Journal of the Royal Statistical Society Series C: Applied Statistics, 2023, **00**, 1–24
https://doi.org/10.1093/jrsssc/qlad102

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society Series C Applied Statistics

Original Article

Multivariate Bayesian structured variable selection for pharmacogenomic studies

Zhi Zhao[1,2], Marco Banterle[3], Alex Lewin[3,†] and Manuela Zucknick[1,†]

[1]Department of Biostatistics, Oslo Centre for Biostatistics and Epidemiology (OCBE), Institute of Basic Medical Sciences, University of Oslo, Oslo 0317, Norway
[2]Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo 0310, Norway
[3]Department of Medical Statistics, Faculty of Epidemiology and Population Healthy, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK

Address for correspondence: Zhi Zhao, Department of Biostatistics, University of Oslo, P.O.Box 1122 Blindern, Oslo 0317, Norway. Email: zhi.zhao@medisin.uio.no

Abstract

Cancer drug sensitivity screens combined with multi-omics characterisation of the cancer cells have become an important tool to determine the optimal treatment for each patient. We propose a multivariate Bayesian structured variable selection model for sparse identification of multi-omics features associated with multiple correlated drug responses. Our model uses known structure between drugs and their targeted genes via a Markov random field (MRF) prior in sparse seemingly unrelated regression. The use of MRF prior can improve the model performance compared to other common priors. The proposed model is applied to the Genomics of Drug Sensitivity in Cancer data.

Keywords: Markov random field prior, precision cancer medicine, random effects, seemingly unrelated regression, spike-and-slab prior

1 Introduction

A large proportion of advanced solid tumours harbour potentially treatable genomic variants (Fontes Jardim et al., 2015; Le Tourneau et al., 2015; Von Hoff et al., 2010), but very few cancer patients actually benefit from genome-informed treatments (Marquart et al., 2018). Thus, there is great potential to improve the use and benefit of therapy for individual patients by better patient stratification and by patient-tailored design of therapies. Precision cancer medicine aims at guiding cancer patient treatment based on detailed molecular characterisation of each patient's disease. One strategy that is rapidly gaining traction is *ex vivo* cancer drug sensitivity screening, which predicts responses to a range of potential therapies in cancer cell lines and patient-derived cells and identifies molecular features that are associated with drug responses. Studies where both, drug sensitivity and molecular (multi-omics), data are available are commonly referred to as pharmacogenomic studies. In this article, we employ a multivariate (multi-response) regression setup with high-dimensional input matrix to analyse pharmacogenomic data, where sensitivities to several drugs are the response variables and molecular (multi-)omics variables are the input features. We analyse data from the Genomics of Drug Sensitivity in Cancer (GDSC) database (Garnett et al., 2012; Yang et al., 2013), which contains the results from drug sensitivity screens to hundreds of cancer drugs for hundreds of cell lines representing diverse cancers in a pan-cancer setup

† A.L. and M.Z. are joint last authors

Received: June 30, 2022. **Revised:** August 15, 2023. **Accepted:** October 28, 2023

and multi-omics characterisation of these cell lines. Our approach can identify important genes affiliated with target pathways of the drugs (i.e. target genes) as well as genes whose dysfunction is known to drive cancer (cancer genes), which may guide personalised cancer therapies and aid discovery of potential new application areas of anti-cancer drugs in additional cancer types based on the identification of both tissue-specific and pan-cancer processes.

Large-scale *in vitro* cancer drug screens produce a large amount of drug sensitivity data which are expected to be correlated for drugs that have similar mechanisms of action or common target genes or pathways. Meanwhile, multi-omics information, including for example transcriptomics (gene expression), genomics (point mutations or copy number variations) or epigenomics (e.g. CpG methylation) data, is measured for the cancer cells, which is expected to guide personalised cancer therapies through prediction of drug sensitivity (Barretina et al., 2012; Garnett et al., 2012). The omics input data are often high-dimensional and are typically sparsely associated with the response variables in a structured manner, where variables corresponding to genes in the same molecular pathway can have similar association patterns with the drugs, for example because a drug targets a molecular signalling pathway which effects the expression of several genes in the pathway. In addition, since multiple omics characterisations reflect different aspects of information of the same system or co-functionality of multiple gene features (Kim et al., 2019), an analysis of joint associations between the correlated multiple phenotypes (e.g. multiple drugs) and high-dimensional molecular features (i.e. multi-omics data) is desired, but poses both theoretical and computational challenges. Finally, it is expected that not all of the heterogeneity between the cancer samples can be explained by the available molecular data. In particular, a pan-cancer pharmacogenomic screen will include samples from multiple cancer types, which adds heterogeneity in the drug sensitivity due to the different tissue and cell types, even if the involved molecular pathways and mechanisms are the same. This leads us to include random effects in the model to reflect heterogeneity between cancer types.

There are a number of statistical and machine learning models developed for predicting drug sensitivity by using omics data (see e.g. Adam et al., 2020; Ballester et al., 2022; Feng et al., 2021; Sharifi-Noghabi et al., 2021). These models are often designed for making accurate predictions, either within a single cancer type (Costello et al., 2014) or using a cancer-agnostic approach (Barretina et al., 2012). Furthermore, while emphasising accurate predictions, many of the models lack effective variable selection options, making such black-box models less practical for biological studies or clinical applications. Huang et al. (2020) developed tissue-guided lasso for integrating cancer tissue of origin with genomic profiles, which just repeats the analysis in each cancer type, rather than jointly modelling the pan-cancer data. Zhao and Zucknick (2020) proposed tree-guided group lasso with integrative penalty factors to jointly model drug-drug similarities and heterogeneity of multi-omics from pan-cancer data, but do not take into account correlation structure across multiple omics data sources.

Bayesian modelling provides flexibility to specify the relationships in such complex data. There have been several Bayesian methods developed to deal with structure in complex data. For example, Bai et al. (2022) and Yang and Narisetty (2020) studied Bayesian group selection of high-dimensional predictors, but for univariate response variables. Liquet et al. (2017) extended the univariate response model to a multivariate model but lack computational efficiency because they used a standard MCMC algorithm. Richardson et al. (2011) proposed hierarchical related regression (HRR) for multivariate response variables. HRR assumes a simple independence prior for the residual covariance matrix, and it applies an efficient Evolutionary Stochastic Search (ESS) algorithm based on Evolutionary Monte Carlo (Bottolo & Richardson, 2010). More complex priors, e.g. inverse Wishart or hyper-inverse Wishart prior, can be used for the residual covariance matrix to learn structures between multivariate response variables (Bhadra & Mallick, 2013; Bottolo et al., 2021; Carvalho et al., 2007; Petretto et al., 2010; Wang, 2010).

Besides imposing different structured priors on the residual covariance matrix, it is necessary to also impose structured variable selection priors for high-dimensional predictors. Although independent spike-and-slab priors for variable selection are often used in high-dimensional multivariate models (Bottolo et al., 2021; Chakraborty et al., 2021; Ha et al., 2021; Jia & Xu, 2007), a structured Markov random field (MRF) prior can also be used for the latent indicator variables to introduce prior dependence between predictors (Chekouo et al., 2015, 2017, 2016) and hyperpriors of the MRF prior can be used to infer the sparsity of the dependence structure. Lee et al. (2017) utilised the residual covariance matrix for the dependence structure in an MRF prior to encourage joint selection of the same predictor across several correlated response variables. In all these articles, an

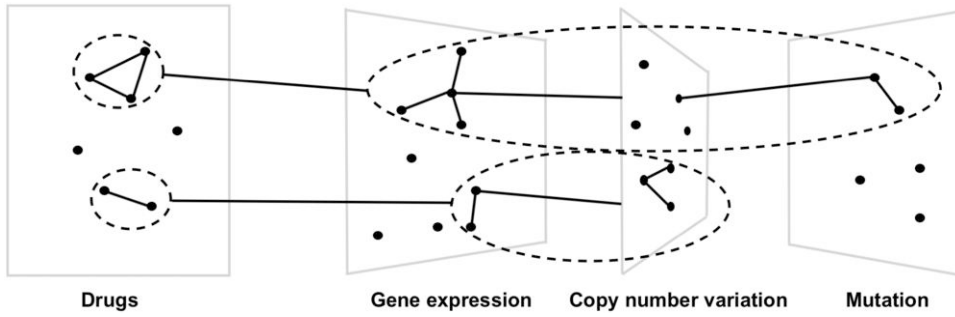

Figure 1. Illustration of the drug groups and omics path.

MRF prior is set for the latent variables of regression coefficients only corresponding to one response variable, which therefore does not allow to learn structures across multiple response variables.

In this article, we propose a multivariate Bayesian structured variable selection approach based on Richardson et al. (2011) and its extension by Bottolo et al. (2021), which can deal with multiple response variables (e.g. the cell lines' sensitivity to multiple cancer drugs) and high-dimensional genomic predictors, and possess computational efficiency through the ESS algorithm. Our proposed approach aims to include a known complex structure between multiple response variables and high-dimensional predictors via a flexible MRF prior for the latent indicator variables of the regression coefficient matrix. That is, we include known biological associations for the dependence structure in an MRF prior rather than doing MRF inference. Our use of the MRF prior has two main advantages:

- it takes into account prior knowledge on inter-relations between predictors including across groups of predictors and across response variables, to improve model performance (i.e. variable selection and prediction), and
- it performs posterior inference for the model in a more computationally efficient manner than the use of data-driven structured priors {e.g. multiplicative prior for the Bernoulli probability of the latent indicator variable (i.e. hotspot prior) by Richardson et al. (2011) and hyperprior for the MRF edge potentials by Chekouo et al. (2017)} would allow.

For example, Figure 1 illustrates two groups of drugs and their corresponding two groups of target genes or pathways across multiple omics characterisations. When using omics data to predict drug responses, the associations between the multiple drugs and omics features can include prior knowledge about the groups of drugs and their target genes or target pathways. An MRF prior is able to address the joint structure by adding the edges for omics features within a group of target genes or pathways that correspond to the group of their targeting drugs. In addition, if the drug responses are measured on cell lines from different cancer types or different tissues, we use random effects to capture the sample heterogeneity arising from these sample groups. An R package BayesSUR (Zhao et al., 2021) is available on the Comprehensive R Archive Network at https://CRAN.R-project.org/package=BayesSUR.

The rest of the article is organised as follows. In Section 2, we introduce the Bayesian sparse seemingly unrelated regression (SSUR) model, propose an MRF prior for the latent indicator variables of the coefficient matrix, and introduce random effects for sample groups. Section 3 compares the performances of Bayesian SSUR models with our MRF prior to the hotspot prior by Bottolo et al. (2011) with respect to (w.r.t.) structure recovery and prediction in simulated data. In Section 4, we analyse a pharmacogenomic dataset from the GDSC database. In Section 5, we conclude the article with a discussion.

2 Methodology

2.1 SSUR model

We study a multivariate regression model with a response matrix $\mathbf{Y}_{n\times m}$ from n samples and m response variables. All response variables are regressed on the same p predictors which are measured

on the n samples, so that the predictor matrix is $\mathbf{X}_{n\times p}$. Associations between the responses $\mathbf{Y}$ and predictors $\mathbf{X}$ are captured by a coefficient matrix $\boldsymbol{B}_{p\times m}$. We first assume correlated response variables, but independent samples. Section 2.3 will then extend the model to allow for correlated samples. The classic seemingly unrelated regression (SUR) model is defined as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{B} + \mathbf{U},\ \text{vec}\{\mathbf{U}\} \sim \mathcal{N}(0,\ \Psi \otimes \mathbb{I}_n), \tag{1}$$

where the residuals have correlated columns with covariance Ψ and independent rows, and vec$\{\cdot\}$ is to vectorise a matrix by column.

In the Bayesian framework, to efficiently sample from the posterior distribution of the regression coefficients from (1), Zellner and Ando (2010) reparametrised the SUR model and proposed a direct Monte Carlo procedure. Bottolo et al. (2021) used the same reparametrisation for the SUR model, but with an inverse Wishart prior $\Psi \sim \mathcal{IW}(v, \tau\mathbb{I}_m)$. Briefly, then model (1) can be rewritten as

$$\boldsymbol{y}_j = \mathbf{X}\boldsymbol{\beta}_j + \sum_{l<j} \boldsymbol{u}_l\rho_{jl} + \boldsymbol{\epsilon}_j,\ \boldsymbol{\epsilon}_j \sim \mathcal{N}(0,\ \sigma_j^2\mathbb{I}_n), \tag{2}$$

where $\boldsymbol{u}_l = \boldsymbol{y}_l - \mathbf{X}\boldsymbol{\beta}_l$. The reparametrised parameters (σ_j^2, ρ_{jl}) have priors

$$\sigma_j^2 \sim \mathcal{IG}\left(\frac{v-m+2j-1}{2},\ \frac{\tau}{2}\right),\ \rho_{jl}|\sigma_j^2 \sim \mathcal{N}\left(0,\ \frac{\sigma_j^2}{\tau}\right),\ j > l, \tag{3}$$

where v is fixed and $\tau \sim \mathcal{G}amma(a_\tau, b_\tau)$. Note that the joint distribution $f(\mathbf{Y}|\mathbf{X}, \boldsymbol{B}, \Psi)$ is the same regardless of the order used for the decomposition since we are simply factorising it by chain-conditioning (Bottolo et al., 2021).

