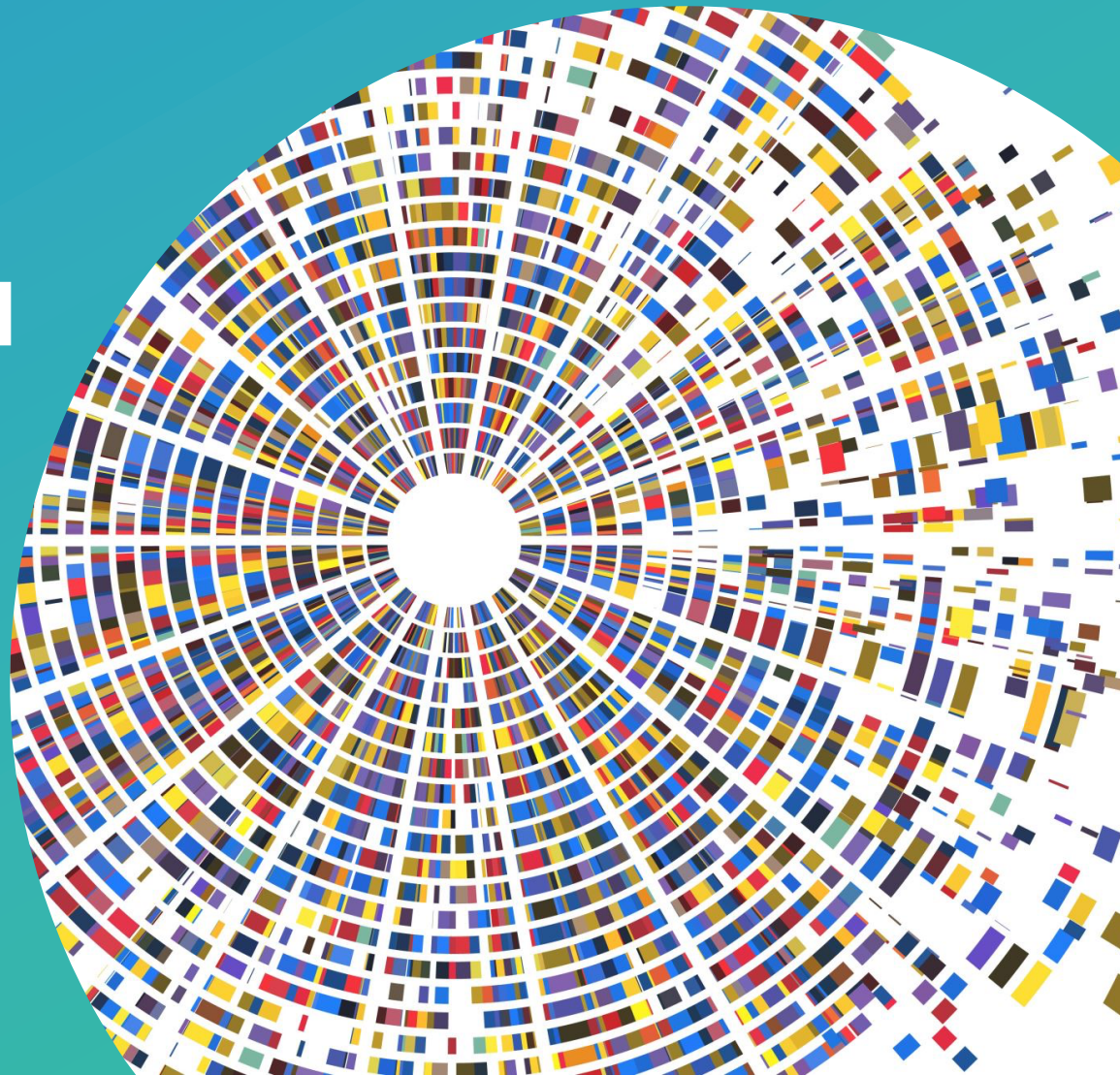
ANALYSIS AND INTEGRATION OF OMICS DATA WITH APPLICATIONS TO BIOMEDICAL RESEARCH

Claudia Angelini

Istituto per le Applicazioni del Calcolo “Mauro Picone”



Preamble: Who am I and What am I Doing Here?

