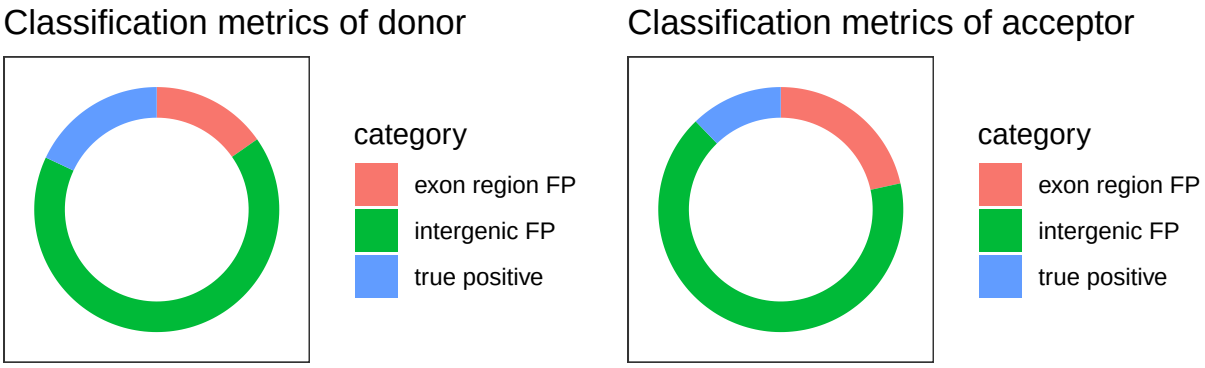
nn-vs-svm

AnhVu

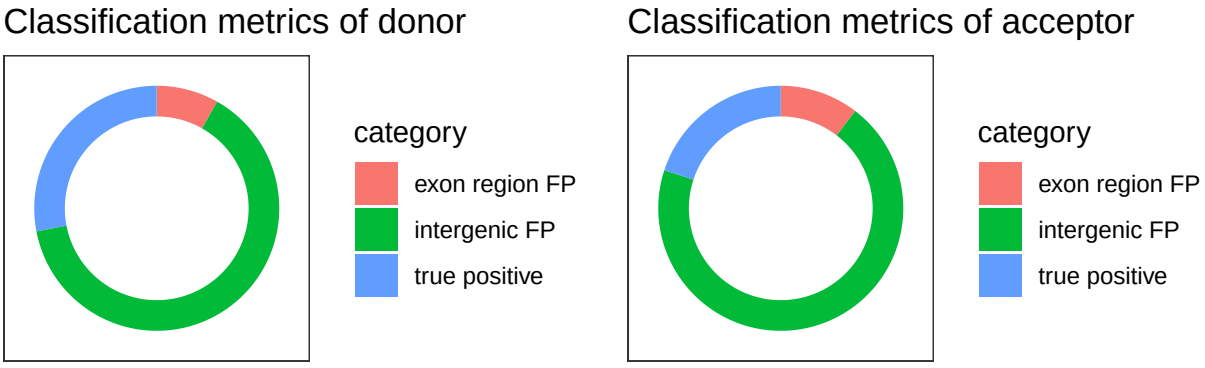
2/18/2020

Comparison of SVM and NN models

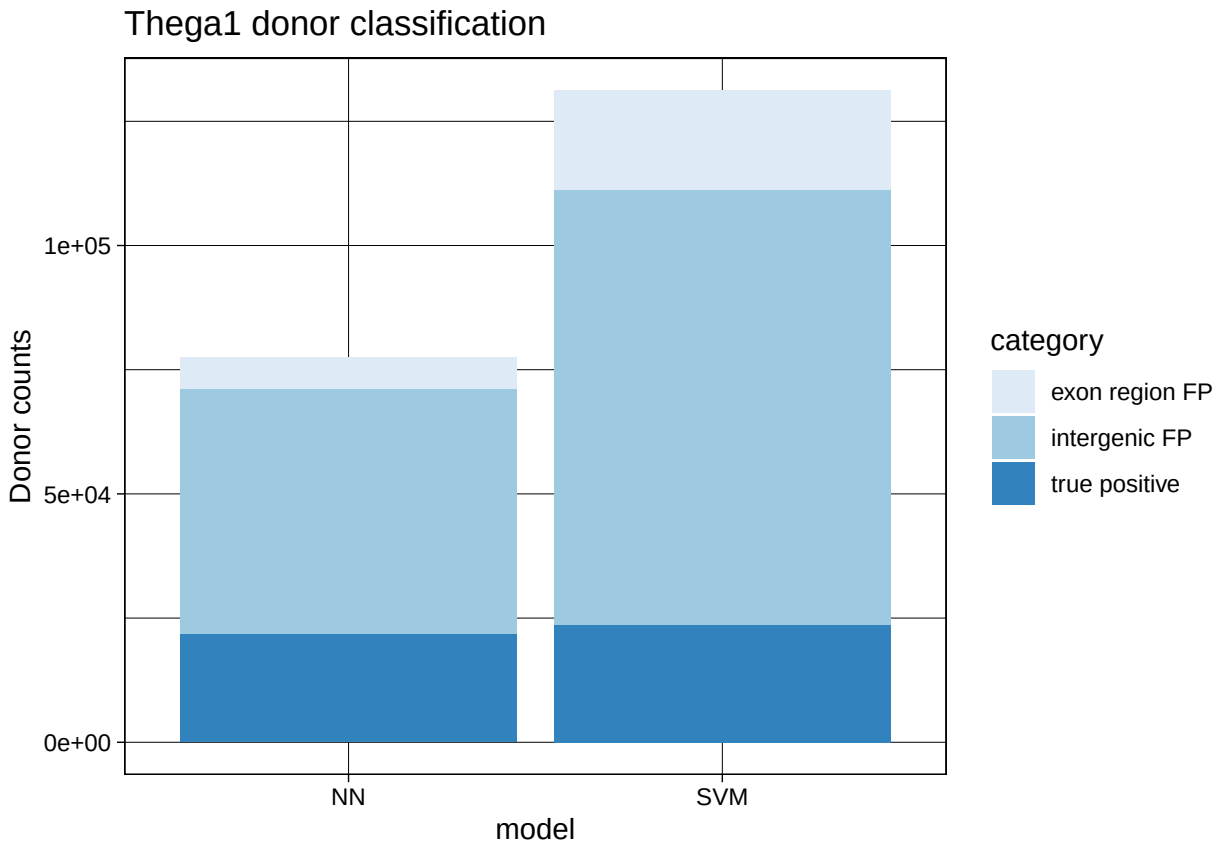
Thega1 splice site classification (SVM models)



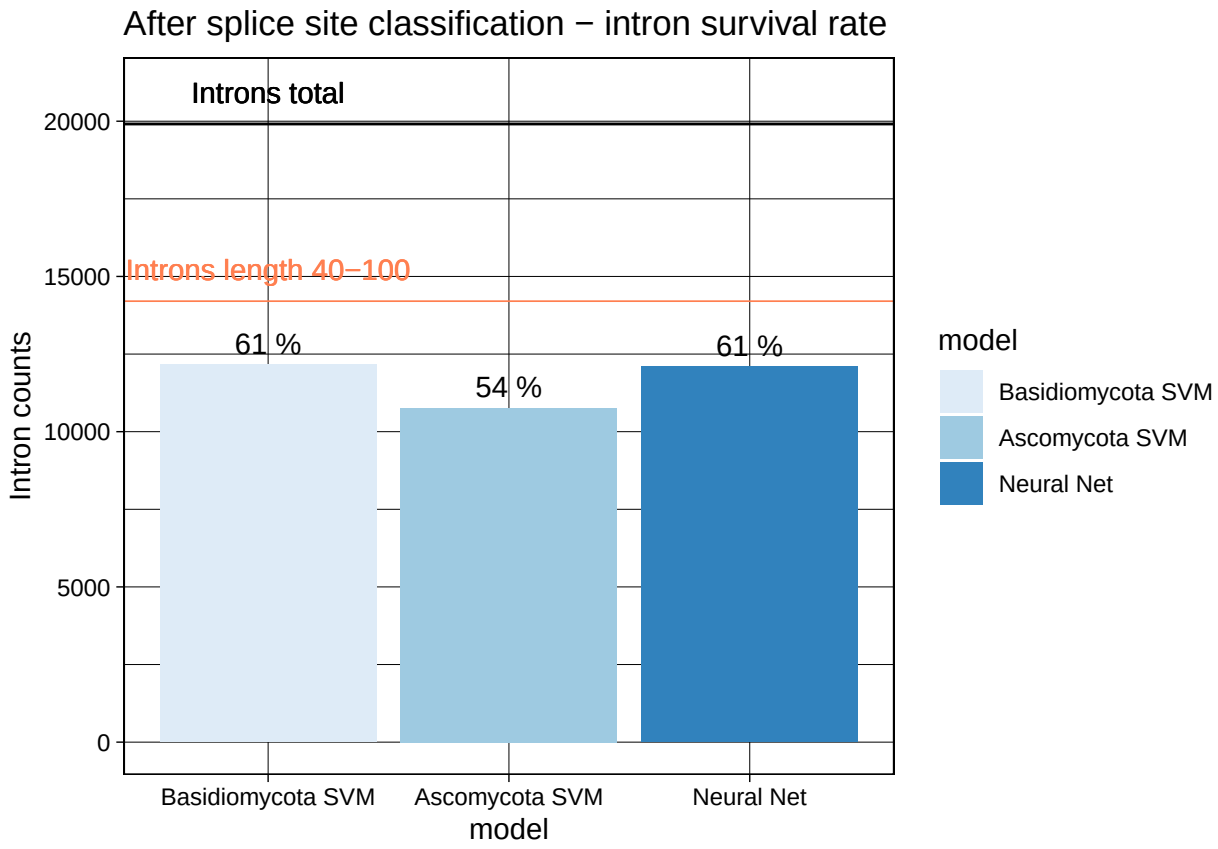
Thega1 splice site classification (NN models)

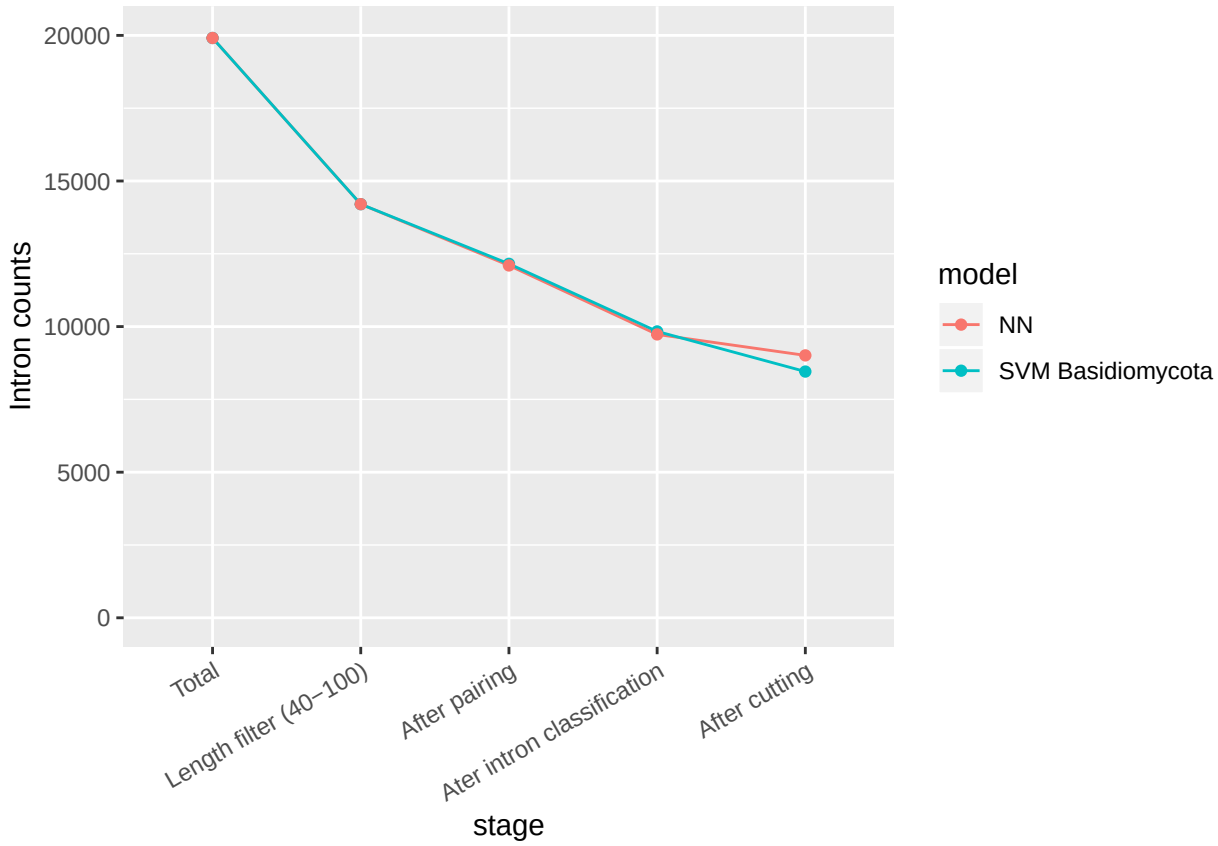


Thega1 splice site classification (comparison)

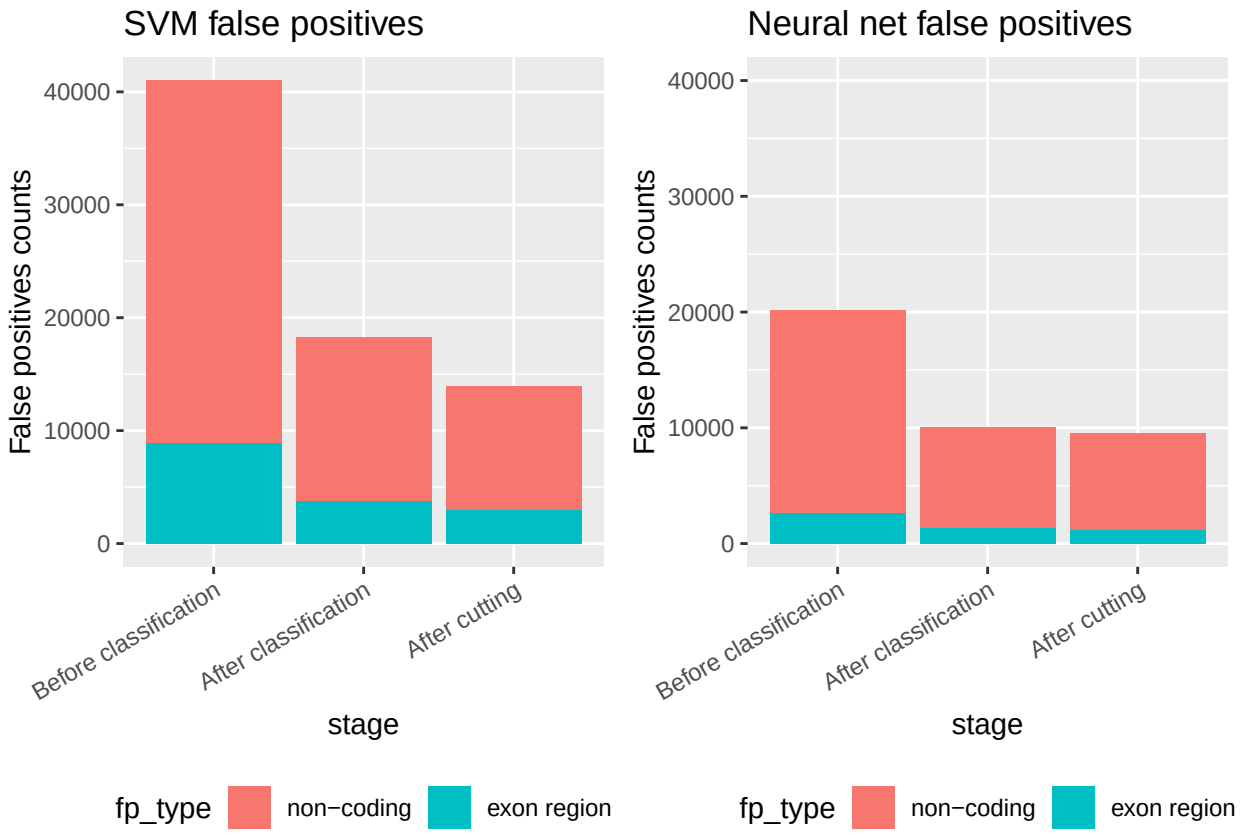


Intron classification (Kocim1 and Thega1)

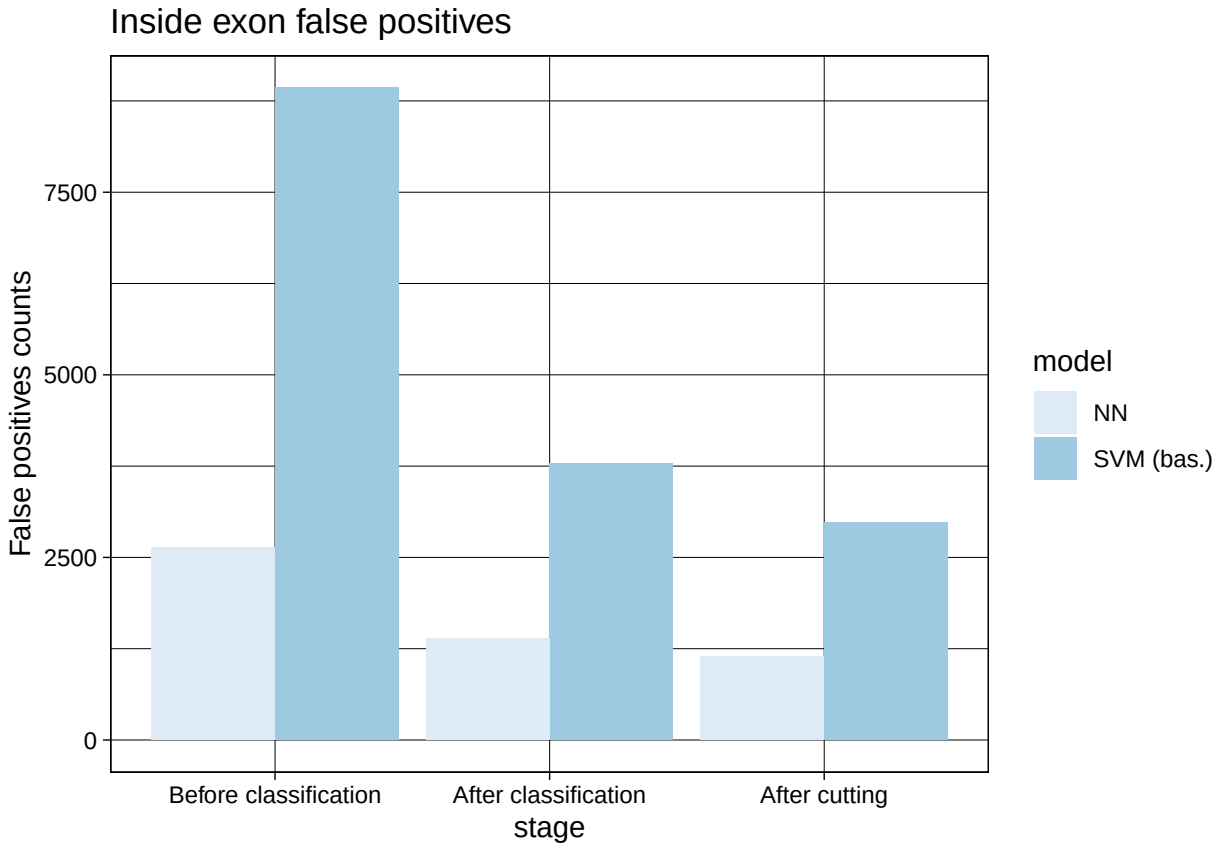