The reparametrisation factorises the likelihood across multiple response variables possible, which especially benefits high-dimensional response variables. If only a few of the p predictor variables are assumed to be associated with any of the response variables, we use a latent indicator matrix $\boldsymbol{\Gamma} = \{\gamma_{kj}\}$ for variable selection. If $\gamma_{kj} = 1$, then $\beta_{kj} \neq 0$ and the kth predictor is regarded as an associated predictor to the jth response variable; otherwise $\gamma_{kj} = 0$ and $\beta_{kj} = 0$. Independent spike-and-slab priors (Brown et al., 1998; George & McCulloch, 1993) for the regression coefficients can be used to find a small subset of predictors that explains the variability of $\mathbf{Y}$, for example:

$$\beta_{kj}|\gamma_{kj}, w \sim \gamma_{kj}\mathcal{N}(0,\ w) + (1-\gamma_{kj})\delta_0(\beta_{kj}), \tag{4}$$

where $w \sim \mathcal{IG}(a_w, b_w)$ and $\delta_0(\cdot)$ is the Dirac delta function.

We may not only introduce sparsity to the high-dimensional coefficient matrix but also sparsity to the precision matrix Ψ^{-1}, which implies that the residuals $\boldsymbol{u}_l = \boldsymbol{y}_l - \mathbf{X}\boldsymbol{\beta}_l$ and $\boldsymbol{u}_j = \boldsymbol{y}_j - \mathbf{X}\boldsymbol{\beta}_j$ for only a few pairs of response variables $l \neq j$ have non-zero partial correlations, assuming a multivariate normal distribution for the residuals. Such a sparse precision matrix can be conceptualised as a graph $\mathcal{G}$, with nodes representing the residual variables $\boldsymbol{u}_l$, and edges between them corresponding to non-zero elements of the precision matrix. Bottolo et al. (2021) used a hyper-inverse Wishart prior for Ψ instead of an inverse Wishart prior, i.e.

$$\Psi \sim \mathcal{HIW}_{\mathcal{G}}(v, \tau\mathbb{I}_m). \tag{5}$$

It assumes an underlying decomposable graph $\mathcal{G}$ between residuals. The hyper-inverse Whishart prior on decomposable graphs greatly enhances computational power since the parameters are updated within each clique and there is no computationally expensive normalisation constant to calculate. Since the fully Bayesian estimation procedure produces edges averaged over many different graphs, the posterior mean graph can well approximate non-decomposable graphs (Fitch et al., 2014). A sparse graph $\mathcal{G}$ can result in sparse Ψ^{-1}. So Bottolo et al. (2021) specified a $\mathcal{B}ernoulli(\eta)$ prior for each edge of the graph. Then, a Binomial prior is on the cardinality edge-set

$$|\mathcal{G}| \sim \mathcal{B}inomial(m(m-1)/2, \eta), \tag{6}$$

where $\eta \sim \mathcal{B}eta(a_\eta, b_\eta)$ controls the sparsity of the graph. Based on (5) and (6), the parameters σ^2 and ρ are indexed across the response variables of each clique of $\mathcal{G}$ rather than all response variables. In addition to sparse covariance selection, Bottolo et al. (2021) also used sparse variable selection for the predictor variables via a hotspot prior (i.e. a multiplicative prior) for the hyper-parameter ω_{kj} in $\gamma_{kj} \sim \mathcal{B}er(\omega_{kj})$. A guideline of prior specifications for the hyper-inverse Wishart prior and spike-and-slab prior can be found in Supplementary S1.

2.2 SSUR model with MRF prior

Figure 1 illustrates known relationships between drug responses and genomic predictors. As an example, imagine a group of drugs with the same mechanism of action, where the response of a cancer cell to these drugs depends on a certain gene to be silenced. Gene silencing can either occur via a genomic alteration (deletion event), missense mutation, or another down-regulation of gene expression. It might thus be observable in one or several omics features, e.g. gene expression, copy number variation, or mutation data. We may include such prior knowledge in the SUR model (1), instead of using independent or hotspot priors (Bottolo et al., 2021; Lewin et al., 2016; Richardson et al., 2011).

We propose to use an MRF prior for the latent indicator vector $\boldsymbol{\gamma} = \text{vec}\{\boldsymbol{\Gamma}\}$ to address prior structure for the associations between response variables and predictors. The MRF prior is

$$f(\boldsymbol{\gamma}|d, e, E) \propto \exp\{d\mathbb{1}^\top\boldsymbol{\gamma} + e\boldsymbol{\gamma}^\top E\boldsymbol{\gamma}\}, \tag{7}$$

where the scalar d controls overall model sparsity, scalar e determines the strength of the structure relationships between responses and predictors, and E is a symmetric $mp \times mp$ (possibly weighted) adjacency matrix representing a graph to include prior structure knowledge. Term $d\mathbb{1}^\top\boldsymbol{\gamma}$ in (7) can be generalised to $\boldsymbol{d}^\top\gamma$, where the vector $\boldsymbol{d}$ will assign different relative contributions to the prior selection probabilities of the predictors. To specify the scalar d, we refer to Lee et al. (2017) by using log-odds of a rough model sparsity (i.e. proportion of non-zero regression coefficients). To specify e, Stingo et al. (2011) suggested a separate simulation from (7) over a grid of values for e to detect the 'phase transition' value e_{pt}, and then specified a Beta prior on e/e_{pt}. However, due to much computational cost in high-dimensional γ, especially in multivariate regressions when searching some large values of e resulting in very dense models, we first estimate a large value $e_{\max}$ (see Supplementary S2 for more details) and then use a grid search for $e \in (0, e_{\max})$ to identify its optimal value with respect to the model's widely applicable information criterion introduced in Section 2.5.

For the E matrix, we assign a positive edge potential $\{k + j(p-1),\ \ k' + j'(p-1)\}$-element if the latent indicator variables γ_{kj} and $\gamma_{k'j'}$ are correlated. To illustrate the idea, we consider a simple case with three response variables (i.e. $\boldsymbol{y}_1, \boldsymbol{y}_2$ and $\boldsymbol{y}_3$) and four predictors (i.e. $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ and $\boldsymbol{x}_4$). When the predictors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are assumed *a priori* to be associated with responses $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, and $\boldsymbol{x}_3$ and $\boldsymbol{x}_4$ are assumed to be associated with $\boldsymbol{y}_3$, then E is a 12×12 matrix given by Equation (8). Any non-zero element in E can be any positive number which indicates a weight for the prior relationship between two latent indicator variables. Here for simplicity, we construct a symmetric E matrix and assume all non-zero weights to be 1.

$$E = \begin{array}{c} \\ \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \\ \gamma_{41} \\ \gamma_{12} \\ \gamma_{22} \\ \gamma_{32} \\ \gamma_{42} \\ \gamma_{13} \\ \gamma_{23} \\ \gamma_{33} \\ \gamma_{43} \end{array} \begin{pmatrix} \gamma_{11} & \gamma_{21} & \gamma_{31} & \gamma_{41} & \gamma_{12} & \gamma_{22} & \gamma_{32} & \gamma_{42} & \gamma_{13} & \gamma_{23} & \gamma_{33} & \gamma_{43} \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{8}$$

Note that we might not know all exact relationships between response variables and predictors, but we still formulate the matrix E based on what we know. For example, if we only know relationships between response variables and relationships between predictors, we can aggregate these relationships by $E_y \otimes E_x - \mathbb{I}$. Here, we use $-\mathbb{I}$ to only allow zero diagonals in E, because non-zero diagonals are already captured by the term $d\mathbb{1}^\top \boldsymbol{\gamma}$. For example, if we assume that $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are related w.r.t. each predictor, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are related w.r.t. each response variable, and $\boldsymbol{x}_3$ and $\boldsymbol{x}_4$ are related w.r.t. each response variable, this translates into the following three Kronecker products. We can then aggregate them by aligning their coordinates into the full matrix E.

$$\underbrace{E_y}_{\text{for } y_1 \text{ and } y_2} \otimes \underbrace{E_x}_{\text{for } x_1, x_2, x_3 \text{ and } x_4} - \mathbb{I} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \mathbb{I}_4 - \mathbb{I}_8.$$

$$\underbrace{E_y}_{\text{for } y_1, y_2 \text{ and } y_3} \otimes \underbrace{E_x}_{\text{for } x_1 \text{ and } x_2} - \mathbb{I} = \mathbb{I}_3 \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} - \mathbb{I}_6.$$

$$\underbrace{E_y}_{\text{for } y_1, y_2 \text{ and } y_3} \otimes \underbrace{E_x}_{\text{for } x_3 \text{ and } x_4} - \mathbb{I} = \mathbb{I}_3 \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} - \mathbb{I}_6.$$

2.3 SSUR model with MRF prior and random effects

The SSUR model with hotspot prior in Section 2.1 and SSUR model with MRF prior in Section 2.2 both assume independent and identically distributed samples conditional on the predictors. However, samples can be heterogeneous, especially in applications with large sample size. For example, large-scale drug screens may include cell line samples from different cancer tissue types. We address the heterogeneity of multiple sample groups by introducing random effects into the model similar to Chekouo et al. (2015).

Let $\mathbf{Z}_{n\times T}$ be indicator variables representing n samples from T heterogeneous groups. We define an SSUR model which includes spike-and-slab priors (4), hyper-inverse Wishart prior (5), MRF prior (7) and random effects, where the random effects $\boldsymbol{B}_0 = \{\beta_{0,tj}: t = 1, \ldots, T; j = 1, \ldots, m\}$, and all priors above are mutually independent:

$$\begin{aligned}
\mathbf{Y} &= \mathbf{Z}\boldsymbol{B}_0 + \mathbf{X}\boldsymbol{B} + \mathbf{U}, \\
\beta_{0,tj}|w_0 &\sim \mathcal{N}(0,\ w_0), \\
\beta_{kj}|\gamma_{kj}, w &\sim \gamma_{kj}\mathcal{N}(0,\ w) + (1-\gamma_{kj})\delta_0(\beta_{kj}), \\
w_0 &\sim \mathcal{IG}(a_{w_0}, b_{w_0}), \\
w &\sim \mathcal{IG}(a_w, b_w), \\
\boldsymbol{\gamma}|d, e, E &\propto \exp\{d\mathbb{1}^\top\boldsymbol{\gamma} + e\boldsymbol{\gamma}^\top E\boldsymbol{\gamma}\}, \\
\text{vec}\{\mathbf{U}\} &\sim \mathcal{N}(\mathbf{0},\ \Psi \otimes \mathbb{I}_n), \\
\Psi &\sim \mathcal{HIW}_{\mathcal{G}}(v, \tau\mathbb{I}_m), \\
\tau &\sim \mathcal{G}amma(a_\tau, b_\tau).
\end{aligned} \tag{9}$$

Here, we use a standard hierarchical prior to produce the random effects $\beta_{0,tj}$, i.e. they are conditionally independent given the variance w_0, and the hierarchical prior induces a marginal correlation within sample group (as with any frequentist or Bayesian random effects model). Let us look into details of the random effects. For any ith sample and jth response variable, we have $y_{ij} = x_{i\cdot}\boldsymbol{\beta}_j + z_{i\cdot}\boldsymbol{\beta}_{0,j} + u_{ij}$. For the ith sample, the covariance between the jth and j'th response variables is $\psi_{jj'}$ that is the jj'-element of Ψ, since

$$\mathbb{C}\text{ov}[y_{ij}, y_{ij'}] = \mathbb{C}\text{ov}[x_{i\cdot}\boldsymbol{\beta}_j + z_{i\cdot}\boldsymbol{\beta}_{0,j} + u_{ij},\ x_{i\cdot}\boldsymbol{\beta}_{j'} + z_{i\cdot}\boldsymbol{\beta}_{0,j'} + u_{ij'}] = \mathbb{C}\text{ov}[u_{ij}, u_{ij'}] = \psi_{jj'}.$$

Although the priors for the coefficients $\boldsymbol{B}_0$ and $\boldsymbol{B}$ in (9) do not provide any correlation between different responses for the same sample, the hyper-inverse Wishart prior on Ψ models correlations

between the response variables, and so does an inverse Wishart prior on Ψ. If we look at the reparametrisation (3) from the inverse Wishart prior, or similarly from the hyper-inverse Wishart prior, correlations between the response variables are contained in the reparametrised parameter $\boldsymbol{\rho}$.

