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

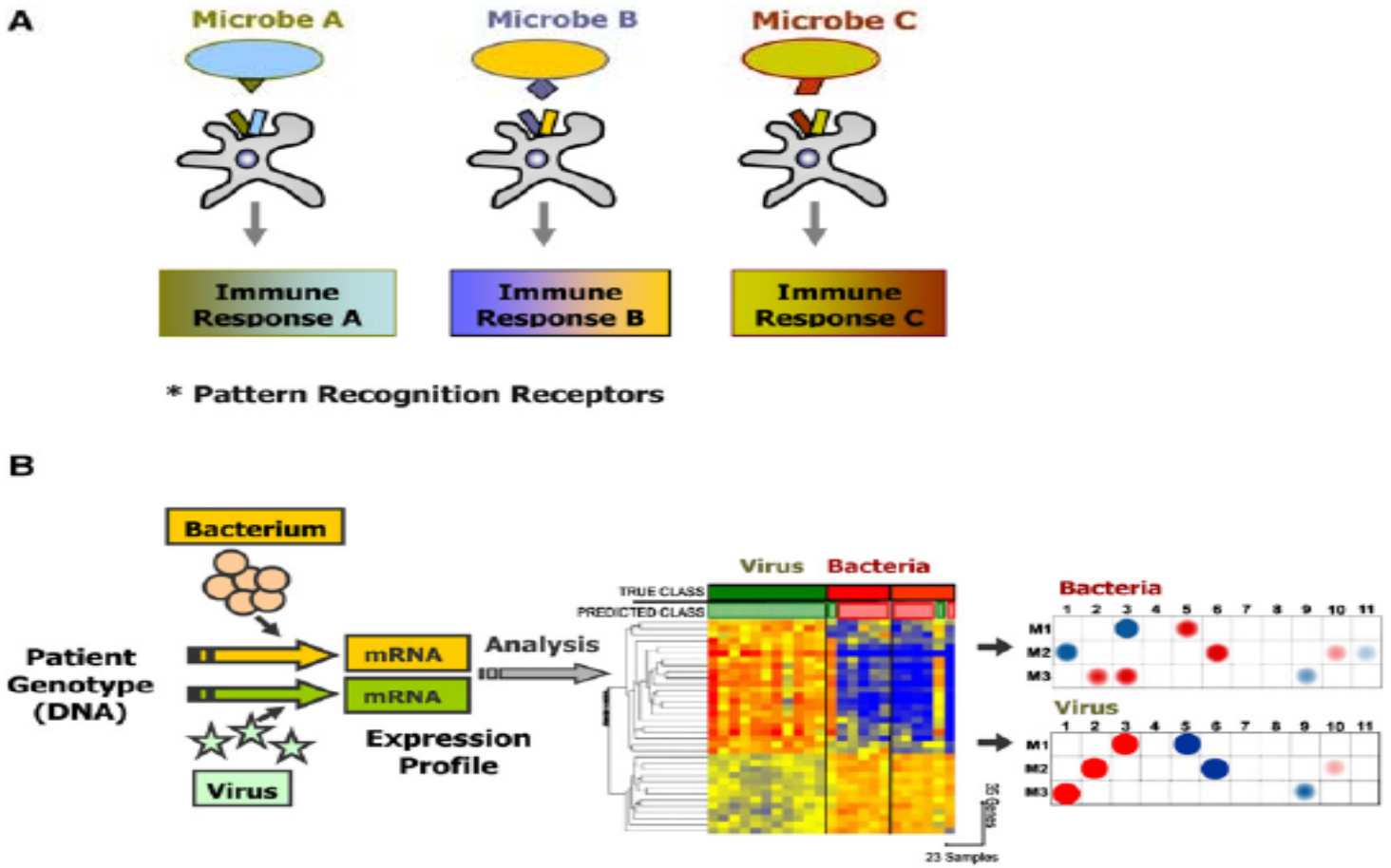
DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

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

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

Gene Expression Can Discriminate Between Pathogens



DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

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

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

Human Viral Challenge Sites: HRV, RSV, Influenza A

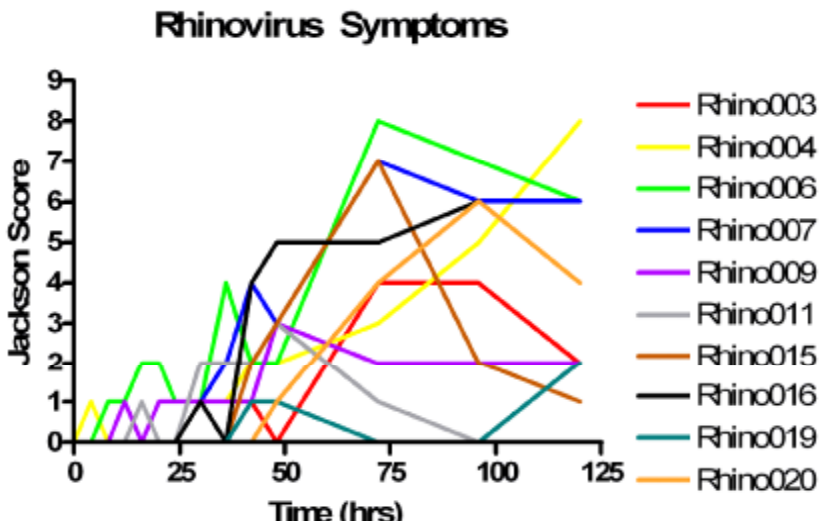
HRV Challenge: Charlottesville, VA 11/2007

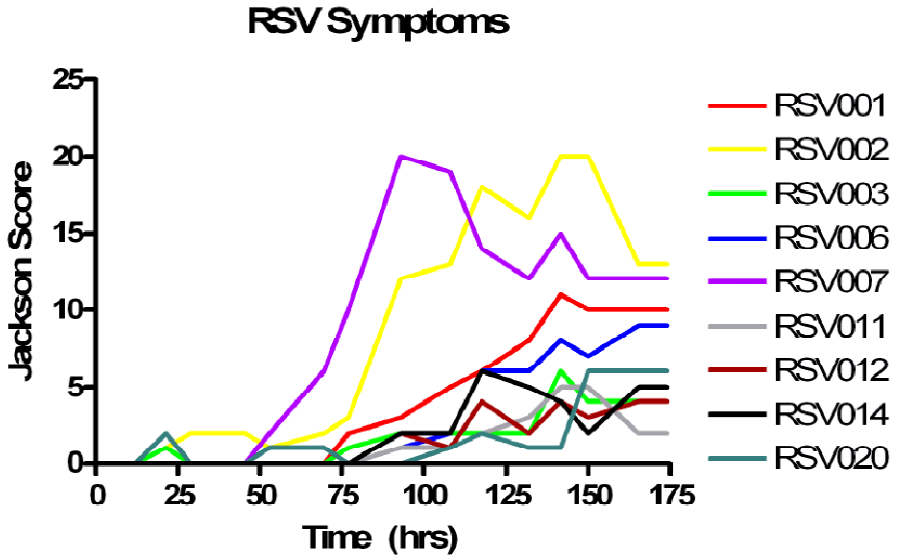
Influenza Challenge (Cambridge, UK 10/2008)

