

MECP2 Data Analysis

TS

December 15, 2017

Contents

Overview of data	1
What is the distribution of genotypes?	2
Spike Statistics	2
Bursting activity	4
Do our networks show bursting activity?	4
Are there any difference in bursting activity between the genotypes over development?	6
Firing Regularity	9
How does firing regularity change with neural network development?	9
Differences in regularity between genotypes?	10
Effective Rank	11
Does effective rank change through development?	11
Is there a difference in effective rank between genotypes?	12
What if we look at some of these statistics only in cultures with bursts?	13
Moving forward	14
Data selection	14
Data analyses	15
<ul style="list-style-type: none">• This document provides an overview of the extracted features from neural networks recorded using micro-electrode arrays (MEA) in wild-type (WT) and MECP2-deficient (HT, KO) mice.• The current implementation provides summary statistics for 1209 and 2309 batch• The current spike statistics are based on spikes detected using a mean-SD based threshold method after filtering with a Butterworth filter (same method in Manuel's Rich Club Topology paper)<ul style="list-style-type: none">– threshold of 5 x SD	

Spike detection and spike analysis were performed in Matlab. R will be used to perform statistical analysis

Overview of data

Let's begin with a preview of the data

```
head(df)
```

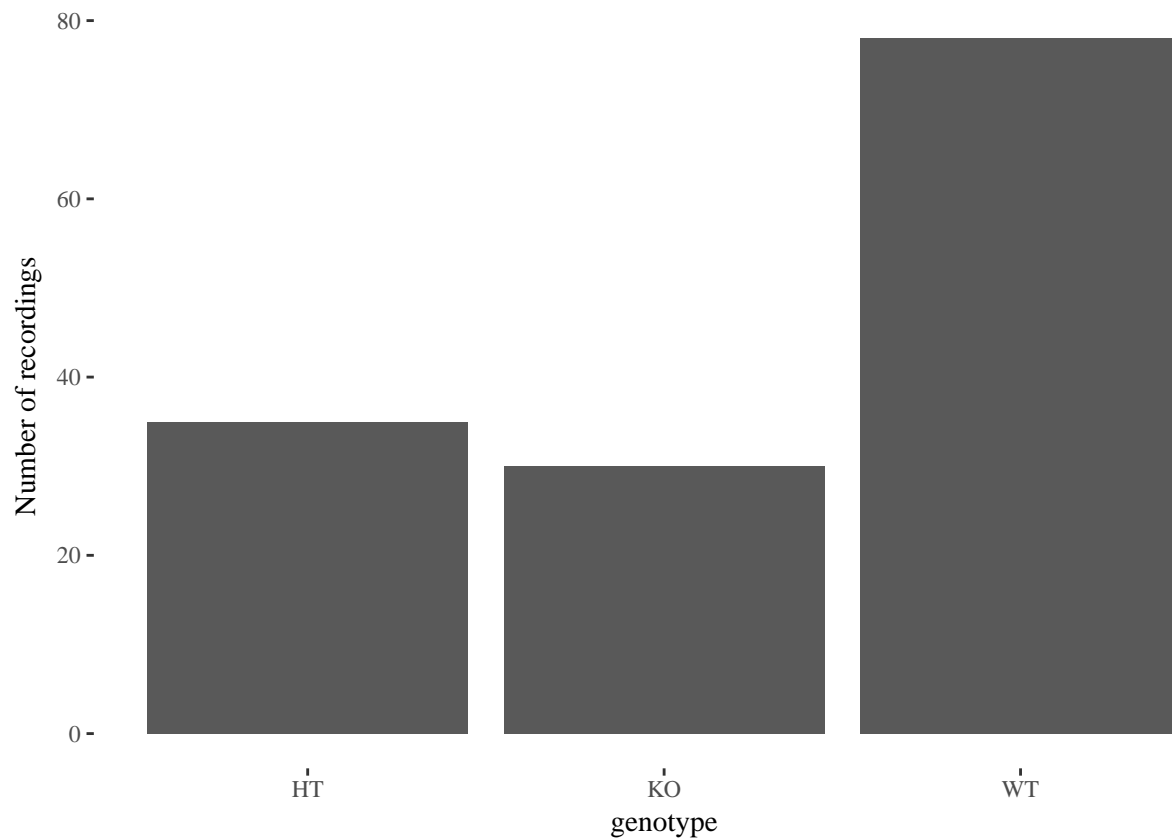
```
##      batch      ID DIV genotype aveFireRate totalBurstCount
## 1.1 12_09_17 12_09_17-3A 14      HT      0.3736111           0
## 1.2 12_09_17 12_09_17-3A 22      HT      0.6541667           3
## 1.3 12_09_17 12_09_17-3A 28      HT      0.3888889           0
## 1.4 12_09_17 12_09_17-3A 35      HT      0.7416667           0
## 1.5 12_09_17 12_09_17-3A 42      HT      0.4388889           0
## 1.6 12_09_17 12_09_17-3A 49      HT      0.5930556           3
##      totalAveBurstSpikeNum      totalReg      effRankCov
```

```
## 1.1      NaN      NaN  52.14387
## 1.2      20 -0.75140880  28.26977
## 1.3      NaN      NaN  51.76948
## 1.4      NaN  0.09791845  55.92442
## 1.5      NaN -0.01229785  50.36599
## 1.6      12 -0.29258737  36.30183
```

What is the distribution of genotypes?

