MLH1 Results Outline

Table of Contents

[Main Figure 2](#_Toc33599013)

[Genome wide recombiantion rate estimates for both sexes 2](#_Toc33599014)

[Evolutionary Framework 3](#_Toc33599015)

[Within Mouse Variance in CO Count per Cell 4](#_Toc33599016)

[Evolution of Genome wide recombination rate associated with evolution of mean precursor number 5](#_Toc33599017)

[Evolution of genome wide recombination rate is reflected at the single chromosome level 8](#_Toc33599018)

[Single Bivalent Level Results 9](#_Toc33599019)

[Q1 Sex Differences in SC-AE Lengths 9](#_Toc33599020)

[Adjusting for XX 14](#_Toc33599021)

[short transition 15](#_Toc33599022)

[Q1. 1CO Position Sex Differences 15](#_Toc33599023)

[Telomere and centromere Distance 24](#_Toc33599024)

[Q1. Sex Differences in CO Interference (IFD) 31](#_Toc33599025)

[Mixed Model Tests, Fixed Effects 36](#_Toc33599026)

[transition Q1 to Q2 38](#_Toc33599027)

[Q2 SC Length 38](#_Toc33599028)

[brief transition, rec landscape 41](#_Toc33599029)

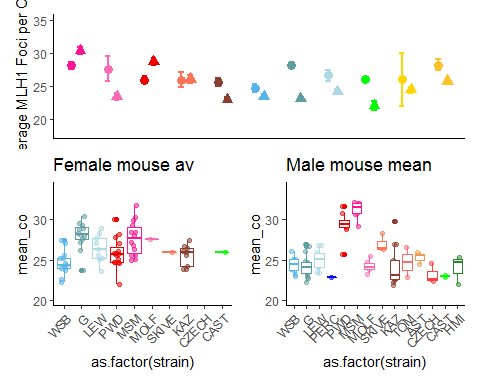
[Q2 Normalized Single CO positions 41](#_Toc33599030)

[Q2 Evolution of interference is associated with genome wide recombination rate evolution 43](#_Toc33599031)

[Note on Heterochiasmy Definition 46](#_Toc33599032)

[References 47](#_Toc33599033)

# 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) 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 are 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).

<add a transition sentence, about the expectations under different evolutionary scenarios>

# 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 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, ( 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,~~ *~~musculus~~~~PWD~~* ~~(glm; p=7.3710^{-10}),~~ *~~musculus~~~~SKIVE~~* ~~(glm; p=0.01) and~~ *~~molossinus~~~~MSM~~* ~~(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. (…for this reason we don’t examine this further)

## 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 linear models support the general qualitative 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, which

~~. 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 of MLH1 foci count per cell in males. While in females, (the significance of strain effects was not consistant between the full data set and higher quality dataset – LEW (glm; p = 0) was a significant effect in the full model and PWD (glm; p =0.04) in the higher quality dataset

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 (almost twice as much as males (Figure 1))~~**

<statement on biological significance – mech, what are the implications of different variance in the two groups of cells>

(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~~DSB number

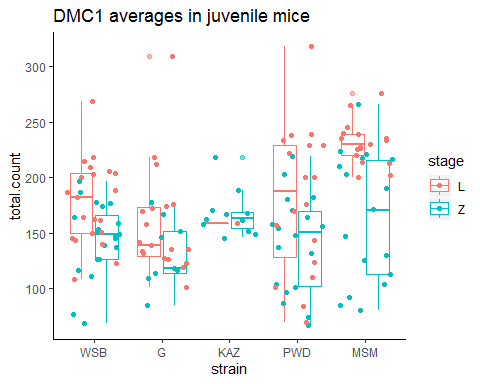


Table X. DMC1 foci counts per cell summary

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 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.

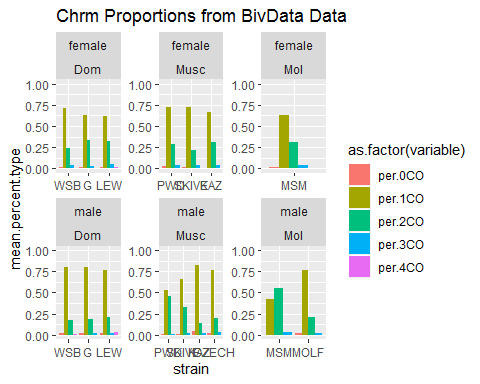
after DSB formation, DSBs are repaired as either NCO or CO (CO designation step) The vast majority of DSBs are repaired as NCOs, with a minority (in most species) of approximately 10% are repaired as crossovers.

-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 the proportion of 1CO:2CO distinguishes the high and low rec males (Figure X). This confirms the intuitive interpretation that high recombination strains are enriched for 2CO bivalents at the expense of 1CO bivalents. Across the female observations the propotions of chromsome classes are not notably different (t-test, pval?).

~~Most of the low rec strains~~ In the high rec strains the 2CO proportions are 0.33 ( *musculusSKIVE* ), 0.44 ( *musculusPWD* ), and 0.53 (*molossinusMSM* ), while in the low rec strains the 2CO proportions range from 0.14 (*musculusKAZ* ) to 0.22 ( *molossinusMOLF* ).

~~While the proportions are significantly different,~~ *musculusPWD* ~~PWD~~ male and *molossinusMSM* ~~MSM~~ male proportions are more similar to each other than strains from the same subspcies ( *musculusPWD* and *musculusKAZ* chi-square test; 1CO p =4.6210^{-26}, 2CO p=3.1510^{-33}) and *molossinusMSM* and *molossinusMOLF* chi-square test; 1CO p = 2.9510^{-12}, 2CO p=4.7210^{-13}).

