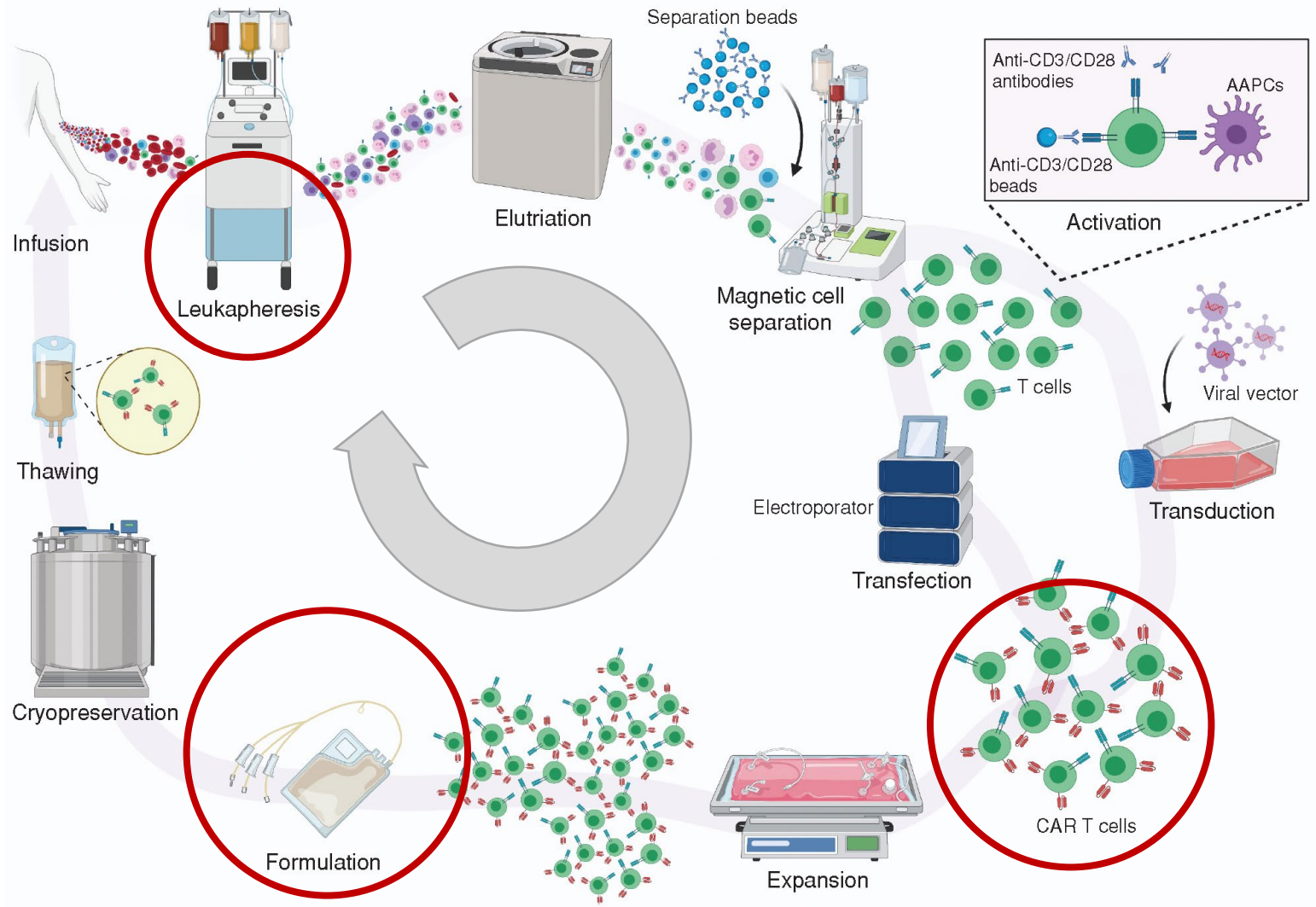
Application of Nanopore Sequencing to Sterility Testing for Cell Therapy Products

James STRUTT

Agenda

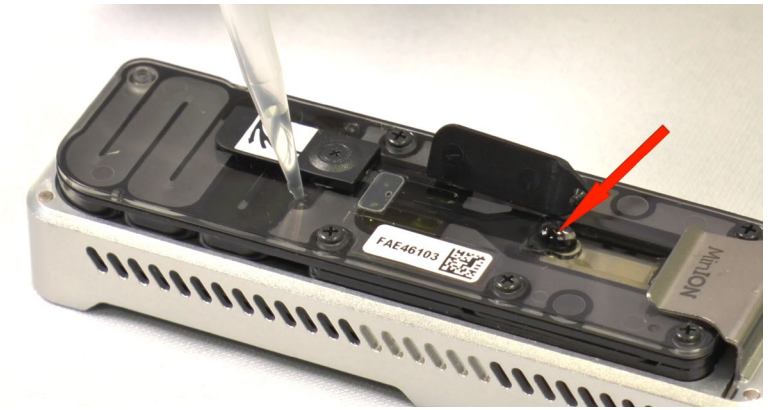
- Overview of sterility testing and key challenges in adopting sequencing technologies
- Application of Nanopore Sequencing for untargeted microbial detection
- Preliminary insights into the kitome and its baseline characterization

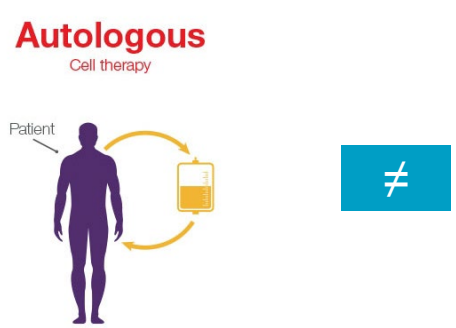
Why Sterility Testing?



Abou-el-Enein et al., 2021,
Blood Cancer Discovery

Why Long Read Sequencing in Particular?

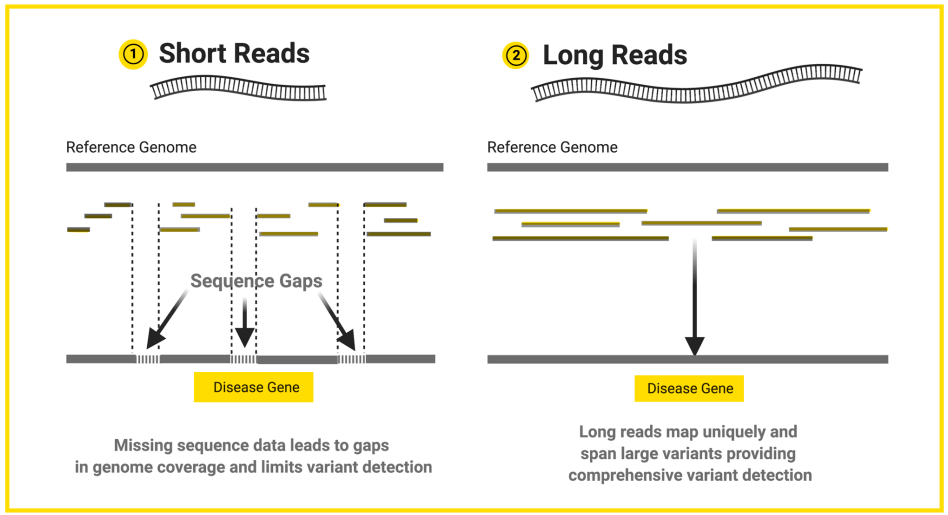




≠

Compendial Testing

- Time to detection is real time
- Agnostic detection of contaminants
- Limit of detection is compliant with existing FDA requirements (Strutt *et al.,*)

