Genetic architecture of individual variation in recombination rate on the X-chromosome in cattle

Junjie Zhang[1], Naveen Kumar Kadri[1,2], Erik Mullaart[3], Richard Spelman[4], Sébastien Fritz[5,6], Didier Boichard[6], Carole Charlier[1], Michel Georges[1] and Tom Druet[1]

[1]Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège, Liège, Belgium

[2]Animal Genomics, Institute of Agricultural Science, ETH Zürich, Zürich, Switzerland

[3]CRV Innovation Department, Arnhem, The Netherlands

[4]Livestock Improvement Corporation, Hamilton, New Zealand

[5]Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas, France

[6]Allice, Paris, France

Corresponding author: Tom Druet / tom.druet@uliege.be

Abstract

Meiotic recombination is an essential biological process that ensures proper chromosome segregation and creates genetic diversity. Individual variation in global recombination rates has been shown to be heritable in several species and variants significantly associated with this trait have been identified. Recombination on the sex chromosome has often been ignored in these studies although this trait may be particularly interesting as it may correspond to a biological process distinct from that on autosomes. For instance, recombination in males is restricted to the pseudo-autosomal region (PAR). We herein used a large cattle pedigree with more than 100,000 genotyped animals to improve the genetic map of the X-chromosome and to study the genetic architecture of individual variation in recombination rate on the sex-chromosome (XRR). The length of the genetic map was 46.4 and 121.2 cM in males and females, respectively, but the recombination rate in the PAR was 6 times higher in males. The heritability of CO counts on the X-chromosome was comparable to that of autosomes in males (0.011) but larger than that of autosomes in females (0.024). XRR was highly correlated (0.76) with global recombination rate (GRR) in females, suggesting that both traits might be governed by shared variants. In agreement, a set of eleven previously identified variants associated with GRR had correlated effects on female XRR (0.86). In males, XRR and GRR appeared to be distinct traits, although more accurate CO counts on the PAR would be valuable to confirm these results.

Introduction

Genetic maps are essential tools with many applications in quantitative and population genetics in humans, animals and plants. They are for instance essential to study complex traits through QTL mapping (Mackay et al. 2009), to model transmission of identity-by-descent (IBD) segments or to perform linkage-based imputation in genotyped pedigrees (Lander and Green 1987; Abecasis et al. 2002; Druet and Farnir 2011). Genetic distances are also used in models to detect signatures of selection (Grossman et al. 2010) or to characterize inbreeding with hidden Markov models (Leutenegger et al. 2003; Druet and Gautier 2017). They are particularly useful to interpret the length of homozygous-by-descent or IBD segments or to estimate the age of alleles (Albers and McVean 2020). Linkage maps can additionally be used to study the recombination process and the genetic architecture of individual recombination rate (RR). As an example, global recombination rate (GRR) on autosomes has been investigated in several species including cattle (Kong et al. 2008; Chowdhury et al. 2009; Sandor et al. 2012; Ma et al. 2015; Johnston et al. 2016; Kadri et al. 2016; Petit et al. 2017). Several studies showed that individual GRR was heritable in humans, cattle and sheep (Fledel-Alon et al. 2011; Sandor et al. 2012; Johnston et al. 2016) and controlled by variants explaining a large fraction of the genetic variance (Kong et al. 2014; Johnston et al. 2016; Kadri et al. 2016; Petit et al. 2017). For instance, variants in *RNF212* have been found to affect GRR in human (Kong et al. 2008), cattle (Sandor et al. 2012) and sheep (Johnston et al. 2016). Johnston et al. (2016) provided evidence that these effects were trans effects likely to affect RR globally. GRR is thus rather an oligogenic trait in these species (e.g., Stapley et al., 2017). Variants with sex-specific effects on GRR are common in humans (Kong et al. 2014) and have also been reported in sheep (Johnston et al. 2016), whereas variants affecting both male and female GRR were most common in cattle (Kadri et al. 2016). Male and female GRR were found to be highly correlated in cattle (Kadri et al. 2016) and sheep (Johnston et al. 2016), contrary to findings in humans (Fledel-Alon et al. 2011). Hotspot usage has also been found to be heritable and controlled by variants with large effects mainly associated with *PRDM9* (e.g., Fledel-Alon et al., 2011; Sandor et al., 2012).

66 The X-chromosome has most often been ignored in these studies due to specific challenges
67 (Khramtsova et al. 2019) including distinct inheritance, lower genotyping accuracy (e.g., lower
68 signal intensity in males) or paucity of SNPs on genotyping arrays (Wise et al. 2013). Yet,
69 recombination of the sex chromosomes may be particularly interesting as it may correspond to a
70 biological process distinct from that on autosomes. For one, gonosomal recombination drastically
71 differs between sexes, being restricted to the small pseudo-autosomal region (PAR) of homology
72 between the X and Y chromosomes in males, whereas females recombine along the entire length of
73 their X-chromosomes (Hinch et al. 2014). Accordingly, RR in the PAR was found to be 10 times
74 greater in males than in females in humans (Rouyer et al. 1986), mouse (Soriano et al. 1987), sheep
75 (Johnston et al. 2016) and red deer (Johnston et al. 2017), consistent with one obligatory gonosomal
76 crossover (CO) per male meiosis. Moreover, studies in the mouse have concluded that *PRDM9*
77 plays a role in CO positioning on autosomes but not in the PAR (Brick et al. 2012). Gonosomes
78 were also shown to pair significantly later during prophase I than the autosomes which may be
79 related to replication timing (Kauppi et al. 2011).
80 A new assembly of the bovine genome has been released recently (Rosen et al. 2020), providing a
81 substantial improvement compared to the previous build, particularly for the X-chromosome. This
82 represents an opportunity to improve the genetic map of the X-chromosome by using updated
83 information on marker ordering. The genetic distances between markers can be estimated with an
84 EM algorithm using coordinates of CO identified in a genealogy (Kong et al. 2010; Druet and
85 Georges 2015; Ma et al. 2015). This approach assumes that all CO have been detected although this
86 might not be true when informativeness is low (e.g., low marker density, parents harboring long
87 homozygous tracks, unphased genotypes). Hidden Markov models (HMM) are also frequently used
88 to estimate genetic distances, notably in the framework of the Lander-Green algorithm (Lander and
89 Green 1987). The likelihood of a map (order and distances) obtained with these models takes into
90 account information of neighboring markers and the variable marker informativeness across
91 families. In addition, genotyping errors or uncertainty in genotype calling (e.g. with low-coverage

sequence data) can also be accounted for. For instance, several methods rely on HMM to estimate genetic maps in full-sib families (Rastas et al. 2013; Bilton et al. 2018) or in multi-parental populations (Zheng et al. 2019). We herein implemented a similar approach in LINKPHASE3 (Druet and Georges 2015) and took advantage of the new bovine build to construct sex-specific genetic maps of the X chromosome using more than 100,000 genotyped animals. These new genetic maps were then used to estimate individual variation in recombination rate on the X-chromosome (XRR). We finally studied the genetic architecture of this new trait in each sex and determined its correlation with GRR measured on all the autosomes.

Materials and methods.

Data. We used genotypes from three dairy cattle populations from France, New Zealand and the Netherlands that were previously used by Kadri et al. (2016) to study global recombination rate on the autosomes. Animals from France (45,348) and the Netherlands (11,831) were Holstein-Friesian whereas those from New-Zealand (58,474) were Holstein-Friesian (24%), Jersey (19%) or crossbred (57%) individuals. For autosomes, 30,127 filtered SNPs were already available whereas for the present study, we selected 853 SNPs common to the BovineSNP50 and the BovineHD genotyping arrays (Illumina) that mapped on the X-chromosome in the newly released ARS-UCD1.2 assembly (Rosen et al. 2020) (based on the same Hereford cow as the previous assembly). Following the findings of Van Laere et al. (2008) and Johnson et al. (2019), we placed the pseudo-autosomal boundary (PAB) at position X:133,300,518. We filtered out monomorphic markers, those with call rate < 0.90 or with homozygosity < 0.98 in males (for X-specific, non-PAR markers only), leaving 744 SNPs mapping to the specific part of the X-chromosome and 73 to the PAR. We subsequently erased residual Mendelian inconsistencies (i.e., parent-offspring pairs with opposite homozygote genotypes and heterozygous offspring with both parents homozygous for the same allele). Among the selected genotyped animals, 90,481 sire-offspring and 26,107 dam-offspring pairs were available for CO identification in gametes transmitted by 2,958 sires and 11,228 dams.

Estimation of the genetic map and CO identification. Since we observed map errors in the sequence-based physical map for the X chromosome in the former genome assembly, we started by validating the new physical map order as described in Supplementary File 1. We then used LINKPHASE3 (Druet and Georges 2015) to phase individuals and identify CO. We start by describing this phasing process on the autosomes. Genotypes from offspring are first phased using Mendelian segregation rules and linkage observed in half-sibs is subsequently used to reconstruct the phases from the common genotyped parent. These are subsequently improved and corrected with an HMM specific to each sub-pedigree in which the haplotypes inherited by the offspring (one haplotype per offspring) are modelled as a mosaic of the two haplotypes of the parent (representing the two hidden states). The initial state probabilities represent the probability to inherit the paternal or the maternal haplotype at the first position and are equal to 0.5. Transition probabilities $\tau_{k-1,k}(i,j)$ represent the probability to inherit paternal or maternal haplotype at marker *k* conditional on the haplotype inherited at marker *k-1* and depend on the recombination rate between markers *k-1* and *k* (see below). Finally, the emission probabilities $\varepsilon_k(i)$ associated with hidden state *i* at marker *k* are functions of the alleles observed on the parental haplotypes, on the inherited haplotype and of the inheritance vector. CO are then identified where the inheritance patterns of an offspring change from paternal to maternal or vice versa.

