

Original Paper

A User-Friendly, Web-Based Integrative Tool (ESurv) for Survival Analysis: Development and Validation Study

Kyoungjune Pak^{1*}, MD, PhD; Sae-Ock Oh^{2*}, MD, PhD; Tae Sik Goh^{3*}, MD, PhD; Hye Jin Heo², PhD; Myoung-Eun Han², PhD; Dae Cheon Jeong⁴, MA; Chi-Seung Lee^{5,6}, PhD; Hokeun Sun⁷, PhD; Junho Kang⁸, MA; Suji Choi⁸, BA; Soohwan Lee⁸, BA; Eun Jung Kwon⁸, BA; Ji Wan Kang⁸, BA; Yun Hak Kim^{2,8}, MD, PhD

¹Department of Nuclear Medicine, Pusan National University Hospital, Busan, Republic of Korea

²Department of Anatomy, School of Medicine, Pusan National University, Yangsan, Republic of Korea

³Department of Orthopaedic Surgery, Pusan National University Hospital, Busan, Republic of Korea

⁴Deloitte Analytics Group, Deloitte Consulting LLC, Seoul, Republic of Korea

⁵Biomedical Research Institute, Pusan National University Hospital, Busan, Republic of Korea

⁶Department of Biomedical Engineering, School of Medicine, Pusan National University, Busan, Republic of Korea

⁷Department of Statistics, Pusan National University, Busan, Republic of Korea

⁸Department of Biomedical Informatics, School of Medicine, Pusan National University, Yangsan, Republic of Korea

*these authors contributed equally

Corresponding Author:

Yun Hak Kim, MD, PhD

Department of Anatomy, School of Medicine

Pusan National University

49 Busandaehak-ro

Yangsan, 50612

Republic of Korea

Phone: 82 515108091

Email: yunhak10510@pusan.ac.kr

Abstract

Background: Prognostic genes or gene signatures have been widely used to predict patient survival and aid in making decisions pertaining to therapeutic actions. Although some web-based survival analysis tools have been developed, they have several limitations.

Objective: Taking these limitations into account, we developed ESurv (Easy, Effective, and Excellent Survival analysis tool), a web-based tool that can perform advanced survival analyses using user-derived data or data from The Cancer Genome Atlas (TCGA). Users can conduct univariate analyses and grouped variable selections using multiomics data from TCGA.

Methods: We used R to code survival analyses based on multiomics data from TCGA. To perform these analyses, we excluded patients and genes that had insufficient information. Clinical variables were classified as 0 and 1 when there were two categories (for example, chemotherapy: no or yes), and dummy variables were used where features had 3 or more outcomes (for example, with respect to laterality: right, left, or bilateral).

Results: Through univariate analyses, ESurv can identify the prognostic significance for single genes using the survival curve (median or optimal cutoff), area under the curve (AUC) with C statistics, and receiver operating characteristics (ROC). Users can obtain prognostic variable signatures based on multiomics data from clinical variables or grouped variable selections (lasso, elastic net regularization, and network-regularized high-dimensional Cox-regression) and select the same outputs as above. In addition, users can create custom gene signatures for specific cancers using various genes of interest. One of the most important functions of ESurv is that users can perform all survival analyses using their own data.

Conclusions: Using advanced statistical techniques suitable for high-dimensional data, including genetic data, and integrated survival analysis, ESurv overcomes the limitations of previous web-based tools and will help biomedical researchers easily perform complex survival analyses.

(*J Med Internet Res* 2020;22(5):e16084) doi: [10.2196/16084](https://doi.org/10.2196/16084)

KEYWORDS

survival analysis; grouped variable selection; The Cancer Genome Atlas; web-based tool; user service

Introduction

The accumulation of large amounts of genomic data following the development of next-generation sequencing techniques is paving the way toward precision medicine [1-4]. In particular, gene expression profiles or signatures have been widely used to predict patient prognosis and assist in deciding therapeutic strategies for the treatment of various cancers [2,5-9].

Genomic data sets are highly variable, with that variability rising with an increasing number of patients, making it high-dimensional in nature. To efficiently link high-dimensional genomic and survival data, statisticians have developed grouped variable selection models, based on the Cox proportional hazards model, including the following: least absolute shrinkage and selection operator (lasso), elastic net regularization (elastic net), and network-regularized high-dimensional Cox-regression (Coxnet, hereon referred to as Net) [2,10-13]. Among these methods, Net has been found to have the fewest overfitting problems and the highest prediction performance in these applications, as it takes into consideration the complexities of biological networks [2,6,8,10,14].

Successfully identifying and verifying prognostic factors using big databases is essential in medical research, but this can be difficult for researchers who are unfamiliar with computer science. To address this unmet clinical need, some web-based survival analysis tools have been developed. Although these tools have some limitations, they have been used in some univariate analyses [15-18]. SurvExpress [19], PROGgene [20], and PrognoScan [21] are popular web-based survival analysis tools that calculate the statistical significance of a prognosis using only messenger RNA (mRNA) expression data [15-18].

The limitations of previous tools include the following: (1) The use of mRNA expression as a simple categorical value to provide Kaplan Meier curves for all patients, regardless of their characteristics. The use of a continuous variable like mRNA as a categorical factor can change the nature of the variable arbitrarily, resulting in serious errors. (2) These tools do not provide gene/variable signatures that are statistically better, in terms of predicting prognosis, than a single gene. (3) They do not consider cancer classifications like histological type. (4) Users cannot use their own data. (5) These tools do not provide high-quality images and tabular results. (6) Users cannot create a risk-scoring system by specifying the genes of interest.

