|  |  |
| --- | --- |
| Import\_alldata() | |
| **Arguments** | **Description** |
| datafiles | Array with sample names to be imported. Expected format of the input matrix is features x cells. |
| treatment | Array with type of treatment of each sample, e.g., control, treated. Note: To be introduced in the same order as sample names. |
| Return Value | Returns a Seurat object from a feature expression matrix. |

|  |  |
| --- | --- |
| **Appendix: Function definition and parameter testing** Description of the implemented functions for the scRNA-seq analysis. Each function has a table with the different adjustable parameters and some possible options and their impact on the output. In bold it is marked the input used for the analysis.  Import\_alldata() | |
| **Arguments** | **Description** |
| datafiles | Array with sample names to be imported. Expected format of the input matrix is features x cells. |
| treatment | Array with type of treatment of each sample, e.g., control, treated. Note: To be introduced in the same order as sample names. |
| **Return Value** | Returns a Seurat object from a feature expression matrix. |

|  |  |
| --- | --- |
| Calculate\_QC () | |
| **Arguments** | **Description** |
| alldata | Seurat object |
| **Return Value** | Returns a Seurat object with the proportion of the feature set stored in metadata. |

|  |  |
| --- | --- |
| Plot\_QC() | |
| **Arguments** | **Description** |
| data | Seurat object |
| **Return Value** | Return 5 QC plots. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Filter\_data() | |  | | |
| **Arguments** | **Description** | **Input value** | | |
| alldata | Seurat object. | **alldata** | alldata | alldata |
| min\_Feature | Filter cells with a minimum number of genes detected. | **20** | 1000 | 5 |
| max\_Feature | Filter cells with a maximum number of genes detected. | **7000** | 4000 | 8000 |
| min\_cells | Filter genes expressed in at least a minimum of cells. | **20** | 200 | 5 |
| min\_mito | Filter cells with a minimum percentage of mitochondrial genes. | **0** | 5 | 0 |
| max\_mito | Filter cells with a maximum percentage of mitochondrial genes. | **40** | 20 | 60 |
| min\_ribo | Filter cells with a minimum percentage of ribosomal genes. | **0.05** | 5 | 0 |
| max\_ribo | Filter cells with a maximum percentage of ribosomal genes. | **50** | 20 | 100 |
| **Return Value** | Returns Seurat object with the counts filtered. | **1727 features and 52 samples removed 8 resulting clusters** | 4268 features and 644 samples removed  3 resulting clusters | 875 features and 9 samples removed  7 resulting clusters |

|  |  |
| --- | --- |
| SCTransformation() | |
| **Arguments** | **Description** |
| data | Seurat object. |
| **Return Value** | Returns Seurat object with normalized data |

|  |  |
| --- | --- |
| PlotPCA() | |
| **Arguments** | **Description** |
| pbmc | Seurat object. |
| **Return Value** | Returns 6 PCA plots. |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ClusterCells() | |  |  |  |  |  |  |  |  |
| **Arguments** | **Description** | **Input test** | | | | | | | |
| pbmc | Seurat object. | data.filt | data.filt | data.filt | **data.filt** | data.filt | data.filt | data.filt | Data.filt |
| PCAdim | Dimensionality of the data set. Number of relevant PCs. | 11 | 6 | 15 | **12** | 10 | 12 | 12 | 12 |
| clusterResol | Cluster resolution. | 0.8 | 0.8 | 0.8 | **0.8** | 0.8 | 0.4 | 1.2 | 0.6 |
| **Return Value** | Returns Seurat object with calculated clusters. | 8 clusters | 7 clusters | 7 clusters | **7 clusters** | 8 clusters | 5 clusters | 9 clusters | 7 clusters |

