

Extraction of context-specific models via XomicsToModel

Author: German Preciat, Analytical BioSciences, Leiden University

Reviewer(s): Ronan M.T. Fleming, School of Medicine, University of Galway

INTRODUCTION

The XomicsToModel pipeline¹ of the COBRA Toolbox v3.4², facilitates the generation a thermodynamic-flux-consistent, context-specific, genome-scale metabolic model in a single command by combining a generic model with bibliomic, transcriptomic, proteomic, and metabolomic data. To ensure the network's quality, several thermodynamic consistency checks are implemented within the function. To generate a thermodynamic-flux-consistent, context-specific, genome-scale metabolic model, the function requires three inputs: a generic COBRA model and two variables containing the context-specific data and technical information defined by the user.

This tutorial shows how to extract a context-specific genome-scale model of a dopaminergic neuron (*iDopaNeuroC*³) from the human generic model Recon3D⁴. The *iDopaNeuroC*³ model is extracted using data from manual curation of a dopaminergic neuron to identify active and inactive genes, reactions, and metabolites, as well as information from in vitro experiments such as exometabolomic quantification and transcriptomic sequencing of a cell culture of pluripotent stem cell-derived dopaminergic neurons.

PROCEDURE

Install MATLAB and then the COBRA Toolbox as described here: <https://opencobra.github.io/cobratoolbox/stable/installation.html>

Select a solver suitable for solving linear (LP) and quadratic (QP) optimisation problems, e.g., mosek, gurobi, ibm_cplex, etc.

```
[~, ~] = changeCobraSolver('mosek', 'all', 0);
```

Generic model

The COBRA model Recon3D⁴, representing human metabolic reconstruction, can be found in a file with the extension ".mat". Recon3D¹, which is found in the VMH database², can be used as a generic model for human metabolism. The thermodynamic consistent human metabolic reconstruction Recon3D⁴.

```
inputFolder = [ '~' filesep 'work' filesep 'sbgCloud' filesep
    'programExperimental' ...
    filesep 'projects' filesep 'xomics' filesep 'data' filesep
    'Recon3D_301' ];
genericmodelName = 'Recon3DModel_301_xomics_input.mat';
load([inputFolder filesep genericmodelName])
```

Context-specific data

This type of information represents the biological system's phenotype and can be obtained through a review of the literature or experimental data derived from the biological system. The context-specific data can be loaded from a spreadsheet or added manually.

Automated data integration

Tables or multiple data sets can be inserted in an external worksheet document so that the `preprocessingOmicsModel` function can include them in the `options` variable. The name of the sheet corresponding to the `options` field must be the same as those specified above and in the manuscript, or they will be omitted.

Bibliomic data. It is derived from manual reconstruction following a review of the literature. This includes data on the activation or inactivation of genes, reactions, or metabolites. Another example is the addition of coupled reactions or the constraints of different reactions based on phenotypic observations.

- `specificData.activeGenes`: List of Entrez ID of genes that are known to be active based on the bibliomic data (Default: empty).
- `specificData.addCoupledRxns`: Logical, should the coupled constraints be added (Default: true).
- `specificData.coupledRxns`: Logical, indicates whether curated data should take priority over omics data (Default: false).
- `specificData.essentialAA`: List exchange reactions of essential amino acid (Default: empty).
- `specificData.inactiveGenes`: List of Entrez ID of genes known to be inactive based on the bibliomics data (Default: empty).
- `specificData.presentMetabolites`: List of metabolites known to be active based on the bibliomics data (Default: empty).
- `specificData.rxns2add`: Table containing the identifier of the reaction to be added, its name, the reaction formula, the metabolic pathway to which it belongs, the gene rules to which the reaction is subject, and the references. (Default: empty).
- `specificData.rxns2constraint`: Table containing the reaction identifier, the updated lower bound, the updated upper bound, a description for the constraint and notes such as references or special cases (Default: empty).

Manually curated data. To read the table and prepare the variable `specificData` it is used the function `preprocessingOmicsModel`. In this tutorial the bibliomic data is contained in the file '`bibliomicData.xlsx`'.

```
dataFolder = [fileparts(which('tutorial_XomicsToModel.mlx')) filesep  
'iDopaNeuro' filesep 'data' filesep];  
bibliomicData = 'bibliomicData.xlsx';  
specificData = preprocessingOmicsModel([dataFolder bibliomicData], 1, 1);
```

```
Reading inputData from : /home/rfleming/work/sbgCloud/programExperimental/projects/xomics/code/tutorials/i  
Reading sheet: activeGenes  
Reading sheet: activeReactions  
Reading sheet: cellCultureData  
Reading sheet: coupledRxns  
Reading sheet: essentialAA  
Reading sheet: inactiveGenes  
Reading sheet: mediaData
```

```
Reading sheet: presentMetabolites
Reading sheet: rxns2add
Reading sheet: rxns2constrain
Reading sheet: rxnsHypothesis
Reading sheet: rxns2remove
Reading sheet: sinkDemand
```

Metabolomic data. Differences in measured concentrations of metabolites within cells, biofluids, tissues, or organisms are translated into flux units of flux ($\mu\text{mol/gDW/h}$).

- **specificData.cellCultureData:** Table containing the cell culture data used to calculate the uptake flux. Includes well volume (L), time interval cultured (hr), average protein concentration (g/L), assay volume (L), protein fraction (g/g dry weight) , and the sign for uptakes (Default: -1).
- **specificData.exoMet:** Table with the fluxes obtained from exometabolomics experiments. It includes the reaction identifier, the reaction name, the measured mean flux, standard deviation of the measured flux, the flux units, and the platform used to measure it.
- **specificData.mediaData:** Table containing the initial media concentrations. Contains the reaction identifier, the maximum uptake ($\mu\text{mol/gDW/h}$) based on the concentration of the metabolite and the concentration (μmol ; Default: empty).

In this tutorial the exometabolomic data is saved in the table 'exoMet'.

```
specificData.exoMet = readtable([dataFolder 'exometabolomicData.txt']);
```

Proteomic data. This information indicates the level of expression of the proteome.

- **specificData.proteomics:** Table with a column with Entrez ID's and a column for the corresponding protein levels (Default: empty).

For this tutorial no proteomic data was used.

Transcriptomic data. Indicates the level of transcriptome expression and can also be used to calculate reaction expression. Transcriptomic data can be analysed in FPKM.

- **specificData.transcriptomicData:** Table with a column with Entrez ID's and a column for the corresponding transcriptomics expresion value (Default: empty).

