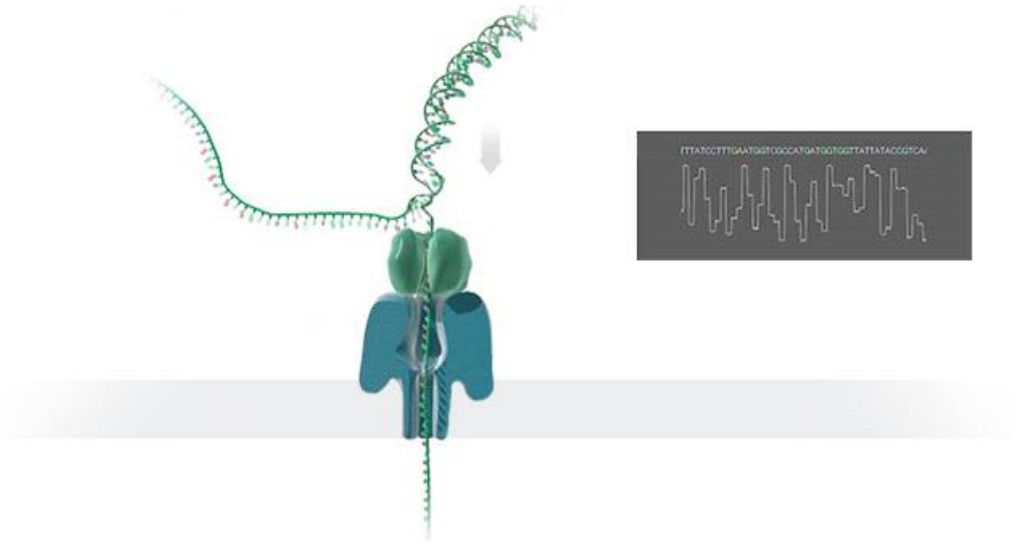
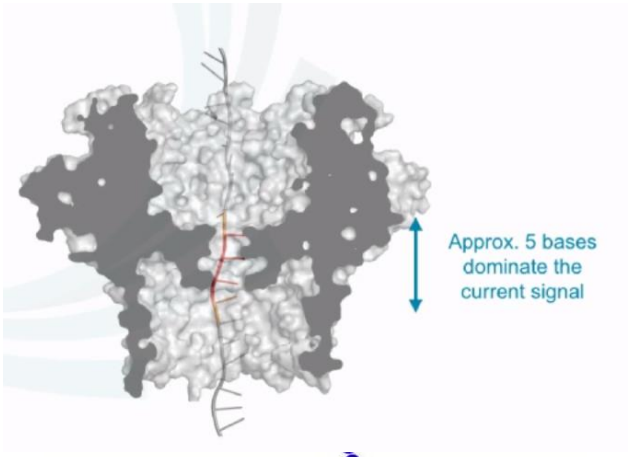
WTAC
RNA Transcriptomics 2019
cDNA sequencing with ONT

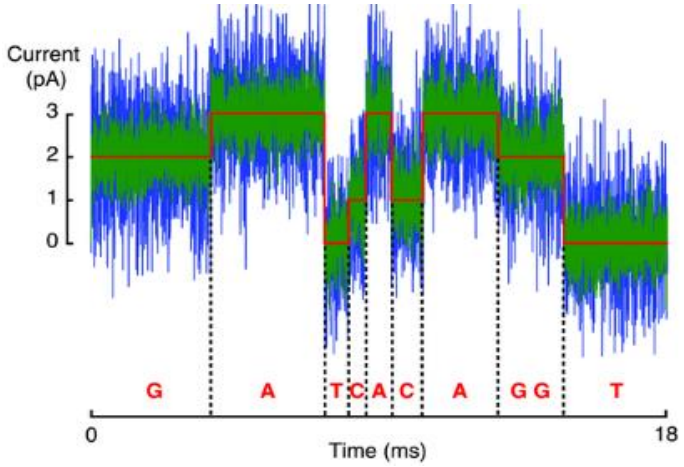
Spyros Oikonomopoulos, Dr
Anthony Bayega, Ph.D Candidate
McGill university

