Clinical Metagenomics for the Diagnosis of Acute Infectious Diseases

Jerome Bouquet, Ph.D.
Postdoctoral scholar, Chiu Lab
Department of Laboratory Medicine, University of California, San Francisco
UCSF-Abbott Viral Diagnostics and Discovery Center
Critical need for better Diagnostic tests

Meningitis/Encephalitis: 60-80% unknown cause

Pneumonia: 15-25% unknown cause

Hemorrhagic Fever: 20% unknown cause

Diarrheal Disease: 50% unknown cause
Metagenomic Sequencing – Casting a Wide Net

Conventional Testing
- Sero
- PCR
- Culture
- Microscopy
...

Metagenomic Sequencing
Accelerating Diagnosis with Next-Gen Testing can Impact Clinical Decision-Making ("Precision Medicine")

Lower healthcare costs

Improved patient outcomes

Cost-effective and actionable information for early treatment
14 y/o male with Meningoencephalitis

- 3 hospitalizations over 4 months
- 44 days in the ICU
- >100 inconclusive tests
- 3 empiric treatments with no effect
- Brain biopsy and induced coma

(Wilson, et al., 2014, New England Journal of Medicine; photo courtesy of the Osborn family)
14 y/o male with Meningoencephalitis

3 hospitalizations over 4 months

44 days in the ICU

>100 inconclusive tests

3 empiric treatments with no effect

Brain biopsy and induced coma

→ Neuroleptospirosis Diagnosed by Metagenomic NGS (mNGS) in 48h

Cured 2 weeks after NGS dx with appropriate treatment

Plan:

1 - Translating a Research pipeline to a Clinical assay: The challenges of mNGS ID diagnostics

2 - Prospective case series to study which cases benefit from mNGS ID diagnostics
1 - Translating a Research pipeline to a Clinical assay: The challenges of mNGS ID diagnostics
mNGS pipeline

Sample → NA extraction

library → Library prep

QC → library

library → Sequencing

Sequencing → Analysis

Analysis → Results

Results → Report
Challenges of unbiased pathogen sequencing

1. Only small % of nucleic acid in clinical samples are from pathogens

2. Pathogens have various types of nucleic acids present

3. Quick turnaround time

4. Controls are needed for validation

5. NGS sensitivity leads to potential contamination

6. Can be performed by clinical labs
Challenges of unbiased pathogen sequencing

1. Only small % of nucleic acid in clinical samples are from pathogens

2. Pathogens have various types of nucleic acids present

3. Quick turnaround time

4. Controls are needed for validation

5. NGS sensitivity leads to potential contamination

6. Can be performed by clinical labs
mNGS pipeline – the needle in the haystack

**Diagram**

- **Sample** → **NA extraction**
- **QC** → **library** → **Library prep**
- **Sequencing**
- **Analysis** → **Results** → **Report**
mNGS pipeline – the needle in the haystack

Not all samples are equal
mNGS pipeline – the needle in the haystack

Sample

QC

Sequencing

Analysis

Results

Report

GI and Food Microbiome background

High host genomic background
mNGS pipeline

CSF

QC

Sequencing

Analysis

Results

Report

Stack

human | bacterial | viral | other

stool

respiratory

cerebrospinal fluid

serum

tissue

0% | 25% | 50% | 75% | 100%

1

2

3 (BAL)
mNGS pipeline – the needle in the haystack

Sample processing:
1 – Depleting host background
   DNAse
   RNAse
   Ribo-zero
   Methylated-DNA pull-down
   Duplex-specific nucleic acid depletion
   ...
2 – Enriching for Pathogens
   Multiplex PCR
   Probe enrichment
   ...

