Multivariate Bayesian variable selection in high-dimensional settings

ZHAO ZHI

Radiumhospitalet, OUS

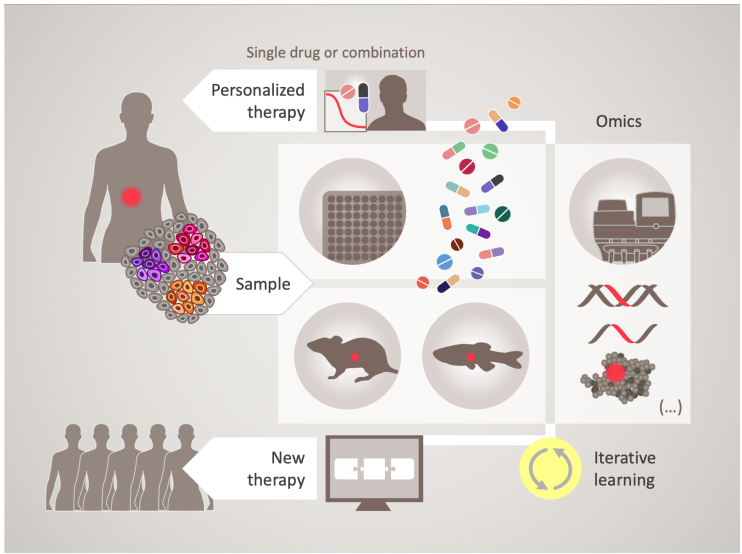
Oslo Centre for Biostatistics and Epidemiology, UiO

Acknowledgements

We present work in collaboration of a number of people:

- Marco Banterle (LSHTM)
- Leonardo Bottolo (Cambridge)
- Alex Lewin (LSHTM)
- Sylvia Richardson (Cambridge)
- Manuela Zucknick (UiO)

Background of pharmacogenomic study

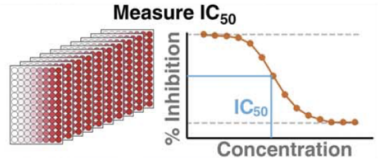


slide by Kjetil Taskén

Data integration

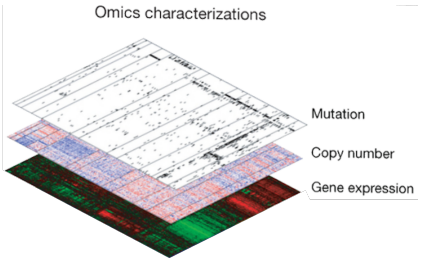
- **Drug dose response**

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



- **Integrative omics**

$$n \text{ cell lines} \begin{bmatrix} \overbrace{X_1}^{\text{gene expression}} & \overbrace{X_2}^{\text{copy number}} & \overbrace{X_3}^{\text{mutation}} \end{bmatrix} = X$$



Yang, et al. 2017; TCGA, 2013

Modelling Data

- The problem can be formulated as a **multivariate linear regression model**

$$\begin{aligned} \mathsf{Y}_{n\times m} &= \mathsf{X}_{n\times p}\boldsymbol{B}_{p\times m} + \mathsf{U}_{n\times m} \\ \text{vec}\{\mathsf{U}\} &\sim \mathcal{N}(0,\ C \otimes \mathbb{I}_n) \end{aligned}$$

where we assume sparse association using independent spike-and-slab prior on the coefficient

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

where γ_{kj} is a latent variable inclusion indicator ($j = 1, \cdots, m; k = 1, \cdots, p$). Denote matrix $\Gamma = \{\gamma_{kj}\}_{kj}$ and vector $\boldsymbol{\gamma} = \text{vec}\{\Gamma\}$

BayesSUR for high-dimensional multivariate Bayesian variable and covariance selection in linear regression

- A unified, efficient and user-friendly implementation of a class of models in the R package BayesSUR using Evolutionary Stochastic Search.
- `https://CRAN.R-project.org/package=BayesSUR`

	$\gamma_{jk} \sim$ Bernoulli	$\gamma_{jk} \sim$ Hotspot	$\gamma \sim$ 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

Tab: Nine models across three priors of C by three priors of indicator variable Γ

