.6038089e+007**	0.356194728	0.298838911	0.501692136	0.405730211	0.499741156	1.000000000	1.000000000	0.799895628	0.751373921	0.463071618	0.494100639
Int29	**1.6038312e+007**	0.368160515	0.361231681	0.501692136	0.702347107	1.000000000	1.000000000	1.000000000	0.799895628	0.751373921	0.506820669	0.574674987
Int30	**1.6038422e+007**	0.374062920	0.392008383	0.501692136	0.702347107	1.000000000	1.000000000	1.000000000	0.799895628	1.000000000	0.510500480	0.638656585
Int31	**1.6038526e+007**	0.539882200	0.392008383	0.501692136	0.702347107	1.000000000	1.000000000	1.000000000	0.849093866	1.000000000	0.622496349	0.638656585
Int32	**1.6038671e+007**	0.579012453	0.406847326	0.501692136	0.702347107	1.000000000	1.000000000	1.000000000	0.917687563	1.000000000	0.658905990	0.646288206
Int33	**1.6038845e+007**	0.625968756	0.424654058	0.501692136	0.702347107	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.702597560	0.655446150
Int34	**1.6039080e+007**	0.689386752	0.448703380	0.501692136	0.702347107	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.742135041	0.667814638
Int35	**1.6039219e+007**	0.726897822	0.462928298	0.501692136	0.702347107	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.765521041	0.675130468
Int36	**1.6039387e+007**	0.772234943	0.480121004	0.501692136	0.751759840	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.793786133	0.692049785
Int37	**1.6039448e+007**	0.788696635	0.486363594	0.501692136	0.769701368	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.804049054	0.698193108
Int38	**1.6039537e+007**	0.812714514	0.495471635	0.501692136	0.795878352	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.819022824	0.707156318
Int39	**1.6039689e+007**	0.853733814	0.511026941	0.501692136	0.840585110	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.844596003	0.722464271

S29 (continued). Cumulative crossover fractions grouped by TD-SNPs - SNP 4&5

		d06AB	d06BA	d07AB	d07BA	d37-AB	d37-BA	d52-BA	d56-AB	d56-BA	sumTG	sumGT
Int40	**1.6039763e+007**	0.873703736	0.518599919	0.501692136	0.862350243	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.857046103	0.729916827
Int41	**1.6040231e+007**	1.000000000	0.566493887	0.501692136	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.935784576	0.777049210
Int42	**1.6040474e+007**	1.000000000	0.827302362	0.751962930	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.968036215	0.911182164
Int43	**1.6040553e+007**	1.000000000	0.827302362	0.751962930	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.968036215	0.911182164
Int44	**1.6040565e+007**	1.000000000	0.827302362	0.751962930	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.968036215	0.911182164
Int45	**1.6040712e+007**	1.000000000	0.827302362	0.751962930	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.968036215	0.911182164
Int46	**1.6040725e+007**	1.000000000	0.827302362	0.751962930	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.968036215	0.911182164
Int47	**1.6040779e+007**	1.000000000	0.827302362	0.751962930	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.968036215	0.911182164
Int48	**1.6040921e+007**	1.000000000	0.827302362	0.751962930	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.968036215	0.911182164
Int49	**1.6040949e+007**	1.000000000	0.827302362	0.751962930	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.968036215	0.911182164
Int50	**1.6041044e+007**	1.000000000	0.827302362	0.751962930	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.968036215	0.911182164
Int51	**1.6041120e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int52	**1.6041236e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int53	**1.6041251e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int54	**1.6041278e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int55	**1.6041353e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int56	**1.6041412e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int57	**1.6041424e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int58	**1.6041578e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int59	**1.6041633e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int60	**1.6041776e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int61	**1.6041780e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int62	**1.6041823e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int63	**1.6041858e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int64	**1.6041885e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int65	**1.6042053e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int66	**1.6042084e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int67	**1.6042085e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int68	**1.6042112e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int69	**1.6042209e+007**	1.000000000	0.827302362	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.911182164
Int70	**1.6042424e+007**	1.000000000	0.886136822	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.941440537
Int71	**1.6042527e+007**	1.000000000	0.914322634	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.955936408
Int72	**1.6042550e+007**	1.000000000	0.914322634	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.955936408
Int73	**1.6042612e+007**	1.000000000	0.914322634	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	0.955936408
Int74	**1.6042665e+007**	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000
Int75	**1.6042739e+007**	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000
Int76	**1.6042767e+007**	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000
Int77	**1.6042943e+007**	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000
Int78	**1.6043032e+007**	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000	1.000000000

S30. Cumulative Crossover fractions of all individuals combined, with complex crossovers excluded. Blue dotted line refers to A→B orientation, Red dotted line to B→A orientation and black solid line represents both orientations combined.

