

Creation of TF and target gene search engine

The steps:

- 1: Collect the targets of transcription factors with TFBS in -1000 to +100 bp around the TSS.
- 2: Create targets of each transcription factors in a separate folder
- 3: Annotate the transcription factors which are mentioned in wpGSA
- 4: Upload the results of #1 to molecularbrain.org to find either targets or transcription factors for any gene.

Materials and Methods

to collect the targets of transcription factors with TFBS in -1000 to +100 bp around the TSS.

Steps

- 1: download the bulk data from wpGSA (<http://wpgsa.org/download>), about 2.2 GB zipped file
- 2: unzip using 7 zip, it becomes 6.6 GB, too big to add geneID and gene Symbols to the targets which are identified only as EnsemblIDs in wpGSA



FileSplitter.zip

- 3: split the 6.6 GB txt file to 27 or so 250000 kb files using FileSplitter
- 4: upload the files to Microsoft ACCESS. Remove columns not needed, kept TF, EnsemblID, distance from TSS and name
- 5: download Ensembl to gene from NCBI, parsed and selected mus only and upload specific relevant info to ACCESS
- 6: downloaded Mus gene info from NCBI gene
- 6: combine the wpGSA data with the GeneID and symbol from NCBI using Ensembl as the common key

7: the resulting files were then queried to identify the TF binding to target genes in the region of ± 1000 around TSS

8: the outputs were combined using combinecsv.py

9: upload to molecular brain.org as a cmv file, which has TF name, the target geneID and gene symbol and distance from TSS

Combinecsv.py

(Nihar Samal)

```
import csv
```

```
# Open result file
```

```
with open('output.txt','wb') as fout:
```

```
    wout = csv.writer(fout,delimiter=',')
```

```
    interesting_files = glob.glob("*.csv")
```

```
    for filename in interesting_files:
```

```
        print 'Processing',filename
```

```
        # Open and process file
```

```
        h = True
```

```
        with open(filename,'rb') as fin:
```

```
            if h:
```

```
                h = False
```

```
            else:
```

```
                fin.next()#skip header
```

```
            for line in csv.reader(fin,delimiter=',):
```

```
wout.writerow(line)
```