CSF

QC

Library prep

Library

NA correction

Sequencing

Analysis

Results

Report
mNGS pipeline – the needle in the haystack

1. CSF
2. Total NA extraction
3. Post-DNAse treatment
4. RNA clean-up
5. ds cDNA synthesis
6. Host background removal
7. Library prep
8. RNA library
9. Sequencing
10. Analysis
11. Results
12. Report
13. QC
mNGS pipeline – the needle in the haystack

CSF → Total NA extraction → Post-DNAs treatment → RNA clean-up → ds cDNA synthesis → Host background removal

QC → Sequencing → Library prep → Sequencing Depth is related to the Limits of Detection (LOD)

Rapid HiSeq
40 million reads/sample

Analysis → Results → Report
mNGS pipeline – the needle in the haystack

**SURPI =**

*Sequencing-based Ultra Rapid Pathogen Identification*

**Host background removal**

- Post-DNase treatment

- RNA clean-up

- ds cDNA synthesis

**Analysis**

- Raw sequence reads
- Preprocessing
  - SNAP alignment (nucleotide) to human DB
  - SNAP alignment (nucleotide) to bacterial DB
  - SNAP alignment (nucleotide) to viral DB
  - De novo contig assembly (ABYSS + Minimo)
  - RAPSearch alignment (translated nucleotide) to viral protein DB
  - BACTERIAL
  - VIRAL
  - Fungal
  - Parasitic
  - Other

**Result**

Naccache et al., 2014. *Genome Research*

**CSF**

Rapid HiSeq

40 million reads/case

**QC**

**Sequencing**

**RNAlibrary**

**TOTAL NA extraction**

SURPI starts with rapid host subtraction
Challenges of unbiased pathogen sequencing

1. Only small % of nucleic acid in clinical samples are from pathogens

2. **Pathogens have various types of nucleic acids present**

3. Quick turnaround time

4. Controls are needed for validation

5. NGS sensitivity leads to potential contamination

6. Can be performed by clinical labs
mNGS pipeline – RNA vs DNA

1. **CSF**
2. **Total NA extraction**
3. **Post-DNAsese treatment**
4. **RNA clean-up**
5. **ds cDNA synthesis**
6. **Library prep**
7. **RNA library**
8. **Library prep**
9. **Analysis**
10. **Results**
11. **Report**
mNGS pipeline – RNA vs DNA

1. CSF
2. Total NA extraction
3. Post-DNAse treatment
4. RNA clean-up
5. ds cDNA synthesis
6. Library prep
7. RNA library
8. Sequencing
9. QC
10. SURPI
11. Results
12. Report

To identify RNA virus

mNGS pipeline – RNA vs DNA
mNGS pipeline – RNA vs DNA

CSF

Total NA extraction

Post-DNAs treatment

RNA clean-up

ds cDNA synthesis

Total NA extraction

Library prep

DNA library

RNA library

Library prep

DNA library

QC

To identify DNA virus, Bacteria and Eukaryotes

To identify RNA virus

Sequencing

Analysis

Results

Report

mNGS pipeline – RNA vs DNA

CSF

Total NA extraction

Post-DNAs treatment

RNA clean-up

ds cDNA synthesis

Total NA extraction

Library prep

DNA library

RNA library

Library prep

DNA library

QC

To identify DNA virus, Bacteria and Eukaryotes

To identify RNA virus

Sequencing

Analysis

Results

Report
Challenges of unbiased pathogen sequencing

1. Only small % of nucleic acid in clinical samples are from pathogens

2. Pathogens have various types of nucleic acids present

3. **Quick turnaround time**

4. Controls are needed for validation

5. NGS sensitivity leads to potential contamination

6. Can be performed by clinical labs
mNGS pipeline – Fast turnaround time

- CSF
- Total NA extraction
- Post-DNAsese treatment
- RNA clean-up
- ds cDNA synthesis
- Library prep
- DNA library
- RNA library
- Library prep
- QC
- Sequencing
- Analysis
- Results
- Report
mNGS pipeline – Fast turnaround time

- **CSF**
  - Total NA extraction
  - Post-DNase treatment
- **DNA library**
  - Nextera XT
  - RNA clean-up
  - ds cDNA synthesis
- **RNA library**
  - Nextera XT
- **QC**
- **Sequencing**
- **Analysis**

1h30
mNGS pipeline – Fast turnaround time

1h30
mNGS pipeline – Fast turnaround time

CSF

Bioanalyzer

Sequencing

Rapid Hiseq 100bp SE

Analysis

Results

Report

SURPI = Sequencing-based Ultra Rapid Pathogen Identification

SURPI uses SNAP and RAPSearch fast aligners Naccache et al., 2014. Genome Research
mNGS pipeline – Fast turnaround time

