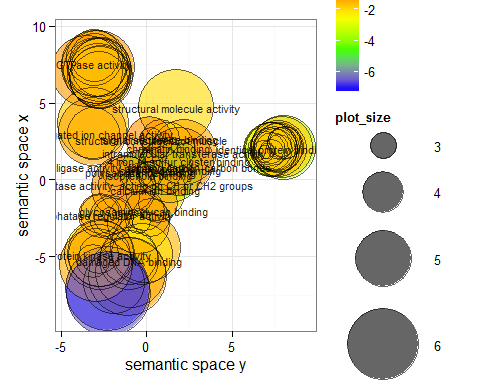
RNASeqV2-BRCA-1-3

Yu Shang [yushang@uga.edu](mailto:yushang@uga.edu)

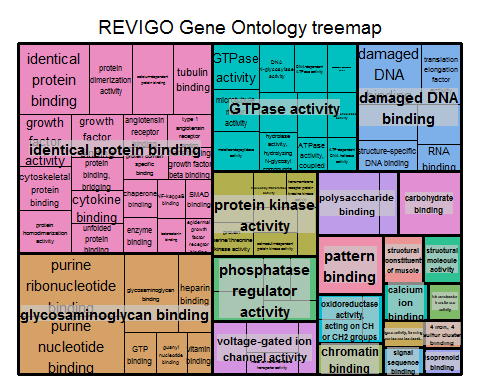
Tuesday, February 10, 2015

## What's New

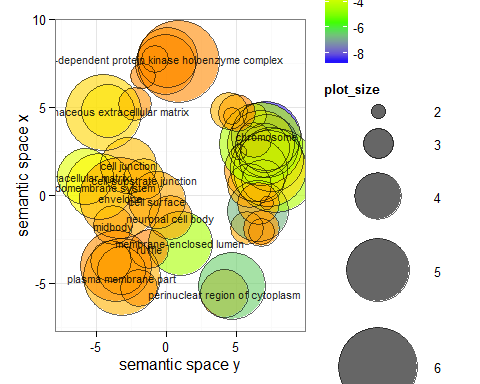
setwd("D:/BRCA\_RNASeq/");  
source("GO-PATHWAY1.r");



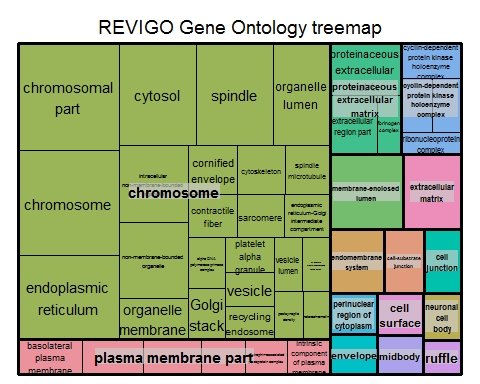
source("GO-PATHWAY01.r");



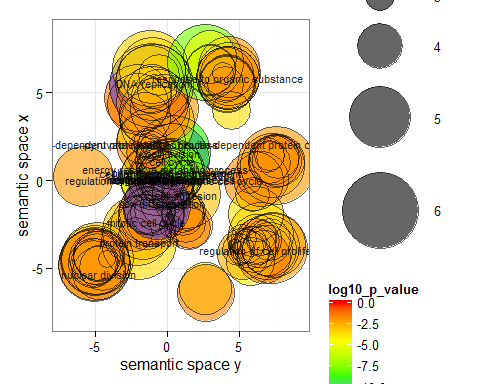
source("GO-PATHWAY2.r");



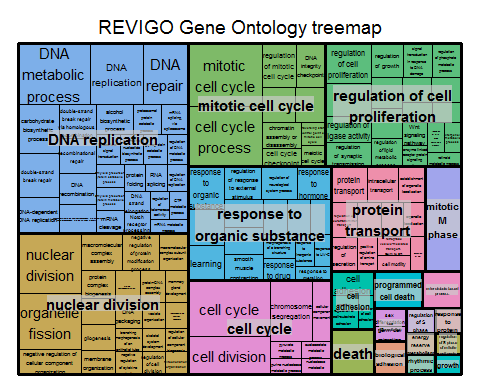
source("GO-PATHWAY02.r");



source("GO-PATHWAY3.r");



source("GO-PATHWAY03.r");



* Goto <https://tcga-data.nci.nih.gov/tcga/>
* Select BRCA
* Download RNASeqV2 (RSEM) by Tumor/Normal

## The First Class by Professor Xu

* Monday, January 26, 2015
* C128 3:00 PM
* Department of Biochemistry and Molecular Biology
* University of Georgia, Athens, GA 30602-7229

#setwd("~/Documents/BCMB6015\_Lab/BRCA\_RNASeq/Normal");  
setwd("D:/BRCA\_RNASeq/Normal/");  
sample.map = read.table("./FILE\_SAMPLE\_MAP.txt", header = T);  
head(sample.map);