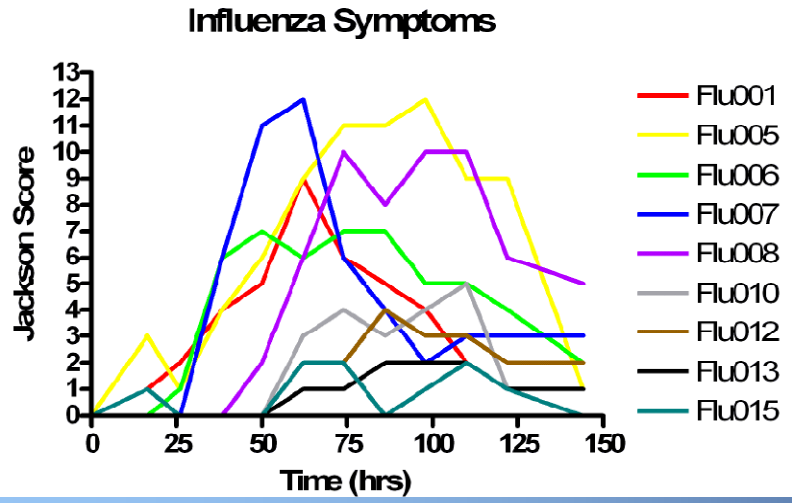
RSV Challenge (Brentwood, UK 7/2008)

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

Human Viral Challenges: Symptom Scores

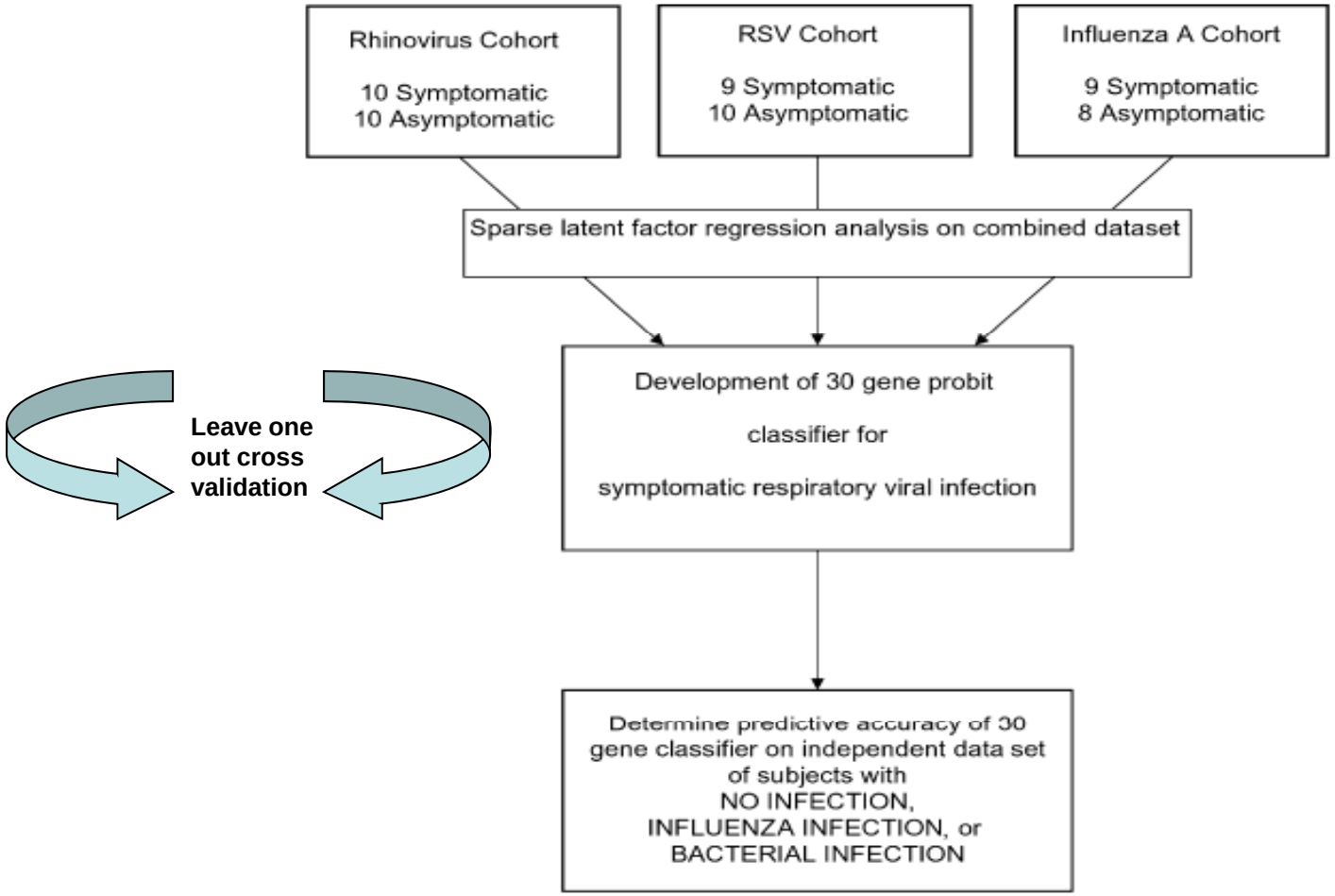






Cohort	Number Exposed	Number Symptomatic	Median Time "T": Time to Peak Symptoms
Rhinovirus	20	10	72 hours
RSV	20	8	141.5 hours
Influenza	17	9	80 hours

