bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Susan E. Johnston[1], Jon Slate[2], Josephine M. Pemberton[1]

[1] Institute of Evolutionary Biology, University of Edinburgh, Charlotte Auerbach Road, Edinburgh, EH9 3FL, United Kingdom.

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

Abstract

Meiotic recombination breaks down linkage disequilibrium and forms new haplotypes, meaning that it is an important driver of diversity in eukaryotic genomes. Understanding the causes of variation in recombination rate is not only important in interpreting and predicting evolutionary phenomena, but also for understanding the potential of a population to respond to selection. Yet, there remains little data on if, how and why recombination rate varies in natural populations. Here, we used extensive pedigree and high-density SNP information in a wild population of Soay sheep (*Ovis aries*) to determine individual crossovers in 3330 gametes from 813 individuals. Using these data, we investigated the recombination landscape and the genetic architecture of individual autosomal recombination rate. The population was strongly heterochiasmic (male to female linkage map ratio = 1.31), driven by significantly elevated levels of male recombination in sub-telomeric regions. Autosomal recombination rate was heritable in both sexes (h^2 = 0.16 & 0.12 in females and males, respectively), but with different genetic architectures. In females, 46.7% of heritable variation was explained by a sub-telomeric region on chromosome 6; a genome-wide association study showed the strongest associations at *RNF212,* with further associations observed at a nearby ~374kb region of complete linkage disequilibrium containing three additional candidate loci, *CPLX1*, *GAK* and *PCGF3*. This region did not affect male recombination rate. A second region on chromosome 7

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

containing *REC8* and *RNF212B* explained 26.2% of heritable variation in recombination rate in both sexes, with further single locus associations identified on chromosome 3. Our findings provide a key empirical example of the genetic architecture of recombination rate in a natural mammal population with male-biased crossover frequency.

Author Summary.

For almost 50 years we have known that genetic linkage can constrain responses to selection, while recombination can offer an escape from this constraint by forming new combinations of alleles. This increases the genetic variance for traits and hence the potential of a population to respond to selection. Therefore, understanding the causes and consequences of variation in the rate of recombination is important, not only in interpreting and predicting evolutionary phenomena, but for applications in genetic improvement of domesticated species. In our study, we used extensive genomic and pedigree information to identify genes underlying individual recombination rate variation in a wild population of Soay sheep on the St Kilda archipelago, NW Scotland. We show that the rate of recombination is partly inherited, that it is 30% higher in males than females, and that the majority of genetic variation in female sheep is likely to be explained by a genomic region containing the gene *RNF212*. This finding shows that recombination rate has the potential to evolve within wild populations.

Introduction.

Recombination is a fundamental process ensuring the proper segregation of homologous chromosomes during meiosis [1,2], but also plays an important role in driving the evolution of eukaryotic genomes, because it rearranges existing allelic variation to create novel haplotypes. Theoretical work predicts that recombination prevents the accumulation of deleterious mutations, uncoupling them from linked beneficial alleles [3,4], and can also lead to an increase in genetic

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

variance for fitness, increasing the efficiency of selection within finite populations [5–9] (experimental examples [10,11]). However, recombination is also associated with fitness costs: higher rates of crossing-over may increase the risk of deleterious chromosomal rearrangements [12], or lead to the break-up of favourable combinations of alleles previously built up by selection, reducing the mean fitness of subsequent generations [7]. Therefore, understanding the causes and consequences of variation in recombination rate is not only important in interpreting and predicting evolutionary phenomena in the context of sex, selection and drift [8,13,14], but also has potential applications in genetic improvement through increased responses to selection [15].

Recent studies of model mammal systems have shown that crossover frequency can vary both within and between chromosomes, individuals, populations and species [16–18]. This may be partially driven by physical mechanisms affecting chiasma positioning (e.g. obligate crossing-over [19], crossover interference [20,21], GC content [22] and/or proximity to telomeres and centromeres [23]), but is also likely to be driven by heritable genetic effects [24–26]. Genome-wide association studies in humans, cattle and mice have attributed some heritable variation to specific genetic variants, including ring finger protein 212 (*RNF212*), meiotic recombination protein REC8 (*REC8*) and PR domain zinc finger protein 9 (*PRDM9*) [26–29]. This relatively simple genetic architecture implies that recombination rate has the potential to respond rapidly to selection; for example, mutations in *PRDM9* in humans and other primates can result in large changes in recombination landscapes over relatively short evolutionary timescales [30,31].

One additional observation across eukaryotes is that recombination rate often varies between the sexes. The absence of recombination in one sex (achiasmy) is almost always in the heterogametic sex, and may be a pleiotropic effect of selection for tight linkage on Y and Z chromosomes (Haldane-Huxley rule, [32–34]). However, in species where recombination rate varies between the sexes (heterochiasmy), this rule no longer holds, as recombination rates can be male or female biased,

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

even within hermaphroditic species or species with no sex chromosomes or sex determining loci [35]. In mammals, females generally have higher numbers of crossovers across the genome as a whole [35]; this may be partially due to differences in chromatid structure [36–38], but there may also be strong selection for increased recombination rate in females as a mechanism to avoid aneuploidy (i.e. incorrect chromosome number) after long periods of meiotic arrest before fertilisation [39–41]. However, some mammal species exhibit higher recombination rates in males, such as domestic sheep [42], macaques [43] and marsupials [43,44]. Several selective processes have been suggested to explain variation in heterochiasmy, including haploid selection [35], meiotic drive [45], sperm competition, sexual dimorphism and dispersal [43,46,47]. Nevertheless, a clear understanding has been limited by a paucity of empirical examples, particularly in systems with male-biased heterochiasmy.

Here, we examine the genetic architecture of recombination rate variation in a wild mammal population with strong male-biased heterochiasmy. Soay sheep (*Ovis aries*) are a Neolithic breed of domestic sheep that have lived unmanaged on the St Kilda archipelago (Scotland, UK, 57º49'N, 8º34'W) since the Bronze age [48]. Individuals from the Village Bay area of archipelago have been subject to a long-term study since 1985, with extensive genomic, pedigree, phenotype and environmental data collected for more than 6,000 individuals. In this study, we integrate genomic and pedigree information to characterise autosomal cross-over positions in more than 3000 gametes in individuals from both sexes. We show that individual recombination rate in Soay sheep is heritable and strongly heterochiasmic, and identified two genomic regions associated with rate variation: one containing *RNF212*, which explained a 46.7% of additive genetic variance in females alone; and another containing *REC8* and *RNF212B*, which explained 26.2% of additive genetic variance across both sexes. These results provide new information on genetic architecture of recombination in system with strongly sexually-dimorphic rate variation.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Results.

Broad-scale variation in recombination landscape.

We used pedigree information and data from 39104 genome-wide SNPs typed on the Ovine SNP50 BeadChip [49] to identify 98420 meiotic crossovers in gametes transferred to 3330 offspring from 813 unique focal individuals. A linkage map of all 26 autosomes had a sex-averaged length of 3304 centiMorgans (cM), and sex-specific lengths of 3748 cM and 2860 cM in males and females, respectively, indicating strong male-biased heterochiasmy in this population (Male:Female linkage map lengths = 1.31; Fig S1, Table S1). There was a linear relationship between the length of autosomes in megabases (Mb) and linkage map lengths (cM; Adjusted $R^2 = 0.991$, $P < 0.001$; Fig 1A). Chromosome-wide recombination rates (cM/Mb) were higher in smaller autosomes (fitted as multiplicative inverse function, adjusted $R^2 = 0.616$, $P < 0.001$, Fig 1B), indicative of obligate crossing over. The degree of heterochiasmy based on autosome length in cM (i.e. differences in male and female recombination rate) was consistent across all autosomes (Adjusted $R^2 = 0.980$, $P < 0.001$, Fig 1C).

