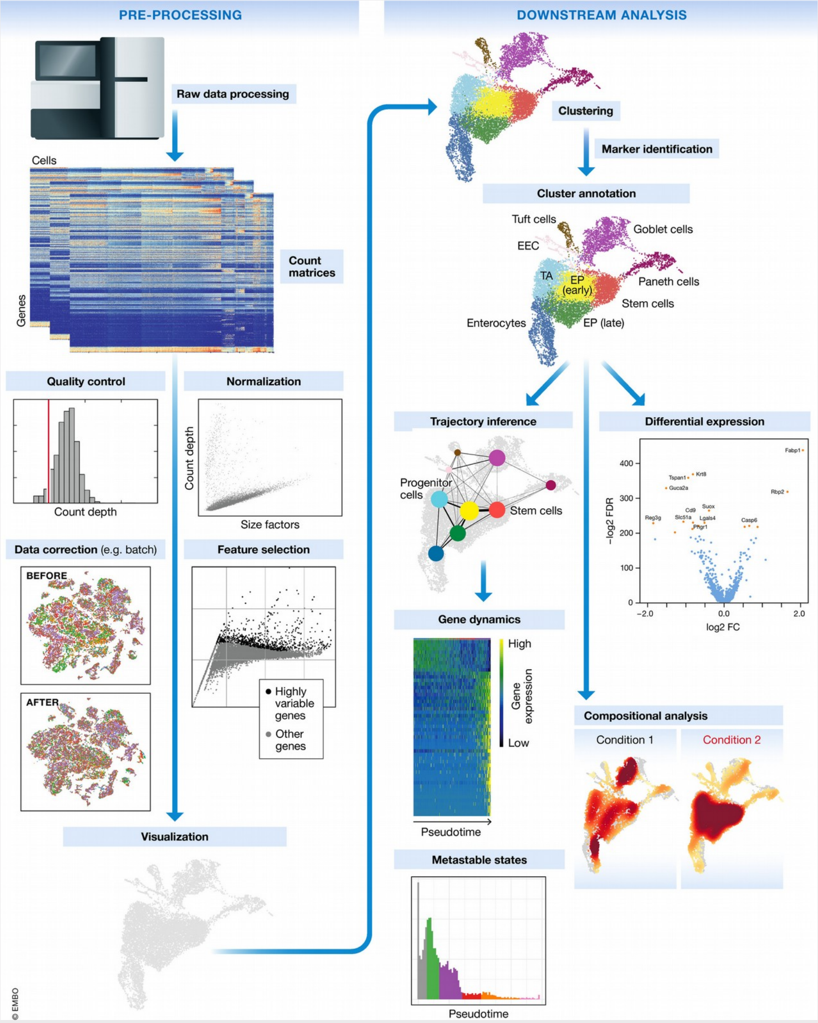
SincellTE 2022

Marine AGLAVE
Rémi MONTAGNE

Mapping, quality control and quantification

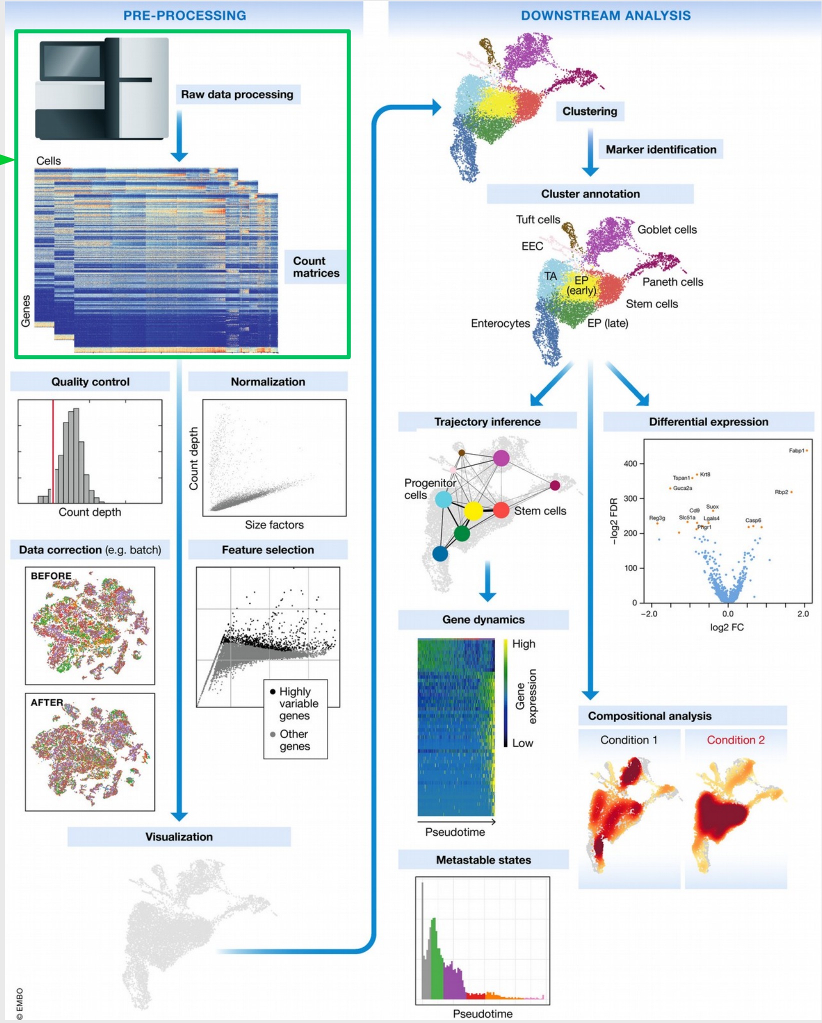
Main steps of single cell data processing



From Luecken and Theis,
Mol Systems Biology 2019

Main steps of single cell data processing

this course

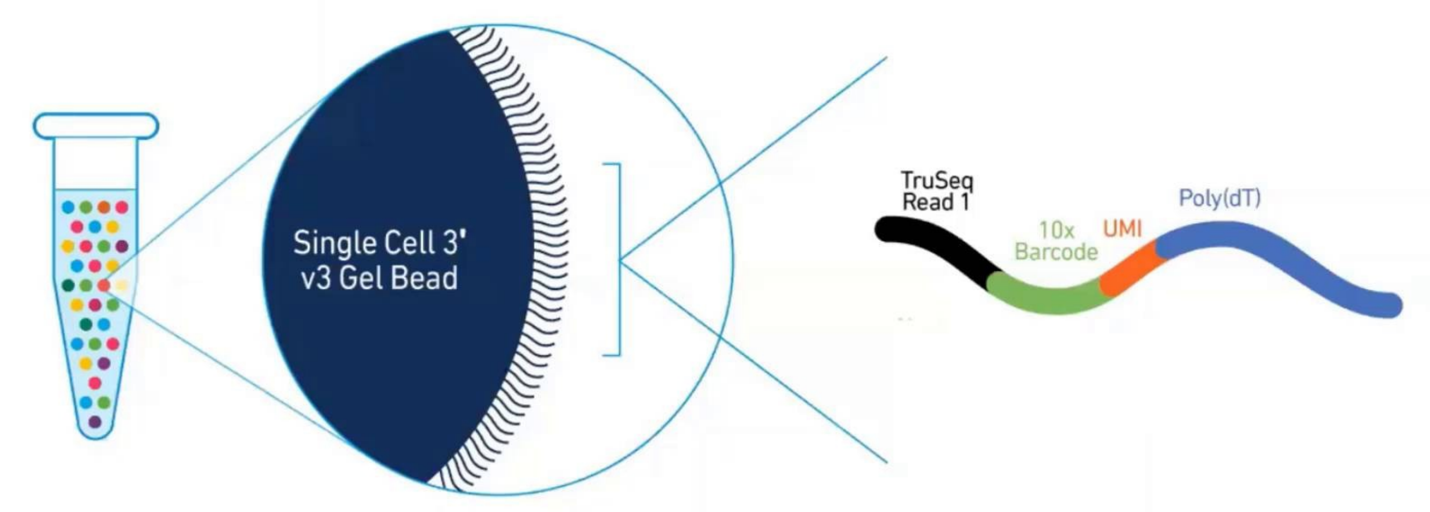


From Luecken and Theis,
Mol Systems Biology 2019

The starting library

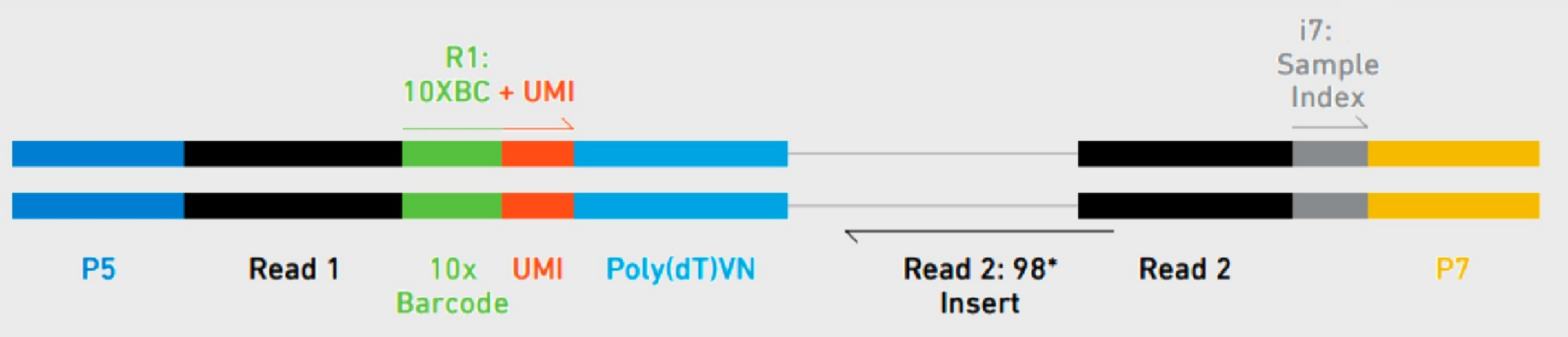
We will use a droplet-based library as an example.

10x GENOMICS®



The starting library

We will use a droplet-based library as an example.



Read1: unique cell barcode (x nt) + UMI (y nt)

Read2: RNA 3’ sequence

I7: sample index

Plan

- Demultiplexing: generating fastqs from bcl
- Quality Check
- Generating a gene x cell count matrix

Demultiplexing

- Illumina’s sequencer output is base call files (bcl).
- Convert them to fastq ?

