



MHCpLogics standalone application

Step-by-step Tutorial & User Manual

©Written and developed in the Immunoproteomics research group, by M. Shahbazy et al.

Lab-page: [Purcell Lab](#)



Infection and Immunity Program, Monash Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Clayton, 3800 Victoria, Australia

Powered by MATLAB Runtime ver. 9.10 and App Designer

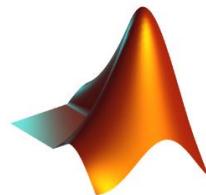


Table of contents

Code availability via GitHub	2
Installation of an executable standalone desktop application	2
1. Data Import.....	5
2. Analysis Setting	9
3. Basic Analysis	11
4. Data Preview	13
5. Cluster Analysis	15
6. Data Visualization	18
7. Multi-Data Comparative Analysis	21
8. Proteome DB search	23

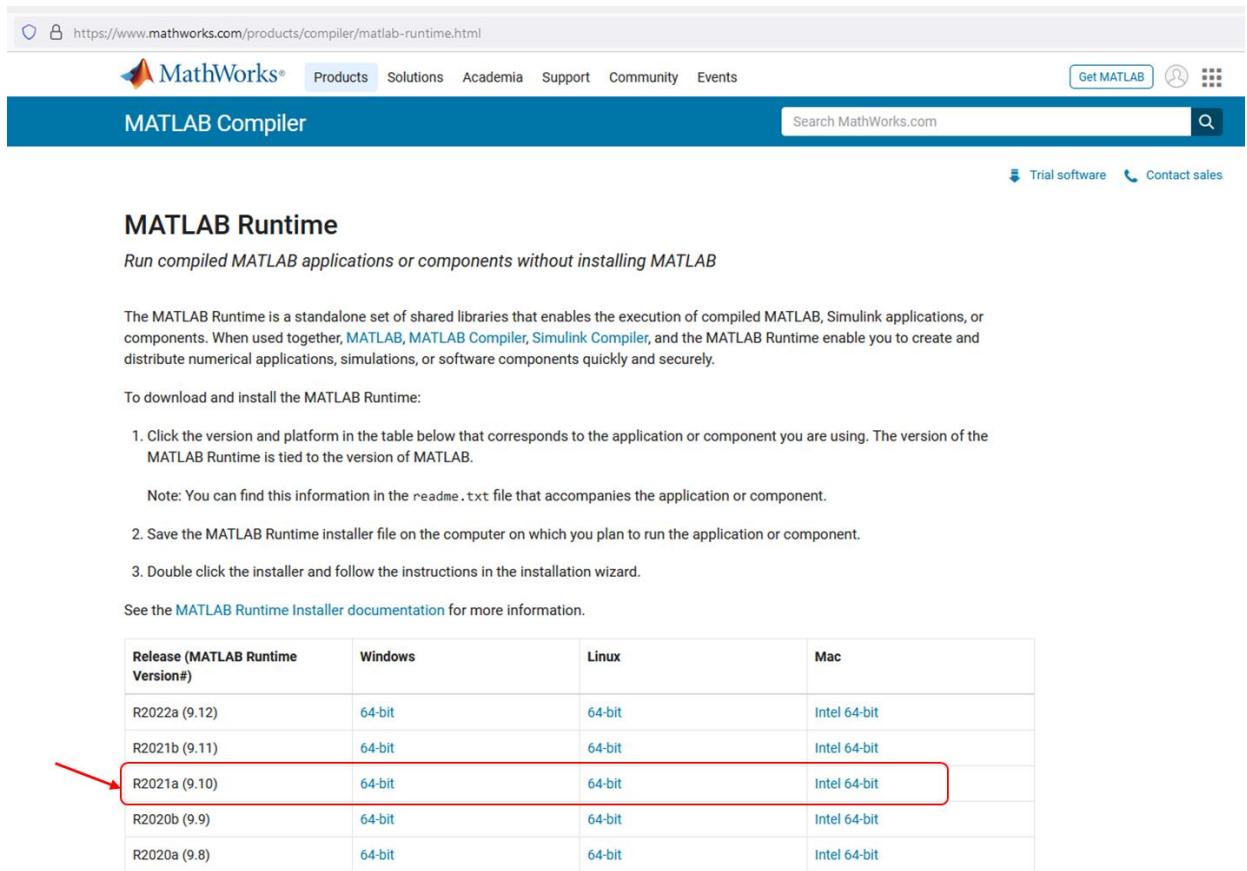
Code availability via GitHub

The MHCpLogics standalone application is available via <https://github.com/PurcellLab/MHCpLogics> GitHub repository that is installable and executable on Windows and Mac computer systems.

Installation of an executable standalone desktop application

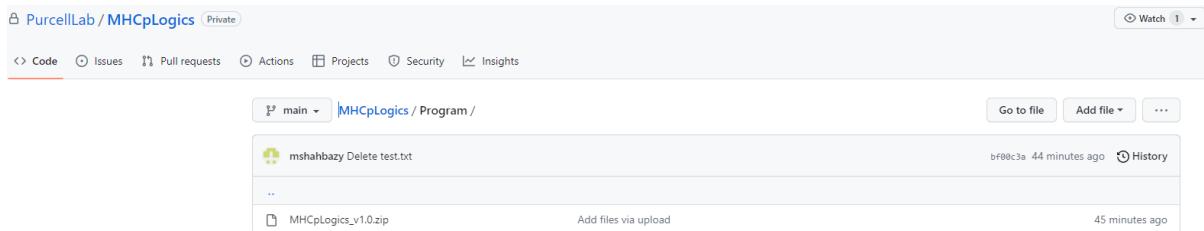
To install the standalone application on a PC (tested on Windows Desktop 10), please follow these step-by-step instructions:

- 1) Please, before the application installation, install the MATLAB Runtime ver. R2021a (9.10) here <https://www.mathworks.com/products/compiler/matlab-runtime.html>. This freely available set of shared libraries and code enables non-MATLAB users to install and execute the packaged application on computer systems without needing an installed version of MATLAB. *Note: End user must install MATLAB Runtime or specify the location of an installed MATLAB Runtime during the installation steps.

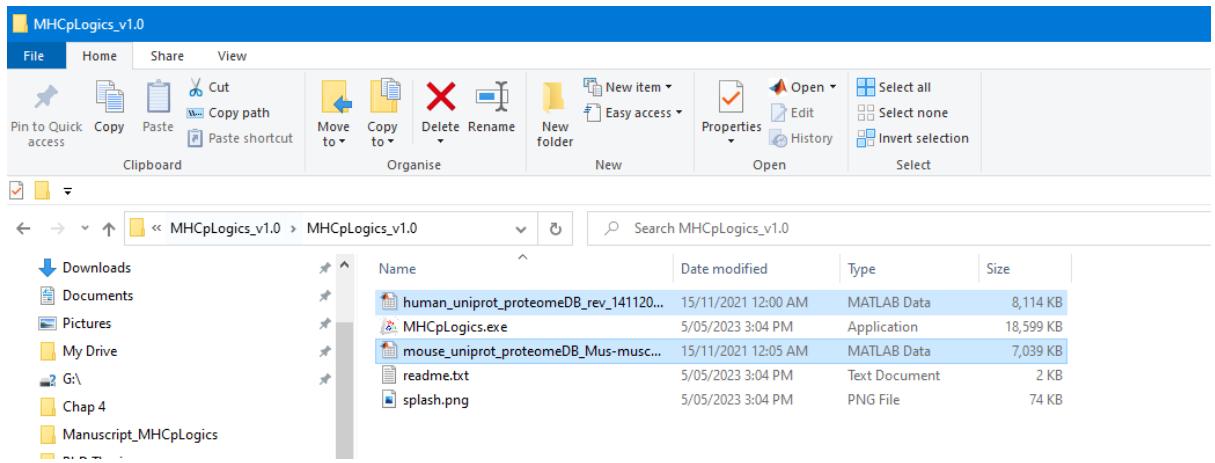
A screenshot of the MATLAB Compiler product page on the MathWorks website. The page title is "MATLAB Compiler". It features a search bar and navigation links for "Products", "Solutions", "Academia", "Support", "Community", and "Events". A "Get MATLAB" button is visible. The main content area is titled "MATLAB Runtime" with the subtitle "Run compiled MATLAB applications or components without installing MATLAB". It explains that the MATLAB Runtime is a standalone set of shared libraries for executing compiled MATLAB, Simulink, and MEX files. Below this, instructions for downloading the MATLAB Runtime are provided, along with a note about finding version information in the application's README file. A table lists MATLAB Runtime versions and their supported platforms. The R2021a (9.10) row is highlighted with a red box and a red arrow pointing to it.

Release (MATLAB Runtime Version#)	Windows	Linux	Mac
R2022a (9.12)	64-bit	64-bit	Intel 64-bit
R2021b (9.11)	64-bit	64-bit	Intel 64-bit
R2021a (9.10)	64-bit	64-bit	Intel 64-bit
R2020b (9.9)	64-bit	64-bit	Intel 64-bit
R2020a (9.8)	64-bit	64-bit	Intel 64-bit

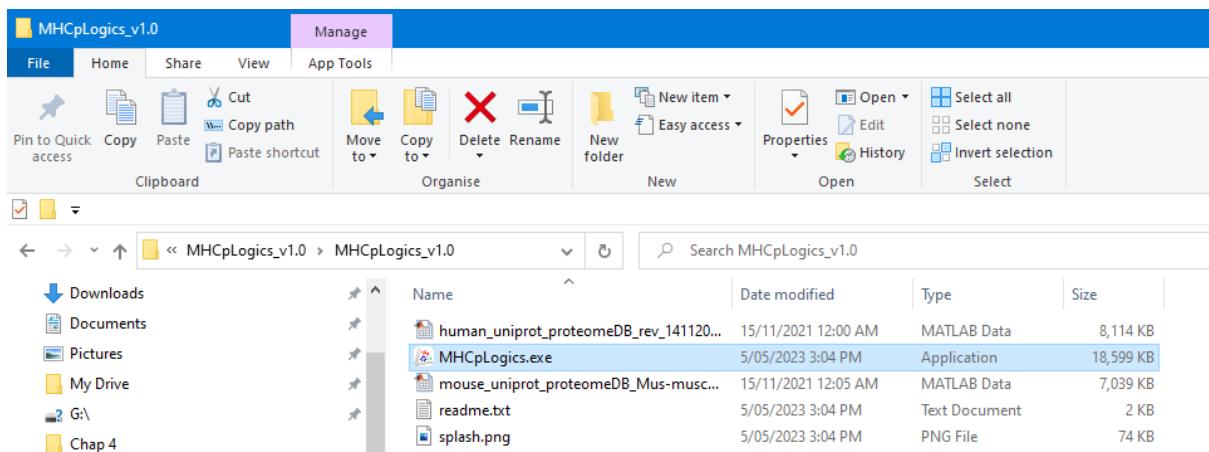
- 2) Now download the application package via the MHCpLogics GitHub repository, <https://github.com/PurcellLab/MHCpLogics>.
<https://github.com/PurcellLab/MHCpLogics/tree/main/Program>

