

standarderrortwo

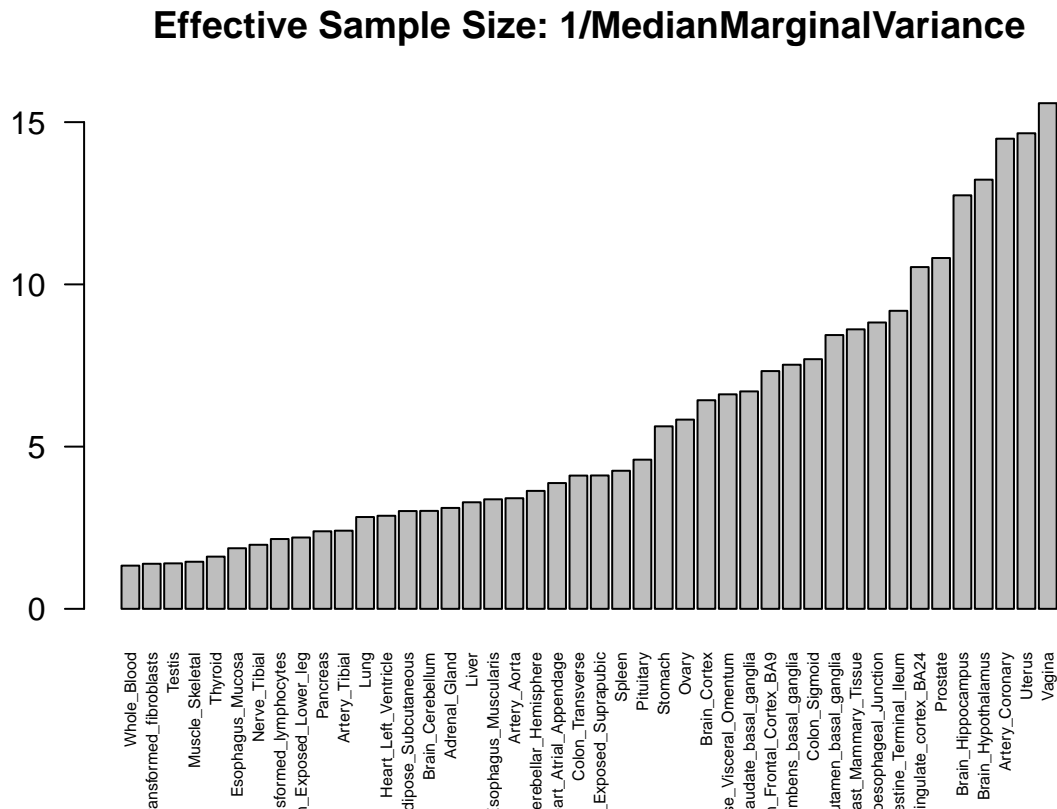
The aim of this document is to investigate the correlation of standard error and sample size, and to show how the presence of small sample sizes and large standard errors in biologically ‘unique’ tissues drives incompatibilities between the fold-size sharing heatmap and significance sharing heatmap.

Look at the ordering of Sample Size and see how it is almost identical to that of standard error, though no sample sizes differ by more than about 4 fold.

Now we look at the median posterior variances:

```
marginal.var=read.table("../Dropbox/Aug12/Aug13withEDmarginal.var.txt")[, -1]

median.mar.var=apply(marginal.var, 2, median)
barplot(sort(1/median.mar.var, decreasing=F), cex.names=0.5, main="Effective Sample Size: 1/MedianMarginalVariance")
```



Now let's plot effective sample size. Recall:

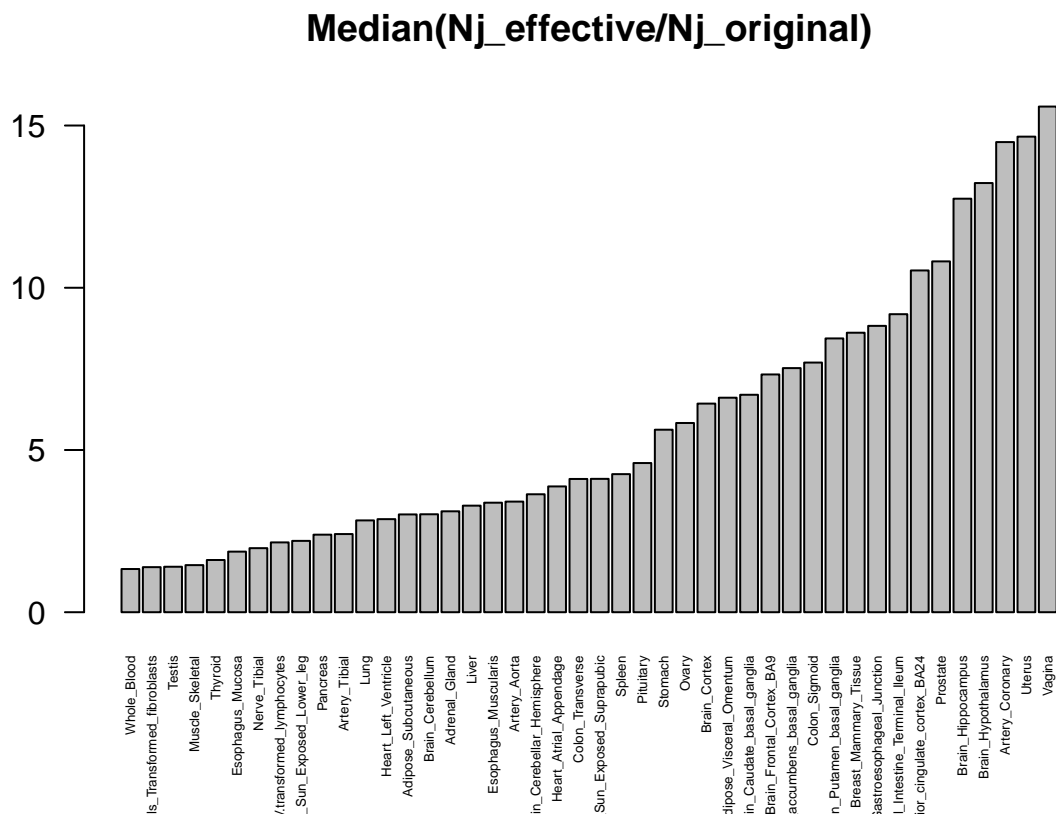
$$n_{eff} = \frac{s_j^2}{s_j^2}$$

```
original.var=as.matrix(standard.error.from.z)^2
#original.var=(standard.error.from.z/standard.error.from.z)^2
size=as.matrix(exp.sort)
post.var=as.matrix(marginal.var)*standard.error.from.z^2
njeffective=size*original.var/post.var
```

```
gtex.colors=read.table('../Data/GTEXColors.txt', sep = '\\t', comment.char = '')[-missing.tissues,2]
missing.tissues=c(7,8,19,20,24,25,31,34,37)

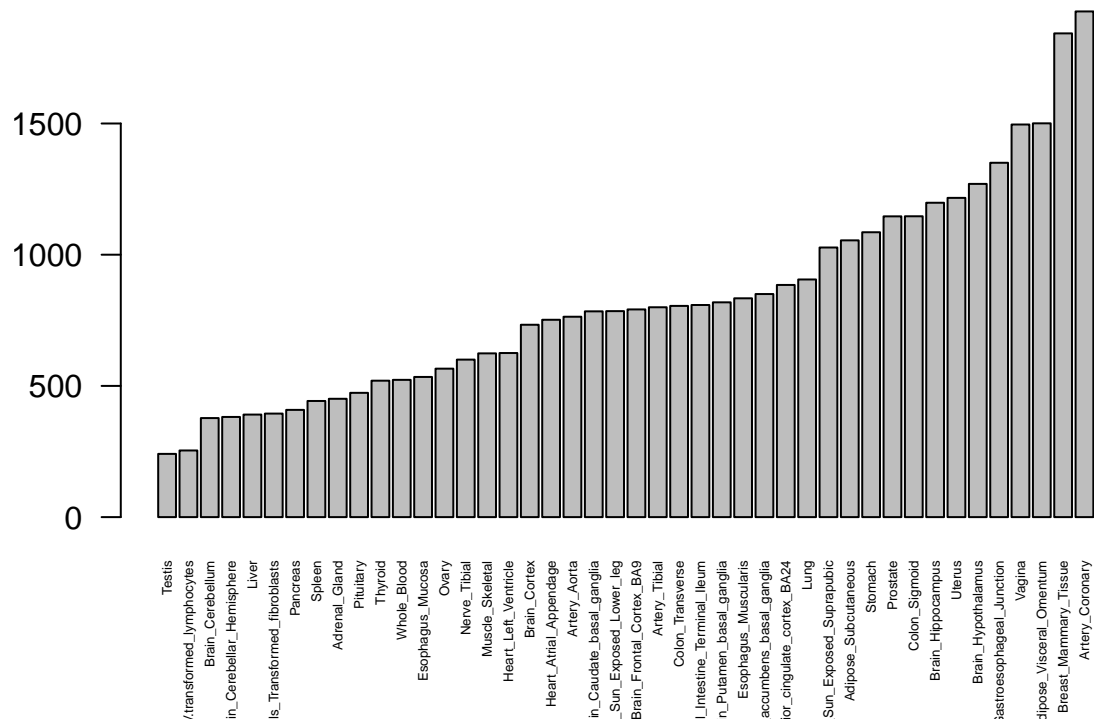
##ask why some of the original variances are smaller than posterior variances, even though calculations
###15093 44
increase=njeffective/size

barplot(sort(apply(increase,2,median),decreasing=F),las=2,cex.names=0.4)
title("Median(Nj_effective/Nj_original)")
```



