

## Supplementary Information

### **Falco: A quick and flexible single-cell RNA-seq processing framework on the cloud**

Andrian Yang<sup>1,2</sup>, Michael Troup<sup>1</sup> and Joshua WK Ho<sup>1,2\*</sup>

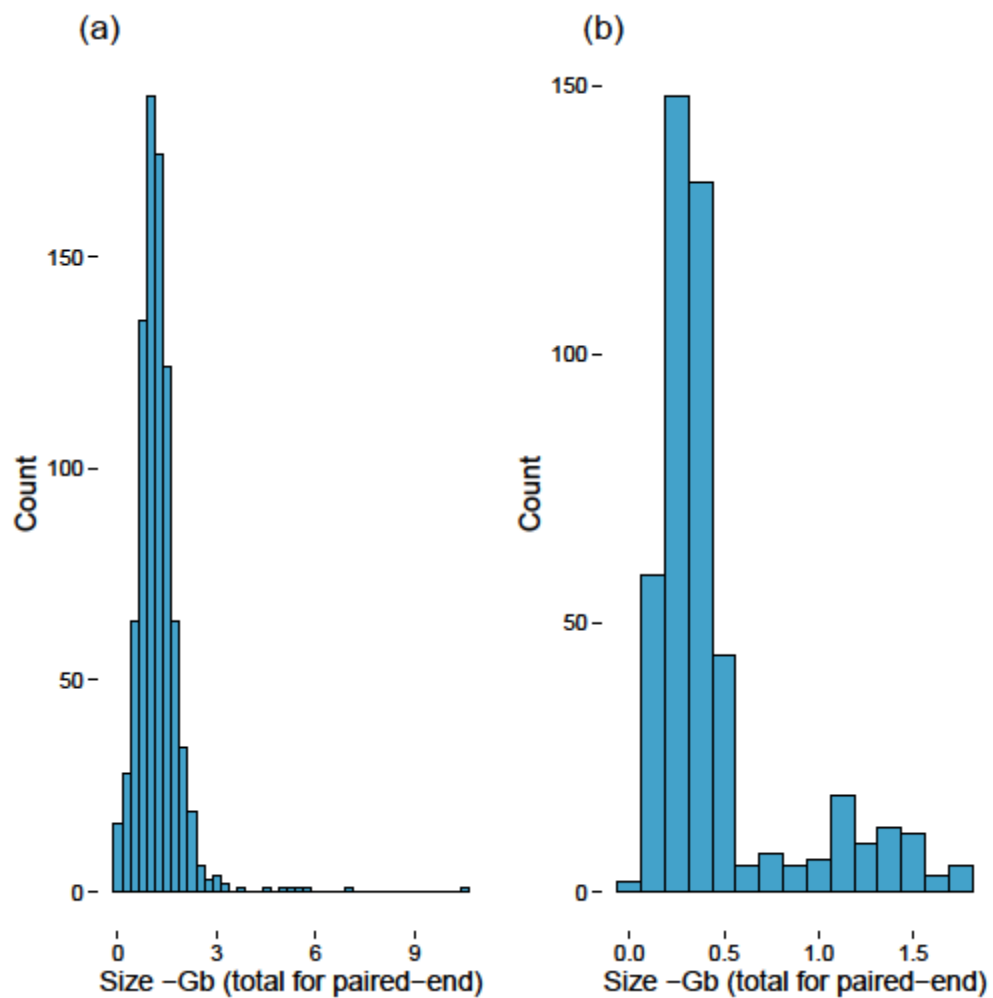
<sup>1</sup>Victor Chang Cardiac Research Institute, Sydney, NSW, Australia

<sup>2</sup>St. Vincent's Clinical School, University of New South Wales, Sydney, NSW, Australia

