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 at the individual level 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 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). The heritability of autosomal recombination rate was low but significant in both sexes (h^2 = 0.16 & 0.12 in females and males, respectively). In females, 46.7% of the 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 the heritable variation in recombination rate in both sexes. Comparative analyses with 40 other sheep breeds showed that haplotypes associated with recombination rates are both old and globally distributed. Both regions have been implicated in rate variation in mice, cattle and humans, suggesting a common genetic architecture of recombination rate variation in mammals.

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, and is an important driver of diversity because it rearranges existing allelic variation to create novel haplotypes. It can prevent the accumulation of deleterious mutations by uncoupling them from linked beneficial alleles (Muller 1964; Crow and Kimura 1965), and can 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 may be associated with fitness costs; higher rates of crossing-over may increase deleterious mutations and chromosomal rearrangements (Inoue and Lupski 2002), or lead to the break-up of favorable combinations of alleles previously built up by selection, reducing the mean

71 fitness of subsequent generations (Charlesworth and Barton 1996). Therefore, the relative
72 costs and benefits of recombination are likely to vary within different contexts, leading to an
73 expectation of variation in recombination rates within and between populations (Barton
74 1998; Burt 2000; Otto and Lenormand 2002).
75
76 Recent studies of model mammal systems have shown that recombination rates vary at an
77 individual level, and that a significant proportion of variance is driven by heritable genetic
78 effects (Kong *et al.* 2004; Dumont *et al.* 2009; Sandor *et al.* 2012). In cattle, humans and mice,
79 the heritability of recombination rate is 0.22, 0.08—0.30 and 0.46. respectively, and genome-
80 wide association studies have repeatedly attributed some heritable variation to specific
81 genetic variants, including *RNF212*, *CPLX1*, *REC8* and *PRDM9,* among others (Kong *et al.* 2008;
82 Baudat *et al.* 2010; Sandor *et al.* 2012; Kong *et al*. 2014; Ma *et al.* 2015). The majority of these
83 loci appear to influence crossover frequency, may have sex-specific or sexually-antagonistic
84 effects on recombination rate (e.g. *RNF212* and *CPLX1* in humans and cattle; Kong *et al.* 2014;
85 Ma *et al.* 2015) and may be dosage dependent (e.g. *RNF212* in mice; Reynolds *et al.* 2013).
86 The locus *PRDM9* is associated with the positioning and proportion of crossovers that occur
87 in mammalian recombination hotspots (i.e. regions of the genome with particularly high
88 recombination rates; Baudat *et al.* 2010; Ma *et al.* 2015), although this locus is not functional
89 in some mammal species, such as canids (Auton *et al.* 2013). These studies suggest that
90 recombination rate has a relatively oligogenic architecture, and therefore has the potential
91 to respond rapidly to selection over relatively short evolutionary timescales.
92
93 Such studies in model systems have provided key insights into the causes of recombination
94 rate variation. However, with the exception of humans, studies have been limited to 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 compared to their progenitors (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, to determine if it is underpinned by similar or different genetic architectures, and ultimately will allow examination of the association between recombination rate and individual fitness, enabling understanding how this trait evolves 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, U.K., since the Bronze age (Clutton-Brock *et al.* 2004). 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 the relative importance of common environment and other individual effects on recombination rates (e.g. age, sex, inbreeding coefficients); 2) to determine if individual recombination rates were heritable; 3) to identify specific genetic variants associated with recombination rate variation; and 4) to determine if the genetic architecture of recombination rate variation is similar to that observed in other mammal species.

MATERIALS AND METHODS

Study population and pedigree.

Soay sheep living within the Village Bay area of the island of Hirta (57º49'N, 8º34'W) 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) > 0.01; individual SNP locus genotyping success > 0.95; individual sheep genotyping success > 0.99; and identity by state (IBS) with another individual < 0.90. Heterozygous genotypes at non-pseudoautosomal X-linked SNPs within males were scored as missing, accounting for 0.022% of genotypes. The genomic inbreeding coefficient (measure $\hat{F}_{III}$ in Yang *et al.* 2011, hereafter $\hat{F}$), was calculated for each sheep in the software GCTA v1.24.3 (Yang *et al.* 2011), using information for all SNP loci passing quality control.

143

144 **Estimation of meiotic autosomal crossover count (ACC).**

145 **Sub-pedigree construction.** To allow unbiased phasing of the SNP data, a standardized
146 pedigree approach was used to identify cross-overs that had occurred within the gametes
147 transferred from a focal individual to its offspring; hereafter, focal individual (FID) refers to
148 the sheep in which meiosis took place. For each FID-offspring combination in the Soay sheep
149 pedigree, a sub-pedigree was constructed to include both parents of the FID (Father and
150 Mother) and the other parent of the offspring (Mate), where all five individuals had been
151 genotyped (Figure 1). This sub-pedigree structure allowed phasing of SNPs within the FID, and
152 thus the identification of autosomal cross-over events in the gamete transferred from the FID
153 to the offspring (Figure 1). Sub-pedigrees were discarded from the analysis if they included
154 the same individual twice (e.g. father-daughter matings; N = 13).

155

156 **Linkage map construction and chromosome phasing.** All analyses in this section were
157 conducted using the software CRI-MAP v2.504a (Green *et al.* 1990). First, Mendelian
158 incompatibilities in each sub-pedigree were identified using the *prepare* function;
159 incompatible genotypes were removed from all affected individuals, and sub-pedigrees
160 containing parent-offspring relationships with more than 0.1% mismatching loci were
161 discarded. Second, sex-specific and sex-averaged linkage map positions (in Kosambi cM) were
162 obtained using the *map* function, where SNPs were ordered relative to their estimated
163 positions on the sheep genome assembly Oar_v3.1 (Genbank assembly ID:
164 GCA_000298735.1; Jiang *et al.* 2014). SNP loci with a map distance of greater than 3 cM to
165 each adjacent marker (10cM for the X chromosome, including PAR) were assumed to be
166 incorrectly mapped and were removed from the analysis, with the *map* function rerun until

all map distances were below this threshold; in total, 76 SNPs were assumed to be incorrectly mapped (these SNP IDs are included in archived data, see Data Availability). Third, the *chrompic* function was used to identify informative SNPs (i.e. those for which the grand-parent of origin of the allele could be determined) on chromosomes transmitted from the FID to its offspring; crossovers were deemed to have occurred where there was a switch in the grandparental origin of a SNP allele (Figure 1).

Quality control and crossover estimation in autosomes. Errors in determining the grand-parental origin of alleles can lead to false calling of double-crossovers (i.e. two adjacent crossovers occurring on the same chromatid) and in turn, an over-estimation of recombination rate. To reduce the likelihood of calling false crossover events, runs of grandparental-origin consisting of a single allele (i.e. resulting in a double crossover either side of a single SNP) were recoded as missing (N = 973 out of 38592 double crossovers, Figure S1). In remaining cases of double crossovers, the base pair distances between immediately adjacent SNPs spanning a double crossover were calculated (hereafter, "span distance"; Figure S1). Informative SNPs that occurred within double-crossover segments with a $\log_{10}$ span distance lower than 2.5 standard deviations 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 chromosomes) 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; File S1).

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, with each bin containing a mean of 15.6 SNPs (SD = 4.04). 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.

Factors affecting autosomal recombination rate, including heritability and cross-sex genetic correlations.

