KSBi-BIML 2024

Bioinformatics & Machine Learning(BIML) Workshop for Life and Medical Scientists

생명정보학 & 머신러닝 워크샵 (오프라인)

Single Cell Multiomics Analysis

김준일 _ 숭실대학교

KSBI KOREAN SOCIETY FOR BIOINFORMATICS | 한국생명정보학회

본 강의 자료는 한국생명정보학회가 주관하는 BIML 2024 워크샵 오프라인 수업을 목적으로 제작된 것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다.

이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없으며 만약 이러한 사항을 위반할 경우 발생하는 **모든 법적 책임은 전적으로 불법 행위자 본인에게 있음을 경고**합니다.

KSBi-BIML 2024

Bioinformatics & Machine Learning(BIML)
Workshop for Life and Medical Scientists

안녕하십니까?

한국생명정보학회가 개최하는 동계 교육 워크샵인 BIML-2024에 여러분을 초대합니다. 생명정보학 분야의 연구자들에게 최신 동향의 데이터 분석기술을 이론과 실습을 겸비해 전달하고자 도입한 전문 교육 프로그램인 BIML 워크샵은 2015년에 시작하여 올해로 벌써 10년 차를 맞이하게 되었습니다. BIML 워크샵은 국내 생명정보학 분야의 최초이자 최고 수준의 교육프로그램으로 크게 인공지능과 생명정보분석 두 개의 분야로 구성되어 있습니다. 올해 인공지능 분야에서는 최근 생명정보 분석에서도 응용이 확대되고 있는 다양한 인공지능 기반 자료모델링 기법들에 대한 현장 강의가 진행될 예정이며, 관련하여 심층학습을 이용한 단백질구조예측, 유전체분석, 신약개발에 대한 이론과 실습 강의가 함께 제공될 예정입니다. 또한 단일세포오믹스, 공간오믹스, 메타오믹스, 그리고 롱리드염기서열 자료 분석에 대한 현장 강의는 많은 연구자의 연구 수월성 확보에 큰 도움을 줄 것으로 기대하고 있습니다.

올해 BIML의 가장 큰 변화는 최근 연구 수요가 급증하고 있는 의료정보자료 분석에 대한 현장 강의를 추가하였다는 것입니다. 특히 의료정보자료 분석을 많이 수행하시는 의과학자 및 의료정보 연구자들께서 본 강좌를 통해 많은 도움을 받으실 수 있기를 기대하고 있습니다. 또한 다양한 생명정보학 분야에 대한 온라인 강좌 프로그램도 점차 증가하고 있는 생명정보 분석기술의 다양화에 발맞추기 위해 작년과 비교해 5강좌 이상을 신규로 추가했습니다. 올해는 무료 강좌 5개를 포함하여 35개 이상의 온라인 강좌가 개설되어 제공되며, 연구 주제에 따른 연관된 강좌 추천 및 강연료 할인 프로그램도 제공되며, 온라인을 통한 Q&A 세션도 마련될 예정입니다. BIML-2024는 국내 주요 연구중심 대학의 전임 교원이자 각 분야 최고 전문가들의 강의로 구성되었기에 해당 분야의 기초부터 최신 연구 동향까지 포함하는 수준 높은 내용의 강의가 될 것이라 확신합니다.

BIML-2024을 준비하기까지 너무나 많은 수고를 해주신 운영위원회의 정성원, 우현구, 백대현, 김태민, 김준일, 김상우, 장혜식, 박종은 교수님과 KOBIC 이병욱 박사님께 커다란 감사를 드립니다. 마지막으로 부족한 시간에도 불구하고 강의 부탁을 흔쾌히 허락하시고 훌륭한 현장 강의와 온라인 강의를 준비하시는데 노고를 아끼지 않으신 모든 강사분들께 깊은 감사를 드립니다.

2024년 2월

한국생명정보학회장 이 인 석

강의 시간표

DAY1 : 2월 24일 (토)

시간	강 의 (자연과학대학 28동 101호)
12:30-12:50	등록
12:50-13:00	공지사항 전달
13:00-14:30	**의료빅데이터/인공지능 총론** 김헌성 교수(가톨릭대학교)
14:30-14:45	휴식
14:45-16:15	**의료영상 인공지능의 이해 및 의료영상 레이블링 실습** 백서연 교수(연석대학교)
16:15-16:30	휴식
16:30-18:00	**의료 정보처리 자동화 실습 / 독자적인 어플리케이션 만들기** 김선근 대표(원닥 주식회사), 서사도 조교

시간	강 의 (자연과학대학 28동 102호)
12:30-12:50	등록
12:50-13:00	공지사항 전달
13:00-14:20	**EMR 데이터를 활용한 머신러닝 기반 예후예측:** **Decision Tree-based Models + EMR 샘플 데이터 실습 (MIMIC sample dataset)** 고태훈 교수(가톨릭대학교)
14:20-14:40	휴식
14:40-16:00	**Chest X-ray 영상을 활용한 딥러닝 기반 폐질환 진단:** **Convolutional Neural Network + 의료영상 샘플 데이터 실습 (NIH Chest X-ray14)** 고태훈 교수(가톨릭대학교)
16:00-16:20	휴식
16:20-17:40	**심전도 데이터를 활용한 딥러닝 기반 부정맥 탐지: Recurrent Neural Network +** **Transformer + 심전도 샘플 데이터 실습 (MIT-BIH Arrhythmia Database)** 고태훈 교수(가톨릭대학교)

DAY1 : 2월 26일 (월)

시간	강 의 (자연과학대학 28동 101호)
09:00-09:20	등록
09:20-09:30	공지사항 전달
09:30-10:50	**DNN (이론)** 이상근 교수(고려대학교)
10:50-11:00	휴식
11:00-12:10	**CNN (이론)** 이상근 교수(고려대학교)
12:10-13:40	점심
13:40-15:10	**RNN, ChatGPT, XAI (이론)** 이상근 교수(고려대학교)
15:10-15:20	휴식
15:20-16:50	**CNN/RNN 모델 구조 정의, 학습 알고리즘 적용, 성능 평가, 시각화 방법 (Tensorflow 실습)** 이정현 조교, 한성민 조교

시간	강 의 (자연과학대학 28동 102호)
09:00-09:20	등록
09:20-09:30	공지사항 전달
09:30-11:00	**Best practice for single-cell data analysis** 박종은 교수(KAIST)
11:00-11:10	휴식
11:10-12:40	**Practice1: Scanpy basic workflow** 정성민 조교, 고용준 조교
12:40-14:10	점심
14:10-15:30	**Public database, data integration, reference mapping, multiomics** 박종은 교수(KAIST)
15:30-15:40	휴식
15:40-16:50	**Practice2: Advanced single-cell analysis (siVI universe)** 정성민 조교, 고용준 조교

DAY1 : 2월 27일 (화)

시간	강 의 (자연과학대학 28동 101호)
09:00-09:20	등록
09:20-09:30	공지사항 전달
09:30-10:50	**AI-based protein structure prediction** **- Intro to protein structure prediction** **- Early AI-based approaches** **- AlphaFold and RoseTTAFold** 백민경 교수(서울대학교)
10:50-11:00	휴식
11:00-12:10	**단백질 구조 예측 실습** **- ColabFold를 활용한 단백질 구조 및 상호작용 예측** **- Tips &Tricks for better structure modeling** 백민경 교수(서울대학교)
12:10-13:40	점심
13:40-15:10	**AI-based protein design** **- Intro to protein design** **- Protein backbone design using RFdiffusion** **- Protein sequence design using ProteinMPNN** 백민경 교수(서울대학교)
15:10-15:20	휴식
15:20-16:50	**단백질 디자인 실습** **- RFdiffusion 및 ProteinMPNN의 활용법 실습** 백민경 교수(서울대학교)

시간	강 의 (자연과학대학 28동 102호)
09:00-09:20	등록
09:20-09:30	공지사항 전달
09:30-11:00	**Introduction to Single-cell biology** 최정민 교수(고려대학교)
11:00-11:10	휴식
11:10-12:40	**i. Unsupervised Spatial transcriptome analysis** **ii. Tumor Boundary Determination in Spatial Transcriptomics** 유광민 조교, 이문영 조교
12:40-14:10	점심
14:10-15:30	**i. Deconvolution Analysis Using Single-cell RNA Sequencing and Spatial Transcriptomics** **ii. Cell-Cell Interaction Analysis in Spatial Transcriptomics** 김지현 조교, 최승지 조교
15:30-15:40	휴식
15:40-16:50	**i. Open Chromatin Region Analysis and Biological Interpretation of Using scATAC-seq Dataset** **ii. Construction of Gene Regulatory Networks Based on Integrated Analysis of scATAC-seq and scRNA-seq Datasets** 천하림 조교, 이호진 조교

DAY1 : 2월 28일 (수)

시간	강 의 (자연과학대학 28동 101호)
09:00-09:20	등록
09:20-09:30	공지사항 전달
09:30-11:00	**Introduction to Transformers (이론)** 전민지 교수 (고려대학교)
11:00-11:10	휴식
11:10-12:40	**Introduction to Transformers (실습)** 봉현수 조교, 임우택 조교
12:40-14:10	점심
14:10-15:40	**Deep learning in Bioinformatics** 노미나 교수(한양대학교)
15:40-15:50	휴식
15:50-17:20	**Deep learning model을 이용한 실습** 박예솔 조교

시간	강 의 (자연과학대학 28동 102호)
09:00-09:20	등록
09:20-09:30	공지사항 전달
09:30-10:50	**마이크로바이옴 기본 이론** 이선재 교수(GIST)
10:50-11:00	휴식
11:00-12:10	**16S rRNA amplicon seq. - DADA2** 조준우 조교, 백재우 조교
12:10-13:40	점심
13:40-14:40	**최신 메타지놈 분석 기법의 현황** 이선재 교수(GIST)
14:40-14:50	휴식
14:50-16:50	**Shotgun metagenome 분석 (Linux)** 조준우 조교, 백재우 조교

DAY1 : 2월 29일 (목)

시간	강 의 (자연과학대학 28동 101호)
09:00-09:20	등록
09:20-09:30	공지사항 전달
09:30-10:50	**화학정보학 기초(Cheminformatics) / 약물특성 및 약물다움(druglikeness)** **Molecular Notations &Descriptors / AI 신약개발을 위한 Databases** **AI 신약개발을 위한 Programming 기초** 김동섭 교수(KAIST)
10:50-11:00	휴식
11:00-12:10	**Google Colab에 RDKit 설치 / 화합물 정보 읽기 실습** **Bioactivity database 검색 및 정보 읽기 실습** **Molecular descriptor (fingerprint) 생성 및 similarity 계산 실습** 정수재 조교, 나민주 조교
12:10-13:40	점심
13:40-15:10	**AI 신약개발을 위한 기계학습법 기초 / QSAR 모델링 기초 / AI 신약개발을 위한 딥러닝 모델** **Virtual screening (ligand-based, structure-based) 및 de novo design** 김동섭 교수(KAIST)
15:10-15:20	휴식
15:20-16:50	**QSAR modeling 전체 과정 실습/ 화합물의 Bioactivity 예측 모델 개발** **Virtual screening 과정을 통한 신약후보물질 발굴 실습** 정수재 조교, 나민주 조교

시간	강 의 (자연과학대학 28동 102호)
09:00-09:20	등록
09:20-09:30	공지사항 전달
09:30-11:00	**Single cell multiomics 이론 / Gene regulatory network 이론** 김준일 교수(숭실대학교)
11:00-11:10	휴식
11:10-12:40	**Seurat/Signac, ArchR, TENET+ 실습** 김현규 조교, 정회빈 조교
12:40-14:10	점심
14:10-15:40	**롱리드 시퀀싱 소개 및 유전체 조립 실습** 김준 교수(충남대학교)
15:40-15:50	휴식
15:50-17:20	**변이 분석 및 시각화 실습** 김준 교수(충남대학교)

강의개요

Single Cell Multiomics Analysis

단일세포에서 전사체를 분석하는 시대에서 멀티모달 데이터 즉, 여러 층위가 섞인 데이터를 분석하는 시대로 넘어가고 있다. 개별 세포에서 mRNA 발현뿐만 아니라 protein level, epigenetic change를 동시에 측정할 수 있게 되면서 세포 유형이나 상태를 다각도에서 특징지을 수 있게 되었다.

본 강의에서는 단일세포 멀티오믹스 데이터 중에서 단일세포전사체(single cell RNA sequencing)와 단일세포염색질접근성(single cell ATAC sequencing) 데이터를 통합 분석하는 방법을 배운다. 또한 이를 바탕으로 유전자조절네트워크를 재구성하여 유전자 발현과 후성유전학적 변화를 조절하는 핵심인자를 발굴하는 방법을 익힌다.

강의는 다음의 내용을 포함한다:

- 단일세포멀티오믹스 개요
- Seurat/Signac을 통한 멀티오믹스 데이터 통합
- ArchR를 통한 멀티오믹스 데이터 통합
- TENET+를 통한 유전자조절네트워크 구축

* 참고강의교재:

  Stuart, T., Srivastava, A., Madad, S. et al. Single-cell chromatin state analysis with Signac. Nat Methods 18, 1333–1341 (2021).