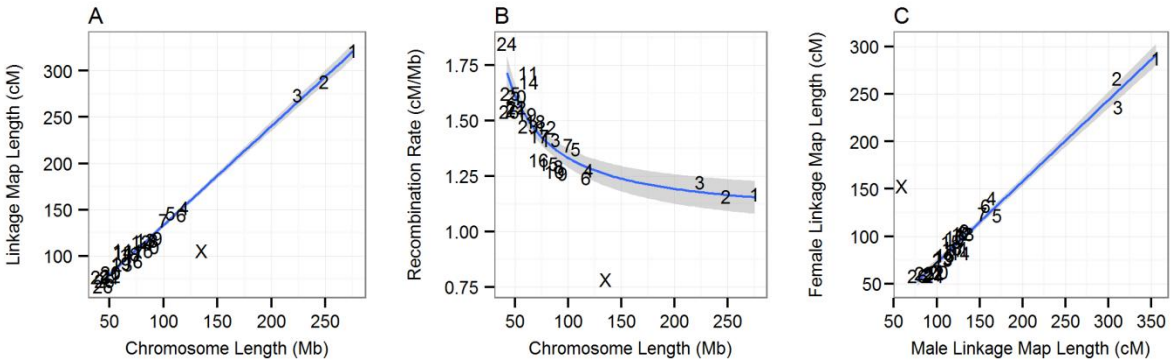

Fig 1. Broad scale variation in recombination rate.
Relationships between: (A) sex-averaged linkage map length (cM) and physical chromosome length (Mb); (B) physical chromosome length (Mb) and recombination rate (cM/Mb); and (C) male and female linkage map lengths (cM). Points are chromosome numbers. Lines and the grey shaded areas indicate the regression slopes and standard errors, respectively, excluding the X chromosome. NB. The male linkage map length for the X chromosome is equivalent to the length of the pseudo-autosomal region.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fine-scale variation in recombination landscape.

Finer-scale probabilities of crossing-over were calculated for 1Mb windows across the genome for each sex, using recombination fractions from their respective linkage maps. Crossover probability was variable relative to proximity to telomeric regions, with a significant interaction between sex and distance to the nearest telomere fitted as a cubic polynomial function (Fig 2A). Males had significantly higher probabilities of crossing-over than females between distances of 0Mb to 12.45Mb from the nearest telomere (Fig 2B, Table S2). Crossover probabilities were significantly higher in windows with higher GC content (General linear model, $P < 0.001$; Table S2). Investigation of the relative distances between crossovers (in cases where two or more crossovers were observed on a single chromatid) indicated that there is crossover interference within this population, with a median distance between double crossovers of 48Mb (Fig S2).

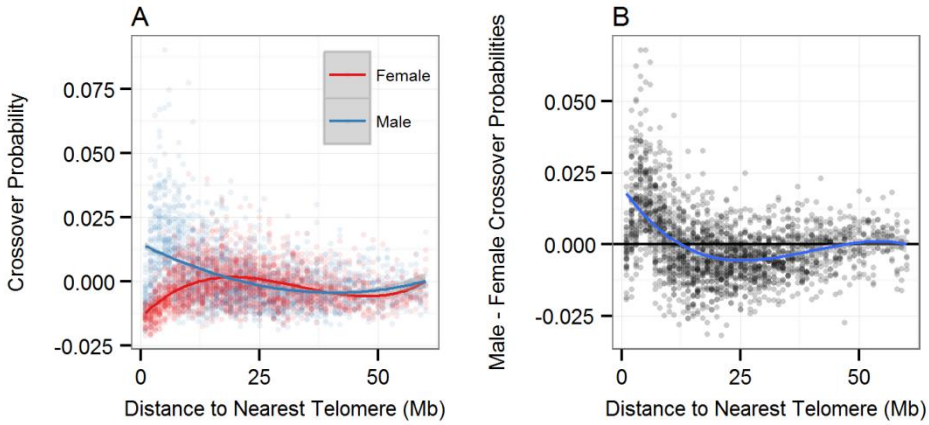

Fig 2. Variation in recombination rate relative to telomeric regions.
Probability of crossing over relative to the nearest telomere (Mb) for (A) female and male linkage maps individually and (B) the difference between male and female crossover probabilities (male minus female). Data points are given for 1Mb windows, and are the residuals of a general linear model with terms fitted for GC % and the number of SNP loci within each window. Lines indicate the function of best fit.

Analysis of individual recombination rate.

Individual recombination rates were defined as the number of autosomal crossover counts observed per gamete (hereafter ACC) transmitted from a focal individual (hereafter FID) to its offspring. To

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

determine the heritability of ACC, variance was partitioned using a restricted maximum likelihood animal model approach [50], and additive genetic variance was modelled with a genomic relatedness matrix constructed using autosomal SNP marker information. Individual ACC was heritable (h^2_g= 0.145, SE = 0.027), with the remainder of the phenotypic variance being explained by the residual error term (Table 1). ACC was significantly higher in males than in females, with 7.376 (SE = 0.263) more crossovers observed per gamete (Animal Model, Z = 28.02, $P_{Wald} < 0.001$). However, females had marginally higher additive genetic variance $P_{LRT} = 0.046$) and higher residual variance ($P_{LRT} = 1.07 \times 10^{-3}$) in ACC than males (Table 1). There was no relationship between ACC and FID age, offspring sex, or the genomic inbreeding coefficients of the FID or offspring; furthermore, there was no variance in ACC explained by common environmental effects such as FID birth year, year of gamete transmission, or maternal/paternal identities of the FID (Animal Models, P > 0.05). A bivariate model of male and female ACC showed that the cross-sex additive genetic correlation (r_a) was 0.808 (SE = 0.202); this correlation was significantly different from 0 ($P_{LRT} < 0.001$) but not different from 1 ($P_{LRT} = 0.308$).

Table 1. Dataset information and animal model results of autosomal crossover count (ACC). N_{OBS}, N_{FIDs} and N_{Xovers} indicate the number of ACC measures, the number of FIDs and the total number of crossovers observed, respectively. Mean is that of the raw data, and V_P and V_A are the phenotypic and additive genetic variances, respectively. The heritability h^2 and residual effect e^2 are the proportions of phenotype variance explained by the additive genetic and residual variances, respectively. *P* (h^2) is the significance of the additive genetic effect (h^2) as determined using a model comparison approach (see text). V_A and heritability were modelled using genomic relatedness. Figures in brackets are standard errors.

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

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Genetic architecture of recombination rate.

Partitioning variance by genomic region. The contribution of specific genomic regions to ACC was determined by partitioning the additive genetic variance in sliding windows (Regional heritability analysis, Table S3 [51]). There was a strong sex-specific association of ACC in females within a sub-telomeric region on chromosome 6 (20 SNP sliding window; Fig 3B). This corresponded to a 1.46 Mb segment containing ~37 protein coding regions, including ring finger protein 212 (*RNF212*), a locus previously implicated in recombination rate variation [26,27,52]. This region explained 8.02% of the phenotypic variance (SE = 3.55%) and 46.7% of the additive genetic variance in females ($P_{LRT} = 9.78 \times 10^{-14}$), but did not contribute to phenotypic variation in males (0.312%, SE = 1.2%, $P_{LRT} = 0.82$; Fig 3C, Table S3). There was an additional significant association between ACC in both sexes and a region on chromosome 7, corresponding to a 1.09Mb segment containing ~50 protein coding regions, including *RNF212B* (a paralogue of *RNF212*) and meiotic recombination protein locus *REC8* ($P_{LRT} = 3.31 \times 10^{-6}$, Fig 3A, Table S3). This region explained 4.12% of phenotypic variation (SE = 2.3%) and 26.2% of the additive genetic variation in both sexes combined; however, this region was not significant associated with ACC within each sex individually after correction for multiple testing (Table S3).

