Construction of a Computational Metabolic Model of the Human Gut Microbiome

Tomer Altman
taltman1@stanford.edu

Biomedical Informatics
Stanford University
# Table of Contents

1. Introduction
2. Pipeline and Model Set Description
3. Comparison of PMGDBs
4. Vignettes
   - Vignette #1: Metabolite Tracing
   - Vignette #2: Chimeric Pathways
   - Vignette #3: Pharmaco-Microbiomics
   - Vignette #4: Analysis of Mass-Spec Data
5. Conclusion
6. Extras
Outline

1. Introduction
2. Pipeline and Model Set Description
3. Comparison of PMGDBs
4. Vignettes
   - Vignette #1: Metabolite Tracing
   - Vignette #2: Chimeric Pathways
   - Vignette #3: Pharmaco-Microbiomics
   - Vignette #4: Analysis of Mass-Spec Data
5. Conclusion
6. Extras
The human body contains and is covered by thousands of microbes
The Human Microbiome

- The human body contains and is covered by thousands of microbes
- Paradigm shift: from pathogenicity to symbiosis ("supra-organism")
The Human Microbiome

The human body contains and is covered by thousands of microbes.

Paradigm shift: from pathogenicity to symbiosis ("supra-organism")

Microbiome involved in obesity, irritable bowel syndrome, gingivitis, cancer, and cardiovascular disease.
The human body contains and is covered by thousands of microbes

Paradigm shift: from pathogenicity to symbiosis ("supra-organism")

Microbiome involved in obesity, irritable bowel syndrome, gingivitis, cancer, and cardiovascular disease

Understanding the function of the microbial communities in health and disease is a grand challenge
The real culprit, they proposed, was a little-studied chemical that is burped out by bacteria in the intestines after people eat red meat."
The Human Microbiome Project and MetaHIT

- US and European multi-year, multi-institution efforts
- Tens of millions of dollars
- Hundreds of volunteers and samples
- Software and data publicly available
Benefits of Modeling Multi-Organism Metabolic Pathways

- Integrate domain knowledge into Pathway/Metagenome Database

(Wikipedia)
Benefits of Modeling Multi-Organism Metabolic Pathways

- Integrate domain knowledge into Pathway/Metagenome Database
- Allow disparate data modalities to be compared: 16S rRNA, (meta)genomics, transcriptomics, metabolomics, etc.
Benefits of Modeling Multi-Organism Metabolic Pathways

- Integrate domain knowledge into Pathway/Metagenome Database
- Allow disparate data modalities to be compared: 16S rRNA, (meta)genomics, transcriptomics, metabolomics, etc.
- Model analysis drives hypothesis generation

(Wikipedia)
# Outline

1. Introduction
2. Pipeline and Model Set Description
3. Comparison of PMGDBs
4. Vignettes
   - Vignette #1: Metabolite Tracing
   - Vignette #2: Chimeric Pathways
   - Vignette #3: Pharmaco-Microbiomics
   - Vignette #4: Analysis of Mass-Spec Data
5. Conclusion
6. Extras
Gut microbiome mediates complex biochemical transformations
- Gut microbiome mediates complex biochemical transformations
- Need for computational database of small molecules, reactions, and pathways
- Gut microbiome mediates complex biochemical transformations
- Need for computational database of small molecules, reactions, and pathways
- Analyzed stool sample data from Human Microbiome Project [2] and MetaHIT [3].
• Gut microbiome mediates complex biochemical transformations
• Need for computational database of small molecules, reactions, and pathways
• Analyzed stool sample data from Human Microbiome Project [2] and MetaHIT [3].
• Metagenome annotations used to build 270+ pathway / metagenome databases (PMGDBs) using Pathway Tools software [1].
Gut microbiome mediates complex biochemical transformations

Need for computational database of small molecules, reactions, and pathways

Analyzed stool sample data from Human Microbiome Project [2] and MetaHIT [3].

Metagenome annotations used to build 270+ pathway/metagenome databases (PMGDBs) using Pathway Tools software [1].

- Duplicate annotations clustered to reduce database size
Gut microbiome mediates complex biochemical transformations

Need for computational database of small molecules, reactions, and pathways

Analyzed stool sample data from Human Microbiome Project [2] and MetaHIT [3].

