HD-eXplosion

hd-eXplosion.utdallas.edu

Developed by Naifu Zhang

D'Arcy Laboratory, The University of Texas at Dallas

Currently running v1.2

<u>Welcome to HD-eXplosion!</u>

HD-eXplosion is a straightforward tool for visualizing HDX data. It can be used to create chiclet and volcano plots with or without a statistical filter applied. It can also extract a list of peptides that are shared between multiple HDX experiments post-processing.

HD-eXplosion has a simple GUI that we hope you find intuitive. This document provides some instructions. If you have questions, problems, or feedback, please email us at darcylabweb@gmail.com. Good, bad, or ugly… we'd love to hear from you!
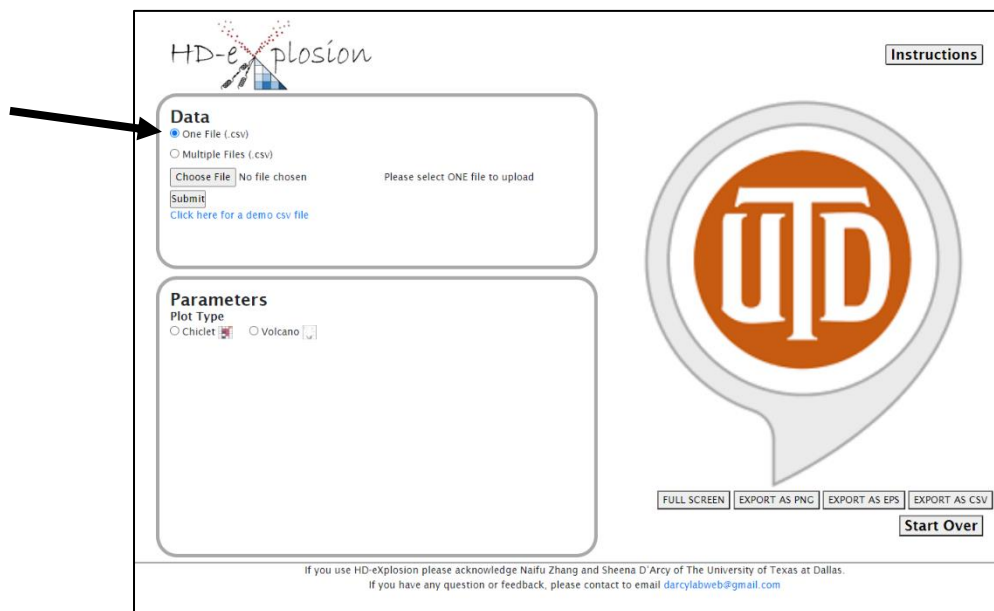
Be assured that we are NOT data spies. The only information we retain is your IP address.

<u>Outline</u>

1. **Data Upload – One file**
2. **Data Upload – Multiple files**
3. **Parameters – Chiclet Plot**
4. **Parameters – Volcano Plot**
5. **Output Files**
6. **Referencing**

1. Data Upload – One File

The GUI is divided into three panels. You submit your precious HDX data in the **Data** panel with 'One File' selected.

Input File Format

The input file must be *.csv. The data must be organized into the following columns with these EXACT headings in this EXACT order. Headings are case-sensitive.

- A. **Protein**
- B. **Start**
- C. **End**
- D. *Sequence*
- E. *Modification*
- F. *Fragment*
- G. *MaxUptake*
- H. *MHP*
- I. **State**
- J. **Exposure**
- K. *Center*
- L. *Center SD*
- M. **Uptake**
- N. **Uptake SD**
- O. RT
- P. RT SD

Protein is the protein name; Start and End are the residue numbers of the peptide boundaries; State is a description of the variable such as ± ligand; Exposure is the time in deuterium in any unit; Uptake is the average deuterium incorporated from replicates; and Uptake SD is the standard deviation of the average deuterium uptake. We wrote the software with uptake being in Daltons, but fractional uptake will also work. Just be sure to double check auto-generated headings!

The input file format mirrors the output file of *DynamX 3.0.0.* Our tool will only use data from columns A. B, C, I, J, M and N (bolded above). So, if you are creating your input file manually you can leave all other columns empty, except for their headings.