We used an EM algorithm to determine the genetic distances between markers that maximizes the likelihood of data under our model. The estimation of the expected number of CO between two markers requires the definition of the forward variables $\alpha_k(i)$ that are equal to the probability to inherit parental haplotype *i* at marker position *k* conditional on observations at markers 1 to *k*, as well as the backward variables $\beta_{k'}(j)$ that are equal to the probability to inherit parental haplotype *j* at marker *k'* conditional on observations at markers *k'+1* to M (where M is the number of markers). The four possible pairs of states at marker pair (*k*, *k*+1) are $\{s_k = 1, s_{k+1} = 1\}$, $\{s_k = 1, s_{k+1} = 2\}$, $\{s_k = 2, s_{k+1} = 1\}$ and $\{s_k = 2, s_{k+1} = 2\}$, where s_k is the hidden state (inherited parental haplotype) at

marker *k* (using 1 and 2 for the paternal and maternal parental haploptypes, respectively). The probability of each pair {i,j} is equal to:

$$\frac{\alpha_k(i)\tau_{k,k+1}(i,j)\varepsilon_{k+1}(j)\beta_{k+1}(j)}{P(O|\lambda)}$$

where $\tau_{k,k+1}(i,j)$ is the transition probability from state *i* to *j* between markers *k* and *k*+1 and is equal to (1-$r_{k,k+1}$) and $r_{k,k+1}$ when *i* and *j* are respectively equal or different, $r_{k,k+1}$ being the recombination rate between the two markers. P(O|λ) is the probability of the observations conditional on the parameters of the model (equal to the sum of probabilities for the four possible configurations). The expected number of CO between markers *k* and *k+1* is equal to the probability of states combination $\{s_k = 1, s_{k+1} = 2\}$ and $\{s_k = 2, s_{k+1} = 1\}$ in the modelled offspring haplotype. Finally, the recombination rate $r_{k,k+1}$ is updated in the maximization step as the sum of expected number of CO across all offspring within a sub-pedigree and across all sub-pedigrees divided by the total number of modeled meioses.

The rules described above need to be modified when modeling the X-chromosome. LINKPHASE3 has therefore been previously modified to account for the segregation of the X-chromosomes (Murgiano et al. 2016). On the X-specific part, males transmit their maternal chromosome without recombination to their daughters whereas a null chromosome is transmitted to their sons (the paternal chromosome of males is modelled as null). Using this model adapted for the X-chromosome, we estimated the sex-specific genetic maps by performing 100 iterations of the EM algorithm described above.

In summary, the genetic map is estimated by finding the recombination rates for each marker interval that maximizes the likelihood of the segregation patterns modelled in half-sib families.

Estimation of genetic parameters for recombination rate on the X-chromosome. We ran LINKPHASE3 with the newly estimated sex-specific map of the X-chromosome to identify CO in

90,481 genotyped sire-offspring and 26,107 genotyped dam-offspring pairs. For each pair, the CO counts were associated to gametes transmitted by the parent to its offspring (each offspring allowing to count CO in one gamete of its parent). We conserved for further analysis only parent-offspring previously selected in the study from Kadri et al. (2016). These 114,254 individual CO counts were considered as repeated phenotypes from the 13,576 parents (2,839 sires and 10,737 dams) and X-chromosome recombination rates (XRR) in males and females were modeled as distinct traits. We first estimated heritability of XRR in each sex separately with the following univariate model (Kadri et al. 2016):

$$\boldsymbol{y} = \mathbf{1}\mu + \mathbf{Pc} + \mathbf{Z}_u\mathbf{u} + \mathbf{Z}_p\mathbf{p} + \mathbf{e}$$

where **y** is the vector of CO counts related to XRR (one record per offspring), μ is the mean effect (**1** is a vector with all elements equal to 1), **P**, $\mathbf{Z}_u$ and $\mathbf{Z}_p$ are incidence matrices relating respectively principal components, random polygenic effects and random permanent environment effects to the phenotypes, **c** is the vector of effects of the first four principal components of genetic variation (using 30,127 autosomal SNPs, see below) fitted to account for breed structure, **u** is the vector of normally distributed random individual polygenic effects $\sim N(0,\mathbf{A}\sigma_g^2)$, where σ_g^2 is the additive genetic variance and **A** is the additive relationship matrix estimated from the pedigree, **p** is the vector of normally distributed random permanent environment effects $\sim N(0,\mathbf{I}\sigma_p^2)$, where σ_p^2 is the permanent environment variance and **e** is the vector of normally distributed error terms $\sim N(0,\mathbf{I}\sigma_e^2)$, where σ_e^2 is the residual variance. We applied the same model to individual RR on each of the 29 bovine autosomes (BTA) for comparisons. These GRR phenotypes were available from a previous study and computed using 30,127 SNPs (Kadri et al. 2016). We estimated the genetic correlations of sex-specific XRR with sex-specific global recombination rate (GRR) estimated on autosomes (also available from Kadri et al. (2016)) with a multivariate linear mixed model including the same effects for the four traits. The random polygenic effects had the following covariance structure:

193

$$var\begin{bmatrix}\boldsymbol{u}_{XM}\\ \boldsymbol{u}_{XF}\\ \boldsymbol{u}_{GM}\\ \boldsymbol{u}_{GF}\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}\sigma^2_{g_{XM}} & \boldsymbol{A}\sigma_{g_{XM},g_{XF}} & \boldsymbol{A}\sigma_{g_{XM},g_{GM}} & \boldsymbol{A}\sigma_{g_{XM},g_{GF}}\\ \boldsymbol{A}\sigma_{g_{XM},g_{XF}} & \boldsymbol{A}\sigma^2_{g_{XF}} & \boldsymbol{A}\sigma_{g_{XF},g_{GM}} & \boldsymbol{A}\sigma_{g_{XF},g_{GF}}\\ \boldsymbol{A}\sigma_{g_{XM},g_{GM}} & \boldsymbol{A}\sigma_{g_{XF},g_{GM}} & \boldsymbol{A}\sigma^2_{g_{GM}} & \boldsymbol{A}\sigma_{g_{GM},g_{GF}}\\ \boldsymbol{A}\sigma_{g_{XM},g_{GF}} & \boldsymbol{A}\sigma_{g_{XF},g_{GF}} & \boldsymbol{A}\sigma_{g_{GM},g_{GF}} & \boldsymbol{A}\sigma^2_{g_{GF}}\end{bmatrix}$$

194

195 where $\sigma_{gi,gj}$ is the genetic covariance between trait *i* and *j* and where subscripts XM, XF, GM and
196 GF refer to male XRR, female XRR, male GRR and female GRR, respectively. Since male and
197 female records are measured on different individuals, the covariance structure for the random
198 permanent environment effect was:

199

$$var\begin{bmatrix}\boldsymbol{p}_{XM}\\ \boldsymbol{p}_{XF}\\ \boldsymbol{p}_{GM}\\ \boldsymbol{p}_{GF}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I}\sigma^2_{p_{XM}} & \boldsymbol{I}\sigma_{p_{XM},p_{XF}} & 0 & 0\\ \boldsymbol{I}\sigma_{p_{XM},p_{XF}} & \boldsymbol{I}\sigma^2_{p_{XF}} & 0 & 0\\ 0 & 0 & \boldsymbol{I}\sigma^2_{p_{GM}} & \boldsymbol{I}\sigma_{p_{GM},p_{GF}}\\ 0 & 0 & \boldsymbol{I}\sigma_{p_{GM},p_{GF}} & \boldsymbol{I}\sigma^2_{p_{GF}}\end{bmatrix}$$

200

201 where $\sigma_{ri,rj}$ is the covariance between permanent environment effects for traits *i* and *j*. Similarly,
202 the covariance structure for the residual error terms was:

203

$$var\begin{bmatrix}\boldsymbol{e}_{XM}\\ \boldsymbol{e}_{XF}\\ \boldsymbol{e}_{GM}\\ \boldsymbol{e}_{GF}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I}\sigma^2_{e_{XM}} & \boldsymbol{I}\sigma_{e_{XM},e_{XF}} & 0 & 0\\ \boldsymbol{I}\sigma_{e_{XM},r_{XF}} & \boldsymbol{I}\sigma^2_{e_{XF}} & 0 & 0\\ 0 & 0 & \boldsymbol{I}\sigma^2_{e_{GM}} & \boldsymbol{I}\sigma_{e_{GM},e_{GF}}\\ 0 & 0 & \boldsymbol{I}\sigma_{e_{GM},e_{GF}} & \boldsymbol{I}\sigma^2_{e_{GF}}\end{bmatrix}$$

204

205 where $\sigma_{ei,ej}$ is the covariance between residual error terms for trait *i* and *j*.

