

Global Thresholding for everyone

Load the data

Load the transcriptomics data, the housekeeping genes obtained in the previous script, and the metabolic model (Human1 version).

```
% Transcriptomics data
data = readtable('Mod_data.xlsx')
```

```
data = 26269x49 table
```

	Ensembl_GeneID	BJ_Y1	BJ_Y2	BJ_Y3	BJ_OLD_1	BJ_OLD_2	BJ_OLD_3
1	'ENSG00000000003'	6.4934	6.7186	7.1946	10.0460	9.4987	9.9429
2	'ENSG00000000005'	0	0	0	0	0	0
3	'ENSG000000000419'	24.2420	21.2693	24.6489	22.6069	23.4380	23.1285
4	'ENSG000000000457'	2.4154	2.4048	2.4530	2.0556	2.3939	2.2430
5	'ENSG000000000460'	1.9736	2.0035	2.2052	0.9924	1.0449	1.0548
6	'ENSG000000000938'	0.0420	0.0450	0.0272	0.0552	0.0554	0.0369
7	'ENSG000000000971'	5.6428	5.5900	5.3046	13.5653	13.1726	13.5571
8	'ENSG000000001036'	36.4484	34.3146	34.8401	40.4961	40.3793	40.6007
9	'ENSG000000001084'	1.5123	1.5729	1.7464	2.0652	2.2006	2.2952
10	'ENSG000000001167'	7.6099	8.6422	8.1970	5.5155	5.9501	5.9592
11	'ENSG000000001460'	1.7292	1.8737	1.7218	1.4539	1.5739	1.5477
12	'ENSG000000001461'	11.7417	12.2397	12.4404	23.1796	23.2848	23.3608
13	'ENSG000000001497'	6.2678	6.1339	6.0014	5.8433	5.7648	5.9670
14	'ENSG000000001561'	1.6520	1.6535	1.5636	4.6577	4.3341	4.3173
:							

```
% Housekeeping genes with the ensembl ids
h_k_g = readtable('housekeeping_ens.csv');
% Human1 metabolic model.
model = readCbModel('Human-GEM_Cobra_v1.01.mat');
```

Each model.subSystems{x} is a cell array, allowing more than one subSystem per reaction.

Metabolic genes

Pick the genes related to metabolism

In the transcriptomics dataset there are a lot of genes, but we are interested just in those ones that take part in the metabolism, so, we are going to compare the ensembl id of both, the transcriptomics dataset and the Human1 metabolic model, and pick those ones that are present.

```

% Create a vector just with the ensembl ids of the genes from the model
model_genes = model.genes;
% Find indexes of the genes that are present in the metabolic model.
index_names = ismember(data.Ensembl_GeneID, model_genes);
% take the rows of the dataset that match the indexes obtained before, with all the
data
data_met = data(index_names, :);

```

For loop to calculate the accuracy for all the samples

```

% Initialise a table to store the results
results = table;
% Get expression data column names
expression_col = data_met.Properties.VariableNames(2:end);
results_colname = []

```

```

results_colname =

[]

```

```

results_percentage_hk = []

```

```

results_percentage_hk =

[]

```

now create the for loop where a threshold will be established based on the 70th percentile of gene expression for each sample, with genes above this threshold being considered core genes. Moreover, the accuracy will also be calculated on the basis of the housekeeping genes expressed as core genes divided by those related to metabolism.

```

% Iterate over each column of expression data
for i = 1:length(expression_col)
    % Select the expression data column
    col_name = expression_col{i};
    expfibroblast = data_met(:,1);
    expfibroblast(:,2) = data_met(:, col_name);
    expfibroblast.Properties.VariableNames{1} = 'gene';
    expfibroblast.Properties.VariableNames{2} = 'FPKMvalue';

    % Perform the calculations
    expfibroblast.logFPKMvalue = log10(expfibroblast{:,2} + 1);
    expfibroblast.value = expfibroblast.logFPKMvalue -
min(expfibroblast.logFPKMvalue);

    % Calculate the threshold, creating a slider form 70 to 100
    percentage = 90;
    up_threshold_fib = prctile(expfibroblast.value, percentage);

    % Filtering the data
    idx = expfibroblast.value >= up_threshold_fib;

```



```
results_FPKM = results
```

```
results_FPKM = 48x2 table
```

	Column	PercentageHK
1	'BJ_Y1'	57.7558
2	'BJ_Y2'	57.7558
3	'BJ_Y3'	57.4257
4	'BJ_OLD_1'	59.0759
5	'BJ_OLD_2'	58.7459
6	'BJ_OLD_3'	58.0858
7	'IMR90_Y1'	60.7261
8	'IMR90_Y2'	60.7261
9	'IMR90_Y3'	60.3960
10	'IMR90_O1'	60.3960
11	'IMR90_O2'	58.7459
12	'IMR90_O3'	61.0561
13	'WI_38_Y1'	63.0363
14	'WI_38_Y2'	62.3762

```
⋮
```

```
% Show results
disp(results_FPKM);
```

