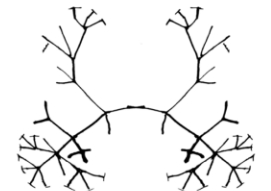


Vincent Rennie, PhD
Annelies Van Rie, MD,
PhD

TORCH
Tuberculosis Omics Research



Why another pipeline?

Mtb pipelines were developed for research

What we need now is:

A clinical bioinformatics pipeline

For personalized DR-TB treatment

And targeted public health interventions

Challenges posed by clinical *Mtb* samples

Low amount of *Mtb* in sputum samples and primary MGIT cultures

- High levels of contaminating sequences (human, NTMs, other bacteria)
- Low amounts of (hard-to-extract) *Mtb* DNA

Mixed infections and hetero-resistance

- Need to disentangle mixed infections
- Need to detect major and minor variants

***Mtb* genome contains conserved regions similar to contaminating sequences**

- Can lead to false positive variants in candidate resistance genes (e.g rrl, rrs)

***Mtb* genome contains repetitive regions**

- Elimination of complex regions by most analysis pipelines (loss of data)

Maximum Accessible Genome for *Mtb* Analysis

Aim One:

Call variants in genomes with low coverage

Aim Two:

Call both major and minor variants

Aim Three:

Call variants in the presence of contaminating sequences

Aim Four:

Retain data from complex regions in *Mtb* genome

Core Software Package



Developed for human genomics

Key features:

1

Joint variant calling
of major variants

2

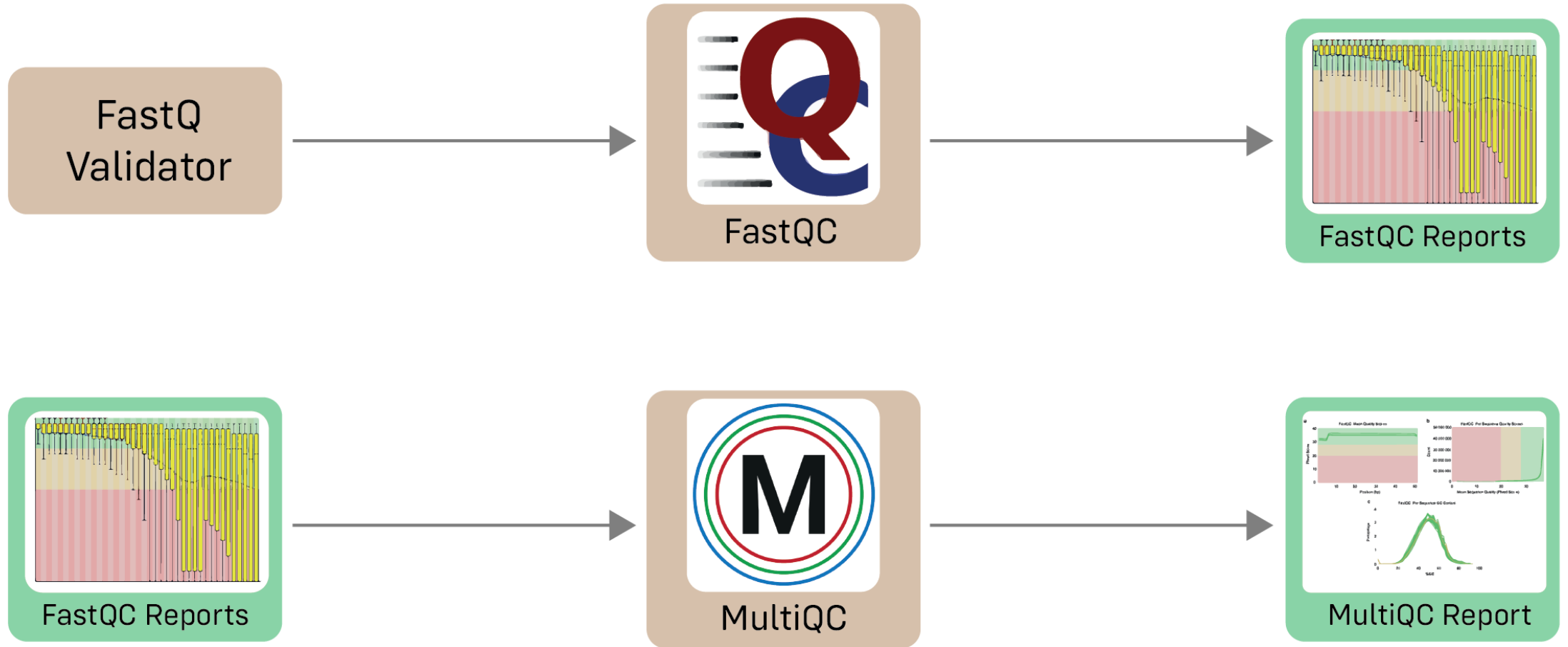
Filtering of artefacts
and contamination

