The Spike-and-Slab Lasso Generalized Linear Models for Prediction and Associated Genes Detection

Zaixiang Tang,*,†,‡ Yueping Shen,*,† Xinyan Zhang,‡ and Nengjun Yi‡,1

*Department of Biostatistics, School of Public Health and †Jiangsu 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 ‡Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Alabama 35294

ABSTRACT Large-scale "omics" data have been increasingly used as an important resource for prognostic prediction of diseases and detection of associated genes. However, there are considerable challenges in analyzing high-dimensional molecular data, including the large number of potential molecular predictors, limited number of samples, and small effect of each predictor. We propose new Bayesian hierarchical generalized linear models, called spike-and-slab lasso GLMs, for prognostic prediction and detection of associated genes using large-scale molecular data. The proposed model employs a spike-and-slab mixture double-exponential prior for coefficients that can induce weak shrinkage on large coefficients, and strong shrinkage on irrelevant coefficients. We have developed a fast and stable algorithm to fit large-scale hierarchal GLMs by incorporating expectation-maximization (EM) steps into the fast cyclic coordinate descent algorithm. The proposed approach integrates nice features of two popular methods, *i.e.*, penalized lasso and Bayesian spike-and-slab variable selection. The performance of the proposed method is assessed via extensive simulation studies. The results show that the proposed approach can provide not only more accurate estimates of the parameters, but also better prediction. We demonstrate the proposed procedure on two cancer data sets: a well-known breast cancer data set consisting of 295 tumors, and expression data of 4919 genes; and the ovarian cancer data set from TCGA with 362 tumors, and expression data of 5336 genes. Our analyses show that the proposed procedure can generate powerful models for predicting outcomes and detecting associated genes. The methods have been implemented in a freely available R package BhGLM (http://www.ssg.uab.edu/bhglm/).

KEYWORDS cancer; double-exponential distribution; generalized linear model; outcome prediction; spike-and-slab lasso

THE growing recognition of precision medicine reflects the emergence of a field that is accelerating rapidly and will help shape new clinical practice in the future (Collins and Varmus 2015; Jameson and Longo 2015). The important basis of precision medicine is to generate knowledge of disease that will enable better assessment of disease risk, understanding of disease mechanisms, and prediction of optimal therapy and prognostic outcome for diseases by using a wide range of biomedical, clinical, and environmental information. Precision medicine needs accurate detection of biomarkers and prognostic prediction (Chin *et al.* 2011; Barillot *et al.* 2012). Traditional clinical prognostic

and predictive factors often provided poor prognosis and prediction (Barillot *et al.* 2012). Modern "omics" technologies can generate robust large-scale molecular data, such as large amounts of genomic, transcriptomic, proteomic, and metabolomics data, which provides extraordinary opportunities to detect new biomarkers, and to build more accurate prognostic and predictive models. However, these large-scale data sets also introduce computational and statistical challenges.

Various approaches have been applied in analyzing large-scale molecular profiling data to address the challenges (Bovelstad *et al.* 2007, 2009; Lee *et al.* 2010; Barillot *et al.* 2012). The lasso, and its extensions, are the most commonly used methods (Tibshirani 1996; Hastie *et al.* 2009, 2015). These methods put an L_1-penalty on the coefficients, and can shrink many coefficients exactly to zero, thus performing variable selection. With the L_1-penalty, the penalized likelihood can be solved by extremely fast optimization algorithms, for example, the lars and the cyclic coordinate

descent algorithms (Efron *et al.* 2004; Friedman *et al.* 2010), making them very popular in high-dimensional data analysis. Recently, these penalization approaches have been applied widely for prediction and prognosis (Rapaport *et al.* 2007; Barillot *et al.* 2012; Sohn and Sung 2013; Zhang *et al.* 2013; Yuan *et al.* 2014; Zhao *et al.* 2014).

However, the lasso uses a single penalty for all coefficients, and thus can either include a number of irrelevant predictors, or over-shrink large coefficients. An ideal method should induce weak shrinkage on large effects, and strong shrinkage on irrelevant effects (Fan and Li 2001; Zou 2006; Zhang 2010). Statisticians have introduced hierarchical models with mixture spike-and-slab priors that can adaptively determine the amount of shrinkage (George and McCulloch 1993, 1997). The spike-and-slab prior is the fundamental basis for most Bayesian variable selection approaches, and has proved remarkably successful (George and McCulloch 1993, 1997; Chipman 1996; Chipman *et al.* 2001; Ročková and George 2014, and unpublished results). Recently, Bayesian spike-and-slab priors have been applied to predictive modeling and variable selection in large-scale genomic studies (Yi *et al.* 2003; Ishwaran and Rao 2005; de los Campos *et al.* 2010; Zhou *et al.* 2013; Lu *et al.* 2015; Shankar *et al.* 2015; Shelton *et al.* 2015; Partovi Nia and Ghannad-Rezaie 2016). However, most previous spike-and-slab variable selection approaches use the mixture normal priors on coefficients and employ Markov Chain Monte Carlo (MCMC) algorithms (*e.g.*, stochastic search variable selection) to fit the model. Although statistically sophisticated, these MCMC methods are computationally intensive for analyzing large-scale molecular data. The mixture normal priors cannot shrink coefficients exactly to zero, and thus cannot automatically perform variable selection. Ročková and George (2014) developed an expectation-maximization (EM) algorithm to fit large-scale linear models with the mixture normal priors.

V. Ročková and E. I. George (unpublished results) recently proposed a new framework, called the spike-and-slab lasso, for high-dimensional normal linear models by using a new prior on the coefficients, *i.e.*, the spike-and-slab mixture double-exponential distribution. They proved that the spike-and-slab lasso has remarkable theoretical and practical properties, and overcomes some drawbacks of the previous approaches. However, the spike-and-slab lasso, and most of the previous methods, were developed based on normal linear models, and cannot be directly applied to other models. Therefore, extensions of high-dimensional methods using mixture priors to frameworks beyond normal linear regression provide important new research directions for both methodological and applied works (Ročková and George 2014, and unpublished results).

In this article, we extend the spike-and-slab lasso framework to generalized linear models, called spike-and-slab lasso GLMs (sslasso GLMs), to jointly analyze large-scale molecular data for building accurate predictive models and identifying important predictors. By using the mixture double-exponential

Table 1 Simulated effect sizes of five nonzero coefficients under different scenarios

Simulated Scenarios		β_5	β_{20}	β_{40}	β_{m-50}	β_{m-5}
Scenario 1	$m = 1000$	0.362	0.395	−0.418	−0.431	0.467
Scenario 2	$m = 1000$	−0.457	−0.491	0.521	0.550	0.585
Scenario 3	$m = 1000$	0.563	−0.610	0.653	−0.672	0.732
Scenario 4	$m = 3000$	0.357	−0.388	0.414	−0.429	0.462
Scenario 5	$m = 3000$	−0.455	0.509	0.528	0.552	−0.592
Scenario 6	$m = 3000$	0.560	−0.618	0.654	−0.673	0.716

m is the number of simulated predictors. For all scenarios, the number of individuals (n) is 500.

priors, sslasso GLMs can adaptively shrink coefficients (*i.e.*, weakly shrink important predictors but strongly shrink irrelevant predictors), and thus can result in accurate estimation and prediction. To fit the sslasso GLMs, we propose an efficient algorithm by incorporating EM steps into the extremely fast cyclic coordinate descent algorithm. The performance of the proposed method is assessed via extensive simulations, and compared with the commonly used lasso GLMs. We apply the proposed procedure to two cancer data sets with binary outcomes and thousands of molecular features. Our results show that the proposed method can generate powerful prognostic models for predicting disease outcome, and can also detect associated genes.

Methods

Generalized linear models

We consider GLMs with a large number of correlated predictors. The observed values of a continuous or discrete response are denoted by $y = (y_1, \cdots, y_n)$. The predictor variables include numerous molecular predictors (*e.g.*, gene expression). A GLM consists of three components: the linear predictor η, the link function h, and the data distribution p (McCullagh and Nelder 1989; Gelman *et al.* 2014). The linear predictor for the i-th individual can be expressed as

$$\eta_i = \beta_0 + \sum_{j=1}^{J} x_{ij}\beta_j = X_i\beta \qquad (1)$$

where β_0 is the intercept, x_{ij} represents the observed value of the j-th variable, β_j is the coefficient, X_i contains all variables, and β is a vector of the intercept and all the coefficients. The mean of the response variable is related to the linear predictor via a link function h:
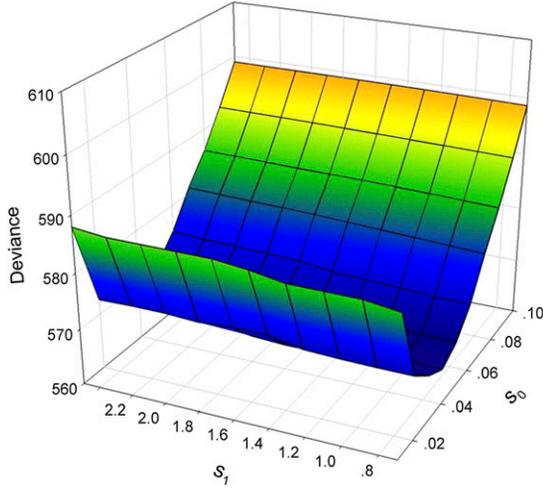
$$E(y_i|X_i) = h^{-1}(X_i\beta) \qquad (2)$$

The data distribution is expressed as

$$p(y|X\beta, \phi) = \prod_{i=1}^{n} p(y_i|X_i\beta, \phi) \qquad (3)$$

where ϕ is a dispersion parameter, and the distribution $p(y_i|X_i\beta, \phi)$ can take various forms, including normal,
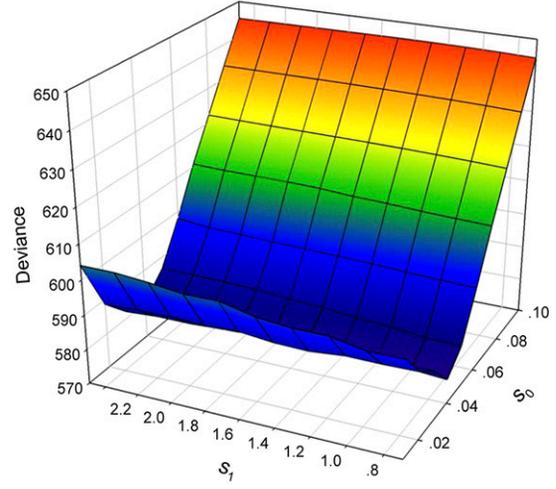
Figure 1 The profiles of deviance under scenarios 3 and 6 over 50 replicated testing datasets. (s_0, s_1) is the prior scale for sslasso GLMs.

binomial, and Poisson distributions. Some GLMs (for example, the binomial distribution) do not require a dispersion parameter; that is, ϕ is fixed at 1. Therefore, GLMs include normal linear and logistic regressions, and various others as special cases.

