

# P-väärtuste uurimine

```
path = "/gpfs/hpchome/a72094/rocket/projects/chromatin_to_splicing/"
```

Kõik kromatiini avatuse külmud, millel on  $FDR < 0.1$  QTL

```
ca = read.table(paste(path, "tabix/cqn_permutations_100000_significant.sorted.txt.gz", sep=""))
```

Kõik Yurii Toma CTCF seondumiskohad

```
ctcf = read.table("/gpfs/hpchome/a72094/rocket/projects/chromatin-QTLs/Blood_ATAC/processed/CTCF/ctltools/ctcf_bindingsites.txt")
bindingsites <- GRanges(seqnames = ctcf$V2, strand = c("*"), ranges = IRanges(start = ctcf$V3, end = ctcf$V4))
```

Kromatiini avatuse külmud, mille QTL paikneb CTCF seondumiskohas

```
caQTL = GRanges(seqnames = ca$V9, strand = c("*"), ranges = IRanges(start = ca$V10, end = ca$V11, names = ca$V12))
caQTL_in_bindingsite = subsetByOverlaps(caQTL, bindingsites)
ca_in_bindingsite = ca[which(ca$V8 %in% names(caQTL_in_bindingsite)), ]
```

Lookused, mis on seotud kromatiini avatuse ja CTCF seondumisega

```
# V4 on ctcf QTLd
r2 = read.table(paste(path, "results/ctcf/cqn_rsq08.txt", sep=""))
# kontroll
ctcfQTL = read.table(paste(path, "CTCF/CTCF.permuted.significant.sorted.txt.gz", sep=""))
#ctcfQTL[which(ctcfQTL$V8 %in% r2$V4),]
```

Kromatiini avatuse külmud, mille QTL seotud CTCF QTLga

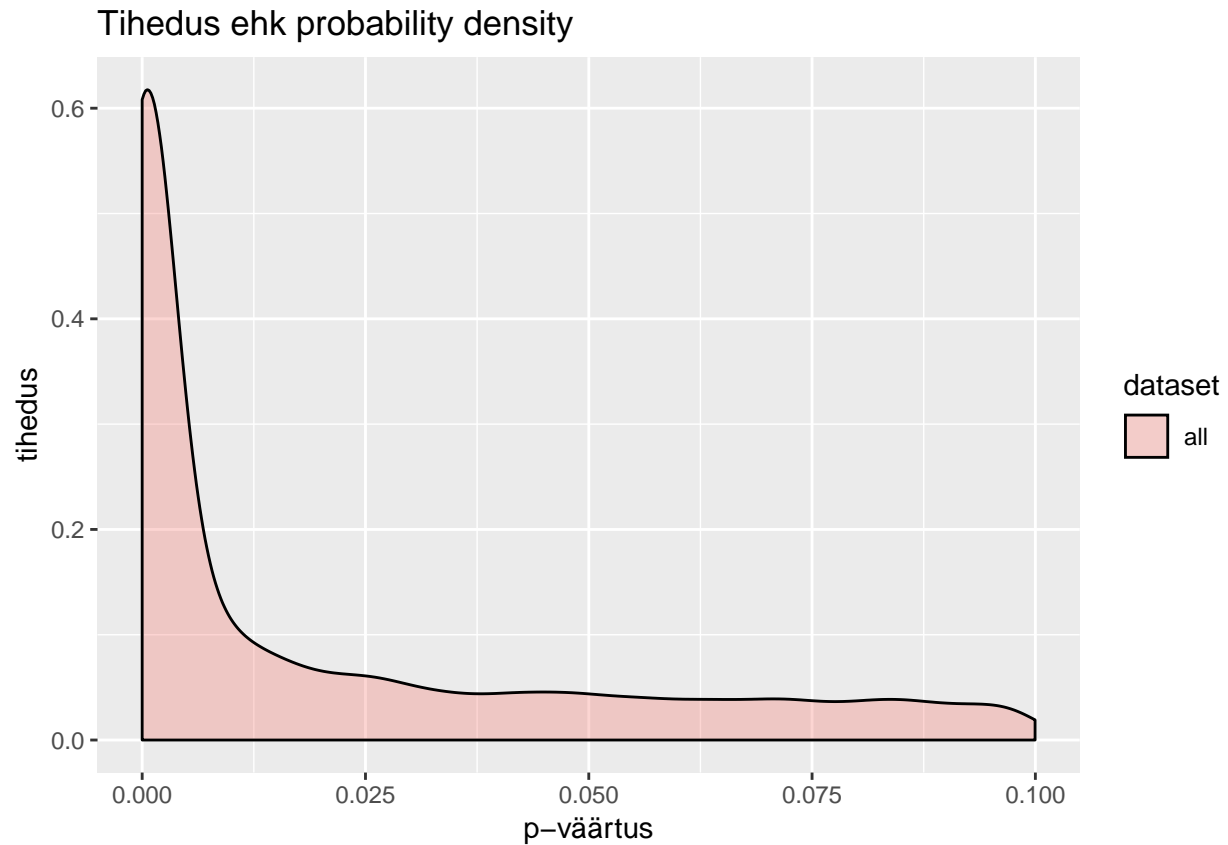
```
ca_r2 = ca[which(ca$V8 %in% r2$V1), ]
```

