Viral Metagenomics

Lori Holtz, MD
Pediatric Gastroenterology
Disclosures

• No conflicts of interest to disclose
What is “Metagenomics”? 

• The “Metagenome”
  – Handelsman et al Chemistry and Biology (1998)

  “…..advances in molecular biology and eukaryotic genomics, which have laid the groundwork for cloning and functional analysis of the collective genomes of soil microflora, which we term the metagenome of the soil.”

• Metagenomics
  – Aggregate nucleic acid sequences in a sample
  – Studies in the absence of specific culture or targeting
  – “Whole community genome analysis”
  – Bacterial
  – Viral
Defining our metagenome

- Stool metagenomics
- Human microbiome project
- Stool viral metagenomics (stool virome)
Uses of metagenomics

• Pathogen discovery (Part 1)
• Define our microbial communities (Part 2)
Why Search for Novel Viruses?

1. Many unexplained diseases
2. Novel insights from novel viruses
Existing Viral Detection/Discovery Methods Have Limitations

- **Viral culture**
  - Many viruses not culturable, specific cell types needed
- **Electron Microscopy**
  - High titers, many viruses with similar morphology
- **Immunoassays**
  - Require special reagents, candidate viruses
- **PCR**
  - Limited breadth (candidate viruses)

*Solution:* Develop comprehensive, unbiased, and high-throughput methods.
*Metagenomics offers key*
Why study diarrhea?

<table>
<thead>
<tr>
<th>Leading Causes of Death Due to Infectious Diseases</th>
<th>Number of Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Respiratory Tract Infections</td>
<td>3.46 million</td>
</tr>
<tr>
<td><em>Diarrheal Diseases</em></td>
<td><strong>2.46 million</strong></td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>1.78 million</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>1.34 million</td>
</tr>
</tbody>
</table>

A significant fraction of diarrhea cases are of unexplained etiology

4 Major viral causes of diarrhea:

- Rotavirus
- Norovirus
- Adenovirus
- Astrovirus

~40% of diarrhea is of unknown etiology

Are these unexplained cases due to novel viruses?

Could complex viromes be responsible for these unknown cases?
How do we study the stool virome?

1. Filter sample
2. Extract nucleic acid
3. Randomly amplify by PCR
4. Next Generation Sequence
5. Custom bioinformatics pipeline

- Eukaryotic Viruses
- Bacteria
- Host
- Fungi
- Phage
- Other
- Unassigned

All viral reads

Novel
Known
Pipeline

Total raw sequences from sample

CD-Hit to define unique sequences

RepeatMasker and quality filter

BLASTn vs. host genome

Hits $e < e^{-10}$ → Host

No hits or hits $e > e^{-10}$

BLASTn vs. NCBI nt DB

Hits $e < e^{-10}$

No hits or hits $e > e^{-10}$

BLASTx vs. NCBI nr DB

Hits $e < e^{-5}$ → Classify using NCBI Taxonomy DB

No hits or hits $e > e^{-10}$

Unassigned, Bacteria, Fungus, Virus, Phage, Ambiguous

Zhao et al, PlosOne 2013
Samples for viral discovery

Pediatric diarrhea

- Negative for known pathogens
- Global collaborations:
Evidence for many novel viruses in diarrhea specimens

- Prior to 2005 the rate of viral discovery was slow
- 1975-2005: <10 viruses discovered associated with diarrhea
- >1000 diarrhea samples sequenced from around the world
- Many novel viruses:
## Evidence for many novel viruses in diarrhea specimens