The classical analysis of generalized linear models is to obtain maximum likelihood estimates (MLE) for the parameters (β, ϕ) by maximizing the logarithm of the likelihood function: $l(\beta, \phi) = \log p(y|X\beta, \phi)$. However, the classical GLMs cannot jointly analyze multiple correlated predictors, due to the problems of nonidentifiability and overfitting. The lasso is a widely used penalization approach to handling high-dimensional models, adding an L_1 penalty to the log-likelihood function, and estimating the parameters by maximizing the penalized log-likelihood (Zou and Hastie 2005; Hastie *et al.* 2009, 2015; Friedman *et al.* 2010):

$$pl(\beta, \phi) = l(\beta, \phi) - \lambda \sum_{j=1}^{J} |\beta_j| \qquad (4)$$

The overall penalty parameter, λ, controls the overall strength of penalty and the size of the coefficients; for a small λ, many coefficients can be large, and for a large λ, many coefficients will be shrunk toward zero. Lasso GLMs can be fit by the extremely fast cyclic coordinate descent algorithm, which successively optimizes the penalized log-likelihood over each parameter, with others fixed, and cycles repeatedly until convergence (Friedman *et al.* 2010; Hastie *et al.* 2015).

With a single penalty parameter, λ, however, the lasso can either overshrink large effects, or include a number of irrelevant predictors. Ideally, one should use small penalty values for important predictors, and large penalties for irrelevant predictors. However, if we have no prior knowledge about the importance of the predictors, we cannot appropriately preset penalties. Here, we propose a new approach, *i.e.*, the

sslasso GLMs, which can induce different shrinkage scales for different coefficients, and allow us to estimate the shrinkage scales from the data.

sslasso GLMs

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

$$\beta_j|s \sim DE(\beta_j|0, s) = \frac{1}{2s} \exp\left(-\frac{|\beta_j|}{s}\right) \qquad (5)$$

where the scale, s, controls the amount of shrinkage; smaller scale induces stronger shrinkage and forces the estimates of β_j toward zero.

We develop sslasso GLMs by extending the double-exponential prior to the spike-and-slab mixture double-exponential prior:

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

or, equivalently

$$\beta_j|\gamma_j, s_0, s_1 \sim DE(\beta_j|0, S_j) = \frac{1}{2S_j} \exp\left(-\frac{|\beta_j|}{S_j}\right) \qquad (6)$$

where γ_j is the indicator variable, $\gamma_j = 1$ or 0, and the scale S_j equals one of two preset positive value 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 as a "spike scale" for modeling irrelevant (zero) 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
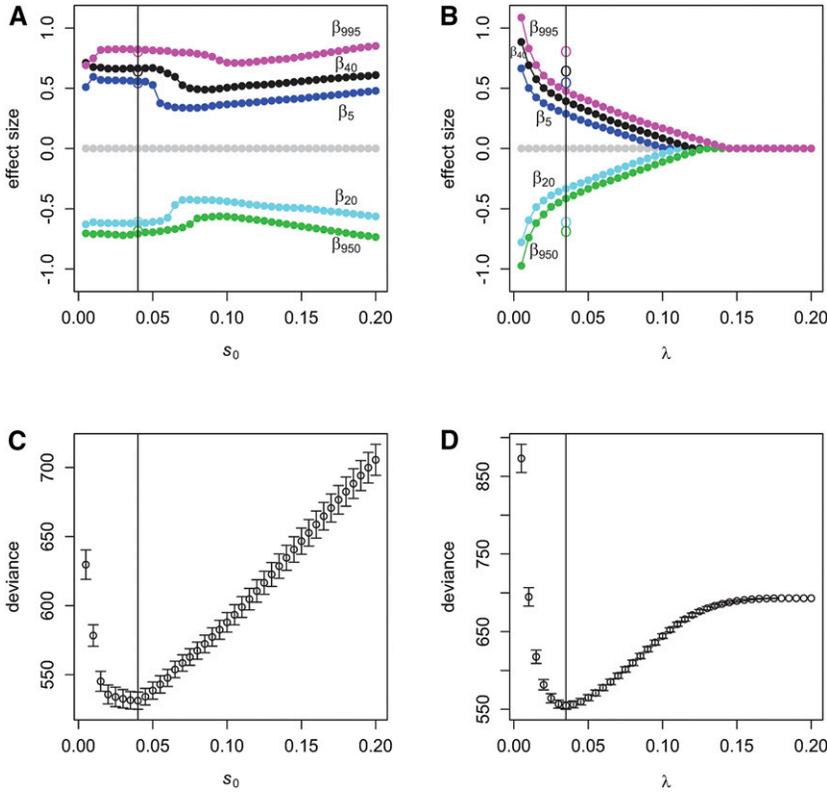
Figure 2 The solution path and deviance profiles of sslasso GLMs (A, C) and the lasso model (B, D) for scenario 3. The colored points on the solution path represent the estimated values of assumed five non-zero coefficients, and the circles represent the true nonzero coefficients. The vertical lines correspond to the optimal models.

no or weak shrinkage on estimation. If we set $s_0 = s_1$, or $\gamma_j = 1$ or 0, 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 in linking the scale parameters with the coefficients. The indicator variables are assumed to follow the independent binomial distribution:

$$\gamma_j | \theta \sim \text{Bin}\left(\gamma_j | 1, \theta\right) = \theta^{\gamma_j}\left(1-\theta\right)^{1-\gamma_j} \tag{7}$$

where θ is the probability parameter. 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 $p\left(\gamma_j = 1 | \theta\right)$. 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.

Algorithm for fitting sslasso GLMs

For high-dimensional data, it is desirable to have an efficient algorithm that can quickly identify important predictors and build a predictive model. We develop a fast algorithm to fit the sslasso GLMs. Our algorithm, called the EM coordinate descent algorithm, incorporates EM steps into the cyclic coordinate descent procedure for fitting the penalized lasso GLMs regression. We derive the EM coordinate descent algorithm based on the log joint posterior density of the parameters $(\beta, \phi, \gamma, \theta)$:

$$\begin{aligned}
\log p\left(\beta, \phi, \gamma, \theta | y\right) &= \log p(y|\beta, \phi) + \sum\nolimits_{j=1}^{J} \log p(\beta_j | S_j) \\
&\quad + \sum\nolimits_{j=1}^{J} \log p(\gamma_j | \theta) + \log p(\theta) \propto l(\beta, \phi) \\
&\quad - \sum_{j=1}^{J} \frac{1}{S_j} |\beta_j| \\
&\quad + \sum_{j=1}^{J} \left(\gamma_j \log \theta + (1-\gamma_j)\log\left(1-\theta\right)\right)
\end{aligned} \tag{8}$$

where $l(\beta, \phi) = \log p(y|X\beta, \phi)o$ and $S_j = (1-\gamma_j)s_0 + \gamma_j s_1$.

The EM coordinate decent algorithm treats the indicator variables γ_j as "missing values," and estimates the parameters (β, ϕ, 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

$$\begin{aligned}
p_j &= p\left(\gamma_j = 1 | \beta_j, \theta, y\right) \\
&= \frac{p\left(\beta_j | \gamma_j = 1, s_1\right)\left(\gamma_j = 1 | \theta\right)}{p\left(\beta_j | \gamma_j = 0, s_0\right)p\left(\gamma_j = 0 | \theta\right) + p\left(\beta_j | \gamma_j = 1, s_1\right)p\left(\gamma_j = 1 | \theta\right)}
\end{aligned} \tag{9}$$

where $p\left(\gamma_j = 1 | \theta\right) = \theta$, $p\left(\gamma_j = 0 | \theta\right) = 1 - \theta$, $p\left(\beta_j | \gamma_j = 1, s_1\right) = DE\left(\beta_j | 0, s_1\right)$, and $p\left(\beta_j | \gamma_j = 0, s_0\right) = DE\left(\beta_j | 0, s_0\right)$. Therefore,

Table 2 Estimates of four measures over 50 replicates under different simulated scenarios

