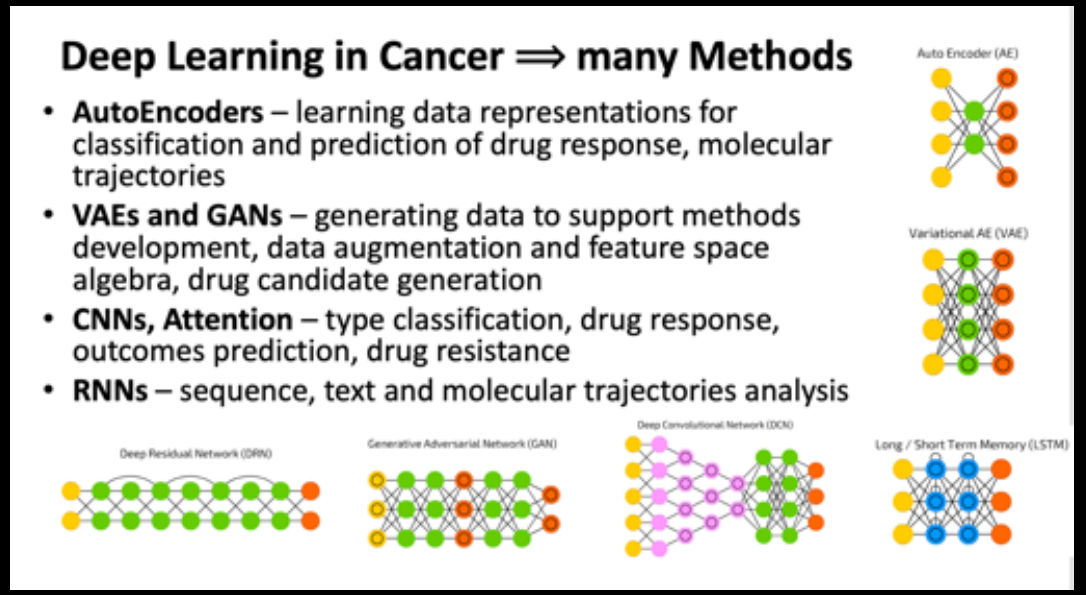
Cancer Drug Response Prediction

Rick Stevens
Argonne National Laboratory
The University of Chicago

Crescat scientia; vita excolatur

Machine Learning In Cancer Research

- Cancer Susceptibility
- Cancer Detection and Diagnosis
- Cancer Recurrence
- Cancer Prognosis and Survival
- Cancer Classification and Clustering
- Cancer Drug Response Prediction
- Cancer Genomics Analysis
- Cancer Medical Records Analysis
- Cancer Biology



Medical Sute presents an infographic look at

CANCER

1 in 4 of us will die from CANCER

has an overall death rate of:

59%

Although this rate has been dropping consistently since the peak in 1991, when 215 people died from one from or cancer or another per 100,000 US residents

1 in 4 people in the USA will die from cancer

Approximately 1500 people die every day from a form of cancer

Cancer death rates (USA):

Heart and soft tissue
Death Rate: 42%

Brain and nervous system
Death Rate: 10%

Digestive system
Death Rate: 42%

Lung and respiratory system
Death Rate: 85%

Eye and orbit
Death Rate: 10%

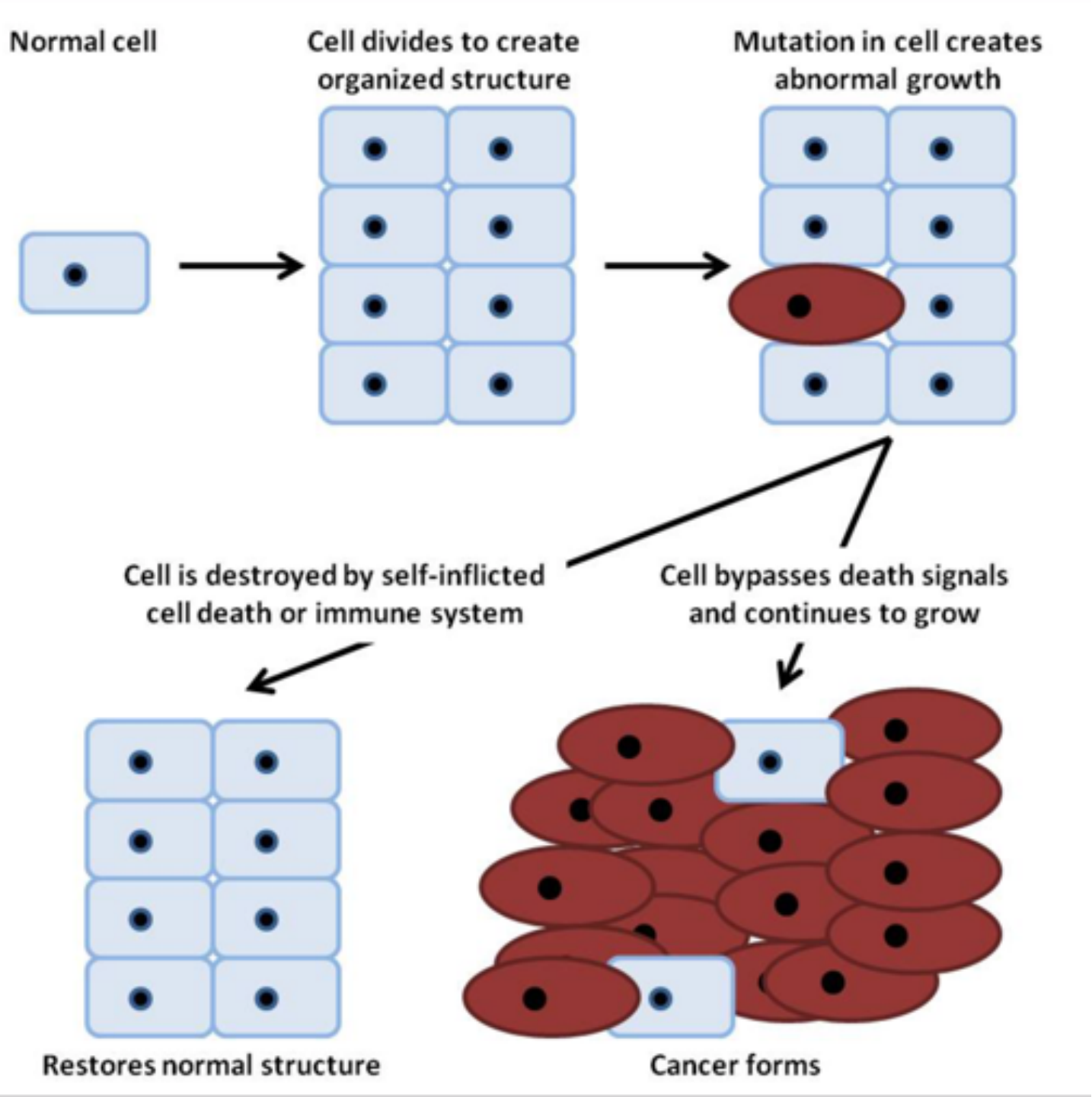
Bones and joints
Death Rate: 48%

Prostate and male genital system
Death Rate: 14%

Female genital system
Death Rate: 32%

What is Cancer?

- Large number of complex diseases
- Each behave differently depending on cell type from which the tumor originates
  - Age on onset, invasiveness, response to treatment
- Common General Properties
  - Abnormal cell growth/division (cell proliferation)
  - Spread to other regions of body (metastasis)
  - Malignant tumors


Normal cell
Cell divides to create organized structure
Mutation in cell creates abnormal growth
Cell is destroyed by self-inflicted cell death or immune system
Cell bypasses death signals and continues to grow
Restores normal structure
Cancer forms

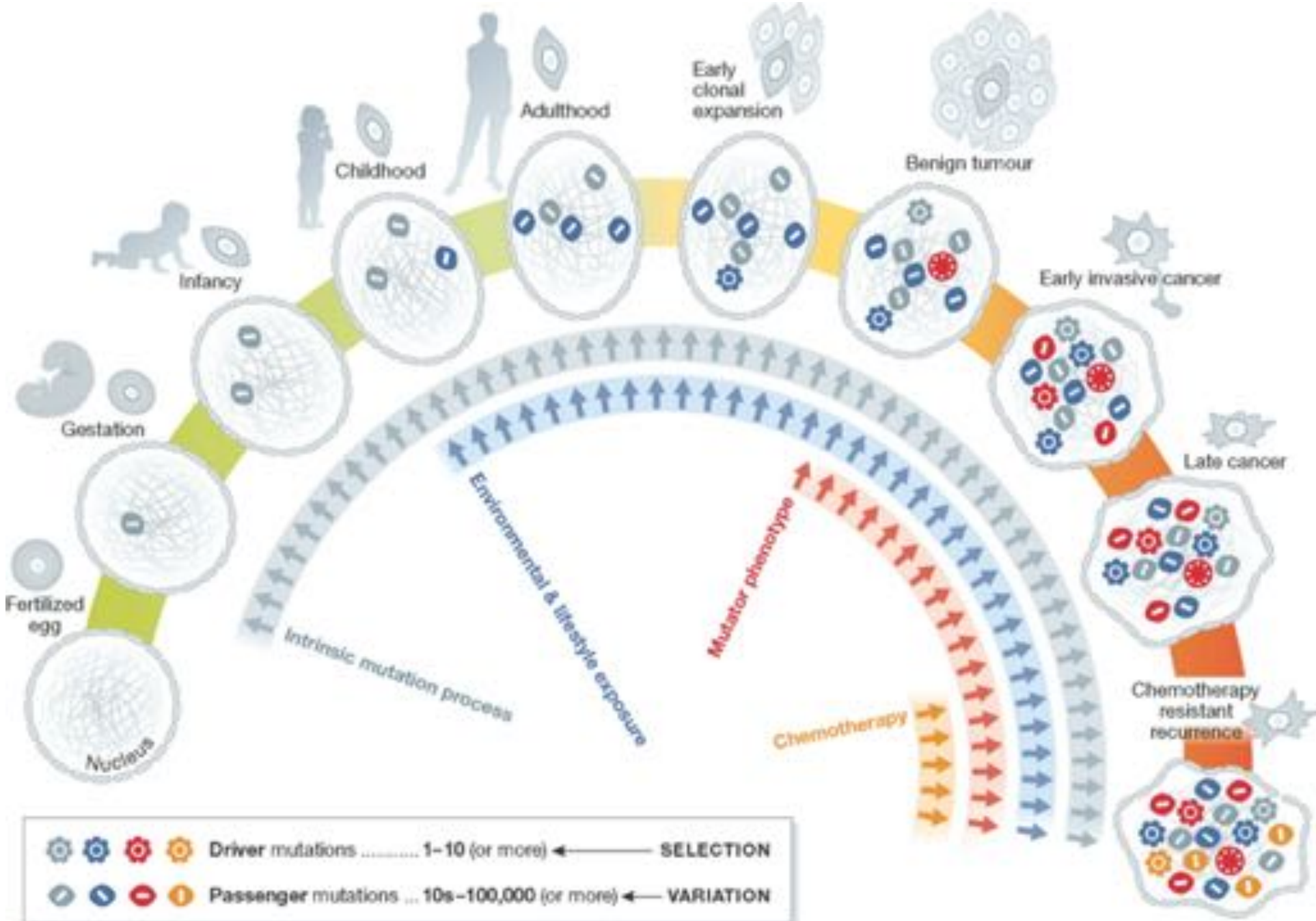

Fertilized egg
Nucleus
Gestation
Infancy
Childhood
Adulthood
Early clonal expansion
Benign tumour
Early invasive cancer
Late cancer
Chemotherapy resistant recurrence
Intrinsic mutation process
Environmental & lifestyle exposure
Mutator phenotype
Chemotherapy
Driver mutations ..... 1–10 (or more) ← SELECTION
Passenger mutations ... 10s–100,000 (or more) ← VARIATION

Table. Last 20 Oncologic Drugs Approved Between 2009 and 2013 by the US Food and Drug Administration