<table>
<thead>
<tr>
<th>Novel virus</th>
<th>Viral family</th>
<th>RNA or DNA virus</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astrovirus MLB1</td>
<td>Astroviridae</td>
<td>RNA</td>
<td>Finkbeiner et al, 2008</td>
</tr>
<tr>
<td>Astrovirus VA1</td>
<td>Astroviridae</td>
<td>RNA</td>
<td>Finkbeiner et al, 2009</td>
</tr>
<tr>
<td>Astrovirus MLB3</td>
<td>Astroviridae</td>
<td>RNA</td>
<td>Jiang et al, 2013</td>
</tr>
<tr>
<td>Astrovirus VA4</td>
<td>Astroviridae</td>
<td>RNA</td>
<td>Jiang et al, 2013</td>
</tr>
<tr>
<td>Astrovirus VA5</td>
<td>Astroviridae</td>
<td>RNA</td>
<td>Meyer et al, in preparation</td>
</tr>
<tr>
<td>Cosavirus E1</td>
<td>Picornaviridae</td>
<td>RNA</td>
<td>Holtz et al, 2008</td>
</tr>
<tr>
<td>Klassevirus</td>
<td>Picornaviridae</td>
<td>RNA</td>
<td>Holtz et al, 2009</td>
</tr>
<tr>
<td>Rosavirus 2</td>
<td>Picornaviridae</td>
<td>RNA</td>
<td>Lim et al, 2014</td>
</tr>
<tr>
<td>Novel Amdovirus</td>
<td>Parvoviridae</td>
<td>DNA</td>
<td>Unpublished</td>
</tr>
</tbody>
</table>
Flowchart Following Novel Virus Discovery

Screen Samples for Novel Viruses

Novel Virus Identified

Epidemiologic Characterization
- Diagnostic assays
- Prevalence
- Case-control association studies
- Seroepidemiology
- Disease association

Virologic Characterization
- Genome sequencing/comparisons
- Characterize individual proteins (e.g., transformation potential, toxin activity)
- Culture virus
- Animal models
Astrovirus background

• Known cause of human diarrhea

• 1 human species (1975)
  – Eight closely related serotypes

• Extent of infection in human populations varies greatly by serotype
  – Astrovirus serotype 1 (90%)
  – Astrovirus serotype 7 (10%)

• We have discovered 8 new astroviruses
8 novel astroviruses found in human diarrhea

Meyer et al, in preparation
Discovery of novel astrovirus MLB1

• Stool sample:
  – collected in 1999
  – 3 yr old boy
  – acute diarrhea
  – Melbourne, Australia
  – negative for known pathogens
Characterization of novel viruses

Determine spectrum of viruses in sample

Characterization of novel virus(es)

Epidemiological studies of novel viruses

Molecular studies of novel viruses

How frequently is the virus found in the population?

Is the virus associated with disease?

How prevalent are antibodies to the virus?
Astrovirus MLB1 is present globally

Overall prevalence is in diarrhea is 1-2%
What are the potential roles of MLB1?

- Agent of diarrhea?
- Dietary finding?
- Is a pathogen outside of the GI tract, but is shed in stool (i.e. polio)?
Is MLB1 associated with diarrhea?

• Volunteer ingestion is no longer allowable
• 2\textsuperscript{nd} best alternative is case control study
  • India
  • Africa
Stool samples: cases and controls

- 3 year longitudinal birth cohort study (n=370)
- 27,000 surveillance samples (controls)
- 1800 diarrhea samples (cases)

- Samples cannot leave India
Each subject serves as their own control

- 27,000 controls and 1800 diarrhea samples available
- Power calculation: 400 diarrhea samples and 400 controls

Subject X

Week  
-6  
-4  
-2  
0

At least 6 weeks between diarrhea and control sample
Results of case control study of astroviruses

<table>
<thead>
<tr>
<th></th>
<th>Classic Astrovirus Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases (n=400)</td>
<td>14</td>
</tr>
<tr>
<td>Controls (n=400)</td>
<td>4</td>
</tr>
<tr>
<td>p-value (McNemar)</td>
<td>0.03</td>
</tr>
</tbody>
</table>

---

2nd MLB1 Case Control Study

- Samples collected from the Global Enteric Multicenter Study (GEMS)
  Field sites: Kenya and The Gambia

- Cases: moderate-severe diarrhea
  \( \geq 3 \) loose stools in prior 24 hours
  New episode (onset after \( > 7 \) diarrhea free days)
  +1 of following: sunken eyes, loss of skin turgor, IV hydration prescribed, blood in stools or admission to the hospital due to diarrhea

- Controls:
  Excluded if episode of diarrhea in prior 7 days
Results of astrovirus case control study

<table>
<thead>
<tr>
<th></th>
<th>Kenya</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases n=184</td>
</tr>
<tr>
<td>HAstV</td>
<td>7 (3.8%)</td>
</tr>
</tbody>
</table>

- Classic astroviruses not associated with diarrhea
- MLB1 is associated with diarrhea

Meyer et al, in preparation
Case control studies fail these pathogens

- *Campylobacter*
- *Giardia*
- Diarrheagenic *E.coli*
Does MLB1 have a role in human disease?

