



Machine Learning for Precision Medicine

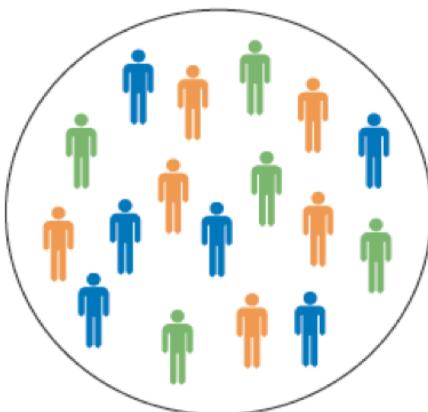
Martin Ester
Simon Fraser University

Precision Medicine

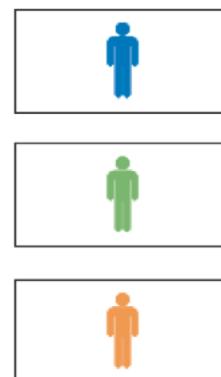
Empirical medicine

Stratified medicine

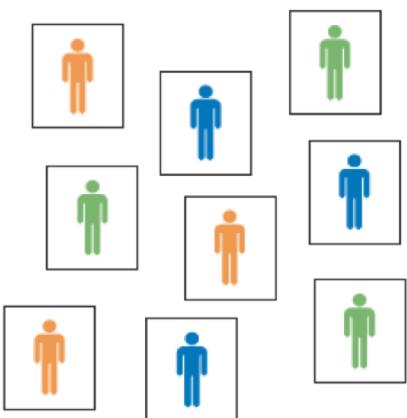
Personalized medicine



- One treatment for all
- Evidence based



- Different treatments for each group
- Evidence based
- Biomarker led

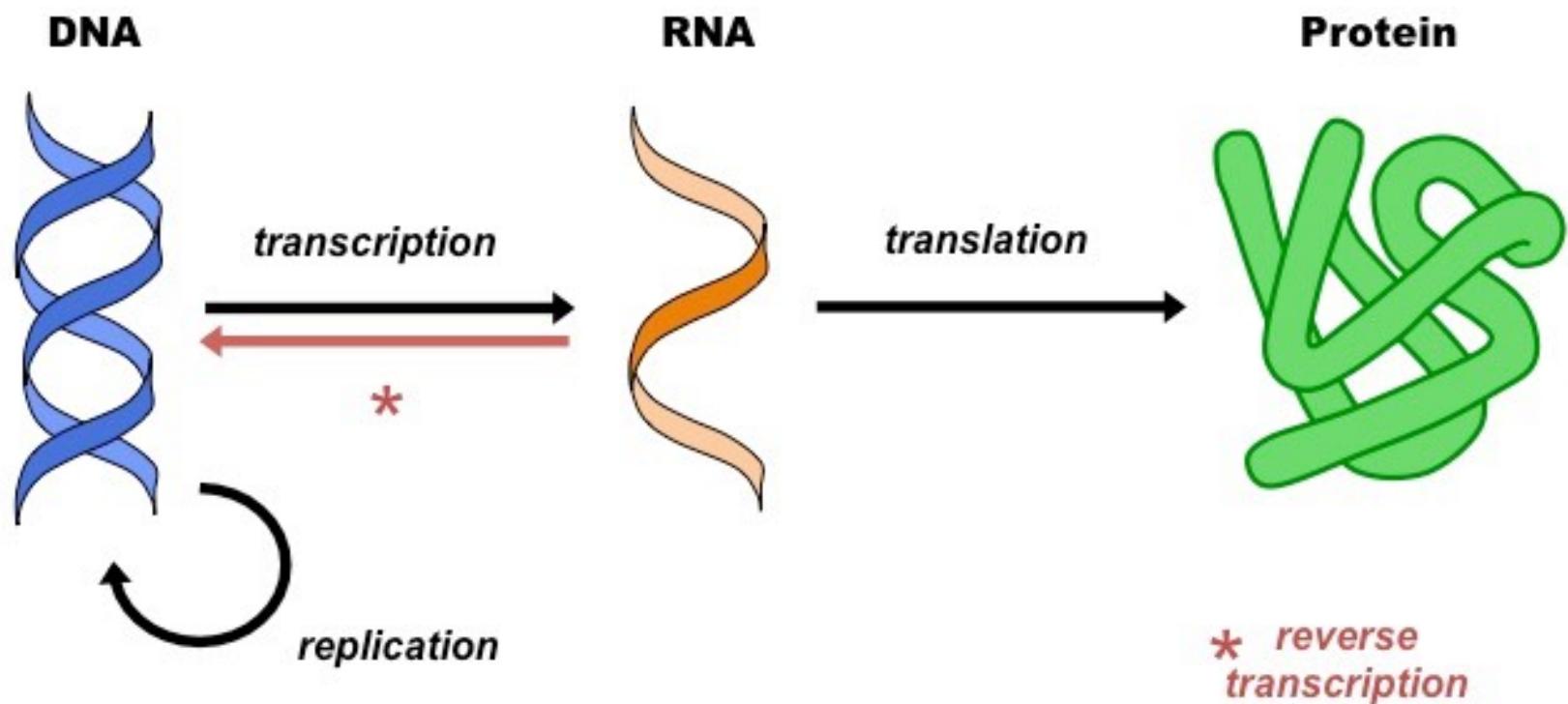


- Individual treatments for each patient
- Evidence based
- Patient derived

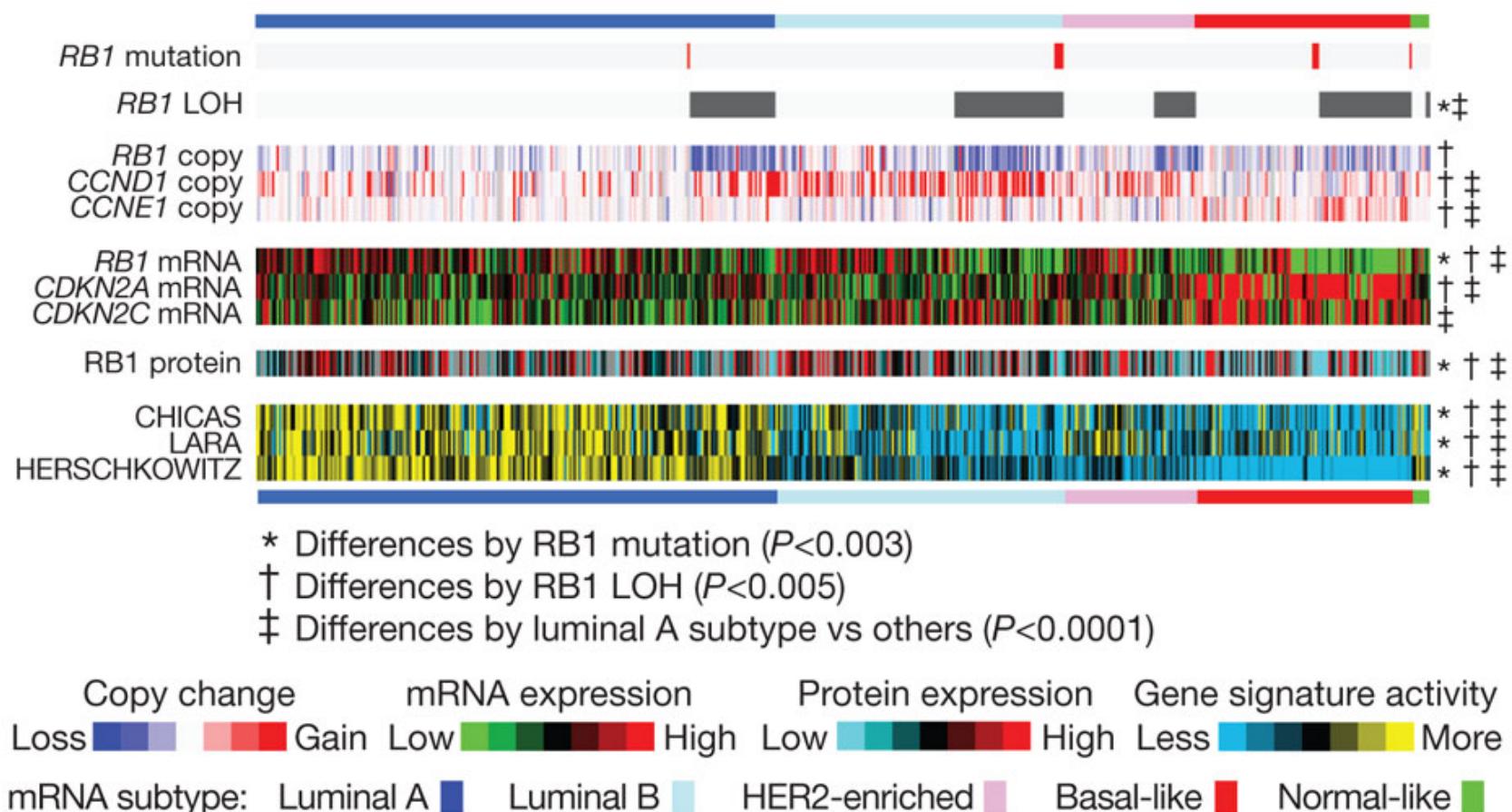
Precision Oncology

- Precision medicine
Diagnose and treat patients taking into account their individual genomic, environmental, and lifestyle factors.
- For precision oncology
in particular the genomic factors.
- Genomic factors
SNVs, CNAs, methylation, gene expression, protein expression . . .
→ multi-omics data

Multi-Omics Data



Multi-Omics Data



Precision Oncology Tasks

- Patient stratification

What groups of patients have similar genomic profiles and can be diagnosed/treated similarly?

- Diagnosis

What disease does the patient have?

- Prognosis

How is the patient's health going to be in the future?

E.g. how is a patient going to respond to a particular drug?

- Treatment recommendation

What is the best treatment for the patient?

