



Kim GSE73121
(with DeepImpute)
Analysis Report

This report is generated by GranatumX

Please cite: Zhu, Xun et al. "Granatum: A Graphical Single-Cell RNA-Seq Analysis Pipeline for Genomics Scientists." *Genome Medicine* 9.1 (2017)

Upload Files

Assay to upload: **assay.csv (17.07 MB)**

Sample metadata file to upload: **sample_meta.csv (2.82 KB)**

File format: **"csv"**

The assay has **18283** genes and **91** samples.

The first few rows and columns:

```
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
0.0, 0.0, 0.0, 5.853313427568552, 0.0, 0.0, 2.320501479664279, 0.0, 0.0, 0.0
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 3.83695943788012, 0.0, 0.0, 0.0
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
0.0, 0.0, 0.0, 3.9854805960385185, 0.0, 6.093811525239639, 1.7368977847225782, 0.0, 4.862837702635269,
3.7060242246373396, 5.1045206838043145, 0.0, 3.3791174307819216, 0.0, 0.0, 1.8114399614997911, 0.0
```

- Sample meta with name **sample_id** is accepted (PDX_pRCC_SC_29_1887263, PDX_mRCC_SC_01_1887216, PDX_pRCC_SC_28_1887262, PDX_pRCC_SC_94_1887301, PDX_mRCC_SC_79_1887240, ... and 86 more entries).
- Sample meta with name **cell_type** is accepted (PDX_pRCC, PDX_mRCC, PDX_pRCC, PDX_pRCC, PDX_mRCC, ... and 86 more entries).

DeepImpute

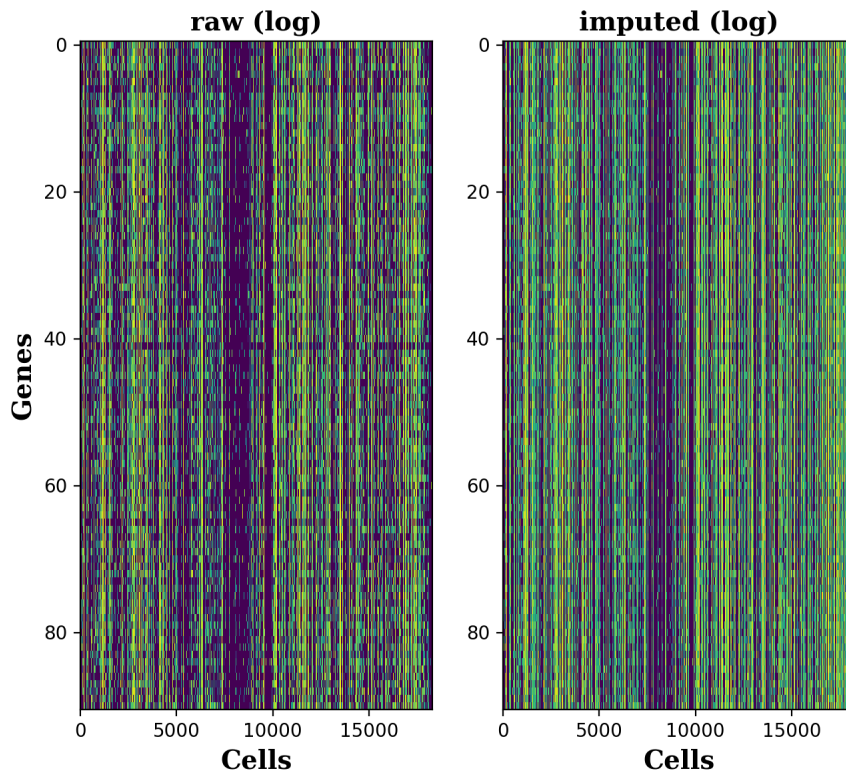
Random seed: **12345**

Use automatic gene imputation limit: **true**

Gene rank limit: **2000**

Cell subset: **1**

Assay: **[A]assay.csv** (from step 1: Upload Files)