Drug and Indication	Cost per Year of Treatment, $[a]	Parent Drug	Mechanism of Action	Clinical Benefit
Sorafenib for papillary thyroid cancer	140 984	NA	First approved VEGFR and RAS tyrosine kinase inhibitor	Median PFS, 10.8 vs 5.8 mo
Cri[illegible]nib [illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Ibr[illegible]nib [illegible]	[illegible]	NA	[illegible]	[illegible]
Obinutuzumab for chronic lymphocytic leukemia	74 3[illegible]	Rituximab	Anti-CD2[illegible] monoclonal antibody	Media[illegible] PFS, 23.0 vs 11.1 mo
Pe[illegible]tuzumab for breast cancer	78 252	Tra[illegible]tuzumab	Anti-her2 monoclonal antibody	Pathologic CR, 39.3% vs 21.6%
N[illegible]	[illegible]	Pa[illegible]	[illegible]	Media[illegible]
Afatinib for non-small-cell lung cancer	79 920	Erlotinib	EGFR tyrosine kinase inhibitor	Median PFS, 11.1 vs 6.9 mo; median OS, NS
Lenalidomide for mantle-cell lymphoma	124 870	Thalidomide	Immunomodulatory drug (thalidomide analogue)	RR, 26%; median DOR, 16.6 mo
Trametinib for malignant melanoma	125 280	NA	First approved mek inhibitor	Median PFS, 4.8 vs 1.5 mo
Dabrefenib for malignant melanoma	109 440	Vemurafenib	BRAF inhibitor	Median PFS, 5.1 vs 2.7 mo; median OS, NS
Radium 223 for prostate cancer	82 800	NA	First approved radiotherapeutic drug	Median OS, 14.0 vs 11.2 mo
Erlotinib for non-small-cell lung cancer	82 827	NA	First approved EGFR [illegible] inhibitor	Median PFS, 10.4 vs 5.2 mo; m[illegible]
[illegible] for breast cancer	113 161	[illegible]	First approved [illegible] drug conjugate	M[illegible] vs 9.4 mo; m[illegible] OS, 29.1 vs 20.9 mo
Pomalidomide for multiple myeloma	150 408	Thalidomide	Immunomodulatory drug (thalidomide analogue)	RR, 29%; median DOR, 7.4 mo
Bevacizumab for colorectal cancer	59 422	NA	First anti-VEGF monoclonal antibody	Median PFS, 5.7 vs 4 mo; median OS, 11.2 vs 9.8 mo
Ponatinib for chronic myeloid leukemia and Ph+ acute lymphoblastic leukemia	137 952	Imatinib	Bcr-abl tyrosine kinase inhibitor	Major cytogenetic response, 54%; median DOR, 3.2-9.5 mo
Abiraterone for prostate cancer	92 092	Ketoconazole	Androgen biosynthesis inhibitor	Median OS, 35.3 vs 30.1 mo
Cabozantinib for medullary thyroid cancer	118 800	NA	First multi[illegible] (medullin, met and VEGF[illegible]	Median PFS, 11.2 vs 4 mo; median OS, NS
Omacetaxine for chronic myeloid leuken[illegible]	8[illegible]	[illegible]	[illegible] protein translation inhibitor	Major cytogenetic response, 14.3%; median DOR, 12.5 mo
Nab-paclitaxel[b] for non-small-cell lung cancer	82 231	Paclitaxel	Albumin-bound paclitaxel (microtubule inhibitor)	RR, 33% vs 25%; median OS, NS
Regorafenib for colorectal cancer	141 372	Sorafenib	Multikinase inhib[illegible]	Median PFS, 2 vs 1.7 mo; m[illegible] OS[illegible]

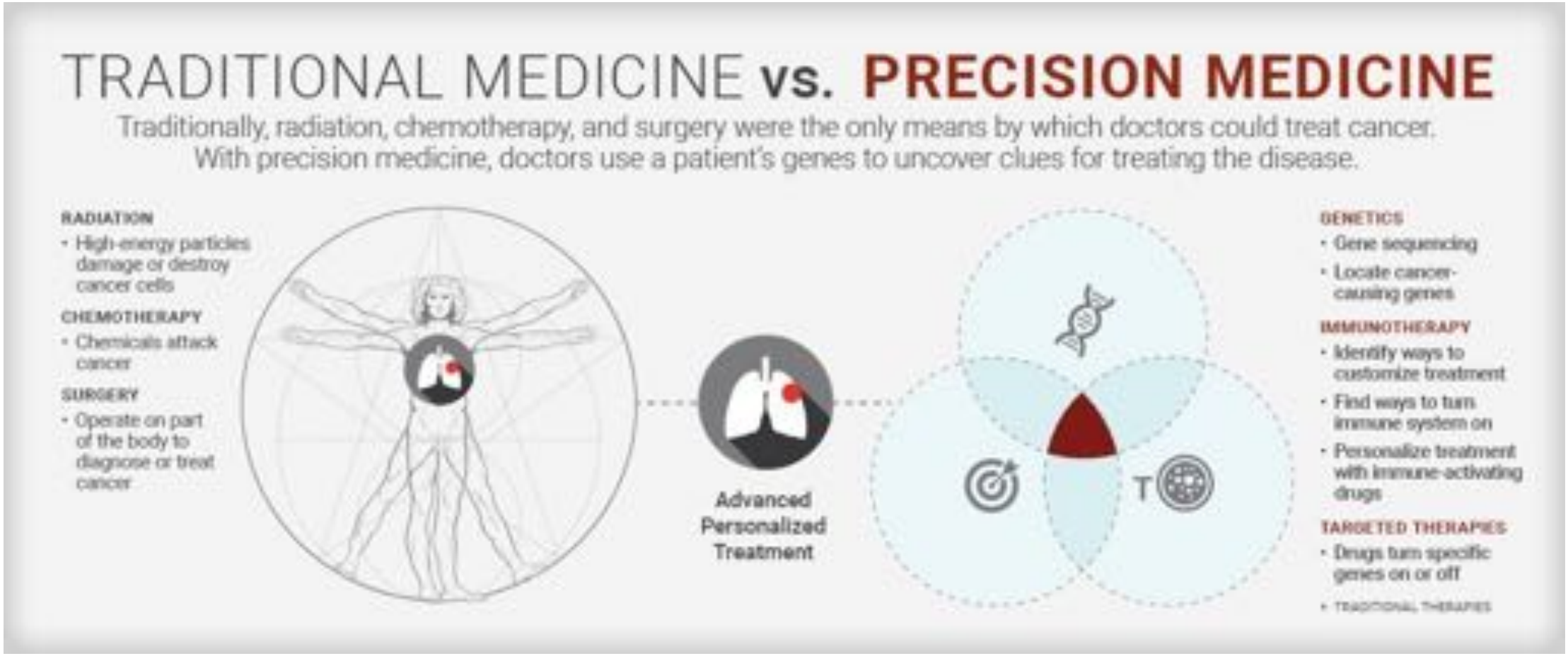
[illegible]ev[illegible] [illegible] applicable; NS, [illegible] significant; OS, overall survival; PFS, progression-free survival; Ph+, Philadelphia chromosome positive; RR, response rate; UA, unavailable; (V)EGF(R), (vascular) endothelial cell growth factor (receptor).

[illegible]erage wholesale p[illegible]ere obtained [illegible] ([subscription required] http://www.redbook.com/redbook/online/).

[b] This drug was approved separately for 2 indications.

20 Oncologic Drugs Approved
Between 2009 and 2013
Treatment cost $100K per year
Average PFS
Improvement < 6 months

50% of Patients do not respond to chemotherapy for some tumors



Drug Response is specific to Cancer type and specific genetic variance in each tumor

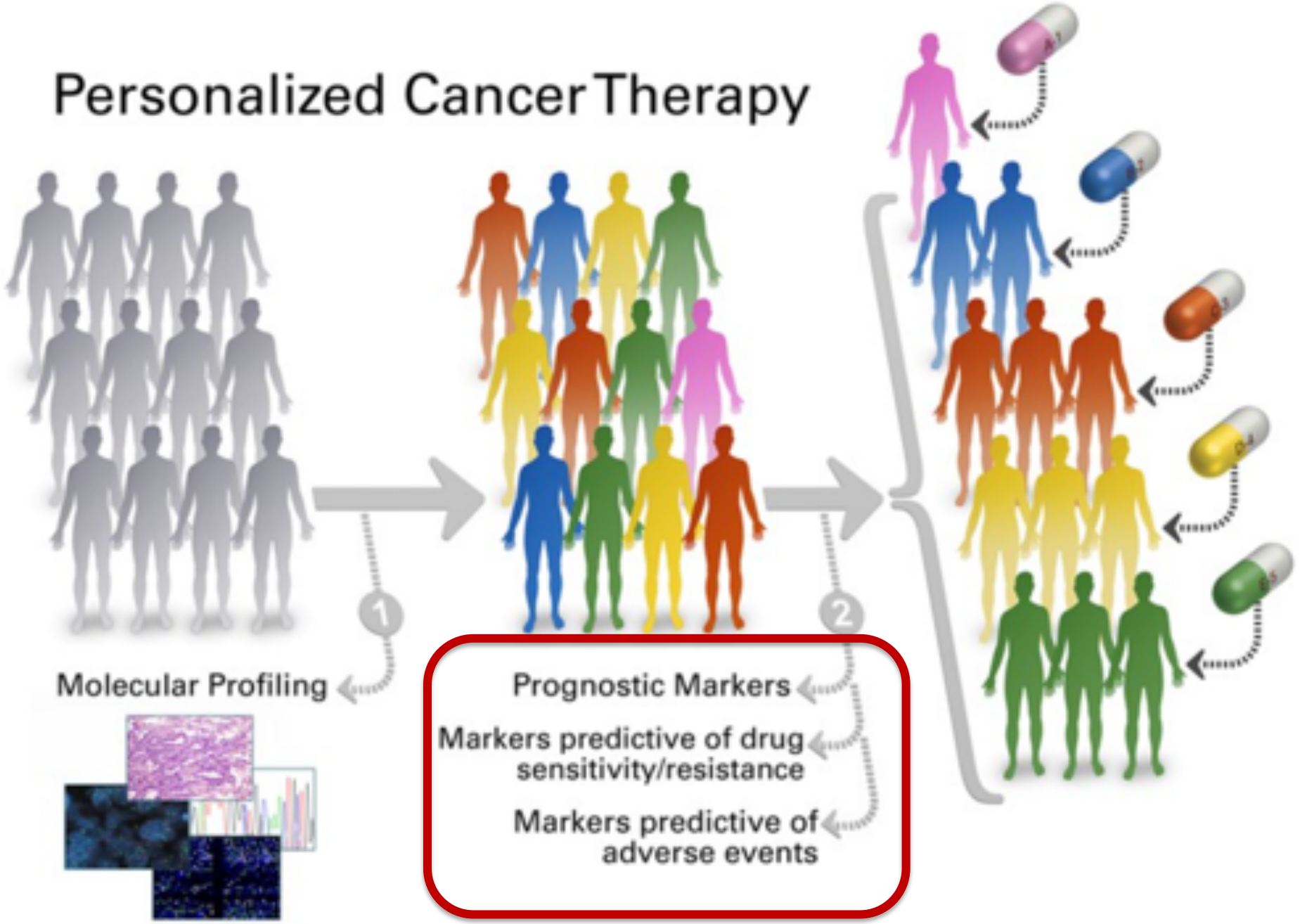
Fig. 2. Heatmap for normalized IC50 values of 75 drugs (columns) on 624 cell lines (rows). Green means the most sensitive, red means the most resistant.

