Iran J Public Health, Vol. 45, No.2, Feb 2016, pp. 239-248

Original Article

Predicting the Survival Time for Bladder Cancer Using an Additive Hazards Model in Microarray Data

Leili TAPAK [1], ****Hossein MAHJUB*** [2], ***Majid SADEGHIFAR*** [3], ***Massoud SAIDIJAM*** [4], ***Jalal POOROLAJAL*** [5]

1. *Dept. of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran*
2. *Research Center for Health Sciences and Dept. of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran*
3. *Dept. of Statistics, Faculty of Science, Bu-Ali Sina University, Hamadan, Iran*
4. *Research Center for Molecular Medicine, Dept. of Molecular Medicine and Genetics, School of Medicine, Hamadan University of Medical Sciences, Hamadan, Iran*
5. *Modeling of Noncommunicable Diseases Research Center, Dept. of Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran*

***Corresponding Author:** Email: mahjub@umsha.ac.ir

(Received 04 May 2015; accepted 11 Oct 2015)

Abstract

Background: One substantial part of microarray studies is to predict patients' survival based on their gene expression profile. Variable selection techniques are powerful tools to handle high dimensionality in analysis of microarray data. However, these techniques have not been investigated in competing risks setting. This study aimed to investigate the performance of four sparse variable selection methods in estimating the survival time.

Methods: The data included 1381 gene expression measurements and clinical information from 301 patients with bladder cancer operated in the years 1987 to 2000 in hospitals in Denmark, Sweden, Spain, France, and England. Four methods of the least absolute shrinkage and selection operator, smoothly clipped absolute deviation, the smooth integration of counting and absolute deviation and elastic net were utilized for simultaneous variable selection and estimation under an additive hazards model. The criteria of area under ROC curve, Brier score and c-index were used to compare the methods.

Results: The median follow-up time for all patients was 47 months. The elastic net approach was indicated to outperform other methods. The elastic net had the lowest integrated Brier score (0.137±0.07) and the greatest median of the over-time AUC and C-index (0.803±0.06 and 0.779±0.13, respectively). Five out of 19 selected genes by the elastic net were significant ($P<0.05$) under an additive hazards model. It was indicated that the expression of RTN4, SON, IGF1R and CDC20 decrease the survival time, while the expression of SMARCAD1 increase it.

Conclusion: The elastic net had higher capability than the other methods for the prediction of survival time in patients with bladder cancer in the presence of competing risks base on additive hazards model.

Keywords: Survival analysis, Microarray data, Additive hazards model, Variable selection, Bladder cancer

Introduction

Urothelial carcinoma of the urinary bladder is the ninth most frequently diagnosed malignancy world-wide (1) and one of the most prevalent, representing 3% of cancers diagnosed globally (2). Bladder cancer accounts for an estimated 386,000 new diagnoses and 150,000 related deaths annually (1). Early detection of bladder cancer remains one of the most urgent issues in many researches (3). Although, there are some improvements in imaging and surgical techniques, the overall mortality of patients with bladder cancers has not been unchanged (4) and outcomes for patients remain suboptimal (2).

Currently, morphologic and pathologic criteria such as histology, stage, and grade are used for conventional diagnosis of bladder cancer, which play an important role in determining treatment (5). However, even though essential prognostic information is provided by these clinical criteria, they have inadequate power to predict patient outcome precisely (5) and there remains significant variability in the prognosis of patients with similar characteristics (2). Thus, the need to identify additional tumor characteristics that predict clinical behavior is highlighted in patients with bladder cancer (2).

Recently, there has been a growing interest in the use of gene expression signatures for prediction disease outcome of patients with cancer. The major objective of these studies is to identify a small subset of genes that their expression levels are significantly correlated with clinical outcomes like time to an event (6). Identification of influential genes may help to better characterize cancers and consequently optimize therapy decisions. Accordingly, the cancer patients' survival time can be estimated based on gene expression profile (7). However, typical methods of analysis are not applicable anymore, because the number of genomic variables, *P*, is much greater than the number of subjects, *n* (8). When the outcome of interest is survival time, microarray data analysis is further complicated due to censorship for a number of subjects, especially in the competing risks setting where there is more than one reason of failure.

Several variable selection methods based on the maximization of a penalized likelihood, originally developed for linear regression, have been adapted to survival models for high dimension low sample size time-to-event data under the Cox proportional hazards model (9-14). For example, the least absolute shrinkage and selection operator (Lasso) (15) smoothly clipped absolute deviation (SCAD) (10), Dantzig selector (9), LARS (11), the smooth integration of counting and absolute deviation (SICA) penalty (12) and the elastic net (13) have been proposed.