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. narrow-sense heritability, h^2, hereafter 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 adjusted using the argument *--grm-adj 0*, which assumes that frequency spectra of genotyped and causal loci

239 are similar; matrices with and without adjustment were highly correlated ($R^2 > 0.997$) and
240 variance components estimated from models with and without adjustment were highly
241 similar, suggesting that adjusting for sampling error in this way did not introduce bias.
242 Matrices were not pruned to remove related individuals (i.e. the *--grm-cutoff* option was not
243 used) as there is substantial relatedness within this population, and models included common
244 environment and parental effects, controlling for some consequences shared environment
245 amongst relatives (see below). Trait variance was analyzed first with the following univariate
246 model:
247
248 $$y = X\beta + Z_1 a + Z_r u_r + e$$
249
250 where y is a vector of the ACC per transferred gamete; X is an incidence matrix relating
251 individual measures to a vector of fixed effects, *beta*; Z_1 and Z_r are incidence matrices relating
252 individual measures with additive genetic effects and random effects, respectively; a and u_r
253 are vectors of additive genetic effects from the genomic relatedness matrix and additional
254 random effects, respectively; and e is a vector of residual effects. The heritability (h^2) was
255 calculated as the ratio of the additive genetic variance to the sum of the variance estimated
256 for all random effects. Model structures were initially tested with a number of fixed effects,
257 including sex, $\hat{F}$ and FID age at the time of meiosis; random effects tested included: individual
258 identity to account for repeated measures within the same FID (sometimes referred to as the
259 permanent environment effect); maternal and paternal identity; and common environment
260 effects of FID birth year and offspring birth year. Significance of fixed effects was determined
261 using a Wald test, whereas significance of random effects was calculated using likelihood ratio
262 tests (LRT) between models with and without the focal random effect. Only sex and additive

263 genetic effects were significant in any model; however, $\hat{F}$ and individual identity were
264 retained in all models to account for potential underestimation of ACC and the effects of
265 pseudoreplication, respectively.
266
267 To investigate if the additive genetic variation underlying male and female ACC was associated
268 with sex-specific variation in ACC, bivariate models were run. The additive genetic correlation
269 r_A was determined using the CORGH error structure function in ASReml-R, (correlation with
270 heterogeneous variances) with r_A set to be unconstrained; models fitted sex-specific
271 inbreeding coefficients and individual identity effects. To test whether the genetic correlation
272 was significantly different from 0 and 1, the unconstrained model was compared to models
273 with r_A fixed at a value of 0 or 0.999. Differences in additive genetic variance in males and
274 females were tested by constraining both to be equal values using the CORGV error structure
275 function in ASReml-R. Models were then compared using likelihood ratio tests with 1 degree
276 of freedom.
277
278 **Genetic architecture of autosomal crossover count.**
279 **Genome-wide association study of variants controlling ACC.** Genome-wide association
280 studies (GWAS) of autosomal recombination rates under different scenarios were conducted
281 using ASReml-R (Butler *et al.* 2009) in R v3.1.1, fitting individual animal models for each SNP
282 locus using the same model structure as above. SNP genotypes were fitted as a fixed effect
283 with two or three levels. The GRM was replaced with a relatedness matrix based on pedigree
284 information to speed up computation; the pedigree and genomic relatedness matrices have
285 been shown to be highly correlated (Bérénos *et al.* 2014). Sex-specific models were also run.
286 Association statistics were corrected for any population stratification not captured by the

287 animal model by dividing them by the genomic control parameter, λ (Devlin *et al.* 1999), when
288 λ > 1, which was calculated as the median Wald test χ_2^2 divided by the median χ_2^2, expected
289 from a null distribution. The significance threshold after multiple testing was determined
290 using a linkage disequilibrium-based method (outlined in Moskvina and Schmidt 2008) using
291 a sliding window of 50 SNPs; the effective number of tests in the GWAS analysis was 22273.61,
292 meaning the significance threshold for P after multiple testing at α = 0.05 was 2.245×10^{-6}.
293 Although sex chromosome recombination rate was not included in the analysis, all GWAS
294 included the X chromosome and SNP markers of unknown position (N=314). The proportion
295 of phenotypic variance attributed to a given SNP was calculated using the following equation
296 (Falconer and Mackay 1996):
297
298 $$V_{SNP} = 2pq[a + d(q - p)]^2$$
299
300 where *p* and *q* are the frequencies of alleles A and B at the SNP locus, *a* is half the difference
301 in the effect sizes estimated for the genotypes AA and BB, and *d* is the difference between *a*
302 and the effect size estimated for genotype AB when fitted as a fixed effect in an animal model.
303 The proportion of heritable variation attributed to the SNP was calculated as the ratio of V_{SNP}
304 to the sum of V_{SNP} and the additive genetic variance estimated from a model excluding the
305 SNP as a fixed effect. Standard errors of V_{SNP} were estimated using a delta method approach.
306 Gene annotations in significant regions were obtained from Ensembl (gene build ID:
307 Oar_v3.1.79; Cunningham *et al.* 2014). The position of a strong candidate locus, *RNF212* is
308 not annotated on Oar_v3.1, but sequence alignment indicated that it is positioned at the sub-
309 telomere of chromosome 6 (see File S2).
310

Genome partitioning of genetic variance (regional heritability analysis).

Although a powerful tool to detect regions of the genome underlying heritable traits, the single locus approach of GWAS has reduced power to detect rare variants and variants with small effect sizes (Yang *et al.* 2011; Nagamine *et al.* 2012). One solution to this is to use a regional heritability approach that incorporates the effects of multiple haplotypes and determines the proportion of phenotypic variance explained by defined regions of the genome. The contribution of specific genomic regions to trait variation was determined by partitioning the additive genetic variance across all autosomes as follows (Nagamine *et al.* 2012):

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

where *v* is the vector of additive genetic effects explained by an autosomal genomic region *i*, and *nv* is the vector of the additive genetic effects explained by all remaining autosomal markers outwith region *i*. Regional heritabilities were determined by constructing genomic relatedness matrices (GRMs) for regions of *i* of increasing resolution (whole chromosome partitioning, sliding windows of 150, 50 and 20 SNPs, corresponding to regions of 9.41 ± 1.42, 3.12 ± 0.60 and 1.21Mb mean ± 0.32 SD length, respectively) and fitting them in models with an additional GRM of all autosomal markers not present in region *i*; sliding windows overlapped by half of their length (i.e. 75, 25 and 10 SNPs, respectively). GRMs were constructed in the software GCTA v1.24.3 and were adjusted using the *–grm-adj 0* argument (see above; Yang *et al.* 2011). Adjusted and unadjusted matrices were highly correlated, but unadjusted matrices had higher incidences of negative pivots at the regional level. In cases where both models with adjusted and unadjusted matrices were used, there was little

335 variation in estimated variance components, again suggesting that estimates were unbiased.
336 The significance of additive genetic variance attributed to a genomic region *i* was tested by
337 comparing models with and without the Z_1v_i term using a likelihood ratio test; in cases where
338 the heritability estimate was zero (i.e. estimated as "Boundary" by ASReml), significant model
339 comparison tests were disregarded. A Bonferroni approach was used to account for multiple
340 testing across the genome, by taking the number of tests and dividing by two to account for
341 the overlap of the sliding windows (since each genomic region was modelled twice).
342
343 **Accounting for cis- and trans- genetic variants associated with recombination rate**. In the
344 above analyses, we wished to separate potential associations with ACC due to cis-effects (i.e.
345 genetic variants that are in linkage disequilibrium with polymorphic recombination hotspots)
346 from those due to trans-effects (i.e. genetic variants in LD with genetic variants that affect
347 recombination rate globally). By using the total ACC within a gamete, we incorporated both
348 cis- and trans- effects into a single measure. To examine trans-effects only, we determined
349 associations between each SNP and ACC minus crossovers that had occurred on the
350 chromosome on which the SNP occurred e.g. for a SNP on chromosome 1, association was
351 examined with ACC summed across chromosomes 2 to 26. We found that in this case,
352 examining trans-variation (ACC minus focal chromosome) obtained similar results to cis- and
353 trans-variation (ACC) for both regional heritability and genome-wide association analyses,
354 leading to the same biological conclusions.
355
356 **Linkage disequilibrium and imputation of genotypes in significant regions.** A reference
357 population of 189 sheep was selected and genotyped at 606,066 SNP loci on the Ovine
358 Infinium® HD SNP BeadChip for imputation of genotypes into individuals typed on the 50K

chip. Briefly, the reference population was selected iteratively to maximize $\sum_{i=1}^{m} p_i$ using the equation $p_m = A_m^{-1} c_m$, where p is a vector of the proportion of genetic variation in the population captured by m selected animals, A_m is the corresponding subset of a pedigree relationship matrix and c is a vector of the mean relationship of the m selected animals (as outlined in Pausch *et al.* 2013 & Goddard and Hayes 2009). This approach should capture the maximum amount of genetic variation within the main population for the number of individuals in the reference population. SNP loci were retained if call rate was > 0.95 and MAF > 0.01 and individuals were retained if more than 95% of loci were genotyped. Linkage disequilibrium (LD) between loci was calculated using Spearman-Rank correlations (r^2) in the 188 individuals passing quality control.