Precision Medicine in Other Fields

- Precision mental health

Causal lifestyle factors of mental/cognitive health
Diagnosis/prognosis for a patient

- Precision agriculture

Discovery of new anti-fungal compounds

Compound-fungus interaction prediction

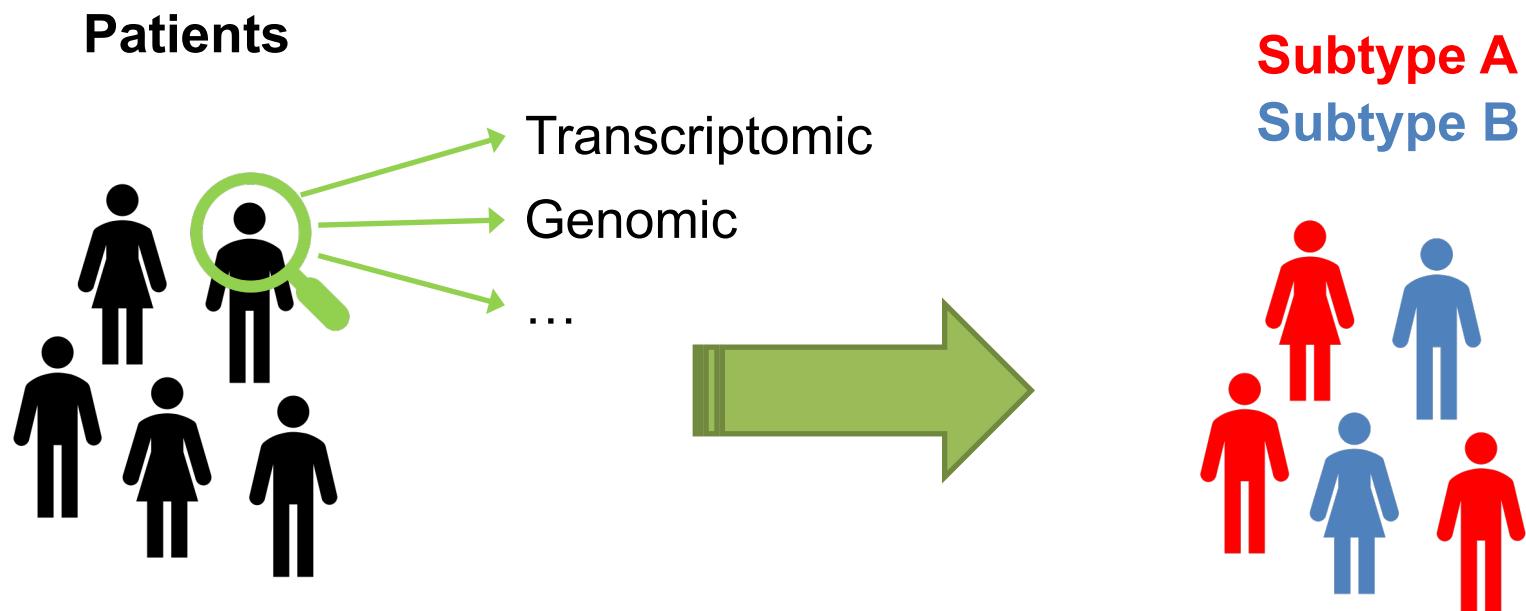
Active learning to systematically explore the space
of all possible combinations

Transfer learning from lab to field

Machine Learning Challenges

- Genomic data is noisy.
 - Probabilistic methods, Probabilistic Graphical Models
- Datasets with clinical information are very small.
 - Transfer learning
- Data has far more columns than rows.
 - Representation learning, Deep Neural Networks
- Results need to be explainable.
- Need for causal patterns.
- Predictions have to be validated in the wet lab or clinic.
 - close collaboration with life scientists

Patient Stratification



Patient Stratification

Existing Methods

Reference	Probabilistic	Technique	Input	Non-param.
Verhaak et al. [135]	No (HC)	Clustering	Expression	No
Hochreiter et al. [56]	Yes (FA)	Biclustering	Expression	No
Shen et al. [114, 115]	Yes (FA)	Clustering	Multiple	No
Zhang et al. [144]	No (NMF)	Biclustering	Multiple	No
Hofree et al. [57]	No (NMF)	Biclustering	Mutation	No
Cho and Przytycka [28]	Yes (PGM)	Clustering	Multiple	No
Sun et al. [121]	No (SVD)	Biclustering	Multiple	No
Raykov et al. [104]	Yes (PGM)	Clustering	Clinical	Yes
Liu et al. [79]	No (CC)	Clustering	Multiple	No

HC: Hierarchical Clustering

FA: Factor Analysis

NMF: Non-negative Matrix Factorization

SVD: Singular Value Decomposition

B2PS: Bayesian Biclustering for Patient Stratification [PSB 2016]

Expression

Expression Genes					
Patients	0	1	...	-1	1
	-1	-1	...	1	0
	0	1	...	1	1
	0	1	...	1	1
	0	1	...	1	1

SNV Genes

Patients	0	1	...	0	0
	0	0	...	1	0
	0	0	...	1	0
	0	0	...	0	1
	0	0	...	0	1

CNV Genes

Patients	2	0	...	1	0
	-1	0	...	2	1
	-1	0	...	-2	0
	-2	0	...	2	1
	-2	0	...	2	1

Single Nucleotide Variation

Copy Number Variation



Expression Gene Clusters

Expression Gene Clusters	$\theta^e \in [0,1]^{K^p \times K^e \times 2}$
Patient Clusters	$\theta^e \in [0,1]^{K^p \times K^e \times 2}$

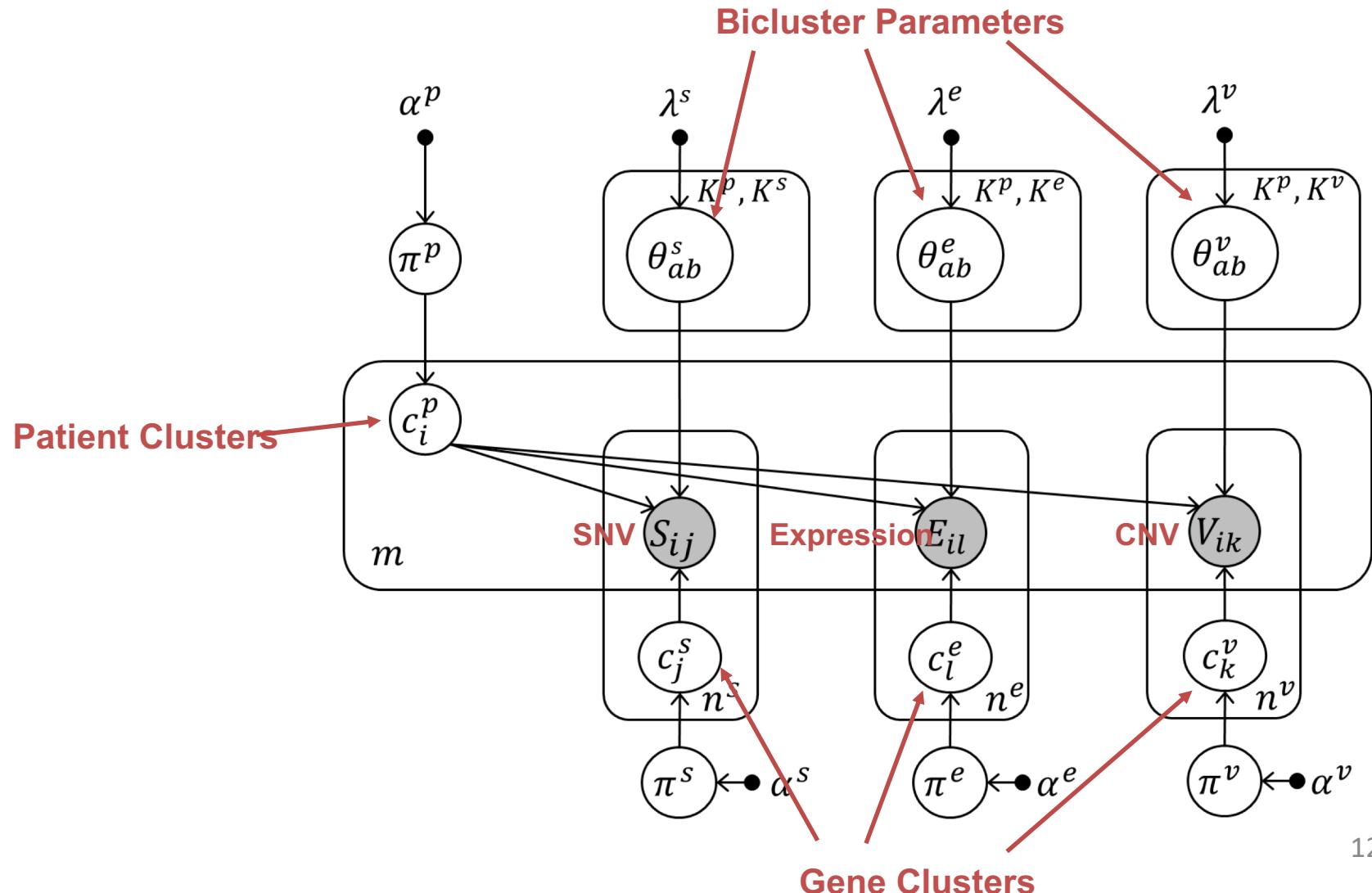
SVN Gene Clusters

SVN Gene Clusters	$\theta^s \in [0,1]^{K^p \times K^s \times 1}$
Patient Clusters	$\theta^s \in [0,1]^{K^p \times K^s \times 1}$

CNV Gene Clusters

CNV Gene Clusters	$\theta^v \in [0,1]^{K^p \times K^v \times 4}$
Patient Clusters	$\theta^v \in [0,1]^{K^p \times K^v \times 4}$

B2PS: Probabilistic Graphical Model



B2PS: Parameter Learning

- Gibbs Sampling

Conditional Probability of c_i^p based on expected parameter values

$$P(c_i^p = q | c_{-i}^p, c^s, c^e, c^v, S, E, V, H) \propto$$

Patient Cluster

$$(|\{l : c_l^p = q, l \neq i\}| + \alpha^p) \times$$

SNV

$$\prod_{t=1}^{K^s} \frac{\prod_{z \in \{0,1\}} (\bar{s}_q^t(z) + \beta_z^s - \mathbf{1}(c_i^p = q) \cdot \hat{s}_i^t(z)) \hat{s}_i^t(z)}{\left(\sum_{z \in \{0,1\}} (\bar{s}_q^t(z) + \beta_z^s - \mathbf{1}(c_i^p = q) \cdot \hat{s}_i^t(z)) \right)^{\sum_{z \in \{0,1\}} \hat{s}_i^t(z)}} \times$$

Expression

$$\prod_{t=1}^{K^e} \frac{\prod_{z \in \{-1,0,1\}} (\bar{e}_q^t(z) + \beta_z^e - \mathbf{1}(c_i^p = q) \cdot \hat{e}_i^t(z)) \hat{e}_i^t(z)}{\left(\sum_{z \in \{-1,0,1\}} (\bar{e}_q^t(z) + \beta_z^e - \mathbf{1}(c_i^p = q) \cdot \hat{e}_i^t(z)) \right)^{\sum_{z \in \{-1,0,1\}} \hat{e}_i^t(z)}} \times$$

CNV

$$\prod_{t=1}^{K^v} \frac{\prod_{z \in \{-2,-1,0,1,2\}} (\bar{v}_q^t(z) + \beta_z^v - \mathbf{1}(c_i^p = q) \cdot \hat{v}_i^t(z)) \hat{v}_i^t(z)}{\left(\sum_{z \in \{-2,-1,0,1,2\}} (\bar{v}_q^t(z) + \beta_z^v - \mathbf{1}(c_i^p = q) \cdot \hat{v}_i^t(z)) \right)^{\sum_{z \in \{-2,-1,0,1,2\}} \hat{v}_i^t(z)}} \times$$

B2PS: Experimental Evaluation

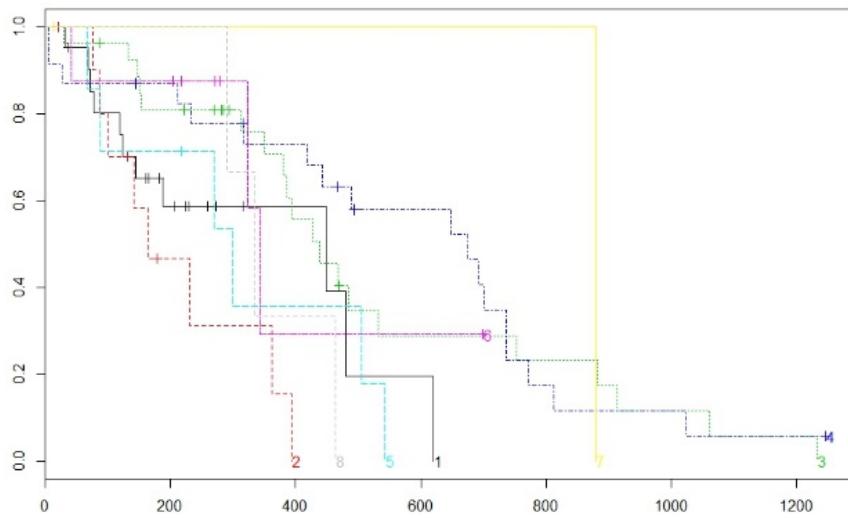
Datasets

Diseases	Samples	Features		
		Point Mutation	CNV	Expression
GBM	102	4117	23082	11874
BRCA	501	13776	23082	17814