To overcome these limitations, we developed ESurv (Easy, Effective, and Excellent Survival analysis tool [22]), which is an online web resource for identifying prognostic biomarkers in pan-cancer from The Cancer Genome Atlas (TCGA) or user data.

Methods

Processing Genetic and Clinical Information From Patients

We performed survival analyses based on multiomics and clinical data from TCGA (Table 1), obtained via Broad GDAC Firebrowse [23] and the GDC Data Portal [24-26]. We used level 3 RNAseq (RNA sequencing), miRNAseq (microRNA sequencing), and methylation array data. The criteria for exclusion were as follows: genes with 0 values of more than 10% (when the amount of missing data is greater than 10%, the results may be biased) [27], patients with insufficient overall survival information (survival time or status), and patients with paired normal tissue and metastasis samples in TCGA.

Table 1. Summary of the data available in the ESurv.

Cancers with available omics data	Messenger RNA (Yes/No)	MicroRNA (Yes/No)	Methylation (Yes/No)	Total patients, n
Acute myeloid leukemia	Yes	Yes	Yes	200
Adrenocortical carcinoma	Yes	Yes	Yes	92
Bladder urothelial carcinoma	Yes	Yes	Yes	412
Brain lower grade glioma	Yes	Yes	Yes	515
Breast invasive carcinoma	Yes	Yes	Yes	1097
Cervical and endocervical carcinoma	Yes	Yes	Yes	307
Cholangiocarcinoma	Yes	Yes	Yes	45
Colon adenocarcinoma	Yes	Yes	Yes	458
Esophageal carcinoma	Yes	Yes	Yes	185
Glioblastoma multiforme	Yes	No	Yes	595
Head and neck squamous cell carcinoma	Yes	Yes	Yes	528
Kidney chromophobe	Yes	Yes	Yes	113
Kidney renal clear cell carcinoma	Yes	Yes	Yes	537
Kidney renal papillary cell carcinoma	Yes	Yes	Yes	291
Liver hepatocellular carcinoma	Yes	Yes	Yes	377
Lung adenocarcinoma	Yes	Yes	Yes	522
Lung squamous cell carcinoma	Yes	Yes	Yes	504
Lymphoid neoplasm diffuse large B cell lymphoma	Yes	Yes	Yes	48
Mesothelioma	Yes	Yes	Yes	87
Ovarian serous cystadenocarcinoma	Yes	Yes	No	591
Pancreatic adenocarcinoma	Yes	Yes	Yes	185
Pheochromocytoma and paraganglioma	Yes	Yes	Yes	179
Prostate adenocarcinoma	Yes	Yes	Yes	499
Rectum adenocarcinoma	Yes	Yes	Yes	171
Sarcoma	Yes	Yes	Yes	261
Skin cutaneous melanoma	Yes	Yes	Yes	470
Stomach adenocarcinoma	Yes	Yes	Yes	443
Testicular germ cell tumor	Yes	Yes	Yes	134
Thymoma	Yes	Yes	Yes	124
Thyroid carcinoma	Yes	Yes	Yes	516
Uterine carcinosarcoma	Yes	Yes	Yes	57
Uterine corpus endometrial carcinoma	Yes	Yes	Yes	548
Uveal melanoma	Yes	Yes	No	80

Processing Clinical Variables in Net

Clinical variables (tumor stage, age, sex, cancer type, blast count, histologic grade, laterality, anatomic neoplasm subdivision, tumor tissue site, and human papillomavirus status) can be included in the Net depending on cancer variety, allowing for sophisticated analyses. Clinical variables were classified as 0 and 1 when there were two categories (for example, chemotherapy: no or yes), and dummy variables were applied when the clinical variable could fall into three or more categories (for example, laterality: right, left, or bilateral).

Grouped Variable Selections for Creating Gene/Variable Signatures

ESurv uses one of the following three methods: least absolute shrinkage and selection operator (lasso), elastic net regularization (elastic net), and network-regularized high-dimensional Cox-regression (Net) using the Coxnet package (version 0.2) in R [2,10,12,14]. In Net analysis, we transformed the topological pathway information of large databases (KEGG [Kyoto Encyclopedia of Genes and Genomes], Biocarta, HumanCyc, Reactome, Panther, and NCI [National

Cancer Institute]) into a gene network matrix using the graphite package (version 3.10) in R. Users can set the mixing parameter alpha, which decides the balance between lasso and ridge regression [10,14]. All grouped variable selections use 10-fold cross-validation. After variable selection, we calculated the prognostic score by multiplying the variable value by the regression coefficient.

Statistical Analysis

To determine the optimal cutoff value, preventing overoptimization, we used the maximal UNO's C-index and a 5-fold cross validation. For the Kaplan-Meier survival curves, patients were divided into two groups, high- and low-risk, based on specific gene expression parameters (median cutoff or optimal cutoff value), with a P value determined by a log-rank test. The C-index and area under the curve (AUC) value were used to evaluate the effect of specific variables on survival [28]. These results can be seen not only in all patients but also in patient subgroups based on sex or stage. The results were obtained using R packages survival (version 2.44-1.1), survMisc (version 0.5.5), coin (version 1.3-0), MASS (Modern Applied

Statistics with S, version 7.3-51.4), edgeR (version 3.24.3), and survAUC (version 1.0-5). All graphical outputs from ESurv were plotted using the plotly (version 4.9.0) R package. All data cleaning and statistical analyses in this study were performed using R statistical software (version 3.6.0, R Core Team, R Foundation for Statistical Computing).

