Statistical analysis of SELDI protein chip data from breast cancer cell lines exposed to chemotherapeutic agents

Ian L. Dryden*[1], Shahid Mian[1,2], William J. Browne[1], Kelly Handley[1], Riccardo di Nisio[3] and Robert Rees[2]

[1] University of Nottingham
[2] Nottingham Trent University
[3] University of Chieti, Pescara, Italy.

1 Introduction

Identifying molecular biochemical pathways that serve to either limit or enhance cellular responsiveness towards chemotherapy agents is a fundamental goal in the drug discovery process. Selectively targeting pathways that result in enhanced killing, and possibly greater specificity, could eventually lead to enhanced chemotherapy efficacy. To address this issue we have conducted a proof of principle study using surface enhanced laser desorption/ionisation time-of-flight mass spectrometry (SELDI-TOF) to examine protein expression patterns of resistant and sensitive breast cancer cell lines that have received/not received a 24 hour exposure to the chemotherapeutic agent Taxol. We consider one resistant breast cancer line (MCF7/ADR) and two sensitive breast cancer cell lines (MCF7 and T47D). The ability to classify cellular populations at a molecular level based upon their phenotypic response to chemotherapeutic agents could provide a first step in the identification of molecules/pathways responsible for chemotherapy sensitivity/resistance.

2 The experiment and SELDI mass spectrometry

The breast cancer cell lines MCF-7, T47D and MCF-7/ADR were seeded at 400,000 cells per T25 cm^2 tissue culture flask and Taxol (Paclitaxel) added to a final concentration of 10^{-7} M. In total 18 flasks were used for control and 18 flasks for drug treatment, consisting of 3 flasks for each treatment regimen which were sacrificed at 24 hour intervals post drug treatment up to and including 96 hours. The experiment was repeated twice. Homogenates were prepared (as described in Mian *et al.*, 2003) and protein concentrations were determined using the Bradford method (Biorad) as per manufacturer's instructions. SELDI mass spectrometry was conducted as outlined previously (Ball *et al.*, 2002; Mian *et al.*, 2003). We remove the m/z values up to 2000 Daltons as background interference from sinapinic acid matrix peaks tend to produce low signal:noise ratios (Ball *et al.*, 2002). In summary there are six groups: ADC (MCF-7/ADR control), ADT (MCF-7/ADR Taxol treated), TDC (T47D control), TDT (T47D Taxol treated), MCC (MCF-7 control), MCT (MCF-7 Taxol treated). SELDI-TOF scans were taken at periods of 24,48,72 and 96 hours, and these times are labelled as Day 1, Day 2, Day 3 and Day 4 respectively. For each group on each day there are 6 observations: replicates A,B,C for experiment 1 and replicates A,B,C for experiment 2.

3 The offset Hotelling T^2 test

We wish to consider which m/z values give rise to differences between Control vs Taxol treated for each cell-line, and which m/z values are different for chemo-resistant (MCF-7/ADR) vs chemo-sensitive (MCF-7 and T47D) cell-lines. Let $\bar{x}_{Ai}$ and $\bar{x}_{Bi}$ be the $q-$vector means in groups A and B respectively, at m/z value i ($i = 1, \ldots, p$), with sample sizes n_A, n_B [for us $q = 4$ is the number of days, and $p = 13951$ is the number of data points between 2000 and 30000 Daltons]. Let S_{xi} be the unbiased pooled within-group $q \times q$ covariance matrix at m/z value i, ($i = 1, \ldots, p$). The conventional two sample Hotelling T^2 test (cf. Mardia *et al.*, 1979, p139) for testing $H_0 : \mu_{Ai} = \mu_{Bi}$ versus $H_1 : \mu_{Ai} \neq \mu_{Bi}$ assumes independent multivariate normal distributed data in each group with means μ_{Ai} and μ_{Bi} and common population covariance matrix. The test statistic for m/z value i is $T^2_{x,i} = (\bar{x}_{Ai} - \bar{x}_{Bi})^T S_{xi}^{-1} (\bar{x}_{Ai} - \bar{x}_{Bi})$ and we reject H_0 in favour of H_1 at the $100\alpha\%$ level if

$$T^2_{x,i} > T_{crit}(\alpha) = \frac{(n_A + n_B)(n_A + n_B - 2)q}{n_A n_B (n_A + n_B - q - 1)} F_{q, n_A + n_B - q - 1}(1 - \alpha),$$

where $F_{\nu_1,\nu_2}(1-\alpha)$ is the $1-\alpha$ quantile of the F_{ν_1,ν_2} distribution.

In a preliminary analysis of the data it is clear that some m/z values have extremely small variability in the samples, especially within an experiment. Hence, only a very small difference in mean value in different groups can lead to a statistically significant difference at that m/z value. This under-dispersion of the data could be due to the small number of separate experiments conducted (just two). It seems highly plausible that there would be more variability if further experiments were carried out on different occasions. In addition, since each scan is high-dimensional we need to carry out a large number of simultaneous tests. Conventional multiple comparisons procedures (e.g. controlling the family wise error rate) lead to far too stringent protection against making a Type I error and so we need to consider alternatives.