- I am a Mathematician
- I like to analyze data arising from omic studies and face real data problem
- I already know some of you

- Design of **complex pipelines** for omics data analysis and multi-omic data integration
- Development of **computational tools** for bioinformatics
- Development of **novel statistical approaches** for the analysis of omics data

Omics data

In the last two decades high-throughput technologies have revolutionized biomedical research

- **Huge amount of data due to** decreasing experimental costs
- **Different omic technologies** allow to elucidate genome-wide **cellular mechanisms**
- The challenges shifted from **data collection** to **data analysis** (and result interpretation)

→ *the dark side of the moon*

Snap-shot of some results

Aim of the work was to study the **regulatory mechanism** between ***Tbx1*** and **H3K4me1** (i.e., to show that TBX1 positively regulate H3K4me1) and show that treatment of cells with ***Tranylcypromine*** (**TCP**) rebalance the loss of H3K4me1 and rescue the expression of a significant number of genes, ameliorating the the cardiovascular phenotype (congenital heart disease), including structural defects.

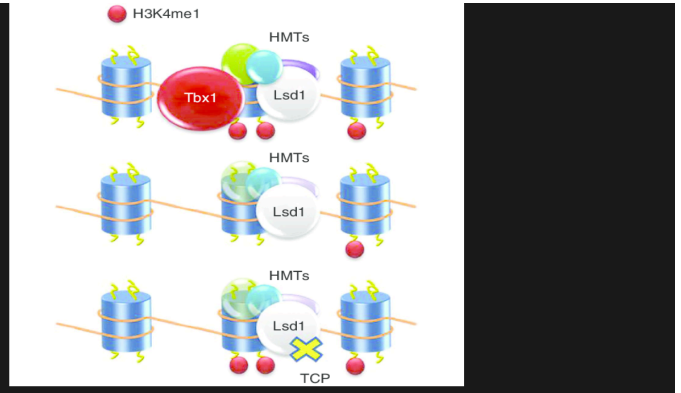


Figure 5 | Cartoon showing a working model for Tbx1 interactions with chromatin and TCP rebalancing effect. In the WT state (top row), Tbx1 promotes H3K4me1 deposition (small red dots) through interaction with histone methyltransferases (HMTs). When Tbx1 is absent or of lowered dosage (middle row), H3K4me1 enrichment is lowered because of reduced recruitment of HMTs. TCP treatment (bottom row) re-establishes H3K4me1 by inhibiting the activity of the Lsd1 demethylase. Blue cylinders indicate histone octamers.

nature COMMUNICATIONS

ARTICLE

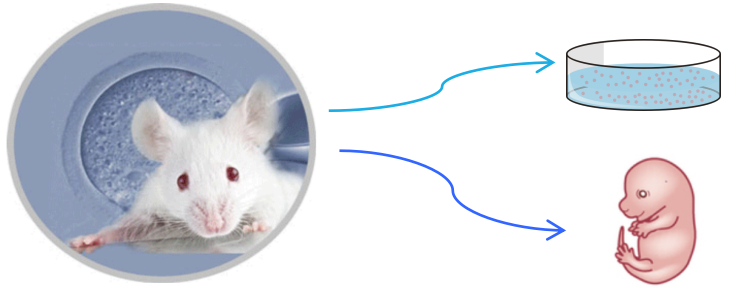
Received 16 Mar 2016 | Accepted 19 Apr 2016 | Published 3 Jun 2016

