BMEG3105: Data analytics for personalized genomics and precision medicine – Lecture 18

Topic: Single cell genomics

Lecturer: Prof. LI Yu

Date: 4 Nov 2022

Content of Lecture 18

- Epigenetics
- Single cell RNA sequencing (RNA-seq)

Epigenetics

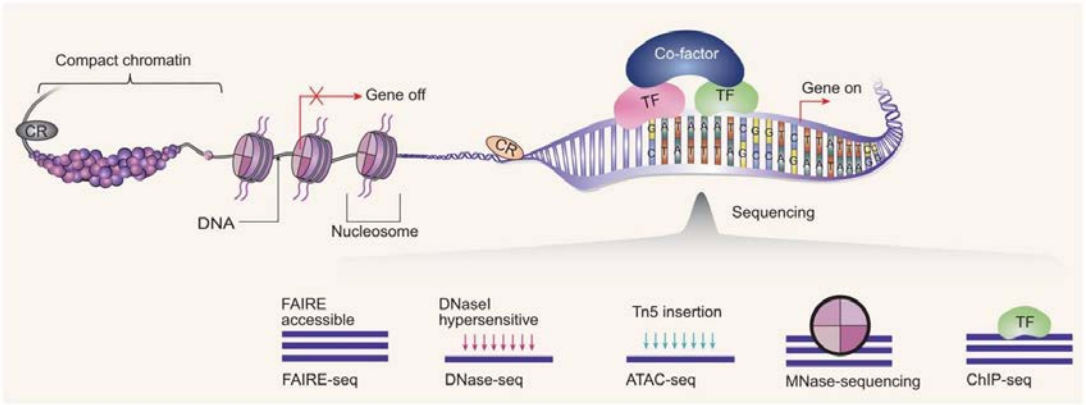
Gene expression regulation: structure and environment

Epigenetics is the study of how your **behaviors and environment** can cause changes that affect the way your genes work. Unlike genetic changes, epigenetic changes are reversible and **do not change your DNA sequence**, but they can **change how your body reads a DNA sequence**.

Mechanisms of epigenetic regulation:

1. DNA Methylation
2. Chromatin Modification
3. DNase I hypersensitive sites – DNA sites bound by a transcription factor of interest

Sequencing protocols:

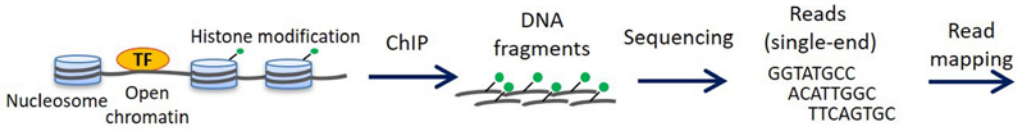


The structure of chromatin (DNA and its organizing proteins) can be regulated. More open or "relaxed" chromatin makes a gene more available for transcription.

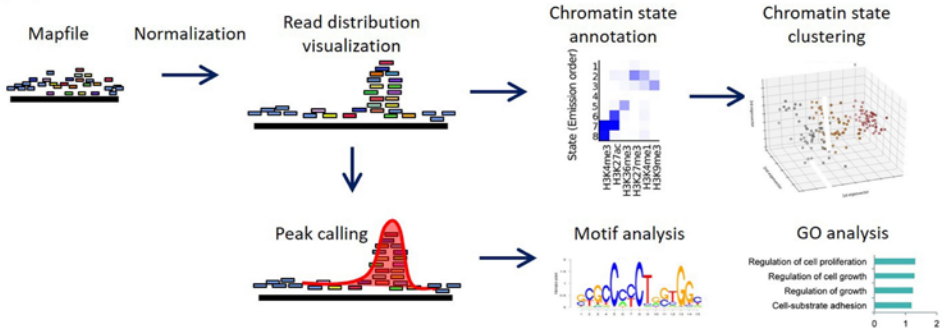
Common sequencing method for Chromatin Changes and Histone Modification:
FAIRE-seq, DNase-seq, ATAC-seq, MNase-sequencing, ChIP-seq

