Evaluation form

The potential of Bayesian modelling for complex biomedical data

Oslo Bioinformatics Workshop Week

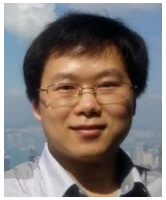
12 December 2025

Manuela Zucknick

Oslo Centre for Biostatistics and Epidemiology, UiO, OUS

www.med.uio.no/imb/english/research/groups/stat-learn-mol-med

Leiv Rønneberg

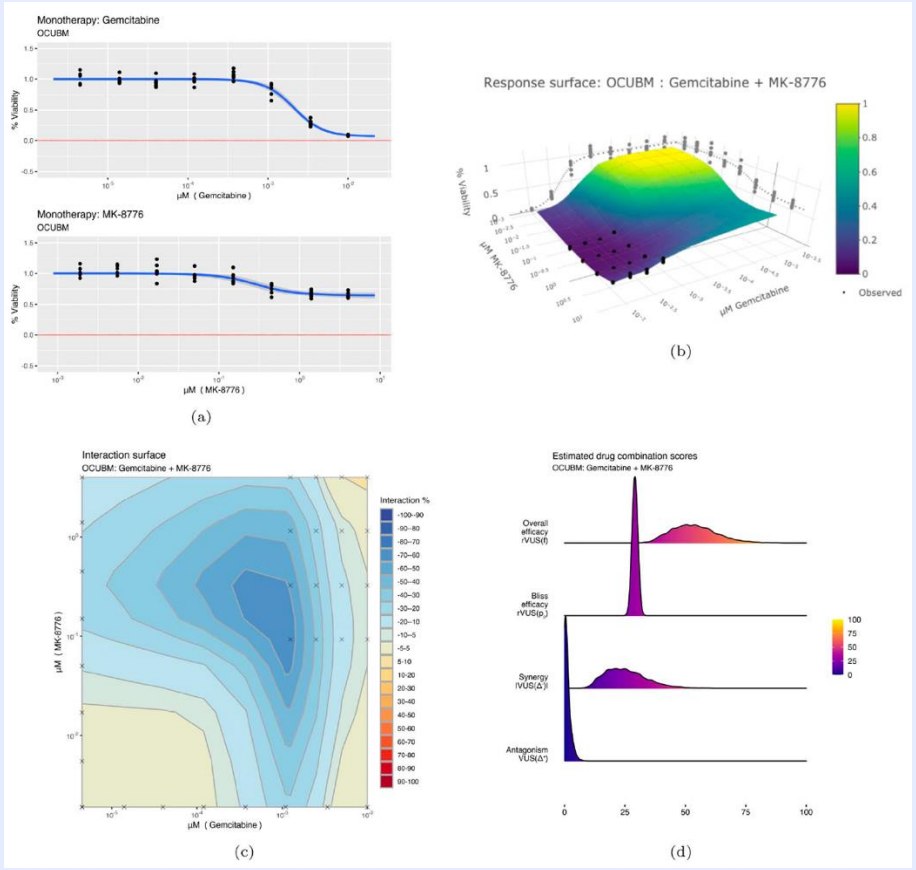


Zhi Zhao

Theo Asenso

UNIVERSITY OF OSLO

Oslo universitetssykehus



Rønneberg et al. (2021), Fig 6

Going back to the basics:
Statistics is the science of uncertainty

Statistical science is the systematic discipline that

- collects, organizes, analyzes, interprets, and presents data in order to
- draw reliable conclusions and
- make informed decisions **under uncertainty**.

Bayesian inference

"Bayesian inference is the process of updating probabilities for hypotheses as more evidence or information becomes available, using **Bayes' theorem** to combine prior beliefs with the likelihood of observed data."

(Gelman et al, 2023. Bayesian Data Analysis, 3rd ed.)

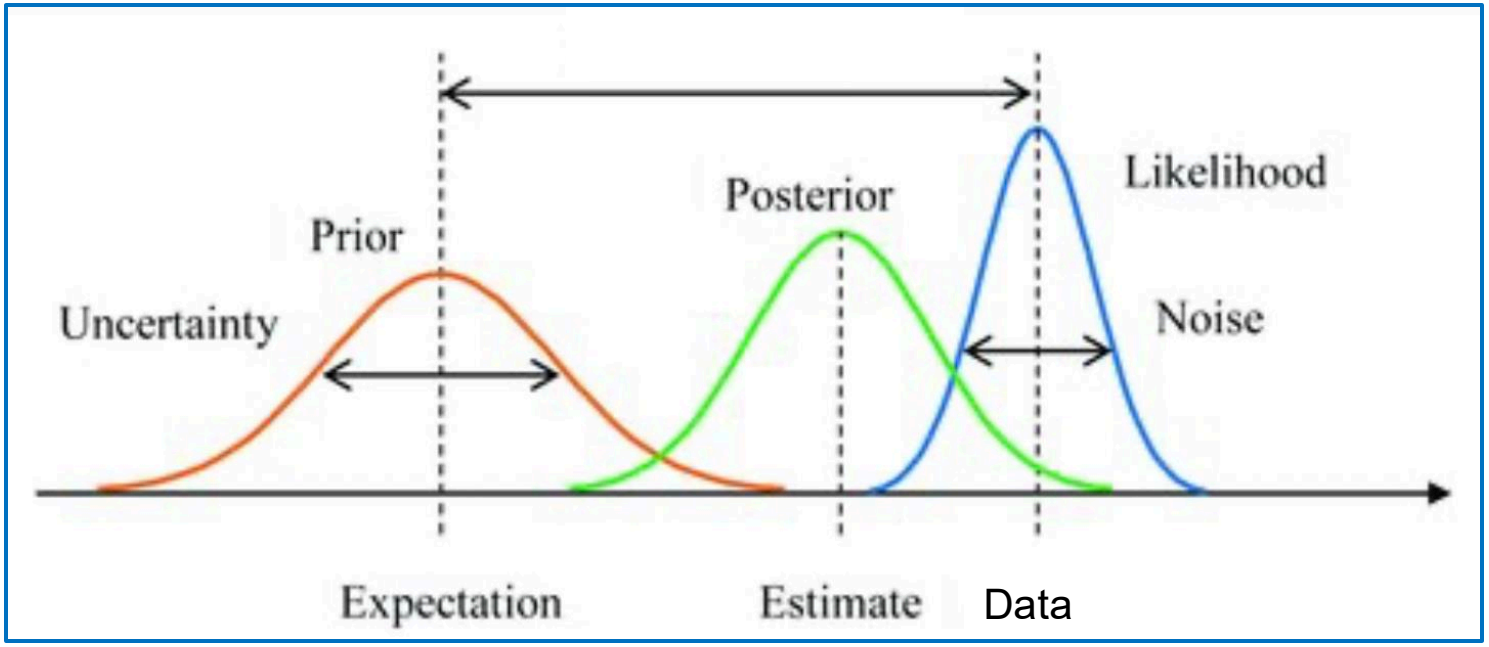
$p(\theta)$ prior distribution for the unknown parameter θ,

