Extraction of context-specific models

Authors: Thomas Pfau, Systems Biochemistry Group, University of Luxembourg - Anne Richelle, Systems Biology and Cell Engineering, University of California San Diego

Reviewer(s): Almut Heinken, Molecular Systems Physiology Group, LCSB, University of Luxembourg

INTRODUCTION

Genome-scale reconstruction of metabolism (GEM) can illuminate the molecular basis of cell phenotypes exhibited by an organism. Since some enzymes are only active in specific cell types or environmental conditions, several algorithms have been developed to extract context-specific models that capture the metabolism of individual tissues or cell types. Therefore, a context-specific model is a subset of the GEM, in which inactive reactions are removed. Reaction removal is determined by the algorithm used, gene expression levels, presence of proteins or metabolites, experimental data availability, literature knowledge, and/or predefined metabolic functions of the cell type that need to be maintained in the extracted model. These decisions on methodology and data processing significantly influence the size, functionality and accuracy of constructed context-specific models. While there is no strong evidence that one model extraction method (MEM) universally gives the most physiologically accurate models, each method has different underlying assumptions that affect the resulting model. Therefore, selection of the MEM and the associated parameters should be done while considering the goals of the study and the available data [1].

Multiple algorithms have been suggested to automatically derive context specific networks from a generic reconstruction and a set of transcriptomic or proteomic data. The COBRA toolbox offers the following selection of extraction agorithms:

- FASTCORE [2] - Define one set of core reactions that is guaranteed to be active in the extracted model and find the minimum number of reactions possible to support the core;
- GIMME [3] - Minimize usage of low-expression reactions while keeping the objective (e.g., biomass) above a certain value. Does not favor inclusion of reactions not related to the objective;
- iMAT [4] - Find the optimal trade-off between including high-expression reactions and removing low-expression reactions;
- INIT [5] - Find the optimal trade-off between including and removing reactions based on their given weights. If desired, accumulation of certain metabolites can be allowed or even forced;
- MBA [6] - Define high-confidence reactions to ensure activity in the extracted model. Medium confidence reactions are only kept when a certain parsimony trade-off is met. In random order, prune other reactions and remove them if not required to support high- or medium- confidence reactions;
- mCADRE [7] - Define a set of core reactions and prune all other reactions based on their expression, connectivity to core and confidence score. Remove reactions not necessary to support the core or defined functionalities. Core reactions are only removed if supported by a certain number of zero-expression reactions.

For convenience, there is a common interface to use these algorithms provided in the createTissueSpecificModel method. The method can be called as follow:

```
tissueModel = createTissueSpecificModel(model, options)
```

PROCEDURE

If necessary, initialise The COBRA Toolbox

```
global TUTORIAL_INIT_CB;
if ~isempty(TUTORIAL_INIT_CB) && TUTORIAL_INIT_CB==1
    initCobraToolbox
end
changeCobraSolver ('glpk', 'all');
```

```
> glpk is compatible and fully tested with MATLAB R2016b on your operating system.
> Solver for LP problems has been set to glpk.
> glpk is compatible and fully tested with MATLAB R2016b on your operating system.
> Solver for MILP problems has been set to glpk.
> Solver glpk not supported for problems of type MIQP. Currently used: gurobi
> Solver glpk not supported for problems of type NLP. No solver set for this problemtype
> Solver glpk not supported for problems of type QP. Currently used: gurobi
```

Load the model that will be used for the extraction. For this tutorial, we have choosen to use E. coli core model as example. Please download the model from http://systemsbiology.ucsd.edu/sites/default/files/Attachments/Images/downloads/Ecoli_core/ecoli_core_model.mat and save it in your preferred folder.

```
modelFileName = 'ecoli_core_model.mat';
modelDirectory = getDistributedModelFolder(modelFileName); %Look up the folder for the distributed Models.
modelFileName= [modelDirectory filesep modelFileName]; % Get the full path. Necessary to be sure, that the right model is loaded
model = readCbModel(modelFileName);
```

