Multi-task learning in the analysis of phenotypic data

Adam Arany
KU Leuven, Belgium

KU LEUVEN

15 May, 2019
Bled, Slovenia

Overview

- Multi-task learning
  - Motivation
  - Recommender systems view
- Matrix factorization
  - Matrix factorization with side information
  - Beyond 2-way relations
  - Connection to Deep Learning
- Application I: Chemogenomics
- Application II: Repurposing Imaging screens
- Application III: Gaining insight about mechanism of action

Multi-task learning

Motivation

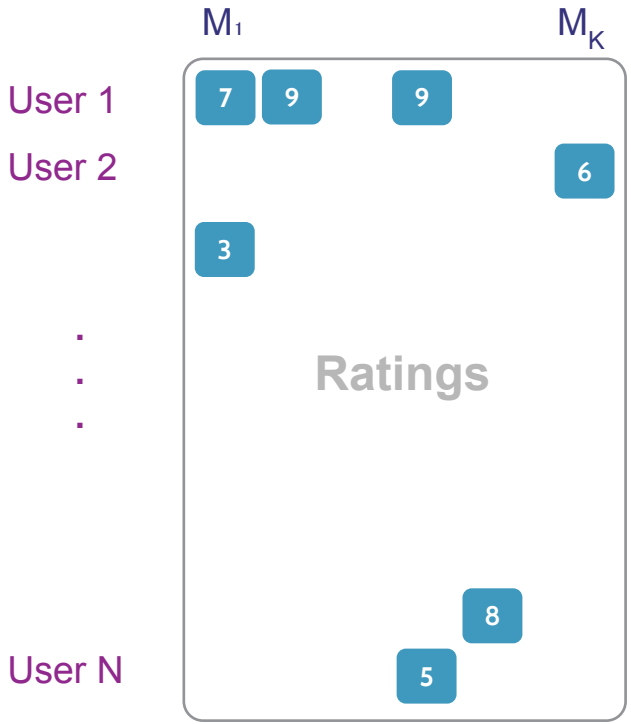
- Biological learning agents are not single-task
  - Playing ball games
  - Studying languages
  - Playing video games
- Drug targets are similar
  - Orthology
  - Paralogy
  - Convergent evolution
- Pervasive features

Recommender systems

Netflix challenge

- **U**sers rating **M**ovies
- Very sparse matrix
- Filling in missing values

"If Alice likes Lord of the Rings and Harry Potter, and Bob likes Lord of the Rings, he will probably likes Harry Potter."

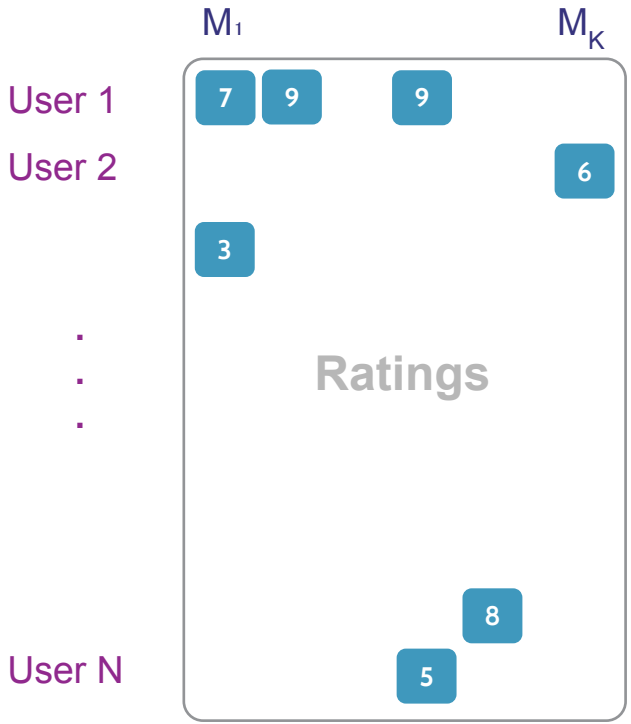


Recommender systems

Netflix challenge

- **Cold start problem**
- A user or a movie have very few or no rating
- Incorporate information on movies and users : **side information**

“ The Hobbit is new in our system. Can we give prediction?”



Matrix Factorization

Matrix factorization

- Represent each user and movie by a **latent vector**

$\mathbf{u}_i = [\, u_{i,1} \;\; u_{i,2} \;\; \ldots \;\; u_{i,D} \,]$

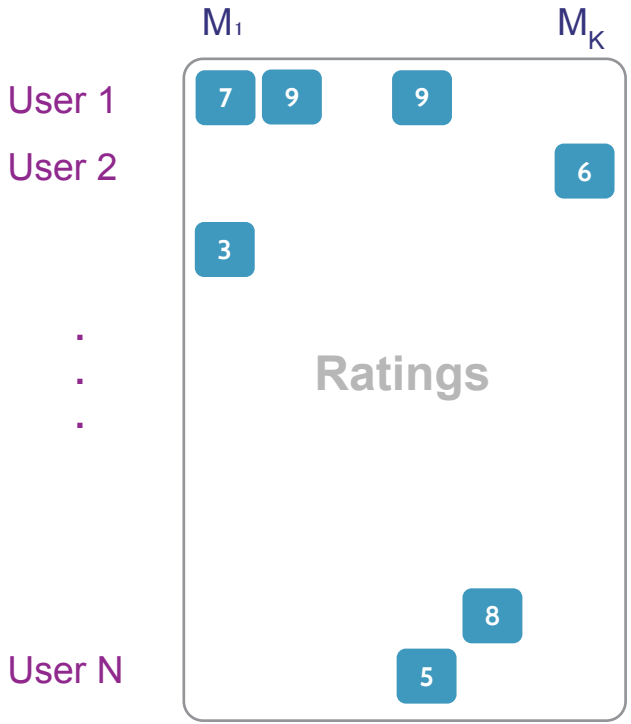
$\mathbf{v}_j = [\, v_{j,1} \;\; v_{j,2} \;\; \ldots \;\; v_{j,D} \,]$

- Prediction is made by

$$\hat{R}_{ij} = \mathbf{u}_i^T \mathbf{v}_j$$

- Learn latent vectors by minimizing error

$$\min \sum_{ij} I_{ij}(R_{ij} - \hat{R}_{ij})^2$$



Matrix factorization

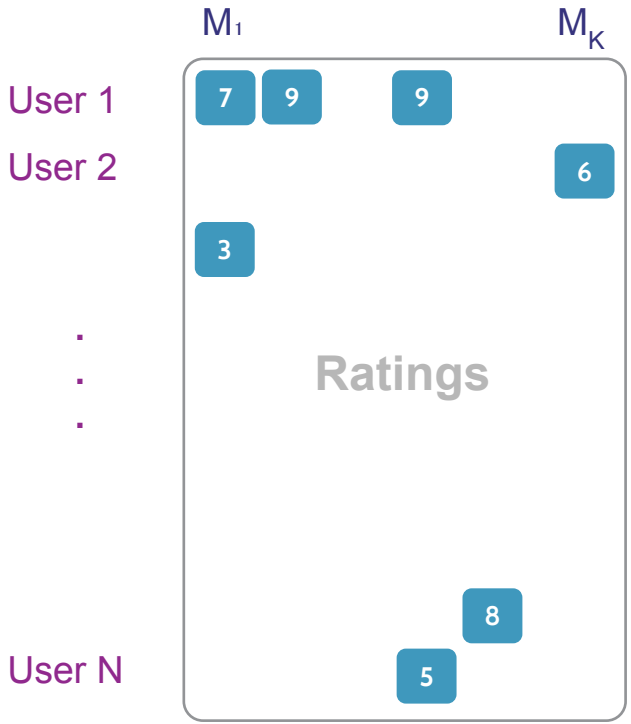
- Each latent dimension will focus on certain aspect of the entities
- For example
  - whether a movie has action
  - whether user likes action movies

$\mathbf{v}_1$ = [2.1 | 0.1 . . . 0.5]

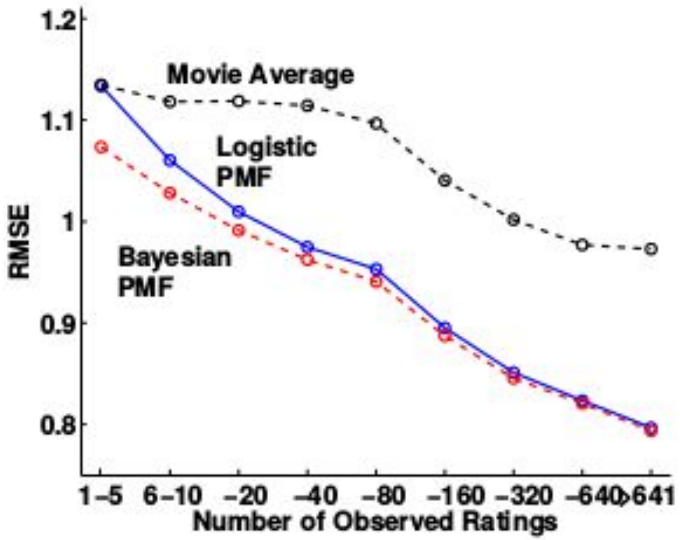
$\mathbf{u}_1$ = [3.7 | -0.1 . . . 1.2]

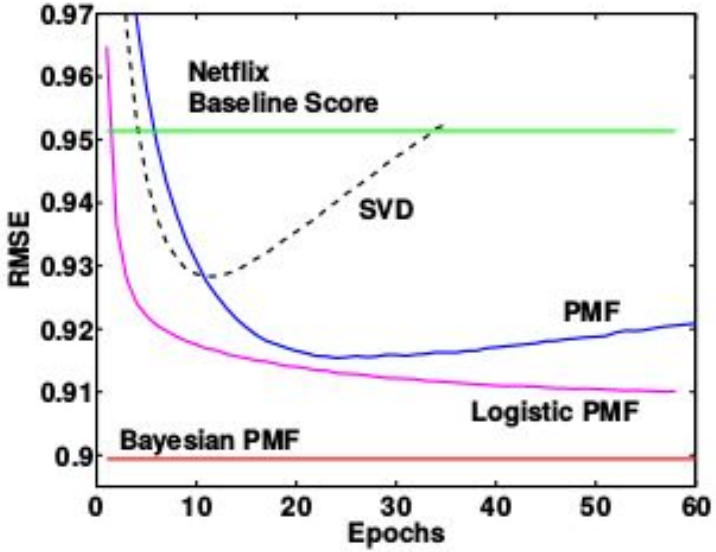
$\mathbf{u}_2$ = [0 | -0.1 . . . 1.2]