|  |  |
| --- | --- |
| Plotnonlinear() | |
| **Arguments** | **Description** |
| pbmc | Seurat object. |
| **Return Value** | Returns 4 plots with the results of the non-linear dimensional reduction. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| FindBioMarkers() | |  |  |  |  |  |
| **Arguments** | **Description** | **Input value** | | | | |
| data | Seurat object. | data.filt | data.filt | data.filt | data.filt | **Data.filt** |
| topN | Number of top markers to calculate per cluster. | 10 | 10 | 1 | 1 | **30** |
| test | Type of test, e.g., wilcox, roc... | wilcox | roc | wilcox | roc | **wilcox** |
| **Return Value** | Returns a dataframe with top genes and their result values. | When comparing both tests, wilcox selects genes with less percentage of expressing but with higher difference of expression in comparison to the rest of the clusters. Roc returns just 1 marker for cluster 1 and 2. | | Top gene per cluster: CDCA8, RC3H2, DHX40, TMEM45A, KRT81, S100A7, ASPM | Top genes per cluster: CDCA8, SCGB1D2, PRKACB, TMEM45A, KRT81, GDF15, ASPM | **Select the 30 top markers for each clusters with p-values below 0.05** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PlotMarkers() | |  |  |  |
| **Arguments** | **Description** | **Input value** |  |  |
| data | Seurat object. | **data.filt** | Data.filt | Data.filt |
| markers | Dataframe with top makers. | **Top gene per cluster: CDCA8, RC3H2, DHX40, TMEM45A, KRT81, S100A7, ASPM** | Genes with higher fold change: S100A7, GDF15, Lower p-value: TPX2, CDCA8 | 12 genes |
| featureVln | List of features to include in the violin plot. | **Top gene per cluster: CDCA8, RC3H2, DHX40, TMEM45A, KRT81, S100A7, ASPM** | Genes with higher fold change: S100A7, GDF15, Lower p-value: TPX2, CDCA8 | 12 genes |
| featureFeaturePlot | List of features to include in the feature plot. | **Top gene per cluster: CDCA8, RC3H2, DHX40, TMEM45A, KRT81, S100A7, ASPM** | Genes with higher fold change: S100A7, GDF15, Lower p-value: TPX2, CDCA8 | 12 genes |
| topN | Number of top expressed features for the Heatmap. | **10** | 30 | 5 |
| featureDot | List of features to include in the Dot plot. | **Top gene per cluster: CDCA8, RC3H2, DHX40, TMEM45A, KRT81, S100A7, ASPM** | Genes with higher fold change: S100A7, GDF15, Lower p-value: TPX2, CDCA8 | 12 genes |
| **Return Value** | Return 4 plots with markers information. |  | With higher number of genes in the Heatmap, its more difficult to see differences between clusters. | With a higher number of genes the plots are difficult to read |

Figure 1. Output 1 for PlotMarkers()