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

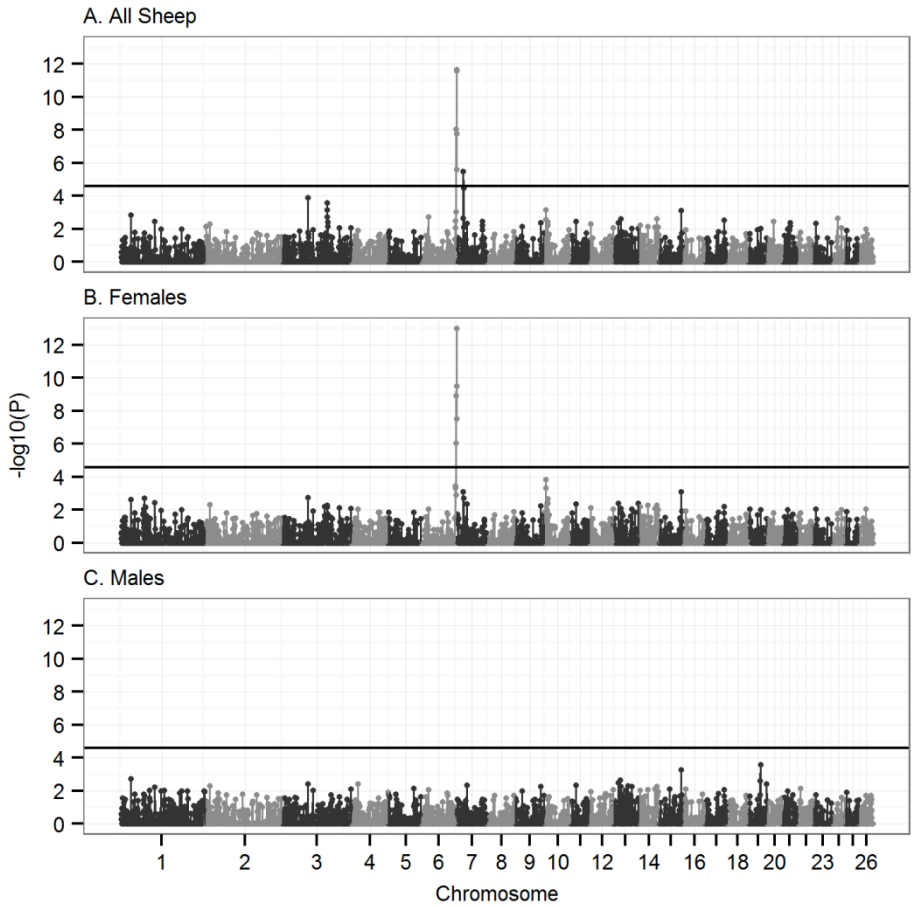

Fig 3. Association analysis of autosomal crossover count using a regional heritability approach. Significance of association analysis in (A) all sheep, (B) females only and (C) males only. The results presented are from a sliding window of 20 SNPs across 26 autosomes, with an overlap of 10 SNPs (see main text). Points represent the median base pair position of all SNPs within the sliding window. The solid black horizontal line is the significance threshold after multiple testing. Underlying data is provided in Table S3.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Genome-wide association study (GWAS). The most significant association between SNP genotype and ACC in both sexes was at s74824.1 in the sub-telomeric region of chromosome 6 ($P = 2.92 \times 10^{-10}$, Table 2). Sex-specific GWAS indicated that this SNP was highly associated with female ACC ($P = 1.07 \times 10^{-11}$), but was not associated with male ACC (P = 0.55; Table 2, Fig 4). This SNP corresponded to the same region identified in the regional heritability analysis, was the most distal typed on the chromosome from the Ovine SNP50 BeadChip (Fig 4), and had an additive effect on female ACC, with a difference of 3.37 (S.E. = 0.49) autosomal crossovers per gamete between homozygotes (Table 2). A SNP on an unmapped genomic scaffold (1.8kb, NCBI Accession: AMGL01122442.1) was also highly associated with female ACC (Fig 4). BLAST analysis indicated that the most likely genomic position of this SNP was at ~113.8Mb on chromosome 6, corresponding to the same sub-telomeric region.

Two regions on chromosome 3 were associated with ACC using the GWAS approach, although their respective regions had not shown a significant association with ACC using a regional heritability approach (see above). A single SNP on chromosome 3, OAR3_51273010.1, was associated with ACC in males, but not in females, and had an approximately dominant effect on ACC ($P = 1.15\times 10^{-6}$, Fig 4, Table 2); This SNP was 17.8kb from the 3' UTR of leucine rich repeat transmembrane neuronal 4 (*LRRTM4*) in an otherwise gene poor region of the genome (i.e. the next protein coding regions are > 1Mb from this SNP in either direction). A second SNP on chromosome 3, OAR3_87207249.1, was associated with ACC in both sexes ($P = 1.95\times 10^{-6}$, Fig 4, Table 2). This SNP was 137kb from the 5' end of an orthologue of WD repeat domain 61 (*WDR61*) and 371kb from the 5' end of an orthologue of ribosomal protein L10 (*RPL10*). No significant SNP associations were observed at the significant regional heritability region on chromosome 7. Full results of GWAS are provided in Table S4.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

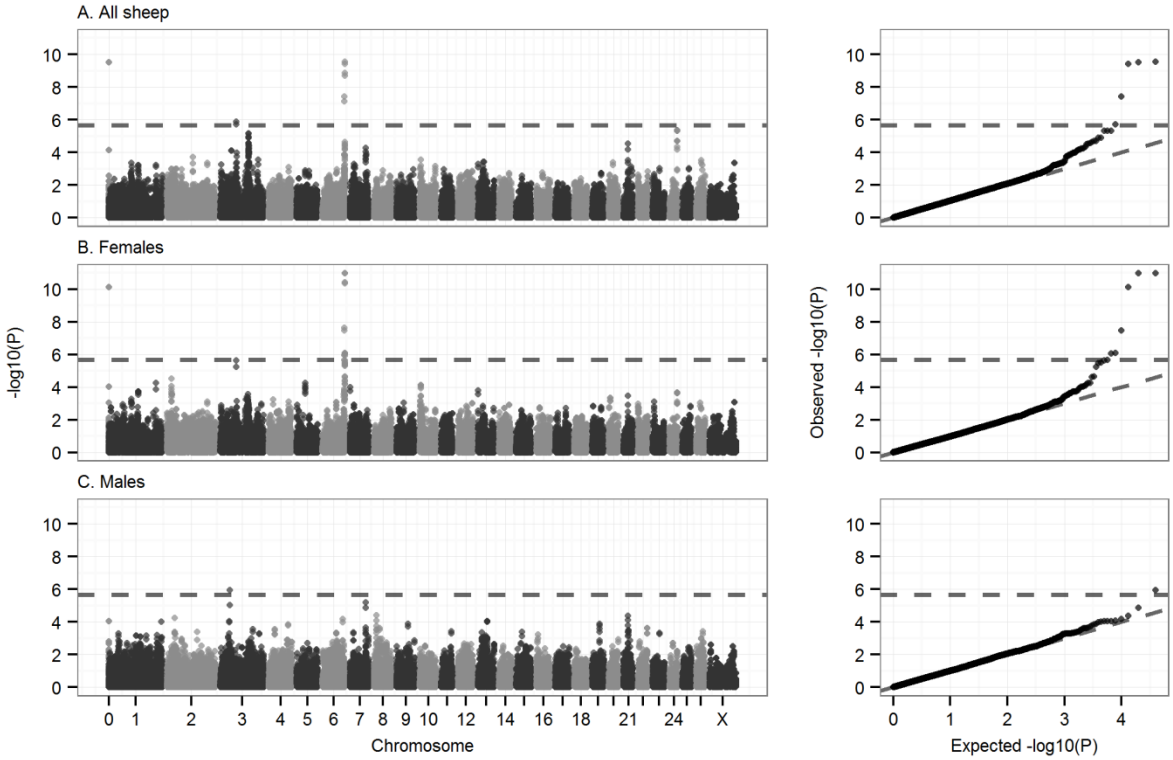

Fig 4. Genome-wide association of autosomal crossover count.

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 colour 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 as λ was less than one for all GWAS (λ = 0.996, 0.933 and 0.900 for plots A, B and C, respectively). Information on the most highly associated SNPs, their genomic positions, and the sample sizes are given in Table S4. The significant SNP in grey 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).

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 2. Top hits for genome-wide association studies of ACC in all sheep, females and males. Results provided are from Ovine SNP50 BeadChip and the most highly associated imputed SNP from chromosome 6[1]. Additional loci that were significantly associated with ACC and in strong LD with these hits are not shown; full GWAS results are provided in Table S4. A and B indicate the reference alleles. P values are given for a Wald test of an animal model with SNP genotype fitted as a fixed effect; those in bold type were genome-wide significant. V_{SNP} is the variance attributed to the SNP and Prop. V_A is the proportion of the additive genetic variance explained by the SNP. Effect AB and BB are the effect sizes of genotypes AB and BB, respectively, relative to the model intercept at genotype AA. The number of unique individuals for both sexes, females and males are approximately N = 813, 586 and 227, respectively. Numbers in brackets are standard errors.

