bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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.

Insights into the genetic determinism and evolution of recombination rates from combining multiple genome-wide datasets in Sheep

Morgane Petit[*], Jean-Michel Astruc[†], Julien Sarry[*], Laurence Drouilhet[*], Stéphane Fabre[*], Carole Moreno[*], Bertrand Servin[*]

*INRA, Génétique, Physiologie et Systèmes d'Elevage, F-31326 Castanet-Tolosan, France

[†] Institut de l'Elevage, F-31326 Castanet-Tolosan, France

Short running title: Genetics of recombination in sheep.

Keywords: recombination rate; genetic maps; QTLs; evolution; sheep

Corresponding author:

Bertrand Servin

GenPhySE, INRA UR631, F-31326 Castanet-Tolosan, France

phone: +33 5 61 28 51 17

Email: bertrand.servin@inra.fr

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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.

Abstract

Recombination is a complex biological process that results from a cascade of multiple events during meiosis. Understanding the genetic determinism of recombination can help to understand if and how these events are interacting. To tackle this question, we studied the patterns of recombination in sheep, using multiple approaches and datasets. We constructed genetic maps in the Lacaune breed at a fine scale by combining meiotic recombination rates from a large pedigree genotyped with a 50K SNP array and historical recombination rates from a sample of unrelated individuals genotyped with a 600K SNP array. This analysis revealed recombination patterns in sheep similar to other mammals but also genome regions that have likely been affected by directional and diversifying selection. We estimated the average recombination rate of Lacaune sheep at 1.5 cM/Mb, identified about 50,000 crossover hotspots on the genome and found a high correlation between population- and pedigree-based recombination rate estimates. A genome-wide association study revealed three major loci affecting inter-individual variation in recombination rate, including the *RNF212*, *HEI10* and *KCNJ15* genes. Finally, we compared our results to those obtained previously in a distantly related population of domestic sheep, the Soay. This comparison revealed on one hand that Soay and Lacaune sheep have similar genetic determinant of total recombination rate while confirming that Soay sheep have experienced strong selection greatly increasing their recombination rate. Taken together these observations highlight that multiple genetically independent pathways affect recombination rate.

Introduction

Meiotic recombination is a fundamental biological process that brings a major contribution to the evolution of populations and species. During meiosis, recombination enables proper chromosomal alignment resulting in proper disjunction and segregation of chromosomes, avoiding deleterious outcomes such as aneuploidy (Hassold *et al.* 2007). Over generations, recombination contributes to shaping genetic diversity in a population by creating new allelic combinations and preventing the

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

accumulation of deleterious mutations. Over large evolutionary timescales, divergence in recombination landscapes can lead to speciation: the action of a key actor in the recombination process in many mammals, the gene *PRDM9*, has been shown to have a major contribution to the infertility between two mouse species, making it the only known speciation gene in mammals today (Davies *et al.* 2016).

Genetics studies on recombination were first used to infer the organisation of genes along the genome (Sturtevant 1913). With the advance in molecular techniques, more detailed physical maps and eventually whole genome assemblies are now available in many species. In the context of mammalian genome cartography, recombination maps are now essentially used only to validate the physical ordering obtained from sequencing experiments (Groenen *et al.* 2012; Jiang *et al.* 2014). However, from an evolutionary perspective, the relevant distance between loci is the genetic distance, which is governed by recombination. Hence, the establishment of highly resolutive recombination maps remains of fundamental importance for genetic studies of a species for estimation of past demography (Li and Durbin 2011; Boitard *et al.* 2016), detection of selection signatures (Sabeti *et al.* 2002; Voight *et al.* 2006), QTL mapping (Cox *et al.* 2009) and imputation of genotypes (Howie *et al.* 2009) for genome-wide association studies (GWAS) or genomic selection. Precise recombination maps can be estimated using different approaches. Meiotic recombination can be estimated from the observation of markers' segregation in families. Although this is a widespread approach, its resolution is limited by the number of meioses that can be collected within a population. A second approach is to take advantage of patterns of correlation between allele frequencies in a population (*i.e.* Linkage Disequilibrium, LD) to infer population-based recombination rates. Because the LD-based approach exploits in essence meioses accumulated over many generations, it provides more precise estimates of recombination rate. For example, until recently (Pratto *et al.* 2014; Lange *et al.* 2016) this was the only approach allowing to detect fine scale patterns of recombination genome-wide. One caveat of LD-based approaches is that their recombination rate estimates are affected by other evolutionary processes, especially selection that affects LD patterns unevenly across the genome.

The LD-based approach does not allow to study individual phenotypes and therefore to identify directly loci influencing interindividual variation in recombination rates. In contrast, family-based studies in human (Kong *et al.* 2008; Chowdhury *et al.* 2009), cattle (Sandor *et al.* 2012; Ma *et al.*

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

2015; Kadri *et al.* 2016) and sheep (Johnston *et al.* 2016) have demonstrated that recombination exhibits interindividual variation and that this variation is partly determined by genetic factors. Two recombination phenotypes have been described: the variation in the number of crossovers per meiosis (Genome-wide Recombination Rate , GRR herein) and the variation in the fine scale localization of crossovers (Individual Hotspot Usage, IHU herein). GRR has been shown to be influenced by several genes. For example, a recent genome-wide association study found evidence for association with 6 genome regions in cattle (Kadri *et al.* 2016). Among them, one of the genomic regions consistently found associated to GRR in mammals is an interval containing the *RNF212* gene. In contrast to GRR, the IHU phenotype seems mostly governed by a single gene in most mammals, *PRDM9*. This zinc-finger protein has a key role in recruiting *SPO11*, thereby directing DNA double-strand breaks (DSBs) that initiate meiotic recombination. Because *PRDM9* recognizes a specific DNA motif, the crossover events happen in small genomic regions carrying this motif, termed recombination hotspots. This *PRDM9* associated process is however not universal, as it is only active in some mammals; canids for example do not carry a functional copy of *PRDM9* and exhibit different patterns for the localization of recombination hotspots (Auton *et al.* 2013).

In this work, we first present a strategy to establish fine-scale recombination maps by combining a pedigree analysis and an LD-based approach. We exploited two different datasets from a single sheep population (the Lacaune breed): a large pedigree genotyped with a medium density genotyping array (about 50K SNP) and a sample of unrelated individuals genotyped with a high-density genotyping array (about 500K SNP). The study of these two datasets and their combination allowed to produce high-resolution recombination maps. Second we exploited family data to understand the genetic determinism of individual variation in GRR in the Lacaune population. Finally, we compared our recombination maps and results of our quantitative traits loci (QTLs) analysis to those obtained from an analysis of pedigree data in Soay sheep (Johnston *et al.* 2016). This comparison informed on how recombination rate has evolved in the Soay population, offering new evidence that recombination rate is genetically determined by multiple independent biological pathways.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Materials and Methods

Study Population and Genotype Data

In this work, we exploited two different datasets of sheep from the Lacaune breed: a pedigree dataset of 8,085 related animals genotyped with the medium density Illumina Ovine Beadchip® including 54,241 SNPs, and a diversity dataset of 64 unrelated Lacaune individuals selected as to represent population genetic diversity, genotyped with the high density Illumina Ovine Infinium® HD SNP Beadchip including 685,734 SNPs (Rochus *et al.* 2017; Moreno-Romieux *et al.* 2017). Standard data cleaning procedures were carried out using plink 1.9 (Chang *et al.* 2015), excluding animals with call rates below 95% and SNPs with call freq below 98%. After quality control, 6,230 animals and 46,813 SNPs were kept in the family dataset and 503,784 autosomal SNPs in the diversity dataset. Among these 6,230 animals, 5,928, had their father known and were kept for the analyses.

Recombination Maps

Family-based recombination maps

Detection of crossovers

We detected the localization of crossovers using LINKPHASE (Druet and Georges 2015) on each chromosome. From the LINKPHASE outputs (*recombination_hmm* files), we extracted crossovers boundaries. We then identified crossovers occurring in the same meiosis less than 3 Mb apart from each other (that we call double crossovers) and considered them as dubious. We applied the following procedure until none of them remained. Given a pair of double crossovers, we set the genotype of the corresponding offspring as missing in the region spanned by the most extreme boundaries and re-run the LINKPHASE analysis. After this quality control step, we used the final

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

set of crossovers identified by LINKPHASE to estimate recombination rates.

Estimation of recombination rates

Based on the inferred crossover locations, we estimated recombination rates in windows of one megabase and between marker intervals of the medium SNP array using the same procedure, described below.

Given an interval j on the genome, we estimated its recombination rate c_j from the number y_{sj} of crossovers observed in the interval for an individual s using the model:

$$y_{sj} | c_j \sim Poisson(0.01 \tfrac{1}{\hat{\mu}} n_s l_j c_j) \quad (1)$$

where n_s is the total number of crossovers in all meioses of individual s, $\hat{\mu}$ is the average number of crossovers per meiosis across all individuals and l_j is the interval length (in megabases). With this parametrization, c_j is expressed in centiMorgans per megabase. To combine crossovers across individuals, the likelihood for c_j is the product of poisson likelihoods from equation (1).

We then specify a prior distribution for c_j:

$$c_j \sim \Gamma(\alpha, \beta) \quad (2)$$

To set α and ß, we first computed raw c_j estimates using the method of (Sandor *et al.* 2012) across the genome and fitted a gamma distribution to the resulting genome-wide distribution (Figure S3). Combining the prior (2) with the likelihoods in equation (1), we obtain that the posterior distribution for c_j is:

$$c_j \sim \Gamma(\alpha + \sum_s y_{sj}, \beta + \sum_s 0.01 \tfrac{1}{\hat{\mu}} n_s l_j) \quad (3)$$

As the localization of crossovers was usually not good enough to assign them with certainty to a single genomic interval, we obtained our final estimates of c_j as follows :

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

(i) for each crossover overlapping interval j and localized within a window of size L, let x_c be an indicator variable that takes value 1 if the crossover occurred in interval j and 0 otherwise. We assume that locally recombination rate is proportional to physical distance and set $P(x_c = 1) = min(l_j/L, 1)$.

(ii) Using the probability in step (i), sample x_c for each crossover overlapping interval j and set $y_{sj} = \sum_c x_c$

(iii) Given y_{sj}, sample c_j from equation (3)

For each interval considered, we performed step (ii) and (iii) above 1000 times to get samples from the posterior distribution of c_j thereby accounting for uncertainty in the localization of crossovers.

Population-based recombination maps

We estimated population-scaled recombination rates using PHASE (Li and Stephens 2003) on the 64 unrelated Lacaune individuals genotyped on the HD SNP array. For computational reasons, we performed the estimations in 2 Mb windows, with an additional 100 Kb on each side overlapping with neighbouring windows, to avoid border-effect in the PHASE inference. PHASE was run on each window with default options, except that we increased the number of main iterations to obtain larger posterior samples for recombination rate estimation (option -X10) as recommended in the documentation.

From the PHASE output, we obtained 1000 samples from the posterior distribution of:

- The background recombination rate: $\rho_w = 4Nc_w$, where N is the effective population size, c_w is the recombination rate comparable to the family-based estimate.
- An interval specific recombination intensity λ_j, for each marker interval j of length l_j in the window, such that the population scaled genetic length of an interval is: $\delta_j = \rho_w \lambda_j l_j$

As point estimates of λ_j and δ_j, we used the median computed over the posterior distributions $\{\lambda_j^{(k)}, \lambda_j^{(k)} \rho_w^{(k)} l_j;\ k \in [1, 1000]\}$.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Identification of intervals harbouring crossover hotspots

Intervals that showed an outlying λ_j value compared to the genome-wide distribution of λ_j were considered as harbouring a crossover hotspot. Specifically, we fitted a mixture of Gaussian distribution to the genome-wide distribution of $log_{10}(\lambda_j)$, considering that the major component of the mixture modelled the background distribution of λ_j in non-hotspots intervals. From this background distribution, we computed a p-value for each interval corresponding to the null hypothesis that it does not harbour a hotspot. Finally, we called hotspot harbouring intervals those for which FDR(λ_j) < 5%, estimating FDR with the (Storey and Tibshirani 2003) method, implemented in the R qvalue package. This procedure is illustrated in Figure S4.

Combination of meiotic- and population-based recombination rates and construction of a high resolution recombination map

To construct a recombination map of the HD SNP array, we needed to scale the population-based estimates by 4 Ne, where Ne is the effective population size. Due to evolutionary pressures, Ne varies along the genome, so it must be estimated locally. This can be done by exploiting the meiotic recombination rate inference obtained from family data as follows.

Consider a window of one megabase on the genome, using the approach described above, we can sample values c_{jk} (window j, sample k) from the posterior distribution of the meiotic recombination rate c_j. Similarly, using output from PHASE we can extract samples ρ_{jk} from the posterior distribution of the population-based recombination rates ρ_j. Now, considering that $\rho_j = 4Ne_jc_j$ where Ne_j is the local effective population size of window j, we get $log(\rho_j) = log(4Ne_j) + log(c_j)$. This justifies using a model on both c_{jk} and ρ_{jk} values:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \nu_{ij} + e_{ijk} \quad (4)$$

where y_{ijk} is $log(c_{jk})$ when i=1 (meiotic-recombination rate sample) and y_{ijk} is $log(\rho_{jk})$ when i=2

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

(population-based recombination rate sample). In this model, with appropriate design matrices, μ estimates the log of the genome-wide recombination rate, α_i estimates log(4Ne), where Ne is the average effective population size of the Lacaune population, $\mu+\beta_j$ estimates log(c_j) combining population and meiotic recombination rates, and $\alpha_i+\nu_{ij}$ estimates log(4Ne_j). μ and α_i were considered as fixed effects while β_j and ν_{ij} were considered as independent random effects. Using this approach allows to combine in a single model population- and family-based inferences, while accounting for their respective uncertainties as we exploit posterior distribution samples.

In this study, we fitted model (4) with an additional fixed effect for each chromosome, using the lme4 R package (Bates *et al.* 2015). We considered windows of one megabase covering the genome, using 20 samples of the posterior distributions of c_j and ρ_j. We omitted windows lying less than 4 Mb from each chromosome end because our inference on c_j was likely biased in these regions (see Results). After estimating this model, we corrected population-based recombination rate estimates of HD intervals within each window by dividing them by their estimated local effective population size (*i.e.* exp($\hat{\alpha}_i+\hat{\nu}_{ij}$) for window *j*). For windows lying within 4 Mb of the chromosome ends, we corrected population-based estimates using the genome-wide average effective population size (*i.e.* exp($\hat{\alpha}_i$)). This led eventually to estimates of the recombination rates, expressed in centiMorgans per megabase, for all intervals of the HD SNP array, which we termed a high resolution recombination map.

Genome-Wide Association Study on Recombination Phenotypes

Genome-wide Recombination Rate (GRR)

We exploited the set of crossovers detected to estimate the genome-wide recombination rate (GRR) of each FID in the family dataset from their observed number of crossovers per meiosis, adjusting for covariates: year of birth of the parent, considered as a cofactor with 14 levels for years spanning from 1997 to 2010, and insemination month of the offspring's ewe, treated as a cofactor with 7 levels for months spanning from February to August. We used a mixed-model for estimating the

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

population average GRR μ, covariates fixed effects β and individual specific deviations u_s:

$$y_{so} = \mu + x_{so}\beta + u_s + e_{so}$$

with $u_s \sim N(0, A\sigma_s^2)$ and $e_{so} \sim N(0, I\sigma_e^2)$, where y_{so} is the number of crossovers in the meiosis between FID s and offspring o, A is the pedigree-based relationship matrix between FIDs and x_{so} the line of the corresponding design matrix for observation y_{so}. We fitted this model using BLUPf90 (Misztal et al., 2002) and extracted: (i) estimates of variance components σ_e^2 and σ_s^2, which allows to estimate the heritability of the trait (calculated as $\sigma_s^2/(\sigma_s^2+\sigma_e^2)$) and (ii) prediction $\tilde{u}_s$ of GRR deviation for each FID.

