A genomic region containing *RNF212 and CPLX1* is associated with sexually-dimorphic recombination rate variation in Soay sheep (*Ovis aries*).

Susan E. Johnston[*], Camillo Bérénos[*], Jon Slate[†], Josephine M. Pemberton[*]

[*] Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Charlotte Auerbach Road, Edinburgh, EH9 3FL, United Kingdom.

[†] Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom.

Running title: Genetic basis of recombination rate

Keywords: Meiotic recombination, genome-wide association study, genomic relatedness, heritability, natural population.

Corresponding Author:

Dr. Susan E. Johnston

Institute of Evolutionary Biology

School of Biological Sciences

University of Edinburgh,

Charlotte Auerbach Road,

Edinburgh, EH9 3FL,

United Kingdom

Tel: +44 131 650 7702

Email: Susan.Johnston@ed.ac.uk

ABSTRACT

Meiotic recombination breaks down linkage disequilibrium and forms new haplotypes, meaning that it is an important driver of diversity in eukaryotic genomes. Understanding the causes of variation in recombination rate is important in interpreting and predicting evolutionary phenomena and for understanding the potential of a population to respond to selection. However, despite attention in model systems, there remains little data on how recombination rate varies in natural populations. Here, we used extensive pedigree and high-density SNP information in a wild population of Soay sheep (*Ovis aries*) to investigate the genetic architecture of individual autosomal recombination rate. Individual recombination rates were high relative to other mammal systems, and were higher in males than in females (autosomal map lengths of 3748 cM and 2860 cM, respectively). Autosomal recombination rate was heritable in both sexes (h^2 = 0.16 & 0.12 in females and males, respectively), but with different genetic architectures. In females, 46.7% of heritable variation was explained by a sub-telomeric region on chromosome 6; a genome-wide association study showed the strongest associations at the locus *RNF212,* with further associations observed at a nearby ~374kb region of complete linkage disequilibrium containing three additional candidate loci, *CPLX1*, *GAK* and *PCGF3*. A second region on chromosome 7 containing *REC8* and *RNF212B* explained 26.2% of heritable variation in recombination rate in both sexes, with further single locus associations identified on chromosome 3. Comparative analyses with 40 other sheep breeds showed that haplotypes associated with recombination rates are both old and globally distributed. These findings provide a key advance in understanding the causes and consequences of recombination rate variation in a natural system.

AUTHOR SUMMARY

Recombination offers an escape from genetic linkage by forming new combinations of alleles, increasing the potential for populations to respond to selection. Understanding the causes and consequences of individual recombination rates are important in studies of evolution and genetic improvement, yet little is known on how rates vary in natural systems. Using data from a wild population of Soay sheep, we show that individual recombination rate is heritable and differs between the sexes, with the majority of genetic variation in females explained by a genomic region containing the genes *RNF212* and *CPLX1*.

INTRODUCTION

Recombination is a fundamental feature of sexual reproduction in nearly all multi-cellular organisms. In many species, it ensures the proper segregation of homologous chromosomes during meiosis, avoiding deleterious outcomes such as aneuploidy (Hassold and Hunt 2001; Fledel-Alon *et al.* 2009). It also plays an important role in driving the evolution of eukaryotic genomes due to its rearrangement of existing allelic variation to create novel haplotypes. Recombination can prevent the accumulation of deleterious mutations by uncoupling them from linked beneficial alleles (Muller 1964; Crow and Kimura 1965). It can also lead to an increase in genetic variance for fitness, allowing populations to respond to selection at a faster rate (McPhee and Robertson 1970; Felsenstein 1974; Charlesworth and Barton 1996; Burt 2000): this is particularly true for small populations under strong selection, where beneficial and deleterious alleles are more likely to be linked (Hill-Robertson Interference), and their relative selective costs and benefits are likely to be stronger (Hill and Robertson 1966; Otto and Barton 2001). However, recombination is also associated with fitness costs:

72 higher rates of crossing-over may increase deleterious mutations and chromosomal
73 rearrangements (Inoue and Lupski 2002), or lead to the break-up of favorable combinations
74 of alleles previously built up by selection, reducing the mean fitness of subsequent
75 generations (Charlesworth and Barton 1996). Therefore, the relative costs and benefits of
76 recombination are likely to vary within different contexts, leading to an expectation of
77 variation in recombination rates across systems (Barton 1998; Burt 2000; Otto and
78 Lenormand 2002).
79
80 Recent studies of model mammal systems have shown that recombination rates vary at an
81 individual level, and that a significant proportion of variance is driven by heritable genetic
82 effects (Kong *et al.* 2004; Dumont *et al.* 2009; Sandor *et al.* 2012). In cattle, humans and
83 mice, the heritability of recombination rate is 0.22, 0.30 and 0.46. respectively, and
84 genome-wide association studies have repeatedly attributed some heritable variation to
85 specific genetic variants, including *RNF212*, *CPLX1*, *REC8* and *PRDM9,* among others (Kong *et*
86 *al.* 2008; Baudat *et al.* 2010; Sandor *et al.* 2012). The majority of these loci appear to
87 influence crossover frequency, may be dosage dependent (e.g. *RNF212* in mice Reynolds *et*
88 *al.* 2013) and/or may have sex-specific or sexually-antagonistic effects on recombination
89 rate (e.g *RNF212* and *CPLX1* in humans and cattle; Kong *et al.* 2014; Ma *et al.* 2015). The
90 locus *PRDM9* is associated with the positioning and proportion of crossovers that occur in
91 mammalian recombination hotspots (i.e. regions of the genome with particularly high
92 recombination rates; Baudat *et al.* 2010; Ma *et al.* 2015), although this locus is not
93 functional in some mammal species, such as canids (Auton *et al.* 2013). These studies
94 suggest that recombination rate has a relatively oligogenic architecture, and therefore has
95 the potential to respond rapidly to selection over relatively short evolutionary timescales.

Such studies in model systems have provided key insights into the causes of recombination rate variation. However, with the exception of humans, studies have been limited to model systems that are likely to have been subject to strong artificial selection in their recent history, a process that will favour alleles that increase recombination rate to overcome Hill-Robertson Interference (Hill and Robertson 1966; Otto and Barton 2001). Some experimental systems show increased recombination rates after strong selection on unrelated characters (Otto and Lenormand 2002), and recombination rates are higher in domesticated plants and animals (Burt and Bell 1987; Ross-Ibarra 2004; but see Muñoz-Fuentes *et al.* 2015). Therefore, artificial selection may result in different genetic architectures than exist in natural populations. Studies examining recombination rate in wild populations will allow dissection of genetic and environmental drivers of recombination rate, and ultimately, the association with rate and individual fitness, to understand how this trait may be evolving in natural systems.

Here, we examine the genetic architecture of recombination rate variation in a wild mammal population. The Soay sheep (*Ovis aries*) is a Neolithic breed of domestic sheep that has lived unmanaged on the St Kilda archipelago (Scotland, UK, 57º49'N, 8º34'W) since the Bronze age (Clutton-Brock *et al.* 2004). Individuals from the Village Bay area of archipelago have been subject to a long-term study since 1985, with extensive genomic, pedigree, phenotype and environmental data collected for more than 6,000 individuals. In this study, we integrate genomic and pedigree information to characterize autosomal cross-over positions in more than 3000 gametes in individuals from both sexes. Our objectives were as follows: 1) to determine if individual recombination rates were heritable within the Soay

sheep system, and to what extent they differ between the sexes; 2) to determine the relative effects of common environment and other individual effects on recombination rates (e.g. age, inbreeding coefficients); and 3) to identify specific genetic variants associated with recombination rate variation.

MATERIALS AND METHODS

Study population and pedigree.

The Soay sheep is a primitive breed of domestic sheep (*Ovis aries*) which has lived unmanaged on the St Kilda archipelago (Scotland, UK, 57º49'N, 8º34'W) since the Bronze age. Sheep living within the Village Bay area of Hirta have been studied on an individual basis since 1985 (Clutton-Brock *et al.* 2004). All sheep are ear-tagged at first capture (including 95% of lambs born within the study area) and DNA samples for genetic analysis are routinely obtained from ear punches and/or blood sampling. All animal work was carried out according to UK Home Office procedures and was licensed under the UK Animals (Scientific Procedures) Act 1986 (License no. PPL60/4211). A Soay sheep pedigree has been constructed using 315 SNPs in low LD, and includes 5516 individuals with 4531 maternal and 4158 paternal links (Bérénos *et al.* 2014).

SNP Dataset.

A total of 5805 Soay sheep were genotyped at 51,135 single nucleotide polymorphisms (SNPs) on the Ovine SNP50 BeadChip using an Illumina Bead Array genotyping platform (Illumina Inc., San Diego, CA, USA; Kijas *et al.* 2009). Quality control on SNP data was carried out using the *check.marker* function in GenABEL v1.8-0 (Aulchenko *et al.* 2007) implemented in R v3.1.1, with the following thresholds: SNP minor allele frequency (MAF) >

144 0.01; individual SNP locus genotyping success > 0.95; individual sheep genotyping success >
145 0.99; and identity by state (IBS) with another individual < 0.90. Heterozygous genotypes at
146 non-pseudoautosomal X-linked SNPs within males were scored as missing. The genomic
147 inbreeding coefficient (measure $\hat{F}_{III}$ in Yang *et al.* 2011, hereafter $\hat{F}$), was calculated for
148 each sheep in the software GCTA v1.24.3 (Yang *et al.* 2011), using information for all SNP
149 loci passing quality control.
150
151 **Estimation of meiotic autosomal crossover count (ACC).**
152 **Sub-pedigree construction.** To allow unbiased phasing of the SNP data, a standardized
153 pedigree approach was used to identify cross-overs that had occurred within the gametes
154 transferred from a focal individual to its offspring; hereafter, focal individual (FID) refers to
155 the sheep in which meiosis took place. For each FID-offspring combination in the Soay sheep
156 pedigree, a sub-pedigree was constructed to include both parents of the FID (Father and
157 Mother) and the other parent of the offspring (Mate), where all five individuals had been
158 genotyped (Figure 1). This sub-pedigree structure allowed phasing of SNPs within the FID,
159 and thus the identification of autosomal cross-over events in the gamete transferred from
160 the FID to the offspring (Figure 1). Sub-pedigrees were discarded from the analysis if they
161 included the same individual twice (e.g. father-daughter matings; N = 13).
162
163 **Linkage map construction and chromosome phasing.** All analyses in this section were
164 conducted using the software CRI-MAP v2.504a (Green *et al.* 1990). First, Mendelian
165 incompatibilities in each sub-pedigree were identified using the *prepare* function;
166 incompatible genotypes were removed from all affected individuals, and sub-pedigrees
167 containing parent-offspring relationships with more than 0.1% mismatching loci were