206 Genetic parameters of univariate and multivariate mixed models were estimated using Average
207 Information Restricted Maximum Likelihood analysis (AI-REML) implemented in AIREMLF90
208 (Misztal et al. 2002). Standard Deviations of heritabilities and correlations were estimated with

AIREMLF90 by repeated sampling of parameter estimates from their asymptotic multivariate normal distribution (Meyer and Houle 2013).

Haplotype-based association study and single variant associations. We performed a haplotype-based GWAS to find variants associated with male or female XRR. Zhang et al. (2012) demonstrated the benefits of such approach in cattle when using 50K SNP arrays. For the autosomes, we used the ancestral haplotype (AHAP) clusters from Kadri et al. (2016) available for 30,127 marker positions (with the number of clusters per position, *K*, set to 60). In that study, haplotypes were clustered at each marker position based on their similarity following the method of Scheet and Stephens (2006) as described in Druet and Georges (2010). For the X-chromosome, the phases estimated by LINKPHASE3 were completed using Beagle 4.1 (Browning and Browning 2007) which uses linkage disequilibrium information while conserving the pre-phasing information obtained from LINKPHASE3. Then, we clustered these haplotypes at each marker position in 60 AHAP using HiddenPHASE (Druet and Georges 2010).

The association between AHAP and XRR was evaluated by the same procedure as in Kadri et al. (2016). At each marker position, two mixed models were compared, one without AHAP effect (H0, no QTL) and one with AHAP fitted as random effects (H1). Both models included a mean, four principal component effects and a random polygenic effect to account for structure. A Likelihood Ratio Test (LRT) was used to test whether H1 was significant. The genome-wide significance threshold was set at 1.67×10^{-6} after Bonferroni correction for 30,000 independent tests (as an approximation for a total of 30,944 tested SNPs).

In addition to the haplotype-based GWAS, we also tested whether 11 variants associated with GRR and previously identified in Kadri et al. (2016) were significantly affecting XRR. To that end, we used the same linear mixed model and simultaneously fitted the variants as fixed effects. Each variant was fitted by a regression on SNP allelic dosage and the association was tested using a Z-test. For comparisons purposes, we also tested these variants for association with individual RR estimated for each of the 29 BTA separately.

To gain further insights into the correlation between global and X-specific individual RR levels, we estimated the correlations between haplotype effects estimated simultaneously at all marker positions (60 haplotypes per position). We fitted a mixed model with random haplotype effects (**h**) assumed to be normally distributed with a common variance σ_h^2. To that end we took advantage that models in which all haplotypes are fitted as random effects can be transformed into models in which genomic breeding values are estimated as the sum of haplotype effects, $\mathbf{g} = \mathbf{Z}_H\mathbf{h}$ (VanRaden 2008; Stranden and Garrick 2009), with $\mathbf{Z}_H$ being the incidence matrix relating random haplotype effects to the random polygenic effects. In this equivalent model, the haplotypes are used to estimate the genomic relationship matrix (GRM) between the random individual polygenic effects (Stranden and Garrick 2009). We thus first estimated an haplotype-based GRM $\mathbf{G}_H$ between 15,107 genotyped individuals (including the parents) by using all the ancestral haplotypes estimated in Kadri et al. (2016) and the method implemented in GLASCOW described in Zhang et al. (2012). Seventy-eight pairs of individuals with a relationship greater than 0.95, corresponding to monozygotic twins, were merged. We then estimated the genetic correlation between XRR and GRR within each sex with a bivariate model where the additive relationship matrix **A** was replaced by $\mathbf{G}_H$. As a result of the equivalence between the two models, the estimated genetic correlations corresponded also to the genetic correlations between random haplotype effects estimated for XRR and GRR (Karoui et al. 2012; Maier et al. 2015).

Finally, to obtain the estimators from the haplotype effects, we applied the following back-transformation (Stranden and Garrick 2009; Wang et al. 2012): $\hat{\mathbf{h}} = \sigma_g^{-2}\sigma_h^2\mathbf{Z}_H'\mathbf{G}_H^{-1}\hat{\mathbf{g}}$, where σ_g^2 is the variance of the random individual polygenic effects **g**.

Results

Estimation of sex-specific genetic distances and identification of crossovers. With the exception of a 2.55 to 3.72 Mb long inversion between positions chrX:51911569 and chrX:54460491, the marker order from the ARS-UCD1.2 bovine assembly appeared reliable (see Supplementary File 1)

and was used in further analyses. We ran LINKPHASE3 on the entire data set to estimate genetic distances with the EM procedure described in methods. We tested two marker orders: 1) the unchanged marker order of the ARS-UCD1.2 bovine assembly and 2) the same map order but with the order of the markers located between positions chrX:51911569 to chrX:54460491 inverted (see Supplementary Material 1). The likelihood of the genetic map estimated in females with the second order (logL = -415350.5) was much higher than the value obtained with the original order (logL = -417498.3). This corresponds to a highly significant LRT of 2 x 2147.8 achieved thanks to the size of our data set. The estimated map lengths in females were 122.85 and 121.17 cM for the first and second map, respectively. The modified map was hence 1.68 cM shorter, providing additional support for the new order as suggested by Groenen et al. (2012) for a similar case. Finally, the recombination rates in the marker intervals flanking the putative order inversion dropped from 2.04 cM and 1.82 cM to 0.57 cM and 0.36 cM with the corrected map.

Figure 1A represents the genetic maps in both-sexes and the corresponding average RR along the X-chromosome. In females, the estimated average RR on the X-chromosome was 0.873 cM/Mb. The lengths of the map for the X-specific part and for the PAR were 113.68 and 7.30 cM, respectively, corresponding to 0.86 and 1.29 cM/Mb. In females, RR along the X-chromosome was higher at both ends and reduced in the center (Figure 1C). In particular, we observed two large regions (~7 Mb) and a third smaller one (3 Mb) with almost no recombination. The first of these regions, extending from 38 to 45 Mb and with a RR below 0.01 cM/Mb for most of the 1Mb windows, corresponds to the centromere. The entire region contained only five SNPs, all of them with a MAF < 0.005. Similarly, the RR remained below 0.05 cM/Mb from 66 to 74 Mb and from 92 to 93 Mb and few informative SNPs were available in these regions (respectively 13 and 6 SNPs, all with MAF < 0.01 except one SNP in the first window with MAF = 0.013). The female RR at the pseudo-autosomal boundary (PAB) located at 133.300 Mb (Johnson et al. 2019) followed the trend of the surrounding marker intervals (e.g., no marked variation). In males, the estimated genetic length of the map was shorter and equal to 46.44 cM but recombination occurred only in the PAR.

This corresponds to 8.17 cM/Mb and represents a six-fold increase (6.3) in RR compared to females. The RR remained high for the entire PAR, yet due to its short size, it was difficult to observe a trend in variation of RR along the PAR (Figure 1B).

We estimated genetic maps for all autosomes with the same methodology. As reported by Housworth and Stahl (2003), or by Johnston et al. (2016, 2017), there was a clear linear relationship (adjusted-r^2 = 0.91 and 0.97 in males and females, respectively) between length of linkage map and physical size of the chromosomes (Figure 2 and Supp. Figure 1 for autosomes only). Chromosome-wide RR were fitted as a multiplicative inverse function (Johnston et al. 2017) consistent with one obligate CO per meiosis (adjusted-r^2 = 0.78 and 0.82 in males and females, respectively). In females, the RR on the X-chromosome and its genetic length were both in range with the relationships estimated for the autosomes. For instance, the X-chromosome had similar values to those observed on BTA2 that has a similar physical size. In males, the genetic length of the PAR in males was higher than expected according to the linear relationship estimated on autosomes. This was however not unique to the PAR since other autosomes presented larger deviations from the fit (e.g., BTA19). The RR estimated for the PAR in males lied also above the trend estimated on autosomes with a multiplicative inverse function (note that the size of the PAR lied outside the range of values for which the trend was fitted).

The distribution of numbers of CO per gamete on the X-chromosome in males and females is represented in Supp. Figure 2 (each parent-offspring pair representing one gamete). In females, the distribution ranged from 0 to 7 (after exclusion of one outlier with 19 CO) but with 99.7% having less than four CO (26.4%, 43.1%, 25.5% and 4.6% with 0, 1, 2 and 3 CO respectively). These values are close to those observed on chromosome 2 (having the same length; see Supp. Figure 2). In males, no CO were identified for most individuals (67.6%) and 1 single CO was identified for 32.2% of the sires. These values are below expectations and suggest that some CO remained undetected in the PAR (see discussion). We used the XOI R package (Broman et al. 2002) to measure the interference levels (ν) associated with this distribution of CO and with the distances

between identified CO. The values of the gamma function are compared to those observed on the autosomes in Supp. Figure 3. In females, higher interference levels were generally observed for shorter chromosomes in agreement with Broman et al. (2002). Indeed, the mean interference parameters on smaller (BTA16-29) and larger chromosomes (BTA1-15) were respectively equal to 5.5 and 5.1, although the lowest level was observed for BTA-22 (in the first group). The X-chromosome presented the second lowest observed value (ν = 3.8). In males, interference levels were rather stable across autosomes (4.9 *vs* 4.9 for short and large chromosomes) whereas the PAR had the highest interference levels (ν = 7.6), with less double CO observed than expected. Finally, average estimated interference levels were lower in males (4.9) than in females (5.3).