• Agent of diarrhea?
• Protective effect?
• Dietary finding?
• Has a role outside of the GI tract, but is shed in stool (i.e. polio)?
Characterization of novel viruses

Determine spectrum of viruses in sample

Characterization of novel virus(es)

Epidemiological studies of novel viruses

How frequently is the virus found in the population?

Is the virus associated with disease?

How prevalent are antibodies to the virus?
What is needed to determine seroprevalence of MLB1?

• Develop assay:
  – Indirect ELISA
  – Capture antigen (protein of interest)
    • Recombinant capsid protein
      (Dr. Schultz-Cherry)
  – Positive and negative controls
    • Pre and post immune rabbit serum

• Human sera to test
  – 400 age stratified healthy volunteer serum samples (Dr. Belshe)
First evidence of human antibody to MLB1

100% seropositivity by adulthood!

Holtz et al, Clin Vacc Immunol, 2014
Uses of metagenomics

- Pathogen discovery (Part 1)
- Define our microbial communities (Part 2)
Why do we care about metagenomics?

- Traditional thought: 1 organism = 1 disease
- Change in paradigm: Metagenome role in disease
- Bacterial Stool Microbiome: Changes with Disease
- Dysbiosis = shift in relative abundance of bacterial taxa

Spor et al, 2011
What Influences Bacterial Microbiome?

• Diet: Bacterial microbiome differs with diet
• Geography: Bacterial microbiome differs with geography
• A variety of diseases states prevalence varies with geography
  – Allergies
  – Autoimmune conditions
  – Inflammatory bowel disease
  – Environmental enteropathy
  – Oral vaccine efficacy

• Does the virome play a role in these processes?
Our understanding of the stool eukaryotic virome is limited

<table>
<thead>
<tr>
<th>Paper</th>
<th>Specimen Type</th>
<th>Number of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al 2006</td>
<td>Healthy adults</td>
<td>3</td>
</tr>
<tr>
<td>Finkbeiner, Allred et al 2008</td>
<td>Pediatric diarrhea</td>
<td>12</td>
</tr>
<tr>
<td>Victoria et al 2009</td>
<td>Nonpolio acute flaccid paralysis</td>
<td>35</td>
</tr>
<tr>
<td>Van Leeuwen et al 2010</td>
<td>Diarrhea</td>
<td>13</td>
</tr>
<tr>
<td>Phan et al 2012</td>
<td>Pediatric diarrhea (West Africa)</td>
<td>98 (in pairs)</td>
</tr>
</tbody>
</table>
Goals of stool viral metagenomic study

• To systematically define known and novel viruses in diarrhea
• Determine if the diarrhea virome differs with geography
  – Large population study
  – State of the art technology
What samples should be examined?

• Prospectively collected, single PI
• Inclusion criteria:
  – Children <5 years old
  – Hospitalized for diarrhea
  – Samples randomly selected from all available cases
• Melbourne, Australia
  – Westernized, urban
• Northern Territory, Australia
  – Sparse, poor access to healthcare
# Demographics of cohorts

<table>
<thead>
<tr>
<th></th>
<th>Melbourne (n=43)</th>
<th>Northern Territory (n=44)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average age (months)</td>
<td>25.4</td>
<td>15.7</td>
<td>0.0013</td>
</tr>
<tr>
<td>Gender (female)</td>
<td>23 (53%)</td>
<td>22(50%)</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Method to study the stool virome

1. Filter sample
2. Extract nucleic acid
3. Randomly amplify by PCR
4. Next Generation Sequence (454)
5. Custom bioinformatics pipeline

- Eukaryotic
- Bacteria
- Host
- Fungi
- Phage
- Other
- Unassigned

All viral reads
- Known
- Novel
Virome Definition

• Viral diversity = presence or absence of sequences from 95 virus families defined by International Committee on Taxonomy of Viruses (ICTV)
• Virome = number and type of viral taxa with at least one representative sequence present in the sample
• Conservative criteria-likely underestimates
What have we found in diarrhea?

