# Looking at the BiG Picture: Incorporating Bipartite Graphs in Drug Response Prediction

David Earl Hostallero[1,2], Yihui Li[1], and Amin Emad[1,2]

[1] Dept. of Electrical and Computer Engineering, McGill University
[2] Mila - Quebec AI Institute

McGill

Emad's Computational Biology and Artificial Intelligence Lab

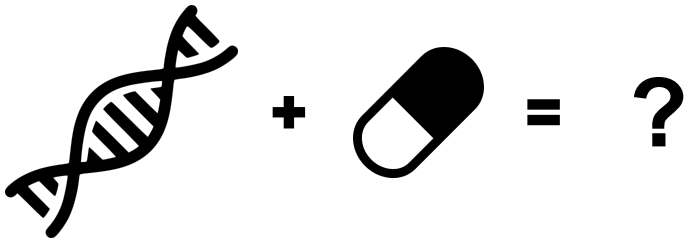
# Cancer is one of the deadliest diseases worldwide

Traditional methods of prescribing cancer drugs **do not ensure positive results**.

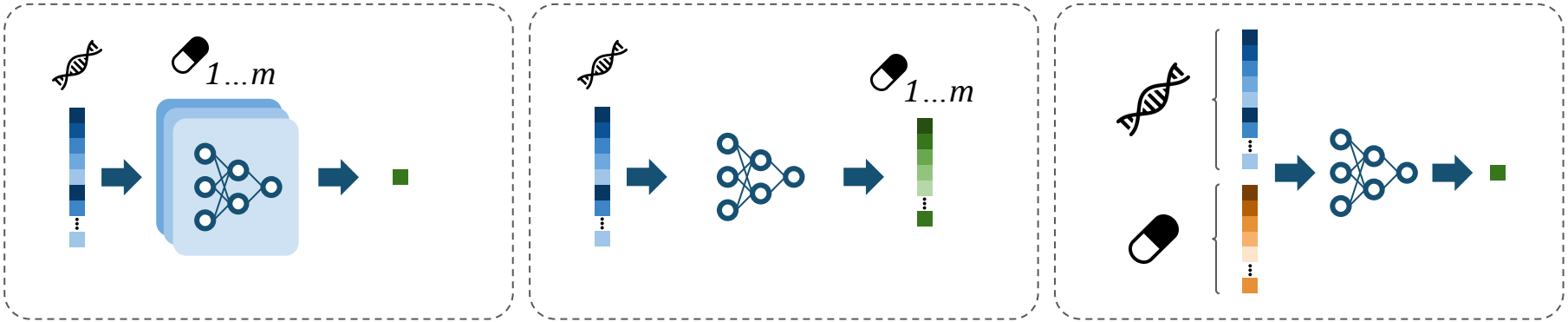
# Increase survival rate through precision medicine

- Prediction of preclinical drug responses is a good step towards individualized medicine
  - more data available
  - many methods are being developed to adapt preclinical models to clinical data

## The drug response prediction (DRP) problem



# Drug Response Prediction




Single model per drug
(e.g., Geeleher et al., 2014)

One output per drug
(multitask)
(e.g., Costello et al., 2014)

Paired-prediction model
(e.g., Liu et al., 2020)

Implicitly learn drug similarities during training

# Pattern Logic

"similar" cancer cell lines (CCLs) → probably similar responses

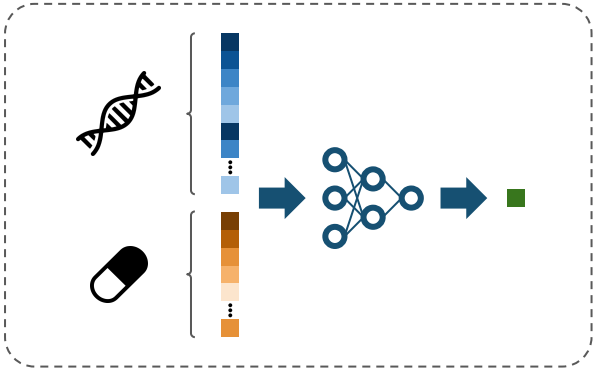
"similar" drugs → probably similar effect

**Similar in terms of what?**

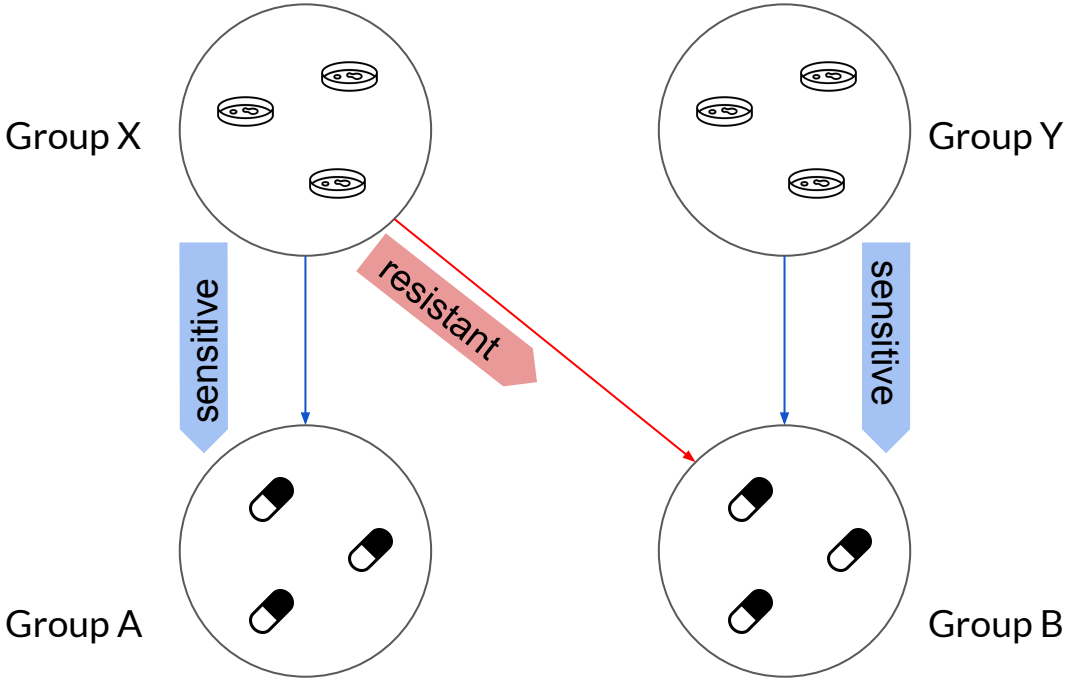
**CCLs:** gene expression, mutation, tissue types