## filename  
## 1 unc.edu.2b11e565-01a8-4054-a200-03e06765bd5e.1169055.junction\_quantification.txt  
## 2 unc.edu.2b11e565-01a8-4054-a200-03e06765bd5e.1173849.rsem.genes.results  
## 3 unc.edu.2b11e565-01a8-4054-a200-03e06765bd5e.1173850.rsem.isoforms.results  
## 4 unc.edu.2b11e565-01a8-4054-a200-03e06765bd5e.1174028.rsem.isoforms.normalized\_results  
## 5 unc.edu.2b11e565-01a8-4054-a200-03e06765bd5e.1174035.rsem.genes.normalized\_results  
## 6 unc.edu.2b11e565-01a8-4054-a200-03e06765bd5e.1764485.bt.exon\_quantification.txt  
## barcode.s.  
## 1 TCGA-BH-A0DQ-11A-12R-A089-07  
## 2 TCGA-BH-A0DQ-11A-12R-A089-07  
## 3 TCGA-BH-A0DQ-11A-12R-A089-07  
## 4 TCGA-BH-A0DQ-11A-12R-A089-07  
## 5 TCGA-BH-A0DQ-11A-12R-A089-07  
## 6 TCGA-BH-A0DQ-11A-12R-A089-07

file.indices = grep("rsem.genes.results", sample.map[,1]);  
head(file.indices);

## [1] 2 8 14 20 26 32

#setwd("~/Documents/BCMB6015\_Lab/BRCA\_RNASeq/Normal/RNASeqV2/UNC\_\_IlluminaHiSeq\_RNASeqV2/Level\_3");  
setwd("D:/BRCA\_RNASeq/Normal/RNASeqV2/UNC\_\_IlluminaHiSeq\_RNASeqV2/Level\_3");  
file1 = read.table(as.character(sample.map[file.indices[1],1]), header = T, sep = "\t");  
head(file1);

## gene\_id raw\_count scaled\_estimate transcript\_id  
## 1 ?|100130426 0.00 0.000000e+00 uc011lsn.1  
## 2 ?|100133144 39.16 1.000290e-06 uc010unu.1,uc010uoa.1  
## 3 ?|100134869 74.84 1.394188e-06 uc002bgz.2,uc002bic.2  
## 4 ?|10357 244.42 1.394645e-05 uc010zzl.1  
## 5 ?|10431 1988.00 5.292537e-05 uc001jiu.2,uc010qhg.1  
## 6 ?|136542 0.00 0.000000e+00 uc011krn.1

normal = file1[,2];  
head(normal);

## [1] 0.00 39.16 74.84 244.42 1988.00 0.00

a1 = as.integer(normal);  
head(a1);

## [1] 0 39 74 244 1988 0

normal = round(file1[,2]);  
head(normal);

## [1] 0 39 75 244 1988 0

for(i in 2:length(file.indices))  
{  
 current.file = read.table(as.character(sample.map[file.indices[i], 1]), header = T, sep = "\t");  
 length(current.file[,2]);  
 normal = cbind(normal, as.integer(current.file[,2]));  
}  
colnames(normal) = paste("normal", 1:10, sep="");  
#setwd("~/Documents/BCMB6015\_Lab/BRCA\_RNASeq/Tumor");  
setwd("D:/BRCA\_RNASeq/Tumor");  
sample.map = read.table("./FILE\_SAMPLE\_MAP.txt", header = T);  
head(sample.map);

## filename  
## 1 unc.edu.0b225e38-d4c8-4cb5-a668-316894ab858f.1155443.junction\_quantification.txt  
## 2 unc.edu.0b225e38-d4c8-4cb5-a668-316894ab858f.1157068.rsem.genes.results  
## 3 unc.edu.0b225e38-d4c8-4cb5-a668-316894ab858f.1157069.rsem.isoforms.results  
## 4 unc.edu.0b225e38-d4c8-4cb5-a668-316894ab858f.1157305.rsem.genes.normalized\_results  
## 5 unc.edu.0b225e38-d4c8-4cb5-a668-316894ab858f.1157308.rsem.isoforms.normalized\_results  
## 6 unc.edu.0b225e38-d4c8-4cb5-a668-316894ab858f.1771044.bt.exon\_quantification.txt  
## barcode.s.  
## 1 TCGA-A7-A0CH-01A-21R-A00Z-07  
## 2 TCGA-A7-A0CH-01A-21R-A00Z-07  
## 3 TCGA-A7-A0CH-01A-21R-A00Z-07  
## 4 TCGA-A7-A0CH-01A-21R-A00Z-07  
## 5 TCGA-A7-A0CH-01A-21R-A00Z-07  
## 6 TCGA-A7-A0CH-01A-21R-A00Z-07

file.indices = grep("rsem.genes.results", sample.map[,1]);  
head(file.indices);

## [1] 2 8 14 20 26 32

#setwd("~/Documents/BCMB6015\_Lab/BRCA\_RNASeq/Tumor/RNASeqV2/UNC\_\_IlluminaHiSeq\_RNASeqV2/Level\_3");  
setwd("D:/BRCA\_RNASeq/Tumor/RNASeqV2/UNC\_\_IlluminaHiSeq\_RNASeqV2/Level\_3");  
file1 = read.table(as.character(sample.map[file.indices[1],1]), header = T, sep = "\t");  
head(file1);

## gene\_id raw\_count scaled\_estimate transcript\_id  
## 1 ?|100130426 0.00 0.000000e+00 uc011lsn.1  
## 2 ?|100133144 13.57 5.458039e-07 uc010unu.1,uc010uoa.1  
## 3 ?|100134869 11.43 3.339613e-07 uc002bgz.2,uc002bic.2  
## 4 ?|10357 175.96 1.605399e-05 uc010zzl.1  
## 5 ?|10431 2259.00 9.472156e-05 uc001jiu.2,uc010qhg.1  
## 6 ?|136542 0.00 0.000000e+00 uc011krn.1