A well-known method for analyzing survival data is additive hazards model, which assumes that the covariates have additive effects on the hazard (16, 17). The additive models have some remarkable features. Particularly, they pertain to the risk difference or excess risk measure, which is especially relevant and informative in epidemiological and clinical studies (8). Regularization techniques for variable selection, in high dimension survival data, have also been extended to the additive hazards model, though the number of studies is limited (8, 18). The objective function of additive hazards model makes least-squares form of estimations computationally easier. This is especially substantial for high dimensional studies where computation cost is of serious concern (19). These techniques have just been used for a single time endpoint and the performance of them has not been investigated for gene selection in the presence of competing risks.

This study aimed to investigate the performance of four renowned variable selection methods of Lasso, SCAD, SICA and elastic net for high-dimensional time-to-event data with competing risks based on an additive hazards model to predict survival time in patients with bladder cancer. The other goal of this study was to identify significant genes among those selected by the better variable selection method and to determine their effects on bladder cancer patient's survivals according to the additive hazards model.

Methods

Data Source

This study used a publicly available bladder cancer data set analyzed by Dyrskjøt et al. (20). The dataset consists gene expression measurements for 1381 genes and survival outcomes on 404 patients with bladder cancer that were operated in the years 1987 to 2000 in hospitals in Denmark, Sweden, Spain, France, and England, with pT*a* and pT1 tumors, with no previous or synchronous muscle-invasive tumors (GEO with series accession no. GSE5479). However, the analysis was limited to n=301 patients with complete information. Two competing events including time to progression or death from bladder cancer (the response of interest) and death from other or unknown causes (the competing event) existed.

Regularization for Additive Hazards Model

For a sample with *k* competing risk types, let T_k be the time to the *k*th type of failure, $T=\min(T_1,\ldots,T_K)$ be the failure time and *C* be the censoring time. Denote the failure indicator by $\Delta = I\,(T \le C)$, where $I\,(\cdot)$ is the indicator function and takes value 1 if the observed time is an event time and value 0 if censoring occurred. Let **Z** be a *P*-dimensional vector of predictable covariate processes and assume that T and *C* are conditionally independent given **Z**. The observed data consist of $(T_i, \Delta_i \varepsilon_i, Z_i)$, where $\varepsilon_i \in \{1,\ldots,K\}$ indicates the (potentially unobserved) cause of failure (8, 21).

The cause-specific hazard function associated with *k*th risk is defined as:

$$\lambda_k(t;Z)=\lim_{\Delta t\to 0}\frac{P\left(t\le T\le t+\Delta t,\varepsilon=k\,|T\ge t,Z\right)}{\Delta t},\qquad k=1,\ldots,K\,.$$

Under the Lin and Ying additive hazards model (16), the hazard function of a failure time *T* conditional on a *P*-vector of possibly time-dependent covariates **Z** is specified as:

$$\lambda_k(t;Z)=\lambda_0(t)+\beta_0^T Z$$

where $\lambda_0(.)$ is an unspecified baseline hazard function which is common to all subjects and β_0 is a *P*-vector of regression coefficients (8).

The penalized estimator $\hat{\beta}$ of regression coefficients in the additive hazards model is a solution to the regularization problem:

$$\hat{\beta}=\underset{\beta\in\mathbb{R}^p}{\arg\min}\left\{Q(\beta)\equiv L(\beta)+\sum_{j=1}^{p}p_\lambda\left(\left|\beta_j\right|\right)\right\}$$

Where $L(\beta)$ is the likelihood function of the additive hazards model (8), and $p_\lambda(\theta),\ \theta\ge 0$, is a penalty function that depends on the regularization parameter $\lambda\ge 0$ and often is rewritten as $p_\lambda(\cdot)=\lambda\rho(\cdot)$ (8).

