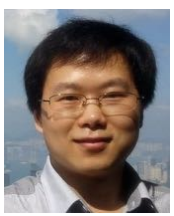
Multivariate structured regression for pharmacogenomic screens

Manuela Zucknick
University of Oslo

Centre for Statistical Methodology - Methods in Integrative Genomics
LSHTM, February 18, 2012

This is joint work with Zhi (George) Zhao, Marco Banterle and Alex Lewin.

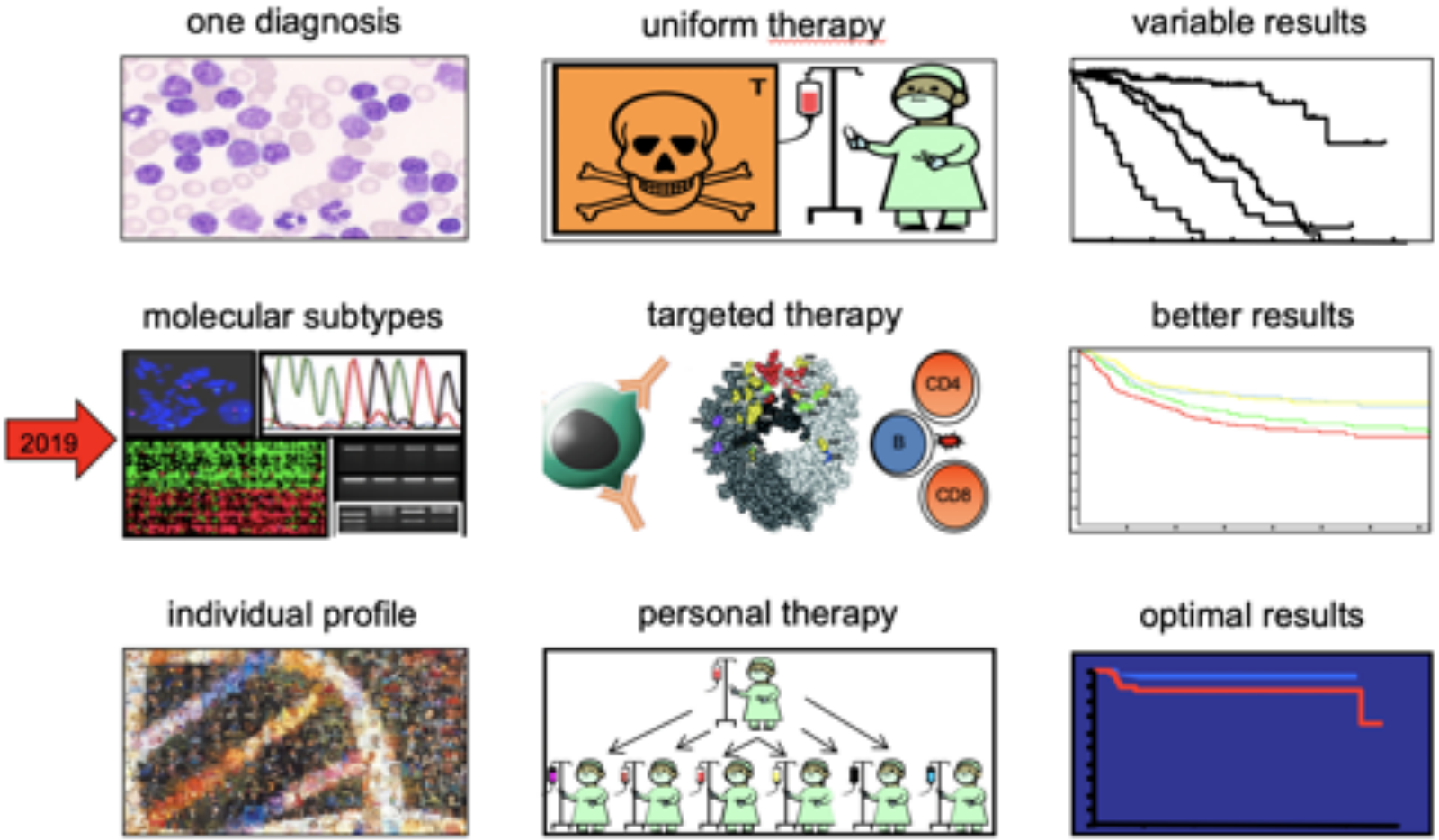


Outline

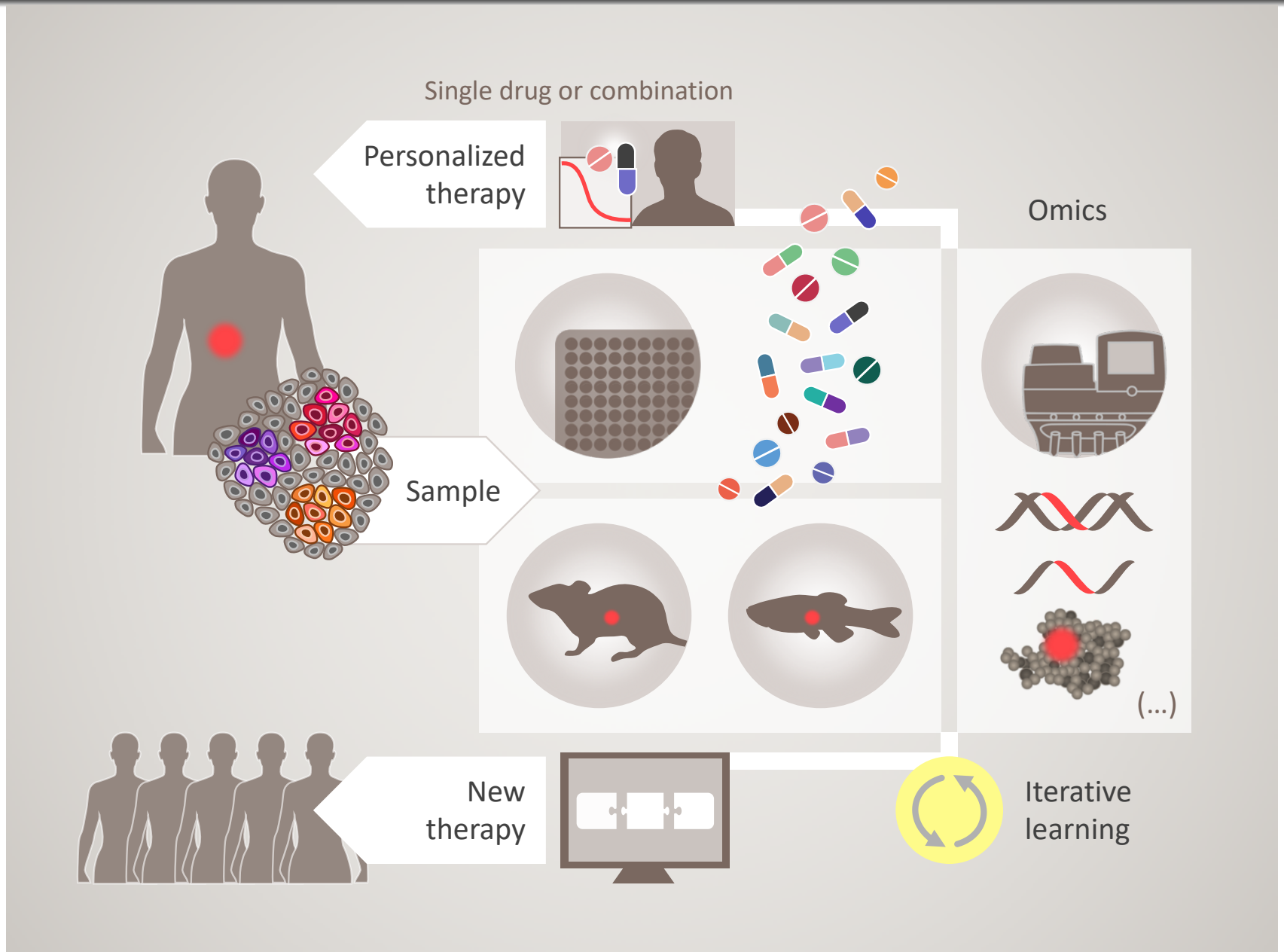
1. Inroduction: Integrative omics for personalized cancer therapy
2. Structured penalties for drug sensitivity prediction
3. Multivariate Bayesian variable selection for structured outcomes and high-dimensional features

Personalized cancer therapy

...aims to find the best therapy for each patient based on data about the patient and tumor (e.g. genomic data).