Genotypes from the HD SNP chip were imputed to individuals typed on the SNP50 chip in the chromosome 6 region significantly associated with ACC, using pedigree information in the software MaCH v1.0.16 (Li *et al.* 2010). This region contained 10 SNPs from the Ovine SNP50 BeadChip and 116 additional independent SNPs on the HD SNP chip. As the software 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,

383 individuals with < 0.99 imputed genotypes scored were removed from the analysis. Loci with
384 < 0.95 of individuals typed were also discarded. Imputation accuracy was calculated for all loci
385 as the proportion of imputed genotypes matching their true genotypes; all remaining loci had
386 imputation accuracies > 0.95.
387
388 **Haplotype sharing of associated regions with domesticated breeds.**
389 A recent study has shown that Soay sheep are likely to have experienced an introgression
390 event with a more modern breed (the Old Scottish Shortwool, or Dunface breed, now extinct)
391 approximately 150 years ago (Feulner *et al.* 2013). Therefore, we wished to determine if
392 alleles at the most highly associated imputed SNP, oar3_OAR6_116402578 (see Results), had
393 recently introgressed into the population by examining haplotype sharing (HS) between Soay
394 sheep and Boreray sheep, a cross between Dunface and Scottish Blackface sheep. We used
395 data from the OvineSNP50 BeadChip data for Soays and a further 2709 individuals from 73
396 different sheep breeds (provided by the International Sheep Genomics Consortium, ISGC; see
397 Kijas *et al.* 2012, Feulner *et al.* 2013 and Table S1). In both the Soay and non-Soay datasets of
398 the Ovine SNP50 BeadChip, we extracted 58 SNPs corresponding to ~4Mb of the sub-
399 telomeric region on chromosome 6 and phased them using Beagle v4.0 (Browning and
400 Browning 2007). We identified core haplotypes of 6 SNP loci that tagged different alleles at
401 oar3_OAR6_116402578. The length of HS between the core Soay haplotypes and non-Soay
402 breeds was then calculated as follows: for each core haplotype *i* and each sheep breed *j*, any
403 haplotypes containing *i* were extracted, and the distance from *i* to the first mismatching SNP
404 downstream *i* was recorded. This was repeated for all pairwise comparisons of Soay and non-
405 Soay haplotypes to determine a mean and standard deviation of HS between *i* and breed *j*.
406

Data availability statement.

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. The [Dryad] repository contains genomic data before and after quality control measures, pedigree information and sub-pedigree structures, autosomal genomic relatedness matrices, population-wide crossover probabilities and individual recombination rate results. All scripts for analysis are provided on a GitHub repository at https://github.com/susjoh/GENETICS_2015_185553.

RESULTS

Broad-scale variation in recombination landscape.

We used information from 3330 sub-pedigrees 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 from 813 unique focal individuals to 3330 offspring; this included 2134 offspring from 586 unique females and 1196 offspring from 227 unique males. 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 varied 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 18.11Mb 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 may be crossover interference within this population, with a median distance between double crossovers of 48Mb (Figure S1).

Variation in individual recombination rate.

Individual autosomal crossover count (ACC) was heritable ($h^2 = 0.145$, 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 ($P_{LRT} = 0.040$) and higher residual variance ($P_{LRT} =$

454 1.11×10^{-3}) in ACC than males (Table 1). There was no relationship between ACC and FID age,
455 offspring sex, and the genomic inbreeding coefficient of the FID or offspring; furthermore,
456 there was no variance in ACC explained by common environmental effects such as FID birth
457 year, year of gamete transmission, or maternal/paternal identities of the FID (Animal Models,
458 $P > 0.05$). A bivariate model of male and female ACC showed that the cross-sex additive
459 genetic correlation (r_A) was 0.826 (SE = 0.260); this correlation was significantly different from
460 0 ($P_{LRT} = 1.14 \times 10^{-3}$) but not different from 1 ($P_{LRT} = 0.551$).
461
462 **Genetic architecture of recombination rate.**
463 **Genome-wide association study (GWAS).** The most significant association between SNP
464 genotype and ACC in both sexes was at s74824.1 in the sub-telomeric region of chromosome
465 6 ($P = 2.92 \times 10^{-10}$, Table 2). Sex-specific GWAS indicated that this SNP was highly associated
466 with female ACC ($P = 1.07 \times 10^{-11}$), but was not associated with male ACC (P = 0.55; Table 2,
467 Figure 4); the SNP had an additive effect on female ACC, with a difference of 3.37 (S.E. = 0.49)
468 autosomal crossovers per gamete between homozygotes (Table 2). This SNP was the most
469 distal typed on the chromosome from the Ovine SNP50 BeadChip at ~116.7Mb (Figure 4,
470 Table 2), and corresponded to a genomic region containing ring finger protein 212 (*RNF212*)
471 and complexin 1 (*CPLX1*), two loci that have previously been implicated in recombination rate
472 variation in humans, cattle and mice (Kong *et al.* 2008, 2014; Sandor *et al.* 2012; Reynolds *et*
473 *al.* 2013; Ma *et al.* 2015). A further SNP on an unmapped genomic scaffold (1.8kb, NCBI
474 Accession: AMGL01122442.1) was also highly associated with female ACC (Figure 4). BLAST
475 analysis indicated that the most likely genomic position of this SNP was at ~113.8Mb on
476 chromosome 6, corresponding to the same sub-telomeric region.
477

Two further regions on chromosome 3 were associated with ACC using the GWAS approach. A single SNP, OAR3_51273010.1, was associated with ACC in males, but not in females, and had an approximately dominant effect on ACC ($P = 1.15\times 10^{-6}$, Figure 4, Table 2); This SNP was 17.8kb from the 3' UTR of leucine rich repeat transmembrane neuronal 4 (*LRRTM4*) in an otherwise gene poor region of the genome (i.e. the next protein coding regions are > 1Mb from this SNP in either direction). A second SNP on chromosome 3, OAR3_87207249.1, was associated with ACC in both sexes ($P = 1.95\times 10^{-6}$, Figure 4, Table 2). This SNP was 137kb from the 5' end of an orthologue of WD repeat domain 61 (*WDR61*) and 371kb from the 5' end of an orthologue of ribosomal protein L10 (*RPL10*). Full results of GWAS are provided in Table S4.

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 S5). There was a strong sex-specific association of ACC in females within a sub-telomeric region on chromosome 6 (20 SNP sliding window; Figure 5B). This corresponded to a 1.46 Mb segment containing ~37 protein coding regions, including *RNF212* and *CPLX1*. The 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% of phenotypic variance, SE = 1.2%, $P_{LRT} = 0.82$; Figure 5C, Table S5). 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 5A, Table S5); this region had not shown any significant associations using the GWAS approach alone. The region explained 4.12% of

502 phenotypic variance (SE = 2.3%) and 26.2% of the additive genetic variance in both sexes
503 combined; however, in sex-specific models, the significant association with ACC did not
504 remain after correction for multiple testing (Table S5). No association was observed in the
505 regional heritability analysis with the two regions on chromosome 3 identified using the
506 GWAS approach. Full results for the regional heritability analysis are provided in Table S5.
507
508 **Accounting for cis- and trans- genetic variants associated with recombination rate**. The
509 results presented above were ACC incorporated both cis- and trans-effects on recombination
510 rate. When repeated with trans-effects only, all variants associated with ACC in both the
511 GWAS and regional heritability analyses remained significant (see Materials and Methods,
512 Tables S4 & S5), meaning that they are likely to affect recombination rate globally (i.e. trans-
513 acting effects), rather than being in LD with polymorphic recombination hotspots.
514
515 **Genotype imputation and association analysis at the sub-telomeric region of chromosome**
516 **6.** Genotyping of 187 sheep at a further 122 loci in the sub-telomeric region of chromosome
517 6 showed that this region has elevated levels of linkage disequilibrium, with the two most
518 significant SNPs from the 50K chip tagging a haplotype block of ~374kB ($r^2 > 0.8$; see File S3,
519 Figure 6, Table S6). This block contained three candidate genes, complexin 1 (*CPLX1*), cyclin-
520 G-associated kinase (*GAK*) and polycomb group ring finger 3 (*PCGF3*) and was 177kb away
521 from the candidate locus *RNF212* (Kong *et al.* 2014). SNP genotypes were imputed for all
522 individuals typed on the 50K chip at these 122 loci, and the association analysis was repeated.
523 The most highly associated SNP (oar3_OAR6_116402578, $P = 1.83 \times 10^{-19}$; Table 2, Figure 6)
524 occurred within an intronic region of an uncharacterized protein orthologous to
525 transmembrane emp24 protein transport domain containing *(TMED11*), 25.2kb from the

putative location of *RNF212* and 13kb from the 3' end of spondin 2 (*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;

550 however, this value is much lower than half that of previously identified introgressed
551 haplotypes (Feulner *et al.* 2013). The three most common haplotypes, H2, H3 and H6 (for
552 high, high and low ACC, respectively) are found in many other sheep breeds across the world
553 (Figure S3), suggesting that both high and low ACC haplotypes are ancient across sheep
554 breeds.

