The virome of the human gut: metagenomic analysis of changes associated with diet

James Lewis
Gary Wu
Frederic Bushman

Diet, Genetic Factors, and the Gut Microbiome in Crohn’s Disease

University of Pennsylvania
Diet, Genetic Factors, and the Gut Microbiome in Crohn’s Disease

COMBO: Cross-Sectional Study of Diet and Stool Microbiome

CAFE: Controlled Feeding Experiment

Study virome under controlled feeding
COMBO

97 subjects, assess diet with food frequency questionnaire, 24 hr. recall questionnaire
454 tag sequencing of 16S regions
Conclusions from COMBO

• At an FDR of 25%, ~30-40 nutrients correlated with changes in bacterial populations

• Correlations between food groups and Pyla detected, but considerable divergence among deeper taxa

• Correlated effects on microbiome among BMI, dietary fat, and percent calories from saturated fatty acids

• Though effects were significant, measured dietary effects accounted for a small fraction of the total variation among subjects

Follow up in CAFÉ: 1) fat versus fiber, 2) total calories
CAFE1: Controlled Feeding Experiment

- 10 healthy volunteers
- Randomized to high fat vs. low fat diet
- 10 day inpatient stay with same meals each day
- Caloric intake adjusted to maintain current weight
- Daily stool sample collection
- Rectal biopsies on days 1 and 10
- Sitz marker study to assess transit time
- Sequencing: 16S tags, shot-gun metagenomic sequencing of total DNA and viral DNA, targeted analysis of Archaea and Eukarya
Longitudinal analysis of microbiome under controlled feeding

Unweighted Unifrac Graphic from QIIME
Conclusions from CAFÉ 1 and 2

• Inter-individual variation predominates.

• Bacterial populations change within 24 hours of initiating controlled diet.

• High fat versus fiber has detectable effect (CAFE1), increased calories has detectable effect (CAFE2).

• Considerable longitudinal drift during stay in hospital in all groups. Specific vitamins? Hospital environment?
Approaches to virome analysis

Multiple possible goals:

• Characterize overall viral communities
• Hunt for new viruses linked to disease
• Characterize populations of a specific virus
Viral Analysis on CAFE1 samples

- Purification of viral particles: filtration, CsCl gradient, DNAse digestion.
- Quantitative recovery of phage $\lambda$ analyzed as a control.
- Greatly reduced 16S rDNA.
- Multiple displacement amplification.
- Shot-gun sequencing, 454 Titanium, 992,309 reads, median length 380 nt.

<table>
<thead>
<tr>
<th>PID</th>
<th>Day 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Diet</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>Low Fat</td>
</tr>
<tr>
<td>2012</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low Fat</td>
</tr>
<tr>
<td>2016</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High Fat</td>
</tr>
<tr>
<td>2020</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low Fat</td>
</tr>
<tr>
<td>2019</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>High Fat</td>
</tr>
<tr>
<td>1013</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ad lib</td>
</tr>
</tbody>
</table>

10^{10} phage per gram of stool

Sam Minot et al.
Assembly of viral sequence reads

- Newbler assembler
  - 40bp overlap
  - 90% sequence identity
- 7,175 contigs >500 bp
- 86.6% of reads in contigs
- Custom code to allow circular assembly
- PHACCS: median species richness 44 (range=19-785)
Phage metagenomics

Showing 46.04 kbp from contig07718, positions 1 to 46,043

[Image of a website page showing a genetic sequence analysis interface with various genetic annotations and identifiers.]
Gut virome characterization

Also found 22 CRISPR arrays, one example of CRISPR spacer targeting another virus in the same individual.
Comparative metagenomics: viruses are parasites
Lysogeny

A

Phage Protein (1,481)
Bacterial Genome (37)
Prophage Proteins (428)
Integrate (73)
VLP Contigs (3,029)

C

Multiple (n=1176)
Other (n=19)
Bacteroidetes (n=24)
Proteobacteria (n=447)
Firmicutes (n=1148)

B

Phage Tail
Phage Baseplate
Reverse Transcriptase

Read count
Position (Mb)
contig09477

Faecalibacterium prausnitzii L2/6

D

Phage integrase
ADP-ribosylglycohydrolase

Read count
Position (Mb)
contig09505

Parabacteroides distasonis ATCC 8503
Interpersonal variation

- Rows: samples
- Columns: contigs
- Clustering by individual
Variation associated with subject and diet

- People are more similar to themselves
- People on the same diet become more similar over time
- Specify the contigs involved

P values by label permutation
Procrustes analysis: covariation of host and viral communities
What determines phage abundance?

- Abundance of host?
- Lysogenic induction?

In some cases the abundance of phage does not correlate with apparent abundance of host: possibly indicative of differential induction?
Summary

• Variation in both bacterial and viral communities dominated by interpersonal variation
• Dietary intervention associated with altered composition of both bacterial and viral communities
• Metagenomic analysis over six subjects yielded 7000 viral contigs; 19-785 types per sample (minimum estimate)
• Both expected and novel functionality, one example being viral CRISPRs targeting other viruses
Credits

UPenn/CHOP Group
Gary Wu*
James Lewis*
Frederic Bushman*
Sam Minot
Emily Charlson
Christian Hoffmann
Scott Sherrill-Mix
Serena Dollive
Kyle Bittinger
Rohini Sinha
Jennifer Hwang
Rebecca Custers-Allen
Stephanie Grunberg

Bob Baldassano
Sue Keilbaugh
Ying-Yu Chen
Lisa Nessel
Jun Chen
Hongzhi Li

* co-PIs

Collaborators
Rob Knight
Catherine Lozupone
Dan Knights