		Deviance	MSE	AUC	Misclassification
Scenario 1	$n = 500$, m $= 1000$				
	Lasso	638.983(17.431)	0.224(0.008)	0.693(0.031)	0.359(0.029)
	sslasso: $s_0 = 0.01$	685.229(18.922)	0.246(0.009)	0.595(0.067)	0.442(0.051)
	sslasso: $s_0 = 0.02$	643.119(23.694)	0.226(0.011)	0.681(0.038)	0.371(0.031)
	sslasso: $s_0 = 0.03$	637.226(22.197)	0.223(0.010)	0.692(0.033)	0.362(0.029)
	sslasso: $s_0 = 0.04$	636.930(18.928)[a]	0.223(0.009)	0.692(0.029)	0.364(0.026)
	sslasso: $s_0 = 0.05$	639.385(16.732)	0.224(0.008)	0.694(0.029)	0.364(0.022)
	sslasso: $s_0 = 0.06$	639.784(17.359)	0.224(0.008)	0.684(0.029)	0.369(0.024)
	sslasso: $s_0 = 0.07$	645.151(19.752)	0.227(0.009)	0.674(0.028)	0.372(0.025)
Scenario 2	$n = 500$, $m = 1000$				
	Lasso	601.872(15.666)	0.207(0.007)	0.753(0.025)	0.320(0.022)
	sslasso: $s_0 = 0.01$	640.816(26.885)	0.224(0.012)	0.686(0.042)	0.367(0.031)
	sslasso: $s_0 = 0.02$	581.940(24.945)	0.199(0.012)	0.761(0.033)	0.308(0.028)
	sslasso: $s_0 = 0.03$	581.661(28.271)[a]	0.198(0.010)	0.765(0.028)	0.306(0.023)
	sslasso: $s_0 = 0.04$	583.037(21.964)	0.199(0.009)	0.764(0.026)	0.307(0.021)
	sslasso: $s_0 = 0.05$	590.185(19.343)	0.202(0.008)	0.755(0.023)	0.314(0.018)
	sslasso: $s_0 = 0.06$	595.879(19.388)	0.204(0.008)	0.751(0.024)	0.328(0.018)
	sslasso: $s_0 = 0.07$	603.756(20.020)	0.208(0.008)	0.738(0.024)	0.333(0.020)
Scenario 3	$n = 500$, $m = 1000$				
	Lasso	561.917(14.623)	0.190(0.006)	0.790(0.021)	0.289(0.020)
	sslasso: $s_0 = 0.01$	585.600(34.703)	0.201(0.014)	0.759(0.035)	0.318(0.030)
	sslasso: $s_0 = 0.02$	531.956(26.214)	0.180(0.010)	0.808(0.021)	0.269(0.022)
	sslasso: $s_0 = 0.03$	532.747(26.343)	0.179(0.010)	0.808(0.021)	0.271(0.022)
	sslasso: $s_0 = 0.04$	530.781(24.638)[a]	0.179(0.010)	0.809(0.020)	0.274(0.020)
	sslasso: $s_0 = 0.05$	541.192(24.496)	0.182(0.010)	0.802(0.020)	0.279(0.019)
	sslasso: $s_0 = 0.06$	550.971(25.065)	0.186(0.010)	0.794(0.020)	0.284(0.019)
	sslasso: $s_0 = 0.07$	559.430(24.311)	0.190(0.009)	0.785(0.020)	0.293(0.019)
Scenario 4	$n = 500$, $m = 3000$				
	Lasso	655.349(11.253)	0.232(0.005)	0.665(0.028)	0.382(0.024)
	sslasso: $s_0 = 0.01$	680.988(16.432)	0.244(0.008)	0.601(0.058)	0.430(0.119)
	sslasso: $s_0 = 0.02$	655.714(23.241)	0.231(0.010)	0.663(0.034)	0.385(0.027)
	sslasso: $s_0 = 0.03$	646.877(20.963)	0.228(0.009)	0.673(0.030)	0.372(0.026)
	sslasso: $s_0 = 0.04$	645.278(16.039)[a]	0.227(0.007)	0.674(0.024)	0.377(0.022)
	sslasso: $s_0 = 0.05$	654.349(16.241)	0.231(0.007)	0.659(0.027)	0.390(0.023)
	sslasso: $s_0 = 0.06$	665.488(18.227)	0.236(0.008)	0.646(0.028)	0.400(0.026)
	sslasso: $s_0 = 0.07$	675.374(20.660)	0.241(0.009)	0.634(0.028)	0.404(0.026)
Scenario 5	$n = 500$, $m = 3000$				
	Lasso	620.034(16.209)	0.215(0.007)	0.726(0.030)	0.334(0.027)
	sslasso: $s_0 = 0.01$	642.083(30.947)	0.225(0.014)	0.683(0.056)	0.363(0.045)
	sslasso: $s_0 = 0.02$	597.547(34.288)	0.205(0.015)	0.745(0.039)	0.322(0.033)
	sslasso: $s_0 = 0.03$	593.701(32.304)[a]	0.205(0.013)	0.746(0.034)	0.318(0.030)
	sslasso: $s_0 = 0.04$	596.421(30.006)	0.205(0.012)	0.746(0.032)	0.324(0.029)
	sslasso: $s_0 = 0.05$	610.549(24.024)	0.211(0.010)	0.731(0.030)	0.333(0.025)
	sslasso: $s_0 = 0.06$	623.014(24.530)	0.217(0.011)	0.715(0.032)	0.347(0.028)
	sslasso: $s_0 = 0.07$	634.536(26.023)	0.222(0.011)	0.701(0.033)	0.355(0.026)
Scenario 6	$n = 500$, $m = 3000$				
	Lasso	570.138(17.989)	0.193(0.008)	0.791(0.026)	0.289(0.026)
	sslasso: $s_0 = 0.01$	568.332(35.346)	0.194(0.015)	0.777(0.036)	0.302(0.029)
	sslasso: $s_0 = 0.02$	537.665(28.103)	0.180(0.011)	0.806(0.025)	0.275(0.028)
	sslasso: $s_0 = 0.03$	530.081(29.097)[a]	0.178(0.012)	0.812(0.025)	0.271(0.027)
	sslasso: $s_0 = 0.04$	530.535(26.149)	0.178(0.011)	0.811(0.023)	0.266(0.025)
	sslasso: $s_0 = 0.05$	542.091(26.825)	0.184(0.011)	0.801(0.024)	0.275(0.024)
	sslasso: $s_0 = 0.06$	557.014(27.697)	0.189(0.011)	0.788(0.025)	0.288(0.022)
	sslasso: $s_0 = 0.07$	572.405(28.018)	0.195(0.011)	0.776(0.025)	0.299(0.024)

Values in parentheses are SE. The slab scales, s_1, are 1 in all scenarios.
[a]The smallest deviance values indicate the optimal model.

the conditional posterior expectation of S_j^{-1} can be obtained by

$$E\left(S_j^{-1}\big|\beta_j\right) = E\left(\frac{1}{(1-\gamma_j)s_0 + \gamma_j s_1}\bigg|\beta_j\right) = \frac{1-p_j}{s_0} + \frac{p_j}{s_1} \quad (10)$$

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

For the M-step, we update (β, ϕ, θ) by maximizing the posterior expectation of the log joint posterior density with γ_j and S_j^{-1} replaced by their conditional posterior
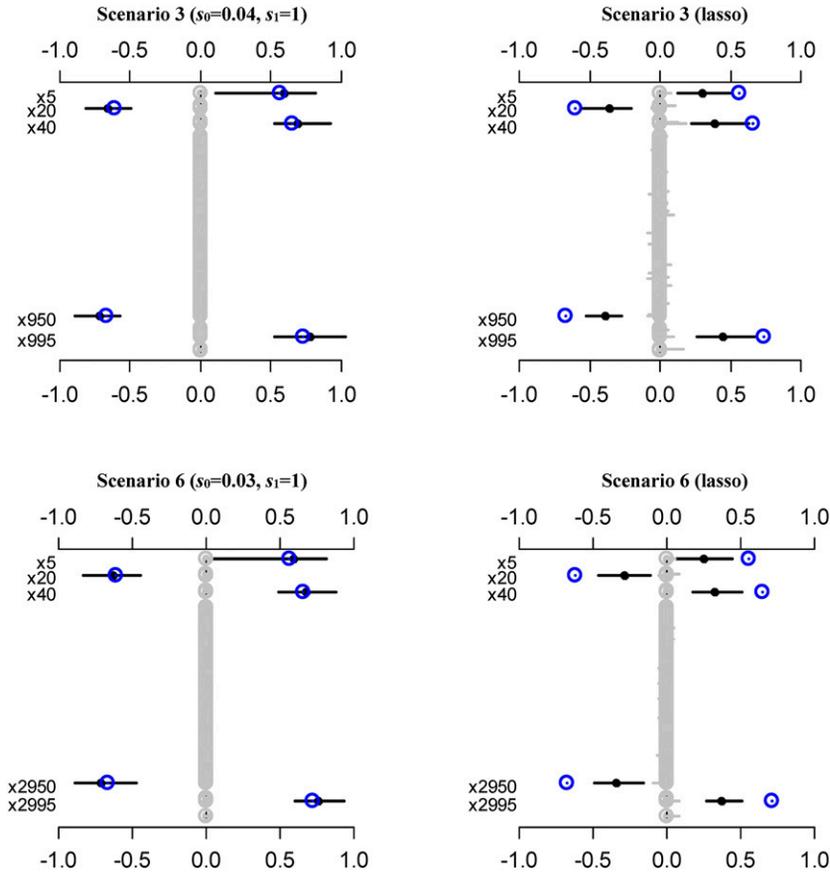
Figure 3 The parameter estimation averaged over 50 replicates for sslasso GLMs and the lasso model under scenarios 3 and 6. The blue circles are the assumed true values. The black points and lines represent the estimated values and the interval estimates of coefficients over 50 replicates.

expectations. From the log joint posterior density, we can see that (β, ϕ) and θ can be updated separately, because the parameters (β, ϕ) are only involved in $l(\beta, \phi) - \sum_{j=1}^{J} S_j^{-1} |\beta_j|$, and the probability parameter θ is only in $\sum_{j=1}^{J} \left[\gamma_j \log\theta + (1 - \gamma_j)\log(1 - \theta) \right]$. Therefore, the parameters (β, ϕ) are updated by maximizing the expression:

$$Q_1(\beta, \phi) = l(\beta, \phi) - \sum_{j=1}^{J} \frac{1}{S_j} |\beta_j| \tag{11}$$

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

$$Q_2(\theta) = \sum_{j=1}^{J} [\gamma_j \log\theta + (1 - \gamma_j)\log(1 - \theta)] \tag{12}$$

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

In summary, the EM coordinate decent algorithm for fitting the sslasso Cox models proceeds as follows:

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

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

1. Update (β, ϕ) using the cyclic coordinate decent algorithm;
2. Update θ.

We assess convergence by the criterion: $\left| d^{(t)} - d^{(t-1)} \right| \Big/ \left(0.1 + \left| d^{(t)} \right| \right) < \varepsilon$, where $d^{(t)} = -2\log l(\beta^{(t)}, \phi^{(t)})$ is the estimate of deviance at the tth iteration, and ε is a small value (say 10^{-5}).

Selecting optimal scale values

The performance of the sslasso approach can depend on the scale parameters (s_0, s_1). Rather than restricting attention to a single model, 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 (e.g., $s_1 = 1$), and consider a sequence of L decreasing values $\{s_0^l\}$: $0 < s_0^1 < s_0^2 < \cdots < s_0^L < s_1$, for the spike
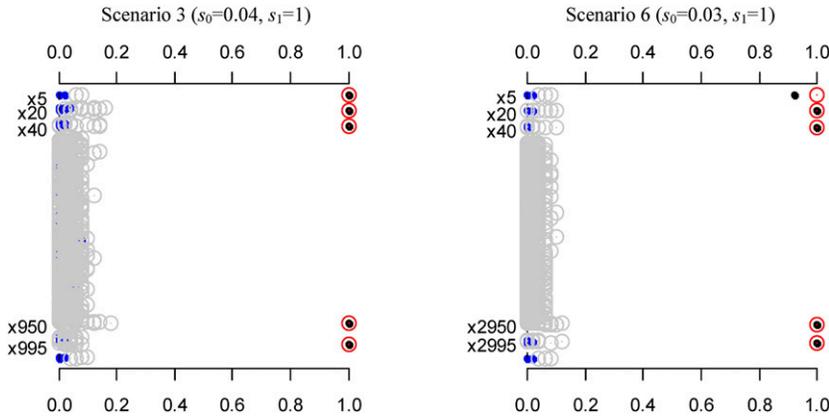
Figure 4 The inclusion proportions of the nonzero and zero coefficients in the model over 50 simulation replicates under scenarios 3 and 6. The black points and red circles represent the proportions of nonzero coefficients for sslasso GLMs and the lasso model, respectively. The blue points and gray circles represent the proportions of zero coefficients for sslasso GLMs and the lasso model, respectively.

scale s_0. We then fit L models with scales $\{(s_0^l, s_1); l = 1, \cdots, L\}$ Increasing the spike scale s_0 tends to include more nonzero coefficients in the model. This procedure is similar to the lasso implemented in the widely used R package glmnet, which quickly fits the lasso model over a grid of values of λ covering its entire range, giving a sequence of models for users to choose from (Friedman *et al.* 2010; Hastie *et al.* 2015).