Ühendan kolme tüüpi kromatiini avatuse külmude p-väärtused. P-väärtus näitab kui tugevalt on seotud genotüüp lookuses ja kromatiini avatus ühes külmus

```
ca_values<-ca[, "V19"]
ca_in_bindingsite_values<-ca_in_bindingsite[, "V19"]
ca_r2_values<-ca_r2[, "V19"]
all<-data.frame(dataset=c(rep('all', length(ca_values)), rep('withinpeak', length(ca_in_bindingsite_values))),
  value=c(ca_values, ca_in_bindingsite_values, ca_r2_values))
```

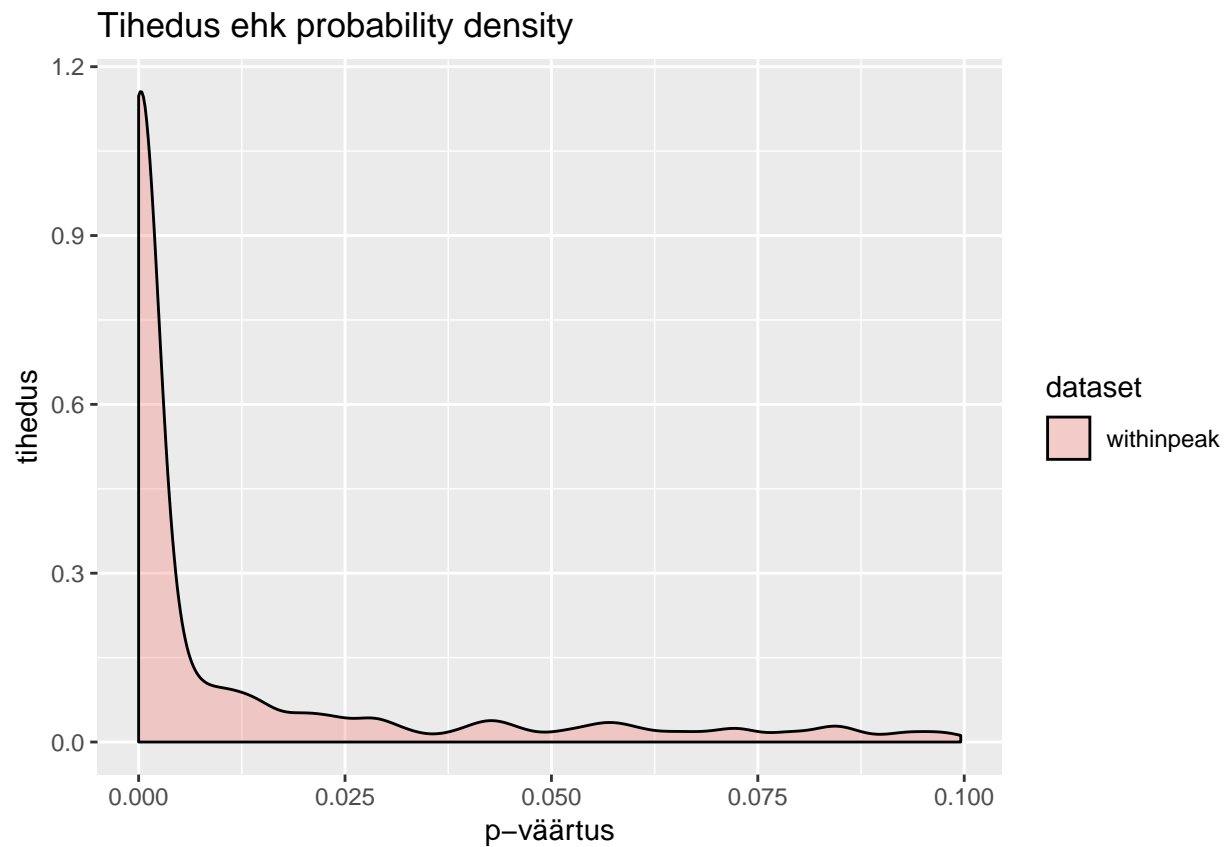
Kõik kromatiini külmud

```
ggplot(all[which(all$dataset=='all'),], aes(x=value, fill=dataset)) +
  geom_density(aes(y=0.01*..density..), alpha=0.3) +
  labs(x='p-väärtus', y='tihedus', title='Tihedus ehk probability density')
```