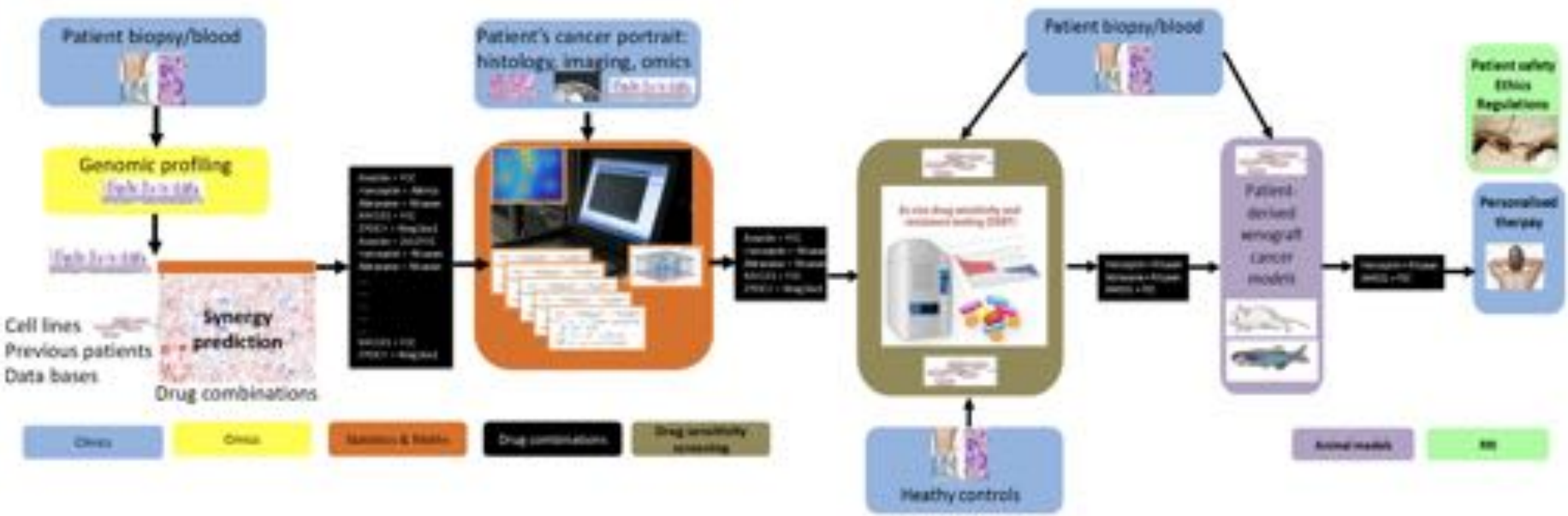
slide by Stephan Pfister



slide by Kjetil Taskén

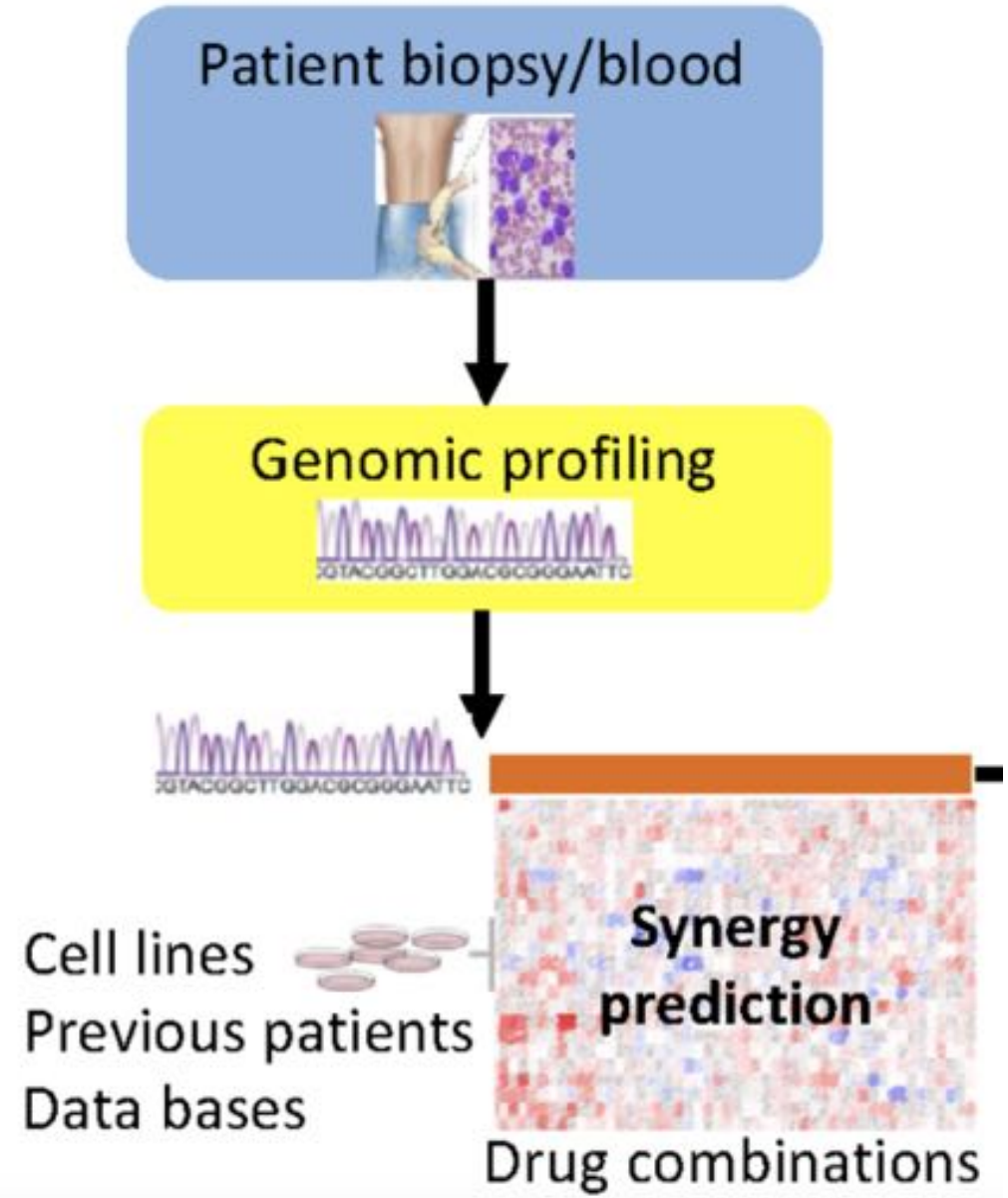
The PerCaThe project at the University of Oslo

- $>$ 25 key collaborators from many disciplines: math & stats, medicine, biology, computer science, physics, ethics
- Consortium lead:
  A Frigessi, K Taskén, V Kristensen, Å Helland, A Köhn-Luque



in vitro screening *in silico* modeling *ex vivo testing*

The PerCaThe project at the University of Oslo

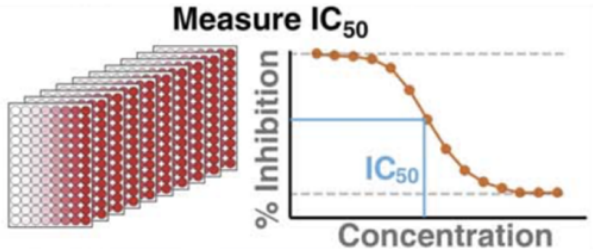


Predict sensitivity to multiple drugs $\boldsymbol{Y}$ from multi-omics $\boldsymbol{X}$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \epsilon$$

- **Drug dose response**

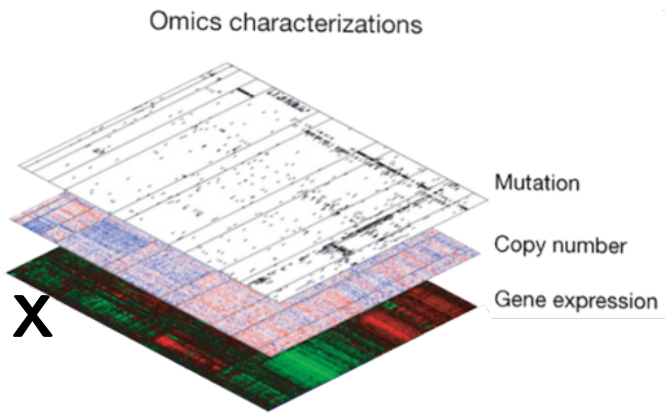
$$n \text{ cell lines} \overbrace{\begin{bmatrix} | & & | \\ \mathbf{y}_{\bullet 1} & \dots & \mathbf{y}_{\bullet m} \\ | & & | \end{bmatrix}}^{\text{drug sensitivity}} = \mathbf{Y}$$



Source: Yang, et al. 2017

- **Integrative omics**

$$n \text{ cell lines} \begin{bmatrix} \overbrace{\quad}^{\text{gene expression}} & \overbrace{\quad}^{\text{copy number}} & \overbrace{\quad}^{\text{mutation}} \\ \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 \end{bmatrix} = \mathbf{X}$$



