

Genome analysis

The spike-and-slab lasso Cox model for survival prediction and associated genes detection

Zaixiang Tang^{1,2,3,†}, Yueping Shen^{1,2,†}, Xinyan Zhang³ and Nengjun Yi^{3,*}

¹Department of Biostatistics, School of Public Health and ²Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, and Center for Genetic Epidemiology and Genomics, Medical College of Soochow University, Suzhou 215123, China and ³Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294, USA

*To whom correspondence should be addressed.

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Associate Editor: John Hancock

Received on February 2, 2017; revised on April 6, 2017; editorial decision on April 28, 2017; accepted on May 5, 2017

Abstract

Motivation: Large-scale molecular profiling data have offered extraordinary opportunities to improve survival prediction of cancers and other diseases and to detect disease associated genes. However, there are considerable challenges in analyzing large-scale molecular data.

Results: We propose new Bayesian hierarchical Cox proportional hazards models, called the spike-and-slab lasso Cox, for predicting survival outcomes and detecting associated genes. We also develop an efficient algorithm to fit the proposed models by incorporating Expectation-Maximization steps into the extremely fast cyclic coordinate descent algorithm. The performance of the proposed method is assessed via extensive simulations and compared with the lasso Cox regression. We demonstrate the proposed procedure on two cancer datasets with censored survival outcomes and thousands of molecular features. Our analyses suggest that the proposed procedure can generate powerful prognostic models for predicting cancer survival and can detect associated genes.

Availability and implementation: The methods have been implemented in a freely available R package BhGLM (<http://www.ssg.uab.edu/bhglm/>).

Contact: nyi@uab.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Precision medicine needs accurate prognostic prediction (predicting the risk of future relapse after an initial treatment or responsiveness to different treatments) (Barillot *et al.* 2012; Chin *et al.* 2011). Many traditional clinical prognostic and predictive factors have been known for diseases (e.g. cancers) for years, however, usually provide poor prognosis and prediction (Barillot *et al.* 2012). There is therefore a need for new prognostic and predictive factors with better reproducibility and better discriminatory power between different prognosis and drug responsiveness groups. Precision medicine is often based on the genetic background of the patients (Barillot *et al.* 2012; Collins and Varmus 2015). Recent high-throughput technologies can easily and robustly generate large-scale molecular

profiling data, offering extraordinary opportunities to search for new biomarkers and to build accurate prognostic and predictive models (Barillot *et al.* 2012; Collins and Varmus 2015). However, significant challenges exist, including: (i) how to select predictive molecular factors among numerous candidates, (ii) how to precisely estimate the effects of predictors and (iii) how to combine many predictive factors into accurate predictive models.

Several methods have been applied to address the challenges, mainly including traditional variable selection, principal components method and penalized regressions. Extensive studies have demonstrated that penalized regressions usually have good predictive performance and can result in powerful and interpretable predictive models (Bovelstad *et al.* 2007, 2009). Various penalized

survival models have recently been proposed to analyze large-scale data, including lasso, ridge or elastic-net Cox proportional hazards models (Hastie *et al.* 2015; Simon *et al.* 2011; Tibshirani 1997; van Houwelingen *et al.* 2006) and parametric models (Li *et al.* 2016; Mittal *et al.* 2013). The penalized survival models can be fit by fast algorithms, and can achieve both predicting survival and identifying important predictors (Hastie *et al.* 2015; Simon *et al.* 2011; Tibshirani 1997). Thus these penalized methods have widely been used for analyzing high-dimensional molecular data (Barillot *et al.* 2012; Gerstung *et al.* 2015; Sohn and Sung 2013; Yuan *et al.* 2014; Zhang *et al.* 2013; Zhao *et al.* 2015). Recently, Bayesian survival models and spike-and-slab variable selection methods have been applied to molecular predictive modeling (Bonato *et al.* 2011; Lee *et al.* 2011, 2015; Li and Zhang 2010; Monni and Li 2010; Peng *et al.* 2013; Stingo *et al.* 2010; Tai *et al.* 2009). These methods use Markov Chain Monte Carlo (MCMC) algorithms to fully explore the joint posterior distribution. Although statistically sophisticated, these MCMC-based methods are computationally intensive for analyzing large-scale molecular data.

In this article, we propose new hierarchical Cox survival models with spike-and-slab double-exponential priors for jointly analyzing large-scale molecular data for building powerful predictive models and identifying important predictors. The spike-and-slab prior is the fundamental basis for most Bayesian variable selection approaches, and has proved remarkably successful (Chipman 1996; Chipman *et al.* 2001; George and McCulloch 1993, 1997; Ročková and George, 2014; Ročková and George, 2016, unpublished manuscript). Most previous spike-and-slab variable selection approaches use the spike-and-slab normal priors on coefficients and employ computationally intensive MCMC algorithms to search for high posterior models. Recently, Ročková and George (2016, unpublished manuscript) proposed a new framework, called the spike-and-slab lasso, for high-dimensional normal linear models, and showed that it has remarkable properties (Ročková and George, 2016, unpublished manuscript).

We extend the spike-and-slab lasso framework to Cox survival models. The proposed spike-and-slab lasso Cox models can adaptively shrink coefficients and thus can result in accurate estimates and predictions. We propose an efficient algorithm to fit the spike-and-slab lasso Cox models by incorporating EM steps (Expectation-Maximization) into the extremely fast cyclic coordinate descent algorithm. The proposed approach integrates two popular methods, i.e. lasso and Bayesian spike-and-slab hierarchical modeling, into one unifying framework, and thus combines the remarkable features of the two popular methods while diminishing their shortcomings. The performance of the proposed method is assessed via extensive simulations and compared with the commonly used lasso Cox regression. We apply the proposed procedure to two cancer datasets with censored survival outcomes and thousands of molecular features: breast cancer and myelodysplastic syndromes (MDS). Our results show that the proposed method can generate powerful prognostic models for predicting cancer survival and can detect survival associated genes.

2 Materials and methods

2.1 Cox proportional hazards models

For censored survival outcomes, we observe a pair of response variables $y_i = (t_i, d_i)$ for each individual, where the censoring indicator d_i takes 1 if the observed survival time t_i for individual i is uncensored and 0 if it is censored. Denote the true survival time by T_i for

individual i . Thus, when $d_i = 1$, $T_i = t_i$, whereas when $d_i = 0$, $T_i > t_i$. Cox proportional hazards model is the most widely used method for studying the relationship between the censored survival response and some explanatory variables X , and assumes that the hazard function of survival time T takes the form (Ibrahim *et al.* 2001; Klein and Moeschberger 2003):

$$h(t|X) = h_0(t) \exp(X\beta). \quad (1)$$

where the baseline hazard function $h_0(t)$ is unspecified, X and β are the vectors of explanatory variables and coefficients, respectively, and $X\beta$ is the linear predictor or called the prognostic index.