In this tutorial the transcriptomic analysis is saved in the table 'transcriptomicData'.

```
specificData.transcriptomicData = readtable([dataFolder
'transcriptomicData.txt']);
specificData.transcriptomicData.genes =
string(specificData.transcriptomicData.genes);
```

Technical parameters

With these options, technical constraints can be added to the model, as well as setting the parameters for model extraction or debugging.

Bounds. They are the instructions that will be set in the boundaries.

- **param.boundPrecisionLimit:** Precision of flux estimate, if the absolute value of the lower bound (`model.lb`) or the upper bound (`model.ub`) are lower than `options.boundPrecisionLimit` but higher than 0 the value will be set to the `boundPrecisionLimit` (Default: primal feasibility tolerance).
- **param.TolMaxBoundary:** The reaction boundary's maximum value (Default: `1e3`).
- **param.TolMinBoundary:** The reaction boundary's minimum value (Default: `-1e3`).
- **param.relaxOptions:** A structure array with the relaxation options (Default: `param.relaxOptions.steadyStateRelax = 0`).

```
param.TolMinBoundary = -1e4;
param.TolMaxBoundary = 1e4;
feasTol = getCobraSolverParams('LP', 'feasTol');
param.boundPrecisionLimit = feasTol * 10;
```

Exchange reactions. They are the instructions for the exchange, demand, and sink reactions.

- **param.addSinksexoMet:** Logical, should sink reactions be added for metabolites measured in the media but without existing exchange reaction (Default: false).
- **param.closeIons:** Logical, it determines whether or not ion exchange reactions are closed. (Default: false).
- **param.closeUptakes:** Logical, decide whether or not all of the uptakes in the generic model will be closed (Default: false).
- **param.nonCoreSinksDemands:** The type of sink or demand reaction to close is indicated by a string (Possible options: 'closeReversible', 'closeForward', 'closeReverse', 'closeAll' and 'closeNone'; Default: 'closeNone').

```
param.closeIons = true;
param.closeUptakes = true;
param.nonCoreSinksDemands = 'closeAll';
param.sinkDMinactive = true;
```

Extraction options. The solver and parameters for extracting the context-specific model.

- **param.activeGenesApproach:** String with the name of the active genes approach will be used (Possible options: 'oneRxnsPerActiveGene' or 'deletModelGenes'; Default: 'oneRxnsPerActiveGene').
- **param.fluxCCmethod:** String with the name of the algorithm to be used for the flux consistency check (Possible options: 'swiftcc', 'fastcc' or 'dc', Default: 'fastcc').
- **param.fluxEpsilon:** Minimum non-zero flux value accepted for tolerance (Default: Primal feasibility tolerance).
- **param.thermoFluxEpsilon:** Flux epsilon used in 'thermoKernel' (Default: feasibility tolerance).
- **param.tissueSpecificSolver:** The name of the solver to be used to extract the context-specific model (Possible options: 'thermoKernel' and 'fastcore'; Default: 'thermoKernel').

```

param.activeGenesApproach = 'oneRxnPerActiveGene';
param.tissueSpecificSolver = 'thermoKernel';
param.fluxEpsilon = feasTol * 10;
param.fluxCCmethod = 'fastcc';

```

Data-specific parameters. Parameters that define the minimum level of transcript/protein to be considered as present in the network (threshold) and whether the transcripts below the set threshold should be removed from the model.

- **param.addCoupledRxns:** Logical, should the coupled constraints be added (Default: false). !
CAUTION If it is TRUE and the table coupledRxns is empty, the step is not performed.
- **param.curationOverOmics:** Logical, indicates whether curated data should take priority over omics data (Default: false).
- **param.inactiveGenesTranscriptomics:** Logical, indicate if inactive genes in the transcriptomic analysis should be added to the list of inactive genes (Default: true).
- **param.metabolomicWeights:** String indicating the type of weights to be applied for metabolomics fitting (Possible options: 'SD', 'mean' and 'RSD'; Default: 'SD').
- **param.setObjective:** Linear objective function to optimise (Default: empty).
- **param.thresholdP:** The proteomic cutoff threshold (in linear scale) for determining whether or not a gene is active (Default: 0).
- **param.transcriptomicThreshold:** The transcriptomic cutoff threshold (in logarithmic scale) for determining whether or not a gene is active (Default: 0)
- **param.weightsFromOmics:** Should gene weights be assigned based on the omics data (Default: 0).

```

param.addCoupledRxns = 1;
param.curationOverOmics = false;
param.inactiveGenesTranscriptomics = true;
param.metabolomicWeights='mean';
param.transcriptomicThreshold = 2;
param.weightsFromOmics = true;

```

Debugging options. The user can specify the function's verbosity level as well as save the results of the various blocks of the function for debugging.

- **param.debug:** Logical, should the function save its progress for debugging (Default: false).
- **param.diaryFilename:** Location where the output be printed in a diary file (Default: 0).
- **param.printLevel:** Level of verbose that should be printed (Default: 0).

```

param.printLevel = 1;
param.debug = true;
if isunix()
    name = getenv('USER');
else
    name = getenv('username');
end
param.diaryFilename = [pwd filesep datestr(now,30) '_' name '_diary.txt'];

```

XomicsToModel function

```
[iDopaNeuro1, modelGenerationReport] = XomicsToModel(model, specificData, param);
```

```
XomicsToModel run, beginning at:16-Dec-2022 08:56:14
```

```
-----
```

```
XomicsToModel input specificData:
```

```
    inputData: '/home/rfleming/work/sbgCloud/programExperimental/projects/xomics/code/tutorials/i
    activeGenes: {239x1 cell}
    activeReactions: {334x1 cell}
    cellCultureData: [1x6 table]
    coupledRxns: [11x5 table]
    essentialAA: [9x1 table]
    inactiveGenes: {61x1 cell}
    mediaData: [56x3 table]
    presentMetabolites: [45x4 table]
        rxns2add: [21x9 table]
    rxns2constrain: [48x5 table]
    rxnsHypothesis: [33x5 table]
        rxns2remove: [233x5 table]
        sinkDemand: [49x8 table]
        exoMet: [49x13 table]
    transcriptomicData: [18530x2 table]
```

```
XomicsToModel input param:
```

```
    TolMinBoundary: -10000
    TolMaxBoundary: 10000
    boundPrecisionLimit: 1e-05
        closeIons: 1
        closeUptakes: 1
    nonCoreSinksDemands: 'closeAll'
        sinkDMinactive: 1
    activeGenesApproach: 'oneRxnPerActiveGene'
    tissueSpecificSolver: 'thermoKernel'
        fluxEpsilon: 1e-05
        fluxCCmethod: 'fastcc'
        addCoupledRxns: 1
        curationOverOmics: 0
    inactiveGenesTranscriptomics: 1
        metabolomicWeights: 'mean'
    transcriptomicThreshold: 2
        weightsFromOmics: 1
        printLevel: 1
        debug: 1
        diaryFilename: '/home/rfleming/20221216T085613_rfleming_diary.txt'
    inactiveReactions: []
        thresholdP: 0
        uptakeSign: -1
        thermoFluxEpsilon: 1e-05
growthMediaBeforeReactionRemoval: 1
    metabolomicsBeforeExtraction: 1
        workingDirectory: '/home/rfleming'
findThermoConsistentFluxSubset: 1
    plotThermoKernelStats: 0
    plotThermoKernelWeights: 0
        finalFluxConsistency: 0
            relaxOptions: [1x1 struct]
        boundsToRelaxExoMet: 'both'
```

Replacing reaction name DM_atp_c_ with ATPM, because it is not strictly a demand reaction.

