1 Variants at *RNF212* and *RNF212B* are associated with 2 recombination rate variation in Soay sheep (*Ovis aries*).

3 Susan E. Johnston^{1,*}, Martin A. Stoffel¹, Josephine M. Pemberton¹

4 ¹Institute of Evolutionary Biology, School of Biological Sciences,
5 University of Edinburgh, Edinburgh, EH9 3FL, United Kingdom

6 *Corresponding author: Susan.Johnston@ed.ac.uk

7 Abstract

8 Meiotic recombination is a ubiquitous feature of sexual reproduction, ensuring proper dis-
9 junction of homologous chromosomes, and creating new combinations of alleles upon which
10 selection can act. By identifying the genetic drivers of recombination rate variation, we can
11 begin to understand its evolution. Here, we revisit an analysis investigating the genetic archi-
12 tecture of gamete autosomal crossover counts (ACC) in a wild population of Soay sheep (*Ovis*
13 *aries*) using a much larger dataset (increasing from 3,300 to 7,235 gametes and from ~39,000 to
14 ~415,000 SNPs for genome-wide association analysis). Animal models fitting genomic relat-
15 edness confirmed that ACC was heritable in both females ($h^2 = 0.18$) and males ($h^2 = 0.12$).
16 Genome-wide association studies identified two regions associated with ACC variation. A
17 region on chromosome 6 containing *RNF212* explained 46% of heritable variation in female
18 ACC, but was not associated with male ACC, confirming the previous finding. A region
19 on chromosome 7 containing *RNF212B* explained 20-25% of variation in ACC in both males
20 and females. Both *RNF212* and *RNF212B* have been repeatedly associated with recombination
21 rate in other mammal species. These findings confirm that moderate to large effect loci can
22 underpin ACC variation in wild mammals, and provide a foundation for further studies on
23 the evolution of recombination rates.

24 Introduction

25 Meiotic recombination is ubiquitous in sexually-reproducing organisms and is an important
 26 driver of genomic diversity in eukaryotes (BARTON and CHARLESWORTH, 1998; OTTO and
 27 LENORMAND, 2002). The process of crossing-over ensures the proper disjunction of homol-
 28 ogous chromosomes during meiosis (HASSOLD and HUNT, 2001), with most species having
 29 a minimum requirement of at least one obligate crossover per chromosome pair (OTTO and
 30 PAYSEUR, 2019). It also has the advantage of generating novel haplotypes and preventing
 31 the accumulation of harmful mutations in the genome, allowing populations to respond to
 32 selection faster and to a greater degree (MULLER, 1964; HILL and ROBERTSON, 1966; BATT-
 33 GIN *et al.*, 2016). However, this can come at the cost of increased mutations at crossover
 34 sites (HALLDORSSON *et al.*, 2019) and the break up of favourable haplotypes (BARTON and
 35 CHARLESWORTH, 1998). Variation in recombination rate is likely to arise across taxa with
 36 variation in genome size and chromosome number (STAPLEY *et al.*, 2017); in addition, the
 37 evolutionary costs and benefits of recombination may vary depending on the selective con-
 38 text, again leading to the expectation of recombination rate variation within and between
 39 populations (OTTO and BARTON, 2001; OTTO and LENORMAND, 2002). Indeed, cytogenetic
 40 and linkage mapping studies have shown that recombination rates can vary by orders of
 41 magnitude within and between chromosomes (MYERS *et al.*, 2005), individuals (KONG *et al.*,
 42 2004), populations (SAMUK *et al.*, 2020), sexes (LENORMAND and DUTHEIL, 2005) and species
 43 (STAPLEY *et al.*, 2017).

44 In mammals, genomic studies in different species (i.e. humans, cattle, sheep, pigs and deer)
 45 have shown that recombination rate is often heritable, with a significant proportion of the
 46 additive genetic variation underpinned by loci of moderate to large effects (SANDOR *et al.*,
 47 2012; KONG *et al.*, 2014; MA *et al.*, 2015; KADRI *et al.*, 2016; PETIT *et al.*, 2017; JOHNSTON *et al.*,
 48 2018; JOHNSON *et al.*, 2020). Genome-wide association studies (GWAS) have shown repeated
 49 associations with variants at ring finger protein 212 *RNF212* and its paralogue *RNF212B*,
 50 whose associated proteins are essential for the crossover designation process during meiosis
 51 (REYNOLDS *et al.*, 2013), as well as other loci with established roles in meiotic processes (e.g.