B2PS: Experimental Evaluation

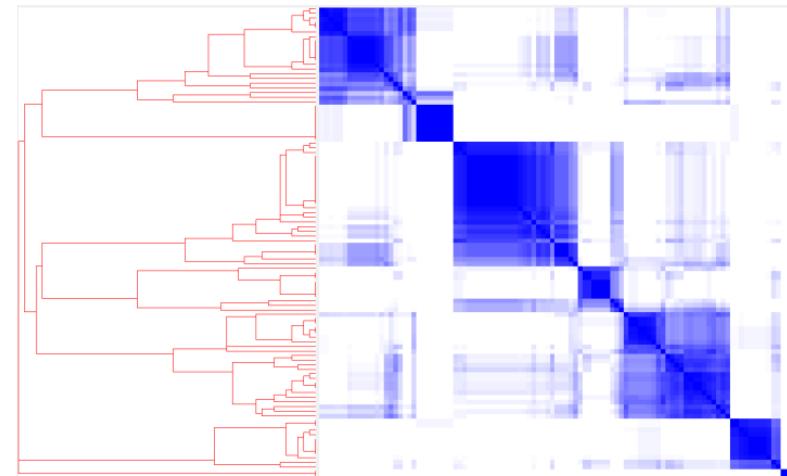
- Challenge: no ground truth patient strata
- Use the following metrics to evaluate performance:

Survival Analysis (log-rank test)



Robustness

(Cophenetic Correlation Coefficient)



B2PS: Experimental Evaluation

Impact of different omics datatypes

Dataset	Data Types	Sample Clusters	Feature Clusters		Log-rank <i>p-value</i>	Cophenetic Corr. Coef.	GOTO	
			Exp.	CNV			Exp.	CNV
GBM	Exp.	8	25	NA	4.0e-3	0.96	3.41	NA
GBM	CNV	19	NA	86	4.1e-1	0.98	NA	1.82
GBM	Both	7	22	68	2.9e-1	0.80	3.40	1.80
BRCA	Exp.	8	69	NA	1.4e-1	0.94	2.60	NA
BRCA	CNV	20	NA	63	3.5e-1	0.91	NA	1.85
BRCA	Both	11	69	68	5.4e-1	0.90	2.58	1.86

B2PS: Experimental Evaluation

Impact of different omics datatypes

Dataset	Data Types	Sample Clusters	Feature Clusters		Log-rank <i>p-value</i>	Cophenetic Corr. Coef.	GOTO	
			Exp.	CNV			Exp.	CNV
GBM	Exp.	8	25	NA	4.0e-3	0.96	3.41	NA
GBM	CNV	19	NA	86	4.1e-1	0.98	NA	1.82
GBM	Both	7	22	68	2.9e-1	0.80	3.40	1.80
BRCA	Exp.	8	69	NA	1.4e-1	0.94	2.60	NA
BRCA	CNV	20	NA	63	3.5e-1	0.91	NA	1.85
BRCA	Both	11	69	68	5.4e-1	0.90	2.58	1.86

B2PS: Experimental Evaluation

Impact of different omics datatypes

Dataset	Data Types	Sample Clusters	Feature Clusters		Log-rank <i>p-value</i>	Cophenetic Corr. Coef.	GOTO	
			Exp.	CNV			Exp.	CNV
GBM	Exp.	8	25	NA	4.0e-3	0.96	3.41	NA
GBM	CNV	19	NA	86	4.1e-1	0.98	NA	1.82
GBM	Both	7	22	68	2.9e-1	0.80	3.40	1.80
BRCA	Exp.	8	69	NA	1.4e-1	0.94	2.60	NA
BRCA	CNV	20	NA	63	3.5e-1	0.91	NA	1.85
BRCA	Both	11	69	68	5.4e-1	0.90	2.58	1.86

B2PS: Experimental Evaluation

Comparison to NMF

Dataset	Method	Sample Clusters	Feature Clusters	Log-rank <i>p-value</i>	Cophenetic Corr. Coef.	GOTO
GBM	B2PS	8	25	0.004	0.96	3.41
GBM	NMF	3	3	0.460	0.97	2.54
GBM	B2PS	3	29	0.047	0.97	3.41
GBM	B2PS	3	6	0.220	1.00	3.39
BRCA	B2PS	8	69	0.140	0.94	2.60
BRCA	NMF	3	3	0.230	0.99	2.54
BRCA	B2PS	3	101	0.120	1.00	2.60
BRCA	B2PS	3	6	0.490	0.98	2.55

B2PS: Experimental Evaluation

Comparison to NMF

Dataset	Method	Sample Clusters	Feature Clusters	Log-rank <i>p-value</i>	Cophenetic Corr. Coef.	GOTO
GBM	B2PS	8	25	0.004	0.96	3.41
GBM	NMF	3	3	0.460	0.97	2.54
GBM	B2PS	3	29	0.047	0.97	3.41
GBM	B2PS	3	6	0.220	1.00	3.39
BRCA	B2PS	8	69	0.140	0.94	2.60
BRCA	NMF	3	3	0.230	0.99	2.54
BRCA	B2PS	3	101	0.120	1.00	2.60
BRCA	B2PS	3	6	0.490	0.98	2.55

B2PS: Experimental Evaluation

Comparison to NMF

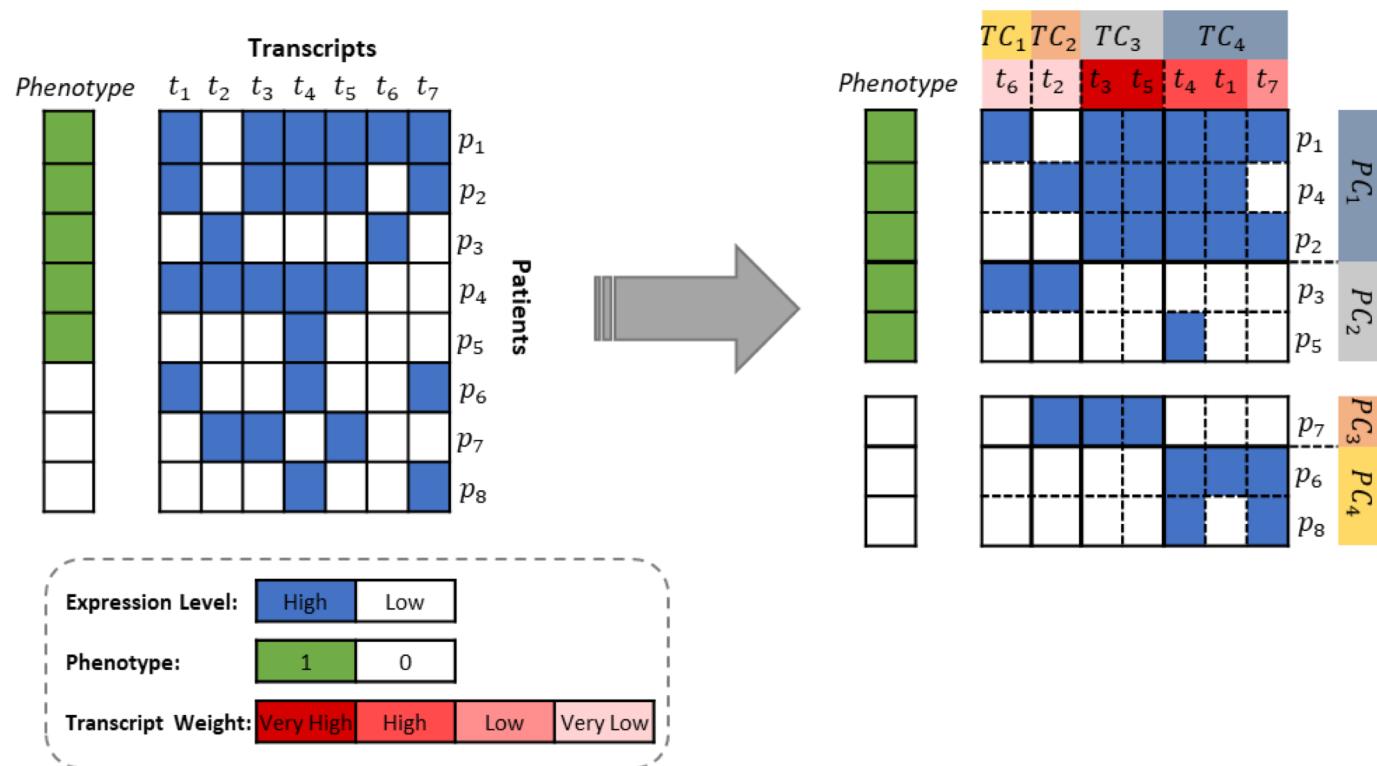
Dataset	Method	Sample Clusters	Feature Clusters	Log-rank <i>p-value</i>	Cophenetic Corr. Coef.	GOTO
GBM	B2PS	8	25	0.004	0.96	3.41
GBM	NMF	3	3	0.460	0.97	2.54
GBM	B2PS	3	29	0.047	0.97	3.41
GBM	B2PS	3	6	0.220	1.00	3.39
BRCA	B2PS	8	69	0.140	0.94	2.60
BRCA	NMF	3	3	0.230	0.99	2.54
BRCA	B2PS	3	101	0.120	1.00	2.60
BRCA	B2PS	3	6	0.490	0.98	2.55