Automation and standardization

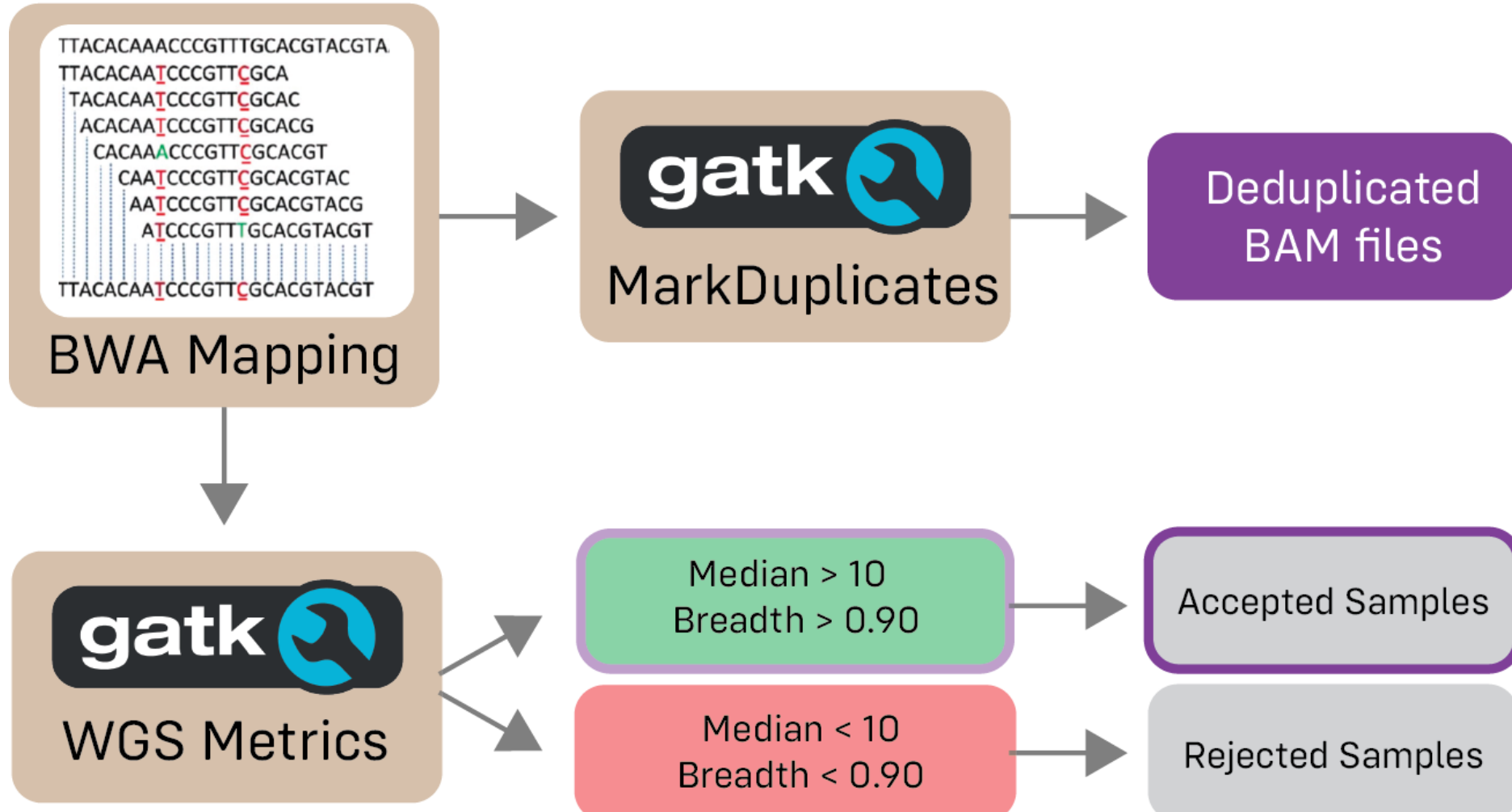
nextflow

Workflow

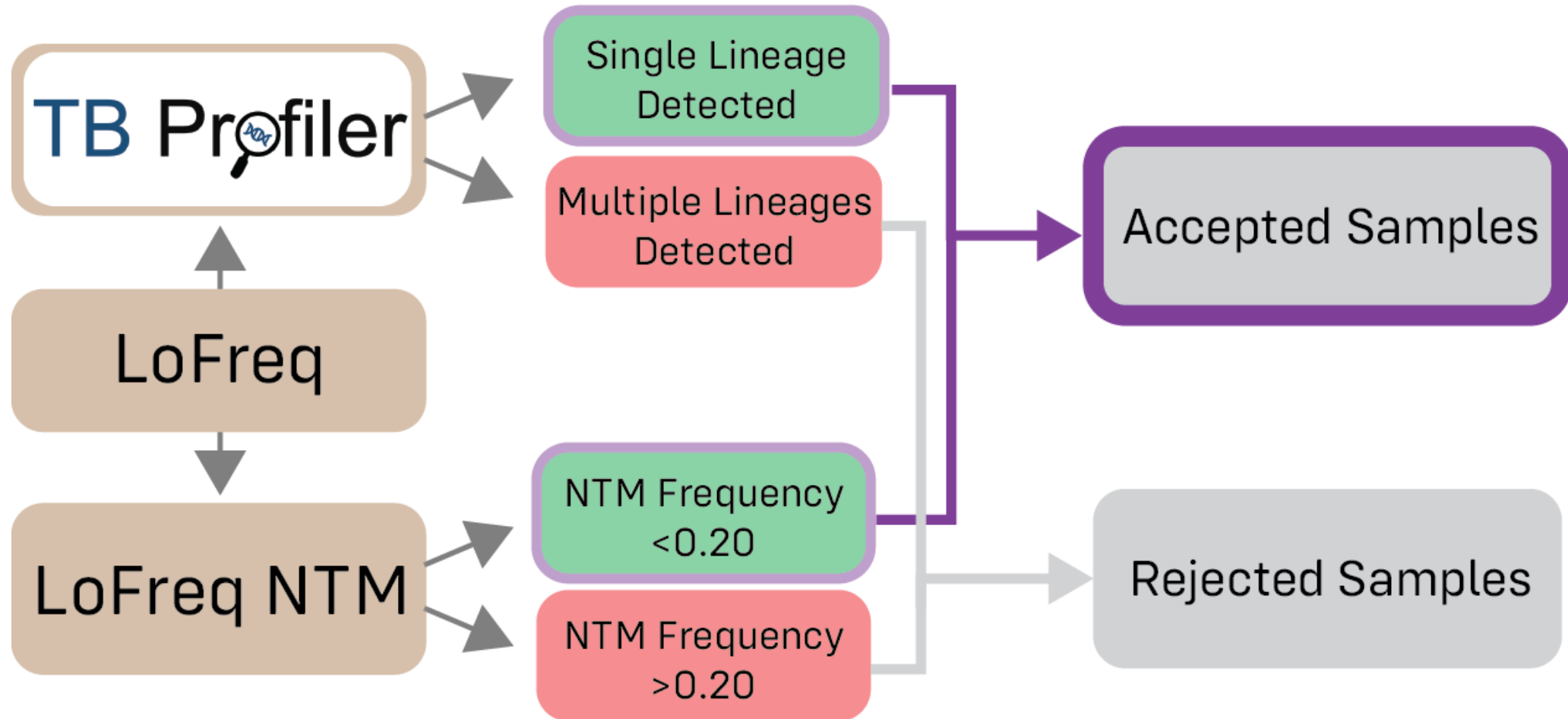
Quality Control



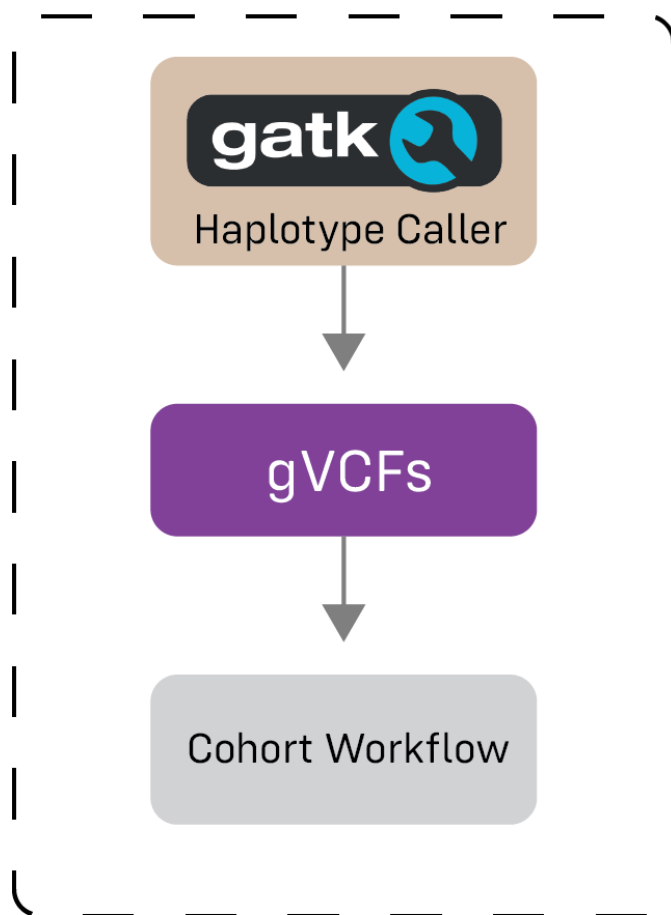
