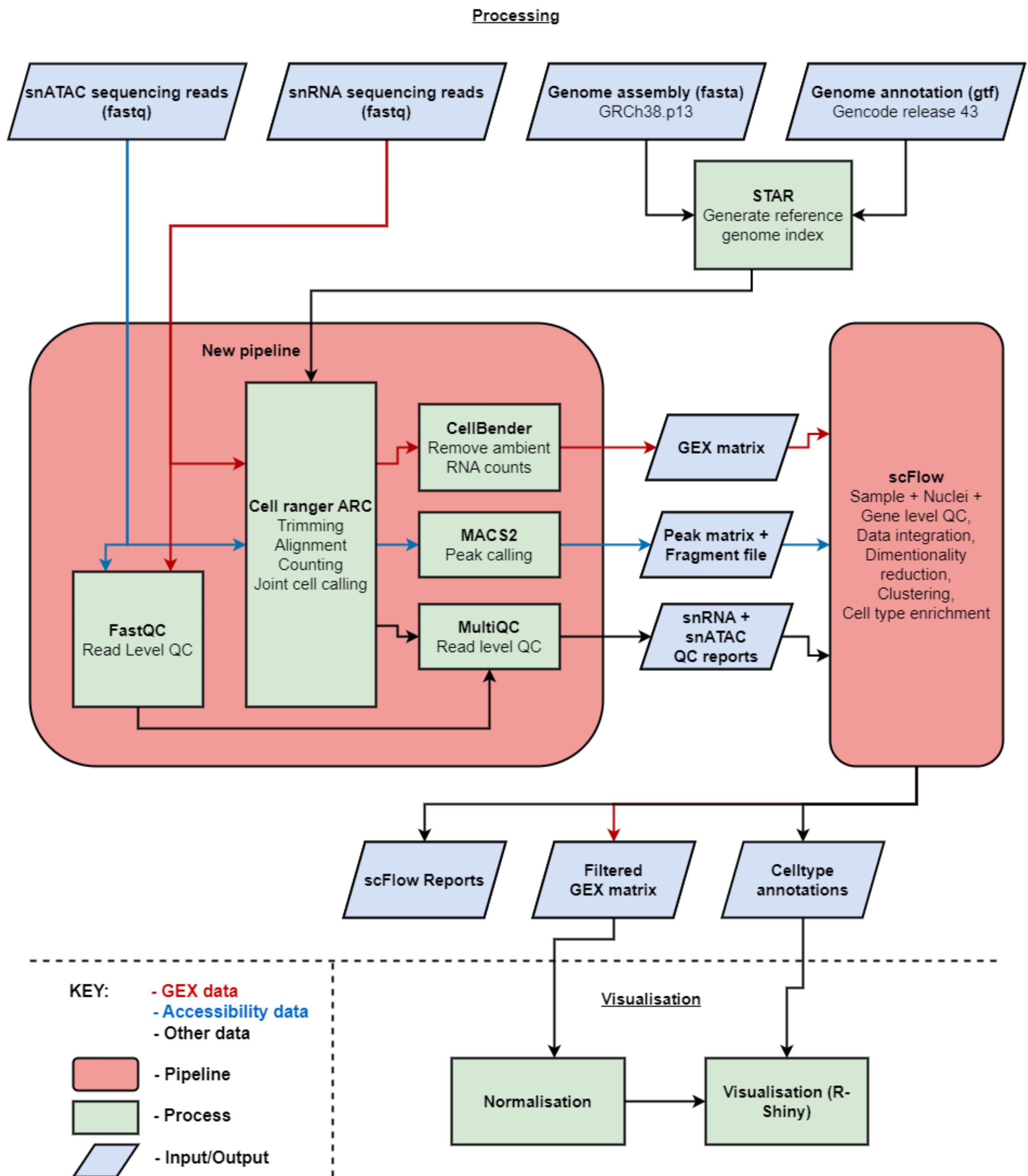
Imperial College London

MULTI-'OMICS ATLAS PROJECT

March 2023

Standard operating procedure: Single Nuclei Multiomics (snATACseq + snRNAseq) – Processing and Visualisation

PROTOCOL SCHEMATIC



Written by: Nurun Fancy & Michael Thomas

PROTOCOL

Processing

Inputs

- Reference Genome built from GENCODE release 43 Basic CHR of the NCBI GRCh38.p13 Assembly (corresponding to Ensembl release 108).
- Fastq sequencing files for snATACseq and snRNAseq.

Process

1. Each Fastq file undergo read QC with **FastQC** and reports are saved as .html files.
2. Sequencing reads are mapped with **cellranger-arc** and a joint snATACseq and snRNAseq cell calling is performed.
3. FastQC reports are collected via **MultiQC**.
4. Peak calling is performed on aligned **.bam** files from snATACseq using **MACS2**.
5. Raw snRNAseq feature-barcode matrices are processed by **CellBender** to remove ambient RNA counts.
6. Output from CellBender is further processed by **nf-core/scFlow** pipeline.

A nextflow based pre-processing pipeline "X" is used for step 1-5 (based on FastQC, cellranger-arc, MultiQC, CellBender and MACS2).

Parameters files are provided for a full list of parameters used.

Outputs (snRNAseq)

- MultiQC report.
- Filtered feature-barcode matrices per sample.
- scFlow reports for sample QC, integration, dimensionality reduction and cell-type annotation and a final feature-barcode matrix with cell-type annotation.

Outputs (snATACseq)

- MultiQC report.
- ATACseq peak fragments.tsv files.

Data quality standards (snRNAseq)

- Nuclei with the following QC measures are included: Total UMI count > 400 and < 40000; total expressed features > 200 and < 8000; mitochondrial read < 5%.
- Any samples with less than 1000 nuclei are excluded.

Data quality standards (snATACseq)- **optional for March deliverable**

- Nuclei with the following QC measures are included: Total number of fragments in peak > 3000 and < 20000; nucleosome signal < 4; TSS.enrichment > 2; Fraction of fragments in peaks > 15; ratio reads in genomic blacklist regions from the ENCODE project < 0.05.
- Any samples with less than 1000 nuclei are excluded.

Visualisation (snRNAseq)

Written by: Nurun Fancy & Michael Thomas

- Data is $\log_2(x + 1)$ normalised for individual gene expression visualisation via Shiny-app.