
Table of Contents

.....	1
part 1	1
Part 2	1
Part 3a version 1 normalize the GO terms by median GO accross tissue	2
Part 3b version 2 normalize protein abundance	3
Part 3c Version 3 normalize GO terms but assume data is already in log10	4

```
% Problem 1
% Approach 1 using GoAmigo to identify goterms
% part 1: creates a binary matrix of 1s and 0s with GoTerms vs Genes
%         GoAmigo could only identify go terms for 3.5k of the genes
%         it matched those genes to 3.5k 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 enerichment of the
data
%         There are 3 normalization approaches that yeilded slight
%         different results.
```

part 1

```
load('martexport3.mat');
martexport3.Properties.VariableNames = {'Genes','GoTerms'};
martexport3.GoTerms = strrep(martexport3.GoTerms,':',' ');
g = unique(martexport3.Genes);
t = unique(martexport3.GoTerms);
mega = zeros(length(g),length(t));
rows = cellstr(g);
cols = cellstr(t);
MEGA = array2table(mega,'VariableNames',cols,'RowNames',rows);
for i = 1:length(martexport3.Genes);
    MEGA(char(martexport3.Genes(i)),char(martexport3.GoTerms(i))) =
    {1};
end
gosums = sum(table2array(MEGA),1);
MEGAcell = table2cell(MEGA);
MEGAflip =
    cell2table(MEGAcell,'RowNames',MEGA.Properties.VariableNames,...
    'VariableNames',MEGA.Properties.RowNames);
```

Part 2

```
MEGAflip.counts = gosums';
MEGA10 = sortrows(MEGAflip,{'counts'},'descend');
```

```

MEGA10(:,3510) = [];
MEGA10(436:3518,:) = [];
gosums4 = sum(table2array(MEGA10),2);
load('proteinConsSk.mat');
proteinConsSk.Properties.VariableNames =
    {'Gen','T1','T2','T3','T4','T5',...
     'T6','T7','T8','T9','T10','T11','T12','T13'};
[lia,locB] = ismember(proteinConsSk.Gen,g);
proteinShort = proteinConsSk(lia,:);
[lia2,locB2] = ismember(proteinShort.Gen,g);
gotgene = table2array(MEGA10);
genetiss = table2array(proteinShort(:,2:14));
%genetissNorm = zeros(3509,13);

```

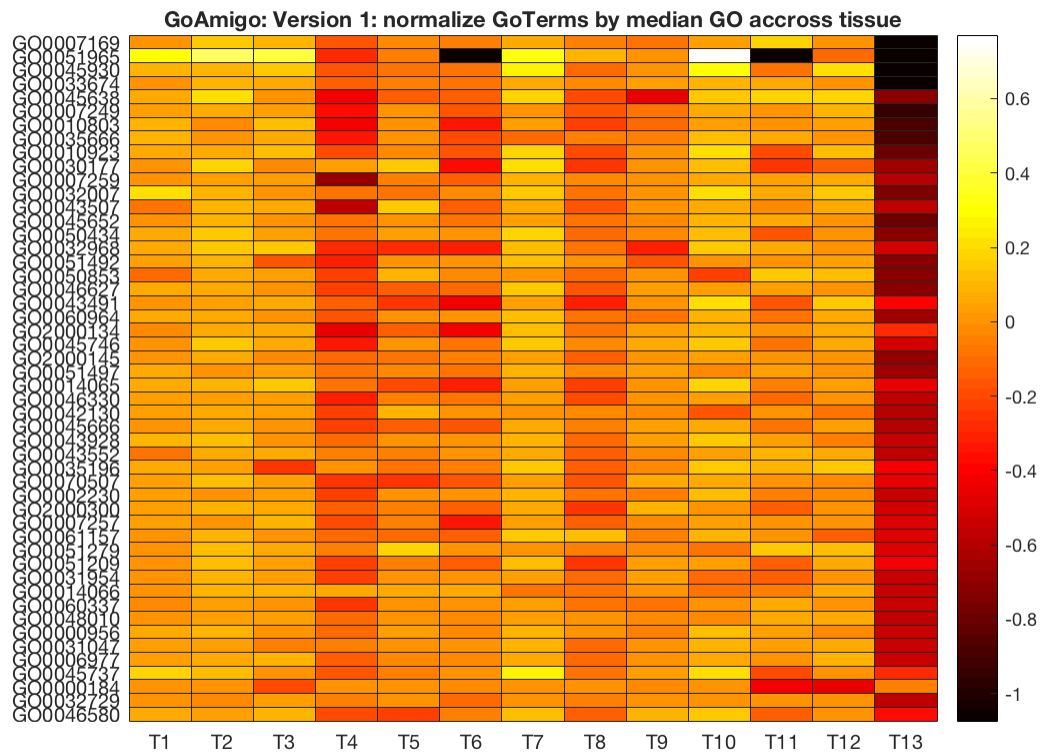
Part 3a version 1 normalize the GO terms by median GO accross tissue

```

enrich = gotgene*genetiss;
enrichNorm = enrich;
for i = 1:13;
    enrichNorm(:,i) = log10(enrich(:,i)./median(enrich,2));
end
var1 = std(enrichNorm,0,2);
tis = {'T1','T2','T3','T4','T5',...
       'T6','T7','T8','T9','T10','T11','T12','T13'};
enrichTab =
    array2table(enrichNorm,'RowNames',MEGA10.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;
map = [0,0,1;1,1,1;1,0,0];
figure(1);
h = heatmap(xlab,ylab,table2array(enrich50),'ColorMap',hot);
title('GoAmigo: Version 1: normalize GoTerms by median GO accross
      tissue');
%%%%%%%%%%%%%%

