Aalto University

Predicting drug combination responses in cancer

Juho Rousu*
Department of Computer Science
Aalto University

AIDD Summer School, 20.3.2023

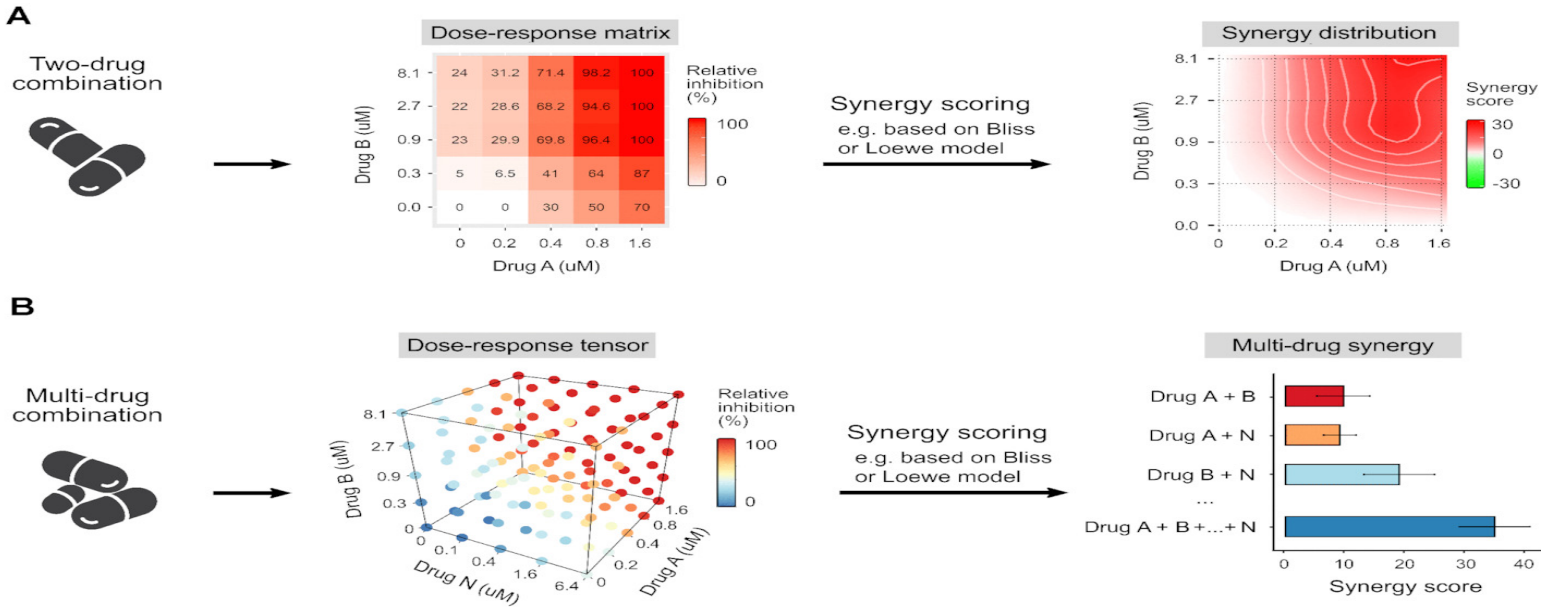
*Joint work with: Heli Julkunen, Tianduanyi Wang, Sandor Szedmak, Anna Cichonska, Prson Gautam, Tapio Pahikkala, Tero Aittokallio

Combinatorial therapies in complex diseases

- Combinatorial treatments involving two or more drugs have become a standard of care for various complex diseases (tuberculosis, malaria, HIV, cancer)
- Their benefits include enhanced treatment efficacy, avoidance of drug resistance and fewer side-effects
- Promising drug combinations are typically search for by High-Throughput-Screening (HTS) in preclinical model systems (cancer cell lines, viral infection models)

Dose-response and synergy of drug combinations

- The quantities of interest are the **dose-response behavior** and the **synergy** of the drug combination – how much more effective the combination is compared to drugs acting alone



https://synergyfinder.fimm.fi/synergy/synfin_docs/

Ianevski, A., Giri, A.K. and Aittokallio, T., 2020. SynergyFinder 2.0: visual analytics of multi-drug combination synergies. *Nucleic Acids Research*.