Old reaction formulas

```
ATPS4mi    adp[m] + pi[m] + 4 h[i]      ->    h2o[m] + 3 h[m] + atp[m]
CYOOm2i    o2[m] + 8 h[m] + 4 focyteC[m]  ->    2 h2o[m] + 4 focyteC[m] + 4 h[i]
CYOOm3i    o2[m] + 7.92 h[m] + 4 focyteC[m] ->    1.96 h2o[m] + 4 focyteC[m] + 0.02 o2s[m] + 4 h[i]
CYOR_u10mi   2 h[m] + 2 focyteC[m] + q10h2[m] ->    q10[m] + 2 focyteC[m] + 4 h[i]
NADH2_u10mi   5 h[m] + nadh[m] + q10[m]     ->    nad[m] + q10h2[m] + 4 h[i]
0x0 empty char array
```

New reaction formulas

```
ATPS4minew   4 h[c] + adp[m] + pi[m]      ->    h2o[m] + 3 h[m] + atp[m]
CYOOm2inew   o2[m] + 8 h[m] + 4 focyteC[m]  ->    2 h2o[m] + 4 h[c] + 4 focyteC[m]
CYOOm3inew   o2[m] + 7.92 h[m] + 4 focyteC[m] ->    1.96 h2o[m] + 4 h[c] + 4 focyteC[m] + 0.02 o2s[m]
CYOR_u10minew 2 h[m] + 2 focyteC[m] + q10h2[m] ->    4 h[c] + q10[m] + 2 focyteC[m]
NADH2_u10minew 5 h[m] + nadh[m] + q10[m]     ->    4 h[c] + nad[m] + q10h2[m]
```

Feasible generic input model.

Generating model without an objective function.

Adding 21 reactions ...

Reaction boundaries not provided. Default (min and max) values will be used based on the reaction formula.

```
acleua    h2o[c] + acleu_L[c]      ->    ac[c] + leu_L[c]
acthra    h2o[c] + acthr_L[c]      ->    ac[c] + thr_L[c]
acileua   h2o[c] + acile_L[c]      ->    ac[c] + ile_L[c]
acglua    h2o[c] + acglu[c]       ->    glu_L[c] + ac[c]
CE1554tm   CE1554[c]           <=>    CE1554[m]
RE2031M    accoa[m] + ala_L[m]    <=>    h[m] + coa[m] + CE1554[m]
RE2642C    h2o[c] + CE1554[c]    <=>    ac[c] + ala_L[c]
CE1554t    CE1554[c]           <=>    CE1554[e]
EX_CE1554[e]  CE1554[e]         <=>
DM_ps_hs[c]  ps_hs[c]          <=>
CYSTS_H2S   cys_L[c] + hcys_L[c]  <=>    HC00250[c] + cyst_L[c]
DHBOX     h2o2[c] + quinonemethide[c] ->    3,4-dihydroxybenzaldehyde[c] + methanimine[c]
DM_clpn_hs[c] clpn_hs[c]        ->
EX_adocbl[e]  adocbl[e]         <=>
EX_ca2[e]   ca2[e]            <=>
EX_cl[e]    cl[e]             <=>
EX_mg2[e]   mg2[e]            <=>
EX_selni[c] selni[c]          <=>
EX_zn2[e]   zn2[e]            <=>
NORCON    dopa[c] + fald[c] + 3,4-dihydroxybenzaldehyde[c] ->    CE2172[c]
Q-METHRED  fe2[c] + CE5276[c]  ->    quinonemethide[c]
0 deleted non-core metabolites, corresponding to generic model.
0 deleted core metabolites, corresponding to generic model.
```

Old model does not contain these core reactions (now removed from core reaction set):

```
{'DM_ca2[c]'      }
{'EX_HC01944[e]'  }
{'EX_adpcbl[e]'   }
{'Htmi'            }
{'RE1917C'         }
```

0 deleted non-core reactions, corresponding to generic model.

0 deleted core reactions, corresponding to generic model.

Identifying the stoichiometrically consistent subset...

```

--- findStoichConsistentSubset START ---
--- Summary of stoichiometric consistency ---
 5843    10620      totals.
   3     1818      heuristically external.
 5840    8802      heuristically internal:
 5840    8800      ... of which are stoichiometrically consistent.
   0      2      ... of which are stoichiometrically inconsistent.
   0      0      ... of which are of unknown consistency.
 5840    8800      Confirmed stoichiometrically consistent by leak/siphon testing.
--- findStoichConsistentSubset END ---
3 deleted non-core metabolites, corresponding to stoichiometric inconsistency.
0 deleted core metabolites, corresponding to stoichiometric inconsistency.
Old model does not contain these core reactions (now removed from core reaction set):
{'DM_ca2[c]'      }
{'EX_HC01944[e]'  }
{'EX_adpcbl[e]'   }
{'Htmi'           }
{'RE1917C'         }

```

0 deleted non-core reactions, corresponding to stoichiometric inconsistency.
 5 deleted core reactions, corresponding to stoichiometric inconsistency.

Reversible_Reaction	Name	lb	ub
{'RE2130C' }	{'RE2130C' }	-1000	1000
{'CYSTS_H2S' }	{'Cystathione Beta-Synthase (sulfide-forming)'}	-10000	10000
{'EX_adocbl[e]'}	{'Exchange of adenosylcobalamin'}	-10000	10000
{'EX_mg2[e]' }	{'Exchange of magnesium'}	-10000	10000
{'EX_selni[c]' }	{'Exchange of selenite'}	-10000	10000

3 stoichiometrically inconsistent metabolites removed.
 5 stoichiometrically inconsistent reactions removed.

Feasible stoichiometrically consistent model with new reactions.