Load the expression data that will be used for the extraction. For this tutorial, we have choosen to use E. coli Microarray-based gene expression data (downloaded from http://systemsbiology.ucsd.edu/InSilicoOrganisms/Ecoli/EcoliExpression2) related to a anaerobic growth on glucose of a wild-type E. coli strain.

```
load('dataEcoli');
```

Depending on the method that will be used for the extraction, different options need to be provided that could depend on a preprocessing step of the gene expression data. Since preprocessing is data dependent and the data used can involve multiple sources, we provide preprocessed data for all the methods. The section "CRITICAL STEP" below briefly explains how the preprocessing has been done for this tutorial and provides some guidelines for the users on how to preprocess their data.

Load the preprocessed data associated with the extraction method. Select one of the following sections to select the algorithm and data you want to use.

For the iMAT method, the available parameter options are (with all options marked with * being optional):

- **options.solver** : 'iMAT'
- **options.expressionRxns** : gene expression data corresponding to model.rxns. Note : If no gene expression data are available for a reaction, set the value to -1
- **options.threshold_lb** : lower bound of gene expression threshold, reactions with expression below this value are "non-expressed"
- **options.threshold_ub** : upper bound of gene expression threshold, reactions with expression above this value are "expressed"
- **options.tol*** : minimum flux value for "expressed" reactions (default - 1e-8)
- **options.core*** : list of reaction names that are associated with a high confidence (default - no core reactions)
- **options.logfile*** : name of the file to save the MILP log (defaut - 'MILPlog')
- **options.runtime*** : maximum solve time for the MILP (default - 7200s)

```
options = 'options_iMAT';
```

Now, load the preprocessed data

```
load(['options_methods' filesep options]);
```

Call createTissueSpecificModel to extract your model

```
iMAT_model = createTissueSpecificModel(model, options);
```

```
GLPK Integer Optimizer, v4.42
170 rows, 175 columns, 547 non-zeros
80 integer variables, all of which are binary
Preprocessing...
6 constraint coefficient(s) were reduced
134 rows, 125 columns, 445 non-zeros
57 integer variables, all of which are binary
Scaling...
 A: min|aij| =  7.090e-02  max|aij| =  1.001e+03  ratio =  1.412e+04
GM: min|aij| =  1.836e-01  max|aij| =  5.448e+00  ratio =  2.968e+01
EQ: min|aij| =  3.406e-02  max|aij| =  1.000e+00  ratio =  2.936e+01
2N: min|aij| =  3.125e-02  max|aij| =  1.000e+00  ratio =  3.200e+01
Constructing initial basis...
Size of triangular part = 126
Solving LP relaxation...
GLPK Simplex Optimizer, v4.42
134 rows, 125 columns, 445 non-zeros
      0: obj =  -8.732017982e+02  infeas =  8.531e+04 (8)
*    47: obj =   4.019237671e+01  infeas =  8.235e-16 (5)
*    82: obj =   5.894902997e+01  infeas =  4.996e-16 (5)
OPTIMAL SOLUTION FOUND
Integer optimization begins...
+    82: mip =     not found yet <=              +inf        (1; 0)
+   502: >>>>>   4.100000000e+01 <=   4.800000000e+01  17.1% (54; 13)
+  2446: >>>>>   4.300000000e+01 <=   4.300000000e+01   0.0% (138; 169)
+  2446: mip =   4.300000000e+01 <=     tree is empty   0.0% (0; 611)
INTEGER OPTIMAL SOLUTION FOUND
```

For GIMME method, the available parameter options are (with all options marked with * being optional):

- **options.solver** : 'GIMME'
- **options.expressionRxns** : gene expression data corresponding to model.rxns. Note : If no gene expression data are available for a reaction, set the value to -1
- **options.threshold** : gene expression threshold, reactions below this are minimized
- **options.obj_frac*** : minimum fraction of the model objective function that need to be maintained in the extracted model (default - 0.9)

