

Melanoma BCR repertoire analysis

Anna Obratsova

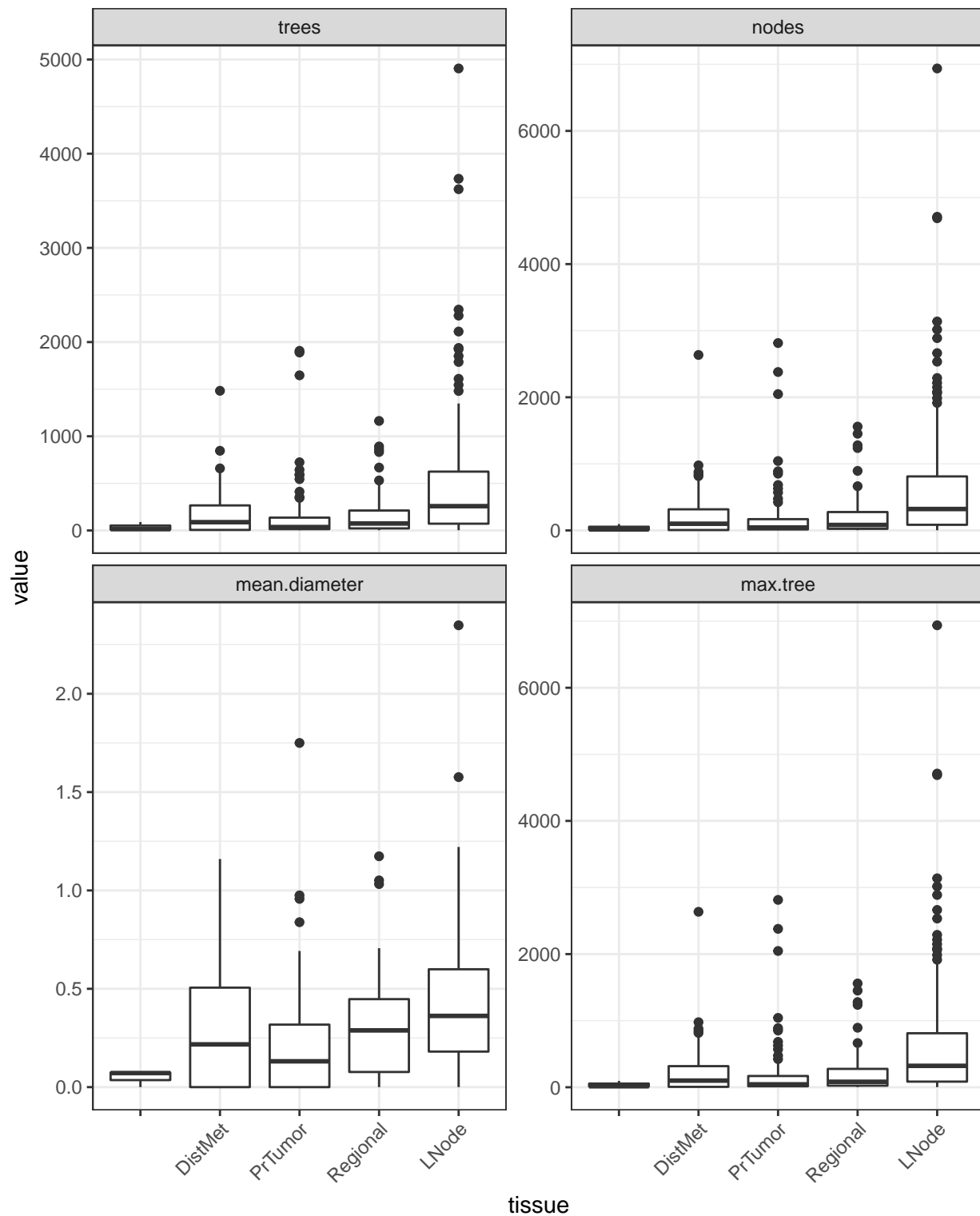
1 June 2017

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(plyr)

