Building novartisA.Rda

Keith A. Baggerly

July 23, 2009

Contents

1 Executive Summary **2**
1.1 Introduction . . . 2
1.2 Methods . . . 2
1.3 Results . . . 2

2 Options and Libraries **2**

3 Loading and Parsing Data **2**
3.1 NovartisAll.Rda . . . 2
3.2 NCI60 Cell Line Info . . . 3
3.3 Lists of Cell Lines in Predictors . . . 3

4 Ordering Cell Line Panels **4**

5 Ordering Lines Within Panels **5**
5.1 Leukemia . . . 5
5.2 Central Nervous System . . . 6
5.3 Breast . . . 6
5.4 Colon . . . 7
5.5 Non-Small Cell Lung Cancer . . . 8
5.6 Melanoma . . . 9
5.7 Ovarian . . . 9
5.8 Prostate . . . 10
5.9 Renal . . . 10

6 The Final Ordering **11**

7 Extracting and Ordering Quantifications and Info **11**
7.1 The Quantifications . . . 11
7.2 The Info . . . 12

8 Save Rda File **12**

9 Appendix **12**
9.1 File Location 12
9.2 Saves 12
9.3 SessionInfo 12

List of Figures

List of Tables

1 Executive Summary

1.1 Introduction

In this report, we extract the "A" set replicates from the complete set of Novartis U95A NCI-60 quantifications.

1.2 Methods

We loaded the previously assembled novartisAll Rda file. We loaded cell line information from the drug sensitivity data. We loaded the most recent lists of cell lines in each drug predictor from the Potti et al. [1] web site.

We extracted the A set quantifications from the full data matrix, and used the orderings of cell lines in the reported signature lists to reorder the data columns in the quantification matrix.

1.3 Results

We created a "novartisA" matrix of array quantifications and an identically ordered "novartisAInfo" data frame of sample information. We stored these in RDataObjects as "novartisA.Rda."

2 Options and Libraries

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

3 Loading and Parsing Data

3.1 NovartisAll.Rda

We begin by loading the novartisAll Rda file assembled earlier.

```
> rdaList <- c("novartisAll")
> for (rdaFile in rdaList) {
+     rdaFullFile <- file.path("RDataObjects", paste(rdaFile, "Rda",
+         sep = "."))
+     if (file.exists(rdaFullFile)) {
+         cat("loading ", rdaFullFile, " from cache\n")
+         load(rdaFullFile)
```

```
+     }
+     else {
+         cat("building ", rdaFullFile, " from raw data\n")
+         Stangle(file.path("RNowebSource", paste("buildRda", rdaFile,
+             "Rnw", sep = ".")))
+         source(paste("buildRda", rdaFile, "R", sep = "."))
+     }
+ }

loading  RDataObjects/novartisAll.Rda  from cache
```

3.2 NCI60 Cell Line Info

Next, we load in information about the NCI-60 cell lines extracted from the drug sensitivity information.

```
> nci60Info <- read.table(file.path("RawData", "NCI60", "nci60Info.csv"),
+     sep = ",", header = TRUE, colClasses = c("character", "character",
+         "numeric", "numeric"), row.names = 1)
> nci60Info[1:3, ]

                       PANEL PANELNBR CELLNBR
NCI-H23   Non-Small Cell Lung        1       1
NCI-H522  Non-Small Cell Lung        1       3
A549/ATCC Non-Small Cell Lung        1       4
```

3.3 Lists of Cell Lines in Predictors

Next, we load in the most recent (August 2008) lists of cell lines in each drug prediction signature available from the Potti et al. [1] web site, `http://data.genome.duke.edu/NatureMedicine.php`. We reorganized the contents of the relevant file, "Celllines_in_each_predictor1.xls," as a csv file for easier loading and parsing.

```
> signatureCellLines <- read.table(file.path("RawData", "PottiNatMed",
+     "cellLinesInEachPredictor1_reorg.csv"), header = TRUE, sep = ",")
> signatureCellLines[1:3, ]

  CellLine    Status      Drug
1     EKVX Resistant Docetaxel
2   IGROV1 Resistant Docetaxel
3  OVCAR-4 Resistant Docetaxel

> which(is.na(match(signatureCellLines[, "CellLine"], rownames(nci60Info))))

[1] 124

> dim(signatureCellLines)

[1] 124   3

> signatureCellLines[123:124, ]
```

```
    CellLine    Status         Drug
123    SN12C Resistant Fluorouracil
124  OVKAR-8 Resistant Fluorouracil
```

```
> signatureCellLines[124, "CellLine"] <- "OVCAR-8"
> signatureCellLines[, "CellLine"] <- as.factor(as.character(signatureCellLines[,
+     "CellLine"]))
```

There is a variant spelling for the name of the last cell line involved, but this is easily fixed.