Heatmaps

#Imputed genes: **9062**

Dropout (before/after): 0.53/0.38

Accuracy (R2 score) on masked data: **0.37**

Gene Filtering

The gene has to be expressed in at least ___ cells: **3**

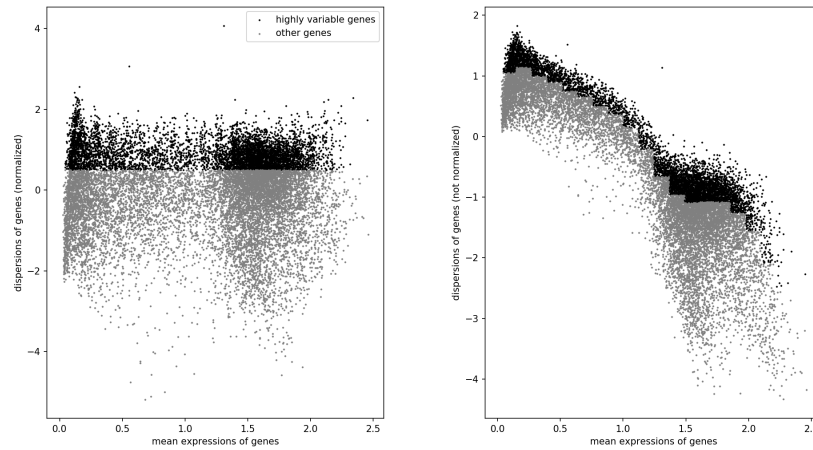
The average expression level of the gene has to be greater than: **1**

The average expression level of the gene has to be less than: **999999**

The dispersion of the gene has to be greater than: **0.5**

The dispersion of the gene has to be less than: **999999**

Assay: **Imputed assay** (from step 2: DeepImpute)



Each dot represent a gene. The gray dots are the removed genes. The x-axis is log-transformed.

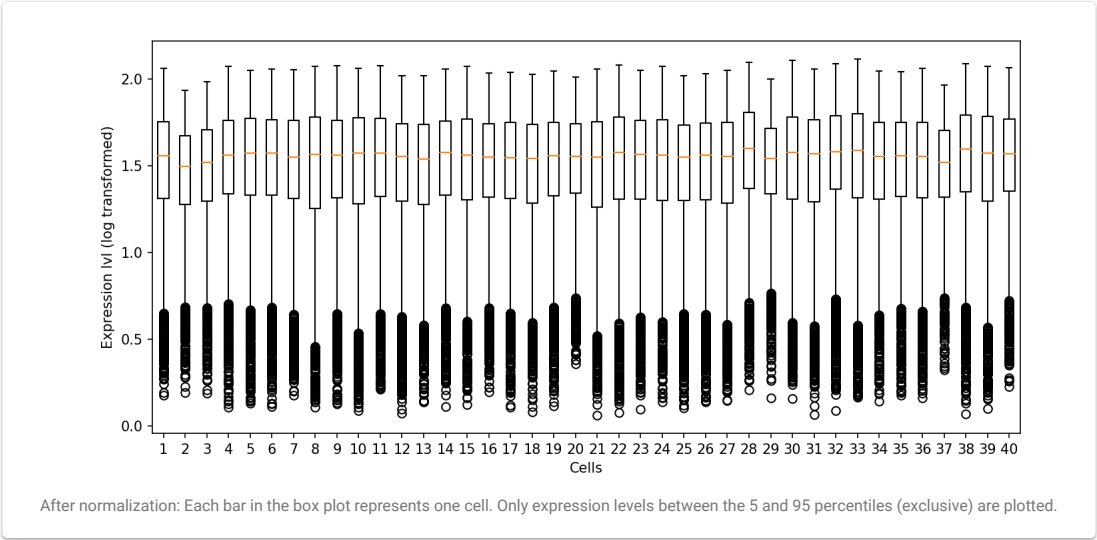
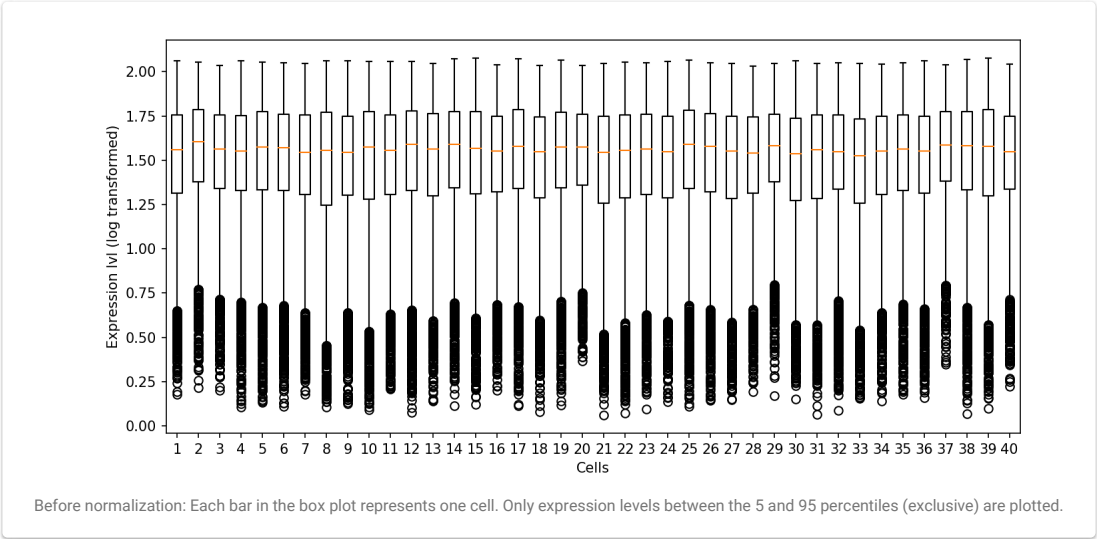
Number of genes before filtering: **18283**

