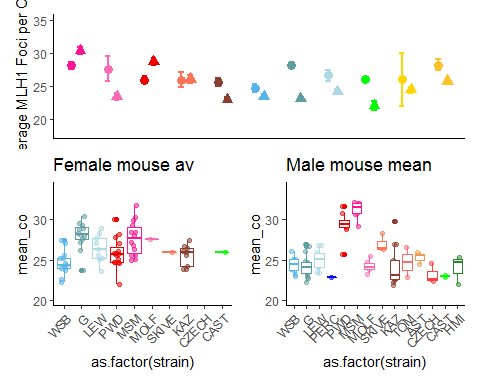
MLH1 Results Outline

3/1/2020

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 for each mouse 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 in females compared to males (female biased heterochiasmy) and ii) the degree of heterochiasmy (Female:Male ratio) is generally low, ranging from 1.17 in *domesticusG* to 1.02 in *domesticusLEW*. Three notable exceptions of male biased heterochiasmy are 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

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).

**M1. Mixed Model**

**M2. Linear Model**

**M3. Linear Model**

**M4. Linear Model**

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, were either weak or not significant (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 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 recombining’ 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 using 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.

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

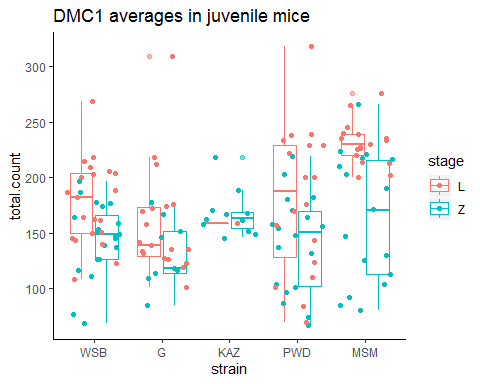


Table X. DMC1 foci counts per cell summary

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 double strand breaks (DSBs), DMC1, in prophase spermatocytes. 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 recombining 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 leptotene 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 DMC1 counts at this stage 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, partially known as the NCO:CO decision. The ratios, calculated for DMC1 means from both stages, are not significantly different between the high and low strain groups (t.test; p = 0.94 and p=0.11 for leptotene and zygotene ratios respectively).

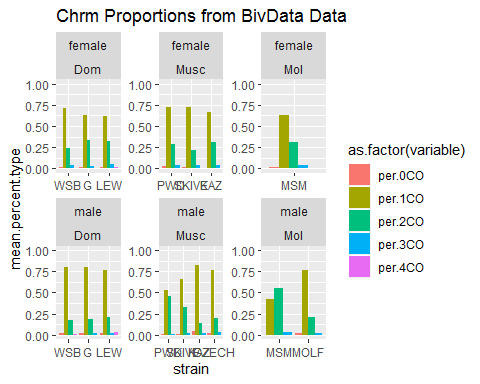
# Single Bivalent Level Results

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 is that not all bivalents per cell are isolated, due to overlapping chromosomes. The range isolation rates per cell in this data set is 0.51 in *molossinusMSM* male to 0.72 *musculusKAZ* female. However, 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.

A limitation of these data from cytological stains are the only chromosome which can be accurately identified is the XY in males. The unpaired XY chromosomes have SYCP3 signal along the chromosome axis, but it forms a structure distinct from the synapsed autosomes, which can be excluded in the curation step. Since it has homologous pair in oocytes, the XX bivalent is indistinguishable from the rest of the autosomes. By physical length in Mb, the X is predicted to be the 3rd longest bivalent.

# 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 of pachytene cells, our chromosome observations are fully synapsed paired homologous chromosomes (4n copies of each chromosomes). Additionally we can not fully distinguish the identity of chromosomes (e.g. Chromosome 1 or Chromosome 2), for these reasons we use the term ‘bivalent’ instead of chromosome.

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 recombining 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 recombining strains the 2CO proportions are 0.33 ( *musculusSKIVE* ), 0.44 ( *musculusPWD* ), and 0.53 (*molossinusMSM* ).

