24.001: Whole genome sequencing for diagnosing infectious diseases

Paul D Brown, PhD
The University of the West Indies, Mona, Jamaica
ICID2018, Buenos Aires, Argentina
Session # 24, Clinical Applications of Whole Genome Sequencing

Disclosures

- Nothing to declare
Main Goals of Clinical Microbiology

• What organism(s) is(are) causing the infection?
  • Species identification
  • Single or multiple aetiological agents

• What drugs (antimicrobials) can be used to treat it?
  • Phenotypic characterization (AMR and virulence)

• How is it related to similar infections?
  • Evolutionary traits and pathogen tracking
    • Part of outbreak investigations; Important for surveillance

Transforming Clinical Microbiology with WGS

• Allows sequencing of the whole genome of numerous pathogens in one sequence run, either from bacterial isolates of (different) patients, or from multiple species present in material from one individual (metagenomics)
• Both the investment- and the running costs have decreased dramatically during the last decade
• Single protocol can be used for multiple pathogens for identification and characterization (typing)
**WGS Strategy**

- A step further than molecular diagnostics

**Platforms**

- WGS
- Whole-exome sequencing (WES)
- Metagenomic sequencing
- Targeted gene sequencing

---

**Properties of current NGS platforms.**

<table>
<thead>
<tr>
<th>Company</th>
<th>Equipment</th>
<th>Output/run (Gb)</th>
<th>Maximum read length (bp)</th>
<th>Reads (x10^6)</th>
<th>Running time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illumina</td>
<td>MiSeq</td>
<td>0.6–7.5</td>
<td>2 × 150</td>
<td>25</td>
<td>4–24 h</td>
</tr>
<tr>
<td>Illumina</td>
<td>Miseq</td>
<td>0.3–1.5</td>
<td>2 × 300</td>
<td>25</td>
<td>5–55 h</td>
</tr>
<tr>
<td>Illumina</td>
<td>NexSeq</td>
<td>20–120</td>
<td>2 × 150</td>
<td>130–400</td>
<td>12–30 h</td>
</tr>
<tr>
<td>Illumina</td>
<td>HiSeq 3000</td>
<td>125–700</td>
<td>2 × 150</td>
<td>2500</td>
<td>&lt;1–3.5 days</td>
</tr>
<tr>
<td>ThermoFisher</td>
<td>Ion PGM™</td>
<td>0.03–2</td>
<td>200–400</td>
<td>6.4–5.5</td>
<td>2–7 h</td>
</tr>
<tr>
<td>ThermoFisher</td>
<td>Ion 55™</td>
<td>0.6–15</td>
<td>200–400</td>
<td>3–80</td>
<td>2.5–4 h</td>
</tr>
<tr>
<td>ThermoFisher</td>
<td>Ion 55™ XL</td>
<td>0.6–15</td>
<td>200–400</td>
<td>3–80</td>
<td>&lt;24 h</td>
</tr>
<tr>
<td>Oxford Nanopore</td>
<td>MinION</td>
<td>21–42</td>
<td>230,000–300,000</td>
<td>2.2–4.4</td>
<td>1 min–48 h</td>
</tr>
<tr>
<td>Pacific Biosciences</td>
<td>Sequel</td>
<td>0.75–1.25</td>
<td>&gt;20,000</td>
<td>379,000</td>
<td>30 min–6 h</td>
</tr>
<tr>
<td>Pacific Biosciences</td>
<td>RSII</td>
<td>0.5–1</td>
<td>&gt;20,000</td>
<td>52,000</td>
<td>30 min–4 h</td>
</tr>
</tbody>
</table>

* The Pacific Biosciences data are per smart cell; both the Sequel and the RSII can run 1–16 smart cells in one run.
**WGS Strategy**

- Fragmentation of template DNA
- Library preparation
- Sequencing
- Processing and data analysis
- (Diagnosis and treatment/interventions)

Lefterova et al., J Mol Diagn 2015

---

**Transforming Clinical Microbiology with WGS**

- Workflow anticipated after adoption of WGS of cultured isolates or directly on the clinical samples, with an expected time scale

