**CSBB COMMAND LINE APPLICATION**

CSBB offers thirteen statistical and bioinformatics modules for several bioinformatics applications. Detailed instructions on how to execute each module is give below. Also refer to additional documents in the package for definition and explanation of each module.

**Steps to make the application executable**.

**Step 1**: Open the terminal/Command prompt in MAC OS & Windows respectively.

**Step 2**: Browse to the directory where you have saved the package.

=================

cd —> change directory

ls —> list directory

=================

**For example**: let’s say you have downloaded the package in Downloads.

====================================================

MACOS & LINUX users:

cd /Users/xxx/Downloads/CSBB/MAC0S\_LINUX\_VERSION/

Windows:

cd C:\Users\xxx\Downloads\CSBB\WINDOWS\_VERSION\

====================================================

**Step 3**: Type in console: perl -v

A detailed message describing perl version should be displayed:

This is perl 5, version 18, subversion 2 (v5.18.2) built for darwin-thread-multi-2level

(with 2 registered patches, see perl -V for more detail)

Copyright 1987-2013, Larry Wall

**This confirms that perl is installed on your system**.

====================================================

**If perl is not installed**:

**Mac Users**: Use this link to install perl <http://learn.perl.org/installing/osx.html> [With MAC OS perl and python comes installed]

**Windows User**: Use this link to install perl <http://learn.perl.org/installing/windows.html>

====================================================

**Step 4**: Install R, python (already comes installed with MAC OS) and Pandoc

**Install R for MAC users**: Use this link to install R <https://cran.r-project.org/bin/macosx/>

**Install R for Windows users**: Use this link to install R <https://cran.r-project.org/bin/windows/base/>

**Install Python for MAC users**: Use this link to install python <https://www.python.org/downloads/mac-osx/> [Already installed with MAC OSX]

**Install Python for Windows users**: Use this link to install python <https://www.python.org/downloads/windows/>

**Install Pandoc for Mac Users**: Use this link to install Pandoc https://github.com/jgm/pandoc/releases/download/1.17.0.2/pandoc-1.17.0.2-osx.pkg

**Install Pandoc for Windows Users**: Use this link to install Pandochttps://github.com/jgm/pandoc/releases/download/1.17.0.2/pandoc-1.17.0.2-windows.msi

**Running each module**

All modules can be run in two ways:

1) Giving all arguments from command line in one line.

2) Providing arguments when prompted by the application.

I) Let’s see how to use **—help** option in the application.

====================================================

./CSBB.exe —help

====================================================

Below is the sample output from —help option

Operating System is darwin

=============================================

COMMAND LINE SUITE FOR BIOINFORMATICIANS & BIOLOGISTS

Version\_1

Use --help for information on running COMMAND LINE SUITE FOR BIOINFORMATICIANS & BIOLOGISTS

Please See README for getting detailed instructions on running the script...

Requirements for MacOSX and Linux version: Please install R and python 2.7+

=============================================

=====Happy to see you ... Lets get started and do some magic with your files =====

Please use options below to run the COMMAND LINE SUITE FOR BIOINFORMATICIANS & BIOLOGISTS

Options::

UpperQuantile ---> for performing upper quantile normalization

BasicStats ---> for obtaining stats like mean, median, standard deviation, variance, Sum, min and max for each Gene Expression profile

ExpressionToZscore ---> for obtaining z-scores for Gene Expression in samples

ExtractGeneInfo ---> for obtaining info/expression of list genes from a huge matrix gene info/expression

InteractiveHeatmap ---> for generating interactive heatmaps for expression data. User has three options on clustering type and four choices on color theme. Please read README for descriptions and run command.

CorrelationProfiles ---> for obtaining genes correlation profile termed as positively and negatively correlated based on User threshold. One can obtain profile for all genes or just genes of interest (For genes of interest user needs to provide the path to gene list file). Heatmap will only be displayed for genes of interest

Biogrid-Gene-Protein-Search ---> for obtaining gene-protein interactions for Human and Mouse for genes of interest

DifferentialExpression ---> for obtaining DE genes in RNA-SEQ experiments. Uses RUVSeq package in R

PCA ---> for performing Principal Component Analysis on Genes and Samples. User needs to provide path to the file and specify Genes or Samples based on which PCA is needed to be performed

NMF ---> for performing Nonnegative Matrix Factorization on Samples in Expression dataset

FetchSRA ---> for downloading Raw SRA data from NCBI SRA for example user can download all the samples under SRA project id or just a single sample based on requirement. Just requires path to the folder where you want to download the data and SRA ID

FetchGEO ---> for downloading GEO expression data for example user can fetch expression of GSE Id, GSM Id, GDS Id. Just requires path to the folder where you want to download the data and GSE/GSM/GSD ID

InteractiveScatterPlot ---> for generating Interactive Scatter plot based on user preference. User needs to provide path to the file, Column number for x-axis values, Column number for y-axis values and Column number which user needs for color factorization (If users provides No color factorization will not be done) Please see README for extensive explanation

**\*\*\*\*\*\* Please note that users can simply drag and drop the file/folder from Folder Window (Windows) and Finder Window (MAC) respectively when path to a folder or a file is required**.

