Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (*Ovis aries*)

Susan E. Johnston,*,[1] Camillo Bérénos,* Jon Slate,[†] and Josephine M. Pemberton*
*Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, EH9 3FL, United Kingdom, and
[†]Department of Animal and Plant Sciences, University of Sheffield, S10 2TN, United Kingdom
ORCID IDs: 0000-0002-5623-8902 (S.E.J.); 0000-0003-3356-5123 (J.S.); 0000-0002-0075-1504 (J.M.P.)

ABSTRACT Meiotic recombination breaks down linkage disequilibrium (LD) 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 in 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 rates. Individual rates were high relative to other mammal systems and were higher in males than in females (autosomal map lengths of 3748 and 2860 cM, respectively). The heritability of autosomal recombination rate was low but significant in both sexes (h^2 = 0.16 and 0.12 in females and males, respectively). In females, 46.7% of the heritable variation was explained by a subtelomeric region on chromosome *6*; a genome-wide association study showed the strongest associations at locus *RNF212*, with further associations observed at a nearby ~374-kb region of complete LD 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.

KEYWORDS meiotic recombination; genome-wide association study; genomic relatedness; heritability; natural population

MEIOTIC recombination is a fundamental feature of sexual reproduction in nearly all multicellular 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 is also 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 breakup of favorable combinations of alleles previously built up by selection, reducing the mean fitness of subsequent generations (Charlesworth and Barton 1996). Therefore, the relative costs and benefits of recombination are

doi: 10.1534/genetics.115.185553
Manuscript received December 1, 2015; accepted for publication March 11, 2016; published Early Online March 29, 2016.
Available freely online through the author-supported open access option.

Supplemental material is available online at www.genetics.org/cgi/data/genetics.115.185553/DC1.
[1]Corresponding author: Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Charlotte Auerbach Road, Edinburgh EH9 3FL, United Kingdom. E-mail: Susan.Johnston@ed.ac.uk

likely to vary within different contexts, leading to an expectation of variation in recombination rates within and between populations (Barton 1998; Burt 2000; Otto and Lenormand 2002).

Recent studies of model mammal systems have shown that recombination rates vary at an individual level and that a significant proportion of variance is driven by heritable genetic effects (Kong *et al.* 2004; Dumont *et al.* 2009; Sandor *et al.* 2012). In cattle, humans, and mice, the heritability of recombination rate is 0.22, 0.08–0.30, and 0.46, respectively, and genome-wide association studies (GWASs) have repeatedly attributed some heritable variation to specific genetic variants, including *ring finger protein 212* (*RNF212*), *complexin 1* (*CPLX1*), *meiotic recombination protein REC8* (*REC8*), and *PR domain zinc finger protein 9* (*PRDM9*), among others (Kong *et al.* 2008, 2014; Baudat *et al.* 2010; Sandor *et al.* 2012; Ma *et al.* 2015). Most of these loci appear to influence crossover frequency, may have sex-specific or sexually antagonistic effects on recombination rate (*e.g.*, *RNF212* and *CPLX1* in humans and cattle) (Kong *et al.* 2014; Ma *et al.* 2015), and may be dosage dependent (*e.g.*, *RNF212* in mice) (Reynolds *et al.* 2013). The locus *PRDM9* is associated with the positioning and proportion of crossovers that occur in mammalian recombination hotspots (*i.e.*, regions of the genome with particularly high recombination rates) (Baudat *et al.* 2010; Ma *et al.* 2015), although this locus is not functional in some mammal species, such as canids (Auton *et al.* 2013). These studies suggest that recombination rate has a relatively oligogenic architecture and therefore has 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 systems that are likely to have been subject to strong artificial selection in their recent history, a process that will favor 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 than in 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 rates in wild populations will allow dissection of genetic and environmental drivers of recombination rate to determine whether 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 of 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 in Scotland since the Bronze Age (Clutton-Brock *et al.* 2004). In this study, we integrate genomic and pedigree information to characterize autosomal crossover 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 to recombination rates (*e.g.*, age, sex, and inbreeding coefficients), (2) to determine whether individual recombination rates were heritable, (3) to identify specific genetic variants associated with recombination rate variation, and (4) to determine whether 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 (57°49'N, 8°34'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 of 1986 (license no. PPL60/4211). A Soay sheep pedigree has been constructed using 315 SNPs in low linkage disequilibrium (LD) and includes 5516 individuals with 4531 maternal and 4158 paternal links (Bérénos *et al.* 2014).

SNP data set

A total of 5805 Soay sheep were genotyped at 51,135 SNPs on the Ovine SNP50 BeadChip using an Illumina Bead Array genotyping platform (Illumina, San Diego) (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 $\widehat{F}_{III}$ in Yang *et al.* (2011), hereafter $\widehat{F}$] 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.

Estimation of autosomal meiotic crossover count

Subpedigree construction: To allow unbiased phasing of the SNP data, a standardized pedigree approach was used to identify crossovers that had occurred within the gametes transferred from a focal individual to its offspring; hereafter, focal individual (FID) refers to the sheep in which meiosis took place. For each FID-offspring combination in the Soay sheep pedigree, a subpedigree was constructed to include both parents of the FID ("Father" and "Mother") and the other parent of the "Offspring" ("Mate"), where all five individuals had been genotyped (Figure 1). This subpedigree structure

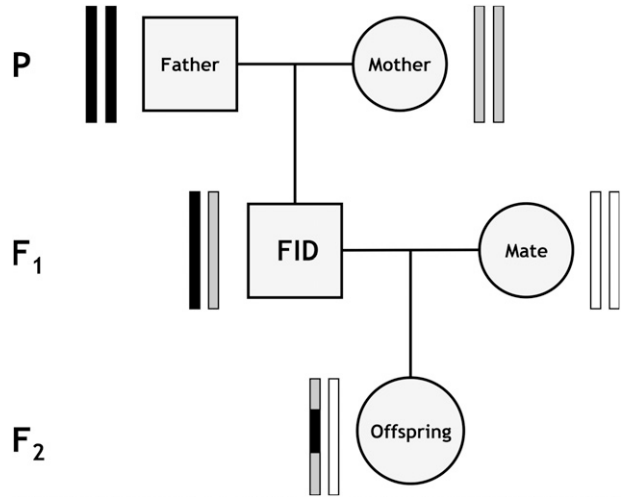

Figure 1 Diagram of the subpedigree structure used to infer crossover events. Rectangle pairs next to each individual represent chromatids, with black and gray shading indicating chromosome or chromosome sections of FID paternal and FID maternal origin, respectively. White shading indicates chromatids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

allowed phasing of SNPs within the FID and thus identification of autosomal crossover events in the gamete transferred from the FID to the offspring (Figure 1). Subpedigrees were discarded from the analysis if they included the same individual twice (*e.g.*, father-daughter matings; $N = 13$).

Linkage map construction and chromosome phasing: All analyses in this section were conducted using the software CRI-MAP v2.504a (Green *et al.* 1990). First, Mendelian incompatibilities in each subpedigree were identified using the *prepare* function; incompatible genotypes were removed from all affected individuals, and subpedigrees containing parent-offspring relationships with more than 0.1% mismatching loci were discarded. Second, sex-specific and sex-averaged linkage map positions (in Kosambi centimorgans) were obtained using the *map* function, where SNPs were ordered relative to their estimated positions on the sheep genome assembly Oar_v3.1 (GenBank assembly ID GCA_000298735.1) (Jiang *et al.* 2014). SNP loci with a map distance of >3 cM to each adjacent marker (10 cM for the *X* chromosome, including the pseudoautosomal regions) were assumed to be 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 the section *Data availability*). Third, the *chrompic* function was used to identify informative SNPs (*i.e.*, those for which the grandparent 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 grandparental 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 overestimation 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 ($N = 973$ of 38,592 double crossovers) (Supplemental Material, Figure S1). In the 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 <2.5 SD from the mean $\log_{10}$ span distance (equivalent to 9.7 Mb) also were recoded as missing ($N = 503$ of 37,619 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) then were 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 ($N_{\text{simulations}} = 100$) (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 $\widehat{F}$. This is likely to be an artifact of long runs of homozygosity as a result of inbreeding, which may prevent detection of double crossovers or crossovers in subtelomeric 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 lengths 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 modeled using a multiplicative inverse ($1/x$) regression in R v3.1.1.

Fine-scale recombination rate: The probability of crossing over was calculated in 1-Mb 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} is the number of bases to the bin boundary, and N_{adjSNP} is 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 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 (percent, obtained using a 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 the 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 60 Mb, equivalent to half the length of the largest acrocentric chromosome (chromosome *4*). Initial models also included a term indicating whether a centromere was present or absent on the 60-Mb region, but this term was not significant in either model.

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

ACC was modeled 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 (GRM) at all autosomal markers was constructed for all genotyped individuals using GCTA v1.24.3 (Yang *et al.* 2011). The GRM was adjusted using the argument *−grm-adj 0*, which assumes that frequency spectra of genotyped and causal loci are similar; matrices with and without adjustment were highly correlated ($R^2 > 0.997$), and variance components estimated from models with and without adjustment were highly similar, suggesting that adjusting for sampling error in this way did not introduce bias. Matrices were not pruned to remove related individuals (*i.e.*, the *−grm-cutoff* option was not used) because there is substantial relatedness within this population, and models included common environment and parental effects, controlling for some consequences of shared environment among relatives (see below). Trait variance was analyzed first with the following univariate model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_\mathrm{r}\mathbf{u}_\mathrm{r} + \mathbf{e}$$

where $\mathbf{y}$ is a vector of the ACC per transferred gamete; $\mathbf{X}$ is an incidence matrix relating individual measures to a vector of fixed effects; $\boldsymbol{\beta}$, $\mathbf{Z}_1$, and $\mathbf{Z}_\mathrm{r}$ are incidence matrices relating individual measures with additive genetic effects and random effects, respectively; $\mathbf{a}$ and $\mathbf{u}_\mathrm{r}$ are vectors of additive genetic effects from the GRM and additional random effects, respectively; and $\mathbf{e}$ 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, $\widehat{F}$, and FID age at the time of meiosis; random effects tested included individual identity to account for repeated measures within the same FID (sometimes referred to as the *permanent environment effect*), maternal and paternal identity, and common environment effects of FID birth year and offspring birth year. The significance of fixed effects was determined using a Wald test, whereas the significance of random effects was calculated using likelihood-ratio tests (LRTs) between models with and without the focal random effect. Only sex and additive genetic effects were significant in any model, but $\widehat{F}$ and individual identity were retained in all models to account for potential underestimation of ACC and the effects of pseudoreplication, respectively.

To investigate whether the additive genetic variation underlying male and female ACC was associated with sex-specific variation in ACC, bivariate models were run. The additive genetic correlation r_A was determined using the CORGH error-structure function in ASReml-R (correlation with heterogeneous variances) with r_A set to be unconstrained; models fitted sex-specific inbreeding coefficients and individual identity effects. To test whether the genetic correlation was significantly different from 0 and 1, the unconstrained model was compared to models with r_A fixed at a value of 0 or 0.999. Differences in additive genetic variance in males and females were tested by constraining both to be equal values using the CORGV error-structure function in ASReml-R. Models then were compared using LRTs with 1 degree of freedom.

Genetic architecture of autosomal crossover count

Genome-wide association study of variants controlling ACC: Genome-wide association studies (GWASs) 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 earlier. 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 genomic relatedness matrices have been shown to be highly correlated (Bérénos *et al.* 2014). Sex-specific models also were 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 LD-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 22,273.61, meaning that 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

GWASs 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 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 subtelomere of chromosome *6* (see File S2).

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):

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{v}_i + \mathbf{Z}_2\mathbf{nv}_i + \mathbf{Z}_r\mathbf{u}_r + \mathbf{e}$$

where $\mathbf{v}$ is the vector of additive genetic effects explained by autosomal genomic region i, and $\mathbf{nv}$ is the vector of the additive genetic effects explained by all remaining autosomal markers outside region i. Regional heritabilities were determined by constructing 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 1.21 Mb 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 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 earlier) (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 variation in estimated variance components, again suggesting that estimates were unbiased. The significance of additive genetic variance attributed to a genomic region i was tested by comparing models with and without the $\mathbf{Z}_1\mathbf{v}_i$ term using a LRT; 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 2 to account for overlap of the sliding windows (because each genomic region was modeled twice).

Accounting for cis and trans genetic variants associated with recombination rate: In the preceding analyses, we wished to separate potential associations with ACC due to *cis* effects (*i.e.*, genetic variants that are in LD 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*, association was examined with ACC summed across chromosomes *2–26*. We found that in this case, examining *trans* variation (ACC minus focal chromosome) obtained similar results to *cis* and *trans* variation (ACC) for both regional heritability and genome-wide association analyses, leading to the same biological conclusions.

LD and imputation of genotypes in significant regions: A reference population of 189 sheep was selected and genotyped at 606,066 SNP loci on the Ovine 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 $\mathbf{p}_m = \mathbf{A}_m^{-1}\mathbf{c}_m$, where $\mathbf{p}$ is a vector of the proportion of genetic variation in the population captured by m selected animals, $\mathbf{A}_m$ is the corresponding subset of a pedigree relationship matrix, and $\mathbf{c}$ is a vector of the mean relationship of the m selected animals [as outlined in Pausch *et al.* (2013) and 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 > 0.95 and MAF > 0.01, and individuals were retained if more than 95% of loci were genotyped. LD between loci was calculated using Spearman's rank correlations (r^2) in the 188 individuals passing quality control.

Genotypes from the high-density 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 high-density SNP chip. Because 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, individuals with <0.99 imputed genotypes scored were removed from the analysis. Loci with <0.95 of individuals typed also were 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 whether 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 for Soays and a further 2709 individuals from 73 different sheep breeds [provided by the International Sheep Genomics Consortium (ISGC)] (Kijas *et al.* 2012; Feulner *et al.* 2013) (Table S1). In both the Soay and non-Soay data sets of the Ovine SNP50 BeadChip, we extracted 58 SNPs corresponding to ~4 Mb of the subtelomeric region on chromosome *6* and phased them using Beagle v4.0 (Browning and Browning 2007). We identified core haplotypes of six SNP loci that tagged different alleles at oar3_OAR6_116402578. The length of HS between the core Soay haplotypes and the non-Soay breeds then was 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 of i was recorded. This was repeated for all pairwise comparisons of Soay and non-Soay haplotypes to determine a mean and SD of HS between i and breed j.