For the jth response variable, the covariance between the ith and i'th samples is

$$\begin{aligned}\mathbb{C}\mathrm{ov}[y_{ij}, y_{i'j}] &= \mathbb{C}\mathrm{ov}[x_{i\cdot}\boldsymbol{\beta}_j + z_{i\cdot}\boldsymbol{\beta}_{0,j} + u_{ij}, x_{i'\cdot}\boldsymbol{\beta}_j + z_{i'\cdot}\boldsymbol{\beta}_{0,j} + u_{i'j}] = w x_{i\cdot}x_{i'\cdot}^{\top} + w_0 z_{i\cdot}z_{i'\cdot}^{\top} \\ &= \begin{cases} w x_{i\cdot}x_{i'\cdot}^{\top}, & \text{if } i\text{th and } i'\text{th samples belong to different groups,} \\ w x_{i\cdot}x_{i'\cdot}^{\top} + w_0, & \text{if } i\text{th and } i'\text{th samples belong to the same group,} \end{cases}\end{aligned}$$

in which the hyper-parameter w_0 in the random effect determines the correlation between two samples from the same group.

We would like a weakly informative prior for $\beta_{0,tj}$ based on previous studies or expert knowledge in applications. In pharmacogenomic studies from multiple cancer tissues, for predicting drug responses a tissue effect is usually stronger than a gene effect. Therefore, it is appropriate to specify a larger hyper-parameter w_0 than w.

2.4 Posterior computation

Posterior inference for the SSUR model with the MRF prior with or without additional random effects can be done in a similar manner to Bottolo et al. (2021). For the SUR model (2) with a hyper-inverse Wishart prior for the residual covariance matrix Ψ and an MRF prior for the latent indicator variables γ, the joint posterior distribution is

$$\begin{aligned} &f(\boldsymbol{B}, \boldsymbol{\Gamma}, w, \boldsymbol{\rho}, \boldsymbol{\sigma}^2, \tau, \mathcal{G}, \eta | \mathbf{Y}, \mathbf{X}) \\ &= f(\mathbf{Y}|\mathbf{X}, \boldsymbol{B}, \boldsymbol{\rho}, \boldsymbol{\sigma}^2) f(\boldsymbol{B}|\boldsymbol{\Gamma}, w) f(\boldsymbol{\Gamma}|E, d, e) f(w) f(\boldsymbol{\rho}|\boldsymbol{\sigma}^2, \tau, \mathcal{G}) f(\boldsymbol{\sigma}^2|\tau, \mathcal{G}) f(\tau) f(\mathcal{G}|\eta) f(\eta) \\ &= \prod_j f(\boldsymbol{y}_j \,|\, \boldsymbol{y}_{-j}, \mathbf{X}, \boldsymbol{B}, \boldsymbol{\rho}, \boldsymbol{\sigma}^2) \prod_{k,j} f(\beta_{kj}|\gamma_{kj}, w) f(\boldsymbol{\gamma}|E, d, e) f(w) \prod_{j,l<j} f(\rho_{jl} \,|\, \sigma_j^2, \tau, \mathcal{G}) \prod_j f(\sigma_j^2|\tau, \mathcal{G}) f(\tau) f(\mathcal{G}|\eta) f(\eta), \end{aligned} \tag{10}$$

where $\boldsymbol{y}_{-j} = \{\boldsymbol{y}_l : l \neq j, l = 1, \ldots, m\}$, $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}^2$ are vectors of $\{\rho_{jl}\}$ and $\{\sigma_j^2\}$, respectively. Since $\boldsymbol{y}_j|\mathbf{X}, \boldsymbol{B}, \boldsymbol{\rho}, \boldsymbol{\sigma}^2$ is normally distributed with mean $\mathbf{X}\boldsymbol{\beta}_j + \sum_{l<j} \boldsymbol{u}_l \rho_{jl}$ and variance $\sigma_j^2 \mathbb{I}_n$, we can obtain the full conditional distributions of the regression coefficients $\boldsymbol{\beta}_j$, w, σ_j^2, ρ_{jl} and τ. The posterior distribution of the latent indicator variable $\gamma = \boldsymbol{vec}\{\boldsymbol{\Gamma}\}$ is estimated by a Metropolis–Hastings sampler. The graph $\mathcal{G}$ of the hyper-inverse Wishart prior, $f(\mathcal{G}|\eta)$, is sampled via a junction tree sampler which is essentially Metropolis–Hastings sampling (Green & Thomas, 2013), see Bottolo et al. (2021) for more details in a SUR model framework. If there are random effects for sample groups as in (9), the joint posterior distribution (10) includes parameters $\boldsymbol{B}_0$ and w_0, i.e.

$$\begin{aligned} &f(\boldsymbol{B}_0, \boldsymbol{B}, \boldsymbol{\Gamma}, w_0, w, \boldsymbol{\rho}, \boldsymbol{\sigma}^2, \tau, \mathcal{G}, \eta | \mathbf{Y}, \mathbf{X}, \mathbf{Z}) \\ &= f(\mathbf{Y}|\mathbf{X}, \mathbf{Z}, \boldsymbol{B}_0, \boldsymbol{B}, \boldsymbol{\rho}, \boldsymbol{\sigma}^2) f(\boldsymbol{B}_0|w_0) f(w_0) f(\boldsymbol{B}|\boldsymbol{\Gamma}, w) f(\boldsymbol{\Gamma}|E, d, e) f(w) f(\boldsymbol{\rho}|\boldsymbol{\sigma}^2, \tau, \mathcal{G}) f(\boldsymbol{\sigma}^2|\tau, \mathcal{G}) f(\tau) f(\mathcal{G}|\eta) f(\eta). \end{aligned}$$

We implement Gibbs samplers to obtain posterior estimates for $\boldsymbol{B}$, $\boldsymbol{\rho}$, and $\boldsymbol{\sigma}^2$, and update the latent indicator variable $\boldsymbol{\Gamma}$ via a Metropolis–Hastings sampler with parallel tempering in the same way as Bottolo et al. (2021). Thompson sampling (Russo et al., 2018) is used to derive the proposal for each latent indicator γ_{kj}. The hyper-parameter τ is updated via a random walk Metropolis sampler as proposed by Bottolo et al. (2021). To overcome the prohibitive computational time in high-dimensional settings, the ESS algorithm (Bottolo & Richardson, 2010; Richardson et al., 2011) is used to update the posteriors. For each iteration of the MCMC sampler, after sampling the latent indicator variables $\boldsymbol{\Gamma}$, we first update the hyper-parameters $(\tau, w, w_0, \mathcal{G})$, then update the parameters $\boldsymbol{\sigma}^2$ and $\boldsymbol{\rho}$, and finally the regression coefficient matrices $\boldsymbol{B}$ and $\boldsymbol{B}_0$ (see Supplementary S3). ESS is an evolutionary Monte Carlo method (Liang & Wong,

2000) which is based on running multiple parallel Markov chains at different temperatures. 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 indicators between any two parallel tempered chains. The temperature across all response variables is adapted based on the acceptance rate of the global exchange operator. The ESS algorithm with parallel tempering is effective in searching a high-dimensional model space with multiple modes (Bottolo & Richardson, 2010).

2.5 Model performance evaluation

To evaluate the performance of the proposed approach, we focus on structure recovery and prediction performance. The structure recovery includes the estimation of the latent indicator variable $\boldsymbol{\Gamma}$ which captures the relationships between response variables and high-dimensional predictors, and the estimation of the graph $\mathcal{G}$ which addresses the residual relationships between response variables. After thresholding the posterior means of $\boldsymbol{\Gamma}$ at a cutoff value 0.5 into 0 and 1, we calculate accuracy, sensitivity, and specificity for evaluating the performance of variable selection. Note that the selection of variables based on this threshold corresponds to the median probability model (MPM) (Barbieri & Berger, 2004). Similarly, we calculate the performance of covariance selection based on the posterior means of $\mathcal{G}$. Accuracy denotes the percentage of both true non-zeros that are correctly estimated as non-zeros and true zeros correctly estimated as zeros, while sensitivity denotes the percentage of true non-zeros estimated as non-zeros, and specificity denotes the percentage of true zeros estimated as zeros. Predictive performance of Bayesian models for new data points can be measured by the expected log pointwise predictive density (elpd), which can be assessed by leave-one-out cross-validation (LOO), or by the widely applicable information criterion (WAIC) (Vehtari et al., 2017). We also calculate the root mean squared prediction error (RMSPE) measured on an independent test dataset for the median probability model (Barbieri & Berger, 2004; Barbieri et al., 2021) in addition to the training data root mean squared error (RMSE).

Vehtari et al. (2017) proposed an efficient computation for the Bayesian LOO estimate of out-of-sample predictive fit $\text{elpd}_{\text{loo}} = \sum_{j=1}^{m}\sum_{i=1}^{n} \log f(y_{ij}|\boldsymbol{y}_{(-i)j})$, where $f(y_{ij}|\boldsymbol{y}_{(-i)j})$ is the leave-one-out predictive density given the data $\boldsymbol{y}_{(-i)j}$ of the jth response variable without the ith observation. The LOO is estimated by

$$\widehat{\text{elpd}}_{\text{loo}} = \sum_{j=1}^{m}\sum_{i=1}^{n} \frac{1}{\frac{1}{N}\sum_{l=1}^{N}\frac{1}{f(y_{ij}|\boldsymbol{\vartheta}^{(l)})}},$$

where N is the length of an MCMC chain and $f(y_{ij}|\boldsymbol{\vartheta}^{(l)})$ is the likelihood conditional on $\boldsymbol{\vartheta}^{(l)}$ which is the MCMC estimates of all related parameters at the lth iteration. The WAIC is estimated by

$$\widehat{\text{elpd}}_{\text{waic}} = \widehat{\text{elpd}}_{\text{loo}} - \sum_{j=1}^{m}\sum_{i=1}^{n} \mathbb{V}\text{ar}_{t=1}^{N}[\log f(y_{ij}|\boldsymbol{\vartheta}^{(t)})],$$

where the second term above is used as a measure of the model complexity.

For future prediction, a single model may be required in some cases, for practical reasons or for simplicity. Barbieri and Berger (2004) suggested the median probability model (MPM), which is defined for each coefficient to be $\mathbb{E}[\beta_{kj}|\gamma_{kj}=1, data]$ if $\mathbb{P}\{\gamma_{kj}=1|data\} > 0.5$, or 0 otherwise. It can be estimated through MCMC estimates:

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

where $\gamma_{kj}^{(l)}$ is the estimate of the lth MCMC iteration for the latent indicator variable of β_{kj}. After obtaining $\hat{\boldsymbol{B}}_{\text{MPM}} = \{\hat{\beta}_{kj,\text{MPM}}\}$, the

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

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

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

Although the posterior probability of inclusion $\frac{1}{N}\sum_{l=1}^{N}\gamma_{kj}^{(l)}$ can measure the uncertainty of variable selection, the uncertainty (posterior distribution) of $\hat{\boldsymbol{B}}_{\text{MPM}}$ cannot be estimated straightforwardly via MCMC. One can use resampling-based methods to report the means and standard deviations of $\hat{\boldsymbol{B}}_{\text{MPM}}$ for stability selection. Similarly, the uncertainty of the out-of-sample predictions based on $\hat{\boldsymbol{B}}_{\text{MPM}}$ can be measured by resampling-based methods as well.

3 Simulation study

In this section, our SSUR model with MRF prior, denoted as SSUR-MRF, is evaluated w.r.t. structure recovery for the regression coefficient matrix and prediction performance of responses. We set up two simulation scenarios: one with independent samples and the other with heterogeneous and correlated samples. In the first scenario, our approach is compared with the SSUR model with a hotspot prior, denoted as SSUR-hotspot which was studied by Bottolo et al. (2021), and also compared with the SSUR model with a Bernoulli prior, denoted as SSUR-Ber which was studied by Richardson et al. (2011). In the second scenario, our approach is compared with the SSUR-MRF model without random effects.

