Metabolomics pathway analysis and network integration

Shuzhao Li, Ph.D.
Assistant Professor, Dept. Medicine,
Emory University School of Medicine
E-mail: shuzhao.li@gmail.com
July 20, 2016

- Metabolomics pathway analysis and *mummichog*
- Examples
  - Snyderome re-visited
  - MWAS of NAFLD
  - PCB Exposure
  - Cross-generation cancer risk
  - Memory T cells
  - VZV systems immunology
Pathway enrichment test

If metabolites are known; red are significant metabolites

Untargeted metabolomics data

<table>
<thead>
<tr>
<th>m/z</th>
<th>Retention time</th>
<th>Sample_1</th>
<th>Sample_2</th>
<th>Sample_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>260</td>
<td>58.23623037</td>
<td>3.06</td>
<td>7.3</td>
<td>16.9</td>
</tr>
<tr>
<td>242</td>
<td>70.46755402</td>
<td>4.28</td>
<td>9.0</td>
<td>21.4</td>
</tr>
<tr>
<td>224</td>
<td>40.31440000</td>
<td>2.56</td>
<td>3.9</td>
<td>7.8</td>
</tr>
<tr>
<td>206</td>
<td>70.46755402</td>
<td>4.28</td>
<td>9.0</td>
<td>21.4</td>
</tr>
<tr>
<td>188</td>
<td>58.23623037</td>
<td>3.06</td>
<td>7.3</td>
<td>16.9</td>
</tr>
</tbody>
</table>

P = (n_p) / (n_t) = 0.01
Uncertainty in matching metabolites - features

Search of m/z 190.1065 in HMDB with accurate matching

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Chemical Formula</th>
<th>Exact m/z Value</th>
<th>Monoisotopic m/z</th>
<th>Relative Error</th>
<th>Match Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Hydroxyisovaleric acid</td>
<td>C6H12O3</td>
<td>190.1065</td>
<td>190.1065</td>
<td>0.004</td>
<td>100</td>
</tr>
<tr>
<td>Acetylpyrroline</td>
<td>C5H10O2</td>
<td>190.1065</td>
<td>190.1065</td>
<td>0.004</td>
<td>100</td>
</tr>
<tr>
<td>Lactobacillic acid</td>
<td>C6H12O3</td>
<td>190.1065</td>
<td>190.1065</td>
<td>0.004</td>
<td>100</td>
</tr>
<tr>
<td>Lactic acid</td>
<td>C3H6O3</td>
<td>190.1065</td>
<td>190.1065</td>
<td>0.004</td>
<td>100</td>
</tr>
<tr>
<td>Asp</td>
<td>C4H7NO2</td>
<td>190.1065</td>
<td>190.1065</td>
<td>0.004</td>
<td>100</td>
</tr>
<tr>
<td>Pyruvic acid</td>
<td>C3H4O3</td>
<td>190.1065</td>
<td>190.1065</td>
<td>0.004</td>
<td>100</td>
</tr>
<tr>
<td>Alanine</td>
<td>C3H6O2</td>
<td>190.1065</td>
<td>190.1065</td>
<td>0.004</td>
<td>100</td>
</tr>
<tr>
<td>Alanine-2-carboxylic acid</td>
<td>C3H6O3</td>
<td>190.1065</td>
<td>190.1065</td>
<td>0.004</td>
<td>100</td>
</tr>
<tr>
<td>Aconitase</td>
<td>C2H3NO</td>
<td>190.1065</td>
<td>190.1065</td>
<td>0.004</td>
<td>100</td>
</tr>
</tbody>
</table>

metabolites + biological activity

? ?
Genome-scale Metabolic model

Mummichog bridging metabolic models

Li et al. 2013. PLoS Computational Biology. 9:e1003123
Module analysis in *mummichog*

For a subgraph $G$, activity score

$$\tilde{A} = Q \cdot \frac{N_{G,1}}{N_{G}}$$

where $N_{G}$ is the number of compounds in $G$, $N_{G,1}$ the number of input compounds in $G$, $Q$ the adjusted Newman-Girvan modularity:

$$Q = \frac{1}{N_{G}} \left( \frac{E_{G}}{m} - \sum_{i,j \in G} \frac{k_{i} k_{j}}{2m} \right), \quad i,j \in G$$

$k_{i}$ the network degree of compound $i$, $m$ the total number of edges in the metabolic network, $E_{G}$ the total number of edges in $G$, $N_{G}$ the number of input compounds.

