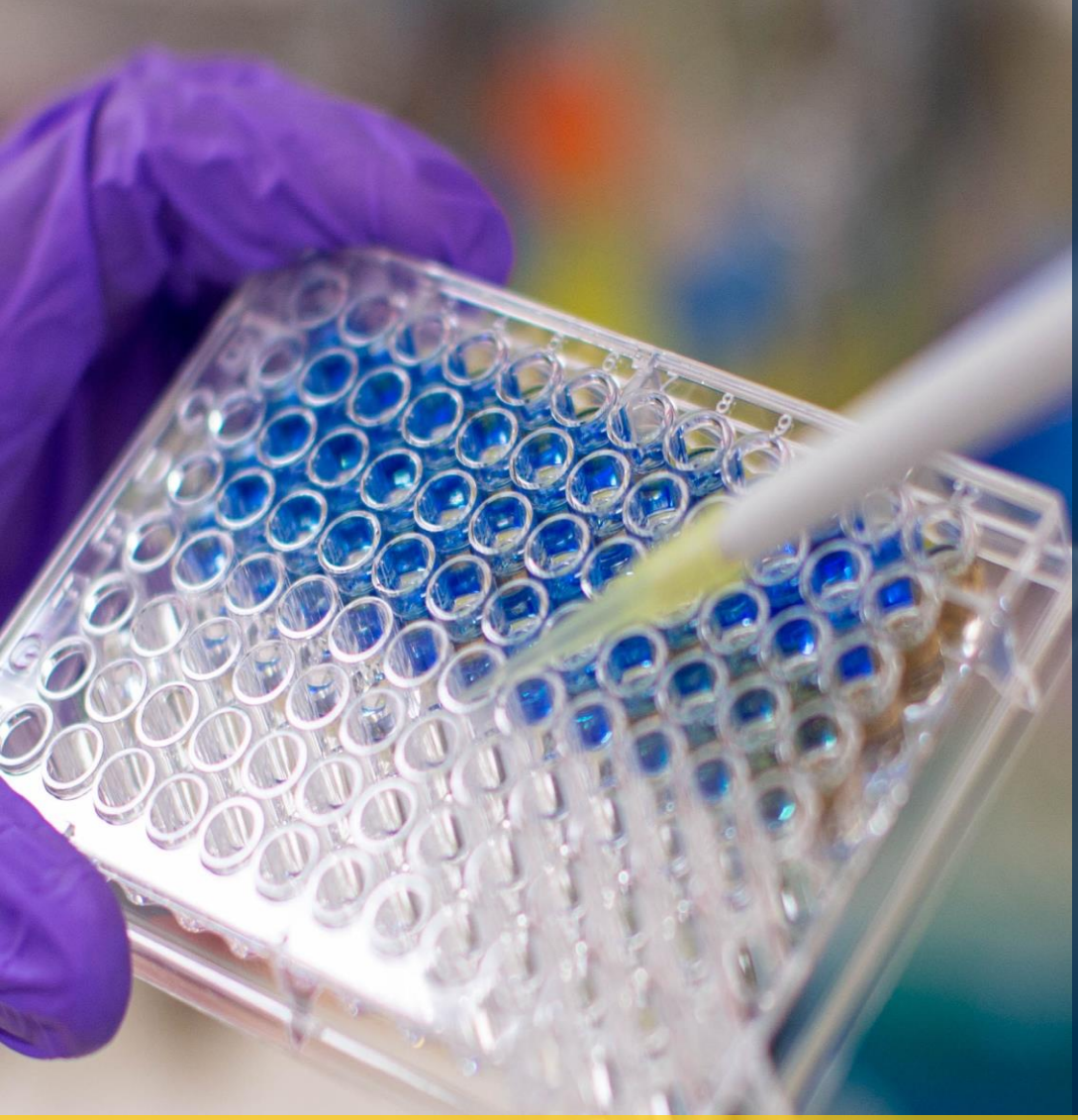


Veera Baladandayuthapani
Department of Biostatistics
Rogel Cancer Center
University of Michigan
http://bayesrx.com/

BIOSTATISTICS

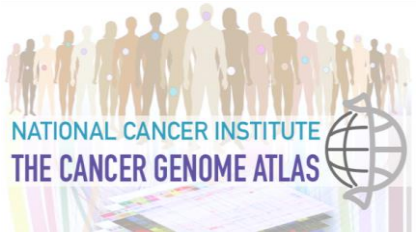
ROGEL CANCER CENTER
UNIVERSITY OF MICHIGAN HEALTH

Bayesian Models for Multi-omic Multi-system Integration

Bayesian Biostatistics Conference |Utrecht | October 26th, 2023

The Cancer Context

- Cancer is one of the **most well-characterized** path-physiological & path-biological disease systems at different molecular levels
- Multiple types of high-throughput data now available on the multiple **model systems**: Patients, Cell-lines, Patent-Derived Xenografts (PDX), Organoids… [growing day by day!]
- Motivates many **precision medicine endeavors**…



International Cancer Genome Consortium

NIH NATIONAL CANCER INSTITUTE Genomic Data Commons

PRECISION HEALTH UNIVERSITY OF MICHIGAN

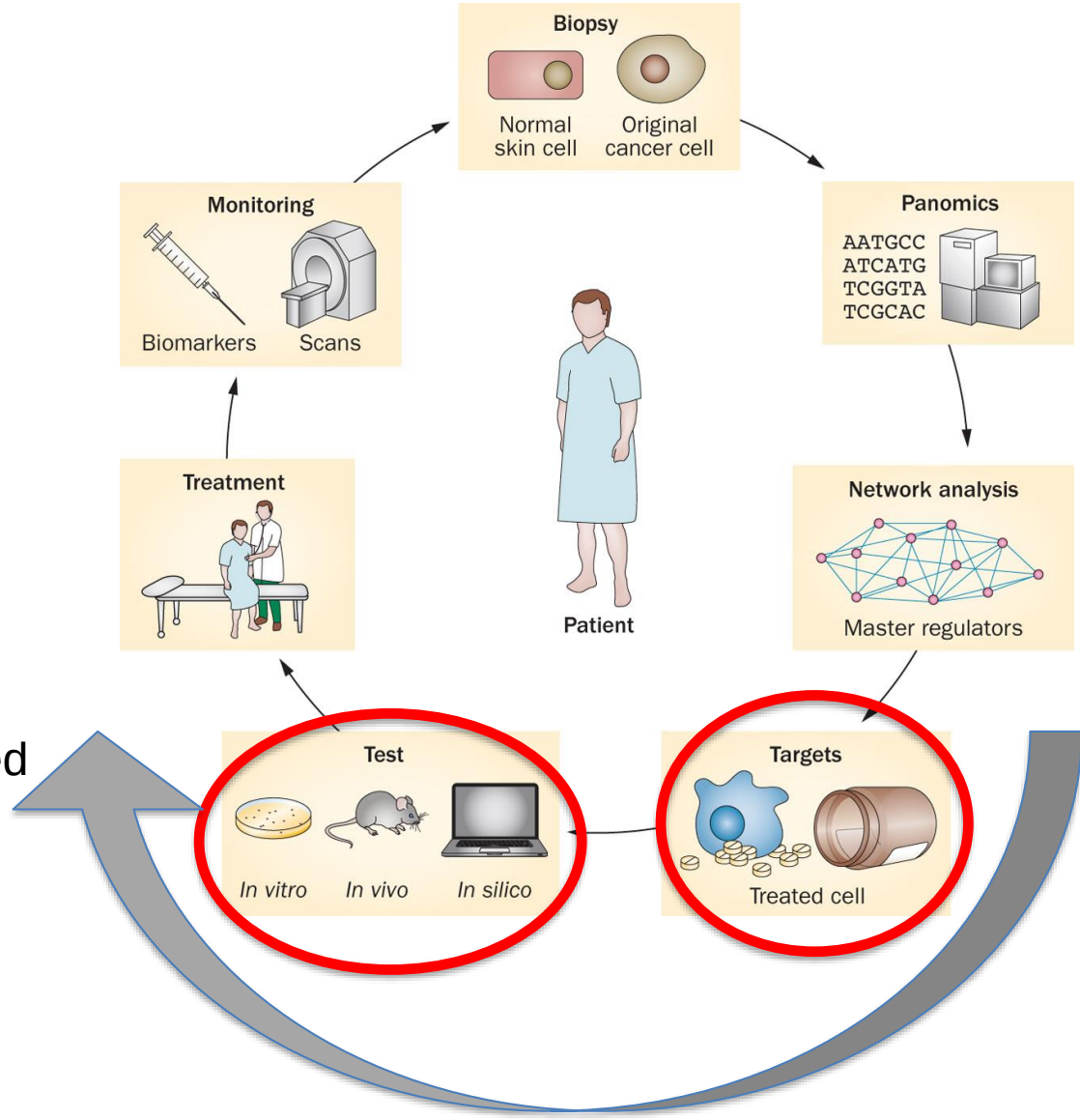
MICHIGAN GENOMICS INITIATIVE

Precision Oncology

- **Precision Oncology 1.0**
  (~5-10 years ago)
  - Small numbers of molecular abnormalities
  - Always constrained by the tissue-of-origin

- **Precision Oncology 2.0**
  (Current)
  - Dozens or 100's of possible mutational hotspots and exomes of cancer-associated genes.
  - Could be tissue-agnostic

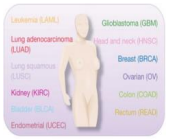
- **Precision Oncology 3.0**
  (Future)
  - Pan-omic analyses
  - Multi-system integration
  - Network analyses



Today's talk: enable discovery to translation

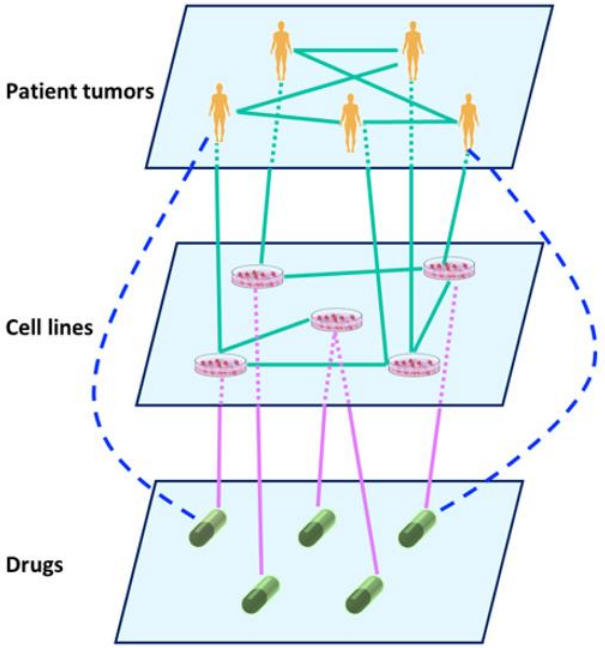
Shrager, J. & Tenenbaum, J. M. *Nat. Rev. Clin. Oncol.*

High-level Goals



Patients Tumor Systems	ICGC, TCGA, TCPA
Pre-clinical Models	CCLE, MCLP, GDSC
Drugs	NCI60, LINCS, DepMap

- Exploit the **conserved biology** between different model systems (patients, cell-lines, PDXs) to calibrate therapeutic response of drugs in patients
- Find optimal pre-clinical **"avatars"** as proxies for patients
- Identify **key genomic drivers and mechanisms** explaining model system similarities



Statistically…
- **Joint probability models across model systems to borrow strength**

The Annals of Applied Statistics
2022, Vol. 16, No. 4, 2055–2082
https://doi.org/10.1214/21-AOAS1550
© Institute of Mathematical Statistics, 2022

A BAYESIAN PRECISION MEDICINE FRAMEWORK FOR CALIBRATING INDIVIDUALIZED THERAPEUTIC INDICES IN CANCER

BY ABHISEK SAHA[1,a], MIN JIN HA[2,b], SATWIK ACHARYYA[3,c] AND VEERABHADRAN BALADANDAYUTHAPANI[3,d]

[1]*Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institute of Health,* [a]*abhisek.saha@uconn.edu*
[2]*Department of Biostatistics, Graduate School of Public Health, Yonsei University, Seoul, Korea,* [b]*minjin.ha@gmail.com*
[3]*Department of Biostatistics, University of Michigan,* [c]*satwika@umich.edu,* [d]*veerab@umich.edu*

Bayesian joint factor-based models

Bayesian nonparametric tree-based models

The Annals of Applied Statistics
2023, Vol. 17, No. 3, 1884–1908
https://doi.org/10.1214/22-AOAS1696
© Institute of Mathematical Statistics, 2023