Number of genes after filtering: **5656**

Cell Normalization

Number of cells to plot in the bar-plot: **40**

Assay: **Filtered Assay** (from step 3: Gene Filtering)

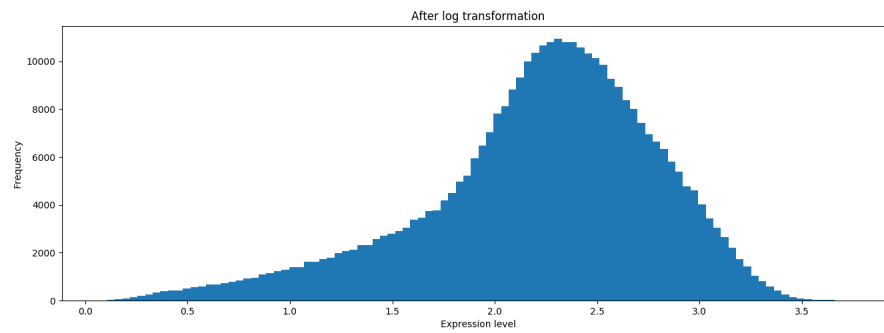
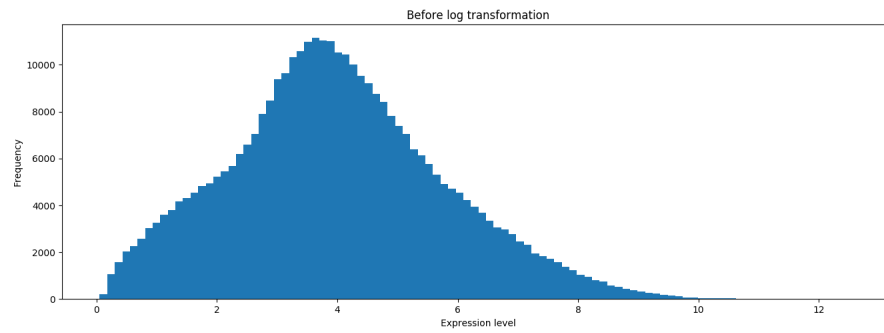


Log transformation

The base used for the log function: **2**

The pseudo counts added before log transformation (to avoid getting $\log(0)$): **1**

Assay including matrix and genets: **Normalized assay** (from step 4: Cell Normalization)

