ProteoCompanion User-Guide

Version 1.0.0 Apr 2019

—

Proteomics and Metabolomics Unit CCHE 57357

Table of Contents

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

ProteoCompanion version 1.0.0 software enables performing many statistical operations on text-based files of Proteomics Data via combining multiple columns of tab delimited text files based on unique identifier in the second index of file. It helps in the protein Quantification pipeline. The software was developed using R-3.6.1. ProteoCompanion is designed to be run on tab-delimited text files. The output is an Excel sheet containing the specified columns after performing operations on them.

Description

ProteoCompanion version 1.0.0 software enables performing many statistical operations on text-based files of Proteomics Data. It helps in the protein Quantification pipeline. The software was developed using R-3.6.1. ProteoCompanion is designed to be run on tab-delimited text files. The output is an Excel sheet containing the specified columns after performing operations on them.

The software workflow passes through;

1. Selection of parent directory that contains smaller subdirectories each having their own text-based sample files 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 operations to be performed on the files so that ;
   a. Average; it computes average value for the specified columns among protein samples in each subcategory.
   b. Average with ceiling; it enables calculating the average for all values within the chosen columns while rounding it to the highest value.
   c. Concatenate: It is being used for merging columns based on the second index.
   d. Not Applicable: It enables inclusion of columns as it is, without any changing in the properties.
   e. Not included: It executes the columns that are not needed for the current analysis.

3. All the previously mentioned operations are presented for each column, identified by its header (i.e. column name). Like; Molecular weight, Go Accession, Gene name, possible coverage, spectrum counting … etc).
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 ProteoCompanion:

III.   Pre-installation steps;

1. Download ProteoCompanion software, it can be downloaded from ;
   https://beta.57357.org/wp-content/themes/57357/programs/ProteoCompanion.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.
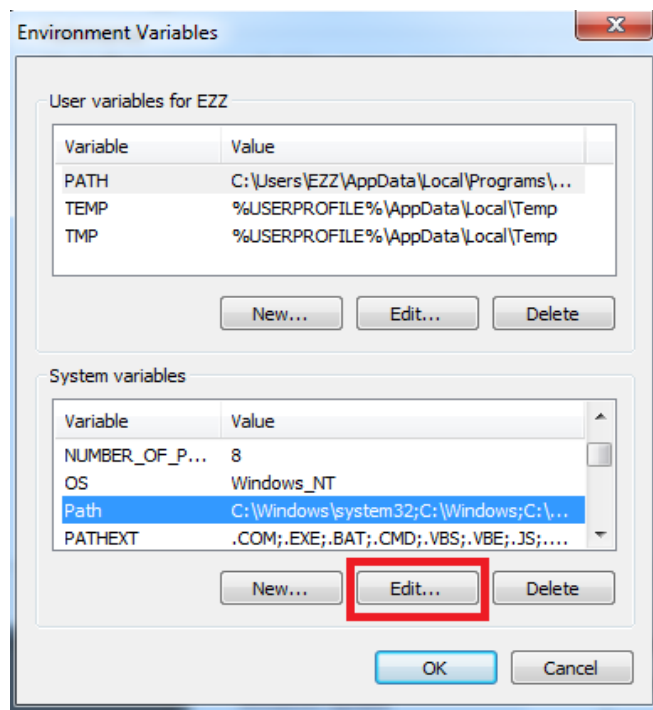   ii. Choose "Advanced System Settings"