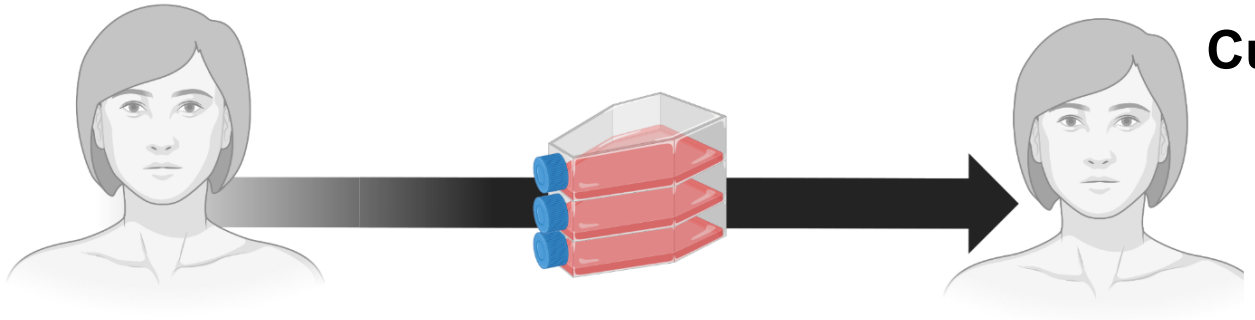


Benefits of long reads as compared to short reads:

- Improved accuracy
- Simpler assembly
- Better detection of structural variants (e.g. conserved rRNA)

Sterility Release Testing

Cell Therapy Sterility - *The problem*

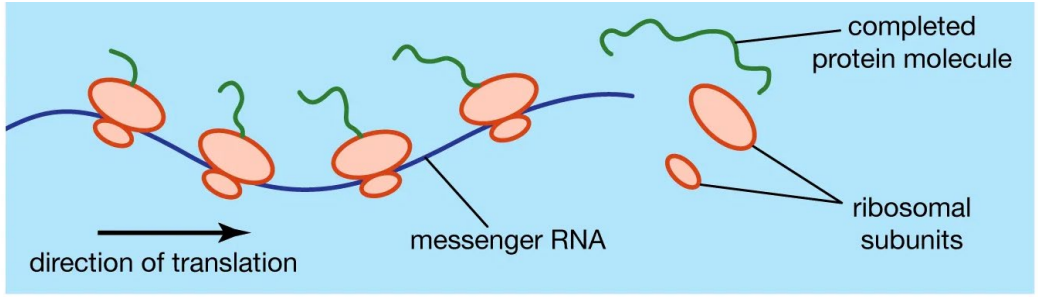


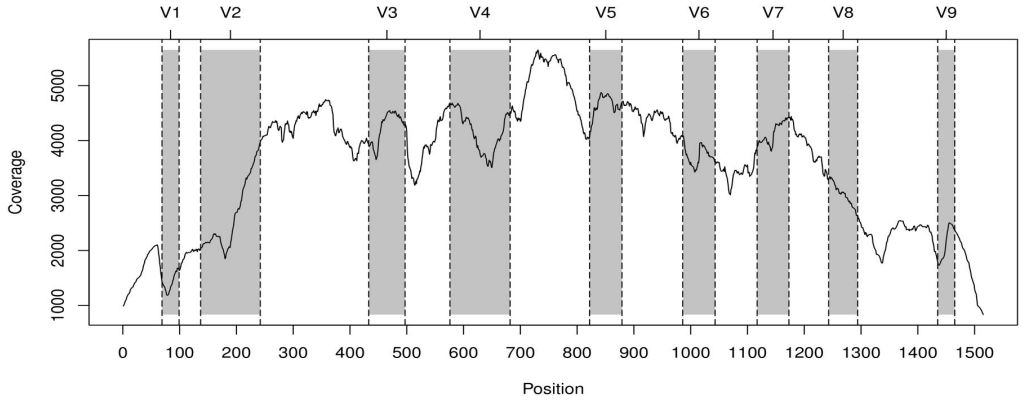
Current Compendial Sterility Tests

- Limit of detection ≤10 CFU / mL
- Time to detection > 7 days
- Sample volume ~10 mL
- No Identification of Species

Why Target Ribosomal RNAs?

V = variable regions

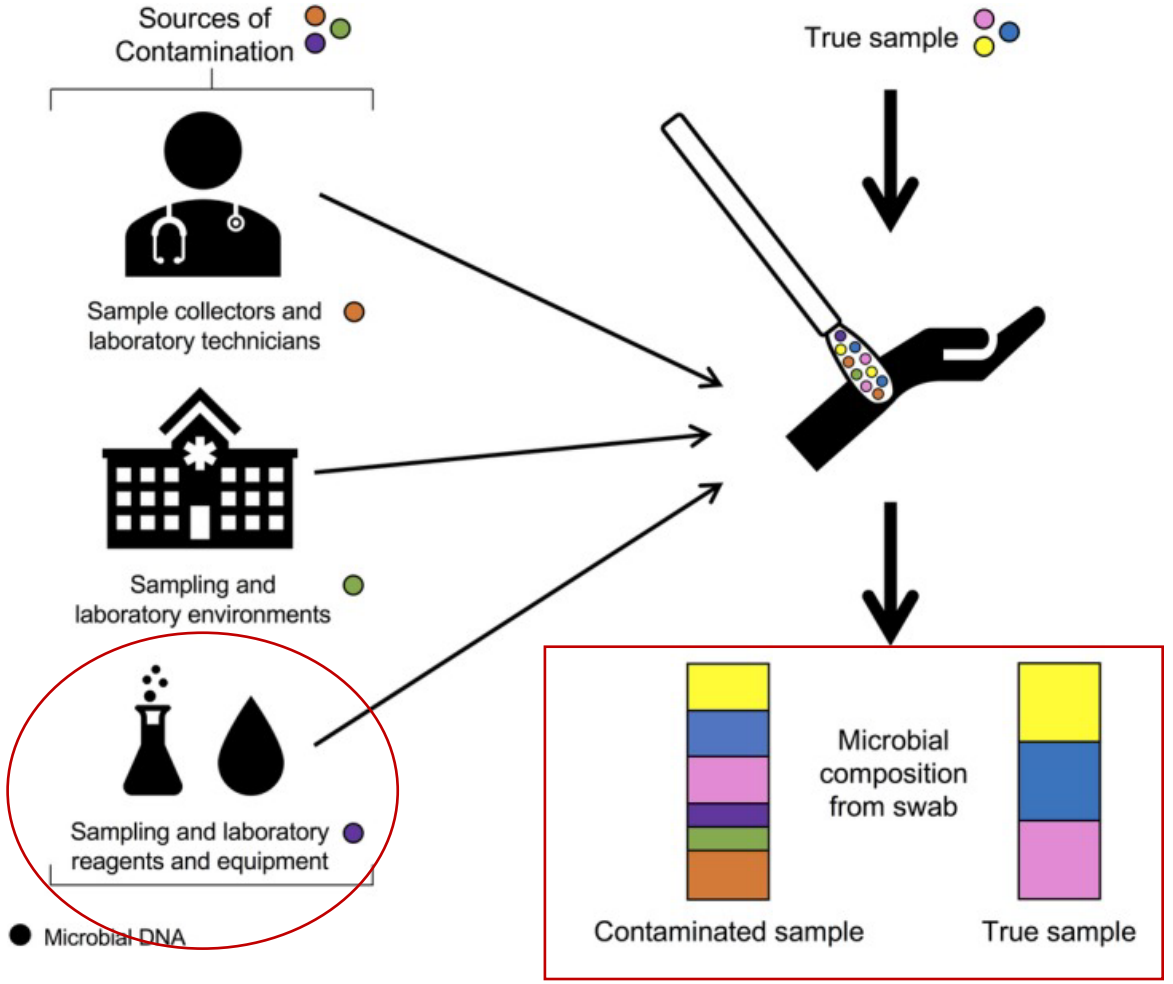




- What is ribosomal RNA?
- Provide culture-free metagenome analysis
- Generate full length amplicons, which permits species level identification
- Variable regions contain species-specific DNA sequence
- Highly conserved regions of prokaryotic and eukaryotic genomes provide universality and make an untargeted approach possible