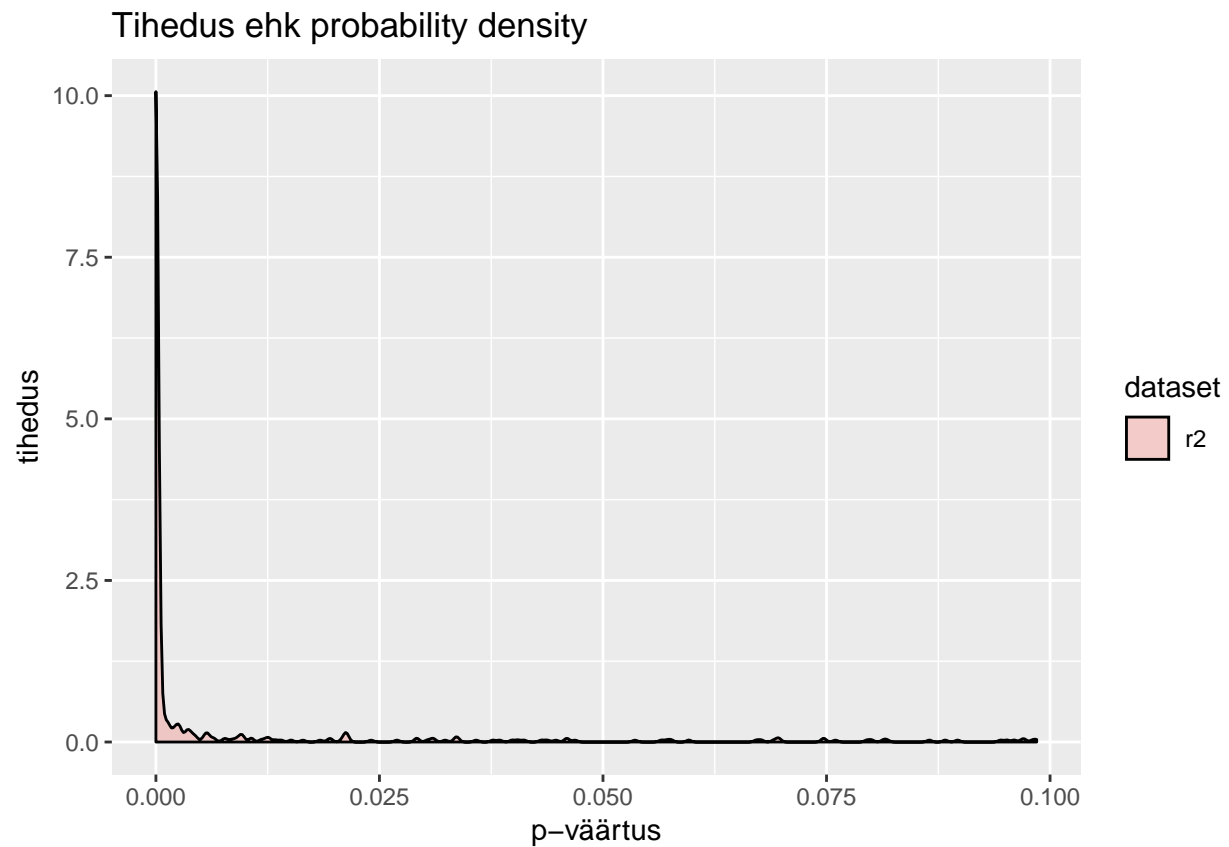
Kühmud, mille QTL paikneb CTCF seondumiskohas. (Äkki peaks vaatama hoopis kühme, mis ise kattuvad CTCF seondumiskohaga?? Raport vist räägib pigem QTL ja seondumiskoha kattumise poolt)

```
ggplot(all[which(all$dataset=='withinpeak'),],aes(x=value,fill=dataset))+  
  geom_density(aes(y=0.01*..density..), alpha=0.3) +  
  labs(x='p-väärtus', y='tihedus', title='Tihedus ehk probability density')
```

