## Analysis of liver immune cellular SEGs

```
mkdir liver
cd ./liver
mkdir raw_SEG
mkdir majorCluster_SEG
mkdir liver_SEGs
```

*# curating the top 1000 genes with largest GESIs within each subcluster of individual liver samples; calculating SEGs of each individual major immune cell cluster within each sample*

*# curating the common SEGs of each major immune cellular cluster among all the liver samples; retrieving the inter-major-cluster common SEG list*

*# identifying the SEGs within each sample among the major immune cellular clusters, while the SEGs should be present in the inter-major-cluster common SEG list.*

# transferring the sub-directories curating top 1000 SEGs of each major immune cell cluster into ./raw_SEG.

# transferring '*uSEGdistMean*', '*clusterSEG.pl*' and '*formatSingleSubclusterSEGs.pl*' into ./raw_SEG and individual sample_cluster sub-directory.

*# calculating SEGs of each individual major immune cell cluster within each sample*

*# for the major clusters with multiple subclusters, using 'uSEGdistMean' and 'clusterSEG.pl'*

*e.g.,*

```
cd ./raw_SEG/Liver-SRA716608-SRS3391633/T_subclusters_1000SEGs
./uSEGdistMean T_subclusters_1000SEGs/ _1000_SEGs.txt >SRS3391633_T_subclusters_1000SEGs_Means.txt
```

*# open 'SRS3391633_T_subclusters_SEGs.txt', see the relationship of the number of SEGs and the number of sub-clusters sharing the SEGs, and determine the minimal cutoff of sub-cluster number (typically ensuring >500 SEGs being retained)*

```
perl clusterSEG.pl SRS3391633_T_subclusters_1000SEGs_Means.txt 2 > SRS3391633_T_subclusters_SEGs.txt
```

*# for the major clusters with single subcluster, using 'formatSingleSubclusterSEGs.pl'*

*e.g.,*

```
cd ./raw_SEG/Liver-SRA716608-SRS3391633/B_subclusters_1000SE
perl formatSingleSubclusterSEGs.pl 9_B_1000_SEGs.txt >SRS3391633_B_subclusters_SEGs.txt
```

```
cd    majorCluster_SEG
mkdir    B
mkdir    T
mkdir    NK
mkdir    M
mkdir    HC    # Hepatocytes

# moving the SEGs of major immune cell clusters of each sample into the same sub-directory of the major immune cell cluster
# moving the 'dist' script into the 'majorCluster_SEG' directory

# calculating the SEGs of each individual major immune cell cluster among all samples
# calculating common SEGs of all the major immune cell clusters among all samples
# for both tasks, using 'dist'
```

*e.g.,*

```
cd    ./majorCluster_SEG
./dist    B    _B_subclusters_SEGs.txt    >B_common_SEGs.txt

mkdir    all
mkdir    all_but_HC

# moving all '*_common_SEGs.txt' files into 'all', and all but 'HC_common_SEGs.txt' files into 'all_but_HC'
./dist    all    _common_SEGs.txt    >all_common_SEGs.txt
./dist    all_but_HC    _common_SEGs.txt    >allButHC_common_SEGs.txt

# save the all_clusters sharing SEGs within 'all_common_SEGs.txt' into 'allClusters_common_SEGs.txt'; save the all_but_HC_clusters sharing SEGs within 'allButHC_common_SEGs.txt' into 'allButHCClusters_common_SEGs.txt'
```

```
cd liver_SEGs
# build a sub-directory for each sample
```

*# identifying the SEGs within each sample among the major immune cellular clusters, while the SEGs should be present in the inter-major-cluster common SEG list*

e.g.,

```
mkdir SRS3391632
mkdir SRS3391633
```

*# moving 'allClusters_common_SEGs.txt' and 'allButHCClusters_common_SEGs.txt' into 'SRS3391632' and 'SRS3391633'; moving 'sampleMajorClusterSEGext.pl' into 'SRS3391632' and 'SRS3391633'; building 'all' and 'all_but_HC' within 'SRS3391632' and 'SRS3391633'; moving the major-cluster SEGs within each sample (including HC) into 'all', and moving the major-cluster SEGs within each sample (not including HC) into 'all_but_HC'.*

In each sub-directory of the samples,

e.g.,

```
perl sampleMajorClusterSEGext.pl all allClusters_common_SEGs.txt >SRS3391632_all_SEGs.txt
perl sampleMajorClusterSEGext.pl all_but_HC allButHCClusters_common_SEGs.txt
    >SRS3391632_allButHC_SEGs.txt
```

*# to find out the TISSUE_SEGs, simply within 'liver_SEGs' directory, put the 'dist' script, and,*

```
mkdir final
cd final
mkdir all
mkdir all_but_HC
```

# put '*_all_SEGs.txt' files in 'all' sub-directory and '_allButHC_SEGs.txt' files in 'all_but_HC' sub-directory

```
./dist all _all_SEGs.txt >liver_all_SEGs.txt
./dist all_but_HC _allButHC_SEGs.txt >liver_allButHC_SEGs.txt
```

*# to find the intersect between 'PBMC_SEGs.txt' and 'liver_all_SEGs.txt' or 'liver_allButHC_SEGs.txt' with 'interSect.pl'*

# put 'interSect.pl' and 'PBMC_SEGs.txt' in 'final' directory

cd final

perl interSect.pl PBMC_SEGs.txt liver_all_SEGs.txt >PBMC_liver_all_intersect_SEGs.txt

perl interSect.pl PBMC_SEGs.txt liver_allButHC_SEGs.txt >PBMC_liver_allButHC_intersect_SEGs.txt