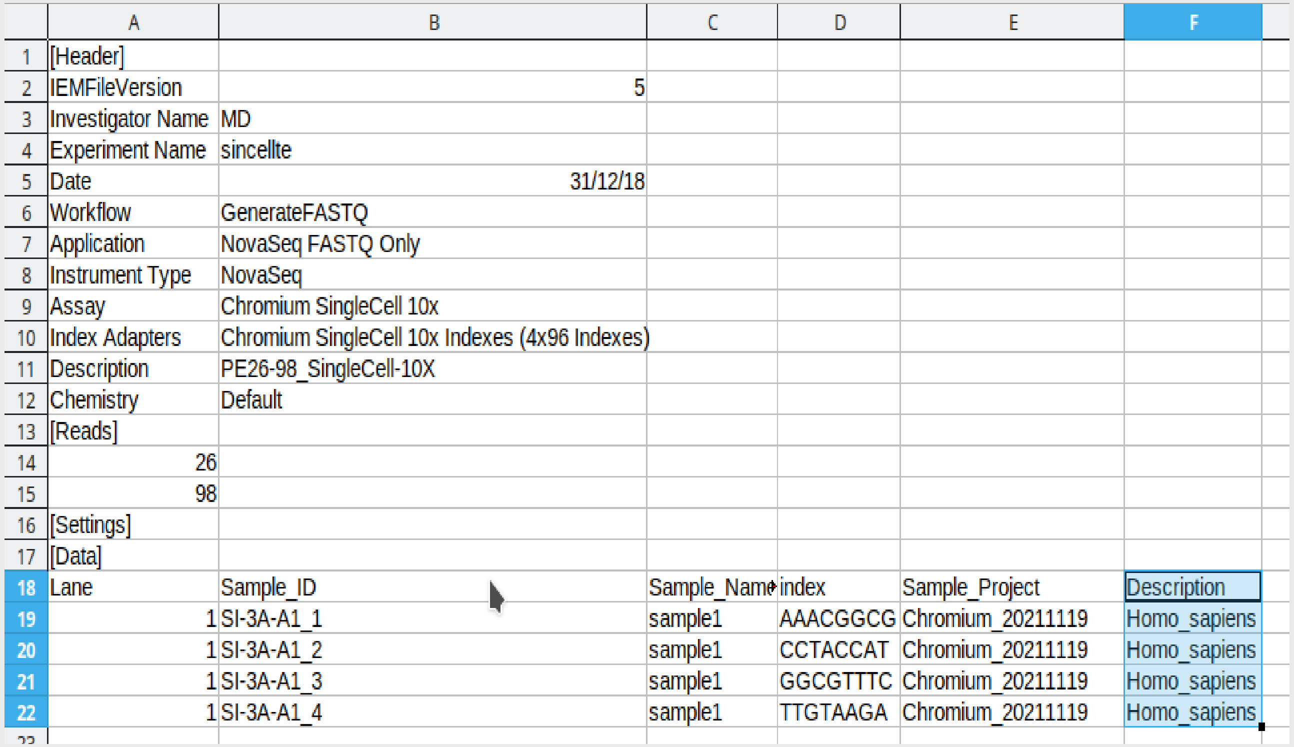
⇨ bcl2fastq

⇨ 10X’s cellranger mkfastq

Demultiplexing

bcl2fastq

- Usual sample sheet
- You must know :
  - i7 (i5) index sequence
  - R1 and R2 lengths
    (depends on technology, version…)
- 10X: 1 index = 4 sequences ⇒ 4 lines



	A	B	C	D	E	F
1	[Header]					
2	IEMFileVersion	5				
3	Investigator Name	MD				
4	Experiment Name	sincellte				
5	Date	31/12/18				
6	Workflow	GenerateFASTQ				
7	Application	NovaSeq FASTQ Only				
8	Instrument Type	NovaSeq				
9	Assay	Chromium SingleCell 10x				
10	Index Adapters	Chromium SingleCell 10x Indexes (4x96 Indexes)				
11	Description	PE26-98_SingleCell-10X				
12	Chemistry	Default				
13	[Reads]					
14	26					
15	98					
16	[Settings]					
17	[Data]					
18	Lane	Sample_ID	Sample_Name	index	Sample_Project	Description
19	1	SI-3A-A1_1	sample1	AAACGGCG	Chromium_20211119	Homo_sapiens
20	1	SI-3A-A1_2	sample1	CCTACCAT	Chromium_20211119	Homo_sapiens
21	1	SI-3A-A1_3	sample1	GGCGTTTC	Chromium_20211119	Homo_sapiens
22	1	SI-3A-A1_4	sample1	TTGTAAGA	Chromium_20211119	Homo_sapiens

https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/bcl2fastq-direct

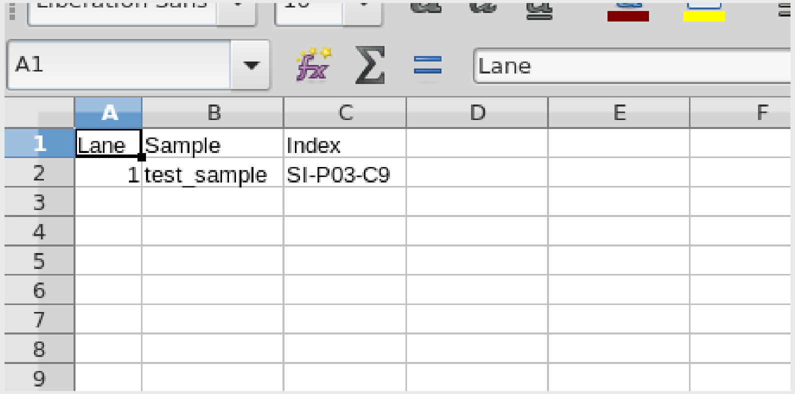
Demultiplexing

Cellranger mkfastq

- A wrapper around bcl2fastq with additional features:
  - Automatic translation of index names to sequences
  - Splitting work into multiple jobs (HPC)

- Simpler samplesheet : csv file, 3 columns

- Additional barcodes QC-metrics



Lane	Sample	Index
1	test_sample	SI-P03-C9

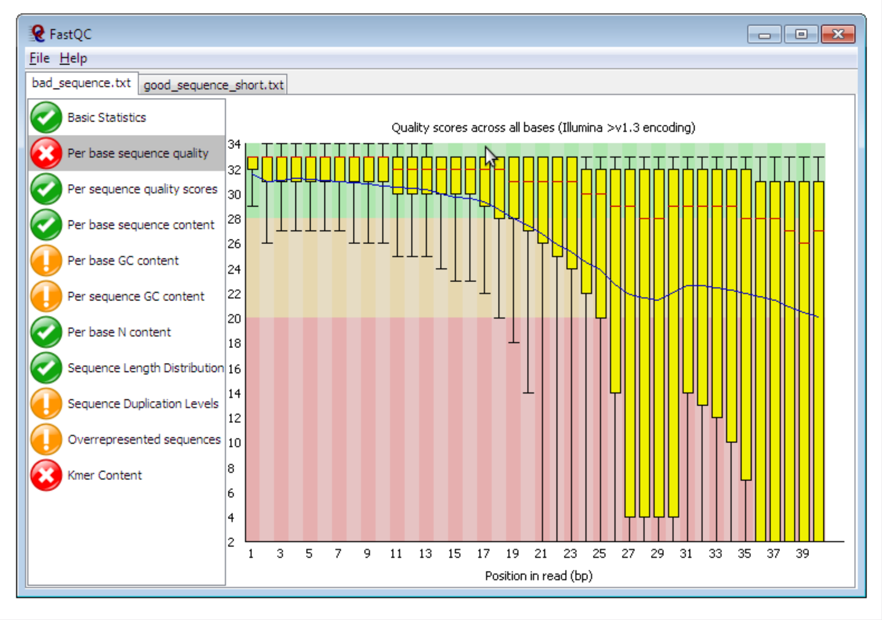
https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/mkfastq#simple_csv

Quality Check

Check reads quality : fastqc

- Performs various basic QC on reads

- For 10X scRNA datasets :
  - R1 (BC + UMI) : QC is mandatory. Watch out for Ns and highly repeated sequences
  - R2 : do as usual

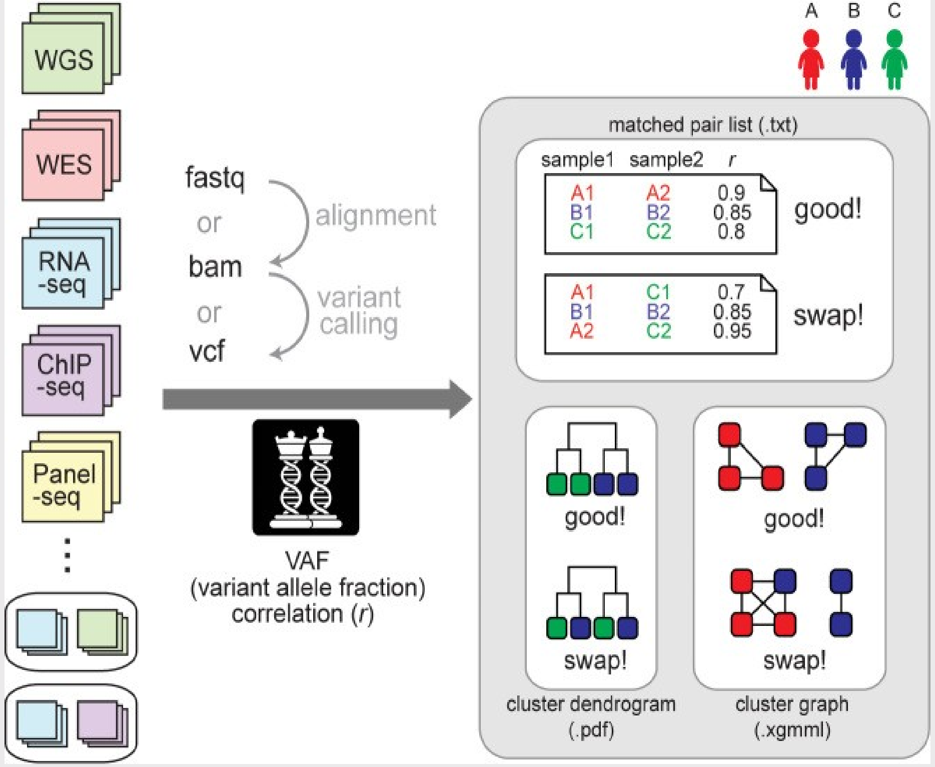


- ✓ Basic Statistics
- ✗ Per base sequence quality
- ✓ Per sequence quality scores
- ✓ Per base sequence content
- ! Per base GC content
- ! Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ! Sequence Duplication Levels
- ! Overrepresented sequences
- ✗ Kmer Content

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Check close samples : NGS CheckMate

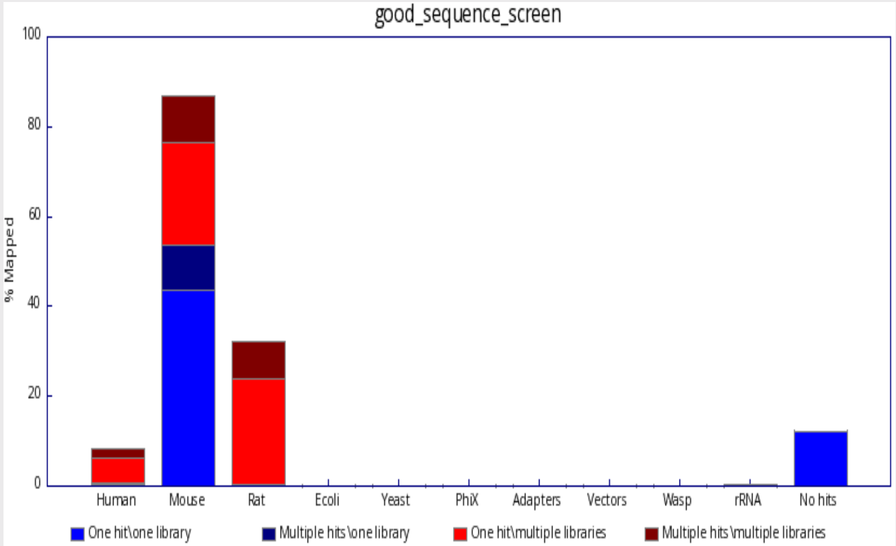
- When expected closeness : family, matched samples (e.g. : healthy-tumor)
- Check samples proximity using a set of known SNPs.
- Many data types : WES, WGS, RNA-Seq, ChIP-Seq, Many input formats : fastq, bam, VCF
- Helps controlling mislabelled samples

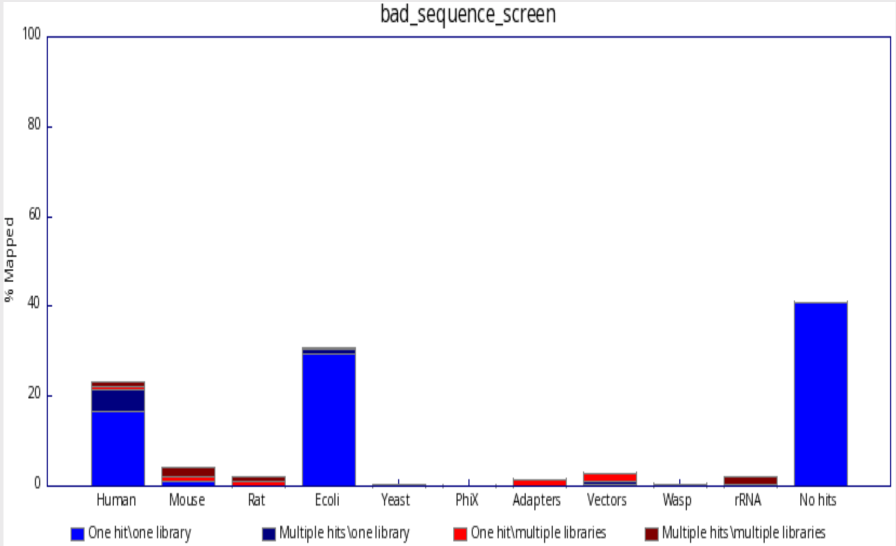


https://github.com/parklab/NGSCheckMate

Quality Check

Check cross-species contaminations: FastQ Screen





https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/

- Quick mapping (bowtie2) of a subset of reads across multiple genomes and common contaminants: human, mouse, rat, E. coli, adapters, vectors…