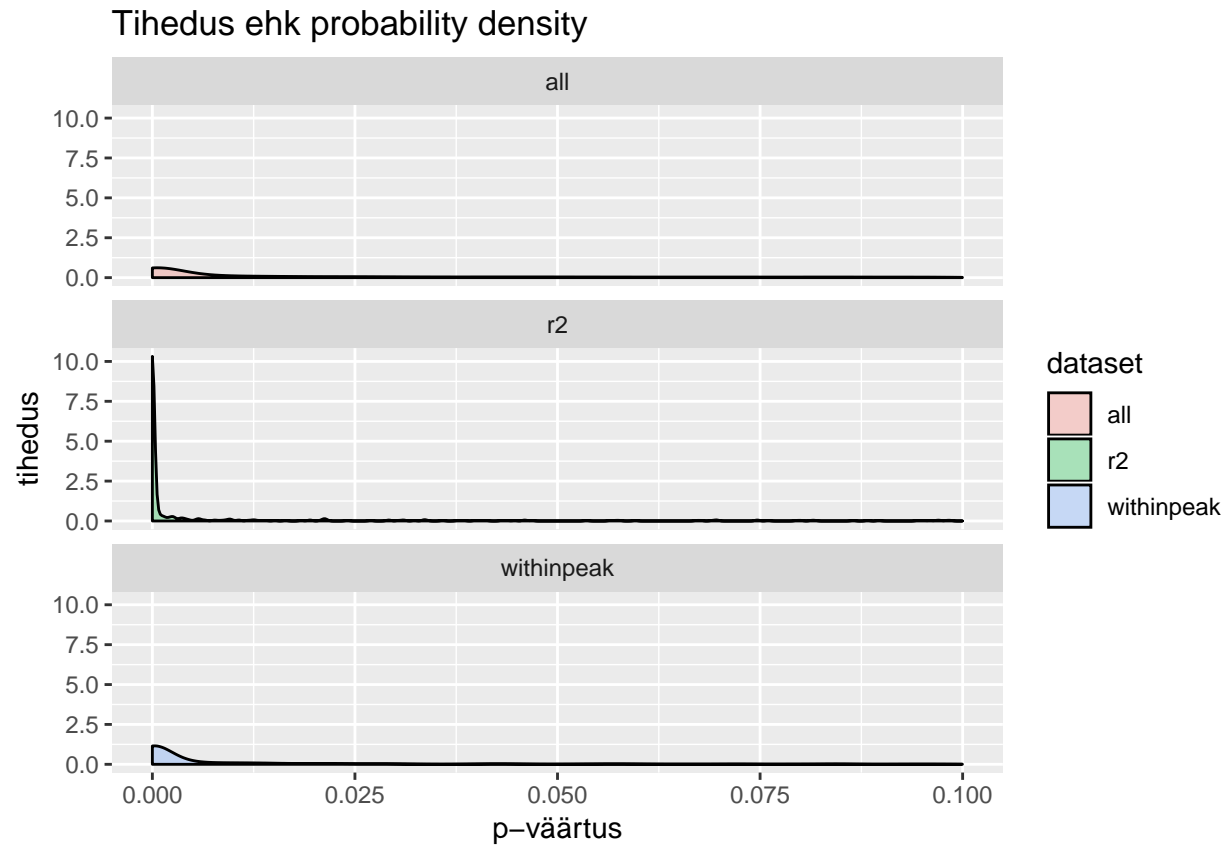


Kühmud, mille QTL on seotud CTCF QTLga

```
ggplot(all[which(all$dataset=='r2'),], aes(x=value, fill=dataset)) +  
  geom_density(aes(y=0.01*..density..), alpha=0.3) +  
  labs(x='p-väärtus', y='tihedus', title='Tihedus ehk probability density')