Metagenome annotations used to build 270+ pathway / metagenome databases (PMGDBs) using Pathway Tools software [1].

- Duplicate annotations clustered to reduce database size
- Pathway Tools PathoLogic module ran with “Taxonomic Pruning” turned off
- Gut microbiome mediates complex biochemical transformations
- Need for computational database of small molecules, reactions, and pathways
- Analyzed stool sample data from Human Microbiome Project [2] and MetaHIT [3].
- Metagenome annotations used to build 270+ pathway / metagenome databases (PMGDBs) using Pathway Tools software [1].
  - Duplicate annotations clustered to reduce database size
  - Pathway Tools PathoLogic module ran with “Taxonomic Pruning” turned off
  - Two days of processing on Stanford FarmShare cluster (10 machines, 8-cores each)
Pipeline

Filter annotations for enzymatic, transport proteins

Run Pathway Tools on Stanford FarmShare cluster

HMP stool samples (148) → MetaHIT stool samples (125) → Filter annotations for enzymatic, transport proteins → Run Pathway Tools on Stanford FarmShare cluster → GutCyc, Sample1_Cyc, Sample(N)_Cyc

Figure: Diagram of method used to build GutCyc PMGDBs.
Bioinformatic Mapping

Figure: Details of annotation canonicalization method.
Summary Statistics

- Meaningful annotations from HMP & MetaHIT: 52,263,551
- Annotation clusters in GutCyc: 20,894

<table>
<thead>
<tr>
<th>P(M)GDB</th>
<th># compounds</th>
<th># reactions</th>
<th># pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>MetaCyc</td>
<td>10,150</td>
<td>11,703</td>
<td>2,440</td>
</tr>
<tr>
<td>GutCyc</td>
<td>6,271</td>
<td>7,080</td>
<td>1,594</td>
</tr>
<tr>
<td>EcoCyc</td>
<td>2,497</td>
<td>2,085</td>
<td>322</td>
</tr>
</tbody>
</table>
Histogram of Model Compound Counts

Sample PMGDB Compound Distribution

- HMP
- MetaHIT

# Compounds
Frequency
0 2000 4000 6000
0 20 40
Model Enzymatic Reaction Histogram

Sample PMGDB Enzymatic Reaction Distribution

- HMP
- MetaHIT

Frequency vs. # Enzymatic Reactions
Model Transport Reaction Histogram

Sample PMGDB Transport Reaction Distribution

- HMP
- MetaHIT

Frequency vs. # Transport Reactions
Model Pathway Histogram

Sample PMGDB Pathway Distribution

- HMP
- MetaHIT

Frequency vs. # Pathways
Figure: Upper-left section of GutCyc Cellular Overview, showing compounds, transporters, reactions, pathways, and pathway classes.
**Figure:** Zoom-out of GutCyc Cellular Overview, showing compounds, transporters, reactions, pathways, and pathway classes.
1. Introduction

2. Pipeline and Model Set Description

3. Comparison of PMGDBs

4. Vignettes
   - Vignette #1: Metabolite Tracing
   - Vignette #2: Chimeric Pathways
   - Vignette #3: Pharmaco-Microbiomics
   - Vignette #4: Analysis of Mass-Spec Data

5. Conclusion

6. Extras
Reaction Principal Coordinate Analysis

Metabolic Reaction PCoA

Legend
- HMP
- MetaHIT CD
- MetaHIT No IBD
- MetaHIT UC
- Unmapped

PC1 (56%) - PC2 (12%)
Clinical Phenotype Comparison

**Figure:** Venn diagram of pathway sets among three groups of MetaHIT subjects: Crohn’s Disease \((n = 12)\), Ulcerative Colitis \((n = 27)\), and those without Inflammatory Bowel Disease \((n = 85)\). The difference in the distribution of pathways is statistically significant \((p < 2.2 \times 10^{-16})\) between all sets except for between UC and non-IBD subjects (Wilcoxon signed rank).
Outline

1. Introduction
2. Pipeline and Model Set Description
3. Comparison of PMGDBs
4. Vignettes
   - Vignette #1: Metabolite Tracing
   - Vignette #2: Chimeric Pathways
   - Vignette #3: Pharmaco-Microbiomics
   - Vignette #4: Analysis of Mass-Spec Data
5. Conclusion
6. Extras
The gut microbiome is capable of transforming excess L-carnitine into trimethylamine, which is further processed by the liver to form the cardiovascular disease-associated metabolite TMAO [4].
Optimal Tracing of Dietary Compounds to Disease Biomarkers

- The gut microbiome is capable of transforming excess L-carnitine into trimethylamine, which is further processed by the liver to form the cardiovascular disease-associated metabolite TMAO [4].
- The Pathway Tools Metabolic Route Search tool was used to find an optimal path between L-carnitine to trimethylamine (TMA),
The gut microbiome is capable of transforming excess L-carnitine into trimethylamine, which is further processed by the liver to form the cardiovascular disease-associated metabolite TMAO [4].

