

Contents

- [addpath to all tools](#)
- [initiate one instance of the "cooccurrence_clustering_analysis" class](#)
- [read data prepared in matlab file](#)
- [filter the data by removing genes and cells \(same as Seurat tutorial on this data\)](#)
- [binarize data](#)
- [cooccurrence clustering](#)
- [re-order pathways for better visualization](#)
- [Visualize the percentages of detection for pathways in individual cells](#)
- [export results to files](#)

addpath to all tools

```
addpath(genpath('./tools/'))
```

initiate one instance of the "cooccurrence_clustering_analysis" class

```
cooc = cooccurrence_clustering_analysis;
```

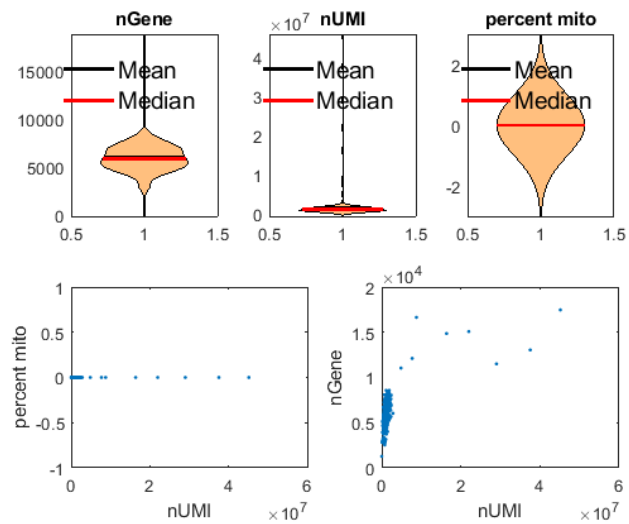
read data prepared in matlab file

```
cooc = cooc.ReadMatlab('Burns.mat');
```

filter the data by removing genes and cells (same as Seurat tutorial on this data)

```
cooc.initial_filtering_min_num_cells = 10;
cooc.initial_filtering_min_num_genes = 0;
cooc.initial_filtering_max_num_genes = Inf;
cooc.initial_filtering_max_percent_mito = 1;
cooc = cooc.initial_filtering_of_data(1);
```

Data after initial filtering 15726 genes * 321 cells.



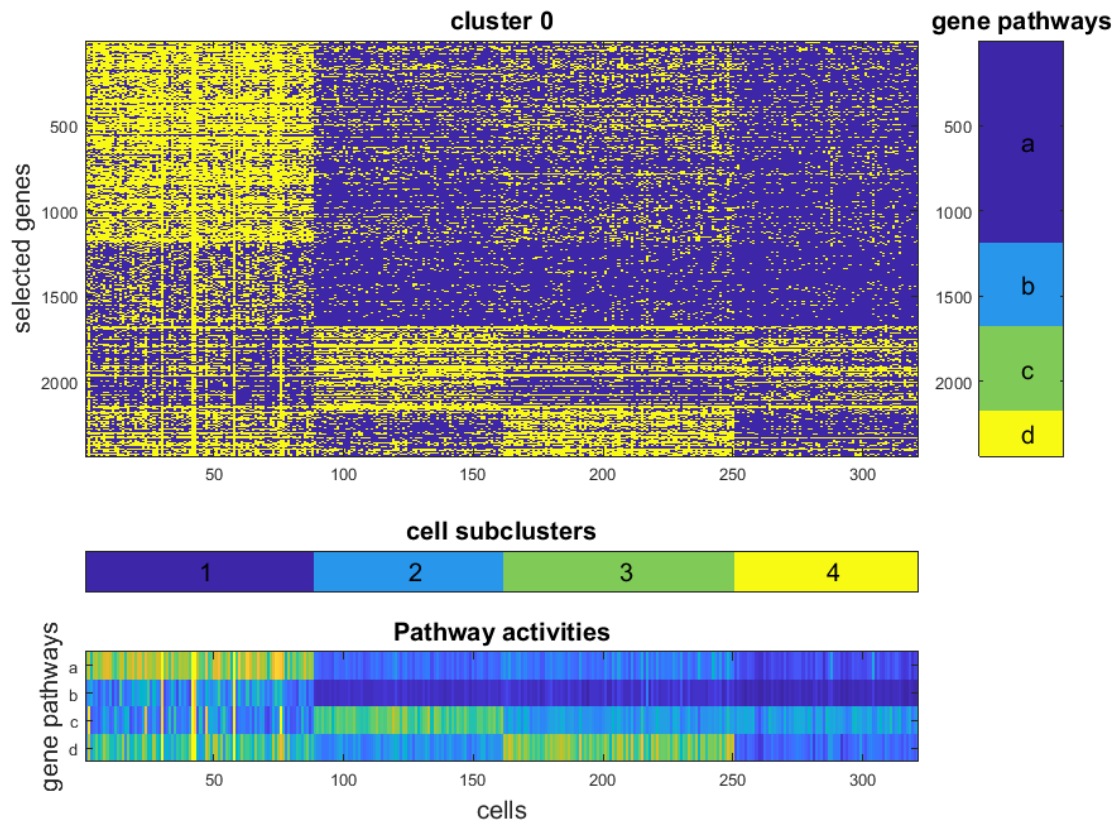
binarize data

```
cooc.binarization_threshold = 0;  
cooc.binary_data = full(double(cooc.data>cooc.binarization_threshold));
```

