

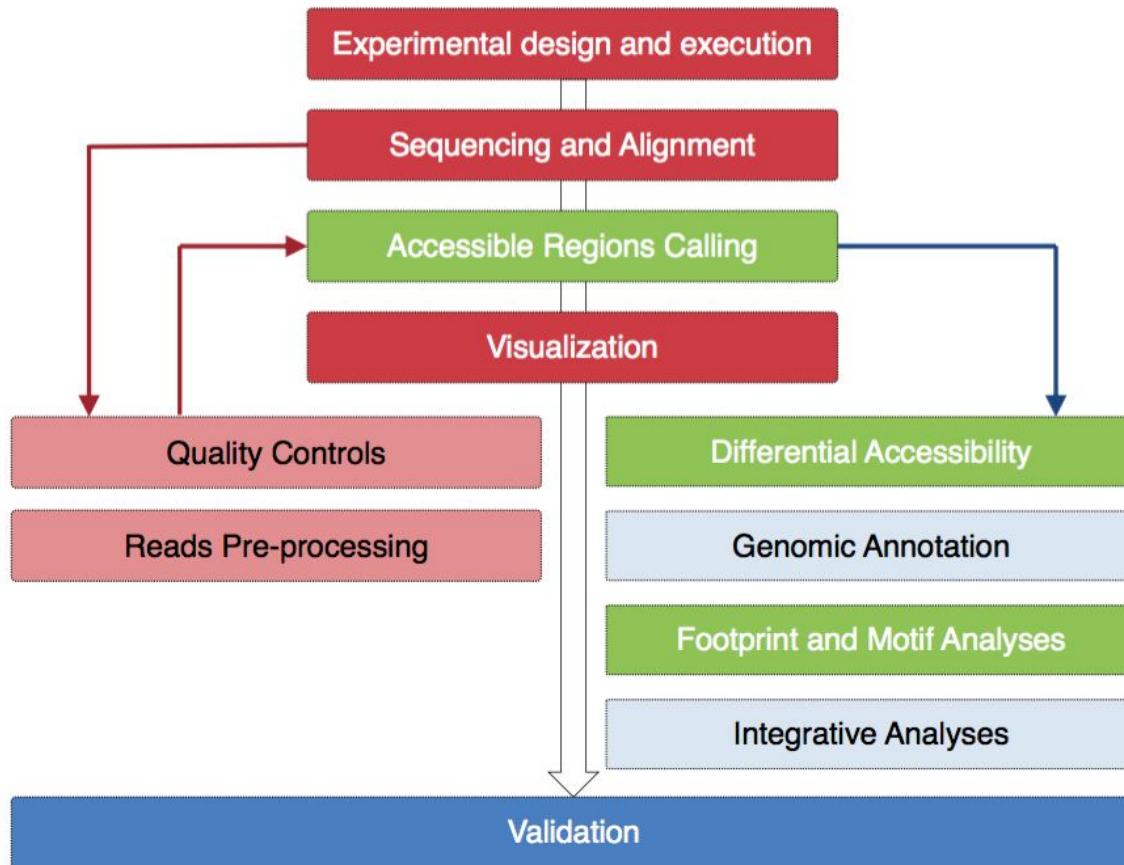
# Single-cell epigenomics

## Practice - sciATAC-seq

*Ecole thématique 2019 / sincellTE*

Morgane Thomas-Chollier,  
Nicolas Servant, Nathalie Lehmann

# Data processing: general workflow



# Single-cell epigenomic public pipelines

**SCREW** - single cell DNA methylation reproducible workflow:

<https://github.com/Epigenomics-Screw/Screw>

**SCASAT** - single cell ATAC-seq preprocessing and analysis pipeline:

<https://github.com/ManchesterBioinference/Scasat>

**Dr.seq2** - QC and analysis pipeline for ATAC-seq and Drop-ChIP (among others)

<https://github.com/ChengchenZhao/DrSeq2>

# Specific bioinformatics tools for scATAC-seq

**Destin** (nov. 2018) <https://www.biorxiv.org/content/10.1101/461905v1>

**Cis-topic** <https://www.biorxiv.org/content/10.1101/370346v1>

**ChromVar** (Schep et al, *Nature Methods*, 2017)

Performs clustering, motifs discovery, differential accessibility analysis

<https://github.com/GreenleafLab/chromVAR>

**scABC** (Zamanighomi et al, *BiorXiv*, 2017)

Unsupervised clustering to classify scATAC-seq data and discover regions of open chromatin specific to cell identity

<https://github.com/timydaley/scABC>

**Cicero** (Pliner et al, *BiorXiv*, 2017)

Statistical method that connects regulatory elements to target genes

# Aim of the practice

- The goal is to get a first try at manipulating scATAC-seq data
- The tutorial was made by the authors of a recent sciATAC-seq paper (E. Furlong's lab)
  - Chromatin Cell Atlas of the Developing Fly Embryo
- WARNING : We have tested and commented this published code

