Introduction to Metagenomics for Clinical Virology

Sarah Buddle

UCL Great Ormond Street Institute of Child Health

Session developed by Dr Cristina Venturini

Session structure

11:00-12:00: Introduction to metagenomics

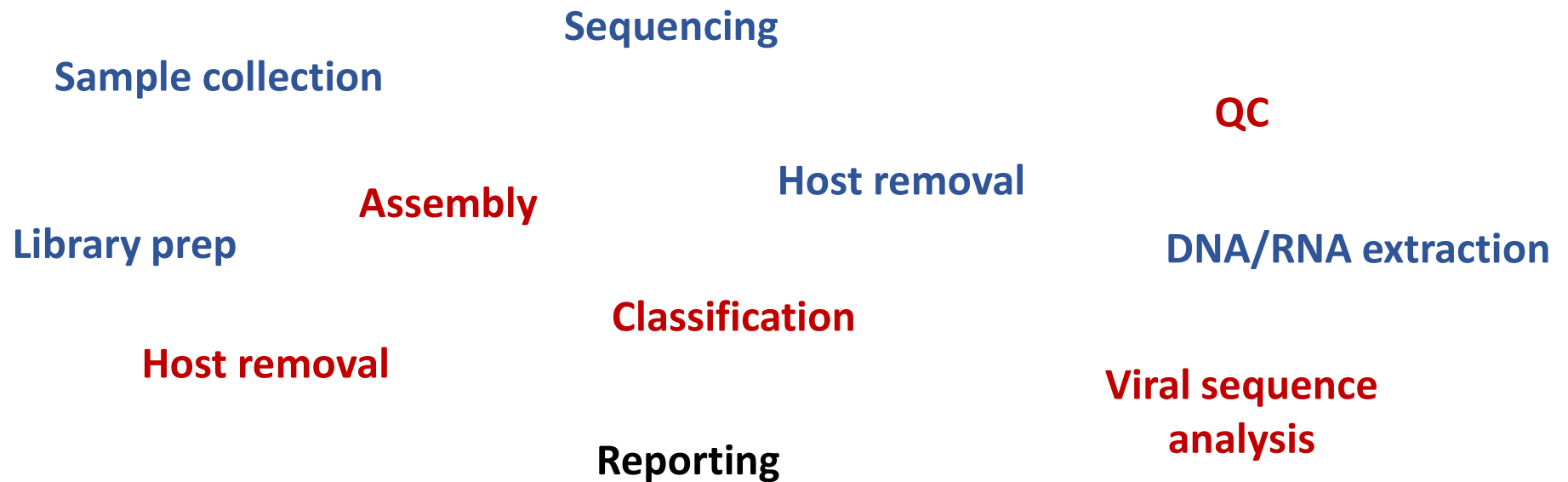12:00-13:00: Metagenomics bioinformatics practical

1. What is metagenomics?

2. What clinical questions can we answer with metagenomics?

3. What are the advantages and disadvantages of metagenomics over other techniques you might use to answer those questions?

Protocol

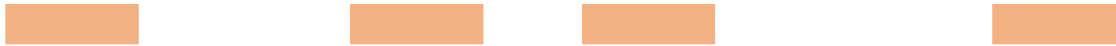What are the key steps in a metagenomics protocol?

What is the purpose of each step?

What methods might you use?