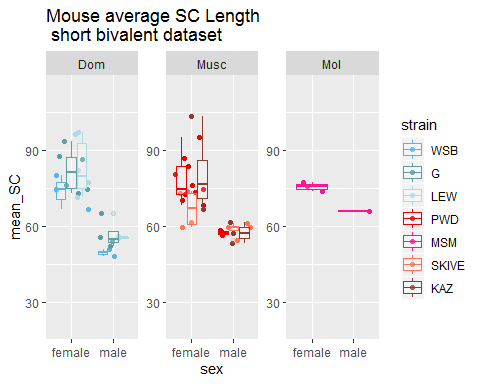
While the proportions are significantly different (chi-square-test; p = 0.06), males *musculusPWD* and *molossinusMSM* have 2CO proportions more similar to each other than strains from the same subspecies ( *musculusPWD* and *musculusKAZ* chi-square test; 2CO p=3.1510^{-33}) and *molossinusMSM* and *molossinusMOLF* chi-square test; 2CO p=4.7210^{-13}). Using this data set of single bivalents, we focus on aspects of the recombination landscape to 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 recombination groups in males and if the differences are significant, our second step is to apply the same sets of models applied to the mean MLH1 foci counts to test for effects due to subspecies or strains.

# Q1 Sex Differences in SC-AE Lengths

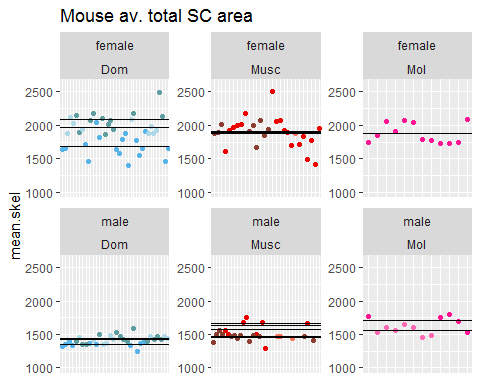
In many mammalian species the SC lengths are longer in females, yet the majority of these observations come from species with female biased heterochiasmy. Our sample of house mice provides an opportunity to test if the sex differences in chromatin compaction (SC length) are reversed in the high recombining strains with male-biased heterochiasmy; *musculusPWD* and *molossinusMSM*). 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 from pooled samples of bivalents, 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. 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 but one strains have significantly longer mouse mean for short bivalents in females (t.test; p <0.05). The differences, (F:M), across strains ranges from 1.15 in *musculusMSM* to 1.49 in *domesticusWSB*. While the female value is longer, it is not significantly different from males in *musculusSKIVE* (t.test; p = 0.11).

For an additional metric of chromatin compaction, we tested a whole cell measure of SC length using a second image analysis algorithm which converts the SC signal into single pixel wide trace of the SC shapes and provides accurate estimates of the summed SC length of bivalents for single cells (ref Wang). We apply the skeletonizing algorithm to all MLH1 meiocyte 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; p < 0.05). To test for strain and subspecies specific effects on the sex difference in SC lengths we apply the model framework.

Longer Sc compaction in females is supporte by the model framework, with sex being the most consistant significant effect (LTR and glm; p < 0.05). Additionally there are some significant subspecies and strain effects indicating that the SC lengths can evolve within strains and subspecies (LTR and glm; p < 0.05).

These combined results of total SC per cell and reduced single bivalent SC metrics, indicate that females have longer SC lengths (chromosome axis), even in strains where males have more MLH1 foci per cell. I the means MLH1 foci per cell and SC area per cell, are considered seperatly, it is as if the males in these strains have more crossovers (more 2CO), but generally less ‘space’ to place them.

This leads us to 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

For most of the house mouse strains, 1COs are the major class of chromosomes observed across cells. There are known sex differences in the relative placement along chromsomes this section tests if there has been evolution in this aspect of the recombination landscape.

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}).