- These patterns are learned from data



Results on Netflix challenge





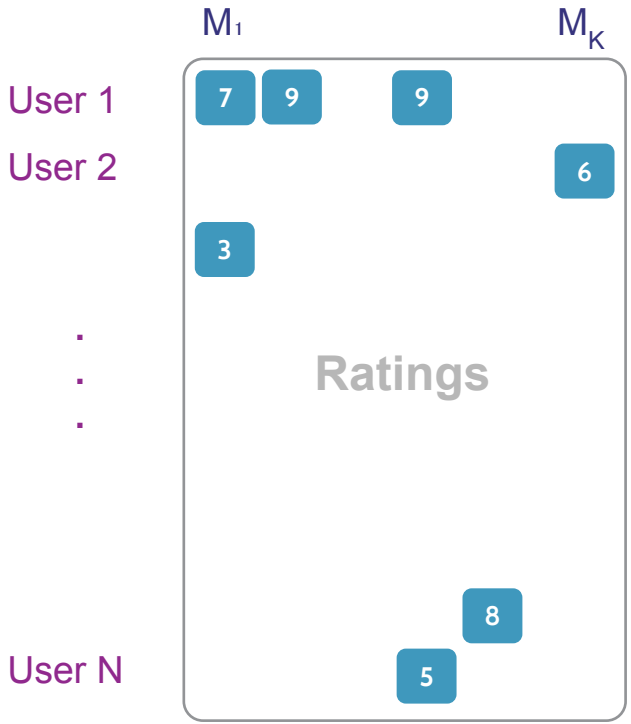
Source: Salakhutdinov, R., & Mnih, A., ICML 2008

Matrix Factorization

Adding Side Information

Cold start problem

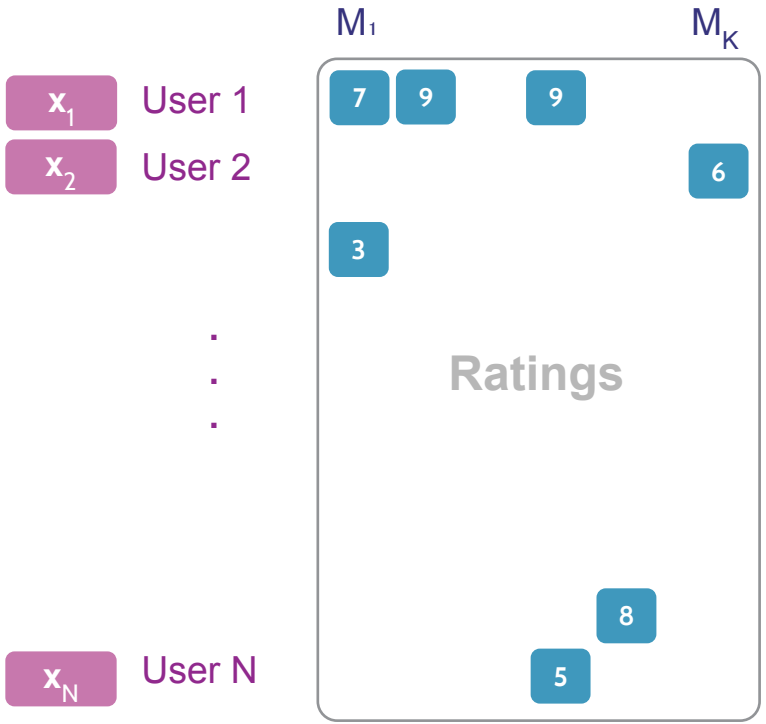
- **Cold start problem**
- A user or a movie have very few or no rating
- **No data** to learn about a given latent vector!
- Incorporate information on movies and users : **side information**



Matrix factorization with side information

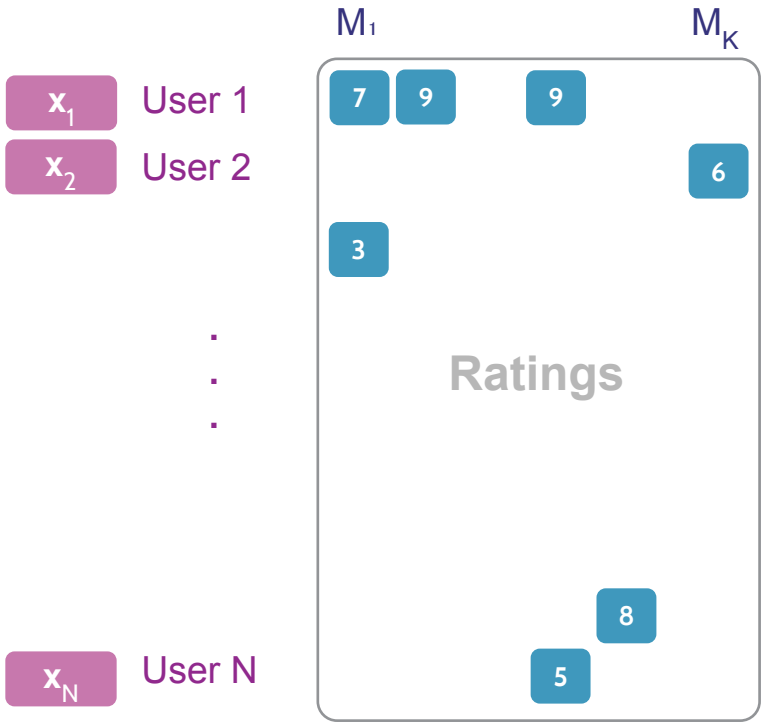
- Side information (**x**)
- Features that might have info on ratings
  - User: age, gender, location
  - Movie: Genre, director, …
- Learn how features predict latents

  - β maps features to latents
  - Learn **u**,**v**,β jointly



Matrix factorization with side information

- Rich side information can be very predictive
- Examples
  - Compounds : structure fingerprints
  - Genes: ontology classification



Beyond 2-way relations

- More than two object relations
- *e.g.*: drug *x* gene *x* cell line

$$\mathbf{d}_i = [\, d_{i,1} \;\; d_{i,2} \;\; \ldots \;\; d_{i,D} \,]$$

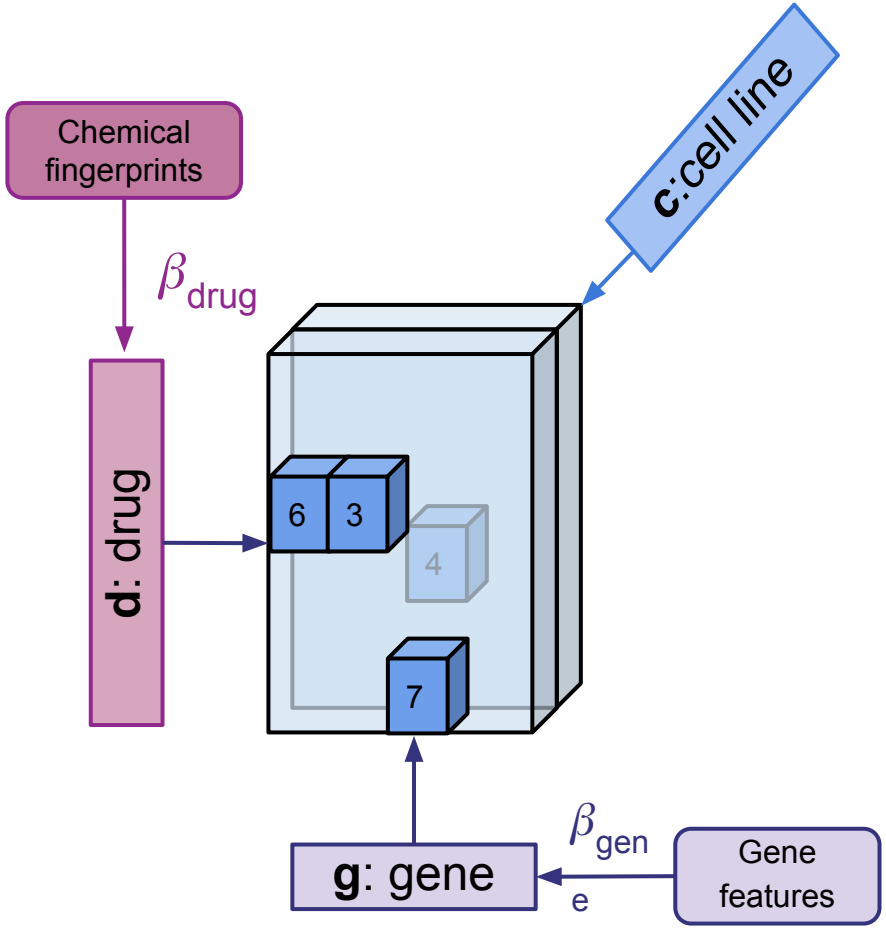
$$\mathbf{g}_j = [\, g_{j,1} \;\; g_{j,2} \;\; \ldots \;\; g_{j,D} \,]$$

$$\mathbf{c}_k = [\, c_{k,1} \;\; c_{j,2} \;\; \ldots \;\; c_{j,D} \,]$$

- The model is:

$$Y_{ijk} = d_{i,1}g_{j,1}c_{k,1} + \ldots + d_{i,D}g_{j,D}c_{k,D}$$

"Treating cell line k with drug i results in differential expression Y_{ijk} in gene j."