52 *HEI10*, *REC8* and *TOP2*; SANDOR *et al.* 2012; KONG *et al.* 2014; MA *et al.* 2015; KADRI *et al.*
 53 2016; PETIT *et al.* 2017; JOHNSTON *et al.* 2018; JOHNSON *et al.* 2020). This heritable, oligogenic
 54 architecture implies that recombination rate has the potential to evolve rapidly to selection
 55 if variation in recombination rate is associated with individual fitness (OTTO and BARTON,
 56 2001; STAPLEY *et al.*, 2017). Therefore, detailed examination of the genetic architecture of
 57 recombination rate is a critical step in understanding how rates are evolving over different
 58 evolutionary timescales.

59 The Soay sheep (*Ovis aries*) are a Neolithic breed of domestic sheep that have been living
 60 wild on the St Kilda archipelago since the Bronze age. These sheep are a model system for
 61 studies of ecology and evolution, and have been intensively studied since 1985 (CLUTTON-
 62 BROCK *et al.*, 2004). A previous study integrated pedigree information with genotypic data
 63 from ~39,000 SNPs to characterise autosomal crossover counts (ACCs) in ~3,300 gametes
 64 transmitted from parents to their offspring (JOHNSTON *et al.*, 2016). A GWAS showed that a
 65 single region on chromosome 6, corresponding to the locus *RNF212*, explained around 47%
 66 of heritable variation in female ACC only. A broader-scale “regional heritability” analysis
 67 identified a tentative association in both sexes combined at a 1.09 Mb region on chromosome
 68 7, containing the candidate loci *REC8* and *RNF212B*, but was not significant in sex-specific
 69 regional heritability models, nor the GWAS analyses. Since then, addition of more SNP geno-
 70 typed individuals and improvements in pedigree construction methodology (HUISMAN, 2017)
 71 now allows us to characterise recombination rate in a further ~3,900 gametes. Furthermore,
 72 advances in genotype imputation methods in complex pedigrees has allowed us to increase the
 73 SNP dataset for conducting genome-wide association study from ~39,000 to ~417,000 SNPs.

74 Here, we revisit our previous study to carry out GWAS of male and female recombination
 75 rate. Our motivation was to: (a) confirm a sex-limited association with *RNF212* on chromo-
 76 some 6; (b) fine-map the association on chromosome 7 and determine the action and effect
 77 size of this locus in both males and females; and (c) to identify any further associations that
 78 could not be detected using the previous dataset.

79 Materials and Methods

80 Study Population.

81 Soay sheep living within the Village Bay area on the island of Hirta, St Kilda, Scotland
 82 (57°49'N, 8°34'W) have been individually studied since 1985 (CLUTTON-BROCK *et al.*, 2004).
 83 All sheep are ear-tagged at their first capture, including 95% of lambs born within the study
 84 area, and the majority of animals are recaptured on an annual basis. DNA samples are rou-
 85 tinely obtained from ear punches and/or blood sampling. All animal work is carried out
 86 according to UK Home Office procedures and was licensed under the UK Animals (Scientific
 87 Procedures) Act 1986 (License no. PPL60/4211).

88 Genotype and Pedigree Data.

89 All sheep were genotyped at 51,135 SNP markers on the Illumina Ovine SNP50 BeadChip
 90 (KIJAS *et al.*, 2009). Quality control was carried out using the check.marker function in
 91 GenABEL v1.8-0 with the following thresholds: SNP minor allele frequency (MAF) > 0.001,
 92 SNP locus genotyping success > 0.99, individual sheep genotyping success > 0.95; a total
 93 of 7,700 sheep and 39,398 SNPs remained. Pedigree relationships were previously inferred
 94 using data from 438 SNP loci in the R package Sequoia v1.02 (HUISMAN, 2017) and from field
 95 observations between mothers and their offspring born within the study area. Autosomal
 96 SNP genotypes were used to calculate the $\hat{F}_{II}$ genomic inbreeding coefficients using GCTA
 97 v1.90.2 (YANG *et al.*, 2011).

98 A further 189 individuals were genotyped at 430,702 polymorphic SNPs on the Ovine In-
 99 finium HD SNP BeadChip for imputation of genotypes into individuals typed on the 50K
 100 chip (see JOHNSTON *et al.* 2016 for method and individual selection criteria and STOFFEL *et al.*
 101 2020 for full imputation method for this dataset). Briefly, SNP genotypes from the HD Chip
 102 were imputed into the SNP50 Chip typed individuals using AlphaImpute v1.98 (HICKEY *et al.*,
 103 2012; ANTOLÍN *et al.*, 2017) resulting in a dataset with 7,691 individuals genotyped at 417,373

104 SNPs, with a mean genotyping rate per individual of 99.5% (range 94.8% - 100%). As there
105 were no X-chromosome SNPs in common between the two SNP Chips, only the SNP50
106 Chip SNPs were used for association analyses on the X chromosome. SNP positions for both
107 chips were known relative to sheep genome assembly Oar_v3.1 (GenBank assembly ID
108 GCA_000298735.1).

