Inversions heterozygosis-related increase in chromosomal aberrations

Ruth Gómez Graciani

In this report, I analyze the realtionship between the amount of chromosomes in heterozygosis and amount of aneuploidies, which is key to distinguish whether the observed reduction in crossover rate in heterozygotes is due to a physical impediment of recombination, or due to a generation of aberrations and later discard from the crossover dataset.

1 Introduction

There are two possible mechanisms that can lead to recombination reduction between opposite orientations of an inversion:

- A physical impediment to pair or recombine.
- Purifying selection against the unbalanced results of a crossover between orientations.

In Bell et al. (2020), they sequence 900 to 2000 gametes from 20 donors, and then the sequenced chromosomes have 3 possible fates:

- Halthy chromosome: goes to the crossover dataset and it is used to elaborate single-individual crossover maps.
- Chromosome with small and medium aberrations: it is discarded.
- Whole chromosome and chromosome arm gains and losses: they are counted in the aberration dataset.

We observe a reduction in crossover rate in heterozygotes compared to homozygotes, but due to the methodology in Bell et al. 2020, this is not enough to confirm whether the inhibition mechanism is physical inhibition or purifying selection. However, we can complement that observation with the measurement of alterations in the aberration count caused by inversions in heterozygosis. If more heterozygous inversions leads to more aberrations in that chromosome, that would be consistent with a purifying selection scenario.

2 Variables to test

As per the independent variable, I initially measured, on each chromosome and individual, how many centi-Morgans were affected by inversions in heterozygosis as a proxy of the probability of having an aberration-generating crossover. In addition, I selected some inversion subsets that may have a larger impact in the aberration count:

- Inversions near the centromere will form a very large acentric spanning almost all the chromosome arm, so we expect them to influence the number of chromosome arm losses. I counted as centromeric regions the 20% of each arm next to the centromere, since any inversion in that region would cause an acentric of at least ~80% of the arm.
- Inversions near the telomere will form a very small acentric fragment, but if the dicentric fragment breaks near the centromere, the results of the crossover will be a chromosome arm gain and a chromosome arm loss. In addition, these regions concentrate most of the crossovers in the cromosome.

I counted as telomeric regions the same regions we selected based on recombination density in the population-based recombination map analysis.

- Big inversions represent a highest chance of crossover just by probability, and we observed a consistent reduction in crossovers in heterozygotes compared to homozygotes. All those discarded crossovers could be contributing to the aberration count. I considered as big inversions those that span more than 1 window in the single-individual crossover maps (>200 kb).

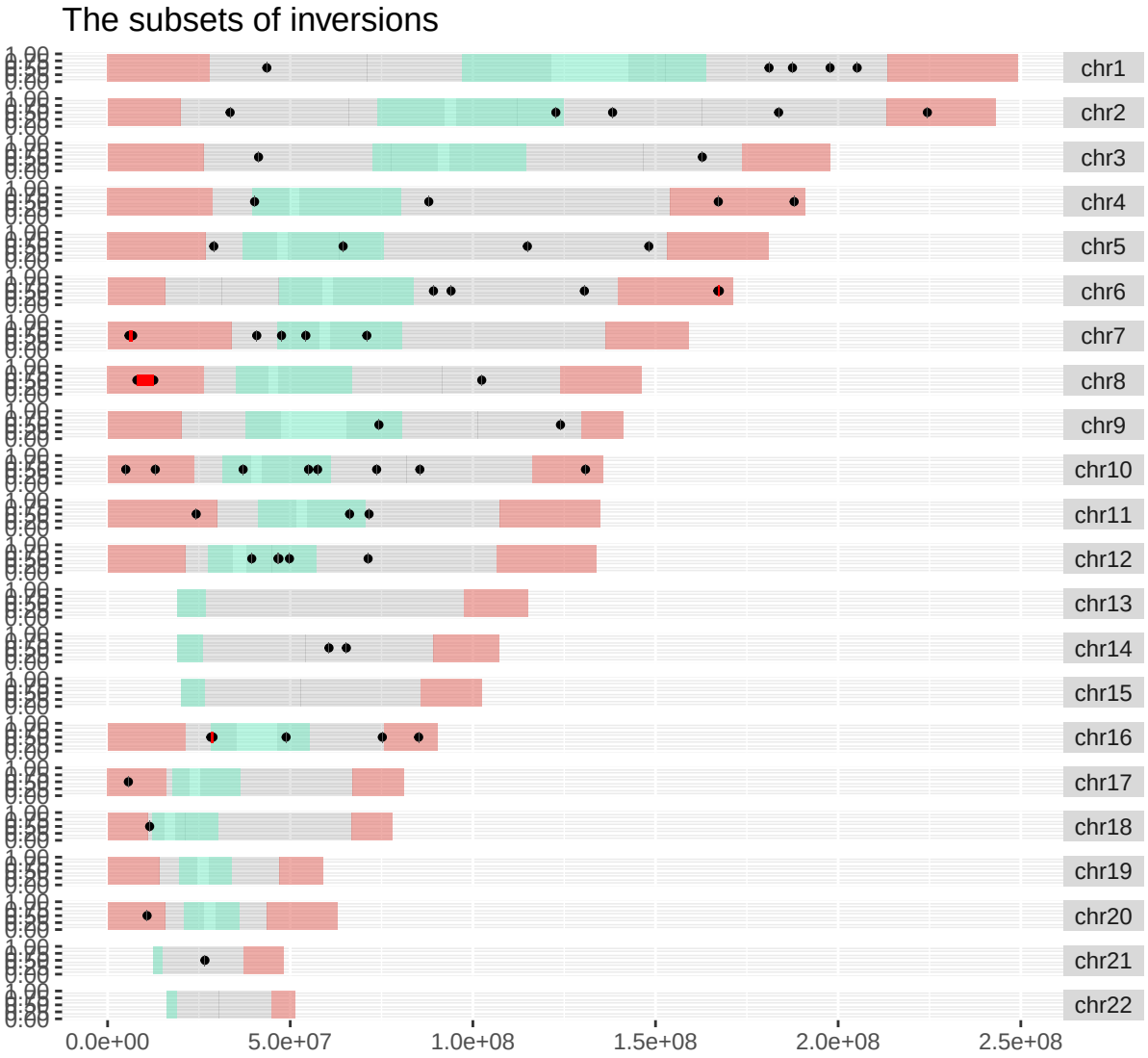


Figure 1: The selection of inversion subsets. The map windows that contain the inversion are marked in as dark gray lines (or red for big inversions), and also with points for easier visualization. Inversions within the green region are considered to be near the centromere. Inversions within the blue region are considered to be near the telomere.

As a dependent variable, I used the number of arm gains and arm losses, which are the types of aberrations expected to be influenced by the independent variable.

In our fist attempts in this analysis I made some correlations, however, as it will be shown later, a big portion of cases had 0 heterozygous inversions and/or 0 aberrations detected, so I opted for converting them into binomial variables and comparing absence and presence of each.