Tackling the combinatorial explosion with machine learning

- The main challenge in finding promising drug combinations is the exponential size (in # combined drugs) of the search space
- To efficiently explore this space, we need **prioritization** of which drug combinations will be tested
- Machine learning is seen as a key tool for focusing the HTS efforts to most promising drug combinations
- NCI-Almanac dataset (Holbeck et al. 2017), with over 5000 drug pairs tested, is the first **large scale public combination response dataset** that enables accurate ML models to be developed

Holbeck et al., 2017. The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer research*, *77*(13), pp.3564-3576.

Prediction scenarios

- Multiple prediction scenarios of different difficulty can be defined on drug combination response prediction
- Scenarios with out-of-tensor drugs and/or cell lines are generally the harder ones

Scenario	Title	Out of training tensor items in the test set			
		drugs	cell line	drug-drug	drug-drug-cell
NEW-ENTRY	Predicting new dose-response matrix entries	no	no	no	no
NEW-MATRIX	Predicting new dose-response matrices	no	no	no	yes
NEW-COMBO	Predicting new drug combinations	no	no	yes	yes
NEW COMBO-NO-MONO	Predicting new drug combinations with no monotherapy measurements	no	no	yes	yes
NEW-CELL-LINE	Predicting dose-response matrices on new cell lines	no	yes	no	yes
NEW-TISSUE-TYPE	Predicting dose-response matrices on new tissue types	no	yes	no	yes
NEW-DRUG	Predicting dose-response matrices on combos with one new drug	yes	no	yes	yes
2-NEW-DRUGs	Predicting dose-response matrices on combos with two new drugs	yes	no	yes	yes

Predicting multi-way interactions

Multi-way interaction problems

Consider prediction problems where the predicted target variable dependes on simultaneous interaction of m objects

- **Pairwise cases (m=2):**
  - Movie recommendation: f(user,movie) – predict if user will like a movies
  - Binding affinity between molecules (f(drug,target), f(protein,DNA), etc.)
- **Higher order interactions (m > 2):**
  - Genotype-phenotype interactions: m SNPs needed to explain a phenotype
  - Drug combination responses (this talk): 2 drugs x 2 doses x target = 5th order interaction

A first attempt of a predictive model

- Suppose we have large amounts of data on drugs and their responses on cancer cells
- Consider an additive model predicting the response of a cancer cell (c) to a drug (d):

$$F(d,c) = F_{Drug}(d) + F_{cell}(c)$$

- F_{drug} scores a drugs potency to kill cancer cells
- F_{cell} scores the cancer cells suceptibility to be killed by a drug
- Q: Would this model give a good basis for finding new drugs?

The need for non-linearity

- **An additive model can only recognize drugs that are effective regardless of the properties of the cancer cell**

  $F(d,c) = F_{Drug}(d) + F_{cell}(c)$

  - 'snake-oil' type cure for everything
  - cannot find targeted therapies for particular type of cancer cell
- **This is true even if F_{drug} and F_{cell} are highly non-linear**
- **Non-linear dependencies between the interacting objects need to be modelled**
- **Example: a polynomial of degree k can model interactions of k objects**
  - i.e. minimum quadratic model is needed for pairwise prediction

Factorization machines & latent tensor reconstruction

Starting point: polynomial regression

- Consider a quadratic regression model $f: R^d \rightarrow R$

$$f(x) = \sum_j w_j x_j + \sum_j \sum_k w_{jk} x_j x_k$$

- $O(d^2)$ parameters to estimate
- Generalize this to m'th degree polynomials

$$f(x) = \sum_j w_j x_j + \sum_j \sum_k w_{jk} x_j x_k + \cdots + \sum_{j_1,\ldots,j_m} w_{j_1,\ldots,j_m} \cdot x_{j_1} \cdot x_{j_2} \cdot x_{j_k}$$

- $O(d^m)$ parameters to estimate!
- Two-fold challenge:
  - Time complexity of estimating the model is exponential in **m**
  - Statistical challenge: in practical scenarios typically not enough data to estimate all parameters when **m** > 2

Factorization machines

- Factorization machines (Rendle, 2010) are an approach to make estimation of polynomial regression models from large dataset feasible
- The polynomial regression model is replaced by a factorized form

$$\hat{y}FM(x) = \sum_{j} w_j x_j + \sum_{j'>j} \langle p_j, p_{j'} \rangle x_j x_{j'}$$

- $p_j \in R^k$ is a vector representing contributions of variable x_j to k latent factors, where typically $k \ll d$ (low-rank)
- The interaction weight $w_{ij} \approx \langle p_j, p_{j'} \rangle = \sum_{s=1}^{k} p_{js}\, p_{j's}$ is represented as a inner product over the factor contributions
- O(dk) parameters to estimate, compare to original O(d^2)

Rendle, S., 2010, December. Factorization machines. In *2010 IEEE International Conference on Data Mining* (pp. 995-1000). IEEE.