$p(x \mid \theta)$ likelihood (sampling model) of the observed data x,

$$p(\theta \mid x) = \frac{p(x \mid \theta)\, p(\theta)}{\int_{\Theta} p(x \mid \vartheta)\, p(\vartheta)\, d\vartheta} \text{ posterior distribution.}$$

Bayesian inference

- Provides a posterior distribution
- Instead of a point estimate (with or without confidence interval) as in frequentist statistics.
- A posterior can become a prior in a new model (with new data).
- Easy to build complex models, which propagate uncertainty consistently throughout all hierarchical layers of the model.



Adapted from DOI: 10.3389/fncom.2019.00002

Bayesian hierarchical modelling

- Model related individual experiments together.
- Assume that the individual-model parameters are related and all come from a joint distribution (= hyper-prior distribution).
- **Borrow information** between individual experiments and achieve more precise results (reduce uncertainty).
- Example: Drug screen with several drugs (and cell lines)

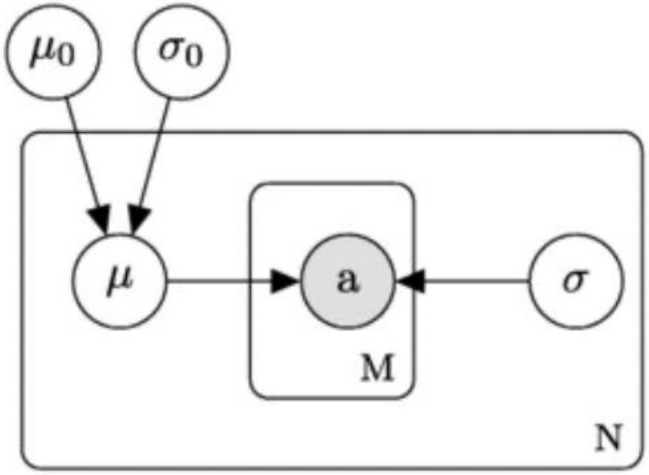


Illustration: Hierarchial model with 2 layers

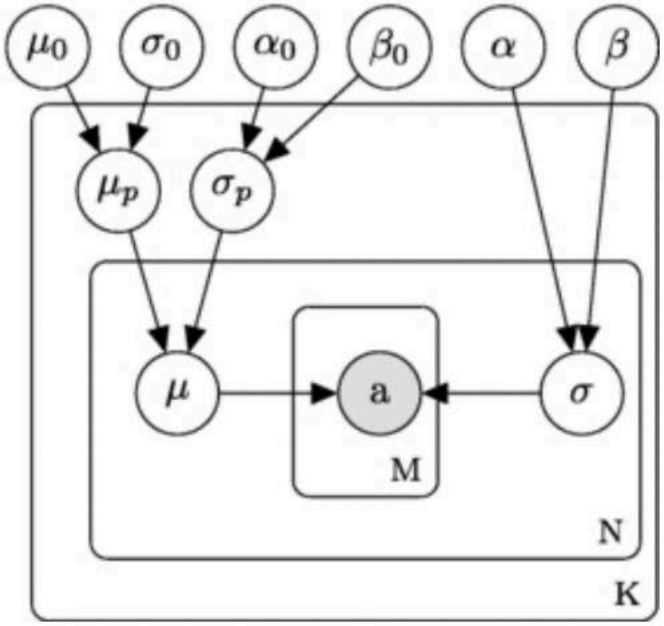
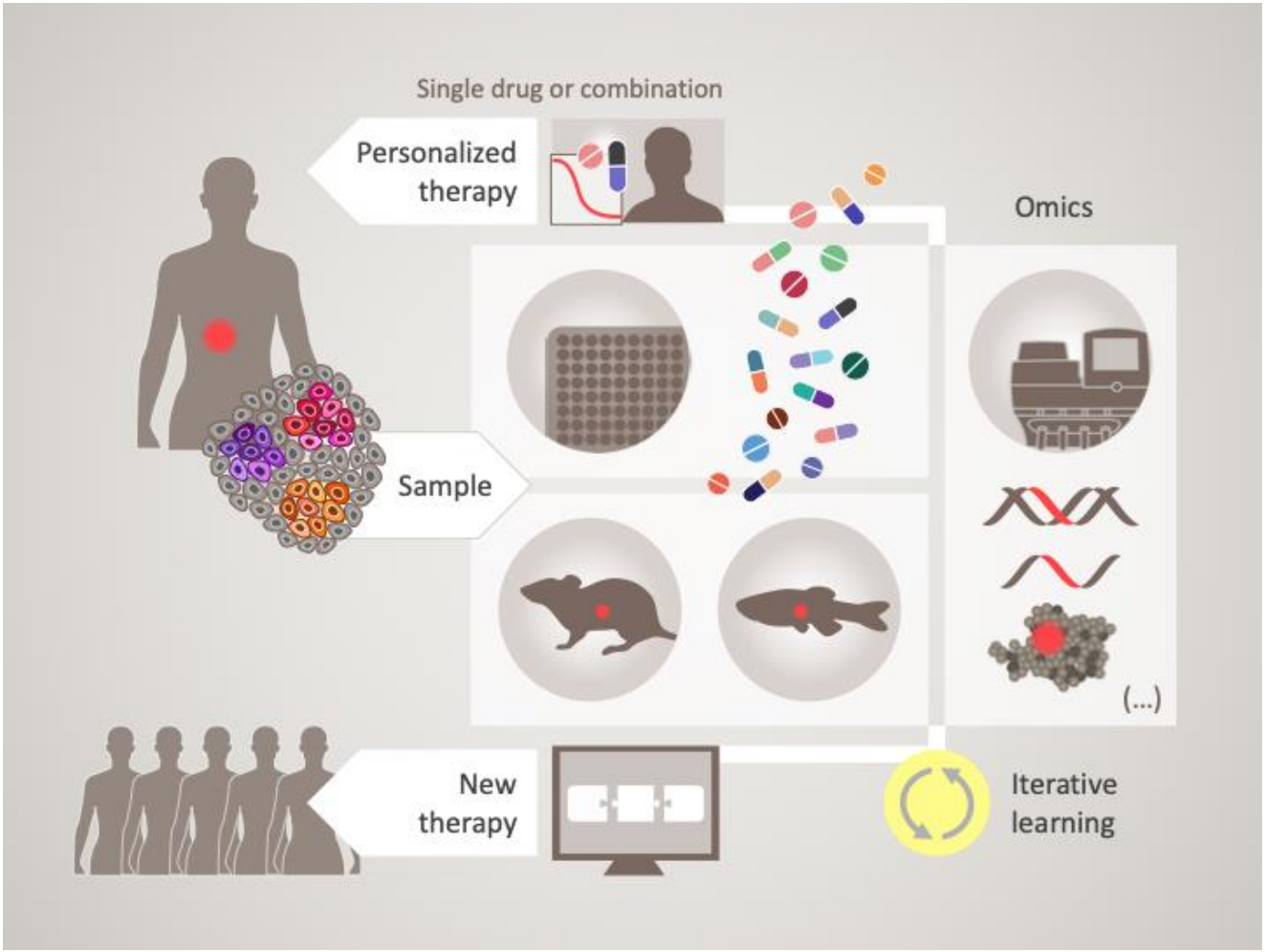