B2PS: Summary

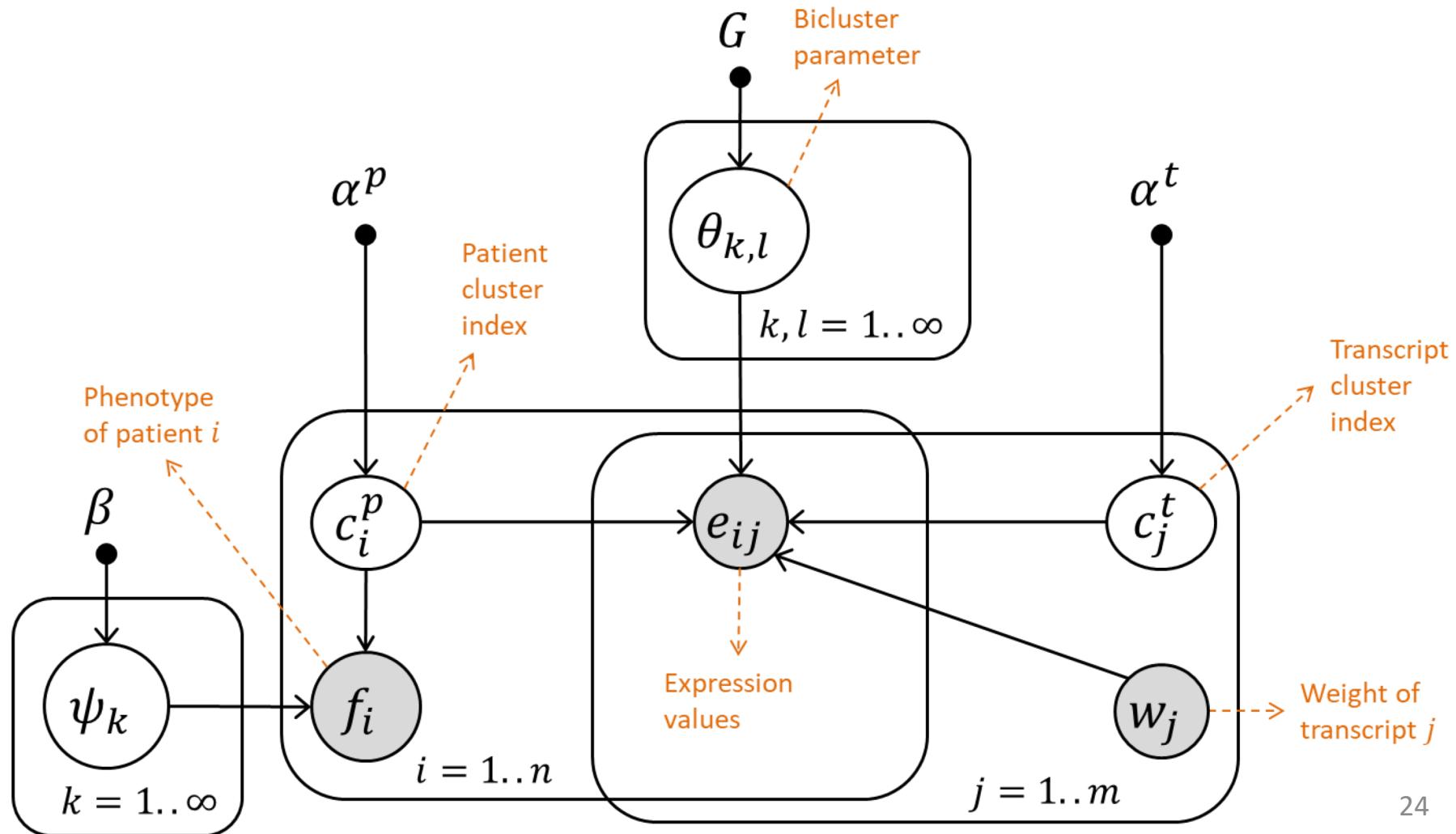
- Novelty
 - Probabilistic + Biclustering + Integrative
- Results
 - Gene expression is the most informative data type for patient stratification.
 - B2PS outperforms NMF both for patient stratification and gene clustering.
 - B2PS detects the number of clusters automatically.

SUBSTRA: Supervised Bayesian Patient Stratification [Bioinformatics 2019]

- Exploit phenotype information
 - Find subtypes (within each phenotype)
 - Find feature groups and feature weights



SUBSTRA: Probabilistic Graphical Model



Discovery of Genetic Causes of ADRs

- Goal: Given a large patient database with genomic and clinical information, detect mutations which cause adverse drug reactions (ADRs).
- Input: patient dataset
 - Output: ranked list of (potentially) causal mutations
- Challenges
 - only observational data,
 - many potentially causes and potential confounders
 - 30,000 genes and 1,000,000 SNVs

Discovery of Genetic Causes

Patient Dataset

#	Patient	Drugs			Reactions			Biomarkers		
		Doxorubicin	Daunorubicin	...	Cardiotoxicity	Mood Swing	...	rs2229774	rs2868177	...
1	CAL600056	1	0	...	1	0	...	A_C	G_G	...
2	CAL600064	0	0	...	0	1	...	A_A	T_G	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
434	WIN450183	1	1		0	0		A_A	T_T	

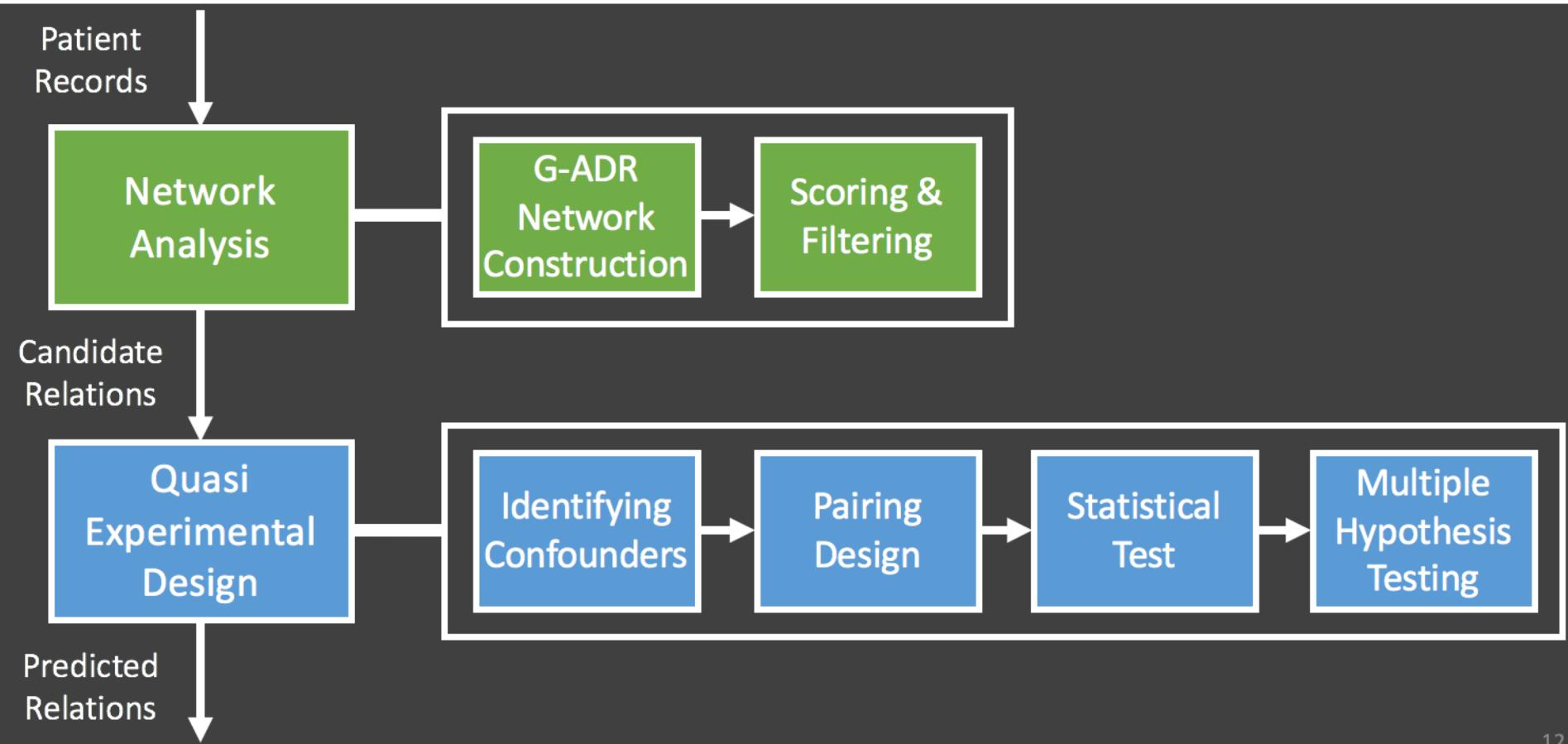
HUME [Bioinformatics 2018]

Approach

- Challenge 1:
many potentially causes and potential confounders
→ network analysis to reduce causes/confounders
- Challenge 2:
only observational data
→ quasi-experimental design
determine a subset of the dataset that corresponds
to a randomized case/control experiment.

HUME

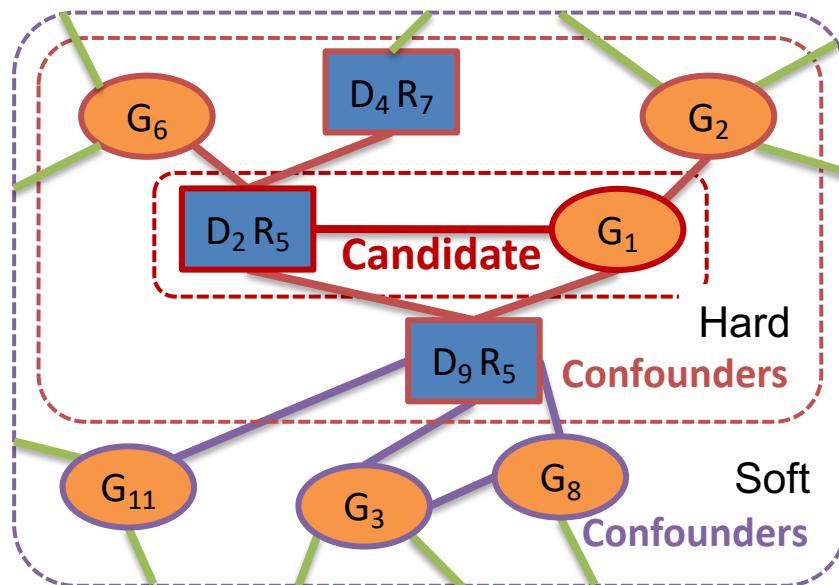
Overview



HUME

Network Analysis

- Construct a network of genes, ADRs, and their relationships
- For potential cause G_1
Consider as potential confounders only genes that are in the network neighborhood of G_1



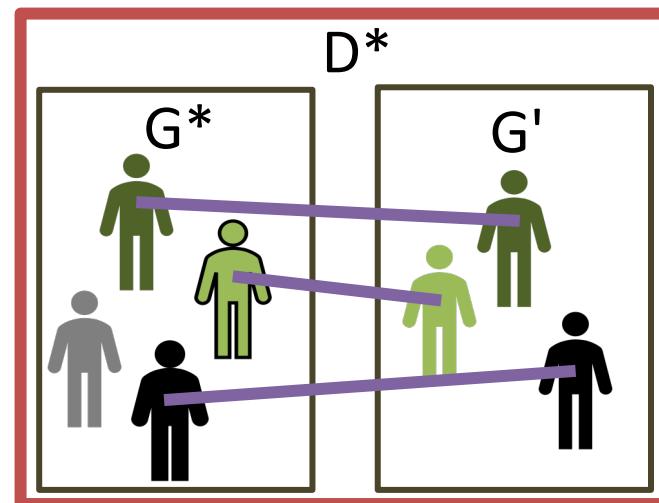
Hard confounders
Nodes at distance of 1 or less from either G or DR

Soft confounders
Nodes in the same connected Component at distance greater than 1

HUME

Quasi-experimental Design

Matched pair design



	drug	confounders	mutation	ADR
Patient 1:	Yes	Yes	No	Yes
Patient 2:	Yes	Yes	No	No

HUME

Quasi-experimental Design

- Task
 - assigning treatment cases to control cases that agree as much as possible on the values of potential confounders
- Constraints
 - matched patients need to agree on hard confounders
- Penalties
 - for matching patients that disagree on soft confounders
- Hungarian Algorithm
 - solves the constrained optimization problem

HUME

Statistical test

- McNemar's test
 - Test on paired nominal data

	Test 2 positive	Test 2 negative	Row total
Test 1 positive	a	b	$a + b$
Test 1 negative	c	d	$c + d$
Column total	$a + c$	$b + d$	n

- Tests whether the two marginal probabilities for each outcome are the same.