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

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Genotype imputation and association analysis at the sub-telomeric region of chromosome 6. Genotyping of 187 sheep at a further 122 loci in the sub-telomeric region of chromosome 6 showed that this region has elevated levels of linkage disequilibrium, with the two most significant SNPs from the 50K chip tagging a haplotype block of ~374kB ($r^2 > 0.8$; see S1 Appendix, Fig 5, Table S5). Although 177kb from the primary candidate locus *RNF212*, this block contained three further candidate genes, complexin 1 (*CPLX1*), cyclin-G-associated kinase (*GAK*) and polycomb group ring finger 3 (*PCGF3*) [29]. 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, Fig 5) occurred within an intronic region of an uncharacterised protein orthologous to transmembrane emp24 protein transport domain containing *(TMED11*), 25.2kb from the putative location of *RNF212* and 13kb from the 3' end of spondin 2 (*SPON2*). A bivariate animal model including an interaction term between ACC in each sex and the genotype at oar3_OAR6_116402578 confirmed that this locus had an effect on female ACC only; this effect was additive, with a difference of 4.91 (S.E. = 0.203) autosomal crossovers per gamete between homozygotes (Fig 6, Tables 2 and S5). There was no difference in ACC between the three male genotypes. Full results for univariate models at imputed SNPs are given in Table S5.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

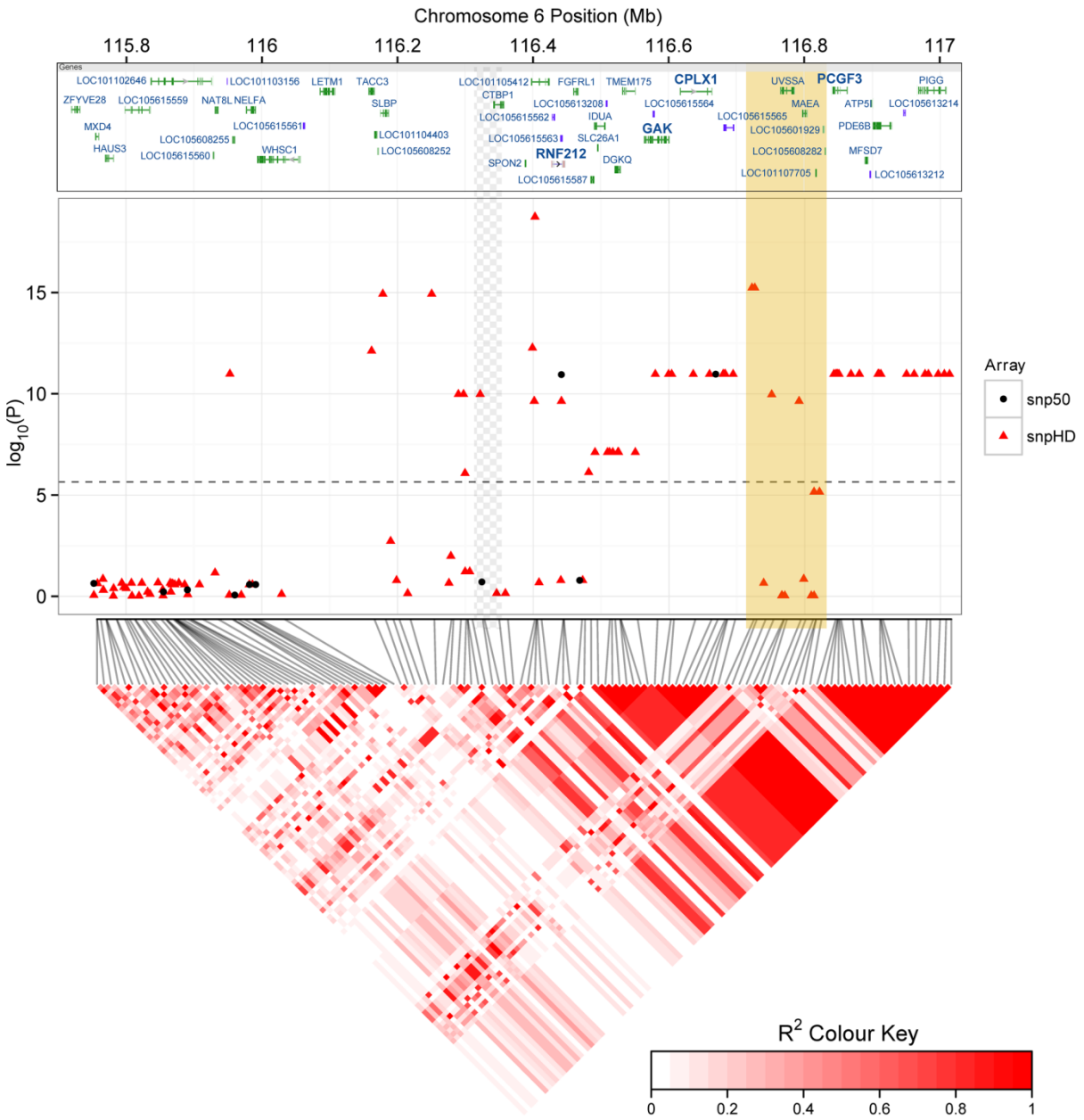

Fig 5. Associations at the sub-telomeric region of chromosome 6.
Local associations of female ACC with Ovine SNP50 BeadChip SNPs (black circles, middle panel) and imputed genotypes from the Ovine HD SNP BeadChip (red triangles). The top panel indicates protein coding regions within this region as provided by the NCBI Graphical Sequence Viewer v3.8, with genes previously implicated in recombination or meiosis given in bold text (see Introduction and [29,53]). The dashed line in the middle panel indicates the significance threshold after multiple testing. The lower panel is a heatmap of linkage disequilibrium in this region calculated for the 188 individuals originally typed on the HD SNP chip using Spearman Rank correlation r^2 created using the R library *LDheatmap* [54]. 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 S1 Appendix); its position is likely to fall within the region indicated by the grey chequered pattern to the left, leaving a large region of very high LD at the distal end of chromosome 6.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

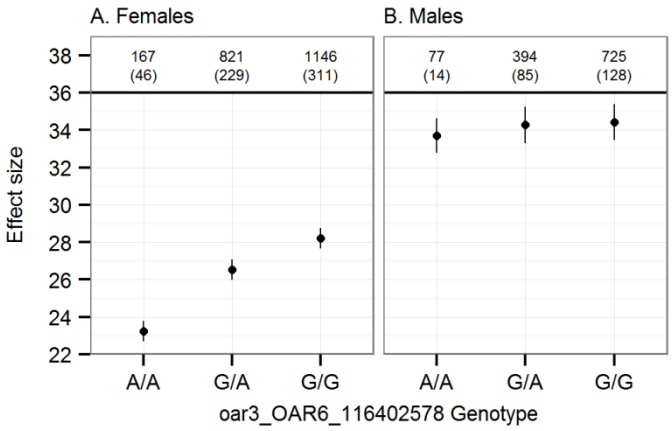

Fig 6. Effect sizes from a bivariate animal model of autosomal crossover count.
Effect sizes are shown for (A) females and (B) males from a single bivariate model including oar3_OAR6_116402578 genotype as a fixed interaction term. Error bars are the standard error around the model intercept (genotype 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.

Discussion.

We characterised autosomal crossover counts in 3,300 gametes transmitted from 813 unique individuals to their offspring, to determine the genetic architecture of recombination rate in Soay sheep. Recombination rate was heritable and had a sexually dimorphic genetic architecture, including variants at loci previously implicated in ACC in other mammal species. By using both regional heritability and GWAS approaches, we were able to identify significant loci otherwise undetected through GWAS alone. This supports the strength of using regional heritability approaches to characterise variation from multiple alleles and/or haplotypes encompassing both common and rare variants that are in linkage disequilibrium with causal loci [51]. Our findings show the benefits of using multiple association approaches, and suggest that (a) ACC can be heritable in natural populations, with the potential to respond to selection, and (b) identification of regions *a priori* associated with ACC indicates that the genetic architecture of recombination rate variation is similar across mammal species.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The strongest association was observed at *RNF212*, with the genomic region accounting for ~47% of heritable variation in female ACC; this region was not associated with male ACC. *RNF212* has been repeatedly implicated in recombination rate variation in mammals [26,52] and has also been shown to have sexually-antagonistic effects on recombination rate in humans [27,29,55]. 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 [52]. The same sub-telomeric region of chromosome 6 also contained three further candidate loci, namely *CPLX1*, *GAK* and *PCGF3* (Fig 5); these loci occurred on a ~374kb block of high LD ($r^2 > 0.8$) and were in moderate LD with the most highly associated SNP at *RNF212* ($r^2 = 0.54$). A rare intronic SNP variant within *CPLX1* has been associated with large differences in linkage map lengths in humans, independently of *RNF212* [29]; GAK forms part of a complex with cyclin-G, a locus involved in meiotic recombination repair in *Drosophila* [56]; and PCGF3 forms part of a PRC1-like complex (polycomb repressive complex 1) which is involved in meiotic gene expression and the timing of meiotic prophase in female mice [53]. High LD within this region meant that it was not possible to test the effects of these loci on ACC independently; however, the co-segregation of several loci affecting meiotic processes may merit further investigation to determine if recombination is suppressed in this region, and if this co-segregation is of adaptive significance.

