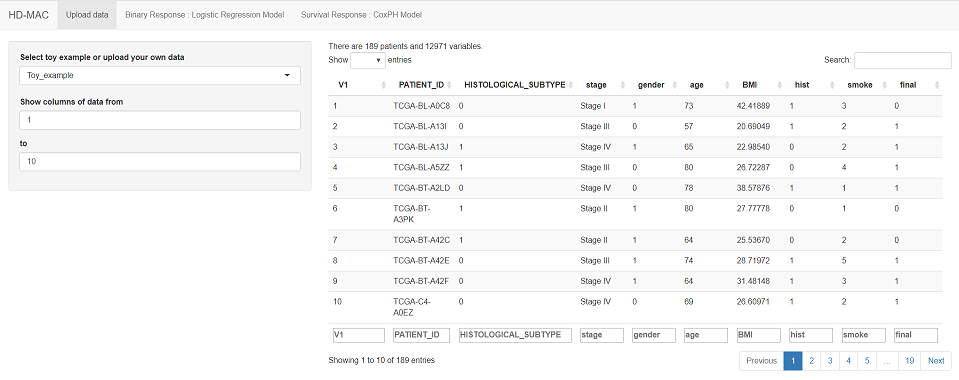
**HDMAC Tutorial**

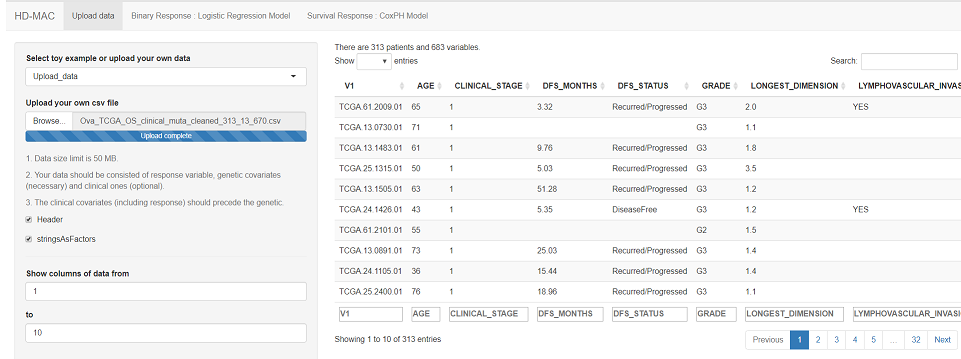
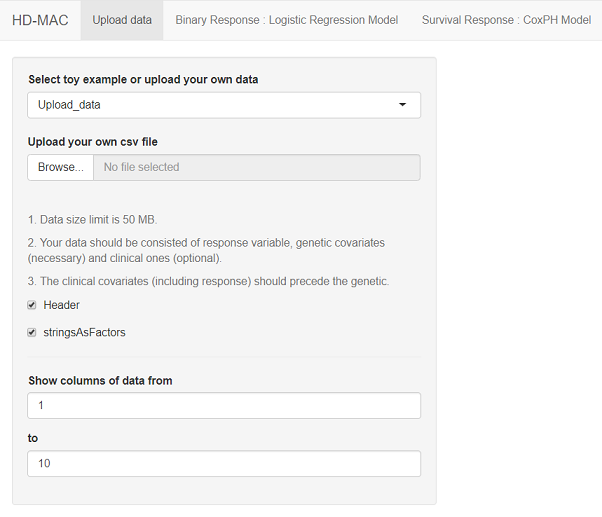
The high-dimensional analysis of cancer-associated genetic alterations, HDMAC, was developed to analyze high-dimensional genetic covariates, with the choice of including clinical covariates, with several regression models suitable for high-dimensional analysis and to identify important genetic alterations which could be used to construct a fitted multivariate regression model while its prediction power could be estimated by cross validation. HDMAC also allows choice of a penalty type for the corresponding penalized regression model for high-dimensional data and a first-step screening to screen out unrelated variables if the multiple-testing problem is of concern via control of the false discovery rate (FDR). Below is a tutorial on how to use HDMAC with respect of both survival data and binary outcome. The platform is available at <http://ripsung26.shinyapps.io/rshiny>.

**Data Upload/Toy Example**

To begin the analysis of your data, go to the website listed above and click the tab “Upload Data”. On the left of the page, you may choose whether to upload your own data or use the toy example we provided on the site. The data you upload or from our toy example are shown on the right. You may also choose which columns are shown on the front page.



The size limit to the data you upload is 50 MB. It may take a few moments for your data to upload dependent on the size.