cooccurrence clustering

```
cooc.cooccurrence_min_expressed_cells = 10;           % genes will only be considered if detected in >= minimum number of cells, and undetected in >= minimum number of cells  
cooc.cooccurrence_min_pathway_size = 20;           % only considered gene clusters of size >= this threshold  
cooc.cooccurrence_min_population_size = 10;        % only considered gene clusters of size >= this threshold  
cooc.cooccurrence_snr_merge_threshold = 1.5;       % threshold for merging Louvain communities, based on snr of average detection of each gene cluster  
cooc.cooccurrence_mean_diff_merge_threshold = 0.5; % threshold for merging Louvain communities  
cooc.cooccurrence_mean_ratio_merge_threshold = 2;  % threshold for merging Louvain communities  
  
cooc = cooc.iterative_cooccurrence_clustering;
```

```
Remaining clusters to partition 1  
Processing cluster 0 now ...  
Processing data subset with 15726 genes and 321 cells:  
Remove genes detected in <10 cells. Remaining 15286 genes. Elapsed time is 0.026132 seconds.  
Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 161.982578 seconds.  
Compute gene-gene similarity ... Elapsed time is 12.750144 seconds.  
Create gene-gene graph for clustering genes ...  
Writing graph into file ... 100%Elapsed time is 2.513890 seconds.  
Running ModularityOptimizer for clustering ...Elapsed time is 4.514023 seconds.  
Gene-gene graph contains 8 pathways, 6006 genes in total  
Elapsed time is 5.155855 seconds.  
Create cell-cell graph for clustering cells ...  
Writing graph into file ... 100%Elapsed time is 0.037040 seconds.  
Running ModularityOptimizer for clustering ...Elapsed time is 0.315499 seconds.  
Cell-cell graph contains 6 cell types by community detection  
Elapsed time is 0.324739 seconds.  
Cell-cell graph contains 5 cell types after merging tiny cell clusters  