168 discarded. Second, sex-specific and sex-averaged linkage map positions (in Kosambi cM)
169 were obtained using the *map* function, where SNPs were ordered relative to their estimated
170 positions on the sheep genome assembly Oar_v3.1 (Genbank assembly ID:
171 GCA_000298735.1; Jiang *et al.* 2014). SNP loci with a map distance of greater than 3 cM to
172 each adjacent marker (10cM for the X chromosome, including PAR) were assumed to be
173 incorrectly mapped and were removed from the analysis, with the *map* function rerun until
174 all map distances were below this threshold; in total, 76 SNPs were assumed to be
175 incorrectly mapped. Third, the *chrompic* function was used to identify informative SNPs (i.e.
176 those for which the grand-parent of origin of the allele could be determined) on chromatids
177 transmitted from the FID to its offspring; crossovers were deemed to have occurred where
178 there was a switch in the grandparental origin of a SNP allele (Figure 1).
179
180 **Quality control and crossover estimation in autosomes.** Errors in determining the grand-
181 parental origin of alleles can lead to false calling of double-crossovers (i.e. two adjacent
182 crossovers occurring on the same chromatid) and in turn, an over-estimation of
183 recombination rate. To reduce the likelihood of calling false crossover events, runs of
184 grandparental-origin consisting of a single allele (i.e. resulting in a double crossover either
185 side of a single SNP) were recoded as missing (N = 973 out of 38592 double crossovers,
186 Figure S1). In remaining cases of double crossovers, the base pair distances between
187 immediately adjacent SNPs spanning a double crossover were calculated (hereafter, "span
188 distance"). The distribution of the span distances indicated that crossover interference is
189 present within the Soay sheep population (Figure S1). Informative SNPs that occurred within
190 double-crossover segments with a $\log_{10}$ span distance lower than 2.5 standard deviations
191 from the mean $\log_{10}$ span distance (equivalent to 9.7Mb) were also recoded as missing (N =

503 out of 37619 double crossovers, Figure S1). The autosomal crossover count (ACC), the number of informative SNPs and the informative length of the genome (i.e. the total distance between the first and last informative SNPs for all chromatids) was then calculated for each FID. A simulation study was conducted to ensure that our approach accurately characterized ACC and reduced phasing errors. Autosomal meiotic crossovers were simulated given an identical pedigree structure and population allele frequencies ($N_{simulations}$ = 100; see File S1 for detailed methods and results). Our approach was highly accurate in identifying the true ACC per simulation across all individuals and per individual across all simulations (adjusted $R^2 > 0.99$), but indicated that accuracy was compromised in individuals with high values of $\hat{F}$. This is likely to be an artefact of long runs of homozygosity as a result of inbreeding, which may prevent detection of double crossovers or crossovers in sub-telomeric regions. To ensure accurate individual estimates of ACC, gametes with a correlation of adjusted $R^2 \leq 0.95$ between simulated and detected crossovers in the simulation analysis were removed from the study (N = 8).

Assessing variation in the recombination landscape.

Broad Scale Recombination Rate. Relationships between chromosome length and linkage map length, and male and female linkage map length were analyzed using linear regressions in R v3.1.1. The relationship between chromosome length and chromosomal recombination rate (defined as cM length/Mb length) was modelled using a multiplicative inverse ($1/x$) regression in R v3.1.1.

Fine Scale Recombination Rate. The probability of crossing-over was calculated in 1MB windows across the genome using information from the male and female linkage maps.

Briefly, the probability of crossing over within a bin was the sum of all recombination fractions, *r*, in that bin; in cases where an *r* value spanned a bin boundary, it was recalculated as $r \times N_{boundary}/N_{adjSNP}$, where $N_{boundary}$ was the number of bases to the bin boundary, and N_{adjSNP} was the number of bases to the closest SNP within the adjacent bin.

Variation in crossover probability relative to proximity to telomeric regions on each chromosome arm was examined using general linear models with a Gaussian error structure. The response variable was crossover-probability per bin; the fitted covariates were as follows: distance to the nearest telomere, defined as the best fit of either a linear (x), multiplicative inverse ($1/x$), quadratic ($x^2 + x$), cubic ($x^3 + x^2 + x$) or a log term ($\log_{10} x$); sex, fitted as a main effect and as an interaction term with distance to the nearest telomere; number of SNPs within the bin; and GC content of the bin (%, obtained using sequence from Oar_v3.1, Jiang *et al.* 2014). The best model was identified using Akaike's Information Criterion (Akaike 1974). An additional model was tested, using ratio of male to female crossover probability as the response variable, with the same fixed effect structure (omitting sex). In both models, the distance to the nearest telomere was limited to 60Mb, equivalent to half the length of the largest acrocentric chromosome (Chr 4). Initial models also included a term indicating if a centromere was present or absent on the 60Mb region, but this term was not significant in either model.

Heritability and cross-sex genetic correlations of autosomal recombination rate.

ACC was modelled as a trait of the FID. Phenotypic variance in ACC was partitioned using a restricted maximum likelihood (REML) Animal Model (Henderson 1975) implemented in ASReml-R (Butler *et al.* 2009) in R v3.1.1. To determine the proportion of phenotypic

variance attributed to additive genetic effects (i.e. heritability), a genomic relatedness matrix at all autosomal markers was constructed for all genotyped individuals using GCTA v1.24.3 (Yang *et al.* 2011); the matrix was shrunk using *--adj 0*. Trait variance of was analyzed first with the following univariate model:

$$y = X\beta + Z_1 a + Z_r u_r + e$$

where *y* is a vector of the ACC per transferred gamete; *X* is an incidence matrix relating individual measures to a vector of fixed effects, *beta*; Z_1 and Z_r are incidence matrices relating individual measures with additive genetic effects and random effects, respectively; *a* and u_r are vectors of additive genetic effects from the genomic relatedness matrix and additional random effects, respectively; and e is a vector of residual effects. The heritability (h^2) was calculated as the ratio of the additive genetic variance to the sum of the variance estimated for all random effects. Model structures were initially tested with a number of fixed effects, including sex, $\hat{F}$ and FID age at the time of meiosis; random effects tested included: individual identity to account for repeated measures within the same FID; maternal and paternal identity; and common environment effects of FID birth year and offspring birth year. Significance of fixed effects was determined using a Wald test, whereas significance of random effects was calculated using likelihood ratio tests (LRT) between models with and without the focal random effect. Only sex and additive genetic effects were significant in any model; however, $\hat{F}$ and individual identity were retained in all models to account for potential underestimation of ACC and the effects of pseudoreplication, respectively.

263 To investigate if the additive genetic variation underlying male and female ACC was
264 associated with sex-specific variation in ACC, bivariate models were run. The additive
265 genetic correlation r_A was determined using the CORGH error structure function in ASReml-
266 R, (correlation with heterogeneous variances) with r_A set to be unconstrained. To test
267 whether the genetic correlation was significantly different from 0 and 1, the unconstrained
268 model was compared to models with r_A fixed at a value of 0 or 0.999. Differences in additive
269 genetic variance in males and females were tested by constraining both to be equal values.
270 Models were then compared using likelihood ratio tests with 1 degree of freedom.
271
272 **Genetic architecture of autosomal crossover count.**
273 **Genome partitioning of genetic variance (regional heritability analysis).** The contribution
274 of specific genomic regions to trait variation was determined by partitioning the additive
275 genetic variance as follows (Nagamine *et al.* 2012):
276

$$y = X\beta + Z_1 v_i + Z_2 nv_i + Z_r u_r + e$$

277
278 where *v* is the vector of additive genetic effects explained by an autosomal genomic region *i*,
279 and *nv* is the vector of the additive genetic effects explained by all remaining autosomal
280 markers outwith region *i*. Regional heritabilities were determined by constructing genomic
281 relatedness matrices (GRMs) for regions of *i* of increasing resolution (whole chromosome
282 partitioning, sliding windows of 150, 50 and 20 SNPs, corresponding to regions of 9.4, 3.1
283 and 1.2Mb mean length, respectively) and fitting them in models with an additional GRM of
284 all autosomal markers not present in region i; sliding windows overlapped by half of their
285 length (i.e. 75, 25 and 10 SNPs, respectively). GRMs were constructed in the software GCTA

v1.24.3 and were shrunk using the *--adj 0* argument (Yang *et al.* 2011). The significance of additive genetic variance attributed to a genomic region i was tested by comparing models with and without the Z_1v_i term using a likelihood ratio test; in cases where the heritability estimate was zero (i.e. estimated as "Boundary" by ASReml), significant model comparison tests were disregarded. A Bonferroni approach was used to account for multiple testing across the genome, by taking the number of tests and dividing by two to account for the overlap of the sliding windows (since each genomic region was modelled twice).

Genome-wide association study of variants controlling ACC. Genome-wide association studies (GWAS) of autosomal recombination rates under different scenarios were conducted using ASReml-R (Butler *et al.* 2009) in R v3.1.1, fitting individual animal models for each SNP locus using the same model structure as above. SNP genotypes were fitted as a fixed effect with two or three levels. The GRM was replaced with a relatedness matrix based on pedigree information to speed up computation; the pedigree and genomics matrices have been shown to be highly correlated (Bérénos *et al.* 2014). Sex-specific models were also run. Association statistics were corrected for any population stratification not captured by the animal model by dividing them by the genomic control parameter, λ (Devlin *et al.* 1999), when $\lambda > 1$, which was calculated as the median Wald test χ_2^2 divided by the median χ_2^2, expected from a null distribution. The significance threshold after multiple testing was determined using a linkage disequilibrium-based method outlined in (Moskvina and Schmidt 2008) using a sliding window of 50 SNPs; the effective number of tests in the GWAS analysis was 22273.61, meaning the significance threshold for P after multiple testing at $\alpha = 0.05$ was 2.245×10^{-6}. Although sex chromosome recombination rate was not included in the analysis, all GWAS included the X chromosome and SNP markers of unknown position (N=314). The

proportion of phenotypic variance attributed to a given SNP was calculated using the using the following equation (Falconer and Mackay 1996):

$$V_{SNP} = 2pq[a + d(q - p)]^2$$

where *p* and *q* are the frequencies of alleles A and B at the SNP locus, *a* is half the difference in the effect sizes estimated for the genotypes AA and BB, and *d* is the difference between a and the effect size estimated for genotype AB when fitted as a fixed effect in an animal model. The proportion of heritable variation attributed to the SNP was calculated as the ratio of V_{SNP} to the sum of V_{SNP} and the additive genetic variance estimated from a model excluding the SNP as a fixed effect. Standard errors of V_{SNP} were estimated using a delta method approach. Gene annotations in significant regions were obtained from Ensembl (gene build ID: Oar_v3.1.79; Cunningham *et al.* 2014). The position of a strong candidate locus, *RNF212* is not annotated on Oar_v3.1, but sequence alignment indicated that it is positioned at the sub-telomere of chromosome 6 (see File S2).