Implementation

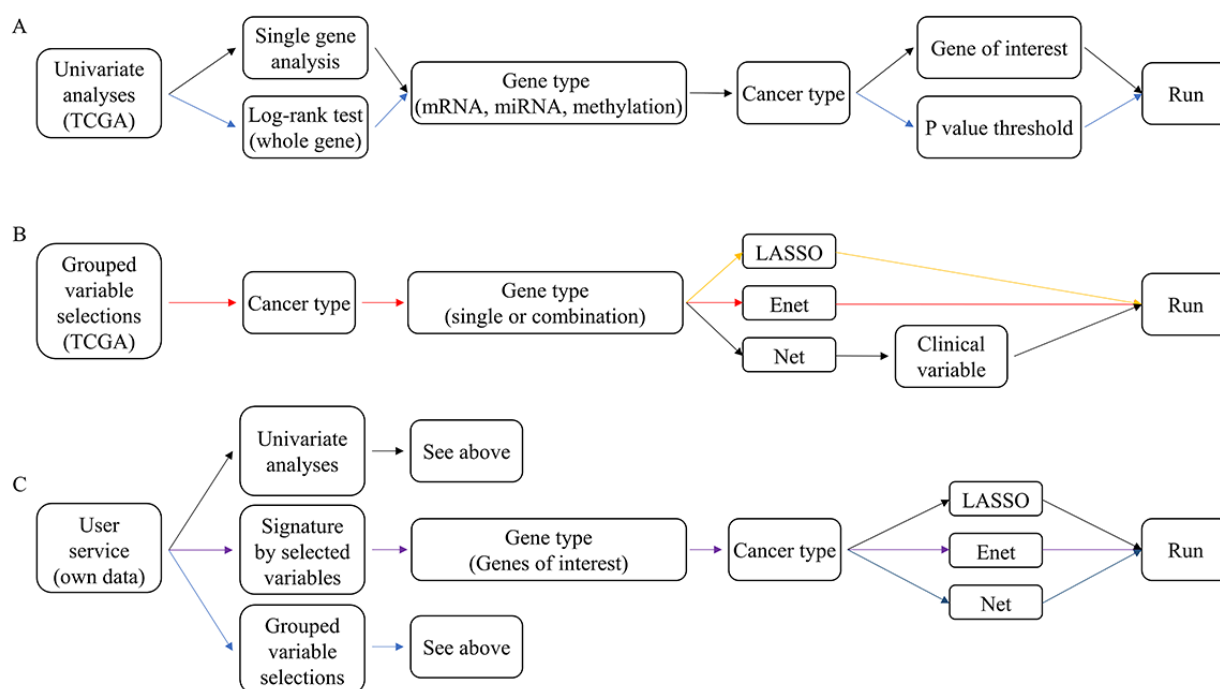
The ESurv web server implements AngularJS with HTML5 to display analyzed data through a web query interface. The results of these analyses were calculated on demand on a backend server running Java Servlet in conjunction with the R statistical program.

Results

Running ESurv

Details of the running procedure for ESurv are described in Figure 1. First, users can choose one of three methods: univariate survival analyses, grouped variable selections, or user service.

Figure 1. The running procedure of ESurv.



After choosing univariate survival analysis, users can select single gene analysis or log-rank test of whole genomes. In single gene analysis, users select the type of cancer, gene data type (mRNA, microRNA [miRNA], or methylation), gene of interest, and time for the receiver operating characteristics (ROC) curve, in that order. For log-rank testing of whole genomes, users choose the type of cancer, gene data type (mRNA, miRNA, or methylation), time for the ROC curve, and P value threshold, in that order.

If users instead choose grouped variable selections, they must select the type of cancer, gene data (mRNA, miRNA, methylation, or integrative analysis), grouped variable selection method (lasso, elastic net, or Net), time for the ROC curve, and

alpha, in that order. Alpha decides the balance between ridge and lasso penalties; the larger the alpha, the closer to lasso (alpha=1), and the fewer variables are chosen. If the users select Net, they can include clinical variables in grouped variable selection.

In the user service, users can perform univariate or grouped variable selections after uploading their own data. Instructions on uploading data are detailed in the manual. Once data are uploaded, all the abovementioned analyses can be performed.

Finally, when developing gene signatures using selected variables, users should choose the cancer type, genes of interest, and time for the ROC curve, in that order.

Univariate Analysis: Single Gene Analysis in Pan-Cancer

ESurv determines the prognostic significance of single genes as categorical (Kaplan-Meier curve with median or optimal cutoff values) or continuous variables (C-index and AUC values at specific time points) in various subgroups including sex and cancer stage (early and late as defined by the American Joint Committee on Cancer). To complement the results from the categorical variables, the C-index and AUC values were

calculated as continuous variables. For example, we performed survival analysis using mRNA (*SLC2A1*) as a biomarker in bladder urothelial carcinoma (n=403; Figure 2). As shown in Figure 2A and B, the discrimination power with optimal cutoff values is much better than that of analyses completed using median cutoff values in all patients. Users easily obtain survival curves in subgroups as well as for all patients. In the time-dependent ROC curve analyses, users can identify the AUC value based on the follow-up time (Figure 2C). ROC curves at selected times can all be calculated (Figure 2D).

Figure 2. An example of mRNA-based survival analysis. Expression levels of genes are classified as low or high (blue or red lines, respectively) based on the comparison of their optimal (A) and median cut-off values (B). (C) Time-dependent area under the curve (AUC) for each of these subgroups. (D) Receiver operating characteristics (ROC) curves for selected years in each of these subgroups.