Data analytics pipeline

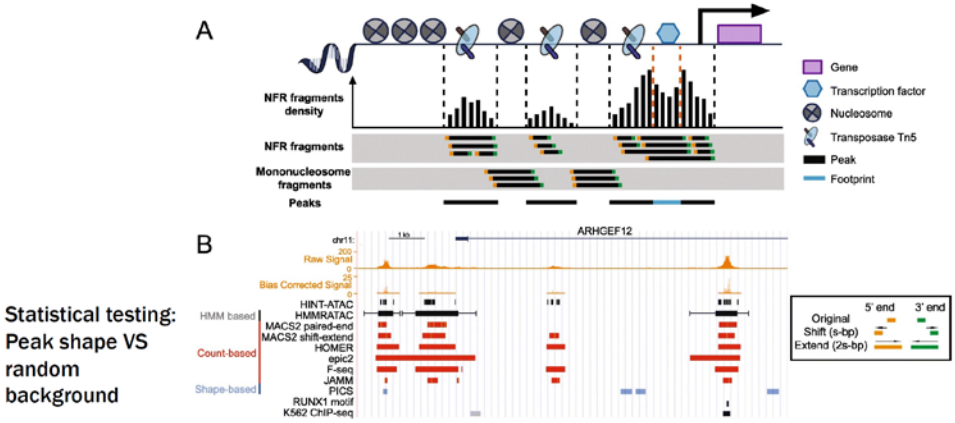
1. Sample preparation and sequencing



2. Computational analysis (downstream analysis)



Peak calling (statistical testing):



Statistical testing: Peak shape VS random background

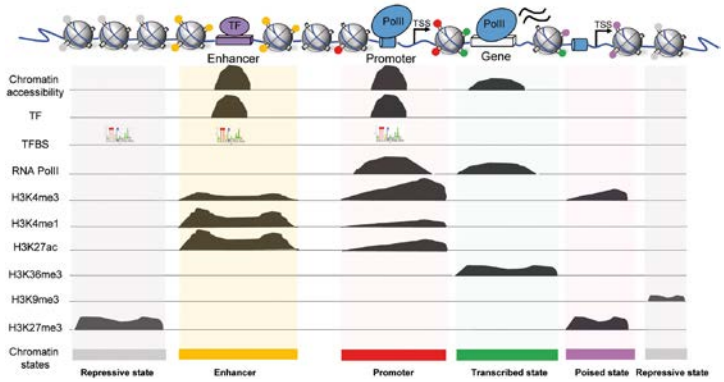
One of the first steps in the ChIP-seq analysis is peak calling. Peak calling is a statistical procedure, which uses coverage properties of ChIP and Input samples to find regions which are enriched due to protein binding. The procedure requires mapped reads, and outputs a set of regions, which represent the putative binding locations. Each region is usually associated with a significance score which is an indicator of enrichment.

Peak calling output-BED file:

```
track name="ItemRGBDemo" description="Item RGB demonstration" visibility=2 itemRgb="On"
chr7    127471196    127472363    Pos1    0    +    127471196    127472363    255,0,0
chr7    127472363    127473530    Pos2    0    +    127472363    127473530    255,0,0
chr7    127473530    127474697    Pos3    0    +    127473530    127474697    255,0,0
chr7    127474697    127475864    Pos4    0    +    127474697    127475864    255,0,0
chr7    127475864    127477031    Neg1    0    -    127475864    127477031    0,0,255
chr7    127477031    127478198    Neg2    0    -    127477031    127478198    0,0,255
chr7    127478198    127479365    Neg3    0    -    127478198    127479365    0,0,255
chr7    127479365    127480532    Pos5    0    +    127479365    127480532    255,0,0
chr7    127480532    127481699    Neg4    0    -    127480532    127481699    0,0,255
```