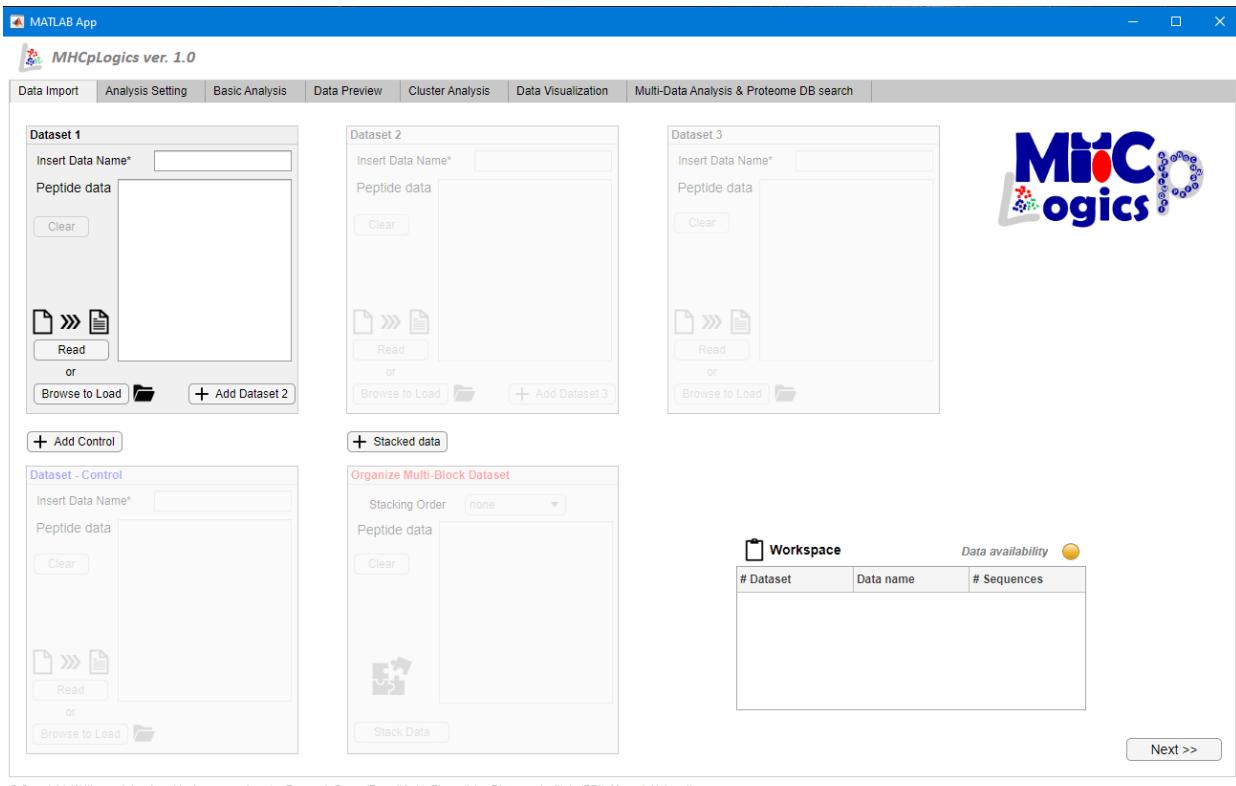
A screenshot of a GitHub repository page for "PurcellLab / MHCpLogics". The repository is private. The "Code" tab is selected. The main branch is "main". The commit history shows a single commit by "mshahbazy" that deleted a file named "test.txt". Below the commit, there is an uploaded file named "MHCpLogics_v1.0.zip". The commit was made 44 minutes ago, and the file upload was made 45 minutes ago.

- 3) Extract the zipped folder of the package.
- 4) Download both human and mouse databases from <https://github.com/PurcellLab/MHCpLogics/tree/main/Database>, and copy them into the unzipped folder: "MHCpLogics_v1.0"



- 5) Double-click on the “MHCpLogics.exe” installer to run the software tool (please make sure that you have already installed the MATLAB Runtime).
- 6) Please wait until seeing the GUI (software tool).

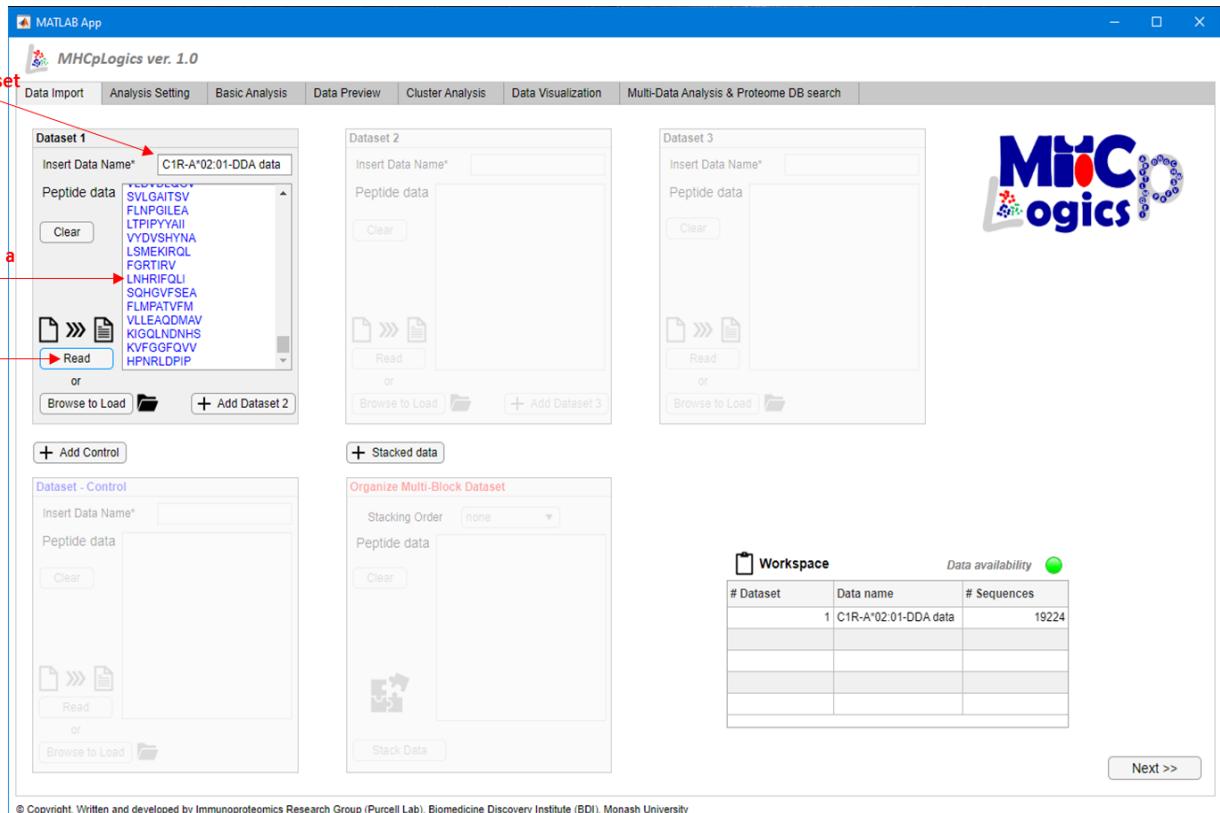




1. Data Import

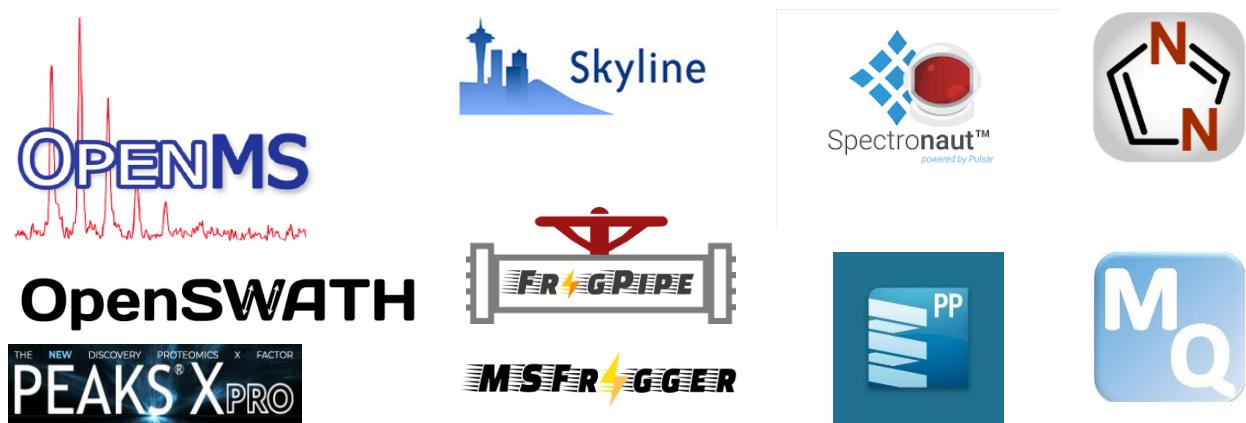
There are two straightforward approaches to importing data:

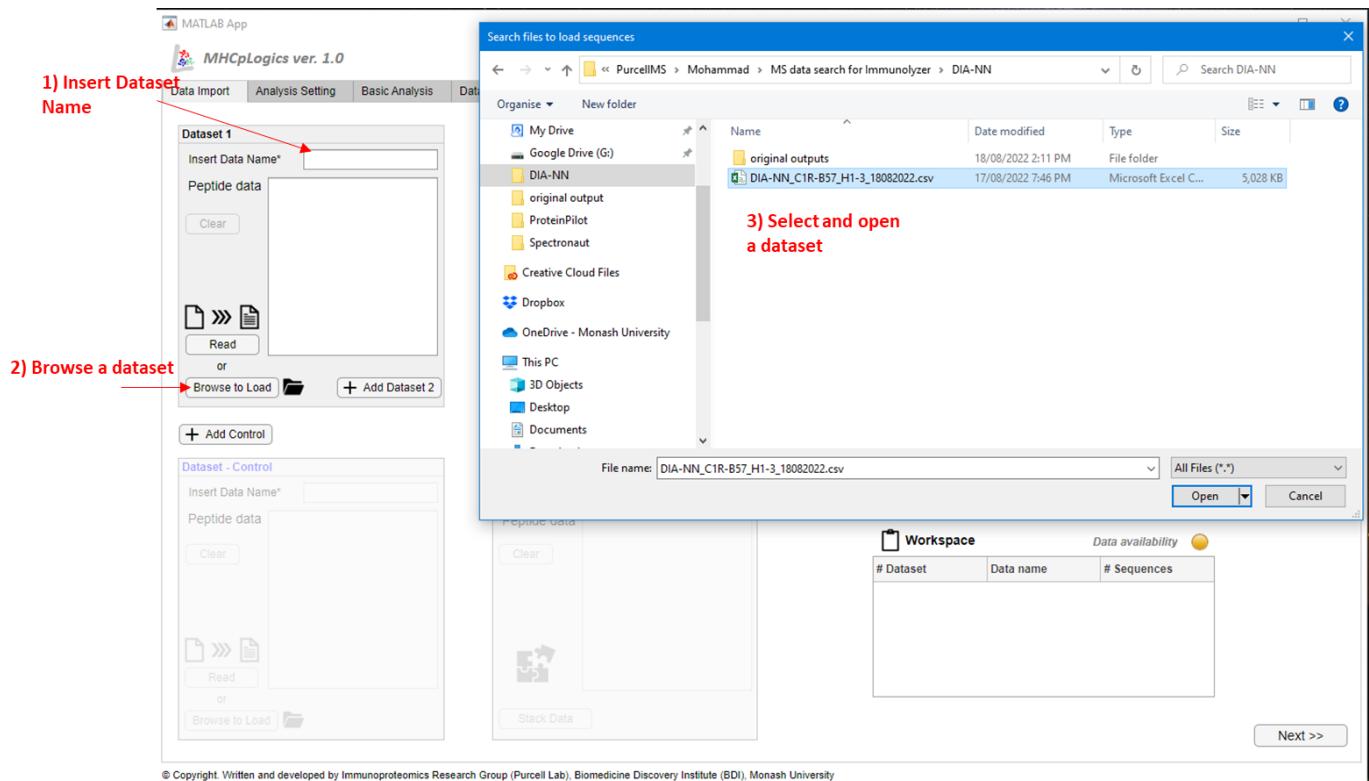
- Copy-and-paste the peptide sequences list and then push the “Read” button. Before importing, you can name the data by inserting it into the “Insert Data Name.”



- B) Push “Browse to load” and select peptide data file as *txt or *csv from your computer to read peptide data automatically.