In this study, four commonly used sparse penalty functions of Lasso, SCAD, SICA and elastic net have been considered (8). Then, the L_1-penalty term, $\rho(\theta)=\theta$, $\theta\ge 0$ is used by the Lasso method. On the other hand, the elastic net method combines the L_1-penalty $\rho(\theta)=\theta$ and the L_2-penalty $\rho(\theta)=\theta^2$ which yields a penalty with the form of $\rho(\theta)=(1\text{-}a)\theta+a\theta^2$ and 0<a<1. The SCAD penalty is given by the derivative $\rho'_\lambda(\theta)=I(\theta\le\lambda)+\frac{(a\lambda\text{-}\theta)_+}{(a\text{-}1)\lambda}I(\theta>\lambda),\ \theta\ge 0$ with some a>2 as a shape parameter and the SICA penalty takes the form $\rho(\theta)=\frac{(a+1)\theta}{a+\theta}$, $\theta\ge 0$ and a>0 is a shape parameter. Estimation of $\hat{\beta}$ was accomplished through the *coordinate descent algorithm* (8).

After a solution path has been produced, selecting the optimal regularization parameter λ is carried out via the use of a cross validation score by *M*-fold cross-validation. The cross-validation score is defined as follows

$$CV(\lambda)=\frac{1}{M}\sum_{m=1}^{M}L^{(m)}\left(\hat{\beta}^{(-m)}(\lambda)\right),$$

where $L^{(m)}(.)$ is the least squares type loss function computed from the *m*th part of the data, and $\hat{\beta}^{(-m)}(\lambda)$ is the estimate from the data with the *m*th part removed (8).

The additional parameter *a* in the elastic net, SCAD and SICA methods was also tuned according to the method used by (8).

Performance Criteria

Assessment of the performance of the four methods was conducted through several criteria. Analysis of predictive performance was performed by using time-dependent receiver operator characteristic (ROC) curves (22) and bootstrap .632+ prediction error curves (21). The present study utilized concordance probability (C-index) that can be applied to measure and compare the discriminative power of a risk prediction models and the Integrated Brier score (21, 23).

Software

Analysis was performed using the R software programming (http://www.r-project.org) by implementing a publically available R package which has been provided by Lin and Lv (2013) (http://162.105.204.96/teachers/linw/software.html). In addition, the "pec" and "survAUC" R packages were utilized to evaluate the performance of used methods.

Results

The median follow-up time for all patients was 47 months. Progression or death from bladder cancer and competing event were observed in 74 and 33 patients, respectively. By the end of the time of follow-up, the number of 194 patients was censored.

Additive hazards models were fitted to microarray bladder cancer data by the Lasso, elastic net, SCAD and SICA penalization techniques for the 'progression or death from bladder cancer' event. Table 1 presents selected genes by the four methods. The procedures were repeated 100 times, each time yielding a different set of genes. The frequency of occurrences of the genes, means of coefficients and standard errors over 100 replicates, were shown in Table 1. The number of selected genes varied between the methods. In addition, there were eight common genes (SEQ265, SEQ279, SEQ1226, SEQ1262, SEQ1384, SEQ213, SEQ34 and SEQ377) among the four feature selection techniques.

Table 1: Selected microarray features by using four variable selection approaches for progression or death from bladder cancer event in Dyrskjøt data set. Values shown are frequency of selected genes, means of coefficients (standard errors) over 100 replicates

Method	Elastic net		LASSO		SCAD		SICA	
Gene ID	**Frequency**	***β (SE)**	**Frequency**	**β (SE)**	**Frequency**	**β (SE)**	**Frequency**	**β (SE)**
SEQ1082	84	5.2(0.3)	88	4.9(0.3)	87	3.9(0.3)	-	-
SEQ1197	90	4.5(0.2)	91	4.3(0.2)	95	3.6(0.2)	-	-
SEQ1226	100	-5.5(0.2)	100	-5.1(0.2)	100	-4.4(0.2)	100	-19(0.2)
SEQ1262	99	-8.9(0.3)	99	-9.0(0.3)	99	-7.9(0.3)	100	-19(0.2)
SEQ1284	-	-	-	-	-	-	56	1.5(0.2)
SEQ1295	-	-	-	-	-	-	83	2.3(0.2)
SEQ1330	54	0.8(0.1)	43	0.6(0.1)	40	0.4(0.1)	-	-
SEQ1384	76	-5.1(0.4)	75	-4.9(0.4)	74	-3.4(0.3)	100	-48(0.3)
SEQ162	98	-1.6(0.1)	96	-1.6(0.1)	98	-1.5(0.1)	-	-
SEQ213	63	1.6(0.2)	43	1.2(0.2)	50	0.8(0.1)	83	3.8(0.3)
SEQ240	32	-0.1(0.3)	-	-	-	-	-	-
SEQ265	100	9.4(0.1)	100	9.5(0.1)	100	9.4(0.1)	97	2.6(0.1)
SEQ279	84	-5.4(0.3)	88	-5.2(0.3)	87	-4.1(0.3)	83	-1.7(0.1)
SEQ287	76	1.0(0.1)	69	0.8(0.1)	74	0.6(0.1)	-	-
SEQ34	100	23(0.3)	100	24(0.3)	100	23(0.3)	100	31(0.1)
SEQ377	99	8.8(0.4)	97	8.1(0.3)	99	7.2(0.3)	100	8.2(0.2)
SEQ408	-	-	-	-	-	-	13	0.4(0.1)
SEQ410	-	-	-	-	-	-	100	6.9(0.2)
SEQ542	-	-	-	-	-	-	56	0.9(0.1)
SEQ820	100	11(0.1)	100	11(0.1)	100	12(0.1)	-	-
SEQ833	100	7.1(0.2)	100	7.1(0.2)	100	6.6(0.2)	-	-
SEQ843	-	-	-	-	-	-	97	-3.3(0.1)
SEQ940	90	-5.2(0.3)	91	-5.1(0.2)	95	-4.2(0.2)	-	-
SEQ948	-	-	-	-	-	-	13	-0.3(0.1)