HUME

Multiple Hypothesis Testing

- Correct for multiple hypothesis testing
 - Benjamini-Hochberg procedure
 - controls the ratio of false discoveries.

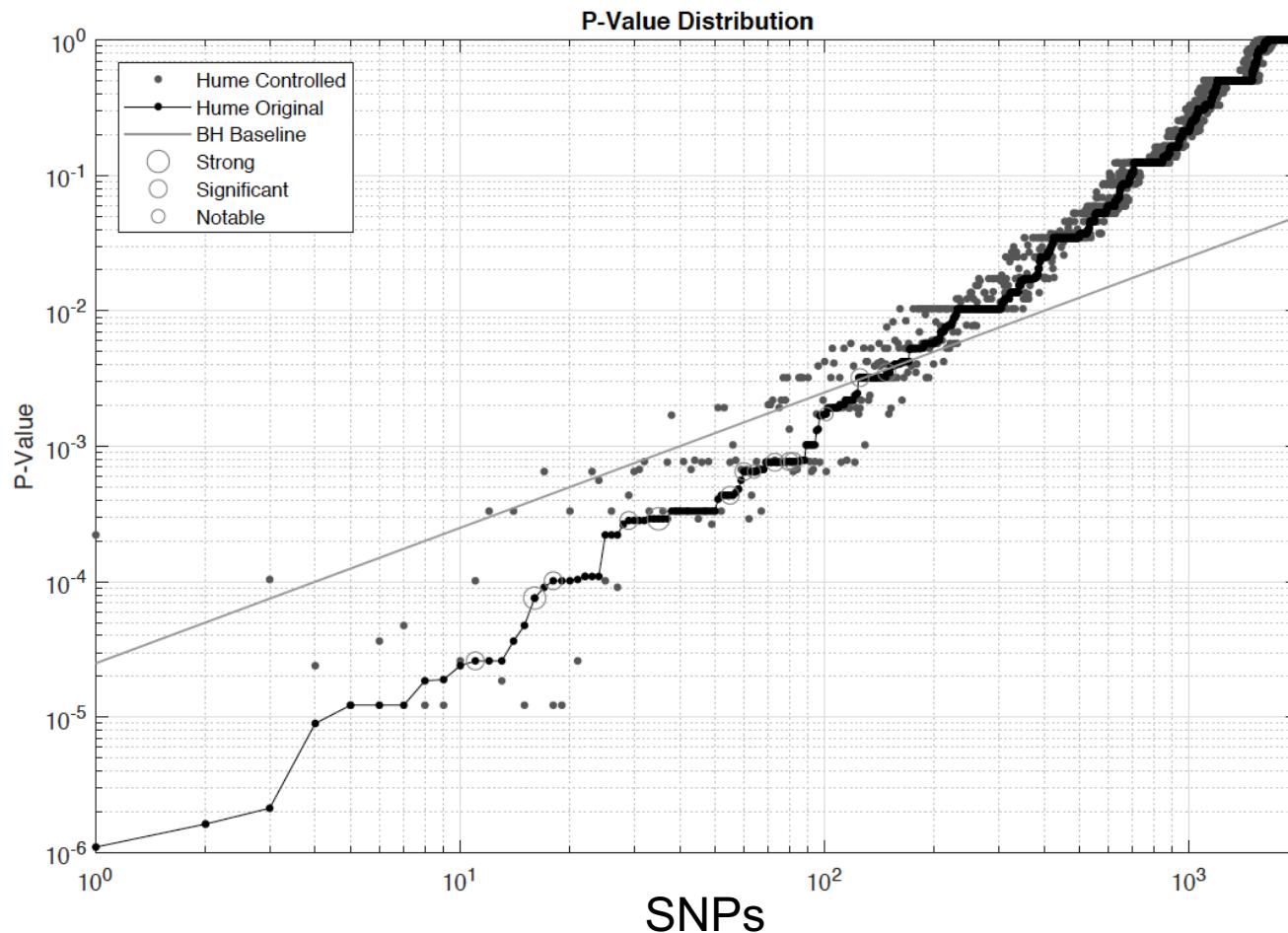
Experiments

Design

- CPNDS dataset
(Canadian Pharmacogenomics Network for Drug Safety)
434 patients, ~1,000,000 SNPs
ADR: anthracycline-induced cardiotoxicity
- Guideline of genetic variants in anthracycline-induced cardiotoxicity
SNPs and genes with various levels of evidence of association
with anthracycline cardiotoxicity in childhood cancer
strong,
significant, and
notable
- How well are the guideline SNPs and genes predicted?

Experiments

Results



Ongoing Work on Causal Discovery

Causal patterns for Subpopulations

- Motivation
 - Same effect can be caused by different mechanisms in different subpopulations.
 - When considering the whole population, the cause may not be significant enough.
- Method: Aristotle
 - Detect subpopulations and relevant sets of features (correlated with the effect in the subpopulation).
 - Use SUBSTRA for this task.
 - Apply quasi-experimental design to subpopulations and their relevant features.

Ongoing Work on Causal Discovery

SNP	CH	POS	REF	ALT	CADD	HM 1	AR 1	HM 2	AR 2
rs2229774	12	53605545	G	A	22.3	✓	✓	✓	✓
rs7853758	9	86900926	G	A	7.917	✓		✓	
rs17863783	2	234602277	G	T	6.729		✓		✓
rs17583889	2	138746039	C	A	6.404	✓	✓	✓	✓
rs13058338	22	37632770	T	A	5.325	✓		✓	
rs1056892	21	37518706	G	A	18.99	✓		✓	
rs7627754	3	183737356	A	T	0.553	✓		✓	
rs4149178	6	43272188	A	G	0.28	✓		✓	
rs1799945	6	26091179	C	G	24.4	✓		✓	
rs10426377	19	48588977	C	A	—	✓	✓		
rs8187694	?	?	?	?	—	✓			
rs1883112	22	36860804	G	A	—	✓			
rs4982753	14	23345360	C	A	—	✓			
rs4673	16	88646828	A	G	—		✓		
rs246221	16	16138322	T	C	0.012				
rs2232228	16	69143577	A	G	9.396				
rs1800562	6	26093141	G	A	25.7				
rs10836235	11	34460704	C	T	9.204	✓	✓	✓	✓
rs2868177	7	75589903	A	G	3.156	✓	✓	✓	✓
rs13240755	7	75606109	G	A	2.17				

	Guideline +++ SNPs
	Guideline ++ SNPs
	Guideline + SNPs
	PharmGKB SNP
✓	Discovered
	Undiscovered
HM	Hume Method
AR	Aristotle Method
1	Original Results
2	New Results

Drug Response Prediction

Introduction

- Problem
 - Given the genomic profile of a patient.
 - Predict the effectiveness of a given drug.
 - Challenge 1
 - Cannot treat patients with many drugs.
 - Clinical trial data are either small or not publicly available.
- Patient training datasets are very small.
- But preclinical datasets (cell lines / PDX) are much larger.

Drug Response Prediction

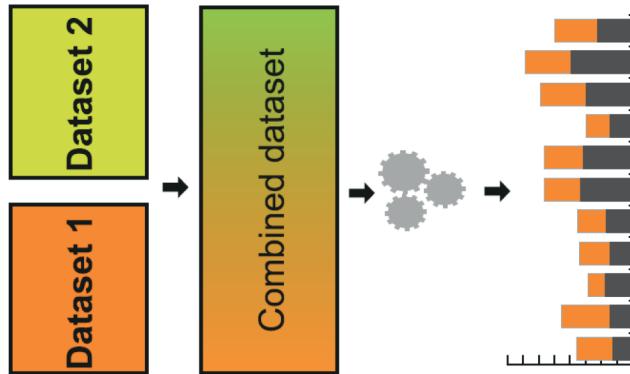
Introduction

- Challenge 2
How to use multi-omics data effectively?
 - Gene expression data
have been shown to be the best data type for drug response prediction.
 - Recent studies suggest that adding other omics data types can improve the prediction performance.
- How to integrate different omics data types?

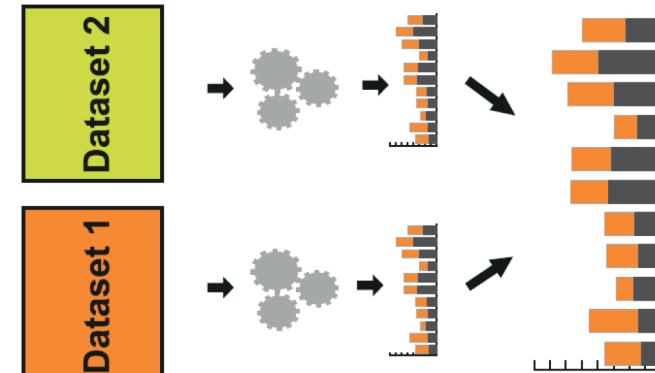
Drug Response Prediction

Multi-Omics Integration

Early integration

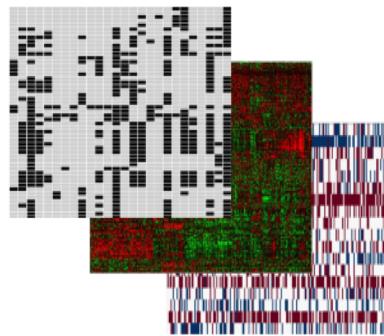


Late integration

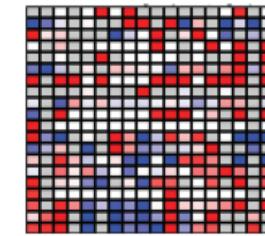
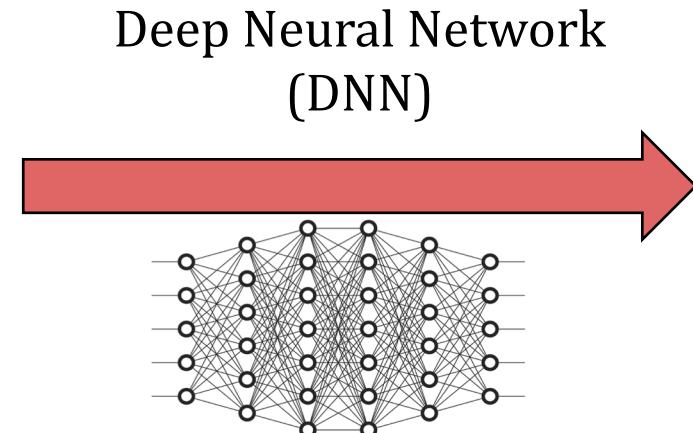


Different datatypes have different characteristics

MOLI: Multi-Omics Late Integration for SFU Drug Response Prediction [ISMB 2019]