Contaminants in Your Low-Biomass Data

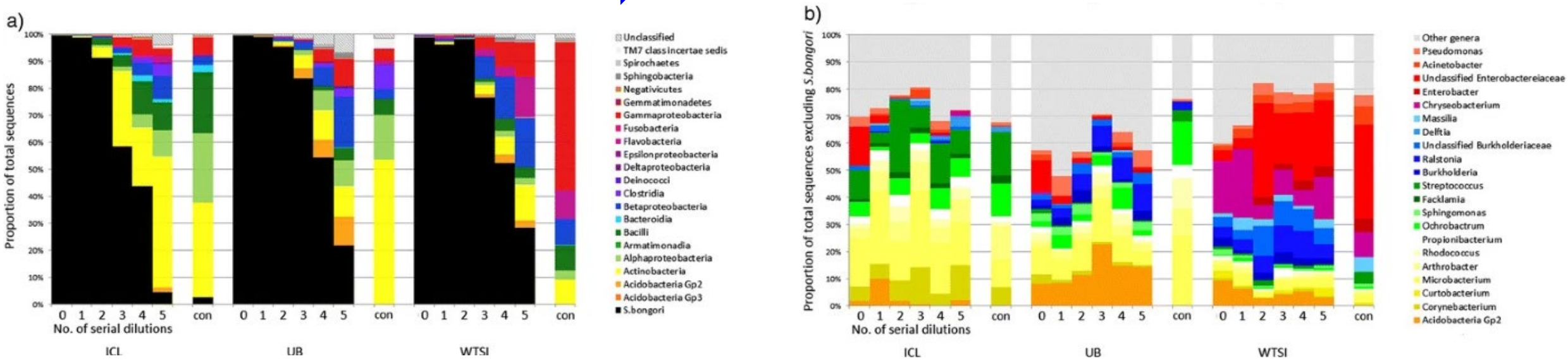
- The kitome refers to background DNA contamination from reagents and extraction kits
- Misclassification events occur when sequencing pipelines report species that aren't actually present in the sample.
- **Kitome reads are real DNA fragments** introduced during sample processing, not by the original sample.
- As a result, **sterile or negative samples can appear contaminated**, even when no microbes are present.
- The kitome signal can overwhelm or mask true contaminants, especially in low-biomass samples.



Selway *et al.*, 2020

The Kitome: Hidden Contaminants in Your Low-Biomass Data

Remove *S. bongori* (black)

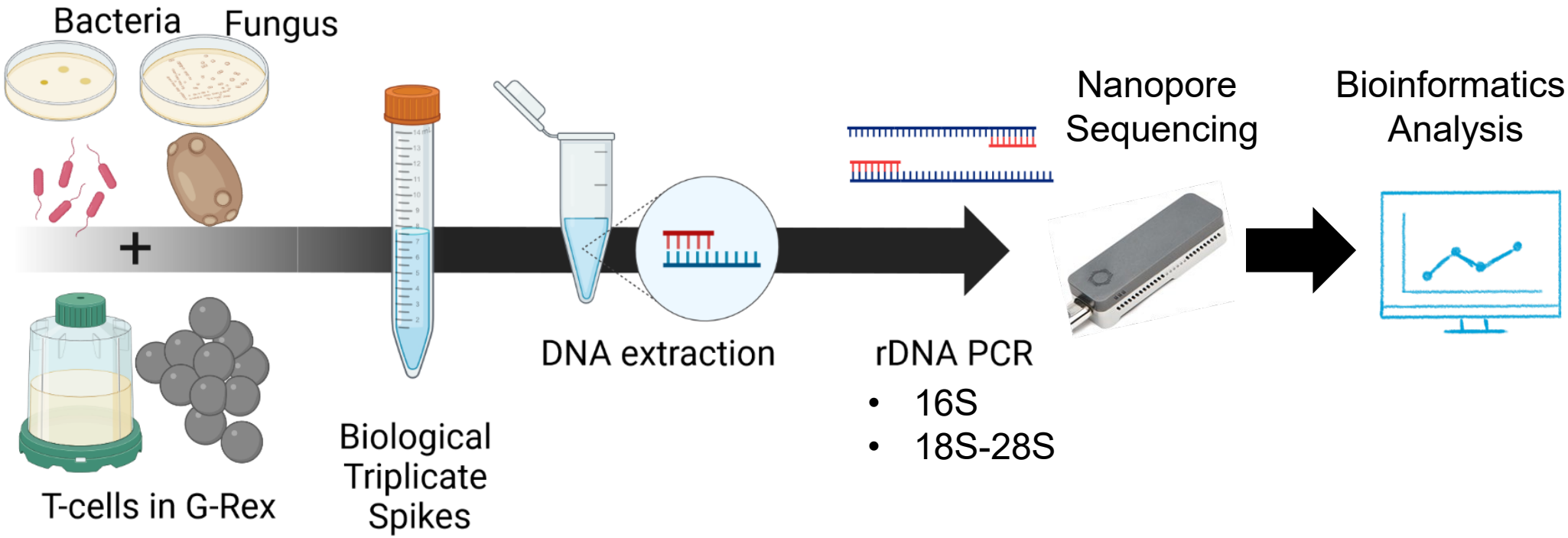


- *S. Bongori* DNA was extracted with FastDNA SPIN Kit for Soil (kit FP).
- Samples were serially diluted
- As sample biomass decreases, contaminant DNA becomes dominant, and contaminant profiles reflect the lab and kit used rather than the true sample.

ICL (Imperial College London), UB (University of Birmingham) and WTSI (Wellcome Trust Sanger Institute); con = NTC control