109 Estimation of autosomal crossover counts and linkage maps.

110 Autosomal crossover positions were estimated using an identical protocol to that outlined in
111 JOHNSTON *et al.* (2016). Briefly, the method uses Ovine SNP50 BeadChip SNP data from a
112 focal individual's parents, mate and offspring to characterise crossovers that occurred in the
113 gametes transmitted from the focal individual to its offspring. Crossovers were determined
114 using the software CRI-MAP v2.504a (GREEN *et al.*, 1990). Here, we estimated ACC for a
115 further 3,908 gametes, leading to a dataset of ACC for 7,235 gametes transmitted from 1,632
116 unique focal individuals. Because of differences in the SNP quality control between analyses,
117 we used 37,853 SNPs in common with the previous recombination rate study. Simulation
118 studies previously showed that this method will identify > 99% of crossovers, meaning that
119 ACC estimated using this method can be used as a proxy for individual recombination rate.
120 The method also provides linkage maps for each chromosome, allowing us to update the
121 existing map for Soay sheep to include information from a much larger number of meioses.

122 Animal Models

123 A restricted maximum likelihood (REML) animal model approach (HENDERSON, 1975) was
124 used to estimate components of phenotypic variance for ACC, including the additive genetic
125 effect (i.e. heritability). A genomic relatedness matrix (GRM) was constructed for all poly-
126 morphic autosomal SNPs from the 50K chip using GCTA v1.90.2 (YANG *et al.*, 2011) and
127 was adjusted with the argument `--grm-adj 0`, which assumes a similar frequency spectra of
128 genotyped and causal loci. Related individuals were not removed from the GRM as there is
129 high relatedness in the study population, and there are were no significant parental effects

130 or common environment effects on individual recombination rate (JOHNSTON *et al.*, 2016).
 131 Animal models were constructed for both sexes combined, females only and males only, and
 132 were run in ASReml-R v4 (BUTLER *et al.*, 2009) in R v3.6.2. Fixed effects included sex (in
 133 the combined model) and the genomic inbreeding coefficient $\hat{F}_{III}$; initial models fit age in
 134 years fit as a linear covariate, but as it was not significant, it was removed from the final mod-
 135 els (Figure 1). The significance of fixed effects was tested using Wald tests. Random effects
 136 included the additive genetic effect (as estimated using the GRM) and identity of the focal
 137 individual (i.e. “permanent environment”) effect to account for repeated measures. Despite
 138 not being significant, the fixed effect $\hat{F}_{III}$ and the random permanent environment effect were
 139 retained in all models to account for potential underestimation of ACC due to long runs of
 140 homozygosity and pseudoreplication, respectively. Significance of the additive genetic effect
 141 was determined by dropping the additive genetic effect from each model and comparing with
 142 the full model using a likelihood ratio test distributed as χ^2_1 . Bivariate models examining
 143 the genetic correlation r_A between male and female ACC were run using the CORGH error
 144 structure in ASReml-R (correlation with heterogeneous variances). Models were set with r_A
 145 to be unconstrained. To test whether r_A was significantly different from 0 and 1, the model
 146 was compared to models with r_A fixed at a value of 0 or 0.999 using likelihood ratio tests.

147 **Genome-wide association studies.**

148 Genome-wide association studies of ACC with the imputed SNP dataset were carried out us-
 149 ing function *rGLS* in the package RepeatABEL v1.1.31 (RÖNNEGÅRD *et al.*, 2016) in R v3.6.2.
 150 This approach fits both repeated measures and the GRM, the latter accounting for population
 151 structure. Association statistics were corrected for any further inflation by dividing by the
 152 genomic control parameter λ , calculated as the median observed χ^2_1 statistic divided by the
 153 median χ^2_1 expected under a null distribution. The significance threshold at $\alpha = 0.05$ was
 154 calculated by STOFFEL *et al.* (2020) as $P < 1.28 \times 10^{-06}$, using a ‘simpleM’ approach using
 155 linkage disequilibrium information to determine the effective number of independent tests
 156 (GAO *et al.*, 2008).

For the most highly-associated SNPs, genotype effect sizes were estimated by fitting the SNP genotype as a three level factor in the animal models described above. Then, in separate models, the proportion of phenotypic variance explained by significant regions was determined using a regional heritability approach (NAGAMINE *et al.*, 2012). GRMs were constructed as above for the 20 SNPs from the SNP50 chip spanning the highest associated SNP in that region (i.e. 10 SNPs from either side of the association). In one case where the association was at the end of the chromosome, the last 20 SNPs on that chromosome were used to construct the GRM. These regional GRMs were then fit as additional random effects in the animal models described above, with their significance determined using likelihood ratio tests distributed as χ^2_1 .