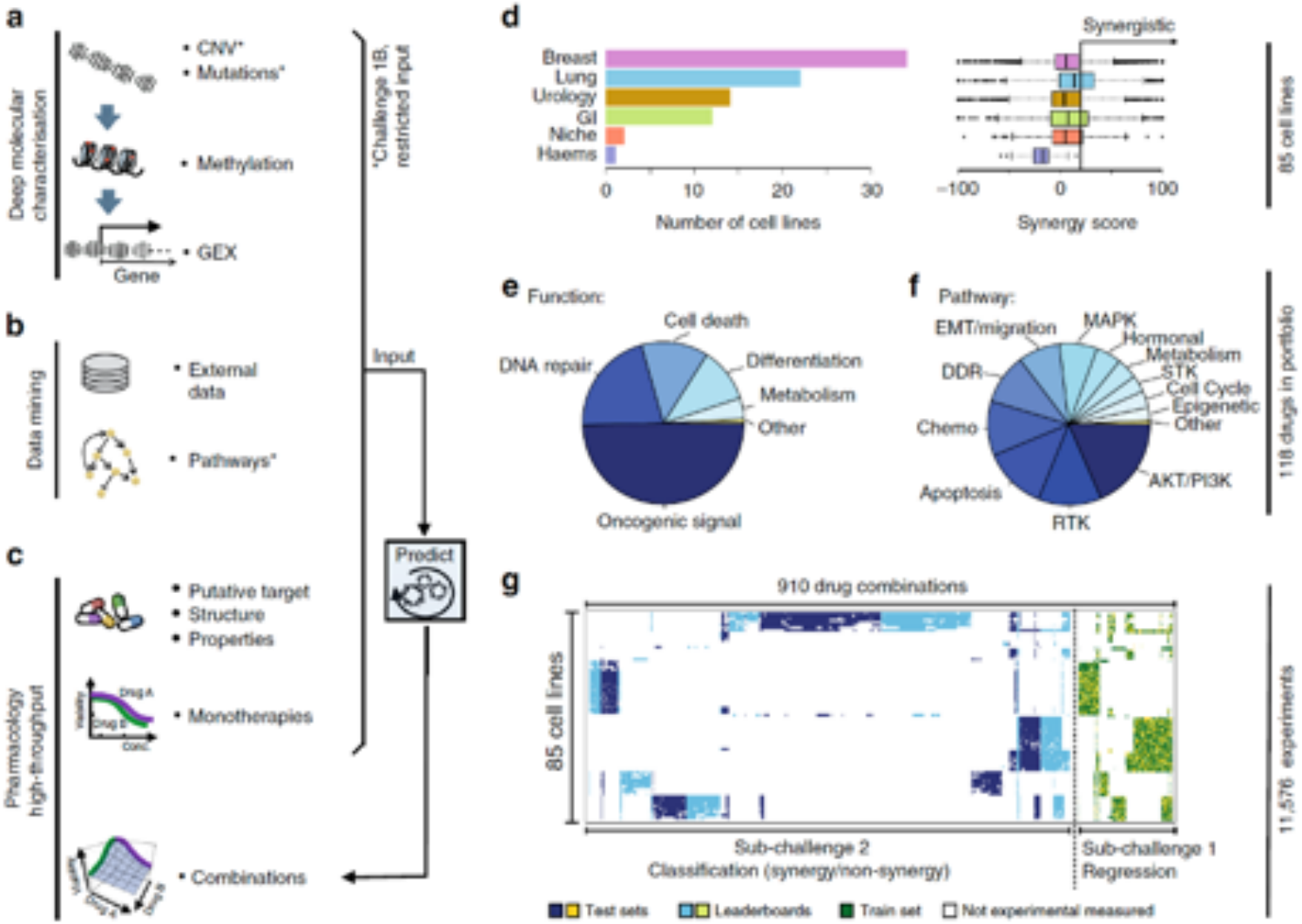
Source: TCGA, 2013

Predict sensitivity to multiple drugs $\boldsymbol{Y}$ from multi-omics $\boldsymbol{X}$

- Penalised regression and Bayesian hierarchical models
- Use structured penalties (for $\boldsymbol{B}$) or hyper-priors (for $\boldsymbol{B}$ or selection indicator $\boldsymbol{\Gamma}$ to:

- Borrow information between experiments
- Use prior knowledge about similarity between $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_k$

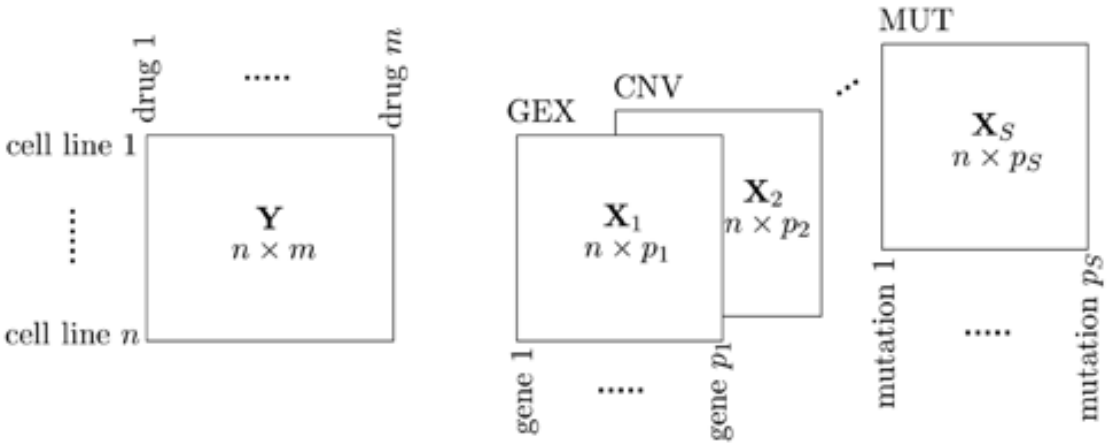
Typically available data (Menden et al. 2019, Figure 1*)



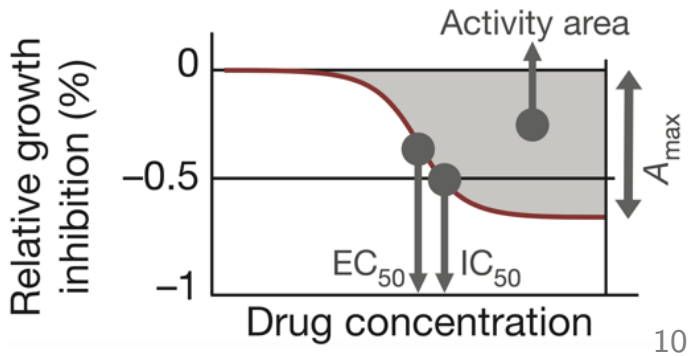
*DREAM AstraZeneca-Sanger Drug Combo Prediction Challenge

Challenges and opportunities (1)

- Small sample size
- Several types of input data $\boldsymbol{X}$:
  E.g., gene expression, copy number, mutation
- Multivariate response $\boldsymbol{Y}$



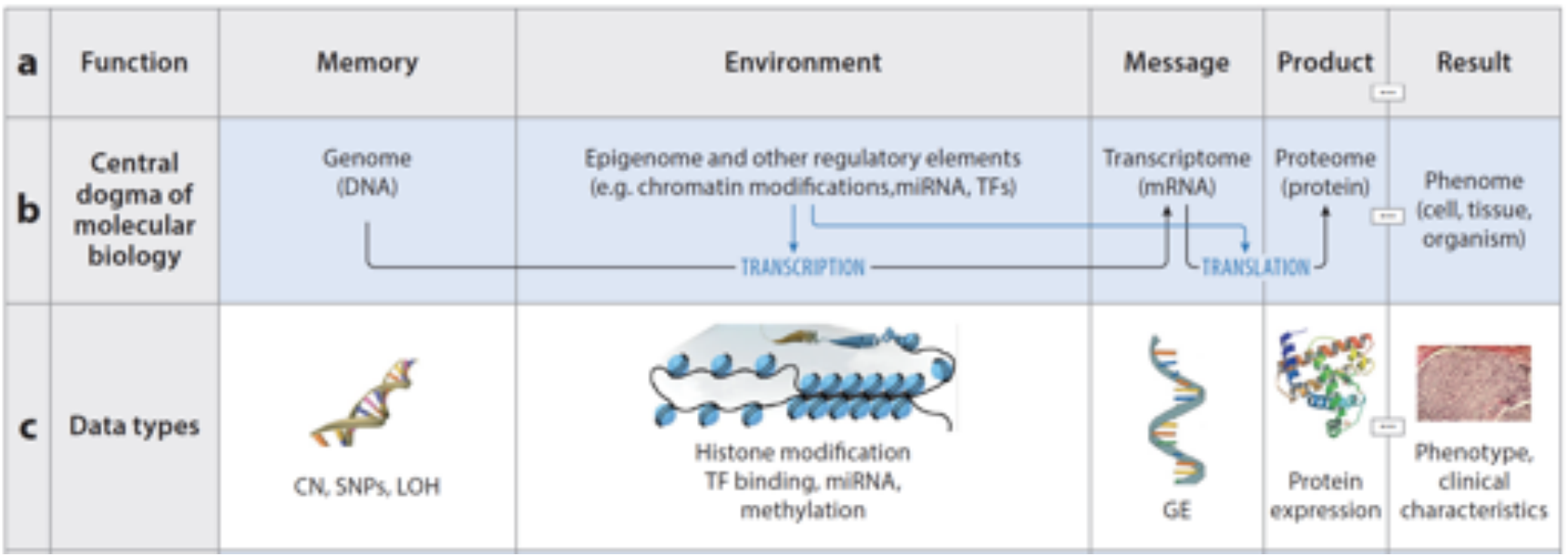
- Unclear how to define $\boldsymbol{Y}$



Challenges and opportunities (2)

The data are highly structured:

1. In $\boldsymbol{Y}$: relationships between drugs, e.g. due to similar chemical drug composition, same target genes/pathways
2. In $\boldsymbol{X}$: relationships between molecular data sources



Ickstadt et al. (2018)

