Flux Analysis in Metabolite Biomarker Discovery

Wai-Ki CHING

Department of Mathematics

The University of Hong Kong

Abstract: Incorporating metabolic and gene expression data to reveal biochemical networks becomes a considerably admitted challenge. We here propose a promising approach to identify metabolic biomarkers through integrating available biomedical data and disease-specific gene expression data. Linear Programming(LP) based method is utilized to determine flux variability interval, therefore enabling the analysis of significant metabolites. This application can also be used for discovering potential novel biomarkers.

A joint work with Limin,Li, Wai-Ki,Ching and other postgraduate student.

Introduction

(1) Motivations and Objectives.

(2) Description of Metabolite Network.

(3) Expression Levels in Reactions.

(4) LP Model to determine flux interval.

(5) Metabolite Biomarker Discovery.

(6) Results.

(7) Concluding Remarks.

1 Motivations and Objectives.

• **Metabolic diseases**, can be directly caused by the lack of essential metabolites. Analysis of metabolite networks has been used in the past decades.

- J.G. Reich and E.E. Selkov, *Energy Metabolism of the Cell: A Theoretical Treatise*, New York: Acad.Press, 1981.

- J.Varner and D.Ramkrishna, *Understanding the Control of Metabolism*, London: Portland.Press, 1996.

- *Metabolic engineering from a cyberneic perspective.1.Theoretical preliminaries*, Biotechnol.Prog.(15)407-425, 1999.

Admitted Challenge

$$\begin{cases} \text{Metabolic Data} \\ \text{Gene Expression Data} \end{cases}$$

Incorporating metabolic and gene expression data to reveal biochemical networks.

Contribution

We propose a promising approach to identify metabolic biomarkers.

2 Metabolite Network Description

- **Stoichiometry.**

(1)M_h2o_c+(1)M_mi134p_c→(1)M_mi34p_c+(1)M_pi_c

	M_h2o_c	M_mi134p_c	M_mi34p_c	M_pi_c
Stoichimetry	−1	−1	1	1

Table 1.

- **Enzymes involved in Reactions**

'LOCUS:_10846*TRANSCRIPT:1*ABBREVIATION:**PDE10A***'

↓

PDE10A

- **Flux lowerbound and upperbound.**

[0, 1000] ([-1000,1000] if reaction is reversible)

Within the human metabolic network, there are

- 3742 reactions
- 2766 metabolites
- 1905 genes

3 Expression Levels in Reactions.

To determine expression levels in reactions, we have to determine expression levels in genes.

- **Gene Expression**

$$\text{New\_Signal} = \frac{\text{signal} - \mu_{\text{signal}}}{\sigma_{\text{signal}}}.$$

μ_{signal} is the average signal value, σ_{signal} is the standard deviation of the signal vector.

Binary Expressions :

$$\begin{cases} 0, & \textbf{New\_Signal} < 0; \\ 1, & \textbf{New\_Signal} \geq 0. \end{cases}$$

- **Reaction Expression**

Those highly expressed reactions are defined if all the participating genes are highly expressed, otherwise, we define the reactions to be lowly expressed.

	Gene1	Gene2	Gene3	Gene4
$R1$	0	1	1	1
$R2$	1	1	1	1

Table 2.

R1 is lowly expressed, R2 is highly expressed.

4 LP Model to determine Flux interval

- The **Mixed Integer Programming Model**

$$
\begin{aligned}
&\max_{y_i^+, y_i^-, \mathbf{v}} \textstyle\sum_{i \in R_H} (y_i^+ + y_i^-) + \sum_{i \in R_L} y_i^+ \\
&\mathbf{S} \cdot \mathbf{v} = 0 \\
&\mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max} \\
&v_i + y_i^+ (v_{min,i} - \epsilon) \geq v_{min,i}, i \in R_H \\
&v_i + y_i^+ (v_{max,i} + \epsilon) \leq v_{max,i}, i \in R_H \\
&v_i (1 - y_i^+) \leq v \leq v_{max,i} (1 - y_i^+), i \in R_L \\
&y_i^+, y_i^- = \{0, 1\}.
\end{aligned}
$$

v: flux for all the reactions;

S : stoichiometric matrix

ϵ : the flux threshold, and it is chosen to be 1 (see Reference)

R_H, R_L : highly and lowly expressed reactions.

y_i^+, y_i^-: reaction i is active or inactive

-L.Li and X.Zhou and W.Ching and P.Wang *Predicting enzyme targets for cancer drugs by profiling human Metabolic reactions in NCI-60 cell lines, BMC Bioinformatics , (11)501, 2010.*

- The **Linear Programming Model**

$$
\begin{aligned}
&\max_{y_i^+, y_i^-, \mathbf{v}} \textstyle\sum_{i \in R_H} (y_i^+ + y_i^-) + \sum_{i \in R_L} y_i^+ \\
&\mathbf{S} \cdot \mathbf{v} = 0 \\
&\mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max} \\
&v_i + y_i^+ (v_{min,i} - \epsilon) \geq v_{min,i}, i \in R_H \\
&v_i + y_i^+ (v_{max,i} + \epsilon) \leq v_{max,i}, i \in R_H \\
&v_i (1 - y_i^+) \leq v \leq v_{max,i} (1 - y_i^+), i \in R_L \\
&0 \leq y_i^+, y_i^- \leq 1.
\end{aligned}
$$

Interpretation: In the LP model, y_i^+ and y_i^- represent the likelihoods for reaction i to be active which are more realistic. Furthermore, the LP model is much easier to handle when compared to the MILP model.