*Coefficients and standard errors (SE) must be multiplied by 10^{-4}

Table 2 shows the mean and standard errors of the Brier score, the median of the area under ROC curve over time (AUC) and C-index for each method. In terms of three criteria, the elastic net penalty outperformed the other three methods. The mean of the integrated Brier score of the elastic net over 100 repetitions was the lowest (0.137±0.07). In addition, the mean of the median of the over-time AUC and C-index were the greatest for the elastic net (0.803±0.06 and 0.779±0.13, respectively).

In order to evaluate prediction performance improvement by including selected microarray features over a purely clinical model, strap .632+ prediction error curves were drawn based on B=100 bootstrap samples drawn without replacement. Fig. 1 shows the estimates for prediction of the four variable selection methods as well as the model with clinical covariates. Including selected microarray features in the models clearly improve over the purely clinical model, indicating that valuable information is contained in the data. In addition, based on these criteria, the elastic net outperformed the other three methods.

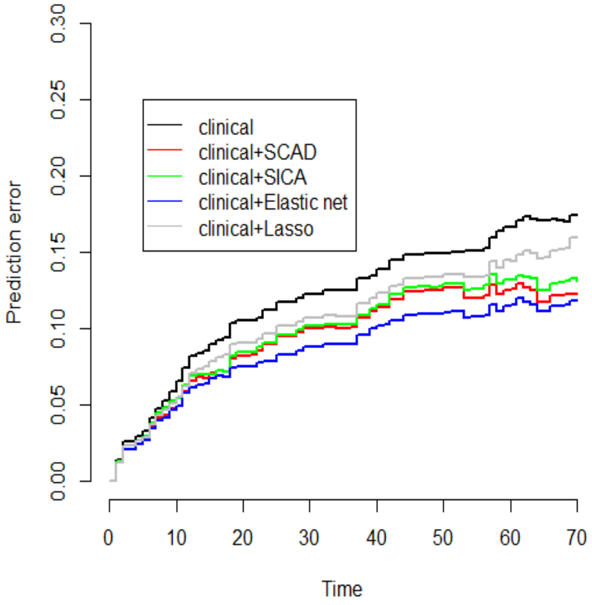

Fig. 1: Bootstrap 0.632+ prediction error curve estimates for prediction of the conditional probability function from bladder cancer microarray data

Table 2: Results of various methods applied to the Bladder cancer microarray data

Method	Integrated Brier score	AUC(t)	C-index
Elastic net	0.137±0.07	0.803±0.06	0.779±0.13
Lasso	0.153±0.09	0.741±0.11	0.693±0.19
SCAD	0.144±0.06	0.763±0.07	0.722±0.16
SICA	0.145±0.07	0.761±0.07	0.717±0.12

AUC is the area under ROC curve

Besides, five out of 19 genes selected by elastic net (SEQ1082, SEQ1197, SEQ1262, SEQ833 and SEQ940) were significant based on additive hazards model (*P*=0.005, 0.016, 0.039, 0.009 and 0.025 respectively). Table 3 shows the coefficients, standard errors and *P*-values for these genes. The expression of gene SEQ940 increased the survival time, whereas the expression of other significant genes (SEQ1082, SEQ1197, SEQ1262 and SEQ833) decreased the survival time.

