MLH1 Results Outline

Table of Contents

[Main Figure 2](#_Toc33634072)

[Genome wide recombiantion rate estimates for both sexes 2](#_Toc33634073)

[Evolutionary Framework 3](#_Toc33634074)

[Within Mouse Variance in CO Count per Cell 4](#_Toc33634075)

[Evolution of Genome wide recombination rate associated with evolution of mean precursor number 5](#_Toc33634076)

[Evolution of genome wide recombination rate is reflected at the single chromosome level 8](#_Toc33634077)

[Single Bivalent Level Results 9](#_Toc33634078)

[Q1 Sex Differences in SC-AE Lengths 9](#_Toc33634079)

[Adjusting for XX 15](#_Toc33634080)

[short transition 15](#_Toc33634081)

[Q1. 1CO Position Sex Differences 15](#_Toc33634082)

[Q1. Sex Differences in CO Interference (IFD) 19](#_Toc33634083)

[Mixed Model Tests, Fixed Effects 23](#_Toc33634084)

[transition Q1 to Q2 25](#_Toc33634085)

[Q2 SC Length 25](#_Toc33634086)

[brief transition, rec landscape 27](#_Toc33634087)

[Q2.1CO rec landscape evolution is decoupled from gwRR evolution 28](#_Toc33634088)

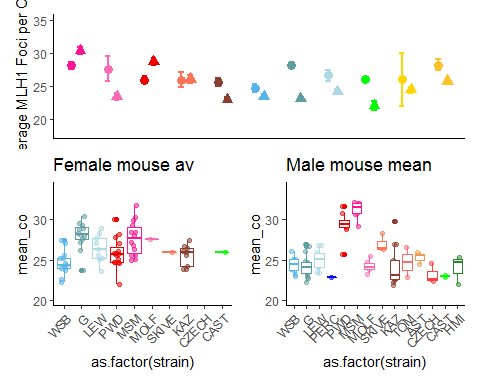
[Q2 Evolution of interference is associated with genome wide recombination rate evolution 29](#_Toc33634089)

[Blank1 30](#_Toc33634090)

[Blank2 32](#_Toc33634091)

[References 33](#_Toc33634092)

# Main Figure



This is a draft of the main figure. Outgroups are not included in the analysis below. (add horizontal lines for denoting subspecies, group means and expected minimum)

# Genome wide recombiantion 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) that the genome wide recombination rates averages are greater females compared to males (female biaded 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 seperately for each sex reveal disctinct patterns of variation (Figure1 B and C). The female means can be discribed as more evenly distributed around the sex-wide mean of approximately 25 MLH1 foci per cell (Figure1 B). While in males, the strain specific means seperate more clearly into two groups of high genome wide recombination rates near 30 MLH1 foci per cell and low genome wide recombiantion 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 a 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 recombiantion 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, ( p = 0.06 and *molossinusMSM* strain 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*. No significant sex effect showing that the sex effect on mean CO number is not uniform across the sample but has synergy with speciefic strains.

A linear model was run on the sex specific data sets (M4), with 192 number of male and 144 female mean MLH1 foci counts were used in each of these models. For 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.

three of the strain effects of 13 strains had significant effects on mean MLH1 per cell, *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).

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 four to one, indicateing specific, but small strain effects for the mean number of MLH1 foci per cell.

## Within Mouse Variance in CO Count per Cell

Variance of a trait can contain information pertinant to it’s mode of evolution and mechanistically could reflect the strength of constraints acting on the trait. 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 dependant variable).

**the general pattern is that females have more variance (alsmot twice as much (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} ).

The sex specific effect linear models indicate that there is no significant difference in the amount of within mouse variance in males. While in females, one strain, *domesticusLEW* has significantly higher within mouse variance (glm; p = 0). However, in the model using the high quality cells, *musculusPWD* has the only signicant strain effect (glm; p =0.04) suggesting a lack of a consistant signal for differences in the amount of within mouse variance in oocytes.

**the general pattern is that females have more variance (alsmot twice as much (Figure 1))**

(combined the summary of the MLH1 distributions evolution in the mean and consistant pattern of sex differences)

1. – the striking differences / patterns of variation across sexes suggest that the genome wide recombination rates have distinct evolutionary trajectories for male and female rates. For female the pattern of Strain averages distributed around a species wide average - fits a model of stabilizing / relaxed / neutral evolution in contrast to the male pattern where there is rapid evolution in a subset of genetic backgrounds / strains which fits a model of directional selection on genome wide recombination rate.