Pathway analysis in *mummichog*

Li et al. 2013. PLoS Computational Biology. 9:e10031323
Testing module/pathway significance in *mummichog*

Distribution of permutated data (null distribution)

Distribution of real data from user’s significant m/z list

---

? metabolites + ?

? biological activity
Case study: viral activation of immune cells

Monocyte derived dendritic cells (moDC)

+ YF-17D
+ mock

6 hrs

QA: total ion counts are similar among samples
**Mummichog: viral activation of immune cells**

- Tandem mass spectrometry confirmed 9/11 metabolites
- Gene expression supported GSH/GSSG depletion and Arg/Cit conversion

**Experimental validation of mummichog prediction**

- Gene expression supported GSH/GSSG depletion and Arg/Cit conversion

Li et al. 2013. PLoS Computational Biology. 9:e10031323
Arginine as master regulator of viral response

Li et al. 2013. PLoS Computational Biology. 9:e1003132

Argininosuccinate synthetase 1 knockdown led to increased replication of HSV-1.

Ravindran et al. 2014. Science 343:313

Arginine as master regulator of viral response

Mummichog demo

Mummichog.org
A few more examples

- Snyderome re-visited
- MWAS of NAFLD
- PCB Exposure
- Cross-generation cancer risk
- Memory T cells
- VZV systems immunology

Snyderome: personal omics (2012)
**Mummichog interpretation of Snyder metabolome**

<table>
<thead>
<tr>
<th>Pathways</th>
<th>overlap_size</th>
<th>pathway_size</th>
<th>p-value (raw)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysinate metabolism</td>
<td>8</td>
<td>10</td>
<td>3e-05</td>
<td>0.00035</td>
</tr>
<tr>
<td>Glycolytic metabolism</td>
<td>19</td>
<td>0.14045</td>
<td>0.02665</td>
<td></td>
</tr>
<tr>
<td>Pyruvate metabolism</td>
<td>5</td>
<td>16</td>
<td>0.1234</td>
<td>0.00585</td>
</tr>
</tbody>
</table>
The "N" in systems medicine

MWAS + mummichog (NAFLD)

MWAS of Polychlorinated biophenyl

CHDS: exposure and risk of breast cancer

Adapted from Perera F, Herbstman J, Reproductive Toxicology PMID: 21256208; Courtesy Barbara Cohn
Maternal metabolome associated with daughters’ breast cancer (I)

Unpublished data removed

Maternal metabolome associated with daughters’ breast cancer (II)

Unpublished data removed
Multi-omics integration

- **DNA**
  - Genome & Epigenome
    - polymorphisms: single nucleotide polymorphisms (SNPs), copy number variation (CNV)
    - whole genome sequencing
    - DNA methylation

- **RNA**
  - Transcriptome & Exome
    - genome-wide expression profiling
    - microarrays
    - RNAseq
    - exomes

- **Protein**
  - Proteome & Interactome
    - MS-based proteomics
    - protein assays
    - phosphorylation state assays
    - protein-protein interactions (PPI)
    - protein:small molecule interactions

- **Metabolite**
  - Metabolome & Fluxome
    - NMR and MS-based metabolomics
    - metabolic pathways
    - metabolic fluxes

- **Phenotype**
  - Phenome & Exposome
    - disease phenotype
    - treatment
    - family history
    - risk factors
    - environmental exposure

**Figure:** Courtesy: Doug Walker

Dumas, 2012. Mol. BioSyst. 8:2494
Autophagy is essential for effector CD8+ T cell survival and memory formation

Here we investigated two issues: the kinetics of autophagy activation during the course of the viral infection and the role of autophagy in the generation of memory CD8+ T cells. Our studies define when autophagy is needed during effector and memory differentiation and warrant reexamination of the relationship between T cell activation and autophagy. It is important to define the role of macroautophagy (called autophagy) in the differentiation process, T cells undergo cellular and metabolic reprogramming to shift from anabolic processes and proliferation to catabolic processes and contraction of cell populations to generate survival and memory formation.