Prior setup

- **Prior knowledge for drug correlations**
  - $C \sim$ indep: assume uncorrelated drugs
  - $C \sim \mathcal{IW}$: assume correlated drugs
  - $C \sim \mathcal{HIW}_{\mathcal{G}}$: assume correlated drugs and estimate their correlations based on a sparse graph

- **Prior knowledge for sparse association**
  - $\gamma_{jk} \sim$ Bernoulli prior: not assume uncorrelated genes or drugs
  - $\gamma_{jk} \sim$ Hotspot prior: assume relationships between one gene and multiple drugs, and relationships between multiple genes and one drug
  - $\boldsymbol{\gamma} \sim$ MRF prior: give known (partial) relationships between genes and drugs

Illustration of our idea

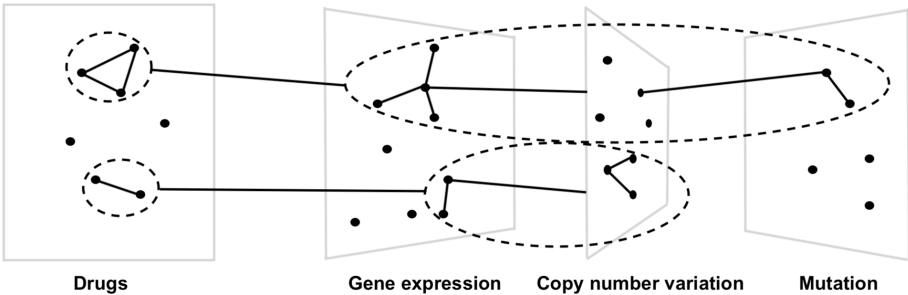


Figure: Illustration of targeted cancer drug groups and omics path

Prior setup

- **Joint graph structure**
  Set Markov random field (MRF) prior on the variable inclusion indicators $\gamma = \text{vec}\{\Gamma\}$,

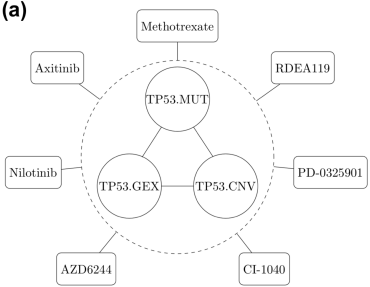
$$\gamma | d, e, G \propto \exp\{d \mathbb{1}^\top \gamma + e \gamma^\top G \gamma\}$$

  where G is an $mp \times mp$ (possibly weighted) adjacency matrix representing a graph to include prior structure knowledge
  - d: control the sparsity of the model
  - e: encourage the selection of related predictors

Prior setup

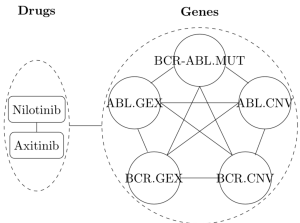
- **Example: Constructing G in the MRF prior**

(a)



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

(b)



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

Full model

$$
\begin{aligned}
\mathsf{Y} &= \mathsf{Z}\boldsymbol{B}_0 + \mathsf{X}\boldsymbol{B} + \mathsf{U}, \\
\beta_{0,tj} | w_0 &\sim \mathcal{N}(0,\ w_0), && \Leftarrow \textit{random effects} \\
\beta_{kj} | \gamma_{kj}, w &\sim \gamma_{kj}\mathcal{N}(0,\ w) + (1-\gamma_{kj})\delta_0(\beta_{kj}), \\
w_0 &\sim \mathcal{IG}(a_{w_0}, b_{w_0}), \\
w &\sim \mathcal{IG}(a_w, b_w), \\
\boldsymbol{\gamma} | d, e, G &\propto \exp\{d\mathbb{1}^\top\boldsymbol{\gamma} + e\boldsymbol{\gamma}^\top G\boldsymbol{\gamma}\}, && \Leftarrow \textit{MRF prior} \\
\mathrm{vec}\{\mathsf{U}\} &\sim \mathcal{N}(0,\ C \otimes \mathbb{I}_n), \\
C &\sim \mathcal{HIW}_{\mathcal{G}}(\nu, \tau\mathbb{I}_m), \\
\tau &\sim \mathcal{G}amma(a_\tau, b_\tau)
\end{aligned}
$$

Sampling steps

- sampling latent indicator variables $\mathbf{\Gamma}$ using Thompson sampling;
- sampling coefficients $\boldsymbol{B}$ (and $\boldsymbol{B}_0$) from the full conditional distributions;
- sampling hyper-parameter τ using a random walk Metropolis sampler;
- sampling hyper-parameter w (and w_0) using Gibbs sampling;
- sampling the graph $\mathcal{G}$ from the junction tree sampler;
- sampling $\boldsymbol{\sigma}^2$ and $\boldsymbol{\rho}$ from the full conditional distributions (reparametrized from C).

