
ProteoSelector User- guide



Version 1.0 Apr 2019

—



Table of Contents

1. Introduction
2. Description
3. Software Requirements
4. How to Download & Install ProteoSelector
5. Graphical User interface steps
6. Example Datasets
7. Citation

Project PI: Dr. Sameh Magdeldin, Proteomics and Metabolomics Unit, Department of Basic Research, Children's Cancer Hospital 57357, Sameh.Magdeldin@57357.org

Lead developer: Mohammed Soudi, Mohamed.Soudy@57357.org, Ain-shams University.



Introduction

ProteoSelector version 1.0.0 software enables selecting multiple columns of interest based on a second unique identifier index, needed for analysis purposes of Proteomics and Metabolomics data. It helps in the protein Identification pipeline. The software was developed using R-3.6.1. ProteoSelector is designed to be run on tab-delimited text files and Comma separated files (CSV) as well. The output is an Excel sheet containing the index column together with columns of interest being selected and arranged column-wise.



Description

ProteoSelector version 1.0.0 software enables selecting multiple columns of interest based on a second unique identifier index needed for analysis purposes of Proteomics and Metabolics data. It helps in the protein Identification pipeline. The software was developed using R-3.6.1 ProteoSelector is designed to be run on tab-delimited text files and Comma separated files (CSV) as well. The output is an Excel sheet containing the index column together with columns of interest being selected and arranged column-wise.

The software workflow passes through;

1. Selection of parent directory that contain smaller subdirectories each having their own text-based sample files or comma separated files (CSV) ready to be analyzed. Note that the input files should contain a directory that has subdirectories containing at least two or more text-based files with the same column length and having the same column names.
2. Selection of columns specified from the user based on a second unique identifier, usually accession numbers (indexed 2)
3. The columns of interest are being selected and concatenated in one file so that the output file will have all the columns of each of the subdirectory folders arranged in column-wise manner and ready for further analysis .
4. The resulted output is being stored in an excel sheet for every subcategory



Software Requirements

I. OPERATING SYSTEM

This software is compatible with Microsoft Windows XP with SP3, Windows XP Professional with SP2 64-bit system, Windows 7 SP1 64-bit system and Microsoft windows server 2008 standard.

II. R programming language

In case R is not installed, R and RStudio can be downloaded from the following link;

<https://www.rstudio.com/products/rstudio/download>

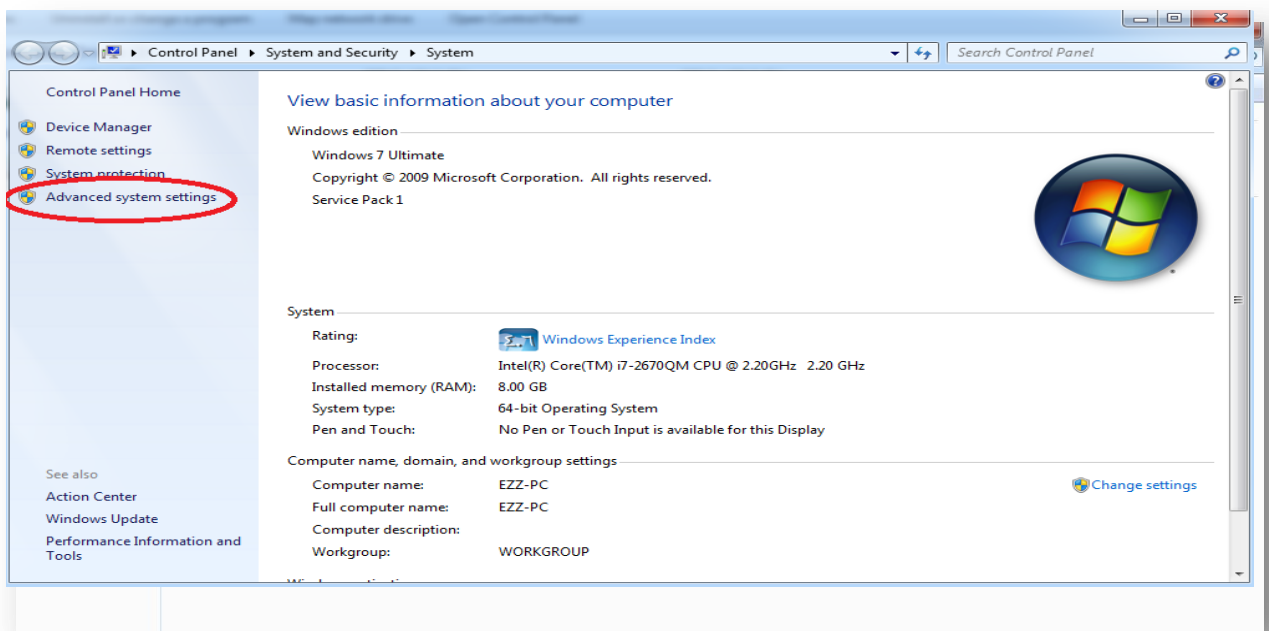
for downloading R either use your shell script via pip or pip3 or visit the website of the R-project

