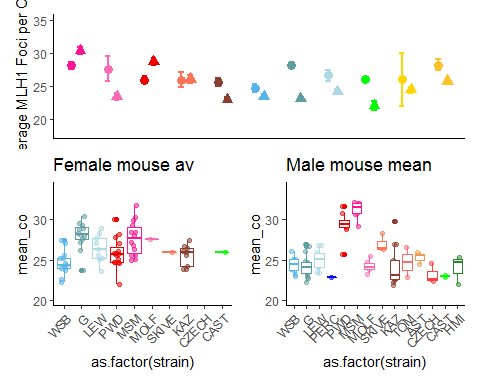
MLH1 Results Outline

Table of Contents

# Main Figure



# Genome wide recombination rate estimates for both sexes

We used counts of MLH1 foci per cell to estimate the genome wide recombination rates across our panel of 14 wild derived inbred strains sampled from three subspecies within the house mouse species complex; *M. m. domesticus,* *M. m. musculus* and *M. m. molossinus*. Mean MLH1 foci counts were quantified from a total of 1867 spermatocytes and 1409 oocytes.

The general patterns of heterochiasmy in house mouse are displayed in Figure 1A. Taking note of the direction and the magnitude of the sex differences, our results confirm two general patterns: i) genome wide recombination rates averages are greater females compared to males (female biased heterochiasmy) and ii) the degree of heterochiasmy (F:M rates) is generally low in these house mouse strains, ranging from 1.17 in *domesticusG* to 1.02 in *domesticusLEW*. Three notable exceptions of male biased heterochiasmy are from the strains *musculusPWD*, *molossinusMSM* and *musculusSKIVE*, with heterochiasmy values of 0.88, 0.9 and 0.96 respectively.

Examining the mouse means of MLH1 foci per cell separately for each sex reveal distinct patterns of variation (Figure1 B and C). The female means are evenly distributed around the sex-wide mean of approximately 25 MLH1 foci per cell (Figure1 B). While in males, the strain specific means separate more clearly into two groups of high genome wide recombination rates near 30 MLH1 foci per cell and low genome wide recombination rates, near 23 MLH1 foci per cell (Figure1 C).

# Evolutionary Framework

**M1. Mixed Model**

**M2. Linear Model**

**M3. Linear Model**

**M4. Linear Model**

We applied a series of models to our dataset of mean MLH1 foci counts from 187 mice to fit the patterns of variation across sex, subspecies and strain into an evolutionary framework. All of the effects within the mixed model M1 were highly significant (LRT: subsp p= 1.7210^{-4}, sex p = 3.9710^{-5}, subspecies \* sex p = 3.110^{-5}) and random LTR; strain p= 0).

The linear model results confirm the qualitative observations of sex-specific evolution and significant difference in the degree of sexual dimorphism in genome wide recombination rates. The subspecies fixed effects tested in M2, had weak to no significance (glm; p=0.22 and p= 0.09 for *M. m. musculus* and *M. m. molossinus* respectively), indicating that most of the variance is due to strain effects. As such, we focus on results from M3. The two most significant strain effects were *domesticusG* and *musculusMSM* (p= 1.0410^{-6} and p= 3.9910^{-6} for *domesticusG* and *molossinusMSM* respectively).

In addition to significant strain effects for *molossinusMSM*, (glm; p = 0.06 and p = 3.9910^{-6}) both *musculusPWD* and *musculusMSM* had significant strain by sex effects (glm; p = 3.8610^{-4} and p = 1.2610^{-4}) *musculusPWD* by male and *musculusMSM* by male.

Combined these results demonstrate that i) the significant sex-specific evolution in MSM and PWD and ii) the larger magnitude of sexual dimorphism/heterochiasmy in *domesticusG*.

A linear model was run on the sex specific data sets (M4), with 192 male and 144 female mean MLH1 foci per cell. There were three significant strain effects in the male specific data; *musculusPWD* (glm; p=7.3710^{-10}), *musculusSKIVE* (glm; p=0.01) and *molossinusMSM* (glm; p=2.2310^{-14} with effect sizes ranging from 5, 7, and 2 foci respectively. Given these results, theses strains are classified as ‘high rec’ strains and point to rapid evolution in the recombiantion rate for spermatocytes.

While the female specific linear model has many more had significant strain effects; *domesticusG* (glm; p=2.510^{-6}), *domesticusLEW* (glm; p=0.01), *domesticusPWD* (glm; p=0.02), *domesticusMSM* (glm; p=6.2410^{-6}), *domesticusMOLF*(glm; p=0.08), and *domesticusKAZ*(glm; p=0.1), the effect sizes have a smaller range of (4 to 1), indicating significant but small strain specific effects on mean MLH1 foci per cell.