**Drugs**: molecular structure, properties, targets

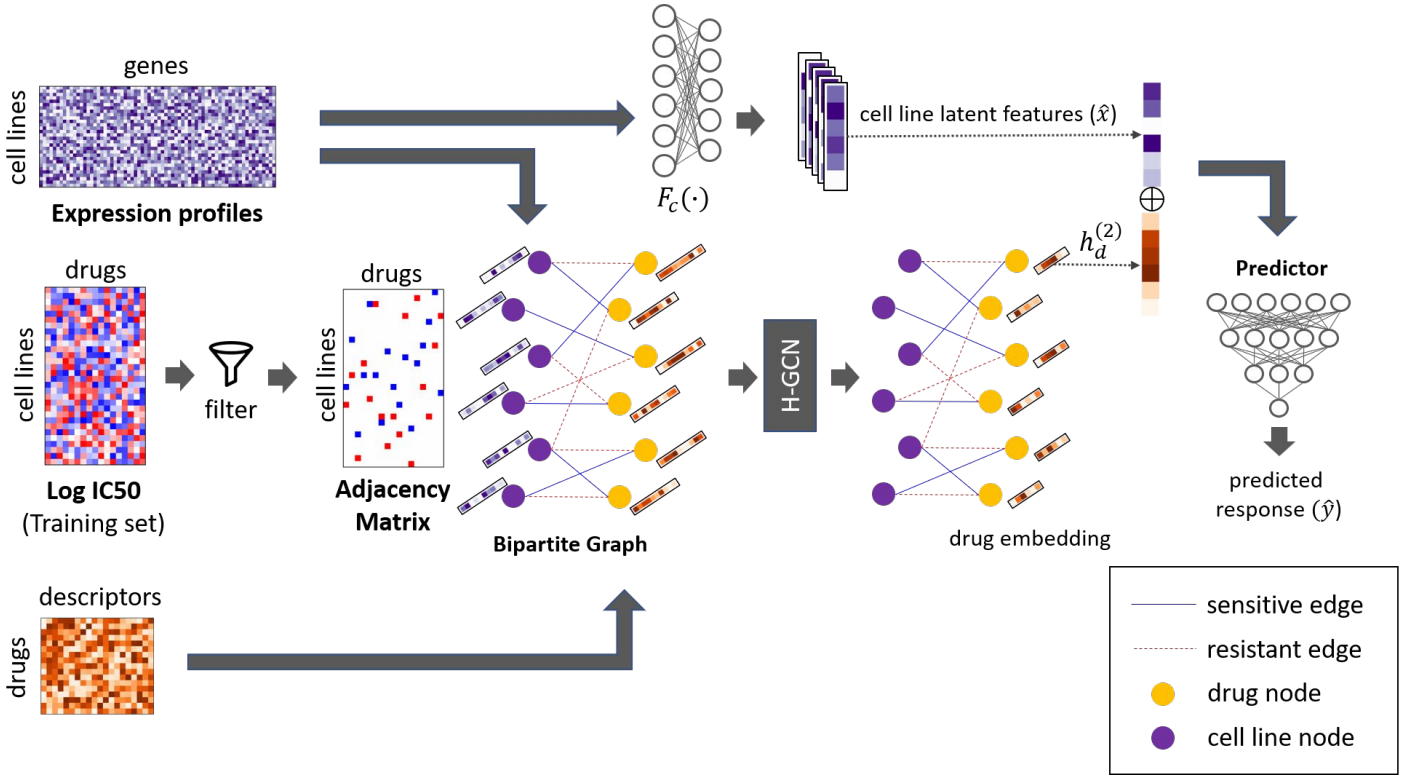
**Are these enough?**



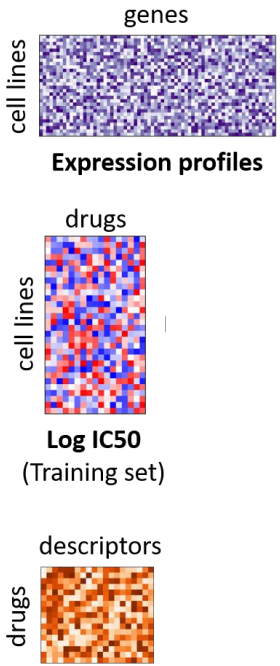
# What if we define representation of the drug according to the **properties of the CCLs that are highly sensitive/resistant to the drug?**