Mapping to reference genome



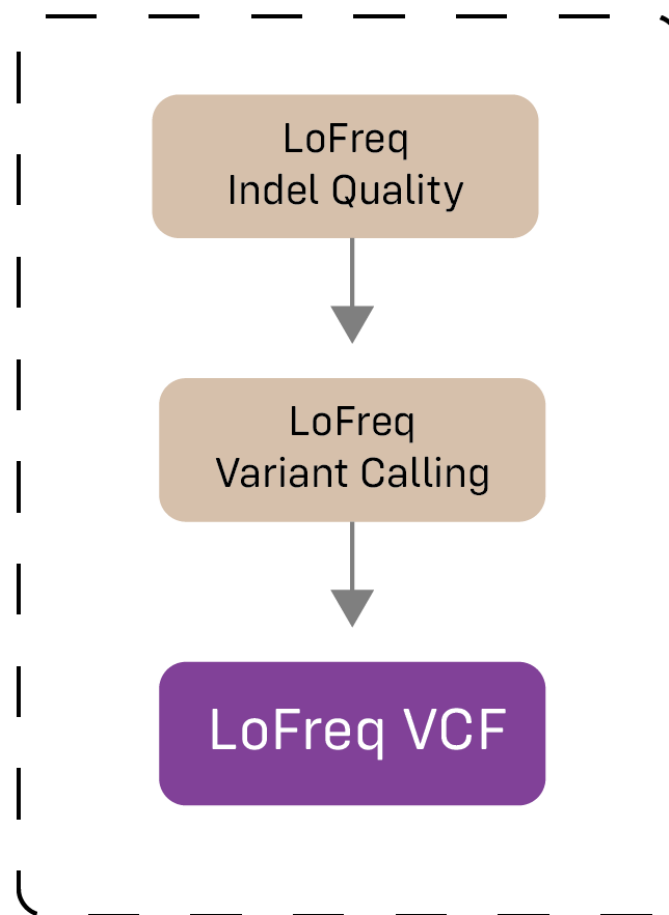
Detection of Multiple Infections



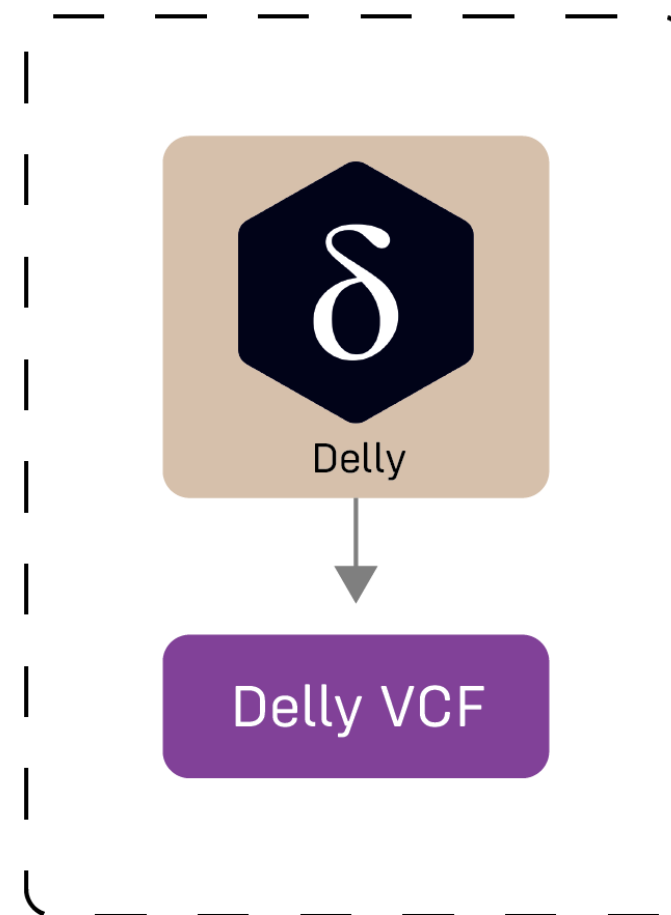
Variant Calling



Major Variants