Row 1 should contain the column headings and not any actual data. The order of the peptides in the file will be the order of the peptides in the chiclet plot. If a peptide is listed more than once (e.g. if charge states are not combined like they are in *DynamX*), only the last entry will be shown in the plot. <u>If there is a blank cell in a used column, it will be read as a zero.</u>

You can download a sample demo csv file by clicking on the blue 'Click here for a demo csv file' link below the 'Submit' button.
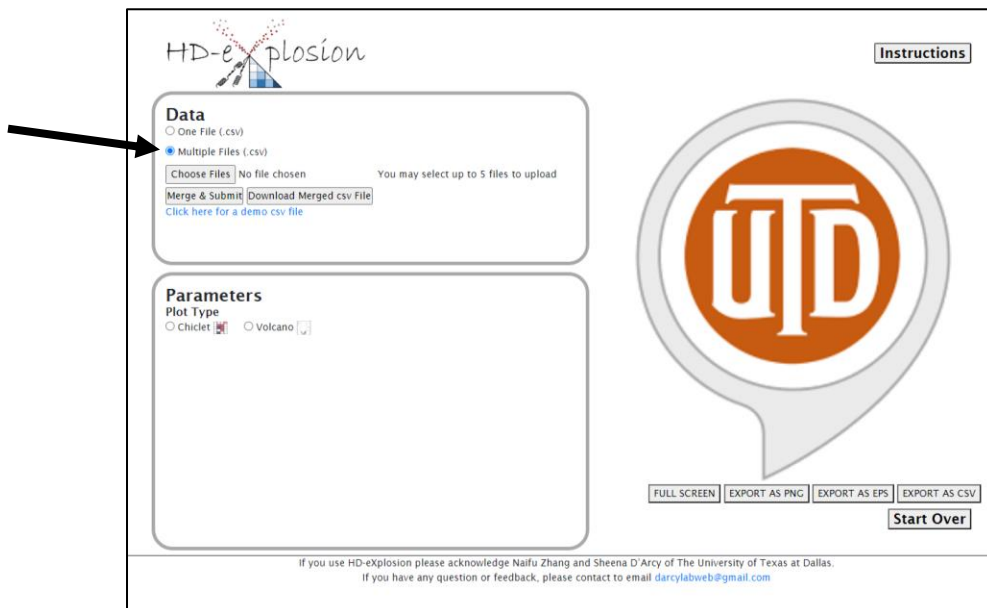
Uploading your Input File
1. Click on 'Choose File'.
2. Navigate to and select your file. It will be listed to the right of the 'Choose File' button.
3. Click 'Submit'. A yellow box saying '*.csv successfully uploaded' should appear. You are now ready to set the parameters for your chiclet plot (see 3) or volcano plot (see 4).

If an error appears, chances are something is up with the format of your input file. Compare it to our sample demo csv file to diagnose the issue… or better… use the demo file but paste your data under the relevant headings.

2. Data Upload – Multiple Files

HD-eXplosion can extract a list of peptides <u>shared</u> between up to five different input files. To do this, select 'Multiple Files' in the **Data** panel. This selection will cause 'Merge & Submit' and 'Download Merged csv File' buttons to appear.

Input File Format
The input files should be formatted as described in 1. They must also be in the same folder.

IMPORTANT… NONE of the states should have the same name. This will cause the plots to be blank. If you have a state repeated make sure you rename it in the input file (easily done with the replace function in excel).

Here is an example.
If you have studied a protein with many ligands and each time have repeated the unbound state you may have the following states; Unbound_1, LigandA_1, Unbound_2, LigandB_2, Unbound_3, LigandC_3 etc. We like using the underscore to keep track of which unbound and bound samples were collected at the same time.

Further, you might have biological replicates. This may lead to the following states; Unbound_1, LigandA_1, Unbound_2, LigandA_2, Unbound_3, LigandB_3, Unbound_4, LigandB_4 etc.

Uploading your Input Files
1. Click on 'Choose Files'.
2. Navigate to and select ALL your files before clicking 'Open'. Multiple files can be selected by holding Ctrl (Windows) or Command (Mac). The number of files will be listed to the right of the 'Choose Files' button.
3. Click 'Merge & Submit'. A yellow box saying 'Your X files successfully uploaded' should appear. You are now ready to set the parameters for your chiclet plot (see 3) or volcano plot (see 4).
4. You can download the merged file by clicking on the 'Download Merged csv File' button.

The csv file of the merged input files will list the shared peptides in column 1 and the theoretical maximum uptake in column 2. The data will then be listed from left to right starting with the time course of the first state, followed by the time course of the second state and so on. The columns alternate between the uptake and the SD of the uptake. All data from the first file will be listed, followed by all data from the second file and so on. Data from different input files are separated by a repeat of the theoretical maximum uptake column.

