

# Drug Synergy Prediction Challenge

Chieh Lin, Po-Wei Wang, Yan Xia

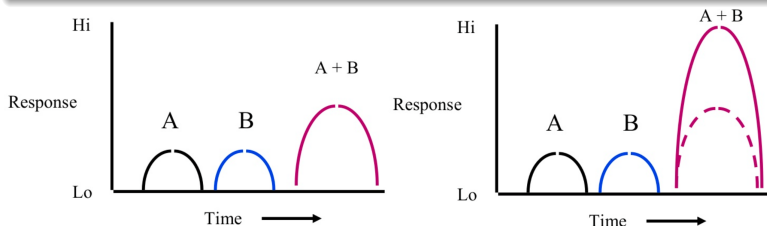
Project Supervisor: Ziv Bar-Joseph

December 13, 2015

# Synergy: What and Why

## What is synergy?

$$\text{Effect}(A,B) = \text{Additive\_Effect}(A,B) + \text{Synergy}(A,B)$$



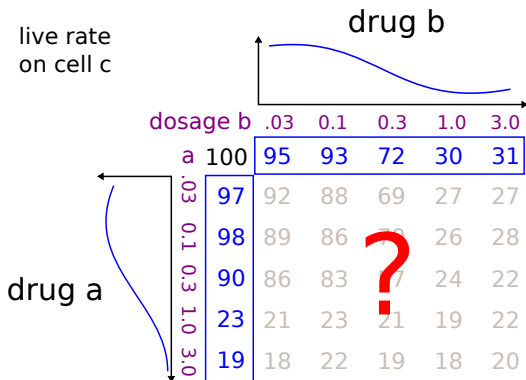
## Why is synergy important?

1. Help prevent deadly drug combination side effects
2. Help develop new drug combination therapies

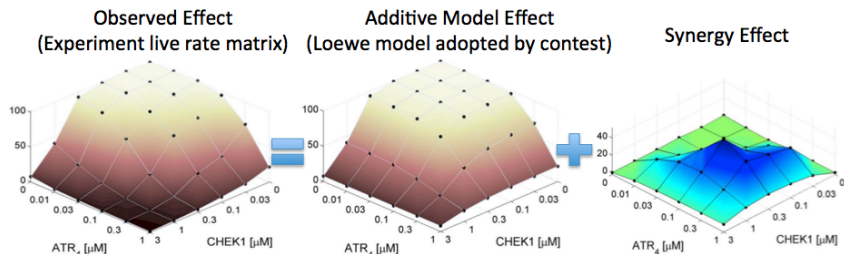
# Dataset: From Synergy Prediction DREAM Challenge

## Given

- Mono therapy: Live rates of **drug a** on **cell c**.
- Combined therapy: Live rates of **drug a**, **drug b** on **cell c**.



# Problem Definition



- Perform **integration** on the **synergy effect** => synergy score (target value of contest)

# Modeling Strategies

## Additive Damage Model (from literature) (baseline)

Fit the mono-therapy curve to predict combined-therapy data

## Naive Regression (baseline)

Use (kernelized) logistic regression models to fit synergy score directly

## Locality Model

Predict the combined-therapy data based on the locality of mono-therapy and combined-therapy.

# Model: Additive Damage Model (baseline)

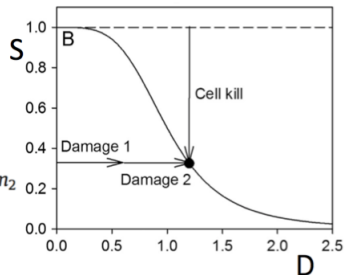
From literature

Proposed by Jones, Leslie Braziel, et al. (2014)

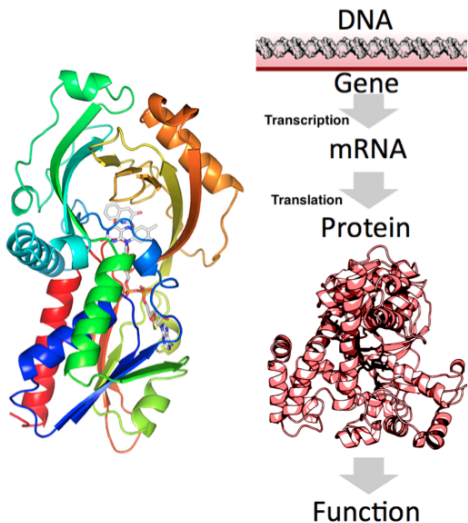
Idea: Curve fitting

Assumption: The curve is **decreasing** and has a **sigmoidal shape**

$$S = \frac{1}{1 + D^n}$$
$$D = \left( \frac{a_1 C_1}{1 + s_1 C_1} \right)^{m_1} + \left( \frac{a_2 C_2}{1 + s_2 C_2} \right)^{m_2}$$



# Background: Gene Expression and Drug



Gene Expression Data				
	Cell1	Cell2	Cell3	Cell4
Gene1	4.80449	8.71356	8.49596	7.81344
Gene2	2.84813	2.83472	2.887	2.87933
Gene3	10.332	10.5047	9.64344	10.2727
Gene4	5.14539	3.79219	4.15447	4.05039
Gene5	3.54519	3.94666	3.30731	3.70149