We consider an adjustment to the Hotelling T^2 test statistic which is designed to account for the additional noise that we believe would be inherent in repetitions of the experiments on different occasions. The extra noise is taken to be i.i.d Gaussian with zero mean and variance σ^2. If x_i is the measured vector at m/z value i then we let the unobserved noisy vector be $w_i = x_i + \epsilon_i$ where $\epsilon_i \sim N_q(0, \sigma^2 I_q)$ independently. Note that

$$\bar{w}_{Ai} = \bar{x}_{Ai} + O_p(\sigma/\sqrt{n}) \ , \ \bar{w}_{Bi} = \bar{x}_{Bi} + O_p(\sigma/\sqrt{n}), S_{wi} = S_{xi} + \sigma^2 I_q + O_p(\sigma/\sqrt{n}).$$

We would like to consider the Hotelling T^2 statistics based on the unobserved w_i's:

$$T^2_{w,i} = (\bar{w}_{Ai} - \bar{w}_{Bi})^T S_{wi}^{-1} (\bar{w}_{Ai} - \bar{w}_{Bi}) = T_i^2(\sigma^2) + O_p(\sigma/\sqrt{n}),$$

where

$$T_i^2(\sigma^2) = (\bar{x}_{Ai} - \bar{x}_{Bi})^T (S_{xi} + \sigma^2 I_q)^{-1} (\bar{x}_{Ai} - \bar{x}_{Bi}).$$

However, $T^2_{w,i}$ is not observable, but we can observe $T_i^2(\sigma^2)$ given σ^2. We call $T_i^2(\sigma^2)$ the offset Hotelling T^2 statistics. We shall use these statistics for inference, rejecting H_0 if $T_i^2(\sigma^2) > T_{crit}(\alpha)$. The assumptions of the test are that both groups are independently multivariate normally distributed with the same covariance matrix, and $\sigma/\sqrt{n}$ is small.

The introduction of σ^2 plays two roles: 1) the additional variability results only in tests with larger mean differences being rejected, and 2) the value of σ^2 can be chosen so that $P(\text{at least one m/z value is significant}) = \alpha$ by calibrating the choice of σ^2 from random permutations of the groups. Hence, the multiple comparison problem is also addressed. The calibrated value $\hat{\sigma}^2$ is given by $\hat{\sigma}^2 = \text{arginf}\left\{P\left(\text{at least one } T_i^2(\sigma^2) > T_{\text{crit}}(\alpha)\right) \leq \alpha\right\}$.

A practical calibration method is as follows:

1. Choose α and initialize $\hat{\sigma}^2 = 0$.

2. For each i of n_I iterations: permute the group labels randomly with n_A observations assigned to group A and n_B observations assigned to group B, and calculate $T_i^2(\hat{\sigma}^2)$.

3. Let R be the observed proportion of times that at least one $T_i^2(\hat{\sigma}^2) > T_{crit}(\alpha)$. If $R > \alpha$ then increase $\hat{\sigma}^2$ by δ and return to step 2, else stop.

Note that for small samples all possible permutations can be used, rather than random permutations. The choice of increment δ is problem specific, and we choose $\delta = 2.5$ for our data.

4 Results and conclusions

We show the results of the offset Hotelling T^2 test for Controls versus Taxol treated in Figure 1. The value $\hat{\sigma}^2 = 7.5$ was chosen from calibration from $n_I = 500$ random permutations of the ADC and ADT groups giving $\alpha = 0.042$ ($\hat{\sigma}^2 = 5$ gave $\alpha = 0.07$). We used $\delta = 2.5$ as the step size for increasing σ^2 in the calibration algorithm, as stated earlier.