Vertical integration: genes as the common biological units

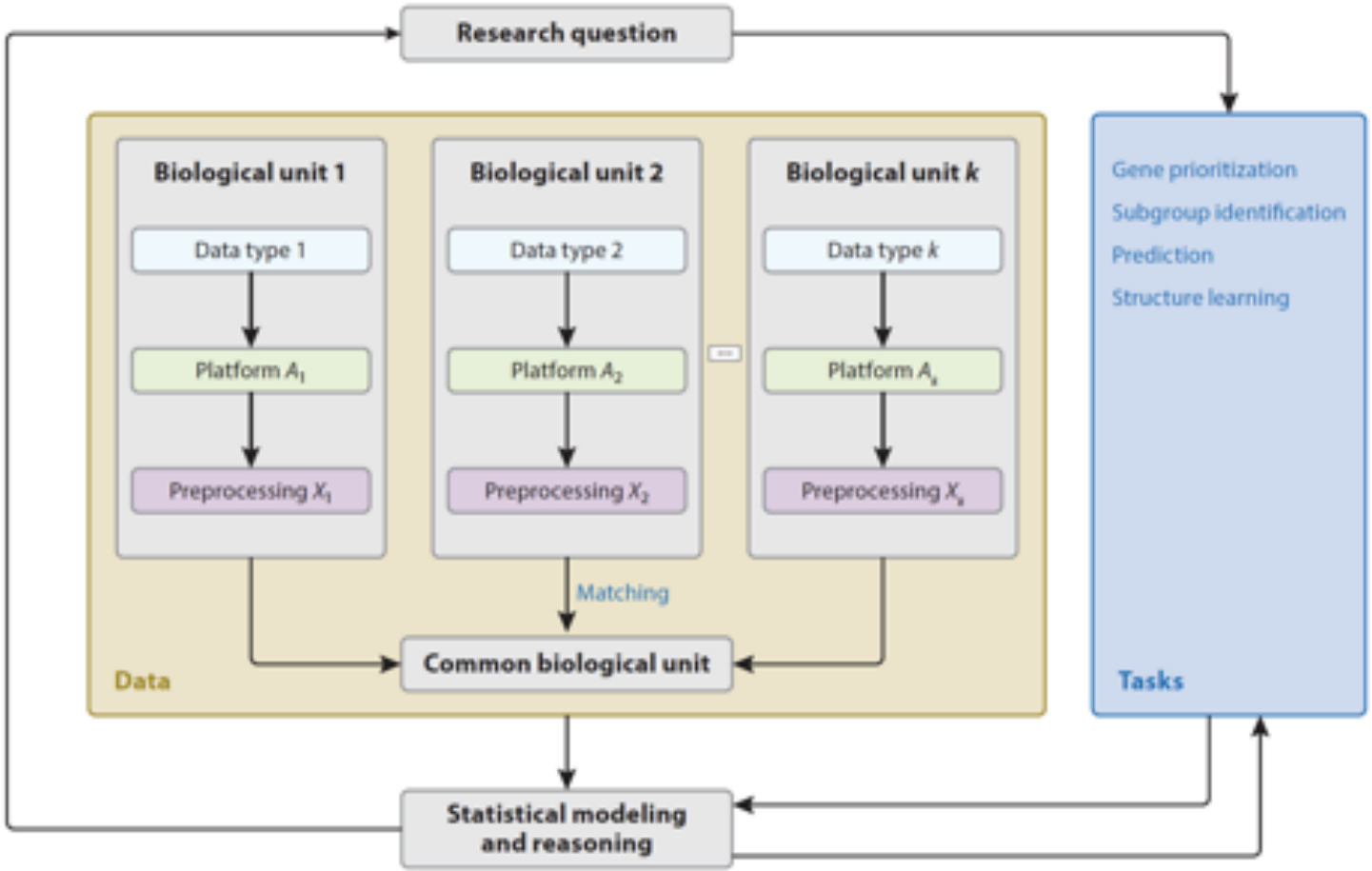


Figure 3

Integrative data analysis approach in molecular biology. Data types, platforms, and/or preprocessing methods may coincide in units $1, 2, \ldots, k$.

Ickstadt et al., 2018

Structured Penalized Regression for Drug Sensitivity Prediction

Zhi (George) Zhao, Manuela Zucknick (arXiv:1902.04996)

Multi-response penalised linear regression

Objective function:

$$\min_{\beta_0, \mathbf{B}} \left\{ \frac{1}{2mn} \|\mathbf{Y} - \mathbf{1}_n \beta_0^T - \mathbf{X}\mathbf{B}\|_F^2 + \text{pen}(\mathbf{B}) \right\}$$

Standard penalised regression assigns the same penalty to all data sources, and treats columns of $\boldsymbol{Y}$ as independent:

- **Lasso**: $\text{pen}(\mathbf{B}) = \lambda \|\mathbf{B}\|_{\ell_1}$
- **Elastic-net**: $\text{pen}(\mathbf{B}) = \lambda(\alpha \|\mathbf{B}\|_{\ell_1} + \frac{1}{2}(1-\alpha)\|\mathbf{B}\|_{\ell_2}^2)$

Integrative LASSO with Penalty Factors (Boulesteix et al. 2017)

- Allow different penalties for different data sources
- Extensions of IPF-lasso to multi-response regression and to the elastic net

$$\textbf{IPF-lasso: } \text{pen}(\mathbf{B}) = \sum_s \lambda_s \|\mathbf{B}_s\|_{\ell_1}$$

$$\textbf{IPF-sEN: } \text{pen}(\mathbf{B}) = \sum_s \lambda_s (\alpha \|\mathbf{B}_s\|_{\ell_1} + \frac{1}{2}(1-\alpha)\|\mathbf{B}_s\|_{\ell_2}^2)$$

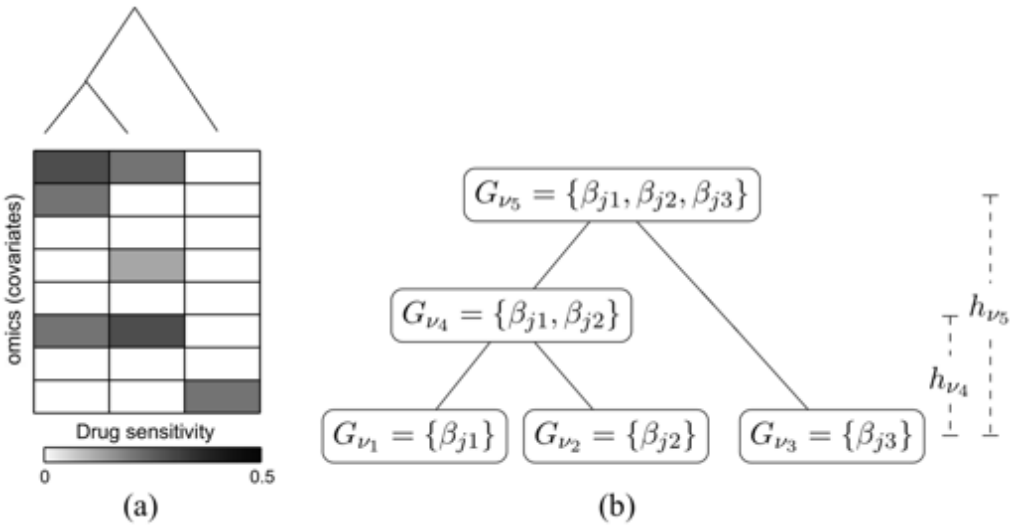
$$\textbf{IPF-EN: } \text{pen}(\mathbf{B}) = \sum_s \lambda_s (\alpha_s \|\mathbf{B}_s\|_{\ell_1} + \frac{1}{2}(1-\alpha_s)\|\mathbf{B}_s\|_{\ell_2}^2)$$

(Multi-response) Tree-guided group lasso (Kim & Xing 2012)

- Include dependencies between columns of $\boldsymbol{Y}$ in a group lasso
- Extension to IPF-tree lasso

$$\textbf{Tree lasso: } \text{pen}(\mathbf{B}) = \lambda \sum_{j=1}^{p} \sum_{\nu \in \{V_{\text{int}}, V_{\text{leaf}}\}} \omega_\nu \|\beta_j^{G_\nu}\|_{\ell_2}$$

$$\textbf{IPF-tree lasso: } \text{pen}(\mathbf{B}) = \sum_{s} \lambda_s \left(\sum_{j_s} \sum_{\nu \in \{V_{\text{int}}, V_{\text{leaf}}\}} \omega_\nu \|\beta_{j_s}^{G_\nu}\|_{\ell_2} \right)$$

