**Program User** **Guide**

User has to make sure he has Perl 5.24.3 or later and Bio Perl are installed and Vis, LWP and LWP.pm are in the running directory as the program requires them to run. Program uses few more modules of Bio Perl too. It can be run from Command line, Perl command line interpreter or on Linux with Perl version 5 installed including all related modules of Bio Perl and the HTML files can be opened on any browser.

Program is designed to give the user information about BRCA1 protein’s sequence, its isoelectric point, its amino acid composition and frequency, it’s string 20 interacting proteins with quality score over 900 and its PDB structure identifiers. It uses the “LWP::Simple” library to get the data from uniprot, string database, and also uses other Bio Perl modules such as “Bio::Seq” and “Bio::Tools::pICalculator” for changing arrays into subjects and calculating the isoelectric point.

It creates HTML links to create a column chart for amino acid composition & frequency but the program had to fetch the required information from previous steps and add JAVA script to it to create HTML link. Program output has the java script included for column chart in step 3. You can see the amino acid composition and frequencies are plotted nicely in a column chart.

Program fetches the information about other 20 interacting proteins with quality score over 900. Quality score confirms that these interactions are proven to be happening within the nucleus. As we can see they are arranged in a way the user can read through it easily and quickly. These proteins work together to repair DNA damages and many other important tasks within the cell.

HTML script provided in step 4 is for network graph BRCA1 and its interacting partners which is provided as a HTML file separately. Our primary protein BRCA1 is highlighted red in the network graph to make sure it’s spotted easily by the user. It shows the network of connections between all interacting partners and how they are connected to BRCA1 too.

Step 6 has the information about PDB structure identifiers for each protein in the network including their gene names.

Bar chart has the structure counts for the proteins only with more than 10 structures available. Program gets the information related to the structure identifiers and only displays the information of the proteins with more than 20 structures available and prints it out as java script. The html file of this bar chart is attached to the combined HTML file. User can see 7 proteins with related information on bar chart. Because only 6 proteins have more than 10 structure counts available within the network. As you’ve notices, BRCA1 also has more than 10 structure counts available.

User can look through the data when they run the program and can test the HTML script by saving it into a file and open it from any browser. But the user must have related libraries installed in the very same directory in order for them to open and display information they have within them.