Illustration: Hierarchial model with 3 layers

In vitro drug screens and (multi-)omics tumor characterisations for personalised cancer therapy development

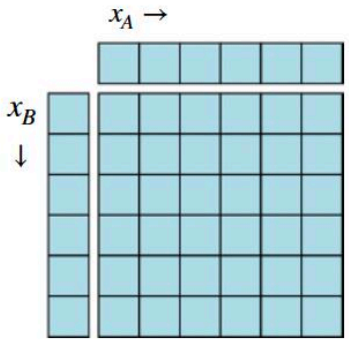


Slide from K Taskén

Example 1: Drug combination screens

Rønneberg et al., (2021, 2023, 2025)

Recap: Base model for individual experiments (bayesynergy)

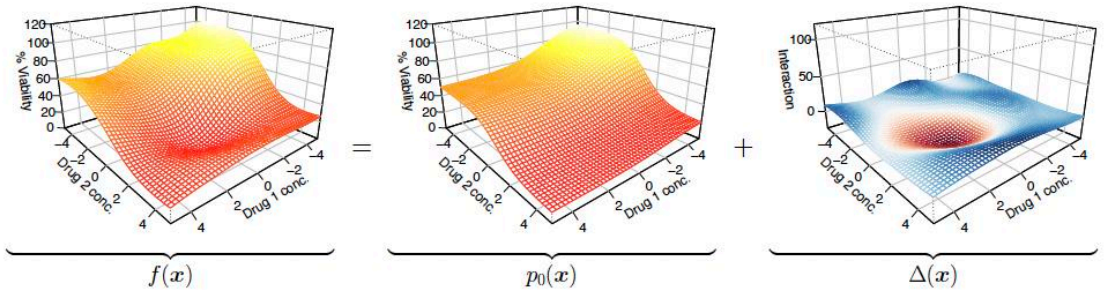


concentrations

$$\mathbf{x}_{ij} = (x_{Ai}, x_{Bj})$$

viability measurements

$$y_{ij} = f(\mathbf{x}_{ij}) + \varepsilon_{ij}$$



