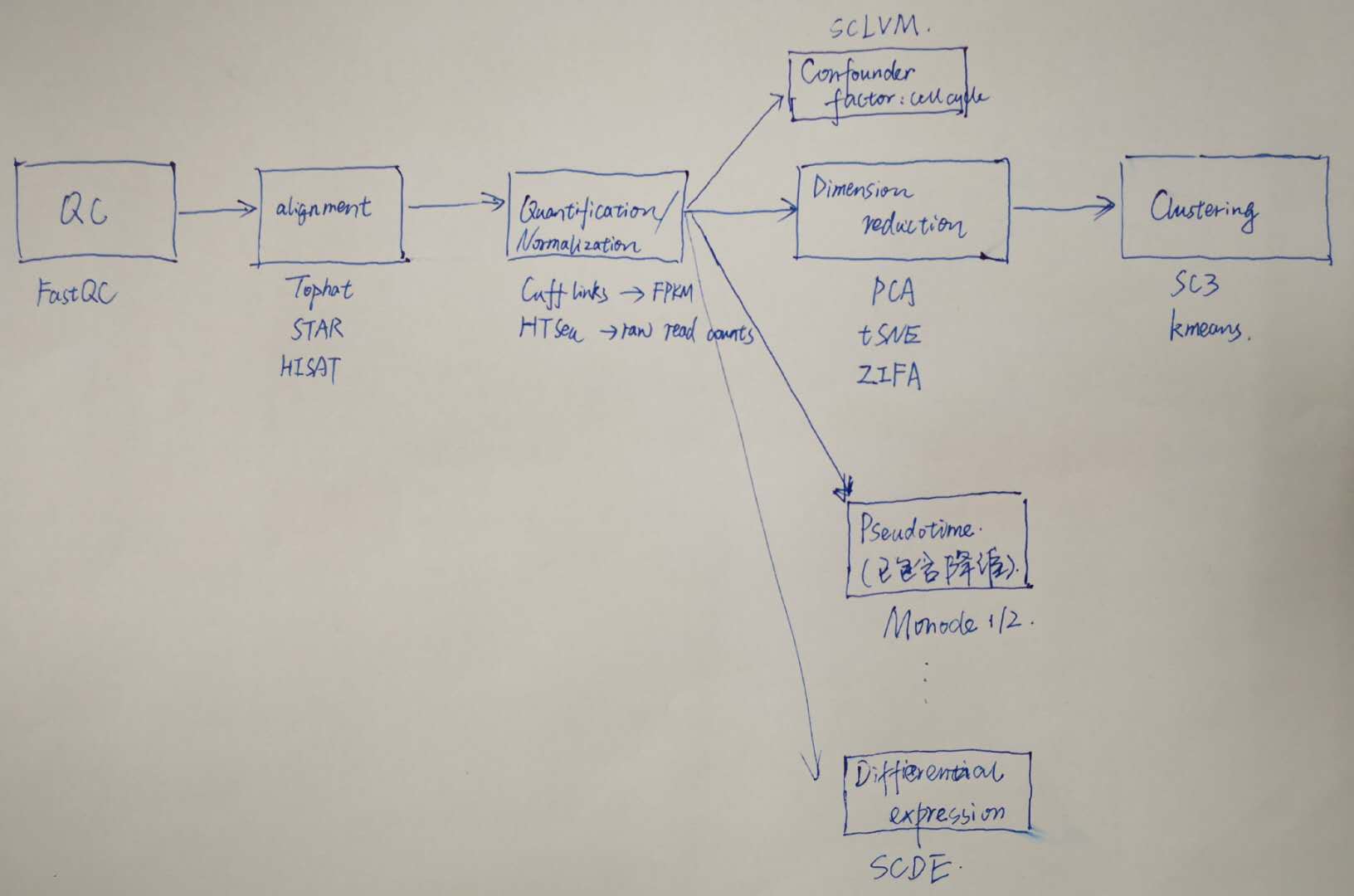
**scRNA-seq data processing steps**

by 李翔宇、苗准



**scATAC-seq data (from Greenleaf lab) processing steps**

by 张威、魏征

|  |  |  |
| --- | --- | --- |
| Steps | Processing | software |
| 1 | Quality Control(GC, adapters) | FASTQC (Java; windows, MAC, Linux) |
| 2 | trimming adapters (detection and trimming) | AdapterRemoval (c++; MAC, Linux) |
| 3 | Mapping | Bowtie2 (c++; windows, MAC, Linux) |
| 4 | Remove Duplicates | Picard (Java; windows, MAC, Linux) |
| 5 | Filter Reads (remove low quality and chrM, Y and unmapped reads) | Scripts |
| 6 | Peak Calling and filter peak (using ENCODE blacklist and custom blacklist) | MACS2 (py, c++; windows, MAC, Linux) |
| 7 | Compute Accessibility Peaks (peak summits were extended +/-250 bps) | Scripts |
| 8 | Fragment Shift (+4 -5) and Tn5 insertion score | Scripts |
| 9 | Deviation Inference (whether ATAC-seq signal varies from cell to cell) | Scripts (algorithms provided) |
| 10 | Genome Annotation (TF, histone, DNA binding motifs, chromatin state) | ENCODE data is required, MEME(c++; linux, MAC) |
| 11 | Inferring TF associated accessibility | chromVAR (R; windows, MAC, Linux) |
| 12 | Regulome Analysis | SCRAT (R; windows, MAC, Linux) |
| 13 | Sub-population Clustering | scABC(R; windows, MAC, Linux) |

Yellow means the essential steps.

**scATAC-seq data (from Shendure lab) processing steps**

|  |  |  |
| --- | --- | --- |
| Steps | Processing | software |
| 1 | Barcode pre-processing (no matching) | Scripts |
| 2 | trimming adapters (detection and trimming) | Trimmomatic (java; MAC, Linux) |
| 3 | Mapping | BWA (c++, zlib; windows, MAC, Linux) |
| 4 | Filter Reads (remove low quality and chrM) | Scripts |
| 5 | Remove Duplicates | Picard (Java; windows, MAC, Linux) |
| 6 | Calculating the barcode collision rate | Scripts |
| 7 | Determining accessible hypersensitive sites in single cells (using data in ENCODE) | Scripts, python, Bedtools, Hotspot, F-seq |
| 8 | Comparing with bulk ATAC-seq |  |
| 9 | Evaluating the complexity of sequenced libraries (PCR duplication rates and library complexity, etc.) | Picard (Java; windows, MAC, Linux) |
| 10 | Identifying DHSs differentially accessible between cell types | VGAM (R; windows, MAC, Linux) |
| 11 | Identifying enriched annotations for differentially accessible DHSs (find related gene) | BedTools, ENCODE data, Piano (R), |
| 12 | Dimensionality reduction of chromatin accessibility data | R (several algorithm provided) |

Yellow means the essential steps.