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

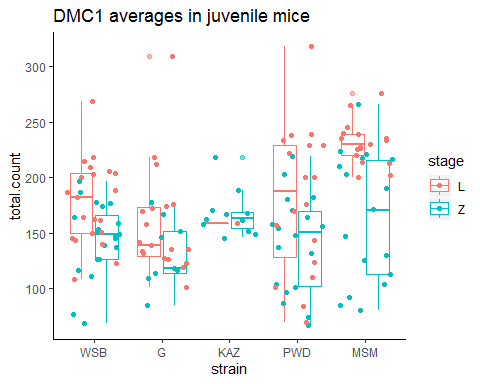


Table X. DMC1 foci counts per cell summary

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 early prophase spermtaocytes. DMC1 foci were scored from a total of 76 leptotene and 75 zygotene staged spermatocytes form juvenille mice (12 to 18 days). A subset of strains were choosen to test for differences between the high and low recombination groups; *musculusPWD*, *molossinusMSM*, *musculusKAZ* , *domesticusWSB* , and *domesticusG* (Figure2). One mouse was sampled for each strain.

For all strains there were significantly more foci from leptotene cells compared to zygotene (t.test; p = 1.0310^{-5}), which reflects the expected patterns across the timing of meiosis. Confirming the prediction that evolution of the mean number of crossovers 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 the differences in DMC1 foci were not significant for (later prophase) zygotene cells between the two groups (t.test; p=0.66, one-way-anova; p=0.15). Indicating that this marker at early prophase is more precitive of the downstream crossover number differences.

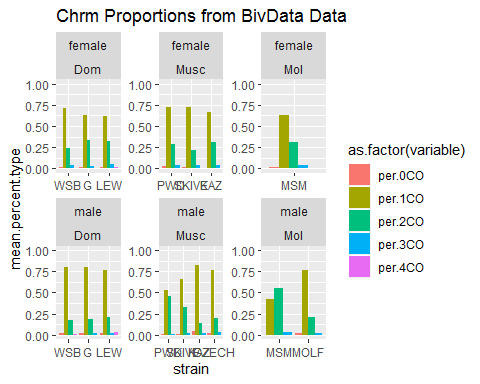
after DSB formation, DSBs are repaired as either NCO or CO (CO designation step) The vast majority of DSBs are repaired as NCOs, with a minority (in most species) of approximately 10% are repaired as crossovers.

-DSBs are either repaired as crossovers or noncrossovers (NCO), with conversion tract lengths (30 to 100bp long). (this is also known as the CO repair deicison )

thus the ratio of CO:DSB is in part an indicator of the proportion of DSBs which are designated into COs – (and the others are repaired as NCO).

The ratios, calculated for DMC1 means from both stages of prophase, are not significantly differet between the strain groups (t.test; p = 0.94 and p=0.11 for leptotene and zygotene based ratios respectively). These results indicate that the targets of evolution which lead to differences in crossover numbers may be established before DSB formation (cite Baier et al 2014).

# 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 chromosomes (bivalent). Because our observations of MLH1 foci are made in pachtyene cells, our chromosome opbservations are fully synapsesed paired homologous chromosomes (i.e. 4n copies of each chromsomes). Additionally we can distinguish the idenetiy of indiviual chromsomes (Chromsome 1 or Chromsome 2) for these reasons we use the term bivalent.

Of the pooled dataset of 34982 single bivalents, 96% of bivalents had either one or two crossovers (Figure X) with proportion of 1CO:2CO distinguishes the high and low rec males (Figure X). THis confirms the intuitive interpertation that high recombination strains are enriched for 2CO bivalents at the expense of 1CO bivalents. Across the female observations the proportions of chromosome classes are not notably different.

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 (p = r), PWD male and MSM male proportions are more similar to each other than strains from the same subspcies ( *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

In order to deconstruct the cell wide MLH1 counts deconstruct we look at the recombination landscape at the single bivalent level.

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 measures were left.

We use these observations to look at the recombination landscapes, here refereing to the i) total length of the SC ii) the position of single crossovers along bivalents and iii) the spacing between multiple COs on the same bivalent.

A potential downside of the image analysis algorithm we used is that not all bivalents per cell are isolated. The ranges of rates of isolating chromosomes per cell are 0.51 in *molossinusMSM* male to 0.72 ) *musculusKAZ* female, we assume that the isolation process is not biased. However, because there are hundreds of observations per category, we assume the set of single bivalents 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?

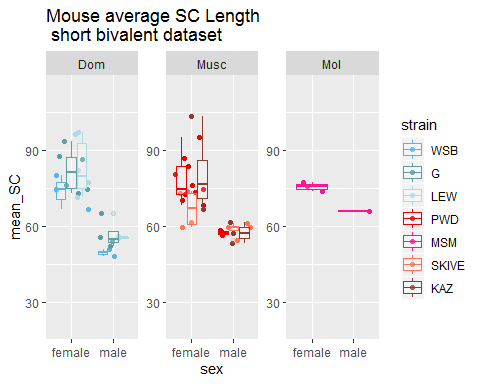
# Q1 Sex Differences in SC-AE Lengths

(in many mammal species – SC lengths are longer in females - yet most all of these observations come from female biased heterochiasmy species) – Our data set provides an opportunity to test (how SC length and chromatin compaction relates to heterochiasmy)

if the canonical female biased heterochiasmy pattern and long SC lengths are uncoupled in instances / our strain with male-biased heterochiasmy ( *musculusPWD* and *molossinusMSM*). A simple model might predict if SC length is the strongest predictor for number of MLH1 foci er cell the high recombining males would have longer SC length proportional to differences in the number of MLH1 foci per cell.

