TELEVIR module

(From reads to virus detection)

Virus Detection

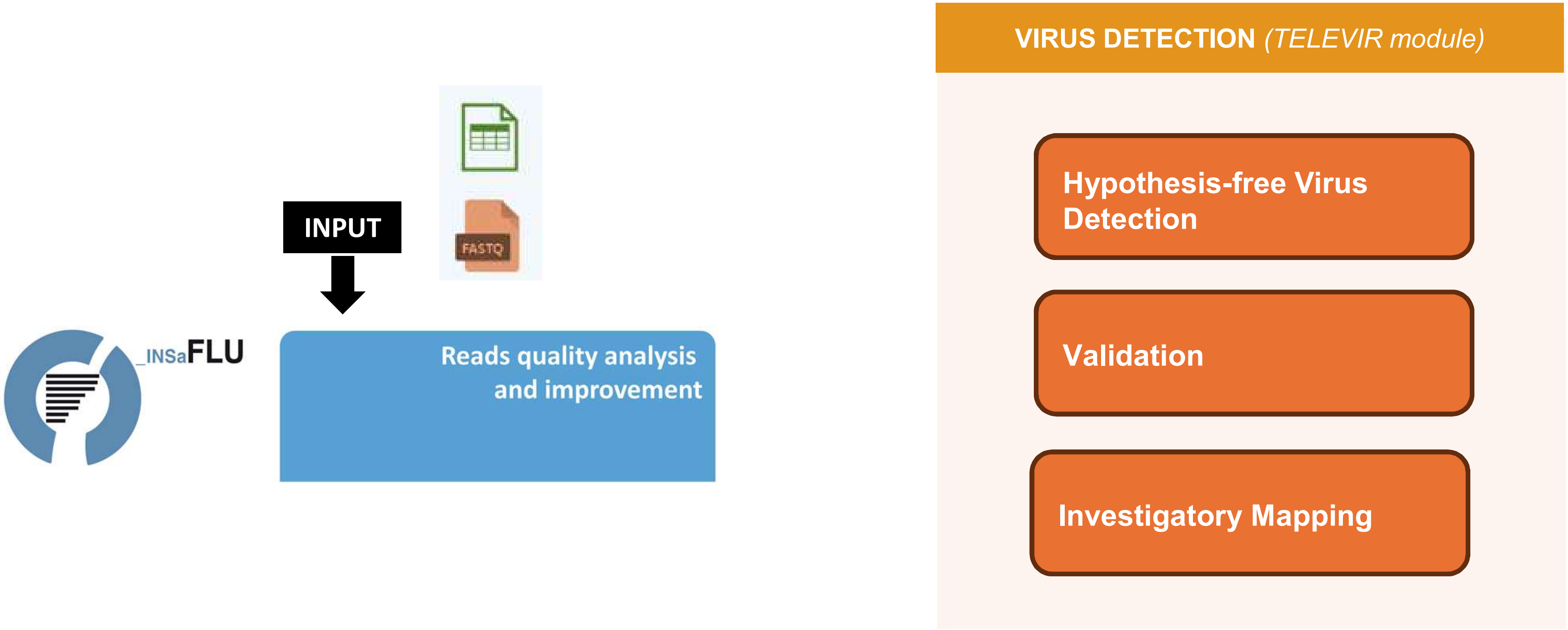
From Reads to viral detection

João Dourado
National Institute of Health Doutor Ricardo Jorge (INSA), Portugal
Department of Infectious Diseases
Genomics and Bioinformatics Unit

Basics

TELEVIR | Basics

Bioinformatics Diagnostics



INPUT

Reads quality analysis and improvement

VIRUS DETECTION *(TELEVIR module)*

Hypothesis-free Virus Detection

Validation

Investigatory Mapping

TELEVIR | Basics

Bioinformatics Diagnostics

VIRUS DETECTION *(TELEVIR module)*

- **Hypothesis-free Virus Detection**
- **Validation**
- **Investigatory Mapping**