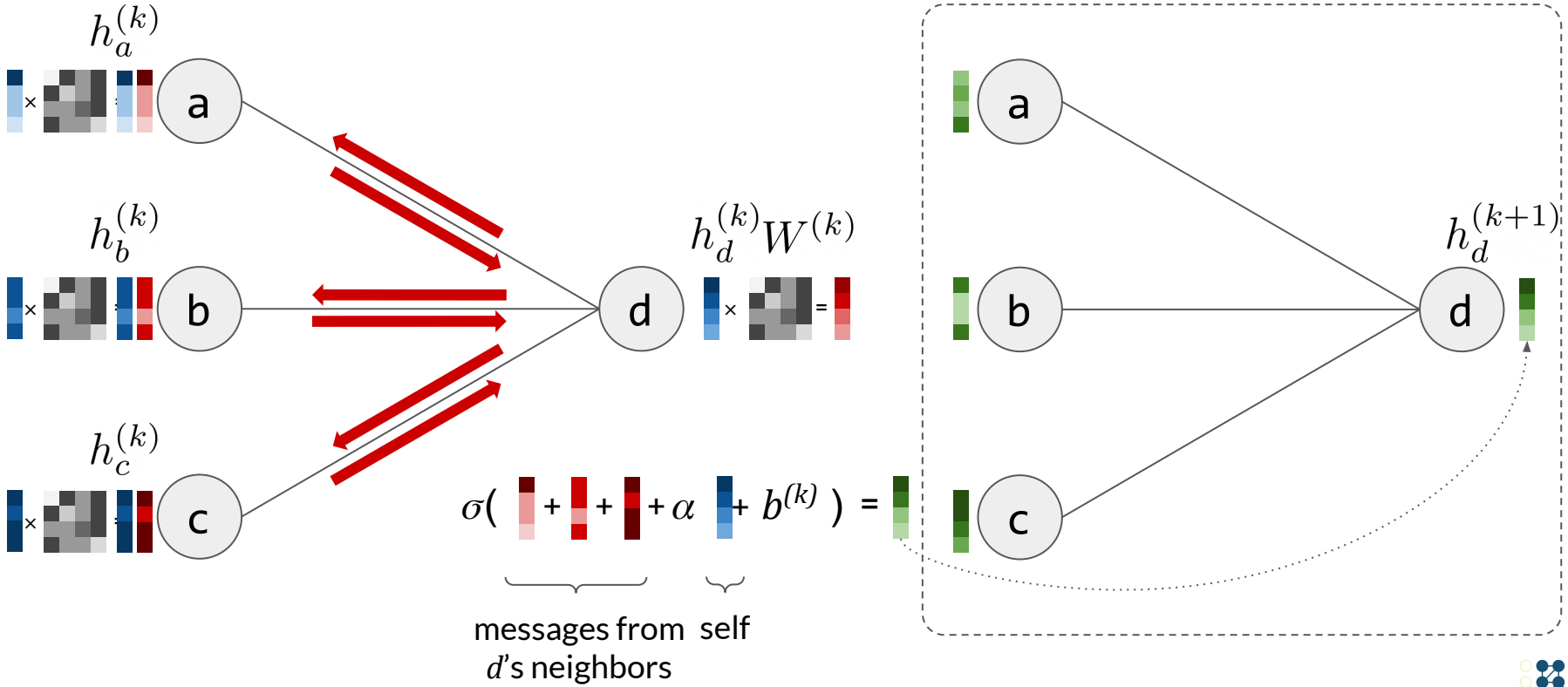
# Bipartite Graph-Represented DR Predictor (BiG-DRP)



