MinION workshop - What to do with basecalled data

pandoc brachy.md –smart –standalone –bibliography test.bib -o brachy.pdf

So we have some basecalled `fast5` files now! How did this change things?

Lets look at a `fast5` file in `HDFView` to see what has changed

poretools

`poretools` is a toolkit for MinION data that can extract relevant data from `fast5` data into formats you know and love.

- poretools github page
- poretools publication in Bioinformatics
- poretools documentation

poretools provides a simple command-line interface to examine your MinION basecalled data, determine overall quality, extract `fastq` or `fasta` sequences for downstream analysis, and more.

Let's examine how `poretools` acts in an analysis pipeline:

```
# get help on poretools in total:
poretools -h
```

We can easily take a directory of basecalled fast5 files and create a `fastq` file:

```
poretools fastq ./path/to/fast5 > my.fastq
```

Or a `fasta` file:

```
poretools fasta ./path/to/fast5 > my.fasta
```

From here, we have sequence data in formats that are more common to date and can be used with other analysis methods (to be discussed later today!)

`poretools` can also provide information regarding your sequencing stats:

```
poretools stats ./path/to/fast5 stats
```

plots of read size histogram:

```
poretools hist --theme-bw ./path/to/fast5
```

examine the overall yield of reads over time

(remember, the fast5 files keep track of all sorts of metadata for these purposes!):

```
poretools yield_plot --plot-type reads ./path/to/fast5
```

or by total basepairs:

```
poretools yield_plot --plot-type basepairs ./path/to/fast5
```

look at the quality score distribution over read position:

```
poretools qpalpos ./path/to/fast5
```

look at the overall performance of the pores on the flowcell (useful for finding positional / technical artifacts from sequencing)

```
poretools occupancy ./path/to/fast5
```

You can use `poretools` to examine the initial signal traces from a single fast5 file as well using `squiggle`

```
poretools squiggle ./path/to/single.fast5
```