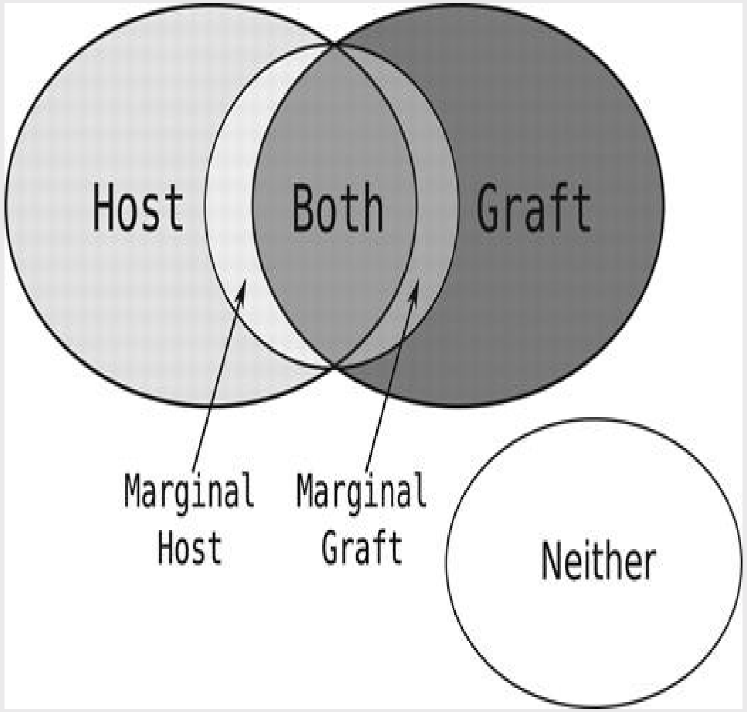
- Identifies 1hit-1library, multi hits-1library, 1hit-multi libraries and multi hits-multi libraries

Multiple species: Xenome

- For xenografts or contaminated samples

- 5 fastq files :
  - Graft
  - Host
  - Both
  - Neither
  - Ambiguous

- For single cell, apply to R2 only and sync R1: e.g. seqkit:
  - *seqkit seq* lists the selected read names.
  - *seqkit grep* filters R1 by keeping only reads in this list.
  - *seqkit pair* pairs filtered R1 with R2.



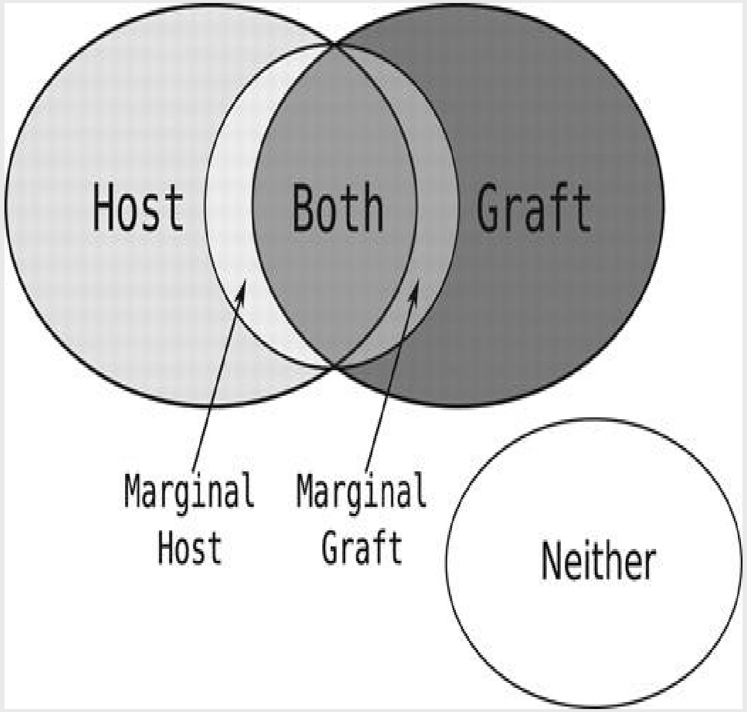
https://github.com/data61/gossamer/blob/master/docs/xenome.md

Quality Check

Multiple species: Xenome

- Xenome version is bugged: patch gossamer
  https://github.com/data61/gossamer

- Alternatives :
  - Xengsort (Zentgraf and Rahmann, S. Mol Biol 2021).
  - XenofilteR (Kluin *et al*, BMC Bioinfo 2018)
  - Bamcmp (Khandelwal *et al.*, MCR 2017).
  - XenoSplit: (https://github.com/goknurginer/XenoSplit Unpublished 2019).



https://github.com/data61/gossamer/blob/master/docs/xenome.md

Quality Check

Trimming

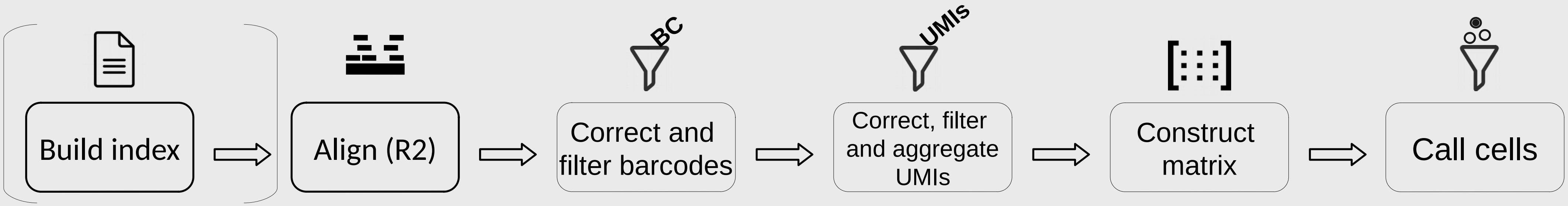
- If QC is not good:
  - Low base quality
  - Remaining adapter sequence
  - Homopolymer tailing
  - Low complexity

- Many tools to trim reads:
  - Trimmomatic (Bolger A.M. *et al.,* Bioinformatics (2014).
  - TrimGalore (Krueger F., https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/, unpublished 2012).
  - Cutadapt (Martin M., EMBnet.journal 2011)
  - Fastp (Chen *et al.,* Bioinformatics 2018).

- For single cell, like with xenome, apply to R2 file, then sync the R1 file.

Building the count matrix

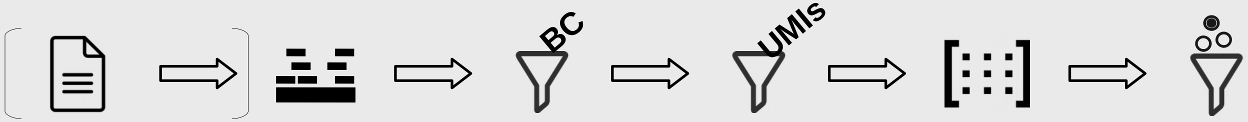
Principle


BC
UMIs
Build index
Align (R2)
Correct and filter barcodes
Correct, filter and aggregate UMIs
Construct matrix
Call cells

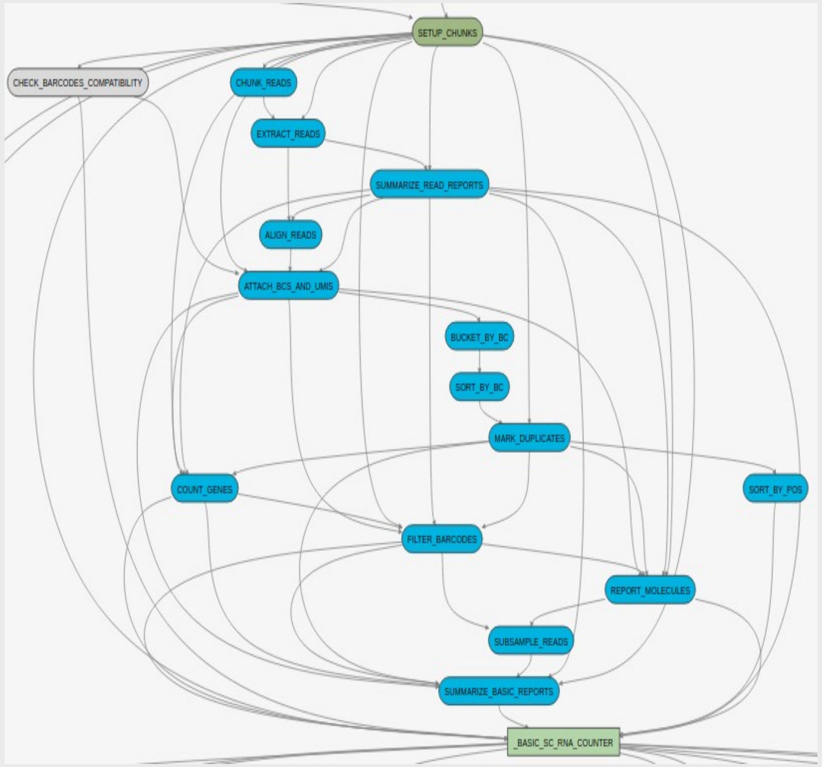
- Various tools have been developed:
  - **Cellranger**: 10X solution for 10X libraries only
  - **STARsolo**: an open source alternative to cellranger
  - **kallisto+bustools:** a pseudomapper and tool suite needing very little resources
  - (**Alevin**: a pseudomapper integrated with the salmon software)

Building the count matrix

Cellranger



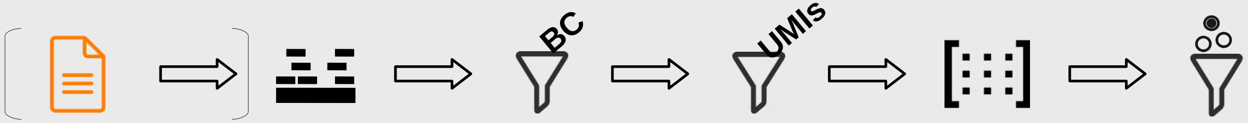
- A set of pipelines for single cell analysis
- Many languages + task scheduler Martian
- Aligner: STAR
- single cell gene expression: *cellranger count*



https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest

Building the count matrix

Cellranger



- Pre-built references: human (hg19, GRCh38), mouse (mm10) or both (xenografts)

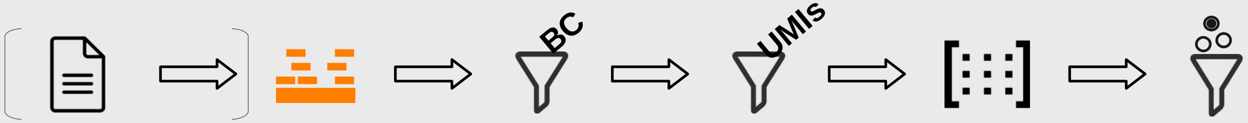
  https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/tutorial_mr

- For custom reference: *cellranger mkgtf* and *cellranger mkref.* Needs:
  - a genome FASTA
  - STAR compatible GTF file (Ensembl)

- Possible filtering according to biotype (lncRNA, protein coding...)

Building the count matrix

Cellranger



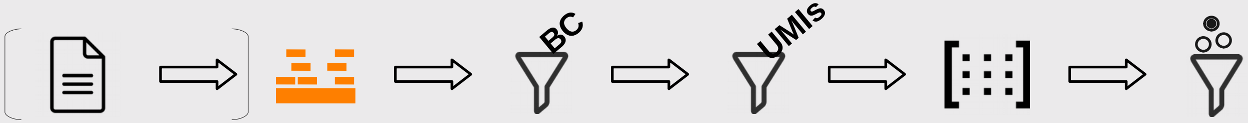
1 Splicing-aware genome alignment by STAR

2 Using gtf file, bucket the reads into:

- exonic : at least 50% mapping on an exon
- intronic : non exonic read intersecting an intron
- intergenic otherwise

3 Mapping quality adjustment: for reads that align on 1 single exon + non-exonic loci, the read is considered confidently mapped to the exon. MAPQ forced to 255.