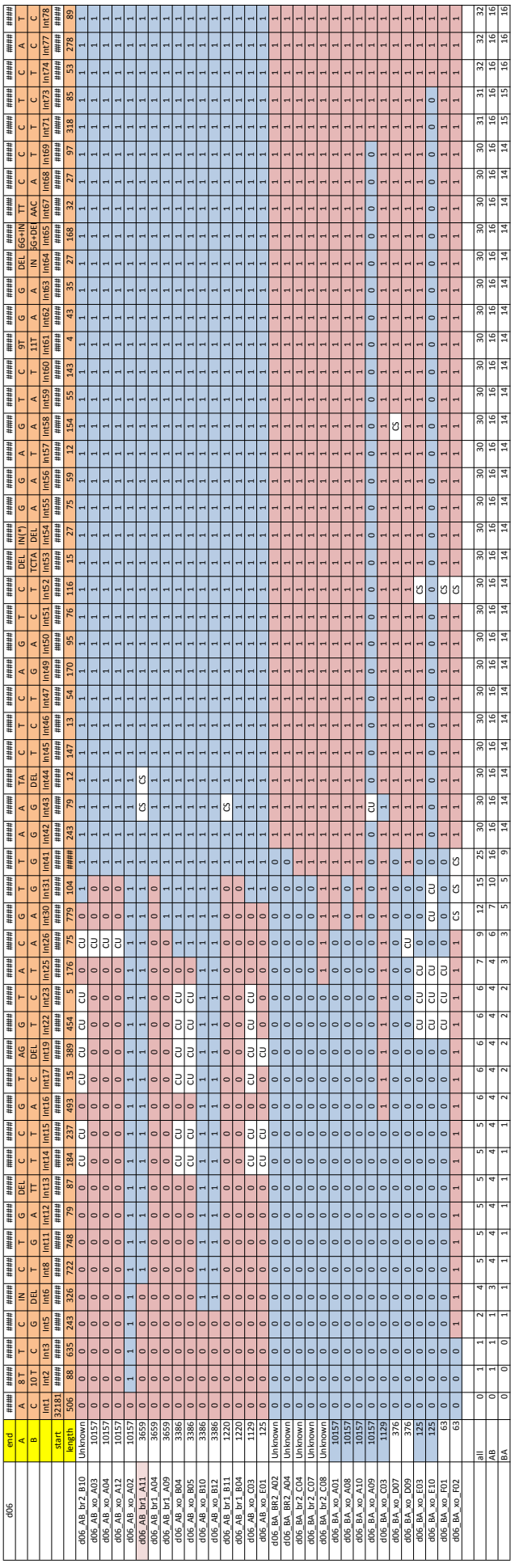


S31. Sanger-Sequencing results. Columns one to 3 represent identifiers for individual PCR reactions (consistent with identifiers in S32-S38). Positions are indicated as chr27:160xxxxx. Haplotypes (A and B) refer to individual haplotypes from S4). “NS” and “SE” refer to insufficient sequencing quality to make a call. Blank squares indicate that the individual does not carry the marker or the marker was not covered in the sequencing assay.