## Within Mouse Variance in CO Count per Cell

We examine the within animal variance in MLH1 foci count per cell within the same models applied to mean MLH1 foci count (replacing mean MLH1 foci count with variance of MLH1 foci across cells as the dependent variable).

The linear models support the general pattern of females having almost twice as much variance in MLH1 foci per cell compared to males (Figure 1)). For both the mixed and linear models, sex was the only significant effect (LRT; p= 0, glm; p= 2.310^{-4}). Since the measures for within mouse variance may be more susceptible to technical error effects from the staining protocol (i.e. background noise), we replicated the model analysis using a subset of cells with higher quality scores. These models replicated the results of the full data set with sex being the most significant effect (LTR; p = 0, and glm; p = 2.310^{-4} ).

There is no significant difference in the amount of within mouse variance of MLH1 foci counts per cell in males according to the linear models. In females, the significance of strain effects was not consistent between the full data set and higher quality dataset *domesticusLEW* (glm; p = 0) was a significant effect in the full model and *musculusPWD* (glm; p =0.04) in the higher quality dataset.

The striking patterns of variation across male and females suggest that the genome wide recombination rates have distinct evolutionary trajectories. For female the pattern of Strain averages distributed around a species wide average fits a model of stabilizing or relaxed evolution in contrast to the male pattern where there is rapid evolution in a subset of genetic strains that is close to a model of directional selection on genome wide recombination rate.

# Evolution of Genome wide recombination rate associated with evolution of mean DSB number

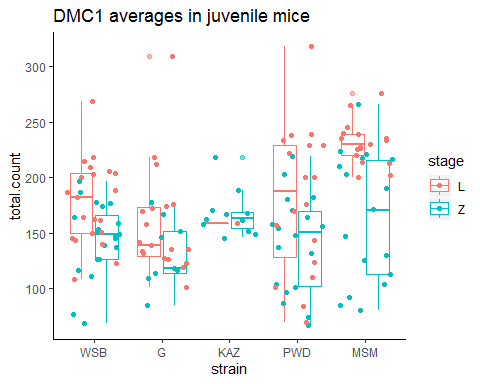


Table X. DMC1 foci counts per cell summary

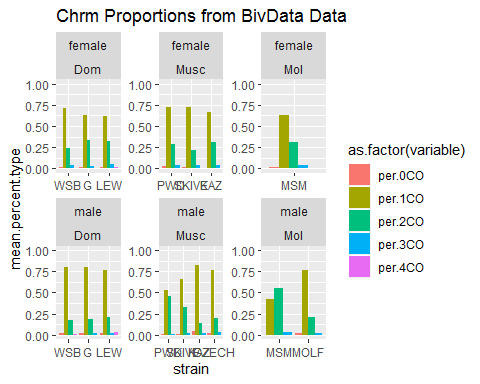
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | category | mean.MLH1 | var.MLH1 | ncells.x | mean.count.x | stage.x | mean.count.L | ncells.y | mean.count.y | stage.y | mean.count.Z | MLH1.L\_ratio | MLH1.Z\_ratio |
| 5 | WSB male | 24 | 0.9 | 21 | 178 | L | 178 | 20 | 144 | Z | 144 | 0.14 | 0.17 |
| 1 | G male | 24 | 1.8 | 19 | 158 | L | 158 | 9 | 132 | Z | 132 | 0.15 | 0.18 |
| 2 | KAZ male | 24 | 5.6 | 1 | 159 | L | 159 | 11 | 167 | Z | 167 | 0.15 | 0.14 |
| 4 | PWD male | 29 | 3.0 | 18 | 180 | L | 180 | 18 | 141 | Z | 141 | 0.16 | 0.21 |
| 3 | MSM male | 31 | 1.1 | 17 | 231 | L | 231 | 17 | 164 | Z | 164 | 0.14 | 0.19 |

In an attempt to localize the above male specific crossover number evolution within the meiotic pathway we quantified a marker for DSBs, DMC1, in prophase spermtaocytes. DMC1 foci were scored from a total of 76 leptotene and 75 zygotene staged spermatocytes from juvenile mice (12 to 18 days) of three low rec, ( *musculusKAZ* , *domesticusWSB* , and *domesticusG*) and two high rec strains ( *musculusPWD* and *molossinusMSM*).