Cellranger



1 Splicing-aware genome alignment by STAR (work only at the genome level)

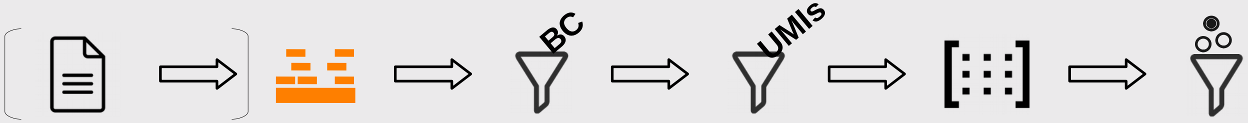
2 Using gtf file, bucket the reads into:

- exonic : when it intersects an exon for at at least 50% of its own length
- intronic : when the read is not exonic and intersects an intron
- ~~intergenic otherwise~~

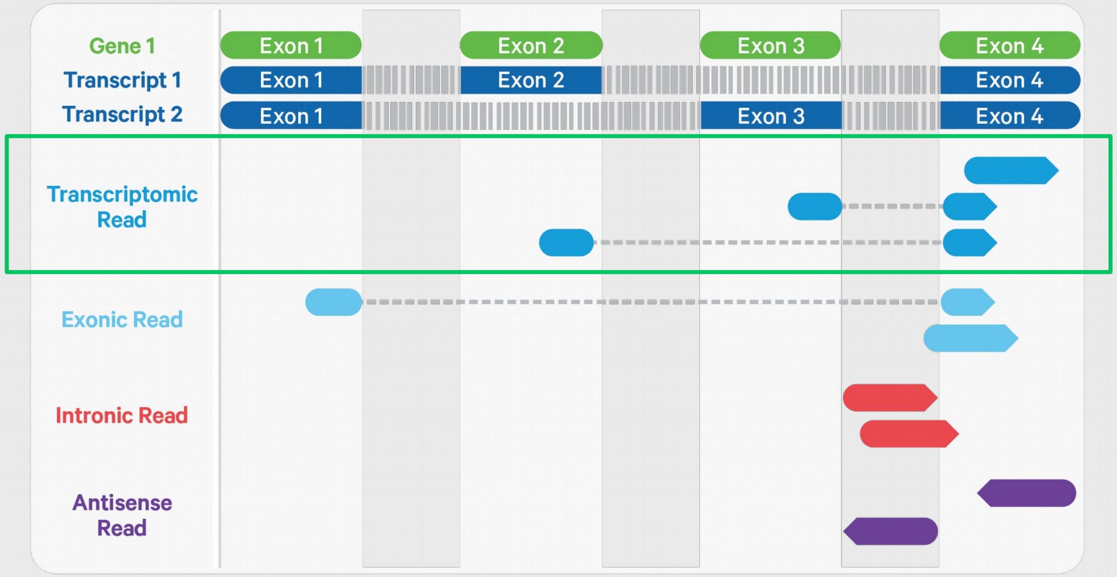
3 Mapping quality adjustment: for reads that align to 1 single exon + non-exonic loci, the read is considered confidently mapped to the exon. MAPQ forced to 255.

Building the count matrix

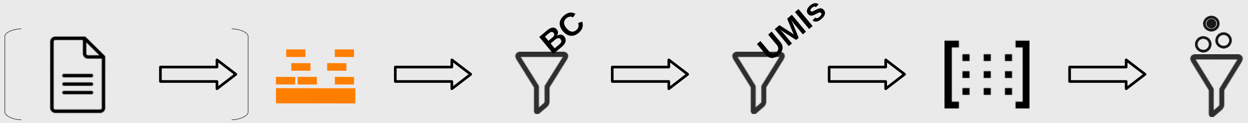
Cellranger



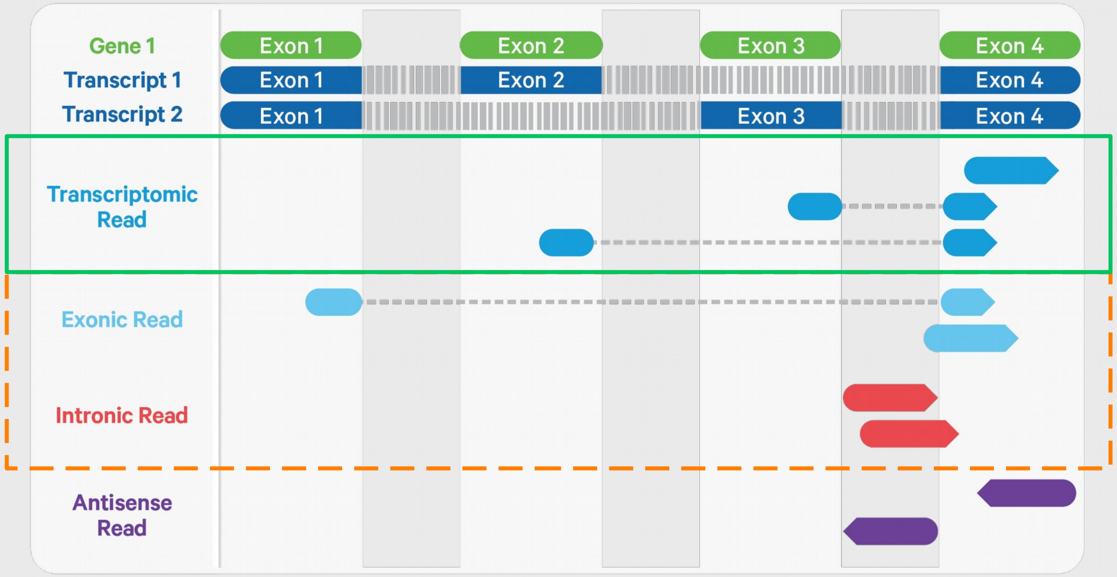
4 Transcriptome alignment of exonic and intronic reads (gtf file). Reads that are exonic, sens and compatible with a known transcript are selected.



Cellranger



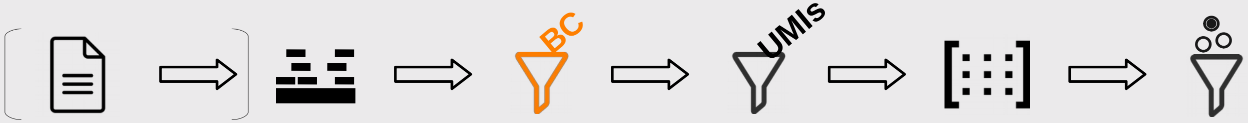
4 Transcriptome alignment (gtf) of exonic and intronic reads. Reads that are sens and compatible with a known transcript are selected.

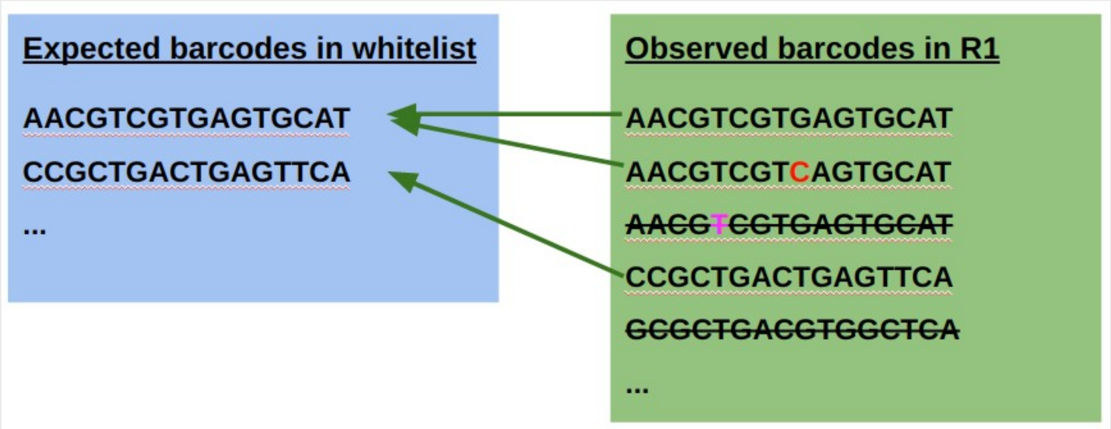


Selection of every sens read can be forced (assays on nuclei, many unspliced reads)

Building the count matrix

Cellranger





<u>Expected barcodes in whitelist</u>

AACGTCGTGAGTGCAT

CCGCTGACTGAGTTCA

...

<u>Observed barcodes in R1</u>

AACGTCGTGAGTGCAT

AACGTCGTCAGTGCAT

~~AACGTCGTGAGTGCAT~~

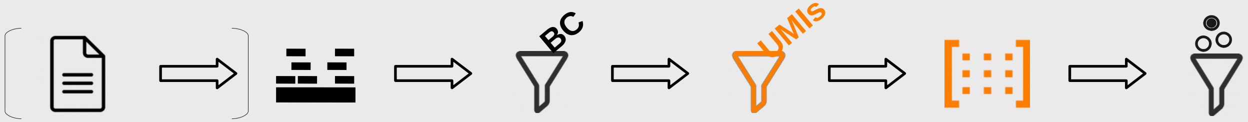
CCGCTGACTGAGTTCA

~~GCGCTGACGTGGCTCA~~

...

- **Attribute each selected read to 1 cell**
- White lists with all possible 10x barcodes
- Correction: barcodes with Hamming distance = 1 from a whitelist BC, ie one mismatch, are corrected (if the mismatch has a low BASEQ).
- Filtering: keep only BC in the whitelist.

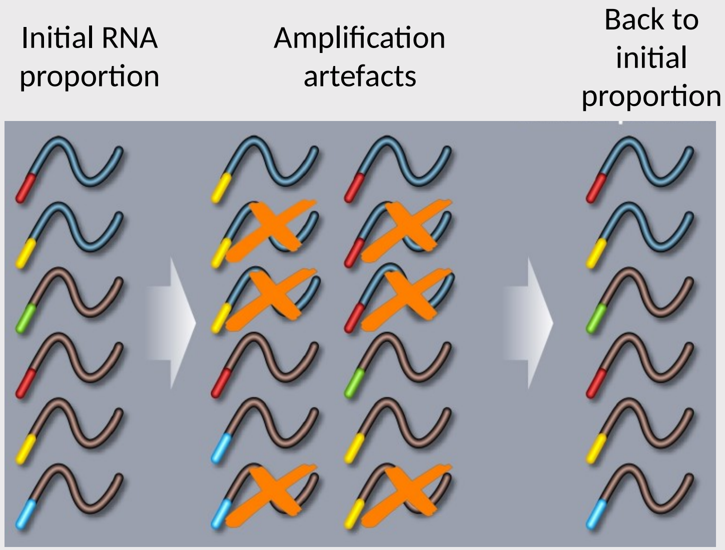
Cellranger



- **Correct amplification artefacts**

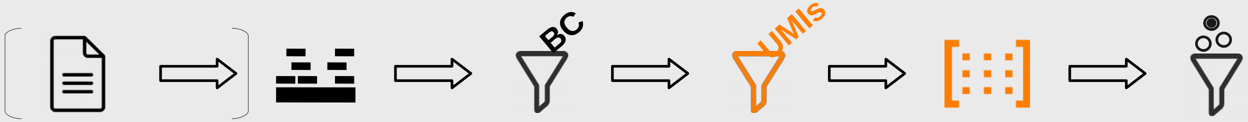
- Filtering : remove incorrect UMIs:
  - homopolymers (e.g. AAAAAAAAAA)
  - Contains 1 or several N
  - contains any base with BASEQ < 10

- Correction: if 2 UMIs have the same cell BC, the same gene alignment and a Hamming distance of 1, the lower-count UMI changed to the higher count UMI.



