Checking the Clinical Information for Docetaxel

Keith A. Baggerly and Kevin R. Coombes

November 13, 2007

1 Introduction

In their reply to our correspondence, Potti and Nevins note that there is now more supplementary info on their website, `http://data.cgt.duke.edu/NatureMedicine.php`, which should hopefully allow some of our disagreements to be resolved.

With respect to docetaxel, one of the first of these new files is "Chang clinical data.doc", which gives the clinical information (including the sensitive/resistant status) that they are applying to the samples at hand. This clinical information is relevant because it supplies an explantion for why Potti et al use 13 sensitive and 11 resistant samples when Chang et al report 11 sensitive and 13 resistant. Potti et al define samples to be "sensitive" if there is less than 40% residual disease, whereas Chang et al use a cutoff of 25% residual disease.

The clinical information that Chang et al provide is found in two places:

1. in Table 1 of Chang et al, and

2. in the table of supplementary information for Chang et al available from the Lancet website, `http://image.thelancet.com/extras/01art11086webtable.pdf`.

The latter table gives the percent residual disease, and expression values for 92 probesets for each of the 24 samples profiled.

Here, we simply load the information to check the results. For all 3 of the files (the clinical info doc file from Duke, and the two Chang et al Lancet tables) we have converted the contents to csv files for easier loading.

2 Options and Libraries

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

3 Checking Data

We begin with the basic clinical info found in Table 1 of Chang et al.

```
> changTable1 <- read.table(file.path("OtherData", "changTable1.csv"),
+     sep = ",", header = TRUE)
> changTable1[1:3, ]
```

```
  Patient Age..years. Menopausal.status Ethnic.origin
1       1          37      Premenopausal       Hispanic
2       2          55     Postmenopausal       Hispanic
3       3          41      Premenopausal          Black
  Bidimensional.tumour.size..cm. Clinical.axillary.nodes
1                          10x10                       No
2                           10x8                      Yes
3                            6x5                      Yes
  Oestrogen..receptor.status Progesterone..receptor.status HER.2 Tumour.type
1                          -                             -     -          IMC
2                          -                             -     +          IDC
3                          +                             +     -          IDC

> changClinicalFromDuke <- read.table(file.path("DukeWebSite",
+     "changClinicalFromDuke.csv"), sep = ",", header = TRUE)
> changClinicalFromDuke[1:3, ]

  Patient Age..years. Menopausal.status Ethnic.origin
1       1          37      Premenopausal       Hispanic
2       2          55     Postmenopausal       Hispanic
3       3          41      Premenopausal          Black
  Bidimensional.tumour.size..cm. Clinical.axillary.nodes
1                          10x10                       No
2                           10x8                      Yes
3                            6x5                      Yes
  Oestrogen..receptor.status Progesterone..receptor.status HER.2 Tumour.type
1                          -                             -     -          IMC
2                          -                             -     +          IDC
3                          +                             +     -          IDC
  Percent.Residual.Tumor Sens.or.Res..S.is.less.than.40pct.
1                      1                                    S
2                      1                                    S
3                      6                                    S

> all(changTable1 == changClinicalFromDuke[, 1:10])

[1] TRUE
```

There are 10 columns of data in Chang et al's Table 1. These are the the first 10 columns in the clinical information from Duke (in the same order), and all of the values agree.

Next we check the information on (a) Sensitive/Resistant status and (b) percent residual disease, both of which are given in the supplementary table.

```
> changSupp <- read.table(file.path("OtherData", "01art11086webtable_rev.csv"),
+     sep = ",")
> dim(changSupp)

[1] 96 29

> changSupp[1:5, ]
```

```
                        V1          V2          V3                    V4
1           Sample number
2       Residual tumour, %
3 Sensitive or resistant
4                 Probe set  GenBank  Locus Link  Official Symbol
5                 1008_f_at  U50648         5610             PRKR
                                                               V5      V6
1                                                                      N1
2                                                                       1
3                                                                       S
4                                                              Gene name
5 protein kinase, interferon-inducible double stranded RNA dependent  247.45
      V7     V8     V9    V10    V11    V12    V13    V14    V15    V16    V17
1     N2     N3     N4     N5     N6     N7     N8     N9    N10    N11    N12
2      1      6      6     13     14     16     17     18     22     25     36
3      S      S      S      S      S      S      S      S      S      S      R
4
5 149.39 321.16 301.74 765.01 659.16 462.4 816.97 385.68 398.58 227.32 355.79
      V18    V19    V20    V21    V22    V23    V24    V25    V26    V27
1     N13    N14    N15    N16    N17    N18    N19    N20    N21    N22
2      38     39     44     45     47     60     64     65     70    100
3      R      R      R      R      R      R      R      R      R      R
4
5 1392.68 1069.56 1118.46 1580.86 739.08 502.23 772.62 929.03 1892.94 406.26
     V28    V29
1    N23    N24
2    100    131
3      R      R
4
5 528.1 638.04
```