Hasman et al., J Clin Microb 2014

---

**Table 1. Summary of In Silico Validation Results, Accuracy**

<table>
<thead>
<tr>
<th>Total</th>
<th>Taxa</th>
<th>Simulated Samples</th>
<th>Accuracy, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA viruses</td>
<td>73</td>
<td>230</td>
<td>100</td>
</tr>
<tr>
<td>DNA viruses</td>
<td>29</td>
<td>70</td>
<td>100</td>
</tr>
<tr>
<td>Bacteria</td>
<td>82</td>
<td>329</td>
<td>99.1</td>
</tr>
<tr>
<td>Fungi</td>
<td>1</td>
<td>6</td>
<td>100</td>
</tr>
</tbody>
</table>

*The table shows numbers of species and viral types (taxa), number of different reference sequences used to generate simulated samples representing the genetic intraspecies or intratype variability (simulated samples), and percent of simulated samples that were correctly identified as positive (accuracy).*

Schlaberg et al, Arch Pathol Lab Med—Vol 141, June 2017
Transforming Clinical Microbiology with WGS

• Potential clinical applications for metagenomics sequencing

Goldberg et al., mBio 2015

Transforming Clinical Microbiology with WGS

• Assessment of the Performance Characteristics of WGS-Based Tests for Clinical Microbiology

Lefterova et al., J Mol Diagn 2015

Performance characteristic | Approach to evaluation
---|---
Accuracy | Use of specimens with known findings and confirmation of additional findings detected during validation by an orthogonal method.
Precision | Reproducibility (between-run precision): sequencing of the same samples on different runs.
Repeatability (within-run precision): sequencing of the same samples in replicates within a run.
Between-library precision: sequencing different library preparations of the same samples on the same sequencing run. 
Analytical sensitivity | Microbial variant detection: mixes of known variant strains and wild-type strains at different percentages and at low, medium, and high levels (eg, viral loads).
Microbial identification: serial dilutions of samples in an appropriate matrix containing a known pathogen(s) coupled with an estimation of the minimum coverage needed to detect the pathogen.
Analytical specificity | Microbial variant detection and microbial identification: estimation of the false-positive rate at various read depths.
Transforming Clinical Microbiology with WGS

• Characterization and surveillance of pathogens, incl. identification in polymicrobial samples
• Correlation of genomic features with strains of clinical importance
• Generating regional phylogenomic sublineages
• Metagenomics approaches on clinical samples
• Outbreak management
• Determination of transmission of animal to human zoonoses
• Tracking of MDR organisms in hospitals and communities

Transforming Clinical Microbiology with WGS

• Outbreak management
From the trenches: **pathogen identification** - **Francisella tularensis** subspecies differentiation

- Potential biosecurity issue; Preventing need for isolation

![Diagram of pathogen identification process](Koruda et al. JCM 2012)

**From the trenches: pathogen identification – Burholderia pseudomallei** association with environmental samples

- Potential biosecurity issue

![Map showing melioidosis cases](Chappelle et al. EID 2015)

Melioidosis cases reported (1945–2017)
Sanchez-Villamil & Torres. Trop Med Inf Dis 2018

![Diagram of pathogen identification process](Chappelle et al. Microb Genom 2016)
From the trenches: *pathogen identification – Melioidosis distribution in the tropics*

Perumal Samy et al. (2017) PLOS Negl Trop Dis

From the trenches: *pathogen identification – Mastoiditis caused by Fusobacterium nucleatum–Actinomyces israelii coinfection*

- Resolving discrepancies in conventional microbiology

Salipante et al JCM 2014

A rare aggregate of branching, Gram-positive rods is depicted, consistent with Nocardia or Actinomyces species
From the trenches: *pathogen identification* – discovery of novel arenavirus

• First study for new pathogen detection and discovery

Palacios et al NEJM 2008

From the trenches: *pathogen identification*

• Reveals *Coccidioides immitis* cluster in organ transplant patients

Engelthaler et al Emerg Infect Dis. 2011
From the trenches: **pathogen identification** - novel genomospecies of *Bacteroides*