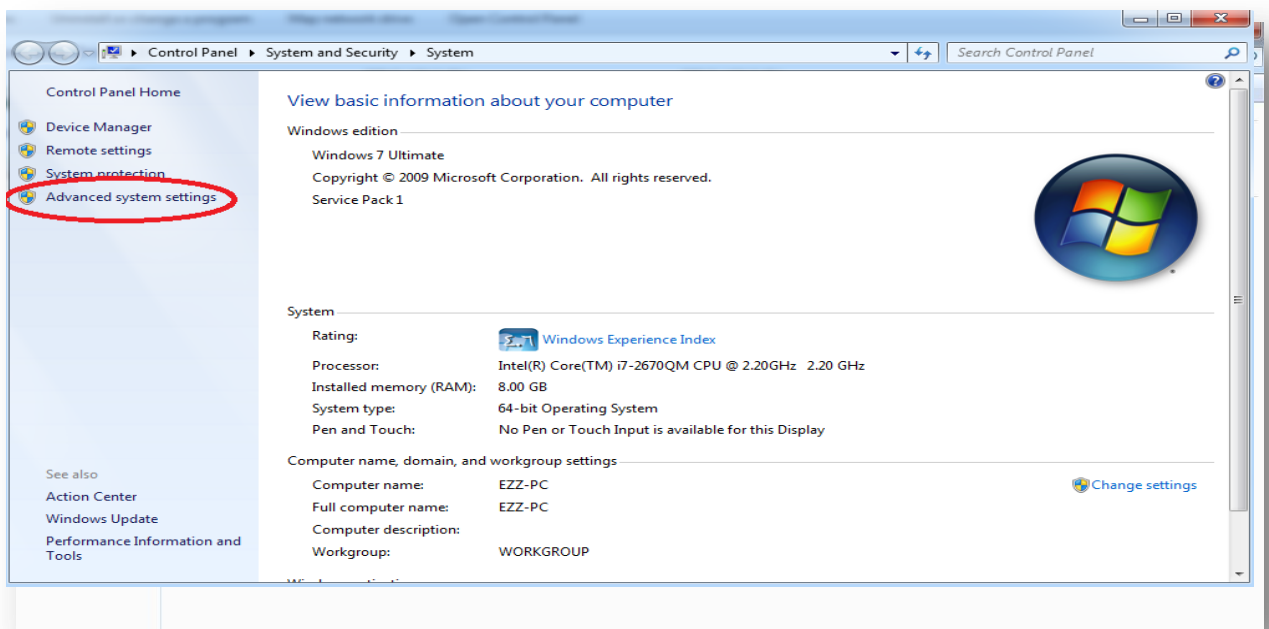
   iii. Choose "Environment Settings"
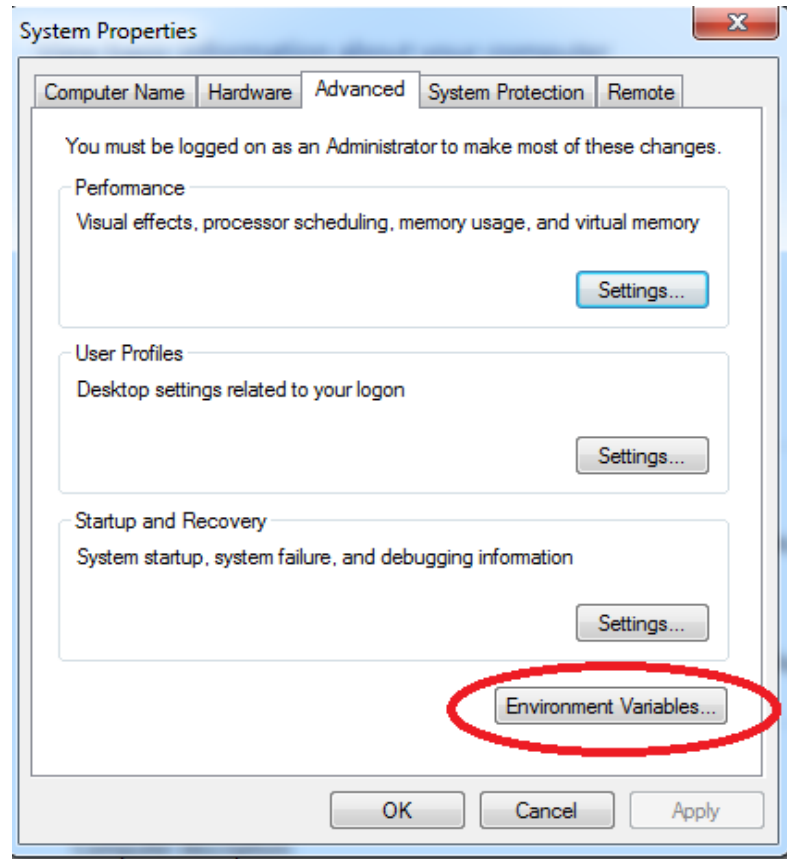
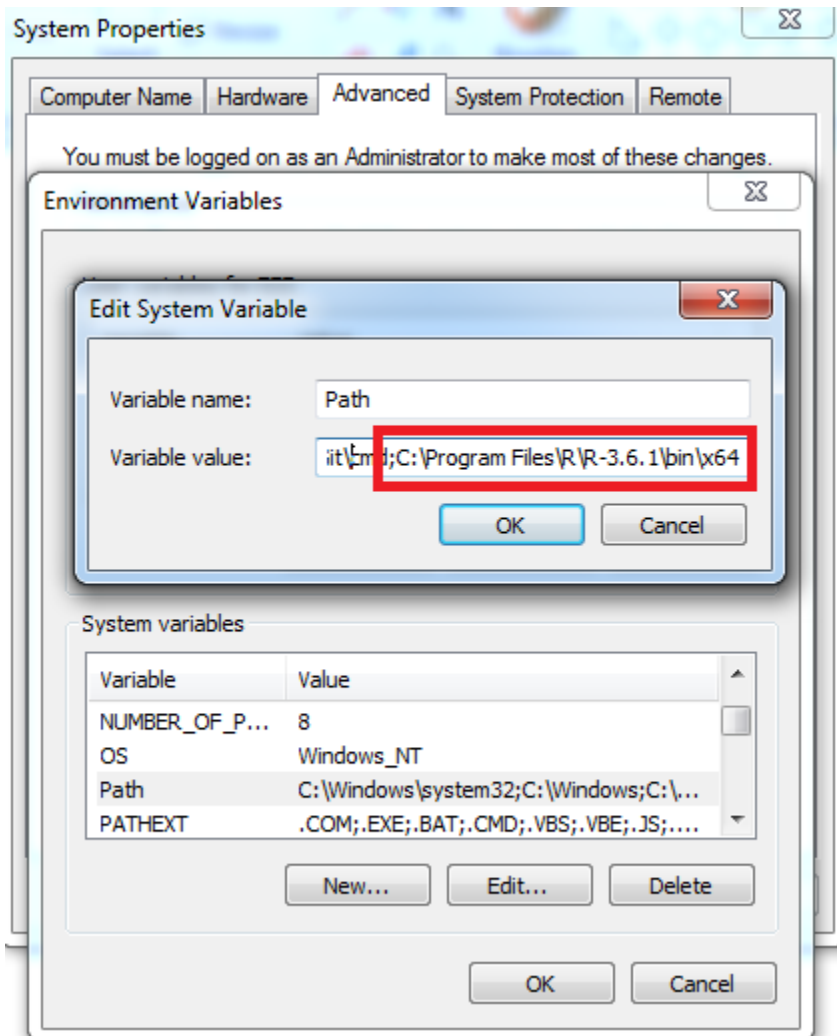a. Get the path of R script located within the local folder and copy it.
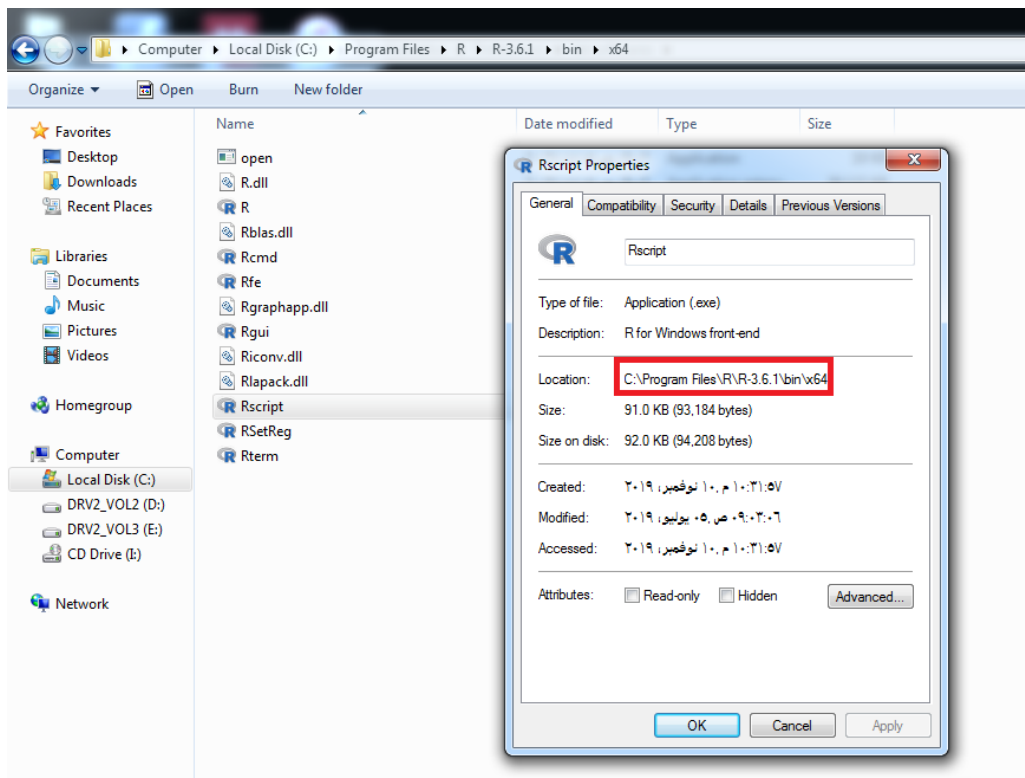
b. In the second frame, select 'Path' then press "edit"

