eQTL power calculations

Matthias Heinig

19 5 2020

Power calculcations

Effect sizes of whole blood trans-eQTL from eQTLgen

Effect sizes of the meta analysis are given as z-scores for the slope estimate.

The original QTL analysis start with a linear regression model:

$$Y = \alpha + \beta x + \epsilon$$

where x is the genotype dosage at each SNP.

From the z-scores we would like to get some effectsizes β. The z-score is defined as

$$z = \beta / se(\beta).$$

The variance of β is:

$$var(\beta) = \frac{var(\epsilon)}{N var(x)}$$

The variance of the SNP dosage is given by its allele frequency p.

$$var(x) = 2p(1-p)$$

This gives:

$$var(\beta) = \frac{var(\epsilon)}{N 2p(1-p)}$$

To compute the variance of the residulas we write the variance of Y in terms of the regression

$$var(Y) = \beta^2 var(x) + var(\epsilon)$$

and assume that $var(Y) = 1$. So that

$$var(\epsilon) = 1 - \beta^2 2p(1-p)$$

and therefore

$$var(\beta) = \frac{1 - \beta^2 2p(1-p)}{N 2p(1-p)}.$$

Plugging in $\beta^2 = z^2 var(\beta)$ and solving yields:

$$var(\beta) = \frac{1}{2p(1-p)(N+z^2)}$$

And therefore

$$\beta = \frac{z}{\sqrt{2p(1-p)(N+z^2)}}$$

With this we can now also compute the proportion of variance explained

$$R^2 = 1 - var(\epsilon) = \beta^2 2p(1-p) = \frac{z^2}{N+z^2}$$

Now we investigate whether the assumption $var(Y) = 1$ is actually required. When we do the calculations with $var(Y)$, we get:

$$var(\beta) = \frac{var(y)}{var(x)(N + z^2)}$$

$$\beta^2 = z^2 var(\beta) = \frac{z^2 var(y)}{(N + z^2)var(x)}$$

$$R^2 = \frac{\beta^2 var(x)}{var(y)} = \frac{z^2}{N + z^2}$$

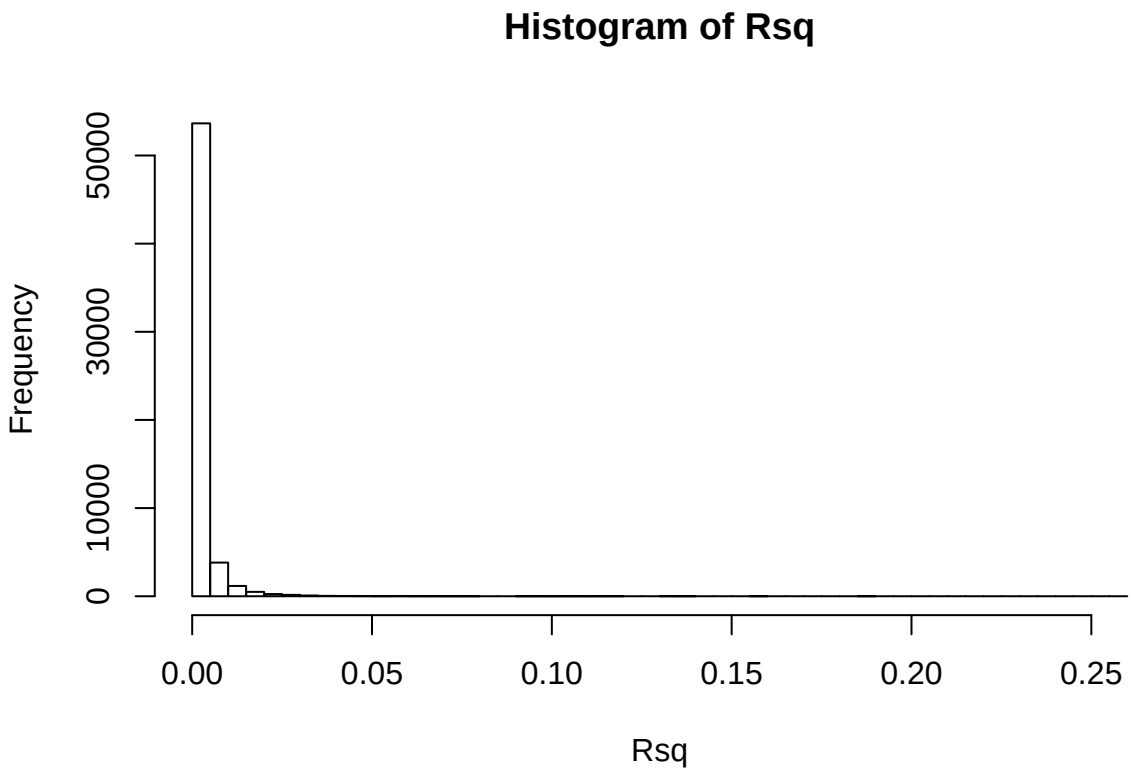
From this result we can conclude that we do not need to standardize to var(y) = 1 for the power calculation.