Feasible model with default bounds.

Assuming gene expression is NaN for 160 genes where no transcriptomic data is provided.

Model statistics:

5840 x 10615 stoichiometric matrix.
 1566 exchange reactions.
 109 exchange reactions in the core reaction set.
 7 exchange reactions in the rxns2Constrain set.

1456 exchange reactions with uptake closed

239 closed non-core sink/demand reactions via param.nonCoreSinksDemands = closeAll
 10 core sink/demand reactions.
 10 open core sink/demand reactions.

Feasible after closing non-core sink/demand reactions.

Adding growth media information...
 The following reactions could not be constrained since they are not present in the model:
 {'EX_HC01944[e]'}

```
{'EX_adpcbl[e]' }
{'EX_mg2[e]' }
{'EX_selni[c]' }
```

Adding constraints on 52 reactions

These reactions have bounds larger than the recommended value = abs(10000)

The bounds for the following reactions have been adjusted:

Number of corrected bounds (to min/max boundary):

2

Reverse_Reaction, 0 bound	Name	lb_before	lb_after	ub_before
{'EX_na1[e]' }	{'Exchange of Sodium' }	-17903.1402873266	-10000	0
{'EX_cl[e]' }	{'Exchange of chloride ion' }	-14916.2499880923	-10000	0

Feasible after application of media constraints.

Adding quantitative metabolomics constraints ...

-0.621631949021564

Feasible after application of metabolomic constraints

Checking for mismatches ...

Adding custom constraints ...

tissueSpecificSolver = thermoKernel. Ignoring specificData.rxnns2constrain for demand reactions, i.e. with
Adding constraints on 38 reactions

Feasible after application of custom constraints.

Adding 11 sets of coupled reactions ...

coupledRxnId	constraints
{'Phosphatidylcholine' }	{'PCHOLP_hs + PLA2_2 + SMS >= 2.025'}
{'Adenosine Monophosphate' }	{'AMPDA + NTD7 >= 0.2265'}
{'Glutamate' }	{'- ALATA_L + GLUCYS + GLUDxm + GLUDym - ASPTA - ILETA - LEUTA - VALTA'}
{'Aspartate' }	{'ARGSS + ASPTA >= 1.1925'}
{'Serine' }	{'GHMT2r + r0060 >= 0.8625'}
{'Arginine' }	{'GLYAMDTRc + r0145 + ARGN >= 0.7245'}
{'Tyrosine' }	{'TYR3MO2 + TYRTA + HMR_6728 + HMR_6874 >= 0.55875'}
{'Histidine' }	{'HISDC + HISD >= 1.095'}
{'Leucine + Isoleucine' }	{'ILETA + LEUTA >= 1.305'}
{'Valine + Methionine' }	{'METAT + VALTA >= 0.705'}
{'Glycine' }	{'GTHS - GHMT2r >= 0.7725'}

coupledRxnId	constraints
{'Phosphatidylcholine'}	{'PCHOLP_hs + PLA2_2 + SMS >= 2.025'}
{'Adenosine Monophosphate'}	{'AMPDA + NTD7 >= 0.2265'}
{'Glutamate'}	{'- ALATA_L + GLUCYS + GLUDxm + GLUDym - ASPTA - ILETA - LEUTA - VALTA >= 1.1925'}
{'Aspartate'}	{'ARGSS + ASPTA >= 0.8625'}
{'Serine'}	{'GHMT2r + r0060 >= 0.7245'}
{'Arginine'}	{'GLYAMDTRc + r0145 + ARGN >= 0.55875'}
{'Tyrosine'}	{'TYR3MO2 + TYRTA + HMR_6728 + HMR_6874 >= 1.095'}
{'Histidine'}	{'HISDC + HISD >= 1.305'}
{'Leucine + Isoleucine'}	{'METAT + VALTA >= 0.705'}
{'Valine + Methionine'}	{'GTHS - GHMT2r >= 0.7725'}
{'Glycine'}	

Feasible model after adding coupling constraints.

Removing 233 reactions ...

The following reaction(s) to be removed is(are) not in the model:

{'HMR_biomass_Renalcancer'}
{'DM_HMR_biomass_renalcancer'}
{'biomass_components'}
{'EX_ser_D[e]'}
{'EX_pro_D[e]'}
{'HMR_1708'}
{'HMR_1934'}
{'r0947'}
{'r1431'}
{'r1432'}
{'RE1096R'}
{'RE1134R'}
{'RE2117M'}
{'RE2768R'}
{'RE2782C'}
{'RE3111M'}
{'RE3338C'}
{'RE3340C'}
{'RE3448C'}
{'RE3564C'}
{'RE3627C'}
{'DM_adchac[c]'}
{'DM_alchac[c]'}

Feasible model after removing inactive reactions.

2 deleted non-core metabolites, corresponding to bibliomic inactive reactions.

0 deleted core metabolites, corresponding to bibliomic inactive reactions.

210 deleted non-core reactions, corresponding to bibliomic inactive reactions.

0 deleted core reactions, corresponding to bibliomic inactive reactions.

Removing 973 inactive genes...

57 manually selected inactive genes have been marked as active by omics data and will be discarded:

{'5834'}
{'10165'}
{'246213'}
{'2571'}
{'26227'}
{'100137049'}

```
{
  '100526794'
}
{
  '2752'
}
{
  '26227'
}
{
  '10060'
}
{
  '10858'
}
{
  '10873'
}
{
  '1160'
}
{
  '125965'
}
{
  '130752'
}
{
  '1346'
}
{
  '1468'
}
{
  '1583'
}
{
  '1588'
}
{
  '1607'
}
{
  '170712'
}
{
  '206358'
}
{
  '2110'
}
{
  '240'
}
{
  '2645'
}
{
  '27165'
}
{
  '2747'
}
{
  '2820'
}
{
  '3099'
}
{
  '3101'
}
{
  '341947'
}
{
  '349565'
}
{
  '366'
}
{
  '374291'
}
{
  '3767'
}
{
  '412'
}
{
  '43'
}
{
  '5053'
}
{
  '5106'
}
{
  '548596'
}
{
  '57084'
}
{
  '622'
}
{
  '64802'
}
{
  '6505'
}
{
  '6529'
}
{
  '6531'
}
{
  '6538'
}
{
  '6571'
}
{
  '6581'
}
{
  '6582'
}
{
  '6818'
}
{
  '6833'
}
{
  '7054'
}
{
  '79751'
}
{
  '83733'
}
{
  '84889'
}
{
  '8659'
}
```

11 inactive genes are not in the model to be removed.