Evaluation of predictive performance

There are several measures to evaluate the performance of a fitted GLM (Steyerberg 2009), including: (1) deviance, which is a generic way of measuring the quality of any model, and is defined as: $d = -2\sum_{i=1}^{n}\log p(y_i|X_i\hat{\beta}, \hat{\phi})$. Deviance measures the overall quality of a fitted GLM, and thus is usually used to choose an optimal model. (2) Mean squared error (MSE), defined as $\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$; (3) For logistic regression, we can use two additional measures: AUC (area under the ROC curve) and misclassification, which is defined as: $\frac{1}{n}\sum_{i=1}^{n}I(|y_i - \hat{y}_i| > 0.5)$, where $I(|y_i - \hat{y}_i| > 0.5) = 1$ if $|y_i - \hat{y}_i| > 0.5$, and $I(|y_i - \hat{y}_i| > 0.5) = 0$ if $|y_i - \hat{y}_i| \leq 0.5$. (4) Prevalidated linear predictor analysis (Tibshirani and Efron 2002; Hastie *et al.* 2015); for all GLMs, we can use the cross-validated linear predictor $\hat{\eta}_i = X_i\hat{\beta}$ as a continuous covariate in a univariate GLM to predict the outcome: $E(y_i|\hat{\eta}_i) = h^{-1}(\mu + \hat{\eta}_i b)$. We then look at the P-value for testing the hypothesis $b = 0$ or other statistics to evaluate the predictive performance. We also can transform the continuous cross-validated linear predictor $\hat{\eta}_i$ into a categorical factor $c_i = (c_{i1}, \cdots, c_{i6})$ based on the quantiles of $\hat{\eta}_i$, for example, 5, 25, 50, 75, and 95% quantiles, and then fit the model: $E(y_i|c_i) = h^{-1}(\mu + \sum_{k=2}^{6} c_{ik}b_k)$. This allows us to compare statistical significance and prediction between different categories.

To evaluate the predictive performance of the proposed model, a general way is to fit the model using a data set, and then calculate the above measures with independent data. We use the prevalidation method, a variant of cross-validation (Tibshirani and Efron 2002; Hastie *et al.* 2015), by randomly splitting the data to K subsets of roughly the same size, and using $(K-1)$ 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 linear predictor $\hat{\eta}_{(k)} = X_{(k)}\hat{\beta}^{(-k)}$ for all individuals in the k-th subset of the data, called the cross-validated or prevalidated predicted index. Cycling through K parts, we obtain the cross-validated linear predictor $\hat{\eta}_i$ for all individuals. We then use $(y_i, \hat{\eta}_i)$ to compute the measures described above. The cross-validated linear predictor for each patient is derived independently of the observed response of the patient, and hence the "prevalidated" dataset $\{y_i, \hat{\eta}_i\}$ can essentially be treated as a "new dataset." Therefore, this procedure provides valid assessment of the predictive performance of the model (Tibshirani and Efron 2002; Hastie *et al.* 2015). To get stable results, we run $10\times$ 10-fold cross-validation for real data analysis.

Data availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

Implementation

We created an R function, bmlasso, for setting up and fitting sslasso GLMs, 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 incorporated these functions into the freely available R package BhGLM (http://www.ssg.uab.edu/bhglm/).

Simulation study

Simulation design: We used simulations to validate the proposed sslasso approach, and to compare with the lasso in the R package glmnet. Although the proposed method can be applied to any GLMs, we focused on sslasso logistic regression, because we analyzed binary outcomes in our real data sets [see *Simulation results*]. In each situation, we simulated two data sets, 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

Table 3 Average number of nonzero coefficients and MAE of coefficient estimates over 50 simulation replicates

Simulation Scenarios	sslasso ($s_1 = 1$)		Lasso	
	Average Number	MAE	Average Number	MAE
Scenario 1, $s_0 = 0.04$	15.800	1.431(0.420)	30.800	2.150(0.726)
Scenario 2, $s_0 = 0.03$	5.780	0.750(0.452)	30.540	2.255(0.667)
Scenario 3, $s_0 = 0.04$	10.420	0.669(0.286)	39.540	2.680(0.729)
Scenario 4, $s_0 = 0.04$	32.900	1.984(0.352)	32.960	2.316(0.770)
Scenario 5, $s_0 = 0.03$	6.460	0.906(0.498)	42.260	2.783(0.908)
Scenario 6, $s_0 = 0.03$	5.920	0.629(0.362)	43.060	2.865(0.842)

Values in parentheses are SE.

predictive values including deviance, MSE, AUC, and misclassification in the test data, and 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 binary response and a vector of m ($= 1000$, 3000) continuous predictors $X_i = (x_{i1}, \cdots, x_{im})$. The vector X_i was generated with 50 elements at a time, $i.e.$, the subvector $(x_{i(50k+1)}, \cdots, x_{i(50k+50)})$, $k = 0, 1, \cdots$, was sampled randomly from multivariate normal distribution $N_{50}(0, \sum)$, where $\sum = (\sigma_{jj'})$ with $\sigma_{jj} = 1$ and $\sigma_{jj'} = 0.6 (j \neq j')$. Thus, the predictors within a group were correlated, and between groups were independent. To simulate a binary response, we first generated Gaussian response z_i from univariate normal distribution $N(\eta_i, 1.6^2)$, where $\eta_i = \beta_0 + \sum_{j=1}^{m} x_{ij}\beta_j$, and then transformed the continuous response to binary data by setting individuals with the 30% largest continuous response Z as "affected" ($y_i = 1$), and the other individuals as "unaffected" ($y_i = 0$). We set five coefficients β_5, β_{20}, β_{40}, β_{m-50}, and β_{m-5} to be nonzero, two of which are negative, and all others to be zero. Table 1 shows the preset nonzero coefficient values for six simulation scenarios.

We analyzed each simulated data set using the lasso logistic model implemented in the R package glmnet, and the proposed sslasso logistic regression in our R package BhGLM. For the lasso approach, we used 10-fold cross-validation to select an optimal value of λ, which determines an optimal lasso model, and reported the results based on the optimal lasso model. For the proposed sslasso GLM approach, we fixed slab scale as $s_1 = 1$, and run a grid value of spike scales: $s_0 = 0.01, 0.02, \ldots, 0.07$.

To fully investigate the impact of the scales (s_0, s_1) on the results, we also fitted the simulated data sets under scenarios 3 and 6 (see Table 1) with 100 combinations when s_0 changed from 0.01 to 0.10, and s_1 from 0.7 to 2.5. We analyzed each scale (s_0, s_1) combination over 50 replicates. We also presented the solution path under simulation scenarios 3 and 6, averaged over 50 replicates, to show the special characteristic of sslasso GLMs.

Simulation results: *Impact of scales (s_0, s_1) and solution path:* We analyzed the data with a grid of values of (s_0, s_1) covering their entire range to fully investigate the impact of the prior scales. Figure 1 show the profiles of the deviance

under scenarios 3 and 6. It can be seen that the slab scale s_1 within the range [0.7, 2.5] had little influence on the deviance, while the spike scale s_0 strongly affected model performance. These results show that our approach with a fixed slab scale s_1 ($e.g.$, $s_1 = 1$) is reasonable.

Figure 2A and Figure 2B present the solution path for scenario 3 by the proposed model and the lasso model, respectively. Figure 2, C and D show the profiles of deviance by $10 \times$ 10-fold cross-validation for the proposed model and the lasso model, respectively. For the proposed sslasso GLMs, the minimum of deviance was achieved by including 10.7 nonzero coefficients (averaged over 50 replicates), when the s_0 scale was 0.04. For the lasso model, the optimal model included 29.8 nonzero coefficients when the λ value was 0.035. Similar to the lasso, the sslasso GLM is a path-following strategy for fast dynamic posterior exploration. However, its solution path is essentially different from that of the lasso model. For the lasso solution, as shown in Figure 2B, the number of nonzero coefficients could be a few, even zero if a strong penalty was adopted. However, a spike-and-slab mixture prior can help larger coefficients escape the gravitational pull of the spike. Larger coefficients will be always included in the model with none or weak shrinkage. Supplemental Material, Figure S1 shows the solution path of the proposed model and the lasso model under scenario 6. Figure S2 shows the adaptive shrinkage amount, along with the different effect size. The same conclusion could be reached, that the spike-and slab prior shows self-adaptive and flexible characteristics.

Predictive performance: Table 2 shows the deviance, MSE, AUC, and misclassification in the test data under different simulated scenarios. The smallest deviance is highlighted in Table 2. As described earlier, the deviance measures the overall quality of a model, and thus is usually used to choose an optimal model. From these results, we can see that the sslasso GLMs with an appropriate value of s_0 performed better than the lasso. Table 2 also shows that the optimal spike-and-slab GLMs usually had higher AUC value than the lasso. The AUC measures the discriminative ability of a logistic model (Steyerberg 2009). Thus, the sslasso GLMs generated better discrimination.

Accuracy of parameter estimates: Figure 3, Figure S3, and Figure S4 show the estimates of coefficients from the best sslasso GLMs, and the lasso model over 50 replicates of

Table 4 Measures of optimal sslasso and lasso models for Dutch breast cancer dataset and TCGA ovarian cancer dataset by 10× 10-fold cross validation

		Deviance	MSE	AUC	Misclassification
Dutch breast cancer	sslasso	336.880(4.668)	0.192(0.003)	0.684(0.011)	0.284(0.014)
	lasso	342.134(3.383)	0.197(0.003)	0.656(0.018)	0.297(0.007)
TCGA ovarian cancer	sslasso	394.796(7.684)	0.179(0.004)	0.647(0.017)	0.258(0.010)
	lasso	393.152(5.392)	0.179(0.003)	0.636(0.017)	0.254(0.005)

Values in parentheses are SE.

testing data. It can be seen that the sslasso GLMs provided more accurate estimation in most situations, especially for larger coefficients. This result is expected, because the spike-and-slab prior can induce weak shrinkage on larger coefficients. In contrast, the lasso employs a single penalty on all the coefficients, and thus can overshrink large coefficients. As a result, five nonzero coefficients were shrunk and underestimated compared to true values in our simulation.

Proportions of coefficients included in the model: We calculated the proportions of the coefficients included in the model over the simulation replicates. Like the lasso model, the proposed sslasso GLMs can estimate coefficients to be zero, and thus can easily return these proportions. Figure 4 and Figure S5 show the inclusion proportions of the nonzero coefficients and the zero coefficients for the best sslasso GLMs and the lasso model. It can be seen that the inclusion proportions of the nonzero coefficients were similar for the two approaches in most situations. However, the lasso included zero coefficients in the model more frequently than the sslasso GLMs. This indicates that the sslasso approach can reduce noisy signals.

We summarized the average numbers of nonzero coefficients and the MAE of coefficient estimates, defined as $\text{MAE} = \sum \left| \hat{\beta}_j - \beta_j \right| / m$, in Table 3. In most simulated scenarios, the average numbers of nonzero coefficients in the sslasso GLMs were much lower than those in the lasso model. We also found that the average numbers of nonzero coefficients detected by the proposed models were close to the number of the simulated nonzero coefficients in most scenarios. However, the lasso usually included many zero coefficients in the model. This suggests that the noises can be controlled by the spike-and-slab prior.