- note that this includes all recording throughout development

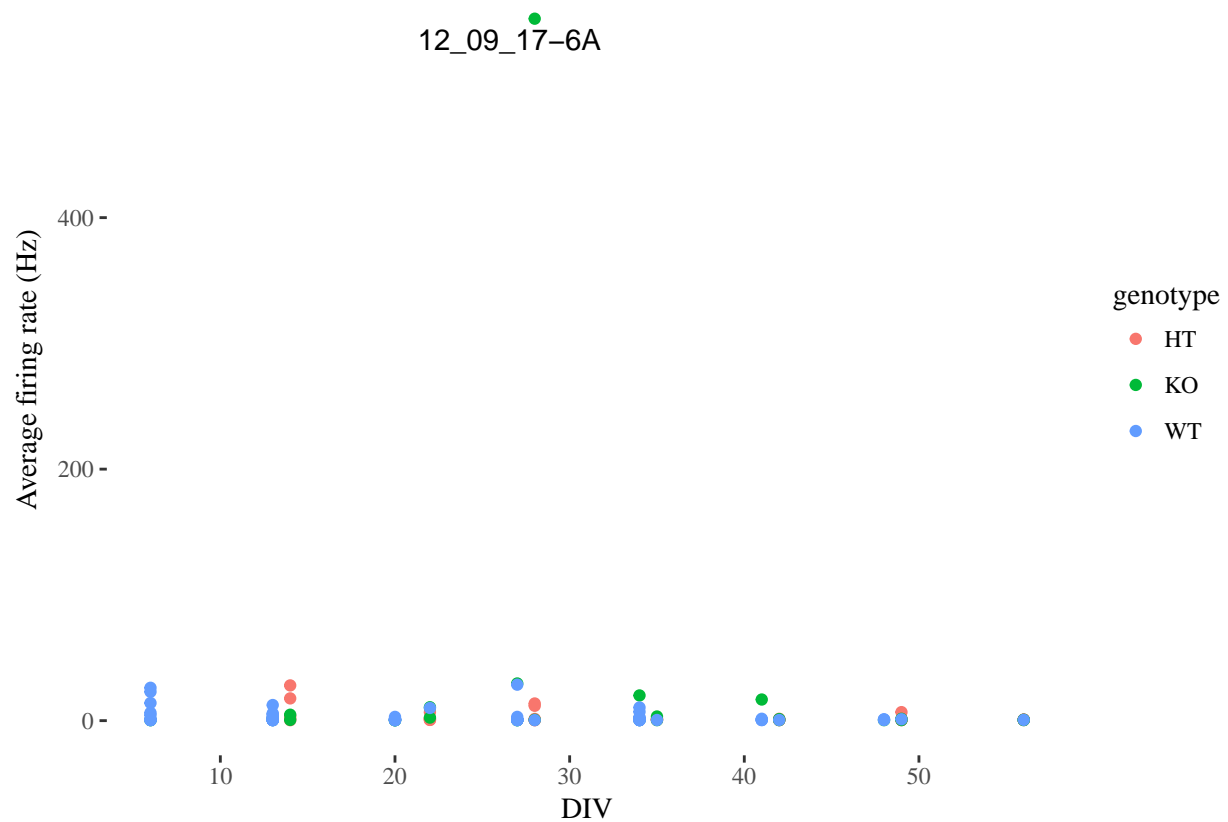
```
ggplot(data = df) + geom_bar(mapping = aes(x = genotype)) + ylab("Number of recordings") +
  th
```



Spike Statistics

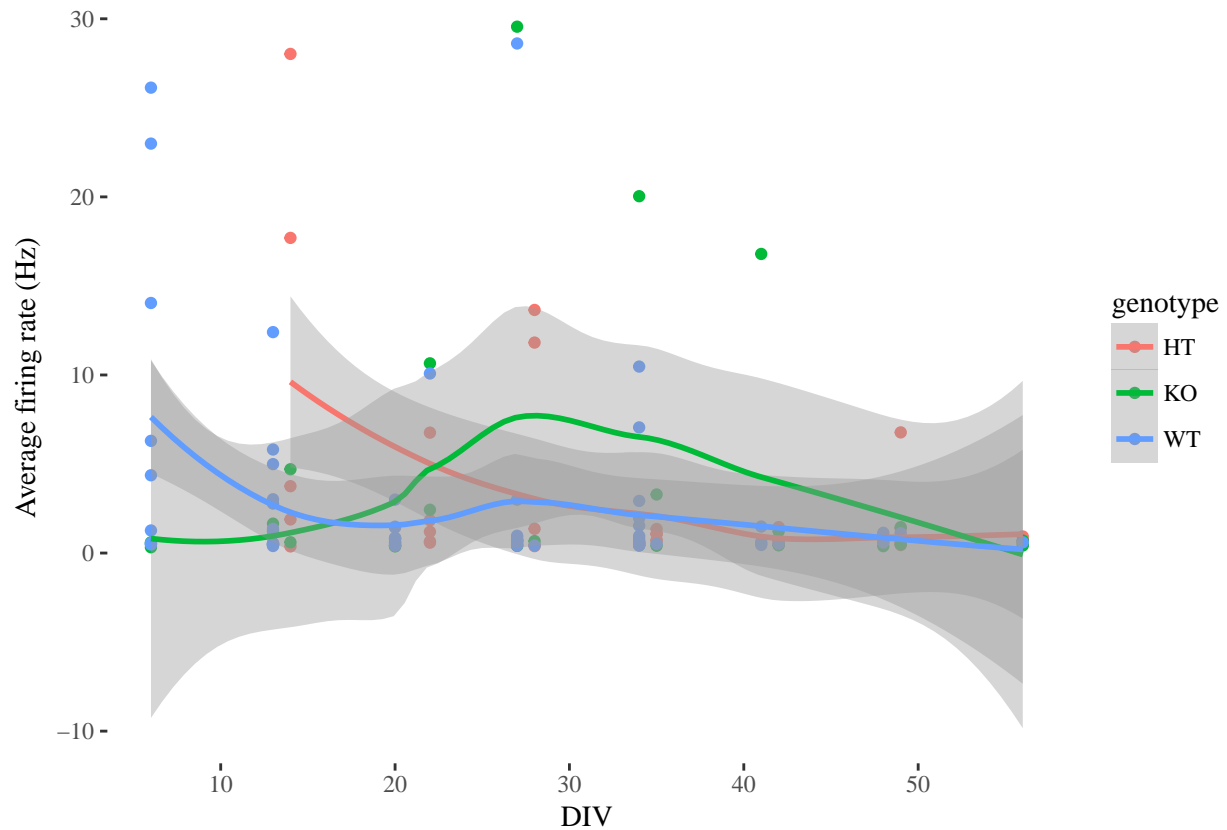
First let's ask the question: do we see any trends in spike firing rate over development?

```
ggplot(data = df) + geom_point(mapping = aes(x = DIV, y = aveFireRate,
  color = genotype)) + geom_text_repel(data = subset(df, aveFireRate >
  200), aes(DIV, aveFireRate, label = ID)) + ylab("Average firing rate (Hz)") +
  th
```



We notice that there is an outlier obscuring the plot, let's look at the overall activity without it.

```
ddf <- df[!(df$aveFireRate > 100), ]
ggplot(data = ddf) + geom_point(mapping = aes(x = DIV, y = aveFireRate,
  color = genotype)) + geom_smooth(mapping = aes(x = DIV, y = aveFireRate,
  color = genotype)) + ylab("Average firing rate (Hz)") + th
```



- Firing rate does not change much over time, or perhaps show a slight decrease
- Firing rate amongst the three genotypes seem quite similar over time

