

---

## Table of Contents

.....	1
part 1 .....	1
Part 2 .....	1
Part 3 .....	1

```
% Problem 1
% Approach 2 using GProfiler to identify goterms
% part 1: imports a binary matrix of 1s and 0s with GoTerms vs Genes
%         GProfiler identified 5766 genes and sorted them into 1034
%         GoTerms
%         Also removed all goTerms with less then 10 genes
% part 2: Imports the genes vs tissues data and sorts the genes
%         according to the same order as the genes vs goterms matrix
% part 3: GoTerms vs Genes * Genes vs tissues = GoTerms vs Tissues
%         Dot product of these 2 matrixes is the Go enierichment of the
%         data
%         Only one method of normalization used here
```

## part 1

```
load('gProfiler1Sk.mat');
load('gProfGenes.mat');
gProfData = gProfiler1Sk(:,3:5768);
goTerms = gProfiler1Sk(:,1);
gProfData.Properties.RowNames = strrep(goTerms.VarName1, '-', '');
gProfData.Properties.VariableNames = cellstr(gProfGenes);
gProfData.count = sum(table2array(gProfData),2);
counts = gProfData.count;
gProfDataSort = sortrows(gProfData,{'count'},'descend');
gProfDataSort(1035,:) = [];
GPD = gProfDataSort(:,1:5766);
```

## Part 2

```
load('proteinCons2.mat');
proteinCons2.Properties.RowNames = proteinCons2.ENSEG00000161057;
proteinCons2(:,1) = [];
[lia,locB] = ismember(proteinCons2.Properties.RowNames,...
    gProfDataSort.Properties.VariableNames);
slavProtShort = proteinCons2(lia,:);
goTermGene = table2array(GPD);
geneTiss = table2array(slavProtShort);
```

## Part 3

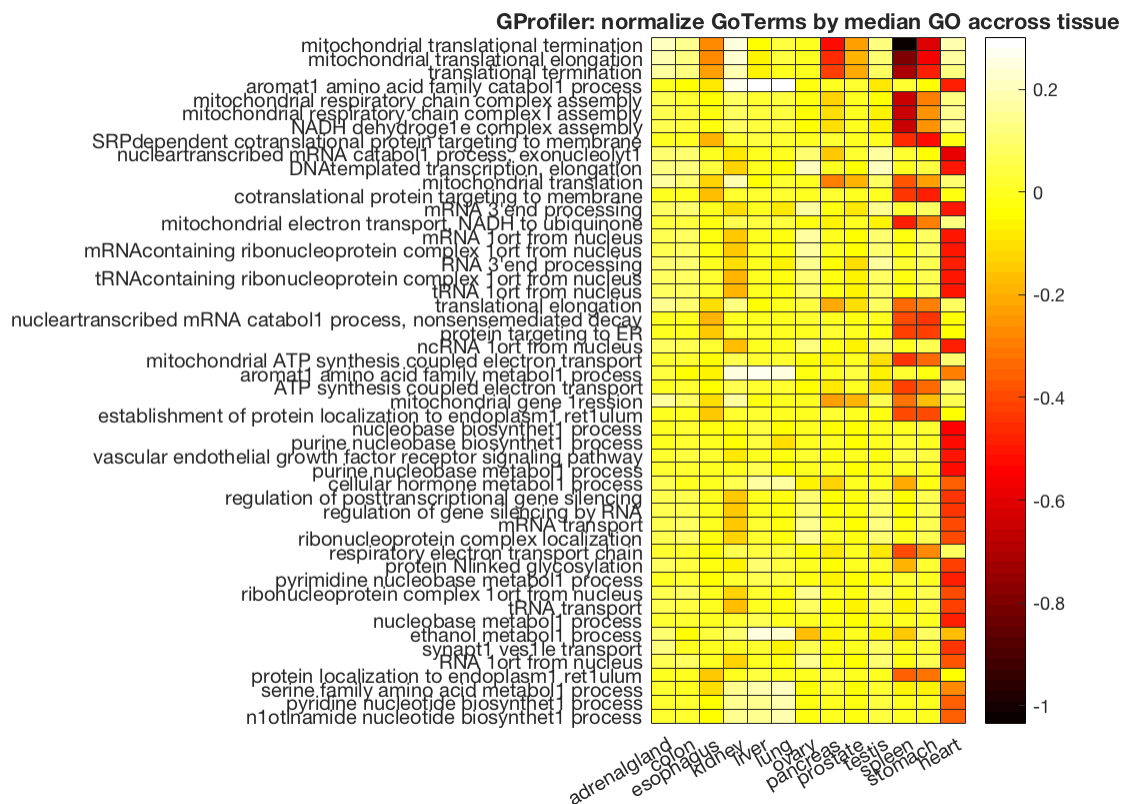
```
enrich = goTermGene*geneTiss;
enrichNorm = enrich;
```

```

for i = 1:13;
    enrichNorm(:,i) = log10(enrich(:,i)./median(enrich,2));
end
var1 = std(enrichNorm,0,2);
tis =
    {'adrenalgland','colon','esophagus','kidney','liver','lung','ovary',...
     'pancreas','prostate','testis','spleen','stomach','heart'};
enrichTab =
    array2table(enrichNorm,'RowNames',GPD.Properties.RowNames...
    , 'VariableNames',tis);
enrichTab.VAR = var1;
enrichTabSort = sortrows(enrichTab,{ 'VAR' }, 'descend');

enrich50 = enrichTabSort(1:50,1:13);
xlab = enrich50.Properties.VariableNames;
ylab = enrich50.Properties.RowNames;
figure(1);
h = heatmap(xlab,ylab,table2array(enrich50),'ColorMap',hot);
title('GProfiler: normalize GoTerms by median GO accross tissue');

```



Published with MATLAB® R2017a