Cell-cell graph contains 4 cell types after merging  
Number of useful pathways is 4
```

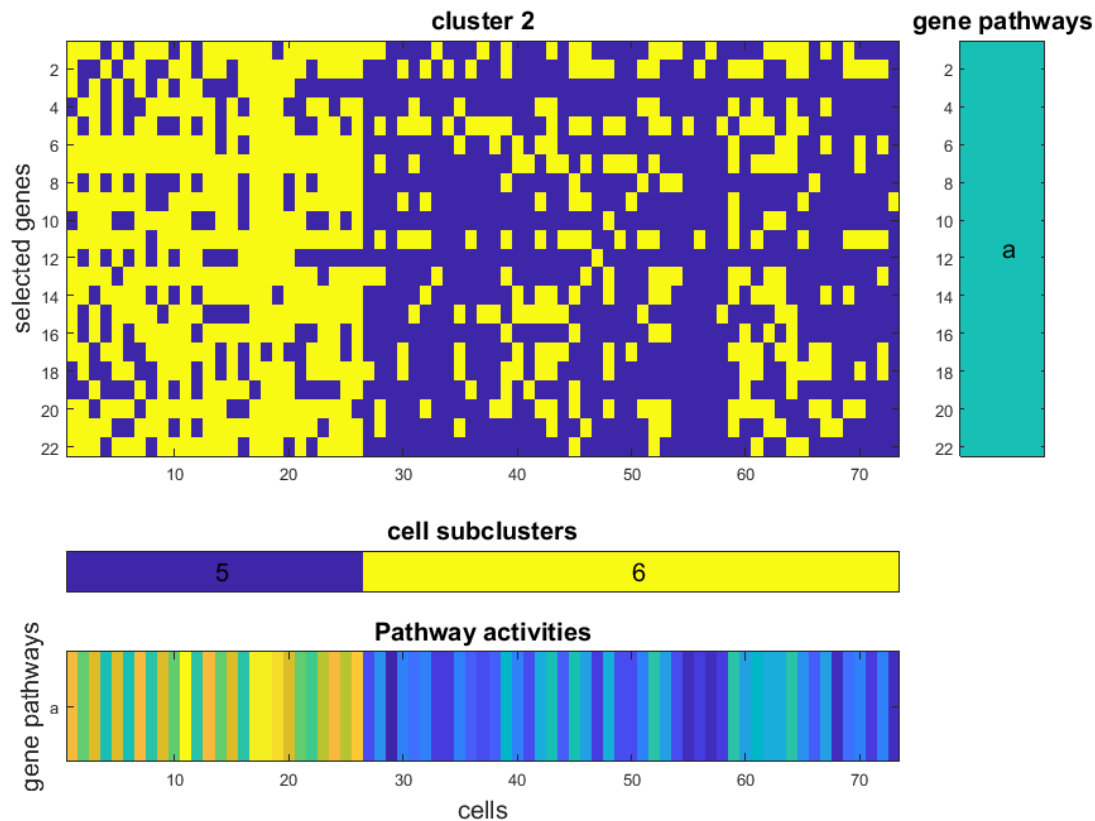


```

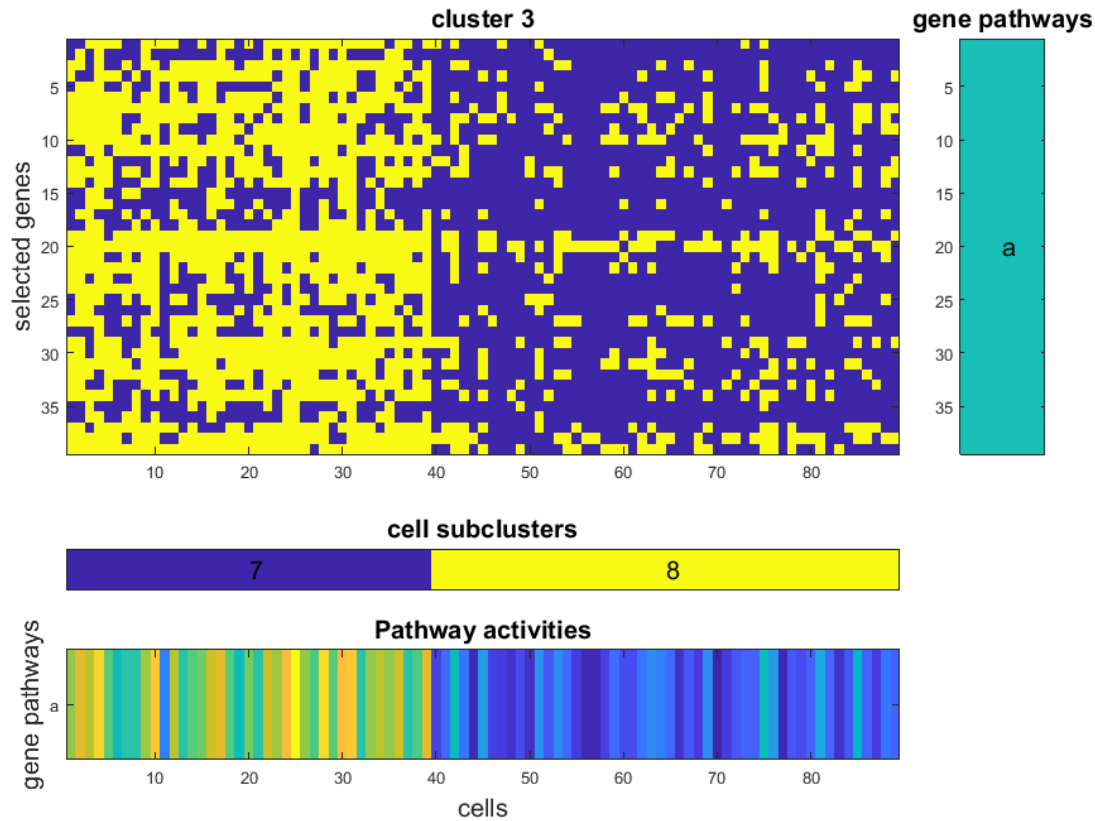
Remaining clusters to partition 4
Processing cluster 1 now ...
Processing data subset with 15726 genes and 88 cells:
Remove genes detected in <10 cells. Remaining 11507 genes. Elapsed time is 0.007131 seconds.
Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 83.715222 seconds.
Compute gene-gene similarity ... Elapsed time is 7.061592 seconds.
Create gene-gene graph for clustering genes ...
Writing graph into file ... 100%Elapsed time is 1.042124 seconds.
Running ModularityOptimizer for clustering ..Elapsed time is 2.655275 seconds.
Gene-gene graph contains 7 pathways, 3088 genes in total
Elapsed time is 3.088986 seconds.
Create cell-cell graph for clustering cells ...
Writing graph into file ... 100%Elapsed time is 0.014736 seconds.
Running ModularityOptimizer for clustering ..Elapsed time is 0.269005 seconds.
Cell-cell graph contains 3 cell types by community detection
Elapsed time is 0.273310 seconds.
Cell-cell graph contains 3 cell types after merging tiny cell clusters
creating a total of 2 edges ... 2
Cell-cell graph contains 1 cell types after merging

Remaining clusters to partition 3
Processing cluster 2 now ...
Processing data subset with 15726 genes and 73 cells:
Remove genes detected in <10 cells. Remaining 9280 genes. Elapsed time is 0.006932 seconds.
Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 54.354745 seconds.
Compute gene-gene similarity ... Elapsed time is 4.621719 seconds.
Create gene-gene graph for clustering genes ...
Writing graph into file ... 100%Elapsed time is 0.595589 seconds.
Running ModularityOptimizer for clustering ..Elapsed time is 0.772872 seconds.
Gene-gene graph contains 6 pathways, 251 genes in total
    
```