```



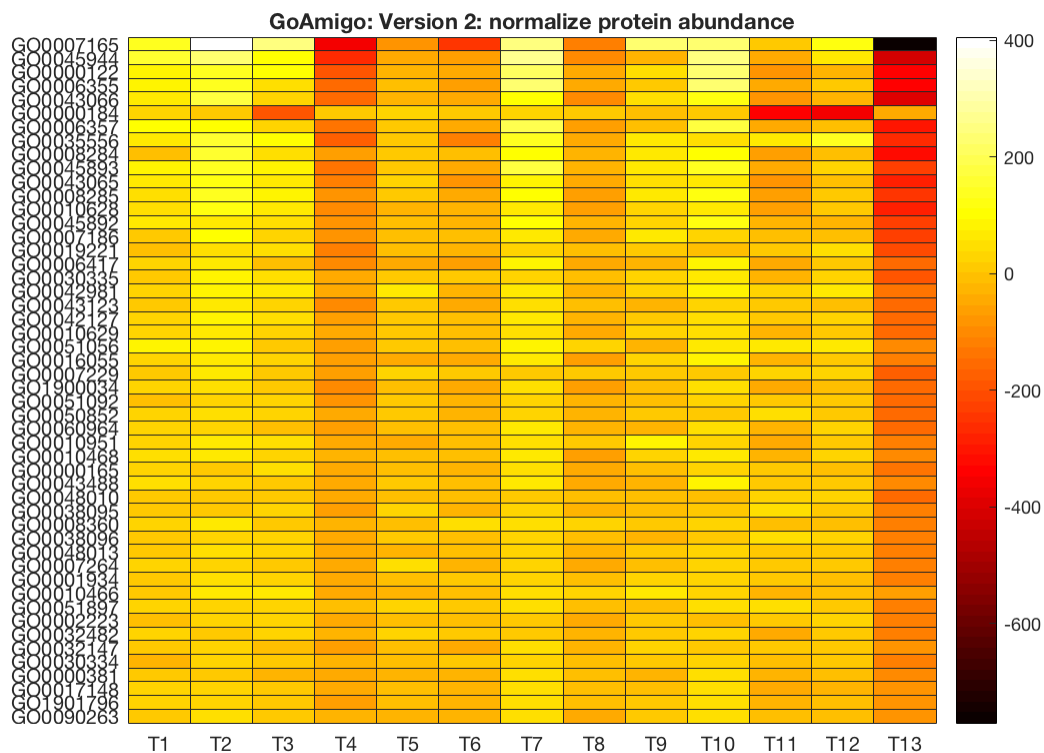
Part 3b version 2 normalize protein abundance

```
genetiss10 = 10.^genetiss;
med = log10(median(genetiss10,2));
genetissNorm = genetiss - med;

enrich2 = gotgene*genetissNorm;

var2 = std(enrich2,0,2);
tis2 = {'T1','T2','T3','T4','T5',...
        'T6','T7','T8','T9','T10','T11','T12','T13'};
enrichTab2 =
    array2table(enrich2,'RowNames',MEGA10.Properties.RowNames...
        , 'VariableNames',tis2);
enrichTab2.VAR = var2;
enrichTabSort2 = sortrows(enrichTab2,{'VAR'}, 'descend');

enrich502 = enrichTabSort2(1:50,1:13);
xlabl2 = enrich502.Properties.VariableNames;
ylabl2 = enrich502.Properties.RowNames;
figure(2);
h2 = heatmap(xlabl2,ylabl2,table2array(enrich502),'ColorMap',hot);
title('GoAmigo: Version 2: normalize protein abundance');
```



```
liaC = ismember(ylab12,ylab);
sum(liaC);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Part 3c Version 3 normalize GO terms but assume data is already in log10

