**Manual for t-SNE Visualization Interactive app**

This manual describes the functionality of a MATLAB-based GUI designed for visualizing gene expression data using principal component analysis (PCA). This interactive GUI allows users to perform PCA computations on their dataset, visualize the results in 3D, cluster data using K-means, and explore gene correlations and proximities. Start the PCA app GUI from by clicking PCA button on main IVCCA GUI.

### 1. Data Input and Initialization

Upon starting, the GUI initializes and prompts the user to input their dataset through MATLAB's workspace. The dataset should include gene expression data with genes as variables. The script checks if the input is a square correlation matrix and prepares the data for PCA.

### 2. PCA Computation

The GUI performs PCA on the input data, visualizing the data points in a 3D scatter plot. It preprocesses the data by filling missing values, performing PCA for dimensionality reduction, and then applying PCA. It produces a Scree Plot to visualize the variance with the first 25 principal components and a 3D Scatter plot with the first three principal components. Hover over points in the scatter plot to see gene names.

### 3. Visualization Controls

3D Scatter Plot: Displays the PCA results. Users can interact with the plot, brush points, visualize gene expression, rotate, and zoom to explore data points. Buttons for Analysis:

**Cluster:** Performs K-means clustering on the PCA results and visualizes clusters.

**Clear Clusters:** Removes clustering results from the visualization.

**Select Pathway(s):** Allows users to highlight specific genes or pathways by selecting files with gene lists.

**Clear Highlights**: Removes any highlighted genes from the visualization.

**Visualize Gene:** Searches for and highlights a specific gene entered in the adjacent input field.

**Find proximity genes**: Highlights a gene and its nearest neighbors based on the number specified by the user.

**Connect to STRING:** Opens a connection or visualizes the selected from the table gene in the STRING database for network analysis.

### 4. Data Interaction and Exploration

Gene Selection: Users can select genes directly from the scatter plot or through the gene list to view specific information or highlight them.

Brushing and Linking: Allows users to brush (select) points on the scatter plot to see their gene names in a table, facilitating detailed analysis of selected genes.

**Saving and Exporting Results:**

The GUI offers functionality to save clustering results and highlighted genes for further analysis, either programmatically or through user interaction with the GUI.

### 5. Advanced Features

Customization and Extensions: Users with programming knowledge can modify the GUI to add new features, integrate with other databases, or adapt the visualization to other types of data.

**Additional outputs:** KL Divergence: View the Kullback-Leibler divergence value, indicating the difference between the original data distribution and the PCA-transformed version.