Importantly, shared peptides will only be extracted for the <u>first protein in the first file</u> selected when you open files (step 2 above). To do it for different proteins, alter their order in your original input file. The shared peptides for the first protein listed can be read to create a chiclet plot (see 3) or volcano plot (see 4).

3. Parameters – Chiclet Plot

The second panel of the GUI is **Parameters**. This is where you specify the details of your plot. One plot is a chiclet plot. A chiclet plot displays differences in deuterium uptake between two states over time. One axis is the list of peptides, while the other axis is time of deuterium exposure. The difference in deuterium uptake is denoted by the color of each rectangle in the grid.

To make a chiclet, upload your data (see 1-2) and select 'Chiclet'. Plot variables will appear. Select options for variables (described below) and then click 'Show'. Chiclet will replace the UTD monogram in the third panel (see 5). All variables MUST be specified.

Plot Variables
- **Protein** – the drop-down box will list all proteins in your input file (or the first protein if you are working from a merged file). Select the protein of interest.
- **Comparison** – the drop-down boxes will list all states in your input file(s). Select the states you wish to compare. The chiclet will show *State1 – State2*.
- **Range** – this is where you adjust the color range of uptake difference shown in the chiclet.
    Max is the maximum value and will be the darkest shade of color.
    Step is the gradient of the color change above zero.
    The minimum is assumed to be zero unless *Below 0* is selected (see below).
    For example, if max = 5 Da and step = 10, the color bar will have 10 shades of color between 0 and 5 Da

    To display a negative difference in uptake, check the box next to *Below 0.*

Min is the minimum value and will be the darkest shade of the second color.
Step is the gradient of the color change below zero.
If you have both positive and negative differences, we advise both steps to be equal.

- **Layout** – select the orientation of your chiclet as either *vertical* or *horizontal.* Peptides will be listed on the y-axis or x-axis, respectively. Also, select your preferred *color* from the drop-down box. The list will be a single color or a pair of colors depending on if *Below 0* is selected. Color are red (r), green (g), blue (b), orange (o), and yellow (y).
- **Statistics** - to turn on a *Significance Filter* check the box. The filter is based on a Welch's *t*-test, which tests whether two populations with different variance have the same means. The *t* variable is calculated as follows:

$$ t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{s_1^2}{N_1} + \dfrac{s_2^2}{N_2}}} $$

Where,
$\bar{X}_1$ and $\bar{X}_2$ are the mean of each population.
s_1 and s_2 are the SD of each population.
N_1 and N_2 are the number of replicate samples.

You define the level of *Significance* by setting the maximum p-value. Typically, the p-value should be less than 0.01. Any differences in uptake that have a p-value greater than the maximum you designate, will be shown as zero difference (white) on the chiclet plot.

You also need to specify the sample size (n), which is the number of replicates of each sample. 'n' on the webpage is equal to N_1 and N_2 in the above equation. It should be between 3 and 999. [Note: this was added later and hence does not appear in the screenshots]
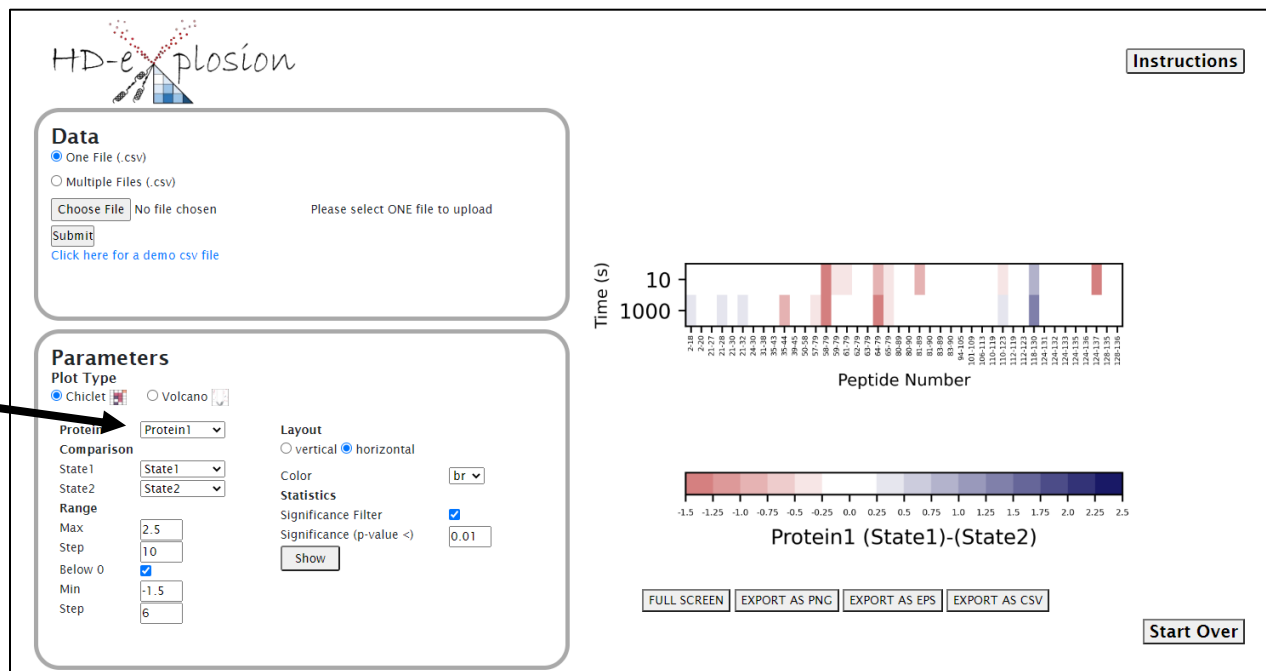
The significance filter is only valid if your input file contains the <u>average uptake calculated from replicate samples, as well as the SD of the uptake.</u>