3.1 Simulation scenarios

We design a network (Figure 2a) to construct a complex structure between 20 response variables and 300 predictors. It assumes that the responses are divided into six groups, and the first 120 predictors are divided into nine groups. The first group of responses ($\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_5\}$) is related to four groups of predictors ($\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_5\}$, $\{\boldsymbol{x}_{30}, \ldots, \boldsymbol{x}_{50}\}$, $\{\boldsymbol{x}_{51}, \ldots, \boldsymbol{x}_{60}\}$, and $\{\boldsymbol{x}_{110}, \ldots, \boldsymbol{x}_{120}\}$). The second group of responses ($\{\boldsymbol{y}_6, \ldots, \boldsymbol{y}_{12}\}$) is also related to four predictor groups ($\{\boldsymbol{x}_{10}, \ldots, \boldsymbol{x}_{20}\}$, $\{\boldsymbol{x}_{51}, \ldots, \boldsymbol{x}_{60}\}$, $\{\boldsymbol{x}_{70}, \ldots, \boldsymbol{x}_{90}\}$, and $\{\boldsymbol{x}_{110}, \ldots, \boldsymbol{x}_{120}\}$). The third group of responses ($\{\boldsymbol{y}_{13}, \ldots, \boldsymbol{y}_{15}\}$) is related to three predictor groups ($\{\boldsymbol{x}_{30}, \ldots, \boldsymbol{x}_{50}\}$, $\{\boldsymbol{x}_{70}, \ldots, \boldsymbol{x}_{90}\}$, and $\{\boldsymbol{x}_{110}, \ldots, \boldsymbol{x}_{120}\}$). The fourth group of responses ($\{\boldsymbol{y}_{16}, \ldots, \boldsymbol{y}_{18}\}$) is related to one predictor $\{\boldsymbol{x}_{121}\}$. The fifth group, a single response variable $\{\boldsymbol{y}_{19}\}$, is related to one predictor $\{\boldsymbol{x}_{122}\}$. The sixth group, a single response variable $\{\boldsymbol{y}_{20}\}$, is related to one predictor $\{\boldsymbol{x}_{123}\}$. Corresponding to this network structure between responses and predictors, a sparse latent indicator variable $\boldsymbol{\Gamma}$ (Figure 2b) reflects the associations between response variables and predictors in the SUR model (1). In addition, we design a decomposable graph $\mathcal{G}$ (Figure 2c) to reflect the residual structure between the response variables. The graph $\mathcal{G}$ has six blocks representing six subgroups of responses that cannot be explained by the linear predictor $\mathbf{XB}$, which makes the modelling more challenging. The information in $\mathcal{G}$ is included in the residuals and can be expected to be recovered by statistical models.

In scenario 1, the response and predictor datasets are generated based on a multivariate linear regression model $\mathbf{Y} = \mathbb{1}\boldsymbol{\alpha}^\top + \mathbf{X}\boldsymbol{B}_{\boldsymbol{\Gamma}} + \mathbf{U}$. The intercepts $\boldsymbol{\alpha} = \{\alpha_j\}$ and input data $\mathbf{X} = \{x_{ik}\}$ ($i = 1, \ldots, 250$; $k = 1, \ldots, 300$; $j = 1, \ldots, 20$) are simulated independently from the standard normal distribution. The regression coefficients $\boldsymbol{B} = \{\beta_{kj}\}$ ($k = 1, \ldots, 300$; $j = 1, \ldots, 20$) are also simulated independently from the standard normal distribution but truncated by the latent indicator variable $\boldsymbol{\Gamma} = \{\gamma_{kj}\}$, i.e. $\boldsymbol{B}_{\boldsymbol{\Gamma}} = \{\beta_{kj}\mathbb{1}_{\{\gamma_{kj}=1\}}\}$. The noise matrix $\mathbf{U}$ is simulated based on the multivariate normal distributed $\tilde{\mathbf{U}}$ and a G-Wishart distribution (Mohammadi & Wit, 2019). We first simulate the G-Wishart distribution $P \sim \mathcal{W}_{\mathcal{G}}(3, M)$ where diagonals of M are 1 and the off-diagonals are 0.5 and then use Cholesky decomposition $\text{chol}(P^{-1})$ to obtain the noise matrix $\mathbf{U} = \tilde{\mathbf{U}} \cdot \text{chol}(P^{-1})$. Independent datasets $\mathbf{X}^*$ and $\mathbf{Y}^*$ are simulated based on the same scenario as

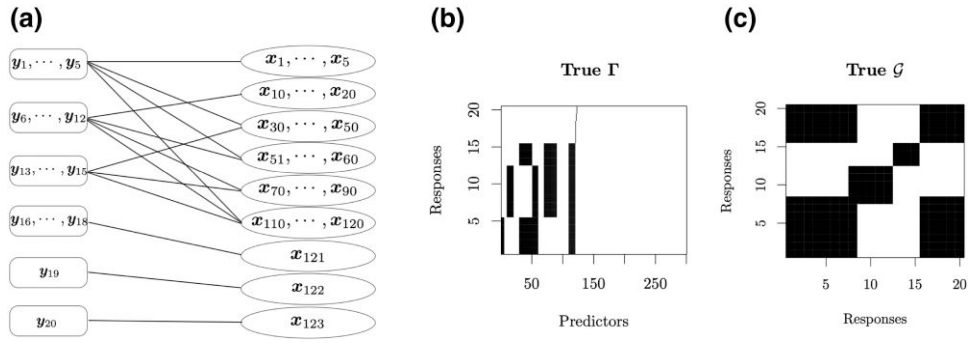

Figure 2. Simulation scenarios: True relationships between response variables and predictors. (a) Network structure between **Y** and **X**. (b) Sparse latent indicator variable $\boldsymbol{\Gamma}$ for the associations between **Y** and **X** in the SUR model. Black blocks indicate non-zero coefficients and white blocks indicate zero coefficients. (c) Additional structure in the residual covariance matrix $\mathcal{G}$ between response variables not explained by $\mathbf{XB}$. Black blocks indicate correlated residuals of the corresponding response variables and white blocks indicate uncorrelated residuals of the corresponding response variables.

validation data. In scenario 2, $\mathbf{X}$, $\boldsymbol{\Gamma}$, $\boldsymbol{B}_{\boldsymbol{\Gamma}}$, and $\mathbf{U}$ are generated in the same manner as scenario 1. We also include group indicators $\mathbf{Z}$ with independent row vectors $z_i \sim multinomial(0.1, 0.2, 0.3, 0.4)$ ($i = 1, \ldots, n$ and the number of groups is set to $T = 4$), and random effects $\boldsymbol{B}_0$ with each group effect from $\mathcal{N}(0, 2^2)$. The response dataset is generated from a linear mixed model $\mathbf{Y} = \mathbf{X}\boldsymbol{B}_{\boldsymbol{\Gamma}} + \mathbf{Z}\boldsymbol{B}_0 + \mathbf{U}$. Independent validation datasets are also simulated based on scenario 2. The algorithms for the two simulation scenarios can be found in Supplementary S4. Both simulation algorithms generate validation datasets independently with the same sample size to evaluate the performance of the proposed methods. Since it is not practical to save all intermediate MCMC estimates for obtaining credible intervals, we repeat every scenario 50 times with different random seeds and every evaluation metric is reported as the mean and standard deviation over the 50 repeated simulations.

3.2 Comparison of the SSUR-hotspot and SSUR-MRF models

We first compare our proposed SSUR-MRF model to the SSUR-hotspot model on simulated data generated with scenario 1. Our approach uses the network in Figure 2a as prior information to construct edge potentials for the MRF prior as illustrated in Section 2.2. Throughout this article, we refer to a predictor as being selected or identified, if the corresponding latent indicator variable has posterior mean larger than 0.5. Figure 3 shows that SSUR-hotspot and our SSUR-MRF both have good recovery for the residual structure between response variables (i.e. $\mathcal{G}$). However, SSUR-MRF has better structure recovery of the latent indicator variable $\boldsymbol{\Gamma}$. Table 1 reports higher accuracy, sensitivity and specificity of the estimate for $\boldsymbol{\Gamma}$ by SSUR-MRF than SSUR-hotspot. The two methods have similar $\widehat{\text{elpd}}_{\text{loo}}$ and $\widehat{\text{elpd}}_{\text{waic}}$, but our approach has smaller RMSE and RMSPE.

3.3 Sensitivity analysis for SSUR-MRF

The MRF prior can be strongly informative, as for example the graph E in the previous subsection was constructed in full correspondence with the true relationships $\boldsymbol{\Gamma}$ between the simulated response variables and predictors, i.e. the assumed biological information in E is completely true. However, in real applications, this given biological information in the graph E can be mis-specified, resulting in the incorrect deletion of edges (false negatives, i.e. non-zero entries in the adjacency matrix being wrongly specified as zero) and/ or incorrect addition of edges (false positives, i.e. zero entries in the adjacency matrix being wrongly specified as non-zero). Here we manipulate the construction of the graph E in different ways for sensitivity analysis, that is to assess how the model performance is affected by such mis-specifications in E. Starting from the previously constructed edge potentials E, we partially delete true edge potentials, either uniformly or non-

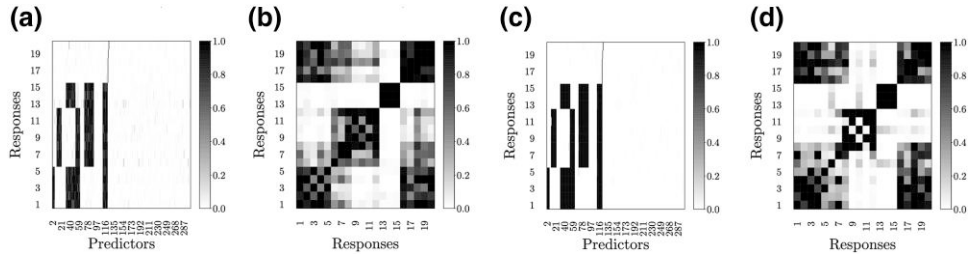

Figure 3. Results for simulation scenario 1: Posterior mean of $\mathbf{\Gamma}$ and $\mathcal{G}$ by the SSUR-hotspot and SSUR-MRF models from one simulated data set. (a) $\hat{\mathbf{\Gamma}}$ from the SSUR-hotspot. (b) $\hat{\mathcal{G}}$ from the SSUR-hotspot. (c) $\hat{\mathbf{\Gamma}}$ from the SSUR-MRF. (d) $\hat{\mathcal{G}}$ from the SSUR-MRF.

uniformly, or add noise edges, or aggregate Kronecker products between the three response groups and six predictor groups as shown in Figure 2a. The four cases are as follows:

- **Case 1:** delete 1%, 10%, 50%, or 90% edges uniformly from the fully informative E. In this case for every block in $\mathbf{\Gamma}$, some corresponding edge potentials in E are kept.
- **Case 2:** delete 1%, 10%, 50%, 90%, or 100% edges non-uniformly in consecutive chunks from the edge list[1] of the fully informative E. In this case, for some blocks in $\mathbf{\Gamma}$ all corresponding edge potentials in E are deleted.
- **Case 3:** add 0.1%, 0.5%, 1%[2] noise edges to the fully informative E.
- **Case 4:** aggregate Kronecker products between response groups and predictor groups (see guidance in Section 2.2).

Table 2 Case 1 shows that our SSUR-MRF model can identify well truly associated predictors w.r.t. accuracy, sensitivity and specificity of the estimated $\mathbf{\Gamma}$, and have stable prediction performance w.r.t. RMSE and RMSPE, when deleting 1%, 10%, 50%, or 90% true edges uniformly. This indicates that our approach can recover a good structure of Γ and good prediction performance of responses, even if only a little true association knowledge across all patterns of $\mathbf{\Gamma}$ is used in the MRF prior. Case 2 (Table 2) where some of the patterns/blocks in $\mathbf{\Gamma}$ are fully unknown (i.e. when the corresponding blocks of edges in E are deleted) in the MRF prior, the sensitivity of variable selection and prediction performance w.r.t. RMSE and RMSPE becomes slightly worse. Figure 4 indicates that the information of the deleted blocks cannot be recovered fully, but will instead be estimated with a sparser $\hat{\mathbf{\Gamma}}$. Supplementary S5 shows slightly worse residual structure recovery (i.e. $\hat{\mathcal{G}}$) when deleting more edges non-uniformly. However, even the worst-case scenario in Case 2, when all edges are deleted, i.e. when the MRF prior with $E = \mathbf{0}$ degenerates to a Bernoulli prior without any known structure information between variables (named as SSUR-Ber), has similar performance to the SSUR-hotspot model in Table 1. Case 3 (Table 2), where adding noise edges, shows similar variable selection and prediction performance to using true potential edges. Finally, Case 4 (Table 2), where aggregating Kronecker products for the edge potentials in the MRF prior, the variable selection remains similar to using true potential edges. Here, $\widehat{\text{elpd}}_{\text{loo}}$ and $\widehat{\text{elpd}}_{\text{waic}}$ do not change much between different cases, but they can be used as the objective function to optimise hyper-parameters.

3.4 Results and discussion of SSUR-MRF with random effects

In the simulation scenario 2, $T = 4$ sample group variables are simulated to assess the performance of our SSUR-MRF model with random effects. Figure 5a, c and Table 3 show similar recovery of the latent indicator variable $\mathbf{\Gamma}$ w.r.t. accuracy, sensitivity, and specificity for both SSUR-MRF with

[1] The coordinates of all non-zero entries of E are put in an edge list in order. Deleting edge potentials uniformly, e.g. deleting 1%, means that the $1 + (1 - 1/|E|)/1\% \cdot \{0{:}(1\% \cdot |E|)\}$th edges of the edge list are deleted. Deleting 1% edges non-uniformly (i.e. blocks of edges) means that the last 1% edges in the edge list are deleted. The edge list includes the edges of each pattern (i.e. association block) together. Adding 1% noise edges means that $1\% \cdot mp(mp-1)/2$ wrong edges are included randomly.

