

# Investigating Microbe Associated Prostate Cancer from Short Read Whole Genome Sequences

Gihawi, A.<sup>1, ID</sup>

@AbrahamGihawi

A.Gihawi@uea.ac.uk

Hurst, R.<sup>1</sup> Cooper, C.S.<sup>1</sup> Brewer, D.S.<sup>1,2</sup>

<sup>1</sup> Bob Champion Research and Education Building, University of East Anglia, Norwich, UK

<sup>2</sup> Earlham Institute, Norwich, UK

## Background

Prostate cancer is the most common cancer in men in the UK with more than 52,000 new diagnoses annually. We recently discovered that a subset of anaerobic bacteria appears to be associated with aggressive prostate cancer – the Anaerobic Bacterial Biomarker Set (ABBS) genera[1]. In this study, we explored whether these findings could be validated in the Pan-Prostate Cancer Group (PPCG,  $N=2,176$ ), using the taxonomic classification of short non-human whole genome sequencing reads. We additionally searched for evidence of the bacterial gene *pks*, which produces genotoxic colibactin, frequently associated with *Escherichia*.

## Methods

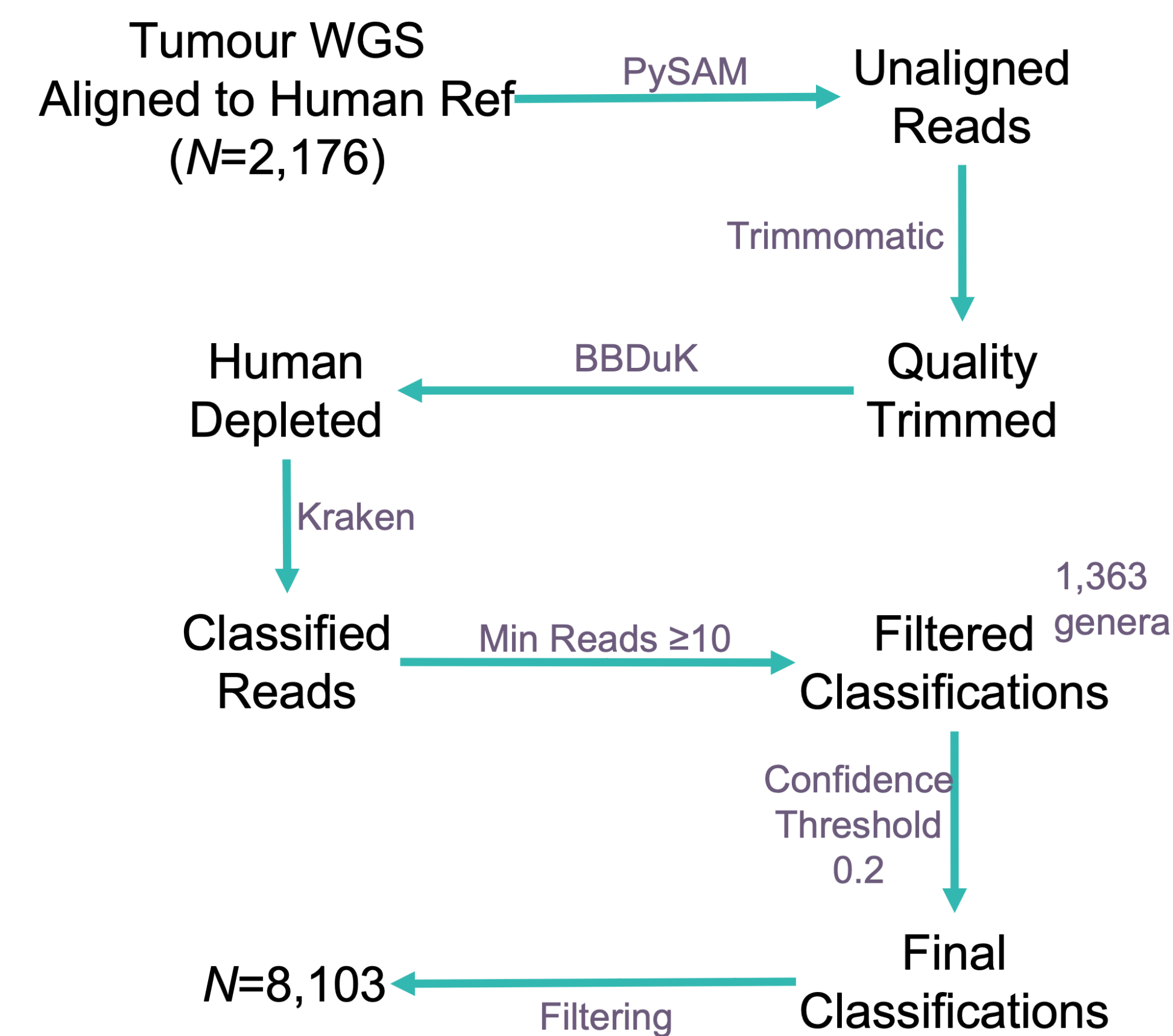


Figure 1: SEPATH was applied to tumour whole genome sequences (WGS,  $N=2,176$ ) from the Pan-Prostate Cancer Group. This pipeline extracts high quality reads unaligned to the human genome, runs additional human read removal and they are subject to taxonomic classification.

To obtain a microbial community matrix for every sample, we applied an adaptation of the (SEPATH [2] pipeline to whole genome sequences from the PPCG ( $N=2,176$ ). Read counts less than 10 were set to 0 to limit false positive classifications.

### Survival Analysis

ABBS<sup>+</sup> samples were identified in primary tumour samples by identifying any sample with evidence for any of the genera: *Fenollaria*, *Ezakiella*, *Peptoniphilus*, *Porphyromonas*, *Anaerococcus* or *Fusobacterium*. Relapse free survival was measured from date of tumour collection and analysed using the `survminer` and `survival` R packages.

### Identifying Bacterial *pks*

To search for evidence of bacterial *pks* we aligned unmapped reads from each sample to *pks* gene (`bwa mem`). The resulting BAM files were sorted, indexed and depth reports were obtained with `samtools`.

## Results

The PPCG reported 1,363 microbial genera with varying levels of evidence. Many of these genera are suspected contaminants, but there some more logical microbial genera with strong levels of evidence (fig. 2).