3 Results

3.1 Data visualization

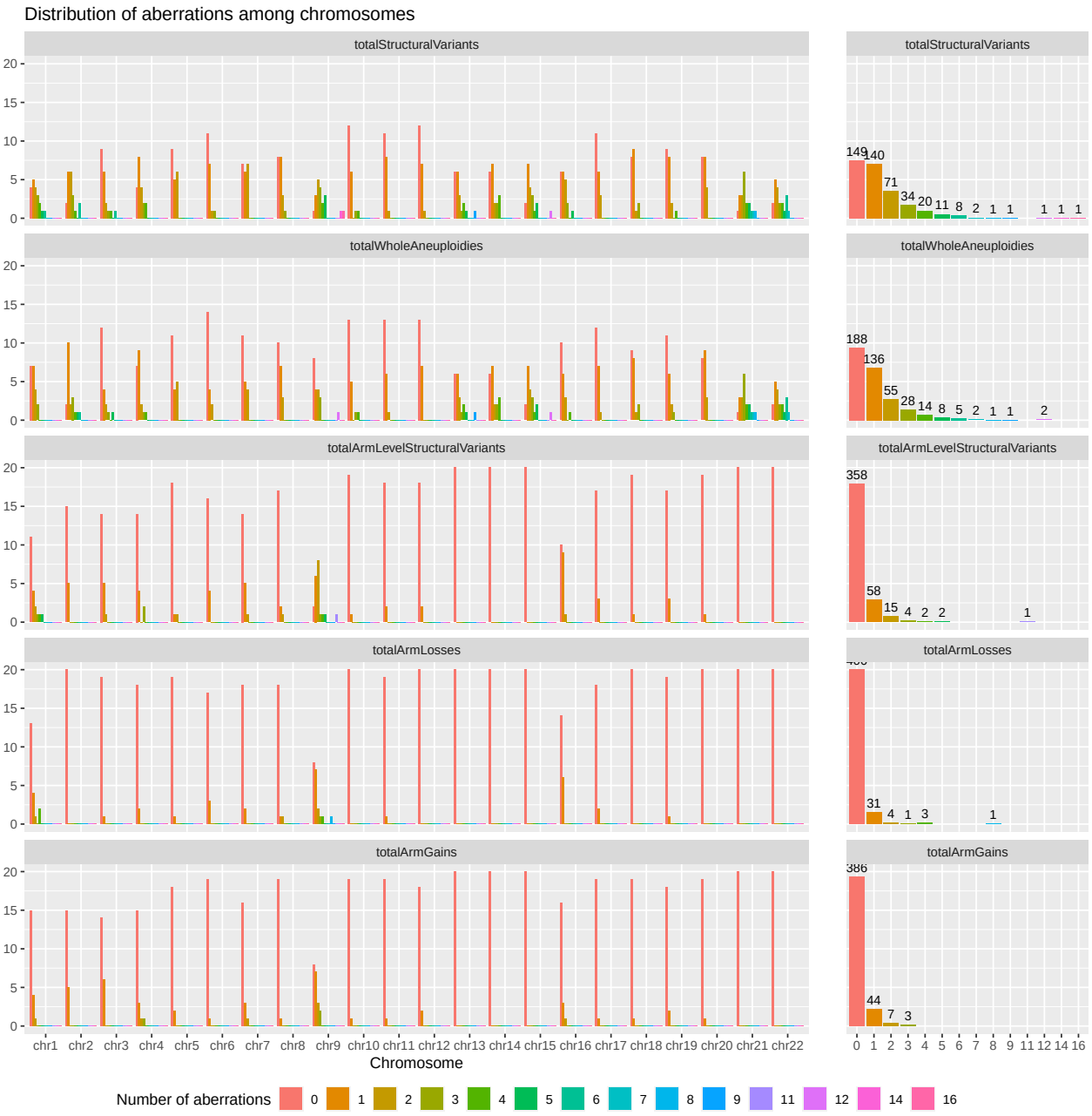
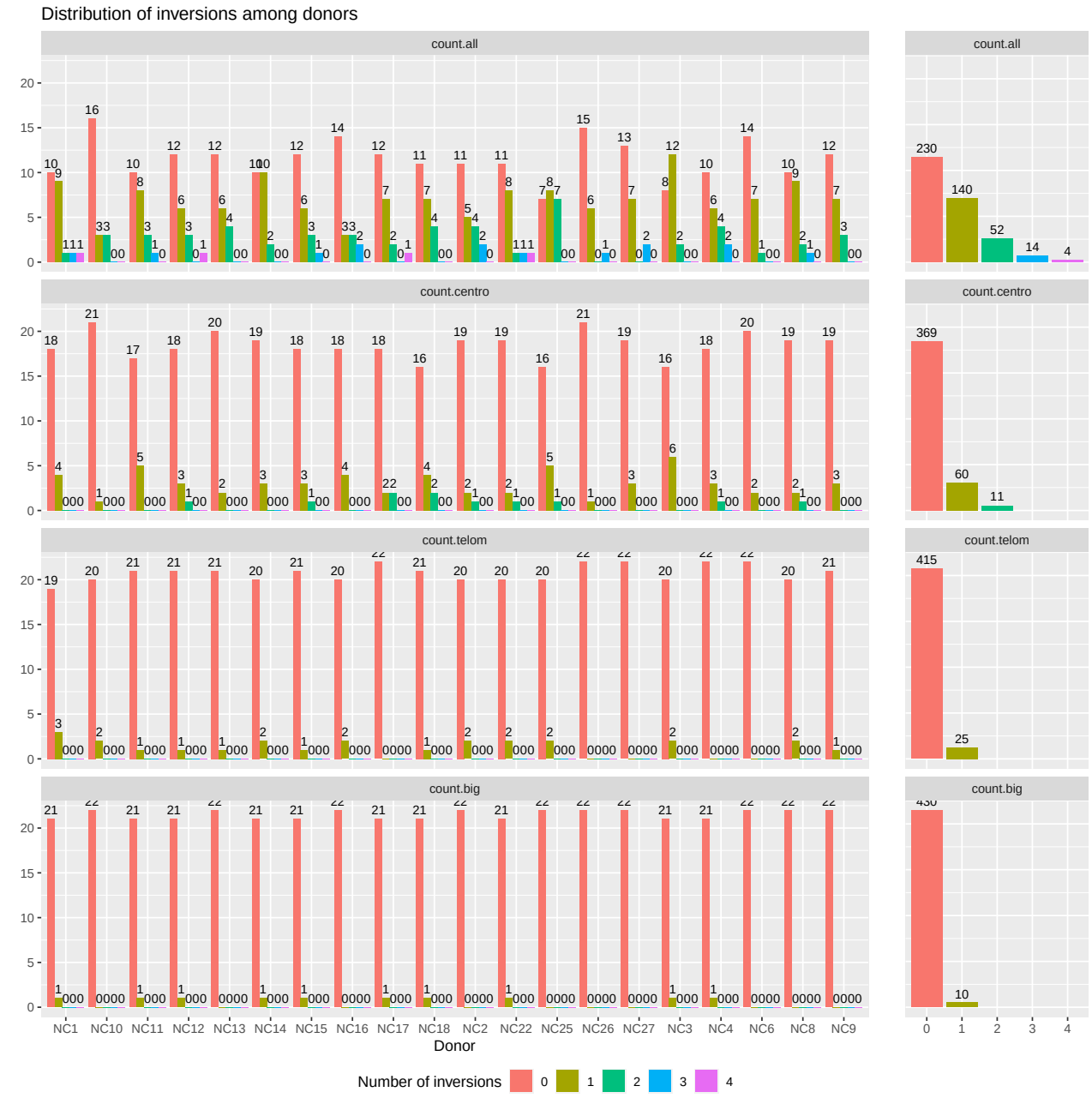


Figure 2: We see that telocentric chromosomes (13, 14, 15, 21, 22) do not have data for chromosome arm aberrations, because they are counted as whole aneuploidies. Since we cannot differentiate between both kinds of aberration, we'll have to discard those chromosomess or analyze them apart from each other.