PROBABILISTIC LEARNING OF TREATMENT TREES IN CANCER

BY TSUNG-HUNG YAO[1,a], ZHENKE WU[1,b], KARTHIK BHARATH[2,e], JINJU LI[1,c] AND VEERABHADRAN BALADANDAYUTHAPANI[1,d]

[1]*Department of Biostatistics, University of Michigan at Ann Arbor,* [a]*yaots@umich.edu,* [b]*zhenkewu@umich.edu,* [c]*lijinju@umich.edu,* [d]*veerab@umich.edu*
[2]*School of Mathematical Sciences, University of Nottingham,* [e]*Karthik.Bharath@nottingham.ac.uk*

SPECIAL SERIES: INFORMATICS TOOLS FOR CANCER RESEARCH AND CARE

Personalized Network Modeling of the Pan-Cancer Patient and Cell Line Interactome

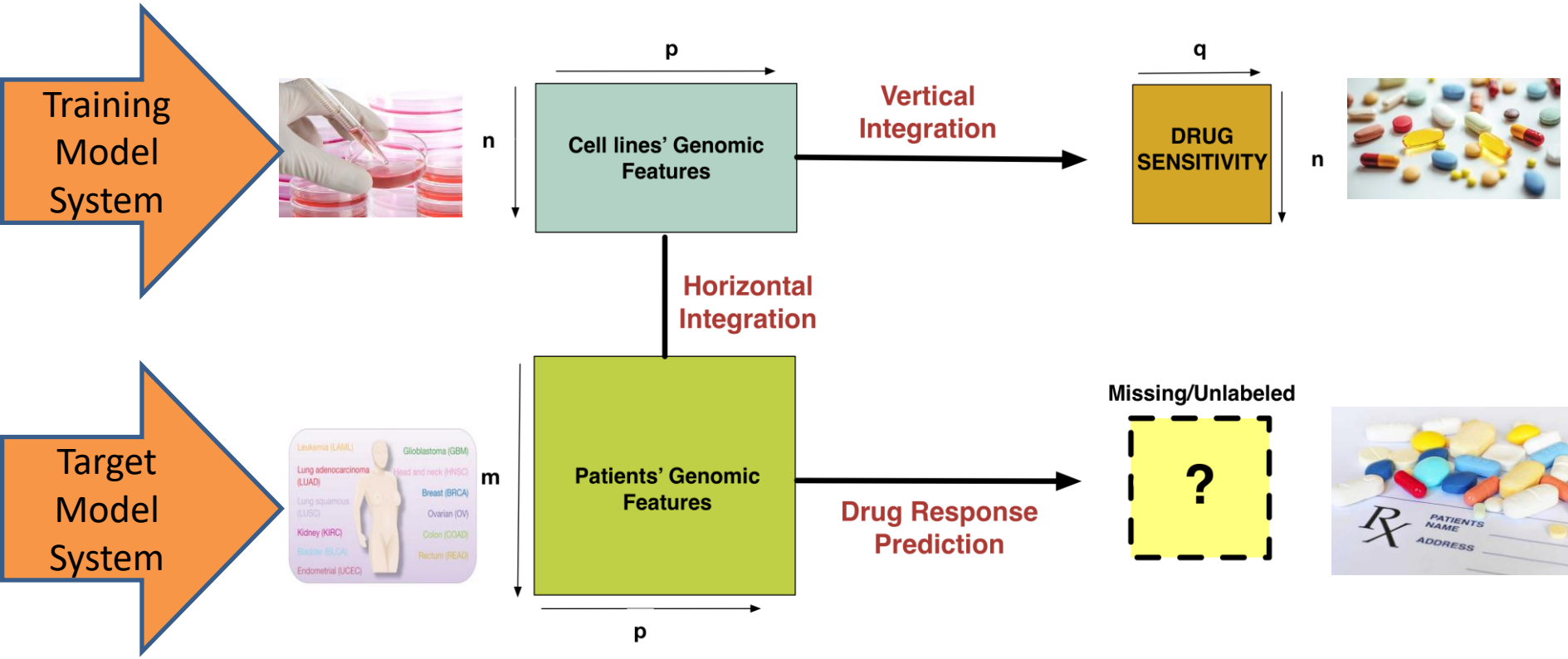
Rupam Bhattacharyya, MStat[1]; Min Jin Ha, PhD[2]; Qingzhi Liu, MS[1]; Rehan Akbani, PhD[3]; Han Liang, PhD[3,4]; and Veerabhadran Baladandayuthapani, PhD[1]

ASCO
JCO® Clinical Cancer Informatics

Bayesian graphical modeling approaches

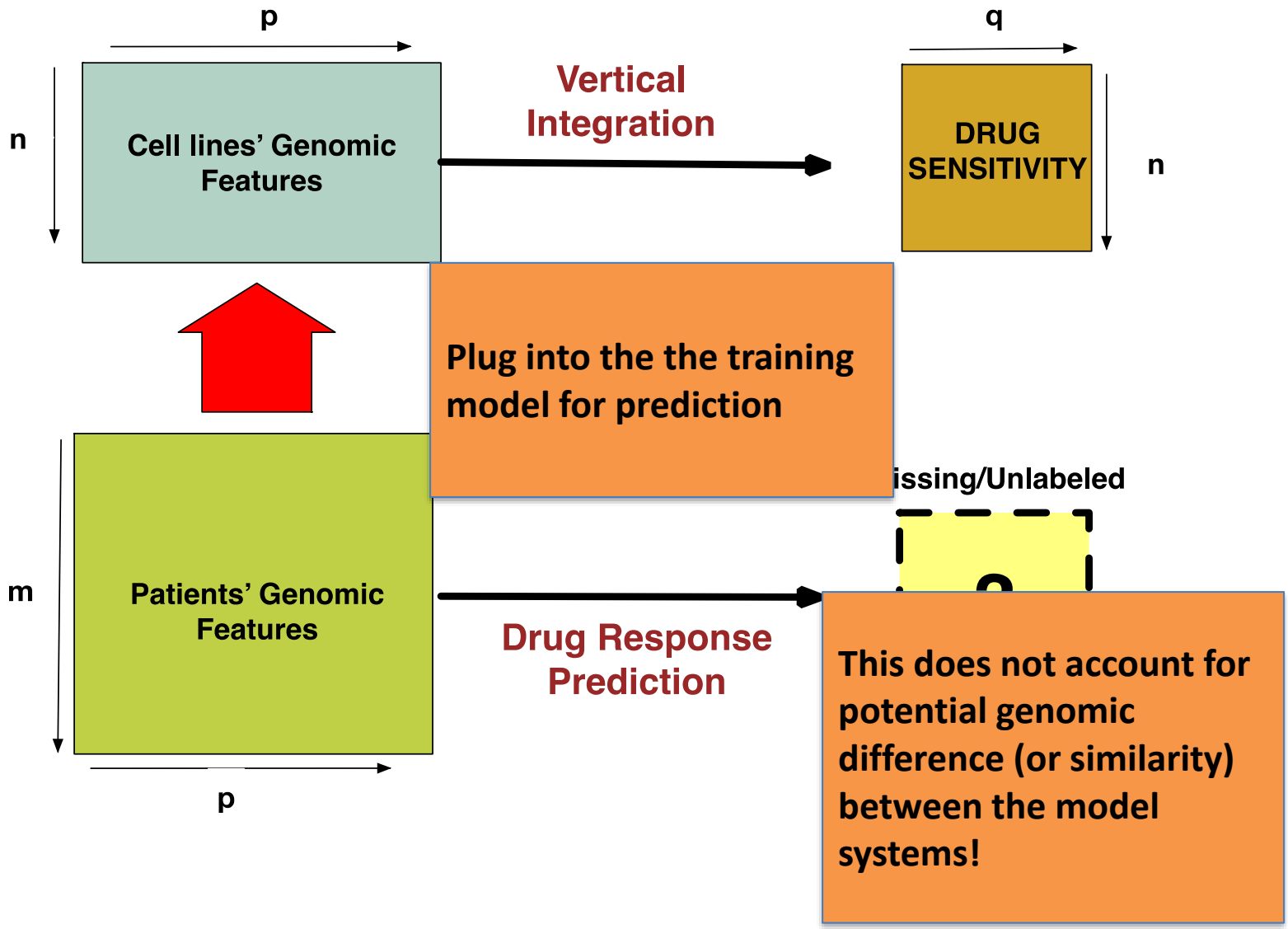
Bayesian Calibration of Therapeutic Indices

Conceptual Integrative Framework



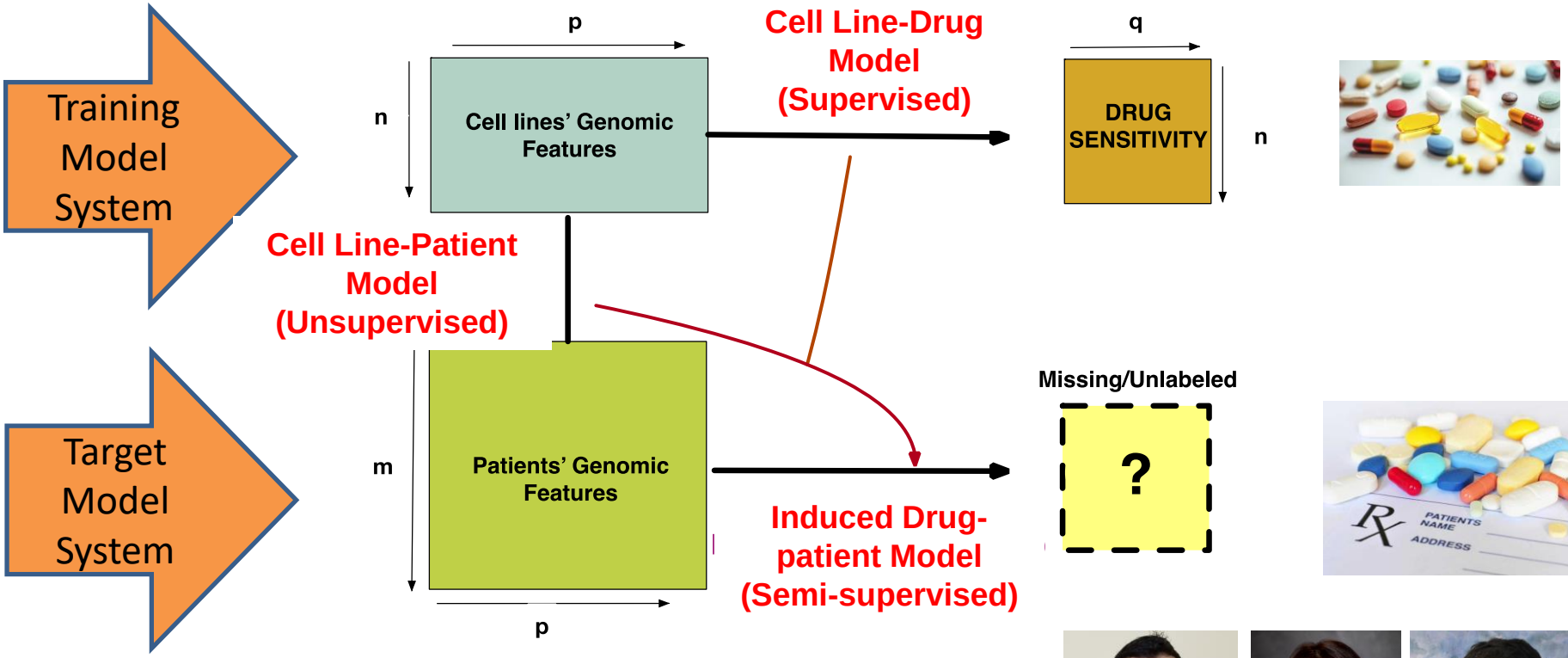
- m,n = # of samples
- p = # of genomic features
- q = # of drugs

Naïve supervised approach



Trained on cell lines and tested (independently) on patient's data

Calibration approach



individualized theRapeutic index (iR_x) model

Saha et al (2022, AOAS)

iR$_x$ Model Formulation

- Labeled data: $(D_j, \boldsymbol{C_j}), j = 1, \ldots, N_C$ (# of cell lines)
- Unlabeled data: $(D_i^*, \boldsymbol{P_i}), i = 1, \ldots, N_P$ (# of patients)
  - $(D_j, \boldsymbol{C_j}, \boldsymbol{P_i})$ = drug response and genomic measurements of j^{th}cell line, i^{th} patient
  - $(\boldsymbol{C_j}, \boldsymbol{P_i})$ – each high-dimensional vector of **G** genes
  - D_i^* unknown

- Key goal: infer distribution of D_i^*
  - $p(D|P, \Theta)$ **<-** **target distribution of interest (iR$_x$ distribution)**
  - Θ (model parameters)

iR_x Distribution

For a generic patient with genomic features $\mathcal{P}$,

$$p(\mathcal{D}|\mathcal{P},\Theta) = \int_{\mathcal{C}} p(\mathcal{D},\mathcal{C}|\mathcal{P},\Theta)d\mathcal{C} = \int_{\mathcal{C}} p(\mathcal{D}|\mathcal{C},\mathcal{P},\Theta)p(\mathcal{C}|\mathcal{P},\Theta)d\mathcal{C}$$

$$= \int_{\mathcal{C}} p(\mathcal{D}|\mathcal{C},\Theta)p(\mathcal{C}|\mathcal{P},\Theta)d\mathcal{C}$$

$$= \int_{\mathcal{C}} \underbrace{p(\mathcal{D}|\mathcal{C},\Theta)}_{\text{[Drug-cell line model]}} \times \underbrace{p(\mathcal{C}|\mathcal{P},\Theta)}_{\text{[Cell line-patient model]}} d\mathcal{C}, \quad (1)$$

Assumption, $p(\mathcal{D}|\mathcal{C},\mathcal{P},\Theta) = p(\mathcal{D}|\mathcal{C},\Theta)$
Motivates two sub-models for $\mathrm{iR_x}$ computation.