  Granja, J.M., Corces, M.R., Pierce, S.E. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet 53, 403–411 (2021).

* 교육생준비물:

  노트북 (메모리 16GB 이상, 디스크 여유공간 30GB 이상)

* 강의 난이도: 중급

* 강의: 김준일교수 (숭실대학교 생명정보학과)

Curriculum Vitae

Speaker Name: Junil Kim, Ph.D.

▶ **Personal Info**

Name	Junil Kim
Title	Assistant Professor
Affiliation	Soongsil University

▶ **Contact Information**

Address	369, Sangdo-Ro, Dongjak-Gu, Seoul, 06978
Email	junilkim@ssu.ac.kr
Phone Number	010-3140-6567

Research Interest

Single Cell Genomics, Systems Biology, Network Biology

Educational Experience

2005	B.S. in Bioinformatics, Soongsil University, Republic of Korea
2008	M.S. in Bioinformatics, Seoul National University, Republic of Korea
2014	Ph.D. in Bio and Brain Engineering, KAIST, Republic of Korea

Professional Experience

2014-2016	Postdoctoral Researcher, CHA Cancer Institute, CHA University, Republic of Korea
2016-2018	Postdoctoral Researcher, Perelman School of Medicine, University of Pennsylvania, USA
2018-2021	Postdoctoral Researcher, BRIC, University of Copenhagen, Denmark
2021-	Assistant Professor, School of Systems Biomedical Science, Soongsil University, Republic of Korea

Selected Publications (5 maximum)

1. **Junil Kim**, Michaela Mrugala Rothová, Esha Madan, Siyeon Rhee, Guangzheng Weng, António Palma, Linbu Liao, Eyal David, Ido Amit, Morteza Chalabi Hajkarim, Vignesh Vudatha, Andrés Gutiérrez-García, Eduardo Moreno, Robert Winn, Jose Trevino, Paul B. Fisher, Joshua M. Brickman, Rajan Gogna, Kyoung Jae Won, "Neighbor-specific gene expression revealed from physically interacting cells during mouse embryonic development", *PNAS* (**IF: 12.777**), Vol. 120, Issue 2, e22053711120, 3 Jan. 2023
2. Dongha Kim*, **Junil Kim***, Young Suk Yu, Yong Ryoul Kim, Sung-hee Baek, Kyoung-Jae Won, "Systemic approaches using single cell transcriptome reveal that C/EBPγ regulates autophagy under amino acid starved condition", *Nucleic Acids Research* (**IF: 19.160**), Vol. 50, Issue 13, 7298-7309, 22 July 2022 (*Co-first authors)

3. Guangzheng Weng, **Junil Kim***, and Kyoung Jae Won*, "VeTra: a tool for trajectory inference based on RNA velocity", Bioinformatics (**IF: 6.937**), btab364, May 2021. (*Co-corresponding authors)
4. **Junil Kim**, Simon T. Jakobsen, Kedar N. Natarajan, Kyoung Jae Won, "TENET: gene network reconstruction using transfer entropy reveals key regulatory factors from single cell transcriptomic data", *Nucleic Acids Research* (**IF: 16.971**), Vol. 49, No. 1, e1-e1, Jan. 2021.
5. **Junil Kim**, Diana E. Stanescu, and Kyoung Jae Won, "CellBIC: Bimodality-based top-down clustering of single-cell RNA sequencing data reveals hierarchical structure of the cell type", *Nucleic Acids Research* (**IF: 16.971**), Vol. 46, Issue 21, e124, Aug. 2018. (Google Scholar Citations: **8** / Web of Science Citations: **4**)

KSBI | 한국생명정보학회

KSBi-BIML 2024

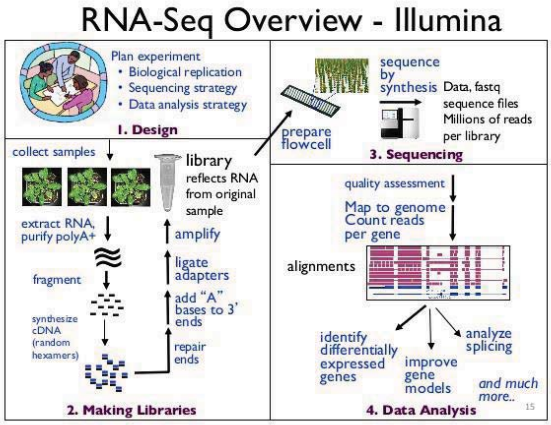
Single cell multiomics
& Gene regulatory network

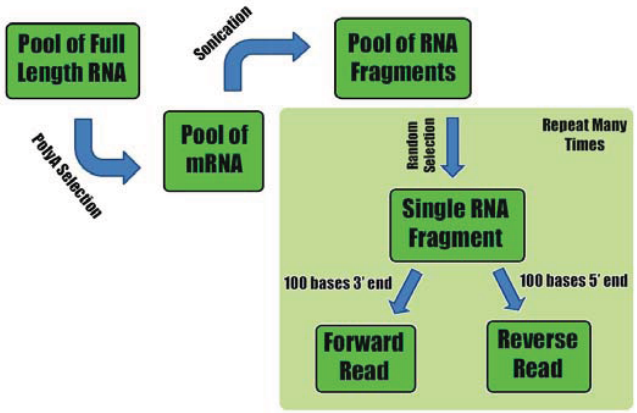
Contents

1. Bulk RNA sequencing
2. Single cell RNA sequencing
3. Single cell ATAC sequencing
4. Single cell multiomics
   - Overview
   - scRNA-seq + scATACseq
   - CITE-seq
5. Data integration strategy
   - Feature-based integration (Horizontal)
   - Cell-based integration (Vertical)
   - No anchor (Diagonal)
6. Gene regulatory network
   - Overview
   - SCENIC+
   - TENET+

2

1. Bulk RNA sequencing : mRNA sequencing exactly measures the quantity of mRNA molecule

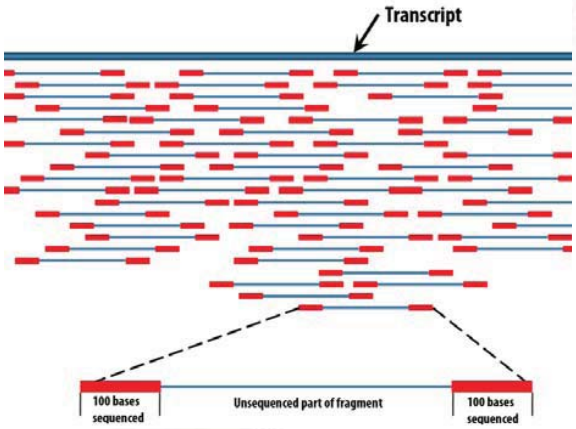




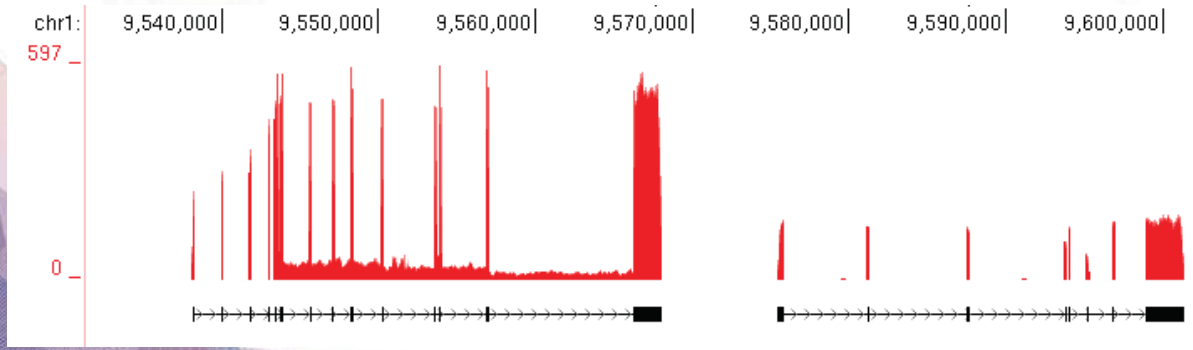
- *Full length* mRNA cannot yet be sequenced routinely (Illumina).
  - Only short fragments can be sequenced accurately and cheaply.
- RNA are fragmented into small pieces, typically 200 - 500 bases.
- Approximately **100 bases** are sequenced from one, or both, ends of the fragments.

3

1. Bulk RNA sequencing



- Reads are aligned to the genome.
- Data are represented as "depth of coverage" plots.
  - The height of the bar over a nucleotide is the number of reads which align across that location.
- The higher a gene is expressed, the more reads we find for that gene.
- The higher the peak, the higher the gene is expressed.



4

1. Bulk RNA sequencing: RNA-seq analysis pipeline

Raw Data (FASTQ file)

Quality Control (FASTQC, TrimGalore)

Alignment (STAR, TOPHAT)

Aligned Data (BAM, SAM file)

Quantification (HTSeq, Samtools)

Read count (text or excel file)

Normalization

RPKM, FPKM, TPM (text or excel file)

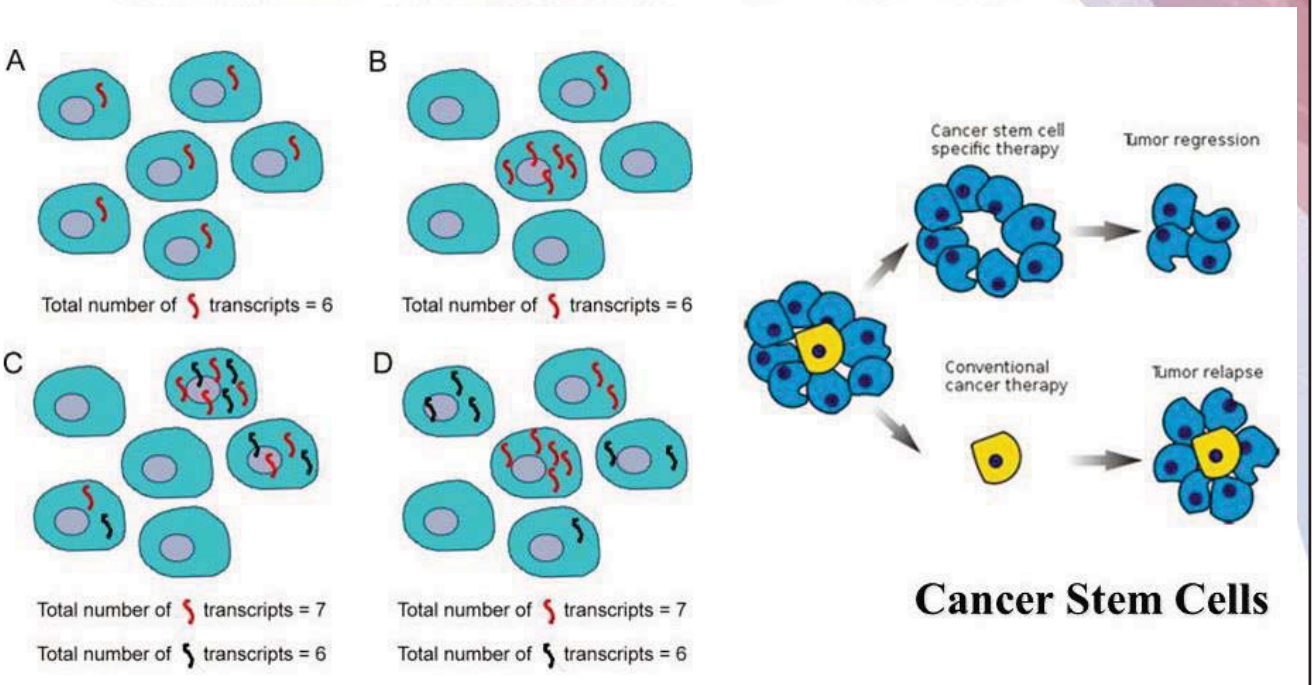
Differential Expression Analysis (DESeq2, edgeR)

DEGs (text or excel file)

Further analysis (GSEA, Pathway, Alternative splicing, Variant)

5

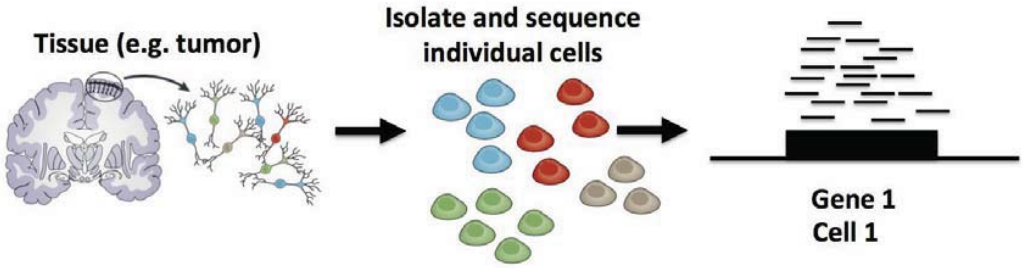
2. Single cell RNA sequencing: Why single cell?