Application to real data

Dutch breast cancer data: We applied our sslasso GLMs to analyze a well-known Dutch breast cancer data set, and compared the results with that of the lasso in the R package glmnet. This data were described in van de Vijver *et al.* (2002), and is publicly available in the R package "breastCancerNKI" (https://cran.r-project.org). This data set contains the microarray mRNA expression measurements of 4919 genes, and the event of metastasis after adjuvant systemic therapy from 295 women with breast cancer (van't Veer *et al.* 2002; van de Vijver *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 logistic model for predicting the metastasis event using the 4919 gene-expression predictors. Prior to fitting the models, we standardized all the predictors. For hierarchical (and also penalized) models, it is important to use a roughly common scale for all predictors.

We fixed the slab scale s_1 to 1, and varied the spike scale s_0 over the grid of values: $0.005 + k \times 0.005$; $k = 0, 1, \cdots,$ 39, leading to 40 models. We performed 10× 10-fold cross-validation to select an optimal model based on the deviance. Figure 5A shows the profile of the prevalidated deviance for the Dutch breast cancer dataset. The minimum value of deviance appears to be 336.880(4.668), when the spike scale s_0 is 0.09. Therefore, the sslasso GLM with the prior scale (0.09, 1) was chosen for model fitting and prediction. We further used 10-fold cross-validation over 10 replicates to evaluate the predictive values of the chosen model with the prior scale (0.09, 1). As a comparison, we fitted the model by the lasso approach, and also performed 10-fold cross-validation over 10 replicates. Table 4 summarizes the measures of performance. The result of the proposed sslasso GLMs was better than the result of the lasso. The cross-validated AUC by the proposed method was estimated to be 0.684(0.011), which is significantly larger than 0.5, showing the discriminative ability of the prognostic model. Totally, 51 genes were detected, and the effect sizes for most of these genes were small (Figure S6).

We further estimated the prevalidated linear predictor, $\eta_i = X_i \hat{\beta}$, for each patient, and then grouped the patients on the basis of the prevalidated linear predictor into categorical factor according to 5, 25, 50, 75, and 95th percentiles, denoted by $c_i = (c_{i1}, \cdots, c_{i6})$. We fitted the univariate model $E(y_i | \hat{\eta}_i) = h^{-1}(\mu + \hat{\eta}_i b)$ and multivariate model $E(y_i | c_i) = h^{-1}(\mu + \sum_{k=2}^{6} c_{ik} b_k)$ by using the prevalidated linear predictor and the categorical factors, respectively. The results are summarized in Table 5. As expected, the two models were significant, indicating that the proposed prediction model was very informative.

TCGA ovarian cancer (OV) dataset: The second data set that we analyzed was microarray mRNA expression data for ovarian cancer (OV) downloaded from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/) (update March 2016). The outcome of interest is a tumor event within 2 years (started the date of initial pathologic diagnosis), including progression, recurrence, and new primary malignancies. Clinical
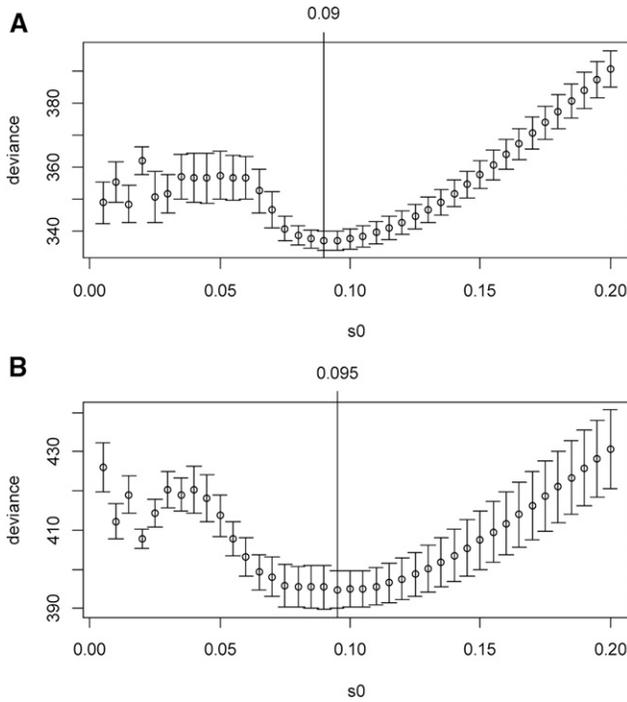
Figure 5 The profiles of prevalidated deviance under varied prior scales s_0 and fixed $s_1 = 1$ for Dutch breast cancer dataset (A) and TCGA ovarian cancer dataset (B).

Table 5 Univariate and multivariate analyses using the prevalidated linear predictors and their categorical factors

	Model	Coefficients	Estimates	SE	P Values
Dutch breast cancer	univariate	b	0.816	0.162	4.76e−07
	multivariate	b_2	−0.619	0.761	0.416
		b_3	0.251	0.700	0.720
		b_4	0.412	0.697	0.555
		b_5	1.556	0.696	0.025
		b_6	1.520	0.827	0.066
TCGA ovarian cancer	univariate	b	0.683	0.162	2.47e−05
	multivariate	b_2	0.347	0.519	0.504
		b_3	0.906	0.518	0.080
		b_4	1.426	0.536	0.008
		b_5	1.719	0.572	0.003
		b_6	2.035	0.877	0.020

data are available for 513 OV patients, where 110 tumors have no interested outcome records. Microarray mRNA expression data (Agilent Technologies platform) includes 551 tumors that have 17,785 feature profiles after removing duplications. Even though we can analyze all these features, considering that the genes with small variance might contribute little to the outcome, we selected the top 30% genes filtered by variance for predictive modeling. We merged the expression data and the tumor with new tumor event. After removing individuals with missing response, totally, 362 tumors with 5336 genes were included in our analysis.

Similar to the above analysis of the Dutch breast cancer dataset, we fixed the slab scale s_1 to 1 and considered the spike scale s_0 over the grid of values: $0.005 + k \times 0.005$; $k = 0, 1, \cdots, 39$. We performed 10-fold cross-validation with 10 replicates to select an optimal model based on the prevalidated deviance. Figure 5B shows the profile of the prevalidated deviance. The minimum value of deviance appears to be 394.796(7.684) when the prior $s_0 = 0.095$. Therefore, the scale (0.095, 1) was chosen for the model fitting and prediction. We also fitted the optimal lasso model using 10-fold cross-validation over 10 replicates. Table 4 summarizes the measures of performance of the proposed sslasso GLMs and the lasso model. We can see that the sslasso was slightly better than the lasso model. The cross-validated AUC was estimated to be 0.647(0.017), which is significantly larger than 0.5, showing the discriminative ability of the prognostic model. Totally, 85 genes were detected, and the effect sizes for most of these genes were small (Figure S7).

We further estimated the cross-validated linear predictor, and performed similar analysis as above Dutch breast cancer dataset. The results are summarized in Table 5. The univariate and multivariate analyses suggested that the proposed prediction model was very informative for predicting new tumor events in ovarian tumors.

Discussion

Omics technologies allow researchers to produce tons of molecular data about cancer and other diseases. These multilevel molecular, and also environmental, data provide an extraordinary opportunity to predict the likelihood of clinical benefit for treatment options, and to promote more accurate and individualized health predictions (Chin *et al.* 2011; Collins and Varmus 2015). However, there are huge challenges in making predictions and detecting associated biomarkers from large-scale molecular datasets. In this article, we have developed a new hierarchical model approach, *i.e.*, sslasso GLMs, for detecting important variables and prognostic prediction (*e.g.*, events such as tumor recurrence or death). Although focusing on molecular profiling data and binary outcome in our simulations and real data analysis, the proposed approach can also be used for analyzing general large-scale data, and other generalized linear models.

The key to our sslasso GLMs is proposing the new prior distribution, *i.e.*, the mixture spike-and-slab double-exponential prior, on the coefficients. The mixture spike-and-slab prior can induce different amounts of shrinkage for different predictors depending on their effect sizes, and thus have the effect of removing irrelevant predictors, while supporting the larger coefficients, thus improving the accuracy of coefficient estimation, and prognostic prediction. The theoretical property of the spike-and-slab prior is characterized by the solution path. Without the slab component, the output would be equivalent or similar to the lasso solution path. Instead, the solution path of the model with the spike-and-slab prior is different from the lasso. Large coefficients are usually included in the model, while irrelevant coefficients are shrunk to zero.

The large-scale sslasso GLMs 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; Simon *et al.* 2011; Hastie *et al.* 2015), to update the coefficients. As shown in our extensive simulation and real data analyses, the EM coordinate descent algorithm converges rapidly, and is capable of identifying important predictors, and building promising predictive models from numerous candidates.

The sslasso retains the advantages of two popular methods for high-dimensional data analysis (V. Ročková and E. I. George, unpublished results), *i.e.*, Bayesian variable selection (George and McCulloch 1993, 1997; Chipman 1996; Chipman *et al.* 2001; Ročková and George 2014), and the penalized lasso (Tibshirani 1996, 1997; Hastie *et al.* 2015), and bridges these two methods into one unifying framework. Similar to the lasso, the proposed method can shrink many coefficients exactly to zero, 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. This self-adaptive strategy is very much in contrast to the lasso model, which shrinks all estimates equally with a constant penalty. These remarkable properties of the mixture spike-and-slab priors have been theoretically proved previously for normal linear models (V. Ročková and E. I. George, unpublished results), and have also been observed in our methodological derivation and empirical studies.

The performance of the sslasso GLMs depends on the scale parameter of the double-exponential prior. Optimally presetting two scale values (s_0, s_1) can be difficult. 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. We performed simulations by a two-dimensional search, and observed that the slab scale s_1 within the range [0.75, 2.5] has little influence on the fitted model, while the spike scale s_0 can strongly affect model performance. Therefore, we suggest a path-following strategy for fast dynamic posterior exploration, which is similar to the approach of Ročková and George (2014, and unpublished results). We fixed the slab scale s_1 (*e.g.*, $s_1 = 1$), and run grid values of spike scale s_0 from a reasonable range, *e.g.*, (0, 0.1), and then selected an optimal according to cross-validation. Therefore, rather than restricting analysis based on a single value for the scale s_0, the speed of the proposed algorithm makes it feasible to consider several, or tens of, reasonable values for the scale s_0.