The code to create the tissue-specific model is as for iMAT.

```
options = 'options_GIMME';
load(['options_methods' filesep options]);
GIMME_model = createTissueSpecificModel(model, options);
```

For MBA method, the available parameter options are (with all options marked with * being optional):

- **options.solver** : 'MBA'
- **options.medium_set** : list of reaction names with medium confidence
- **options.high_set** : list of reaction names with high confidence
- **options.tol*** : minimum flux value for "expressed" reactions (default - 1e-8)

```
options = 'options_MBA';
load(['options_methods' filesep options]);
MBA_model = createTissueSpecificModel(model, options);
```

```
fastcc.m: The input model is entirely flux consistent.\n
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
```

For FASTCORE method, the available parameter options are (with all options marked with * being optional):

- **options.solver** : 'fastCore'
- **options.core** : indices of reactions in the model that are part of the core set of reactions
- **options.epsilon*** : smallest flux value that is considered nonzero (default - 1e-4)
- **options.printLevel*** : 0 = silent, 1 = summary, 2 = debug (default - 0)

```
options = 'options_fastCore';
load(['options_methods' filesep options]);
FastCore_model = createTissueSpecificModel(model, options);
```

For mCADRE method, the available parameter options are (with all options marked with * being optional):

- **options.solver** : 'mCADRE'
- **options.ubiquityScore** : ubiquity scores, vector of the size of 'model.rxns' quantifying how often a gene is expressed accross samples
- **options.confidenceScores** : literature-based evidence for generic model, vector of the size of 'model.rxns'
- **options.protectedRxns*** : cell array with reactions names that are manually added to the core reaction set (default - no reactions)
- **options.checkFunctionality*** : Boolean variable that determines if the model should be able to produce the metabolites associated with the '*protectedRxns*' (0: don't use functionality check (default value) - 1: include functionality check)
- **options.eta*** : tradeoff between removing core and zero-expression reactions (default - 1/3)
- **options.tol*** : minimum flux value for "expressed" reactions (default - 1e-8)

```
options = 'options_mCADRE';
load(['options_methods' filesep options]);
mCADRE_model = createTissueSpecificModel(model, options);
```

```
Processing inputs and ranking reactions...
Generic model passed precursor metabolites test
Pruning reactions...
Reaction no. 1
Attempting to remove reaction O2t...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 11 (0 core, 11 non-core); Num. remaining: 75
Reaction no. 2
Attempting to remove reaction ACALDt...
Warning: No metabolites defined to check the model function
```

```
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 13 (0 core, 13 non-core); Num. remaining: 73
Reaction no. 3
Attempting to remove reaction EX_co2(e)...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 15 (0 core, 15 non-core); Num. remaining: 71
Reaction no. 4
Attempting to remove reaction EX_h2o(e)...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 17 (0 core, 17 non-core); Num. remaining: 69
Reaction no. 5
Attempting to remove reaction EX_etoh(e)...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 17 (0 core, 17 non-core); Num. remaining: 68
Reaction no. 6
Attempting to remove reaction EX_ac(e)...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 21 (0 core, 21 non-core); Num. remaining: 64
Reaction no. 7
Attempting to remove reaction EX_pyr(e)...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 23 (0 core, 23 non-core); Num. remaining: 62
Reaction no. 8
Attempting to remove reaction EX_succ(e)...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 30 (0 core, 30 non-core); Num. remaining: 55
Reaction no. 9
Attempting to remove reaction EX_lac_D(e)...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 33 (0 core, 33 non-core); Num. remaining: 52
Reaction no. 10
Attempting to remove reaction EX_glu_L(e)...
Warning: No metabolites defined to check the model function
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 33 (0 core, 33 non-core); Num. remaining: 51
Reaction no. 11
Attempting to remove reaction EX_akg(e)...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 33 (0 core, 33 non-core); Num. remaining: 50
Reaction no. 12
Attempting to remove reaction EX_glc(e)...
Warning: No metabolites defined to check the model function