Higher-order factorization machines

- Higher-order factorization machines (HOFM) (Blondel, 2016) can represent polynomial models of arbitrary maximum degree m

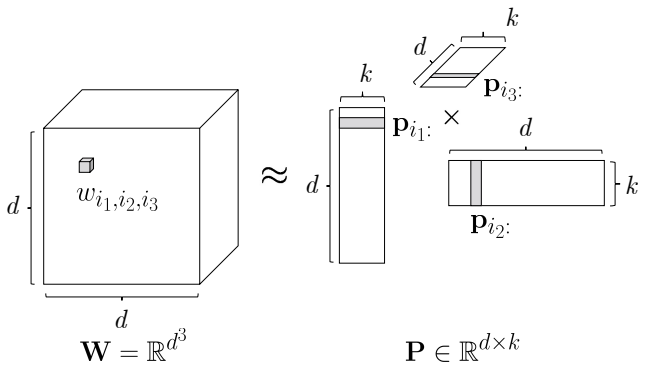
$$\hat{y}HOFM(x) = \sum_{j} w_j x_j + \sum_{j\prime>j} \langle p_j{}^{(2)}, p_{j\prime}{}^{(2)} \rangle x_j x_{j\prime} + \cdots + \sum_{j_m>\cdots>j_1} \langle p_{j_1}{}^{(m)}, \ldots, p_{j_m}{}^{(m)} \rangle x_{j_1} \cdots x_{j_m}$$

- The interaction weights are given by generalized inner products

$$\mathrm{w}_{\mathrm{j}_1 \ldots \mathrm{j}_\mathrm{m}} \approx \langle \mathrm{p}_{\mathrm{j}_1}{}^{(\mathrm{m})}, \ldots, \mathrm{p}_{\mathrm{j}_\mathrm{m}}{}^{(\mathrm{m})} \rangle = \sum_{\mathrm{s}=1}^{\mathrm{k}} \mathrm{p}_{\mathrm{j}_1 \mathrm{s}}{}^{(\mathrm{m})} \cdots \mathrm{p}_{\mathrm{j}_\mathrm{m} \mathrm{s}}{}^{(\mathrm{m})}$$

- O(dkm) parameters to estimate, compared to O(d^m) of the full model

Figure right: factorization of a 3rd order weight tensor



Blondel, M., Fujino, A., Ueda, N. and Ishihata, M., 2016. Higher-order factorization machines. In *Advances in Neural Information Processing Systems* (pp. 3351-3359).

Learning HOFMs

- The objective function of learning HOFMs is given by

$$\frac{1}{n}\sum_{i=1}^{n}\ell\left(y_i,\hat{y}HOFM(x_i)\right)+\frac{\beta_1}{2}\|w\|^2+\frac{\beta_2}{2}\|P\|^2$$

- It is a differentiable non-convex functional; can be trained by stochastic gradient descent (SGD)
- The challenging part is the exponential number of terms in the expression for $\hat{y}HOFM(x)$
- However by making use of the repetitive structure of the factor combinations, dynamic programming can be used to compute $\hat{y}HOFM(x)$ and its gradients in linear time
- The SGD algorithm runs in time O(dkmn) per epoch
- GPU acceleration can be used for further speed-up

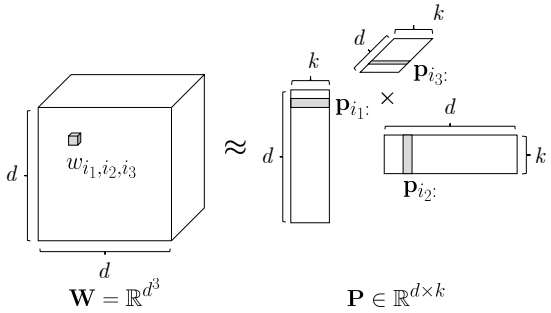
Blondel, M., Fujino, A., Ueda, N. and Ishihata, M., 2016. Higher-order factorization machines. In *Advances in Neural Information Processing Systems* (pp. 3351-3359).