- Many choices available for both sub-models

iR_x model specifications: Drug-Cell Line Model

Drug-cell line model

$$\boldsymbol{D} = \mathbf{C}'\boldsymbol{\beta} + \boldsymbol{\epsilon}, \;\; \boldsymbol{\epsilon} \sim N(\mathbf{0}, \tau^{-1}\mathbf{I}),$$

- $\mathbf{C}$ = genomic data from cell-lines
- $\boldsymbol{\beta}$ = corresponding regression coefficients
- Need to deal with high-dimensional setting ($G >> N_c$)

- Drug-cell line model serves as our (labeled) training model
- Follows a penalized linear (ridge) regression model
- Can be more general: [insert your favorite prediction model]
  - linear/parametric (e.g. lasso, horse-shoe…) or nonlinear/non-parametric (e.g. trees, additive models)

iR_x model specifications: Cell Line – Patient Model

Cell line-patient model

$$[\mathbf{C}, \mathbf{P}] = \mathcal{A}\mathcal{F} + \varepsilon, \ \mathcal{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_N],$$
$$\varepsilon = [\varepsilon_1^1, \ldots, \varepsilon_1^{N_c}, \varepsilon_2^1, \ldots, \varepsilon_2^{N_p}],$$

where $\mathbf{f}_i$'s: vector of K latent factors, $\mathcal{A}$: loading matrix. $\varepsilon_1^j, \varepsilon_2^i \sim N(\mathbf{0}, \Psi_1^{-1}), N(\mathbf{0}, \Psi_2^{-1})$ respectively (K is unknown)

- Latent factor models: captures the underlying genomic similarity using low-dimensional factors (K unknown)
- Explicitly quantify --
  - Source-specific (**patients vs. cell line**) &
  - Shared variations separately (**patients + cell lines**)
  - Dependence between genes
  - Useful for clustering as well

iR_x Distribution and Calibration

Cell line-patient model factorization

$$\mathcal{C}|\mathbf{f} = \mathcal{A}\mathbf{f} + \varepsilon_1, \;\; \mathcal{P}|\mathbf{f} = \mathcal{A}\mathbf{f} + \varepsilon_2; \;\; \varepsilon_l \sim N(\mathbf{0}, \Psi_l^{-1}), \; l = 1, 2,$$

where $\mathbf{f}$, the common K-dimensional vector of factors, follows $N(\mathbf{0}, \mathbf{I}_K)$. $\varepsilon_1, \varepsilon_2$ and $\mathbf{f}$ are mutually independent.

iR_x distribution

Under the two Gaussian sub-models the target iR_x distribution $p(\mathcal{D}|\mathcal{P})$ is,

$$p(\mathcal{D}|\mathcal{P}) = N(\beta' \mathcal{A}\mathcal{A}^T(\mathcal{A}\mathcal{A}^T + \Psi_2^{-1})^{-1}\mathcal{P}, \tau^{-1} + \beta^T V \beta)$$

where V is $var(\mathcal{C}|\mathcal{P})$

Any distributional summary can be used to compute the iR_x scores (mean, variance, quantiles etc..)

iR_x Scores

iR_x

The theoretical *i*ndividualized the*R*apeutic inde*x*, or $\mathrm{iR_x}$, for a patient with genomic profile, $\boldsymbol{\mathcal{P}}$, is defined as $E(\mathcal{D}|\boldsymbol{\mathcal{P}})$, as given in following proposition.

Proposition

Under the proposed model in (2), (3) and the assumption (4), for a patient with genomic profile $\boldsymbol{\mathcal{P}}$, $\mathrm{iR_x}(\boldsymbol{\mathcal{P}}) = \beta^{\mathrm{T}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^{\mathrm{T}}(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^{\mathrm{T}} + \boldsymbol{\Psi}_2^{-1})^{-1}\boldsymbol{\mathcal{P}} = \beta^{\mathrm{T}}\boldsymbol{\Psi}_2\boldsymbol{\mathcal{A}}(\mathbf{I} + \boldsymbol{\mathcal{A}}^{\mathrm{T}}\boldsymbol{\Psi}_2\boldsymbol{\mathcal{A}})^{-1}\boldsymbol{\mathcal{A}}^{\mathrm{T}}\boldsymbol{\mathcal{P}}$

In practice,
$$\mathrm{iR_x}(\boldsymbol{P}_{\mathrm{i}}) = \widehat{\beta}^{\mathrm{T}}\widehat{\boldsymbol{\mathcal{A}}}\widehat{\boldsymbol{\mathcal{A}}}^{\mathrm{T}}\left(\widehat{\mathrm{Var}}(\boldsymbol{P}_{\mathrm{i}})\right)^{-1}\boldsymbol{P}_{\mathrm{i}} \quad (5)$$

where $\widehat{Var}(\boldsymbol{P}_i) = \widehat{\boldsymbol{\mathcal{A}}}\widehat{\boldsymbol{\mathcal{A}}}^{T} + \widehat{\boldsymbol{\Psi}}_2^{-1}$, and $\widehat{\beta}$, $\widehat{\boldsymbol{\mathcal{A}}}$ and $\widehat{\boldsymbol{\Psi}}_2$ are the corresponding estimates.

Note dependence on shared variation

Naïve Scores

NI

Alternative to $\mathrm{iR_x}$, naive index (NI), solely based on the drug-cell line model, is defined as $E(\mathcal{D}|\boldsymbol{\mathcal{C}} = \boldsymbol{P}_i) = \boldsymbol{\beta}^T \boldsymbol{P}_i$. Thus, $\mathrm{NI}(\boldsymbol{P}_\mathrm{i}) = \widehat{\boldsymbol{\beta}}^\mathrm{T} \boldsymbol{P}_\mathrm{i}$

Claim: Under the assumption of shared genomic variation across cell lines and patients, $\mathrm{MSPE}(\boldsymbol{P}_i)_{\mathrm{iR_x}} \leq \mathrm{MSPE}(\boldsymbol{P}_i)_{\mathrm{NI}}$. In addition, gain in accuracy is given by $(\mathrm{MSPE}(\boldsymbol{P}_i)_{\mathrm{NI}} - \mathrm{MSPE}(\boldsymbol{P}_i)_{\mathrm{iR_x}}) = (\boldsymbol{\beta}^T \mathbf{B} \boldsymbol{P}_i)^2$, $\mathbf{B} = [\boldsymbol{\Psi}_2 \boldsymbol{\mathcal{A}} (\mathbf{I}_K + \boldsymbol{\mathcal{A}}^T \boldsymbol{\Psi}_2 \boldsymbol{\mathcal{A}})^{-1} \boldsymbol{\mathcal{A}}^T - \mathbf{I}_G]$

Bottom-line

- **Better predictive accuracy** in comparison to naïve plug-in approaches
- Can **explicitly quantify accuracy gain** based on shared variation (and shown in simulations)

Bayesian Estimation

Let Θ, Θ_1 and Θ_2: set of all parameters, Drug-cell line model parameters and Cell line-patient model parameters respectively. Then

$$p(\Theta|\boldsymbol{D}, \mathbf{C}, \mathbf{P}) = \int_{\mathcal{F}} p(\Theta, \mathcal{F}|\boldsymbol{D}, \mathbf{C}, \mathbf{P}) d\mathcal{F} \propto p(\Theta_1|\mathbf{C}, \boldsymbol{D}) p(\Theta_2|\mathbf{C}, \mathbf{P})$$

if Θ_1 and Θ_2 are assumed independent to each other. Suffices to independently define priors for each model and conduct MCMC steps in parallel.

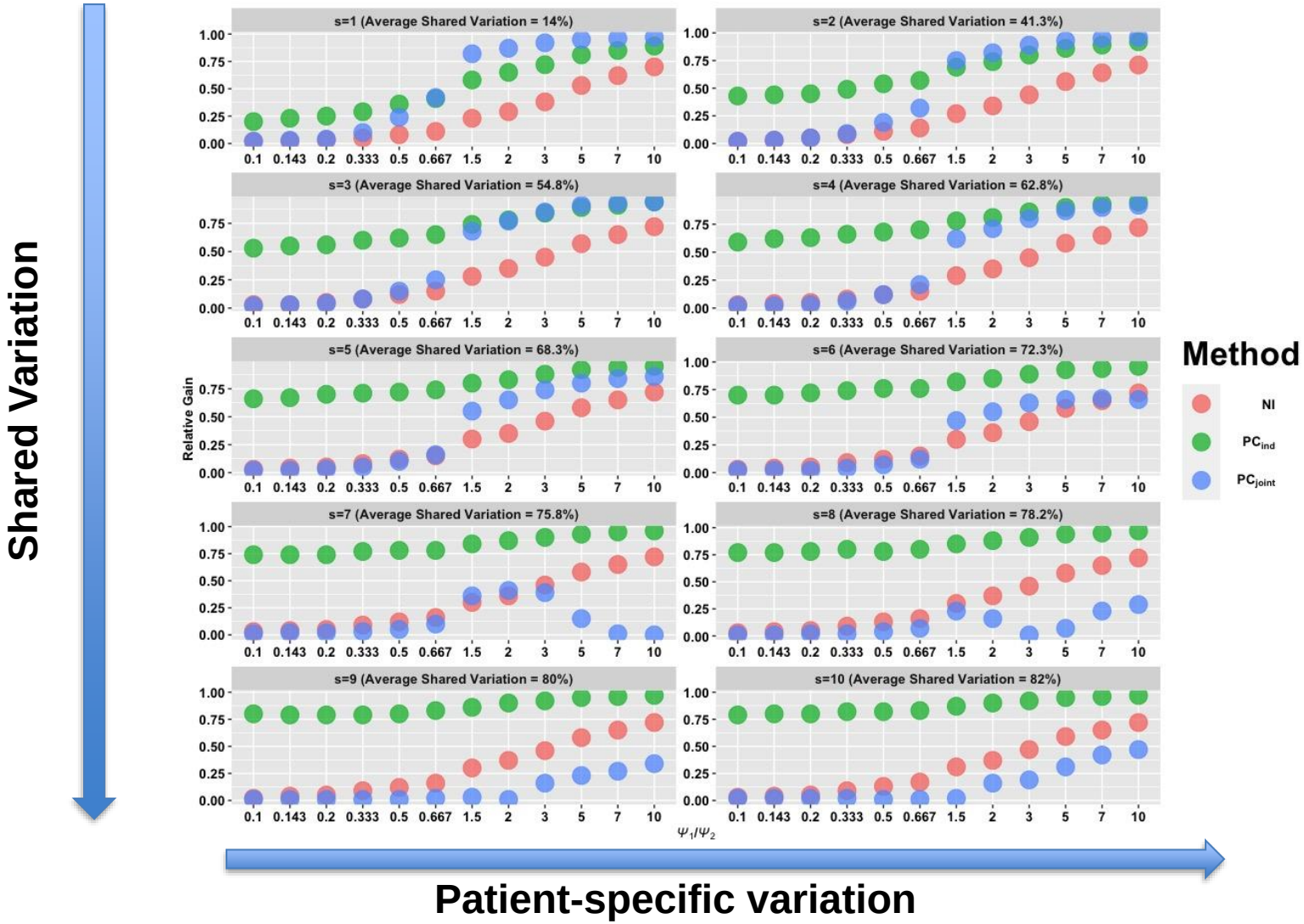
Prior for drug-cell line model

- $\Theta_1 = \{\boldsymbol{\beta}, \tau\}$, $p(\boldsymbol{\beta}|\tau) \sim N(\mathbf{0}, [\lambda\tau]^{-1}\boldsymbol{I})$, and $p(\tau) \sim \mathcal{G}a(a_\tau, b_\tau)$
- A normal prior for $\boldsymbol{\beta}$ automatically applies ridge penalization.
- λ, hyper parameter controlling shrinkage, assumed a weakly informative gamma.