Estimation of genetic parameters. The genetic parameters estimated with the multivariate model are reported in Table 1. The heritability of female XRR was equal to 0.024 (± 0.006) and small compared to the value estimated for GRR (0.079). The contrast was even higher in males with heritabilities equal to 0.011 (± 0.003) and 0.137 (± 0.012) for XRR and GRR, respectively. Larger heritabilities for GRR were expected since this phenotype was obtained from observations on multiple chromosomes (similar to an average of multiple records). Interestingly, heritability for female XRR was higher than the heritability for female RR measured for the other chromosomes individually (Figure 3). The heritability for male XRR was in the range of values observed for other chromosomes despite the smaller map length. The repeatabilities were equal to 0.178 (± 0.006), 0.087 (± 0.007), 0.041 (± 0.002) and 0.050 (± 0.006) for male GRR, female GRR, male XRR and female XRR, respectively.

The estimated genetic correlations between female traits (XRR and GRR) was particularly high and equal to 0.759 (± 0.094) suggesting that these traits might be under the control of the same variants. Conversely, the genetic correlations between male traits was much lower (0.168 ± 0.124) with less evidence that these traits might be regulated by the same variants. The relatively high genetic correlation between male and female GRR (0.511 ± 0.078), previously reported by Kadri et al.

(2016), suggests that these two traits might be affected by shared variants. The genetic correlation between male and female XRR was moderate with a large standard deviation (0.399 ± 0.192) and indicates that these two traits would share less variants. Importantly, all parameters associated with male XRR presented higher standard errors.

Genome-wide association study and correlation between SNP or haplotype effects across traits. None of the tested positions in the haplotype-based GWAS for male or female XRR reached the significance threshold (Supp. Figure 4). One hypothesis to explain this lack of significant association is the lower power of detection when working with RR on a single chromosome compared to GRR. In addition to these GWAS approaches, we also tested whether eleven variants previously found to be significantly associated with male GRR, female GRR or both (Kadri et al. 2016) had also significant effects on male or female XRR (Supp. File 2). The significance threshold was set at 4.55×10^{-3} after Bonferroni correction for 11 independent tests. With the exception of the rs110203897 marker in the proximity of PRDM9, these variants were absent from the initial dataset and imputed (Kadri et al. 2016). Most of these variants were significant for male or female GRR (9 and 7, respectively). In contrast, none of the eleven candidate variants was associated with male XRR after correction for eleven independent tests (Supp. File 2). As a matter of comparison, 5.5 significant associations were found on average for these eleven variants and male RR measured for single autosomes (Supp. File 2), suggesting that the lack of significant association with male XRR was not due to a low number of records. Only one variant was significantly associated with female XRR (and three additional had uncorrected p-values < 0.05). Association with female RR on other single autosomes resulted in 0 to 4 significant tests per trait (2.1 on average). We also compared the effects of these eleven variants on XRR and GRR estimated in the same sex with a univariate model (e.g., independent analyses) and found a high correlation (0.86) in females whereas a moderate value (0.30) was observed in males (Figure 4). The high correlations between the effects on female

XRR and GRR suggests that all variants might be shared although some were not significantly associated with female XRR.

We finally estimated correlations between 1,807,620 random haplotype effects fitted simultaneously at all marker positions (60 ancestral haplotypes fitted at each marker position) as described in methods. When estimated with a REML procedure, the correlations between random haplotype effects on XRR and GRR were equal to 0.22 and 0.66 in males and females, respectively, consistent with the pedigree-based genetic correlations. To further illustrate this correlation, haplotype-effects estimates were obtained by transformation of the breeding values (see methods). The correlation between these estimated haplotypes effects on female XRR and GRR was equal to 0.90 (Supp. Figure 5).

Discussion

Thanks to a large genotyped pedigree containing more than 100,000 meioses, we were able to validate the map order and to estimate the genetic map on the X-chromosome in cattle. For the map order, we took advantage of highly informative sub-pedigrees in which dams were phased and the haplotypes they transmitted to their sons were known without errors (see Supp. File 1 for more details). In contrast with the previous build, there was no evidence for errors in the new bovine assembly on the X-chromosome with the exception of an inversion in the middle of the chromosome. The identified order inversion might result from a true inversion segregating in the population but we did not find clear evidence based on local principal components analysis (Ma and Amos 2012) or on inspection of whole genome-sequence data. The corrected map was more likely in both pure Holstein and Jersey populations, suggesting that the inversion is not a segregating variant. The Hereford cow used to realize the bovine assembly might also carry an inversion when compared to the breeds used in the present study or the order inversion might result from an assembly error. Note also that we searched only for errors associated with markers mapped at relative large distances from their correct position (more than 1 cM). It is indeed more difficult to

validate errors at shorter distances because there are no or very few CO between such markers. In addition, the informative meioses are distinct for each pair of markers making comparisons more challenging. Noise associated with genotyping errors is also more problematic at short distances. Consequently, small map errors might still be present.

We relied on a HMM to estimate the genetic distances between markers. Such an approach presents the advantage to account for informativeness and for potential genotyping errors. For instance, relying as previously only on the identified CO (Ma et al. 2015; Kadri et al. 2016) resulted in an 8.5% shorter map in females. Similarly, the map was inflated by 1.6% when genotyping errors were ignored. LINKPHASE3 extracts the information from different family structures. In a first step, Mendelian segregation rules are used to phase parent-offspring pairs whereas linkage information is used in a second step, within half-sibs families. Single parent-offspring pairs are used if genotypes from the parents were phased in the first step (i.e., if one of the grand-parents was also genotyped). As a result, reliable genetic maps were estimated by modelling 90,481 male and 26,107 female meioses. Although the physical map remains a reasonable approximation in females (since the RR is 0.873 cM / Mb), the male RR is much higher in the PAR. It is therefore recommended to use sex-specific genetic maps rather than relying on physical maps. Overall, our study indicated that the new bovine assembly is reliable and provides genetic maps that are important to perform QTL mapping, association studies or imputation on the X-chromosome. It is important to include the X-chromosome in such applications since it is a gene rich chromosome and contains important genes related to fertility, reproduction and potentially recombination (Demars et al. 2013; Fernandez et al. 2014; Arishima et al. 2017; Pacheco et al. 2020).

At the chromosome level, recombination rates from autosomes had a multiplicative inverse relationship with their physical length (r^2 = 0.82 and 0.78 in males and females). These high adjusted-r^2 values represented a good fit, but some chromosomes deviated from the estimated relationship. Similar or even larger deviations were observed in previous studies (Housworth and Stahl 2003; Johnston et al. 2016; Johnston et al. 2017), indicating that additional factors influence

RR at the chromosome level. These might for instance include chromosome and chromatin structure (e.g., proportion of heterochromatic regions), gene density, GC content, density of repetitive elements, presence of structural variants or hotspot motifs density (Coop and Przeworski 2007; Stapley et al. 2017). The recombination rates we herein estimated for the X-chromosome indicate that, at the chromosome level, it had similar properties to autosomes in females. For instance, RR was close to the RR estimate on BTA2, a chromosome with almost the same physical size. The distribution of number of CO were close too. Note that more markers were genotyped on BTA2 (1589), providing higher informativeness although the number of markers on the X-chromosome was sufficient to estimate the genetic length of the chromosome. Conversely, the PAR had a much higher RR in males, as previously observed in other species (Rouyer et al. 1986; Soriano et al. 1987; Johnston et al. 2016; Johnston et al. 2017) and in agreement with the obligate CO per meiosis hypothesis. Higher RR had consistently been reported for other short chromosomes such as micro-chromosome in birds (Rodionov 1996; Megens et al. 2009). Very few double CO were detected as in human where they were found to be rare (Schmitt et al. 1994). As a result, the PAR presented higher interference levels in males than on other chromosomes consistent with previous reports of higher interference on shorter chromosomes (Broman et al. 2002; Wang et al. 2016). For autosomes, this trend of higher interference levels in shorter chromosomes was present only in females. Campbell et al. (2015) or Wang et al. (2016) observed this relationship in both sexes, but more pronounced in females as in our study. As pointed out by Wang et al. (2016), cows presented thus lower recombination rates and higher interference levels than bulls whereas opposite patterns were observed in human (e.g., Campbell et al., 2015). Finally, chromosomes with lower interference levels in our study, such as BTA-4 or BTA-22, were previously found to have higher levels in comparable cattle populations (Wang et al. 2016). This suggests that some technical aspects as data and CO cleaning might cause these differences.

Importantly, a certain number of CO remained undetected in males on the PAR (more CO occurred but we did not identify them). Indeed, one obligate CO per meiosis would result in a length larger

than 50 cM whereas the length estimated with our data remained below 47 cM. This suggests that even when using an HMM approach, there is still a lack of informativeness on such a small chromosome (for instance at the borders of the chromosome that have higher influence in small chromosomes). Furthermore, the number of identified CO per gamete (0.36) indicates that only a fraction of CO is unambiguously identified (around 20% of expected CO missing). The lower informativeness on the PAR is also due to a lower marker density per cM compared to other chromosomes. Although 73 SNPs were available for a 5 Mb region, this corresponded only to 1.6 SNPs per cM. A more detailed comparison of different informativeness measures is available in Supp. File 5.

