Computers in Biology and Medicine 137 (2021) 104793

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

SmulTCan: A Shiny application for multivariable survival analysis of TCGA data with gene sets

Ayse Ozhan [a], Melike Tombaz [b], Ozlen Konu [a,b,c,*]

[a] UNAM-National Nanotechnology Research Center, Institute of Material Science and Nanotechnology, Bilkent University, Ankara, Turkey
[b] Department of Molecular Biology and Genetics, Faculty of Science, Bilkent University, Ankara, Turkey
[c] Interdisciplinary Graduate Program in Neuroscience, Aysel Sabuncu Brain Research Center, Bilkent University, Ankara, Turkey

ARTICLE INFO

Keywords:
Survival
CPH
Shiny
Netrins
K-M
Elastic net
Prognosis
TCGA

ABSTRACT

Background: Survival analysis is widely used in cancer research, and although several methods exist in R, there is the need for a more interactive, flexible, yet comprehensive online tool to analyze gene sets using Cox proportional hazards (CPH) models. The web-based Shiny application (app) SmulTCan extends existing tools to multivariable CPH models of gene sets—as exemplified using the netrins and their receptors (netrins-receptors). It can be used to identify survival gene signatures (GSs) and select the best subsets of input gene, microRNA, methylation level, and copy number variation sets from the Cancer Genome Atlas (TCGA).
Objectives: To create a tool for CPH model building and best subset selection, using survival data from TCGA with input gene expression files from UCSC Xena. Furthermore, we aim to analyze the input TSV file of netrins-receptors in SmulTCan and discuss our findings.
Methods: SmulTCan uses Shiny's reactivity with built-in R functions from packages for CPH model analysis and best subset selection including "survminer", "riskRegression", "rms", "glmnet", and "BeSS".
Results: Results from the SmulTCan app with the netrins-receptors gene set indicated unique hazard ratio GSs in certain renal and neural cancers, while the best subsets for this gene set, obtained via the app, could differentiate between prognostic outcomes in these cancers.
Availability: SmulTCan is available at http://konulabapps.bilkent.edu.tr:3838/SmulTCan/. The input file for netrins-receptors is available in the online version of this paper. TCGA dataset folders containing survival files are available through https://github.com/aozh7/SmulTCan/.
Supplementary information: The supplementary information (SI) accompanies the online version of this article.

1. Introduction

In recent years, web-based tools developed using the Shiny package from R [1] have become prevalent due to their availability, interactivity, robustness, and relative simplicity of development. They are frequently used as environments for database communication and analysis—with applications in a multitude of research areas. Recent examples of Shiny applications (apps) include tools for a variety of subjects, such as the IPDmada tool for individual patient data meta-analyses [2]; Metabolite-Investigator, which incorporates metabolomics data and facilitates metabolite identification in user-input disease datasets [3]; and Quickomics that allows for multidimensional statistical analysis of omics data [4].

Survival analysis is widely used in biomedical research, particularly for identifying the effects of continuous and categorical factors on survival rates for different types of cancers, as well as calculating and comparing their hazard ratios (HRs). Currently, available tools include the Shiny apps ECCDIA—used for Kaplan-Meier (K-M) and Cox regression of esophageal cancer samples [5]—and MEPHAS, for generalized statistical analyses with a wide range of user-input data [6]. In addition, survival analysis has also been used disease-specifically, e.g., the Shiny app brain-coX offers gene prioritization for neuropsychiatric disorders [7].

The Cancer Genome Atlas (TCGA), a large-scale omics project for cancer, contains transcriptomics, genomics, and epigenomics datasets along with clinical and survival data (https://portal.gdc.cancer.gov/) [8]. Multiple outlets have been developed for TCGA data visualization and reuse, including the cBioPortal for Cancer Genomics (https://www.

[*] Corresponding author. Department of Molecular Biology and Genetics, Faculty of Science, Bilkent University, Ankara, Turkey.
E-mail address: konu@fen.bilkent.edu.tr (O. Konu).

https://doi.org/10.1016/j.compbiomed.2021.104793
Received 13 June 2021; Received in revised form 20 August 2021; Accepted 21 August 2021
Available online 25 August 2021

cbioportal.org) [9], University of California, Santa Cruz (UCSC) Xena (http://xena.ucsc.edu) [10], and Firehose (https://gdac.broadinstitute.org/) [11].

Moreover, several apps using TCGA data specific for survival-analysis of a single input gene are also present in the literature. GEPIA (http://gepia.cancer-pku.cn/) is one such web-based tool that provides univariableand interactive tool for univariableKaplan Meier (K-M) plots along with exploratory graphics, e.g., principal components analysis (PCA), built with HTML5 and JavaScript [12]. PROGgeneV2 (http://www.progtools.net/gene/) [13]—which accepts genes, gene ratios, or gene signatures (GSs) as input to generate expression-based K-M plots—is another tool for analyzing TCGA and the Gene Expression Omnibus (GEO) [14] datasets. The online tool KM plotter (https://kmplot.com/analysis/) [15] is widely used for generating K-M plots using gene and microRNA (miRNA) expression levels with TCGA datasets—so is UALCAN (http://ualcan.path.uab.edu) [16]. TRGAted (https://nborcherding.shinyapps.io/TRGAted/) [17] is another useful and interactive tool for univariate protein-level based survival analysis, which can be used to determine biomarkers across TCGA Pan-Cancer (TCGA--PANCAN). CVCDAP (https://omics.bjcancer.org/cvcdap/home.do) [18] performs survival analysis with a selected gene, while the app can build CPH models when clinical variables are added.

In addition to the above, SurvExpress (http://bioinformatica.mty.itesm.mx/SurvExpress) [19] and SurvMicro (http://bioinformatica.mty.itesm.mx/SurvMicro) [20] are tools that can be used for the multivariable survival analysis of gene and miRNA sets, respectively. Nevertheless, to our knowledge, there is no interactive online app in the literature that can perform multivariable survival analysis, which incorporates recent best subset selection methods with genes and miRNAs, and gene-level copy number variations (CNVs) and methylation β-values, with data from TCGA-PANCAN.

In this study, we developed a web-based Shiny app called SmulTCan (available online from http://konulabapps.bilkent.edu.tr:3838/SmulTCan/) that accepts gene expression (as well as miRNA expression and gene-level CNV and methylation) files as input, including cancer and sample type information downloaded from UCSC Xena by the user. Once a Xena TSV file is uploaded to the app, users can analyze their genes of interest's multivariable survival and HR profiles across 33 TCGA-PANCAN datasets whose survival data are embedded in the app. SmulTCan makes use of the Cox proportional hazards (CPH) model to analyze the HR GSs of input genes interactively with respect to a selected TCGA dataset from a drop-down menu. The CPH model can then be validated, and the best subset can be used in the prognostic index (PI) and K-M analyses for low vs. high-risk prognoses.

For the use case with SmulTCan, we chose the netrin family of axon guidance molecules and their receptors from the DCC and UNC-5 families [21] since their involvement in cancer progression has gained recent attention. Netrins promote cell survival, proliferation, and differentiation; they are also involved in migration, invasion, and angiogenesis [22]. In a study of netrin mutation, expression, and methylation profiles across TCGA-PANCAN, several netrins were identified as potential diagnostic markers for endocrine tumors [23]. However, an approach focusing on multivariable, multi-cancer survival analysis remains unexplored with this gene family. The gene expression dataset of netrins and their receptors (netrins-receptors) consisted of the six netrins *NTN1*, *NTN2*, *NTN3*, *NTN4*, *NTN5*, *NTNG1*, and *NTNG2*; the receptors of the UNC-5 family *UNC5A*, *UNC5B*, *UNC5C*, and *UNC5D*; and receptors *DCC* and *DSCAM* [21]. Altogether the dataset comprised twelve genes. The use of the netrins-receptor gene set led to identifying novel cancer-specific prognostic HR GSs using SmulTCan. In addition, selecting the most important gene subset for cancer prognosis out of the twelve genes can provide potential target genes for *in vivo* validation experiments.