The proposed framework is highly extensible. The benefits of hierarchical modeling are flexible, and it is easy to incorporate structural information about the predictors into predictive modeling. With the spike-and-slab priors, the biological information (for example, biological pathways, molecular interactions of genes, genomic annotation, etc.) and multi-level molecular profiling data (*e.g.*, clinical, gene expression, DNA methylation, somatic mutation, etc.) could be effectively integrated. This prior biological knowledge will help improve prognosis and prediction (Barillot *et al.* 2012). Another important extension is to incorporate a polygenic random component for modeling small-effect predictors, or relationships among individuals, into the sslasso GLMs. Zhou *et al.* (2013) proposed a Bayesian sparse linear mixed model that includes numerous predictors with mixture normal priors and a polygenic random effect and developed a MCMC algorithm to fit the model. We will incorporate the idea of Zhou *et al.* (2013) into the framework of our sslasso GLMs to develop faster model-fitting algorithms.
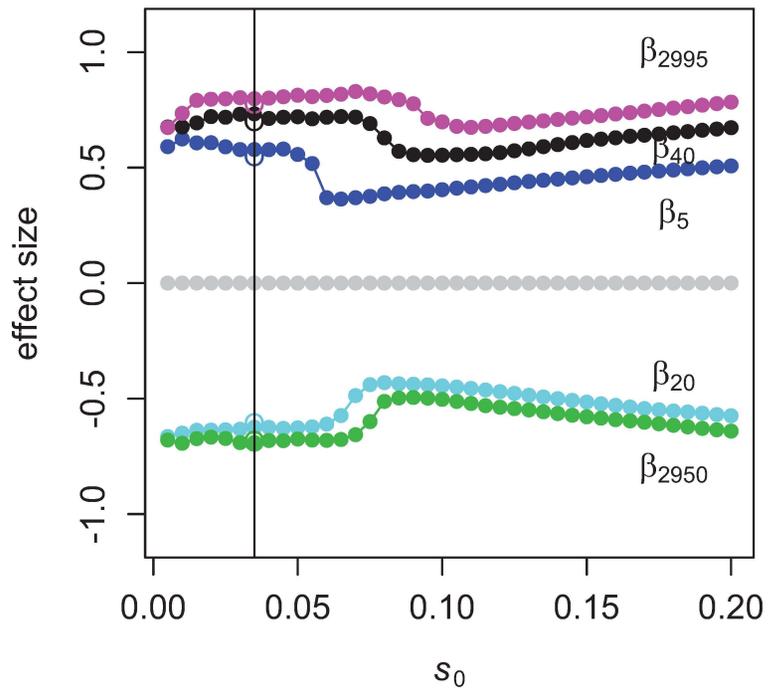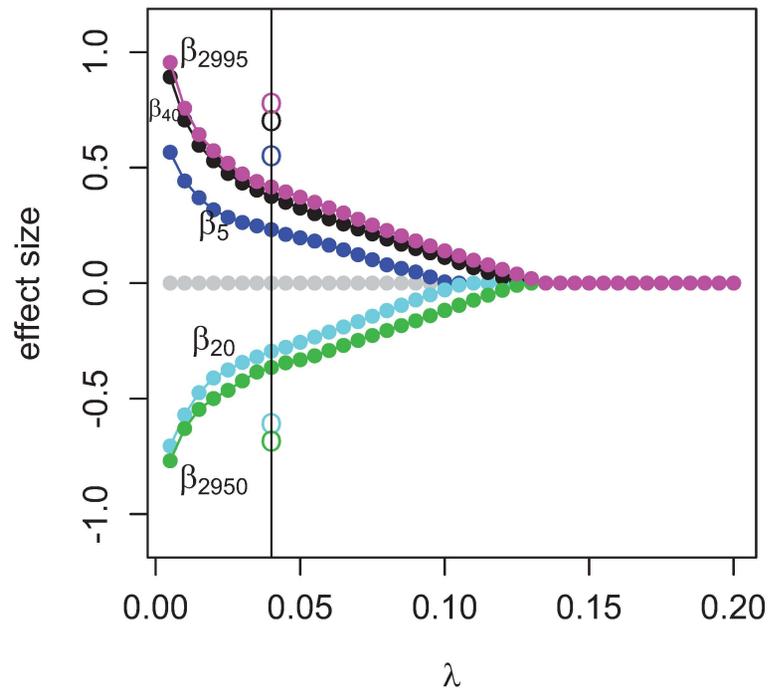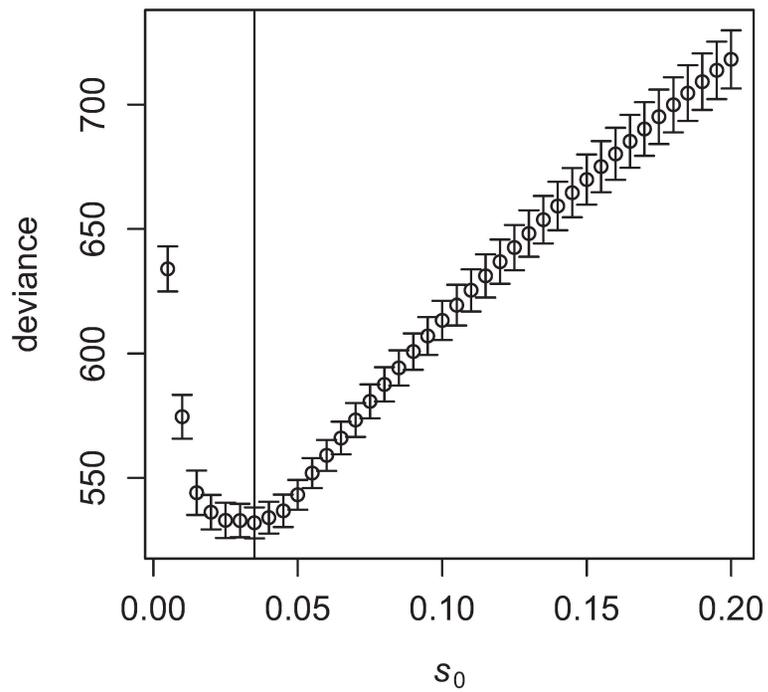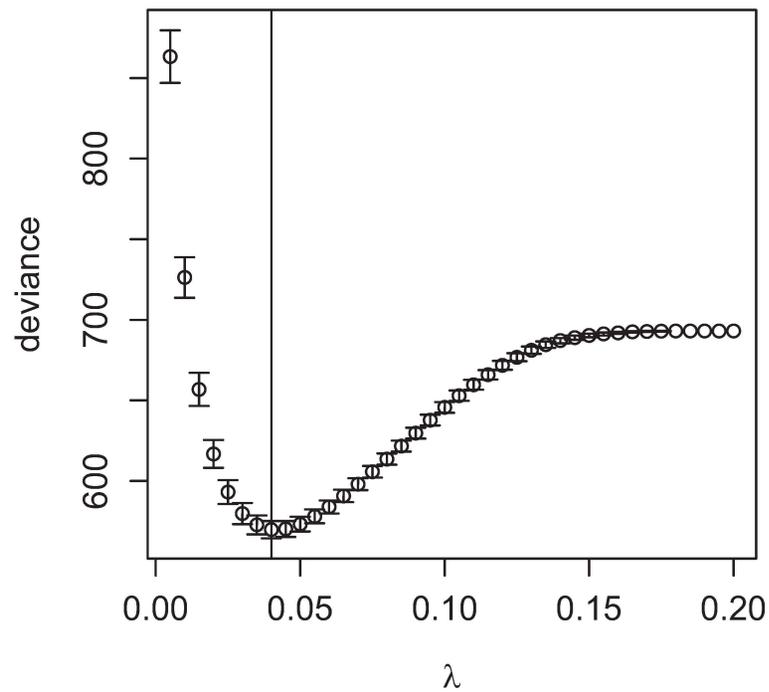
Acknowledgments