6

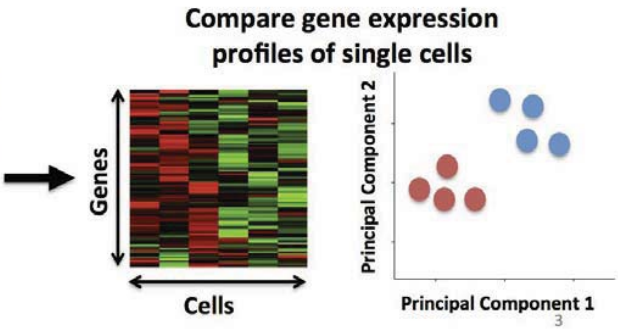
2. Single cell RNA sequencing: single cell technology

Single-cell RNA-Seq (scRNA-Seq)



Read Counts

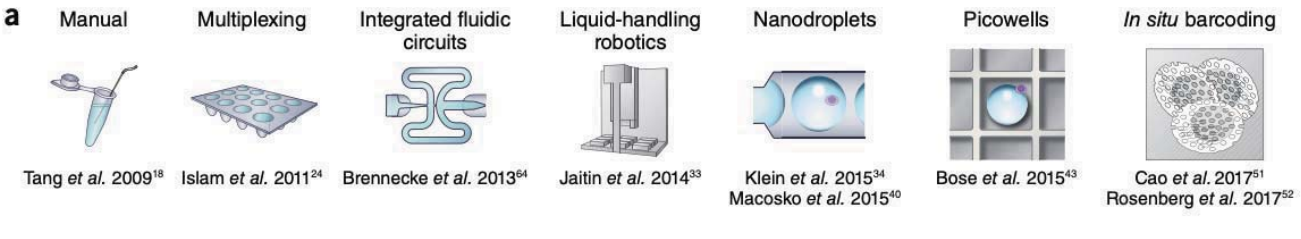
	Cell 1	Cell 2	...
Gene 1	18	0	
Gene 2	1010	506	
Gene 3	0	49	
Gene 4	22	0	
...			

