# Title:

# Abstract (150 words)

Cancer relapse is a major complication after treatment of patients. The cellular heterogeneity and the intrinsic drug-resistance mechanisms are potential reasons for the failure of a single drug to eradicate the whole tumor. The combination of synergistic monotherapies is a powerful strategy to increase the efficacy, decreasing the toxicity and avoiding resistance and recurrence. Here, we present a novel algorithm which aims at delineating cancer subpopulations, relating their respective sensitivities to each drug and ultimately predicting optimal combinations that are complementary in terms of subpopulation targeting. Our ability to understand tumor response offers the possibility to anticipate the escape mechanism that tumors use.

# Introduction

There are currently about 300 clinical drugs approved for cancer treatment. The treatment decision depends on various relevant factors, such as the localization, histological features, biological subtypes and the toxicities. Therapies based on a single drug are generally given to patients. However, despite showing some improvements, cancer can reappear years later with higher aggressiveness and increased resistance to treatments. To account for these clinical observations, three models were described to explain tumor relapse after treatment. First, the acquisition of molecular resistance decreases the sensitivity of tumor cells to the treatment. Alternatively, the abnormally high mutation rate leads to high cellular heterogeneity. A single subpopulation with prior intrinsic resistance is positively selected and repopulates the tumor. Finally, tumor harbors a hierarchical organization with a cancer-stem cell population that are resistant to therapies and fuels the tumor bulk. Although different, these models are not mutually exclusive and can be valid depending on the mode of action of the drug. A common feature between these models is the fitness landscape that describes the adaptation of cells to its environment, based on hills and valleys.

A large number of studies have investigated molecular mechanisms responsible to tumor resistance for a single drug. Several cell lines, representative of the panel of human patients, are characterized by high-throughput technologies and exposed to clinically relevant drugs. The strategy relies on linking a particular cell line with its sensitivity profile to a human patient. Another innovative approach is to cluster independently human patients and cell lines into biologically relevant subtypes and then assign patients to its relevant subtype where treatments are optimally designed. Instead of, the efficacy of two drugs combination is evaluated by integrating biological knowledge with the cellular response of each drug separately. The identification of critical soft spots by modeling molecular network could lead to a synergistic effect of the combination treatment.

Overall, available data for *in silico* prediction are snapshot of the tumor biology that offers insight about inter-tumor heterogeneity. The intra-tumor heterogeneity is rarely addressed in patient. In this aspect, time-course analysis is valuable at describing the dynamic and the components of the tumor heterogeneity.

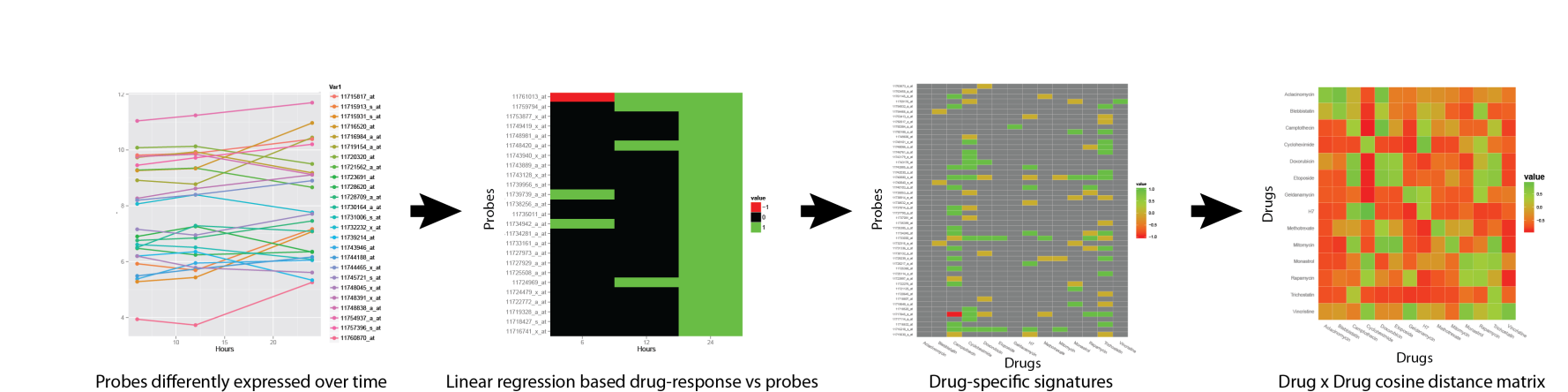
NCI-DREAM challenge provides time-series drug response data that can be used to assess the dynamic of the system. In this study, a cell line has been exposed to 14 different drugs at different concentrations. The transcriptome of cells have been measured at different time points for each drug. We hypothesize that the change in microarray profiles is predominantly the result of a shift in subpopulation(s), which are positively selected during drug exposure. By subjecting the whole to a variety of landscapes (eg. Drug), we delineate subpopulations that are described by signatures. Finally a drug-specific landscape map with hills and valleys are draw for all subpopulations. The rational is to fill valleys (resistant) by combining maps with additive topologies.