555

556 ## DISCUSSION

557 In this study, we have shown that autosomal crossover count (ACC) is heritable in Soay sheep
558 and that variation in female ACC is strongly influenced by a genomic region containing *RNF212*
559 and *CPLX1*, loci that have previously been implicated in recombination rate variation in other
560 species. The narrow sense heritability (h^2) was 0.15 across both sexes, and was lower than
561 estimates in some mammal species (h^2 = 0.22 and 0.46 in cattle and mice, respectively;
562 Dumont *et al.* 2009; Sandor *et al.* 2012) and similar to recent estimates in humans (0.13 and
563 0.08 in females and males, respectively; Kong *et al*. 2014). ACC was 1.3 times higher in males,
564 but females had a higher proportion of heritable variation than males (h^2 = 0.16, compared to
565 h^2 = 0.12). Here, we discuss the genetic architecture of the trait in more detail, the observation
566 of sexual dimorphism and male-biased recombination rates, and how our findings inform the
567 broader topic of understanding the genetic architecture of recombination rates in mammals.

568

569 **Genetic variants associated with individual recombination rate.**

570 The majority of variants associated with ACC in this study have previously been implicated in
571 recombination rate variation in other mammal species, suggesting a shared genetic
572 architecture across taxa. The strongest association was observed at the locus *RNF212*,

573 occurring 88.4kB from a ~374kb block of high LD ($r^2 > 0.8$) containing three further candidate
574 loci, *CPLX1*, *GAK* and *PCGF3* (Figure 6). Both *RNF212* and *CPLX1* have been associated with
575 recombination rate variation in mammals (Kong *et al.* 2008; Sandor *et al.* 2012; Reynolds *et*
576 *al.* 2013; Ma *et al.* 2015) and mouse studies have established that the protein RNF212 is
577 essential for the formation of crossover-specific complexes during meiosis, and that its effect
578 is dosage-sensitive (Reynolds *et al.* 2013). We observed an additive effect of the *RNF212*
579 region on female recombination rate (Figure 7), suggesting that dosage dependence could be
580 a plausible mechanism driving rate differences in Soay sheep. GAK forms part of a complex
581 with cyclin-G, a locus involved in meiotic recombination repair in *Drosophila* (Nagel *et al.*
582 2012), and PCGF3 forms part of a PRC1-like complex (polycomb repressive complex 1) which
583 is involved in meiotic gene expression and the timing of meiotic prophase in female mice
584 (Yokobayashi *et al.* 2013). High LD within this region meant that it was not possible to test the
585 effects of these loci on recombination rate independently; however, the co-segregation of
586 several loci affecting meiotic processes may merit further investigation to determine if
587 recombination is suppressed in this region, and if this co-segregation is of adaptive
588 significance.
589
590 Additional genomic regions associated with recombination rate included: two loci at 48.1Mb
591 and 82.4Mb on chromosome 3 (identified using GWAS) with effects on males only and both
592 sexes, respectively; and a 1.09Mb region of chromosome 7 affecting rates in both sexes
593 (identified using regional heritability analysis). Although the chromosome 7 region was large
594 and specific loci cannot be pinpointed, it contained *REC8*, the protein of which is required for
595 the separation of sister chromatids and homologous chromosomes during meiosis (Parisi *et*
596 *al.* 1999); and *RNF212B*, a paralogue of *RNF212*. The same region is also associated with

597 recombination rate in cattle (Sandor *et al.* 2012). The chromosome 3 variants identified were
598 novel to this study, and occurred in relatively gene poor regions of the genome (see above).
599
600 Although there are homologues of *PRDM9* on chromosomes 1, 5, 18 and X, it is not currently
601 known whether any of these copies are functional in sheep. Here, we did not identify any
602 association between recombination rate and any of these regions using either GWAS or
603 regional heritability approaches. This may not be surprising, as this locus is primarily
604 associated with recombination hotspot usage. Nevertheless, *PRDM9* has been associated
605 with recombination rate in cattle and male humans (Kong *et al*. 2014; Ma *et al.* 2015) and is
606 likely to be consequence of differences in the abundance of motifs recognized by the PRDM9
607 protein in hotspots rather than the locus itself affecting rate. In the current study, it was not
608 possible to examine hotspot usage, as crossovers could only be resolved to a median interval
609 of 800Mb. This is unlikely to be of the fine scale required to characterize hotspot variation, as
610 they typically occur within 1-2kb intervals in mammals (Paigen and Petkov 2010). Further
611 studies would require higher densities of markers to determine crossover positions at a
612 greater resolution and to determine the functionality and/or relative importance of *PRDM9*
613 within this system.
614
615 **Sexual dimorphism in genetic architecture of recombination rate.**
616 This study identified sexual dimorphism in the genetic architecture of recombination rate in
617 Soay sheep. Using a classical quantitative genetic approach, the between-sex genetic
618 correlation was not significantly different from 1, indicating that male and female
619 recombination rate variation had a shared genetic basis – albeit with a relatively large error
620 around this estimate. However, females had significantly higher additive genetic and residual

621 variance in the trait in comparison to males, and GWAS and regional heritability showed that
622 the *RNF212/CPLX1* region was associated with female recombination rate only. This is
623 consistent with previous studies, where this region was associated with sexually dimorphic
624 and sexually antagonistic variation in recombination rate in cattle and humans, respectively
625 (Kong *et al.* 2014; Ma *et al.* 2015). Therefore, our findings suggest that variation in
626 recombination rate has *some* degree of a shared and distinct genetic architecture between
627 the sexes, which may be expected due to various similarities of this process of meiosis, but
628 differences in its implementation within each sex (discussed further below). There were some
629 differences in sample sizes between the sexes, with twice as many meioses characterized in
630 females than males, so it could be argued that low sample sizes in males may have had less
631 power to identify specific loci. Whilst possible, it is unlikely that the absence of associations
632 between the *RNF212*/*CPLX1* region and male recombination rate is due to low power to
633 detect the effect, as (a) models including both sexes showed reduced, rather than increased
634 significance in this region; (b) bivariate models accounting for variation in *RNF212* as a fixed
635 effect supported a sexually dimorphic genetic effect with a lower degree of error than the
636 bivariate approach (Figure 7) and (c) repeating the association analysis at the most highly
637 associated SNP using sampled datasets of identical size in males and females found
638 consistently higher association at this locus in females (Figure S4).
639
640 **How much phenotypic variation in recombination rate is explained?**
641 The approaches used in this study were successful in characterizing several regions of the
642 genome contributing to the additive genetic variance in recombination rate. The regional
643 heritability approach demonstrated some potential for characterizing variation from multiple
644 alleles and/or haplotypes encompassing both common and rare variants that are in linkage

disequilibrium with causal loci that were not detectable by GWAS alone (Nagamine *et al.* 2012). However, whilst some of the genetic contribution to phenotypic variance was explained by specific genomic regions, the overall heritability of recombination rate was low, and a substantial proportion of the heritable variation was of unknown architecture (i.e. "missing heritability"; Manolio *et al.* 2009). In females, 64% of additive genetic variance was explained by the *RNF212/CLPX1* and *RNF212B/REC8* regions combined, but this only accounted for 11% of the phenotypic variance (Table S9), leaving the remaining additive genetic and phenotypic variance unexplained. Our sample size is small relative to such studies in model systems, and there may have been reduced power to detect genetic variants, particularly in males which were under-represented in the dataset. However, our finding of low heritability and unexplained additive genetic variance are consistent with recent results from Icelandic humans, where despite a higher sample size (N = 15,253 males and 20,674 females) and marker density (N = 30.3×10^6), the fraction of phenotypic variance explained by specific loci remained small; identified variants including *RNF212* and *CPLX1* explained just 2.52 and 3.15% of male and female phenotypic variance, respectively, accounting for 29 and 24.8% of the additive genetic variance (Kong *et al*. 2014). Therefore, despite evidence of a conserved genetic architecture across mammal systems, there remains a very large proportion of both the additive genetic and phenotypic variance unexplained.