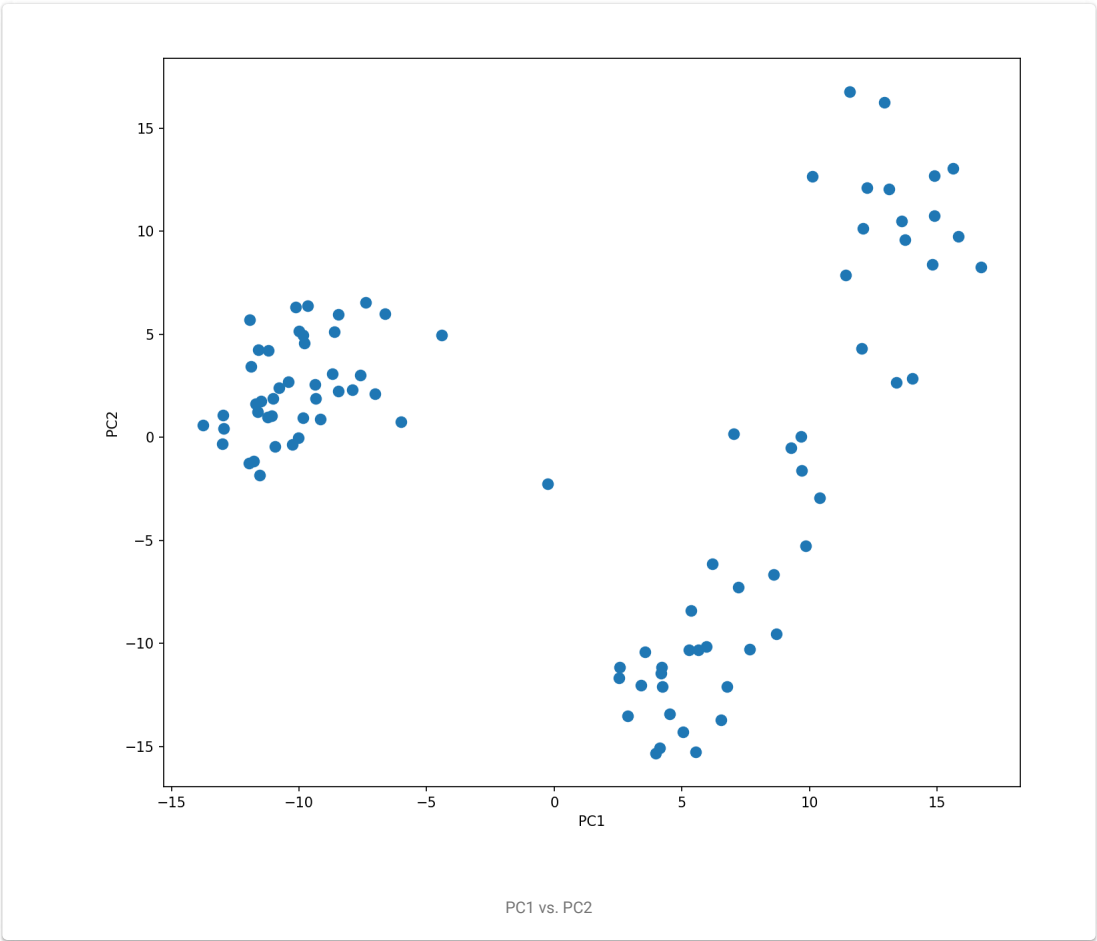
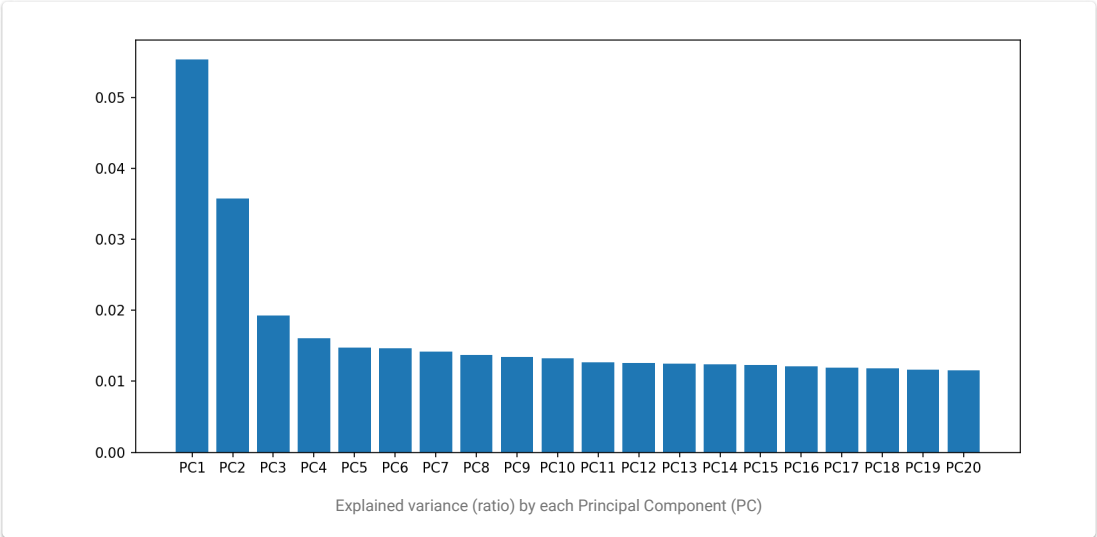


The distribution of expression level before and after log transformation. Only the values greater than the 5 percentile (usually zero in single-cell data) and lower than 95 percentile are considered.