**Note: MHCpLogics is compatible with reading and importing the outlined results/search files (as CSV) directly exported from popular mass spectrometry (MS) data processing software tools like OpenMS, OpenSWATH, Skyline, Spectronaut, MaxQuant, MSFagger, DIA-NN, Protein Pilot, and PEAKS.*





To add the next datasets, the user can easily push the “Add dataset” in each panel to add the subsequent data panel.

The workspace (placed right-bottom) is an updatable table sheet describing a summary of the imported dataset. Once one dataset is imported successfully, the orange lamp (on the top of the workspace) turns green to show data availability.

The screenshot shows the MHCpLogics software interface with three datasets imported:

- Dataset 1:** Insert Data Name: HLA-B*57.01. Peptide data sequence: GKVKGVINGFGRGRGLVTR AAFNSGKVDIVAIND GDEDEAESATGKRAED DEDDOVDTKQTKTDEO QIEPEVPSERLPEKIPENV PVKARPROAELVAAS FIGNYASTFTWYDRIFGTD SOYNAYNEKRKXFEKTE PFTONAEATTGGYOPPVY PNTAYASPGAVSVDNYQOP SSKEVKKSVEPSEVKQATS TSGPASAVADPPSTEKEID
- Dataset 2:** Insert Data Name: HLA-B*57.03. Peptide data sequence: LSLPAKAOPOSRSDDSSDS DSSSEEEEEEKTSK PALKRKARREREAK/KFEEERY KTGKKNVYIFFOKLRF FODVAONPRANMSKQYOSNP KVMNLISLAKSFQGQA PKGKKAKGKVKVAPAVPK KQEAKKVNVPLFEKRPK AYGYTFV/SISDERFLDELE DEAKAARALARASGS SSKEVKKSVEPSEVKQATS TSGPASAVADPPSTEKEID
- Dataset 3:** Insert Data Name: HLA-B*58.01. Peptide data sequence: TPEVNGDPPVQVYV KTGKKNVYIFFOKLRF YSWWIGGSILASLSLTFQW WISKOEYDESPGSIVHR FIGNYASTFTWYDRIFGTD SOYNAYNEKRKXFEKTE KYSVWIGGSILASLSLTFQW MWSKOEYDESPGSIVHR SSKEVKKSVEPSEVKQATS TSGPASAVADPPSTEKEID KSGSSPKWTHDKYQGDGI VDEEETMENNEEKDRR KEEKE

Add another dataset button is highlighted with a red arrow.

The workspace shows a summary of imported datasets

This green light is an indicator showing the user imported at least one dataset to analyze

Workspace table:

# Dataset	Data name	# Sequences
1	HLA-B*57.01	2673
2	HLA-B*57.03	3168
3	HLA-B*58.01	2526
Control	HLA-C*04.01/B*35.03	4542
Stacked	multi-block data	8367

Next >>

MHCpLogics can stack different datasets in a customized order for multi-data comparative analysis.

Please follow the instructions below.

The screenshot shows the MHCpLogics software interface with three datasets imported and stacked:

- Dataset 1:** Insert Data Name: C1R-A*02.01-DDA data. Peptide data sequence: FLNPOILEA LTRIPYVAIL YDVDFSYHNA LSMEKIRQL FGRTIRV LNIHRIFOLI SOHGVFSEA FLMPATVFM VLLEAQDMAV KIGOLNDNHS KVFGGFQVV HPNRLDPIP
- Dataset 2:** Insert Data Name: C1R-B*07.02. Peptide data sequence: RPAQNNATF SPSSLFSAL FPTSLGQAEAL GGPVSVLCTRLY IPTEQVNEL RNQOEAVGL SPSSKHQLL APNWAEVLV PSELORPSL APLYTPSL SPRPQTEL KPKYPPNKM
- Dataset 3:** Insert Data Name: HLA-C*04.01. Peptide data sequence: TPEVNGDPPVQVYV KTGKKNVYIFFOKLRF YSWWIGGSILASLSLTFQW WISKOEYDESPGSIVHR FIGNYASTFTWYDRIFGTD SOYNAYNEKRKXFEKTE KYSVWIGGSILASLSLTFQW MWSKOEYDESPGSIVHR SSKEVKKSVEPSEVKQATS TSGPASAVADPPSTEKEID KSGSSPKWTHDKYQGDGI VDEEETMENNEEKDRR KEEKE

1) Organize stacked dataset button is highlighted with a red arrow.

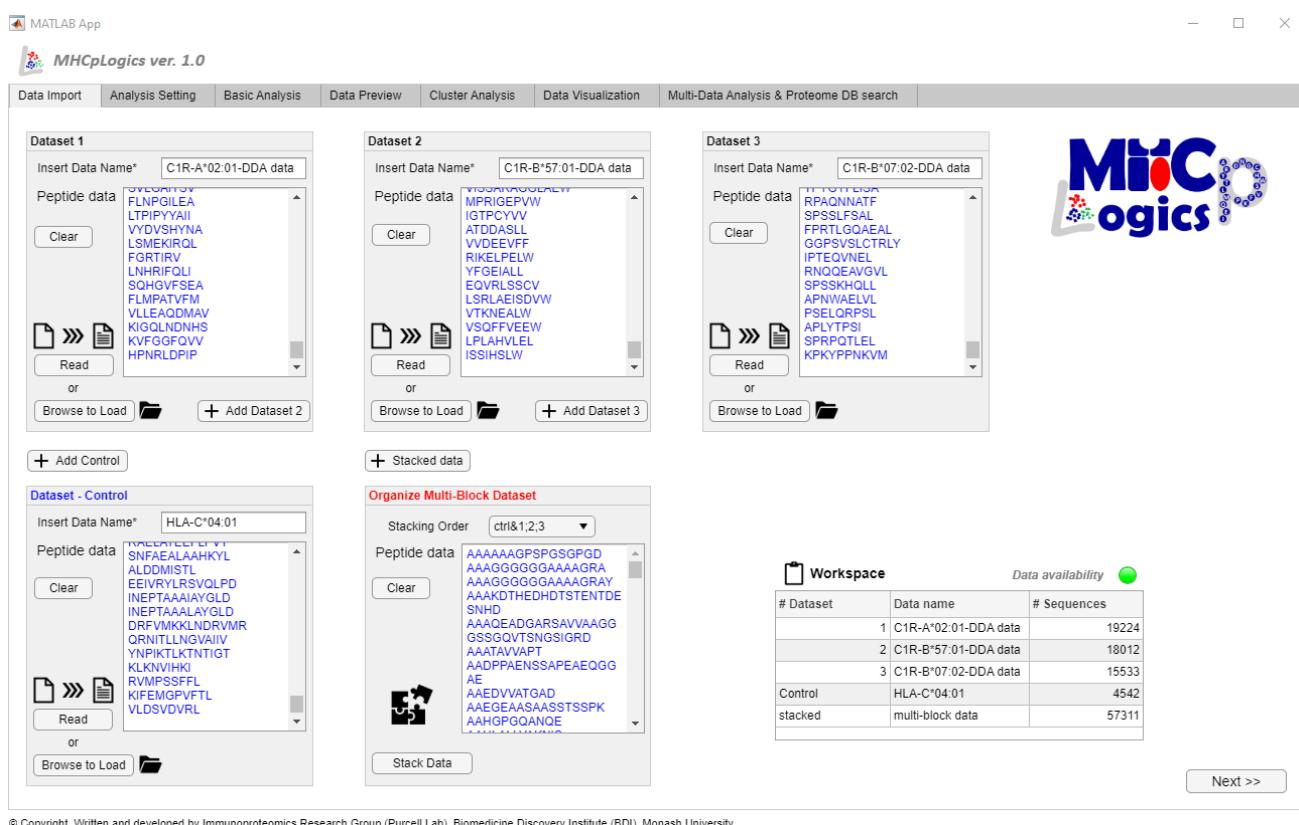
2) Specify stacking order dropdown menu is highlighted with a red arrow. Options include: none, 1&2, 1&3, 2&3, 2&1, 3&1, 3&2, 1:2, 1:3, 2:1, 2:3, 3:1, 3:2, 1:2, 1:3, 2:1, 2:3, 3:1, 3:2, ctrl&1, ctrl&2, ctrl&3, ctrl&1:2, ctrl&1:3, ctrl&2:3, ctrl&1;2, ctrl&1;3, ctrl&2;3, ctrl&1;2;3.

3) Stack the datasets button is highlighted with a red arrow.

Workspace table:

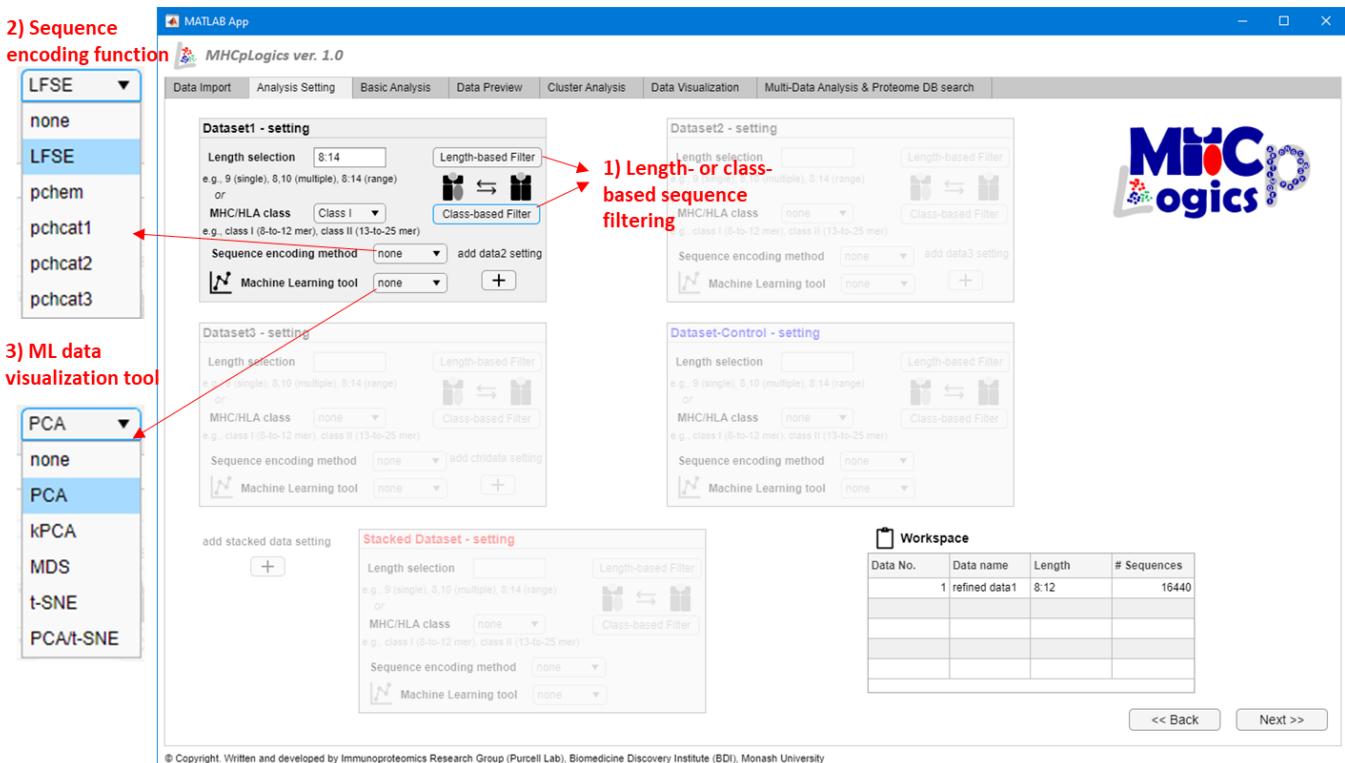
# Dataset	Data name	# Sequences
1	C1R-A*02.01-DDA data	19224
2	C1R-B*07.02-DDA data	18012
3	C1R-B*07.02	15533
Control	HLA-C*04.01	4542

Next >>



2. Analysis Setting

After data importing, the user can set the analysis parameters in the next tab. This setting includes sequence length-based filtering (as a single length, multiple lengths, and a customized range), class-based MHC binders filtering with predefined ranges (i.e., MHC class I [8-12 mer] and class II [12-25 mer]), sequence scoring functions to generate a numerical data matrix according to the peptide sequence residues for HLA-I allotypes encoding functions (i.e., *LFSE*, *pchem*, *pchcat1*, *pchcat2*, and *pchcat3*), and ML tool (i.e., PCA, kPCA, MDS, t-SNE, PCA-t-SNE) for the data visualization.