To account of the obvious confounding effect of sex chromosome differences in females and males, we use a reduced bivalent data set made up of chromsomes (below the 2nd quartile for SC length of single cells) to make single bivalent groups which exclude the XX homologs from the female data. A total of 678 ‘short’ bivalents were isolated from 140 oocytes and spermatocytes. Even though this smaller data set has decreased power due to redcued sample size, it offers a mors comparable set of SC-AE meterics across the sexes.

This ‘short biv’ data set has advantages in that it is closer to sampling across a similar set of chromosomes (ideally Chromosomes 15 to 19) and it excludes the XX in females, which is predicted to be the 3rd longest based on the physical (Mb) length.



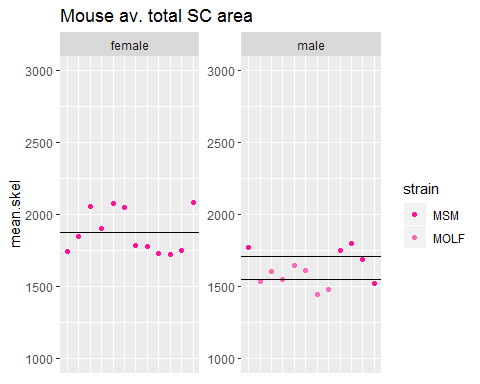
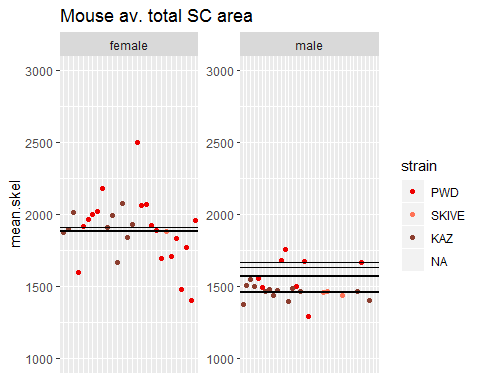
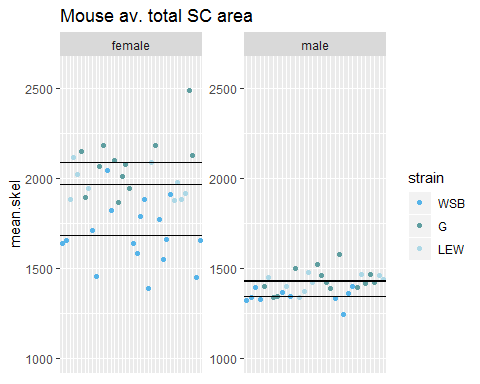
For all but one strain, *musculusSKIVE*, the mouse mean for the short bivalents are significantly longer   
**r the magnitude differences between the mouse mean short bivalent SC range from X to X**

(t.test; p = 0.02, p =0.00049, and p=0.0016 for *domesticusWSB*, *domesticusG* and *domesticusLEW* respectively. This pattern is also true for the musculus strins (t.test; p= 0.00011, 0.06, and 0.11 for *musculusPWD* , *musculusKAZ* , *musculusSKIVE* respectively.

An additional measure of a smaller subset of single bivalents (which excludes the sex chromsoomes), we tested/compared a measure of the SC lengths for the whole genome (whole cells) using a image nalysis algorithm which measures the skeletonized bivalent and provides an acurate measure of the lengths of single SC/bivalents (ref RWang)

We apply a skeletonizing image analysis algorithm to all MLH1 meiocyte spreads. A total of X images were run trhought the image analysis pipeline, and after outliers removed, X remained. Mouse means were calculated for their total SC area per cell and are displayed in figure X.

The comparison of this metric is somewhat complicated across sexes due to the sex chromosomes. Breifly/in effect the oocytes have an additional autosome, and the spermatocytes have a mostly recombinationly inactive unit of SC (the XY).



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, KAZ p = 0.00000087, *molossinusMSM*  0.027

<**t.test results for total.SC** all values (F vs M) (p= 1.4710^{-26}) Dom (F vs M) – (ttest; p=4.1910^{-15}) Musc (F vs M) – (ttest; p=1.1510^{-9}) (note, missing SKIVE, not enough female measures)  
MSM (F vs M) – (ttest; p= 0.03)>

We follow up the general result of females have longer SC lengths – by applying the same model frameworks which we used to analyze the mean MLH1 foci per cell – to understsand the evolutionary framework.