Study Design

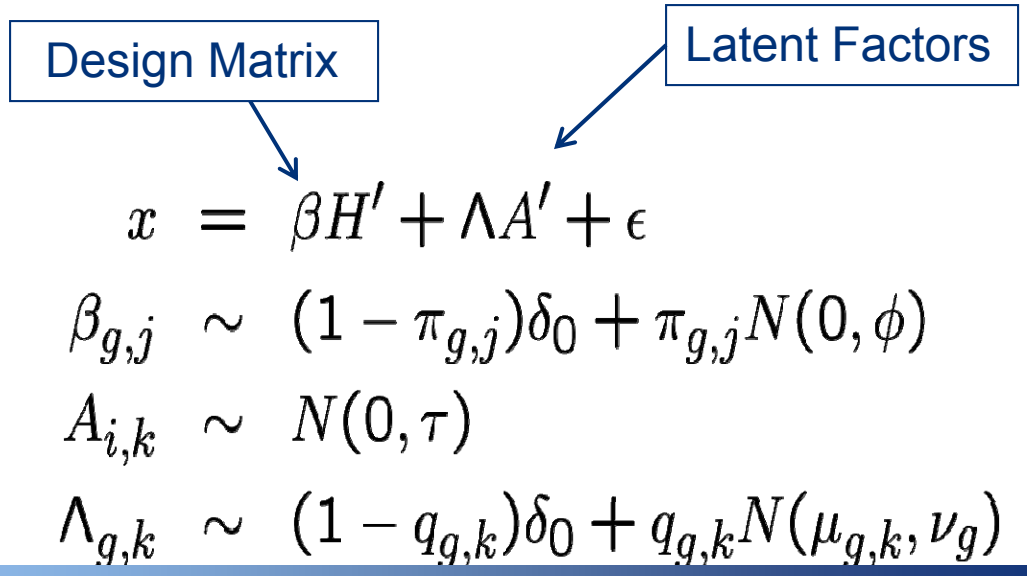


Zaas AK, *et al*. Cell Host Microbe 2009.

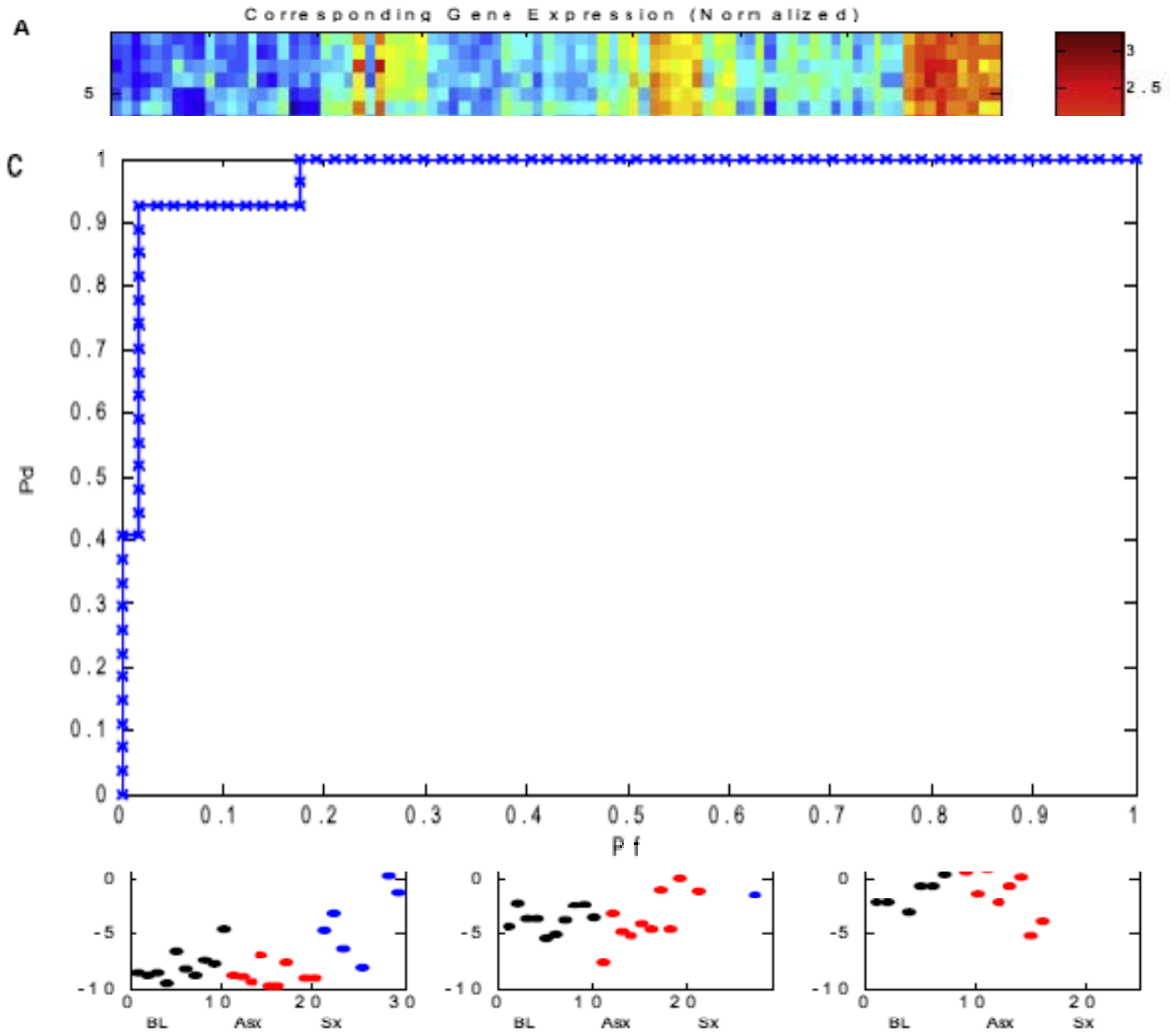
DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