[2] Note that 1% already exceeds the total number of true edges that are $\sim 0.3\%$ of all possible edges.

Table 1. Results for simulation scenario 1: Performance (mean/standard deviation) of variable selection and prediction of models SSUR-hotspot and SSUR-MRF prior

	Accuracy	Sensitivity	Specificity	RMSE	RMSPE
SSUR-hotspot				0.549 (0.3454)	0.618 (0.3528)
Γ	0.998 (0.0010)	0.985 (0.0077)	1.000 (0.0004)		
$\mathcal{G}$	0.868 (0.0422)	0.804 (0.0671)	0.933 (0.0633)		
SSUR-MRF				0.394 (0.2487)	0.423 (0.2230)
Γ	0.991 (0.0001)	0.998 (0.0005)	0.990 (0.0001)		
$\mathcal{G}$	0.865 (0.0364)	0.736 (0.0718)	0.997 (0.0061)		

and without random effects. However, SSUR-MRF model without random effects is difficult to recover the residual graph structure $\mathcal{G}$ (Figure 5d), while the model with random effects can recover well the true structure (Figure 5b). See also Table 3, which reports the recovery performance of $\mathcal{G}$ w.r.t. accuracy, sensitivity, and specificity when thresholding its posterior mean at 0.5. For the response prediction, SSUR-MRF with random effects has smaller RMSE and RMSPE than SSUR-MRF without random effects (Table 3). In addition, for the accuracy of estimated regression coefficients, $\frac{1}{\sqrt{mp}}\|\hat{\boldsymbol{B}}_{\text{MPM}} - \boldsymbol{B}\|_{\ell_2}$ by SSUR-MRF without random effects has larger error (0.0326, SD = 0.2095) than by SSUR-MRF with random effects (0.006, SD = 0.0050). Furthermore, in the simulation scenario 2 we set random effects $\boldsymbol{B_0} = \boldsymbol{0}$ (i.e. there are no random effects in the ground truth), and applied the SSUR-MRF with random effects model. Table S6.1 in the supplementary materials indicates that the estimated random effects are fairly negligible.

4 Analysis of the pharmacogenomic screen

4.1 Pharmacogenomic data

We apply our approach to the Genomics of Drug Sensitivity in Cancer (GDSC) database (Garnett et al., 2012; Yang et al., 2013) to study the relationships between multiple cancer drugs and high-dimensional genomic features characterising cancer cell lines. The pharmacological and genomic data are from the archived dataset release 5 (https://www.cancerrxgene.org) preprocessed by Garnett et al. (2012). We would like to investigate how the MRF prior can help to improve inference for groups of drugs that are known to have correlated response; we therefore select two groups of cancer drugs with similar molecular targets and the generic non-targeted chemotherapy agent Methotrexate: four MAPK inhibitors (RDEA119, PD-0325901, CI-1040 AZD6244), two Bcr–Abl tyrosine kinase inhibitors (Nilotinib, Axitinib), and one chemotherapy agent (Methotrexate).

The seven drugs were tested on 499 cell lines from 13 cancer tissue types with complete drug sensitivity values. The drug sensitivity of the cell lines was summarised by the $\log_{10}$ of the half maximal inhibitory concentration IC_{50}, which is the drug concentration that caused inhibition of 50% cell viability as determined from the drug concentration-response curve for each in vitro experiment. Note that smaller $\log_{10}(IC_{50})$ values indicate higher sensitivity of a cell line to the drug; therefore a negative regression coefficient indicates that a positive increment of the value of a feature is associated with an increase in drug sensitivity. In order to explore the relationships between the three groups of drugs and the genomic profiles of the cell lines, we preselect mutation and copy number features by following Garnett et al. (2012). Garnett et al. (2012) sequenced 64 of the most frequently mutated cancer genes and determined seven of the most commonly rearranged cancer genes. We removed three of the 71 mutation features that did not have any variation across the 499 cell lines in our analysis, which resulted in 68 mutation features (binary). Garnett et al. (2012) filtered copy number data of 426 cancer genes according to the Cancer Genome Project. To make a trade-off between the computational efficiency and amount of information from gene expression data, we preselect the most variable gene expression features which together explain a prespecified proportion of the cumulative variance across the cell lines. For sensitivity analysis, we choose three subsets of gene expression data explaining 10%, 30%, and 50% of the cumulative variance, which

Table 2. Results for simulation scenario: Sensitivity analysis of SSUR-MRF with different MRF priors

	Accuracy	Sensitivity	Specificity	$\widehat{\text{elpd}}_{\text{loo}}$	$\widehat{\text{elpd}}_{\text{waic}}$	RMSE	RMSPE
Case 1							
delete edges uniformly							
1%	0.991 (0.0002)	0.999 (0.0005)	0.990 (0.0002)	−16308.8 (117.28)	−16309.9 (117.28)	0.389 (0.2504)	0.428 (0.2216)
10%	0.991 (0.0002)	0.998 (0.0006)	0.990 (0.0002)	−16308.1 (117.69)	−16309.3 (117.60)	0.390 (0.2505)	0.429 (0.2225)
50%	0.991 (0.0001)	0.998 (0.0006)	0.990 (0.0001)	−16308.7 (118.19)	−16309.8 (118.03)	0.386 (0.2527)	0.424 (0.2252)
90%	0.995 (0.0004)	0.992 (0.0021)	0.996 (0.0003)	−16308.0 (118.95)	−16309.3 (118.71)	0.416 (0.2694)	0.464 (0.2358)
Case 2							
delete edges non-uniformly							
1%	0.991 (0.0001)	0.998 (0.0007)	0.990 (0.0001)	−16309.8 (116.52)	−16311.1 (116.39)	0.407 (0.2559)	0.452 (0.2267)
10%	0.991 (0.0006)	0.990 (0.0044)	0.991 (0.0001)	−16308.7 (117.84)	−16309.8 (117.70)	0.454 (0.2662)	0.504 (0.2365)
50%	0.993 (0.0012)	0.982 (0.0092)	0.994 (0.0002)	−16264.3 (179.95)	−16267.1 (179.53)	0.472 (0.2835)	0.528 (0.2667)
90%	0.994 (0.0029)	0.968 (0.0223)	0.998 (0.0002)	−16364.3 (118.97)	−16369.0 (118.90)	0.488 (0.3001)	0.570 (0.2642)
100% (i.e. SSUR-Ber)	0.997 (0.0023)	0.974 (0.0180)	1.000 (0.0002)	−16278.3 (133.61)	−16281.1 (131.92)	0.547 (0.3315)	0.624 (0.3168)
Case 3							
add noise edges							
0.1%	0.991 (0.0001)	0.998 (0.0005)	0.990 (0.0001)	−16308.9 (121.33)	−16310.2 (121.10)	0.413 (0.2501)	0.460 (0.2149)
0.5%	0.991 (0.0001)	0.999 (0.0004)	0.989 (0.0001)	−16308.2 (116.67)	−16309.5 (116.66)	0.391 (0.2503)	0.431 (0.2222)
1%	0.991 (0.0002)	0.998 (0.0009)	0.990 (0.0002)	−16307.0 (118.23)	−16308.3 (118.12)	0.406 (0.2518)	0.451 (0.2216)
Case 4							
aggregate Kronecker products	0.992 (0.0003)	0.997 (0.0016)	0.992 (0.0001)	−16308.3 (117.64)	−16309.6 (117.53)	0.411 (0.2592)	0.457 (0.2289)

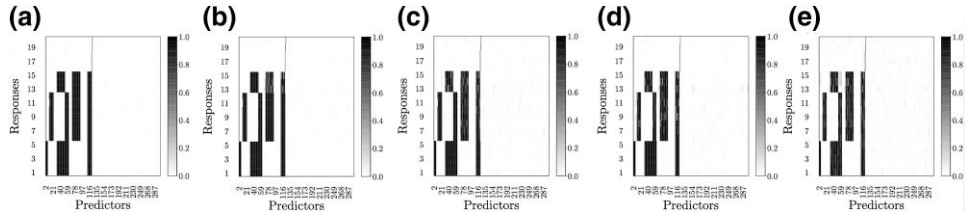

Figure 4. Results for simulation scenario 1: Sensitivity analysis for case 2, i.e. when blocks of edges are deleted from one simulated data set. (a) Delete 1% edges non-uniformly. (b) Delete 10% edges non-uniformly. (c) Delete 50% edges non-uniformly. (d) Delete 90% edges non-uniformly. (e) Delete 100% edges.

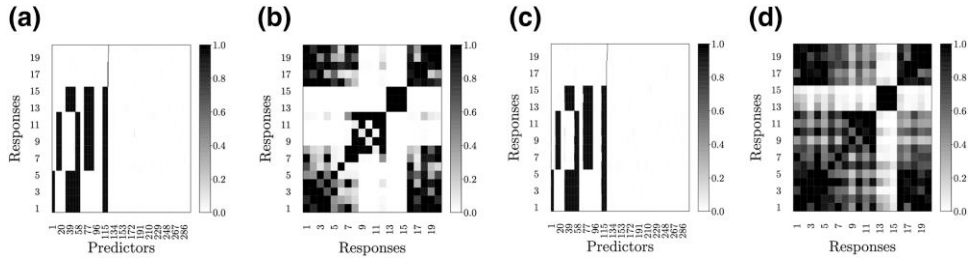

Figure 5. Results for simulation scenario 2: Posterior mean of $\mathbf{\Gamma}$ and $\mathcal{G}$ by the SSUR-MRF with random effects based on one simulated data set from scenario 2. (a) $\hat{\mathbf{\Gamma}}$ from the SSUR-MRF with random effects. (b) $\hat{\mathcal{G}}$ from the SSUR-MRF with random effects. (c) $\hat{\mathbf{\Gamma}}$ from the SSUR-MRF without random effects. (d) $\hat{\mathcal{G}}$ from the SSUR-MRF without random effects.

Table 3. Results for simulation scenario 2: Performance (mean/standard deviation) of variable selection and prediction by SSUR-MRF with and without random effects

	Accuracy	Sensitivity	Specificity	RMSE	RMSPE
With random effects				0.389 (0.2575)	0.436 (0.2296)
Γ	0.991 (0.0002)	1.000 (0.0008)	0.990 (0.0002)		
$\mathcal{G}$	0.872 (0.0378)	0.996 (0.0748)	0.996 (0.0082)		
Without random effects				2.751 (0.1087)	2.767 (0.1072)
Γ	0.990 (0.0008)	0.990 (0.0054)	0.990 (0.0002)		
$\mathcal{G}$	0.771 (0.0939)	0.870 (0.0439)	0.669 (0.2095)		

results in 269, 1,175, and 2,602 gene expression features, respectively. This creates three data sets including both gene expression, copy number variation and mutation information, to predict drug sensitivity responses, i.e. Feature set I with 763 predictors, Feature set II with 1,669 predictors, Feature set III with 3,096 predictors.

4.2 Prior specification and model setup

To construct edge potentials for the MRF prior in the proposed model (9) in Section 2.3, we summarise some known biological relationships between the drugs and genomic information. First, all features (gene expression, copy number variation, mutation) corresponding to the same gene are assumed to be related. Such group of features are likely to be identified together corresponding to each drug, i.e. if one feature for a certain gene is a predictor of drug sensitivity, then the other

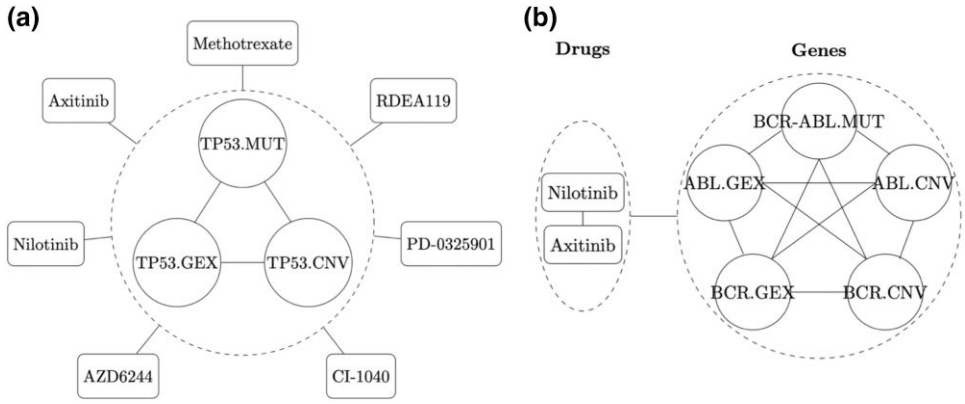

Figure 6. GDSC data application: Illustration of the assumed relationships between drugs and related gene features, which are used for the MRF prior (Zhao et al., 2021). (a) Illustration of gene TP35 as one example with its corresponding features presented in three data sources. All seven drugs are shown to indicate that the relationship between the three gene features is valid in relation to all drugs in the dataset. (b) Illustration of Bcr–Abl fusion gene with its corresponding features related to the two Bcr–Abl tyrosine kinase inhibitor drugs. The rectangles indicate drugs, solid circles indicate gene features and dashed circles indicate that the elements inside are related. The edges between drugs and dashed circled gene features indicate assumed associations between the gene features and the drug sensitivity measurements for the drugs. The names with suffix '.GEX', '.CNV', and '.MUT' indicate features of expression, copy number variation, and mutation, respectively.

features corresponding to the same gene are more likely to be predictors as well. This is illustrated in Figure 6a and results in a Kronecker product for the edge potentials

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

Second, the two Bcr–Abl tyrosine kinase inhibitors were developed to inhibit Bcr–Abl tyrosine kinase activity and proliferation of Bcr–Abl expressing cells, so the point mutation BCR–ABL, and features associated with genes BCR and ABL are related and likely to be identified together corresponding to the two Bcr–Abl inhibitors. This is illustrated in Figure 6b and results in a Kronecker product for the edge potentials

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

Third, the four MAPK inhibitors were developed to reduce the activity of the MAPK pathway, so genes representing the MAPK pathway are likely to be identified together as potential predictor variables for drug sensitivity of the four MAPK inhibitors. Based on the set of 267 genes involved in the MAPK pathway from the Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY database, Feature sets I, II, and III include 27, 39, and 63 genes of the MAPK pathway, respectively. Correspondingly, we can use the Kronecker product to construct the edge potentials for each feature set. Here, an edge potential of the E matrix, i.e. an edge weight, is 2, if the corresponding two features are both from the same gene and also belong to one group of drug target genes. Finally, we aggregate the individual Kronecker products by aligning their coordinates in the final E matrix for the MRF prior.

