A combined Bayesian and Cosine Similarity Matrix Based Analysis of Time-Series Drug Response

Jean-Paul Abbuehl[$], Arvind Sridhar[*], Jonathan Bernard[$], Krisztian Homicsko[$] and Anguraj Sadanandam[$,#]

[$] Swiss Institute for Experimental Cancer Research (ISREC), [*]Embedded Systems Laboratory (ESL), Institute of Electrical Engineering, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland

[#]Correspondence to Anguraj.sadanandam@epfl.ch

Summary Sentence:

BETR-NTP based analysis of time series drug response data.

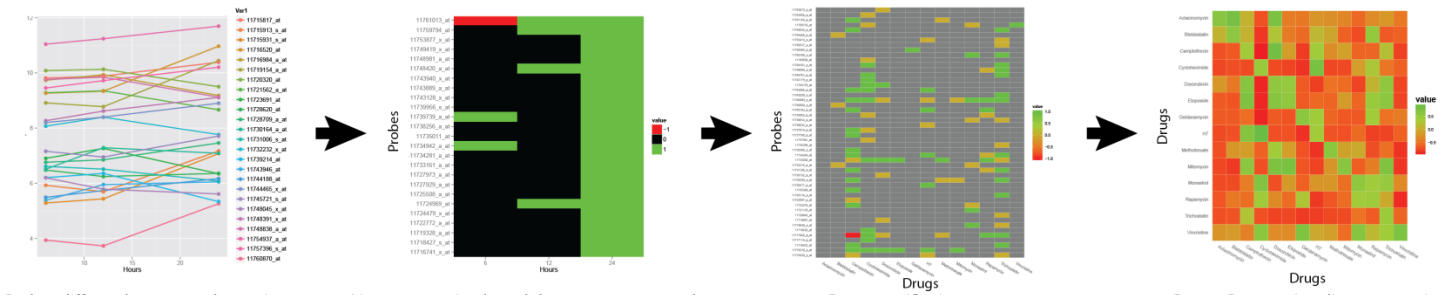
Background/Introduction

In order to determine combination therapies that are effective, we developed an algorithm based on the following assumptions:

1. Cancer cells are intra-heterogeneous and they have distinct population of cells with different transcriptome profiles. Among the distinct population of cells, there are drug-resistant and -sensitive populations that can be identified after exposure to drug.
2. There are a population of cancer cells (probably cancer stem cells; CSC) that are intrinsically resistant to drug treatments whereas other population of cells (dubbed as non-CSC) are sensitive to the same treatment.
3. The change in transcriptome profiles is predominantly the result of a dominant population of CSC cells that are resistant to treatment.

NCI-DREAM team sub-challenge 2 time-series drug response and molecular data can be used to identify transcriptome profiles of CSC and non-CSC subpopulations of cells. We reasoned that combinations of two drugs that affect one way or the other both the populations of CSC and non-CSC cells are an effective pair. Our method did not use SNP profiles or IC20 values.

Methods



Probes differently expressed over time | Linear regression based drug-response vs probes | Drug-specific signatures | Drug x Drug cosine distance matrix

1. Identification of differentially expressed genes

The time-series transcriptome data provide the advantage to improve the sensitivity of detecting the changes over time, while noisy signals sustains across time point and can be removed. Common static gene expression analysis methods are not useful for these data. On the other hand, bayesian estimation of temporal regulation (BETR) correlates transcriptiome data and identifies genes that are

differentially expressed between two consecutive time points[1]. R based BETR package[1] was used to select probes with differential expression for each drug against DMSO with a confidence of >95%.

2. Filtering of probes for each time point

In order to select the most relevant probes for each drug, we assume that a linear correlation between drug dosage (0, 1/10 of IC20 and IC20) and a subset of probes exists. Linear regression for each probe was performed at a given time and the quality of the fitted model was determined by one-way ANOVA.