Accounting for cis- and trans- genetic variants affecting recombination rate. In the above analyses, we wished to separate potential associations with ACC due to cis-effects (i.e. genetic variants that are in linkage disequilibrium with polymorphic recombination hotspots) from those due to trans-effects (i.e. genetic variants in LD with genetic variants that affect recombination rate globally). By using the total ACC within a gamete, we incorporated both cis- and trans- effects into a single measure. To examine trans-effects only, we determined associations between each SNP and ACC minus crossovers that had occurred on the chromosome on which the SNP occurred e.g. for a SNP on chromosome 1,

333 association was examined with ACC summed across chromosomes 2 to 26. We found that
334 examining trans-variation (ACC minus focal chromosome) obtained similar results to cis- and
335 trans-variation (ACC) for both regional heritability and genome-wide association analyses,
336 leading to the same biological conclusions.
337
338 **Linkage disequilibrium and imputation of genotypes in significant regions.** A reference
339 population of 189 sheep was selected and genotyped at 606,066 SNP loci on the Ovine
340 Infinium® HD SNP BeadChip for imputation of genotypes into individuals typed on the 50K
341 chip. Briefly, the reference population was selected iteratively to maximize $\sum_{i=1}^{m} p_i$ using the
342 equation $p_m = A_m^{-1} c_m$, where *p* is a vector of the proportion of genetic variation in the
343 population captured by *m* selected animals, A_m is the corresponding subset of a pedigree
344 relationship matrix and *c* is a vector of the mean relationship of the *m* selected animals (as
345 outlined in (Pausch *et al.* 2013) & (Goddard and Hayes 2009)). This approach should capture
346 the maximum amount of genetic variation within the main population for the number of
347 individuals in the reference population. SNP loci were retained if call rate was > 0.95 and
348 MAF > 0.01 and individuals were retained if more than 95% of loci were genotyped. Linkage
349 disequilibrium (LD) between loci was calculated using Spearman-Rank correlations (r^2) in the
350 188 individuals passing quality control.
351
352 Genotypes from the HD SNP chip were imputed to individuals typed on the SNP50 chip in
353 the chromosome 6 region significantly associated with ACC, using pedigree information in
354 the software MaCH v1.0.16 (Li *et al.* 2010). This region contained 10 SNPs from the Ovine
355 SNP50 BeadChip and 116 additional independent SNPs on the HD SNP chip. As the software
356 requires both parents to be known for each individual, cases where only one parent was

known were scored as both parents missing. Genotypes were accepted when the dosage probability was between 0 and 0.25, 0.75 and 1.25, or 1.75 and 2 (for alternate homozygote, heterozygote and homozygote, respectively). The accuracy of genotyping at each locus was tested using 10-fold cross-validation within the reference population: genotypes were imputed for 10% of individuals randomly sampled from the reference population, using genotype data for the remaining 90%; this cross-validation was repeated 1000 times, to compare imputed genotypes with true genotypes. Cross-validation showed a relationship between number of missing genotypes and number of mismatching genotypes within individuals; therefore, individuals with < 0.99 imputed genotypes scored were removed from the analysis. Loci with < 0.95 of individuals typed were also discarded. Imputation accuracy was calculated for all loci as the proportion of imputed genotypes matching their true genotypes; all remaining loci had imputation accuracies > 0.95.

Haplotype sharing of associated regions with domesticated breeds.

A recent study has shown that Soay sheep are likely to have experienced an introgression event with a more modern breed (the Old Scottish Shortwool, or Dunface breed, now extinct) approximately 150 years ago (Feulner *et al.* 2013). Therefore, we wished to determine if alleles at the most highly associated imputed SNP, oar3_OAR6_116402578 (see Results), had recently introgressed into the population by examining haplotype sharing (HS) between Soay sheep and Boreray sheep, a cross between Dunface and Scottish Blackface sheep. We used data from the OvineSNP50 BeadChip data for Soays and a further 2709 individuals from 73 different sheep breeds (provided by the International Sheep Genomics Consortium, ISGC; see Kijas *et al.* 2012, Feulner *et al.* 2013 and Table S1). In both the Soay and non-Soay datasets of the Ovine SNP50 Beadchip, we extracted 58 SNPs corresponding

to ~4Mb of the sub-telomeric region on chromosome 6 and phased them using Beagle v4.0 (Browning and Browning 2007). We identified core haplotypes of 6 bases that tagged different alleles at oar3_OAR6_116402578. The length of HS between the core Soay haplotypes and non-Soay breeds was then calculated as follows: for each core haplotype *i* and each sheep breed *j*, any haplotypes containing *i* were extracted, and the distance from *i* to the first mismatching SNP downstream *i* was recorded. This was repeated for all pairwise comparisons of Soay and non-Soay haplotypes to determine a mean and standard deviation of HS between *i* and breed *j*.

Data availability statement.

[Data and scripts will be uploaded to a public data repository if accepted for publication]. The supplementary information contains information on additional analyses conducted and is referenced within the text. Table S1 contains the sex-averaged and sex-specific linkage map positions and genomic positions of SNP loci. Tables S3, S4 and S5 contain full detailed results and effect sizes of the regional heritability, genome-wide association and imputed association studies, respectively.

RESULTS

Broad-scale variation in recombination landscape.

We used pedigree information and data from 39104 genome-wide SNPs typed on the Ovine SNP50 BeadChip (Kijas *et al.* 2009) to identify 98420 meiotic crossovers in gametes transferred to 3330 offspring from 813 unique focal individuals. A linkage map of all 26 autosomes had a sex-averaged length of 3304 centiMorgans (cM), and sex-specific lengths

of 3748 cM and 2860 cM in males and females, respectively, indicating strong male-biased recombination rates in this population (Male:Female linkage map lengths = 1.31; Figure S2, Table S2). There was a linear relationship between the length of autosomes in megabases (Mb) and linkage map lengths (cM; Adjusted $R^2 = 0.991$, $P < 0.001$; Figure 2A). Chromosome-wide recombination rates (cM/Mb) were higher in smaller autosomes (fitted as multiplicative inverse function, adjusted $R^2 = 0.616$, $P < 0.001$, Figure 2B), indicative of obligate crossing over. The degree of sex-differences in recombination rate based on autosome length in cM (i.e. differences in male and female recombination rate) was consistent across all autosomes (Adjusted $R^2 = 0.980$, $P < 0.001$, Figure 2C).

Fine-scale variation in recombination landscape.

Finer-scale probabilities of crossing-over were calculated for 1Mb windows across the genome for each sex, using recombination fractions from their respective linkage maps. Crossover probability was variable relative to proximity to telomeric regions, with a significant interaction between sex and distance to the nearest telomere fitted as a cubic polynomial function (Figure 3A). Males had significantly higher probabilities of crossing-over than females between distances of 0Mb to 12.45Mb from the nearest telomere (Figure 3B, Table S3). Increased crossover probabilities were associated with higher GC content (General linear model, $P < 0.001$; Table S3). Investigation of the relative distances between crossovers (in cases where two or more crossovers were observed on a single chromatid) indicated that there is likely to be crossover interference within this population, with a median distance between double crossovers of 48Mb (Figure S1).

Analysis of individual recombination rate.

Individual autosomal crossover count (ACC) was heritable (h^2_g= 0.145, SE = 0.027), with the remainder of the phenotypic variance being explained by the residual error term (Table 1). ACC was significantly higher in males than in females, with 7.376 (SE = 0.263) more crossovers observed per gamete (Animal Model, Z = 28.02, $P_{Wald} < 0.001$). However, females had marginally higher additive genetic variance P_{LRT} = 0.046) and higher residual variance ($P_{LRT} = 1.07 \times 10^{-3}$) in ACC than males (Table 1). There was no relationship between ACC and FID age, offspring sex, or the genomic inbreeding coefficients of the FID or offspring; furthermore, there was no variance in ACC explained by common environmental effects such as FID birth year, year of gamete transmission, or maternal/paternal identities of the FID (Animal Models, P > 0.05). A bivariate model of male and female ACC showed that the cross-sex additive genetic correlation (r_a) was 0.808 (SE = 0.202); this correlation was significantly different from 0 ($P_{LRT} < 0.001$) but not different from 1 (P_{LRT} = 0.308).

Genetic architecture of recombination rate.

Partitioning variance by genomic region. The contribution of specific genomic regions to ACC was determined by partitioning the additive genetic variance in sliding windows (Regional heritability analysis, Table S4). There was a strong sex-specific association of ACC in females within a sub-telomeric region on chromosome 6 (20 SNP sliding window; Figure 4B). This corresponded to a 1.46 Mb segment containing ~37 protein coding regions, including ring finger protein 212 (*RNF212*) and complexin 1 (*CPLX1*) loci previously implicated in recombination rate variation in humans, cattle and mice (Kong *et al.* 2008, 2014; Sandor *et al.* 2012; Reynolds *et al.* 2013; Ma *et al.* 2015). This region explained 8.02% of the phenotypic variance (SE = 3.55%) and 46.7% of the additive genetic variance in females ($P_{LRT} = 9.78 \times 10^{-14}$), but did not contribute to phenotypic variation in males (0.312%, SE = 1.2%, P_{LRT} = 0.82; Figure 4C, Table S4). There was an additional significant

association between ACC in both sexes and a region on chromosome 7, corresponding to a 1.09Mb segment containing ~50 protein coding regions, including *RNF212B* (a paralogue of *RNF212*) and meiotic recombination protein locus *REC8* ($P_{LRT} = 3.31 \times 10^{-6}$, Figure 4A, Table S4). This region explained 4.12% of phenotypic variation (SE = 2.3%) and 26.2% of the additive genetic variation in both sexes combined; however, this region was not significant associated with ACC within each sex individually after correction for multiple testing (Table S4). Full results for the regional heritability analysis are provided in Table S4.