$$y_{ij} = f(\mathbf{x}_{ij}) + \varepsilon_{ij}$$

$$f(\mathbf{x}_{ij}) = p_0(\mathbf{x}_{ij}) + \Delta(\mathbf{x}_{ij})$$

$$p_0(\mathbf{x}_{ij}) = m_A(x_{Ai}) m_B(x_{Bj})$$

$$\Delta(\mathbf{x}_{ij}) = g(z(\mathbf{x}_{ij}))$$

$$z \sim \mathcal{GP}(0, \kappa(\mathbf{x}, \mathbf{x}'))$$

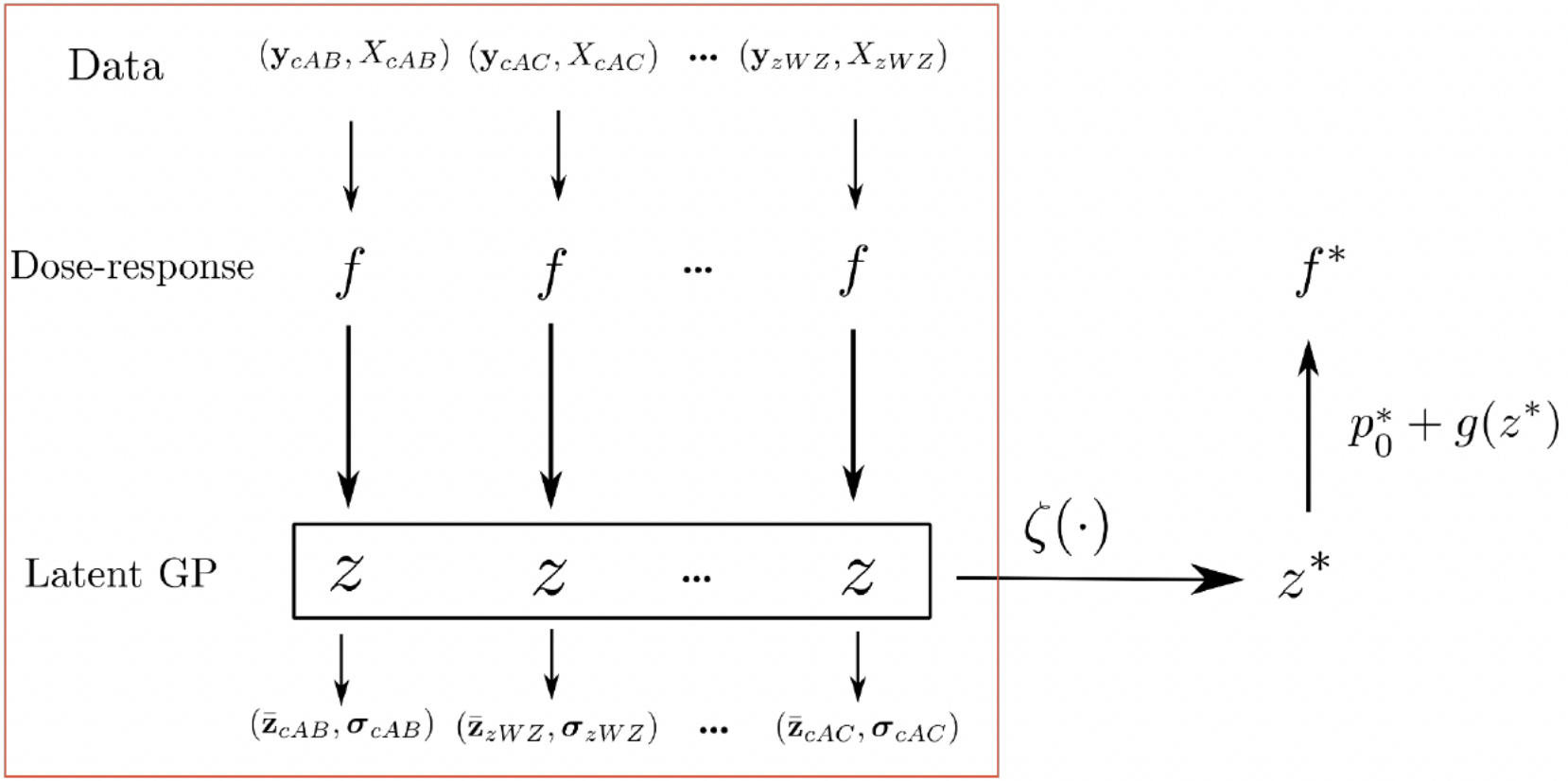
- $g(\cdot)$ is a squeezing function

$$f \in [0,1]$$

$$\mathbb{E}[f(\mathbf{x})] \approx p_0(\mathbf{x})$$

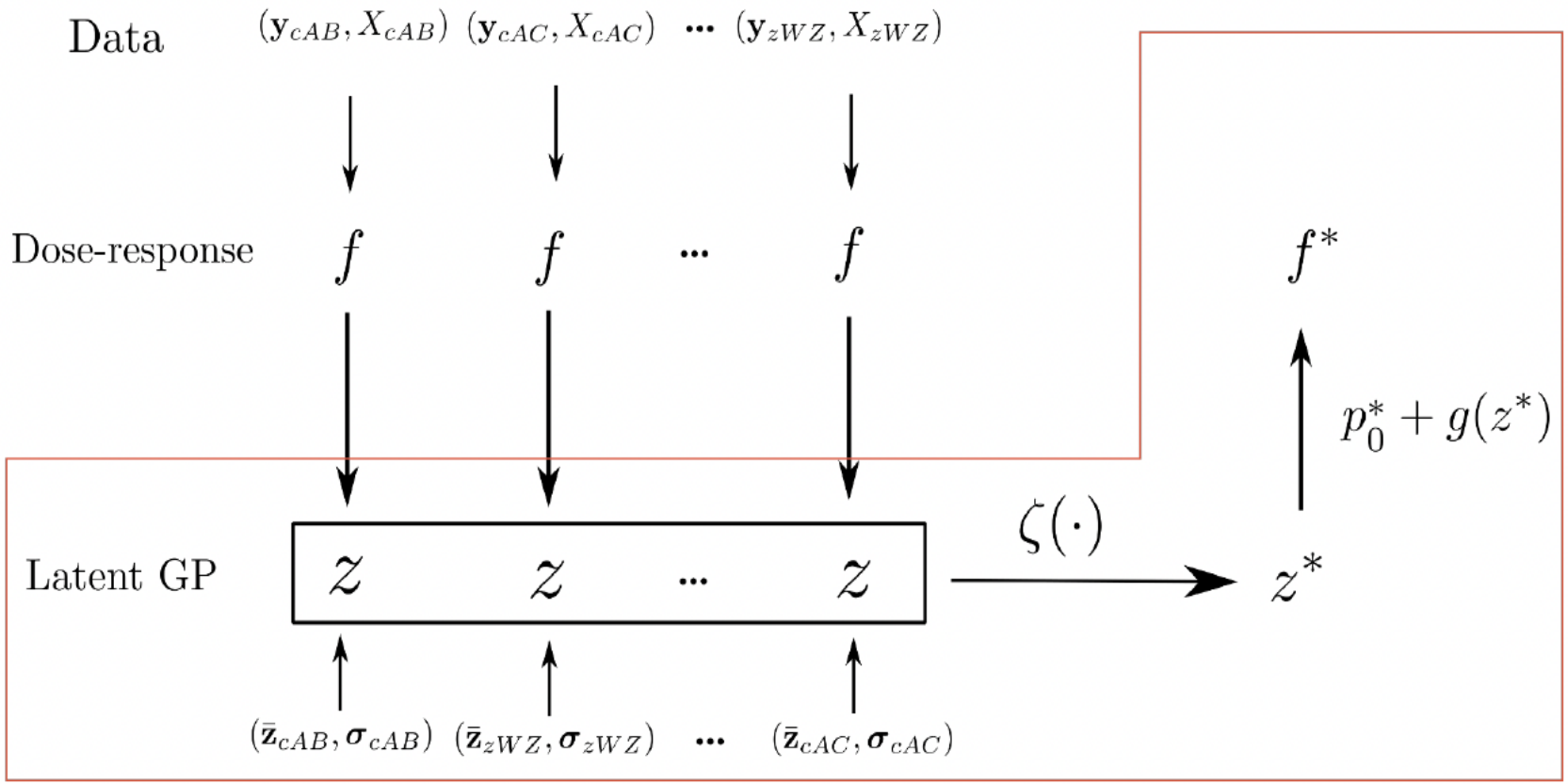
Example 1: Drug combination screens

Rønneberg et al., (2021, 2023, 2025)