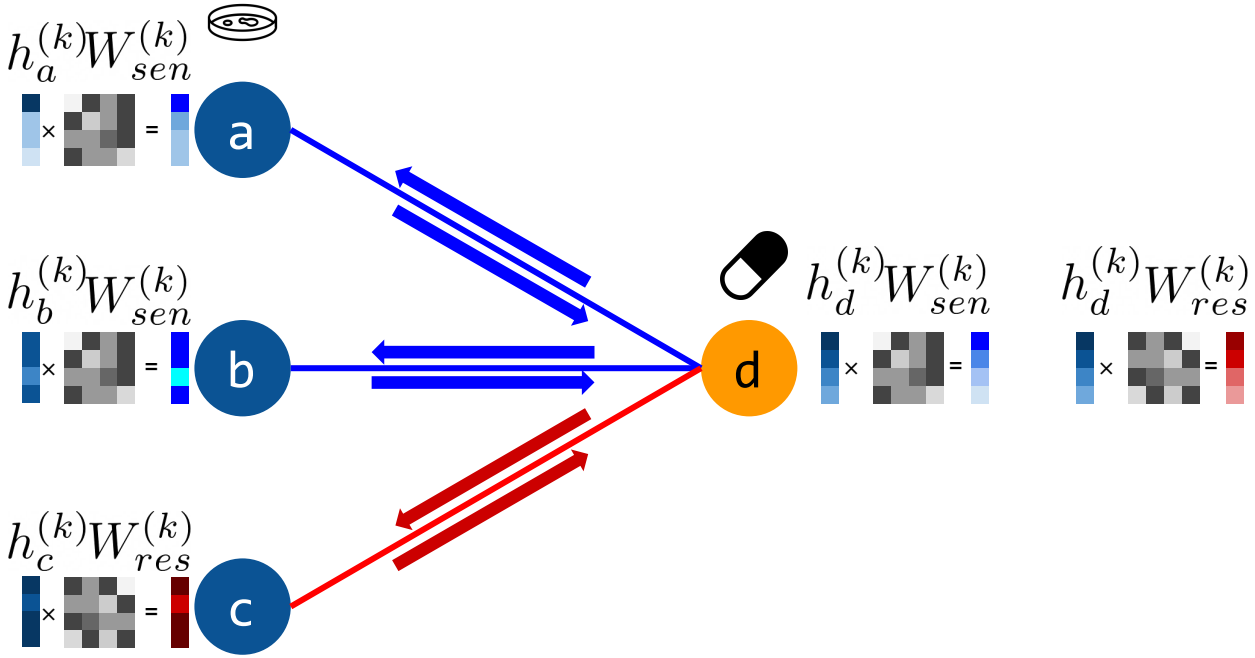
# BiG-DRP



# Graph Convolutional Network (GCN)



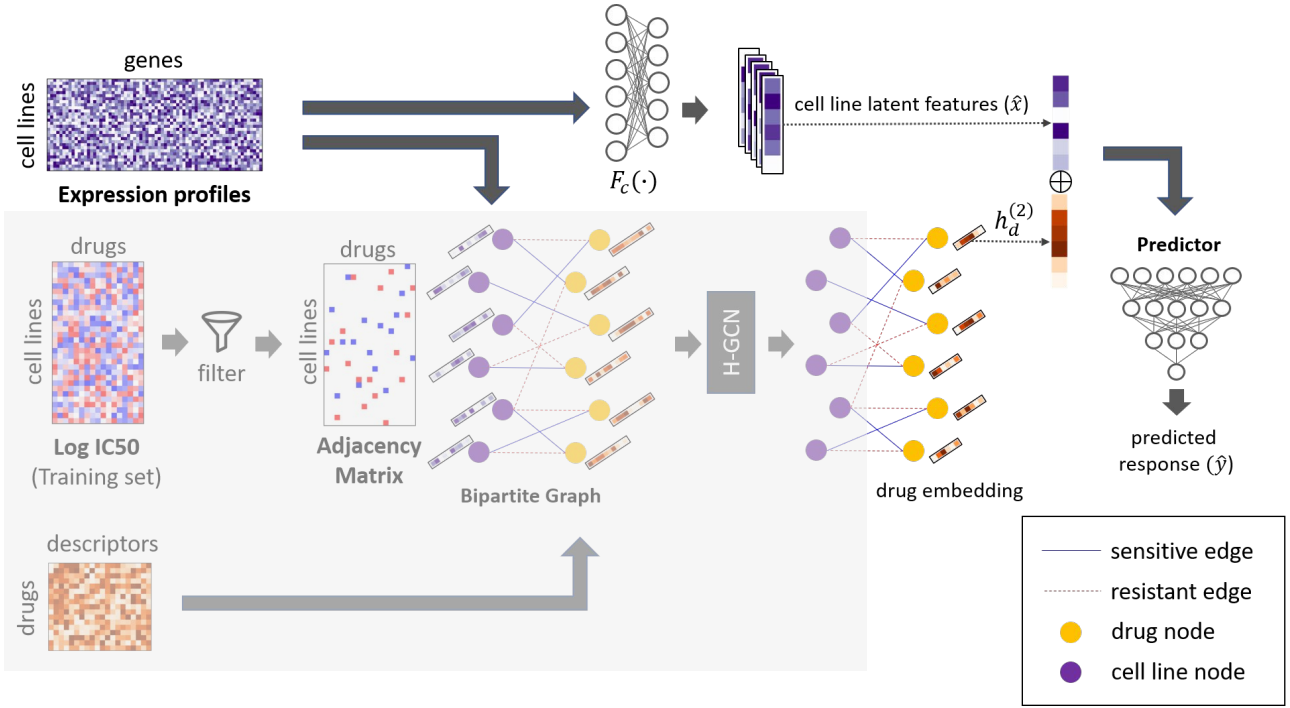
# Heterogenous GCN (H-GCN)




$$h_v^{(k+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\left(b_r^{(k)} + \frac{1}{\sqrt{|\mathcal{N}(v,r)|}}\sum_{u\in\mathcal{N}(v,r)} h_u^{(k)}W_r^{(k)}\right) + \alpha h_v^{(k)}\right)$$

# BiG-DRP+

- Preserve (i.e. freeze) the embeddings
- Lower the learning rate (to avoid overfitting)

