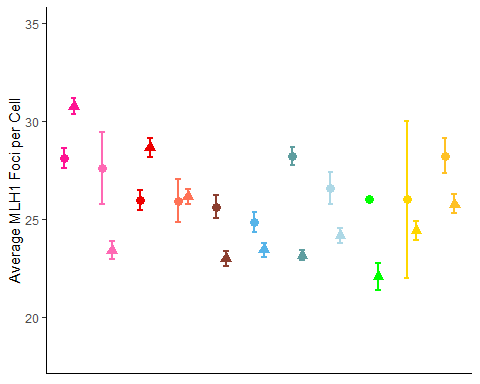
MLH1 Results Outline

13 February 2020

Table of Contents

# Main Figure

## Warning: NAs introduced by coercion



This is a draft of the main figure. Outgroups are not included in the analysis below.

# Genome wide recombiantion rate estimates

Table 1. MLH1 counts per cell summary statistics by strain

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| subsp | strain | sex | Nmice | Ncells | mean\_co | cV | var | sd | se |
| Dom | WSB | female | 14 | 201 | 25 | 14 | 13.0 | 3.6 | 0.25 |
| Dom | WSB | male | 12 | 235 | 23 | 11 | 7.2 | 2.7 | 0.18 |
| Dom | G | female | 12 | 318 | 28 | 15 | 17.5 | 4.2 | 0.24 |
| Dom | G | male | 18 | 355 | 23 | 11 | 6.9 | 2.6 | 0.14 |
| Dom | LEW | female | 9 | 147 | 27 | 18 | 23.3 | 4.8 | 0.40 |
| Dom | LEW | male | 10 | 253 | 24 | 13 | 9.6 | 3.1 | 0.20 |

In order to estimate the genome wide recombination rates across our panel of 17 strains, MLH1 foci were quantified from a total of 2359 spermatocytes and 1525.

The strain means for females were pretty similar. For female Domesticus strains the means are 24.84, 28.21, 26.59 for *M. m. domesticusWSB*, *M. m. domesticusG*, and *M. m. domesticus$LEW* respectively. In the three Musculus strains the means are 25.98, 25.94, 25.62 for *M. m. musculus^ PWD}^*, *M. m. musculus^ SKIVE^*, and *M. m. musculusKAZ* respectively. In the molossinus strains the means are 28.12, 27.62 for *M. m. molossinus^ MSM^* and *M. m. molossinusMOLF* respectively.

While in two of the subspecies, the means for males show a greater range across means, notable in the mmmolossinus strains 30.77, 23.42 for *M. m. molossinusMSM* and *M. m. molossinusMOLF* respectively. In the musculus the strain means are 28.67, 26.16, 22.99, 23.7, 24.41, 22.31 for *M. m. musculusPWD*, *M. m. musculusSKIVE*, *M. m. musculusKAZ*, *M. m. musculusTOM*, *M. m. musculusAST*, and *M. m. musculusCZECH* respectively. In the male Dom strains the means for MLH1 count per cell are 23.43, 23.16, 24.16, 21.81 for *M. m. domesticusWSB*, *M. m. domesticusG*, *M. m. domesticusLEW* and *M. m. domesticusPERC* respectively. These

The ranges of mean MLH1 counts per cell are

# Analysis for Evolutionary Patterns

In order to test the effects of subspecies, sex and strain on the mean and variance of MLH1 counts per cell, we fit a data set of 137 mouse averages for MLH1 foci per cell. We constructed the model to set subspecies, sex and their interaction as fixed effects, which strain was coded as a random effect. The predicted coeffecients and pvalues are reporter below in table X.

## Estimate Std. Error t value  
## (Intercept) 26.35 1.04 25.37  
## subspMol -0.52 1.72 -0.30  
## subspMusc -0.72 1.50 -0.48  
## sexmale -1.63 0.48 -3.39  
## subspMol:sexmale 3.27 1.04 3.14  
## subspMusc:sexmale 2.91 0.74 3.91

## refitting model(s) with ML (instead of REML)  
## refitting model(s) with ML (instead of REML)

Because all of the mixed model effects had significant effects, (p= 0, 610^{-4}, and 210^{-4} for the sex, subspecies and interaction fixed effects and 0 for the random strain effect), we followed up with two general linear models to investigate the specific strain effects.

and

In the first glm, the fixed subspecies effects were not significant, so we only report for values for the second model which is limited to sex and strain fixed effects.

## Estimate Pr(>|t|)  
## (Intercept) 24.71 0.00000  
## sexmale -0.16 0.82201  
## strainG 3.30 0.00000  
## strainLEW 1.70 0.01941  
## strainPWD 1.16 0.06536  
## strainMSM 2.98 0.00001  
## strainMOLF 2.91 0.09633  
## strainSKIVE 1.23 0.48097  
## strainKAZ 0.86 0.23441  
## sexmale:strainG -3.17 0.00165  
## sexmale:strainLEW -1.19 0.27809  
## sexmale:strainPWD 3.59 0.00054  
## sexmale:strainMSM 3.72 0.00196  
## sexmale:strainMOLF -3.22 0.09905  
## sexmale:strainSKIVE 1.45 0.46010  
## sexmale:strainKAZ -0.81 0.43176

In both models strain by sex interaction effects were significant for (p = 5.3510^{-4} and 0 for the second mixed model) for *M. m. musculusPWD* and *M. m. molossinusMSM* respectively. These results confirm the qualitative observation that within these two strains there has been sex-specific in the average MLH1 count per cell.

Additionally, *M. m. domesticusG* had significant strain and sex by strain effects (p =1.9510^{-6} and 0 for strain and strain by sex effect). These results confirm the qualtative pattern of the *M. m. domesticusG* strain having the largest degree of sexual dimorphism (heterochiasmy) in the dataset.

The second model were strain was (nested and random), the fixed effects, sex and interaction of (subsp and sex), are much more significant. The coefficients indicate, males in general have 1 less in average, and the musc and molf subsp have ~3 more on average.

# Within Mouse Variance in CO Count per Cell

In order to more fully discribe the distributions we examined the variance in addition to the mean applying the same mixed models and linear models, with *within mouse variance* of MLH1 counts per cell as the dependant variable.

## sub.sp sex.results interaction.pval rand.pval  
## 1 0.87 1.6e-10 0.76 0.043

## sub.sp sex.results interaction.pval rand.pval  
## 1 0.86 5e-12 0.88 0.046

## sub.sp sex.results interaction.pval rand.pval  
## 1 0.021 2.1e-11 0.03 1

## sub.sp sex.results interaction.pval rand.pval  
## 1 0.073 1.3e-12 0.36 0.39

These are the p values for the effects of the mixed models of the foloowing mixed models: VAR.FULL, CV.FULL, VAR.Q12 and cv.Q12

We followed up the mixed models with glms. The coefficients are listed below for the following glms: VAR.FULL, CV.FULL, VAR.Q12 and cv.Q12

## Estimate Pr(>|t|)  
## (Intercept) 11.62 0.00000  
## sexmale -4.65 0.04418  
## strainG 2.70 0.21567  
## strainLEW 9.15 0.00017  
## strainPWD 0.93 0.64987  
## strainMSM 3.03 0.14847  
## strainMOLF 6.33 0.26968  
## strainSKIVE -1.81 0.75143  
## strainKAZ 3.20 0.17753  
## sexmale:strainG -3.37 0.30124  
## sexmale:strainLEW -8.02 0.02761  
## sexmale:strainPWD 0.23 0.94562  
## sexmale:strainMSM -0.79 0.83783  
## sexmale:strainMOLF -7.32 0.25341  
## sexmale:strainSKIVE 1.75 0.78734  
## sexmale:strainKAZ -4.14 0.22143

## Estimate Pr(>|t|)  
## (Intercept) 13.128 0.0000  
## sexmale -2.605 0.0224  
## strainG -0.017 0.9875  
## strainLEW 3.405 0.0041  
## strainPWD -0.069 0.9454  
## strainMSM 0.524 0.6111  
## strainMOLF 2.210 0.4341  
## strainSKIVE -1.058 0.7080  
## strainKAZ 1.638 0.1614  
## sexmale:strainG -0.656 0.6824  
## sexmale:strainLEW -2.923 0.1022  
## sexmale:strainPWD -1.010 0.5391  
## sexmale:strainMSM -1.421 0.4583  
## sexmale:strainMOLF -2.728 0.3878  
## sexmale:strainSKIVE 0.164 0.9590  
## sexmale:strainKAZ -2.514 0.1331

## Estimate Pr(>|t|)  
## (Intercept) 19.0 0.00000  
## sexmale -13.4 0.00023  
## strainG -3.8 0.25744  
## strainLEW 4.6 0.20970  
## strainPWD -8.7 0.01069  
## strainMSM -3.6 0.27589  
## strainMOLF -0.8 0.84163  
## strainSKIVE -1.1 0.87359  
## strainKAZ -5.6 0.10575  
## sexmale:strainG 5.1 0.25579  
## sexmale:strainLEW -4.3 0.39673  
## sexmale:strainPWD 10.2 0.03484  
## sexmale:strainMSM 2.9 0.57392  
## sexmale:strainSKIVE 1.4 0.86102  
## sexmale:strainKAZ 7.1 0.12924

## Estimate Pr(>|t|)  
## (Intercept) 15.916 0.0e+00  
## sexmale -7.059 3.2e-05  
## strainG -2.584 9.2e-02  
## strainLEW 0.630 7.1e-01  
## strainPWD -4.365 5.7e-03  
## strainMSM -2.646 8.0e-02  
## strainMOLF -0.721 6.9e-01  
## strainSKIVE -0.198 9.5e-01  
## strainKAZ -2.101 1.9e-01  
## sexmale:strainG 4.026 5.3e-02  
## sexmale:strainLEW -0.213 9.3e-01  
## sexmale:strainPWD 4.150 6.1e-02  
## sexmale:strainMSM 0.412 8.6e-01  
## sexmale:strainSKIVE -0.028 9.9e-01  
## sexmale:strainKAZ 2.528 2.4e-01

Across all four data sets, the sex effect has the most significant p value. The highest p value is for the sex effect for VAR and Cv in the full models. This might be due to extra noise from technical sources.

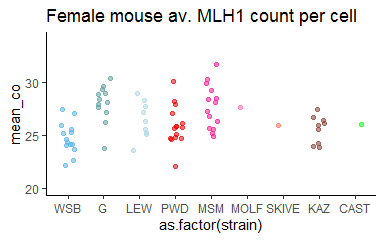
For both the mixed model and linear model, sex was the most signifcant effect (p= 1.6210^{-10} p= 2.2810^{-4} (for the high quality dataset) ) respectively.

The measures for variance are more supciptibal to thechnical error effects (noise). For this reasons, we analysze the models by restricting the data to a subset cells which were scored as highest quality. Sex was also the most significant effect (p)

s

## Female Specific Analysis

In order to more finely discirbe the sex specific patterns, we foucsed on data sets including single sexes.



## strain Nmice Ncells mean\_co mean.mouse\_cV mean.mouse\_var mean.mouse\_sd  
## 1 WSB 14 201 25 13 11.6 3.3  
## 2 G 12 318 28 13 14.3 3.7  
## 3 LEW 9 147 26 17 20.8 4.4  
## 4 PWD 15 222 26 13 12.5 3.4  
## 5 MSM 14 300 28 14 14.6 3.8  
## 6 MOLF 1 21 28 15 17.9 4.2  
## 7 SKIVE 1 32 26 12 9.8 3.1  
## 8 KAZ 9 184 26 15 14.8 3.8  
## 9 CAST 1 1 26 NaN NaN NaN  
## mean.mouse\_se  
## 1 1.03  
## 2 0.77  
## 3 1.17  
## 4 0.94  
## 5 0.85  
## 6 0.92  
## 7 0.55  
## 8 0.84  
## 9 NaN

For some strains, we were only able to quantified or include one mouse and the Cast observation comes from 1 1 cell. To analyze the female specific dataset,

In the dom subspecies the ranges of mouse means are 22.17 to 27.43 in *M. m. domesticusG*

23.53 to 28.9 in *M. m. domesticus$LEW*

In the musc subspecies the ranges of mouse means are 22 to 30.09 in *M. m. musculusKAZ*

The range is 24.9 to 31.73 in *M. m. molossinus$MSM*

Two glm models were run to test the effects of strain and subspecies on mean MLH1 coount per cell per mouse. The coefficients for the following model are below.

## Estimate Pr(>|t|)  
## (Intercept) 24.71 0.00000  
## subspCast 1.29 0.48177  
## subspMusc 0.86 0.25905  
## subspMol 2.91 0.11511  
## strainG 3.30 0.00001  
## strainLEW 1.70 0.02729  
## strainPWD 0.30 0.68472  
## strainMSM 0.07 0.96952  
## strainSKIVE 0.37 0.84220

## Estimate Pr(>|t|)  
## (Intercept) 24.71 0.00000  
## strainG 3.30 0.00001  
## strainLEW 1.70 0.02729  
## strainPWD 1.16 0.08104  
## strainMSM 2.98 0.00003  
## strainMOLF 2.91 0.11511  
## strainSKIVE 1.23 0.50303  
## strainKAZ 0.86 0.25905  
## strainCAST 1.29 0.48177

For mean CO count, *M. m. domesticusG* strain effects in both models (p = ) *M. m. musculusMSM* was a significant strain effect on female mean MLH1 counts per cell for only the second model (p= ).

Above are the coefficents for the two glm’s of the female specific mouse MLH1 averages, which include subspecies and strain as fixed effects. G has the most consistant significant strain effects in both models. MSM has a pretty low p value in the second model, LEW has a slightly significant pvalue.

*M. m. domesticusG* is 1.1 higher the the other means. *M. m. domesticuLEW* is 1.04 higher and *M. m. musculusMSM* is 1.09. These three will be designated as ‘moderate high rec’ strains for later sex specific analysis.

Within mouse variance was assess to follow up on the results from the full dataset.

## Estimate Pr(>|t|)  
## (Intercept) 11.62 1.6e-08  
## strainG 2.70 3.1e-01  
## strainLEW 9.15 2.3e-03  
## strainPWD 0.93 7.1e-01  
## strainMSM 3.03 2.4e-01  
## strainMOLF 6.33 3.7e-01  
## strainSKIVE -1.81 8.0e-01  
## strainKAZ 3.20 2.7e-01

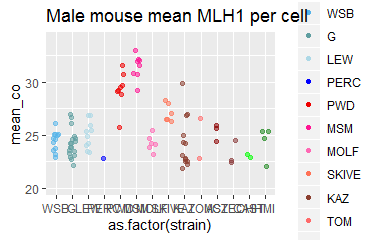
## Estimate Pr(>|t|)  
## (Intercept) 19.0 2.7e-07  
## subsp3 -5.6 1.8e-01  
## subspMol -3.6 3.7e-01  
## strainG -3.8 3.5e-01  
## strainLEW 4.6 3.0e-01  
## strainPWD -3.1 3.8e-01  
## strainSKIVE 4.5 5.9e-01

## Estimate Pr(>|t|)  
## (Intercept) 11.62 1.6e-08  
## strainG 2.70 3.1e-01  
## strainLEW 9.15 2.3e-03  
## strainPWD 0.93 7.1e-01  
## strainMSM 3.03 2.4e-01  
## strainMOLF 6.33 3.7e-01  
## strainSKIVE -1.81 8.0e-01  
## strainKAZ 3.20 2.7e-01

The variance across mouse means for each strain is also variable across strains. With *M. m. domesticus$LEW* having the most significant strain effect in the second model (p= ).

## Male Specific Analysis

The male specific patterns were also investigated.



## strain Nmice Ncells mean\_co mean.mouse\_cV mean.mouse\_var mean.mouse\_sd  
## 1 WSB 12 235 24 10.9 7.5 2.7  
## 2 G 18 355 24 9.8 6.1 2.4  
## 3 LEW 10 253 25 11.2 8.3 2.8  
## 4 PERC 1 26 23 9.3 4.5 2.1  
## 5 PWD 8 161 29 9.4 8.1 2.8  
## 6 MSM 8 219 31 9.7 9.4 3.0  
## 7 MOLF 6 119 24 10.0 6.0 2.4  
## 8 SKIVE 6 211 27 9.8 7.1 2.7  
## 9 KAZ 13 264 24 9.4 5.6 2.3  
## 10 TOM 2 10 25 13.6 11.2 3.4  
## 11 AST 3 63 25 10.1 6.6 2.6  
## 12 CZECH 3 62 23 10.0 5.4 2.3  
## 13 CAST 2 44 23 9.9 5.2 2.3  
## 14 HMI 4 44 24 8.1 4.6 2.0  
## mean.mouse\_se  
## 1 0.66  
## 2 0.57  
## 3 0.57  
## 4 0.41  
## 5 0.67  
## 6 0.62  
## 7 0.62  
## 8 0.46  
## 9 0.56  
## 10 1.50  
## 11 0.57  
## 12 0.51  
## 13 0.50  
## 14 0.63

The same model framwrok was applied to the male specific data. The male specific analysis was done to assess the variance across strains. The plots below illustrate mouse level means for MLH1 per cell serperated by subspecies.

