Metagenomic Sequencing Approaches for Diagnosis and Genomic Characterization of Blood-borne Infections

Charles Chiu, MD / PhD
Associate Professor, Department of Laboratory Medicine and Medicine / Infectious Diseases
Director, UCSF-Abbott Viral Diagnostics and Discovery Center
Associate Director, UCSF Clinical Microbiology Laboratory
University of California, San Francisco

Surveillance and Screening of Blood-Borne Pathogens

2018 IPFA /PEI 25th International Workshop
Major Diagnostic Challenges in Infectious Diseases

**Pneumonia**

15 – 62% unknown cause


**Fever / Sepsis**

~20% unknown cause


**Meningitis / Encephalitis**

40 – 60% unknown cause


Failure to obtain a timely diagnosis leads to delayed / inappropriate therapy, increased mortality, and excess healthcare costs.
All Microbes can be Uniquely Identified by mNGS

Bacteria

Viruses

Fungi

Parasites
The SURPI Bioinformatics Pipeline

“Sequence-based ultra-rapid pathogen identification” (minutes – hours)

- **Directly addresses computational analysis bottleneck**

- **SURPI+ (clinical version) – automated analysis**

Naccache, et al., 2014, Genome Research 24(7):1180-1192
Precision Diagnosis of a Mysterious Infection

3 hospitalizations over 4 months

44 days in the ICU

>100 inconclusive tests

Brain biopsy and induced coma

Leptospira santarosai
Leptospira borgpetersenii
unclassified
Leptospira interrogans
Propionibacterium acnes
The Precision Diagnosis of Acute Infectious Diseases (PDAID) Study

Meningitis/Encephalitis

7 hospitals in CA and nationwide
Enroll/consent patients
204 total completed study
CSF collected
Clinical chart review

mNGS assay validated in CLIA lab
86% analytic sensitivity, 98% specificity

Clinical report in patient EMR

Miller, et al., 2018 (manuscript under review);
Wilson, et al., 2018 (manuscript under review)
58 y/o immunosuppressed woman with fever, headache, nausea/vomiting

- History of idiopathic pulmonary fibrosis status post bilateral lung transplant in 2011, multiple sclerosis, on chronic immunosuppression
- Admitted to hospital in October 2016 with 8 days of fever, headache (“worst in my life”), nausea/vomiting, neck stiffness, and photophobia
- Neurological symptoms: admitted to “5 years” of word-finding difficulty and slurred speech, 1 year of dizziness / falls, and 1 month of leg weakness; also had first-time seizure in March 2016
- Resident of Orange County; no sick contacts; travel to mountains in Utah in August 2016, Caribbean in 2010, and throughout Europe decades ago
- Fever to 38.3°C, pancytopenic, transaminitis (negative for hepatitis A,B, and C); MRI – white matter intensities related to MS
- Started on empiric antimicrobials: IV vancomycin, ceftazidime, acyclovir, and voriconazole

- Lumbar puncture done, showing a lymphocytic pleocytosis

- WBC 10, 88% lymphocytes, protein 29, glucose 48
Hepatitis E virus
mapped to GenBank AB089824, 7,262 bp (Hepatitis E virus genomic RNA, complete genome, isolate: HE-JA10)
assembly 930% complete, 90.2% average pairwise identity
• Patient treated with ribavirin and is clinically improved
• This is a likely case of donor-transmitted HEV (positive anti-HEV antibody testing of donor’s serum)
15 y/o child from Massachusetts with meningoencephalitis
Phylogeny of tickborne Powassan virus

YFV X03700.1, 17D vaccine strain

HM440563.1, 1964, USA: New York, 64-7062
NC_003687.1, LB
L06436.1, LB
KT224351.1, 1970, Russia: Far East, LEIV-5530
KT224350.1, 1977, Russia: Primorsky krai (Far East of Russia), LEIV-3070Prm
EU770575.1, 15-Apr-1975, Russia, Spassk-9
EU543649.1, 2006, Russia, Partizansk/2006
EU670438.1, Russia, Nadezdinsk-1991
HQ231415.1, 2006, Russia, Ternay
HQ231414.1, 2006, Russia, Ulysses

MGH Patient

MG647781.1, 2016, USA
KJ746872.2, Apr-2013, USA
MG647780.1, 2016, USA
MG647779.1, 2016, USA
HM440559.1, 1996, USA: Nantucket, MA, NFS001
AF311056.1, ctb30
MG647783.1, 2016, USA
KU886216.1, 2010, USA: Bridgeport, CT, P0375
MG647782.1, 2016, USA
MF688929.1, Dec-2016, USA
HM440558.1, 1999, USA: Chippewa Falls, WI, wicf9901
HM440562.1, 2008, USA: Spooner, WI
HM440560.1, 2008, USA: Spooner, WI
HM440561.1, 2008, USA: Spooner, WI
Enhancing Metagenomic Detection Sensitivity