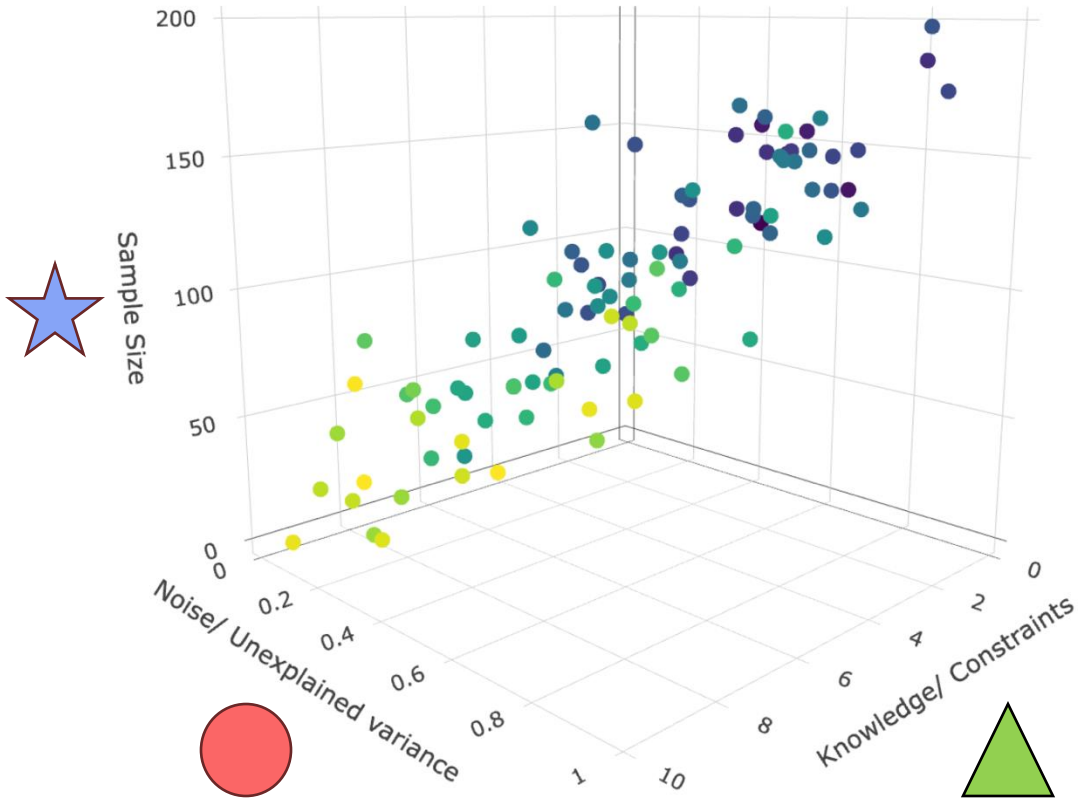
Example 1: Drug combination screens

Rønneberg et al., (2021, 2023, 2025)



- Individual experiments (bayesynergy): fully acknowledge all uncertainty
- Bayesian hierarchical model: Multi-output Gaussian processes to model individual latent GPs together and borrow information across experiments

Biomedical data are heterogeneous (and sample sizes too small)



Biomedical data are heterogeneous (and sample sizes too small)
Bayesian hierarchical models to the rescue

1. **Explain biological variance/ heterogeneity** through diverse data sources
   - Multi-omics/ multi-modal, and other data (subgroups, clinical)
   - Can integrate different data sources by tailoring priors to the different data types

2. **Acknowledge technical variance (noise)**
   - Full uncertainty propagation in modelling

3. **Increase sample size**
   - Samples do not need to be i.i.d. (but exchangeable)

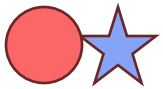
4. **Borrow information**
   - Joint model for related responses
   - Borrow across related features, e.g. with shrinkage priors

5. **Restrict the model space** through knowledge-based constraints
   - By including prior knowledge in the priors
   - Example: graph priors for drug target pathways

Examples 2: Bayesian variable selection with structured selection priors

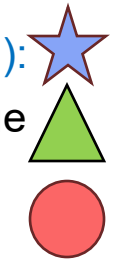
a. Multi-response Bayesian Variable Selection (BVS) (BayesSUR) (Zhao et al, JSS 2021):

- General setup for multi-response high-dimensional Bayesian linear model with variable and covariance selection (BayesSUR)
- Unified software (https://CRAN.R-project.org/package=BayesSUR)
  - Efficient implementations for several models in C++
  - R package with unified interface and graphics



b. BayesSUR with Markov random field (MRF) selection prior (Zhao et al, JRSSC 2024):

- Systematic investigation of the MRF selection prior for incorporating prior knowledge about molecular pathways and drug targets
- Allow for mandatory covariates (e.g. cancer subtype)



c. Pliable BVS: BayesSUR with interactions (Asenso et al, to be submitted):

- Allow interactions between omics effects and cancer types (adapt pliable lasso idea)

d. Cox BVS with MRF selection prior (Hermansen et al, arXiv:2503.13078):

- Cox-like Bayesian survival model with variable selection, including MRF selection prior
- R package: https://CRAN.R-project.org/package=BayesSurvive

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 ($p_3 = 68$)

a. **BayesSUR** (Zhao et al, JSS 2021)

Bayesian seemingly unrelated regression for variable and covariance selection (Bottolo et al. 2021; Zhao et al. 2021)