Alternate Optimal Solutions to Determine Flux Interval

Multiple feasible solutions exist in the LP model.

Emphasis:
Investigate the multiplicity of solutions to find the lower bound and upper bound for each flux.

Technical Support:
All the flux ranges can be determined through solving a series of LP problems(See Reference).

- R.Mahadevan and C.H.Schilling *The effects of alternate optimal solutions in constraint-based genome-scale metabolic models, Metabolic Engineering, (5)264-276, 2003.*

Flux Profiles in Disease/Normal Sample

Using Expression Levels for Normal Sample, we obtain Flux interval for the Normal with **LP** model.
Using Expression Levels for Disease Sample, we obtain Flux interval for the Disease with **LP** model.

$$\Downarrow$$

$$\begin{cases} FN\_Lower, & \text{lower bound of flux profiles in normal sample;} \\ FN\_Upper, & \text{upper bound of flux profiles in normal sample;} \\ FD\_Lower, & \text{lower bound of flux profiles in disease sample;} \\ FD\_Upper, & \text{upper bound of flux profiles in disease sample.} \end{cases}$$

5 Metabolite Biomarker Discovery

- **Significant Reaction Identification**

$[l1, u1]$: flux intervals for disease sample.
$[l2, u2]$ flux intervals for normal sample.

$$\Downarrow$$

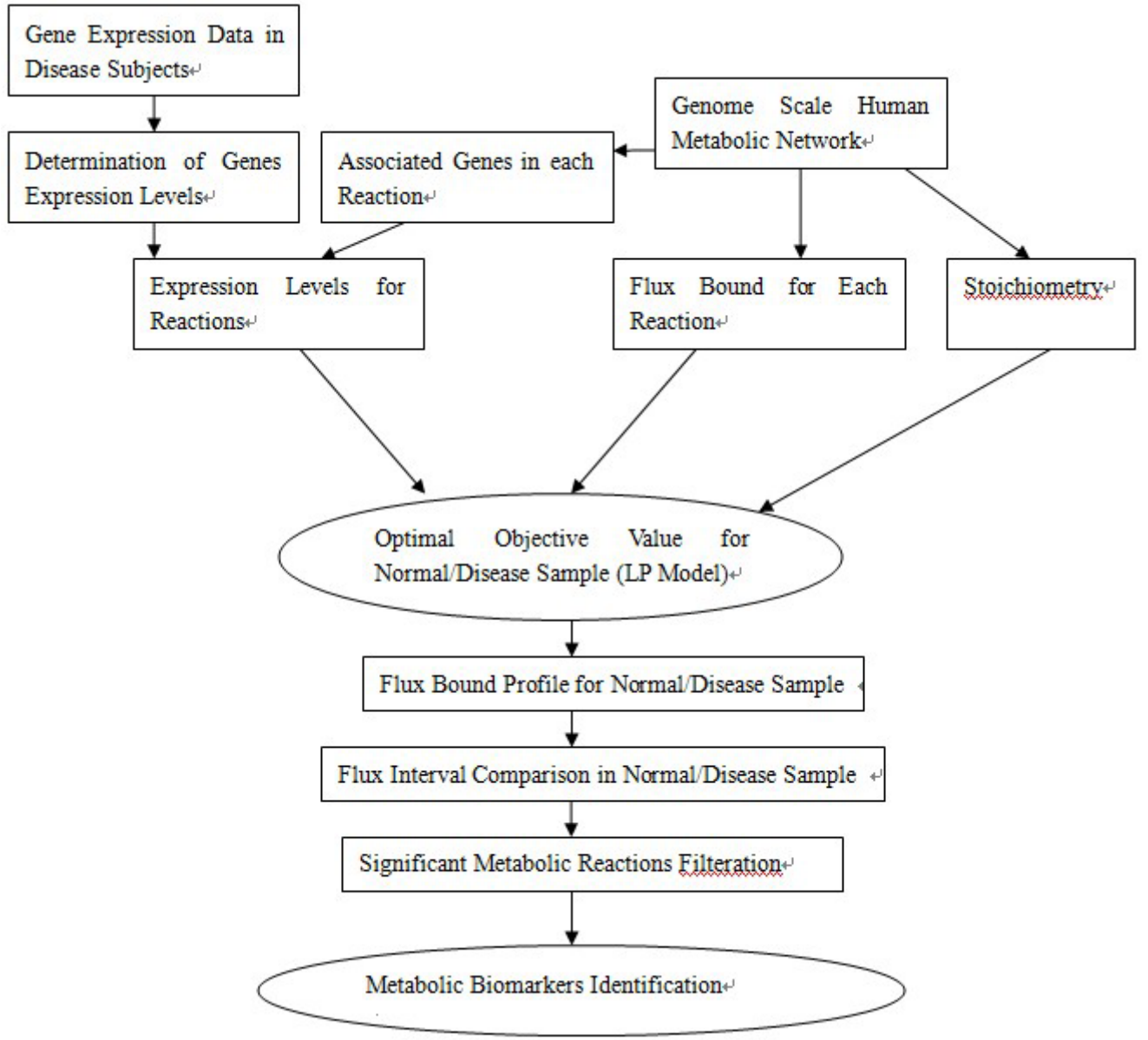
Significant Reactions will be selected if $u1 \leq l2$ or $u2 \leq l1$.