~~These results confirm the intuitive conclusion that the rapid male specific evolution in gwRR proceeds though the chromosome level/though increasing the number of 2CO bivalents at the expense of 1CO bivalents.~~

~~<(Also comparing the low group together … Six strain have been designated low rec groups, WSB, LEW, G, KAZ, CZECH, and MOLF. They have significantly different proportions for 1CO and 2CO (chi-square; 1CO p = 0.05 and 2CO p= 0.01). This significant p value indicates that at least one of the six strains had significant differences).>~~

# Single Bivalent Level Results

~~– Confirmation that it’s the increase in 2CO bivalents – that comprises a higher cell wide number of COs (total number), Let’s look at the Rec landscape –~~

<Transition sent here> (sex differences –rec landscape, Sardell

In order to better deconstruct the genome wide recombiation rate, we decided to deconstruct cell wide average by examining features of the recombiantion landscape at the single bivalent level. recombination landscape refers to the i) total length of the SC-AE (potential area for COs), ii) the position of single crossovers along bivalents and iii) the spacing between multiple COs on the same bivalent.

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. (make sure these are the hand.foci bivalents).

The ranges of 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. Because there are hundreds of observations per category, we assume that each of the 19 or 20 chromosomes are equally represented in male and female single bivalent data datasets. We assume the bivalent datasets across all categories are equally eqivilent for representing general patterns.

<transition sent here>

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

# Q1 Sex Differences in SC-AE Lengths

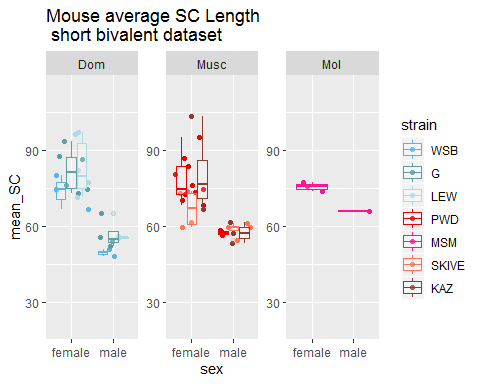
DO THE HIGH QUALITY dataset tests

Our data set provides an opportunity to test if the canonical pattern of females having higher gwRR and more SC area) have been uncoupled in the instance of rapid male specific evolution in the gwRR of *musculusPWD* and *molossinusMSM* A simple model might predict if SC area is the strongest predictor for gwRR / CO number, the high recombining males would have greater SC area proportional to the difference in CO number per cell.

To test this prediction – we use a combination of SC based metrics (and do our best to make comparable tests across male and female cells (oocytes and spermatocytes) / do our best to over come the confounding effects of the sex chromosomes. (in oocytes, the paired X chromosome is indistinguishable (effectively and autosome) while in the spermatocytes the XY SC has a distinct unpaired morphology.

In the first solution We use a reduced bivalent data set made up of the shortest bivalents from a single cell (see methods). This reduced set of short bivalents excludes the XX (3rd longest by Mb) in ooctytes – and (more similar set of chromosome identities compared across spermatocytes and oocytes) (ideally Chromosomes 15 to 19)

A total of 678 short bivalents were isolated from 140 oocytes and spermatocytes (add sex specific numbers) and mouse means were calculated (Figure X).



For all but one strain, *musculusSKIVE*, the mouse mean for the short bivalents are significantly longer (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. The lack of significance in the SKVIE strain –might be due to lower sampling of bivalents from this strain.

(the second dataset is opposite to summary of small single bivalents – whole genome for many cells) An additional measure of the SC-AE we tested is the total SC area per cell (the whole genome).**to get a comparison of the SC-AE area for the whole genome (not just a subset of chromosomes)**

We use a metric of SC area for whole cells, while this measure does include the confounding factors of the sex chromosomes (it’s good – in a sense it compare the same units, whole genomes.

(the XX in on average 10% of the total sc area per cell, the XY has much more variant morphology – (on average the skeletonize area of the XY is XX% of the total SC area per spermatocyte)

(additional lines of evidence which support – female SC lengths being longer:

(permutations): To test if the sex differences in SC area per cell exceed (the expected difference of an additional autosome (in females) – we ran permutations (for each strain) calculating the mean bivalent length (also write to calq mean total cell SC area) – replicating ‘in silico’ nuclei by sampling 19 female bivalents (instead of 20) and 20 male bivalents instead of (19).

the effects of an additional autosome (XX)

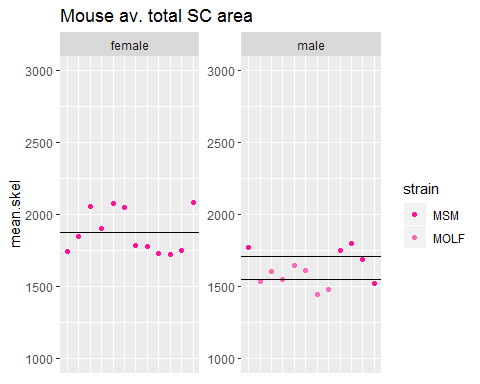
We confirm that all of the bivalents have longer

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. / cell based measures were left. Mouse means were calculated for their total SC-AE area and are displayed in figure X with the strain means shown as horizontal lines (below).

Insert p.values for the t.tests of **total SC**!!

Total, F:M (p value and amount of difference (for total)

Dom, Musc,



(The male mouse means for total SC in domesticus strains are significantly shorter compared to Musc (t.test; p= 1.5210^{-7}) and Molossinus (t.test; p = 3.8310^{-5}, indecating evolution (divergence) of SC length and (chromatin compaction) – we follow up with models (which focus on sex differences). )

The above t.tests were done to establish the sex differences for SC-AE (chromatin condensation), we now apply models (for understanding within an evolutionary framework) – we apply the same models to both mean mouse total sc and mouse mean short bivalent.