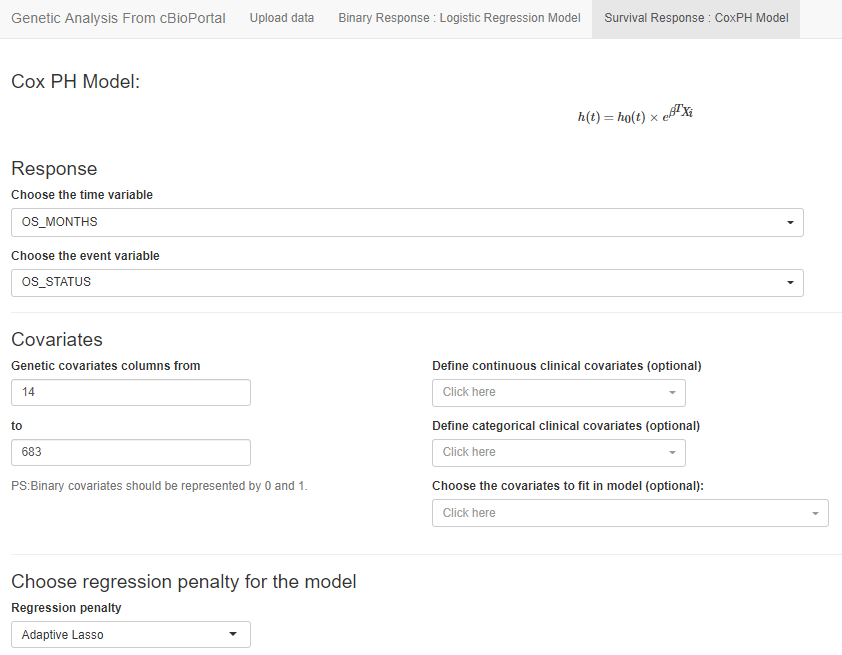


Once the data are uploaded, you may begin your analysis. For survival data, click the tab “Survival Regress: CoxPH Model”. For binary outcomes, click the tab “Binary Regression: Logistic Regression Model”.

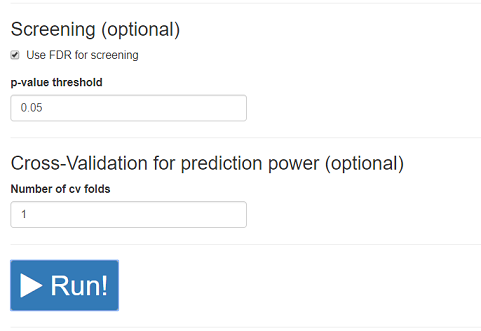
**Survival Analysis**

The data used here to illustrate how to run survival analysis on HDMAC contain the information of 8,310 mutated genes from 316 patients with serous type, high grade ovarian cancer from TCGA. The aim is to relate gene mutations to the patients’ overall survival.

1. Choose the tab “Survival Response: CoxPH Model” and locate all the variables needed for the analysis in the data uploaded. Inclusion of clinical variables is optional. Then choose the Cox regression method desired. Three are available: ridge, Lasso and adaptive Lasso.



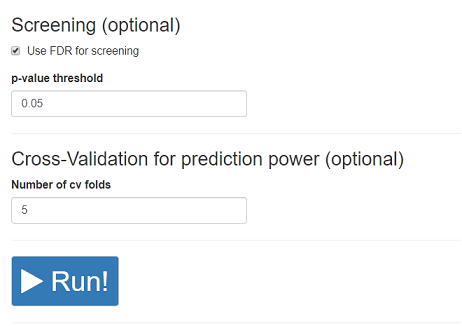
2. Print the gene list. [Optional:] Choose whether the initial screening to control the FDR is desired. If chosen (the box before “Use FDR for screening” is checked as shown below), the p-value threshold is set at 0.05. The number of cross-validation (CV) folds for testing the prediction power (C-index in the case of survival analysis) is set at 1 as the default for printing out gene lists. It is possible to change the CV fold for statistical tests, which is illustrated in the next step.



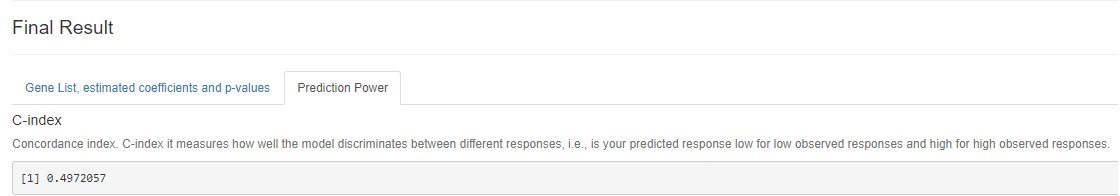
Hit “Run” and the gene list with each gene’s coefficient and p value will be printed on the page.



3. (Optional) To test the prediction power of the results, set the CV folds at 5 and hit RUN.