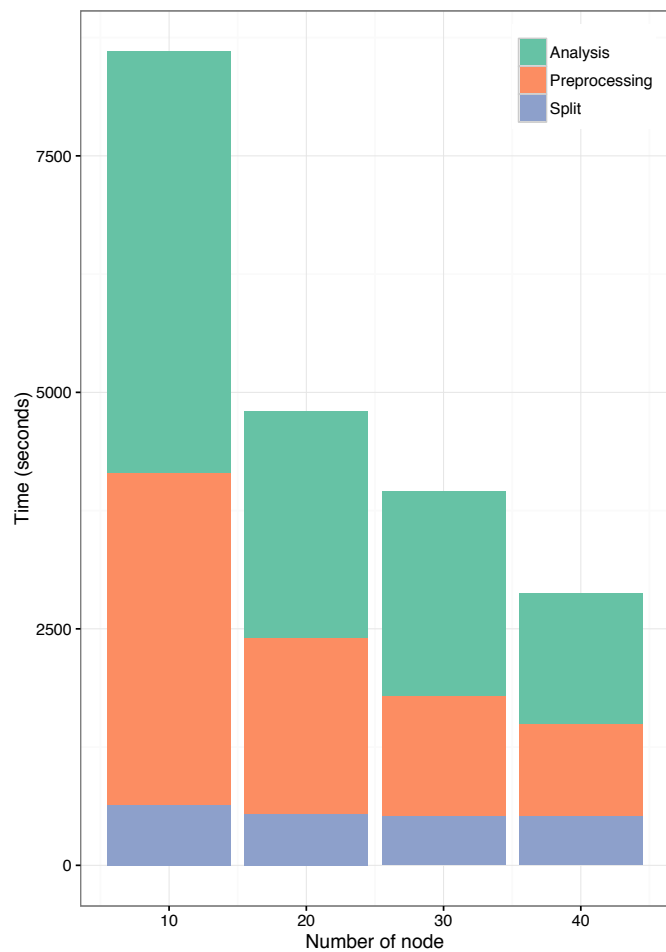
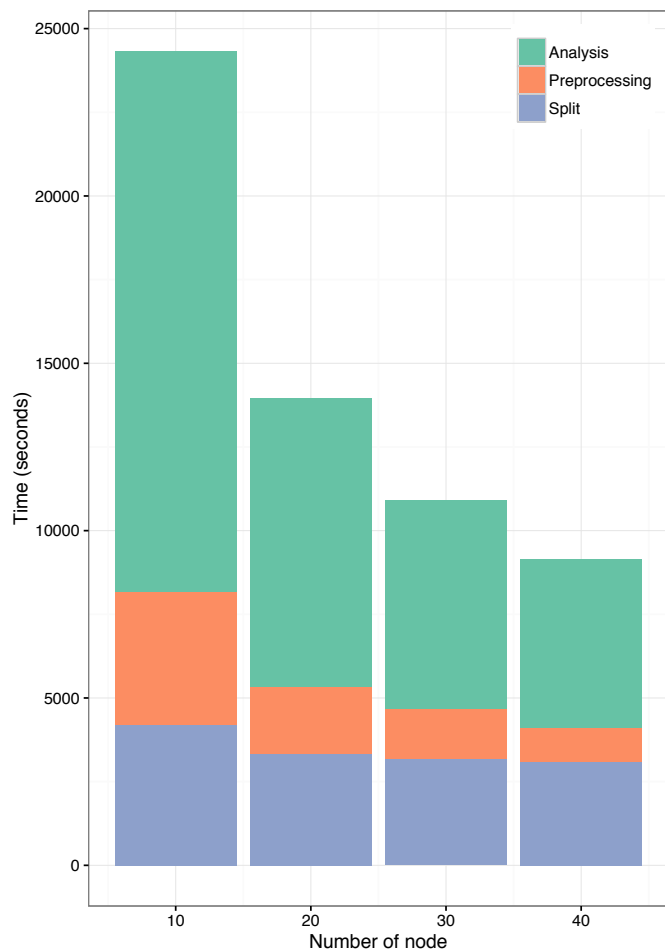
\*To whom correspondence should be addressed: [j.ho@victorchang.edu.au](mailto:j.ho@victorchang.edu.au)

Availability: <https://github.com/VCCRI/Falco>

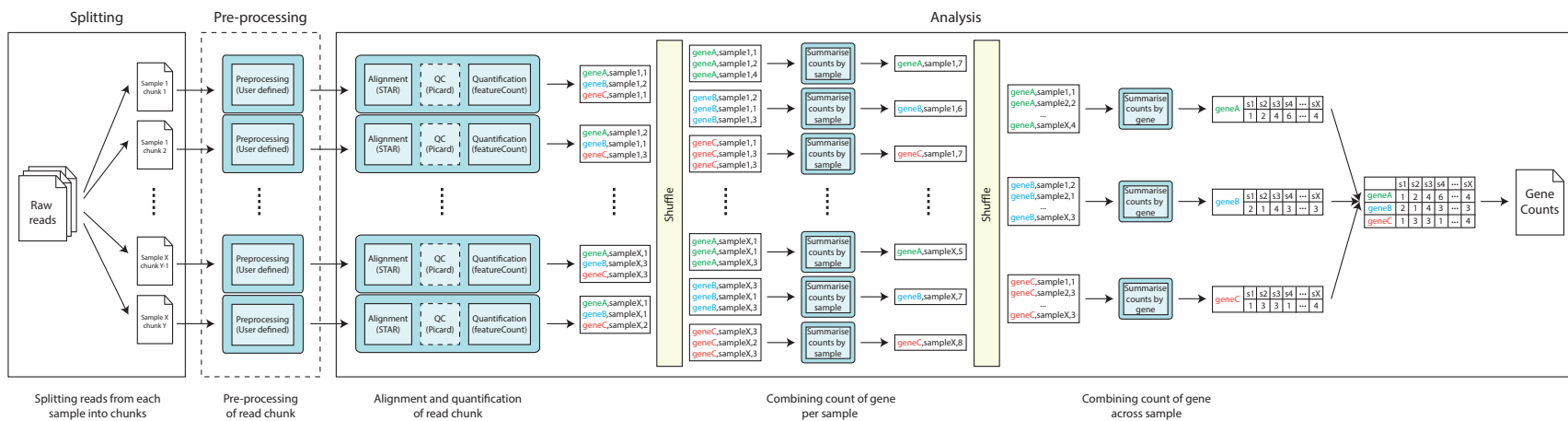
---



**Supplementary Figure 1.** Size of the FASTQ file (in terms of gigabytes) of individual cells in each single-cell RNA-seq data set. **(a)** Mouse embryonic stem cells (SRA accession: [ERP005988](https://www.ncbi.nlm.nih.gov/sra/ERP005988)). **(b)** Human brain cells (SRA accession: [SRP057196](https://www.ncbi.nlm.nih.gov/sra/SRP057196)).



**Supplementary Figure 2.** Falco processing time split by steps for the Mouse **(a)**, and Human **(b)** data sets



**Supplementary Figure 3.** The system architecture of the Falco framework. In the splitting step, reads from one or multiple FASTQ files are split into multiple chunks of size 256 Mb uncompressed. This step makes use of Apache Hadoop MapReduce. A pre-processing step is executed if the data require pre-processing using MapReduce. In the main analysis step, sequence alignment and gene expression quantification are carried out in a highly parallelised fashion using the Apache Spark framework.