Direction	Dog	well (XO plate)	32687	32775	33410	33468	33653	33979	35163	35299	35449	35528	35,615-16	36529	36544	36933	38036	38089	38312	38422	39080	39219	40474	40553	40,712&725	40779	40,921&949	41044	41120	41236	41251	42424	42,527&550	42665	42767	42943	43032
brAB	d06	A2	A	A																									A				A	A				
brAB	d06	A3	A	A	NS	NS	NS	A	A	A	A	A	NS	A	A		NS	NS	NS	NS	A	NS		B	B	B	B	NS	NS	NS	NS	B	B	B			
brAB	d06	A4	A	A	NS	NS	NS	A	A	A	A	A	NS	A	A	NS	NS	NS	NS	NS	A	B	N	NS	B	B	B	B	B	B	B	B						
brAB	d06	A9	A	NS	NS	NS	NS	A	A	A	A	NS	NS	A	A	NS	NS	NS	NS	NS	A	B	N	B	B	B	B	B	B	B	NS	B	B					
brAB	d06	B4						A	A	A	A	A	NS	NS	NS	NS	NS					NS	N	B	B														
brAB	d06	B8						A		A	A	A	A	A	A	A	NS	NS	NS	NS		B	N	B	B	B	B	NS	NS	NS	NS	B	B	B				
brAB	d06	B11	A	A	NS	NS	NS	A	A	A	A	A	A	A	A	A	NS	NS	NS	NS		B	N	B	B	B	B	B	B	B	B	B	B					
brAB	d06	B12						B															N															
brBA	d06	A2						B															N										M	M				
brBA	d06	A3	B	N	NS	NS	NS	NS	NS	NS	NS	NS	NS	B	B	SE	NS	NS	NS	NS		B	N	NS	A	A	A	A	A	A	A	SE	A	A				
brBA	d06	A4	B	N	NS	NS	NS	NS			B	B	B	B	B	SE	A		A	SE	NS	NS	N	NS	NS	NS	NS	NS	NS	NS	NS	NS	B	B				
brBA	d06	A5	B		NS	NS	NS	NS			B	B	B	B	B		NS	NS	NS	NS		B	N	NS	NS	NS	NS	NS	NS	NS	NS	NS		B	B	NS		
brBA	d06	A6	B																				N											B				
brBA	d06	A7												B									N											B				
brBA	d06	A8												B									N											M				
brBA	d06	A9	B																				N											M				
brBA	d06	A10	B																				N											B	B			
brBA	d06	A11	B																				N											B	B			
brBA	d06	A12	B																				N											B	B			
brBA	d06	B1	B											B									N															
brBA	d06	B2																					N											M	M			
brBA	d06	B3	B																				N											B	B			
brBA	d06	B4	B																				N											M	M			
brBA	d06	B5						B															N											B	B			
brBA	d06	B6	B																				N		B													
brBA	d06	B7	B		NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS		B	N	NS	A	A	A	A	A	A	A		A	A				
brBA	d06	B8	B	N	NS	NS	NS	NS	NS	NS	NS	NS	NS	A	A	A	A	A	A	A	NS	NS	N	NS	A	A	A	A	A	A	A		A	A				
brBA	d06	B9																					N											M				
brBA	d06	B10	B													B							N															
brBA	d06	B11												B									N											B	B			
brBA	d06	B12	B																				N											B	B			
br2AB (?)	d06	A1																					N															
br2AB (?)	d06	A3																					N															
br2AB (?)	d06	A6	B	NS	NS	NS	NS				B	B	B	B	B	NS						B	N	A	A	A	A	A	A	A			A	A	A			
br2AB (?)	d06	A9																					N															
br2BA	d06	A9																					N															
br2BA	d06	A10						B															N				B											
br2BA	d06	A12																					N															
br2BA	d06	B12	B																				N				B											
br2BA	d06	C6						B															N	B														
brAB	d07	A5	M																																			
brAB	d07	A6	A	A	NS	NS	NS	A	NS	NS	NS	NS	NS			A		A	A	A	A	NS	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
brAB	d07	A9	A	A	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS		A	A	A	A	NS	A	A	A	A	A	A	B	A	B	B	B	B			
brBA	d07	A1	B	B	NS	NS	NS	A	NS	NS	NS	NS	NS	SE	SE	NS		A	B	B	A	A	A	A	A	A	B&A	A	A	NS	NS	B	A	A	B	A	A
brBA	d07	A2																																				
brBA	d07	A3																																				
brBA	d07	A4																																				
brBA	d07	A7																																				
brBA	d07	A9																																				
brBA	d07	A10	B																													B						
brBA	d07	A11	B	B	NS	NS	NS	NS		B	B		B				NS	NS	NS	NS	B	B	NS	NS	B	B	B	B	B	B	B	B	B	B	B			
brBA	d07	A12	B																													B						
brBA	d07	B2	B																															B	B			
brBA	d07	B3	B	B	NS	NS	NS	NS		B	B		B				NS	NS	NS	NS	NS	NS	NS	NS	B	B	NS	NS	NS	NS	NS	B	B	B	B	B	B
brBA	d07	B4	B																													B	A	A				
brBA	d07	B5	B																		M																	
brBA	d07	B6						B																								A	A					
brBA	d07	B8	B																													B	B					
brBA	d07	B10	B	B	NS	NS	NS		B	B	B		B				NS	NS	NS	NS	A	A	NS	A	A	A	A	A	A	A		A	A	A	A	A	A
brBA	d07	B11																																				
brBA	d07	B12																																				
brAB	d07	F1																																				
brAB	d07	F5	B																												B							
brAB	d07	F11	B	B	NS	NS	NS		B		B	SE 1	B	SE 1	SE C			A	B	B	A	A	A	A	A	A	B&A	A	A	A	SE	B	A	A	B	A	A
brAB	d26	A1						A			NS	NS	NS	B	B	NS	NS	NS	NS	NS	B		B	B	B	B		B	B	NS	NS		B	B		B	B
brAB	d26	A2	B	B																														B	B				
brAB	d26	A3	A	A	NS	NS	NS	A			A	A	A	A	A						NS	A		NS	A	B	B		B	B		B		B	B		B	B
brAB	d26	A4	A	A	NS	NS	NS	A			A	A	A	A	A						NS	A		B	B	B	B		B	B		B		B	B		B	B
brAB	d26	A5 2	A	A	NS	NS	NS	A			A	B	A	A	A						NS	B		B	B	B	B		B	B		B		B	B		B	B
brAB	d26	A8	A	A	NS	NS	NS	A			A	A	A	A	A						NS	B		B	B	B	B		B	B		B		B	B		B	B
brAB	d26	A9	A	A	NS	NS	NS	A			A	B	A	B	B						NS	B		B	B	B	B	B	B	B		B		B	B		B	B
brAB	d26	A11	A	A	NS	NS	NS	A			A	A	A	A	A						NS	B		NS	NS	B	B	B	B	B		B		B	B		B	B
brAB	d26	A12	A	A	NS	NS	NS	A			A	A	A	A	A						NS	B		B	B	B	B	B	B	B		B		B	B		B	B
brAB	d26	B2	A	A	NS	NS	NS	A			A	A	A	A	A						NS	B		B	B	B	B	B	NS	NS		B		B	B		B	B
brAB	d26	B5	A	B	NS	NS	NS	A			A	A	NS	NS	NS						NS	B		B	B	B	B	B	B	B		B		NS	NS		NS	NS
brAB	d26	B6	NS	NS	NS	NS	NS	A																														
brAB	d26	B7	A	SE	NS	NS	NS	A																										A	A		A	
brAB	d26	B9	NS	NS	NS	NS	NS				B	B																						B	B			
brAB	d26	B11	A	A	NS	NS	NS	A			A	A	A	B	B						NS	B		B	B	B	B	B	B	B		B		B	B		B	B
brBA	d26	A1	B	B																														B	B				
brBA	d26	A2	B	B																														B	B		B		
brBA	d26	A3	B	B	NS	NS	NS	SE			B	B	B	A	A						NS	A		A	A	NS	NS		NS	NS	NS	NS		A	A		A	SE
brBA	d26	A4	B	B	NS	NS	NS	A			B	B	B	B	B						NS	B		B	B	B	B		B	B		SE		A	A		A	A
brBA	d26	A5									B	B																						B	B				
brBA	d26	A6	B	B	NS	NS	NS	SE			NS	NS	NS	B	B						NS	B		A	A	A&B	B							A&B	A		B	A
brBA	d26	A7									B	B	B	B	B						NS	B		B	B	B		B	B	B		NS		B	B		B	B
brBA	d26	A8	B	B																														B	B				
brBA	d26	A9	B	B	NS	NS	NS				B	B	B	A	A						NS	NS		NS	NS	NS	NS	NS	NS	NS	NS	NS		A	A		A	A
brBA	d26	A10	B	NS							NS	NS	NS	NS	B																			A	B		B		
brBA	d26	A11																																				
brBA	d26	A12									B	B																						B	NS				
brBA	d26	B1																																				
brBA	d26	B2	NS	NS							NS	NS	NS	NS	NS							NS		NS	NS	B	B	B	NS	NS	NS	NS		B	NS		NS	NS
brBA	d26	B4	B	B	NS	NS	NS				B	B	B	A	A						NS	NS		A	A	A	A	A	A	A	A			A	A		A	A
brBA	d26	B5												B	B							B																
brBA	d26	B6	NS	NS	NS	NS	NS				NS	NS	NS	NS	NS						NS	B		NS	NS	B	B	B			NS	NS		NS		NS	NS
brBA	d26	B8	B	B	NS	NS	NS				B	B	NS	NS	NS						NS	B		NS	NS	NS	NS	B	B	NS	NS	NS		NS	NS		NS	NS
brBA	d26	B9	NS	NS	NS	NS	NS				B	B	B																B	B								
brBA	d26	B11																																				
brBA	d26	B12																																				
brAB	d30	C3	A	A	NS	NS	NS	A			A	A	A	A	A	A	B		B	B	B		B	B	B	B		B	B		B		B	B		B	B