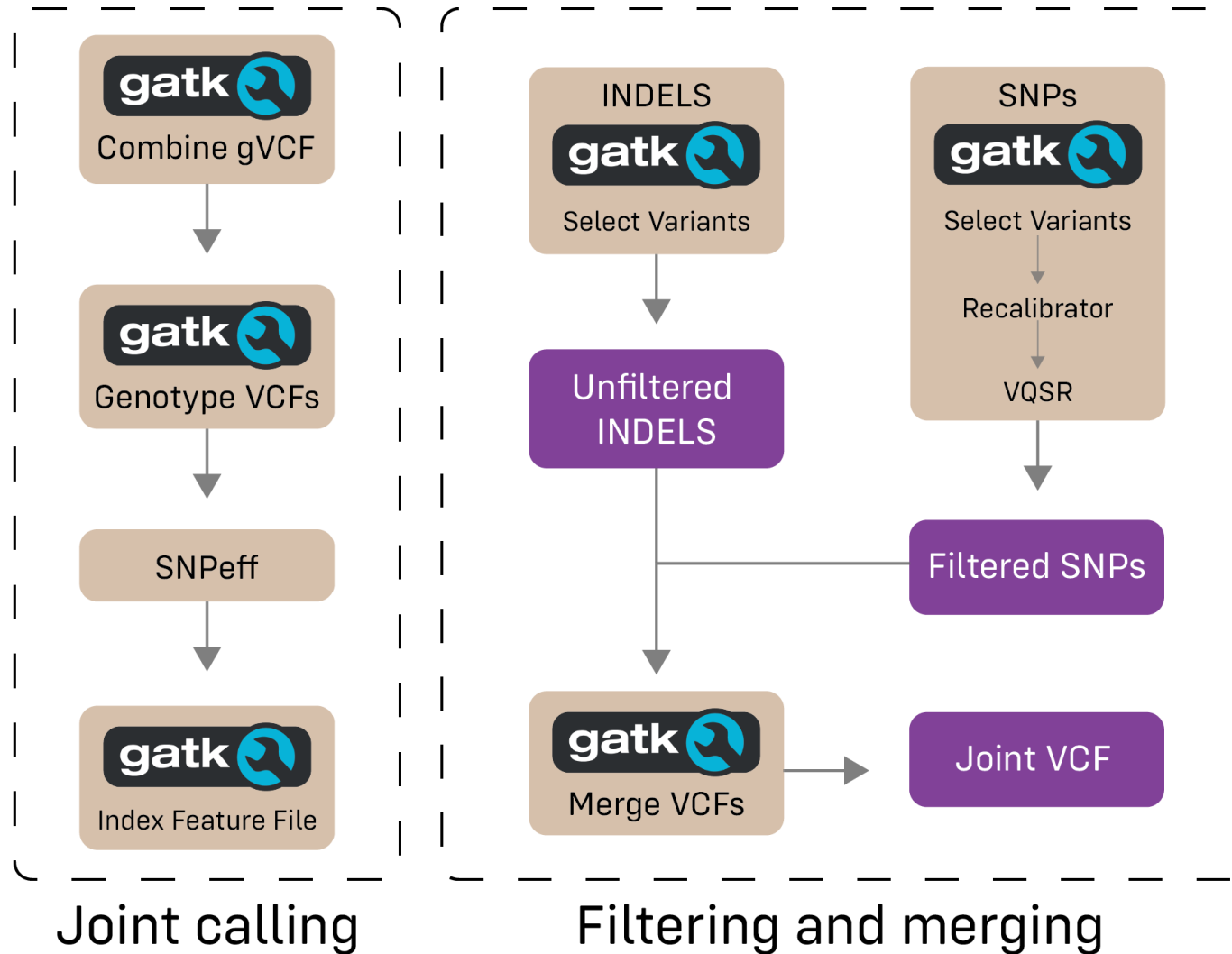


Minor Variants

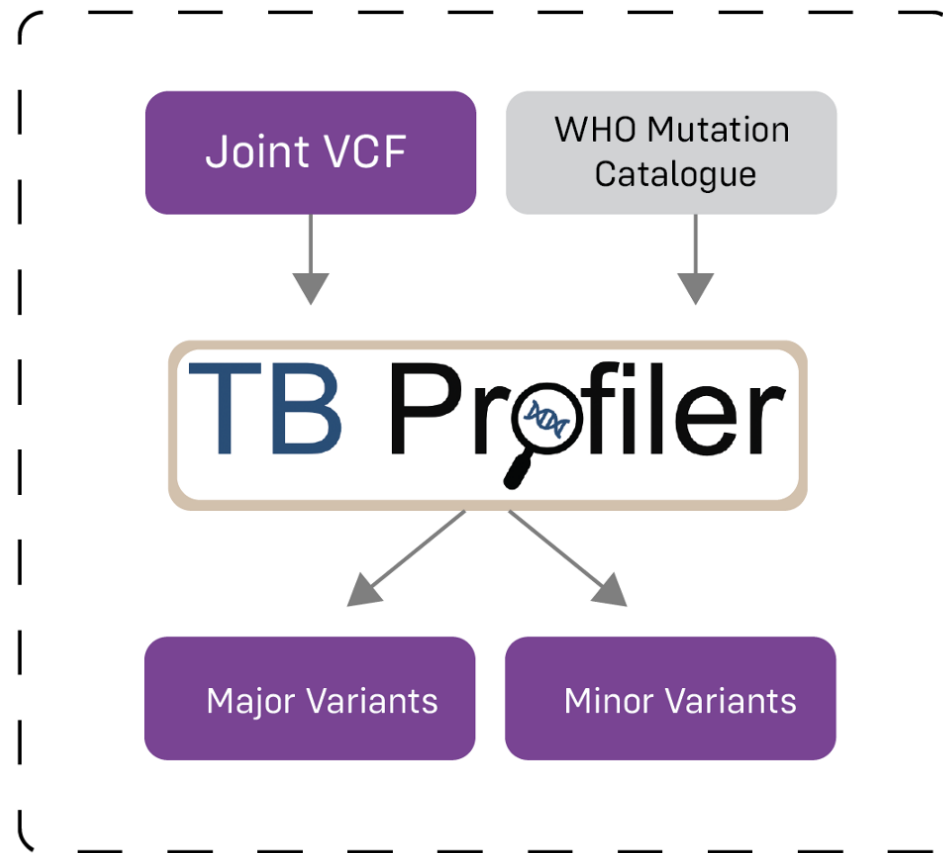


Structural Variants

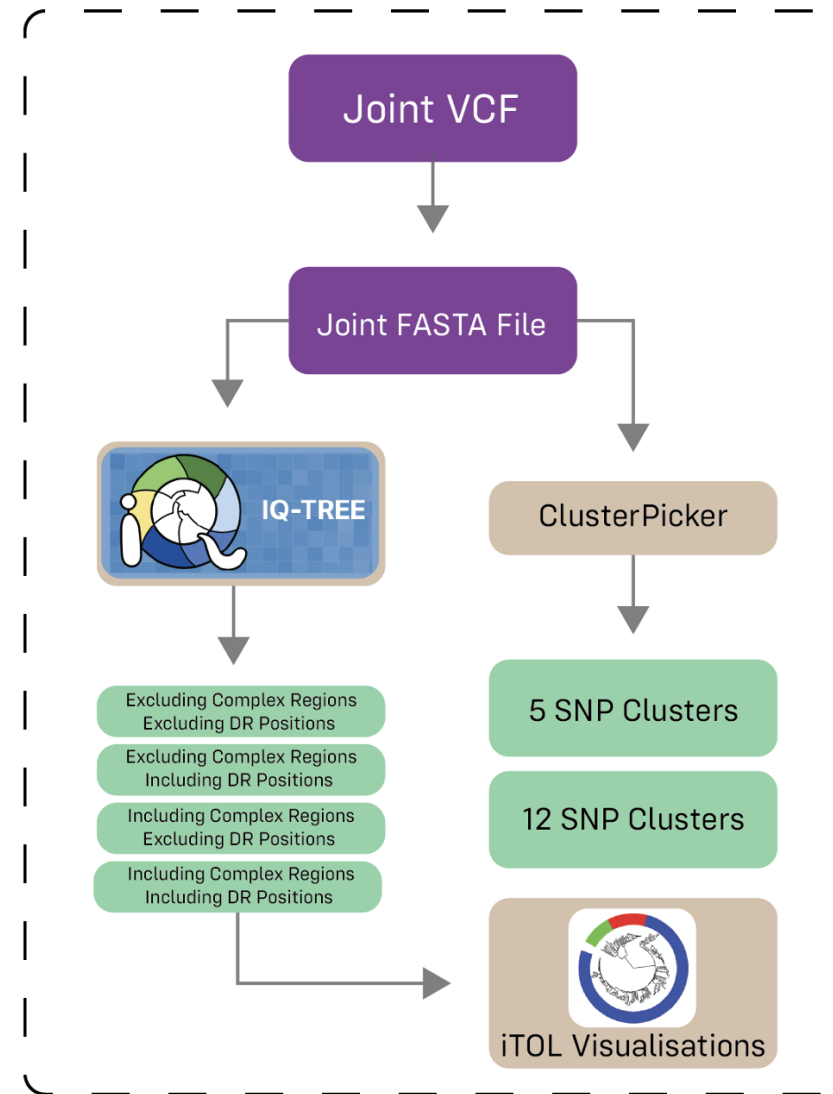
Cohort calling and filtering



Primary outputs



DR Variant Characterisation



Phylogenetic Analysis

MAGMA Performance

How does MAGMA compare with other Mtb pipelines?

Heupink TH, Verboven L, Warren RM, Van Rie A. Comprehensive and accurate genetic variant identification from contaminated and low-coverage Mycobacterium tuberculosis whole genome sequencing data. Microb Genom. 2021