Tumor = file1[,2];  
head(Tumor);

## [1] 0.00 13.57 11.43 175.96 2259.00 0.00

a1 = as.integer(Tumor);  
head(a1);

## [1] 0 13 11 175 2259 0

Tumor = round(file1[,2]);  
head(Tumor);

## [1] 0 14 11 176 2259 0

for(i in 2:length(file.indices))  
{  
 current.file = read.table(as.character(sample.map[file.indices[i], 1]), header = T, sep = "\t");  
 length(current.file[,2]);  
 Tumor = cbind(Tumor, as.integer(current.file[,2]));  
}  
colnames(Tumor) = paste("Tumor", 1:10, sep="");  
all.samples = cbind(normal, Tumor);  
row.names(all.samples) = current.file[,1];  
head(all.samples);

## normal1 normal2 normal3 normal4 normal5 normal6 normal7  
## ?|100130426 0 0 0 0 0 0 0  
## ?|100133144 39 13 0 6 50 5 11  
## ?|100134869 75 12 9 1 32 8 5  
## ?|10357 244 248 178 118 178 208 142  
## ?|10431 1988 3082 1591 1351 2081 2656 2581  
## ?|136542 0 0 0 0 0 0 0  
## normal8 normal9 normal10 Tumor1 Tumor2 Tumor3 Tumor4 Tumor5  
## ?|100130426 0 0 0 0 0 0 0 0  
## ?|100133144 11 19 6 14 2 42 5 7  
## ?|100134869 29 19 8 11 13 37 6 10  
## ?|10357 137 143 208 176 215 215 160 392  
## ?|10431 1529 1555 1803 2259 2796 2658 1080 2434  
## ?|136542 0 0 0 0 0 0 0 0  
## Tumor6 Tumor7 Tumor8 Tumor9 Tumor10  
## ?|100130426 0 0 0 0 0  
## ?|100133144 19 2 20 7 32  
## ?|100134869 13 7 21 3 10  
## ?|10357 209 319 421 275 384  
## ?|10431 2381 3546 3634 3275 3755  
## ?|136542 0 0 0 0 0

setwd("D:/BRCA\_RNASeq/");  
write.table(all.samples, "./all\_samples.txt", quote = F, sep = "\t");

## The Second Class by Professor Xu

* Monday, February 02, 2015
* C128 3:00 PM
* Department of Biochemistry and Molecular Biology
* University of Georgia, Athens, GA 30602-7229

if(!require(DESeq2))  
{  
 source("http://bioconductor.org/biocLite.R");  
 biocLite("DESeq2");  
}

## Loading required package: DESeq2  
## Loading required package: S4Vectors  
## Loading required package: stats4  
## Loading required package: BiocGenerics  
## Loading required package: parallel  
##   
## Attaching package: 'BiocGenerics'  
##   
## The following objects are masked from 'package:parallel':  
##   
## clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,  
## clusterExport, clusterMap, parApply, parCapply, parLapply,  
## parLapplyLB, parRapply, parSapply, parSapplyLB  
##   
## The following object is masked from 'package:stats':  
##   
## xtabs  
##   
## The following objects are masked from 'package:base':  
##   
## anyDuplicated, append, as.data.frame, as.vector, cbind,  
## colnames, do.call, duplicated, eval, evalq, Filter, Find, get,  
## intersect, is.unsorted, lapply, Map, mapply, match, mget,  
## order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,  
## rbind, Reduce, rep.int, rownames, sapply, setdiff, sort,  
## table, tapply, union, unique, unlist, unsplit  
##   
## Loading required package: IRanges  
## Loading required package: GenomicRanges  
## Loading required package: GenomeInfoDb  
## Loading required package: Rcpp  
## Loading required package: RcppArmadillo

library(DESeq2);  
#setwd("~/Documents/BCMB6015\_Lab/BRCA\_RNASeq/");  
setwd("D:/BRCA\_RNASeq/");  
all.samples = read.table("./all\_samples.txt", header = T, row.names = 1);  
colData = data.frame(condition = c(rep("normal", times = 10), rep("tumor", times = 10)));  
colData;

## condition  
## 1 normal  
## 2 normal  
## 3 normal  
## 4 normal  
## 5 normal  
## 6 normal  
## 7 normal  
## 8 normal  
## 9 normal  
## 10 normal  
## 11 tumor  
## 12 tumor  
## 13 tumor  
## 14 tumor  
## 15 tumor  
## 16 tumor  
## 17 tumor  
## 18 tumor  
## 19 tumor  
## 20 tumor

row.names(colData) = names(all.samples);  
#colnames(all.samples);  
brca.dds = DESeqDataSetFromMatrix(countData = all.samples, colData = colData, design = ~condition);  
#brca.dds = DESeq(brca.dds);  
#brca.results = results(brca.dds);  
head(brca.dds);

## class: DESeqDataSet   
## dim: 1 20   
## exptData(0):  
## assays(1): counts  
## rownames(1): ?|100130426  
## rowData metadata column names(0):  
## colnames(20): normal1 normal2 ... Tumor9 Tumor10  
## colData names(1): condition

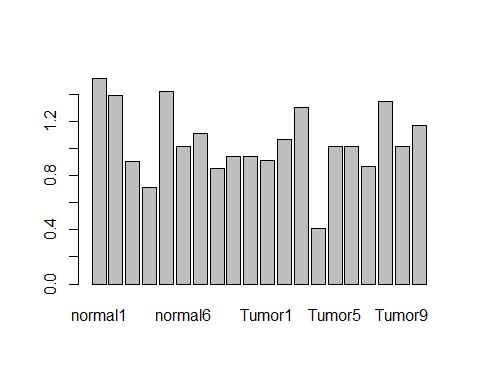
brca.dds = estimateSizeFactors(brca.dds);  
head(brca.dds);

