**Methods:**

The scatacseqflow pipeline consists of 2 major functional module groups: preprocess and downstream analysis (with ArchR).

For preprocess, it supports 3 options: 10xgenomics, Biorad, and default. The implemented modules will be listed below with the first appearance in bold (some modules may be re-used across different option).

Specifically, the 10xgenomics option consists of the following modules:

1. **GET\_10XGENOMICS\_FASTQ**: used to prepare sample fastq by parsing the sample\_sheet.csv file and concatenating data from different lane into one single file.
2. **DOWNLOAD\_FROM\_UCSC**: used to download genome file given ucsc genome name.
3. **GET\_PRIMARY\_GENOME**: extract the primary genome from downloaded ucsc geome.
4. **DOWNLOAD\_FROM\_UCSC\_GTF**: used to download ucsc gtf file given ucsc genome name.
5. **FIX\_UCSC\_GTF**: used to repair the problematic gtf files.
6. **DOWNLOAD\_FROM\_ENSEMBL:** used to download genome file from ensemble given ensemble genome name.
7. **DOWNLOAD\_FROM\_ENSEMBL\_GTF**: used to download ensembl gtf file given ensembl genome name.
8. **CELLRANGER\_INDEX**: create index used for cellranger.
9. **CELLRANGER\_ATAC\_COUNT**: perform scATAC-seq analysis with cellranger\_atac software.

The Biorad scATAC-seq analysis tools come with numerous Docker containers, and in order to be able to use singularity in our pipeline, we rebuilt the images. The modules are listed below:

1. **GET\_BIORAD\_FASTQ**: used to prepare fastq into Biorad-compatible format.
2. **BIORAD\_FASTQC**: wrapper of the fastqc module from Biorad scATAC-seq analysis software.
3. **BIORAD\_ATAC\_SEQ\_DEBARCODE**: wrapper of the debarcode module from Biorad scATAC-seq analysis software.
4. **BIORAD\_ATAC\_SEQ\_TRIM\_READS**: wrapper of the trim\_reads module from Biorad scATAC-seq analysis software.
5. **BIORAD\_ATAC\_SEQ\_BWA**: wrapper of the bwa module from Biorad scATAC-seq analysis software.
6. **BIORAD\_ATAC\_SEQ\_ALIGNMENT\_QC**: wrapper of the alignment\_qc module from Biorad scATAC-seq analysis software.
7. **BIORAD\_ATAC\_SEQ\_FILTER\_BEADS (planning)**: wrapper of the filter\_beads module from Biorad scATAC-seq analysis software.
8. **BIORAD\_ATAC\_SEQ\_BEADS\_DECONVOLUTE (planning)**: wrapper of the beads\_deconvolute module from Biorad scATAC-seq analysis software.
9. **BIORAD\_ATAC\_SEQ\_CELL\_FILTER (planning)**: wrapper of the cell\_filter module from Biorad scATAC-seq analysis software.

The default option uses the following modules:

1. GET\_10XGENOMICS\_FASTQ
2. **FASTQC**: wrapper around FastQC.
3. **ADD\_BARCODE\_TO\_READS**: custom Python code that leverages sinto add barcode.
4. **CORRECT\_BARCODE**: custom R code to correct barcode.
5. **MATCH\_READS**: custom Python code that leverages seqkit to match read pairs.
6. **CUTADAPT**: wrapper around cutadapt.
7. **BWA\_INDEX**: wrapper around bwa.
8. **BWA\_MAP**: wrapper around bwa.
9. **MINIMAP2\_INDEX**: wrapper around minimap2.
10. **MINIMAP2\_MAP**: wrapper around minimap2.
11. **QUALIMAP**: wrapper around qualimap.
12. DOWNLOAD\_FROM\_UCSC
13. GET\_PRIMARY\_GENOME
14. DOWNLOAD\_FROM\_ENSEMBL

For downstream analysis with ArchR, it currently supports the following modules:

1. **ARCHR\_CREATE\_ARROWFILES**: wrapper to create ArchR arrowfiles.
2. **ARCHR\_ADD\_DOUBLETSCORES**: wrapper to detect doublet.
3. **ARCHR\_ARCHRPROJECT**: wrapper to create archrproject object.
4. **ARCHR\_ARCHRPROJECT\_QC**: wrapper to perform qc on archrproject.
5. **ARCHR\_DIMENSION\_REDUCTION**: wrapper to perform dimension reduction.
6. **ARCHR\_BATCH\_CORRECTION**: wrapper to perform batch correction.
7. **ARCHR\_CLUSTERING**: wrapper to perform clustering.
8. **ARCHR\_EMBEDDING**: wrapper to perform single cell embedding.
9. **ARCHR\_MARKER\_GENE**: wrapper to find marker genes.
10. **ARCHR\_PSEUDO\_BULK\_CLUSTERS**: wrapper to find pseudo bulks based on scATAC-seq data.
11. **ARCHR\_PSEUDO\_BULK\_CLUSTERS2**: wrapper to find pseudo bulks based on scRNA-seq data.
12. **ARCHR\_SCRNASEQ\_UNCONSTRAINED**: wrapper to perform unconstrained integration of scRNA-seq data.
13. **ARCHR\_SCRNASEQ\_CONSTRAINED**: wrapper to perform constrained integration of scRNA-seq data.
14. **ARCHR\_CALL\_PEAKS\_CLUSTERS**: wrapper to call peaks based on scATAC-seq clusterings.
15. **ARCHR\_CALL\_PEAKS\_CLUSTERS2**: wrapper to call peaks based on scRNA-seq clusterings.
16. **ARCHR\_GET\_MARKER\_PEAKS\_CLUSTERS**: wrapper to get marker peaks based on scATAC-seq data.
17. **ARCHR\_GET\_MARKER\_PEAKS\_CLUSTERS2**: wrapper to get marker peaks based on scRNA-seq data.
18. **ARCHR\_MARKER\_PEAKS\_IN\_TRACKS\_CLUSTERS**: wrapper to visualize marker peaks for scATAC-seq data.
19. **ARCHR\_MARKER\_PEAKS\_IN\_TRACKS\_CLUSTERS2**: wrapper to visualize marker peaks for scRNA-seq data.
20. **ARCHR\_PAIRWISE\_TEST\_CLUSTERS**: wrapper to perform pairwise test based on scATAC-seq data.
21. **ARCHR\_PAIRWISE\_TEST\_CLUSTERS2**: wrapper to perform pairwise test based on scRNA-seq data.
22. **ARCHR\_MOTIF\_DEVIATIONS\_CLUSTERS**: wrapper to find motif deviations based on scATAC-seq data.
23. **ARCHR\_MOTIF\_DEVIATIONS\_CLUSTERS2**: wrapper to find motif deviations based on scRNA-seq data.
24. **ARCHR\_COACCESSIBILITY\_CLUSTERS**: wrapper to detect co-accessibility using scATAC-seq data.
25. **ARCHR\_COACCESSIBILITY\_CLUSTERS2**: wrapper to detect co-accessibility using scRNA-seq data.
26. **ARCHR\_PEAK2GENELINKAGE\_CLUSTERS2**: wrapper to find peak2genelinkage using scRNA-seq data.
27. **ARCHR\_GET\_POSITIVE\_TF\_REGULATOR\_CLUSTERS**: wrapper to find positive TF regulators using scATAC-seq data.
28. **ARCHR\_GET\_POSITIVE\_TF\_REGULATOR\_CLUSTERS2**: wrapper to find positive TF regulators using scRNA-seq data.
29. **ARCHR\_TRAJECTORY\_CLUSTERS2**: predict trajectory using scRNA-seq data.