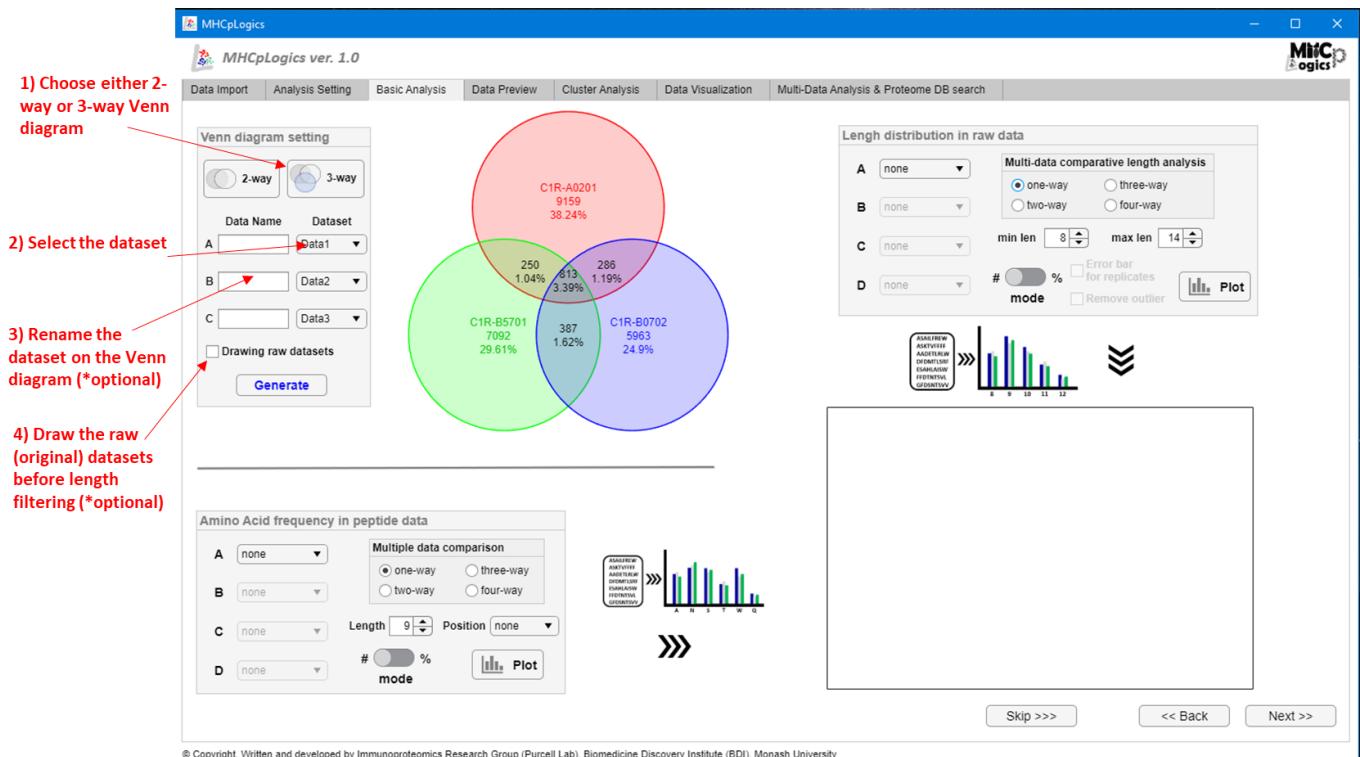


3. Basic Analysis

In the “Basic Analysis” tab, three analyses are available to compare the imported datasets:

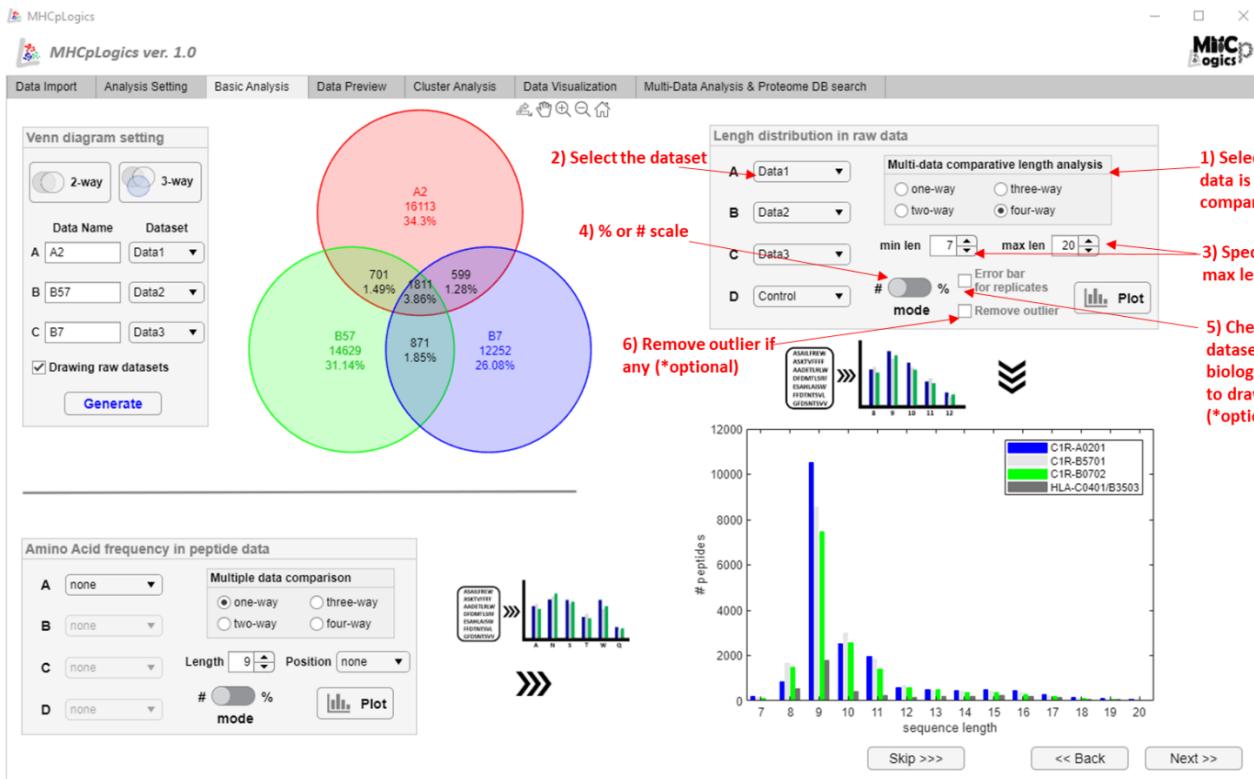
- A) Venn diagram to analyze overlap between datasets.

- A.1. For this first push, either “2-way” or “3-way” select the Venn diagram type to compare two or three datasets, respectively.
- A.2. Select the dataset.
- A.3. Rename the dataset on the Venn diagram (*optional).
- A.4. Draw the original datasets (raw data without length- or class-based filtering - *optional).



- B) Length distribution of immunopeptides

- B.1. Multi-data comparative length analysis >> Select how many datasets are involved in comparative analysis (e.g., two-way means that the user is interested in analyzing and comparing two datasets).
- B.2. Select the dataset (e.g., Data1).
- B.3. Specify the minimal and maximal lengths of the sequences to draw.
- B.4. Toggle the number (#) or percentage (%) scale of the length distribution graph.
- B.5. Check the error bar box if datasets are technical or biological replicates to draw error bars (*optional).
- B.6. Check the outlier removal box to remove outliers, if any (*optional).



© Copyright. Written and developed by Immunoproteomics Research Group (Purcell Lab), Biomedicine Discovery Institute (BDI), Monash University

C) Amino acid frequency in peptide data

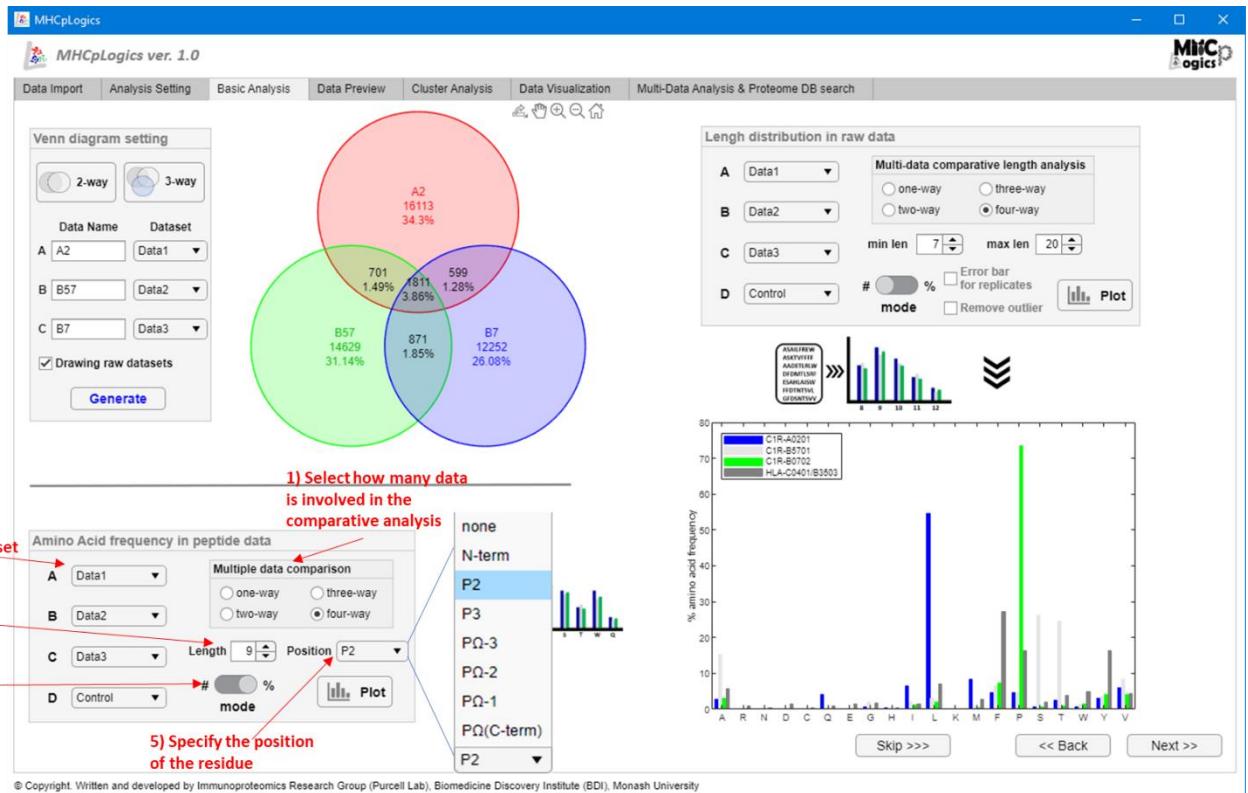
C.1. Multi-data comparison >> Select how many datasets are involved in comparative analysis (e.g., two-way means that the user is interested in analyzing and comparing two datasets).

C.2. Select the dataset (e.g., Data1).