- **Matrix formulation of the model:**

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

  - $\mathbf{Y}$ $n \times m$ matrix of outcomes with $m \times m$ covariance matrix C,
  - $\mathbf{X}$ $n \times p$ matrix of predictors for all outcomes,
  - $\mathbf{B}$ $p \times m$ matrix of regression coefficients.
- In addition: Variable selection indicator matrix $\mathbf{\Gamma}$

	$\gamma_{jk} \sim$ 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

a. BayesSUR

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

- Independent Bernoulli prior:

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

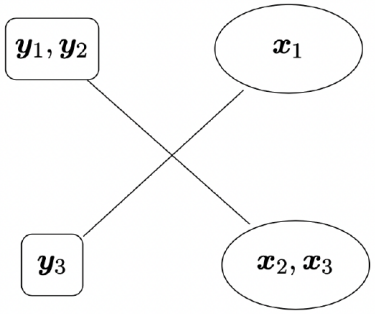
- Hotspot prior: (Bottolo et al. 2021)

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

- Markov Random Field (MRF) prior: (e.g. Chekouo et al. 2015.)

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

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

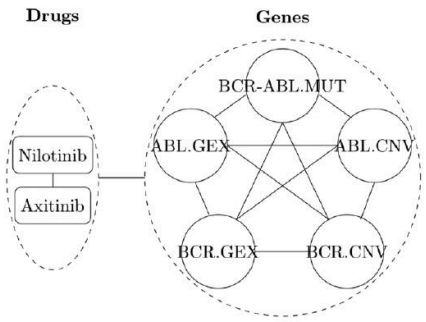


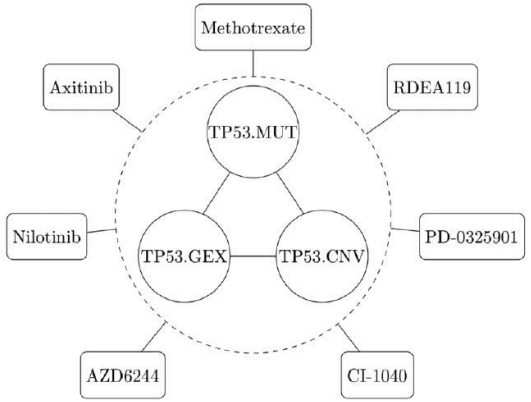
$$G = \begin{array}{c} \\ \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \\ \gamma_{12} \\ \gamma_{22} \\ \gamma_{32} \\ \gamma_{13} \\ \gamma_{23} \\ \gamma_{33} \end{array} \begin{array}{c} \begin{array}{ccccccccc} \gamma_{11} & \gamma_{21} & \gamma_{31} & \gamma_{12} & \gamma_{22} & \gamma_{32} & \gamma_{13}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

b. **BayesSUR with MRF prior** (Zhao et al, JRSSC 2023)

- Genomics of Drug Sensitivity in Cancer (Garnett et al., 2012)
- Illustration with 7 drugs

- **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 (targets of Group1)
  - genes in the Bcr-Abl fusion gene (targets 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)





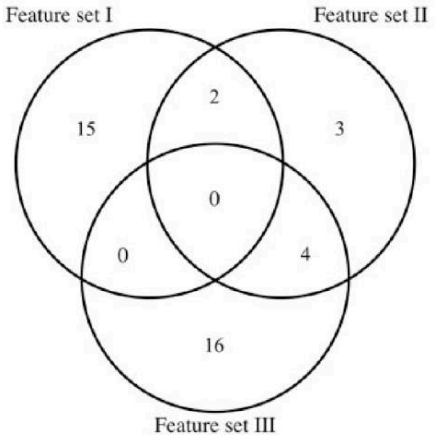
$$\underbrace{G_y}_{2\text{ drugs}} \otimes \underbrace{G_x}_{5\text{ features}} - \mathbb{I}_{10} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} - \mathbb{I}_{10}$$

$$\underbrace{G_y}_{7\text{ drugs}} \otimes \underbrace{G_x}_{3\text{ features}} - \mathbb{I}_{21} = \mathbb{I}_7 \otimes \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} - \mathbb{I}_{21}$$

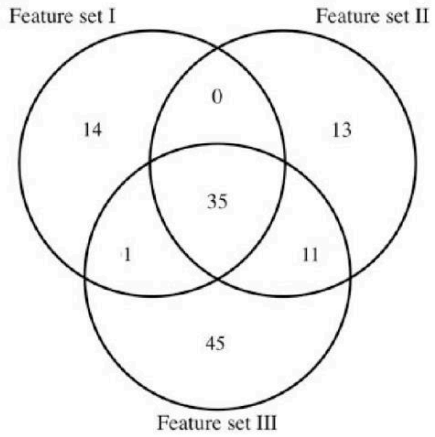
b. BayesSUR with MRF prior: what does this give us?

More stable feature selection w/ MRF prior

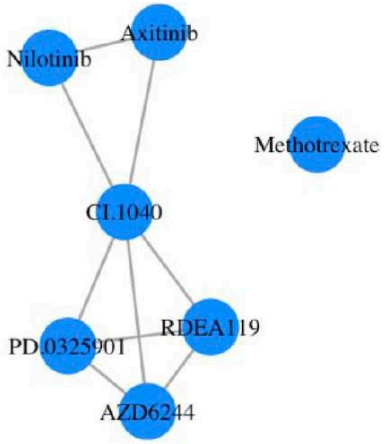
- Independent Bernoulli prior:



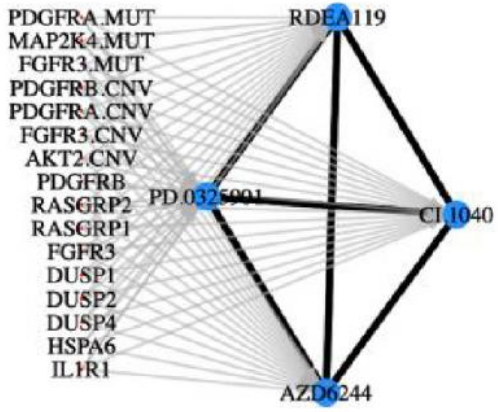
- Markov Random Field prior:



Graph of the joint (residual) covariances between drugs …



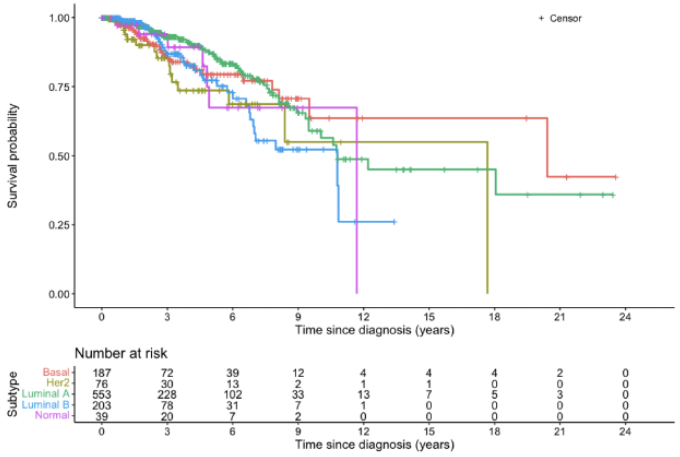
… and identify which features are associated with each sub-graph



d. Cox BVS with MRF selection prior (Hermansen et al, arXiv:2503.13078)

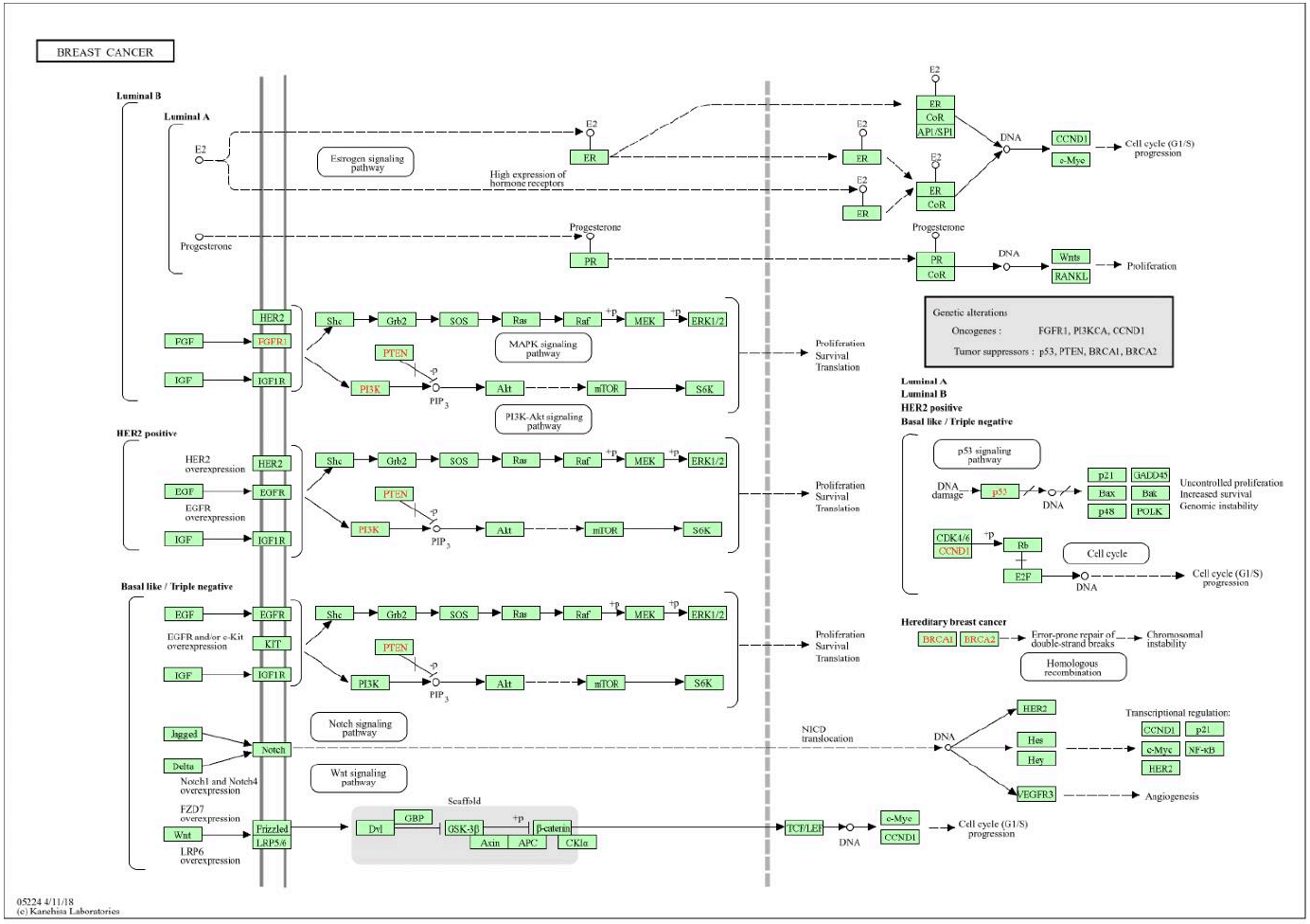
Application: Basal breast cancers (TCGA)

- Cohort of 187 women with basal breast cancer from The Cancer Genome Atlas (TCGA)
- Censoring rate: 87%, all with survival of at least 30 days
- Gene expression (p=414), gene mutation (p=75) and clinical (age, treatment)