**M1. Mixed Model**

**M2.Linear Model**

**M3. Linear Model**

In the mixed models for the sex differences mean **short bivalents** length as the dependant 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).

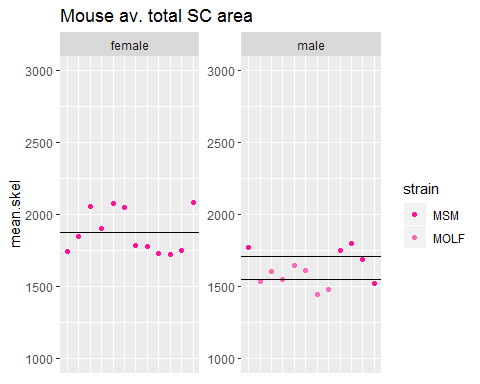
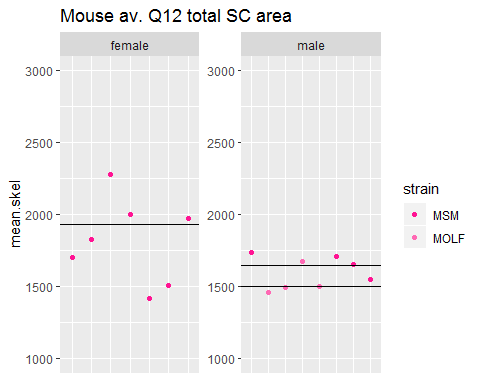
For the mixed model run on the mean total SC metric, there are more significant effects for strain and subsp for the mixed model using mean **total sc** per cell as the dependant variable, the three fixed effects and one random effect were significant (LTR; subsp p= 0, sex p = 1.210^{-31}, and interaction p = 7.1210^{-4}, (random LRT; strain p = 0).

In linear models (M2 and M3) for **mean short bivalent length**, male is most significant (M2 glm; p =0 and M3 glm; p = 0). In the linear model M2 the skive strain and strain by sex effects are slightly significant (glm; p=0.03, p =0.09 respectively), however these could be due to low sampling of skive females.

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 / quantified at this level.

<1. the longer female sc’s is consistant and seen across all strains, even in the high rec strains, suggest a decoupling of sc and gwRR - when comparing across sexes – or the sc area is used differently for CO in oocytes and spermatocytes. 2. when the total sc per cell is compared, domesticus males have significantly lower SC arae compared to molossinus and musculus strains.Except in Q2 I Say this isn’t true.>



## Adjusting for XX

Summarize for now, refine this section later.

1. illustrate problem(affects mostly SC length)
2. Expected impact on sex comparisons, estimated effect size of the X
3. (prove general pattern that ALL bivalents are longer), chrms sorted by bin comparisons
4. permutations of 19 female, 19male, 20female 20 male

# short transition

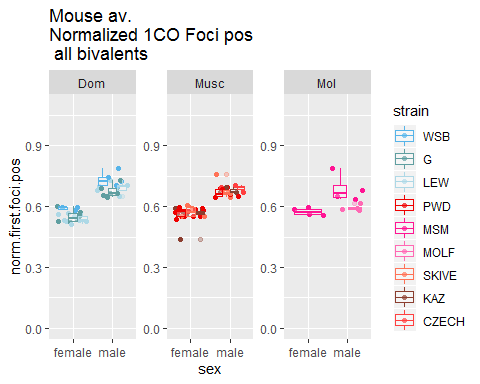
-following up on SC area - with how COs are place - placement and number of COs on single bivalents

1. Transition, — the SC-AE area can be thought of as the ‘area / available real-estate for COs – these results suggest that this area is ’utilized’ in different ways across sexes and strains - so we next investigate the recombination landscape - (ie the relationship between the placement and number of COs along chromosomes).
2. We focus on two metrics/ aspects i) the placement of single foci along a bivalent (just 1CO) and ii) the placement of two foci on the same bivalent / the interfocal distance of 2CO bivalents (reflects crossover interference)

# Q1. 1CO Position Sex Differences

We focus on the foci positions from one crossover (1CO n= r) bivalents since the landscape patterns for multi crossover bivalents (2COs) will be highly influenced by crossover interference.

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 the domesticus strains the mean normalized foci positions were significantly different between sexes and fit the predictions of the ‘typical recombination landscape’ (with female positions being medially placed and terminally placed in males). The p.values across the strains are; (t.test; p = 4.0610^{-4}, 1.6710^{-5}, and 6.9110^{-5} for WSB, G and LEW respectively).

With the musculus and molossinus strains displaying a similar pattern (t.test p = 9.2710^{-5} and 0.04) PWD and MSM. Two musculus strains had a weaker signal, but it was still in the same direction (t.test; p = 0.01, and 0 for KAZ and SKIVE respectively).