Matrix Factorization

Bayesian variant

J. Simm, A. Arany, et al. IEEE MLSP, Tokyo, Japan, September 2017
URL: https://ieeexplore.ieee.org/document/8168143

Probabilistic Model

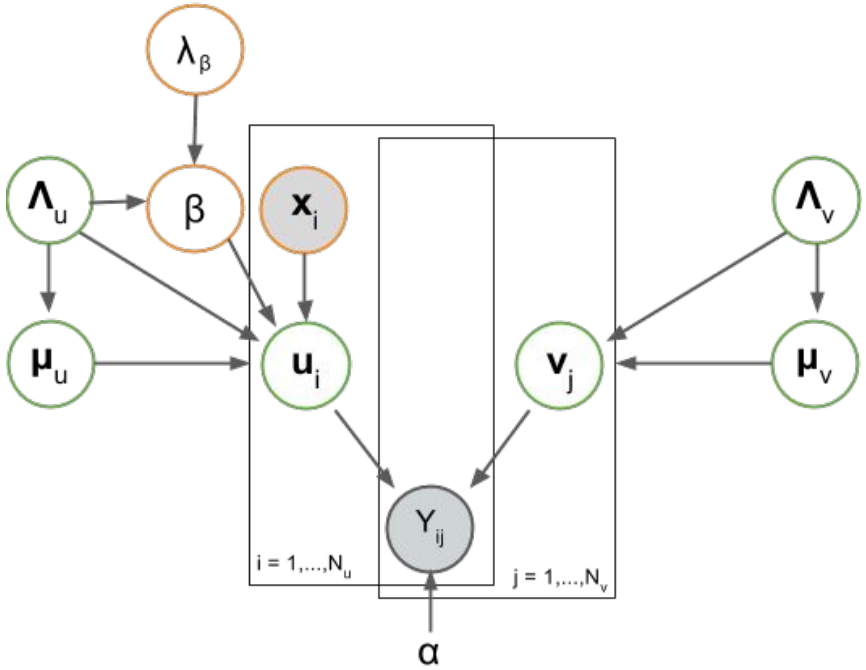
- We learn **u** and **v** such that

$$Y_{ij} \sim \mathcal{N}(\mathbf{u}_i^\top \mathbf{v}_j, \alpha^{-1})$$

- side information In the prior mean of latent vectors

$$\mathbf{u}_i \sim \mathcal{N}(\mu_U + \beta^\top \mathbf{x}_i, \Lambda_U)$$

- Gibbs sampling

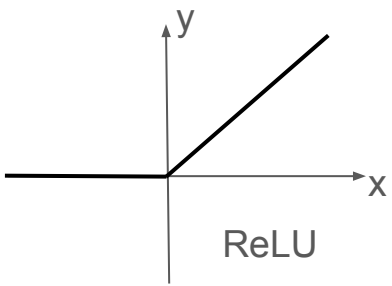


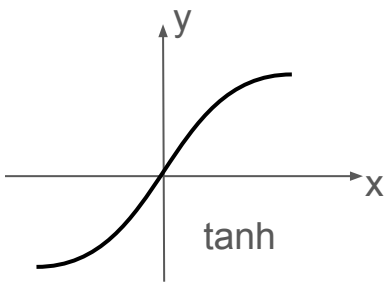
Matrix Factorization

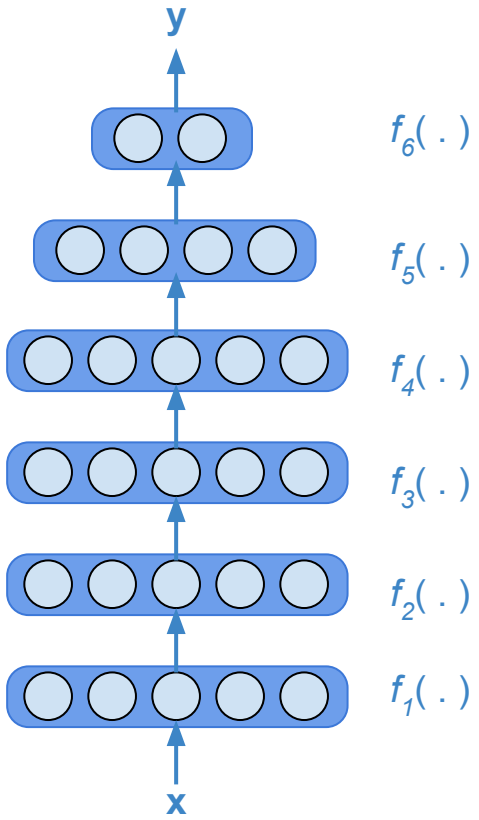
Connection to Deep Learning

Deep Learning

- Learning representation with nested function composition
- A usual choice for $f_n(.)$ is
  $f(\mathbf{x}) = \sigma(\mathbf{W}^T\mathbf{x} + \mathbf{b})$
  Where σ is a nonlinearity, called the **activation function**, applied elementwise



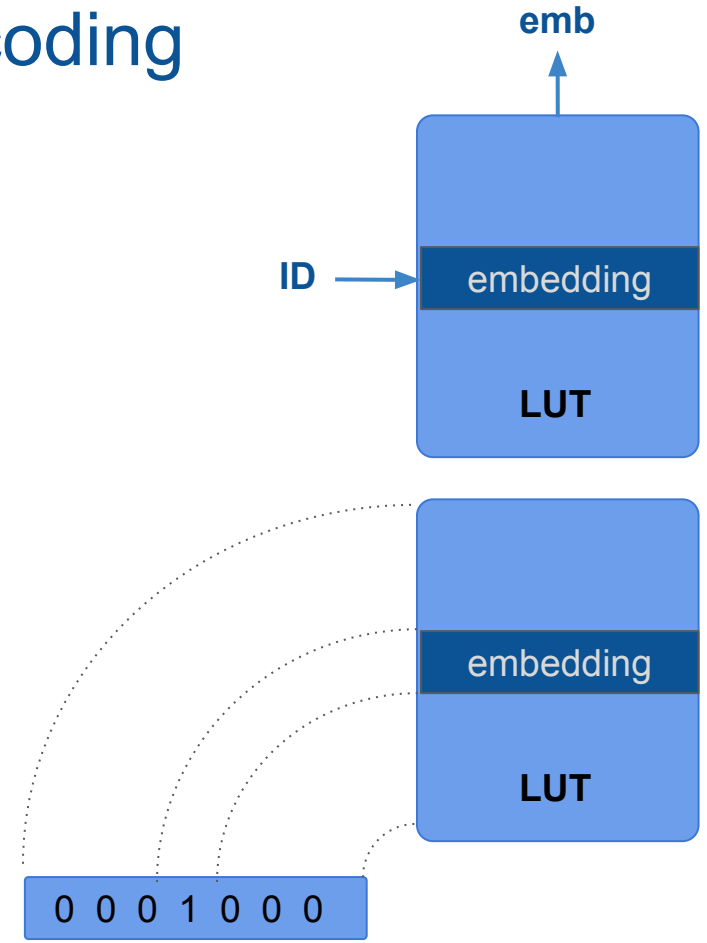




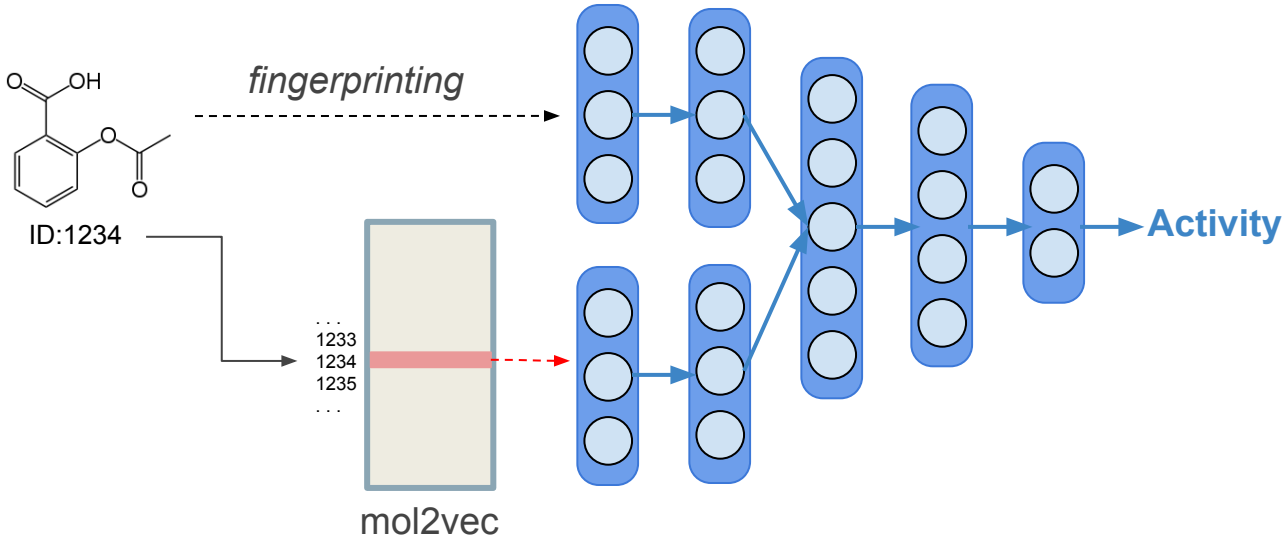
$\mathbf{y} = f_6(f_5(f_4(f_3(f_2(f_1(\mathbf{x}))))))$

Embedding and one-hot encoding

- We learn an **embedding** to every entity (*e.g.* every user)
- Feed it into the deep network
- Address a **lookup table** (LUT) by ID
- Equivalently multiplying a weight matrix with a **one-hot encoding**
- This introduces entity level reasoning into the network



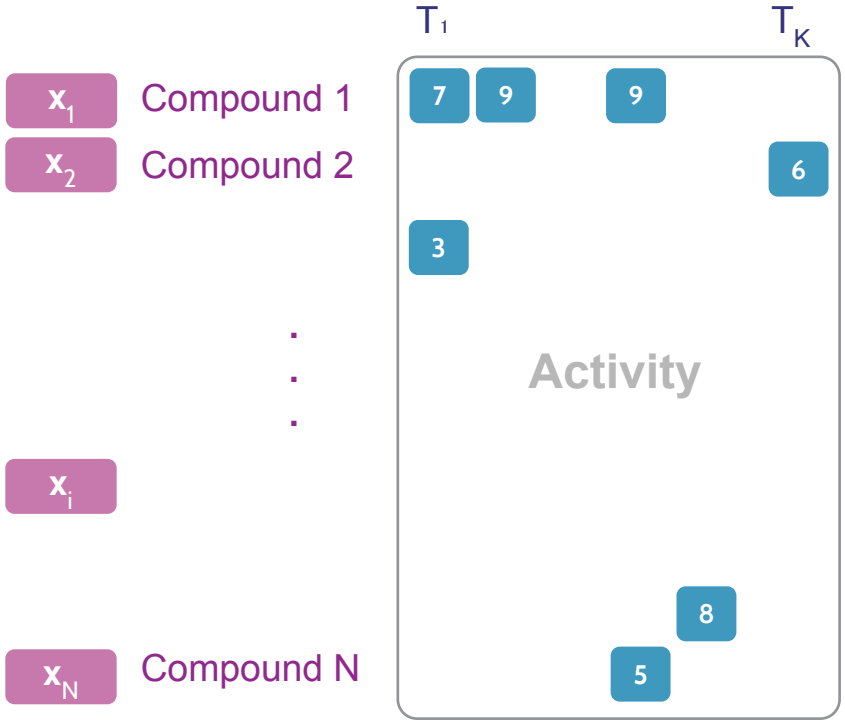
Deep factorization model



Application: Chemogenomics

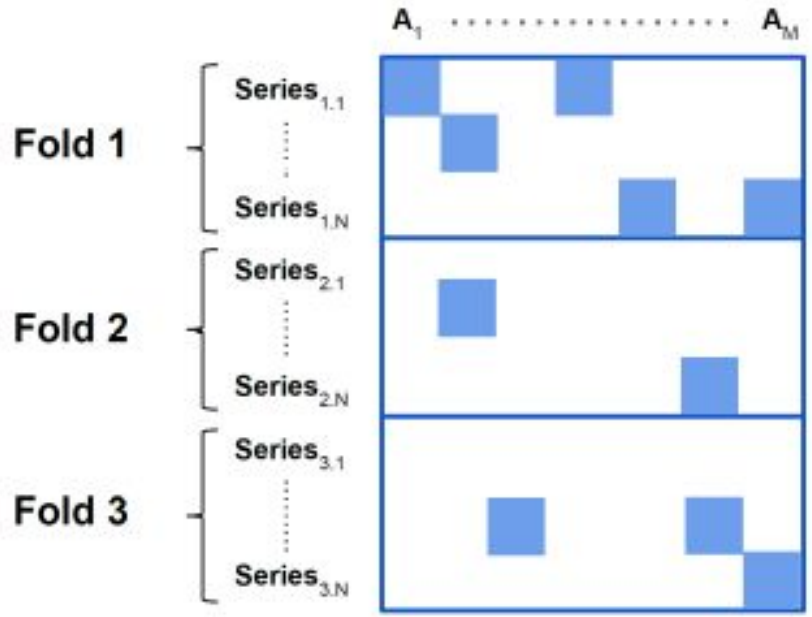
Chemogenomics

- Compound-target activity matrix
- Side information on compounds
  - Fingerprint of structure
  - Phenotypic effect
  - Gene expression changes
- Sparsely filled
- Protein side information
  - Proteochemometrics