Discussion

The present study considered a cause specific additive hazard approach for high-dimensional time-to-event data with competing risks and compared the performance of four penalized variable selection techniques of Lasso, elastic net, SCAD and SICA. Four variable selection techniques were applied in a real dataset containing microarray competing risks data related to the bladder cancer patients.

Table 3: Influential genes on bladder cancer patient's survival based on additive hazards model from selected genes by elastic net

Gene ID	Gene number	Gene name	Coefficient (SE)	*P*-value
SEQ1082	NM_207521.1	Homo sapiens reticulon 4 (RTN4)	0.0025 (0.0009)	0.005
SEQ1197	NM_003103.5	Homo sapiens (human) SON DNA binding protein (SON)	0.0034 (0.0014)	0.016
SEQ1262	NM_000875.2	Homo sapiens insulin-like growth factor 1 receptor (IGF1R), mRNA	0.0029 (0.0014)	0.039
SEQ833	NM_001255.1	Homo sapiens CDC20 cell division cycle 20 homolog (S. cerevisiae) (CDC20), mRNA	0.0018 (0007)	0.009
SEQ940	NM_020159.1	Homo sapiens SWI/SNF-related, matrix-associated actin-dependent regulator of chromatin, subfamily a, containing DEAD/H box 1 (SMARCAD1), transcript variant 3, mRNA	-0.0023 (0.0010)	0.025

Despite the existence of the variability in the selected genes by different methods due to the low sample size and high dimensionality, there was some consistency across the methods. However, the elastic net method showed better performance. The results also showed that selected microarray features by the four methods could improve the prediction of survival over a purely clinical model in bladder cancer patients.

Besides, based on an additive hazards model five out of 19 selected genes by the elastic net were indicated as influential genes on bladder cancer survival. Consequently, these genes can predict the survival time of the patients with bladder cancer. The expression of genes RTN4, SON, IGF1R and CDC20 can decrease the survival time, while the expression of gene SMARCAD1 might increase survival time.

The significant genes are related to types of cancers. Nogo proteins, encoded by gene reticulon-4 (RTN4), are myelin associated endoplasmic reticulum proteins, have been suggested by recent studies to play an important role in apoptosis, especially in cancer cells like bladder and lung cancer (24, 25). They were apoptosis-inducing proteins, and involved in the process of apoptosis through some classical apoptotic signal pathways (24, 25). A potent neurite outgrowth inhibitor is the product of this gene and is useful in blocking the regeneration of the central nervous system (25). Besides, the SON protein regulates alternative splicing of RNAs from the genes involved in apoptosis and epigenetic modification (26). In addition, SON-mediated splicing is essential for proper processing of selective transcripts related to cell cycle, microtubules, centrosome maintenance, and genome stability (26). In addition, the absence of this gene involved in the regulation of gene expression will result in a disruption in gene expression and is effective in the editing process. This gene has an important role in cancer and other types of human disease (27, 28).

The insulin-like growth factor 1 receptor (IGF1R) signaling pathway plays important roles in regulating cellular proliferation and apoptosis and changes in expression of IGF1R may be a risk factor of cancer incidence (29). This gene functions as an anti-apoptotic agent by enhancing cell survival, besides, it is highly overexpressed in most malignant tissues (30). The overexpression of the IGF1R in invasive bladder cancer tissues and promotes motility and invasion of urothelial carcinoma cells have been confirmed by several studies (31-35). The other gene, CDC20, acts as a regulatory protein interacting with several other proteins at multiple points in the cell cycle. The overexpression of CDC20 is related to poor prognosis of urothelial carcinoma of the human bladder (36-38). The gene SMARCAD1's function is as binding transcriptional start sites of many genes involved in transcriptional regulation and the end resection. This gene encodes a member of the SNF subfamily of helicase proteins, which plays a critical role in the restoration of heterochromatin organization and propagation of epigenetic patterns following DNA replication by mediating histone H3/H4 deacetylation (39). Heterochromatin maintenance and proper chromosome segregation needs SMARCAD1 and it is related to several types of cancer including bladder, breast, colorectal, and gastric cancers (40-42). This data set was first analyzed by Dyrskjøt et al. (20) and they identified 88 highly significantly correlated genes with progression-free survival by a univariate Cox regression strategy. Based on the results of the present study, there were only four common microarray feature (SEQ213, SEQ833, SEQ820 and SEQ843). In addition, Binder et al. performed a study on the same dataset using Cox proportional hazards likelihood based boosting method (21). There were also five common genes (SEQ34, SEQ162, SEQ265, SEQ820 and SEQ1384) with their study.