In order to identify potential candidate genes for ACC, sheep gene IDs SNPs in the broad associated regions were extracted from Ensembl (gene build ID Oar_v3.1.100) using the function *getBM* in the R package biomaRt v2.42.1 (DURINCK *et al.*, 2009) in R v3.6.2. Gene orthologues in humans (*Homo sapiens*), cattle (*Bos taurus*), mouse (*Mus musculus*) and rat (*Rattus norvegicus*) were extracted using the function *getLDS*. The associated gene ontology (GO annotations) information for all genes and orthologues were then extracted using the *getBM* function. All gene names, phenotype descriptions, GO terms and definitions were then queried for terms associated with meiosis and recombination, using the R command *grep* with the text strings *meio* and *recombin*.

Data Availability

Raw data will be publicly archived. Code for the analysis is archived at https://github.com/susjoh/2020_Soay_Recomb_GWAS.

179 Results

180 Variation and heritability of autosomal crossover count.

181 Males had higher recombination rates than females, with 7.24 more autosomal crossovers
 182 observed per gamete ($SE = 0.18$. Wald $\chi^2_1 = 1613.6$, $P < 0.001$; Figure 1). There was no
 183 association with ACC and age or the inbreeding coefficient $\hat{F}_{III}$ ($P > 0.05$; Figure 1). Females
 184 had higher phenotypic variance than males (female $V_P = 32.41$, male $V_P = 24.162$; Table 1).
 185 ACC was heritable in both sexes ($h^2 = 0.148$) and in males and females separately ($h^2 =$
 186 0.118 and 0.181 , respectively; Table 1); likelihood ratio tests confirmed that including the
 187 additive genetic effect significantly improved the animal models ($\chi^2_1 = 139.4$, 19.72 and 119.9
 188 for both sexes, males, and females, respectively; $P < 0.001$). The permanent environment
 189 contribution to phenotypic variance was not significant in any models ($P > 0.05$), with the
 190 remaining variance attributed to residual effects (Table 1). Bivariate models of ACC between
 191 males and females showed a positive genetic correlation ($r_A = 0.555$, $SE = 0.138$) that was
 192 significantly different from both $r_A = 0$ and 1 ($\chi^2_1 = 11.95$ and 17.91 , respectively; $P <$
 193 0.001). Linkage map lengths from the ACC analysis in this study were strongly correlated
 194 with the previous maps (linear regression, adjusted $R^2 > 0.999$, Figure S1). The updated map
 195 had 15,787 unique centiMorgan positions compared with 12,359 in the previous study; it is
 196 provided in Table S2 and Figure S2.

197 Genome-wide Association Study.

198 There was a strong association between SNPs at the sub-telomeric region of chromosome 6
 199 and female ACC (*oar3_OAR6_116402578*, $P = 2.45 \times 10^{-17}$; Figures 2 & 4; Tables 2 & S3).
 200 This locus explained 45.9% of the heritable variation in female ACC, with a difference of 5
 201 crossovers per gamete between the two homozygotes (Figure 3, Table 2). There was no sig-
 202 nificant association with this locus and male ACC, confirming that its action is likely to be
 203 sex-limited. The same SNP showed the highest genome-wide association in the previous study

(JOHNSTON *et al.*, 2016) and occurs 25.2 kb from the putative location of *RNF212*. It should be noted that *RNF212* is not annotated on the Oar_v3.1 sheep genome due to a gap in the sequence, but is predicted to be positioned between 116,427,802 and 116,446,904 bp (JOHNSTON *et al.*, 2016); *RNF212* is present in this position on the Oar_rambouillet_v1.0 genome assembly (GenBank assembly ID GCA_002742125.1). A second association was observed in both sexes on chromosome 7 (*oar3_OAR7_21368818* and *oar3_OAR7_21347355*, $P = 6.11 \times 10^{-13}$, 1.83×10^{-10} and 7.20×10^{-7} in both sexes combined, females and males, respectively); the association statistic in males was also identical for locus *oar3_OAR7_21116299* (Figures 2 & 5; Tables 2 & S3). The action of this locus was similar between males and females, with a difference of ~4.5 crossovers per gamete between the two homozygotes and explaining between 19.7% and 24.8% of the heritable variation in ACC (Figure 3, Table 2). These SNPs spanned the candidate locus *RNF212B* (21,241,831 to 21,273,165 bp). GO term analysis indicated that both *RNF212* and *RNF212B* were the closest genes to each hit that were associated with meiotic processes (Table S4; Figures 4 & 5).