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 33 (0 core, 33 non-core); Num. remaining: 49
Reaction no. 13
Attempting to remove reaction EX_pi(e)...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
```

```
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 33 (0 core, 33 non-core); Num. remaining: 48
Reaction no. 14
Attempting to remove reaction EX_nh4(e)...
Warning: No metabolites defined to check the model function
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 33 (0 core, 33 non-core); Num. remaining: 47
Reaction no. 15
Attempting to remove reaction EX_for(e)...
Warning: No metabolites defined to check the model function
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 33 (0 core, 33 non-core); Num. remaining: 46
Reaction no. 16
Attempting to remove reaction EX_h(e)...
Warning: No metabolites defined to check the model function
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 33 (0 core, 33 non-core); Num. remaining: 45
Reaction no. 17
Attempting to remove reaction ETOHt2r...
Warning: No metabolites defined to check the model function
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
0 = solution.stat
110 = solution.origStat
Warning: LP solution may not be optimal
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 33 (0 core, 33 non-core); Num. remaining: 44
Reaction no. 18
Attempting to remove reaction SUCCt3...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 35 (0 core, 35 non-core); Num. remaining: 42
Reaction no. 19
Attempting to remove reaction ATPM...
Warning: No metabolites defined to check the model function
fastcc.m: The input model is entirely flux consistent.\n
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 36 (0 core, 36 non-core); Num. remaining: 41
Reaction no. 20
Attempting to remove reaction Biomass_Ecoli_core_w_GAM...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 36 (0 core, 36 non-core); Num. remaining: 40
Reaction no. 21
Attempting to remove reaction RPE...
Warning: No metabolites defined to check the model function
```

```
fastcc.m: The input model is entirely flux consistent.\n
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 37 (0 core, 37 non-core); Num. remaining: 39
Reaction no. 22
Attempting to remove reaction GLNS...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 37 (0 core, 37 non-core); Num. remaining: 38
Reaction no. 23
Attempting to remove reaction SUCDi...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 39 (0 core, 39 non-core); Num. remaining: 36
Reaction no. 24
Attempting to remove reaction FORti...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 39 (0 core, 39 non-core); Num. remaining: 35
Reaction no. 25
Attempting to remove reaction FORt2...
Warning: No metabolites defined to check the model function
fastcc.m: The input model is entirely flux consistent.\n
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 40 (0 core, 40 non-core); Num. remaining: 34
Reaction no. 26
Attempting to remove reaction NH4t...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 40 (0 core, 40 non-core); Num. remaining: 33
Reaction no. 27
Attempting to remove reaction PFL...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 40 (0 core, 40 non-core); Num. remaining: 32
Reaction no. 28
Attempting to remove reaction GLCpts...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 40 (0 core, 40 non-core); Num. remaining: 31
Reaction no. 29
Attempting to remove reaction ACONTa...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 40 (0 core, 40 non-core); Num. remaining: 30
Reaction no. 30
Attempting to remove reaction ACONTb...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 40 (0 core, 40 non-core); Num. remaining: 29
Reaction no. 31
Attempting to remove reaction GLUN...
Warning: No metabolites defined to check the model function
fastcc.m: The input model is entirely flux consistent.\n
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 41 (0 core, 41 non-core); Num. remaining: 28
Reaction no. 32
Attempting to remove reaction ME2...
Warning: No metabolites defined to check the model function
fastcc.m: The input model is entirely flux consistent.\n
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 42 (0 core, 42 non-core); Num. remaining: 27
Reaction no. 33
Attempting to remove reaction GLUt2r...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 42 (0 core, 42 non-core); Num. remaining: 26
Reaction no. 34
Attempting to remove reaction PPCK...
Warning: No metabolites defined to check the model function
fastcc.m: The input model is entirely flux consistent.\n
Warning: No metabolites defined to check the model function
```

Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 43 (0 core, 43 non-core); Num. remaining: 25
Reaction no. 35
Attempting to remove reaction PPS...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 43 (0 core, 43 non-core); Num. remaining: 24
Reaction no. 36
Attempting to remove reaction RPI...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 43 (0 core, 43 non-core); Num. remaining: 23
Reaction no. 37
Attempting to remove reaction TKT1...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 43 (0 core, 43 non-core); Num. remaining: 22
Reaction no. 38
Attempting to remove reaction TKT2...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 43 (0 core, 43 non-core); Num. remaining: 21
Reaction no. 39
Attempting to remove reaction AKGt2r...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 43 (0 core, 43 non-core); Num. remaining: 20
Reaction no. 40
Attempting to remove reaction ME1...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 45 (0 core, 45 non-core); Num. remaining: 18
Reaction no. 41
Attempting to remove reaction FBP...
Warning: No metabolites defined to check the model function
fastcc.m: The input model is entirely flux consistent.\n
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 46 (0 core, 46 non-core); Num. remaining: 17
Reaction no. 42
Attempting to remove reaction PGL...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 46 (0 core, 46 non-core); Num. remaining: 16
Reaction no. 43
Attempting to remove reaction PIt2r...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 46 (0 core, 46 non-core); Num. remaining: 15
Reaction no. 44
Attempting to remove reaction PGM...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 46 (0 core, 46 non-core); Num. remaining: 14
Reaction no. 45
Attempting to remove reaction NADTRHD...
Warning: No metabolites defined to check the model function
fastcc.m: The input model is entirely flux consistent.\n
Warning: No metabolites defined to check the model function
Removed non-core inactive reactions
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 13
Reaction no. 46
Attempting to remove reaction THD2...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 12
Reaction no. 47
Attempting to remove reaction ATPS4r...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 11
Reaction no. 48

```
Attempting to remove reaction CS...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 10
Reaction no. 49
Attempting to remove reaction PDH...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 9
Reaction no. 50
Attempting to remove reaction G6PDH2r...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 8
Reaction no. 51
Attempting to remove reaction PFK...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 7
Reaction no. 52
Attempting to remove reaction PPC...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 6
Reaction no. 53
Attempting to remove reaction PGI...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 5
Reaction no. 54
Attempting to remove reaction PYK...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 4
Reaction no. 55
Attempting to remove reaction ICDHyr...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 3
Reaction no. 56
Attempting to remove reaction TALA...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 2
Reaction no. 57
Attempting to remove reaction FBA...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 1
Reaction no. 58
Attempting to remove reaction TPI...
Warning: No metabolites defined to check the model function
Warning: No metabolites defined to check the model function
No reactions removed
Num. removed: 47 (0 core, 47 non-core); Num. remaining: 0
```

For INIT method, the available parameter options are (with all options marked with * being optional):