Distribution of inversions among donors
count.all
count.centro
count.telom
count.big
NC1
NC10
NC11
NC12
NC13
NC14
NC15
NC16
NC17
NC18
NC2
NC22
NC25
NC26
NC27
NC3
NC4
NC6
NC8
NC9
Donor
Number of inversions
0
1
2
3
4

3.2 Determine outliers

As I mentioned before, chromosomes 13, 14, 15, 21 and 22 are telocentric and are not counted in the chromosome arm aberration dataset.

3.2.1 All chromosomes together

- chr9 is a structural variant outlier (for arm gains and losses as well)
- NC2 (arm losses) and NC3 (arm gains) have more structural variants than expected for the number of cells analyzed.
- chr 1 and 16 are outliers for arm losses besides chr9

- NC25 is an inversion count outlier
- chr8 (and its carriers) are inversion length outliers
- chr7 and chr16 are inversion length outliers in the centromeric group
- chr7 and 6 are length outliers within all inversions, besides chr8

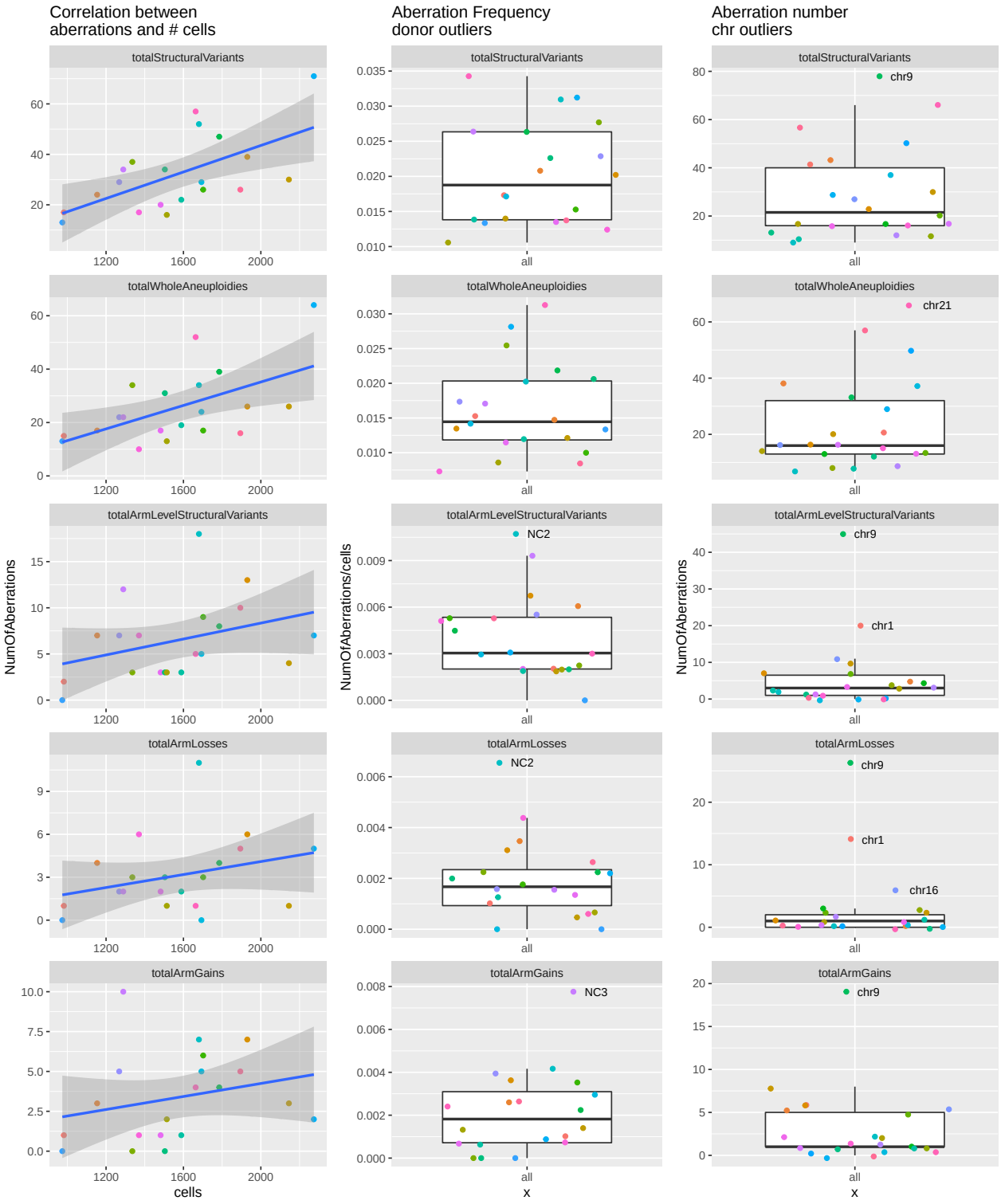


Figure 3: Scanning for aberration outliers

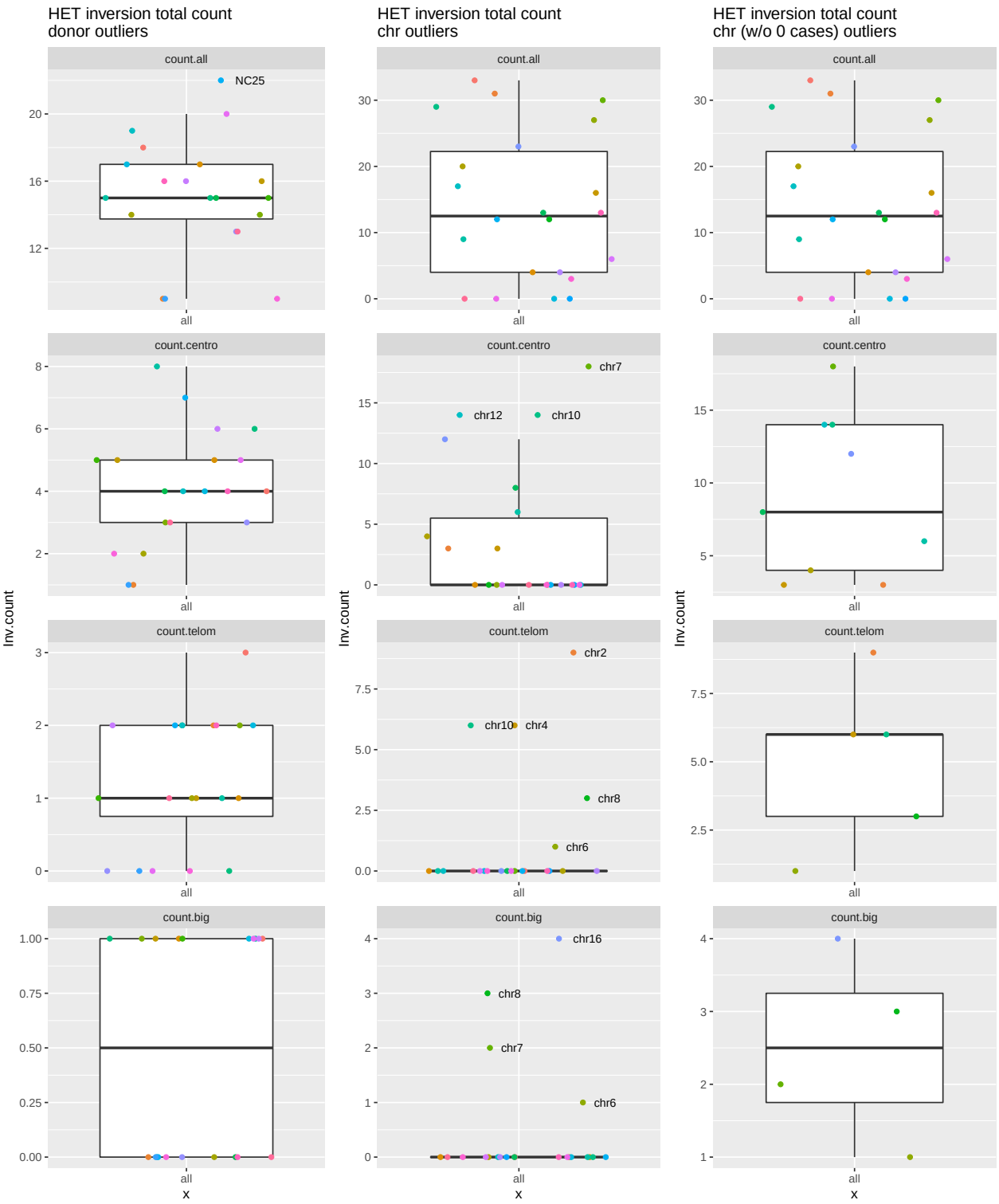


Figure 4: Scanning for inversion count outliers

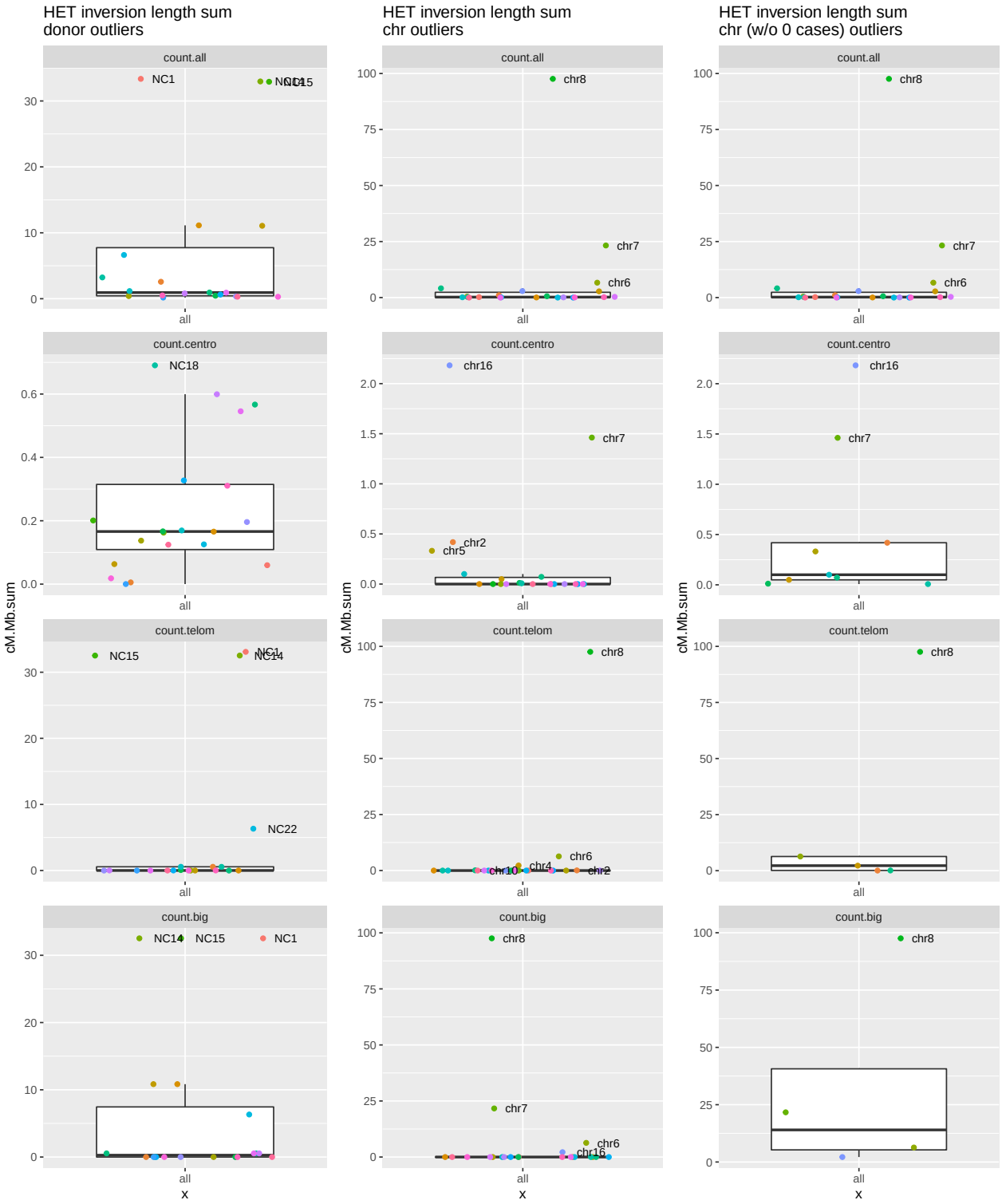


Figure 5: Scanning for inversion length sum outliers

3.2.2 Without the telocentric chromosomes

- chr9 is a structural variant outlier (for arm gains and losses as well)
- NC2 (arm losses) and NC3 (arm gains) have more structural variants than expected for the number of cells analyzed.
- chr 1 and 16 are outliers for arm losses besides chr9
- NC25 is an inversion count outlier
- chr8 (and its carriers) are inversion length outliers
- chr7 and chr16 are inversion length outliers in the centromeric group
- chr7 and 6 are length outliers within all inversions, besides chr8