Elapsed time is 1.082240 seconds.
 Create cell-cell graph for clustering cells ...
 Writing graph into file ... 100%Elapsed time is 0.009137 seconds.
 Running ModularityOptimizer for clustering ...Elapsed time is 0.242460 seconds.
 Cell-cell graph contains 3 cell types by community detection
 Elapsed time is 0.246409 seconds.
 Cell-cell graph contains 3 cell types after merging tiny cell clusters
 Cell-cell graph contains 2 cell types after merging
 Number of useful pathways is 1

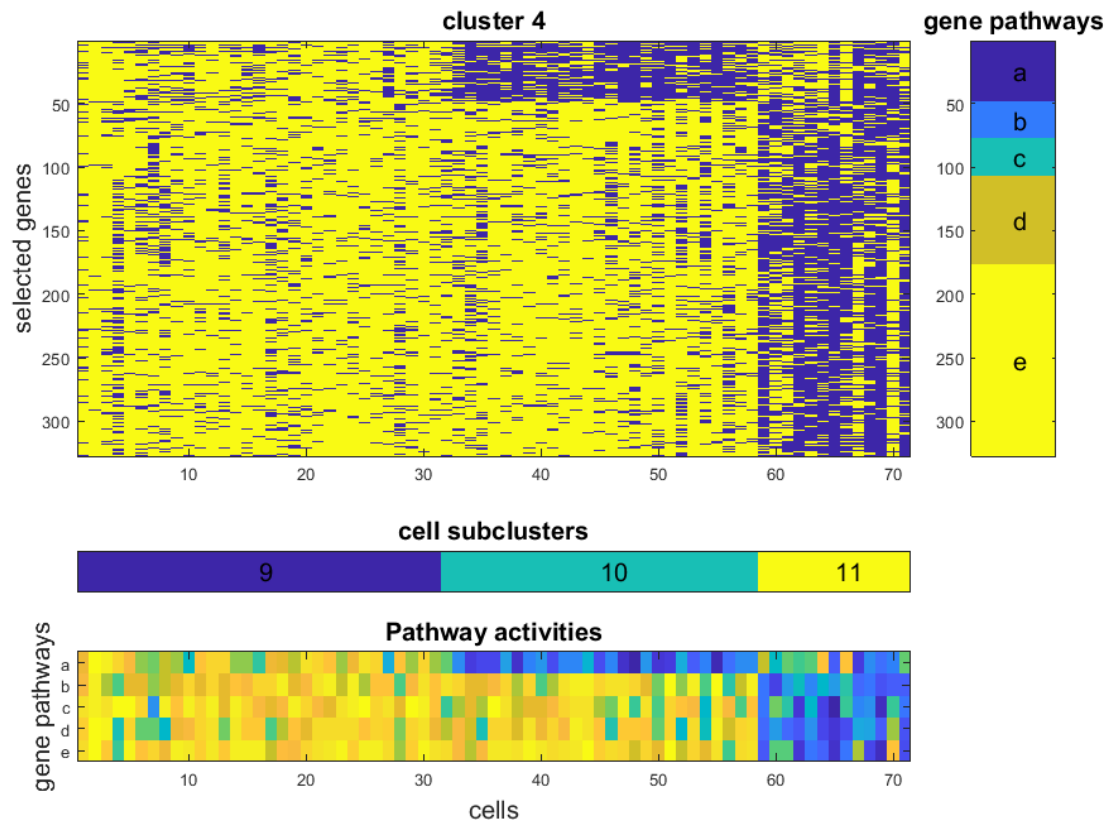


Remaining clusters to partition 4
 Processing cluster 3 now ...
 Processing data subset with 15726 genes and 89 cells:
 Remove genes detected in <10 cells. Remaining 10185 genes. Elapsed time is 0.008578 seconds.
 Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 65.266880 seconds.
 Compute gene-gene similarity ... Elapsed time is 5.561086 seconds.
 Create gene-gene graph for clustering genes ...
 Writing graph into file ... 100%Elapsed time is 0.711840 seconds.
 Running ModularityOptimizer for clustering ...Elapsed time is 0.714317 seconds.
 Gene-gene graph contains 8 pathways, 345 genes in total
 Elapsed time is 1.084309 seconds.
 Create cell-cell graph for clustering cells ...
 Writing graph into file ... 100%Elapsed time is 0.010831 seconds.
 Running ModularityOptimizer for clustering ...Elapsed time is 0.241295 seconds.
 Cell-cell graph contains 2 cell types by community detection
 Elapsed time is 0.245622 seconds.
 Cell-cell graph contains 2 cell types after merging tiny cell clusters
 Cell-cell graph contains 2 cell types after merging
 Number of useful pathways is 1



```

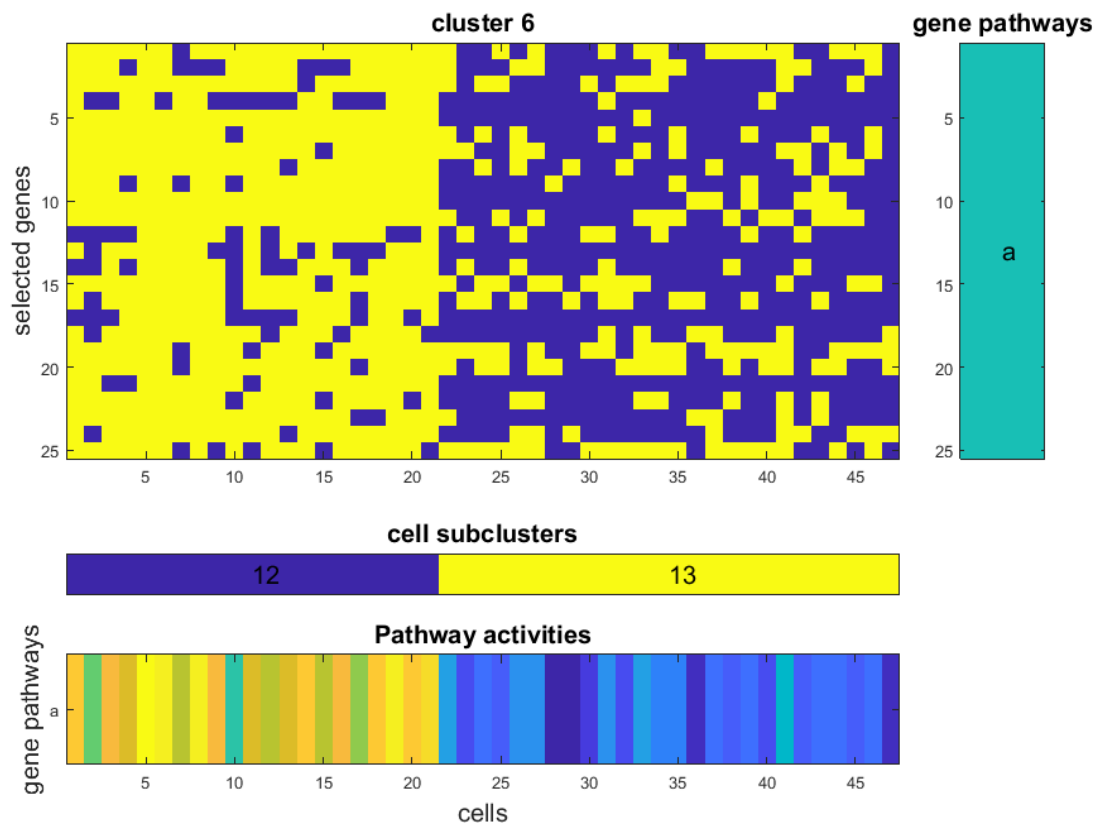
Remaining clusters to partition 5
Processing cluster 4 now ...
Processing data subset with 15726 genes and 71 cells:
Remove genes detected in <10 cells. Remaining 8638 genes. Elapsed time is 0.005994 seconds.
Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 46.857498 seconds.
Compute gene-gene similarity ... Elapsed time is 4.097090 seconds.
Create gene-gene graph for clustering genes ...
Writing graph into file ... 100%Elapsed time is 0.462738 seconds.
Running ModularityOptimizer for clustering ...Elapsed time is 0.705146 seconds.
Gene-gene graph contains 8 pathways, 537 genes in total
Elapsed time is 0.980547 seconds.
Create cell-cell graph for clustering cells ...
Writing graph into file ... 100%Elapsed time is 0.009228 seconds.
Running ModularityOptimizer for clustering ...Elapsed time is 0.208774 seconds.
Cell-cell graph contains 3 cell types by community detection
Elapsed time is 0.212607 seconds.
Cell-cell graph contains 3 cell types after merging tiny cell clusters
Cell-cell graph contains 3 cell types after merging
Number of useful pathways is 5
    
```



Remaining clusters to partition 7
 Processing cluster 5 now ...
 Processing data subset with 15726 genes and 26 cells:
 Remove genes detected in <10 cells. Remaining 3079 genes. Elapsed time is 0.001234 seconds.
 Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 5.509668 seconds.
 Compute gene-gene similarity ... Elapsed time is 0.497762 seconds.
 Create gene-gene graph for clustering genes ...
 Writing graph into file ... 100%Elapsed time is 0.061258 seconds.
 Running ModularityOptimizer for clustering ... Elapsed time is 0.407234 seconds.
 Gene-gene graph contains 0 pathways, 0 genes in total
 Elapsed time is 0.489395 seconds.