Browser Extensible Data (BED) format: Chromosome, Start, End, Label, …

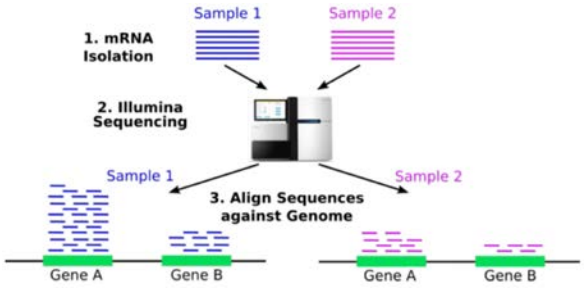
Histone marks and chromatin accessibility:



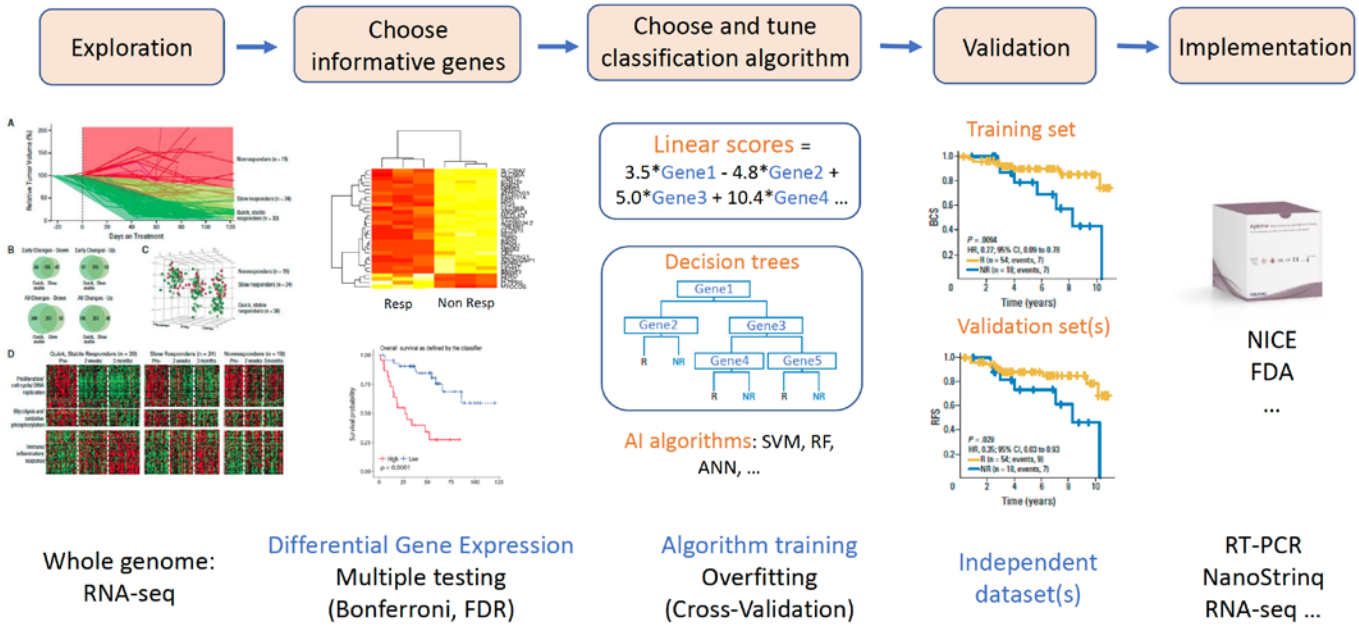
Enhancer, Promoter, Gene is opened area and, hence having peak.

Single cell RNA sequencing

Why single cell?