*M. m. domesticusLEW*

* There is a low degree of strain varaince in Dom with the range of mouse means 22.17 to 30.36
* Musc and Mol have a much larger amount of variance across means with the range in mouse means being 21.87 to 31.63 in Musc and 23.18 to 33.04 and Molossinus.
* While there is alot of variance within strains, a general pattern is that PWD, SKIVE and MSM can be classified as ‘high rec’ strains, there strain averages are

Two models were fitted to the male specific data to test the effects of strain and subspecies.

## Estimate Pr(>|t|)  
## (Intercept) 24.453 0.00000  
## subspCast -0.099 0.90777  
## subspMusc -1.260 0.18934  
## subspMol -0.213 0.77317  
## strainG -0.256 0.64213  
## strainLEW 0.591 0.35227  
## strainPERC -1.645 0.28680  
## strainPWD 6.113 0.00000  
## strainMSM 7.214 0.00000  
## strainSKIVE 3.882 0.00036  
## strainKAZ 1.040 0.27413  
## strainTOM 1.506 0.26629  
## strainAST 2.105 0.08403  
## strainCAST -1.345 0.29514

## Estimate Pr(>|t|)  
## (Intercept) 24.453 0.00000  
## strainG -0.256 0.64213  
## strainLEW 0.591 0.35227  
## strainPERC -1.645 0.28680  
## strainPWD 4.853 0.00000  
## strainMSM 7.001 0.00000  
## strainMOLF -0.213 0.77317  
## strainSKIVE 2.622 0.00063  
## strainKAZ -0.220 0.71046  
## strainTOM 0.247 0.82703  
## strainAST 0.846 0.37670  
## strainCZECH -1.260 0.18934  
## strainCAST -1.444 0.20325  
## strainHMI -0.099 0.90777

the coefficients for glms of the CO means are listed above.

## Estimate Pr(>|t|)  
## (Intercept) 7.50 0.000  
## subspCast -2.88 0.141  
## subspMusc -2.10 0.335  
## subspMol -1.52 0.367  
## strainG -1.40 0.265  
## strainLEW 0.82 0.568  
## strainPERC -3.02 0.390  
## strainPWD 2.73 0.233  
## strainMSM 3.43 0.062  
## strainSKIVE 1.69 0.478  
## strainKAZ 0.20 0.925  
## strainTOM 5.85 0.059  
## strainAST 1.16 0.673  
## strainCAST 0.57 0.845

## Estimate Pr(>|t|)  
## (Intercept) 7.50 0.00  
## strainG -1.40 0.27  
## strainLEW 0.82 0.57  
## strainPERC -3.02 0.39  
## strainPWD 0.63 0.68  
## strainMSM 1.91 0.22  
## strainMOLF -1.52 0.37  
## strainSKIVE -0.41 0.81  
## strainKAZ -1.90 0.16  
## strainTOM 3.75 0.15  
## strainAST -0.94 0.67  
## strainCZECH -2.10 0.33  
## strainCAST -2.31 0.37  
## strainHMI -2.88 0.14

## Estimate Pr(>|t|)  
## (Intercept) 10.86 0.000  
## strainG -1.10 0.216  
## strainLEW 0.34 0.740  
## strainPERC -1.58 0.524  
## strainPWD -1.42 0.194  
## strainMSM -1.17 0.281  
## strainMOLF -0.86 0.472  
## strainSKIVE -1.04 0.382  
## strainKAZ -1.44 0.133  
## strainTOM 2.77 0.130  
## strainAST -0.73 0.634  
## strainCZECH -0.86 0.574  
## strainCAST -0.97 0.595  
## strainHMI -2.78 0.046

## Estimate Pr(>|t|)  
## (Intercept) 5.55 0.0018  
## subsp3 1.54 0.5069  
## subspMol -0.80 0.7841  
## strainG 1.35 0.5402  
## strainLEW 0.28 0.9142  
## strainPWD -0.10 0.9656  
## strainMSM 0.14 0.9679  
## strainSKIVE -1.26 0.6342

## Estimate Pr(>|t|)  
## (Intercept) 5.55 0.0018  
## strainG 1.35 0.5402  
## strainLEW 0.28 0.9142  
## strainPWD 1.43 0.5609  
## strainMSM -0.66 0.8200  
## strainMOLF -0.80 0.7841  
## strainSKIVE 0.27 0.9196  
## strainKAZ 1.54 0.5069

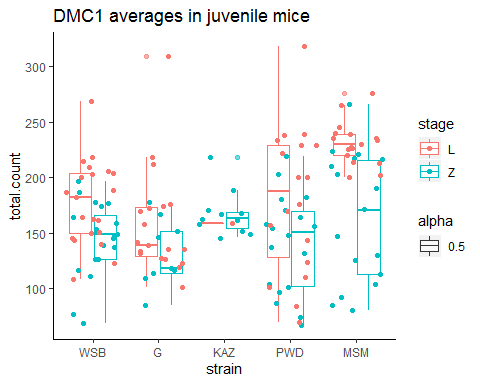
The coefficients for the variance models are listed above

The strain average for *M. m. musculusPWD* , *M. m. molossinusMSM* and *M. m. musculusSKIVE* are 1.2, 1.29, and 1.11 higher than the other strain means respectively. Due to this and the significant strain effects they will be designedated into the high rec group for later analyses.

# Variation in DSB number per cell

## category ncells mean.count stage  
## 1 WSB male 21 178 L  
## 2 G male 19 158 L  
## 3 PWD male 18 180 L  
## 4 MSM male 17 231 L  
## 5 KAZ male 1 159 L  
## 6 WSB male 20 144 Z  
## 7 G male 9 132 Z  
## 8 PWD male 18 141 Z  
## 9 MSM male 17 164 Z  
## 10 KAZ male 11 167 Z

In order to get a good idea on wether variation with mean CO number per cell is associated with the number of percursors upstream in prophase, we quantified DMC1 foci, a marker for DSBs. DMC1 foci were scored from X 76 leptotene and 75 zygotene staged spermatocytes of juvenile mice (12-14-18 days). A subset of strains were quantified *M. m. musculusPWD*, *M. m. molossinusMSM*, *M. m. musculusKAZ* , *M. m. domesticusWSB* , and *M. m. domesticusG*. Leptotene and zygotene cells were staged based on SC-AE and centromere morphology.



(discription of the nice boxplot figure)

* overall the range is very high (CO homeostasis, technical noise, process is noiser continous repair and new DSBs)
* exclude KAZ?
* expected difference between cell stages for all strains
* pattern of musc strains being higher than the dom strains (subspe pooled t-test)
* MSM is much higher
* the ratios of DMC1 : MLH1 counts!!

the pvalues for the differences between time points are 1.0310^{-5}, 1.110^{-4} , and 0.02 for all observations, the high rec group and the low group respectively.

## Warning: `cols` is now required.  
## Please use `cols = c(p)`

## Warning: Ignoring unknown aesthetics: xmin, xmax, annotations, y\_position

The correlation with MLH1 and leptotene cells is 0.87.

The correlation with MLH1 and zygotene cells is 0.28.

Why calculate ratios using the mean Leptotene cells – they represent the a metric for the CO : NCO decision ( DSB / COs = NCO)

ratios are .. not that different (slightly suprisingly)

7.27, 6.54, 7.34, 6.15 for *M. m. domesticusWSB* , *M. m. domesticusG*, , *M. m. molossinusMSM* and *M. m. musculusPWD* respecitively.

These are the DSB:CO ratios for the L means 6.54, 6.56, 7.34, 6.15, 7.27 and Z means 5.45, 6.91, 5.23, 4.8, 5.9.

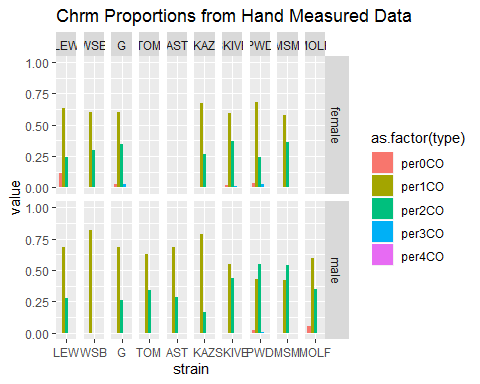
(maybe ignore KAW.L? 1 cell)

Since there is evidence for non-equal variance across strains for zygotene cells, so don’t rely on lm()s that estimate the effect of strains.

The p-values from the t.tests of the high vs low groups, indicate that the high recombining are significantly higher for the L cells (p value = 0) while the zygotene cells, there is not a significant difference across the high and low groups (value = 0.66).

# Chromosome Class Proportions

In order to decompose the cell wide rate, we decided to look at the proportion or chromosomes with different numbers of COs. The two plots show the chromosome class proportions from hand measured and the curated BivData.



These results are meant to compare the proportions of bivalents with 0,1,2 or 3 chromosomes. Most of the variation in gwRR across strains in is due to more 2COs at the ‘expense’ / trade off of 1COs.

Most all the the p values for the proportion tests are significant, indicating there are slight but significant shifts across the classes of chromosomes. However the most striking male pattern is the propotion of 2COs

(A previously reported for house mouse, the most prevelent class of chromosomes is the 1CO class. The high rec group of males are the exception indicating, which fits with the conclusion that higher cell wide CO counts are due to more chromosomes(bivalents) have 2 instead of 1 CO.

* High female strains, G and Lew have significantly more 2CO bivalents.
* The overall male pattern that’s most striking is the gradient of 2CO proportions

**MSM 60% PWD 50% SKIVE 30% 20 - 10% remaining (low) strains**

# Single Bivalent Level Results

In order to better deconstruct the genome wide recombiation rate, we decided to deconstruct cell wide average by examining single chromosome patterns.

From our total set of cell images 83975 chromosome objects were isolated by the image analysis software. After the human curation step, 83975 single bivalent measures were left.

## Var1 Freq  
## 1 WSB female 767  
## 2 WSB male 592  
## 3 G female 726  
## 4 G male 1004  
## 5 LEW female 714  
## 6 LEW male 533  
## 7 PWD female 1031  
## 8 PWD male 678  
## 9 MSM female 550  
## 10 MSM male 173  
## 11 MOLF male 437  
## 12 SKIVE female 488  
## 13 SKIVE male 775  
## 14 KAZ female 485  
## 15 KAZ male 615  
## 16 CZECH male 239

Table X illustrates the break down of single bivalent observations by category

Error compared to human measures and other details have been discribed in ((**???**)).

While the automated software doesn’t isolate all bivalents/chromosomes from each cell (on average 17), we assume that the isolation process is not biased. Because there are hundreds of observations per category, we assume that each of the 19 autosomes (chromosomes) is equally represented in the dataset of single bivalents.

The mouse averages for 3 bivalent level metrics will be analyzed across these questions:  
1)SC lengths (or total.SC)/chromatin compaction (using all cells regaurdless of CO number) 2) Interfocal Distance of 2CO bivalents and 3) the normalized foci position from 1CO bivalents.

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

These questions use different datasets: Q1 uses a dataset with only sex matched strains and Q2 uses the data set with all male observations with including strains including those not in the Q1 set.

(We will first list the results for the Q1)

These are our predictions for the sex specifici patterns on the single bivalent lanscape

# Q1 Analysis, Predictions for Heterochiasmy Q1

Two bivalent level traits are predicted to display heterochiasmy (ie significant effects of sex);

1. SC length will be sexually dimorphic (sex effect will be significant)(cite Lynn)
2. Normalized 1CO positions will be sexually dimorphic (sedell and Kirkpatrick).
3. Interference / IFD will not be sexually dimorphic. Previous physical measures of interference were not different between sexes (deBoer et al 1996, petkov 2001).

The same basic models from the MLH1 counts per cell will be used. In addition to basic t.tests and logistic regression models for Q2 to distinguish betwee high and low recombining strains.

Q1. Model 1, mixed model lmer()

Q1. Model 2, glm()

Q2 Model 3, logistic regression

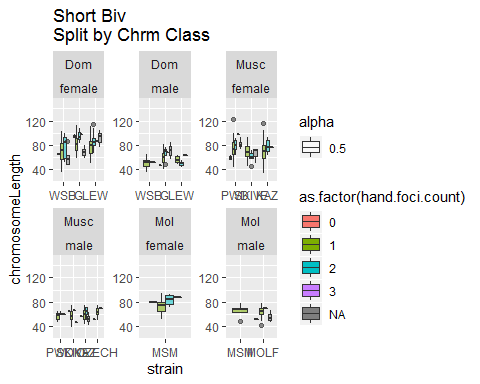
In the chunk above the mouse averages table is made – may need to add all the extra metrics (IFD, .

# Q1 SC Lengths

We expect female SC lengths to be longer (refs). In the plot above the SC length ~ higher hand foci cells. Any of the chromosome classes above 3, that don’t have a higher mean are likely due to low data number. For the 0 class chromosomes most all are around the same size of the 1CO distribution. Add in the code for the plot under 2COs males and females

* Caveats XX

The most convincing should be the short biv data set



the models for testing

M1. subsp \* sex + (strain)

M2. subsp \* sex \* strain (slightly redundant model)

M3. sex \* strain

Below’s the code for Mixed model results – try to organize them into a table or something

Because of confonding effects across chromsomes of different physical lenghts, we attempt to … reduce the error from mixing chromosomes by constructing reduced bivalent sets, for the longest 5 and shortest 5 bivalents from the same cell. The mouse means for these two data sets were then calculated.

## refitting model(s) with ML (instead of REML)  
## refitting model(s) with ML (instead of REML)

## sub.sp sex.results interaction.pval rand.pval  
## 1 0.25 1.4e-11 0.087 0.25

## refitting model(s) with ML (instead of REML)  
## refitting model(s) with ML (instead of REML)

## sub.sp sex.results interaction.pval rand.pval  
## 1 0.25 1.4e-11 0.087 0.25

Above are the p values for the components of the mixed models for the short and long biv mouse means. In both sets of models the sex effect is the most significant p value.

## refitting model(s) with ML (instead of REML)  
## refitting model(s) with ML (instead of REML)

## sub.sp sex.results interaction.pval rand.pval  
## 1 0.69 6.3e-24 0.88 0.011

## boundary (singular) fit: see ?isSingular  
## refitting model(s) with ML (instead of REML)  
## refitting model(s) with ML (instead of REML)

## sub.sp sex.results interaction.pval rand.pval  
## 1 0.74 5.6e-16 0.46 0.31

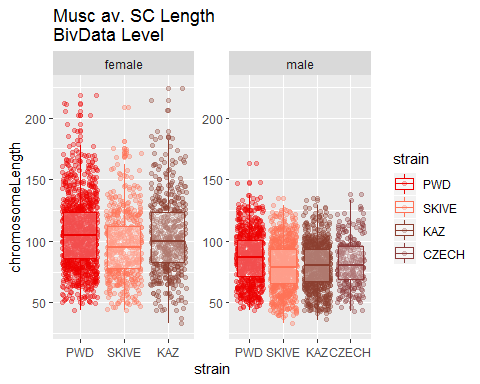
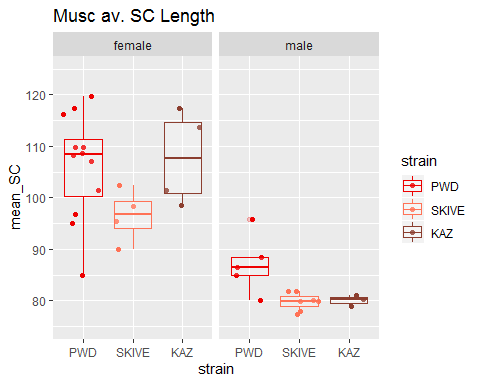
Above are the mixed model results seperated by mouse averages for 1 crossover and 2 crossover bivalents respectively.

The fixed sex effect is highly significant for the mouse means of SC length. The data sets tested were all pooled SC lengths, the long bivalents, 1CO and 2CO seperately.