Prior for cell line-patient model

- $\Theta_2 = \{\boldsymbol{\mathcal{A}}, K, \boldsymbol{\Psi_1}, \boldsymbol{\Psi_2}\}$, $p(\boldsymbol{\mathcal{A}}_{n\times\infty}) \sim$ multiplicative gamma process (MGP) shrinkage prior {Bhattacharya and Dunson's, 2011}, K determined by keeping significant columns from left.
- $p(\psi_{1,i}), p(\psi_{2,i}) \sim \mathcal{G}a(a, b)$, $i = 1 \ldots G$, i.i.d.

Impact of variability on relative accuracy gain



Bottomline: iR_x dominates NI, PC-based models in all scenarios; attributable to shared variation

Clinical validation using existing trial data

- **Hypothesis:** pooling information across labeled and unlabeled data (by using Cell line-patient model) **improves calibration accuracy**

- **Validate** **iR_x (Semi-supervised)** versus **independent methods** while estimating true drug response obtained from two clinical trials with drug-response data available

- Natural clustering due to latent factors can be used to
  - Find optimal cell line **"avatars"** as proxies for patients
  - Identify **key genomic drivers** explaining cell line-patient similarities; conserved biology

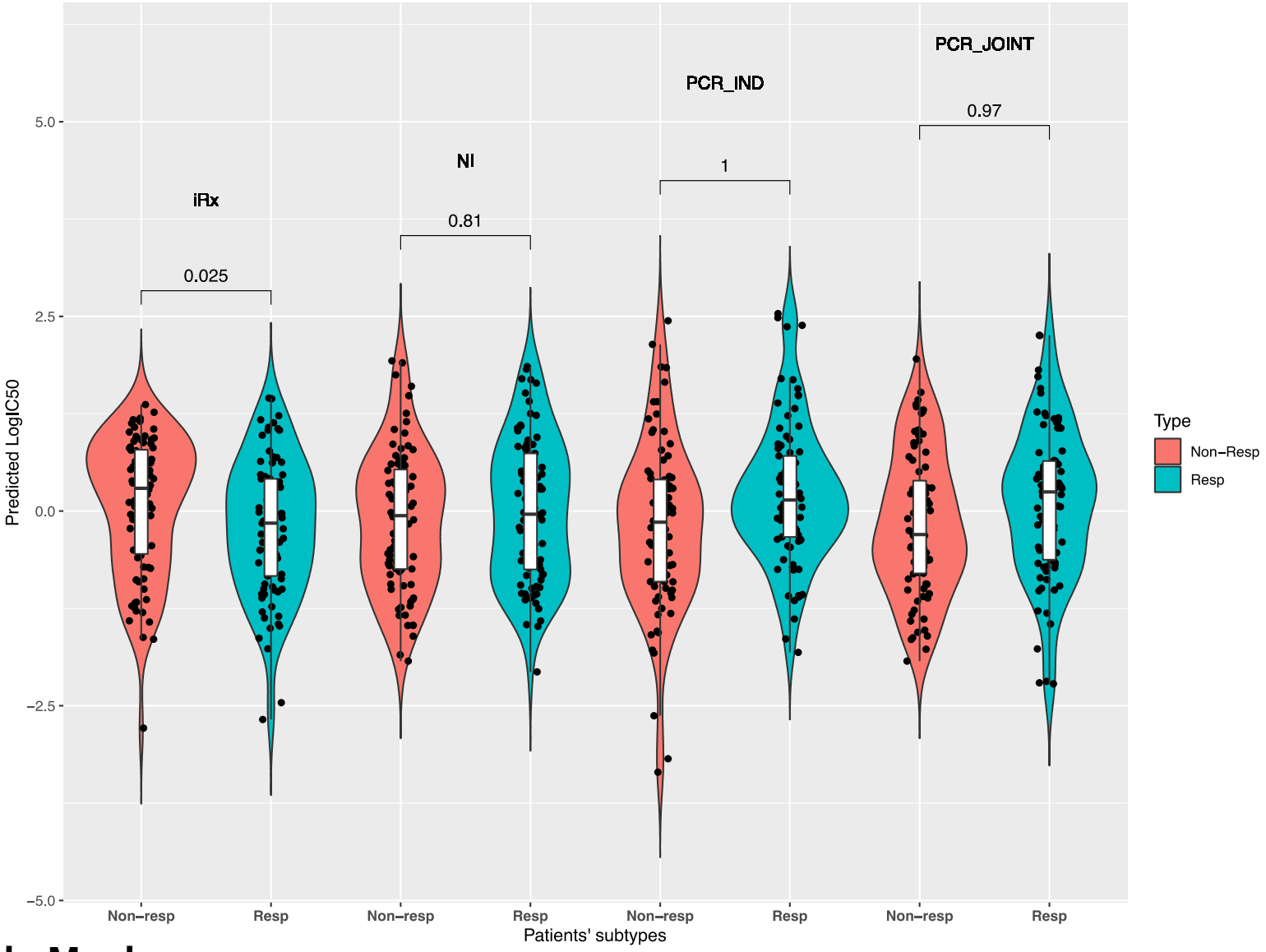
Data

- A panel of cell lines from GDSC database (training model system); log-IC50 as responses

- Target Clinical Data (drug responses are masked for validation)
  - **Multiple Myeloma** Phase II and III clinical trials of **bortezomib** on relapsed/refractory multiple myeloma patients, clinical response – Responder (CR, PR, MR; n=85) vs. non-responder (PD, NC; n=84) (Mulligan et al, 2007)
  - **Breast Cancer** **Docetaxel** study for 24 breast cancer tumor, % reduction of tumor after four cycles. Responder (25% reduction, n=10) vs. non-responder(n=14) (Chang et al, 2003)

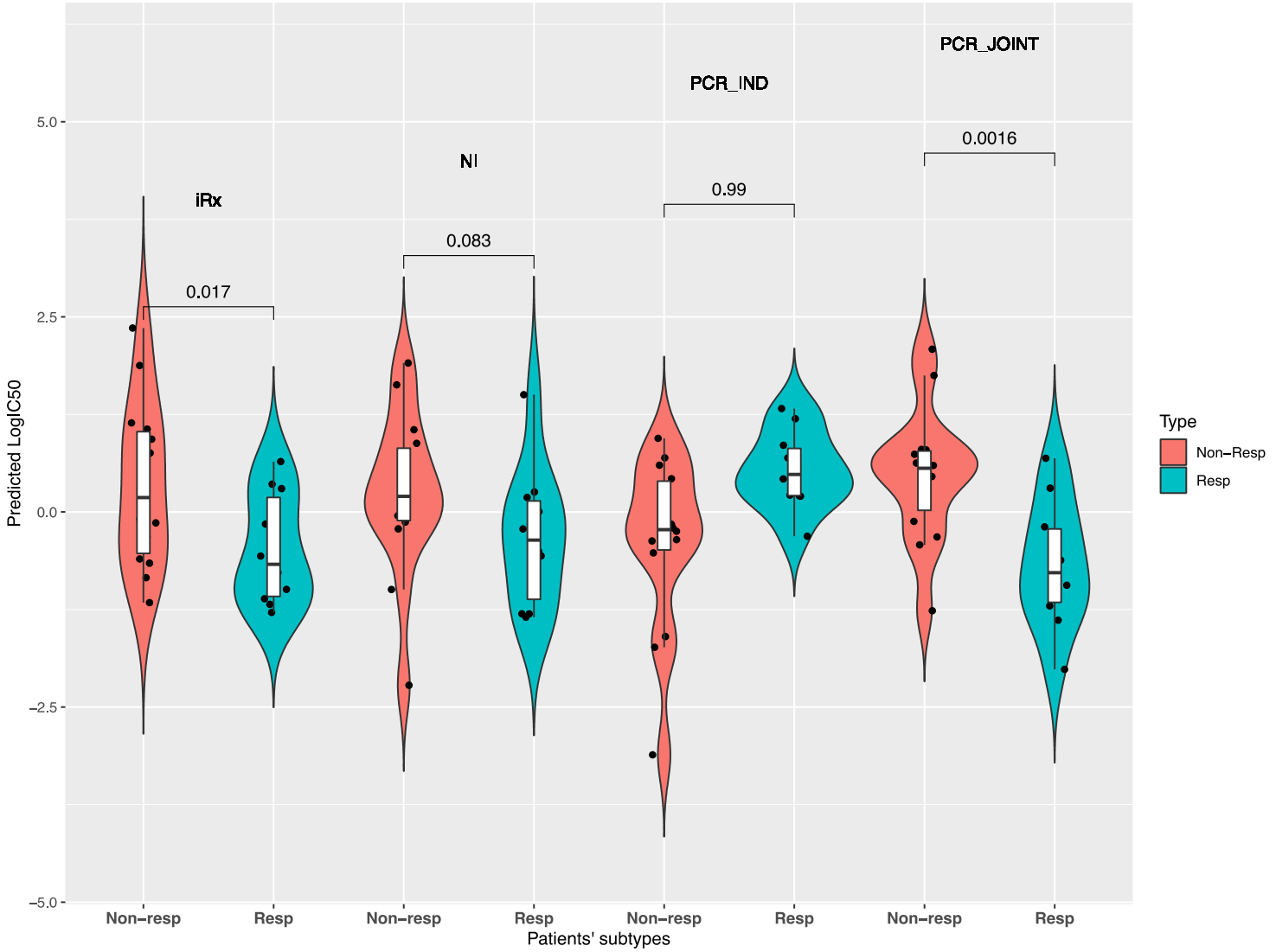
- Used mRNA/gene-expression data
- Some pre-processing and normalization (~1000 genes for fitting)

Therapeutic response calibration



Multiple Myeloma

Therapeutic response calibration



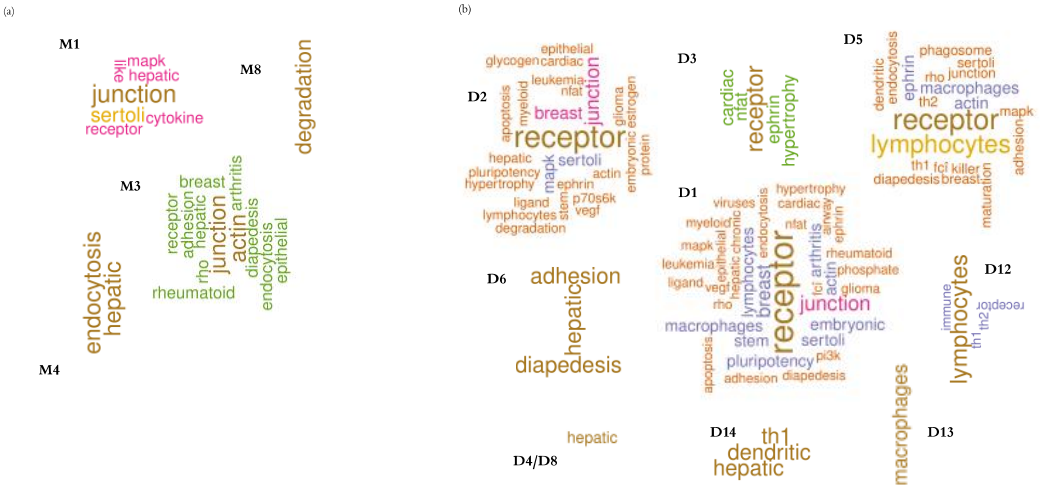
Breast Cancer

Top Pathway Drivers of Shared Variation

Table 3: The table lists the top five canonical pathways, as given by Ingenuity Pathway Analysis (IPA) software, that may be associated with the latent factors obtained from breast cancer patients. The significant pathways are selected based on adjusted Benjamini-Hochberg p-values. Only factors with at least one significant pathway are mentioned here.