Genotype Imputation

Using the 64 unrelated Lacaune as a panel, we imputed the genotypes of 345 animals using BIMBAM (Guan and Stephens, 2008) on 507,784 SNPs. To impute, BIMBAM uses the fastPHASE model (Scheet and Stephens, 2006), which relies on methods using cluster of haplotypes to estimate missing genotypes and reconstruct haplotypes from unphased SNPs of unrelated animal. BIMBAM was run with 10 expectation-maximization (EM) starts, each EM was run 20 steps on panel data alone, and an additional 1 step on cohort data, with a number of clusters of 15. From the BIMBAM results, in subsequent analyses, we used the mean genotypes of the 345 FIDs at all markers of the HD SNP array. To validate the quality of genotype imputation, 10 markers of the HD SNP array, 1 in chromosome 6 associated region and 9 in the chromosome 7 associated region were genotyped for 266 FIDs. We evaluated the quality of imputation by comparing the posterior genotype probabilities at these 10 markers to the error rate if we were to call them. We observed a very good agreement between the two measures (Figure S6), which denoted good calibration of the imputed genotypes.

Single- and multi-QTLs GWAS on GRR

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

We first tested association of GRR with mean genotypes at 503,784 single SNPs imputed with BIMBAM. We tested these associations using the univariate mixed-model approach implemented in the Genome-wide Efficient Mixed Model Association (GEMMA) software (Zhou and Stephens, 2012). In the mixed-model, we used the centered genomic relationship matrix calculated from the mean genotypes. The p-values reported in the results correspond to the Wald test.

To go beyond single SNP association tests, we also estimated a Bayesian sparse linear mixed-model (Zhou *et al.* 2013) as implemented in GEMMA. This method allows to consider multiple QTLs in the model, together with polygenic effects at all SNPs. Inference of the model parameters was done using an iterative MCMC algorithm ; we set the number of iterations to 10 millions and extracted samples every 100 iterations.

Variant Discovery and Additional Genotyping in *RNF212*

Identification and assignation of the *RNF212* sheep genome sequence

The *RNF212* gene was not annotated on the *Ovis aries* v3.1 reference genome. Nevertheless, a full sequence of *RNF212* was found in the scaffold01089 of *Ovis orientalis* (assembly Oori1, NCBI accession NW_011943327). By BLAST alignment of this scaffold, ovine *RNF212* could be located with confidence on chromosome 6 in the interval OAR6:116426000-116448000 of Oari3.1 reference genome (Figure S11). This location was confirmed by BLAST alignment with the bovine *RNF212* gene sequence. We also discovered that the Oari3.1 unplaced scaffold005259 (NCBI accession JH922970) contained the central part of *RNF212* (exons 4-9) and it could be placed within a large assembly gap. Moreover, we also observed that automatically annotated non-coding RNA in the *RNF212* interval matched exonic sequence of *RNF212* (Figure S11).

Variant discovery in *RNF212* in the Lacaune population

Based on the genomic sequence and structure of the *RNF212* gene annotated in *Ovis orientalis* (NCBI accession NW_011943327), a large set of primers were designed using PRIMER3 software

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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 S1) for amplification of each annotated exon and some intron part corresponding to exonic region annotated in *Capra hircus* (Chir_v1.0). PCR amplification (GoTaq, Promega) with each primer pair was realized on 50ng of genomic DNA from 4 selected homozygous Lacaune animals exhibiting the GG and AA (non imputed) genotypes at the most significant SNP of the medium density SNP array of the chromosome 6 QTL (rs418933055, p-value 2.56e-17). Each PCR product was sequenced via the BigDye Terminator v3.1 Cycle Sequencing kit and analyzed on an ABI3730 sequencing machine (Applied Biosystems). Sequenced reads were aligned against the *Ovis orientalis RNF212* gene using CLC Main Workbench Version 7.6.4 (Qiagen Aarhus) in order to identify polymorphisms.

Genotyping of mutations in *RNF212*

The genotyping of 266 genomic DNA from Lacaune animals for the four identified polymorphisms within the ovine *RNF212* gene was done by Restriction Fragment Length Polymorphism (RFLP) after PCR amplification using dedicated primers (Table S1) (GoTaq, Promega), restriction enzyme digestion (BsrBI for SNP_14431_AG; RsaI for SNP_18411_GA; and Bsu36I for both SNP_22570_CG and SNP_22594_AG; New England Biolabs) and resolution on 2% agarose gel.

Comparison of Meiotic Recombination Maps between Soay males and Lacaune males

Using data and results from (Johnston *et al.* 2016), we could create genetic maps based on common markers in Soay and Lacaune (36,651 SNPs). We computed the genetic distance (in cM) for the Soay males and the Lacaune males while removing the 4 Mb at each chromosomal extremity as explained above. Finally, for each chromosome, we fitted a linear regression of the Lacaune male genetic distance on the Soay male genetic distance and collected all the slopes for each chromosome and for each mean chromosomal recombination rate.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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

High-Resolution Recombination Maps

Meiotic recombination maps: genome-wide recombination patterns

We studied meiotic recombination using a pedigree of 6,230 individuals, genotyped for a medium density SNP array (50K) comprising around 54,000 markers. After quality control we exploited genotypes at 46,813 SNPs and identified 213,118 crossovers in 5,928 meioses divided among 345 male parents which we called focal individuals (FIDs) hereafter (see methods). The pedigree information available varied among focal individuals (Figure 1): 281 FIDs had their father genotype known while the remaining 64 did not. However, having a missing parent genotype did not affect our ability to detect crossovers as the average number of crossovers per meiosis in the two groups was similar (36.0 with known father genotype and 35.9 otherwise) and the statistical effect of the number of offspring on the average number of crossovers per meiosis was not significant ($p>0.23$). We explained this by the fact that individuals that lacked father genotype information typically had a large number of offspring (11.2 on average, ranging from 4 to 104), allowing to infer correctly their haplotype phase from their offspring genotypes only. Overall, given that the physical genome size covered by the medium density SNP array is 2.45 gigabases, we estimate that the mean recombination rate in our population is about 1.5 cM/megabase.

Based on the crossovers identified, we developed a statistical model to estimate family-based recombination maps (see methods) and constructed recombination maps at two different scales : for windows of one megabase and for each interval of the medium density SNP array. As our statistical approach allowed to evaluate the uncertainty in our recombination rate estimates, we provide respectively in File S1 and S2, along with the recombination rate estimates in each interval, their posterior variance and 90% credible intervals. Graphical representation of the meiotic recombination maps of all autosomes are given in File S3.

The recombination rate on a particular chromosome region was found to depend highly on its

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

position relative to the telomere and to the centromere for metacentric chromosomes, *i.e.* chromosomes 1, 2 and 3 in sheep (Figure S1). Specifically, for acrocentric and metacentric chromosomes, recombination rate estimates were elevated near telomeres and centromeres, but very low within centromeres. In our analysis, recombination rate estimates were found low in intervals lying within 4 megabases of chromosome ends. This is most likely due to crossovers being undetected in our analysis as only few markers are informative to detect crossovers at chromosome ends. In the following analyses, we therefore did not consider regions lying within 4 Mb of the chromosomes ends.

From local recombination rate estimates in 1 Mb windows or medium SNP array intervals, we estimated chromosome specific recombination rates (Figure S2). Difference in recombination rates between chromosomes was relatively well explained by their physical size, larger chromosomes exhibiting smaller recombination rates. Even after accounting for their sizes, some chromosomes showed particularly low (chromosomes 9, 10 and 20) or particularly high (chromosomes 11 and 14) recombination rates. In low recombining chromosomes, large regions had very low recombination, between 9 and 14 Mb on chromosome 9, 36 and 46 Mb on chromosome 10 and between 27 and 31 Mb on chromosome 20. In highly recombining chromosomes, recombination rates were globally higher on chromosome 14, while chromosome 11 exhibited two very high recombination windows between 7 Mb and 8 Mb and between 53 and 54 Mb. In addition, we found, consistent with the literature, that GC content was quite significantly positively correlated with recombination rate (p-value $< 10^{-16}$, r=0.23) in medium SNP array intervals.

Population-based recombination maps: identification of crossover hotspots

We used a different dataset, with 64 unrelated individuals from the same Lacaune population genotyped for the Illumina HD SNP array (600K) comprising 503,784 autosomal SNPs after quality control. Using a multipoint model for LD patterns (Li and Stephens 2003), we estimated, for each marker interval of the HD SNP array, population-scaled recombination rates ρ (see Methods). Compared to pedigree-based maps, these estimates offer a greater precision as they in essence exploit meioses cumulated over many generations. However, the recombination rates obtained are scaled by the effective population size ($\rho = 4N_e c$ where Ne is the effective population size and c

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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 meiotic recombination rate) which is unknown, and may vary along the genome due to evolutionary pressures, especially selection. Thanks to the higher precision in estimation of recombination rate, population-based recombination maps offer the opportunity to detect genome intervals likely to harbour crossover hotspots. We considered about 50,000 intervals exhibiting elevated recombination intensities (Figure S4) as recombination hotspots, corresponding to an FDR of 5%. From our population-based recombination map, we could conclude that 80% crossover events occurred in 40% of the genome and that 60% of crossover events occurred in only 20% of the genome (Figure S5).

High-resolution recombination maps combining family and population data

Having constructed recombination maps with two independent approaches and datasets in the same population of Lacaune sheep allowed first to evaluate to which extent population-based crossover hotspots explain meiotic recombination, and second to estimate the impact of evolutionary pressures in the Lacaune population. We present our results on these questions in turn.

We studied whether variation in meiotic recombination can be attributed to the crossover hotspots detected from population data. For each interval between two adjacent SNPs of the medium density array, we (i) extracted the number of significant hotspots and (ii) calculated the crossover hotspot density (in number of hotspots per unit of physical distance). We found both covariates to be highly associated with meiotic recombination rate estimated on family data, the hotspot density even more than the number of hotspots (respectively $r=0.15$ ($p < 10^{-16}$) and $r=0.19$ ($p=1.8\ 10^{-4}$)). Figure 2 illustrates this finding in two one-megabase intervals from chromosome 24, one that exhibits a very high recombination rate (7.08 cM/Mb) and the second a low one (0.46 cM/Mb). In this comparison, the highly recombining window carries 36 recombination hotspots while the low recombinant one exhibits none. As the population-based background recombination rates in the two windows are similar (0.7/Kb for the one with a high recombination rate, and 0.2/Kb for the other), the difference in recombination rate between these two regions is largely due to their contrasted number of crossover hotspots.

In order to study more precisely the relationship between population and meiotic recombination

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

rates, we estimated a linear mixed model (see Methods) that allowed to estimate the average effective population size of the population, the correlation between meiotic and population-based recombination rates and to identify genome regions where population- and family-based estimates were significantly different. We found the effective population size of the Lacaune population to be about 7,000 individuals and a correlation of 0.73 between meiotic- and population-based recombination rates (Figure 3). We discovered 6 regions where population recombination rates were much lower than meiotic ones and 4 regions where they were much higher (Table 1, Figure S8).

Three of the regions with a low population-based recombination rate corresponded to previously identified selection signatures in sheep: a region on chromosome 6 spanning 3 intervals between 35 and 38 megabases contains the *ABCG2* gene, associated to milk production (Cohen-Zinder *et al.* 2005), and the *LCORL* gene associated to stature (recently reviewed in (Takasuga 2015)). This region has been shown to have been selected in the Lacaune breed (Fariello *et al.* 2014; Rochus *et al.* 2017); a region spanning one interval on chromosome 10, between 29 and 30 megabases contains the *RXFP2* gene, associated to polledness and horn phenotypes (Johnston *et al.* 2013) and found to be under selection in many sheep breeds (Fariello *et al.* 2014) ; and a region on chromosome 13 between 63 and 64 megabases that contains the *ASIP* gene responsible for coat color phenotypes in many breeds of sheep (Norris and Whan 2008) , again previously demonstrated to have been under selection. For these three regions, we explain the low population recombination estimates by a local reduction of the effective population size due to selection.

One of the four regions with a high population-based recombination rate, on chromosome 20 between 28 and 29 megabases, harbours a cluster of olfactory receptors genes and could be explained by selective pressure for increased genetic diversity in these genes (*i.e.* diversifying selection), a phenomenon which has been shown in other species (*e.g.* pig (Groenen *et al.* 2012), human (Ignatieva *et al.* 2014), rodents (Stathopoulos *et al.* 2014)). The seven other regions of low and high population-based recombination rates correspond to regions of very low (resp. very high) meiotic recombination intensities, and the discrepancy between meiotic recombination and population recombination estimates could come from the fact that we used a genome-wide prior in our model to estimate meiotic recombination rates that had the effect of shrinking our estimates toward the mean. Because population estimates were not shrunk in the same way, for these few outlying regions the two estimates did not concur and it is possible that our meiotic recombination

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

rate estimates were slightly over (resp. under) estimated. Finally, we used the meiotic recombination rates to scale the population-based estimates and produce high-resolution recombination maps on the HD SNP array (Supporting File S4).

Genetic Determinism of Genome-wide Recombination Rate in Lacaune sheep

Our dataset provides information on the number of crossovers for a set of 5,928 meioses among 345 male individuals. Therefore, it allows to study the number of crossovers per meiosis (GRR) as a recombination phenotype.

Genetic and environmental effects on GRR

We used a linear mixed-model to study the genetic determinism of GRR. The contribution of additive genetic effects was estimated by including a random FID effect with covariance structure proportional to the matrix of kinship coefficients calculated from pedigree records (see Methods). We also included environmental fixed effects in the model: year of birth of the FID and insemination month of the ewe for each meiosis. We did not find significant differences between the FID year of birth, however the insemination month of the ewe was significant ($p = 1.7\ 10^{-3}$). Based on the estimated variance components (Table 2), we estimated the heritability of GRR in our population at 0.23.

Genome-wide association study identifies three major loci affecting GRR in Lacaune sheep

The additive genetic values of FIDs, predicted from the above model were used as phenotypes in a genome-wide association study. Among the 345 FIDs with at least two offsprings, the distribution of the phenotype was found to be approximately normally distributed (Figure S9). To test for association of this phenotype with SNPs markers, we used a mixed-model approach correcting for

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

relatedness effects with a genomic relationship matrix (see Methods). Using our panel of 64 unrelated Lacaune, we imputed the 345 FIDs for markers of the HD SNP array. With these imputed genotypes, we performed two analyses. The first was an association test with univariate linear mixed models, which tested the effect of each SNP in turn on the phenotype, the second fitted a Bayesian sparse linear mixed model, allowing multiple QTLs to be included in the model.

Figure 4 illustrates the GWAS results: the top plot shows the p-values of the single SNP analysis and the bottom plot, the posterior probability that a region harbours a QTL, calculated on overlapping windows of 20 SNPs. The single SNP analysis revealed five significant regions (FDR = 5%): two on chromosome 1, one on chromosome 6, one on chromosome 7 and one chromosome 19. Regions of chromosome 6 and 7 exhibited very low p-values whereas the other three showed less intense association signals. The multi-QTLs Bayesian analysis was conclusive for three regions: the rightmost region on chromosome 1, and the regions on chromosome 6 and chromosome 7. Using the multi-QTL approach of (Zhou *et al.* 2013) allowed to estimate that together, these three QTLs explain about 40% of the additive genetic variance for GRR, with a 95% credible interval ranging from 28 to 53 %.