SHIFT 1
- CSF
- Total NA extraction
- Post-DNase treatment

SHIFT 2
- DNA library
- Nextera XT
- RNA cleaning-up
- ds cDNA synthesis
- Bioanalyzer
- RNA library
- Nextera XT
- 1h30

SHIFT 3
- Analysis
- Results
- Report
- 56h turnaround time

SURPI uses SNAP and RAPSearch fast aligners
Challenges of unbiased pathogen sequencing

1. Only small % of nucleic acid in clinical samples are from pathogens

2. Pathogens have various types of nucleic acids present

3. Quick turnaround time

4. **Controls are needed for validation**

5. NGS sensitivity leads to potential contamination

6. Can be performed by clinical labs
mNGS pipeline - Validation

1. **CSF**
   - Total NA extraction
     - Post-DNAses treatment
   - RNA clean-up
   - ds cDNA synthesis

2. **Bioanalyzer**
   - DNA library
     - Nextera XT
     - RNA library
     - Nextera XT

3. **Sequencing**
   - Analysis
     - Results
     - Report
mNGS pipeline - Validation

Positive control

HIV (RNA virus)
CMV (DNA virus)
*Streptococcus agalactiae* (gram “+”)
*Klebsiella pneumoniae* (gram “-”)
*Cryptococcus neoformans* (yeast)
*Aspergillus niger* (mold)
*Toxoplasma gondii* (parasite)

→ LOD validated for 7 Organisms
mNGS pipeline

Post-DNase treatment

RNA clean-up

ds cDNA synthesis

Analysis

Results

Report

SURPI+ automatically compares samples to controls to call ID positive

Positive control

CSF

Total NA extraction

HIV (RNA virus)
CMV (DNA virus)
*Streptococcus agalactiae* (gram“+”)
*Klebsiella pneumoniae* (gram“-”)
*Cryptococcus neoformans* (fungus)
*Aspergillus niger* (mold)
*Toxoplasma gondii* (parasite)
Challenges of unbiased pathogen sequencing

1. Only small % of nucleic acid in clinical samples are from pathogens

2. Pathogens have various types of nucleic acids present

3. Quick turnaround time

4. Controls are needed for validation

5. **NGS sensitivity leads to potential contamination**

6. Can be performed by clinical labs
mNGS pipeline - Contaminations

CSF

Post-DNase treatment

Total NA extraction

Bioanalyzer

DNA library

Nextera XT

RNA library

Nextera XT

ds cDNA synthesis

RNA clean-up

Positive control

Sequencing

Analysis

Results

Report

mNGS pipeline - Contaminations

Bioanalyzer

DNA library

Nextera XT

RNA library

Nextera XT

ds cDNA synthesis

RNA clean-up

Positive control

Sequencing

Analysis

Results

Report
mNGS pipeline - Contaminations

CSF

Positive control

Negative control

Total NA extraction

Post-DNase treatment

RNA clean-up

ds cDNA synthesis

Bioanalyzer

DNA library

Nextera XT

RNA library

Nextera XT

Sequencing

Analysis

Results

Report
mNGS pipeline - Contaminations

- **CSF**
  - Positive control
  - Negative control

- Total NA extraction
- Post-DNAse treatment
- RNA clean-up
- ds cDNA synthesis
- Bleach surfaces Single use aliquots

- Bioanalyzer
- DNA library
- RNA library
- Nextera XT

- Sequencing
- Analysis
- Results
- Report
mNGS pipeline - Contaminations

- CSF
- Positive control
- Negative control
- Pre-PCR room Formaldehyde
- Total NA extraction
- Post-DNAs treatment
- RNA clean-up
- ds cDNA synthesis
- Double barcoding
- Bioanalyzer
- DNA library
- Nextera XT
- RNA library
- Nextera XT
- Sequencing
- Analysis
- Results
- Report
mNGS pipeline - Contaminations

