

tSNE-cosmic-gene-census

February 1, 2018

1 t-SNE of COSMIC Genes in TCGA and GTEx

Dimensionality reductions of selected tissues from The Cancer Genome Atlas (TCGA) and The Genotype Tissue Expression Consortium (GTEx) using a gene subset from The Cancer Gene Census, which is collated by the Catalogue of Somatic Mutations in Cancer (COSMIC).

```
In [1]: from __future__ import division
import rnaseq_lib as r

import numpy as np
import pandas as pd
import holoviews as hv
hv.extension('bokeh', logo=False)
```

<IPython.core.display.HTML object>

```
In [2]: ## Synapse ID: syn11515015
df_path = '/mnt/rnaseq-cancer/Objects/tcga-gtex-metadata-expression.tsv'
df = pd.read_csv(df_path, sep='\t', index_col=0, dtype=r.tissues.dtype)
```

```
In [3]: # Plotting wrapper for dataframe
h = r.plot.Holoview(df)
# Load cosmic data and get genes
cosmic_path = '/mnt/rnaseq-cancer/Metadata/cosmic_all_1-26-2018.tsv'
cosmic = pd.read_csv(cosmic_path, sep='\t', index_col=0)
cosmic_genes = [x for x in cosmic.index.tolist() if x in df.columns]
```

```
In [4]: # Run tsne and create plot object
title = ' of TCGA and GTEx Subset by {} Cosmic Genes'.format(len(cosmic_genes))
tsne = h.tsne(genes=cosmic_genes)
```

```
In [6]: %%opts Overlay [tabs=True] Scatter [width=800 height=700]
color_indices = ['tissue', 'labels', 'type']
hv.Overlay([tsne.opts(dict(Scatter=dict(plot=dict(color_index=x))))).relabel('{}').format(
```

```
Out[6]: :Overlay
      .Scatter.Tissue :Scatter    [x]    (y,id,tissue,dataset,tumor,type,labels)
      .Scatter.Labels :Scatter    [x]    (y,id,tissue,dataset,tumor,type,labels)
      .Scatter.Type   :Scatter    [x]    (y,id,tissue,dataset,tumor,type,labels)
```