Genome-wide association study (GWAS). The most significant association between SNP genotype and ACC in both sexes was at s74824.1 in the sub-telomeric region of chromosome 6 ($P = 2.92 \times 10^{-10}$, Table 2). Sex-specific GWAS indicated that this SNP was highly associated with female ACC ($P = 1.07 \times 10^{-11}$), but was not associated with male ACC (P = 0.55; Table 2, Figure 5). This SNP corresponded to the same region identified in the regional heritability analysis, was the most distal typed on the chromosome from the Ovine SNP50 BeadChip (Figure 5), and had an additive effect on female ACC, with a difference of 3.37 (S.E. = 0.49) autosomal crossovers per gamete between homozygotes (Table 2). A SNP on an unmapped genomic scaffold (1.8kb, NCBI Accession: AMGL01122442.1) was also highly associated with female ACC (Figure 5). BLAST analysis indicated that the most likely genomic position of this SNP was at ~113.8Mb on chromosome 6, corresponding to the same sub-telomeric region.

Two regions on chromosome 3 were associated with ACC using the GWAS approach, although their respective regions had not shown a significant association with ACC using a regional heritability approach (see above). A single SNP on chromosome 3,

476 OAR3_51273010.1, was associated with ACC in males, but not in females, and had an
477 approximately dominant effect on ACC (P = 1.15× 10^{-6}, Figure 5, Table 2); This SNP was
478 17.8kb from the 3' UTR of leucine rich repeat transmembrane neuronal 4 (*LRRTM4*) in an
479 otherwise gene poor region of the genome (i.e. the next protein coding regions are > 1Mb
480 from this SNP in either direction). A second SNP on chromosome 3, OAR3_87207249.1, was
481 associated with ACC in both sexes (P = 1.95× 10^{-6}, Figure 5, Table 2). This SNP was 137kb
482 from the 5' end of an orthologue of WD repeat domain 61 (*WDR61*) and 371kb from the 5'
483 end of an orthologue of ribosomal protein L10 (*RPL10*). No significant SNP associations were
484 observed at the significant regional heritability region on chromosome 7. Full results of
485 GWAS are provided in Table S5.
486
487 **Genotype imputation and association analysis at the sub-telomeric region of**
488 **chromosome 6.** Genotyping of 187 sheep at a further 122 loci in the sub-telomeric region of
489 chromosome 6 showed that this region has elevated levels of linkage disequilibrium, with
490 the two most significant SNPs from the 50K chip tagging a haplotype block of ~374kB (r^2 >
491 0.8; see File S3, Figure 6, Table S6). This block contained three candidate genes, complexin 1
492 (*CPLX1*), cyclin-G-associated kinase (*GAK*) and polycomb group ring finger 3 (*PCGF3*) and was
493 177kb away from the candidate locus *RNF212* (Kong *et al.* 2014). SNP genotypes were
494 imputed for all individuals typed on the 50K chip at these 122 loci, and the association
495 analysis was repeated. The most highly associated SNP (oar3_OAR6_116402578, P = 1.83 ×
496 10^{-19}; Table 2, Figure 6) occurred within an intronic region of an uncharacterized protein
497 orthologous to transmembrane emp24 protein transport domain containing *(TMED11*),
498 25.2kb from the putative location of *RNF212* and 13kb from the 3' end of spondin 2
499 (*SPON2*). A bivariate animal model including an interaction term between ACC in each sex

and the genotype at oar3_OAR6_116402578 confirmed that this locus had an effect on female ACC only; this effect was additive, with a difference of 4.91 (S.E. = 0.203) autosomal crossovers per gamete between homozygotes (Figure 7, Tables 2 and S6). There was no difference in ACC between the three male genotypes. Full results for univariate models at imputed SNPs are given in Table S6.

Haplotype sharing of associated regions with domesticated breeds.

Seven core haplotypes of six SNPs in length tagged different alleles at oar3_OAR6_116402578 at the sub-telomeric region of chromosome 6. Two were perfectly associated with the A allele at oar3_OAR6_116402578 (conferring reduced ACC) and five were perfectly associated with the G allele (conferring increased ACC; Table S8). The extent of HS between Soays and non-Soays was low, and there was no evidence of long-range HS between Soays and Boreray in comparison to other domesticated breeds (Figure S3). This test is not definitive due to the relatively small sample size of the Boreray sample (N = 20), meaning that it is possible that either allele occurs in Boreray sheep but has not been sampled. Nevertheless, low levels of haplotype sharing with other breeds throughout the sample suggest that allele at oar3_OAR6_116402578 have not been recently introduced to the Soay sheep population. For example, HS of core haplotypes with Boreray sheep for coat colour, coat pattern (Feulner *et al.* 2013) and normal horn development (Johnston *et al.* 2013) extended to longer distances, of up to 5.7, 6.4 and 2.86Mb respectively. In contrast, the maximum HS observed here was 0.38Mb. A shorter haplotype may be expected, as the core haplotype occurs at the end of the chromosome, and so haplotype sharing is only calculated downstream of the core haplotype; however, this value is much lower than half

of previously identified shared haplotypes. The three most common haplotypes, H2, H3 and H6 (for high, high and low ACC, respectively) are found in many other sheep breeds across the world (Figure S3), suggesting that both high and low ACC haplotypes are ancient within the population.

DISCUSSION

In this study, we used the complementary approaches of regional heritability analysis, genome-wide association and genotype imputation to identify the genetic architecture of recombination rate in Soay sheep. We found that recombination rate was higher in males, but that females had more heritable variation, with female recombination rates strongly influenced by a single genomic region containing *RNF212* and *CPLX1*. We found no association between common environment effects, individual age or inbreeding and recombination rate; rather, the majority of variation was attributed to residual effects. This suggests a strong stochastic element driving recombination rates, with a small but significant heritable component. Although Soay sheep have undergone a period of domestication in their early history, examination of haplotypes around *RNF212* suggest that variation has been maintained in Soay sheep throughout their long history on St Kilda, and none are likely to have been introduced by recent introgression. Here, we discuss the interpretation and implications of our findings for studies of recombination rate variation in the Soay sheep and other systems.

Genetic variants associated with individual recombination rate.

The strongest association was observed at *RNF212*, with the genomic region accounting for ~47% of heritable variation in female recombination rate; this region was not associated

546 with male recombination rate. *RNF212* has been repeatedly implicated in recombination
547 rate variation in mammals (Sandor *et al.* 2012; Reynolds *et al.* 2013) and has also been
548 shown to have sexually-antagonistic effects on recombination rate in humans (Kong *et al.*
549 2008, 2014; Chowdhury *et al.* 2009). Mouse studies have established that the protein
550 RNF212 is essential for the formation of crossover-specific complexes during meiosis, and
551 that its effect is dosage-sensitive (Reynolds *et al.* 2013). The same sub-telomeric region of
552 chromosome 6 also contained three further candidate loci, namely *CPLX1*, *GAK* and *PCGF3*
553 (Figure 6); these loci occurred on a ~374kb block of high LD ($r^2 > 0.8$) and were in moderate
554 LD with the most highly associated SNP at *RNF212* ($r^2 = 0.54$). SNP variants in *CPLX1* have
555 been associated with large differences in linkage map lengths and with sex-differences in
556 recombination rates in both cattle and humans (Kong *et al.* 2014; Ma *et al.* 2015). GAK
557 forms part of a complex with cyclin-G, a locus involved in meiotic recombination repair in
558 *Drosophila* (Nagel *et al.* 2012), and PCGF3 forms part of a PRC1-like complex (polycomb
559 repressive complex 1) which is involved in meiotic gene expression and the timing of meiotic
560 prophase in female mice (Yokobayashi *et al.* 2013). High LD within this region meant that it
561 was not possible to test the effects of these loci on recombination rate independently;
562 however, the co-segregation of several loci affecting meiotic processes may merit further
563 investigation to determine if recombination is suppressed in this region, and if this co-
564 segregation is of adaptive significance.
565
566 Additional genomic regions associated with recombination rate included a 1.09Mb region of
567 chromosome 7 affecting rates in both sexes (identified using regional heritability analysis),
568 and two loci at 48.1Mb and 82.4Mb on chromosome 3 (identified using GWAS) with effects
569 on males only and both sexes, respectively. The chromosome 7 region contained two loci

570 associated with recombination phenotypes: *REC8*, the protein of which is required for the
571 separation of sister chromatids and homologous chromosomes during meiosis (Parisi *et al.*
572 1999); and *RNF212B*, a paralogue of *RNF212*. This region is also associated with
573 recombination rate in cattle (Sandor *et al.* 2012). This association was not detected through
574 GWAS but corresponds with an *a priori* candidate locus, indicating the potential for regional
575 heritability approaches to characterize variation from multiple alleles and/or haplotypes
576 encompassing both common and rare variants that are in linkage disequilibrium with causal
577 loci (Nagamine *et al.* 2012). The chromosome 3 variants identified were novel to this study,
578 and occurred in relatively gene poor regions of the genome. These variants remained
579 significant when repeating association analysis on recombination rate excluding their own
580 chromosome (see Materials and Methods, Tables S4 & S5), meaning that they are likely to
581 affect recombination rate globally (i.e. trans-acting effects), rather than being in LD with
582 polymorphic recombination hotspots.
583
584 After accounting for significant regions on chromosomes 6 and 7, between 36 and 61% of
585 heritable variation in recombination rate was attributed to polygenic variation independent
586 of these regions (Table S9). This "missing heritability" (i.e. heritable variation of unknown
587 architecture) suggests that the genetic architecture of this trait is comprised of several loci
588 of large effects on phenotype, but with a significant polygenic component i.e. variance
589 attributed to few or many loci with small effects on phenotype in this population.
590
591 In this study, there was no association between recombination rate and known homologues
592 of *PRDM9* within the sheep genome. This may not be surprising, as although *PRDM9* is
593 associated with recombination rate in cattle (Ma *et al.* 2015), this may be a consequence of

differences in the abundance of motifs recognized by the PRDM9 protein in hotspots rather than the locus itself affecting rate. At present, it is not known whether *PRDM9* is functional in sheep; furthermore, it was not possible to examine hotspot usage in the current study, as such analyses would require higher density of SNP loci to determine crossover positions at a greater resolution. Further studies with higher densities of markers will be required to determine the functionality and/or relative importance of *PRDM9* within this system.