Green means Sensitive

Red means Resistant

IEEE_J_Biomed_Health_Inform_2015_Sheng.pdf

Personalized Cancer Therapy

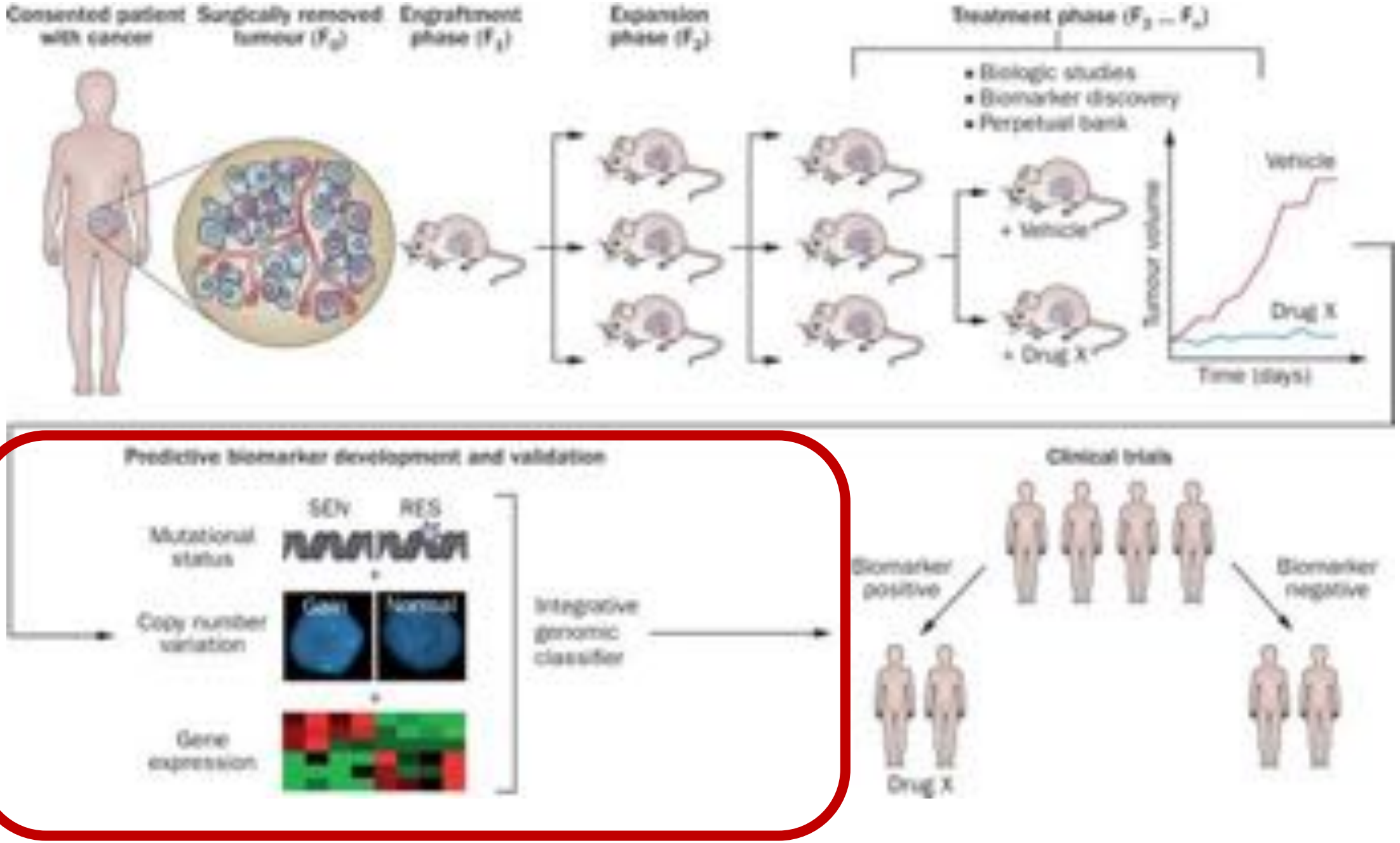


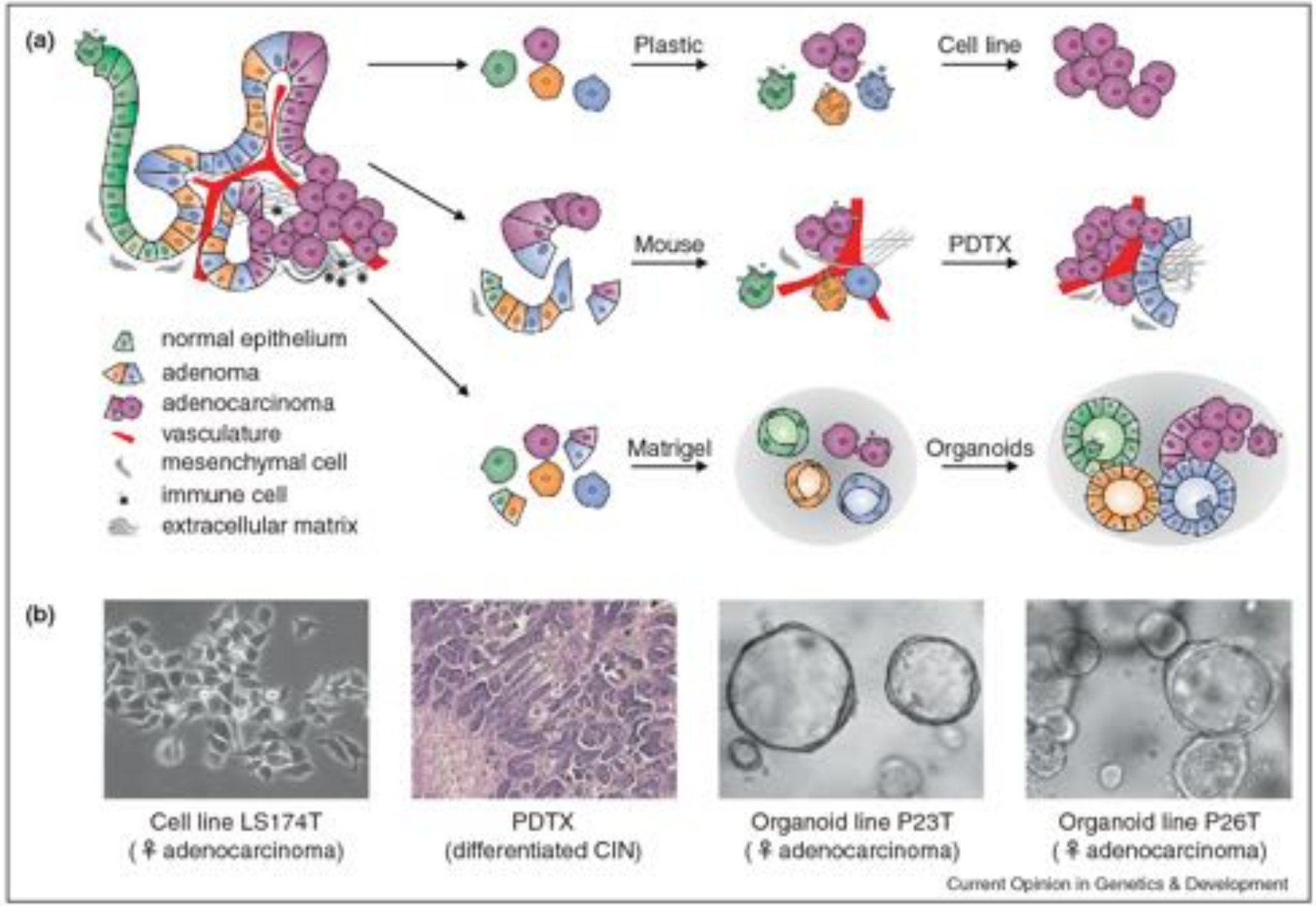
PDX models to the rescue

Patient-derived xenograft (PDX) models, however, offer the potential to transform translational oncology [5]. PDX models are created by transplanting fresh tumour tissues from human patients directly into mice.

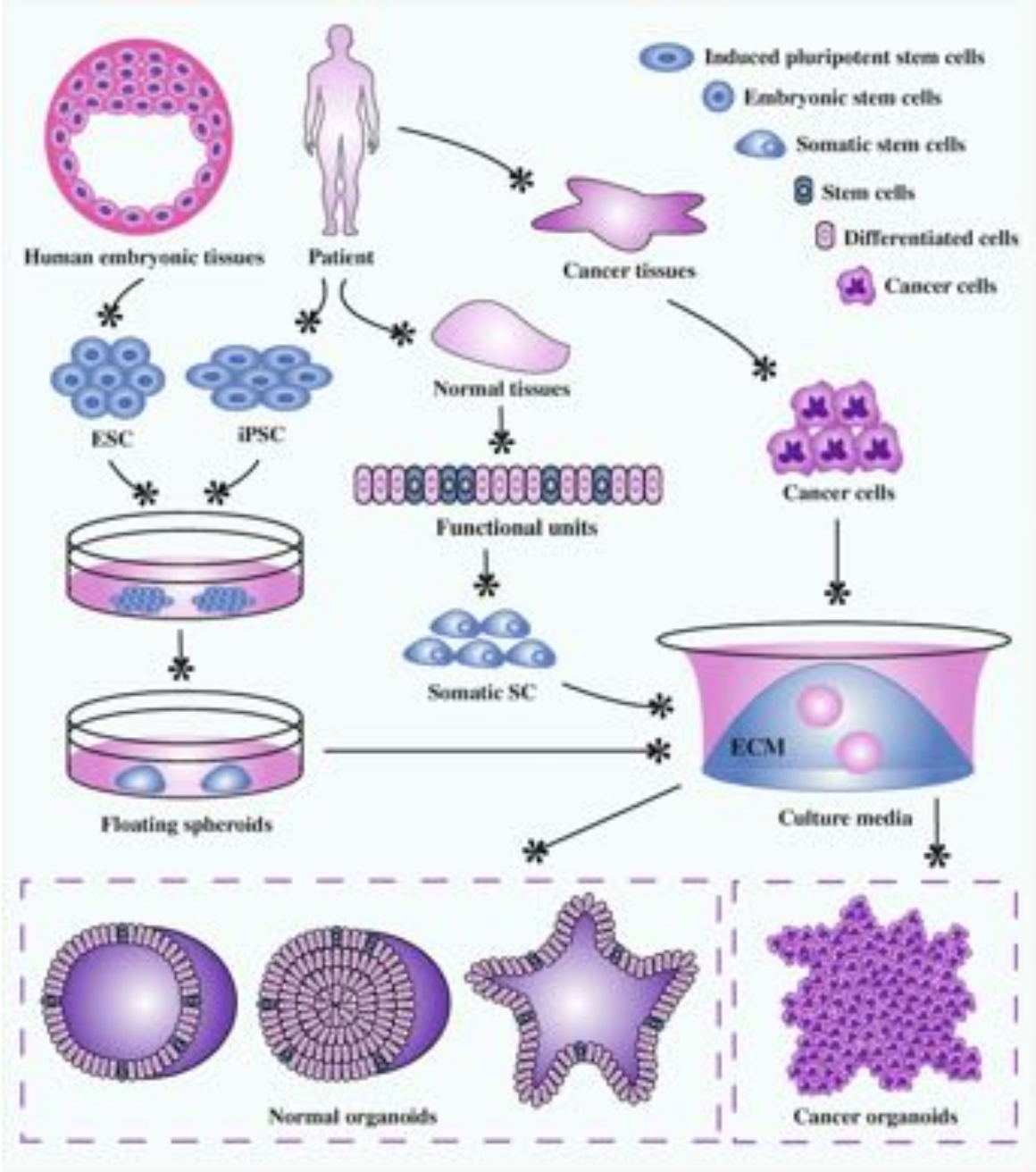
How PDX models are made and characterised

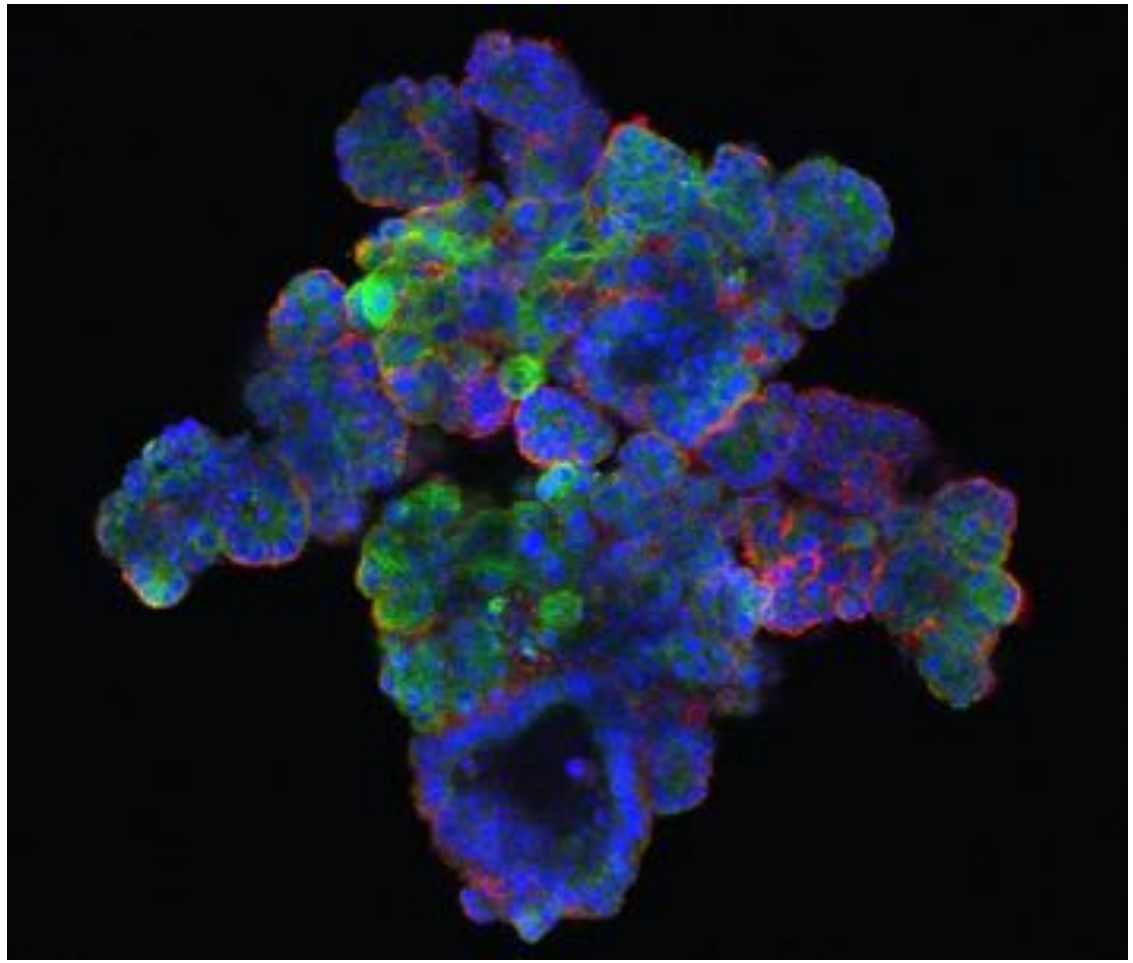
PDX tumours have been shown to preserve key features of a specific cancer and these models have been shown to be predictive of clinical outcomes – unlike cell culture models. Furthermore, PDX models have been used to predict biomarkers of drug susceptibility and drug resistance, which is crucial for clinical trial phases of development where multiple drugs fail. They are becoming the preferred preclinical tool to improve the drug development process.