Building the count matrix

Cellranger



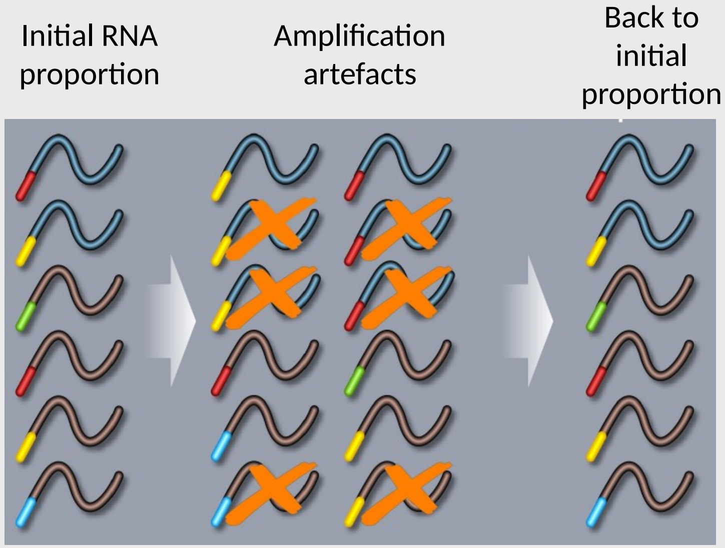
- **Correct amplification artefacts**

- Filtering : remove incorrect UMIs:
  - homopolymers (e.g. AAAAAAAAAA)
  - Contains 1 or several N
  - contains any base with BASEQ < 10

- Correction: if 2 UMIs have the same cell BC, the same gene alignment and a Hamming distance of 1, the lower-count UMI changed to the higher count UMI.

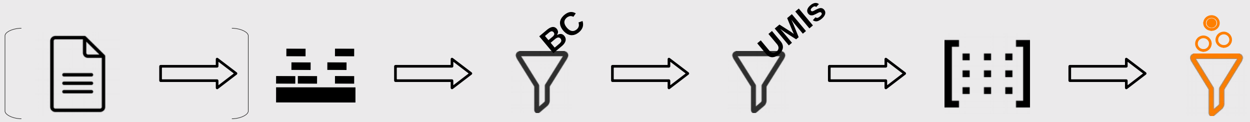
- Aggregation: 1 BC+UMIs = 1 unique RNA molecule (filter excess)

- Finally, construct matrix with selected reads: *genes x barcodes*

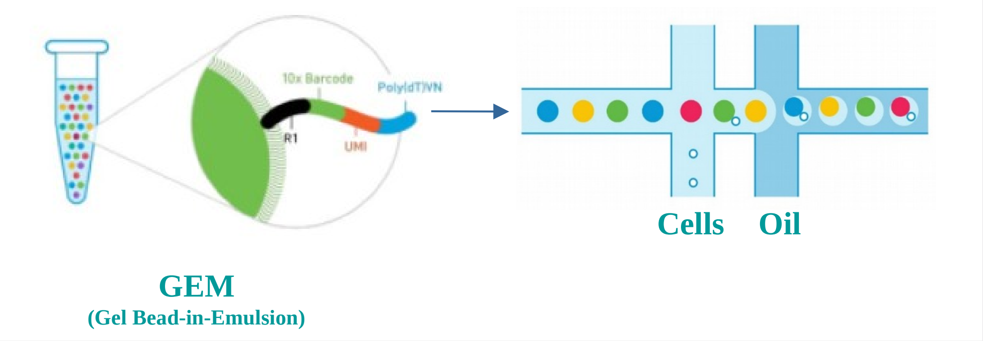


Building the count matrix

Cellranger



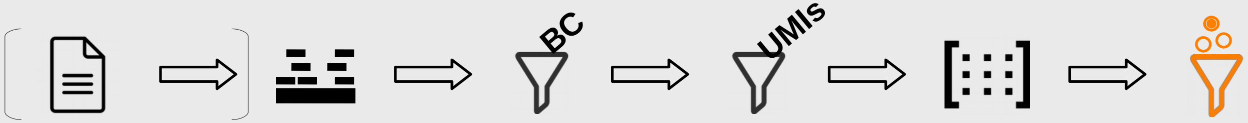
- Most droplets contain no cell:
  - ~ 10 000 cells
  - ~ 100 000 droplets



- Call the actual cells

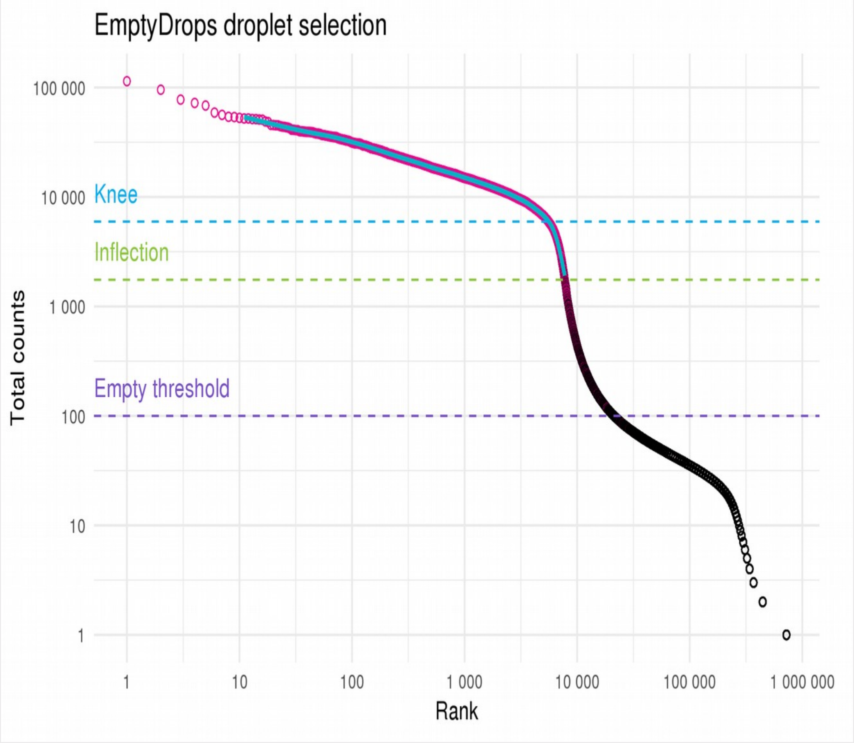
- But they contain circulating RNA from dead cells, i.e. a meaningless ambient ‘soup’.

Cellranger



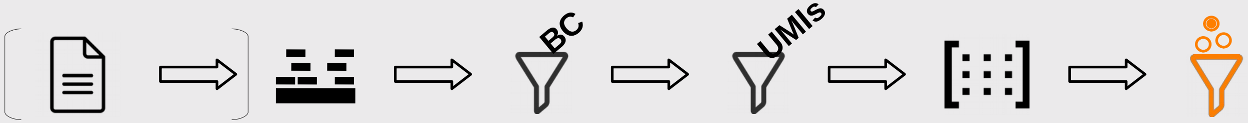
- 1st version of cell calling algorithm was specific to cellranger: a simple threshold:

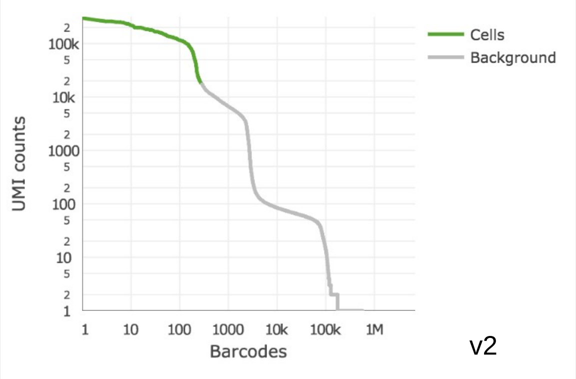
1. Rank droplets by decreasing count: kneeplot.
2. Take nb UMIs in one of the most populated droplets (99th percentil): m.
3. Select droplets where nb UMIs ≥ m/10

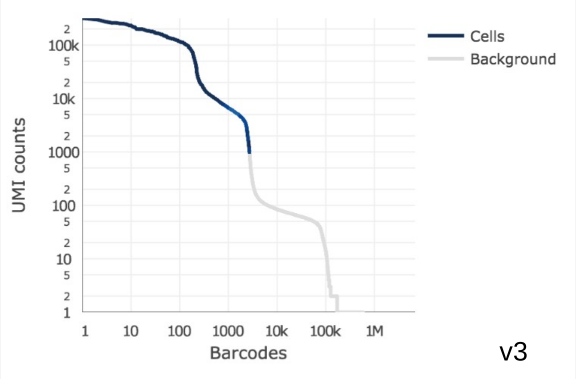


Building the count matrix

Cellranger





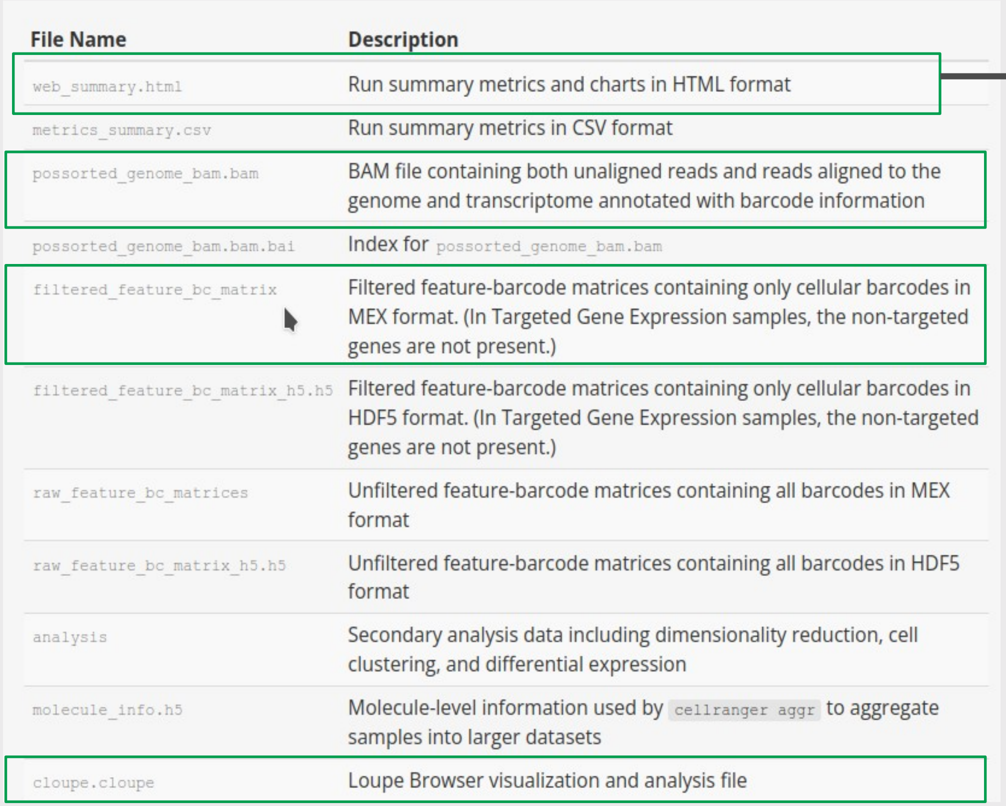


- Problems for complex libraries: e.g. tumors with high RNA content tumor cells + low RNA content tumor infiltrating lymphocytes
- Cellrangerv3 added a 2nd step (re-implementation of open source EmptyDroplets):
  - deduce background from low content droplets
  - select droplets with very different composition
- During next steps, a last filtering will generally be needed.