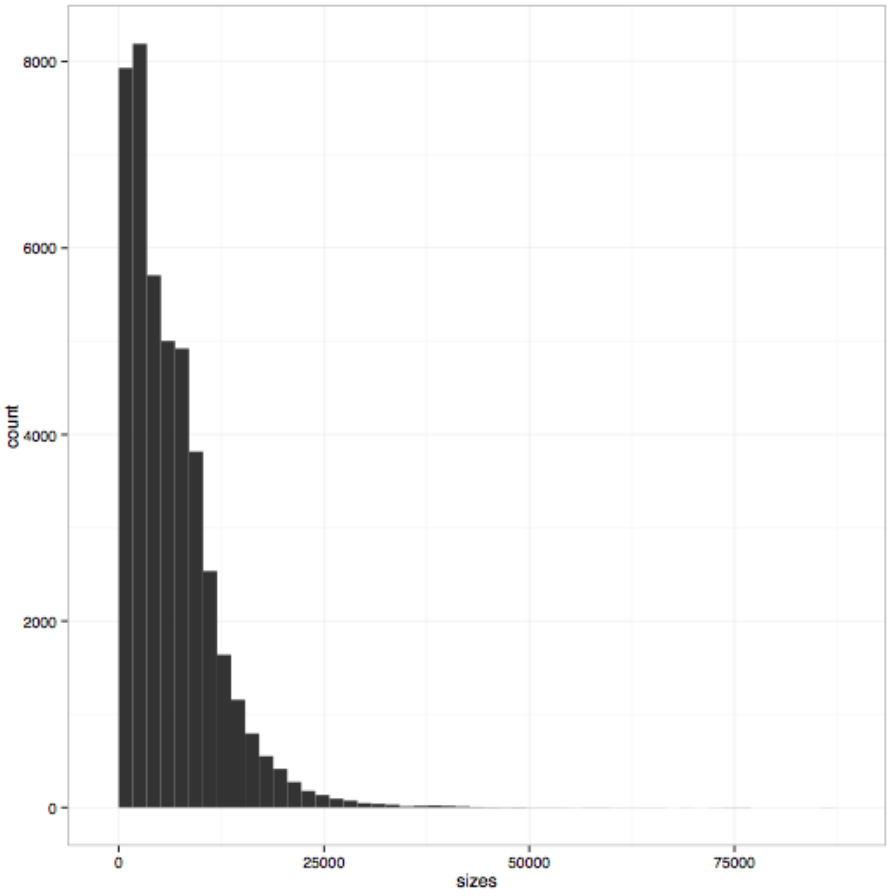

Figure 1: hist from poretools docs

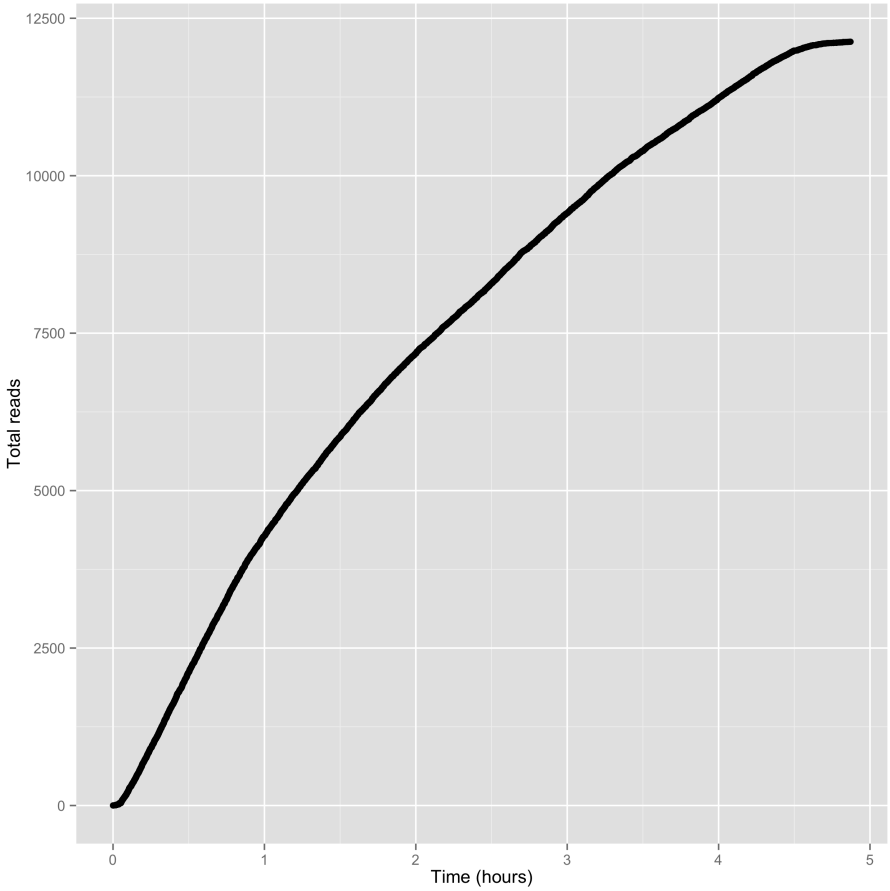

Figure 2: yield plot reads

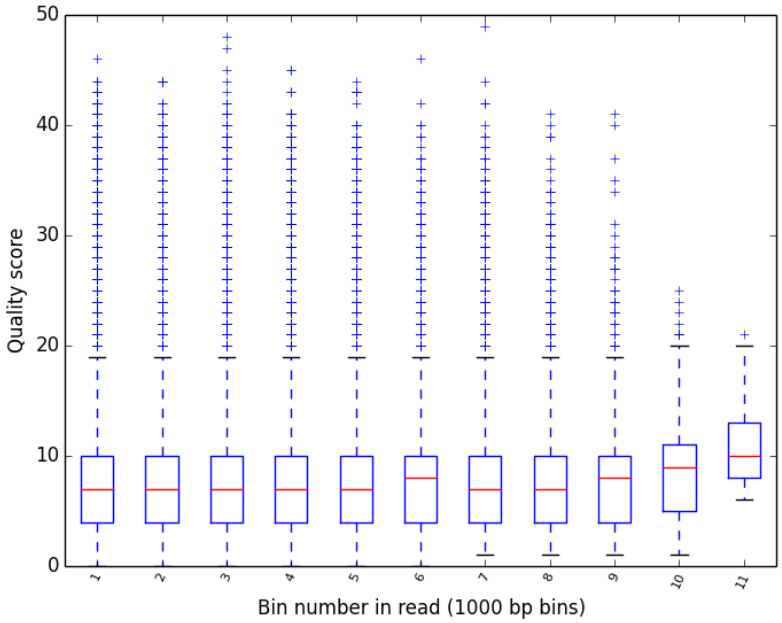

Figure 3: qualpos plot

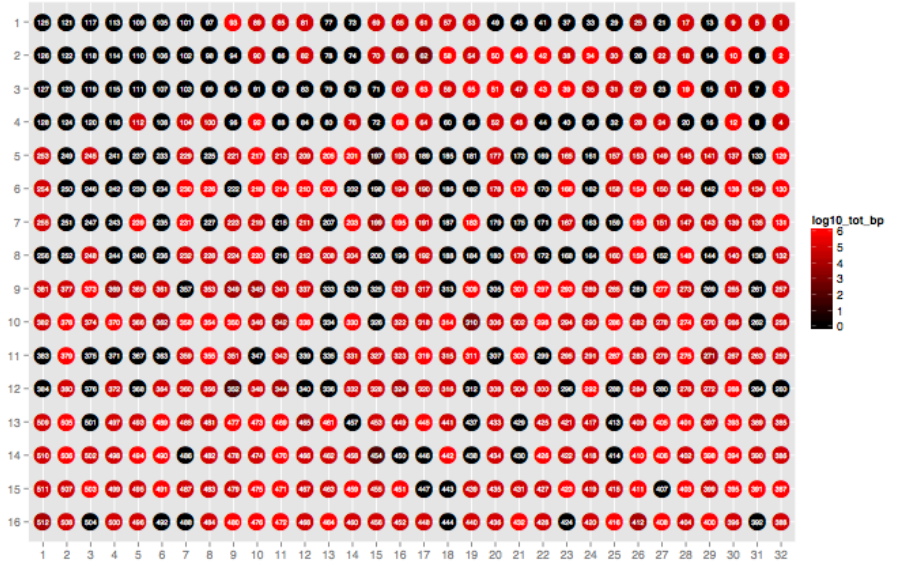

Figure 4: occupancy plot

and most important, report the `winner` (longest read you got!)

```
poretools winner ./path/to/fast5
```

There are a few other commands and options for these to create slightly different plots of to pull additional metrics from your basecalled fast5 files. Take a bit of time to explore our training reads, as well as your first set of real data, today!