Pipeline_simple

March 13, 2018

In [2]:
```
%matplotlib inline
%load_ext autoreload
%autoreload 2
import pipeline1 as pi
import matplotlib.pyplot as plt
```

```
The autoreload extension is already loaded. To reload it, use:
  %reload_ext autoreload
```

In [3]:
```
Genes = pi.Genes()
gene = Genes.genevect['YAL019W'] #we get the one gene dataset we have in ou
Genes.preprocess(gene)
centroids, labels = Genes.clusterize(gene)
```

```
YAL019W
```

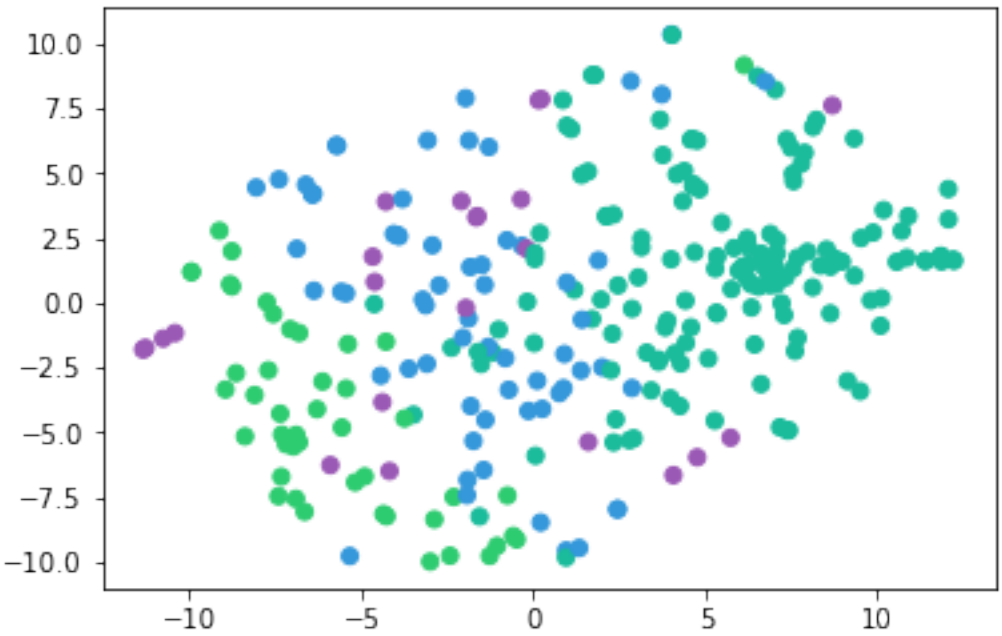
```
---------------------------------------------------------------------------

NameError                                 Traceback (most recent call last)

<ipython-input-3-879d69a16f72> in <module>()
----> 1 Genes = pi.Genes()
      2 gene = Genes.genevect['YAL019W'] #we get the one gene dataset we have i
      3 Genes.preprocess(gene)
      4 centroids, labels = Genes.clusterize(gene)


/Users/jeremie/Documents/Projects/Python/MasterProj/pipeline1.py in __init_
     42         for gene in self.genelist:
     43             try:
---> 44                 gentab,genreg, species = self.readcods_gene(folder, gen
     45                 if len(species) >  minspecies:
     46                     self.genereg.update({gene:genreg})
```

```
/Users/jeremie/Documents/Projects/Python/MasterProj/pipeline1.py in readcoc
 94             if amino != 'ror':
 95                 struct = pd.read_csv(file).dropna().reset_index()
---> 96                 gentab, species = check_nans(struct, species, gentab)
 97                 gentab[i,:] = struct['entropyLocation'].values
 98                 i+=1


NameError: global name 'check_nans' is not defined
```