The most significant region was located on the distal end of chromosome 6 and corresponded to a locus including the *RNF212* gene, associated with recombination rate variation in human (Chowdhury *et al.* 2009),(Kong *et al.* 2008), in bovine (Sandor *et al.* 2012; Ma *et al.* 2015; Kadri *et al.* 2016) and in mouse (Reynolds *et al.* 2013). *RNF212* is not annotated in the sheep genome assembly oviAri3, however this chromosome 6 region corresponds to the bovine region that contains *RNF212* (Figure S11). We found an unassigned scaffold (scaffold01089, NCBI accession NW_011943327) of *Ovis orientalis musimon* (assembly Oori1) that contained the full *RNF212* sequence and that could be placed confidently in the QTL region. To confirm *RNF212* as a valid positional candidate, we studied further the association of its polymorphisms with GRR in results presented below.

The second most significant region was located between 22.5 and 23.1 megabases on chromosome 7. All significant SNPs in the region were imputed, *i.e.* the association would not have been found based on association of the medium density array alone. It matched an association signal on GRR in Soay sheep (Johnston *et al.* 2016). Consistent with our finding, in the Soay sheep study, this

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

association was only found using regional heritability mapping and not using single SNP associations with the medium density SNP array. This locus could match previous findings in cattle (association on chromosome 10 at about 20 Mb on assembly btau3.1), however the candidate genes mentioned in this species (*REC8* and *RNF212B*) were located respectively 2 and 1.5 megabases away from our strongest association signal. In addition, none of the SNPs located around these two candidate genes in cattle were significant in our analysis. However another functional candidate, *CCNB1IP1*, also named *HEI10*, was located between positions 23,946,971 and 23,951,850 bp, about 500 Kb from our association peak. This gene is a good functional candidate as it has been shown to interact with *RNF212*: *HEI10* allows to eliminate the *RNF212* protein from early recombination sites and to recruit other recombination intermediates involved in crossover maturation (Qiao *et al.* 2014; Rao *et al.* 2016). Again SNPs located at the immediate proximity of *HEI10* did not exhibit significant associations with GRR. Hence, our association signal did not allow to pinpoint any clear positional candidate among these functional candidates (see Figure S7). However, it was difficult to rule them out completely for three reasons. First, with only 345 individuals, our study may not be powerful enough to localize QTLs with the required precision. Second, the presence of causal regulatory variants, even at distances of several hundred kilobases is possible. Finally, the associated region of *HEI10* exhibited apparent rearrangements with the human genome, possibly due to assembly problems in oviAri3. These assembly problems could be linked to the presence of genomic sequences coding for the T-cell receptor alpha chain. This genome region is in fact rich in repeated sequences making its assembly challenging. Overall, identifying a single positional and functional candidate gene in this gene-rich misassembled genomic region was not possible based on our data alone.

Finally, our third associated locus was located on chromosome 1 between 268,600 and 268,700 kilobases. In cattle, the homologous region, located at the distal end of cattle chromosome 1, has also been shown to be associated with GRR (Ma *et al.* 2015; Kadri *et al.* 2016). In these studies the *PRDM9* gene has been proposed as a potential candidate gene, especially because it is a strong functional candidate given its proven effect on recombination phenotypes. In sheep, *PRDM9* is located at the extreme end of chromosome 1, around 275 megabases, 7 megabases away from our association signal (Ahlawat *et al.* 2016). Hence, *PRDM9* was not a good positional candidate for association with GRR in our sheep population. However, the associated region on chromosome 1 contains a single gene, *KCNJ15*, which has been associated with DNA double-strand breaks repair

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

in human cells (Słabicki *et al.* 2010).

Mutations in the *RNF212* gene are strongly associated to Genome-wide Recombination Rate variation in Lacaune sheep

The QTL with the largest effect in our association study corresponded to a locus associated to GRR variation in other species and harbouring the *RNF212* gene. As it was a clear positional and functional candidate gene, we carried out further experiments to interrogate specifically polymorphisms within this gene. As stated above, we used the sequence information available for the *RNF212* gene from *Ovis orientalis* which revealed that *RNF212* spanned 23,7 Kb on the genome and may be composed of 12 exons by homology with bovine *RNF212*. However, mRNA annotation indicated multiple alternative exons. Surprisingly, the genomic structure of ovine *RNF212* was not well conserved with goat, human and mouse syntenic *RNF212* genes (Figure S11). As a first approach, we designed primers for PCR amplification (see Methods) and sequencing of all annotated exons and some intronic regions corresponding to exonic sequences of *Capra hircus RNF212*. By sequencing *RNF212* from 4 carefully chosen Lacaune animals homozygous GG or AA at the most significant SNP of the medium density SNP array on chromosome 6 QTL (rs418933055, p-value 2.56 10^{-17}), we evidenced 4 polymorphisms within the ovine *RNF212* gene (2 SNPs in intron 9, and 2 SNPs in exon 10). The 4 mutations were genotyped in 266 individuals of our association study. We then tested their association with GRR using the same approach as explained above and computed their linkage disequilibrium (genotypic r^2) with the most associated SNPs of the high-density genotyping array (see Figure S10) (Table 4). Two of these mutations were found highly associated with GRR, their p-values being of the same order of magnitude ($p<10^{-14}$) as the most associated SNP (rs412583165), although none of them were more significant. Indeed, we found a clear agreement between the amount of LD between a mutation and the most associated SNPs and their association p-value (see Figure S10). Overall, these results showed that polymorphisms within the *RNF212* gene were strongly associated with GRR, and likely tagged the same causal mutation as the most associated SNP. This confirmed that *RNF212*, a very strong functional candidate, was also a very strong positional candidate gene underlying our association signal.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Comparison of Recombination Rate and its Genetic Determinism between Soay and Lacaune Sheep

Recently, a genome-wide association study on GRR was conducted in the Soay sheep. Soay sheep is a feral population of ancestral domestic sheep living on an island located northwest of Scotland. The Lacaune and Soay populations are genetically very distant, their genome-wide Fst, calculated using the sheephapmap data (Kijas *et al.* 2012), being about 0.4. Combining our results with results from the Soay offered a rare opportunity to study the evolution of recombination over a relatively short time scale as the two populations can be considered separated at most dating back to domestication, about 10 Kya.

We compared the average recombination rates of the two populations for each autosomes (See Methods). Although the recombination rates found in both studies were highly correlated, we found a consistent pattern of elevated recombination in Soay males compared to Lacaune males (Figure 5 left), with a median 20% and average 50% increase in recombination rate. This result was consistent with the conclusions from the Soay study that recombination rate was increased in this population due to selection on male fertility. Interestingly, the increase in recombination in Soay was not homogeneous among chromosomes. Larger differences were found in chromosomes with higher recombination rates in Lacaune (Figure 5 right): for example chromosome 10, which had a very low recombination rate, did not show much increased recombination in Soay while the most recombining chromosome, chromosome 24, showed the highest increase (~ 80%).

GWAS in the Soay identified two major QTLs for GRR, with apparent sex-specific effects. These two QTLs were located in the same genomic regions as our QTLs on chromosome 6 and chromosome 7. The chromosome 6 QTL was only found significant in Soay females, while we detect it in Lacaune males. Although the QTL was located in the same genomic region, the most significant SNPs were different in the two GWAS (Figure 6). Two possible explanations could be offered for these results: either the two populations have the same QTL segregating and the different GWAS hits correspond to different LD patterns between SNPs and QTLs in the two populations, or the two populations have different causal mutations in the same region. Given that

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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 two populations are both domestic sheeps and were not separated that long ago, we tend to favour the former hypothesis, however denser genotyping data would be needed to have a clear answer. For the chromosome 7 QTL, the signal was only found using regional heritability mapping (Nagamine *et al.* 2012) in the Soay, and after genotype imputation in our study, which makes it even more difficult to discriminate between a shared causal mutation or different causal mutations at the same location in the two populations.

Discussion

In this study, our aims were to produce fine-scale genetic maps of the sheep genome and to elucidate some components of the genetic determinism of recombination rate variation in sheep. In addition, combining three datasets, two pedigree datasets in distantly related domestic sheep populations and a densely genotyped sample of unrelated animals, allowed to partly unravel how recombination rate could evolve at short time scales, as we discuss below.

Fine-scale Recombination Maps

In this work, we were able to construct fine-scale genetic maps of the sheep autosomes by combining two independent inferences on recombination rate.

Our study on meiotic recombination from a large pedigree dataset revealed that sheep recombination exhibits general patterns similar to other mammals (Shifman *et al.* 2006; Chowdhury *et al.* 2009; Tortereau *et al.* 2012). First, sheep recombination rates were elevated at the chromosome ends, both on acrocentric and metacentric chromosomes. In the latter, our analysis revealed a clear reduction in recombination near centromeres. Second, recombination rate depended on the chromosome physical size, consistent with an obligate crossover per meiosis irrespective of the chromosome size.

Our LD-based maps revealed patterns of recombination at the kilobase scale, with small highly recombining intervals interspaced by more wide, low recombining regions. This result was consistent with the presence of recombination hotspots in the highly recombinant intervals. A

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

consequence was that, as observed in other species, the majority of recombination took place in a small portion of the genome: we estimated that 80% of recombination takes place in 40% of the genome. This proportion was likely an underestimate due to our limited resolution (a few kilobases on the HD SNP array) compared to the typical hotspot width (a few hundred base-pairs). Overall, we identified 50,000 hotspot intervals which was twice the estimated number of hotspots in humans (International HapMap Consortium *et al.* 2007). This difference can be explained by different non mutually exclusive reasons. First, it is possible that what we detect as crossover hotspots are due to genome assembly errors and we indeed found a significant albeit moderate effect (OR ≃ 1.4) of the presence of assembly gaps in an interval on its probability of being called a hotspot. Second, our method to call hotspots could be too liberal. Indeed, a more stringent threshold (FDR=0.1%) would lead to about 25,000 hotspots, which would be similar to what is found in humans. Finally, there exists the possibility that historically sheep exhibits more recombination hotspots than humans. In any case, the strong association between meiotic recombination rate and density in LD-based hotspots showed that our LD-based maps were generally accurate. Ultimately better resolution of crossover hotspots should be addressed in the future from LD-based studies on resequencing data.

We combined, using a formal statistical approach, meiotic- and LD-based recombination rate estimates. This led us to assess the impact of selection events on the sheep genome, in particular revealing a new potential signal for diversifying selection at olfactory receptors genes. Based on this comparison, the correlation between LD-based and meiotic estimates was found to be high ($r \simeq 0.7$), but less than could be expected from previous results in humans (Myers *et al.* 2005). Again, more precise estimates of both meiotic- and LD-based recombination rates could change this number but it is also likely that intense selective pressure due to domestication and later artificial breeding had the impact of modifying extensively LD patterns on the sheep genome, degrading the correlation between the two approaches. Indeed, the LD-based recombination estimates summarize ancestral recombinations that took place in the past and it is possible that recombination hotspots that were present in an ancestral sheep population are not longer active in today's Lacaune individuals. This could arise, for example, if domestication led to a reduction in the diversity of hotspots defining genes, such as *PRDM9*. Further studies on the determinism of hotspots in the sheep, its related genetic factors and their diversity would be needed to elucidate this question.

Finally our combined analysis enabled us to scale the LD-based recombination rates and produced fine-scale recombination maps of the HD SNP array. As an illustration of the importance of fine-scale recombination maps for genetic studies, we found an interesting example in a recent

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

study on adaptation of sheep and goat (Kim *et al.* 2016). In this study, a common signal of "selection" was found using the iHS statistic (Voight *et al.* 2006) in these two species (Figure 5 in (Kim *et al.* 2016)). This signature matches precisely the low recombining regions we identified on chromosome 10. Because fine-scale recombination maps were not available at the time, physical distance was used as a proxy for genetic distance in the iHS calculation. In the case of chromosome 10, this proxy failed and extended shared haplotypes, resulting from the low recombination rate, were then interpreted as selection signatures. This kind of false positives could be avoided by using genetic distances from our maps rather than physical distances.

Determinism of Recombination Rate in the Lacaune population

Our approach to study the genetic determinism of recombination rate in the Lacaune population was first to estimate its heritability, using a classical analysis in a large pedigree. This analysis also allowed to extract additive genetic values (EBVs) for the trait in 345 male parents, which we used for a GWAS in a second step. The EBVs are by definition, only determined by genetic factors, as environmental effects on GRR are averaged out. Indeed, we found that the proportion of variance in EBVs explained by genetic factors in the GWAS was essentially one. A consequence was that, although this sample size could be deemed low in current standards, the power of our GWAS was greatly increased by the high precision on the phenotype.

We estimated the heritability of GRR at 0.23, which was similar to estimates from studies on the same phenotype in ruminants (*e.g.* 0.22 in cattle (Sandor *et al.* 2012) or 0.15 in Soay sheep (Johnston *et al.* 2016), but see below for a discussion on the comparison with Soay sheep). We had little information on the environmental factors that could influence recombination rate, but did find a suggestive effect of the month of insemination on GRR, especially we found increased GRR at the month of May. Confirmation and biological interpretation of this result would need dedicated studies, but it was consistent with the fact that fresh (*i.e.* not frozen) semen is used for insemination in sheep and that the reproduction of this species is seasonal (Rosa and Bryant 2003).

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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 genetic determinism of GRR discovered in our study closely resembles what has been found in previous studies, especially in ruminants. Three major loci affected recombination rate in Lacaune, two of them common to cattle and Soay sheep. The underlying genes and mutations for these two QTLs are not yet resolved but the fact that the two regions harbour interacting genes (*RNF212* and *HEI10* (Qiao *et al.* 2014; Rao *et al.* 2016)) involved in the maturation of crossovers, make these two genes likely functional candidates. The third gene identified here, *KCNJ15*, is a novel candidate, and its role and mechanism of action in the repair of DSBs needs to be confirmed and elucidated. Interestingly, these three genes are linked to the reparation of DSBs and crossover maturation processes. This is only one of the stages of recombination which is a complicated process (Baudat *et al.* 2013). In our study, sixty percents of the additive genetic variance in GRR remained unexplained by large effect QTLs and were due to polygenic effects. This could be interpreted in the light of recent evidence that has shown that other mechanisms, involved in chromosome conformation during meiosis, explain a substantial part of the variation in recombination rate between mouse strains (Baier *et al.* 2014) and bovids (Ruiz-Herrera *et al.* 2017). Our results indicate that the genetic determinism of such processes is most likely of polygenic nature. Elucidating this genetic determinism would thus require much larger sample sizes or different experimental approaches (Baier *et al.* 2014; Ruiz-Herrera *et al.* 2017).

Mechanism of Evolution of Recombination Rates in Soay sheep as informed by the Lacaune population

Our study is the second work on the genetic determinism of recombination rate in the sheep. Recently, a thorough study in Soay sheep highlighted how recombination rate variation is genetically determined in this population of feral sheep leaving unmaintained on an island (Johnston *et al.* 2016). By comparing the genetic maps of the two populations, our analysis showed that Soay males have an increase of 50% of recombination compared to Lacaune males. Thus it is clear that Soay sheep have experienced strong selection that had the effect of increasing their recombination rate, especially in males.

One of the most striking difference between our two studies is that the two QTLs that were detected in common had no effect in Soay males, whereas they had strong effects in Lacaune males.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

However, the two populations had very similar polygenic heritability: accounting for the fact that the Lacaune QTLs explain about 40% of the additive genetic variance, we could estimate the polygenic additive genetic variance in Lacaune males at 0.16, very similar to the 0.15 found in Soay males.

Hence, interpreting the Soay results in the light of our study reveals that (i) Soay sheep have experienced selection on increased recombination rate and (ii) that this happened without fixation of large effect QTLs that affect recombination rate. If selection acted on recombination rate through the processes affected by *RNF212* and *HEI10*, namely the maturation of DSBs into crossovers, these two phenomenons would be unlikely to happen: selection should have fixed the increasing alleles at the QTLs. Thus, we can put forward the hypothesis that increased recombination rate in Soay is a consequence of selection on processes that are not genetically linked to maturation of crossovers. This is another piece of evidence that recombination rate is governed by multiple, independent biological processes that have distinct genetic determinisms.