Data availability

The Supplemental Material 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. Table S3, Table S4, and Table S5 contain full detailed results and effect sizes of the regional heritability, genome-wide association, and imputed association studies, respectively. A Dryad repository (doi: 10.5061/dryad.pf4b7) contains genomic data after quality control measures, pedigree information and subpedigree structures, autosomal GRMs, population-wide crossover probabilities, and individual recombination rate results. All scripts used for the analysis are provided at https://github.com/susjoh/GENETICS_2015_185553.

Results

Broad-scale variation in recombination landscape

We used information from 3330 subpedigrees and data from 39,104 genome-wide SNPs typed on the Ovine SNP50 BeadChip (Kijas *et al.* 2009) to identify 98,420 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 cM and sex-specific lengths of 3748 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 and Table S2). There was a linear relationship between the length of autosomes in megabases and linkage map lengths (adjusted $R^2 = 0.991$, $P < 0.001$; Figure 2A). Chromosome-wide recombination rates (cM/Mb) were higher in smaller autosomes (fitted as a 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 centimorgans (*i.e.*, differences in male and female recombination rates) 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 1-Mb 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 0–18.11 Mb from the nearest telomere (Figure 3B and 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 48 Mb (Figure S1).

Variation in individual recombination rate

Individual 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

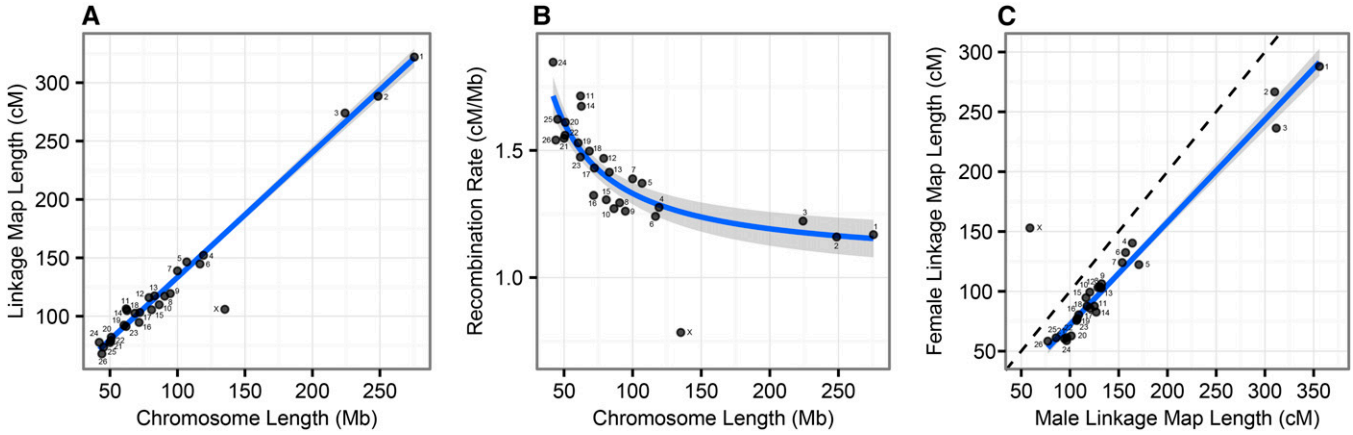

Figure 2 Broad-scale variation in recombination rate. Relationships between (A) sex-averaged linkage map length (cM) and physical chromosome length (Mb), (B) physical chromosome length (Mb) and recombination rate (cM/Mb), and (C) male and female linkage map lengths (cM). Points are chromosome numbers. Lines and the gray-shaded areas indicate the regression slopes and standard errors, respectively, excluding the *X* chromosome. The dashed line in C indicates where male and female linkage maps are of equal length. Note that the male linkage map length for the *X* chromosome is equivalent to the length of the pseudoautosomal region.

higher in males than in females, with 7.376 (SE = 0.263) more crossovers observed per gamete (animal model, $Z = 28.02$, $P_{\text{Wald}} < 0.001$). However, females had marginally higher additive genetic variance ($P_{\text{LRT}} = 0.040$) and higher residual variance ($P_{\text{LRT}} = 1.11 \times 10^{-3}$) in ACC than males (Table 1). There was no relationship between ACC and FID age, offspring sex, and the genomic inbreeding coefficient 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 0.826 (SE = 0.260); this correlation was significantly different from 0 ($P_{\text{LRT}} = 1.14 \times 10^{-3}$) but not different from 1 ($P_{\text{LRT}} = 0.551$).

Genetic architecture of recombination rate

Genome-wide association study: The most significant association between SNP genotype and ACC in both sexes was at s74824.1 in the subtelomeric region of chromosome *6* ($P = 2.92 \times 10^{-10}$; Table 2). Sex-specific GWASs 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 and Figure 4); the SNP had an additive effect on female ACC, with a difference of 3.37 (SE = 0.49) autosomal crossovers per gamete between homozygotes (Table 2). This SNP was the most distal typed on the chromosome from the Ovine SNP50 BeadChip at ~116.7 Mb (Figure 4 and Table 2) and corresponded to a genomic region containing *RNF212* and *CPLX1*, two loci that have previously been 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). A further SNP on an unmapped genomic scaffold (1.8 kb; NCBI accession no. AMGL01122442.1) also was highly associated with female ACC (Figure 4). BLAST analysis indicated that the most likely genomic position of this SNP was at ~113.8 Mb on chromosome *6*, corresponding to the same subtelomeric region.

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 and Table 2). This SNP was 17.8 kb 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 >1 Mb 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 and Table 2). This SNP was 137 kb from the 5′ end of an ortholog of *WD repeat domain 61* (*WDR61*) and 371 kb from the 5′ end of an ortholog of *ribosomal protein L10* (*RPL10*). Full results of GWASs 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 subtelomeric 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 ($P_{\text{LRT}} = 9.78 \times 10^{-14}$) but did not contribute to phenotypic variation in males (0.312% of phenotypic variance; SE = 1.2%, $P_{\text{LRT}} = 0.82$; Figure 5C and Table S5). There was an additional significant association between ACC in both sexes and a region on chromosome *7* corresponding to

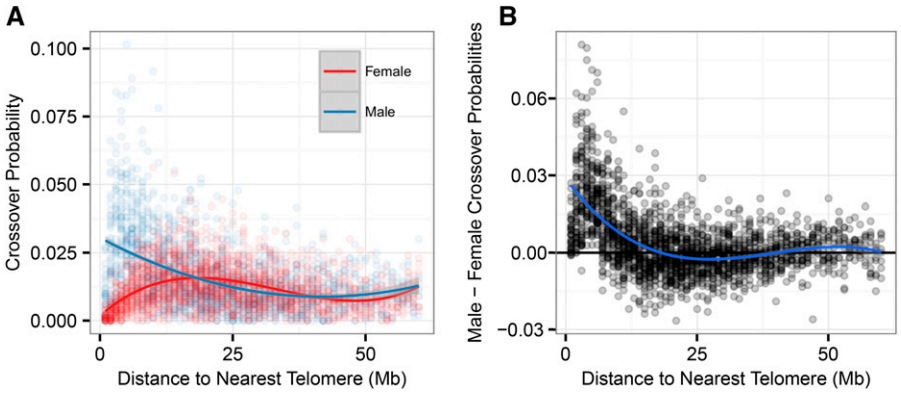

Figure 3 Variation in recombination rate relative to telomeric regions. Probability of crossing over relative to the nearest telomere (Mb) for (A) female and male linkage maps individually and (B) the difference between male and female crossover probabilities (male minus female). Data points are given for 1-Mb windows. Lines indicate the function of best fit.

a 1.09-Mb segment containing ~50 protein-coding regions, including *RNF212B* (a paralogue of *RNF212*) and *REC8* ($P_{LRT} = 3.31 \times 10^{-6}$; Figure 5A and Table S5); this region had not shown any significant associations using the GWAS approach alone. The region explained 4.12% of phenotypic variance (SE = 2.3%) and 26.2% of the additive genetic variance in both sexes combined; however, in sex-specific models, the significant association with ACC did not remain after correction for multiple testing (Table S5). No association was observed in the regional heritability analysis with the two regions on chromosome *3* identified using the GWAS approach. Full results for the regional heritability analysis are provided in Table S5.

Accounting for cis and trans genetic variants associated with recombination rate: The results presented earlier were ACC incorporating both *cis* and *trans* effects on recombination rate. When repeated with *trans* effects only, all variants associated with ACC in both the GWAS and regional heritability analyses remained significant (see *Materials and Methods* and Table S4 and Table S5), meaning that they are likely to affect recombination rate globally (*i.e.*, *trans*-acting effects) rather than being in LD with polymorphic recombination hotspots.

Genotype imputation and association analysis at the subtelomeric region of chromosome 6: Genotyping of 187 sheep at a further 122 loci in the subtelomeric region of chromosome *6* showed that this region has elevated levels of LD, with the two most significant SNPs from the 50K chip tagging a haplotype block of ~374 kB ($r^2 > 0.8$; File S3, Figure 6, and Table S6). This block contained three candidate genes, *CPLX1*, *cyclin-G-associated kinase* (*GAK*), and *polycomb group ring finger 3* (*PCGF3*), and was 177 kb away from the candidate locus *RNF212* (Kong *et al.* 2014). SNP genotypes were imputed for all individuals typed on the 50K chip at these 122 loci, and the association analysis was repeated. The most highly associated SNP (oar3_OAR6_116402578, $P = 1.83 \times 10^{-19}$; Table 2 and Figure 6) occurred within an intronic region of an uncharacterized protein orthologous to *transmembrane emp24 protein transport domain containing TMED11*, 25.2 kb from the putative location of *RNF212* and 13 kb 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 (SE = 0.203) autosomal crossovers per gamete between homozygotes (Figure 7, Table 2, and Table 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.

HS of associated regions with domesticated breeds

Seven core haplotypes of six SNPs in length tagged different alleles at oar3_OAR6_116402578 at the subtelomeric 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 sheep was low, and there was no evidence of long-range HS between Soays and Boreray sheep compared with other domesticated breeds (Figure S3). This test is not definitive owing to the relatively small 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 HS with other breeds throughout the sample suggest that the alleles at oar3_OAR6_116402578 have not been introduced recently to the Soay sheep population. For example, HS of core haplotypes with Boreray sheep for coat color, coat pattern (Feulner *et al.* 2013), and normal horn development (Johnston *et al.* 2013) extended to longer distances, up to 5.7, 6.4, and 2.86 Mb, respectively. In contrast, the maximum HS observed here was 0.38 Mb. A shorter haplotype may be expected because the core haplotype occurs at the end of the chromosome, so HS is only calculated downstream of the core haplotype; however, this value is much lower than half that of previously identified introgressed haplotypes (Feulner *et al.* 2013). 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 across sheep breeds.

Table 1 Data set information and animal model results of ACC

Sex	N_{OBS}	N_{FIDs}	N_{Xovers}	Mean	V_P	V_A	h^2	e^2	P (h^2)
Both	3330	813	98,420	29.56 (0.11)	29.56 (0.83)	4.28 (0.85)	0.15 (0.03)	0.85 (0.02)	6.88×10^{-15}
Female	2134	586	57,613	27.00 (0.10)	31.71 (1.06)	5.04 (0.82)	0.16 (0.02)	0.84 (0.02)	4.76×10^{-12}
Male	1196	227	40,807	34.12 (0.09)	25.21 (1.16)	2.97 (0.84)	0.12 (0.03)	0.88 (0.03)	0.022

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 modeled using genomic relatedness. Figures in parentheses are standard errors.

Discussion

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

Genetic variants associated with individual recombination rate

Most of the variants associated with ACC in this study have been implicated previously in recombination rate variation in other mammal species, suggesting a shared genetic architecture across taxa. The strongest association was observed at locus *RNF212*, occurring 88.4 kb from an ~374-kb block of high LD ($r^2 > 0.8$) containing three further candidate loci, *CPLX1*, *GAK*, and *PCGF3* (Figure 6). Both *RNF212* and *CPLX1* have been associated with recombination rate variation in mammals (Kong *et al.* 2008; Sandor *et al.* 2012; Reynolds *et al.* 2013; Ma *et al.* 2015), and mouse studies have established that the protein RNF212 is essential for the formation of crossover-specific complexes during meiosis and that its effect is dosage sensitive (Reynolds *et al.* 2013). We observed an additive effect of the *RNF212* region on female recombination rate (Figure 7), suggesting that dosage dependence could be a plausible mechanism driving rate differences in Soay sheep. GAK forms part of a complex with cyclin G, a locus involved in meiotic recombination repair in *Drosophila* (Nagel *et al.* 2012), and PCGF3 forms part of a PRC1-like complex (polycomb repressive complex 1) that is involved in meiotic gene expression and the timing of meiotic prophase in female mice (Yokobayashi *et al.* 2013). High LD within this region meant that it was not possible to test the effects of these loci on recombination rate independently; however, the cosegregation of several loci affecting meiotic processes may merit further investigation to determine whether recombination is suppressed in this region and whether this cosegregation is of adaptive significance.

Additional genomic regions associated with recombination rate included two loci at 48.1 and 82.4 Mb on chromosome *3* (identified using GWAS) with effects on males only and both sexes, respectively, and a 1.09-Mb region of chromosome *7* affecting rates in both sexes (identified using regional heritability analysis). Although the chromosome *7* region was large and specific loci cannot be pinpointed, it contained *REC8*, which codes a protein required for the separation of sister chromatids and homologous chromosomes during meiosis (Parisi *et al.* 1999), and *RNF212B*, a paralogue of *RNF212*. The same region is also associated with recombination rate in cattle (Sandor *et al.* 2012). The chromosome *3* variants identified were novel to this study and occurred in relatively gene-poor regions of the genome (see earlier).