```
kocim.NN.fp.all <- c(20141, 10103, 9495)
kocim.NN.fp.intragen <- c(2640, 1383, 1143)
kocim.NN.fp.intergen <- kocim.NN.fp.all - kocim.NN.fp.intragen

kocim.SVM.fp.all <- c(41001, 18228, 13934)
kocim.SVM.fp.intragen <- c(8927, 3792, 2975)
kocim.SVM.fp.intergen <- kocim.SVM.fp.all - kocim.SVM.fp.intragen

dat_svm <- build_plot_df(kocim.SVM.fp.intergen, kocim.SVM.fp.intragen)

svm.plot <- ggplot(data=dat_svm, aes(x=stage, y=fp_counts, fill=fp_type)) +
    geom_bar(stat="identity") +
    ggtitle("SVM false positives") +
    ylab("False positives counts") +
    theme(legend.direction="horizontal", legend.position="bottom", axis.text.x = element_text(angle = 3
# ------------------------------------------------------------------------------------------------
dat_nn <- build_plot_df(kocim.NN.fp.intergen, kocim.NN.fp.intragen)

nn.plot <- ggplot(data=dat_nn, aes(x=stage, y=fp_counts, fill=fp_type)) +
    geom_bar(stat="identity") +
    ggtitle("Neural net false positives") +
    ylab("False positives counts") +
    ylim(0,40000) +
    theme(legend.direction="horizontal", legend.position="bottom", axis.text.x = element_text(angle = 3