Nanopore sequencing basics

- **Concept**: pass a single strand of DNA through a membrane via a nanopore and apply a voltage difference across the membrane. The nucleotides present in the pore will affect the pore's electrical resistance, so current measurements over time can indicate the sequence of DNA bases passing through the pore. This electrical current signal (a.k.a. the 'squiggle' due to its appearance when plotted







Base-calling

- The computational process of translating raw electrical signal to nucleotide sequence
- Base-calling based on comparing known signals to observed signals
  - Hidden Markov models (HMM)
  - Recurrent neural network (RNN)
- Performance of any particular base-caller is influenced by the data used to train
- not a trivial task;
  - the electrical signals come from single molecules, making for noisy and stochastic data.
  - large number of possible states: 4^5=1024 for a standard four-base model or 5^5=3125 for 5-methylcytosine
- Base-calling accuracy assessed;
  - read accuracy
  - consensus accuracy
- Examples of base-callers
  - ONT: Albacore, Guppy (Flip-flop option), Scrappie (events & raw), and Flappie (CTC decorder)
  - Research base-caller; Chiron
  - Discontinued: Nanonet, DeepNano, basecRAWller
- sss

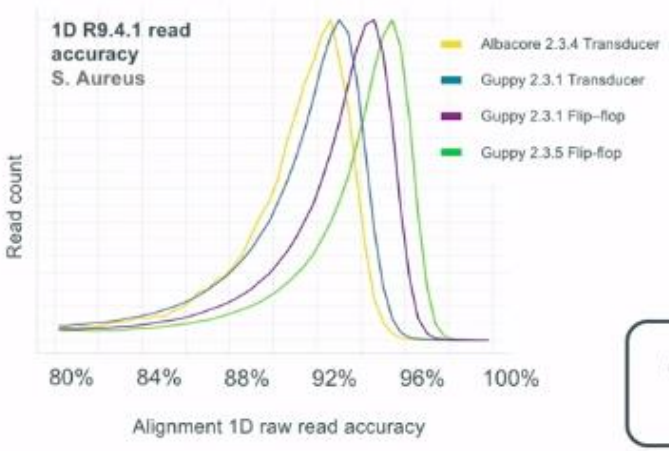
Evolution of ONT base-callers

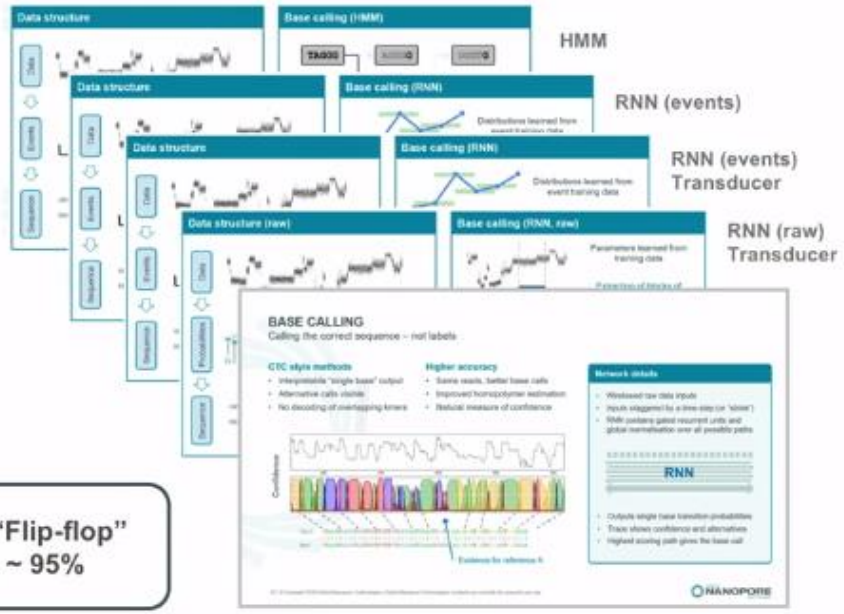
BASE CALLING

Generations of algorithms – 1D raw reads

Strong history of improving accuracy

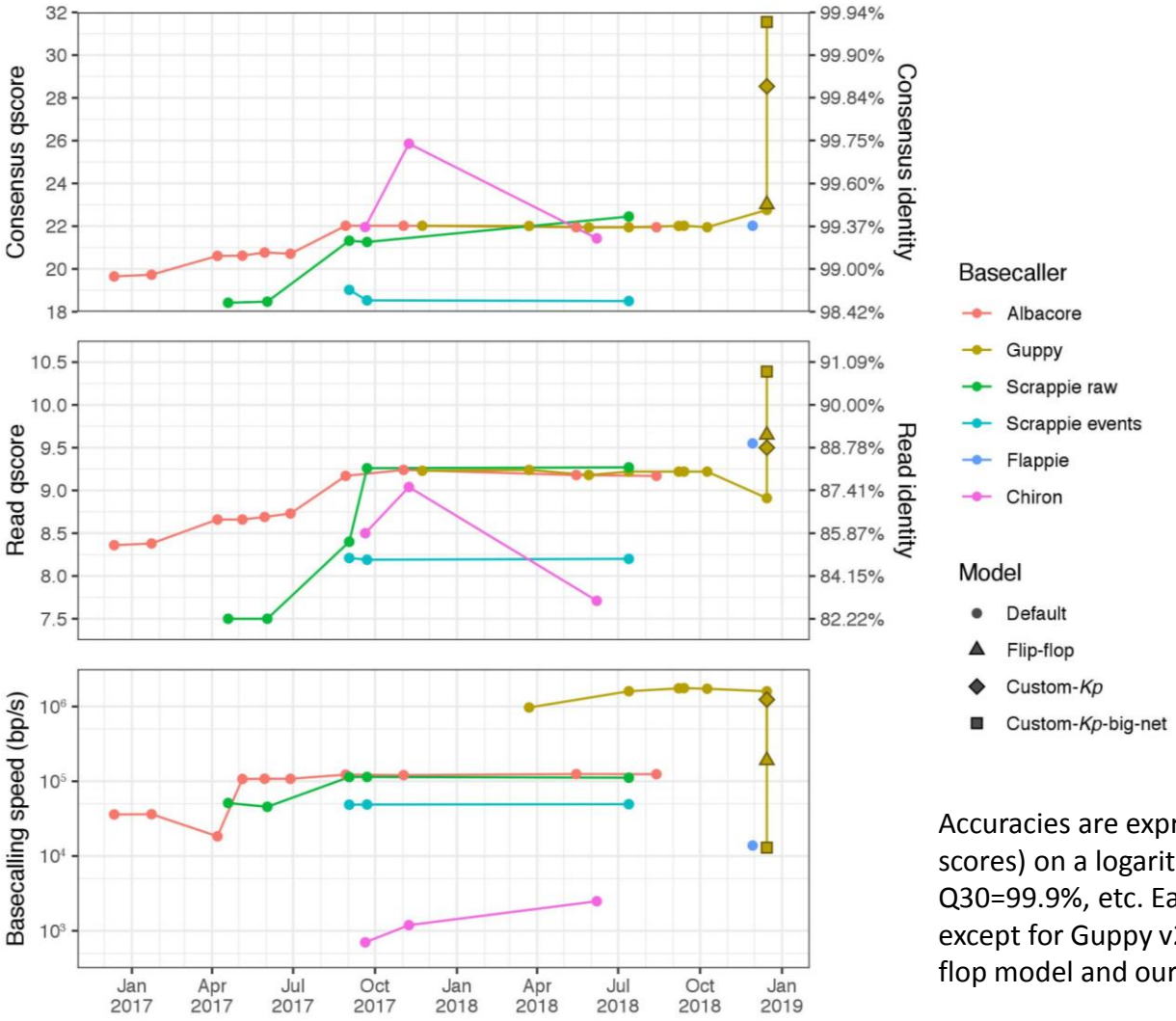
- Algorithms for base calling improving continuously
- Track record of cutting-edge base calling
- New methods can be applied to old “raw” data