CSF

Positive control

Negative control

Pre-PCR room
Formaldehyde

Total NA extraction

Post-DNAs treatment

Bleach surfaces
Single use aliquots

CSF to Total NA extraction

DNA library

Nextera XT

RNA library

Nextera XT

Double barcoding

ds cDNA synthesis

RNA clean-up

Bioanalyzer

DNA library

Sequencing

RNA library

Analysis

Results

Report

More washes
Barcode cycling
Reduced multiplexing
mNGS pipeline - Contaminations

CSF

Positive control

Total NA extraction

Post-DNase treatment

RNA clean-up

DNA library

Nextera XT

Double barcoding

ds cDNA synthesis

Bioanalyzer

RNA library

Nextera XT

Analysis

Results

Report

More washes
Barcode cycling
Reduced multiplexing

SURPI+ uses Neg control to subtract contamination background
Challenges of unbiased pathogen sequencing

1. Only small % of nucleic acid in clinical samples are from pathogens

2. Pathogens have various types of nucleic acids present

3. Quick turnaround time

4. Controls are needed for validation

5. NGS sensitivity leads to potential contamination

6. Can be performed by clinical labs
mNGS pipeline—In the clinic

- performed in CLIA lab
- by CLS personnel
- according to approved LDT protocol
mNGS pipeline—In the clinic

CSF
Recorded in PHI

Positive control

Total NA extraction

Post-DNase treatment

RNA clean-up

ds cDNA synthesis

CSF

Bioanalyzer

DNA library

Nextera XT

Nextera XT

RNA library

Sequencing

Analysis

Results

Report
mNGS pipeline—In the clinic

- **CSF**
  - Recorded in PHI
  - **Positive control**
  - **Negative control**
  - Total NA extraction
  - Post-DNAlase treatment
  - RNA clean-up
  - ds cDNA synthesis
  - Lot numbers

- **Bioanalyzer**
  - DNA library
  - Nextera XT
  - RNA library
  - Nextera XT

- **Sequencing**
  - Analysis
  - Results
  - Report
mNGS pipeline – In the clinic

CSF Recorded in PHI

Positive control

Total NA extraction

Post-DNAs treatment

RNA clean-up

ds cDNA synthesis

Lot numbers

Bioanalyzer

DNA library

Nextera XT

RNA library

Nextera XT

Sequencing

RNA library

Results

SURPIClin

SURPIviz

Report

Reviewed by clinician
# SURPIviz

The visual home of SURPI+, a cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing for diagnostics and research.

View results from the SURPI+ runs below:

<table>
<thead>
<tr>
<th>#</th>
<th>SURPI+ Run</th>
<th>Sequencing</th>
<th>SURPI+ files</th>
<th>Heat maps</th>
<th>Krona plots</th>
<th>Old Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Demo Clinical</td>
<td>Sequencing</td>
<td>SURPI+ files</td>
<td>Heat maps</td>
<td>Krona plots</td>
<td>Old Coverage</td>
</tr>
<tr>
<td>1</td>
<td>Demo Spike-in</td>
<td>Sequencing</td>
<td>SURPI+ files</td>
<td>Heat maps</td>
<td>Krona plots</td>
<td>Old Coverage</td>
</tr>
<tr>
<td>2</td>
<td>1PilotBoardMeeting</td>
<td>Sequencing</td>
<td>SURPI+ files</td>
<td>Heat maps</td>
<td>Krona plots</td>
<td>Old Coverage</td>
</tr>
<tr>
<td>3</td>
<td>131025 JLG A1N2DNAseTest</td>
<td>Sequencing</td>
<td>SURPI+ files</td>
<td>Heat maps</td>
<td>Krona plots</td>
<td>Old Coverage</td>
</tr>
</tbody>
</table>
Assembled from 236615 reads; plotting average coverage over 23 bp segments.

**Graph:**
- **X-axis:** Base pairs
- **Y-axis 1:** Mapped reads
- **Y-axis 2:** Percent identity

**Table:**

<table>
<thead>
<tr>
<th>Reference length bp</th>
<th>Coverage in bp</th>
<th>Percent coverage</th>
<th>Avg coverage depth</th>
<th>Covered pct ident</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>10936</td>
<td>10875</td>
<td>99.44</td>
<td>2936.98</td>
</tr>
<tr>
<td>Displayed</td>
<td>10936</td>
<td>10875</td>
<td>99.44</td>
<td>2936.98</td>
</tr>
</tbody>
</table>
St. Louis encephalitis virus (gi\[537790028\], 10936 bp)

St. Louis encephalitis virus strain CbaAr-4005, complete genome

Assembled from 236615 reads; plotting average coverage over 23 bp segments.

