Building hsuReportedGenelists.Rda

Keith A. Baggerly

July 30, 2009

Contents

List of Figures

List of Tables

1 Executive Summary

1.1 Introduction

In this report, we import and format the gene lists reported by Hsu et al. [1] for cisplatin and pemetrexed.

1.2 Methods

Hsu et al. [1] report the probesets used in their drug sensitivity signatures for cisplatin and pemetrexed in their Supplementary Tables 1 and 2, respectively. These are available as 10593_Supplementary_Table_1.doc and 10593_Supplementary_Table_2.doc at `http://jco.ascopubs.org/cgi/content/full/25/28/4350/DC1`. We acquired these tables and reformatted them as csv files for easier loading.

1.3 Results

We created a "cisplatinTable" data frame containing the data in supplementary table 1, and an extracted data frame, "cisplatinReportedProbesets", containing just the non-null probeset ids and their associated gene symbols. Similarly, we created a "pemetrexedTable" data frame containing the data in supplementary table 2, and an extracted data frame, "pemetrexedReportedProbesets", containing just the non-null probeset ids and their associated gene symbols. We stored these in RDataObjects as "hsuReportedGeneLists.Rda."

2 Options and Libraries

```
> options(width = 80)
```

3 Loading and Parsing Data

3.1 Cisplatin

We begin with the table for cisplatin.

```
> cisplatinTable <- read.table(file.path("RawData", "HsuJCO", "cisplatinGeneList.csv"),
+     header = TRUE, skip = 1, sep = ",")
> dim(cisplatinTable)

[1] 182    5

> cisplatinTable[1:8, ]

  Gene.Title                              Gene.Symbol
1 200075_s_at                       guanylate kinase 1
2                                   guanylate kinase 1
3
4
5
6
7
8 200718_s_at S-phase kinase-associated protein 1A (p19A)
  GO.Biological.Process.Description GO.Molecular.Function.Description
1                              GUK1                  GTP biosynthesis
2
3
4
5
6
7
8                             SKP1A                    ubiquitin cycle
  GO.Cellular.Component.Description
1            nucleotide binding
2      guanylate kinase activity
3                    ATP binding
4                   drug binding
```

```
5                     kinase activity
6                 transferase activity
7         guanylate kinase activity
8                     protein binding
```

The cisplatin signature only contains 45 probesets, but the table has 182 rows. This is because some of the genes have multiple entries for GO function or component. We want to extract the non-null entries in the column containing probeset ids (Gene.Title) and the gene symbols in the corresponding rows (in the column labeled GO.Biological.Process.Description).

```
> cisplatinReportedProbesets <- cisplatinTable[as.character(cisplatinTable[,
+     "Gene.Title"]) != "", c("Gene.Title", "GO.Biological.Process.Description")]
> colnames(cisplatinReportedProbesets) <- c("probesetID", "geneSymbol")
> cisplatinReportedProbesets[, "probesetID"] <- as.character(cisplatinReportedProbesets[,
+     "probesetID"])
> cisplatinReportedProbesets[, "geneSymbol"] <- as.character(cisplatinReportedProbesets[,
+     "geneSymbol"])
> rownames(cisplatinReportedProbesets) <- cisplatinReportedProbesets[,
+     "probesetID"]
> dim(cisplatinReportedProbesets)

[1] 45  2

> cisplatinReportedProbesets[1:5, ]

             probesetID geneSymbol
200075_s_at 200075_s_at       GUK1
200718_s_at 200718_s_at      SKP1A
201014_s_at 201014_s_at      PAICS
201199_s_at 201199_s_at      PSMD1
201923_at     201923_at      PRDX4
```

3.2 Pemetrexed

Next, we turn to the table for pemetrexed.

```
> pemetrexedTable <- read.table(file.path("RawData", "HsuJCO",
+     "pemetrexedGeneList.csv"), header = TRUE, skip = 1, sep = ",")
> dim(pemetrexedTable)

[1] 396    5

> pemetrexedTable[1:5, ]

  Gene.Title                              Gene.Symbol
1    1100_at interleukin-1 receptor-associated kinase 1
2
3
4
5
```

```
   GO.Biological.Process.Description
1                              IRAK1
2
3
4
5
                                               GO.Molecular.Function.Description
1                                                               defense response
2                                                             signal transduction
3 transmembrane receptor protein serine/threonine kinase signaling pathway
4                                       activation of NF-kappaB-inducing kinase
5                                         positive regulation of transcription
        GO.Cellular.Component.Description
1                       nucleotide binding
2                    magnesium ion binding
3 protein serine/threonine kinase activity
4        NF-kappaB-inducing kinase activity
5                              ATP binding
```

The pemetrexed signature only contains 85 probesets, but the table has 396 rows. As with cisplatin, this is because some of the genes have multiple entries for GO function or component. We want to extract the non-null entries in the column containing probeset ids (Gene.Title) and the gene symbols in the corresponding rows (in the column labeled GO.Biological.Process.Description).

```
> pemetrexedReportedProbesets <- pemetrexedTable[as.character(pemetrexedTable[,
+     "Gene.Title"]) != "", c("Gene.Title", "GO.Biological.Process.Description")]
> colnames(pemetrexedReportedProbesets) <- c("probesetID", "geneSymbol")
> pemetrexedReportedProbesets[, "probesetID"] <- as.character(pemetrexedReportedProbesets[,
+     "probesetID"])
> pemetrexedReportedProbesets[, "geneSymbol"] <- as.character(pemetrexedReportedProbesets[,
+     "geneSymbol"])
> rownames(pemetrexedReportedProbesets) <- pemetrexedReportedProbesets[,
+     "probesetID"]
> dim(pemetrexedReportedProbesets)

[1] 85  2

> pemetrexedReportedProbesets[1:5, ]

          probesetID geneSymbol
1100_at      1100_at      IRAK1
1227_g_at  1227_g_at     ADAM17
1318_at      1318_at      RBBP4
1355_g_at  1355_g_at      NTRK2
241_g_at    241_g_at        SRM
```

4 Save Rda File

Finally, we save the full tables and extracted probeset lists.

```
> save(cisplatinTable, cisplatinReportedProbesets, pemetrexedTable,
+     pemetrexedReportedProbesets, file = file.path("RDataObjects",
+         "hsuReportedGenelists.Rda"))
```

5 Appendix

5.1 File Location

```
> getwd()

[1] "/Users/kabagg/ReproRsch/WebSite"
```

5.2 Saves

5.3 SessionInfo

```
> sessionInfo()

R version 2.8.1 (2008-12-22)
i386-apple-darwin8.11.1

locale:
en_US.UTF-8/en_US.UTF-8/C/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
[1] tools_2.8.1
```

References

[1] Hsu DS, Balakumaran BS, Acharya CR, et al.: Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer. *J Clin Oncol*, **25**:4350-4357, 2007