<https://www.r-project.org/>.

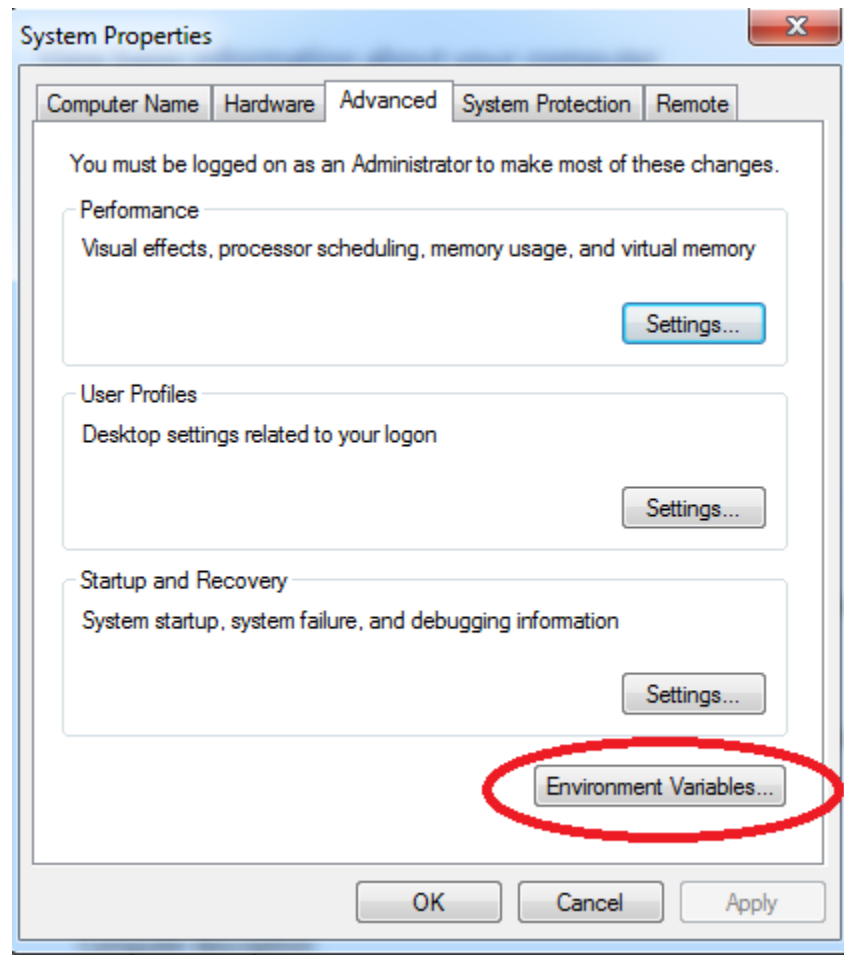
How to Download & Install ProteoSelector:

III. Pre-installation steps;

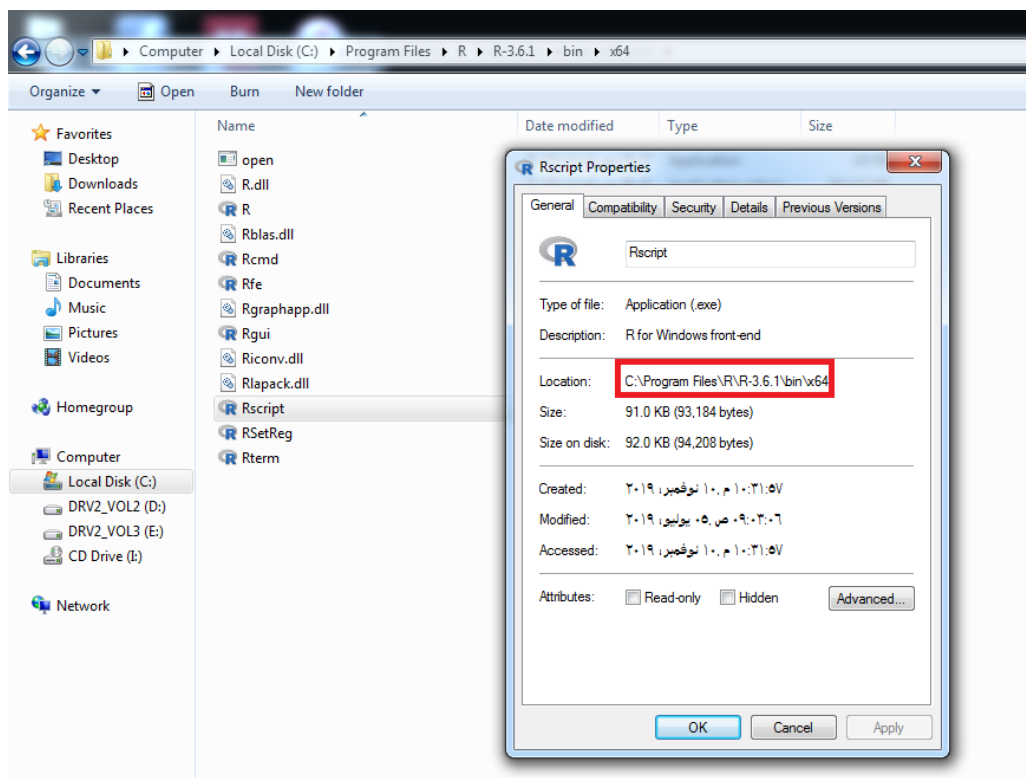
1. Download ProteoSelector software, it can be downloaded from ;
<https://beta.57357.org/wp-content/themes/57357/programs/ProteoSelector.rar>
2. In case R Software is not added to the path, it needs to be initially added. This can be achieved by adding Rscript to Environment variables through following the upcoming steps.
 - i. From my computer, right click and choose properties.
 - ii. Choose “Advanced System Settings”



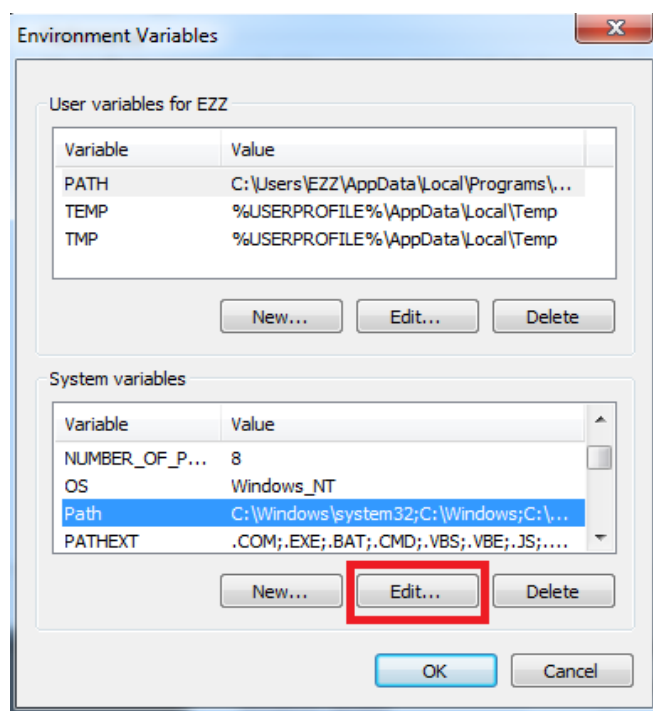
-
-
- iii. Choose “Environment Settings”