Column	PercentageHK
{ 'BJ_Y1' }	57.756
{ 'BJ_Y2' }	57.756
{ 'BJ_Y3' }	57.426
{ 'BJ_OLD_1' }	59.076
{ 'BJ_OLD_2' }	58.746
{ 'BJ_OLD_3' }	58.086
{ 'IMR90_Y1' }	60.726
{ 'IMR90_Y2' }	60.726
{ 'IMR90_Y3' }	60.396
{ 'IMR90_O1' }	60.396
{ 'IMR90_O2' }	58.746
{ 'IMR90_O3' }	61.056
{ 'WI_38_Y1' }	63.036
{ 'WI_38_Y2' }	62.376
{ 'WI_38_Y3' }	63.036
{ 'WI_38_O1' }	60.726
{ 'WI_38_O2' }	61.386
{ 'WI_38_O3' }	60.726
{ 'HFF_PD16_1' }	60.726
{ 'HFF_PD16_2' }	60.396
{ 'HFF_PD16_3' }	60.066
{ 'HFF_PD74_1' }	59.406
{ 'HFF_PD74_2' }	59.076

{ 'HFF_PD74_3' }	58.746
{ 'MRC_5_PD32_1' }	57.756
{ 'MRC_5_PD32_2' }	58.086
{ 'MRC_5_PD32_3' }	58.746
{ 'MRC_5_PD72_1' }	62.706
{ 'MRC_5_PD72_2' }	62.706
{ 'MRC_5_PD72_3' }	62.706
{ 'HFF_PD26_1' }	58.746
{ 'HFF_PD26_2' }	58.416
{ 'HFF_PD26_3' }	58.086
{ 'HFF_PD46_1' }	59.076
{ 'HFF_PD46_2' }	58.416
{ 'HFF_PD46_3' }	62.706
{ 'HFF_PD64_1' }	57.756
{ 'HFF_PD64_2' }	58.746
{ 'HFF_PD64_3' }	59.076
{ 'MRC_5_PD42_1' }	59.736
{ 'MRC_5_PD42_2' }	59.736
{ 'MRC_5_PD42_3' }	59.736
{ 'MRC_5_PD52_1' }	59.406
{ 'MRC_5_PD52_2' }	58.746
{ 'MRC_5_PD52_3' }	59.076
{ 'MRC_5_PD62_1' }	58.086
{ 'MRC_5_PD62_2' }	60.396
{ 'MRC_5_PD62_3' }	59.736

The results are shown according to the different samples, but we can see that the values are around 60%, which is not bad data, but we expected better.

Save the results table in a csv file

```
% Save as CSV
writetable(results_FPKM, 'results.csv');
```

```
% Create a table with the results
f = figure('Name', 'Results', 'NumberTitle', 'off');
t = uitable(f, 'Data', table2cell(results), 'ColumnName',
results_FPKM.Properties.VariableNames);
t.Position = [20 20 f.Position(3)-40 f.Position(4)-40];
```

	Column	PercentageHK
1	BJ_Y1	57.7558
2	BJ_Y2	57.7558
3	BJ_Y3	57.4257
4	BJ_OL...	59.0759
5	BJ_OL...	58.7459
6	BJ_OL...	58.0858
7	IMR90_Y1	60.7261
8	IMR90_Y2	60.7261
9	IMR90_Y3	60.3960
10	IMR90_...	60.3960
11	IMR90_...	58.7459
12	IMR90_...	61.0561
13	WI_38_Y1	63.0363
14	WI_38_Y2	62.3762
15	WI_38_Y3	63.0363
16	WI_38_...	60.7261
17	WI_38_...	61.3861
18	WI_38_...	60.7261
19	HFF_P...	60.7261
20	HFF_P...	60.3960

With TPM

```
clear all;
close all;
```

Load the data again

Load the transcriptomics data, the housekeeping genes obtained in the previous script, and the metabolic model (Human1 version).

```
% Transcriptomics data
data = readtable('Mod_data.xlsx');
% Housekeeping genes with the ensembl ids
h_k_g = readtable('housekeeping_ens.csv');
% Human1 metabolic model.
model = readCbModel('Human-GEM_Cobra_v1.01.mat');
```

Each `model.subSystems{x}` is a cell array, allowing more than one subSystem per reaction.