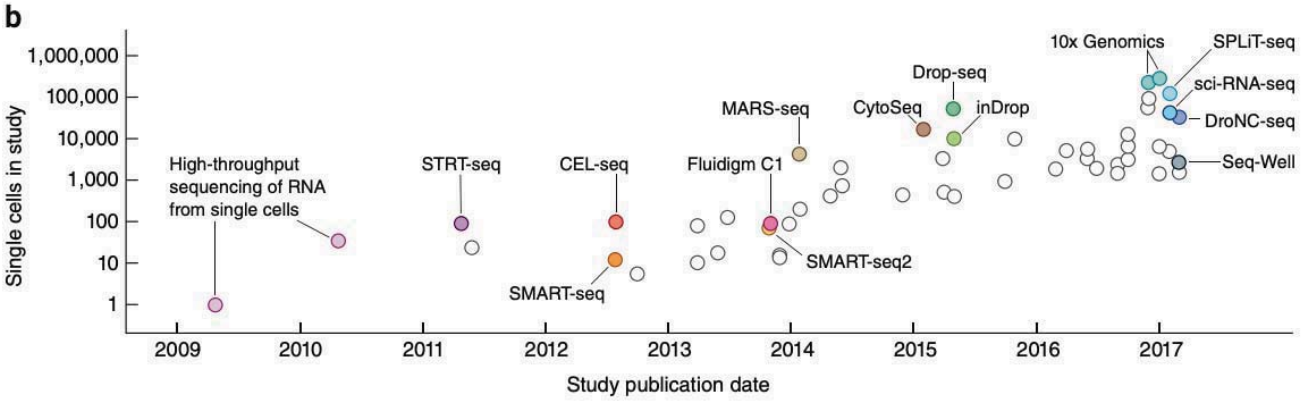


3

7

2. Single cell RNA sequencing: single cell technology

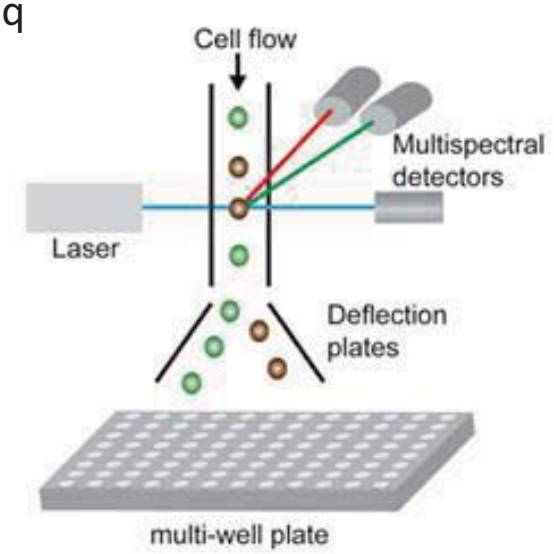




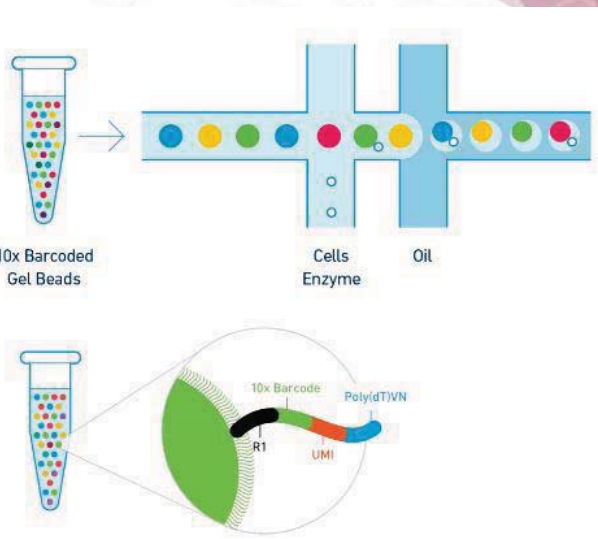
8

2. Single cell RNA sequencing: single cell technology

Smart-seq: Well-based scRNA-seq

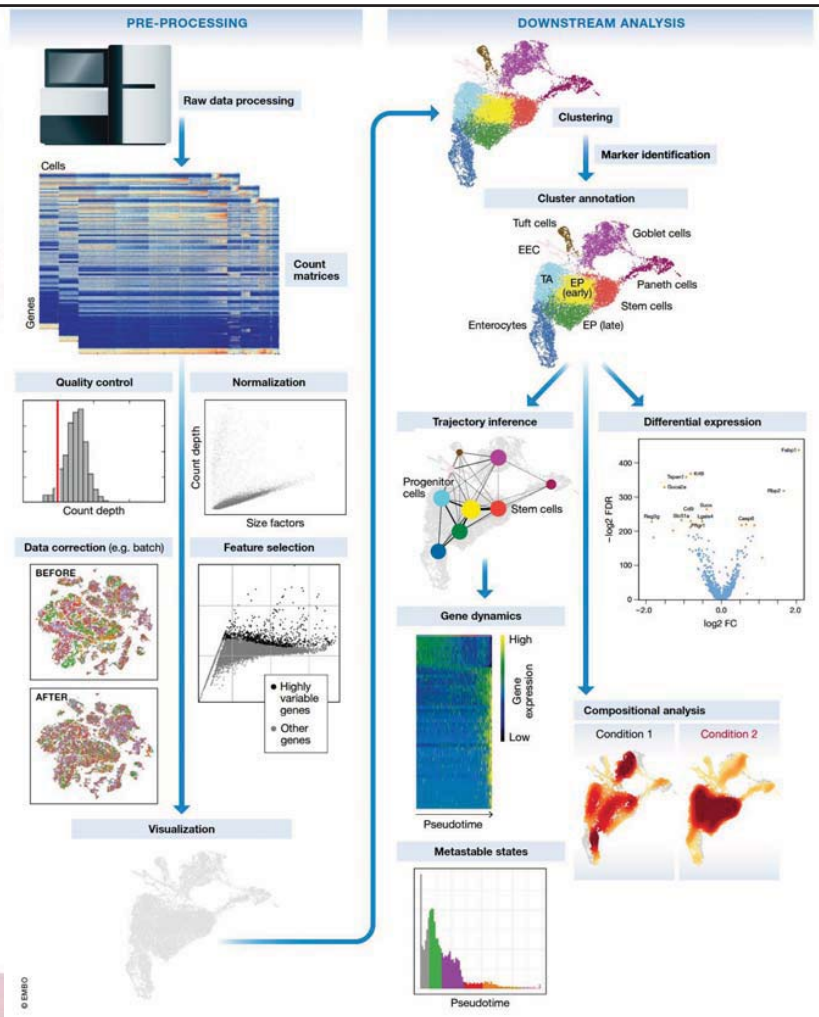


10x Genomics: Microfluidic droplet-based scRNA-seq

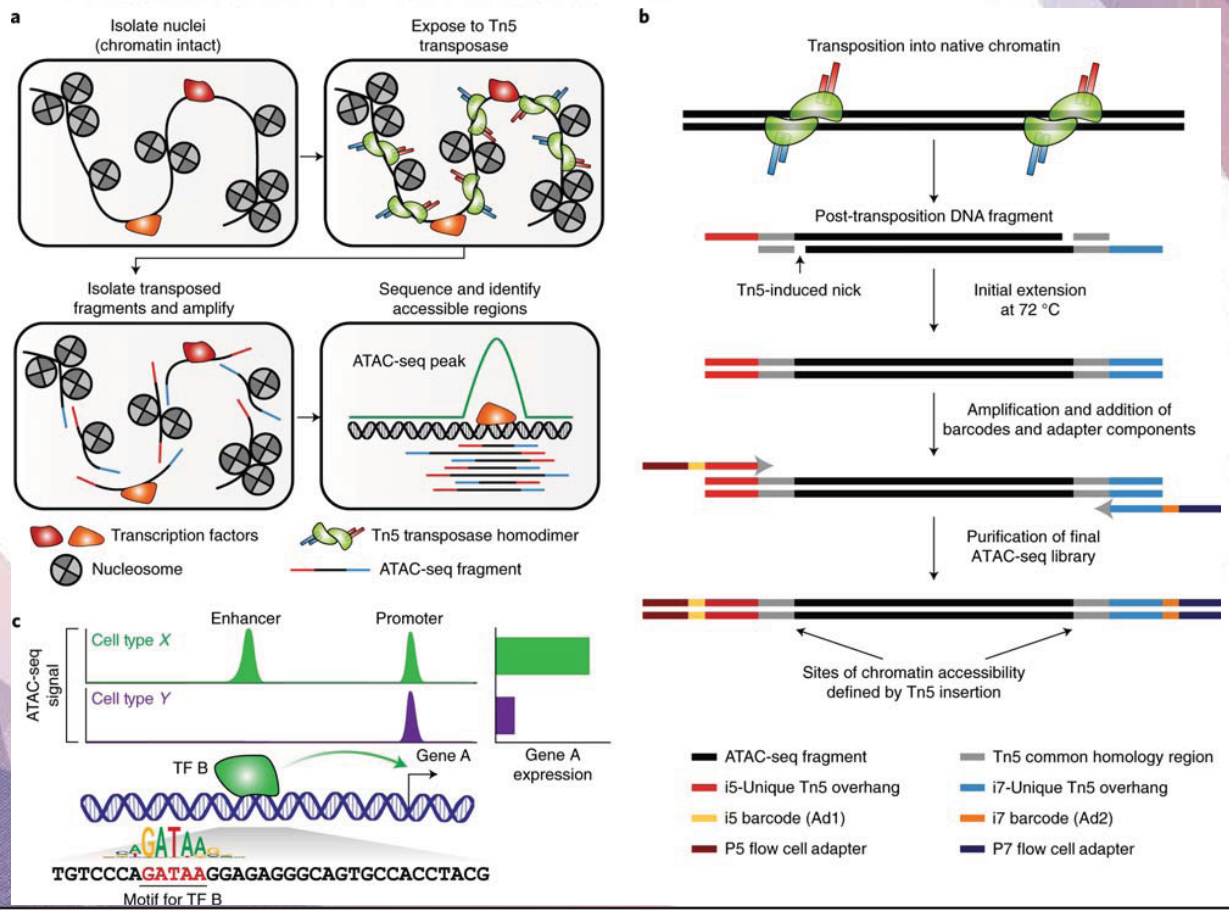


9

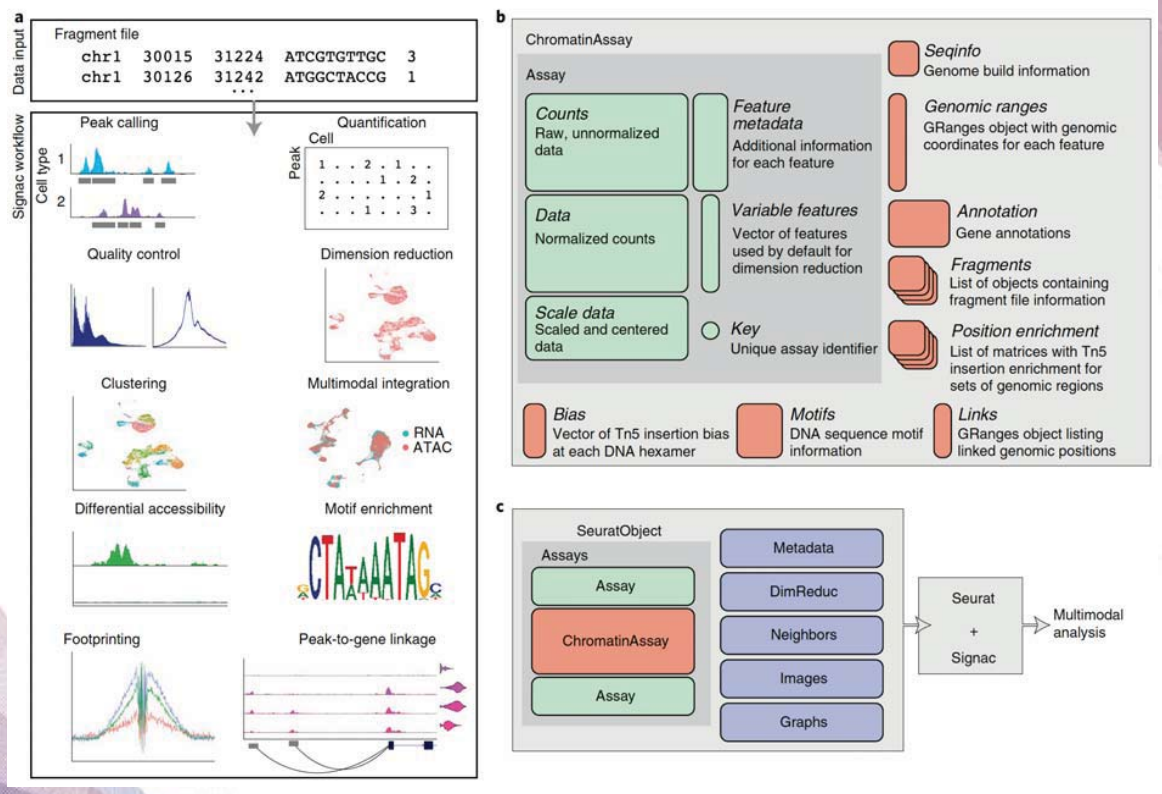
2. Single cell RNA sequencing: Analysis pipeline



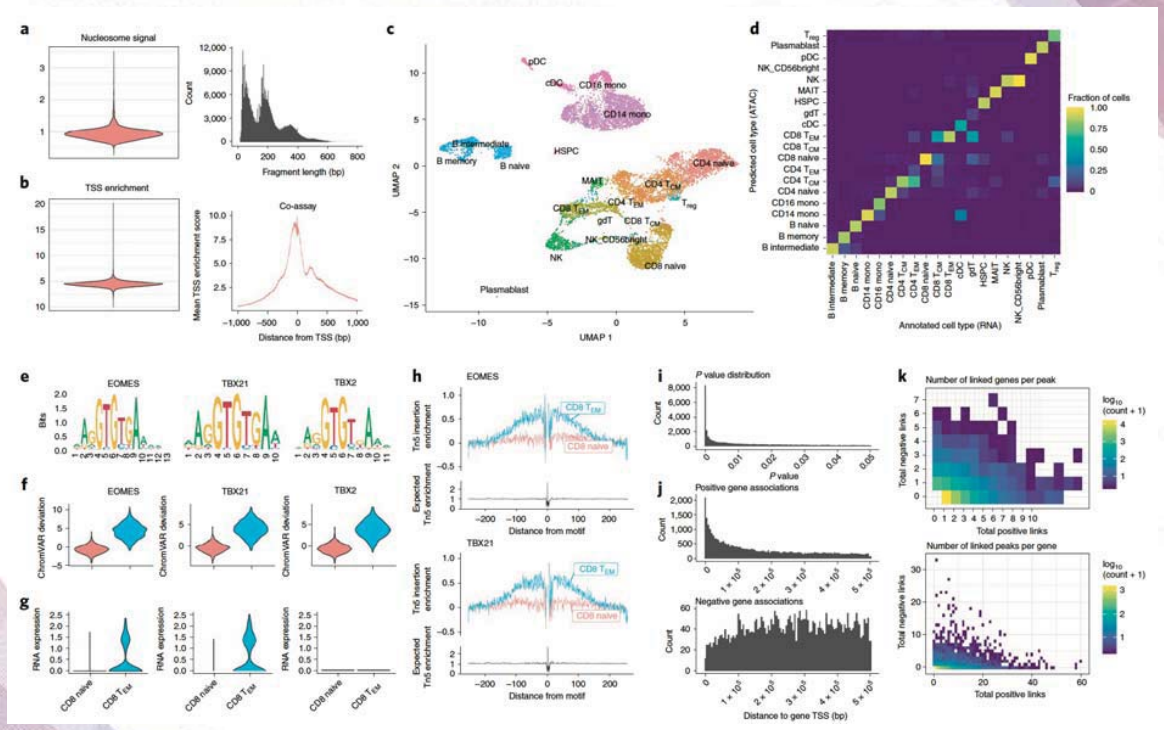
3. Single cell ATAC sequencing : Overview