# Results

## Identification of subpopulations

We first identify time-relevant probesets by Bayesian estimation of temporal regulation (BETR) for each drug over time with a probability >95% [REF]. We observe that the distribution of the number of probesets is heterogeneous among drugs (Fig1A). A linear correlation (Anova, pval<0.05) between drug dosage and probeset intensity was used as a criteria for features selection. For a given drug and time point, candidate probesets were clustered into two signatures depending on the drug concentration – probeset intensities relationship (Fig1B). In total, 84 signatures were obtained (14 drugs x 3 time points x 2 up/down probesets). 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***). Each DES and DSS is used to describe subpopulations.

## Derivation of the drug-drug interaction matrix



For all timepoints, samples that were treated with a particular drug (eg., drug A) were matched to either DES or DSS of another drug (eg., drug B) by Nearest Template Prediction (NTP) [REF]. The cosine distance is used as quantitative information about drug-drug interaction. After matching all DES and DSS with all drug-treated samples, we obtained the drug-drug interaction matrix (Fig2A). A positive score means the DES of drug A is found overrepresented in the treated samples of drug B. . Hierarchical clustering of the matrix regroup class of drugs with similar correlation trends. The interaction matrix shows X% of correlation are positive, X% negative and X% could not be significantly attributed to any DSS nor DES. If we cluster possible combinations by sign correlation, we observe that positive symetry is a feature of top ranked combination while negative symetry is a feature of worst combination, independantly of each drug combination (Fig2A sup).

Because the drug-drug interaction matrix is derived from DES,DSS and treated-samples at all timepoint, we sought to determine which timepoint for all drug was the most informative for DES and DSS (Fig2B and C). The similarity between any treated samples and any DES is most important if there are temporarly distinct. On the other hand, DSS are closer to treated samples at same timepoint. The latest timepoint (24hr) DSS cannot be matched significantly at 24hr drug-treated samples.

Prediction of drug-drug combination and synergism.

We used the experimental results provided by the Dream 7 Challenge to assess the prediction of the algorithm. The scoring metric used is based on the Bliss independence, which estimates the synergy against the additictive effects of two drugs. All predicted pairwise combinations were ranked according to the similarity matrix by additioning element ij and ji. The predictive score obtained is 0.56 (0.01 pvalue). The synergy is determined when Xij + Xji > Xii+Yjj.