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



DUKE INSTITUTE FOR GENOME SCIENCES & POLICY



DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

An "Acute Respiratory Viral" Signature Dominates at Time T

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

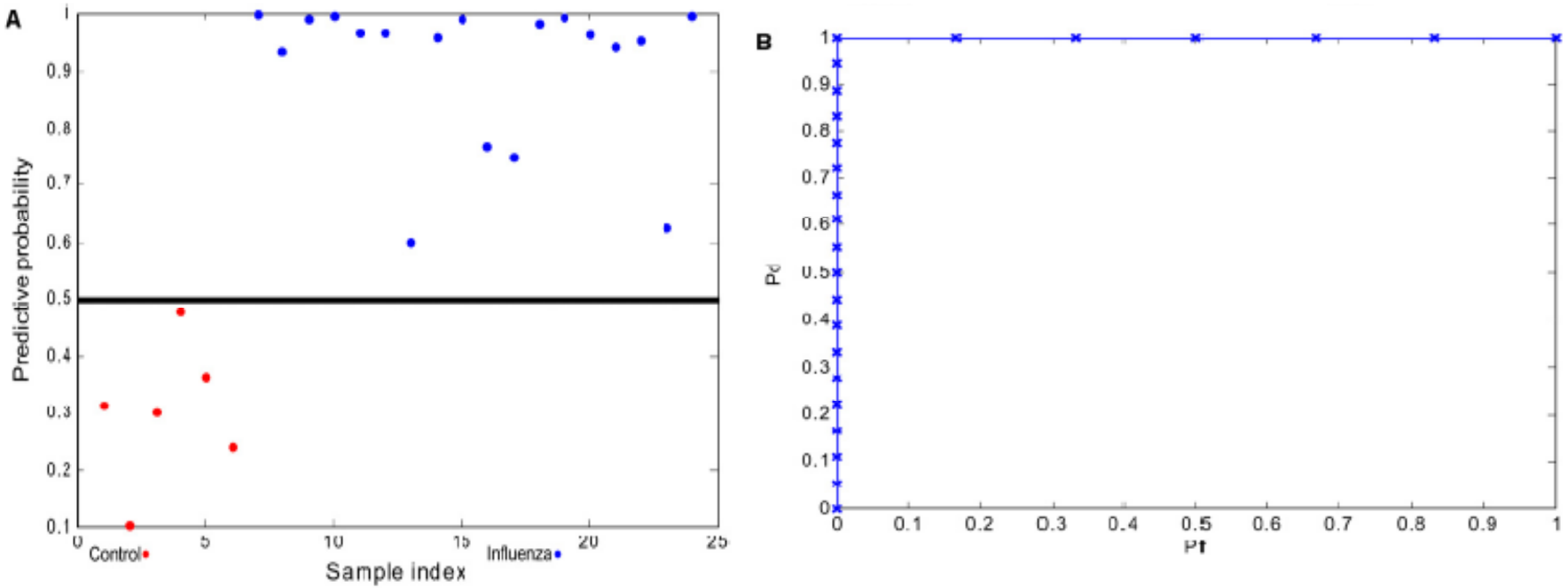
	Test: HRV	Test: RSV	Test: Influenza
Train: HRV	1/30 (*RSAD2*)	2/29 (*RTP4*)	0/25 (*ISG15*)
Train: RSV	1/30 (*RSAD2*)	2/29 (*RTP4*)	0/25 (*ISG15*)
Train: Influenza	1/30 (*RSAD2*)	2/29 (*RTP4*)	0/25 (*ISG15*)

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.

Zaas AK, *et al*. Cell Host Microbe 2009.

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

The Acute Respiratory Viral Signature Validates in a Historical Cohort

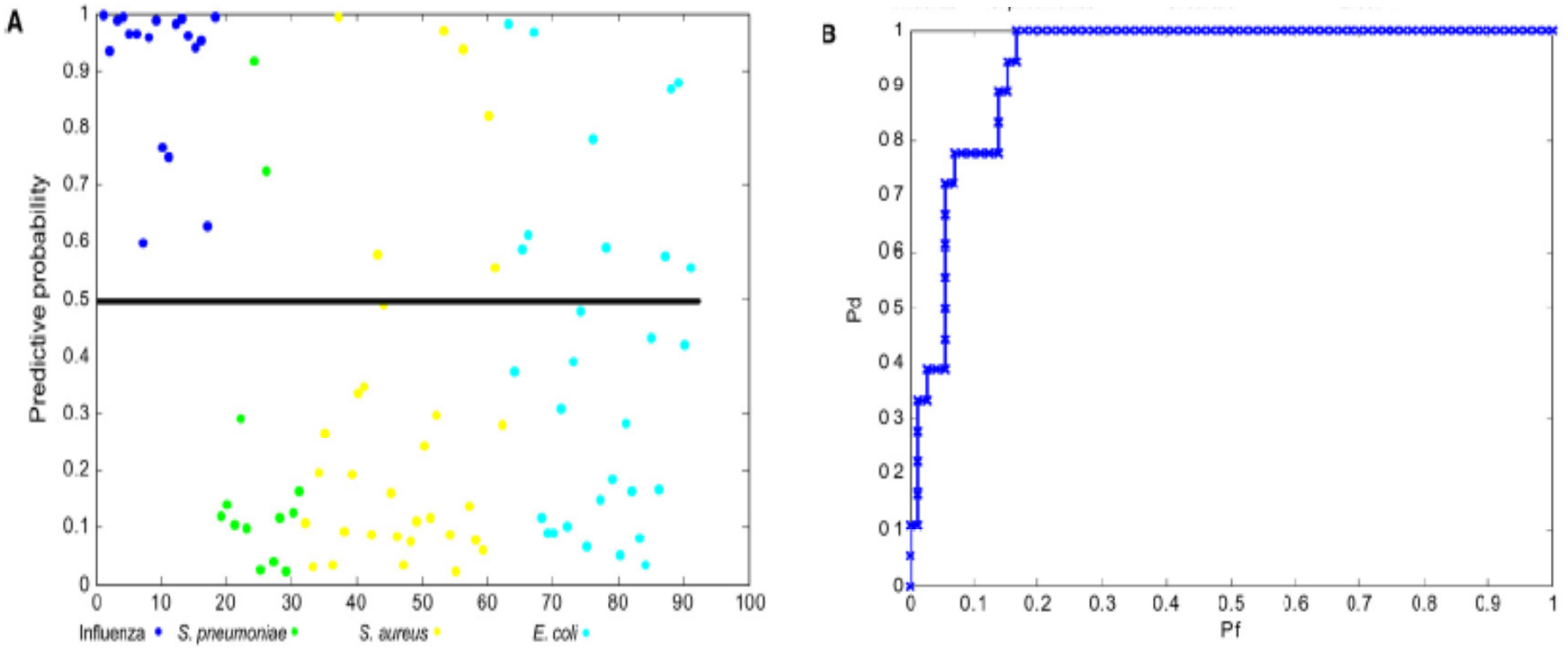


Perfect classification of pediatric subjects with Influenza A (blue) versus hospitalized controls (red)

Zaas AK, *et al*. Cell Host Microbe 2009.

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

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*)

Zaas AK, *et al*. Cell Host Microbe 2009.