Although there are homologs of *PRDM9* on chromosomes *1*, *5*, *18*, and *X*, it is not currently known whether any of these copies are functional in sheep. Here we did not identify any association between recombination rate and any of these regions using either GWAS or regional heritability approaches. This may not be surprising because this locus is primarily associated with recombination hotspot usage. Nevertheless, *PRDM9* has been associated with recombination rate in cattle and male humans (Kong *et al.* 2014; Ma *et al.* 2015) and is likely to be a consequence of differences in the abundance of motifs recognized by the PRDM9 protein in hotspots rather than the locus itself affecting rate. In this study, it was not possible to examine hotspot usage because crossovers could only be resolved to a median interval of 800 kb. This is unlikely to be of a scale fine enough to characterize hotspot variation because hotspots typically occur within 1- to 2-kb intervals in mammals (Paigen and Petkov 2010). Further studies would require higher densities of markers to determine crossover positions at a greater resolution and to determine the functionality and/or relative importance of *PRDM9* within this system.

Sexual dimorphism in the genetic architecture of recombination rate

This study identified sexual dimorphism in the genetic architecture of recombination rate in Soay sheep. Using a classical quantitative genetics approach, the between-sex genetic correlation was not significantly different from 1,

Table 2 Top hits from genome-wide association studies of ACC in all sheep, females and males

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)
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)

Results provided are from the Ovine SNP50 BeadChip and (below the line) the most highly associated imputed SNP from chromosome *6* (oar3_OAR6_116402578). 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 Table 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 boldface 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 numbers of unique individuals for all sheep, females, and males are approximately N = 813, 586, and 227, respectively. Numbers in parentheses are standard errors

indicating that male and female recombination rate variation had a shared genetic basis—albeit with a relatively large error around this estimate. However, females had significantly higher additive genetic and residual variance in the trait in comparison with males, and GWAS and regional heritability analyses showed that the *RNF212/CPLX1* region was associated with female recombination rate only. This is consistent with previous studies, where this region was associated with sexually dimorphic and sexually antagonistic variation in recombination rates in cattle and humans, respectively (Kong *et al.* 2014; Ma *et al.* 2015). Therefore, our findings suggest that variation in recombination rate has *some* degree of a shared and distinct genetic architecture between the sexes, which may be expected as a result of various similarities of this process of meiosis, but differences in its implementation within each sex (discussed further later). There were some differences in sample sizes between the sexes, with twice as many meioses characterized in females than in males, so it could be argued that low sample sizes in males may have had less power to identify specific loci. While possible, it is unlikely that the absence of associations between the *RNF212/CPLX1* region and male recombination rate is due to low power to detect the effect because (1) models including both sexes showed reduced, rather than increased, significance in this region, (2) 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), and (3) repeating the association analysis at the most highly associated SNP using sampled data sets of identical size in males and females found consistently higher association at this locus in females (Figure S4).

How much phenotypic variation in recombination rate is explained?

The approaches used in this study were successful in characterizing several regions of the genome contributing to the additive genetic variance in recombination rate. The regional heritability approach demonstrated some potential for characterizing variation from multiple alleles and/or haplotypes encompassing both common and rare variants that are in LD with causal loci that were not detectable by GWAS alone (Nagamine *et al.* 2012). However, while 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,

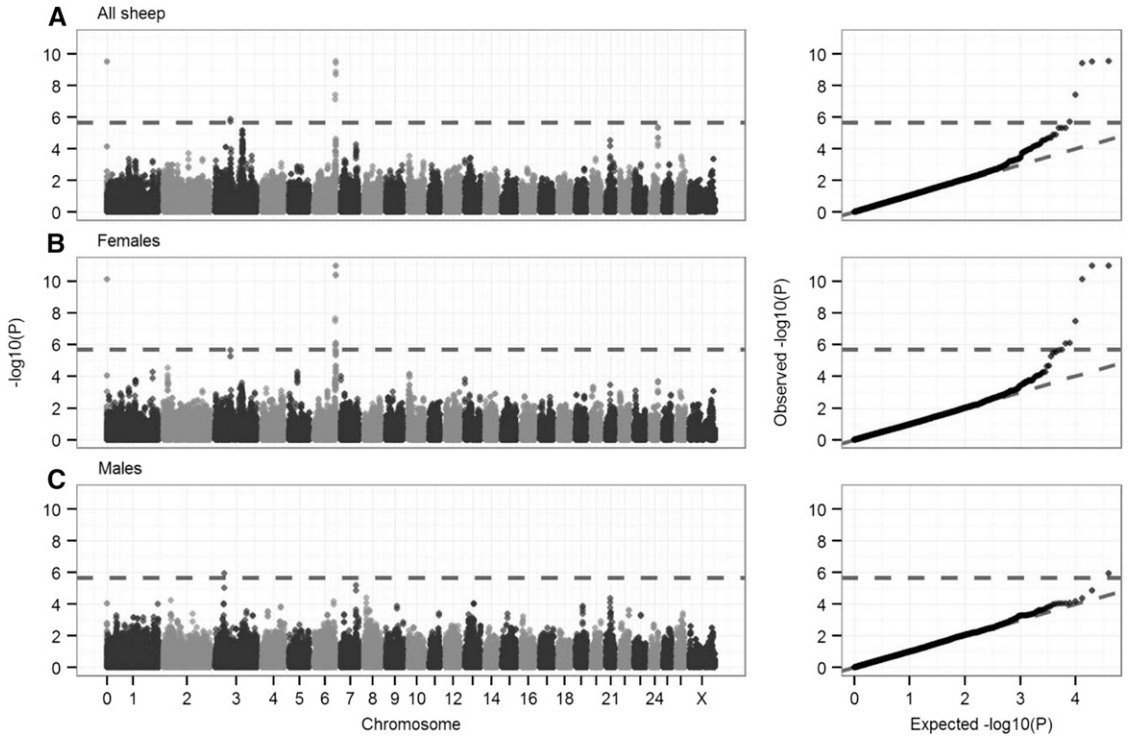

Figure 4 Genome-wide association of autosomal crossover count. Genome-wide association statistics in (A) all sheep, (B) females only, and (C) males only. The dotted line indicates the threshold for statistical significance after multiple testing (equivalent to an experiment-wide threshold of $P = 0.05$). The left column shows association statistics relative to genomic position; points are color-coded by chromosome. The right column shows the distribution of observed *P*-values against those expected from a null distribution. Association statistics were not corrected using genomic control because λ was less than 1 for all GWASs (λ = 0.996, 0.933, and 0.900 for plots A, B, and C, respectively). Underlying data on associations at the most highly associated SNPs, their genomic positions, and the sample sizes are given in Table S5. The significant SNP in gray at position zero in A and B occurs on an unmapped contig that is likely to correspond to the distal region of chromosome 6 (see text).

and there may have been reduced power to detect genetic variants, particularly in males, which were under-represented in the data set. However, our findings of low heritability and unexplained additive genetic variance are consistent with recent results from Icelandic humans, where despite a larger sample size (N = 15,253 males and 20,674 females) and greater marker density ($N = 30.3 \times 10^6$), 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, a very large proportion of both the additive genetic and phenotypic variance remains unexplained.

Variation and sexual dimorphism in the recombination landscape

Males had considerably higher recombination rates than females, which was driven mainly by large differences in crossover frequencies in the subtelomeric regions between 0 and 18.11 Mb (Figure 3); recombination was reduced in males if the centromere was present in the subtelomeric region (*i.e.*, in autosomes *4–26*, which are acrocentric) but, unlike in cattle, was still significantly higher than that in females (Ma *et al.* 2015) (Figure S5). Outside the subtelomeric region, recombination rates were more similar between the sexes, with females showing slightly higher recombination rates between 18.1 and 40 Mb from the telomere (Figure 3B). This observation of increased subtelomeric recombination in males is consistent with studies in humans, cattle, and mice (Kong *et al.* 2002; Shifman *et al.* 2006; Ma *et al.* 2015), although the magnitude of the difference is much greater in the Soay sheep population. Within females, the rate differences associated with different genotypes at *RNF212* were most clear in regions likely to be euchromatic, whereas there was no difference in rate in regions likely to be heterochromatic, such as the subtelomeric and centromeric regions (Figure S6 and Table S10).

Why is the recombination rate higher in males?

In placental mammals, females usually exhibit higher recombination rates than males (Lenormand and Dutheil 2005), and it has been postulated that this is a mechanism to avoid aneuploidy after long periods of meiotic arrest (Koehler *et al.* 1996; Morelli and Cohen 2005; Nagaoka *et al.* 2012). However, Soay sheep exhibited male-biased recombination rates to a greater degree than observed in any placental mammal to date

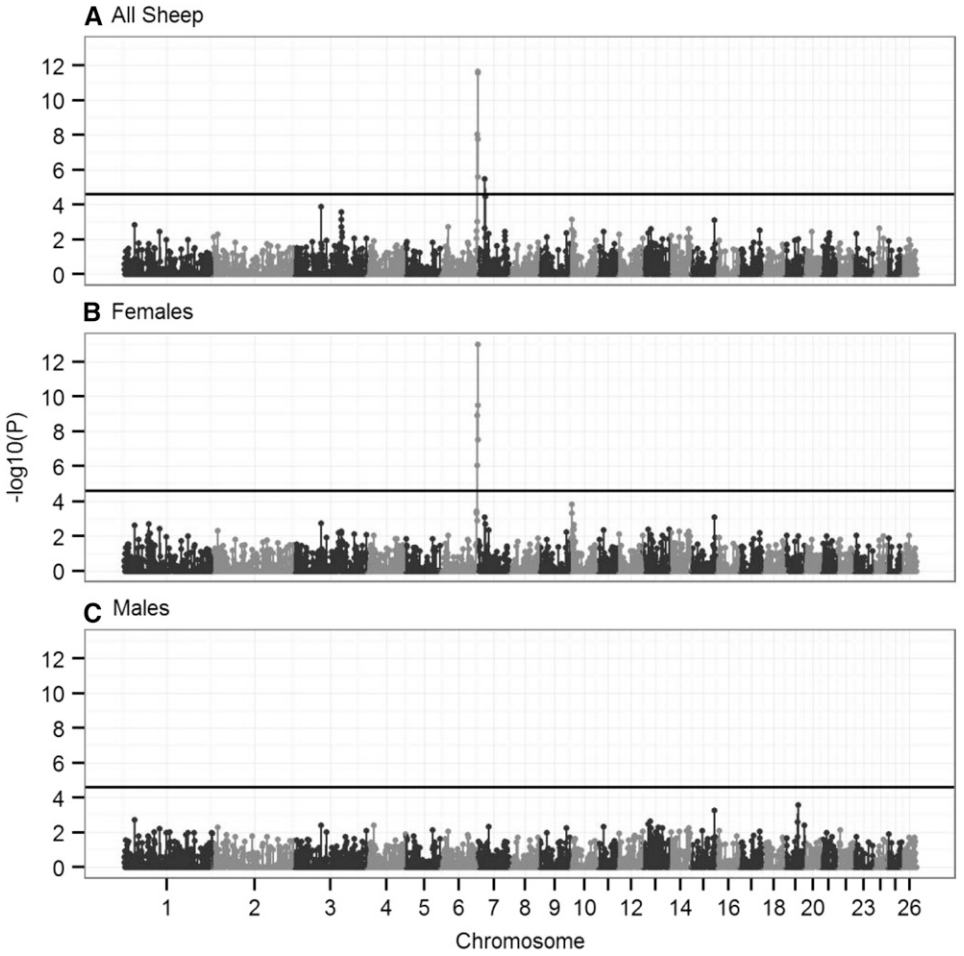

Figure 5 Regional heritability analysis of autosomal crossover count. Significance of association analysis in (A) all sheep, (B) females only, and (C) males only. The results presented are from a sliding window of 20 SNPs across 26 autosomes, with an overlap of 10 SNPs (see text). Points represent the 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 are provided in Table S4.

(male-female linkage map ratio = 1.31). The biological significance of this remains unclear, although a number of mechanisms have been proposed to explain variation in sex differences more generally, including haploid selection (Lenormand and Dutheil 2005), meiotic drive (Brandvain and Coop 2012), sperm competition, and sexual dimorphism and dispersal (Trivers 1988; Burt *et al.* 1991; Mank 2009). Nevertheless, testing these ideas has been limited by a paucity of empirical data.

