



Differential Accessible Regions analysis of single-cell 10X Genomics multiome data

Dario Righelli, PhD

PostDoc

Department of Statistical Sciences, University of Padova, Padua, Italy

dario.righelli@gmail.com

The 10X Multiome protocol

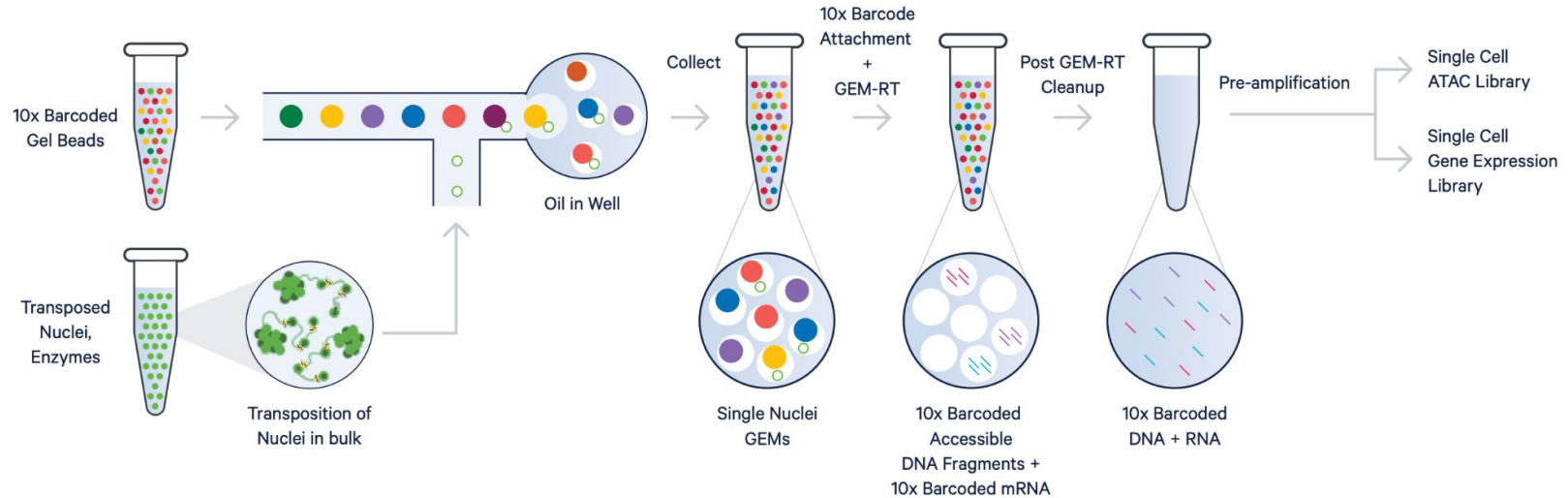
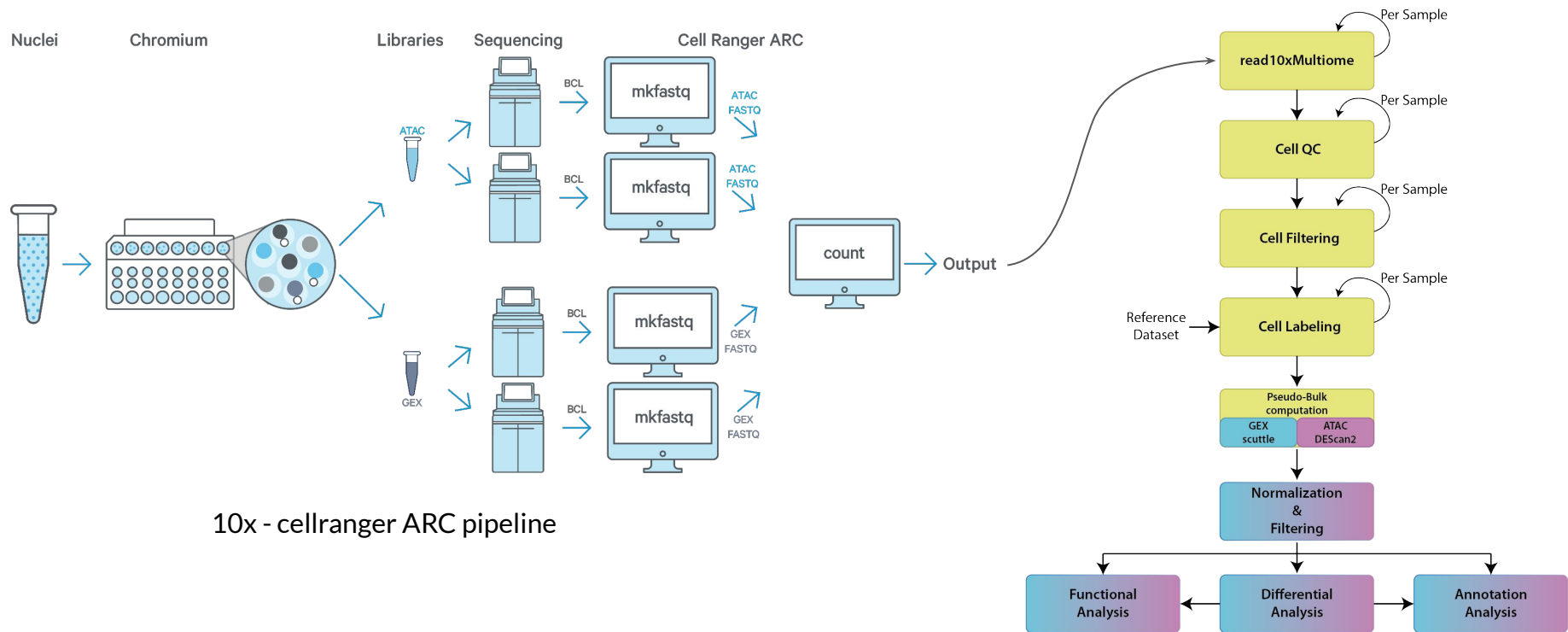