2. Methods & app architecture

2.1. Input files

The required input file format for SmulTCan is presented in the first ten rows of Table 1, which shows the TSV file containing gene expressions of netrins downloaded from UCSC Xena. The first two columns of the file house sample names, followed by columns for each of the selected genes containing $\log_2(norm\_value + 1)$ expressions for each of the selected genes. Lastly, are the phenotypic input columns "cancer type abbreviation" and "sample_type". Our example file comprises expressions for the full set of twelve netrins-receptors, and the study we selected from UCSC Xena was TCGA-PANCAN. This column structure should also be followed when working with miRNA, CNV, and methylation β-value sets (all downloadable from UCSC Xena). When working with methylation β-values that range between 0 and 1, we recommend the values to be multiplied by 100 before uploading the TSV file on SmulTCan.

2.2. Workflow & functionality

At the core of SmulTCan are the TCGA survival data embedded in the app and used to build CPH models for the analysis tabs (Fig. 1). Four types of survival data: overall survival (OS), disease-specific survival (DSS), disease-free interval (DFI), and the progress-free interval (PFI), are stored as TXT files for each of the 33 TCGA-PANCAN datasets (https://www.cancer.gov/tcga). These files have been downloaded using the "UCSCXenaTools" package from Bioconductor [24] and are stored in separate folders within the app. First, the TXT files are read into the app with the "readr" [25] function read_delim() upon initiation and the custom global set.surv() function is used to select the chosen dataset's survival file in the app. Next, CPH model data are built within the selected sub-tab out of the four survival types and then merged with selected gene expressions from the input file. The CPH model data are updated whenever the user changes the dataset from the ''Dataset'' drop-down menu, or the number or names of input genes from the side-bar panel; this, in turn, updates the plot and table outputs from all the main tabs.

The app can be divided into two modules: model analysis and best subset selection. The model analysis "Forest plot", "Schoenfeld plots", "ROC", "ANOVA", and "Validation" tabs all rely on the CPH model created by the function coxph() from the "survival" package in R [26]. This reliance also applies to the main "CPH" tab of the best subset selection part of the app. These tabs are reactively connected to the same CPH model (Fig. 1) that can be interactively displayed and analyzed. Default parameters of the coxph() function, including the Efron approximation for handling ties, are used. Input gene expressions from the CPH model can be viewed reactively as a table in the "Data table" tab in the main panel, while displays of the expression distributions are found in the "Boxplot" tab (Fig. 1).

In the "Forest plot" tab (Fig. 1), the ggforest() function from the "survminer" package [27] is used to display the HR profile of the CPH model. The model is used inside the cox.zph() function, and residuals are plotted for the input genes using the ggcoxzph() function from the "survminer" in the "Schoenfeld plots" tab. In the "ROC" tab, the Score.list() function of the "riskRegression" package [28] is used for generating the receiver operating characteristic (ROC) plots with area under the ROC curve (AUC) metrics (Fig. 1). The CPH model from coxph() is used reactively inside the cph() function with the "surv" parameter set to TRUE, which in turn is used inside anova() and validate() (all three are functions from the "rms" package [29]), in the "ANOVA" and "Validation" tabs, respectively. The default parameters for the validate() function include bootstrapping with 40 repetitions and the Akaike information criterion (AIC) with cutoff 0 as the stopping rule. The "ROC", "Validation" and "ANOVA" tabs contain the "Plot" and "Stats" sub-tabs for the CPH model's results (Fig. 1). ANOVA plots and plots of

Table 1
First ten TCGA samples of the required input file, as downloaded from UCSC Xena with $\log_2(norm\_value + 1)$ expressions of netrins, as well as columns for the TCGA dataset which the sample belongs to and the type of the tumor sample.

sample	samples	NTN1	NTN3	NTN4	NTN5	NTNG1	NTNG2	cancer type abbreviation	sample_type
TCGA-DX-A48L-01	TCGA-DX-A48L-01	15.27	3.72	7.91	1.52	0	3	SARC	Primary Tumor
TCGA-DX-AB2H-01	TCGA-DX-AB2H-01	14.98	3.9	7.36	4.14	1.88	5.39	SARC	Primary Tumor
TCGA-DU-6404-02	TCGA-DU-6404-02	14.29	4.97	10	3	8.38	11.47	LGG	Recurrent Tumor
TCGA-DX-A2IZ-01	TCGA-DX-A2IZ-01	14.29	3.15	8.96	4.95	4.37	10.36	SARC	Primary Tumor
TCGA-DX-A48R-01	TCGA-DX-A48R-01	14.15	3.52	6.84	0	7.71	1.82	SARC	Primary Tumor
TCGA-3B-A9HR-01	TCGA-3B-A9HR-01	13.95	2.25	8.09	1.32	1.32	1.53	SARC	Primary Tumor
TCGA-PC-A5DN-01	TCGA-PC-A5DN-01	13.8	4.1	9.46	4.17	5.3	5.28	SARC	Primary Tumor
TCGA-S9-A6WN-01	TCGA-S9-A6WN-01	13.73	4.39	5.63	2.85	7.19	8.46	LGG	Primary Tumor
TCGA-ZB-A969-01	TCGA-ZB-A969-01	13.71	2.94	10.51	5.05	6.67	5.16	THYM	Primary Tumor
TCGA-CQ-6221-01	TCGA-CQ-6221-01	13.7	2.44	9.96	3.79	10.26	4.62	HNSC	Primary Tumor

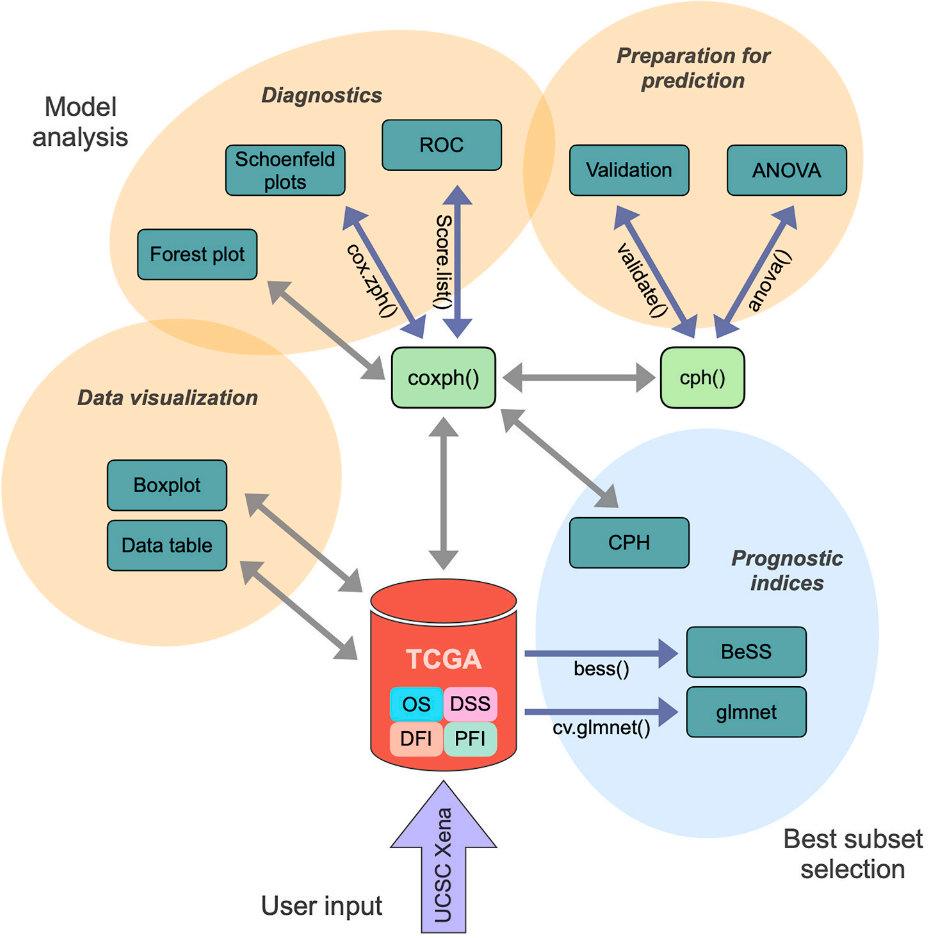


Fig. 1. Diagram of a simplified version of SmulTCan's architecture, consisting of the two modules: model analysis and best subset selection. The model analysis part can in turn be thought of as made up of three smaller submodules, which are for: displaying the data inside the CPH model built with coxph(), diagnostics of this CPH model and preparation for prediction using this model. Each main tab, within olive green boxes, is indicated inside the submodule they belong to in the figure. The app accepts UCSC Xena output files containing normalized gene expressions and incorporates the information of the input file with its embedded data from TCGA upon upload. The coxph() function, with its default parameters, is at the center of the diagnostics and preparation for prediction parts, through which the model analysis main tabs "Forest plot", "Schoenfeld plots" and "ROC" are reactively connected to their CPH model and its associated data; while the "Validation" and "ANOVA" tabs connect to the central function through the additional cph() function. Additionally, the "CPH" tab of the best subset selection part is also reactively connected to its model through the central coxph(). Reactivity in the figure is indicated with double-headed arrows. The best subset selection tabs "BeSS" and "glmnet" build their individual CPH models internally from the embedded TCGA data and user-input file. Reactivity is stopped in the "BeSS" and "glmnet" tabs, indicated with single-headed arrows. Main functions for the required analyses of the main tabs are indicated under the blue arrows the tab names are connected to. The packages which the functions in the figure belong to are indicated in the text and listed in the References, as well as additional visualization and data analysis functions from existing packages and built-in functions that make use of them omitted in the figure. Detailed information about the outputs and results that can be produced from the main tabs is given in the text.