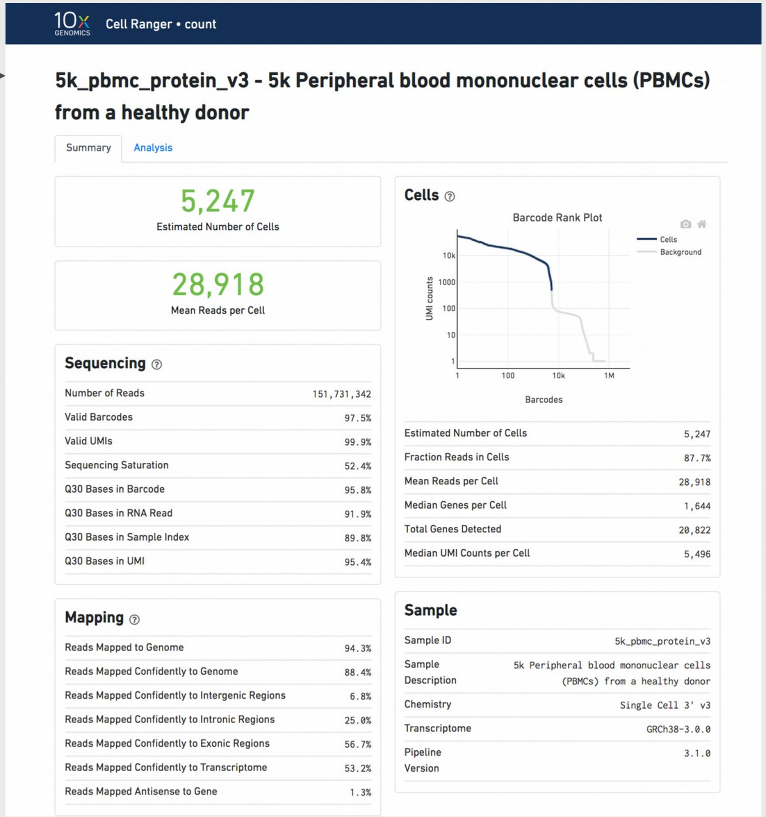
Building the count matrix

Cellranger

Outputs



File Name	Description
`web_summary.html`	Run summary metrics and charts in HTML format
`metrics_summary.csv`	Run summary metrics in CSV format
`possorted_genome_bam.bam`	BAM file containing both unaligned reads and reads aligned to the genome and transcriptome annotated with barcode information
`possorted_genome_bam.bam.bai`	Index for `possorted_genome_bam.bam`
`filtered_feature_bc_matrix`	Filtered feature-barcode matrices containing only cellular barcodes in MEX format. (In Targeted Gene Expression samples, the non-targeted genes are not present.)
`filtered_feature_bc_matrix_h5.h5`	Filtered feature-barcode matrices containing only cellular barcodes in HDF5 format. (In Targeted Gene Expression samples, the non-targeted genes are not present.)
`raw_feature_bc_matrices`	Unfiltered feature-barcode matrices containing all barcodes in MEX format
`raw_feature_bc_matrix_h5.h5`	Unfiltered feature-barcode matrices containing all barcodes in HDF5 format
`analysis`	Secondary analysis data including dimensionality reduction, cell clustering, and differential expression
`molecule_info.h5`	Molecule-level information used by `cellranger aggr` to aggregate samples into larger datasets
`cloupe.cloupe`	Loupe Browser visualization and analysis file



10x GENOMICS Cell Ranger • count

5k_pbmc_protein_v3 - 5k Peripheral blood mononuclear cells (PBMCs) from a healthy donor

Summary | Analysis

5,247
Estimated Number of Cells

28,918
Mean Reads per Cell

Sequencing

Number of Reads	151,731,342
Valid Barcodes	97.5%
Valid UMIs	99.9%
Sequencing Saturation	52.4%
Q30 Bases in Barcode	95.8%
Q30 Bases in RNA Read	91.9%
Q30 Bases in Sample Index	89.8%
Q30 Bases in UMI	95.4%

Mapping

Reads Mapped to Genome	94.3%
Reads Mapped Confidently to Genome	88.4%
Reads Mapped Confidently to Intergenic Regions	6.8%
Reads Mapped Confidently to Intronic Regions	25.0%
Reads Mapped Confidently to Exonic Regions	56.7%
Reads Mapped Confidently to Transcriptome	53.2%
Reads Mapped Antisense to Gene	1.3%

Cells

Estimated Number of Cells	5,247
Fraction Reads in Cells	87.7%
Mean Reads per Cell	28,918
Median Genes per Cell	1,644
Total Genes Detected	20,822
Median UMI Counts per Cell	5,496

Sample

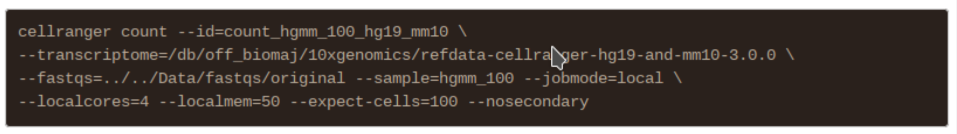
Sample ID	5k_pbmc_protein_v3
Sample Description	5k Peripheral blood mononuclear cells (PBMCs) from a healthy donor
Chemistry	Single Cell 3' v3
Transcriptome	GRCh38-3.0.0
Pipeline Version	3.1.0

Building the count matrix

Cellranger

🙂

- Turnkey solution



```
cellranger count --id=count_hgmm_100_hg19_mm10 \
--transcriptome=/db/off_biomaj/10xgenomics/refdata-cellranger-hg19-and-mm10-3.0.0 \
--fastqs=../../Data/fastqs/original --sample=hgmm_100 --jobmode=local \
--localcores=4 --localmem=50 --expect-cells=100 --nosecondary
```

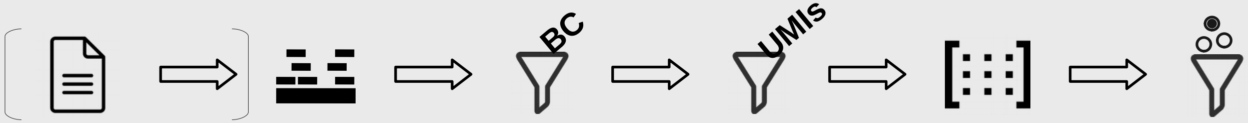
- Many QC-metrics, results summarized in 1 html.
- Some secondary analysis
- More complex experiences: VDJ analysis, feature-barcoding
- Versions for ATAC-Seq, multiomics

☹

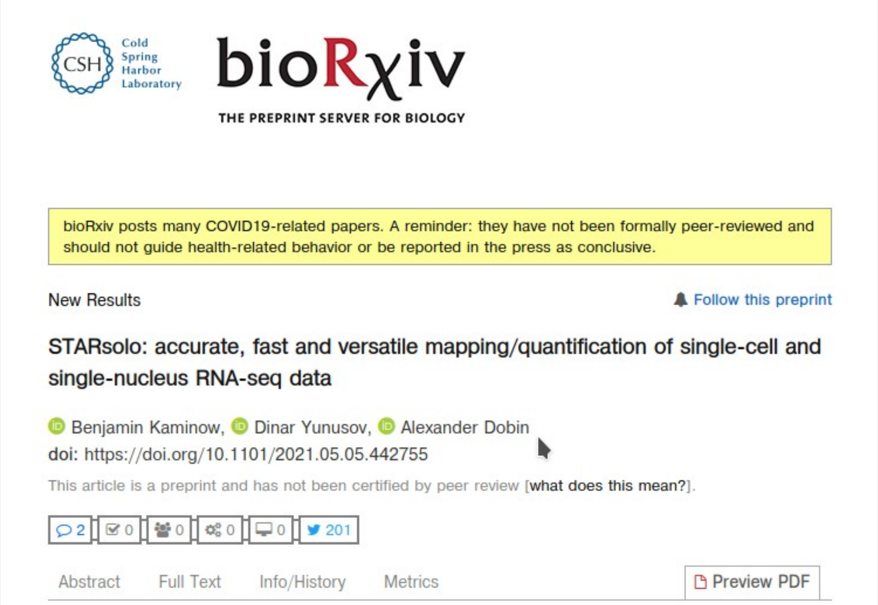
- Proprietary
- Analyze only 10X product (cannot customize BC and UMI)
- A lot of resource and time
- Has its own scheduler: hard to include in another pipeline
- Compatibility not guaranteed with all HPC managers

Building the count matrix

STARsolo



- Turnkey solution
- Implemented in STAR aligner
- Drop-in replacement for cellranger
- Same steps as cellranger:
  - Splice-aware genome alignment
  - Cell barcodes and UMI correction, filtering and aggregation
  - Matrix creation
  - Cell calling


CSH Cold Spring Harbor Laboratory
bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

🔔 Follow this preprint

STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data

Benjamin Kaminow, Dinar Yunusov, Alexander Dobin

doi: https://doi.org/10.1101/2021.05.05.442755

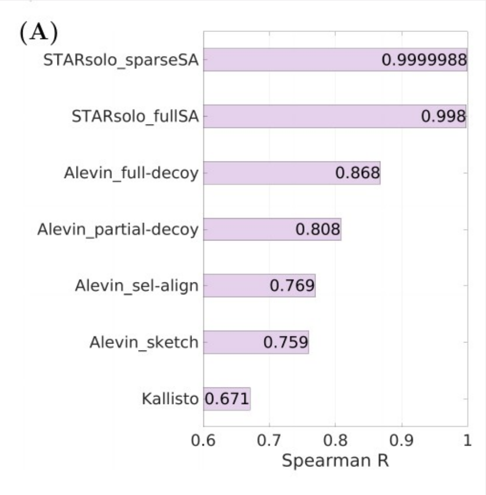
This article is a preprint and has not been certified by peer review [what does this mean?].

https://github.com/alexdobin/STAR

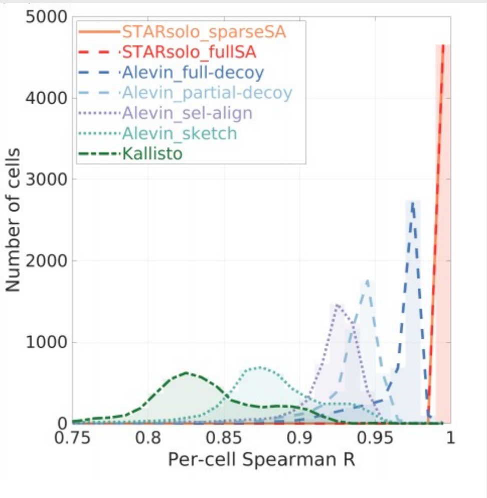
Building the count matrix

STARsolo

- Designed to give results as similar to Cellranger’s results as possible with the right set of parameters



Element-wise comparison of a gene-cell matrix with cellranger results



Per cell comparison with cellranger results

From Kaminow *et al*., BioRxiv 2020

Building the count matrix

STARsolo

- But highly configurable:
  - Alignment parameter
  - Read-to-gene assignment rule: e.g.: keep reads with several targets help keeping signal for paralogs
  - R1 structure (CB + UMI geometry)   **Allows analysis of non 10X technologies**
  - Rules for CB and UMI filtering

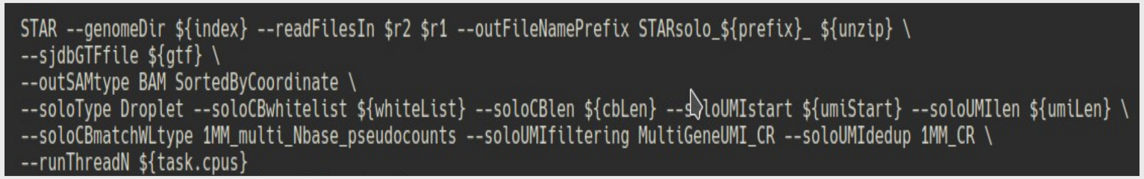
- Possibility to work at the transcripts level

Building the count matrix

STARsolo

🙂

- Turnkey solution



```
STAR --genomeDir ${index} --readFilesIn $r2 $r1 --outFileNamePrefix STARsolo_${prefix}_ ${unzip} \
--sjdbGTFfile ${gtf} \
--outSAMtype BAM SortedByCoordinate \
--soloType Droplet --soloCBwhitelist ${whiteList} --soloCBlen ${cbLen} --soloUMIstart ${umiStart} --soloUMIlen ${umiLen} \
--soloCBmatchWLtype 1MM_multi_Nbase_pseudocounts --soloUMIfiltering MultiGeneUMI_CR --soloUMIdedup 1MM_CR \
--runThreadN ${task.cpus}
```

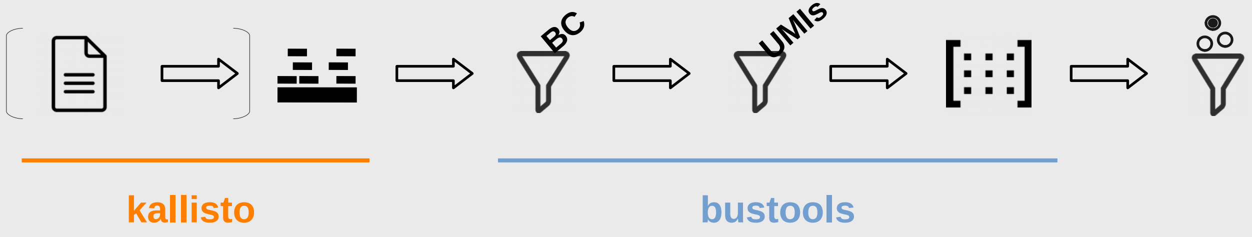
- Not proprietary
- Highly configurable (allows analysis of non 10X technologies)
- Needs less resource than cellranger
- Easy to include in a pipeline
- Compatible with HPC managers

☹

- Many QC files but not summarized
- No secondary analysis
- Does not take in charge more complex experiences (feature barcoding), ATAC-Seq...

