Data Mining and Modeling of the Human Gut Microbiota using Pathway Tools

Tomer Altman
taltman1@stanford.edu

Biomedical Informatics, Stanford University
The human body contains and is covered by thousands of microbes.
Introduction

The Human Microbiome

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

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

Microbiome involved in obesity, irritable bowel syndrome, gingivitis, and cancer.
The Human Microbiome

- The human body contains and is covered by thousands of microbes
- Paradigm shift: from pathogenicity to symbiosis ("super-organism")
- Microbiome involved in obesity, irritable bowel syndrome, gingivitis, and cancer
- Understanding the function of the microbial communities in health and disease is a grand challenge
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)
Specific Aims:

1. Develop data mining methods for analyzing human distal gut high-throughput datasets
Preliminary Dissertation Proposal

Specific Aims:

1. Develop data mining methods for analyzing human distal gut high-throughput datasets
2. Construct a metabolic bioreactor model of the human distal gut
Preliminary Dissertation Proposal

Specific Aims:

1. Develop data mining methods for analyzing human distal gut high-throughput datasets
2. Construct a metabolic bioreactor model of the human distal gut
3. Apply flux balance analysis to the reconstructed metabolic model
Specific Aim #1: Data Mining

Develop data mining methods for analyzing human distal gut high-throughput datasets

Example: A novel enzymatic distance measure for analyzing metagenomic data. Complements 16S-based measures such as UniFrac.
## Scale of HMP Metagenomic Data

<table>
<thead>
<tr>
<th>Data</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples</td>
<td>139</td>
</tr>
<tr>
<td>Annotation Files</td>
<td>33G</td>
</tr>
<tr>
<td>Genes</td>
<td>$27.8 \times 10^6$</td>
</tr>
<tr>
<td>Unique MetaCyc Reactions</td>
<td>3388</td>
</tr>
</tbody>
</table>
MetaCyc Reactions As Distance Measure

Figure: PCoA with cosine similarity over enzyme abundance: First two components as axes.
Enzyme Copy Number Variation

Figure: Exponential distribution of enzyme copy numbers.
HMP Stool Sample PGDB

**Figure:** Cellular Overview of Pathway/Genome Database built from HMP metagenome sample SRS011405.
Figure: Neighboring fermentation pathways have contrasting robustness to enzyme copy number variation.
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.

(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.
- Analysis of model drives hypothesis generation

(Wikipedia)
Specific Aim #2: Model Construction

Construct a bioreactor model of the human distal gut

A coarse-grained description of the major in-flows and out-flows of a gut microbe commonly used to analyze bioreactors:

glucose and ammonia → biomass, carbon dioxide, water, and a short-chain fatty acid

\[ C_6H_{12}O_6 + bNH_3 \rightarrow cCH_{1.79}O_{0.5}N_{0.2} + dCO_2 + eH_2O + gCH_7\textsubscript{4}O_{\frac{1}{2}} \]

For \( b = 0.26, \ c = 2.6, \ d = 0.67, \ e = 2.9, \) and \( g = 1.3, \) colonic bacteria consume \( 197 \frac{kcal}{day} \), or 8% to 9% of daily diet.
Specific Aim #3: Flux Balance Analysis

Apply flux balance analysis to the reconstructed metabolic model

Figure: Flux balance analysis modeling the first several reactions of the glycolysis pathway (Wikipedia)
Introduction

Questions?
Specific Aim #2: Model Construction

Parameterize a bioreactor model of the human distal gut using physiological data and metabolic modeling:

An *in silico* model of the human distal gut:

Figure: An analogous model: Simulator of the Human Intestinal Microbial Ecosystem (SHIME). Nutrition.org.