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

Examining recombination rates in the wild

A principal motivation for this study was to determine how recombination rate and its genetic architecture may vary

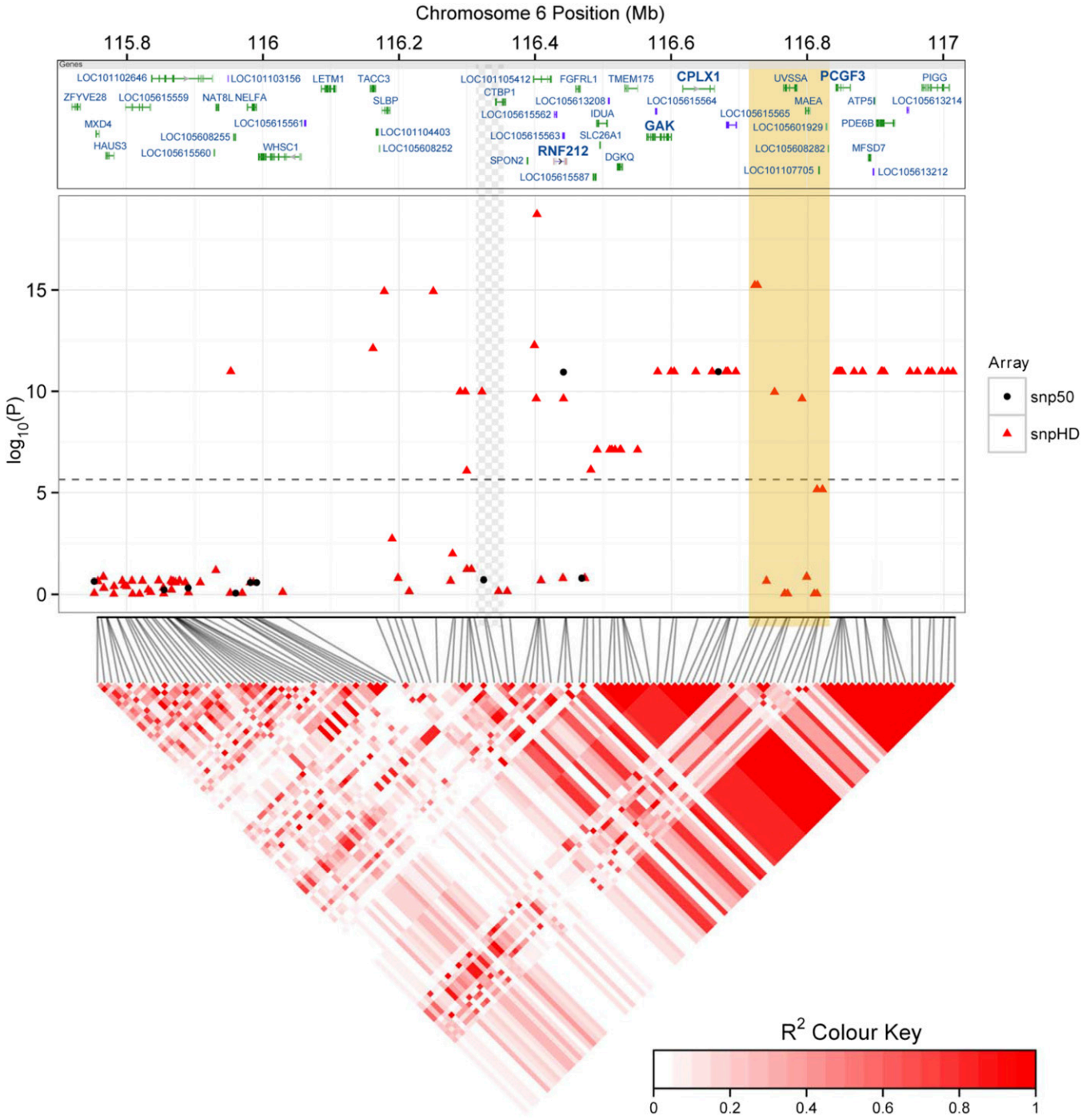

Figure 6 Associations at the subtelomeric region of chromosome *6*. Local associations of female ACC with Ovine SNP50 BeadChip SNPs (black circles, middle panel) and imputed genotypes from the Ovine HD SNP BeadChip (red triangles). The top panel indicates protein-coding regions within this region, as provided by the NCBI Graphical Sequence Viewer v3.8, with genes previously implicated in recombination or meiosis given in boldface type [see introduction and Yokobayashi *et al.* (2013) and Kong *et al.* (2014)]. The dashed line in the middle panel indicates the significance threshold after multiple testing. The lower panel is a heat map of LD in this region calculated for the 188 individuals typed on the high-density SNP chip using Spearman's rank correlation r^2 created using the R library *LDheatmap* (Shin *et al.* 2006). The superimposed beige block indicates a region that is likely to be incorrectly assembled on the sheep genome assembly (Oar v3.1) based on sequence comparison with the homologous region of the cattle genome assembly (vUMD3.1) (see File S3); its position is likely to fall within the region indicated by the gray-checkered pattern to the left, leaving a large region of very high LD at the distal end of chromosome *6*.

relative to model species that have undergone strong selection in their recent history. We found that the heritability of recombination rate in Soay sheep was much lower than in cattle and mice and was comparable with recent estimates in humans, which also can be considered a wild population (Kong *et al.* 2014). Despite these differences, variants affecting

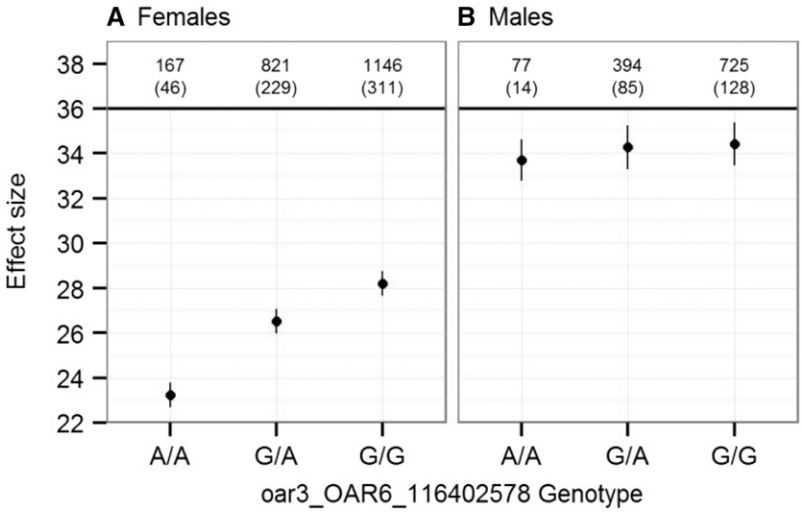

Figure 7 Effect sizes from a bivariate animal model of autosomal crossover count. Effect sizes are shown for (A) females and (B) males from a single bivariate model including oar3_OAR6_116402578 genotype as a fixed interaction term. Error bars are the standard error around the model intercept (genotype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

recombination rate on sheep chromosomes *6* and *7* have been associated previously with recombination rates in other mammalian populations (see earlier), and only the two relatively gene-poor regions identified on chromosome *3* are novel. Furthermore, examination of haplotypes around *RNF212* suggests that variation in the *RNF212/CPLX1* region affecting recombination rate has been segregating for a long time in sheep. Examining variation in the wild also allowed us to quantify the effects of the individual and the environment on recombination rate; however, we found no effect of common environment (*i.e.*, birth year and year of gamete transmission), individual age, or inbreeding on recombination rate; rather, most of the variation was attributed to residual effects. This contrasts with the observation in humans that recombination rates increase with age in females (Kong *et al.* 2004). Overall, our findings suggest a strong stochastic element driving recombination rates, with a small but significant heritable component that has a similar architecture to other mammal systems regardless of their selective background. The Soay sheep system is one of the most comprehensive wild data sets in terms of genomic resources and sampling density, making it one of the most suitable in terms of quantifying and analyzing recombination rate variation in the wild. A future ambition is to investigate the fitness consequences of phenotypic variation in ACC and more specifically the relationship between variants identified in this study and individual life-history variation to determine whether the maintenance of genetic variation for recombination rates is due to selection, sexually antagonistic effects, or stochastic processes.

Conclusions

In this study we have shown that recombination rates in Soay sheep are heritable and have a sexually dimorphic genetic architecture. The variants identified have been implicated in recombination rates in other mammal species, indicating a conserved genetic basis across distantly related taxa. However, the proportion of phenotypic variation explained by identified variants was low; this was consistent with studies in humans and cattle, in which although genetic variants were identified, and most of both additive genetic and phenotypic variance has remained unexplained. Similar studies in both mammalian and nonmammalian wild systems may provide a broader insight into the genetic and nongenetic drivers of recombination rate variation, as well its evolution in contemporary populations. However, the question remains as to whether variation in recombination rate is adaptive or merely a by-product of other biological processes. Overall, the approaches and findings presented here provide an important foundation for studies examining the evolution of recombination rates in contemporary natural populations.

Acknowledgments

We thank Jill Pilkington, Ian Stevenson, and all Soay Sheep Project members and volunteers for collection of data and samples. Discussions and comments from Jarrod Hadfield, Bill Hill, Craig Walling, Jisca Huisman, John Hickey, and two anonymous reviewers greatly improved the analysis. James Kijas, Yu Jiang, and Brian Dalrymple responded to numerous queries on the genome assembly and 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, with the SNP genotyping and research presented here supported by an ERC Advanced Grant to J.M.P.

Literature Cited

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

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

Auton, A., Y. Rui Li, J. Kidd, K. Oliveira, J. Nadel *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 281: 1986–1990.

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

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

Brandvain, Y., and G. Coop, 2012 Scrambling eggs: meiotic drive and the evolution of female recombination rates. Genetics 190: 709–723.

Browning, S. R., and B. L. Browning, 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., 2000 Sex, recombination, and the efficacy of selection: was Weismann right? Evolution 54: 337–351.

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

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

Butler, D. G., B. R. Cullis, A. R. Gilmour, and B. J. Gogel, 2009 *Mixed Models for S Language Environments: ASReml-R Reference Manual*. State of Queensland, Department of Primary Industries and Fisheries, Brisbane, Australia.

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

Clutton-Brock, T., J. Pemberton, T. N. Coulson, I. R. Stevenson, and A. D. C. MacColl, 2004 The sheep of St. Kilda, pp. 17–51 in Soay Sheep: Dynamics and Selection in an Island Population, Vol. 2, edited by T. H. Clutton-Brock, and J. M. Pemberton. Cambridge University Press, Cambridge, UK.

Crow, J. F., and M. Kimura, 1965 Evolution in sexual and asexual populations. Am. Nat. 99: 439–450.

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

Devlin, A. B., K. Roeder, and B. Devlin, 1999 Genomic control for association. Biometrics 55: 997–1004.

Dumont, B. L., K. W. Broman, and B. A. Payseur, 2009 Variation in genomic recombination rates among heterogeneous stock mice. Genetics 182: 1345–1349.

Falconer, D. S., and T. F. C. Mackay, 1996 Variance, pp. 122–144 in Introduction to Quantitative Genetics, Ed. 4. Pearson Prentice Hall, Essex, UK.

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

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

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

Goddard, M. E., and B. J. Hayes, 2009 Genomic selection based on dense genotypes inferred from sparse genotypes. Proc. Assoc. Advmt. Anim. Breed. Genet 18: 26–29.

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

Hassold, T., and P. Hunt, 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., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. 8: 269–294.

Inoue, K., and J. R. Lupski, 2002 Molecular mechanisms for genomic disorders. Ann. Rev. Genomics Hum. Genet. 3: 199–242.

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

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

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

Kijas, J. W., J. A. Lenstra, B. Hayes, S. Boitard, L. R. Porto Neto *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., R. S. Hawley, S. Sherman, and T. Hassold, 1996 Recombination and nondisjunction in humans and flies. Hum. Mol. Genet. 5: 1495–1504.

Kong, A., D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson *et al.*, 2002 A high-resolution recombination map of the human genome. Nat. Genet. 31: 241–247.

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

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

Kong, A., G. Thorleifsson, M. L. Frigge, G. Masson, D. F. Gudbjartsson *et al.*, 2014 Common and low-frequency variants associated with genome-wide recombination rate. Nat. Genet. 46: 11–16.

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

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

Ma, L., J. R. O'Connell, P. M. VanRaden, B. Shen, A. Padhi *et al.*, 2015 Cattle sex-specific recombination and genetic control from a large pedigree analysis. PLoS Genet. 11: e1005387.

Maddox, J. F., K. P. Davies, A. M. Crawford, D. J. Hulme, D. Vaiman *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. 91: 355–363.

Manolio, T. A., F. S. Colling, N. J. Cox, D. B. Goldstein, L. A. Hindorff *et al.*, 2009 Finding the missing heritability of complex diseases. Nature 461: 747–753.

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

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

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

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

Muñoz-Fuentes, V., M. Marcet-Ortega, G. Alkorta-Aranburu, C. Linde Forsberg, J. M. Morrell *et al.*, 2015 Strong artificial selection in domestic mammals did not result in an increased recombination rate. Mol. Biol. Evol. 32: 510–523.

Nagamine, Y., R. Pong-Wong, P. Navarro, V. Vitart, C. Hayward *et al.*, 2012 Localising loci underlying complex trait variation using regional genomic relationship mapping. PLoS One 7: e46501.

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

Nagel, A. C., P. Fischer, J. Szawinski, M. K. L. Rosa, and A. Preiss, 2012 Cyclin G is involved in meiotic recombination repair in Drosophila melanogaster. J. Cell Sci. 125: 5555–5563.

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

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

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

Parisi, S., M. J. Mckay, M. Molnar, M. A. Thompson, P. J. van der Spek *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., B. Aigner, R. Emmerling, C. Edel, K.-U. Götz *et al.*, 2013 Imputation of high-density genotypes in the Fleckvieh cattle population. Genet. Sel. Evol. 45: 3.

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

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

Reynolds, A., H. Qiao, Y. Yang, J. K. Chen, N. Jackson *et al.*, 2013 RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. Nat. Genet. 45: 269–278.

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

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

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

Shin, J., S. Blay, B. McNeney, and J. Graham, 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., B. Marrow, B. T. Preston, J. M. Pemberton, and K. Wilson, 2004 Adaptive reproductive strategies, pp. 243–275 in *Soay Sheep: Dynamics and Selection in an Island Population*, Vol. 2, edited by T. H. Clutton-Brock, and J. M. Pemberton. Cambridge University Press, Cambridge, UK.

Tankimanova, M., M. A. Hultén, and C. Tease, 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, pp. 270–286 in *The Evolution of Sex*, edited by R. Michod, and B. Levin. Sinauer Press, Sunderland, MA.