Consented patient with cancer
Surgically removed tumour (F_0)
Engraftment phase (F_1)
Expansion phase (F_2)
Treatment phase (F_3 – F_n)
• Biologic studies
• Biomarker discovery
• Perpetual bank
+ Vehicle
+ Drug X
Vehicle
Drug X
Tumour volume
Time (days)
Predictive biomarker development and validation
SEN
RES
Mutational status
Copy number variation
Gain
Normal
Gene expression
Integrative genomic classifier
Clinical trials
Biomarker positive
Biomarker negative
Drug X


(a)
Plastic
Cell line
Mouse
PDTX
Matrigel
Organoids
normal epithelium
adenoma
adenocarcinoma
vasculature
mesenchymal cell
immune cell
extracellular matrix
(b)
Cell line LS174T
(♀ adenocarcinoma)
PDTX
(differentiated CIN)
Organoid line P23T
(♀ adenocarcinoma)
Organoid line P26T
(♀ adenocarcinoma)
Current Opinion in Genetics & Development

Cancer Organoids

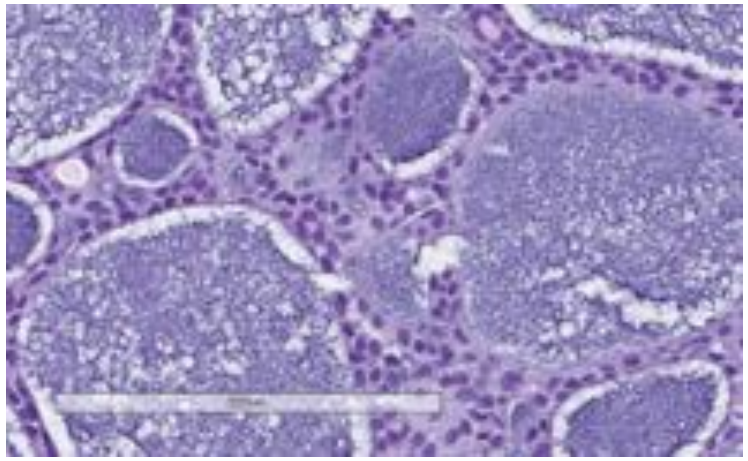




Overarching Goal

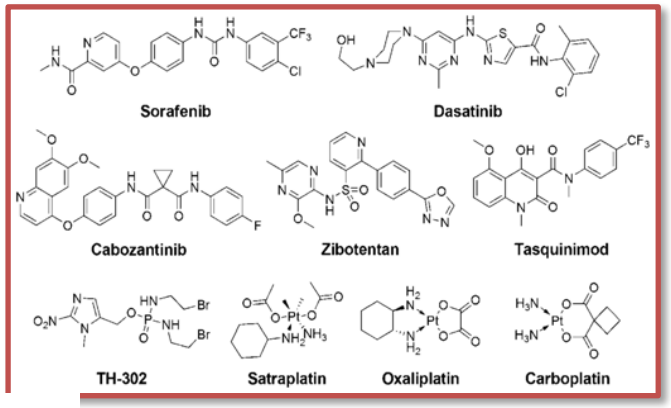
A single ML model trained on data from many cancer samples and many drugs that can predict drug response across wide range of tumors and drug combinations

Modeling Cancer Drug Response

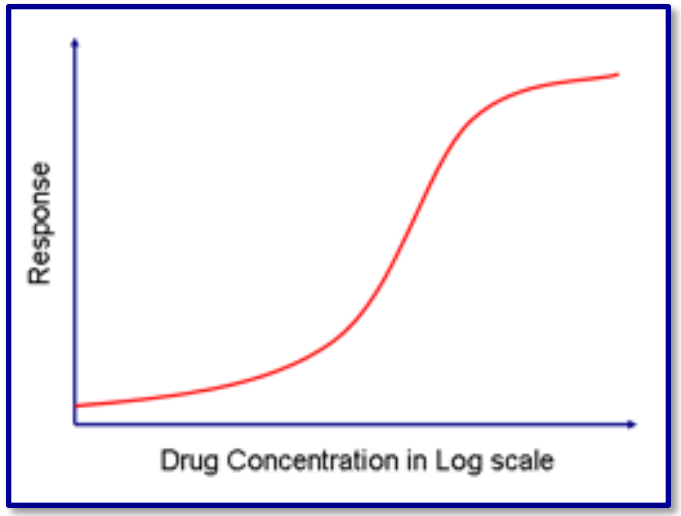


Drug (s)

- descriptors
- fingerprints
- structures
- SMILES
- dose



$$\mathcal{R} = f(\mathcal{T}, \mathcal{D}_1, \mathcal{D}_2)$$

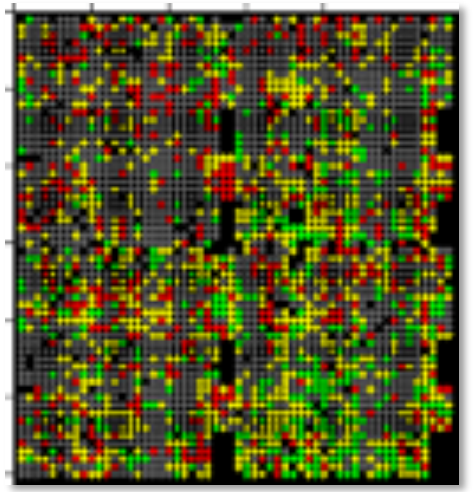


Response

- IC50
- AUC
- GI50
- % growth
- Z-score

Tumor

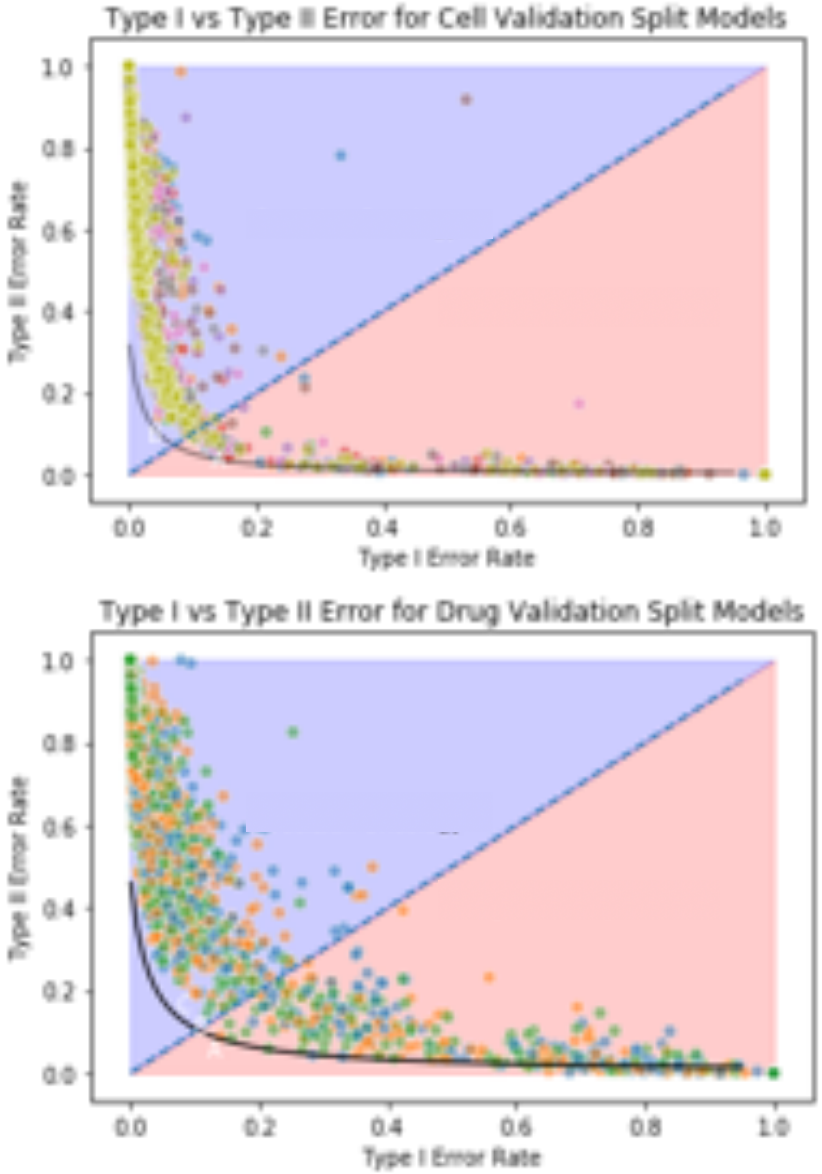
- gene expression levels
- SNPs
- protein abundance
- microRNA
- methylation



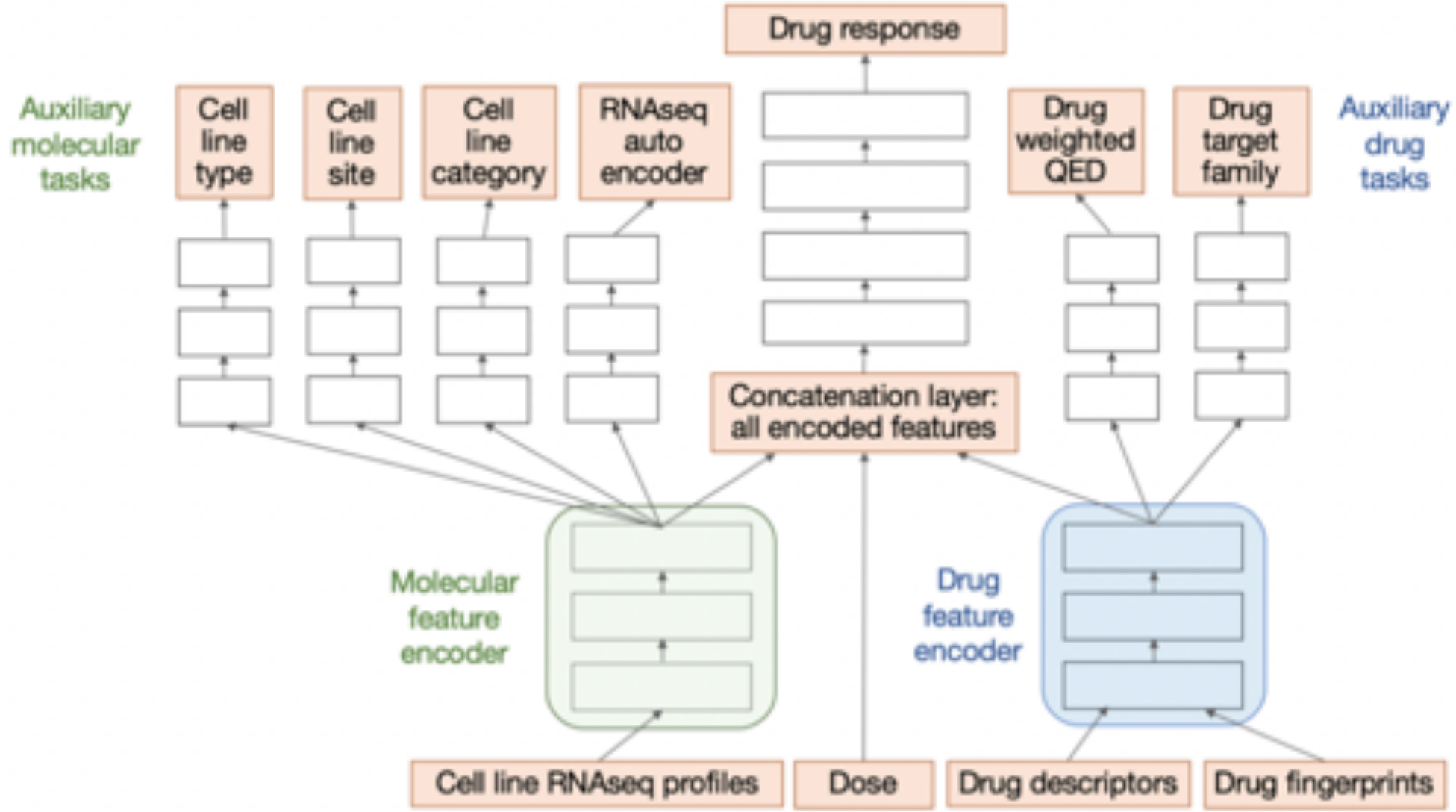
Two General Use Cases for Models

- **Predictive Oncology**
  - Predicting outcomes of experiments or patient treatments
  - Fixed Drugs .. Sweep tumors
- **Drug Discovery**
  - In silico screening novel compounds for activity
  - Fixed Tumors .. Sweep drugs