The drug-drug interaction matrix is not Write down what you observe in figure 2. Tell sth about the symetry of the matrix

Fig2A sup symmetry

Fig2B and C : Timepoint

Fig2D sup symmetry

Show DES (Fig2A), Fig(2B), (delayed drug effect) separatly and combined Fig(2C)

What is the most informative timepoint for DSS and DES generation Fig(2D) At fixed timepoint of DSS and DES

1. Relevant timepoint for DSS definition and DSS targeting 3x3 matrix 🡪 Clinical point of view
2. Relevant timepoint for DES definition and DES targeting 3x3 matrix 🡪 Clinical point of view

DSS total<- DSS 6hr + DSS 12hr + DSS 24hr

Make table with score in fct of (to test robustness)

nb timepoint (more informative?)

matching time point

IC20 vs IC20low

# Drugs

# Prediction Network graph fig3.

x

# Valley filling in biological landscapes

# Discussion

# Methods

# Study highlights

## What is the current knowledge on the topic?

## What question this study addressed?

## What this study adds to our knowledge?

## How this might change clinical pharmacology and therapeutics?

**DREAM Project’s challenge number 2**

The second challenge of The DREAM Project seeks to determine which drugs combination are potent at preventing relapse over time. This report exposes our approach to this question. It begins by stating our hypothesis before presenting our methodology. An example is then given in order to illustrate the algorithm we developped.

In order to answer the second challenge for the DREAM Project, we developped an algorithm based on the following assumptions:

1. *Cancer cells always harbor heterogeneity in their transcription profiles. Upon selective pressure (eg Drug exposure), a drug-resistant and a drug-sensitive populations are identified.*
2. *Prior drug exposure in vitro, rare cancer cells are already intrinsically resistant.*
3. *The dynamic of transcription profiles are predominantly the result of changes in populations distribution, rather than molecular response of each cancer cell.*

Our approach consists of combining drugs that are the most complementary in terms of population targeting.

Firstly, differentialy expressed genes for each compound over DMSO are identified by using Bayesian Estimation of Temporal Regulation bioconductor package (Aryee, 2011). Secondly, linear regression is used to select the ones with linear correlation between expression level and drug exposure over time. Thirdly, probes for each compounds are clustered into two groups, based on the sign of the slope. The first group includes genes downregulated over time and is used to predict the efficiency of the drug, according to a given population. The second group includes genes overregulated over time and is used to predict which population is unlikely to be affected. With these 2 x 14 signatures, we use the nearest template prediction algorithm (Hoshida, 2010) to obtain cosine distances of all treated-samples with all compound-specific signatures.

Scores are attributed as follows:

* Pairs receives a high positive score when a treated sample is positively correlated to a drug-resistant signature of another drug.
* Pairs receives a high negative score when a treated sample is negatively correlated to a drug-sensitive signature of another drug.

The score is calculated for each timepoint and sum up together, with weighting coefficients. The synergestic effect is predicted when the combination of two different drugs gives a higher score than each drug alone.

Scenario:

Drug 1 is efficient only against population A (99% frequency) and drug 2 is efficient only against population B (1% frequency). In term of drug potencies, drug 1 is higher as most cells are affected by it.

* After drug 1 treatment, upregulated genes reflect the raise in population B (1% 🡪 100%), while downregulated genes reflects the loss of population A (99% 🡪 0%).
* After drug 2 treatment, upregulated genes reflect the raise in population A (99% 🡪 100%), while downregulated genes reflects the loss of population B (1% 🡪 0%).

When combining drug 1 and drug 2, all cells are eliminated because drug 1 targets population A while drug 2 targets population B. Alone, none of these drugs would have prevented relapse, but drug 2 has the ability to reduce heterogeneity, prior treatment with the highly effective drug 2.

**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.

# 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

**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).

In order to determine which drugs combination are potent at preventing relapse over time, we developed an algorithm based on the following assumptions:

1. Cancer cells always harbor heterogeneity in their transcription profiles. Upon selective pressure (eg Drug exposure), a drug-resistant and a drug-sensitive populations are identified.
2. Prior drug exposure in vitro, rare cancer cells are already intrinsically resistant.
3. The dynamic of transcription profiles are predominantly the result of changes in populations distribution, rather than molecular response of each cancer cell.