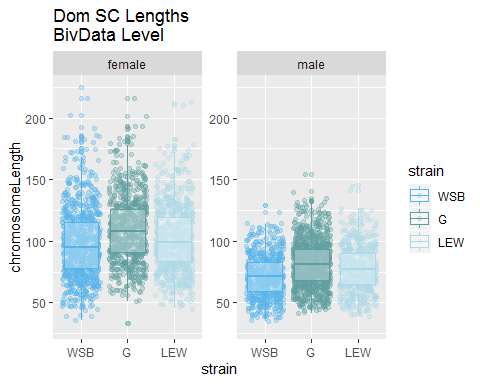
All fixed and all interactions

## Warning: Removed 1 rows containing non-finite values (stat\_boxplot).

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



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



I think the strange thing about the Musc is that Skive females have shorter SC lengths.

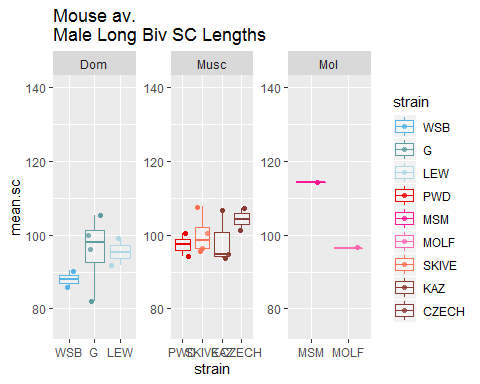
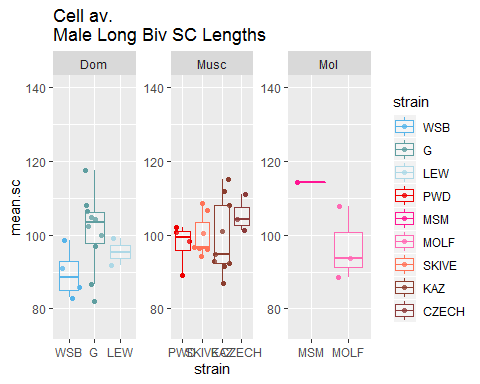
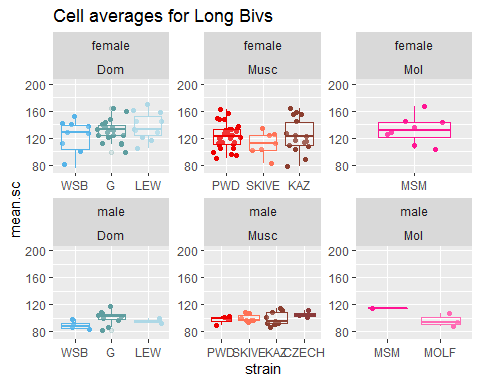
#### Q1. SC Length t.test

#### Strain Comparisons of SC Length

Dive deeper into the sex specific pattern for each strain. (I think for all the within strain comparisons across sexes the general pattern is met of females being longer)

(adding all of these blocks to – the setup script); strain.Bivalents.DF

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



#### XX adjustment

Generate inteligible summaries for the permutation / subsampling results.

# Q1. IFD

Brief lit review for IFD / interference expectations  
**Petkov et al 2007, CO interference underlies sex differences in RR**

“Here we show that in mice, this is because of a shorter genomic interference distance in females than in males, measured in Mb. However, the interference distance is the same in terms of bivalent length. We propose a model in which the interference distance in the two sexes reflects the compaction of chromosomes at the pachytene stage of meiosis.”

Chrm1 genetic map,

(other human and mice refs

Tease n Hulten 2004: – no difference between MLH1 foci in males and females

DeBoer 2006, 13 – measured chrm 1 in males and females, both were 2.8 microns

))

Still working on the best way to display the general IFD patterns.

### Mixed Model Tests, Fixed Effects

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

## boundary (singular) fit: see ?isSingular  
## boundary (singular) fit: see ?isSingular  
## boundary (singular) fit: see ?isSingular

## refitting model(s) with ML (instead of REML)  
## refitting model(s) with ML (instead of REML)

## boundary (singular) fit: see ?isSingular

## 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

## refitting model(s) with ML (instead of REML)

## refitting model(s) with ML (instead of REML)

## boundary (singular) fit: see ?isSingular

## 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

## Estimate Pr(>|t|)  
## (Intercept) 53.66 0.000  
## sexmale -5.72 0.258  
## strainG 6.58 0.177  
## strainLEW 2.63 0.560  
## strainPWD 2.11 0.634  
## strainMSM 6.73 0.207  
## strainMOLF 0.10 0.984  
## strainSKIVE -1.29 0.808  
## strainKAZ 8.78 0.084  
## sexmale:strainG -2.77 0.665  
## sexmale:strainLEW 1.28 0.845  
## sexmale:strainPWD 9.95 0.129  
## sexmale:strainMSM -0.23 0.975  
## sexmale:strainSKIVE 13.18 0.059  
## sexmale:strainKAZ -8.70 0.244

## Estimate Pr(>|t|)  
## (Intercept) 0.4619 0.000  
## sexmale 0.0807 0.008  
## strainG 0.0258 0.367  
## strainLEW 0.0062 0.817  
## strainPWD 0.0142 0.588  
## strainMSM 0.0128 0.682  
## strainMOLF -0.0130 0.661  
## strainSKIVE 0.0199 0.525  
## 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 0.0192 0.656  
## sexmale:strainSKIVE 0.0788 0.055  
## 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 ansD 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 consistantly significant effect, which also increases the nrm.IFD measure.

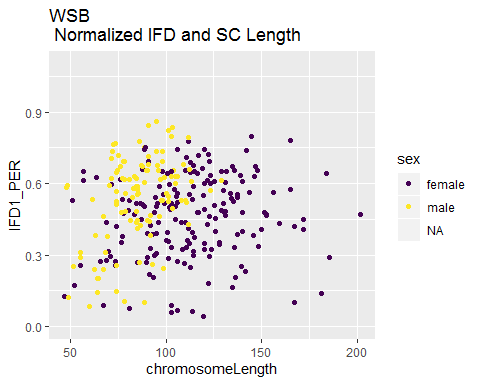
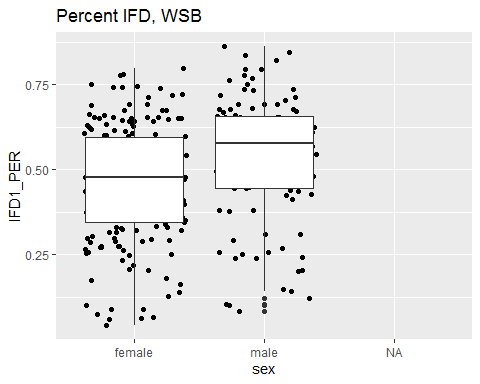
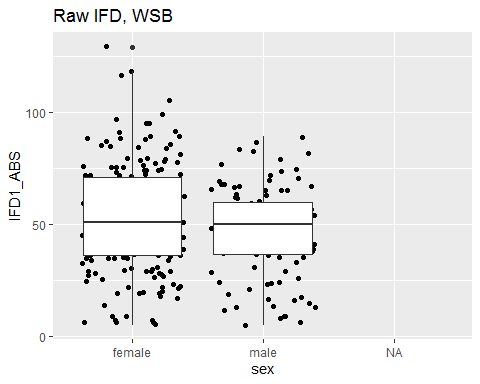
Overall There’s a low amount of significant effects across the 2CO IFD measures. This might be an indication that interference is conserved across these samples and/or that there is too much noise across from chromosome specific effects.

## Strain Comparisons

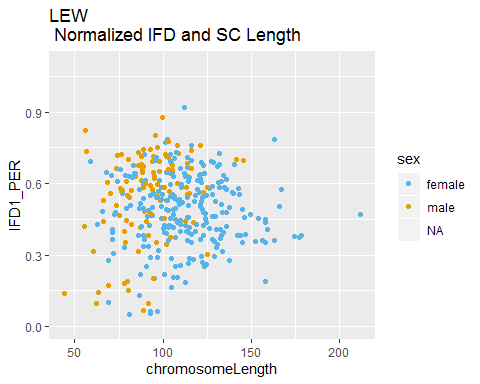
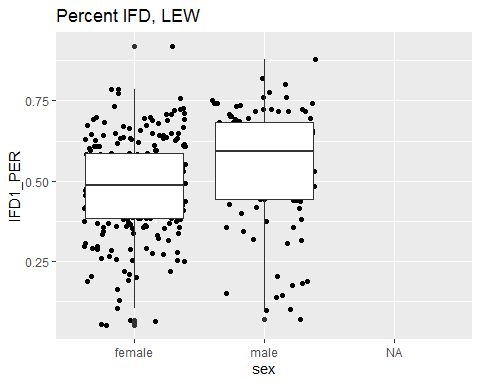
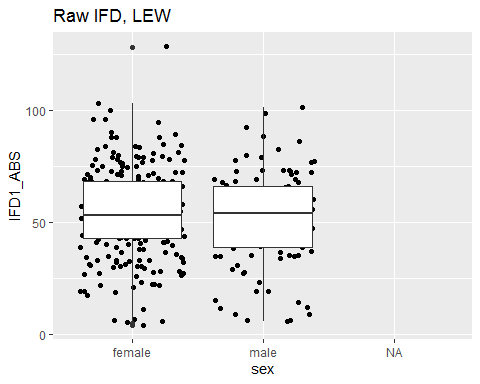
Dive deeper into the sex specific pattern for each strain. Below are code chunks which show the unusual sex specific results for IFD measures. The general pattern is that, female raw IFD > male IFD and female PER IFD < male PER IFD. The scatter plots show that female raw measures are longer than male 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?

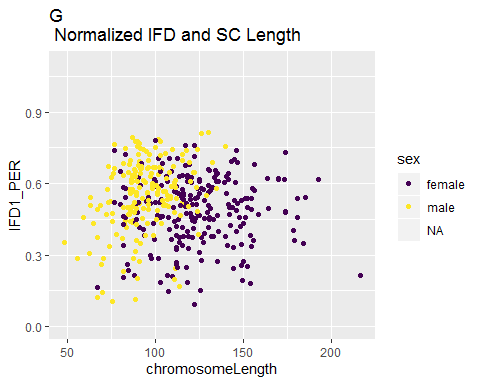
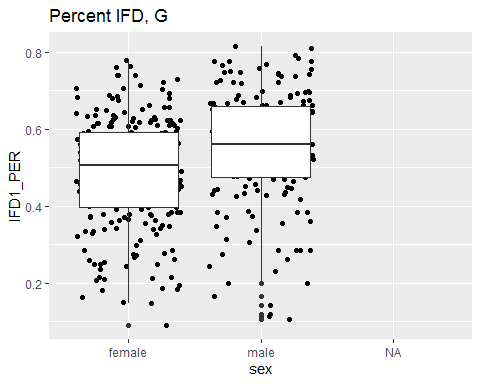
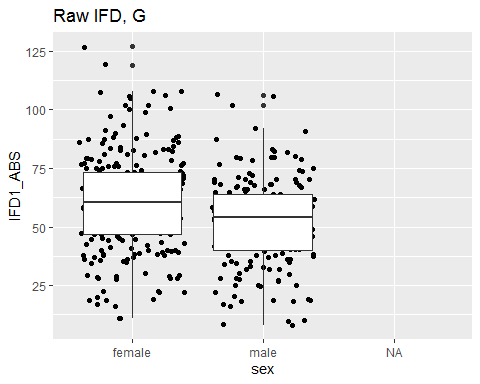
Above is a table of the proportion of 2Co bivalents which have a norm IFD below 30%, For all strains but KAZ, the females have a greater proportion of these shorted IFD values.

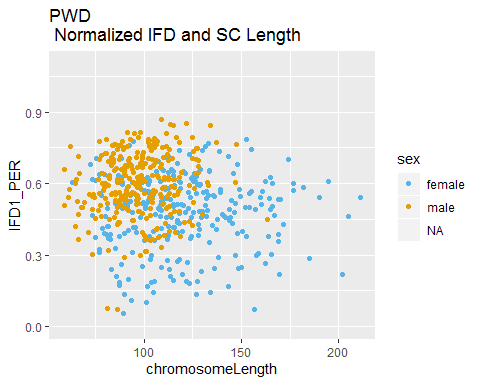
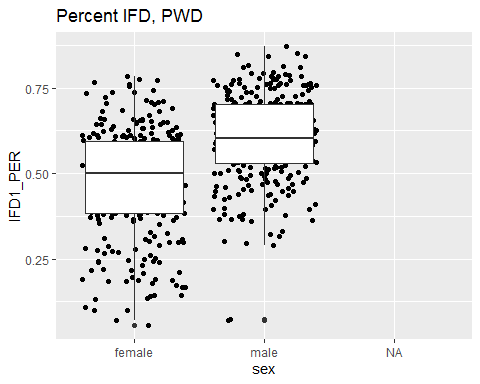
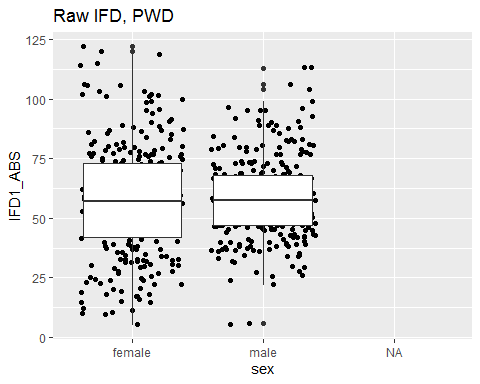


The range of normalized IFDs overlap closer in males and females in the WSB data.

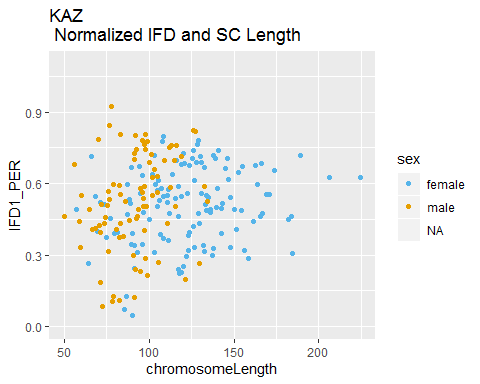
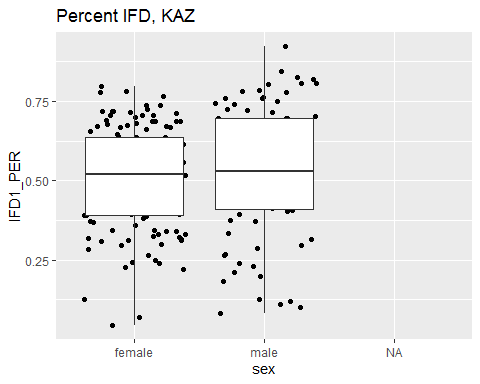
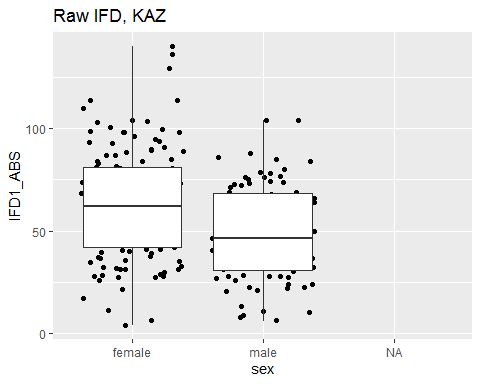


The Lew pattern doesn’t have a clean cut off of nrm.IFD. the range of male and females overlap, but there are more female observations below.