The mean number of DMC1 foci per cell is associated with evolution of the number of precursors (DSBs), the two high recombining strains tested have significantly more DMC1 foci compared to low recombining strains in leptoene cells (t.test; p=0, one-way-anova; p=0.00027). However the differences in DMC1 foci were not significant for zygotene cells (later prophase) between the two groups (t.test; p=0.66, one-way-anova; p=0.15). Indicating that this marker at early prophase is more predictive of the downstream crossover number differences.

After DSB formation, DSBs are repaired as either non-crossover (NCO) or crossovers (COs). with the vast majority being repaired as NCOs. Thus the ratio of CO:DSB is a partial indicator of the proportion of DSBs which are designated into COs. The ratios, calculated for DMC1 means from both stages of prophase, are not significantly different between the high and low strain groups (t.test; p = 0.94 and p=0.11 for leptotene and zygotene based ratios respectively).

# Evolution of genome wide recombination rate is reflected at the single chromosome level



We distill the variance observed in the mean MLH1 foci counts per cell to number of MLH1 foci per bivalent. Because our observations of MLH1 foci are made in pachtyene cells, our chromosome observations are fully synapsed paired homologous chromosomes (4n copies of each chromosomes). Additionally we can distinguish the identity of chromsomes (e.g. Chromosome 1 or Chromosome 2), for these reasons we use the ‘term bivalent’.

Ninety six percent of the pooled dataset of single bivalents (n=34982) have either one or two crossovers (Figure X). The proportion of 1CO:2CO distinguishes the high and low rec strains (Figure X). This confirms the intuitive interpretation that high recombination strains are enriched for 2CO bivalents at the expense of 1CO bivalents. In the high rec strains the 2CO proportions are 0.33 ( *musculusSKIVE* ), 0.44 ( *musculusPWD* ), and 0.53 (*molossinusMSM* ).

While the proportions are significantly different (chi-square-test; p = 0.06), males *musculusPWD* and *molossinusMSM* have 2CO proportions more similar to each other than strains from the same subspecies ( *musculusPWD* and *musculusKAZ* chi-square test; 2CO p=3.1510^{-33}) and *molossinusMSM* and *molossinusMOLF* chi-square test; 2CO p=4.7210^{-13}).

# Single Bivalent Level Results

For each strain, we deconstructed the estimates of genome wide recombination rate by isolating single bivalents and quantifying the recombination landscape at the single bivalent level using an previously developed image analysis pipeline. The error of the image analysis algorithm is measuring chromosome features is low and highly similar to manual measures (Peterson 2019). A limitation of our image analysis algorithm we used is that not all bivalents per cell are isolated. The range isolation rates per cell in this data set is 0.51 in *molossinusMSM* male to 0.72 *musculusKAZ* female.

Due to overlap of chromosomes most cells do not have measurements for all bivalents, but we assume that the isolation of chromsomes within cell images is not biased. From our total set of cell images 10458 chromosome objects were isolated by the image analysis software. After the human curation step, 9829 single bivalent observations remained. Given the large number of single chromosome observations, we assume that each of the data sets are equally representative of general patterns.

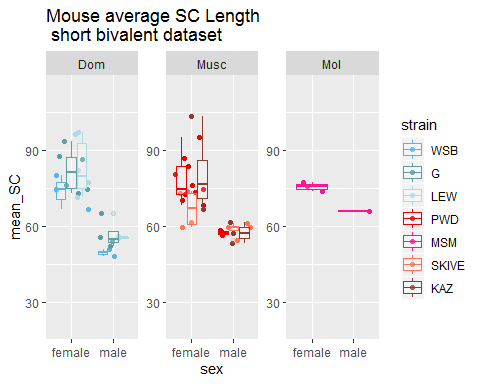
Using this data set, we address two main questions 1) Which traits are sexually dimorphic? and 2) Which traits fit distinguish the high and low recombining strains for males?

In our analysis our first step is to test for measurable differences, between sexes or rec.groups and is the differences are significant, our second step is to apply the same sets of models applied to the mean MLH1 foci counts to test / describe the effects due to subspecies or strains.

# Q1 Sex Differences in SC-AE Lengths

In many mammalian species the SC lengths are longer in females, yet most all of these observations come from species with female biased heterochiasmy. Our panel of house mice provides an opportunity to test if the sex differences in chromatin compaction (SC length) are reversed in the high rec strains (with male-biased heterochiasmy; *musculusPWD* and *molossinusMSM*). A simple model might predict if SC length is the strongest predictor for number of MLH1 foci per cell males of the high recombining strains would have longer SC length leading to the difference in total MLH1 foci per cell.