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

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

Communicating editor: S. F. Chenoweth

GENETICS

Supporting Information
www.genetics.org/cgi/data/genetics.115.185553/DC1

Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (*Ovis aries*)

Susan E. Johnston, Camillo Bérénos, Jon Slate, and Josephine M. Pemberton

DOI: 10.1534/genetics.115.185553

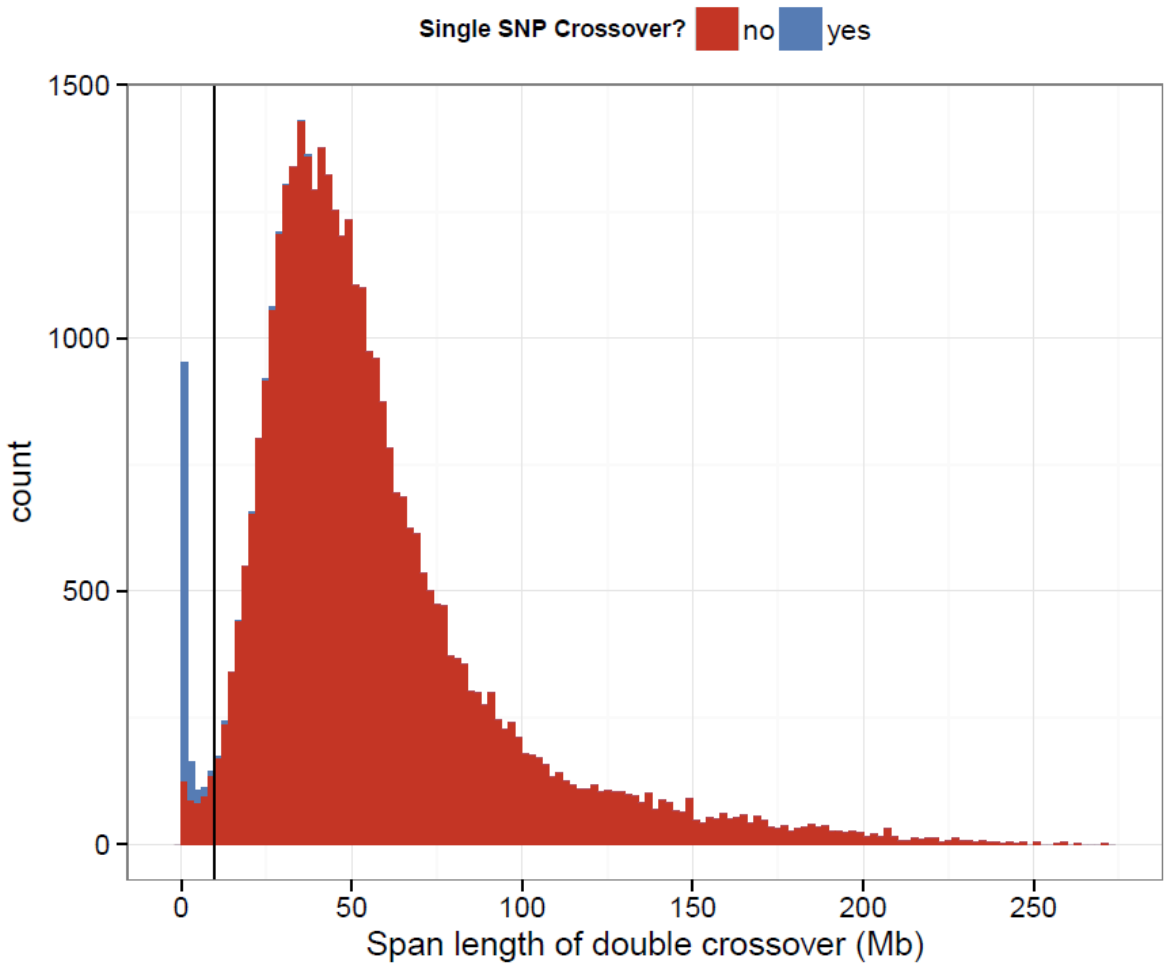

Figure S1. Histogram of the span distances (in Mb) between double crossovers on autosomal chromatids. Bar segments are color coded as double crossovers spanning a single SNP locus (blue) and those spanning more than one SNP (red). All double crossovers across a single SNP were discarded from the dataset, as they are likely to be the result of a genotyping error at that SNP. Short crossovers below a certain threshold (indicated by the solid vertical line) were also discarded (see text for rationale).

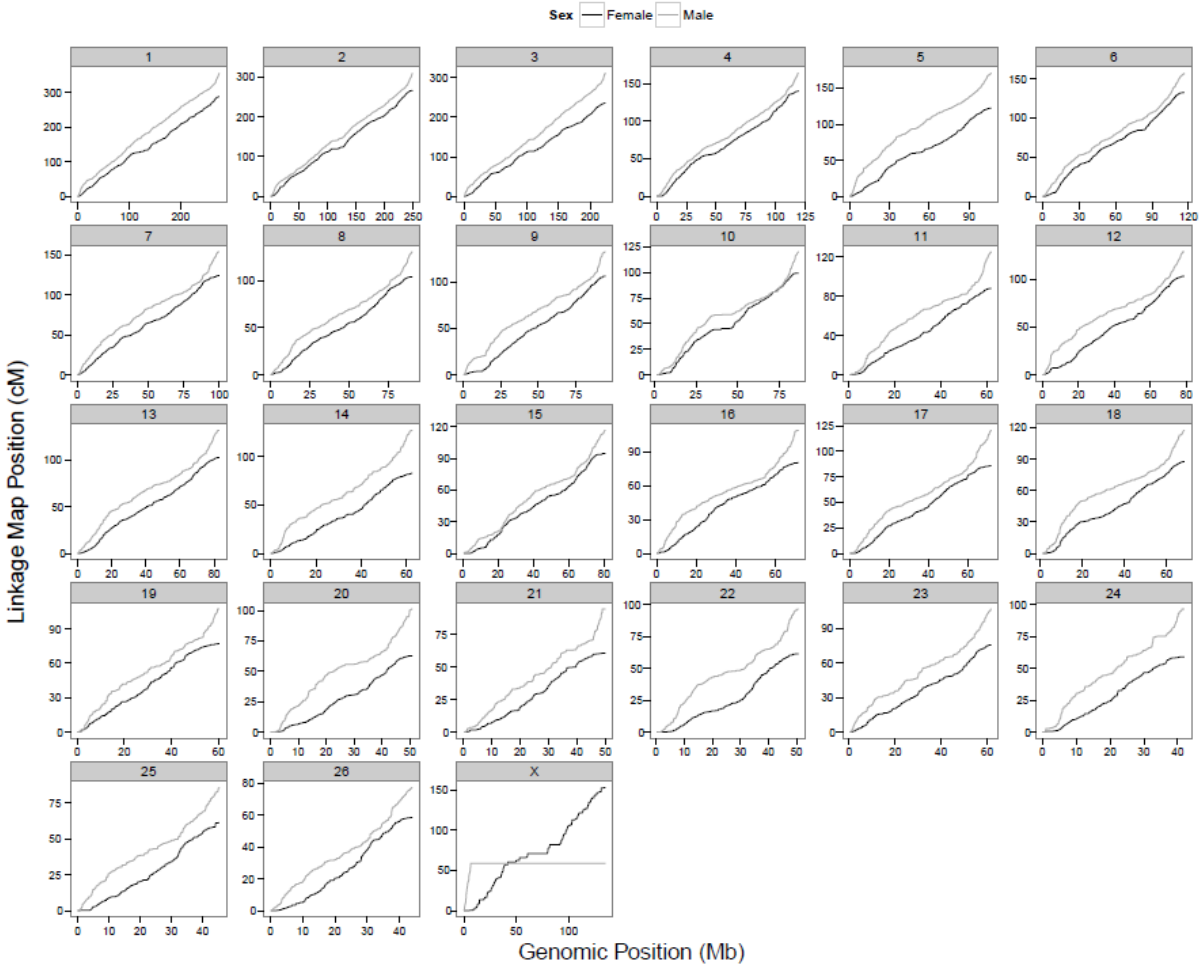

Figure S2. Comparison of sex-specific linkage map positions (cM) and genomic positions (Mb) relative to the Oar_v3.1 genome build. Female and male maps are plotted as black and grey lines, respectively. Numbers above each panel are the chromosome number. Chromosome 1, 2, 3 and X are metacentric chromosomes, whereas chromosomes 4 - 26 are acrocentric chromosomes, with the centromere located towards the left-hand end of the x-axis in each plot (Maddox *et al.* 2001).

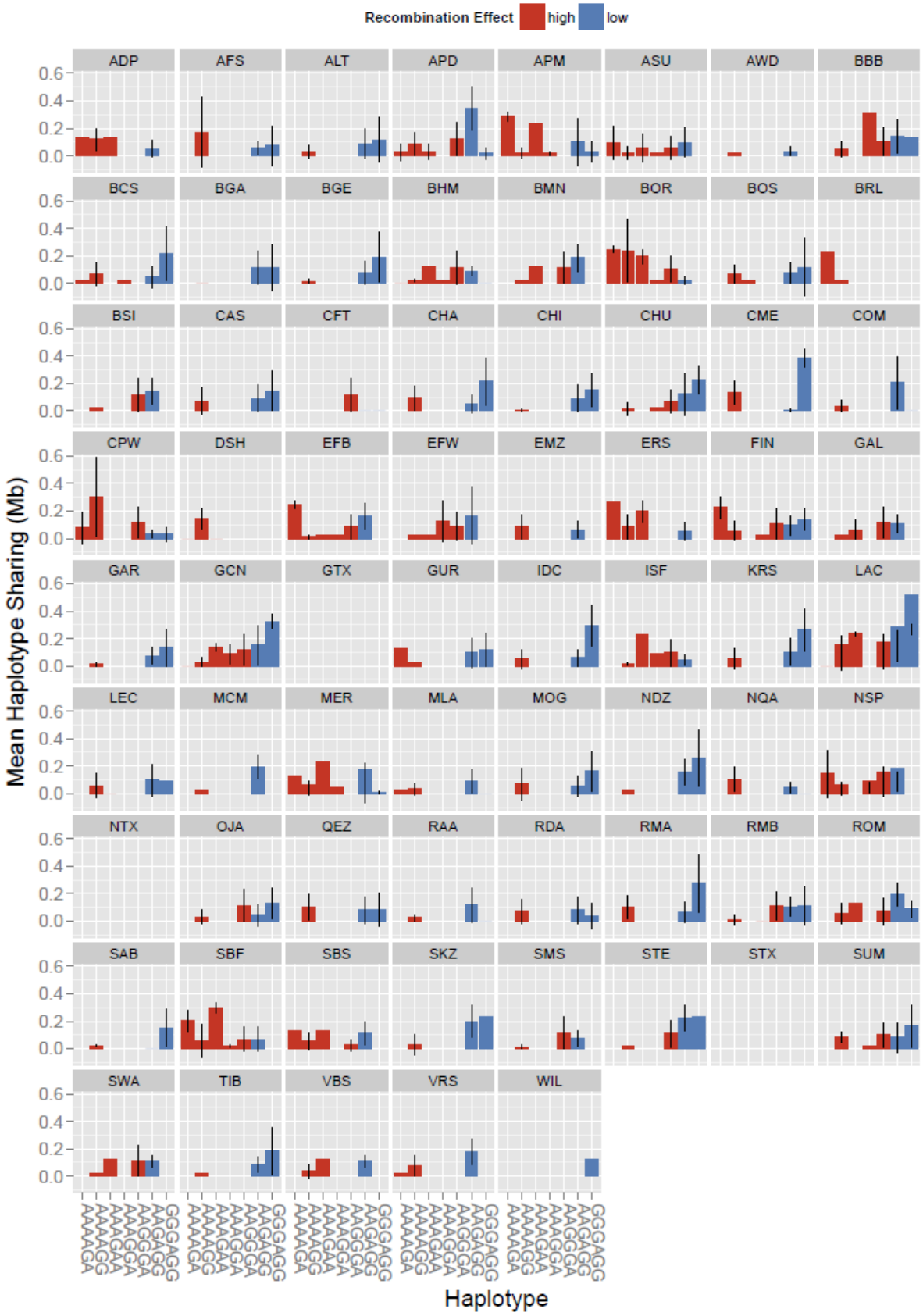
Recombination Effect
high
low
ADP
AFS
ALT
APD
APM
ASU
AWD
BBB
BCS
BGA
BGE
BHM
BMN
BOR
BOS
BRL
BSI
CAS
CFT
CHA
CHI
CHU
CME
COM
CPW
DSH
EFB
EFW
EMZ
ERS
FIN
GAL
GAR
GCN
GTX
GUR
IDC
ISF
KRS
LAC
LEC
MCM
MER
MLA
MOG
NDZ
NQA
NSP
NTX
OJA
QEZ
RAA
RDA
RMA
RMB
ROM
SAB
SBF
SBS
SKZ
SMS
STE
STX
SUM
SWA
TIB
VBS
VRS
WIL
0.0
0.2
0.4
0.6
Mean Haplotype Sharing (Mb)
AAAAGA
AAAAGG
AAAGAA
AAAGGA
AAGAGG
AAGGGA
GGGAGG
Haplotype

Figure S3. Mean and standard deviation of haplotype sharing (HS in Mb) between Soay sheep and 73 domestic sheep breeds for core haplotypes outlined in Table S8, for breeds listed in Table S1. Red and blue coloration indicates haplotypes tagging the high (G) and low (A) recombination alleles at oar3_OAR6_116402578, respectively.

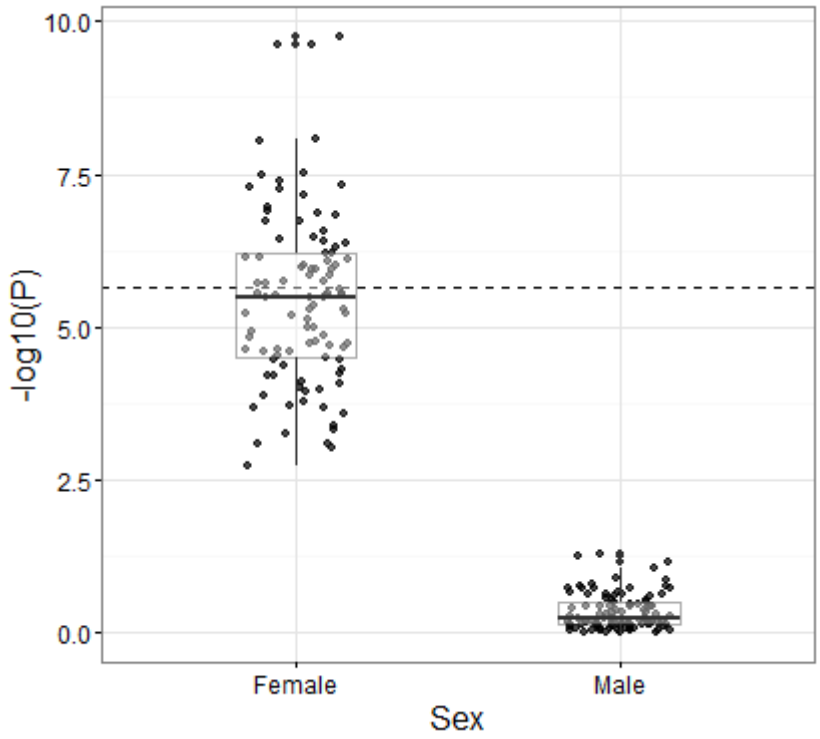

Figure S4. Differences in male and female association statistics from sampled datasets at the most highly associated locus in the genome-wide association study, s74824.1. A total of 1196 autosomal crossover count measures (i.e. the number of male samples in the study) were sampled with replacement from the male and female dataset, and association was tested with s74824.1 genotype. Points show the $-\log_{10}P$ output from 100 iterations of sampling. Female and male $-\log_{10}P$ values were significantly different (Welch's t-test, $P < 0.001$). Sampling was run in R v3.2.3.