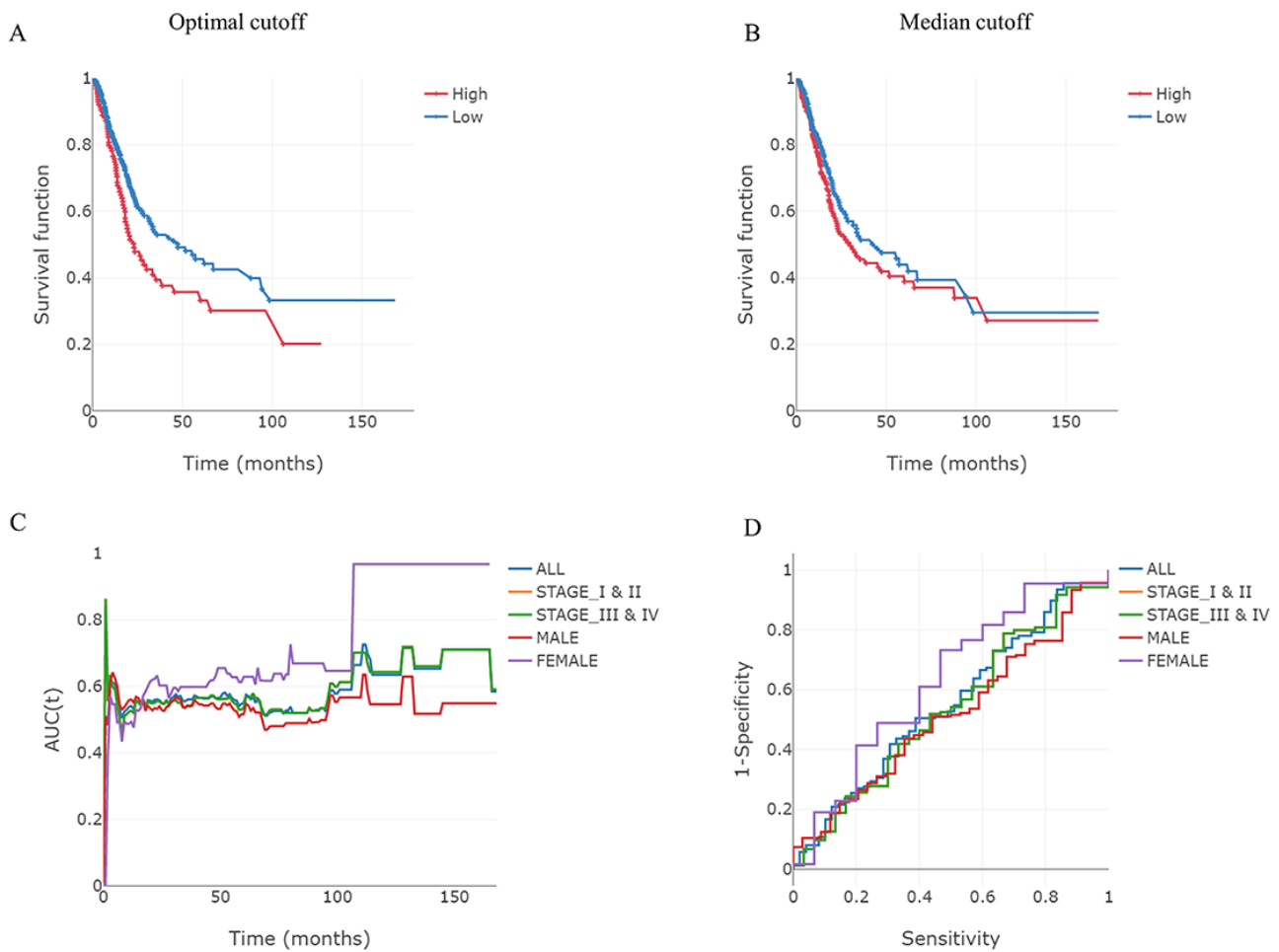
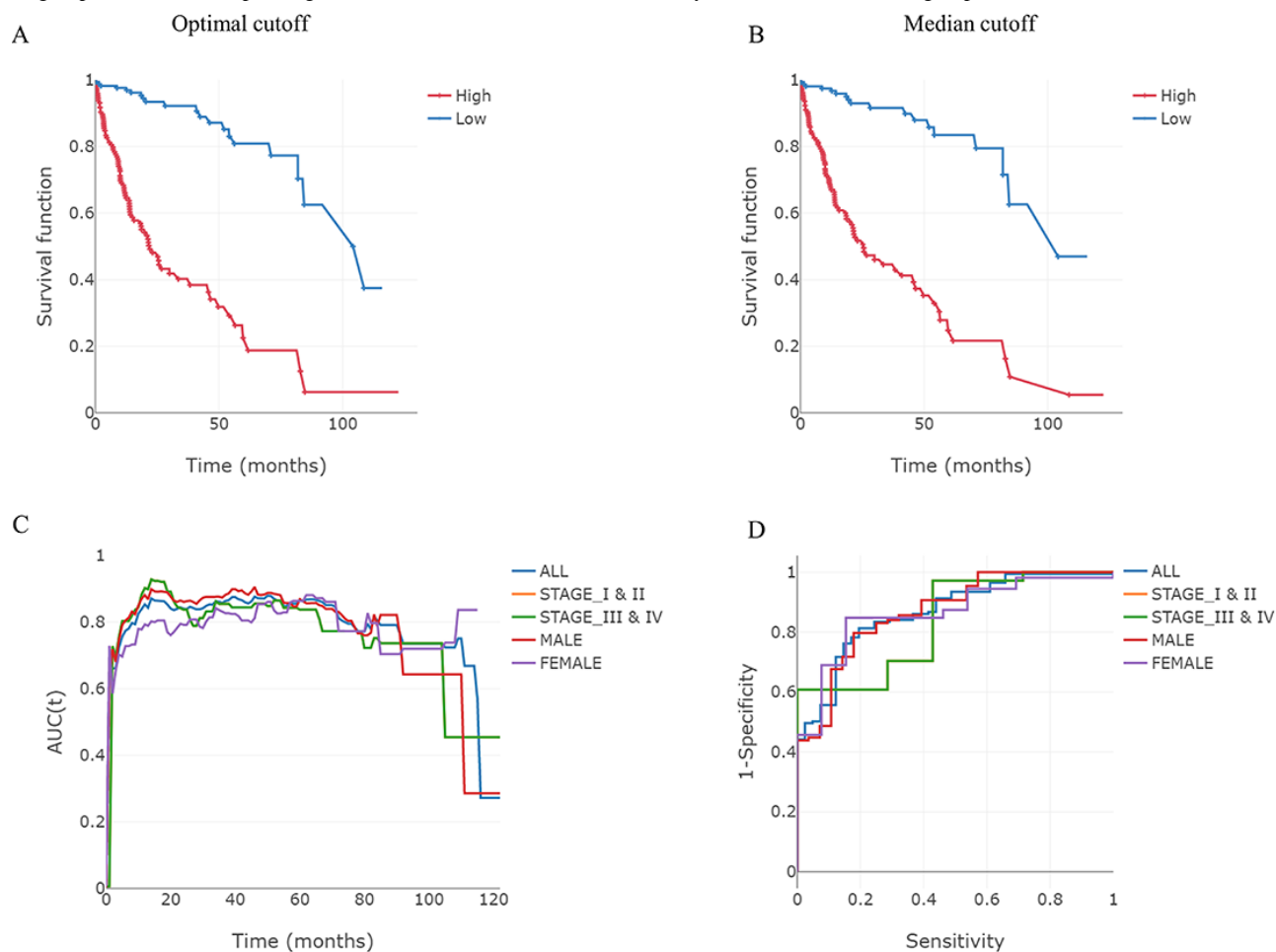