Comparison with two MTB pipelines

UVP pipeline



MTBseq pipeline

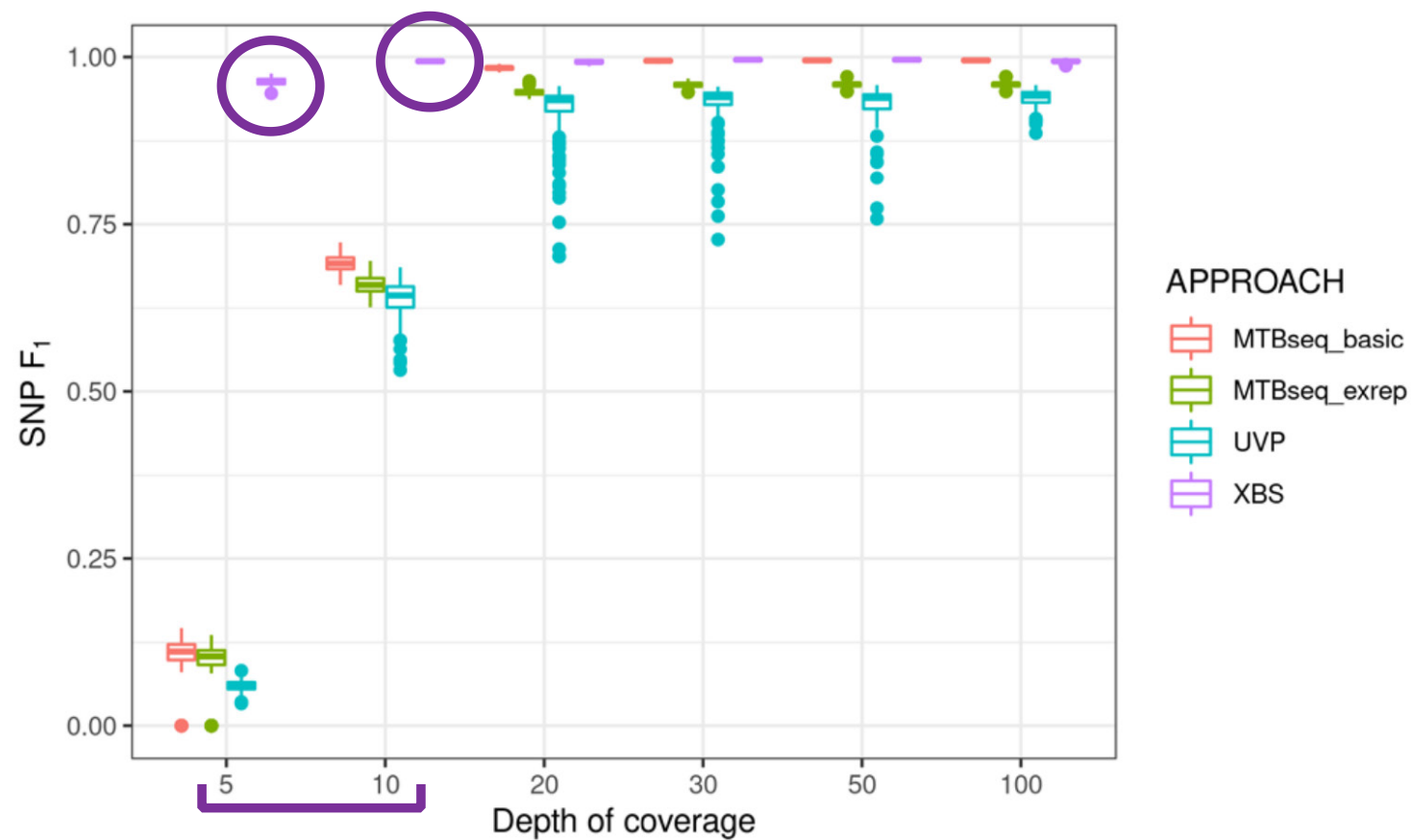
ngs-fzb/
MTBseq_source



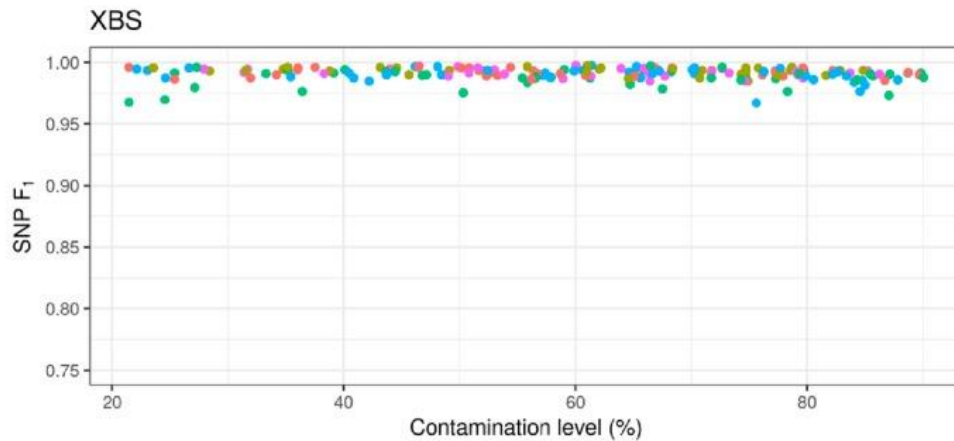
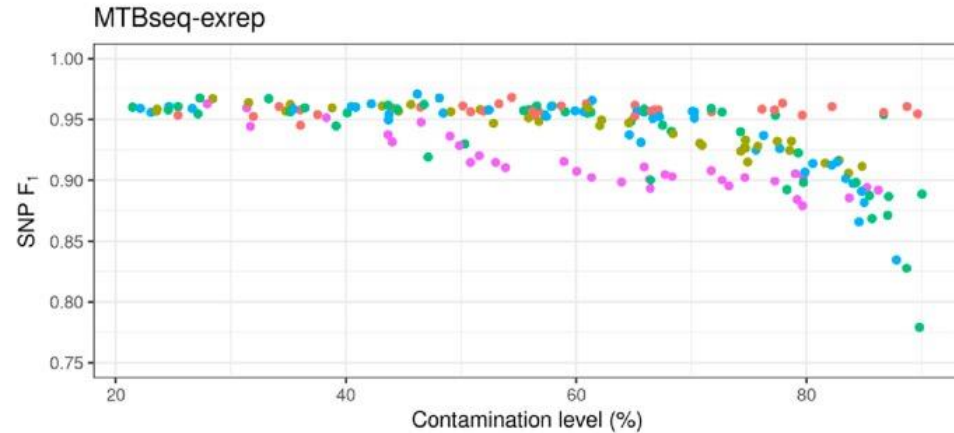
UVP: Ezewudo, M., Borens, A., Chiner-Oms, Á. et al. Integrating standardized whole genome sequence analysis with a global Mycobacterium tuberculosis antibiotic resistance knowledgebase. Sci Rep 2018