Fitting classical Cox models is to estimate β by maximizing the partial log-likelihood (Cox 1972):

$$pl(\beta) = \sum_{i=1}^n d_i \log \left(\frac{\exp(X_i\beta)}{\sum_{r' \in R(t_i)} \exp(X_{r'}\beta)} \right), \quad (2)$$

where $R(t_i)$ is the risk set at time t_i . In the presence of ties, the partial log-likelihood can be approximated by the Breslow or the Efron methods (Breslow 1972; Efron 1977). The standard algorithm for maximizing the partial log-likelihood is the Newton–Raphson algorithm (Klein and Moeschberger 2003; van Houwelingen and Putter 2012).

The lasso uses the L_1 penalty to the partial log-likelihood function and estimates the parameters β by maximizing the penalized log-likelihood (Friedman *et al.* 2010; Hastie *et al.* 2015; Simon *et al.* 2011; Tibshirani 1997; Zou and Hastie 2005):

$$ppl(\beta) = pl(\beta) - \lambda \sum_{j=1}^J |\beta_j|. \quad (3)$$

The overall penalty parameter λ controls the overall strength of penalty and the size of the all coefficients. The lasso Cox model can be fit by the extremely fast cyclic coordinate descent algorithm (Simon *et al.*, 2011).

Ideally, one should use small penalty values for important predictors and large penalties for irrelevant predictors. We here propose a new approach, i.e., the spike-and-slab lasso, which can induce different shrinkage scales for different coefficients and allows us to estimate the shrinkage scales from the data.

2.2 Spike-and-slab lasso Cox models

The spike-and-slab lasso Cox models are more easily interpreted and handled from Bayesian hierarchical modeling framework (Ročková and George, unpublished manuscript; Ročková and George, 2016). It is well known that the lasso can be expressed as a hierarchical model with double-exponential prior on coefficients (Kyung, *et al.*, 2010; Park and Casella, 2008; Tibshirani, 1996; Yi and Xu, 2008):

$$\beta_j|s \sim \text{DE}(\beta_j|0, s) = \frac{1}{2s} \exp\left(-\frac{|\beta_j|}{s}\right), \quad (4)$$

where the scale parameter s equals λ^{-1} . The scale s controls the amount of shrinkage; smaller scale induces stronger shrinkage and forces the estimates of β_j towards 0.

We develop the spike-and-slab lasso Cox models by extending the double-exponential prior to the following spike-and-slab mixture double-exponential prior:

$$\beta_j|\gamma_j, s_0, s_1 \sim (1 - \gamma_j) \text{DE}(\beta_j|0, s_0) + \gamma_j \text{DE}(\beta_j|0, s_1)$$

or equivalently

$$\beta_j|\gamma_j, s_0, s_1 \sim \text{DE}(\beta_j|0, S_j) = \frac{1}{2S_j} \exp\left(-\frac{|\beta_j|}{S_j}\right). \quad (5)$$

where γ_j is the indicator variable, $\gamma_j = 1$ or 0 , and the scale S_j equals one of two preset positive values s_0 and s_1 ($s_1 > s_0 > 0$), i.e. $S_j = (1 - \gamma_j)s_0 + \gamma_j s_1$. The scale value s_0 is chosen to be small and serves a ‘‘spike scale’’ for modeling irrelevant (0) coefficients and inducing strong shrinkage on estimation, and s_1 is set to be relatively large and thus serves as a ‘‘slab scale’’ for modeling large coefficients and inducing no or weak shrinkage on estimation. If we set $s_0 = s_1$, the spike-and-slab double-exponential prior becomes the double-exponential prior. Therefore, the spike-and-slab lasso includes the lasso as a special case.

The indicator variables γ_j play an essential role on linking the scale parameters with the coefficients. For our development, we confine attention to exchangeable priors on the indicator variables, that is, the indicator variables are assumed to follow the independent binomial distribution:

$$\gamma_j|\theta \sim \text{Bin}(\gamma_j|1, \theta) = \theta^{\gamma_j} (1 - \theta)^{1 - \gamma_j}, \quad (6)$$

where θ is the probability parameter. However, the scope of the spike-and-slab mixture priors is greatly enhanced by the flexibility of the model space prior for the indicator variables γ_j , which can be used to concentrate on preferred and meaningful models.

For the probability parameter θ , we assume the uniform prior: $\theta \sim U(0, 1)$. The probability parameter θ can be viewed as the overall shrinkage parameter that equals the prior probability $p(\gamma_j = 1|\theta)$. The prior expectation of the scale S_j equals $E(S_j) = (1 - \theta)s_0 + \theta s_1$, which lies in the range $[s_0, s_1]$. As will be seen, the scale S_j for each coefficient can be estimated, leading to different shrinkage for different predictors.

2.3 Algorithm for fitting the spike-and-slab lasso Cox models

We here develop a fast algorithm to fit the spike-and-slab lasso Cox models by finding the posterior modes of the parameters, i.e. estimating the parameters by maximizing the posterior density. Our algorithm, called the EM coordinate descent algorithm, incorporates EM steps into the cyclic coordinate descent procedure for fitting the penalized lasso Cox regression. We derive the EM coordinate descent algorithm based on the log joint posterior density of the parameters (β, γ, θ) :

$$\begin{aligned} \log p(\beta, \gamma, \theta|t, d) &= \log p(t, d|\beta, h_0) + \sum_{j=1}^J \log p(\beta_j|S_j) \\ &\quad + \sum_{j=1}^J \log p(\gamma_j|\theta) + \log p(\theta) \end{aligned}$$

The log-likelihood function, $\log p(t, d|\beta, h_0)$, is proportional to the partial log-likelihood $pI(\beta)$ defined in Equation (1) or the Breslow or the Efron approximation in the presence of ties (Breslow 1972; Efron 1977), if the baseline hazard function h_0 is replaced by the Breslow estimator (Breslow 1974; van Houwelingen and Putter 2012). Therefore, the log joint posterior density can be expressed as

$$\begin{aligned} \log p(\beta, \gamma, \theta|t, d) &\propto pI(\beta) - \sum_{j=1}^J \frac{1}{S_j} |\beta_j| \\ &\quad + \sum_{j=1}^J \left(\gamma_j \log \theta + (1 - \gamma_j) \log (1 - \theta) \right), \quad (7) \end{aligned}$$

where $pI(\beta)$ is the partial likelihood described in (2), and $S_j = (1 - \gamma_j)s_0 + \gamma_j s_1$.