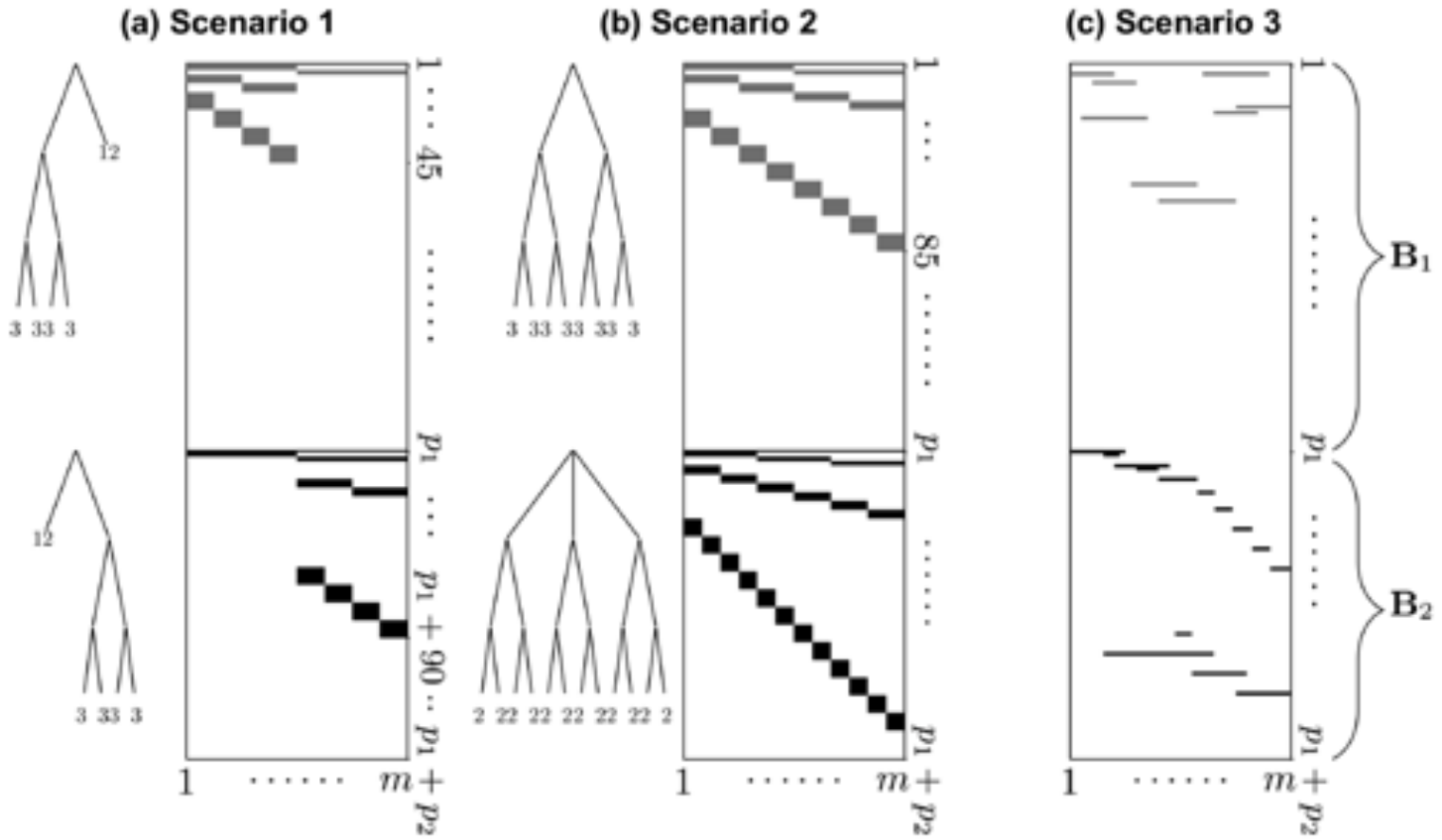


Simulation setup (similar to Boulesteix et al. 2017)

- Two data sources ($p_1 > n, p_2 > n$).

$$\mathbf{Y} = [\mathbf{X}_1, \tilde{\mathbf{X}}_2] \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} + \mathbf{E}.$$

- Let the final $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ after dichotomizing $\mathbf{X}_2 = \mathbf{1}_{\{\tilde{\mathbf{x}}_2 > 0\}}$.
- Coefficient matrix $\mathbf{B}$:



Some simulation results (m=24 drugs, m=100 samples)

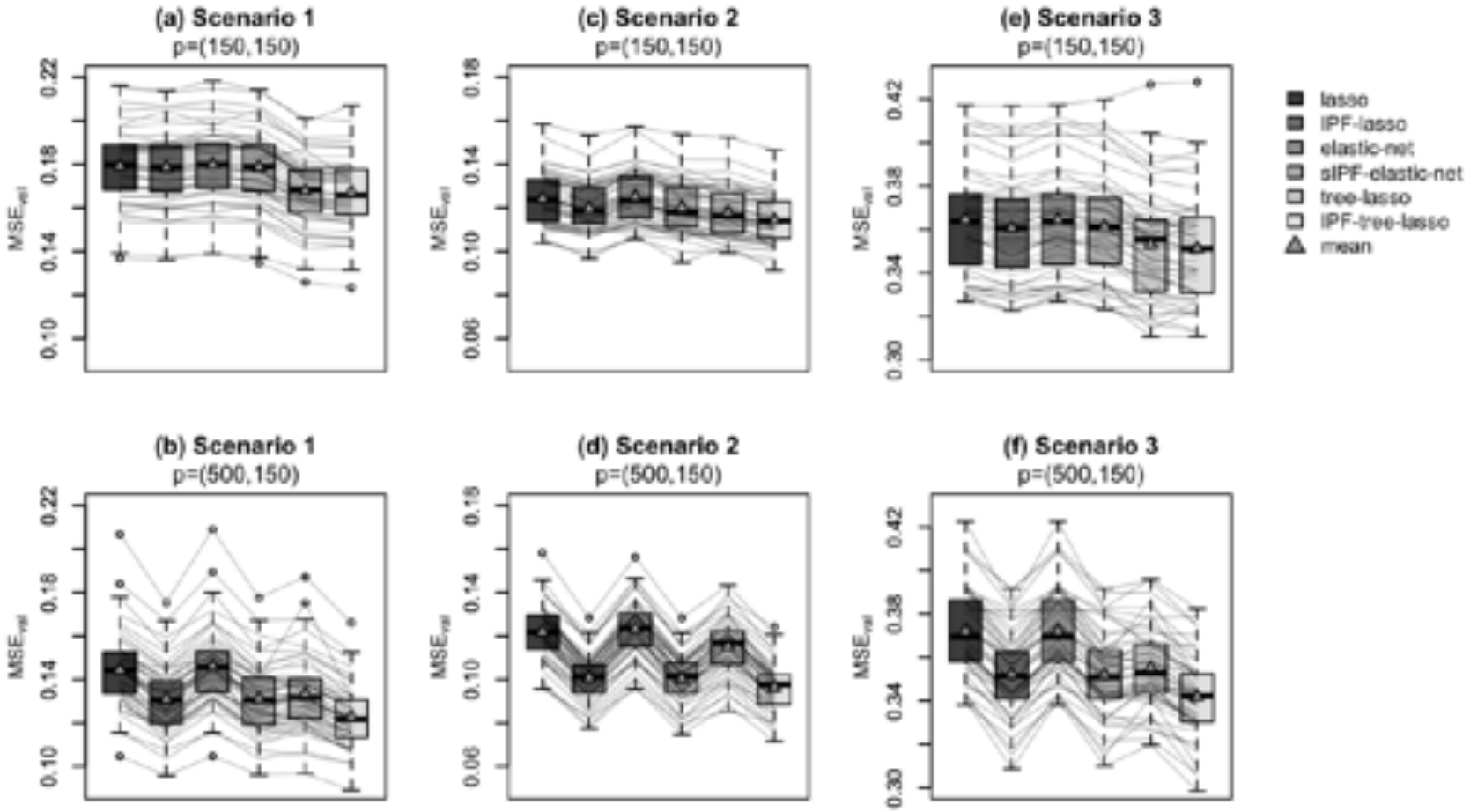


Fig 5: Comparison of MSE_{val} between different approaches (summary of 50 simulations runs). The three top panels are based on $p_1 = p_2 = 150$. The three bottom panels are based on $p_1 = 500$ and $p_2 = 150$.

Application to Genomics of Drug Sensitivity in Cancer data

(Garnett et al., 2012)

- Large-scale pharmacogenomic study with n=498 cell lines and m=97 drugs.
- Outcome data: $\log(IC_{50})$ from dose-response experiments
- Random draws of 80% cell lines as training data and 20% as validation data.
- Input data:
  - cancer type ($p_0 = 13$)
    $\rightarrow$ mandatory covariates not included in the penalty term,
  - mRNA expression ($p_1 = 2602$),
  - copy numbers ($p_2 = 426$) and
  - DNA mutations ($p_3 = 68$)

Results: Average performance across all drugs

TABLE 2
Prediction and the numbers of selected features in the GDSC data analysis

Method	NULL[‡‡]	Lasso	elastic net	Tree-lasso
$VS^{\star\dagger}$	-	302+1+92[♯]	315+1+93	21928+8149+1
$\frac{1}{mp}VS^{\star\ddagger}$	-	0.1%	0.1%	10.0%
MSE_{CV} (SD)[§]	3.360(0.027)	3.200 (0.040)	3.198 (0.039)	3.138 (0.040)
MSE_{val} (SD)	3.368 (0.107)	3.151 (0.077)	3.149 (0.077)	3.069 (0.079)
R^2_{val}[¶] (SD)	-0.014 (0.008)	0.051 (0.012)	0.052 (0.014)	0.076 (0.019)
	OLS	IPF-lasso	sIPF-elastic-net	IPF-tree-lasso
$VS^{\star}$	-	774+11+74	252394+41322+6596	30567+515+452
$\frac{1}{mp}VS^{\star}$	-	0.3%	100.0%	10.5%
MSE_{CV} (SD)	3.013 (0.016)	3.182 (0.037)	3.179 (0.036)	3.068 (0.035)
MSE_{val} (SD)	3.199 (0.074)	3.134 (0.078)	3.130 (0.076)	3.025 (0.074)
R^2_{val} (SD)	0.036 (0.016)	0.056 (0.014)	0.057 (0.015)	0.089 (0.018)