3. Single cell ATAC sequencing : Analysis with Signac

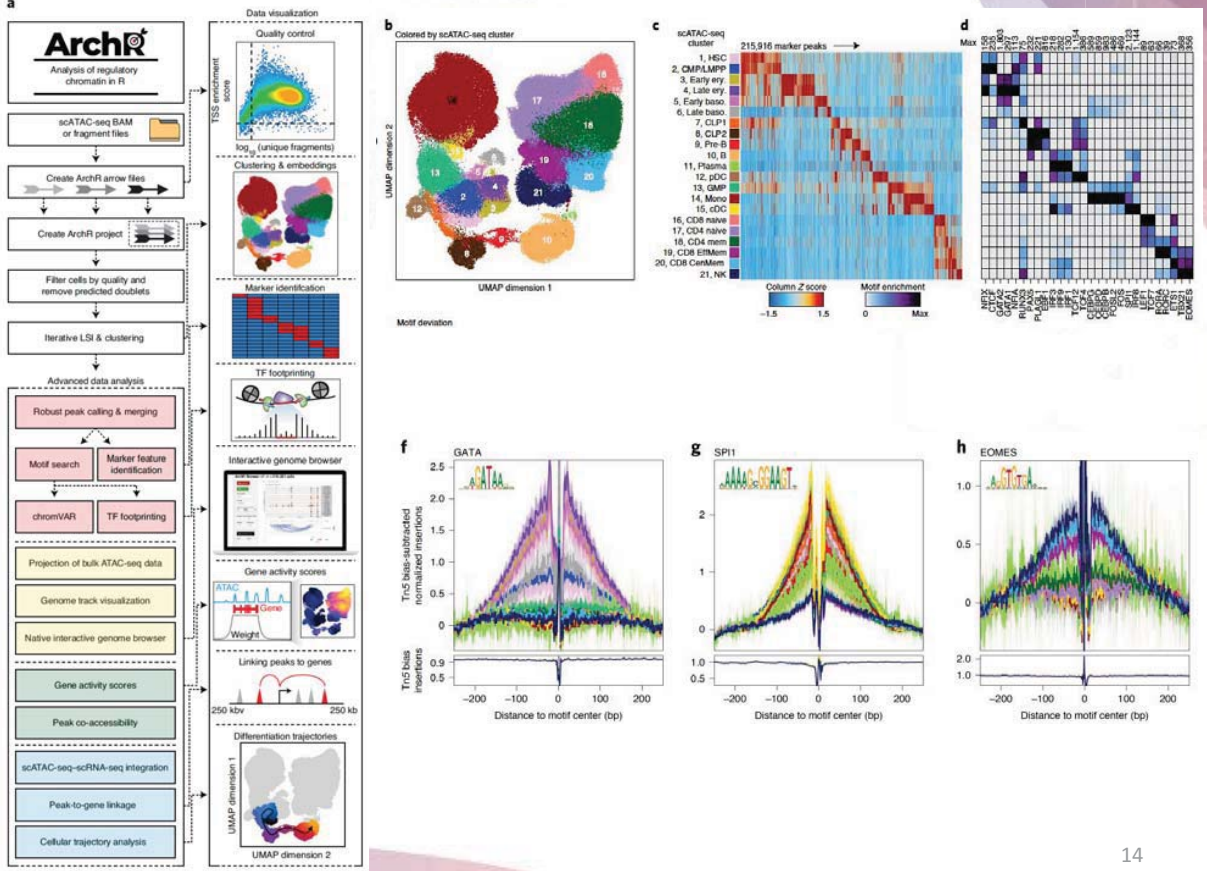


3. Single cell ATAC sequencing : Analysis with Signac



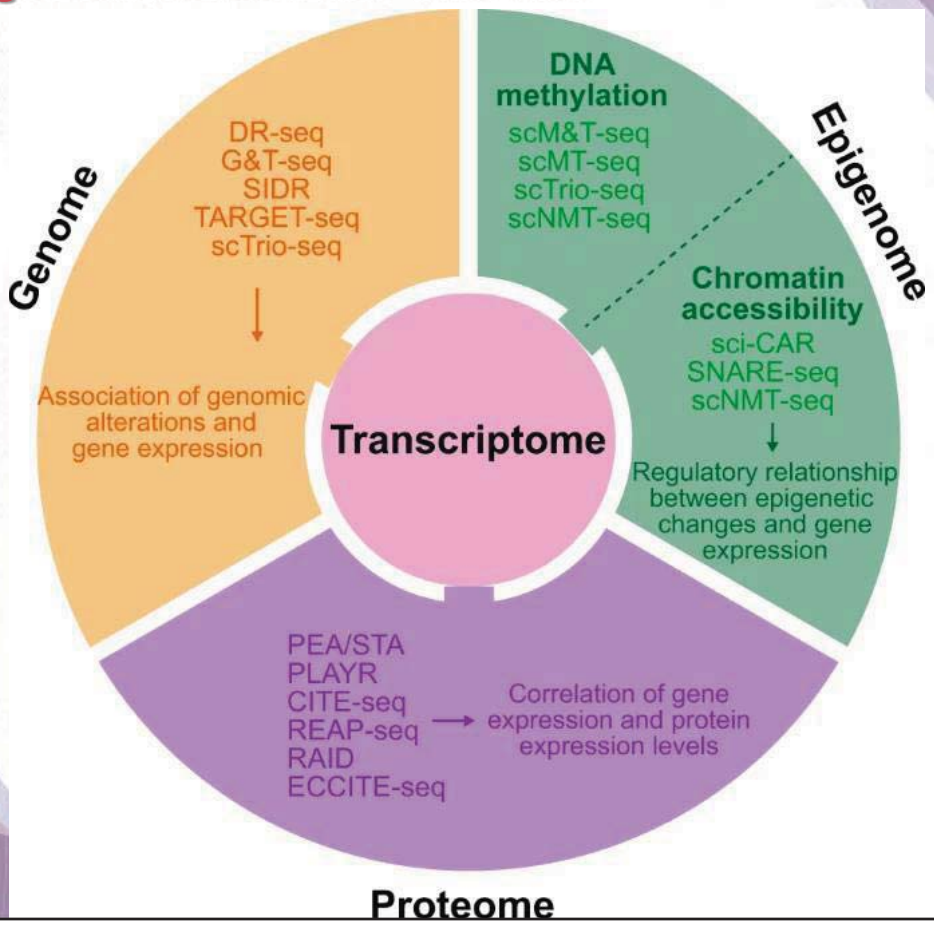
13

4. Single cell ATAC sequencing : analysis with ArChr



14

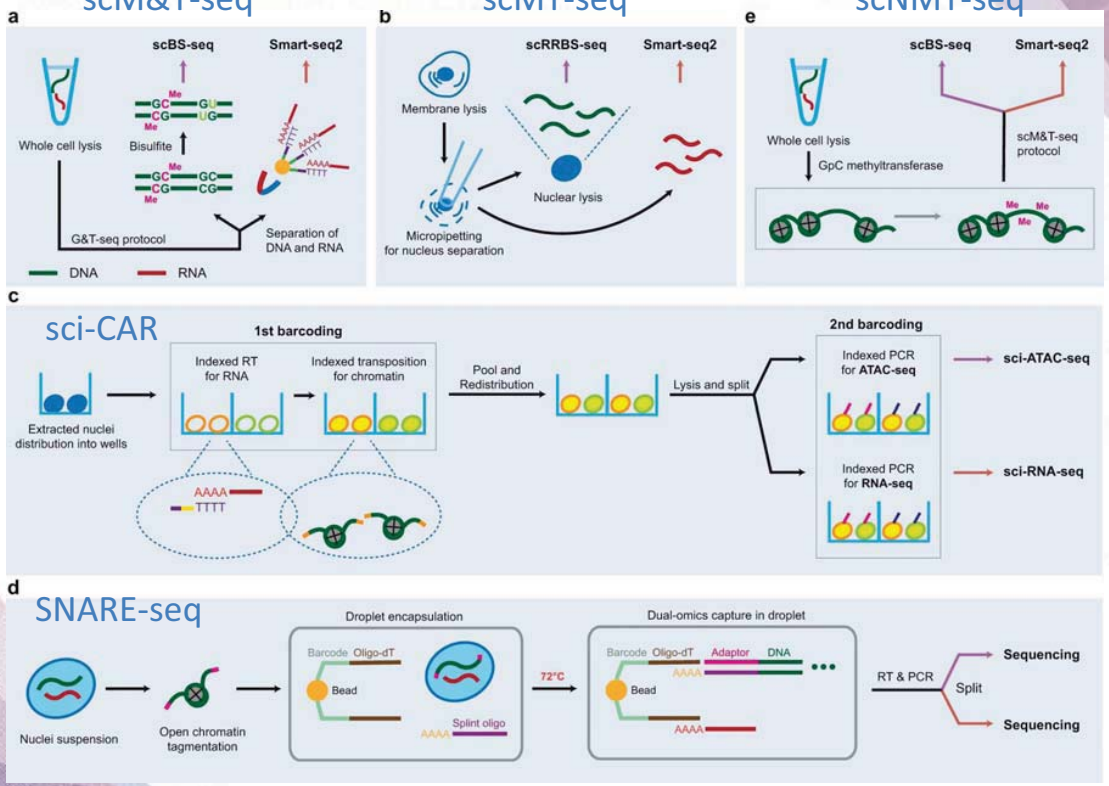
4. Single cell multiomics: Overview



15

4. Single cell multiomics: RNA+epigenome

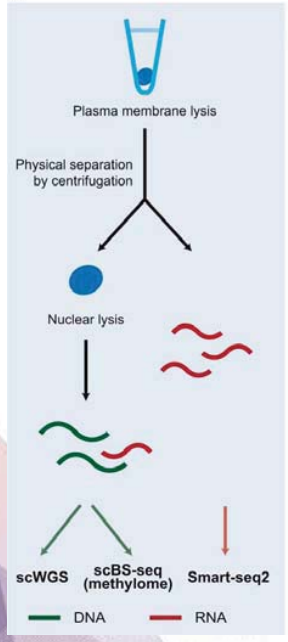
scM&T-seq scMT-seq scNMT-seq



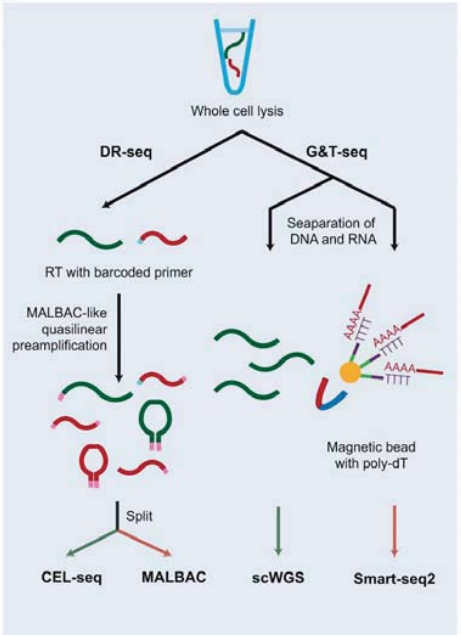
16

4. Single cell multiomics: RNA+DNA

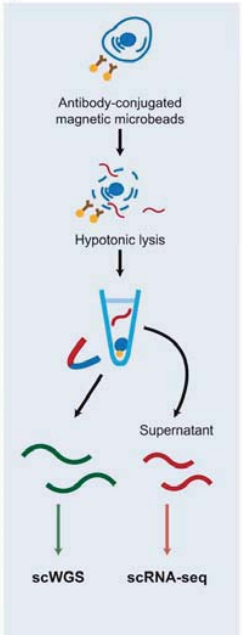
scTrio-seq



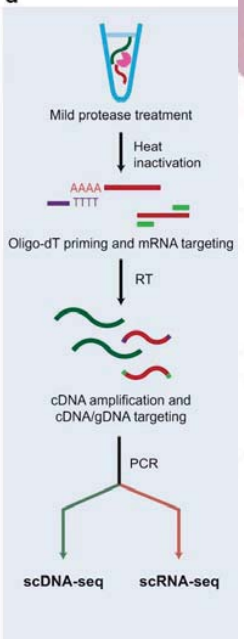
DR-seq and G&T-seq



SIDR

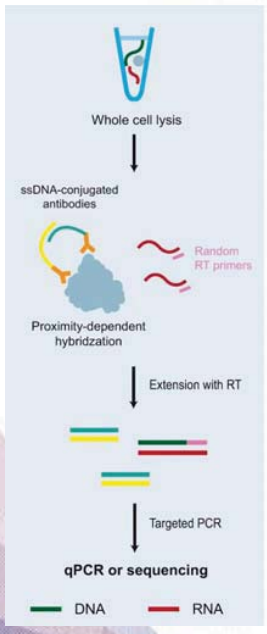


TARGET-seq

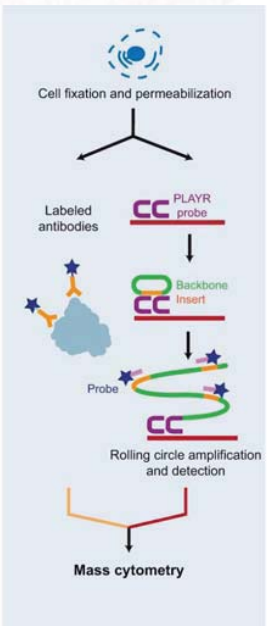


4. Single cell multiomics: RNA+Protein

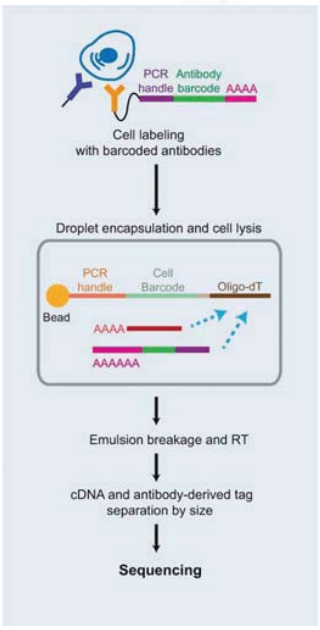
PEA/STA



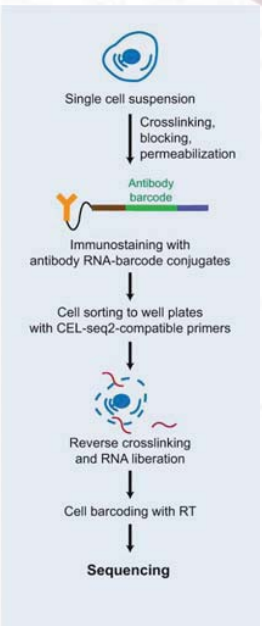
PLAYR



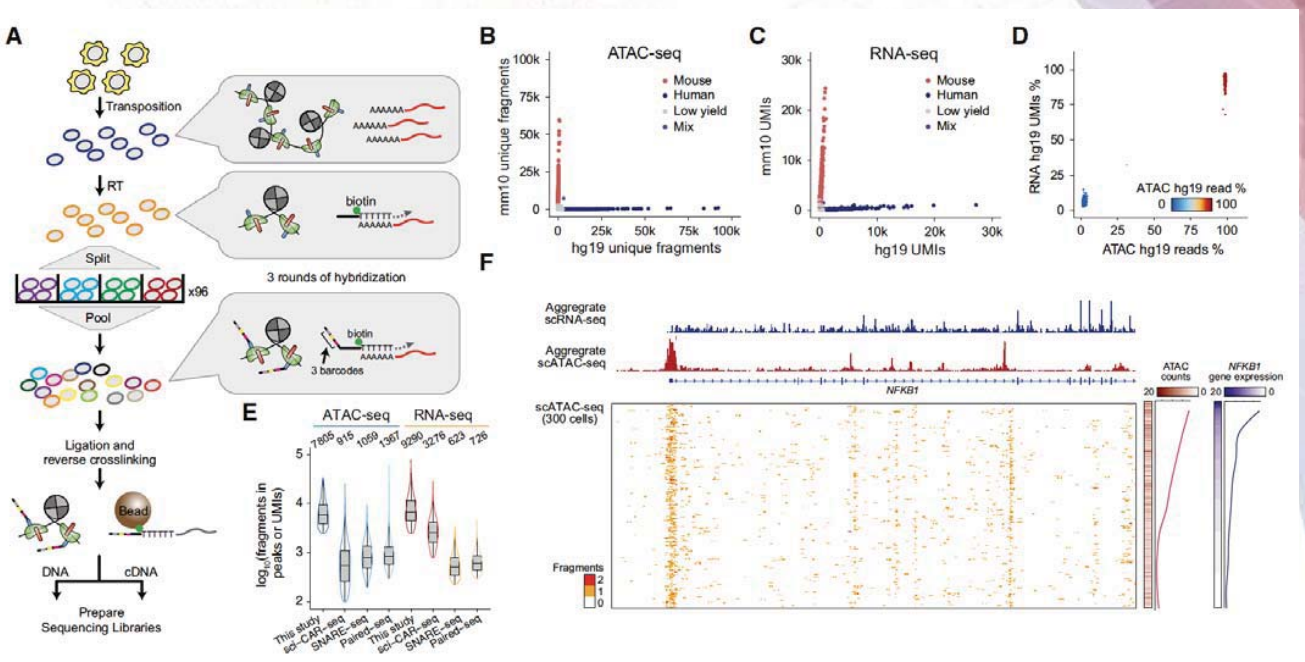
CITE-seq



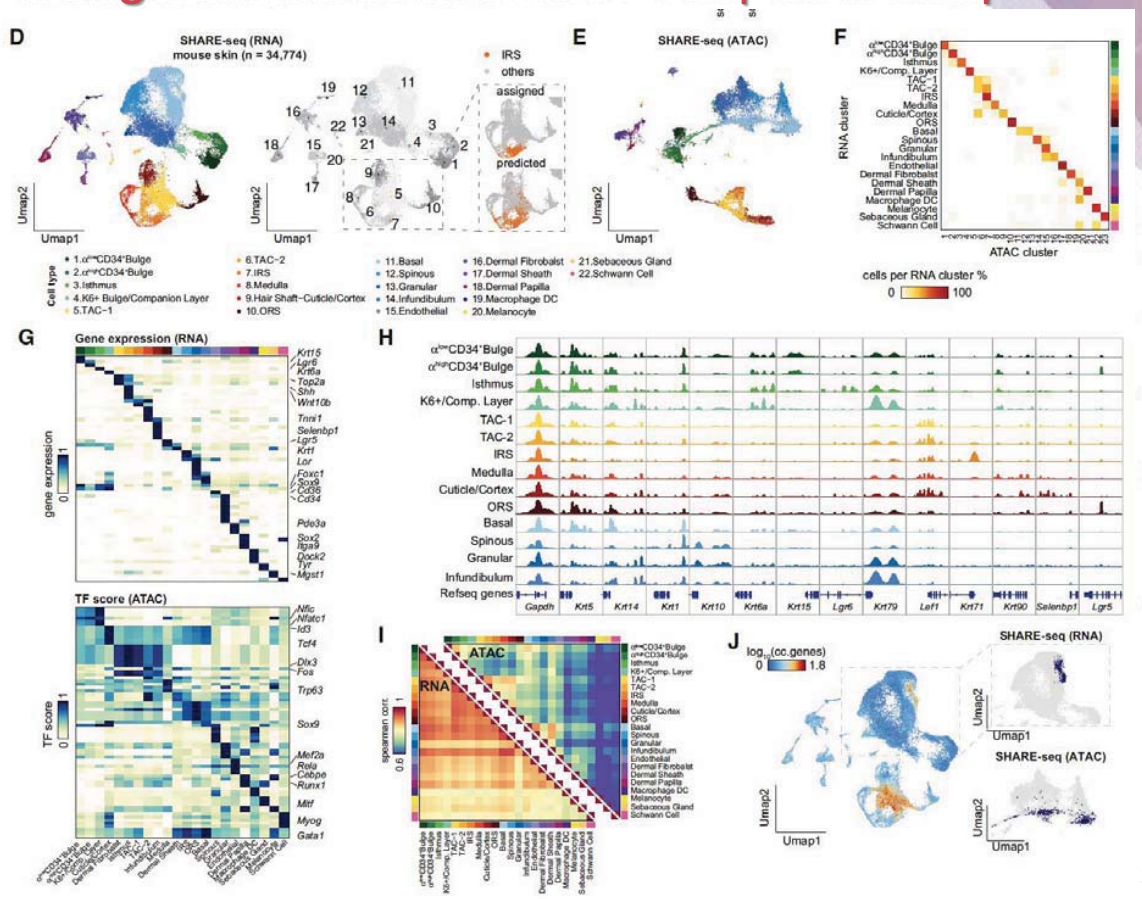
RAID



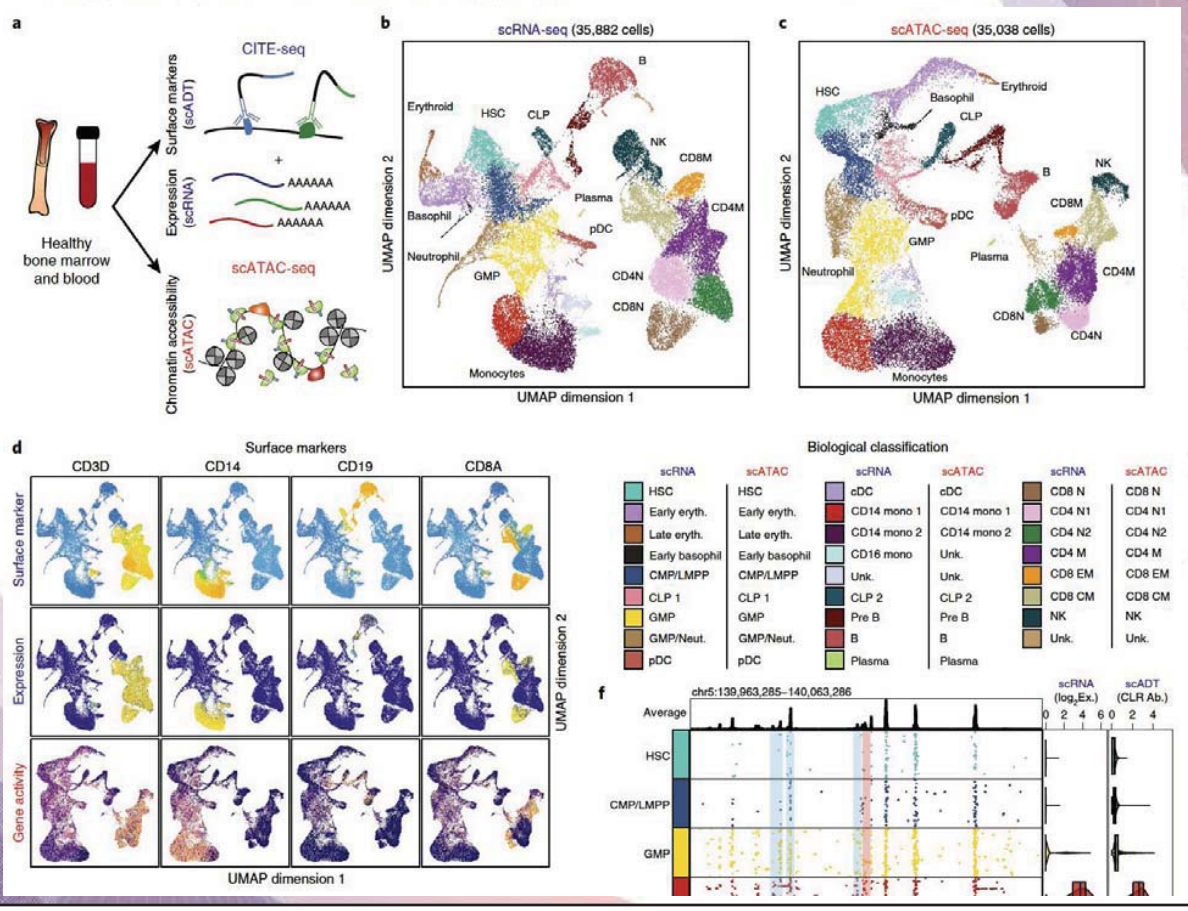
4. Single cell multiomics: scRNA-seq+scATACseq



4. Single cell multiomics: scRNA-seq+scATACseq



4. Single cell multiomics: CITE-seq



5. Data integration strategy

- Three data integration strategies
  1. Integration by genomic features (gene) as the anchor (same data modality from different experiments)
  2. Integration by cells as the anchor (multiple data modalities from the same cells)
  3. No anchors (different data modalities from different experiments)

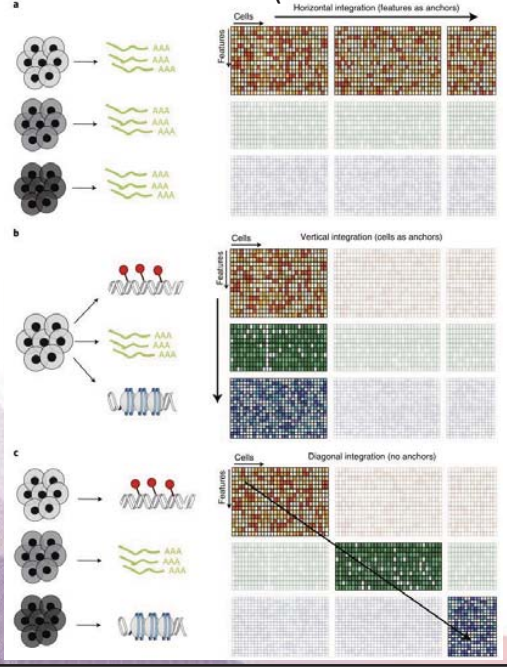