Ingenuity canonical pathways (breast cancer patients)		
F1*	F2*	F5*
Hepatic Fibrosis / Hepatic Stellate Cell Activation	ERK5 Signaling	Leukocyte Extravasation Signaling
GP6 Signaling Pathway	Ephrin Receptor Signaling	iCOS-iCOSL Signaling in T Helper Cells
Role of Tissue Factor in Cancer	Non-Small Cell Lung Cancer Signaling	CD28 Signaling in T Helper Cells
Osteoarthritis Pathway	G Beta Gamma Signaling	Integrin Signaling
Oncostatin M Signaling	Actin Nucleation by ARP-WASP Complex	T Cell Receptor Signaling
F12	F4	F6
NFAT in Regulation of the Immune Response	Hepatic Fibrosis / Hepatic Stellate Cell Activation	Hepatic Fibrosis / Hepatic-
		-Stellate Cell Activation
B Cell Development	Role of Tissue Factor in Cancer	Granulocyte Adhesion and Diapedesis
Th1 and Th2 Activation Pathway	Osteoarthritis Pathway	
iCOS-iCOSL Signaling in T Helper Cells		
B Cell Receptor Signaling		
F8	F14	
Hepatic Fibrosis / Hepatic Stellate Cell Activation	T Helper Cell Differentiation	
	Crosstalk between Dendritic Cells-	
	-and Natural Killer Cells	

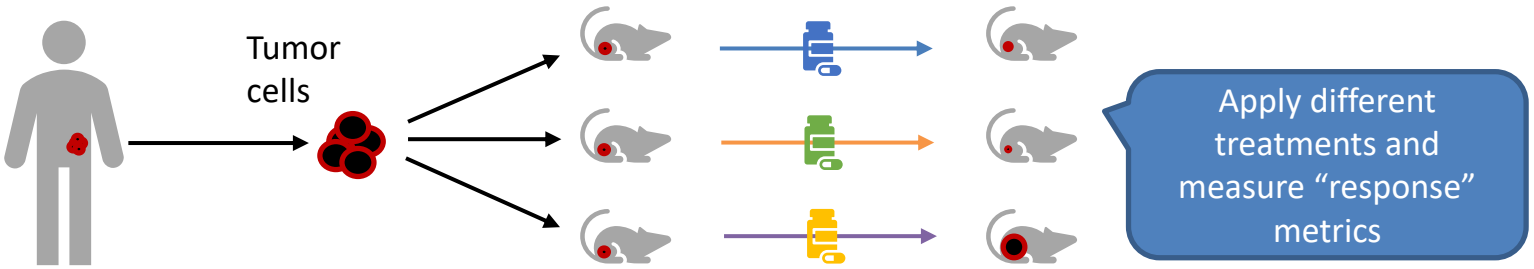
Top two are **MAPK signaling** pathways and the endocytic pathways. **NKG2D** (a receptor activating natural killer cells (NK)) expression, in multiple myeloma (MM) cells after cord blood derived **natural killer cell (CB-NK) treatment** correlates with lower MM progression, and NKG2D and NKP30 contribute more to the cytotoxicity of MM cells.



Probabilistic Learning of Treatment Trees in Cancer

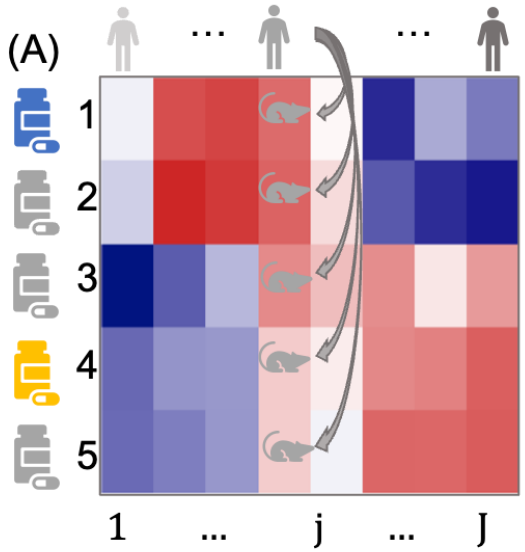
Patient Derived Xenografts (PDX)

- Broad context: preclinical models in cancer
- Create “avatar” for patients to test different plausible treatments
- Typical choices:
  - Tumor-derived cell lines: *in vitro*; cheaper but lower fidelity to human tumors
  - Immuno-deficient mice: *in vivo*; more expensive with high clinical relevance
- **Patient Derived Xenograft** (PDX): implant cancer cell from a single patient to multiple immuno-deficient mice
  - Capture better micro-environment with higher clinical relevance



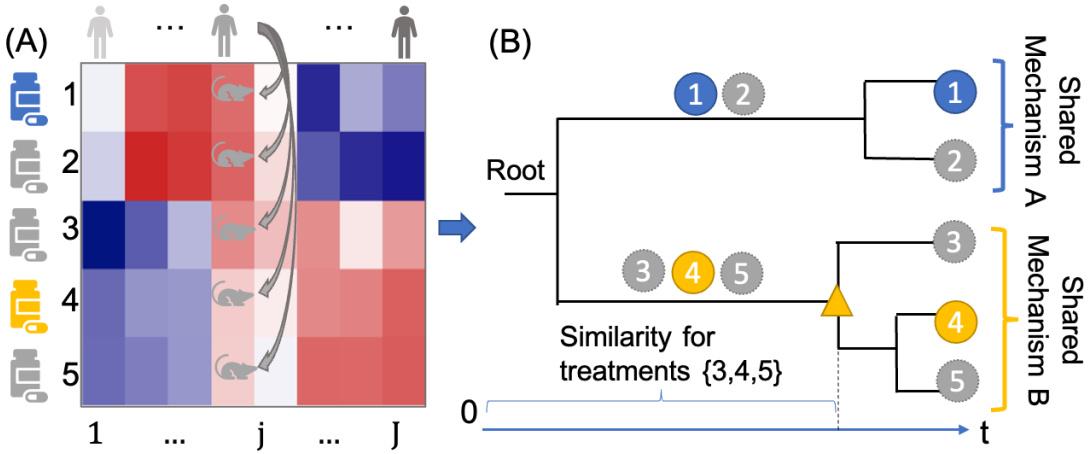
PDX Experiments

- Multiple PDXs from same patients with set of treatments (usually same)
- Response: some metric of difference in tumor size (esp. in cancer)
- Use PDX data as a pre-clinical trial to screen different treatments; "co-clinical trials"
- Key scientific question(s):
  - Evaluate the effectiveness of multi-drug combinations
  - Identify underlying plausible biological mechanisms



Tree-based representations

- Key concept: Treatments with the same target or biological mechanism should induce similar responses
  - Engender **mechanism-related clustering** among treatments
- Infer hierarchy among treatments: partition + how clusters relate to each other
  - "Flat" clustering (e.g. k-means) only shows partition patterns



- Known entities at the leaves, i.e., the different treatments
- Unknown tree to be inferred from the treatment responses
- **We call this a treatment tree (RxTree)**

Treatment Trees (RxTree) on PDXs

- Rxtree empirically characterizes the mechanism similarity
- Rooted tree: root + internal nodes + branch lengths + leaves

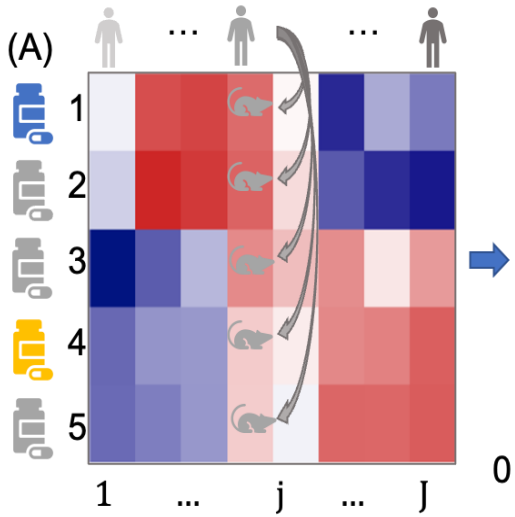
clusters of treatments

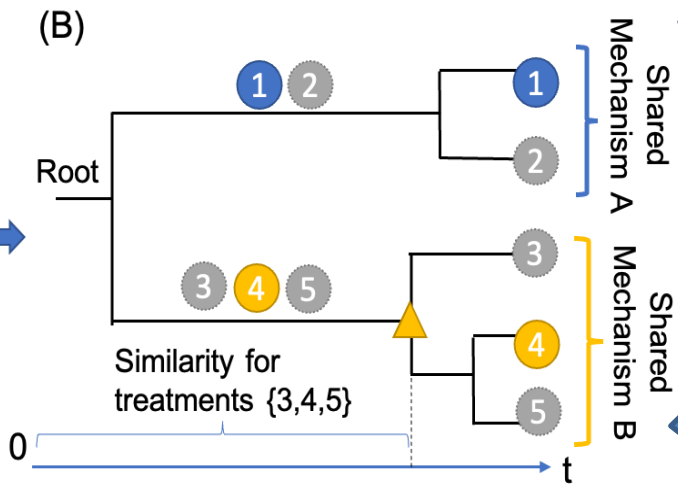
mechanism similarities

Only leaves are observable

- Key idea: Treatments that stay clustered "longer" have higher mechanism similarities

Treatments 1 and 4: different but known biological mechanisms; the rest treatments have unknown mechanisms





The tree suggests two treatment groups corresponding to two different mechanisms

The horizontal position of "△" measures the mechanism similarity for treatments {3,4,5}.

Basic Construction of RxTrees

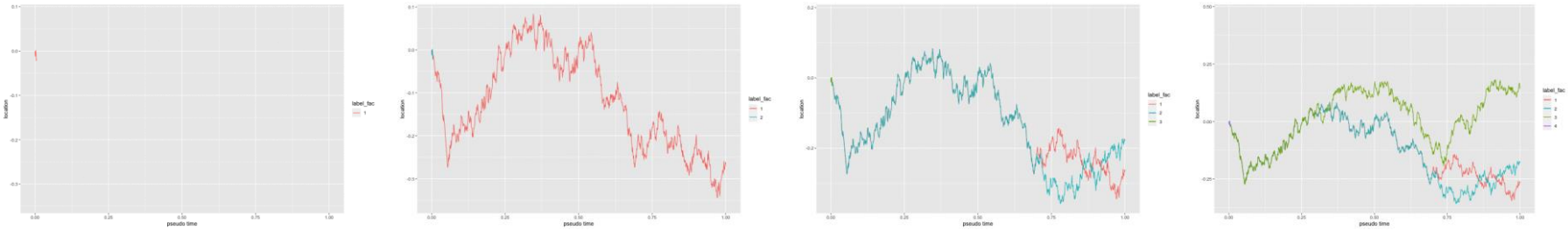
- Observed data i.e., responses matrix $\mathbf{X}_{\mathrm{I}\times\mathrm{J}} = [\boldsymbol{X}_1 \dots \boldsymbol{X}_{\mathrm{I}}]^{\mathrm{T}}$ for I treatments and J patients
  - $\boldsymbol{X}_{.,\mathrm{j}} = [\mathrm{X}_{1\mathrm{j}} \dots \mathrm{X}_{\mathrm{Ij}}]^{\mathrm{T}}$ observed response column for j-th patient across I treatments

- Model the responses through a generative model that results in a Gaussian likelihood:
  - $\boldsymbol{X}_{.,\mathrm{j}} | \boldsymbol{\Sigma}^{\mathcal{T}} \sim^{\mathrm{iid}} \mathbf{N}_{\mathrm{I}}\left(\mathbf{0}, \boldsymbol{\Sigma}^{\mathcal{T}}\right), \mathrm{j} = 1, \dots, \mathrm{J}$

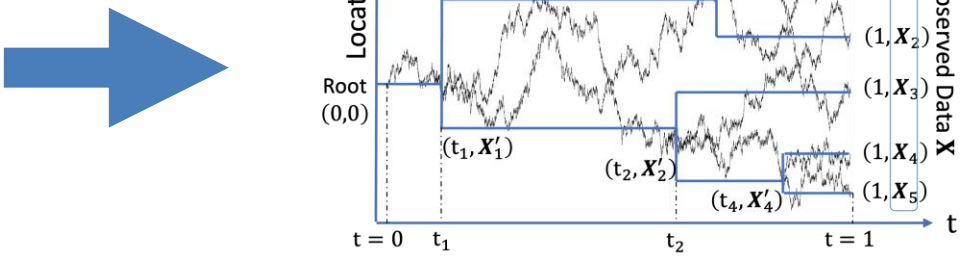
- $\boldsymbol{\Sigma}^{\mathcal{T}}_{\mathrm{i},i'}$: covariance between treatments i and i′ and measures their similarity
- $\boldsymbol{\Sigma}^{\mathcal{T}}$ has a special structure: tree-based covariance (not usual covariance)
- Needs specific constraints; non-trivial to estimate