- $25/87$ (29%) samples with sequence from the 4 canonical diarrhea families
- $57/87$ (66%) samples have evidence of virus
- $35/87$ (40%) samples have evidence for 2 or more viruses
- 22 different viral families detected
What kind of viruses are present?

- Classic diarrhea viruses
  - Rotavirus, noroviruses, adenoviruses, astroviruses
- Known human enteric viruses
  - Enteroviruses, circoviruses, anelloviruses
- Known, but unexpected human viruses
  - Influenza, parainfluenza
- Plant viruses
- Novel virus!
  - Dicistrovirus (28% AA identity to insect virus Israeli acute paralysis virus)
The number of viral families is greater in the Northern Territory

Wilcoxon p=0.0002
Adjusting for age p=0.001

Are there specific viral families driving this difference?

Holtz et al, Virology 2014
Specific viral families differ in Melbourne vs. Northern Territory

* p<0.05
Enterovirus most commonly detected

<table>
<thead>
<tr>
<th>Virus Class</th>
<th>Northern Territory n=44</th>
<th>Melbourne n=43</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovirus</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Enterovirus</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>Klassevirus</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Kobuvirus</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Parechovirus</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Enterovirus statistically significant (p<0.001 (Fisher’s Exact))
Virome difference confirmed in extended cohorts by PCR

• Question: Can the differences in the virome be seen in more samples from Melbourne and the Northern Territory?

• Samples:
  – Melbourne cohort (n=159)
  – Northern Territory cohort (n=165)

• Assays:
  – Pan-enterovirus qRT-PCR (Verstrepen et al, 2001)
  – Pan-adenovirus qPCR (Jothikumar et al, 2005)

<table>
<thead>
<tr>
<th></th>
<th>Adenovirus Positive</th>
<th>Enterovirus Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melbourne (n=159)</td>
<td>10 (6.3%)</td>
<td>1 (0.6%)</td>
</tr>
<tr>
<td>Northern Territory (n=165)</td>
<td>31 (18.8%)</td>
<td>29 (18.0%)</td>
</tr>
</tbody>
</table>
Are these differences in the virome seen in other locations?

• **Goal**: To determine if the difference in enteroviruses and adenoviruses is seen in other geographic locations?

• **Assays**:
  – Pan-enterovirus qRT-PCR (Verstrepen et al, 2001)
  – Pan-adenovirus qPCR (Jothikumar et al, 2005)

• **Samples to test**:
  – Seattle cohort (n=80) (Tarr)
    • Children <5y/o with diarrhea evaluated in ER 2003-2005
  – Gambia cohort (n=160) (Stine)
    • Children <5 y/o with diarrhea in Upper River Division (rural) 2008-2009
Results of PCR screening

<table>
<thead>
<tr>
<th>Region</th>
<th>Adenovirus Positive</th>
<th>Enterovirus Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melbourne (n=159)</td>
<td>10 (6.3%)</td>
<td>1 (0.6%)</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>31 (18.8%)</td>
<td>29 (18.0%)</td>
</tr>
</tbody>
</table>

Enteroviruses more prevalent in Northern Territory and The Gambia compared to Melbourne and Seattle

Adenovirus most prevalent in The Gambia and least prevalent in Melbourne

Adenovirus had comparable frequencies in Seattle and Northern Territory
What are possible reasons for these differences?

• Diet?
• Living conditions?
• Water quality?
• Hygiene?
Future Questions

• What is stool virome of healthy controls?
• How does our virome develop and change over time?
• What is the role of our virome in environmental enteropathy?
• What is the role of the virome in oral polio vaccine efficacy?