Regression on GO Terms

Luis Pedro Coelho

Instituto de Medicina Molecular

**Summary Sentence**: The assumption is that two drugs will act in a synergistic fashion if they disturb different genes with similar functionality.

**Background/Introduction**

Because only the single drug data was made available there is no training data.

**Methods**

For the implementation of this idea, I relied, on the molecular function GO terms.

First, the media and DMSO data was used to determine a baseline variation and genes were considered disturbed if they moved more than 1.5 standard deviations away from the mean at any time point. This is a noisy measurement, but it will only be used very indirectly. As above, genes were also grouped together by GO terms.

Each drug was thus characterized by two signatures:

1. A set of disturbed genes
2. A set of disturbed GO terms

For each pair of drugs, I computed the correlation between its disturbed genes and its disturbed GO terms.

There is a roughly linear correlation between these two values, but the distance to the regression line is our measure of synergy. Drugs pairs for which the GO terms are more correlated than predicted by the correlation at the gene level were predicted to be synergistic.

**Conclusion/Discussion**

With the benefit of the testing data, which was not available at the time of the competition, we can test a few variations and measure whether they would have been better than the submission. In particular, the restriction to the Molecular Function vocabulary was somewhat arbitrary and we can test other GO vocabularies. In total, there are 6 possible combinations of vocabularies. The best result is the combination of the molecular function and the biological process vocabularies, but the p-value is 16%. The best single vocabulary is not molecular function, but biological process (with a p-value of 21%). In retrospect, this may have been a better embodiment of the idea that was underlying this method, but the result is still of limited value.

**References**

**Authors Statement**

LPC developed the methodology, implemented it, and wrote the report.