I don’t know what I have

Any guess needs to be looked into

Suspected, Contaminants and False Positives

TELEVIR | Basics

Bioinformatics Diagnostics

VIRUS DETECTION *(TELEVIR module)*

Hypothesis-free Virus Detection

Validation

Investigatory Mapping

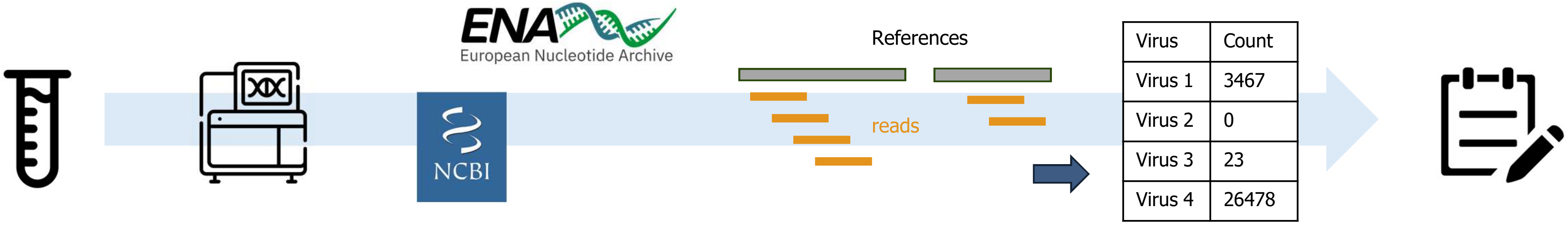
I don’t know what I have

Any guess needs to be looked into

Suspected, Contaminants and False Positives

TELEVIR | Basics

Bioinformatics Metagenomics – Broad View



Virus	Count
Virus 1	3467
Virus 2	0
Virus 3	23
Virus 4	26478

Sequencing

Mappings to Genomes

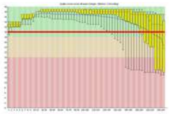
Exploratory,
Statistical analysis

“Best Hits”

Clinical metagenomic NGS (mNGS)
Bioinformatics workflow

FASTQ

input



Quality Control & Entropy Filters
Read trimming, adapter removal, contamination check

Host / Background Removal
- *Depletion of host (or abundant background) or pathogen enrichment*

de novo Assembly
- *To increase classification specificity*

Taxonomic Classification / Functional Annotation
- Pathogen identification, microbial community profiling, detection of AMR genes..

Confirmatory Mapping, Interpretation & Reporting
- Use of multiple bioinformatics metrics, read confirmation, and in-depth false-positive checking to ensure accuracy of the result.

TELEVIR | Basics

Bioinformatics Metagenomics – Key Steps

- Quality Control?
- Assembly?
- Nucleotide or Protein?
- Different Databases
- Different Algorithms
- End-User Reports?
- False Positives?

de Vries, J.J., Brown, J.R., Fischer, N., Sidorov, I.A., Morfopoulou, S., Huang, J., Oude Munnink, B.B., Sayiner, A., Bulgurcu, A., Rodriguez, C. and Gricourt, G., 2021. Benchmark of thirteen bioinformatic pipelines for metagenomic virus diagnostics using datasets from clinical samples. *Journal of Clinical Virology*, *141*, p.104908. https://doi.org/10.1016/j.jcv.2021.104908

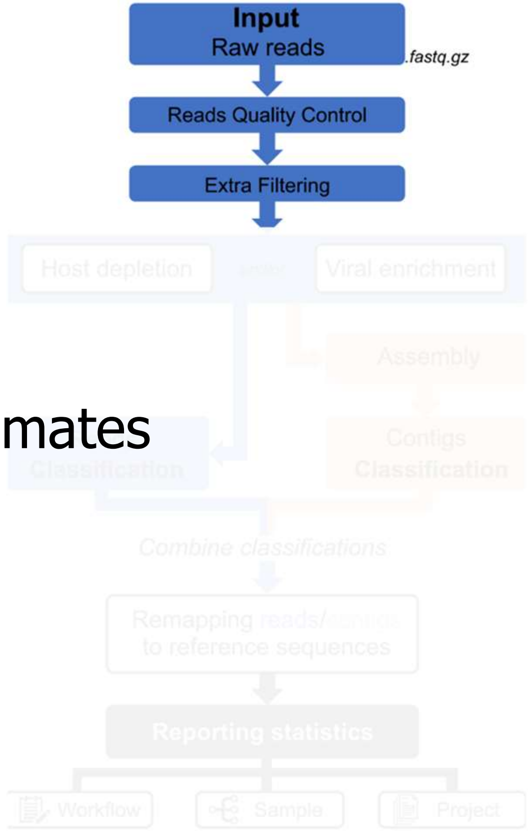
TELEVIR | Basics

Pre-processing | Quality Control

Remove Adaptors, Low Quality Reads, Trimming.

