

This is a file that describes how to make the matrix that indicates individuals in rows and tissues in columns. It can then be used to generate the appropriate heatmap.

Sarah Urbut

September 27, 2014

Contents

```
cd individuals
for i in $(ls | grep "expr"); do
b='basename $i ".expr.txt"'
echo $b
cat $i | head -1 > ./d/$b
done

for i in $(cat list.txt); do b='basename $i ".expr.txt"'
';echo $b; done
```

```
Adipose_Subcutaneous_Analysis
Adipose_Visceral_Omentum_Analysis
Adrenal_Gland_Analysis
Artery_Aorta_Analysis
Artery_Coronary_Analysis
Artery_Tibial_Analysis
Brain_Anterior_cingulate_cortex_BA24_Analysis
Brain_Caudate_basal_ganglia_Analysis
Brain_Cerebellar_Hemisphere_Analysis
```

```

Brain_Cerebellum_Analysis
Brain_Cortex_Analysis
Brain_Frontal_Cortex_BA9_Analysis
Brain_Hippocampus_Analysis
Brain_Hypothalamus_Analysis
Brain_Nucleus_accumbens_basal_ganglia_Analysis
Brain_Putamen_basal_ganglia_Analysis
Breast_Mammary_Tissue_Analysis
Cells_EBV-transformed_lymphocytes_Analysis
Cells_Transformed_fibroblasts_Analysis
Colon_Sigmoid_Analysis
Colon_Transverse_Analysis

```

```

setwd("./individuals/d")
files=list.files()[1:45]
tissues=read.table("tissuenames.txt",row.names=
  NULL)
  names <- list()
  for(i in 1:length(files)){
    names[[i]]=data.frame(read.table(files[i],
      header=FALSE, sep="\t"))}

```

###To determine length:

```

(max.inds <- max(unlist(lapply(names,function(x)
  {ncol(x)}))))
(tissue.max=tissues$V1[which.max(unlist(lapply(
  names,function(x){ncol(x)})))]))
namevector=t(data.frame(read.table(files[tissue.max
  ], header=FALSE, sep="\t")))

```

```

mat=matrix(NA,nrow=length(files),ncol=max.inds)

```

```

for(i in 1:44){
  r=data.frame(names[[i]])
  l=ncol(r)

```

```

    if (l <= 362)
    {
        nulls = 362 - l
        n = matrix(NA, ncol = nulls, nrow = 1)
        r = as.matrix(cbind(r, n))
    }
    mat[i,] = r[1,]
}

rownames(mat) = tissues$V1

ind.rows = t(mat)
rownames(ind.rows) = namevector[,1]

matched.mat = matrix(NA, ncol = 44, nrow = 362)
for (i in 1:ncol(ind.rows)) {
    matched.mat[,i] = ind.rows[match(namevector[,1],
    ind.rows[,i]),i]
}

rownames(matched.mat) = namevector[,1]
colnames(matched.mat) = tissues$V1

boolean.mat = matrix(NA, ncol = 44, nrow = 362)

matched.mat[1:10, 1:10]

for (i in 1:nrow(matched.mat)) {
    row = matched.mat[i,]
    t = sapply(row, function(x) {
    if (is.na(x)) {x=0}
    else if (!is.na(x)) {x=1}
    })
    boolean.mat[i,] = t
}

rownames(boolean.mat) = namevector[,1]

```

```
colnames(boolean.mat)=tissues$V1  
boolean.mat=data.frame(boolean.mat)  
heatmap(boolean.mat)
```