Remaining clusters to partition 6
 Processing cluster 6 now ...
 Processing data subset with 15726 genes and 47 cells:
 Remove genes detected in <10 cells. Remaining 7107 genes. Elapsed time is 0.003269 seconds.
 Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 31.713502 seconds.
 Compute gene-gene similarity ... Elapsed time is 2.728581 seconds.
 Create gene-gene graph for clustering genes ...
 Writing graph into file ... 100%Elapsed time is 0.304955 seconds.
 Running ModularityOptimizer for clustering ... Elapsed time is 0.503480 seconds.
 Gene-gene graph contains 1 pathways, 25 genes in total
 Elapsed time is 0.722700 seconds.
 Create cell-cell graph for clustering cells ...
 Writing graph into file ... 100%Elapsed time is 0.004826 seconds.
 Running ModularityOptimizer for clustering ... Elapsed time is 0.191876 seconds.
 Cell-cell graph contains 2 cell types by community detection
 Elapsed time is 0.195402 seconds.
 Cell-cell graph contains 2 cell types after merging tiny cell clusters
 Cell-cell graph contains 2 cell types after merging

Number of useful pathways is 1



Remaining clusters to partition 7
 Processing cluster 7 now ...
 Processing data subset with 15726 genes and 39 cells:
 Remove genes detected in <10 cells. Remaining 6173 genes. Elapsed time is 0.002855 seconds.
 Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 23.158681 seconds.
 Compute gene-gene similarity ... Elapsed time is 1.996613 seconds.
 Create gene-gene graph for clustering genes ...
 Writing graph into file ... 100% Elapsed time is 0.241264 seconds.
 Running ModularityOptimizer for clustering ... Elapsed time is 0.507929 seconds.
 Gene-gene graph contains 0 pathways, 0 genes in total
 Elapsed time is 0.691853 seconds.

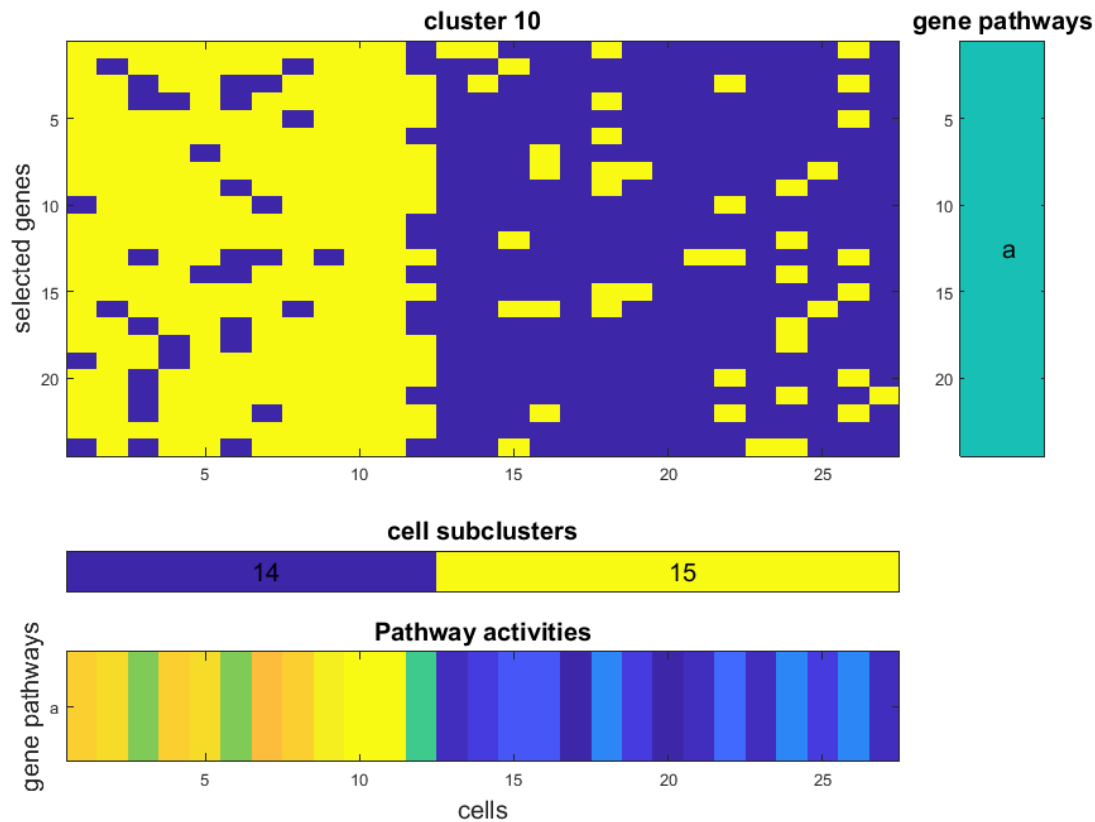
Remaining clusters to partition 6
 Processing cluster 8 now ...
 Processing data subset with 15726 genes and 50 cells:
 Remove genes detected in <10 cells. Remaining 7358 genes. Elapsed time is 0.003744 seconds.
 Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 33.522333 seconds.
 Compute gene-gene similarity ... Elapsed time is 2.879052 seconds.
 Create gene-gene graph for clustering genes ...
 Writing graph into file ... 100% Elapsed time is 0.327440 seconds.
 Running ModularityOptimizer for clustering ... Elapsed time is 0.533027 seconds.
 Gene-gene graph contains 1 pathways, 35 genes in total
 Elapsed time is 0.756172 seconds.
 Create cell-cell graph for clustering cells ...
 Writing graph into file ... 100% Elapsed time is 0.005766 seconds.
 Running ModularityOptimizer for clustering ... Elapsed time is 0.194528 seconds.
 Cell-cell graph contains 2 cell types by community detection