218 Discussion

219 This study revisited a previous analysis investigating the genetic architecture of ACC in Soay
 220 sheep, with more than double the number of gametes and more than ten times the number
 221 of SNP loci. We confirmed that ACC is ~1.27 times higher in males than in females, and is
 222 not associated with age or inbreeding in neither males nor females. We identified two candi-
 223 date genes associated with ACC: *RNF212* on chromosome 6, which is associated with female
 224 recombination rate only; and its paralogue *RNF212B* on chromosome 7, which is associated
 225 with recombination rate in both sexes. Both loci have repeatedly been associated with recom-
 226 bination rate variation in mammals, with all studies conducted to date implicating either one
 227 or both of these loci; these include humans (KONG *et al.*, 2008, 2014), cattle (SANDOR *et al.*,
 228 2012; MA *et al.*, 2015; KADRI *et al.*, 2016), pigs (JOHNSON *et al.*, 2020), deer (JOHNSTON *et al.*,
 229 2018) and domestic sheep (PETIT *et al.*, 2017). Functional studies in mice indicate that the
 230 protein RNF212 is essential for crossover formation during meiosis, and that it has a dosage
 231 sensitive effect (REYNOLDS *et al.*, 2013). We observe an additive effect of alleles on ACC at
 232 both loci in this study, suggesting that a similar dosage sensitive effect is the mechanism driv-
 233 ing rate differences in Soay sheep.

234 The current study confirms the previous association at *RNF212* and does not gain any novel
 235 insights into the nature of its association with ACC. However, the observation at *RNF212B*
 236 builds on previously tentative evidence of an association in this genomic region. In the 2016
 237 study, a genome-wide regional heritability analysis identified an association in both sexes
 238 combined in a 1.09Mb region, which contained *RNF212B* and another candidate locus, mei-
 239 otic recombination protein *REC8*. However, there was no corresponding association using
 240 GWAS and there was no significant effect observed in the sex-specific regional heritability
 241 analyses, providing limited insight into the action of this region on ACC variation. Here,
 242 we have shown that adding a larger number of gametes and SNPs has increased the power
 243 to detect an association at *RNF212B*, and indicating that ACC variation is more likely to be
 244 attributed to *RNF212B* than *REC8* (although see further discussion below). The MAF of
 245 *RNF212B* is relatively low (0.082), suggesting that lower sample sizes in the 2016 study were

246 not sufficient to detect the effects of rare loci within the population. Using the SNP50 Chip
247 alone, the increased number of gametes in current study would have shown a significant hit
248 on chromosome 7 in females (but not in males; Figure 5), however, the highest association
249 would be around 345Kb away from the that observed in the current study. Our work shows
250 that increasing marker density using imputation is likely to be an increasingly important
251 approach in conducting GWAS in natural populations.

252 Soay sheep have extensive linkage disequilibrium (LD) throughout the genome due to recent
253 population bottlenecks, a small effective population size and a high prevalence of inbreeding
254 (CLUTTON-BROCK *et al.*, 2004; BÉRÉNOIS *et al.*, 2014; STOFFEL *et al.*, 2020). An advantage of
255 this is that there is a higher chance of typing SNP loci that are in LD with causal loci, and
256 that genotype imputation can be carried out with high accuracy, increasing the likelihood of
257 detecting associated regions. However, a disadvantage is that high LD within the population
258 makes it difficult to separate the effects of linked loci that may also contribute to phenotypic
259 variation in the population. The associated regions on both chromosomes 6 and 7 extend
260 over several megabases, with several other loci associated with meiotic processes occurring
261 within these regions (Figures 4 & 5, Table S4); we cannot rule out that these loci also affect
262 variation in ACC in Soay sheep. Functional validation of the role of *RNF212* and *RNF212B*
263 is not possible within this wild system as experimental manipulation and invasive sampling
264 are prohibited. However, further studies in domestic sheep (e.g. PETIT *et al.* 2017) and related
265 species may further elucidate the role of loci within these regions and the relative importance
266 of regulatory and protein-coding variation in driving recombination rate variation.

267 Overall, this study has highlighted the merits of genotype imputation and increasing sample
268 sizes to determine the genetic architecture of recombination rate variation in a natural system.
269 Identifying candidate loci and their sex-specific effect sizes provides a stepping stone for future
270 studies investigating the evolution of recombination rates within this and related systems,
271 through modelling the temporal dynamics of these loci, their origin and their association
272 with individual fitness.