Additional genomic regions associated with ACC included a 1.09Mb region of chromosome 7 affecting both sexes (identified using regional heritability analysis), and two loci at 48.1Mb and 82.4Mb on chromosome 3 (identified using GWAS) with effects on males only and both sexes, respectively. The chromosome 7 region contained two loci associated with recombination phenotypes: *REC8*, the protein of which is required for the separation of sister chromatids and homologous chromosomes during meiosis [57]; and *RNF212B*, a paralogue of *RNF212*. This region is also associated with recombination rate in cattle [26]. The chromosome 3 variants identified were novel to this study, and occurred in relatively gene poor regions of the genome. These variants

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

remained significant when repeating association analysis on ACC excluding their own chromosome (see Materials and Methods, Tables S3 & S4), meaning that they are likely to affect recombination rate globally (i.e. trans-acting effects), rather than being in LD with polymorphic recombination hotspots. After accounting for significant regions on chromosomes 6 and 7, between 36 and 61% of heritable variation in ACC was attributed to polygenic variation independent of these regions (Table S6). This "missing heritability" (i.e. heritable variation of unknown architecture) suggests that the genetic architecture of this trait is comprised of several loci of large effects on phenotype, but with a significant polygenic component i.e. variance attributed to few or many loci with small effects on phenotype in this population.

Meiotic recombination in mammals occurs during distinctly different periods in each sex: male meiosis occurs during spermatogenesis, whereas female meiosis starts in the foetal ovary, but is arrested during crossing-over between prophase and metaphase I and completed after fertilisation [40]. The between-sex genetic correlation of ACC detected in this study was not significantly different from 1, indicating that male and female ACC variation had a shared genetic basis – albeit with a relatively large error around this estimate. Further investigation through regional heritability mapping and GWAS indicated that variation in ACC has *some* degree of a shared and distinct genetic architecture between the sexes, which may be expected due to the similarities of this process, but differences in implementation. It is unlikely that the absence of associations between the *RNF212* region and male ACC is due to low power to detect the effect, as (a) this locus had a sexually dimorphic effect *a priori*, and (b) bivariate models accounting for variation in *RNF212* as a fixed effect supported a sexually dimorphic genetic effect with a lower degree of error (Fig 6).

The strong sex difference in recombination rate in this study was manifested in increased male recombination in sub-telomeric regions, allowing further crossovers to occur on the same chromatid despite crossover interference. Although variation at *RNF212* resulted in markedly different ACCs in

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

361 females, the relative positions of crossovers did not differ between females with different genotypes
362 (Fig S4), suggesting that differences in the action of *RNF212* on female ACC may be due to protein
363 function or dosage dependence, such as is observed in mouse systems [52]. The degree of
364 heterochiasmy observed in Soay sheep was similar to that of domestic sheep (*Ovis aries,*
365 Male:Female linkage map = 1.19 [42]), but not to that of bighorn sheep (*Ovis canadensis,*
366 male:female ratio = 0.89, [58], divergence ~2.8Mya). This suggests that large changes in
367 heterochiasmy can occur over relatively short evolutionary timescales; indeed, to date there is no
368 clear phylogenetic signal of heterochiasmy within mammals [35,47]. If recombination rate is indeed
369 controlled by a small number of loci with relatively large and/or sexually-dimorphic effects, then the
370 introduction of novel variants could lead to rapid changes in recombination landscape within and
371 between species. This phenomenon has been observed at *PRDM9* in humans and other species,
372 where non-synonymous mutations can lead to changes in recombination hot-spots due to
373 differential motif recognition [30,59]. Although more empirical studies on other species are
374 required, our findings add to a compelling case for a role for *RNF212* in driving heterochiasmy in
375 mammal systems.
376
377 The reasons as to why Soay sheep and domestic sheep have strong male-biased heterochiasmy are
378 not yet well understood; however, there are several factors which may favour increased
379 recombination rate in males and variation in females. Soay sheep have a highly promiscuous mating
380 system, with the largest testes to body size ratio within ruminants [60] and high levels of sperm
381 competition, with dominant rams suffering from sperm depletion towards the end of the rut [61].
382 Increased recombination may allow more rapid sperm production through formation of meiotic
383 bouquets [62]. However, when considering the large degree of variation in ACC within females, it is
384 not clear why variants with a large effect exist within the population. Examination of haplotypes
385 around *RNF212* suggest that variation has been maintained in Soay sheep throughout their long
386 history on St Kilda, and are unlikely to have been introduced by recent introgression events (S4

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Appendix, [63]). One explanation for the maintenance of variation may be that this locus is under selection due to higher investment in their offspring by females, who produce one or two lambs per year. If so, then there may be a trade-off in females between the risks of aneuploidy versus maintaining genome integrity (i.e. through balancing selection). Further empirical studies of recombination rate and heterochiasmy in this and other natural systems will be required to determine if recombination rate is under selection, and to determine the drivers of differences in recombination rate and heterochiasmy across eukaryotes.

Materials and Methods.

Study population and pedigree.

The Soay sheep is a primitive breed of domestic sheep (*Ovis aries*) which has lived unmanaged on the St Kilda archipelago (Scotland, UK, 57º49'N, 8º34'W) since the Bronze age. Sheep living within the Village Bay area of Hirta have been studied on an individual basis since 1985 [48]. 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. A Soay sheep pedigree has been constructed using 315 SNPs in low LD, and includes 5516 individuals with 4531 maternal and 4158 paternal links [64].

SNP Dataset.

A total of 5805 Soay sheep were genotyped at 51,135 single nucleotide polymorphisms (SNPs) on the Ovine SNP50 BeadChip using an Illumina Bead Array genotyping platform (Illumina Inc., San Diego, CA, USA; [49]). Quality control on SNP data was carried out using the *check.marker* function in GenABEL v 1.8-0 [65] 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

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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. The genomic inbreeding coefficient (measure $\hat{F}_{III}$ in [66], hereafter $\hat{F}$, was calculated for each sheep in the software GCTA v1.24.3 [66], using information for all SNP loci passing quality control.

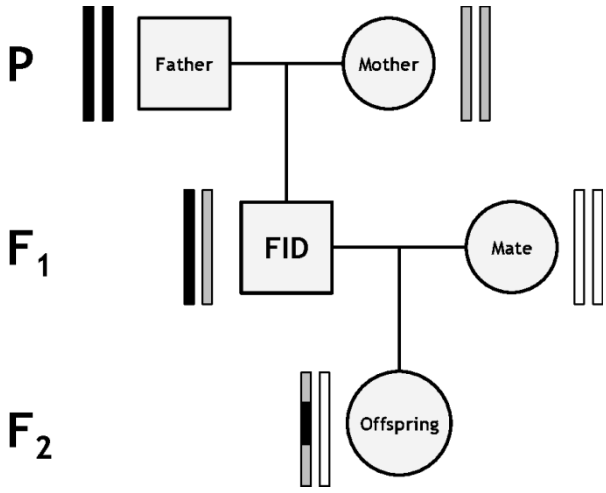

Fig 7. Diagram of the sub-pedigree structure used to infer crossover events. Rectangle pairs next to each individual represent chromatids, with black and grey shading indicating chromatids or chromatid sections of FID paternal and FID maternal origin, respectively. White shading indicates chromatids 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 grey arrow) can be distinguished at the points where origin of alleles origin flips from FID paternal to FID maternal and *vice versa*.

Estimation of meiotic autosomal crossover count (ACC).