Figure 3. An example of survival analysis using a variable signature. Expression levels of genes were classified as either low or high (blue or red lines, respectively) based on a comparison of their optimal (A) and median cut-off values (B). (C) Time-dependent area under the curve (AUC) for each of the subgroups. (D) Receiver operating characteristics (ROC) curves for selected years in each of these subgroups.



Univariate Analysis: Log-Rank Test of Whole Genomes

If the users want to calculate the prognostic value of all gene variants in a particular cancer, they can choose to do a log-rank test for the whole genome. Here, we performed a log-rank test of whole genomes in colon adenocarcinoma with a $P < .05$. ESurv provided the results (gene name, C-index, AUC value at user-selected time, P value of log-rank test) as an Excel file ([Multimedia Appendix 1](#)).

Grouped Variable Selections

Grouped variable selection methods have been developed to take advantage of advances in biological technology, and can be used in statistical models to accurately predict the prognosis for a patient [2,10]. To select a prognostic gene set from high-dimensional data, it is necessary to select and minimize the number of variables in a systematic and statistically sound manner. ESurv provides representative grouped variable selection methods (lasso, elastic net, and Net), which can reduce the number of variables by considering the relationships between them. To develop a risk-scoring system, we used a linear combination of the expression values and regression coefficients of the selected variables. Users can develop a powerful prognostic signature in pan-cancer using any combination of these options. To illustrate this, we performed Net using all the clinical variables in hepatocellular carcinoma ($n=307$; [Figure](#)

3). Selected variables and regression coefficients can be downloaded from the summary tab in ESurv as Excel files. The results for all possible subgroups were recorded, just as in the case of single gene analysis. Outputs for the analysis can be in the form of Kaplan-Meier curves, time-dependent ROC curves, and AUC values at specific time points ([Figure 3](#)).

User Service

Users can conduct univariate analysis, log-rank tests, and grouped variable selections by uploading their own data ([Figure 1C](#)). The manual is provided in [Multimedia Appendices 2](#) and [3](#). Users can generate all survival results, as described above, using their own data sets. To protect data integrity, user-derived data sets are password-protected and connected to a unique user ID. In addition, as users refine their needs, they can request additional survival analysis packages via email. The most-requested package will be added to ESurv the following year.

Discussion

Principal Findings

To overcome the limitations of existing survival analysis tools, we developed a web-based user-friendly tool called ESurv. Kaplan-Meier curves are a common method of conducting survival analysis in the medical field, which involves

categorizing patients into groups based on their risk profile. There is no clear criterion to classify continuous variables, like gene expression, in these categorical analysis methods. For this reason, median and quartile cutoff values used in previous tools may miss the prognostic significance of individual genes. As shown in [Figure 2B](#), the median cutoff is not a suitable parameter for risk stratification, whereas optimal cutoff is a good parameter for risk stratification. Additionally, survival analysis as a continuous variable should be accompanied by these results to evaluate the prognostic significance of the gene of interest. Researchers can identify the prognostic significance of a gene (mRNA, miRNA, and methylation) as both a categorical and continuous variable in ESurv.