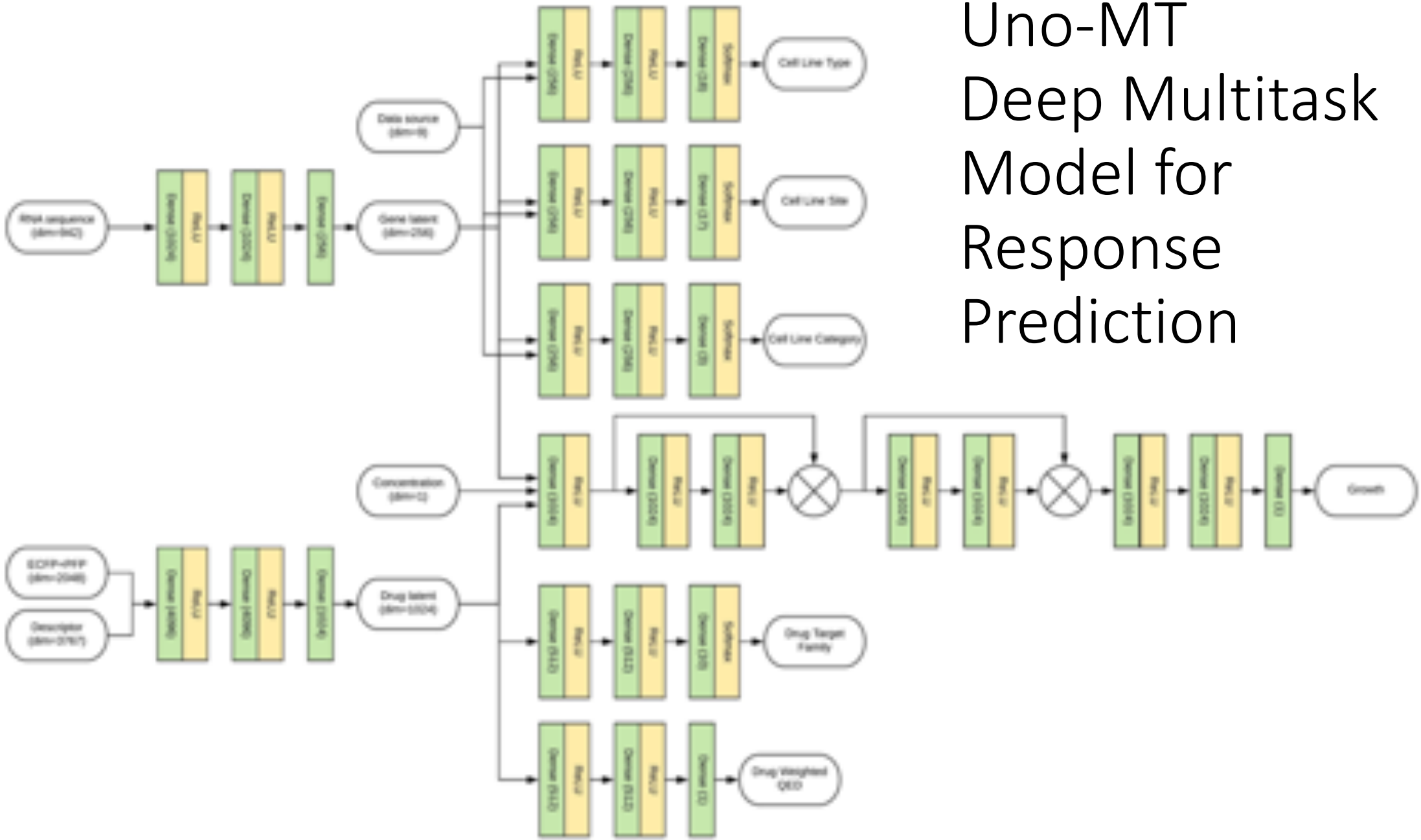
- Validation strategies different
- The models are subtly different
- We can tune models for each case



Uno-MT


Drug response
Auxiliary molecular tasks
Cell line type
Cell line site
Cell line category
RNAseq auto encoder
Drug weighted QED
Drug target family
Auxiliary drug tasks
Concatenation layer: all encoded features
Molecular feature encoder
Drug feature encoder
Cell line RNAseq profiles
Dose
Drug descriptors
Drug fingerprints

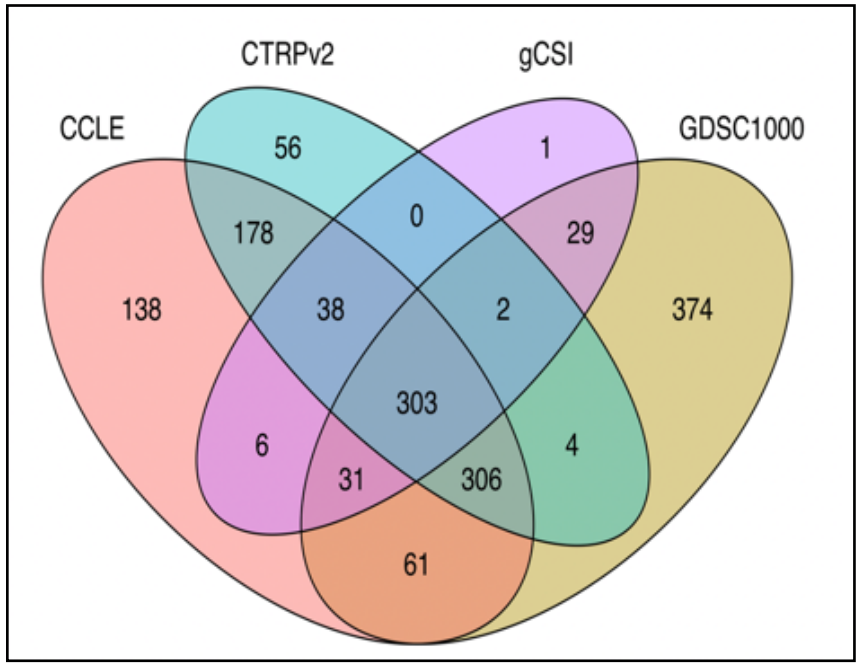
Uno-MT Deep Multitask Model for Response Prediction



RANK	DI COUNT	TYPE
1	58051	Skin_Cutaneous_Melanoma
2	56693	Colon_Adenocarcinoma
3	56534	Lung_Adenocarcinoma
4	40203	Kidney_Renal_Clear_Cell_Carcinoma
5	38921	Ovarian_Serous_Cystadenocarcinoma
6	36595	Breast_Invasive_Carcinoma
7	35932	Lymphoid_Leukemia
8	27292	Glioblastoma_Multiforme
9	25044	Lung_Small_Cell_Carcinoma
10	20226	Lung_Non-Small_Cell_Carcinoma
11	19751	Sarcoma
12	16862	Pancreatic_Adenocarcinoma
13	16086	Brain_NOS
14	15985	Acute_Myeloid_Leukemia
15	15980	Lung_Squamous_Cell_Carcinoma
16	15912	Acute_Lymphoblastic_Leukemia
17	14725	Head_and_Neck_Squamous_Cell_Carcinoma
18	13719	Uterine_Corpus_Endometrial_Carcinoma
19	13153	Esophageal_Carcinoma
20	13060	Lymphoid_Neoplasm_Diffuse_Large_B-cell_Lymphoma
21	12744	Ovary_NOS

DI $\Longrightarrow$ dose independent

+ NCI60



Can we build models that are predictive of drug response?

Dose Independent, Top 6. Top21, cancers, Attention MLP (Means from 10-fold CV)

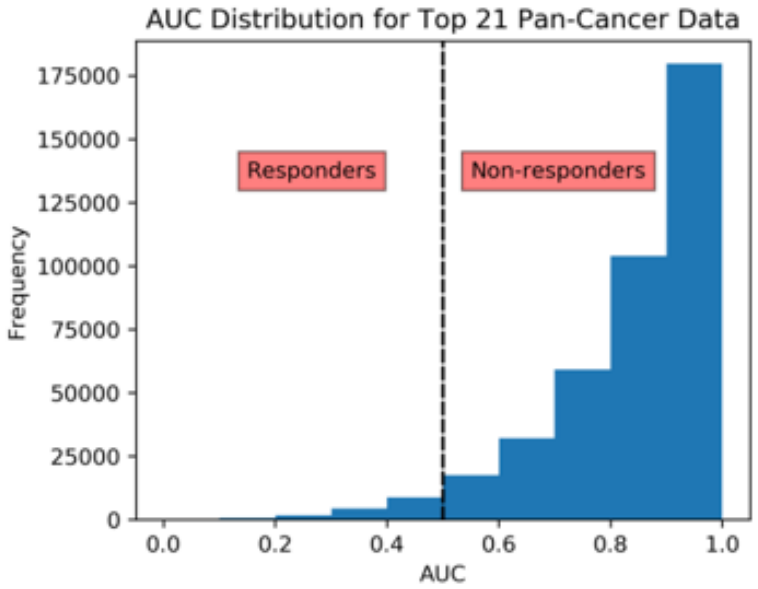
Top 6 Cancer Types

Precision	Recall	f1-score	
0.917	0.790	0.837	Mordred, Lincs1000 (bin.3)
0.933	0.853	0.882	Dragon7, Lincs1000 (bin.3)
0.933	0.855	0.884	Dragon7, Lincs1000 (bin.1)

Top 21 Cancer Types

Precision	Recall	f1-score	
0.95	**0.927**	**0.935**	Dragon7, Lincs1000 (bin.3)

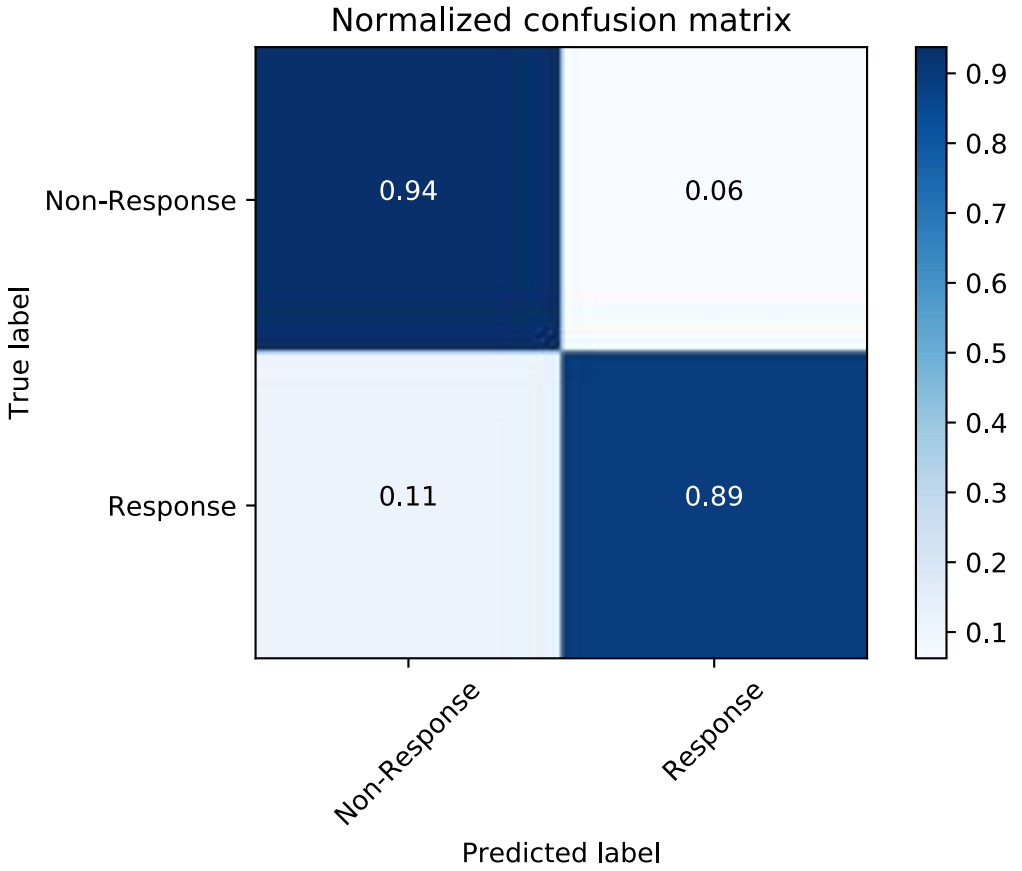
(~6,200 features)



Single Drug Response

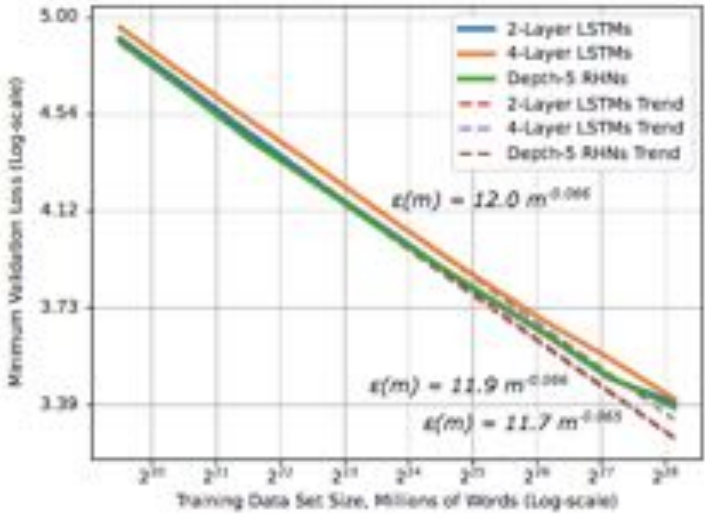
Drug	R^2	MAE	AUC	Accuracy
Afatinib	0.4369	0.0737	0.9248	**0.9679**
Bortezomib	0.3871	0.0752	**0.9429**	0.9569
Docetaxel	**0.5748**	0.1154	0.9158	0.8853
Doxorubicin	0.3749	0.1103	0.7794	0.7105
Etoposide	0.3787	0.1108	0.8855	0.8768
GDC-0941	0.3294	0.0744	0.6924	0.9478
Navitoclax	0.4329	0.0982	0.9035	0.9295
Paclitaxel	0.5299	0.1285	0.8471	0.7626
Selumetinib	0.2944	0.1056	0.8831	0.9115
SN-38	0.3415	0.1150	0.8269	0.8361
Temsirolimus	0.2048	0.1136	0.7406	0.8912
Tipifarnib	0.3187	0.1115	0.8474	0.8981
Vinorelbine	0.1407	0.1289	0.7605	0.8367
Vorinostat	0.4041	**0.0627**	0.9134	0.9532
mean	**0.3678**	**0.1017**	**0.8474**	**0.8832**