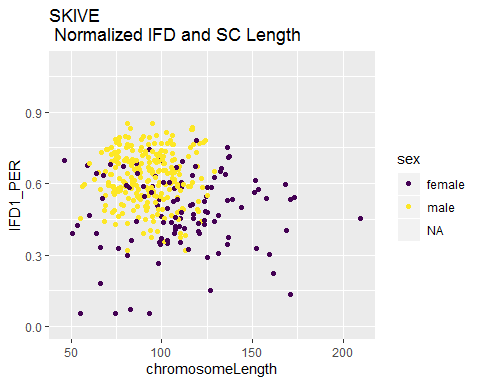
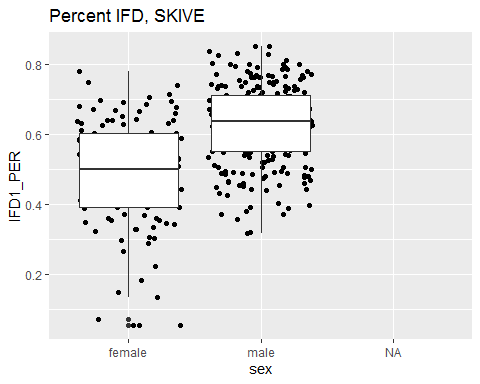
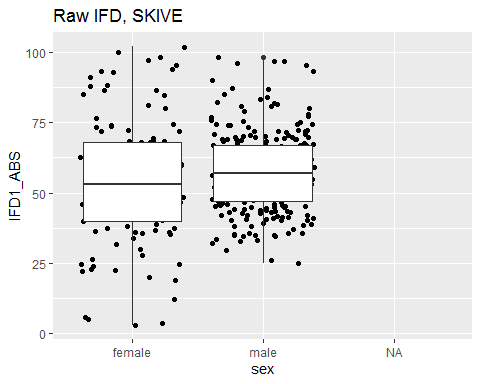




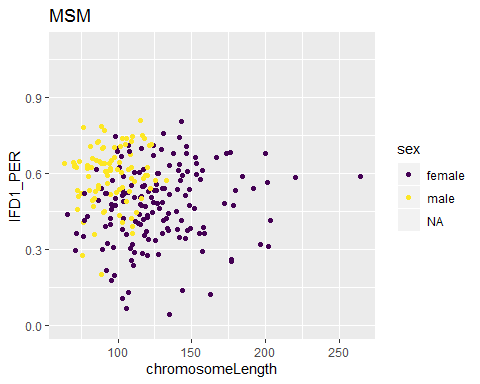
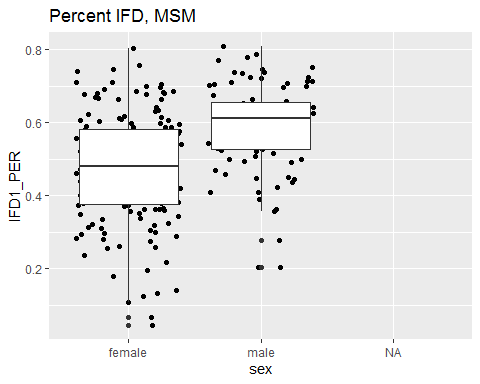
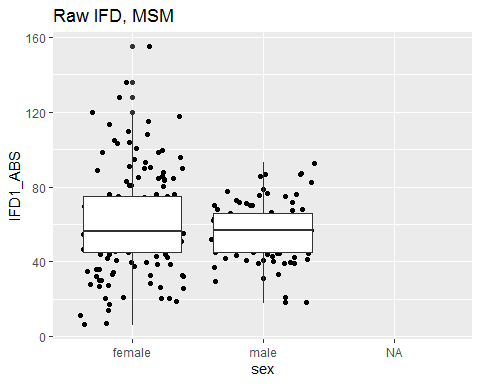
For PWD, there are a few observations of the short IFDs for males, but there seems to be a cut-off / threshold at .3



For the KAZ, pattern the distinction between the male and female pattern is less distinct. There are fewer instances of females with v close IFD distances.



In the Skive data, it could be the case that the v. short IFD measures in females are rare / another class of observations.



The MSM pattern has a short range and longer range of nrm.IFD in males and females respectively.

##   
## 0 1 2 3  
## female 0 0 0 0  
## male 9 326 97 5

The strains which show a clean “30% threshold” for normalized IFD in males are: PWD, SKIVE, and MSM (which are the 2 high Rec and a intermediate strain). The other strains which have more overlap between males and females are the Dom strains and KAZ.

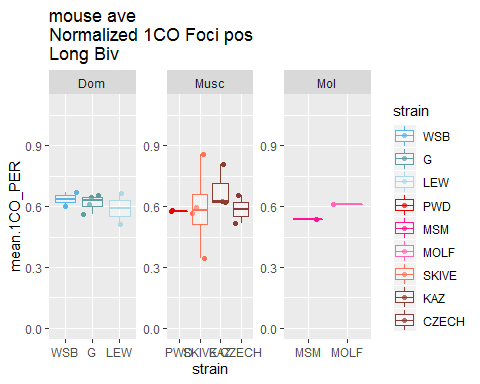
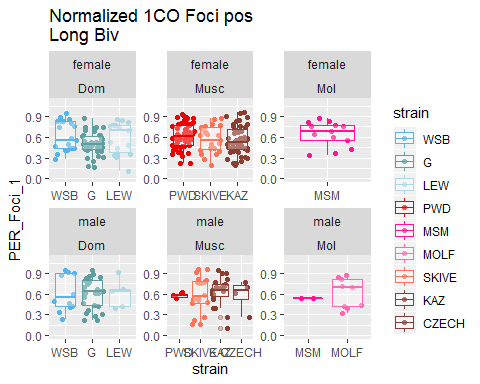
#### IFD 3CO bivalents

Run comparisons for 3CO bivalents.

# Q1. General CO Positions

Try accounting for chrm size effects. **Also i didn’t notice that I’m using density plots and scatter+boxplots across sections**

density plots



## refitting model(s) with ML (instead of REML)

## Data: Mouse.Table\_BivData\_4MM  
## Models:  
## lmer.Q1.nrm1CO\_M1reduce.sex: norm.first.foci.pos ~ subsp + (1 | strain)  
## lmer.Q1.nrm1CO\_M1: norm.first.foci.pos ~ subsp \* sex + (1 | strain)  
## Df AIC BIC logLik deviance Chisq Chi Df  
## lmer.Q1.nrm1CO\_M1reduce.sex 5 -201 -189 105 -211   
## lmer.Q1.nrm1CO\_M1 8 -311 -291 163 -327 116 3  
## Pr(>Chisq)   
## lmer.Q1.nrm1CO\_M1reduce.sex   
## lmer.Q1.nrm1CO\_M1 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## refitting model(s) with ML (instead of REML)

## Data: Mouse.Table\_BivData\_4MM  
## Models:  
## lmer.Q1.nrm1CO\_M1reduce.subp: norm.first.foci.pos ~ sex + (1 | strain)  
## lmer.Q1.nrm1CO\_M1: norm.first.foci.pos ~ subsp \* sex + (1 | strain)  
## Df AIC BIC logLik deviance Chisq Chi Df  
## lmer.Q1.nrm1CO\_M1reduce.subp 4 -312 -302 160 -320   
## lmer.Q1.nrm1CO\_M1 8 -311 -291 163 -327 7.02 4  
## Pr(>Chisq)  
## lmer.Q1.nrm1CO\_M1reduce.subp   
## lmer.Q1.nrm1CO\_M1 0.13

## refitting model(s) with ML (instead of REML)

## Data: Mouse.Table\_BivData\_4MM  
## Models:  
## lmer.Q1.nrm1CO\_M1reduce.sex: norm.first.foci.pos ~ subsp + (1 | strain)  
## lmer.Q1.nrm1CO\_M1: norm.first.foci.pos ~ subsp \* sex + (1 | strain)  
## Df AIC BIC logLik deviance Chisq Chi Df  
## lmer.Q1.nrm1CO\_M1reduce.sex 5 -201 -189 105 -211   
## lmer.Q1.nrm1CO\_M1 8 -311 -291 163 -327 116 3  
## Pr(>Chisq)   
## lmer.Q1.nrm1CO\_M1reduce.sex   
## lmer.Q1.nrm1CO\_M1 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## refitting model(s) with ML (instead of REML)

## Data: Mouse.Table\_BivData\_4MM  
## Models:  
## lmer.Q1.nrm1CO\_M1reduce.subp: norm.first.foci.pos ~ sex + (1 | strain)  
## lmer.Q1.nrm1CO\_M1: norm.first.foci.pos ~ subsp \* sex + (1 | strain)  
## Df AIC BIC logLik deviance Chisq Chi Df  
## lmer.Q1.nrm1CO\_M1reduce.subp 4 -312 -302 160 -320   
## lmer.Q1.nrm1CO\_M1 8 -311 -291 163 -327 7.02 4  
## Pr(>Chisq)  
## lmer.Q1.nrm1CO\_M1reduce.subp   
## lmer.Q1.nrm1CO\_M1 0.13

## Single term deletions  
##   
## Model:  
## norm.first.foci.pos ~ subsp \* sex + (1 | strain)  
## Df AIC LRT Pr(Chi)  
## <none> -311   
## subsp:sex 2 -310 4.43 0.11

##   
## simulated finite sample distribution of RLRT.  
##   
## (p-value based on 10000 simulated values)  
##   
## data:   
## RLRT = 10, p-value <2e-16

## Estimate Pr(>|t|)  
## (Intercept) 0.5894 0.00000  
## subspMusc -0.0423 0.07674  
## subspMol -0.1116 0.00142  
## strainG -0.0373 0.10652  
## strainLEW -0.0487 0.02662  
## strainPWD 0.0248 0.15343  
## strainMSM 0.0971 0.00006  
## strainSKIVE 0.0301 0.16845  
## sexmale 0.1422 0.00000  
## subspMusc:sexmale -0.0099 0.76666  
## subspMol:sexmale -0.0284 0.38886  
## strainG:sexmale -0.0247 0.39850  
## strainLEW:sexmale 0.0096 0.74727  
## strainPWD:sexmale -0.0356 0.22582  
## strainSKIVE:sexmale -0.0298 0.34596

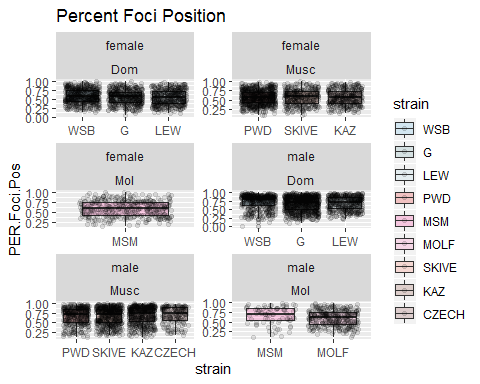
The model results aren’t as clear as I’d like. For the Mol, both the sex effect and MOLF strain effect are significant (this means both female and male rec landscapes are effected - most of the time towards the middle

GENERALLY - Sex is the biggest effect for the 1CO landscape (female middle male telomereic) molf Not sure I understand how MSM is supper significant for There must be some strange effect due to the Mol subsp - MOLF isn’t even in M2 ( Male is consistently significant across the two models.

Male and MOLF are the most significant effects for M3

Above plot focuses on the 1CO bivalent normalized positions since CO interference controls the general position of COs when there are multiple COs. This plot shows the sexual dimorphism in the density plots.

Consider adding annotate\_text for the number of observations in each category. think about adding a vertical line for centromere, for the position means. Think about removing the extra Musc strains.



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.

the distribution of SC lengths and sis-coten seems very different across sexes

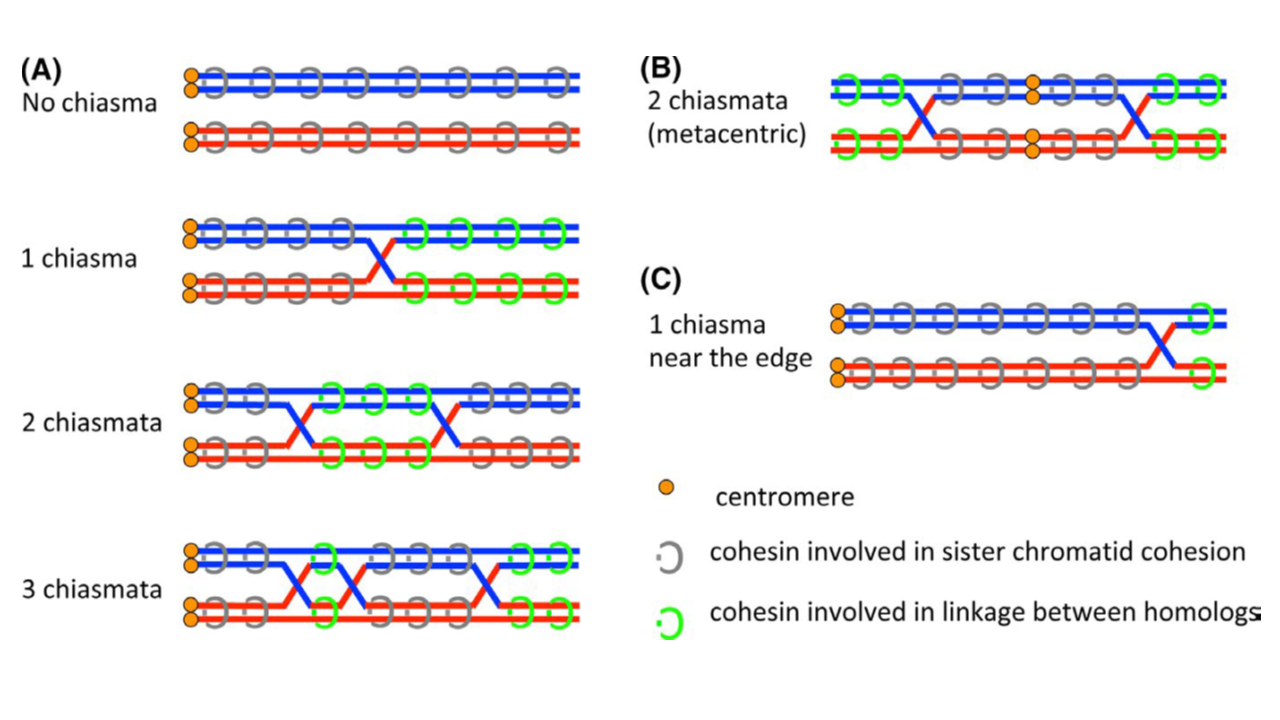
The mixed model data should only come from 1CO bivalent data.

the mouse average foci1 pos is more significant in t.test, but not log regression… (is something wrong?) Check the mouse averages for the F1\_pos, there might be an outlier or mouse with v.few observations.

### Siscoten

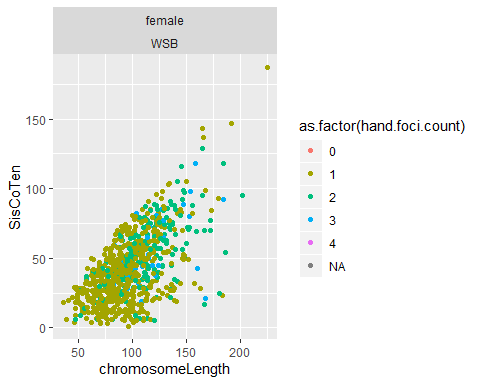
The metric Sis-co-ten measures the amount of sister cohesion connected to the other pole.

The logic of how the sis-co-ten metric is outlined in the figure below. The goal is to use this metric to model different tension active cohesion amounts as a consequence of different numbers and placements of chiasmata/CO. This metric is calculated using SC area as a proxy to the amount of cohesion at metaphase.

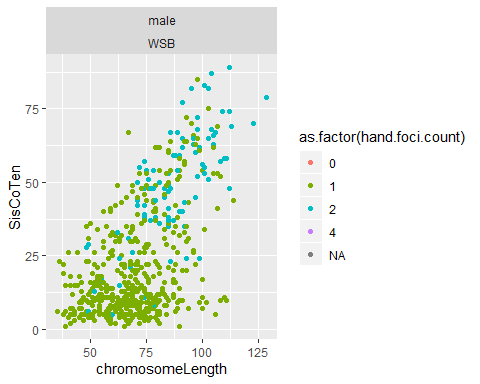


from (Lee, J. (2019). Is age-related increase of chromosome segregation errors in mammalian oocytes caused by cohesin deterioration?. Reproductive Medicine and Biology.)

## Warning: Removed 133 rows containing missing values (geom\_point).



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

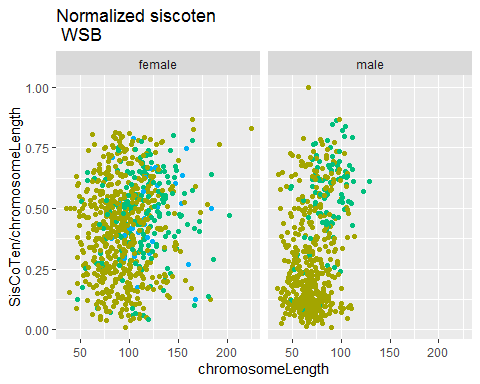


Males have much clearer separation of siscoten across chrm classes. This is emphasized when SC length is also plotted. It seems like musc males have higher amounts of this metric compared to Dom males.