The supplementary table has four header lines of explanatory information, but not all headers are relevant for all columns. The first five columns of the supplementary table give identifying information for the Affymetrix probesets described (Probe Set, GenBank, Locus Link, Official Symbol, Gene Name) and the last 24 columns give the sample specific information. What we want here are the first three rows of header information for the last 24 (sample specific) columns.

```
> temp <- t(changSupp[1:3, 6:29])
> colnames(temp) <- t(as.character(changSupp[1:3, 1]))
> temp[c(1, 2, 23, 24), ]

    Sample number  Residual tumour, %  Sensitive or resistant
V6  "N1 "          "1"                 "S "
V7  "N2 "          "1"                 "S "
V28 "N23 "         "100"               "R "
V29 "N24 "         "131"               "R "

> temp[, 3] <- substr(temp[, 3], 1, 1)
> temp[, 1] <- substr(temp[, 1], 1, nchar(temp[, 1]) - 1)
> temp[c(1, 2, 23, 24), ]
```

```
    Sample number  Residual tumour, %  Sensitive or resistant
V6  "N1"           "1"                 "S"
V7  "N2"           "1"                 "S"
V28 "N23"          "100"               "R"
V29 "N24"          "131"               "R"

> changSuppClinical <- as.data.frame(temp, stringsAsFactors = FALSE)
> changSuppClinical[, 2] <- as.numeric(changSuppClinical[, 2])
> rownames(changSuppClinical) <- NULL
> changSuppClinical[1:3, ]

  Sample number  Residual tumour, %  Sensitive or resistant
1           N1                   1                        S
2           N2                   1                        S
3           N3                   6                        S
```

The supplementary table labels the samples N1 through N24, as opposed to 1 through 24 in Table 1, but other than that these look equivalent.

```
> all(changSuppClinical[, "Sample number"] == paste("N", changClinicalFromDuke[,
+     "Patient"], sep = ""))

[1] TRUE
```

For the Sensitive/Resistant calls, we expect to see two differences between the tables, corresponding to the two relabeled samples.

```
> which(changSuppClinical[, "Sensitive or resistant "] != changClinicalFromDuke[,
+     "Sens.or.Res..S.is.less.than.40pct."])

[1] 12 13
```

We expect the percent residual tumor values to line up across tables.

```
> which(changSuppClinical[, "Residual tumour, % "] != changClinicalFromDuke[,
+     "Percent.Residual.Tumor"])

[1] 14

> changSuppClinical[11:15, ]

   Sample number  Residual tumour, %  Sensitive or resistant
11           N11                  25                        S
12           N12                  36                        R
13           N13                  38                        R
14           N14                  39                        R
15           N15                  44                        R

> changClinicalFromDuke[11:15, c(1, 11, 12)]
```

```
    Patient Percent.Residual.Tumor Sens.or.Res..S.is.less.than.40pct.
11       11                     25                                    S
12       12                     36                                   S*
13       13                     38                                   S*
14       14                     40                                    R
15       15                     44                                    R
```

Unexpectedly, there is a discrepancy. The percent residual tumor for patient 14 has changed from 39 percent (Chang) to 40 percent (Potti). This does have a minor effect on classification, as 40 percent was set as the cutoff for the Sensitive/Resistant divide. Using the Duke cutoff with the numbers from Chang et al says there should be 14 Sensitive patients, not 13.

4 Appendix

4.1 Saves

4.2 SessionInfo

```
> sessionInfo()

R version 2.5.1 (2007-06-27)
i386-pc-mingw32

locale:
LC_COLLATE=English_United States.1252;LC_CTYPE=English_United States.1252;LC_MONETARY=English_United Sta

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

other attached packages:
R.matlab     R.oo
 "1.1.3"  "1.3.0"
```