Series effect

- **Chemical series** effect
  - Same scaffold small modification
  - They need to end up in the same fold
- **Clustering** the structures
  - *e.g.* Tanimoto > 0.5
  - using ECFP6
- **Nested cross-validation** over the clusters
  - For hyperparameter tuning



Evaluation results

- Model: Non-linear version
- Dataset: ChEMBL v23
  - #**Assays**: 702
  - #**Compounds**: 296k
  - #Activity threshold: 4
  - Clustering threshold: 0.5
- **AUC: 0.8574**
- (Non clustered: 0.9195)

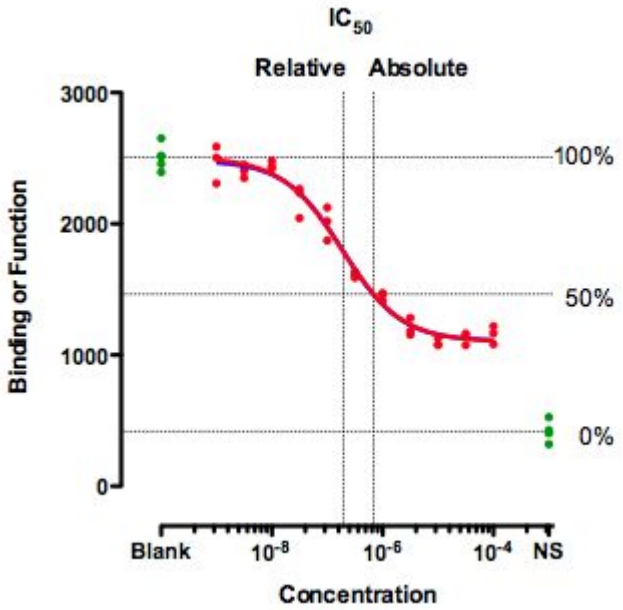
Quality	#Assay - threshold pair
> 0.9 AUC	550 / 2808
> 0.8 AUC	1156 / 2808
> 0.7 AUC	1355 / 2808

Application: Binding mode

A. Arany, J. Simm, et al. MLCB/MLSB 2015, NIPS Workshop, December 2015
URL: https://arxiv.org/abs/1512.00315

Affinity and potency

- **Affinity:** Amount required to occupy a given proportion of the target
  - e.g: Ki
- **Potency**: Amount required to produce an effect of given intensity
  - e.g.: IC50



Competitive vs. non-competitive inhibition

For non-competitive:

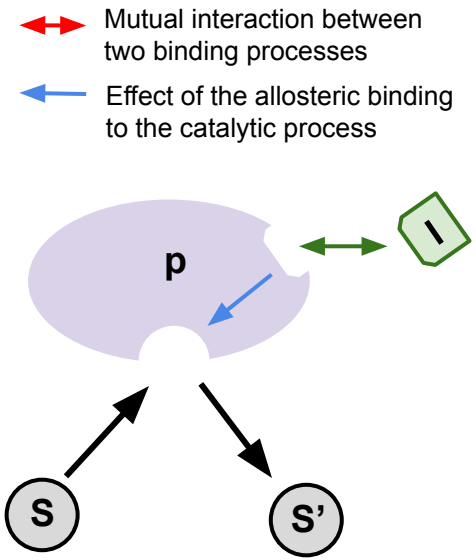
$$pK_i = pIC_{50}$$

For competitive:

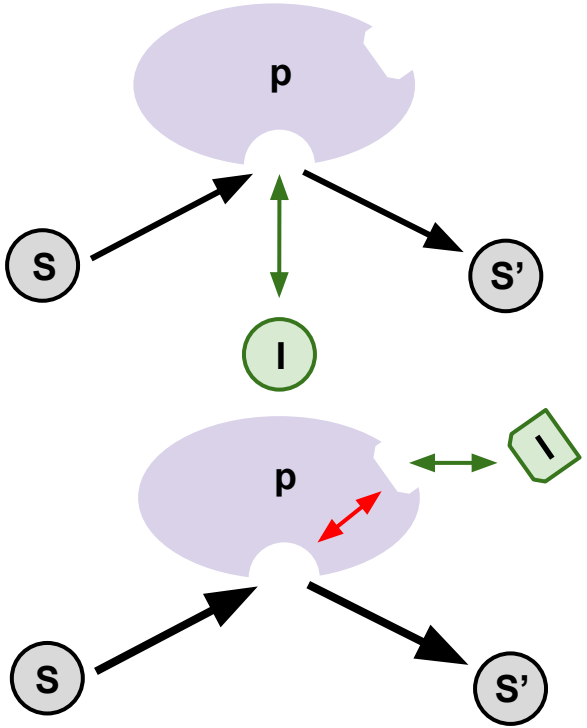
$$pK_i = pIC_{50} + C\left([S], K_m\right)$$

Assumption:
Enzymes follow Michaelis-Menten kinetics

Goal:
Predict which model is true



Non-competitive interaction



Competitive interaction

Identification of binding mode

- drug *x* target *x* measurement type

$$\mathbf{d}_i = [\, d_{i,1} \; d_{i,2} \; \ldots \; d_{i,D} \,]$$

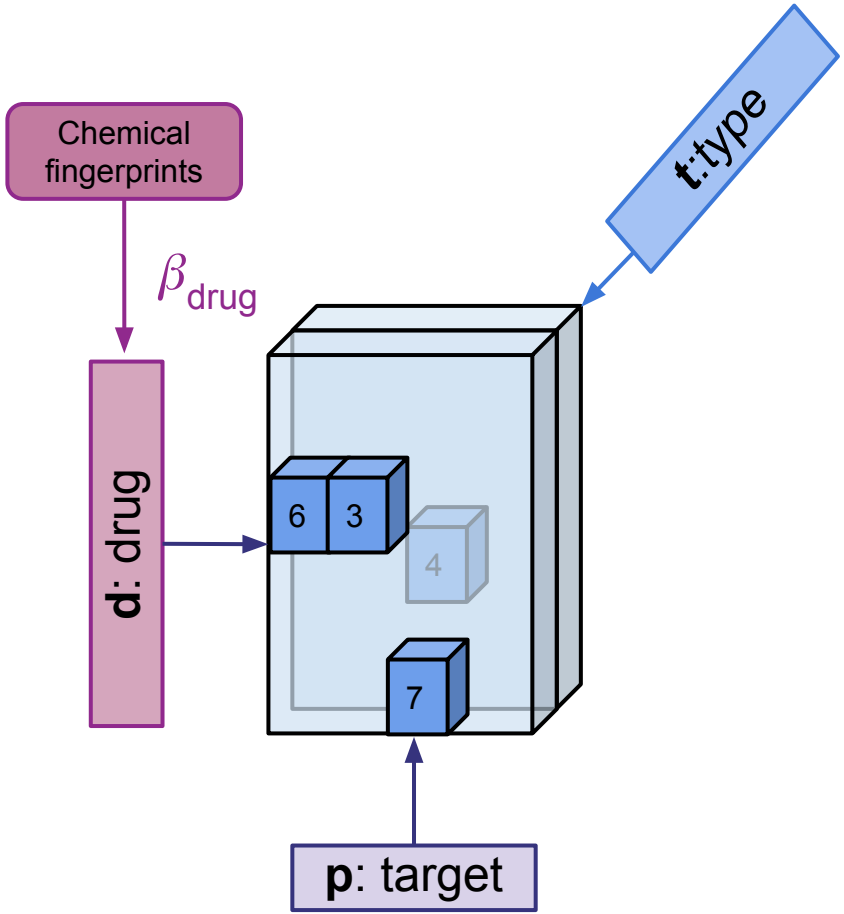
$$\mathbf{p}_j = [\, p_{j,1} \; p_{j,2} \; \ldots \; p_{j,D} \,]$$

$$\mathbf{t}_k = [\, t_{k,1} \; t_{j,2} \; \ldots \; t_{j,D} \,]$$

- The model is:

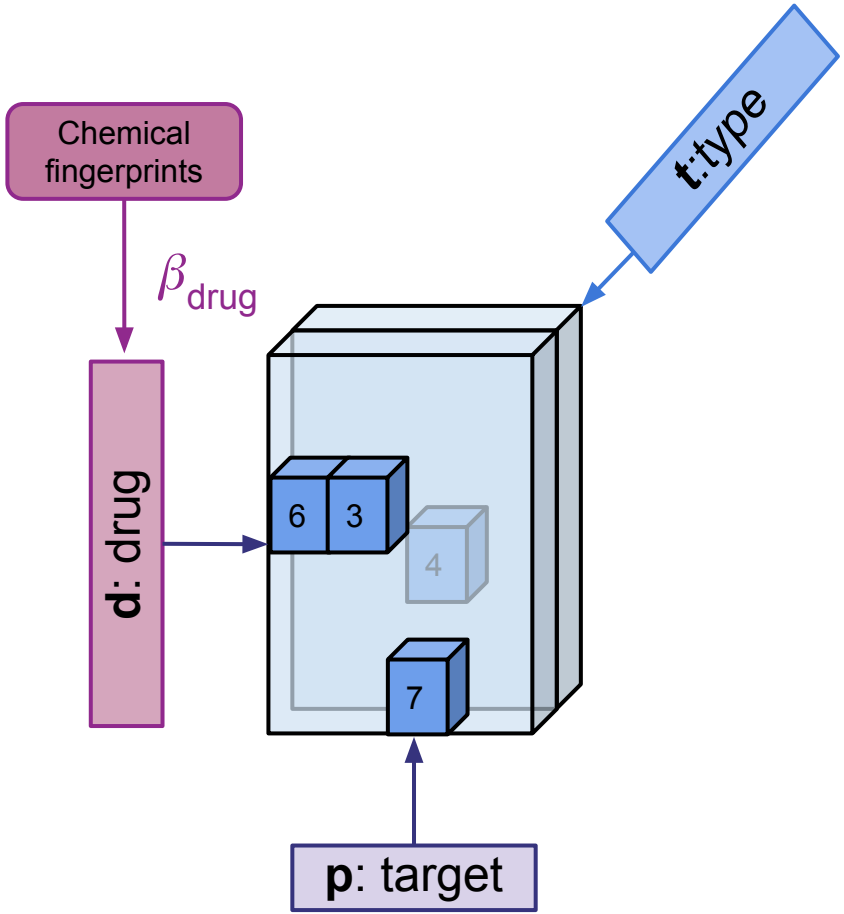
$$Y_{ijk} = d_{i,1}p_{j,1}t_{k,1} + \ldots + d_{i,D}p_{j,D}t_{k,D}$$