Sequencing

Sample collection

QC

Host removal

Assembly

Library prep

DNA/RNA extraction

Classification

Host removal

Viral sequence analysis

Reporting

Host removal: alignment
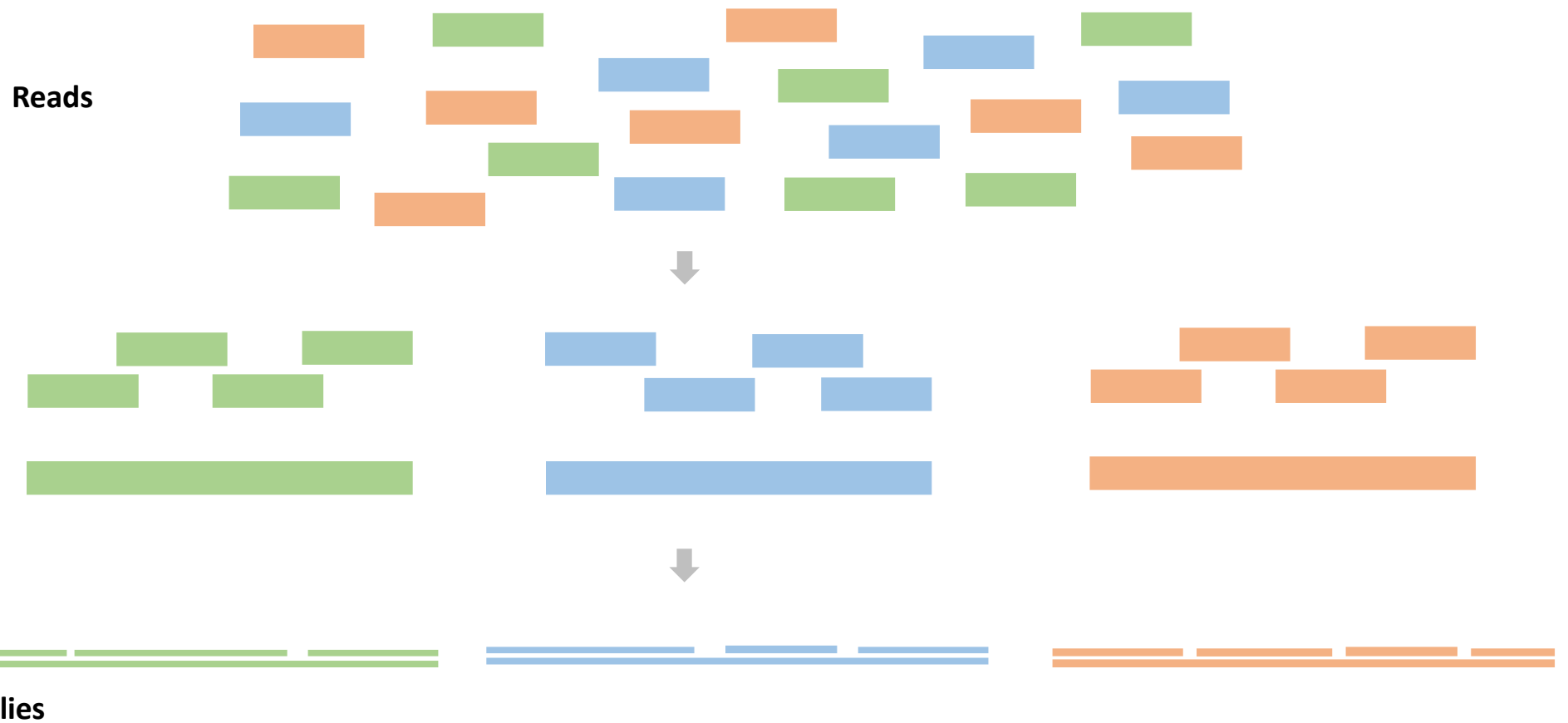
Sequencing reads

Human genome

Preliminary round with a quick classifier also an option

Assembly

Classification

Classification is deciding which species (or other taxonomic group) a read corresponds to

Reads are classified by comparison to a reference databases containing known genome sequences

Challenge: some parts of DNA are similar in different organisms

Classification tools

Alignment-based
E.g. BLAST, DIAMOND

K-mer-based
E.g. Kraken2, Centrifuge

Marker gene-based
E.g. mOTU, MetaPhlAn

Nucleotide-based
E.g. BLASTN

Protein-based
E.g. DIAMOND, Kaiju

Classification

What factors should we consider when choosing:

1: a classifier

2: sequences to include in your database

Contamination

1. Where might contamination come from?

2. How can we reduce/deal with contamination?

Practical: Metagenomics for diagnosis

Background:

We'll be analysing a clinical specimen from a 41-year-old patient who reported fever, chest tightness, cough, pain and weakness, and was admitted to hospital 6 days after the onset of the disease. The patient with no history of hepatitis, tuberculosis or diabetes. Preliminary investigations excluded the presence of influenza virus, *Chlamydia pneumoniae* and *Mycoplasma pneumoniae* and this was confirmed by PCR. Other common respiratory pathogens, including human adenoviruses, also tested negative. To investigate other possible infectious causes of disease, bronchoalveolar lavage fluid (BALF) was collected and deep metatranscriptomic sequencing was performed.

Task:

You need to analyse the data to determine the cause of the disease and determine what treatment options may be available.

Choosing bioinformatics protocols for metagenomics

The protocol shown in the practical is probably not the best one for your research or clinical question!

Some other tools: a non-exhaustive list

nf-core is a set of community-curated best practice bioinformatics pipelines built in Nextflow.
Taxprofiler Includes Kraken2/Bracken, DIAMOND, Centrifuge etc

Online, cloud-based, user-friendly tool

Illumina Dragen Metagenomics / Nanopore EPI2ME labs wf-metagenomics
Illumina and Nanopore's tools. Simple to run and can be automated.

Check benchmarking papers for lots of other options!