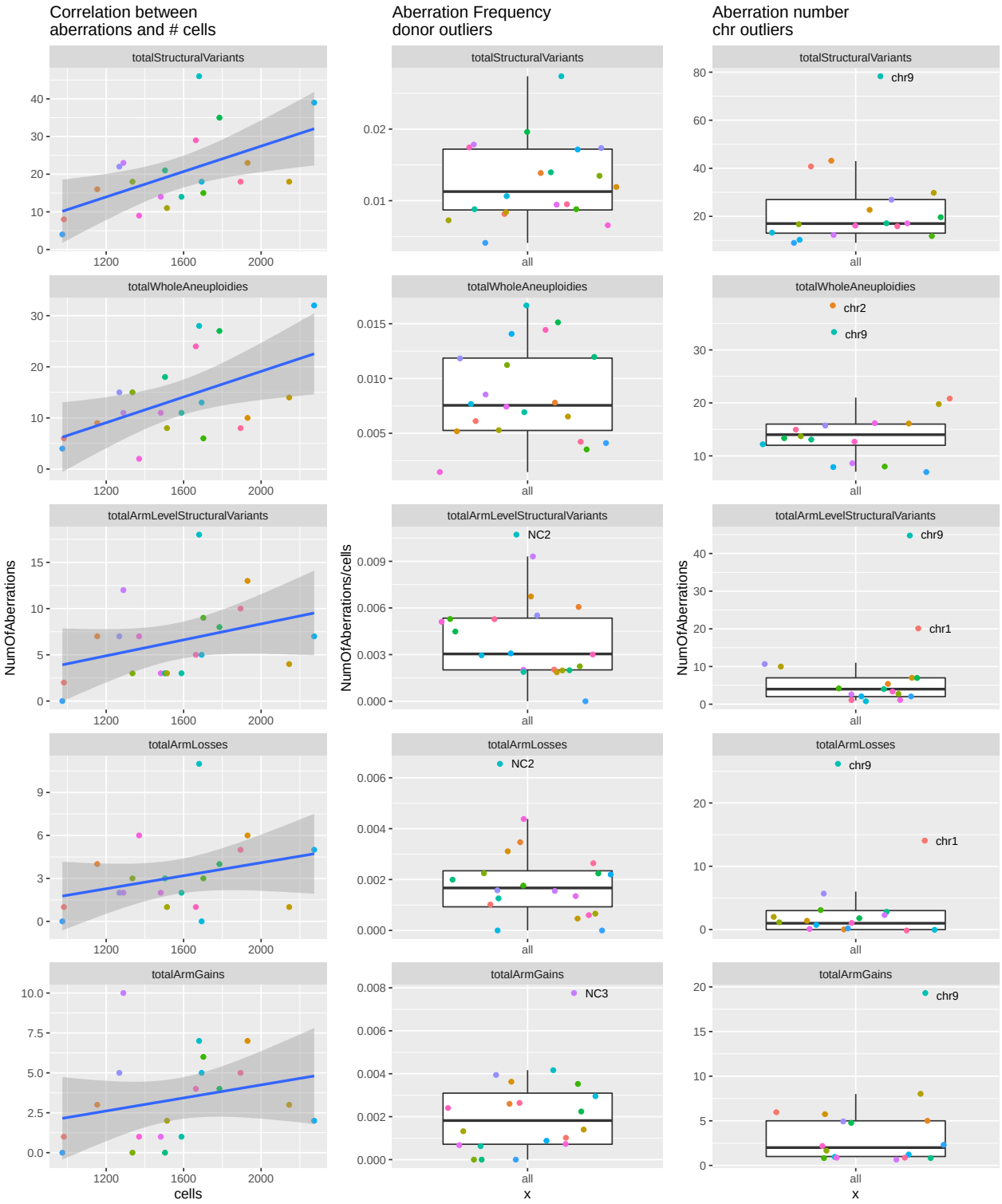


Figure 6: Scanning for aberration outliers

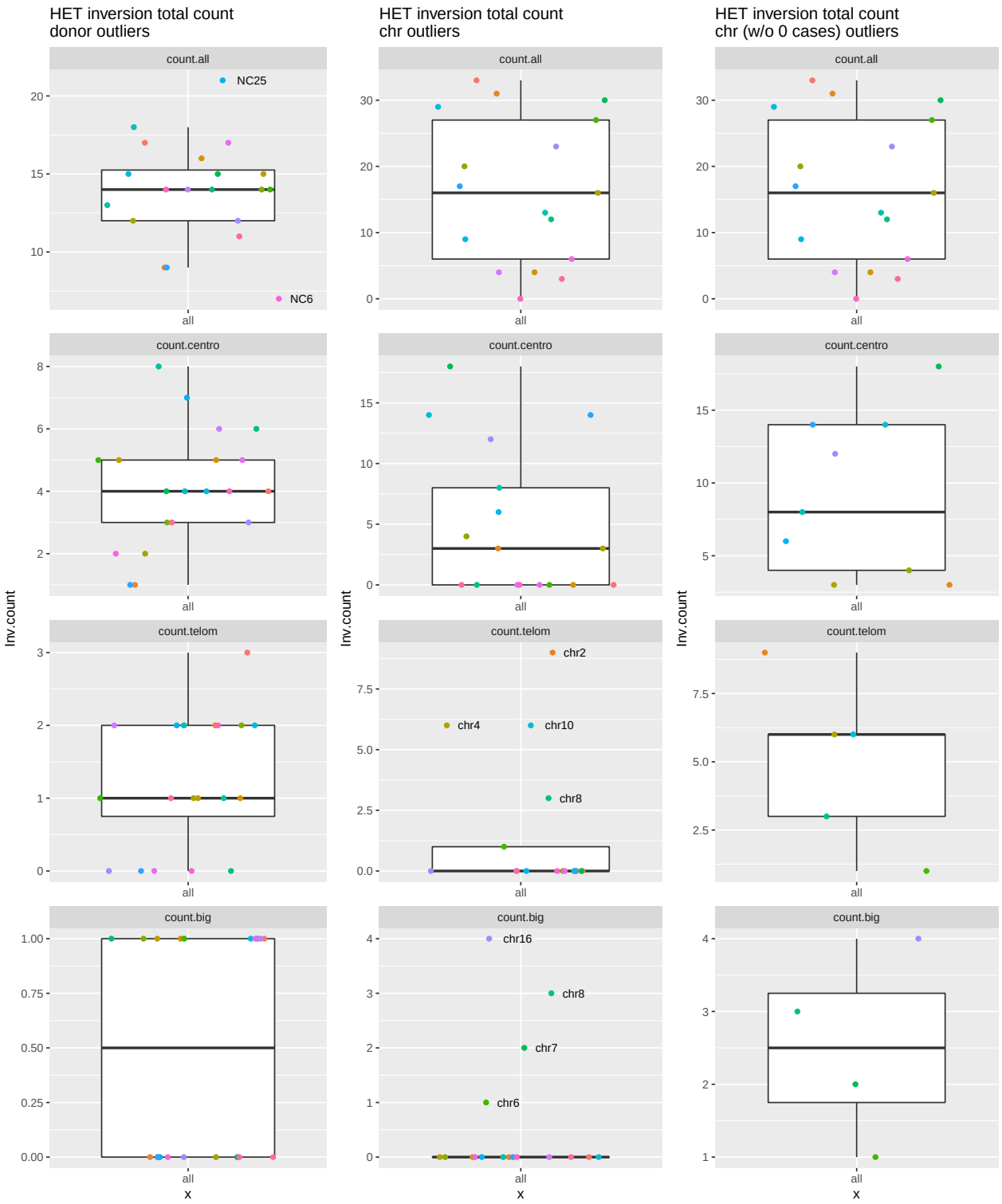


Figure 7: Scanning for inversion count outliers

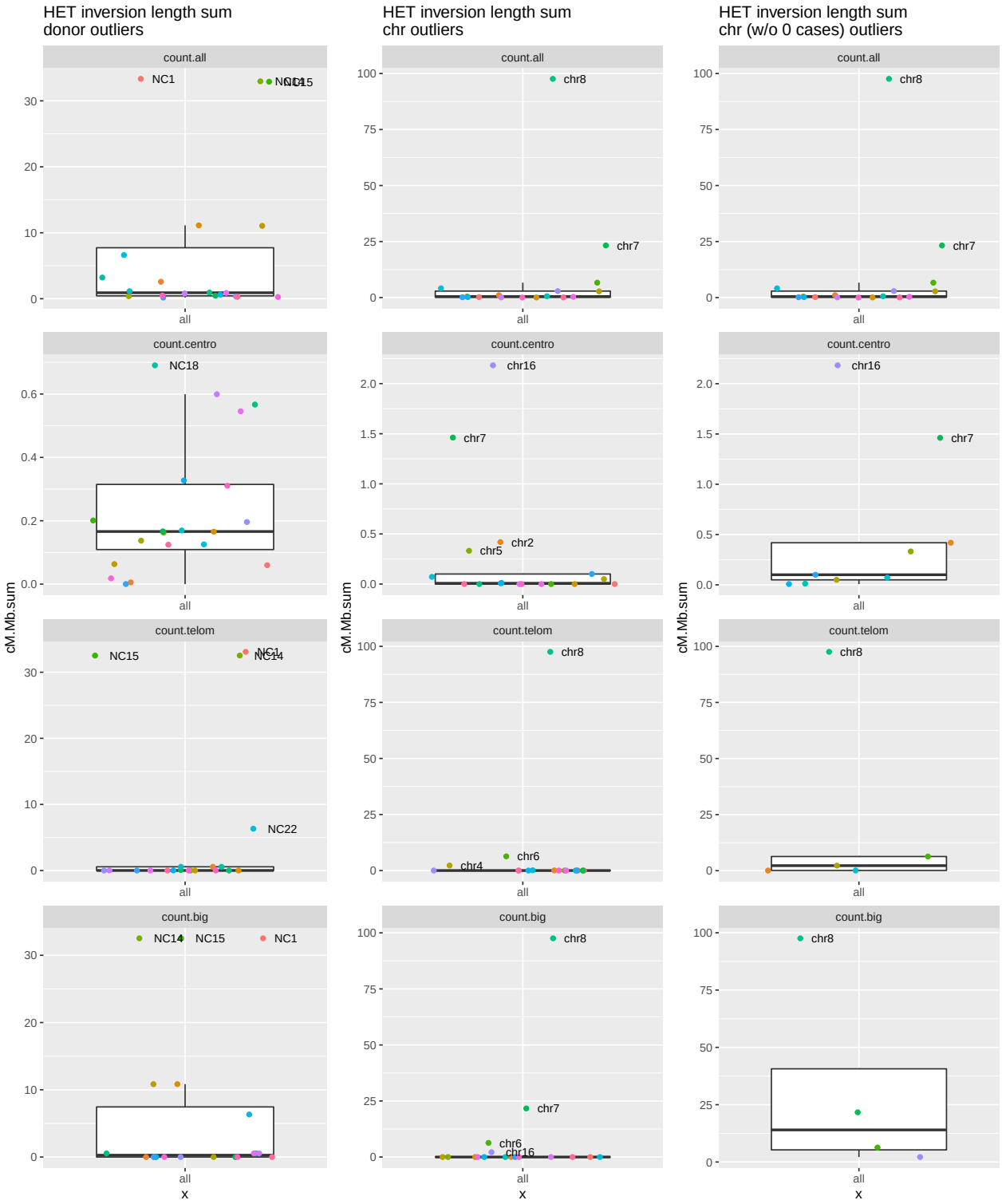