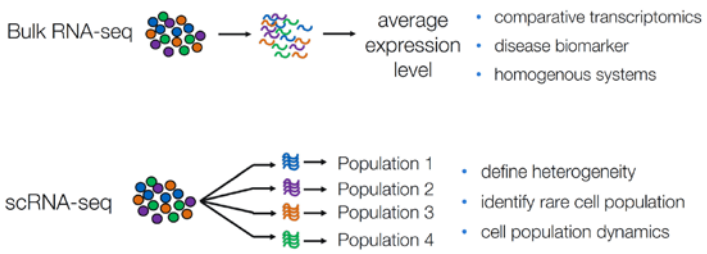


With single-cell RNA sequencing, it is now possible to analyze the transcriptome at single-cell level for over millions of cells in a single study. This allows us to classify, characterize and distinguish each cell at the transcriptome level, which leads to identify rare cell population but functionally important.



The above diagram shows the transcriptome pipeline.

Bulk RNA-seq vs scRNA-seq :



Since using bulk RNA-seq for a heterogeneous system will not give you the right information about a type of cell within the sample. Therefore, we need single-cell RNA-seq (scRNA-seq).

What is single-cell RNA-seq?

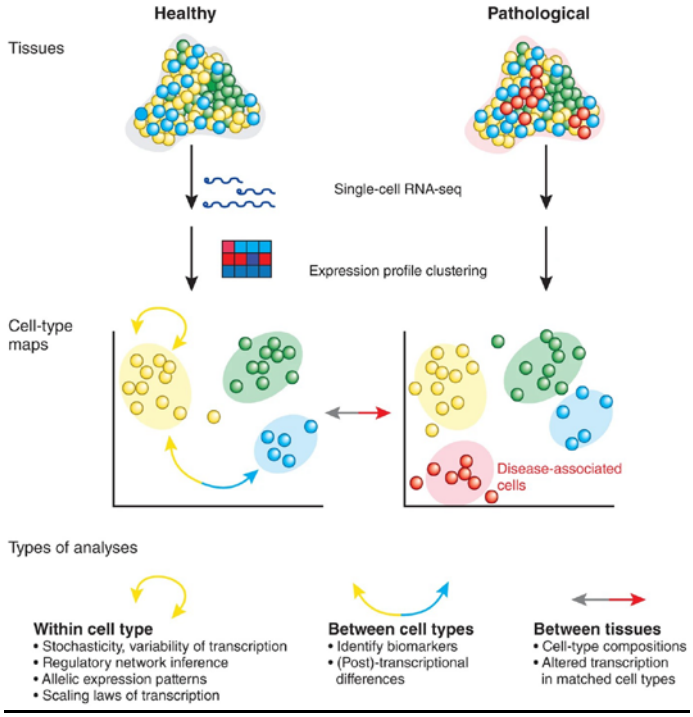
Single cell sequencing examines the sequence information from individual cells with optimized next-generation sequencing (NGS) technologies, providing a higher resolution of cellular differences and a better understanding of the function of an individual cell in the context of its microenvironment.

What can single-cell sequencing do?

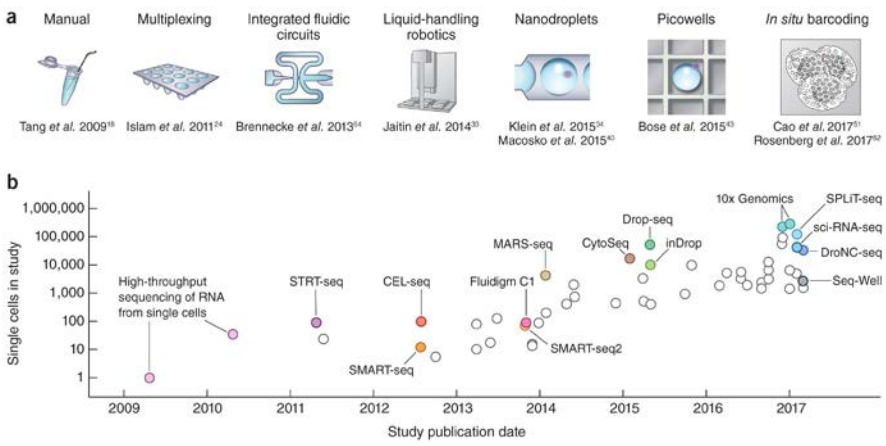
- Understanding heterogeneous tissues
- Identification and analysis of rare cell types
- Changes in cellular composition
- Dissection of temporal changes