selected gene coefficients are drawn using the "ggplot2" package and found in the best subset selection tabs [30].

The second module of SmulTCan involves selecting gene subsets for best predicting survival (module with light blue background in Fig. 1). In the "CPH" tab of the main panel, coefficients of the CPH model with $p < 0.05$ can be used to determine the best subsets from the input gene combinations in prognostic outcome analyses. In the "glmnet" tab, we use an R package with the same name to interactively visualize the best subset of input genes using the elastic net method (Fig. 1). The elastic net slider starts at 1 by default, corresponding to lasso; the user can then adjust the slider values 0–1 where 0 is equal to ridge regression [31]. By default, the "glmnet" package uses 10-fold cross-validation, where the number of folds can be set to the sample size from the "Folds" menu.

The cv.glmnet() function with the "family" parameter set to "cox" [31] is used inside the built-in function in global.R. For the default 10-fold cross-validation, the createFolds() function from the "caret" package [32] is implemented inside this built-in function to generate a "foldid" parameter with the seed set to 123 in the cv.glmnet() function. This step is added to prevent variability of the results between sub-tabs when 10-fold cross-validation is used. It is important to note that the "glmnet" tab builds and analyzes its own CPH model using Breslow approximation for tied events [31].

In the "BeSS" tab, a method for best subset selection is implemented from the "Bioconductor" package with the same name [33]; the bess() function with "family" parameter set to "cox" is used for best subset selection with ridge regression (Fig. 1). Users can select between the default method "gsection" and "sequential" for determining the optimal model size. Additionally, users can filter coefficients from the "BeSS" tab based on significance (i.e., keep only those with $p < 0.05$). The "BeSS" tab also builds its own CPH model, though the package internally relies on coxph() with the function's default parameters.

The isolate() function of Shiny is used to stop reactivity in the "BeSS" and "glmnet" tabs to stabilize their responsiveness. An individual PI value is calculated for each sample in a selected dataset based on the formula from Xue et al. [34], using coefficients associated with the genes of the best subset in each of the selection tabs. Samples with a PI value greater than the median PI of the dataset are labeled "High" risk, while remaining samples are labeled "Low" risk. ROC plots in the "Risk ROC" sub-tabs while the Low:High risk ratios in the "K-M data" sub-tabs in the best subset selection main tabs are calculated with the coxph() function with respect to the computed PI covariate of the selected dataset and genes.

2.3. *Warnings, downloads & help*

All analyses in SmulTCan are carried out using only primary (or primary blood-derived from acute myeloid leukemia [LAML]) tumor samples for the selected TCGA dataset. The correct format of the UCSC Xena output TSV file is required to initiate SmulTCan. Warnings in the SmulTCan app are intended to protect users from erroneous calculations that might arise from CPH models not converging or leading to infinite coefficients. For example, users are warned when a selected dataset cannot be found, when it lacks information for a survival type or when there are missing expressions for one or more input genes in the TSV file for the selected dataset. In addition, users would also be informed if a best subset selection method cannot find any genes for a selected dataset.

SmulTCan currently supports TSV for downloading tables; and PDF or high-resolution PNG formats, depending on the tab, for downloading figures. There are download buttons in each of the four survival sub-tabs of all tabs. In addition, the SmulTCan app contains a "Walk-though" button at the top left of its interface to guide users from file upload to gene selection. Information relating to the app can be found in the "About" tab of the main panel. Moreover, this paper's supplementary information (SI) can be used as a manual (Supplementary Figs. S1–S6) and includes a demonstration of the app with the netrins-receptor example input file.

3. Results

In this section, we present the SmulTCan features comparison with existing multivariable survival analysis tools and briefly present the results of our demonstration of the SmulTCan app. We used a TSV file containing expression data from twelve netrins-receptors as input (the demo TSV file). All twelve genes were used as input in the app's "Genes" menu. Several interesting findings are presented.

3.1. *Comparison of SmulTCan with existing univariable and multivariable tools*

We compared the main features of relevant univariable survival analysis tools with respect to expression levels in the literature with those of SmulTCan. Our comparison revealed that the most distinguishing characteristic of our app is its multivariable nature in terms of expression levels and its reactivity (Table 2; see Zheng et al. [35] for a more detailed review). We further demonstrated that SmulTCan has novel features that are complementary to existing online multivariable survival analysis tools like SurvExpress and SurvMicro, which take sets of genes and miRNAs as input and use the coxph() function from the "survival" package. SurvExpress and SurvMicro multivariable tools rely on a PI generated from the CPH model coefficients from all input molecules; it can be visualized through K-M plots, heatmaps, ROC plots, risk group expression analyses, and risk optimization plots (Table 3). On the other hand, SmulTCan offers direct interactive survival analyses with expression levels of selected genes, visualized through forest, ROC, ANOVA, and validation plots (Table 3).

Both SurvExpress and SurvMicro can prioritize gene and miRNA subsets, respectively, from significant coefficients of CPH models constructed with coxph(), whereas SmulTCan also offers the interactive elastic net and the BeSS algorithm for best subset selection. Moreover, only SmulTCan performs downstream risk analyses of the best subset's PI, which can then be visualized through K-M, cumulative hazard, and ROC plots (Table 3). Moreover, SmulTCan allows the use of mRNA or miRNA expressions and gene-level CNV and methylation data from the same app. On the other hand, SurvExpress and SurvMicro provide options for visualization of gene expression distributions in low vs. high-risk sample groups, determined by the PI of the coxph() function's coefficients, with boxplots, clustering heatmaps, and risk group optimization curves. They also incorporate expression datasets other than TCGA.

Users can view and download survival data tables with TCGA samples and expressions from the "Data table" tab of SmulTCan, which they have worked on with other model analysis tabs in the app. This output table generated by SmulTCan for a selected cancer provides a useful resource for users who would like to use it as a training set in their command-line prediction analyses with their cancer survival test data. However, this data table with survival and expression values is neither viewable nor downloadable from SurvExpress or SurvMicro.