The EM coordinate descent algorithm treats the indicator variables γ_j as ‘missing values’ and estimates the parameters (β, θ) by averaging the missing values over their posterior distributions. For the E-step, we calculate the expectation of the log joint posterior density with respect to the conditional posterior distributions of the missing data γ_j . The conditional posterior expectation of the indicator variable γ_j can be derived as

$$\begin{aligned} p_j &= p(\gamma_j = 1|\beta_j, \theta, t, d) \\ &= \frac{p(\beta_j|\gamma_j = 1, s_1) p(\gamma_j = 1|\theta)}{p(\beta_j|\gamma_j = 0, s_0) p(\gamma_j = 0|\theta) + p(\beta_j|\gamma_j = 1, s_1) p(\gamma_j = 1|\theta)}, \quad (8) \end{aligned}$$

where $p(\beta_j|\gamma_j = 1, s_1) = \text{DE}(\beta_j|0, s_1)$, $p(\beta_j|\gamma_j = 0, s_0) = \text{DE}(\beta_j|0, s_0)$, $p(\gamma_j = 1|\theta) = \theta$ and $p(\gamma_j = 0|\theta) = 1 - \theta$. Therefore, the conditional posterior expectation of S_j^{-1} can be obtained by

$$E(S_j^{-1}|\beta_j) = E\left(\frac{1}{(1 - \gamma_j)s_0 + \gamma_j s_1} \mid \beta_j\right) = \frac{1 - p_j}{s_0} + \frac{p_j}{s_1}. \quad (9)$$

It can be seen that the estimates of p_j and S_j are larger for larger coefficients β_j , leading to different shrinkage for different coefficients.

For the M-step, we update (β, θ) by maximizing the posterior expectation of the log joint posterior density with γ_j and S_j^{-1} replaced by their conditional posterior expectations. From the log joint posterior density, we can see that β and θ can be updated separately, because the coefficients β are only involved in $pI(\beta) - \sum_{j=1}^J S_j^{-1} |\beta_j|$ and the probability parameter θ is only in $\sum_{j=1}^J (\gamma_j \log \theta + (1 - \gamma_j) \log (1 - \theta))$. Therefore, the coefficients β are updated by maximizing the expression:

$$Q_1(\beta) = pI(\beta) - \sum_{j=1}^J S_j^{-1} |\beta_j|, \quad (10)$$

where S_j^{-1} is replaced by its conditional posterior expectation derived above. Given the scale parameters S_j , the term $\sum_{j=1}^J S_j^{-1} |\beta_j|$ serves as the L_1 lasso penalty with S_j^{-1} as the penalty factors, and thus the coefficients can be updated by maximizing $Q_1(\beta)$ using the cyclic coordinate descent algorithm. Thus, the coefficients can be estimated to be 0. The probability parameter θ is updated by maximizing the expression:

$$Q_2(\theta) = \sum_{j=1}^J \left(\gamma_j \log \theta + (1 - \gamma_j) \log (1 - \theta) \right) \quad (11)$$

We can easily obtain: $\theta = \frac{1}{J} \sum_{j=1}^J p_j$.

In summary, the EM coordinate descent algorithm for fitting the spike-and-slab lasso Cox models proceeds as follows:

1. Choose a starting value, β^0 and θ^0 . For example, we can initialize $\beta^0 = 0$ and $\theta^0 = 0.5$.
2. For $t = 1, 2, 3, \dots$,

E-step: Update γ_j and S_j^{-1} by their conditional posterior expectations.

M-step:

- a. Update β using the cyclic coordinate descent algorithm;
- b. Update θ .

We assess convergence by the criterion: $|d^{(t)} - d^{(t-1)}| / (0.1 - |d^{(t)}|) < \varepsilon$, where $d^{(t)} = pI(\beta^{(t)})$ is the estimate of deviance at the t th iteration, and ε is a small value (say 10^{-5}).

2.4 Selecting optimal scale values

The performance of the spike-and-slab lasso Cox model approach can depend on the scale parameters (s_0, s_1). We can usually analyze data with reasonably preset scale values, for example, setting the slab scale s_1 to a value (say $s_1 = 0.5$) that induces weak shrinkage and the spike scale s_0 to a small value (say $s_0 = 0.04$) that induces strong shrinkage. Rather than restricting attention to a single model, however, our fast algorithm allows us to quickly fit a sequence of models, from which we can choose an optimal one based on some criteria. Our strategy is to fix the slab scale $s_1 = 0.5$, and consider a sequence of L decreasing values $\{s_0^l\}$: $s_1 > s_0^1 > s_0^2 > \dots > s_0^L > 0$, for the spike scale s_0 . We then fit L models with scales $\{(s_0^l, s_1); l = 1, \dots, L\}$. Increasing the spike scale s_0 tends to include more non-zero coefficients in the model. This procedure is similar to the lasso implemented in the widely-used R package `glmnet`, which quickly fits the lasso Cox models over a grid of values of λ covering its entire range, giving a sequence of models for users to choose from (Hastie et al. 2015; Simon et al. 2011).

There are several ways to measure the performance of a fitted Cox model, including *the partial log-likelihood (PL)*, *the concordance index (C-index)*, *the survival curves* and *the survival prediction error* (van Houwelingen and Putter 2012). The partial log-likelihood function measures the overall quality of a fitted Cox model, and thus is usually used to choose an optimal model (Simon et al. 2011; van Houwelingen and Putter 2012; van Houwelingen et al. 2006). To evaluate the predictive performance of a Cox model, a general way is to fit the model using a dataset and then calculate the above measures with independent data. We use the pre-validation method, a variant of cross-validation (Hastie et al. 2015; Tibshirani and Efron 2002). We randomly split the data to K subsets of roughly the same size, and use $(K-1)$ subsets to fit a model. Denote the estimate of coefficients from the data excluding the k th subset by $\hat{\beta}^{(-k)}$. We calculate the prognostic indices $\hat{\eta}_{(k)} = X_{(k)} \hat{\beta}^{(-k)}$ for all individuals in the k th subset of the data, called the cross-validated or pre-validated prognostic index. Cycling through K parts, we obtain the cross-validated prognostic indices $\hat{\eta}_i$ for all individuals. We then use $(t_i, d_i, \hat{\eta}_i)$ to compute the measures described above. The cross-validated prognostic score for each patient is derived independently of the observed response of the patient, and hence the ‘pre-validated’ dataset $\{t_i, d_i, \hat{\eta}_i\}$ can essentially be treated as a ‘new dataset’. Therefore, this procedure provides valid assessment of the predictive performance of the model (Hastie et al. 2015; Tibshirani and Efron 2002).

We also use an alternative way to evaluate the partial log-likelihood, i.e., the so-called cross-validated partial likelihood (CVPL), defined as (Simon et al. 2011; van Houwelingen and Putter 2012; van Houwelingen et al. 2006)

$$\text{CVPL} = \sum_{k=1}^K \left[pL(\hat{\beta}^{(-k)}) - pL_{(-k)}(\hat{\beta}^{(-k)}) \right], \quad (12)$$

where $\hat{\beta}^{(-k)}$ is the estimate of β from all the data except the k th part, $pL(\hat{\beta}^{(-k)})$ is the partial likelihood of all the data points and $pL_{(-k)}(\hat{\beta}^{(-k)})$ is the partial likelihood excluding part k of the data. By subtracting the log-partial likelihood evaluated on the non-left out data from that evaluated on the full data, we can make efficient use of the death times of the left out data in relation to the death times of all the data.