After confirming the significan sex differences for 1CO position, we apply the model frameworks

**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). These results confirm the sex difference establish in the t.tests above – and suggesting variation across the strains (in the sexual dimorphism for this pattern).

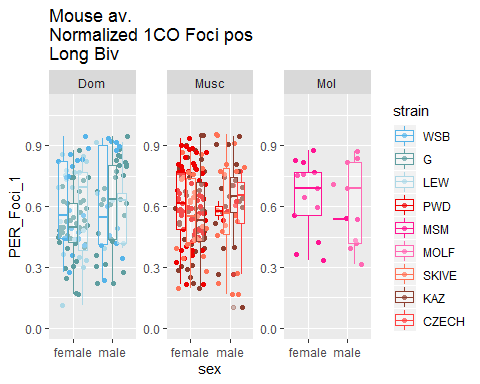
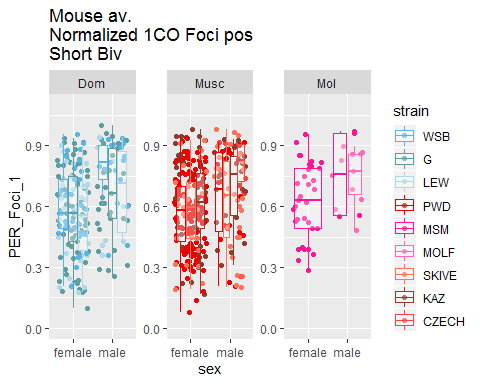
The male effects were the most significant (glm M2; p =1.3310^{-7}, glm M3; p = 1.3310^{-7}). While the musc subspecies, LEW and KAZ strains had slightly signifcant effects (glm M2; Musc subsp p = 0.07, LEW strain p = 0.03) and (glm M3; LEW strain p = 0.03, KAZ strain p = 0.07).

(these results will be explored more fully in the Q2 section). The general pattern that emerges across these models is the sex effect being the most significant, (with the male single foci position being more telomeric compared to females).

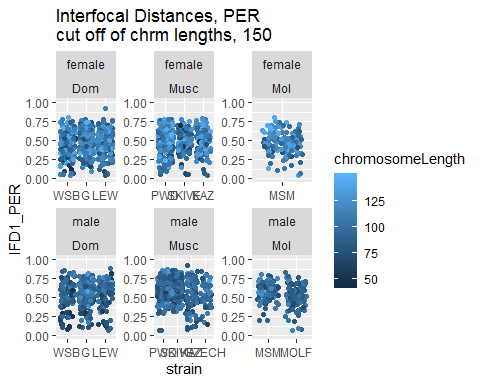
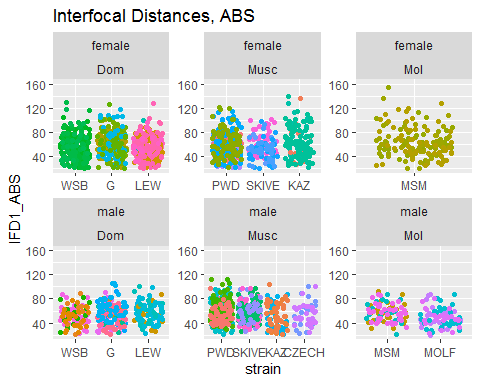
REMOVE THE SISCOTEN PLOTs…

## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 2 rows containing missing values (geom\_point).



# Q1. Sex Differences in CO Interference (IFD)



Crossover interference is one of the major determinates – of the CO landscape. We use X number of 2CO bivalents, to test if there are diffferences between males and females in interference. The mean interfocal distance (IFD) for raw and normalized values was calculated from X female and X male mice.

The raw IFD measures – are usuful for thinking of crossover interference as a mechnical force mediated through the SC. Previous measures found no sex difference in IFD in B6 mice (de Boer et al )

In the abosolute plots, you can start to see clustering/ enrichment of IFD values between 40 and 80 pixels for PWD and SKIVE, (and maybe slight clustering for MSM).

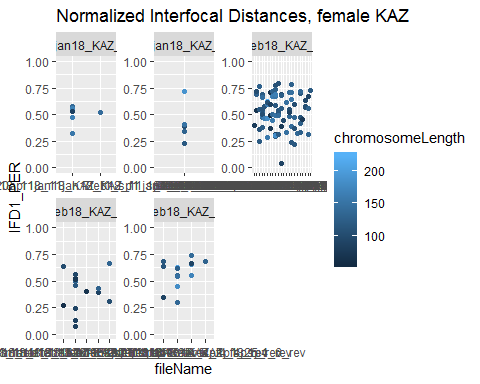
t.tests are performed on the normalized metrics for establish if there are sex differences in the IFD meterics (raw and normalized).

The raw IFD measures for 2CO bivalents are slightly significant sexes (t.test; p = 0.07). BUT WHEN G IS REMOVE, the pvalue is no longer significant (t.test; p = 0.27).