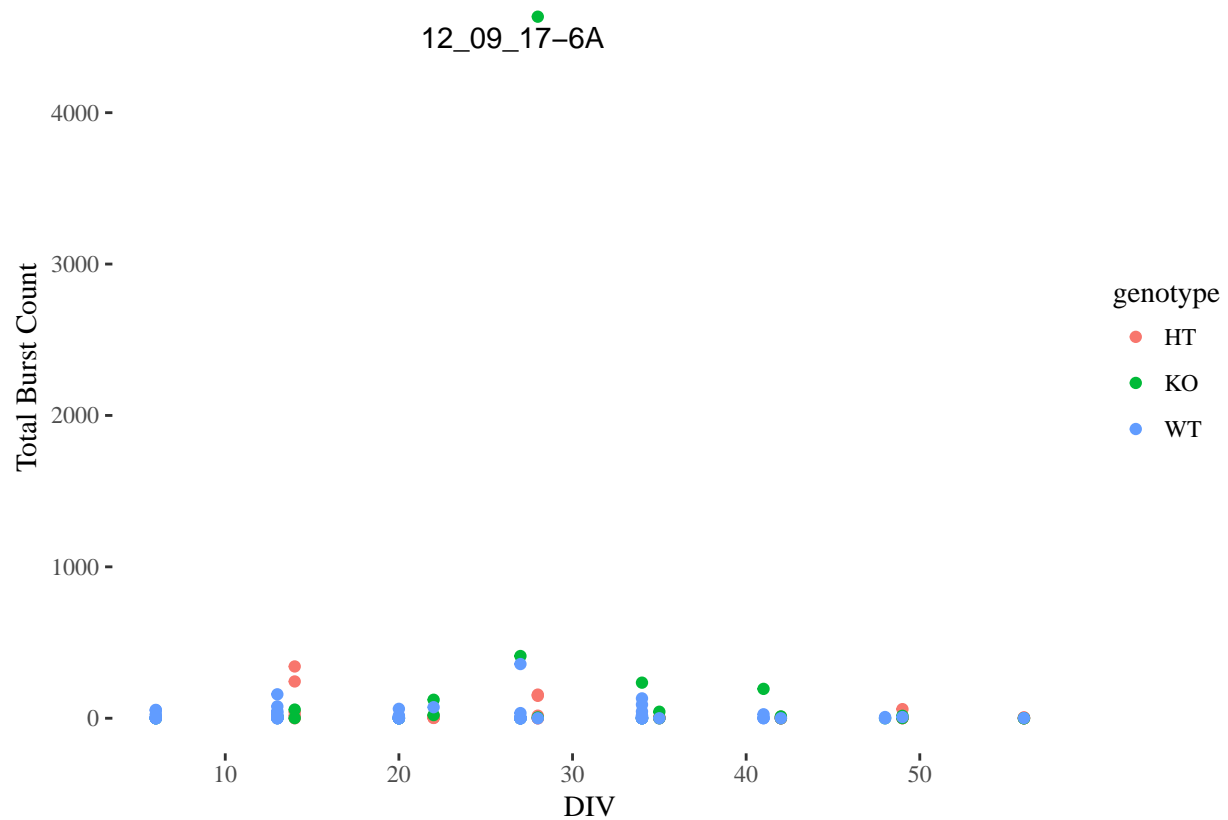
Bursting activity

Do our networks show bursting activity?

We first look at how many of our neural networks showed bursting activity:

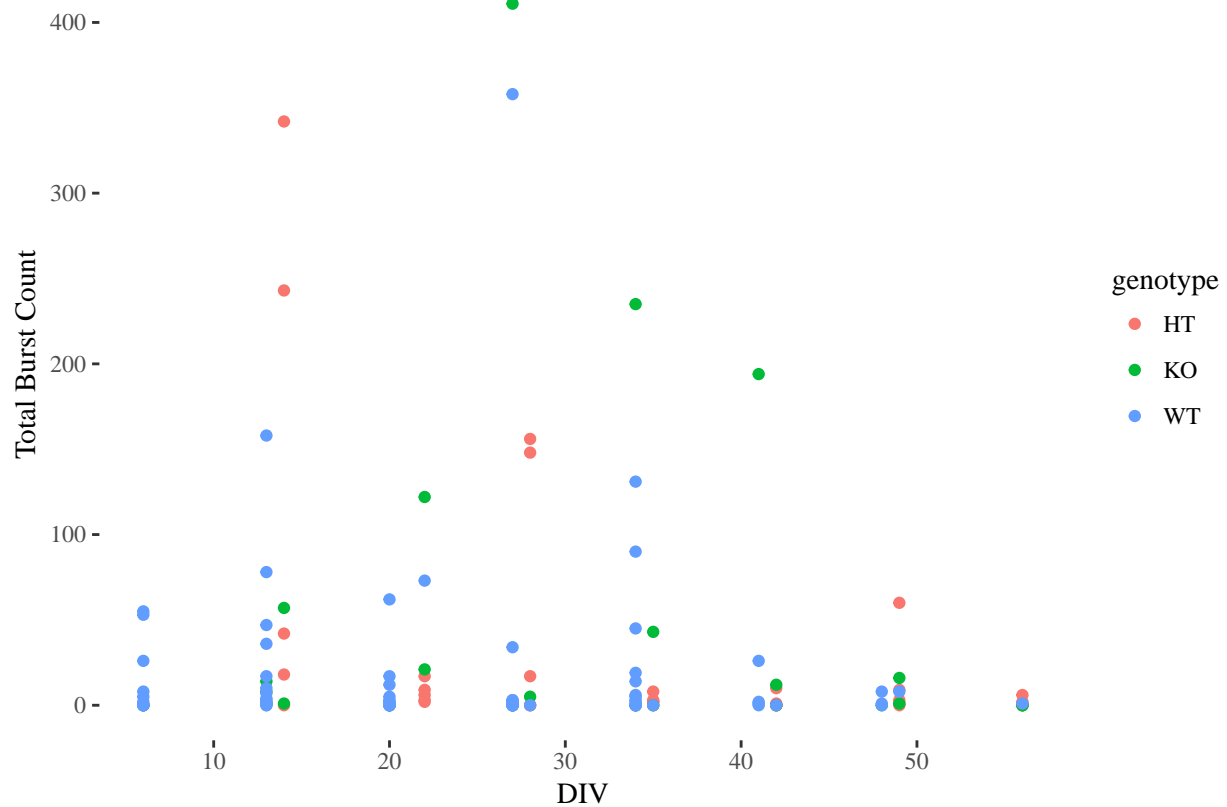
- bursting activity calculate using the gamma distribution method (Used by Eglén)
- this calculates the bursting activity of each electrode independently
- total burst count is the sum of the number of bursts observed in all electrodes

```
ggplot(data = df) + geom_point(mapping = aes(x = DIV, y = totalBurstCount,
  color = genotype)) + ylab("Total Burst Count") + geom_text_repel(data = subset(df,
  totalBurstCount > 1000), aes(DIV, totalBurstCount, label = ID)) +
  th
```



Again we see outlier 1209-6A, the same outlier for average firing rate. There may be positive correlation between fire rate and burst rate. We will consider that later, but for now let's see if we can find any patterns in bursting activity by removing the outlier.

