In this analysis, we utilized the Seurat package in R to process and analyze single-cell RNA sequencing data from a 20k Mixture of NSCLC DTCs dataset (3' v3.1) provided by 10x Genomics. The workflow commenced with loading the dataset from an HDF5 file using Read10X\_h5, creating a Seurat object with raw counts data while filtering out cells with fewer than 3 detected genes and fewer than 200 features. Quality control (QC) was performed by assessing mitochondrial gene content and visualizing distributions through violin plots and scatter plots. Subsequently, data normalization was applied using Seurat’s default method, followed by the identification of highly variable features to retain 2000 significant genes. The data was then scaled, and principal component analysis (PCA) was executed to perform linear dimensionality reduction. To determine the optimal number of dimensions for downstream analysis, an elbow plot was used. Clustering was performed with varying resolutions to identify distinct cell populations, and the clustering results were visualized using UMAP for non-linear dimensionality reduction. The final clustering results were visualized with a UMAP plot, allowing for the identification and analysis of distinct cell clusters within the dataset.