Conclusion

Recombination is a complex biological process that results from the cumulation of several steps leading eventually to the formation of crossovers that can be detected experimentally. Our study exemplifies how a genetical approach can help to separate out these different stages and their genetic determinisms. Further work is needed to get a more detailed picture of the genetic control of recombination and will likely require combining multiple inferences from genetics, cytogenetics, molecular biology and bioinformatics analyses.

Acknowledgments

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Institut de l'Elevage (JM. Astruc) and breed organizations (Ovitest and Confédération de Roquefort) provided the SNP genotypes and pedigree information. This work was partially funded by the BoDeliRe grant of the INRA Selgen Metaprogram and by Région Midi-Pyrénées. We are thankful to Tom Druet, Laurent Duret, Alain Pinton and Pierre Sourdille for their helpful comments on the manuscript.

Literature Cited

Ahlawat S., Sharma P., Sharma R., Arora R., Verma N. K., Brahma B., Mishra P., De S., 2016 Evidence of positive selection and concerted evolution in the rapidly evolving PRDM9 zinc finger domain in goats and sheep. Anim. Genet. **47**: 740–751.

Auton A., Rui Li Y., Kidd J., Oliveira K., Nadel J., Holloway J. K., Hayward J. J., Cohen P. E., Greally J. M., Wang J., Bustamante C. D., Boyko A. R., 2013 Genetic recombination is targeted towards gene promoter regions in dogs. PLoS Genet. **9**: e1003984.

Baier B., Hunt P., Broman K. W., Hassold T., 2014 Variation in genome-wide levels of meiotic recombination is established at the onset of prophase in mammalian males. PLoS Genet. **10**: e1004125.

Bates D., Mächler M., Bolker B., Walker S., 2015 Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. **67**: 1–48.

Baudat F., Imai Y., Massy B. de, 2013 Meiotic recombination in mammals: localization and regulation. Nat. Rev. Genet. **14**: 794–806.

Boitard S., Rodríguez W., Jay F., Mona S., Austerlitz F., 2016 Inferring Population Size History from Large Samples of Genome-Wide Molecular Data - An Approximate Bayesian Computation Approach. PLoS Genet. **12**: e1005877.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Chang C. C., Chow C. C., Tellier L. C., Vattikuti S., Purcell S. M., Lee J. J., 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience **4**: 7.

Chowdhury R., Bois P. R. J., Feingold E., Sherman S. L., Cheung V. G., 2009 Genetic analysis of variation in human meiotic recombination. PLoS Genet. **5**: e1000648.

Cohen-Zinder M., Seroussi E., Larkin D. M., Loor J. J., Everts-van der Wind A., Lee J.-H., Drackley J. K., Band M. R., Hernandez A. G., Shani M., Lewin H. A., Weller J. I., Ron M., 2005 Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res. **15**: 936–944.

Cox A., Ackert-Bicknell C. L., Dumont B. L., Ding Y., Bell J. T., Brockmann G. A., Wergedal J. E., Bult C., Paigen B., Flint J., Tsaih S.-W., Churchill G. A., Broman K. W., 2009 A new standard genetic map for the laboratory mouse. Genetics **182**: 1335–1344.

Davies B., Hatton E., Altemose N., Hussin J. G., Pratto F., Zhang G., Hinch A. G., Moralli D., Biggs D., Diaz R., Preece C., Li R., Bitoun E., Brick K., Green C. M., Camerini-Otero R. D., Myers S. R., Donnelly P., 2016 Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. Nature **530**: 171–176.

Druet T., Georges M., 2015 LINKPHASE3: an improved pedigree-based phasing algorithm robust to genotyping and map errors. Bioinformatics **31**: 1677–1679.

Fariello M.-I., Servin B., Tosser-Klopp G., Rupp R., Moreno C., International Sheep Genomics Consortium, San Cristobal M., Boitard S., 2014 Selection signatures in worldwide sheep populations. PLoS One **9**: e103813.

Groenen M. A. M., Archibald A. L., Uenishi H., Tuggle C. K., Takeuchi Y., Rothschild M. F., Rogel-Gaillard C., Park C., Milan D., Megens H.-J., Li S., Larkin D. M., Kim H., Frantz L. A. F., Caccamo M., Ahn H., Aken B. L., Anselmo A., Anthon C., Auvil L., Badaoui B., Beattie C. W.,

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Bendixen C., Berman D., Blecha F., Blomberg J., Bolund L., Bosse M., Botti S., Bujie Z., Bystrom M., Capitanu B., Carvalho-Silva D., Chardon P., Chen C., Cheng R., Choi S.-H., Chow W., Clark R. C., Clee C., Crooijmans R. P. M. A., Dawson H. D., Dehais P., De Sapio F., Dibbits B., Drou N., Du Z.-Q., Eversole K., Fadista J., Fairley S., Faraut T., Faulkner G. J., Fowler K. E., Fredholm M., Fritz E., Gilbert J. G. R., Giuffra E., Gorodkin J., Griffin D. K., Harrow J. L., Hayward A., Howe K., Hu Z.-L., Humphray S. J., Hunt T., Hornshøj H., Jeon J.-T., Jern P., Jones M., Jurka J., Kanamori H., Kapetanovic R., Kim J., Kim J.-H., Kim K.-W., Kim T.-H., Larson G., Lee K., Lee K.-T., Leggett R., Lewin H. A., Li Y., Liu W., Loveland J. E., Lu Y., Lunney J. K., Ma J., Madsen O., Mann K., Matthews L., McLaren S., Morozumi T., Murtaugh M. P., Narayan J., Nguyen D. T., Ni P., Oh S.-J., Onteru S., Panitz F., Park E.-W., Park H.-S., Pascal G., Paudel Y., Perez-Enciso M., Ramirez-Gonzalez R., Reecy J. M., Rodriguez-Zas S., Rohrer G. A., Rund L., Sang Y., Schachtschneider K., Schraiber J. G., Schwartz J., Scobie L., Scott C., Searle S., Servin B., Southey B. R., Sperber G., Stadler P., Sweedler J. V., Tafer H., Thomsen B., Wali R., Wang J., Wang J., White S., Xu X., Yerle M., Zhang G., Zhang J., Zhang J., Zhao S., Rogers J., Churcher C., Schook L. B., 2012 Analyses of pig genomes provide insight into porcine demography and evolution. Nature **491**: 393–398.

Hassold T., Hall H., Hunt P., 2007 The origin of human aneuploidy: where we have been, where we are going. Hum. Mol. Genet. **16**: R203–R208.

Howie B. N., Donnelly P., Marchini J., 2009 A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. **5**: e1000529.

Ignatieva E. V., Levitsky V. G., Yudin N. S., Moshkin M. P., Kolchanov N. A., 2014 Genetic basis of olfactory cognition: extremely high level of DNA sequence polymorphism in promoter regions of the human olfactory receptor genes revealed using the 1000 Genomes Project dataset. Front. Psychol. **5**: 247.

International HapMap Consortium, Frazer K. A., Ballinger D. G., Cox D. R., Hinds D. A., Stuve L. L., Gibbs

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

R. A., Belmont J. W., Boudreau A., Hardenbol P., Leal S. M., Pasternak S., Wheeler D. A., Willis T. D., Yu F., Yang H., Zeng C., Gao Y., Hu H., Hu W., Li C., Lin W., Liu S., Pan H., Tang X., Wang J., Wang W., Yu J., Zhang B., Zhang Q., Zhao H., Zhao H., Zhou J., Gabriel S. B., Barry R., Blumenstiel B., Camargo A., Defelice M., Faggart M., Goyette M., Gupta S., Moore J., Nguyen H., Onofrio R. C., Parkin M., Roy J., Stahl E., Winchester E., Ziaugra L., Altshuler D., Shen Y., Yao Z., Huang W., Chu X., He Y., Jin L., Liu Y., Shen Y., Sun W., Wang H., Wang Y., Wang Y., Xiong X., Xu L., Waye M. M. Y., Tsui S. K. W., Xue H., Wong J. T.-F., Galver L. M., Fan J.-B., Gunderson K., Murray S. S., Oliphant A. R., Chee M. S., Montpetit A., Chagnon F., Ferretti V., Leboeuf M., Olivier J.-F., Phillips M. S., Roumy S., Sallée C., Verner A., Hudson T. J., Kwok P.-Y., Cai D., Koboldt D. C., Miller R. D., Pawlikowska L., Taillon-Miller P., Xiao M., Tsui L.-C., Mak W., Song Y. Q., Tam P. K. H., Nakamura Y., Kawaguchi T., Kitamoto T., Morizono T., Nagashima A., Ohnishi Y., Sekine A., Tanaka T., Tsunoda T., Deloukas P., Bird C. P., Delgado M., Dermitzakis E. T., Gwilliam R., Hunt S., Morrison J., Powell D., Stranger B. E., Whittaker P., Bentley D. R., Daly M. J., Bakker P. I. W. de, Barrett J., Chretien Y. R., Maller J., McCarroll S., Patterson N., Pe'er I., Price A., Purcell S., Richter D. J., Sabeti P., Saxena R., Schaffner S. F., Sham P. C., Varilly P., Altshuler D., Stein L. D., Krishnan L., Smith A. V., Tello-Ruiz M. K., Thorisson G. A., Chakravarti A., Chen P. E., Cutler D. J., Kashuk C. S., Lin S., Abecasis G. R., Guan W., Li Y., Munro H. M., Qin Z. S., Thomas D. J., McVean G., Auton A., Bottolo L., Cardin N., Eyheramendy S., Freeman C., Marchini J., Myers S., Spencer C., Stephens M., Donnelly P., Cardon L. R., Clarke G., Evans D. M., Morris A. P., Weir B. S., Tsunoda T., Mullikin J. C., Sherry S. T., Feolo M., Skol A., Zhang H., Zeng C., Zhao H., Matsuda I., Fukushima Y., Macer D. R., Suda E., Rotimi C. N., Adebamowo C. A., Ajayi I., Aniagwu T., Marshall P. A., Nkwodimmah C., Royal C. D. M., Leppert M. F., Dixon M., Peiffer A., Qiu R., Kent A., Kato K., Niikawa N., Adewole I. F., Knoppers B. M., Foster M. W., Clayton E. W., Watkin J., Gibbs R. A., Belmont J. W., Muzny D., Nazareth L., Sodergren E., Weinstock G. M., Wheeler D. A., Yakub I., Gabriel S. B., Onofrio R. C., Richter D. J., Ziaugra L., Birren B. W., Daly M. J., Altshuler D., Wilson R. K., Fulton L. L., Rogers J., Burton J., Carter N. P., Clee C. M., Griffiths M., Jones M. C., McLay K., Plumb R. W., Ross M. T.,

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Sims S. K., Willey D. L., Chen Z., Han H., Kang L., Godbout M., Wallenburg J. C., L'Archevêque P., Bellemare G., Saeki K., Wang H., An D., Fu H., Li Q., Wang Z., Wang R., Holden A. L., Brooks L. D., McEwen J. E., Guyer M. S., Wang V. O., Peterson J. L., Shi M., Spiegel J., Sung L. M., Zacharia L. F., Collins F. S., Kennedy K., Jamieson R., Stewart J., 2007 A second generation human haplotype map of over 3.1 million SNPs. Nature **449**: 851–861.

Jiang Y., Xie M., Chen W., Talbot R., Maddox J. F., Faraut T., Wu C., Muzny D. M., Li Y., Zhang W., Stanton J.-A., Brauning R., Barris W. C., Hourlier T., Aken B. L., Searle S. M. J., Adelson D. L., Bian C., Cam G. R., Chen Y., Cheng S., DeSilva U., Dixen K., Dong Y., Fan G., Franklin I. R., Fu S., Fuentes-Utrilla P., Guan R., Highland M. A., Holder M. E., Huang G., Ingham A. B., Jhangiani S. N., Kalra D., Kovar C. L., Lee S. L., Liu W., Liu X., Lu C., Lv T., Mathew T., McWilliam S., Menzies M., Pan S., Robelin D., Servin B., Townley D., Wang W., Wei B., White S. N., Yang X., Ye C., Yue Y., Zeng P., Zhou Q., Hansen J. B., Kristiansen K., Gibbs R. A., Flicek P., Warkup C. C., Jones H. E., Oddy V. H., Nicholas F. W., McEwan J. C., Kijas J. W., Wang J., Worley K. C., Archibald A. L., Cockett N., Xu X., Wang W., Dalrymple B. P., 2014 The sheep genome illuminates biology of the rumen and lipid metabolism. Science **344**: 1168–1173.

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

Johnston S. E., Bérénos C., Slate J., Pemberton J. M., 2016 Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (Ovis aries). Genetics **203**: 583–598.

Kadri N. K., Harland C., Faux P., Cambisano N., Karim L., Coppieters W., Fritz S., Mullaart E., Baurain D., Boichard D., Spelman R., Charlier C., Georges M., Druet T., 2016 Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle. Genome Res. **26**: 1323–1332.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

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

Kim E.-S., Elbeltagy A. R., Aboul-Naga A. M., Rischkowsky B., Sayre B., Mwacharo J. M., Rothschild M. F., 2016 Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity **116**: 255–264.

Kong A., Thorleifsson G., Stefansson H., Masson G., Helgason A., Gudbjartsson D. F., Jonsdottir G. M., Gudjonsson S. A., Sverrisson S., Thorlacius T., Jonasdottir A., Hardarson G. A., Palsson S. T., Frigge M. L., Gulcher J. R., Thorsteinsdottir U., Stefansson K., 2008 Sequence Variants in the RNF212 Gene Associate with Genome-Wide Recombination Rate. Science **319**: 1398–1401.

Lange J., Yamada S., Tischfield S. E., Pan J., Kim S., Zhu X., Socci N. D., Jasin M., Keeney S., 2016 The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair. Cell **167**: 695–708.e16.

Li N., Stephens M., 2003 Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics **165**: 2213–2233.

Li H., Durbin R., 2011 Inference of human population history from individual whole-genome sequences. Nature **475**: 493–496.

Ma L., O'Connell J. R., VanRaden P. M., Shen B., Padhi A., Sun C., Bickhart D. M., Cole J. B., Null D. J., Liu G. E., Da Y., Wiggans G. R., 2015 Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. PLoS Genet. **11**: e1005387.

Myers S., Bottolo L., Freeman C., McVean G., Donnelly P., 2005 A fine-scale map of recombination rates and hotspots across the human genome. Science **310**: 321–324.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Nagamine Y., Pong-Wong R., Navarro P., Vitart V., Hayward C., Rudan I., Campbell H., Wilson J., Wild S., Hicks A. A., Pramstaller P. P., Hastie N., Wright A. F., Haley C. S., 2012 Localising Loci underlying Complex Trait Variation Using Regional Genomic Relationship Mapping. PLoS One **7**: e46501.

Norris B. J., Whan V. A., 2008 A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. Genome Res. **18**: 1282–1293.

Pratto F., Brick K., Khil P., Smagulova F., Petukhova G. V., Daniel Camerini-Otero R., 2014 Recombination initiation maps of individual human genomes. Science **346**: 1256442.

Qiao H., Prasada Rao H. B. D., Yang Y., Fong J. H., Cloutier J. M., Deacon D. C., Nagel K. E., Swartz R. K., Strong E., Holloway J. K., Cohen P. E., Schimenti J., Ward J., Hunter N., 2014 Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination. Nat. Genet. **46**: 194–199.

Rao H. B. D. P., Qiao H., Bhatt S. K., Bailey L. R. J., Tran H. D., Bourne S. L., Qiu W., Deshpande A., Sharma A. N., Beebout C. J., Pezza R. J., Hunter N., 2016 A SUMO-Ubiquitin Relay Recruits Proteasomes to Chromosome Axes to Regulate Meiotic Recombination. bioRxiv.