So we can compute the R^2 for the Vosa et al trans eQTL.

```
efile <- "2018-09-04-trans-eQTLsFDR0.05-CohortInfoRemoved-BonferroniAdded.txt.gz"
if (!file.exists(efile)) {
  download.file(
    paste0("https://molgenis26.gcc.rug.nl/downloads/eqtlgen/trans-eqtl/",
           "2018-09-04-trans-eQTLsFDR0.05-CohortInfoRemoved-BonferroniAdded.txt.gz"),
    dest=efile
    )
}
trans.eqtl <- read.csv(efile, sep="\t", stringsAsFactors=FALSE)
Rsq <- trans.eqtl$Zscore^2 / (trans.eqtl$Zscore^2 + trans.eqtl$NrSamples)

trans.eqtl <- cbind(trans.eqtl, Rsq)
```

Plot the R^2 distribution

```
smry <- with(trans.eqtl, {
  hist(Rsq, breaks=50)
  summary(Rsq)
  })
```




```
smry
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0006267 0.0008121 0.0011051 0.0024671 0.0020684 0.2571029
```

Power to detect trans eQTL

Here we are intersted to assess the power of a targeted trans only for AF SNPs ('n.snps.per.gene.trans' in the code) with a limited sample size ('Nind.trans' in the code). The approach is to get effect sizes from the Vosa trans eQTLs and estimate how many tests we can afford to still be able to detect these effects in our sample size. The number of tests performed will influence the alpha level. We will use Bonferroni adjustment for the sake of simplicity.

The significance threshold used for trans-eQTL:

```
n.snps.per.gene.trans <- 109
n.genes <- 1:500
alpha.trans <- 0.05 / (n.genes * n.snps.per.gene.trans)
summary(alpha.trans)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 9.174e-07 1.222e-06 1.831e-06 6.232e-06 3.648e-06 4.587e-04
```