- **options.solver** : 'INIT'
- **options.weights** : column with positive (high expression) and negative (low expression) weights for each reaction
- **options.tol*** : minimum flux value for "expressed" reactions (default - 1e-8)
- **options.logfile*** : name of the file to save the MILP log (defaut - 'MILPlog')
- **options.runtime*** : maximum solve time for the MILP (default - 7200s)

```
options = 'options_INIT';
load(['options_methods' filesep options]);
INIT_model = createTissueSpecificModel(model, options);
```

```
GLPK Integer Optimizer, v4.42
198 rows, 181 columns, 590 non-zeros
86 integer variables, all of which are binary
Preprocessing...
6 constraint coefficient(s) were reduced
144 rows, 136 columns, 471 non-zeros
69 integer variables, all of which are binary
Scaling...
 A: min|aij| =  7.090e-02  max|aij| =  1.001e+03  ratio =  1.412e+04
GM: min|aij| =  1.836e-01  max|aij| =  5.448e+00  ratio =  2.968e+01
EQ: min|aij| =  3.406e-02  max|aij| =  1.000e+00  ratio =  2.936e+01
2N: min|aij| =  3.125e-02  max|aij| =  1.250e+00  ratio =  4.000e+01
Constructing initial basis...
Size of triangular part = 134
Solving LP relaxation...
GLPK Simplex Optimizer, v4.42
144 rows, 136 columns, 471 non-zeros
      0: obj =   5.648220556e+02  infeas =  1.019e+05 (10)
*    60: obj =   5.614599110e+01  infeas =  1.029e-14 (5)
*   100: obj =   7.458402071e+01  infeas =  5.551e-16 (5)
OPTIMAL SOLUTION FOUND
Integer optimization begins...
+   100: mip =     not found yet <=              +inf        (1; 0)
+   765: >>>>>   4.926985032e+01 <=   6.489465815e+01  31.7% (88; 17)
+  3373: >>>>>   4.987940265e+01 <=   5.581258238e+01  11.9% (217; 297)
+  3780: >>>>>   5.106322395e+01 <=   5.531164139e+01   8.3% (256; 353)
+  3940: >>>>>   5.331563786e+01 <=   5.493999008e+01   3.0% (238; 428)
+  4001: mip =   5.331563786e+01 <=     tree is empty   0.0% (0; 1257)
INTEGER OPTIMAL SOLUTION FOUND
```

Note that additional options are available when extracting your model (`funcModel` and `exRxnRemove`). `funcModel` allows to define whether the extracted model will be/ or not flux consistent (i.e., an extracted model that contains only reactions that can carry fluxes). `exRxnRemove` allows to predefine a list of reactions that will be automatically removed in the extracted model. Example, if you want to extract a consistent model but you do not have any predefined list of reactions to remove, your call will be:

```
funcModel=1;
exRxnRemove={};
tissueModel = createTissueSpecificModel(model, options,funcModel,exRxnRemove);
```

CRITICAL STEPS

Gene Expression Preprocessing

When integrating transcriptomic data, the selection of options related to each method is critical and algorithmic performance often strongly depends on these choices. In GIMME [3], a fixed threshold value was used to distinguish between unexpressed and expressed reactions (threshold = 12 of log2 expression). For iMAT [4], the authors used a data dependent cutoff of half a standard deviation above the mean for active reactions, and half a standard deviation below for inactive reactions. In contrast to the selection by automated processing of expression data, the authors of FASTCORE [2] and MBA [5] used manually assembled sets of high confidence reactions along with data derived information as input. In INIT [5], the selection is based on the data in the Human Protein Atlas [8] using the indication from the proteomics data. For mCADRE [7], proprietary software was used to obtain presence/absence calls.

In the context of this tutorial, the options have been defined following the original paper for iMAT and GIMME and arbitrary preprocessing as used in [1] has been done for the four other methods.

The quality of the results also strongly depends on data preprocessing, and the COBRA toolbox does not provide an automatic preprocessing pipeline to derive expression values from raw measurements as multipe methods can be used. We therefore strongly suggest, that all preprocessing and discretization is performed prior to the call to `createTissueSpecificModel`.

Gene to Reaction Mapping

As with Preprocessing, we intentionally left the gene to reaction mapping separate from the tissue specific model extraction. The most common approach to obtain reaction expression values from geen expression values is to evaluate the Gene-Protein-Reaction rules associated with each reaction as follows:

Replace 'and' bin 'min' and 'or' by 'max'. E.g. for a reaction R1 with GPR 'A and (B or C)' with expression of Genes A = 2, B = 7 and C = 0, 'B or C' would evaluate to 7 and the whole would evaluate to 2.

This type of mapping can be applied to iMAT and GIMME, and is provided as a function in the toolbox:

```
model = getDistributedModel('ecoli_core_model.mat')
load('dataEcoli');
[expressionRxns parsedGPR] = mapExpressionToReactions(model, expression);
```

INIT and mCADRE use similar approaches, but require the expression to be protein expression, or ubiquityscores respectively, i.e. the function can be used for them, but the inputs need to be adjusted according to the required data for INIT and mCADRE.