Infeasible model after temporarily closing reactions corresponding to inactive genes, relaxing...

itn	obj	obj_old	err(obj)	err(x)	card(v)	card(r)	card(p)	card(q)
0	8.4066	1332.5	1324.1	2.8021e+05	1272	0	1	4
1	8.2951	8.4066	0.11146	5.1996e+05	1242	0	1	4
2	8.4569	8.2951	0.16175	5.362e+05	1283	0	1	4
3	8.4867	8.4569	0.029796	5.3872e+05	1281	0	1	4
4	8.6194	8.4867	0.13272	5.3926e+05	1384	0	1	4
5	8.285	8.6194	0.33442	5.3742e+05	1246	0	1	4
6	8.7531	8.285	0.46819	5.398e+05	1322	0	1	4
7	8.264	8.7531	0.48915	5.4022e+05	1295	0	1	4

8	8.6802	8.264	0.4162	5.3896e+05	1326	0	1	4
9	8.4196	8.6802	0.26057	5.4105e+05	1277	0	1	4
10	8.5304	8.4196	0.11073	5.3927e+05	1277	0	1	4
11	8.3826	8.5304	0.14776	5.397e+05	1263	0	1	4
12	8.6824	8.3826	0.29983	5.4096e+05	1335	0	1	4
13	8.4864	8.6824	0.19599	5.4213e+05	1273	0	1	4
14	8.4448	8.4864	0.041662	5.4075e+05	1273	0	1	4
15	8.3701	8.4448	0.074667	5.4132e+05	1282	0	1	4
16	8.7422	8.3701	0.37209	5.4318e+05	1311	0	1	4
17	8.5835	8.7422	0.1587	5.4349e+05	1293	0	1	4
18	8.6017	8.5835	0.018147	5.4035e+05	1310	0	1	4
19	8.4429	8.6017	0.15872	5.3978e+05	1300	0	1	4
itn	obj	obj_old	err(obj)	err(x)	card(v)	card(r)	card(p)	card(q)

Relaxed model is feasible.

Statistics:

1 lower bound relaxation(s)

4 upper bound relaxation(s)

0 steady state relaxation(s)

The lower bound of these reactions had to be relaxed:

Closed_Reaction	Name	lb_before	lb_after
{'URAt'}	{'Uracil Transport via Faciliated Diffusion'}	0	-0.00701902520631847

The upper bound of these reactions had to be relaxed:

Closed_Reaction	Name	lb_before	lb_after	ub_befor
{'KHK'}	{'Ketohexokinase'}	0	0	0
{'MMEm'}	{'Methylmalonyl Coenzyme A Epimerase/Racemase'}	0	0	0
{'OIVD1m'}	{'2-Oxoisovalerate Dehydrogenase (Acylating'}	0	0	0
{'SERPT'}	{'Serine C-Palmitoyltransferase'}	0	0	0

... done.

5 reaction(s) were not deleted based on inactive genes as their removal would cause the model to be infeasible.

480 genes were specified as inactive but not removed as they are involved in reactions that may be catalysed by other genes.

123 deleted non-core metabolites, corresponding to inactive genes.

0 deleted core metabolites, corresponding to inactive genes.

1747 deleted non-core reactions, corresponding to inactive genes.

0 deleted core reactions, corresponding to inactive genes.

Feasible model after removing inactive genes (that do not affect core reactions).

Identifying flux consistent reactions ...

5715 x 8658 stoichiometric matrix, before flux consistency.

--- findFluxConsistentSubset START ---

2965 flux consistent metabolites

2750 flux inconsistent metabolites

4983 flux consistent reactions

3675 flux inconsistent reactions

--- findFluxConsistentSubset END ---

2739 deleted non-core metabolites, corresponding to flux inconsistency.

11 deleted core metabolites, corresponding to flux inconsistency.

mets	metNames
{'Tyr_ggn[c]'}	{'Tyr-194 Of Apo-Glycogenin Protein (Primer For Glycogen Synthesis)'}
{'pre_prot[r]'}	{'Glycophosphatidylinositol (Gpi)-Anchored Protein Precursor'}
{'retfa[c]'}	{'Fatty Acid Retinol'}

{'thm[m]'}	{'Thiamin'}	}
{'no2[c]'} }	{'Nitrite'}	}
{'CE1273[c]'} }	{'5Beta-Cholestane-3Alpha,7Alpha,12Alpha,24S,25-Pentol'}	}
{'pail35p_hs[n]'} }	{'1-Phosphatidyl-1D-Myo-Inositol 3,5-Bisphosphate'}	}
{'c101coa[c]'} }	{'Decenoyl Coenzyme A'}	}
{'fe3[c]'} }	{'Iron (Fe3+)'}	}
{'6hddopaqn[c]'} }	{'6-Hydroxydopamine-Quinone'}	}
{'gml_hs[n]'} }	{'Ganglioside Gml'}	}

3630 deleted non-core reactions, corresponding to flux inconsistency.

45 deleted core reactions, corresponding to flux inconsistency.

Forward_Reaction, 0 bound	Name	lb
---------------------------	------	----

{'ACHee'}	{'Acetylcholinesterase'}	0
{'APOC_LYS_BTNPm'}	{'Proteolysis of ApoC-Lys-Biotin, Mitochondrial'}	0
{'ARGNm'}	{'Arginase, Mitochondrial'}	0
{'CLS_hs'}	{'Cardiolipin Synthase (Homo Sapiens)'}	0
{'DURIK1m'}	{'Deoxyuridine Kinase (ATP:Deoxyuridine), Mitochondrial'}	0
{'EX_co[e]'}	{'Exchange of Carbon Monoxide'}	0
{'G3PD2m'}	{'Glycerol-3-Phosphate Dehydrogenase (FAD), Mitochondrial'}	0
{'GLYKm'}	{'Glycerol Kinase'}	0
{'OCOAT1m'}	{'3-Oxoacid Coa-Transferase'}	0
{'P45011A1m'}	{'Cytochrome P450 11A1, Mitochondrial [Precursor]'}	0
{'P45027A11m'}	{'5-Beta-Cholestane-3-Alpha, 7-Alpha, 12-Alpha-Triol 27-Hydrox'}	0
{'P45027A14m'}	{'5-Beta-Cytochrome P450, Family 27, Subfamily A, Polypeptide '}	0
{'RBK_D'}	{'D-Ribulokinase'}	0
{'SARDHm'}	{'Sarcosine Dehydrogenase, Mitochondrial'}	0
{'STS1'}	{'Seryl-Sulfatase'}	0
{'r0321'}	{'Acetoacetate:Coa Ligase (AMP-Forming)'}	0
{'FE2DMT1'}	{'Uptake of Food Iron by Dmt1 Transporter'}	0
{'ARGN'}	{'Arginase'}	0
{'RBK'}	{'Ribokinase'}	0
{'DOPAOQNOX'}	{'Dopamine-O-Quinone Oxidase'}	0
{'HMR_9726'}	{'5-Formyltetrahydrofolate:L-Glutamate N-Formiminotransferase'}	0
{'DHBOX'}	{'3,4-dihydroxybenzaldehyde oxidase'}	0
{'DM_clpn_hs[c]'}	{'Demand of cardiolipin'}	0
{'NORCON'}	{'Norsalsolinol condensation'}	0
{'Q-METHRED'}	{'Quinonemethide reductase'}	0

