Searching for Evidence of Microbial Nucleic Acids in Cancer Sequencing Data

(…and emerging challenges)

International Conference on Clinical Metagenomics

Abraham Gihawi

21st Oct 2021

University of East Anglia

BIG

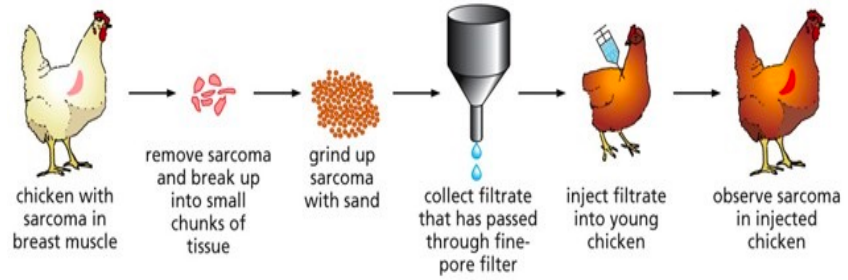
Genomics england

Outline

- Background
- Benchmarking to form a pipeline
- Genomics England
  - Microbial Structure
  - HPV Detection
  - Infectious Diseases
- Conclusions

Background

- Francis Peyton Rous – 1911
- Howard Temin – 1975
- Pathogens attributed to 15% of all cancers globally in 2012[1]



<u>Estimated cancers:</u>

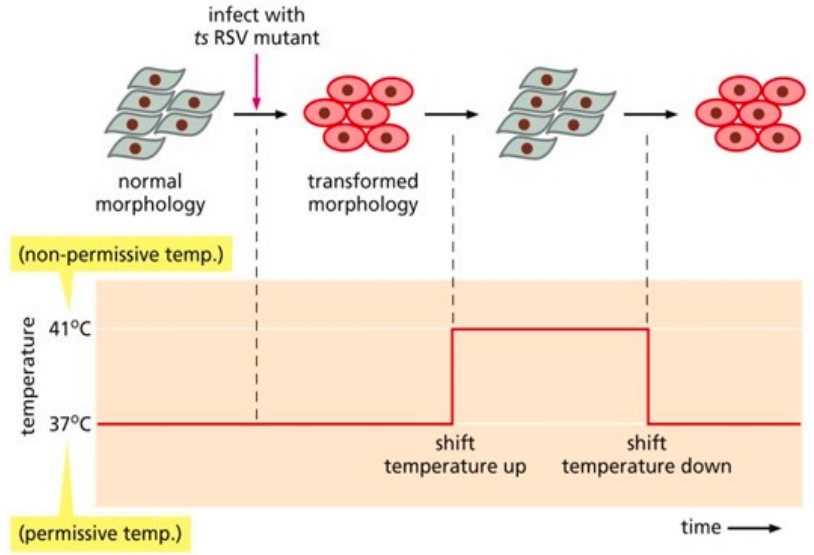
Helicobacter pylori - 770,000
HPV - 640,000
Hepatitis B - 420,000
Hepatitis C - 170,000
Epstein-Barr - 120,000

<u>Others:</u>

HHV-8, HTLV, *Opisthorchis viverrini*, *Schistosoma haematobium*

<u>Emerging or indirect causes:</u>

HIV, polyomaviruses, *Fusobacterium nucleatum, Porphyromonas, Streptococci, Enterococci faecium, Salmonella typhimurium, Human Endogenous Retroviruses*



Range of therapeutics and public health interventions

1.Plummer, M. *et al. The Lancet Global Health* 2016

Fig: Weinberg RA. The Biology of Cancer. 2014

Background – *Helicobacter pylori*

- *H. pylori* CagA injected into gastric epithelium:

- Increased proliferation:
  - Increased half-life of GTP bound ras
  - Accumulation of β-catenin

- Chronic inflammation:
  - Increased replication stress
  - Increased genome instability

- Production of free radicals
  - DNA damage

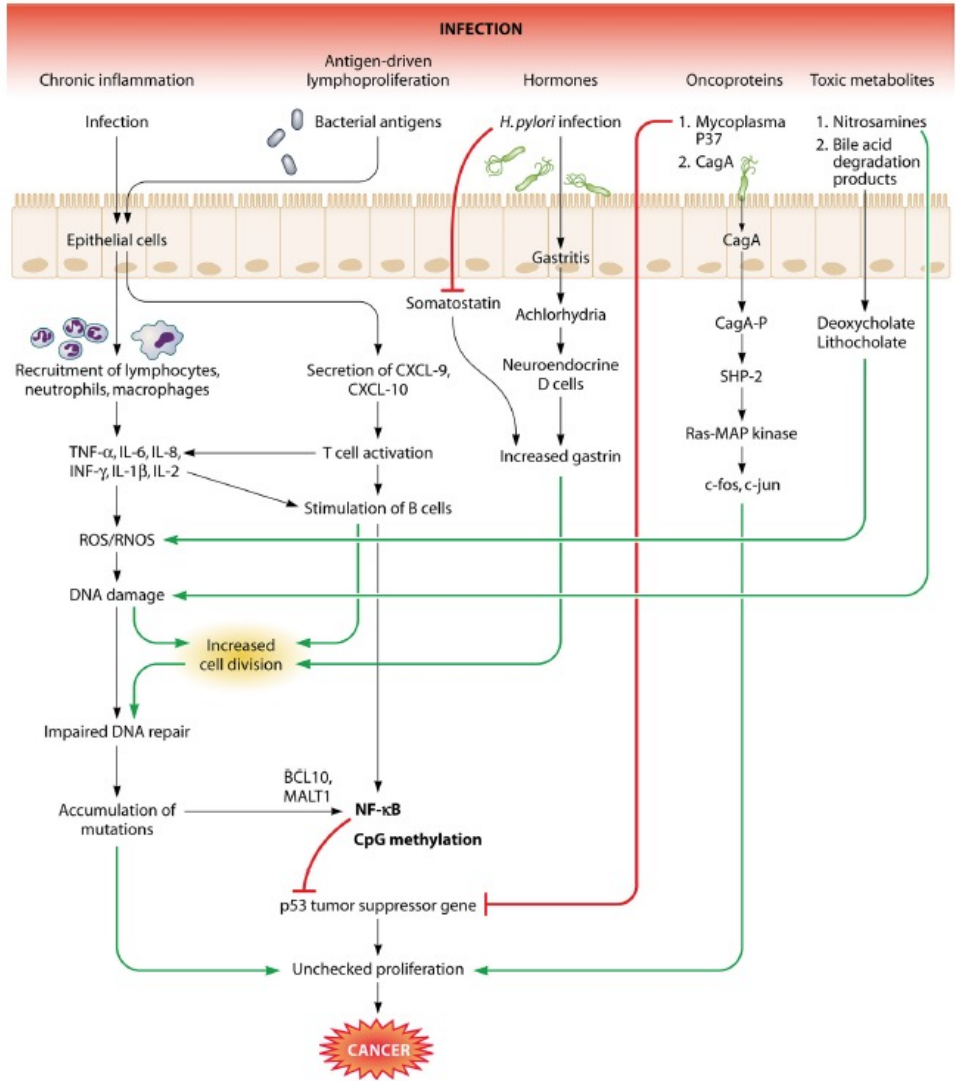


Fig: Chang and Parsonnet. *Clin. Microbiol. Rev.* 2010

Background - HPV

- Associated with most cervical cancers

- Also linked with some cases of:
  - Head & neck cancer
  - Anal cancer
  - Penile cancer
  - Vaginal cancer

- Hundreds of subtypes, a few tumourigenic
- 16 & 18 most common high-risk subtypes

- Viral proteins E5, E6, E7
- E6 stimulates P53 degradation (prevents apoptosis)
- E7 inactivates the retinoblastoma tumour suppressor
- E5, E6 and E7 assist the host cell in immune escape