Example – Demo File
The demo csv file contains peptides from two proteins (Protein1 and Protein2) in two states (State1 and State2). It is an artificial data set that has both positive and negative

differences in uptake. The best parameters and the resulting chiclets plots are shown below.

In this example, you can see the peptides with the largest positive difference in uptake are a darker blue color, while those with the largest negative difference in uptake are a darker red color. For a new dataset, it may take a few attempts to get the perfect parameters, especially the range.
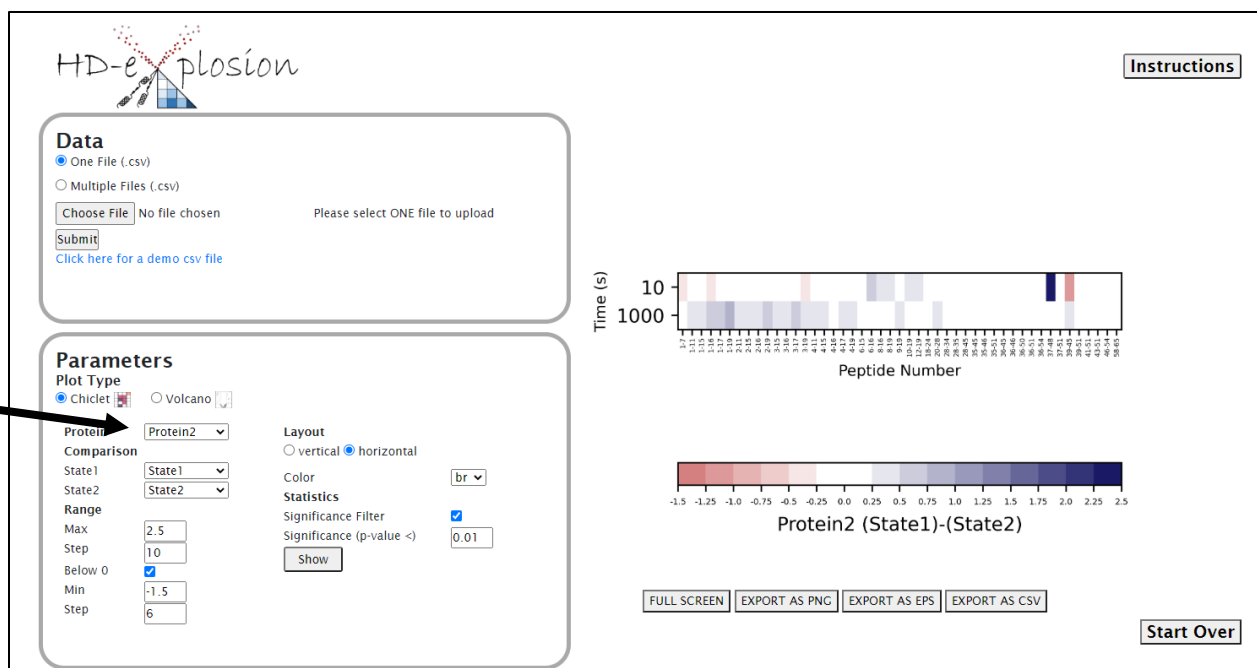
4. Parameters – Volcano Plot