(for glms below, combine both the short biv and total SC per cell)

**M1. Mixed Model**

**M2.Linear Model**

**M3. Linear Model**

For the mixed model run on the total sc metric, there are more significant effects for strain and subsp for the mixed model of mouse.av total sc.

**Total SC and short biv glm results**

(total or short biv)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). Mixed models for the sex differences in mouse averages for **short bivalents** sex is the most significant factor for the mixed model of mean short bivalents (LTR; p =6.910^{-11}). The interaction (subspecies by sex) effect also significant (p = 0.13).

These effects were followed up the glms, again applied to both mouse means for short biv SC lengths and total sc area per cell.

The following effects had significant p values for the **total SC** from the above model

subspecies musc p = 2.1410^{-4}, subspecies mol p =6.2410^{-4}, sex male p =4.310^{-8}, strain lew p =5.910^{-6}, strain g p =1.7410^{-11},

(subsp\*sex) mol.male p =0.08, sex.g p =2.1110^{-4}, lew.male p =0.03

Fr the **short bivalent** sc the model results are listed below. From this (M2) model sex effect, male is most significant (glm; p =0).

while the skive strain effect (glm; p=0.03) and and SKIVE\*male (glm; p =NA) are slightly significant.

In the M3 model (strains and sex as fixed effects), the Sex effect (male) is the only significant p value from this model (glm; p= 0.93).

<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

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

< - Of all female single bivalents observations, 5% are XX (1 of 20). - The XX is large, likely within the top 25% longest bivalents of the cell (3rd largest by Mb). - The average % of XX for whole cell SC (sum(all bivalents)) can be calculated from the whole.cell data set. Lets guess 12% of a cell’s total SC area is XX. - rate of bivalent segmentation /\* rate of XX, 5% /\* mean SC length for 3rd longest bivalent / total SC area (by bivalent) = proportion of SC area due to XX>

# 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)

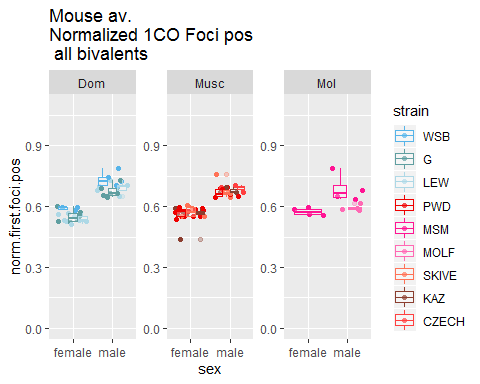
# Q1. 1CO Position Sex Differences

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

We anchor the position measures at the centromere (and normalize by deviding the position by the length of the chromosome, so the value range from 0 to 1, with smaller values representing positions closer to the centromere end of the bivalent, and larger values meaning foci positions closer to the opposite end (refered here as terminal or telomeric position).

We plot the mouse averages across all of the single bivalent measures, with the strain averages shown as boxplots…

in order to get a better picture of the sex differences in this foci-position metric



In the domesticus strains the mouse averages for normalized single foci positions were significantly sexually dimorphic (fitting the predictions/pattern of the ‘typical recombination landscape’ (Sardell)). The single CO positions are medially placed in females and terminall placed in males.

The pavlues 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 familiar models

**M1. Mixed 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).

**M2. Linear model**

**M3. Linear model**

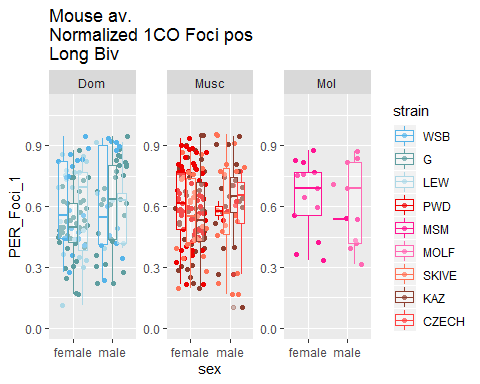
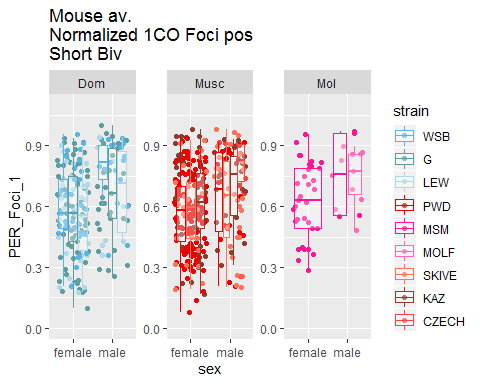
The male effects were the most significant (glm M2; p =1.3310^{-7}, glm M3; p = 1.3310^{-7}). While the musc subspecies, LEW and KAZ strains had slightly signifcant effects (glm M2; Musc subsp p = 0.07, LEW strain p = 0.03) and (glm M3; LEW strain p = 0.03, KAZ strain p = 0.07).

(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).

Chromosome size effects are a confounding factor for CO position, with the expectation that shorter chromosomes will have a more uniform landscape compared to larger/longer chromosomes (cite). To account for this we again use the reduced bivalent sets (long.biv and short.biv)

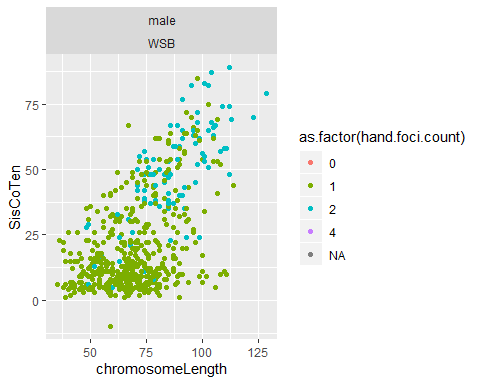
## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).