Q1: *Dominant interaction mode for a target*
Q2: *Interaction mode for a target-ligand pair*



The setup

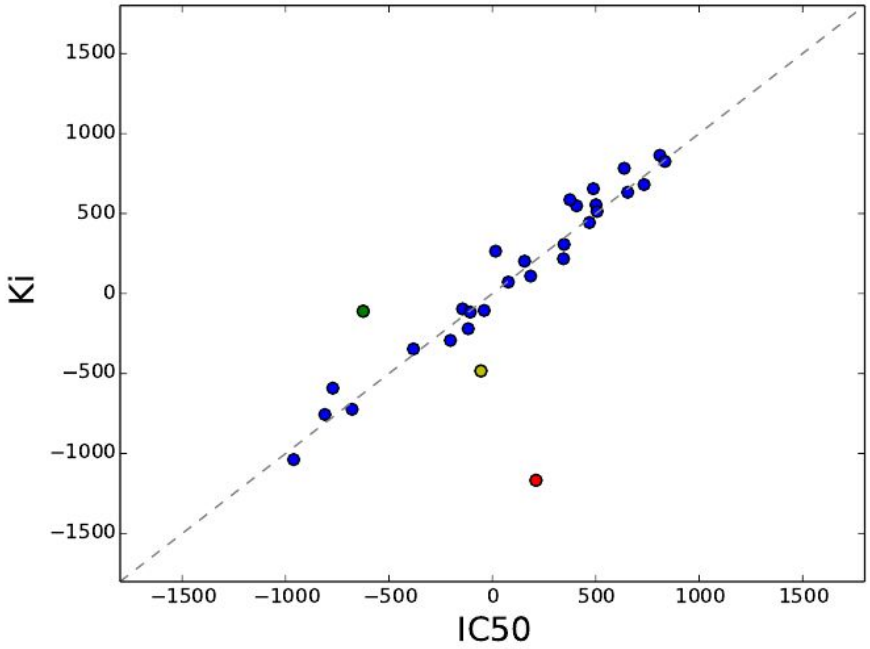
- Model: Bayesian version
  - 30 latent dimensions
- Dataset: ChEMBL v19
  - Proteins: 346
  - Compounds: 15k
  - **59k IC50** observations
  - **3121 Ki** observations
  - Fingerprint: ECFP6 (dim: 106k)



Latent dimensions for IC_{50} and K_i

- Compared latent vectors for IC_{50} and K_i (t_0 and t_1)
- There are 3 latent dimensions encoding their differences

$$pK_i = pIC_{50} + C\left([S], K_m\right)$$



$$\bar{t}_{d,IC50} = t_{d,IC50} \cdot |\mathbf{U}_d| \cdot |\mathbf{V}_d|$$

Predicting interaction of pairs

- We chose 30 protein-compound pairs with both Ki and IC50 measurements
  - Randomly 10 from top 100 pairs with highest pKi - pIC50
  - Randomly 10 from top 100 pairs with highest pIC50 - pKi
  - Randomly 10 from 100 pairs with smallest abs(pIC50 - pKi)
- Removed them to test set and tried to predict
- Predicted abs(pKi - pIC50) for all test pairs

Protein-compounds	Mean abs(pKi - pIC50)
Competitive pairs	0.671
Non-competitive pairs	0.234

Predicting dominant behaviour

TOP10 Predicted competitive

ChEMBL ID	Protein name
CHEMBL284	Dipeptidyl peptidase IV
CHEMBL325	Histone deacetylase 1
CHEMBL260	MAP kinase p38 alpha
CHEMBL1865	Histone deacetylase 6
CHEMBL1937	Histone deacetylase 2
CHEMBL289	Cytochrome P450 2D6
CHEMBL4005	PI3-kinase p110-alpha subunit
CHEMBL1978	Cytochrome P450 19A1
CHEMBL2581	Cathepsin D
CHEMBL4793	Dipeptidyl peptidase IX

TOP10 Predicted uncompetitive

ChEMBL ID	Protein name
CHEMBL240	HERG
CHEMBL3772	Metabotropic glutamate receptor 1
CHEMBL5145	Serine/threonine-protein kinase B-raf
CHEMBL3663	Growth factor receptor-bound protein 2
CHEMBL4641	Voltage-gated T-type calcium channel alpha-1G subunit
CHEMBL3230	Sphingosine 1-phosphate receptor Edg-6
CHEMBL2001	Purinergic receptor P2Y12
CHEMBL1785	Endothelin receptor ET-B
CHEMBL287	Sigma opioid receptor
CHEMBL3227	Metabotropic glutamate receptor 5

Possible future work

We are interested to collaborate on analysis of different type of **compound - protein interaction** data:

- Unprocessed measurement series for IC_{50} determination
- Hill slope
- **k_{on}, k_{off}**
- ...

Cell Chemical Biology

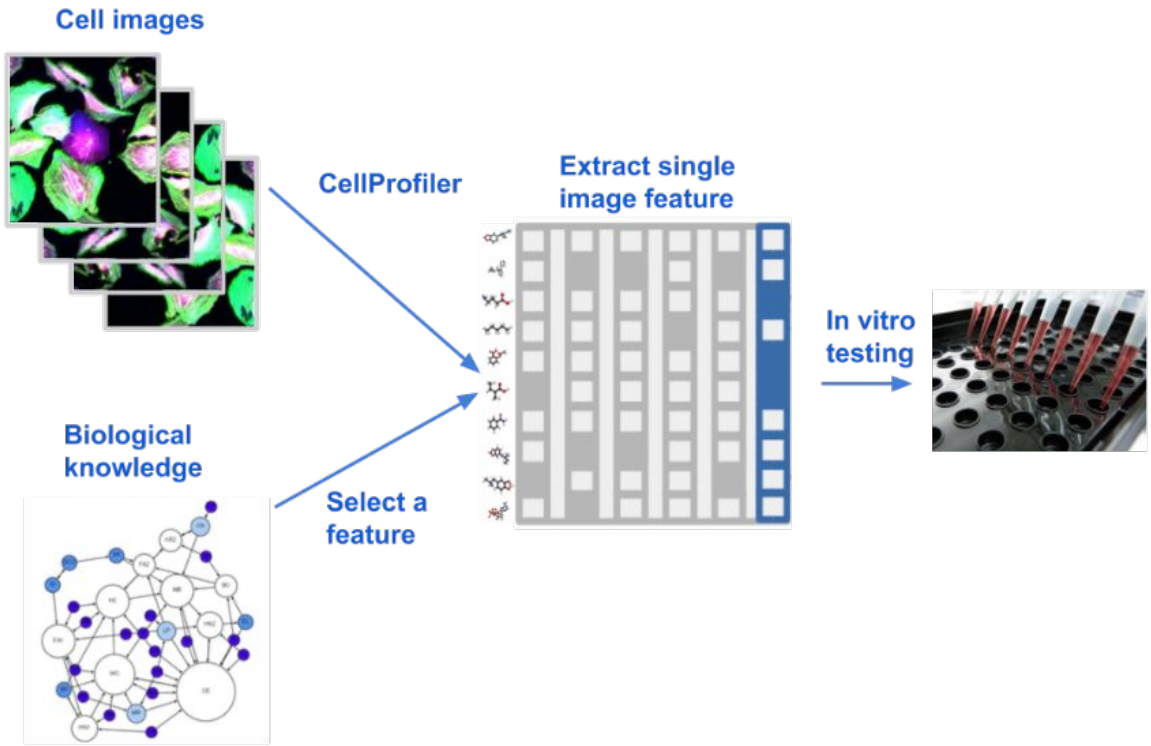
RESOURCE | VOLUME 25, ISSUE 5, P611-618.E3, MAY 17, 2018

Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery

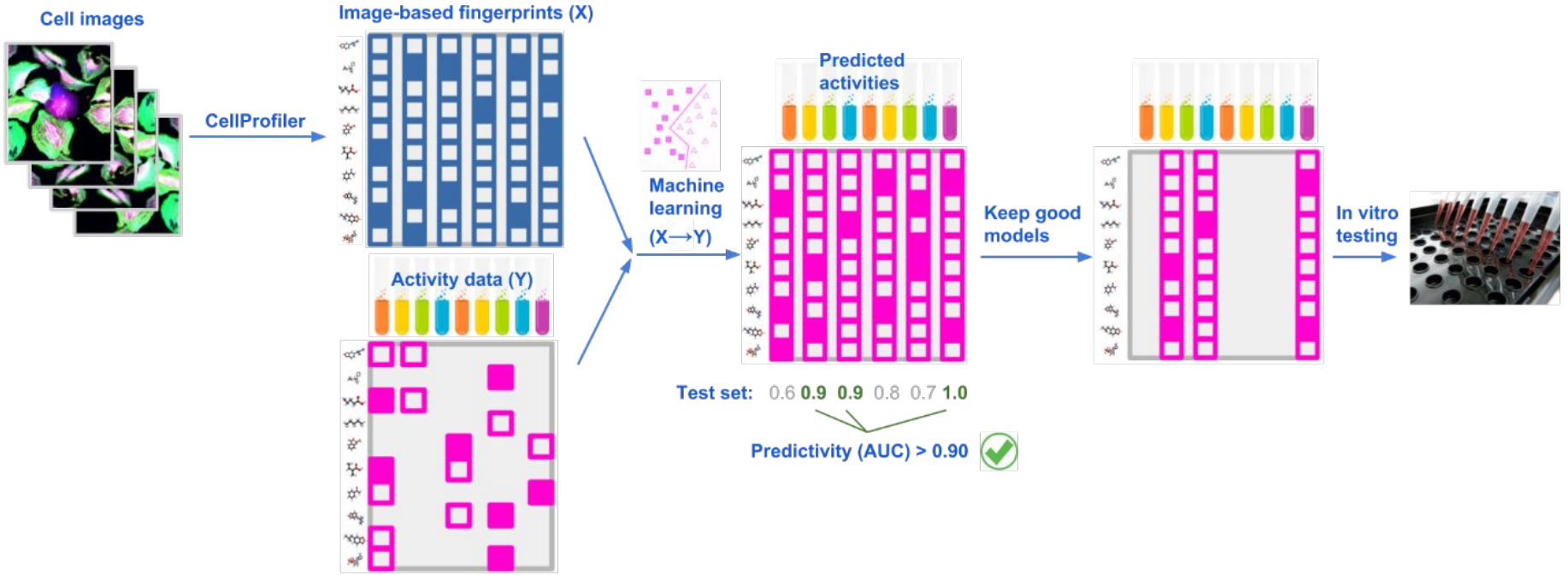
Jaak Simm [8] • Günter Klambauer [8] • Adam Arany [8] • Marvin Steijaert • Jörg Kurt Wegner • Emmanuel Gustin • Vladimir Chupakhin • Yolanda T. Chong • Jorge Vialard • Peter Buijnsters • Ingrid Velter • Alexander Vapirev • Shantanu Singh • Anne E. Carpenter • Roel Wuyts • Sepp Hochreiter [9] • Yves Moreau [9] • Hugo Ceulemans [9, 10] • Show less • Show footnotes