C.3. Specify the length of the sequences if there are peptides with different lengths in the data.

C.4. Toggle the number (#) or percentage (%) scale of the amino acid frequency graph.

C.5. Specify the position of residues to analyze amino acid frequency in a specific position (i.e., N-terminus, P2, P3, PΩ-3, PΩ-2, PΩ-1, PΩ [C-terminus]).



4. Data Preview

- A) In the “Data Preview” tab, the user can evaluate the immunopeptide data space by using unsupervised data visualization and dimensionally reduction tools selected in the “Analysis Setting” tab. It gives insight into how many clusters exist visually.

A.1. Select the dataset to preview.

A.2. Run ML-dimensionally reduction method for data visualization.

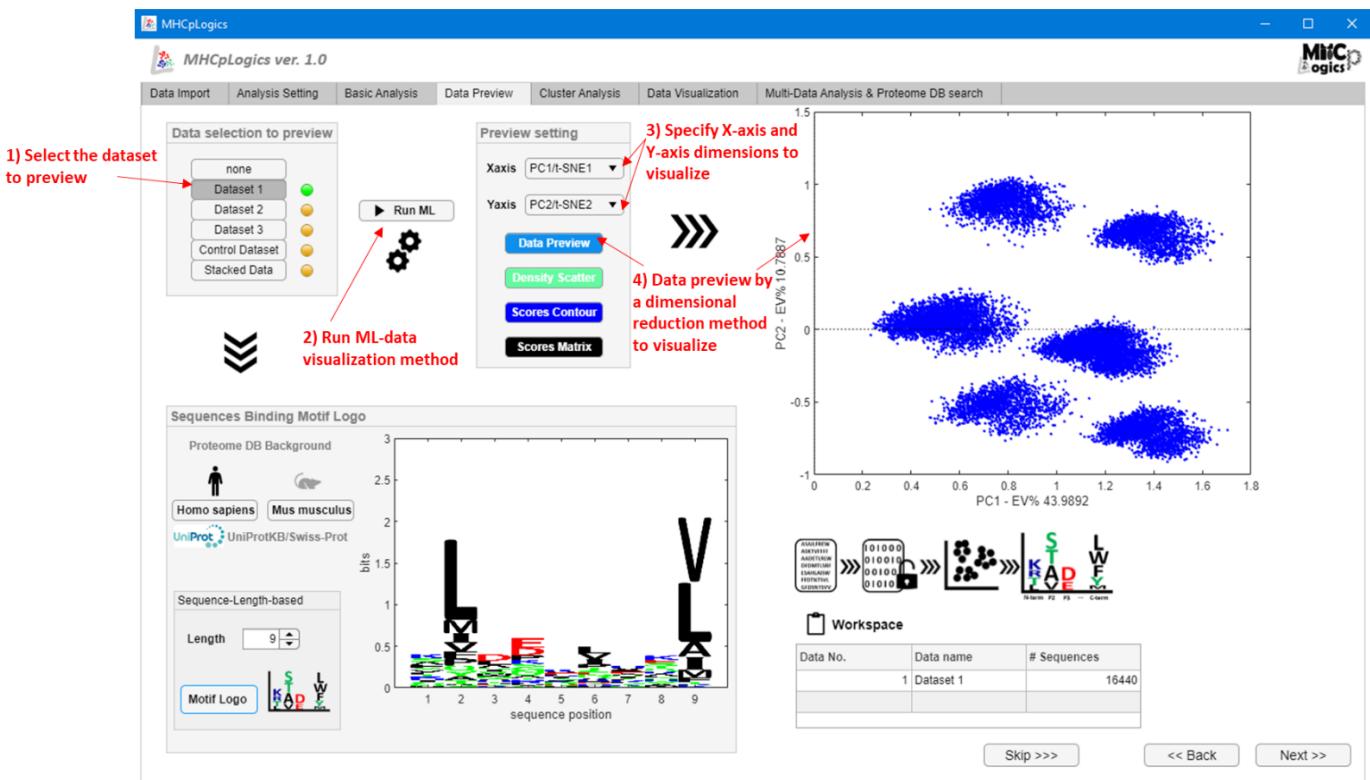
A.3. Specify X-axis and Y-axis dimensions and select the new transformed variables generated by the dimensional reduction, like PCs from PCA to visualize.

A.4. Data preview by a dimensional reduction method to visualize.

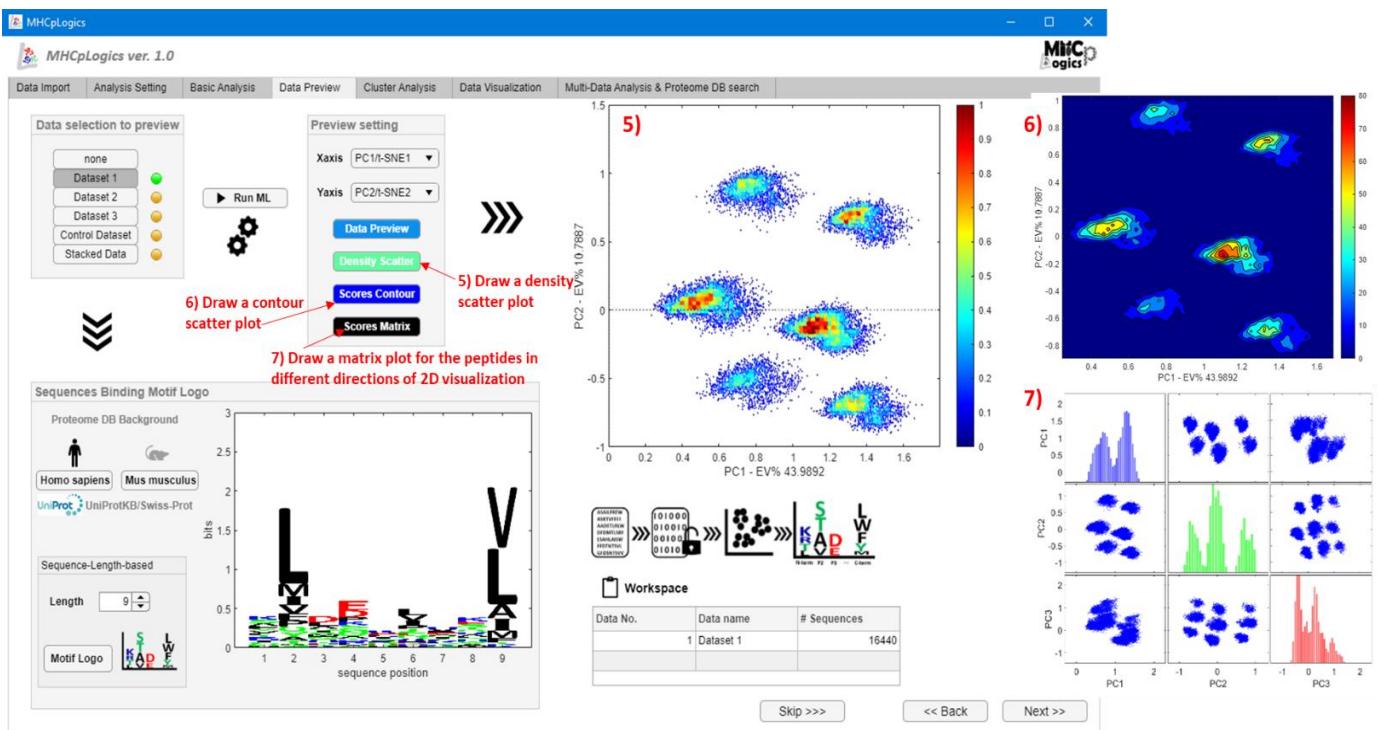
A.5. Draw a density scatter plot.

A.6. Draw a contour scatter plot.

A.7. Draw a matrix plot for the peptides in different directions of 2D visualization to explore data space.

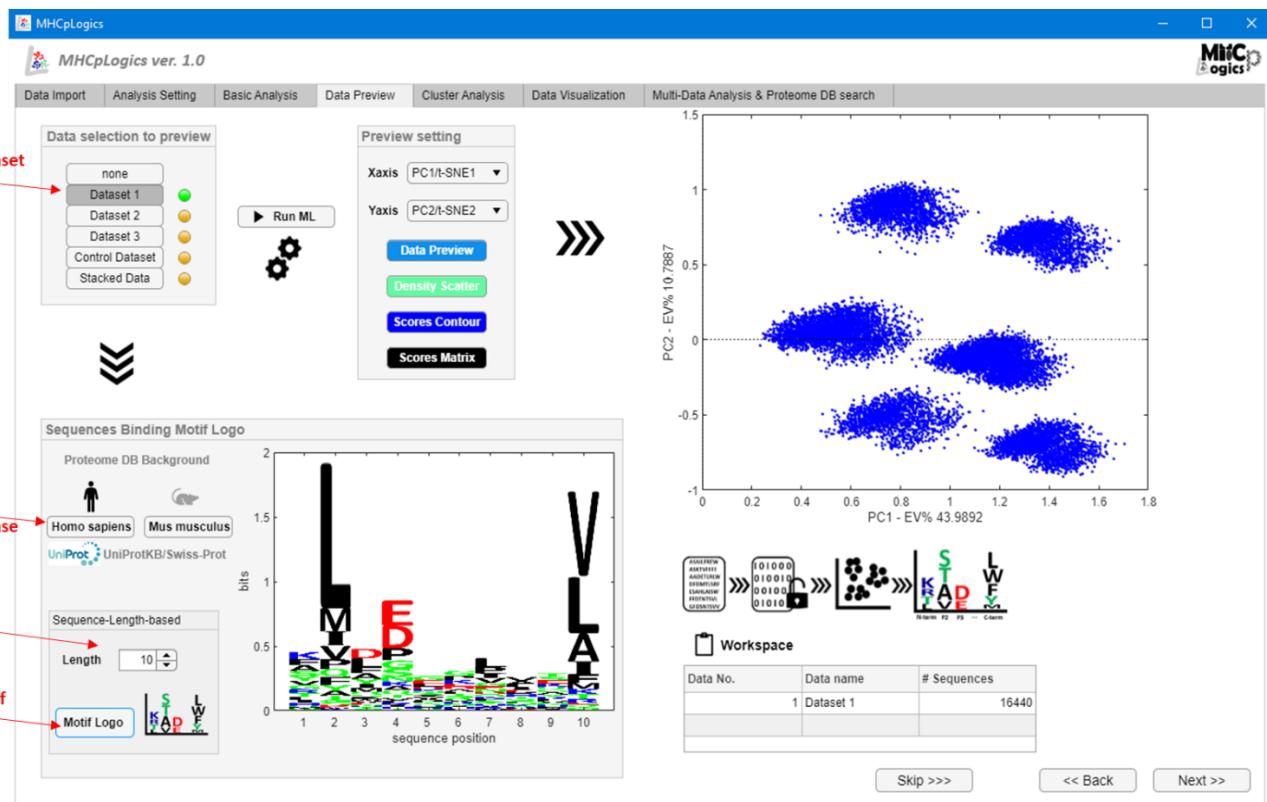


© Copyright. Written and developed by Immunoproteomics Research Group (Purcell Lab), Biomedicine Discovery Institute (BDI), Monash University