2.5 Implementation

We have created an R function `bmlasso()` for setting up and fitting the spike-and-slab lasso Cox, and several other R functions (e.g.

`summary.bh`, `plot.bh`, `predict.bh`, `cv.bh`) for summarizing the fitted models and for evaluating the predictive performance. We have incorporated these functions into the freely available R package `BhGLM` (<http://www.ssg.uab.edu/bhglm/>).

3 Simulation study

3.1 Simulation design

We used simulations to validate the proposed spike-and-slab lasso approach, and to compare with the lasso in the R package `glmnet`. In each situation, we simulated two datasets, and used the first one as the training data to fit the models and the second one as the test data to evaluate the predictive values. For each simulation setting, we replicated the simulation 50 times and summarized the results over these replicates. We reported the results on the predictive values including partial log-likelihood and C-index in the test data, the accuracy of parameter estimates and the proportions of coefficients included in the model.

For each dataset, we generated n ($=500$) observations, each with a survival response, consisting of an observed censored survival time t_i and a censoring indicator d_i , and a vector of m ($=200, 1000$) continuous predictors $X_i = (x_{i1}, \dots, x_{im})$. The vector X_i , was generated with 50 elements at a time, i.e. the sub-vector $(x_{i(50k+1)}, \dots, x_{i(50k+50)})$, $k = 0, 1, \dots$, was randomly sampled from multivariate normal distribution $N_{50}(0, \Sigma)$, where $\Sigma = (\sigma_{ij})$ with $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.6$ ($j \neq i$). Thus, the predictors within a group were correlated and between groups were independent. We generated ‘true’ survival time T_i for each individual from the exponential distribution: $T_i \sim \exp(\sum_{j=1}^m x_{ij} \beta_j)$, where the coefficients β_j were preset as described below. We then generated censoring time C_i for each individual from the exponential distribution: $C_i \sim \exp(r_i)$, where r_i were randomly sampled from a standard normal distribution. The observed censored survival time t_i was set to be the minimum of the ‘true’ survival and censoring times, $t_i = \min(T_i, C_i)$, and the censoring indicator d_i was set to be 1 if $C_i > T_i$ and 0 otherwise.

We set five coefficients $\beta_5, \beta_{20}, \beta_{40}, \beta_{m-50}$ and β_{m-5} to be non-zero, two of which were negative, and all others to be 0. Table 1 shows the preset non-zero coefficient values for six simulation scenarios.

We analyzed each simulated dataset using the lasso Cox model implemented in the R package `glmnet` and the proposed spike-and-slab lasso Cox model. For the lasso Cox approach, we used 10-fold cross-validation to select an optimal value of λ , which determines an optimal lasso Cox model, and reported the results based on the optimal lasso Cox model. For the spike-and-slab lasso Cox approach, we mainly considered a slab scale $s_1 = 0.5$, and six spike scales: $s_0 = s_\lambda + 0.02$, $s_\lambda + 0.01$, $s_\lambda, s_\lambda - 0.01$, $s_\lambda - 0.02$, $s_\lambda - 0.03$, where s_λ is the scale of the double-exponential prior corresponding to the optimal lasso Cox model, i.e. $s_\lambda = 1/\lambda$ as described in Equation (4). To investigate the impact of the scales (s_0, s_1) on

Table 1. The simulated effect sizes of five coefficients under different scenarios ($n = 500$)

		β_5	β_{20}	β_{40}	β_{m-50}	β_{m-5}
Scenario 1	$m = 200$	0.214	0.261	0.262	-0.296	-0.422
Scenario 2	$m = 200$	0.359	0.451	-0.445	-0.518	0.725
Scenario 3	$m = 200$	-0.522	-0.635	0.630	0.744	1.017
Scenario 4	$m = 1000$	-0.209	-0.257	0.258	0.298	0.422
Scenario 5	$m = 1000$	0.366	0.450	0.446	-0.518	-0.725
Scenario 6	$m = 1000$	0.514	-0.625	0.631	0.722	-1.037

Table 2. The partial log-likelihood and C-index values over 50 simulated replicates under different simulated scenarios

Simulated scenarios	Methods	Partial log-likelihood	C-index
Scenario 1	lasso: $s_\lambda = 0.045^a$	-1200.49(62.769)	0.680(0.022)
	ssllasso: $s_0 = s_\lambda$	-1201.039(63.520)	0.679(0.022)
Scenario 2	lasso: $s_\lambda = 0.052^a$	-1204.975(49.157)	0.735(0.019)
	ssllasso: $s_0 = s_\lambda - 0.02$	-1198.005(49.900)	0.740(0.020)
Scenario 3	lasso: $s_\lambda = 0.057^a$	-1150.189(60.427)	0.795(0.016)
	ssllasso: $s_0 = s_\lambda - 0.03$	-1141.872(61.211)	0.800(0.016)
Scenario 4	lasso: $s_\lambda = 0.029^a$	-1195.046(52.336)	0.624(0.025)
	ssllasso: $s_0 = s_\lambda + 0.01$	-1192.936(51.836)	0.625(0.028)
Scenario 5	lasso: $s_\lambda = 0.034^a$	-1153.503(53.632)	0.785(0.015)
	ssllasso: $s_0 = s_\lambda - 0.02$	-1140.097(54.464)	0.790(0.015)
Scenario 6	lasso: $s_\lambda = 0.042^a$	-1159.421(55.534)	0.792(0.018)
	ssllasso: $s_0 = s_\lambda - 0.02$	-1143.262(57.398)	0.801(0.018)

Note: Scenario 1, 2 and 3: $n = 500, m = 200$; Scenario 4, 5 and 6: $n = 500, m = 1000$. The slab scales, s_1 , are 0.5 in all scenarios. The optimal models for different scenarios are summarized here. More results can be found in Supplementary Table S1

^a s_λ denotes the average value of penalty parameters s_λ ($s_\lambda = 1/\lambda$) of 50 repeated samples, where λ is the optimal value of the penalty parameter for the lasso. 'ssllasso' represents the spike-and-slab lasso Cox model

the results, we fit the simulated datasets under scenarios 3 and 6 with 81 combinations when s_0 changed from 0.01 to 0.09 and s_1 from 0.1 to 0.9.