## class: DESeqDataSet   
## dim: 1 20   
## exptData(0):  
## assays(1): counts  
## rownames(1): ?|100130426  
## rowData metadata column names(0):  
## colnames(20): normal1 normal2 ... Tumor9 Tumor10  
## colData names(2): condition sizeFactor

sizeFactors(brca.dds);

## normal1 normal2 normal3 normal4 normal5 normal6 normal7   
## 1.5130464 1.3927871 0.9056733 0.7106308 1.4210639 1.0161559 1.1090607   
## normal8 normal9 normal10 Tumor1 Tumor2 Tumor3 Tumor4   
## 0.8478863 0.9431461 0.9416870 0.9102059 1.0663467 1.3011372 0.4105615   
## Tumor5 Tumor6 Tumor7 Tumor8 Tumor9 Tumor10   
## 1.0172048 1.0169926 0.8666803 1.3426949 1.0124927 1.1712151

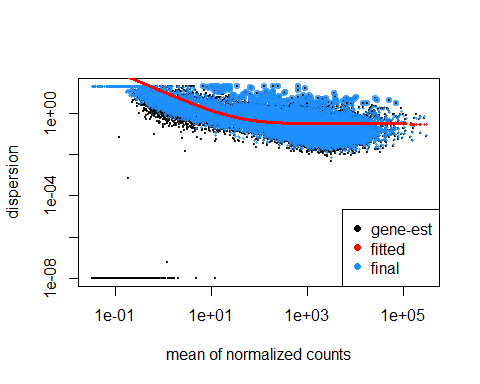
barplot(sizeFactors(brca.dds));



brca.dds = estimateDispersions(brca.dds);

## gene-wise dispersion estimates  
## mean-dispersion relationship  
## final dispersion estimates

plotDispEsts(brca.dds);



brca.dds = nbinomWaldTest(brca.dds);  
brca.results = results(brca.dds);  
dim(brca.results);

## [1] 20531 6

mode(brca.results);

## [1] "S4"

head(brca.results);

## log2 fold change (MAP): condition tumor vs normal   
## Wald test p-value: condition tumor vs normal   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ?|100130426 0.00000 NA NA NA NA  
## ?|100133144 13.58922 0.0513905 0.5138814 0.1000046 9.203407e-01  
## ?|100134869 14.62155 -0.4005714 0.4594789 -0.8717948 3.833203e-01  
## ?|10357 225.79420 0.7268437 0.1804444 4.0280756 5.623526e-05  
## ?|10431 2335.32854 0.5465976 0.1451822 3.7649091 1.666097e-04  
## ?|136542 0.00000 NA NA NA NA  
## padj  
## <numeric>  
## ?|100130426 NA  
## ?|100133144 0.9529287139  
## ?|100134869 0.5316065848  
## ?|10357 0.0004450373  
## ?|10431 0.0011118932  
## ?|136542 NA

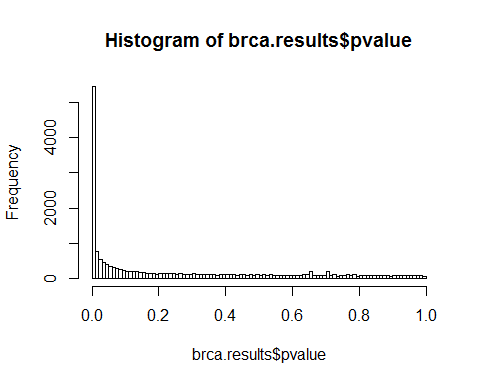
mcols(brca.results)$description;

## [1] "mean of normalized counts for all samples"   
## [2] "log2 fold change (MAP): condition tumor vs normal"  
## [3] "standard error: condition tumor vs normal"   
## [4] "Wald statistic: condition tumor vs normal"   
## [5] "Wald test p-value: condition tumor vs normal"   
## [6] "BH adjusted p-values"

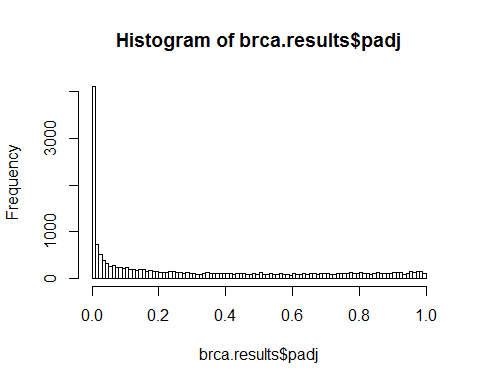
summary(brca.results);

##   
## out of 19420 with nonzero total read count  
## adjusted p-value < 0.1  
## LFC > 0 (up) : 3836, 20%   
## LFC < 0 (down) : 3444, 18%   
## outliers [1] : 923, 4.8%   
## low counts [2] : 952, 4.9%   
## (mean count < 0.5)  
## [1] see 'cooksCutoff' argument of ?results  
## [2] see 'independentFiltering' argument of ?results

hist(brca.results$pvalue, breaks = 100);