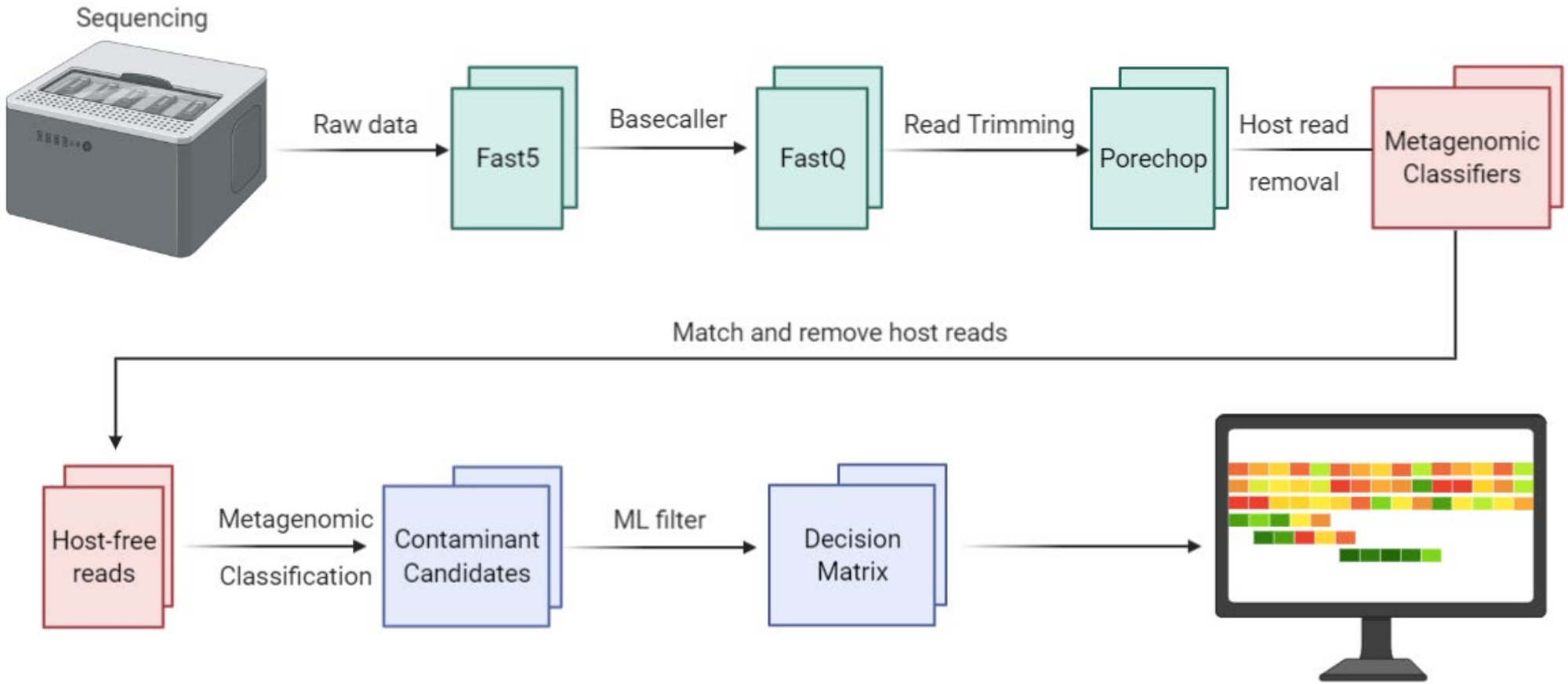
Salter *et al.*, 2014

The Nanopore Sequencing Workflow



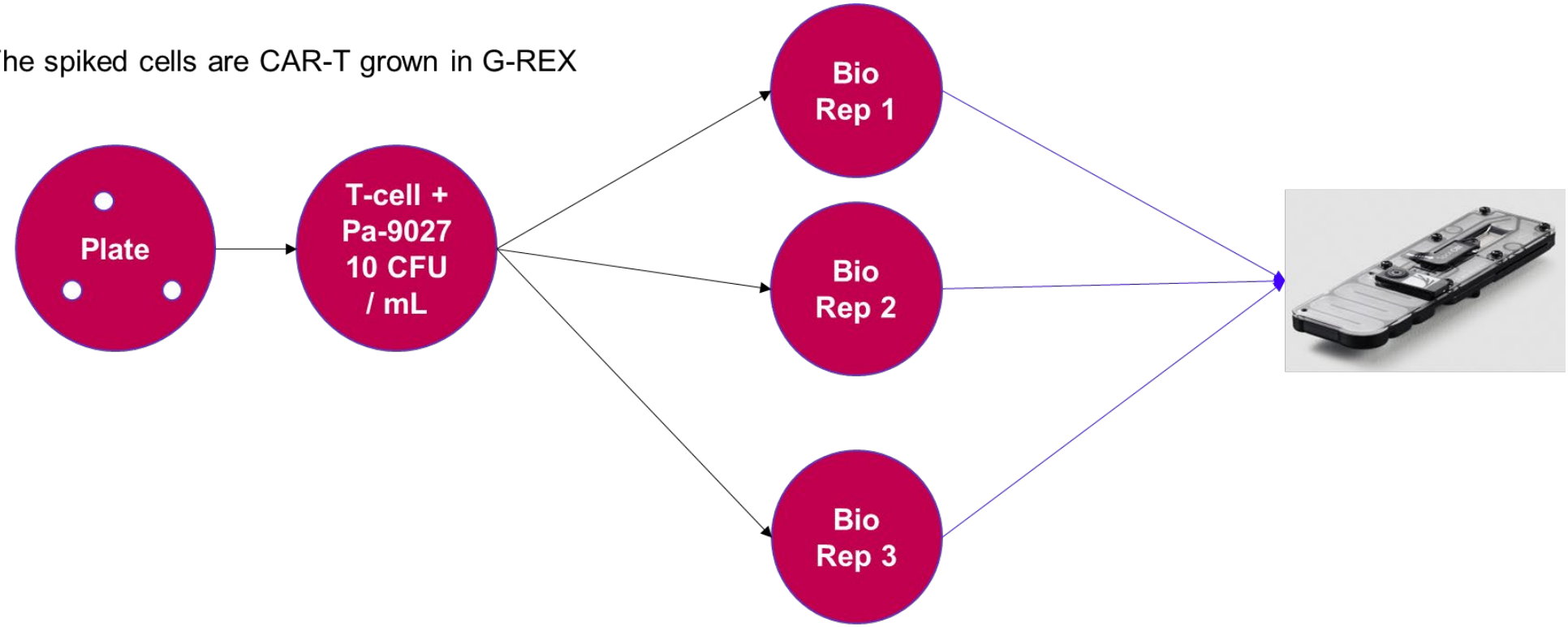
15 m → ~ 3 h → < 12 h → 1 h / sample → **~ 16 h**

The Bioinformatics Workflow



Experimental Setup: Understanding Limit of Detection and Sources of Reliability and Variance

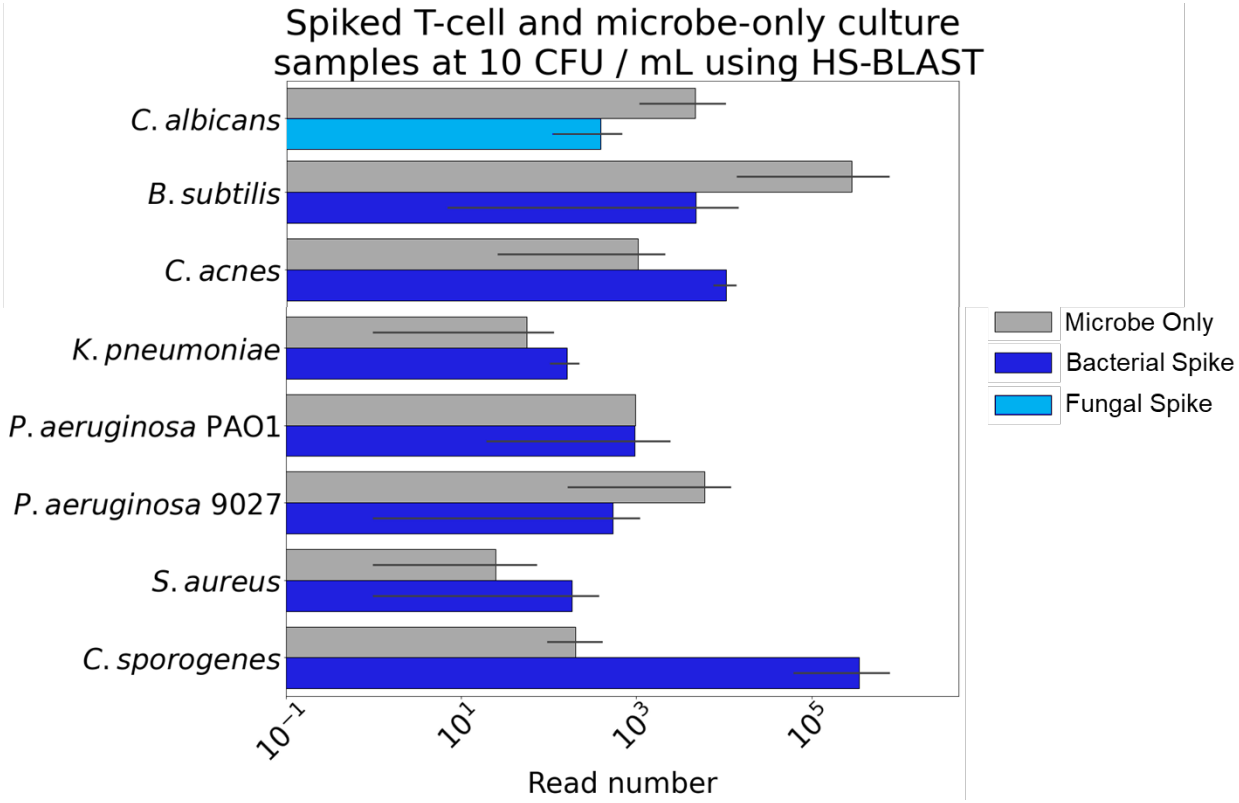
The spiked cells are CAR-T grown in G-REX