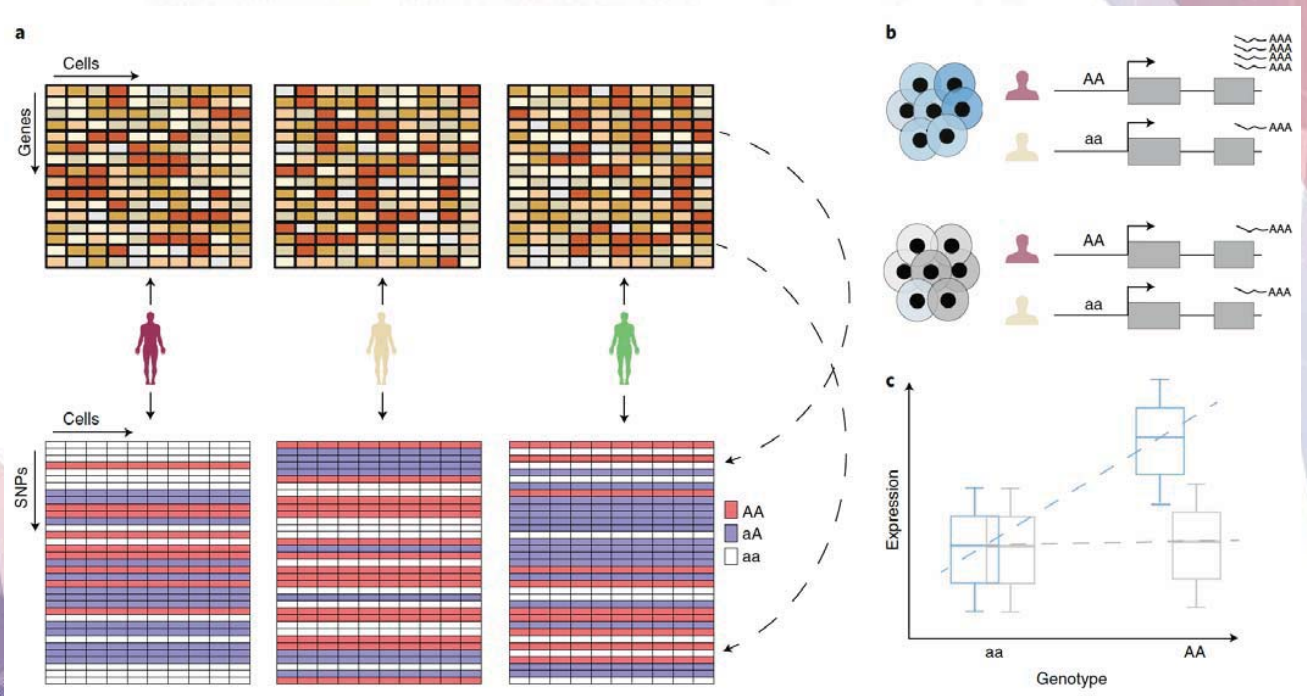
Table 1 | Overview of common data integration methods classified according to their anchor choice

Integration task	Method	Ref.
Vertical (global)	CCA	112
Vertical (global)	JIVE	70
Vertical (global)	PLS	71
Vertical (global)	MCIA	113
Vertical (global)	MOFA+	65
Vertical (global)	scAI	114
Vertical (global)	iNMF	38
Vertical (global)	Seurat v4	11
Vertical (local)	Spearman's rank correlation coefficient	50
Vertical (local)	LMM	51
Horizontal	MNN	21
Horizontal	Seurat v3	22
Horizontal	LIGER	23
Horizontal	Harmony	24
Horizontal	Scanorama	29
Horizontal	BBKNN	25
Horizontal	scVI	26
Horizontal	scmap	28
Horizontal	conos	27
Diagonal	MATCHER	77
Diagonal	MMD-MMA	78
Diagonal	SCIM	115
Diagonal	UnionCom	116
Diagonal	coupledNMF	117

22

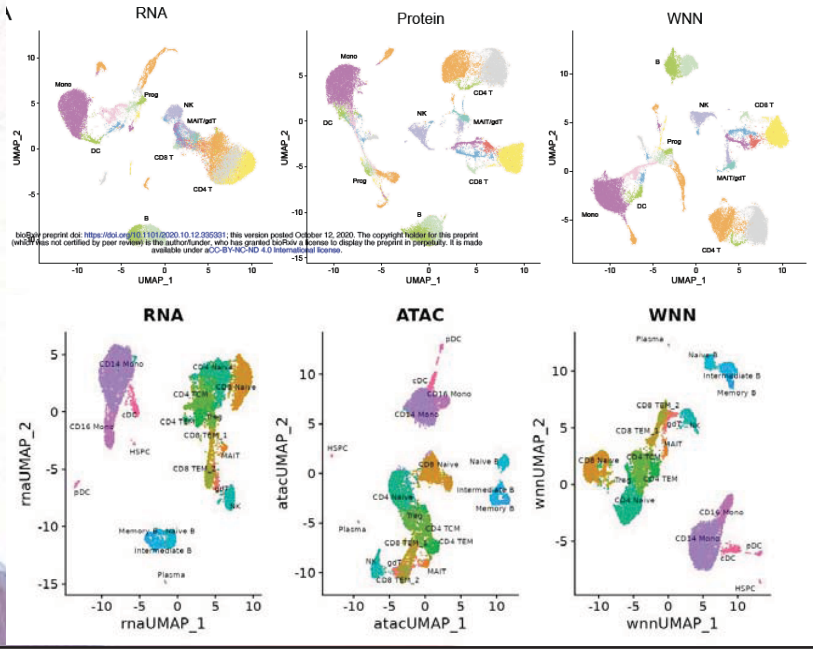
5. Data integration strategy – vertical (local)

Single cell eQTL

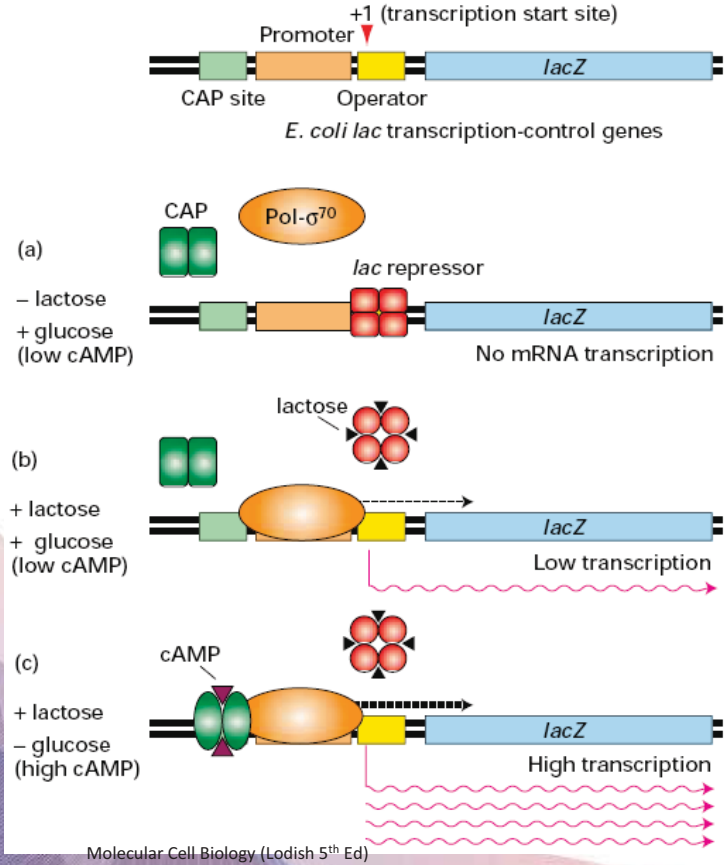


5. Data integration strategy – vertical (global)

- Challenges in multimodal analysis
  1. Molecular readouts collected using different techniques generally have heterogeneous statistical properties
  2. Different data modalities can have different numbers of features
  3. All matrix factorization methods report a solution, but its quality is hard to assess
- Seurat v4 – weighted nearest neighbor (WNN) analysis



6. Gene regulatory network



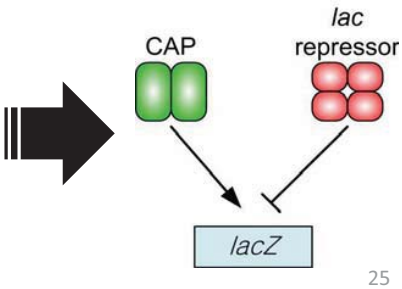
Molecular Cell Biology (Lodish 5th Ed)

lacZ: a protein which is needed to digest lactose

(a) If there is no lactose, e.coli does not need to express lacZ gene.