DOI: 10.1038/ncomms11688 OPEN

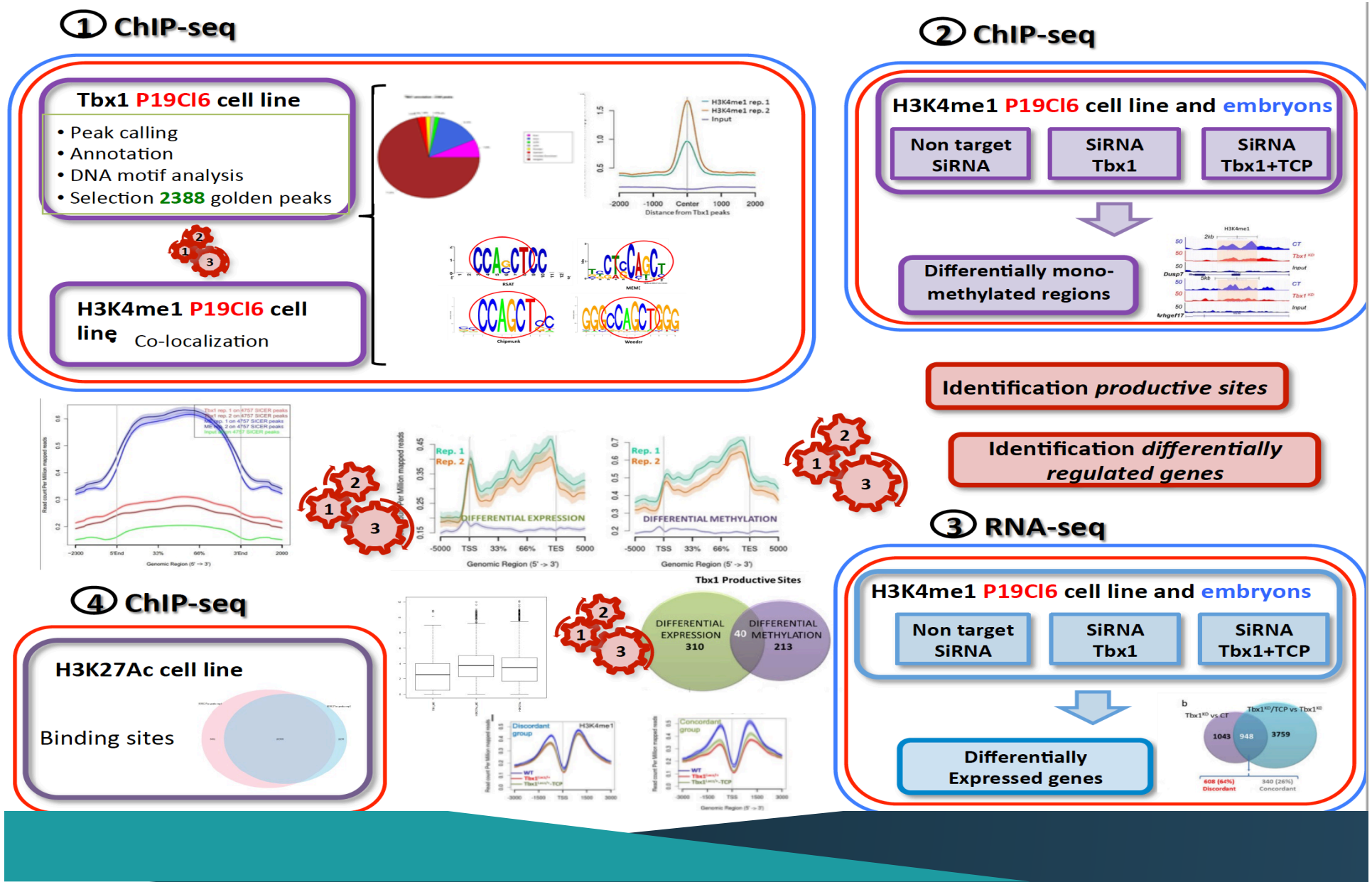
Rebalancing gene haploinsufficiency *in vivo* by targeting chromatin

Filomena Gabriella Fulcoli[1], Monica Franzese[2], Xiangyang Liu[3], Zhen Zhang[3], Claudia Angelini[2] & Antonio Baldini[1,4]

Congenital heart disease (CHD) affects eight out of 1,000 live births and is a major social and health-care burden. A common genetic cause of CHD is the 22q11.2 deletion, which is the basis of the homonymous deletion syndrome (22q11.2DS), also known as DiGeorge syndrome. Most of its clinical spectrum is caused by haploinsufficiency of *Tbx1*, a gene encoding a T-box transcription factor. Here we show that Tbx1 positively regulates monomethylation of histone 3 lysine 4 (H3K4me1) through interaction with and recruitment of histone methyltransferases. Treatment of cells with tranylcypromine (TCP), an inhibitor of histone demethylases, rebalances the loss of H3K4me1 and rescues the expression of approximately one-third of the genes dysregulated by *Tbx1* suppression. In Tbx1 mouse mutants, TCP treatment ameliorates substantially the cardiovascular phenotype. These data suggest that epigenetic drugs may represent a potential therapeutic strategy for rescue of gene haploinsufficiency phenotypes, including structural defects.