Sex differences in the genetic architecture of recombination rate.

The between-sex genetic correlation of recombination rate detected in this study was not significantly different from 1, indicating that male and female recombination rate variation had a shared genetic basis – albeit with a relatively large error around this estimate. Further investigation through regional heritability mapping and GWAS indicated that variation in recombination rate has *some* degree of a shared and distinct genetic architecture between the sexes, which may be expected due to the similarities of this process, but differences in implementation (see next section). It is unlikely that the absence of associations between the *RNF212* region and male recombination rate is due to low power to detect the effect, as (a) this locus had a sexually dimorphic effect *a priori*, and (b) bivariate models accounting for variation in *RNF212* as a fixed effect supported a sexually dimorphic genetic effect with a lower degree of error than the bivariate approach (Figure 7). The strong sex difference in recombination rate in this study was manifested in increased male recombination in sub-telomeric regions, allowing further crossovers to occur on the same chromatid despite crossover interference. Although variation at *RNF212* resulted in markedly different recombination rates in females, the relative positions of crossovers did not differ between females with different genotypes (Figure S4, Table S10), suggesting that differences in the

618 action of *RNF212* on female recombination rate may be due to protein function or dosage
619 dependence, such as is observed in mouse systems (Reynolds *et al.* 2013).
620
621 **Why is recombination rate higher in males?**
622 Meiotic recombination occurs at distinctly different times in each sex: male meiosis occurs
623 during spermatogenesis, whereas female meiosis starts in the fetal ovary, but is arrested
624 during crossing-over between prophase and metaphase I and completed after fertilization
625 (Morelli and Cohen 2005). In placental mammals, females usually exhibit higher
626 recombination rates than males (Lenormand and Dutheil 2005); this is likely to be a
627 mechanism to avoid aneuploidy after long periods of meiotic arrest (Koehler *et al.* 1996;
628 Morelli and Cohen 2005; Nagaoka *et al.* 2012). Nevertheless, the difference between male
629 and female recombination rates vary to a large degree across species, with some species,
630 such as domestic sheep, cattle, macaques and marsupials showing increased recombination
631 rates in males (Burt and Bell 1987; Maddox *et al.* 2001a; Ma *et al.* 2015). A number of
632 mechanisms have been proposed to explain variation in sex differences, including haploid
633 selection (Lenormand and Dutheil 2005), meiotic drive (Brandvain and Coop 2012), sperm
634 competition, sexual dimorphism and dispersal (Trivers 1988; Burt *et al.* 1991; Mank 2009),
635 although testing these ideas is limited by a paucity of empirical data.
636
637 In this study, Soay sheep had male-biased recombination rates to a greater degree than
638 observed in any placental mammal to date (Male:Female linkage map ratio = 1.31). The
639 reason for this is unclear. One possibility is that Soay sheep have a highly promiscuous
640 mating system, with the largest testes to body size ratio within ruminants (Stevenson *et al.*
641 2004) and high levels of sperm competition, with dominant rams suffering from sperm

642 depletion towards the end of the rut (Preston *et al.* 2001). Increased recombination may
643 allow more rapid sperm production through formation of meiotic bouquets (Tankimanova
644 *et al.* 2004). However, the finding of male-biased recombination is in contrast to that
645 observed in wild bighorn sheep (*Ovis canadensis,* male:female ratio = 0.89; Poissant *et al.*
646 2010, divergence ~2.8Mya). Given our finding that the *RNF212*/*CPLX1* region is involved in
647 the sex difference in Soay sheep, there is a compelling case for a role for this region in
648 driving sex-differences in mammal systems over relatively short evolutionary timescales.
649 Regardless, more empirical studies are required to elucidate the specific drivers of sex-
650 differences in recombination rate at both a mechanistic and inter-specific level.
651
652 **Conclusions**
653 In this study, we have shown that recombination rates in Soay sheep are heritable and have
654 a sexually dimorphic genetic architecture. The variants identified have been implicated in
655 recombination rates in other mammal species, indicating a conserved genetic basis across
656 distantly related taxa. Future studies in this system will investigate the relationship between
657 variants identified in this study and individual life-history variation to determine if the
658 maintenance of genetic variation for recombination rates is due to selection or stochastic
659 processes. Overall, the approaches and findings presented here provide an important
660 foundation for studies examining the evolution of recombination rates in contemporary
661 natural populations.
662
663 ## ACKNOWLEDGEMENTS

664 We thank Ian Stevenson and all Soay sheep project members and volunteers for collection
665 of data and samples. Discussions and comments from Jarrod Hadfield, Bill Hill, Craig Walling,

Jisca Huisman and John Hickey greatly improved the analysis. James Kijas, Yu Jiang and Brian Dalrymple responded to numerous queries on the genome assembly, annotation and SNP genotyping. Philip Ellis prepared DNA samples, and Louise Evenden, Jude Gibson and Lee Murphy carried out SNP genotyping at the Wellcome Trust Clinical Research Facility Genetics Core, Edinburgh. This work has made extensive use of the resources provided by the Edinburgh Compute and Data Facility (http://www.ecdf.ed.ac.uk/). Permission to work on St Kilda is granted by The National Trust for Scotland and Scottish Natural Heritage, and logistical support was provided by QinetiQ and Eurest. The Soay sheep project is supported by grants from the UK Natural Environment Research Council and the SNP genotyping and this research was supported by an ERC Advanced Grant to JMP.

REFERENCED LITERATURE

Akaike H., 1974 A new look at the statistical model identification. IEEE Trans. Automat. Contr. **19**: 716–723.

Aulchenko Y. S., Ripke S., Isaacs A., Duijn C. M. van, 2007 GenABEL: an R library for genome-wide association analysis. Bioinformatics **23**: 1294–1296.

Auton A., Rui Li Y., Kidd J., Oliveira K., Nadel J. et al 2013 Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs. PLoS Genet. **9**: e1003984.

Barton N. H., 1998 Why sex and recombination? Science (80-.). **281**: 1986–1990.

Baudat F., Buard J., Grey C., Fledel-Alon A., Ober C., *et al.*, 2010 PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science **327**: 836–840.

Bérénos C., Ellis P., Pilkington J. G., Pemberton J. M., 2014 Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. Mol. Ecol. **23**: 3434–3451.

Brandvain Y., Coop G., 2012 Scrambling eggs: meiotic drive and the evolution of female recombination rates. Genetics **190**: 709–23.

Browning S. R., Browning B. L., 2007 Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. **81**: 1084–1097.

Burt A., Bell G., 1987 Mammalian chiasma frequencies as a test of two theories of recombination. Nature **326**: 803–805.

Burt A., Bell G., Harvey P. H., 1991 Sex differences in recombination. J. Evol. Biol. **4**: 259–277.

Burt A., 2000 Sex, Recombination, and the Efficacy of Selection - was Weismann Right? Evolution (N. Y). **54**: 337–351.

Butler D. G., Cullis B. R., Gilmour A. R., Gogel B. J., 2009 Mixed Models for S language Environments: ASReml-R reference manual.

Charlesworth B., Barton N. H., 1996 Recombination load associated with selection for increased recombination. Genet. Res. Cambr. **67**: 27–41.

Chowdhury R., Bois P. R. J., Feingold E., Sherman S. L., Cheung V. G., 2009 Genetic analysis of variation in human meiotic recombination. PLoS Genet. **5**.

Clutton-Brock T., Pemberton J., Coulson T. N., Stevenson I. R., MacColl A. D. C., 2004 The sheep of St Kilda. In: *Soay Sheep: Dynamics and Selection in an Island Population.2*,, pp. 17–51.

Crow J. F., Kimura M., 1965 Evolution in Sexual and Asexual Populations. Am. Nat. **99**: 439–450.

Cunningham F., Amode M. R., Barrell D., Beal K., Billis K. *et al.*, 2014 Ensembl 2015. Nucleic Acids Res. **43**: D662–D669.

Devlin A. B., Roeder K., Devlin B., 1999 Genomic Control for Association. Biometrics **55**: 997–1004.

Dumont B. L., Broman K. W., Payseur B. A., 2009 Variation in genomic recombination rates among heterogeneous stock mice. Genetics **182**: 1345–9.

Falconer D. S., Mackay T. F. C., 1996 Variance. In: *Introduction to Quantitative Genetics*,, pp. 122–144.

Felsenstein J., 1974 The evolutionary advantage of recombination. Genetics **78**: 737–756.

Feulner P. G. D., Gratten J., Kijas J. W., Visscher P. M., Pemberton J. M., Slate J., 2013 Introgression and the fate of domesticated genes in a wild mammal population. Mol. Ecol. **22**: 4210–4221.

Fledel-Alon A., Wilson D. J., Broman K., Wen X., Ober C. *et al.*, 2009 Broad-scale recombination patterns underlying proper disjunction in humans. PLoS Genet. **5**: e1000658.

Goddard M. E., Hayes B. J., 2009 Genomic Selection Based on Dense Genotypes Inferred From Sparse Genotypes. Proc. Assoc. Advmt. Anim. Breed. Genet **18**: 26–29.

Green P., Falls K., Crooks S., 1990 *Documentation for CRIMAP, version 2.4*. Washington University School of Medicine, St Louis, MO, USA.

Hassold T., Hunt P., 2001 To err (meiotically) is human: the genesis of human aneuploidy. Nat. Rev. Genet. **2**: 280–291.

Henderson C. R., 1975 Best linear unbiased estimation and prediction under a selection model. Biometrics **31**: 423–447.

Hill W. G., Robertson A., 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8**: 269–294.

Hinch A. G., Tandon A., Patterson N., Song Y., Rohland N., *et al.*, 2011 The landscape of

recombination in African Americans. Nature **476**: 170–175.

Inoue K., Lupski J. R., 2002 Molecular mechanisms for genomic disorders. Ann. Rev. Genomics Hum. Genet.: 199–242.

Jiang Y., Xie M., Chen W., Talbot R., Maddox J. F., *et al.*, 2014 The sheep genome illuminates biology of the rumen and lipid metabolism. Science (80-.). **344**: 1168–1173.

Johnston S. E., Gratten J., Berenos C., Pilkington J. G., Clutton-Brock T. H. *et al.*, 2013 Life history trade-offs at a single locus maintain sexually selected genetic variation. Nature **502**: 93–95.