Principal Component Analysis

Number of top components to calculate: 2

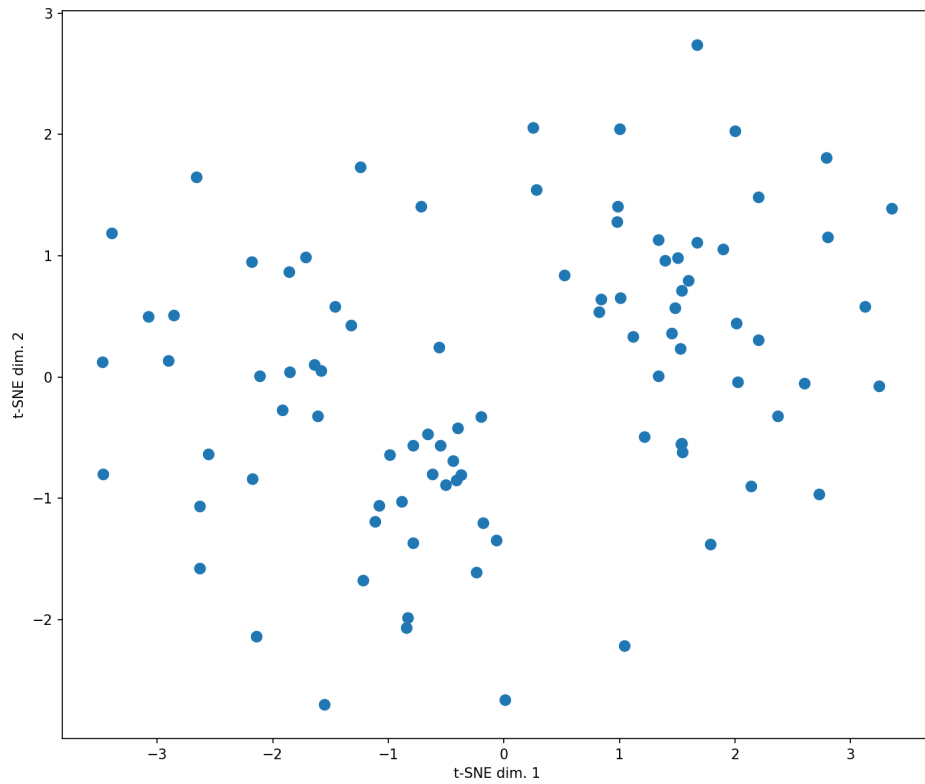
Assay: **Log transformed assay** (from step 5: Log transformation)



t-Distributed Stochastic Neighbor Embedding

Random seed: 0

Assay: **Log transformed assay** (from step 5: Log transformation)