3.2 Simulation results

3.2.1 Predictive performance

Supplementary Table S1 shows the partial log-likelihood and C-index in the test data from the lasso and the spike-and-slab lasso under different s_0 values. Table 2 shows the selected optimal model under different simulated scenarios. From these results, we can see that the spike-and-slab lasso Cox with an appropriate value of s_0 performed better than the lasso. Table 2 also shows that the spike-and-slab lasso had slightly higher C-index value than the lasso. Thus, the spike-and-slab lasso approach can generate better discrimination.

From Table 2 and Supplementary Table S1, we found that the best models depended on the spike scale s_0 . Supplementary Figure S1 shows the profiles of the partial log-likelihood function on a grid of values of (s_0, s_1) for scenarios 3 and 6. It can be seen that the slab scale s_1 within the range [0.1, 0.9] had little influence on the partial log-likelihood function, while the spike scale s_0 strongly affected the model performance. These results show that our approach with a fixed slab scale s_1 is reasonable.

3.2.2 Solution path and adaptive property for the spike-and-slab lasso

The performance of the proposed method strongly depended on the spike scale s_0 . To fully investigate the impact of the spike scale s_0 on the results, Figure 1A and B presents the solution paths under scenario 3 by the proposed model with $s_1 = 0.5$ and the lasso model, respectively. Figure 1C and D shows the profiles of CVPL by 10-fold cross-validation for the proposed model and the lasso model, respectively. Similar to the lasso, the spike-and-slab lasso Cox is a path-following strategy for fast dynamic posterior exploration. However, the solution path is essentially different from that of the lasso model. For the lasso solution as shown in Figure 1B, the number of non-zero coefficient could be a few, even zero if a strong penalty is adopted. However, in the spike-and-slab lasso, larger coefficients will be always included in the model with none or weak shrinkage, while irrelevant coefficients are removed. Supplementary

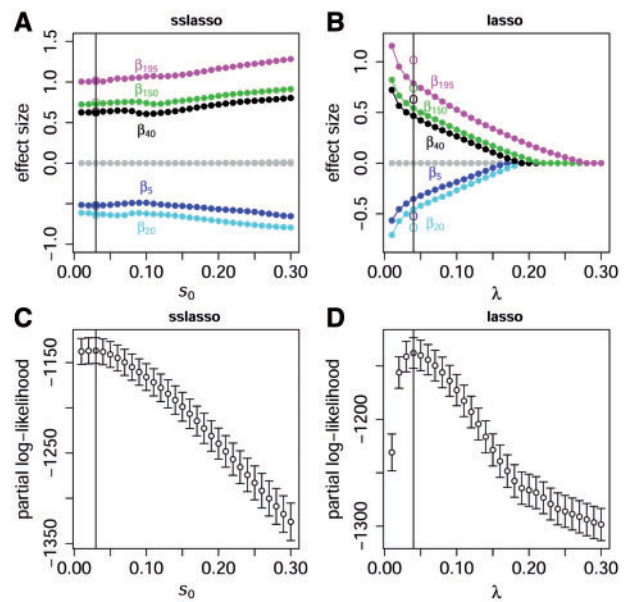


Fig. 1. The solution paths and partial log-likelihood profiles of the proposed spike-and-slab lasso Cox model and the lasso Cox model under scenario 3 with $n = 500$ and $m = 200$. (A, C) Present the solution path and the CVPL profile for proposed spike-and-slab lasso Cox model, respectively. (B, D) Present the solution path and the CVPL profile for the lasso model, respectively. The points and circles on the solution paths represent the estimated values of five simulated non-zero coefficients over 50 replicates and the true non-zero coefficients. The vertical lines correspond to the optimal models with the largest CVPL

Figure S2 shows the solution paths of the proposed model with $s_1 = 0.5$ and the lasso model under scenario 6.

To show the adaptive property of the proposed method, we performed additional simulation study with $(n, m) = (500, 1000)$ and $(s_0, s_1) = (0.05, 0.5)$. Supplementary Figure S3 shows the adaptive shrinkage amount, along with the different effect size. It clearly shows that the proposed spike-and-slab lasso Cox model approach has self-adaptive and flexible characteristics.

3.2.3 Accuracy of parameter estimates

Figure 2 and Supplementary Figures S4 and S5 show the estimates of coefficients from the spike-and-slab lasso Cox model and the lasso

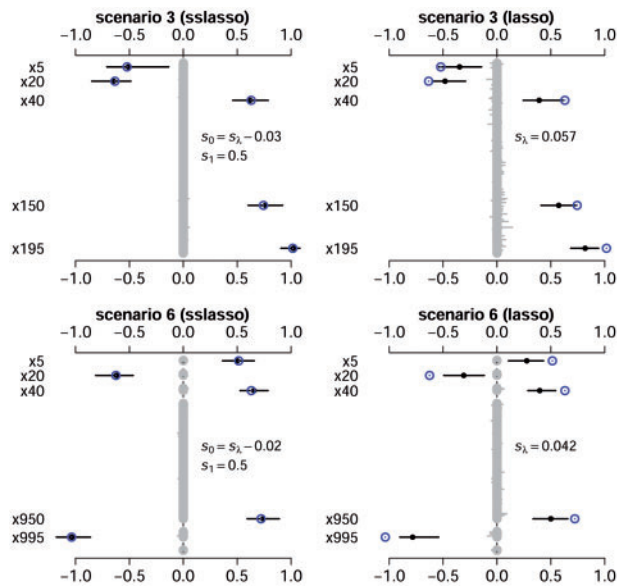


Fig. 2. The parameter estimation averaged over 50 replicates for the spike-and-slab lasso Cox model and the lasso Cox model. The numbers of simulated predictors are $m=200$ for scenario 3 and $m=1000$ for scenario 6. The cycles are the assumed true values. The black points and lines represent the estimated values and the interval estimates of coefficients over 50 replicates

Cox model over 50 replicates of training data. It can be seen that the spike-and-slab lasso provided more accurate estimation in most situations, especially for larger coefficients. In contrast, the lasso can over-shrink large coefficients, due to employing a single penalty on all the coefficients.

We also summarized the mean absolute errors (MAE) of coefficient estimates, defined as $\sum |\hat{\beta}_j - \beta_j|/m$, in Table 3. A smaller MAE suggests more accurate parameter estimation. It can be seen that the proposed models gave smaller MAE in all scenarios.

3.2.4 Proportions of coefficients included in the model

We calculated the proportions for the simulated non-zero coefficients included in the model over the simulation replicates. Like the lasso approach, the proposed spike-and-slab lasso can estimate coefficients to be 0, and thus can easily return these proportions. Figure 3 and Supplementary Figure S6 show the inclusion proportions of the non-zero coefficients and the zero coefficients for the spike-and-slab lasso and the lasso Cox models. It can be seen that in most situations the inclusion proportions of the non-zero coefficients were similar for the two approaches. However, the lasso included zero coefficients in the model more frequently than the spike-and-slab lasso. This indicates that the spike-and-slab lasso approach can reduce noisy signals.