Other prior specifications, MCMC settings and diagnostics can be found in Supplementary S7 and S8. For comparison, we also run almost the same model as SSUR-MRF but with

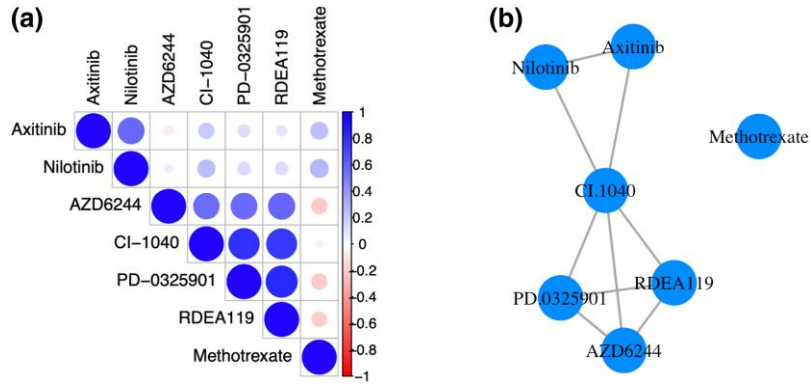

Figure 7. GDSC data application: (a) Drug responses' Pearson correlations where the colour bar indicates the correlation coefficient and the size of a circle indicates the absolute value of the correlation coefficient and (b) estimated residual structure by the SSUR-MRF model based on features set III with $\hat{\mathcal{G}}$ thresholded at 0.5.

hyper-parameter $e = 0$ in the MRF prior, which degenerates to a Bernoulli prior, named as SSUR-Ber. We choose SSUR-Ber instead of SSUR-hotspot as the comparison model, since it is easier to use than SSUR-hotspot, which has many tuning hyper-parameters of the hotspot prior, and because SSUR-Ber has shown similar model performance as SSUR-hotspot shown in Section 3.

4.3 Results and discussion

Figure 7b shows an estimated residual structure between the seven drugs by our SSUR-MRF model based on Feature set III with the most genomic information. It does not only estimate residual correlation between any two MAPK inhibitors and between the two Bcr–Abl inhibitors, but also separates the chemotherapy drug Methotrexate from the other drugs. Supplementary S9 shows the residual structures between the seven drugs as estimated by the SSUR-MRF and SSUR-Ber models with feature sets I–III, respectively. We find that the structure estimated by our SSUR-MRF model based on feature set III is closest to our knowledge about the relationships between the seven drugs.

To look at variable selection, a gene feature is considered to be identified if the estimated marginal selection probability of its coefficient is larger than 0.5, i.e. if the corresponding latent indicator variable has posterior mean larger than 0.5. To measure the uncertainty of variable selection $\boldsymbol{B}_{\text{MPM}}$ (i.e. stability selection), we randomly selected 90% of the 499 cell lines for fitting a model 10 times. Table 4 reports the mean and standard deviation of the numbers of identified features over the seven drugs by the SSUR-Ber and SSUR-MRF models. SSUR-Ber results in very sparse models and identifies a similar number of genomic features for each drug. In contrast, our SSUR-MRF model identifies more genomic features and finds a different model sparsity for the three drug groups, in particular relatively denser models for the four MAPK inhibitors. This indicates that our model is able to distinguish variable selection corresponding to different response variables. For the group with the two BCR–ABL inhibitors, i.e. Nilotinib and Axitinib, both SSUR-Ber and SSUR-MRF identify the mutation BCR–ABL associated with drug Nilotinib, as expected.

For the group of the four MAPK inhibitors, Figure 8 displays the numbers of identified features appeared at least 2 out of 10 repetitions by SSUR-Ber and SSUR-MRF. For Feature sets I, II, and III, SSUR-Ber identifies quite different features (Figure 8a), i.e. there is not much overlap. However, our SSUR-MRF model identifies 35 common features over the three feature sets. This reflects more stable variable selection due to using prior knowledge via the MRF prior. Table 5 further shows that the SSUR-Ber model identifies in average <1 target feature for the MAPK inhibitors. Supplementary Table S10.1 shows the identified feature names for the MAPK inhibitors by SSUR-Ber. Overall, our SSUR-MRF model is able to identify many more features than SSUR-Ber, and identifies more known target features for the MAPK inhibitors.

Figure 9 shows the names of features that were identified for the MAPK inhibitors by the SSUR-MRF model. The seven copy number variation and mutation features in Figure 9a are

Table 4. GDSC data application: Number of identified genomic features (mean/standard deviation of 10 repetitions) corresponding to each drug by the SSUR-Ber and SSUR-MRF models

	Nilotinib	Axitinib	RDEA119	PD-0325901	CI-1040	AZD6244	Methotrexate
SSUR-Ber							
Feature set I	3.1 (1.52)	2.9 (1.52)	1.9 (1.10)	1.3 (0.95)	4.1 (0.99)	3.2 (1.03)	2.3 (1.57)
Feature set II	1.5 (0.97)	2.0 (1.05)	1.7 (1.16)	1.0 (0.67)	1.4 (0.97)	2.0 (1.83)	0.6 (0.52)
Feature set III	4.6 (3.41)	4.5 (3.89)	5.0 (2.98)	2.3 (2.36)	5.2 (4.26)	4.3 (3.09)	5.3 (4.79)
SSUR-MRF							
Feature set I	3.5 (0.53)	2.0 (0.82)	42.3 (1.42)	42.0 (1.15)	40.4 (0.52)	40.6 (0.70)	1.9 (1.29)
Feature set II	8.9 (3.35)	8.7 (2.83)	50.5 (17.41)	50.4 (17.26)	50.1 (17.02)	50.8 (17.86)	8.4 (3.41)
Feature set III	35.3 (1.89)	34.9 (2.28)	82.4 (1.65)	81.8 (1.48)	82.1 (1.66)	83.2 (2.04)	35.1 (2.42)

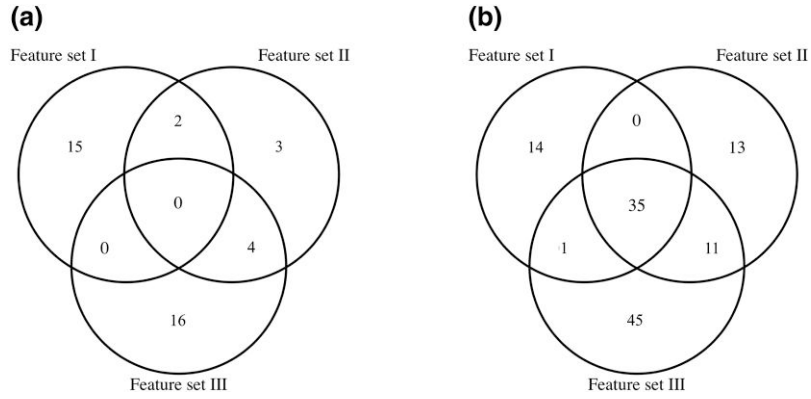

Figure 8. GDSC data application: A Venn diagram for the numbers of identified features appeared at least 2 out of 10 repetitions for the MAPK inhibitors by SSUR-Ber (a) and SSUR-MRF (b) models and overlaps between the models fitted with feature sets I, II, and III.

Table 5. GDSC data application: Numbers of genomic features (mean/standard deviation of 10 repetitions) selected as predictors for the MAPK inhibitors in the SSUR-Ber and SSUR-MRF models

	Known targets	SSUR-Ber	SSUR-MRF
Feature set I	40	0.9 (0.88)	7.4 (0.52)
Feature set II	55	0.1 (0.32)	7.3 (2.58)
Feature set III	81	0.5 (0.85)	16.2 (0.42)

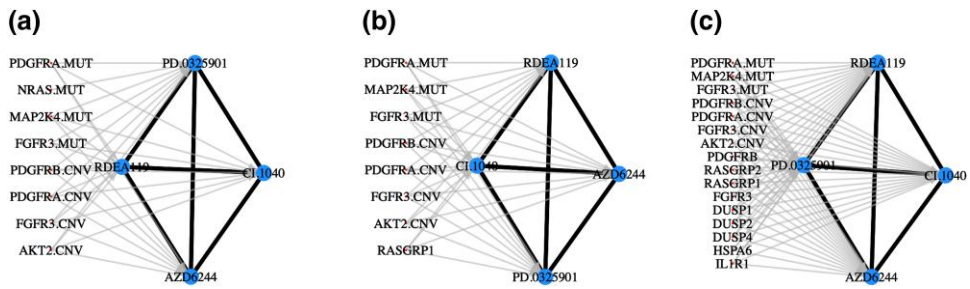

Figure 9. GDSC data application: Estimated network between the MAPK inhibitors and identified target genes based on at least 2 out of 10 repetitions with $\hat{\mathcal{G}}$ and $\hat{\mathbf{\Gamma}}$ thresholded at 0.5 by SSUR-MRF corresponding to Feature set I (a), Feature set II (b), and Feature set III (c), respectively.

also in Figure 9b, c, because only more gene expression features are selected by the models using Feature sets II and III, but no additional mutation or copy number variation features. As more target gene expression features are used to construct the edge potentials in the MRF prior in the models built with Feature sets II and III, our approach can identify more of them. As Figure 8b shows, we have identified 35 common features with the SSUR-MRF model over Feature sets I, II, and III, but only seven of these common features belong to known target genes of the corresponding drugs as shown in Figure 9. We find that the 28 other common identified features (listed in Supplementary Table S10.2) are cancer genes, i.e. genes that are known to be deregulated in cancer. The Cancer Gene Census summarises how dysfunction of these genes drives cancer (Sondka et al., 2018).

Table 6. GDSC data application: Prediction performance (mean/standard deviation of 10 repetitions) of the SSUR-Ber and SSUR-MRF models based on Feature set III

	SSUR-Ber	SSUR-MRF	*p*-value
elpd.LOO	−7,357.5 (59.19)	−7,393.3 (27.31)	0.0488
elpd.WAIC	−7,366.2 (50.04)	−7,390.3 (19.58)	0.0488
RMSE	2.101 (0.1546)	1.847 (0.0631)	0.0020
RMSPE	2.062 (0.1123)	2.121 (0.0731)	0.1309

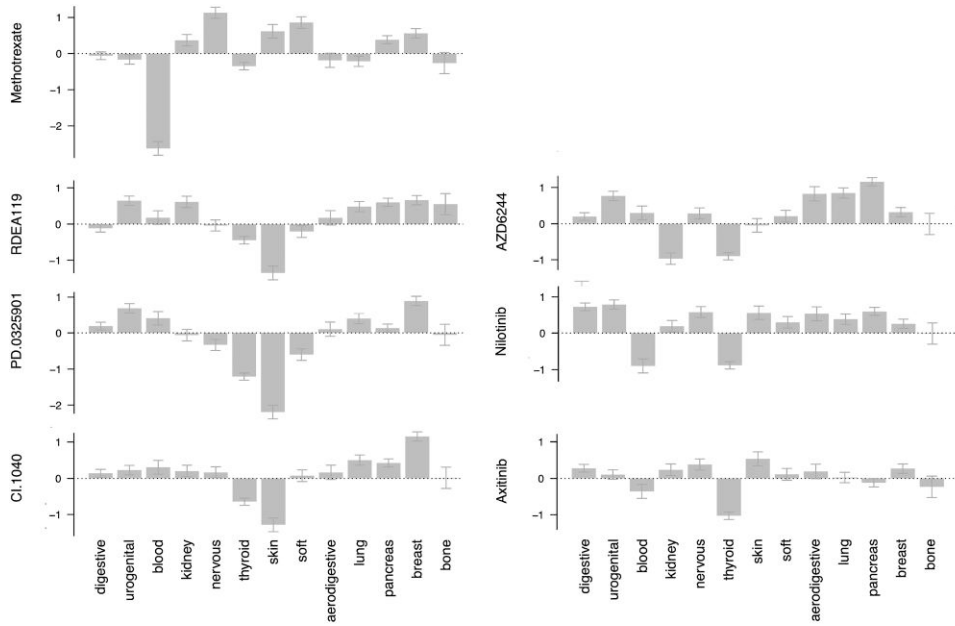

Figure 10. GDSC data application: Posterior estimates (mean and standard deviation of 10 repetitions) for the cancer tissue random effects for all drugs based on the median probability model. Random effects are centred around zero. Error bars are ± standard deviation of the posterior mean over 10 repetitions.