Convert the data from FPKM to TPM

It was decided to explore new alternatives for the normalisation of gene expression data by switching from FPKM (Fragments Per Kilobase Million) to TPM (Transcripts Per Million), which has been shown in other studies to facilitate comparison between different samples.

```
% Extract matrix from table
data_matrix = data{:, 2:end};
```

```
% Calculate the sum of each column
column_sums = sum(data_matrix, 1);

% Normalize each element by the sum of its column and then multiply by 10^6
normalized_matrix = (data_matrix ./ column_sums) * 1e6;

% Convert the normalized matrix back to a table
normalized_table = array2table(normalized_matrix, 'VariableNames',
data.Properties.VariableNames(2:end));

% Replace the relevant columns in the original table
data(:, 2:end) = normalized_table;
data
```

data = 26269x49 table

	Ensembl_GeneID	BJ_Y1	BJ_Y2	BJ_Y3	BJ_OLD_1	BJ_OLD_2	BJ_OLD_3
1	'ENSG000000000003'	25.3088	26.0030	28.0444	37.7252	35.9194	37.7420
2	'ENSG000000000005'	0	0	0	0	0	0
3	'ENSG000000000419'	94.4859	82.3190	96.0804	84.8942	88.6313	87.7934
4	'ENSG000000000457'	9.4143	9.3073	9.5616	7.7193	9.0525	8.5142
5	'ENSG000000000460'	7.6922	7.7543	8.5958	3.7268	3.9512	4.0041
6	'ENSG000000000938'	0.1637	0.1740	0.1062	0.2074	0.2095	0.1402
7	'ENSG000000000971'	21.9935	21.6351	20.6772	50.9410	49.8125	51.4614
8	'ENSG000000001036'	142.0617	132.8084	135.8053	152.0724	152.6952	154.1160
9	'ENSG000000001084'	5.8945	6.0876	6.8073	7.7551	8.3217	8.7123
10	'ENSG000000001167'	29.6603	33.4479	31.9516	20.7120	22.5003	22.6203
11	'ENSG000000001460'	6.7398	7.2518	6.7116	5.4596	5.9518	5.8749
12	'ENSG000000001461'	45.7646	47.3714	48.4922	87.0448	88.0520	88.6751
13	'ENSG000000001497'	24.4295	23.7401	23.3934	21.9430	21.7997	22.6502
14	'ENSG000000001561'	6.4387	6.3997	6.0950	17.4908	16.3895	16.3881

⋮

Metabolic genes

Pick the genes related to metabolism

In the transcriptomics dataset there are a lot of genes, but we are interested just in those ones that take part in the metabolism, so, we are going to compare the ensembl id of both, the transcriptomics dataset and the Human1 metabolic model, and pick those ones that are present.

```
% Create a vector just with the ensembl ids of the genes from the model
model_genes = model.genes;
% Find indexes of the genes that are present in the metabolic model.
```



```

index_names = ismember(data.Ensembl_GeneID, model_genes);
% take the rows of the dataset that match the indexes obtained before, with all the
data
data_met = data(index_names, :);

```

For loop to calculate the percentage of all the samples

```

% Initialise a table to store the results
results = table;
% Get expression data column names
expression_col = data_met.Properties.VariableNames(2:end);
results_colname = []

```

```

results_colname =

[]

```

```

results_percentage_hk = []

```

```

results_percentage_hk =

[]

```

now create the for loop where a threshold will be established based on the 70th percentile of gene expression for each sample, with genes above this threshold being considered core genes. Moreover, the accuracy will also be calculated on the basis of the housekeeping genes expressed as core genes divided by those related to metabolism.

```

% Iterate over each column of expression data
for i = 1:length(expression_col)
    % Select the expression data column
    col_name = expression_col{i};
    expfibroblast = data_met(:,1);
    expfibroblast(:,2) = data_met(:, col_name);
    expfibroblast.Properties.VariableNames{1} = 'gene';
    expfibroblast.Properties.VariableNames{2} = 'FPKMvalue';

    % Perform the calculations
    expfibroblast.logFPKMvalue = log10(expfibroblast{:,2} + 1);
    expfibroblast.value = expfibroblast.logFPKMvalue -
min(expfibroblast.logFPKMvalue);

    % Calculate the threshold, creating a slider form 70 to 100
    percentage = 90;
    up_threshold_fib = prctile(expfibroblast.value, percentage);

    % Filtering the data
    idx = expfibroblast.value >= up_threshold_fib;
    filtered_data = expfibroblast(idx, :);

    % Select the ensembl id of the core genes
    genes_table = data_met.Ensembl_GeneID;

```

```

%% housekeeping
% See how many housekeeping genes in the list have metabolic functions.
index_names = ismember(genes_table, h_k_g.converted_alias);
hkg_met = data_met(index_names, :);

% Now just the ensembl id
hkg_met_ens = hkg_met(:, "Ensembl_GeneID");

% Finding the housekeeping genes among the filtered genes
ens_filt_dat = filtered_data.gene;
matching_names_idx = ismember(ens_filt_dat, hkg_met_ens.Ensembl_GeneID);
matching_names = filtered_data(matching_names_idx, 1);

% Calculate the percentage of maintenance genes
percentage_hk = (length(matching_names.gene) / length(ens_filt_dat)) * 100;

% Save the results in the table
results{i, 'Column'} = {col_name};
results{i, 'PorcentageHK'} = percentage_hk;

```

end

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

Warning: The assignment added rows to the table, but did not assign values to all of the table's existing variables. Those variables are extended with rows containing default values.