```

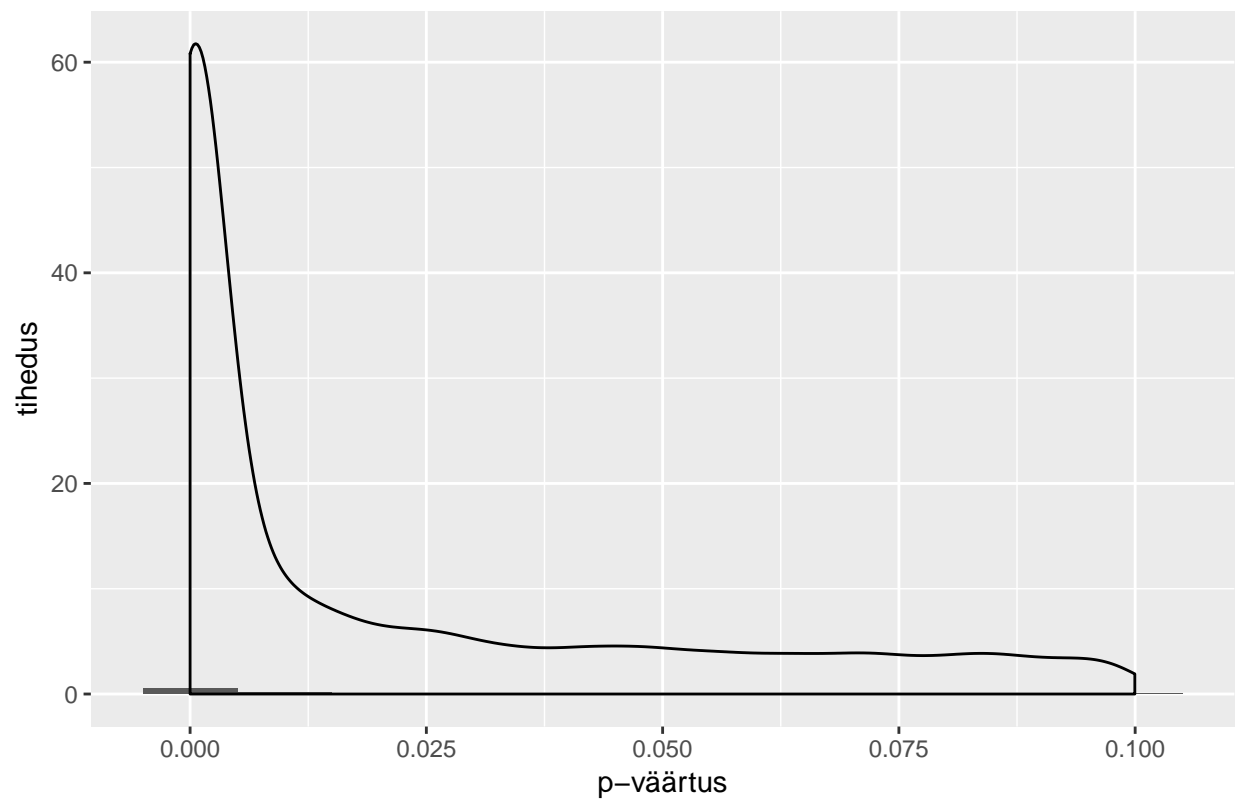


```
ggplot(all,aes(x=value,fill=dataset))+  
  geom_density(aes(y=0.01*..density..), alpha=0.3) +  
  facet_wrap(~dataset,nrow=3) +  
  labs(x='p-väärtus', y='tihedus', title='Tihedus ehk probability density')
```