Reverse_Reaction, 0 bound	Name	lb	ub
---------------------------	------	----	----

{'r0245'}	{'Glycerol:NADP+ Oxidoreductase'}	-10000	0	{'nadp[...]
{'EX_fe3[e]'}	{'Exchange of Iron (Fe3+) '}	-1.63928248314438	0	{'fe3[e]'}
{'EX_k[e]'}	{'Exchange of Kalium'}	-733.702544787044	0	{'k[e]'}
{'EX_na1[e]'}	{'Exchange of Sodium'}	-10000	0	{'na1[e]'}
{'EX_ca2[e]'}	{'Exchange of calcium'}	-220.501332473743	0	{'ca2[e]'}
{'EX_cl[e]'}	{'Exchange of chloride ion'}	-10000	0	{'cl[e]'}
{'EX_zn2[e]'}	{'Exchange of zinc (II) ion'}	-0.283998779119298	0	{'zn2[e]'}

Reversible_Reaction	Name	lb	ub
---------------------	------	----	----

{'EX_i[e]'}	{'Exchange of Iodide '}		
{'RE1530M'}	{'Thymidine Kinase'}		
{'C02712tm'}	{'Transport of N-Acetylmethionine, Intracellular'}		
{'ACGLUtM'}	{'Transport of N-Acetyl-L-Glutamate, Mitochondrial'}		
{'r2535m'}	{'Transport of L-Homoserine, Mitochondrial'}		
{'EX_fe2[e]'}	{'Exchange of Iron (Fe2+)'}		
{'EX_pnto_R[e]'}	{'Exchange of (R)-Pantothenate '}		-1.01132
{'EX_hxan[e]'}	{'Exchange of Hypoxanthine '}		-1.16204
{'EX_thm[e]'}	{'Exchange of Thiamin'}		-1.41539
{'EX_pydxn[e]'}	{'Exchange of Pyridoxine'}		-2.26640

```

{'EX_btn[e]'}      {'Exchange of Biotin '}          }
{'CE2172t'}        {'Transport of 6, 7-Dihydroxy-1, 2, 3, 4-Tetrahydroisoquinolin'} }
{'EX_CE2172[e]'}   {'Exchange of 6, 7-Dihydroxy-1, 2, 3, 4-Tetrahydroisoquinoline'} }

-0.00110892

```

2965 x 4983 stoichiometric matrix, after flux consistency.

Identifying thermodynamically flux consistent subset ...

```

1 model.C constraints removed
    solver: 'mosek'
    algorithm: 'default'
    stat: 0
    origStat: 'PRIMAL_INFEASIBLE_CER'
origStatText: []
    time: 0.00464499999999646
    basis: []
        f: NaN
        v: []
        y: []
        w: []
        s: []
    ctrs_y: []
    ctrs_s: []
    x: []

```

Warning: findThermoConsistentFluxSubset: thermoConsistModel is not feasible.

0 deleted non-core metabolites, corresponding to thermodynamic flux inconsistency.

0 deleted core metabolites, corresponding to thermodynamic flux inconsistency.

1 deleted non-core reactions, corresponding to thermodynamic flux inconsistency.

12 deleted core reactions, corresponding to thermodynamic flux inconsistency.

Forward_Reaction, 0 bound	Name	lb	ub	equat
{'KHK'}	{'Ketohexokinase'}	0	0.287664725528884	{'atp[c] + fru[c] -> h[...]

Forward_Reaction, non-0 bound	Name			
{'ALATA_L'}	{'L-Alanine Transaminase'}			1
{'LYSOXp'}	{'Transport of L-Lysine Oxidase, Peroxisomal'}			
{'PIK4'}	{'Phosphatidylinositol 4-Kinase'}			
{'ATPM'}	{'Demand for ATP, Cytosolic'}			
{'GHMT2r'}	{'Glycine Hydroxymethyltransferase, Reversible'}			
{'ILETA'}	{'Isoleucine Transaminase'}			
{'LEUTA'}	{'Leucine Transaminase'}			
{'PROD2'}	{'Proline Dehydrogenase'}			0
{'NTD2'}	{'5'-Nucleotidase (UMP)'}			0
{'THRD_L'}	{'L-Threonine Deaminase'}			0
{'HMR_0653'}	{'S-Adenosyl-L-Methionine:Phosphatidylethanolamine N-Methyltra'}			0

2965 x 4970 stoichiometric matrix, after thermodynamic flux consistency.

Infeasible after extraction of thermodynamically feasible subset, relaxing...

itn	obj	obj_old	err(obj)	err(x)	card(v)	card(r)	card(p)	card(q)
0	8.5431	651.14	642.59	2.8354e+05	1252	0	3	0
1	7.9475	8.5431	0.59563	5.2136e+05	1172	0	3	0
2	8.4058	7.9475	0.45835	5.3075e+05	1279	0	3	0
3	8.5054	8.4058	0.099608	5.3921e+05	1213	0	3	0
4	8.3797	8.5054	0.12571	5.3952e+05	1284	0	3	0
5	8.6121	8.3797	0.23242	5.3995e+05	1276	0	3	0
6	8.5337	8.6121	0.078416	5.4113e+05	1299	0	3	0
7	8.5236	8.5337	0.010169	5.411e+05	1241	0	3	0

	8	8.5332	8.5236	0.0096319	5.4235e+05	1293	0	3	0
9	8.5411	8.5332	0.0079158	5.4318e+05	1242	0	3	0	0
10	8.6421	8.5411	0.10104	5.4241e+05	1243	0	3	0	0
11	8.4761	8.6421	0.16606	5.4348e+05	1225	0	3	0	0
12	8.7017	8.4761	0.22564	5.4381e+05	1295	0	3	0	0
13	8.6743	8.7017	0.027477	5.4587e+05	1272	0	3	0	0
14	8.6722	8.6743	0.0020358	5.4543e+05	1269	0	3	0	0
15	8.4168	8.6722	0.25544	5.4252e+05	1224	0	3	0	0
16	8.8788	8.4168	0.46204	5.4502e+05	1283	0	3	0	0
17	8.3864	8.8788	0.49246	5.4402e+05	1240	0	3	0	0
18	8.8933	8.3864	0.50697	5.4341e+05	1301	0	3	0	0
19	8.5211	8.8933	0.37227	5.4603e+05	1221	0	3	0	0
itn	obj	obj_old	err(obj)	err(x)	card(v)	card(r)	card(p)	card(q)	

Relaxed model is feasible.