273 Acknowledgements

274 We thank Jill Pilkington, Ian Stevenson and all Soay sheep project members and volunteers
 275 for collection of data and samples. Camillo Bérénos, Philip Ellis, Hannah Lemon and Penny
 276 Jack prepared DNA samples for Ovine SNP50 BeadChip genotyping and Jisca Huisman as-
 277 sisted with the pedigree construction. Genotyping was performed by Lee Murphy and col-
 278 leagues at the Wellcome Trust Clinical Research Facility Genetics Core in Edinburgh. This
 279 work has made use of the resources provided by the Edinburgh Compute and Data Facil-
 280 ity (<http://www.ecdf.ed.ac.uk/>). Permission to work on St Kilda is granted by The National
 281 Trust for Scotland, and logistical support is provided by QinetiQ and Eurest. The Soay sheep
 282 project has been supported by grants from the UK Natural Environment Research Council,
 283 with the SNP genotyping funded by a European Research Council Advanced Grant to JMP.
 284 SEJ is supported by a Royal Society University Research Fellowship.

285 Author Contributions

286 S.E.J and J.M.P. conceived the study. J.M.P organised the collection of samples. M.S. con-
 287 ducted the genotype imputation. S.E.J. analysed the data and wrote the paper. All authors
 288 contributed to revisions.

289 Figures and Tables

Figure 1: Association between autosomal crossover count and: A. sex; B. inbreeding coefficient $\hat{F}_{III}$; and C. age in years. Plots were made using the raw data. Points are red for females and blue for males. Lines in B. and C. are general additive model smoothing parameters as fit by the plotting library ggplot2 v3.3.2 (WICKHAM, 2016) in R v3.6.2. Points in plot C are slightly jittered between the sexes on the x-axis to more clearly visualise the sex differences with age. Sample sizes are provided in Table 1.

Figure 2: Genome-wide association plots of autosomal crossover counts for A. both sexes combined, B. females only and C. males only. The dashed line is the significant threshold equivalent to $\alpha < 0.05$. The left panels show the association for individual SNPs relative to their genomic position, with points colour-coded by chromosome. The right plots show the distribution of the observed $-\log_{10}P$ values against those expected under a null distribution. All association statistics have been corrected for genomic control using the λ parameter. All results are provided in Table S3. Sample sizes are provided in Table 1.

Figure 3: Effect sizes for the most highly associated SNPs shown for males and females, as calculated from animal models including the SNP genotype as a fixed factor. Genotype AA is the model intercept. Numbers indicate the number of observations at each genotype, with the number of unique focal individuals in parentheses.

Figure 4: Regional association statistics for the significant region on chromosome 6. Each point represents a single SNP locus, coloured by their origin on the SNP50 (red) or SNPHD Chips (black, i.e. imputed genotypes). The positions of genes associated with meiotic processes are shown as black bars at the top of each plot (Gene information in Table S4).

Figure 5: Regional association statistics for the significant region on chromosome 7. Each point represents a single SNP locus, coloured by their origin on the SNP50 (red) or SNPHD Chips (black, i.e. imputed genotypes). The positions of genes associated with meiotic processes are shown as black bars at the top of each plot (Gene information in Table S4).

Table 1: Data information and animal model results for autosomal crossover count (ACC) in Soay sheep. Numbers in parentheses are the standard error unless otherwise stated. N_{OBS} is the observed number of gametes, with numbers in parentheses the number of unique focal individuals. N_{xovers} are the total number of crossovers in the dataset. The mean ACC and the variance (V_{OBS}) were calculated from the raw data. V_P is phenotypic variance. h^2 , pe^2 and e^2 are the narrow-sense heritability, the permanent environment effect, and the residual effect, respectively; all are calculated as the proportion of V_P .

Analysis	N_{OBS}	N_{xovers}	Mean	V_{OBS}	V_P	h^2	pe^2	e^2
Both Sexes	7626 (1483)	227646	29.851	41.32	29.514 (0.588)	0.148 (0.019)	0.025 (0.013)	0.827 (0.014)
Males	3043 (434)	103914	34.149	23.974	24.162 (0.72)	0.118 (0.032)	0 (0.024)	0.882 (0.02)
Females	4583 (1049)	123732	26.998	32.439	32.41 (0.797)	0.181 (0.024)	0 (0.017)	0.819 (0.017)

Table 2: The most highly associated SNPs from GWAS of ACC in Soay sheep. MAF is the minor allele frequency (corresponding to the A allele for each locus). Sex indicates if the model was run in both sexes combined, or in females or males only. P is the P-value for the association statistic after correction using genomic control λ . Effect sizes are provided for animal models where the genotype fit as a fixed factor, and are given relative to the model intercept of 0 at genotype AA. Prop. V_A is the proportion of the additive genetic variance explained by the genomic region containing the SNP locus. All values in parentheses are standard errors. Full GWAS results are provided in Table S3. Effect sizes and associated sample sizes for each genotype are shown in Figure 3.