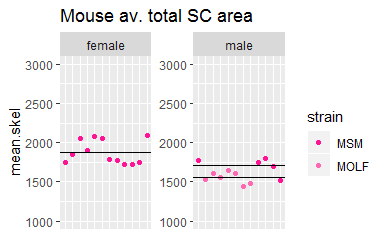
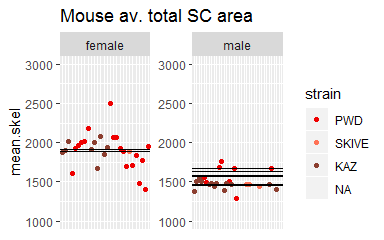
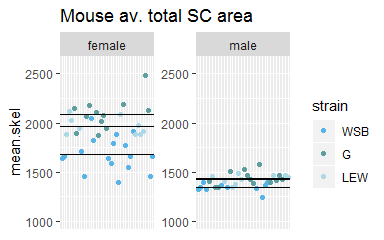
To account for confounding effects of sex chromosomes on measures of chromatin compaction (SC area) we use a reduced bivalent data set made up of bivalents with SC lengths below the 2nd quartile for SC length from a single cell to compare across sexes. This results in a dataset of the shortest 4 to 5 chromosomes and excludes the X bivalent in the oocytes, which is predicted to be the third longest chromosome by Mb. A total of 678 ‘short’ bivalents were isolated from 103 oocytes and 37 spermatocytes. Even though this smaller data set has decreased power, it offers a more comparable set of single bivalents to compare between sexes.



All strains have longer mouse mean for short bivalents in females (Figure X) and all but one strain, SKIVE, had significant differences. The difference, (F:M), across strains ranges from 1.15 in MSM to 1.49 in WSB.

<(t.test; p = 0.02, p =0.00049, and p=0.0016 for *domesticusWSB*, *domesticusG* and *domesticusLEW* respectively. the musculus strins (t.test; p= 0.00011, 0.06, and 0.11 for *musculusPWD* , *musculusKAZ* , *musculusSKIVE* respectively.>

For an addtional metric of chromatin compaction, we tested a whole cell measure of SC length using an second image analysis algorithm which converts the SC signal (into single pixel wide skeleton) and provides accurate estimates of the summed SC length of bivalents for single cells (ref Wang). We apply a skeletonizing image analysis algorithm to all MLH1 meiocyte spreads images. Mouse means were calculated from 2984 cell wide total SC measures (figure X). (where as 3680 number of cells had MLH1 measures ).



In all of the strains the means of total SC per cell was significantly greater in females than males (ttest; *domesticuesWSB* p = 0.00000032, *domesticuesG* p = 0.0000000027, *domesticuesLEW* p = 0.0000000019, *musculusPWD* p = 0.00085, *musculusKAZ* p = 0.00000087, *molossinusMSM* 0.027

To test for strain and subspecies specific effects on the sex difference in SC lengths we apply the following set of models.

**M1. Mixed Model**

**M2.Linear Model**

**M3. Linear Model**

In the mixed models for the sex differences mean **short bivalents** length as the dependent variable, sex is the most significant effect (LTR; p =6.910^{-11}). The interaction (subspecies by sex) effect was also slightly significant (p = 0.13).

In the mixed model run on the mean total SC metric all of the coefficients were significant, (LTR; subsp p= 0, sex p = 1.210^{-31}, and interaction p = 7.1210^{-4}, (random LRT; strain p = 0).

In linear models (M2 and M3) for **mean short bivalent length**, sex is the most significant coefficient (M2 glm; p =0 and M3 glm; p = 0) with slight significant strain and interaction effects (glm; p=0.03, p =0.09 respectively).

The following effects had significant p values for the **total SC** from the above model (for the linear models M2 and M3), subspecies *musculus* (glm M2; p = 2.1410^{-4}), subspecies *molossinus* (glm M2; p =6.2410^{-4}), sex male (glm M2; p =4.310^{-8}), strain lew p =5.910^{-6}, strain g p =1.7410^{-11},

For M3 in the mean total SC lengths there were many strain effects are significant, but the largest are sex (glm M3; p = 4.310^{-8}) but also some strain effects, G strain (glm M3; p = 1.7410^{-11}), LEW strain (glm M3; p = 5.910^{-6}), PWD strain (glm M3; p = 6.2710^{-5}). (this means…). Supporting the strong sex difference and also indicating that SC length has evolved across strains within this panel, when chromatin condensation is summarized at this level.

The SC length per chromsomes can be thought of as the ‘area / available real-estate for COs. These results suggest that this area is ’utilized’ in different ways between sexes, espicially in the strains with the high rec males, which have less ‘area’ for crossovers compared to females, but have more crossovers per cell on average.