step_02_cooccurrence_clustering_pipeline

```
Elapsed time is 0.198064 seconds.
Cell-cell graph contains 2 cell types after merging tiny cell clusters
creating a total of 1 edges ... 1
Cell-cell graph contains 1 cell types after merging

Remaining clusters to partition 5
Processing cluster 9 now ...
Processing data subset with 15726 genes and 31 cells:
Remove genes detected in <10 cells. Remaining 4003 genes. Elapsed time is 0.001523 seconds.
Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 9.563860 seconds.
Compute gene-gene similarity ... Elapsed time is 0.843559 seconds.
Create gene-gene graph for clustering genes ...
Writing graph into file ... 100%Elapsed time is 0.098213 seconds.
Running ModularityOptimizer for clustering ...Elapsed time is 0.345061 seconds.
Gene-gene graph contains 0 pathways, 0 genes in total
Elapsed time is 0.451224 seconds.

Remaining clusters to partition 4
Processing cluster 10 now ...
Processing data subset with 15726 genes and 27 cells:
Remove genes detected in <10 cells. Remaining 2843 genes. Elapsed time is 0.001309 seconds.
Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 4.719851 seconds.
Compute gene-gene similarity ... Elapsed time is 0.445212 seconds.
Create gene-gene graph for clustering genes ...
Writing graph into file ... 100%Elapsed time is 0.051407 seconds.
Running ModularityOptimizer for clustering ...Elapsed time is 0.343791 seconds.
Gene-gene graph contains 1 pathways, 24 genes in total
Elapsed time is 0.416527 seconds.
Create cell-cell graph for clustering cells ...
Writing graph into file ... 100%Elapsed time is 0.002542 seconds.
Running ModularityOptimizer for clustering ...Elapsed time is 0.161416 seconds.
Cell-cell graph contains 2 cell types by community detection
Elapsed time is 0.164378 seconds.
Cell-cell graph contains 2 cell types after merging tiny cell clusters
Cell-cell graph contains 2 cell types after merging
Number of useful pathways is 1
```

```

Remaining clusters to partition 5
Processing cluster 11 now ...
Processing data subset with 15726 genes and 13 cells:
Remove genes detected in <10 cells. Remaining 0 genes. Elapsed time is 0.002375 seconds.

Remaining clusters to partition 4
Processing cluster 12 now ...
Processing data subset with 15726 genes and 21 cells:
Remove genes detected in <10 cells. Remaining 1063 genes. Elapsed time is 0.003191 seconds.
Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 0.734368 seconds.
Compute gene-gene similarity ... Elapsed time is 0.063500 seconds.
Create gene-gene graph for clustering genes ...
Writing graph into file ... 101%Elapsed time is 0.009782 seconds.
Running ModularityOptimizer for clustering ...Elapsed time is 0.273663 seconds.
Gene-gene graph contains 0 pathways, 0 genes in total
Elapsed time is 0.301217 seconds.

Remaining clusters to partition 3
Processing cluster 13 now ...
Processing data subset with 15726 genes and 26 cells:
Remove genes detected in <10 cells. Remaining 3050 genes. Elapsed time is 0.002114 seconds.
Iterate 10 random permutations for gene-gene similarity threshold ... 10 Elapsed time is 5.592817 seconds.
Compute gene-gene similarity ... Elapsed time is 0.508726 seconds.
Create gene-gene graph for clustering genes ...
Writing graph into file ... 100%Elapsed time is 0.060390 seconds.
Running ModularityOptimizer for clustering ...Elapsed time is 0.344409 seconds.
Gene-gene graph contains 0 pathways, 0 genes in total
Elapsed time is 0.423228 seconds.

Remaining clusters to partition 2
    
```

step_02_cooccurrence_clustering_pipeline

```
Processing cluster 14 now ...
Processing data subset with 15726 genes and 12 cells:
Remove genes detected in <10 cells. Remaining 0 genes. Elapsed time is 0.000560 seconds.

Remaining clusters to partition 1
Processing cluster 15 now ...
Processing data subset with 15726 genes and 15 cells:
Remove genes detected in <10 cells. Remaining 0 genes. Elapsed time is 0.000647 seconds.
```