Figure 8: Scanning for inversion length sum outliers

3.3 Inversions vs. aberrations, exploratory analysis

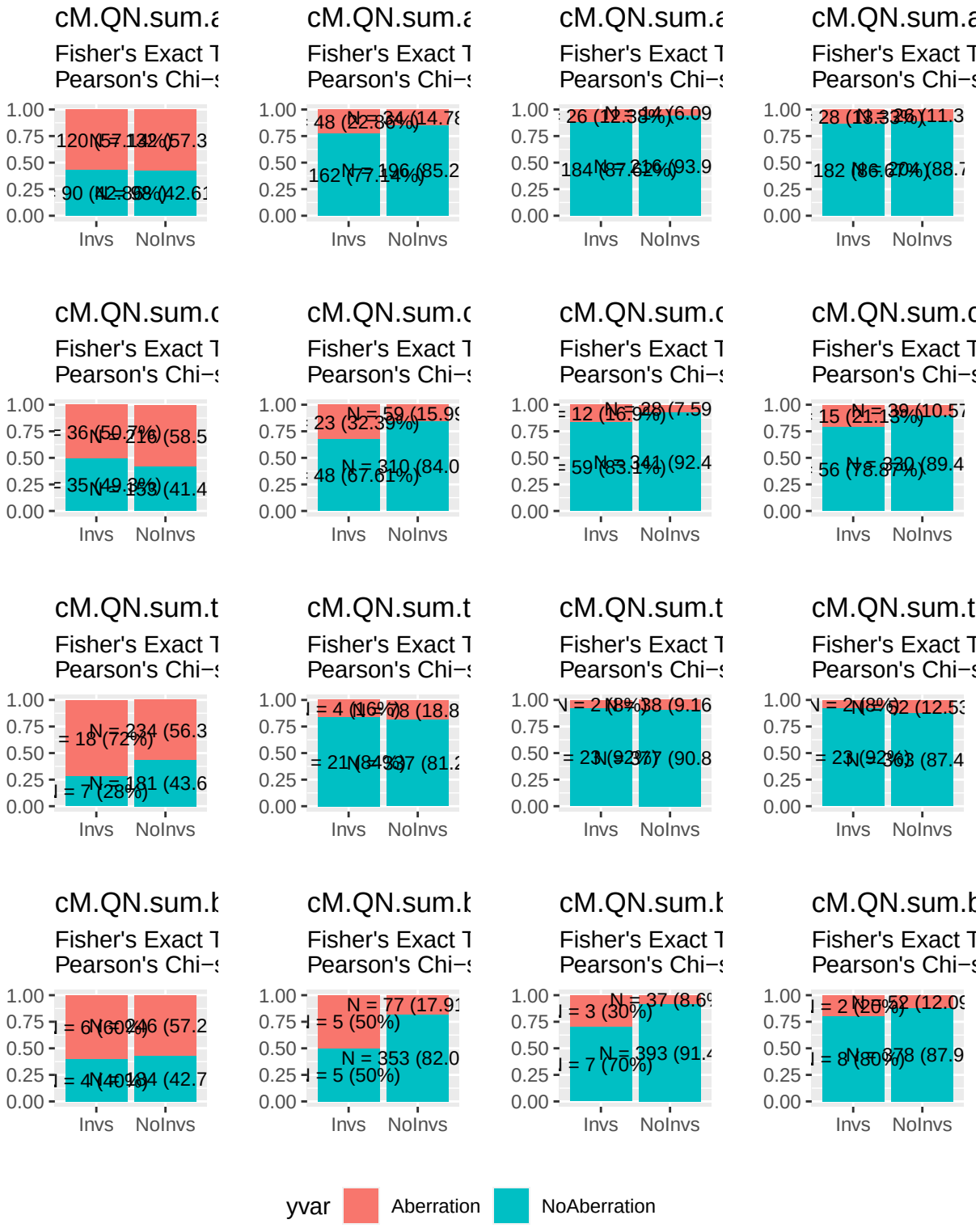
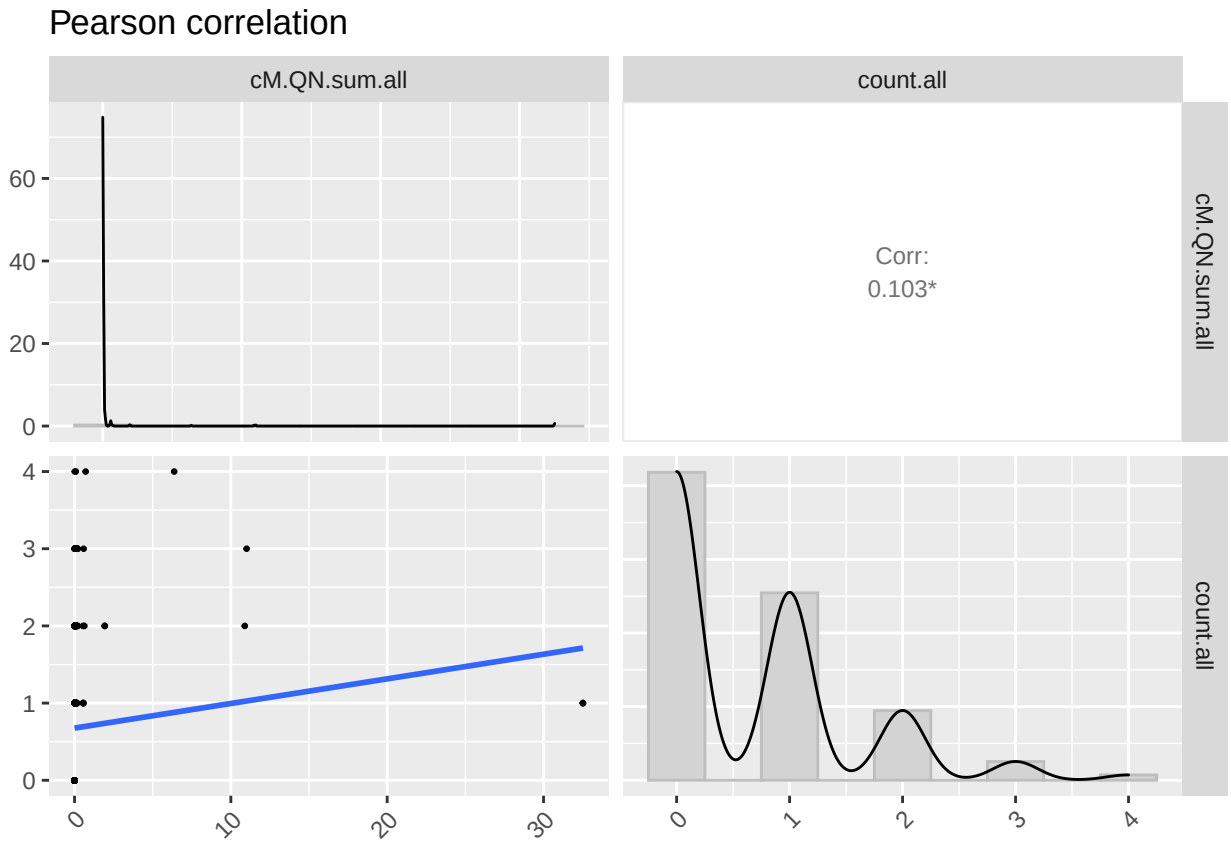