We focus on aspects of the recombination landscape for the two major classes of bivalents; i) the placement of single foci along a single crossover bivalent and ii) the spacing of two foci on the same bivalent, or the interfocal distance (IFD).

# Q1. 1CO Position Sex Differences

The mean normalized foci positions per mouse from 6829 single crossover bivalents. The CO positions are normalized by the SC length and anchored at the centromere with the normalized position values ranging from 0 to 1, (refered here as terminal or telomeric position).

In all strains, the 1CO landscape had significant differences between males and female which followed the “typical landscape” of sex difference observed in many other species (Sardell Kirkpatrick). The foci position was more medially places in females while males have more terminal normalized foci positions (ttest; p = 2.9210^{-22}).

**M1. Mixed Model**

**M2. Linear model**

**M3. Linear model**

For the Mixed model, sex is the most significant effect (LTR; p =1.2610^{-25}).The random effect of strain is also significant (LRT; p =0.01).

The linear models were used to detect if specific strains differed in the sex differences in the 1CO landscape. While sex was the most significant effect for both M2 and M3 (there were some strain specific effects…)

The male effects were the most significant (glm M2; p =1.3310^{-7}, glm M3; p = 1.3310^{-7}). While the musc subspecies, LEW and KAZ strains had slightly signifcant effects (glm M2; Musc subsp p = 0.07, LEW strain p = 0.03) and (glm M3; LEW strain p = 0.03, KAZ strain p = 0.07).

# Q1. Sex Differences in CO Interference (IFD)

Crossover interference is one of the major determinates of the recombination landscape. It generates a distribution of evenly spaced crossover along chromosomes instead of a random and more uniform distribution.

We isolated 1360 and 1272 2CO bivalents in females and males to test describe sex differences in crossover interference. Mean interfocal distances (IFD) were calculated from 42 female and 45 male mice. We examined both raw IFDraw and normalized by SC length (IFDPER). The mean IFDraw measures of crossover interference as a mechanical force mediated through the SC. We compared the IFDraw to test the hypothesis of interference acting as a mechanical force transmitted through the SC. Previous measures found no sex difference in measures of IFDraw in mice (de Boer et al ).

The difference in mean IFDraw are slightly significant sexes from the full data set (t.test; p = 0.07). However *domesticusG* might be an outlier (by having a larger than average degree of sexual dimorphism). When the *domesticusG* observations are removed, the difference is not longer significant (t.test; p = 0.27). This indicates there is no general pattern of sex differences in crossover interference measured in physical (SC) units. (confirming deBoer and other models of crossover interference acting through physical/mechanical force).

While the comparisons of the IFDPER metrics can reveal more general recombination landscape patterns while controlling for the underlying differences in chromatin compaction and SC length. Also this metric is closer to measures of interference from linkage maps which are also removed from physical scales in that they measure frequencies of crossovers. Crossover interference is stronger in male specific linkage maps compared to female maps in a variety of species (ref).

The difference in mean IFDPER are highly significant between sexes. The mean IFDPER are significantly long in six of seven strains (MSM t.test; p-value = 0.01), *musculusSKIVE* p-value = 4.310^{-5}), *musculusPWD* p-value = 2.2810^{-4}), *molossinusMSM* p-value =0.01), *domesticusLEW* p-value = 3.5710^{-5}), *domesticusG* p-value = 0.05), and *domesticusWSB* (t.test; p-value =0.03) ). Only *musculusKAZ* did not have a difference in mean IFDPER (t.test; p-value = 0.33). These results indicate a general pattern of two foci are seperated by more area (hence stronger interference) in males. These results indicate a general pattern of two foci are separated by more area (hence stronger interference) in males.

GENERAL RESTULTS DIFFERENCES between sexes

In examining the IFDPER distributions, we note that the female norm IFD are centered at ~50% â but they also seem to have slight an enrichment of short (<25%) normalized IFDs

While the general pattern in the male distributions are enrichment longer IFDPER. (however we note that there are slight differences in the IFDPER distributions between the high and low rec males to be followed up in later section)

THE male vs female pattern… the female norm IFD are centered at ~50% but they also seem to have a higher rate of short (<25%) normalized IFDs. There seems to be a 25% norm.IFD cutoff (in females) but some mice seem to have less of this cut off. Another way to describe these results is in terms of thresholds (where the distributions end) with the general pattern of this sex difference is that the IFDPER distribution is pushed higher (females have mean of .5 IFDPER where as males have a IFDPER mean of ~.6?)