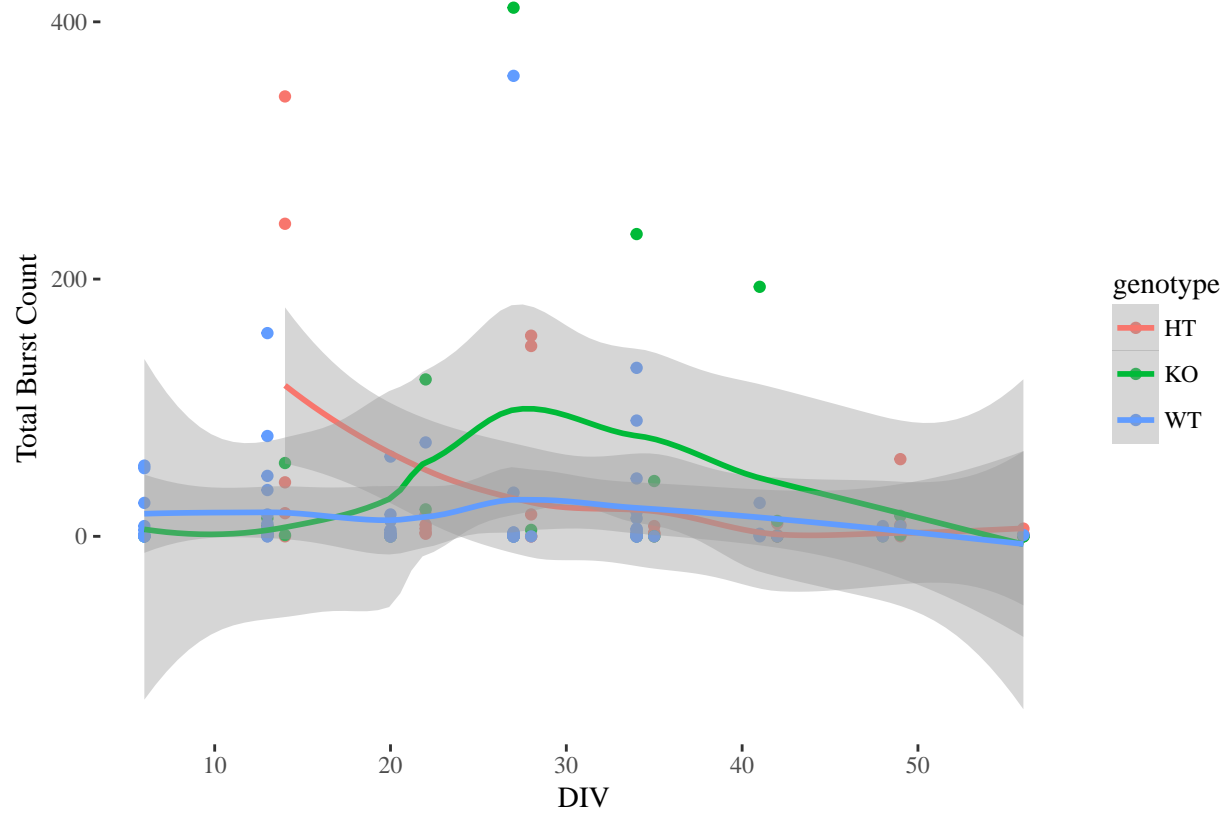
```
ggplot(data = subset(df, totalBurstCount < 1000)) + geom_point(mapping = aes(x = DIV,
  y = totalBurstCount, color = genotype)) + ylab("Total Burst Count") +
  th
```



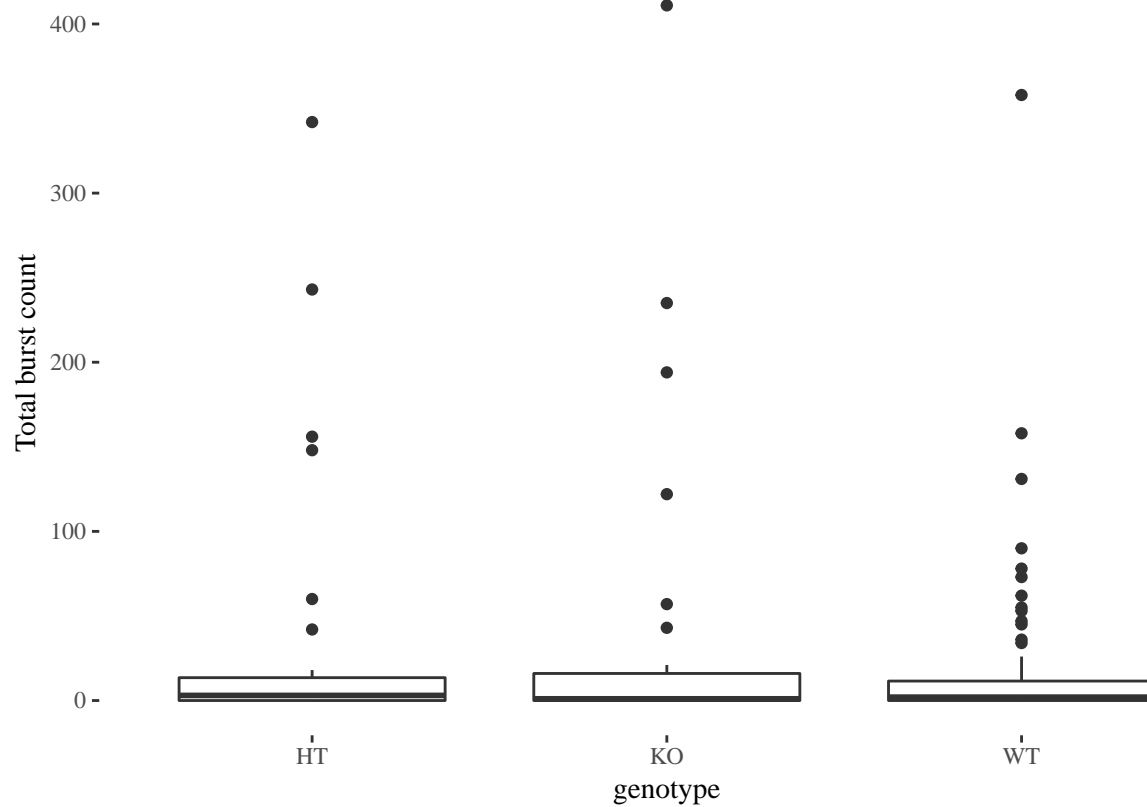
Are there any difference in bursting activity between the genotypes over development?

```
ggplot(data = subset(df, totalBurstCount < 1000)) + geom_point(mapping = aes(x = DIV,
  y = totalBurstCount, color = genotype)) + geom_smooth(mapping = aes(x = DIV,
  y = totalBurstCount, color = genotype)) + ylab("Total Burst Count") +
  th
```

```
## `geom_smooth()` using method = 'loess'
```