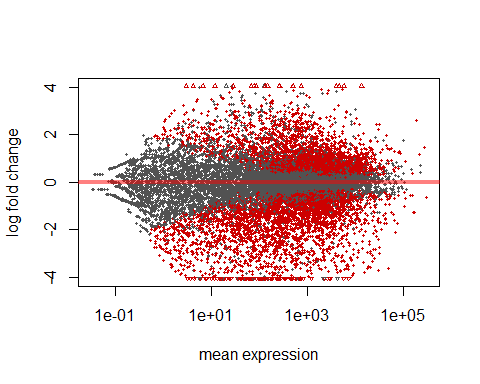
hist(brca.results$padj, breaks = 100);



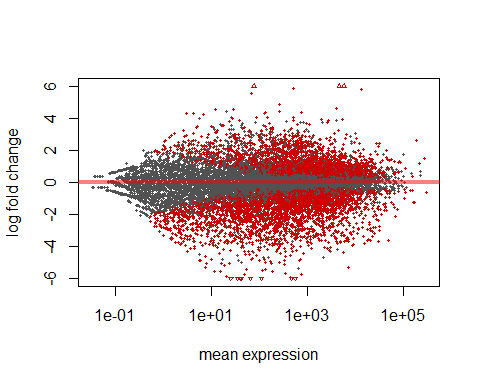
#sort the results by adjusted p-value  
sorted.results = brca.results[order(brca.results$padj),];  
head(sorted.results);

## log2 fold change (MAP): condition tumor vs normal   
## Wald test p-value: condition tumor vs normal   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## COL10A1|1300 4646.4100 7.509235 0.4765896 15.75618 6.228105e-56  
## AKR1B10|57016 471.1286 -6.715717 0.4835530 -13.88827 7.461101e-44  
## NTRK2|4915 8932.1881 -3.942006 0.3060717 -12.87935 5.882627e-38  
## MMP11|4320 13814.5646 5.822099 0.4716054 12.34528 5.165635e-35  
## HLF|3131 1225.8499 -3.830037 0.3126958 -12.24844 1.712502e-34  
## PAMR1|25891 2304.1978 -4.076160 0.3628675 -11.23319 2.801680e-29  
## padj  
## <numeric>  
## COL10A1|1300 1.092721e-51  
## AKR1B10|57016 6.545251e-40  
## NTRK2|4915 3.440357e-34  
## MMP11|4320 2.265777e-31  
## HLF|3131 6.009171e-31  
## PAMR1|25891 8.192580e-26

plotMA(brca.results);



plotMA(brca.results, ylim = c(-6,6));



head(sorted.results);

## log2 fold change (MAP): condition tumor vs normal   
## Wald test p-value: condition tumor vs normal   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## COL10A1|1300 4646.4100 7.509235 0.4765896 15.75618 6.228105e-56  
## AKR1B10|57016 471.1286 -6.715717 0.4835530 -13.88827 7.461101e-44  
## NTRK2|4915 8932.1881 -3.942006 0.3060717 -12.87935 5.882627e-38  
## MMP11|4320 13814.5646 5.822099 0.4716054 12.34528 5.165635e-35  
## HLF|3131 1225.8499 -3.830037 0.3126958 -12.24844 1.712502e-34  
## PAMR1|25891 2304.1978 -4.076160 0.3628675 -11.23319 2.801680e-29  
## padj  
## <numeric>  
## COL10A1|1300 1.092721e-51  
## AKR1B10|57016 6.545251e-40  
## NTRK2|4915 3.440357e-34  
## MMP11|4320 2.265777e-31  
## HLF|3131 6.009171e-31  
## PAMR1|25891 8.192580e-26

setwd("D:/BRCA\_RNASeq/");  
write.table(sorted.results, "./deseq\_results\_sorted.txt", quote = F, sep = "\t");  
head(brca.results);

## log2 fold change (MAP): condition tumor vs normal   
## Wald test p-value: condition tumor vs normal   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## ?|100130426 0.00000 NA NA NA NA  
## ?|100133144 13.58922 0.0513905 0.5138814 0.1000046 9.203407e-01  
## ?|100134869 14.62155 -0.4005714 0.4594789 -0.8717948 3.833203e-01  
## ?|10357 225.79420 0.7268437 0.1804444 4.0280756 5.623526e-05  
## ?|10431 2335.32854 0.5465976 0.1451822 3.7649091 1.666097e-04  
## ?|136542 0.00000 NA NA NA NA  
## padj  
## <numeric>  
## ?|100130426 NA  
## ?|100133144 0.9529287139  
## ?|100134869 0.5316065848  
## ?|10357 0.0004450373  
## ?|10431 0.0011118932  
## ?|136542 NA

significant.results = brca.results[!is.na(brca.results$padj) && (brca.results$padj < 0.1), ];  
#significant.results = brca.results[which(is.nan(brca.results$padj)),][brca.results$padj < 0.1, ];  
head(significant.results);

## log2 fold change (MAP): condition tumor vs normal   
## Wald test p-value: condition tumor vs normal   
## DataFrame with 0 rows and 6 columns