grid.arrange(svm.plot, nn.plot, ncol = 2)
```



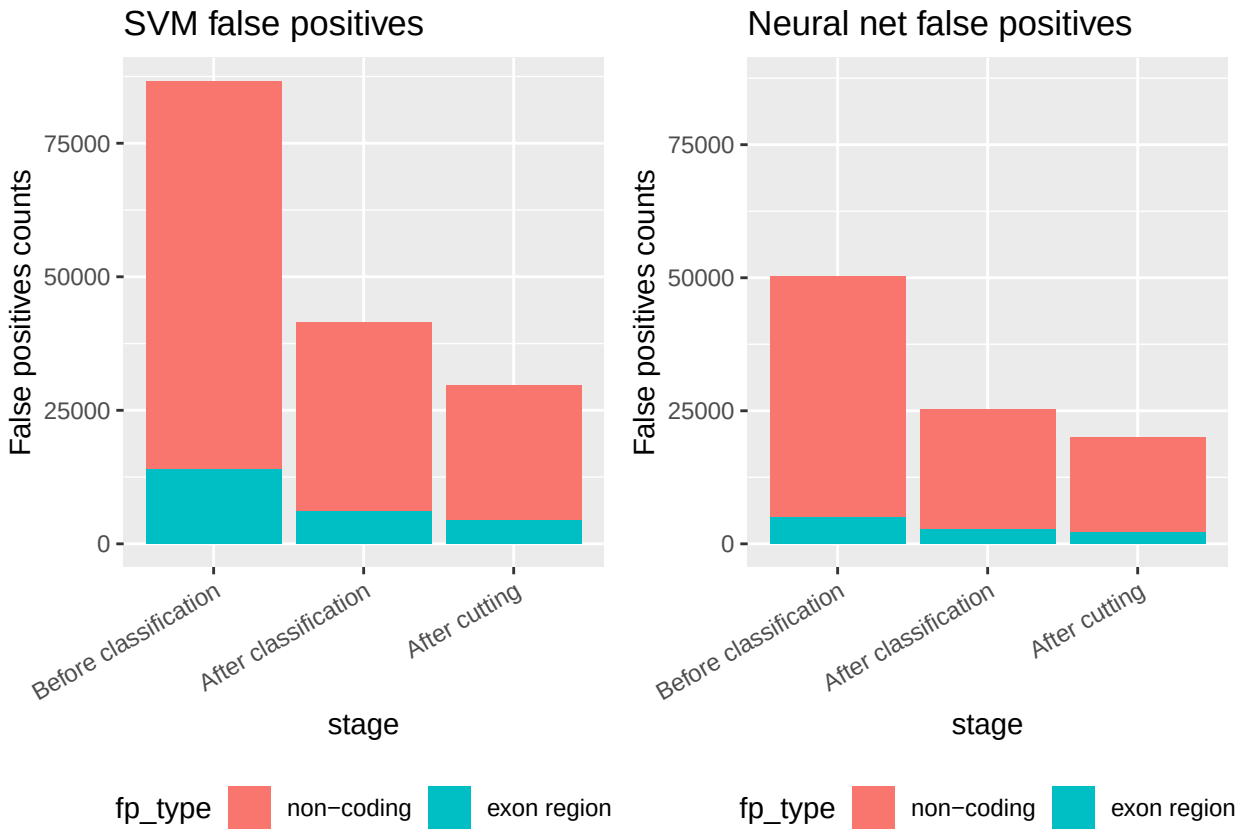

```
dat3 <- data.frame(
  stage=rep(progress.groups, times=2),
  model=rep(c("SVM (bas.)", "NN"), each=3),
  intragen_fp=c(kocim.SVM.fp.intragen, kocim.NN.fp.intragen)
  )

dat3$stage <- factor(dat3$stage, levels = progress.groups)

exon.fp.plot <- ggplot(data=dat3, aes(x=stage, y=intragen_fp, fill=model)) +
    geom_bar(stat="identity", position = position_dodge()) +
    ggtitle("Inside exon false positives") +