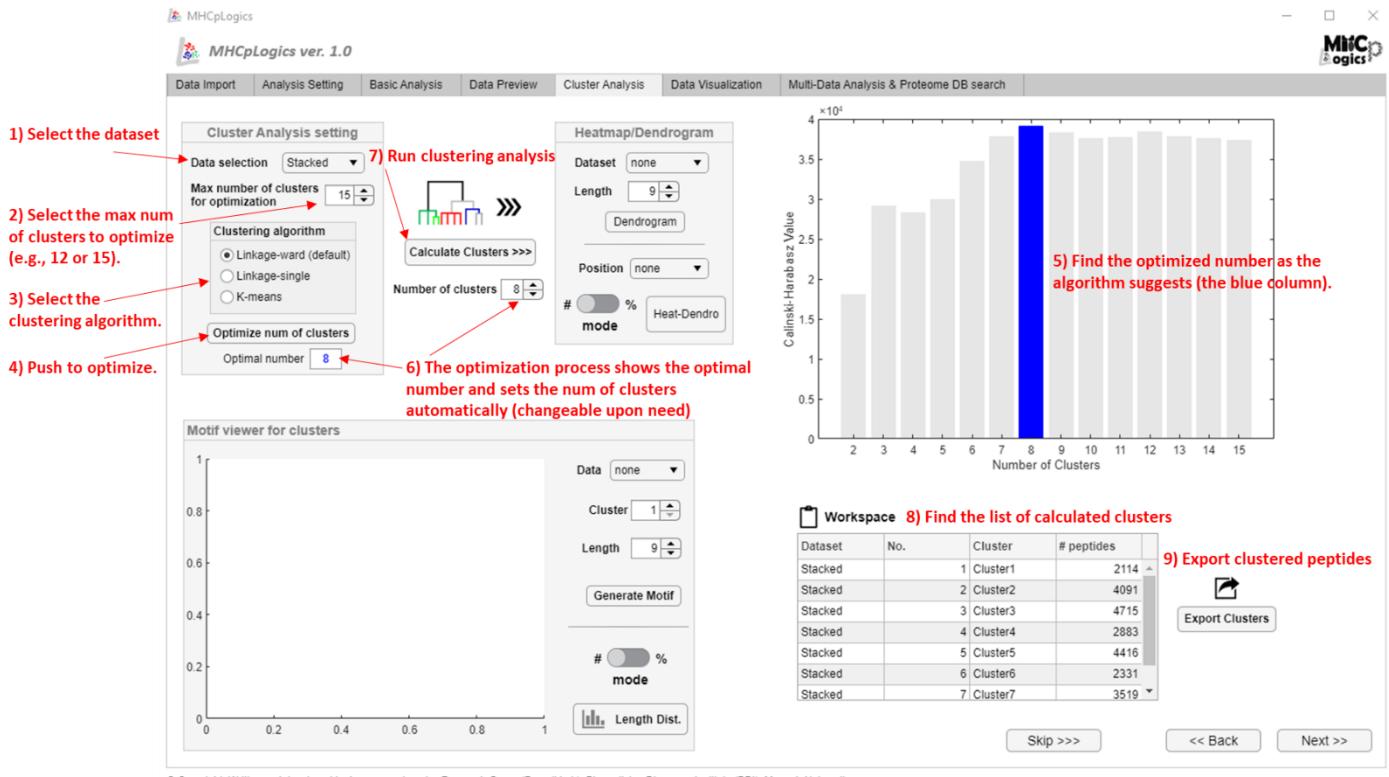
© Copyright. Written and developed by Immunoproteomics Research Group (Purcell Lab), Biomedicine Discovery Institute (BDI), Monash University

- B) Moreover, a motif viewer allows checking sequence binding motif logos per dataset.
- Select the dataset to preview.
 - Select the background proteome database.
 - Select a sequence length.
 - Draw the Motif logo.



5. Cluster Analysis

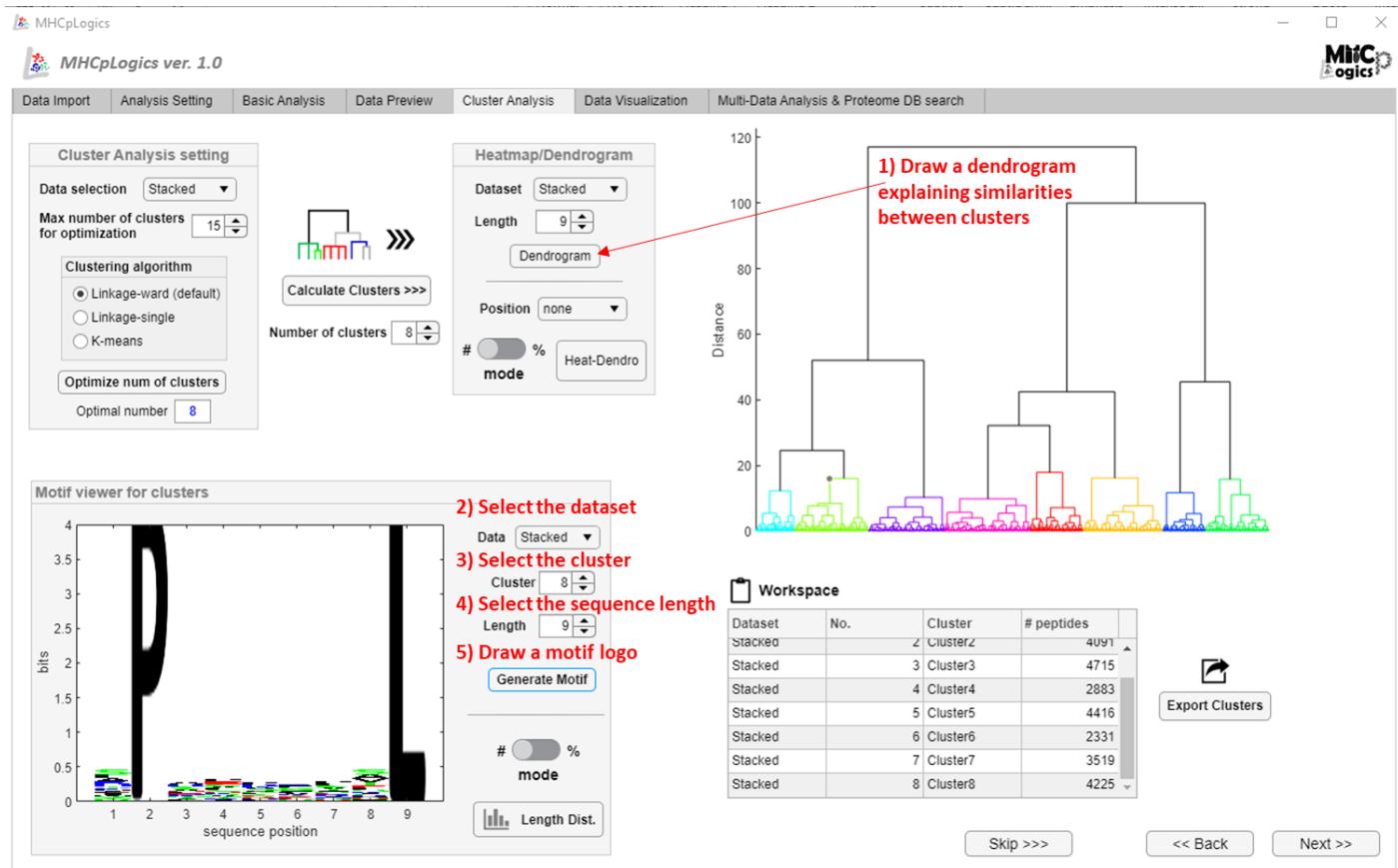
- A) To perform clustering analysis, please follow the step-by-step instructions here:
- Select the dataset for clustering.
 - Select the max num of clusters to optimize (e.g., 12 or 16).
 - Select the clustering algorithm, including Linkage Ward (default and recommended in most cases), Linkage Single, and K-means.
 - Push to run the optimization process.
 - Find the optimized number as the algorithm suggests (the blue column on the top-right bar graph).
 - The optimization process shows the optimal number and automatically sets the number of clusters (changeable upon need).
 - Run clustering analysis.
 - Find a list of calculated clusters with the number of peptide IDs per each.
 - Export clustered peptides as CSV file.



© Copyright. Written and developed by Immunoproteomics Research Group (Purcell Lab), Biomedicine Discovery Institute (BDI), Monash University

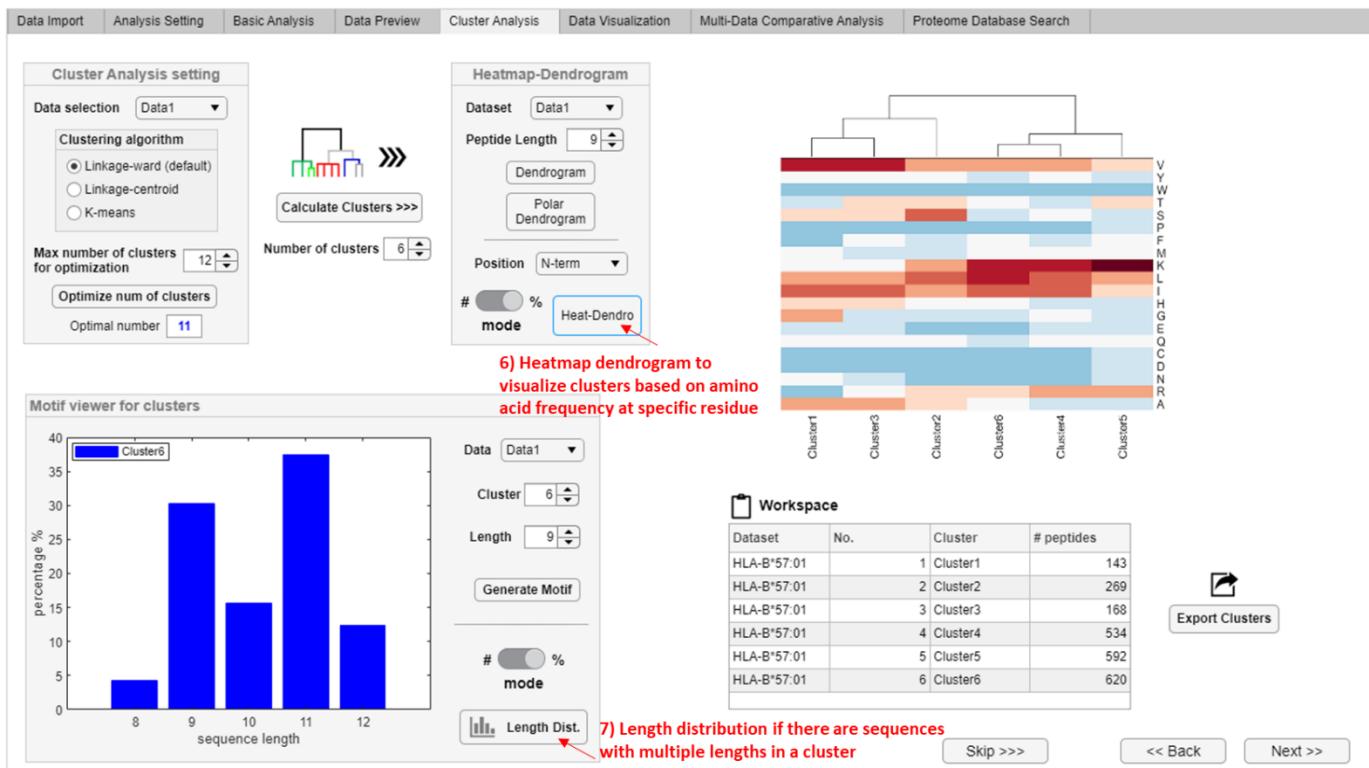
B) To further analysis of the clusters, we can do the following instructions:

1. Drawing a hierarchical dendrogram to visualize the clusters showing the relationship between the clusters.
2. First, select the dataset to check the motif per cluster.
3. Select the (sub)cluster to view its binding (sub)motif.
4. Select the length of the peptides for the clusters with multiple lengths.
5. Draw the motif.
6. Draw a heatmap-dendrogram to visualize clusters based on the amino acid frequency at specific residues (position). Select the dataset, length, residue position, and mode, then draw.
7. Analysis of the length distribution if peptides in the cluster have multiple lengths in two modes (i.e., # and %).