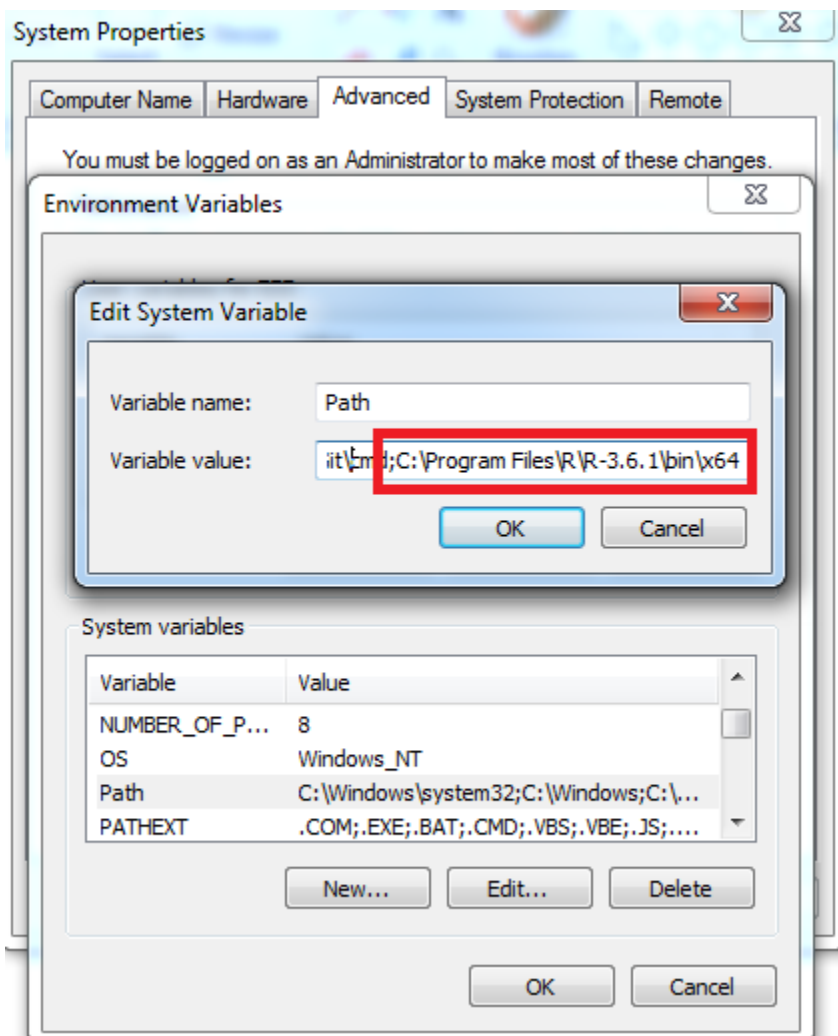
- a. Get the path of R script located within the local folder and copy it.



- b. In the second frame, select 'Path' then press "edit"



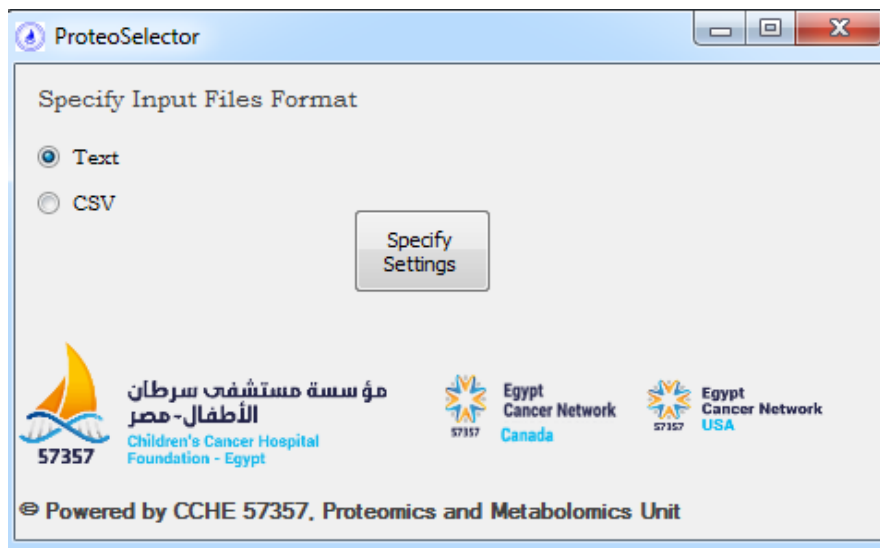
- c. Add semicolon “;” followed by the copied path. After which path will be added & installation is finished.



