Genome-wide Precision Metabolic Modeling Protocol

Gonghua Li. Ph.D

Email: ligonghua@mail.kiz.ac.cn

or ligonghua@vip.163.com

Genome-wide Precision Metabolic Modeling (GPMM) is aim to perform genome-wide metabolic modeling using single or multi generalized omics inputs. In the recent version, GPMM can systematically model the human complex metabolic network only using the transcriptome (e.g. RNA seq data). Any transcriptome data (including single cell sequencing data) can be directly used to perform genome-wide metabolic modeling using GPMM toolbox.

Part I: Required software and toolbox

1). Install MATLAB. MATLAB is the software environment for engineers and scientists. You can get detailed installation information by visiting http://www.mathworks.com/products/matlab/

2). Install Cobra toolbox in Matlab. Cobra toolbox is a matlab toolbox that used to implement metabolic modeling. Users can download and install this toolbox from: https://opencobra.github.io/cobratoolbox/

3). Optimal: Install FastMM package. FastMM is a C/C++ package to implement Flux viarly analysis (FVA) and genome-wide Knock-in and knockout analysis of constraint-based metabolic models. Users can download and install this package from: https://github.com/GonghuaLi/FastMM/. Make sure the binaries of FastMM are in your path environment.

Part II: Required datasets

1). Curated Recon3. Curated global human metabolic reconstruction (Reocn3). This file can be found in: './data/ Recon3_v1.mat'

2). Expression dataset. This file should be saved in: './data/'

3). Nutrition uptake file. In this case, the nutrition uptake of human muscle can be found in: './data/'

Part III: set parameters for GPMM

All of the parameters can be set in the file of "pars.txt" . These parameters contains:

Path of curated Recon3 file, the default is:

curatedRecon3 = './data/Recon3_v1.mat';

Cobra optimization solver, GPMM supports　'cplex'　and 'gurobi5'
cobrasolver = 'cplex';

Use FastMM　to accelerate metabolic modeling?　yes or no
isFastMM = 'no';

Path of nutrition uptake file, the default is muscle nutrition uptake:
exchangeFile = './data/exchange_rxns_white_blood_final_recentmodel.txt';
Unit of nutrition uptake, recently support two types: uM/100ml/min　or mM/L/min,
uptakeUnite = 'mM/L/min';

Path of expression file,
expressionFile = './data/protein_coding_matrix.txt';

Expression type, this version of GPMM supports three expression data type:
　　RMA: microarray, normalized by RMA method
　　FPKM:　RNAseq, normalized by FPKM method
　　RSEM:　RNAseq, normalized by RSEM method
expressionType = 'FPKM';

If the model contains protein degradation reaction. In this case, burned muscle contains protein degradation:
degradation = 'yes';
But in most cases, you should set as
degradation = 'no';

number of threading used for FVA, only used when FastMM was installed
fva_threading = '6';

Number of threading for MCMC sampling. MCMC sampling is very time costing, especially for hundreds and thousands of samples. Set multiple threading will largely save time. Default is 4, but you can set more, for example , cpu = '8' or cpu = '16'.
cpu = '4';

Part IV: Run GPMM

After you have prepared the datasets, and set proper parameters, Run GPMM will be very easy. You just need to type one command in matlab command window and wait for the results.

run_GPMM

This command will automatically reconstruct the personalized metabolic models and implement the Markov Chain Monte Carlo(MCMC) sampling.

Part V: Interpret the GPMM results.

The GPMM results are saved in two subdirectories: ./Mat_PQM and ./out.

./Mat_PQM subdirectory saved the reconstructed individual models where the model names are the same as the sample names described in expression matrix. This subdirectory also saved the result of MCMC sampling results for each individual where the name is begin with "MCMC".

./out subdirectory saved the summary results from the ./Mat_PQM and is ready for further statistical analysis.

'PQMM_fluxRxnsMax.txt',' PQMM_fluxRxnsMin.txt', 'PQMM_fluxRxnsMean.txt' and 'PQMM_fluxRxnsMedian.txt' represent matrix of maximum, minimum, mean and median fluxes for each sample using the MCMC sampling methods, respectively.

'PQM_productionFluxMetMax.txt', 'PQM_productionFluxMetMin.txt', 'PQM_productionFluxMetMean.txt', and 'PQM_productionFluxMetMedian.txt' represent matrix of maximum, minimum, mean and median of total flux of metabolite production by enzyme reaction for each sample, respectively.

'PQM_consumptionFluxMetMax.txt', 'PQM_consumptionFluxMetMin.txt', 'PQM_consumptionFluxMetMean.txt', and 'PQM_consumptionFluxMetMedian.txt' represent matrix of maximum, minimum, mean and median of total flux of metabolite consumption by enzyme reaction for each sample, respectively.

'PQM_detaFluxMetMax.txt', 'PQM_detaFluxMetMin.txt', 'PQM_detaFluxMetMean.txt', and 'PQM_detaFluxMetMedian.txt' represent matrix of maximum, minimum, mean and median of net flux of metabolite by enzyme reaction for each sample, respectively. The net flux of metabolite is equal to the sum of transport metabolites.

In steady state, for any metabolite, production + consumption - transport(deta flux) = 0.