The second panel of the GUI is **Parameters**. This is where you specify the details of your plot. One plot is a volcano plot. A volcano plot depicts each peptide at a specific time as a single point. The x-axis is the uptake difference (usually in Daltons) and the y-axis is the p-value (calculated as described in 3). A dashed box is placed on the plot to indicate significance cutoffs for both axes. Volcano plots are only valid if your input file contains the <u>average uptake calculated from replicate samples, as well as the SD of the uptake.</u>

To make a volcano plot, upload your data (see 1-2) and select 'Volcano'. Plot variables will appear. Select options for variables (described below) and then click 'Show'. Volcano will replace the UTD monogram in the third panel (see 5). All variables MUST be specified.

Plot Variables

- **Protein** – the drop-down box will list all proteins in your input file (or the first protein if you are working from a merged file). Select the protein of interest.
- **Time points** – the drop-down box will list all time points in your input file. Select the time point of interest, or *All* to display all time points on one graph.
- **Comparison** – the drop-down boxes will list all states in your input file(s). Select the states you wish to compare. The volcano will show *State1 – State2*.
- **Axis** – this is where you define the scale of your x- and y-axes.

  For the *X-Scale*, specify the minimum and maximum values, as well as the *Major interval* for tick spacing. The x-axis will be a linear axis.

  For the *Y-Scale*, specify the minimum p-value as the exponent of 10. A typical value will be -5 to -7. The maximum will be 0.5 by default. The y-axis will be a logarithmic axis.
- The top of the next column of variables relates to graph presentation.

  Color is the color of the points. If you select only one time point, the first color from the series will be used. Otherwise, the series will be used from left to right according to time point order in the input file. <u>Only points within your significance cutoffs will be colored. All other points will be grey</u>.

  Symbol size allows you to configure the size of the dots. 1 is the smallest and 9 is the largest.

  Text size allows you to designate the font size of peptide labels if you choose to include them on the plot (see below). We recommend 6 pt.

  Size of plot box allows you to set the actual length of the *X* axis and the *Y* axis in the final figure. The units are inches.
- **Significance Cutoffs** – significance cutoffs will be indicated on the volcano plot by dashed black lines. You need to specify the exponent of the maximum *p-value* and the *Min-difference.* The *Min-difference* is the absolute value and will be shown as

plus or minus. See Hageman and Weis (Anal. Chem. 2019, 91, 8008-8016) for how to use the pooled standard deviation to determine the *Min-difference*.

To label the peptides that pass significance cutoffs on the plot, select box next to 'Show list on plot'.

After showing the volcano plot an additional button will appear ('Download List') that enables you to download the list of significant peptides.

You also need to specify the sample size (n), which is the number of replicates of each sample. 'n' on the webpage is equal to N_1 and N_2 in the previously described equation. It should be between 3 and 999. [Note: this was added later and hence does not appear in the screenshots]

Example – Demo File
The demo csv file contains peptides from two proteins (Protein1 and Protein2) in two states (State1 and State2). It is an artificial data set that has both positive and negative differences in uptake. The best parameters and the resulting volcano plots are shown below.