Sub-pedigree construction. To allow unbiased phasing of the SNP data, a standardised pedigree approach was used to identify cross-overs 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 sub-pedigree 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 (Fig 7). This sub-pedigree structure allowed phasing of SNPs within the FID, and thus the identification of autosomal cross-over events in

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

435 the gamete transferred from the FID to the offspring (Fig 7). Sub-pedigrees were discarded from the
436 analysis if they included the same individual twice (e.g. father-daughter matings; N = 13).
437
438 **Linkage map construction and chromosome phasing.** All analyses in this section were conducted
439 using the software CRI-MAP v2.504a [67]. First, Mendelian incompatibilities in each sub-pedigree
440 were identified using the *prepare* function; incompatible genotypes were removed from all affected
441 individuals, and sub-pedigrees containing parent-offspring relationships with more than 0.1%
442 mismatching loci were discarded. Second, sex-specific and sex-averaged linkage map positions (in
443 Kosambi cM) were obtained using the *map* function, where SNPs were ordered relative to their
444 estimated positions on the sheep genome assembly Oar_v3.1 (Genbank assembly ID:
445 GCA_000298735.1; [68]). SNP loci with a map distance of greater than 3 cM to each adjacent marker
446 (10cM for the X chromosome, including PAR) were assumed to be incorrectly mapped and were
447 removed from the analysis, with the *map* function rerun until all map distances were below this
448 threshold; in total, 76 SNPs were assumed to be incorrectly mapped. Third, the *chrompic* function
449 was used to identify informative SNPs (i.e. those for which the grand-parent of origin of the allele
450 could be determined) on chromatids transmitted from the FID to its offspring; crossovers were
451 deemed to have occurred where there was a switch in the grandparental origin of a SNP allele (Fig
452 7).
453
454 **Quality control and crossover estimation in autosomes.** Errors in determining the grand-parental
455 origin of alleles can lead to false calling of double-crossovers (i.e. two adjacent crossovers occurring
456 on the same chromatid) and in turn, an over-estimation of recombination rate. To reduce the
457 likelihood of calling false crossover events, runs of grandparental-origin consisting of a single allele
458 (i.e. resulting in a double crossover either side of a single SNP) were recoded as missing (N = 973 out
459 of 38592 double crossovers, Fig S2). In remaining cases of double crossovers, the base pair distances
460 between immediately adjacent SNPs spanning a double crossover were calculated (hereafter, "span

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

distance"). The distribution of the span distances indicated that crossover interference is present within the Soay sheep population (Fig S2). Informative SNPs that occurred within double-crossover segments with a $\log_{10}$ span distance lower than 2.5 standard deviations from the mean $\log_{10}$ span distance (equivalent to 9.7Mb) were also recoded as missing (N = 503 out of 37619 double crossovers, Fig S2). The autosomal crossover count (ACC), the number of informative SNPs and the informative length of the genome (i.e. the total distance between the first and last informative SNPs for all chromatids) was then calculated for each FID. A simulation study was conducted to ensure that our approach accurately characterised ACC and reduced phasing errors. Autosomal meiotic crossovers were simulated given an identical pedigree structure and population allele frequencies ($N_{simulations}$ = 100; see S2 Appendix for detailed methods and results). Our approach was highly accurate in identifying the true ACC per simulation across all individuals and per individual across all simulations (adjusted $R^2 > 0.99$), but indicated that accuracy was compromised in individuals with high values of $\hat{F}$. This is likely to be an artefact of long runs of homozygosity as a result of inbreeding, which may prevent detection of double crossovers or crossovers in sub-telomeric regions. To ensure accurate individual estimates of ACC, gametes with a correlation of adjusted $R^2 \leq 0.95$ between simulated and detected crossovers in the simulation analysis were removed from the study (N = 8).

Assessing variation in the recombination landscape.

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

Fine Scale Recombination Rate. The probability of crossing-over was calculated in 1MB windows across the genome using information from the male and female linkage maps. Briefly, the probability of crossing over within a bin was the sum of all recombination fractions, *r*, in that bin; in cases where

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

an *r* value spanned a bin boundary, it was recalculated as $r \times N_{boundary}/N_{adjSNP}$, where $N_{boundary}$ was the number of bases to the bin boundary, and N_{adjSNP} was the number of bases to the closest SNP within the adjacent bin.

Variation in crossover probability relative to proximity to telomeric regions on each chromosome arm was examined using general linear models with a Gaussian error structure. The response variable was crossover-probability per bin; the fitted covariates were as follows: distance to the nearest telomere, defined as the best fit of either a linear (x), multiplicative inverse ($1/x$), quadratic ($x^2 + x$), cubic ($x^3 + x^2 + x$) or a log term ($\log_{10} x$); sex, fitted as a main effect and as an interaction term with distance to the nearest telomere; number of SNPs within the bin; and GC content of the bin (%, obtained using sequence from Oar_v3.1 [68]). The best model was identified using Akaike's Information Criterion [69]. An additional model was tested, using ratio of male to female crossover probability as the response variable, with the same fixed effect structure (omitting sex). In both models, the distance to the nearest telomere was limited to 60Mb, equivalent to half the length of the largest acrocentric chromosome (Chr 4). Initial models also included a term indicating if a centromere was present or absent on the 60Mb region, but this term was not significant in either model.

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

ACC was modelled as a trait of the FID. Phenotypic variance in ACC was partitioned using a restricted maximum likelihood (REML) Animal Model [50] implemented in ASReml-R [70] in R v3.1.1. To determine the proportion of phenotypic variance attributed to additive genetic effects (i.e. heritability), a genomic relatedness matrix at all autosomal markers was constructed for all genotyped individuals using GCTA v1.24.3 [66]; the matrix was shrunk using *--adj 0*. Trait variance of was analysed first with the following univariate model:

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

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

To investigate if 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. 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. Models were then compared using likelihood ratio tests with 1 degree of freedom.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Genetic architecture of autosomal crossover count.

Genome partitioning of genetic variance (regional heritability analysis). The contribution of specific genomic regions to trait variation was determined by partitioning the additive genetic variance as follows [51]:

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

where *v* is the vector of additive genetic effects explained by an autosomal genomic region *i*, and *nv* is the vector of the additive genetic effects explained by all remaining autosomal markers outwith region *i*. Regional heritabilities were determined by constructing genomic relatedness matrices (GRMs) for regions of *i* of increasing resolution (whole chromosome partitioning, sliding windows of 150, 50 and 20 SNPs, corresponding to regions of 9.4, 3.1 and 1.2Mb mean length, respectively) and fitting them in models with an additional GRM of all autosomal markers not present in region i; sliding windows overlapped by half of their length (i.e. 75, 25 and 10 SNPs, respectively). GRMs were constructed in the software GCTA v1.24.3 and were shrunk using the *--adj 0* argument [66]. The significance of additive genetic variance attributed to a genomic region i was tested by comparing models with and without the $Z_1 v_i$ term using a likelihood ratio test; in cases where the heritability estimate was zero (i.e. estimated as "Boundary" by ASReml), significant model comparison tests were disregarded. A Bonferroni approach was used to account for multiple testing across the genome, by taking the number of tests and dividing by two to account for the overlap of the sliding windows (since each genomic region was modelled twice).

Genome-wide association study of variants controlling ACC. Genome-wide association studies (GWAS) of autosomal recombination rates under different scenarios were conducted using ASReml-R [70] in R v3.1.1, fitting individual animal models for each SNP locus using the same model structure as above. SNP genotypes were fitted as a fixed effect with two or three levels. The GRM was replaced with a relatedness matrix based on pedigree information to speed up computation; the

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

pedigree and genomics matrices have been shown to be highly correlated [64]. Sex-specific models were also run. Association statistics were corrected for any population stratification not captured by the animal model by dividing them by the genomic control parameter, λ [71], when λ > 1, which was calculated as the median Wald test χ_2^2 divided by the median χ_2^2, expected from a null distribution. The significance threshold after multiple testing was determined using a linkage disequilibrium-based method outlined in [72] using a sliding window of 50 SNPs; the effective number of tests in the GWAS analysis was 22273.61, meaning the significance threshold for P after multiple testing at α = 0.05 was 2.245×10^{-6}. Although sex chromosome recombination rate was not included in the analysis, all GWAS included the X chromosome and SNP markers of unknown position (N=314). The proportion of phenotypic variance attributed to a given SNP was calculated using the using the following equation [73]:

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

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; [74]). The position of a strong candidate locus, *RNF212* is not annotated on Oar_v3.1, but sequence alignment indicated that it is positioned at the sub-telomere of chromosome 6 (see S3 Appendix).

Accounting for cis- and trans- genetic variants affecting recombination rate. In the above analyses, we wished to separate potential associations with ACC due to cis-effects (i.e. genetic variants that are in linkage disequilibrium with polymorphic recombination hotspots) from those due to trans-effects (i.e. genetic variants in LD with genetic variants that affect recombination rate globally). By

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 to 26. We found that 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. All results presented above are for ACC, but results from both types of analysis are available in Tables S3 and S4.