Kijas J. W., Townley D., Dalrymple B. P., Heaton M. P., Maddox J. F. *et al.*, 2009 A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One **4**: e4668.

Kijas J. W., Lenstra J. A., Hayes B., Boitard S., Porto Neto L. R. *et al*.2012 Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. **10**: e1001258.

Koehler K. E., Hawley R. S., Sherman S., Hassold T., 1996 Recombination and nondisjunction in humans and flies. Hum. Mol. Genet. **5 Spec No**: 1495–1504.

Kong A., Barnard J., Gudbjartsson D. F., Thorleifsson G., Jonsdottir G. *et al*.2004 Recombination rate and reproductive success in humans. Nat. Genet. **36**: 1203–1206.

Kong A., Thorleifsson G., Stefansson H., Masson G., Helgason A. *et al*.., 2008 Sequence variants in the RNF212 gene associate with genome-wide recombination rate. Science **319**: 1398–1401.

Kong A., Thorleifsson G., Frigge M. L., Masson G., Gudbjartsson D. F. *et al*.2014 Common and low-frequency variants associated with genome-wide recombination rate. Nat. Genet. **46**: 11–6.

Lenormand T., Dutheil J., 2005 Recombination difference between sexes: a role for haploid selection. PLoS Biol. **3**: e63.

Li Y., Willer C. J., Ding J., Scheet P., Abecasis G. R., 2010 MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. **34**: 816–834.

Ma L., O'Connell J. R., VanRaden P. M., Shen B., Padhi A. *et al*.., 2015 Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. PLoS Genet **11**: e1005387.

Maddox J. F., Davies K. P., Crawford A. M., Hulme D. J., Vaiman D. *et al.*, 2001 An Enhanced Linkage Map of the Sheep Genome Comprising More Than 1000 Loci. Genome Res. **11**: 1275–1289.

Mank J. E., 2009 The evolution of heterochiasmy: the role of sexual selection and sperm competition in determining sex-specific recombination rates in eutherian mammals. Genet. Res. (Camb). **91**: 355–363.

McPhee C. P., Robertson A., 1970 The effect of suppressing crossing-over on the response to selection in Drosophila melanogaster. Genet. Res. **16**: 1–16.

Morelli M. A., Cohen P. E., 2005 Not all germ cells are created equal: Aspects of sexual dimorphism in mammalian meiosis. Reproduction **130**: 761–781.

781 Moskvina V., Schmidt K. M., 2008 On multiple-testing correction in genome-wide
782 association studies. Genet. Epidemiol. **32**: 567–573.

783 Muller H., 1964 The relation of recombination to mutational advance. Mutat. Res. **1**: 2–9.

784 Muñoz-Fuentes V., Marcet-Ortega M., Alkorta-Aranburu G., Linde Forsberg C., Morrell J. M
785 *et al*., 2015 Strong Artificial Selection in Domestic Mammals Did Not Result in an
786 Increased Recombination Rate. Mol. Biol. Evol. **32**: 510–523.

787 Nagamine Y., Pong-Wong R., Navarro P., Vitart V., Hayward C., *et al*., 2012 Localising Loci
788 underlying Complex Trait Variation Using Regional Genomic Relationship Mapping.
789 PLoS One **7**.

790 Nagaoka S. I., Hassold T. J., Hunt P. A., 2012 Human aneuploidy: mechanisms and new
791 insights into an age-old problem. Nat. Rev. Genet. **13**: 493–504.

792 Nagel A. C., Fischer P., Szawinski J., Rosa M. K. L., Preiss A., 2012 Cyclin G is involved in
793 meiotic recombination repair in Drosophila melanogaster. J. Cell Sci.

794 Otto S. P., Barton N. H., 2001 Selection for recombination in small populations. Evolution (N.
795 Y). **55**: 1921–1931.

796 Otto S. P., Lenormand T., 2002 Resolving the paradox of sex and recombination. Nat. Rev.
797 Genet. **3**: 252–261.

798 Parisi S., McKay M. J., Molnar M., Thompson M. A., Spek P. J. van der *et al*.., 1999 Rec8p, a
799 meiotic recombination and sister chromatid cohesion phosphoprotein of the Rad21p
800 family conserved from fission yeast to humans. Mol. Cell. Biol. **19**: 3515–3528.

801 Pausch H., Aigner B., Emmerling R., Edel C., Götz K.-U., Fries R., 2013 Imputation of high-
802 density genotypes in the Fleckvieh cattle population. Genet. Sel. Evol. **45**: 3.

803 Poissant J., Hogg J. T., Davis C. S., Miller J. M., Maddox J. F., Coltman D. W., 2010 Genetic
804 linkage map of a wild genome: genomic structure, recombination and sexual
805 dimorphism in bighorn sheep. BMC Genomics **11**: 524.

806 Ponting C. P., 2011 What are the genomic drivers of the rapid evolution of PRDM9? Trends
807 Genet. **27**: 165–71.

808 Preston B. T., Stevenson I. R., Pemberton J. M., Wilson K., 2001 Dominant rams lose out by
809 sperm depletion. Nature **409**: 681–682.

810 Reynolds A., Qiao H., Yang Y., Chen J. K., Jackson N *et al*., 2013 RNF212 is a dosage-sensitive
811 regulator of crossing-over during mammalian meiosis. Nat. Genet. **45**: 269–78.

812 Ross-Ibarra J., 2004 The evolution of recombination under domestication: a test of two
813 hypotheses. Am. Nat. **163**: 105–112.

814 Sandor C., Li W., Coppieters W., Druet T., Charlier C., Georges M., 2012 Genetic variants in
815 REC8, RNF212, and PRDM9 influence male recombination in cattle. PLoS Genet. **8**:
816 e1002854.

817 Shin J., Blay S., Brad McNeney, 2006 LDheatmap: an R function for graphical display of
818 pairwise linkage disequilibria between single nucleotide polymorphisms. J. Stat. Softw.
819 **16**: Code Snippet 3.

820 Stevenson I. R., Marrow B., Preston B. T., Pemberton J. M., Wilson K., 2004 Adaptive
821 reproductive strategies. In: Clutton-Brock TH, Pemberton JM (Eds.), *Soay Sheep:*

Dynamics and Selection in an Island Population., Cambridge Univ. Press, Cambridge, U.K., pp. 243–275.

Tankimanova M., Hultén M. A., Tease C., 2004 The initiation of homologous chromosome synapsis in mouse fetal oocytes is not directly driven by centromere and telomere clustering in the bouquet. Cytogenet. Genome Res. **105**: 172–181.

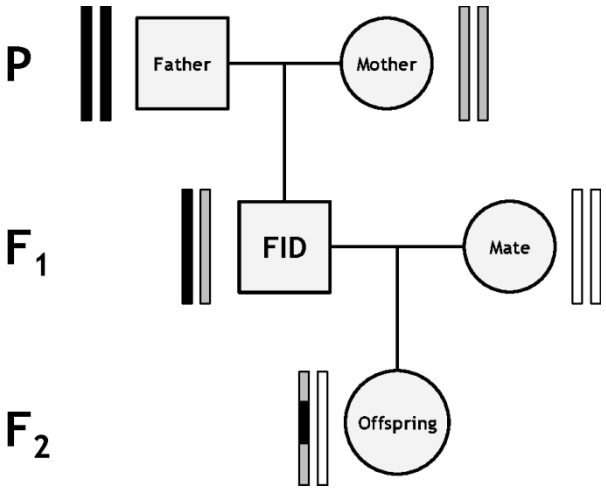
Trivers R., 1988 Sex differences in rates of recombination and sexual selection. In: Michod R, Levin B (Eds.), *The evolution of sex.*, Sinauer Press., Sunderland, MA, USA, pp. 270–286.

Yang J., Lee S. H., Goddard M. E., Visscher P. M., 2011 GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. **88**: 76–82.

Yokobayashi S., Liang C. Y., Kohler H., Nestorov P., Liu Z., *et al*.2013 PRC1 coordinates timing of sexual differentiation of female primordial germ cells. Nature **495**: 236–240.

835 # FIGURES AND TABLES

836

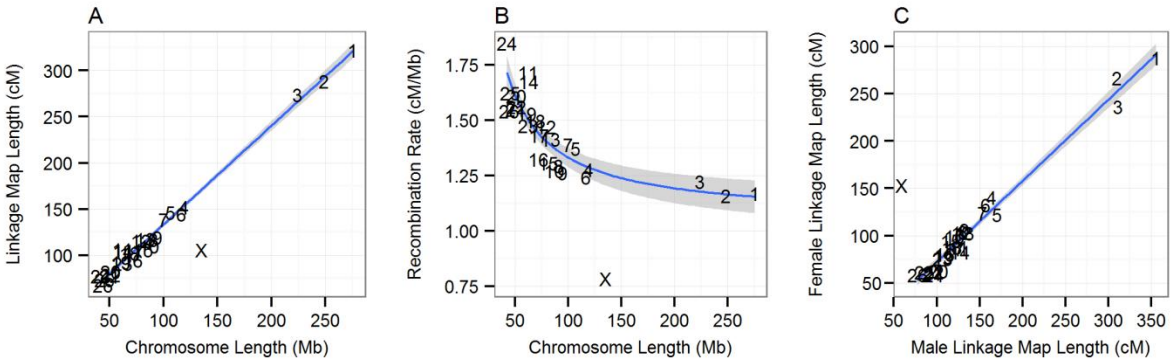


837

838

839 **Figure 1. Diagram of the sub-pedigree structure used to infer crossover events.** Rectangle
840 pairs next to each individual represent chromatids, with black and grey shading indicating
841 chromatids or chromatid sections of FID paternal and FID maternal origin, respectively.
842 White shading indicates chromatids for which the origin of SNPs cannot be determined.
843 Crossovers in the gamete transferred from the focal individual (FID) to its offspring
844 (indicated by the grey arrow) can be distinguished at the points where origin of alleles origin
845 flips from FID paternal to FID maternal and *vice versa*.