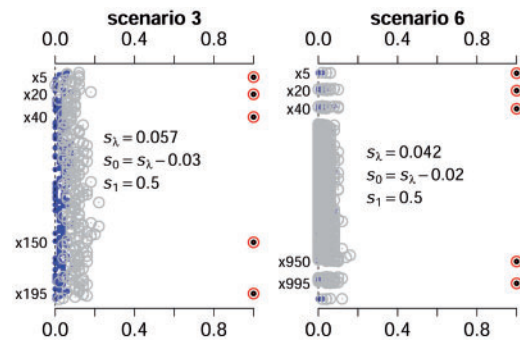


Fig. 3. The inclusion proportions of the non-zero and zero coefficients in the model over the simulation replicates under scenarios 3 and 6. The points and cycles nearby the right part of the figure represent the proportions of non-zero coefficients for the spike-and-slab lasso Cox model and the lasso Cox model, respectively. The points and grey cycles nearby left axis represent the proportions of zero coefficients for the spike-and-slab lasso Cox model and the lasso Cox model, respectively

We summarized the average numbers of non-zero coefficients in Table 3. In most simulated scenarios, except for scenario 4, the average numbers of non-zero coefficients in the spike-and-slab lasso Cox models were much lower than those in the lasso Cox models. We also found that the average numbers of non-zero coefficients detected by the proposed models were close to the number of the simulated non-zero coefficients in most scenarios. However, the lasso always included many zero coefficients in the model.

4 Application to real data

4.1 Dutch breast cancer data

We applied our spike-and-slab lasso Cox model to analyze the well-known Dutch breast cancer data set. This dataset contains the microarray mRNA expression measurements of 4919 genes and the time of metastasis after adjuvant systemic therapy (a censored survival outcome) from 295 women with breast cancer (van de Vijver, et al. 2002; van't Veer et al. 2002). The 4919 genes were selected from 24,885 genes, for which reliable expression is available (van't Veer et al. 2002). Among 295 tumors, 88 had distant metastases. Our analysis was to build a survival model for predicting the metastasis time using the 4919 gene-expression predictors. Prior to fitting the models, we standardized all the predictors.

We fixed the slab scale s_1 to 0.5, and varied the spike scale s_0 over the grid of values: $0.005 + k \times 0.005$; $k=0, 1, \dots, 17$, leading to 18 models. For each model, we performed 10-fold cross-validation with 10 replicates to select an optimal model based on the cross-validated partial log-likelihood (CVPL). Supplementary Figure S7 shows the profiles of CVPL. The largest value of CVPL appeared to be

Table 3. Average number of non-zero coefficients and mean absolute error (MAE) of coefficient estimates over 50 simulations for the best spike-and-slab lasso Cox model and the best lasso model

	slasso ($s_1 = 0.5$)		lasso	
	Average number	MAE	Average number	MAE
Scenario 1, $s_0 = s_\lambda$	18.68	1.285(0.402)	20.10	1.348(0.378)
Scenario 2, $s_0 = s_\lambda - 0.02$	7.82	0.605(0.319)	20.52	1.531(0.308)
Scenario 3, $s_0 = s_\lambda - 0.03$	6.92	0.458(0.253)	23.58	1.618(0.340)
Scenario 4, $s_0 = s_\lambda + 0.01$	35.08	1.708(0.432)	16.34	1.374(0.257)
Scenario 5, $s_0 = s_\lambda - 0.02$	5.42	0.422(0.252)	24.68	1.480(0.284)
Scenario 6, $s_0 = s_\lambda - 0.02$	8.22	0.457(0.316)	37.36	2.183(0.404)

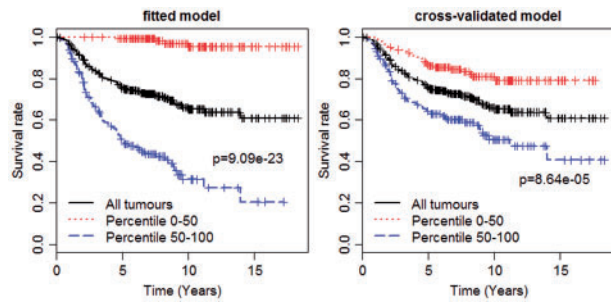


Fig. 4. The Kaplan–Meier survival curves of the fitted spike-and-slab lasso Cox model (left panel) and the cross-validated model (right panel) for Dutch breast cancer data set. The Log-rank tests are performed to compare the two curves and obtain P values

–541.237, when the spike scale s_0 was 0.05. Therefore, the spike-and-slab lasso Cox model with the scale (0.05, 0.5) was chosen for the prediction. This proposed model detected 42 genes, and the effect sizes for most genes were small (Supplementary Fig. S8).

We performed 10-fold cross-validation with 10 replicates to evaluate the predictive values of the chosen model. The cross-validated C-index was estimated to be 0.670, which was significantly larger than 0.5, showing the discriminative ability of the final prognostic model. We estimated the cross-validated prognostic index, $\eta_i = X_i\beta$, for each patient, and then grouped the patients on the basis of the prognostic index into two groups of equal size by 50th percentile. The Kaplan–Meier survival curve in each subgroup was calculated to assess the predictive usefulness of the proposed model. Figure 4 shows the Kaplan–Meier survival curves for the chosen model fitted using the entire data (left panel) and the cross-validation (right panel). The Log-rank tests were performed to compare the survival estimates between two groups. As expected, the separation of the curves from cross-validation was not as strong as that of the model fitted using the entire data. However, it was still significant, indicating that the proposed model was very informative on prediction.

Measuring survival prediction error is an alternative important way to evaluate the predictive performance of a survival model. Supplementary Figure S9 presents the prediction error curves for the fitted model, the cross-validated model and the null model (i.e. with no predictors). As expected, the proposed model had much lower prediction error than the cross-validated model and the null model. It also can be seen that the cross-validated model had lower prediction error than the null model. Thus, the detected genes provided valuable predictive information.

4.2 MDS data

The second dataset that we analyzed was a recently published data on MDS (Gerstung *et al.* 2015). Clinical data are available for 142 MDS patients, where 24 of them had 0 survival time and were excluded from the analysis. The outcome of interest is acute myeloid leukaemia (AML)-free survival. Among the 118 samples, the number of dead patients was 40. For gene expression (from CD34+ bone marrow cells), 124 samples have 21,762 features profiled. Even though we can analyze all the 21,762 features, considering that the number of genes related to AML free survival is not expected to be too large, we filtered the expression data using variance with cutoff 0.9 and selected 2177 gene expression for predictive modeling.