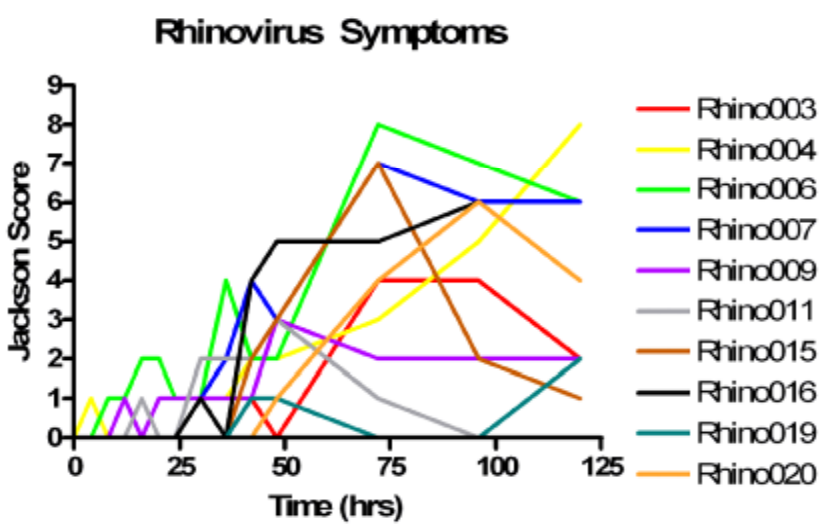
DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

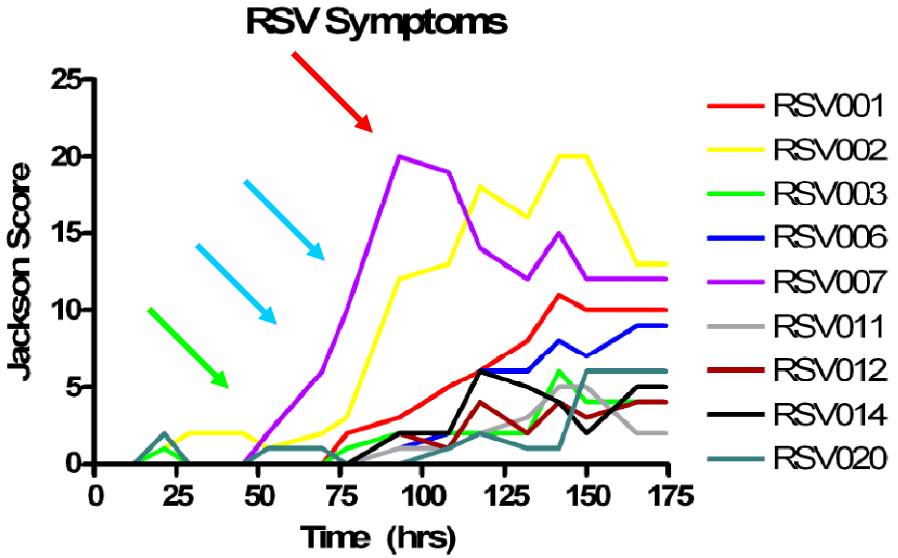
Conclusions: Classification is Highly Accurate at Maximal Symptoms

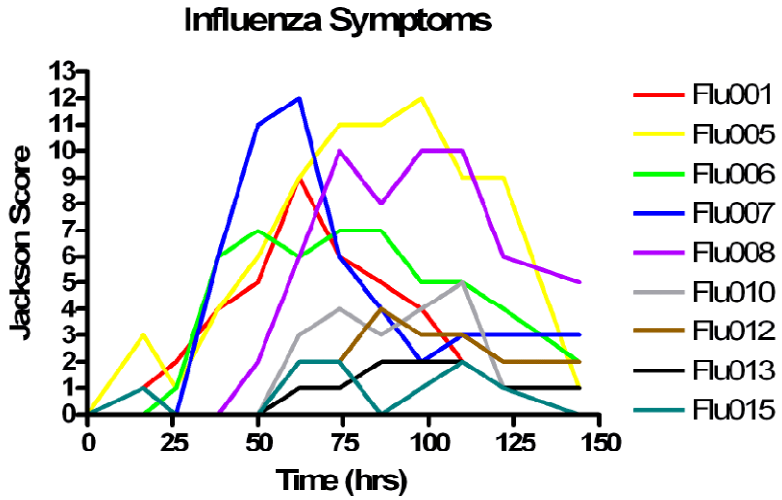
- Sparse latent regression analysis identifies a gene expression signature that accurately classifies experimentally infected individuals with symptomatic viral respiratory infection at time of maximal symptoms
- Genes contained in this signature have direct relationship to known viral response pathways
- 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

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

What About Earlier Than Maximal Symptoms?