<table>
<thead>
<tr>
<th></th>
<th>Reference length bp</th>
<th>Coverage in bp</th>
<th>Percent coverage</th>
<th>Avg coverage depth</th>
<th>Covered pct ident</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>10936</td>
<td>10875</td>
<td>99.44</td>
<td>2936.98</td>
<td>99.37</td>
</tr>
<tr>
<td>Displayed</td>
<td>10936</td>
<td>10875</td>
<td>99.44</td>
<td>2936.98</td>
<td>99.37</td>
</tr>
</tbody>
</table>
2 - Prospective case series to study which cases benefit from mNGS ID diagnostics
Prospective Case series

ID specialist → UCSF Clinical Micro → Patient Consented → Excess Sample Sequestered

Research report

Chiu Lab mNGS
Prospective Case series

ELIGIBLE PATIENTS
n=64

EXCLUDED
n=37

INCLUDED
n=27

INPATIENT
n=20

OUTPATIENT
n=7

Incomplete consent n=30
No excess sample n=3
Inadequate sequencing n=4

27 Consecutive patients
22 Months

Samples

35

CSF
plasma/serum
blood
FFPE
NP
tissue
culture
BAL
stool

(Naccache, et al., 2016, manuscript in preparation)
Will you be the next Dr. House?
3 case reviews
Case Report I: 70 y/o male with fever and pancytopenia

- Evaluated for 2 months of fevers, chills, diarrhea, and fatigue

- Pancytopenic [WBC 2,000 / mm\(^3\), HgB 8 g/dL (12.4 – 14.9 g/dL), platelets 51 / mm\(^3\)] with elevated liver enzymes (AST = 294 U/L, ALT = 212 U/L)

- On imaging, hepatomegaly, splenomegaly, and liver cysts

- Bone marrow biopsy and lymph node biopsy performed; treated empirically with antibiotics, blood transfusion, G-CSF

- Concern for HLH (hemophagocytic lymphohistocytosis) with negative infectious workup → transferred from community to tertiary care hospital

- 6-week hiking in Spain 10 months prior to admission (500 mile Camino de Santiago)
Case Report I: 70 y/o male with fever and pancytopenia

FFPE bone marrow → SURPI
Case Report I: 70 y/o male with fever and pancytopenia

→ mNGS Dx: Visceral Leishmaniasis

Treated with 2 weeks of IV amphotericin

→ clinically improved, now doing well
Case Report II: 15 y/o female with hemorrhagic encephalitis

• 15 year-old girl with type 1 diabetes
• 7 days of headache, vomiting, arm weakness, and confusion
• Elevated WBC count in CSF (pleocytosis); hemorrhagic frontal / occipital lobe brain lesions on MRI
• No international travel, no sick contact, no insect bites
• Contact with alpacas, swimming in freshwater pond 9 months prior
• Extensive diagnostic testing negative
• “Classic” presentation for hemorrhagic form of ADEM → given IV steroids x 5 days
Case Report II: 15 y/o female with hemorrhagic encephalitis

CSF → SURPI
Case Report II: 15 y/o female with hemorrhagic encephalitis

→ mNGS Dx: *Balamuthia mandrillaris* encephalitis

- Brain biopsy pathology from HD 6 → hemorrhagic necrotizing process with numerous amoebae
- Despite combination amoebicidal therapy, she developed intracranial hypertension, cardiac arrest, and died
- Had a complete *B. mandrillaris* genome been available in Genbank NT, it would have been detected in 1st LP

(Greninger, et al., 2015, *Genome Medicine*, 7:113)
Case Report III: 55 y/o male with deafness and behavioral change

- 55 y/o male with bone marrow transplant May, 2013. Developed rapidly progressive hearing loss over 2-3 weeks in October.

- CSF unremarkable; MRI negative; PCR for HSV and enterovirus negative; treated empirically with high-dose valacyclovir, antibiotics, IVIg and steroids daily to no effect.