II) **Running UpperQuantile**:

====================================================

./CSBB.exe UpperQuantile Path\_to\_file

====================================================

Example Mac Users: ./CSBB.exe UpperQuantile /Users/xx/Desktop/FIND.txt

Example Windows Users: ./CSBB.exe UpperQuantile C:\Users\xx\Desktop\FIND.txt

Now if you did not provide the path to file as an argument

Example: ./CSBB.exe UpperQuantile

— > Application will prompt to provide user input

Operating System is darwin

=============================================

COMMAND LINE SUITE FOR BIOINFORMATICIANS & BIOLOGISTS

Version\_1

Use --help for information on running COMMAND LINE SUITE FOR BIOINFORMATICIANS & BIOLOGISTS

Please See README for getting detailed instructions on running the script ...

Requirements for MacOSX and Linux version version : Please install R and python 2.7+

=============================================

=====Happy to see you ... Lets get started and do some magic with your files =====

Upper Quantile Normalization module loaded

You have forgot to give path to file from command line

Please provide the path to the file

path\_to\_file [User input is required]

III) **Running BasicStats**:

====================================================

./CSBB.exe BasicStats Path\_to\_file

====================================================

Example Mac Users: ./CSBB.exe BasicStats /Users/xx/Desktop/FIND.txt

Example Windows Users: ./CSBB.exe BasicStats C:\Users\xx\Desktop\FIND.txt

IV**) Running ExpressionToZscore**:

====================================================

./CSBB.exe ExpressionToZscore Path\_to\_file

====================================================

Example Mac Users: ./CSBB.exe ExpressionToZscore /Users/xx/Desktop/FIND.txt

Example Windows Users: ./CSBB.exe ExpressionToZscore C:\Users\xx\Desktop\FIND.txt

V) **Running ExtractGeneInfo**:

====================================================

./CSBB.exe ExtractGeneInfo Path\_to\_Expression/Info\_file Gene\_List\_File

====================================================

Example Mac Users: ./CSBB.exe ExtractGeneInfo /Users/xx/Desktop/FIND.txt /Users/xx/Desktop/FIND1.txt

Example Windows Users: ./CSBB.exe ExpressionToZscore C:\Users\xx\Desktop\FIND.txt C:\Users\xx\Desktop\FIND1.txt

VI) **Running InteractiveHeatmap**:

====================================================

./CSBB.exe InteractiveHeatmap Path\_to\_File Clustering\_Option [Row\_Clust, Col\_Clust or Row\_Col\_Clust] Color\_theme [YellowGreenOrange, BlueWhiteRed, YellowBlackBlue or GreenWhitePurple]

====================================================

VII) **Running CorrelationProfiles**:

====================================================

./CSBB.exe CorrelationProfiles Path\_to\_File Correlation\_Threshold [-1 to 1] all/Path to Gene list

====================================================

**\*\*use all to calculate correlation profiles all the genes in the matrix.**

**\*\*using path to gene list file to calculate correlation profile for a specified set of genes.**

VIII) **Running Biogrid-Gene-Protein-Search**:

====================================================

./CSBB.exe Biogrid-Gene-Protein-Search Human/Mouse Path\_to\_gene\_list

====================================================

IX) **Running DifferentialExpression**

====================================================

./CSBB.exe DifferentialExpression Path\_to\_Counts\_File Number\_of\_Controls Number\_of\_Treatments Counts\_Threshold\_for\_filtering Number\_of\_Samples\_for\_Filtering\_per\_Gene

====================================================

Example: ./CSBB.exe DifferentialExpression Path\_to\_Counts\_File 10 10 5 8

X) **Running PCA**

====================================================

./CSBB.exe PCA Path\_to\_File Genes/Samples

====================================================

XI) **Running NMF**

====================================================

./CSBB.exe NMF Path\_to\_File

====================================================

XII) **Running FetchSRA**

====================================================

./CSBB.exe FetchSRA Path\_to\_Folder\_for\_downloading\_files SRA-id

====================================================

XIII) **Running FetchGEO**

====================================================

./CSBB.exe FetchGEO Path\_to\_Folder\_for\_downloading\_files GEO-id

====================================================

XIV) **Running InteractiveScatterPlot**

====================================================

./CSBB.exe InteractiveScatterPlot Path\_to\_File Column\_x\_axis\_values Column\_y\_axis\_values Column\_for\_Color\_Factorization/No

====================================================