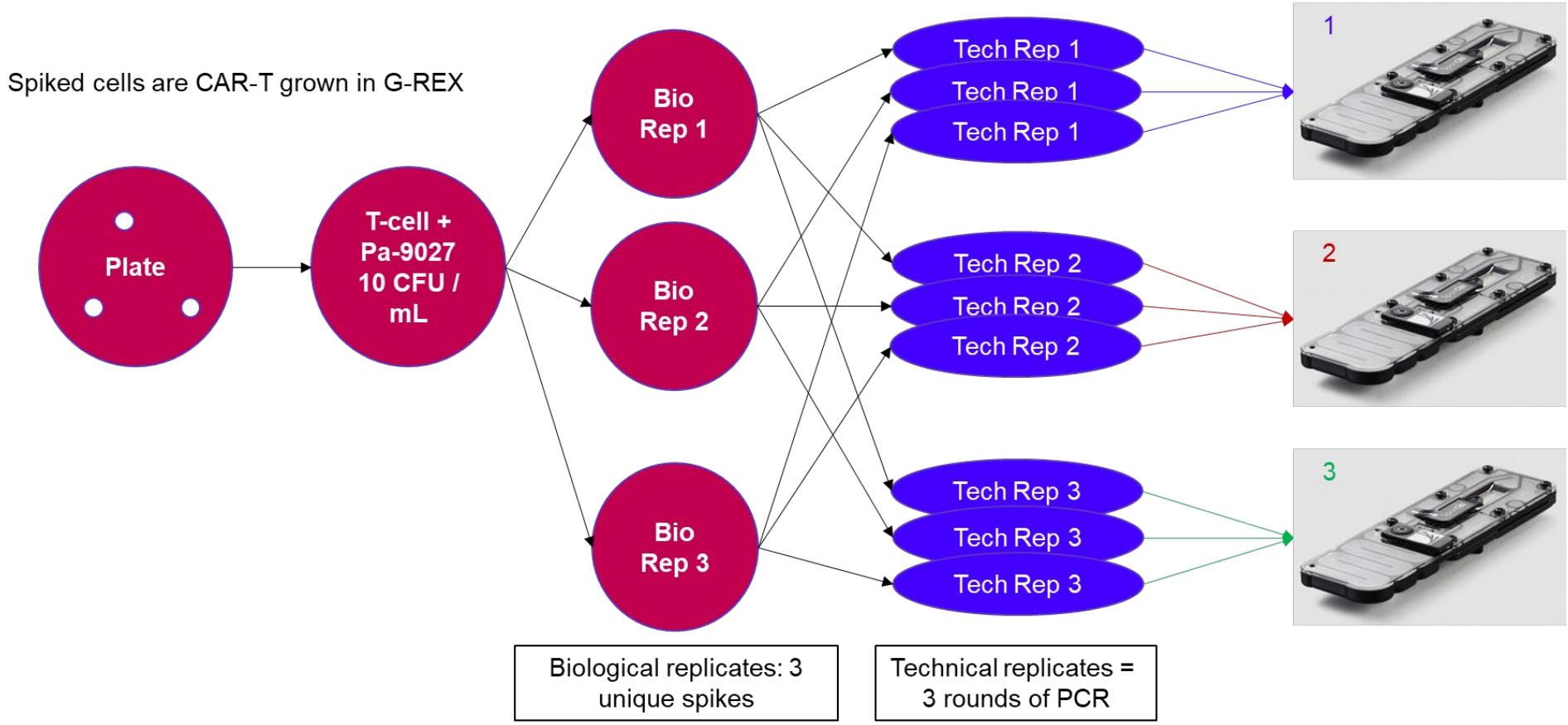
Biological replicates: 3 unique spikes

Spiking baselines: detection to 10 CFU / mL

- Generated pure culture and spiked T-cell cultures
- Assayed using either bacterial or fungal species
- **16 hour from sampling to sequenced result**
- Used either 16S or 18S-28S amplicon sequencing
- Can detect aerobic and anaerobic species at 10 CFU

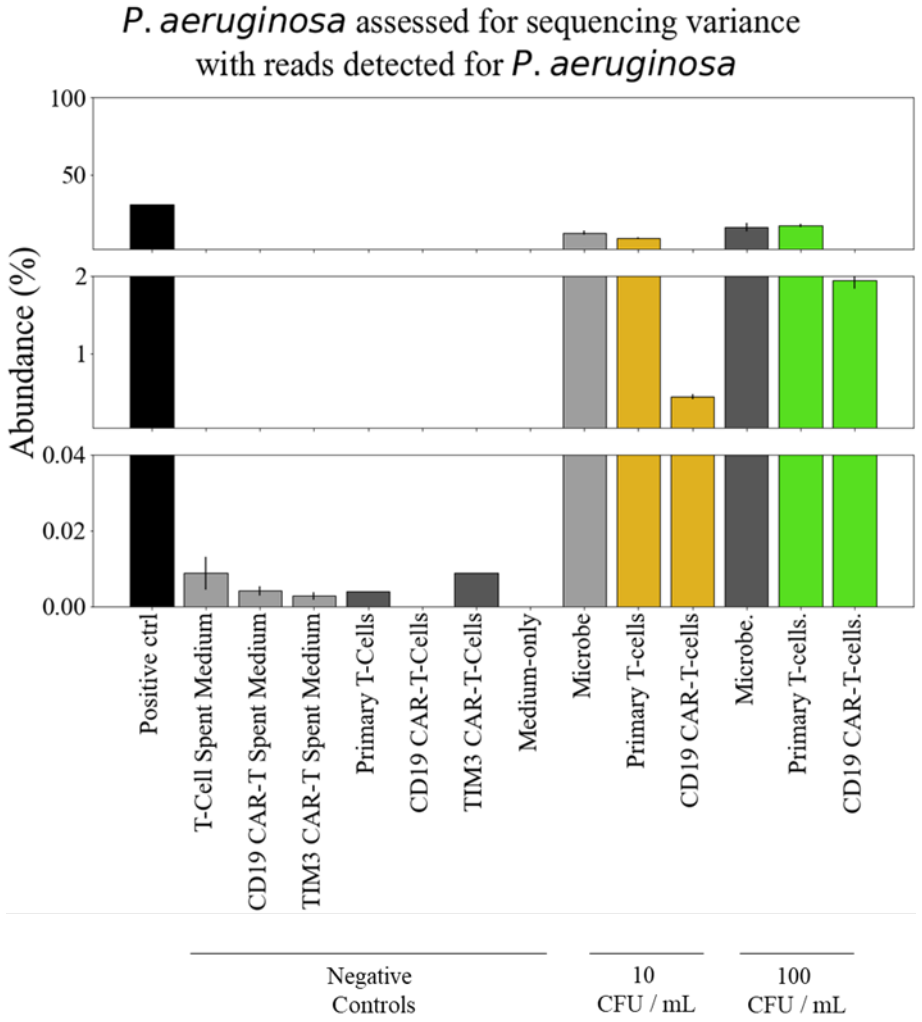


Understanding LOD and sources of reliability and variance



Proof of principle validation: Spiked *P. a* is detected in ≥1 technical replicate at 10 CFU / mL