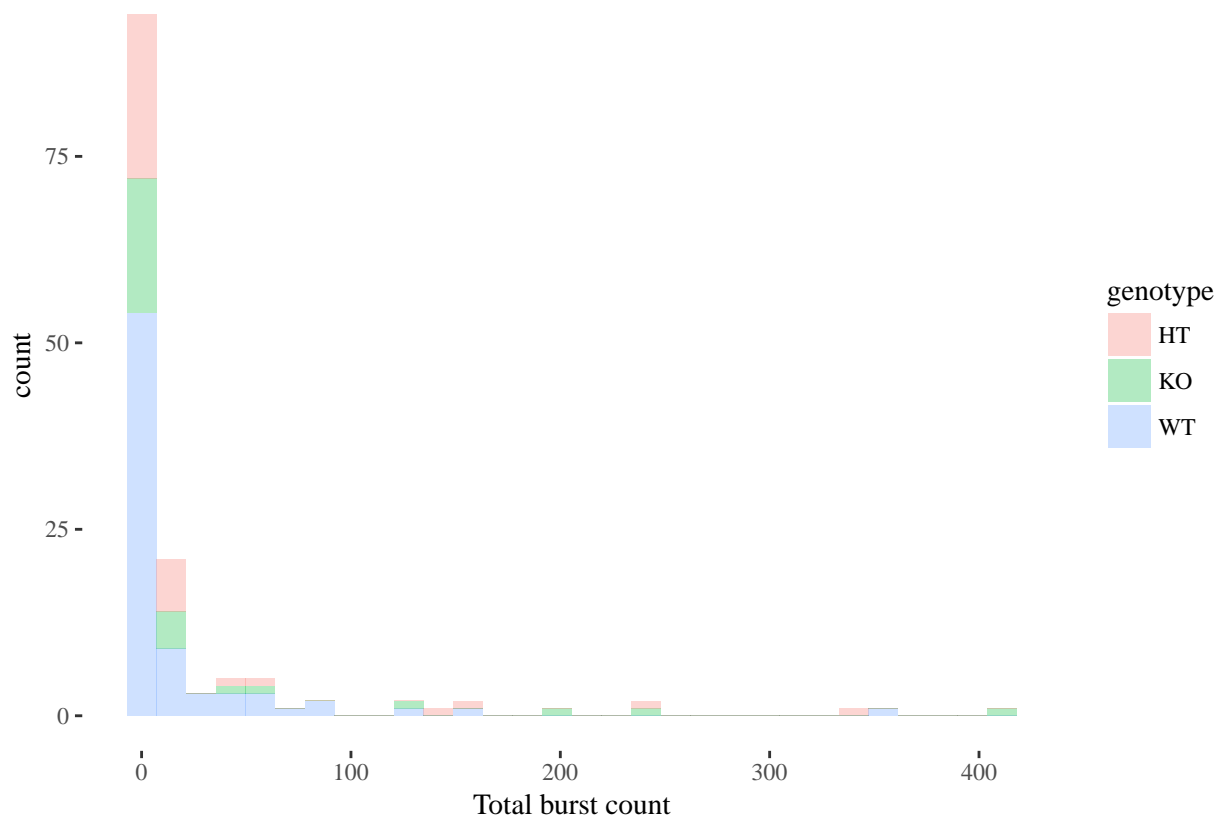
```
ggplot(data = subset(df, totalBurstCount < 1000)) + geom_boxplot(mapping = aes(genotype,
totalBurstCount)) + ylab("Total burst count") + theme_tufte()
```