The differences in 1CO landscapes due to sexes is confirmed by the model framework, with sex is the most significant effect in the mixed model (M1) (LTR; p =1.2610^{-25}) and the two linear models (glm M2; p =1.3310^{-7}, glm M3; p = 1.3310^{-7}).

# 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. To quantifiy how crossover interference differs between the typical recombination landscapes of males and females, we isolated 1360 and 1272 2CO bivalents in females and males respectively. Mean interfocal distances (IFD) were calculated from 42 female and 45 male mice. We examined both raw IFDraw and normalized by SC length (IFDnorm). 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 2004).

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.

While the comparisons of the IFDnorm metrics can reveal more general recombination landscape patterns while controlling for the underlying differences in chromatin compaction and SC length. 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 IFDnorm are highly significant between sexes. The mean IFDnorm 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 IFDnorm (t.test; p-value = 0.33).

In examining the IFDnorm distributions, we note that the female IFDnorm observations are centered at approximately 50% and have slight an enrichment of short (<25%) IFDnorm observations. Comparitively, the IDFnorm distributions in males are enriched for longer IFDnorm.

The three sets of models support the result of stronger interference in males, sex was the most significant significant effect (LRT: p = 6.7410^{-14}). When interference is measured in physical SC units (IFDraw), the differences between sexes is low and only slightly significant (data not shown). These results support the model of the physical measures of interference (in SC units) is conserved between sexes inwithin the same species.

# Q1 Summary, consequences of sex differences in the recombination landscape

Our results confirm sex differences in recombination landscapes described in several other species and we note that these differences have distinct consequences on the potential patterns of genetic variation resulting from the distinct recombination landscapes.  
The terminal position of single foci and the greater distance between two foci on the same chromosome will have a consequence of larger sections of linked sites segregating together in male gametes compared to female gametes (Veller).

The next section is meant to focus on the greater aspect of variation in mean MLH1 counts per cell the high recombining strains ( *musculusPWD* , *musculusSKIVE*, and *molossinusMSM*) to the low recombing strains. The main objective of this section is to test for significant correlations between features of the recombination landscape and the evolution of mean MLH1 foci per cell.

In comparing the male specific single bivalent based metrics the first creteria for analysis is a significant differences between the high and low recombining groups and the second step is testing for significant subspecies and strain effects in the same models which used mean MLH1 count per cell.

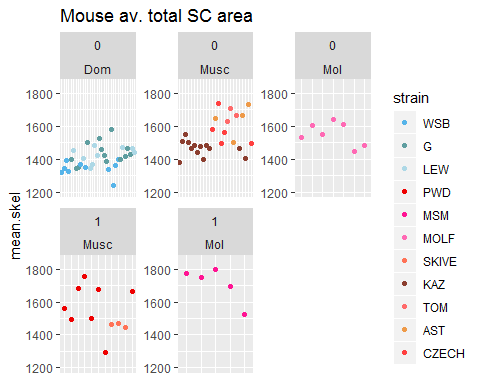
**Linear model M1**

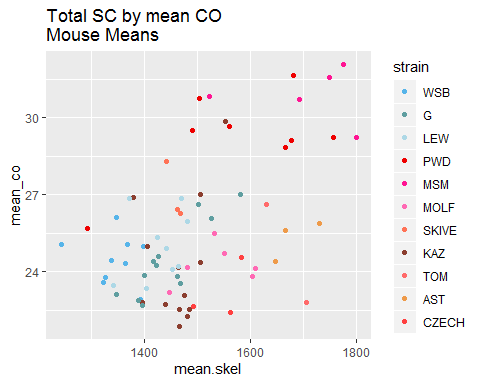
**Linear model M2**

The literature provides basic predictions for the direction of the relationship the genome wide recombination rate and two aspects of the recombination landscape; SC length (positive) 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

Given the differences in the DSB numbers between a subset of the high and low recombining strains we predict that there will also be differences in SC lengths proportional to the increase in the mean genome wide recombination rate. 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.