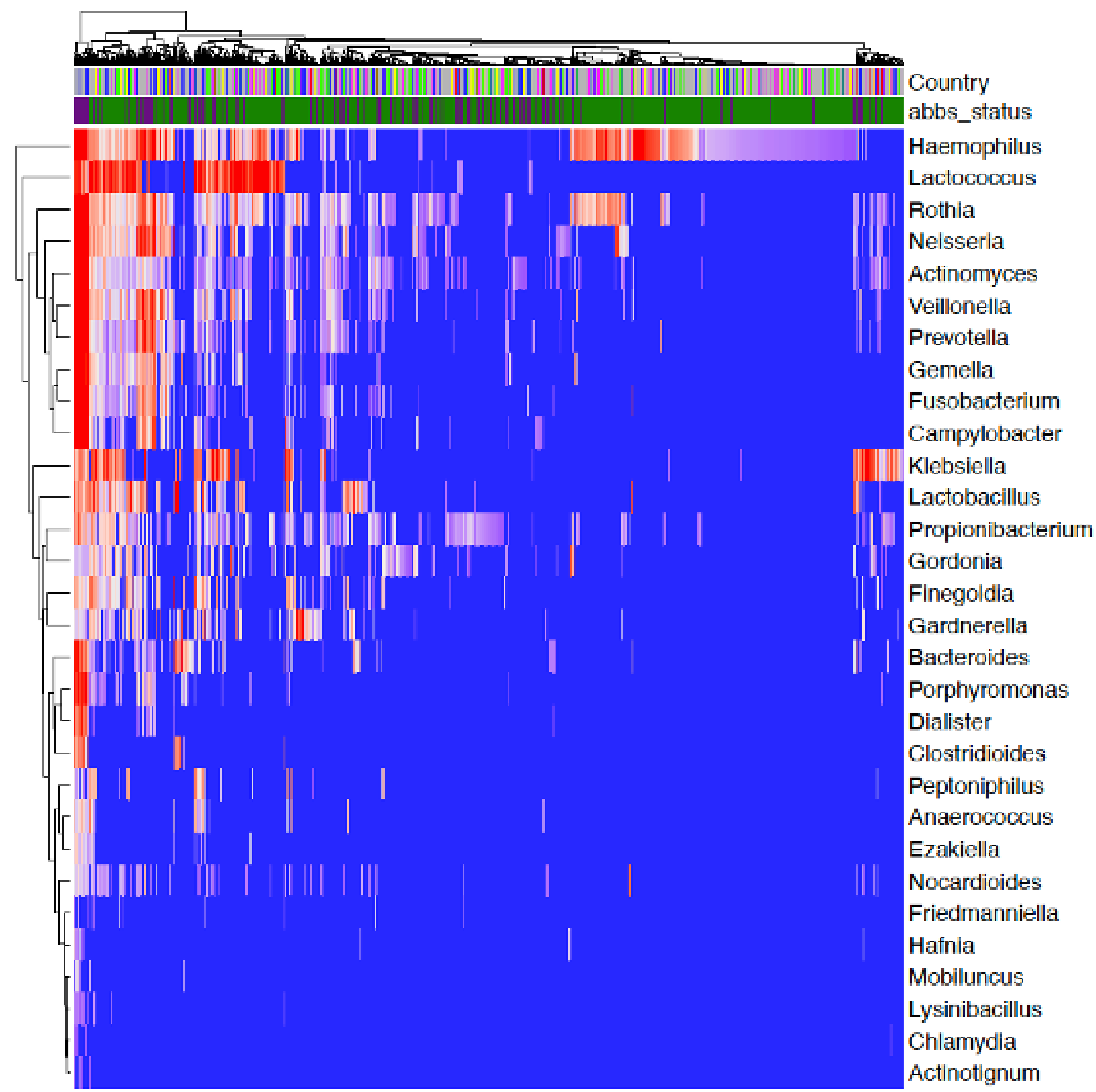


Figure 2: A heatmap demonstrating a selection of bacteria identified in the urine across the PPCG. Top bars indicate the country from which a sample was obtained and ABBS status (purple: positive, green: negative).

### ABBS is Associated with Poor Relapse Free Survival in the PPCG

We investigated the association between patients with ABBS<sup>+</sup> primary tumour samples and relapse-free survival (fig. 3). ABBS<sup>+</sup> patients demonstrated more aggressive disease in terms of relapse free survival ( $p = 0.023$ , log-rank test). We also repeated another finding, that ABBS<sup>+</sup> samples are often identified in older participants ( $p = 5.5 \times 10^{-11}$ , Wilcoxon signed-rank test)