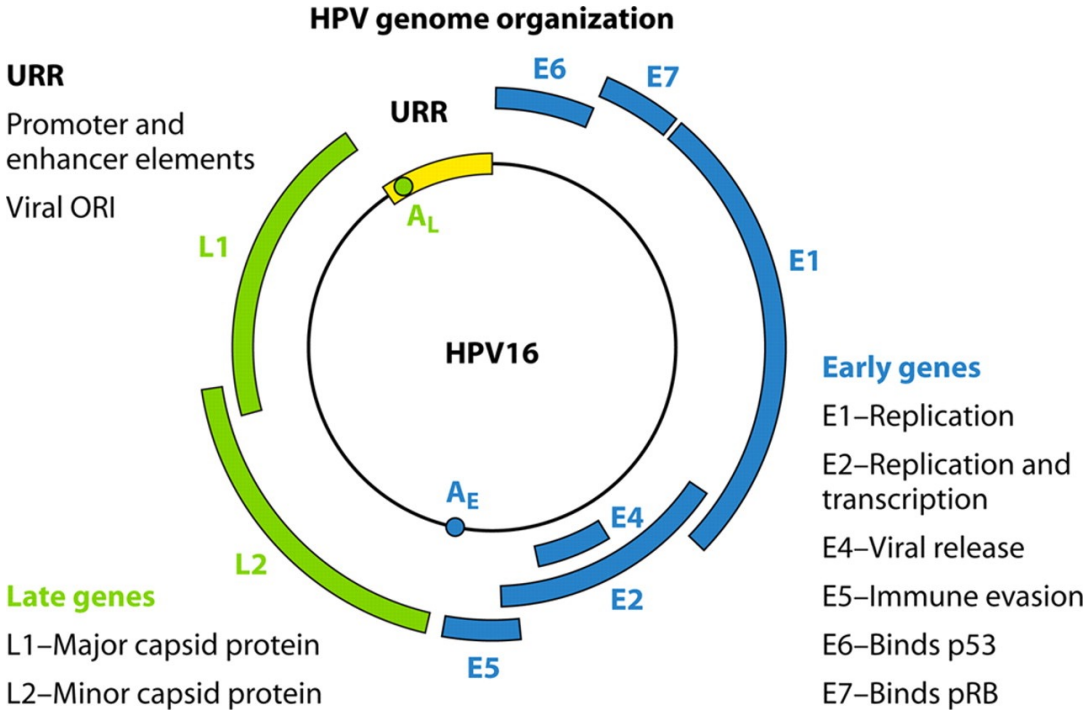


Fig: Stanley. *Clin Microbiol Rev*. 2012

Background - HPV

- HPV vaccines prepared from virus-like particles

- Vaccines available:
  - Cervarix® - subtypes 16, 18
  - Gardasil® - 6, 11, 16, 18
  - Gardasil 9® - 6, 11, 16, 18, 31, 33, 45, 52, 58
  - Cecolin® - 16, 18

- Vaccination started in girls 2008

- Vaccination extended to boys from 2018

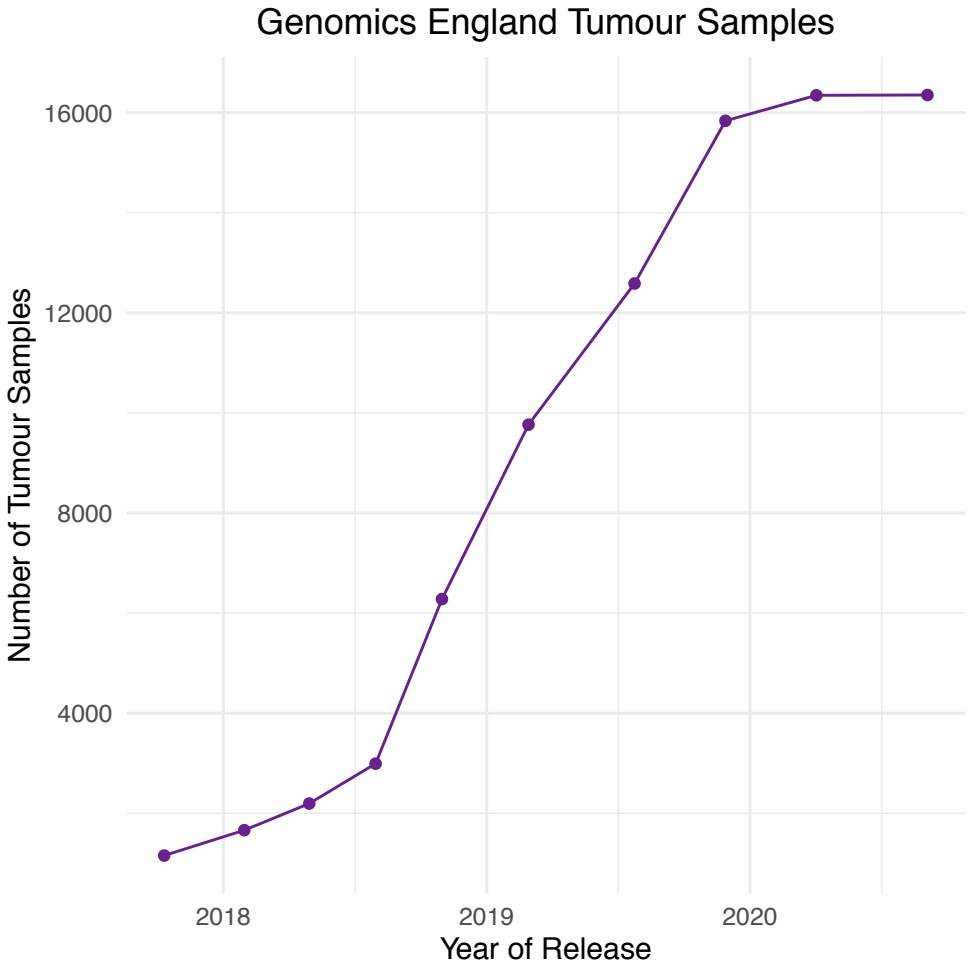
- Sweden: 88% lower risk of cervical cancer in vaccinated <17 years old vs unvaccinated[2]

2. Lei J, Ploner A, Elfström KM, *et al. New England Journal of Medicine.* 2020

Fig: UK Health Security Agency

Advent of Genomic Medicine

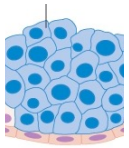
- National sequencing initiatives releasing huge quantities of cancer whole genome sequences
- Unparalleled power for cancer research
- Huge scope for patient benefit
- Genomics England's 100,000 Genomes Project
- Plans for 5 million genomes



Advent of Genomic Medicine

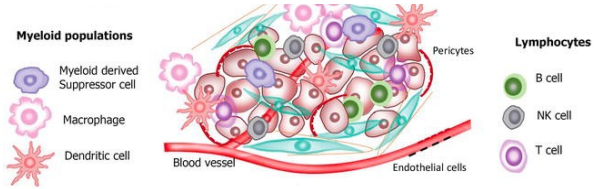
What we like to think happens:

Tumour + Sequencer = Tumour DNA sequences



```
TGACATGGGTACACATGACGGG
|||||||||||||||||||||||
ACTGTACCCATGTGTACTGCCC
```

Tumour + Surrounding Cells + Laboratory Contamination + Sequencer = *Mostly* Human DNA Mix



Microbial Reads – Bonus Insights?

- Whole genome sequencing moving into clinical practice
- Opportunity for insights into microbial composition at no extra cost
- Particularly as metadata improves with time
- Not a substitute for:
  - Well designed microbiome studies
  - Clinically validated tests
- Limitations → Hypothesis generation

Can we reliably detect microbes within human tissue whole genome sequences?