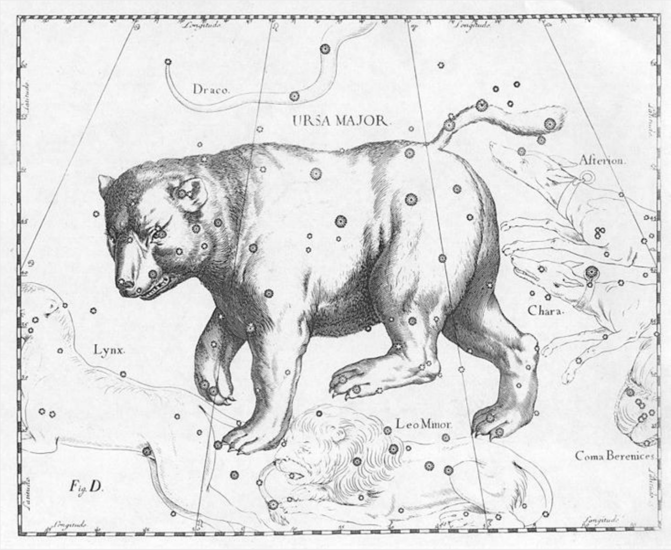
Building the count matrix

Kallisto/bustools



- Make use of the pseudo-aligner kallisto and the toolsuite bustools

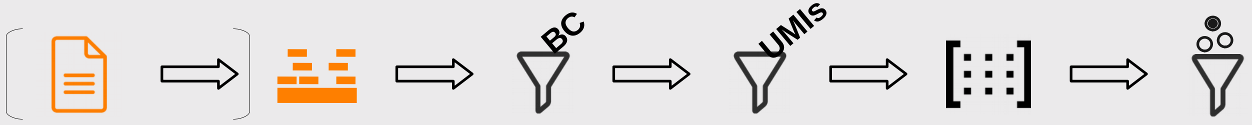
- Very good time and memory performance.



https://pachterlab.github.io/kallisto/download

Building the count matrix

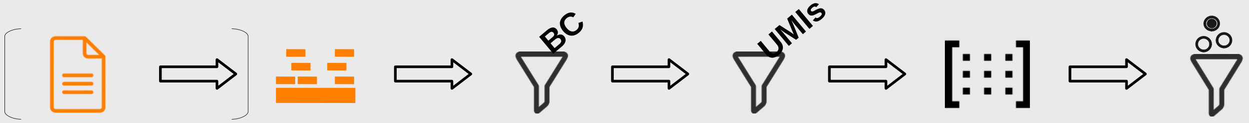
Kallisto/bustools



- Kallisto is a pseudo aligner: fast, low memory
- Working with a reference transcriptome, not genome

Building the count matrix

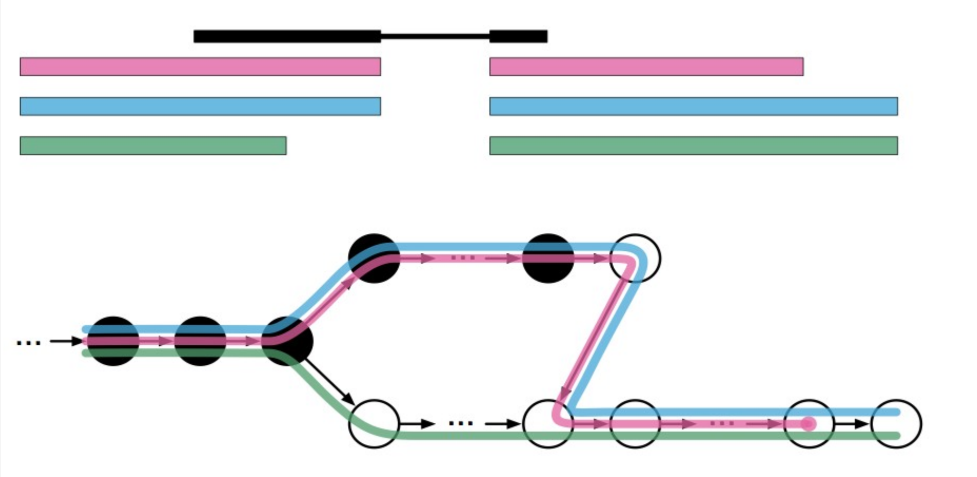
Kallisto/bustools



- Kallisto is a pseudo aligner: fast, low memory

- Working with a reference transcriptome, not genome

- Principle:
  - reference chunked into k-mers ==> de Bruijn Graph
  - Reads chunked into k-mers and assigned to the transcript(s) they overlap with
  - 1 read generally compatible with several transcripts: proportion of transcripts computed by Expectation Maximization from all reads

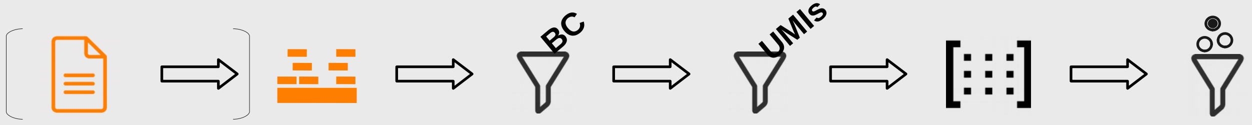


A very nice explanation of kallisto: https://bioinfo.iric.ca/fr/comprendre-comment-kallisto-fonctionne

From Bray *et al.*, *Nat Biointechno* 2016

Building the count matrix

Kallisto/bustools

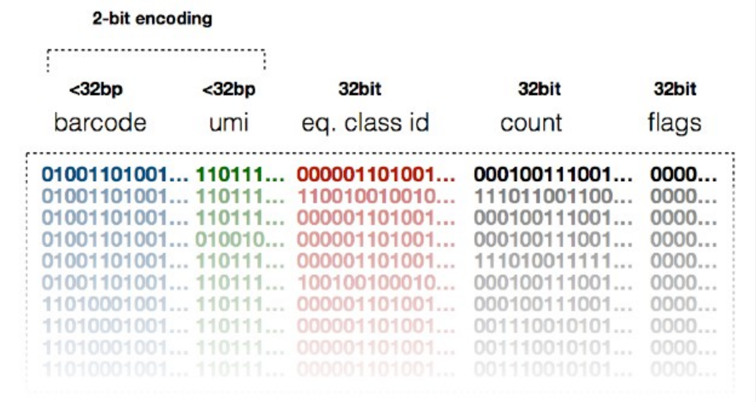


- Many technologies already accepted, the CB + UMI geometry is configurable

Allows analysis of non 10X technologies

- Gives relative abundance, not absolute counts

- Output format in a specific, compressed format: bus instead of sam or bam files.

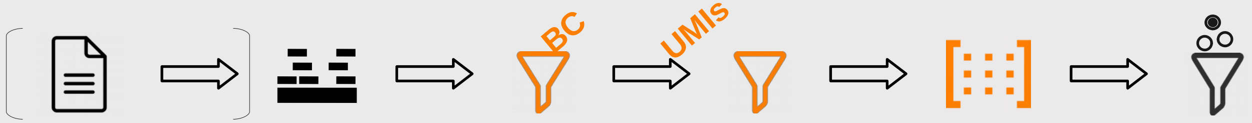


2-bit encoding				
<32bp	<32bp	32bit	32bit	32bit
barcode	umi	eq. class id	count	flags
01001101001...	110111...	000001101001...	000100111001...	0000...
01001101001...	110111...	110010010010...	111011001100...	0000...
01001101001...	110111...	000001101001...	000100111001...	0000...
01001101001...	010010...	000001101001...	000100111001...	0000...
01001101001...	110111...	000001101001...	111010011111...	0000...
01001101001...	110111...	100100100010...	000100111001...	0000...
11010001001...	110111...	000001101001...	000100111001...	0000...
11010001001...	110111...	000001101001...	001110010101...	0000...
11010001001...	110111...	000001101001...	001110010101...	0000...
11010001001...	110111...	000001101001...	001110010101...	0000...

From Melsted *et al.*, *Bioinformatics* 2019

Building the count matrix

Kallisto/bustools



- Next steps: bustools

Bus file + BC whitelist

↓

bustools correct: correct and filter BC

↓

bustools sort: sort results by BC, UMIs and gene

↓

bustools count: correct and filter UMIs, construct matrix

↓

raw gene x barcodes matrix

- The matrix must then be filtered: e.g. EmptyDrops (Lun *et al.*, Genome Biol 2019).

Building the count matrix

Kallisto/bustools

🙂

- For modular pipeline construction
- Not proprietary
- Allows analysis of non 10X technologies
- The fastest and less resource consuming (can run on a laptop)
- Easy to include in a pipeline
- Compatible with HPC managers

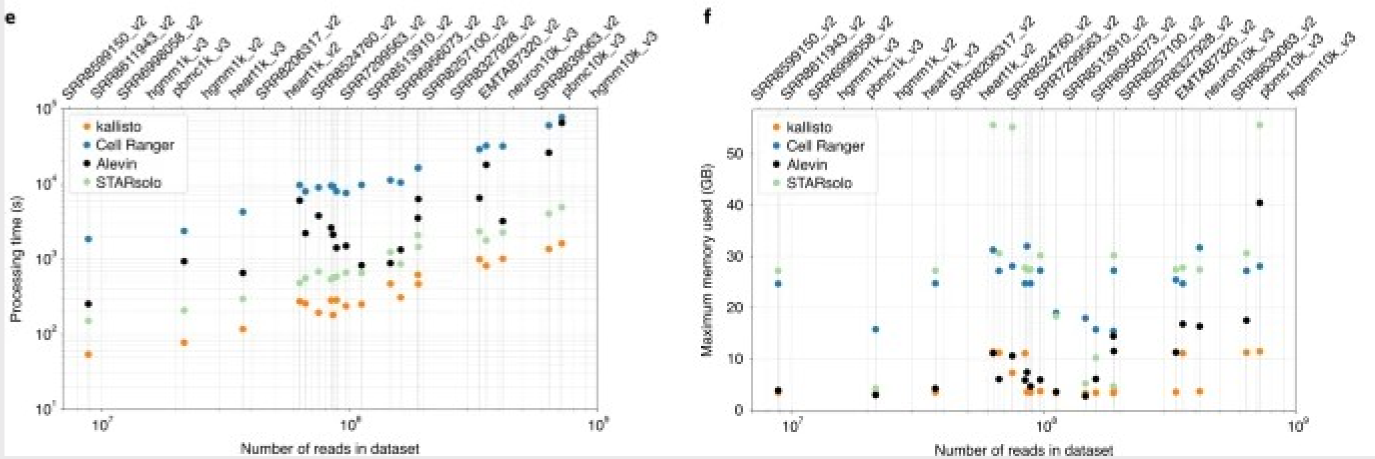
☹

- Not a turnkey solution
- No secondary analysis
- Gap with cellranger
- No empty droplets filtering

Building the count matrix

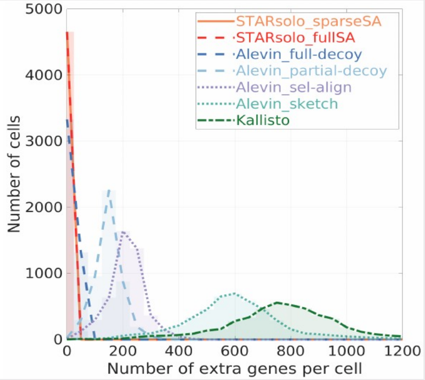
Which alternative to cellranger ?