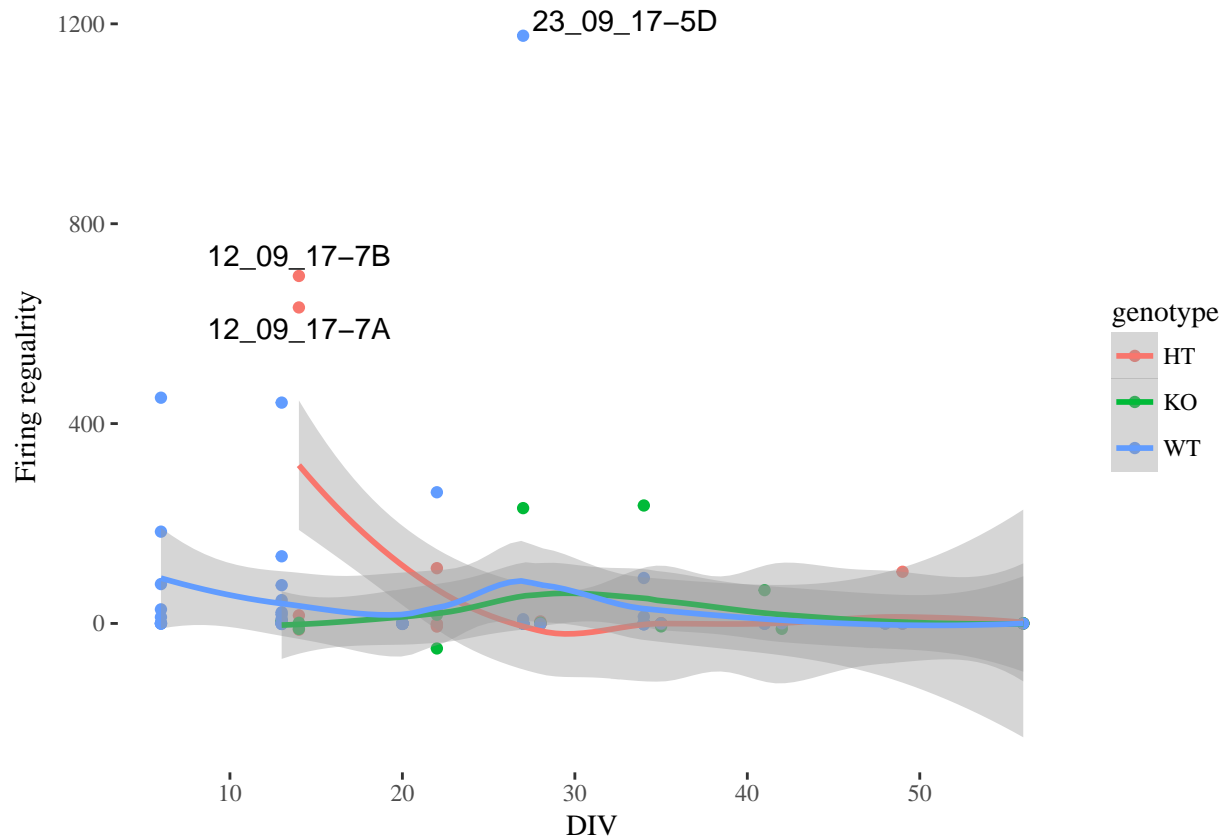
Looking more specifically at the distribution of bursting activities

```
ggplot(data = subset(df, totalBurstCount < 1000)) + geom_histogram(mapping = aes(x = totalBurstCount,
  fill = genotype), alpha = 0.3) + xlab("Total burst count") +
  th
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
ggplot(data = subset(df, is.finite(totalReg))) + geom_point(mapping = aes(x = DIV,
y = totalReg, color = genotype)) + geom_smooth(mapping = aes(x = DIV,
y = totalReg, color = genotype)) + ylab("Firing regularity") +
geom_text_repel(data = subset(df, totalReg > 500 & is.finite(totalReg)),
aes(x = DIV, y = totalReg, label = ID)) + th
```

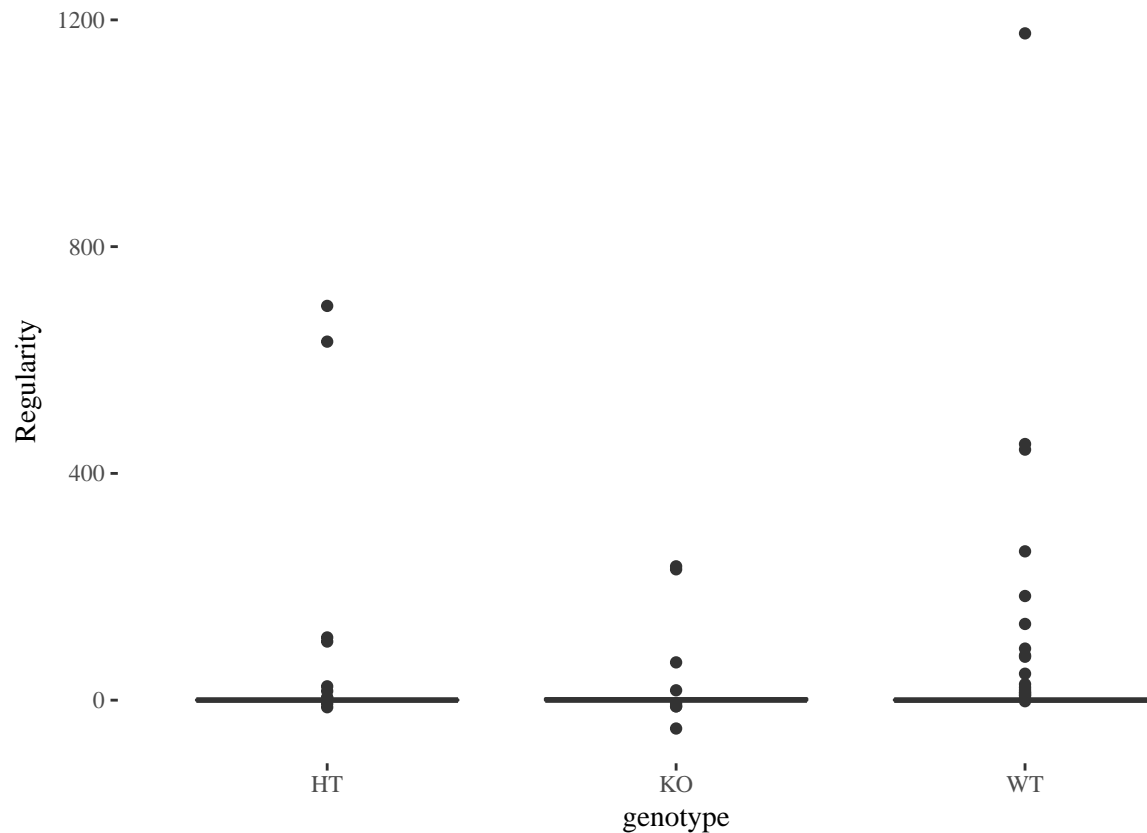


Thoughts:

- it appears that all three genotypes converge to a regularity value of 0 over development
- there are two particular outliers in the HT group
- (possible) high regularity value at start of development for HT and WT group, but number of samples too small to draw any conclusions

Differences in regularity between genotypes?

```
ggplot(data = subset(df, is.finite(totalReg))) + geom_boxplot(mapping = aes(genotype,
totalReg)) + ylab("Regularity") + th
```