d. Cox BVS with MRF selection prior

All KEGG breast cancer pathways incl. basal breast cancer



d. Cox BVS with MRF selection prior

How to translate a pathway for proteins to an MRF prior for gene-level features

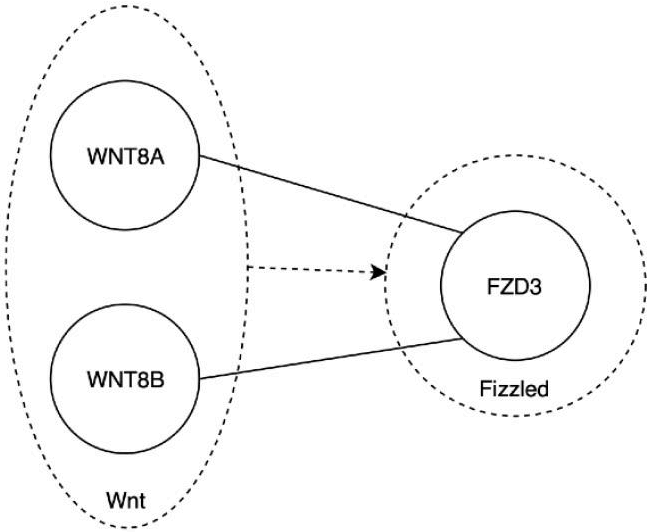
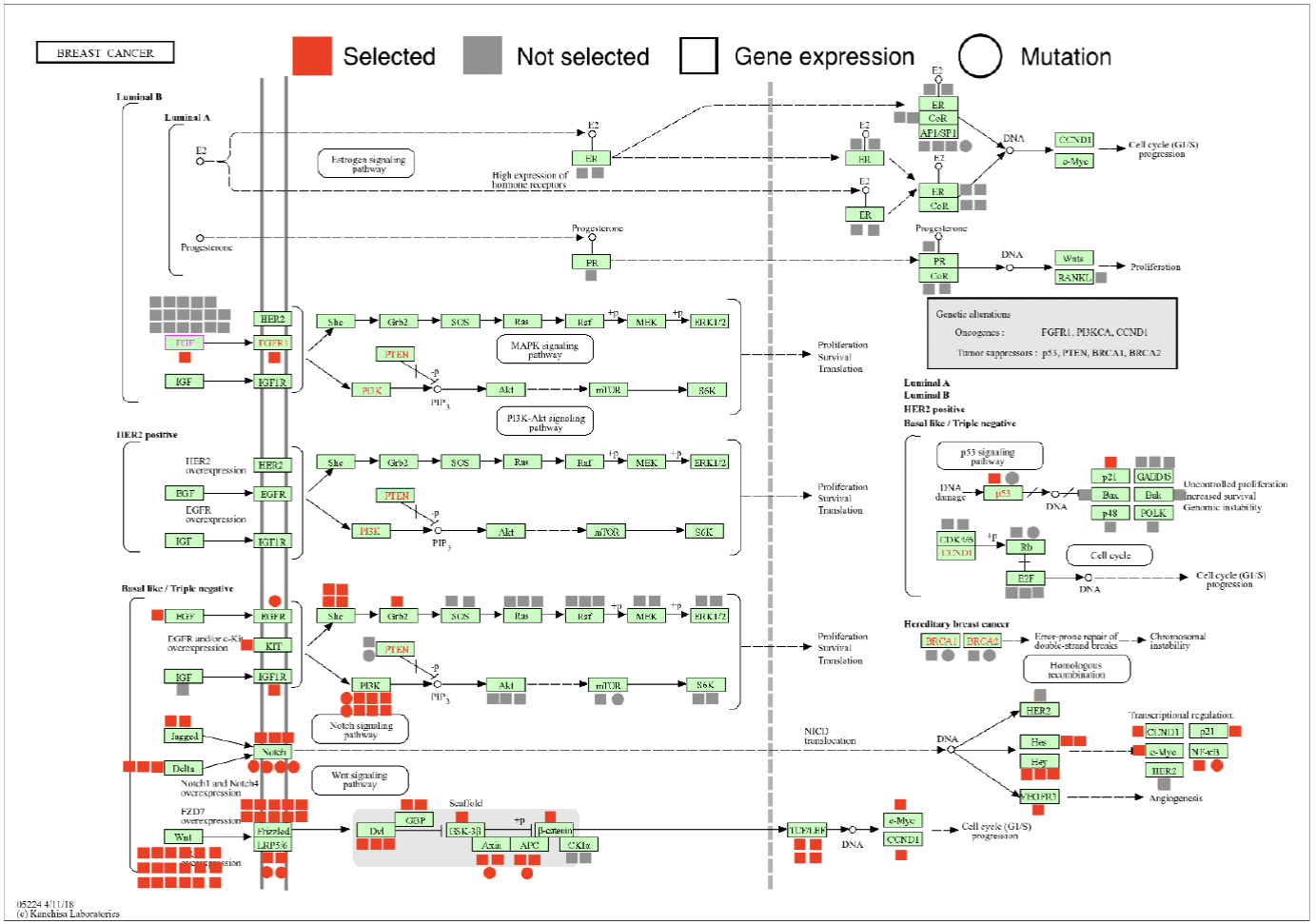


Figure S13: Illustration of how the graph is constructed from having proteins as nodes (the dotted lines) in the KEGG LRP6-overexpression to Wnt signaling pathway (Wnt $\rightarrow$ Fizzled), to having gene covariates as nodes (solid lines) in the graph used in the MRF prior. Note that while only two genes are chosen as representatives here (WNT8A and WNT8B), our data set actually includes a total of 18 gene expression features (and no mutations), see Figure 7. The edges are also changed from being directed in the KEGG pathways to being undirected in the graph for the MRF prior.

d. Cox BVS with MRF selection prior

Selecting mostly features from basal breast cancer pathway



Acknowledgements

Oslo Centre for Biostatistics and Epidemiology,
University of Oslo & Oslo University Hospital
George Zhi Zhao
Andrea Cremaschi
Leiv Rønneberg
Maren-Helene Langeland Degnes
Fekadu Bayisa
Theophilus Quachie Asenso
Fatima Heinicke
Matteo d'Alessandro
Tobias Østmo Hermansen

London School of Hygiene & Tropical Medicine, UK
Alexandra Lewin, Marco Banterle

University of Cambridge, UK
Leiv Rønneberg, Paul Kirk,
Leonardo Bottolo, Sylvia Richardson

University of Dortmund, Germany
Katrin Madjar, Katja Ickstadt, Jörg Rahnenführer

Oslo University Hospital
Mary-Ann Jallad, Pilar Ayuda-Durán, Robert Hanes,
Jorrit Enserink, Kjetil Taskén, Sigve Nakken, Eivind Hovig,
Vessela Kristensen






BigInsight

RESCUER

UNIVERSITY OF OSLO

The Research Council of Norway