Remove low complexity reads – dust masking, entropy estimates

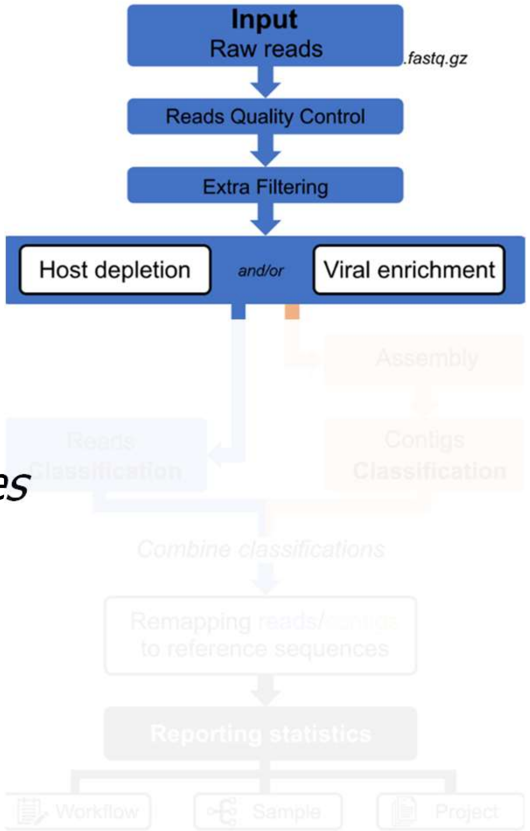
- Repetitive regions, e.g. GC-rich sequences.
- **Reduce risk of False Positives**



Input
Raw reads
.fastq.gz
Reads Quality Control
Extra Filtering
Host depletion
Viral enrichment
Assembly
Classification
Contigs Classification
Combine classifications
Remapping reads/contigs to reference sequences
Reporting statistics
Workflow
Sample
Project

TELEVIR | Basics

Pre-processing | Host Filtering



Host Depletion	Single, known host *15 hosts available (TELEVIR)*	e.g. *Blood samples*
Viral Enrichment	Unknown hosts	e.g. *wastewater samples*

- Accelerates downstream processing
- facilitates assembly

⚠ **May filter out pathogenic reads.**
⚠ **Host analysis not yet available, under evaluation.**