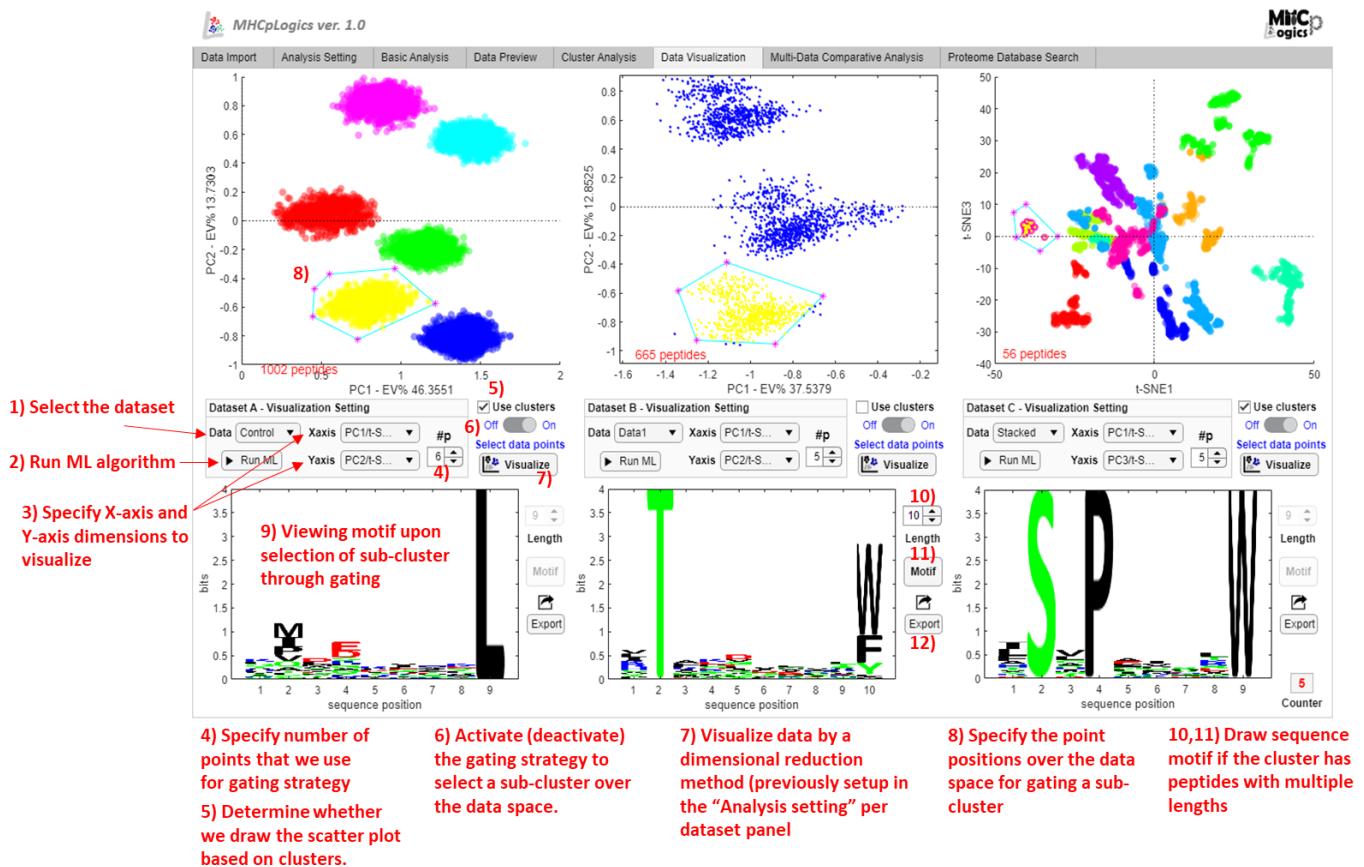
© Copyright. Written and developed by Immunoproteomics Research Group (Purcell Lab), Biomedicine Discovery Institute (BDI), Monash University

Heat dendrogram and length distribution analysis



6. Data Visualization

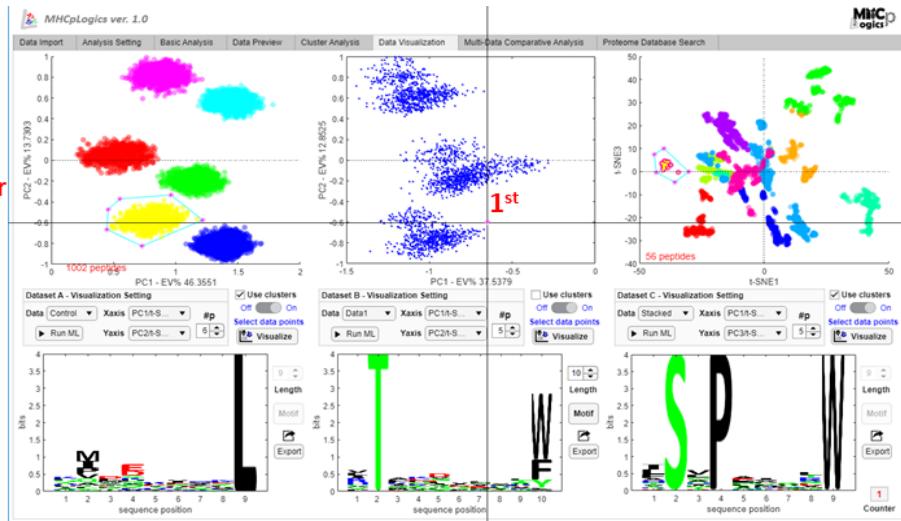
- A) To visualize data, please follow the step-by-step instructions here:
- A.1. Select the dataset.
 - A.2. Run the ML algorithm.
 - A.3. Specify x-axis and y-axis to visualize data by a scatter plot (e.g., PC1 vs. PC2).
 - A.4. Specify the number of points we use for the gating strategy (e.g., 4, 5, etc.).
 - A.5. Determine whether we draw the scatter plot based on clusters.
 - A.6. Activate (deactivate) the gating strategy to select a sub-cluster over the data space (please see the next section – 6B to familiarize yourself with the gating tool).
 - A.7. Visualize data by a dimensional reduction method (previously set up in the “Analysis setting” per dataset panel).
 - A.8. Specify the positions of the points over the data space for gating a sub-cluster.
 - A.9. Viewing motif upon selection of sub-cluster through the gating method.
 - A.10-11. Draw a sequence motif if the cluster has peptides with multiple lengths (for a gate with single-length peptides, the motif is drawn automatically).
 - A12. Export the peptides in the selected gate.



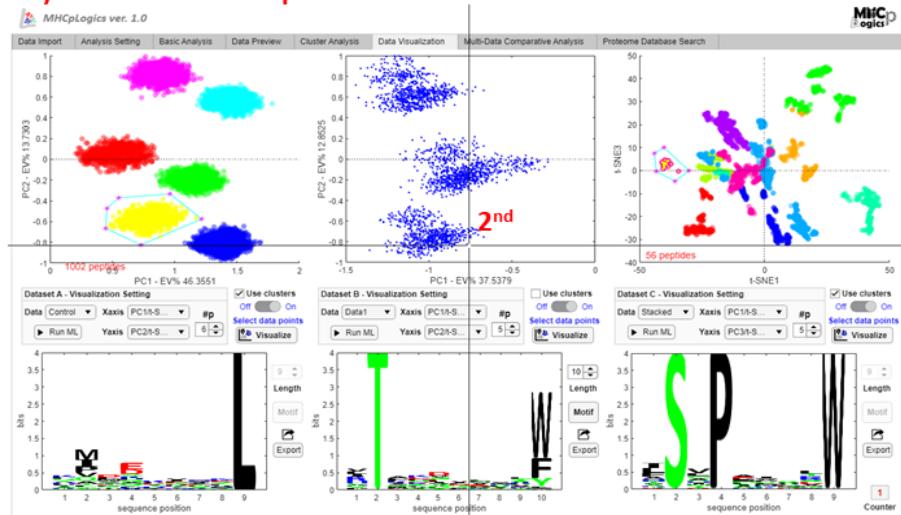
- B) Gating strategy to select a specific (sub)cluster over the data space
- B.1. Ensure you activate the select data points (see section 6A.6).
 - B.2. After pushing the Visualize button, a prompt appears to mention that you please select the data points over the data space. Please push “OK” and wait until seeing the sniper tool.

- B.3. After appearing in the sniper tool, please select the first point over the data space.
- B.4. Select the second point.
- B.5. Select the following points up to #p (e.g., here, #p was set to 5).
- B.6. One message is prompted to announce that you completed the gating selection.
- B.7. Now, you can check the motifs and proportional diagrams (in comparative analysis).

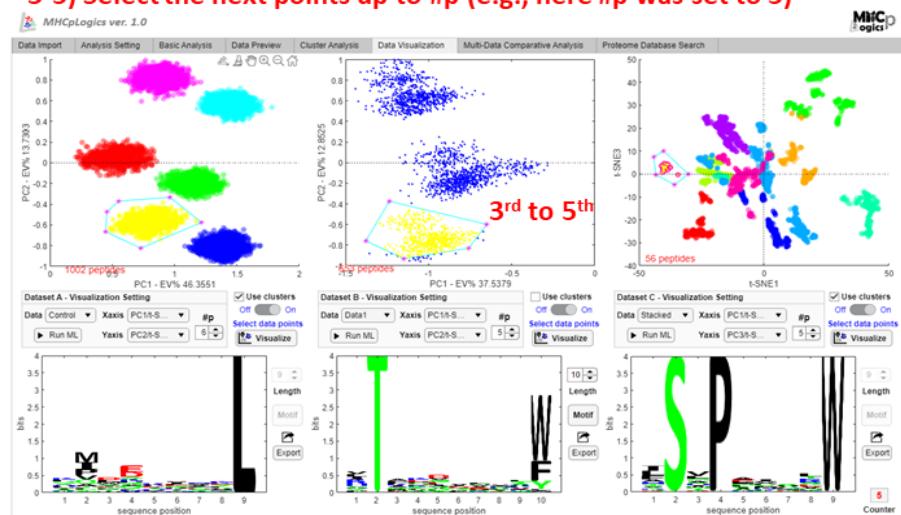
1) After appearing the sniper (please wait until seeing the sniper), select the first point over the data space