**EBOV**


**ZIKV**

(Naccache, Thévé, et al., 2016, *EID*, 22:10)
Spiked Primer Strategy for Metagenomic Target Enrichment

- Set of reference genomes
  - Multiple sequence alignment
  - Partition aligned genomes into overlapping 250-nt segments
    - HIV (3,571 genomes)
    - ZIKV (44 genomes)
  - Generate consensus sequence and partition into 250-nt segments
  - Select forward and reverse spiked 13-nt primers
    - Filter primers by Tm (<2 SD from mean)
    - Remove self-dimers / cross-dimers
    - Remove homopolymer repeats (≥ 25 nt)

Cell Host & Microbe, in press
Manuscript in preparation
MSSP (Metagenomic Sequencing with Spiked Primers) Enrichment

Thézé, et al., 2018, Cell Host & Microbe, in press
Deng, et al., 2017, manuscript in preparation
Thézé, et al., 2018, Cell Host & Microbe, in press
BloodSeq Panel

- **Borrelia sp.**
- **Babesia sp.**
- **Anaplasma/Ehrlichia sp.**
- Tick-borne viruses (Heartland virus, Powassan virus, etc.)
- Bloodborne viruses (dengue virus, Zika virus, West Nile virus, Ebola virus, Lassa virus, Japanese encephalitis virus, etc.)

**Diagram:**
- Set of reference genomes from targeted pathogens
- Multiplex sequence alignment
- Select forward and reverse spiked 13-nt primers
- 50-nt window
- Generate consensus sequence and partition into 250-nt segments
- Spiked 13-nt primer with 18-nt adapter
Whole-Genome Sequencing of Asymptomatic "Babesia microti" Infection in US Blood Donors

All samples to date have M134I mutation associated with atovaquone resistance.

Concatenated COX1-CYTB-COX3
Usutu Virus Detection in an African Blood Sample using HIV Primers

<table>
<thead>
<tr>
<th>Sequencing Stats</th>
<th>Run1</th>
<th>Run2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw Data</td>
<td>19,969,548</td>
<td>690,200,472</td>
</tr>
<tr>
<td>Human immunodeficiency virus 1</td>
<td>69374</td>
<td>2384992</td>
</tr>
<tr>
<td>Usutu virus</td>
<td>201</td>
<td>6502</td>
</tr>
<tr>
<td>% Usutu virus</td>
<td>0.001%</td>
<td>0.001%</td>
</tr>
</tbody>
</table>

Collaboration with Mary Rodgers and John Hackett, Abbott Laboratories
Usutu Virus Detection using HIV Primers

Original sequencing (20%)

Deeper sequencing (29%)

With Sanger sequencing (34%)
Metagenomics for Diagnosis of Sepsis in Point-of-Care Settings

![Diagram showing the percentage of patients with effective antibiotic therapy and patient survival rate over time to effective antibiotics. The x-axis represents time in hours (0, 1, 2, 3, 4, 5, 6, 9, 12, 24, 36), and the y-axis represents the percentage of patients with effective antibiotic therapy and patient survival rate.]
Nanopore Sequencing – Sample-to-Answer Diagnosis in <3 hr

Chikungunya Virus

Ebola Virus

(Greninger, et al., 2015, Genome Medicine 7:99)
## Differential Diagnosis of Tropical Febrile Illness

**BACTERIAL**

- Rickettsioses
- Bacillary Dysentery
- Plague
- Meningococcemia
- Typhoid fever
- Other bacterial septicemia
- Leptospirosis
- Ehrlichiosis
- Tuberculosis
- Bartonellosis
- Brucellosis

**VIRAL**

- Arboviral infections
- Viral hepatitis
- Enterovirus
- Measles
- Rubella
- Acute Retroviral Syndrome (HIV)
- Epstein-Barr virus
- Parvovirus
- Roseola vairus
- Filovirus infection (Ebola, Marburg)
- Lassa virus
- bunyaviruses

**OTHER**

- Malaria / Babesiosis
- Amebiasis
- Visceral leishmaniasis
- Acute schistosomiasis
- Filarial fever
- Trypanosomiasis
Protocol Optimization on the MinION Nanopore Sequencer

Can analyze 1 million reads in 10 minutes on a laptop

Wayne Deng, PhD
Dianna Ng, MD
Scot Federman, BA

Deng, et al., 2017, manuscript in preparation
Pan-Bacterial Detection in <30 Min by CRISPR-CAS12