Current “Flip-flop” Mode ~ 95%

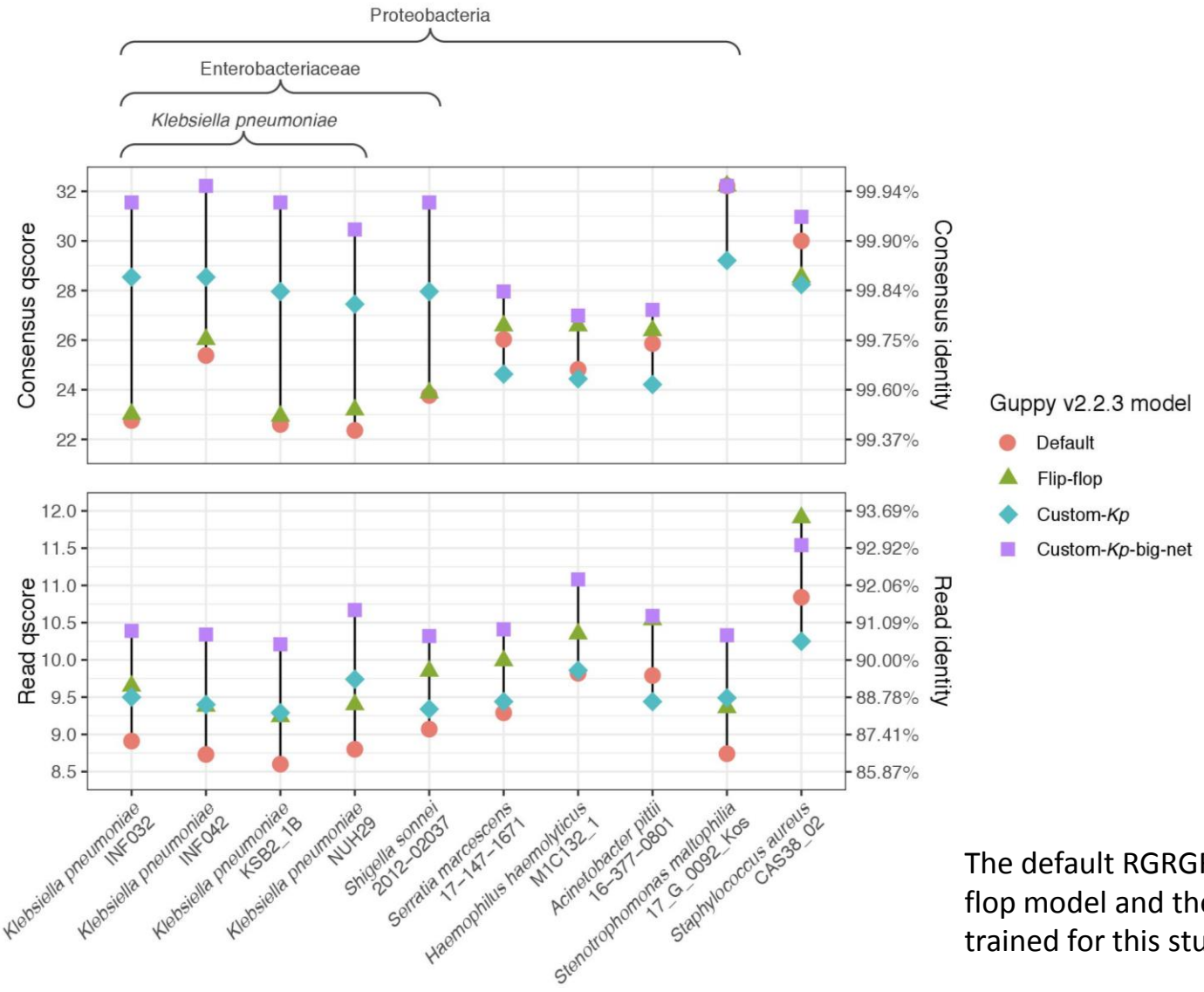
Read accuracy, consensus accuracy and speed performance for basecallers, plotted against the release date



