Metagenomic NGS
A new promising diagnostic tool?

Valérie CARO
Institut Pasteur Paris
Pôle de Génotypage des Pathogènes
Cellule d'Intervention Biologique d'Urgence
Unité Environnement et Risques Infectieux

valerie.caro@pasteur.fr

A promising single universal pathogen detection method for infectious diseases diagnostics?

- Challenge of accurate diagnosis due to a wide variety of pathogens, causing clinically similar diseases
- Current laboratory methods requiring a battery of tests
- Syndromic multiplex PCR, 16S rDNA sequencing, MALDI-TOF MS
- Slow turnaround (from several days to weeks)
- Etiology still unknown in up to 60% of infectious diseases cases
- Accurate information regarding pathogen identification leads to favorable clinical outcomes

→ Unbiased metagenomic NGS offer hypothesis-free, culture-independent, pathogen detection directly from clinical specimens

A brief story of sequencing

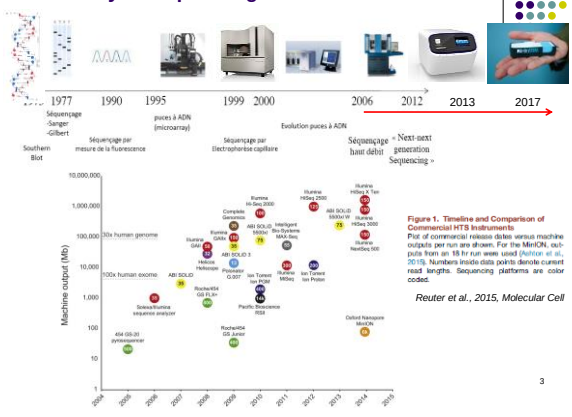
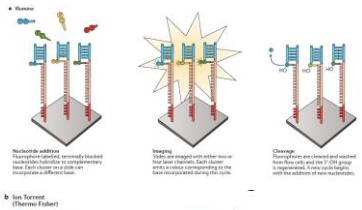


Figure 1. Timeline and Comparison of Commercial HTS Instruments
Plot of commercial release dates versus machine outputs per run are shown. For the MinION, outputs from an 18 hr run were used (Ashton et al., 2015). Numbers inside data points denote current read lengths. Sequencing platforms are color coded.

Reuter et al., 2015, Molecular Cell

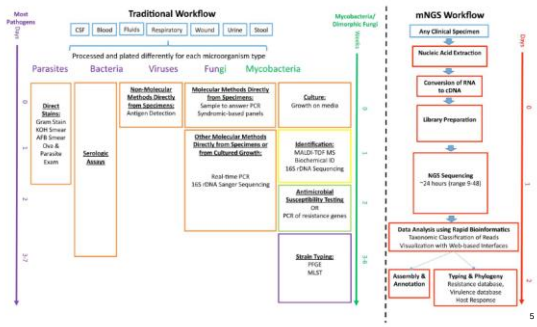
NGS 2nd generation (Sequencing by synthesis)



- Illumina
- Solid-phase bridge amplification (Cluster generation)
- Cyclic reversible termination (CRT) approach

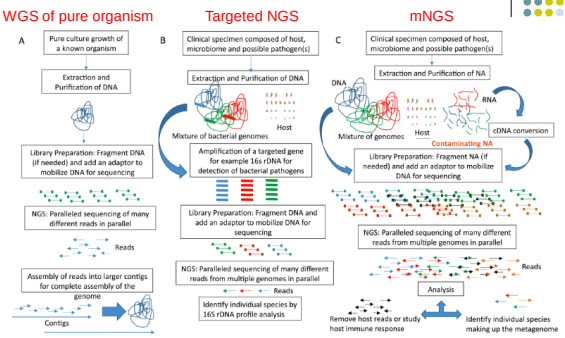
- Ion Torrent (Thermofisher)
- Emulsion PCR (micelle droplets), on-bead amplification
- Single-nucleotide addition (SNA) approach
- H+ release results in a 0,02 unit change in pH

Timeline and workflow in diagnostic medical microbiology laboratories



5

Different applications of NGS analysis



6

Interpretations of mNGS analysis

- Host nucleic acid amplification
  - Human genome 1000x larger than bacterial genomes
  - 99% host reads
  - <1% used reads to make successful diagnoses
  - Host nucleic acid depletion or pathogen reads enrichment
  - Sequencing depth (more reads = higher sensitivity and higher cost)
- Colonization vs Infection
  - Normal microbiota, transient colonizers, sample contamination, infection
  - Initial mNGS focused on « normally » sterile specimens
  - Quantify pathogens reads as a percentage of total number of sequence reads
  - Procedures to distinguish true pathogens from colonizers to be develop for mNGS

7

Interpretations of mNGS analysis

- Other exogenous sources of nucleic acid
  - Nucleic acid contamination occurs at several steps in the process from specimen collection to processing
  - handling, containers DNA/RNA free, specimen collection guidelines
  - « kit-ome » ubiquitous DNA in commonly used reagents
  - No template control should be included (reads filter)
  - Reads localization : spanning the genome vs restricted area
- Methods
  - No standardized protocol
  - DNA approach indicates the presence of organisms but RNA approach the transcription activity
  - Extraction methods (nucleic acid recovery not equal) : critical step
  - Internal control (negative mNGS)