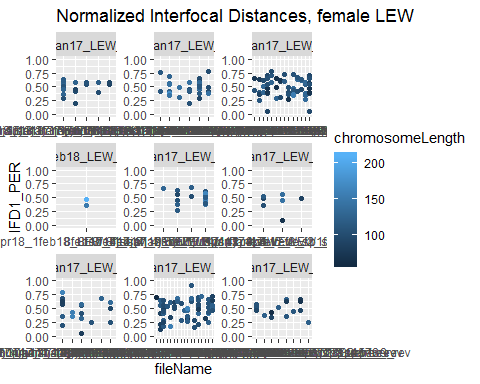
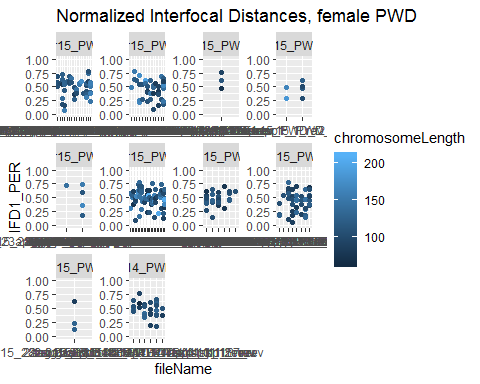
The normalized IFD measures are highly significant (t.test; p = 1.4910^{-12}) between sexes, indicating there a

When the t.tests were performed within each strain, all but KAZ show (p-value = 0.3), significant sex differences in the normalized IFD measures MSM r SKIVE r PWD r MSM r LEW r G r WSB r

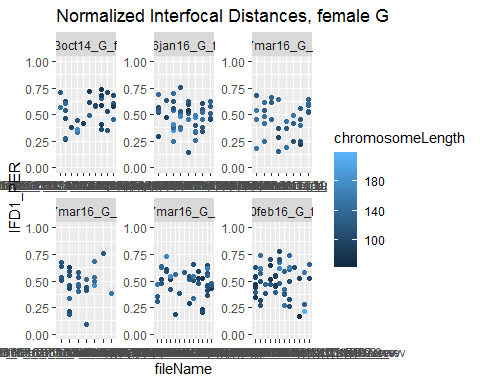
This indicates a general pattern of two foci are seperated by more area (hance have stronger interference) in males, this pattern is also for genetic maps.



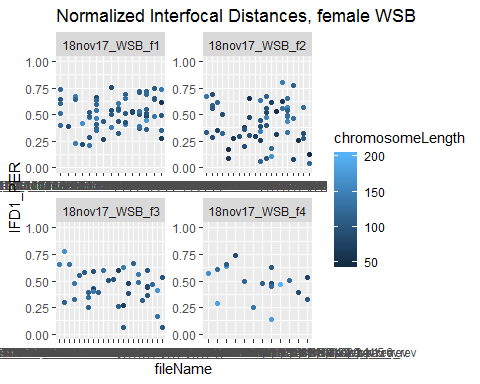
## Warning: Removed 1 rows containing missing values (geom\_point).



## Warning: Removed 2 rows containing missing values (geom\_point).



## Warning: Removed 1 rows containing missing values (geom\_point).



see the reference plots within this dir, –

this section – all strains but KAZ, males have a higher mean normalized IFDs

the female norm IFD are centered at ~50% – but they also seem to have a higher rate of short (<25%) normalized IFDs

The KAZ strain is the only strain without significant p value for the normalized IFDs, due to KAZ males having a higher rate (enrichment of short normalized IFDs.

There seems to be a 25% norm.IFD cutoff (in females) – but some mice seem to have less of this cut off. Is there a default expectation for why .25 would be the cut off…(given female uniform CO positioning?

**THE male female pattern** female close to 50% – which fits with uniform rec landscape / weak interference

to rule out chromosome size effects, (espically in KAZ) I re-run the ttests using the short and long biv data sets. Our predictions for these are that longer chromsomes will show more of the ‘typical male pattern’ while short chromosomes often have a more uniform positing (hence tighter)

Plotting the single bivalent data for the short and the long data set show in KAZ – there are two observations near the lowerest IFD1.PER, (even lower than female). These come from 2 different males and I looked up the images to confirm that the measures are accurate (by the bivalent).

Considering that the short-long bivData is much smaller than the full data set

### Mixed Model Tests, Fixed Effects

Since the most significant t.test were significantly different across sexes for the NORMALIZED effects – we follow up with the model framework to –check out the potential evolutionary patterns–

We applied the same frame work of models, Mixed model analysis for IFD (interference), the first set of models are made with the lme() functions.