Linkage disequilibrium 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 maximise $\sum_{i=1}^{m} p_i$ using the equation $p_m = A_m^{-1} c_m$, where p is a vector of the proportion of genetic variation in the population captured by m selected animals, A_m is the corresponding subset of a pedigree relationship matrix and c is a vector of the mean relationship of the m selected animals (as outlined in [75] & [76]). This approach should capture the maximum amount of genetic variation within the main population for the number of individuals in the reference population. SNP loci were retained if call rate was > 0.95 and MAF > 0.01 and individuals were retained if more than 95% of loci were genotyped. Linkage disequilibrium (LD) between loci was calculated using Spearman-Rank correlations (r^2) in the 188 individuals passing quality control.

Genotypes from the HD SNP chip were imputed to individuals typed on the SNP50 chip in the chromosome 6 region significantly associated with ACC, using pedigree information in the software MaCH v1.0.16 [77]. This region contained 10 SNPs from the Ovine SNP50 BeadChip and 116 additional independent SNPs on the HD SNP chip. As the software requires both parents to be

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 mis-matching genotypes within individuals; therefore, individuals with < 0.99 imputed genotypes scored were removed from the analysis. Loci with < 0.95 of individuals typed were also discarded. Imputation accuracy was calculated for all loci as the proportion of imputed genotypes matching their true genotypes; all remaining loci had imputation accuracies > 0.95.

Acknowledgements

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, Camillo Bérénos, Jisca Huisman and John Hickey greatly improved the analysis. James Kijas, Yu Jiang and Brian Dalrymple provided information for numerous queries on the genome assembly, annotation and SNP genotyping. Philip Ellis and Camillo Bérénos 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.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Referenced Literature

1. Hassold T, Hunt P. To err (meiotically) is human: the genesis of human aneuploidy. Nat Rev Genet. 2001;2: 280–291. doi:10.1038/35066065

2. Fledel-Alon A, Wilson DJ, Broman K, Wen X, Ober C, Coop G, et al. Broad-scale recombination patterns underlying proper disjunction in humans. PLoS Genet. 2009;5: e1000658. doi:10.1371/journal.pgen.1000658

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

4. Crow JF, Kimura M. Evolution in Sexual and Asexual Populations. Am Nat. 1965;99: 439–450.

5. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966;8: 269–294. doi:10.1017/S001667230800949X

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

7. Charlesworth B, Barton NH. Recombination load associated with selection for increased recombination. Genet Res Cambr. 1996;67: 27–41. doi:10.1017/S0016672300033450

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

9. Comeron JM, Williford A, Kliman RM. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity (Edinb). 2008;100: 19–31. doi:10.1038/sj.hdy.6801059

10. McPhee CP, Robertson A. The effect of suppressing crossing-over on the response to selection in Drosophila melanogaster. Genet Res. 1970;16: 1–16. doi:10.1017/S0016672300002238

11. Rice WR, Chippindale AK. Sexual recombination and the power of natural selection. Science. 2001;294: 555–559. doi:10.1126/science.1061380

12. Inoue K, Lupski JR. Molecular mechanisms for genomic disorders. Ann Rev Genomics Hum Genet. 2002; 199–242. doi:10.1146/annurev.genom.3.032802.120023

13. Barton NH. Why sex and recombination? Science (80-). 1998;281: 1986–1990. doi:10.1101/sqb.2009.74.030

14. Otto SP, Lenormand T. Resolving the paradox of sex and recombination. Nat Rev Genet. 2002;3: 252–261. doi:10.1038/nrg761

15. Mészáros G, Gorjanc G, Jenko J, Cleveland MA, Woolliams JA, Hickey JM. Selection on Recombination Rate to Increase Genetic Gain. Proceedings, 10th World Congr Genet Appl to Livest Prod. 2014;

16. Coop G, Przeworski M. An evolutionary view of human recombination. Nat Rev Genet. 2007;8: 23–34. doi:10.1038/nrg1947

17. Dumont BL, Payseur BA. Evolution of the genomic rate of recombination in mammals. Evolution. 2008;62: 276–94. doi:10.1111/j.1558-5646.2007.00278.x

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

18. Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, et al. Fine-scale recombination rate differences between sexes, populations and individuals. Nature. Nature Publishing Group; 2010;467: 1099–103. doi:10.1038/nature09525

19. Pardo-Manuel de Villena F, Sapienza C. Recombination is proportional to the number of chromosome arms in mammals. Mamm Genome. 2001;12: 318–322. doi:10.1007/s003350020005

20. Hillers KJ. Crossover interference. Curr Biol. 2004;14: R1036–R1037. doi:10.1016/j.cub.2004.11.038

21. Hultén MA. On the origin of crossover interference: A chromosome oscillatory movement (COM) model. Mol Cytogenet. 2011;4: 10. doi:10.1186/1755-8166-4-10

22. Fullerton SM, Bernardo Carvalho A, Clark AG. Local rates of recombination are positively correlated with GC content in the human genome. Mol Biol Evol. 2001;18: 1139–1142. doi:10.1093/oxfordjournals.molbev.a003886

23. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, et al. A high-resolution recombination map of the human genome. Nat Genet. 2002;31: 241–247. doi:10.1038/ng917

24. Kong A, Barnard J, Gudbjartsson DF, Thorleifsson G, Jonsdottir G, Sigurdardottir S, et al. Recombination rate and reproductive success in humans. Nat Genet. 2004;36: 1203–1206. doi:10.1038/ng1445

25. Dumont BL, Broman KW, Payseur BA. Variation in genomic recombination rates among heterogeneous stock mice. Genetics. 2009;182: 1345–9. doi:10.1534/genetics.109.105114

26. Sandor C, Li W, Coppieters W, Druet T, Charlier C, Georges M. Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle. PLoS Genet. 2012;8: e1002854. doi:10.1371/journal.pgen.1002854

27. Kong A, Thorleifsson G, Stefansson H, Masson G, Helgason A, Gudbjartsson DF, et al. Sequence variants in the RNF212 gene associate with genome-wide recombination rate. Science. 2008;319: 1398–1401. doi:10.1126/science.1152422

28. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, et al. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science. 2010;327: 836–840. doi:10.1126/science.1183439

29. Kong A, Thorleifsson G, Frigge ML, Masson G, Gudbjartsson DF, Villemoes R, et al. Common and low-frequency variants associated with genome-wide recombination rate. Nat Genet. 2014;46: 11–6. doi:10.1038/ng.2833

30. Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, et al. The landscape of recombination in African Americans. Nature. 2011;476: 170–175. doi:10.1038/nature10336

31. Stevison LS, Woerner AE, Kidd JM, Kelley JL, Veeramah KR, McManus KF, et al. The time-scale of recombination rate evolution in Great Apes. BioRxiv. 2015;Preprint. . doi:http://dx.doi.org/10.1101/013755

32. Haldane J. Sex ratio and unisexual sterility in hybrid animals. J Genet. 1922;12.

33. Huxley J. Sexual difference in linkage in Gammarus chevreuxi. J Genet. 1928;20: 145–156.

34. Nei M. Linkage Modification and Sex Difference in Recombination. Genetics. 1969;63: 681–699.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

35. Lenormand T, Dutheil J. Recombination difference between sexes: a role for haploid selection. PLoS Biol. 2005;3: e63. doi:10.1371/journal.pbio.0030063

36. Tease C, Hultén M a. Inter-sex variation in synaptonemal complex lengths largely determine the different recombination rates in male and female germ cells. Cytogenet Genome Res. 2004;107: 208–215. doi:10.1159/000080599

37. De Boer E, Stam P, Dietrich AJJ, Pastink A, Heyting C. Two levels of interference in mouse meiotic recombination. P Natl Acad Sci USA. 2006;103: 9607–9612. doi:10.1073/pnas.0600418103

38. Petkov PM, Broman KW, Szatkiewicz JP, Paigen K. Crossover interference underlies sex differences in recombination rates. Trends Genet. 2007;23: 539–542.