- Sterile samples did not generate many *Pseudomonas aeruginosa* reads
- Positive controls were positive
- 9 technical replicates per biological replicate



N=3 [Biological replicates]
n=3 [Technical replicates]

Intuition for Machine Learning Binary Classification Predictions

Confusion Matrix

			Ground Truth Label: *has disease*	Ground Truth Label: *no disease*
	Total Observations (n)		Condition Positive (CP)	Condition Negative (CN)
Predicted Label	*test positive*	Test Outcome Positive (TOP)	True Positive (TP)	False Positive (FP)
Predicted Label	*test negative*	Test Outcome Negative (TON)	False Negative (FN)	True Negative (TN)

Sample (amplicon-seq)	Assessment
S .aureus, 10 CFU	Contaminated
T-cell *S .aureus*, 10 CFU	Contaminated
T-cell *K. pneumoniae,* 10 CFU	Contaminated
Cell-free medium	Sterile
Plain medium	Sterile
T-cell only	Sterile

Paired Machine Learning Models For Unbiased Identification of Contaminated Samples

Sample (amplicon-seq)	Assessment
S .aureus, 10 CFU	Contaminated
T-cell *S .aureus*, 10 CFU	Contaminated
T-cell *K. pneumoniae,* 10 CFU	Contaminated
Cell-free medium	Sterile
Plain medium	Sterile
T-cell only	Sterile

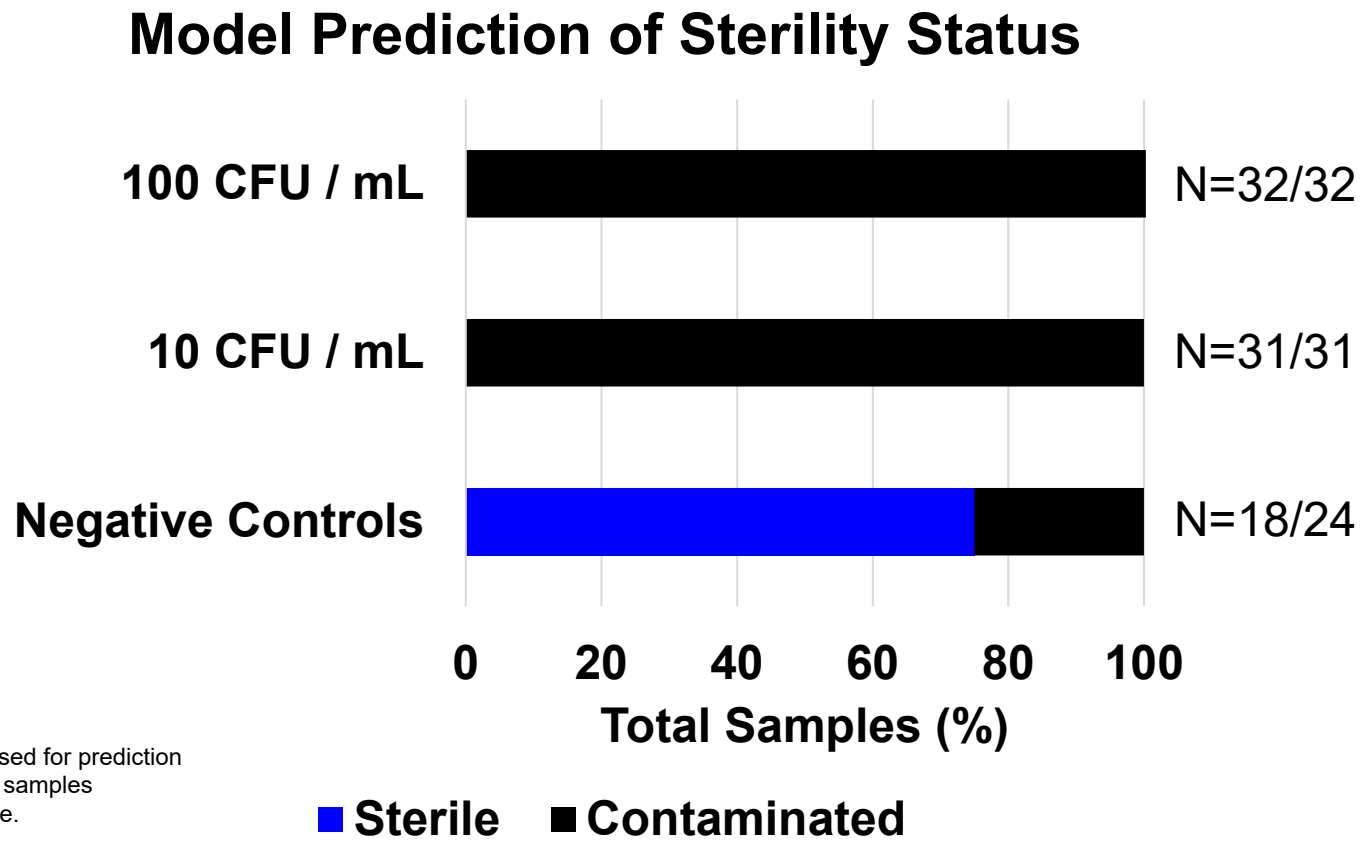
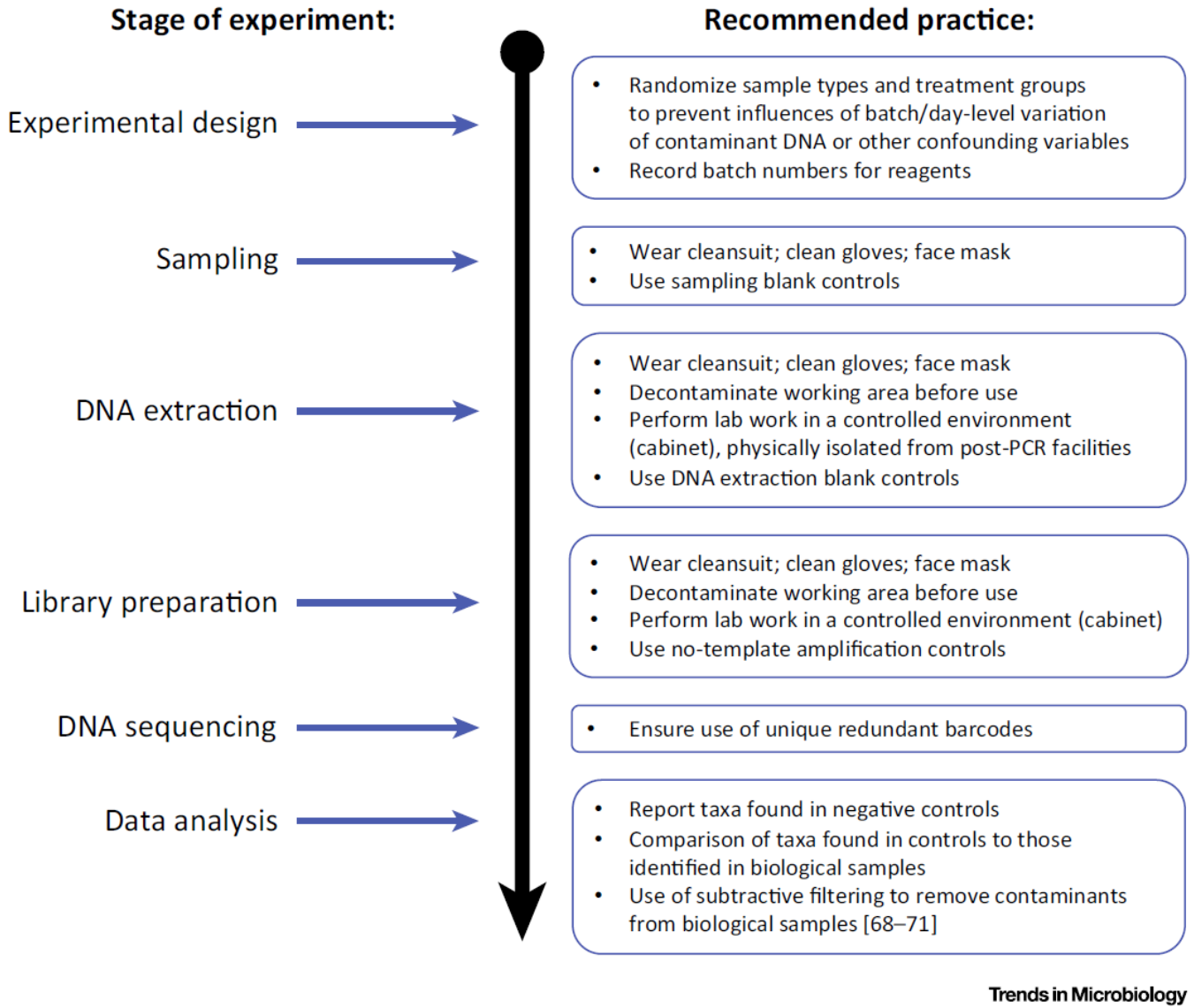