- Kallisto has the best performances



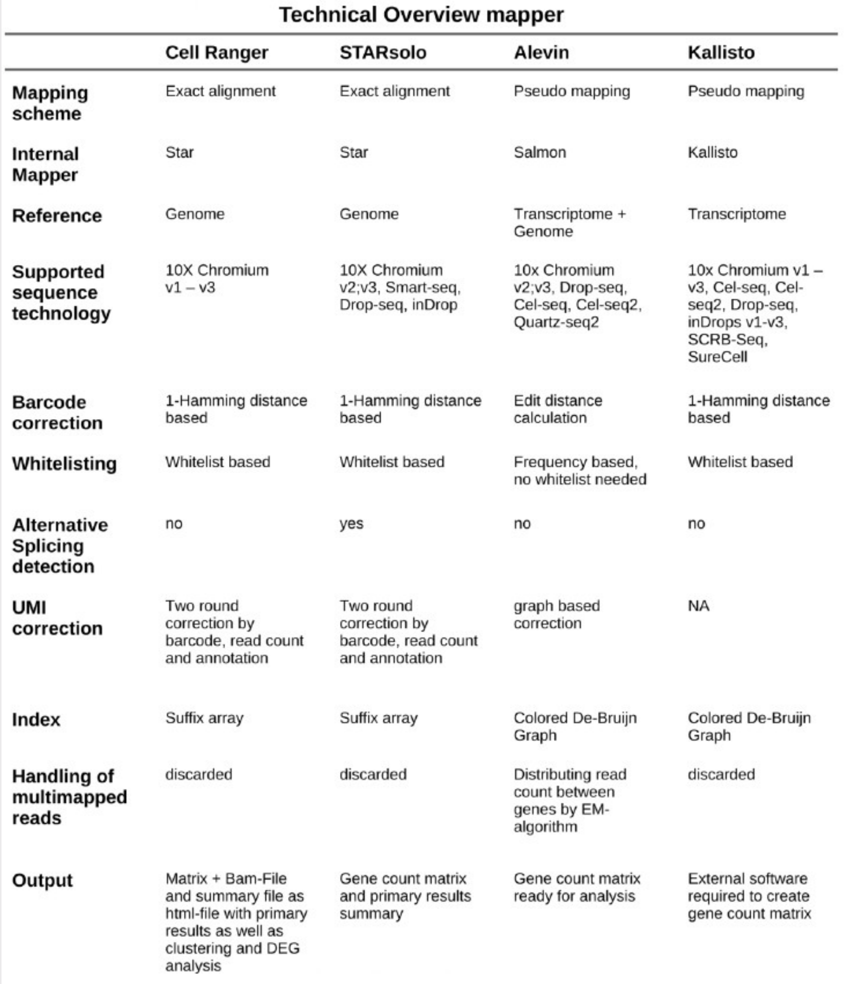
Melsted P. *et al.*, Nat. Biotech. (2021)

- Specificity: Brüning *et al.* and Kaminow *et al.* report more genes per cells and more false positive with pseudomappers (kallisto)



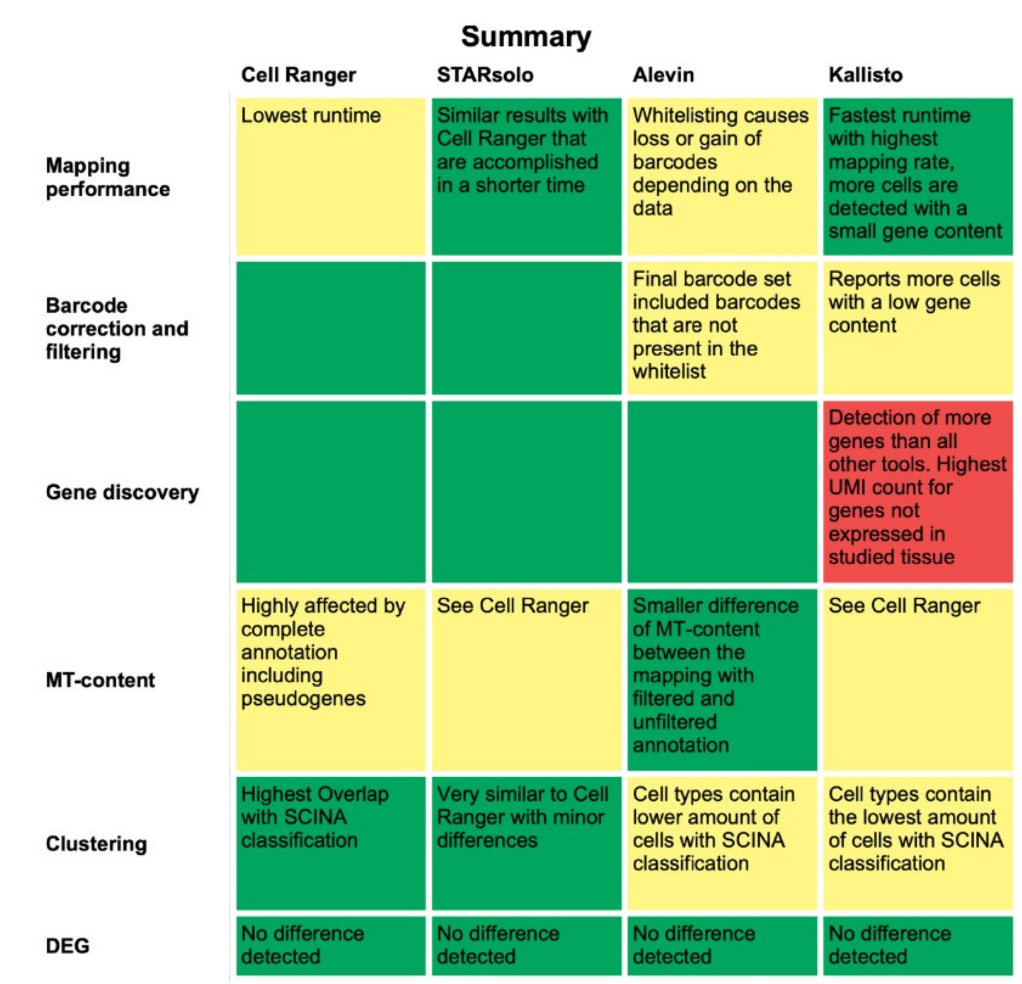
Kaminow *et al.*, BioRxiv (2020)

Conclusion



Technical Overview mapper

	Cell Ranger	STARsolo	Alevin	Kallisto
Mapping scheme	Exact alignment	Exact alignment	Pseudo mapping	Pseudo mapping
Internal Mapper	Star	Star	Salmon	Kallisto
Reference	Genome	Genome	Transcriptome + Genome	Transcriptome
Supported sequence technology	10X Chromium v1 – v3	10X Chromium v2;v3, Smart-seq, Drop-seq, inDrop	10x Chromium v2;v3, Drop-seq, Cel-seq, Cel-seq2, Quartz-seq2	10x Chromium v1 – v3, Cel-seq, Cel-seq2, Drop-seq, inDrops v1-v3, SCRB-Seq, SureCell
Barcode correction	1-Hamming distance based	1-Hamming distance based	Edit distance calculation	1-Hamming distance based
Whitelisting	Whitelist based	Whitelist based	Frequency based, no whitelist needed	Whitelist based
Alternative Splicing detection	no	yes	no	no
UMI correction	Two round correction by barcode, read count and annotation	Two round correction by barcode, read count and annotation	graph based correction	NA
Index	Suffix array	Suffix array	Colored De-Bruijn Graph	Colored De-Bruijn Graph
Handling of multimapped reads	discarded	discarded	Distributing read count between genes by EM-algorithm	discarded
Output	Matrix + Bam-File and summary file as html-file with primary results as well as clustering and DEG analysis	Gene count matrix and primary results summary	Gene count matrix ready for analysis	External software required to create gene count matrix

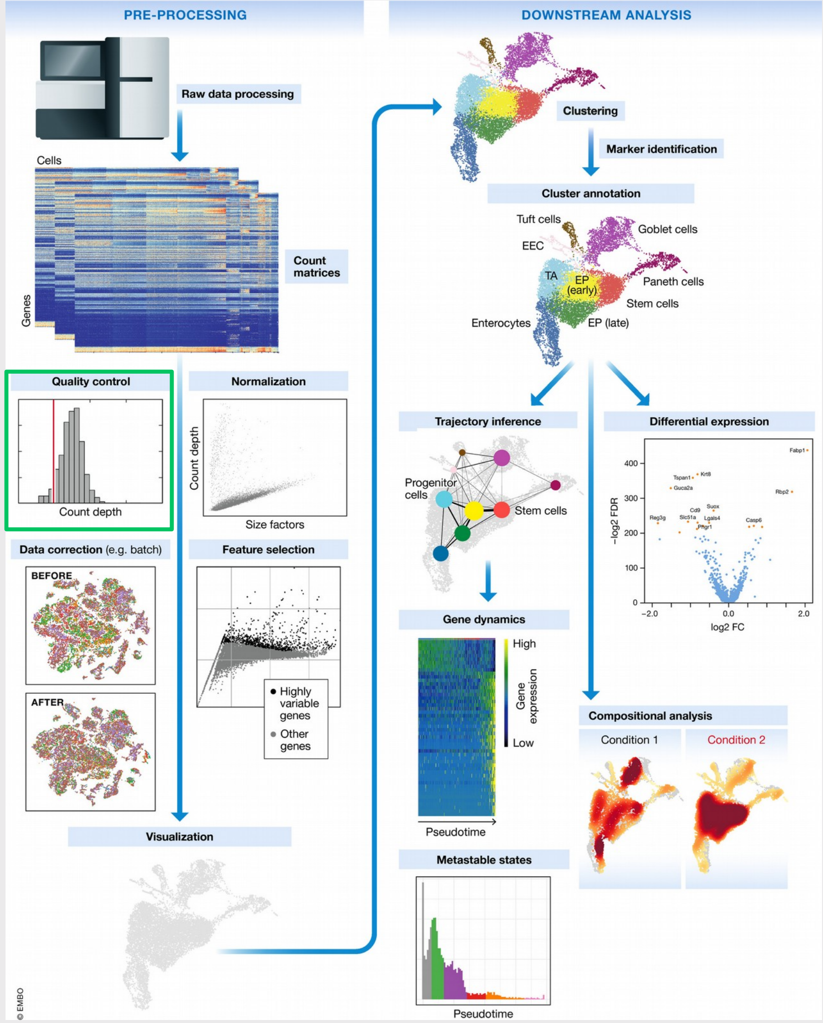


Summary

	Cell Ranger	STARsolo	Alevin	Kallisto
Mapping performance	🟨 Lowest runtime	🟩 Similar results with Cell Ranger that are accomplished in a shorter time	🟨 Whitelisting causes loss or gain of barcodes depending on the data	🟩 Fastest runtime with highest mapping rate, more cells are detected with a small gene content
Barcode correction and filtering	🟩	🟩	🟨 Final barcode set included barcodes that are not present in the whitelist	🟨 Reports more cells with a low gene content
Gene discovery	🟩	🟩	🟩	🟥 Detection of more genes than all other tools. Highest UMI count for genes not expressed in studied tissue
MT-content	🟨 Highly affected by complete annotation including pseudogenes	🟨 See Cell Ranger	🟩 Smaller difference of MT-content between the mapping with filtered and unfiltered annotation	🟨 See Cell Ranger
Clustering	🟩 Highest Overlap with SCINA classification	🟩 Very similar to Cell Ranger with minor differences	🟨 Cell types contain lower amount of cells with SCINA classification	🟨 Cell types contain the lowest amount of cells with SCINA classification
DEG	🟩 No difference detected	🟩 No difference detected	🟩 No difference detected	🟩 No difference detected

R. S. Brüning *et al.*, bioRXiv (2021)

Next step



Malte D Luecken & Fabian J Theis
Molecular Systems Biology (2019)

SincellTE 2022

Marine AGLAVE
Rémi MONTAGNE

Thank you for your attention!

Additional resources

A very handy training session about scRNAseq :

- Main page (2020 edition) : https://hbctraining.github.io/scRNA-seq_online/schedule/
- Quantification matrix QC (2018 edition) : https://hbctraining.github.io/In-depth-NGS-Data-Analysis-Course/sessionIV/lessons/SC_quality_control_analysis.html

Thanks to Bastien Job

institut Curie