Reproducibility of Dressman 2007

VJ Carey

July 20, 2010

The dispute between Baggerly and Coombes (BC) and Dressman, Potti and Nevins (DPN) in J Clin Oncology, 2008; 26(7):1186-1187, is of broad interest. Because general reproducibility of genome-scale data analysis has been questioned at high levels (cite Ioannidis and others), specific debates on difficulties of reproduction of analytic findings should be examined carefully. Study of these debates can reveal patterns of analysis and interpretation that should be avoided to reduce risks of conflicts over reproducibility, or promoted to increase ease with which new findings can be prioritized for application and extension. I have reviewed BC's supplementary materials with some care, to see whether a concise exposition of the conflict can be achieved. Briefly, in their original paper, Dressman et al. (JCO 25(5):517-25) showed that among individuals with ovarian tumors non-responsive to platinum, activation of Src or E2F3 pathways was significantly associated with survival (Figs 2B and 2C, p521).

Below I present figures illustrating partial reproducibility of Dressman's original analyses, computed independently of the materials assembled by BC. Figure 1(a) shows the association between Src dysregulation and survival among non-responders using Dressman et al.'s 'corrected RMA' from the supplementary web site `http://data.cgt.duke.edu/platinum.php` (retrieved April 4 2009) with sample identifiers given in that file. Figure 1(b) shows the same association after sample identifiers are redefined using the method of BC to map Dressman's quantifications to the original CEL files which are assumed to be accurately labeled to correspond to records in the clinical data. Figure 1(b) is extremely similar to Dressman's original Figure 2B, and shows that BC are correct when they say that the labels in the 'corrected RMA' archive need to be revised. The mild discrepancies between Figure 1(b) and Dressman's original 2B might be explained through differences in tumor scoring coefficients, or through differences in exact numbers of patients/events available – only 116 of 119 transcript profiles could be reliably mapped to CEL files for relabeling. In any case, Figure 1(b) suggests that we can use the public archive, with some adjustments, to technically reproduce Dressman et al.'s original findings to a reasonable approximation.

Figures 1(c) and 1(d) are more troubling. Figure 1(c) reproduces the methods for Figure 1(b) in application to E2F3 activation. Among patients labeled as nonresponders by Dressman in the public archive, there is no association between pathway activation and survival. Figure 1(d) shows that there is an association, but among the responders.

These findings are congruent with those of BC, who explored more variations on the data sources and could not recover the finding of the original Figure 2C.

We now have two basic problems. First, ignoring concerns about confounding, and using only (and all) quantifications provided by Dressman, and thereby 'following their methods', we cannot reproduce Figure 2C. Second, the significance of the association indicated in my Figure 1(b) is lost when a simple allowance for batch effects is made in the test for different survival distributions by pathway activation.

```
Call:
survreg(formula = Surv(Survival, dead) ~ sdys + poly(chron(rundate),
    2), subset = CR == 0)
                          Value Std. Error      z        p
(Intercept)               3.767      0.298 12.657 1.02e-36
sdys                     -0.292      0.402 -0.726 4.68e-01
poly(chron(rundate), 2)1  5.805      1.929  3.009 2.62e-03
poly(chron(rundate), 2)2 -1.567      1.711 -0.916 3.60e-01
Log(scale)               -0.250      0.168 -1.489 1.36e-01

Scale= 0.779

Weibull distribution
Loglik(model)= -109   Loglik(intercept only)= -115.2
        Chisq= 12.32 on 3 degrees of freedom, p= 0.0064
Number of Newton-Raphson Iterations: 6
n= 34
```
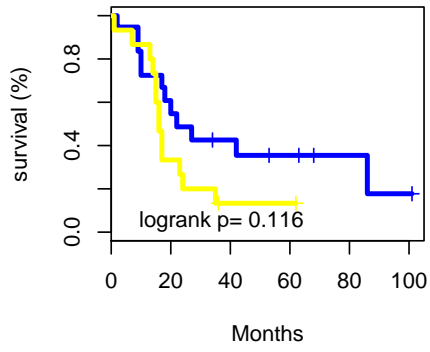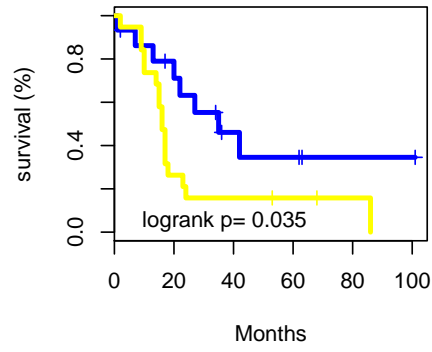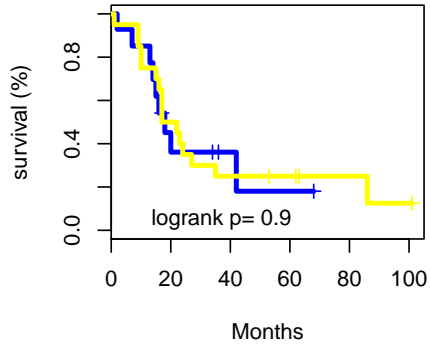
(a)

survival (%)

logrank p= 0.116

Months

(b)

survival (%)

logrank p= 0.035

Months

(c)

survival (%)

logrank p= 0.9

Months

(d)

survival (%)

logrank p= 0.001

Months

3