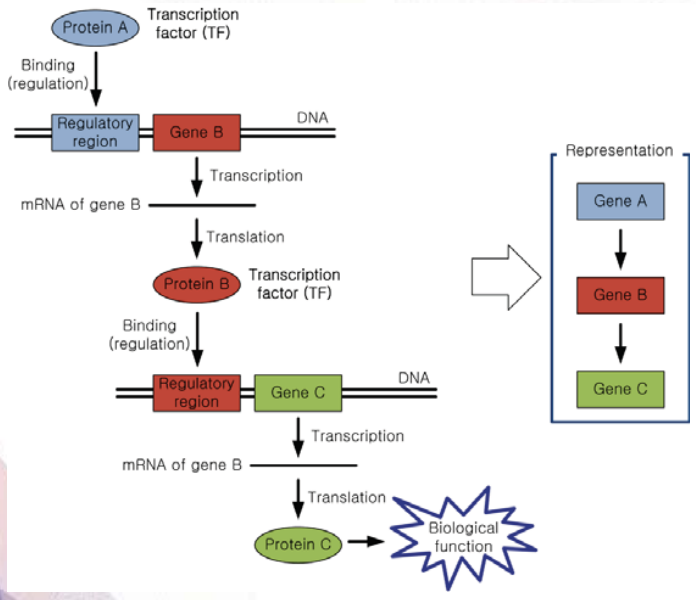
(b) Even if there exist lactose, the existence of glucose make e.coli express lacZ very low level because glucose is easy to make energy.

(c) If lactose is available and glucose does not, e.coli desperately express lacZ in a large amount because there is no other option.



25

6. Gene regulatory network



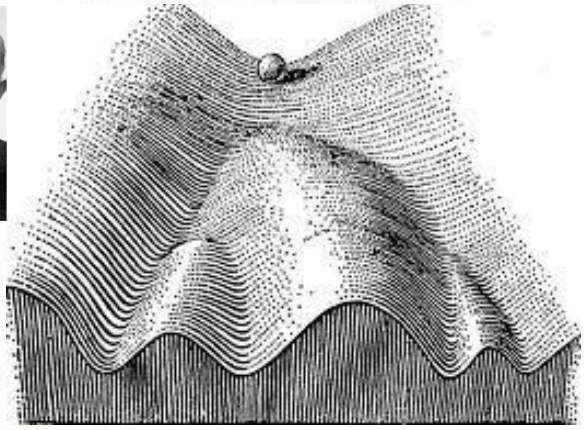
- GRN is the main biological circuit which **control** many biological functions by **governing the amount of the mRNA** and finally the protein
- A transcriptional regulation depends on the binding specificity of **transcription factors (TFs)** and a regulatory **sequence region** (proximal or distal)
- GRN is a directed weighted graph

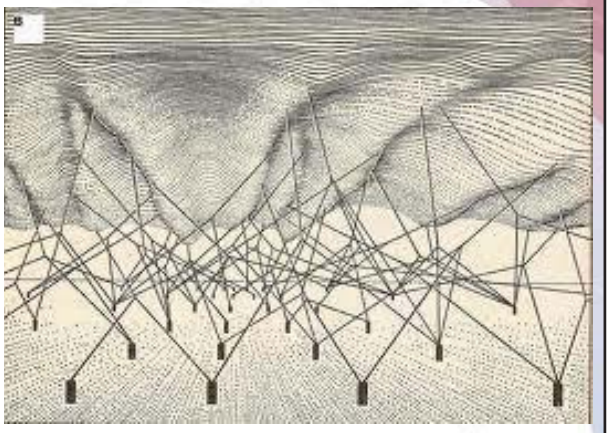
26

6. Gene regulatory network

Cellular development can be explained by gene regulatory circuits

Conrad Waddington

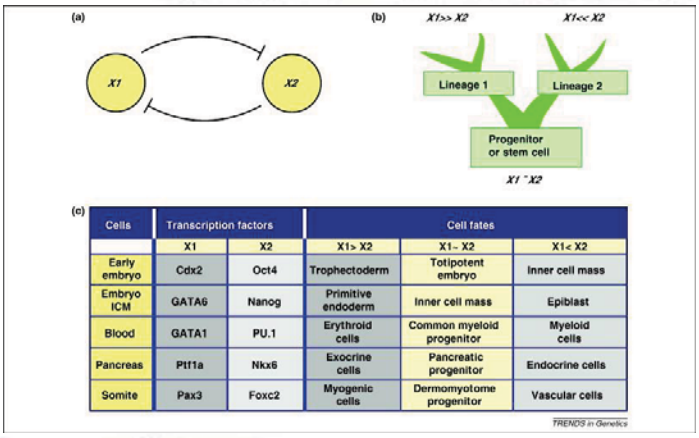




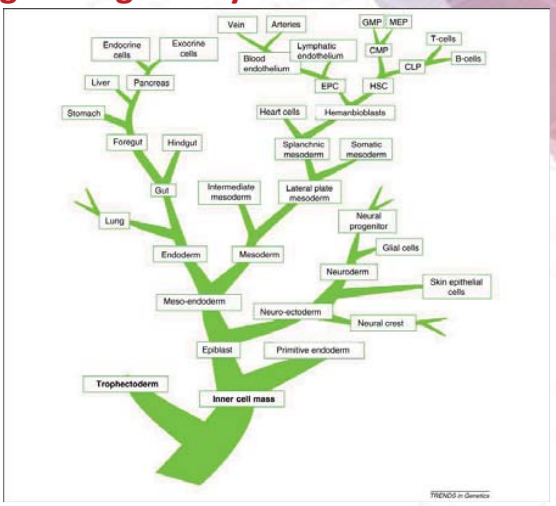
Waddington CH, The Strategy of the Genes (1957)

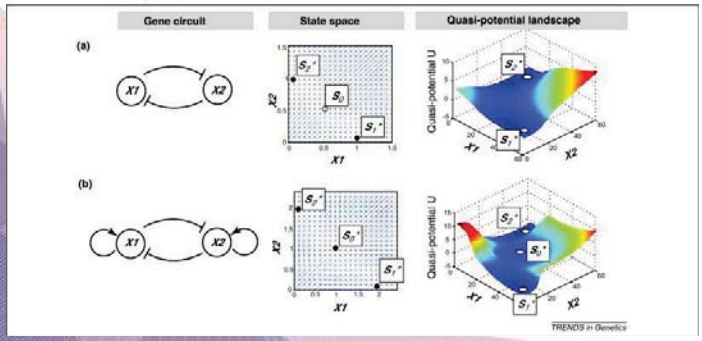
6. Gene regulatory network

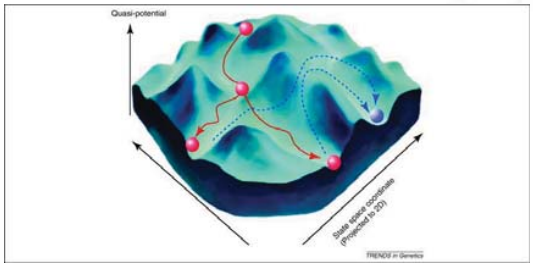
Cellular development can be explained by gene regulatory circuits



Cells	Transcription factors		Cell fates		
	X1	X2	X1> X2	X1~ X2	X1< X2
Early embryo	Cdx2	Oct4	Trophectoderm	Totipotent embryo	Inner cell mass
Embryo ICM	GATA6	Nanog	Primitive endoderm	Inner cell mass	Epiblast
Blood	GATA1	PU.1	Erythroid cells	Common myeloid progenitor	Myeloid cells
Pancreas	Ptf1a	Nkx6	Exocrine cells	Pancreatic progenitor	Endocrine cells
Somite	Pax3	Foxc2	Myogenic cells	Dermomyotome progenitor	Vascular cells

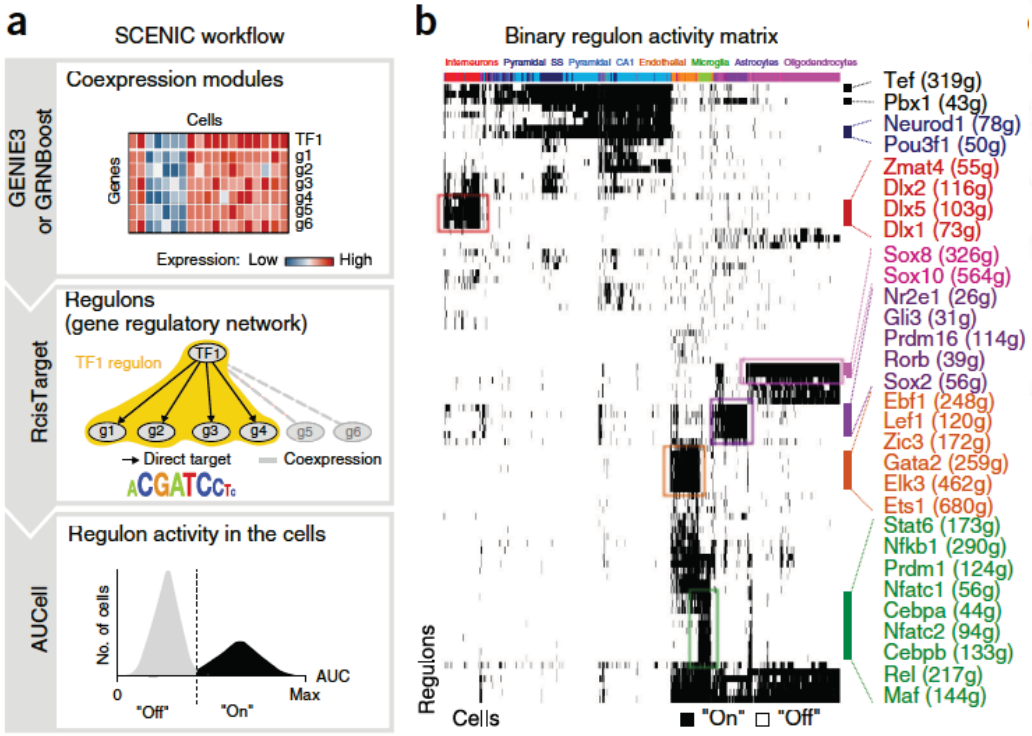




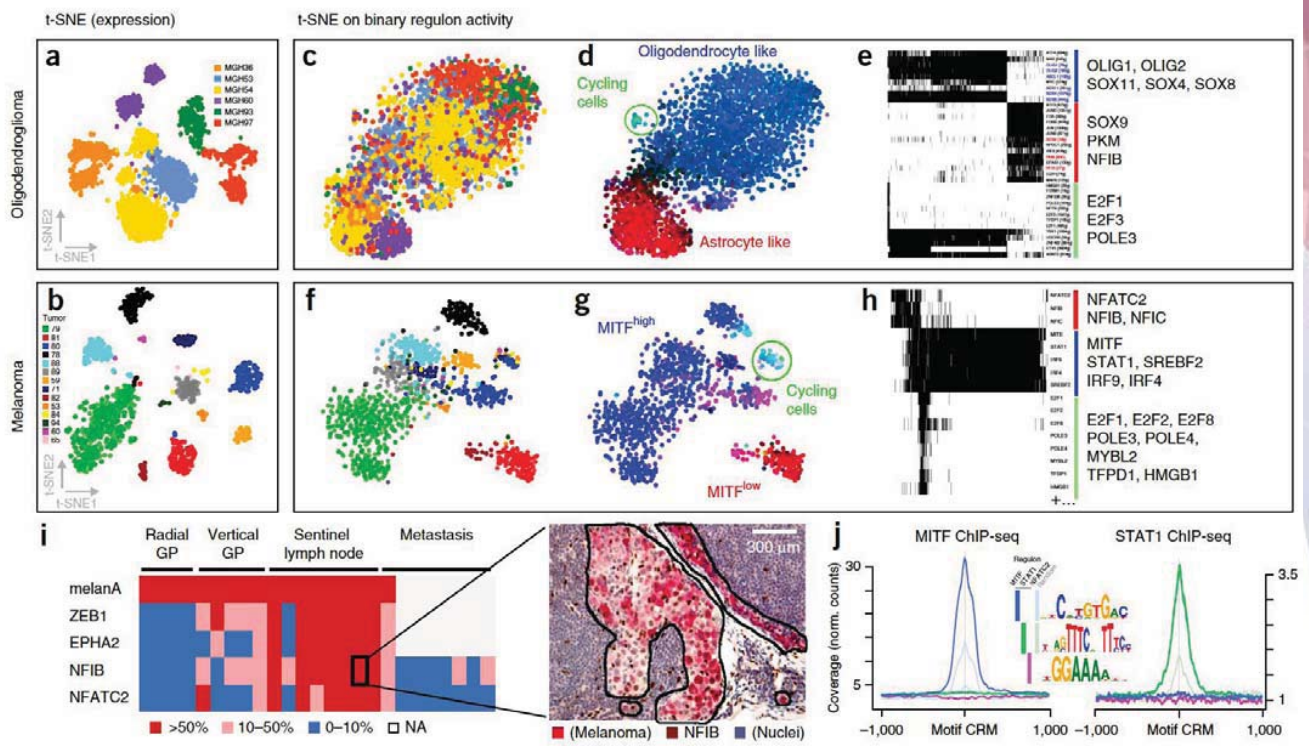


Zhou JX and Huang S, Trends in Genetics (2010)