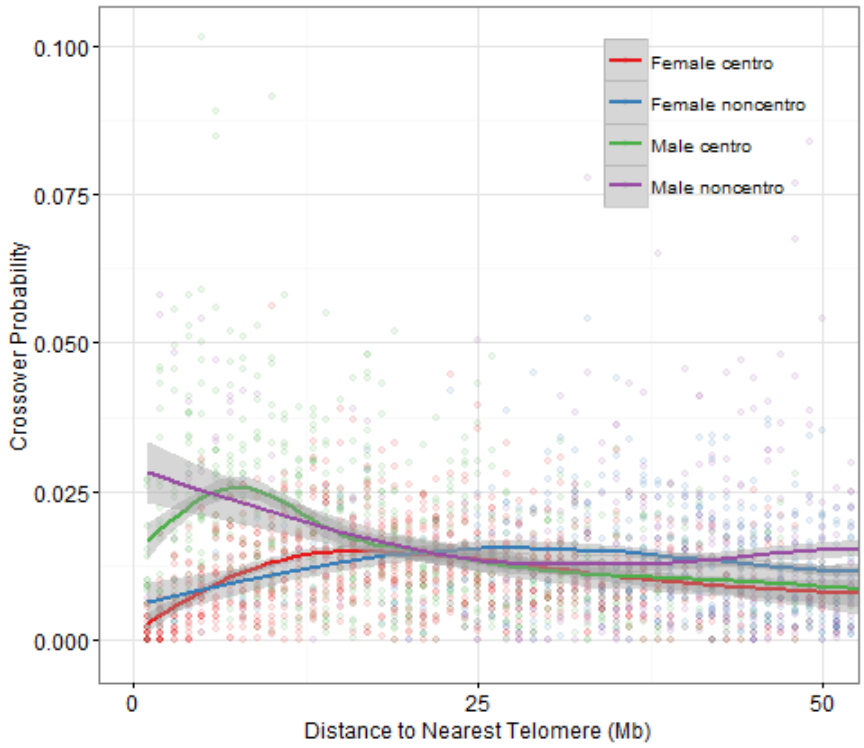

Figure S5. Variation in recombination rate relative to telomeric regions accounting for presence of the centromere on acrocentric chromosomes. Probability of crossing over is given relative to the nearest telomere (Mb) for 1Mb windows. Lines are a generalized additive model smoothing function and are colour coded by sex and presence or absence of centromere ("centro" and "noncentro", respectively). Shaded areas around the line indicate the standard error.

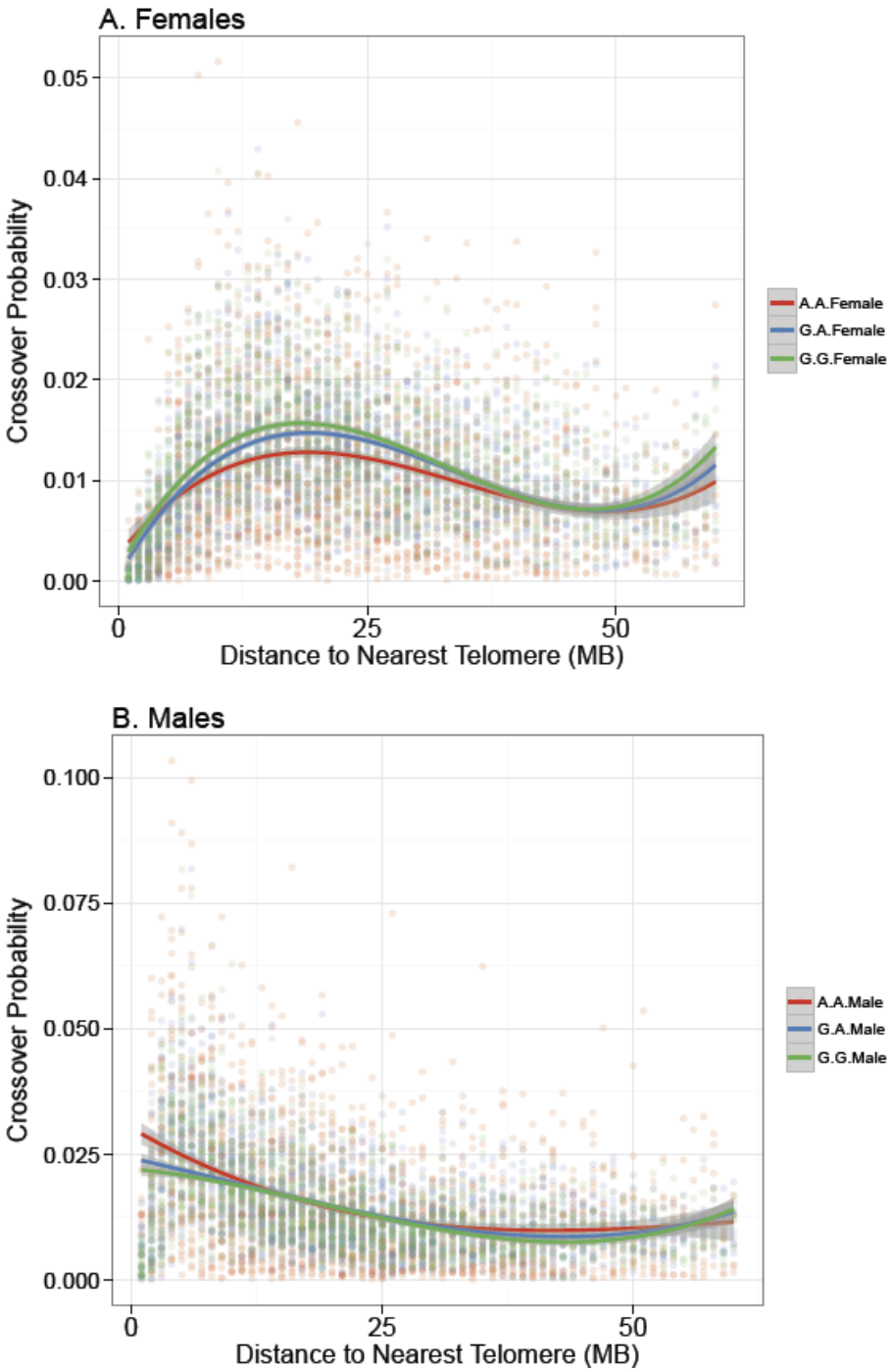

Figure S6. Probability of crossing over relative to the nearest telomere (Mb) for oar3_OAR6_116402578 genotypes in (A) female and (B) male linkage maps. Data points are given for 1Mb windows. Lines indicate the function of best fit. Full model results are given in Table S10.

TABLE S1. CODES AND SAMPLE SIZES FOR DOMESTIC SHEEP BREEDS UTILIZED IN THE HAPLOTYPE SHARING ANALYSIS. Information obtained from the Ovine HapMap project (Kijas *et al.* 2012).

Breed	Code	N
African Dorper	ADP	21
African White Dorper	AWD	6
Afshari	AFS	37
Altamurana	ALT	24
Australian Coopworth	CPW	19
Australian Industry Merino	MER	88
Australian Merino	MER	50
Australian Poll Dorset	APD	108
Australian Poll Merino	APM	98
Australian Suffolk	ASU	109
Bangladeshi BGE	BGE	24
Bangladeshi Garole	BGA	24
Barbados Black Belly	BBB	24
Black Headed Mountain	BHM	24
Border Leicester	BRL	48
Boreray	BOR	20
Brazilian Creole	BCS	23
Bundner Oberlander Sheep	BOS	24
Castellana	CAS	23
Changthangi	CHA	29
Chinese Merino	CME	23
Chios	CHI	23
Churra	CHU	120
Comisana	COM	24
Cyprus Fat Tail	CFT	30
Deccani	IDC	24
Dorset Horn	DSH	21
East Friesian Brown	EFB	39
East Friesian White	EFW	9
Engadine Red Sheep	ERS	24
Ethiopian Menz	EMZ	34
Finnsheep	FIN	99
Galway	GAL	49
Garut	GUR	22
German Texel	GTX	46
Gulf Coast Native	GCN	94
Indian Garole	GAR	26
Irish Suffolk	ISF	55
Karakas	KRS	18
Leccese	LEC	24
Macarthur Merino	MCM	10
Meat Lacaune	LAC	78
Merinolandschaf	MLA	24
Milk Lacaune	LAC	103
Moghani	MOG	34
Morada Nova	BMN	22

Continued…

Breed	Code	N
Namaqua Afrikaner	NQA	12
New Zealand Romney	ROM	24
New Zealand Texel	NTX	24
Norduz	NDZ	20
Ojalada	OJA	24
Old Norwegian Spaelsau	NSP	15
Qezel	QEZ	35
Rambouillet	RMB	102
Rasaaragonesa	RAA	22
Red Maasai	RMA	45
Ronderib Afrikaner	RDA	17
Sakiz	SKZ	22
SantaInes	BSI	47
Sardinian Ancestral Black	SAB	20
Scottish Blackface	SBF	56
Scottish Texel	STX	80
Spael-coloured	NSP	3
Spael-white	NSP	32
St Elizabeth	STE	10
Sumatra	SUM	24
Swiss Black-Brown Mountain Sheep	SBS	24
Swiss Mirror Sheep	SMS	24
Swiss White Alpine Sheep	SWA	24
Tibetan	TIB	37
Valais Blacknose Sheep	VBS	24
Valais Red Sheep	VRS	24
Wiltshire	WIL	23

Kijas J. W., Lenstra J. A., Hayes B., Boitard S., Porto Neto L. R., San Cristobal M., Servin B., McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan J., Dalrymple B., 2012 Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. **10**: e1001258.

Table S2. Sex-averaged and sex specific linkage map positions (CM) and genomic positions (BP) for SNP loci passing quality control. (.txt, 1,834 KB)

Available for download as a .txt file at

www.genetics.org/cgi/data/genetics.115.185553/DC1/11

TABLE S3. FIXED EFFECTS SUMMARY OF GENERAL LINEAR MODELS OF (A) CROSSOVER PROBABILITY AND (B) DIFFERENCE BETWEEN MALE AND FEMALE CROSSOVER PROBABILITIES PER 1MB WINDOW OF THE GENOME. See Materials and Methods for a detailed description of the model. *Dist* is defined as the distance of the window from the nearest telomeric region, estimated relative to the Oar_v3.1 genome build; here it is modelled with a polynomial function, therefore $Dist^3$, $Dist^2$ and *Dist* refer to the cubic, square and linear terms of the model. SNP Count is the number of SNP loci contained within the 1Mb window. The polynomial function of model residuals rooted at 12.45MB.

Model	Effect	Estimate	S.E.	t	P
Crossover	Female (Intercept)	-0.014	2.02×10^{-03}	-6.69	2.49×10^{-11}
Probability	Female $Dist^3$	6.54×10^{-07}	5.84×10^{-08}	11.19	1.13×10^{-28}
	Female $Dist^2$	-6.55×10^{-05}	5.02×10^{-06}	-13.04	4.10×10^{-38}
	Female Dist	1.78×10^{-03}	1.22×10^{-04}	14.58	5.89×10^{-47}
	Male	0.029	1.11×10^{-03}	25.91	3.29×10^{-137}
	Male $Dist^3$	-6.56×10^{-07}	8.24×10^{-08}	-7.96	2.24×10^{-15}
	Male $Dist^2$	7.78×10^{-05}	7.07×10^{-06}	11.01	8.64×10^{-28}
	Male Dist	-2.78×10^{-03}	1.71×10^{-04}	-16.33	4.25×10^{-58}
	SNP Count	2.81×10^{-04}	3.65×10^{-05}	7.70	1.69×10^{-14}
	GC content (%)	0.024	3.74×10^{-03}	6.50	9.25×10^{-11}
Model	Effect	Estimate	S.E.	t	P
Male:Female	Intercept	0.012	2.99×10^{-03}	4.02	6.07×10^{-05}
Crossover	$Dist^3$	-6.38×10^{-07}	6.38×10^{-08}	-10.00	5.18×10^{-23}
Probability	$Dist^2$	7.57×10^{-05}	5.49×10^{-06}	13.78	2.37×10^{-41}
	Dist	-2.68×10^{-03}	1.34×10^{-04}	-19.98	2.76×10^{-81}
	SNP Count	1.41×10^{-04}	5.62×10^{-05}	2.50	0.012
	GC content (%)	0.032	5.76×10^{-03}	5.48	4.74×10^{-08}

Table S4. Full results of genome-wide association studies for ACC in both sexes and males and females only. Type indicates whether ACC was modelled as the total number of crossovers across all autosomes ("cistrans") or the total number on all autosomes minus the number of crossovers on the focal chromosome ("trans"; see Materials and Methods). Df, Wald.statistic and Pr.Chisq. are the statistical results of the GWAS. A and B are the reference alleles. N is the number of unique individuals in the analysis. CallRate and MAF are the call rate and minor allele frequency of the SNP locus. Solution, std.error and z.ratio are the effect size, standard error and the Z ratio of each genotype relative to the model intercept for genotypes AA, AB and BB; cases where z values are 'NA' indicate the model intercept. (.txt, 38 MB)