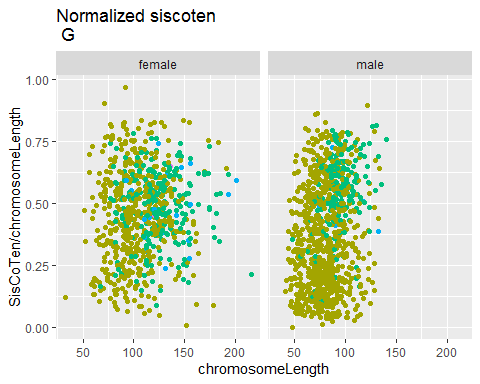
To formally test the differences in sis-co-ten I plan to write a sub sampling / permutation loop to compare the mean(sis.co.ten) of the same numbers of bivalents of the same class.

BUT females have a greater range – so maybe it’s just a scale issue.

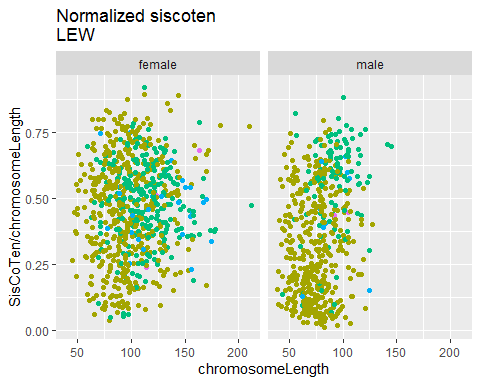
## Warning: Removed 24 rows containing missing values (geom\_point).



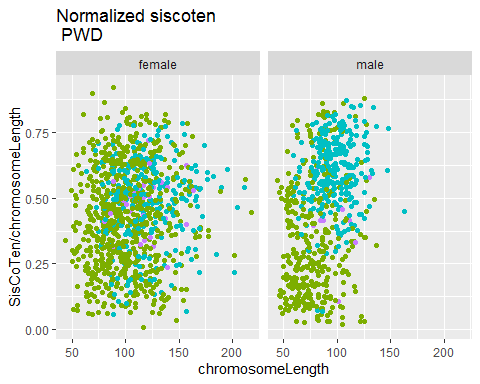
## Warning: Removed 21 rows containing missing values (geom\_point).



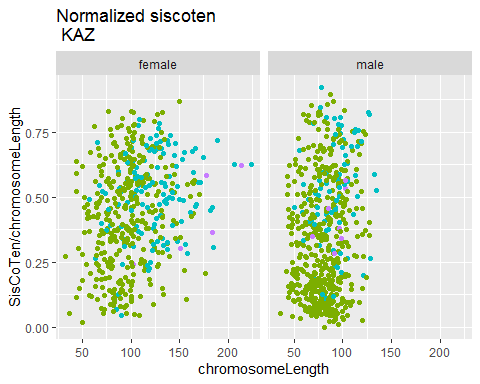
## Warning: Removed 12 rows containing missing values (geom\_point).



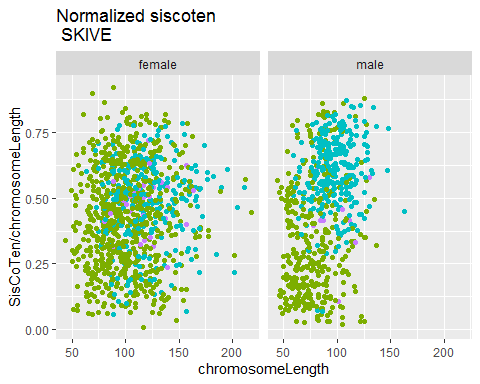
## Warning: Removed 16 rows containing missing values (geom\_point).



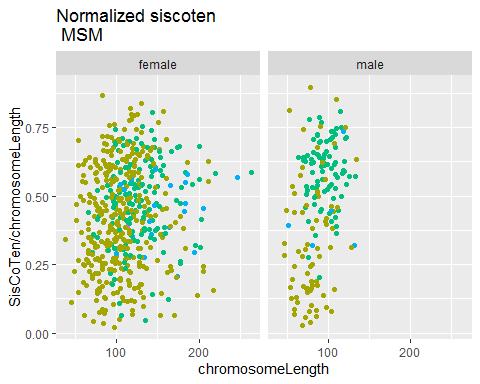
## Warning: Removed 23 rows containing missing values (geom\_point).



## 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.

The fixed effects, sex and sex\*subsp are significant. The random strain effect is also significant.

Is the heterochiasmy prediction met?

Yes, model predicting the mouse average siscoten, sex and sex-subp interaction are significant factors. The Random strain effect is also significant.

##   
## 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   
## -1.5256 -0.0661 0.0042 0.0773 1.4045   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -47.208 30.908 -1.53 0.13  
## mean.siscoten 1.451 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: 5.2122 on 16 degrees of freedom  
## AIC: 9.212  
##   
## Number of Fisher Scoring iterations: 9

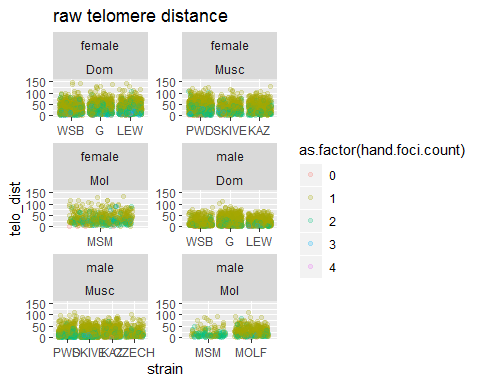
##   
## Call:  
## glm(formula = Rec.group ~ SisCoTen, family = binomial(link = "logit"),   
## data = Curated\_BivData[(Curated\_BivData$subsp == "Musc") &   
## (Curated\_BivData$sex == "male"), ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.239 -0.862 -0.725 1.356 1.792   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.41070 0.08759 -16.1 < 2e-16 \*\*\*  
## SisCoTen 0.01494 0.00194 7.7 1.4e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2761.4 on 2269 degrees of freedom  
## Residual deviance: 2701.2 on 2268 degrees of freedom  
## (37 observations deleted due to missingness)  
## AIC: 2705  
##   
## 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.

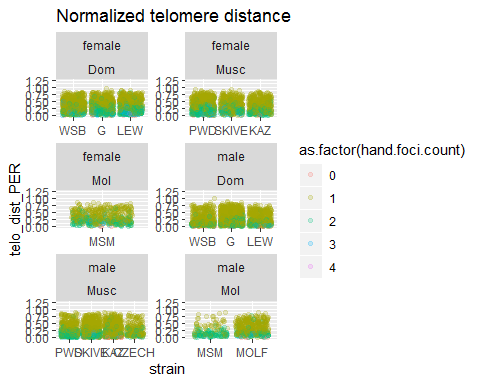
### Telomere and centromere Distance

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 78 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 1.03 0.6

## Estimate Std. Error t value  
## (Intercept) 34.2 2.1 16.06  
## subspMusc 1.2 3.0 0.41  
## subspMol 5.4 4.2 1.30  
## sexmale -14.5 1.6 -8.88  
## subspMusc:sexmale -2.2 2.4 -0.91  
## subspMol:sexmale -3.2 3.7 -0.87

## Single term deletions  
##   
## Model:  
## mean.telo.dist ~ subsp \* sex + (1 | strain)  
## Df AIC LRT Pr(Chi)  
## <none> 540   
## subsp:sex 2 537 1.03 0.6

##   
## simulated finite sample distribution of RLRT.  
##   
## (p-value based on 10000 simulated values)  
##   
## data:   
## RLRT = 7, p-value = 0.004

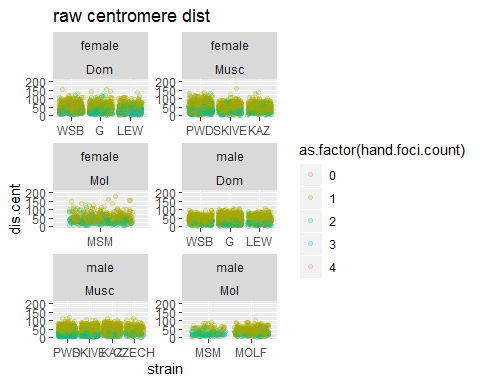
Mixed model result summary:

##   
## 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 0.443 0.595 0.743 0.851   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 4.588 5.996 0.77 0.44  
## mean.telo.dist -0.172 0.315 -0.54 0.59  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 15.012 on 14 degrees of freedom  
## Residual deviance: 14.673 on 13 degrees of freedom  
## AIC: 18.67  
##   
## Number of Fisher Scoring iterations: 4

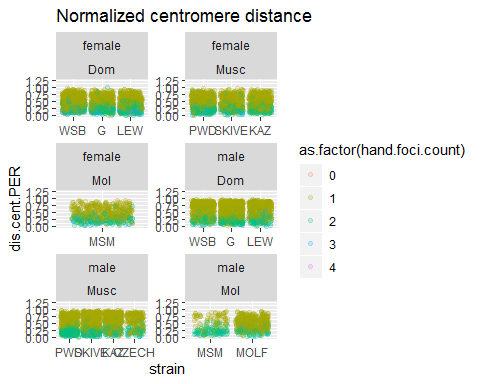
##   
## Call:  
## glm(formula = Rec.group ~ telo\_dist, family = binomial(link = "logit"),   
## data = Curated\_BivData[(Curated\_BivData$subsp == "Musc") &   
## (Curated\_BivData$sex == "male"), ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.870 -0.853 -0.823 1.533 1.763   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.77560 0.06987 -11.10 <2e-16 \*\*\*  
## telo\_dist -0.00492 0.00261 -1.89 0.059 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2790.4 on 2303 degrees of freedom  
## Residual deviance: 2786.7 on 2302 degrees of freedom  
## (3 observations deleted due to missingness)  
## AIC: 2791  
##   
## 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   
## -0.902 -0.867 -0.805 1.503 1.775   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.6882 0.0721 -9.54 <2e-16 \*\*\*  
## telo\_dist\_PER -0.7449 0.2263 -3.29 0.001 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2790.4 on 2303 degrees of freedom  
## Residual deviance: 2779.2 on 2302 degrees of freedom  
## (3 observations deleted due to missingness)  
## AIC: 2783  
##   
## Number of Fisher Scoring iterations: 4

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



## Warning: Removed 126 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   
## -1.042 -0.881 -0.768 1.420 2.034   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.37679 0.09306 -4.05 5.1e-05 \*\*\*  
## dis.cent -0.01291 0.00223 -5.80 6.7e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2764.5 on 2271 degrees of freedom  
## Residual deviance: 2729.5 on 2270 degrees of freedom  
## (35 observations deleted due to missingness)  
## AIC: 2733  
##   
## 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   
## -1.117 -0.890 -0.715 1.358 1.864   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.2106 0.0935 -2.25 0.024 \*   
## dis.cent.PER -1.3810 0.1795 -7.69 1.4e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2764.5 on 2271 degrees of freedom  
## Residual deviance: 2703.5 on 2270 degrees of freedom  
## (35 observations deleted due to missingness)  
## AIC: 2707  
##   
## 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.

### Heterochiasmy Prediction

Is sex a significant effect for the 1CO normalized CO position? (as predicted)

is random strtain effect significant?

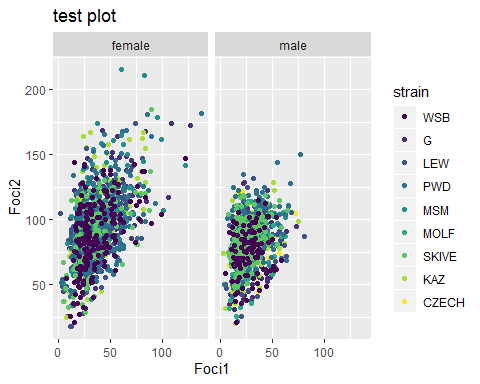
The random strain effect seems very significant.

remember to use the mouse average table **mouse.avs\_4MM** (I don’t think I need the MELT data frame)

#make point plots / boxplots which show differences in mean positions   
  
#scatter + boxplot for t.tests

Dr. Broman suggested that the Smirnov K /(curve comparison) wasn’t the best test to differences in general CO position. He suggested doing simple t-test for the positions

#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")  
F1.x.F2



#what is the pattern of variance  
#run analyses for each subsp\*sex  
#use non-melt DF  
  
#how is the variance partioned across  
#cell, mouse, strain  
  
female.Dom <- Curated\_BivData[Curated\_BivData$sex == "female",]  
female.Dom <- female.Dom[female.Dom$subsp == "Dom",]  
  
female.Dom$Foc1.PER <- female.Dom$Foci1 / female.Dom$chromosomeLength  
  
#unorder strain and mouse  
  
female.Dom$mouse <- as.factor(female.Dom$mouse)  
  
  
female.Dom$strain <- unclass(female.Dom$strain)  
female.Dom$strain <- as.factor(female.Dom$strain)  
  
female.Dom\_1CO <- female.Dom[female.Dom$hand.foci.count == 1,]  
female.Dom\_1CO <- female.Dom\_1CO[(!is.na(female.Dom\_1CO$hand.foci.count)),]  
  
#1CO first  
modo <- lm(Foc1.PER ~ fileName + mouse + strain, data=female.Dom\_1CO)  
  
#can't get mouse and strain to have sum of square  
#residual size decreases with per.F1  
#residuals much larger than fileName, mouse and strain no   
  
#model <- lm(breaks ~ wool \* tension,   
# data = warpbreaks,   
# contrasts = list(wool = "contr.sum", tension = "contr.poly"))  
  
male.Dom <- Curated\_BivData[Curated\_BivData$sex == "male",]  
male.Dom <- male.Dom[male.Dom$subsp == "Dom",]  
  
male.Dom$mouse <- as.factor(male.Dom$mouse)  
  
male.Dom$strain <- unclass(male.Dom$strain)  
male.Dom$strain <- as.factor(male.Dom$strain)  
  
male.Dom <- male.Dom[male.Dom$hand.foci.count == 1,]  
male.Dom <- male.Dom[(!is.na(male.Dom$hand.foci.count)),]  
  
male.Dom$Foc1.PER <- male.Dom$Foci1 / male.Dom$chromosomeLength  
  
male.modo <- lm(Foc1.PER ~ fileName | mouse | strain, data=male.Dom)  
summary(aov(male.modo))  
  
#only file name is registering as effect  
#Review ANOVA frameworks  
#http://www.biostathandbook.com/nestedanova.html

## Q2 Analysis Predictions, Male Polymorphism and (High vs Low Rec strains)

The general predictions across the males and subspecies based on th above MLH1 results.

For positive correlation traits/metrics

1. in DOM strains, low to no difference across strains
2. in Musc, PWD > SKIVE > KAZ, CZECH all the others
3. in Mol, MSM > MOLF
4. **SC lengths** will be longer for high Rec strains.

B.1) **Interfernce/IFD** will be shorter in high Rec strains. Use IFD\_PER to account for SC length differences.

C.1) not enough is known about variation within species for the **1CO normalized positions**. Null prediction, no difference in the ‘telomeric pattern’.

The mouse averages for the other position metrics will be highly influenced by proportions of the 1CO and 2CO bivalents. When class of chromosome and SC length is account for, there won’t be a difference, however, not enough is known about these patterns.

C.2) sis-co-ten metric … (what about the clustering?)