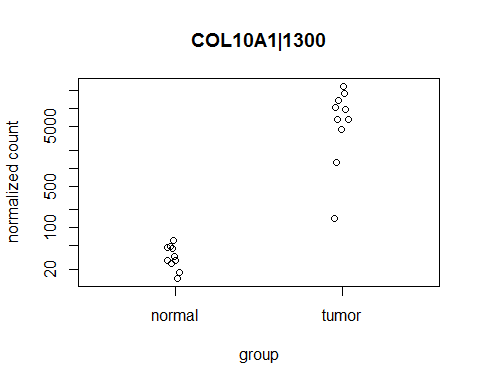
###Prepare data for DEB  
gene.names = row.names(all.samples);  
all.samples.deb= cbind(gene = gene.names, all.samples);  
setwd("D:/BRCA\_RNASeq/");  
write.table(all.samples.deb, "./all\_samples\_deb.txt", row.names = F, quote = F, sep = "\t");  
#Read edgeR output from DEB  
setwd("D:/BRCA\_RNASeq/");  
edger.results = read.table("./ori\_e.txt", header = T);  
head(edger.results);

## Gene logConc logFC P.Value FDR  
## 1 GNAT3|346562 12.429372 3.756826 8.849391e-59 1.816868e-54  
## 2 COL10A1|1300 7.915415 6.601932 1.653811e-43 1.697719e-39  
## 3 CGA|1081 8.821348 3.742481 2.682097e-43 1.835537e-39  
## 4 SULT1C3|442038 -13.373628 1.433742 8.917195e-43 4.576973e-39  
## 5 DLK1|8788 -8.343266 3.553024 8.569674e-42 3.518880e-38  
## 6 TRY6|154754 8.590074 2.084256 3.457267e-41 1.183019e-37

dim(significant.results);

## [1] 0 6

#take consensus  
consensus.genes = intersect(row.names(sorted.results[1:100,]),edger.results$Gene);  
plotCounts(brca.dds, gene = which.min(brca.results$padj), intgroup = "condition");



## The Third Class by Professor Xu

* Monday, February 09, 2015
* C128 3:00 PM
* Department of Biochemistry and Molecular Biology
* University of Georgia, Athens, GA 30602-7229

#read in previously saved results  
#setwd("~/Documents");  
setwd("D:/BRCA\_RNASeq/");  
sorted.results = read.table("./deseq\_results\_sorted.txt", header = T, sep = "\t");  
head(sorted.results);

## baseMean log2FoldChange lfcSE stat pvalue  
## COL10A1|1300 4646.4100 7.509235 0.4765896 15.75618 6.228105e-56  
## AKR1B10|57016 471.1286 -6.715717 0.4835530 -13.88827 7.461101e-44  
## NTRK2|4915 8932.1881 -3.942006 0.3060717 -12.87935 5.882627e-38  
## MMP11|4320 13814.5646 5.822099 0.4716054 12.34528 5.165635e-35  
## HLF|3131 1225.8499 -3.830037 0.3126958 -12.24844 1.712502e-34  
## PAMR1|25891 2304.1978 -4.076160 0.3628675 -11.23319 2.801680e-29  
## padj  
## COL10A1|1300 1.092721e-51  
## AKR1B10|57016 6.545251e-40  
## NTRK2|4915 3.440357e-34  
## MMP11|4320 2.265777e-31  
## HLF|3131 6.009171e-31  
## PAMR1|25891 8.192580e-26

#Remove rows with NAs in them.  
sorted.results = sorted.results[complete.cases(sorted.results), ];  
head(sorted.results);

## baseMean log2FoldChange lfcSE stat pvalue  
## COL10A1|1300 4646.4100 7.509235 0.4765896 15.75618 6.228105e-56  
## AKR1B10|57016 471.1286 -6.715717 0.4835530 -13.88827 7.461101e-44  
## NTRK2|4915 8932.1881 -3.942006 0.3060717 -12.87935 5.882627e-38  
## MMP11|4320 13814.5646 5.822099 0.4716054 12.34528 5.165635e-35  
## HLF|3131 1225.8499 -3.830037 0.3126958 -12.24844 1.712502e-34  
## PAMR1|25891 2304.1978 -4.076160 0.3628675 -11.23319 2.801680e-29  
## padj  
## COL10A1|1300 1.092721e-51  
## AKR1B10|57016 6.545251e-40  
## NTRK2|4915 3.440357e-34  
## MMP11|4320 2.265777e-31  
## HLF|3131 6.009171e-31  
## PAMR1|25891 8.192580e-26

#result.genes = row.names(sorted.results);  
#gene.info = sapply(result.genes, strsplit, "|", fixed = T);  
#head(result.genes);  
#head(gene.info);  
#gene.info = data.frame(matrix(unlist(gene.info)), nrow = length(result.genes));  
#gene.info = data.frame(matrix(unlist(gene.info)), nrow = length(result.genes), byrow = T);  
#head(gene.info);  
#dim(gene.info);  
significant.results = sorted.results[sorted.results$padj < 0.1, ];  
dim(significant.results);

## [1] 7280 6

head(significant.results);

## baseMean log2FoldChange lfcSE stat pvalue  
## COL10A1|1300 4646.4100 7.509235 0.4765896 15.75618 6.228105e-56  
## AKR1B10|57016 471.1286 -6.715717 0.4835530 -13.88827 7.461101e-44  
## NTRK2|4915 8932.1881 -3.942006 0.3060717 -12.87935 5.882627e-38  
## MMP11|4320 13814.5646 5.822099 0.4716054 12.34528 5.165635e-35  
## HLF|3131 1225.8499 -3.830037 0.3126958 -12.24844 1.712502e-34  
## PAMR1|25891 2304.1978 -4.076160 0.3628675 -11.23319 2.801680e-29  
## padj  
## COL10A1|1300 1.092721e-51  
## AKR1B10|57016 6.545251e-40  
## NTRK2|4915 3.440357e-34  
## MMP11|4320 2.265777e-31  
## HLF|3131 6.009171e-31  
## PAMR1|25891 8.192580e-26

dim(sorted.results);