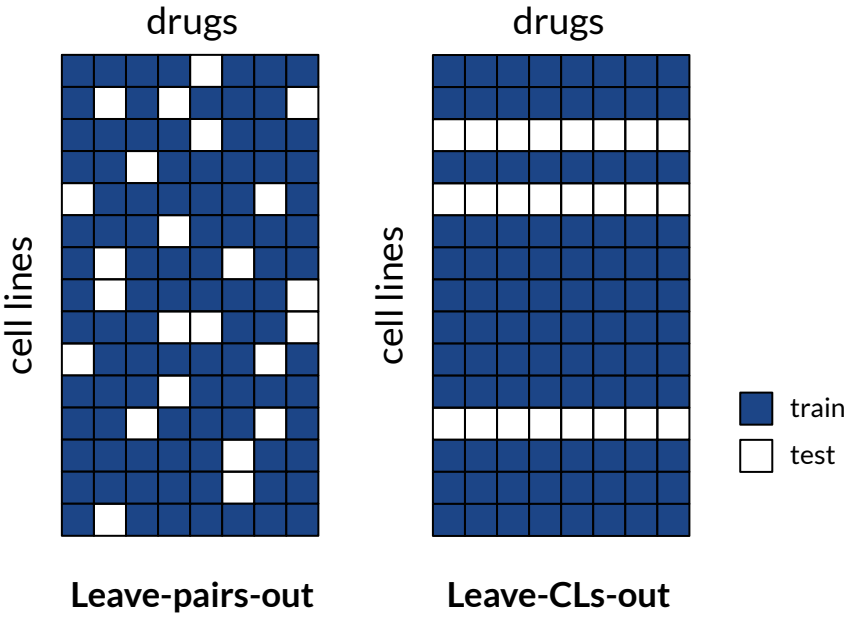


# Data

Genomics of Drug Sensitivity in Cancer (GDSC) Database (Yang et al., 2012)

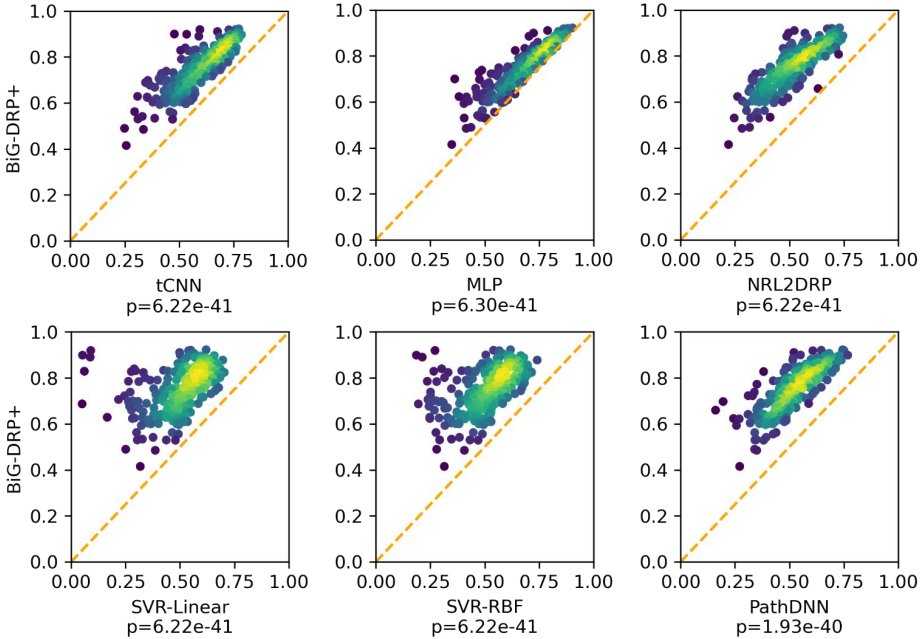
- 990 unique cell lines (RNAseq from Sanger Cell Model Passports)
- 238 unique compounds (descriptors from RDkit)
- ~200k drug responses (z-scored per drug)

# Performance Evaluation & Comparison




| | Drug Features | Other input features |
|---|---|---|
| **BiG-DRP+** | Descriptors | Gene expression |
| **BiG-DRP** | Descriptors | Gene expression |
| **MLP** | Descriptors | Gene expression |
| **SVR-RBF** | Descriptors | Gene expression |
| **SVR-Linear** | Descriptors | Gene expression |
| **PathDNN** (Deng et al., 2020) | Drug Targets | Gene expression, pathway information |
| **tCNN** (Liu et al., 2019) | One-hot SMILES encoding | Genetic Features (mutations) |
| **NRL2DRP** (Yang et al., 2019) | N/A | Drug-CCL-Gene network |