846



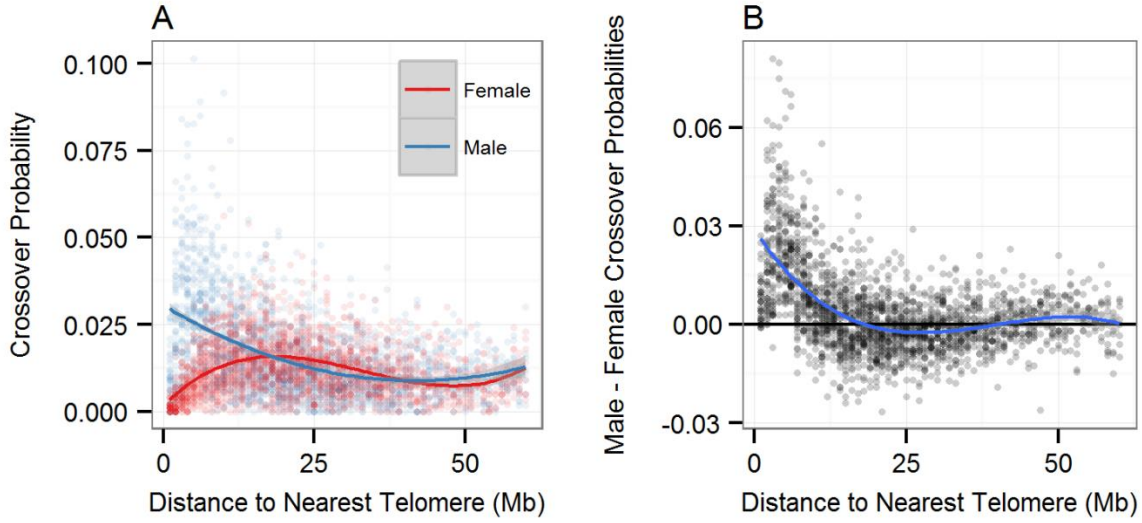
847

848

849 **Figure 2. Broad scale variation in recombination rate.**

850 Relationships between: (A) sex-averaged linkage map length (cM) and physical chromosome
851 length (Mb); (B) physical chromosome length (Mb) and recombination rate (cM/Mb); and
852 (C) male and female linkage map lengths (cM). Points are chromosome numbers. Lines and
853 the grey shaded areas indicate the regression slopes and standard errors, respectively,
854 excluding the X chromosome. NB. The male linkage map length for the X chromosome is
855 equivalent to the length of the pseudo-autosomal region.

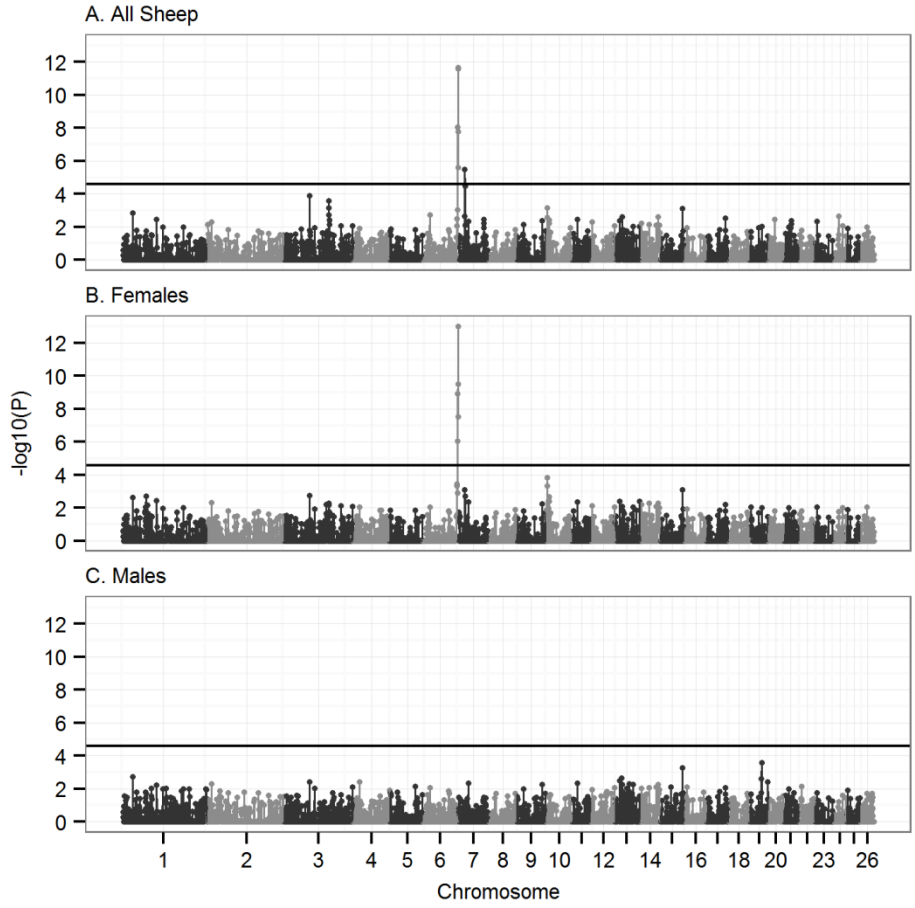
856



857

858 **Figure 3. Variation in recombination rate relative to telomeric regions.**

859 Probability of crossing over relative to the nearest telomere (Mb) for (A) female and male
860 linkage maps individually and (B) the difference between male and female crossover
861 probabilities (male minus female). Data points are given for 1Mb windows. Lines indicate
862 the function of best fit.



863

864 **Figure 4. Association analysis of autosomal crossover count using a regional heritability**

865 **approach.**

866 Significance of association analysis in (A) all sheep, (B) females only and (C) males only. The

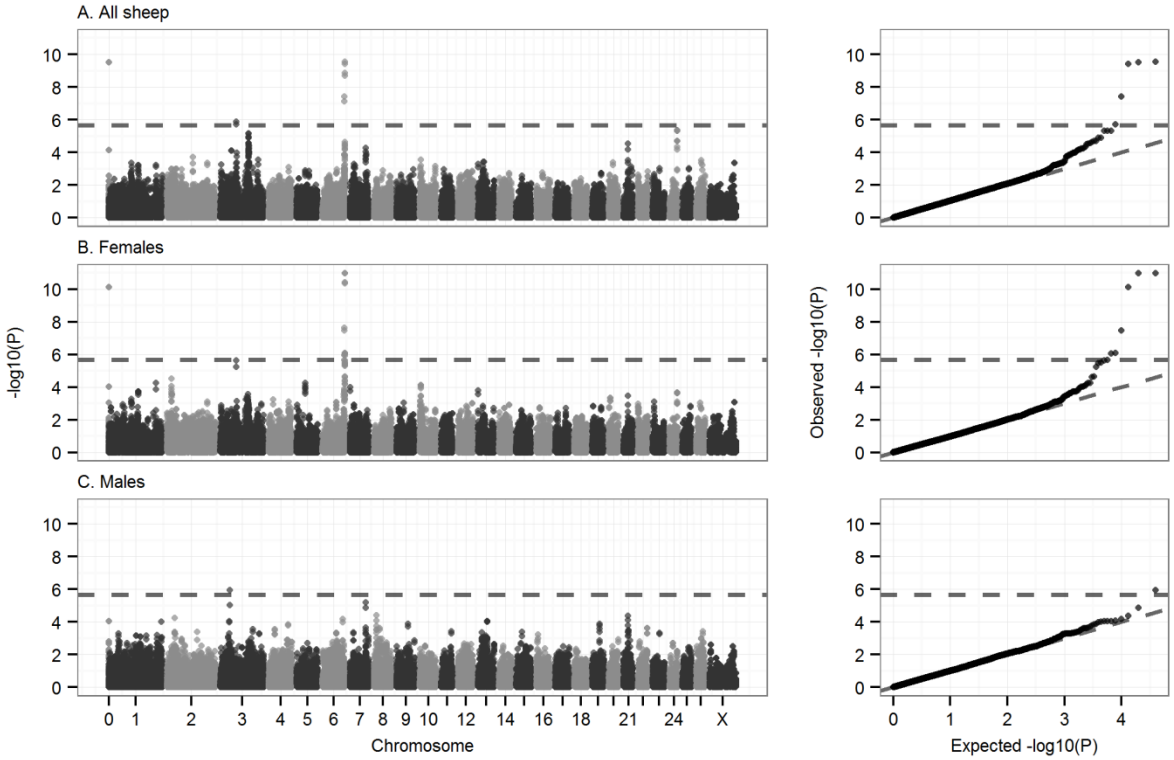
867 results presented are from a sliding window of 20 SNPs across 26 autosomes, with an

868 overlap of 10 SNPs (see main text). Points represent the median base pair position of all

869 SNPs within the sliding window. The solid black horizontal line is the significance threshold

870 after multiple testing. Underlying data is provided in Table S4.