The Pathway Tools Metabolic Route Search tool was used to find an optimal path between L-carnitine to trimethylamine (TMA),

Path found used MetaCyc “carnitine degradation II” pathway (PWY-3602, expected in *Proteobacteria*) along with a betaine reductase reaction (EC 1.21.4.4; found in *Clostridium sticklandii* and *Eubacterium acidaminophilum*), minimizing the number of enzymes involved and chemical bond rearrangements.
**Optimal Tracing of Dietary Compounds to Disease Biomarkers**

Figure: Optimal path between L-carnitine and trimethylamine as found in GutCyc using the Pathway Tools Metabolic Route Search.
**Searching for Chimeric Pathways**

- **Definition:** known metabolic pathways with enzymes from two or more organisms.
- **Important for identifying** emergent metabolic properties of microbial communities
- **Further understanding of** eventual fate of ingested nutrients and pharmaceuticals
1. Start with GutCyc pathways with two or more enzymatic reactions that are not pathway holes \((n = 1,057)\)
Pipeline Progress

1. Start with GutCyc pathways with two or more enzymatic reactions that are not pathway holes (n = 1,057)

2. Remove all pathways present in 180 HMP Gut Isolate PGDBs (remaining: n = 512)
Pipeline Progress

1. Start with GutCyc pathways with two or more enzymatic reactions that are not pathway holes ($n = 1,057$)
2. Remove all pathways present in 180 HMP Gut Isolate PGDBs (remaining: $n = 512$)
3. Keep pathways catalyzed by two or more HMP Gut Isolate PGDBs (remaining: $n = 189$)
Keep pathways where two or more enzymes have divergent nucleotide composition
Chimeric Pathways, Next Steps

1. Keep pathways where two or more enzymes have divergent nucleotide composition
2. Keep pathways that have supporting transporter evidence
Chimeric Pathways, Next Steps

1. Keep pathways where two or more enzymes have divergent nucleotide composition
2. Keep pathways that have supporting transporter evidence
3. Rank remaining pathways for tracer experiment priority
The Role of the Microbiome in Drug Metabolism

- Accurate drug dosing stymied by inter-patient variability
The Role of the Microbiome in Drug Metabolism

- Accurate drug dosing stymied by inter-patient variability
- Fields like pharmacogenetics seek to explore sources of variability
The Role of the Microbiome in Drug Metabolism

- Accurate drug dosing stymied by inter-patient variability
- Fields like pharmacogenetics seek to explore sources of variability
- Hypothesis: Microbial enzymes modulate GI drug absorption
Set Analysis

1. Drugbank: 7,678 entries
   - Entries with drug route information: 1,419
     - ... with oral or rectal routes: 899
2. GutCyc compounds: 6,271
   - Entries with links to DrugBank: 198
3. Intersection: 29 compounds
Six Examples

- Hydrocortisone
- Bethanidine
- Morphine
- Pyrazinamide
- Flucytosine
Hydrocortisone

- **DrugBank**: DB00741
- **MetaCyc**: CORTISOL
- **Indications**: Immune & allergy disorders (e.g., ulcerative colitis & Crohn’s disease)
- **Annotation frequency**: 173
Reaction: Steroid 11beta-monooxygenase
Hydrocortisone Enzyme Abundance

Distribution of Hydrocortisone–related Enzyme Abundance

Enzyme(s) as KEGG KO IDs

K07433

K00497

Enzyme Abundance (HMMRC)
High throughput mass-spectrometry data

- Mass-spec data from Wei-Ting’s analysis of Les’ samples
- Data file with 54,980 unique masses
- Only 1,132 masses identified
Pathway Tools and High-Throughput Data

Relevant features of Pathway Tools:

- High-level visualizations
- In-pathway data display
- Data set storage and manipulation
Visualization: Cellular Overview Highlight

Human gut metagenome (Human Microbiome Project)
Enrichment / Depletion Analysis

- Answers the question, “do I see more or less than what I expect from chance?”
- Fisher’s Exact Test
- P-Value Cutoff: 0.001
- Bonferroni correction
Enrichment Analysis Results

- superpathway of mycolate biosynthesis ($7 \times 10^{-11}$)
- Hormones biosynthesis ($9 \times 10^{-8}$)
- Secondary Metabolites Biosynthesis ($9 \times 10^{-7}$)
Outline

1. Introduction
2. Pipeline and Model Set Description
3. Comparison of PMGDBs
4. Vignettes
   - Vignette #1: Metabolite Tracing
   - Vignette #2: Chimeric Pathways
   - Vignette #3: Pharmaco-Microbiomics
   - Vignette #4: Analysis of Mass-Spec Data
5. Conclusion
6. Extras
Understanding of human microbiome metabolism critical for understanding many diseases
Summary

- Understanding of human microbiome metabolism critical for understanding many diseases
- GutCyc is an interactive pathway / metagenome database enabling metabolic analyses
Summary

- Understanding of human microbiome metabolism critical for understanding many diseases
- GutCyc is an interactive pathway / metagenome database enabling metabolic analyses
- Crohn’s Disease subjects have distinct pathway distributions from Ulcerative Colitis and non-IBD subjects.
Summary

- Understanding of human microbiome metabolism critical for understanding many diseases
- GutCyc is an interactive pathway / metagenome database enabling metabolic analyses
- Crohn’s Disease subjects have distinct pathway distributions from Ulcerative Colitis and non-IBD subjects.
- Chimeric pathways
Summary

- Understanding of human microbiome metabolism critical for understanding many diseases
- GutCyc is an interactive pathway / metagenome database enabling metabolic analyses
- Crohn’s Disease subjects have distinct pathway distributions from Ulcerative Colitis and non-IBD subjects.
- Chimeric pathways
- Trace metabolic route between dietary compounds and disease biomarkers
Summary

- Understanding of human microbiome metabolism critical for understanding many diseases
- GutCyc is an interactive pathway / metagenome database enabling metabolic analyses
- Crohn’s Disease subjects have distinct pathway distributions from Ulcerative Colitis and non-IBD subjects.
- Chimeric pathways
- Trace metabolic route between dietary compounds and disease biomarkers
- Pharmaco-Microbiomics
Summary

- Understanding of human microbiome metabolism critical for understanding many diseases
- GutCyc is an interactive pathway / metagenome database enabling metabolic analyses
- Crohn’s Disease subjects have distinct pathway distributions from Ulcerative Colitis and non-IBD subjects.
- Chimeric pathways
- Trace metabolic route between dietary compounds and disease biomarkers
- Pharmaco-Microbiomics
- Mass Spec data analysis
Summary

- Understanding of human microbiome metabolism critical for understanding many diseases
- GutCyc is an interactive pathway / metagenome database enabling metabolic analyses
- Crohn’s Disease subjects have distinct pathway distributions from Ulcerative Colitis and non-IBD subjects.
- Chimeric pathways
- Trace metabolic route between dietary compounds and disease biomarkers
- Pharmaco-Microbiomics
- Mass Spec data analysis
What Next for GutCyc?

- Positive response at Keystone Symposium
- Drafting manuscript describing resources
- Feedback?
Acknowledgments

- Dill Group
- Relman Lab
- Dr. Craig Criddle
- Bioinformatics Research Group, SRI
- Funding from NIH Biotechnology Grant & KAUST
Thank you for your time!
taltman1@stanford.edu
www.gutcyc.org
Bibliography


Guiding Metaphor

Modeling the human gut as a bioreactor provides a novel perspective for the analysis of digestion, disease, and the design of medical interventions.

Figure: (Wikipedia)
Construction of a Computational Metabolic Model of the Human Gut Microbiome
Tomer Altman, David A. Relman, and David L Dill.
The Guilded Age: A Chalk Talk

- Bioreactor vs. Waste-Water treatment plant
- Modeling granularity
- Which Guilds?
- Abundance of Guilds?
- Stoichiometric Analysis
- Simulation