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



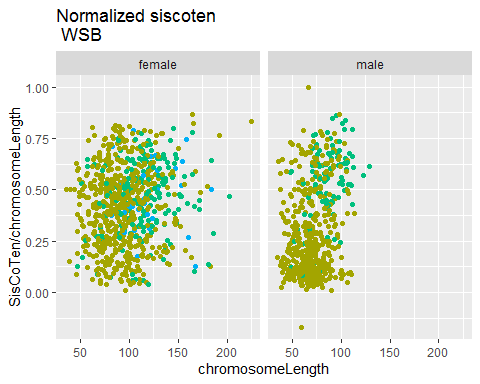
These box plot show that females have a much more medial position of single foci bivalents, (much closer to 50% compared to males). They also show that Musc males’ Foci1 position is slightly more central / medial compared to the same type of positions in the Dom male strains. MOLF males have much more medial positions than other strains.

## Warning: Removed 46 rows containing missing values (geom\_point).

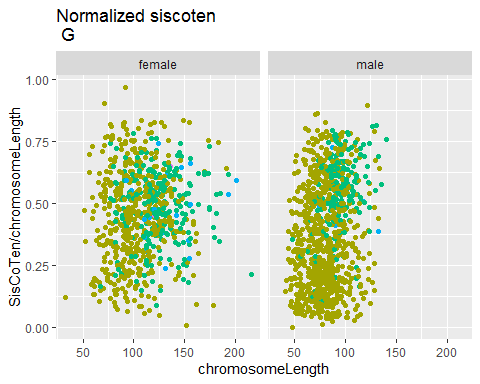


Males have much clearer separation of siscoten across chrm classes. This is emphasized when SC length is also plotted. It seems like musc males have higher amounts of this metric compared to Dom males. To formally test the differences in sis-co-ten I plan to write a sub sampling / permutation loop to compare the mean(sis.co.ten) of the same numbers of bivalents of the same class. BUT females have a greater range – so maybe it’s just a scale issue.

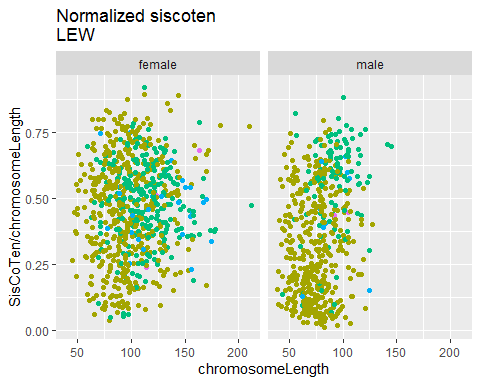
## Warning: Removed 24 rows containing missing values (geom\_point).



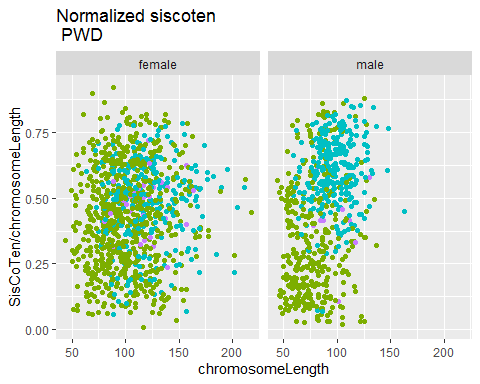
## Warning: Removed 22 rows containing missing values (geom\_point).



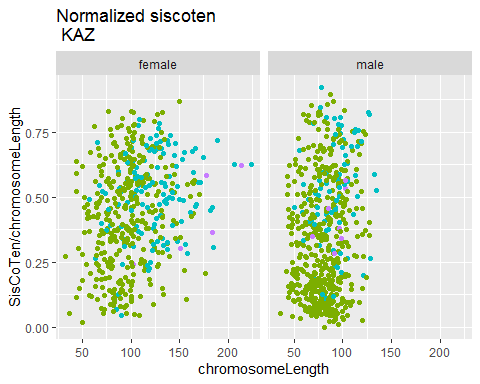
## Warning: Removed 12 rows containing missing values (geom\_point).



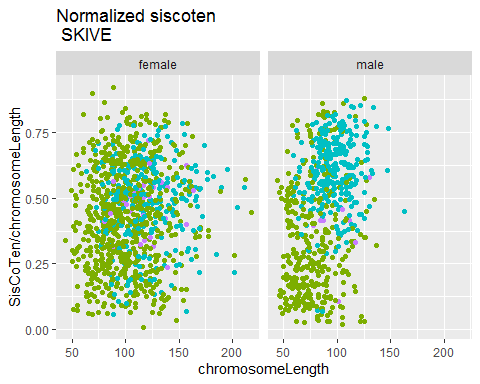
## Warning: Removed 16 rows containing missing values (geom\_point).



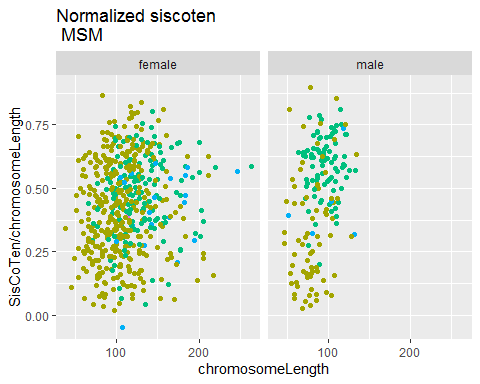
## Warning: Removed 23 rows containing missing values (geom\_point).



## Warning: Removed 16 rows containing missing values (geom\_point).



## Warning: Removed 8 rows containing missing values (geom\_point).



I think the the normalized sis.co.ten plots also show that the there is more clustering of the sis.co.ten for the males.