Published: March 01, 2018 • DOI: https://doi.org/10.1016/j.chembiol.2018.01.015 •

Classical high-content imaging



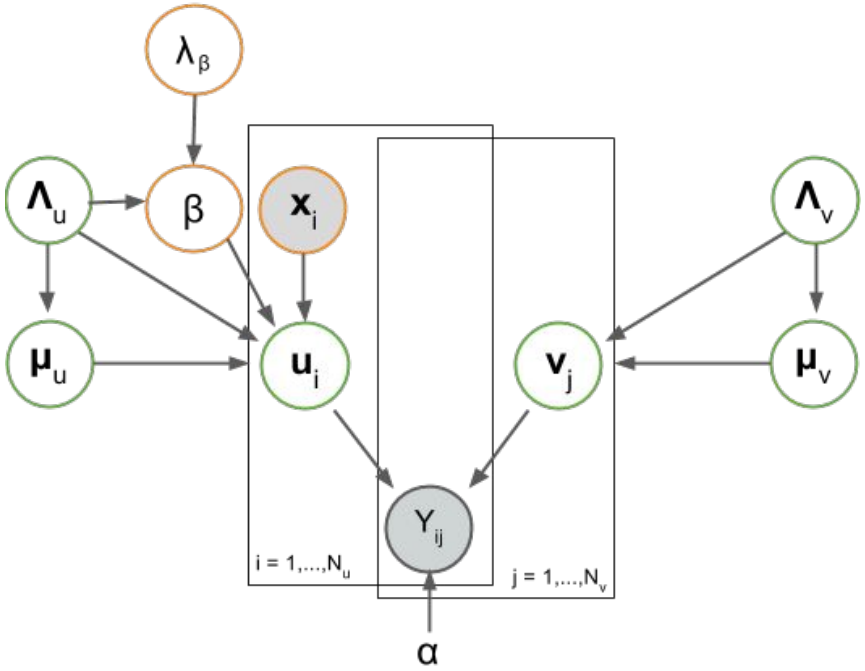
Predicting unrelated protein assays



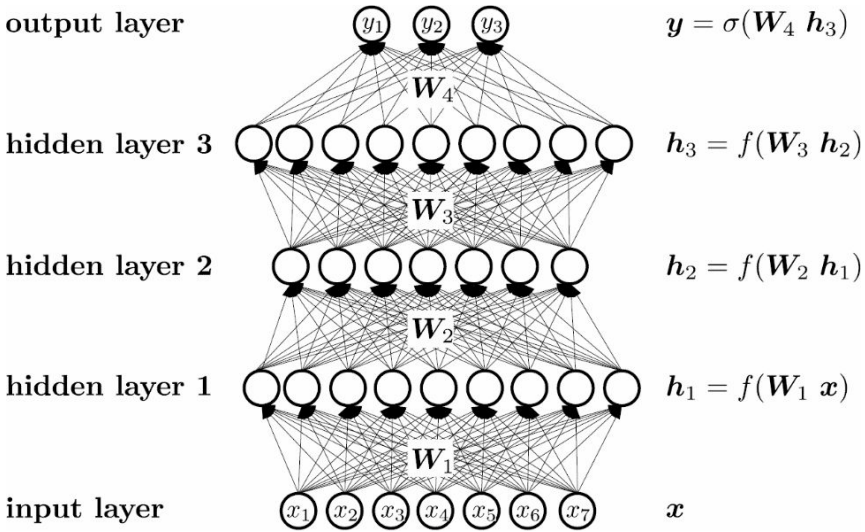
Assay preparation

- Cell-line: **H4** - *Homo sapiens* brain neuroglioma
- Assay was designed for detecting Glucocorticoid receptor nuclear translocation and **repurposed**
- Compounds applied @ **10μM** and incubated **1 hour**
- 1μM hydrocortisone for 1 hour
- Fixed, **stained**, imaged
  - Hoechst, CellMask deep red, indirect immunofluorescence on GCR
- 842 dimensional **feature extracted** via CellProfiler pipeline

Machine Learning Approaches



Matrix Factorization
(MACAU)



Deep Learning

Experimental Results

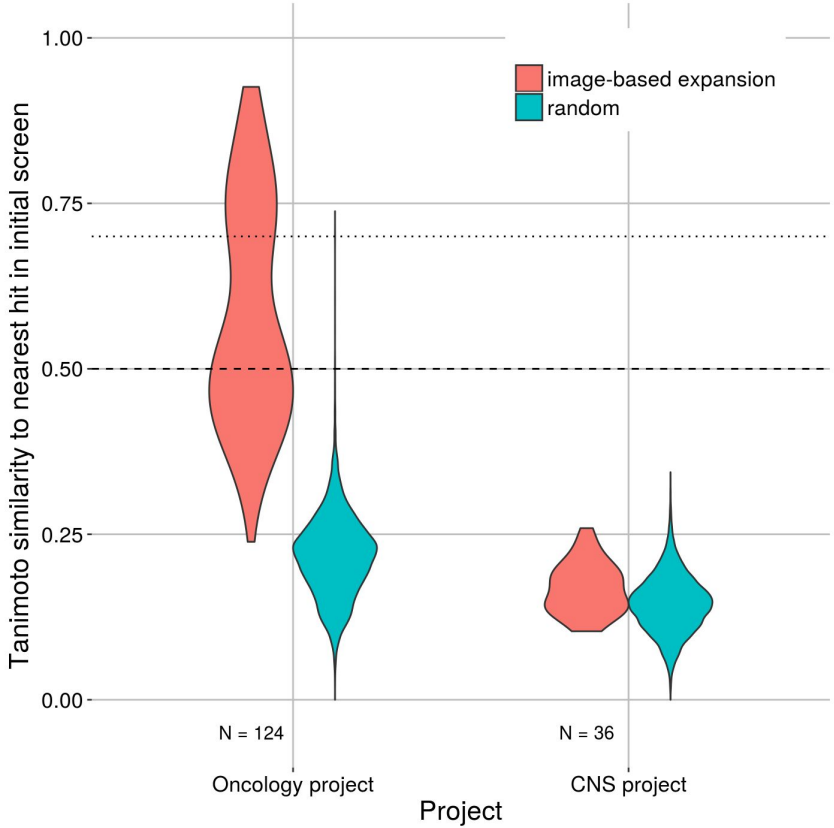
- Imaging assay for glucocorticoid receptor (GR)
  - **500k** compounds
  - Repurposed for **535** protein targets
  - Using **15M IC_{50}** measurements
- Estimated AUC-ROC for targets with at least **25 actives**, **25 inactives**

AUC-ROC threshold	Macau	DNN	Common
0.9	31 (5.8%)	43 (8.0%)	26 (4.9%)
0.7	218 (40.7%)	245 (45.8%)	209 (39.1%)

- **6-8%** of protein assays are predictable (AUC-ROC > 0.9)

In vitro validation

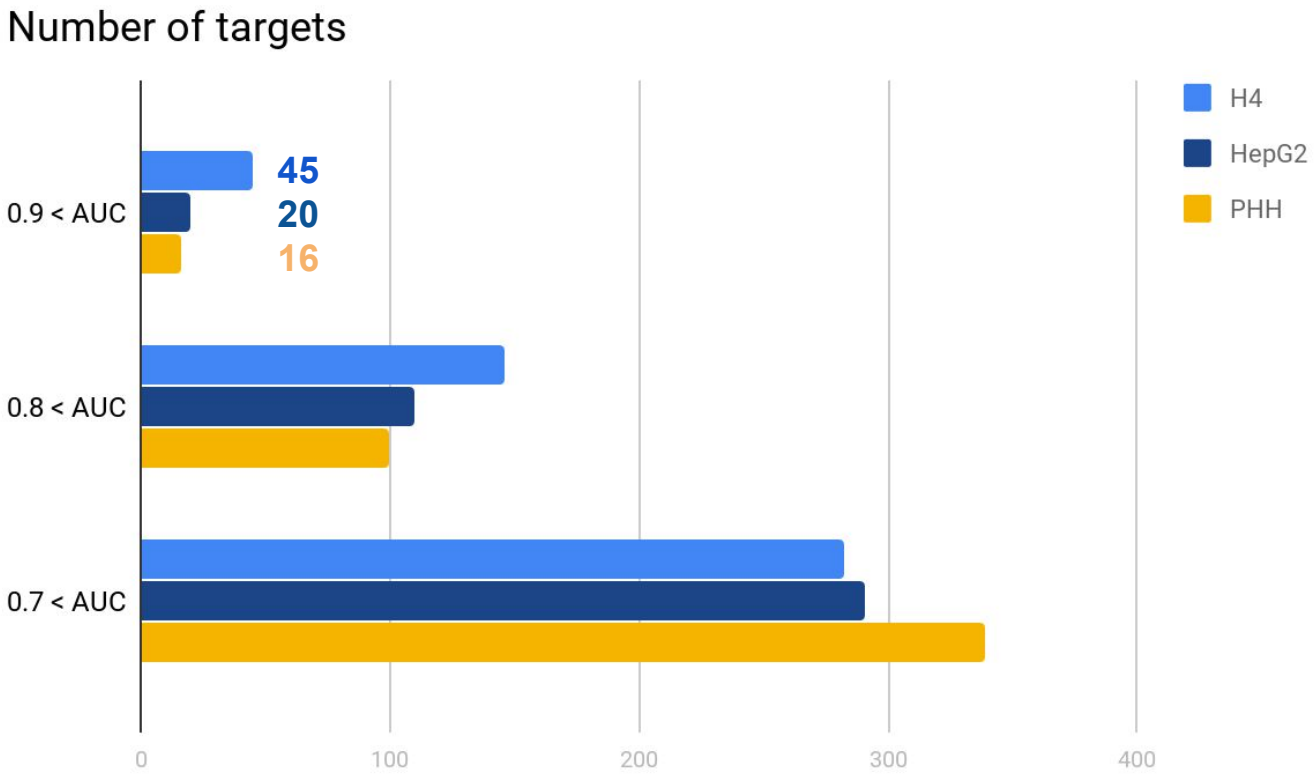
- Two targets (with AUC-ROC > 0.9) were followed up.
- Oncology
  - kinase (unrelated to GR)
  - **36.3% hit rate** (124 submicromolar hits)
  - **50-fold** enrichment
- CNS
  - non-kinase enzyme
  - We used diversity maximization
  - **25.5% hit rate** (36 submicromolar hits)
  - **280-fold** enrichment