Prognosis using multiple genes yields superior results compared to using a single gene. There are a number of ways to select gene signatures, but among these, genes selected using grouped variable selection have proven to be the most versatile and reproducible [2,5,9,10,13,29,30]. Grouped variable selection methods select and shrink variables from high-dimensional data sets while considering multicollinearity, which is especially valuable when considering biological pathways, making Net options ideal for these applications. Despite this, grouped variable selection has not been applied to many studies because of the difficulty surrounding its computer programming (R, Python, and Matlab). To address this gap, we added grouped variable selection methods (lasso, elastic net, and Net) to ESurv. When selecting a prognostic gene set from high-dimensional data, researchers have to minimize the number of variables. In order to reduce variables effectively, we must provide as much information as possible linking genes and their relevant pathways [2,10]. It is easier to complete external validation with Net than with other methods because it performs variable selection using information about each gene, derived from databases hosting information pertaining to genetic pathways (Reactome, HumanCyc, KEGG, Biocarta, NCI, and Panther) [2,10,29,30]. Like with the univariate analyses, users can obtain

the results as continuous and categorical data. ESurv is the first web-based tool to provide grouped variable selection using multiomics data.

The prognosis of patients may vary based on clinical information, such as sex and stage [9,31-35]. For these reasons, subgroup analysis is required to identify stage- and sex-specific prognostic genes. ESurv shows the results of survival analyses by taking into consideration tumor heterogeneity based on several classifications (cancer type, stages, and sex).

Limitations and Future Work

This software does have some limitations that will be addressed upon further development. Here we used only one cancer database but there are many more, which we plan to add to ESurv as we continue to develop the software. Although users can upload their own data, this still requires users to be computer savvy; this will be addressed in future versions of the software. In addition, ESurv currently only accesses cancer databases, but this type of analysis is valuable in other diseases, including several vascular and degenerative diseases; we aim to add these as well. Finally, ESurv is not exhaustive in its analyses, but it is possible for users to request additional survival analysis packages for R via email. We will then select the most requested package and add it to ESurv on an annual basis.

Conclusions

The most important functionality provided by ESurv is that users can use this software to analyze their own data. As more medical data is produced, the demand for survival analyses increases. Analysis of data created by individual institutions is as important as big data analysis, but there have been no survival analysis tools available to conduct this type of analysis. Using advanced statistical methods and comprehensive survival analyses, ESurv overcomes the limitations of previous tools, and allows users to work on their own data sets. We strongly believe that ESurv is an ideal tool to meet the growing demand for increased survival analysis in both small and large data sets.

Acknowledgments

This work was supported by grants from the Medical Research Center (MRC) Program and the Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the government of Korea (NRF-2018R1A5A2023879 and NRF-2019R1A2B5B01070163). This study was supported by a Biomedical Research Institute Grant (2018B032) from the Pusan National University Hospital.

Authors' Contributions

KP and TSG designed and wrote the manuscript. HJH, MEH, CSL, JK, SJC, and SL collected and preprocessed the data. DCJ and HS corrected the error in the algorithm. SOO and YHK designed and supervised the project. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The univariate results of whole significant genes.

[\[XLSX File \(Microsoft Excel File\), 339 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

The manual of univariate analyses.

[\[DOCX File , 67 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

The manual of grouped variable selection analyses.

[\[DOCX File , 72 KB-Multimedia Appendix 3\]](#)