**M1. Mixed Model**

**M2. Linear Model**

**M3. Linear model**

## Warning in optwrap(optimizer, devfun, x@theta, lower = x@lower, calc.derivs  
## = TRUE, : convergence code 3 from bobyqa: bobyqa -- a trust region step  
## failed to reduce q  
  
## Warning in optwrap(optimizer, devfun, x@theta, lower = x@lower, calc.derivs  
## = TRUE, : convergence code 3 from bobyqa: bobyqa -- a trust region step  
## failed to reduce q

## Estimate Pr(>|t|)  
## (Intercept) 0.4619 0.000  
## sexmale 0.0807 0.008  
## strainG 0.0258 0.368  
## strainLEW 0.0062 0.817  
## strainPWD 0.0142 0.588  
## strainMSM 0.0107 0.732  
## strainMOLF -0.0130 0.661  
## strainSKIVE 0.0199 0.526  
## strainKAZ 0.0196 0.510  
## sexmale:strainG -0.0283 0.453  
## sexmale:strainLEW 0.0059 0.879  
## sexmale:strainPWD 0.0484 0.209  
## sexmale:strainMSM 0.0213 0.622  
## sexmale:strainSKIVE 0.0788 0.056  
## sexmale:strainKAZ -0.0421 0.339

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

(IFD.PER) sex MM p = r

subsp p = r

the interaction effects were slightly significant for both raw and nrm.IFD. The most significant value was the sex effect for nrm.IFDs. The the random strain effect was not significant for either model.

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.

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.

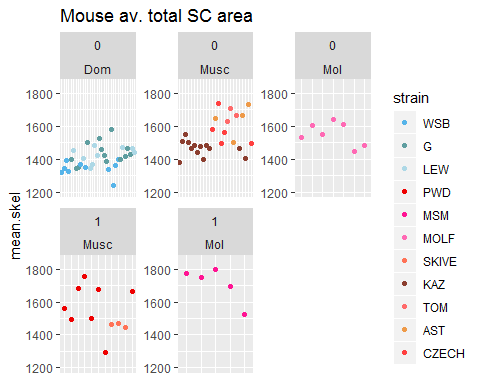
# transition Q1 to Q2

we tested the sex differences – now we focus on the largest aspect of variation, the high rec males vs the low rec males.

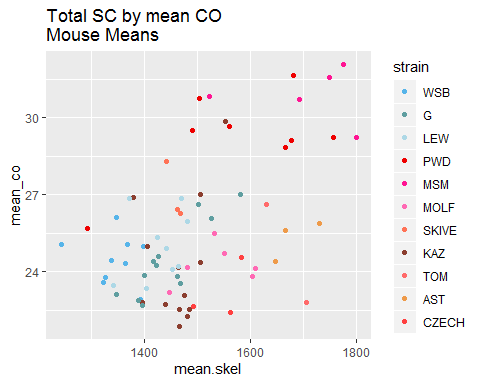
# Q2 SC Length

<motivation: the positive correlation, Our prediction for this metric is that the high rec strain will have longer sc lengths (axis lengths)>

while testing the sex differences, there were results (the significant subspecies effect) which suggest that the musculus and molossinus strains had longer SC compared to the domesticus strains - these two strains also contain the high rec strains - motivating a model/hypothesis that sc length / chromatin compaction evolution may proceed - facilitate the rapid male specific evolution in gwRR



## Warning: Removed 4 rows containing missing values (geom\_point).



**mouse means of total SC area for Higher rec strains are significantly higher compared to low rec strains (ttest; p = 0.01).**

Molossinus have significant difference between mouse means for total sc, between the high and low strains (mol p = 0.03). The total sc mouse means are not significant for the high and low strains in musc (musc p= 0.87).

**For the total SC glm**

SKIVE > the rest 3. Msm > mol>

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).

**The mouse means for the reduced bivalent datasets (short and long) are not significantly different between the high and low rec strains (for the total pooled data) (ttest; short p = 0.88 and long p = 0.18).**

**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).**

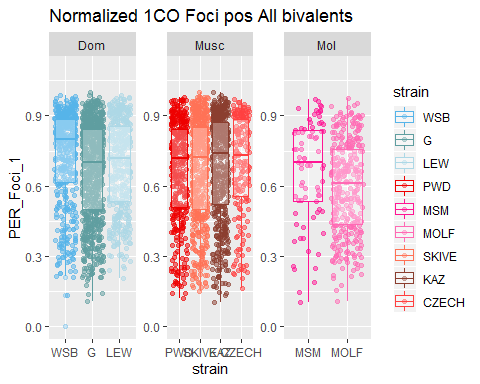
**Summary Points**

* confirm that the positive correlation of SC and CO number is maintained
* Predicted differences, between High and low strains OR between Dom and Musc/Mol, only have significant p values when the single bivalent levels are used.
* (the glms, indecate there are many strain effects), Using the mouse average SC lengths, in the full data set all strain effects are significant. This is an indication that there is more variation for the SC lengths than for gwRR / CO counts.
* For the M3 models of mouse av.s across SC meterics, for many of the SC metrics, most all of the strain effects are significant
* for the M2 models of the mouse av, musc - mol are significant fixed effects
* many strain effects, complications in the SC meterics; short/long single bivalents have lower sample size and the total SC might have noise due to the algorithm / differences in the imaging
* th predictions for M3 (strain) differencs (motivated by the gwrr variation are not met, (correlation of SC means and CO means?))