C.3) telomere and centromere distances …

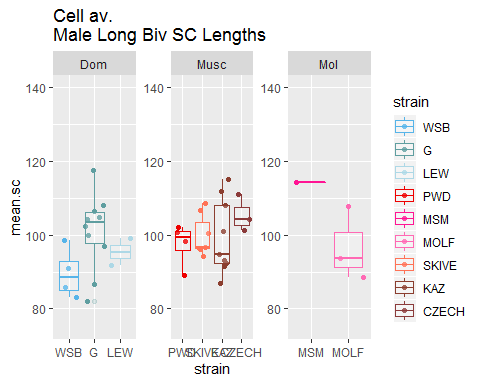
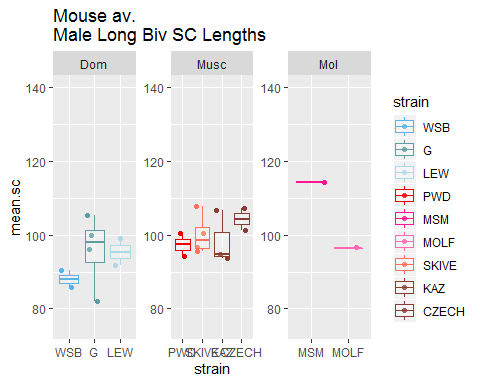
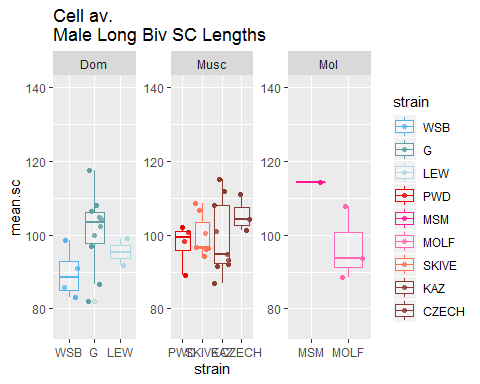
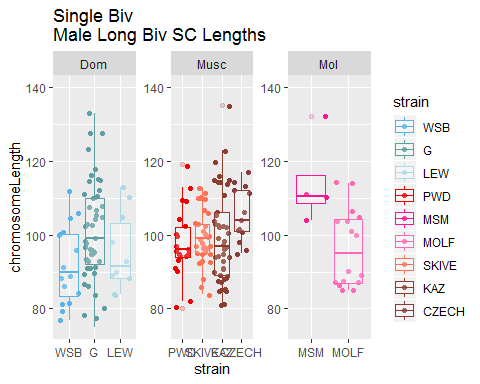
M1 glm model for fixed strain effect across male averages.

# Q2 SC

Focus on comparing the differences between long.biv (and short biv once I make that dataset)

## Warning: Removed 3 rows containing non-finite values (stat\_boxplot).

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



Not the pattern I expected, there might be more noise than I thought in this dataset.

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.

The general pattern is that the high rec strain have a greater mean SC length for all pooled bivalents (5046 males bivalent observations).

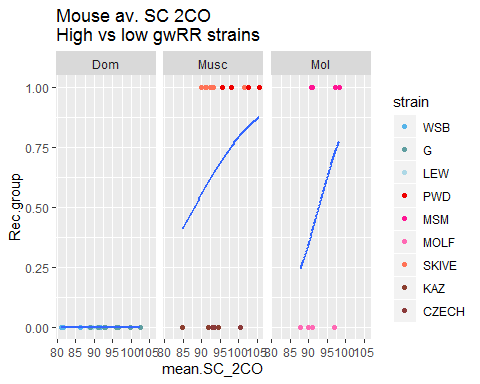
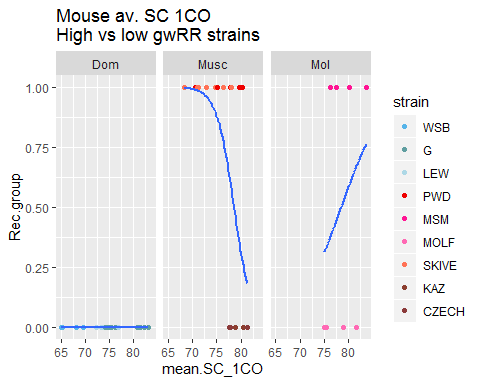
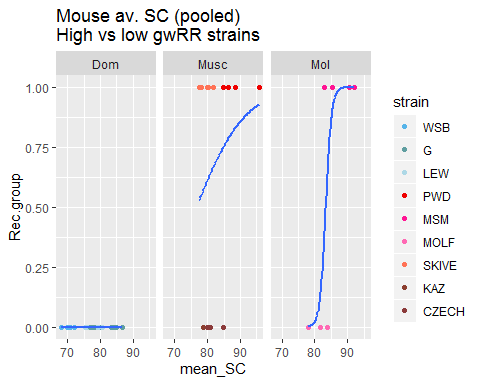
#### Q2 SC by class

The predicted pattern for SC length becomes more nuanced when the data are split up by Chromosome class. Mainly that 1CO are shorter in the high rec strains then low rec strains. This is likely due to the fact that more physically longer chromosomes have 1CO in the low rec strains, which pushes the mean SC length up. Where as in the High rec strains, physically longer chromosomes are more likely to be in the 2CO group. This supports a general pattern of tighter clustering of SC lengths across chromosome classes in the high rec group. They have a lower probability of 2COs below a certain SC length threshold.

I think these comparisons should be done with the **Bivalent observations** The logistic regression showed rec groups could be predicted by SC length - re-run them while separating out by chrm class

##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength \* strain, family = binomial(link = "logit"),   
## data = Bivalent\_1o2)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.324 -0.690 -0.470 0.579 2.684   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.26744 0.62115 -10.09 <2e-16 \*\*\*  
## chromosomeLength 0.06202 0.00752 8.25 <2e-16 \*\*\*  
## strainG 0.48653 0.77719 0.63 0.531   
## strainLEW 0.15970 0.88808 0.18 0.857   
## strainPWD 1.13825 0.77419 1.47 0.141   
## strainMSM 2.16716 1.10023 1.97 0.049 \*   
## strainMOLF 1.16195 0.90555 1.28 0.199   
## strainSKIVE 0.27723 0.78686 0.35 0.725   
## strainKAZ 1.33186 0.87191 1.53 0.127   
## strainCZECH 0.57808 1.10203 0.52 0.600   
## chromosomeLength:strainG -0.01208 0.00907 -1.33 0.183   
## chromosomeLength:strainLEW -0.00445 0.01042 -0.43 0.669   
## chromosomeLength:strainPWD -0.00537 0.00913 -0.59 0.556   
## chromosomeLength:strainMSM -0.01347 0.01269 -1.06 0.289   
## chromosomeLength:strainMOLF -0.01662 0.01046 -1.59 0.112   
## chromosomeLength:strainSKIVE 0.00200 0.00937 0.21 0.831   
## chromosomeLength:strainKAZ -0.02511 0.01008 -2.49 0.013 \*   
## chromosomeLength:strainCZECH -0.01172 0.01233 -0.95 0.342   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 5617.8 on 4909 degrees of freedom  
## Residual deviance: 4553.7 on 4892 degrees of freedom  
## AIC: 4590  
##   
## Number of Fisher Scoring iterations: 5

These are the plots of the logistic regression.



the plot of log regression for 1COs is wacky/reversed. Musc – the mean SC lengths for 1COs are longer in the low group

## [1] 0.0024

## [1] 1

## [1] 0.034

The top t.tests indicate that when all the mice averages are pooled, there’s a significant difference in SC lengths. But When the means are compared across chrm class, the mouse averages are no longer significant.

### Q2 SC length Chrm class prediction

How well does SC length predict chromosome class? Prediction, High rec strains will have more significant p value, given the lower overlap in SC lengths across chrm class. (Note this test can only be done with bivalent level observations, because 1CO or 2CO)

##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength, family = binomial(link = "logit"),   
## data = Bivalent\_1o2)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.970 -0.746 -0.507 0.760 2.549   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.5459 0.1874 -29.6 <2e-16 \*\*\*  
## chromosomeLength 0.0531 0.0021 25.3 <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: 5617.8 on 4909 degrees of freedom  
## Residual deviance: 4806.9 on 4908 degrees of freedom  
## AIC: 4811  
##   
## Number of Fisher Scoring iterations: 4

##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength, family = binomial(link = "logit"),   
## data = Bivalent\_1o2\_WSB)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.532 -0.619 -0.451 -0.254 2.579   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.26744 0.62115 -10.09 <2e-16 \*\*\*  
## chromosomeLength 0.06202 0.00752 8.25 <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: 554.21 on 578 degrees of freedom  
## Residual deviance: 469.39 on 577 degrees of freedom  
## AIC: 473.4  
##   
## Number of Fisher Scoring iterations: 5

##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength, family = binomial(link = "logit"),   
## data = Bivalent\_1o2\_G)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.558 -0.642 -0.467 -0.297 2.614   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.78091 0.46712 -12.38 <2e-16 \*\*\*  
## chromosomeLength 0.04995 0.00508 9.82 <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: 943.65 on 979 degrees of freedom  
## Residual deviance: 829.78 on 978 degrees of freedom  
## AIC: 833.8  
##   
## Number of Fisher Scoring iterations: 5

##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength, family = binomial(link = "logit"),   
## data = Bivalent\_1o2\_LEW)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.713 -0.651 -0.460 -0.271 2.684   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.10774 0.63471 -9.62 < 2e-16 \*\*\*  
## chromosomeLength 0.05758 0.00721 7.98 1.4e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 523.62 on 514 degrees of freedom  
## Residual deviance: 442.66 on 513 degrees of freedom  
## AIC: 446.7  
##   
## Number of Fisher Scoring iterations: 5

##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength, family = binomial(link = "logit"),   
## data = Bivalent\_1o2\_PWD)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.324 -0.873 -0.487 1.004 1.971   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.12919 0.46211 -11.1 <2e-16 \*\*\*  
## chromosomeLength 0.05665 0.00518 10.9 <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: 917.11 on 665 degrees of freedom  
## Residual deviance: 752.69 on 664 degrees of freedom  
## AIC: 756.7  
##   
## Number of Fisher Scoring iterations: 4

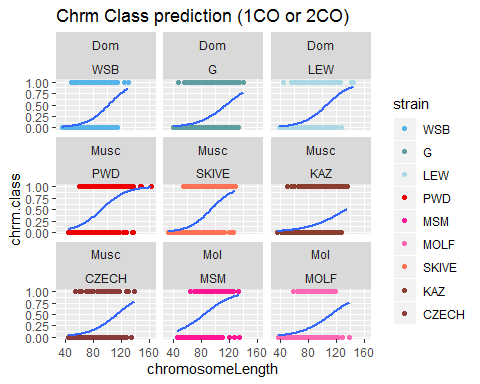
##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength, family = binomial(link = "logit"),   
## data = Bivalent\_1o2\_SKIVE)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.095 -0.815 -0.463 0.958 2.259   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.99021 0.48301 -12.4 <2e-16 \*\*\*  
## chromosomeLength 0.06402 0.00559 11.4 <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: 970.27 on 763 degrees of freedom  
## Residual deviance: 788.24 on 762 degrees of freedom  
## AIC: 792.2  
##   
## Number of Fisher Scoring iterations: 4

##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength, family = binomial(link = "logit"),   
## data = Bivalent\_1o2\_KAZ)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.074 -0.593 -0.466 -0.338 2.504   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.93558 0.61188 -8.07 7.3e-16 \*\*\*  
## chromosomeLength 0.03691 0.00671 5.50 3.8e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 481.68 on 585 degrees of freedom  
## Residual deviance: 448.88 on 584 degrees of freedom  
## AIC: 452.9  
##   
## Number of Fisher Scoring iterations: 5

##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength, family = binomial(link = "logit"),   
## data = Bivalent\_1o2\_CZECH)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.621 -0.660 -0.500 -0.301 2.459   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.68936 0.91018 -6.25 4.1e-10 \*\*\*  
## chromosomeLength 0.05030 0.00977 5.15 2.6e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 240.36 on 228 degrees of freedom  
## Residual deviance: 208.49 on 227 degrees of freedom  
## AIC: 212.5  
##   
## Number of Fisher Scoring iterations: 4

##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength, family = binomial(link = "logit"),   
## data = Bivalent\_1o2\_MSM)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.233 -1.048 0.585 1.031 1.617   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.1003 0.9081 -4.52 6.3e-06 \*\*\*  
## chromosomeLength 0.0486 0.0102 4.75 2.0e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 231.73 on 167 degrees of freedom  
## Residual deviance: 203.73 on 166 degrees of freedom  
## AIC: 207.7  
##   
## Number of Fisher Scoring iterations: 4

##   
## Call:  
## glm(formula = chrm.class ~ chromosomeLength, family = binomial(link = "logit"),   
## data = Bivalent\_1o2\_MOLF)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.693 -0.718 -0.533 -0.309 2.260   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.10549 0.65891 -7.75 9.3e-15 \*\*\*  
## chromosomeLength 0.04541 0.00727 6.25 4.2e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 455.53 on 422 degrees of freedom  
## Residual deviance: 409.79 on 421 degrees of freedom  
## AIC: 413.8  
##   
## Number of Fisher Scoring iterations: 4

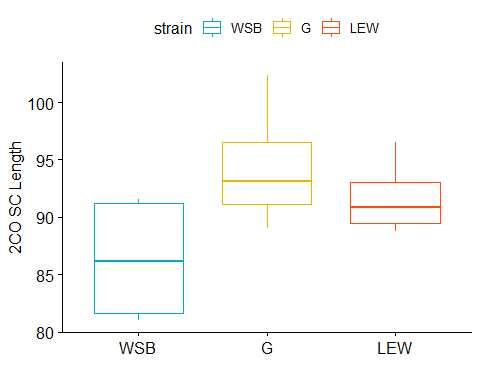
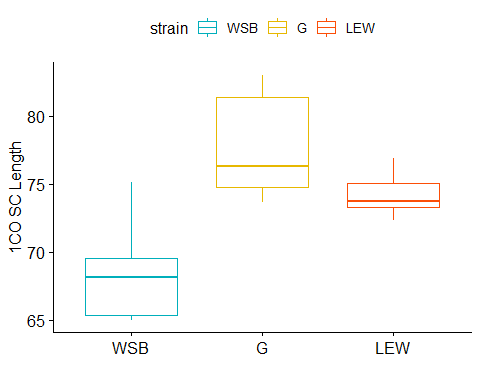
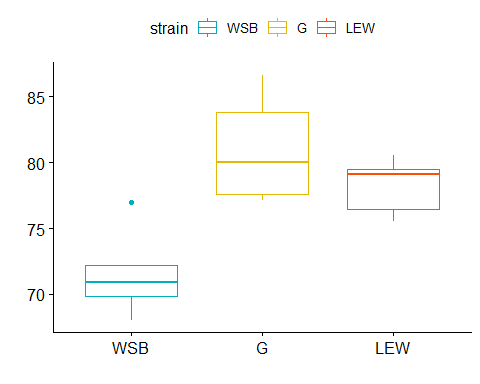
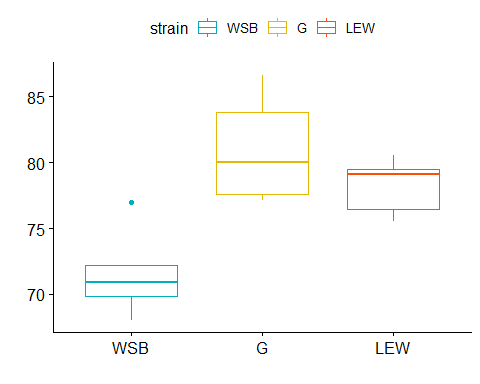


**Summary**

When the SC lengths are divided by co number / class, – difference in direction of the logistic regression. I think the shorter 1CO bivalents in the high rec strain reflects the clustering of groups by SC length. (IE the scatter plots of SC length, binned by chrm class on the X).

Each point is a mouse average SC. When all bivalents are pooled, the logistic regression SC lengths predict the groups. When the data is divided by CO number, 1CO’s have a higher SC length in the low group - in the low group more chrms, including the longer ones have 1CO - whereas in the high group, only shorter chrms (SC) are in the 1CO class. For the 2CO class of bivalents - all the SC are longer, the high group averages were longer than the low.

and the mean 2COs SC are higher in the high strains (this must have something to do with the physical size effect)



If I want to try to use mouse nested in strain – I should use cell level metrics (but those have there own flaws)

Ideas for the above 3 tests,

1. no difference in DOM

* no sig logistic regression test
* glm, strain effect?

Is sex a significant effect for SC length? (as predicted)

* Sex is a significant effect for SC length The results seem to indicate that sex is a significant factor. Consider writing a sub sampling approach (randomize / permute a data set of BivData)
* According to anova, sex effect explains most of the variance in single bivalent SC lengths.
* The Long Biv Data set largely agrees with the full curated dataset
* I caveat I haven’t addressed yet, is the XX in the female Biv data averages —

The mean SC logistic regression model for mouse averages

-10.2, 0.13

and for the single bivalent levels

0.21, 0

When all male mice are used, the predictive power is greater, than when just the Musc strains are used. When, just the Musc strain are used, The mouse mean SC is slightly significant in predicting if a mouse is in the high or low (should I consider running on female too?)

Is the prediction, high rec musc male strains have long SC met?

In a logistic regression, mouse average SC length is slightly predictive telling if a mouse is in a high or low Rec strain. I couldn’t get the Mixed models working for the male polymorphism predictions…

# Q2 Normalized CO positions

A main / biggish cavest to address for this section is the chromosome size effect, use real.long.bivData.

or have chromosome lenght as a effect? would only work for models using the single bivalents, not mouse averages.