Latent tensor reconstruction

- Latent tensor reconstruction (LTR) is another machine learning method for learning multi-way interaction models
- It is motivated by higher-order SVD problem
- The LTR model is given by

$$f(\mathbf{x}) = \sum_{t=1}^{n_t} \lambda_t \left\langle \mathbf{p}_1^{(t)} \otimes \cdots \otimes \mathbf{p}_{n_d}^{(t)}, \mathbf{x} \otimes \cdots \otimes \mathbf{x} \right\rangle .$$

$$= \sum_{t=1}^{n_t} \lambda_t \left\langle \mathbf{p}_1^{(t)}, \mathbf{x} \right\rangle \cdot \ldots \cdot \left\langle \mathbf{p}_{n_d}^{(t)}, \mathbf{x} \right\rangle ,$$



- <.> is the inner product of two tensors and $\otimes$ is the tensor product
- Enjoys a similar exponential reduction in the number of parameters as HOFM, but is a more flexible model

Wang, T., Szedmak, S., Wang, H., Aittokallio, T., Pahikkala, T., Cichonska, A. and Rousu, J., 2021. Modeling drug combination effects via latent tensor reconstruction. Bioinformatics, 37(Supplement_1), pp.i93-i101.

Learning LTR

The basic rank wise algorithm

Given: a sample: $\{(\mathbf{x}_i, y_i),\ i = 1, \ldots, m\}$,
Let $y_i^{(1)} = y_i,\ i = 1, \ldots, m$.
For $t = 1$ **to** n_t **do**
Solve rank-one subproblem

$$\begin{aligned} \min \quad & \sum_{i=1}^{m} ||y_i^{(t)} - \lambda_t f^{(t)}(\mathbf{x}_i)||^2 \\ & + \frac{C_p}{n_d n} \sum_{d=1}^{n_d} ||\mathbf{p}_d^{(t)}||^2 \\ \text{w.r.t.} \quad & \lambda_t,\ \mathbf{p}_1^{(t)}, \ldots, \mathbf{p}_{n_d}^{(t)}, \end{aligned}$$

Optimum solution:
$\lambda_t^*,\ \mathbf{p}_1^{(t)*}, \ldots, \mathbf{p}_{n_d}^{(t)*}$
Deflation of the output:

$$y_i^{(t+1)} = y_i^{(t)} - \lambda_t^* \prod_{d=1}^{n_d} \langle \mathbf{p}_d^{(t)*}, \mathbf{x}_i \rangle, \forall i$$

- **The LTR model can be estimated as a series of rank-one problems**
- **Each subproblem finds a rank-one parameter tensor that best fits the residuals of the output**
- **The subproblem can be solved by gradient approaches (e.g. ADAM)**

Wang, T., Szedmak, S., Wang, H., Aittokallio, T., Pahikkala, T., Cichonska, A. and Rousu, J., 2021. Modeling drug combination effects via latent tensor reconstruction. Bioinformatics, 37(Supplement_1), pp.i93-i101.

LTR vs HOFM

- HOFMs are limited to symmetric polynomials i.e. f(i,j) = f(j,i) while LTR is not → LTR is better as a general regression method
- In LTR, the input data can be fed through an activation function to give rise to a learnable embedding (e.g. non-linear dimensionality reduction): $\phi(\mathbf{x}) = \mathcal{A}(\mathbf{U}^{(d)T}\mathbf{x})$
- LTR requires more training data than HOFMs due to the more flexible function class
- LTR is arguably more interpretable due to the simpler functional form

LTR:

$$f(\mathbf{x}) = \sum_{t=1}^{n_t} \lambda_t \left\langle \mathbf{p}_1^{(t)} \otimes \cdots \otimes \mathbf{p}_{n_d}^{(t)}, \mathbf{x} \otimes \cdots \otimes \mathbf{x} \right\rangle .$$
$$= \sum_{t=1}^{n_t} \lambda_t \left\langle \mathbf{p}_1^{(t)}, \mathbf{x} \right\rangle \cdot \ldots \cdot \left\langle \mathbf{p}_{n_d}^{(t)}, \mathbf{x} \right\rangle ,$$

HOFM:

$$\hat{y}HOFM(x)$$
$$= \sum_j w_j x_j + \sum_{j\prime>j} \langle {p_j}^{(2)}, {p_{j\prime}}^{(2)} \rangle x_j x_{j\prime} + \cdots$$
$$+ \sum_{j_m>\cdots>j_1} \langle {p_{j_1}}^{(m)}, \ldots, {p_{j_m}}^{(m)} \rangle x_{j_1} \cdots x_{j_m}$$

Experiments on drug combination response prediction

Compared methods

- **ComboFM** (Julkunen et al., 2020): relies on a recent machine learning technology called **higher-order factorization machines** (HOFM) that allow capturing the multi-way interactions between drug combinations and their targets
- **ComboLTR** (Wang et al. Bioinformatics, 2021): a **latent tensor reconstruction** method, which removes some limitations of HOFMs
- **Random forest regressor (RF).** Strong baseline e.g. winner of the AstraZeneca-Sanger DREAM Challenge.

