Host Gene Expression for Diagnosis of Infectious Diseases

Aimee K. Zaas, MD, MHS
Dept of Medicine/Institute for Genome Sciences and Policy
Duke University Medical Center
Outline

• Background

• Peripheral Blood Gene Expression for Diagnosis of Acute Respiratory Viral Illness
  • Hypotheses
  • Study Design
  • Statistical Analysis
  • Results
  • Conclusions

• Peripheral Blood Gene Expression for Diagnosis of Candidemia
  • Study Design
  • Results
  • Future Directions

• Conclusions
Gene Expression Can Discriminate Between Pathogens

A

- Microbe A
- Microbe B
- Microbe C

* Pattern Recognition Receptors

B

- Bacterium
- mRNA
- mRNA
- Expression Profile

Patient Genotype (DNA)

Analysis

TRUE CLASS
PREDICTED CLASS

Virus Bacteria

Bacteria

Virus

1 2 3 4 5 6 7 8 9 10 11

23 Samples
Can We Classify Acute Respiratory Viral Illness?

- **Hypotheses:**

  Peripheral blood gene expression at time of peak symptoms in experimentally infected cohorts can differentiate between symptomatic and asymptomatic subjects.

  The above derived peripheral blood gene expression signatures can accurately classify other subjects with viral respiratory infection and differentiate viral from bacterial infection.

- **Methodology:**

  Serial sampling and symptom scores of experimentally infected individuals.

  Unsupervised analysis of peripheral blood gene expression data at time of peak symptoms.
Human Viral Challenge Sites: HRV, RSV, Influenza A

HRV Challenge: Charlottesville, VA 11/2007

RSV Challenge (Brentwood, UK 7/2008)

Influenza Challenge (Cambridge, UK 10/2008)
Human Viral Challenges: Symptom Scores

### Rhinovirus Symptoms
- Median Time “T”: Time to Peak Symptoms
- Rhinovirus: 72 hours

### RSV Symptoms
- RSV001, RSV002, RSV003, RSV006, RSV007, RSV011, RSV012, RSV014, RSV015

### Influenza Symptoms
- Flu001, Flu005, Flu006, Flu007, Flu008, Flu010, Flu012, Flu013, Flu015

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Number Exposed</th>
<th>Number Symptomatic</th>
<th>Median Time “T”: Time to Peak Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhinovirus</td>
<td>20</td>
<td>10</td>
<td>72 hours</td>
</tr>
<tr>
<td>RSV</td>
<td>20</td>
<td>8</td>
<td>141.5 hours</td>
</tr>
<tr>
<td>Influenza</td>
<td>17</td>
<td>9</td>
<td>80 hours</td>
</tr>
</tbody>
</table>
Study Design

Rhinovirus Cohort
10 Symptomatic
10 Asymptomatic

RSV Cohort
9 Symptomatic
10 Asymptomatic

Influenza A Cohort
9 Symptomatic
8 Asymptomatic

Sparse latent factor regression analysis on combined dataset

Development of 30 gene probit classifier for symptomatic respiratory viral infection

Determine predictive accuracy of 30 gene classifier on independent data set of subjects with NO INFECTION, INFLuenza INFECTION, or BACTERIAL INFECTION

Sparse Latent Factor Regression Analysis

- Latent Factor = Co-expressing genes = Signature
- Assumes MOST genes on array do not have differential expression between varied conditions (“sparseness”)
- “Unsupervised”: does not use class information to derive factors
- Signature can be used to classify new samples as they become available

Design Matrix Latent Factors

\[ x = \beta H' + \Lambda A' + \epsilon \]

\[ \beta_{g,j} \sim (1 - \pi_{g,j})\delta_0 + \pi_{g,j}N(0, \phi) \]

\[ A_{i,k} \sim N(0, \tau) \]

\[ \Lambda_{q,k} \sim (1 - q_{q,k})\delta_0 + q_{q,k}N(\mu_{q,k}, \nu_g) \]
An “Acute Respiratory Viral” Signature Dominates at Time T