re-order pathways for better visualization

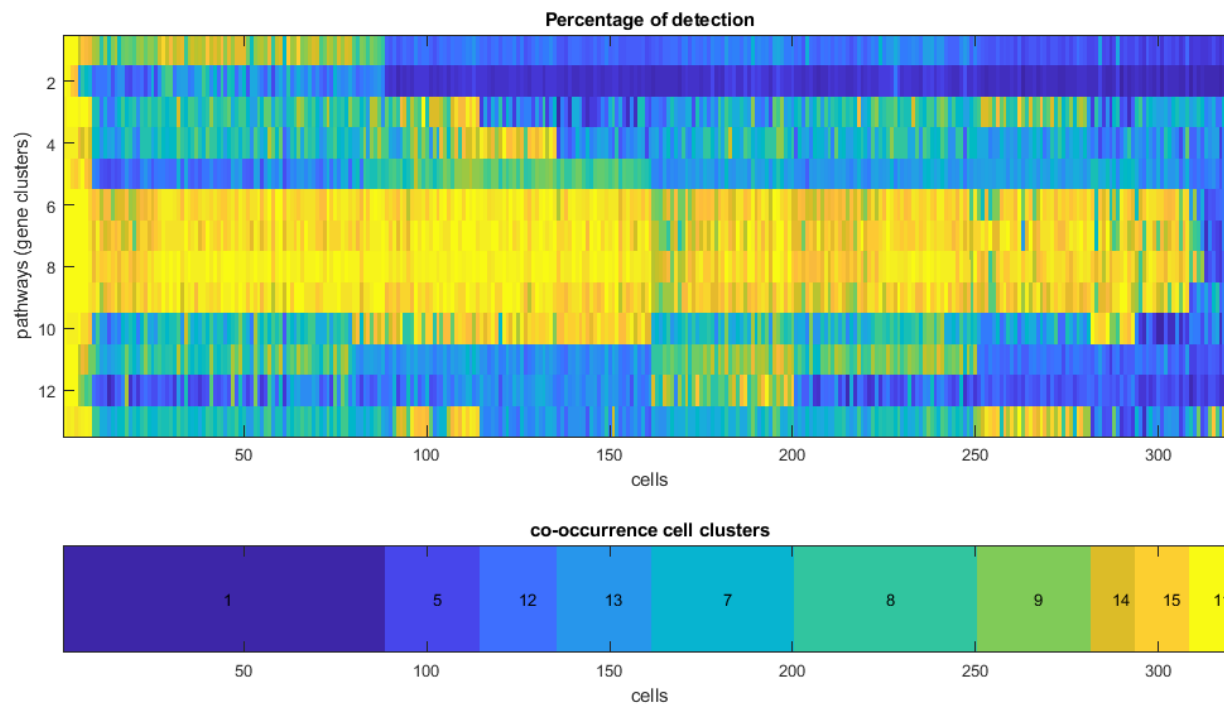
```
pathway_in_cell_detection_rate = cooc.pathway_in_cell_detection_rate;
scores = zeros(size(pathway_in_cell_detection_rate,1),max(cooc.cell_labels));
for k=1:max(cooc.cell_labels)
    for l=k+1:max(cooc.cell_labels)
        tmp = get_correlations(pathway_in_cell_detection_rate(:,ismember(cooc.cell_labels,[k,l])),cooc.cell_labels(ismember(cooc.cell_labels,[k,l])), 'snr');
        scores(:,k) = scores(:,k) - tmp;
        scores(:,l) = scores(:,l) + tmp;
    end
end
[max_score,signature_for_which_celltype] = max(scores,[],2);
pathway_order = [];
for k=1:max(cooc.cell_labels)
    tmp_pathway = find(signature_for_which_celltype==k);
    [~,I] = sort(max_score(tmp_pathway), 'descend');
    pathway_order = [pathway_order;tmp_pathway(I)];
end
cooc.pathway_in_cell_detection_rate = cooc.pathway_in_cell_detection_rate(pathway_order,:);
cooc.gene_pathway_indicator = cooc.gene_pathway_indicator(:,pathway_order);

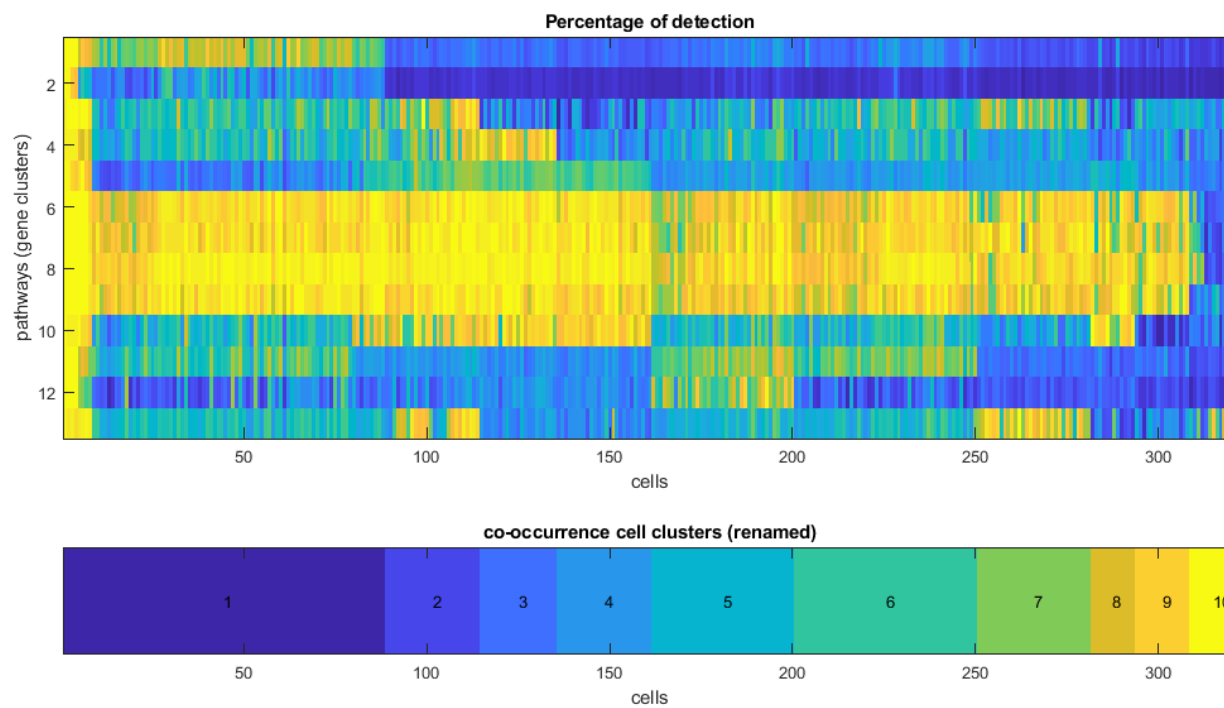
save([mfilename, '_result.mat']);
```