Results: MSE_{val} of individual drugs

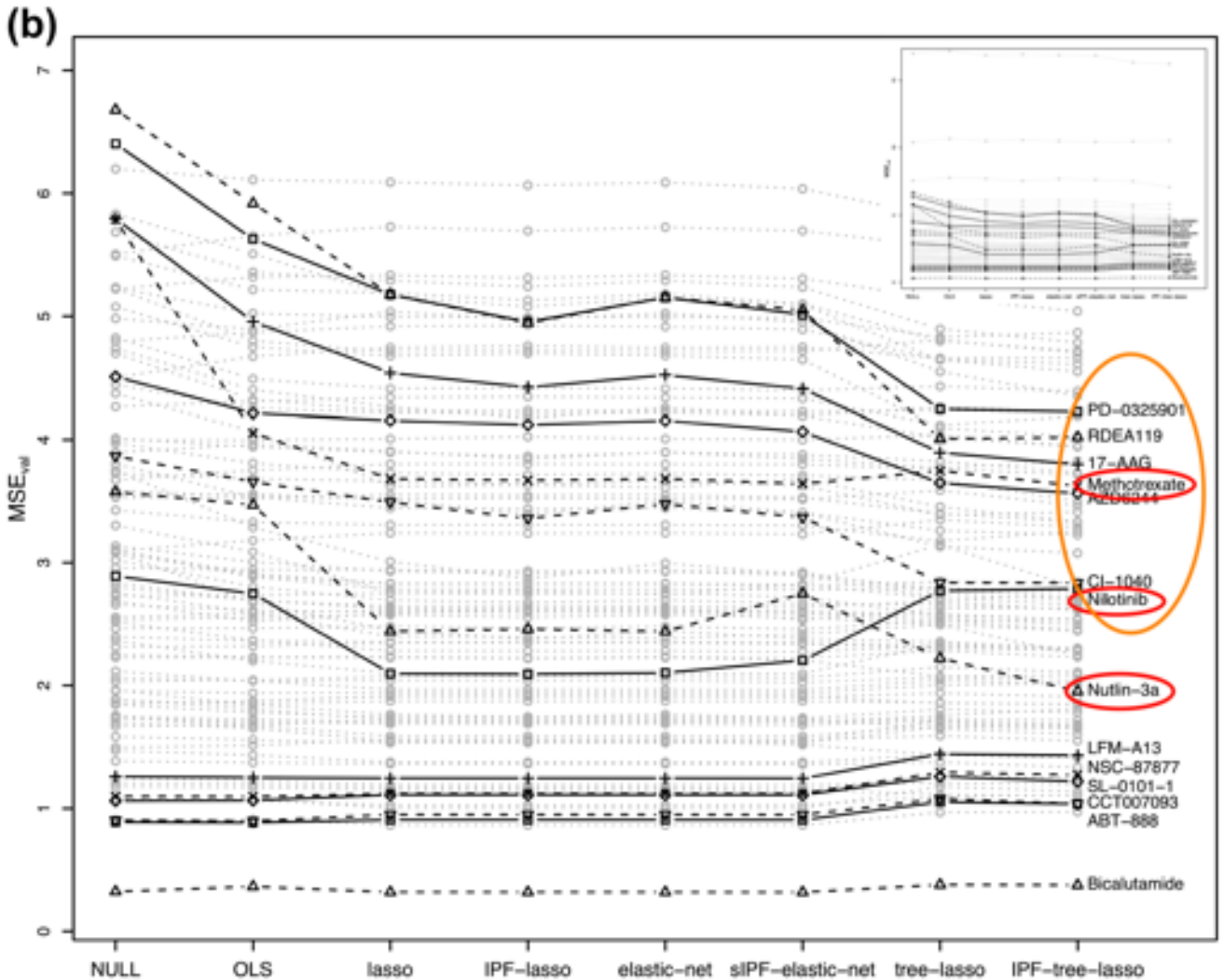


Fig 6: MSE_{val} of individual drugs (average of 10 random training-test splits).

Results: Some examples

- **Nutlin-3α:**
  mechanism of action involves the p53-pathway, affected in ca. 50% of cancers in most cancer types ($R^2_{\text{val}} = 0.45$)
- **PD-0325901, RDEA119, CI-1040, AZD6244:**
  MEK1 inhibitors with highly correlated IC_{50} values.
- **Methotrexate:**
  general cytotoxic drug not targeted to specific genes/pathways
- **Nilotinib:**
  inhibits the BCR-ABL fusion gene characteristic for chronic myeloid leukemia. Related to Axitinib (smaller effect).

IPFStructPenalty summary

- Combination of IPF-penalty (for multiple omics data sources) and tree-lasso (for hierarchical correlation structure in drug responses) to IPF-tree-lasso
- Development of IPF-elastic net

- **Computational aspects:** joint optimisation of > 2 penality parameters is challenging.
- We employ the Interval-search algorithm EPSGO (Frohlich & Zell 2005; Sill et al. 2014): learning a Gaussian process model of loss function surface from visited points
- `https://github.com/zhizuio/IPFStructPenalty`

- **Advantage:** simple implementation, computational speed
- **Disadvantage:** limited options for manipulation of penalty

Multivariate Bayesian variable selection for structured outcomes and high-dimensional features:

Structured seemingly unrelated regression with MRF priors

Zhi Zhao, Marco Banterle, Alex Lewin, Manuela Zucknick

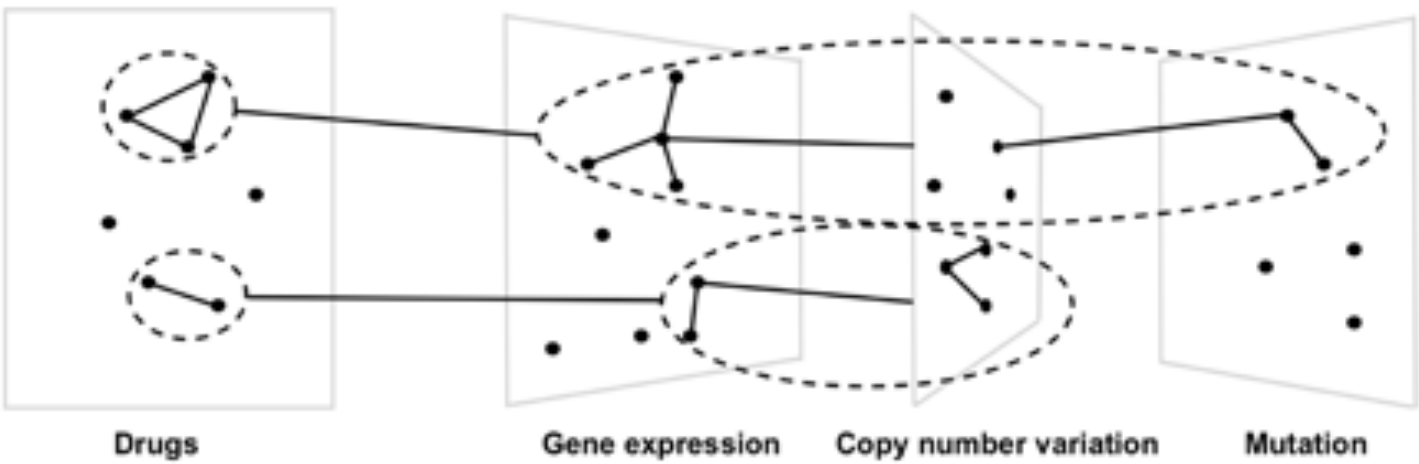


Figure 1: Illustration of the drug groups and omics path (reproduced from [Ruffieux, 2019]).

BayesSUR (https://CRAN.R-project.org/package=BayesSUR)

- Bayesian seemingly unrelated regression for variable and covariance selection (Banterle et al., 2018)

- **Matrix formulation of the model:**

$$\mathbf{Y} = \mathbf{X}\boldsymbol{B} + \mathbf{U},$$
$$\text{vec}(\mathbf{U}) \sim \mathcal{N}(\mathbf{0},\ C \otimes \mathbb{I}_n)$$

  - $\mathbf{Y}$ $n \times m$ matrix of outcomes with $m \times m$ covariance matrix C,
  - $\mathbf{X}$ $n \times p$ matrix of predictors for all outcomes,
  - $\boldsymbol{B}$ $p \times m$ matrix of regression coefficients.

- **Introduce sparsity:**

$$\beta_{kj}|\gamma_{kj}, w \sim \gamma_{kj}\mathcal{N}(0,\ w) + (1-\gamma_{kj})\delta_0(\beta_{kj})$$

  - Binary latent indicator matrix $\Gamma = \{\gamma_{jk}\}$ for variable selection
  - Spike-and-slab prior on vectorised $\boldsymbol{\beta} = \text{vec}(\boldsymbol{B})$ and $\boldsymbol{\gamma} = \text{vec}(\Gamma)$
  - and $w \sim \mathcal{IG}(a_w, b_w)$ and $\delta_0(\cdot)$ is the Dirac delta function.

BayesSUR (https://CRAN.R-project.org/package=BayesSUR)

- Bayesian seemingly unrelated regression for variable and covariance selection (Banterle et al., 2018)

- **Matrix formulation of the model:**

$$\mathbf{Y} = \mathbf{X}\boldsymbol{B} + \mathbf{U},$$
$$\text{vec}(\mathbf{U}) \sim \mathcal{N}(\mathbf{0},\ C \otimes \mathbb{I}_n)$$

  - $\mathbf{Y}$ $n \times m$ matrix of outcomes with $m \times m$ covariance matrix C,
  - $\mathbf{X}$ $n \times p$ matrix of predictors for all outcomes,
  - $\boldsymbol{B}$ $p \times m$ matrix of regression coefficients.

- **Introduce sparsity:**

$$\beta_{kj}|\gamma_{kj}, w \sim \gamma_{kj}\mathcal{N}(0,\ w) + (1-\gamma_{kj})\delta_0(\beta_{kj})$$

  - Binary latent indicator matrix $\Gamma = \{\gamma_{jk}\}$ for variable selection
  - Spike-and-slab prior on vectorised $\boldsymbol{\beta} = \text{vec}(\boldsymbol{B})$ and $\boldsymbol{\gamma} = \text{vec}(\Gamma)$
  - and $w \sim \mathcal{IG}(a_w, b_w)$ and $\delta_0(\cdot)$ is the Dirac delta function.