Gerstung *et al.* (2015) built a prognostic model for predicting AML-free survival using the first 20 principle components of all

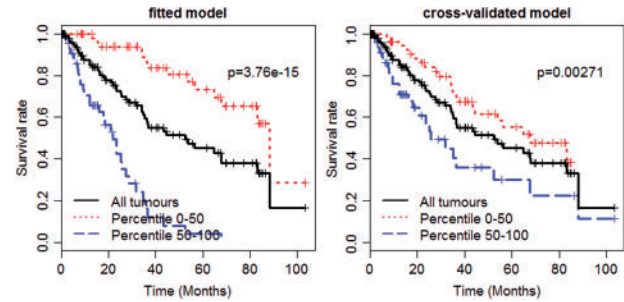


Fig. 5. The Kaplan–Meier survival curves of the fitted spike-and-slab lasso Cox model (left panel) and the cross-validated models (right panel) for MDS dataset. The Log-rank tests are performed to compare the two survival curves and obtain P values

21,762 gene expressions, and obtained an impressive prediction with C-index=0.76. However, their model failed to identify important genes. Our analysis had advantages of not only providing prediction but also detecting associated genes.

Similar to the above analysis of the breast cancer dataset, we fixed the slab scale s_1 to 0.5 and considered the spike scale s_0 over the grid of values: $0.005 + k \times 0.005$; $k = 0, 1, \dots, 17$, and performed 10-fold cross-validation with 10 replicates to select an optimal model based on the CVPL. From Supplementary Figure S10, we can see that $s_0 = 0.075$ resulted in the largest CVPL. Therefore, the scale (0.075, 0.5) was chosen for our spike-and-slab lasso Cox model. This model detected 19 genes, and the effect sizes for most genes were small (Supplementary Fig. S11).

The cross-validated C-index was estimated to be 0.695, which is close that of Gerstung *et al.* (2015). However, our prognostic model only included 19 genes, and thus was much more easily interpreted and clinically useful. We further estimated the Kaplan–Meier survival curves for two groups of equal size by 50th percentile of the cross-validated prognostic index values. Figure 5 shows that the Kaplan–Meier survival curves for the two groups were significantly different. We also estimated the prediction error curves for the fitted model, the cross-validated model and the null model. Supplementary Figure S12 shows that the fitted model and the cross-validated model had lower prediction error than the null model. Thus, the detected genes provided valuable predictive information.

We also used the lasso Cox model to analyze these two datasets and compared the results of the lasso Cox model with the above results. The measures of predictive performance by the two methods were close, which may result from the small effect sizes of the detected genes. However, the lasso models included 46 and 31 genes for Dutch breast cancer and MDS datasets, respectively, while the proposed models included 42 and 19 genes for the two data, respectively, to achieve similar prediction accuracy. Most of the detected genes by the two approaches were overlapped. Although having not greatly improved the prediction, the proposed method generated simpler and more clinically useful prognostic models.

5 Discussion

We have developed a new hierarchical model approach, i.e. the spike-and-slab lasso Cox models, for detecting important variables and predicting survival outcomes (e.g. the time until an event such as tumor recurrence or death). Although focusing on molecular profiling data, the proposed approach also can be used for analyzing general large-scale survival data.

The key to our spike-and-slab lasso Cox models is proposing the new prior distribution, i.e. the spike-and-slab double-exponential prior, on the coefficients. The spike-and-slab double-exponential priors can induce different amounts of shrinkage for different predictors depending on their effect sizes, thus reducing the noises of irrelevant predictors and improving the accuracy of coefficient estimates and prognostic predictions. The spike-and-slab lasso Cox models can be effectively fitted by the proposed EM coordinate descent algorithm, which incorporates EM steps into the cyclic coordinate descent algorithm. The E-steps involve calculating the posterior expectations of the indicator variable γ_j and the scale S_j for each coefficient, and the M-steps employ the existing fast algorithm, i.e. the cyclic coordinate descent algorithm (Friedman et al. 2010; Hastie et al. 2015; Simon et al. 2011), to update the coefficients. As observed in our extensive data analyses, the EM coordinate descent algorithm converges rapidly, and is capable of identifying important predictors and building promising predictive models from numerous candidates. In our real data analysis, it took 0.014 and 0.004 min for Dutch breast cancer dataset and MDS dataset, respectively. The 10-fold cross-validation procedure took a little longer, but still less than one minute for both the two datasets.

The spike-and-slab lasso bridges two popular methods for high-dimensional data analysis, i.e. Bayesian variable selection (Chipman 1996; Chipman et al. 2001; George and McCulloch 1993, 1997; Ročková and George 2014) and the penalized lasso (Hastie et al. 2015; Tibshirani 1996, 1997), into one unifying framework, and thus remains the nice features of these two methods while diminishing their shortcomings (Ročková and George, 2016 unpublished manuscript). Similar to the penalized lasso, the spike-and-slab lasso can shrink irrelevant coefficients exactly to 0, thus automatically achieving variable selection and yielding easily interpretable results. More importantly, due to using the spike-and-slab mixture prior, the shrinkage scale for each predictor can be estimated from the data, yielding weak shrinkage on important predictors but strong shrinkage on irrelevant predictors and thus diminishing the well-known estimation bias of the lasso.

The spike-and-slab lasso Cox models include two scale values (s_0, s_1) which should be preset. A comprehensive approach is to perform a two-dimensional search on all plausible combinations of (s_0, s_1), and then to select an optimal model based on cross-validation. However, this approach can be time-consuming and inconvenient to use. Usually, the slab scale s_1 has little influence on the fitted model, while the spike scale s_0 can strongly affect the model performance. Similar to the approach of Ročková and George (2014, 2016), therefore, our strategy is to fix the slab scale s_1 to be relatively large, and select an optimal spike scale s_0 from a reasonable range, e.g. (0, 0.1), in real data analysis. One also can use the lasso model to guide the selection of s_0 . We found that the optimal value of s_0 is usually close to $s_2 (=1/\lambda)$ obtained from the lasso. These findings can largely simplify our selection of the optimal spike scale.

The proposed spike-and-slab lasso approach may be easily extended to frameworks beyond normal linear regression and Cox survival models. We have already extended the spike-and-slab lasso to generalized linear models (Tang et al. 2017). With the spike-and-slab double-exponential prior, the conditional posterior expectations of missing values (i.e. the indicator variable γ_j and the scale S_j) depend only on the coefficients β , and thus the E-step developed can be applied to other models. Another important avenue for future research will be incorporating external information about the importance of predictors and the relationship between predictors into the proposed spike-and-slab lasso framework. The spike-and-slab mixture priors provide flexible and easy ways to incorporate

structural information about the predictors into predictive modeling. These extensions would be essential for effectively integrating important biological information, and multi-level molecular profiling data.

Acknowledgements

We thank the three reviewers and the associate editor for their constructive suggestions and comments that have improved the manuscript.