Combining Multiple cell lines

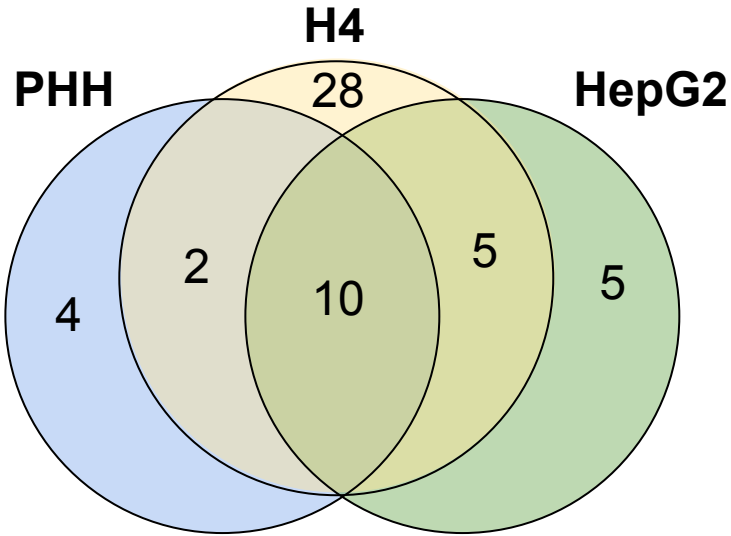
Cell line	Description	Incubation time	Compounds tested
H4	Human brain neuroglioma	1h	500k
HepG2	Human hepatocellular carcinoma	72h	500k
PHH	Primary human hepatocytes	168h	150k

Combining Multiple cell lines

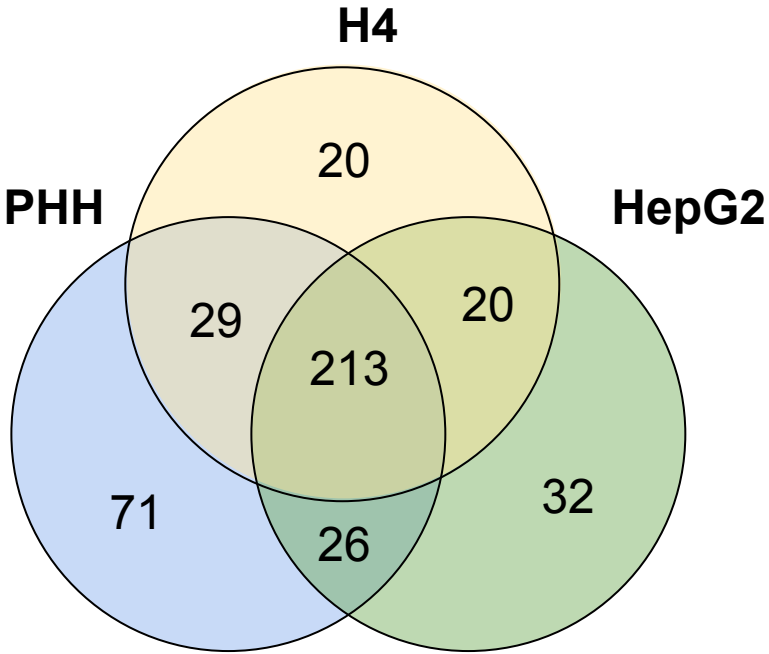


Number of well predicted targets

AUC-ROC > 0.9

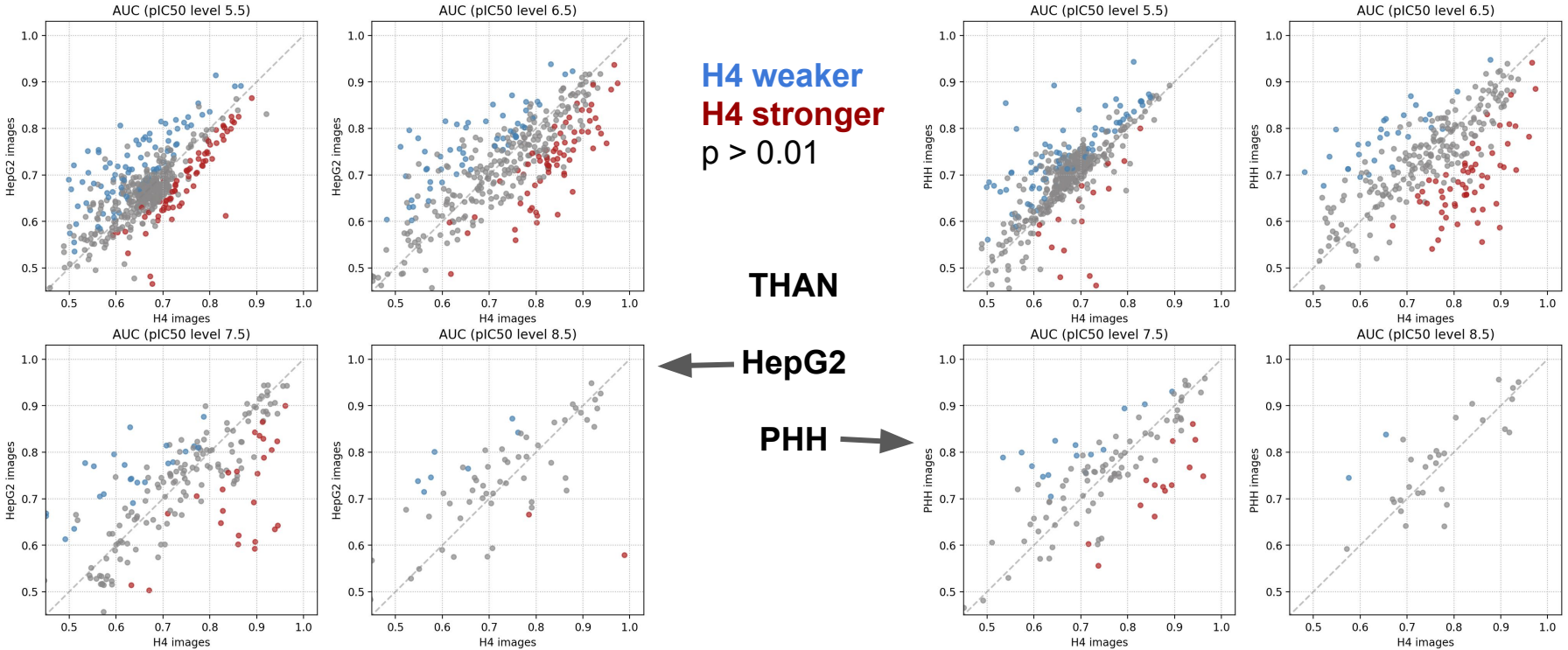


AUC-ROC > 0.7

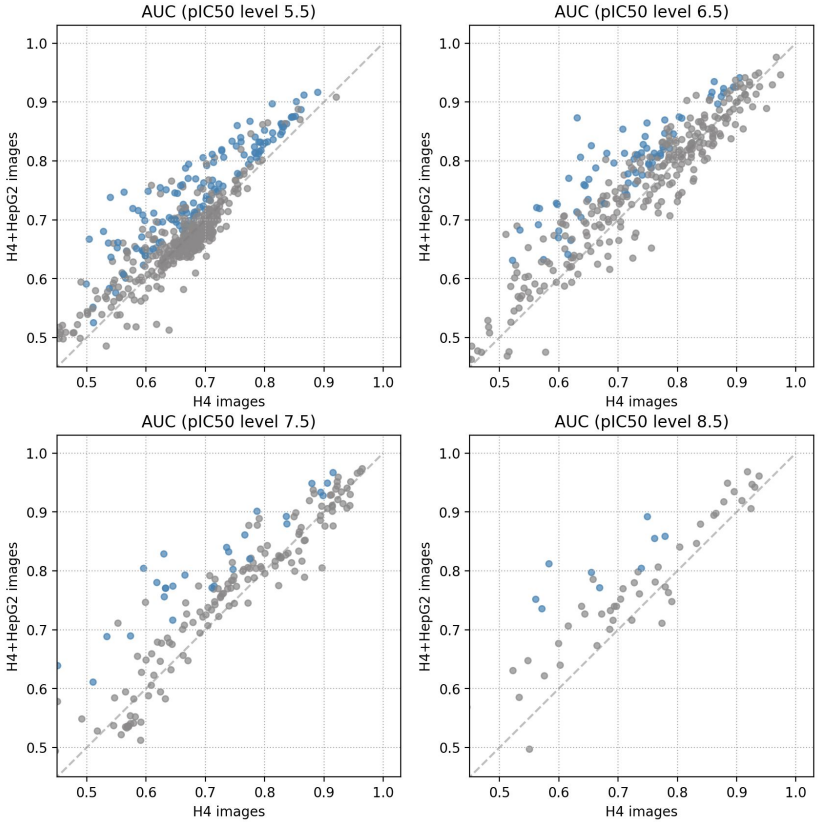


Note that the PHH dataset approximately 4 times smaller.

Significant differences in recognized assays



Data fusion results



H4 weaker - fusion performance is better

H4 stronger - fusion performance is worse

$p > 0.01$

- Fusion is implemented by concatenating features
- 400k common compounds
- There are no assays with significant loss
- Average AUC improves from 0.707 to 0.732

Discussion and future work

- Established a proof of concept that image-based feature can be used for **unrelated targets**.
- Advantages
  - Scaffold hopping: does not depend on chemical similarity
  - Generalizes to other treatments (antibodies, RNAi, etc)
- Future work
  - Using assays for general use (cell painting assay)
  - Machine learning side
    - **Single cell** based pipelines
    - **convolutional neural networks**
- There is a need for public data!