Author contributions: N.Y. and Z.T., study conception and design; N.Y., Z.T., X.Z., and Y.S., simulation study and data summary; Z.T., X.Z., and Y.S., real data analysis; N.Y. and Z.T., drafting of manuscript.
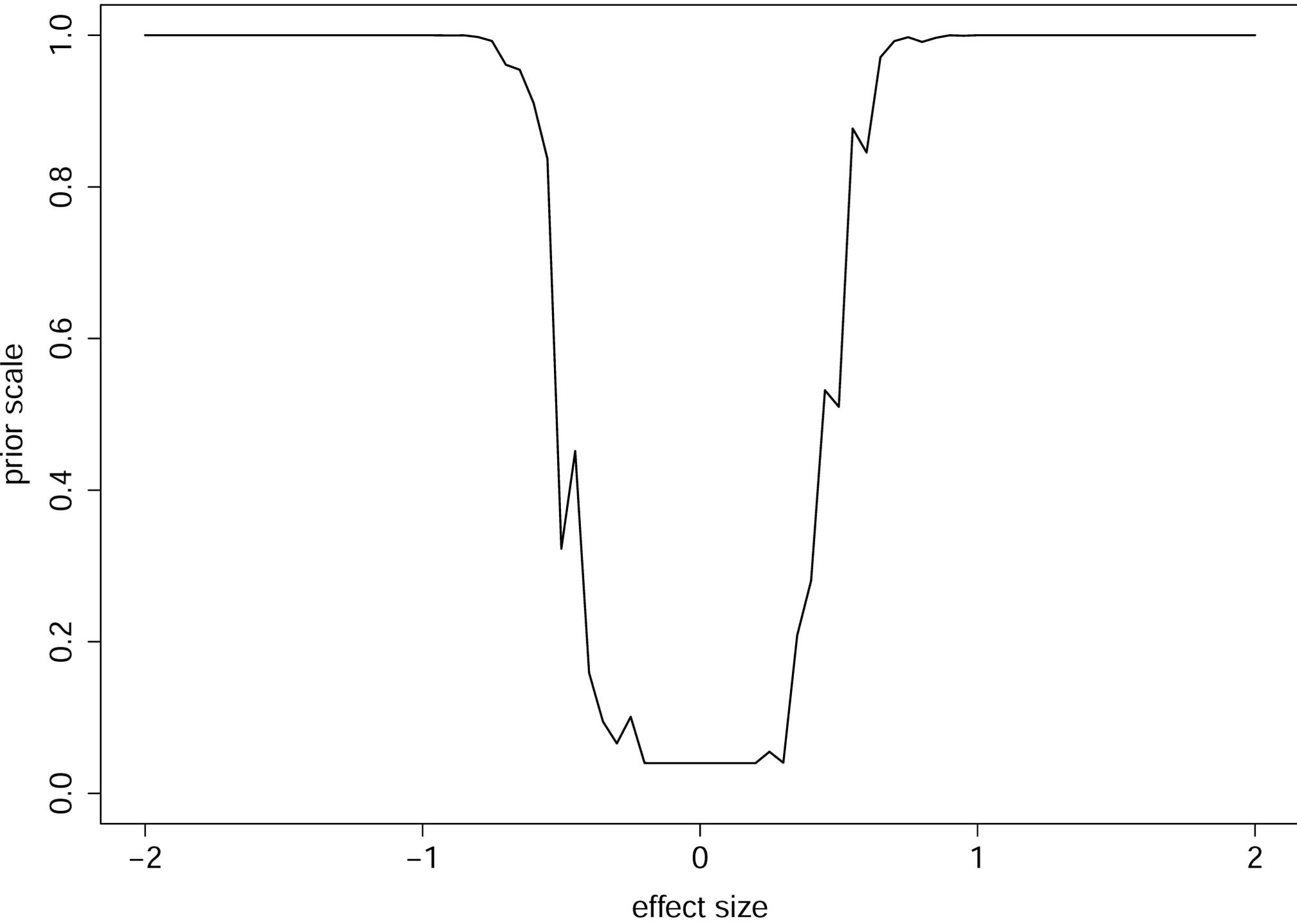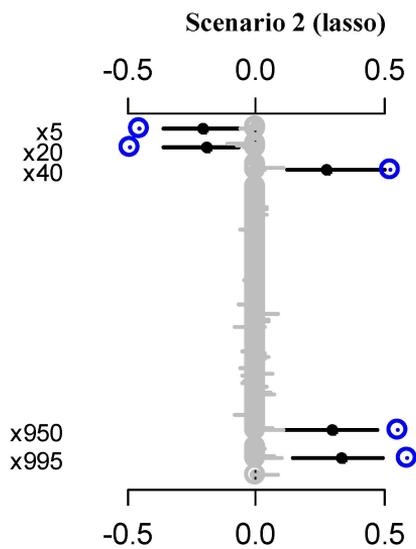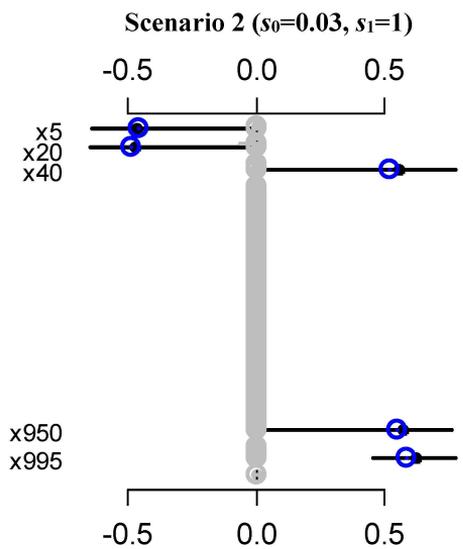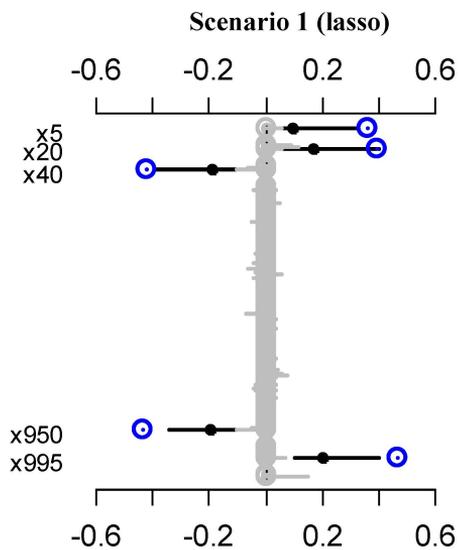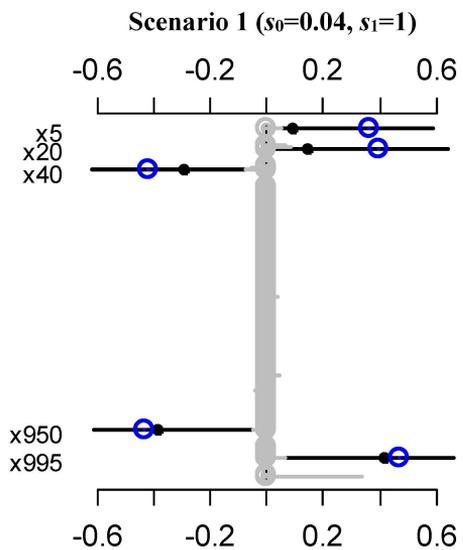
Literature Cited

Barillot, E., L. Calzone, P. Hupe, J. P. Vert, and A. Zinovyev, 2012 *Computational Systems Biology of Cancer*. Chapman & Hall/CRC Mathematical and Computational Biology, Boca Raton, FL.

Bovelstad, H. M., S. Nygard, H. L. Storvold, M. Aldrin, O. Borgan *et al.*, 2007 Predicting survival from microarray data—a comparative study. Bioinformatics 23: 2080–2087.

Bovelstad, H. M., S. Nygard, and O. Borgan, 2009 Survival prediction from clinico-genomic models—a comparative study. BMC Bioinformatics 10: 413.

Chin, L., J. N. Andersen, and P. A. Futreal, 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., E. I. Edwards, and R. E. McCulloch, 2001 The practical implementation of Bayesian model selection, in *Model Selection*, edited by P. Lahiri. Institute of Mathematical Statistics, Beachwood, OH.

Collins, F. S., and H. Varmus, 2015   A new initiative on precision medicine. N. Engl. J. Med. 372: 793–795.

de los Campos, G., D. Gianola, and D. B. Allison, 2010   Predicting genetic predisposition in humans: the promise of whole-genome markers. Nat. Rev. Genet. 11: 880–886.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani, 2004   Least angle regression, pp. 407–451 in *The Annals of Statistics*. Institute of Mathematical Statistics, Beachwood, OH.

Fan, J., and R. Li, 2001   Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. 96: 1348–1360.

Friedman, J., T. Hastie, and R. Tibshirani, 2010   Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33: 1–22.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari *et al.*, 2014   *Bayesian Data Analysis*. Chapman & Hall/CRC Press, New York.

George, E. I., and R. E. McCulloch, 1993   Variable selection via Gibbs sampling. J. Am. Stat. Assoc. 88: 881–889.

George, E. I., and R. E. McCulloch, 1997   Approaches for Bayesian variable selection. Stat. Sin. 7: 339–373.

Hastie, T., R. Tibshirani, and J. Friedman, 2009   *The Elements of Statistical Learning*. Springer-Verlag, New York, NY.

Hastie, T., R. Tibshirani, and M. Wainwright, 2015   *Statistical Learning with Sparsity - The Lasso and Generalization*. CRC Press, New York.

Ishwaran, H., and J. S. Rao, 2005   Spike and slab gene selection for multigroup microarray data. J. Am. Stat. Assoc. 100: 764–780.

Jameson, J. L., and D. L. Longo, 2015   Precision medicine—personalized, problematic, and promising. N. Engl. J. Med. 372: 2229–2234.

Kyung, M., J. Gill, M. Ghosh, and G. Casella, 2010   Penalized regression, standard errors, and Bayesian Lassos. Bayesian Anal. 5: 369–412.

Lee, D., W. Lee, Y. Lee, and Y. Pawitan, 2010   Super-sparse principal component analyses for high-throughput genomic data. BMC Bioinformatics 11: 296.

Lu, Z. H., H. Zhu, R. C. Knickmeyer, P. F. Sullivan, S. N. Williams *et al.*, 2015   Multiple SNP set analysis for genome-wide association studies through Bayesian latent variable selection. Genet. Epidemiol. 39: 664–677.

McCullagh, P., and J. A. Nelder, 1989   *Generalized Linear Models*. Chapman and Hall, London.

Park, T., and G. Casella, 2008   The Bayesian Lasso. J. Am. Stat. Assoc. 103: 681–686.

Partovi Nia, V., and M. Ghannad-Rezaie, 2016   Agglomerative joint clustering of metabolic data with spike at zero: a Bayesian perspective. Biom. J. 58: 387–396.

Rapaport, F., A. Zinovyev, M. Dutreix, E. Barillot, and J.-P. Vert, 2007   Classification of microarray data using gene networks. BMC Bioinformatics 8: 1–15.

Ročková, V., and E. I. George, 2014   EMVS: The EM approach to Bayesian variable selection. J. Am. Stat. Assoc. 109: 828–846.

Shankar, J., S. Szpakowski, N. V. Solis, S. Mounaud, H. Liu *et al.*, 2015   A systematic evaluation of high-dimensional, ensemble-based regression for exploring large model spaces in microbiome analyses. BMC Bioinformatics 16: 31.

Shelton, J. A., A. S. Sheikh, J. Bornschein, P. Sterne, and J. Lucke, 2015   Nonlinear spike-and-slab sparse coding for interpretable image encoding. PLoS One 10: e0124088.

Simon, N., J. Friedman, T. Hastie, and R. Tibshirani, 2011   Regularization paths for Cox's proportional hazards model via coordinate descent. J. Stat. Softw. 39: 1–13.

Sohn, I., and C. O. Sung, 2013   Predictive modeling using a somatic mutational profile in ovarian high grade serous carcinoma. PLoS One 8: e54089.

Steyerberg, E. W., 2009   *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updates*. Springer, New York.

Tibshirani, R., 1996   Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. 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 B. Efron, 2002   Pre-validation and inference in microarrays. Stat. Appl. Genet. Mol. Biol. 1: Article1.

van de Vijver, M. J., Y. D. He, L. J. van't Veer, H. Dai, A. A. Hart *et al.*, 2002   A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 347: 1999–2009.

van't Veer, L. J., H. Dai, M. J. van de Vijver, Y. D. He, A. A. Hart *et al.*, 2002   Gene expression profiling predicts clinical outcome of breast cancer. Nature 415: 530–536.

Yi, N., and S. Xu, 2008   Bayesian LASSO for quantitative trait loci mapping. Genetics 179: 1045–1055.

Yi, N., V. George, and D. B. Allison, 2003   Stochastic search variable selection for mapping multiple quantitative trait loci. Genetics 165: 867–883.

Yuan, Y., E. M. Van Allen, L. Omberg, N. Wagle, A. Amin-Mansour *et al.*, 2014   Assessing the clinical utility of cancer genomic and proteomic data across tumor types. Nat. Biotechnol. 32: 644–652.

Zhang, C. H., 2010   Nearly unbiased variable selection under minimax concave penalty. Ann. Stat. 38: 894–942.

Zhang, W., T. Ota, V. Shridhar, J. Chien, B. Wu *et al.*, 2013   Network-based survival analysis reveals subnetwork signatures for predicting outcomes of ovarian cancer treatment. PLOS Comput. Biol. 9: e1002975.

Zhao, Q., X. Shi, Y. Xie, J. Huang, B. Shia *et al.*, 2014   Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. Brief. Bioinform. 16: 291–303.

Zhou, X., P. Carbonetto, and M. Stephens, 2013   Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet. 9: e1003264.

Zou, H., 2006   The adaptive Lasso and its oracle properties. J. Am. Stat. Assoc. 101: 1418–1429.