Figure to right: All spikes and negative control model predictions were assessed for prediction accuracy regarding whether the sample assayed is sterile. Black bars depict samples assigned as likely contaminated, blue bars depict samples identified as sterile.

How do we handle common contaminants without dismissing them out of hand?

Best Practises to Minimise Kitome Contaminants



Optimisation the Workflow to Minimise False Positives (A Sterile Sample is Labelled as Contaminated)

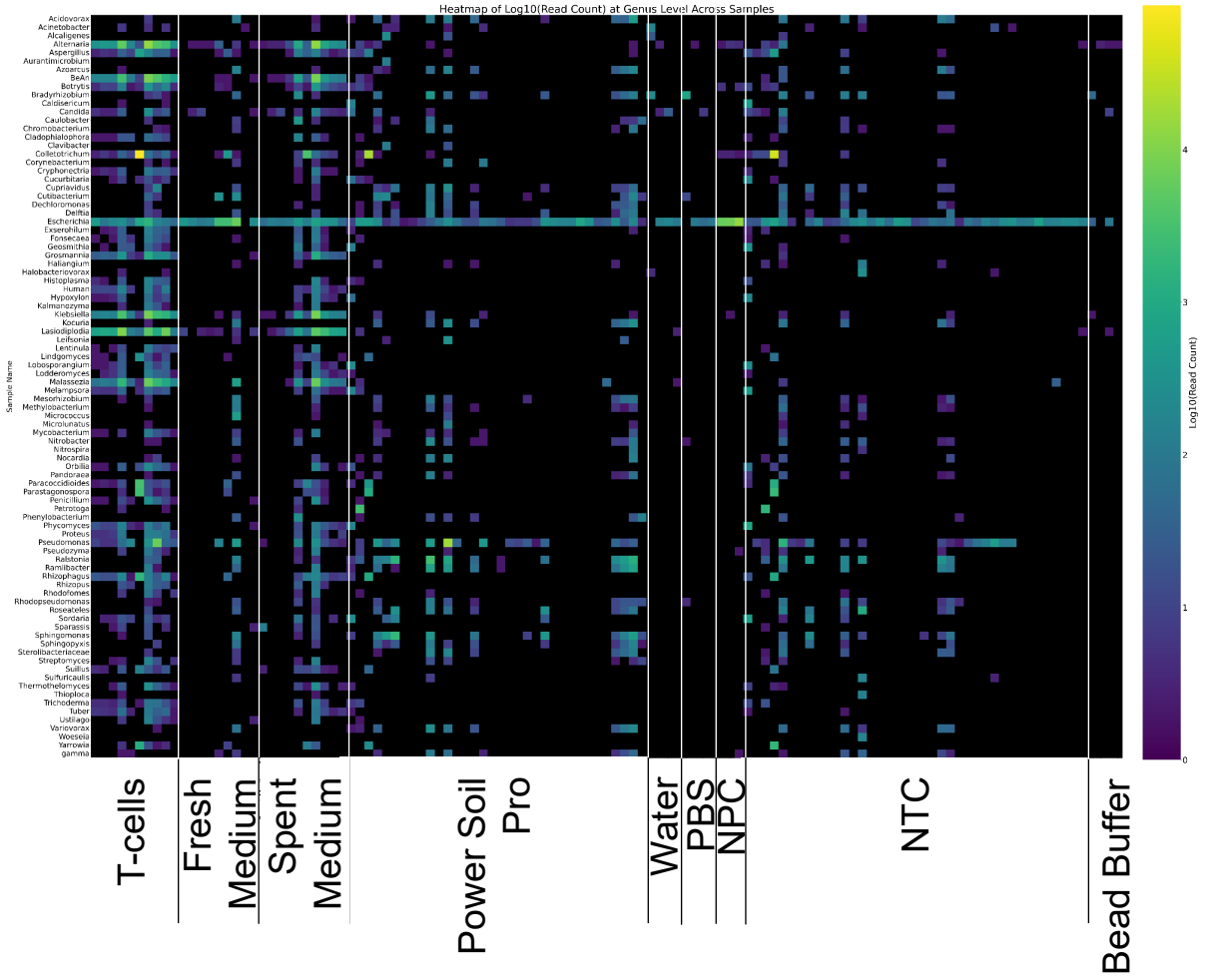
- Sample preparation optimisation

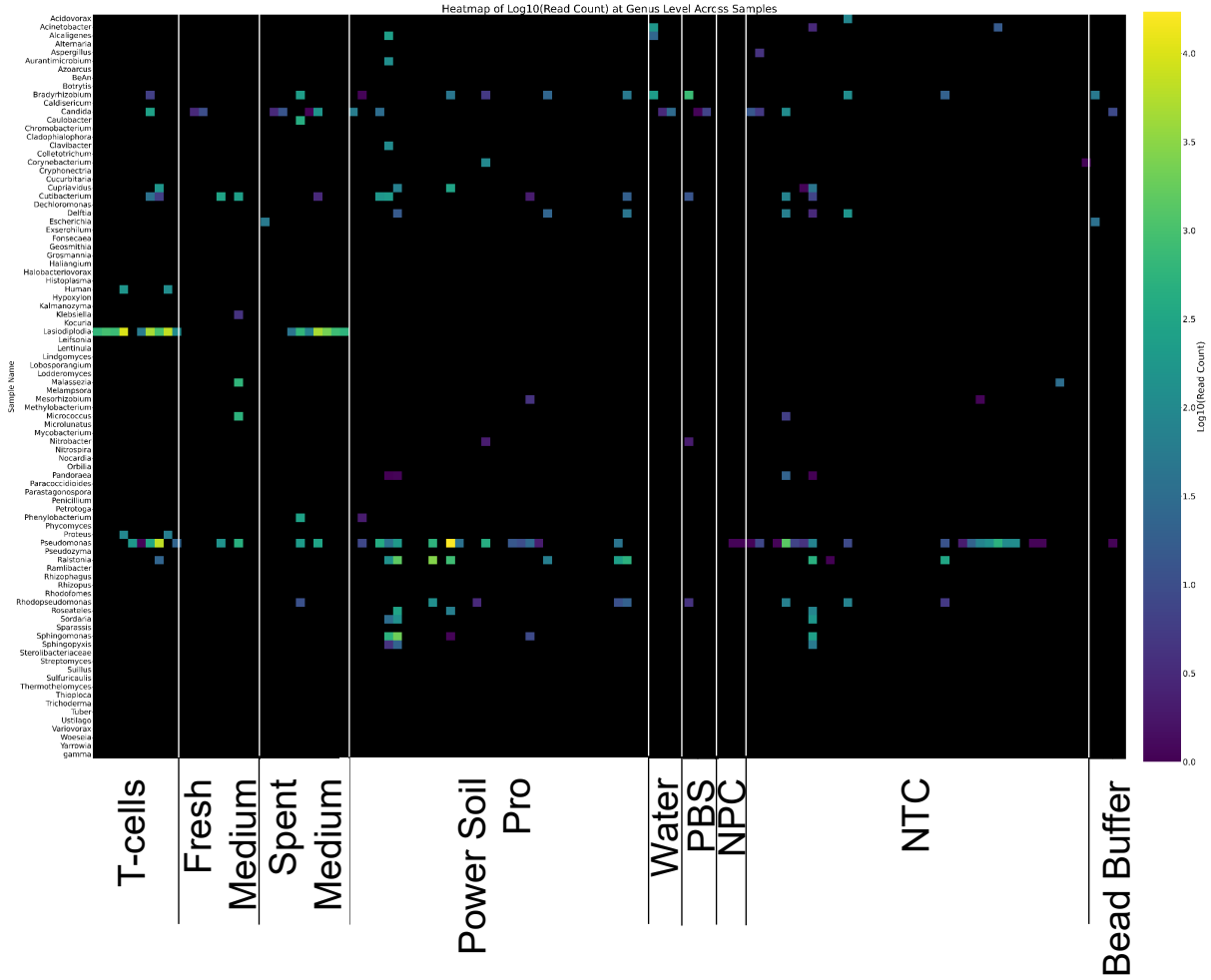
Spiking Species	Cell Pre-Filtering	Nucleic Acid Extraction	PCR conditions	Library Preparation
Fungus, bacteria, virus	Detergent, centrifugation	Kit choice, DNA, RNA	Amplicon, melting temperature, extension time	Multiplexing, Monoplexing

- Bioinformatics optimisations

Sequencing Basecalling	Read Pre-Filtering	Metagenomic Classification	Coverage Analysis	Filtering