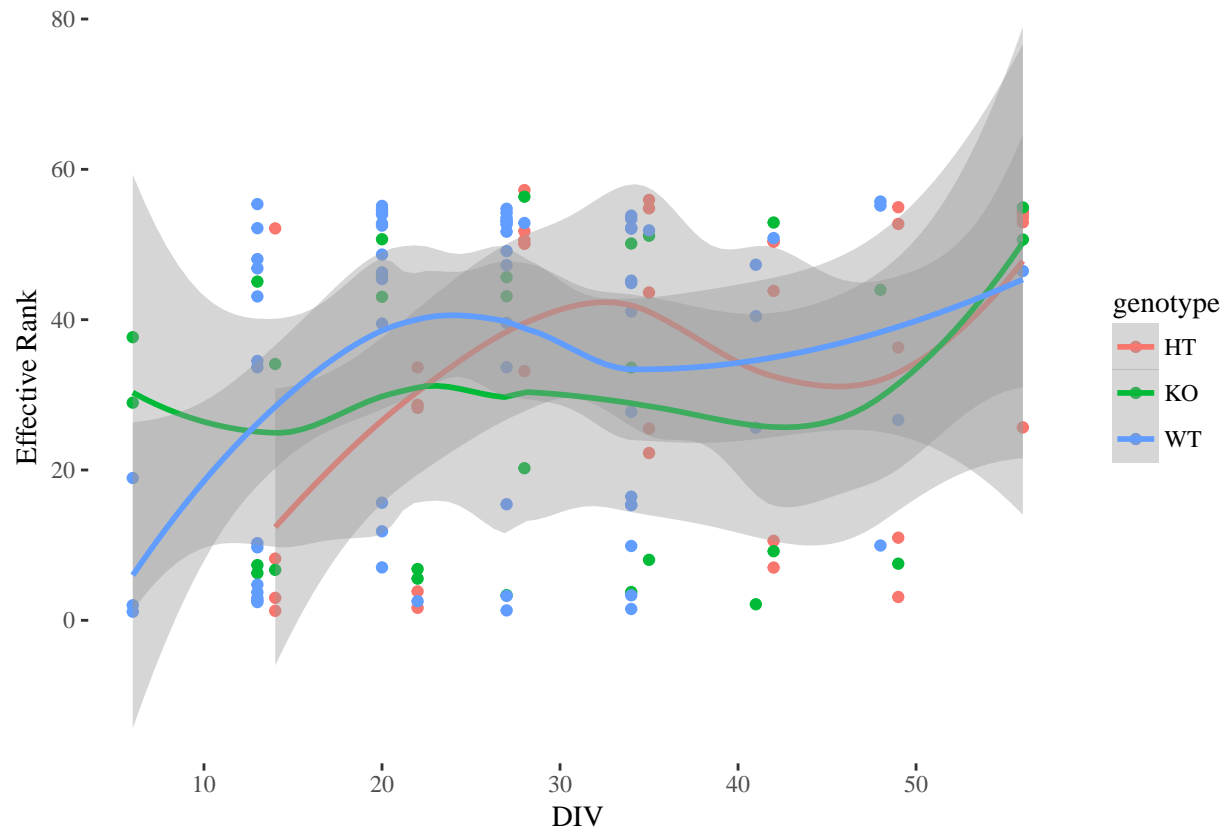
- hard to say, most of regularity values very close to 0 (it can't be 0 because we are taking the log)

Effective Rank

Does effective rank change through development?

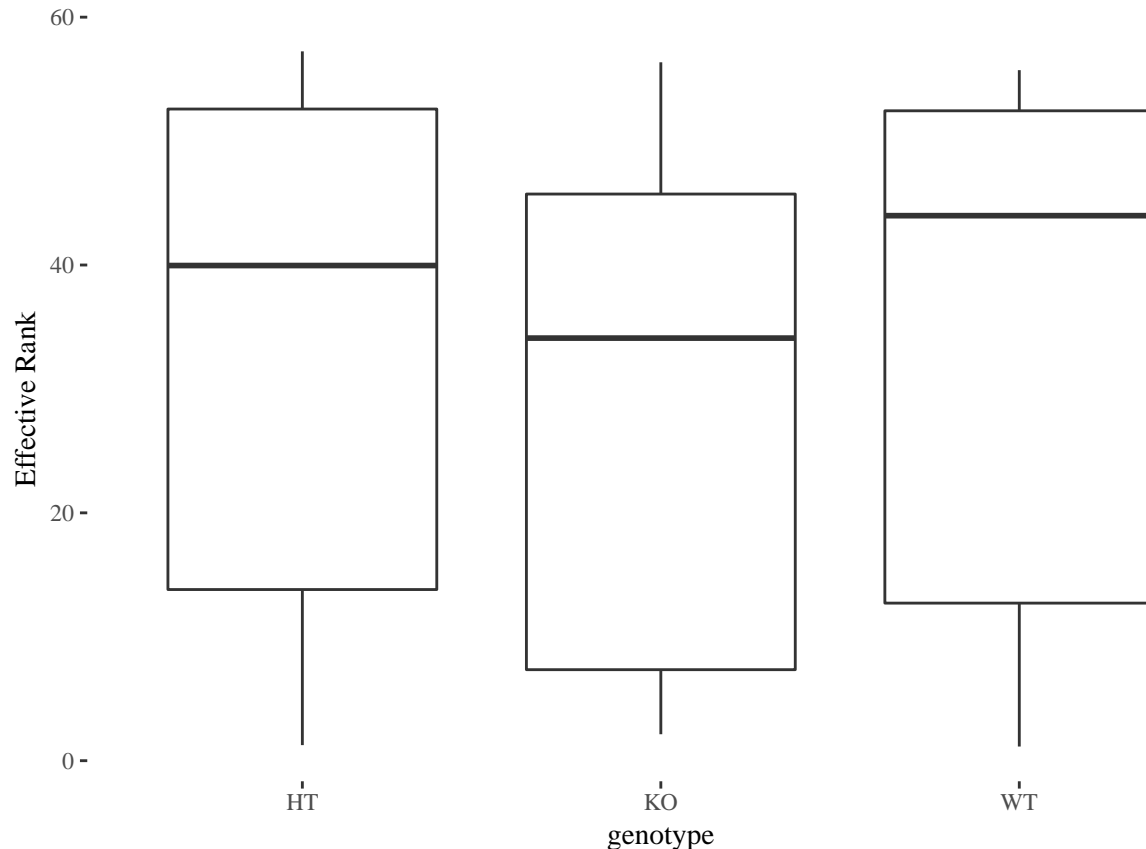
```
ggplot(data = subset(df, is.finite(effRankCov))) + geom_point(mapping = aes(x = DIV,
  y = effRankCov, color = genotype)) + geom_smooth(mapping = aes(x = DIV,
  y = effRankCov, color = genotype)) + ylab("Effective Rank") +
th
```

```
## `geom_smooth()` using method = 'loess'
```



Is there a difference in effective rank between genotypes?

```
ggplot(data = subset(df, is.finite(effRankCov))) + geom_boxplot(mapping = aes(genotype,
  effRankCov)) + ylab("Effective Rank") + th
```



- Although not significant, effective rank seems to be picking up some differences compared to spike and burst statistics
- looking at the 2309 and 1209 combined, it seems that there is some sigmoidal pattern of the effective rank, but the phase (eg. peak time) of that pattern is different between genotypes

What if we look at some of these statistics only in cultures with bursts?