39. Koehler KE, Hawley RS, Sherman S, Hassold T. Recombination and nondisjunction in humans and flies. Hum Mol Genet. 1996;5 Spec No: 1495–1504. doi:10.1093/hmg/5.Supplement_1.1495

40. Morelli MA, Cohen PE. Not all germ cells are created equal: Aspects of sexual dimorphism in mammalian meiosis. Reproduction. 2005;130: 761–781. doi:10.1530/rep.1.00865

41. Nagaoka SI, Hassold TJ, Hunt PA. Human aneuploidy: mechanisms and new insights into an age-old problem. Nat Rev Genet. 2012;13: 493–504. doi:10.1038/nrg3245

42. Maddox JF, Davies KP, Crawford AM, Hulme DJ, Vaiman D, Cribiu EP, et al. An Enhanced Linkage Map of the Sheep Genome Comprising More Than 1000 Loci. Genome Res. 2001;11: 1275–1289. doi:10.1101/gr.GR-1350R

43. Burt A, Bell G, Harvey PH. Sex differences in recombination. J Evol Biol. 1991;4: 259–277. doi:10.1046/j.1420-9101.1991.4020259.x

44. Zenger KR, McKenzie LM, Cooper DW. The first comprehensive genetic linkage map of a marsupial: The tammar wallaby (Macropus eugenii). Genetics. 2002;162: 321–330.

45. Brandvain Y, Coop G. Scrambling eggs: meiotic drive and the evolution of female recombination rates. Genetics. 2012;190: 709–23. doi:10.1534/genetics.111.136721

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

47. Mank JE. The evolution of heterochiasmy: the role of sexual selection and sperm competition in determining sex-specific recombination rates in eutherian mammals. Genet Res (Camb). 2009;91: 355–363. doi:10.1017/S0016672309990255

48. Clutton-Brock T, Pemberton J, Coulson TN, Stevenson IR, MacColl ADC. The sheep of St Kilda. Soay Sheep: Dynamics and Selection in an Island Population2. 2004. pp. 17–51.

49. Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, et al. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One. 2009;4: e4668. doi:10.1371/journal.pone.0004668

50. Henderson CR. Best linear unbiased estimation and prediction under a selection model. Biometrics. International Biometric Society; 1975;31: 423–447. Available: http://www.jstor.org/stable/2529430

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

51. Nagamine Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Rudan I, et al. Localising Loci underlying Complex Trait Variation Using Regional Genomic Relationship Mapping. PLoS One. 2012;7. doi:10.1371/journal.pone.0046501

52. Reynolds A, Qiao H, Yang Y, Chen JK, Jackson N, Biswas K, et al. RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. Nat Genet. 2013;45: 269–78. doi:10.1038/ng.2541

53. Yokobayashi S, Liang CY, Kohler H, Nestorov P, Liu Z, Vidal M, et al. PRC1 coordinates timing of sexual differentiation of female primordial germ cells. Nature. 2013;495: 236–240. doi:10.1016/j.juro.2013.07.027

54. Shin J, Blay S, Brad McNeney. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. J Stat Softw. 2006;16: Code Snippet 3. Available: ftp://131.211.84.186/mirror/CRAN/web/packages/LDheatmap/vignettes/LDheatmap.pdf

55. Chowdhury R, Bois PRJ, Feingold E, Sherman SL, Cheung VG. Genetic analysis of variation in human meiotic recombination. PLoS Genet. 2009;5. doi:10.1371/journal.pgen.1000648

56. Nagel AC, Fischer P, Szawinski J, Rosa MKL, Preiss A. Cyclin G is involved in meiotic recombination repair in Drosophila melanogaster. J Cell Sci. 2012; doi:10.1242/jcs.113902

57. Parisi S, McKay MJ, Molnar M, Thompson MA, van der Spek PJ, van Drunen-Schoenmaker E, et al. Rec8p, a meiotic recombination and sister chromatid cohesion phosphoprotein of the Rad21p family conserved from fission yeast to humans. Mol Cell Biol. 1999;19: 3515–3528.

58. Poissant J, Hogg JT, Davis CS, Miller JM, Maddox JF, Coltman DW. Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorphism in bighorn sheep. BMC Genomics. 2010;11: 524. doi:10.1186/1471-2164-11-524

59. Ponting CP. What are the genomic drivers of the rapid evolution of PRDM9? Trends Genet. Elsevier Ltd; 2011;27: 165–71. doi:10.1016/j.tig.2011.02.001

60. Stevenson IR, Marrow B, Preston BT, Pemberton JM, Wilson K. Adaptive reproductive strategies. In: Clutton-Brock TH, Pemberton JM, editors. Soay Sheep: Dynamics and Selection in an Island Population. 1st ed. Cambridge, U.K.: Cambridge Univ. Press; 2004. pp. 243–275.

61. Preston BT, Stevenson IR, Pemberton JM, Wilson K. Dominant rams lose out by sperm depletion. Nature. 2001;409: 681–682. doi:10.1038/35055617

62. Tankimanova M, Hultén MA, Tease C. 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. 2004;105: 172–181. doi:10.1159/000078187

63. Feulner PGD, Gratten J, Kijas JW, Visscher PM, Pemberton JM, Slate J. Introgression and the fate of domesticated genes in a wild mammal population. Mol Ecol. 2013;22: 4210–4221. doi:10.1111/mec.12378

64. Bérénos C, Ellis P, Pilkington JG, Pemberton JM. Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. Mol Ecol. 2014;23: 3434–3451. doi:10.1111/mec.12827

65. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007;23: 1294–1296. doi:10.1093/bioinformatics/btm108

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

66. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. The American Society of Human Genetics; 2011;88: 76–82. doi:10.1016/j.ajhg.2010.11.011

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

68. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. Science (80-). 2014;344: 1168–1173. doi:10.1126/science.1252806

69. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr. 1974;19: 716–723. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1100705

70. Butler DG, Cullis BR, Gilmour AR, Gogel BJ. Mixed Models for S language Environments: ASReml-R reference manual. 2009.

71. Devlin AB, Roeder K, Devlin B. Genomic Control for Association. Biometrics. 1999;55: 997–1004.

72. Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies. Genet Epidemiol. 2008;32: 567–573. Available: http://dx.doi.org/10.1002/gepi.20331

73. Falconer DS, Mackay TFC. Variance. Introduction to Quantitative Genetics. 1996. pp. 122–144.

74. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. Nucleic Acids Res. 2014;43: D662–D669. doi:10.1093/nar/gku1010

75. Pausch H, Aigner B, Emmerling R, Edel C, Götz K-U, Fries R. Imputation of high-density genotypes in the Fleckvieh cattle population. Genet Sel Evol. 2013;45: 3. doi:10.1186/1297-9686-45-3

76. Goddard ME, Hayes BJ. Genomic Selection Based on Dense Genotypes Inferred From Sparse Genotypes. Proc Assoc Advmt Anim Breed Genet. 2009;18: 26–29.

77. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010;34: 816–834. doi:10.1002/gepi.20533

78. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 2012;10: e1001258. doi:10.1371/journal.pbio.1001258

Supplementary Information

S1 Appendix. BLAST comparison of the sub-telomeric region of chromosome 6 with the homologous region of the cattle genome.

S2 Appendix. Simulation of recombination events.

S3 Appendix. Predicted location of *RNF212* on the Sheep genome Oar_v3.1

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

S4 Appendix. Haplotype sharing of the sub-telomeric region of chromosome 6 with domesticated breeds.

Fig S1. 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 [42].

Fig S2. Histogram of the span distances (in Mb) between double crossovers on autosomal chromatids. Bar segments are colour 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).

Fig S3. 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 S7.

Fig S4. 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 S8.

Fig S5. 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.

Fig S6. 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 colour indicates the percentage sequence match for each segment of aligned sequence.

Fig S7. 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/)

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fig S8. Mean and standard deviation of haplotype sharing (HS in Mb) between Soay sheep and 73 domestic sheep breeds for core haplotypes outlined in Table S11, for breeds listed in Table S10. Red and blue colouration indicates haplotypes tagging the high (G) and low (A) recombination alleles at oar3_OAR6_116402578, respectively.

Table S1. Sex-averaged and sex specific linkage map positions (cM) and genomic positions (bp) for SNP loci passing quality control.

Table S2. 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.

Table S3. Full results of the regional heritability analysis of ACC for sliding window sizes of 150, 50 and 20s across all autosomes. Models were run in both sexes, 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 χ^2 and P values are from likelihood ratio tests.

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.

bioRxiv preprint doi: https://doi.org/10.1101/024869; this version posted August 17, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S5. 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 S4. 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.

Table S6. 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.)

Table S7. 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.

Table S8. Results of the BLAST analysis outlined in S1 Appendix. 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.

Table S9. General linear model of the individual correlations and slopes of detected vs. simulated ACC per individual over 100 simulations (Table in S2 Appendix).

Table S10. Codes and sample sizes for domestic sheep breeds utilised in the haplotype sharing analysis. Information obtained from the Ovine HapMap project [78]. (Table in S4 Appendix).

Table S11. 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. (Table in S4 Appendix).