S31 (continued). Sanger-Sequencing results. Columns one to 3 represent identifiers for individual PCR reactions (consistent with identifiers in S32-S38). Positions are indicated as chr27:160xxxxx. Haplotypes (A and B) refer to individual haplotypes from S4). “NS” and “SE” refer to insufficient sequencing quality to make a call. Blank squares indicate that the individual does not carry the marker or the marker was not covered in the sequencing assay.

brAB	d30	C4	B		NS	NS	NS	A			B	A	A	A	A	NS	NS	NS	B	A	A		A	A	A&B	A		A	A				A	NS		NS	NS
brAB	d30	C6	NS	NS	NS	NS	NS	NS	NS		A	NS	NS	A	B		NS	NS	NS	NS	NS	NS	NS	NS	B	B	B	B	B		B		B	B		B	B
brAB	d30	C7	A		NS	NS	NS	A			A	A	A	A	A	A	A		A	A	B		B	B	B	B	B	B	B		B		B	B		B	B
brAB	d30	C9	A	NS	NS	NS	NS	NS			A	NS	NS	A	A		NS	NS	NS	NS	NS	NS	NS	NS	B	B	NS	NS	NS	NS	NS		B	B		NS	NS
brAB	d30	C10																																			
brAB	d30	C11																																			
brAB	d30	D4	A	NS	NS	NS	NS	A			NS	NS	NS	A	A	NS	NS	NS	NS	NS	A	NS	B	B	B	B	B	B	B	NS	NS		B	B		NS	NS
brAB	d30	D5	A		NS	NS	NS	A			A	B	A	A	A	NS	NS	NS	B	A	NS	NS	B	B	B	B	B	B	NS	NS	NS		A&B	A		B	B
brAB	d30	D6	A		NS	NS	NS	A			A	A	NS	A	A	A	NS	NS	NS	NS	A		B	B	B	B	B	B	B		B		B	B		B	B
brAB	d30	D7	A	A	NS	NS	NS	A			A	A	A	A	A		A		A	A	A		NS	NS	B	B	B	B	B		B		B	B		B	B
brAB	d30	D11	A	A	NS	NS	NS	A			B	B	B	B	B		B		B	B	B		B	B	B	B	B	B	B		B		B	B		B	B
brAB	d30	D12	B		NS	NS	NS	B			B	B	B	B	B	B	NS		NS	NS	NS		NS	NS	NS	NS	B	B	B		NS		B	B		B	B
brBA	d30	C2	B		NS	NS	NS	NS			B	B	B																				B	B			
brBA	d30	C3	B		NS	NS	NS	B			B	B	B													B	B						B	A		NS	NS
brBA	d30	C4	B		NS	NS	NS	B			B	A	A	A	A	NS	NS	NS	B	A	A		A	A		A	B	A	A				A	NS		NS	NS
brBA	d30	C5	NS	NS	NS	NS	NS	B			B	B	B	B	B		NS	NS	NS	NS	A		NS	A	A	A	A	A	A		SE		A	A		A	A
brBA	d30	C6	B		NS	NS	NS	B			B	B	B	B	B		B		B	B	NS		NS	NS	NS	NS	NS	NS	NS		NS		B	B		B	B
brBA	d30	C7	B	B	NS	NS	NS	B			B	NS	NS	B							B			B	NS	NS	NS	NS	NS		NS		NS	NS		NS	NS
brBA	d30	C8	NS	NS	NS	NS	NS	B				B	B	B	B	NS	NS	NS	B	B	A	NS	A	A	A	M	A	NS	NS	NS	NS		A	A		A	A
brBA	d30	C8 unten																						A													
brBA	d30	C9	NS	NS	NS	NS	NS	B			B																						B	B		B	B
brBA	d30	C10	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	B	B																		B	B		B	B
brBA	d30	C11	B	B	NS	NS	NS	B			B	B	B	B	B	A	A		A	A	A		A	A	A	A	A	A	A		A		A	A		A	A
brBA	d30	C12	B	B	NS	NS	NS	B																									B	B		B	
brBA	d30	D1	NS	NS	NS	NS	NS	B			NS	NS	NS	NS	NS	NS	NS	NS	NS	B	NS		B										B	B			
brBA	d30	D2																		B			B	M													
brBA	d30	D3	B	B	NS	NS	NS	B			B	B	B	B	B	B	B		B	B	NS		B	B	B	A	A	A	A		A		A	A		A	A
brBA	d30	D4	B	B	NS	NS	NS	B			B	B	B	B	B	B	B		B	B	NS		B	B	B	B	B	B	B		B		B&A	A		A	A
brBA	d30	D5	B	B	NS	NS	NS	B			NS	NS	NS	B	B	B	A		A	A	NS		A	A	A	A	A	A	A		A		A	A		A	A
brBA	d30	D6	B	B	NS	NS	NS	A			A	A	A	A	A	A	A		A	A			A	A	A	A	A	A	A		A		A	A		A	A
brBA	d30	D7	B	B	NS	NS	NS																										B	B		B	
brBA	d30	D8	B	B	NS	NS	NS	B																									B	B			
brBA	d30	D9	B	B	NS	NS	NS	B			B	B	B	B	B	B	B		B	B			B	B	B	B	B	B	B		B		A	NS	NS	NS	NS
brBA	d30	D10	B	B	NS	NS	NS	B																				B	B		B						
brBA	d30	D11	B	B	NS	NS	NS	B																				B	B		B		B	NS			
brBA	d30	D12	NS	NS	NS	NS	NS	B			NS	NS	NS	NS	B	NS	NS		NS	NS			B	B	A&B	B	B	B	B		B		B	B		B	
xoBA	d30	A1						B						B	B										B	B	B	B	B		B		NS	NS		NS	
xoBA	d30	A2	NS	NS	NS	NS	NS	B			NS	NS	NS	NS	NS	NS	NS		B	A	A		A	A	A	A	A	A	A		NS		A	A		A	A
xoBA	d30	A10	B	B	NS	NS	NS	B			NS	NS	NS	SE	B	B	NS		NS																		
xoBA	d30	A11						B																													