Variation and sexual dimorphism in recombination landscape.

Males had considerably higher recombination rates than in females, which was driven mainly by large differences in crossover frequencies in the sub-telomeric regions between 0 and 18.11Mb (Figure 3); recombination was reduced in males if the centromere was present in the sub-telomeric region (i.e. in autosomes 4 to 26, which are acrocentric), but unlike in cattle,

669 was still significantly higher than that of females (Figure S5; Ma *et al.* 2015). Outside the sub-
670 telomeric region, recombination rates were more similar between the sexes, with females
671 showing slightly higher recombination rates between 18.1 and 40Mb from the telomere
672 (Figure 3B). This observation of increased sub-telomeric recombination in males is consistent
673 with studies in humans, cattle and mice (Kong *et al.* 2002; Shifman *et al*. 2006; Ma *et al*. 2015),
674 although the magnitude of the difference is much greater in the Soay sheep population.
675 Within females, the rate differences associated with different genotypes at *RNF212* were
676 most clear in regions likely to be euchromatic, whereas there was no difference in rate in
677 regions likely to be heterochromatic, such as the sub-telomeric and centromeric regions
678 (Figure S6, Table S10).
679
680 **Why is recombination rate higher in males?**
681 In placental mammals, females usually exhibit higher recombination rates than males
682 (Lenormand and Dutheil 2005), and it has been postulated that this is a mechanism to avoid
683 aneuploidy after long periods of meiotic arrest (Koehler *et al.* 1996; Morelli and Cohen 2005;
684 Nagaoka *et al.* 2012). However, Soay sheep exhibited male-biased recombination rates to a
685 greater degree than observed in any placental mammal to date (Male to Female linkage map
686 ratio = 1.31). The biological significance of this remains unclear, although a number of
687 mechanisms have been proposed to explain variation in sex differences more generally,
688 including haploid selection (Lenormand and Dutheil 2005), meiotic drive (Brandvain and Coop
689 2012), sperm competition, sexual dimorphism and dispersal (Trivers 1988; Burt *et al.* 1991;
690 Mank 2009). Nevertheless, testing these ideas has been limited by a paucity of empirical data.
691

692 One possible explanation for elevated recombination in males is that Soay sheep have a highly
693 promiscuous mating system. Males have the largest testes to body size ratio within ruminants
694 (Stevenson *et al.* 2004) and experiences high levels of sperm competition, with dominant
695 rams becoming sperm-depleted towards the end of the annual rut (Preston *et al.* 2001).
696 Increased recombination may allow more rapid sperm production through formation of
697 meiotic bouquets (Tankimanova *et al.* 2004). Another argument made by Ma *et al*. (2015) to
698 explain increased recombination in male cattle (M:F linkage map ratio = 1.1) is that stronger
699 selection in males may have indirectly selected for higher recombination rates in bulls and
700 may be a consequence of domestication (Burt and Bell 1987; Ross-Ibarra 2004). Soay sheep
701 underwent some domestication before arriving on St Kilda, and have comparable levels of
702 male-biased recombination to domestic sheep (M:F linkage map ratio = 1.19; Maddox *et al.*
703 2001). In contrast, wild bighorn sheep (*Ovis canadensis*) have not undergone domestication
704 and have female-biased recombination rates (M:F linkage map ratio = 0.89; divergence
705 ~2.8Mya; Poissant *et al.* 2010). However, low marker density (N = 232 microsatellites) in the
706 bighorn sheep study may have failed to resolve crossovers in the sub-telomeric regions;
707 furthermore, a recent study of chiasma count in wild progenitor vs. domestic mammal species
708 found that recombination rates had not increased with domestication in sheep (Muñoz-
709 Fuentes *et al.* 2015). In addition, wild cattle and sheep may have higher levels of sperm
710 competition than other mammal species, supporting the former argument. Regardless, more
711 empirical studies are required to elucidate the specific drivers of sex-differences in
712 recombination rate at both a mechanistic and inter-specific level. Given our finding that the
713 *RNF212*/*CPLX1* region is involved in the sex difference in Soay sheep, there is a compelling
714 case for a role for this region in driving sex-differences in mammal systems over relatively
715 short evolutionary timescales.

716

717 **Examining recombination rates in the wild.**

718 A principle motivation for the current study was to determine how recombination rate and
719 its genetic architecture may vary relative to model species that have undergone strong
720 selection in their recent history. We found that the heritability of recombination rate in Soay
721 sheep was much lower than in cattle and mice, and was comparable with recent estimates in
722 humans, which can also be considered a wild population (Kong *et al*. 2014). Despite these
723 differences, variants affecting recombination rate on sheep chromosomes 6 and 7 have
724 previously been associated with recombination rate in other mammalian populations (see
725 above) and only the two relatively gene-poor regions identified on chromosome 3 are novel.
726 Furthermore, examination of haplotypes around *RNF212* suggests that variation at the
727 *RNF212/CPLX1* region affecting recombination rate has been segregating for a long time in
728 sheep. Examining variation in the wild also allowed us to quantify the effects of the individual
729 and environment on recombination rate; however, we found no effect of common
730 environment effects (i.e. birth year and year of gamete transmission), individual age or
731 inbreeding on recombination rate; rather, the majority of variation was attributed to residual
732 effects. This contrasts with the observation in humans that recombination rates increase with
733 age in females (Kong *et al.* 2004). Overall, our findings suggest a strong stochastic element
734 driving recombination rates, with a small but significant heritable component that has a
735 similar architecture to other mammal systems, regardless of their selective background. The
736 Soay sheep system is one of the most comprehensive wild datasets in terms of genomic
737 resources and sampling density, making it one of the most suitable in terms of quantifying
738 and analyzing recombination rate variation in the wild. A future ambition is to investigate the
739 fitness consequences of phenotypic variation in ACC and more specifically the relationship

740 between variants identified in this study and individual life-history variation, to determine if
741 the maintenance of genetic variation for recombination rates is due to selection, sexually
742 antagonistic effects, or stochastic processes.
743
744 **Conclusions**
745 In this study, we have shown that recombination rates in Soay sheep are heritable and have
746 a sexually dimorphic genetic architecture. The variants identified have been implicated in
747 recombination rates in other mammal species, indicating a conserved genetic basis across
748 distantly related taxa. However, the proportion of phenotypic variation explained by
749 identified variants was low; this was consistent with studies in humans and cattle, in which
750 although genetic variants were identified, the majority of both additive genetic and
751 phenotypic variance has remained unexplained. Similar studies in both mammalian and non-
752 mammalian wild systems may provide a broader insight into the genetic and non-genetic
753 drivers of recombination rate variation, as well its evolution in contemporary populations.
754 However, the question remains as to whether variation in recombination rate is adaptive or
755 merely a by-product of other biological processes. Overall, the approaches and findings
756 presented here provide an important foundation for studies examining the evolution of
757 recombination rates in contemporary natural populations.
758

759 ## ACKNOWLEDGEMENTS

760 We thank Jill Pilkington, Ian Stevenson and all Soay sheep project members and volunteers
761 for collection of data and samples. Discussions and comments from Jarrod Hadfield, Bill Hill,
762 Craig Walling, Jisca Huisman, John Hickey and two anonymous reviewers greatly improved
763 the analysis. James Kijas, Yu Jiang and Brian Dalrymple responded to numerous queries on

764 the genome assembly, annotation and SNP genotyping. Philip Ellis prepared DNA samples,
765 and Louise Evenden, Jude Gibson and Lee Murphy carried out SNP genotyping at the
766 Wellcome Trust Clinical Research Facility Genetics Core, Edinburgh. This work has made
767 extensive use of the resources provided by the Edinburgh Compute and Data Facility
768 (http://www.ecdf.ed.ac.uk/). Permission to work on St Kilda is granted by The National Trust
769 for Scotland and Scottish Natural Heritage, and logistical support was provided by QinetiQ
770 and Eurest. The Soay sheep project is supported by grants from the UK Natural Environment
771 Research Council, with the SNP genotyping and research presented here supported by an ERC
772 Advanced Grant to JMP.