Since the most significant t.test were significantly different across sexes for IFDPER and we follow up with the model framework to.

To test for strain specific effects we applied the model framework

**M1. Mixed Model**

**M2. Linear Model**

**M3. Linear model**

For the Mixed models of IFDs, sex is a significant effect for both raw and nrmIFD.

(IFD.PER), sex is most significant effect (LRT: p = 6.7410^{-14}) subspecies, interaction are slightly significant.

For the normalized values in both M2 and M3, sex is a significant effect, increasing nrm.IFD in males. SKIVE\*male is the only other consistently significant effect, which also increases the nrm.IFD measure (p = male-SKIVE p = )

Overall There’s a low amount of significant effects across the 2CO IFD measures. This might be an indication that interference is conserved across these samples and/or that there is too much noise across from chromosome specific effects.

Our comparison of interference across sexes supports the general patterns from the literature.

IFD raw is not significantly different (t.tests and models?). indicating that the…

When interference is measured in physical SC units, the differences between sexes is low / only slightly significant -leading support that the phyisical measures of interference (in SC units) is conserved within species.

While when the variance in chromatin compaction is controlled males have stronger interference (like genetic maps) the impact on the rec landscapes being.

The results from the previous section – confirm (known sex differences for the ‘typical’ recombination landscape (Sardell Kirkpatrick). The next section is meant to focus on the greater aspect of variation in mean MLH1 counts per cell the high rec males males strains ( *musculusPWD* , *musculusSKIVE*, and *molossinusMSM*) and the rest of the males with the main objective of this section is to test for significant correlations with the evolution of mean MLH1 foci per cell.

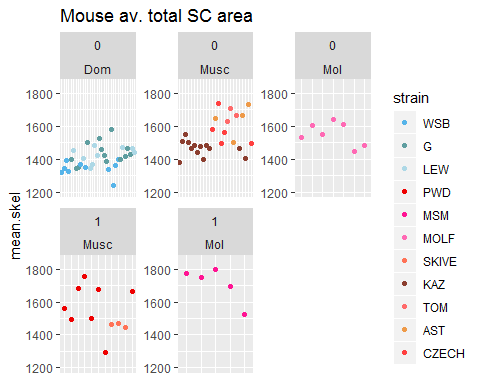
In comparing the male specific single bivalent based metrics the first task/level is to tests for significant differences between the high and low rec groups. The second being to follow up with linear models that test for effects of subspecies and strains.

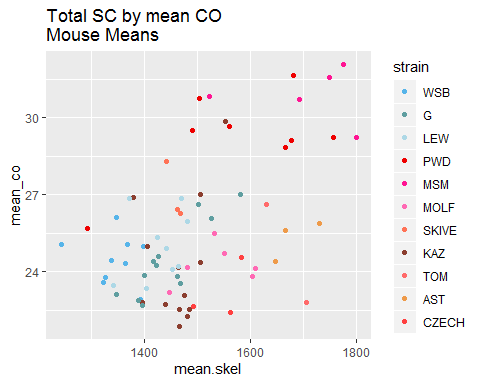
The literature provides basic predictions for the direction of the relationship the genome wide recombination rate and two aspects of the recombiantion landscape; SC length (positve) and crossover interference (negative). Following this logic we predict i) no strain effects for SC length or crossover interference within *domesticus*, ii) *musculusPWD* will have greater SC length and weaker interference than *musculusSKIVE*, which will have longer SC and weaker interference compared to the other *musculus* strains and iii) *molossinusMSM* will have longer SC and weaker interference compared to *molossinusMOLF*.

# Q2 SC Length

Our prediction for this section is that the high rec strains will have longer SC lengths proportional to the increase in the mean genome wide recombination rate.

The confounding effects of sex chromosomes are removed for this section that compares across males from different inbred strains. We compared mouse means of three SC based metrics; total SC, mean short bivalent length and mean long bivalent length. The long bivalent data set was isolated in the same manner as the short bivalent data set.





There is a positive correlation between mean MLH1 foci per cell and total SC, however the variation in mean total SC does not completely predict the high and low rec strains (Figure X). While for the total data set of mean total SC, the high rec strains have significantly (ttest; p = 0.01) when the high and low rec strains are compared within subspecies only the *molossinus* strains have significant difference (ttest mol p = 0.03), while the *musculus* strains do not have significant differences (ttest; musc p= 0.87).

Additionally, the means for the reduced bivalent datasets, short and long bivalents, are not significantly different between the high and low rec strains (ttest; short p = 0.88 and long p = 0.18).