xoBA	d30	A12						B													B		A	A	M												
xoBA	d30	B2															B		B	M	NS	NS	NS	NS	NS	B	B	B	SE		NS						
brAB	d37	C1	B	B	NS	NS	NS	B	NS	NS	NS	NS	NS	B																			B	B	B		
brAB	d37	C9	B	B	NS	NS	NS	B			B	A	B	A	A	A	NS	NS	NS	NS	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
xoBA	d52	A1						A			M	M	A																				M	M	M		
xoBA	d52	A3	B	B				B			NS	NS	NS			A	NS	NS	NS	NS	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	NS	
xoBA	d52	A4	B	B				B			B														M	M	M	M	B				M				
xoBA	d52	A6	B	B				B																B	NS	NS	NS	NS	NS	NS	NS	B					
xoBA	d52	A7	B	B				B			B	B	B	B				A			A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
xoBA	d52	A8	B	B				B			B	NS	NS	NS	NS	NS	NS	NS			NS	NS	NS	NS	B	B	B	B									
xoBA	d52	A9						B			NS	NS	NS	B		NS	NS	NS			B	B	B	B	B	B	B	B	B	B	B	NS					
xoBA	d52	A10	B	B														B			B	B	NS	NS	NS												
xoBA	d52	A12	B	B				B			NS	NS	NS	NS	NS	NS	NS	B			B	B	NS	NS	B	B	B	B	B	B	B	NS	NS				
brAB	d56	A1								NS	NS	NS	NS	B	B	B	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	B	B	B	NS		
brAB	d56	A2	NS	NS	NS	NS	NS	SE	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	A	M	NS	NS	NS	NS	NS	NS	B	NS	NS	NS	NS	NS
brAB	d56	A3	NS	NS	NS	NS	NS	B	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	B	B	NS	NS	NS	B	B	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
brAB	d56	A4	NS	NS	NS	NS	NS	B		NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	B	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
brAB	d56	A6	NS	NS	NS	NS	NS	A			A	M	NS	NS	NS	NS	NS	NS	NS	NS	M	NS	NS	NS	NS	M	M	NS	NS	NS	NS	A	A	A			
brAB	d56	A10						A			A	A	A	A	A	A	NS	NS	NS	NS	B	B	NS	NS		B	B	B	B	B	B	B	B	B	B	B	B
brAB	d56	A11												B							B	B			B	B	B	B	B	B	B						
brAB	d56	A12	B	B	NS	NS	NS	B																									B	B			
brAB	d56	B6	NS	NS	NS	NS	NS	A			NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	B	B	B	B	B	NS	NS	NS	NS	B	B	B	B		
brBA	d56	A1	NS	NS	NS	NS	NS	NS			B	B	B																				B	B	B		
brBA	d56	A2	B	B	NS	NS	NS	B			B	B	B	A	A	SE	NS	NS	NS	NS	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
brBA	d56	A3	NS	NS	NS	NS	NS	B			NS	NS	NS	B	B																		B	B	B		
brBA	d56	A4						B			B	B	B																				B	B	B		
brBA	d56	A5	B	B	NS	NS	NS	B			B	B																					B	B	B	B	
brBA	d56	A6	B	B	NS	NS	NS	B			B	B	B	B	B	A	NS	NS	NS	NS	A	A	A	M	A	A	A	A	A	A	A	A	A	A	A	A	A
brBA	d56	A7	B	B	NS	NS	NS	B			B																						B	M	M		
brBA	d56	A8	B	B	NS	NS	NS	B			A	M	A	A	A	SE	NS	NS	NS	NS	A	M	A	A	A&B	M	A&B	A	A			A	A&B	A	B		
brBA	d56	A9	B	B	NS	NS	NS	NS			NS	NS	NS	B	B	NS	NS	NS	NS	NS	NS	NS	NS	B	NS	NS	NS	NS	NS	NS	NS	M	M	B	M		
brBA	d56	A10	B	B	NS	NS	NS	B			B	NS	NS	B	B	NS	NS	NS	NS	NS	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
brBA	d56	A11						B			B	B																					B	NS			
brBA	d56	A12	B	B	NS	NS	NS	B			B	B																					B	B	B		
brBA	d56	B2												B	B																						
brBA	d56	B3	B	B	NS	NS	NS	B			B																						B	B	NS		
brBA	d56	B4						B			NS	NS	NS	B		NS	NS	NS	NS	NS	NS	NS	B	B	B	B	B	NS	NS	NS	NS	B	B				
brBA	d56	B6									B	B	B											B	M	M											
brBA	d56	B7	B	B				NS			NS	NS	NS		B	B							B	B	M												
brBA	d56	B8	B	B	NS	NS	NS	B			B																						B	B			
brBA	d56	B9	B	B	NS	NS	NS	B			B																						B	B	B	B	
brBA	d56	B10						B			B														B	B	B										
brBA	d56	B11	A					A			A	A																					A	A			
brBA	d56	B12						B						B	B	B																	B	B			