Example of applications:
Differentiation paths; Cancer heterogeneity; Neural cell classification;
Embryonic development; Drug treatment response

Steps of single-cell sequencing:

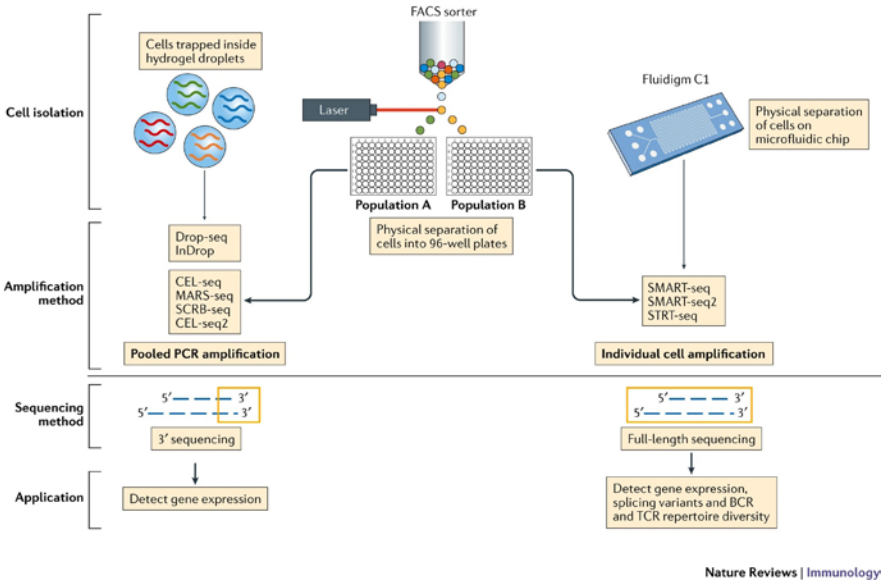


How to get the single-cell RNA-seq data?



How to get the single-cell RNA-seq data?

The following shows one of ways to get single cell sequence.



The table compare different approaches with cell capture method, no. of cell per experiment, cost and sensitivity.

Paplexi et al. 2017	FACS	CyTOF	qPCR	Plate-based protocols (STRT-seq, SMART-seq, SMART-seq2)	Fluidigm C1	Pooled approaches (CEL-seq, MARS-seq, SCRB-seq, CEL-seq2)	Massively parallel approaches (Drop-seq, InDrop)
Cell capture method	Laser	Mass cytometry	Micropipettes	FACS	Microfluidics	FACS	Microdroplets
Number of cells per experiment	Millions	Millions	300–1,000	50–500 ★	48–96 ✖	500–2,000 ★	5,000–10,000 ★
Cost	$0.05 per cell	$35 per cell	$1 per cell	$3–6 per well	$35 per cell	$3–6 per well	$0.05 per cell
Sensitivity	Up to 17 markers ✖	Up to 40 markers ✖	10–30 genes per cell ✖	7,000–10,000 genes per cell for cell lines; 2,000–6,000 genes per cell for primary cells	6,000–9,00 genes per cell for cell lines; 1,000–5,000 genes per cell for primary cells	7,000–10,000 genes per cell for cell lines; 2,000–6,000 genes per cell for primary cells	5,000 genes per cell for cell lines; 1,000–3,000 genes per cell for primary cells