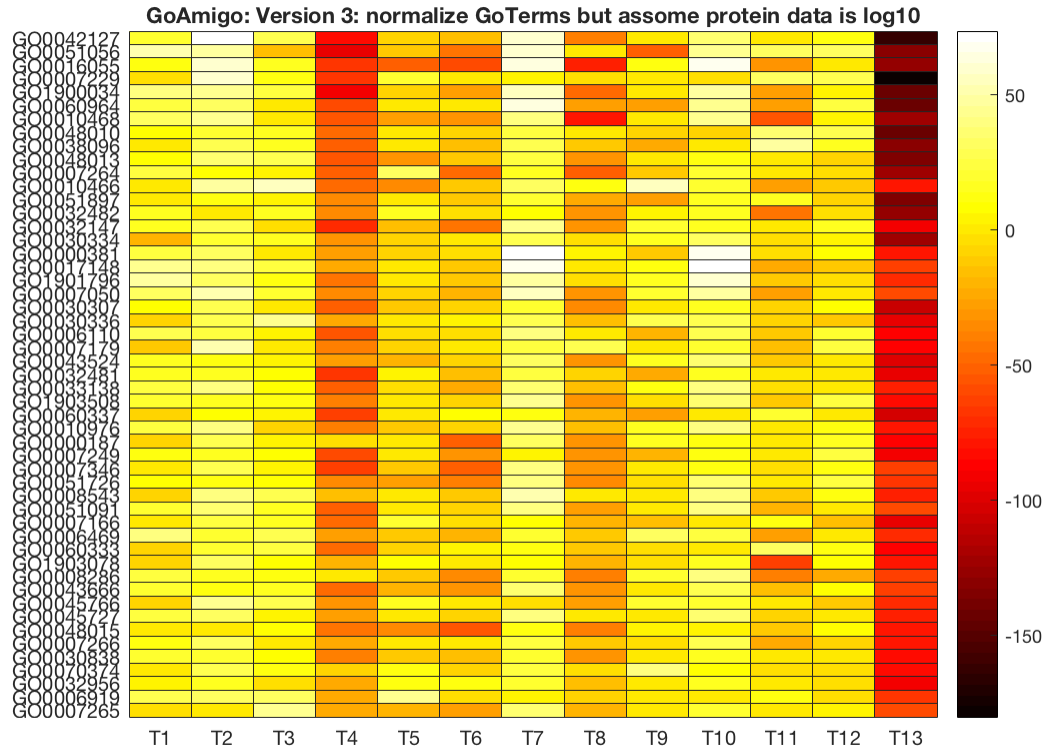
```
enrich3 = gotgene*genetiss;
enrichNorm3 = enrich3 - log10(median(10.^enrich3,2));
var3 = std(enrichNorm3,0,2);
tis = {'T1','T2','T3','T4','T5',...
      'T6','T7','T8','T9','T10','T11','T12','T13'};
enrichTab3 =
    array2table(enrichNorm3,'RowNames',MEGA10.Properties.RowNames...
      ,'VariableNames',tis);
enrichTab3.VAR = var3;
enrichTabSort3 = sortrows(enrichTab3,{'VAR'},'descend');

enrich503 = enrichTabSort3(39:89,1:13);
xlab3 = enrich503.Properties.VariableNames;
ylab3 = enrich503.Properties.RowNames;
map = [0,0,1;1,1,1;1,0,0];
figure(3);
c = heatmap(xlab3,ylab3,table2array(enrich503),'ColorMap',hot);
```

```

title('GoAmigo: Version 3: normalize GoTerms but assome protein data
      is log10');
lia3 = ismember(ylab3,ylab12);
sum(lia3);

```



Published with MATLAB® R2017a