    ylab("False positives counts") +
    theme(axis.text.x = element_text(angle = 30, hjust = 1)) +
    scale_fill_brewer() + theme_linedraw()
exon.fp.plot
```



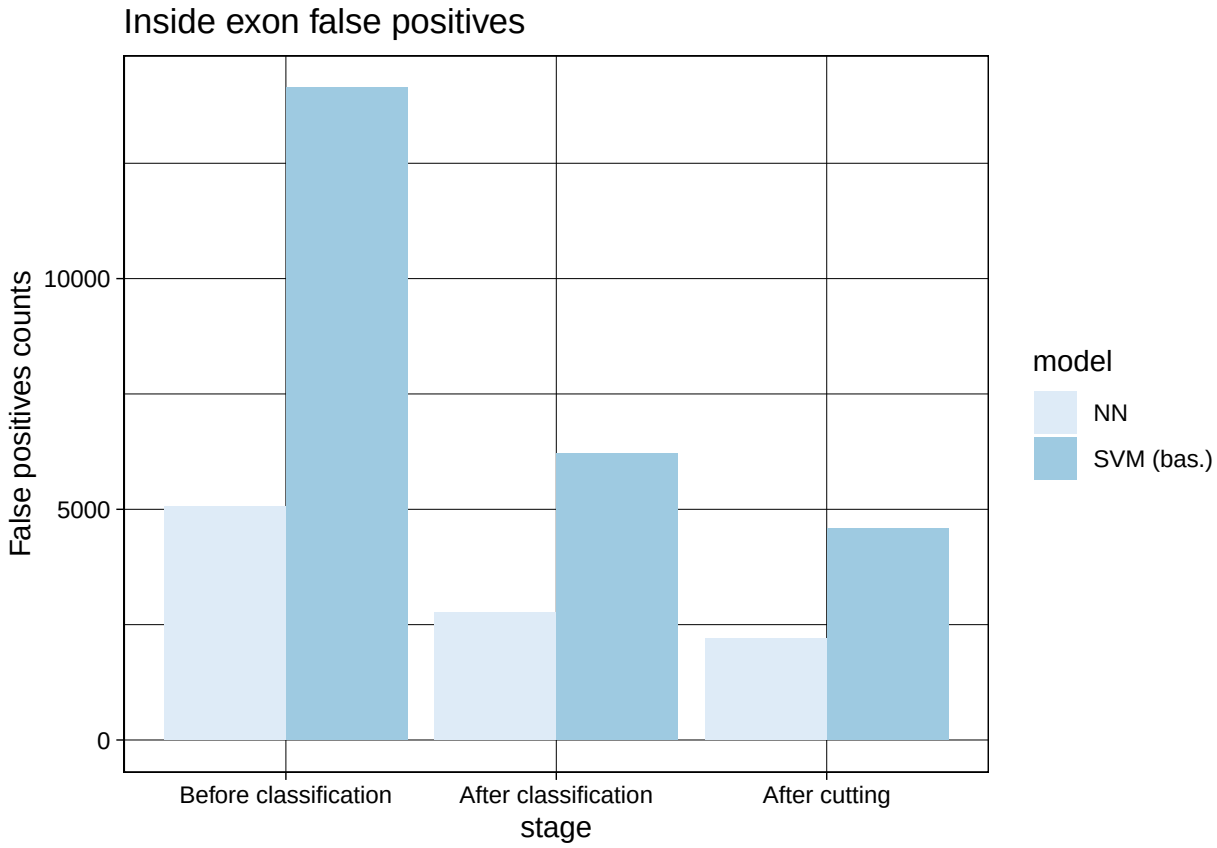
Thega1 experiments

```
thega.NN.fp.all <- c(50331, 25251, 20008)
thega.NN.fp.intragen <- c(5071, 2759, 2195)
thega.NN.fp.intergen <- thega.NN.fp.all - thega.NN.fp.intragen

thega.SVM.fp.all <- c(86685, 41416, 29680)
thega.SVM.fp.intragen <- c(14131, 6204, 4577)
thega.SVM.fp.intergen <- thega.SVM.fp.all - thega.SVM.fp.intragen