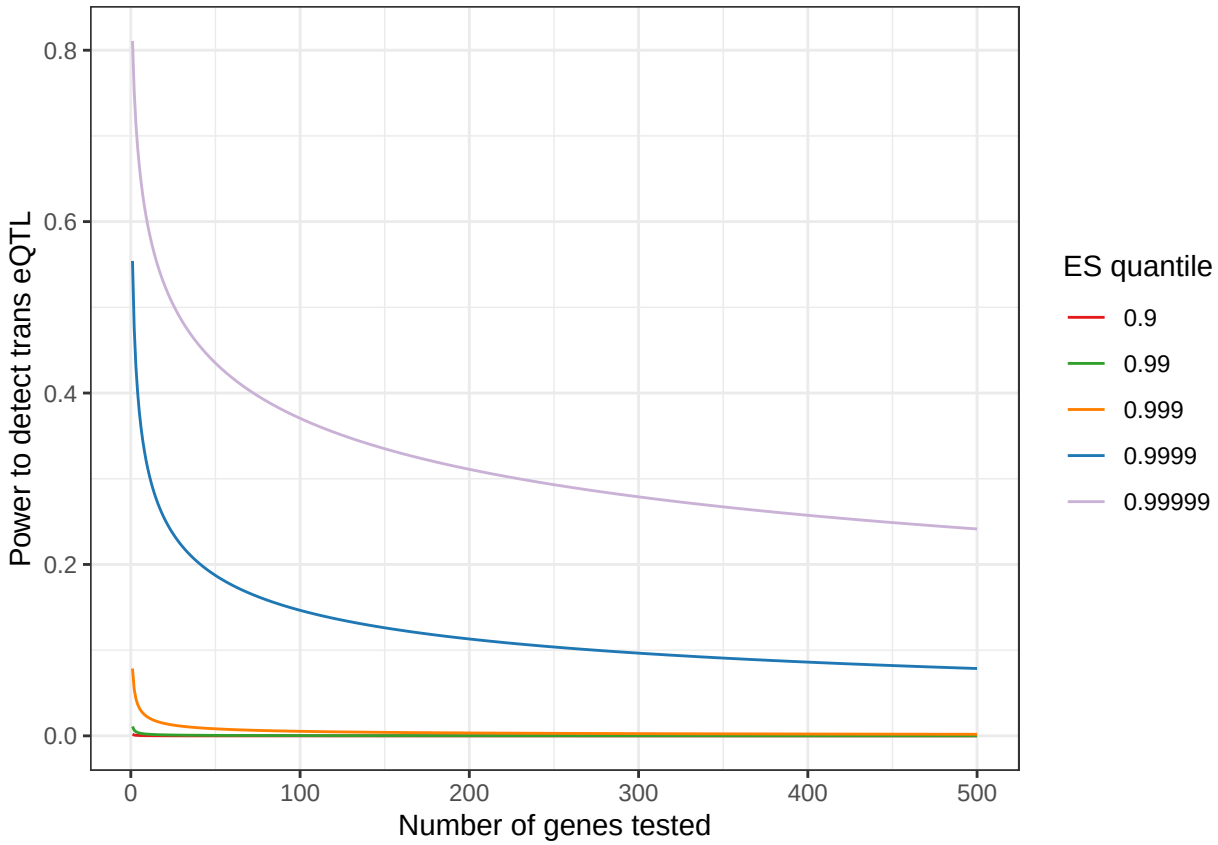
Select an effect size from the quantiles of the trans eQTL effect sizes of Vosa et al.

```
es.quantile <- 1-10^-(1:5)
Rsq <- quantile(trans.eqtl$Rsq, p=es.quantile)
Rsq
```

```
##         90%         99%       99.9%      99.99%     99.999%
## 0.005097604 0.020834419 0.059548736 0.161386799 0.218361197
```

Compute the power for the F test

```
Nind.trans <- 75

plot.data.trans <- NULL

for (i in 1:length(Rsq)) {
  f2 <-  Rsq[i] / (1 - Rsq[i])
  power.f2.per.test.trans <- sapply(alpha.trans, function(alpha) {
    df.num <- 1 ## dfs of the full model
    df.denom <- Nind.trans - df.num - 1 ## error dfs
    pwr.f2.test(u=df.num, v=df.denom, f2=f2, sig.level=alpha)$power
  })

  plot.data.trans <- rbind(plot.data.trans,
                           data.frame(sample_size=Nind.trans,
                                      power=power.f2.per.test.trans,
                                      n.genes,
                                      alpha.trans,
                                      Rsq=Rsq[i],
                                      es.quantile=es.quantile[i],
                                      cell_type="Whole blood"))

}


g3 <- ggplot(data=plot.data.trans,
             aes(x=n.genes, y=power, col=factor(es.quantile))) +
  geom_line() +
  labs(col="ES quantile",
       y="Power to detect trans eQTL",
       x="Number of genes tested") +
  scale_color_manual(values=col.set)

print(g3)
```




```
ggsave("power_trans_eQTLs.png", g3, width=14, height=9, units="cm")
```