A number of studies have evaluated the performance of variable selection methods based on both additive and proportional hazards approaches (9, 10, 18, 19). In a study conducted by Lin and Lv (8), the performance of the elastic net, Lasso, SCAD and SICA was compared based on a simulation study for a single point time-to-event data in high-dimensional low sample size setting. Their results showed that the SICA outperformed other methods, which was inconsistent with the results of the present study. In another study Engler and Li (6), compared the elastic net and Lasso variable selection methods, for non-competing risks time-to-event data in high-dimensional low-sample size setting based on Cox proportional hazards. The results of their simulation studies and real data set demonstrated that the elastic net outperform the Lasso, which was in agreement with the present study. Based on MSE criteria the performance of elastic net in linear regression was superior to Lasso, which was also similar to the results of the present study (13). The performance of elastic net and Lasso was compared via a simulation study for a single point time-to-event data in high-dimensional low sample size setting based on additive hazards method (8). Their results showed that the elastic net outperformed the Lasso, which was similar to the result of the present study for competing risks setting. Ogutu et al. reported similar accuracies for Lasso and elastic net in handling linear regression (43).

The performance of variable selection methods and the different models depends on the used data with no method dominating the others (44, 45). The present study focused on evaluation of the performance of four well-known variable selection methods of Lasso, elastic net, SCAD and SICA in an additive manner to analyze microarray competing risks data. The additive hazards model and used variable selection approaches provide a useful alternative to existing dimension reduction techniques based on Cox's model for competing risks survival data with high-dimensional covariates.

The present study also introduced a new set of influential microarray features in bladder cancer patients' survival from an additive perspective, which is different from proportional hazards point of view. According to the result of the present study, despite the small number of selected genes all the methods of Lasso, elastic net, SCAD and SICA showed reasonable performance in additive manner and the selected genes improved prediction performance over a purely clinical model.

The expression levels of influential genes play an important role on survival time as either risk factors or preventive factors. Therefore, determining the expression levels of such genes might help in primary prevention programs (46).

Conclusion

The elastic net penalty has higher capability than the Lasso, SCAD and SICA in the prediction of survival time in patients with bladder cancer in high-dimensional competing risk settings based on the additive hazards model. Besides a combination of appropriate statistical methods and gene expression data can help detecting influential genes in survival time.

Ethical considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgements

This paper is part of a PhD thesis in Biostatistics. We would like to thank the Vice Chancellor of Research and Technology, Hamadan University of Medical Sciences, Iran, for logistic and financial support. The authors declare that there is no conflict of interests.

References

1. Gurung P, Veerakumarasivam A, Williamson M, Counsell N, Douglas J, Tan WS, Feber A, Crabb SJ, Short SC, Freeman A (2014). Loss of expression of the tumour suppressor gene AIMP3 predicts survival following radiotherapy in muscle-invasive bladder cancer. *Int J Cancer,* 136(3): 709-720.
2. Riester M, Taylor JM, Feifer A, Koppie T, Rosenberg JE, Downey RJ, Bochner BH, Michor F (2012). Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer. *Clin Cancer Res,* 18:1323-1333.
3. Rosser CJ, Liu L, Sun Y, Villicana P, McCullers M, Porvasnik S, Young PR, Parker AS, Goodison S (2009). Bladder Cancer–Associated Gene Expression Signatures Identified by Profiling of Exfoliated Urothelia. *Cancer Epidemiol Biomarkers Prev.,* 18:444-453.
4. Meeks JJ, Bellmunt J, Bochner BH, Clarke NW, Daneshmand S, Galsky MD, Hahn NM, Lerner SP, Mason M, Powles T (2012). A systematic review of neoadjuvant and adjuvant chemotherapy for muscle-invasive bladder cancer. *Eur Urol,* 62:523-533.
5. Blaveri E, Simko JP, Korkola JE, Brewer JL, Baehner F, Mehta K, DeVries S, Koppie T, Pejavar S, Carroll P (2005). Bladder cancer outcome and subtype classification by gene expression. *Clin Cancer Res,* 11:4044-4055.
6. Engler D, Li Y (2009). Survival analysis with high-dimensional covariates: an application in microarray studies. *Stat Appl Genet Molec Biol,* 8:1-22.
7. Bøvelstad HM, Nygård S, Størvold HL, Aldrin M, Borgan Ø, Frigessi A, Lingjærde OC (2007). Predicting survival from microarray data—a comparative study. *Bioinformatics,* 23:2080-2087.
8. Lin W, Lv J (2013). High-dimensional sparse additive hazards regression. *JASA,* 108:247-264.
9. Antoniadis A, Fryzlewicz P, Letué F (2010). The Dantzig selector in Cox's proportional hazards model. *Scand J Stat,* 37:531-552.
10. Fan J, Li R (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *JASA,* 96:1348-1360.
11. Efron B, Hastie T, Johnstone I, Tibshirani R (2004). Least angle regression. *Ann Stat,* 32:407-499.
12. Lv J, Fan Y (2009). A unified approach to model selection and sparse recovery using regularized least squares. *Ann Stat,* 37:3498-3528.
13. Zou H, Hastie T (2005). Regularization and variable selection via the elastic net. *J R Stat Soc.: Series B (Statistical Methodology),* 67:301-320.
14. Park MY, Hastie T (2007). L1-regularization path algorithm for generalized linear models. *J R Stat Soc: Series B (Statistical Methodology),* 69:659-677.
15. Tibshirani R (1996). Regression shrinkage and selection via the lasso. *J R Stat Soc: Series B (Methodological)*:267-288.
16. Lin D, Ying Z (1994). Semiparametric analysis of the additive risk model. *Biometrika,* 81:61-71.
17. Xie X, Strickler HD, Xue X (2013). Additive hazard regression models: an application to the natural history of human papillomavirus. *Comput Math Methods Med,* 2013: 1-7.
18. Martinussen T, Scheike TH (2009). The additive hazards model with high-dimensional regressors. *Lifetime Data Anal,* 15:330-342.