Available for download as a .txt file at

www.genetics.org/cgi/data/genetics.115.185553/DC1/13

Table S5. Full results of the regional heritability analysis of ACC for sliding window sizes of 150, 50 and 20 SNPs across all autosomes. Models were run in all sheep, and in males and females only. Type indicates whether ACC was modelled as the total number of crossovers across all autosomes ("cistrans") or the total number on all autosomes minus the number of crossovers on the focal chromosome ("trans"; see Materials and Methods). First.SNP.Pos and Last.SNP.Pos indicate the positions of the first and last SNP within the window. RestofGenome.GRM.h2 and Window.GRM.h2 are the proportions of phenotypic variance attributed to the rest of the genome and the sliding window, respectively (SE is their standard errors). ASReml.Error indicates cases where models had no errors ("none") or abnormal termination ("A.T."). Window.GRM.h2.constraint indicates cases where the estimate of heritability was positive or fixed at the boundary (i.e. not significantly different from zero). LL indicates the log likelihoods of animal models without and with the window, respectively. The $\chi 2$ and P values are from likelihood ratio tests. (.txt, 4,561 KB)

Available for download as a .txt file at

www.genetics.org/cgi/data/genetics.115.185553/DC1/14

Table S6. Full results of association statistics at imputed SNPs from the ovine HD SNP chip in the sub-telomeric region of chromosome 6 for ACC. Description and headers are the same as for Table S5. Array indicates which SNPs were also present on the Ovine SNP50 BeadChip. Q2 indicates the minor allele frequencies of the imputed SNPs and MAF.Original indicates the minor allele frequencies in the 187 individuals typed on the HD chip. (.txt, 101 KB)

Available for download as a .txt file at

www.genetics.org/cgi/data/genetics.115.185553/DC1/15

Table S7. Results of the blast analysis outlined in File S3. PC.Match is the percentage of sequence matching between the aligned segments. Matches, mismatches and gaps are the number of each term in the sequence, respectively. *Bos* and *Ovis* indicate cattle and sheep, respectively; start and stop positions are given for each aligned segment; mean indicates the mean value of the start and stop positions. (.txt, 23 KB)

Available for download as a .txt file at

www.genetics.org/cgi/data/genetics.115.185553/DC1/16

TABLE S8. DETAILS OF CORE HAPLOTYPES (~700KB, 6 SNPS) IN SOAY SHEEP SPANNING OAR3_OAR6_116402578 ON CHROMOSOME 6, AND THE EXTENT OF HAPLOTYPE SHARING (HS) WITH DOMESTIC SHEEP. HS distances are given in MB. SD is the standard deviation.

			Distance from oar3_OAR6_116402578 (Kb)										
Haplotype	Recombination Allele	Haplotype Frequency	-420.6	-411.5	-78.3	38.9	65.8	266.3	# ISGC Breeds	Global HS Mean	SD	Maximum HS Breed	Maximum HS Mean
H1	High	0.026	A	A	A	A	G	A	22	0.10	0.11	Australian Poll Merino	0.29
H2	High	0.311	A	A	A	A	G	G	67	0.04	0.07	Australian Coopworth	0.30
H3	High	0.030	A	A	A	G	A	A	23	0.14	0.10	Scottish Blackface	0.29
H4	High	0.015	A	A	A	G	G	A	20	0.04	0.06	Barbados BlackBelly	0.31
H5	High	0.355	A	A	G	G	G	A	29	0.09	0.10	Australian Poll Dorset	0.12
H6	Low	0.220	A	A	G	A	G	G	66	0.10	0.10	Australian Poll Dorset	0.34
H7	Low	0.042	G	G	G	A	G	G	46	0.14	0.15	Chinese Merino	0.38

TABLE S9. VARIANCE IN ACC EXPLAINED BY THE GENOMIC RELATEDNESS ($V_{A\ POLY}$), THE 20 SNP WINDOW CONTAINING RNF212 (V_{AREG1}) AND THE 20 SNP WINDOW CONTAINING REC8 (V_{AREG2}) AND RESIDUAL EFFECTS (V_E). Sample sizes are for Table 1. Effect sizes are given as the proportion of the phenotypic variance explained by the variance effects (h^2, $reg1^2$ and $reg2^{2,}$ respectively.)

	$V_{A\ poly}$	V_{Areg1}	V_{Areg2}	V_E	h^2_{poly}	$reg1^2$	$reg2^2$	e^2
Both Sexes	2.071	1.699	1.177	24.903	0.069	0.057	0.039	0.834
	(0.49)	(0.824)	(0.634)	(0.655)	(0.016)	(0.026)	(0.021)	(0.028)
Male	1.902	0.190	1.044	22.252	0.075	0.007	0.041	0.876
	(0.841)	(0.375)	(0.889)	(0.961)	(0.033)	(0.015)	(0.034)	(0.034)
Females	1.976	2.589	0.945	26.363	0.062	0.081	0.030	0.827
	(0.602)	(1.198)	(0.6)	(0.877)	(0.019)	(0.035)	(0.019)	(0.034)

TABLE S10. FIXED EFFECTS SUMMARY OF GENERAL LINEAR MODELS OF CROSSOVER PROBABILITY IN (A) FEMALES AND (B) MALES PER 1MB WINDOW OF THE GENOME. Crossover probabilities are modelled in males and females and with an interaction term between genotype at oar3_OAR6_116402578 and Dist, which is defined as the distance of the window from the nearest telomeric region, estimated relative to the Oar_v3.1 genome build. Dist is modelled with a polynomial function; therefore $Dist^3$, $Dist^2$ and Dist refer to the cubic, square and linear terms of the model. The symbol × indicates an interaction term. SNP Count accounts for the number of SNP loci within the bin.

Sex	Model.Term	Effect Size	S.E.	t	P
(A) Female	A/A (Intercept)	-3.98×10^{-03}	1.19×10^{-03}	-3.36	7.96×10^{-04}
	G/A	-2.00×10^{-03}	8.30×10^{-04}	-2.41	0.016
	G/G	-1.38×10^{-03}	8.18×10^{-04}	-1.68	0.093
	$Dist^3$	4.41×10^{-07}	4.06×10^{-08}	10.8	3.96×10^{-27}
	$Dist^2$	-4.50×10^{-05}	3.59×10^{-06}	-12.6	1.06×10^{-35}
	Dist	1.25×10^{-03}	9.07×10^{-05}	13.7	2.81×10^{-42}
	SNP Count	1.59×10^{-04}	1.94×10^{-05}	8.17	3.67×10^{-16}
	GC content (%)	9.12×10^{-03}	2.03×10^{-03}	4.5	6.98×10^{-06}
	G/A : $Dist^3$	1.73×10^{-07}	5.58×10^{-08}	3.1	1.95×10^{-03}
	G/G : $Dist^3$	2.55×10^{-07}	5.54×10^{-08}	4.61	4.17×10^{-06}
	G/A : $Dist^2$	-1.72×10^{-05}	4.88×10^{-06}	-3.53	4.11×10^{-04}
	G/G : $Dist^2$	-2.37×10^{-05}	4.83×10^{-06}	-4.9	1.00×10^{-06}
	G/A : Dist	4.72×10^{-04}	1.21×10^{-04}	3.89	1.00×10^{-04}
	G/G : Dist	5.81×10^{-04}	1.20×10^{-04}	4.84	1.32×10^{-06}
(B) Male	A/A (Intercept)	0.014	1.92×10^{-03}	7.37	1.95×10^{-13}
	G/A	-5.81×10^{-03}	1.28×10^{-03}	-4.55	5.41×10^{-06}
	G/G	-7.94×10^{-03}	1.27×10^{-03}	-6.26	4.00×10^{-10}
	$Dist^3$	-9.60×10^{-08}	6.79×10^{-08}	-1.41	0.157
	$Dist^2$	1.92×10^{-05}	5.87×10^{-06}	3.27	1.08×10^{-03}
	Dist	-1.09×10^{-03}	1.44×10^{-04}	-7.55	5.01×10^{-14}
	SNP Count	2.63×10^{-04}	3.22×10^{-05}	8.15	4.36×10^{-16}
	GC content (%)	2.59×10^{-02}	3.31×10^{-03}	7.82	6.46×10^{-15}
	G/A : $Dist^3$	2.56×10^{-07}	9.18×10^{-08}	2.79	5.21×10^{-03}
	G/G : $Dist^3$	3.68×10^{-07}	9.15×10^{-08}	4.02	5.94×10^{-05}
	G/A : $Dist^2$	-2.50×10^{-05}	7.92×10^{-06}	-3.15	1.62×10^{-03}
	G/G : $Dist^2$	-3.53×10^{-05}	7.89×10^{-06}	-4.48	7.67×10^{-06}
	G/A : Dist	7.06×10^{-04}	1.93×10^{-04}	3.66	2.56×10^{-04}
	G/G : Dist	9.71×10^{-04}	1.92×10^{-04}	5.06	4.32×10^{-07}

File S1 - Simulation of recombination events.

To verify the accuracy of our estimation of autosomal crossover count (ACC), we simulated meiotic crossovers within the Soay sheep pedigree using information from linkage map positions and allele frequencies, and then compared the observed and true crossover counts in simulated datasets.

Simulations were conducted using the entire Soay sheep pedigree. Each individual was assigned to an ordered cohort using the *kindepth* function in the package kinship2 v1.6.0 (Therneau *et al.* 2014) implemented in R v3.1.1. For all pedigree links where a parent was unknown (i.e. founder individuals), haplotypes were simulated for each chromosome, where each SNP allele was sampled with a probability equal to its minor allele frequency across the entire genomic dataset. In order to simulate recombination events in transmitted haplotypes, we used the following approach sequentially across each ordered cohort. Positions of meiotic crossovers were determined by sampling the likelihood of a crossover between each pair of adjacent markers with probability *r*, where *r* was the sex-specific recombination fraction determined from the linkage map. As cross-over interference was apparent in the main analysis, positions were re-sampled in cases where two crossover events occurred within 10cM of each other. Recombinant haplotypes were constructed based on the simulated crossover positions between the two haplotypes within the parent, and one recombinant haplotype was sampled with a probability of 0.5 and transmitted to the offspring. The positions and counts of crossovers per chromosome were recorded. Each haplotype pair was then condensed into an unphased SNP genotype dataset. Missing genotypes and genotype errors were assigned to genotypes at random, with probabilities of

1×10^{-3} and 1×10^{-4}, respectively. An R function for this process has been archived at http://github.com/susjoh/simperSNP. Individual ACCs were then calculated in FIDs in each simulated dataset as using the same analysis outlined in the methods section "Estimation of meiotic crossovers".

Our approach was highly accurate in identifying the true simulated crossover count across all individuals (linear regression of observed counts as a function of true simulated counts, $\bar{b}$ = 1.014 ± 1.51 × 10^{-3} (s.d.), mean adjusted r^2 = 0.991 ± 6.7 × 10^{-4}). The per-individual correlations across all simulations were also high ($\bar{b}$ = 0.992 ± 0.013, mean adjusted r^2 = 0.99 ± 0.011). A general linear model of per individual mean adjusted r^2 was fitted, including sex and genomic inbreeding coefficient $\hat{F}$ as fixed effects. Adjusted r^2 values and slopes were significantly higher in females than in males (P < 0.001, Table 1, Figure 1); This may be due to increased crossover rates in males resulting in some crossovers not being identified; the removal of single SNP runs may also have more acute effects at telomeric regions of chromosomes, where crossovers may be more likely to occur in males. Adjusted R^2 values and slopes were also negatively correlated with the genomic inbreeding coefficient $\hat{F}$ (P < 0.001, Table 1, Figure 1), which may be due to an increased chance of runs of homozygous genotypes, meaning that substantial regions of the genome cannot be assigned to a particular grandparent of origin. Double crossovers occurring within such regions are therefore less likely to be identified. In both cases, slopes are lower than 1, indicating that crossover counts are consistently under-estimated, albeit to a small degree, in males and/or in more inbred individuals (Table 1).

References

Therneau T., Atkinson E., Sinnwell J., Schaid D., McDonnell S., 2014 kinship2 v1.6.0.

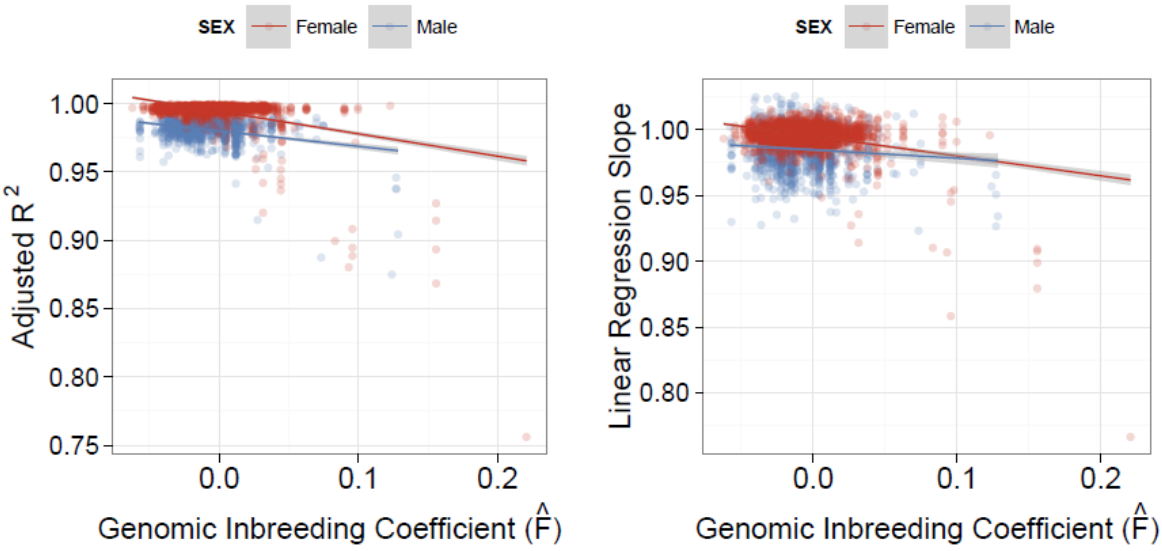

Fig 1. Linear regressions of individual genomic inbreeding coefficient and (A) the correlation between and (B) the regression slopes between simulated numbers of crossover counts and the number of crossovers detected per individual, where the detected counts were modelled as a function of true simulated counts. Red and blue points indicate those observed for males and females, respectively.