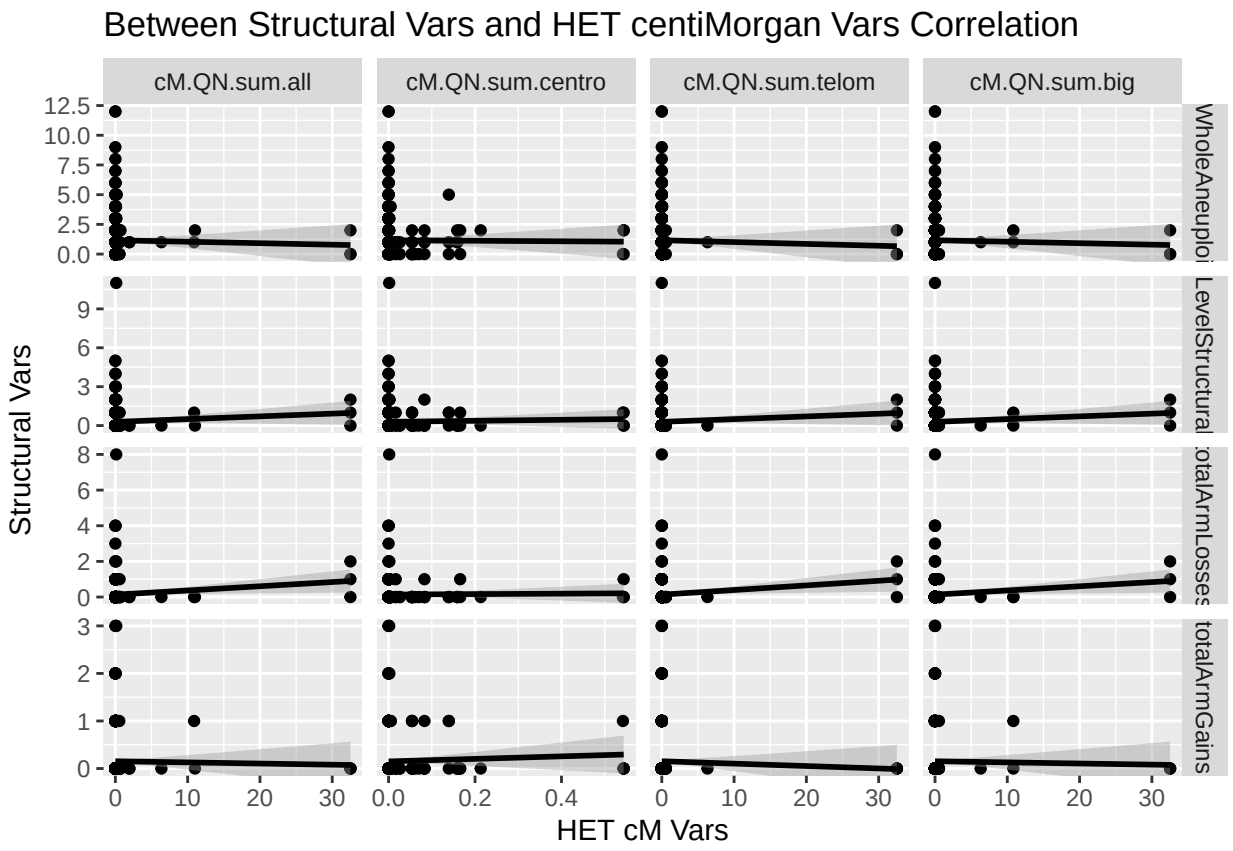
Figure 9: Analysis for the relationship between inversions and aberrations

```
pacman::p_load(GGally)

# Setup corrplot
  diagonal <- function(data, mapping, ...){
    p<-ggplot(data=data, mapping = mapping)+
      geom_histogram(aes(y= ..density..), bins = 9, fill = "lightgray", color = "gray")+
      geom_density()
  }

 lines <- function(data, mapping, ...){
    p<-ggplot(data=data, mapping = mapping)+
      geom_smooth(method=lm, alpha=0)+
      geom_point(size=0.5)
 }

ggpairs(TestDataset[,c("cM.QN.sum.all", "count.all")],
        lower = list(continuous = lines),
              diag = list(continuous = diagonal),
              upper = list(continuous = wrap("cor", method = "pearson", size = 3)))+
   ggtitle("Pearson correlation")+theme(axis.text.x = element_text(angle = 45, hjust = 1))+
   scale_fill_manual(values=c("#00AFBB", "#E7B800", "#FC4E07")) +scale_color_manual(values=c("#00AFBB",
```