19. Ma S, Huang J (2007). Additive risk survival model with microarray data. *BMC Bioinformatics,* 8: 192. doi: 10.1186/1471-2105-8-192.
20. Dyrskjøt L, Zieger K, Real FX, Malats N, Carrato A, Hurst C, Kotwal S, Knowles M, Malmström P-U, de la Torre M (2007). Gene expression signatures predict outcome in non–muscle-invasive bladder carcinoma: a multicenter validation study. *Clin Cancer Res,* 13:3545-3551.
21. Binder H, Allignol A, Schumacher M, Beyersmann J (2009). Boosting for high-dimensional time-to-event data with competing risks. *Bioinformatics,* 25:890-896.
22. Heagerty PJ, Lumley T, Pepe MS (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics,* 56:337-344.
23. Wolbers M, Blanche P, Koller MT, Witteman JC, Gerds TA (2014). Concordance for prognostic models with competing risks. *Biostatistics*:15(3):526-39
24. Chen C-L, Lai Y-F, Tang P, Chien K-Y, Yu J-S, Tsai C-H, Chen H-W, Wu C-C, Chung T, Hsu C-W (2012). Comparative and targeted proteomic analyses of urinary microparticles from bladder cancer and hernia patients. *J Proteome Res,* 11:5611-5629.
25. Zhang K, Bai P, Shi S, Zhou B, Wang Y, Song Y, Rao L, Zhang L (2013). Association of Genetic Variations in RTN4 3'-UTR with Risk of Uterine Leiomyomas. *Pathol Oncol Res,* 19:475-479.
26. Hickey CJ, Kim JH, Ahn EYE (2014). New Discoveries of Old SON: A Link Between RNA Splicing and Cancer. *J Cell Biochem,* 115:224-231.
27. Ahn E-Y, DeKelver RC, Lo M-C, Nguyen TA, Matsuura S, Boyapati A, Pandit S, Fu X-D, Zhang D-E (2011). SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Mol Cell,* 42:185-198.
28. Furukawa T, Tanji E, Kuboki Y, Hatori T, Yamamoto M, Shimizu K, Shibata N, Shiratori K (2012). Targeting of MAPK-associated molecules identifies SON as a prime target to attenuate the proliferation and tumorigenicity of pancreatic cancer cells. *Mol Cancer,* 11(88):1-10.
29. Quan H, Tang H, Fang L, Bi J, Liu Y, Li H (2014). IGF1 (CA) 19 and IGFBP-3-202A/C gene polymorphism and cancer risk: a meta-analysis. *Cell Biochem Biophys,* 69:169-178.
30. Åhlén J, Wejde J, Brosjö O, von Rosen A, Weng W-H, Girnita L, Larsson O, Larsson C (2005). Insulin-like growth factor type 1 receptor expression correlates to good prognosis in highly malignant soft tissue sarcoma. *Clin Cancer Res,* 11:206-216.
31. Moreira A, Meira-Machado L (2012). survivalBIV: Estimation of the Bivariate Distribution Function for Sequentially Ordered Events Under Univariate Censoring. *J Stat Softw,* 46(13):1-16.
32. Pineda S, Milne RL, Calle ML, Rothman N, de Maturana EL, Herranz J, Kogevinas M, Chanock SJ, Tardón A, Márquez M (2014). Genetic Variation in the TP53 Pathway and Bladder Cancer Risk. A Comprehensive Analysis. *PloS One,* 9(5):e89952.
33. Morrione A, Neill T, Iozzo RV (2013). Dichotomy of decorin activity on the insulin-like growth factor-I system. *FEBS J,* 280:2138-2149.
34. Metalli D, Lovat F, Tripodi F, Genua M, Xu S-Q, Spinelli M, Alberghina L, Vanoni M, Baffa R, Gomella LG (2010). The insulin-like growth factor receptor I promotes motility and invasion of bladder cancer cells through Akt-and mitogen-activated protein kinase-dependent activation of paxillin. *Am J Pathol,* 176:2997-3006.
35. Rochester MA, Patel N, Turney BW, Davies DR, Roberts IS, Crew J, Protheroe A, Macaulay VM (2007). The type 1 insulin-like growth factor receptor is over-expressed in bladder cancer. *BJU Int,* 100:1396-1401.
36. Dudziec E, Miah S, Choudhry HM, Owen HC, Blizard S, Glover M, Hamdy FC, Catto JW (2011). Hypermethylation of CpG islands and shores around specific microRNAs and mirtrons is associated with the phenotype and presence of bladder cancer. *Clin Cancer Res,* 17:1287-1296.
37. Choi J-W, Kim Y, Lee J-H, Kim Y-S (2013). High expression of spindle assembly checkpoint proteins CDC20 and MAD2 is associated with poor prognosis in urothelial bladder cancer. *Virchows Arch,* 463:681-687.