Stem cell specific and conventional cancer therapies

Source: Wikipedia

# Methods

## Identification of differentially expressed genes over times following drug exposure

Static methods, such as Limma and SAM, are used to identify differentially expressed genes in samples with multiple biological conditions. When time-series data are available, the time-dependent structure offer the advantage to improve the sensitivity of detection. Noisy signal is sustained across time point and can be eliminated. This improves detection of small signals that cannot be identified when using only static methods.

Bayesian Estimation of Temporal Regulation (BETR) is a new algorithm that takes into account the correlation between successive time points and estimate the probability that a gene is differentially expressed between two conditions over time, without the need to have balanced sample sizes. We used BETR R package to select probes with differential expression for each drug against DMSO. The cut-off used was >95%.

## Filtering of probes with a dose-response relationship 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 subset of probes exists. We perform a linear regression for each probe at a given time and determine quality of the fitted model by one-way ANOVA.

## Clustering of drug-specific probes into drug-response signatures

For a given drug and 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). We considered signatures with down-regulated genes as representative of the drug effect (drug-effect signature, ***DES***) and signatures with up-regulated genes as representative of the compensatory mechanism the tumor cell uses to survive (drug-surviving signature, ***DSS***).

## Derivation of the drug-drug interaction matrix

We cross-compared all drug-specific signatures against all samples by Nearest Template Prediction and used cosine distances (FDR<0.05) to build an interaction matrix

* Cosine distances values are summed up when a treated sample is associated to a given ***DSS***
* Cosine distances values are subtracted when a treated sample is associated to a given ***DES***

Scores are calculated for each timepoint and averaged together to produce a final interaction matrix, describing which combinations are the most potent.

The synergistic effect is predicted when the combination of two different drugs gives a higher score than the drug-specific signature with the same drug-treated sample.

# 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 |

We observe that BETR algorithm was more efficient to detect more subtle differently expressed genes, compared to commonly used methods. Since drug responses are delayed in time, we suggest that using 6 hour as the baseline could have improved the identification of more relevant genes.

The filtering of genes with linear relationship between probe level and drug concentration has severely affected the predictive power of our algorithm. Indeed, few genes were found in DESs (a.i. positive correlation to drug concentration. We suggest that a logistic model would be more suitable for the dose-response relationship. Additionally, the range of concentration used in NCBI-DREAM data could have been too narrowed to capture those important probes (a.i. IC20 and IC50 instead of IC20 and 1/10th of IC20). The unbalanced signatures sizes must have affected the predictive power of NTPez algorithm.

Then, DSSs could have been reduced by keeping probes with sustained pattern over time, since the cells were keep in the drug over the full timecourse. This could have increased the specificity by eliminating false positive probes and probes with circadian regulation.

Finally, our usage of NTPez algorithm could be improved by pre-processing the drug-treated sample by differential expression against all others-drug treated samples. We suggest this approach improves the ranking prediction of drug pairs without synergistic nor antagonistic effects. We could have also used NTPez without time-matching restriction between DES/DSS and treated samples. Indeed, the drug responses could be delayed in time and pre-conditioning tumor with one drug could potentiate the toxicity of a second drug later on.

# Conclusion

Our ability to understand tumor response offers the possibility to anticipate the escape mechanism that tumors use. We are grateful to NCBI-DREAM organizers for the stimulating challenges they provided and the opportunity to gain insights about cancer resistance.

# References

Betr, NTPez

# Authors Statement

4000 words excluding abstract,references, tables and figures

Abstract 150 words

References : 50 maximum

Figures/table 7 maximum

Substantial novel research

Title page

Abstract

Introduction

Results

Discussion

Methods

Each figure must be provivded as an individual filte.

Single image : 500 pixel width, resolution 125 dpi

Multipartimages : 900 pixels width, select constrain proportions, resolution 125 dpi

Acknowledgements