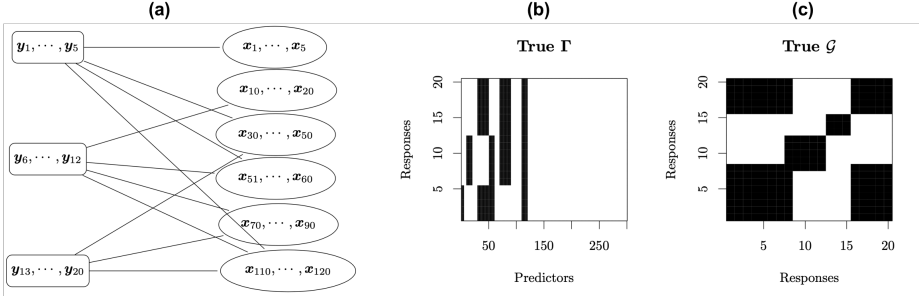
NOTE: At each iteration, the ESS algorithm implements a local move to add/delete and swap the latent indicator variables within each chain, and then a global move to exchange and crossover the latent indicator variables between any two parallel tempered chains.

Simulation setup

- **Sample**
  - $\#\{responses\} = m = 20$
  - $\#\{subjects\} = n = 250$
  - $\#\{predictors\} = p = 300$
- **Network**

- **Generate data**

$$\mathsf{Y}_{n\times m} = \mathsf{X}_{n\times p}\boldsymbol{B}_{\Gamma} + \mathsf{U}$$



Model performance evaluation

- **Variable selection**
  - Accuracy, sensitivity and specificity of nonzero coefficients based on posterior inclusion probability $\gamma_{ij} > 0.5$
- **Response prediction**
  - Median probability model (Barbieri & Berger 2004; Barbieri, Berger, George, Ročková 2021)

$$\mathbb{E}[\beta_{kj}|\gamma_{kj} = 1, data], \quad \text{if } \mathbb{P}\{\gamma_{kj} = 1|data\} > 0.5$$

$$\hat{\beta}_{kj,MPM} = \begin{cases} \frac{\sum_{t=1}^{N} \beta_{kj}^{(t)}}{\sum_{t=1}^{N} \gamma_{kj}^{(t)}}, & \text{if } \frac{\sum_{t=1}^{N} \gamma_{kj}^{(t)}}{N} > 0.5, \\ 0, & \text{otherwise}, \end{cases}$$

  - Prediction errors

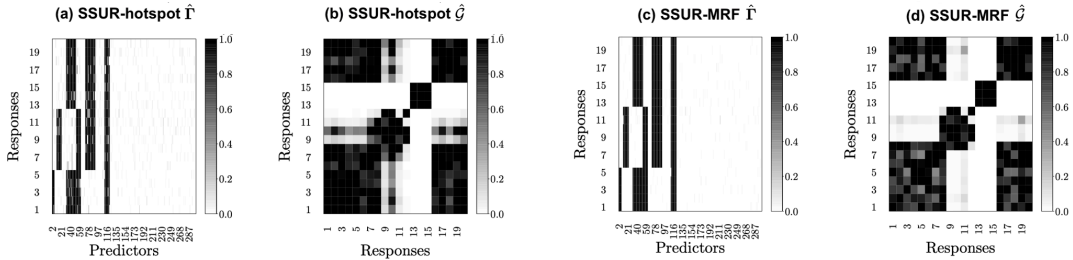
$$\text{RMSE} = \frac{1}{\sqrt{mn}} \|\mathrm{Y} - \mathrm{X}\hat{\boldsymbol{B}}_{MPM}\|_2,$$

$$\text{RMSPE} = \frac{1}{\sqrt{mn'}} \|\mathrm{Y}^* - \mathrm{X}^*\hat{\boldsymbol{B}}_{MPM}\|_2,$$

where $\hat{\boldsymbol{B}}_{MPM} = \{\hat{\beta}_{kj,MPM}\}$, $\mathrm{Y}_{n\times m}$ and $\mathrm{X}_{n\times p}$ were used to estimate $\hat{\boldsymbol{B}}_{MPM}$, and $\mathrm{Y}^*_{n'\times m}$ and $\mathrm{X}^*_{n'\times p}$ are new data.

