Building gyorffyAll.Rda

Keith A. Baggerly

July 29, 2009

Contents

List of Figures

List of Tables

1 Executive Summary

1.1 Introduction

In this report, we assemble the matrix of cell line quantifications provided by Györffy et al. [1] and couple it with their tabulated assessments of drug sensitivity (sensitive, intermediate, or resistant) for each cell line.

1.2 Methods

Györffy et al. [1] provided array quantifications for the 30 cell lines they examined in their first supplementary table, available as `http://www.interscience.wiley.com/jpages/0020-0136/suppmat/2006/jws-ijc.21570.tbl1.xls`. We reformatted this data as a csv file for easier loading. They also assessed each of their 30 cell lines as sensitive, intermediate, or resistant in response to each of 11 drugs, tabulating the values in their Figure 2. Again, we reorganized the data into a csv file for easier loading.

1.3 Results

We created a "gyorffyAll" matrix of array quantifications and a "gyorffyAllInfo" data frame of the sample drug sensitivities. We stored these in RDataObjects as "gyorffyAll.Rda."

2 Options and Libraries

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

3 Loading and Parsing Data

3.1 Quantification Data

First, we load the array quantifications.

```
> gyorffyAll <- read.table(file.path("RawData", "GyorffyIntJCanc",
+     "gyorffyData.csv"), header = TRUE, sep = ",", row.names = 1)
> dim(gyorffyAll)

[1] 22283    30

> gyorffyAll[1:3, 1:3]

              X181P     X257P      A375
1007_s_at  9.977162  8.830360  9.232587
1053_at    8.509234  7.975630  8.601956
117_at     7.320390  7.674556  7.571382
```

3.2 Sample Information

Next, we load information about which cell lines are sensitive to which drug.

```
> gyorffyAllInfo <- read.table(file.path("RawData", "GyorffyIntJCanc",
+     "gyorffyDrugInfo.csv"), header = TRUE, sep = ",", row.names = 1)
> dim(gyorffyAllInfo)

[1] 30 12

> gyorffyAllInfo[1:3, ]
```

```
           origin X5.FU Cisplatin Cyclophosphamide Doxorubicin Etoposide
181/85p pancreas     S        R                 S           M         R
257p      gastric    R        R                 S           S         R
A375     melanoma    M        R                 R           M         R
        Methotrexate MytomycinC Paclitaxel Vinblastine TopotecanC3
181/85p            M          R          S           S           M
257p               M          R          S           S           R
A375               M          M          S           S           R
        MitaxantroneC1
181/85p              R
257p                 R
A375                 R
```

4   Matching Names

The sample names are formatted differently in the two files.

```
> cbind(colnames(gyorffyAll), rownames(gyorffyAllInfo))
```

```
       [,1]       [,2]
 [1,] "X181P"    "181/85p"
 [2,] "X257P"    "257p"
 [3,] "A375"     "A375"
 [4,] "BT20"     "BT20"
 [5,] "C8161"    "C8161"
 [6,] "colo699"  "Colo699"
 [7,] "Cx2"      "CX-2"
 [8,] "Du145"    "DU145"
 [9,] "DV90"     "DV-90"
[10,] "ES2"      "ES-2"
[11,] "FUOV1"    "FU-OV-1"
[12,] "Hep3B"    "Hep3B"
[13,] "HRT18"    "HRT-18"
[14,] "HT29"     "HT-29"
[15,] "mda231"   "MDA-231"
[16,] "me43"     "ME-43"
[17,] "MeWo"     "MeWo"
[18,] "OAW42"    "OAW42"
[19,] "OVKAR"    "OVCAR3"
[20,] "R103"     "R103"
[21,] "R193"     "R193"
[22,] "SKBR3"    "SKBR3"
[23,] "SKMel13"  "SKMel13"
[24,] "SKMel19"  "SKMel19"
[25,] "Skov3"    "SKOV-3"
[26,] "SNU182"   "SNU182"
[27,] "SNU423"   "SNU423"
[28,] "SNU449"   "SNU449"
```

```
[29,] "SNU475"   "SNU475"
[30,] "Sw13"     "SW13"
```

Of the two formulations, we have a slight arbitrary preference for the latter, so we alter the column names
in the quantification matrix to match the row names in the drug sensitivity matrix.

```
> colnames(gyorffyAll) <- rownames(gyorffyAllInfo)
```

5   Save Rda File

Finally, we save the quantification matrix and the annotation information.

```
> save(gyorffyAll, gyorffyAllInfo, file = file.path("RDataObjects",
+     "gyorffyAll.Rda"))
```

6   Appendix

6.1   File Location

```
> getwd()
```

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

6.2   Saves

6.3   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] Györffy B, Surowiak P, Kiesslich O, et al.: Gene expression profiling of 30 cancer cell lines predicts
    resistance towards 11 anticancer drugs at clinically achieved concentrations. *Int J Cancer*, **118**:1699-
    712, 2006.