773

774 ## REFERENCED LITERATURE

775 Akaike H., 1974 A new look at the statistical model identification. IEEE Trans. Automat.
776 Contr. **19**: 716–723.
777 Aulchenko Y. S., Ripke S., Isaacs A., Duijn C. M. van, 2007 GenABEL: an R library for genome-
778 wide association analysis. Bioinformatics **23**: 1294–1296.
779 Auton A., Rui Li Y., Kidd J., Oliveira K., Nadel J. et al 2013 Genetic Recombination Is Targeted
780 towards Gene Promoter Regions in Dogs. PLoS Genet. **9**: e1003984.
781 Barton N. H., 1998 Why sex and recombination? Science (80-.). **281**: 1986–1990.
782 Baudat F., Buard J., Grey C., Fledel-Alon A., Ober C., *et al*., 2010 PRDM9 is a major
783 determinant of meiotic recombination hotspots in humans and mice. Science **327**: 836–
784 840.
785 Bérénos C., Ellis P., Pilkington J. G., Pemberton J. M., 2014 Estimating quantitative genetic
786 parameters in wild populations: a comparison of pedigree and genomic approaches.
787 Mol. Ecol. **23**: 3434–3451.
788 Brandvain Y., Coop G., 2012 Scrambling eggs: meiotic drive and the evolution of female
789 recombination rates. Genetics **190**: 709–23.
790 Browning S. R., Browning B. L., 2007 Rapid and accurate haplotype phasing and missing-data
791 inference for whole-genome association studies by use of localized haplotype
792 clustering. Am. J. Hum. Genet. **81**: 1084–1097.
793 Burt A., Bell G., 1987 Mammalian chiasma frequencies as a test of two theories of
794 recombination. Nature **326**: 803–805.
795 Burt A., Bell G., Harvey P. H., 1991 Sex differences in recombination. J. Evol. Biol. **4**: 259–
796 277.
797 Burt A., 2000 Sex, Recombination, and the Efficacy of Selection - was Weismann Right?
798 Evolution (N. Y). **54**: 337–351.
799 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.

847 Kijas J. W., Lenstra J. A., Hayes B., Boitard S., Porto Neto L. R. *et al*.2012 Genome-wide
848 analysis of the world's sheep breeds reveals high levels of historic mixture and strong
849 recent selection. PLoS Biol. **10**: e1001258.
850 Koehler K. E., Hawley R. S., Sherman S., Hassold T., 1996 Recombination and nondisjunction
851 in humans and flies. Hum. Mol. Genet. **5 Spec No**: 1495–1504.
852 Kong A., Gudbjartsson D. F., Sainz J., Jonsdottir G. M., Gudjonsson S. A. *et al*., 2002 A high-
853 resolution recombination map of the human genome. Nat Genet. **31**:241—247.
854 Kong A., Barnard J., Gudbjartsson D. F., Thorleifsson G., Jonsdottir G. *et al*.2004
855 Recombination rate and reproductive success in humans. Nat. Genet. **36**: 1203–1206.
856 Kong A., Thorleifsson G., Stefansson H., Masson G., Helgason A. *et al*.., 2008 Sequence
857 variants in the RNF212 gene associate with genome-wide recombination rate. Science
858 **319**: 1398–1401.
859 Kong A., Thorleifsson G., Frigge M. L., Masson G., Gudbjartsson D. F. *et al*.2014 Common and
860 low-frequency variants associated with genome-wide recombination rate. Nat. Genet.
861 **46**: 11–6.
862 Lenormand T., Dutheil J., 2005 Recombination difference between sexes: a role for haploid
863 selection. PLoS Biol. **3**: e63.
864 Li Y., Willer C. J., Ding J., Scheet P., Abecasis G. R., 2010 MaCH: Using sequence and
865 genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol.
866 **34**: 816–834.
867 Ma L., O'Connell J. R., VanRaden P. M., Shen B., Padhi A. *et al*.., 2015 Cattle Sex-Specific
868 Recombination and Genetic Control from a Large Pedigree Analysis. PLoS Genet **11**:
869 e1005387.
870 Maddox J. F., Davies K. P., Crawford A. M., Hulme D. J., Vaiman D. *et al*., 2001 An Enhanced
871 Linkage Map of the Sheep Genome Comprising More Than 1000 Loci. Genome Res. **11**:
872 1275–1289.
873 Mank J. E., 2009 The evolution of heterochiasmy: the role of sexual selection and sperm
874 competition in determining sex-specific recombination rates in eutherian mammals.
875 Genet. Res. (Camb). **91**: 355–363.
876 Manolio T. A., Colling F. S., Cox N. J., Goldstein D. B., Hindorff L. A. *et al*., 2009 Finding the
877 missing heritability of complex diseases. Nature **461**: 747-753.
878 McPhee C. P., Robertson A., 1970 The effect of suppressing crossing-over on the response to
879 selection in Drosophila melanogaster. Genet. Res. **16**: 1–16.
880 Morelli M. A., Cohen P. E., 2005 Not all germ cells are created equal: Aspects of sexual
881 dimorphism in mammalian meiosis. Reproduction **130**: 761–781.
882 Moskvina V., Schmidt K. M., 2008 On multiple-testing correction in genome-wide
883 association studies. Genet. Epidemiol. **32**: 567–573.
884 Muller H., 1964 The relation of recombination to mutational advance. Mutat. Res. **1**: 2–9.
885 Muñoz-Fuentes V., Marcet-Ortega M., Alkorta-Aranburu G., Linde Forsberg C., Morrell J. M
886 *et al*., 2015 Strong Artificial Selection in Domestic Mammals Did Not Result in an
887 Increased Recombination Rate. Mol. Biol. Evol. **32**: 510–523.
888 Nagamine Y., Pong-Wong R., Navarro P., Vitart V., Hayward C., *et al*., 2012 Localising Loci
889 underlying Complex Trait Variation Using Regional Genomic Relationship Mapping.
890 PLoS One **7**.
891 Nagaoka S. I., Hassold T. J., Hunt P. A., 2012 Human aneuploidy: mechanisms and new
892 insights into an age-old problem. Nat. Rev. Genet. **13**: 493–504.
893 Nagel A. C., Fischer P., Szawinski J., Rosa M. K. L., Preiss A., 2012 Cyclin G is involved in

meiotic recombination repair in Drosophila melanogaster. J. Cell Sci.

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

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

Paigen K., Petkov P., 2010 Mammalian recombination hot spots: properties, control and evolution. Nat. Rev. Genet. **11**: 221–233.

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

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

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

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

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

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

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

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

Shifman S., Bell J. T., Copley R. R., Taylor M. S., Williams R. W. *et al*., 2006 A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol. **4**:e395.

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

Stevenson I. R., Marrow B., Preston B. T., Pemberton J. M., Wilson K., 2004 Adaptive 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.

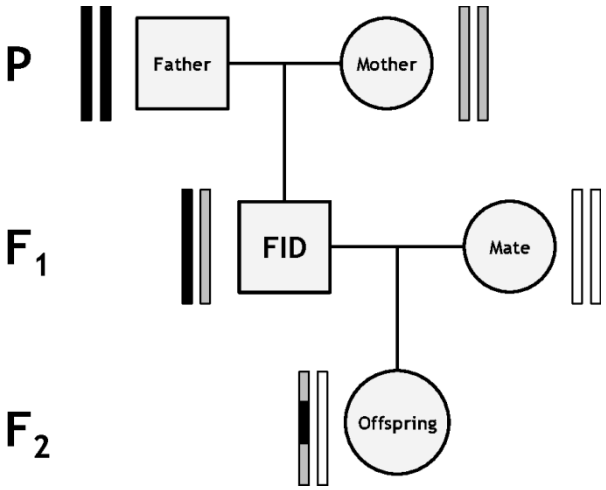
Trivers R., 1988 Sex differences in rates of recombination and sexual selection. In: Michod R, Levin B (Eds.), *The evolution of sex.*, Sinauer Press., Sunderland, MA, USA, pp. 270–286.

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.