BayesSUR (2)

	$\gamma_{jk} \sim$ Bernoulli	$\gamma_{jk} \sim$ Hotspot	$\boldsymbol{\gamma} \sim$ MRF
$C \sim$ indep	HRR-B	HRR-H	HRR-M
$C \sim \mathcal{IW}$	dSUR-B	dSUR-H	dSUR-M
$C \sim \mathcal{HIW}_{\mathcal{G}}$	SSUR-B	SSUR-H	SSUR-M

We can introduce structure/ sparsity in two places:

1. Prior for covariance matrix: $C \sim \mathcal{HIW}_{\mathcal{G}}$ with further hyper-prior on graph $\mathcal{G}$ (Banterle et al. 2018)
   - Graph $\mathcal{G}$ encodes conditional dependence between responses. Sparse $\mathcal{G}$ implies sparse precision matrix C^{-1}.
   - Sparse Seemingly Unrelated Regression (SSUR)
2. Prior for variable selection indicator $\boldsymbol{\gamma}$.
   - Sparsity: which covariates are associated with which responses.

BayesSUR (3)

Options for covariance matrix structure (Banterle et al. 2018)

- diagonal: Hierarchical Related Regression (Richardson et al 2011)
- dense: dense Seemingly Unrelated Regressions (dSUR)
- sparse: Sparse Seemingly Unrelated Regressions (SSUR)

Options for variable selection ($j = 1, \cdots, p, k = 1, \cdots, m$)

- Independent Bernoulli prior:

$$\gamma_{jk}|\omega_{jk} \sim \mathcal{B}er(\omega_j), \quad \text{with } \omega_j \sim \mathcal{B}eta(a_\omega, b_\omega).$$

- Hotspot prior:

$$\gamma_{jk}|\omega_{jk} \sim \mathcal{B}er(\omega_{jk}), \quad \text{with } \omega_{jk} = o_k \times \pi_j,$$
$$o_k \sim \mathcal{B}eta(a_o, b_o), \pi_j \sim \mathcal{G}amma(a_\pi, b_\pi).$$

- Markov Random Field (MRF) prior:

$$f(\gamma|d, e, G) \propto \exp\{d\mathbf{1}^\top\gamma + e \cdot \gamma^\top G\gamma\}$$

BayesSUR (3)

Options for covariance matrix structure (Banterle et al. 2018)

- diagonal: Hierarchical Related Regression (Richardson et al 2011)
- dense: dense Seemingly Unrelated Regressions (dSUR)
- sparse: Sparse Seemingly Unrelated Regressions (SSUR)

Options for variable selection ($j = 1, \cdots, p, k = 1, \cdots, m$)

- Independent Bernoulli prior:

$$\gamma_{jk}|\omega_{jk} \sim \mathcal{B}er(\omega_j), \quad \text{with } \omega_j \sim \mathcal{B}eta(a_\omega, b_\omega).$$

- Hotspot prior:

$$\gamma_{jk}|\omega_{jk} \sim \mathcal{B}er(\omega_{jk}), \quad \text{with } \omega_{jk} = o_k \times \pi_j,$$
$$o_k \sim \mathcal{B}eta(a_o, b_o), \pi_j \sim \mathcal{G}amma(a_\pi, b_\pi).$$

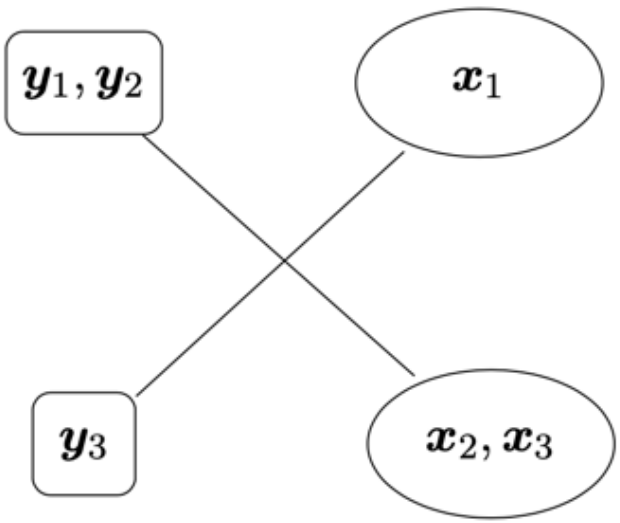
- Markov Random Field (MRF) prior:

$$f(\boldsymbol{\gamma}|d, e, G) \propto \exp\{d\mathbf{1}^\top\boldsymbol{\gamma} + e \cdot \boldsymbol{\gamma}^\top G\boldsymbol{\gamma}\}$$

MRF prior for pharmacogenomics

$$f(\gamma|d, e, G) \propto \exp\{d\mathbf{1}^\top\gamma + e \cdot \gamma^\top G\gamma\}$$

- d controls the model sparsity,
- e the strength of relations between responses and predictors,
- G is an adjacency matrix of the structure prior knowledge.

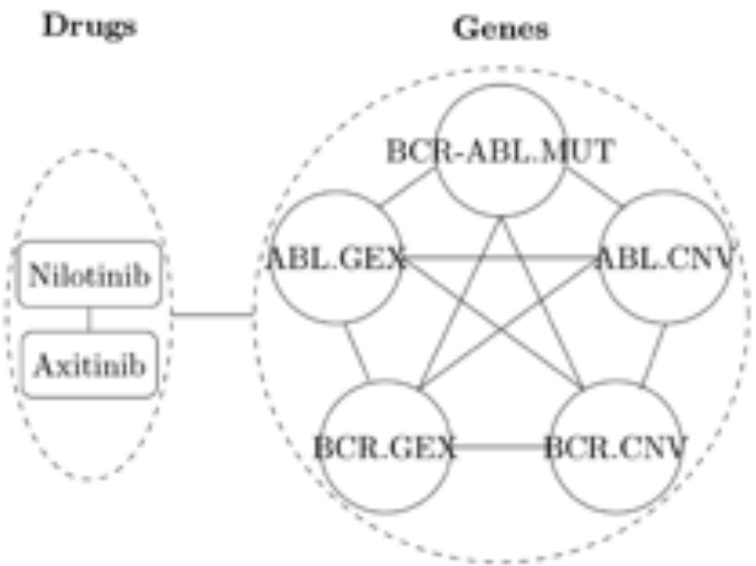


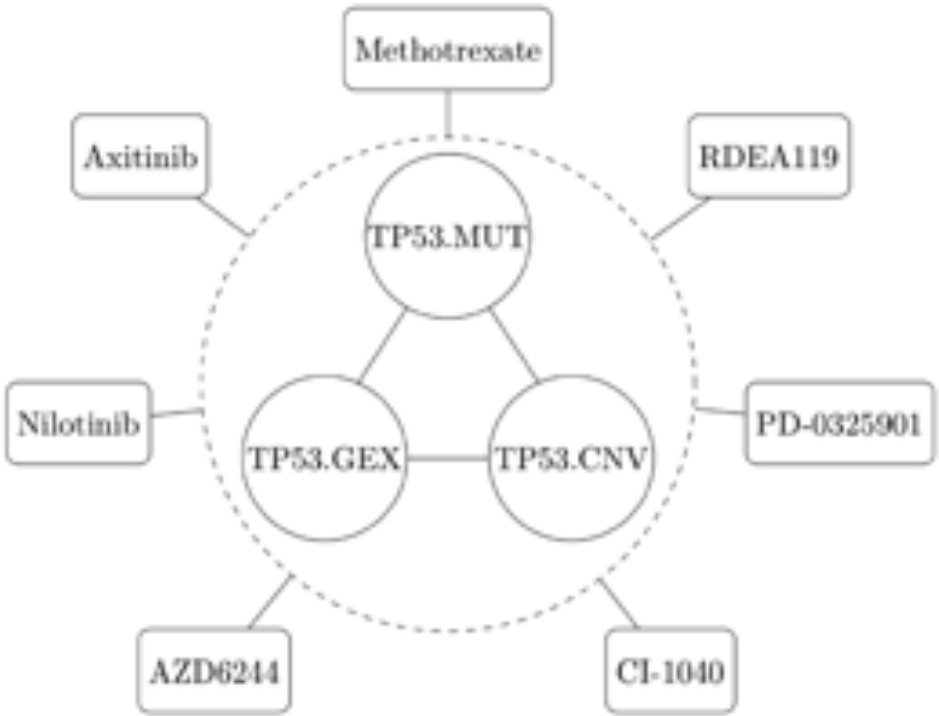
$$G = \begin{array}{c} \\ \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \\ \gamma_{12} \\ \gamma_{22} \\ \gamma_{32} \\ \gamma_{13} \\ \gamma_{23} \\ \gamma_{33} \end{array} \begin{array}{c} \begin{array}{ccccccccc} \gamma_{11} & \gamma_{21} & \gamma_{31} & \gamma_{12} & \gamma_{22} & \gamma_{32} & \gamma_{13} & \gamma_{23} & \gamma_{33} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Application to Genomics of Drug Sensitivity in Cancer data

- Same data as before, but now only use $m = 7$ cancer drugs
- **Use MRF prior to include structure**, with edges between:
  - drugs: Group1 ("RDEA119","PD-0325901","CI-1040" and "AZD6244"); Group2 ("Nilotinib","Axitinib")
  - genes in MAPK/ERK pathway (target of Group1)
  - genes in the Bcr-Abl fusion gene (target of Group2)
  - genes of MAPK/ERK pathway and Group1
  - genes of the Bcr-Abl fusion gene and Group2
  - each gene feature in different data sources (GEX, CNV, MUT)

Application to Genomics of Drug Sensitivity in Cancer data





Results (Γ): Which covariates are important?