Reynolds A., Qiao H., Yang Y., Chen J. K., Jackson N., Biswas K., Holloway J. K., Baudat F., Massy B. de, Wang J., Höög C., Cohen P. E., Hunter N., 2013 RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. Nat. Genet. **45**: 269–278.

Rochus C. M., Tortereau F., Plisson-Petit F., Restoux G., Moreno-Romieux C., Tosser-Klopp G., Servin B., 2017 High density genome scan for selection signatures in French sheep reveals allelic heterogeneity and introgression at adaptive loci. bioRxiv.

Rosa H. J. D., Bryant M. J., 2003 Seasonality of reproduction in sheep. Small Rumin. Res. **48**: 155–171.

Ruiz-Herrera A., Vozdova M., Fernández J., Sebestova H., Capilla L., Frohlich J., Vara C., Hernández-Marsal A., Sipek J., Robinson T. J., Rubes J., 2017 Recombination correlates with

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

synaptonemal complex length and chromatin loop size in bovids-insights into mammalian meiotic chromosomal organization. Chromosoma.

Sabeti P. C., Reich D. E., Higgins J. M., Levine H. Z. P., Richter D. J., Schaffner S. F., Gabriel S. B., Platko J. V., Patterson N. J., McDonald G. J., Ackerman H. C., Campbell S. J., Altshuler D., Cooper R., Kwiatkowski D., Ward R., Lander E. S., 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature **419**: 832–837.

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

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

Słabicki M., Theis M., Krastev D. B., Samsonov S., Mundwiller E., Junqueira M., Paszkowski-Rogacz M., Teyra J., Heninger A.-K., Poser I., Prieur F., Truchetto J., Confavreux C., Marelli C., Durr A., Camdessanche J. P., Brice A., Shevchenko A., Pisabarro M. T., Stevanin G., Buchholz F., 2010 A genome-scale DNA repair RNAi screen identifies SPG48 as a novel gene associated with hereditary spastic paraplegia. PLoS Biol. **8**: e1000408.

Stathopoulos S., Bishop J. M., O'Ryan C., 2014 Genetic signatures for enhanced olfaction in the African mole-rats. PLoS One **9**: e93336.

Storey J. D., Tibshirani R., 2003 Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences **100**: 9440–9445.

Sturtevant A. H., 1913 The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zool. **14**: 43–59.

Takasuga A., 2015 PLAG1 and NCAPG-LCORL in livestock. Anim. Sci. J.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Tortereau F., Servin B., Frantz L., Megens H.-J., Milan D., Rohrer G., Wiedmann R., Beever J., Archibald A. L., Schook L. B., Groenen M. A. M., 2012 A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. BMC Genomics **13**: 586.

Voight B. F., Kudaravalli S., Wen X., Pritchard J. K., 2006 A map of recent positive selection in the human genome. PLoS Biol. **4**: e72.

Zhou X., Carbonetto P., Stephens M., 2013 Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet. **9**: e1003264.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Tables

Chromosome	Window span (Mb)	p-value	ρ^s/c
3	109-110	$6.4\ 10^{-5}$	0.31
6	35-38	$2.3\ 10^{-7}$	0.22
10	29-30	$1.0\ 10^{-4}$	0.32
10	36-37	$2.7\ 10^{-7}$	0.22
10	42-44	$3.8\ 10^{-12}$	0.13
13	63-64	$9.3\ 10^{-5}$	0.32
10	72-73	$6.0\ 10^{-10}$	6.2
12	4-5	$7.4\ 10^{-6}$	3.8
20	28-29	$3.3\ 10^{-6}$	3.9
23	10-11	$7.7\ 10^{-7}$	4.3

Table 1. Genome regions were meiotic and population-based recombination rates differ significantly. ρ^s: population-based recombination rate, c: meiotic recombination rate. ρ^s/c: ratio of population to meiotic recombination rate. Details on the estimation of these parameters are given in the text. Regions with p-values $\leq 10^{-4}$ were considered outliers (FDR = 0.02).

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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. Decomposition of the inter-individual variation in Genome-wide Recombination Rate.

Phenotype	Number of sires	Genetic Variance	Phenotypic Variance	Heritability
GRR	345	6.86 (0.75)	29.73 (0.84)	0.23 (0.02)

Genetic variance, phenotypic variance and heritability of GRR among 345 male individuals. Figures in brackets are standard errors.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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 3 : SNPs associated with GRR.

SNP name (n° rs)	Chromosome	Position (pb)	Minor allele	Frequency	β effect	Pvalue	pQTL
rs419107626	1	268604660	T	0.24	-2.94	$2.83\ 10^{-6}$	0.18
rs411987057	6	116517201	C	0.22	-2.44	$1.12\ 10^{-18}$	0.31
rs401206888	6	116525109	T	0.22	-2.44	$1.08\ 10^{-18}$	0.29
rs412583165	6	116525709	C	0.25	-2.53	$8.96\ 10^{-19}$	0.24
rs430346236	7	22730150	C	0.22	4.14	$2.18\ 10^{-8}$	0.44
rs413147562	7	22798236	A	0.17	2.17	$5.86\ 10^{-8}$	0.51

3 SNPs, whose the most significant, from chromosome 6 are highly associated with GRR. Additional loci were found on chromosome 1 and 7. P-values are given for a Wald test of an animal model with SNP genotype fitted as a fixed effect. β corresponds to the effect of SNP on GRR and pQTL corresponds to the probability for the SNP to be a QTL.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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 4 : Detected mutations in the RNF212 gene.

Mutation name	Bases change	Positions on v3.1 Sheep genome	Positions on *OA Musimon* genome	Rebuilt positions on v3.1 Sheep genome	Frequency	β effect	Pvalue	pQTL
RNF212_14431_AG	A>G	contig Un_JH922970:5925	Scaffold01089:132229	6:11678613	0.14	-2.15	$5.77\ 10^{-16}$	0.34
RNF212_18411_GA	G>A	6:116438363	Scaffold01089:136209	6:11682593	0.14	-2.16	$2.76\ 10^{-15}$	0.29
RNF212_22570_CG	C>G	6:116442942	Scaffold01089:140368	6:11686752	0.13	-2.24	$2.54\ 10^{-15}$	0.33
RNF212_22594_AG	A>G	6:116442966	Scaffold01089:140392	6:11686776	0.12	2.38	0.443	0.18

Detected polymorphisms in the RNF212 gene, after sequencing of key animals. Then, their association with the phenotype was tested through a GWAS on the 266 animals for which the mutations were genotyped. The positions on v3.1 Sheep genome correspond to the real mutations positions on the genome before the integration of the *OA Musimon* scaffold, the first mutation is the only one which is localized in a non assembly contig. On the other hand, rebuilt positions correspond to the supposed mutations positions after the integration of the scaffold in the Sheep genome. As for the Table 3, p-values are given for a Wald test, β corresponds to the effect of the mutation on GRR and pQTL corresponds to the probability for the mutation to be a QTL.

bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Figures

Figure 1 : Families used to infer crossovers events. Crossover events were identified in meioses of 345 focal individuals (FIDs). 281 FID had their father known (left) while 64 FID did not (right).

Figure 2 : Comparison between population-based recombination rate and meiotic recombination rate for two 1 Mb windows on Sheep chromosome 24. Top: meiotic recombination rate along chromosome 24. Two windows with high (left,red) and low (right,blue) meiotic recombination rates estimates are zoomed in. Each panel represents, from top to bottom: meiotic recombination rate estimates (c) in SNP intervals of the 50K SNP array, population-based recombination rate estimates (ρ) in SNP intervals of the 50K SNP array and population-based recombination rate estimates (ρ) in SNP intervals of the HD (~600K) SNP array.

Figure 3: Population-based and meiotic recombination rates in windows of one megabase. The dashed line is the regression for population recombination rate on the family recombination rate. Values are shown on a logarithmic scale.

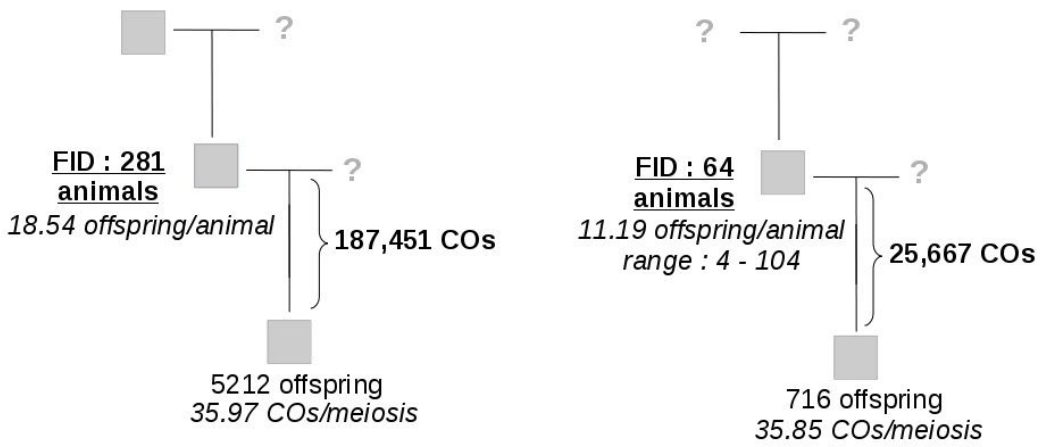
Figure 4: Genome-wide association study identifies three main QTLs for GRR. Top: -log10 (p-value) for single SNP tests for association. The genome-wide significance level (FDR=5%) is represented by the horizontal dotted line. Bottom: posterior probability that a region of 20 SNPs harbors a QTL, using a Bayesian multi-QTL model.

Figure 5 : Recombination intensity in Soay males compared to Lacaune males, for each chromosome. For each chromosome a linear regression of the Lacaune (male) recombination rate on the Soay (male) recombination rate was fitted. Left plot shows the resulting slopes for each chromosome. The solid vertical line correspond to a slope of 1 (identical recombination rates). The dashed line to the median across chromosomes. Right plot shows the resulting slopes as a function of the average recombination rate in Lacaune.

Figure 6 : Comparison of GWAS results for the OAR 6 QTL between Soay (top) and Lacaune (bottom). Top : -log10(p-value) for single SNP test association for the sex-averaged Soay. Bottom : the same representation for Lacaune males. The star point corresponds to the most associated SNP on Soay, the square point corresponds to the most associated SNP on Lacaune. The vertical dark line highlights the position of the *RNF212* gene.

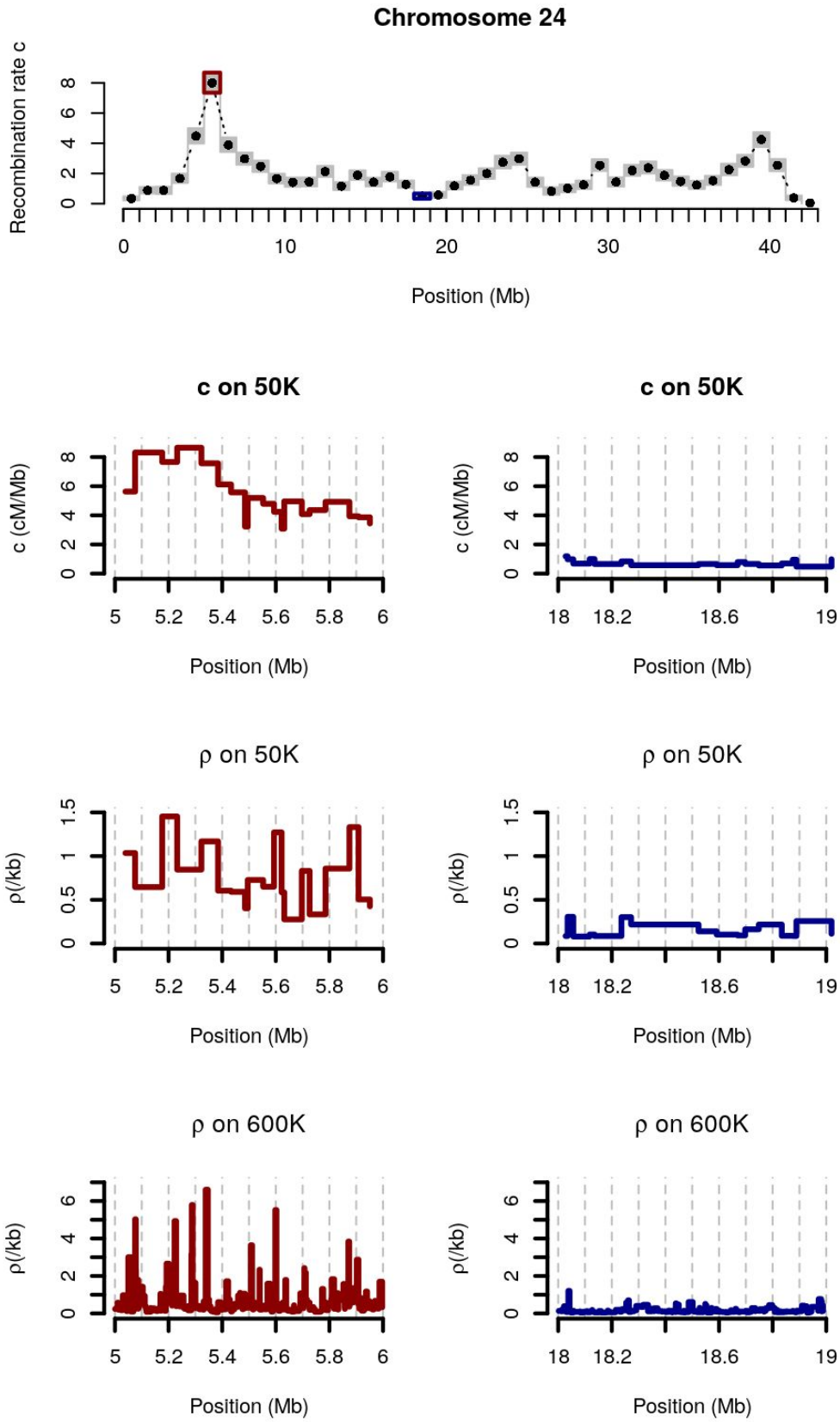
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 1.