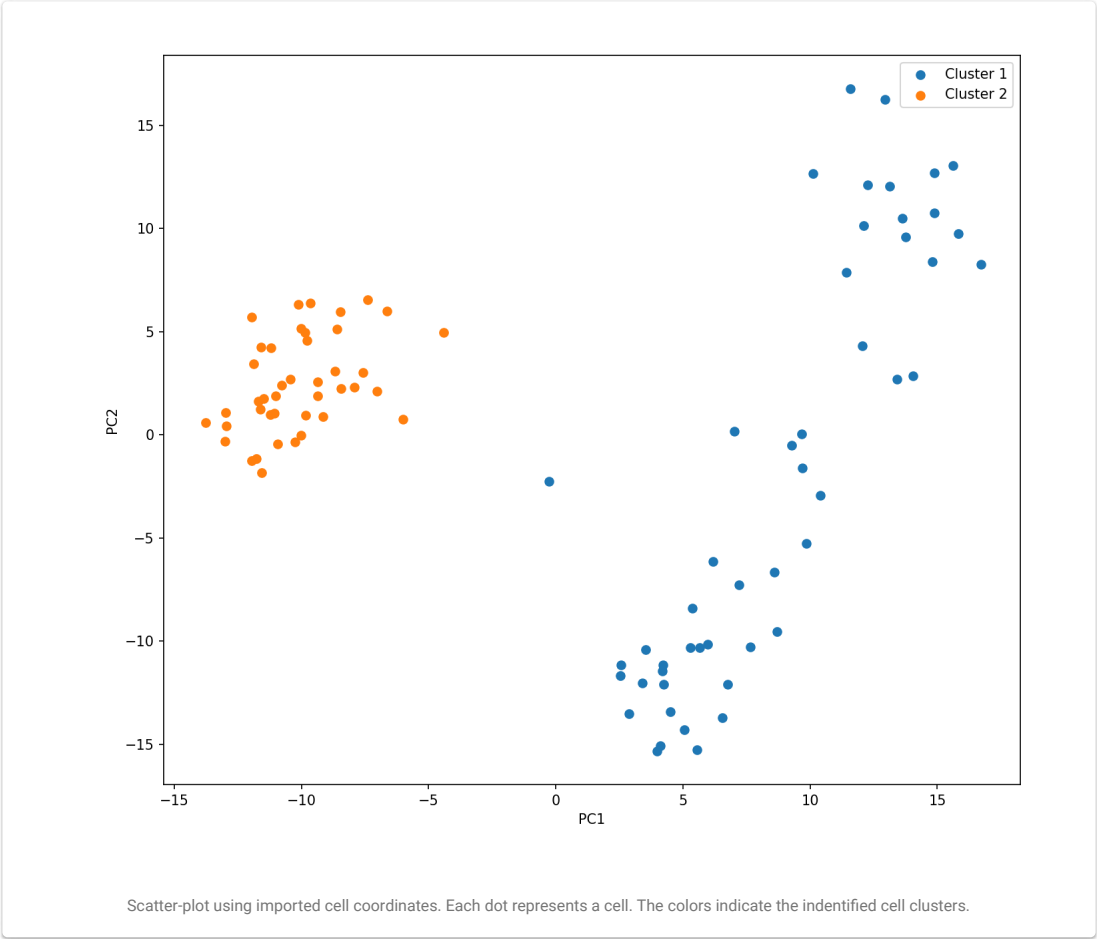
t-SNE plot: each dot represents a cell

Clustering

Random seed: 0

Cell coordinates for visualization: **PC1 vs. PC2** (from step 6: Principal Component Analysis)

Assay including matrix and geneds: **Log transformed assay** (from step 5: Log transformation)



Clustering Quality Measure

Predicted clusters: **Cluster assignment** (from step 8: Clustering)

True clusters: **[M]cell_type** (from step 1: Upload Files)

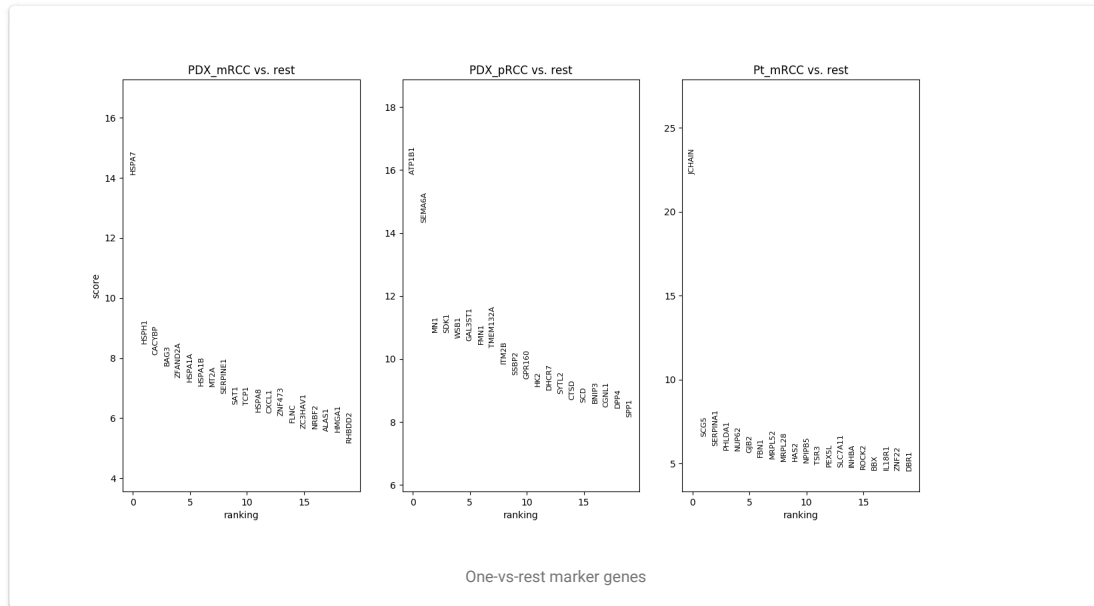
Adjusted Rand score: **0.6747317963935174**

Adjusted mutual information score: **0.5994358709303271**

Marker Genes Identification

Assay including matrix and genelds: **Log transformed assay** (from step 5: Log transformation)

Group vector: **[M]cell_type** (from step 1: Upload Files)