Julkunen, H., Cichonska, A., Gautam, P., Szedmak, S., Douat, J., Pahikkala, T., Aittokallio, T. and Rousu, J., 2020. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nature communications*, *11*(1), pp.1-11.

Wang, T., Szedmak, S., Wang, H., Aittokallio, T., Pahikkala, T., Cichonska, A. and Rousu, J., 2021. Modeling drug combination effects via latent tensor reconstruction. Bioinformatics, 37(Supplement_1), pp.i93-i101.

Dataset

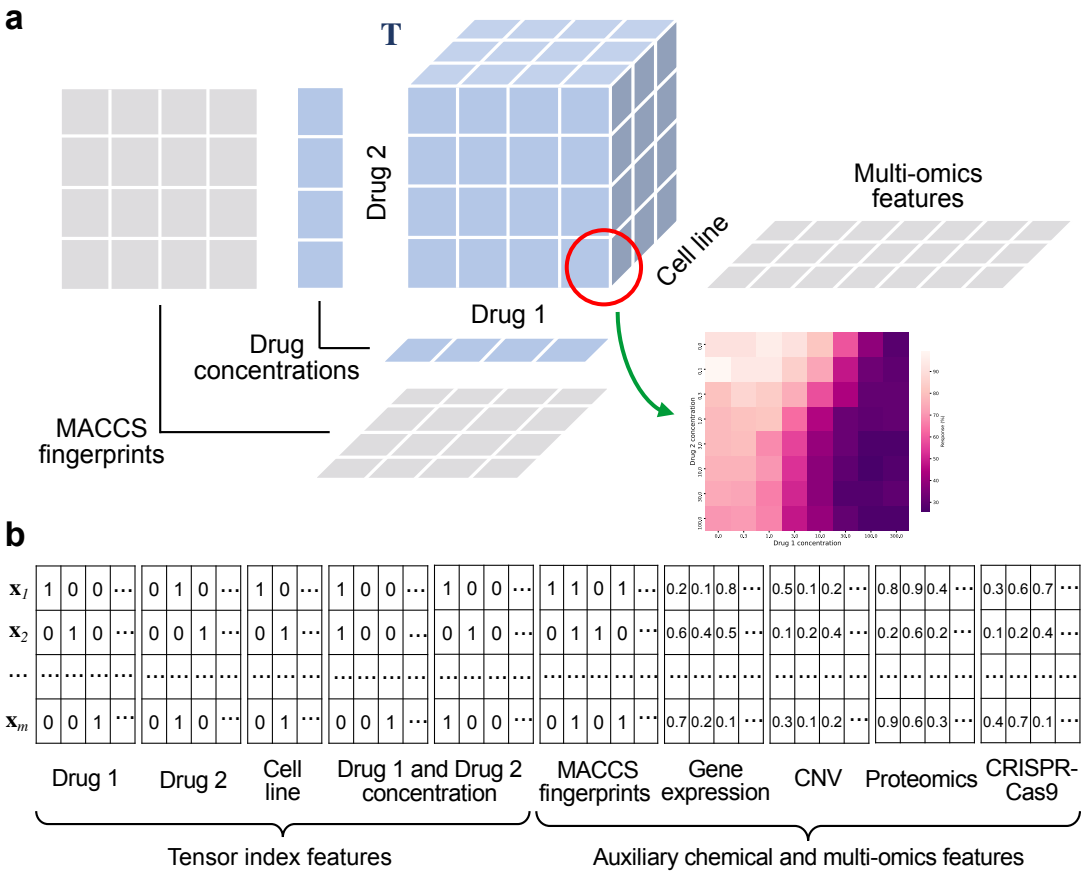
- We used the NCI-ALMANAC data, large drug combinations screening data by NCI
- We used a subset of this data where omics data for cell lines was availables:
  - 828 324 response measurements of 5 035 drug combinations and
  - 15 396 monotherapies in
  - 19 cancer cell lines originating from 9 tissue types.

https://dtp.cancer.gov/ncialmanac/

Holbeck et al., 2017. The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer research*, *77*(13), pp.3564-3576.