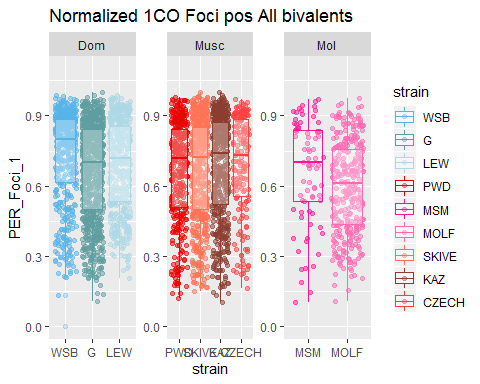




Confirming the basic predictions, there is a positive correlation between mean MLH1 foci per cell and total SC, however the variation in mean total SC does not completely predict the high and low recombining strains (Figure X). While for the total data set of mean total SC, the high recombining strains have significantly more SC area (t.test; p = 0.01), only the *molossinus* strains are significantly different between the high and low recombining strains (t.test mol p = 0.03). The *musculus* are not significantly different across the groups (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 recombining strains (t.test; short p = 0.88 and long p = 0.18).

For the linear model testing subspeices and strain (M1) and total SC as the dependant variable, the two subspecies effects are significant (glm; p= *Musculus* 1.2410^{-6} and *Molossinus* p= 10^{-6}). Whereas the model with reduced bivalent means as the dependant variables, a number of subspeicies and strains effects reach significance, (glm; p < 0.05) they are not consistant across the sets of models tested, indecateing to some extent the chromation compaction evolution is decoupled from evolution in mean MLH1 foci per cell.

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



The normalized 1CO position is not significantly different between the high and low recombining groups for the total pooled data (t.test; p = 0.24) and also when examined within subspecies (t.test; 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. Hence we don’t follow up with the model framework.

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

Unlike total SC length and single foci landscapes, our metrics of interference from 2COs are significant predictors of evolution in genome wide recombination rates. However, the pattern is in the opposite direction to our prediction; the high recombining strains have significantly stronger interference (instead of being predicted weaker interference). In the 2CO landscape, instead of observing weaker interference strength and two foci placed closer together on a single chromosomes, the two foci are spaced further apart.

The mouse averages for both IFDraw and IFDnorm were significantly longer in the high recombining groups (t.test; p = 7.7410^{-7} for IFDnorm and p = 8.7810^{-6} for IFDraw). The pattern is confirmed with t.test comparing strains within *musculus* and *molossinus* (t.tests p =2.0410^{-5} and p= 0.17 for IFDnorm and 1.9410^{-4} 0.08 for IFDraw).

The greater IFD measures in high recombing strains is confirmed by fitting both the IFDraw and IFDnorm into the linear model framework, where only the strain effects for the high recombining strains had significant p values (glm; p < 0.05). The similar patterns across the raw and normalized IFD values rule out differences in IFDraw measures, SC lengths and chromosome size as the major driver of this pattern. We determined that the main difference in IFDPER distribution across the high and low groups is an enrichment of IFDnorm observations under 30% in low recombining strains. The rate of IFDnorm below 30% range from 8.2% ( *domesticusG*) to 16% ( *musculusKAZ*) in the low recombining strains, while the high strains all had rates under 5% (0%, 1.3%, and 3.3% for *musculusSKIVE*, *molossinusMSM*, and *musculusPWD* respectively).

# Q2 Summary

Our results show that the greater crossover interference is the strongest single bivalent-based predictor for the observed rapid evolution of mean MLH1 foci per cell. While these results do not conform initial predictions for how a recombination landscapes would accommodate more crossovers, the increased strength of interference aligns with our results on sex differences. The typical recombination landscapes for males and females results in divergence in the proportion of linked sites along chromosomes which segregate together. The stronger interference of the 2CO bivalents in the high recombining strains accentuates this effect. The measures of DSB and some comparisons of SC length between high and low recombining strains suggest that the SC length have evolved to be longer in high recombining strains, however this evolution of SC length is partially decoupled from the number of crossovers since similar amounts of SC length evolution are seen in low recombining strains.

# References