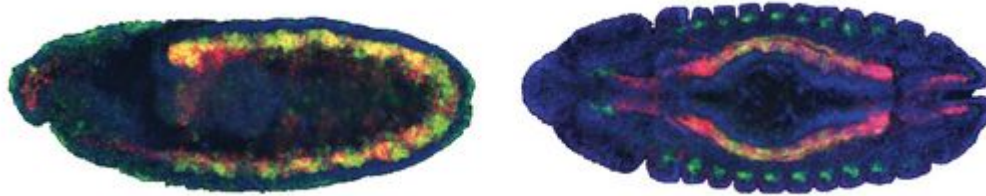
## LETTER

\*Cusanovich et al, Nature, 2018

doi:10.1101/251111

### The *cis*-regulatory dynamics of embryonic development at single-cell resolution

Darren A. Cusanovich<sup>1\*</sup>, James P. Reddington<sup>2\*</sup>, David A. Garfield<sup>2†\*</sup>, Riza M. Daza<sup>1</sup>, Delasa Aghamirzaie<sup>1</sup>, Raquel Marco-Ferreres<sup>2</sup>, Hannah A. Pliner<sup>1</sup>, Lena Christiansen<sup>3</sup>, Xiaojie Qiu<sup>1</sup>, Frank J. Steemers<sup>3</sup>, Cole Trapnell<sup>1</sup>, Jay Shendure<sup>1,4§</sup> & Eileen E. M. Furlong<sup>2§</sup>



sciATAC-seq

20,000 cells - 3 embryonic stages

*Drosophila melanogaster*

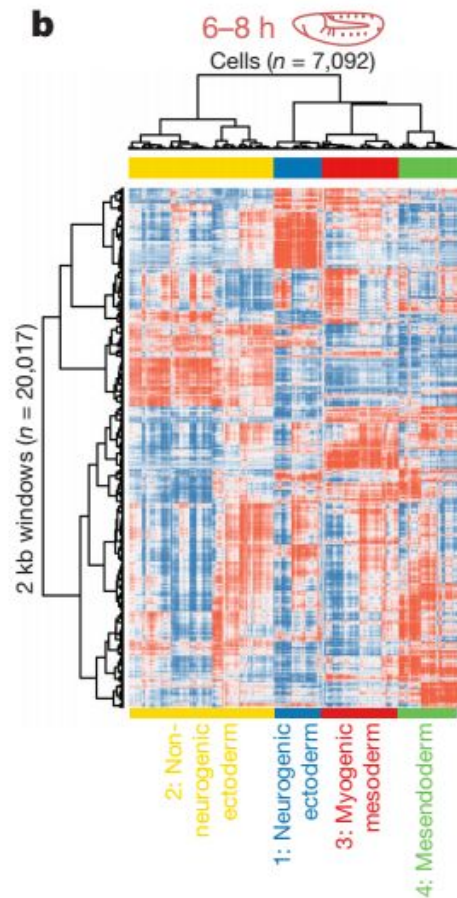
Web access to the Chromatin Cell Atlas of the Developing Fly Embryo (shiny app):

<http://shiny.furlonglab.embl.de/scATACseqBrowser/>

# Let's have a closer look on the data

- Data originates from the Chromatin Cell Atlas of the Developing Fly Embryo
  - 3 different stages of development ( 2-4, 6-8, 10-12 hours after egg laying)
  - More than 20,000 nuclei among the 3 stages
  - From the binary matrix (genomic loci x cells) of 6-8h developmental stage
- Analyses that can be reproduced:
  - preprocessing (fastq, barcodes processing, alignment, count reads overlapping windows => produce window matrix):  
<https://github.com/shendurelab/fly-atac>
  - Most figures of the paper: <https://shendurelab.github.io/fly-atac/docs/>

# Use case 1 : Cell clustering based on patterns of chromatin accessibility



open this website :