Data structure

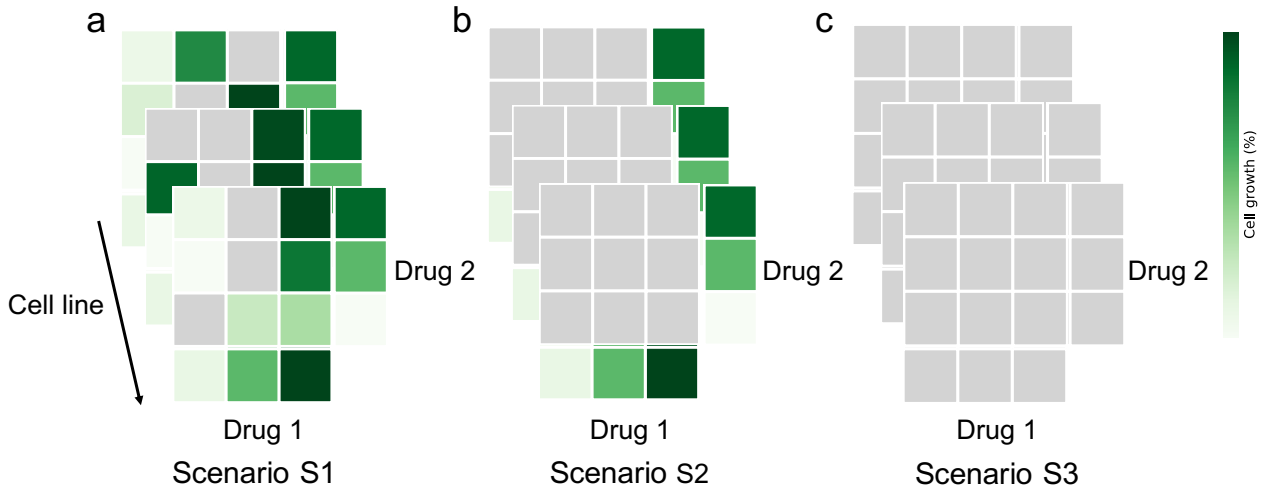
a. 5-order tensor containing the drug combination responses
b. flattened representation for the learning algorithms



Wang, T., Szedmak, S., Wang, H., Aittokallio, T., Pahikkala, T., Cichonska, A. and Rousu, J., 2021. Modeling drug combination effects via latent tensor reconstruction. Bioinformatics, 37(Supplement_1), pp.i93-i101.

Prediction scenarios

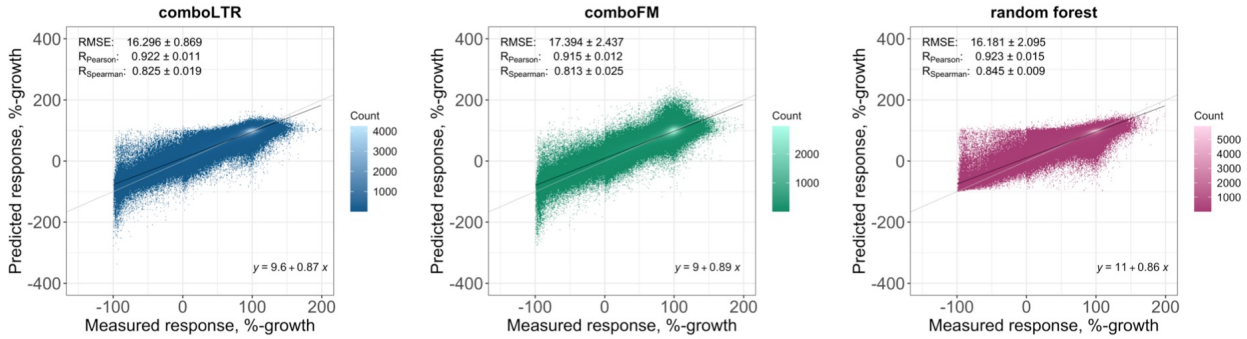
- **Scenario S1: Predicting new response matrix entries**
- **Scenario S2: Predicting new drug combination responses (monotherapy responses known)**
- **Scenario S3: Predicting new combination responses w.o. monotherapy measurements**