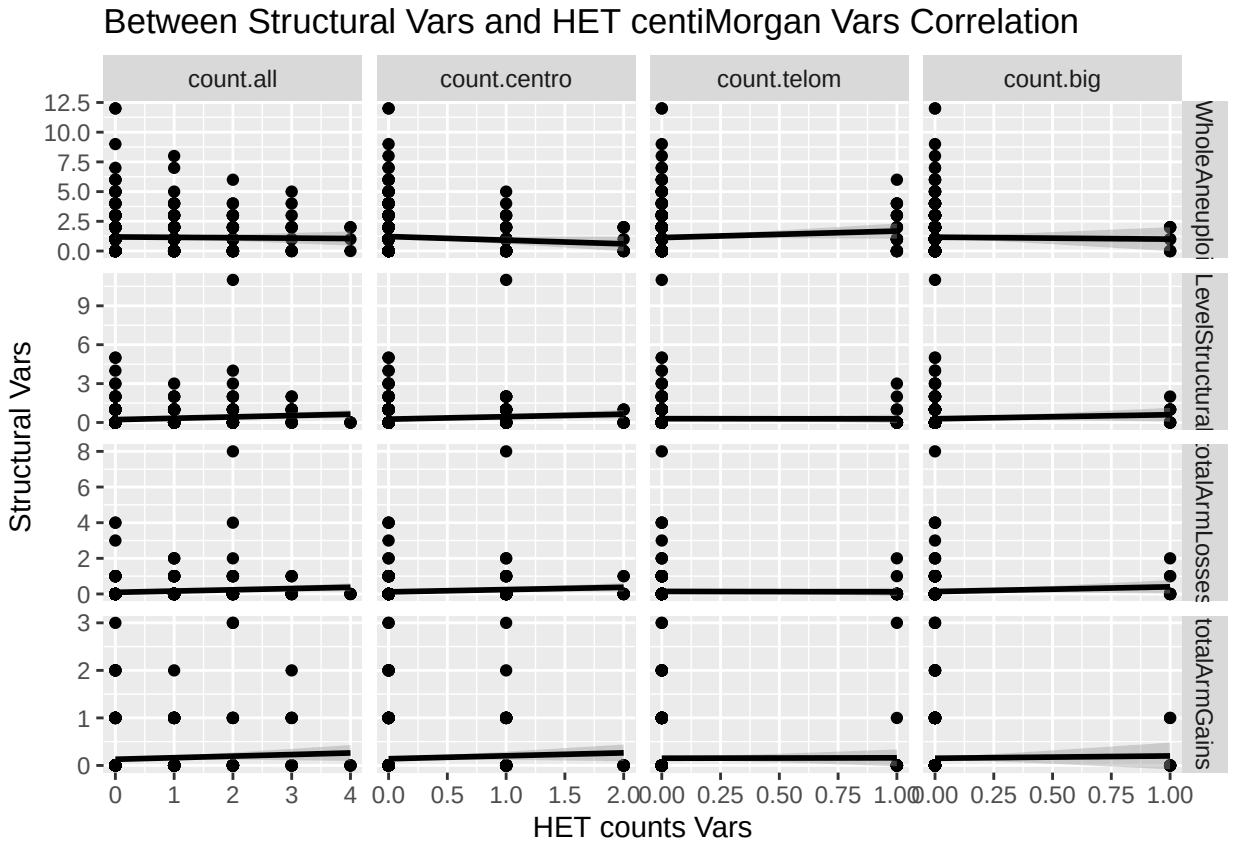



```
# They are not correlated so I'll analyze them separately
# SUMS
ab<- c("totalWholeAneuploidies", "totalArmLevelStructuralVariants", "totalArmLosses", "totalArmGains")
inv<-   as.character(unique(TestDataset.long$InvSubset.sum))
```

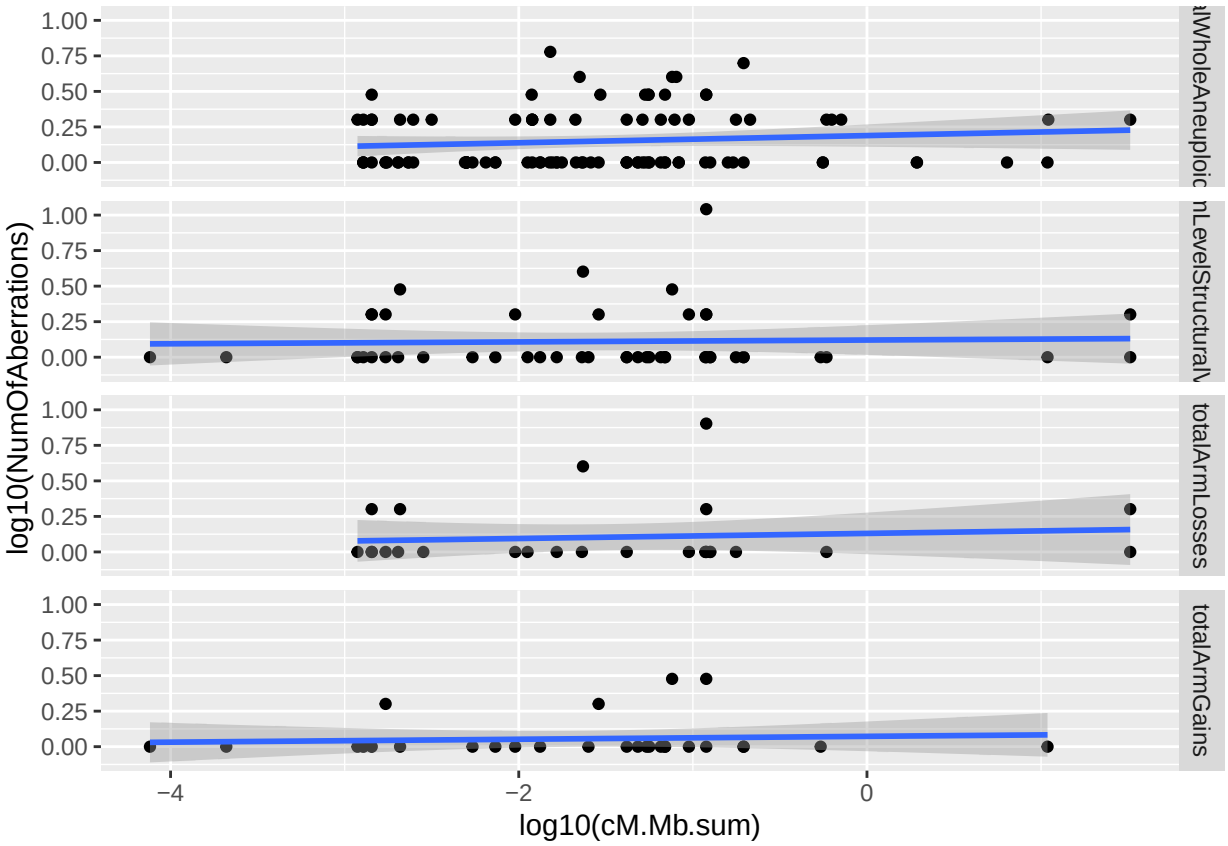
```
ggduo(
  TestDataset[,c(ab, inv)],inv,ab,
  types = list(continuous = "smooth_lm"),
  title = "Between Structural Vars and HET centiMorgan Vars Correlation",
  ylab = "Structural Vars",
  xlab = "HET cM Vars"
)
```




```
# COUNTS
ab<- c("totalWholeAneuploidies", "totalArmLevelStructuralVariants", "totalArmLosses", "totalArmGains")
inv<-   as.character(unique(TestDataset.long$InvSubset.count))
ggduo(
  TestDataset[,c(ab, inv)],inv,ab,
  types = list(continuous = "smooth_lm"),
  title = "Between Structural Vars and HET centiMorgan Vars Correlation",
  ylab = "Structural Vars",
  xlab = "HET counts Vars"
)
```




```
# ONLY POSITIVES

# They are not correlated so I'll analyze them separately
# SUMS
TestDataset.long$InvPresence<-ifelse(TestDataset.long$Inv.count > 0, "WithInvs", "WithoutInvs")
ggplot(TestDataset.long[which((TestDataset.long$TypeOfAberration %in% ab )& TestDataset.long$InvPresenc
       aes(y = log10(NumOfAberrations), x = log10(cM.Mb.sum)  ))+
  geom_point()+geom_smooth(method = "lm")+
  facet_grid(TypeOfAberration ~ .  )
```



3.4 Inversions vs. aberrations, advanced analysis

4 Discussion

The big inversions and the telomeric inversions subsets raised a warning because they did not have enough sample size to use a Chi-squared test, although there wasn't any significant result from the Fisher's exact test for these subsets either.

When considering all inversions together, chromosomes with heterozygous inversions have significantly more arm losses, which is consistent with the fact that any crossover between opposite orientations will generate an acentric that is lost, so there is always less genetic material than usual in the resulting chromosomes.

Centromeric inversions in heterozygosis are significantly associated to both arm losses and arm gains. The relationship with arm losses was expected, but I do not understand where do arm gains come from, unless there are cases in which the dicentric does not break.

Finding an increase in aberrations related to the presence of inversions in heterozygosis is consistent with the purifying selection theory and with other experiments regarding this same question.