940 # FIGURES AND TABLES

941

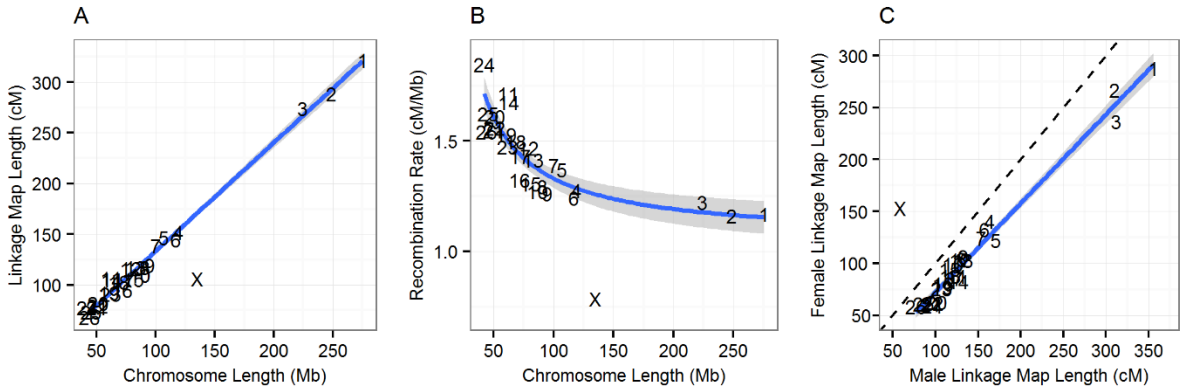


942

943

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

951



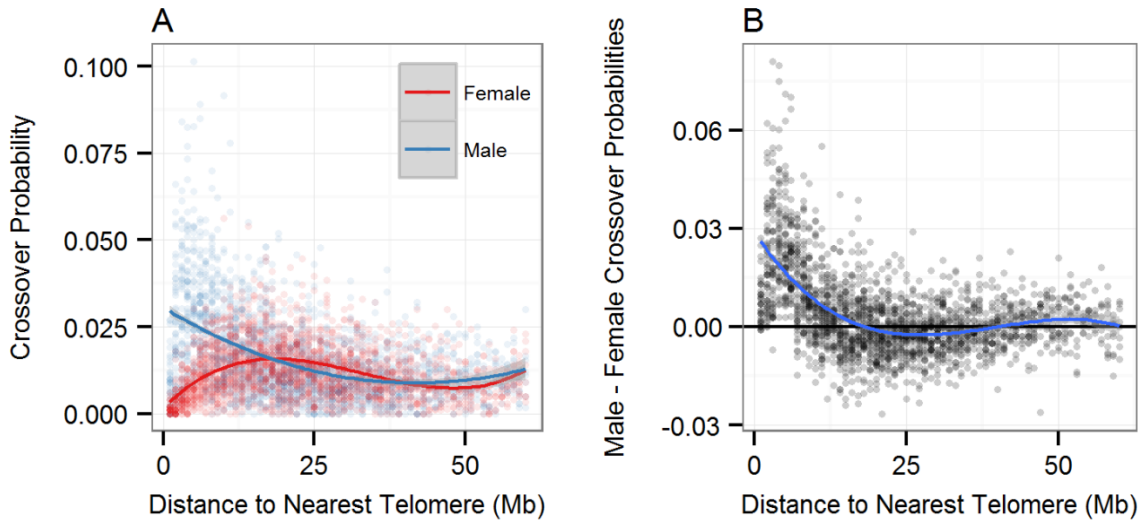
952

953

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

955 Relationships between: (A) sex-averaged linkage map length (cM) and physical chromosome
956 length (Mb); (B) physical chromosome length (Mb) and recombination rate (cM/Mb); and (C)
957 male and female linkage map lengths (cM). Points are chromosome numbers. Lines and the
958 grey shaded areas indicate the regression slopes and standard errors, respectively, excluding
959 the X chromosome. The dashed line in (C) indicates the where male and female linkage maps
960 are of equal length. NB. The male linkage map length for the X chromosome is equivalent to
961 the length of the pseudo-autosomal region.

962

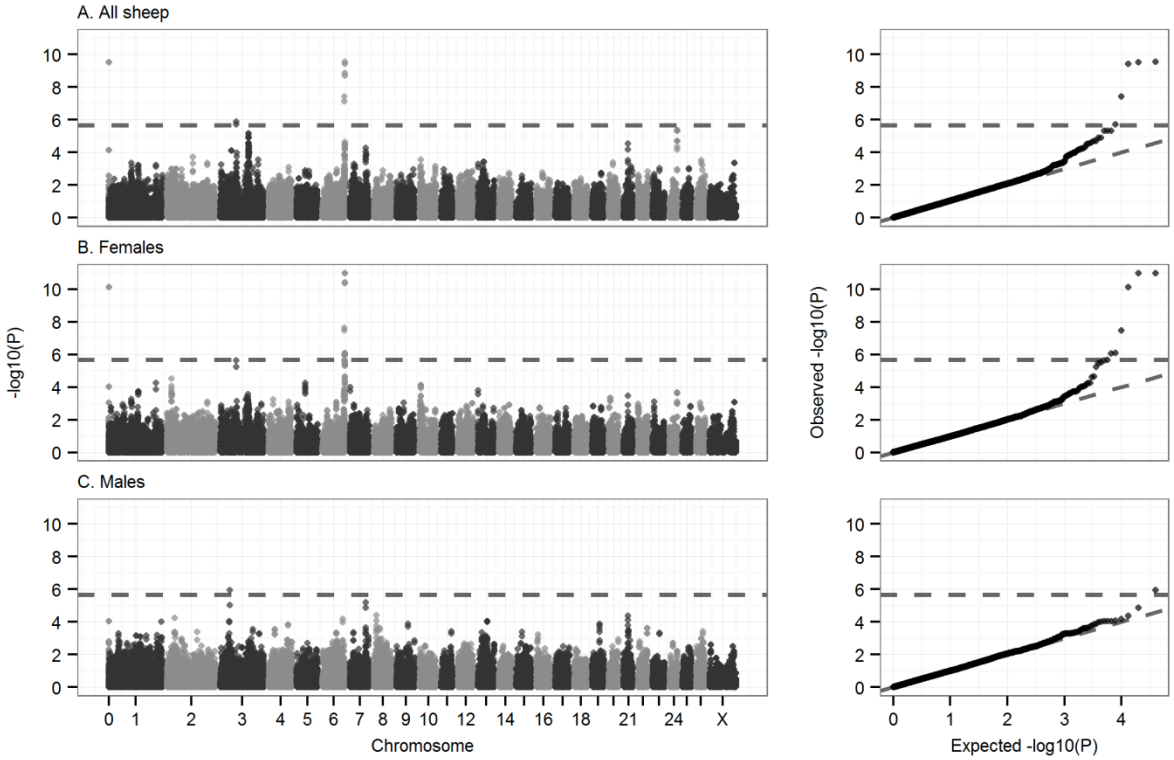


963

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

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

969



970

971 **Figure 4. Genome-wide association of autosomal crossover count.**

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

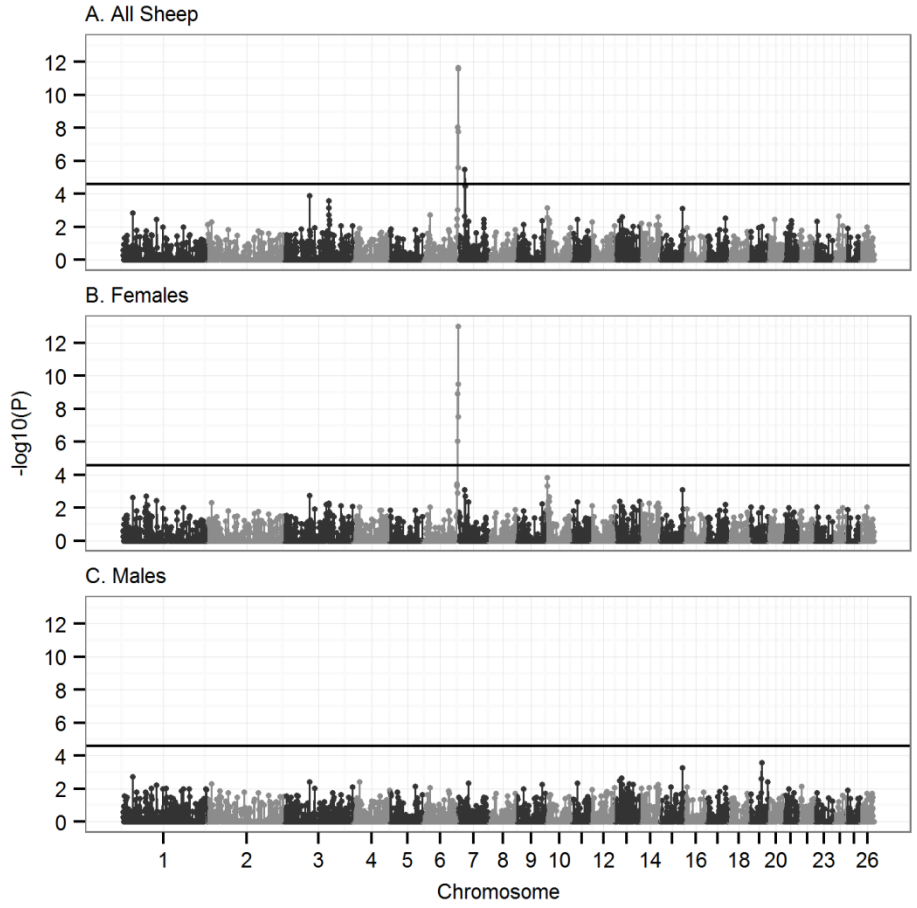


Figure 5. Regional heritability analysis of autosomal crossover count.