Table 1. General linear model of the individual correlations and slopes of detected vs. Simulated acc per individual over 100 simulations.

Value	Effect	Estimate	S.E.	t	P
Adjusted r^2	Female (Intercept)	0.994	1.8×10^{-04}	5536	<0.001
	Male	-0.015	2.9×10^{-04}	-49.66	<0.001
	$\hat{F}$	-0.147	6.0×10^{-03}	-24.39	<0.001
Slope	Female (Intercept)	0.9953	2.6×10^{-04}	3768.8	<0.001
	Male	-0.011	4.4×10^{-04}	-25.35	<0.001
	$\hat{F}$	-0.123	8.9×10^{-03}	-13.83	<0.001

File S2 - Predicted location of *RNF212* on the Sheep genome Oar_v3.1

The locus *RNF212* has not been annotated on the domestic sheep genome (Oar_v3.1). Here, we present evidence that *RNF212* occurs within the region most highly associated with autosomal crossover count (ACC). Sheep sequences for *RNF212* were obtained from Genbank (Accession Numbers provided in Figure 1 and aligned against the Oar_v3.1 genome sequence using BLAST v2.2.27+ using the default parameters and a word size of 15 to determine if the locus was likely to be present within this region. Matches of length < 50 bp and matching bases of < 95% were discarded .The strongest associations on the assembled genome sequence were at the sub-telomeric region of chromosome 6 identified in the main results, falling between positions 116,427,802 and 116,446,904 (Figure 1). This positioned *RNF212* between an uncharacterised protein (orthologous to *Tmed11*) and *FGFRL1*, and suggested that its lack of annotation on the domestic sheep genome is due to its occurrence across a gap in contiguous sequences (Figure 2). Gene order in this region showed high similarity to gene orders obtained for homologous genome regions in cow (UMD3.1), dog (CanFam3.1), human (GRCh38) and mouse (GRCm38); all annotations were obtained from Ensembl build v1.79 (Cunningham *et al.* 2014). The alignment of *RNF212* shown in Figure 5 (main text) and Figure 2 is to sequence XM_012167676.1.

References

Cunningham F., Amode M. R., Barrell D., Beal K., Billis K., Brent S., Carvalho-Silva D., Clapham P., Coates G., Fitzgerald S., Gil L., Giron C. G., Gordon L., Hourlier T., Hunt S. E., Janacek S. H., Johnson N., Juettemann T., Kahari a. K., Keenan S., Martin F. J., Maurel T., McLaren W., Murphy D. N., Nag R., Overduin B., Parker a., Patricio M., Perry E., Pignatelli M., Riat H. S., Sheppard D., Taylor K., Thormann a., Vullo a., Wilder S. P.,

Zadissa a., Aken B. L., Birney E., Harrow J., Kinsella R., Muffato M., Ruffier M., Searle S. M. J., Spudich G., Trevanion S. J., Yates a., Zerbino D. R., Flicek P., 2014 Ensembl 2015. Nucleic Acids Res. **43**: D662–D669.

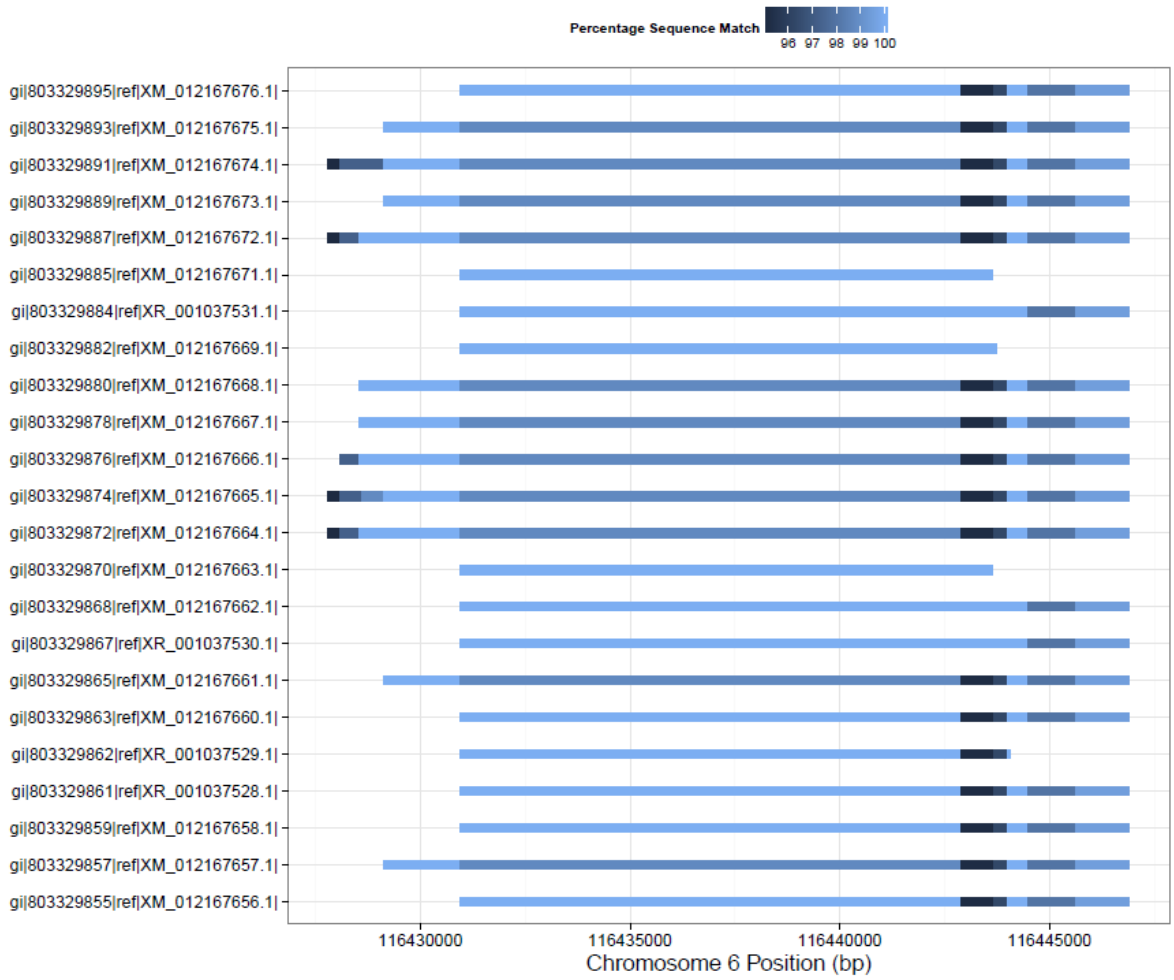

Fig 1. Alignments of 23 Ovis *RNF212* nucleotide sequences (obtained from Genbank) against the sub-telomeric regions of Oar_v3.1 chromosome 6. Genbank IDs for each sequence are given on the left axis. The color indicates the percentage sequence match for each segment of aligned sequence.

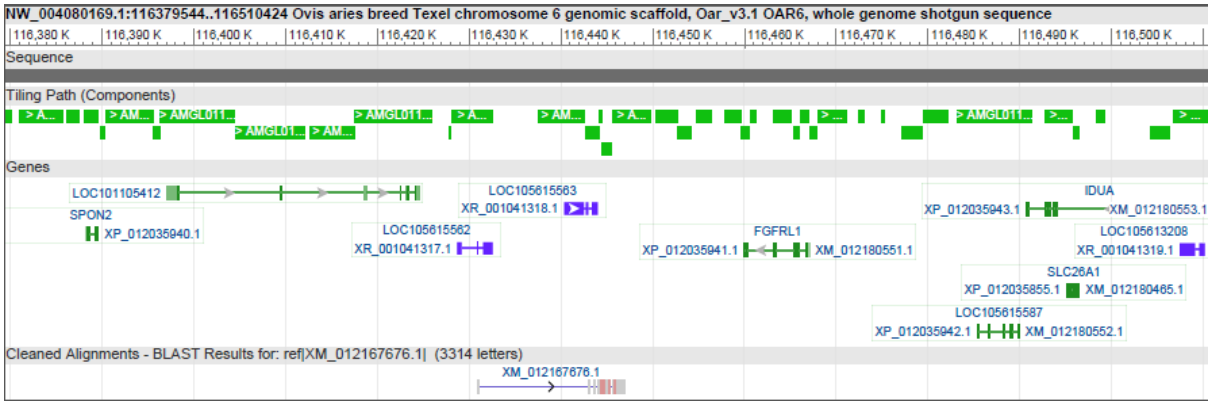

Fig 2. Putative positioning of *RNF212* (Accession XM_012167676.2) against the sub-telomeric region of Oar_v3.1 chromosome 6. The tiling path in green indicates the contiguous sequence that contributes to the genome sequence used in this study. Other annotated loci within this area are indicated in the segment above the putative *RNF212* position. Figure obtained using the NCBI Genome Viewer (http://www.ncbi.nlm.nih.gov/)

File S3 - BLAST comparison of the sub-telomeric region of chromosome 6 with the homologous region of the cattle genome.

As the sheep genome assembly is still ongoing, we wished to determine if a region of reduced linkage disequilibrium within an otherwise complete LD haplotype block at the sub-telomeric region on chromosome 6 was the consequence of the region being assembled incorrectly. We compared DNA sequence from a 2.7Mb region of cattle chromosome 6 (positions 107,699,493 to 110,430,052, cattle genome vUMD3.1) homologous to the 2.1Mb sub-telomeric region of sheep chromosome 6 (positions 115,000,000 to 117,031,472, sheep genome Oar_v3.1) containing candidate regions associated with recombination rate (see main text results and Figure 5). Sequence comparison was carried out using BLAST v2.2.27+ with the default parameters and a word size of 20. BLAST hits of < 2000 bases and < 80% match were discarded. Regions with more than two hits per ~2000bp window were also discarded as they were indicative of repetitive sequence. Our results indicate that a sheep region from 116.725Mb to 116.823Mb is likely to have been incorrectly assembled, and is likely to occur at a position between 116.30 Mb and 116.36Mb (Figure 1, Table S7).

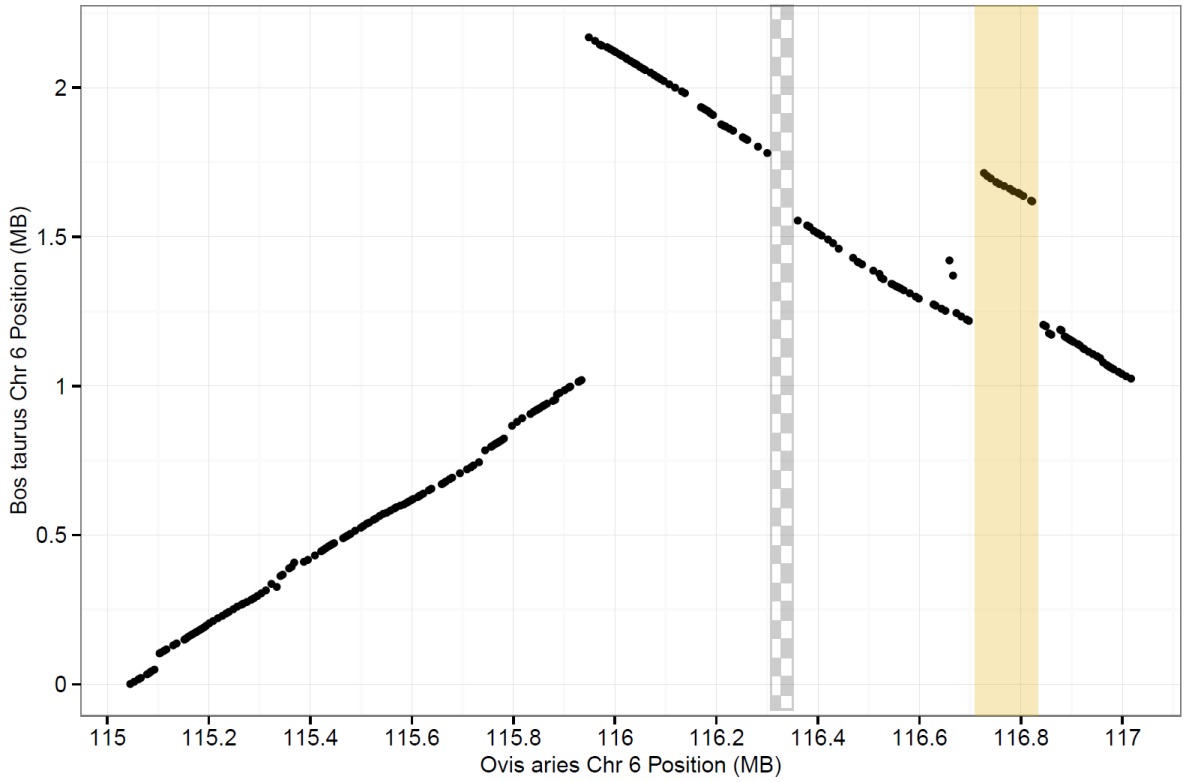

Fig 1. BLAST results comparing a 2.7Mb region of cattle chromosome 6 (*Bos taurus*, positions 107,699,493 to 110,430,052, cattle genome vUMD3.1) with a homologous 2.1Mb sub-telomeric region of sheep chromosome 6 (*Ovis aries*, positions 115,000,000 to 117,031,472, sheep genome Oar_v3.1). Points indicate the mid position of each BLAST hit on their respective query sequence. The beige block indicates the position of a fragment on the sheep genome which is likely to have been incorrectly placed – its most likely position is indicated by the grey chequered block. BLAST results are provided in Table S7.