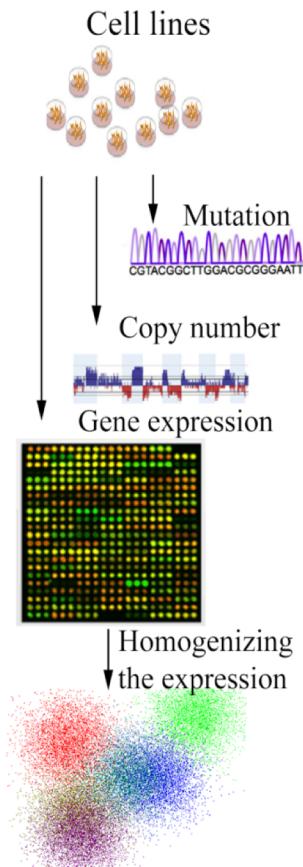
Multi-
omics data



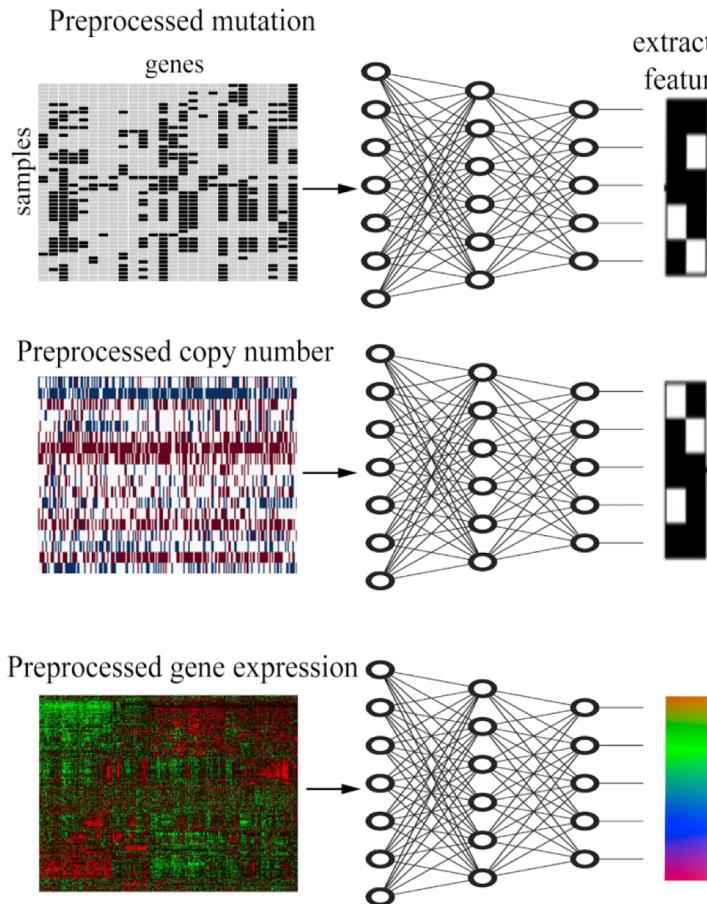
Drug response
(binarized IC50)

MOLI: Overview

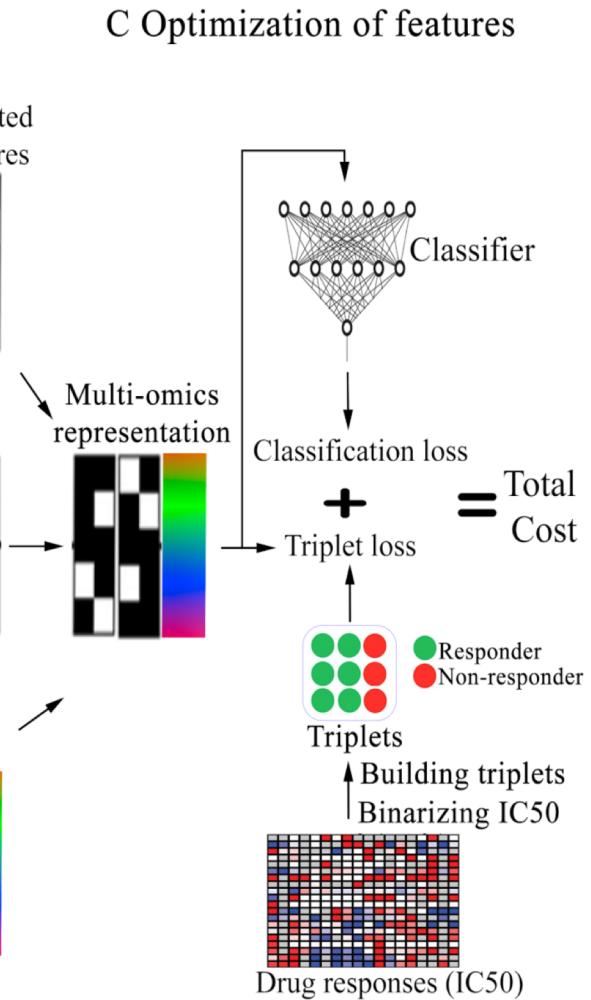
A Preprocessing the input data



B Encoding subnetworks



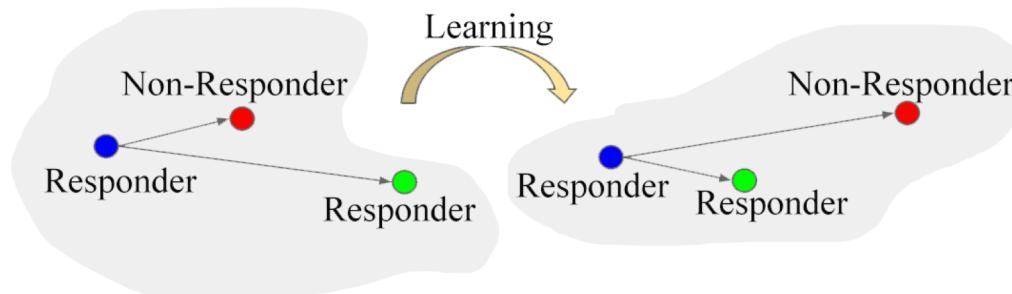
C Optimization of features



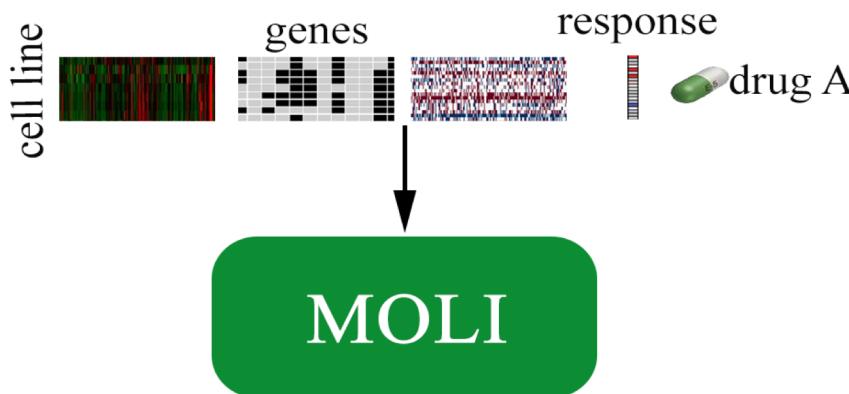
MOLI: Triplet Loss Function

Representation Learning

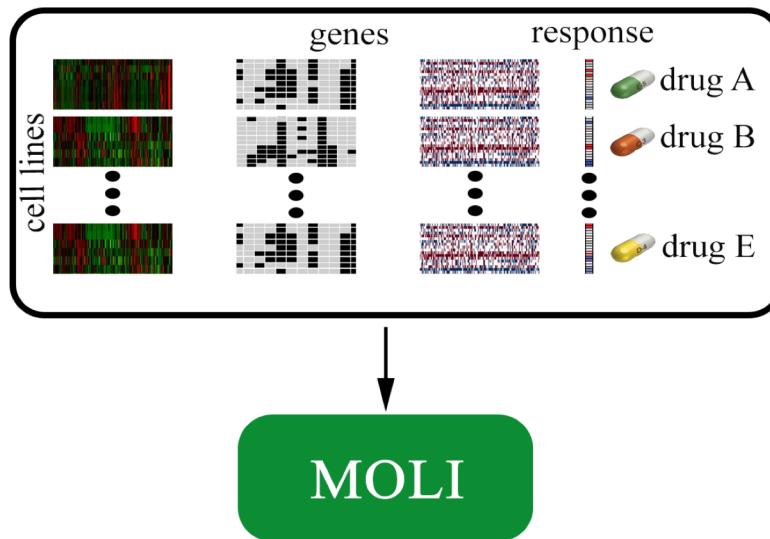
- Inner layers of DNN learn a (lower-dimensional!) representation of the input
input: ~30,000 genes
representation: ~100 dimensions capturing the essence
- In the space of representations, responders should be more similar to each other than to non-responders.
- Construct triplets
(responder 1, responder 2, non-responder)



MOLI: Transfer Learning



Drug-specific DNN



Pan-drug DNN

Transfer between
drugs that target
same pathway

Experimental Evaluation

Datasets

- Cell line data for training

~1000 cell lines with multi-omics data screened with 265 drugs (Iorio et al., 2016 Cell)

- Pre-clinical data for external validation

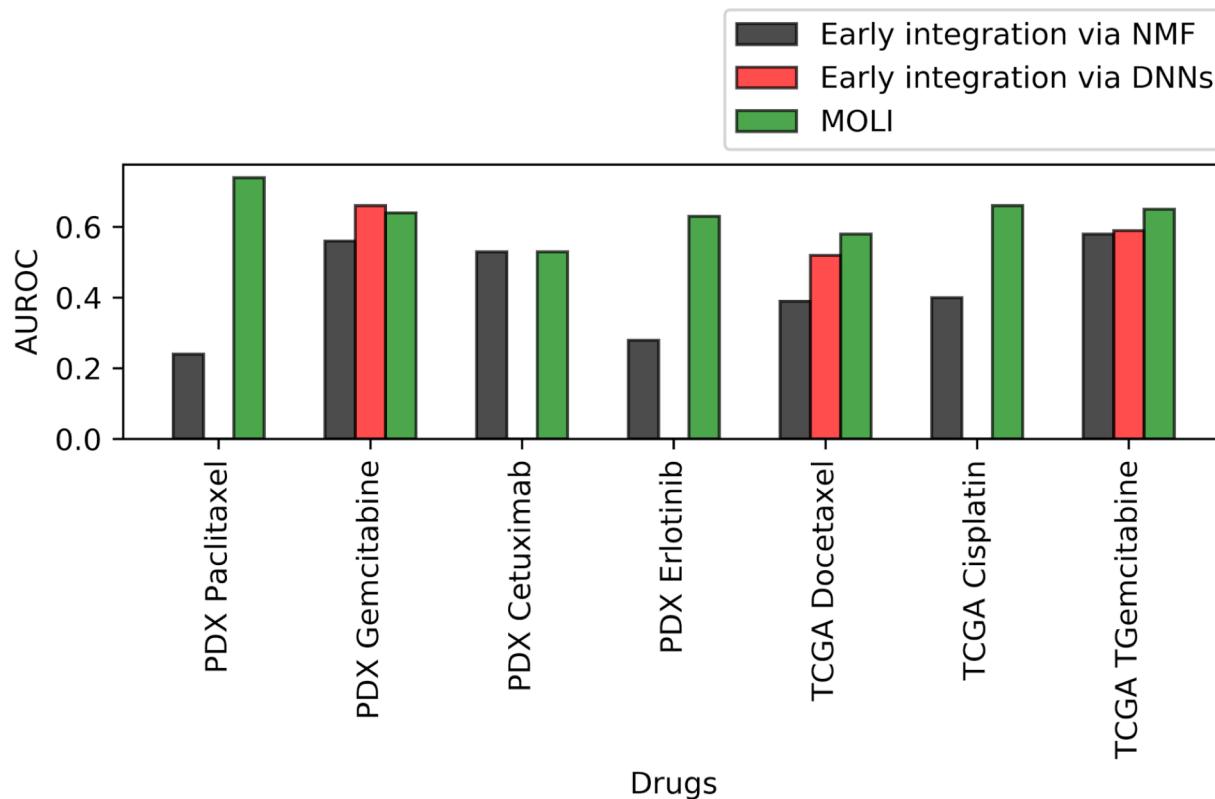
~400 PDX models with multi-omics data screened with 34 drugs (Gao et al., 2015 Nature Medicine)