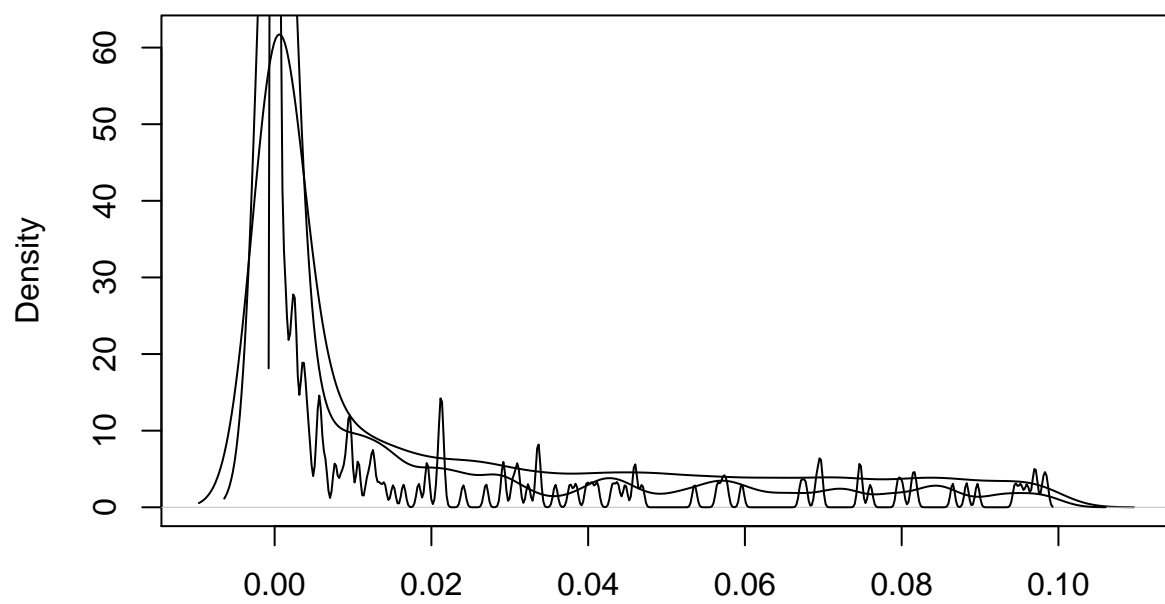
```
ggplot(all[which(all$dataset=='all'),], aes(x=value))+
  geom_histogram(aes(y=0.01*..density..), binwidth = 0.01)+
  geom_density() +
  labs(x='p-väärtus', y='tihedus', title='Tihedus ehk probability density')
```

### Tihedus ehk probability density



```
plot(density(ca_values))  
lines(density(ca_in_bindingsite_values))  
lines(density(ca_r2_values))
```

**density.default(x = ca\_values)**



N = 25872 Bandwidth = 0.003223