SNP Information	MAF	Sex	P	Effect AG	Effect GG	Prop. V_A
oar3_OAR6_116402578	0.275	Both Sexes	2.45×10^{-17}	2.201	3.837	0.399
Chr. 6				(0.332)	(0.361)	(0.118)
Position 116,402,578		Females	1.68×10^{-18}	3.038	4.985	0.459
Candidate Gene: <i>RNF212</i>				(0.397)	(0.423)	(0.115)
		Males	0.0317	0.168	0.884	0.049
				(0.577)	(0.607)	(0.067, ns)
oar3_OAR7_21368818*	0.082	Both Sexes	6.11×10^{-13}	-2.383	-4.520	0.242
Chr. 7				(0.922)	(0.952)	(0.102)
Position 21,368,818		Females	1.83×10^{-10}	-2.705	-4.757	0.197
Candidate Gene: <i>RNF212B</i>				(1.152)	(1.194)	(0.09)
		Males	7.20×10^{-07}	-2.479	-4.394	0.248
				(1.454)	(1.447)	(0.145)

* This association was identical to hits at SNP loci oar3_OAR7_21347355 at position 21,347,355 (in both sexes, males and females) and oar3_OAR7_21116299 at position 21,116,299 (in males).

290 References

- 291 ANTOLÍN, R., C. NETTELBLAD, G. GORJANC, D. MONEY, and J. M. HICKEY, 2017 A hybrid
292 method for the imputation of genomic data in livestock populations. *Genetics Selection*
293 *Evolution* : 1–17.
- 294 BARTON, N. H., and B. CHARLESWORTH, 1998 Why sex and recombination? *Science* **74**:
295 187–195.
- 296 BATTAGIN, M., G. GORJANC, A.-M. FAUX, S. E. JOHNSTON, and J. M. HICKEY, 2016 Effect of
297 manipulating recombination rates on response to selection in livestock breeding programs.
298 *Genet. Select. Evol.* **48**: 44.
- 299 BÉRÉNOIS, C., P. A. ELLIS, J. G. PILKINGTON, and J. M. PEMBERTON, 2014 Estimating quan-
300 titative genetic parameters in wild populations: a comparison of pedigree and genomic
301 approaches. *Mol. Ecol.* **23**: 3434–3451.
- 302 BUTLER, D. G., B. R. CULLIS, A. R. GILMOUR, and B. J. GOGEL, 2009 Mixed Models for S
303 language Environments: ASReml-R reference manual.
- 304 CLUTTON-BROCK, T., J. PEMBERTON, T. N. COULSON, I. R. STEVENSON, and A. D. C. MAC-
305 COLL, 2004 The sheep of St Kilda. In *Soay Sheep: Dynamics and Selection in an Island*
306 *Population*.2. Cambridge University Press, 17–51.
- 307 DURINCK, S., P. T. SPELLMAN, E. BIRNEY, and W. HUBER, 2009 Mapping identifiers for the inte-
308 gration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols*
309 **4**: 1184–1191.
- 310 GAO, X., J. STARMER, and E. R. MARTIN, 2008 A multiple testing correction method for
311 genetic association studies using correlated single nucleotide polymorphisms. *Genetic epi-*
312 *demiology* **32**: 361–369.
- 313 GREEN, P., K. FALLS, and S. CROOKS, 1990 *Documentation for CRIMAP, version 2.4*. Washing-
314 ton University School of Medicine.

315 HALLDORSSON, B. V., G. PALSSON, O. A. STEFANSSON, H. JONSSON, M. T. HARDARSON, *et al.*,
316 2019 Characterizing mutagenic effects of recombination through a sequence-level genetic
317 map. *Science* **363**: eaau1043.

318 HASSOLD, T., and P. HUNT, 2001 To err (meiotically) is human: the genesis of human aneu-
319 ploidy. *Nat. Rev. Genet.* **2**: 280–291.

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

322 HICKEY, J. M., B. P. KINGHORN, B. TIER, J. H. J. V. D. WERF, and M. A. CLEVELAND, 2012
323 A phasing and imputation method for pedigreed populations that results in a single-stage
324 genomic evaluation. *Genet Sel Evol* **44**: 9.

325 HILL, W. G., and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection.
326 *Genet. Res.* **8**: 269–294.

327 HUISMAN, J., 2017 Pedigree reconstruction from SNP data: parentage assignment, sibship
328 clustering and beyond. *Mol. Ecol. Resour.* **17**: 1009–1024.

329 JOHNSSON, M., A. WHALEN, R. ROS-FREIXEDES, G. GORJANC, C.-Y. CHEN, *et al.*, 2020 Genetics
330 of recombination rate variation in the pig. *bioRxiv* .

