A step-by-step guide to execute MATLAB code is provided below:

1. From the **'01_MATLAB Script'** folder, open the **'Main.m'** file in MATLAB 2020b. Run the code by pressing the **'Run'** button in MATLAB editor or by pressing **'F5'**. The process initiates with **filtering of the sample sheet** obtained from the TCGA program based on the following keywords: **'normal'**, **'tumor'**, **'primary'** and **'metastatic'** to include all types of normal and tumor samples, except **'recurrent tumors'**.
   Note: The data files used as input to the code are given as "Supplementary Dataset" at https://data.mendeley.com/datasets/rm7vm2kf4s/1

2. In the subsequent step, specific case amongst Cases 1, 2, and 3 as (i) "paired", (ii) "unpaired", and (iii) "cancer samples only" is selected. For that, a dialogue box appears with the message **'Enter 1 for Paired, 2 for Unpaired, 3 for Only Tumor'**. For selecting '**Paired'** option, enter **'1',** for selecting **'Unpaired'** option**,** enter **'2'** and for selecting **'Only Tumor'** option, enter **'3'**. After selecting an option**,** press **'OK'** to proceed further.

   a) In case of selecting the **'Paired'** option; a dialogue box appears with the message **'You have selected paired option'**. Press **'OK'** to confirm. Upon confirmation, a new sample sheet is assembled for both normal and cancer samples that include patient/sample IDs along with the RNA-seq gene expression file names mapped against normal and cancer samples.
   b) In case of selecting the **'Unpaired'** option; a dialogue box appears with the message **'You have selected unpaired option'**. Press **'OK'** to confirm. Upon confirmation, two separate sample sheets are formed for the normal and cancer case. Each sample sheet contains patient/sample IDs with their corresponding RNA-seq gene expression file names.
   c) In case of selecting the **'Only Tumor'** option; a dialogue box appears with the message **'You have selected only tumor option'**. Press **'OK'** to confirm. Upon confirmation, the sample sheet is assembled for cancer samples only.

3. In the third step, sample size is selected in terms of **random sample** or **complete dataset**. A dialogue box appears with the message **'Enter 1 for random sample selection and 2 for all samples'**.

   a) For selecting **random samples**, enter **'1'** and then press **'OK'** to proceed further. In case the entered **'random sample number'** falls within the range of total number of available samples, it proceeds to the next step. In case the entered **sample size** exceeds the total number of available samples, another dialogue box pops up with the message **'Your random sample size exceeds the number of normal patients. Reselect a smaller sample'.** Press **'OK'** and it takes you to the previous step for re-entering sample size for random sample selection.
   b) For selecting **all samples** option, enter **'2'** and then press **'OK'**. It takes you to the next step.

4. After filtering samples, the next step is **"extraction of gene expressions"**. The script uses the network's nodes list to extract and align RNA-seq based gene expressions of patients with the respective genes and patient IDs. While the expressions are extracted from within the RNA-seq gene expression files and mapped with the respective genes and patient IDs, a dialogue box with progress bar appears with the message **'extracting gene expressions'**.

5. In the fifth step, **outliers are detected**. For the outlier detection in the dataset, 2 options are provided: median absolute deviation (MAD) and inter quartile range (IQR). A dialogue box appears with the message **'Enter 1 for median absolute deviation (MAD) or 2 for inter quartile range (IQR) to detect outliers'**. For selecting MAD option, enter **'1'**, and for selecting IQR, enter **'2'**. After entering 1 or 2, press **'OK'**. Next, the confirmation prompt appears. To confirm, press **'OK'**.

6. In the sixth step, Copy Number Variations (CNVs), Somatic Mutations (SMs) and Genomic Structural Variations (SVs) are processed depending upon the availability of the data. A dialogue box with progress bar appears on the screen.

    a) In case, the CNVs, SMs and SVs are available, the progress bar appears with the message **'loading CNV data', 'loading SM data'** or **'loading SV data'** one after another.
    b) In case, any of the following CNVs, SMs and SVs is missing, the progress bar appears with the respective message e.g. '**CNV data is missing'**, '**SM data is missing'** or **'SV data is missing'**.

Note: Processing includes saving deep deletions (-2) and amplifications (+2), and encodes remaining CNV data as 0. Next, the filtered SMs (for selected patients and network genes) are transformed into logical arrays, where 0 and 1 represent the absence and presence of mutation, respectively. Similarly, patient-specific SVs are also transformed into logical arrays representing the absence (0) and presence (1) of genomic SVs. Lastly, the detected outliers are super-imposed with the CNVs, SMs, and SVs to retain the highly altered RNA-seq gene expressions resulting from CNVs, SMs, and SVs (from within the detected outliers). The remaining outlier gene expressions are removed after which the normal and cancer samples are combined.

7. In the last step, **normalization** is executed. The combined samples are normalized between 0 and 1 using the highest gene expression across patients. Go to the **'Results'** folder to check the output files.