Can we leverage these incredible datasets to uncover
more about the relationship between microbes and cancer?

Are there opportunities for translational utility?

Microbial Reads – The Needle in the Haystack

Several key steps:

- Ensuring good quality data
- Removing human reads
- Accurate taxonomic classification

Cancer metagenome simulations:

- 1-77 bacterial species
- 300 million paired end reads each
- 87%-99.99% human
- 0-40 million bacterial reads

Tested a selection of tools:

- mOTUs2
- Kraken
- MetaPhlAn2
- Kaiju
- Gottcha
- Centrifuge

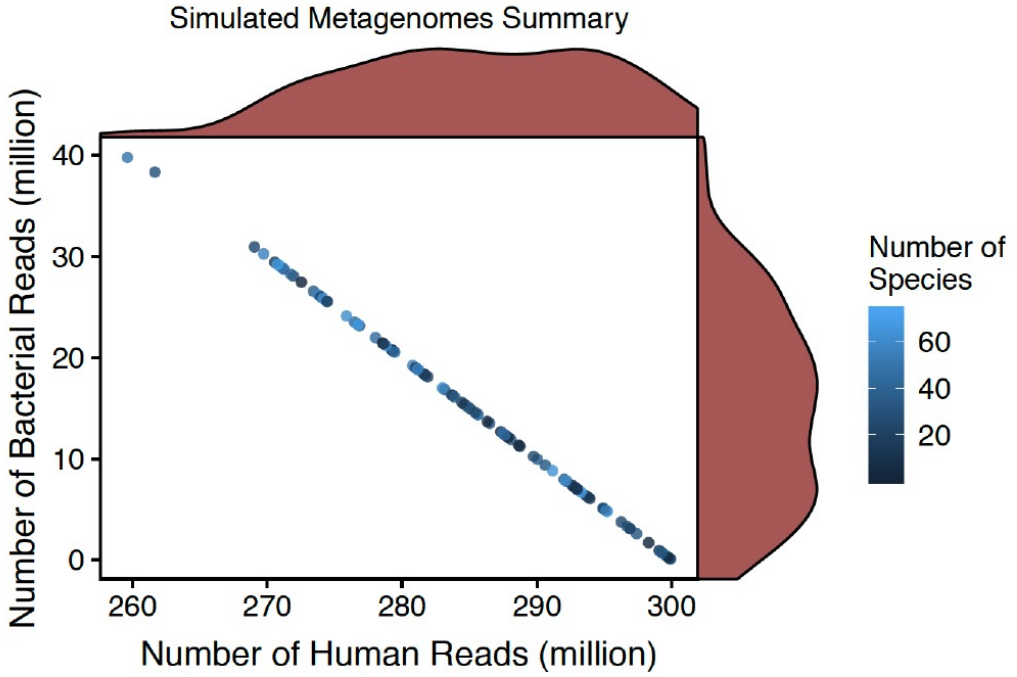
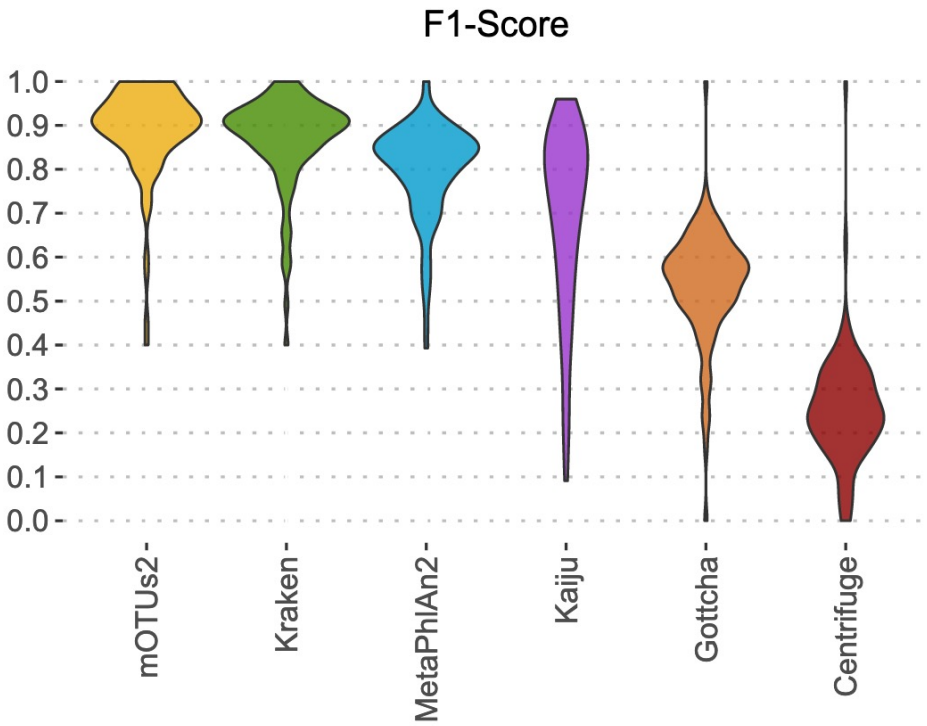


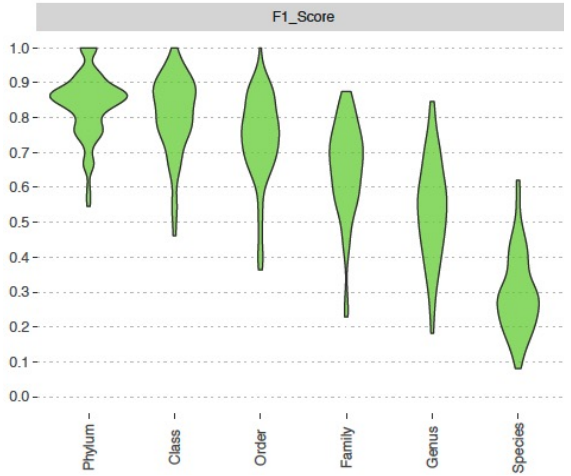
Fig: Gihawi, A. *et al.* (2019). *Genome Biol*

Benchmarking Taxonomic Classification

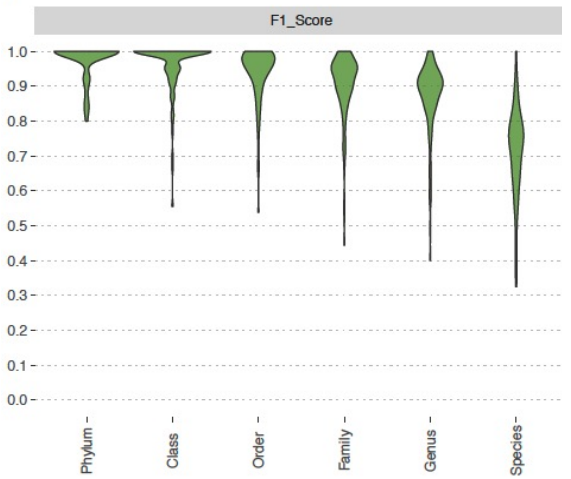
- Genus level → good performance
- Most tools performed well - mOTUs2 and Kraken in particular
- Kraken minimum read threshold is critical