Dirichlet Diffusion Trees

- Model $\boldsymbol{\Sigma}^{\mathcal{T}}$ through Dirichlet diffusion tree (DDT) model (Neal, 2001)
- Generate tree randomly through scaled Brownian motion

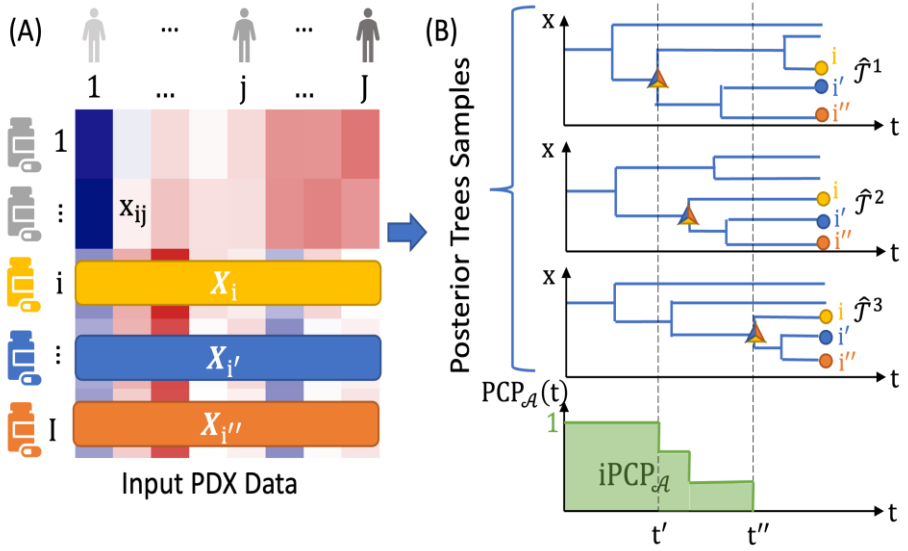


Represent the tree with the backbone



Inferential summaries

- Given posterior samples, we can compute
  - A global MAP of the Rx-tree that represents the overall hierarchy
  - Local uncertainty estimates of **posterior co-clustering probabilities (PCP)** for any subset of treatments

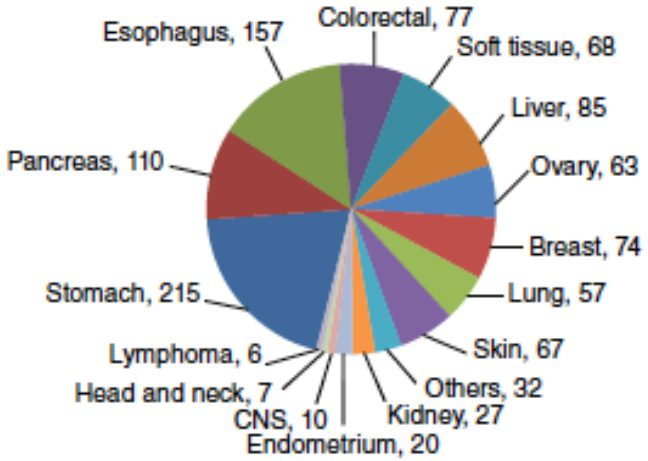


- $PCP_{\mathcal{A}}(t) \in [0,1]$ quantifies propensity among treatments to cluster
- Integrated PCP ($iPCP \in [0,1]$): area under the PCP curve as a scalar summary of PCP

- Several advantages of iPCP:
  - Interpretable metric: expected (or average) chance of co-clustering for treatments
  - Can compute multiway "correlations" (2,3,4… treatments) **<- useful to find combination therapies**

Motivating data: NIBR-PDXE

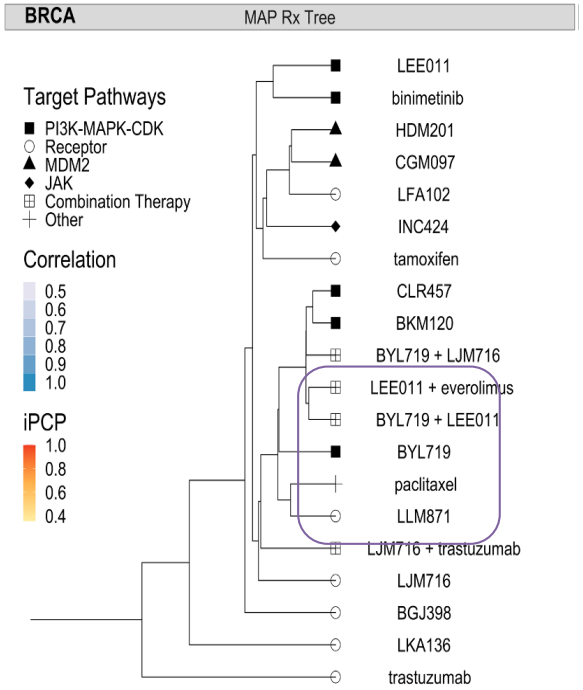
- **Novartis Institutes for BioMedical Research - PDX Encyclopedia (NIBR-PDXE)**
  - High-throughput treatment screening using PDX
  - 1,075 PDX lines with 1 × 1 × 1 design (one animal per PDX model per treatment)
  - Across ~16 cancers and 62 treatments
  - 38 unique therapeutic entities
  - Used in 36 monotherapies or in 26 combination therapies

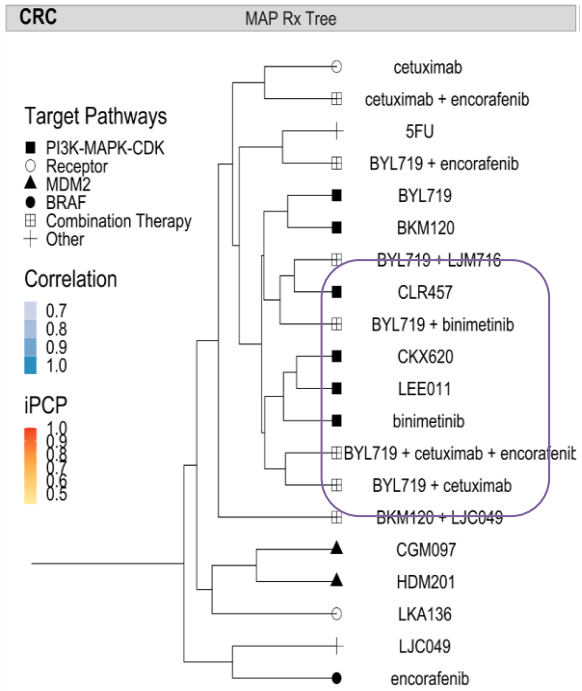


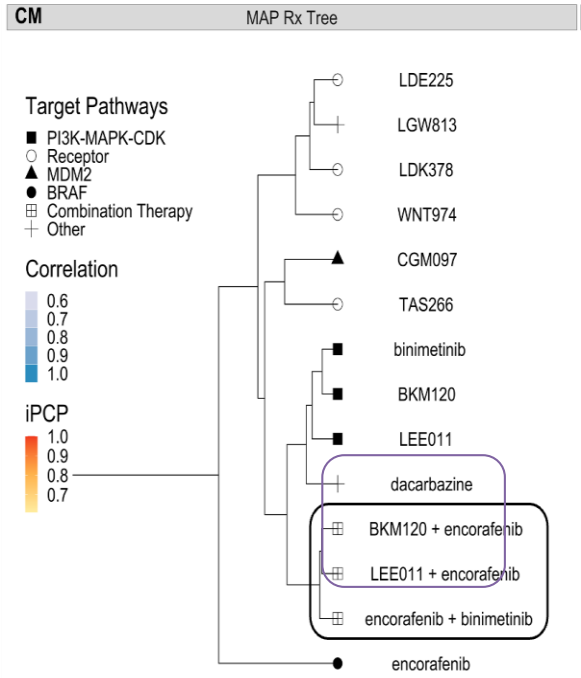
Gao, H. et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* **21**, 1318–1325 (2015)

- Focus on five cancers with more complete responses
  - Breast carcinoma (BRCA); Cutaneous melanoma (CM); Colorectal cancer (CRC); Non-small cell lung carcinoma (NSCLC); Pancreatic ductal carcinoma (PDAC)

RxTrees



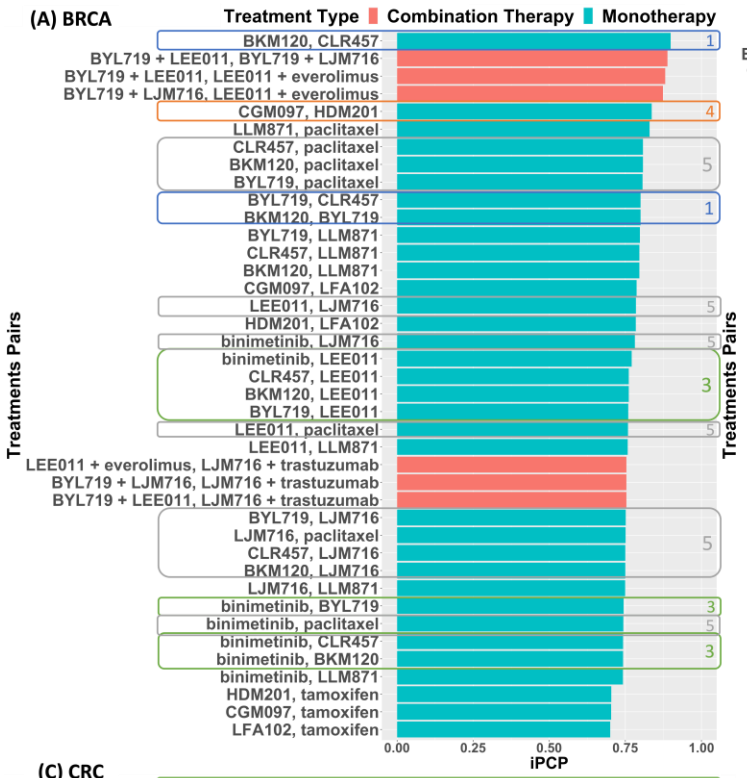




PI3K, MAPK and CDK inhibitors belong to a tighter subtree across cancers

iPCPs

iPCP further quantifies the similarity



Box Colors	Treatment Target Pathways
Blue (1)	PI3K
Yellow (2)	MAPK
Green (3)	(PI3K, MAPK) or (CDK, MAPK) or (PI3K, CDK)
Orange (4)	MDM2
Grey (5)	(Tubulin, PI3K-MAPK-CDK) or (ERBB3, PI3K-MAPK-CDK)

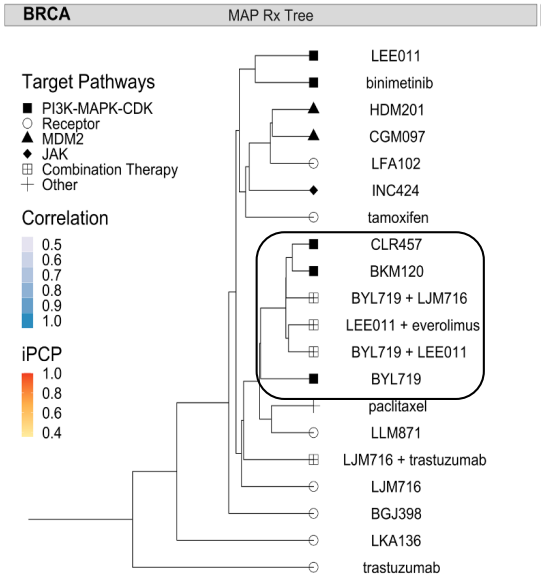
Mechanistic Similarities in Monotherapies