An overview of the analysis



Snap-shot of some results

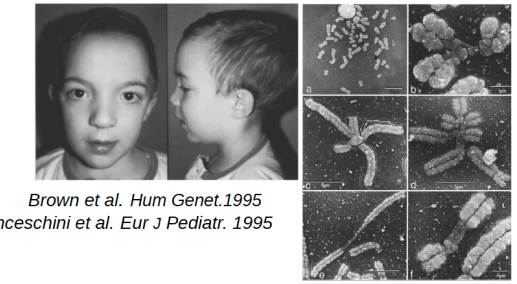
We performed a **transcriptomic and epigenomic study** in patient-derived B-cell lines to investigate the genome-scale effects of **DNMT3B dysfunction**. We highlighted that **altered intragenic CpG-methylation** impairs multiple aspects of **transcriptional regulation**, like alternative TSS usage, antisense transcription and exon splicing. These defects preferentially associate with changes of intragenic **H3K4me3** and at lesser extent of **H3K27me3** and **H3K36me3.**

ICF syndrome: a rare autosomal recessive immunological/neurological disorder

- **Immunodeficiency:** agammaglobulinemia, respiratory tract infections (cause of premature death), variable mental retardation
- **Centromeric instability:** gaps, breaks, deletions, isochromosomes, multiradial figures
- **Facial anomalies:** hypertelorism, flat nasal bridge, low set ears, epicanthic folds

Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome

PAOLA MARASCHIO*, ORSETTA ZUFFARDI*†, TIZIANA DALLA FIORE‡, AND LUCIANO TIEPOLO*
*From *Biologia Generale e Genetica Medica, University of Pavia; †Cattedra di Biologia Generale, Facoltà di Medicina, University of Firenze; and ‡Ospedale S Chiara, Trento, Italy.*



Brown et al. Hum Genet.1995
Franceschini et al. Eur J Pediatr. 1995

Published online 9 March 2017

Nucleic Acids Research, 2017, Vol. 45, No. 10 5739–5756
doi: 10.1093/nar/gkx163

ICF-specific DNMT3B dysfunction interferes with intragenic regulation of mRNA transcription and alternative splicing

Sole Gatto[1,2,†], Miriam Gagliardi[1,3,*,†], Monica Franzese[3], Sylwia Leppert[1], Mariarosaria Papa[1], Marco Cammisa[1,4], Giacomo Grillo[5], Guillame Velasco[5], Claire Francastel[5], Shir Toubiana[6], Maurizio D'Esposito[1,7], Claudia Angelini[3] and Maria R. Matarazzo[1,*]

[1]Institute of Genetics and Biophysics 'Adriano Buzzati-Traverso', CNR, Naples 80131, Italy, [2]Sanford Burnham Prebys Medical Discovery Research Institute, La Jolla, CA, USA, [3]Institute for Applied Mathematics 'Mauro Picone', CNR, Naples 80131, Italy, [4]Department of Environmental, Biological and Pharmaceutical Sciences and Technologies, Second University of Naples, Caserta 81100, Italy, [5]CNRS UMR7216, Epigenetics and Cell Fate, Université Paris Diderot, Sorbonne Paris Cité, Paris 75205, France, [6]Molecular Medicine Laboratory, Rambam Health Care Campus and Rappaport Faculty of Medicine, Technion, Haifa, Israel and [7]IRCCS Neuromed, Pozzilli, Italy

Received July 01, 2016; Revised February 07, 2017; Editorial Decision February 24, 2017; Accepted February 28, 2017