Table 2. Intra-Data Set Probit Classification Cross-Validation Results

<table>
<thead>
<tr>
<th></th>
<th>Test: HRV</th>
<th>Test: RSV</th>
<th>Test: Influenza</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train: HRV</td>
<td>1/30 (RSAD2)</td>
<td>2/29 (RTP4)</td>
<td>0/25 (ISG15)</td>
</tr>
<tr>
<td>Train: RSV</td>
<td>1/30 (RSAD2)</td>
<td>2/29 (RTP4)</td>
<td>0/25 (ISG15)</td>
</tr>
<tr>
<td>Train: Influenza</td>
<td>1/30 (RSAD2)</td>
<td>2/29 (RTP4)</td>
<td>0/25 (ISG15)</td>
</tr>
</tbody>
</table>

The error rate is shown based on the top gene (noted in parentheses) selected from the training set probit classifier. For this model, the top 40 genes from the training set discriminative factor were used to build the probit classifier for testing in the validation data set.

The Acute Respiratory Viral Signature Validates in a Historical Cohort

Perfect classification of pediatric subjects with Influenza A (blue) versus hospitalized controls (red)

The Acute Respiratory Viral Signature Validates in a Historical Cohort

73/91 [80%] subjects accurately classified for Influenza A versus Bacterial Infection (S. pneumonia, S. aureus or E. coli)

Conclusions: Classification is Highly Accurate at Maximal Symptoms

- Sparse latent regression analysis identifies a gene expression signature that accurately classifies experimentally infected individuals with symptomatic viral respiratory infection at time of maximal symptoms.
- Genes contained in this signature have direct relationship to known viral response pathways.
- At time of maximal symptoms, a “pan-viral” signature is dominant.
- This methodology, and other methodologies, can be used to develop classifiers that function at earlier timepoints.
What About Earlier Than Maximal Symptoms?

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Number Exposed</th>
<th>Number Symptomatic</th>
<th>Median Time “T”: Time to Peak Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhinovirus</td>
<td>20</td>
<td>10</td>
<td>72 hours</td>
</tr>
<tr>
<td>RSV</td>
<td>20</td>
<td>8</td>
<td>141.5 hours</td>
</tr>
<tr>
<td>Influenza</td>
<td>17</td>
<td>9</td>
<td>80 hours</td>
</tr>
</tbody>
</table>
Early Emergence of the “Acute Respiratory Viral” Factor

Emergence of the “Acute Respiratory Viral Factor” Prior to Time of Peak Symptoms in Experimental Cohorts

- Flu
- Rhino
- RSV

Sample Index Across Time From Inoculation

Inoculation
Time “T” (maximal symptoms)
Can We Move Detection Earlier?

- 1) Improve on the sensitivity of detection
  - RT-PCR
    - Dynamic range of gene expression greater than Affymetrix array
    - Potential to build classifier on basis of degree of gene expression
    - Potential to reduce the number of genes in the classifier

- 2) Use additional statistical methods to achieve earlier classification
  - Bayesian Elastic Net

- 3) Use combination of clinical (i.e. symptoms/physical signs) and molecular (i.e. gene expression) data to achieve earlier classification
From Healthy Adults to Immunocompromised?

• Paradigm increases in complexity as host increases in complexity

• Perhaps can extrapolate healthy young adults to healthy kids

• Important parameters to consider
  • Effect of immunosuppressive regimens on “baseline” gene expression
  • Difficulties with specimen procurement in neutropenia/leukopenia (adequate cells for RNA extraction)
  • Co-infection, effect of herpesvirus reactivation
The DARPA Team

Infectious disease

Christopher W. Woods, MD MPH
(Duke Medicine)
epidemiology, microbiology
Brad Nicholson
Daphne Jones
Kyle Breitschwerdt
Stephanie Dobos

Aimee K. Zaas, MD, MHS
(Duke Medicine)
host-pathogen interaction

L. Brett Caram, MD
(Duke Medicine)
clinical respiratory virology

Stephen Kingsmore, MD
(NCGR)
biomarker development

Jay B. Varkey, MD
(Duke Medicine)
epidemiology, infection control

Project leader

Geoffrey S. Ginsburg, MD PhD
(IGSP)
genomic medicine

Christopher W. Woods, MD MPH
(Duke Medicine)
epidemiology, microbiology
Brad Nicholson
Daphne Jones
Kyle Breitschwerdt
Stephanie Dobos