At more local levels, the RR in females on the X-chromosome presented differences with autosomes. In bovine, the X-chromosome is metacentric whereas the autosomes are acrocentric. We previously observed on autosomes that recombination was higher in telomeres and lower in the center of the chromosome (Kadri et al. 2016) in agreement with findings in other species (Broman et al. 1998; Liu et al. 2014; Venn et al. 2014). The X-chromosome presented a somehow distinct pattern in females, with higher recombination rates at both extremities, more so at the proximal part (opposite to the PAR). In the middle of the X-chromosome, three large regions from 3 to 8 Mb long presented very low recombination rates (below 0.05 cM/Mb) and were characterized by the absence of informative SNPs (few SNPs and low MAFs). The first of these regions corresponds to the centromere and few similar regions of low RR were observed on autosomes. In males, although recombination was not constant in the PAR, the trend was difficult to characterize given the small physical size of the region.

Genetic parameters estimation and GWAS were realized using a linear mixed model (LMM), although XRR was almost a binary trait in males. Generalized linear mixed models (GLMM) with a Poisson distribution might be more appropriate to model XRR. However, differences might be limited for individuals with multiple records, and when count frequencies are not too extreme as in the present study (e.g., not too close from 0 or 1 for a binary trait). For comparison purposes, we

applied such models, and genetic parameters were in line with those obtained with the LMM. The individual polygenic effects estimated with LMM and GLMM were highly correlated (0.9883 and 0.9996 in males and females, respectively), and GWAS did not reveal significant associations with a GLMM neither (a correlation of 0.90 between LRT values). Results from the LMM were thus presented as they are on the same scale as GRR, making comparisons more interpretable (for instance, a Poisson GLMM would assume multiplicative effects whereas effects are additive in the LMM framework).

Heritabilities for XRR were as expected lower than for GRR obtained with multiple chromosomes. More interestingly, female XRR was more heritable (0.024) than RR on any other single chromosome, whereas male XRR (0.011) had a value similar to other chromosomes. We are not aware of similar estimates in cattle or other species. In absolute terms, these heritabilities were low, and similar to those observed for fertility traits such as non-return rate (Jansen and Lagerweij 1987). The higher female heritability might result from a more accurate phenotype since haplotypes from males are known on the X-chromosome and can help to infer the haplotypes from their daughters (see Supp. File 1 for more details). Conversely, male XRR might be more difficult to study than RR on other chromosomes. As we discussed above, more CO might have remained undetected compared to larger chromosomes and the estimated number of CO per individual was below expectations. More information (higher marker density per cM) is needed to identify male CO in the PAR with high power. Given the presence of interference and the size of the region, there is less room for variation between individuals. The genetic correlations between female GRR and female XRR were high, indicating that these might be controlled by the same genetic variants. Despite a low power for testing associations with RR on a single chromosome in females, a few variants previously identified for their effect on GRR were associated with female XRR and the effects of these variants on XRR were correlated to the effects on GRR. We also estimated that the correlation between the effects of haplotypes on XRR and GRR was high. Overall, recombination patterns at the chromosome levels and genetic correlations with female GRR indicate that female XRR might

494 follow the same process as on the other chromosomes. Correlation between male GRR and male
495 XRR were lower suggesting more different traits as might be expected from the very different
496 recombination patterns. The variants affecting male GRR were not significant for male XRR
497 although many significant associations were found for individual autosomes, suggesting that the
498 number of records was not an issue. However, the male XRR phenotype might be less informative
499 due to the elements discussed earlier. More data or more accurate phenotypes (with a higher
500 heritability), including CO obtained with higher density maps or from molecular techniques such as
501 MLH1 foci counts (e.g., Peterson, et al., 2019; Wang, et al., 2019), are required to clarify how
502 strong are these correlations and to identify eventually shared variants.
503 As illustrated in recent studies, recombination in the PAR has specific mechanisms (Acquaviva et al.
504 2020). We herein showed that genetic correlation between male XRR and GRR was limited. We
505 estimated however recombination rates in broad intervals (larger than 50 kb). Complementary
506 analyses with whole-genome sequence data, and with LD-based approaches in particular, could
507 estimate recombination rate at finer-scale. Such high-density recombination maps would help to
508 determine whether genomic features associated with recombination are the same on autosomes, the
509 X-chromosome and the PAR. For instance, CO occur in hotspots associated with PRDM9 binding
510 motifs in mammals such as human or mice (e.g., Baudat, et al., 2010). However, Brick et al. (2012)
511 observed that *PRDM9* was not associated with CO positions on the PAR in mouse. Therefore, it
512 would be interesting to determine whether such hotspots are present on the X-chromosome
513 (representing female-specific recombination) and the PAR in cattle, and whether they are associated
514 with *PRDM9* binding motifs. Depending on the species, CO frequency might also be higher close to
515 the transcription start site or at CpG island. Recombination has additionally been associated with
516 different families of repetitive elements. Fine-scale recombination maps would thus allow to
517 determine whether this landscape is the same on autosomes, the X-specific part and the PAR
518 (undergoing different constraints during meiosis).
519

Acknowledgements

This work was funded by the Fonds de la Recherche Scientifique-FNRS (F.R.S.-FNRS) under grant T.0080.20 ("LoCO motifs" research project) and supported by the Damona European Research Council grant ERC AdG-GA323030 to Michel Georges. Carole Charlier and Tom Druet are Senior Research Associate from the Fonds de la Recherche Scientifique–FNRS (F.R.S.-FNRS). We also thank Livestock Improvement Corporation (Hamilton, New Zealand), Allice (Paris, France), INRA (Jouy-en-Josas, France), and CRV (Arnhem, the Netherlands) for providing us with the SNP genotype and pedigree data. We used the supercomputing facilities of the Consortium des Equipements de Calcul Intensitf en Fédération Wallonie Bruxelles (CECI), funded by the F.R.S.-FNRS.

References

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30:** 97-101.

Acquaviva L, Boekhout M, Karasu ME, Brick K, Pratto F, Li T et al. (2020) Ensuring meiotic DNA break formation in the mouse pseudoautosomal region. *Nature* **582:** 426-431.

Albers PK, McVean G (2020) Dating genomic variants and shared ancestry in population-scale sequencing data. *PLoS Biol* **18:** e3000586.

Amaral ME, Kata SR, Womack JE (2002) A radiation hybrid map of bovine X chromosome (BTAX). *Mamm Genome* **13:** 268-271.

Arishima T, Sasaki S, Isobe T, Ikebata Y, Shimbara S, Ikeda S et al. (2017) Maternal variant in the upstream of FOXP3 gene on the X chromosome is associated with recurrent infertility in Japanese Black cattle. *BMC Genet* **18:** 103.

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

Bilton TP, Schofield MR, Black MA, Chagné D, Wilcox PL, Dodds KG (2018) Accounting for Errors in Low Coverage High-Throughput Sequencing Data When Constructing Genetic Maps Using Biparental Outcrossed Populations. *Genetics* **209:** 65-76.

Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV (2012) Genetic recombination is directed away from functional genomic elements in mice. *Nature* **485:** 642-645.

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* **63:** 861-869.

Broman KW, Rowe LB, Churchill GA, Paigen K (2002) Crossover interference in the mouse. *Genetics* **160:** 1123-1131.

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81:** 1084-1097.

557 Campbell CL, Furlotte NA, Eriksson N, Hinds D, Auton A (2015) Escape from crossover
558 interference increases with maternal age. *Nat Commun* **6:** 6260.
559 Chowdhury R, Bois PR, Feingold E, Sherman SL, Cheung VG (2009) Genetic analysis of variation
560 in human meiotic recombination. *PLoS Genet* **5:** e1000648.
561 Coop G, Przeworski M (2007) An evolutionary view of human recombination. *Nat Rev Genet* **8:**
562 23-34.
563 Demars J, Fabre S, Sarry J, Rossetti R, Gilbert H, Persani L et al. (2013) Genome-wide association
564 studies identify two novel BMP15 mutations responsible for an atypical hyperprolificacy phenotype
565 in sheep. *PLoS Genet* **9:** e1003482.
566 Druet T, Farnir FP (2011) Modeling of identity-by-descent processes along a chromosome between
567 haplotypes and their genotyped ancestors. *Genetics* **188:** 409-419.
568 Druet T, Gautier M (2017) A model-based approach to characterize individual inbreeding at both
569 global and local genomic scales. *Mol Ecol* **26:** 5820-5841.
570 Druet T, Georges M (2010) A hidden markov model combining linkage and linkage disequilibrium
571 information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics* **184:**
572 789-798.
573 Druet T, Georges M (2015) LINKPHASE3: an improved pedigree-based phasing algorithm robust
574 to genotyping and map errors. *Bioinformatics* **31:** 1677-1679.
575 Fernandez AI, Munoz M, Alves E, Folch JM, Noguera JL, Enciso MP et al. (2014) Recombination
576 of the porcine X chromosome: a high density linkage map. *BMC Genet* **15:** 148.
577 Fledel-Alon A, Leffler EM, Guan Y, Stephens M, Coop G, Przeworski M (2011) Variation in
578 human recombination rates and its genetic determinants. *PLoS One* **6:** e20321.
579 Frohlich J, Kubickova S, Musilova P, Cernohorska H, Muskova H, Vodicka R et al. (2017)
580 Karyotype relationships among selected deer species and cattle revealed by bovine FISH probes.
581 *PLoS One* **12:** e0187559.