Accuracies are expressed as qscores (also known as Phred quality scores) on a logarithmic scale where Q10=90%, Q20=99%, Q30=99.9%, etc. Each basecaller was run using its default model, except for Guppy v2.2.3 which was also run with its included flip-flop model and our two custom-trained models.

Wick et al., 2019 bioRxiv

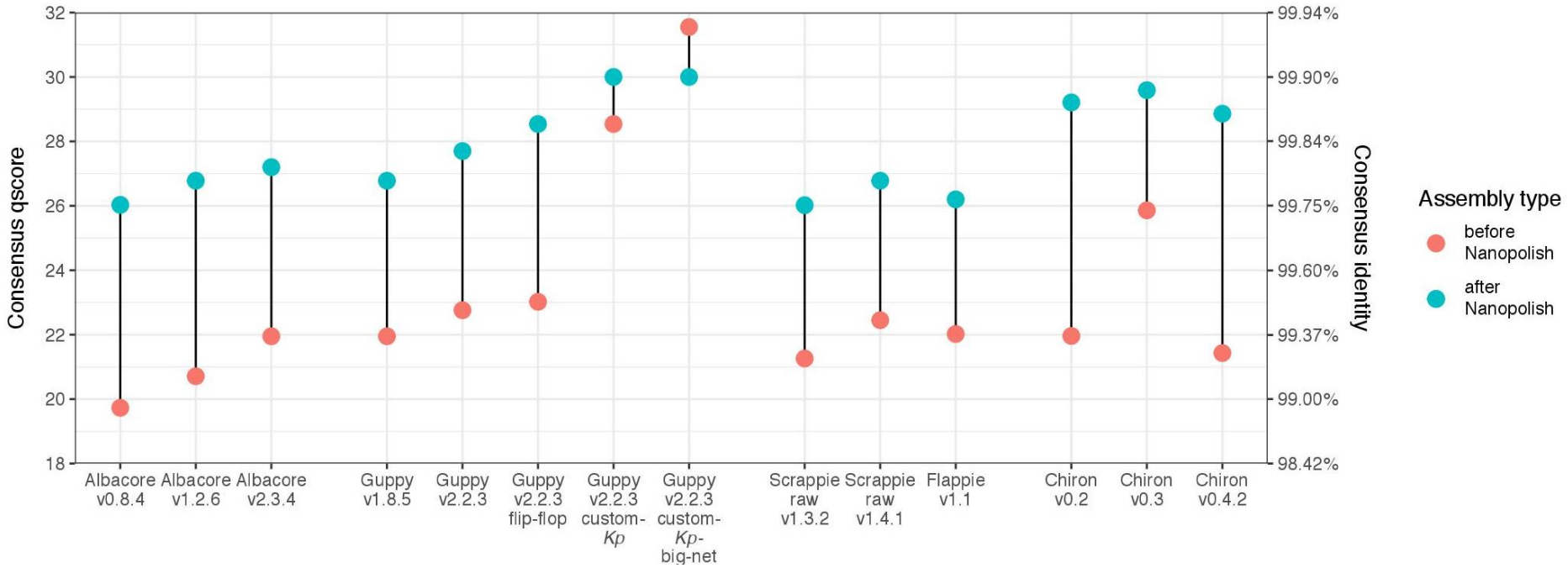
Read and consensus accuracy from Guppy v2.2.3 for a variety of genomes using different models



The default RGRGR model, the included flip-flop model and the two custom models we trained for this study.