Enzymes associated with significant metabolites

The metabolic and transcriptomic profiles of antigen-specific T cells during the course of the viral infection were examined. The data show that autophagy is upregulated in proliferating T cells just before the contraction phase, which is consistent with our findings. Deletion of the Atg5 gene, which encodes a component of the autophagy machinery, results in reduced autophagy and impaired T cell function. This suggests that autophagy plays a critical role in T cell differentiation and survival.

The importance of autophagy in the generation of memory CD8+ T cells has been studied mainly with genetic approaches, which have shown that autophagy is upregulated in proliferating T cells. However, questions remain about why and how proliferating T cells upregulate autophagy with cell growth and proliferation. It is important to define the role of macroautophagy (called autophagy) in the differentiation process, T cells undergo cellular and metabolic reprogramming to shift from anabolic processes and proliferation to catabolic processes and contraction of cell populations to generate survival and memory formation.

In summary, our findings suggest that autophagy is an important regulator of T cell differentiation and memory formation. Further studies are needed to understand the molecular mechanisms underlying autophagy regulation during T cell activation and proliferation.
Enzyme genes significantly enriched towards KO

Expression of genes corresponding to related enzymes are enriched for KO cells, DNA microarray data, GSEA (Gene Set Enrichment Analysis). Nominal p = 0, FWER, p = 0.024.

Comprehensive profiling of VZV immunization

Li, Sullivan, et al. To be submitted.
Summary and future directions

- Advancing of mass spectrometry enables deep sequencing of metabolome and exposome; filling gap for G x E


- MWAS + mummichog is a powerful approach to understand health and disease

- Combining multiple omics is critical to small “N”, human studies. Their integration can be driven by data mining or by knowledge models.
Old workflow

Metabolite identification
MS/MS
Pathway mapping

New workflow

Metabolite identification
MS/MS

Mummichog pathway/network analysis

Omics integration

Emory University
Dept. Medicine
Dean P. Jones
Young-Mi Go
Douglas Walker
Vinh Tran
Bill Liang
Karan Upfal
Ken Liu
Louka Lili-Williams
Sophia Banton
Andrei Todor
Michael Hart
David Green
Annette Esper
Sushma Cribbs
Mark Mulligan
Nadine Rouphael
Anesh Mehta
Mary Galinski
Dept. Genetics
Judy Fedorovitch-Kel

Emory Vaccine Center
Bali Pulendran
Helder Nakaya
Mohan Maddur
Sathyanarayana
Sai Duraisingham
Rafi Ahmed
Nicole Sullivan
Xiaojin Xu
Koichi Araki
Megan McCausland
Miriam B. Vos
Ran Jin
William Hu

Vanderbilt University
Milam Brantley

Georgia State University
Ruiyan Luo

University of Colorado
Adriana Weinberg
Myron Levin
Nancy Lang

CHDS, Public Health Institute
Barbara Cohn
Piera Cirillo
Nicki Krigbaum
Lauren Zimmermann

National Jewish
Russ Bowler

Thank you!

La Jolla Institute for Allergy and Immunology
Alex Sette
Bjoern Peters
John Sidney

Merck Inc.
Kalpit Vora
Sheri Dubey
Janet Sai
Joseph Antonello
Eberhard Durr
Kelly Collins

UCSD
Shankar Subramaniam
Thomas Hagen

Dana Farber Cancer Institute
Nick Haining

Acknowledgement of research supports from US National Institute of Health (NIHES P30 ES019116, 1U2CES026560-01, 1P50ES026071-01, NIAID HHSN272201200031C, 2U19AI090023-06, HHSN27220130001B, NIDDK 1RO1DK107900-01), Department of Defense (HT9404-13-1-0030) and California Breast Cancer Research Program (21UB-8002).