Statistics:

3 lower bound relaxation(s)

0 upper bound relaxation(s)

0 steady state relaxation(s)

The lower bound of these reactions had to be relaxed:

Forward_Reaction, non-0 bound	Name	lb_before	
{'EX_2hb[e]'}	{'Exchange of 2-Hydroxybutyrate'}	0.0327019285314542	-8.789
{'EX_glyc_R[e]'}	{'Exchange of D-Glycerate'}	0.287664725528066	-8.185
{'EX_3hivac[e]'}	{'Exchange of 3-Hydroxy-Isovalerate'}	0.922557070436904	

... done.

184 active genes not present in model.genes, so they are ignored.

Extracting tissue specific model ...

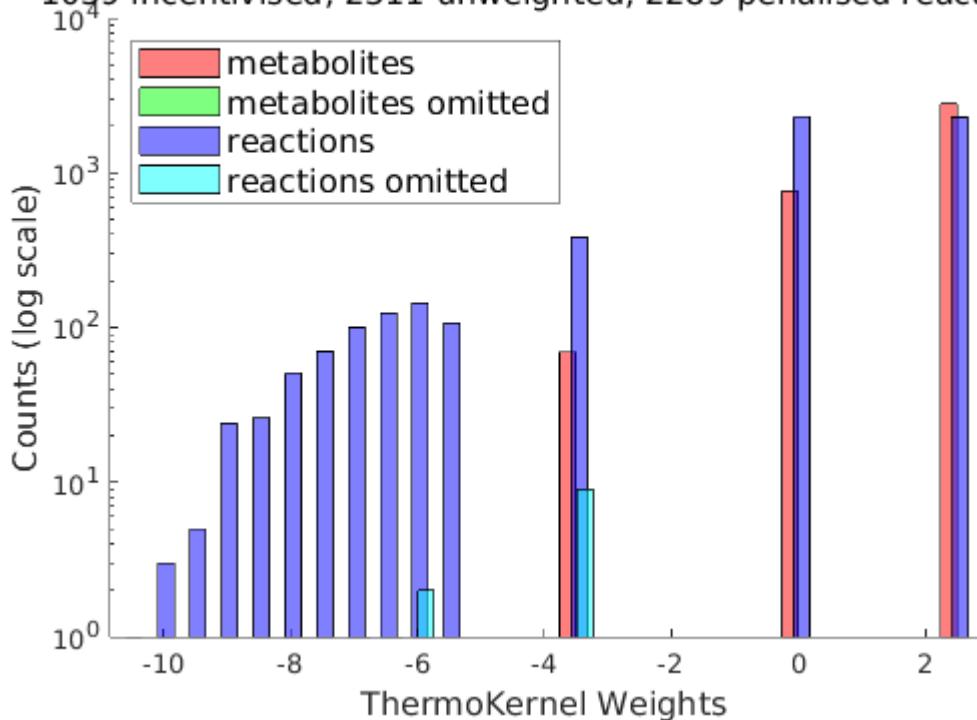
Using real valued weights on metabolites and reactions as input to thermoKernel.

Using real valued weights from omics on dummy reactions as input to thermoKernel.

13 forced internal reactions and relaxation of bounds so cycleFreeFlux only determines thermodynamic feasibility
[mosek] reports OPTIMAL but Primal optimality condition in solveCobraQP not satisfied, residual = 0.000803
[mosek] reports OPTIMAL but Primal optimality condition in solveCobraQP not satisfied, residual = 0.000265
[mosek] reports OPTIMAL but Primal optimality condition in solveCobraQP not satisfied, residual = 0.000379
[mosek] reports OPTIMAL but Primal optimality condition in solveCobraQP not satisfied, residual = 0.000216
[mosek] reports OPTIMAL but Primal optimality condition in solveCobraQP not satisfied, residual = 0.000430
[mosek] reports OPTIMAL but Primal optimality condition in solveCobraQP not satisfied, residual = 0.000396
[mosek] reports OPTIMAL but Primal optimality condition in solveCobraQP not satisfied, residual = 0.000359
[mosek] reports OPTIMAL but Primal optimality condition in solveCobraQP not satisfied, residual = 0.001583
[mosek] reports OPTIMAL but Primal optimality condition in solveCobraQP not satisfied, residual = 0.000423
All incentivised metabolites are produced.

Incentivised(-ve), ambivalent(0), and penalised(+ve)

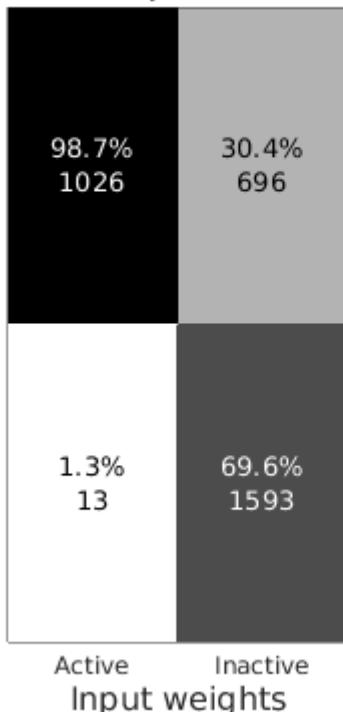
69 incentivised, 764 unweighted, 2801 penalised metabolites
 1039 incentivised, 2311 unweighted, 2289 penalised reactions



Reactions

Accuracy = 0.78696

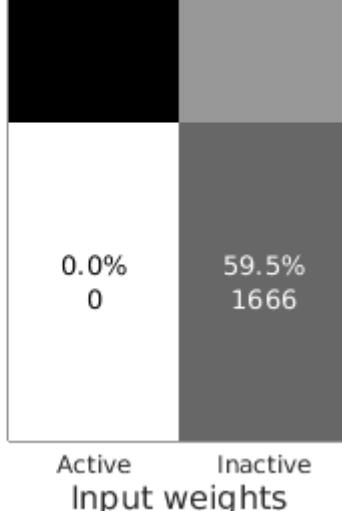
Output model



Metabolites

Accuracy = 0.60453

Output model



2335 deleted non-core metabolites, corresponding to removal by createTissueSpecificModel.
 0 deleted core metabolites, corresponding to removal by createTissueSpecificModel.