Markers are shown by Interval identifier (Int1-Int78) and only the ones heterozygous in the respective individual are shown. “CU” refers to a haplotype switch of one or a few bases before the final crossover switch (defined as the first consecutive 2 kilobases switched to the other haplotype). “CS” refers to an additional haplotype switch of on or a few bases downstream of the final haplotype switch. Column 2 shows the respective number of input molecules for the pool from which the recombinant was recovered. For individual d07 (S33) no final haplotype switch could be detected in some recombinants so these were excluded from analysis.

S33. MinION sequencing results d07

d07	end	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
	A	A	T	C	IN	T	C	T	T	DEL	A	AG	A	A	A	C	T	A	C	G	C	T	A
	B	C	C	G	DEL	G	T	G	G	TT	C	DEL	G	T	C	A	C	G	A	A	T	G	G
		Int1	Int3	Int5	Int6	Int7	Int8	Int10	Int11	Int13	Int18	Int19	Int20	Int21	Int26	Int28	Int29	Int30	Int32	Int34	Int35	Int41	Int42
	start	####	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
	length	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
d07_AB_br1_A05	5952	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
d07_AB_br1_A06	5952	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
d07_AB_br1_A09	5952	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d07_AB_xo_A01	2031	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
excl complex																							
d07_BA_br1_A07	5952	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
excl complex																							
d07_BA_br1_B10	1984	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
d07_BA_xo_A03	2031	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
d07_AB_br2_F11		0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	1	0	1
d07_BA_br1_A01	5952	0	0	0	0	0	0	CU	0	0	0	0	CU	0	0	CU	0	0	0	CU	CU	0	CU
d07_BA_br1_B02	1984	0	0	0	0	0	0	CU	0	0	0	0	CU	0	0	CU	0	0	0	CU	CU	0	CU
	all	0	2	2	2	2	2	2	2	2	2	2	2	2	2	3	4	4	4	4	4	5	6
	AB	0	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3
	BA	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	3	3

d07	end	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
	A	A	TA	C	T	C	A	G	T	C	DEL	IN(*)	G	G	A	G	T	C	G	G	1T	6G+IN	TT	C	C	G	C	G	C	C	A	T
	B	G	DEL	T	C	T	G	A	C	T	TCTA	DEL	A	A	T	A	A	T	A	A	9T	G+DE	AAC	A	T	A	T	A	T	T	C	C
		Int43	Int44	Int45	Int46	Int47	Int49	Int50	Int51	Int52	Int53	Int54	Int55	Int56	Int57	Int58	Int59	Int60	Int62	Int63	Int64	Int65	Int67	Int68	Int69	Int70	Int71	Int72	Int74	Int76	Int77	Int78
	start	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
	length	79	12	###	13	54	###	95	76	###	15	27	75	59	12	###	55	###	47	35	27	###	32	27	97	###	###	23	###	###	###	89
d07_AB_br1_A05	5952	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d07_AB_br1_A06	5952	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d07_AB_br1_A09	5952	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d07_AB_xo_A01	2031	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
excl complex																																
d07_BA_br1_A07	5952	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
excl complex																																
d07_BA_br1_B10	1984	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d07_BA_xo_A03	2031	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d07_AB_br2_F11		1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1
d07_BA_br1_A01	5952	CU	CU	CU	CU	CU	CU	CU	CU	0	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	0	CU	CU	CU	0	1	1
d07_BA_br1_B02	1984	CU	CU	CU	CU	CU	CU	CU	CU	0	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	CU	0	CU	CU	CU	0	1	1
	all	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
	AB	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	BA	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Markers are shown by Interval identifier (Int1-Int78) and only the ones heterozygous in the respective individual are shown. “CU” refers to a haplotype switch of one or a few bases before the final crossover switch (defined as the first consecutive 2 kilobases switched to the other haplotype). “CS” refers to an additional haplotype switch of on or a few bases downstream of the final haplotype switch. Column 2 shows the respective number of input molecules for the pool from which the recombinant was recovered. For individual d07 (S33) no final haplotype switch could be detected in some recombinants so these were excluded from analysis.