3.2. *Model analysis with netrins-receptors*

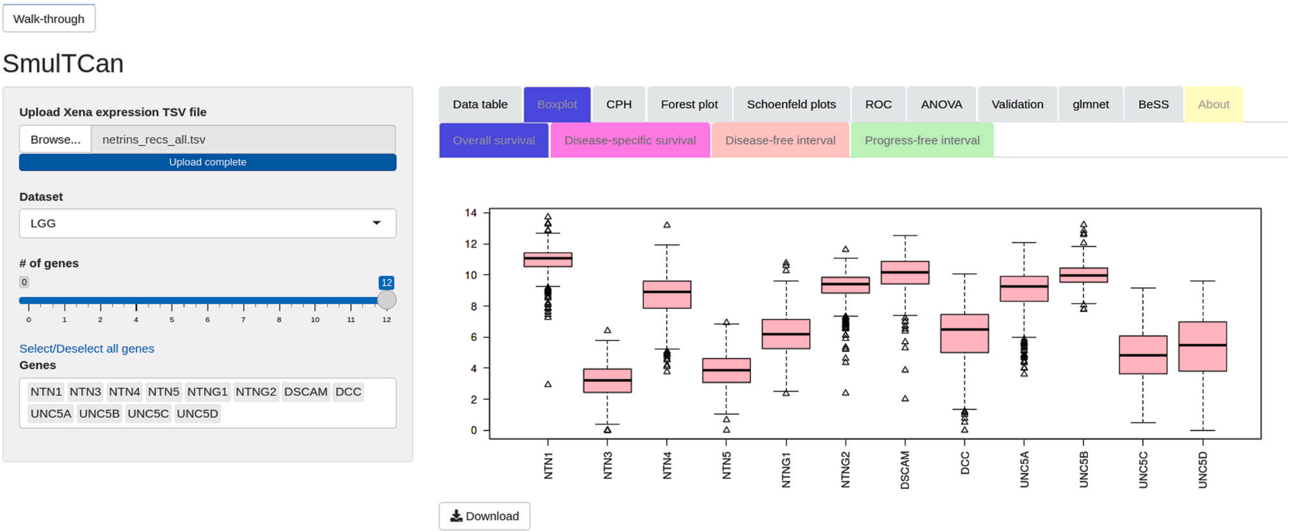
SmulTCan allows for visualization, model development and preparation for prediction using different submodules (those with a light-orange background in Fig. 1). For example, a screenshot from the "Boxplot" tab displays the expression distributions of the netrins-receptors gene set in the low grade glioma (LGG) dataset, which represents the gene expressions in the associated CPH model built using coxph() (Fig. 2). In addition, the TSV file obtained from UCSC Xena uploaded at the top left of the side-bar panel provides the names of all

Table 2
Comparison of SmulTCan features and functionalities with existing online tools for univariable survival analysis with expression levels and TCGA data. Abbreviations in the table: TCGA-PANCAN, the Cancer Genome Atlas Pan-Cancer; CPH, Cox proportional hazards; OS, overall survival; DSS, disease-specific survival; DFI, disease-free interval; PFI, progress-free interval, K-M, Kaplan-Meier.

	cBioPortal [9]	UCSC Xena [10]	GEPIA [12]	PROGgeneV2 [13]	KM plotter [15]	UALCAN [16]	TRGAted [17]	CVCDAP [18]	SmulTCan
TCGA-PANCAN	Yes, others	Yes	Yes	Yes, others	In development	Yes, others	Yes	Yes, others	Yes
Input file required	No	No	No	No	No	No	No	No	Yes
Survival types (Pan-Cancer)	OS, DSS, DFI, PFI	OS, DSS, DFI, PFI	OS, DFI	OS	OS, DFI	OS	OS, DSS, DFI, PFI	OS, DSS, DFI, PFI	OS, DSS, DFI, PFI
Main survival analyses	K-M	K-M	K-M	K-M	K-M	K-M	K-M, CPH	K-M, CPH	K-M, CPH, others

Table 3
Comparison of SmulTCan with currently available online multivariable survival analysis tools SurvMicro and SurvExpress, which work with gene and miRNA expressions, respectively. Abbreviations in the table: K-M, Kaplan-Meier; ROC, receiver operating characteristic; miRNA, microRNA; CNV, copy number variation; ANOVA, analysis of variance; PI, prognostic index.

	Input file required	Interactivity & reactivity	Survival analysis	Survival visualization	Inputs	Best subset selection	PI analysis using best subset
SurvExpress [19] SurvMicro [20]	No	No	PI from coxph()	K-M, heatmaps, ROC, risk group optimization curves	Gene/miRNA expressions	coxph()	No
SmulTCan	Yes	Yes	expressions, CNVs, β-values × 100	Forest, ROC, ANOVA, validation plots	Gene/miRNA expressions, gene-level CNVs, methylation β-values	coxph(), glmnet, BeSS	Yes, visualized with K-M, ROC, cumulative hazard plots



Upload Xena expression TSV file: netrins_recs_all.tsv

Dataset: LGG

# of genes: 0 – 12

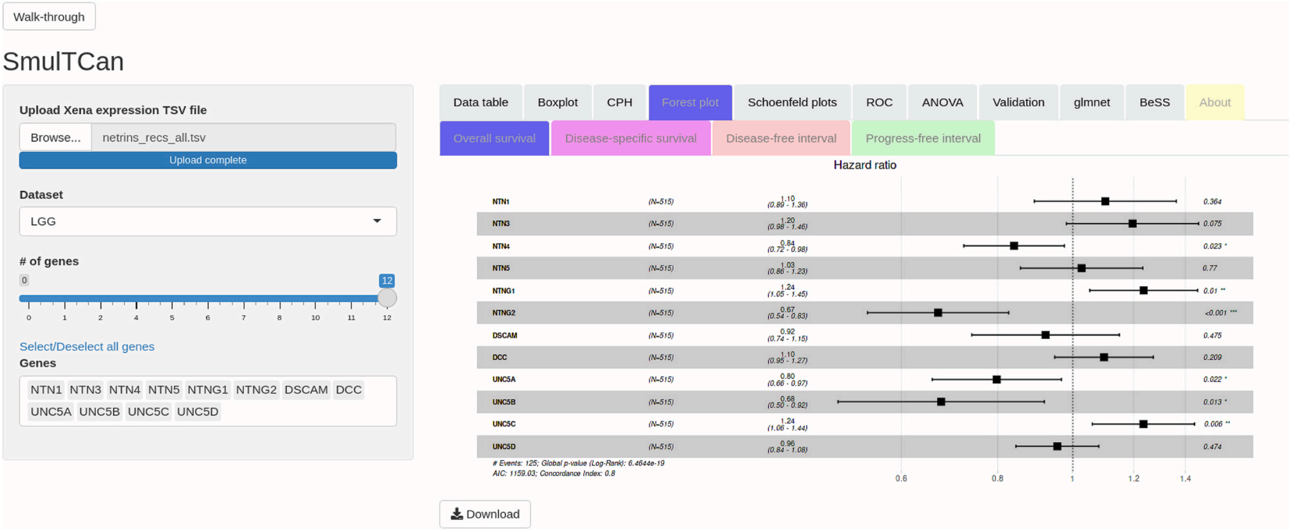
Genes: NTN1 NTN3 NTN4 NTN5 NTNG1 NTNG2 DSCAM DCC UNC5A UNC5B UNC5C UNC5D

Fig. 2. Screenshot of the "Boxplot" tab showing distributions of expressions of netrins-receptors for OS in LGG.

twelve genes of netrins-receptors in the "Genes" menu.

The screenshot for the "Forest plot" tab shown in Fig. 3 for OS in the LGG dataset reveals the HR GS of netrins-receptors. The "Forest plot" tab is highly useful for determining which genes of the gene set have significance based on the given confidence intervals and log-rank statistics. Among netrins-receptors, *NTNG2* is strongly positively associated with OS, while *NTNG1* and *UNC5C* are negatively associated. *UNC5A*, *UNC5B,* and *NTN4* are also found to be positively associated with OS in LGG. The AIC for this multivariable model is 1159.03, while its log-rank *p*-value is approximately 0.65 × 10^{-18}.

The model analysis module of SmulTCan can also provide ROC plots, which indicate the overall robustness of multivariable models [36]. For



Upload Xena expression TSV file: netrins_recs_all.tsv

Dataset: LGG

# of genes: 0 – 12

Genes: NTN1 NTN3 NTN4 NTN5 NTNG1 NTNG2 DSCAM DCC UNC5A UNC5B UNC5C UNC5D

NTN1	(N=515)	1.10 (0.89 - 1.36)	0.364
NTN3	(N=515)	1.20 (0.98 - 1.46)	0.075
NTN4	(N=515)	0.84 (0.72 - 0.98)	0.023 *
NTN5	(N=515)	1.03 (0.86 - 1.23)	0.77
NTNG1	(N=515)	1.24 (1.05 - 1.45)	0.01 **
NTNG2	(N=515)	0.67 (0.54 - 0.83)	<0.001 ***
DSCAM	(N=515)	0.92 (0.74 - 1.15)	0.475
DCC	(N=515)	1.10 (0.95 - 1.27)	0.209
UNC5A	(N=515)	0.80 (0.66 - 0.97)	0.022 *
UNC5B	(N=515)	0.68 (0.50 - 0.92)	0.013 *
UNC5C	(N=515)	1.24 (1.06 - 1.44)	0.006 **
UNC5D	(N=515)	0.96 (0.84 - 1.08)	0.474

Events: 125; Global p-value (Log-Rank): 6.4644e-19
AIC: 1159.03; Concordance Index: 0.8

Fig. 3. Screenshot of the forest plot of netrins-receptors in LGG indicates positive association with OS for genes *NTNG2*, *UNC5A*, *UNC5B*, and *NTN4* as well as negative association for *NTNG1* and *UNC5C*.

example, the ROC plot of the netrins-receptors CPH model for OS in LGG, generated from the “ROC” tab (Fig. 4), revealed the AUC% score, as 87.4, indicating that this CPH model is a good candidate for use in prediction algorithms as a training set. Accordingly, the “ROC” tab can be used to assess the predictive power of CPH models from SmulTCan, while detailed AUC% results of the models can be extracted from its “Stats” sub-tab.