Funding

This work was supported by the National Institutes of Health [NIH 2R01GM069430, R03-DE024198, R03-DE025646], and was also supported by grants from China Scholarship Council, the National Natural Science Foundation of China [81573253], and project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions at Soochow University.

Conflict of Interest: none declared.

References

- Barillot, E. et al. (2012) *Computational Systems Biology of Cancer*. Chapman & Hall/CRC Mathematical & Computational Biology.
- Bonato, V. et al. (2011) Bayesian ensemble methods for survival prediction in gene expression data. *Bioinformatics*, **27**, 359–367.
- Bovelstad, H.M. et al. (2007) Predicting survival from microarray data—a comparative study. *Bioinformatics*, **23**, 2080–2087.
- Bovelstad, H.M. et al. (2009) Survival prediction from clinico-genomic models—a comparative study. *BMC Bioinform.*, **10**, 413.
- Breslow, N. (1974) Covariance analysis of censored survival data. *Biometrics*, **30**, 89–99.
- Breslow, N.E. (1972) Contribution to the discussion of the paper by D.R. Cox. *J. R. Stat. Soc. B*, **34**, 216–217.
- Chin, L. et al. (2011) Cancer genomics: from discovery science to personalized medicine. *Nat. Med.*, **17**, 297–303.
- Chipman, H. (1996) Bayesian variable selection with related predictions. *Can. J. Stat.*, **24**, 17–36.
- Chipman, H. et al. (2001) The practical implementation of Bayesian model selection. In: Lahiri, P. (ed), *Model Selection*. Institute of Mathematical Statistics, Beachwood, Ohio.
- Collins, F.S., and Varmus, H. (2015) A new initiative on precision medicine. *N. Engl. J. Med.*, **372**, 793–795.
- Cox, D.R. (1972) Regression models and life tables. *J. R. Stat. Soc.*, **34**, 187–220.
- Efron, B. (1977) The efficiency of Cox's likelihood function for censored data. *J. Am. Stat. Assoc.*, **72**, 557–565.
- Friedman, J. et al. (2010) Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.*, **33**, 1–22.
- George, E.I., and McCulloch, R.E. (1993) Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.*, **88**, 881–889.
- George, E.I., and McCulloch, R.E. (1997) Approaches for Bayesian variable selection. *Statistica Sinica*, **7**, 339–373.
- Gerstung, M. et al. (2015) Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat. Commun.*, **6**, 5901.
- Hastie, T. et al. (2015) *Statistical Learning with Sparsity—the Lasso and Generalization*. CRC Press, New York.
- Ibrahim, J. et al. (2001) *Bayesian Survival Analysis*. Springer, New York.
- Klein, J., and Moeschberger, M. (2003) *Survival Analysis*. Springer, New York.
- Kyung, M. et al. (2010) Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal.*, **5**, 369–412.
- Lee, K.H. et al. (2011) Bayesian variable selection in semiparametric proportional hazards model for high dimensional survival data. *Int. J. Biostat.*, **7**, 21.

- Lee, K.H. *et al.* (2015) Survival prediction and variable selection with simultaneous shrinkage and grouping priors. *Stat. Anal. Data Min.*, **8**, 114–127.
- Li, F., and Zhang, N.R. (2010) Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *J. Am. Stat. Assoc.*, **105**, 1202–1214.
- Li, Y. *et al.* (2016) Regularized Parametric Regression for High-dimensional Survival Analysis. *Siam International Conference on Data Mining*. pp. 765–773.
- Mittal, S. *et al.* (2013) Large-scale parametric survival analysis. *Stat. Med.*, **32**, 3955–3971.
- Monni, S., and Li, H. (2010) Bayesian methods for network-structured genomics data. In: Chen, M.-H. *et al.* (eds) *Frontiers of Statistical Decision Making and Bayesian Analysis In Honor of James O Berger*. Springer, New York.
- Park, T., and Casella, G. (2008) The Bayesian Lasso. *J. Am. Stat. Assoc.*, **103**, 681–686.
- Peng, B. *et al.* (2013) An integrative framework for Bayesian variable selection with informative priors for identifying genes and pathways. *PLoS One*, **8**, e67672.
- Ročková, V., and George, E.I. (2014) EMVS: the EM approach to Bayesian variable selection. *J. Am. Stat. Assoc.*, **109**, 828–846.
- Ročková, V. *et al.* (2016) Bayesian penalty mixing: the case of a non-separable penalty. In: Frigessi, A. *et al.* (eds.), *Statistical Analysis for High-Dimensional Data: The Abel Symposium 2014*. Cham, Springer International Publishing, pp. 233–254.
- Simon, N. *et al.* (2011) Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.*, **39**, 1–13.
- Sohn, I., and Sung, C.O. (2013) Predictive modeling using a somatic mutational profile in ovarian high grade serous carcinoma. *PLoS One*, **8**, e54089.
- Stingo, F. *et al.* (2010) A Bayesian graphical modeling approach to microRNA regulatory network inference. *Annal. Appl. Stat.*, **4**, 2024–2048.
- Tai, F. *et al.* (2009) Bayesian variable selection in regression with networked predictors. University of Minnesota Biostatistics Technical Report. <http://www.sph.umn.edu/faculty1/wpcontent/uploads/2012/11/rr2009-008.pdf>.
- Tang, Z. *et al.* (2017) The spike-and-slab lasso generalized linear models for prediction and associated genes detection. *Genetics*, **205**, 77–88.
- Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B*, **58**, 267–288.
- Tibshirani, R. (1997) The lasso method for variable selection in the Cox model. *Stat. Med.*, **16**, 385–395.
- Tibshirani, R.J., and Efron, B. (2002) Pre-validation and inference in microarrays. *Stat. Appl. Genet. Mol. Biol.*, **1**, 1–18.
- van de Vijver, M.J., *et al.* (2002) A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.*, **347**, 1999–2009.
- van Houwelingen, H.C. *et al.* (2006) Cross-validated Cox regression on microarray gene expression data. *Stat. Med.*, **25**, 3201–3216.
- van Houwelingen, H.G., and Putter, H. (2012) *Dynamic Prediction in Clinical Survival Analysis*. CRC Press, Boca Raton.
- Van't Veer, L.J. *et al.* (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**, 530–536.
- Yi, N., and Xu, S. (2008) Bayesian LASSO for quantitative trait loci mapping. *Genetics*, **179**, 1045–1055.
- Yuan, Y. *et al.* (2014) Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat. Biotechnol.*, **32**, 644–652.
- Zhang, W. *et al.* (2013) Network-based survival analysis reveals subnetwork signatures for predicting outcomes of ovarian cancer treatment. *PLoS Comput. Biol.*, **9**, e1002975.
- Zhao, Q. *et al.* (2015) Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. *Brief. Bioinform.*, **16**, 291–303.
- Zou, H., and Hastie, T. (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B*, **67**, 301–320.