3520 deleted non-core reactions, corresponding to removal by createTissueSpecificModel.

9 deleted core reactions, corresponding to removal by createTissueSpecificModel.

Forward_Reaction, 0 bound

Name

1b

ub

{'OIVD1m'}	{'2-Oxoisovalerate Dehydrogenase (Acylating'}	0	0.922557070436596					
{'EX_3hivac[e]'}	{'Exchange of 3-Hydroxy-Isovalerate'}	0	10000					
Reverse_Reaction, 0 bound	Name	lb						
{'URAt'}	{'Uracil Transport via Faciliated Diffusion'}	-0.00701902520631847						
Reversible_Reaction	Name	lb						
{'EX_2hb[e]'}	{'Exchange of 2-Hydroxybutyrate '}	-8.78928874126217e-13	10000					
{'EX_lipoate[e]'}	{'Exchange of Lipoate '}	-0.0394044564234044	10000					
{'EX_ncam[e]'}	{'Exchange of Nicotinamide '}	-3.81469600757283	10000					
{'EX_prgstrn[e]'}	{'Exchange of Progesterone '}	-0.00161329356250689	10000					
{'EX_CE2028[e]'}	{'Exchange of Beta-Hydroxy-Beta-Methylbutyrate'}	-10000	10000					
{'EX_glyc_R[e]'}	{'Exchange of D-Glycerate'}	-8.18511924904897e-13	10000					
Infeasible tissue specific model. Trying relaxation...								
itn	obj	obj_old	err(obj)	err(x)	card(v)	card(r)	card(p)	card(q)
0	5.3659	276.96	271.6	2.2509e+05	888	0	1	0
1	5.5265	5.3659	0.16051	4.1938e+05	901	0	1	0
2	5.764	5.5265	0.23751	4.3585e+05	959	0	1	0
3	5.5763	5.764	0.18767	4.3948e+05	928	0	1	0
4	5.8715	5.5763	0.29517	4.4035e+05	932	0	1	0
5	5.7167	5.8715	0.15471	4.426e+05	944	0	1	0
6	5.8007	5.7167	0.083906	4.4116e+05	940	0	1	0
7	5.6403	5.8007	0.1604	4.4086e+05	927	0	1	0
8	5.8213	5.6403	0.18102	4.4043e+05	957	0	1	0
9	5.5644	5.8213	0.25686	4.3868e+05	929	0	1	0
10	5.7181	5.5644	0.15373	4.3992e+05	938	0	1	0
11	5.6189	5.7181	0.099209	4.408e+05	929	0	1	0
12	5.6687	5.6189	0.049738	4.3927e+05	931	0	1	0
13	5.6075	5.6687	0.061117	4.3911e+05	947	0	1	0
14	5.7732	5.6075	0.16563	4.3951e+05	957	0	1	0
15	5.7726	5.7732	0.00057859	4.4213e+05	939	0	1	0
16	5.8206	5.7726	0.048032	4.4014e+05	954	0	1	0
17	5.7132	5.8206	0.10742	4.3954e+05	939	0	1	0
18	5.6906	5.7132	0.022593	4.4018e+05	951	0	1	0
19	5.7724	5.6906	0.081751	4.4154e+05	933	0	1	0
itn	obj	obj_old	err(obj)	err(x)	card(v)	card(r)	card(p)	card(q)
Relaxed model is feasible.								
Statistics:								
1 lower bound relaxation(s)								
0 upper bound relaxation(s)								
0 steady state relaxation(s)								
The lower bound of these reactions had to be relaxed:								
Forward_Reaction, non-0 bound	Name	lb_before	lb_after					
{'EX_ura[e]'}	{'Exchange of Uracil '}	0.00701902520608372	-2.34757525918727e-					

... done.
... relaxation worked.
1299 x 2110 stoichiometric matrix after model extraction.

Feasible at end of XomicsToModel.

```
debugXomicsToModel:
```

#Active_genes		#Active_rxns		#Active_mets		Stage		
909		432		45		Active list		
#Active_genes	#Model_genes	#Active_rxns	#Model_rxns	#Active_mets	#Model_mets	lb_obj	Obj	Message
885	2248	401	10600	45	5835	0	755.003	Feasible
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
882	1892	422	10615	45	5840	NaN		0 Fea
873	1620	422	8658	45	5715	NaN		0 Fea
719	1307	377	4983	34	2965	NaN		0 Fea
719	1305	366	4970	34	2965	NaN		0 Fea
719	1305	366	5639	34	3634	NaN		0 Fea
715	1224	359	2110	34	1299	NaN		0 Fea

```
-----  
Diary written to: /home/rfleming/20221216T085613_rfleming_diary.txt  
XomicsToModel run is complete at:16-Dec-2022 09:09:37
```

Examining when active metabolites, reactions and genes were added or removed during the model generation process

```
debugXomicsToModel(model, pwd, modelGenerationReport)
```

#Active_genes		#Active_rxns		#Active_mets		Stage		
909		432		43		Active list		
#Active_genes	#Model_genes	#Active_rxns	#Model_rxns	#Active_mets	#Model_mets	lb_obj	Obj	Message
885	2248	401	10600	43	5835	0	755.003	Feasible
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
882	1892	422	10615	43	5840	NaN		0 Fea
873	1620	422	8658	43	5715	NaN		0 Fea
719	1307	377	4983	32	2965	NaN		0 Fea
719	1305	366	4970	32	2965	NaN		0 Fea
719	1305	366	5639	32	3634	NaN		0 Fea
715	1224	359	2110	32	1299	NaN		0 Fea

TIMING

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

Bibliography

1. Preciat, G., Wegrzyn, A. B., Luo, X., Thiele, I., Hankemeier, T., & Fleming, R. M. T. (2025). XomicsToModel: omics data integration and generation of thermodynamically consistent metabolic models. *Nature Protocols*. <https://doi.org/10.1038/s41596-025-01288-9>
2. Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, et al., "Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0", *Nature protocols* (2019).
3. German Preciat, Edinson Lucumi Moreno, Agnieszka B. Wegrzyn, et al., "Mechanistic model-driven exometabolomic characterisation of human dopaminergic neuronal metabolism", *bioRxiv* (2021)
4. Elizabeth Brunk, Swagatika Sahoo, Daniel C. Zielinski, et al., "Recon3D enables a three-dimensional view of gene variation in human metabolism", *Nature biotechnology* (2018)