The output of the “Plot” sub-tab of the “ANOVA” main tab for the input gene set’s CPH model for the DFI in LGG is given in Fig. 5. The plot indicates that *NTN4* is the strongest predictor with the lowest *p*-value ($p < 0.05$) and the highest χ^2-df value out of the netrins-receptors gene set in the DFI. The χ^2, df, and *p*-values resulting from the CPH model’s ANOVA, for all selected input genes, can be viewed as a table from the “Stats” sub-tab of the same main tab.

3.3. Best subset selection with netrins-receptors

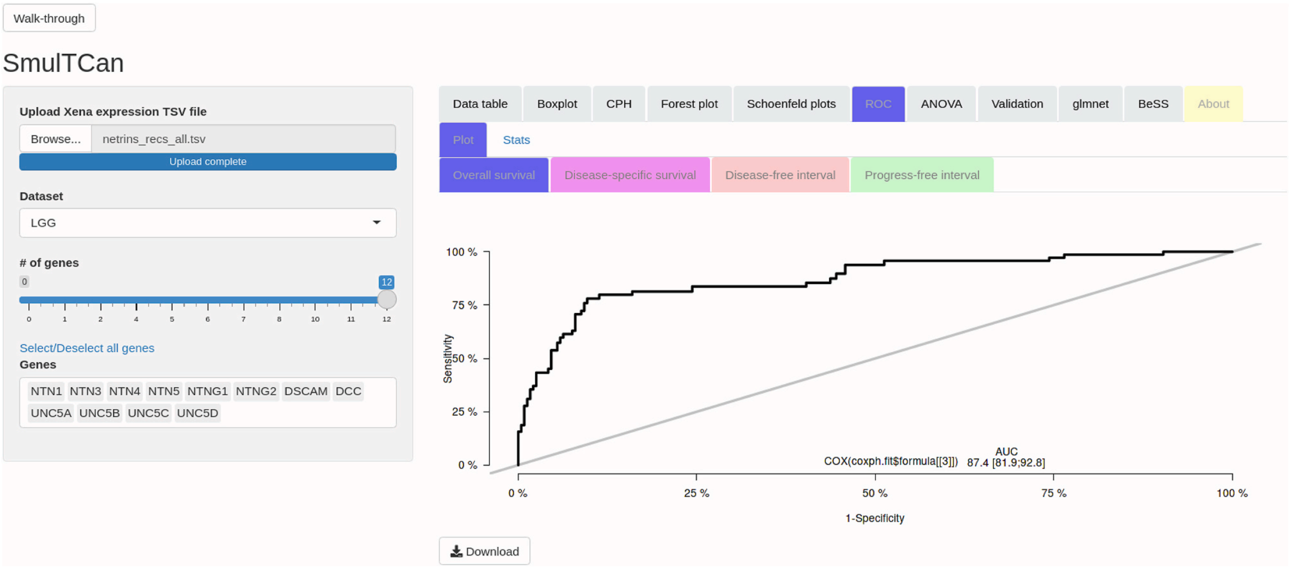
All main tabs for best subset selection contain sub-tabs with a plot of coefficients in the selection and a table of the coefficients of the best subset. Additionally, the K-M and cumulative hazards plots generated from the PI computations of the best subset (module with light blue background in Fig. 1) are included. Other sub-tabs include a table of K-M statistics with Low:High risk ratios, a table of PI calculations, and the ROC plot with the PI of the best subset. The “CV plot” sub-tab of the “glmnet” tab shows the distribution of $\log_{10}\lambda$ values, indicating the amount of regularization applied. The “GIC” sub-tab of the “BeSS” tab plots the growth incidence curve (GIC) of the L (representing the cost of inaccuracy in prediction) of the coefficients in the regression model [33].

As an example, the cumulative hazards plot generated from the “CumHz” sub-tab within the main panel’s “glmnet” tab is given in Fig. 6. The plot shows DSS for the kidney renal clear cell carcinoma KIRC dataset, in which the netrins-receptors gene set seemed to be informative (also check the SI’s model analysis part). According to this, coefficients determined with the default 10-fold cross-validation and lasso could differentiate between high-risk and low-risk prognostic outcomes ($p < 0.0001$) over approximately 11 years. The best subset producing this result was *NTN4* and *NTNG2*, with the former contributing positively and the latter negatively to prognosis. According to this best subset model, high-risk samples had a prognosis about two-times worse than low-risk samples ($p < 0.0001$), as seen from the “K-M data” sub-tab.

The plot of the coefficients found with the lasso option of the “glmnet” tab for the DSS data from KIRC, with “Folds” set to the sample size of the dataset indicated that *NTN4* and *UNC5D* contributed to low-risk prognosis, while *NTNG2* contributed to high-risk prognosis (Supplementary Fig. S7A). These three genes out of the twelve netrins-receptors were able to differentiate between the low- and high-risk prognostic outcomes ($p < 0.0001$), as seen in the K-M plot downloaded from SmulTCan in Supplementary Fig. S7B. Based on this best subset model, high-risk samples in KIRC had a prognosis about three times worse than low-risk samples ($p < 0.0001$).

The plot of the coefficients found for DSS in KIRC also using the “BeSS” tab is represented in a screenshot obtained from the app (Fig. 7). According to ridge regression results of the BeSS algorithm using the default method, the best subset of netrins-receptors consisted of genes *DSCAM*, *NTN3*, *NTN4*, *NTNG1*, *NTNG2*, *UNC5C*, and *UNC5D*. While the first six of these genes contributed to a low-risk prognosis for DSS in KIRC, the seventh contributed to a high-risk prognosis. BeSS, by default, does not exclude the insignificant coefficients from ROC analysis, for which users can include only those that are significant. Coefficients of the best subset, this time using the PFI for kidney renal papillary cell carcinoma (KIRP), computed with lasso when the number of folds is equal to the sample size, can be downloaded from the “Coeffs” sub-tab of the “glmnet” tab (Supplementary Fig. S8A). The best subset, in this case, consisted of the genes *NTN4 and NTNG1* that contributed to low-risk prognosis, and *NTNG2* and *UNC5B,* which contributed to high-risk prognosis. These four genes could differentiate between prognostic outcomes ($p = 0.0076$) though with some overlap in confidence intervals, which can be observed in the cumulative hazards plot in Supplementary Fig. S8B, downloaded from the “CumHz” sub-tab of this best subset selection method. This best subset model predicted a prognosis about two times worse for high-risk samples ($p < 0.05$) with a concordance index (CI) of 0.62 ($p < 0.05$), which could be seen from the “K-M data” sub-tab.

When applied to LGG, coefficients of the best subset for OS, selected with lasso regression and the default 10-fold cross-validation from the “glmnet” tab, showed that *NTN4*, *NTNG1,* and *UNC5A* contributed to low-risk prognosis, while *NTNG1* had a slightly positive coefficient (Supplementary Fig. S9A). Furthermore, this best subset could significantly differentiate between the two prognostic outcomes, as shown in the K-M plot in Supplementary Fig. S9B, also obtained from the app. Accordingly, SmulTCan predicted different sets of netrins-receptors associated with prognostic outcomes in different cancers.