load('clones.rda')
codes <- read.table('SKCM-codes-samples-for-patients-tss.txt', header=F, sep="\t")
colnames(codes) <- c('patient_id', 'c1', 'c2', 'c3', 'tissue')
clones <- merge(clones, codes)
clones$tissue <- revalue(clones$tissue, c("Distant Metastasis" = "DistMet",
                                          "Primary Tumor" = "PrTumor",
                                          "Regional Cutaneous or Subcutaneous Tissue (includes satellite)" = "LNNode"))
clones$LN <- ifelse(clones$tissue == "LNNode", 'LNNode', 'Other')

c <- clones %>% group_by(patient_id, tissue) %>%
  dplyr::summarise(trees = n(), nodes = sum(nodes), mean.diameter = mean(diameter),
                  max.tree = max(nodes)) %>%
  dplyr::select(tissue, patient_id, trees, nodes, mean.diameter, max.tree) %>%
  melt(id.vars = c('tissue', 'patient_id'))

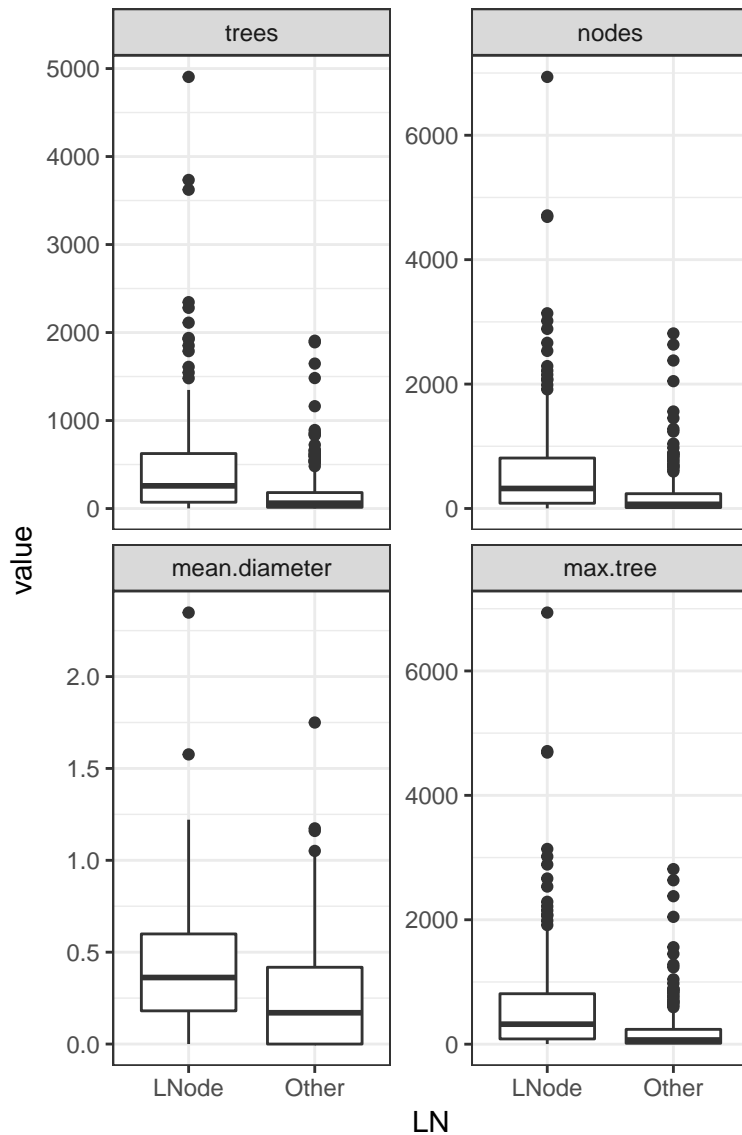
ggplot(c, aes(x = tissue, y = value)) +
  geom_boxplot() +
  facet_wrap(~variable, scales="free_y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
c1 <- clones %>% group_by(patient_id, LN) %>%
  dplyr::summarise(trees = n(), nodes = sum(nodes), mean.diameter = mean(diameter),
                  max.tree = max(nodes)) %>%
  dplyr::select(LN, patient_id, trees, nodes, mean.diameter, max.tree)

ggplot(melt(c1, id.vars = c('LN', 'patient_id')), aes(x = LN, y = value)) +
```

```
geom_boxplot() +
facet_wrap(~variable, scales="free_y") +
theme_bw()
```



```
for (i in c('trees', 'nodes', 'mean.diameter', 'max.tree')){
  cat(i, 't-Test p-value\n', t.test(c1[[i]] ~ c1[['LN']])$p.value, '\n')
}
```

