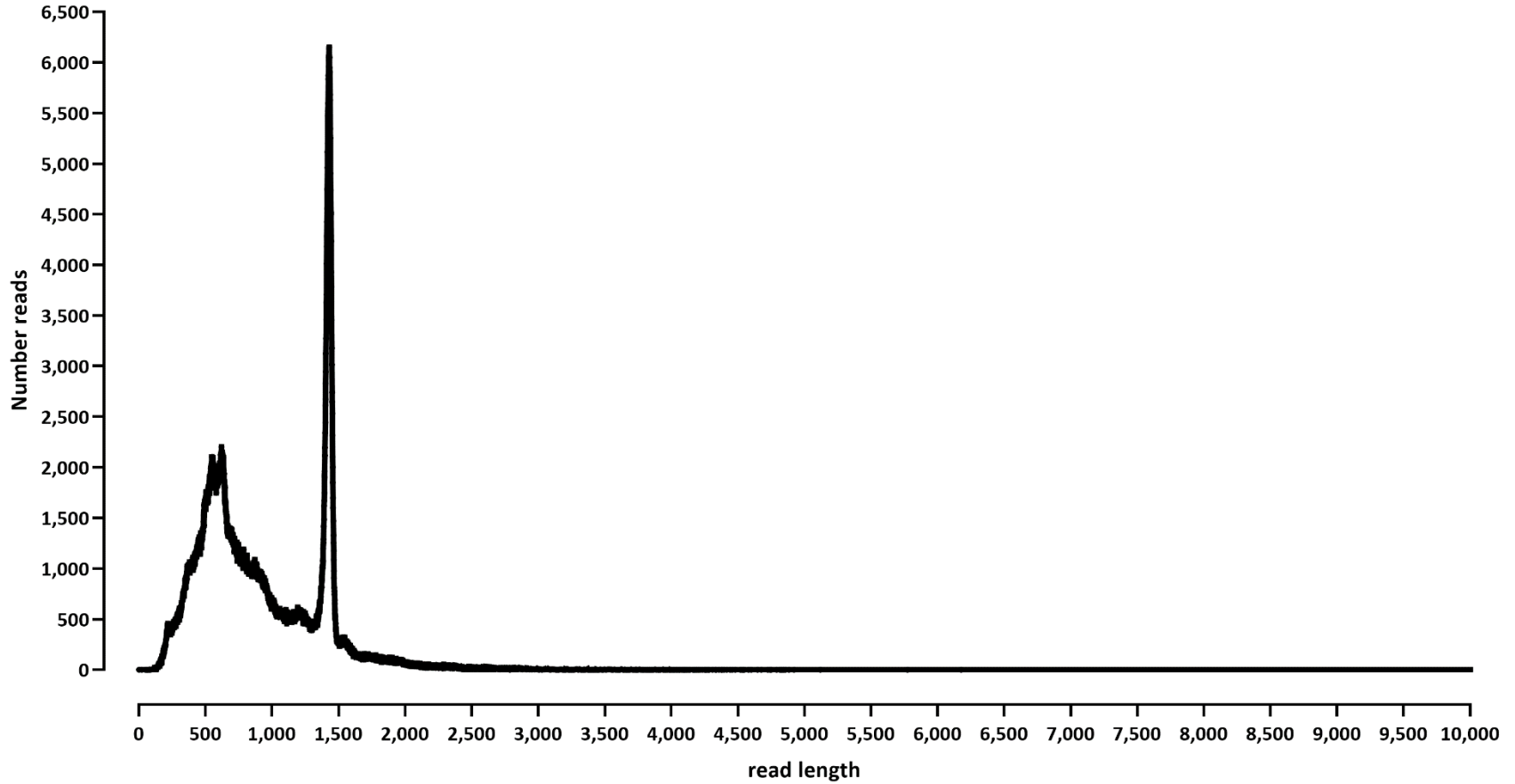


# READ LENGTH DISTRIBUTION

read length distribution



## PIPELINE:

- guppy for rebasecalling
- tailfindr on the Guppy FAST5 files
- Minimap2 alignment against the transcriptome
- tailfindr for annotating tails with transcripts
- Post-processing by me to get TPMs.
  - Note that TPMs include the reads that aligned, but for which there was no tail length output by tailfindr (that was 1.5% of reads)

Note: when counting transcripts, the ENO2 gene came up at 325,861 TPMs, i.e. a third of the reads

The length of its RNA is 1,447 nt in the annotation, so it matches the peak in the read length distribution

TPM vs tail length