Since there aren’t many good predictions for how the 1CO normalized landscape relates to gwRR variation, these will be the main questions

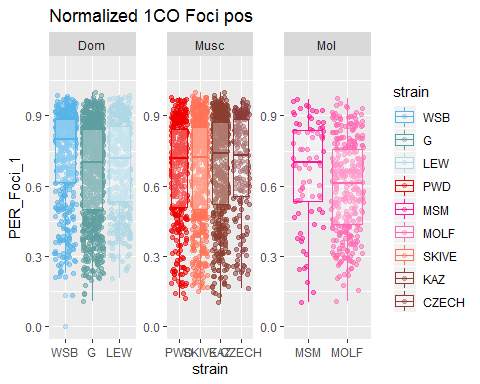
* Has there been evolution in the 1CO pattern?
* Is there a consistent pattern with the high rec strains?

Normalized 1CO foci positions will be used. (F1\_PER). Check fixed effects of subsp and strains.

sis-co-ten, telomere and centromere distance of foci are metrics which draw from a wider pool of samples. (So maybe I will use them… but avoiding focus on them)

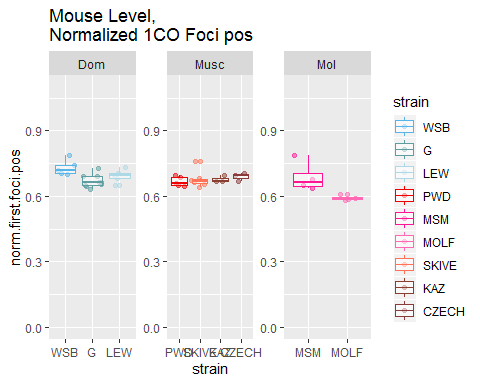
## Warning: Removed 1 rows containing non-finite values (stat\_boxplot).

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

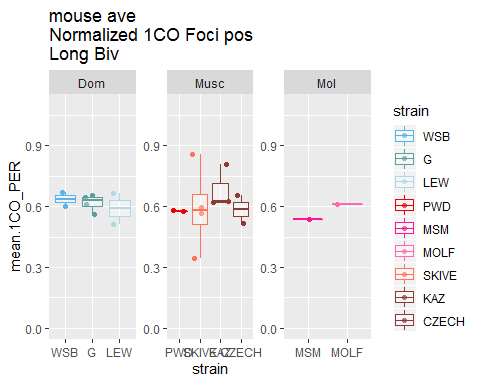
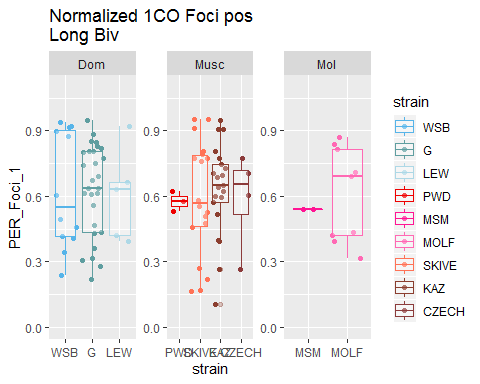


## Warning: Removed 1 rows containing non-finite values (stat\_boxplot).

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



Remade the Position table. and MSM has a more telomeric distribution. MOLF has the most central distribution.



So the plot for the long bivs, shows that the PWD, SKIVE and MSM have on average more centrally placed F1 positions. It’s kinda werid that the glms/models suggest that WSB has the most terminal F1 – but, out of all the strains, for the long bivs, WSB has the lowest mean.

But there are v few observations. I think the plot shows that there are fewer 1CO for the High Rec group. not sure there would be enough for feeding into models/

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

## [1] 0.91

## [1] 0.24

The normalized Foci Positions are not significantly different between the high and low rec groups (full data set, mouse av and single bivalent level).

#### Summary

Not really a clear summary for the normalized 1CO position and rec groups.

# Q2 IFD

this is the Q2 prediction for IFD / interference

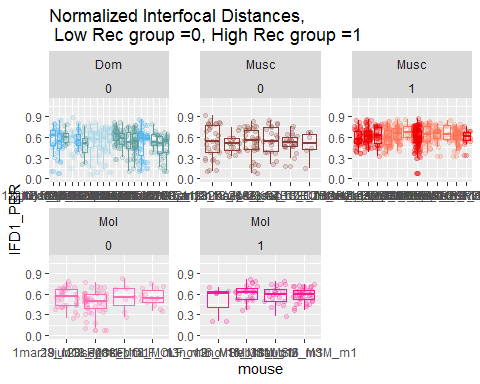
prediction A, Is the Male polymorphism Prediction met? High rec strains have shorter IFDs?

B.1) **Interfernce/IFD** will be shorter in high Rec strains. Use IFD\_PER to account for SC length differences.

1. glm across subsp
2. anova within subsp
3. logistic regression for musc (or all males)

## Warning: Removed 4 rows containing non-finite values (stat\_boxplot).

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



The above plots are not the best at showing the pattern of higher rec group having slightly higher raw IFD - because their SC are longer on average. The normalized plot displays the pattern of higher rec group having slightly higher longer IFDs.

##   
## Call:  
## glm(formula = mean\_IFD.2CO\_ABS ~ subsp \* strain, data = Male.poly.Mouse.Table\_BivData\_4MM)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -7.761 -3.329 -0.345 2.788 12.633   
##   
## Coefficients: (18 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 47.94 2.21 21.75 <2e-16 \*\*\*  
## subspMusc 3.55 3.60 0.99 0.331   
## subspMol 0.10 3.31 0.03 0.976   
## strainG 3.82 2.75 1.39 0.174   
## strainLEW 3.91 3.12 1.25 0.218   
## strainPWD 8.51 3.60 2.36 0.024 \*   
## strainMSM 6.40 3.49 1.84 0.074 .   
## strainMOLF NA NA NA NA   
## strainSKIVE 8.33 3.40 2.45 0.019 \*   
## strainKAZ -3.47 4.03 -0.86 0.395   
## strainCZECH NA NA NA NA   
## subspMusc:strainG NA NA NA NA   
## subspMol:strainG NA NA NA NA   
## subspMusc:strainLEW NA NA NA NA   
## subspMol:strainLEW NA NA NA NA   
## subspMusc:strainPWD NA NA NA NA   
## subspMol:strainPWD NA NA NA NA   
## subspMusc:strainMSM NA NA NA NA   
## subspMol:strainMSM NA NA NA NA   
## subspMusc:strainMOLF NA NA NA NA   
## subspMol:strainMOLF NA NA NA NA   
## subspMusc:strainSKIVE NA NA NA NA   
## subspMol:strainSKIVE NA NA NA NA   
## subspMusc:strainKAZ NA NA NA NA   
## subspMol:strainKAZ NA NA NA NA   
## subspMusc:strainCZECH NA NA NA NA   
## subspMol:strainCZECH NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 24)  
##   
## Null deviance: 1781.5 on 44 degrees of freedom  
## Residual deviance: 875.0 on 36 degrees of freedom  
## AIC: 281.2  
##   
## Number of Fisher Scoring iterations: 2

##   
## Call:  
## glm(formula = mean\_IFD.2CO\_PER ~ subsp \* strain, data = Male.poly.Mouse.Table\_BivData\_4MM)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.06962 -0.01833 -0.00294 0.01864 0.08498   
##   
## Coefficients: (18 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.54260 0.01641 33.06 < 2e-16 \*\*\*  
## subspMusc -0.01708 0.02680 -0.64 0.5280   
## subspMol -0.01303 0.02462 -0.53 0.5999   
## strainG -0.00255 0.02047 -0.12 0.9016   
## strainLEW 0.01203 0.02321 0.52 0.6075   
## strainPWD 0.07965 0.02680 2.97 0.0053 \*\*   
## strainMSM 0.04505 0.02595 1.74 0.0911 .   
## strainMOLF NA NA NA NA   
## strainSKIVE 0.11575 0.02533 4.57 5.5e-05 \*\*\*  
## strainKAZ -0.00542 0.02997 -0.18 0.8576   
## strainCZECH NA NA NA NA   
## subspMusc:strainG NA NA NA NA   
## subspMol:strainG NA NA NA NA   
## subspMusc:strainLEW NA NA NA NA   
## subspMol:strainLEW NA NA NA NA   
## subspMusc:strainPWD NA NA NA NA   
## subspMol:strainPWD NA NA NA NA   
## subspMusc:strainMSM NA NA NA NA   
## subspMol:strainMSM NA NA NA NA   
## subspMusc:strainMOLF NA NA NA NA   
## subspMol:strainMOLF NA NA NA NA   
## subspMusc:strainSKIVE NA NA NA NA   
## subspMol:strainSKIVE NA NA NA NA   
## subspMusc:strainKAZ NA NA NA NA   
## subspMol:strainKAZ NA NA NA NA   
## subspMusc:strainCZECH NA NA NA NA   
## subspMol:strainCZECH NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.0013)  
##   
## Null deviance: 0.122045 on 44 degrees of freedom  
## Residual deviance: 0.048495 on 36 degrees of freedom  
## AIC: -159.8  
##   
## Number of Fisher Scoring iterations: 2

##   
## Call:  
## glm(formula = mean\_IFD.2CO\_ABS ~ strain, data = Male.poly.Mouse.Table\_BivData\_4MM)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -7.761 -3.329 -0.345 2.788 12.633   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 47.9445 2.2048 21.75 < 2e-16 \*\*\*  
## strainG 3.8170 2.7499 1.39 0.17365   
## strainLEW 3.9088 3.1180 1.25 0.21806   
## strainPWD 12.0610 3.1180 3.87 0.00044 \*\*\*  
## strainMSM 6.5047 3.3072 1.97 0.05694 .   
## strainMOLF 0.1004 3.3072 0.03 0.97595   
## strainSKIVE 11.8852 2.8867 4.12 0.00021 \*\*\*  
## strainKAZ 0.0818 3.6004 0.02 0.98200   
## strainCZECH 3.5507 3.6004 0.99 0.33062   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 24)  
##   
## Null deviance: 1781.5 on 44 degrees of freedom  
## Residual deviance: 875.0 on 36 degrees of freedom  
## AIC: 281.2  
##   
## Number of Fisher Scoring iterations: 2

##   
## Call:  
## glm(formula = mean\_IFD.2CO\_ABS ~ strain, data = Male.poly.Mouse.Table\_BivData\_4MM[Male.poly.Mouse.Table\_BivData\_4MM$subsp ==   
## "Dom", ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -7.76 -3.37 -1.20 4.78 12.63   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 47.94 2.60 18.47 3.2e-12 \*\*\*  
## strainG 3.82 3.24 1.18 0.26   
## strainLEW 3.91 3.67 1.06 0.30   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 34)  
##   
## Null deviance: 593.53 on 18 degrees of freedom  
## Residual deviance: 538.90 on 16 degrees of freedom  
## AIC: 125.5  
##   
## Number of Fisher Scoring iterations: 2

##   
## Call:  
## glm(formula = mean\_IFD.2CO\_ABS ~ strain, data = Male.poly.Mouse.Table\_BivData\_4MM[Male.poly.Mouse.Table\_BivData\_4MM$subsp ==   
## "Musc", ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -5.793 -3.225 0.406 2.289 8.742   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 60.005 1.878 31.95 1.8e-14 \*\*\*  
## strainSKIVE -0.176 2.459 -0.07 0.9440   
## strainKAZ -11.979 3.067 -3.91 0.0016 \*\*   
## strainCZECH -8.510 3.067 -2.77 0.0149 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 18)  
##   
## Null deviance: 676.54 on 17 degrees of freedom  
## Residual deviance: 246.95 on 14 degrees of freedom  
## AIC: 108.2  
##   
## Number of Fisher Scoring iterations: 2

##   
## Call:  
## glm(formula = mean\_IFD.2CO\_ABS ~ strain, data = Male.poly.Mouse.Table\_BivData\_4MM[Male.poly.Mouse.Table\_BivData\_4MM$subsp ==   
## "Mol", ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -7.74 -0.25 1.03 2.28 2.84   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 54.45 1.93 28.25 1.3e-07 \*\*\*  
## strainMOLF -6.40 2.73 -2.35 0.057 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 15)  
##   
## Null deviance: 171.183 on 7 degrees of freedom  
## Residual deviance: 89.151 on 6 degrees of freedom  
## AIC: 47.99  
##   
## Number of Fisher Scoring iterations: 2

##   
## Call:  
## glm(formula = mean\_IFD.2CO\_PER ~ strain, data = Male.poly.Mouse.Table\_BivData\_4MM)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.06962 -0.01833 -0.00294 0.01864 0.08498   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.54260 0.01641 33.06 < 2e-16 \*\*\*  
## strainG -0.00255 0.02047 -0.12 0.902   
## strainLEW 0.01203 0.02321 0.52 0.607   
## strainPWD 0.06257 0.02321 2.70 0.011 \*   
## strainMSM 0.03202 0.02462 1.30 0.202   
## strainMOLF -0.01303 0.02462 -0.53 0.600   
## strainSKIVE 0.09867 0.02149 4.59 5.2e-05 \*\*\*  
## strainKAZ -0.02250 0.02680 -0.84 0.407   
## strainCZECH -0.01708 0.02680 -0.64 0.528   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.0013)  
##   
## Null deviance: 0.122045 on 44 degrees of freedom  
## Residual deviance: 0.048495 on 36 degrees of freedom  
## AIC: -159.8  
##   
## Number of Fisher Scoring iterations: 2

##   
## Call:  
## glm(formula = mean\_IFD.2CO\_PER ~ strain, data = Male.poly.Mouse.Table\_BivData\_4MM[Male.poly.Mouse.Table\_BivData\_4MM$subsp ==   
## "Dom", ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.06962 -0.02219 -0.00414 0.02426 0.08498   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.54260 0.01901 28.54 3.8e-15 \*\*\*  
## strainG -0.00255 0.02371 -0.11 0.92   
## strainLEW 0.01203 0.02689 0.45 0.66   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.0018)  
##   
## Null deviance: 0.029624 on 18 degrees of freedom  
## Residual deviance: 0.028915 on 16 degrees of freedom  
## AIC: -61.35  
##   
## Number of Fisher Scoring iterations: 2

##   
## Call:  
## glm(formula = mean\_IFD.2CO\_PER ~ strain, data = Male.poly.Mouse.Table\_BivData\_4MM[Male.poly.Mouse.Table\_BivData\_4MM$subsp ==   
## "Musc", ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.03875 -0.01485 -0.00084 0.01429 0.04264   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.6052 0.0115 52.40 < 2e-16 \*\*\*  
## strainSKIVE 0.0361 0.0151 2.39 0.03162 \*   
## strainKAZ -0.0851 0.0189 -4.51 0.00049 \*\*\*  
## strainCZECH -0.0796 0.0189 -4.22 0.00085 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.00067)  
##   
## Null deviance: 0.0559623 on 17 degrees of freedom  
## Residual deviance: 0.0093366 on 14 degrees of freedom  
## AIC: -75.07  
##   
## Number of Fisher Scoring iterations: 2

##   
## Call:  
## glm(formula = mean\_IFD.2CO\_PER ~ strain, data = Male.poly.Mouse.Table\_BivData\_4MM[Male.poly.Mouse.Table\_BivData\_4MM$subsp ==   
## "Mol", ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.06213 -0.00841 0.01417 0.02132 0.04058   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.5746 0.0207 27.81 1.4e-07 \*\*\*  
## strainMOLF -0.0451 0.0292 -1.54 0.17   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.0017)  
##   
## Null deviance: 0.014303 on 7 degrees of freedom  
## Residual deviance: 0.010244 on 6 degrees of freedom  
## AIC: -24.58  
##   
## Number of Fisher Scoring iterations: 2

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

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

##   
## Call:  
## glm(formula = Rec.group ~ mean\_IFD.2CO\_PER, family = binomial(link = "logit"),   
## data = Male.poly.Mouse.Table\_BivData\_4MM)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.015 -0.487 -0.199 0.394 2.803   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -30.30 8.78 -3.45 0.00056 \*\*\*  
## mean\_IFD.2CO\_PER 51.51 15.03 3.43 0.00061 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 58.574 on 44 degrees of freedom  
## Residual deviance: 30.626 on 43 degrees of freedom  
## AIC: 34.63  
##   
## Number of Fisher Scoring iterations: 6