(collaboration with Janice Chen, Lucas Harrington, and Jennifer Doudna, UC Berkeley)
# Real-Time Pathogen Detection in Infected Body Fluids from Patients

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Nanopore Findings</th>
<th>Pathogen</th>
<th>Total # of Reads</th>
<th>Real-time Detection</th>
<th>Nanopore RPM</th>
<th>HiSeq RPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>JDL052* (BAL)</td>
<td><em>Streptococcus pneumoniae</em></td>
<td>5,816</td>
<td>400,000</td>
<td>2 min (320/50,000)</td>
<td>13,819</td>
<td>22,663</td>
</tr>
<tr>
<td>132 (plasma)</td>
<td><em>Haemophilus influenzae</em></td>
<td>7</td>
<td>240,000</td>
<td>2 hr (7 reads)</td>
<td>28</td>
<td>69</td>
</tr>
<tr>
<td>133 (BAL)</td>
<td><em>Haemophilus influenzae</em></td>
<td>436</td>
<td>280,000</td>
<td>2 min (25/4,000)</td>
<td>2,718</td>
<td>11,096</td>
</tr>
<tr>
<td>81S (joint fluid)</td>
<td><em>Staphylococcus aureus</em></td>
<td>140</td>
<td>107,600</td>
<td>2 min (1/4,000)</td>
<td>1,300</td>
<td>2,792</td>
</tr>
<tr>
<td>78S* (pleural fluid) **</td>
<td><em>Staphylococcus lugdunensis</em></td>
<td>24</td>
<td>60,000</td>
<td>10 min (8/20,000)</td>
<td>400</td>
<td>3,737</td>
</tr>
</tbody>
</table>

*blinded sample

**90 minutes sample-to-result (direct sequencing)

Deng, et al., 2017, manuscript in preparation
# Real-Time Pathogen Detection in Infected Body Fluids from Patients

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Nanopore Findings</th>
<th>Pathogen</th>
<th>Total # of Reads</th>
<th>Real-time Detection</th>
<th>Nanopore RPM</th>
<th>HiSeq RPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>69S* (BAL)</td>
<td><em>Streptococcus pneumoniae</em></td>
<td>8</td>
<td>600,000</td>
<td>30min (3 reads)</td>
<td>13.3</td>
<td>209</td>
</tr>
<tr>
<td>83* (abscess fluid)</td>
<td><em>Staphylococcus aureus</em></td>
<td>5</td>
<td>3 million</td>
<td>&gt;2hr</td>
<td>1.7</td>
<td>?</td>
</tr>
<tr>
<td>87* (abdominal fluid)</td>
<td><em>Candida glabrata</em></td>
<td>10</td>
<td>1.98 million</td>
<td>&gt;2hr</td>
<td>5.1</td>
<td>12.9</td>
</tr>
<tr>
<td>89* (perihepatic fluid)</td>
<td><em>Candida glabrata</em></td>
<td>9</td>
<td>1 million</td>
<td>&gt;2hr</td>
<td>9</td>
<td>82.9</td>
</tr>
<tr>
<td>NP15* (plasma, HCV 10^2 copies/mL)</td>
<td>HCV</td>
<td>9</td>
<td>200,000</td>
<td>30 min</td>
<td>45</td>
<td>–</td>
</tr>
<tr>
<td>NP28* (plasma, DENV 10^3 copies/mL)</td>
<td>DENV</td>
<td>39</td>
<td>66,454</td>
<td>20 min</td>
<td>586</td>
<td>–</td>
</tr>
</tbody>
</table>

*Deng, et al., 2017, manuscript in preparation*
# Nanopore Sequencing of Clinical Samples

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Nanopore Findings</th>
<th>Pathogen</th>
<th>Total # of Reads</th>
<th>Real-time Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM1 (plasma, 40 parasites/µl)</td>
<td>Babesia microti</td>
<td>31</td>
<td>40,000</td>
<td>20 min</td>
</tr>
<tr>
<td>PF1 (plasma, 10 parasites/µl)</td>
<td>Plasmodium falciparum</td>
<td>15</td>
<td>50,000</td>
<td>20 min</td>
</tr>
<tr>
<td>Mex1 (plasma, 10^3 copies/mL)</td>
<td>DENV</td>
<td>7</td>
<td>50,000</td>
<td>20 min</td>
</tr>
<tr>
<td>Mex2 (nasal swab, 10^4 copies/mL)</td>
<td>Influenza B</td>
<td>4</td>
<td>50,000</td>
<td>20 min</td>
</tr>
<tr>
<td>DRC1 (nasal swab)</td>
<td>Influenza A</td>
<td>2</td>
<td>40,000</td>
<td>20 min</td>
</tr>
</tbody>
</table>