Low, High, Super Accurate	DNABERT-2 read encoding	Viral, fungal bacterial database; Human filtering	Alignment of classified reads against reference genomes	Identify thresholds to filter false positives and retain true positives

Kitomes are Dependent on the Type of Negative Control



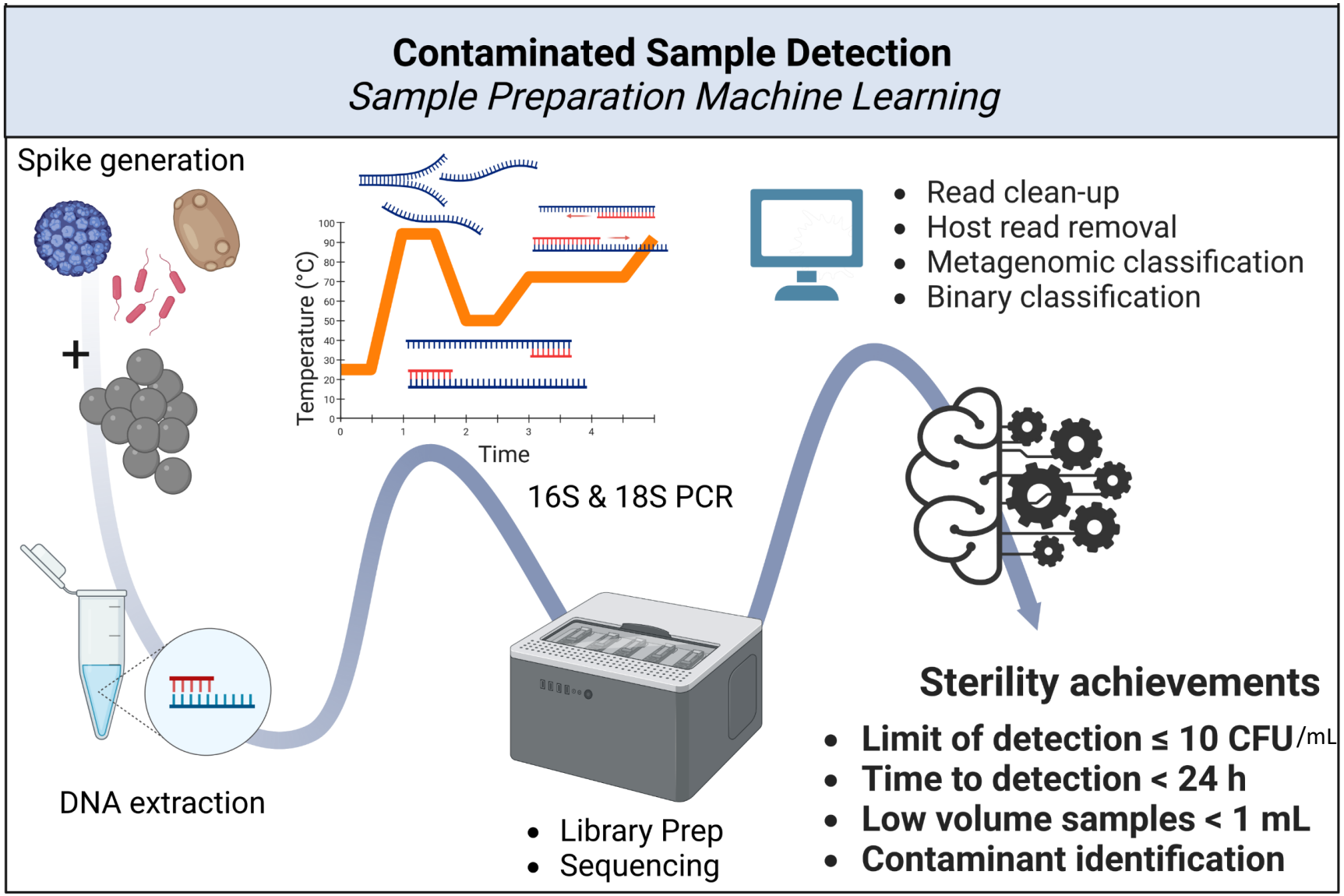


Metagenomic analysis of negative controls returns the highest number of genera

Coverage analysis (using metagenomic hits) of negative controls reduces genera hits

Each sequencing run requires negative controls run in parallel to identify sources of contamination

Summary and Conclusions



Acknowledgements

Faye Ka-Wai Cheung

Wei-Xiang Sin

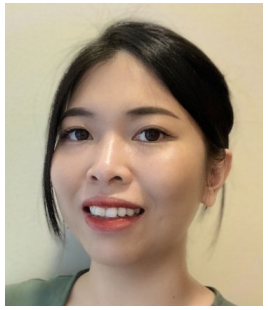
Rohan B. H. Williams

Stacy L. Springs

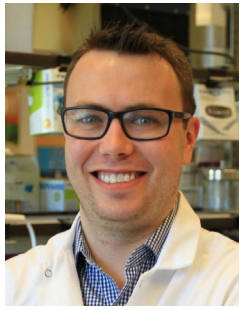
Scott A. Rice

Meenubharathi Natarajan

Denise Teo



Elizabeth Lee



Charley Swofford

Emmanuel Tsinda

Paul Barone

Jacqueline Wolfrum

Marvin Chew

Khaing Thazin

Suiyuan Xia

- Supported by A*STAR under the IAF-PP grant **"ASPHALT" (H24J4a0031)**.
- Supported by NRF Singapore under the CREATE programme through **SMART CAMP**.