##   
## Call:  
## glm(formula = Rec.group ~ mean.siscoten, family = binomial(link = "logit"),   
## data = Male.poly.Mouse.Table\_BivData\_4MM[(Male.poly.Mouse.Table\_BivData\_4MM$subsp ==   
## "Musc"), ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5257 -0.0661 0.0042 0.0773 1.4045   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -47.214 30.904 -1.53 0.13  
## mean.siscoten 1.452 0.948 1.53 0.13  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 22.9145 on 17 degrees of freedom  
## Residual deviance: 5.2123 on 16 degrees of freedom  
## AIC: 9.212  
##   
## Number of Fisher Scoring iterations: 9

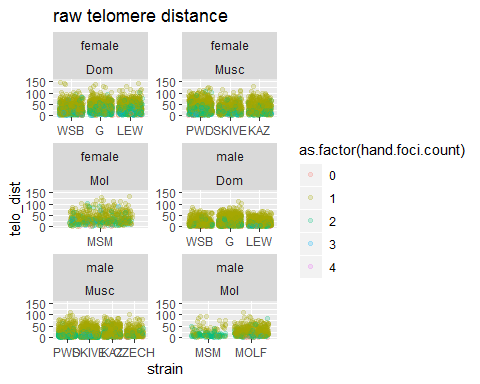
##   
## Call:  
## glm(formula = Rec.group ~ SisCoTen, family = binomial(link = "logit"),   
## data = Curated\_BivData[(Curated\_BivData$subsp == "Musc") &   
## (Curated\_BivData$sex == "male"), ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.912 -1.276 0.809 0.994 1.167   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.02398 0.07745 0.31 0.76   
## SisCoTen 0.01566 0.00195 8.01 1.1e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2977.8 on 2268 degrees of freedom  
## Residual deviance: 2910.1 on 2267 degrees of freedom  
## (38 observations deleted due to missingness)  
## AIC: 2914  
##   
## Number of Fisher Scoring iterations: 4

All the sis.co.ten tests are highly significant. Maybe I should consider running a normalized sis.co.ten? I think nrm\_siscoten would still reflect the differing cohesion structure/outcome.

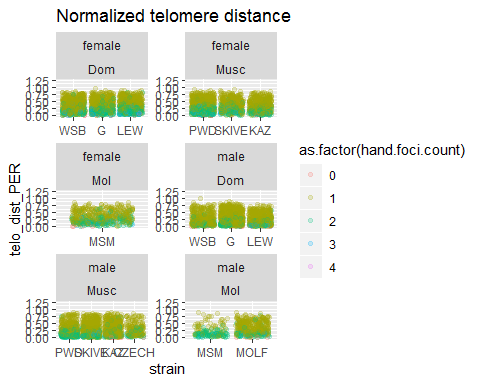
### Telomere and centromere Distance

My metric for telomere and centromere distance measure the distance of the nearest foci to the ends of the bivalent (SC). In the plots below each point is a single bivalent. I choose not to use the mark for centromere because it seems noisy and inconsistent…

## Warning: Removed 82 rows containing missing values (geom\_point).



## Warning: Removed 85 rows containing missing values (geom\_point).



Males on average have much lower raw telomere distance (reflects the telomere bias) compared to females. In Males, 2CO bivalents have very low telomere distances, while the 1CO bivalents have a greater range. In females the ranges of telomere distances have much more overlap.

## Single term deletions  
##   
## Model:  
## mean.telo.dist ~ subsp \* sex + (1 | strain)  
## Df AIC LRT Pr(Chi)  
## <none> 540   
## subsp:sex 2 537 1.06 0.59

## Estimate Std. Error t value  
## (Intercept) 34.2 2.1 16.04  
## subspMusc 1.2 3.0 0.42  
## subspMol 5.4 4.2 1.30  
## sexmale -14.5 1.6 -8.86  
## subspMusc:sexmale -2.2 2.4 -0.93  
## subspMol:sexmale -3.2 3.7 -0.88

## Single term deletions  
##   
## Model:  
## mean.telo.dist ~ subsp \* sex + (1 | strain)  
## Df AIC LRT Pr(Chi)  
## <none> 540   
## subsp:sex 2 537 1.06 0.59

##   
## simulated finite sample distribution of RLRT.  
##   
## (p-value based on 10000 simulated values)  
##   
## data:   
## RLRT = 7, p-value = 0.003

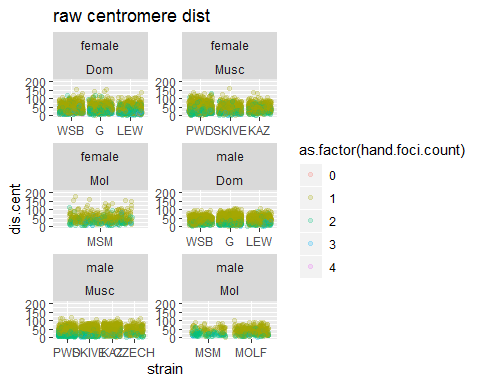
Mixed model result summary:

##   
## Call:  
## glm(formula = Rec.group ~ mean.telo.dist, family = binomial(link = "logit"),   
## data = Mouse.Table\_BivData\_4MM[(Mouse.Table\_BivData\_4MM$subsp ==   
## "Musc") & (Mouse.Table\_BivData\_4MM$sex == "male"), ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.815 0.435 0.592 0.747 0.860   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 4.73 6.09 0.78 0.44  
## mean.telo.dist -0.18 0.32 -0.56 0.57  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 15.012 on 14 degrees of freedom  
## Residual deviance: 14.650 on 13 degrees of freedom  
## AIC: 18.65  
##   
## Number of Fisher Scoring iterations: 4