4 Ordering Cell Line Panels

Now we want to check the ordering of the cell line panels in the signatures used by Potti et al. [1]. We use the panel numbers for this purpose as opposed to the panel names, because the panel names for some cell lines (e.g., MDA-MB-435) have changed over time.

```
> panelNumbers <- tapply(nci60Info[as.character(signatureCellLines[,
+     "CellLine"]), "PANELNBR"], signatureCellLines[, c("Status",
+     "Drug")], FUN = function(x) {
+     x
+ })
> panelNumbers["Resistant", ]

$Cyclophosphamide
[1] 12  5  5  5  1 10 10 10

$Docetaxel
[1] 1 6 6 9 9 9 9

$Doxorubicin
 [1] 5 4 4 1 1 6 6 6 6 6 6 9

$Etoposide
[1]  5  5  4  4  1 10  6  6

$Fluorouracil
[1]  5  5  4  1 10  9  6

$Paclitaxel
[1]  7  4  1  1 10 10  6  9

$Topotecan
 [1]  7  7  5  5  4  4  4  1 10  4

> panelNumbers["Sensitive", ]

$Cyclophosphamide
[1] 7 7 7 5 4 4 1 9
```

```
$Docetaxel
[1]  7 12  4  1 10 10  1

$Doxorubicin
 [1] 12 12  5  1 10 10 10 10 10 10

$Etoposide
[1] 12  5  5  4  1  1 10 11  9

$Fluorouracil
[1]  5  4  4  1 10 10  9  9

$Paclitaxel
[1] 12 12  5  5  4  4  4  6  1

$Topotecan
 [1] 12 12 12  5  1  1  1 10  6  9  9  9  9
```

But for a few exceptions, the panel order used appears to be

$$7 > 12 > 5 > 4 > 1 > 10 > (6, 11) > 9,$$

corresponding to Leukemia (Blood), Central Nervous System (Brain), Breast, Colon, Non-Small Cell Lung (Lung), Melanoma, Ovarian, Prostate, Renal.

5 Ordering Lines Within Panels

5.1 Leukemia

```
> nci60Info[nci60Info[, "PANELNBR"] == 7, ]
```

```
             PANEL PANELNBR CELLNBR
CCRF-CEM  Leukemia        7       3
K-562     Leukemia        7       5
MOLT-4    Leukemia        7       6
HL-60(TB) Leukemia        7       8
RPMI-8226 Leukemia        7      10
SR        Leukemia        7      19
```

```
> signatureCellLines[nci60Info[as.character(signatureCellLines[,
+     "CellLine"]), "PANELNBR"] == 7, ]
```

```
    CellLine    Status            Drug
8  HL-60(TB) Sensitive        Docetaxel
45     K-562 Resistant        Topotecan
46 RPMI-8226 Resistant        Topotecan
64  CCRF-CEM Resistant       Paclitaxel
94     K-562 Sensitive Cyclophosphamide
95    MOLT-4 Sensitive Cyclophosphamide
96 HL-60(TB) Sensitive Cyclophosphamide
```

The ordering is consistent with CELLNBR.

5.2 Central Nervous System

```
> nci60Info[nci60Info[, "PANELNBR"] == 12, ]
```

```
                       PANEL PANELNBR CELLNBR
SNB-19 Central Nervous System       12       2
SNB-75 Central Nervous System       12       5
U251   Central Nervous System       12       9
SF-268 Central Nervous System       12      14
SF-295 Central Nervous System       12      15
SF-539 Central Nervous System       12      16
```

```
> signatureCellLines[nci60Info[as.character(signatureCellLines[,
+     "CellLine"]), "PANELNBR"] == 12, ]
```

```
    CellLine    Status             Drug
9     SF-539 Sensitive        Docetaxel
15    SF-539 Sensitive        Etoposide
32    SF-539 Sensitive        Topotecan
33    SNB-75 Sensitive        Topotecan
34      U251 Sensitive        Topotecan
55    SF-295 Sensitive       Paclitaxel
56    SF-539 Sensitive       Paclitaxel
72    SF-539 Sensitive      Doxorubicin
73    SNB-75 Sensitive      Doxorubicin
102   SNB-19 Resistant Cyclophosphamide
```

Here, CELLNBR doesn't seem to track very well. Alphabetic ordering does, so we'll use that.