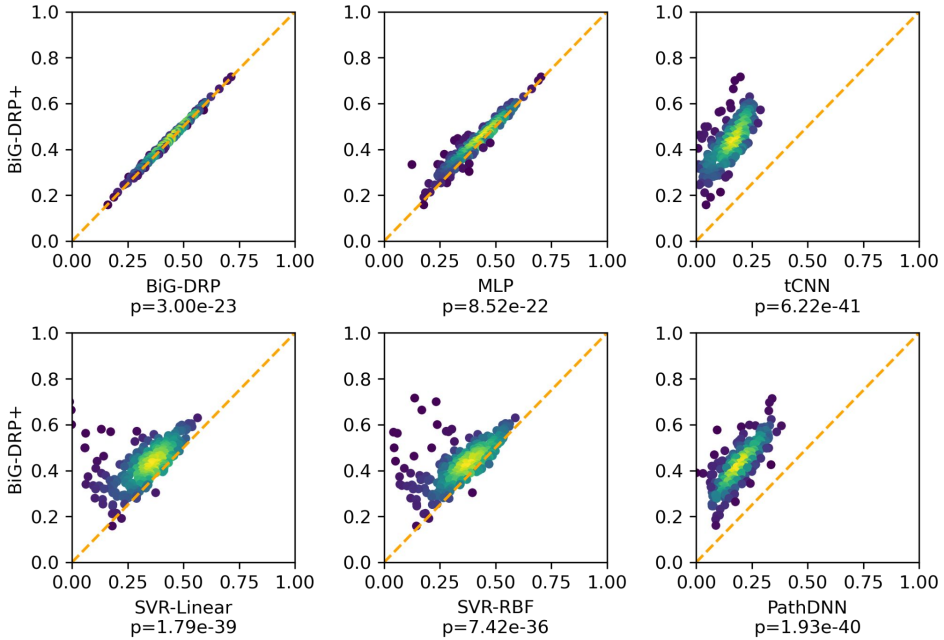
# Leave-pairs-out 5-fold CV




| method | mean SCC (± std.) | mean RMSE (± std.) |
|---|---|---|
| **BiG-DRP+** | **0.748 (± 0.100)** | **0.843 (± 0.241)** |
| BiG-DRP | 0.742 (± 0.100) | 0.855 (± 0.244) |
| MLP | 0.675 (± 0.120) | 0.954 (± 0.274) |
| tCNN (Liu et al., 2019) | 0.587 (± 0.119) | 1.086 (± 0.336) |
| PathDNN (Deng et al., 2020) | 0.516 (± 0.115) | 1.165 (± 0.355) |
| NRL2DRP (Yang et al., 2019) | 0.516 (± 0.119) | 1.153 (± 0.345) |
| SVR-RBF | 0.502 (± 0.123) | 1.181 (± 0.383) |
| SVR-Linear | 0.494 (± 0.129) | 1.184 (± 0.393) |

Drug-wise comparison of Spearman Correlation
(p := p-values of Wilcoxon signed rank test)

# Leave-cell lines-out 5-fold CV




| method | mean SCC (± std.) | mean RMSE (± std.) |
|---|---|---|
| **BiG-DRP+** | **0.431 (± 0.094)** | **1.205 (± 0.367)** |
| BiG-DRP | 0.426 (± 0.095) | 1.210 (± 0.368) |
| MLP | 0.413 (± 0.100) | 1.219 (± 0.374) |
| SVR-RBF | 0.348 (± 0.120) | 1.278 (± 0.403) |
| SVR-Linear | 0.324 (± 0.119) | 1.292 (± 0.420) |
| PathDNN (Deng et al., 2020) | 0.193 (± 0.074) | 2.201 (± 0.698) |
| tCNN (Liu et al., 2019) | 0.147 (± 0.068) | 1.369 (± 0.427) |

Drug-wise comparison of Spearman Correlation
(p := p-values of Wilcoxon signed rank test)