## [1] 17545 6

significant.upreg = significant.results[significant.results$log2FoldChange > 0, ];  
head(significant.upreg);

## baseMean log2FoldChange lfcSE stat pvalue  
## COL10A1|1300 4646.4100 7.509235 0.4765896 15.75618 6.228105e-56  
## MMP11|4320 13814.5646 5.822099 0.4716054 12.34528 5.165635e-35  
## COL11A1|1301 5846.8550 6.124961 0.5468804 11.19982 4.085698e-29  
## GABRD|2563 151.8099 4.069079 0.3793971 10.72512 7.758757e-27  
## NEK2|4751 634.4589 3.642473 0.3479038 10.46977 1.189317e-25  
## INHBA|3624 755.2826 4.370453 0.4196600 10.41427 2.134222e-25  
## padj  
## COL10A1|1300 1.092721e-51  
## MMP11|4320 2.265777e-31  
## COL11A1|1301 1.024051e-25  
## GABRD|2563 1.512527e-23  
## NEK2|4751 1.738881e-22  
## INHBA|3624 2.674637e-22

dim(significant.upreg);

## [1] 3836 6

nrow(significant.upreg);

## [1] 3836

significant.dnreg = significant.results[significant.results$log2FoldChange < 0, ];  
dim(significant.dnreg);

## [1] 3444 6

head(significant.dnreg);

## baseMean log2FoldChange lfcSE stat pvalue  
## AKR1B10|57016 471.1286 -6.715717 0.4835530 -13.88827 7.461101e-44  
## NTRK2|4915 8932.1881 -3.942006 0.3060717 -12.87935 5.882627e-38  
## HLF|3131 1225.8499 -3.830037 0.3126958 -12.24844 1.712502e-34  
## PAMR1|25891 2304.1978 -4.076160 0.3628675 -11.23319 2.801680e-29  
## KLHL29|114818 728.6962 -3.377582 0.3095793 -10.91023 1.029933e-27  
## ALDH3A1|218 440.2601 -5.944863 0.5557482 -10.69704 1.050767e-26  
## padj  
## AKR1B10|57016 6.545251e-40  
## NTRK2|4915 3.440357e-34  
## HLF|3131 6.009171e-31  
## PAMR1|25891 8.192580e-26  
## KLHL29|114818 2.258772e-24  
## ALDH3A1|218 1.843571e-23

if(!require(RColorBrewer))  
{  
 install.packages("RColorBrewer");  
}

## Loading required package: RColorBrewer

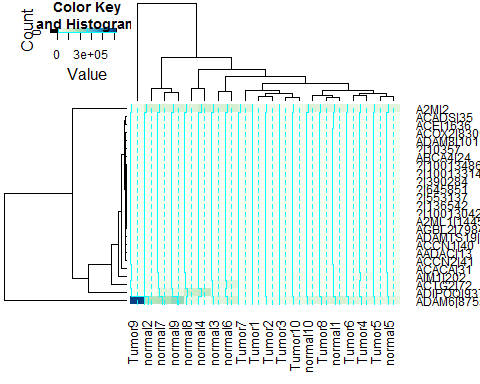
library(RColorBrewer);  
if(!require(gplots))  
{  
 install.packages("gplots");  
}

## Loading required package: gplots  
##   
## Attaching package: 'gplots'  
##   
## The following object is masked from 'package:IRanges':  
##   
## space  
##   
## The following object is masked from 'package:stats':  
##   
## lowess

library(gplots);  
if(!require(DESeq2))  
{  
 source("http://bioconductor.org/biocLite.R");  
 biocLite("DESeq2");  
}  
library(DESeq2);  
upreg.25 = order(significant.upreg$log2FoldChange, decreasing = T)[1:25];  
dnreg.25 = order(significant.dnreg$log2FoldChange)[1:25];  
hmcol = colorRampPalette(brewer.pal(9, "GnBu"))(100);  
#setwd("~/Documents/BCMB6015\_Lab/BRCA\_RNASeq/");  
setwd("D:/BRCA\_RNASeq/");  
all.samples = read.table("./all\_samples.txt", header = T, row.names = 1);  
colData = data.frame(condition = c(rep("normal", times = 10), rep("tumor", times = 10)));  
row.names(colData) = names(all.samples);  
brca.dds = DESeqDataSetFromMatrix(countData = all.samples, colData = colData, design = ~condition);  
brca.dds = DESeq(brca.dds);

## estimating size factors  
## estimating dispersions  
## gene-wise dispersion estimates  
## mean-dispersion relationship  
## final dispersion estimates  
## fitting model and testing  
## -- replacing outliers and refitting for 1044 genes  
## -- DESeq argument 'minReplicatesForReplace' = 7   
## -- original counts are preserved in counts(dds)  
## estimating dispersions  
## fitting model and testing

heatmap.2(counts(brca.dds, normalized = T)[upreg.25, ], col = hmcol);