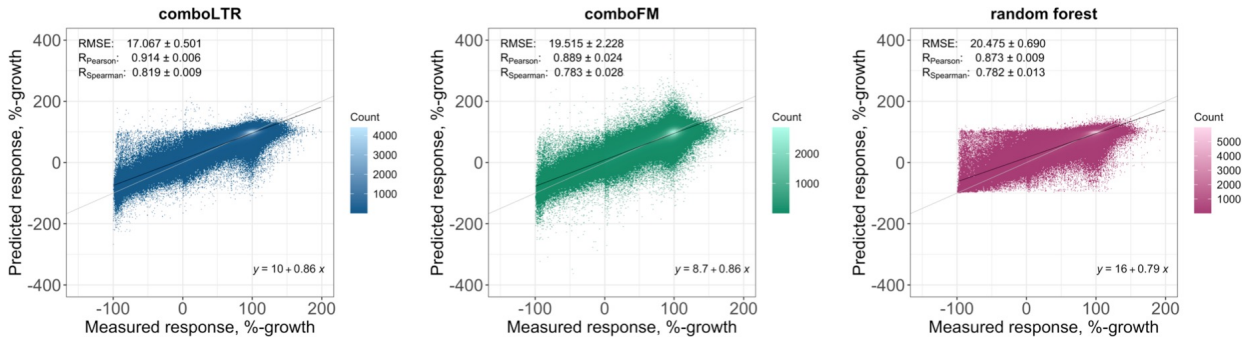
Wang, T., Szedmak, S., Wang, H., Aittokallio, T., Pahikkala, T., Cichonska, A. and Rousu, J., 2021. Modeling drug combination effects via latent tensor reconstruction. Bioinformatics, 37(Supplement_1), pp.i93-i101.

Prediction performance

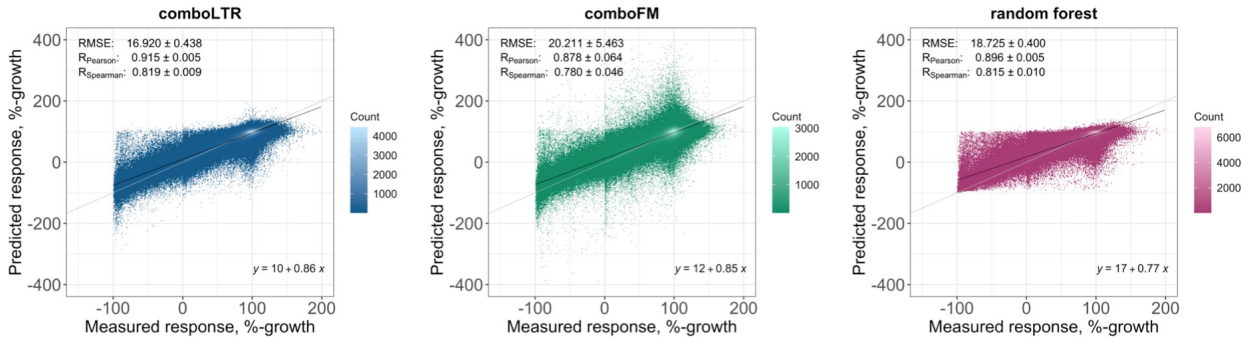
a **S1: Filling in the gaps in partially measured dose-response matrices**



b **S2: Predicting dose-response matrices of new drug combinations**



c **S3: Predicting dose-response matrices of new drug combinations w/o monotherapy responses**



Effect of different data sources

- The predictive performance of the models using different input data sources was studied
- In all prediction scenarios, the primary data (drug features, cell line features) was shown to be very small
- The response measurements from similar drugs and cell lines ('Tensor indices') dominates

Table 2. Performance of *comboLTR*, *comboFM*, and random forest (RF) under different prediction scenarios and using different features. Pearson correlations between predicted and measured drug combination responses, reported as averages across 5 cross validation folds ± standard deviations.