ABSTRACT

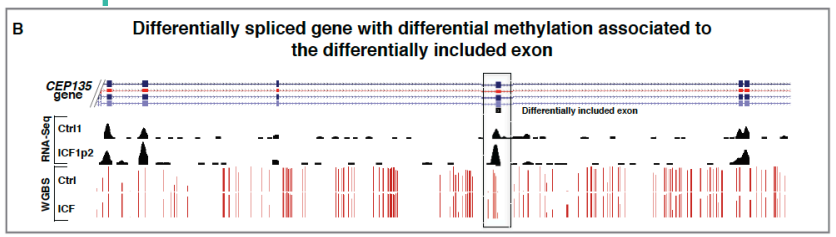
Hypomorphic mutations in DNA-methyltransferase DNMT3B cause majority of the rare disorder Immunodeficiency, Centromere instability and Facial anomalies syndrome cases (ICF1). By unspecified mechanisms, mutant-DNMT3B interferes with lymphoid-specific pathways resulting in immune response defects. Interestingly, recent findings report that DNMT3B shapes intragenic CpG-methylation of highly-transcribed genes. However, how the DNMT3B-dependent epigenetic network modulates transcription and whether ICF1-specific mutations impair this process remains unknown. We performed a transcriptomic and epigenomic study in patient-derived B-cell lines to investigate the genome-scale effects of DNMT3B dysfunction. We highlighted that altered intragenic CpG-methylation impairs multiple aspects of transcriptional regulation, like alternative TSS usage, antisense transcription and exon splicing. These defects preferentially associate with changes of intragenic H3K4me3 and at lesser extent of H3K27me3 and H3K36me3. In addition, we highlighted a novel DNMT3B activity in modulating the self-regulatory circuit of sense-antisense pairs and the exon skipping during alternative splicing, through interacting with RNA molecules. Strikingly, altered transcription affects disease relevant genes, as for instance the memory-B cell marker *CD27* and *PTPRC* genes, providing us with biological insights into the ICF1-syndrome pathogenesis. Our genome-scale approach sheds light on the mechanisms still poorly understood of the intragenic function of DNMT3B and DNA methylation in gene expression regulation.

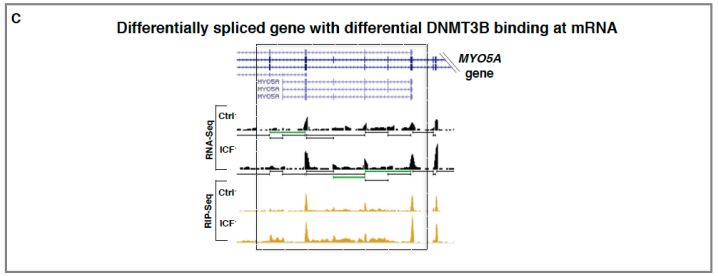
INTRODUCTION

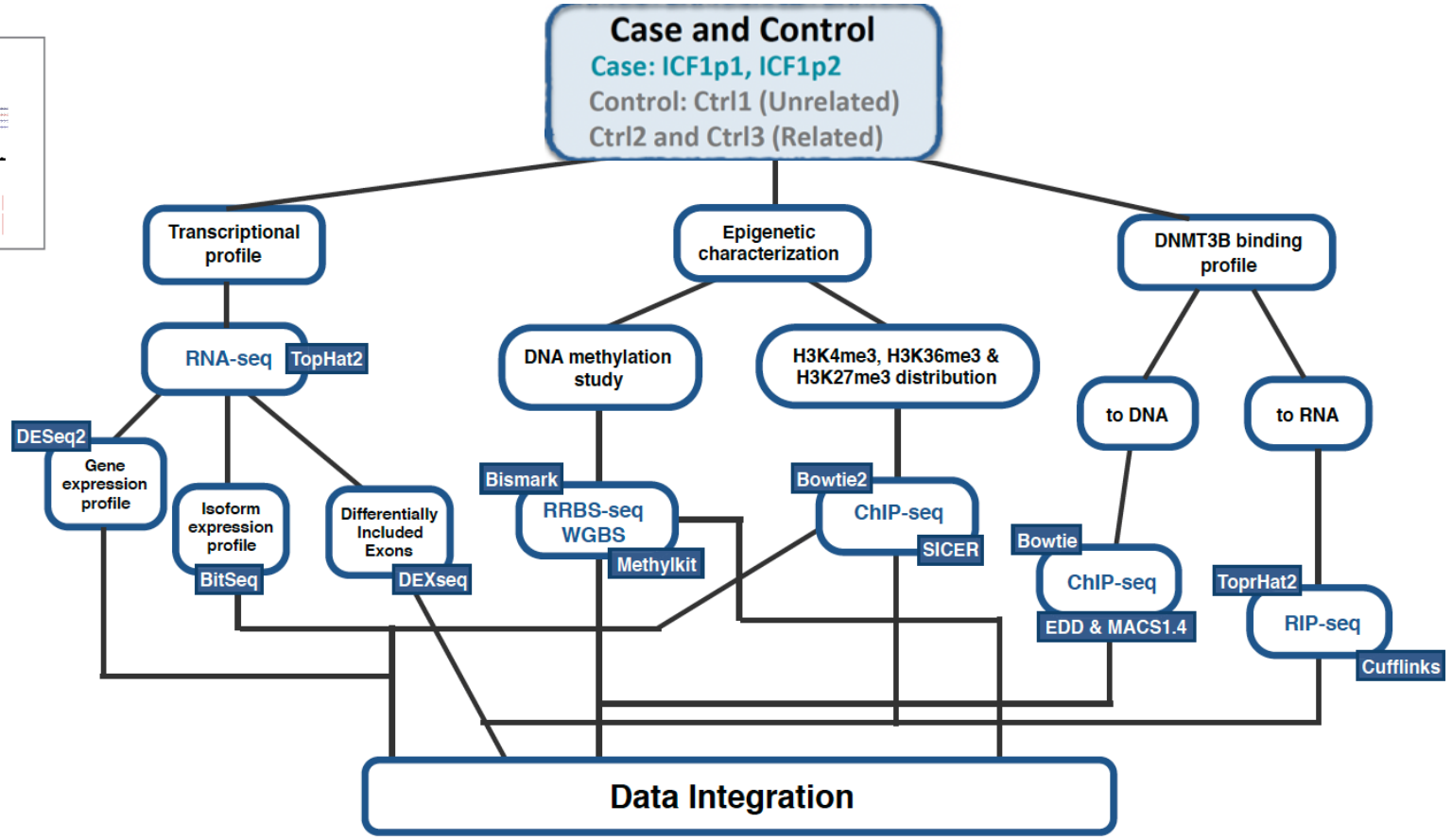
DNA methylation plays an important role in epigenetic signaling, having an impact on gene regulation, chromatin structure, development and disease. Generally, most mammalian genomes are largely methylated except at active or 'poised' promoters, enhancers and CpG islands, where it has a repressive effect. Nevertheless, gene body DNA methylation has been associated with high expression levels (1).