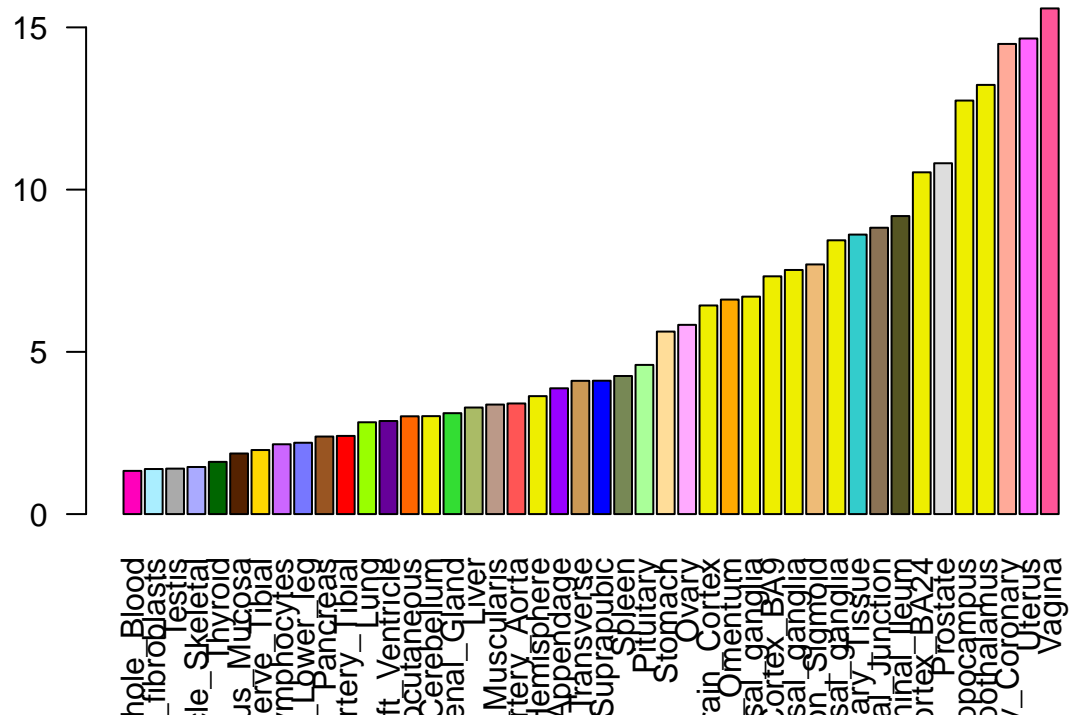
```
barplot(sort(apply(njeffective,2,median),decreasing=F),cex.names=0.4,las=2)
title("MedianNj_effective")
```