5.3 Breast

```
> nci60Info[nci60Info[, "PANELNBR"] == 5, ]
```

```
                  PANEL PANELNBR CELLNBR
MCF7             Breast        5       1
NCI/ADR-RES      Ovarian       5       2
MDA-MB-231/ATCC   Breast       5       5
HS 578T           Breast       5       6
MDA-MB-435      Melanoma       5      11
BT-549            Breast       5      13
T-47D             Breast       5      14
```

```
> signatureCellLines[nci60Info[as.character(signatureCellLines[,
+     "CellLine"]), "PANELNBR"] == 5, ]
```

```
           CellLine    Status             Drug
16           BT-549 Sensitive        Etoposide
17  MDA-MB-231/ATCC Sensitive        Etoposide
24             MCF7 Resistant        Etoposide
25      NCI/ADR-RES Resistant        Etoposide
```

```
35          HS 578T Sensitive        Topotecan
47       MDA-MB-435 Resistant        Topotecan
48  MDA-MB-231/ATCC Resistant        Topotecan
57          HS 578T Sensitive       Paclitaxel
58       MDA-MB-435 Sensitive       Paclitaxel
74       MDA-MB-435 Sensitive      Doxorubicin
82      NCI/ADR-RES Resistant      Doxorubicin
97             MCF7 Sensitive Cyclophosphamide
103         HS 578T Resistant Cyclophosphamide
104 MDA-MB-231/ATCC Resistant Cyclophosphamide
105      MDA-MB-435 Resistant Cyclophosphamide
110            MCF7 Sensitive     Fluorouracil
118     NCI/ADR-RES Resistant     Fluorouracil
119      MDA-MB-435 Resistant     Fluorouracil
```

Again, alphabetic order looks better than CELLNBR. We have conflicting information about the relative placement of MDA-MB-231/ATCC and MDA-MB-435 from the Topotecan Resistant group ordering and the Cyclophosphamide Resistant group ordering, so we leave that one alone. We place NCI/ADR-RES before MDA-MB-435 based on the Fluorouracil Resistant ordering.

5.4 Colon

```
> nci60Info[nci60Info[, "PANELNBR"] == 4, ]
```

```
         PANEL PANELNBR CELLNBR
HT29     Colon        4       1
HCC-2998 Colon        4       2
HCT-116  Colon        4       3
SW-620   Colon        4       9
COLO 205 Colon        4      10
HCT-15   Colon        4      15
KM12     Colon        4      17
```

```
> signatureCellLines[nci60Info[as.character(signatureCellLines[,
+     "CellLine"]), "PANELNBR"] == 4, ]
```

```
    CellLine    Status            Drug
10      HT29 Sensitive       Docetaxel
18  HCC-2998 Sensitive       Etoposide
26    HCT-15 Resistant       Etoposide
27    SW-620 Resistant       Etoposide
49  HCC-2998 Resistant       Topotecan
50   HCT-116 Resistant       Topotecan
51    HCT-15 Resistant       Topotecan
54  COLO 205 Resistant       Topotecan
59  COLO 205 Sensitive      Paclitaxel
60  HCC-2998 Sensitive      Paclitaxel
61      HT29 Sensitive      Paclitaxel
65    SW-620 Resistant      Paclitaxel
```

```
83    HCT-15 Resistant      Doxorubicin
84      HT29 Resistant      Doxorubicin
98  HCC-2998 Sensitive Cyclophosphamide
99   HCT-116 Sensitive Cyclophosphamide
111 COLO 205 Sensitive     Fluorouracil
112  HCT-116 Sensitive     Fluorouracil
120   SW-620 Resistant     Fluorouracil
```

There are ambiguities here, so we choose an alphabetic ordering.