Instituto Nacional de Saúde Doutor Ricardo Jorge

ecdc EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

TELEVIR | Basics

Classification | Metagenomics

|Reads classification

Quantity over quality
Multiple Software

|Contig classification

Precision

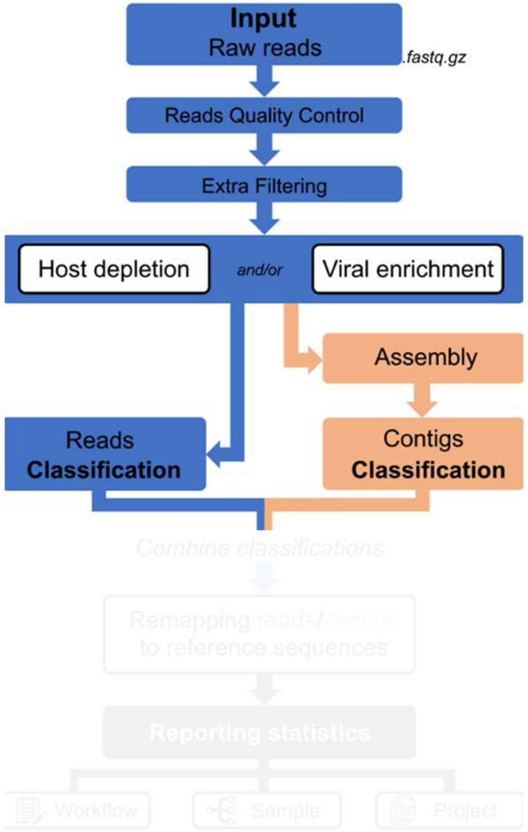
Cross Validation

Different algorithms

Multiple Databases

Divergent Pathogens

Protein based search



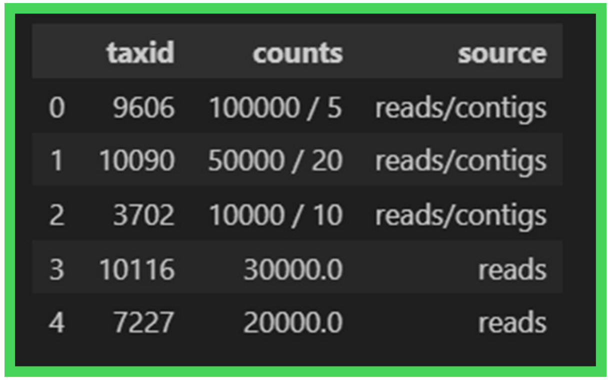
TELEVIR | Basics

Classification | Combining classifications

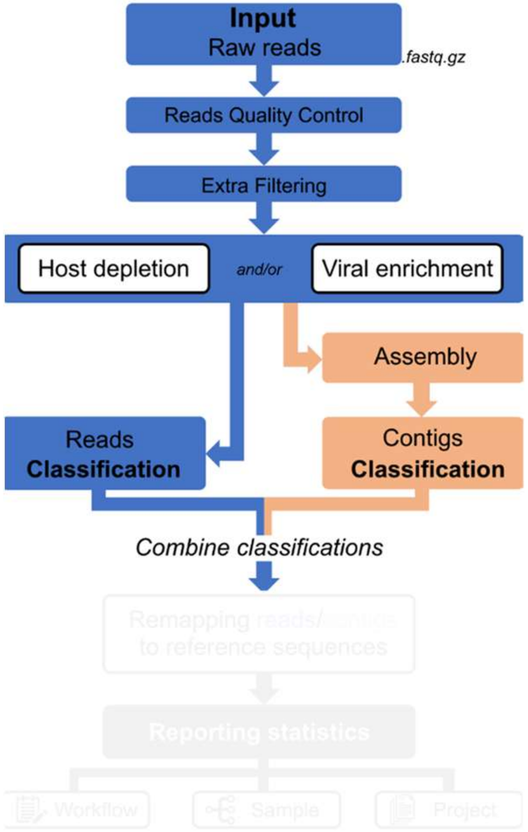
> Combined sort using contig and read **“hits” per taxid**

> Prioritize hits obtained using **contigs**

20 % precision increase



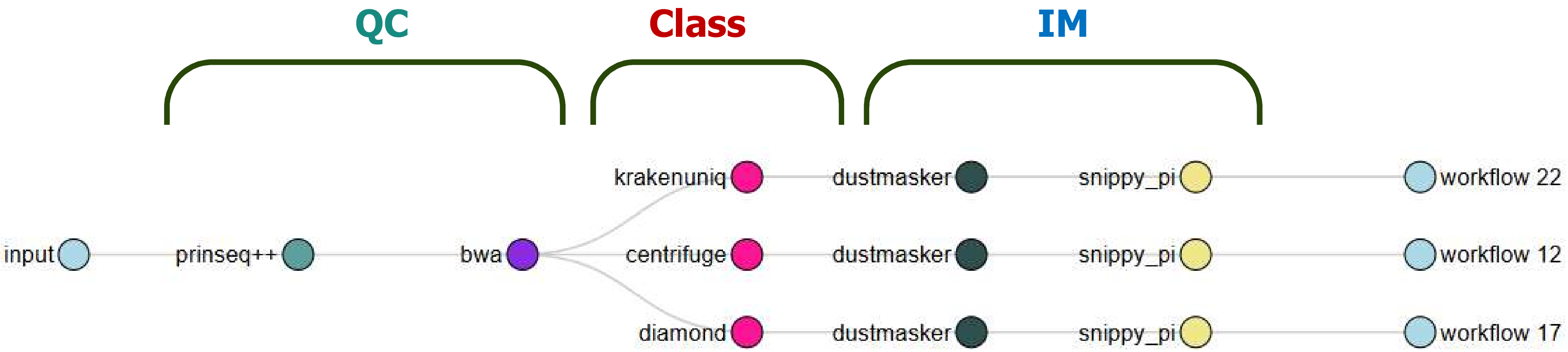
	taxid	counts	source
0	9606	100000 / 5	reads/contigs
1	10090	50000 / 20	reads/contigs
2	3702	10000 / 10	reads/contigs
3	10116	30000.0	reads
4	7227	20000.0	reads



Dourado Santos J et al., (2024) Genome Medicine; 16:61
https://doi.org/10.1186/s13073-024-01334-3