2) Select the second point

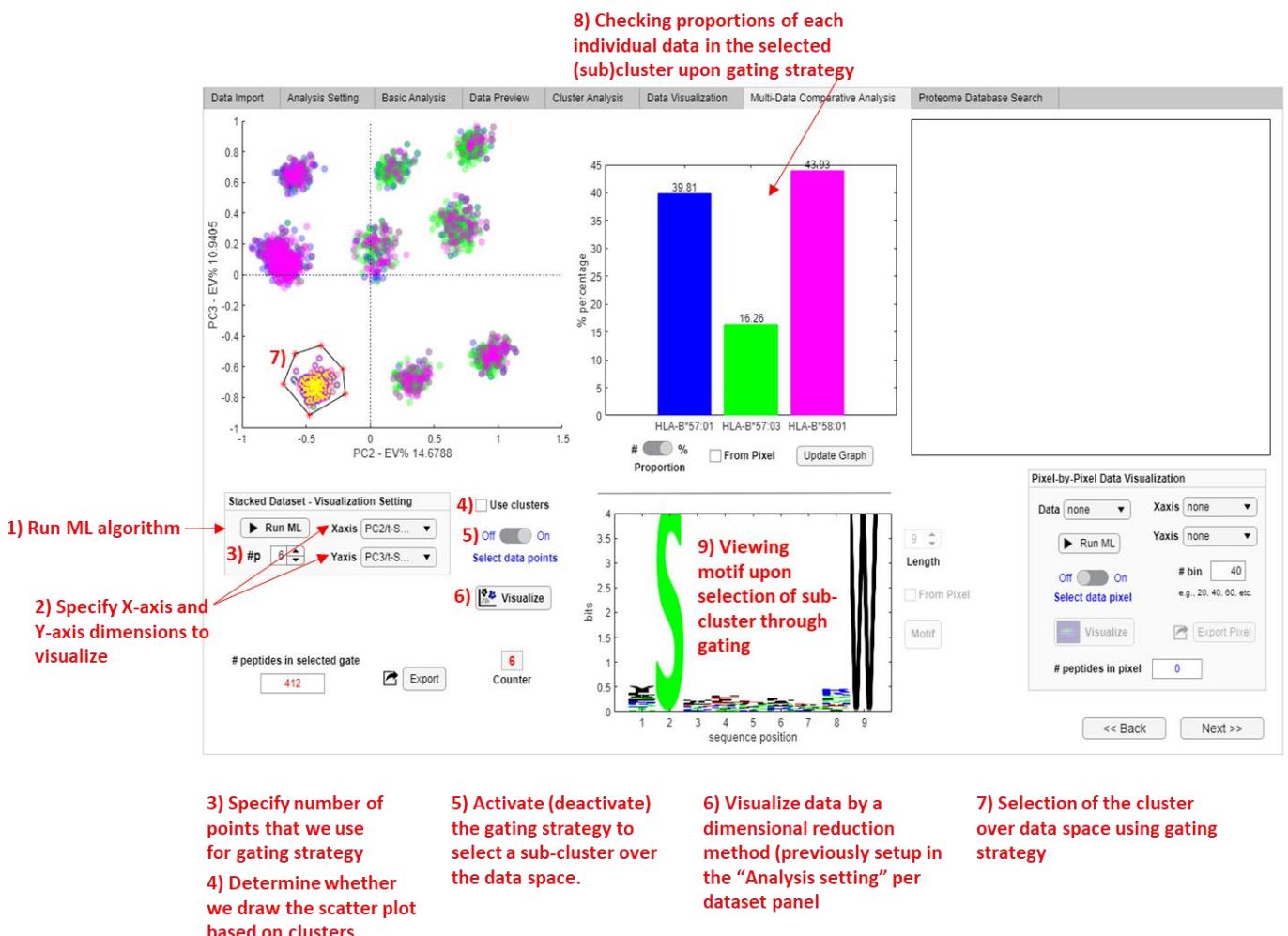


3-5) Select the next points up to #p (e.g., here #p was set to 5)



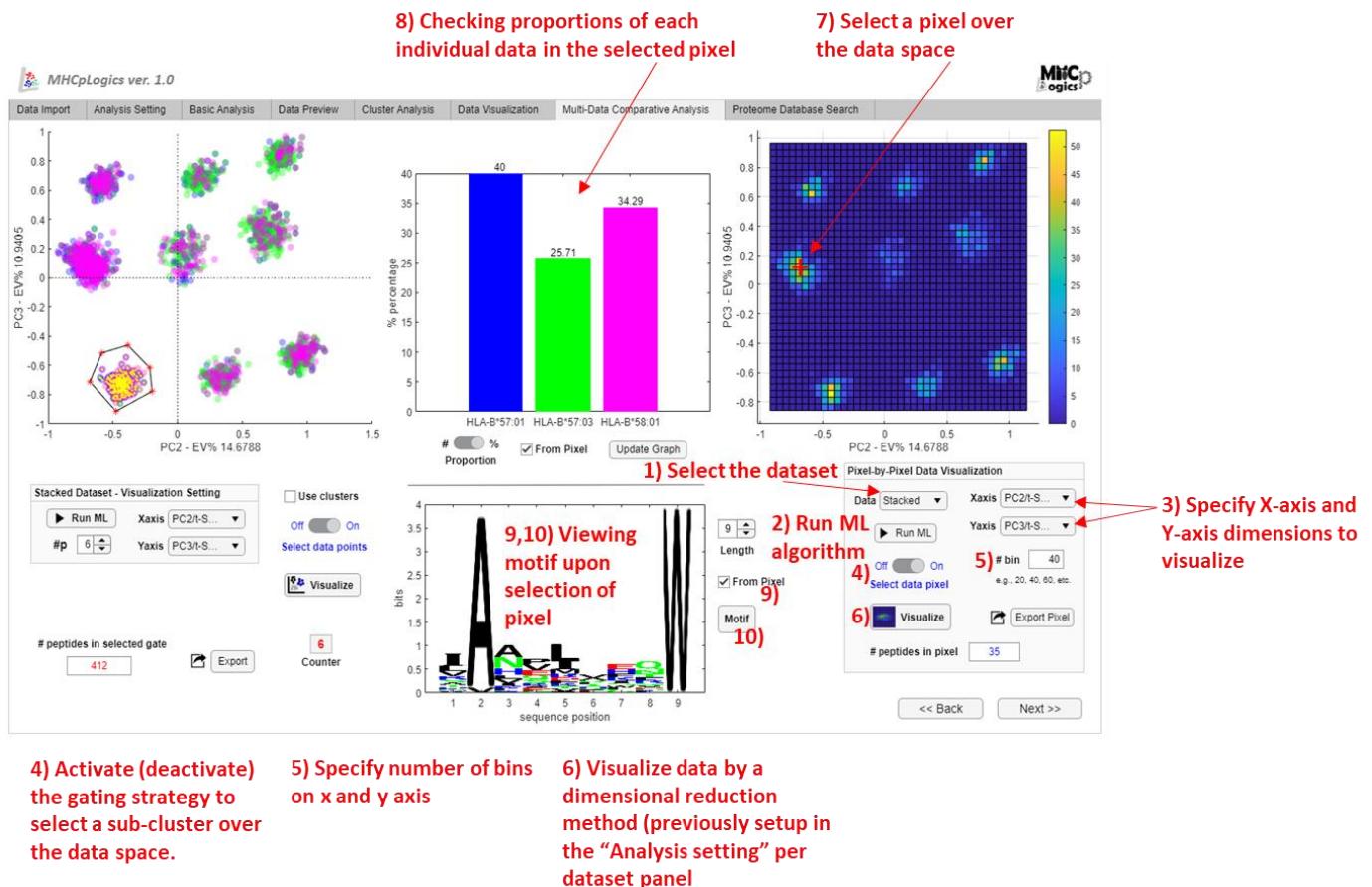
7. Multi-Data Comparative Analysis

- A) To perform a comparative analysis between multiple datasets, we stack the datasets in the first tab (Data import) by a desired order and then refine this stacked data in the “Analysis setting” section. Users can run basic analysis, data preview, and cluster analysis before data visualization and comparative analysis in the multi-data comparative analysis.
- A.1. First, we run an ML algorithm to visualize data.
- A.2. Specify the x-axis and y-axis to visualize data by a scatter plot (e.g., PC1 vs. PC2).
- A.3. Specify the number of points we use for the gating strategy (e.g., 5, 6, etc.).
- A.4. Determine whether we draw the scatter plot based on clusters derived from the cluster analysis.
- A.5. Activate (deactivate) the gating strategy to select a sub-cluster over the data space (please see the previous section – 6B to familiarize yourself with the gating tool).
- A.6. Visualize data by a dimensional reduction method (previously set up in the “Analysis setting” per dataset panel).
- A.7. Specify the positions of the points over the data space for gating a sub-cluster.
- A.8. Checking proportions of individual data in the selected (sub)cluster upon gating strategy.
- A.9. Viewing motif upon selection of sub-cluster through gating.



B) Pixel-by-pixel-based micro visualization of immunopeptidome data

- B.1. On the right panel (on the Multi-data comparative analysis tab), first select the dataset.
- B.2. Run ML algorithm for pixel-by-pixel micro-visualization.
- B.3. Specify x-axis and y-axis to visualize data by a scatter plot (e.g., PC1 vs. PC2).
- B.4. Activate (deactivate) the gating strategy to select a sub-cluster over the data space.
- B.5. Specify number of bins on x and y axis.
- B.6. Visualize data by a dimensional reduction method (previously setup in the “Analysis setting” per dataset panel).
- B.7. Select a pixel by a single click over the data space after appearing the pixel-sniper.
- B.8. Checking proportions of each individual data in the selected (sub)cluster upon gating strategy.
- B.9-10. Viewing motif upon the pixel selection. **Note: Make sure that check “From pixel” on the right side of the motif viewer.*



8. Proteome DB search

MHCpLogics can assign the peptides to their original proteins in the selected (sub)cluster (by gating strategy) or clusters (derived from the cluster analysis) for determining protein accession by searching the peptide data against the reviewed human (*Homo sapiens*) or mouse (*Mus musculus*) proteome databases (exported from [UniProtKB/Swiss-Prot](#)).

A.1. Specify the proteome database to search peptide data against.

A.2. Select the dataset.

A.3. Select either the cluster or the (sub)cluster (selected through the gating over the data space – make sure that you check “From selected subclusters”).

A.4. Run the database search.

A.5. View the results and protein accessions per peptide.

A.6. View the entire protein sequence upon selecting the “DB search results.”

A.7. Export the DB search.

1) Specify the proteome database to search peptide data against

2) Select the dataset

3) Select the cluster

3) or select the (sub)cluster

4) Run the DB search

5) View the results and protein accessions per peptide

Prot. Accession	Datab...	Protein Group (Ur)
sp P06239 LCK_HUMAN	sp	P06239
sp A5PLL7 PDES1_HUMAN	sp	A5PLL7
sp Q9UHE8 STEAT1_HUMAN	sp	Q9UHE8
sp Q9UM54 MYO6_HUMAN	sp	Q9UM54
sp Q9NWS9 ZN446_HUMAN	sp	Q9NWS9
sp Q12834 CDC20_HUMAN	sp	Q12834
sp Q6Q0C0 TRAF7_HUMAN	sp	Q6Q0C0
sp O14578 DC111_HUMAN	sp	O14576
sp Q5JSH3 WDR44_HUMAN	sp	Q5JSH3
sp Q8WT0 JARCH_HUMAN	sp	Q8WT0
sp Q16576 RBBP7_HUMAN	sp	Q16576
sp Q8TEQ6 GEMI5_HUMAN	sp	Q8TEQ6
sp Q6N021 TE12_HUMAN	sp	Q6N021
sp Q9UM11 FZR1_HUMAN	sp	Q9UM11
sp Q09028 RBBP4_HUMAN	sp	Q09028
sp P78406 RAE1L_HUMAN	sp	P78406
sp Q9Q71 IMDR12_HUMAN	on	Q9Q71

6) Protein full sequence

MAQFAFESDLHSLLOLDAPIPNAPPARWQRKA
KEAAGPAPSPMRAANRSHSAGRTGRTPGKS
SSKVQTPSKPGGDYRIPHRSAAQMEVASFLS
KENONENSTPTKKEHQKAVALNLNGFDVEEA
KILRLSGKPQNAPEGDNPRRLKVLYSQKATPGSS
RKTCRYIFSLPDRILDAPFIRNDYLNLVWDWSSG
NVLAVALDNSVLYWASSODIOLLLQMECPGEY
ISSWAIVKEGNYLAVGTSSAEVQLW/DVQQKRL
RNMTSHSARVGQLSLWNSVYLSSGRSGHIIHH
DVRVAEHHVATLSGHSEEVGLRV/APDGRHLA
SGGNDNLVNVWPSAPGEFGGW/PLQTFQHO
GAVKAVAWCPWQSNCNLATGGGTSRDRHIRWNV
CSGACLSDAVHSQVCSILWSPHYKELISHGF
AQNLQVIVKYPYTMKAVALKGHTSRVLSLTMSP
DGATVASAADETLRLWRCFELDPARREREK
ASAAKSSLIHQGIR

7) Export

MHCpLogics

Please cite

Mohammad Shahbazy, Sri H. Ramarathinam, Chen Li, Patricia T. Illing, Pouya Faridi, Nathan P. Croft, and Anthony W. Purcell (2023) MHCpLogics: an interactive machine learning-based tool for unsupervised data visualization and cluster analysis of immunopeptidomes. bioRxiv.

Contact

For technical support or any question, please send an email to the following developers:

mohammad.shahbazy@monash.edu

nathan.croft@monash.edu

chen.li@monash.edu