5.5 Non-Small Cell Lung Cancer

```
> nci60Info[nci60Info[, "PANELNBR"] == 1, ]
```

```
                       PANEL PANELNBR CELLNBR
NCI-H23   Non-Small Cell Lung        1       1
NCI-H522  Non-Small Cell Lung        1       3
A549/ATCC Non-Small Cell Lung        1       4
EKVX      Non-Small Cell Lung        1       8
NCI-H226  Non-Small Cell Lung        1      13
NCI-H322M Non-Small Cell Lung        1      17
NCI-H460  Non-Small Cell Lung        1      21
HOP-62    Non-Small Cell Lung        1      26
HOP-92    Non-Small Cell Lung        1      29
```

```
> signatureCellLines[nci60Info[as.character(signatureCellLines[,
+     "CellLine"]), "PANELNBR"] == 1, ]
```

```
     CellLine    Status            Drug
1        EKVX Resistant        Docetaxel
11     HOP-62 Sensitive        Docetaxel
14   NCI-H522 Sensitive        Docetaxel
19     HOP-62 Sensitive        Etoposide
20   NCI-H226 Sensitive        Etoposide
28  NCI-H322M Resistant        Etoposide
36     HOP-62 Sensitive        Topotecan
37   NCI-H226 Sensitive        Topotecan
38    NCI-H23 Sensitive        Topotecan
52  NCI-H322M Resistant        Topotecan
63   NCI-H522 Sensitive       Paclitaxel
66  A549/ATCC Resistant       Paclitaxel
67       EKVX Resistant       Paclitaxel
75    NCI-H23 Sensitive      Doxorubicin
85       EKVX Resistant      Doxorubicin
86  NCI-H322M Resistant      Doxorubicin
100  NCI-H460 Sensitive Cyclophosphamide
106  NCI-H226 Resistant Cyclophosphamide
113  NCI-H460 Sensitive     Fluorouracil
121      EKVX Resistant     Fluorouracil
```

The data is consistent with alphabetic ordering, so we use that.

5.6 Melanoma

```
> nci60Info[nci60Info[, "PANELNBR"] == 10, ]
```

```
             PANEL PANELNBR CELLNBR
LOX IMVI  Melanoma       10       1
MALME-3M  Melanoma       10       2
SK-MEL-2  Melanoma       10       5
SK-MEL-5  Melanoma       10       7
SK-MEL-28 Melanoma       10       8
M14       Melanoma       10      14
UACC-62   Melanoma       10      20
UACC-257  Melanoma       10      21
```

```
> signatureCellLines[nci60Info[as.character(signatureCellLines[,
+     "CellLine"]), "PANELNBR"] == 10, ]
```

```
     CellLine    Status             Drug
12   SK-MEL-2 Sensitive        Docetaxel
13   SK-MEL-5 Sensitive        Docetaxel
21        M14 Sensitive        Etoposide
29   UACC-257 Resistant        Etoposide
39   LOX IMVI Sensitive        Topotecan
53  SK-MEL-28 Resistant        Topotecan
68   MALME-3M Resistant       Paclitaxel
69  SK-MEL-28 Resistant       Paclitaxel
76        M14 Sensitive      Doxorubicin
77   MALME-3M Sensitive      Doxorubicin
78   SK-MEL-2 Sensitive      Doxorubicin
79  SK-MEL-28 Sensitive      Doxorubicin
80   SK-MEL-5 Sensitive      Doxorubicin
81    UACC-62 Sensitive      Doxorubicin
107       M14 Resistant Cyclophosphamide
108  MALME-3M Resistant Cyclophosphamide
109  SK-MEL-2 Resistant Cyclophosphamide
114  LOX IMVI Sensitive     Fluorouracil
115  SK-MEL-5 Sensitive     Fluorouracil
122       M14 Resistant     Fluorouracil
```

Alphabetic looks fine here.

5.7 Ovarian

```
> nci60Info[nci60Info[, "PANELNBR"] == 6, ]
```

```
          PANEL PANELNBR CELLNBR
OVCAR-3 Ovarian        6       1
OVCAR-4 Ovarian        6       2
OVCAR-5 Ovarian        6       3
OVCAR-8 Ovarian        6       5
```

```
IGROV1  Ovarian        6      10
SK-OV-3 Ovarian        6      11
```

```
> signatureCellLines[nci60Info[as.character(signatureCellLines[,
+     "CellLine"]), "PANELNBR"] == 6, ]
```

```
    CellLine    Status         Drug
2     IGROV1 Resistant    Docetaxel
3    OVCAR-4 Resistant    Docetaxel
30   OVCAR-4 Resistant    Etoposide
31   OVCAR-5 Resistant    Etoposide
40   OVCAR-8 Sensitive    Topotecan
62   OVCAR-3 Sensitive   Paclitaxel
70   OVCAR-8 Resistant   Paclitaxel
87    IGROV1 Resistant  Doxorubicin
88   OVCAR-3 Resistant  Doxorubicin
89   OVCAR-4 Resistant  Doxorubicin
90   OVCAR-5 Resistant  Doxorubicin
91   OVCAR-8 Resistant  Doxorubicin
92   SK-OV-3 Resistant  Doxorubicin
124  OVCAR-8 Resistant Fluorouracil
```

Alphabetic looks fine here.