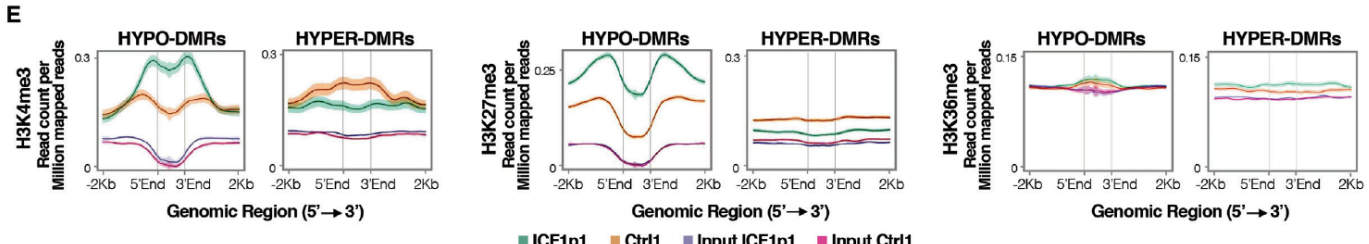
DNA methylation is established and maintained by the combined function of three active DNA methyltransferases DNMT3A, DNMT3B and DNMT1 (2). Although it has been largely studied, much remains unknown regarding how genomic DNA methylation patterns are determined in human cells, and which are the mechanisms that guide recruitment and activity of DNMTs *in vivo* (1). Mouse models suggest that although exhibiting overlapping functions, DNMT3A and DNMT3B have unique expression patterns and genomic targets during development (3–5). In line with