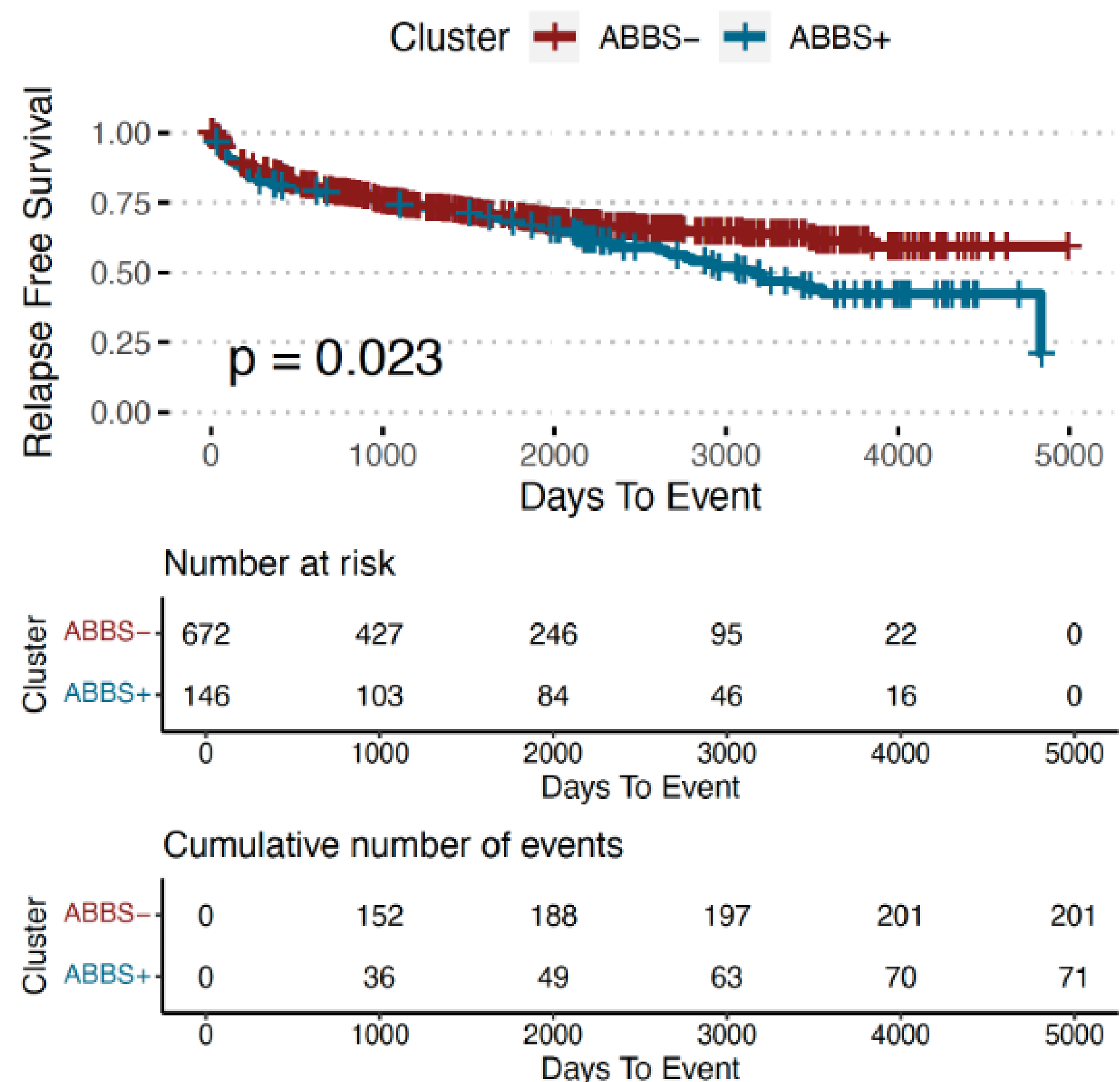


Figure 3: Survival curves demonstrating relapse free survival in ABBS positive primary tumour samples (blue,  $n=146$ ) compared with ABBS negative samples (red,  $n=672$ ). Relapse free survival is significantly worse in patients that demonstrated evidence for ABBS genera ( $p=0.023$ , log-rank test).

### Identifying Bacterial *pks*

A separate pipeline was developed to search for the genotoxic bacterial *pks* gene, which is commonly identified in *Escherichia* (fig. 4). We compared the evidence for *Escherichia* to the total number of bases aligned to the *pks* gene which revealed four clusters:

- *pks*<sup>+</sup> and *Escherichia*<sup>+</sup> ( $N=935$ )
- *pks*<sup>-</sup> and *Escherichia*<sup>+</sup> ( $N=213$ )
- *pks*<sup>+</sup> and *Escherichia*<sup>-</sup> ( $N=41$ )
- *pks*<sup>-</sup> and *Escherichia*<sup>-</sup> ( $N=987$ )