Pathway enrichment analysis (Reimand et al., 2019) maps the 35 common features to four KEGG categories related to cancer in general (in Supplementary Figure S11a), such as central metabolism in cancer, microRNAs in cancer, pathways in cancer and choline metabolism in cancer, and related to the biological functions in some specific cancers (e.g. prostate, renal cell carcinoma, glioma, melanoma, leukaemia), and also directly related to the MAPK inhibitors' targets. Interestingly, the phospholipase D (PLD) signalling pathway can be a potential therapeutic target against cancer (Hwang et al., 2022). The PLD signalling pathway was enriched through our identified genes AKT2, PDGFRA, PDGFRB, PIK3CA, and TSC1. While the first three genes AKT2, PDGFRA, and PDGFRB are targeted by the MAPK inhibitors, PIK3CA and TSC1 are not target genes. Online Supplementary Material, Figure S11b shows the enriched biological processes based the Gene Ontology (GO) database, which are either connected to the MAPK inhibitors' activities (e.g. ERK1 and ERK2 cascade, response to oestrogen) or some general signalling pathways as expected.

In Table 6, prediction performances of the SSUR-Ber and SSUR-MRF models are reported based on Feature set III which has the most genomic information. Overall, prediction performance is very similar between the two models. As for $\widehat{\text{elpd}}_{\text{loo}}$ or $\widehat{\text{elpd}}_{\text{waic}}$, our SSUR-MRF model is better than SSUR-Ber (both Wilcoxon signed-rank tests p-value <0.05). To assess the prediction performance of the median probability model, we need an independent data set for out-of-sample

prediction to obtain RMSPE. For this purpose, we gathered 46 cell lines with complete pharmacogenomic data from the updated GDSC data set by Smirnov et al. (2016), that were not included in our training data. Table 6 shows that SSUR-MRF has similar RMSPE to SSUR-Ber (Wilcoxon signed-rank test p-value =0.1309) on this independent data set.

Our approach also estimates the tissue-specific effects of 13 cancer types, which may indicate relationships between drug responses and cancer types. Figure 10 shows the estimated random effects by SSUR-MRF using the genomic Feature set III. Negative effect estimates can indicate especially high effectiveness of a drug to kill cancer cells of the corresponding cancer type. We focus on the strongest (negative) effects. Methotrexate has the strongest average effect in blood cancer samples; it is known to be an effective chemotherapeutic agent in leukaemia (Powell et al., 2010). Supplementary S12 shows that Methotrexate has much lower $\log(IC_{50})$ values (i.e. more effectiveness) on cell lines from blood tissue type compared with other tissue types. Three of the four MAPK inhibitors (RDEA119, PD-0325901, CI-1040) have their strongest effect in skin cancer cell lines. Supplementary S12 also shows that these drugs have lower $\log(IC_{50})$ values on skin tissue cell lines compared with other tissue types, while AZD6244 shows more variation. Nilotinib and Axitinib are common targeted therapies for chronic myelogenous leukaemia with a BCR–ABL mutation (Halbach et al., 2016). We can observe the strongest effect of Nilotinib on blood cancer samples in Figure 10, and the quite low $\log(IC_{50})$ on the only four BCR–ABL mutated blood cancer cell lines are shown in Supplementary S12.

5 Conclusion

In this work, we have developed a multivariate Bayesian structured variable selection model for analysing data from pharmacogenomic studies. Our model exploits the relationships between multiple correlated response variables (drug sensitivity measurements) and high-dimensional structured multi-omics input data for variable selection and to improve prediction. With our approach we want to (a) be able to borrow known information between response variables and predictors, (b) learn associations between response variables and predictors, and (c) understand the residual covariance of response variables. The proposed approach allows us to make use of known network information on the relationships between responses and predictors in an MRF prior for the variable selection indicator $\boldsymbol{\Gamma}$, and to further simultaneously select predictors in a sparse manner and learn the residual covariance matrix between the response variables. In addition, we can take into account sample heterogeneity through random effects which are excluded from the variable selection. Guidance for specifying (weakly) informative hyper-parameters has been provided in the Supplement.

Through the simulation studies, we have demonstrated that the proposed approach can recover the network structure (i.e. latent indicator variable $\boldsymbol{\Gamma}$) between multiple response variables and predictors, and predict responses well. We have found that including only a small amount of prior knowledge for most patterns/network groups (Figure 4d) will improve model performance over a model that does not any include prior knowledge. Our approach is also robust to noise in the prior information (i.e. false edge potentials) in the MRF prior (Table 2). Even if there is no prior association knowledge between drugs and genes/pathways (i.e. subgraphs), our approach has similar model performance as SSUR-hotspot.

In the pharmacogenomic data application, our approach robustly identified molecular targets of the targeted therapies, and also validated other known cancer-related genes. The use of known information in the MRF prior improved the prediction performance in the independent validation data compared to SSUR-Ber when applied to the largest input data set (Feature set III). Through the random effects in our approach, cancer tissue effects were estimated, which could indicate potential relationships between drugs and cancer types. Nevertheless, there was still remaining heterogeneity within cancer types, e.g. reflecting molecular cancer sub-types. To address this, our model could be extended to multilevel random effects or a mixture approach could be employed for the random effects, e.g. by a flexible Dirichlet process prior (Heinzl et al., 2012; Li et al., 2010).

Although our approach has been successfully applied in scenarios with multiple correlated response variables and high-dimensional predictors, it might become too computationally demanding if the model is not assumed to be very sparse (i.e. if the number of associated features is not

assumed to be much smaller than mp). An alternative is to change our MCMC sampling approach to approximate inference, e.g. variational inference (Blei et al., 2017; Münch et al., 2021; Zhang et al., 2019). Also, our SSUR-MRF model assumes linearity and normality. Alexopoulos and Bottolo (2020) proposed sparse Gaussian copula regression (non-linear) to explore the high-dimensional model space and estimate the structure among multiple responses of diverse types (i.e. Gaussian, low-intensity counts, binary, ordinal, and continuous variables). For our SSUR-MRF model, if the residuals have heavy tails (e.g. t-distributed), online Supplementary Material, Table S6.2 shows good variable selection and slightly worse prediction performance, but large sample size can slightly improve the prediction performance. Although our SSUR-MRF model used random effects to allow heterogeneity between groups of samples, it cannot model sample-specific random effects for individual observations (i.e. random intercepts). Zhao (2020) discussed random intercepts and random slopes in a general setting of high-dimensional SUR models.

Acknowledgments

We would like to thank the editor, the associate editor, and two anonymous referees for their constructive suggestions that improved the quality of the manuscript. We also would like to thank Leonardo Bottolo, Jorrit Enserink, Aram Andersen, and Shixiong Wang for discussions.

Conflict of interests: None declared.

Funding

This work was supported by Research Council of Norway project No. 237718 'Big Insight' (ZZ), European Union Horizon 2020 grant agreements No. 847912 'RESCUER' (MZ) and No. 633595 'DynaHealth' (AL), UK Medical Research Council grants MR/M013138/1 (MB, AL).

Data availability

The Genomics of Drug Sensitivity in Cancer data are publicly available at https://www.cancerrxgene.org.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series C*.

References

Adam G., Rampášek L., Safikhani Z., Smirnov P., Haibe-Kains B., & Goldenberg A. (2020). Machine learning approaches to drug response prediction: Challenges and recent progress. *NPJ Precision Oncology*, *4*(1), 19. https://doi.org/10.1038/s41698-020-0122-1

Alexopoulos A., & Bottolo L. (2020). Bayesian variable selection for Gaussian copula regression models. *Journal of Computational and Graphical Statistics*, *30*(3), 578–593. https://doi.org/10.1080/10618600.2020.1840997

Bai R., Moran G. E., Antonelli J. L., Chen Y., & Boland M. R. (2022). Spike-and-slab group lassos for grouped regression and sparse generalized additive models. *Journal of the American Statistical Association*, *117*(537), 184–197. https://doi.org/10.1080/01621459.2020.1765784

Ballester P. J., Stevens R., Haibe-Kains B., Huang R. S., & Aittokallio T. (2022). Artificial intelligence for drug response prediction in disease models. *Briefings in Bioinformatics*, *23*(1), bbab450. https://doi.org/10.1093/bib/bbab450

Barbieri M. M., & Berger J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, *32*(3), 870–897. https://doi.org/10.1214/009053604000000238

Barbieri M. M., Berger J. O., George E. I., & Ročková V. (2021). The median probability model and correlated variables. *Bayesian Analysis*, *16*(4), 1085–1112. https://doi.org/10.1214/20-BA1249

Barretina J., Caponigro G., Stransky N., Venkatesan K., Margolin A. A., Kim S., Wilson C. J., Lehár J., Kryukov G. V., Sonkin D., Reddy A., Liu M., Murray L., Berger M. F., Monahan J. E., Morais P., Meltzer J., Korejwa A., Jané-Valbuena J., Mapa F. A., Thibault J., Bric-Furlong E., Raman P., Shipway A., Engels I. H., Cheng J., Yu G. K., Yu J., Aspesi P., de Silva M., Jagtap K., Jones M. D., Wang L., Hatton C., Palescandolo E., Gupta S., Mahan S., Sougnez C., Onofrio R. C., Liefeld T., MacConaill L., Winckler W., Reich M., Li N., Mesirov J. P., Gabriel S. B., Getz G., Ardlie K., Chan V., Myer V. E., Weber B. L., Porter J., Warmuth M., Finan P., Harris

J. L., Meyerson M., Golub T. R., Morrissey M. P., Sellers W. R., Schlegel R., & Garraway L. A. (2012). The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, *483*(7391), 603–607. https://doi.org/10.1038/nature11003

Bhadra A., & Mallick B. (2013). Joint high-dimensional Bayesian variable and covariance selection with an application to eQTL analysis. *Biometrics*, *69*(2), 447–457. https://doi.org/10.1111/biom.v69.2

Blei D. M., Kucukelbir A., & McAuliffe J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Bottolo L., Banterle M., Richardson S., Ala-Korpela M., Järvelin M.-R., & 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. https://doi.org/10.1111/rssc.12490

Bottolo L., Petretto E., Blankenberg S., Cambien F., Cook S. A., Tiret L., & Richardson S. (2011). Bayesian detection of expression quantitative trait loci hot-spots. *Genetics*, *189*(4), 1449–1459. https://doi.org/10.1534/genetics.111.131425

Bottolo L., & Richardson S. (2010). Evolutionary stochastic search for Bayesian model exploration. *Bayesian Analysis*, *5*(3), 583–618. https://doi.org/10.1214/10-BA523

Brown P., Vannucci M., & Fearn T. (1998). Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *60*(3), 627–641. https://doi.org/10.1111/1467-9868.00144

Carvalho C. M., Massam H., & West M. (2007). Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika*, *94*(3), 647–659. https://doi.org/10.1093/biomet/asm056

Chakraborty M., Baladandayuthapani V., Bhadra A., & Ha M. J. (2021). 'Bayesian robust learning in chain graph models for integrative pharmacogenomics', arXiv, arXiv:10.48550/ARXIV.2111.11529.