- Clinical data for external validation

TCGA patients with known drug response

Experimental Evaluation

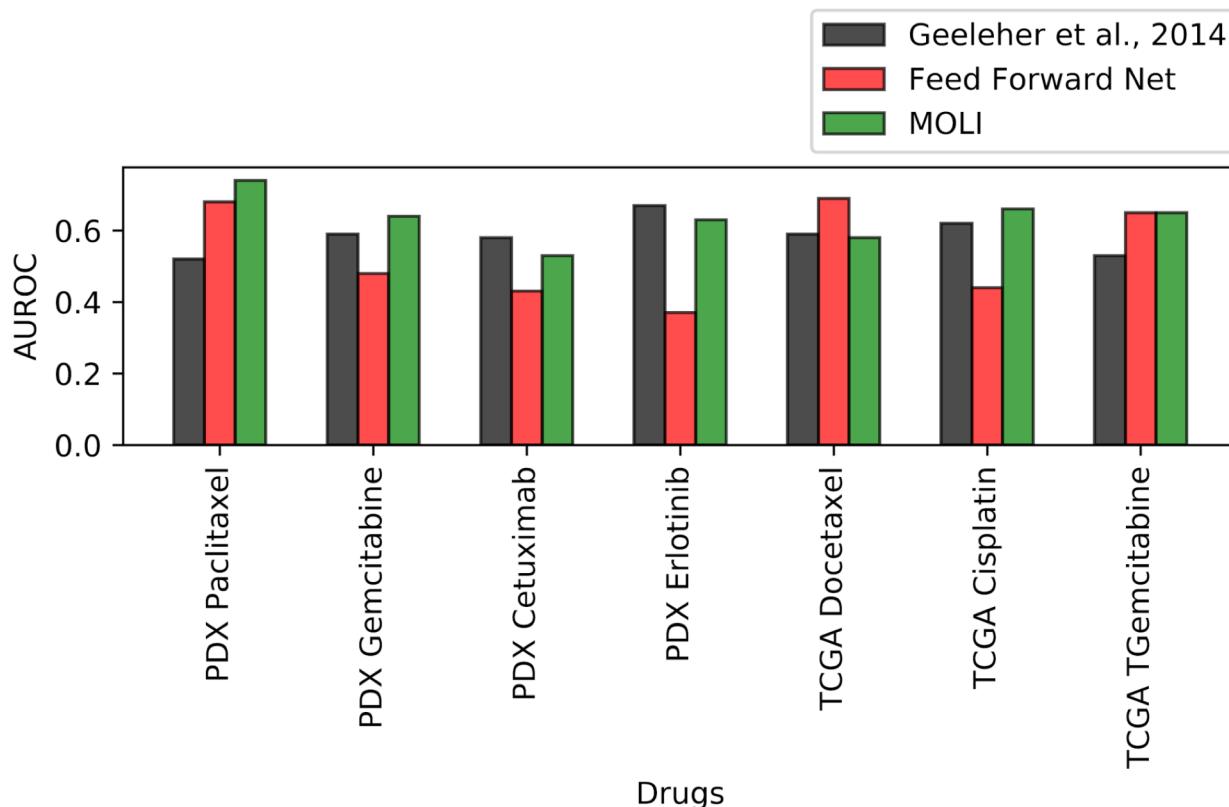
Impact of Integration Type



→ Late integration performs better

Experimental Evaluation

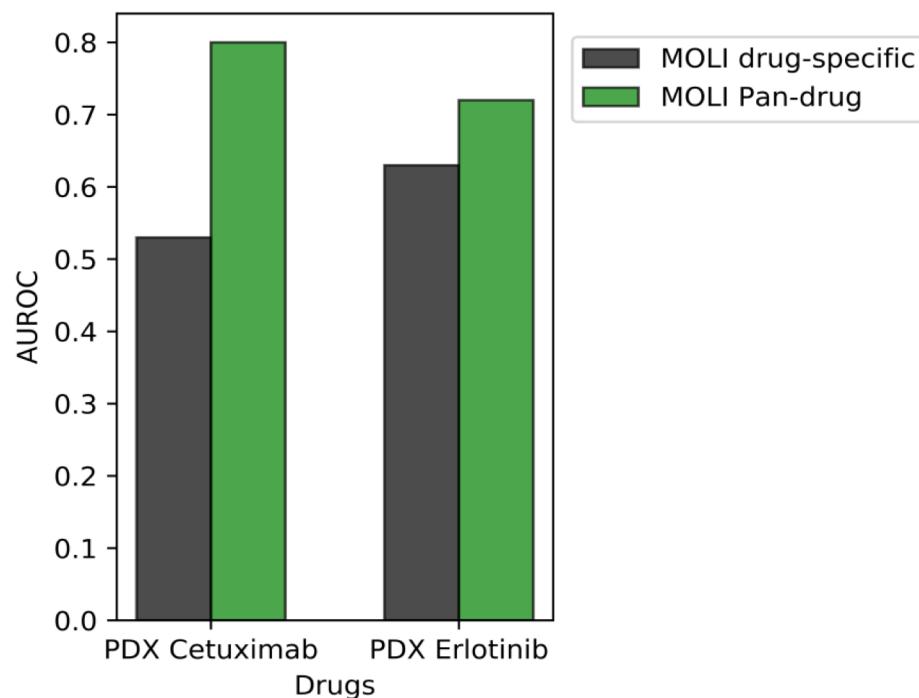
Multi-Omics vs. Single Omics



→ Multi-omics can help

Experimental Evaluation

Pan drug vs. Single drug



→ Transfer learning helps

Ongoing and Future Work

Transfer learning

- Underlying assumption of data mining/machine learning
Training and test data from same probability distribution
- But that assumption is often violated in practice.
- Transfer knowledge from a source domain to a target domain
- Transferred knowledge: model parameters
- How much to transfer from the source domain,
how to adapt to the target domain?
- Especially challenging if no / few labels in the target domain.

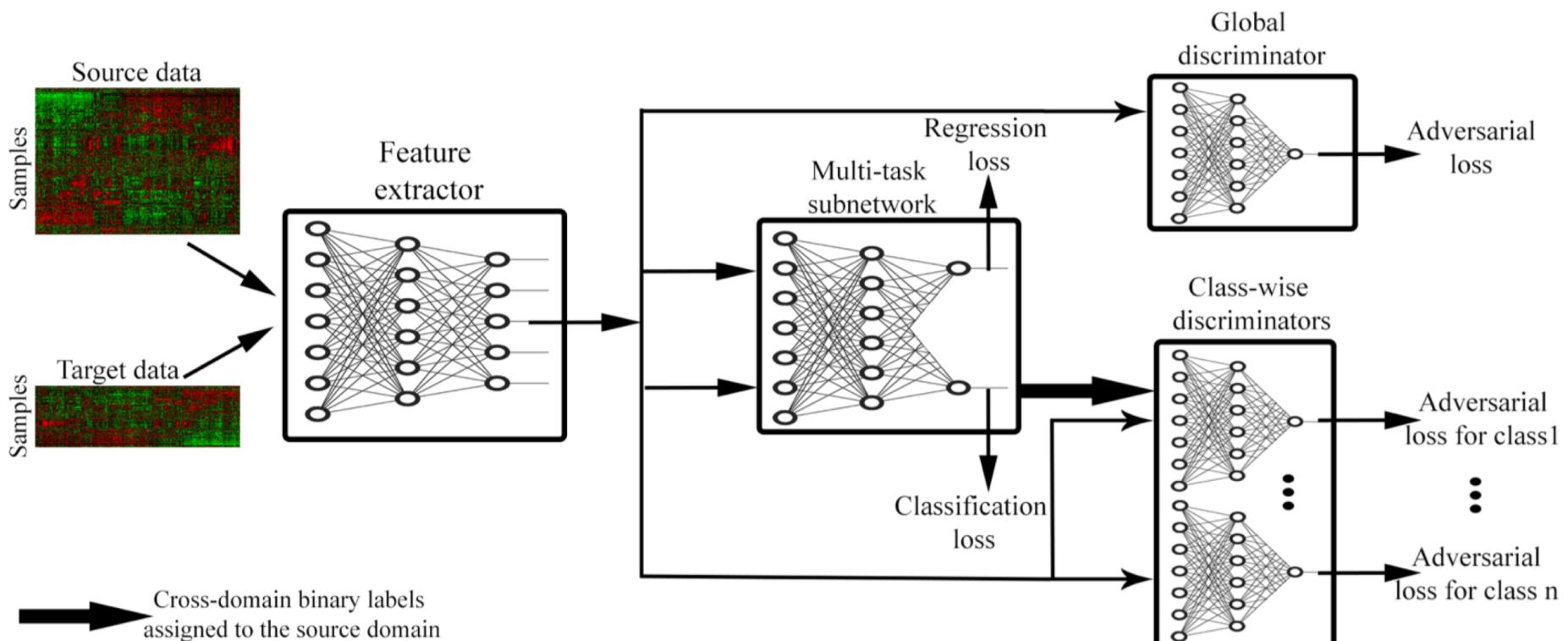
Ongoing and Future Work

Transfer learning

- Source domain: cell lines/PDX
Target domain: patients
- Discrepancy in the input space (X)
genomic data of pre-clinical and clinical datasets
- Discrepancy in the output space (Y)
different measures of drug response:
IC50 vs. change in tumour size
- Adversarial Inductive Transfer Learning (AITL)
[under review at ISMB 2020]
- Learn a feature space in which the source and target domain examples cannot be distinguished.