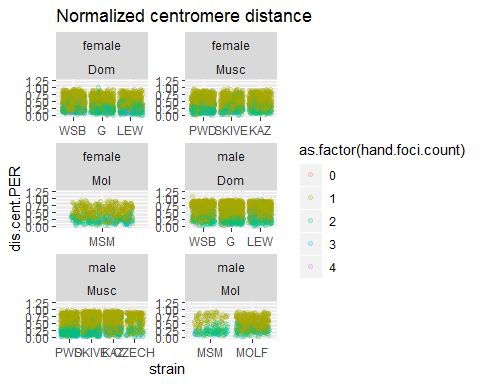
##   
## Call:  
## glm(formula = Rec.group ~ telo\_dist, family = binomial(link = "logit"),   
## data = Curated\_BivData[(Curated\_BivData$subsp == "Musc") &   
## (Curated\_BivData$sex == "male"), ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.490 -1.388 0.916 0.952 1.274   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.70982 0.06653 10.67 <2e-16 \*\*\*  
## telo\_dist -0.00850 0.00235 -3.61 3e-04 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3038.1 on 2303 degrees of freedom  
## Residual deviance: 3025.1 on 2302 degrees of freedom  
## (3 observations deleted due to missingness)  
## AIC: 3029  
##   
## Number of Fisher Scoring iterations: 4

##   
## Call:  
## glm(formula = Rec.group ~ telo\_dist\_PER, family = binomial(link = "logit"),   
## data = Curated\_BivData[(Curated\_BivData$subsp == "Musc") &   
## (Curated\_BivData$sex == "male"), ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.488 -1.382 0.916 0.953 1.128   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.7057 0.0692 10.2 < 2e-16 \*\*\*  
## telo\_dist\_PER -0.6689 0.2027 -3.3 0.00097 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3038.1 on 2303 degrees of freedom  
## Residual deviance: 3027.2 on 2302 degrees of freedom  
## (3 observations deleted due to missingness)  
## AIC: 3031  
##   
## Number of Fisher Scoring iterations: 4

## Warning: Removed 126 rows containing missing values (geom\_point).



## Warning: Removed 127 rows containing missing values (geom\_point).



## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

##   
## Call:  
## glm(formula = Rec.group ~ mean.cent.dist, family = binomial(link = "logit"),   
## data = Mouse.Table\_BivData\_4MM[(Mouse.Table\_BivData\_4MM$subsp ==   
## "Musc") & (Mouse.Table\_BivData\_4MM$sex == "male"), ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -6.31e-05 2.00e-08 2.00e-08 2.00e-08 6.41e-05   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 3373.0 1607241.5 0 1  
## mean.cent.dist -79.9 38084.8 0 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1.5012e+01 on 14 degrees of freedom  
## Residual deviance: 8.0864e-09 on 13 degrees of freedom  
## AIC: 4  
##   
## Number of Fisher Scoring iterations: 25

##   
## Call:  
## glm(formula = Rec.group ~ dis.cent, family = binomial(link = "logit"),   
## data = Curated\_BivData[(Curated\_BivData$subsp == "Musc") &   
## (Curated\_BivData$sex == "male"), ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.790 -1.280 0.793 0.950 1.606   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.30679 0.09495 13.76 <2e-16 \*\*\*  
## dis.cent -0.01884 0.00206 -9.15 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2980.5 on 2271 degrees of freedom  
## Residual deviance: 2894.0 on 2270 degrees of freedom  
## (35 observations deleted due to missingness)  
## AIC: 2898  
##   
## Number of Fisher Scoring iterations: 4

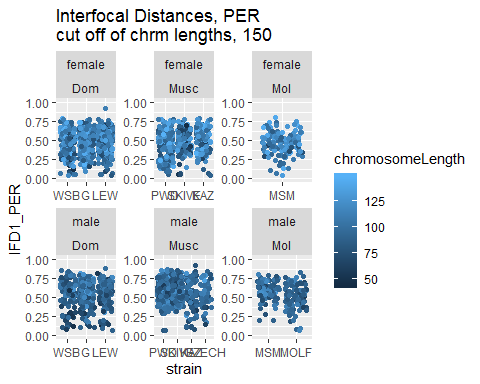
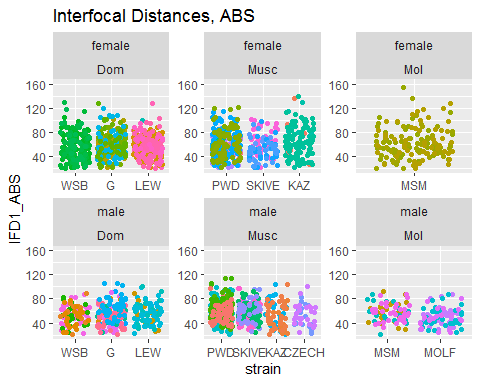
##   
## Call:  
## glm(formula = Rec.group ~ dis.cent.PER, family = binomial(link = "logit"),   
## data = Curated\_BivData[(Curated\_BivData$subsp == "Musc") &   
## (Curated\_BivData$sex == "male"), ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.790 -1.259 0.783 0.973 1.233   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.3060 0.0988 13.21 <2e-16 \*\*\*  
## dis.cent.PER -1.4865 0.1705 -8.72 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2980.5 on 2271 degrees of freedom  
## Residual deviance: 2902.1 on 2270 degrees of freedom  
## (35 observations deleted due to missingness)  
## AIC: 2906  
##   
## Number of Fisher Scoring iterations: 4

The normalized centromere plots show that in Musc males, on 2CO bivalents the 1st CO is closer to the centromere end than in Dom males. Females have more overlap in the distributions of centromere distances across chromosome class compared to males.–>

# Q1. Sex Differences in CO Interference (IFD)

In order to test if there are sex differences between crossover interference, a major determinant of the positioning of crossovers along chromosomes, we examined … ran models with the interfocal distance (IFD) of two foci on the same bivalent.