We follow up the general pattern of a weak signal of high rec strains having longer SC lengths by running the model framework.

**Linear model M1**

**Linear model M2**

For total SC in M1 (glm including subsp), the two subspecies effects are significant (glm; p= *Musculus* 1.2410^{-6} and *Molossinus* p= 10^{-6})

In the short bivalent data set, analyzed by M1, both subspecies are sig (glm; p= *Musculus* 0 and *Molossinus* p= 0.01).

long bivalent set in M1 –

Musc 0.04

MSM strain 0.09

**M3 – tests within subspecies!!** SO the M3 model can be run within species specific data – to test the predictions (based on the variation in mean MLH1 counts), (use all 3 metircs) \*\*the within subspecies M3 models don’t fully meet the predictions

for total.Sc : there are significant Dom strain effects: in musc kaz is significant negative, several other strains have higher total SC, (MOL in the intercept is significant).

short biv: Dom: Significant G effect, MUSC no significant – strain effects

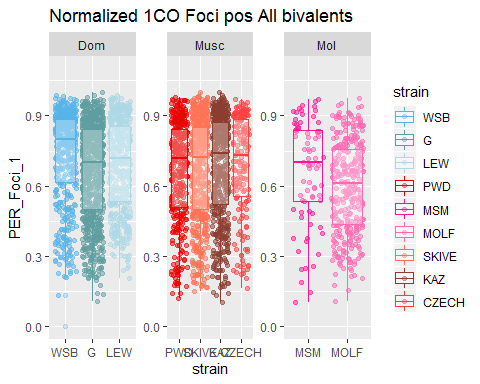
long.biv: Dom: no strain effects, MUSC no significant strain effects, MOL - not enough observations

**For the long bivalent dataset, In M1 where the subsp are tested, only Musc is significant (glm; Musc p= 0.04). The significant strain effects in M1 are slight just MSM (glm; MSM p= 0.09).**

**For the long bivalent dataset, In M2 (limited to just the strain effects) a few strain effects are significant (glm; MSM p = 0.01, SKIVE 0.07 and CZECH p = 0.04).**

**THERE IS a weak signal of the high rec strains having longer SC (having evolved) longer SC (different chromatin compaction) While the mouse mean total sc are significantly different between the two groups — this difference — translated to the reduced bivalent data sets (p = long p = short)**

# Q2.1CO rec landscape evolution is decoupled from gwRR evolution



The normalized 1CO position is not significantly different between the high and low rec groups for the total pooled data (t.test; p = 0.24) and also when examined within subspecies (ttest; p = 0.41 and p = 0.07 for *musculus* and *molossinus* respectively). While there were significant strain effects for *domesticus^WSB* and *molossinusMOLF* (figure X), this evolution of the 1CO positioning is decoupled from the total genome wide recombination rate.

# Q2 Evolution of interference is associated with genome wide recombination rate evolution

Unlike total SC length and 1CO pattern, IFD metrics are significant predictors of the rapid genome wide recombination rate evolution. However, instead of a negative correlation between interference strength and number genome wide recombiantion rate (higher gwRR associated with weaker interference, interference strength is stronger in the high rec strains.

Unexpectedly, the mouse averages for both IFDraw and IFDPER were significantly longer in the high rec groups (t.test; p = 7.7410^{-7} for IFDPER and p = 8.7810^{-6} for IFDraw). The pattern is confirmed with t.test comparing groups within *musculus* and *molossinus* the same subepcies (add these ttests and p values.) (t.tests p =2.0410^{-5} and p= 0.17).

We test the within subspecies predictions using the model framework

**Linear model M1**

**Linear model M2**

In M1, the model build on the mouse averages of IFDPER metrics with subspecies and strain as fixed effects, the high rec strains had significant strain effects for (glm; p = 0.01, p = 0.1 and p = 7.0510^{-5} for *musculusPWD*, *molossinusMSM* and *musculusSKIVE*, respectively).

In the M2 model results, *musculusPWD*, and *musculusSKIVE*, were the only significant strain effects, (glm; p = 0.01 and p = 6.6210^{-5}).

Running M2 within subspecies â confirms the within subepcies predictions across strains â (no significant strain effects

Q2.glm.M2\_IFD.PER\_Dom

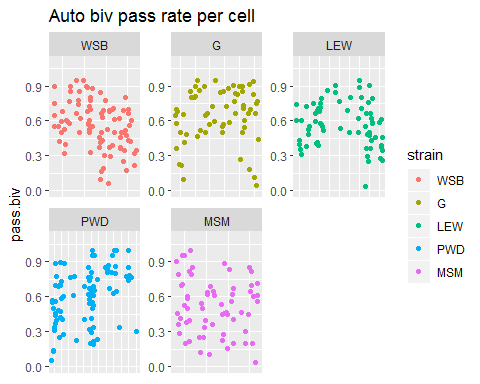
Q2.glm.M2\_IFD.PER\_Musc

Q2.glm.M2\_IFD.PER\_Mol

We determined that the main difference in IFDPER disbribution across the high and low groups is an enrichment of IFDPER observations under 30% in low rec strains.