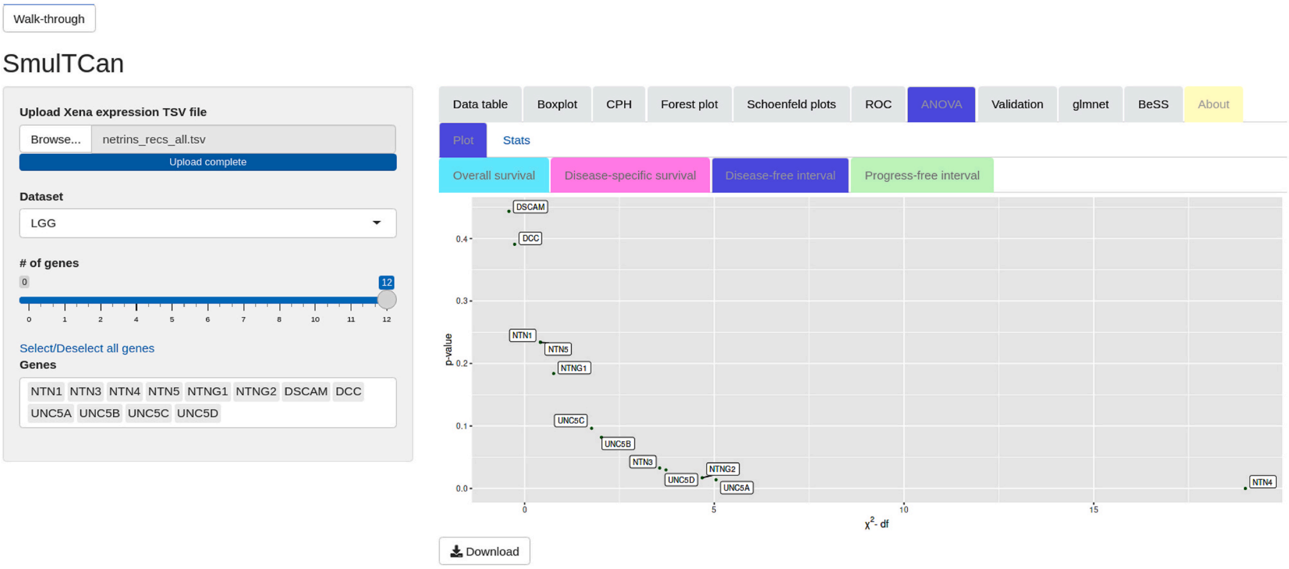
Upload Xena expression TSV file: netrins_recs_all.tsv
Upload complete

Dataset: LGG

# of genes: 0 – 12

Genes: NTN1 NTN3 NTN4 NTN5 NTNG1 NTNG2 DSCAM DCC UNC5A UNC5B UNC5C UNC5D

Fig. 4. ROC plot of the full set of netrins-receptors in LGG for OS gives an AUC% score of 87.4.



SmulTCan

Upload Xena expression TSV file

Browse... netrins_recs_all.tsv

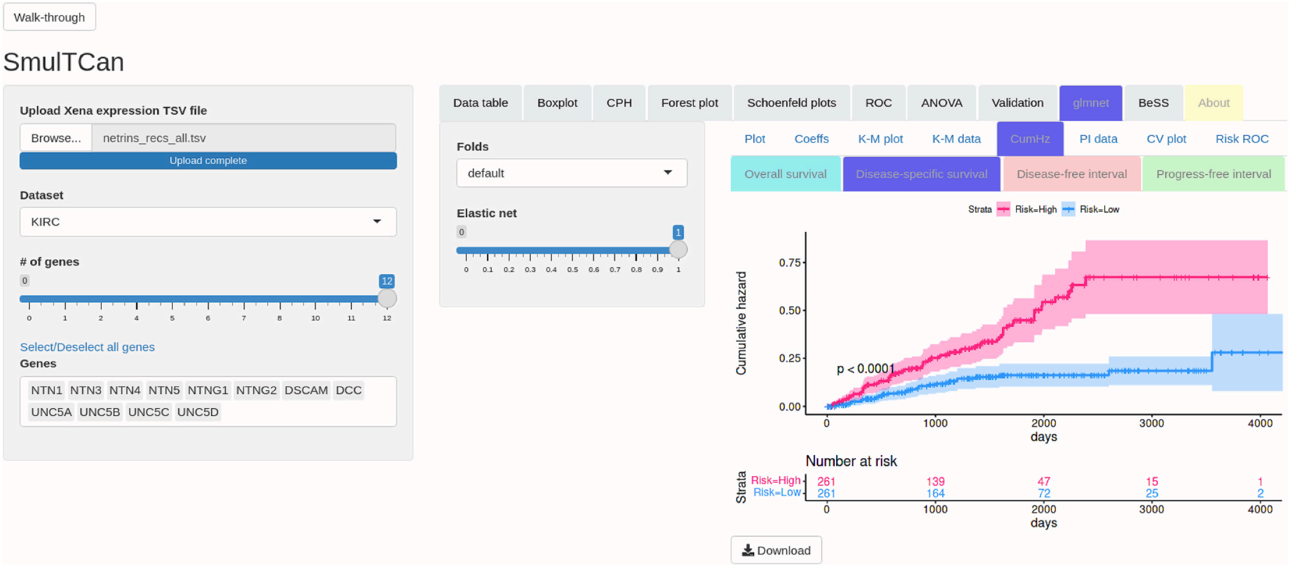
Upload complete

Dataset: LGG

# of genes: 0 – 12

Genes: NTN1 NTN3 NTN4 NTN5 NTNG1 NTNG2 DSCAM DCC UNC5A UNC5B UNC5C UNC5D

Fig. 5. Screenshot of the “Plot” sub-tab of the “ANOVA” tab for DFI in LGG of the netrins-receptors CPH model. The plot can be used to rank genes in the CPH model with respect to their χ^2-df and p-values. According to the plot, *NTN4* is the best predictor of this survival model.



SmulTCan

Upload Xena expression TSV file

Browse... netrins_recs_all.tsv

Upload complete

Dataset: KIRC

# of genes: 0 – 12

Genes: NTN1 NTN3 NTN4 NTN5 NTNG1 NTNG2 DSCAM DCC UNC5A UNC5B UNC5C UNC5D

Folds: default

Elastic net: 0 – 1

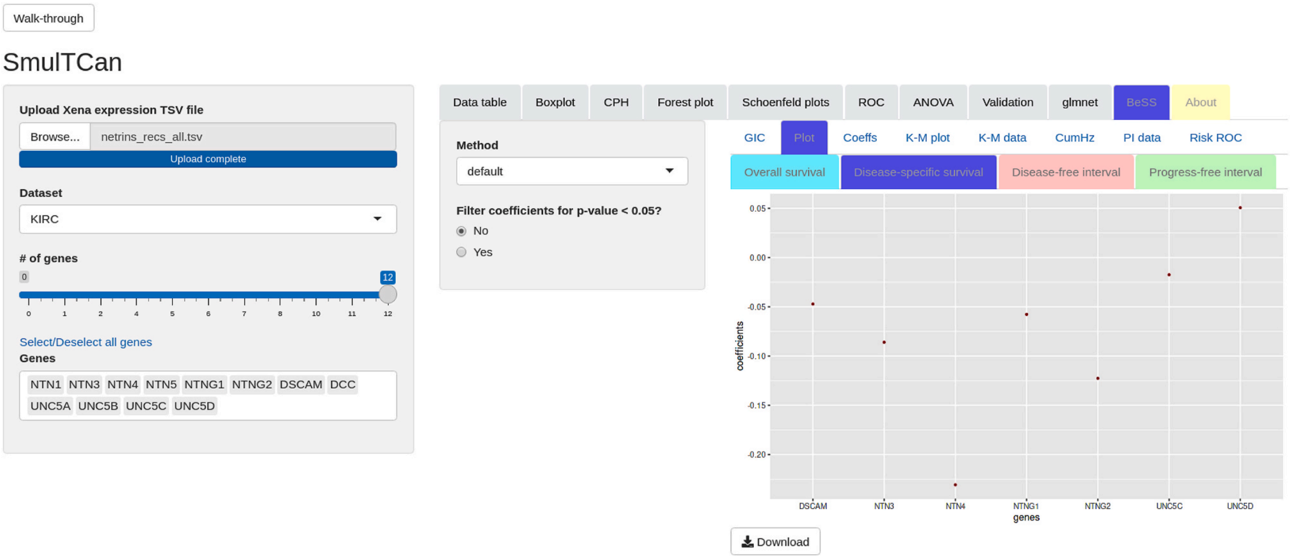
Fig. 6. Screenshot of the cumulative hazard plot of KIRC for DSS indicates distinct prognostic outcomes for low and high-risk sample groups.