0.1 T-sne reveals features :

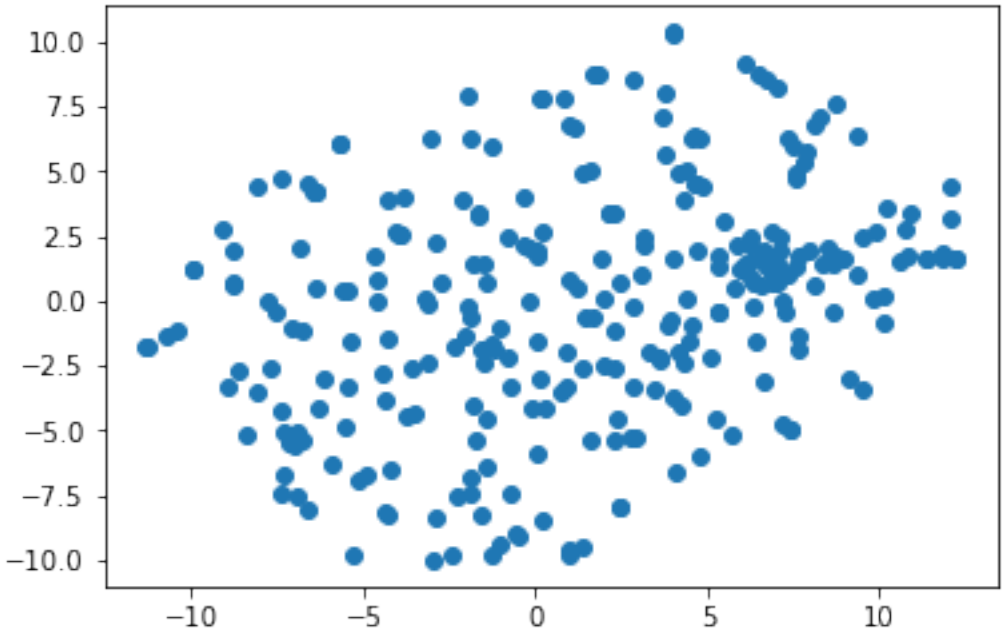
We can see that whatever is the input of tsne it seems that we have a blob of related gene through their bias.

```
In [10]: colormap = ["#1abc9c","#3498db","#2ecc71","#9b59b6",'#34495e','#f1c40f','#

         tsne = Genes.reduce_dim(gene, 2)
         colors= [colormap[x] for x in labels]

         fig = plt.figure()
         ax = fig.add_subplot(111)
         ax.scatter(tsne[:,0],tsne[:,1], c = colors )
         plt.show()
```



```
In [11]: plt.scatter(tsne[:,0],tsne[:,1])
```

```
Out[11]: <matplotlib.collections.PathCollection at 0x111b12650>
```



0.2 Interactive plot :

```
In [6]: from bokeh.plotting import *
        from bokeh.models import HoverTool
        from bokeh.io import output_notebook
        output_notebook()
```

```
In [8]: GA = pi.Genes()
        GA.plot_gene(tsne,Genes.specieslist[0],centroids=centroids,labels=labels)
```

```
Out[8]: Figure(id='e80b7ff9-4850-459d-a4dc-9e6f558291d3', ...)
```

debug and trials

```
In [36]: import numpy as np
         import pandas as pd

         a = ['a','e','b','c']
         b = ['a','c','v','e']
         set(a) | set(b)
```

```
Out[36]: {'a', 'b', 'c', 'e', 'v'}
```

```
In [133]: b = pd.DataFrame()

In [134]: df2 = pd.DataFrame([[5, 6], [7, 8]], columns=list('AB'))
          b = b.append(df2)
          b

Out[134]:    A  B
          0  5  6
          1  7  8

In [138]: a = b.loc[1,:]
          c = pd.DataFrame(a).transpose()

In [139]: a

Out[139]: A    7
          B    8
          Name: 1, dtype: int64

In [140]: c

Out[140]:    A  B
          1  7  8

In [141]: c.rename(index={1: 'Ahaha'})

Out[141]:        A  B
          Ahaha  7  8

In [142]: c

Out[142]:    A  B
          1  7  8

In [144]: b = b.append(c.rename(index={1: 'Ahaha'}))
          b

Out[144]:        A  B
          0      5  6
          1      7  8
          1      7  8
          Ahaha  7  8

In [ ]:
```