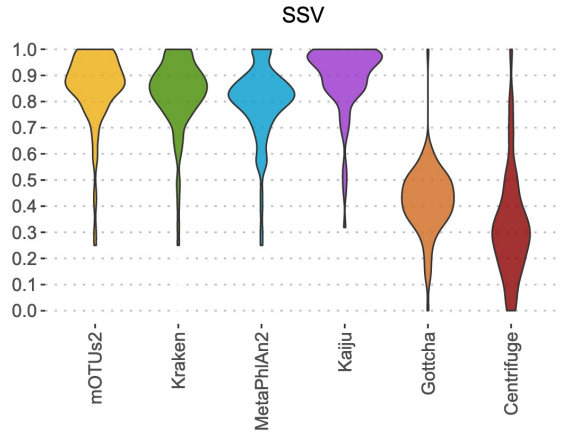


No Read Threshold



> 100 Reads





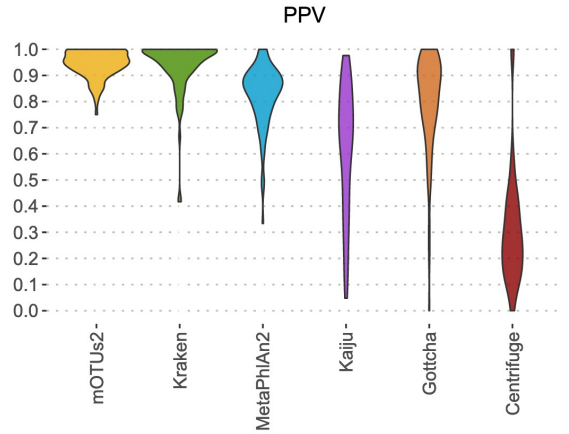
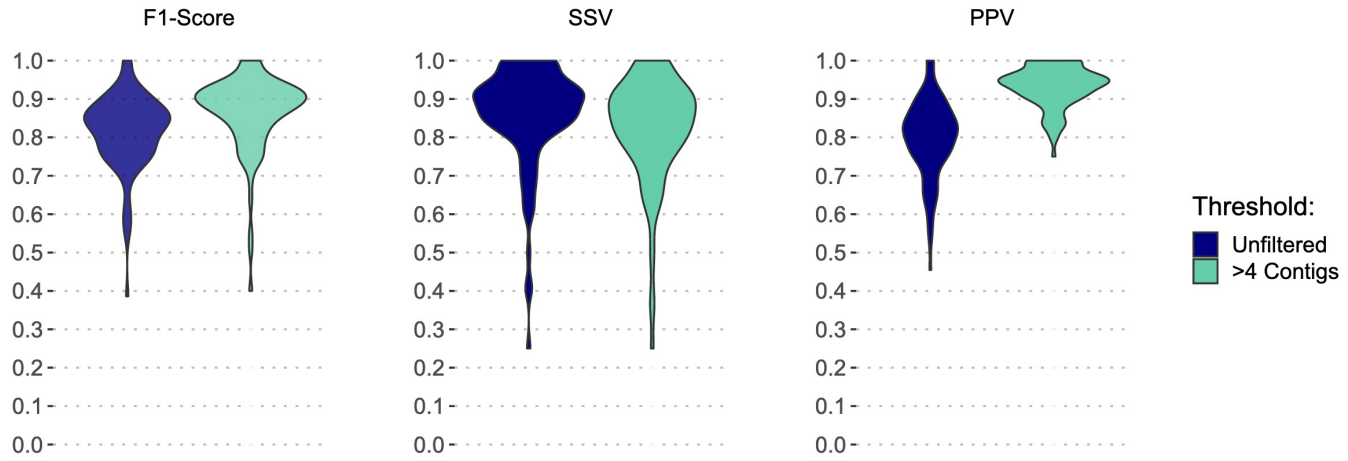
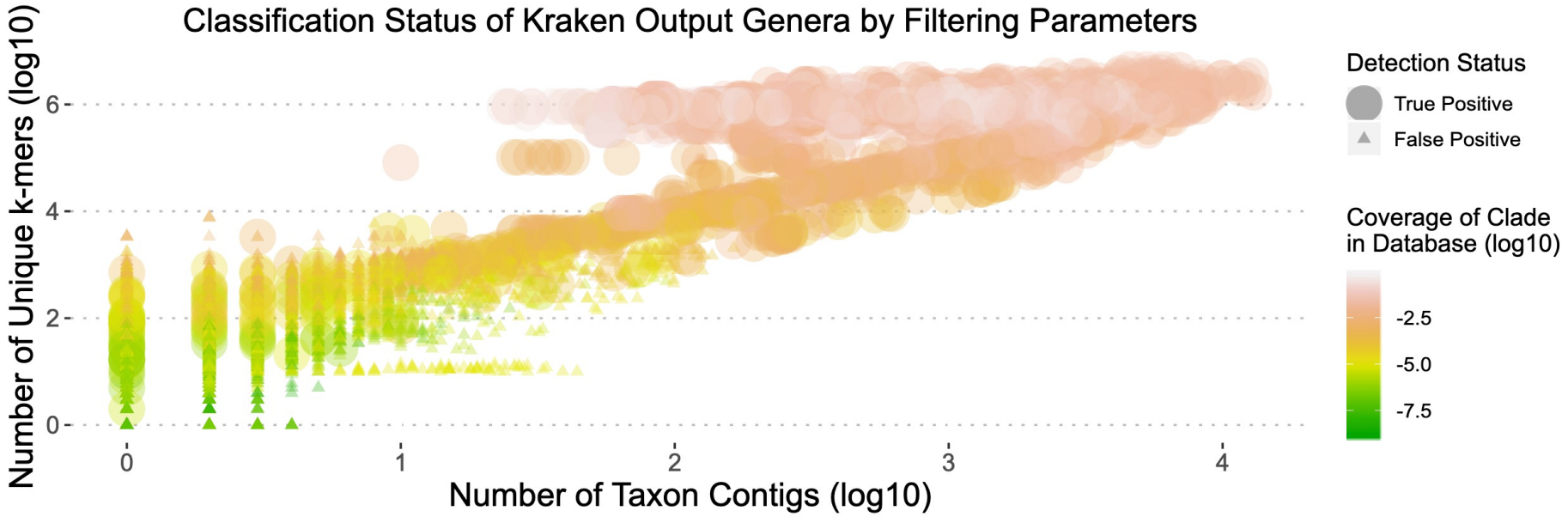


Fig: Gihawi, A. *et al.* (2019). *Genome Biol*

Benchmarking Metagenomic Assembly

- Assembled non-human reads in each sample
- Classification with Kraken still good
- Less requirement for a minimum read threshold
- Perfect classification impossible





Benchmarking – Real Data

- Applying to real data suggests high performance

Cervical:
- Alphapapillomavirus in 9/10 samples, high reads
- mOTUs2 doesn't detect viruses