871

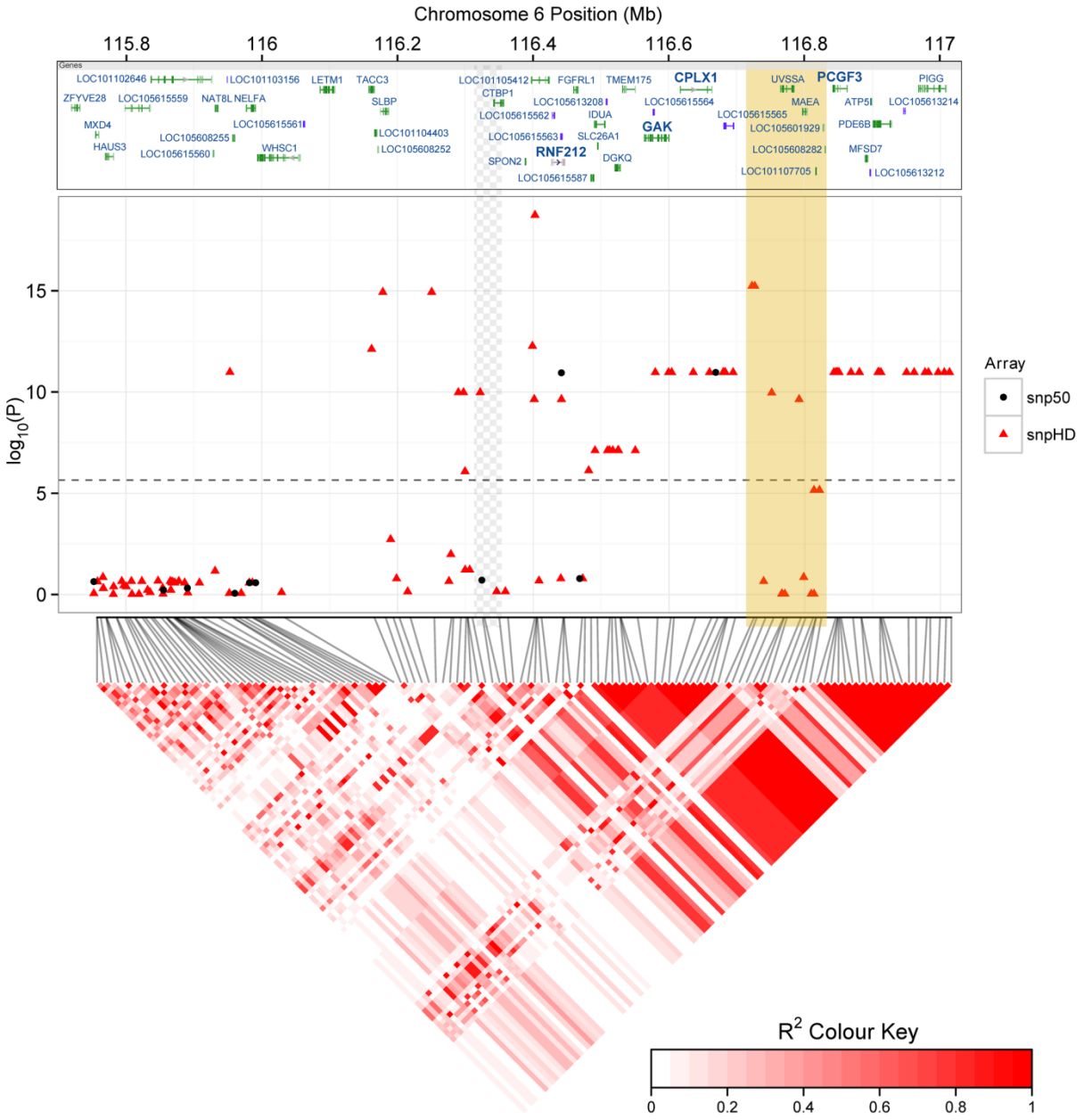


872

873 **Figure 5. Genome-wide association of autosomal crossover count.**

874 Genome-wide association statistics in (A) all sheep, (B) females only and (C) males only. The
875 dotted line indicates the threshold for statistical significance after multiple testing
876 (equivalent to an experiment-wide threshold of P = 0.05). The left column shows association
877 statistics relative to genomic position; points are color coded by chromosome. The right
878 column shows the distribution of observed P values against those expected from a null
879 distribution. Association statistics were not corrected using genomic control as λ was less
880 than one for all GWAS (λ = 0.996, 0.933 and 0.900 for plots A, B and C, respectively).
881 Underlying data on associations at the most highly associated SNPs, their genomic positions,
882 and the sample sizes are given in Table S5. The significant SNP in grey at position zero in (A)
883 and (B) occurs on an unmapped contig that is likely to correspond to the distal region of
884 chromosome 6 (see text).

885

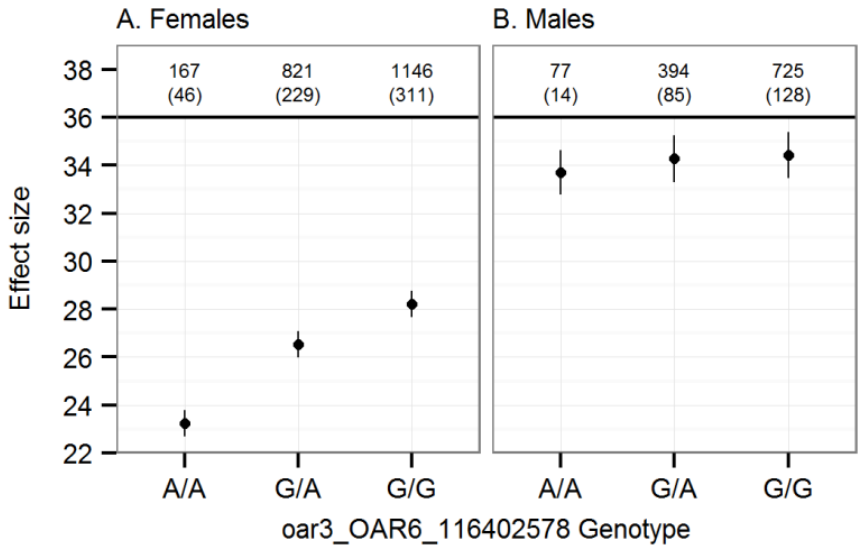


886

887 **Figure 6. Associations at the sub-telomeric region of chromosome 6.**

888 Local associations of female ACC with Ovine SNP50 BeadChip SNPs (black circles, middle
889 panel) and imputed genotypes from the Ovine HD SNP BeadChip (red triangles). The top
890 panel indicates protein coding regions within this region as provided by the NCBI Graphical
891 Sequence Viewer v3.8, with genes previously implicated in recombination or meiosis given
892 in bold text (see Introduction and (Yokobayashi *et al.* 2013; Kong *et al.* 2014)). The dashed
893 line in the middle panel indicates the significance threshold after multiple testing. The lower

894 panel is a heatmap of linkage disequilibrium in this region calculated for the 188 individuals
895 originally typed on the HD SNP chip using Spearman Rank correlation r^2 created using the R
896 library *LDheatmap* (Shin *et al.* 2006). The superimposed beige block indicates a region that is
897 likely to be incorrectly assembled on the sheep genome assembly (Oar v3.1) based on
898 sequence comparison with the homologous region of the cattle genome assembly
899 (vUMD3.1; see File S3); its position is likely to fall within the region indicated by the grey
900 chequered pattern to the left, leaving a large region of very high LD at the distal end of
901 chromosome 6.



902

903 **Figure 7. Effect sizes from a bivariate animal model of autosomal crossover count.**
904 Effect sizes are shown for (A) females and (B) males from a single bivariate model including
905 oar3_OAR6_116402578 genotype as a fixed interaction term. Error bars are the standard
906 error around the model intercept (genotype A/A) or the effect size relative to the intercept
907 (genotypes G/A and G/G). Numbers above the points indicate the number of observations
908 and the number of FIDs (in parentheses) for each genotype.

909

910 **TABLE 1. DATASET INFORMATION AND ANIMAL MODEL RESULTS OF AUTOSOMAL CROSSOVER COUNT**
911 **(ACC).** N_{OBS}, N_{FIDs} and N_{Xovers} indicate the number of ACC measures, the number of FIDs and the total number of
912 crossovers observed, respectively. Mean is that of the raw data, and V_P and V_A are the phenotypic and additive
913 genetic variances, respectively. The heritability h^2 and residual effect e^2 are the proportions of phenotype
914 variance explained by the additive genetic and residual variances, respectively. *P* (h^2) is the significance of the
915 additive genetic effect (h^2) as determined using a model comparison approach (see text). V_A and heritability
916 were modelled using genomic relatedness. Figures in brackets are standard errors.
917

Sex	N_{OBS}	N_{FIDs}	N_{Xovers}	Mean	V_P	V_A	h^2	e^2	*P* (h^2)
Both	3330	813	98420	29.56	29.56	4.28	0.15	0.85	6.88×10^{-15}
				(0.11)	(0.83)	(0.85)	(0.03)	(0.02)	
Female	2134	586	57613	27.00	31.71	5.04	0.16	0.84	4.76×10^{-12}
				(0.10)	(1.06)	(0.82)	(0.02)	(0.02)	
Male	1196	227	40807	34.12	25.21	2.97	0.12	0.88	0.022
				(0.09)	(1.16)	(0.84)	(0.03)	(0.03)	

918
919
920 **TABLE 2. TOP HITS FOR GENOME-WIDE ASSOCIATION STUDIES OF ACC IN ALL SHEEP, FEMALES AND MALES.**
921 Results provided are from Ovine SNP50 BeadChip and the most highly associated imputed SNP from
922 chromosome 6[1]. Additional loci that were significantly associated with ACC and in strong LD with these hits are
923 not shown; full GWAS results are provided in Table S5 and S6. A and B indicate the reference alleles. P values
924 are given for a Wald test of an animal model with SNP genotype fitted as a fixed effect; those in bold type
925 were genome-wide significant. V_{SNP} is the variance attributed to the SNP and Prop. V_A is the proportion of the
926 additive genetic variance explained by the SNP. Effect AB and BB are the effect sizes of genotypes AB and BB,
927 respectively, relative to the model intercept at genotype AA. The number of unique individuals for all sheep,
928 females and males are approximately N = 813, 586 and 227, respectively. Numbers in brackets are standard
929 errors.
930

SNP Information	A	B	MAF	Data	P	V_{SNP}	Prop. V_A	Effect AB	Effect BB
OAR3_51273010.1	A	G	0.44	All sheep	0.22	0.02	< 0.01	-0.58	-0.46
Chr 3						(0.03)		(0.34)	(0.41)
Position: 48,101,207				Females	0.89	0.01	< 0.01	0.04	0.19
						(0.03)		(0.42)	(0.49)
				Males	**1.15×10^{-6}**	0.32	0.18	-2.82	-2.05
						(0.18)		(0.54)	(0.6)
OAR3_87207249.1	A	C	0.25	All sheep	**1.95×10^{-6}**	0.38	0.09	2.00	2.67
Chr 3						(0.27)		(0.5)	(0.52)
Position: 82,382,182				Females	5.82×10^{-6}	0.33	0.07	2.86	3.18
						(0.37)		(0.63)	(0.65)
				Males	0.04	0.28	0.08	0.36	1.38
						(0.27)		(0.82)	(0.83)
s74824.1	A	G	0.43	All sheep	**2.92×10^{-10}**	0.84	0.19	-1.46	-2.70
Chr 6						(0.26)		(0.36)	(0.42)
Position: 116,668,852				Females	**1.07×10^{-11}**	1.36	0.25	-1.68	-3.37
						(0.4)		(0.43)	(0.49)
				Males	0.55	0.03	0.01	-0.72	-0.69
						(0.09)		(0.67)	(0.72)
[1]oar3_OAR6_116402578	A	G	0.27	All sheep	**2.62×10^{-16}**	1.14	0.26	2.46	3.89
Chr 6						(0.4)		(0.46)	(0.49)
Position: 116,402,578				Females	**1.83×10^{-19}**	1.80	0.35	3.30	4.98
						(0.57)		(0.54)	(0.56)
				Males	0.73	0.03	0.01	0.58	0.74
						(0.13)		(0.97)	(0.97)

931

932

933