```
## trees t-Test p-value
## 1.686637e-09
## nodes t-Test p-value
## 6.496936e-09
## mean.diameter t-Test p-value
## 1.282486e-06
## max.tree t-Test p-value
## 6.496936e-09
```

```
library(tidyr)
```

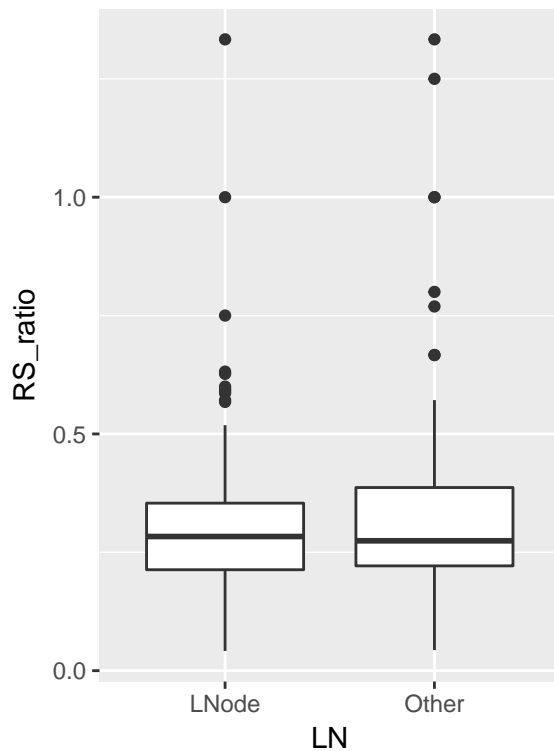
```
##
## Attaching package: 'tidyr'
## The following object is masked from 'package:reshape2':
##
## smiths
library(reshape2)

load('shm.rda')

s1 <- s %>% group_by(patient_id, LN, mut.type) %>%
  dplyr::summarise(n = n()) %>%
  dcast(patient_id + LN ~ mut.type, value.var = "n") %>%
  mutate(RS_ratio = R/S)

ggplot(s1, aes(x = LN, y = RS_ratio)) + geom_boxplot()

## Warning: Removed 54 rows containing non-finite values (stat_boxplot).
```

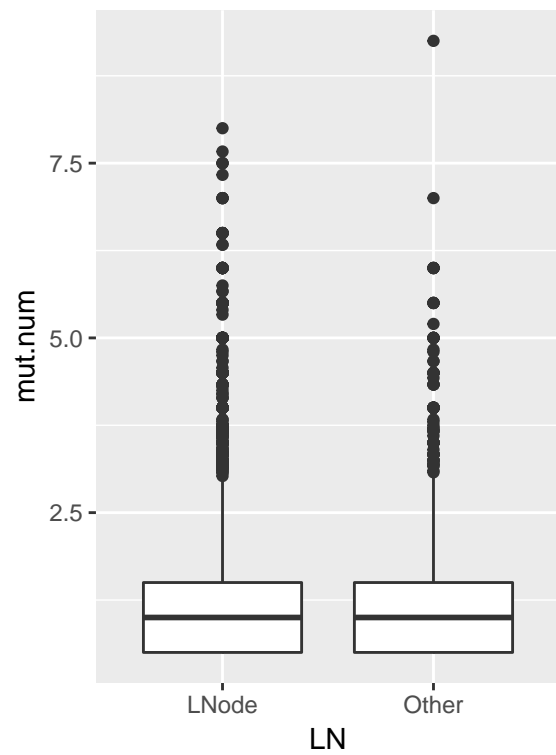


```
print(t.test(s1$RS_ratio ~ s1$LN)$p.value)

## [1] 0.2063957

s2 <- s %>% group_by(patient_id, LN, tree.id, child) %>%
  dplyr::summarise(mut.num = n()) %>%
  group_by(patient_id, LN, tree.id) %>%
  dplyr::summarise(mut.num = sum(mut.num)/(n()+1))

ggplot(s2, aes(x = LN, y = mut.num)) + geom_boxplot()
```



```
print(t.test(s2$mut.num ~ s2$LN)$p.value)
```

```
## [1] 0.1951248
```