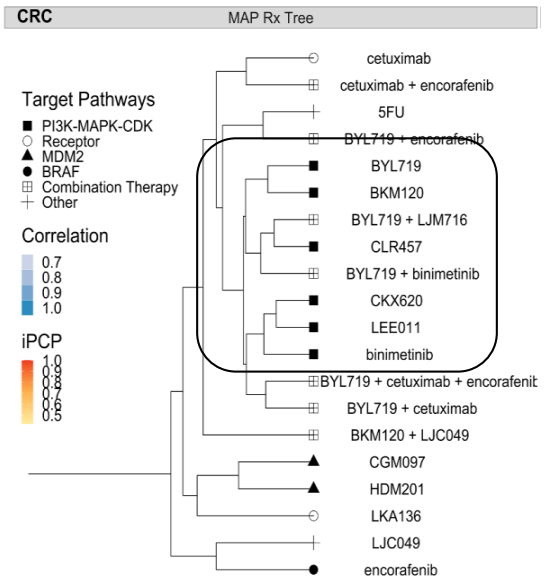
- Treatments have different targets might have a high mechanistic similarity
  - Might share a common downstream mechanism (hypothesis)

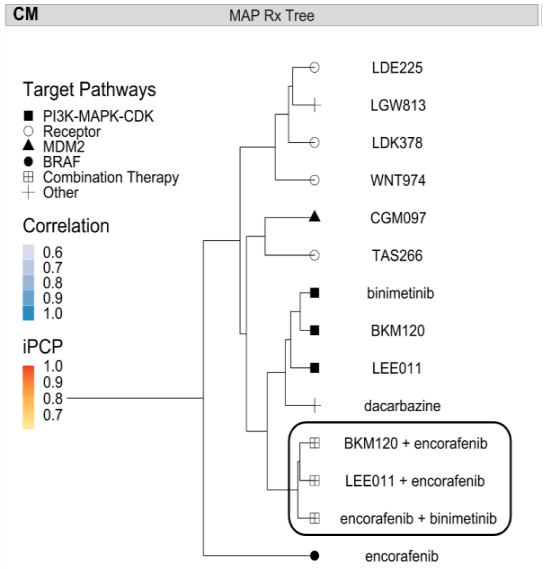
- PI3K and MAPK
  - High pairwise iPCP
    - BBRCA: (binimetinib, BKM120): **0.743**; (binimetinib, BYL719): **0.744**; (binimetinib, CLR457): **0.743**
    - CRC: (binimetinib, BKM120): **0.737**; (binimetinib, BYL719): **0.739**; (binimetinib, CLR457): **0.754;** (CKX620, BKM120): **0.737**, (CKX620, BYL719): **0.736**, (CKX620, CLR457): **0.768**
    - CM: (binimetinib, BKM120): **0.8882**
  - High multi-way iPCPs in BRCA (**0.7422**), CRC (**0.7300**) and CM (**0.8882**)
  - Plausible biological explanation: PI3K and MAPK can be induced by ERBB3 phosphorylation (Balko et al., 2012)

Combination Therapies

- Investigate combination therapies to identify <u>synergistic mechanisms</u>
  - Combination therapies tend to form a tighter subtree
  - Mechanisms under combination therapies are similar to each other and are closer to the PI3K-MAPK-CDK pathways.

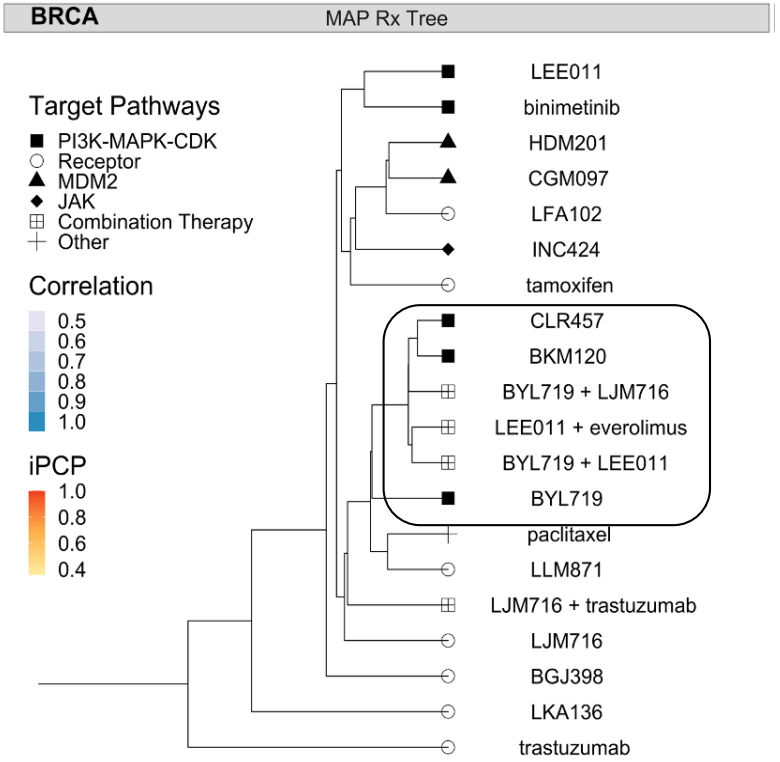




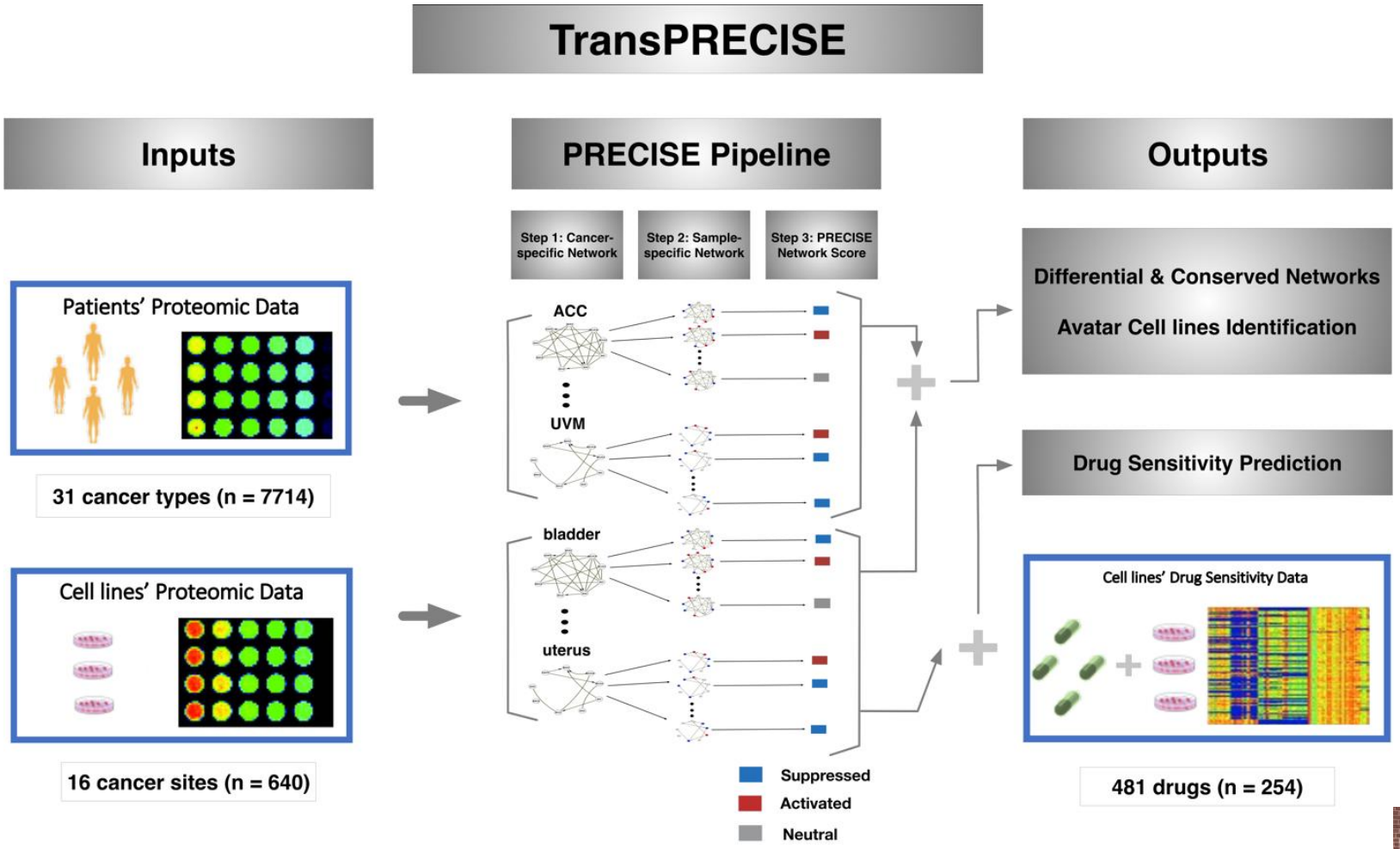


Breast Cancer

- Four combination therapies were tested in BRCA
  - Three therapies targeting PI3K-MAPK-CDK
  - {BYL719 + LJM716, BYL719 + LEE011 and LEE011 + everolimus} form a subtree with a high three-way iPCP (**0.8719**)
- **Clinical relevance**: PI3K-CDK inhibitor, BYL719 + LEE011, has synergistic regulation (Vora et al., 2014; Bonelli et al., 2017; Yuan et al., 2019)
- High three-way iPCP suggest mechanistic synergy for combination therapies targeting:
  - PI3K-ERBB3 (BYL719 + LJM716)
  - CDK-MTOR (LEE011 + everolimus)



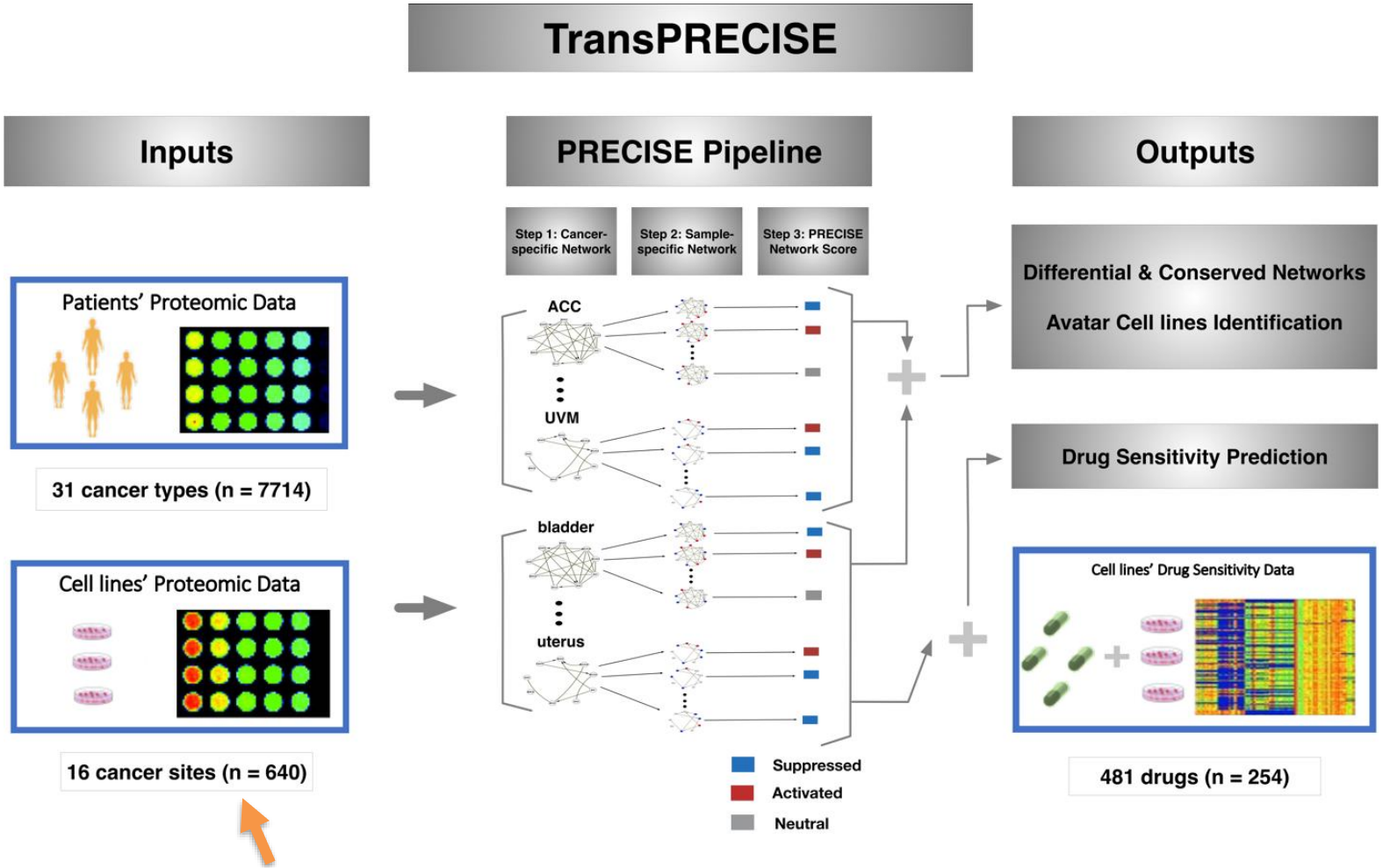
Pan-Cancer Models



Personalized Network Modeling of the Pan-Cancer Patient and Cell Line Interactome