The rate of IFDPER below 30% range from 8.2% ( *domesticusG*) to 16% ( *musculusKAZ*) in the low rec strains, while the high strains all had rates under 5% (0%, 1.3%, and 3.3% for *musculusSKIVE*, *molossinusMSM*, and *musculusPWD* respectively).

# Blank1



For the Automated data set, I like to measure the rate of passing bivalent per cell. The mean pass rate will be multiplied to the estimated XX mean\_SC.

The table above shows the number of bivalents from the same strains as in the manual whole cell data. The plot shows the bivalent passing rate across all of the individual cells from this female data set. For each strain, I’ll calculate the mean bivalent passing rate (maybe I should look at the mouse levels).

(some of the mice have different ranges of per cell passing rate) - given this ranges, i think the xx adjustment factor should be called on the mouse level. (it could even be extended to cell level – except i don’t think the XX SC length estimates wont be good.)

strain.XX.adjustment.factor = per\_cell\_passing rate \* 1 of 20 random biv will be XX \*

\*\* It might be simplier to compare the male and female means, and test it they are greater than the whole cell proprotion of the XX in females cells.\*\* The XX in a whole female cell contributes ~ 7% of total SC, if the female means for a type of total SC measure are from XX. But I am not using ‘whole cell’ summaries to compare female and males.

**What is the effect of an extra XX-autosome on single bivalent means?**

use a permutation approach: Make a True data set to start with, same(similar) number of cells, mice and bivalents. Make fake data sets which sample 19 bivalents, for ‘in silico’ cells for males and females. Also Run cntrl-female data set, where 20 bivalents are sampled, but randomly. Run the same bivalent level summaries for each ‘permuted data set’; male avSC, 19Female\_avSC, and rand.20\_Female\_avSC. The difference between the rand.20 and rand.19 female -permuted data sets should indicate the influence of having an extra ‘XX-autosome’ in the total data set.

# Blank2

I present heterochiasmy as a comparison of oocyte to spermatocyte MLH1 counts, but the sex chromosomes/bivalents complicate this comparison. In females the XX bivalent is indistinguishable from the autosomes. To the meiotic recombination machinery, it is an autosome and has a similar REC landscape. Whereas in spermatocytes the XY bivalent is visually distinct and any MLH1 where not included in the count). (I note if the and Y are paired, which they are at a high rate). The XY pair triggers a response to un-paired chromosomes and only has MLH1 foci within the PAR (the the tips of X and Y). To make a more equivalent comparison I will estimate which bivalent is the XX in oocytes, and subtract that average REC from the category average of each strain.

1. Compile full-cell data from females (all 20 bivalents measured)
2. Look at the SC length -ranked data, extract the 3rd longest estimate average REC for this bivalent,
3. check how variable the REC is across the 1st,2nd,4th, and 5th are.

According to mouse genome website, the X is the 3rd largest chromosome by total amount of DNA (Mb).

(Put the XX adjustment section here)

There is now MOLF, which has female biased hetC 3 of my Musc strains have male biased patter; SKIVE, PWD and MSM. 1 of the musc strains has female biased heterochiasmy, KAZ.

The mouse specific scatter plots aren’t show here because there are too bulky. These plots are in a different document.

Making all of these scatter plots, allows us to look at the whole distributions of the data for each mouse. The distance of the red line from the black could be a indicator of slides or mice with slide specific technical noise.

#try remaking the plot Megan suggested  
# for 2CO positions, Foci1, Position on x and Foci 2 position on y  
  
CurBivData\_2CO <- Curated\_BivData[Curated\_BivData$hand.foci.count == 2,]  
  
CurBivData\_2CO <- CurBivData\_2CO[!(is.na(CurBivData\_2CO$Foci2) | CurBivData\_2CO$Foci2==""), ]  
  
#isolate 2COs  
#facet by sex and subsp  
  
F1.x.F2 <- ggplot(CurBivData\_2CO, aes(x=Foci1,y=Foci2, color=strain) ) + geom\_point()+ facet\_wrap(~sex)+ggtitle("test plot")

–>

# References