Visualize the percentages of detection for pathways in individual cells

```
[~,I] = sort(cooc.cell_labels);
h = figure(30); set(h, 'position', [100 100 1200 600]);
subplot(6,1,1:4); imagesc(cooc.pathway_in_cell_detection_rate(:, I)./max(cooc.pathway_in_cell_detection_rate(:, I),[],2)); ylabel('pathways (gene clusters)'); xlabel('cells'); title('Percentage of detection')
h = subplot(4,1,4); imagesc(cooc.cell_labels(I)); h.YTick=[];
for k=1:max(cooc.cell_labels)
    text(mean(find(cooc.cell_labels(I)==k)), 1, num2str(unique(cooc.cell_labels_history(cooc.cell_labels==k,end))));
end
title('co-occurrence cell clusters'); xlabel('cells')

h = figure(40); set(h, 'position', [100 100 1200 600]);
subplot(6,1,1:4); imagesc(cooc.pathway_in_cell_detection_rate(:, I)./max(cooc.pathway_in_cell_detection_rate(:, I),[],2)); ylabel('pathways (gene clusters)'); xlabel('cells'); title('Percentage of detection')
h = subplot(4,1,4); imagesc(cooc.cell_labels(I)); h.YTick=[];
for k=1:max(cooc.cell_labels)
    text(mean(find(cooc.cell_labels(I)==k)), 1, num2str(k));
end
title('co-occurrence cell clusters (renamed)'); xlabel('cells')
```

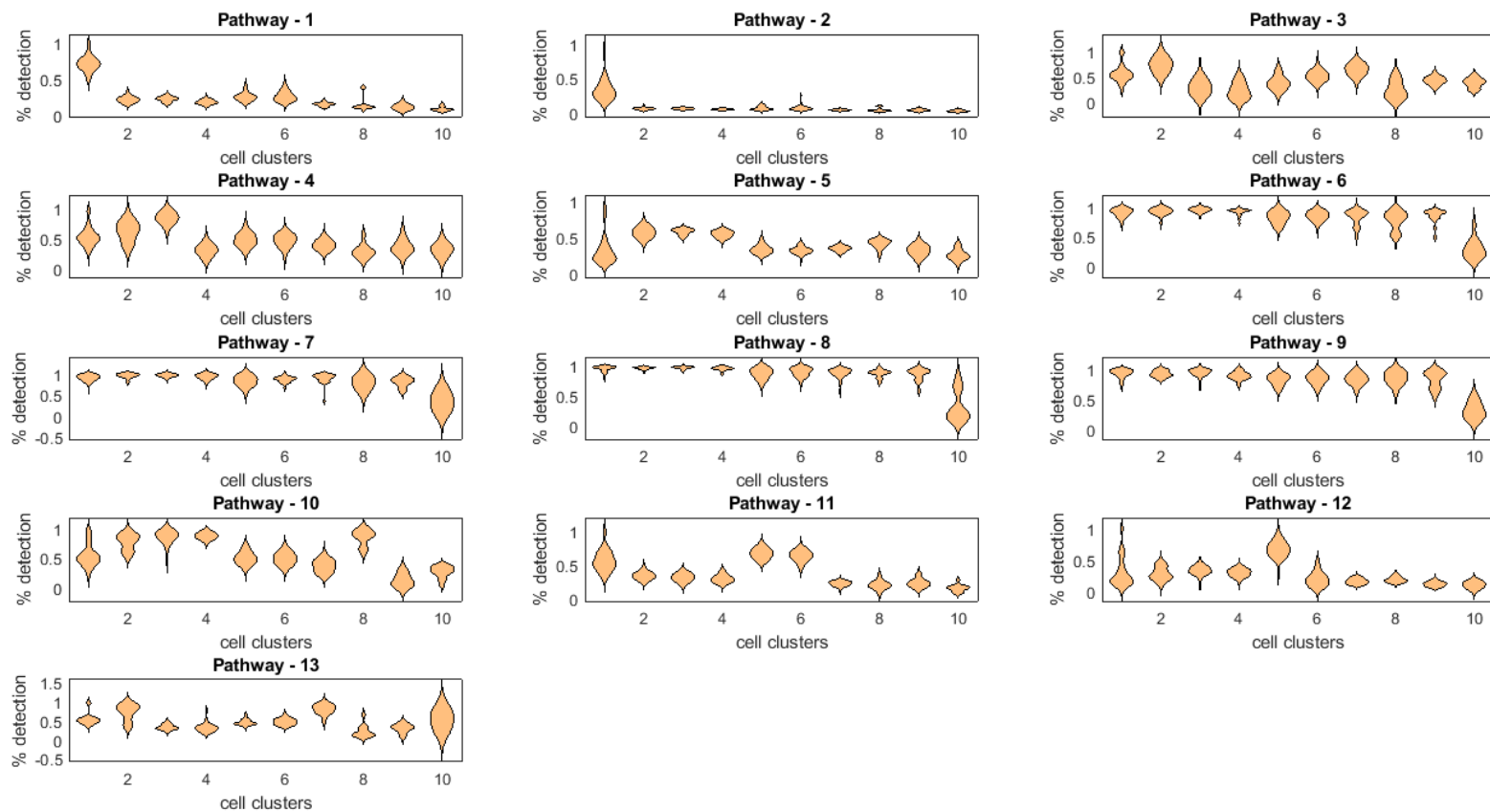




```

h = figure(5); set(h, 'units','normalized','outerposition',[0 0 1 1])
for i=1:size(cooc.pathway_in_cell_detection_rate,1)
    tmp = cell(0);
    for k=1:max(cooc.cell_labels)
        tmp{k} = cooc.pathway_in_cell_detection_rate(i,cooc.cell_labels==k);
    end
    subplot(5, ceil(size(cooc.pathway_in_cell_detection_rate,1)/5),i); violin(tmp,'mc',[],'medc',[]);
    title(num2str(i,'Pathway - %d'))
    xlabel('cell clusters');
    ylabel('% detection');
    drawnow
end

```



export results to files

```
% pathway_names = strcat('pathway-', regexp(num2str(1:size(cooc.gene_pathway_indicator,2),'%05d'),'','split')); pathway_names(end)=[];
% write_to_txt_v2([filename,'_result_gene_pathway_indicator.csv'], [{' '}, pathway_names], cooc.gene_names, cooc.gene_pathway_indicator, '');
% write_to_txt_v2([filename,'_result_pathway_in_cell_detection_rate.csv'], [{' '}, cooc.cell_names], pathway_names, cooc.pathway_in_cell_detection_rate, '');
% write_to_txt_v2([filename,'_result_cell_clusters.csv'], [{' '}, cooc.cell_names], {'cluster idx'}, cooc.cell_labels, '');
```