References

1. Friedman AA, Letai A, Fisher DE, Flaherty KT. Precision medicine for cancer with next-generation functional diagnostics. *Nat Rev Cancer* 2015 Dec;15(12):747-756 [[FREE Full text](#)] [doi: [10.1038/nrc4015](https://doi.org/10.1038/nrc4015)] [Medline: [26536825](https://pubmed.ncbi.nlm.nih.gov/26536825/)]
2. Kim YH, Jeong DC, Pak K, Goh TS, Lee C, Han M, et al. Gene network inherent in genomic big data improves the accuracy of prognostic prediction for cancer patients. *Oncotarget* 2017 Sep 29;8(44):77515-77526 [[FREE Full text](#)] [doi: [10.18632/oncotarget.20548](https://doi.org/10.18632/oncotarget.20548)] [Medline: [29100405](https://pubmed.ncbi.nlm.nih.gov/29100405/)]
3. Han M, Kim J, Kim GH, Park SY, Kim YH, Oh S. SAC3D1: a novel prognostic marker in hepatocellular carcinoma. *Sci Rep* 2018 Oct 23;8(1):15608 [[FREE Full text](#)] [doi: [10.1038/s41598-018-34129-9](https://doi.org/10.1038/s41598-018-34129-9)] [Medline: [30353105](https://pubmed.ncbi.nlm.nih.gov/30353105/)]
4. Shaikh AR, Butte AJ, Schully SD, Dalton WS, Khoury MJ, Hesse BW. Collaborative biomedicine in the age of big data: the case of cancer. *J Med Internet Res* 2014 Apr 07;16(4):e101 [[FREE Full text](#)] [doi: [10.2196/jmir.2496](https://doi.org/10.2196/jmir.2496)] [Medline: [24711045](https://pubmed.ncbi.nlm.nih.gov/24711045/)]
5. Chen H, Yu S, Chen C, Chang G, Chen C, Yuan A, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 2007 Jan 04;356(1):11-20. [doi: [10.1056/NEJMoa060096](https://doi.org/10.1056/NEJMoa060096)] [Medline: [17202451](https://pubmed.ncbi.nlm.nih.gov/17202451/)]
6. Zemmour C, Bertucci F, Finetti P, Chetrit B, Birnbaum D, Filleron T, et al. Prediction of early breast cancer metastasis from DNA microarray data using high-dimensional cox regression models. *Cancer Inform* 2015;14(Suppl 2):129-138 [[FREE Full text](#)] [doi: [10.4137/CIN.S17284](https://doi.org/10.4137/CIN.S17284)] [Medline: [25983547](https://pubmed.ncbi.nlm.nih.gov/25983547/)]
7. Nault J, De Reyniès A, Villanueva A, Calderaro J, Rebouissou S, Couchy G, et al. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. *Gastroenterology* 2013 Jul;145(1):176-187. [doi: [10.1053/j.gastro.2013.03.051](https://doi.org/10.1053/j.gastro.2013.03.051)] [Medline: [23567350](https://pubmed.ncbi.nlm.nih.gov/23567350/)]
8. Zhang J, Song W, Chen Z, Wei J, Liao Y, Lei J, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol* 2013 Dec;14(13):1295-1306. [doi: [10.1016/S1470-2045\(13\)70491-1](https://doi.org/10.1016/S1470-2045(13)70491-1)] [Medline: [24239208](https://pubmed.ncbi.nlm.nih.gov/24239208/)]
9. Lau SK, Boutros PC, Pintilie M, Blackhall FH, Zhu C, Strumpf D, et al. Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol* 2007 Dec 10;25(35):5562-5569. [doi: [10.1200/JCO.2007.12.0352](https://doi.org/10.1200/JCO.2007.12.0352)] [Medline: [18065728](https://pubmed.ncbi.nlm.nih.gov/18065728/)]
10. Sun H, Lin W, Feng R, Li H. Network-Regularized High-Dimensional Cox Regression for Analysis of Genomic Data. *Stat Sin* 2014 Jul;24(3):1433-1459 [[FREE Full text](#)] [doi: [10.5705/ss.2012.317](https://doi.org/10.5705/ss.2012.317)] [Medline: [26316678](https://pubmed.ncbi.nlm.nih.gov/26316678/)]
11. Musoro JZ, Zwinderman AH, Puhan MA, ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. *BMC Med Res Methodol* 2014 Oct 16;14:116 [[FREE Full text](#)] [doi: [10.1186/1471-2288-14-116](https://doi.org/10.1186/1471-2288-14-116)] [Medline: [25323009](https://pubmed.ncbi.nlm.nih.gov/25323009/)]
12. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997 Feb 28;16(4):385-395. [Medline: [9044528](https://pubmed.ncbi.nlm.nih.gov/9044528/)]
13. Fulton L, Kruse CS. Hospital-Based Back Surgery: Geospatial-Temporal, Explanatory, and Predictive Models. *J Med Internet Res* 2019 Oct 29;21(10):e14609 [[FREE Full text](#)] [doi: [10.2196/14609](https://doi.org/10.2196/14609)] [Medline: [31663856](https://pubmed.ncbi.nlm.nih.gov/31663856/)]
14. Wu Y. ELASTIC NET FOR COX'S PROPORTIONAL HAZARDS MODEL WITH A SOLUTION PATH ALGORITHM. *Stat Sin* 2012;22:27-294 [[FREE Full text](#)] [doi: [10.5705/ss.2010.107](https://doi.org/10.5705/ss.2010.107)] [Medline: [23226932](https://pubmed.ncbi.nlm.nih.gov/23226932/)]
15. Aguirre-Gamboa R, Gomez-Rueda H, Martínez-Ledesma E, Martínez-Torteya A, Chacolla-Huaringa R, Rodriguez-Barrientos A, et al. SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. *PLoS One* 2013;8(9):e74250 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0074250](https://doi.org/10.1371/journal.pone.0074250)] [Medline: [24066126](https://pubmed.ncbi.nlm.nih.gov/24066126/)]
16. Goswami CP, Nakshatri H. PROGgene: gene expression based survival analysis web application for multiple cancers. *J Clin Bioinforma* 2013 Oct 28;3(1):22 [[FREE Full text](#)] [doi: [10.1186/2043-9113-3-22](https://doi.org/10.1186/2043-9113-3-22)] [Medline: [24165311](https://pubmed.ncbi.nlm.nih.gov/24165311/)]
17. Goswami CP, Nakshatri H. PROGgeneV2: enhancements on the existing database. *BMC Cancer* 2014 Dec 17;14:970 [[FREE Full text](#)] [doi: [10.1186/1471-2407-14-970](https://doi.org/10.1186/1471-2407-14-970)] [Medline: [25518851](https://pubmed.ncbi.nlm.nih.gov/25518851/)]
18. Mizuno H, Kitada K, Nakai K, Sarai A. PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics* 2009 Apr 24;2:18 [[FREE Full text](#)] [doi: [10.1186/1755-8794-2-18](https://doi.org/10.1186/1755-8794-2-18)] [Medline: [19393097](https://pubmed.ncbi.nlm.nih.gov/19393097/)]
19. SurvExpress. URL: <http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp> [accessed 2020-04-08]
20. PROGgene. URL: <http://www.compbio.iupui.edu/proggene> [accessed 2020-04-08]
21. PrognoScan. URL: <http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html> [accessed 2020-04-08]
22. ESurv. URL: <https://easysurv.net> [accessed 2020-04-08]
23. Broad GDAC Firebrowse. URL: <https://firebrowse.org> [accessed 2020-04-08]