Features	Method	S1	S2	S3
Tensor indices	*comboLTR*	0.915±0.009	0.894±0.002	0.893±0.003
	comboFM	0.920±0.010	0.914±0.003	0.907±0.004
	RF	0.886±0.019	0.853±0.010	0.858±0.010
Tensor indices + MACCS	*comboLTR*	0.921±0.010	0.908±0.003	0.910±0.003
	comboFM	0.923±0.012	0.923±0.005	0.913±0.005
	RF	0.921±0.016	0.872±0.009	0.894±0.005
Tensor indices + Multi-omics	*comboLTR*	0.908±0.014	0.909±0.007	0.911±0.005
	comboFM	0.910±0.027	0.904±0.014	0.870±0.064
	RF	0.895±0.019	0.859±0.010	0.865±0.010
Tensor indices + MACCS + Multi-omics	*comboLTR*	0.922±0.011	0.914±0.006	0.915±0.005
	comboFM	0.915±0.012	0.889±0.024	0.878±0.064
	RF	0.923±0.015	0.873±0.009	0.896±0.005

Wang, T., Szedmak, S., Wang, H., Aittokallio, T., Pahikkala, T., Cichonska, A. and Rousu, J., 2021. Modeling drug combination effects via latent tensor reconstruction. Bioinformatics, 37(Supplement_1), pp.i93-i101.

Time and space complexity

- **The time consumption of comboLTR is on par with RF, while comboFM is significantly less time-efficient**
- **All three methods have similar memory-requirements**

Table 3. The time (h) and memory (GB) usage of *comboLTR*, *comboFM* and random forest in 5-fold cross validation.

	Time (h) / Memory (GB)		
Features	*comboLTR*	*comboFM*	RF
One-hot encoding	1.1 / 24	1.2 / 20	1.2 / 24
One-hot encoding + MACCS	2.1 / 35	5.9 / 35	0.4 / 38
One-hot encoding + Multi-omics	3.2 / 44	31.5 / 63	5.6 / 54
One-hot encoding + MACCS + Multi-omics	3.1 / 74	39.0 / 73	2.1 / 74

Wang, T., Szedmak, S., Wang, H., Aittokallio, T., Pahikkala, T., Cichonska, A. and Rousu, J., 2021. Modeling drug combination effects via latent tensor reconstruction. Bioinformatics, 37(Supplement_1), pp.i93-i101.

Summary

Summary

- **The search for new combinatorial therapies requires prioritization due to the huge search space of drug combinations**
- **Machine learning tools can be used to predict the drug combination responses and synergies, and thus help to prioritize the search**
- **The predictive accuracy of ML models depends strongly on the assumed prediction scenarios**
- **comboFM and comboLTR uses factorization machine technology to learn pairwise drug combination responses**
  - monotherapy responses of the drugs alone, and
  - responses of similar drug combinations in similar cell lines

Thank you for your attention!

ACADEMY OF FINLAND

Cancer Society of Finland

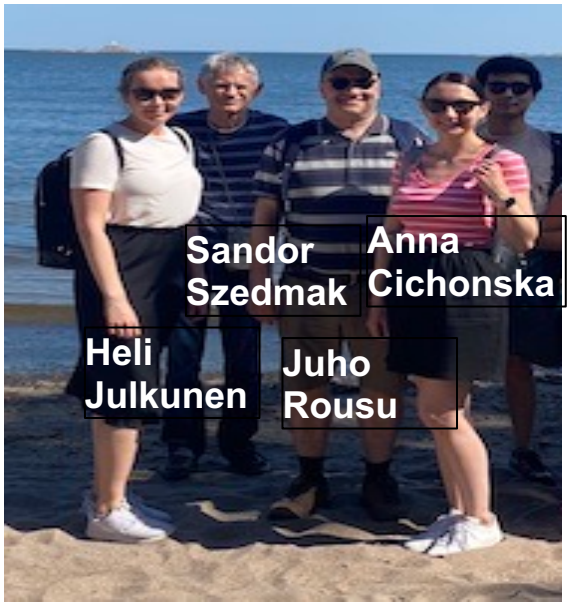
ORION Research Foundation

SIGRID JUSELIUS FOUNDATION

FIMM Institute for Molecular Medicine Finland
Nordic EMBL Partnership for Molecular Medicine

UNIVERSITY OF TURKU

UNIVERSITY OF HELSINKI



Tianduanyi Wang

Prson Gautam

Tero Aittokallio

Tapio Pahikkala

Julkunen, H., et al. 2020. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nature communications*, *11*(1), pp.1-11.

Wang, T., et al. 2021. Modeling drug combination effects via latent tensor reconstruction. Bioinformatics, 37(Supplement_1), pp.i93-i101.

Aalto University

comboFM: https://github.com/aalto-ics-kepaco/comboFM
comboLTR: https://github.com/aalto-ics-kepaco/ComboLTR