Stomach:
- *H. pylori* in 4-5 samples

- Range of other expected microbes reported

Medulloblastoma:
- Not much…

- Likely contamination throughout

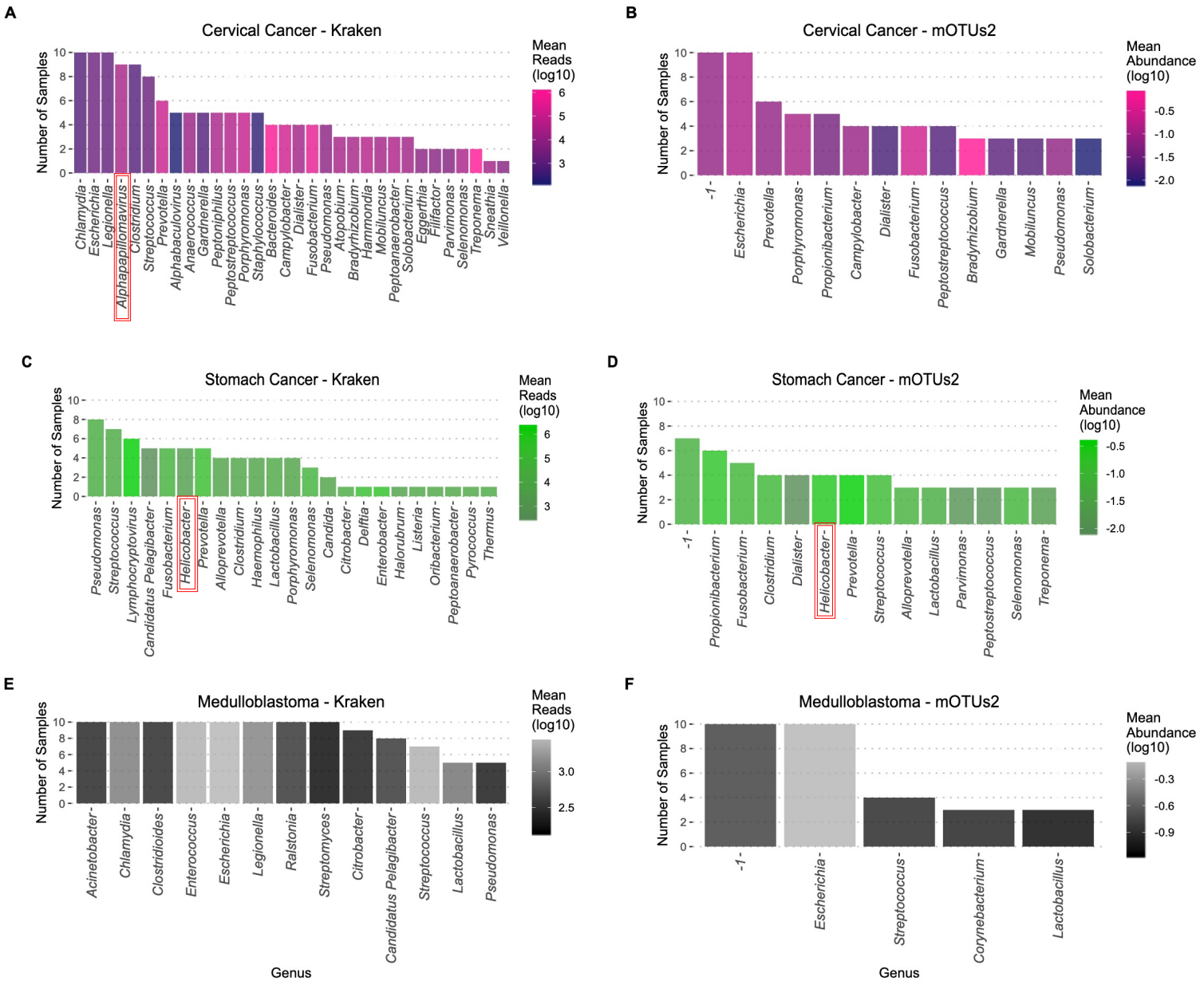
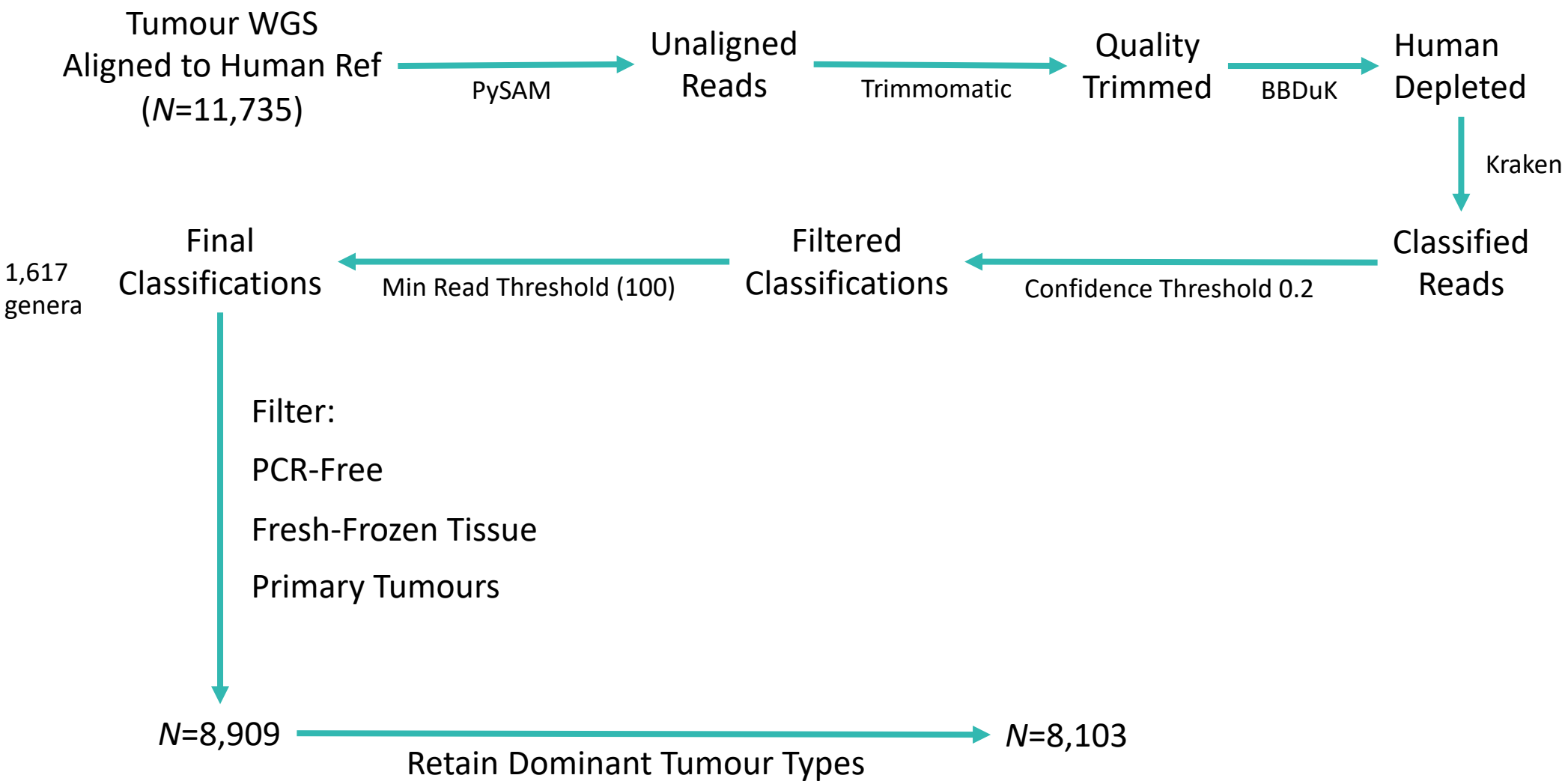