MedianNj_effective

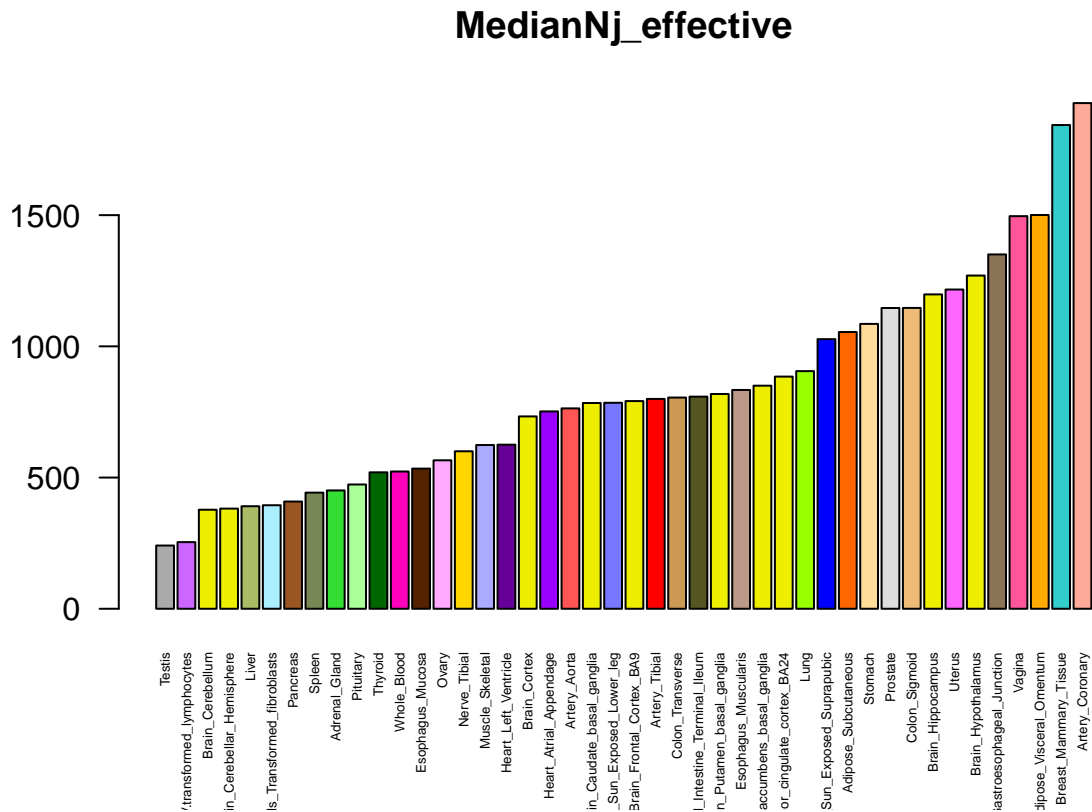


```
barplot(sort(apply(increase,2,median),decreasing=F),las=2,col=as.character(gtex.colors[order(apply(increase,2,median))]),
title("Median(Nj_effective/Nj_original)"))
```

Median(Nj_effective/Nj_original)