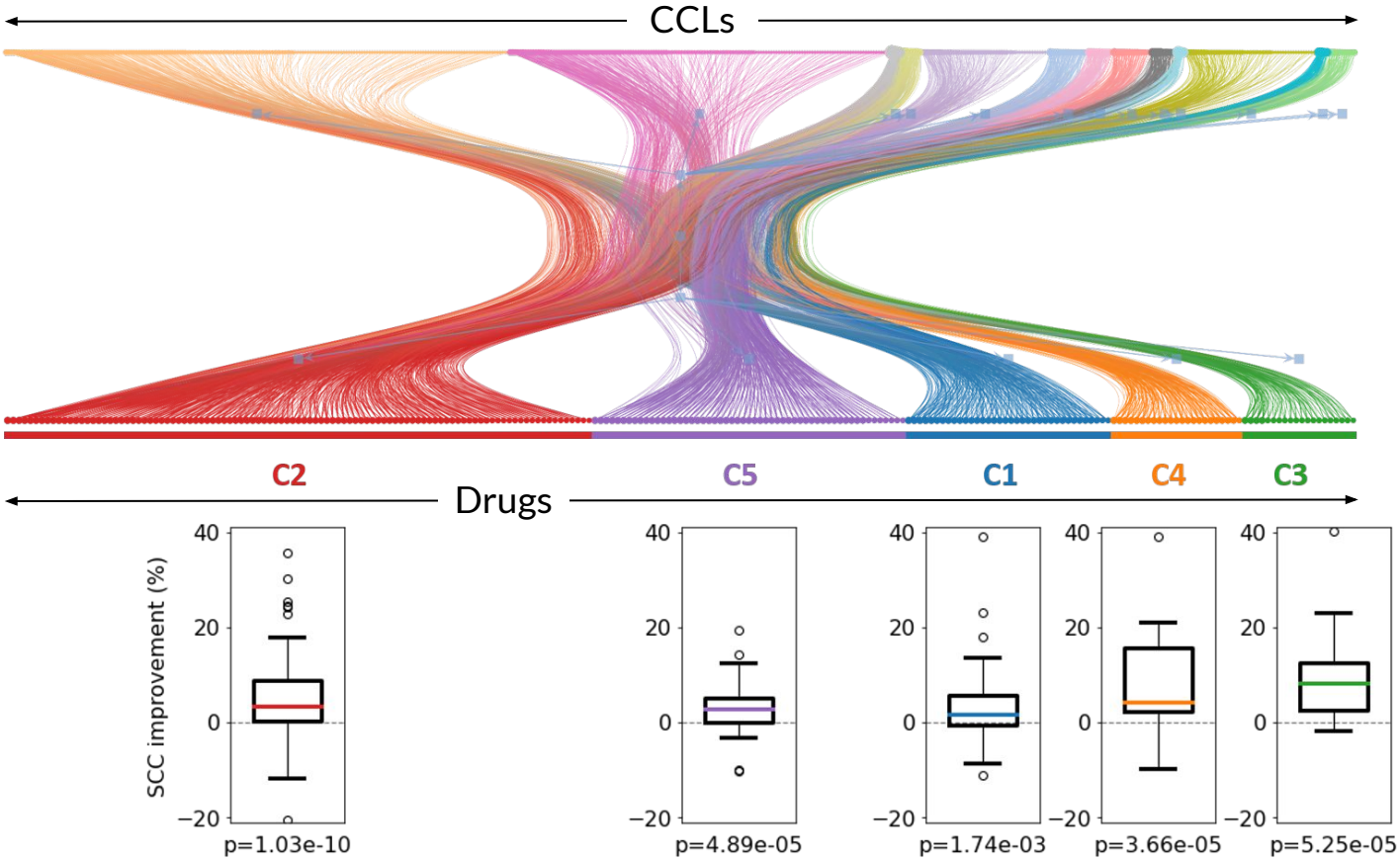
# Drug Feature Assessment

| Method | Drug Attribute | leave-pairs-out | | leave-CLs-out | |
|---|---|---|---|---|---|
| | | AUROC* mean (± std.) | SCC mean (± std.) | AUROC* mean (± std.) | SCC mean (± std.) |
| BiG-DRP+ | Descriptors | 0.878 (±0.068) | 0.748 (±0.100) | 0.746 (±0.077) | 0.431 (±0.094) |
| | Morgan FP | 0.878 (±0.068) | 0.748 (±0.100) | 0.743 (±0.080) | 0.426 (±0.098) |
| | Both | 0.879 (±0.068) | 0.748 (±0.099) | 0.746 (±0.077) | 0.433 (±0.095) |

The method is not sensitive to the drug features

*calculated using continuous value predictions vs binarized labels provided in GDSC