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

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

Graphical User interface steps:

1. The following screen will pop up:
    a. Choose "browse Samples Directory". The user shall enter the parent directory path that has subdirectories containing reports files. they should be at least two or more text-based files with the same column length and having the same column names.
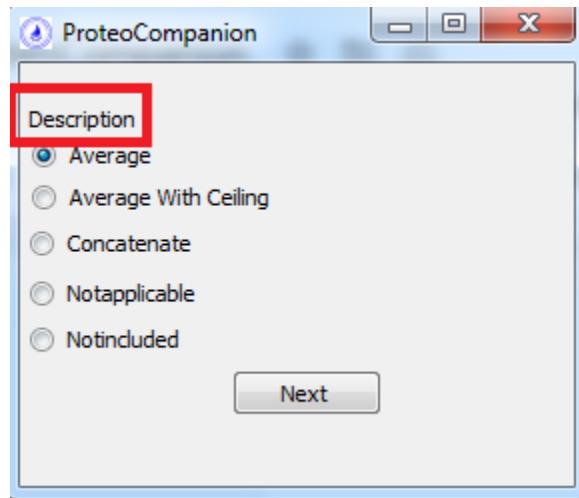

   b. After defining the directory path, "press specify settings"

c. The user will be opted to choose the operation that is needed to be performed upon each column. The column header is present at the top left .

d. List of operations are being listed beneath each column [average, average with ceiling, concatenate, not applicable, not included] as previously explained in details in the Description part.
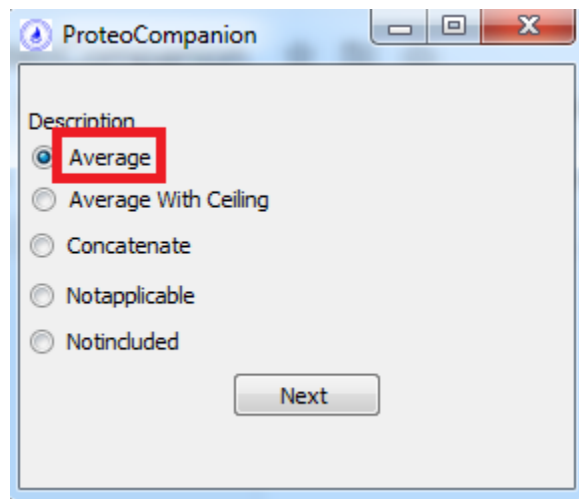
e. After specifying the operations for the needed headers, a message of "Done" is being shown at the end.

f. A file in excel format is being produced for every subdirectory folder.

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