Significance of association analysis in (A) all sheep, (B) females only and (C) males only. The results presented are from a sliding window of 20 SNPs across 26 autosomes, with an overlap of 10 SNPs (see main text). Points represent the median base pair position of all SNPs within the sliding window. The solid black horizontal line is the significance threshold after multiple testing. Underlying data is provided in Table S4.

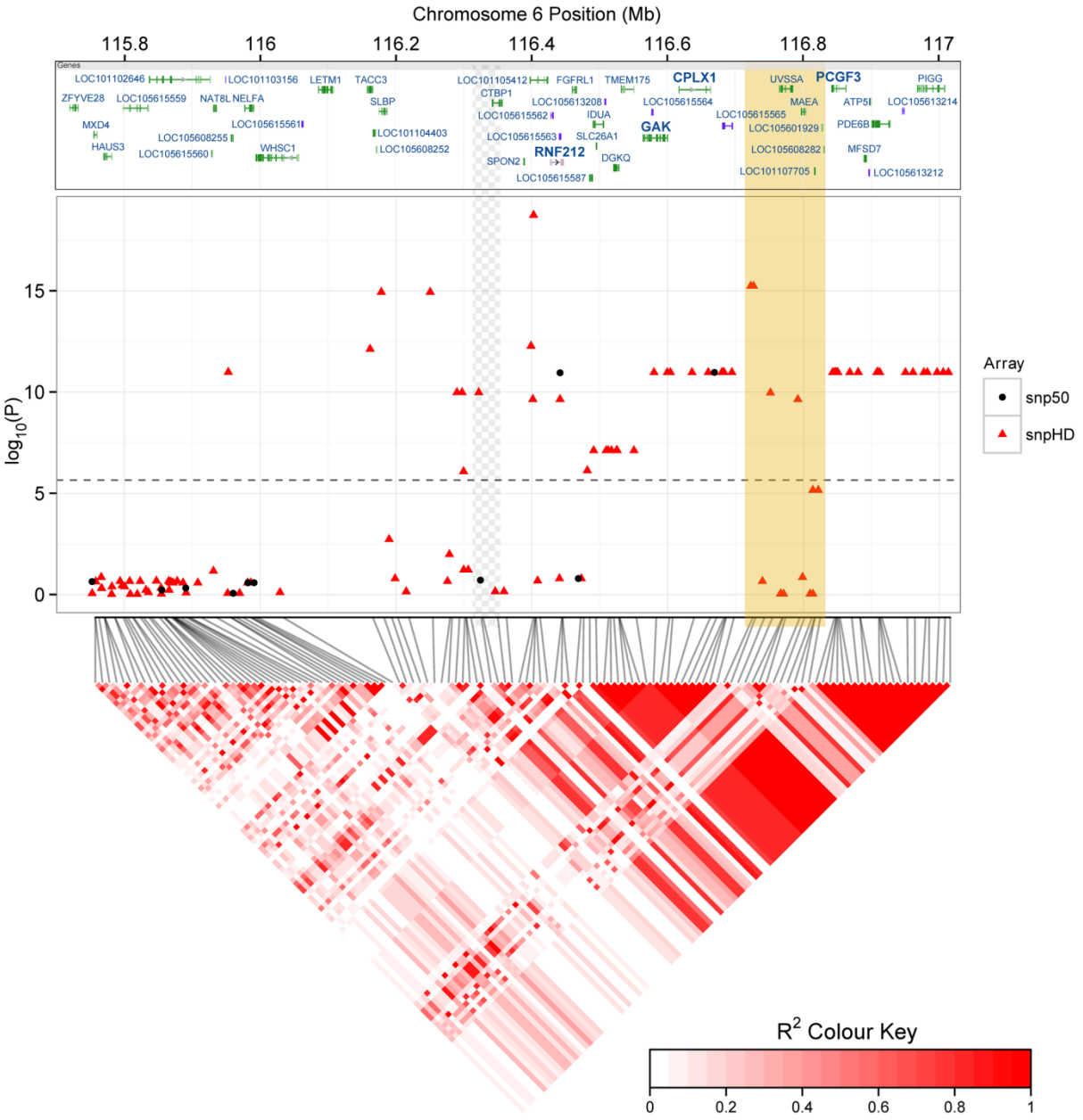
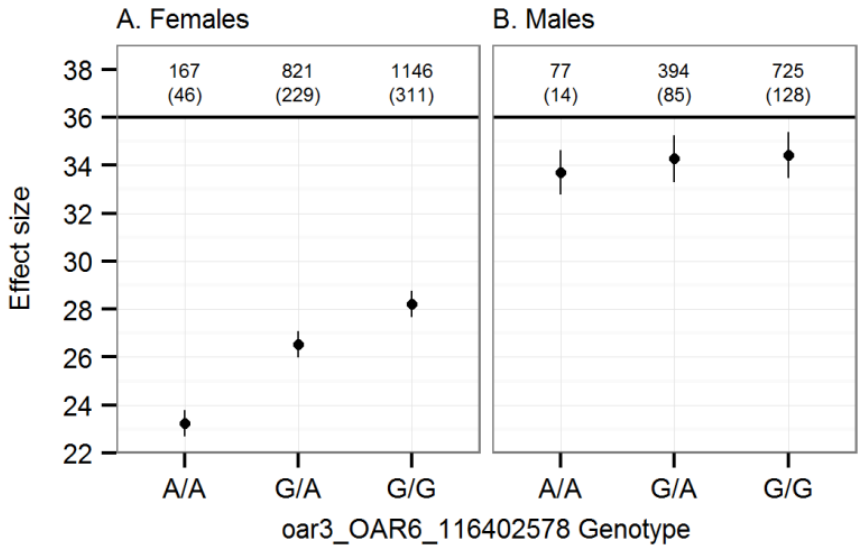


Figure 6. Associations at the sub-telomeric region of chromosome 6.

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

998 is a heatmap of linkage disequilibrium in this region calculated for the 188 individuals typed
999 on the HD SNP chip using Spearman Rank correlation r^2 created using the R library *LDheatmap*
1000 (Shin *et al.* 2006). The superimposed beige block indicates a region that is likely to be
1001 incorrectly assembled on the sheep genome assembly (Oar v3.1) based on sequence
1002 comparison with the homologous region of the cattle genome assembly (vUMD3.1; see File
1003 S3); its position is likely to fall within the region indicated by the grey chequered pattern to
1004 the left, leaving a large region of very high LD at the distal end of chromosome 6.



1005

1006 **Figure 7. Effect sizes from a bivariate animal model of autosomal crossover count.**
1007 Effect sizes are shown for (A) females and (B) males from a single bivariate model including
1008 oar3_OAR6_116402578 genotype as a fixed interaction term. Error bars are the standard
1009 error around the model intercept (genotype A/A) or the effect size relative to the intercept
1010 (genotypes G/A and G/G). Numbers above the points indicate the number of observations
1011 and the number of FIDs (in parentheses) for each genotype.

1012

TABLE 1. DATASET INFORMATION AND ANIMAL MODEL RESULTS OF AUTOSOMAL CROSSOVER COUNT (ACC).

N_{OBS}, N_{FIDs} and N_{Xovers} indicate the number of ACC measures, the number of FIDs and the total number of crossovers observed, respectively. Mean is that of the raw data, and V_P and V_A are the phenotypic and additive genetic variances, respectively. The heritability h^2 and residual effect e^2 are the proportions of phenotype variance explained by the additive genetic and residual variances, respectively. *P* (h^2) is the significance of the additive genetic effect (h^2) as determined using a model comparison approach (see text). V_A and heritability were modelled using genomic relatedness. Figures in brackets are standard errors.

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)	

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

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)