[illegible]

```
results_TPM = results
```

```
results_TPM = 48x2 table
```

	Column	PercentageHK
1	'BJ_Y1'	57.7558
2	'BJ_Y2'	57.7558
3	'BJ_Y3'	57.4257
4	'BJ_OLD_1'	59.0759
5	'BJ_OLD_2'	58.7459
6	'BJ_OLD_3'	58.0858
7	'IMR90_Y1'	60.7261
8	'IMR90_Y2'	60.7261
9	'IMR90_Y3'	60.3960
10	'IMR90_O1'	60.3960
11	'IMR90_O2'	58.7459
12	'IMR90_O3'	61.0561
13	'WI_38_Y1'	63.0363
14	'WI_38_Y2'	62.3762

⋮

```
% Show results
disp(results_TPM);
```

Column	PercentageHK
{ 'BJ_Y1' }	57.756
{ 'BJ_Y2' }	57.756
{ 'BJ_Y3' }	57.426
{ 'BJ_OLD_1' }	59.076
{ 'BJ_OLD_2' }	58.746
{ 'BJ_OLD_3' }	58.086
{ 'IMR90_Y1' }	60.726
{ 'IMR90_Y2' }	60.726
{ 'IMR90_Y3' }	60.396
{ 'IMR90_O1' }	60.396
{ 'IMR90_O2' }	58.746
{ 'IMR90_O3' }	61.056
{ 'WI_38_Y1' }	63.036
{ 'WI_38_Y2' }	62.376
{ 'WI_38_Y3' }	63.036
{ 'WI_38_O1' }	60.726
{ 'WI_38_O2' }	61.386
{ 'WI_38_O3' }	60.726
{ 'HFF_PD16_1' }	60.726
{ 'HFF_PD16_2' }	60.396
{ 'HFF_PD16_3' }	60.066
{ 'HFF_PD74_1' }	59.406
{ 'HFF_PD74_2' }	59.076
{ 'HFF_PD74_3' }	58.746
{ 'MRC_5_PD32_1' }	57.756

{'MRC_5_PD32_2'}	58.086
{'MRC_5_PD32_3'}	58.746
{'MRC_5_PD72_1'}	62.706
{'MRC_5_PD72_2'}	62.706
{'MRC_5_PD72_3'}	62.706
{'HFF_PD26_1' }	58.746
{'HFF_PD26_2' }	58.416
{'HFF_PD26_3' }	58.086
{'HFF_PD46_1' }	59.076
{'HFF_PD46_2' }	58.416
{'HFF_PD46_3' }	62.706
{'HFF_PD64_1' }	57.756
{'HFF_PD64_2' }	58.746
{'HFF_PD64_3' }	59.076
{'MRC_5_PD42_1'}	59.736
{'MRC_5_PD42_2'}	59.736
{'MRC_5_PD42_3'}	59.736
{'MRC_5_PD52_1'}	59.406
{'MRC_5_PD52_2'}	58.746
{'MRC_5_PD52_3'}	59.076
{'MRC_5_PD62_1'}	58.086
{'MRC_5_PD62_2'}	60.396
{'MRC_5_PD62_3'}	59.736

The results are the same as those obtained by analysis with FPKM data.

Save the results table in a csv file

```
% Save the table in a CSV file
writetable(results_TPM, 'results_TPM.csv');
```

```
% Create a table with the results
f = figure('Name', 'Results', 'NumberTitle', 'off');
t = uitable(f, 'Data', table2cell(results), 'ColumnName',
results_TPM.Properties.VariableNames);
t.Position = [20 20 f.Position(3)-40 f.Position(4)-40];
```

	Column	PercentageHK
1	BJ_Y1	57.7558
2	BJ_Y2	57.7558
3	BJ_Y3	57.4257
4	BJ_OL...	59.0759
5	BJ_OL...	58.7459
6	BJ_OL...	58.0858
7	IMR90_Y1	60.7261
8	IMR90_Y2	60.7261
9	IMR90_Y3	60.3960
10	IMR90_...	60.3960
11	IMR90_...	58.7459
12	IMR90_...	61.0561
13	WI_38_Y1	63.0363
14	WI_38_Y2	62.3762
15	WI_38_Y3	63.0363
16	WI_38_...	60.7261
17	WI_38_...	61.3861
18	WI_38_...	60.7261
19	HFF_P...	60.7261
20	HFF_P...	60.3960