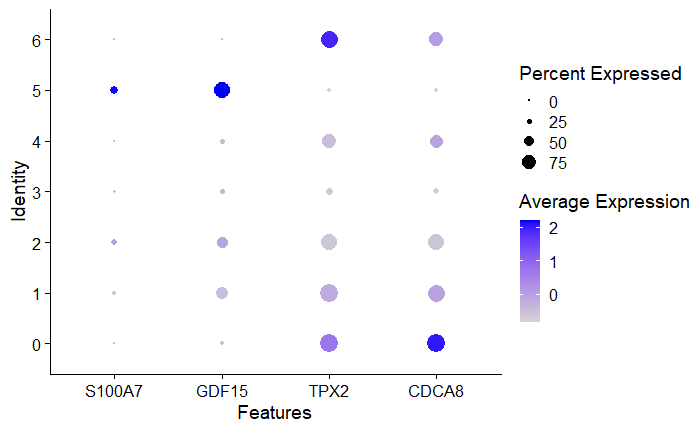
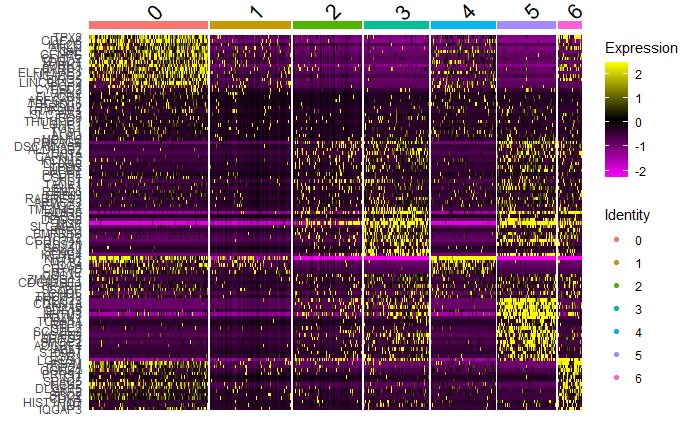
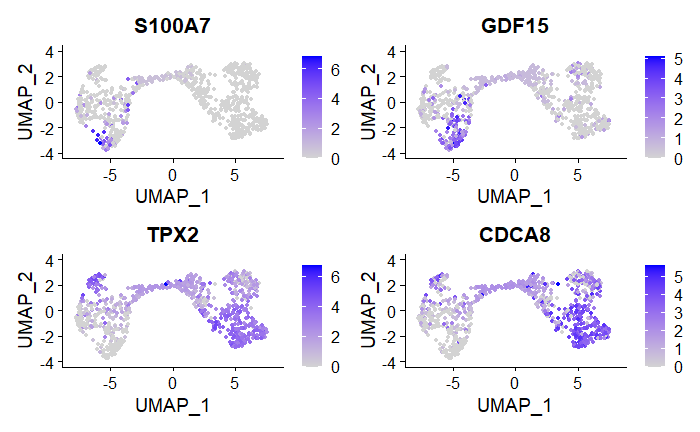
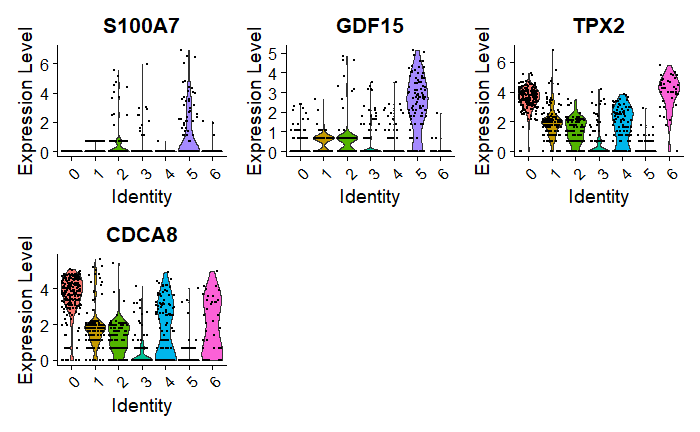
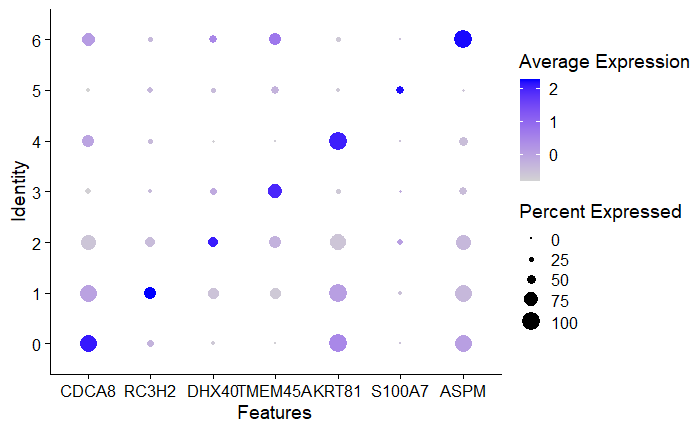
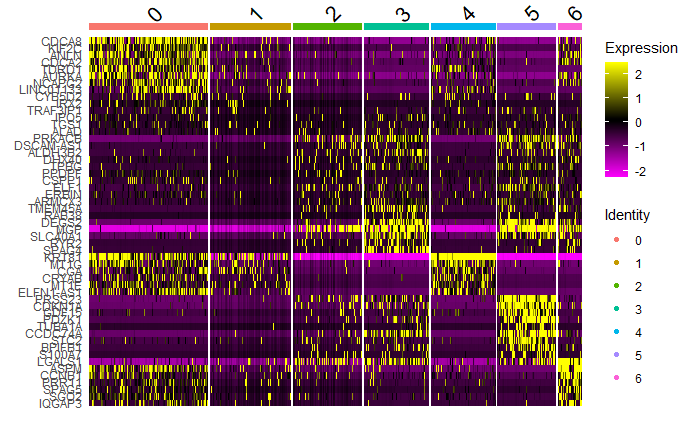
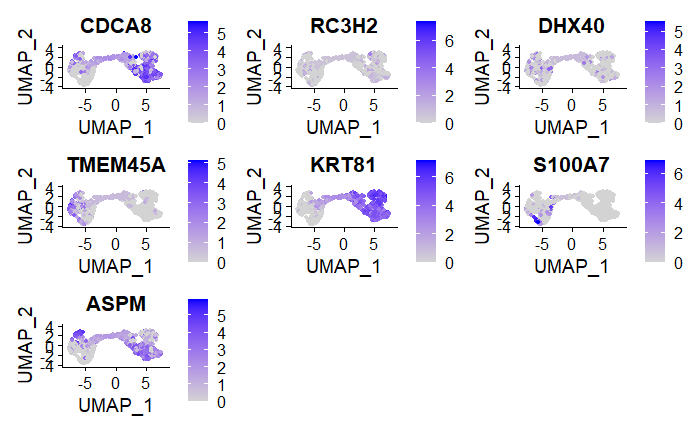
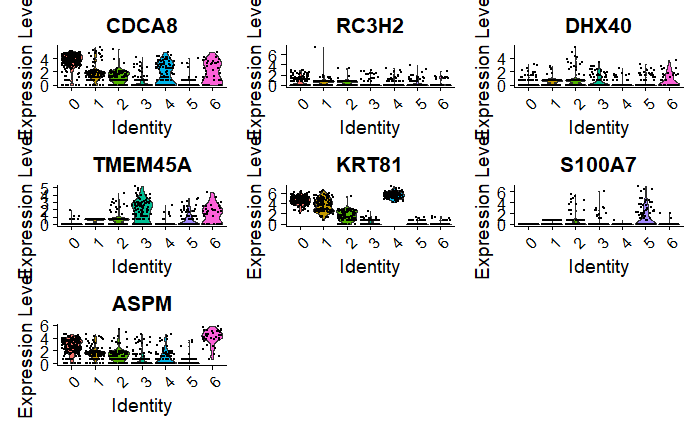