Top 21 Cancer Types in MD DI formulation



Models are best of RF, LGB, GB, LR, etc.; features are RNAseq and D7 descriptors

Learning Curve Power Law



NLP Learning Curves

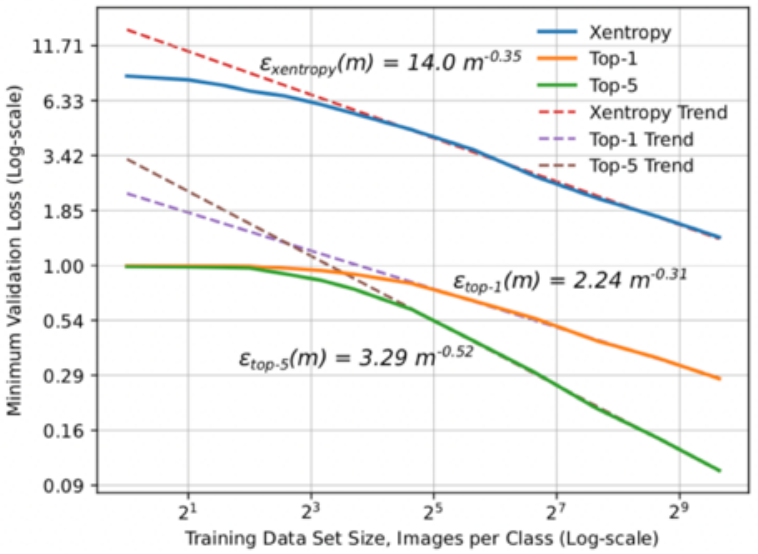
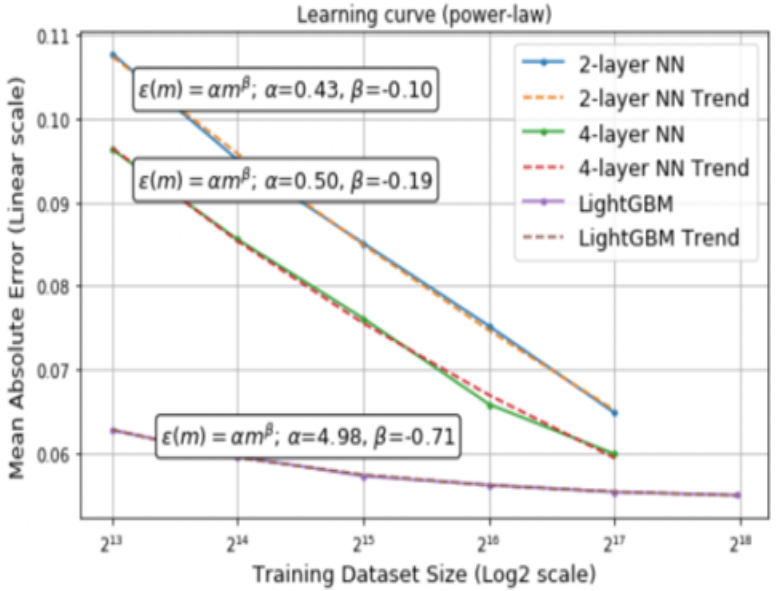


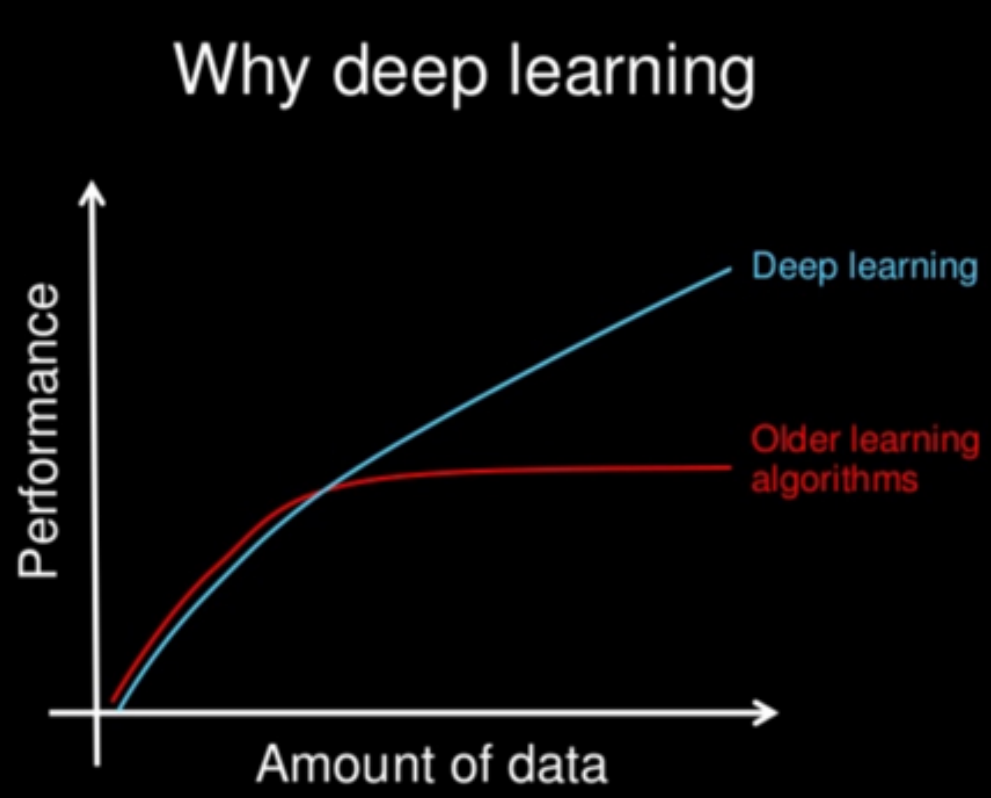
Image Classification

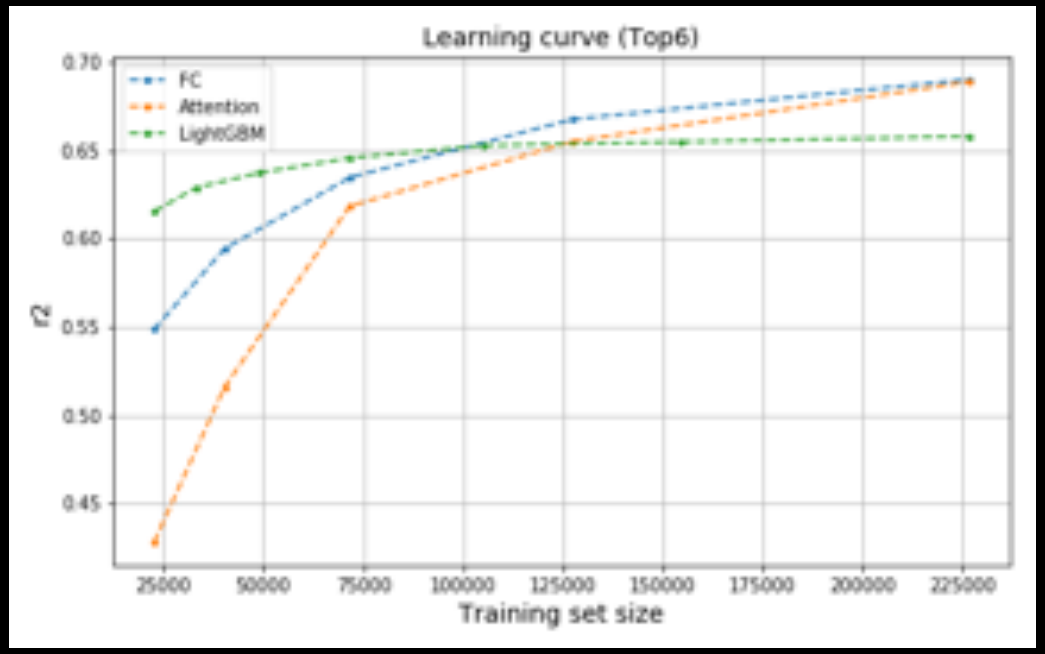


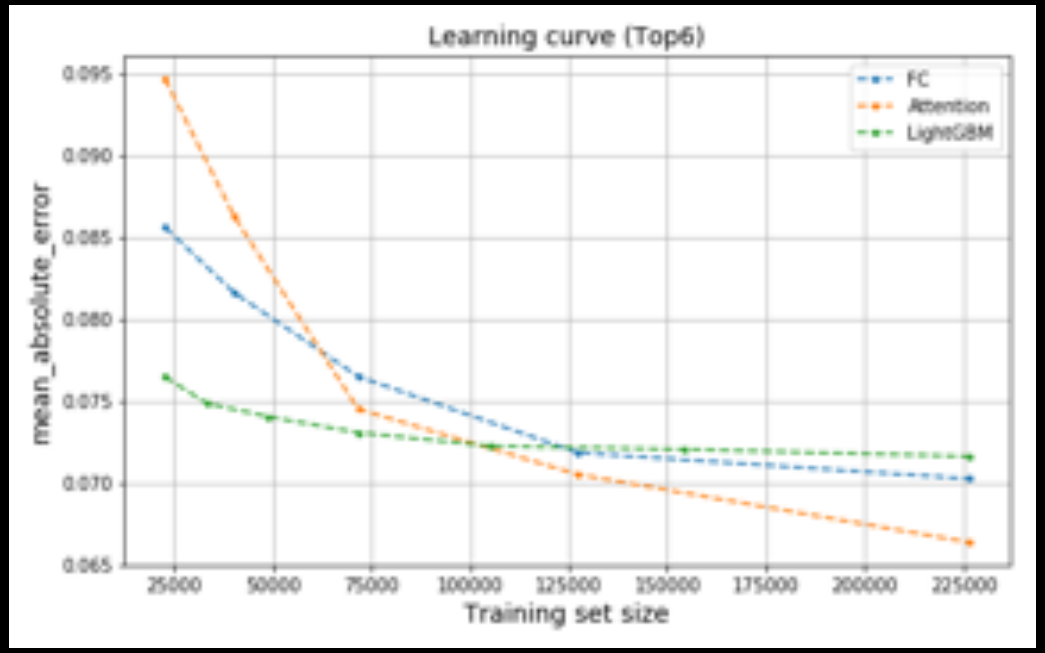
Top6 Cancer Response

It seems that the advent of models that beat the power-law exponent — that get **more data efficient as they learn —** might be an important empirical milestone on that path.

https://arxiv.org/pdf/1712.00409.pdf





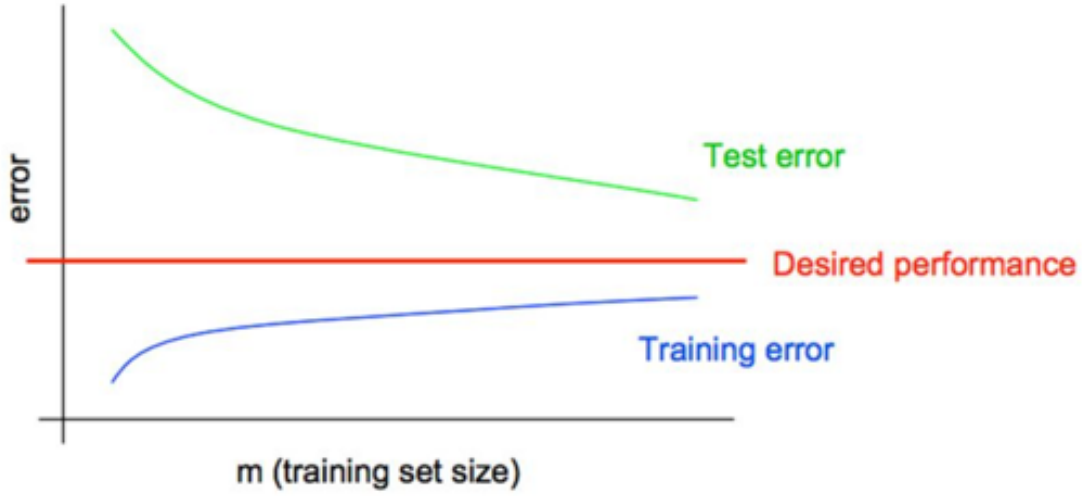


More training data ⟹ Lower Error

Learning Curves – Variance and Bias

Learning curve – Variance

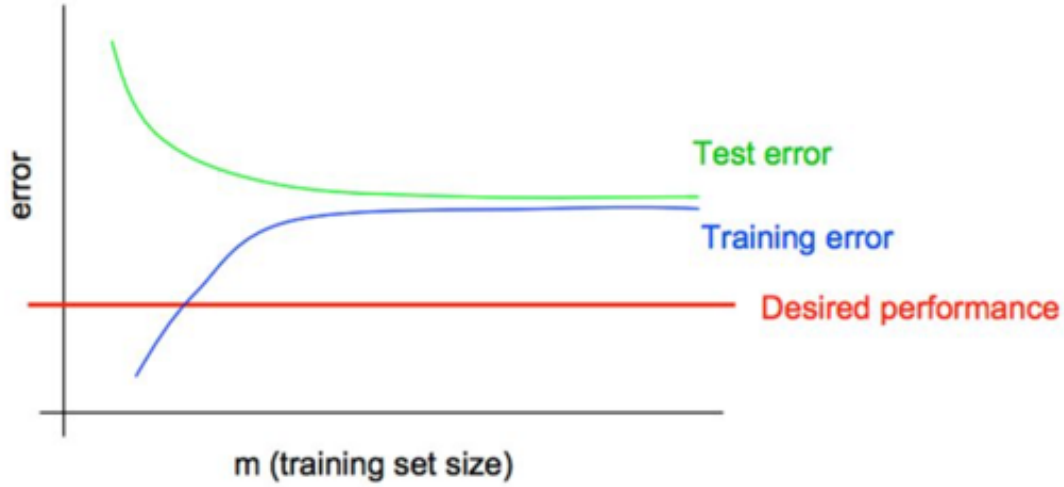
$$h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{100} x^{100}$$



Do we have enough data?