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

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

##   
## Call:  
## glm(formula = Rec.group ~ mean\_IFD.2CO\_PER, 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   
## -2.42e-05 -2.10e-08 2.10e-08 2.10e-08 2.72e-05   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -1180 1283767 0 1  
## mean\_IFD.2CO\_PER 2038 2212381 0 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2.2915e+01 on 17 degrees of freedom  
## Residual deviance: 1.3284e-09 on 16 degrees of freedom  
## AIC: 4  
##   
## Number of Fisher Scoring iterations: 25

##   
## Call:  
## glm(formula = Rec.group ~ mean\_IFD.2CO\_PER, family = binomial(link = "logit"),   
## data = Male.poly.Mouse.Table\_BivData\_4MM[Male.poly.Mouse.Table\_BivData\_4MM$subsp ==   
## "Mol", ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.1390 -1.0618 0.0705 0.7469 1.8185   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -19.2 15.1 -1.27 0.2  
## mean\_IFD.2CO\_PER 34.7 27.1 1.28 0.2  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 11.0904 on 7 degrees of freedom  
## Residual deviance: 8.4622 on 6 degrees of freedom  
## AIC: 12.46  
##   
## Number of Fisher Scoring iterations: 4

##   
## Call:  
## glm(formula = Rec.group ~ IFD1\_PER, family = binomial(link = "logit"),   
## data = male.bivdata.2CO\_IFD)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.705 -1.159 0.834 1.115 1.916   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.894 0.244 -7.75 8.9e-15 \*\*\*  
## IFD1\_PER 3.339 0.410 8.15 3.7e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1757.6 on 1267 degrees of freedom  
## Residual deviance: 1683.9 on 1266 degrees of freedom  
## (4 observations deleted due to missingness)  
## AIC: 1688  
##   
## Number of Fisher Scoring iterations: 4

##   
## Call:  
## glm(formula = Rec.group ~ IFD1\_PER, family = binomial(link = "logit"),   
## data = male.bivdata.2CO\_IFD[male.bivdata.2CO\_IFD$subsp ==   
## "Musc", ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.405 0.448 0.562 0.679 1.454   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.913 0.391 -2.33 0.02 \*   
## IFD1\_PER 4.059 0.684 5.94 2.9e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 677.59 on 685 degrees of freedom  
## Residual deviance: 640.56 on 684 degrees of freedom  
## (2 observations deleted due to missingness)  
## AIC: 644.6  
##   
## Number of Fisher Scoring iterations: 4

##   
## Call:  
## glm(formula = Rec.group ~ IFD1\_PER, family = binomial(link = "logit"),   
## data = male.bivdata.2CO\_IFD[male.bivdata.2CO\_IFD$subsp ==   
## "Mol", ])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.611 -1.118 -0.558 1.099 1.794   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.154 0.651 -3.31 0.00094 \*\*\*  
## IFD1\_PER 3.799 1.136 3.35 0.00082 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 259.1 on 186 degrees of freedom  
## Residual deviance: 246.5 on 185 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 250.5  
##   
## Number of Fisher Scoring iterations: 4

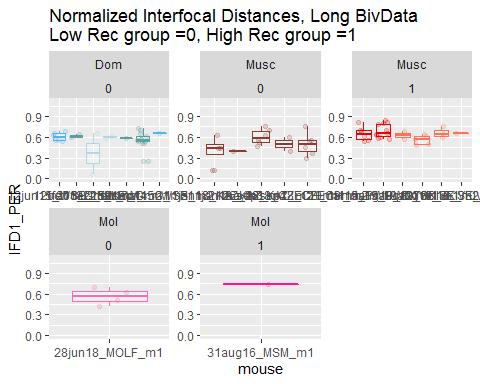
## [1] 2e-17

## [1] 6.9e-07

The t.tests return very significant p values for the normalized IFD between the high and low rec groups (for both mouse av and single bivalent observations levels).

### Q2 Long Bivalent Dataset

Examine the pattern in the Long bivalent data set to get a feel for if the chromosome size effect skews the pattern. In the long bivalent data set, the longest bivalents (top 25%) from each cell are isolated, so they are more likely to be the same chromosome identities (Chrm1 to Chrm5). The section on IFD only looks at the 2CO bivalents.



table(droplevels(male.long.biv.2CO\_IFD$strain) )

The plot above shows the normalized IFD measures from 2CO bivalents

There are many fewer long bivalent data observations, but the same positive correlation with

-table of long bivalents

* glms
* logistic regression

There are not enough observations for sub setting the data into strains or use the mouse averages.

Neither t.test are significant for both the ABS and PER when I test just the Musc strains. The above t.tests are breaking the knitr

The t.tests for IFD1s at the bivalent level for the high and low musc males are significant.

None of the logistic regression models for ABS or PER IFD lengths are significant, even when just the Musc strains are used.

Preliminary results from an independent data set indicated that PWD had longer IFDs, which goes against the simple prediction of more COs ~ denser spacing of foci on the same bivalent. This also indicated that interference distance may evolved in the house mouse complex.

# Caveats

put all of the code chunks/analysis for caveats here

## Chromosome Size Effect

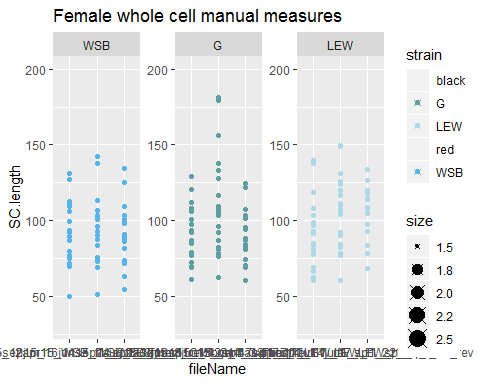
I tried to isolate bivalents which are in the top quartille for SC length from their cells. (re think where this section should go)

Below are examples of plots of SC length distributions across cells. The top figure shows whole cell hand measured data and the bottom shows the Automated bivData from cells with at least 15 bivalents measured. Most plots excluded for space.

Each point is a bivalent plotted by cell on the x axis. X’s are the 4th quartile, big point is the mean and smaller black point is the median. I’m using these to compare the patterns of these statistics in the automated data set which is missing some bivalent data. (the extra stats are not correctly mapped)

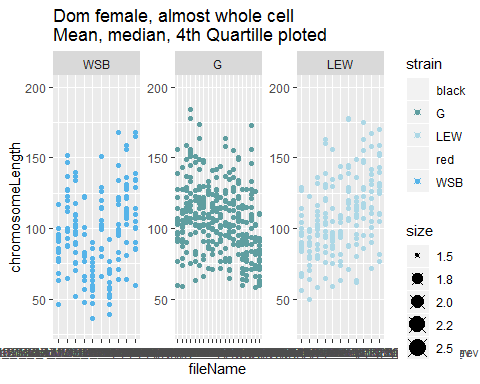
## Warning: Removed 3 rows containing missing values (geom\_point).

## Warning: Removed 9 rows containing missing values (geom\_point).  
  
## Warning: Removed 9 rows containing missing values (geom\_point).  
  
## Warning: Removed 9 rows containing missing values (geom\_point).



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

## Warning: Removed 41 rows containing missing values (geom\_point).  
  
## Warning: Removed 41 rows containing missing values (geom\_point).  
  
## Warning: Removed 41 rows containing missing values (geom\_point).



All the plots above show the distributions of manually whole cell measured SC lengths compared to the SC length distributions from the automated bivalent data. It shows the amount of within cell variance across strains. There is a bit of variance across the SC length distributions in the PWD females.

This data set might be noisy, given the amount of variance in the SC length distributions across cells (PWD females, WSB females).

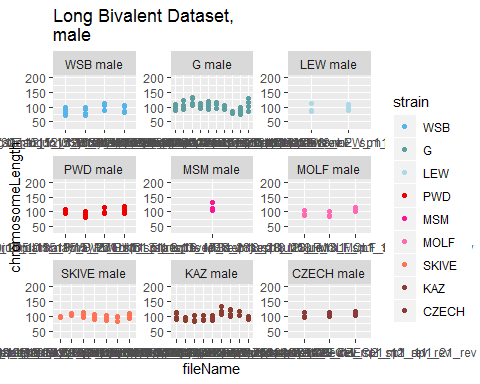
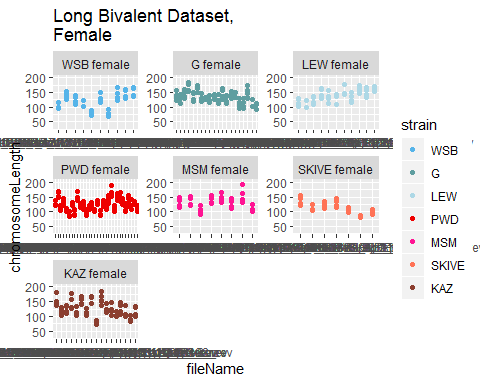
The DF **real.long.bivData** contains 713 bivalent measures. The full data set is 9807. This is the breakdown of bivalent observations by category for this long dataset are

**- Try to merge this DF with the whole.cell manual measures.**

* Try estimating which ‘loose’ bivalent observations might be within the long class of bivalents.

in code chunk above I ran the mouse averages for the longest bivalents. (680 bivalents, from 54 mice. 10202 bivalents from 86 mice.

## Warning: Removed 3 rows containing missing values (geom\_point).



These plots show the SC lengths for the ‘long SC data set’. They are supposed to be the longest 4-5 SC from cells where I could get good measures. These longer bivalents are useful because their patterns shouldn’t be affect by chromosome size effect (which effects, CO position). Hopefully this data set will have less noise from chromosome identity, but there was still data missing (they don’t come from whole cell measures).

## Adjusting for XX

(rethink where this section should go) new outline

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. 19 female, 19male, 20female 20 male

The female mouse averages should have adjustments for the XX. working on code to estimate the SC length from 3rd largest bivalent from female whole cell data across strains. Subtract this amount from the female mouse averages … This isn’t the best solution – since I can’t determine what proportion of cells for female mouse averages include the XX, (most cells are missing at least 3 bivalents)

* 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.

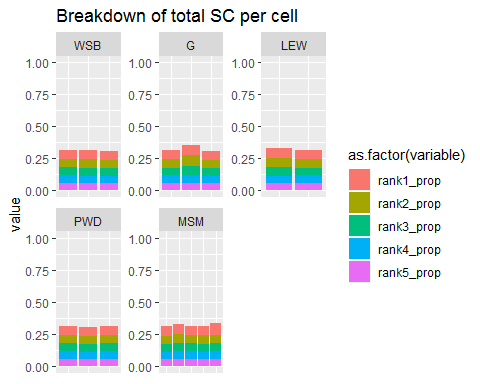
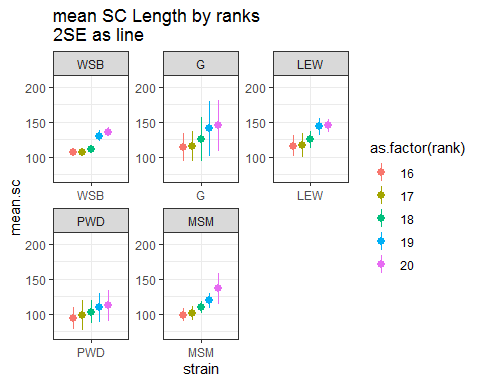
I think a formula something like this can be applied to adjust for XX

* 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 /\*

#### Whole Cell Manual Comparison

## Var1 Freq  
## 1 WSB female 60  
## 2 G female 60  
## 3 LEW female 60  
## 4 PWD female 60  
## 5 MSM female 119

## Var1 Var2 Freq  
## 1 WSB female 1 3  
## 2 G female 1 3  
## 3 LEW female 1 3  
## 4 PWD female 1 3  
## 5 MSM female 1 6  
## 6 WSB female 2 3  
## 7 G female 2 3  
## 8 LEW female 2 3  
## 9 PWD female 2 3  
## 10 MSM female 2 6  
## 11 WSB female 3 3  
## 12 G female 3 3  
## 13 LEW female 3 3  
## 14 PWD female 3 3  
## 15 MSM female 3 6  
## 16 WSB female 4 3  
## 17 G female 4 3  
## 18 LEW female 4 3  
## 19 PWD female 4 3  
## 20 MSM female 4 6  
## 21 WSB female 5 3  
## 22 G female 5 3  
## 23 LEW female 5 3  
## 24 PWD female 5 3  
## 25 MSM female 5 6  
## 26 WSB female 6 3  
## 27 G female 6 3  
## 28 LEW female 6 3  
## 29 PWD female 6 3  
## 30 MSM female 6 6  
## 31 WSB female 7 3  
## 32 G female 7 3  
## 33 LEW female 7 3  
## 34 PWD female 7 3  
## 35 MSM female 7 6  
## 36 WSB female 8 3  
## 37 G female 8 3  
## 38 LEW female 8 3  
## 39 PWD female 8 3  
## 40 MSM female 8 6  
## 41 WSB female 9 3  
## 42 G female 9 3  
## 43 LEW female 9 3  
## 44 PWD female 9 3  
## 45 MSM female 9 6  
## 46 WSB female 10 3  
## 47 G female 10 3  
## 48 LEW female 10 3  
## 49 PWD female 10 3  
## 50 MSM female 10 6  
## 51 WSB female 11 3  
## 52 G female 11 3  
## 53 LEW female 11 3  
## 54 PWD female 11 3  
## 55 MSM female 11 6  
## 56 WSB female 12 3  
## 57 G female 12 3  
## 58 LEW female 12 3  
## 59 PWD female 12 3  
## 60 MSM female 12 6  
## 61 WSB female 13 3  
## 62 G female 13 3  
## 63 LEW female 13 3  
## 64 PWD female 13 3  
## 65 MSM female 13 6  
## 66 WSB female 14 3  
## 67 G female 14 3  
## 68 LEW female 14 3  
## 69 PWD female 14 3  
## 70 MSM female 14 6  
## 71 WSB female 15 3  
## 72 G female 15 3  
## 73 LEW female 15 3  
## 74 PWD female 15 3  
## 75 MSM female 15 6  
## 76 WSB female 16 3  
## 77 G female 16 3  
## 78 LEW female 16 3  
## 79 PWD female 16 3  
## 80 MSM female 16 6  
## 81 WSB female 17 3  
## 82 G female 17 3  
## 83 LEW female 17 3  
## 84 PWD female 17 3  
## 85 MSM female 17 6  
## 86 WSB female 18 3  
## 87 G female 18 3  
## 88 LEW female 18 3  
## 89 PWD female 18 3  
## 90 MSM female 18 6  
## 91 WSB female 19 3  
## 92 G female 19 3  
## 93 LEW female 19 3  
## 94 PWD female 19 3  
## 95 MSM female 19 6  
## 96 WSB female 20 3  
## 97 G female 20 3  
## 98 LEW female 20 3  
## 99 PWD female 20 3  
## 100 MSM female 20 5



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).

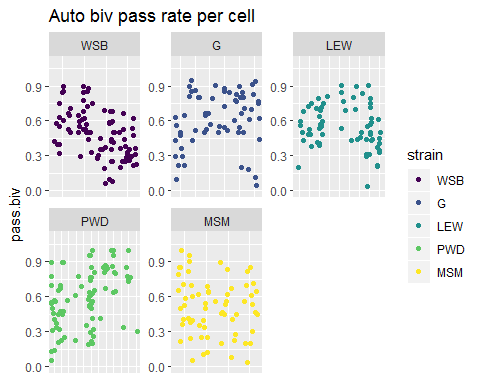
(use the value for the 3rd bivalent to adjust the single bivalent traits for XX – then compare to males values – or re-run in the MM).

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.

#### Automated BivData Comparison

##   
## WSB G LEW PERC PWD MSM MOLF SKIVE KAZ TOM   
## 767 726 714 0 1031 550 0 0 0 0   
## AST CZECH CAST HMI SPRET SPIC CAROLI F1 other   
## 0 0 0 0 0 0 0 0 0



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.

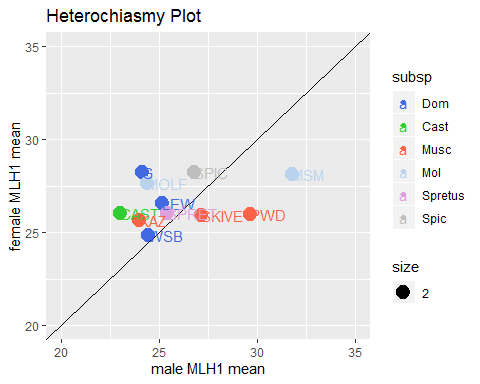
# Note on Heterochiasmy Definition

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.

# Deleted notes

# References