- **Significant Metabolite Discovery**

In each disease, two pairs of control and disease sample are used.

$$\Downarrow$$

Two sets of Significant Reactions $\Rightarrow$ **Reaction Markers(Overlap of the Two!!)** $\Rightarrow$ **Metabolite Biomarkers: Boundary Metabolites**


Gene Expression Data in Disease Subjects
Determination of Genes Expression Levels
Associated Genes in each Reaction
Genome Scale Human Metabolic Network
Expression Levels for Reactions
Flux Bound for Each Reaction
Stoichiometry
Optimal Objective Value for Normal/Disease Sample (LP Model)
Flux Bound Profile for Normal/Disease Sample
Flux Interval Comparison in Normal/Disease Sample
Significant Metabolic Reactions Filteration
Metabolic Biomarkers Identification

- **Materials**

Gene Expression Data: GEO Data sets(log_2 transformed)

- Diabetes: 12558 genes
  Platform: GPL8300

- Obesity: 54675 genes
  Platform: GPL570

Human Metabolite Network Data: BiGG Database

6 Results

$Index$	$Reactions$
1238	$`[e] : ac \rightleftharpoons ac'$
1951	$`gcald[c] + h2o[c] + nad[c] \rightarrow glyclt[c] + (2)h[c] + nadh[c]'$
2297	$`eandrstrn[r] + h[r] + nadph[r] \rightarrow andrstandn[r] + h2o[r] + nadp[r]'$
2357	$`atp[c] + xylu - D[c] \rightarrow adp[c] + h[c] + xu1p - D[c]'$
2700	$`dcdp[c] + h2o[c] \rightarrow dcmp[c] + h[c] + pi[c]'$

Table 3.Significant Reactions for Diabetes

$Index$	$Reactions$
158	$`4abut[m] + akg[m] \rightleftharpoons glu - L[m] + sucsal[m]'$
248	$`ac[m] + atp[m] + coa[m] \rightarrow accoa[m] + amp[m] + ppi[m]'$
582	$`apoC - Lys[c] + btamp[c] \rightarrow amp[c] + apoC - Lys\_btn[c] + h[c]'$
1506	$`[e] : pydam \rightleftharpoons pydam'$
3682	$`(2)na1[e] + uri[e] \rightarrow (2)na1[c] + uri[c]'$

Table 5.Significant Reactions for Obesity

$ReactionsIndex$	Genes
1238	'NONE'
1951	'ALDH1A1' 'ALDH1A2' 'ALDH1A3' 'ALDH3A1' 'ALDH3A2' 'ALDH3B1' 'ALDH3B2' 'ALDH7A1' 'ALDH9A1'
2297	'HSD3B2'
2357	'KHK'
2700	'NONE'

Table 4. Significant Genes for Diabetes

$ReactionsIndex$	Genes
158	'ABAT'
248	'ACAS2L'
582	'HLCS'
1506	'NONE'
3682	'SLC28A3'

Table 6. Significant Genes for Obesity

Diabetes(For Illustration)

'ALDH' : related to the increasing risk of large vessel disease in diabetes.

'HSD3B2' : highly expressed with regulation of FXR (farnesoid X receptor) where FXR agonists are emerging therapeutic treatment of diabetes.

The value of KHK as a pharmacological target needs testing : potential biomarker in diabetes treatment.

'ac[e]' : an inhibitor in clinical trials.

'nadph', 'nadp' : significant in that l-xylulose is intensively used in diabetes diagnosis.

'pi' : a key component in the disturbance of diabetes.

Concluding Remarks

Linear Programming (LP) based strategy was used to obtain flux profiles in both disease and normal sample.

Gene expression data for two pairs of sample in both disease and normal status strengthens the significance of discovered genes or metabolites which can be deemed as potential biomarkers.

The integration of gene expression levels with genome-scale human metabolic network data provides a new way to systematically analyze potential biomarkers.