Solutions for chapter   Hypergeometric Testing Used for Gene Set Enrichment Analysis

Exercise 1

```
a > numSamp = length(ALL_bcrneg$mol.biol)
  > table(ALL_bcrneg$mol.biol)

  BCR/ABL     NEG
       37      42
```

```
b > annotation(ALL_bcrneg)

  [1] "hgu95av2"

  > length(featureNames(ALL_bcrneg))

  [1] 12625
```

Exercise 2

See the description of the `remove.dupEntrez` argument in the manual page for `nsFilter`.

Exercise 3

The `filter.log` component of the `nsFilter` return value provides information about the number of probe sets removed by each step of the filter.

Exercise 4

```
> chrN = mget(featureNames(ALLfilt_bcrneg), envir=hgu95av2CHR)
> onY = sapply(chrN, function(x) any(x == "Y"))
> onY[is.na(onY)] = FALSE
> ALLfilt_bcrneg = ALLfilt_bcrneg[!onY, ]
```

Exercise 5

```
> ## an alternate universe based on the entire chip
> chipAffyUniverse = featureNames(ALLfilt_bcrneg)
> chipEntrezUniverse = mget(chipAffyUniverse, hgu95av2ENTREZID)
> chipEntrezUniverse = unique(unlist(chipEntrezUniverse))
```

Exercise 6

```
> sumpv = sum(smPV)
```

There are 646 probe sets with p-values less than 0.05.

Exercise 7

```
a > df = summary(hgOver)
  > names(df)

  [1] "GOBPID"    "Pvalue"     "OddsRatio" "ExpCount"
  [5] "Count"     "Size"       "Term"
```

```
b > df = summary(hgOver, pvalue=0.05, categorySize=350)
  > nrow(df)
  [1] 23
```

```
c > ? HyperGResult-accessors
```

Exercise 8

```
> browseURL("ALL_hgo.html")
```

Exercise 9

```
> numG = length(sigSub)
> sizes = sapply(sigSub, numNodes)
> sizes
1 2 3 4
7 1 2 1
```

are displayed above.

Exercise 10

```
a > dfcond = summary(hgCond, categorySize=50)
  > ## trim the term names for display purposes
  > trimTerm = function(x) {
        if (nchar(x) <= 20)
            x
        else
            paste(substr(x, 1, 20), "...", sep="")
    }
  > dfcond$Term = sapply(dfcond$Term, trimTerm)
  > sizeOrd = order(dfcond$Size, decreasing=TRUE)
  > dfcond[sizeOrd, c("Count", "Size", "Term")]

    Count Size                  Term
  2   388 1975 regulation of biolog...
  3   214  999 developmental proces...
  5   166  806     signal transduction
  1    62  220 regulation of signal...
  4    54  197           cell adhesion
  6    29   93 regulation of cell m...
```

```
b > stdIds = sigCategories(hgOver)
  > condIds = sigCategories(hgCond)
  > setdiff(stdIds, condIds)
  [1] "GO:0007154" "GO:0010646" "GO:0050794" "GO:0065007"
  [5] "GO:0022610"
```

Exercise 11

```
> params = new("ChrMapHyperGParams",
      conditional=FALSE, testDirection="over",
      universeGeneIds=entrezUniverse,
      geneIds=selectedEntrezIds,
```

```
        annotation="hgu95av2", pvalueCutoff=0.05)
> paramsCond = params
> conditional(paramsCond) = TRUE
```

```
> hgans = hyperGTest(params)
> hgansCond = hyperGTest(paramsCond)
```

```
> summary(hgans, categorySize=10)
  ChrMapID   Pvalue OddsRatio ExpCount Count Size
1     7p15 0.000852      5.44     2.94     9   17
2      7p1 0.005214      2.61     6.92    14   40
3       7p 0.007046      2.18    10.03    18   58
4     8q24 0.014872      2.68     4.84    10   28
5     1q21 0.017153      2.25     7.09    13   41
6     3q25 0.028263      4.01     1.90     5   11
7    14q22 0.032103      3.21     2.59     6   15
8        7 0.045438      1.41    30.09    39  174
```

Exercise 12

```
> kparams = new("KEGGHyperGParams",
      geneIds=selectedEntrezIds,
      universeGeneIds=entrezUniverse,
      annotation="hgu95av2",
      pvalueCutoff=0.05,
      testDirection="over")
> kans = hyperGTest(kparams)
```

```
> summary(kans)
  KEGGID Pvalue OddsRatio ExpCount Count Size
1  04360 0.0184      2.04     9.49    16   53
2  04810 0.0404      1.60    17.01    24   95
3  04510 0.0428      1.63    15.40    22   86
                             Term
1                     Axon guidance
2 Regulation of actin cytoskeleton
3                    Focal adhesion
> kparamsUnder = kparams
> testDirection(kparamsUnder) = "under"
```

```
> kansUnder = hyperGTest(kparamsUnder)
```

```
> summary(kansUnder)
  KEGGID Pvalue OddsRatio ExpCount Count Size
1  05219 0.0154     0.000     3.76     0   21
2  00020 0.0188     0.000     3.58     0   20
3  04120 0.0194     0.384    11.28     5   63
4  04664 0.0225     0.249     6.80     2   38
5  05211 0.0262     0.257     6.62     2   37
6  05214 0.0413     0.281     6.09     2   34
7  00510 0.0419     0.000     2.86     0   16
                             Term
1                     Bladder cancer
2         Citrate cycle (TCA cycle)
3    Ubiquitin mediated proteolysis
4 Fc epsilon RI signaling pathway
5             Renal cell carcinoma
```

```
6                          Glioma
7           N-Glycan biosynthesis
```

Exercise 13

```
> pparams = new("PFAMHyperGParams",
      geneIds=selectedEntrezIds,
      universeGeneIds=entrezUniverse,
      annotation="hgu95av2",
      pvalueCutoff=hgCutoff,
      testDirection="over")
> pans = hyperGTest(pparams)
```

```
> summary(pans)
    PFAMID    Pvalue OddsRatio ExpCount Count Size    Term
1 PF01833 0.000439      6.29    2.741     9   16 PF01833
2 PF00788 0.000733      8.55    1.885     7   11 PF00788
3 PF01023 0.000750     24.37    1.028     5    6 PF01023
4 PF08337 0.000855       Inf    0.685     4    4 PF08337
```