582 Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF et al. (2012)
583 Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491:** 393-
584 398.
585 Grossman SR, Shlyakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G et al. (2010) A
586 composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*
587 **327:** 883-886.
588 Hinch AG, Altemose N, Noor N, Donnelly P, Myers SR (2014) Recombination in the human
589 Pseudoautosomal region PAR1. *PLoS Genet* **10:** e1004503.
590 Housworth EA, Stahl FW (2003) Crossover interference in humans. *Am J Hum Genet* **73:** 188-197.
591 Jansen J, Lagerweij GW (1987) Adjustment of non-return rates of AI technicians and dairy bulls.
592 *Livest Prod Sci* **16:** 363-372.
593 Johnson T, Keehan M, Harland C, Lopdell T, Spelman RJ, Davis SR et al. (2019) Short
594 communication: Identification of the pseudoautosomal region in the Hereford bovine reference
595 genome assembly ARS-UCD1.2. *J Dairy Sci* **102:** 3254-3258.
596 Johnston SE, Berenos C, Slate J, Pemberton JM (2016) Conserved Genetic Architecture Underlying
597 Individual Recombination Rate Variation in a Wild Population of Soay Sheep (Ovis aries). *Genetics*
598 **203:** 583-598.
599 Johnston SE, Huisman J, Ellis PA, Pemberton JM (2017) A High-Density Linkage Map Reveals
600 Sexual Dimorphism in Recombination Landscapes in Red Deer (Cervus elaphus). *G3: Genes,*
601 *Genomes, Genetics* **7:** 2859-2870.
602 Kadri NK, Harland C, Faux P, Cambisano N, Karim L, Coppieters W et al. (2016) Coding and
603 noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination
604 rate in cattle. *Genome Res* **26:** 1323-1332.
605 Karoui S, Carabano MJ, Diaz C, Legarra A (2012) Joint genomic evaluation of French dairy cattle
606 breeds using multiple-trait models. *Genet Sel Evol* **44:** 39.

607 Kauppi L, Barchi M, Baudat F, Romanienko PJ, Keeney S, Jasin M (2011) Distinct properties of
608 the XY pseudoautosomal region crucial for male meiosis. *Science* **331:** 916-920.
609 Khramtsova EA, Davis LK, Stranger BE (2019) The role of sex in the genomics of human complex
610 traits. *Nat Rev Genet* **20:** 173-190.
611 Kong A, Thorleifsson G, Frigge ML, Masson G, Gudbjartsson DF, Villemoes R et al. (2014)
612 Common and low-frequency variants associated with genome-wide recombination rate. *Nature*
613 *Genetics* **46:** 11-16.
614 Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A et al. (2010)
615 Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467:**
616 1099-1103.
617 Kong A, Thorleifsson G, Stefansson H, Masson G, Helgason A, Gudbjartsson DF et al. (2008)
618 Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science*
619 **319:** 1398-1401.
620 Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl*
621 *Acad Sci USA* **84:** 2363-2367.
622 Leutenegger AL, Prum B, Genin E, Verny C, Lemainque A, Clerget-Darpoux F et al. (2003)
623 Estimation of the inbreeding coefficient through use of genomic data. *Am J Hum Genet* **73:** 516-523.
624 Liu EY, Morgan AP, Chesler EJ, Wang W, Churchill GA, Pardo-Manuel de Villena F (2014) High-
625 resolution sex-specific linkage maps of the mouse reveal polarized distribution of crossovers in
626 male germline. *Genetics* **197:** 91-106.
627 Ma J, Amos CI (2012) Investigation of inversion polymorphisms in the human genome using
628 principal components analysis. *PLoS ONE* **7:** e40224.
629 Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, Sun C et al. (2015) Cattle Sex-Specific
630 Recombination and Genetic Control from a Large Pedigree Analysis. *PLoS Genet* **11:** e1005387.
631 Mackay TF, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and
632 prospects. *Nat Rev Genet* **10:** 565-577.

633 Maier R, Moser G, Chen GB, Ripke S, Coryell W, Potash JB et al. (2015) Joint analysis of
634 psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and
635 major depressive disorder. *Am J Hum Genet* **96:** 283-294.
636 Megens HJ, Crooijmans RP, Bastiaansen JW, Kerstens HH, Coster A, Jalving R et al. (2009)
637 Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in
638 chicken. *BMC Genet* **10:** 86.
639 Meyer K, Houle D (2013) Sampling based approximation of confidence intervals for functions of
640 genetic covariance matrices. *Proc Assoc Adv Anim Breed Genet* **20:** 523-526.
641 Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, D. L (2002) BLUPF90 and related programs
642 (BGF90). *In Proceedings of the 7th world congress on genetics applied to livestock production* **28:**
643 21-22.
644 Murgiano L, Shirokova V, Welle MM, Jagannathan V, Plattet P, Oevermann A et al. (2016)
645 Hairless Streaks in Cattle Implicate TSR2 in Early Hair Follicle Formation. *PLoS Genet* **11:**
646 e1005427.
647 Pacheco HA, Rezende FM, Peñagaricano F (2020) Gene mapping and genomic prediction of bull
648 fertility using sex chromosome markers. *J Dairy Sci* **103:** 3304-3311.
649 Peterson AL, Miller ND, Payseur BA (2019) Conservation of the genome-wide recombination rate
650 in white-footed mice. *Heredity* **123:** 442-457.
651 Petit M, Astruc JM, Sarry J, Drouilhet L, Fabre S, Moreno CR (2017) Variation in Recombination
652 Rate and Its Genetic Determinism in Sheep Populations. *Genetics* **207:** 767-784.
653 Rastas P, Paulin L, Hanski I, Lehtonen R, Auvinen P (2013) Lep-MAP: fast and accurate linkage
654 map construction for large SNP datasets. *Bioinformatics* **29:** 3128-3134.
655 Rodionov AV (1996) Micro vs. macro: structural-functional organization of avian micro- and
656 macrochromosomes. *Genetika* **32:** 597-608.

Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E et al. (2020) De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **9**. doi: 10.1093/gigascience/giaa021.

Rouyer F, Simmler MC, Vergnaud G, Johnsson C, Levilliers J, Petit C et al. (1986) The pseudoautosomal region of the human sex chromosomes. *Cold Spring Harb Symp Quant Biol* **51 Pt 1:** 221-228.

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.

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* **78:** 629-644.

Schmitt K, Lazzeroni LC, Foote S, Vollrath D, Fisher EM, Goradia TM et al. (1994) Multipoint linkage map of the human pseudoautosomal region, based on single-sperm typing: do double crossovers occur during male meiosis? *Am J Hum Genet* **55:** 423-430.

Soriano P, Keitges EA, Schorderet DF, Harbers K, Gartler SM, Jaenisch R (1987) High rate of recombination and double crossovers in the mouse pseudoautosomal region during male meiosis. *Proc Natl Acad Sci USA* **84:** 7218-7220.

Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM (2017) Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philos Trans R Soc Lond B Biol Sci* **372**:20160455**.**

Stranden I, Garrick DJ (2009) Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J Dairy Sci* **92:** 2971-2975.

Van Laere AS, Coppieters W, Georges M (2008) Characterization of the bovine pseudoautosomal boundary: Documenting the evolutionary history of mammalian sex chromosomes. *Genome Res* **18:** 1884-1895.

VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* **91:** 4414-4423.

Venn O, Turner I, Mathieson I, de Groot N, Bontrop R, McVean G (2014) Nonhuman genetics. Strong male bias drives germline mutation in chimpanzees. *Science* **344:** 1272-1275.

Wang H, Misztal I, Aguilar I, Legarra A, Muir WM (2012) Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res (Camb)* **94:** 73-83.

Wang RJ, Dumont BL, Jing P, Payseur BA (2019) A first genetic portrait of synaptonemal complex variation. *PLOS Genetics* **15:** 1008337.

Wang Z, Shen B, Jiang J, Li J, Ma L (2016) Effect of sex, age and genetics on crossover interference in cattle. *Sci Rep* **6:** 37698.

Wise AL, Gyi L, Manolio TA (2013) eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am J Hum Genet* **92:** 643-647.

Zhang Z, Guillaume F, Sartelet A, Charlier C, Georges M, Farnir F et al. (2012) Ancestral haplotype-based association mapping with generalized linear mixed models accounting for stratification. *Bioinformatics* **28:** 2467-2473.

Zheng C, Boer MP, van Eeuwijk FA (2019) Construction of Genetic Linkage Maps in Multiparental Populations. *Genetics* **212:** 1031-1044.

Table 1. Genetic parameters of male and female X-chromosome recombination rate (XRR) and male and female global recombination rate (GRR) estimated with a multitrait linear mixed model. estimated correlations between the global recombination rate (GRR) and the X chromosome recombination rate (XRR) in males and females. Diagonal elements represent heritabilities, upper and lower off-diagonal elements are respectively genetic and phenotypic correlations (± standard deviations).