Learning curve – Bias

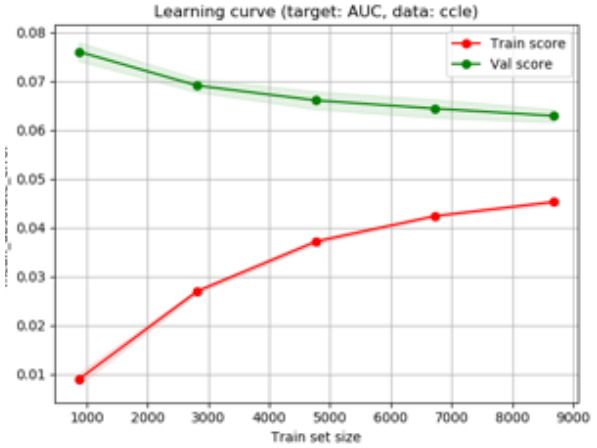
$$h_\theta(x) = \theta_0 + \theta_1 x$$

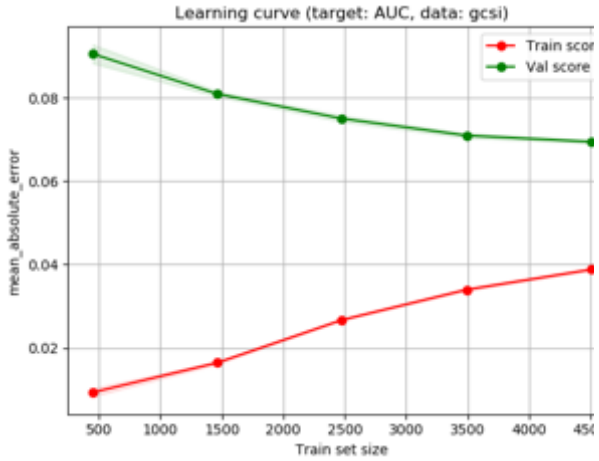


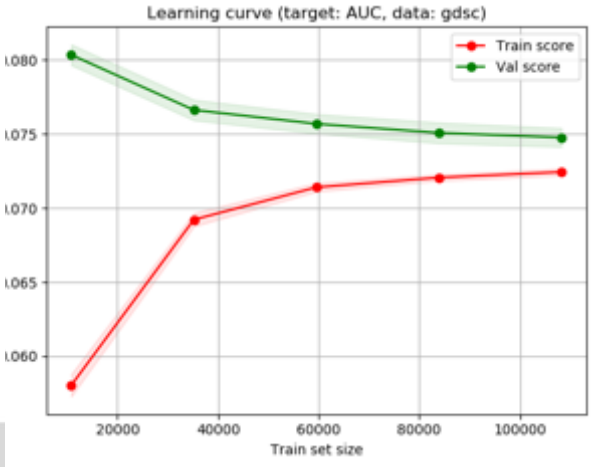
Do we have the right data?

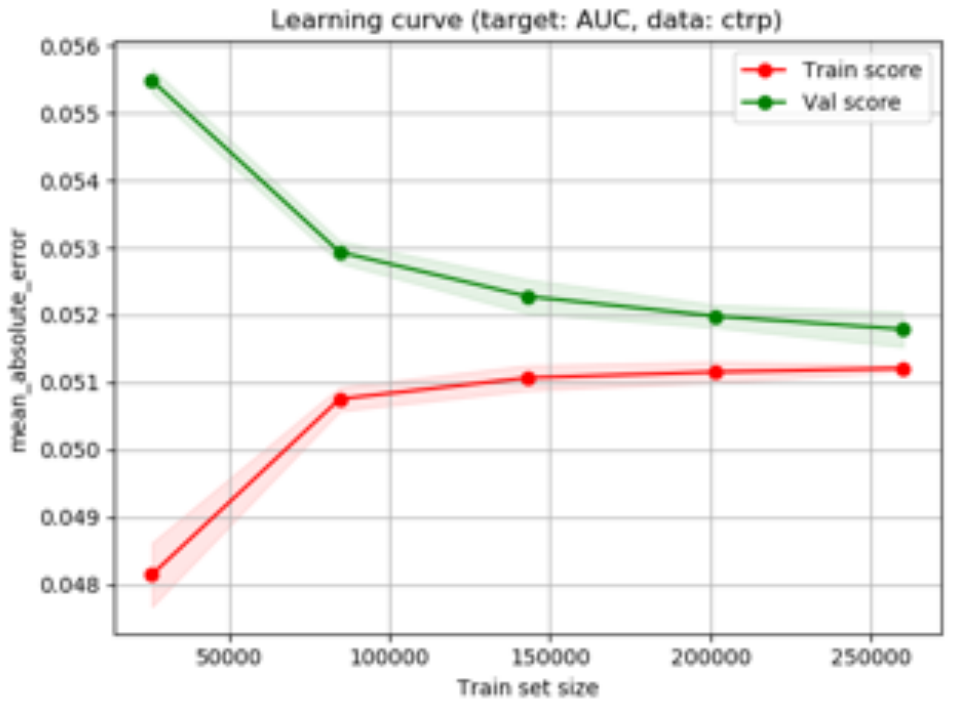
Learning Curves

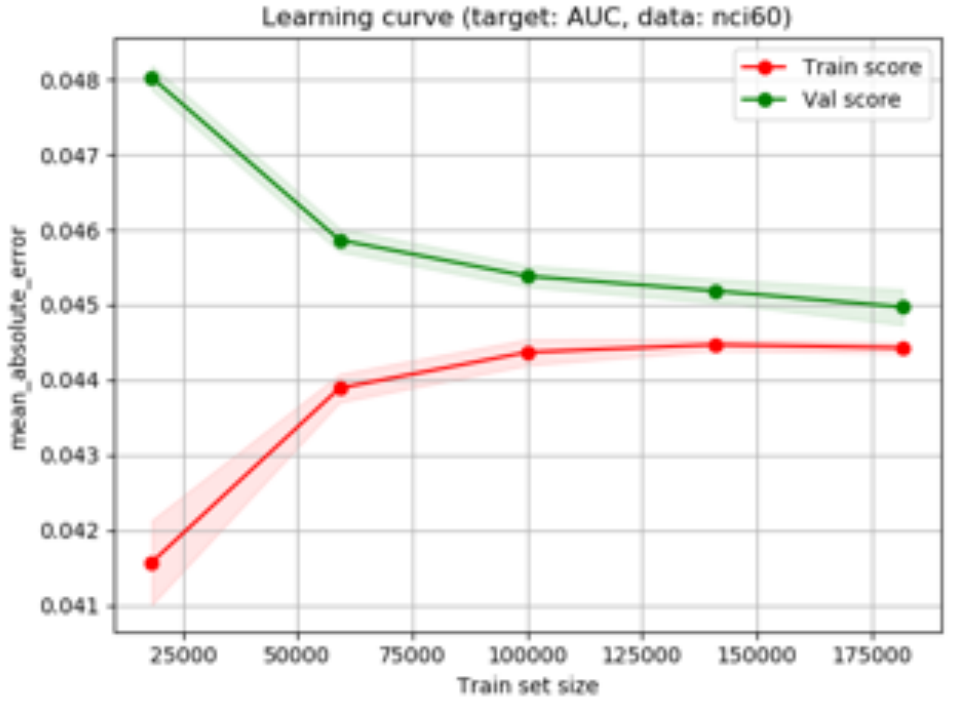
MAE











Can we build models that generalize across studies?

UnoMT Multitask Deep Learning Cross-Study

Best out of Study $R^2 = 0.61$

Table 6. Best cross study validation results with a 3-task UnoMT

		Testing set							
		NCI60	CTRP	GDSC	CCLE	gCSI	N/T Cat Acc	Site Acc	Type Acc
Training set	NCI60	R2 = 0.81 MAE = 17.1	**R2 = 0.38 MAE = 32.2**	**R2 = 0.24 MAE = 35.3**	**R2 = 0.48 MAE = 33.4**	**R2 = 0.46 MAE = 33.4**	**99.43%**	**96.75%**	**96.97%**
	CTRP	**R2 = 0.44 MAE = 29.8**	R2 = 0.68 MAE = 22.7	**R2 = 0.23 MAE = 34.4**	**R2 = 0.61 MAE = 28.3**	**R2 = 0.60 MAE = 28.5**	**99.56%**	96.62%	96.58%
	GDSC	R2 = 0.32 MAE = 34.0	**R2 = 0.25 MAE = 36.7**	R2 = 0.53 MAE = 27.2	R2 = 0.50 MAE = 32.6	**R2 = 0.60 MAE = 29.2**	**99.43%**	**96.93%**	**96.97%**
	CCLE	**R2 = 0.27 MAE = 36.9**	**R2 = 0.20 MAE = 39.2**	**R2 = 0.11 MAE = 38.9**	R2 = 0.68 MAE = 25.4	R2 = 0.39 MAE = 34.2	**99.12%**	**96.36%**	**96.36%**
	gCSI	R2 = 0.00 MAE = 44.9	**R2 = 0.11 MAE = 43.1**	R2 = 0.05 MAE = 42.8	**R2 = 0.33 MAE = 40.6**	R2 = 0.80 MAE = 192	**99.43%**	**96.84%**	**96.62%**

MAE = Mean Absolute Error (in percent growth)

Why is NCI-60 better in the diagonals?

source_scale loss=MSE		Testing set				
		NCI60	CTRP	GDSC	CCLE	gCSI
Training set	NCI60	R2 = 0.80 MAE = 18.0	R2 = 0.36 MAE = 32.9	R2 = 0.21 MAE = 36.2	R2 = 0.45 MAE = 34.6	R2 = 0.42 MAE = 35.6
	CTRP	R2 = 0.39 MAE = 31.3	R2 = 0.67 MAE = 23.3	R2 = 0.19 MAE = 35.4	R2 = 0.58 MAE = 29.4	R2 = 0.56 MAE = 29.3
	GDSC	R2 = 0.31 MAE = 35.0	R2 = 0.23 MAE = 36.9	R2 = 0.52 MAE = 27.6	R2 = 0.51 MAE = 27.6	R2 = 0.57 MAE = 30.1
	CCLE	R2 = -0.03 MAE = 46.0	R2 = 0.02 MAE = 44.8	R2 = -0.04 MAE = 43.6	R2 = 0.67 MAE = 25.6	R2 = 0.46 MAE = 33.5
	gCSI	R2 = -0.07 MAE = 46.0	R2 = -0.01 MAE = 45.6	R2 = -0.06 MAE = 45.7	R2 = 0.30 MAE = 39.9	R2 = 0.78 MAE = 20.01

Variability in Replicates Sets the R^2 Ceiling

Table 1: Characteristics of drug response datasets included in the cross-study analysis

Dataset	Cells	Drugs	Dose Response Samples	Measured Response Groups	Viability Assay
NCI60	60	52,671	18,862,308	3,780,148	CellTiter Glo
CTRP	887	544	6,171,005	395,263	CellTiter Glo
CCLE	504	24	93,251	11,670	CellTiter Glo
GDSC	1,075	249	1,894,212	225,480	Syto60
gCSI	409	16	58,094	6,455	CellTiter Glo

Table 2: Dose response variability among replicates in the same study

Study	Samples with Replicates	Replicates per Group	Mean Response S.D. in Group	R^2 Explaining Response with Group Mean	R^2 for Samples with Replicates
NCI60	41.56%	2.62	0.145	0.959	0.931
CTRP	4.09%	2.05	0.188	0.996	0.862
GDSC	2.62%	2.00	0.219	0.996	0.810

Why is Model Transfer from CTRP => GDSC so Poor?

source_scale loss=MSE		Testing set				
		NCI60	CTRP	GDSC	CCLE	gCSI
Training set	NCI60	R2 = 0.80 MAE = 18.0	R2 = 0.36 MAE = 32.9	R2 = 0.21 MAE = 36.2	R2 = 0.45 MAE = 34.6	R2 = 0.42 MAE = 35.6
	CTRP	R2 = 0.39 MAE = 31.3	R2 = 0.67 MAE = 23.3	R2 = 0.19 MAE = 35.4	R2 = 0.58 MAE = 29.4	R2 = 0.56 MAE = 29.3
	GDSC	R2 = 0.31 MAE = 35.0	R2 = 0.23 MAE = 36.9	R2 = 0.52 MAE = 27.6	R2 = 0.51 MAE = 27.6	R2 = 0.57 MAE = 30.1
	CCLE	R2 = -0.03 MAE = 46.0	R2 = 0.02 MAE = 44.8	R2 = -0.04 MAE = 43.6	R2 = 0.67 MAE = 25.6	R2 = 0.46 MAE = 33.5
	gCSI	R2 = -0.07 MAE = 46.0	R2 = -0.01 MAE = 45.6	R2 = -0.06 MAE = 45.7	R2 = 0.30 MAE = 39.9	R2 = 0.78 MAE = 20.01