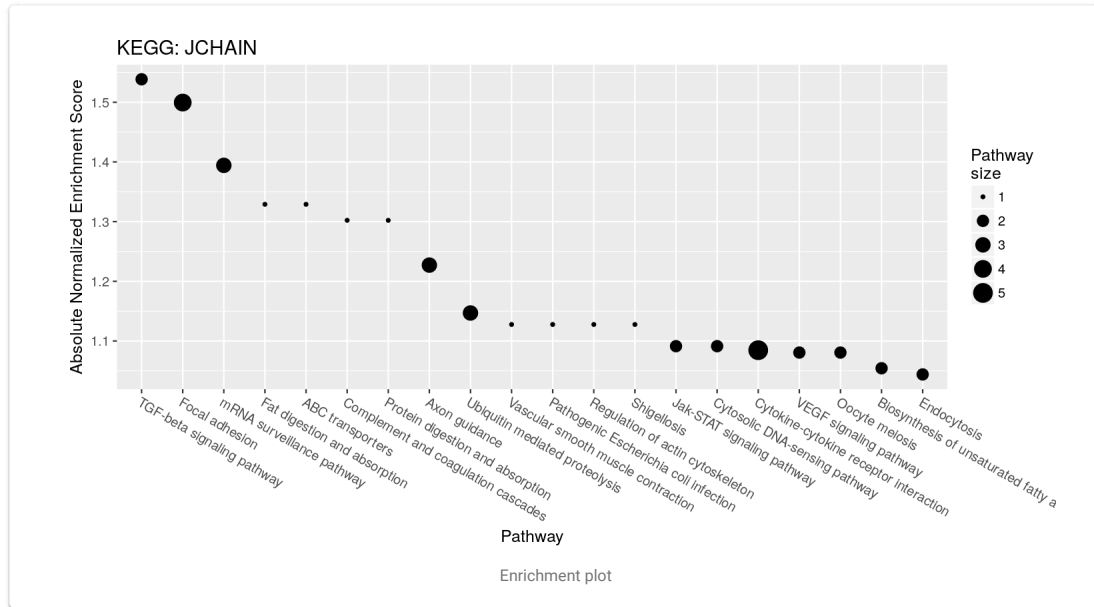


Fast GSEA (R)

The species: **"human"**

The database for the enrichment analysis: **"KEGG"**

A list of genes with their scores: **Marker score (Pt_mRCC vs. rest)** (from step 10: Marker Genes Identification)



Rank	Identifier	Score
1	TGF-beta signaling pathway	1.5385
2	Focal adhesion	1.4995
3	mRNA surveillance pathway	1.3943
4	Fat digestion and absorption	1.329
5	ABC transporters	1.329
6	Complement and coagulation cascades	1.3022
7	Protein digestion and absorption	1.3021
8	Axon guidance	1.2271
9	Ubiquitin mediated proteolysis	1.1469
10	Vascular smooth muscle contraction	1.1277

PREVIOUS PAGE 1 / 9 NEXT PAGE

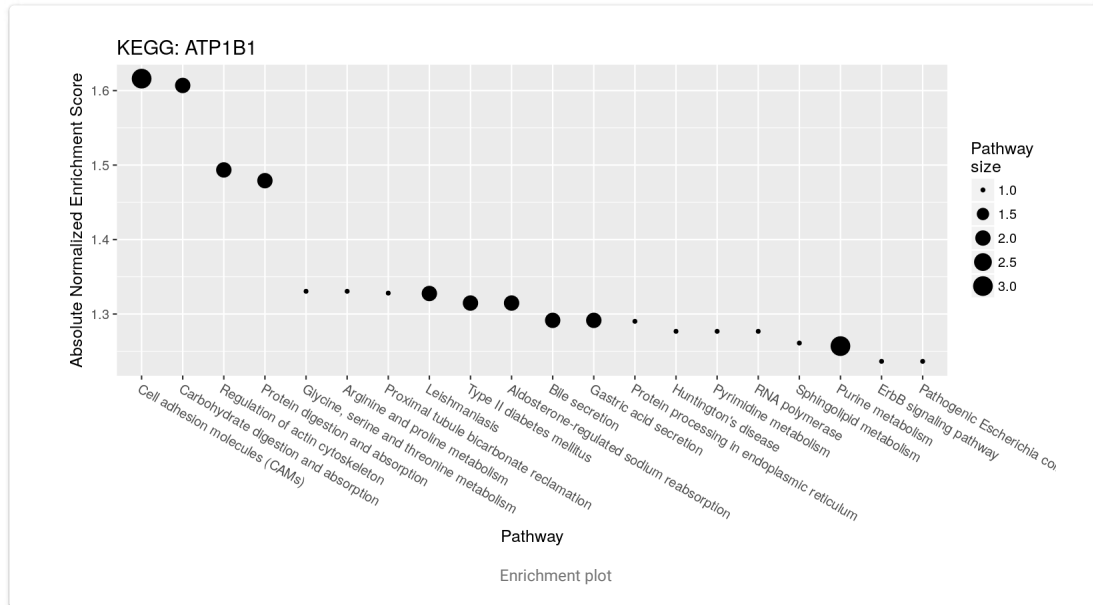
Enrichment table

Fast GSEA (R)