	Male GRR	Female GRR	Male XRR	Female XRR
Male GRR	**0.137 ± 0.012**	0.511 ± 0.078	0.168 ± 0.124	0.154 ± 0.125
Female GRR	0.053 ± 0.009	**0.079 ± 0.010**	0.120 ± 0.151	0.759 ± 0.094
Male XRR	0.007 ± 0.004	0.004 ± 0.005	**0.011 ± 0.003**	0.399 ± 0.192
Female XRR	0.009 ± 0.007	0.150 ± 0.006	0.007 ± 0.003	**0.024 ± 0.006**

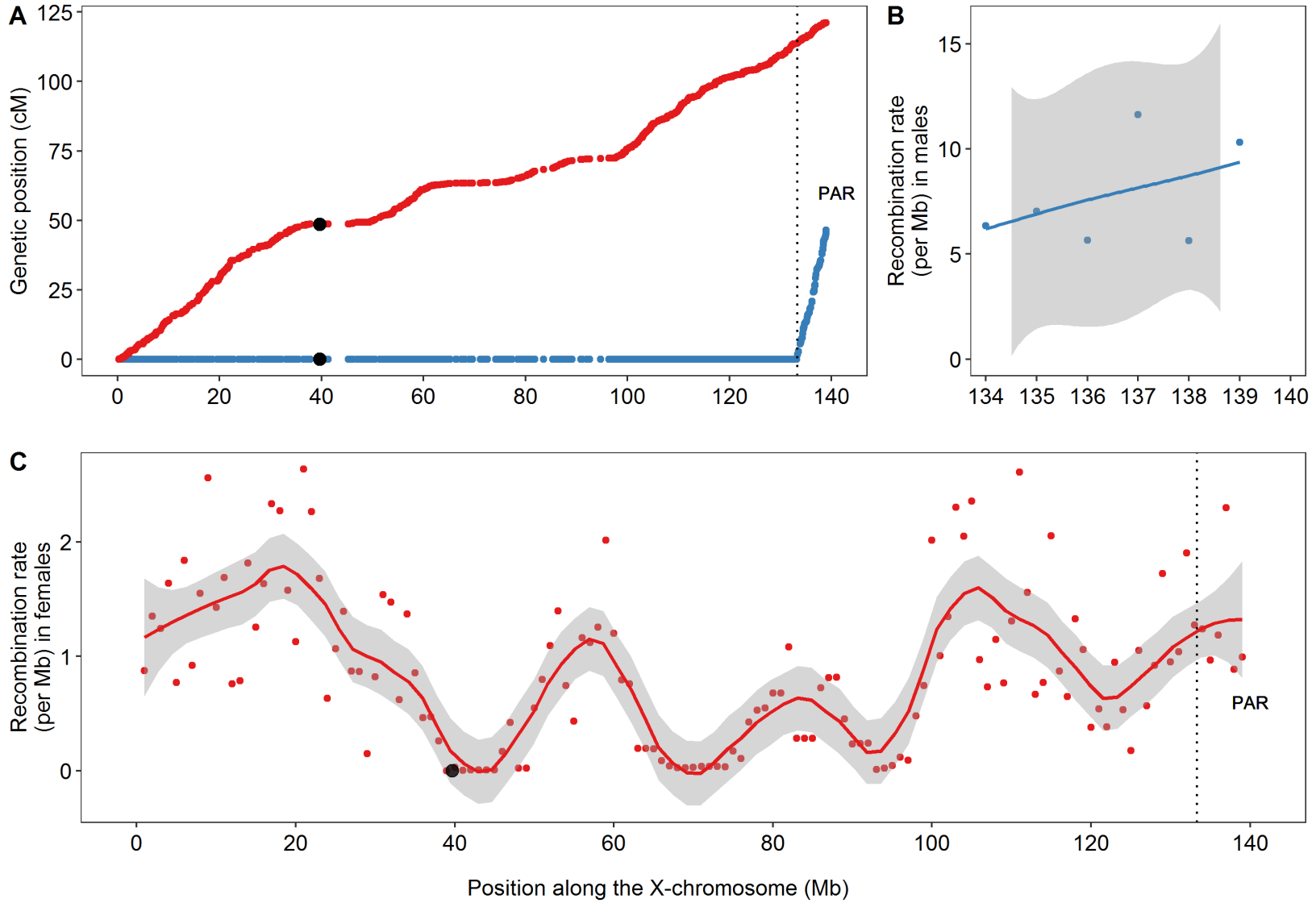
Figure legends

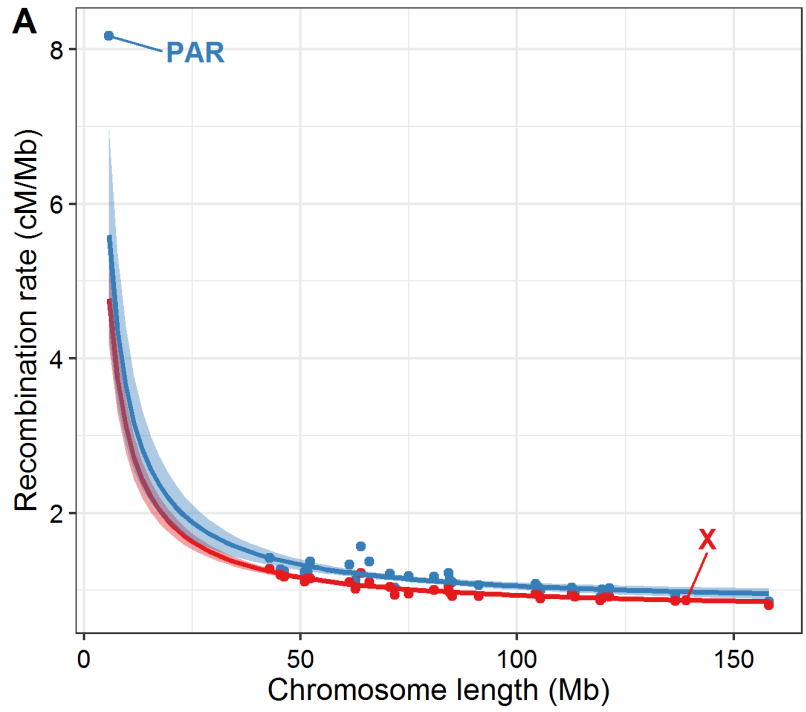
Figure 1. A) Sex-specific genetic maps of the X-chromosome; B) and C) Recombination rate per Mb along the X-chromosome estimated per sex. Recombination rate was first estimated in 1 Mb successive non-overlapping bins. The blue and red loess smoothed lines were fit for males and females with a span parameter of 0.5 and 2.5 respectively, using the ggplot2 (v.3.2.1) package in R (v.3.4.4). In males, we represented the rates only for the PAR. The black circle represents the position of the centromere (according to flanking microsatellites and BAC libraries reported in Amaral et al. (2002) and Frohlich et al. (2017)), whereas the vertical line indicates the limit of the PAR.

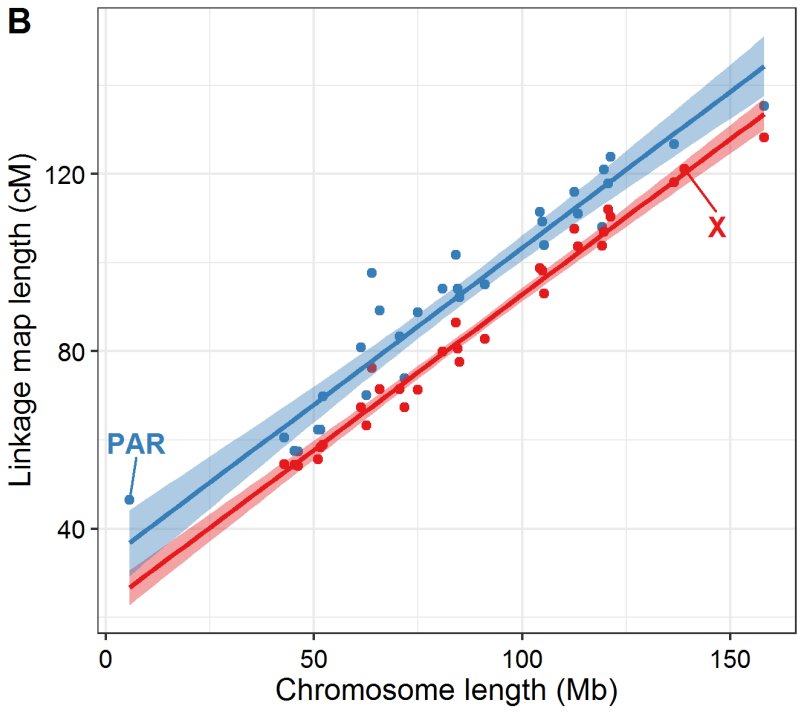
Figure 2. Relationship between recombination rate, genetic length and physical chromosome length on autosomes. For both panels, the curves were fitted per sex and using autosomes only (males in blue, females in red). The X-chromosome in females and the PAR in males were subsequently added. A) The relationship between recombination rate per Mb and physical length was fitted with a multiplicative inverse function. B) The relationship between genetic and physical lengths was fitted with a linear model. The curves were fitted with the stat_smooth function of ggplot2 and shaded areas indicate 95% confidence interval.