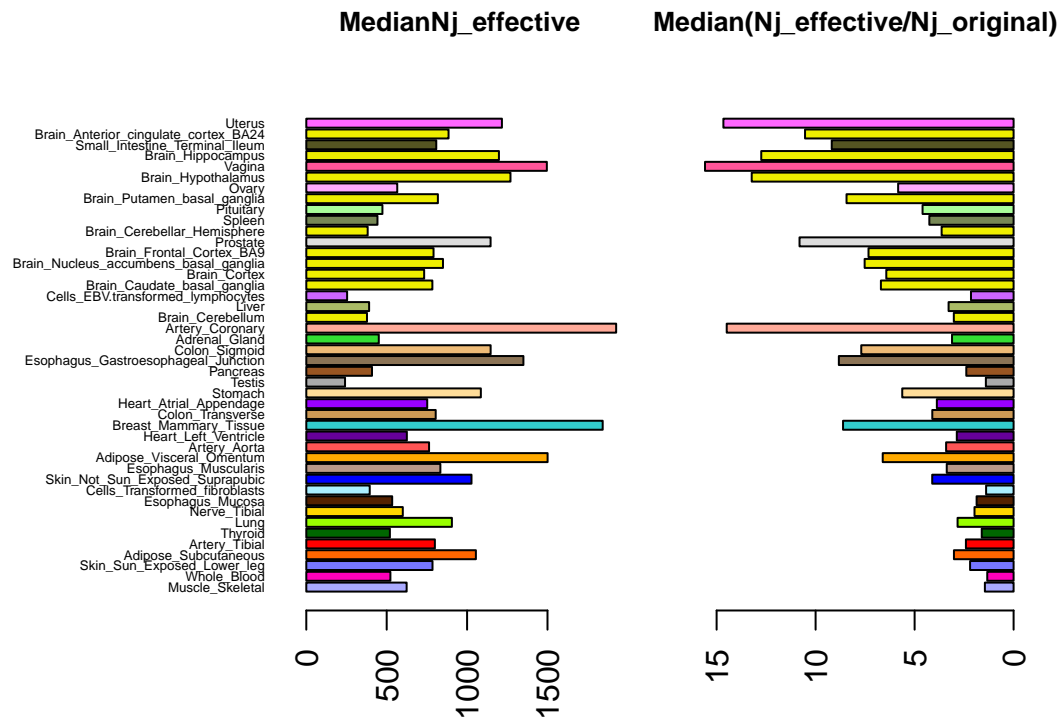
```
barplot(sort(apply(njeffective,2,median),decreasing=F),cex.names=0.4,las=2,col=as.character(gtex.colors),
title("MedianNj_effective"))
```



Let's plot again with order by original sample size:

```
par(mfrow=c(1,2))
samplesize=apply(size,2,function(x){unique(x)})
sampleorder=order(samplesize,decreasing = T)
median.nj.effective=apply(njeffective,2,median)
median.nj.increase=apply(increase,2,median)

par(mar=c(5.1,8,4.1,0.1))
barplot(median.nj.effective[sampleorder],cex.names=0.4,las=2,col=as.character(gtex.colors[sampleorder]))
title("MedianNj_effective",cex.main=0.8)
par(mar=c(5.1,2,4.1,6))
barplot(median.nj.increase[sampleorder],cex.names=0.4,las=2,col=as.character(gtex.colors[sampleorder]),l
title("Median(Nj_effective/Nj_original)",cex.main=0.8)
```



```

par(mfrow=c(1,2))
samplesize=apply(size,2,function(x){unique(x)})
sampleorder=order(samplesize,decreasing = T)
median.nj.effective=apply(njeffective,2,median)
median.nj.increase=apply(increase,2,median)

par(mar=c(5.1,8,4.1,0.1))
barplot(samplesize[sampleorder],cex.names=0.4,las=2,col=as.character(gtex.colors[sampleorder]),horiz = T,
title("Nj",cex.main=0.8))

par(mar=c(5.1,2,4.1,6))
barplot(median.nj.effective[sampleorder],cex.names=0.4,las=2,col=as.character(gtex.colors[sampleorder]),
title("Median(Nj_effective)",cex.main=0.8))

```