The species: **"human"**

The database for the enrichment analysis: **"KEGG"**

A list of genes with their scores: **Marker score (PDX_pRCC vs. rest)** (from step 10: Marker Genes Identification)



Rank	Identifier	Score
1	Cell adhesion molecules (CAMs)	1.616
2	Carbohydrate digestion and absorption	1.6069
3	Regulation of actin cytoskeleton	1.4935
4	Protein digestion and absorption	1.4791
5	Glycine, serine and threonine metabolism	1.3305
6	Arginine and proline metabolism	1.3305
7	Proximal tubule bicarbonate reclamation	1.3281
8	Leishmaniasis	1.3275
9	Type II diabetes mellitus	1.3147
10	Aldosterone-regulated sodium reabsorption	1.3147

PREVIOUS PAGE 1 / 8 NEXT PAGE

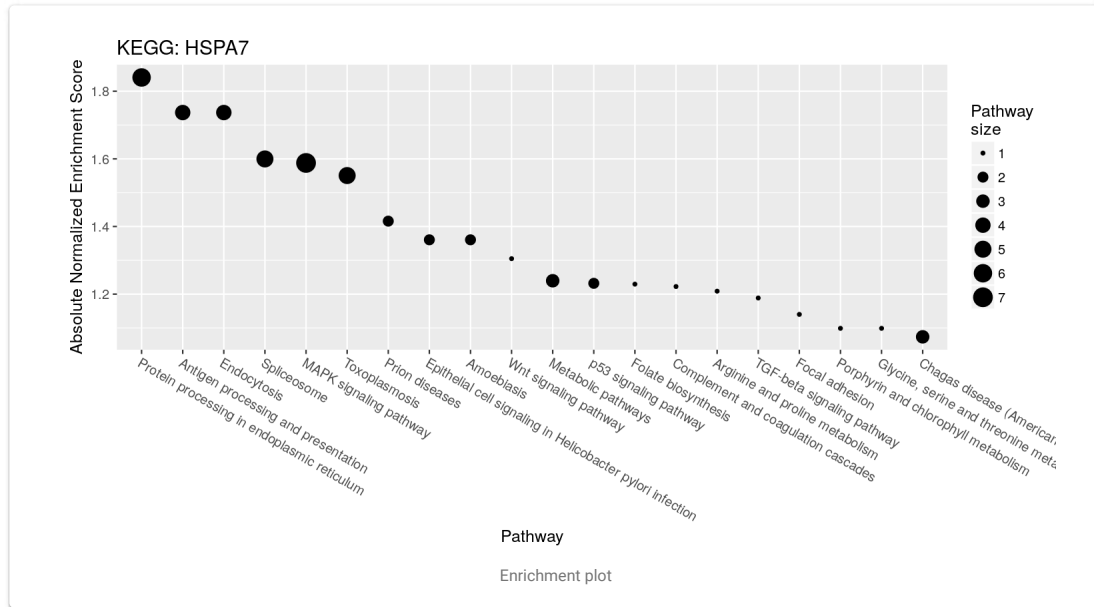
Enrichment table

Fast GSEA (R)

The species: **"human"**

The database for the enrichment analysis: **"KEGG"**

A list of genes with their scores: **Marker score (PDX_mRCC vs. rest)** (from step 10: Marker Genes Identification)



Rank	Identifier	Score
1	Protein processing in endoplasmic reticulum	1.8406
2	Antigen processing and presentation	1.7369
3	Endocytosis	1.7369
4	Spliceosome	1.5998
5	MAPK signaling pathway	1.5879
6	Toxoplasmosis	1.5506
7	Prion diseases	1.4158
8	Epithelial cell signaling in Helicobacter pylori infection	1.3607
9	Amoebiasis	1.3607
10	Wnt signaling pathway	1.3049

PREVIOUS PAGE 1 / 5 NEXT PAGE

Enrichment table

Pseudo-time construction

Random seed: 0

The sample metadata for cell coloring in plot: **[M]cell_type** (from step 1: Upload Files)

The assay to use for pseudo-time cell ordering: **Log transformed assay** (from step 5: Log transformation)