Zou, H., and T. Hastie, 2005   Regularization and variable selection via the elastic net. J. R. Stat. Soc. B 67: 301–320.
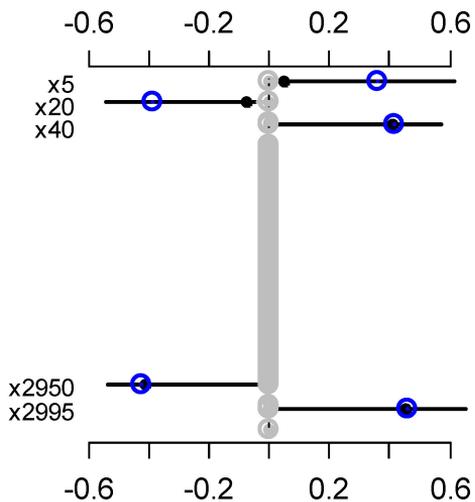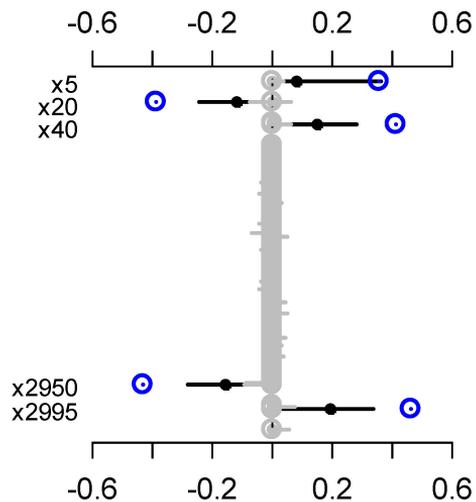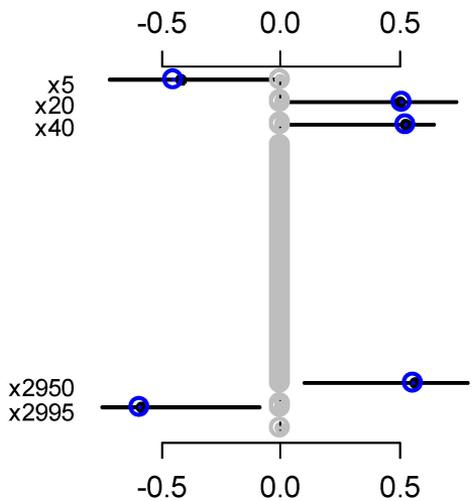
Communicating editor: W. Valdar

(A) effect size vs s_0; **(B)** effect size vs λ; **(C)** deviance vs s_0; **(D)** deviance vs λ.
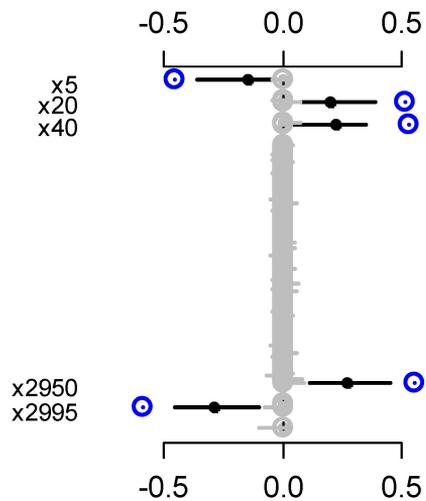
Scenario 1 (s₀=0.04, s₁=1) — Scenario 1 (lasso) — Scenario 2 (s₀=0.03, s₁=1) — Scenario 2 (lasso)

Scenario 4 (s_0=0.04, s_1=1)
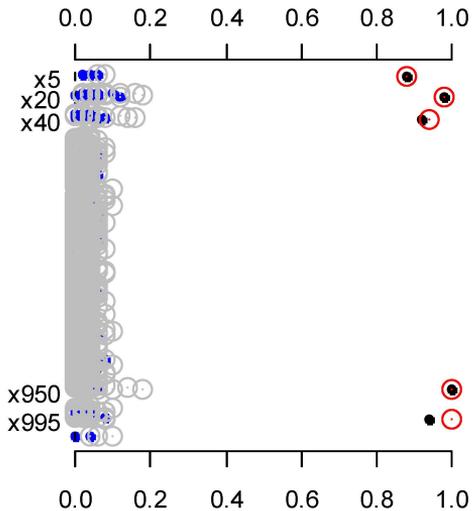
Scenario 4 (lasso)

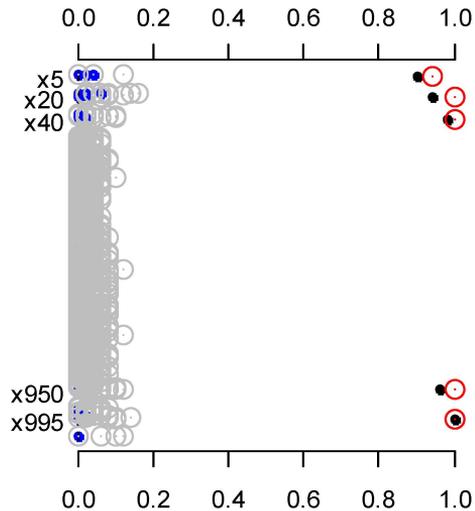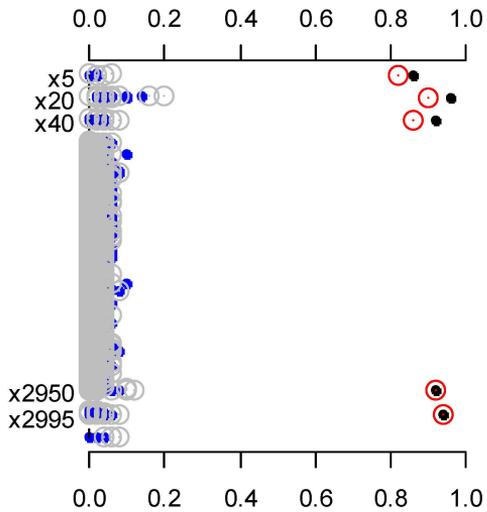**Scenario 5 (s_0=0.03, s_1=1)**

Scenario 5 (lasso)

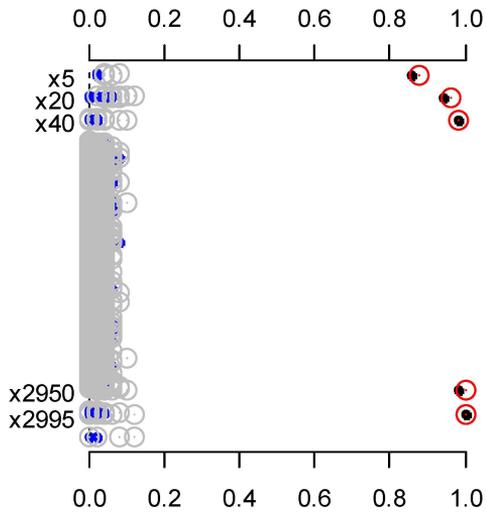Scenario 1 (s_0=0.04, s_1=1)
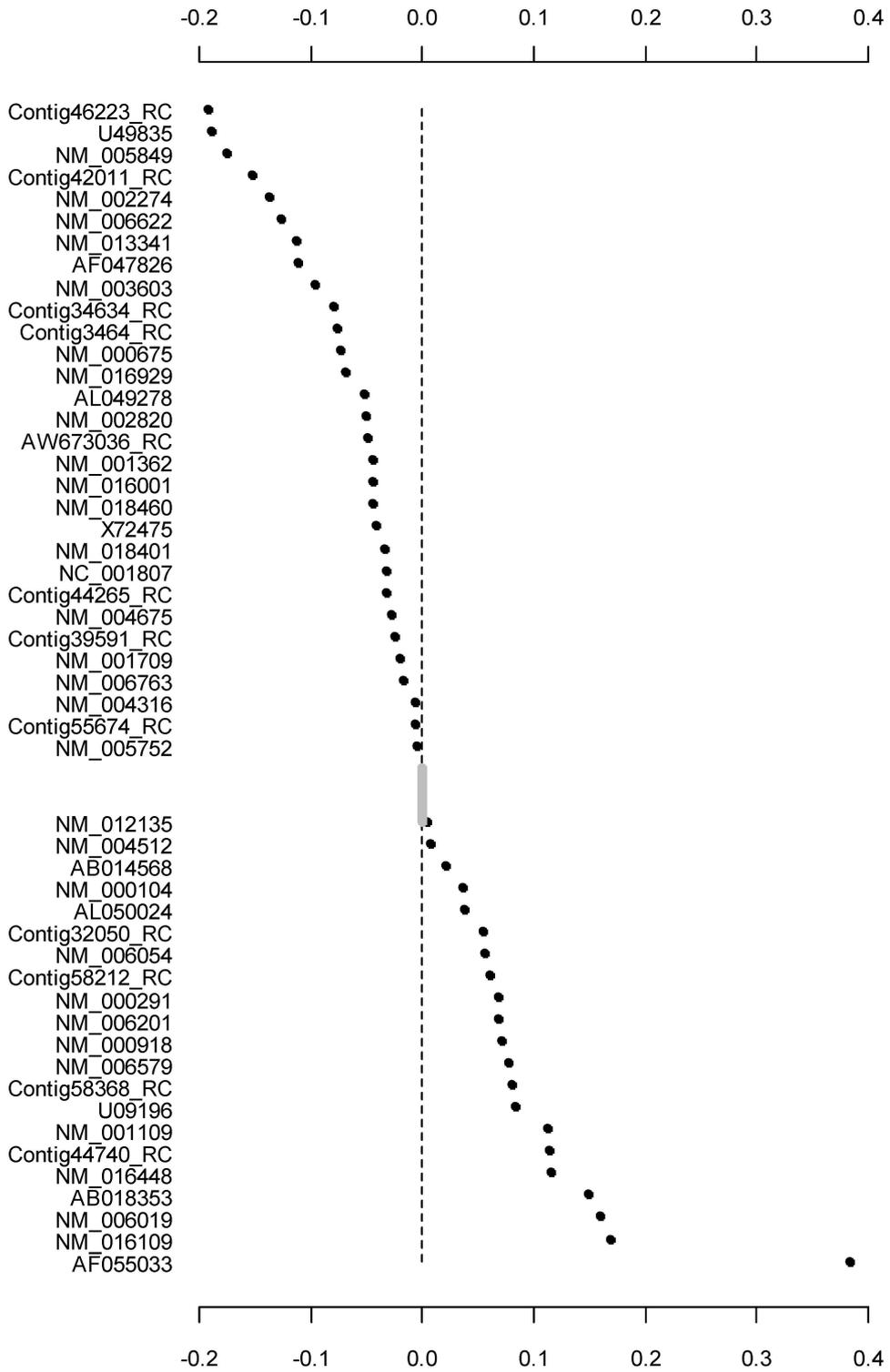
Scenario 2 (s_0=0.03, s_1=1)

Scenario 4 (s_0=0.04, s_1=1)

Scenario 5 (s_0=0.03, s_1=1)

Standardized effect size

Standardized effect size