Figure 2. Output 2 for Plotmarkers()

|  |  |
| --- | --- |
| Obtain\_pathway() | |
| **Arguments** | **Description** |
| data | Seurat object. |
| **Return Value** | Returns the maximum differently expressed pathways. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Plot\_Pathway() | |  |  |  |
| **Arguments** | **Description** | **Input value** |  |  |
| data | Seurat object | **data.filt** | data.filt | data.filt |
| max\_diff | Dataset with maximum differently expressed pathways. | **max\_difference** | max\_difference | max\_difference |
| num\_path | Number of pathways to plot in the Heatmap. | **15** | 30 | 50 |
| **Return Value** | Returns 3 plots with the pathways results. | **It is possible to see a DE between clsuters** |  | Less DE between clusters |

|  |  |  |  |
| --- | --- | --- | --- |
| GOEnrichment() | |  |  |
| **Arguments** | **Description** | **Input value** |  |
| markers | Markers for each cluster | **data.filt** | data.filt |
| sign | Significance threshold | **0.3** | 1 |
| **Return Value** | Enrichment matrix for the GOterms and clusters | **More enriched GO term.** | Less enriched GOterms |
| performHierarchicalClustering() | | | |
| **Arguments** | **Description** | **Input value** |  |
| distance\_method | method to measure distance between GO terms clusters. | **correlation** | Eucledian |
| clustering\_method | agglomeration method to perform the hierarchical clustering | **average** | complete |
| **Return Value** | Hierarchical clustering tree | **Smaller hierarchical tree relating the 6 output GOterms clusters.** | Big hierarchical clustering tree, more resulting cluster |
| groupClusters() | |  |  |
| **Arguments** | **Description** | **Input value** |  |
| distance\_threshold | Group GOterms into clusters | **0.05** | 0.5 |
| **Return Value** |  | **7 GOterms clusters. More clusters but fewer GO terms.** | 5 GOterms clusters. Fewer clusters, but with more GO terms |