Simulation results: comparison †

- **Variable selection**



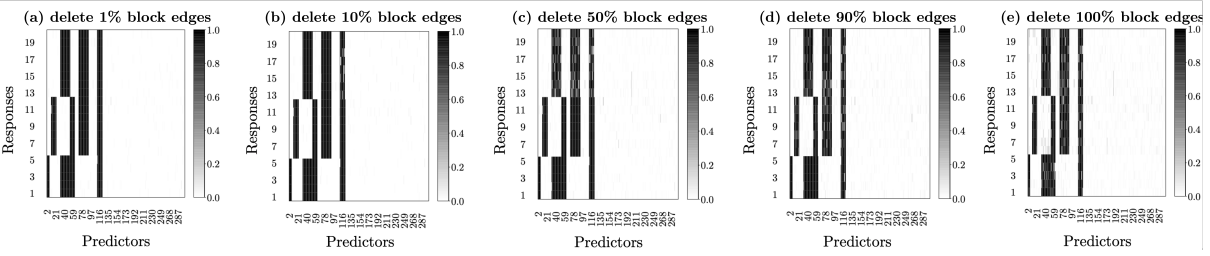
- **Variable selection and prediction performance**

	accuracy	sensitivity	specificity	RMSE	RMSPE
SSUR-hotspot	0.988	0.936	0.999	0.800	0.693
SSUR-MRF	0.989	0.998	0.986	0.643	0.412

†Comparison with an alternative method: Bayesian sparse SUR with (multiplicative) hotspot prior for Γ (Bottolo, Banterle, Richardson et al., 2021, JRSSC)

Simulation results: sensitivity analysis of MRF prior

- More known prior information, more accurate structure recovery



GDSC data

- **Genomics of Drug Sensitivity in Cancer** (Garnett et al. 2012)
  - **Pharmacological profiling**

    $m = 7$ drugs, $n = 499$ cell lines, $T = 13$ tumor types
    - Group1 drugs: MAPK inhibitors (RDEA119, PD-0325901, CI-1040, AZD6244)
    - Group2 drugs: Bcr-Abl tyrosine-kinase inhibitors (Nilotinib, Axitinib)
    - Chemotherapy agent: Methotrexate
  - **Genomic information**

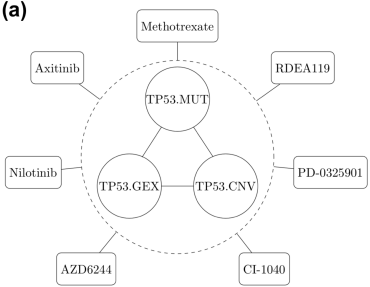
	Feature set I $\subset$	Feature set II $\subset$	Feature set III
# {gene expression features}	783	1175	2602
# {copy number features}	426	426	426
# {mutation features}	68	68	68

  - **Edge potentials for MRF prior**
    - edges between genes in MAPK pathway corresponding to Group1 drugs
    - edges between genes in the Bcr-Abl fusion gene corresponding to Group2 drugs
    - edges between the representations of each gene in different data sources

GDSC data

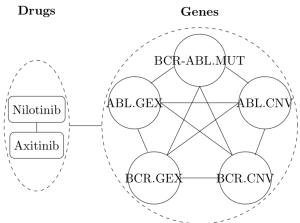
- **Constructing G in the MRF prior**

(a)



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

(b)



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

GDSC data

- **Variable selection for the 4 MAPK inhibitors**
  - SSUR-Ber has unstable variable selection when including more (or less) genomic information
  - SSUR-MRF has quite stable variable selection, and always identifies some common key target genes