-We focus on observations from two-crossover bivalents for more comparable observations. -comparisons of all IFDs of mulit crossover bivalents is also included Interference is a major determinant of the positioning of chromosomes



The raw IFD measures – are usuful for thinking of crossover interference as a mechnical - force mediated through the SC – In some models this metric should be thought of as the best metric for charaecterizing crossover interference – bc it reflect the physcial –

The above plots display the raw and normalized area which seperate two foci on the same chromosome. In the normalized plot, 9 observations fall above the length threshold of 150. The female measures fall more uniformly across the range of (0 to 1).

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).

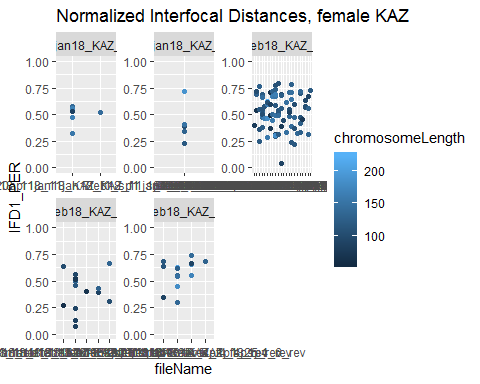
t.tests are performed on the normalized metrics for establish if there are sex differences in the IFD meterics (raw and normalized).

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).

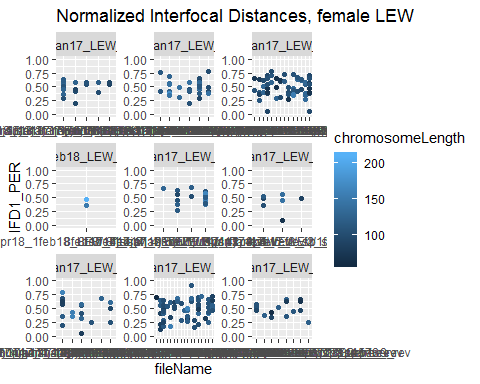
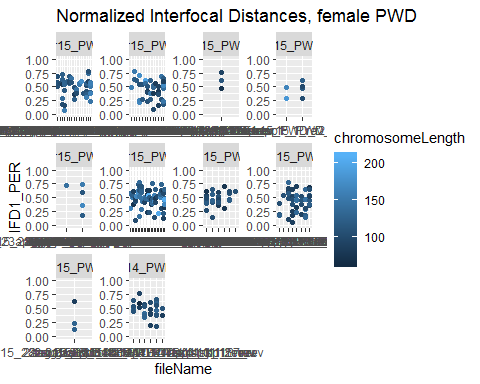
The normalized IFD measures are highly significant (t.test; p = 1.4910^{-12}) between sexes, indicating there a

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

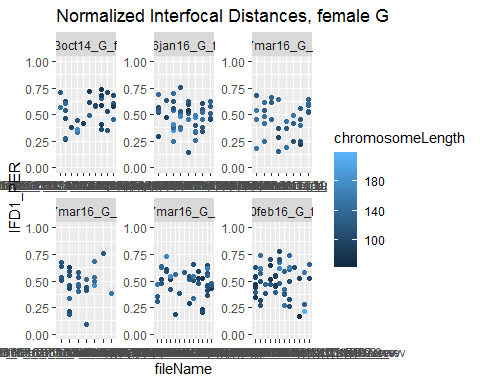
This indicates a general pattern of two foci are seperated by more area (hance have stronger interference) in males, this pattern is also for genetic maps.



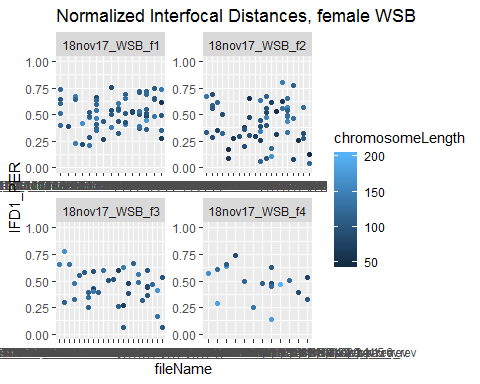
## Warning: Removed 1 rows containing missing values (geom\_point).



## 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)

– run these comparisons to see if the small NRM IFD are due to shorter chromosomes, which are predicted to have weaker interference.

Looked at the KAZ pattern

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

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

**M1. Mixed 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

The table above should display the slightly unusual pattern, where the coefficients for the significant sex fixed effect are positive and negative in the raw and normalized values respectively. That is for the raw IFD values, females are significantly longer but the normalized IFD values, males are significantly longer.

I tested 2 versions of the mixed model for this flavor of trait, raw IFD and normalized IFD measure. The tables below are from anova( for the lmer model ). Random effect of strain is not significant for ABS IFD, and only slightly significant for the IFD.PER

**M2. Linear Model**

**M3. Linear model**

## 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.

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 raw measures, in M2, I think the 2 SKIVE effects mean that the female raw IFD is shorter than the male IFD.raw. The other effects are for PWD and SKIVE (larger raw IFD from intercept.) I think the MSM and MOLF interaction effects were too far down the list to sop up any variance. For the M3, only SKIVE\*male effect is close to 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.

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.

Dive deeper into the sex specific pattern for each strain. Below are code chunks which show the unusual sex specific results for IFD measures. The general pattern is that, female raw IFD > male IFD and female PER IFD < male PER IFD. The scatter plots show that female raw measures are longer than male and for the PER values, the female mean is brought down by an enrichment of short IFDs.

For some strains, PWD, MSM and SKIVE there’s a 30% threshold in the male PER IFD distributions. (What does that mean?). How do I test / quantify this pattern? Cluster metric?

Above is a table of the proportion of 2Co bivalents which have a norm IFD below 30%, For all strains but KAZ, the females have a greater proportion of these shorted IFD values.