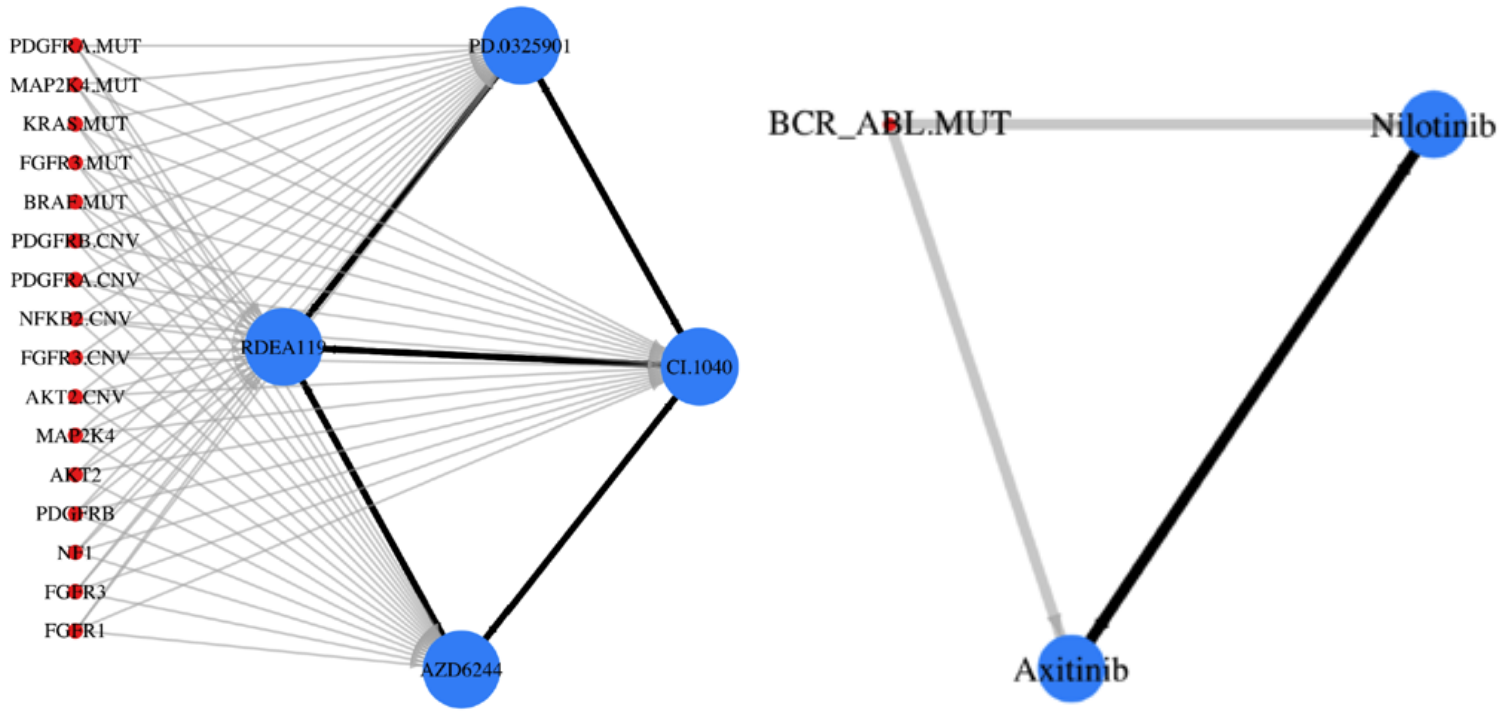


Fig: Important covariates related to the MEK inhibitors (left) or Bcr-Abl inhibitors (right) based on threshold for posterior marginal inclusion probabilities (mPIP > 0.5).

Results ($\mathcal{G}$): Residual covariance structure between drugs

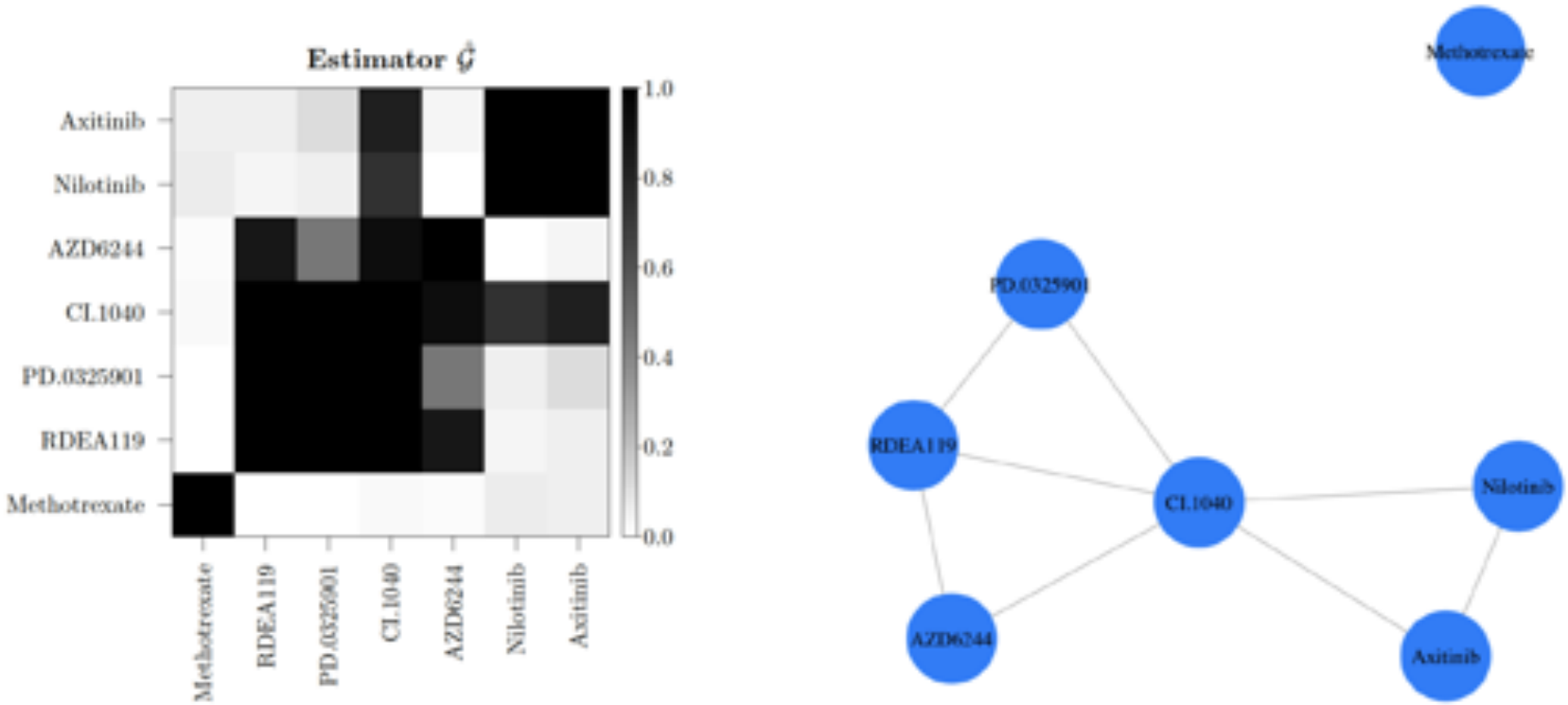


Fig: Posterior mean of $\mathcal{G}$ (left) and resulting graph based on threshold for marginal posterior edge inclusion probabilities (mEPIP) > 0.5 (right).

BayesSUR summary

- Development of "Structured Seemingly Unrelated Regression" (SSUR-M) as one model in a class of flexible multivariate linear models with variable and covariance selection.
- A unified, efficient and user-friendly implementation of all these models in the R package BayesSUR using Evolutionary Stochastic Search (ESS).
- Very flexible options for implementation of structure in covariance matrix C and MRF prior for Γ
- `https://CRAN.R-project.org/package=BayesSUR`
- See Banterle et al (2018) and `vignette("BayesSUR")` for the model details.


Ickstadt K, Schäfer M, Zucknick M (2018). Toward integrative Bayesian analysis in molecular biology. ARSIA. 5:141–167.


Menden MP, et al (2019). Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. Nature Communications 10:2674.


Boulesteix AL, et al. (2017). IPF-LASSO: Integrative-Penalized Regression with Penalty Factors for Prediction Based on Multi-Omics Data. Comput Math Methods Med.


Kim S and Xing E (2012). Ann Appl Stat. 6:1095–1117.


Sill M, Hielscher T, Becker N, Zucknick M (2014). c060: Extended inference with lasso and elastic-net regularized Cox and generalized linear models. JSS 62:1–22.


Zhao Z, Zucknick M (2019). Structured Penalized Regression for Drug Sensitivity Prediction. JRSSC, accepted. arXiv:1902.04996v.


Banterle M, Bottolo L, Richardson S, Ala-Korpela M, Järvelin MR, Lewin A (2018). Sparse variable and covariance selection for high-dimensional seemingly unrelated Bayesian regression. BioRxiv, 10.1101/467019.


Banterle M, Zhao Z, Bottolo L, Richardson S, Lewin A, Zucknick M (2019). BayesSUR: An R package for high-dimensional multivariate Bayesian variable and covariance selection in linear regression. `https://CRAN.R-project.org/package=BayesSUR`.

Thank you!