- Potential to enhance diagnostic accuracy in identifying clinical microbial isolates

Salipante et al Emerg Inf Dis 2015

---

From the trenches: **pathogen identification** - Determination of anthrax outbreak among European heroin users

- Confirmation of accidental heroin contamination into supply along trafficking route

Price et al Emerg Inf Dis 2012
From the trenches: *pathogen identification* - Encephalitis by neuroinvasive astrovirus

- Capability of detecting all potential pathogens simultaneously

Naccache et al. Clin Inf Dis 2012

---

From the trenches: *pathogen identification*

- SURPI (sequence-based ultra-rapid pathogen identification)

Naccache et al. Genome Res 2014
From the trenches: **pathogen identification** – gene flow and heterogeneity during Legionnaires disease outbreak

McAdam et al. BMC Genom Biol 2014. LD (ST191) outbreak in Edinburgh, UK

From the trenches: **pathogen identification**

Ferdous et al. CMI 2016; Dutch STEC isolates
From the trenches: *pathogen identification*

- Plasmids sequenced from 26 Shiga toxin-producing and non-producing *E. coli* strains
- Identified 39 plasmids that carried various virulence and putative virulence genes; 8 with resistance to various antibiotics and some also had resistance genes for heavy metals.

---

*Chiu et al. EID 2015*
From the trenches: **pathogen identification**

- Pathogen diversity (HIV and Pegivirus)
  - HPgV has a worldwide distribution and is transmitted sexually, parenterally and by mother-to-child transmission, and as a consequence HIV patients are often co-infected
  - HPgV sequences were detected in 9 of 35 (26%) specimen libraries, with 8 yielding at least 99% of the viral genome.

The vision of PulseNet International is for WGS to be used in all public health laboratories to identify, characterize and subtype food-borne pathogens.
From the trenches: pathogen identification - Parvovirus

- NGS data sets processed from 2008 to 2013. Bar graphs show the percentage of PHV reads in each data set, with the number of PHV reads given above the bars.
- Sample sets extracted using Qiagen spin columns (red); those extracted using other methods (black)
From the trenches: *from pathogen to pathobiome*

- **Pathobiome concept**
- **Objectives:**
  - Accurately describe the biodiversity of relevant microbial communities
  - Shift from descriptive to functional metagenomics
  - Understand the impact of microbial communities on the persistence, transmission, and evolution of pathogens

From the trenches: *treatment and management implications*

- For clinical WGS (where there is cohort studies biobanking), pertinent ELSI issues include the interpretation of data, data storage, data sharing, informed consent and identifiability/privacy

- For infectious disease WGS, issues include who (else) is at risk, and potential benefits/harms of healthcare policy that accrue to the entire population (source and contact tracing)

Vayssier-Taussat et al. 2014 Front Cell Infect Microbiol
From the trenches: treatment and management implications

Geller et al. Genome Medicine 2014, 6:106

From the trenches: treatment and management implications

AAM 2016
Is it feasible to routinely use WGS in the clinical microbiology laboratory?

- Every isolate that was recovered from culture on a single day at Houston Methodist Hospital was sequenced.
- 130 samples, including 107 aerobic cultures, 9 anaerobic cultures, and 14 acid-fast bacillus/mycology samples
- 115 isolates were correctly identified using WGS.
- 12 hours to extract DNA and prepare the sequencing libraries, 39 hours to perform the sequencing, and 2 to 4 hours for de novo assembly of contigs
- Isolates that were unable to be identified by sequencing were due to low read counts, insufficient sample preparation, or lack of a reference genome

To the future...

- Optimism for the future to deliver tangible clinical benefits to improve patient care, patient outcomes, and public health
- Could be proactive in identifying potential pathogenic outbreak species and lineages
- Pre-emptive use to monitor, detect, and control threatening agents locally, nationally, and internationally
- Will have profound implications in critical decision support systems and clinical care pathways
- Could become a cornerstone of clinical care