Deng, et al., 2017, manuscript in preparation
INRB (Institut National de Recherche Biomedicale)
Anne Rimoin and UCLA-DRC Team

Anne Rimoin, PhD

Matt Bramble, PhD

Russell Williams, PhD
Identification of Bloodborne Infections by Host Response Analysis

Antibody testing misses the window period for diagnosing early Lyme disease
Machine Learning of RNA-Seq Data to Predict Causes of Infection
Machine Learning-Based Analyses of Host Response

90% accuracy in discriminating bacterial from viral infection from CSF (preliminary analysis)

- Training set: 25 bacterial positive cases, 48 viral positive cases
- Test set: 6 bacterial positive cases, 9 viral positive cases

<table>
<thead>
<tr>
<th>Method</th>
<th>CV RSME Score</th>
<th>CV RSME STD</th>
<th>CV Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear SVM</td>
<td>0.221034</td>
<td>0.226150</td>
<td>0.900000</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.250920</td>
<td>0.265278</td>
<td>0.866667</td>
</tr>
<tr>
<td>Polynomial SVM</td>
<td>0.323491</td>
<td>0.220651</td>
<td>0.846667</td>
</tr>
<tr>
<td>Stochastic Gradient Descent</td>
<td>0.241841</td>
<td>0.248018</td>
<td>0.840000</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>0.274579</td>
<td>0.290871</td>
<td>0.806667</td>
</tr>
</tbody>
</table>

(with Matt Massie and Anthony Joseph, UC Berkeley)
Whole transcriptome analysis (n=72) CA, MD
41 clinical Lyme patients 31 controls

Targeted RNA expression (TREx) sequencing (n=220 total)
90 clinical Lyme patients 130 controls

Machine learning (ML) based gene pruning (n=190, 86 genes) BC, CA, MD

random 50:50 split, blinded analysis
Training set (n=95)
30 Lyme seropositive 65 controls

10X cross-validation
comparison of 10 ML methods
Reduced gene panel (20 genes) and best-performing ML algorithm

Validation set (n=95)
30 Lyme seropositive 65 controls

seronegative patients with presumptive Lyme disease (n=30)

Bouquet, et al., 2018, under review
Accuracy and Kappa Statistics for 10 Different Machine Learning Methods

- generalized linear models
- radial support vector machine
- linear support vector machine
- random forest
- naïve Bayes
- neural networks
- nearest shrunken centroids
- classification and regression trees
- k-nearest neighbor
- linear discriminant analysis

(Bouquet, et al., 2018, under review)
20-Gene Lyme Classifier Panel for Diagnosis of Early Lyme Disease

- Best validation set results
  - 94% sensitivity
  - 93% specificity (seropositive)
  - 91% accuracy (including seronegative patients)

(Bouquet, et al., 2018, under review)
Acknowledgements

UCSF Chiu Lab and VDDC
Calla Martyn, BS
Scot Federman, BA
Asmeeta Achari, BS
Shaun Arevalo, CLS
Jerome Bouquet, PhD
Guixia Yu, BS
Dianna Ng, MD
Wayne Deng, PhD
Doug Stryke, MS
Matt Massie, BS
Tony Li, BS
Guixia Yu, BS
Steve Miller, MD, PhD

UCSF Neurology
Michael Wilson, MD

Boston Children’s Hospital/Harvard
Lise Nigrovic, MD
Lyme Disease Biobank
Elizabeth Horn, PhD

Columbia University
Brian Fallon, MD

Johns Hopkins University
John Aucott, MD
Mark Soloski, MD

Duke University
Micah McClain, MD, PhD
Geoffrey Ginsburg, MD, PhD
Christopher Woods, MD

Oxford University, UK
Julien Theze, PhD
Oliver Pybus, PhD
Nuno Faria, PhD

CDPH
Sharon Messenger, PhD
Shigeo Yagi, PhD
Debra Wadford, PhD

American Red Cross
Susan Stramer, PhD
Roger Dodd, PhD

Blood Systems Research Institute
Michael Busch, PhD

Funding
• NIH R01 HL105701-01 and R21 AI120977
• UC Discovery Award
• Translational Research Institute / NASA Grant
• Abbott Pathogen Discovery Award
• California Initiative to Advance Precision Medicine
• Charles and Helen Schwab Foundation
• George and Judy Marcus Innovation Fund
• Steven and Alexandra Cohen Foundation
• Bay Area Lyme Disease Foundation
• Global Lyme Alliance

E-mail: charles.chiu@ucsf.edu