24. GDC Data Portal. URL: <https://portal.gdc.cancer.gov/> [accessed 2020-04-08]
25. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013 Oct;45(10):1113-1120 [FREE Full text] [doi: [10.1038/ng.2764](https://doi.org/10.1038/ng.2764)] [Medline: [24071849](https://pubmed.ncbi.nlm.nih.gov/24071849/)]
26. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Cancer Genome Atlas Research Network, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 2018 Apr 05;173(2):400-416.e11 [FREE Full text] [doi: [10.1016/j.cell.2018.02.052](https://doi.org/10.1016/j.cell.2018.02.052)] [Medline: [29625055](https://pubmed.ncbi.nlm.nih.gov/29625055/)]
27. Bennett D. How can I deal with missing data in my study? *Aust N Z J Public Health* 2001 Oct;25(5):464-469. [Medline: [11688629](https://pubmed.ncbi.nlm.nih.gov/11688629/)]
28. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011 May 10;30(10):1105-1117 [FREE Full text] [doi: [10.1002/sim.4154](https://doi.org/10.1002/sim.4154)] [Medline: [21484848](https://pubmed.ncbi.nlm.nih.gov/21484848/)]
29. Goh TS, Lee JS, Il Kim J, Park YG, Pak K, Jeong DC, et al. Prognostic scoring system for osteosarcoma using network-regularized high-dimensional Cox-regression analysis and potential therapeutic targets. *J Cell Physiol* 2019 Aug;234(8):13851-13857. [doi: [10.1002/jcp.28065](https://doi.org/10.1002/jcp.28065)] [Medline: [30604867](https://pubmed.ncbi.nlm.nih.gov/30604867/)]
30. Pak K, Kim YH, Suh S, Goh TS, Jeong DC, Kim SJ, et al. Development of a risk scoring system for patients with papillary thyroid cancer. *J Cell Mol Med* 2019 Apr;23(4):3010-3015 [FREE Full text] [doi: [10.1111/jcmm.14208](https://doi.org/10.1111/jcmm.14208)] [Medline: [30729678](https://pubmed.ncbi.nlm.nih.gov/30729678/)]
31. Benedix F, Kube R, Meyer F, Schmidt U, Gastinger I, Lippert H, Colon/Rectum Carcinomas (Primary Tumor) Study Group. Comparison of 17,641 patients with right- and left-sided colon cancer: differences in epidemiology, perioperative course, histology, and survival. *Dis Colon Rectum* 2010 Jan;53(1):57-64. [doi: [10.1007/DCR.0b013e3181c703a4](https://doi.org/10.1007/DCR.0b013e3181c703a4)] [Medline: [20010352](https://pubmed.ncbi.nlm.nih.gov/20010352/)]
32. Hansen IO, Jess P. Possible better long-term survival in left versus right-sided colon cancer - a systematic review. *Dan Med J* 2012 Jun;59(6):A4444. [Medline: [22677242](https://pubmed.ncbi.nlm.nih.gov/22677242/)]
33. Kim S, Paik HY, Yoon H, Lee JE, Kim N, Sung M. Sex- and gender-specific disparities in colorectal cancer risk. *World J Gastroenterol* 2015 May 07;21(17):5167-5175 [FREE Full text] [doi: [10.3748/wjg.v21.i17.5167](https://doi.org/10.3748/wjg.v21.i17.5167)] [Medline: [25954090](https://pubmed.ncbi.nlm.nih.gov/25954090/)]
34. Lieberman DA, Williams JL, Holub JL, Morris CD, Logan JR, Eisen GM, et al. Race, ethnicity, and sex affect risk for polyps >9 mm in average-risk individuals. *Gastroenterology* 2014 Aug;147(2):351-8; quiz e14-5 [FREE Full text] [doi: [10.1053/j.gastro.2014.04.037](https://doi.org/10.1053/j.gastro.2014.04.037)] [Medline: [24786894](https://pubmed.ncbi.nlm.nih.gov/24786894/)]
35. Tian S. Identification of Subtype-Specific Prognostic Genes for Early-Stage Lung Adenocarcinoma and Squamous Cell Carcinoma Patients Using an Embedded Feature Selection Algorithm. *PLoS One* 2015;10(7):e0134630 [FREE Full text] [doi: [10.1371/journal.pone.0134630](https://doi.org/10.1371/journal.pone.0134630)] [Medline: [26226392](https://pubmed.ncbi.nlm.nih.gov/26226392/)]

Abbreviations

- AUC:** area under the curve
Elastic net: elastic net regularization
ESurv: Easy, Effective, and Excellent Survival analysis tool
Lasso: least absolute shrinkage and selection operator
miRNA: microRNA
mRNA: messenger RNA
Net: network-regularized high-dimensional Cox-regression
ROC: receiver operating characteristics
TCGA: The Cancer Genome Atlas

Edited by G Eysenbach; submitted 02.09.19; peer-reviewed by S Suh, N Borcharding; comments to author 22.10.19; revised version received 13.11.19; accepted 25.03.20; published 05.05.20

Please cite as:

Pak K, Oh SO, Goh TS, Heo HJ, Han ME, Jeong DC, Lee CS, Sun H, Kang J, Choi S, Lee S, Kwon EJ, Kang JW, Kim YH
A User-Friendly, Web-Based Integrative Tool (ESurv) for Survival Analysis: Development and Validation Study
J Med Internet Res 2020;22(5):e16084
URL: <https://www.jmir.org/2020/5/e16084>
doi: [10.2196/16084](https://doi.org/10.2196/16084)
PMID: [32369034](https://pubmed.ncbi.nlm.nih.gov/32369034/)

©Kyoungjune Pak, Sae-Ock Oh, Tae Sik Goh, Hye Jin Heo, Myoung-Eun Han, Dae Cheon Jeong, Chi-Seung Lee, Hokeun Sun, Junho Kang, Suji Choi, Soohwan Lee, Eun Jung Kwon, Ji Wan Kang, Yun Hak Kim. Originally published in the Journal of

Medical Internet Research (<http://www.jmir.org>), 05.05.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.