3. Identification of drug-response specific signatures

For a given drug and a time point, probes were clustered into two signatures depending on the sign of the slope. In total, 84 signatures were obtained (14 drugs x 3 time points x 2 up/down probes). The down-regulated genes were considered as representative of the drug effect (drug-effect signature, ***DES***) whereas up-regulated genes were considered as representative of the compensatory mechanisms the tumor cells used to survive (drug-surviving signature, ***DSS***).

4. Derivation of the drug-drug interaction matrix

Nearest Template Prediction (NTP) algorithm[2] with cosine distance metric was used to associate all the samples that were treated with a particular drug (eg., drug A) to either DES or DSS of another drug (eg., drug B). Only those associations with Benjamini and Hochberg (BH) false discovery rate greater than 0.05 were considered. A single score of cosine distance (*CD*) was calculated as below:

where, CD_{DSS} and CD_{DES} represent distance (cosine distance) for each sample that were treated with drug A to that of the DSS and DES, respectively, from drug B. $n=i+j$, where i represents number of samples associated DSS and j represents the number of samples associated with DES.

Furthermore, the weighted sum of the *CD* scoring across 3 time points was performed as below:

$$CD = CD_6x(4/7) + CD_{12}x(2/7) + CD_{24}x(1/7)$$

where, CD_6 represents *CD* scoring for 6 h time point, CD_{12} for 12 h and CD_{24} for 24 h. This entire scoring system was repeated for all possible drug combinations leading to a final 14x14 interaction matrix. A final score was calculated as below:

$$S = CD_{AB} + CD_{BA}$$

Where CD_{AB} represents *CD* for drug A and the association of its samples with DSS and DES of drug B where CD_{BA} represents *CD* for drug B and the association of its samples with DSS and DES of drug A. The drug combination was considered synergistic if S is greater than $CD_{AA} + CD_{BB}$ and it is additive otherwise.

Discussion

	BETR	Linear drug-gene at 6h		Linear drug-gene at 12h		Linear drug-gene at 24h	
	# Probe	DES	DSS	DES	DSS	DES	DSS
Aclacinomycin	633	0	52	0	346	0	6
Blebbistatin	6519	0	413	5	1910	0	294
Camptothecin	14701	51	6240	1	5231	71	4177
Cycloheximide	18177	16	4671	2	4329	421	5602
Doxorubicin	10211	3	1439	27	2408	2	2779
Etoposide	2761	0	701	2	621	3	636
Geldanamycin	2460	1	958	0	597	7	520
H7	13368	9	3772	5	3017	50	759
Methotrexate	2912	17	166	2	270	12	1511
Mitomycin	2656	0	218	0	268	27	554
Monastrol	9334	1	1190	6	2160	178	2035
Rapamycin	8850	9	1515	2	1975	8	1217
Trichostatin	17440	7	5557	2	4982	1012	2404
Vincristine	573	0	59	0	189	0	105

Table 1. A summary of probes (DSS and DES) that were selected for each drug at different time points.

We observed that there was significantly less number of genes in DES compared to DSS in all the drugs, and this could have affected algorithm. In this case, logistic model may be suitable for the dose-response relationship. In addition, the association study using NTP algorithm could be improved by comparing each transcriptome data of each drug to that of the other drugs.

Conclusion

Our ability to understand tumor response offers the possibility to anticipate the escape mechanism that tumors use.

Authors Statement. Jean-Paul Abbuehl conceived the idea of using this algorithm, wrote the R scripts, performed the analyses, interpreted the data and co-wrote the description. Arvind Sridhar, Jonathan Bernard and Krisztian Homicsko participated in critical discussions. Anguraj Sadanandam conceived the idea of using NTP algorithm, helped with analysis, co-wrote the description and supervised the project.

References

1. Aryee, M.J., Gutierrez-Pabello, J.A., Kramnik, I., Maiti, T. & Quackenbush, J. An improved empirical bayes approach to estimating differential gene expression in microarray time-course data: BETR (Bayesian Estimation of Temporal Regulation). *BMC bioinformatics* **10**, 409 (2009).
2. Hoshida, Y. Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. *PloS one* **5**, e15543 (2010).