S34. MinION sequencing results d26

d26	end	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
	A	A	8T	T	G	C	IN	T	C	T	G	DEL	G	T	A	A	G	G	G	3T	C	A	A	TA	C	T	C	A	G	T	DEL	IN(*)	G	G	A	G	T	C	G	G	1T	G+A+I	DEL	TT	C	C	C	G	C	T	A	T
	B	C	10T	C	A	G	DEL	G	T	G	A	TT	A	C	C	G	A	A	C	6T	T	G	G	DEL	T	C	T	G	A	C	TCTA	DEL	A	A	T	A	A	T	A	A	9T	G+DE	G	AAC	A	T	T	A	T	C	C	C
		Int1	Int2	Int3	Int4	Int5	Int6	Int7	Int8	Int11	Int12	Int13	Int16	Int17	Int26	Int30	Int34	Int36	Int37	Int38	Int39	Int42	Int43	Int44	Int45	Int46	Int47	Int49	Int50	Int51	Int53	Int54	Int55	Int56	Int57	Int58	Int59	Int60	Int62	Int63	Int64	Int65	Int66	Int67	Int68	Int69	Int71	Int72	Int74	Int75	Int77	Int78
	start	####	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
	length	###	88	###	58	###	###	###	###	###	79	87	###	15	###	###	###	###	61	89	###	###	79	12	###	13	54	###	95	76	###	27	75	59	12	###	55	###	47	35	27	###	31	1	27	97	###	23	###	74	###	89
AB_BR1_A01	5489	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_BR1_A03	5489	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_BR1_A04	5489	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_BR1_A05	5489	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_BR1_A08	5489	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_BR1_A09	5489	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_BR1_A11	5489	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_BR1_A12	5489	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_BR1_B02	2744	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_BR1_B05	2744	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_BR1_B11	2744	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_XO_A08	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AB_XO_A10	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_AB_xo2_A04	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_AB_xo2_A06--F	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	CS	CS	CS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_AB_xo2_A08	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_AB_xo2_C01	273	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_AB_xo2_C04	273	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	CS	CS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_AB_xo2_C06--F	273	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_AB_xo2_E04	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_br1_A04	5489	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_br1_A09	5489	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_A01	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_A03	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_A04	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CU	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_A05	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_A06	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_A07	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CU	0	CU	CU	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_A08	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_A09	2453	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_B03	818	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_B04	818	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
d26_BA_xo2_B05	818	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d26_BA_xo2_B08	818	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_B12	818	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d26_BA_xo2_E02--F	30	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	all	0	0	0	0	0	0	0	0	1	1	1	4	4	6	19	25	26	26	26	26	29	29	29	31	31	31	32	32	32	33	33	33	33	33	33	33	34	34	34	34	34	34	34	34	34	34	34	35	35	35	35
	AB	0	0	0	0	0	0	0	0	0	0	0	1	1	1	9	15	16	16	16	16	18	17	17	19	19	19	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
	BA	0	0	0	0	0	0	0	0	1	1	1	3	3	5	10	10	10	10	10	10	11	12	12	12	12	12	12	12	12	13	13	13	13	13	13	13	14	14	14	14	14	14	14	14	14	14	14	15	15	15	15

Markers are shown by Interval identifier (Int1-Int78) and only the ones heterozygous in the respective individual are shown. "CU" refers to a haplotype switch of one or a few bases before the final crossover switch (defined as the first consecutive 2 kilobases switched to the other haplotype). "CS" refers to an additional haplotype switch of on or a few bases downstream of the final haplotype switch. Column 2 shows the respective number of input molecules for the pool from which the recombinant was recovered. For individual d07 (S33) no final haplotype switch could be detected in some recombinants so these were excluded from analysis.

S35. MinION sequencing results d30

d30	end	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
	A	A	T	C	IN	C	T	G	DEL	C	C	G	T	AG	C	C	T	A	G	G	G
	B	C	C	G	DEL	T	G	A	TT	T	T	A	C	DEL	A	T	C	G	T	T	A
		Int1	Int3	Int5	Int6	Int8	Int11	Int12	Int13	Int14	Int15	Int16	Int17	Int19	Int26	Int27	Int29	Int30	Int31	Int33	Int34
	start	####	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
	length	###	###	###	###	###	###	79	87	###	###	###	15	###	###	###	###	###	###	###	###
d30_AB_br1_C06	2530	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d30_AB_br1_C07	2530	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
d30_AB_br1_C09	2530	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
d30_AB_br1_D04	1265	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d30_AB_br1_D05	1265	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d30_AB_br1_D06	1265	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CU	0	0	0	0
d30_AB_br1_D07	1265	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d30_AB_br1_D11	1265	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_BA_br1_C04	2530	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_BA_br1_C05	2530	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
d30_BA_br1_C08	2530	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	CS	1	1	1
d30_BA_br1_C11	2530	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
d30_BA_br1_D03	1265	0	0	0	0	0	0	0	0	0	0	0	0	0	CU	0	0	0	0	0	0
d30_BA_br1_D04	1265	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d30_BA_br1_D05	1265	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
d30_BA_br1_D06	1265	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	all	0	0	1	1	1	3	3	3	3	3	3	3	4	4	7	8	8	9	9	9
	AB	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	2	2	3	3	3
	BA	0	0	1	1	1	2	2	2	2	2	2	2	3	3	6	6	6	6	6	6

d30	end	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
	A	A	TA	C	T	C	A	G	T	DEL	IN(*)	G	G	A	G	T	C	G	G	1T	G+A+I	TT	C	C	C	G	C	T	A	T
	B	G	DEL	T	C	T	G	A	C	TCTA	DEL	A	A	T	A	A	T	A	A	9T	G+DE	AAC	A	T	T	A	T	C	C	C
		Int42	Int44	Int45	Int46	Int47	Int49	Int50	Int51	Int53	Int54	Int55	Int56	Int57	Int58	Int59	Int60	Int62	Int63	Int64	Int65	Int67	Int68	Int69	Int71	Int72	Int74	Int75	Int77	Int78
	start	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
	length	###	91	###	13	54	###	95	76	###	27	75	59	12	###	55	###	47	35	27	###	32	27	97	###	23	###	74	###	89
d30_AB_br1_C06	2530	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_AB_br1_C07	2530	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_AB_br1_C09	2530	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_AB_br1_D04	1265	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_AB_br1_D05	1265	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
d30_AB_br1_D06	1265	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_AB_br1_D07	1265	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_AB_br1_D11	1265	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_BA_br1_C04	2530	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_BA_br1_C05	2530	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_BA_br1_C08	2530	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_BA_br1_C11	2530	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_BA_br1_D03	1265	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_BA_br1_D04	1265	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
d30_BA_br1_D05	1265	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d30_BA_br1_D06	1265	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	all	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	15	15	15	16
	AB	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	8
	BA	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	8	8	8	8

Markers are shown by Interval identifier (Int1-Int78) and only the ones heterozygous in the respective individual are shown. “CU” refers to a haplotype switch of one or a few bases before the final crossover switch (defined as the first consecutive 2 kilobases switched to the other haplotype). “CS” refers to an additional haplotype switch of on or a few bases downstream of the final haplotype switch. Column 2 shows the respective number of input molecules for the pool from which the recombinant was recovered. For individual d07 (S33) no final haplotype switch could be detected in some recombinants so these were excluded from analysis.