Cross-study Response Consistency

- CTRP and CCLE use the same viability assay
- GDSC uses a different one: the inconsistency is well documented

- Different dose-independent aggregation metrics work differently (R^2 in the table is based on direct mapping)

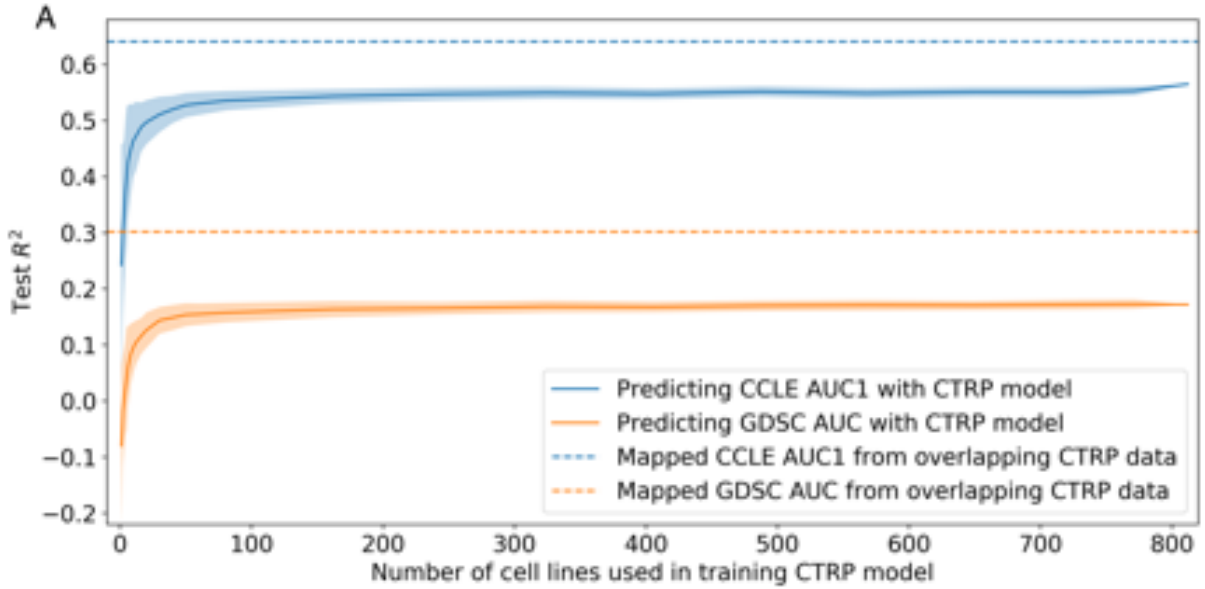
Table 3: Cross-study response consistency in identical cell-drug pairs

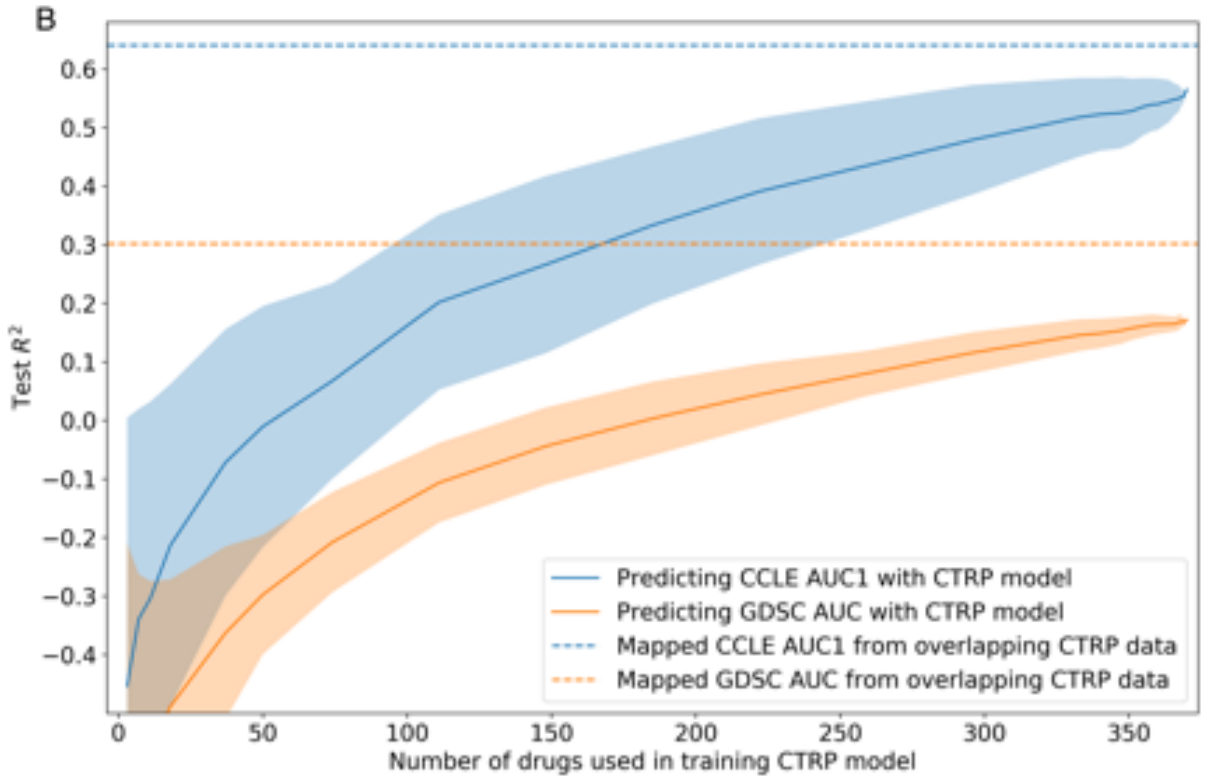
Source Study	Target Study	Overlapping Cell-Drug Groups	R2 on AUC	R2 on AUC1	R2 on DSS
CTRP	CCLE	2,338	0.5944	0.6408	0.6347
CTRP	GDSC	17,259	0.3016	0.0186	0.0062

Is drug diversity more important than cell diversity?

Apparently yes.

What does this mean for PDM and PDO experiments?





Comparison on PDX Prediction Performance With and Without Transfer Learning

Analysis name	R^2	P-value (R^2)	Spearman rank correlation coefficient	P-value (Spearman rank correlation coefficient)
PDX-Only	0.064(0.031)		0.372(0.022)	
CCLE-TL	0.042(0.016)	8.01E-02	0.355(0.013)	7.28E-02
gCSI-TL	0.100(0.016)	8.29E-03	0.389(0.017)	7.55E-02
NCI60-TL	0.102(0.013)	5.16E-03	0.407(0.016)	1.43E-03
CTRP-TL	0.092(0.019)	3.35E-02	0.415(0.013)	1.51E-04
GDSC-TL	0.110(0.017)	1.50E-03	0.419(0.013)	7.22E-05

PDX-only is the analysis without transfer learning. -TL in analysis name indicates transfer learning from a CCL dataset.

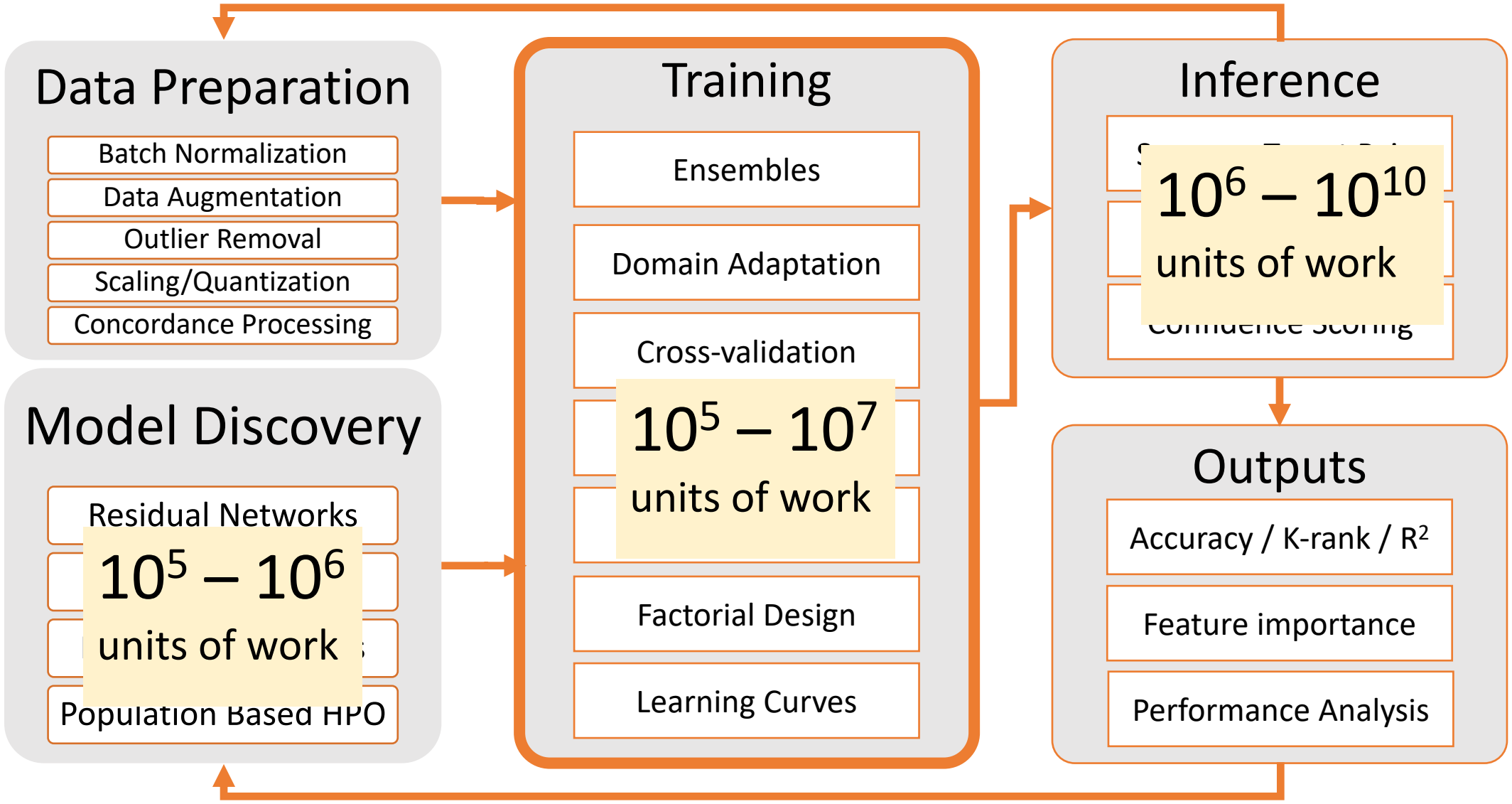
- Mean (standard deviation) of prediction performance is evaluated through 10 times of 10-fold cross-validations on PDXs
- Four out of the five transfer learning analyses show a prediction performance statistically significantly better than that of PDX-only analysis, evaluated by the p-value of t-test ≤ 0.05

CANDLE Project

- **ECP-CANDLE GitHub Organization:**
- https://github.com/ECP-CANDLE

- **CANDLE Python Library** – make it easy to run on DOE Big Machines, scale for HPO, UQ, Ensembles, Data Management, Logging, Analysis
- **CANDLE Benchmarks** – exemplar codes/models and data representing the three primary challenge problems
- **Runtime Software** – Supervisor, Reporters, Data Management, Run Data Base
- **Tutorials** – Well documented examples for engaging the community
- **Contributed Codes** – Examples outside of Cancer, including Climate Research, Materials Science, Imaging, Brain Injury
- **Frameworks** – Leverage of TensorFlow, Keras, Horovod, PyTorch, etc.
- **LL Libraries** – CuDNN, MKL, etc. (tuned to DOE machines)

Scope of CANDLE workflows



How are we using Large-Scale Computing

- Deep Sweeps on Features/Feature Combinations
  - Recently ran 16K model jobs on Summit
- Hyperparameter Optimization
  - Tuning model settings (Big runs on Cori, Theta, Summit, Titan)
- Neural Architecture Search (Model Discovery)
  - Big runs on Theta (SC19 Paper)
- Hierarchical Cross Validation Study > 500K models
  - Bayesian approach to online learning (accelerated convergence)
- Data Augmentation and Generation Networks
  - Exploring strategies for “Low Data” learning
- Uncertainty Quantification
  - Bootstrapping, parameter sweeps

Acknowledgements

Material in this talk comes from Argonne, Oak Ridge, Livermore and Los Alamos national laboratories, National Cancer Institute, Frederick National Laboratory for Cancer Research, Computation Institute, and the University of Chicago

Special thanks to:

Gina Tourassi (Oak Ridge)

Fred Streitz,

Jonathan Allen (Livermore),

Frank Alexander (Brookhaven),

Marian Anghel (Los Alamos),

Eric Stahlberg,

Yvonne Evrard,

Susan Holbeck,

James Doroshow (NCI)

Bob Grossman,

Monisha Gosh (UChicago)

Fangfang Xia, Maulik Shukla,

Tom Brettin,

Emily Dietrich (Argonne)

Acknowledgements

Many thanks to DOE, NSF, NIH, DOD, ANL, UC, Moore Foundation, Sloan Foundation, Apple, Microsoft, Cray, Intel and IBM for supporting my research group over the years