Bhattacharya et al (JCO, 2020)

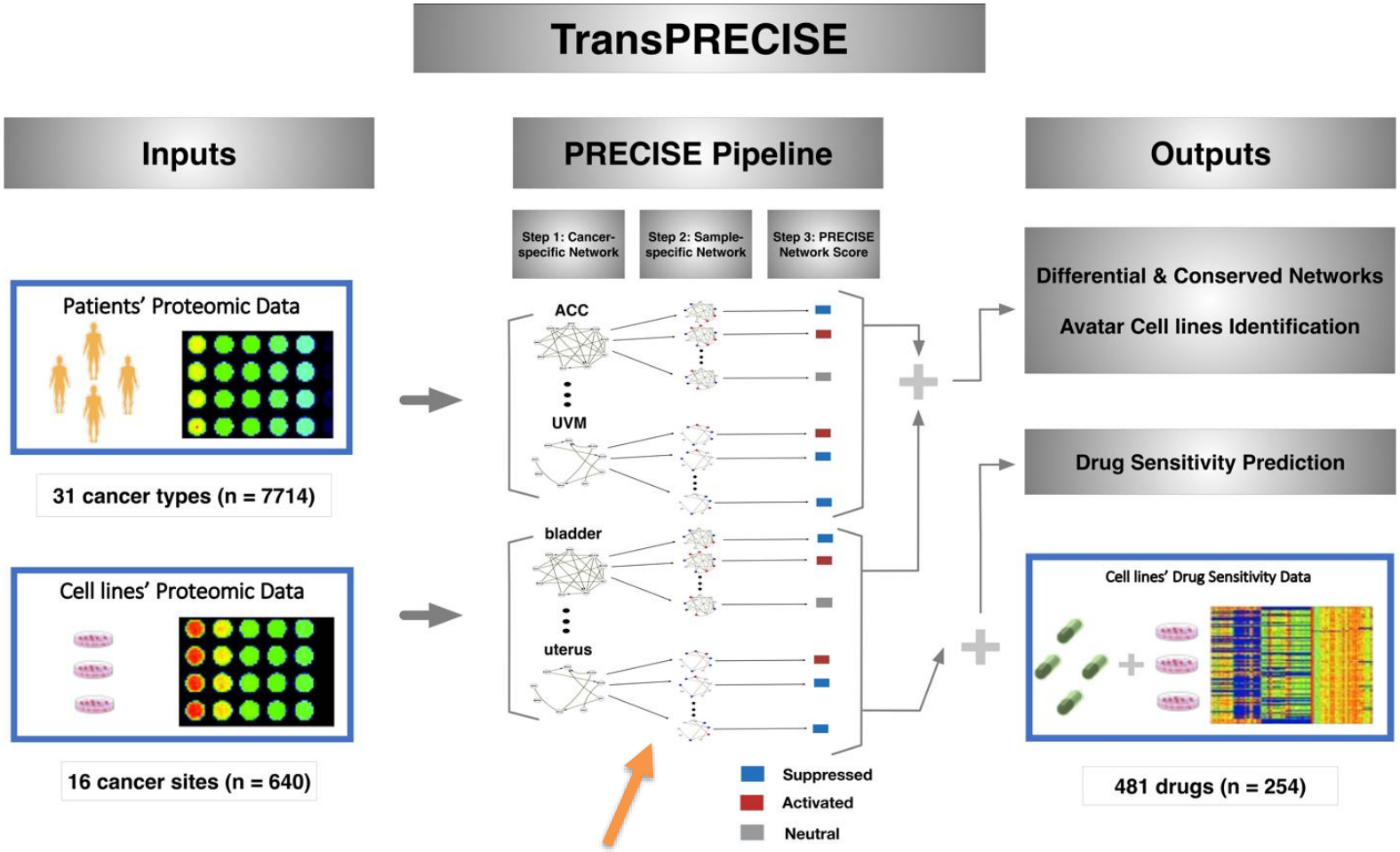
TransPRECISE



Input
RPPA-based proteomics data from the patients and cell lines.

Bhattacharya et al (JCO, 2020)

TransPRECISE

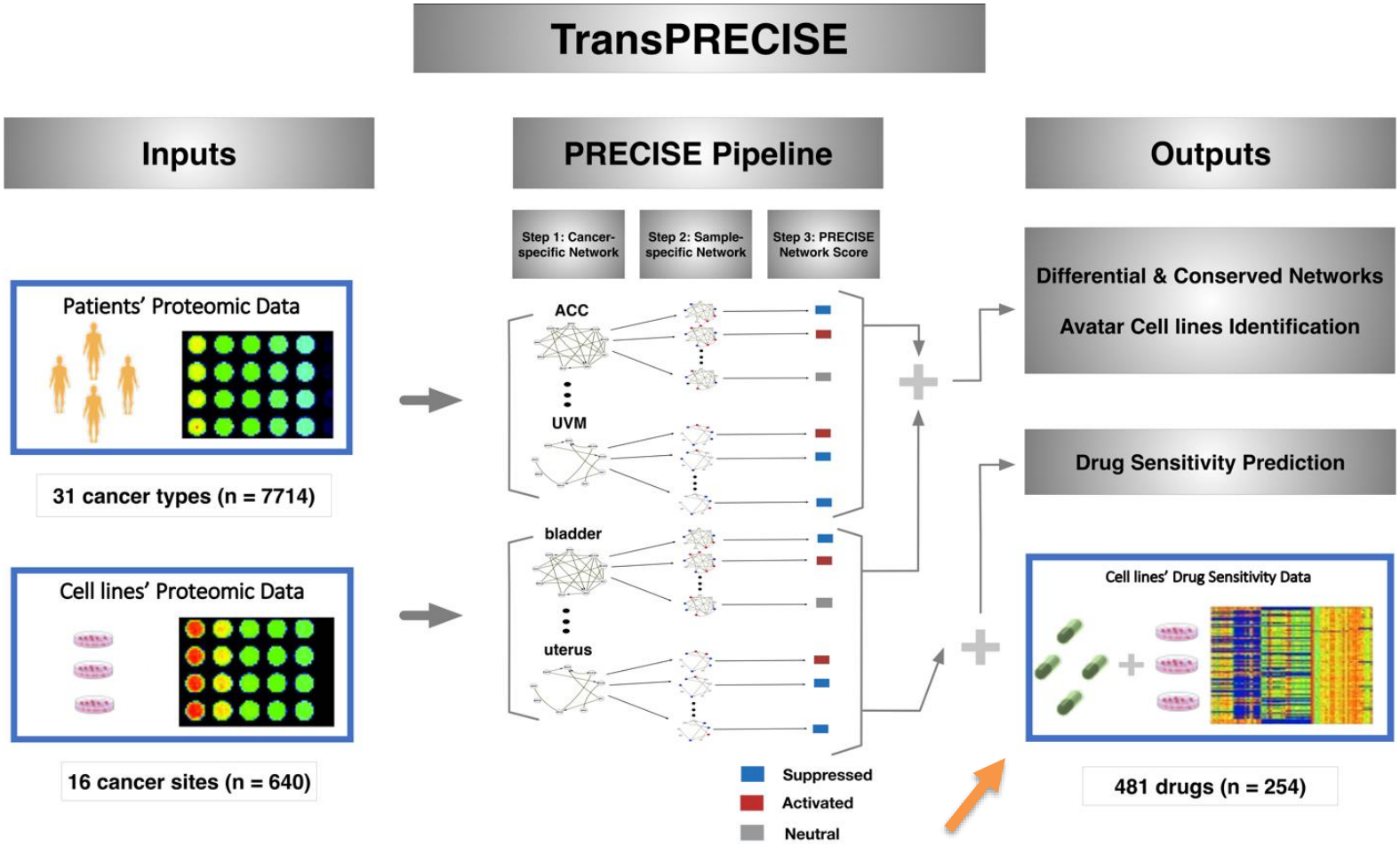


Modeling

1. Bayesian graphical regression to estimate the cancer-specific pathway network structure.
2. De-convolving population-level networks to sample-specific networks.
3. Summarize networks and quantify pathway activity status.

Bhattacharya et al (JCO, 2020)

TransPRECISE

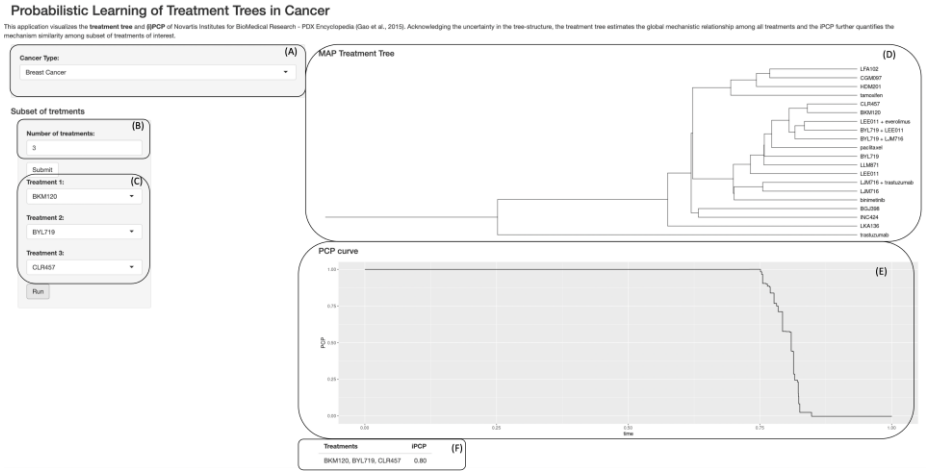


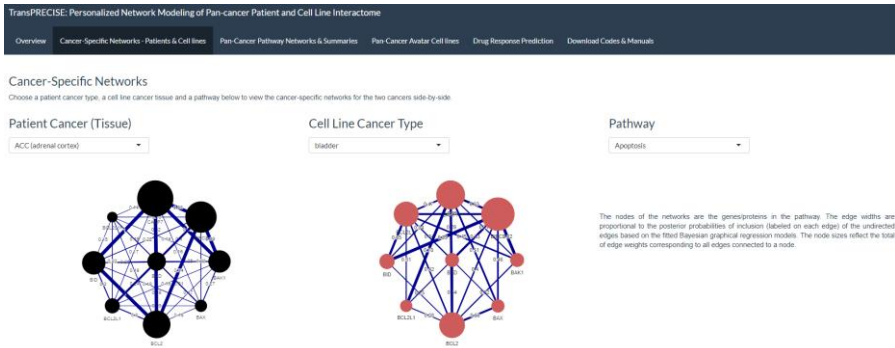
Outputs

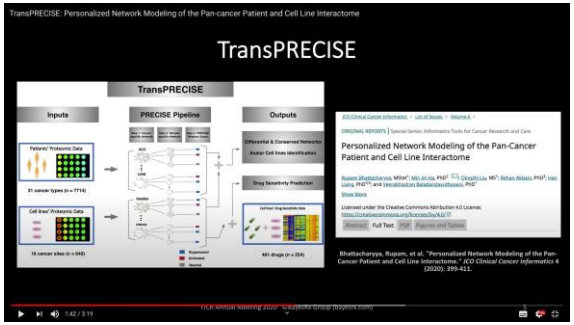
1. Network comparison across lineages and model systems.
2. Matching cell lines to patient profiles.
3. Predicting patient drug responses.

Bhattacharya et al (JCO, 2020)

Resources







Software: https://github.com/bayesrx
Shiny Apps: bayesrx.com

Summary

- Efficient harnessing of information from pre-clinical data
  - Potential uses: re-purpose existing drugs, IND, FDA-approved agents; drug-screening
  - Find potentially useful combination therapies
- Incorporate multi-omic data (e.g., epigenomics, proteomics, metabolomics, microbiome)
- Extension to other model systems e.g., Organoids
- Relax linearity/Gaussian assumptions; non-par Bayes!

If you can't convince them, confuse them.

– Harry Truman