# Model: Naive Regression (baseline)

## Naive Feature vector

A and B are drug indices, C is cell index

Feature:  $[Drug_A; Drug_B; GeneExpression_C]$

## Idea: drug vector should be sparse

Use lasso regression to learn non-zero elements of  $Drug_A, Drug_B$

Solve:  $Drug_A \cdot GeneExpression_C = \min\_live\_rate(\text{drug A})$

## Kernel

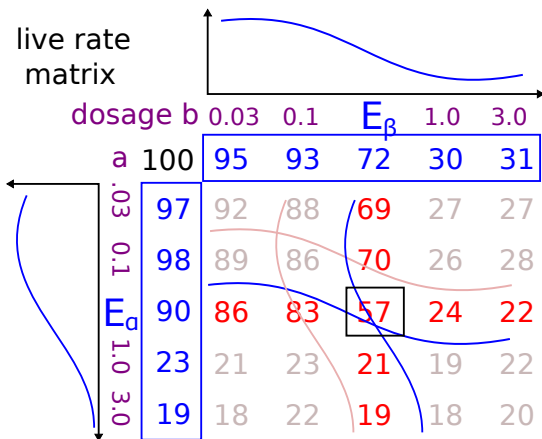
1. RBF kernel
2. Polynomial kernel with degree 2



# Model: Locality

## Observation

The shape of combined therapy is similar to nearby mono-therapy.



$$E_{ref} = \min(E_\alpha, E_\beta) = 72$$

$$damage = (100 - E_\alpha) = 10$$

$$\hat{E}_{\alpha,\beta} = E_{ref} - z * damage$$

$$= 72 - 10(\text{set } z = 1)$$

damage  
\* scale

# The Algorithm of Locality Model

## The Model

Predict live rate  $\hat{E}_{\alpha,\beta}$  by

$$\hat{E}_{\alpha,\beta} := E_{\alpha} - \left( \sum a_k b_k c_k + \mathbf{b} \cdot \mathbf{c} \right) \cdot (E_{\beta} - E_{\beta 0}),$$

where  $\alpha, \beta$  are the dosage<sup>k</sup> of drug  $a$  and  $b$ , and  $E_{\alpha} \leq E_{\beta}$

- Drug vector should be sparse.
- Want the integrated difference instead of every  $E_{\alpha,\beta}$

Solve

$$\min_{drug} \|drug\|_1 + \lambda^{mo} \sum_{i \in mo} \sum_{\alpha} (\hat{E}_{\alpha}^i - E_{\alpha}^i)^2 + \lambda^{co} \sum_{i \in co} \left( \sum_{\alpha, \beta} w_{\alpha, \beta} (\hat{E}_{\alpha, \beta}^i - E_{\alpha, \beta}^i) \right)^2,$$

where  $w_{\alpha, \beta}$  is the numeric integral constant.

## The Algorithm of Locality Model (Cont.)

$$\min_{drug} \|drug\|_1 + \lambda^{mo} \sum_{i \in mo} \sum_{\alpha} (\hat{E}_{\alpha}^i - E_{\alpha}^i)^2 + \lambda^{co} \sum_{i \in co} \left( \sum_{\alpha, \beta} w_{\alpha, \beta} (\hat{E}_{\alpha, \beta}^i - E_{\alpha, \beta}^i) \right)^2,$$

$$\text{where } \hat{E}_{\alpha, \beta}^i := E_{\alpha}^i - \left( \sum_k a_k^i b_k^i c_k^i + \mathbf{b}^i \cdot \mathbf{c}^i \right) \cdot (E_{\beta}^i - E_{\beta 0}^i),$$

$$\hat{E}_{\alpha}^i := E_{\alpha 0}^i - (\mathbf{a}^i \cdot \mathbf{c}^i)(E_{\alpha}^i - E_{\alpha 0}^i).$$

- Non-convex, but convex quadratic in every single coordinate.
- Developed a proximal coordinate descent algorithm to solve it.
- Maintain  $\hat{E}_{\alpha, \beta}$  to save computation.

Naive:  $O(\#data(\#feat)^2) \rightarrow$  DP:  $O(\#data\#feat)$  per iteration, where  $\#feat$  can be all genes or poly2 expansion on cell features.

# Results

Important note provided by contest holder

Redo the experiments will have  $\sim 0.5$  pearson correlation

<b>Regression: 5-fold CV</b>	<b>Mean Correlation</b>
Poly2 kernel	0.22
RBF kernel	0.23

<b>Use only mono therapy</b>	<b>Mean Correlation</b>
Additive Damage Model	0.16
Naïve Locality	0.21

<b>Locality: 5-fold CV</b>	<b>Mean Correlation</b>	<b>Training</b>	<b>Average Non-zero terms</b>
Gene Groups: 1000 iterations	0.28	0.48	8.32
Poly2 Expansion: 10 iterations	0.28	0.45	1.03

# Conclusions

- From fitting parametric curve to locality model
  - New way to construct live rate by mono-therapy.
  - No need to worry about the choice of curve.
- L1 + Poly2 expansion can leads to good result with controllable non-zero terms
  - our best model w/ avg **1.03** non-zero feat/drug.
- New proximal coordinate algorithm with **linear complexity** to *#feat*.
- A systematic study to model drug synergy effect with little domain knowledge (pathway, protein protein interaction, etc.)

## References

- AstraZeneca-Sanger drug combination prediction DREAM challenge:  
<https://www.synapse.org/#!Synapse:syn4231880/wiki/235645>
- Jones, Leslie Braziel, et al. "The additive damage model: A mathematical model for cellular responses to drug combinations." *Journal of theoretical biology* 357 (2014): 10-20.
- <http://www.slideshare.net/tejamba/pharmacology-45460271>
- <https://en.wikipedia.org/wiki/Gene>
- <http://chalice.horizondiscovery.com/chalice-portal/documentation/analyzer/models/model`add.jsp>