Using baseline gene expression for multi-Compound screening in early drug development experiments.

Adetayo Kasim[1], Dan Lin[1], Ziv Shkedy[1], Willem Talloen[2] and Luc Bijnens [2]

[1] Center for Statistics, Hasselt University, Belgium
[2] Johnson & Johnson PRD, Beerse, Belgium

The discovery of new active compounds is a big challenge for pharmaceutical research when differential response can be expected, like for oncology compounds. Here, we propose a method to analyse screening studies on oncology compound libraries on several tumor cell lines. The task is threefold. First, subsets of compounds that reveal some very distinct tumor growth inhibition profiles need to be identified . Second, cell lines need to be classified as either responding or not responding to the specific compounds. And third, baseline gene expression is used to differentiate responding from non-responding tumors in order to discover genetic signatures of response. This task is equivalent to simultaneous clustering of the compounds and cell lines in order to find groups of compounds that are active on responder cell lines and inactive on non-responder cell lines. We propose a method based on finite components mixture models to find clusters of compounds that are active on responder cell lines and inactive on non-responder cell lines. In addition, we suggest using supervised principal components and activity region finder to discover genetic signatures of response from baseline gene expression of the cell lines.

References

[1] Amaratunga D. and Cabrera J. (2004). Mining data to find subsets of high activity. Journal of Statistical Planning and Inference 122, No. 1-2, 23-41

[2] Bair E; Hastie T; Paul D; Tibshirani R. (2006). Prediction by Supervised Principal Components. Journal of American Statistical Society 101, No. 473, 119-137

[3] Gagarin A., Makarenkov V. and Zentilli P. (2006). Using clustering techniques to improve hit selection in high-throughput screening. Journal of Biomolecular Screening. 11, No. 8, 903 - 914

[4] Ma S. (2006). Empirical study of supervised gene screening. BMC Bioinformatics 7, 537