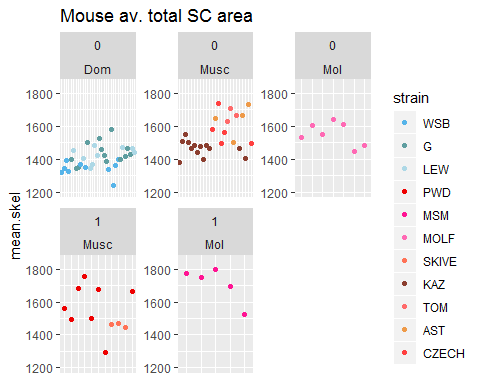
# transition Q1 to Q2

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

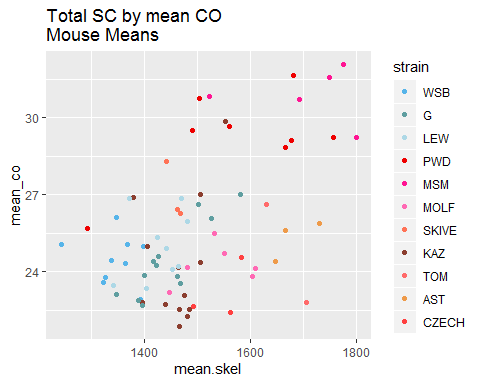
# Q2 SC Length

<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).**

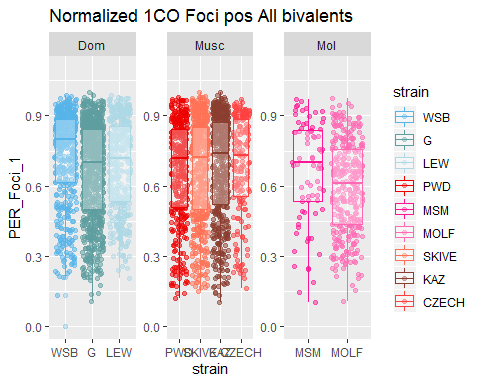
**Summary Points**

* 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 Normalized Single CO positions



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).

in addition models which use mouse average of foci position as a dependat variable, we decided to run models for the single bivalent level, and which include chromosome length as a fixed effect.

in the above chunk I tried making another glm’s which uses the single bivalent measures and include chromosome size for the switching the model to include interaction (instead of additive), all of the pvalues get lower. .. the chromosome length has the most significant effect, LEW by chrm effect are significant

In the reduced model, chrm length is the most significant, strain SKIVE, MOLF, G and LEW are significant strain effects

**For the M2 (subsp)** There are differences for the dom strains: G strain(p = 0.01).

LEW (slight) (p =0.08).

But big effects from Mol (p = 8.1410^{-7}).

and MSM strain (p = 3.9510^{-4})

**For M3 (all strains)**

strain G (p = 0.01).

strain Lew (p = 0.08).

strain PWD ( p = 0.01).

slight strain MSM (p = 0.07).

very sig strain MOLF (p = 8.1410^{-7}).

strain SKIVE (p = 0.02).

strain KAZ (p = 0.04).

from the main figure above, you can see that G and MOLF 1CO foci positions are the most different

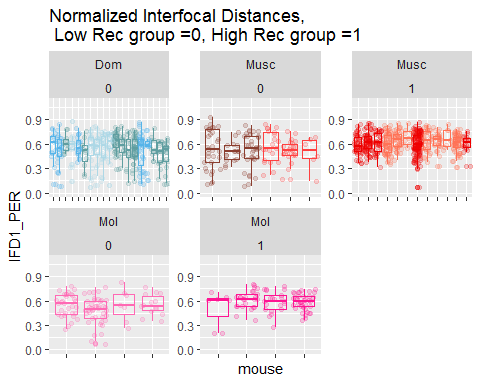
For MOLF 326 1CO bivalents, 79, 126, 62, 59 1CO observations across the 4 MOLF mice

there are only 77 MSM 1COs across 4 mice in MSM (MSM is less central than MOLF) (motivation for MOLF female observations and more male MSM biv measures)

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

There are significant difference across the 1CO positions, yet they don’t match fit any of the predictions. there are also complications due to the chromsome class proportions that

# 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.**

The plots above show the mean SC lengths and 2SE error bars for single bivalents which have been given within cell rank.

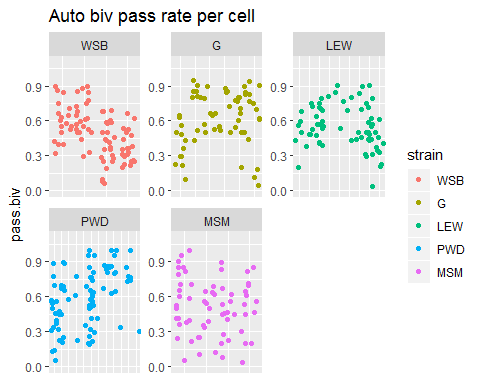
The first plot showing the mean SC lengths by the rank (most all of these cells have 3, MSM has 5 cells (observations)).

The purpose of these plots is to display the variance of single bivalents when they are assigned a within cell rank. For the longest bivalents, XX is predicted to be the 3rd longest (according to physical length Mb).

(use the value for the 3rd bivalent to adjust the single bivalent traits for XX – then compare to males values – or re-run in the MM).

The other figure shows of each single bivalent contributes to the total SC area. Each column is a cell and each color is the percent of total SC area for the longest 5 bivalents in that cell. on average, each of the top longest bivalents make up ~10% of the cell’s total SC area. So for cells all 20 bivalents, of it’s total SC area, 5-7% is due to a XX,

* Is the difference between cell averages for males and females less that 10%?
* also interesting, the pwd and msm don’t have longer SC, compared to other strains.



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.

# Note on Heterochiasmy Definition

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