Figure 2. Efficient and robust workflow. Access a unified view of transcription and the chromatin landscape by combining gene expression and ATAC-seq data from the same single cell with a simple, streamlined workflow. Starting with a single nuclei suspension, transposition is performed in bulk before individual nuclei are captured in GEMs (Gel Bead-in-emulsion), where DNA fragments and the 3' ends of mRNA are barcoded. Generate two complementary libraries from each sample, and link gene expression and open chromatin profiles back to the same cell with certainty.

10X Cell Ranger pipeline

Output Files format details →		
Loupe Browser visualization file (CLOUPE)	1.40 GB	49d0621186974c368442ee10008de461
Secondary analysis outputs (DIR)	383.16 MB	6c1b8f17fc38b457588b565846438597
Run summary (HTML)	5.68 MB	36f5b9f0d27b12ae194476fea4f47ea7
Run summary metrics (CSV)	2 KB	755d8e040aaf159e9d99aefcc52bc953
Per barcode summary metrics (CSV)	85.51 MB	356b41d11800316657573798e48ec18f
Filtered feature barcode matrix MEX (DIR)	396.43 MB	b99400e8000dec92c1b6c9b080e9b01a
Filtered feature barcode matrix (HDF5)	158.62 MB	cee000a98c8a05d0456add3ff864cdb0
Raw feature barcode matrix MEX (DIR)	483.21 MB	f1f78862adaccb6a614803239bd3bd70
Raw feature barcode matrix (HDF5)	195.95 MB	8b27a5fff23f82b60b062399ca29917e
GEX Position-sorted alignments (BAM)	31.76 GB	29c8674d7e45229f0a3eb777f3bfd78c
GEX Position-sorted alignments (BAM index)	10.69 MB	d4e6b5b4720c75c14322ee262d9e16e5
GEX Per molecule information file (HDF5)	182.94 MB	aadd4323de6f46f00076d4989fa2357a
ATAC Position-sorted alignments (BAM)	57.87 GB	965f5f3df8167d342b55a9ad2e8f99bc
ATAC Position-sorted alignments (BAM index)	8.47 MB	fa7f7688587338b9fce01cc845ff444a
ATAC Per fragment information file (TSV.GZ)	3.30 GB	d0046f31a25b076ec389dcf9c017763b
ATAC Per fragment information index (TSV.GZ index)	1.44 MB	7837beb5ed0dc2e31d972cc012d962cf
ATAC peak locations (BED)	2.55 MB	22bf997edb4ce5c6e28490c632b3e2d4
ATAC smoothed transposition site track (BIGWIG)	2.56 GB	6171046154084e2c8c76d16a8da0aff0
ATAC peak annotations based on proximal genes (TSV)	5.90 MB	f6fd80ca3b5c33d50e9f1e25cad38b1b

The 10X Multiome protocol and analysis



The Data

Already available official datasets from the 10x Genomics website.