Ongoing and Future Work

AITL Overview



$$J = L_{BCE} + L_{MSE} + \lambda_G L_{advD_G} + \lambda_{DC} \sum_i L_{advDC_i}$$

Ongoing and Future Work

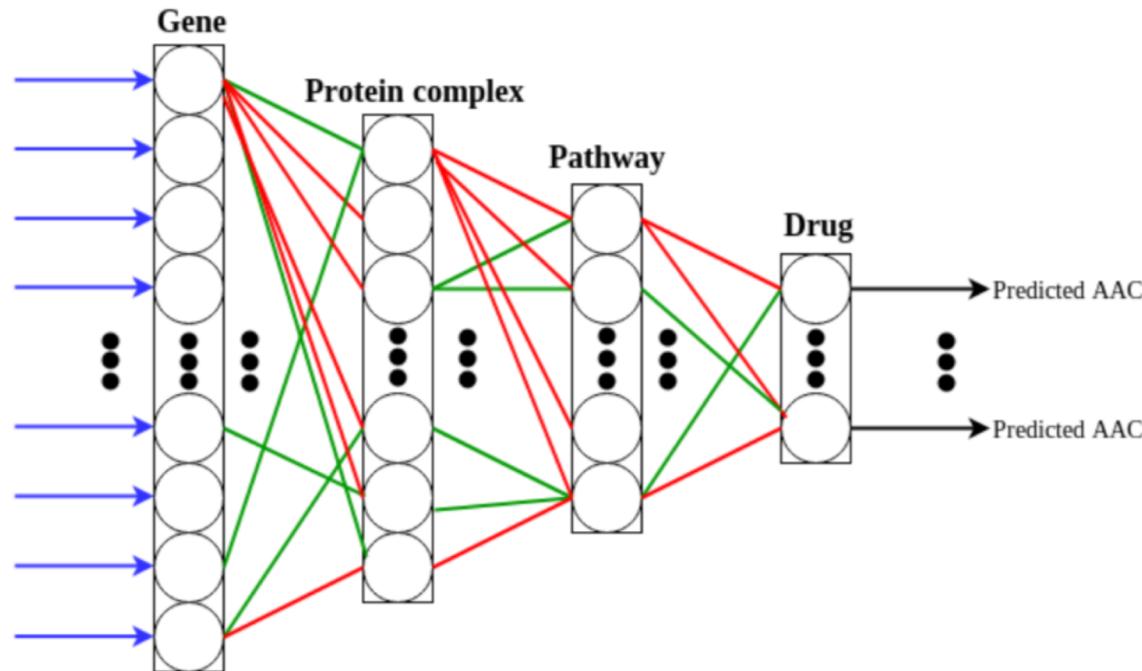
Experimental Evaluation

Method/Drug	Bortezomib	Cisplatin	Docetaxel	Paclitaxel
(Geeleher et al., 2014)	0.48	0.58	0.55	0.53
MOLI (Sharifi-Noghabi et al., 2019b)	0.57	0.54	0.54	0.53
(Chen et al., 2017)	0.54±0.07	0.60±0.14	0.52±0.02	0.58±0.04
ADDA (Tzeng et al., 2017)	0.51±0.06	0.56±0.06	0.48±0.06	did not converge
ProtoNet (Snell et al., 2017)	0.49±0.009	0.42±0.007	0.45±0.02	0.46±0.01
AITL	0.74±0.02	0.66±0.02	0.61±0.03	0.61±0.04

Ongoing and Future Work

Knowledge-based DNNs

- How to incorporate domain knowledge into a DNN?
- Approach: use biological networks to define a sparse DNN architecture (BDKANN) [to be submitted to ECCB 2020]



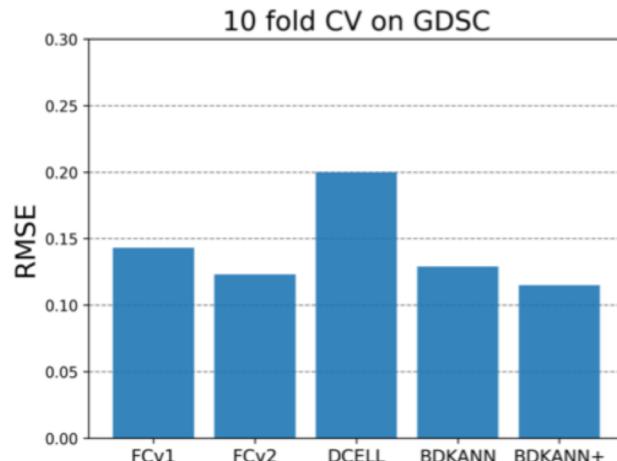
Ongoing and Future Work

Knowledge-based DNNs

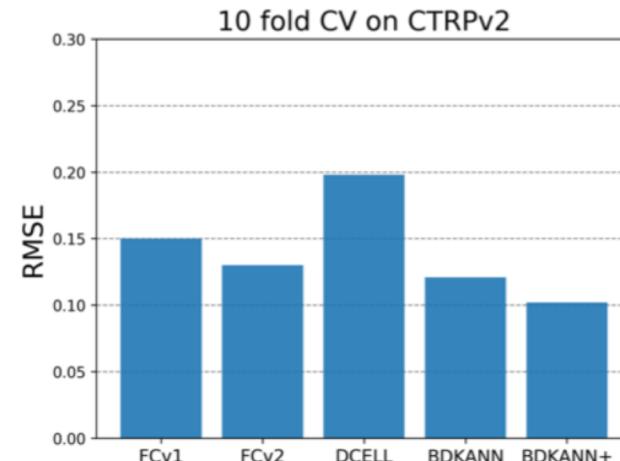
- Biological domain knowledge is incomplete.
- Version BDKANN⁺
Complete the sparse DNN of BDKANN.
- Add l_1 regularization to make the weights of most of the added edges equal to 0.

$$J(\Theta) = MSE(f(X), Y) + \beta l_1(\theta_D)$$

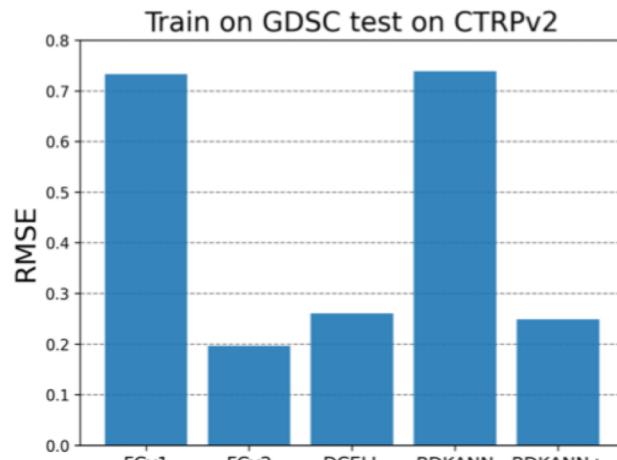
Ongoing and Future Work



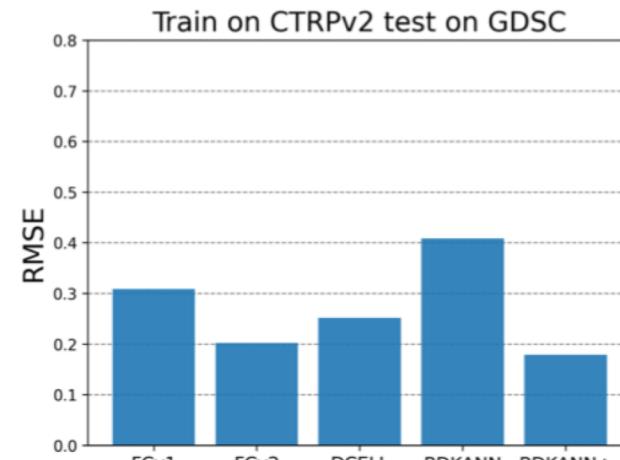
(a)



(b)



(c)



(d)

Ongoing and Future Work

Interpretation of DNNs

- DNN computes highly non-linear function.
- How to explain a single prediction made by a DNN?
- How to explain the overall working of a DNN?
- Trade-off between accuracy and interpretability.

??? Interpretability as post-processing or built-in?

??? Inner nodes of the DNN have no biological meaning.

- Typically, interpretation in terms of input features.
- Knowledge-based DNN can also use inner nodes for interpretation.

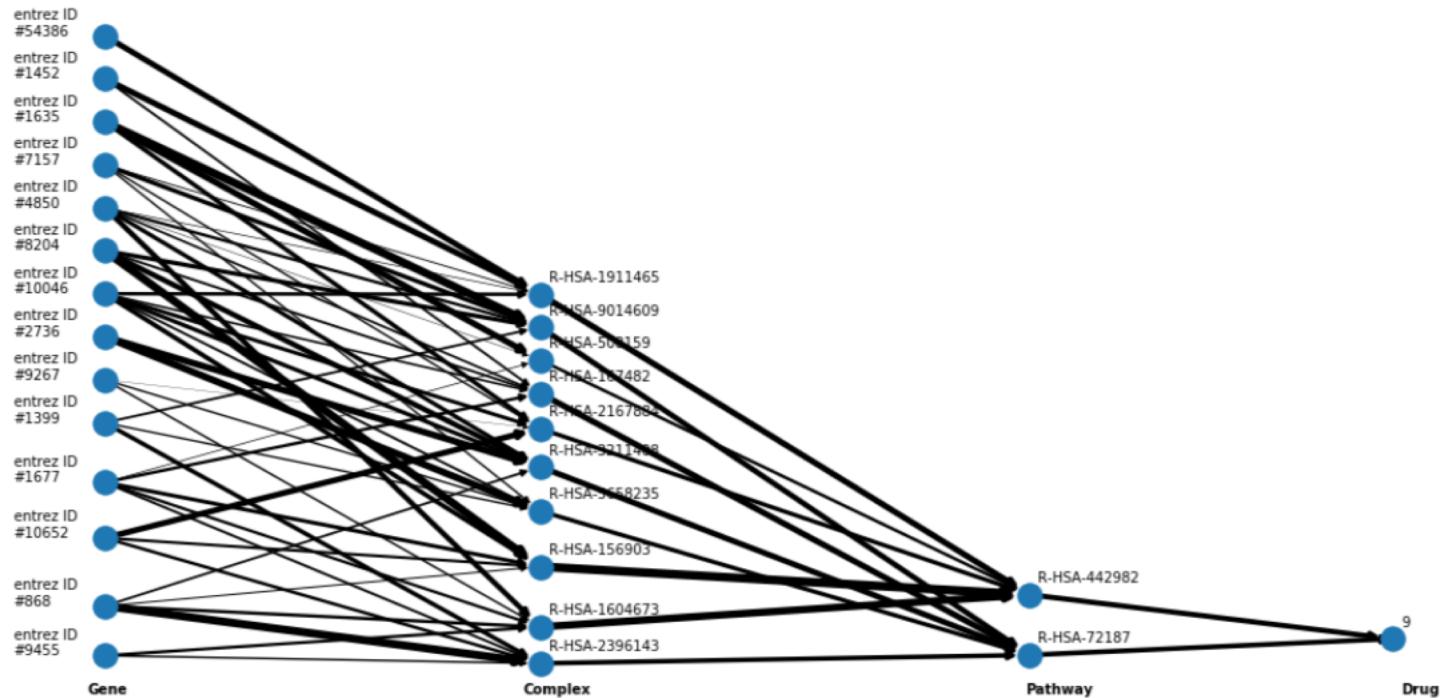
Ongoing and Future Work

Interpretation of BDKANN

- For every training example, compute the relevance (attribution) of every node (at any layer) for classification.
- Aggregate over all examples, obtaining node weights.
- Compute weights of edges based on the weights of their corresponding nodes.
- Extract heavy (high weight) subnetworks that connect the input layer to the output layer.

Ongoing and Future Work

Interpretation of BDKANN



References

Sahand Khakabimamaghani, Martin Ester: Bayesian Bioclustering for Patient Stratification, PSB 2016.

Sahand Khakabimamaghani, Yogeshwar Kelkar, Bruno Grande, Ryan Morin, Martin Ester, Daniel Ziemek: SUBSTRA: Supervised Bayesian Patient Stratification, Bioinformatics, 2019.

Mehrdad Mansouri, Bowei Yuan, Colin Ross, Bruce Carleton, Martin Ester: HUME: Large-scale Detection of Causal Genetic Factors of Adverse Drug Reactions, Bioinformatics, 2018.

Hossein Sharifi-Noghabi, Olga Zolotareva, Colin Collins, Martin Ester: MOLI: Multi-Omics Late Integration with deep neural networks for drug response prediction, ISMB 2019.

Hossein Sharifi-Noghabi, Olga Zolotareva, Colin Collins, Martin Ester: Adversarial Inductive Transfer Learning for Pharmacogenomics Datasets, under review at ISMB 2020.

Oliver Snow, Hossein Sharifi Noghabi, Jialin Lu, Olga Zolotareva, Mark Lee, Martin Ester: BDKANN - Biological Domain Knowledge-based Artificial Neural Network for drug response Prediction, to be submitted to ECCB 2020.