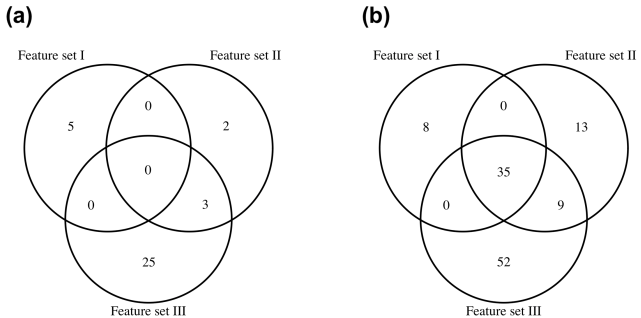


Figure: A Venn diagram for the numbers of identified features for the MAPK inhibitors by SSUR-Ber (panel (a)) and SSUR-MRF (panel (b)) models and overlaps between the models fitted with feature sets I, II, and III.

GDSC data

- **Variable selection for the 4 MAPK inhibitors**
  - 35 common features
    = drug target genes (MAPK pathway) + cancer genes (Cancer Gene Census)

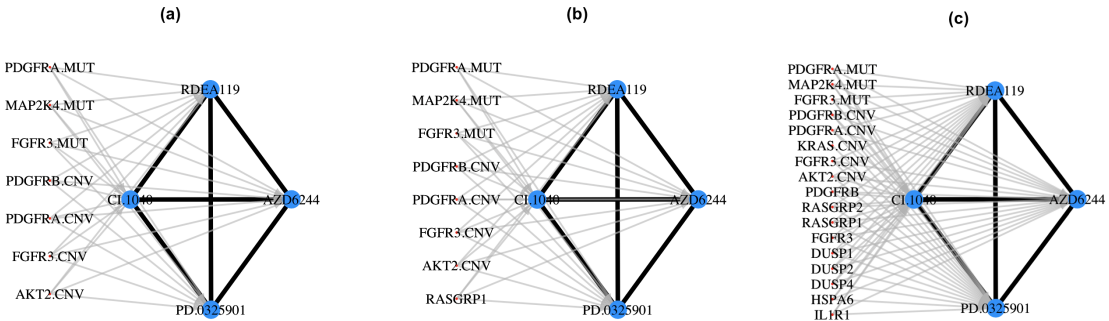


Figure: Estimated network between the MAPK inhibitors and identified target genes based on $\hat{\mathcal{G}}$ and $\hat{\mathbf{\Gamma}}$ thresholded at 0.5 by SSUR-MRF corresponding to feature set I, II and III respectively.

Summary

- **Pros**
  - Integrate multi-omics data in one model for the variable selection and prediction of multivariate response variables
  - Take into account relationships between multiple response variables and high-dimensional predictors
  - Improve performance of variable selection by using known prior knowledge
- **Cons**
  - Our Bayesian method cannot provide cancer tissue-specific gene effects estimation/variable selection
  - The proposed Bayesian model might need long computing time if the model is not assumed to be very sparse (i.e. number of true associated features $\ll mp$)

References

- Alexopoulos A, Bottolo L (2021). Bayesian variable selection for Gaussian copula regression models. Journal of Computational and Graphical Statistics, 30(3): 578–593
- Bottolo L, Banterle M, Richardson S, Ala-Korpela M, Järvelin MR, Lewin A (2021). A computationally efficient Bayesian seemingly unrelated regressions model for high-dimensional quantitative trait loci discovery. Journal of the Royal Statistical Society: Series C (Applied Statistics), 70(4):886-908
- Lee K, Tadesse MG, Baccarelli AA, Schwartz J, Coull BA (2017). Multivariate Bayesian variable selection exploiting dependence structure among outcomes: Application to air pollution effects on DNA methylation. Biometrics, 73(1): 232–241
- Zhao Z, Banterle M, Bottolo L, Richardson S, Lewin A, Zucknick M (2021). BayesSUR: An R package for high-dimensional multivariate Bayesian variable and covariance selection in linear regression. Journal of Statistical Software, 100(11):1-32
- Zhao Z, Banterle M, Lewin A, Zucknick M (2021). *Structured Bayesian variable selection for multiple correlated response variables and high-dimensional predictors*. arXiv: 2101.05899