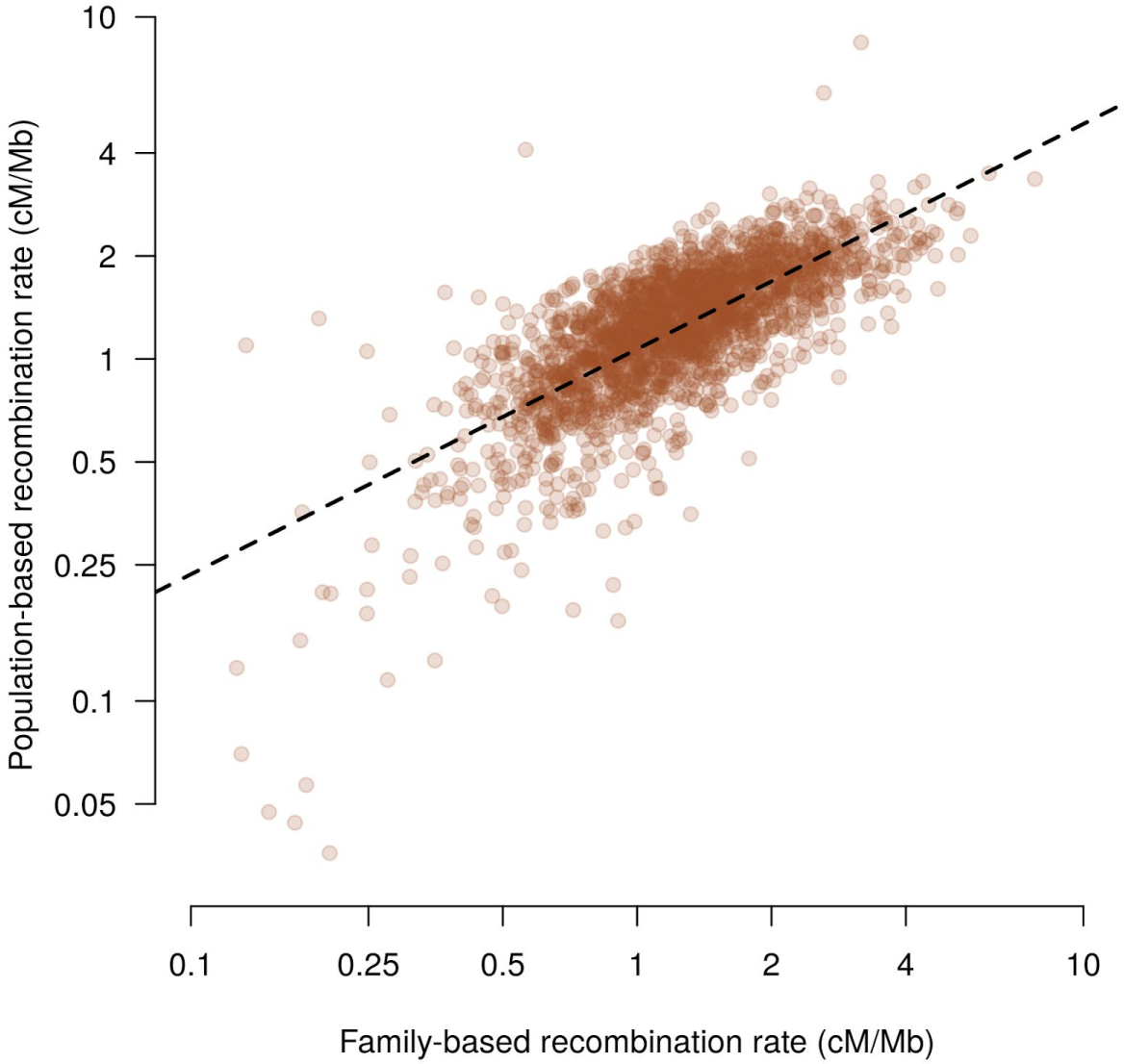
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 2.



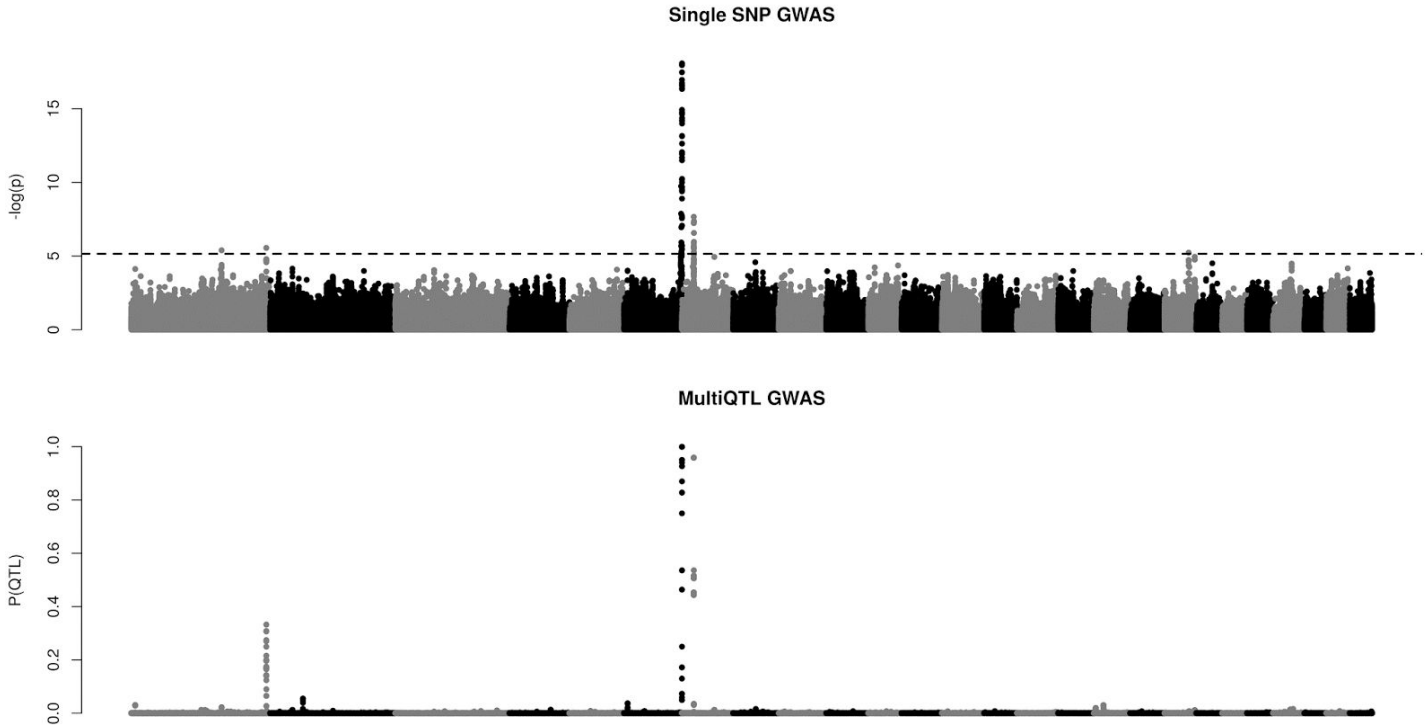
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 3.



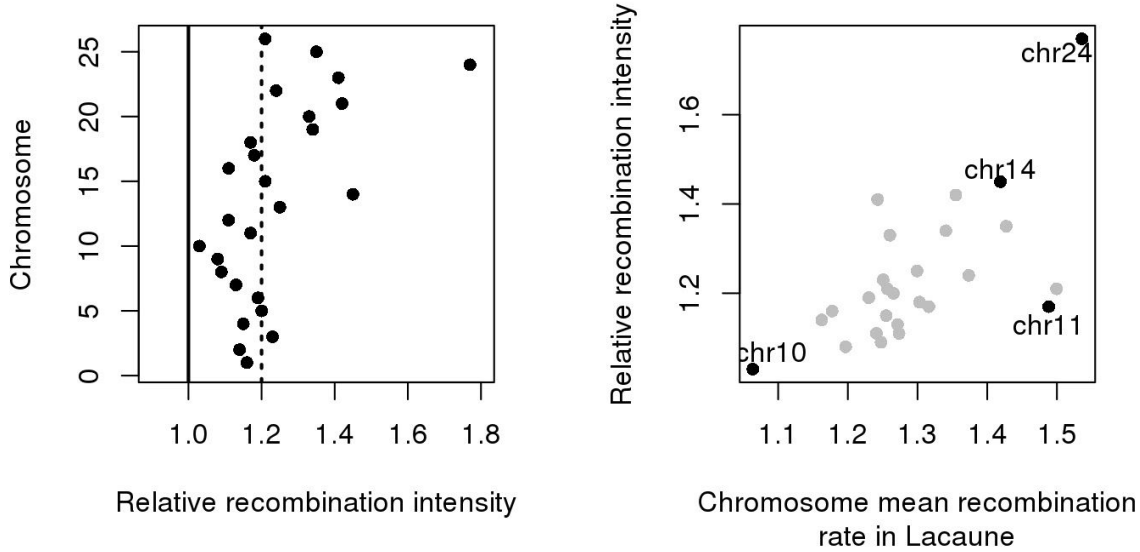
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 4.



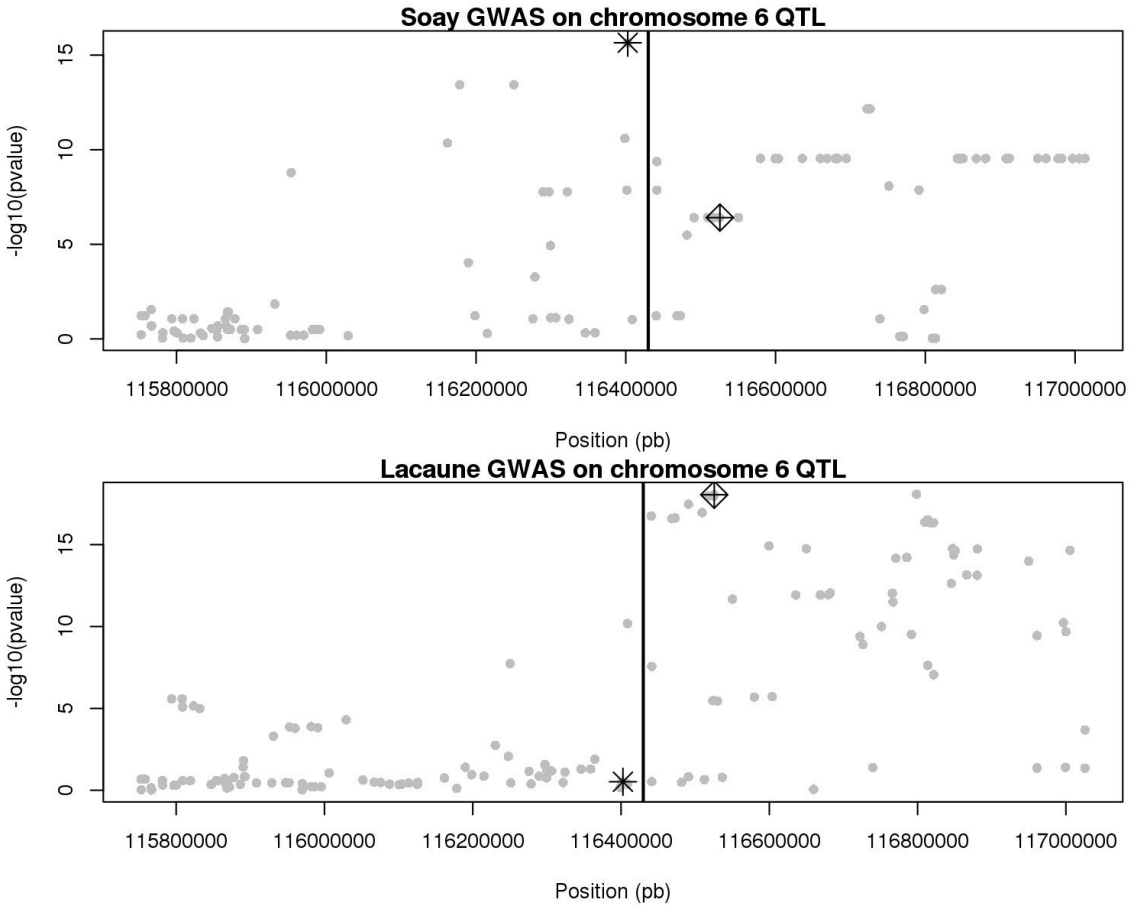
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 5.



bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Figure 6.



bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Supplementary Figures

Figure S1. Patterns of recombination along Sheep autosomes. Left: recombination rate of one megabase windows along metacentric chromosomes (1,2,3). Center: recombination rate of one megabase windows along acrocentric chromosomes (4-26). Right: recombination rate of one megabase windows against distance to nearest chromosome end.

Figure S2. Recombination rates of Sheep autosomes. Left: from recombination rate estimates in windows of one megabase. Right: from recombination rate estimates in SNP array intervals. Top: for each chromosome. Bottom: as a function of chromosome physical size. Dotted line: c = f(log(size)), dashed line: c = f(1/size)**.**

Figure S3. Genomewide distribution of recombination rate using the approach of Sandor et al. (2012). These estimates were calculated from observed crossover frequencies. Fitting a gamma distribution on the observed estimates (red line) provided a prior distribution for the subsequent Bayesian inference on recombination rates (parameters given in the box). See methods for details.

Figure S4. Distribution of recombination intensities among intervals of the HD SNP array. The green curve represents the distribution under the null hypothesis that there is no hotspots ($\log10(\lambda_i) = 0$). It is estimated by fitting a mixture of Gaussian distribution to the observed distribution and extracting the relevant component. Intervals where recombination intensity was particularly high (FDR = 5%) were considered as harbouring recombination hotspots and are shown in red.

Figure S5 : Distribution of recombination on the genome. The figure represents the proportion of the physical genome size affected by recombination, for increasing coverage of the genetic map. 60% of recombination events occur in only 20% of the genome (in green) and 80% of recombination events occur in about 40% of the genome (in red).

Figure S6 : Validation of imputed genotypes for the GWAS. The figure shows the proportion of correct genotype calls as a function of their posterior probability calculated with BIMBAM

Figure S7: Local alignments of the Sheep and Human genome around the OAR7 QTL region. Dotplot of the alignements of sheep OAR7 on human HSA14. Vertical cyan bars are located at significant SNP positions. Three functional candidate genes surrounding the association signal (shaded) are indicated.

Figure S8: Relative intensity of population to meiotic recombination rates in windows of 1 Mb along the sheep genome.

Figure S9: Individual variation in recombination rates among Lacaune Males. Additive genetic values on Genome-wide Recombination Rate genetic for all Lacaune sires of our dataset (in black) and for the 345 FID (in grey). The vertical black line is placed at the mean.

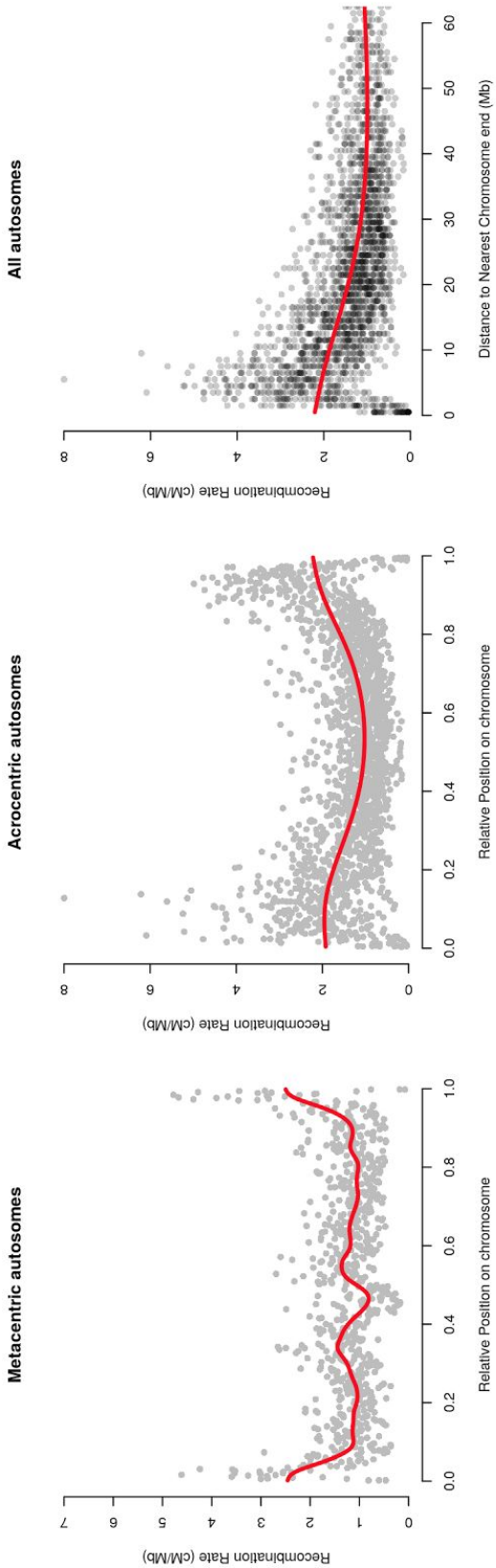
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S10 : Linkage disequilibrium between *RNF212* polymorphisms and chromosome 6 QTL SNPs. The top figure represents the mRNA and the protein of *RNF212*. The four genotyped mutations are indicated : the two first are intronic and the two others are exonic. We replace the gene on a zoom on the chromosome 6 QTL (middle figure). The four left solid lines highlight the mutations, whereas the dashed lines represent the 3 most significant SNPs. Middle points show the intermediate SNPs between the mutations and the significant SNPs. Finally, the figure at the bottom indicates the pairwise LD between the mutations and all the SNPs presented on the middle figure. It highlights two haplotype blocks : one between the 3 most significant mutations and another between the 3 most significant SNPs.