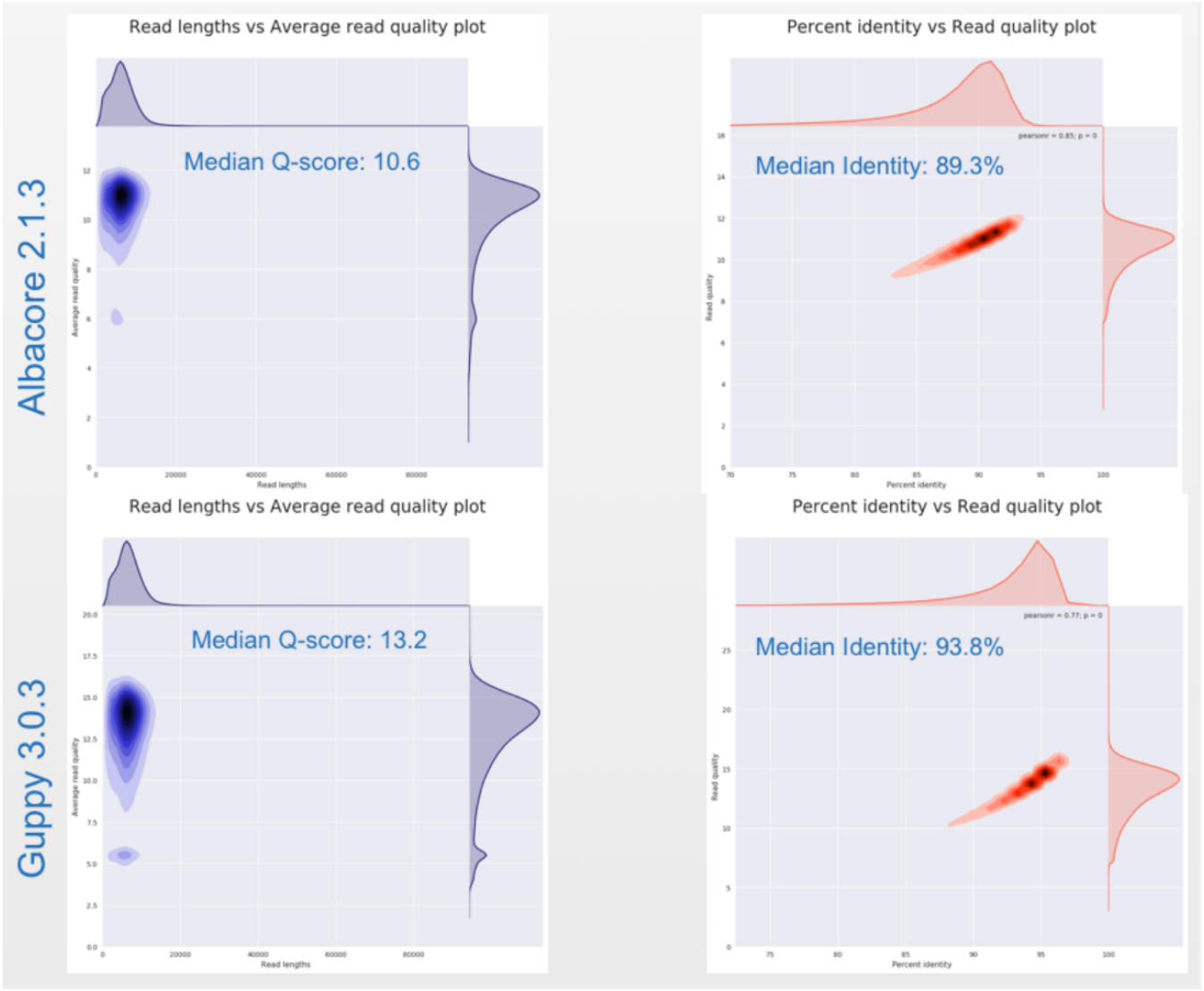
Wick et al., 2019 bioRxiv

Consensus accuracy before (red) and after Nanopolish (blue) for the assemblies of K. pneumoniae benchmarking set.



Wick et al., 2019 bioRxiv

More comparisons



In this lab

- We be basecalling with Guppy using default settings