Aimee K. Zaas, MD, MHS
(Duke Medicine)
host-pathogen interaction

L. Brett Caram, MD
(Duke Medicine)
clinical respiratory virology

Stephen Kingsmore, MD
(NCGR)
biomarker development

Management

Timothy Veldman, PhD
(IGSP)
project management

Lawrence Carin, PhD
(Duke Engineering)
statistical signal processing
machine learning

Joseph Lucas, PhD
(IGSP)
computational biology
Bayesian machine learning

Alfred O. Hero, PhD
(U. Michigan)
statistical inference
bioinformatics

Coordination

N. Christine Øien, MS CGC
(IGSP)
clinical research
genetic counseling

Virology Consultant

John DeVincenzo, MD
(U. Tennessee)

Ron Turner, MD
(UVA)
Invasive Candidiasis – An Important Medical Problem

Candidemia is common and life-threatening

- 4th most common nosocomial BSI\(^1\)
- Excess mortality rates: 10% – 49%
- Average total cost of candidemia: $44,536\(^2\)

Current diagnostic paradigms inadequate

- Variable and nonspecific presentation
- Gold standard for diagnosis: blood culture
  - Sensitivity approximately 50%
- Delay in appropriate therapy increases mortality\(^3\)
Hypothesis

Predictive models based on global changes in gene expression of peripheral blood immune cells can distinguish between infectious causes of illness, particularly candidemia vs. bacteremia.
Study Design

C. albicans:
  discovery (n = 28)
  validation (n=12)
PBS (n=12;5)
S. aureus (n = 12)

Collections at 24, 48, 72, and 96 hrs

RNA Extraction & Globin Reduction

Is there a Signature?
A Disease-Defining Factor
Gene Expression Can Distinguish Between Candidemia and S. aureus Bacteremia
Proposed Clinical Application

High Risk Host
(ICU, Abdominal Surgery, Broad Spectrum Antibiotics)

Point of Care
- Change in clinical status (fever, hypothermia, hypotension, a WBC suspicion of BSI)
  - Evaluate Gene Expression plus Standard Microbiology Results: Candidemia vs. S. aureus bacteremia vs. not infected
    - Not Candidemia: Appropriate Treatment
    - Candidemia: How severe is illness? Early? Moderate? Pre-Morbid?

Screening
- Serial monitoring during high risk time period q 3 days
  - Candidiasis? Consider other markers (1,3 B-D-glucan)
    - Empiric treatment probable candidiasis
    - Continue monitoring; modify risk factors
**Candida: Conclusions**

- Distinct gene expression signatures can identify murine candidemia
- Gene expression signatures change with disease severity
- Genes contained in signature ("factor") are involved in host-pathogen response
- Validation vs. bacteremia AND in human cohort needed

**Specimen Procurement**
- Admission SICU ≥ 72 hours and expected to stay an additional 48 hours
  - IV access
  - Age >18

**Other Data Collection**
- Baseline and ICU d/c oral fungal culture
- Weekly clinical evaluation

**Specimen Analysis**
- Retrospective RNA isolation from patients with blood cultures + for Candida, OR S. aureus OR negative cultures and negative B-glucan testing (controls)

**CLINICAL DATA:**
- Age, gender, underlying illness, medications, Immune suppression, surgery, central line, TPN, Microbiologic culture data

---

**Gene expression signatures**

- Can identify murine candidemia
- Change with disease severity
- Involved in host-pathogen response

**Validation**

- Validation vs. bacteremia AND in human cohort needed
Conclusions

• Diagnosis of infectious diseases can be enhanced by “breaking tradition” from pathogen-based diagnostics

• Combining host and pathogen findings may provide optimal means of classifying infected individuals

• Future directions: Prediction of therapeutic successes or failures
The Candida Team

- Hamza Aziz, MSIII
- Joseph Lucas, PhD
- John Perfect, MD
- Holly Dressman, PhD
- Geoff Ginsburg, MD, PhD