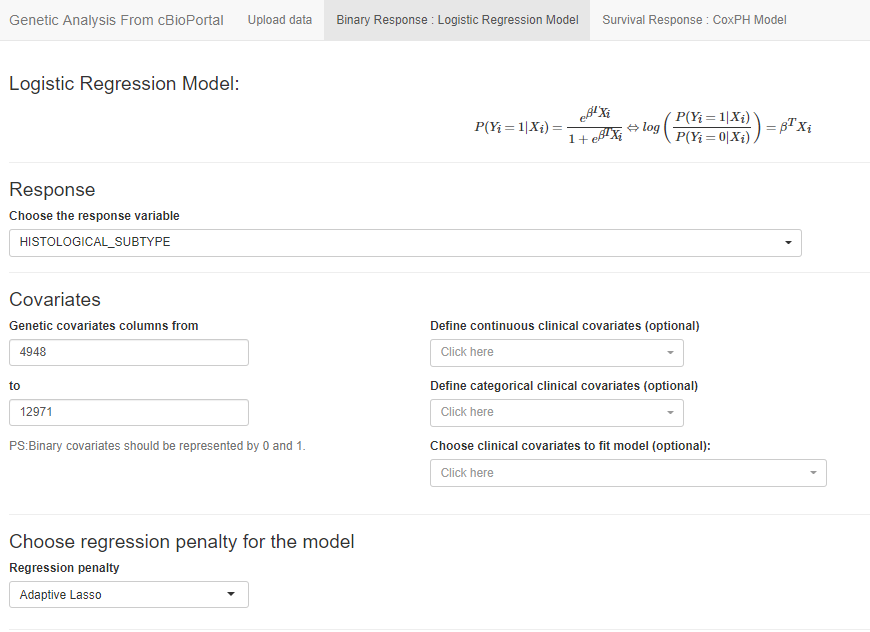
The concordance index, C-index, will be calculated to show the prediction power.



**Binary Outcome**

The data used here to illustrate how to run logistic regression in response to a binary outcome on HDMAC contain the information of 18,335 entries of mRNA expression of 189 patients with bladder cancer from TCGA. The aim is to relate abnormal mRNA expression to the patients’ cancer subtype, invasive vs. non-invasive.

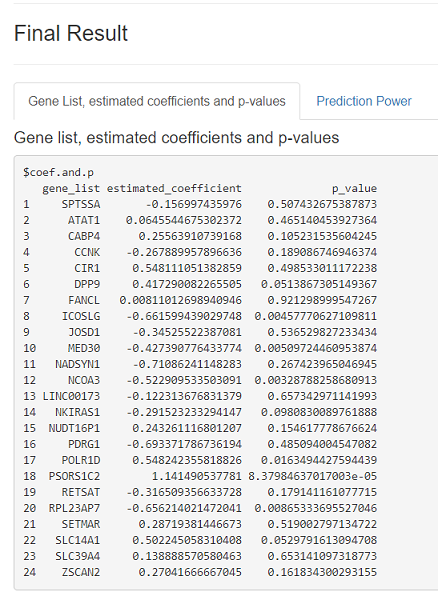
1. Choose the tab “Binary Response: Logistic Regression Model” and locate all the variables needed for the analysis in the data uploaded. Inclusion of clinical variables is optional. Then choose the logistic regression method desired. Three are available: ridge, Lasso and adaptive Lasso.



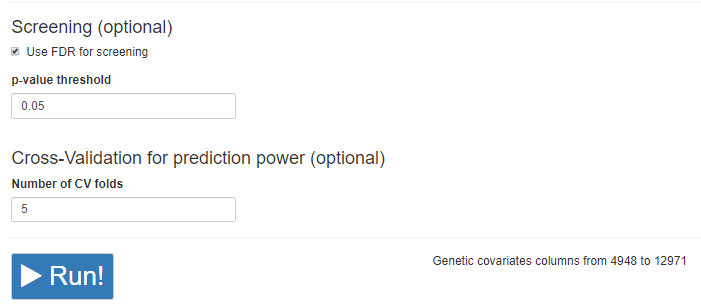
2. Print the gene list. [Optional:] Choose whether the initial screening to control the FDR is desired. If chosen (the box before “Use FDR for screening” is checked as shown below), the p-value threshold is set at 0.05. The number of cross-validation (CV) folds for testing the prediction power (sensitivity, specificity, accuracy and area under curve (AUC) in the case of logistic regression) is set at 1 as the default for printing out gene lists. It is possible to change the CV fold for statistical tests, which is illustrated in the next step.



Hit “Run” and the gene list with each gene’s coefficient and p value will be printed on the page.



3. (Optional) To test the prediction power of the results, set the CV folds at 5 and hit RUN.



The sensitivity, specificity, accuracy, and AUC will be calculated to show the prediction power.