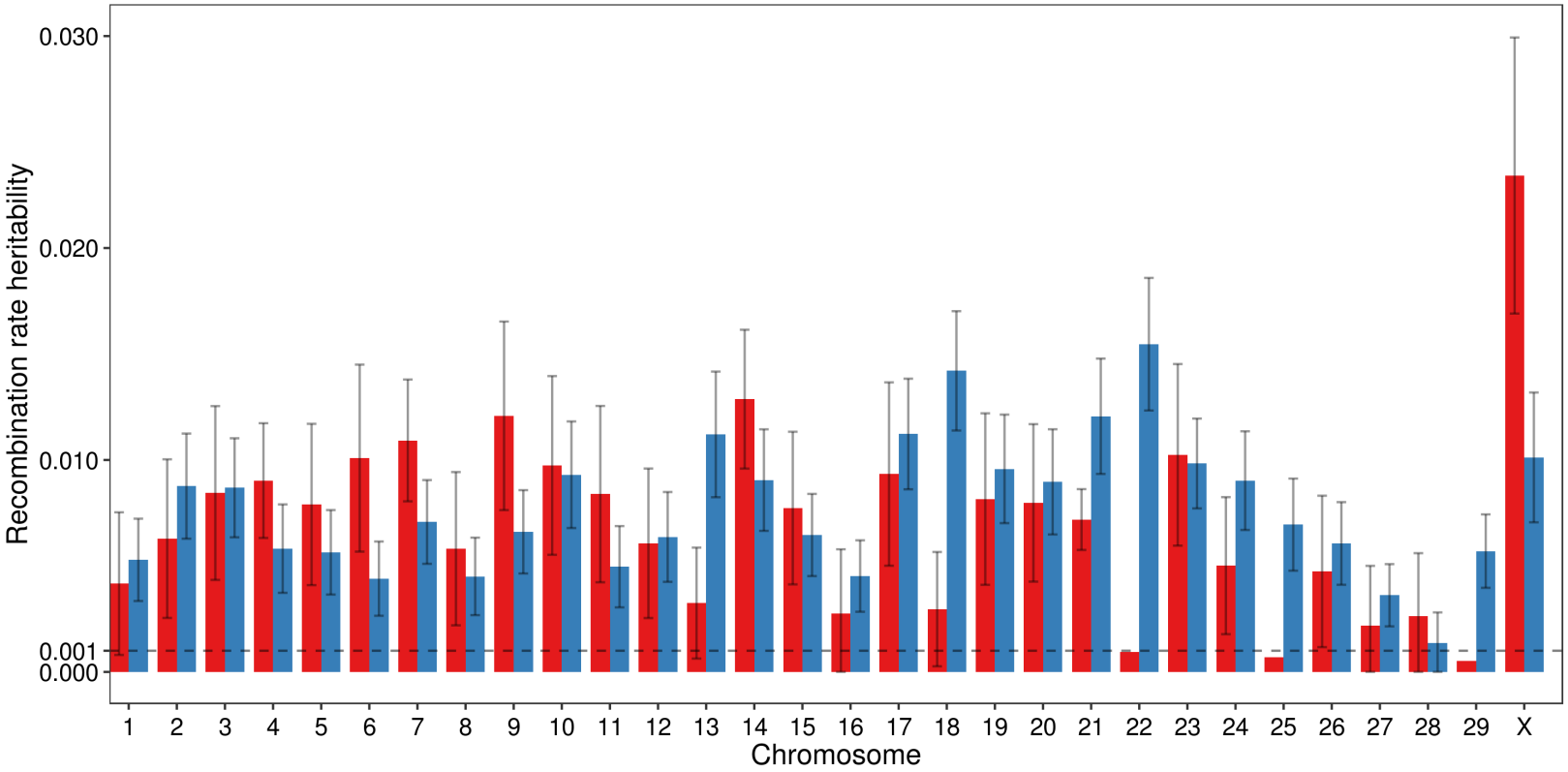
Figure 3. Comparison of heritabilities estimated with univariate models for XRR and RR measured on each autosome in males and females. Error bars represent standard errors (SE). SE were not reported for heritabilities below 0.001.

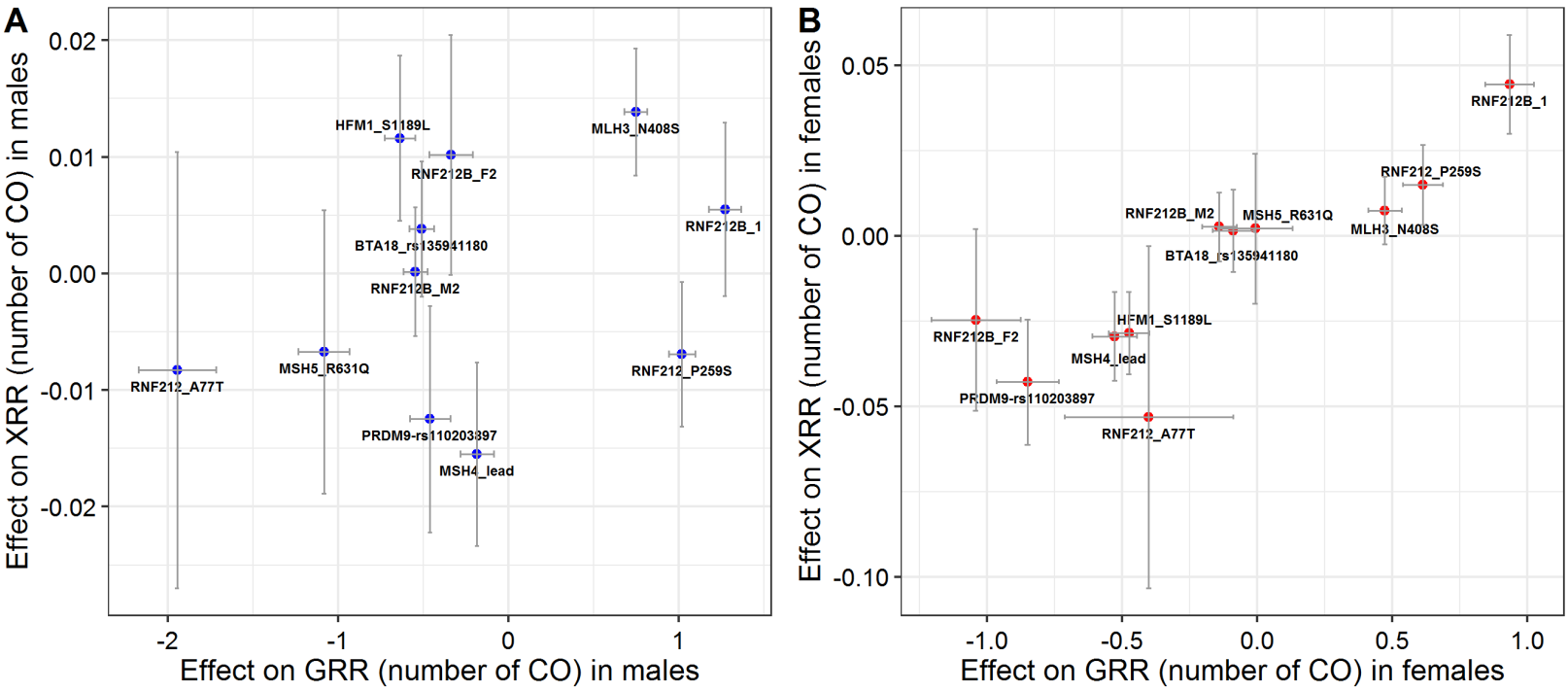
Figure 4. Comparison of estimated effects on global and X-chromosome recombination rates for a set of eleven previously identified variants (Kadri et al. 2016). Comparisons were realized for male (A) and female (B) recombination. The error bars represent the standard errors from the estimates.


A
Genetic position (cM)
125
100
75
50
25
0
PAR
0
20
40
60
80
100
120
140
B
Recombination rate
(per Mb) in males
15
10
5
0
134
135
136
137
138
139
140
C
Recombination rate
(per Mb) in females
2
1
0
PAR
0
20
40
60
80
100
120
140
Position along the X-chromosome (Mb)


A
PAR
X
Recombination rate (cM/Mb)
8
6
4
2
0
50
100
150
Chromosome length (Mb)


B
PAR
X
Linkage map length (cM)
120
80
40
0
50
100
150
Chromosome length (Mb)


Recombination rate heritability
0.030
0.020
0.010
0.001
0.000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
X
Chromosome


A
Effect on XRR (number of CO) in males
0.02
0.01
0.00
-0.01
-0.02
HFM1_S1189L
MLH3_N408S
RNF212B_F2
RNF212B_1
BTA18_rs135941180
RNF212B_M2
MSH5_R631Q
RNF212_A77T
RNF212_P259S
PRDM9-rs110203397
MSH4_lead
-2
-1
0
1
Effect on GRR (number of CO) in males
B
Effect on XRR (number of CO) in females
0.05
0.00
-0.05
-0.10
RNF212B_1
RNF212_P259S
RNF212B_M2
MSH5_R631Q
MLH3_N408S
BTA18_rs135941180
HFM1_S1189L
RNF212B_F2
MSH4_lead
PRDM9-rs110203897
RNF212_A77T
-1.0
-0.5
0.0
0.5
1.0
Effect on GRR (number of CO) in females