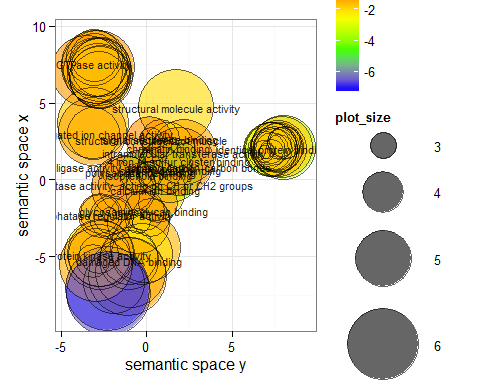
#heatmap.2(counts(brca.dds, normalized = T)[upreg.25, ], col = hmcol, Rowv = F, Colv = F, scale = "none", dendragram = "none", margin = c(10,6));  
Cytoscape\_GO = rownames(significant.results);  
cc = c();  
for(j in 1:length(Cytoscape\_GO))  
{  
 c = c();  
 temp = unlist(strsplit(as.character(Cytoscape\_GO[j]), ""));  
 index = 0;  
 for(i in 1:length(temp))  
 {  
 if(temp[i] == "|")  
 {  
 temp[i] = "\_";  
 index = index + 1;  
 }  
 if(index < 1)  
 c = paste(c, temp[i], sep = "");  
 }  
 cc = c(cc, c);  
}  
head(Cytoscape\_GO);

## [1] "COL10A1|1300" "AKR1B10|57016" "NTRK2|4915" "MMP11|4320"   
## [5] "HLF|3131" "PAMR1|25891"

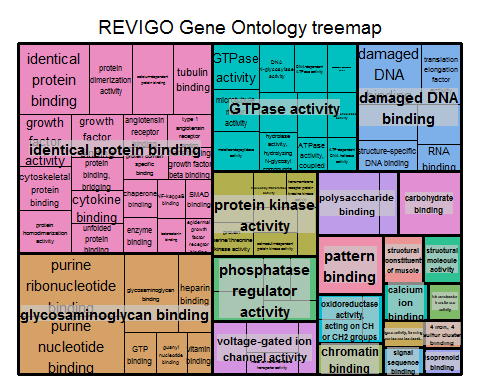
head(cc);

## [1] "COL10A1" "AKR1B10" "NTRK2" "MMP11" "HLF" "PAMR1"

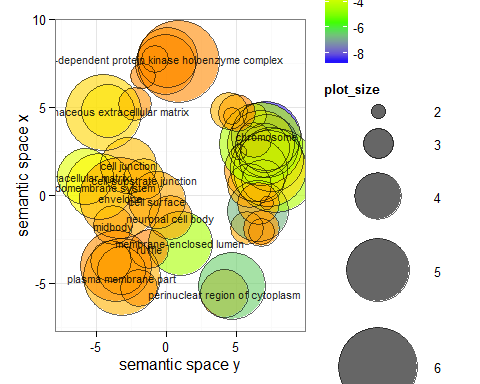
cc = as.matrix(cc);  
#setwd("~/Documents");  
setwd("D:/BRCA\_RNASeq/");  
write.table(cc, "Cytoscape\_GO.txt", quote = FALSE, row.names =FALSE, col.names =FALSE, sep = "\t");  
#Using DAVID, Gene Ontology (GO), KEGG PATHWAY, Cytoscape, REVIGO  
#<CLICK> Make R script for plotting  
source("GO-PATHWAY1.r");



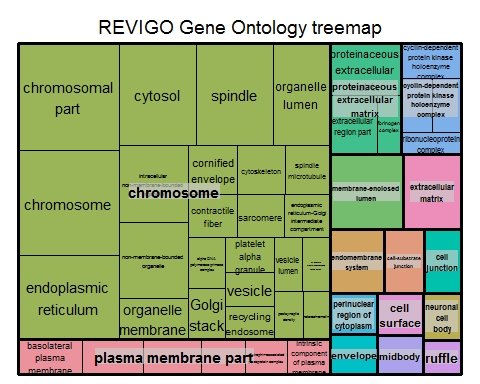
source("GO-PATHWAY01.r");



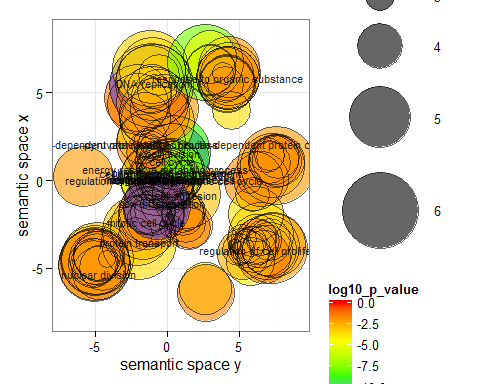
source("GO-PATHWAY2.r");



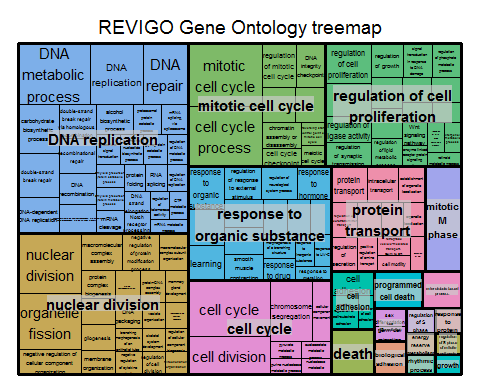
source("GO-PATHWAY02.r");



source("GO-PATHWAY3.r");



source("GO-PATHWAY03.r");



## The Fourth Class by Professor Xu

* Monday, February 16, 2015
* C128 3:00 PM
* Department of Biochemistry and Molecular Biology
* University of Georgia, Athens, GA 30602-7229

cat("To Be ...\n");

## To Be ...