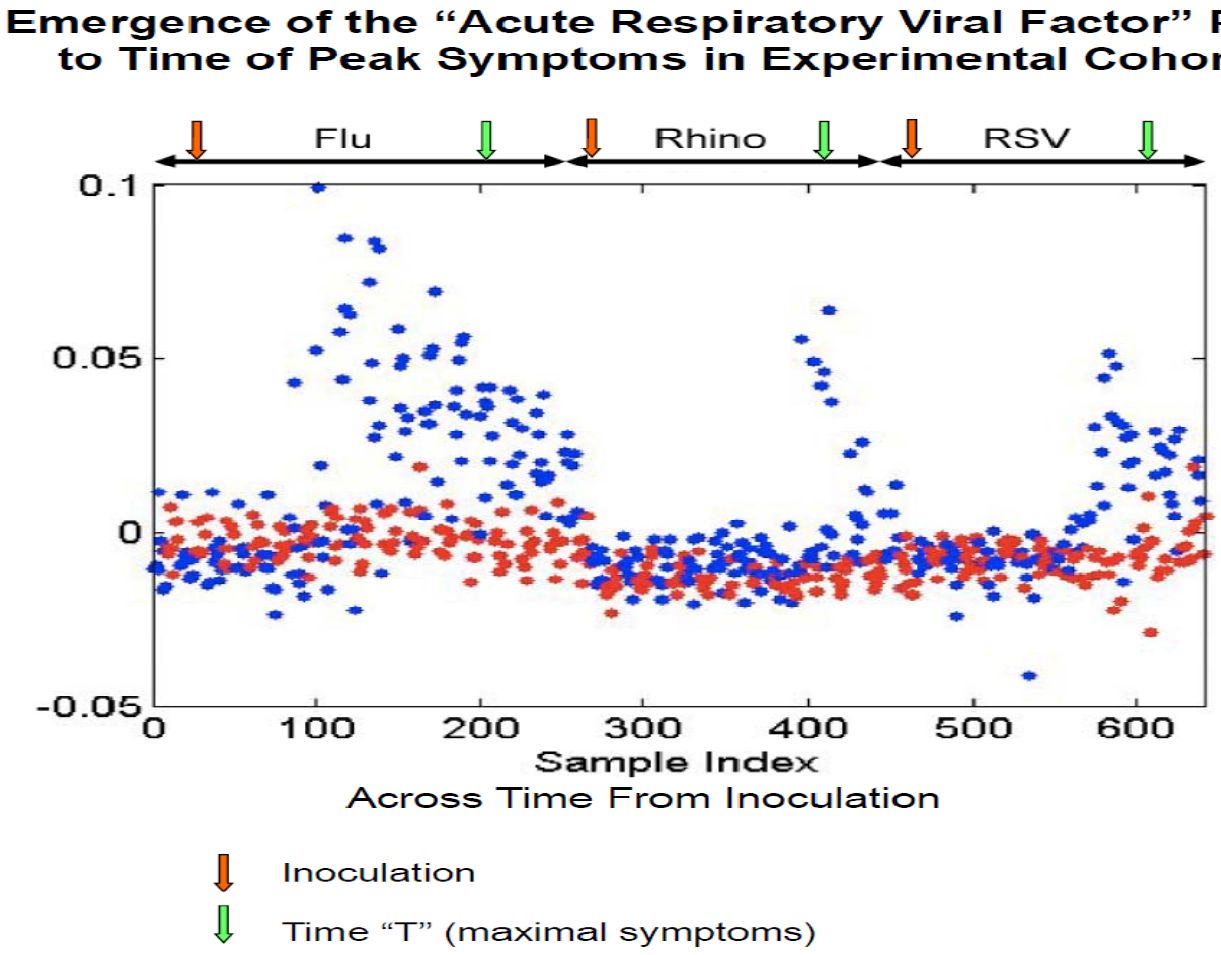




Cohort	Number Exposed	Number Symptomatic	Median Time "T": Time to Peak Symptoms
Rhinovirus	20	10	72 hours
RSV	20	8	141.5 hours
Influenza	17	9	80 hours

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

Early Emergence of the "Acute Respiratory Viral" Factor



DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

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

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

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

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

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

Management

Timothy Veldman, PhD
(IGSP)
project management

Coordination

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

Virology Consultant

John DeVincenzo, MD
(U. Tennessee)

Ron Turner, MD
(UVA)

Data Analysis/Modeling

Lawrence Carin, PhD
Minhua Chen, BS
(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

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

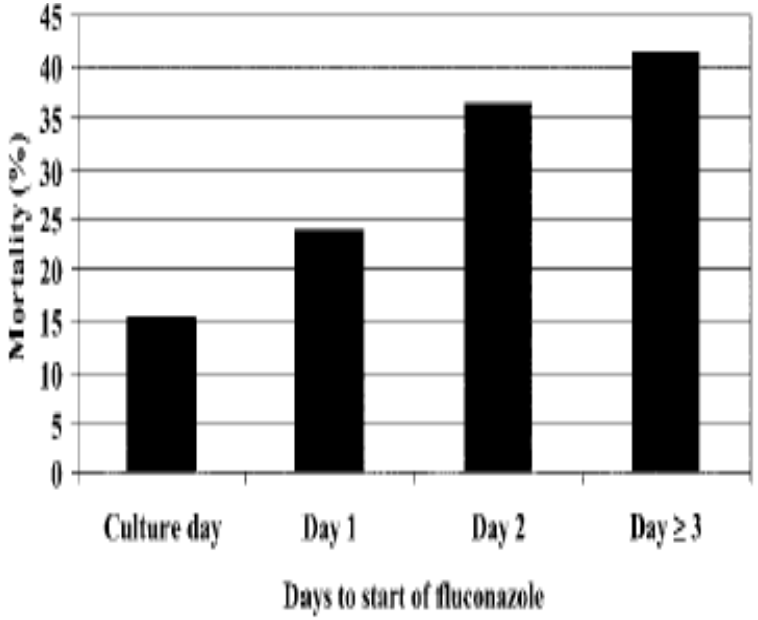
Invasive Candidiasis – An Important Medical Problem

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]

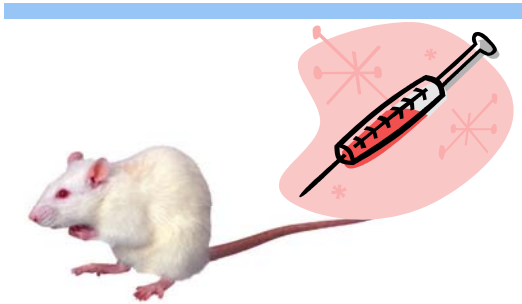


DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

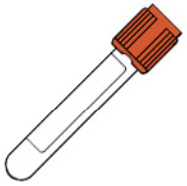
Hypothesis

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

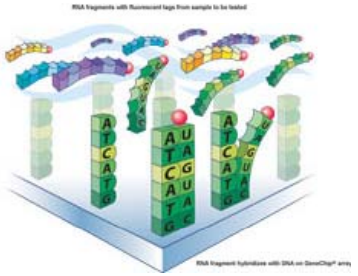


C. albicans:
discovery (n = 28)
validation (n=12)
PBS (n=12;5)
S. aureus (n = 12)

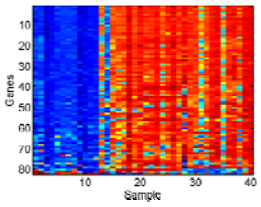


Collections at
24, 48, 72, and
96 hrs

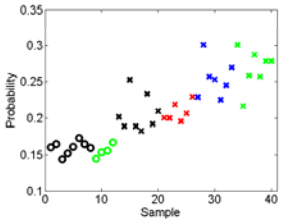
RNA Extraction &
Globin Reduction



Is there a Signature?