<https://shendurelab.github.io/fly-atac/docs>

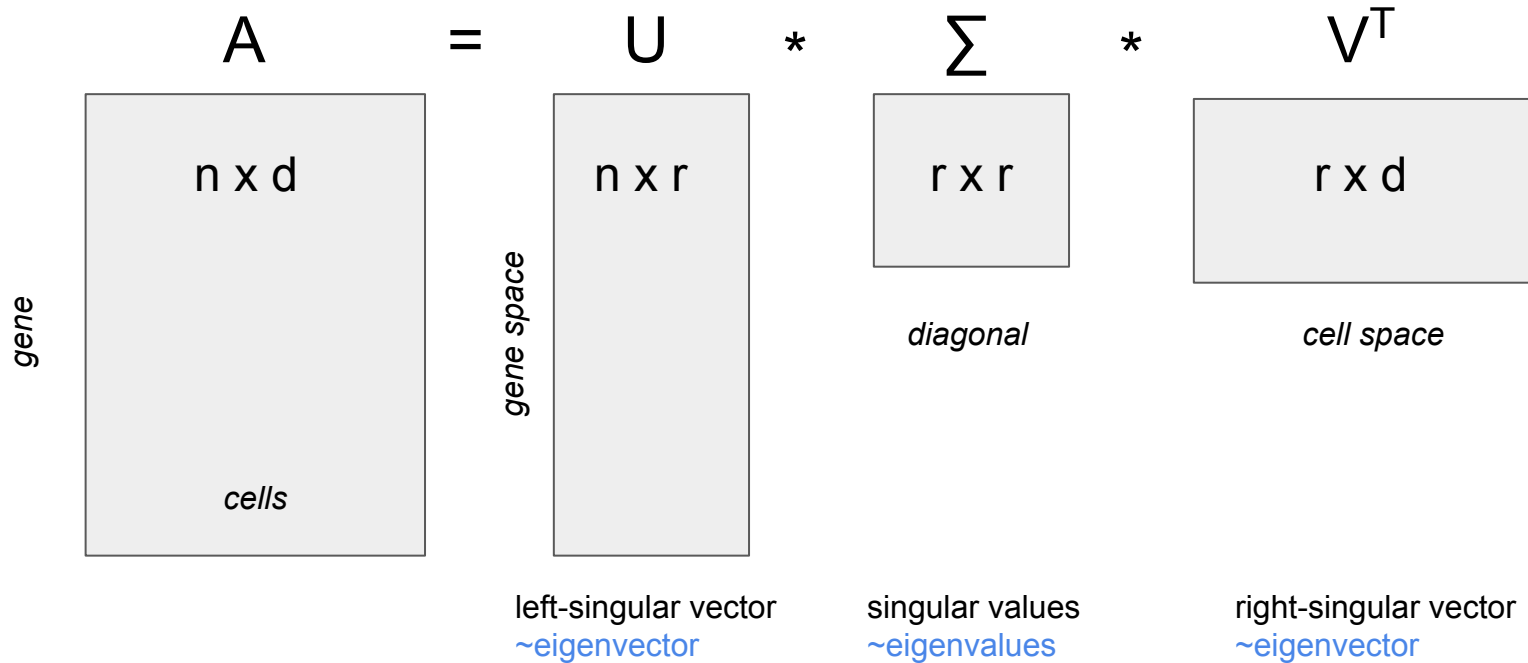
connect to R :

<http://rstudio-singlecell.sb-roscoff.fr/>



# What is SVD ?

Single Value Decomposition = dimension reduction method use in [Principal Component Analysis](#)



# Clustering and LSI

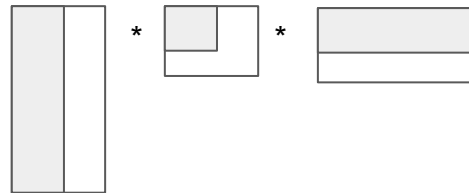


Weight the peaks according to their significance

**TF** = Term Frequency  
*How frequent is a peak in a cell*

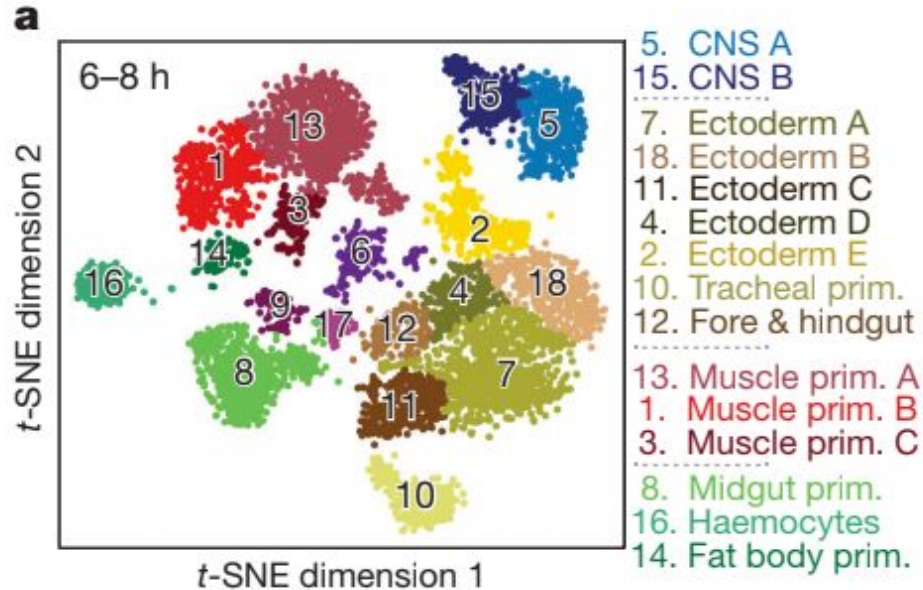
**IDF** = Inverse Document Frequency  
*How important are the peaks in the population of cells*

Partial SVD (fast approximation for sparse matrix)



More info on LSI in dimension reduction: <https://www.youtube.com/watch?v=Bxlzbck51SU>

## Use case 2 : Dimension reduction with t-SNE



# Another single cell R object: CellDataSet (monocle)

- The CellDataSet object was derived from the ExpressionSet class

- E.g.:

```
cds <- newCellDataSet(exprs(...),  
                      phenoData = ..., featureData = ...)
```

```
slotNames(cds)
```

```
?CellDataSet
```

```
head(fData(cds))
```

```
head(pData(cds))
```