Figure S11 : *RNF212* gene structure in various species. *RNF212* gene is not annotated on the ovine reference genome Oar_v3.1, but can be located at the telomeric end of OAR6 (116,4Mb) by homologies (dashed lines) with the *RNF212* gene from Ovis aries musinon (Oor1_1.0). Some anotated predicted non-coding RNA sequence (nc_RNA in brown) were part of the *RNF212* sequence. The ovine *RNF212* gene is also partly present in the unplaced scaffold005259, that can be virtually located in the largest assembly gap (in blue). In Ovis orientalis musimon, the *RNF212* gene exibited 14 putative exons with alternative splicing (mRNA models in green). Homology analysis (dashed lines) with annotated *RNF212* gene in other ruminant species (bovine on BTA6 and caprine on CHI6) indicated a good gene structure conservation between ovine and bovine *RNF212*, but only a partial conservation with goat *RNF212*, where the six last predicted exons match with intronic region in ovine. When compared to human *RNF212* on HSA4 and mouse *RNF212* on MMU5 chromosomes, only four to five exons are conserved with ruminants indicated a non-conserved gene structure. Red lines located SNP associated with global recombination rate (GRR) in the present study, and those previously shown in bovine (Sandor et al. 2012; Kadri et al. 2016) , in human (Kong et al. 2008; Chowdhury et al. 2009; Fledel-Alon et al. 2011; Kong et al. 2014) and in mouse (Fujiwara et al. 2014). Gene scales are in base pair and gene structures were constructed with CLC Main Workbench software v7.7.3 using the NCBI query module (Qiagen Aarhus).

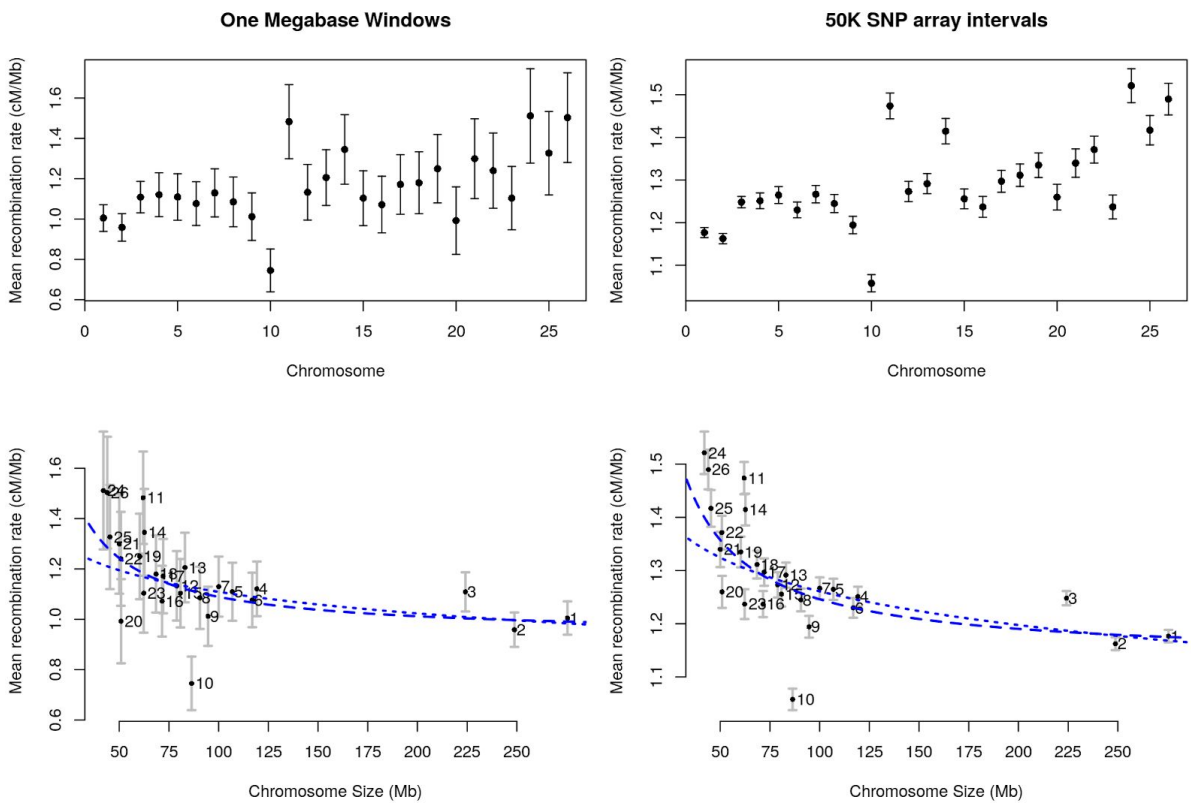
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S1



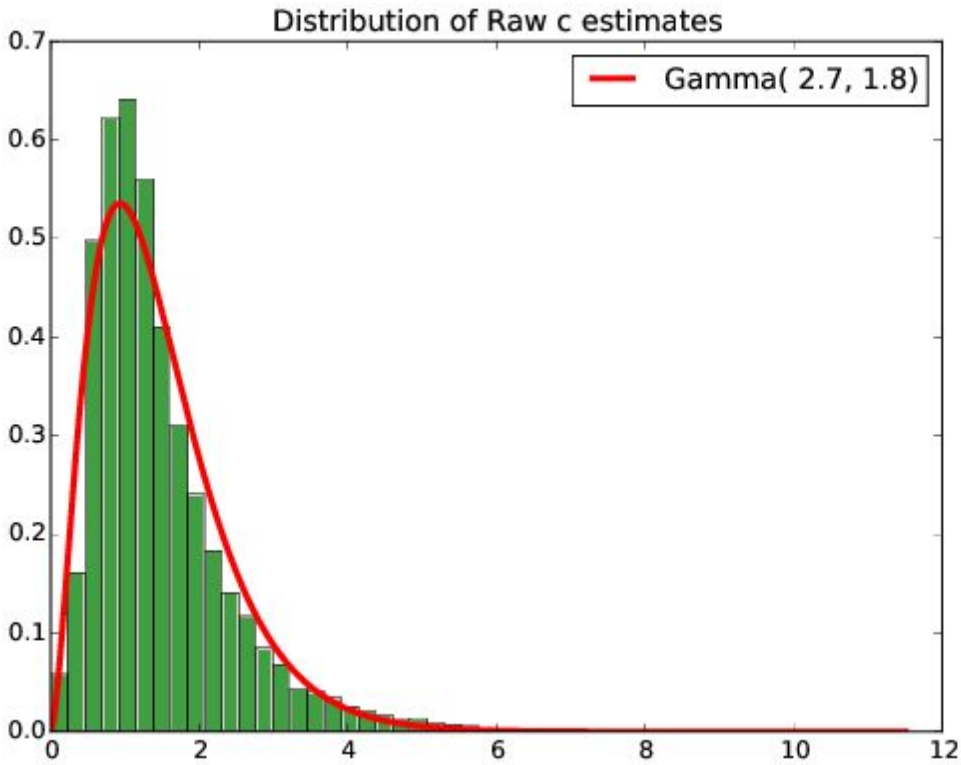
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S2.



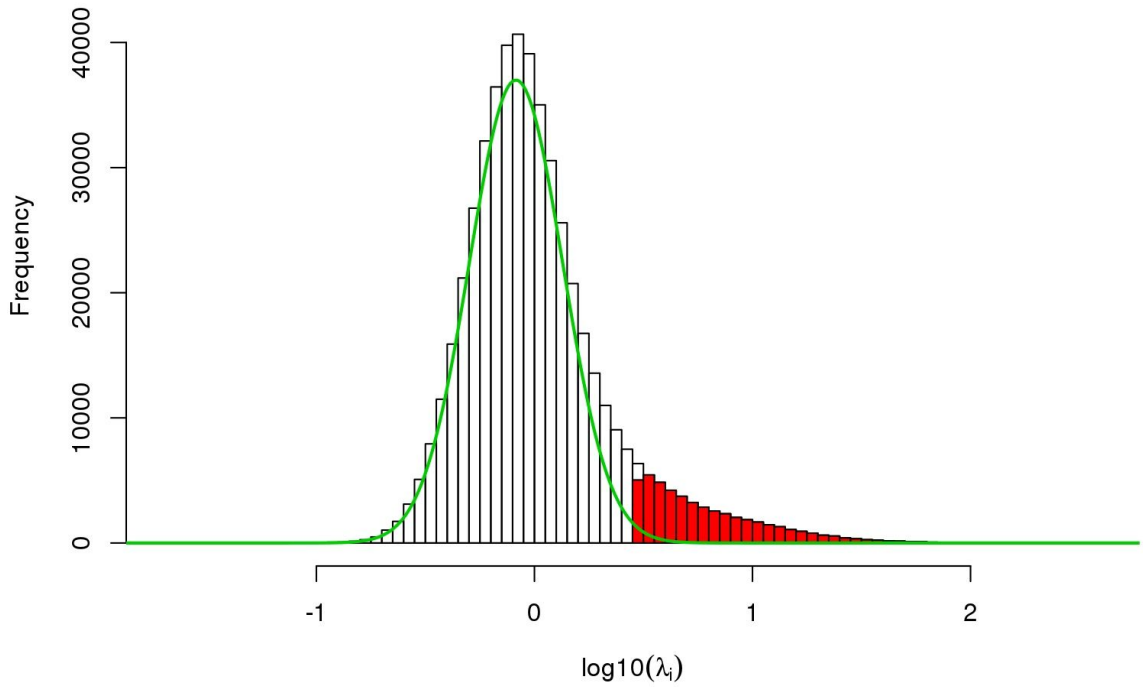
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S3.



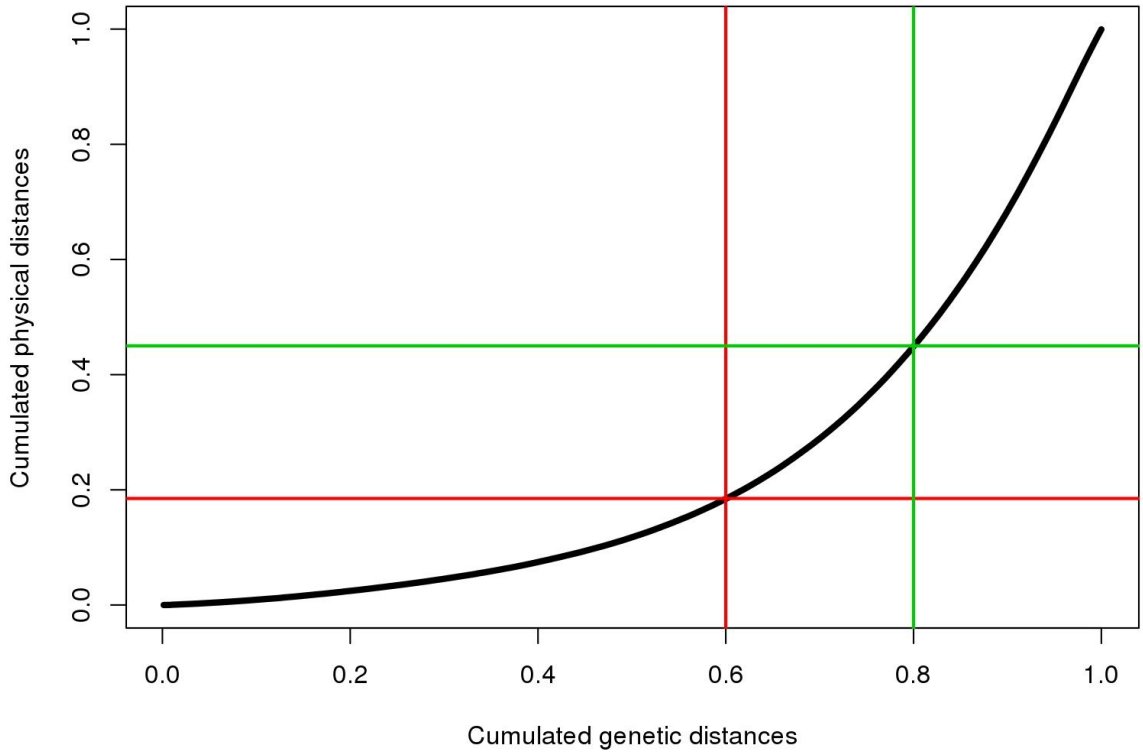
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S4.



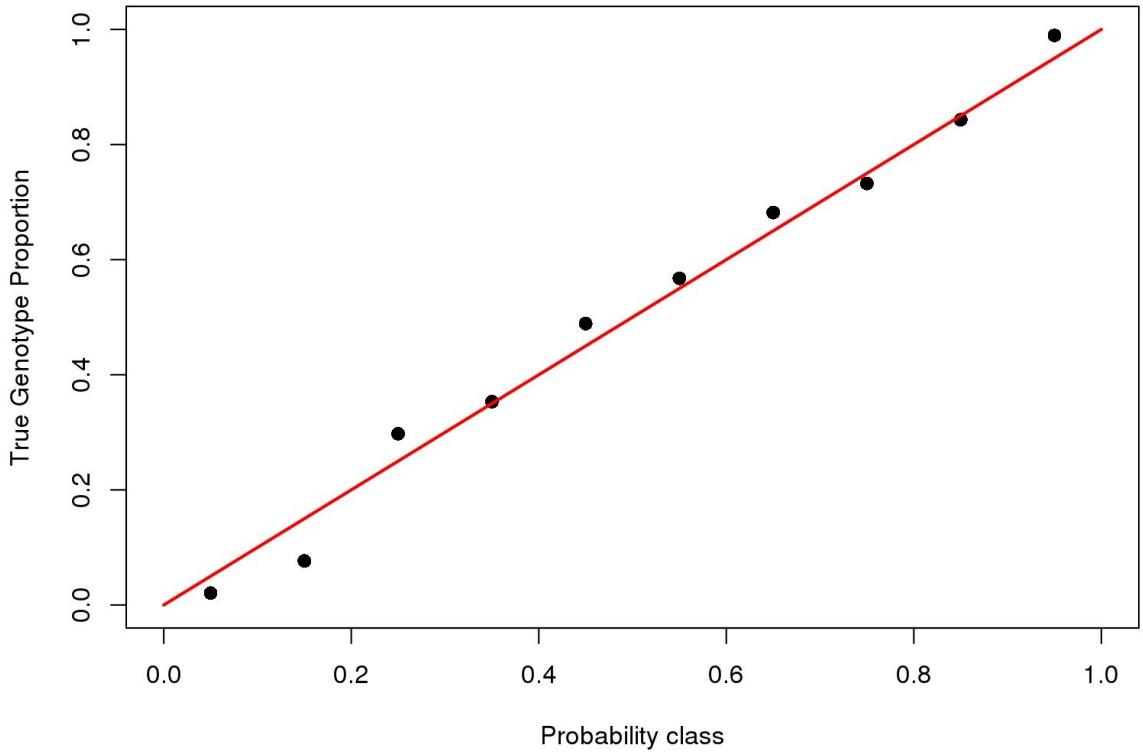
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S5.