TELEVIR | Basics

Settings | Cross Validation



|Modular Deployment

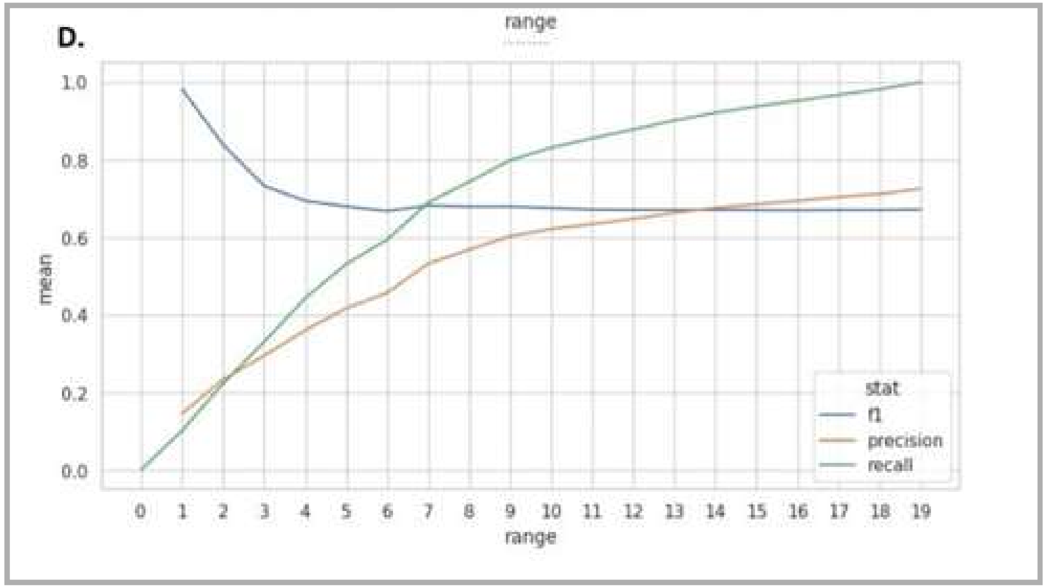
All possible combinations are deployed

Ensemble Ranking

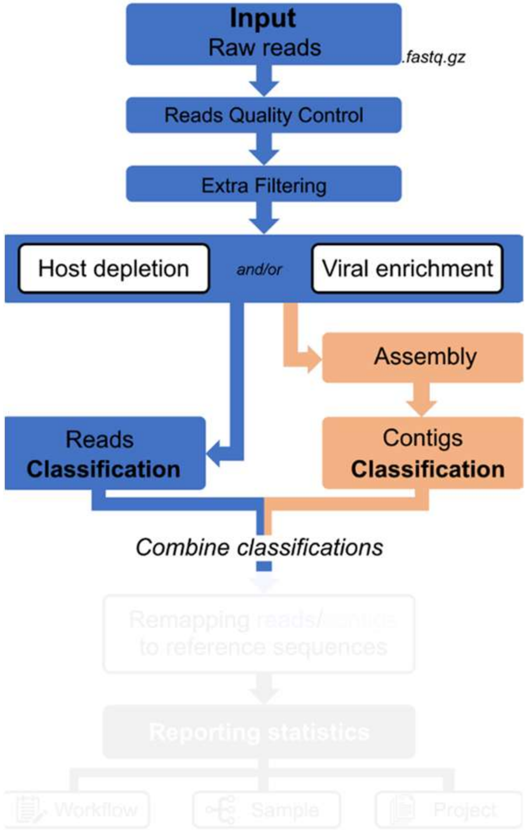
TELEVIR | Basics

Remapping | Confirmatory Mapping

False
Positives
25 %



Classification Ranking


Input
Raw reads
.fastq.gz

Reads Quality Control

Extra Filtering

Host depletion *and/or* Viral enrichment

Assembly

Reads **Classification**

Contigs **Classification**

Combine classifications

Remapping reads/contigs to reference sequences

Reporting statistics

Workflow

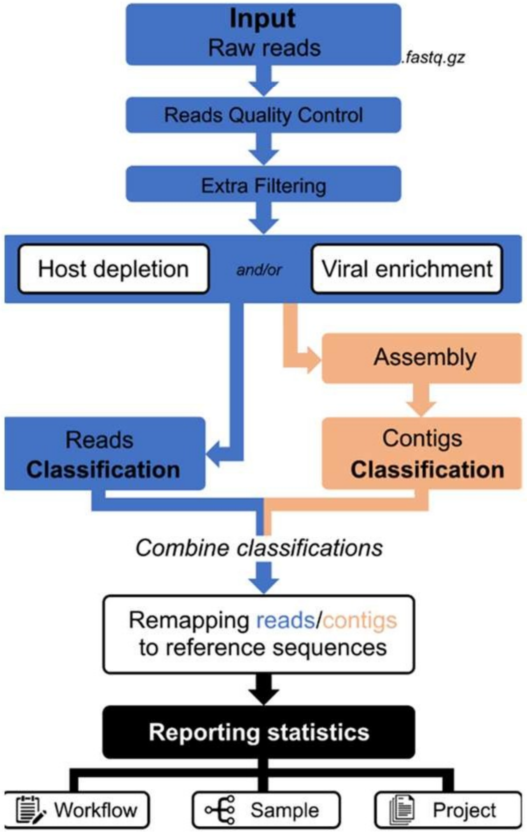
Sample

Project

Dourado Santos J et al., (2024) Genome Medicine; 16:61
https://doi.org/10.1186/s13073-024-01334-3