331 JOHNSTON, S. E., C. BÉRÉNOIS, J. SLATE, and J. M. PEMBERTON, 2016 Conserved genetic ar-
332 chitecture underlying individual recombination rate variation in a wild population of soay
333 sheep (*Ovis aries*). *Genetics* **203**: 583–598.

334 JOHNSTON, S. E., J. HUISMAN, and J. M. PEMBERTON, 2018 A genomic region containing
335 *REC8* and *RNF212B* is associated with individual recombination rate variation in a wild
336 population of red deer (*Cervus elaphus*). *BioRxiv* .

337 KADRI, N. K., C. HARLAND, P. FAUX, N. CAMBISANO, L. KARIM, *et al.*, 2016 Coding and
338 noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect re-
339 combination rate in cattle. *Genome research* **26**: 1323–1332.

340 KIJAS, J. W., D. TOWNLEY, B. P. DALRYMPLE, M. P. HEATON, J. F. MADDOX, *et al.*, 2009 A
341 genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS
342 ONE 4: e4668.

343 KONG, A., J. BARNARD, D. F. GUDBJARTSSON, G. THORLEIFSSON, G. JONSDOTTIR, *et al.*, 2004
344 Recombination rate and reproductive success in humans. Nat. Genet. 36: 1203–1206.

345 KONG, A., G. THORLEIFSSON, M. L. FRIGGE, G. MASSON, D. F. GUDBJARTSSON, *et al.*, 2014
346 Common and low-frequency variants associated with genome-wide recombination rate.
347 Nat. Genet. 46: 11–6.

348 KONG, A., G. THORLEIFSSON, H. STEFANSSON, G. MASSON, A. HELGASON, *et al.*, 2008 Se-
349 quence variants in the RNF212 gene associate with genome-wide recombination rate. Sci-
350 ence 319: 1398–1401.

351 LENORMAND, T., and J. DUTHEIL, 2005 Recombination difference between sexes: a role for
352 haploid selection. PLoS biology 3: e63.

353 MA, L., J. R. O’CONNELL, P. M. VANRADEN, B. SHEN, A. PADHI, *et al.*, 2015 Cattle Sex-Specific
354 Recombination and Genetic Control from a Large Pedigree Analysis. PLoS Genet. 11:
355 e1005387.

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

357 MYERS, S., L. BOTTOLO, C. FREEMAN, G. McVEAN, and P. DONNELLY, 2005 A fine-scale map
358 of recombination rates and hotspots across the human genome. Science 310: 321–4.

359 NAGAMINE, Y., R. PONG-WONG, P. NAVARRO, V. VITART, C. HAYWARD, *et al.*, 2012 Localising
360 Loci underlying Complex Trait Variation Using Regional Genomic Relationship Mapping.
361 PLoS ONE 7: e46501.

362 OTTO, S. P., and N. H. BARTON, 2001 Selection for recombination in small populations.
363 Evolution 55: 1921–1931.

364 OTTO, S. P., and T. LENORMAND, 2002 Resolving the paradox of sex and recombination. Nat.
365 Rev. Genet. 3: 252–261.

OTTO, S. P., and B. A. PAYSEUR, 2019 Crossover Interference: Shedding Light on the Evolution of Recombination. *Annual review of genetics* **53**: 19–44.

PETIT, M., J.-M. ASTRUC, J. SARRY, L. DROUILHET, S. FABRE, *et al.*, 2017 Variation in Recombination Rate and Its Genetic Determinism in Sheep Populations. *Genetics* **207**: 767–784.

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**: 269–78.

RÖNNEGÅRD, L., S. E. MCFARLANE, A. HUSBY, T. KAWAKAMI, H. ELLEGREN, *et al.*, 2016 Increasing the power of genome wide association studies in natural populations using repeated measures - evaluation and implementation. *Methods in Ecology and Evolution* **7**: 792–799.

SAMUK, K., B. MANZANO-WINKLER, K. R. RITZ, and M. A. F. NOOR, 2020 Natural Selection Shapes Variation in Genome-wide Recombination Rate in *Drosophila pseudoobscura*. *Current Biology* **30**: 1517–1528.e6.

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**: 1–13.

STAPLEY, J., P. G. D. FEULNER, S. E. JOHNSTON, A. W. SANTURE, and C. M. SMADJA, 2017 Variation in recombination frequency and distribution across Eukaryotes: patterns and processes. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **372**: 20160455.

STOFFEL, M., S. JOHNSTON, J. PILKINGTON, and J. PEMBERTON, 2020 Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. *bioRxiv*.

WICKHAM, H., 2016 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

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.