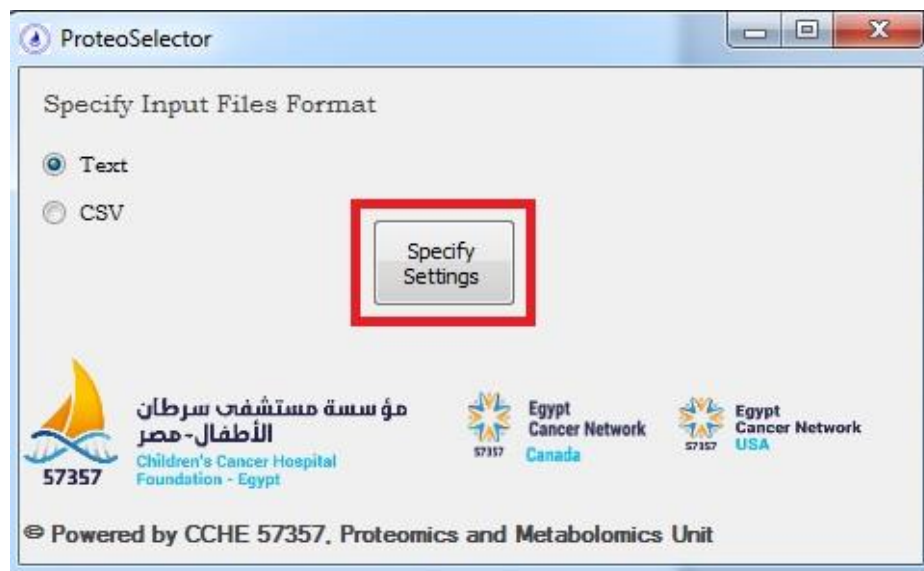
2. Extract the file named ProteoSelector.
3. Run the software from ProteoSelector icon, it shall run smoothly now.

Graphical User interface steps:

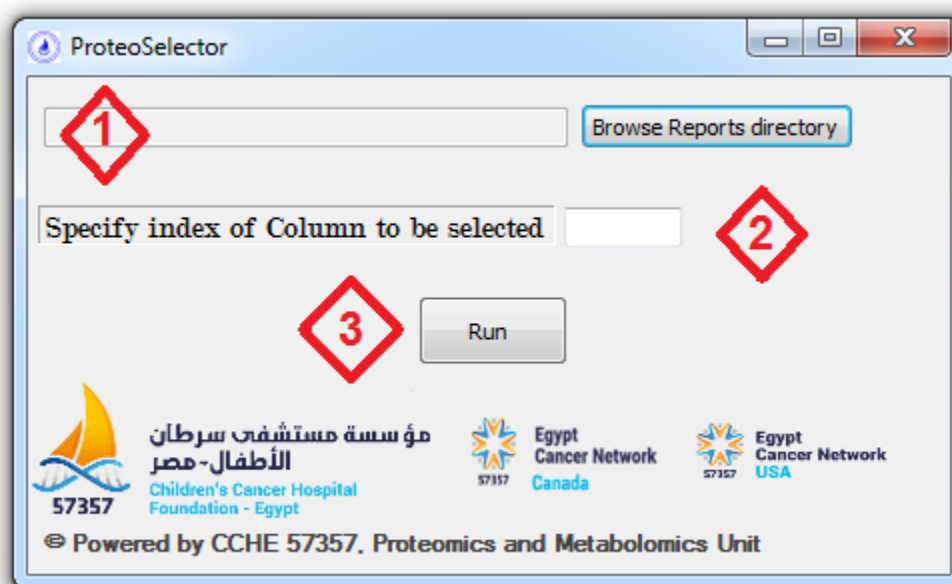
4. After pressing the Proteoselector software icon, The following screen will pop up:



- a. The user will be opted to choose;
1. “Text” if the files are Text-based files
 2. “CSV” if the files are Comma Separated files. Choose “browse Samples Directory”.
- b. After defining the file type, press “specify settings”



- c. Then the user will be opted to enter the following;
1. Define the Reports Directory path
 2. Specify the index of the columns that needs to be selected.
If single column needs to be selected, like 3 or 4 or whatever, it can be entered as it is.
if multiple columns needs to be selected, like both 3 and 4 , then it can be inserted and separated by a comma “,” like (3,4)
 3. Press Run to start the selection process.



- d. After the selection process finish, a message of “Done” is shown at the end.
- e. A file in excel format is being produced for every subdirectory folder containing the selected columns of interest concatenated columns-wise based on the unique index.

Example Datasets:

- I. An example of input files for the software in a folder named “Reports” , containing subfolders A and B each of which contain text-based files for protein identification data.
- II. It can be downloaded from here;
<https://drive.google.com/drive/folders/1eCsLwQEnBuwIKktssdjeSu3P3saBc-pt?usp=sharing>

Citations

Not Yet

Copyright (C) 2019-2020 Children’s Cancer Hospital 57357
{proteomics.lab@57357.org }