(I think this is what Manuel did in his Rich Club Topology paper; looking only at neural networks with burst, and then only using burst period to do his analysis rather than the whole 12 minutes)

Firing rate

```
p0 <- ggplot(data = subset(df, totalBurstCount > 0)) + geom_point(mapping = aes(x = DIV,
  y = aveFireRate, color = genotype)) + geom_smooth(mapping = aes(x = DIV,
  y = aveFireRate, color = genotype)) + ylab("Average firing rate (Hz)") +
  ggtitle("Average firing rate") + th
```

Burst

```
p1 <- ggplot(data = subset(df, totalBurstCount < 1000 & totalBurstCount >
  0)) + geom_point(mapping = aes(x = DIV, y = totalBurstCount,
  color = genotype)) + geom_smooth(mapping = aes(x = DIV, y = totalBurstCount,
  color = genotype)) + ylab("Total Burst Count") + ggtitle("Total Burst Count") +
  th
```

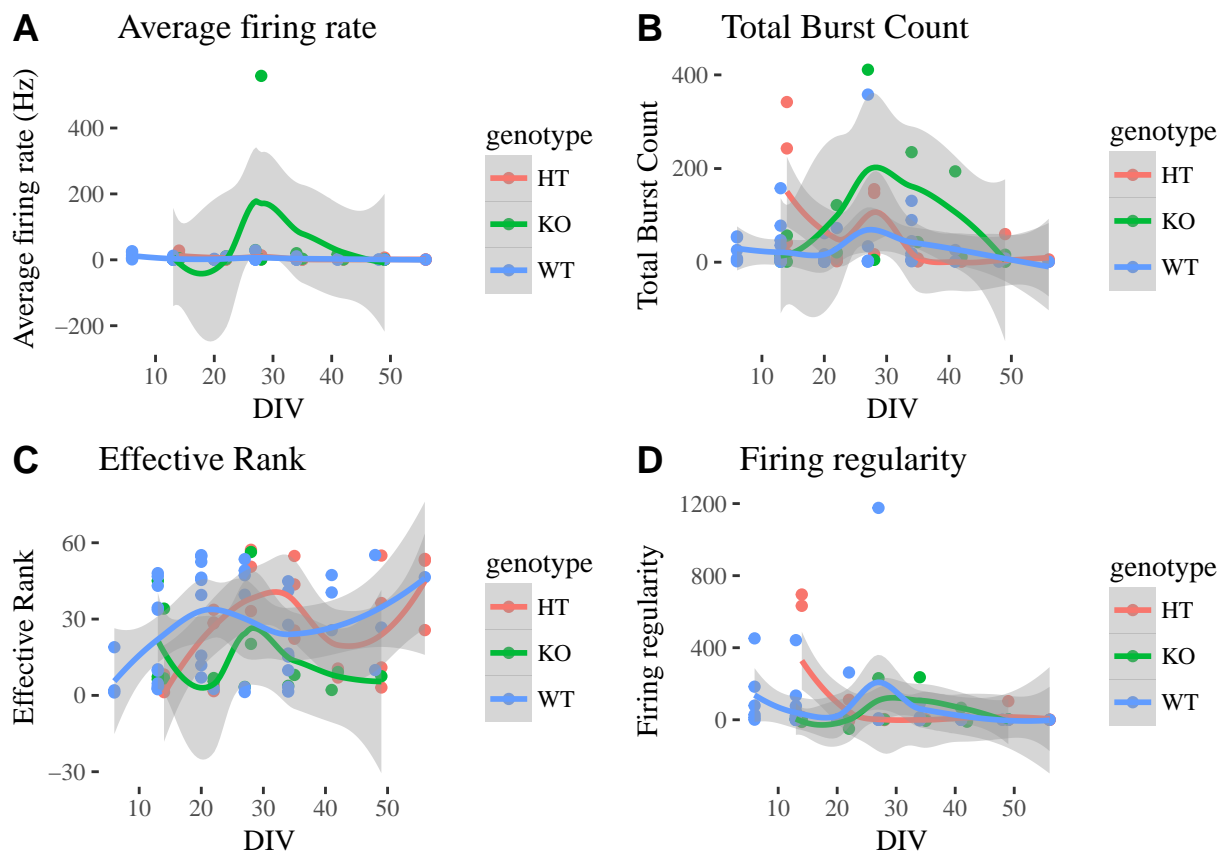
```

# Effective Rank
p2 <- ggplot(data = subset(df, is.finite(effRankCov) & totalBurstCount >
0)) + geom_point(mapping = aes(x = DIV, y = effRankCov, color = genotype)) +
geom_smooth(mapping = aes(x = DIV, y = effRankCov, color = genotype)) +
ylab("Effective Rank") + ggtitle("Effective Rank") + th

# Firing regularity
p3 <- ggplot(data = subset(df, is.finite(totalReg) & totalBurstCount >
0)) + geom_point(mapping = aes(x = DIV, y = totalReg, color = genotype)) +
geom_smooth(mapping = aes(x = DIV, y = totalReg, color = genotype)) +
ylab("Firing regularity") + ggtitle("Firing regularity") +
th

ggarrange(p0, p1, p2, p3, labels = c("A", "B", "C", "D"), ncol = 2,
nrow = 2)

```



Still hard to tell... Will have to look at their distribution.

Moving forward

Data selection

- I have created a spreadsheet to record down which files we are not going to include in our analyses

- Once we have that sorted I can run through the analyses again

Data analyses

- network burst detection
- there are only a few appropriate methods I found for this, the most promising one is Valkki et al 2017
- I contacted the corresponding author of this paper and he says he will email his student to see if she can make the code share-able
- Another way is to use Maneull’s method, but I think we need to email him to clarify what is done as I can’t recreate it just by reading his paper
- calculating similar metrics within burst (effective rank, regularity)
- I am also thinking of quantifying the distribution of spikes across electrodes within individual MEAs, as this is what we were qualitatively trying to do by looking at the heatmaps and spike count distributions
- One way is to do normality (or other probability distribution) tests, eg. KS or SW tests
- it may be interesting to see if there is any relationship between this and effective rank
- I am thinking of calculating more measures that characterise the “controllability” of the system, particular in Gu et al 2015
- but this will require creating a linear model of the network and this may be non-trivial, I am asking Guillaume about it.
- try the feature matrix on machine learning workflow just to make sure it works first

And of course any comments/ideas what we want to do next is welcome :)