TELEVIR | Basics

Remapping | Confirmatory Mapping

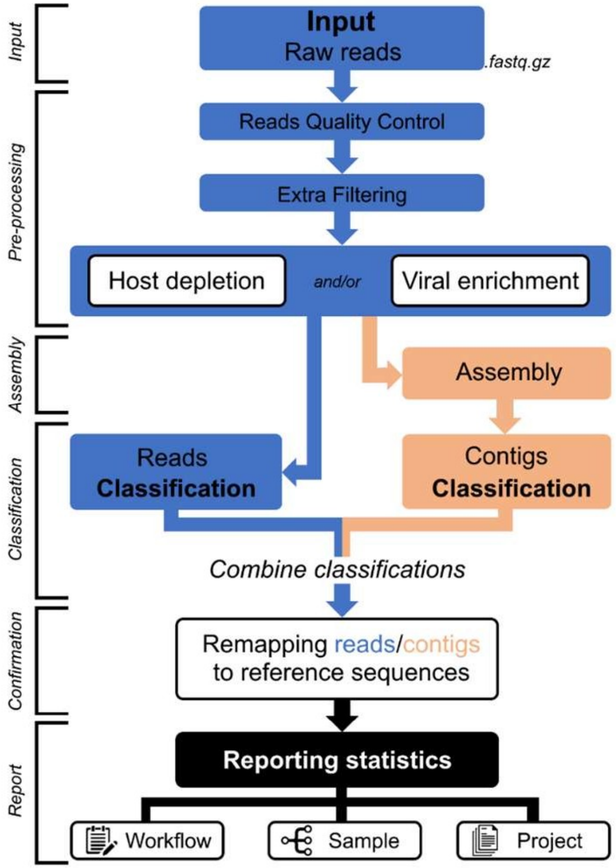

Input
Raw reads
.fastq.gz
Reads Quality Control
Extra Filtering
Host depletion
and/or
Viral enrichment
Assembly
Reads
Classification
Contigs
Classification
Combine classifications
Remapping reads/contigs
to reference sequences
Reporting statistics
Workflow
Sample
Project

Automatically deployed in TELEVIR

TELEVIR | Basics

Bioinformatics Diagnosis

- ✓ Designed under a “**no single best approach**” principle.
  - ✓ Multiple **workflow combinations** (with/without enrichment, host depletion, etc).
  - ✓ Incorporates **multiple classifiers** (Kraken2, KrakenUniq, Centrifuge, BLAST, Diamond, FastViromeExplorer, Kaiju etc.).
  - ✓ Integrates **several databases** (Virosaurus, RefSeq etc).
- ✓ Balances sensitivity and precision via **confirmatory mapping**.
- ✓ Delivers results in **interactive, evidence-oriented, user-friendly dashboards**.



Dourado Santos J et al., (2024) Genome Medicine; 16:61
https://doi.org/10.1186/s13073-024-01334-3

Instituto Nacional de Saúde
Doutor Ricardo Jorge

ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

TELEVIR | Basics

INPUTS

- ✓ **Sample metadata table**

+

- ✓ **NGS data ***

**Compatible with distinct sequencing technologies:*
Illumina, Ion Torrent and Oxford Nanopore

- ✓ **References fasta**

+

- ✓ **Reference Metadata**

**Compatible with distinct sequencing technologies:*
Illumina, Ion Torrent and Oxford Nanopore

_TELEVIR

OUTPUTS

VIRUS DETECTION *(TELEVIR module)*

- ✓ Read quality analysis and improvement

- ✓ **State of the art software**
- ✓ **Modular Pipeline**

- ✓ **Multiple Classification methods**
- ✓ **Multiple Viral databases**

- ✓ **Summary statistics and intuitive and interactive end-user reports**

- ✓ **Confirmation Mapping**
- ✓ **Investigatory Mapping**