Fig: Gihawi, A. *et al.* (2019). *Genome Biol*

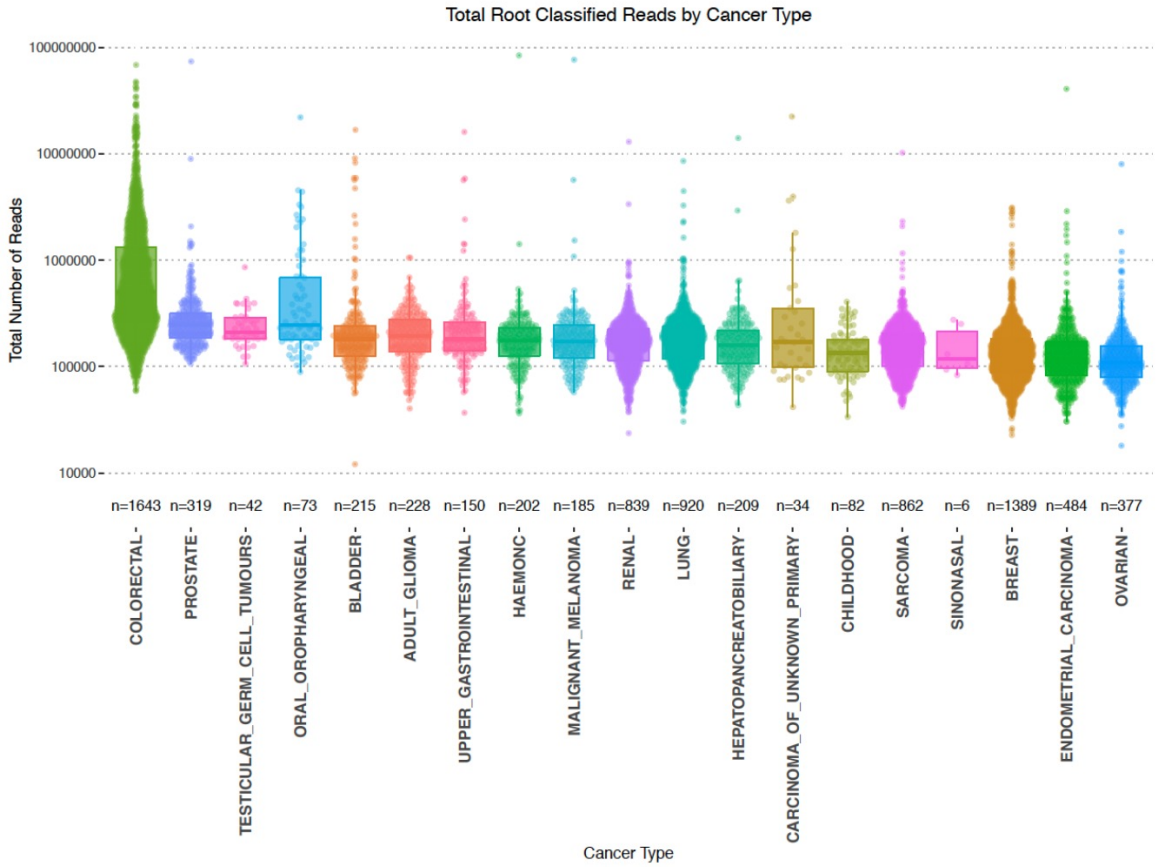
100,000 Genomes Project

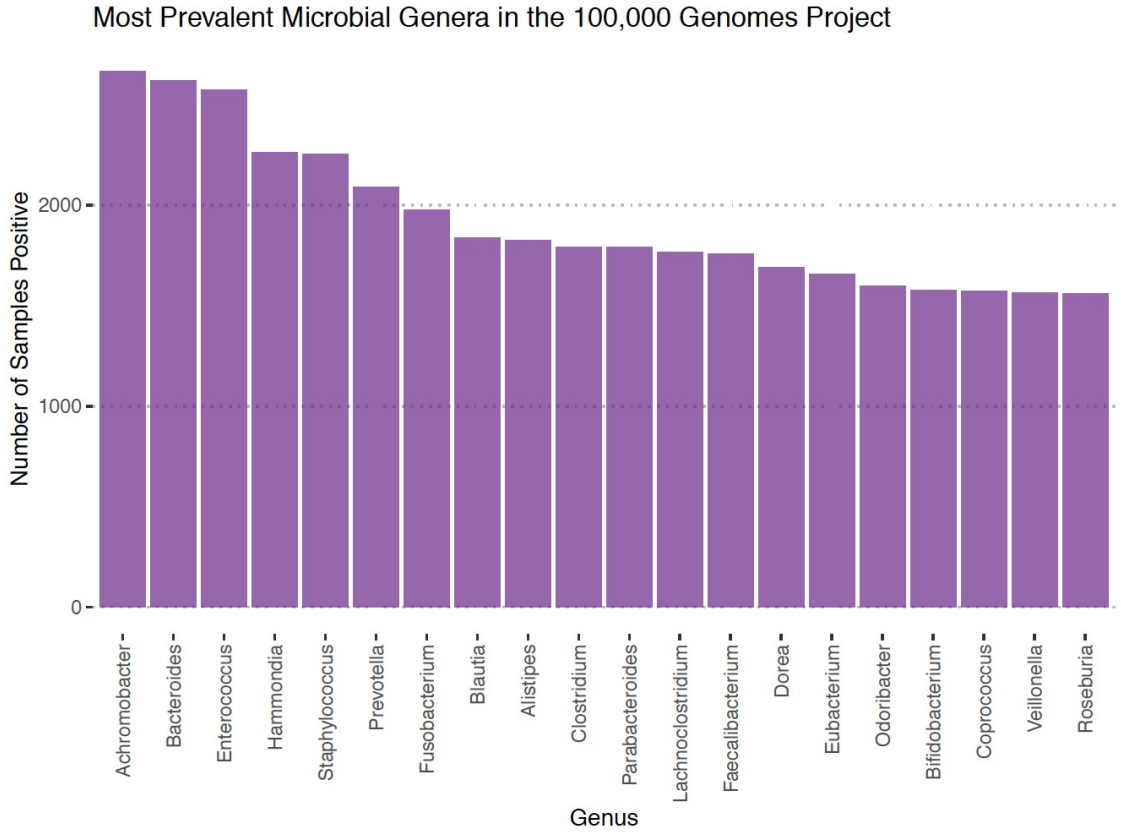
Genomics England – Technicalities



Genomics England - Overview

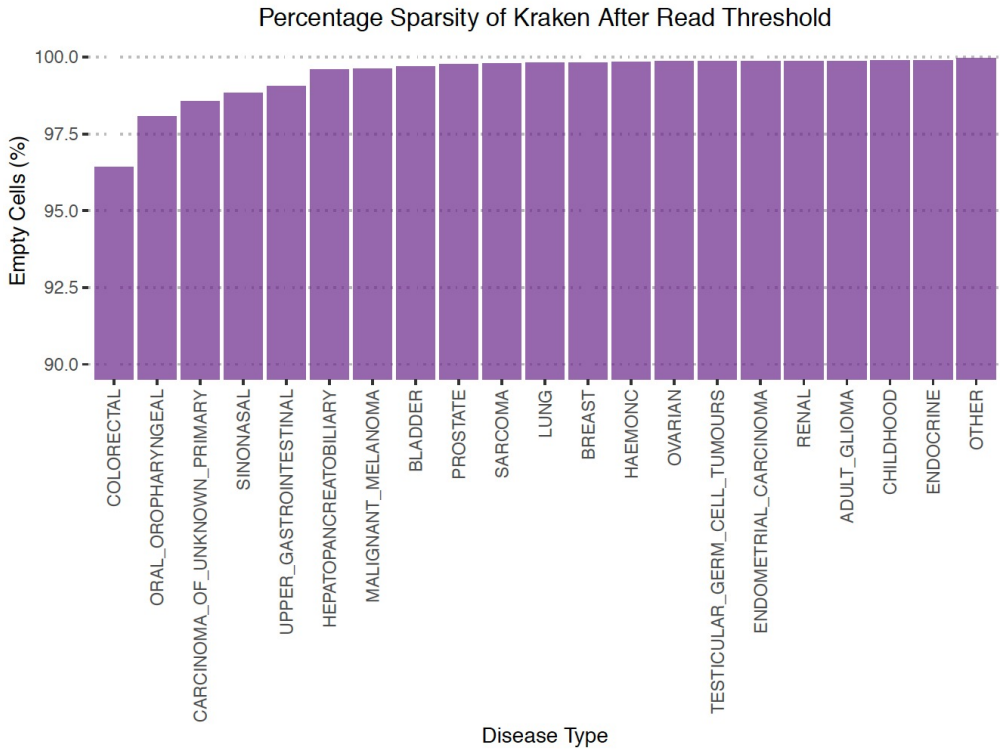
- Colorectal/Oral Dominate
- *Homo* - most common genus despite 2x human depletion
- Contamination