5.8 Prostate

```
> nci60Info[nci60Info[, "PANELNBR"] == 11, ]
```

```
          PANEL PANELNBR CELLNBR
PC-3   Prostate       11       1
DU-145 Prostate       11       3
```

```
> signatureCellLines[nci60Info[as.character(signatureCellLines[,
+     "CellLine"]), "PANELNBR"] == 11, ]
```

```
   CellLine    Status      Drug
22     PC-3 Sensitive Etoposide
```

No real evidence here, but we'll go with alphabetic in keeping with most of the rest.

5.9 Renal

```
> nci60Info[nci60Info[, "PANELNBR"] == 9, ]
```

```
        PANEL PANELNBR CELLNBR
UO-31   Renal        9       4
SN12C   Renal        9       8
A498    Renal        9      13
CAKI-1  Renal        9      15
RXF 393 Renal        9      16
```

```
786-0   Renal        9      18
ACHN    Renal        9      23
TK-10   Renal        9      24
```

```
> signatureCellLines[nci60Info[as.character(signatureCellLines[,
+     "CellLine"]), "PANELNBR"] == 9, ]
```

```
    CellLine    Status             Drug
4      786-0 Resistant        Docetaxel
5     CAKI-1 Resistant        Docetaxel
6      SN12C Resistant        Docetaxel
7      TK-10 Resistant        Docetaxel
23     786-0 Sensitive        Etoposide
41      A498 Sensitive        Topotecan
42      ACHN Sensitive        Topotecan
43    CAKI-1 Sensitive        Topotecan
44     UO-31 Sensitive        Topotecan
71     786-0 Resistant       Paclitaxel
93    CAKI-1 Resistant      Doxorubicin
101    TK-10 Sensitive Cyclophosphamide
116     A498 Sensitive     Fluorouracil
117    UO-31 Sensitive     Fluorouracil
123    SN12C Resistant     Fluorouracil
```

Alphabetic works here.

6 The Final Ordering

```
> tempOrder <- c("CCRF-CEM", "K-562", "MOLT-4", "HL-60(TB)", "RPMI-8226",
+     "SR", "SF-268", "SF-295", "SF-539", "SNB-19", "SNB-75", "U251",
+     "BT-549", "HS 578T", "MCF7", "MDA-MB-231/ATCC", "NCI/ADR-RES",
+     "MDA-MB-435", "T-47D", "COLO 205", "HCC-2998", "HCT-116",
+     "HCT-15", "HT29", "KM12", "SW-620", "A549/ATCC", "EKVX",
+     "HOP-62", "HOP-92", "NCI-H226", "NCI-H23", "NCI-H322M", "NCI-H460",
+     "NCI-H522", "LOX IMVI", "M14", "MALME-3M", "SK-MEL-2", "SK-MEL-28",
+     "SK-MEL-5", "UACC-257", "UACC-62", "DU-145", "PC-3", "IGROV1",
+     "OVCAR-3", "OVCAR-4", "OVCAR-5", "OVCAR-8", "SK-OV-3", "786-0",
+     "A498", "ACHN", "CAKI-1", "RXF 393", "SN12C", "TK-10", "UO-31")
```

7 Extracting and Ordering Quantifications and Info

7.1 The Quantifications

```
> novartisA <- novartisAll[, grep("A$", colnames(novartisAll))]
> colnames(novartisA) <- substr(colnames(novartisA), 1, nchar(colnames(novartisA)) -
+     2)
> novartisA <- novartisA[, tempOrder]
> dim(novartisA)
```

```
[1] 12625    59
```

```
> novartisA[1:3, 1:3]
```

```
          CCRF-CEM     K-562    MOLT-4
36460_at  41.16584  95.75866  68.12313
36461_at  80.50482  98.03310  97.47627
36462_at 113.68166 200.20106 248.60211
```

7.2 The Info

```
> novartisAInfo <- nci60Info[colnames(novartisA), ]
```

8 Save Rda File

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

```
> save(novartisA, novartisAInfo, file = file.path("RDataObjects",
+     "novartisA.Rda"))
```

9 Appendix

9.1 File Location

```
> getwd()
```

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

9.2 Saves

9.3 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
```

References

[1] Potti A, Dressman HK, Bild A, et al: Genomic signatures to guide the use of chemotherapeutics. *Nat Med*, **12**:1294-1300, 2006.