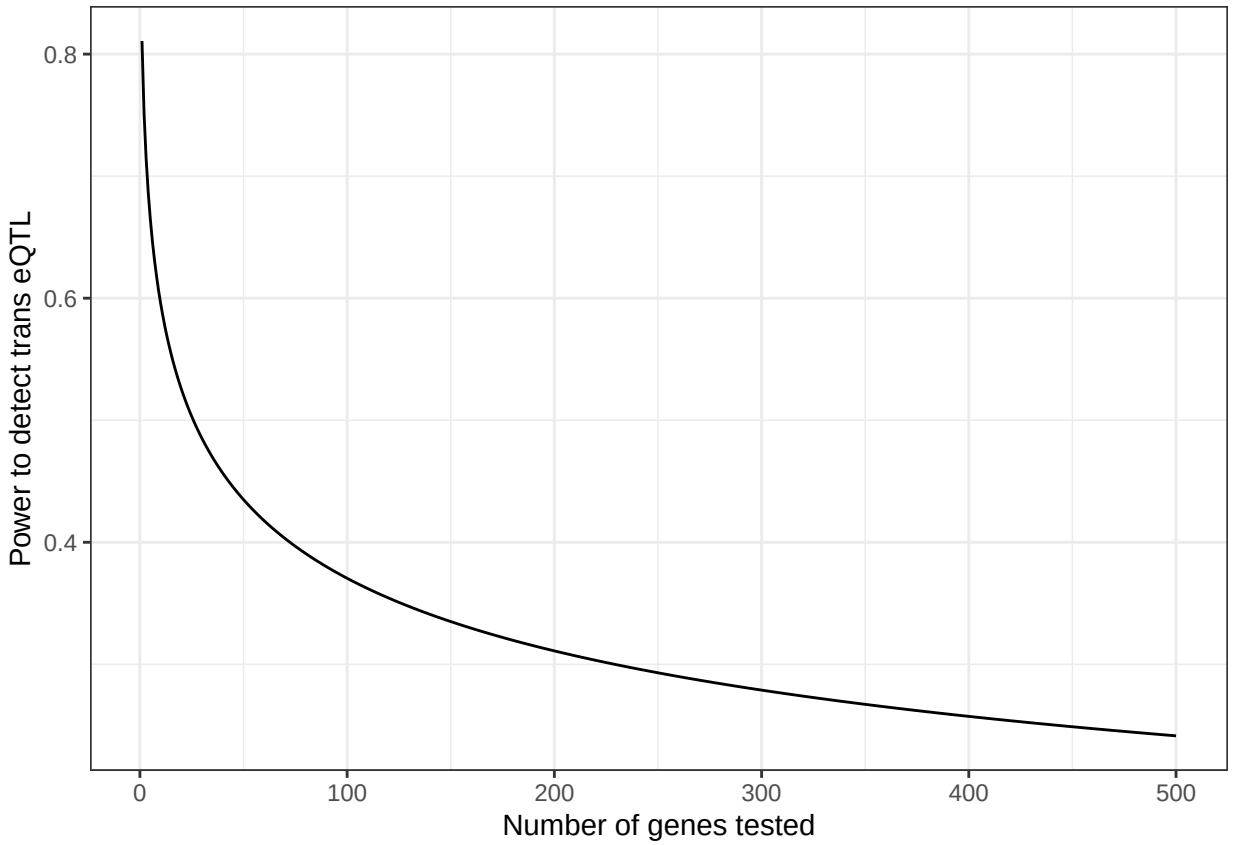
Results for 50% power

```
plot.data.trans %>% group_by(es.quantile) %>% filter(rank(abs(0.5 - power)) == 1)
```

```
## # A tibble: 5 x 7
## # Groups:   es.quantile [5]
##   sample_size   power n.genes alpha.trans     Rsq es.quantile cell_type
##         <dbl>   <dbl>   <int>       <dbl>   <dbl>       <dbl> <fct>
## 1          75 0.00183       1   0.000459  0.00510       0.9   Whole blood
## 2          75 0.0109        1   0.000459  0.0208        0.99  Whole blood
## 3          75 0.0786        1   0.000459  0.0595        0.999 Whole blood
## 4          75 0.476         2   0.000229  0.161         1.000 Whole blood
## 5          75 0.499        26   0.0000176 0.218         1.000 Whole blood
```

Plot for the supplement (just highest effect size)

```
g4 <- ggplot(data=plot.data.trans[plot.data.trans$es.quantile ==
                                    es.quantile[length(es.quantile)],],
             aes(x=n.genes, y=power)) +
  geom_line() +
  labs(y="Power to detect trans eQTL",
       x="Number of genes tested")

print(g4)
```




```
ggsave("power_trans_eQTLs.pdf", g4, width=14, height=9, units="cm")
```