Methods

A) The reads were trimmed using Cutadapt v2.3 (Martin, 2011).
B) FastQC v0.11.8 was used to ensure the quality of data (Andrews, 2010).
C) Fastq Screen v0.13.0 was used to screen for various types of contamination (Wingett and Andrews, 2018).
D) Reads were mapped to the reference genome GRCm38 (ENSEMBL), using STAR v2.7.8a (Dobin *et al.*, 2013) and assigned count estimates to genes with RSEM v1.3.3 (Li and Dewey, 2011). Alignment options followed ENCODE standards for RNA-seq.
E) Multiqc v1.7 compiled the results from several of these tools and provided a detailed and comprehensive quality control report (Ewels *et al.*, 2016).

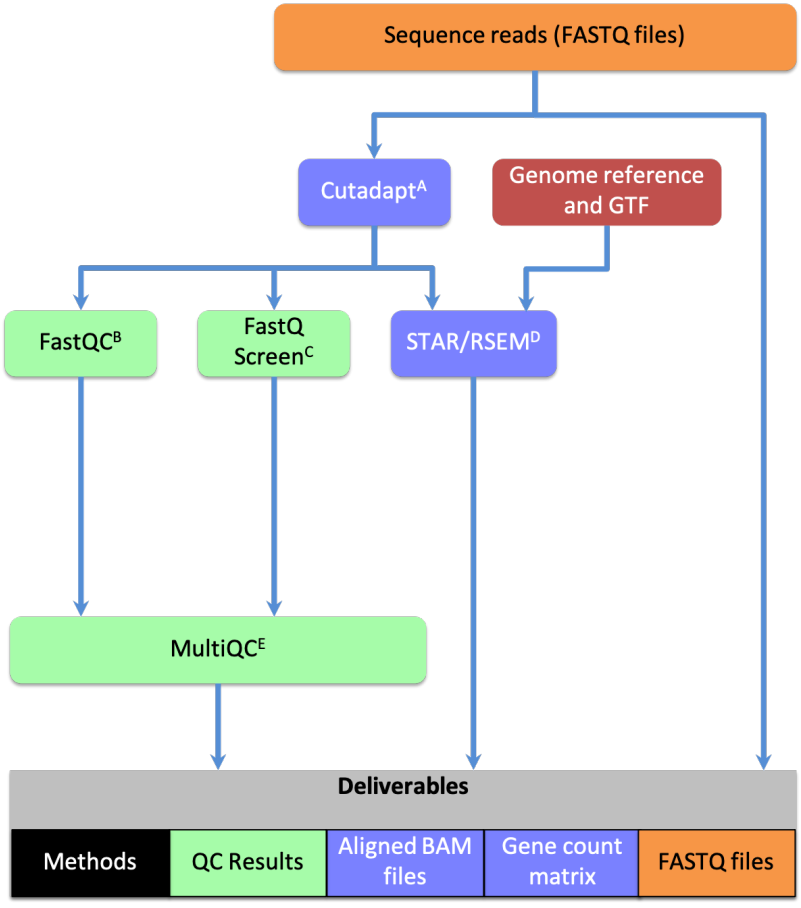

Figure 1: **Figure:** Methods overview

Andrews,S. (2010) FastQC: A quality control tool for high throughput sequence data.
Dobin,A. *et al.* (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
Ewels,P. *et al.* (2016) MultiQC: Summarize analysis results for multiple tools and samples in a single report.

Bioinformatics, **32**, 3047.

Li,B. and Dewey,C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*.

Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.

Wingett,S.W. and Andrews,S. (2018) Fastq screen: A tool for multi-genome mapping and quality control [version 1; referees: 3 approved, 1 approved with reservations]. *F1000Research*, **7**, 1–13.