S36. MinION sequencing results d37

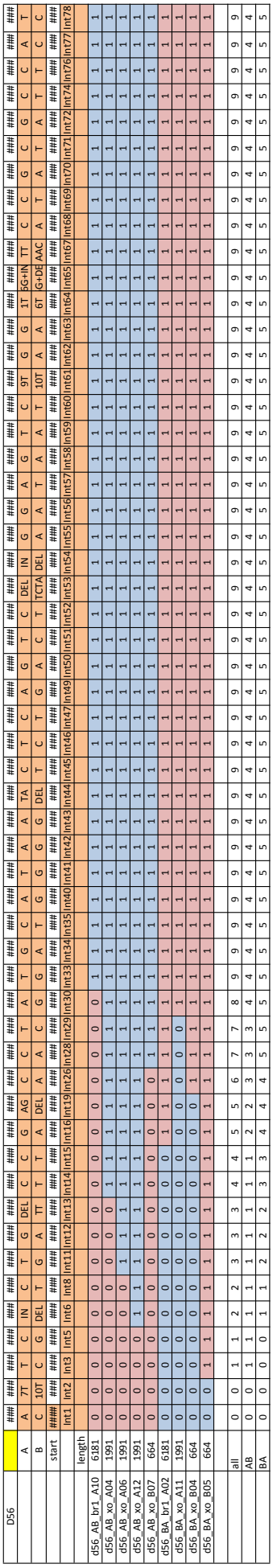
Markers are shown by Interval identifier (Int1-Int78) and only the ones heterozygous in the respective individual are shown. “CU” refers to a haplotype switch of one or a few bases before the final crossover switch (defined as the first consecutive 2 kilobases switched to the other haplotype). “CS” refers to an additional haplotype switch of on or a few bases downstream of the final haplotype switch. Column 2 shows the respective number of input molecules for the pool from which the recombinant was recovered. For individual d07 (S33) no final haplotype switch could be detected in some recombinants so these were excluded from analysis.

S37. MinION sequencing results d52

d52			###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
		A	A	8 T	T	C	IN	C	T	G	DEL	C	C	G	T	AG	C	T	G	C	A	T	A	A	TA	C	T	C	A
		B	C	10 T	C	G	DEL	T	G	A	TT	T	T	A	C	DEL	A	G	A	T	G	G	G	G	DEL	T	C	T	G
			Int1	Int2	Int3	Int5	Int6	Int8	Int11	Int12	Int13	Int14	Int15	Int16	Int17	Int19	Int28	Int33	Int34	Int35	Int40	Int41	Int42	Int43	Int44	Int45	Int46	Int47	Int49
		start	####	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
		length	###	88	###	###	###	###	###	79	87	###	###	###	15	###	###	###	###	###	###	###	###	79	12	###	13	54	###
d52_BA_xo_A03	2254	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

d52			###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
		A	G	T	C	DEL	IN(*)	G	G	A	G	T	C	G	G	1T	6G+IN	DEL	TT	C	C	G	C	G	C	C	A	T
		B	A	C	T	TCTA	DEL	A	A	T	A	A	T	A	A	9T	G+DE	G	AAC	A	T	A	T	A	T	T	C	C
			Int50	Int51	Int52	Int53	Int54	Int55	Int56	Int57	Int58	Int59	Int60	Int62	Int63	Int64	Int65	Int66	Int67	Int68	Int69	Int70	Int71	Int72	Int74	Int76	Int77	Int78
		start	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###
		length	95	76	###	15	27	75	59	12	###	55	###	47	35	27	###	31	1	27	97	###	###	23	###	###	###	89
d52_BA_xo_A03	2254	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Markers are shown by Interval identifier (Int1-Int78) and only the ones heterozygous in the respective individual are shown. “CU” refers to a haplotype switch of one or a few bases before the final crossover switch (defined as the first consecutive 2 kilobases switched to the other haplotype). “CS” refers to an additional haplotype switch of on or a few bases downstream of the final haplotype switch. Column 2 shows the respective number of input molecules for the pool from which the recombinant was recovered. For individual d07 (S33) no final haplotype switch could be detected in some recombinants so these were excluded from analysis.



Markers are shown by Interval identifier (Int1-Int78) and only the ones heterozygous in the respective individual are shown. “CU” refers to a haplotype switch of one or a few bases before the final crossover switch (defined as the first consecutive 2 kilobases switched to the other haplotype). “CS” refers to an additional haplotype switch of on or a few bases downstream of the final haplotype switch. Column 2 shows the respective number of input molecules for the pool from which the recombinant was recovered. For individual d07 (S33) no final haplotype switch could be detected in some recombinants so these were excluded from analysis.

S39. Repeat-element count data per 10 kb bin. Dog genome (CanFam 3.1) was split into bins of 10 kb and number of annotated repetitive elements and CpG-islands counted per bin. The table shows the count histogram (number of bins containing 0, 1, 2, 3…, 12 annotated elements).

Bin Center	Repeats	CpGislands
0	97322	231650
1	57422	6777
2	43238	2410
3	25191	1147
4	11876	410
5	4950	138
6	1783	35
7	553	8
8	178	1
9	51	0
10	10	0
11	1	0
12	1	0