MTBSeq: Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, Niemann S. MTBseq: a comprehensive pipeline for whole genome sequence analysis of Mycobacterium tuberculosis complex isolates. PeerJ. 2018

High accuracy at low coverage



Performance in presence of contamination



Contaminant: ● H.sap ● H.sap/P.aer./S.Epi ● NTM mixture ● P.aer. ● S.epi

UVP

Does not analyse samples containing >20% contamination

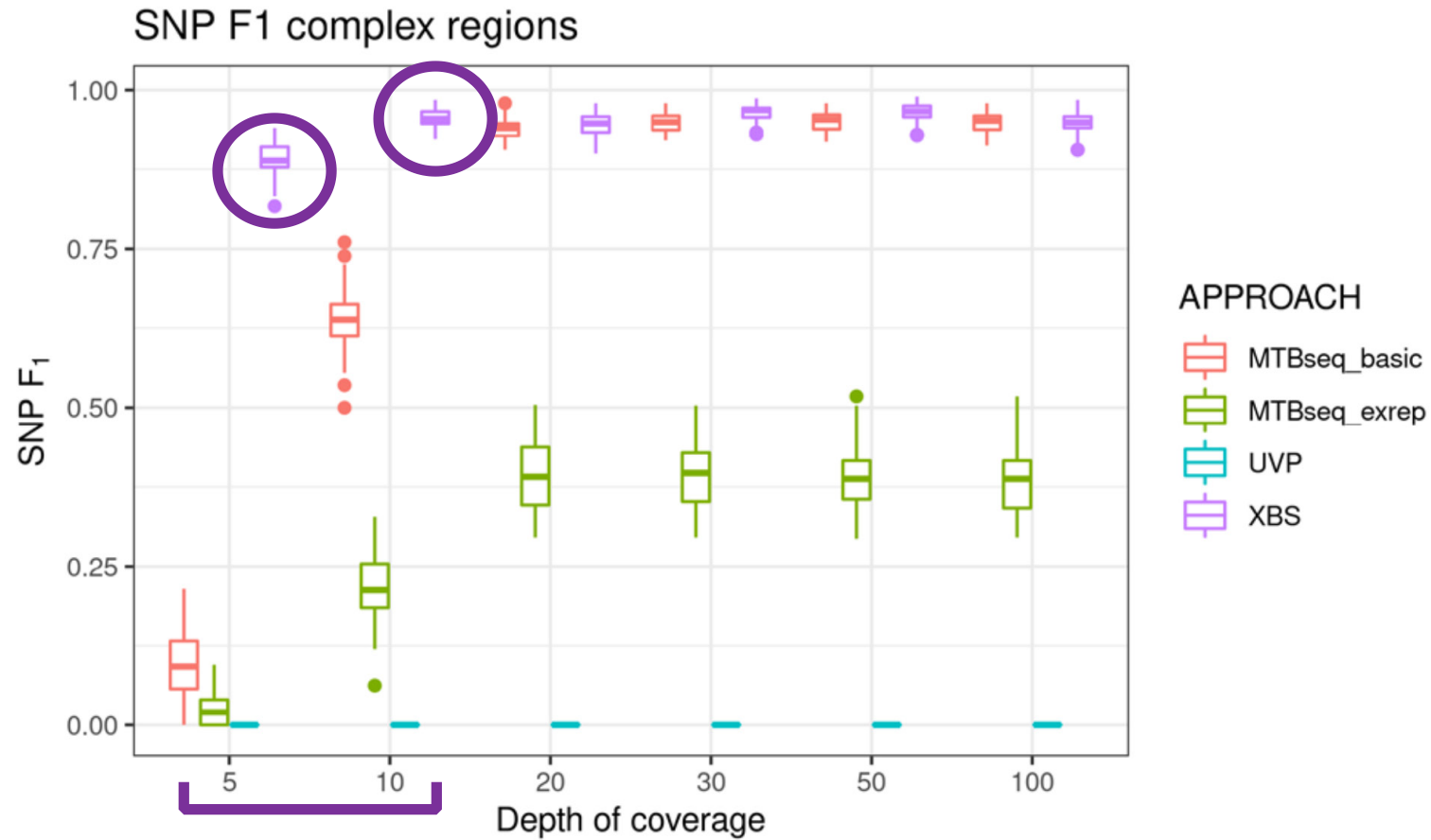
MTBSeq

Does not perform well at high levels of contamination

MAGMA

Performance never drops below a 0.95 F_1 score

High accuracy at complex regions



Conclusion



Works with low *Mtb* genome coverage



Retains data from complex genomic regions in *Mtb*



Still calls variants in the presence of high levels of contamination