FastCore and MBA again, rely on clearly defined core sets of reactions, the quality of which strongly influences the resulting models.

TIMING

TIMING: 15 minutes to hours (computation) - days (interpretation)

ANTICIPATED RESULTS

The size of extracted models varies depending of the MEM used

```
expected_results = {'ecoli_core_model',95,72;...
                    'FastCore_model',45,48;...
                    'GIMME_model',74,69;...
                    'INIT_model',25,32;...
                    'MBA_model',24,41;...
                    'iMAT_model',59,58;...
                    'mCADRE_model',48,49};
```

```
common_reactions_obtained = model.rxns;
obtained_results = cell(size(expected_results,1),2);
obtained_results(1,:) = {numel(model.rxns),numel(model.mets)};
for i = 2:size(expected_results,1)
    obtained_results(i,:) = {eval(['numel(' expected_results{i,1} '.rxns)']),eval(['numel(' expected_results{i,1} '.mets)'])};
    common_reactions_obtained = eval(['intersect(common_reactions_obtained,' expected_results{i,1} '.rxns)']);
end

%Set the result table
result_table = table(expected_results(:,1),expected_results(:,2),obtained_results(:,1), expected_results(:,3),obtained_results(:,2),...
                    'VariableNames',{'ModelName','Reactions_expected','Reactions_obtained','Metabolited_expected','Metabolited_obtained'});

%And display it, any values that do not match indicate some form of problem.
disp(result_table)
```

14 reactions are common to the 6 extracted models:

```
%These are the expected reactions
common_reacs = {'ACALD';'ALCD2x';'ATPS4r';'ENO';'FBA';'GAPD';'GLUDy';'GLUSy';'PDH';'PFK';'PGI';'PGK';'PYK';'TPI'}

%We will test if those are present and inform if not.
if isempty(setxor(common_reacs,common_reactions_obtained))
    fprintf('The expected common reacs where obtained:\n');
    printRxnFormula(model,common_reacs);
else
    not_in_all = setdiff(common_reacs,common_reactions_obtained);
    not_in_expected = setdiff(common_reactions_obtained,common_reacs);
    if ~isempty(not_in_all)
        fprintf('The following reactions were not in all models but should be contained in them:\n');
        printRxnFormula(model,not_in_all);
    end
    if ~isempty(not_in_expected)
        fprintf('The following reactions were unexpectedly present in all tissue models:\n');
        printRxnFormula(model,not_in_expected);
    end
end
```

REFERENCES

1. Opdam, S,, Richelle, A., Kellman, B., Li, S., Zielinski, D.C., Lewis, N.E. A systematic evaluation of methods for tailoring genome-scale models. Cell Syetms, 4:1-12 (2017).

2. Vlassis, N., Pacheco, M.P., and Sauter, T. Fast reconstruction of compact context-specific metabolic network models. PLoS Comput. Biol. 10, e1003424 (2014).

3. Becker, S.A., and Palsson, B.O. Context-specific metabolic networks are consistent with experiments. PLoS Comput. Biol. 4, e1000082 (2008).

4. Zur, H., Ruppin, E., and Shlomi, T. iMAT: an integrative metabolic anal- ysis tool. Bioinformatics 26, 3140–3142 (2010).

5.Agren, R. et al. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. PLoS Comput. Biol. 8, e1002518 (2012).

6. Jerby, L., Shlomi, T., and Ruppin, E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Mol. Syst. Biol. 6, 401 (2010).

7. Wang, Y., Eddy, J.A., and Price, N.D. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC Syst. Biol. 6, 153 (2012).

8. Uhlén M, et al. Tissue-based map of the human proteome. Science, 347(6220):1260419 (2015).