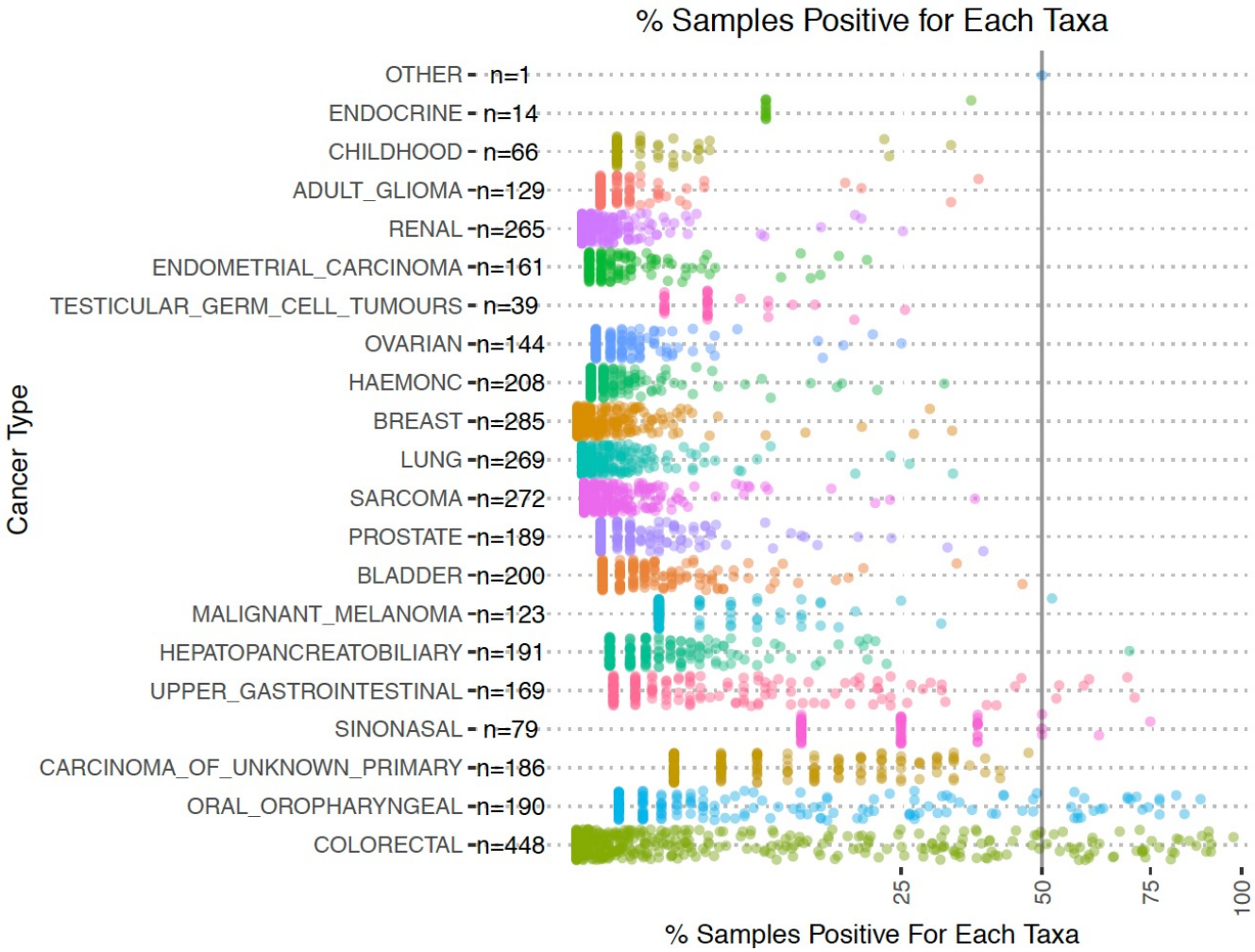




Genomics England - Sparsity

- Overall dataset incredibly sparse

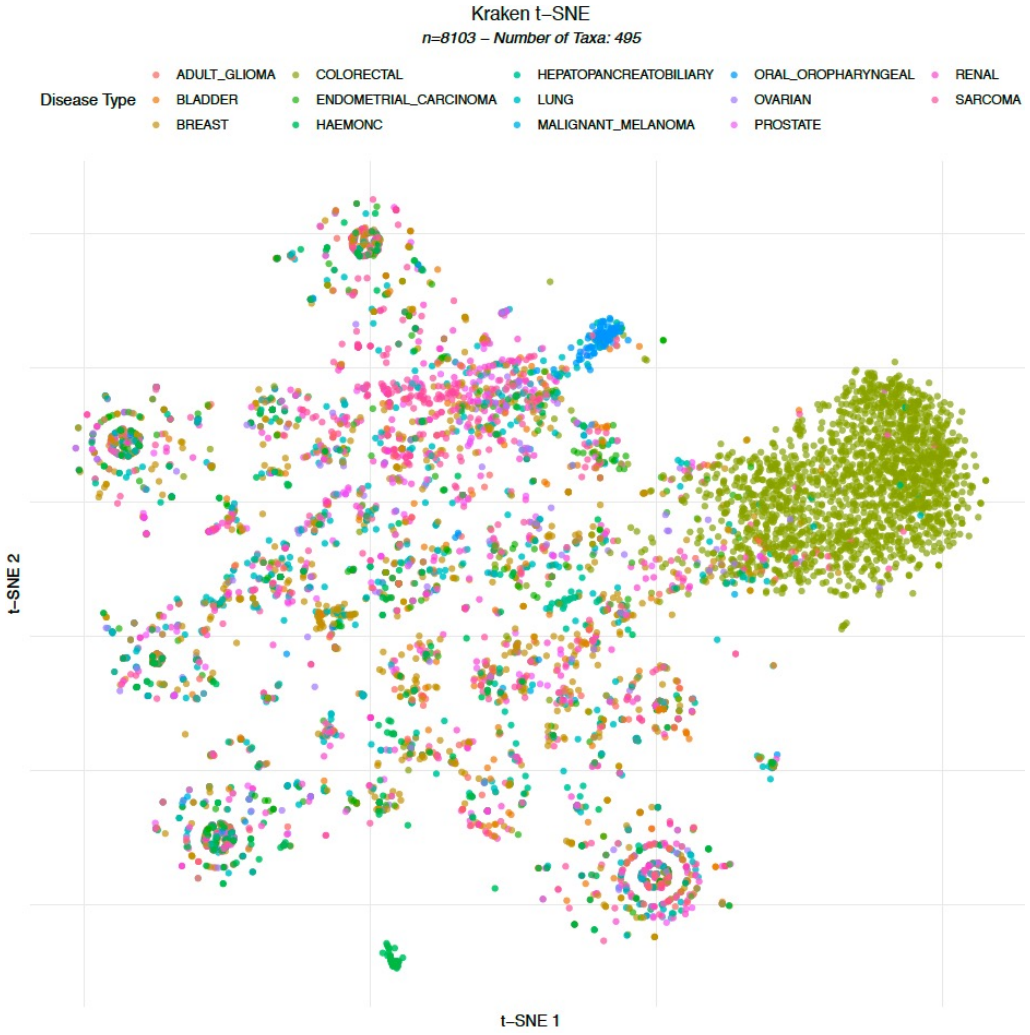




Min read threshold = 20

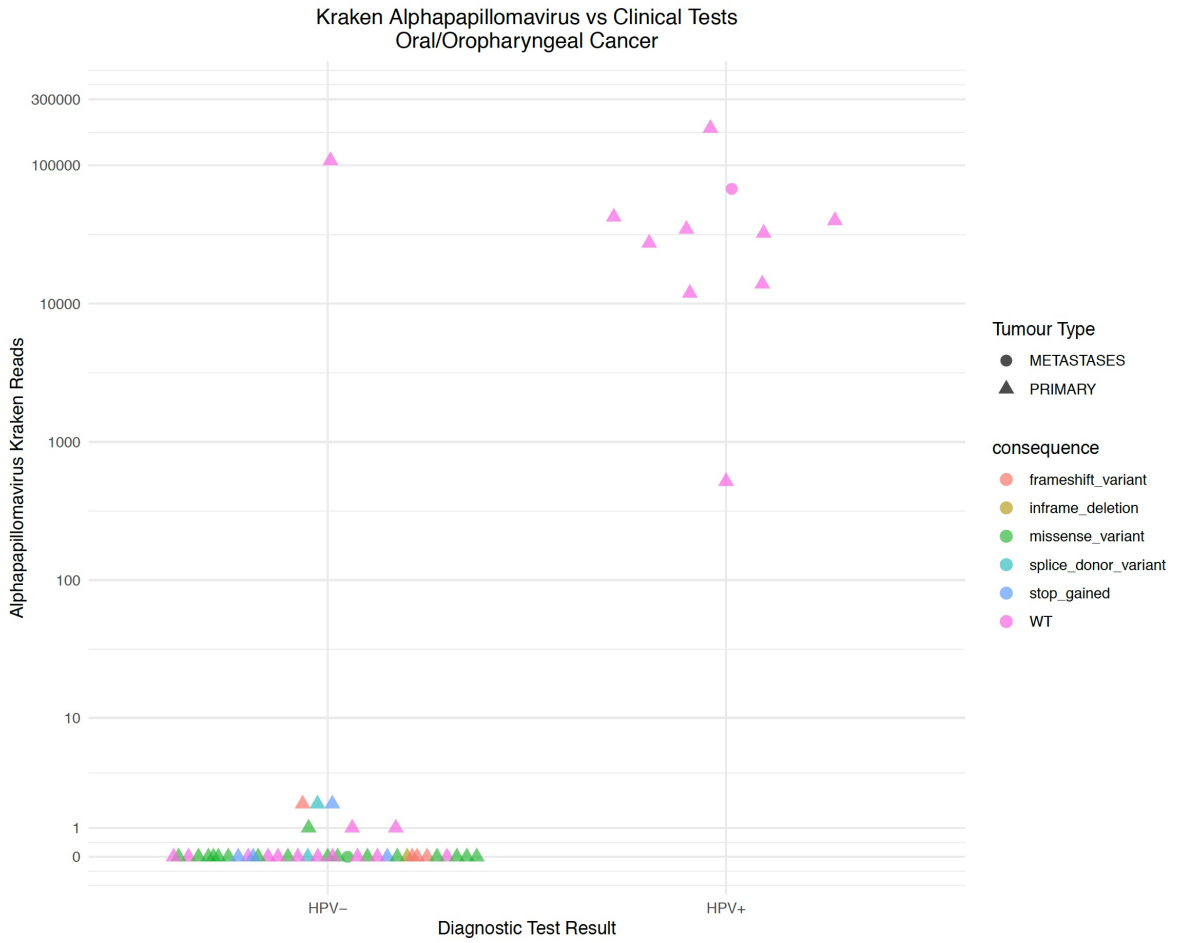
Genomics England – Overall Structure

- Boruta highlights genera informative of tumour type
- Most distinctive structure obtained →
- Oral/Colorectal appear distinct
- Limited structure in other tumours



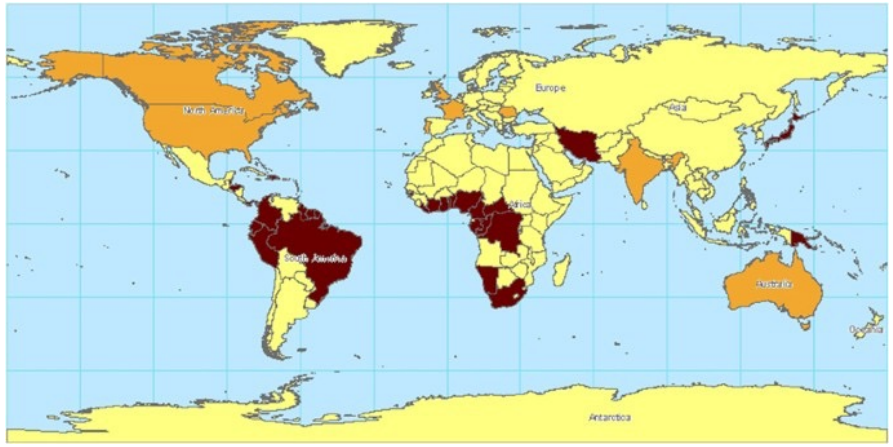
Genomics England – HPV

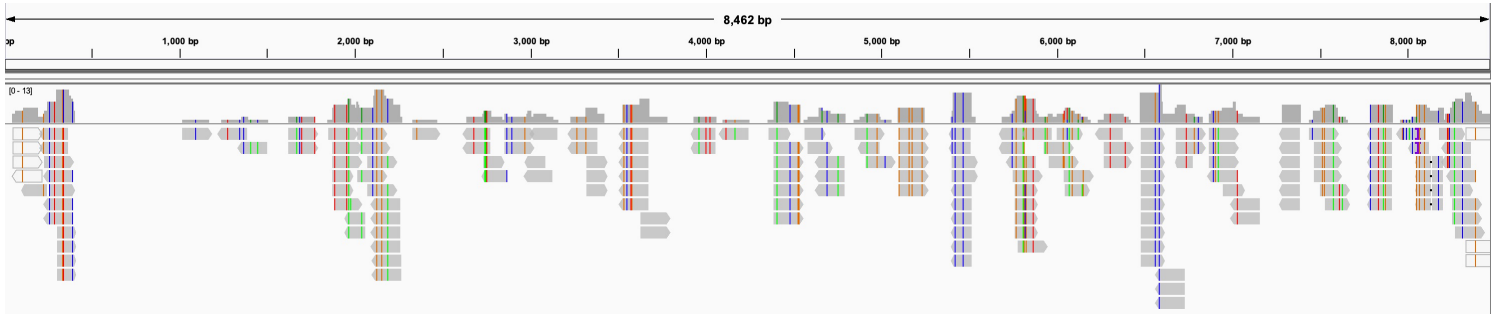
- Great Alphapapillomavirus detection
- HPV^{-} contain *TP53* mutations as expected
- Additional HPV+ samples without diagnostic test result
- More complete approach to HPV stratification
- Alphapapillomavirus reads directly associated with tumour sample
- Clear benefit of read threshold
- Reads > Abundance for presence/absence tests



Genomics England – HTLV-1

- Previously defined 4 viruses to report[3] - HIV, HBV, HCV, HTLV-1
- Benchmarking showed hepatitis filtered
- HTLV-1 in one participant (breast cancer)
- Retested + also identified in germline blood
- Ethnicity may align with HTLV-1 endemic regions
- BLAST confirmed unique matches
- 172 classified reads align across HTLV-1 genome
- Subject to independent validation





3. Magiorkinis, G. *et al.* (2019). *Wellcome Open Research*

Fig: Proietti *et al*. 2005. Oncogene

Can we reliably detect microbes within human tissue whole genome sequences?- Mostly

Can we leverage these incredible datasets to uncover
more about the relationship between microbes and cancer? – Yes

Are there opportunities for translational utility? - Yes

Future Directions

- Definite scope for translational utility
- Preparing tumour community matrix for release
- Great resource for hypothesis generation
- Clinical associations released soon

Acknowledgements

- Big C, Prostate Cancer UK
- Genomics England & Participants
- CRTU teams across UK (NNUH teams in Biorepository, Histopathology, Cancer surgical teams)

Supervisors:
Prof Daniel Brewer
Dr Richard Leggett
Dr Ghanasyam Rallapalli
Dr Rachel Hurst
Prof Colin Cooper