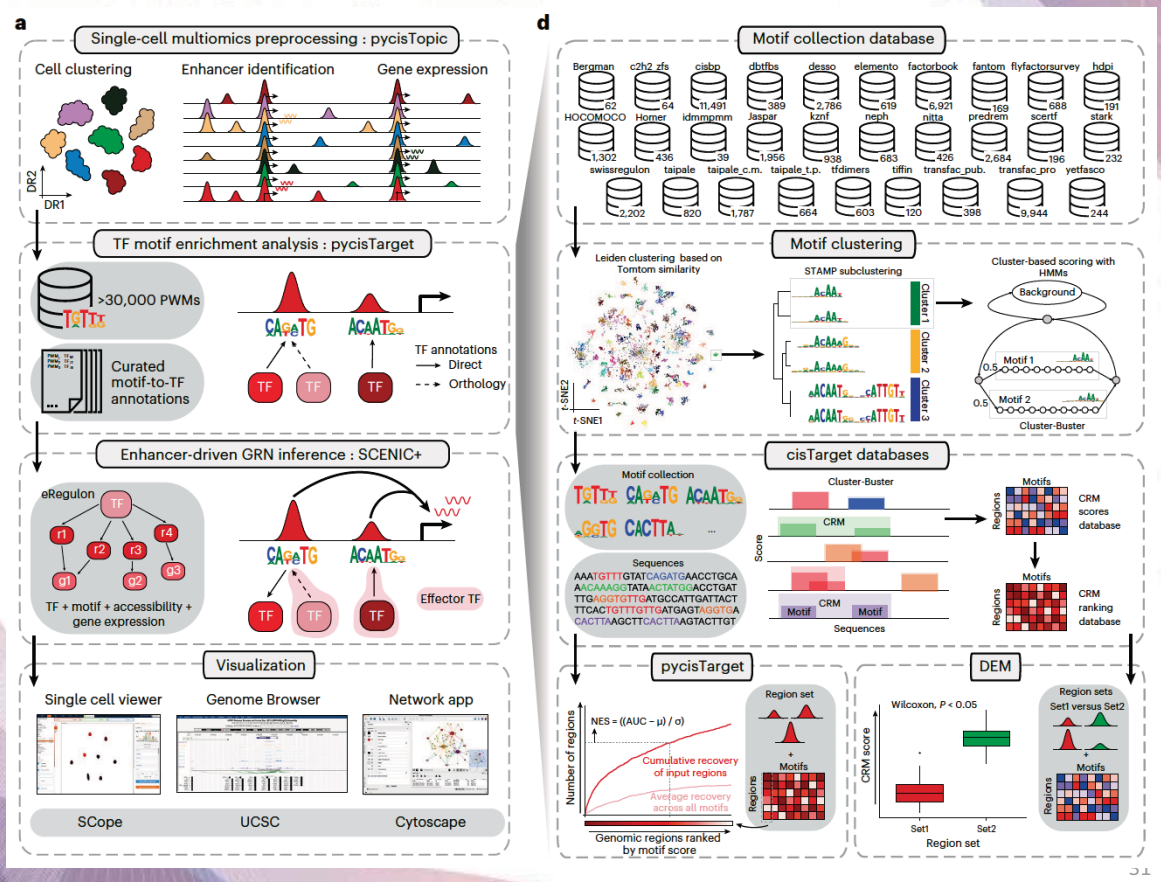
6. Gene regulatory network - SCENIC



6. Gene regulatory network - SCENIC



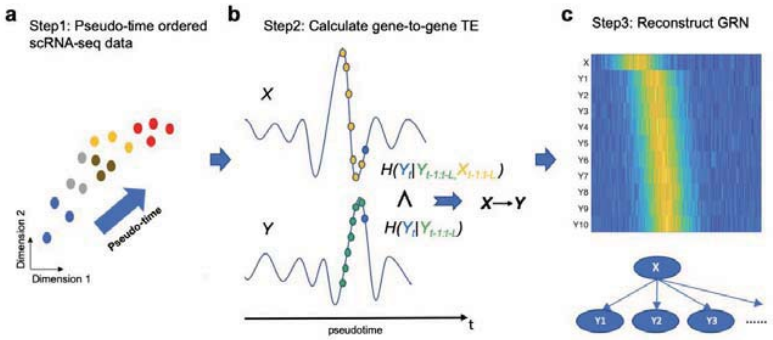
6. Gene regulatory network – SCENIC+



6. Gene regulatory network – TENET+

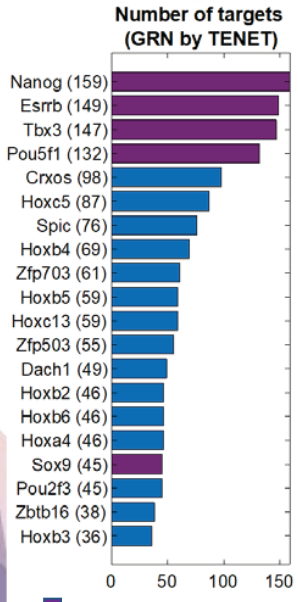
TENET

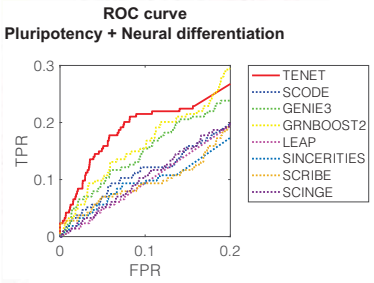
A tool for reconstructing Transfer Entropy-based causal gene NETwork from pseudo-time ordered single cell transcriptomic data

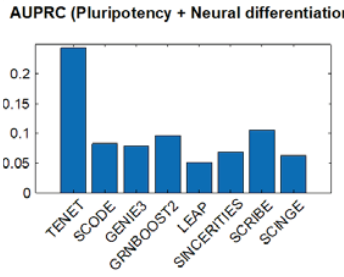


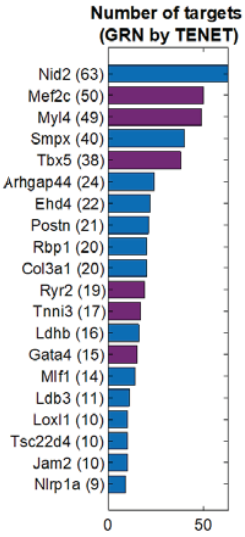
Kim et al., Nucleic Acids Research. (2021)

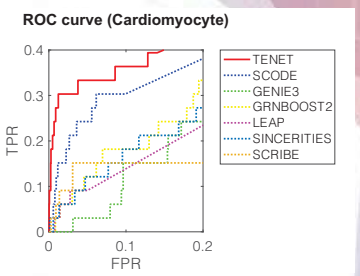
6. Gene regulatory network – TENET+

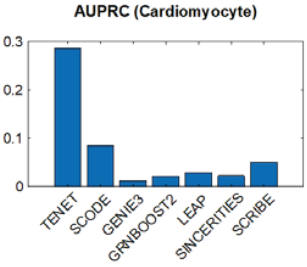






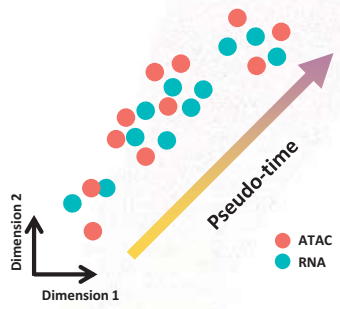


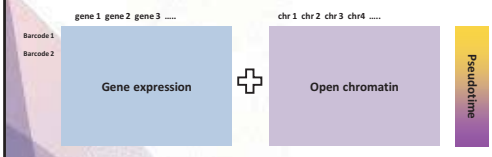




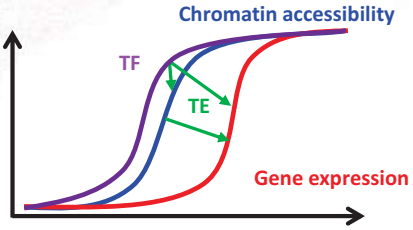
6. Gene regulatory network – TENET+

Step 1 : Pseudot-time ordered scRNA-seq & scATAC-seq data





Step 2 : Calculate TF-to-gene, peak TE



$$H(TF_t|TF_{t-1:t-L}, G_{t-1:t-L})$$
$$\wedge$$
$$H(TF_t|TF_{t-1:t-L})$$

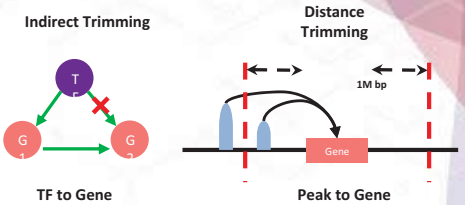
TF -> Gene expression

$$H(TF_t|TF_{t-1:t-L}, C_{t-1:t-L})$$
$$\wedge$$
$$H(TF_t|TF_{t-1:t-L})$$

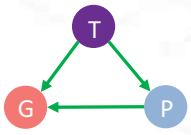
TF -> Chromain accessibility

Chromain accessibility -> Gene expression

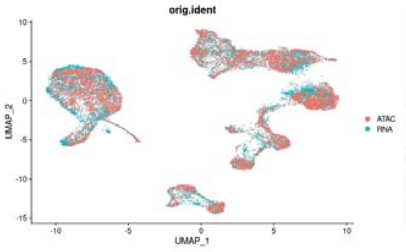
Step 3 : Trimming the targets and Downstream analysis

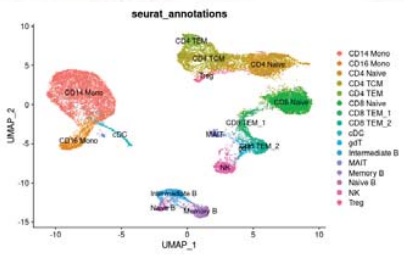


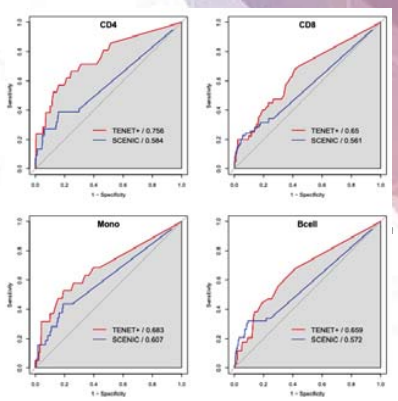
TF-gene-peak Triplet

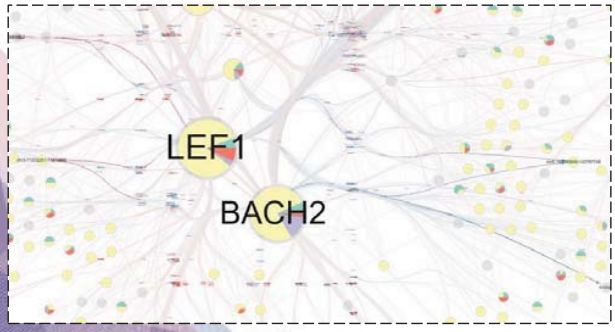


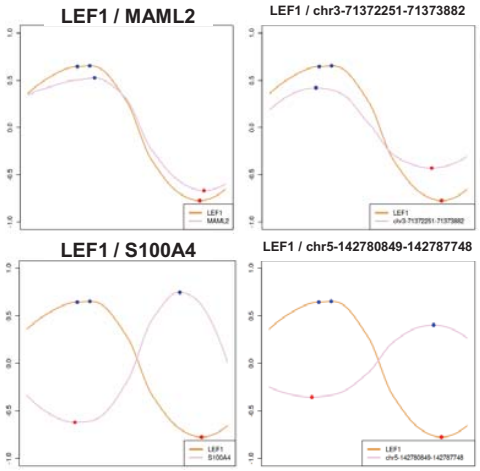
6. Gene regulatory network – TENET+











35

7. References

- Lee et al., “Single-cell multiomics: technologies and data analysis methods”, Exper imental & Molecular Medicine 2020
- Argelaguet et al., “Computational principles and challenges in single-cell data inte gration”, Nature Biotechnology 2021
- Stuart et al., “Integrative single-cell analysis”, Nature Review Genetics 2019
- Macaulay et al., “Single-cell multiomics: multiple measurements from single cells” , Trends in Genetics 2017
- Argelaguet et al., “Multi-omics profiling of mouse gastrulation at single-cell resolu tion” Nature 2019
- Ma et al., “Chromatin potential identified by shared single-cell profiling of RNA an d chromatin”, Cell 2020
- Cao et al., “A human cell atlas of fetal gene expression”, Science 2020
- Domcke et al., “A human cell atlas of fetal chromatin accessibility”, Sceince 2020
- Granja et al., “Single-cell multi-omics analysis identifies regulatory programs in mi xed-phenotype acute leukemia”, Nature Biotechnology 2019
- Hao et al., “Integrated analysis of multimodal single-cell data”, BioRxiv 2020

36