- Over next few weeks, developed nausea, fatigue, ataxia, persistent hearing loss, then depressed, irritable mood (unusual for patient per his wife).

- Repeat MRI late December → abnormal signal in thalamus and midbrain bilaterally; frontal lobe biopsy performed (only accessible region).
Case Report III: 55 y/o male with deafness and behavioral change

Brain Biopsy ➔ SURPI
Case Report III: 55 y/o male with deafness and behavioral change  
→ Genome Assembly of an Astrovirus

- Origin of virus unknown, but presumably community-acquired
- Patient started on ribavirin and IVIg
- Despite treatment, he continued to deteriorate, and passed away 4 months after NGS diagnosis

(Naccache, et al., 2015, Clinical Infectious Diseases)
Prospective Case series

INCLUDED n=27

- Confirmation of presumptive organism n=8
  - Angiostrongylus
  - Leishmania
  - Sporothrix
  - MTB
  - H1N1
  - HHV7
  - Histoplasma
- Suspcion of Infxn n=9
  - Astroivirus
  - Leptospira
  - Aspergillus
  - WNV
  - HHV6
  - HHV6
- Rule out Infxn n=10
  - EBV

Legend:
- Neuro
- Flu
- SIRS
- Pulm
- Epidrm
- GI
- Hep
- Card

- No ID Dx
- Micro and mNGS Dx
- mNGS Dx only
- Micro Dx only

(Naccache, et al., 2016, manuscript in preparation)
Identification of Use Cases for the Precision Medicine Initiative

Prospective metagenomic NGS case series reveals “high-yield” instances:

1. Confirmation of ambiguous or suggestive laboratory results
2. Epidemiologic or clinical links pointing to infection
3. Broad differential diagnosis
4. “rule-out” assay in patients remaining undiagnosed despite extensive testing

(Naccache, et al., 2016, manuscript in preparation)
Clinical Implementation of Metagenomic Next-Generation Sequencing for Precision Diagnosis of Infectious Diseases
Timeline – Clinical CSF mNGS assay in full speed

- **CLIA laboratory validation**
  August 2015 - present

- **Clinical CSF mNGS assay launch**
  May, 2016

- **Clinical BAL mNGS assay launch (pneumonia)**
  End of 2016 / Early 2017

- **Clinical plasma mNGS assay launch (fever, sepsis)**
  August, 2016

- **UC CIAPM prospective study launch**
  May, 2016

- **Clinical CSF mNGS assay in full speed**

  - High yield cases chosen by PM consult services
  - 6 prospective CSF samples sequenced each week
  - Results reported in 56h
  - Cost-benefit and clinical outcomes analysis after 1 year, 300 samples
Acknowledgements

UCSF Chiu Lab and VDDC
Charles Chiu, MD/PhD
Erik Samayoa, BS, CLS
Samia Naccache, PhD
Jerome Bouquet, PhD
Guixia Yu, BS
Scot Federman, BA
Alex Greninger, MD/PhD
Sneha Somasekar, BS
Doug Stryke, MS
Steve Miller, MD/PhD
Elizabeth Pham, BS
Shaun Arevalo, BS, CLS
Becky Fung, BS, CLS
Lauri Greene, BS
Tony Li, BS

Abbott Diagnostics
John Hackett, PhD

UCSF DeRisi Laboratory
Michael Wilson, MD
Joseph DeRisi, PhD

CDPH
Shigeo Yagi, PhD
Carol Glaser, MD
Dongxiang Xia, PhD
Sharon Messenger, PhD
Debra Wadford, PhD

University of Wisconsin
James Gern, MD
Sheryl Henderson, MD/PhD
Christine Seroogy, MD

Children’s Hospital, Washington DC
Joseph Campos, PhD
Brittany Goldberg, MD

American Red Cross
Susan Stramer, PhD

UCL Hospitals
Karl S Peggs, MD/PhD
Vanya Gant, MD/PhD

Funding
• NIH/NHLBI R01, NIH/NIAID R21
• UC Discovery Award
• Abbott Pathogen Discovery Award
• CIAPM Award
• Sandler / Bowes Foundation Award