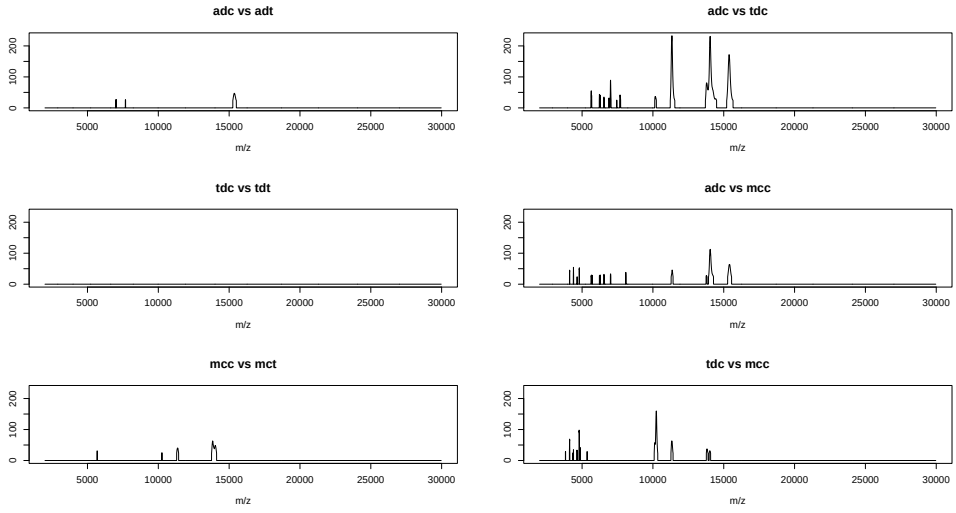

Figure 1: Significant offset Hotelling statistics in comparisons: (left) between the Control versus Taxol treated groups, and (right) between the Control groups.

In each plot in Figure 1 the values of $T_i^2(\sigma^2)$ are displayed only for the significant m/z values, giving an indication of which m/z values are significant in the pairwise group comparisons. The peaks are not just single spikes but rather peaks over a range of values with widths up to about 100 Daltons. Note that there is no overlap in the significant values for the ADC/ADT and MCC/MCT comparisons. Also, none of the TDC/TDT m/z values is significant. A number of significant peaks are indicated on the plots between the chemo-resistant versus chemo-sensitive comparisons (ADC/TDC and ADC/MCC). Also, some significant peaks are observed in the TDC vs MCC tests. Strong candidates for biomarkers for chemo-resistance versus chemo-sensitivity are in the range $11352 - 11369$, $14048 - 14055$, $15391 - 15402$.

Baggerly *et al.* (2004) have recently commented on the reproducibility of SELDI-TOF methodology. They raise many issues, including caution against interpretation of significance for the noisy low m/z values. We have removed the most problematic region in our study (1-2000 Daltons). The variability in the data is higher for low m/z values and lower for high m/z

values. An alternative to adding a single σ^2 offset throughout the range would be to add a decreasing function of m/z. This will be explored in further work.

We are currently exploring the use of Bayesian hierarchical modelling, where the SELDI TOF expression data are fitted with parametric functions with important geometrical features (peaks) captured through key parameters. Bayesian analysis can then be carried out using Markov chain Monte Carlo simulation, and this work will be reported in the poster later in this volume by Handley *et al* (p. 138). The Bayesian approach is considerably more computationally intensive than the procedure of the current paper. Note that both the method of the current paper and the hierarchical Bayesian approach are broadly similar in spirit to the 'peak probability contrast' method of Tibshirani *et al.* (2004), where large peaks are found in the data and then ranked in terms of discriminatory power.

Alternative paradigms for dealing with multiple comparisons include the False Discovery Rate (FDR) (Benjamini and Hochberg, 1995). For the six pairwise comparisons that we have carried out, the number of significant m/z values using a standard Hotelling T^2 test with $\sigma^2 = 0, FDR = 0.05$ gives 13 (ADC/ADT), 1 (TDT/TDC), 6527 (MCC/MCT), 8273 (ADC/TDC), 7128 (ADC/MCC) and 5931 (TDC/MCC) significant m/z values. Note that many of these significant values correspond to very small differences between the groups. Using $\sigma^2 = 15, FDR = 0.05$ leads to no significant values. Some judicial combination of the additional noise and FDR is well worth exploring in future work.

References

Baggerly, K. A., Morris, J. S., and Coombes, K. R. (2004). Reproducibility of SELDI-TOF protein patterns in serum: comparing data sets from different experiments. *Bioinformatics*, **20**, 777-785.

Ball, G., Mian, S., Holding, F., Allibone, R. O., Lowe, J., Ali, S., Li, G., McArdle, S., Ellis, I. O., Creaser, C., and Rees, R. C. (2002). An integrated approach utilizing artificial neural networks and seldi mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. *Bioinformatics*, **18**, 395-404.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, **57(1)**, 289-300.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press, London.

Mian, S., Ball, G., Hornbuckle, J., Holding, F., Carmichael, J., Ellis, I., Ali, S., Li, G., McArdle, S., Creaser, C., and Rees, R. C. (2003). A prototype methodology combining surface-enhanced laser desorption/ionization protein chip technology and artificial neural network algorithms to predict the chemoresponsiveness of breast cancer cell lines exposed to Paclitaxel and Doxorubicin under in vitro conditions. *Proteomics*, **3**, 1725-1737.

Tibshirani, R., Hastie, T., Narasimhan, B., Soltys, S., Shi, G., Koong, A., and Le, Q.-T. (2004). Sample classification from protein mass spectrometry, by 'peak probability contrasts'. *Bioinformatics*, **20**, 3034-3044.

Acknowledgement: This work was partially supported by EPSRC and The Dowager Countess Eleanor Peel Trust.