4. Discussion

The SmulTCan app incorporates several multivariable survival analysis tools in a single reactive Shiny app to analyze and visualize CPH models constructed from input gene sets. Users can interactively analyze their input gene sets’ expression distributions within the CPH model and identify and analyze HR GSs from the model. They can also validate the model for further prediction analyses in selected cancers for different types of survival and predict the best subsets of genes using different methods. Moreover, a comparison of SmulTCan with existing online tools that perform survival analysis using TCGA datasets demonstrated SmulTCan’s unique, modular, and complementary nature (Tables 2 and 3).

One of the outstanding attributes of SmulTCan is its ability to reduce the gene set input to highly predictive sets using different methods. Indeed, best subset selection methods are frequently used in identifying the most important genes or biomarkers in multivariable regression models, including the CPH model. By incorporating different best subset selection methods in a single app, SmulTCan allows users to compare the methods among each other and visualize results from parameter adjustments. Additionally, in SmulTCan, we implemented the elastic net, which combines lasso and ridge regression [37]. For example, lasso has been previously used successfully in identifying a nine-gene signature for non–small cell lung cancer (NSCLC) by Gentles et al. [38], and more recently in identifying an eight-gene signature in bladder cancer for OS [39]. In general, the lasso method is stricter than ridge regression due to its diamond-shaped constraint region instead of the elliptical constraint region of ridge regression, which does not set the coefficients to zero [40]. An alternative for ridge regression in the app involves the BeSS algorithm that has been recently used in the literature [41]. The “CPH” main tab in the app extracts the best subsets of input genes directly from the CPH model, and the results can be compared with those from the “glmnet” and “BeSS” tabs.

Using SmulTCan for predicting survival with netrins-receptors, our



Upload Xena expression TSV file

Browse... netrins_recs_all.tsv

Upload complete

Dataset: KIRC

# of genes: 0 – 12

Genes: NTN1 NTN3 NTN4 NTN5 NTNG1 NTNG2 DSCAM DCC UNC5A UNC5B UNC5C UNC5D

Method: default

Filter coefficients for p-value < 0.05?

● No

○ Yes

Fig. 7. Screenshot of the "BeSS" main tab's "Plot" sub-tab shows the coefficients of the best subset of genes for DSS in KIRC. According to the BeSS algorithm's default model size selection model, the best subset genes *DSCAM*, *NTN3*, *NTN4*, *NTNG1*, *NTNG2*, and *UNC5C* have negative coefficients, while *UNC5D* has a positive coefficient.

findings indicated unique HR GSs for this gene set, especially in neural and renal cancers such as LGG and KIRC (see also Supplementary Fig. S10). In addition, the best subsets of netrins-receptors identified with lasso could significantly differentiate between high- and low-risk prognostic outcomes. Given the role of netrins and their receptors in the nervous system as axon guidance molecules, their involvement in LGG is expectable. Our results with kidney cancers KIRP and KIRC seem to corroborate those by Hao et al. [23], who studied the impact of netrins on survival via methylation studies, e.g., for *NTN4* and *NTNG1*. However, our unique multivariable approach also allows for understanding the proportional contribution of genes within the netrins-receptors gene set on survival. Therefore, the netrins-receptors and their best subsets can be analyzed further experimentally in renal cancers and LGG to determine their biological role in these disorders.

SmulTCan is a potential guide for researchers in identifying significant gene, miRNA, or gene-level CNV or methylation β-value datasets and subsets in specific TCGA datasets. It would be particularly useful in understanding the comparative role of each gene in an input gene set. The user-friendly, interactive, and layered design of the app allows for the quick comparison of different analysis results, methods, and survival types. While the app also works with user-uploaded miRNA and gene-level CNV and processed methylation files from UCSC Xena, the same database's exon expression files can only be used with the best subset selection tabs "BeSS" and "glmnet", due to independent model building within these tabs. We hope SmulTCan aids researchers in understanding the roles of gene/miRNA/CNV/methylation sets for survival.

5. Limitations & future perspectives

As with computational programs that rely on coxph(), SmulTCan might have limitations with respect to the input gene/miRNA/CNV/methylation set's combination with specific TCGA datasets' survival files causing singularity errors. In these cases, the app displays an error message. Currently, SmulTCan requires the upload of a TSV file retrieved from the UCSC Xena by the user. Future versions will allow automatic access to expression data upon entry of the gene names. We also plan to integrate expression data from multiple sources so that SmulTCan is not limited to TCGA-PANCAN datasets. Incorporation of additional functions allowing the user to filter or stratify the data based on clinical and molecular variables are also planned, along with the inclusion of newly emerging best subset methods.

6. Conclusions

We have created the online tool SmulTCan using Shiny for researchers who want to analyze the cancer survival profiles of their gene, miRNA, or gene-level CNV or methylation β-value sets, in a multivariable manner. To this end, we provide a means for interactive use of CPH models built from TCGA data embedded in the app for four different types of survival. The app's user-friendly design and reactive functionality provide interactive and fast model visualization and analysis with input gene/miRNA expression, and gene-level CNV and methylation level sets. The app contains additional tabs for different methods of best subset selection from the input gene set, including the interactive elastic net, with incorporated PI analyses of the best subset. Controls in the app provide robustness and accuracy. SmulTCan is an app that can be used by researchers from a variety of backgrounds with ranging interests. Demonstration of the app with the netrins-receptors gene set revealed prognostic HR GSs in neural cancer and several renal cancers while distinct best subsets selected from this set of twelve genes could significantly differentiate between prognostic outcomes in each of these cancers.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

Authors declare no conflict of interest.

Acknowledgments

We thank Ayse Gokce Keskus and Can Ugur Ayfer from Bilkent University for helping in setting up and maintaining the server from which SmulTCan is shared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2021.104793.

References

[1] W. Chang, J. Cheng, J.J. Allaire, C. Sievert, B. Schloerke, Y. Xie, J. Allen, J. McPherson, A. Dipert, B. Borges, Shiny: web application framework for r, 2021. https://CRAN.R-project.org/package=shiny.

[2] J. Wang, W.R. Keusters, L. Wen, M.M.G. Leeflang, IPDmada: an R Shiny tool for analyzing and visualizing individual patient data meta-analyses of diagnostic test accuracy, Res. Synth. Methods 12 (1) (2021) 45–54, https://doi.org/10.1002/jrsm.1444.

[3] C. Beuchel, H. Kirsten, U. Ceglarek, M. Scholz, Metabolite-Investigator: an integrated userfriendly workflow for metabolomics multi-study analysis, Bioinformatics (2020), https://doi.org/10.1093/bioinformatics/btaa967.

[4] B. Gao, J. Zhu, S. Negi, X. Zhang, S. Gyoneva, F. Casey, R. Wei, B. Zhang, Quickomics: exploring omics data in an intuitive, interactive and informative manner, Bioinformatics (2021), https://doi.org/10.1093/bioinformatics/btab255.

[5] J. Yang, J. Shang, Q. Song, Z. Yang, J. Chen, Y. Yu, L. Shi, ECCDIA: an interactive web tool for the comprehensive analysis of clinical and survival data of esophageal cancer patients, BMC Canc. 20 (1) (2020) 985, https://doi.org/10.1186/s12885-020-07479-9.

[6] Y. Zhou, S.W. Leung, S. Mizutani, T. Takagi, Y.S. Tian, MEPHAS: an interactive graphical user interface for medical and pharmaceutical statistical analysis with R and Shiny, BMC Bioinf. 21 (1) (2020) 183, https://doi.org/10.1186/s12859-020-3494-x.

[7] S. Freytag, R. Burgess, K.L. Oliver, M. Bahlo, brain-coX: investigating and visualising gene co-expression in seven human brain transcriptomic datasets, Genome Med. 9 (1) (2017) 55, https://doi.org/10.1186/s13073-017-0444-y.

[8] C. Hutter, J.C. Zenklusen, The Cancer Genome Atlas: creating lasting value beyond its data, Cell 173 (2) (2018) 283–285, https://doi.org/10.1016/j.cell.2018.03.042.