# brief transition, rec landscape

the lack of predictive power of SC length to predict CO number indicates that there might be different patterns of ‘SC usage’ / different patterns in the layout of the recombination landscape / these patterns could stem from placement or spacing of foci.

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

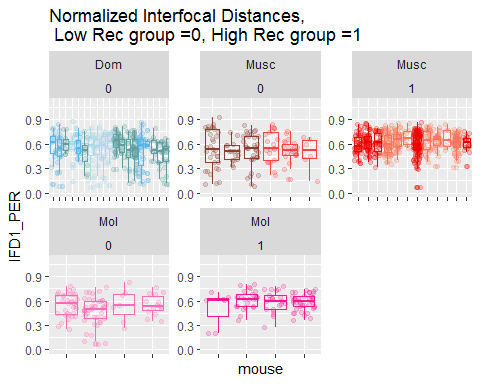


The previous section found that the single crossover landscape is significantly different between males and females. In this section we focus on differences between the high and the low rec male strains In order to test if if there are distinguishing meterics for these two groups. **The figure above shows, WSB has significantly more terminal positioning, while MOLF has significantly more medial position.**

T tests of the single crossover positions reveal that the normalized single foci positions are not significantly different between the total pooled high and low rec male strains (t.test; p = 0.24). The high rec musc strains, *musculusPWD* and *musculusSKVIE*, do not have signifcantly different single foci landscapes from the other musc low rec strains (ttest; p = 0.41). While *molossinusMOLF* is significantly more centrally positioned compared to *molossinusMSM* (t.test; p = 0.07). This result is somewhat puzzeling and it doesn’t match up with the strains with high recombination rates.

given that the t.test don’t return significant pvalues so we are not motivated to follow up with linear models. Males from WSB and MOLF strains have ’strang’unique patterns for the single foci positions, which don’t fit with predicted patterns from the genome wide recombination rate variation

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



The mouse averages for IFDs 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 the same subepcies.

The following model framework was applied to further investigate these patterns using mouse averages of IFDraw and IFDPER.

**Linear model M1**

**Linear model M2**

In M1, the model build on the mouse averages of IFD metrics with subspecies and strain as fixed effects, the high rec strains had significant strain effects (glm; p = 0.01, p = 0.1 and p = 7.0510^{-5} for PWD, MSM and SKIVE respectively). In M2, the mouse average for normalzied IFD is modeled with only strain effects. Only PWD and SKIVE IFDPER had significant strain effects (glm; p = 0.01 and p = 6.6210^{-5} for PWD and SKIVE respectively).

We determined that the main difference in IFD disbribution across these groups is an enrichment of ‘close foci’, IFDPERs less than 30% of the total SC length in low rec strains. The rate of these close foci range from 8.2% (G) to 16% (KAZ) in the low rec strains, while the high strains all had rates under 5% (0%, 1.3%, and 3.3% for skive, MSM and PWD respectively).

**(unlike total SC length and 1CO pattern, IFD metreics are a significant predictors of rapid genome wide recombination rate evolution. However, they go agaisnt a standard perdiction for the relationship between the chromosome level recombination landscape and genome wide recombiantion rate.**

BE brief,

final summary, highlight the themes of 1. more variance in female traits (MLH1 foci per cell, rec landscape / interference)

while the male traits – more focused (less variance in MLH1 counts per cell, shorter chromatin comparction? bias position, stronger interference)

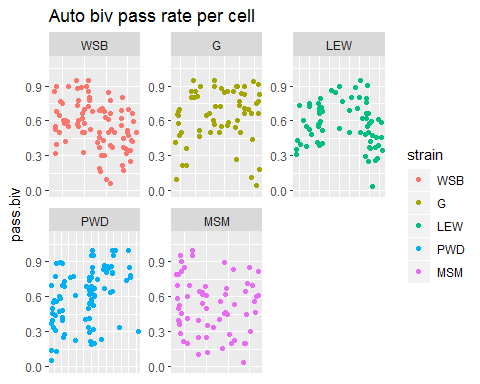
(should I measure the var of all traits in Q1)

1. rapid evolution of genome wide rec rates can fit into this -

* evolution of more COs requires more DSBs (evidence points to a earlier step which is likely the chromatin compaction (amount of axis))
* the strongest single bivalent predictor/correlate of rapid gwRR for males is interferece strength in 2COs

AND somewhat puzzeling result of (having stronger interference), meaning there is a greater proportion of a bivalent that seperates two foci (2 foci push to the ends of chrms (maybe decreased centromere supression))

# 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