In this example, you can see there are peptides from different time points that pass the significance cutoffs. Some have a positive different in uptake, while others have a negative difference in uptake. For a new dataset, it may take a few attempts to get the perfect parameters, especially the significance cutoffs.
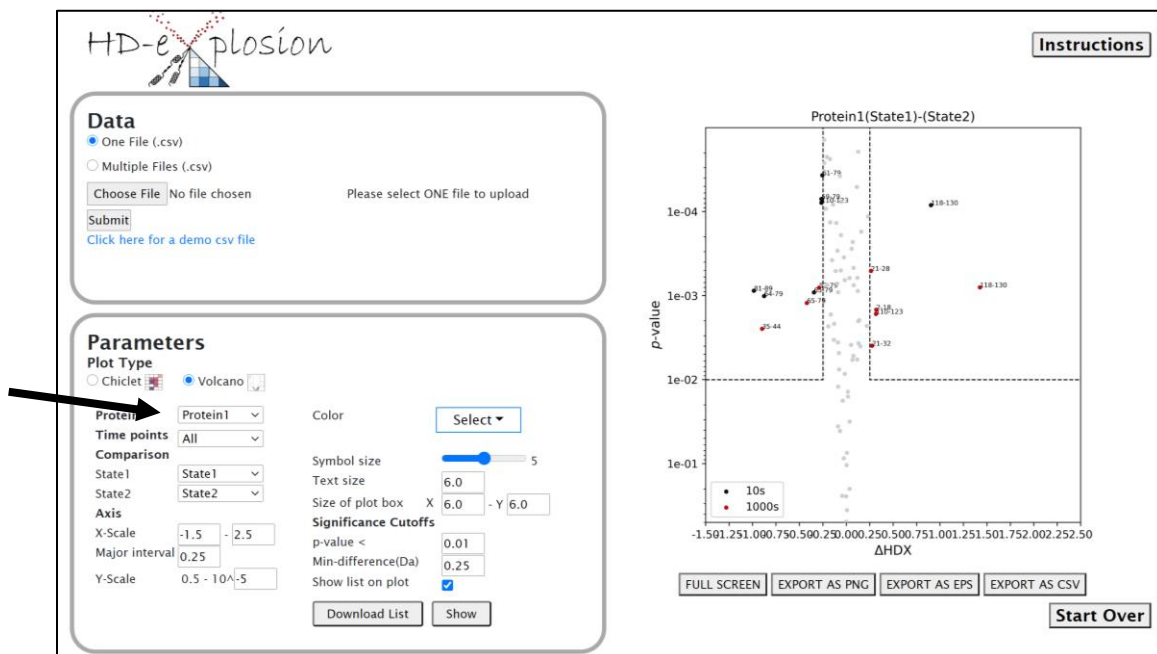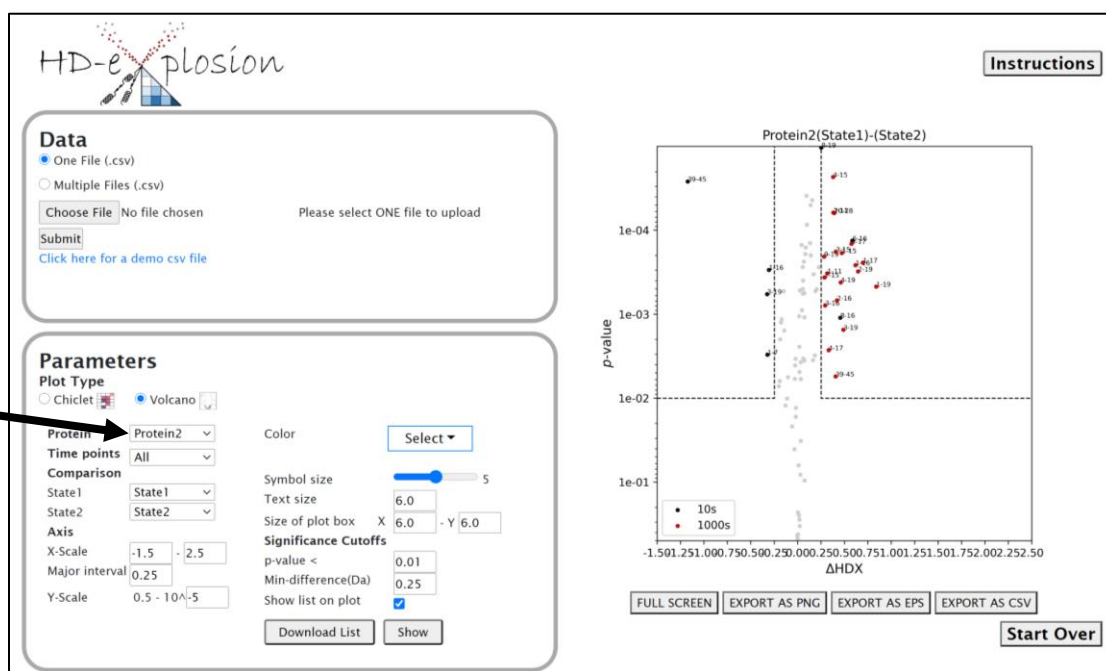
5. Output Files

The third panel of the GUI is where the chiclet or volcano plot will be displayed. You can enlarge the plot to fill the screen by clicking the 'FULL SCREEN' button. The plot can also be exported as a png, eps, or csv file. The eps file is useful if you wish to further beautify your plot in software such as Illustrator.

Once you have created the perfect plot for your input file, click the 'Start Over' button to do it all again for your next input file! The 'Start Over' button will clear all fields.

6. Referencing

Until we get this published… please acknowledge Naifu Zhang and Sheena D'Arcy from the University of Texas at Dallas.