8

Analytic Sensitivity of mNGS analysis

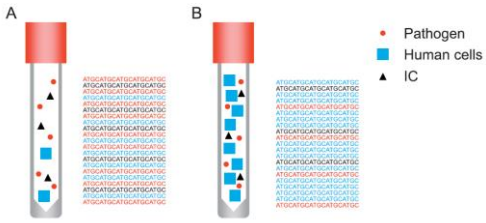
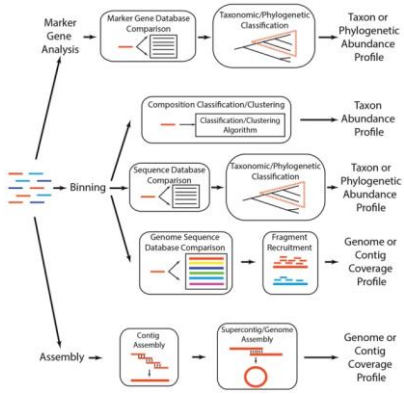


Figure 1. *Conceptual illustration of how cellularity and overall nucleic acid yield affect analytic sensitivity in metagenomic next-generation sequencing tests. At the same pathogen load (red circles), specimens with fewer patient cells (blue) have higher analytic sensitivity (A) than specimens with a greater number of human cells (B). Because DNA/RNA from both pathogens and patient cells are sequenced, the number of pathogen sequences (red) will be lower and the number of human sequences (blue) will be be higher in samples with greater numbers of patient cells (B). Internal controls spiked into patient samples at known concentrations (IC, black) can be used to quantify this effect because the number of IC sequences (black) will also be reduced in samples with greater numbers of patient cells (B). During test validation, cutoffs can then be established for minimal numbers of IC sequences.*

- Specimen collection : cellularity vs relative abundance of pathogens, other microorganisms, and patient cells
- Pathogen load may result in different sequence coverage depending on the total nucleic acid yield

9

Bioinformatics tools for mNGS data



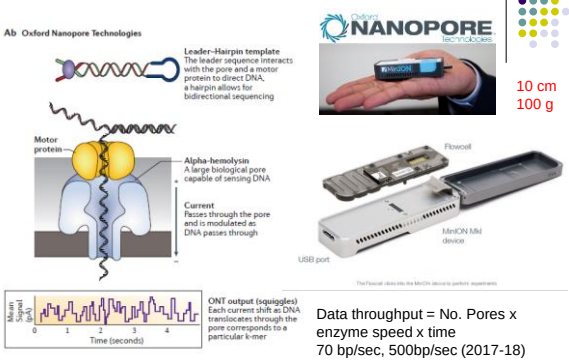
10

Bioinformatics tools for mNGS data

- Large quantity of data : storage and analysis challenges
- Pieplines rapidly align the reads to NCBI nt reference database and use taxonomic classification for more accurate read assignments

11

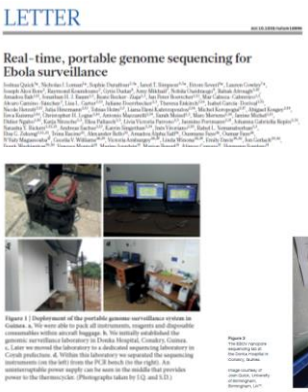
Nanopore sequencing ONT



Data throughput = No. Pores x enzyme speed x time
70 bp/sec, 500bp/sec (2017-18)

12

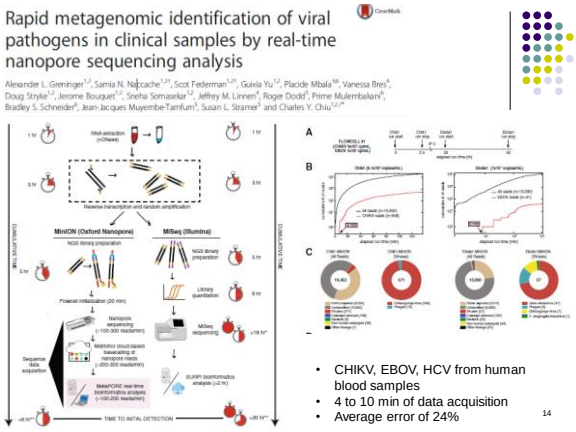
Nanopore sequencing ONT


LETTER

Real-time, portable genome sequencing for Ebola surveillance


MinION in space

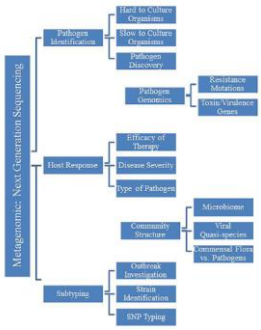
13


Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis

- CHIKV, EBOV, HCV from human blood samples
- 4 to 10 min of data acquisition
- Average error of 24%

14

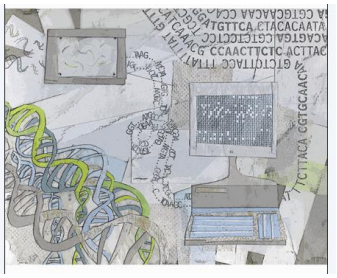
A promising single universal pathogen detection method for infectious diseases diagnostics?



- Characterization of pathogens without a priori knowledge directly from clinical specimens
- Differentiation of colonization from infection
- Laboratory and data analysis workflows still complex
- Specimen preparation, rapidly evolving data analysis algorithms, incomplete reference sequence databases
- Promised diagnostic tool, especially in immunocompromised and critically ill patients

15

Merci de votre attention !



16