SCIENTIFIC REPORTS

Article | OPEN | Published: 29 May 2018

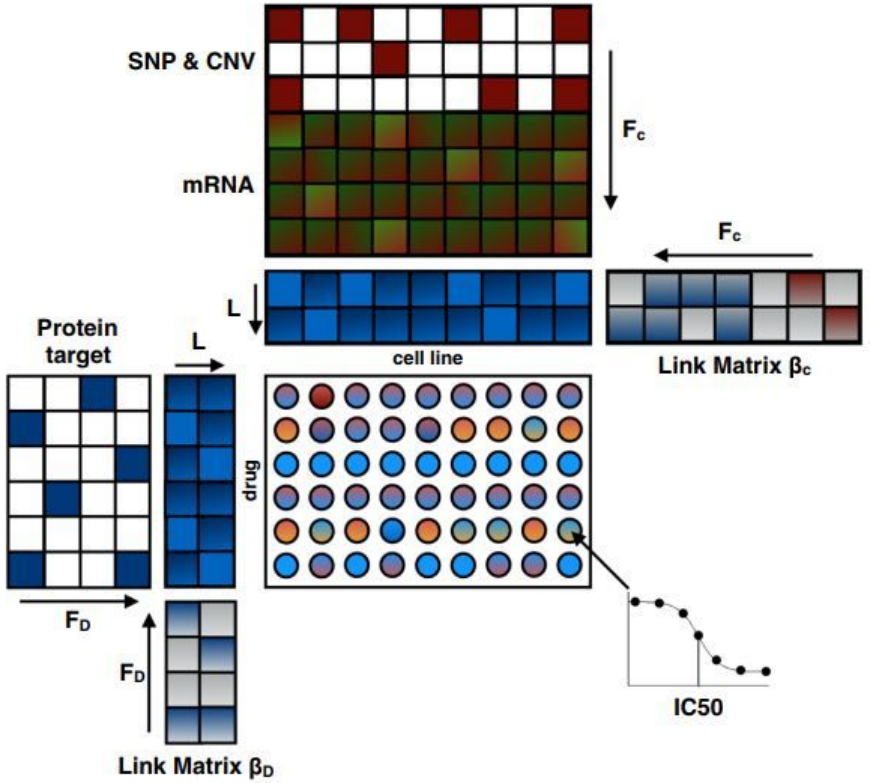
Linking drug target and pathway activation for effective therapy using multi-task learning

Mi Yang, Jaak Simm, Chi Chung Lam, Pooya Zakeri, Gerard J. P. van Westen, Yves Moreau & Julio Saez-Rodriguez ✉

Scientific Reports **8**, Article number: 8322 (2018) | Download Citation ⬇

Computational setup

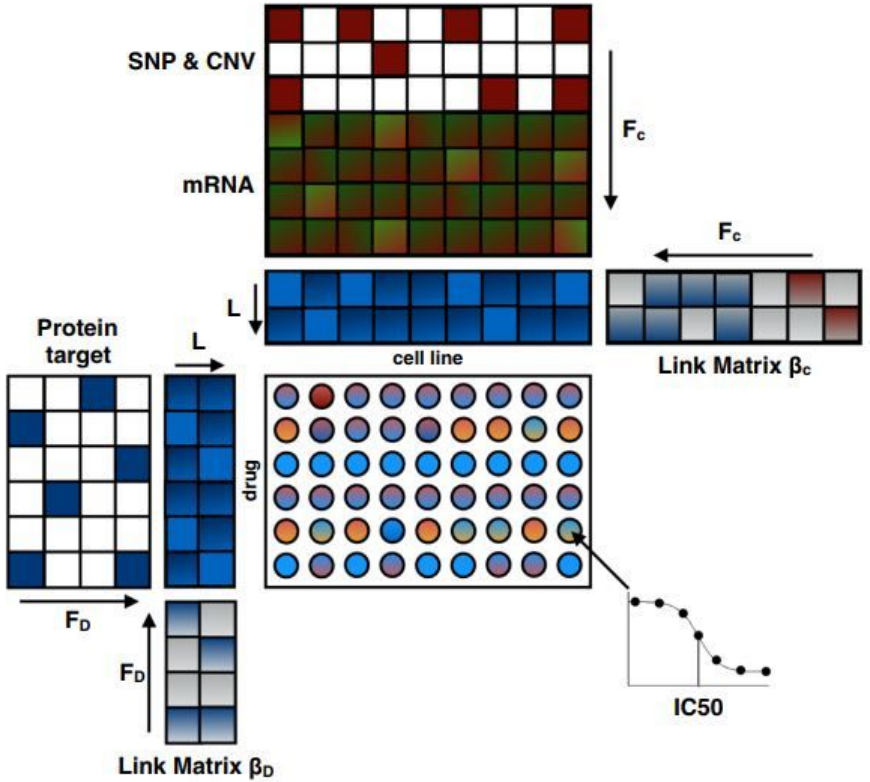
- Model: Bayesian version
- Side information
  - **Drug**: known targets
  - **Cell line**: transcriptomics
  - Pathway level aggregation
- Data
  - 265 drugs
  - 990 cell lines
  - 228 targets
  - 11 pathways



Possible questions

What will be the effect

- **Q1**: given a well characterized drug, and a new cell line knowing its expression profile
- **Q2**: given a known cell line, and a new drug given its targets
- **Q3**: if both are known, but not tested in this combination
- **Q4**: if both are new

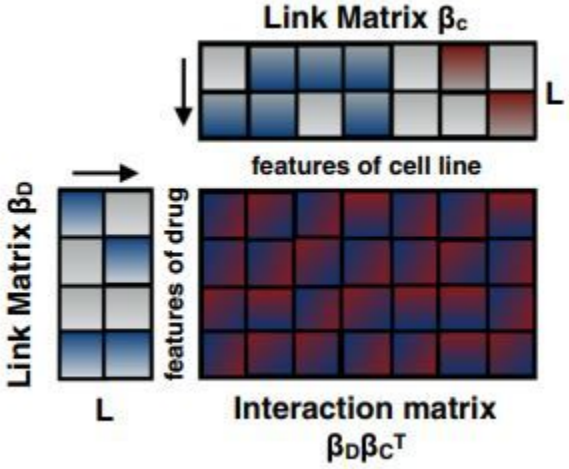


Feature interaction analysis

Only taking into acount the predictable part of the latents:

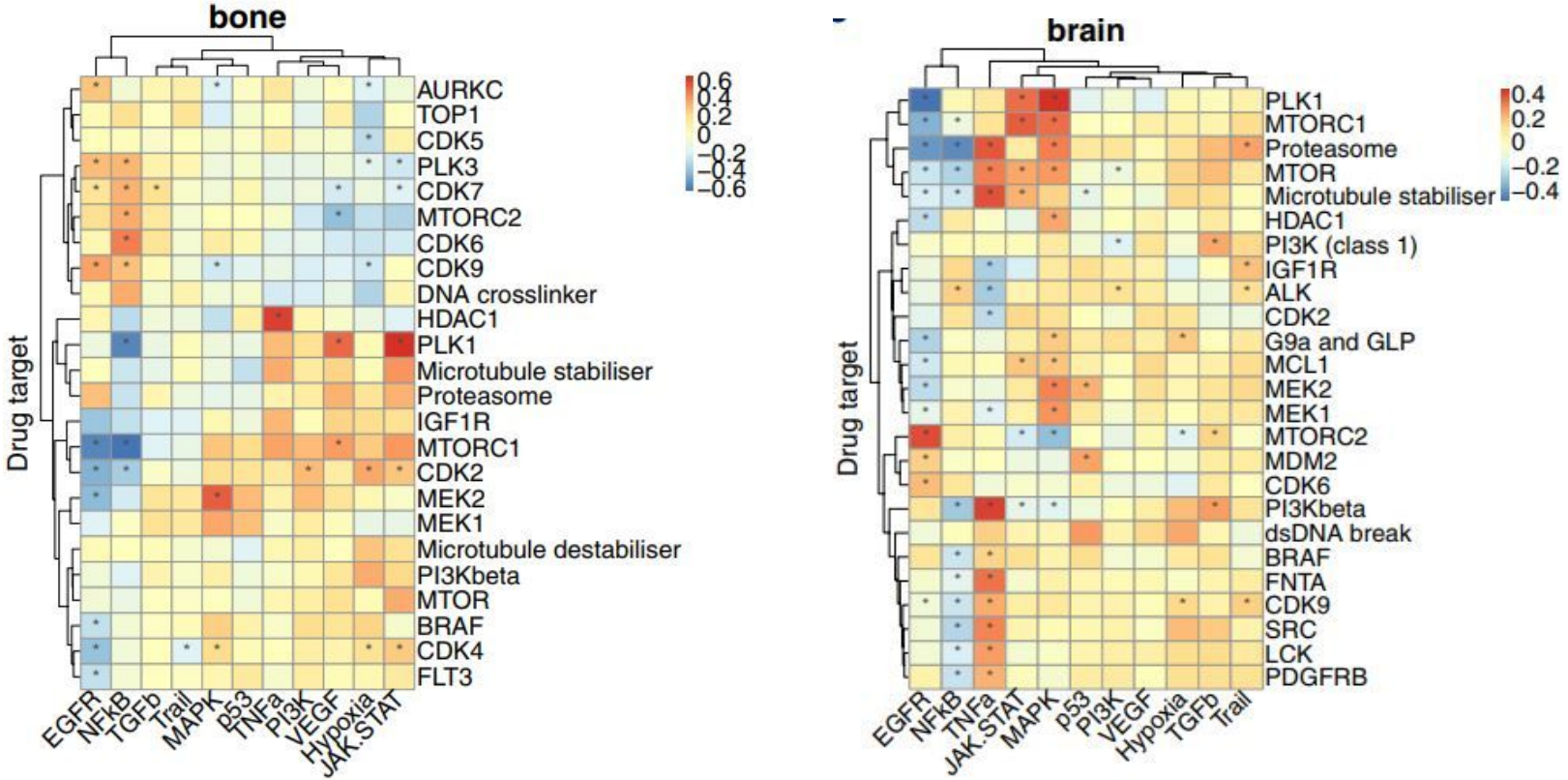
$$IC50 = (\beta_D{}^T x)^T \beta_C{}^T z$$

$$IC50 = x^T \underbrace{\beta_D \beta_C{}^T}_{\text{Interaction matrix}} z$$



"Upregulating of gene Y correlates with drug sensitivity when targeting protein X"

Feature interaction analysis



Packages

License: MIT License

- **Macau**
  - https://github.com/jaak-s/macau
- **SparseFlow**
  - https://gitlab.com/biolearning/sparseflow

Thank you fo your attention!

Jaak Simm
Yves Moreau
Pooya Zakeri

Günter Klambauer
Sepp Hochreiter

Anne E. Carpenter
Shantanu Singh

Hugo Ceulemans
Jörg Kurt Wegner
Marvin Steijaert
Alexander Vapirev
Emmanuel Gustin
Vladimir Chupakhin
Yolanda T. Chong
Jorge Vialard
Peter Buijnsters
Ingrid Velter
Roel Wuyts

Mi Yang
Chi Chung Lam
Gerard J. P. van Westen
Julio Saez-Rodriguez