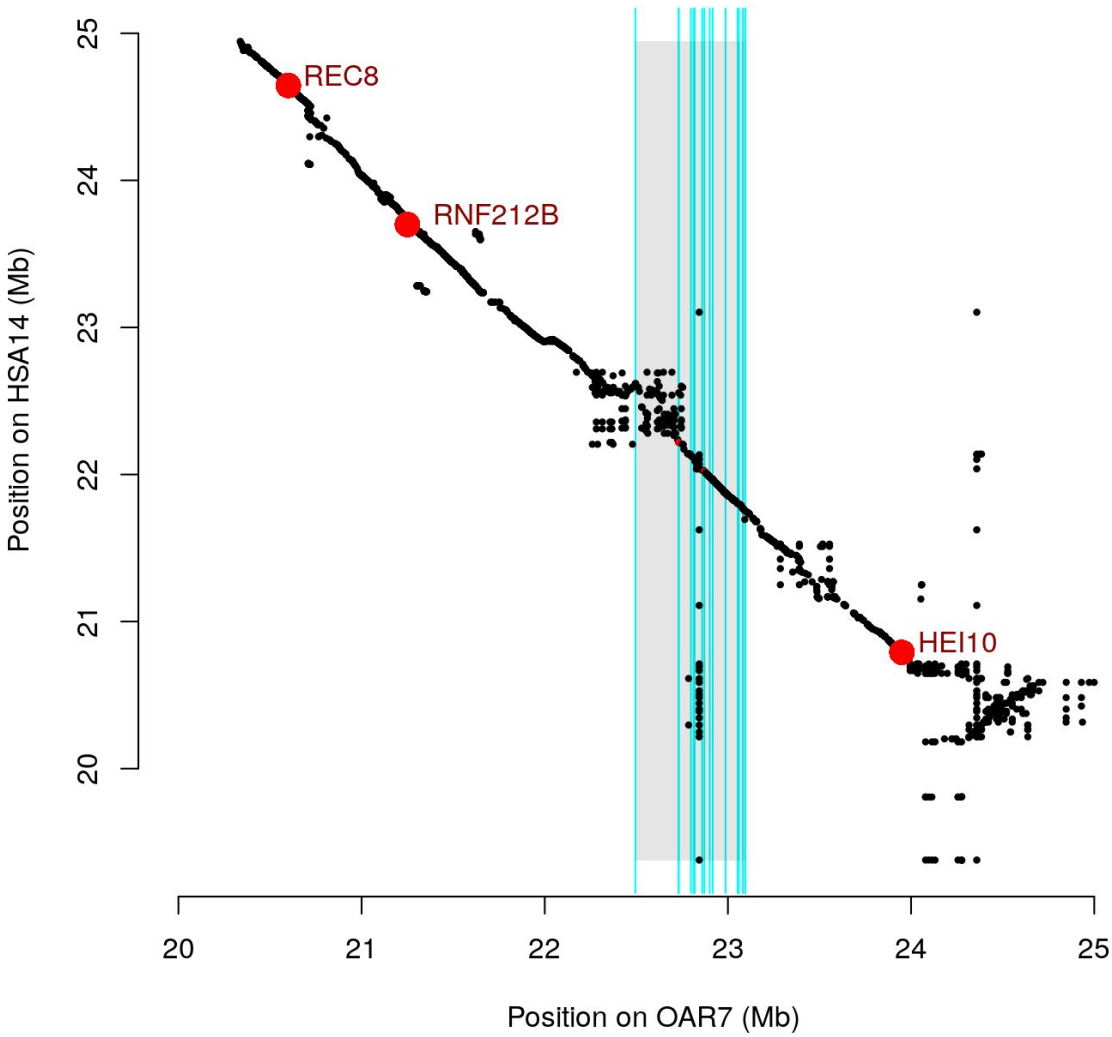
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S6.



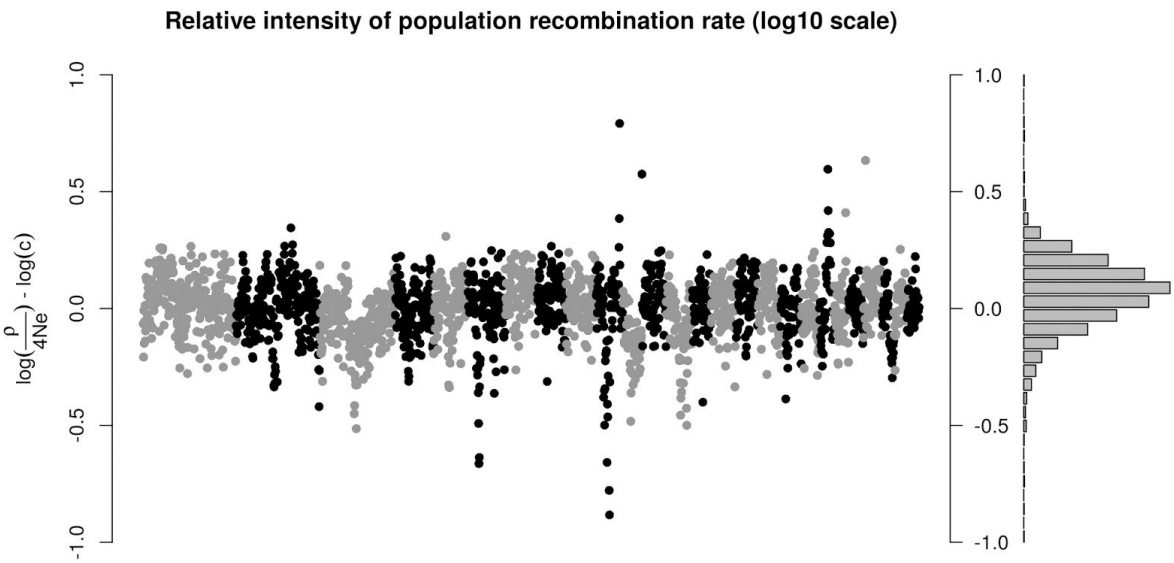
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S7.



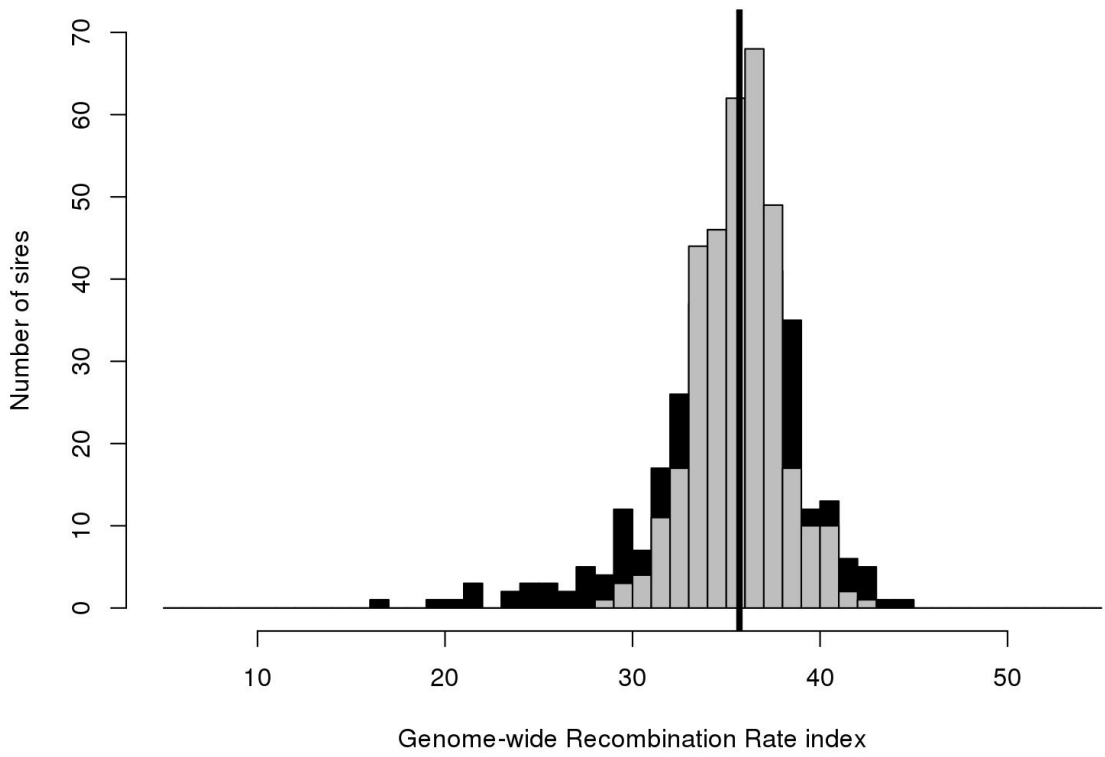
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S8.



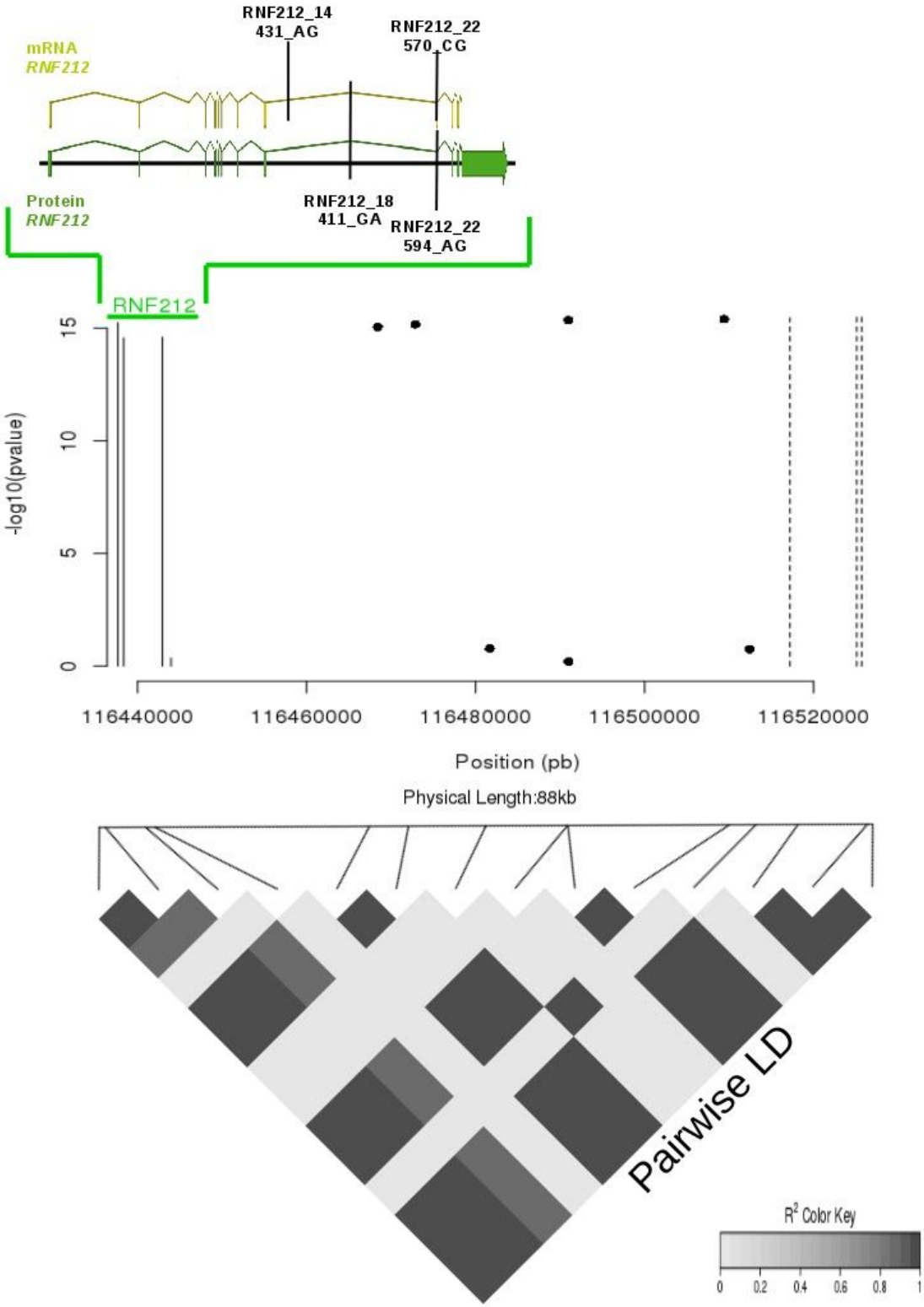
bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S9.



bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

Figure S10.



bioRxiv preprint doi: https://doi.org/10.1101/104976; this version posted February 2, 2017. 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.

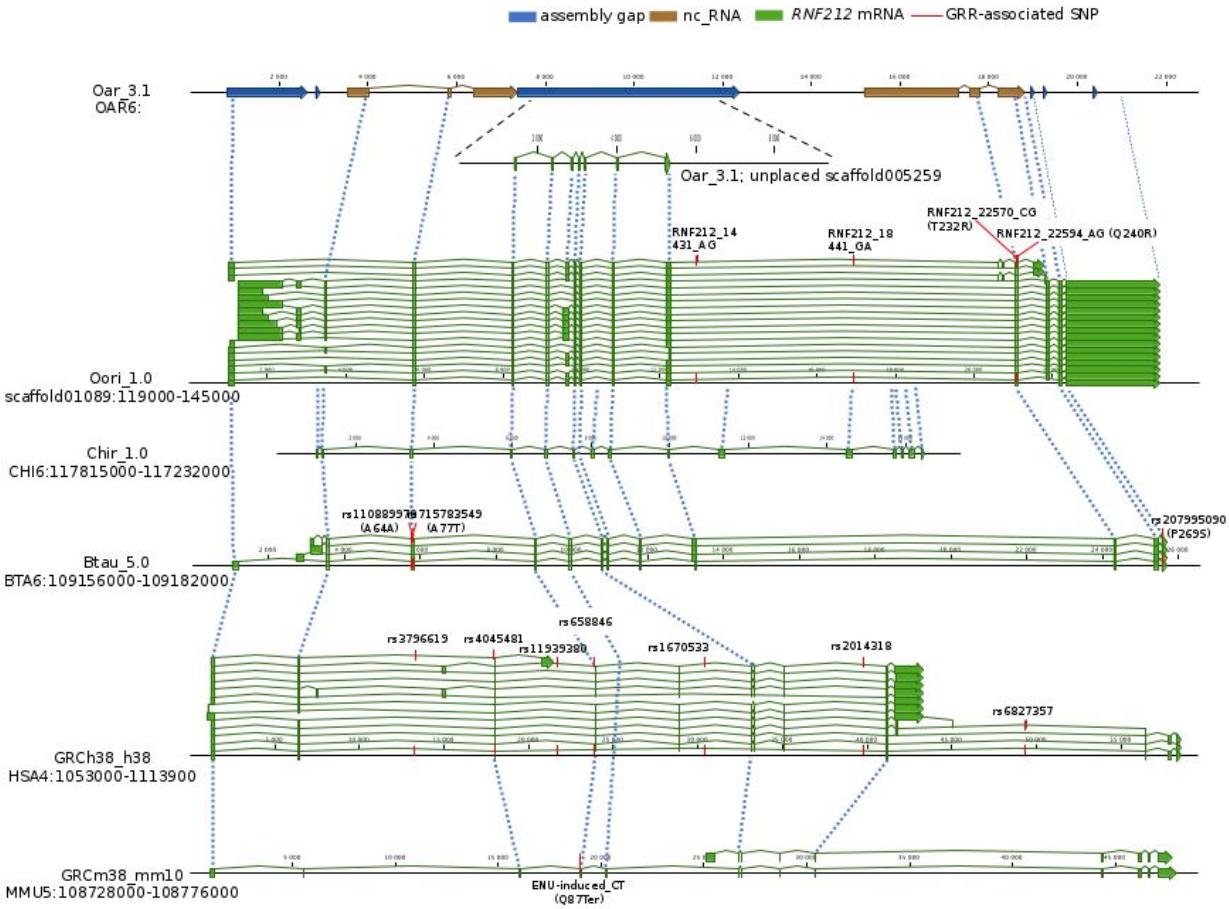



Figure S11.