38. Lambrou GI, Adamaki M, Delakas D, Spandidos DA, Vlahopoulos S, Zaravinos A (2013). Gene expression is highly correlated on the chromosome level in urinary bladder cancer. *Cell Cycle,* 12(10):1544-59.
39. Adra CN, Donato J-L, Badovinac R, Syed F, Kheraj R, Cai H, Moran C, Kolker MT, Turner H, Weremowicz S (2000). SMARCAD1, a novel human helicase family-defining member associated with genetic instability: cloning, expression, and mapping to 4q22–q23, a band rich in breakpoints and deletion mutants involved in several human diseases. *Genomics,* 69:162-173.
40. Rowbotham SP, Barki L, Neves-Costa A, Santos F, Dean W, Hawkes N, Choudhary P, Will WR, Webster J, Oxley D (2011). Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Mole Cell,* 42:285-296.
41. O'Donnell PH, Stark AL, Gamazon ER, Wheeler HE, McIlwee BE, Gorsic L, Im HK, Huang RS, Cox NJ, Dolan ME (2012). Identification of novel germline polymorphisms governing capecitabine sensitivity. *Cancer,* 118:4063-4073.
42. Tappenden DM, Hwang HJ, Yang L, Thomas RS, LaPres JJ (2013). The Aryl-Hydrocarbon Receptor Protein Interaction Network (AHR-PIN) as Identified by Tandem Affinity Purification (TAP) and Mass Spectrometry. *J Toxicol,* 2013:1-12.
43. Ogutu JO, Schulz-Streeck T, Piepho H-P (2012) Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. BMC proceedings, BioMed Central Ltd, 6: pp. S10.
44. Hamidi O, Tapak L, Jafarzadeh Kohneloo A, Sadeghifar M (2014). High-Dimensional Additive Hazards Regression for Oral Squamous Cell Carcinoma Using Microarray Data: A Comparative Study. *BioMed Res Int,* 2014.
45. Ma S, Kosorok MR, Fine JP (2006). Additive Risk Models for Survival Data with High-Dimensional Covariates. *Biometrics,* 62:202-210.
46. Khoshhali M, Mahjub H, Saidijam M, Poorolajal J, Soltanian AR (2012). Predicting the survival time for diffuse large B-cell lymphoma using microarray data. *J Mol Genet Med,* 6:287-292.