dat_svm_thega <- build_plot_df(thega.SVM.fp.intergen, thega.SVM.fp.intragen)

svm.plot <- ggplot(data=dat_svm_thega, aes(x=stage, y=fp_counts, fill=fp_type)) +
    geom_bar(stat="identity") +
    ggtitle("SVM false positives") +
    ylab("False positives counts") +
    theme(legend.direction="horizontal", legend.position="bottom", axis.text.x = element_text(angle = 3
# ------------------------------------------------------------------------------------------------------
dat_nn_thega <- build_plot_df(thega.NN.fp.intergen, thega.NN.fp.intragen)

nn.plot <- ggplot(data=dat_nn_thega, aes(x=stage, y=fp_counts, fill=fp_type)) +
    geom_bar(stat="identity") +
    ggtitle("Neural net false positives") +
    ylab("False positives counts") +
    ylim(0,87000) +
```

```
    theme(legend.direction="horizontal", legend.position="bottom", axis.text.x = element_text(angle = 30
grid.arrange(svm.plot, nn.plot, ncol = 2)
```




```
dat3 <- data.frame(
  stage=rep(progress.groups, times=2),
  model=rep(c("SVM (bas.)", "NN"), each=3),
  intragen_fp=c(thega.SVM.fp.intragen, thega.NN.fp.intragen)
  )

dat3$stage <- factor(dat3$stage, levels = progress.groups)

exon.fp.plot <- ggplot(data=dat3, aes(x=stage, y=intragen_fp, fill=model)) +
    geom_bar(stat="identity", position = position_dodge()) +
    ggtitle("Inside exon false positives") +
    ylab("False positives counts") +
    theme(axis.text.x = element_text(angle = 30, hjust = 1)) +
    scale_fill_brewer() + theme_linedraw()
exon.fp.plot
```


Inside exon false positives
False positives counts
10000
5000
0
Before classification
After classification
After cutting
stage
model
NN
SVM (bas.)