Chekouo T., Stingo F. C., Doecke J. D., & Do K.-A. (2015). miRNA-target gene regulatory networks: A Bayesian integrative approach to biomarker selection with application to kidney cancer. *Biometrics*, *71*(2), 428–438. https://doi.org/10.1111/biom.v71.2

Chekouo T., Stingo F. C., Doecke J. D., & Do K.-A. (2017). A Bayesian integrative approach for multi-platform genomic data: A kidney cancer case study. *Biometrics*, *73*(2), 615–624. https://doi.org/10.1111/biom.v73.2

Chekouo T., Stingo F. C., Guindani M., & Do K.-A. (2016). A Bayesian predictive model for imaging genetics with application to schizophrenia. *The Annals of Applied Statistics*, *10*(3), 1547–1571. https://doi.org/10.1214/16-AOAS948

Costello J. C., Heiser L. M., Georgii E., Gönen M., Menden M. P., Wang N. J., Bansal M., Khan S. A., Mpindi J. -P., Kallioniemi O., Honkela A., Aittokallio T., Wennerberg K., Collins J. J., Gallahan D., Singer D., Saez-Rodriguez J., Kaski S., Gray J. W., & Stolovitzky G. (2014). A community effort to assess and improve drug sensitivity prediction algorithms. *Nature Biotechnology*, *32*(12), 1202–1212. https://doi.org/10.1038/nbt.2877

Feng F., Shen B., Mou X., Li Y., & Li H. (2021). Large-scale pharmacogenomic studies and drug response prediction for personalized cancer medicine. *Journal of Genetics and Genomics*, *48*(7), 540–551. https://doi.org/10.1016/j.jgg.2021.03.007

Fitch A., Jones M., & Massam H. (2014). The performance of covariance selection methods that consider decomposable models only. *Bayesian Analysis*, *9*(3), 659–684. https://doi.org/10.1214/14-BA874

Fontes Jardim D. L., Schwaederle M., Wei C., Lee J. J., Hong D. S., Eggermont A. M., Schilsky R. L., Mendelsohn J., Lazar V., & Kurzrock R. (2015). Impact of a biomarker-based strategy on oncology drug development: A meta-analysis of clinical trials leading to FDA approval. *JNCI: Journal of the National Cancer Institute*, *107*(11), djv253. https://doi.org/10.1093/jnci/djv253

Garnett M. J., Edelman E. J., Heidorn S. J., Greenman C. D., Dastur A., Lau K. W., Greninger P., Thompson I. R., Luo X., Soares J., Liu Q., Iorio F., Surdez D., Chen L., Milano R. J., Bignell G. R., Tam A. T., Davies H., Stevenson J. A., Barthorpe S., Lutz S. R., Kogera F., Lawrence K., McLaren-Douglas A., Mitropoulos X., Mironenko T., Thi H., Richardson L., Zhou W., Jewitt F., Zhang T., O'Brien P., Boisvert J. L., Price S., Hur W., Yang W., Deng X., Butler A., Choi H. G., Chang J. W., Baselga J., Stamenkovic I., Engelman J. A., Sharma S. V., Delattre O., Saez-Rodriguez J., Gray N. S., Settleman J., Futreal P. A., Haber D. A., Stratton M. R., Ramaswamy S., McDermott U., & Benes C. H. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, *483*(7391), 570–575. https://doi.org/10.1038/nature11005

George E. I., & McCulloch R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*(423), 881–889. https://doi.org/10.1080/01621459.1993.10476353

Green P. J., & Thomas A. (2013). Sampling decomposable graphs using a Markov chain on junction trees. *Biometrika*, *100*(1), 91–110. https://doi.org/10.1093/biomet/ass052

Ha M. J., Stingo F. C., & Baladandayuthapani V. (2021). Bayesian structure learning in multilayered genomic networks. *Journal of the American Statistical Association*, *116*(534), 605–618. https://doi.org/10.1080/01621459.2020.1775611

Halbach S., Hu Z., Gretzmeier C., Ellermann J., Wöhrle F. U., Dengjel J., & Brummer T. (2016). Axitinib and sorafenib are potent in tyrosine kinase inhibitor resistant chronic myeloid leukemia cells. *Cell Communication and Signaling*, *14*(1), 6. https://doi.org/10.1186/s12964-016-0129-y

Heinzl F., Fahrmeir L., & Kneib T. (2012). Additive mixed models with Dirichlet process mixture and P-spline priors. *Asta Advances in Statistical Analysis*, *96*(1), 47–68. https://doi.org/10.1007/s10182-011-0161-6

Huang E. W., Bhope A., Lim J., Sinha S., & Emad A. (2020). Tissue-guided lasso for prediction of clinical drug response using preclinical samples. *PLoS Computational Biology*, *16*(1), e1007607. https://doi.org/10.1371/journal.pcbi.1007607

Hwang W. C., Song D., Lee H., Oh C., Lim S. H., Bae H. J., Kim N. D., Han G., & Min D. S. (2022). Inhibition of phospholipase D1 induces immunogenic cell death and potentiates cancer immunotherapy in colorectal cancer. *Experimental & Molecular Medicine*, *54*(9), 1563–1576. https://doi.org/10.1038/s12276-022-00853-6

Jia Z., & Xu S. (2007). Mapping quantitative trait loci for expression abundance. *Genetics*, *176*(1), 611–623. https://doi.org/10.1534/genetics.106.065599

Kim E., Dede M., Lenoir W. F., Wang G., Srinivasan S., Colic M., & Hart T. (2019). A network of human functional gene interactions from knockout fitness screens in cancer cells. *Life Science Alliance*, *2*(2), e201800278. https://doi.org/10.26508/lsa.201800278

Lee K. H., Tadesse M. G., Baccarelli A. A., Schwartz J., & Coull B. A. (2017). Multivariate Bayesian variable selection exploiting dependence structure among outcomes: Application to air pollution effects on DNA methylation. *Biometrics*, *73*(1), 232–241. https://doi.org/10.1111/biom.v73.1

Le Tourneau C., Delord J.-P., Gonçalves A., Gavoille C., Dubot C., Isambert N., Campone M., Trédan O., Massiani M.-A., Mauborgne C., Armanet S., Servant N., Bièche I., Bernard V., Gentien D., Jezequel P., Attignon V., Boyault S., Vincent-Salomon A., ...for the SHIVA Investigators. (2015). Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): A multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *The Lancet Oncology*, *16*(13), 1324–1334. https://doi.org/10.1016/S1470-2045(15)00188-6

Lewin A., Saadi H., Peters J. E., Moreno-Moral A., Lee J. C., Smith K. G. C., Petretto E., Bottolo L., & Richardson S. (2016). MT-HESS: An efficient Bayesian approach for simultaneous association detection in omics datasets, with application to eQTL mapping in multiple tissues. *Bioinformatics*, *32*(4), 523–532. https://doi.org/10.1093/bioinformatics/btv568

Li Y., Lin X., & Müller P. (2010). Bayesian inference in semiparametric mixed models for longitudinal data. *Biometrics*, *66*(1), 70–78. https://doi.org/10.1111/biom.2010.66.issue-1

Liang F., & Wong W. H. (2000). *Statistica Sinica* evolutionary Monte Carlo: Application to c_p model sampling and change point problem. https://www3.stat.sinica.edu.tw/statistica/J10n2/j10n21/j10n21.htm

Liquet B., Mengersen K., Pettitt A. N., & Sutton M. (2017). Bayesian variable selection regression of multivariate responses for group data. *Bayesian Analysis*, *12*(4), 1039–1067. https://doi.org/10.1214/17-BA1081

Marquart J., Chen E. Y., & Prasad V. (2018). Estimation of the percentage of US patients with cancer who benefit from genome-driven oncology. *JAMA Oncology*, *4*(8), 1093–1098. https://doi.org/10.1001/jamaoncol.2018.1660

Mohammadi R., & Wit E. (2019). BDgraph: An R package for Bayesian structure learning in graphical models. *Journal of Statistical Software*, *89*(3), 1–30. https://doi.org/10.18637/jss.v089.i03

Münch M. M., van de Wiel M. A., Richardson S., & Leday G. G. R. (2021). Drug sensitivity prediction with normal inverse Gaussian shrinkage informed by external data. *Biometrical Journal*, *63*(2), 289–304. http://dx.doi.org/10.1002/bimj.201900371

Petretto E., Bottolo L., Langley S. R., Heinig M., McDermott-Roe C., Sarwar R., Pravenec M., Hübner N., Aitman T. J., Cook S. A., & Richardson S. (2010). New insights into the genetic control of gene expression using a Bayesian multi-tissue approach. *PLoS Computational Biology*, *6*(4), e1000737. https://doi.org/10.1371/journal.pcbi.1000737

Powell B. L., Moser B., Stock W., Gallagher R. E., Willman C. L., Stone R. M., Rowe J. M., Coutre S., Feusner J. H., Gregory J., Couban S., Appelbaum F. R., Tallman M. S., & Larson R. A. (2010). Arsenic trioxide improves event-free and overall survival for adults with acute promyelocytic leukemia: North American leukemia intergroup study C9710. *Blood*, *116*(19), 3751–3757. https://doi.org/10.1182/blood-2010-02-269621

Reimand J., Isserlin R., Voisin V., Kucera M., Tannus-Lopes C., Rostamianfar A., Wadi L., Meyer M., Wong J., Xu C., Merico D., & Bader G. D. (2019). Pathway enrichment analysis and visualization of omics data using g: Profiler, GSEA, Cytoscape and Enrichmentmap. *Nature Protocols*, *14*(2), 482–517. https://doi.org/10.1038/s41596-018-0103-9

Richardson S., Bottolo L., Rosenthal J. S., Bernardo J. M., Bayarri M. J., Berger J. O., Dawid A. P., Heckerman D., Smith A. F. M., & West M. (2011). Bayesian models for sparse regression analysis of high dimensional data. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, & M. West (Eds.), *Bayesian statistics* (Vol. *9*, pp. 539–568). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199694587.003.0018

Russo D. J., Roy B. V., Kazerouni A., Osband I., & Wen Z. (2018). A tutorial on Thompson sampling. *Foundations and Trends in Machine Learning*, *11*(1), 1–96. https://doi.org/10.1561/2200000070

Sharifi-Noghabi H., Jahangiri-Tazehkand S., Smirnov P., Hon C., Mammoliti A., Nair S. K., Mer A. S., Ester M., & Haibe-Kains B. (2021). Drug sensitivity prediction from cell line-based pharmacogenomics data: Guidelines for developing machine learning models. *Briefings in Bioinformatics*, *22*(6), bbab294. https://doi.org/10.1093/bib/bbab294

Smirnov P., Safikhani Z., El-Hachem N., Wang D., She A., Olsen C., Freeman M., Selby H., Gendoo D. M., Grossmann P., Beck A. H., Aerts H. J., Lupien M., Goldenberg A., & Haibe-Kains B. (2016). PharmacoGx: An R package for analysis of large pharmacogenomic datasets. *Bioinformatics*, *32*(8), 1244–1246. https://doi.org/10.1093/bioinformatics/btv723

Sondka Z., Bamford S., Cole C. G., Ward S. A., Dunham I., & Forbes S. A. (2018). The COSMIC Cancer Gene Census: Describing genetic dysfunction across all human cancers. *Nature Review Cancer*, *18*(11), 696–705. https://doi.org/10.1038/s41568-018-0060-1

Stingo F. C., Chen Y. A., Tadesse M. G., & Vannucci M. (2011). Incorporating biological information into linear models: A Bayesian approach to the selection of pathways and genes. *The Annals of Applied Statistics*, *5*(3), 1978–2002. https://doi.org/10.1214/11-AOAS463

Vehtari A., Gelman A., & Gabry J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Von Hoff D. D., Stephenson J. J., Rosen P., Loesch D. M., Borad M. J., Anthony S., Jameson G., Brown S., Cantafio N., Richards D. A., Fitch T. R., Wasserman E., Fernandez C., Green S., Sutherland W., Bittner M., Alarcon A., Mallery D., & Penny R. (2010). Pilot study using molecular profiling of patients' tumors to find potential targets and select treatments for their refractory cancers. *Journal of Clinical Oncology*, *28*(33), 4877–4883. https://doi.org/10.1200/JCO.2009.26.5983

Wang H. (2010). Sparse seemingly unrelated regression modelling: Applications in finance and econometrics. *Computational Statistics & Data Analysis*, *54*(11), 2866–2877. https://doi.org/10.1016/j.csda.2010.03.028

Yang W., Soares J., Greninger P., Edelman E. J., Lightfoot H., Forbes S., Bindal N., Beare D., Smith J. A., & Thompson I. R. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Reserch*, *41*(Database issue), D955–D961. https://doi.org/10.1093/nar/gks1111

Yang X., & Narisetty N. N. (2020). Consistent group selection with Bayesian high dimensional modeling. *Bayesian Analysis*, *15*(3), 909–935. https://doi.org/10.1214/19-BA1178

Zellner A., & Ando T. (2010). A direct Monte Carlo approach for Bayesian analysis of the seemingly unrelated regression model. *Journal of Econometrics*, *159*(1), 33–45. https://doi.org/10.1016/j.jeconom.2010.04.005

Zhang C., Butepage J., Kjellstrom H., & Mandt S. (2019). Advances in variational inference. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *41*(8), 2008–2026. https://doi.org/10.1109/TPAMI.34

Zhao Z. (2020). *Multivariate structured penalized and Bayesian regressions for pharmacogenomic screens* [PhD dissertation]. University of Oslo, Oslo.

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. https://doi.org/10.18637/jss.v100.i11

Zhao Z., & Zucknick M. (2020). Structured penalized regression for drug sensitivity prediction. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *69*(3), 525–545. https://doi.org/10.1111/rssc.12400