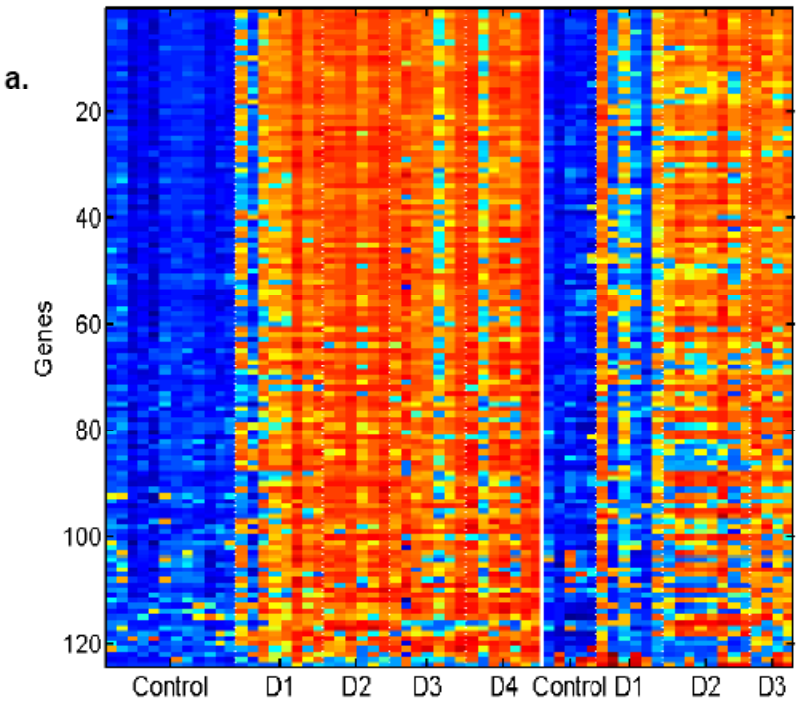


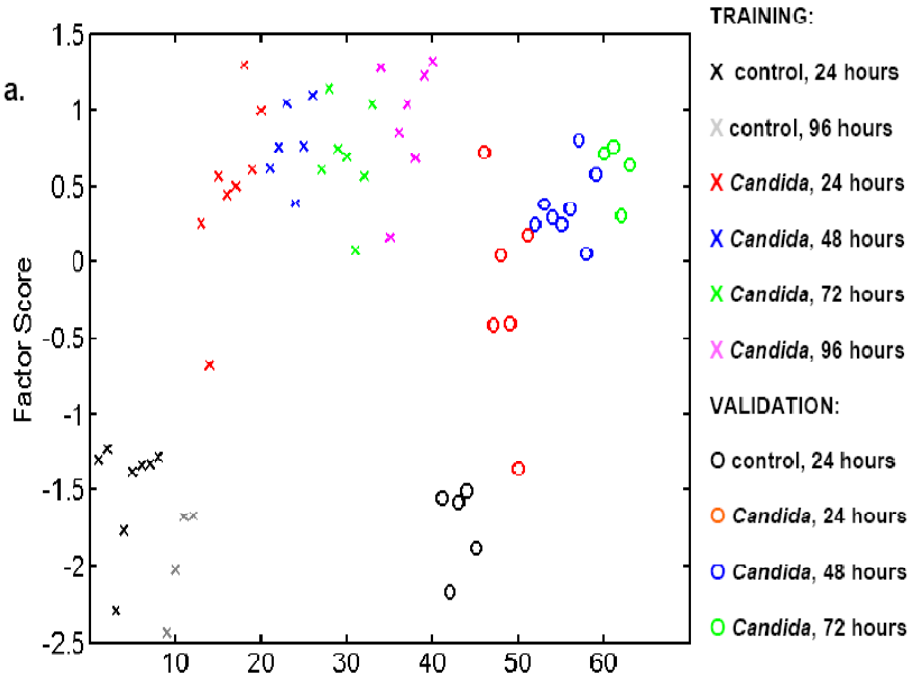
Validation



DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

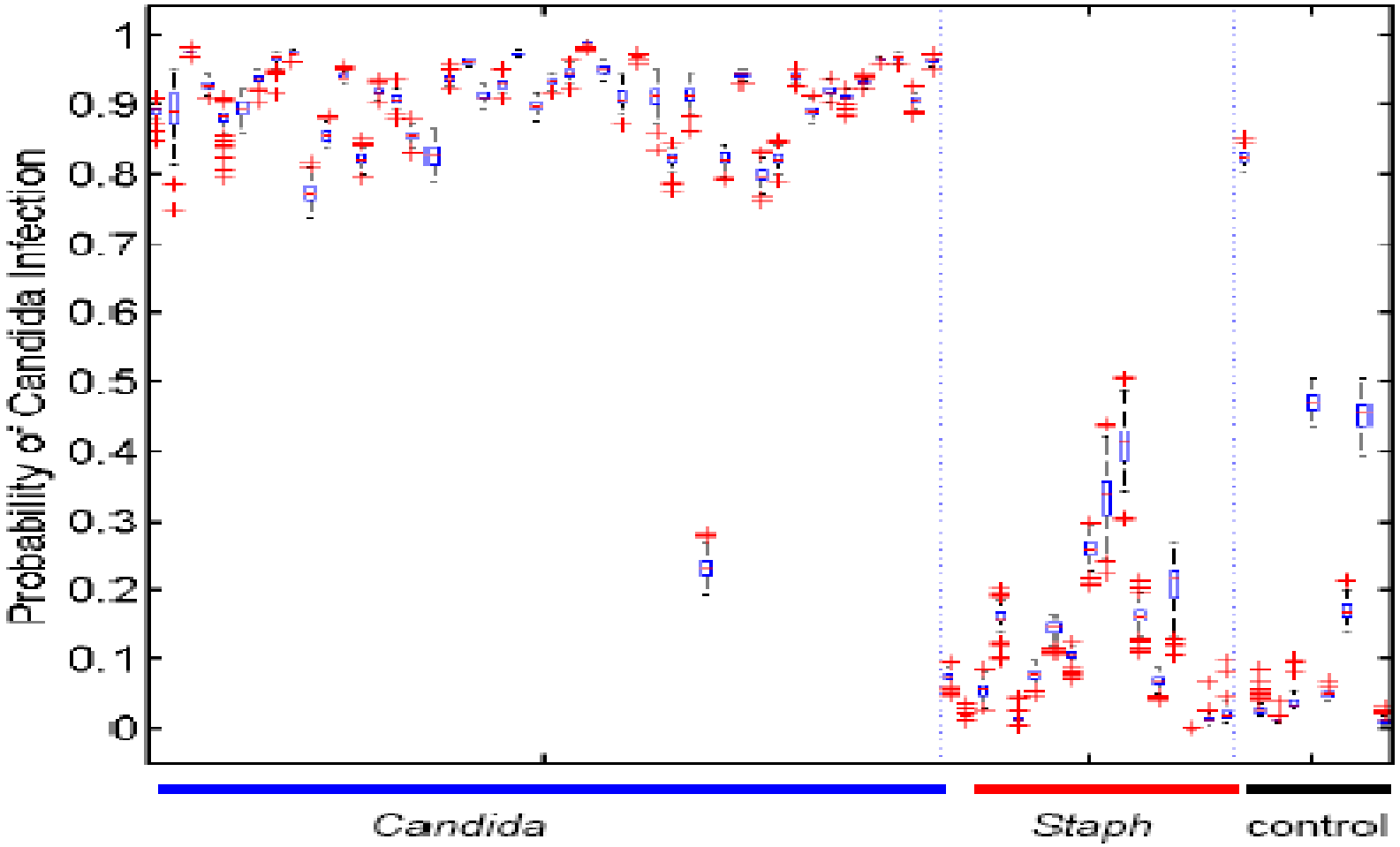
A Disease-Defining Factor





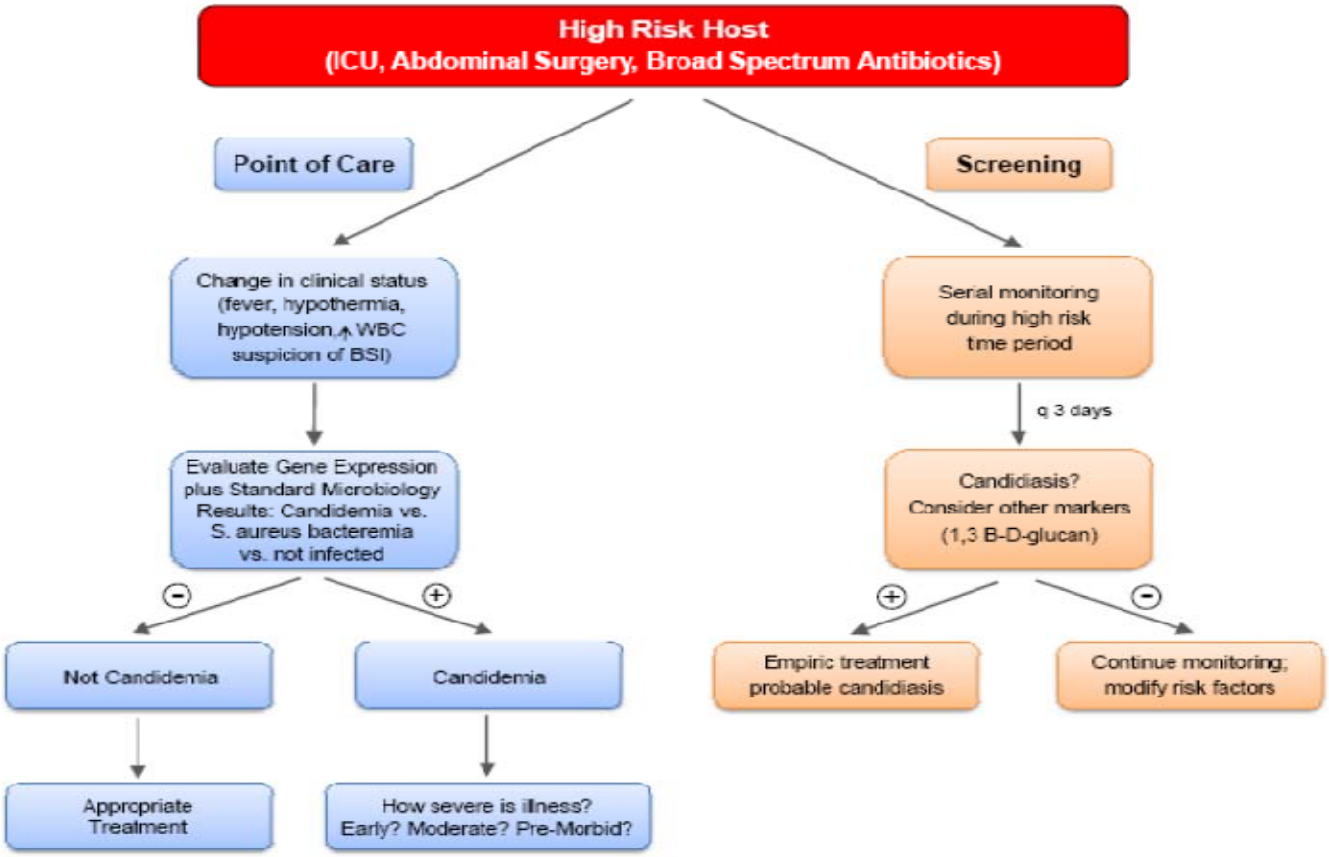
DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

Gene Expression Can Distinguish Between Candidemia and *S. aureus* Bacteremia



DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

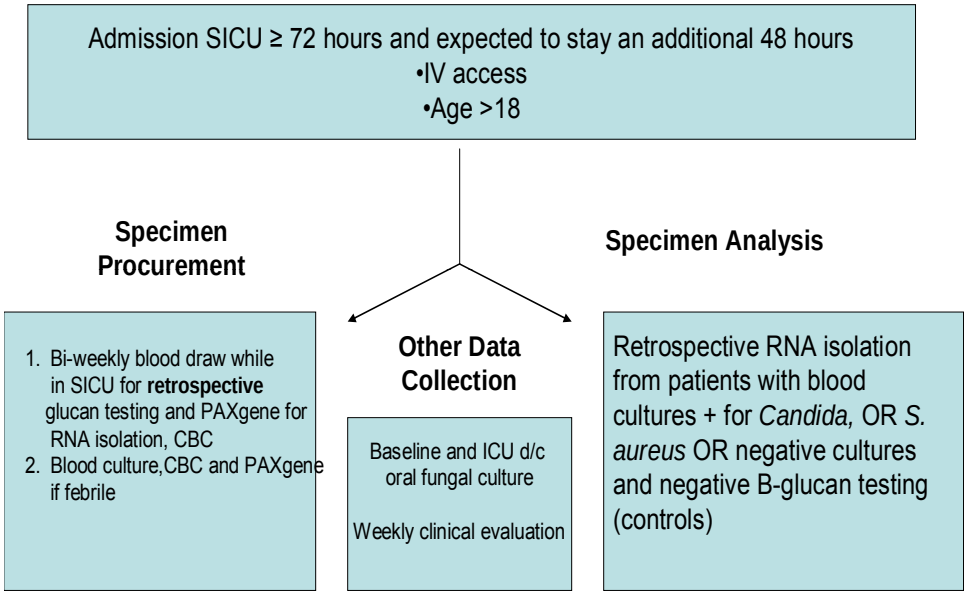
Proposed Clinical Application



DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

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



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

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

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

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY

The *Candida* Team

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

DUKE INSTITUTE FOR GENOME SCIENCES & POLICY