[9] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C. J. Byrne, M.L. Heuer, E. Larsson, Y. Antipin, B. Reva, A.P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Canc. Discov. 2 (5) (2012) 401–404, https://doi.org/10.1158/2159-8290.CD-12-0095. Erratum in: Canc. Discov. 2(10) (2012) 960.

[10] M.J. Goldman, B. Craft, M. Hastie, K. Repecka, F. McDade, A. Kamath, A. Banerjee, Y. Luo, D. Rogers, A.N. Brooks, J. Zhu, D. Haussler, Visualizing and interpreting cancer genomics data via the Xena platform, Nat. Biotechnol. 38 (6) (2020) 675–678, https://doi.org/10.1038/s41587-020-0546-8.

[11] N. Gehlenborg, M.S. Noble, G. Getz, L. Chin, P.J. Park, Nozzle: a report generation toolkit for data analysis pipelines, Bioinformatics 29 (8) (2013) 1089–1091, https://doi.org/10.1093/bioinformatics/btt085.

[12] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, Z. Zhang, GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses, Nucleic Acids Res. 45 (W1) (2017) W98–W102, https://doi.org/10.1093/nar/gkx247.

[13] C.P. Goswami, H. Nakshatri, PROGgeneV2: enhancements on the existing database, BMC Canc. 14 (1) (2014) 970, https://doi.org/10.1186/1471-2407-14-970.

[14] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K. A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C.L. Robertson, N. Serova, S. Davis, A. Soboleva, NCBI GEO: archive for functional genomics data sets—update, Nucleic Acids Res. 41 (D1) (2013) D991–D995, https://doi.org/10.1093/nar/gks1193.

[15] A. Nagy, G. Munkacsy, B. Gyorffy, Pancancer survival analysis of cancer hallmark genes, Sci. Rep. 11 (1) (2021) 6047, https://doi.org/10.1038/s41598-021-84787-5.

[16] D.S. Chandrashekar, B. Bashel, S. Balasubramanya, C.J. Creighton, I. Ponce-Rodriguez, B. Chakravarthi, S. Varambally, UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses, Neoplasia 19 (8) (2017) 649–658, https://doi.org/10.1016/j.neo.2017.05.002.

[17] N. Borcherding, N.L. Bormann, A.P. Voigt, W. Zhang, TRGAted: a web tool for survival analysis using protein data in the Cancer Genome Atlas, F1000Res. 7 (1235) (2018), https://doi.org/10.12688/f1000research.15789.2.

[18] X. Guan, M. Cai, Y. Du, E. Yang, J. Ji, J. Wu, CVCDAP: an integrated platform for molecular and clinical analysis of cancer virtual cohorts, Nucleic Acids Res. 48 (W1) (2020), https://doi.org/10.1093/nar/gkaa423. W463–W471.

[19] R. Aguirre-Gamboa, H. Gomez-Rueda, E. Martínez-Ledesma, A. Martínez-Torteya, R. Chacolla-Huaringa, A. Rodriguez-Barrientos, J.G. Tamez-Peña, V. Treviño, SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis, PloS One 8 (9) (2013), e74250, https://doi.org/10.1371/journal.pone.0074250.

[20] R. Aguirre-Gamboa, V. Trevino, SurvMicro: assessment of miRNA-based prognostic signatures for cancer clinical outcomes by multivariate survival analysis, Bioinformatics 30 (11) (2014) 1630–1632, https://doi.org/10.1093/bioinformatics/btu087.

[21] R. Meijers, R.G. Smock, Y. Zhang, J.H. Wang, Netrin synergizes signaling and adhesion through DCC, Trends Biochem. Sci. 45 (1) (2020) 6–12, https://doi.org/10.1016/j.tibs.2019.10.005.

[22] V. Cirulli, M. Yebra, Netrins: beyond the brain, Nat. Rev. Mol. Cell Biol. 8 (4) (2007) 296–306, https://doi.org/10.1038/nrm2142.

[23] W. Hao, M. Yu, J. Lin, B. Liu, H. Xing, J. Yang, D. Sun, F. Chen, M. Jiang, C. Tang, X. Zhang, Y. Zhao, Y. Zhu, The pan-cancer landscape of netrin family reveals potential oncogenic biomarkers, Sci. Rep. 10 (1) (2020) 5224, https://doi.org/10.1038/s41598-020-62117-5.

[24] S.L. Wang, X. Li, The UCSCXenaTools r package: a toolkit for accessing genomics data from UCSC Xena platform, from cancer multi-omics to single-cell RNA-seq, JOSS 4 (40) (2019) 4750, https://doi.org/10.21105/joss.01627.

[25] H. Wickham, J. Hester, Readr: read rectangular text data, 2021. https://CRAN.R-project.org/package=readr.

[26] T.M. Therneau, P.M. Grambsch, Modeling Survival Data: Extending the Cox Model, Statistics for Biology and Health, Springer, New York, 2000, p. 1, https://doi.org/10.1007/978-1-4757-3294-8, online resource (xiii, 350 pages).

[27] A. Kassambara, M. Kosinski, P. Biecek Survminer, Drawing survival curves using 'ggplot2', 2021. https://CRAN.R-project.org/package=survminer.

[28] T.A. Gerds, B. Ozenne, riskRegression: risk regression models and prediction scores for survival analysis with competing risks, 2020. https://CRAN.R-project.org/package=riskRegression.

[29] F.E. Harrell Jr., Rms: regression modeling strategies, 2021. https://CRAN.R-project.org/package=rms.

[30] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer -Verlag, New York, 2016.

[31] N. Simon, J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for cox's proportional hazards model via coordinate descent, JSS 39 (5) (2011) 1–13, https://doi.org/10.18637/jss.v039.i05.

[32] M. Kuhn, Caret: classification and regression training, 2021. https://CRAN.R-project.org/package=caret.

[33] C. Wen, A. Zhang, S. Quan, X. Wang, BeSS: an r package for best subset selection in linear, logistic and cox proportional hazards models, JSS 94 (4) (2020) 1627, https://doi.org/10.21105/joss.01627.

[34] M. Xue, J. Shang, B. Chen, Z. Yang, Q. Song, X. Sun, J. Chen, J. Yang, Identification of prognostic signatures for predicting the overall survival of uveal melanoma patients, J. Canc. 10 (20) (2019) 4921–4931, https://doi.org/10.7150/jca.30618.

[35] H. Zheng, G. Zhang, L. Zhang, Q. Wang, H. Li, Y. Han, L. Xie, Z. Yan, Y. Li, Y. An, H. Dong, W. Zhu, X. Guo, Comprehensive review of web servers and bioinformatics tools for cancer prognosis analysis, Front. Oncol. 10 (68) (2020), https://doi.org/10.3389/fonc.2020.00068.

[36] A.N. Kamarudin, T. Cox, R. Kolamunnage-Dona, Time-dependent ROC curve analysis in medical research: current methods and applications, BMC Med. Res. Methodol. 17 (53) (2017), https://doi.org/10.1186/s12874-017-0332-6.

[37] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Statist. Soc. B (Statistical Methodology) 65 (5) (2005), https://doi.org/10.1111/j.1467-9868.2005.00503.x, 768-768.

[38] A.J. Gentles, S.V. Bratman, L.J. Lee, J.P. Harris, W. Feng, R.V. Nair, D.B. Shultz, V. S. Nair, C.D. Hoang, R.B. West, S.K. Plevritis, A.A. Alizadeh, M. Diehn, Integrating tumor and stromal gene expression signatures with clinical indices for survival stratification of early-stage non-small cell lung cancer, J. Natl. Cancer Inst. 107 (10) (2015), https://doi.org/10.1093/jnci/djv211.

[39] X. Yan, X. Fu, Z.X. Guo, X.P. Liu, T.Z. Liu, S. Li, Construction and validation of an eight-gene signature with great prognostic value in bladder cancer, J. Canc. 11 (7) (2020) 1768–1779, https://doi.org/10.7150/jca.38741.

[40] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Statist. Soc. B (Statistical Methodology) 58 (1) (1996) 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[41] M. Lachota, A. Lennikov, K.J. Malmberg, R. Zagozdzon, Bioinformatic analysis reveals central role for tumor-infiltrating immune cells in uveal melanoma progression, J. Immunol. Res. 9920234 (2021), https://doi.org/10.1155/2021/9920234.