# Drugs with the same MoAs may form clusters

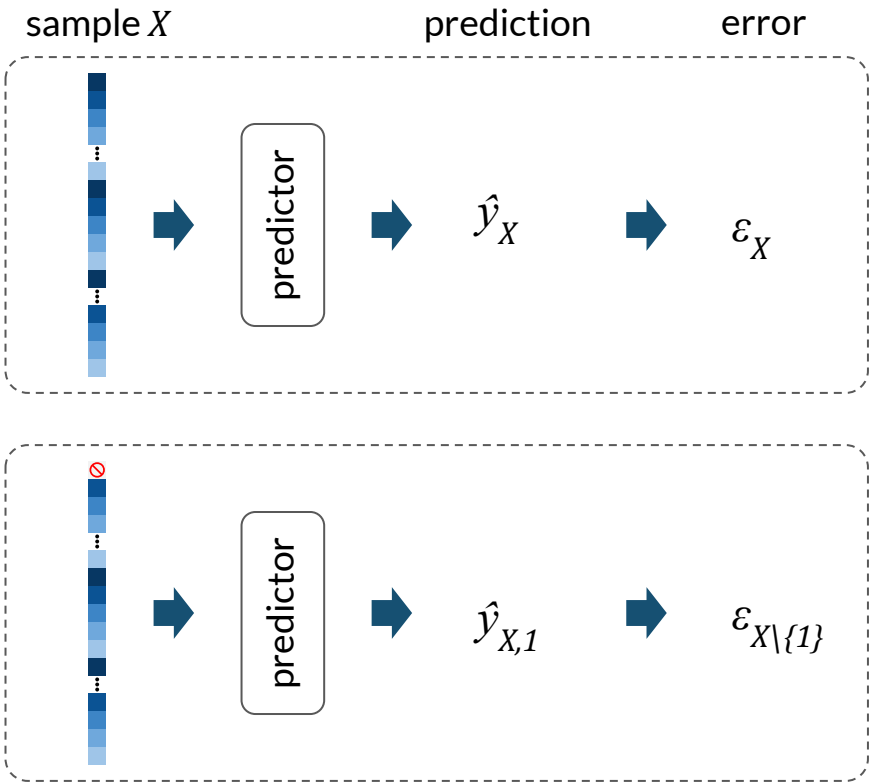



**13/20 protein kinase inhibitors**

8 - serine/threonine protein kinase family

5 - tyrosine kinase family

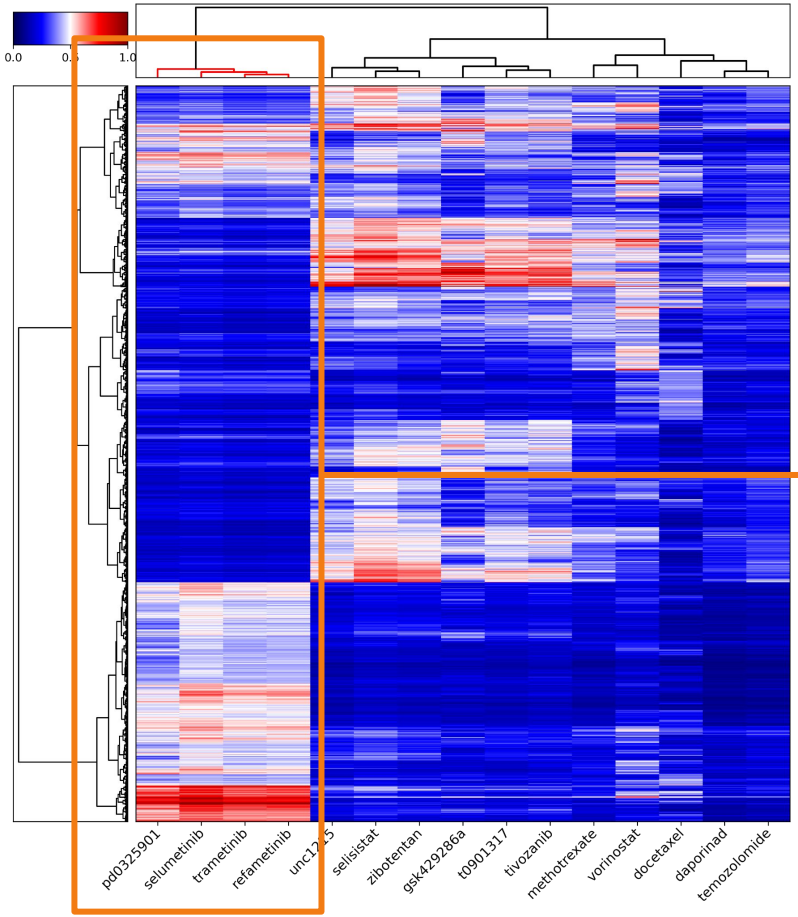
# Gene (feature) attributions




$$\omega_i(X) = \frac{\Delta\varepsilon_{X,i}}{\sum_{j=1\ldots d}\varepsilon_{X,j}}$$

$$\Delta\varepsilon_{X,i} = \varepsilon_{X\setminus\{i\}} - \varepsilon_X$$

(Schwaab et al., 2019)




Identifying and clustering **top-performing drugs** and their **most predictive genes**

inhibit the mitogen-activated protein kinase kinase enzymes (i.e., MEK inhibitors)

# ETV4 and ETV5 are the most predictive genes for Trametinib

- part of the ETS family of oncogenic* transcription factors
- (Sizemore et al., 2017) Upregulated in solid tumors and involved in:
  - Tumor progression
  - Tumor metastasis
  - Chemoresistance

*causes development of tumors

# Clinical Drug Response Prediction

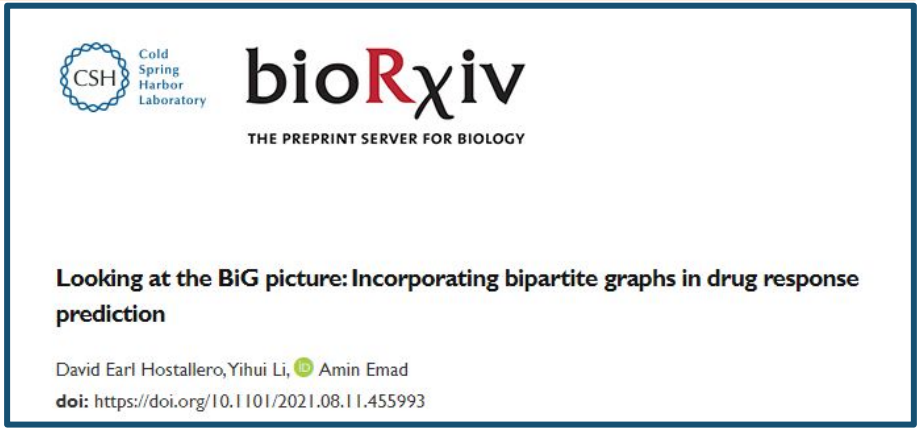
Tested on The Cancer Genome Atlas (TCGA) Database

- Only drugs with at least 150 patients (samples)

| | sensitive | resistant | 1-sided Mann Whitney U p-value | |
|---|---|---|---|---|
| | | | BiG-DRP+ | BiG-DRP |
| cisplatin | 238 | 71 | 2.66e-6 | 2.01e-2 |
| gemcitabine | 74 | 84 | 2.25e-6 | 1.58e-2 |

# Summary

- Presented a drug response prediction method that incorporates bipartite graphs
- BiG-DRP and BiG-DRP+ creates drug representation through the propagation of drug and cell line information using graph convolutions
- Our models surpassed baselines and other competing models in different data-splitting scenarios
- The bipartite graph could provide similarities beyond the molecular structure/properties of the drug




`Code: github.com/ddhostallero/BiG-DRP`

# Future/ongoing work

- Combinational drug therapy
- Preclinical-to-clinical drug response prediction
- Conditional molecule generation

# Thank you

Questions?

# References


- Geeleher P, Cox NJ, Huang RS: Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome Biol 2014, 15:R47
- Costello JC, Heiser LM, Georgii E, et al.: A community effort to assess and improve drug sensitivity prediction algorithms. Nat Biotechnol 2014
- Liu P, Li H, Li S, Leung KS: Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. BMC Bioinformatics 2019
- Deng L, Cai Y, Zhang W, Yang W, Gao B, Liu H: Pathway-Guided Deep Neural Network toward Interpretable and Predictive Modeling of Drug Sensitivity. J Chem Inf Model 2020
- Yang J, Li A, Li Y, Guo X, Wang M: A novel approach for drug response prediction in cancer cell lines via network representation learning. Bioinformatics 2019
- Yang W, Soares J, Greninger P, et al.: Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic acids research 2012, 41:D955-D961
- Schwab P, Karlen W: CXPlain: Causal Explanations for Model Interpretation under Uncertainty. In Advances in Neural Information Processing Systems (NeurIPS). 2019
- Sizemore GM, Pitarresi JR, Balakrishnan S, Ostrowski MC: The ETS family of oncogenic 628 transcription factors in solid tumours. Nat Rev Cancer 2017, 17:337-351