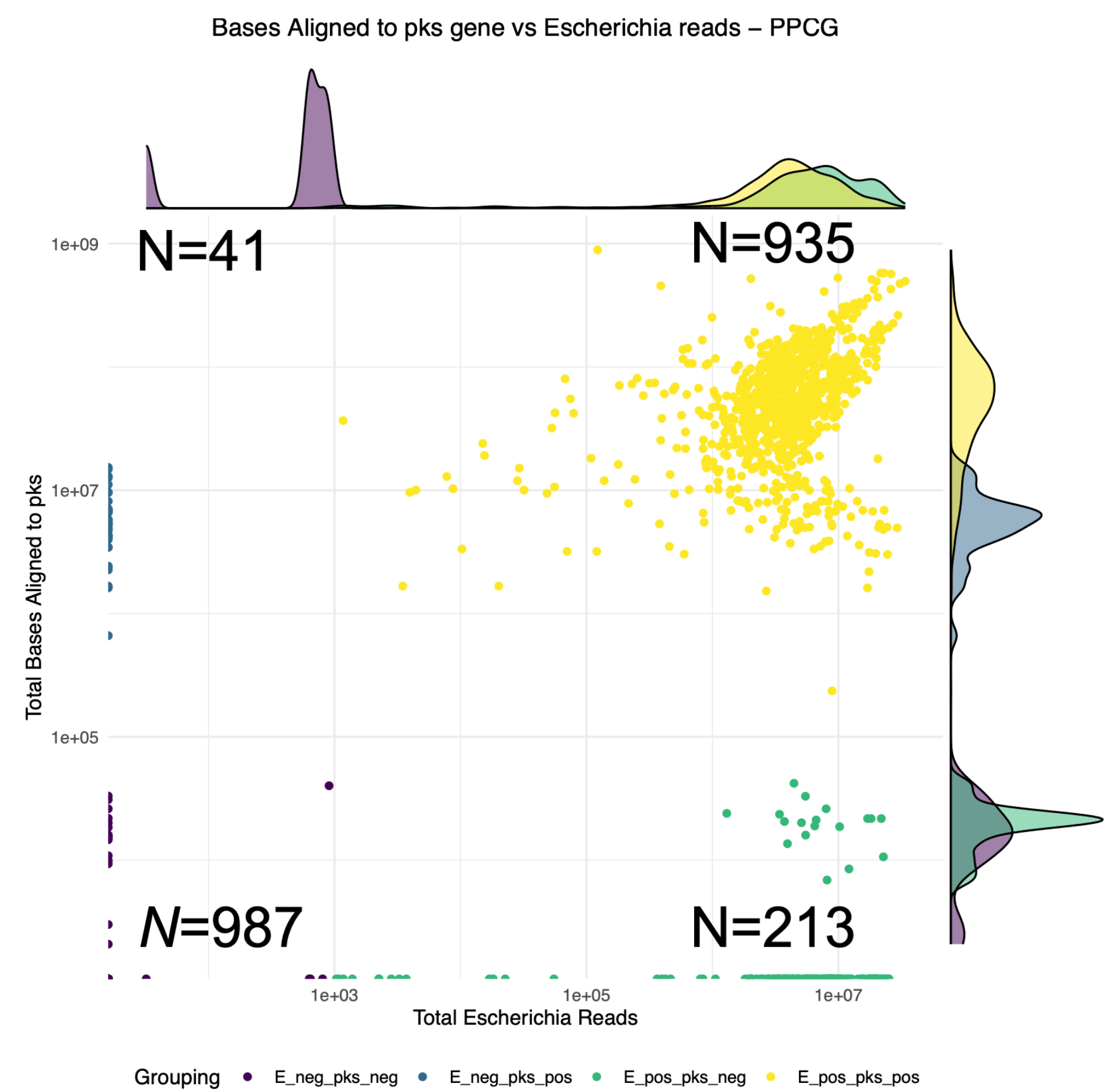


Figure 4: The number of reads associated with *Escherichia* (x-axis) compared with the total number of bases aligned to the *pks* gene (y-axis) for  $N=2,176$  whole genome sequencing samples in the PPCG cohort.

## Conclusions

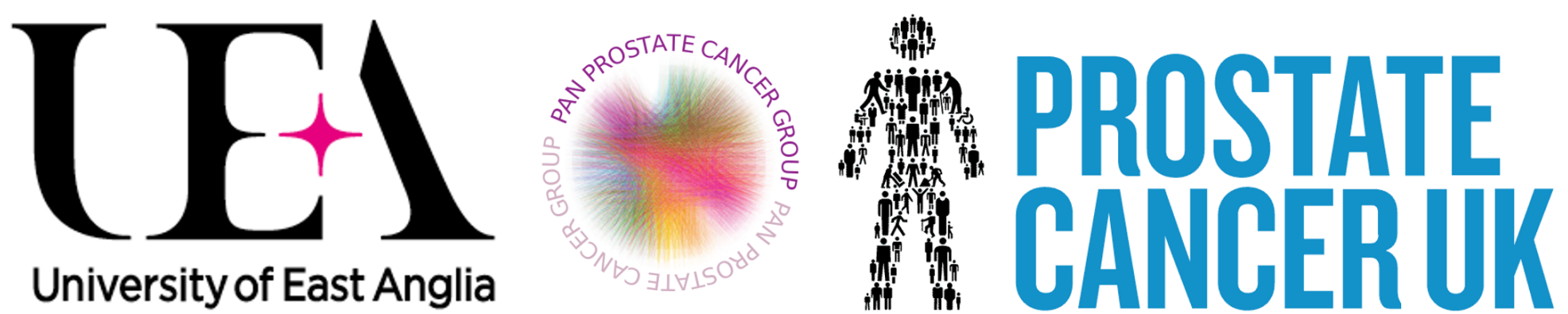
- A range of interesting microbial genera are identified in the PPCG cohort
- ABBS are identified in patients with aggressive prostate cancer
- ABBS are again identified in older participants
- Genotoxic *pks* is identified in 44.8% of samples

## Future Directions

- ABBS is useful at genus level, but what the picture like at species level?
- *Echerichia* can be both a contaminant and also biologically relevant. Is there any evidence for *pks* influencing tumours?

## Acknowledgements

Thanks to participants, the PPCG and to Prostate Cancer UK for making this research possible.



Note: This poster was produced in R. Details and code to reproduce this poster can be found: [github.com/Agihawi/ncrn2023](https://github.com/Agihawi/ncrn2023)

### References

1. Hurst R, Meader E, Gihawi A, Rallapalli G, Clark J, Kay GL, et al. Microbiomes of urine and the prostate are linked to human prostate cancer risk groups. *European Urology*. 2022;5:412–9. doi:10.1016/j.eur.2022.03.008.
2. Gihawi A, Rallapalli G, Hurst R, Cooper CS, Leggett RM, Brewer DS. SEPATH: Benchmarking the search for pathogens in human tissue whole genome sequence data leads to template pipelines. *Genome Biol*. 2019;20:208. doi:10.1186/s13059-019-1819-8.





