Microbiome Analysis in Kawasaki Disease

Kristine M. Wylie

Microbiome Analysis in Kawasaki Disease

No relevant financial relationships exist
Evidence for a viral etiology for Kawasaki Disease

- **Epidemiologic**
  - Epidemics with geographic spread
  - Seasonal predominance in non-temperate climates

- **Clinical**
  - History of preceding respiratory symptoms
  - Failure of antibiotic therapy

- **Ultrastructural**
  - Intracytoplasmic inclusion bodies and virus-like particles detected in the bronchial epithelium of KD tissues but not controls (oligoclonal IgA)

- **Gene expression**
  - Up-regulated cytotoxic T-cell and type I IFN pathways in coronary arteries from KD patients compared with controls
    (Presented at this meeting by Anne Rowley)

Kristine Wylie (wylie_k@kids.wustl.edu)
High-throughput metagenomic shotgun sequencing (MSS)

- A powerful approach for identifying pathogens associated with diseases of unknown etiologies
  - No *a priori* knowledge of the pathogen required
  - No culturing needed
  - Insensitive to minor sequence variation that can confound molecular tests
Sample collection and processing

- Frozen and FFPE tissues used
- Nucleic acid extracted (can use DNA/RNA/TNA)
- We focused on RNA due to previous data that suggested the etiologic agent is an RNA virus

Library construction and sequencing

- cDNA made from RNA
- Sheared to uniform size
- Sequencing library adapters ligated to ends
- Data generated on Illumina HiSeq platform

Sequence analysis

- Described on next slide

Interpretation and additional follow up

2x100 paired-end sequencing: 100 bp sequenced from each end of the fragment
MSS analysis overview

Initial processing
100 base sequences
Trim low quality, mask low complexity, etc.

Assembly of overlapping sequences

Alignment against reference genomes

Classification:
human, bacterial, fungal, viral, other

Known sequences
Unclassified sequences
Related to known sequences

Alignments
- Nucleotide
- Translated amino acid
- HMM

Short sequence reads
Contig (consensus sequence)

Unclassified sequences
Identify unclassified sequences that are shared among samples

Extended analysis (reference genome independent)

Unclassified reads
- Case sample 1
- Case sample 2
- Case sample 3
- Control sample 1
- Control sample 2
- Control sample 3

Sequence alignments with RTG

New reference database:
- All unclassified contigs from all samples and all unclassified reads from all samples

Identification of sequences associated with KD

<table>
<thead>
<tr>
<th></th>
<th>Seq A</th>
<th>Seq B</th>
<th>Seq C</th>
<th>Seq D</th>
<th>Seq E</th>
<th>Seq F</th>
</tr>
</thead>
<tbody>
<tr>
<td>KD tissue 1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>KD tissue 2</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KD tissue 3</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ctr tissue 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ctr tissue 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ctr tissue 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute KD urine 1</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recovery KD urine 1</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Application of MSS to identifying viral pathogens in diseases of unknown etiologies

- Subject was 20-month-old boy with fever and petechial rash, no other pathogenic agents detected
- Plasma sequencing yielded sequences with remote similarity to astrovirus MLB-1
- While this was being investigated, we learned astrovirus MLB-2 had been discovered and was being characterized by Dave Wang’s group at WU.
- We obtained the complete genome sequence of MLB-2 from them.
- 238 reads were identified with > 80% identity to MLB-2, confirmed by specific RT-PCR
- MLB-2 was not found in 188 other samples screened by consensus astrovirus PCR
- First detection of an astrovirus in blood - unexpected pathogen

Application to Kawasaki Disease - samples sequenced

<table>
<thead>
<tr>
<th>Tissue type</th>
<th>Cases number of sample</th>
<th>Controls number of sample</th>
<th>Tissue type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary artery (CA)</td>
<td>7</td>
<td>7</td>
<td>FFPE</td>
</tr>
<tr>
<td>Serum</td>
<td>11</td>
<td>10</td>
<td>Frozen</td>
</tr>
<tr>
<td>Throat</td>
<td>10</td>
<td>4</td>
<td>Frozen</td>
</tr>
<tr>
<td>Lung</td>
<td>8</td>
<td>0</td>
<td>FFPE</td>
</tr>
<tr>
<td>Heart</td>
<td>3</td>
<td>1</td>
<td>FFPE</td>
</tr>
<tr>
<td>Spleen</td>
<td>1</td>
<td>0</td>
<td>Frozen</td>
</tr>
</tbody>
</table>

- 4.9 billion sequence reads generated
  - Minimum 32 million
  - Maximum 500 million
- Non-human sequences ranged from <1% (serum) to 99% (throat)
• Metastats does not find any differentially represented viral candidates
  • Human viruses do not associate with KD
  • Reagent or environmental contaminants
    • Parvo-like hybrid (Naccache, et al., PNAS)
• Many plant viruses
  • Also likely contaminants

Kristine Wylie (wylie_k@kids.wustl.edu)
Bacterial/fungal discovery in sequence data

- Metastats does not find any differentially represented bacterial or fungal candidates
- Many low level “background” bacteria/fungi (a smaller representative set shown here)
- Some known pathogens, but not associated with KD
- Many organisms that look like environmental or kit contaminants, but not associated with KD

Kristine Wylie (wylie_k@kids.wustl.edu)
Discovery in unclassified contig data

- Metastats used to identify contigs more common in KD than controls
- Subset only in KD and not in controls shown here
- Prioritizing for follow up
Current candidates

- Best candidates are sequences predominantly in patient samples but not controls
  - Could be highly divergent from known viruses
  - Could be very low abundance in samples
- Current work with existing data:
  - Continue to try to classify sequences with updated databases
  - Validate candidates with real-time RT-PCR assays on a larger set of cases controls
- Proceed cautiously due to confounding factors associated with pathogen discovery in MSS data
Confounding factors associated with MSS analysis

- Can find microbes (known or novel) that are contaminants from reagents, water, enzymes, etc.
- Contaminants can sometimes associate with a clinical phenotype by chance
  - Example: If control samples are typically higher biomass (signal), case samples will have a higher representation of contaminant (noise)
- While we aim for relatively unbiased representation of the metagenome, biases are introduced during sample preparation
- Classifying sequences based on similarity to reference genomes is limited by the quality of the databases
  - Missing entries
  - Mis-labeled entries
Summary and conclusions

• Most comprehensive ultra-deep sequencing study to identify a causative agent of KD to date
• No etiologic agent of KD identified yet, but we are pursuing candidates
  • We suspect that sequences of the agent are presently unclassified
  • Prioritizing ORFs associated with KD samples for further study
• Possible challenges
  • the agent is highly divergent from sequenced reference genomes
  • causes a very early viremia
  • is highly cell-associated
  • is present in very low quantity in patient tissues
  • may also be present in some controls.
• Our work highlights the challenges of identifying a new viral agent
Support for this work

- National Institutes of Health HL63771, HL109955, AI106030 (to AHR), the Max Goldenberg Foundation, and the Center for Kawasaki Disease at the Ann & Robert H. Lurie Children's Hospital of Chicago.