- A “general” brain dataset:
 - two replicates of the same brain (Brain1, Brain2)
 - cortex, hippocampus, ventricular zone, brain
 - differs for a few cells
 - they are practically the same sample
 - we use them to illustrate the general usage of the TENxMultiomeTools pipeline
- An Alzheimer Disease dataset:
 - twelve samples of two genotypes
 - Alzheimer Disease (AD) and Wild Type (WT)
 - Three different time-points (tp)
 - At least 2 replicates per each tp
 - the released dataset is already in an aggregated form
 - we don't have row data files (i.e. BAM)
 - we use it to show the DARs detection



The Data

Already available official datasets from the 10x Genomics website.

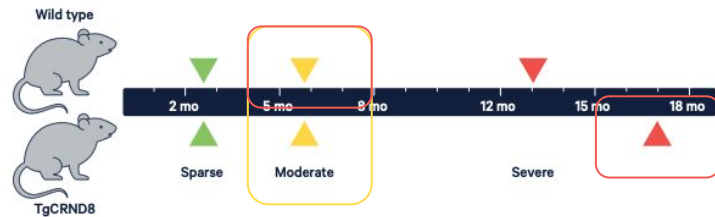
- A “general” brain dataset:
 - two replicates of the same brain (Brain1, Brain2)
 - cortex, hippocampus, ventricular zone, brain
 - differs for a few cells
 - they are practically the same sample
 - we use them to illustrate the general usage of the TENxMultiomeTools pipeline
- An Alzheimer Disease dataset:
 - twelve samples of two genotypes
 - Alzheimer Disease (AD) and Wild Type (WT)
 - Three different time-points (tp)
 - At least 2 replicates per each tp
 - the released dataset is already in an aggregated form
 - we don't have row data files (i.e. BAM)
 - we use it to show the DARs detection



The Data

Already available official datasets from the 10x Genomics website.

- A “general” brain dataset:
 - two replicates of the same brain (Brain1, Brain2)
 - cortex, hippocampus, ventricular zone, brain
 - differs for a few cells
 - they are practically the same sample
 - we use them to illustrate the general usage of the TENxMultiomeTools pipeline
- An Alzheimer Disease dataset:
 - twelve samples of two genotypes
 - Alzheimer Disease (AD) and Wild Type (WT)
 - Three different time-points (tp)
 - At least 2 replicates per each tp
 - the released dataset is already in an aggregated form
 - we don't have row data files (i.e. BAM)
 - we use it to show the DARs detection



Future Plans

- Facilitate the execution of the package with multiple replicates
 - list of SingleCellExperiments or single object with all replicates
- Reduce the number of external dependencies for BAM generation
 - i.e. use the fragment files for computing the pseudobulk
 - already suggested in <https://github.com/drighelli/TENxMultiomeTools/issues/1> by Dr. Michael Stadler
- Add other packages for better comparison during the analysis
 - azimuth for the cell type labeling
 - doublets identification both on RNA and ATAC
 - ...
- submit package to bioconductor

Additional links

- package: <https://github.com/drighelli/TENxMultiomeTools>
- analysis rmd/html file:
https://github.com/drighelli/EuroBioc23_ATAC_multome
- 10x multiome datasets from official website
 - Brain rep1: [here](#)
 - Brain rep2: [here](#)
 - Alzheimer Aggregated: [here](#)

Acknowledgements

Risso Lab @ University of Padova
Dr. Davide Risso
Elena Zuin

Peixoto Lab @ Washington State University
Dr. Lucia Peixoto

Pombo Lab @ Max Delbrück Center, Berlin
Dr. Ana Pombo
Dominika Vojtasova

Chan
Zuckerberg
Initiative 