An overview of the analysis



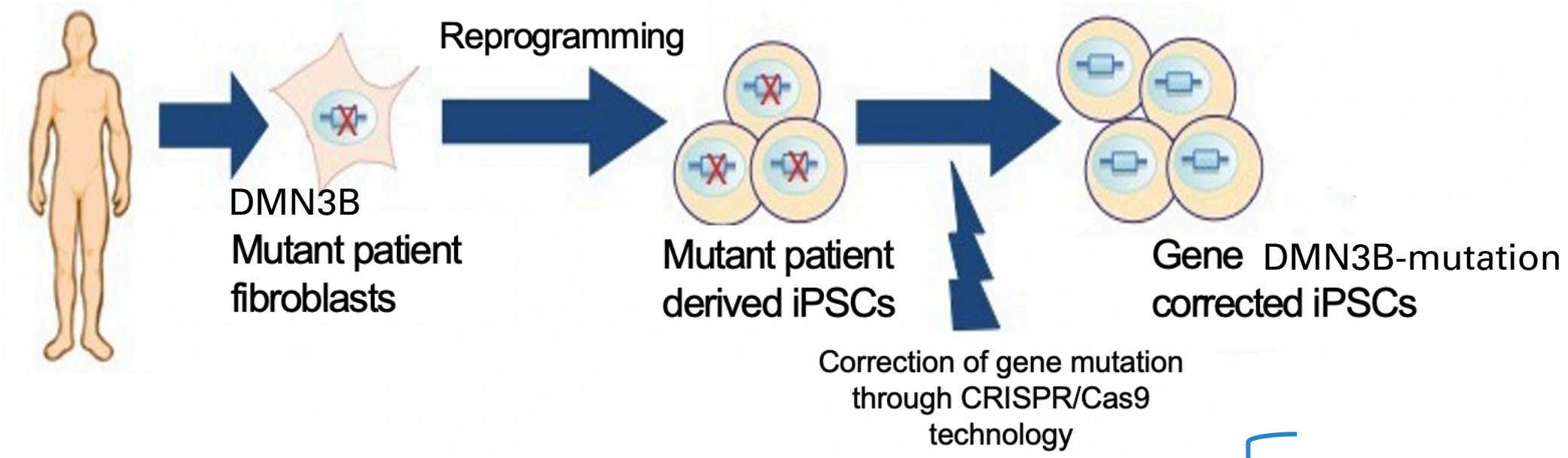






Snap-shot of some ongoing projects

Patient-derived iPSCs and CRISPR-corrected isogenic iPSCs as a model system to study ICF syndrome

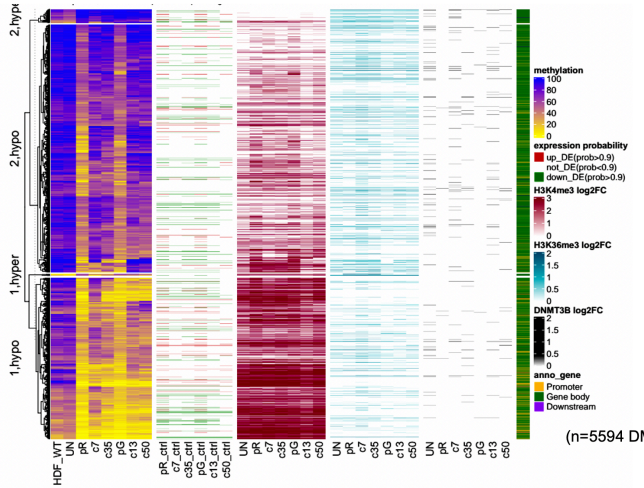


Control iPSCs	Patient-derived iPSCs	CRISPR/Cas9 corrected iPSCs
UN HDF-Wt (Public dataset)	**pR** (homozygous mutation)	**c7; c35**
	pG (compund heterozygous mutation)	**c13; c50**

- ***WGBS-seq***
- ***RNA-seq***
- ***H3K4me3 and H3K36me3***
- ***DMNT3B***

An overview of the analysis

Results will be shown tomorrow at **BBCC2020**



(n=5594 DMRs in genic regions)

Cases &Controls

Control-iPSCs: UN, HDF_Wt (Public dataset)
ICF-patient iPSCs: pR, pG
Corrected-iPSCs: c7,c35 (pR) ;c13,c50 (pG)

Methylation profile

WGBS (PE)

Bismark

MethylKit

Differential Methylation (DMR)

(Diff – 25; q < 0.01)

Gene expression study

RNA-seq (PE)

TopHat2

FeatureCounts

NOIseq

Differential Gene Expression(DE)

(posterior probability > 0.9)

Histone mark enrichment

H3K4me3; H3K36me3

ChIP-seq (SE)

Bowtie2

EPIC2-DF

Differential Enrichment (DER)

H3K4me3 DER
(FDRC- 1e-4
FC- 2)

H3K36me3 DER
(FDRC- 1e-4
FC- 1.3)

DNMT3B binding profile

ChIP-seq (SE)

Bowtie2

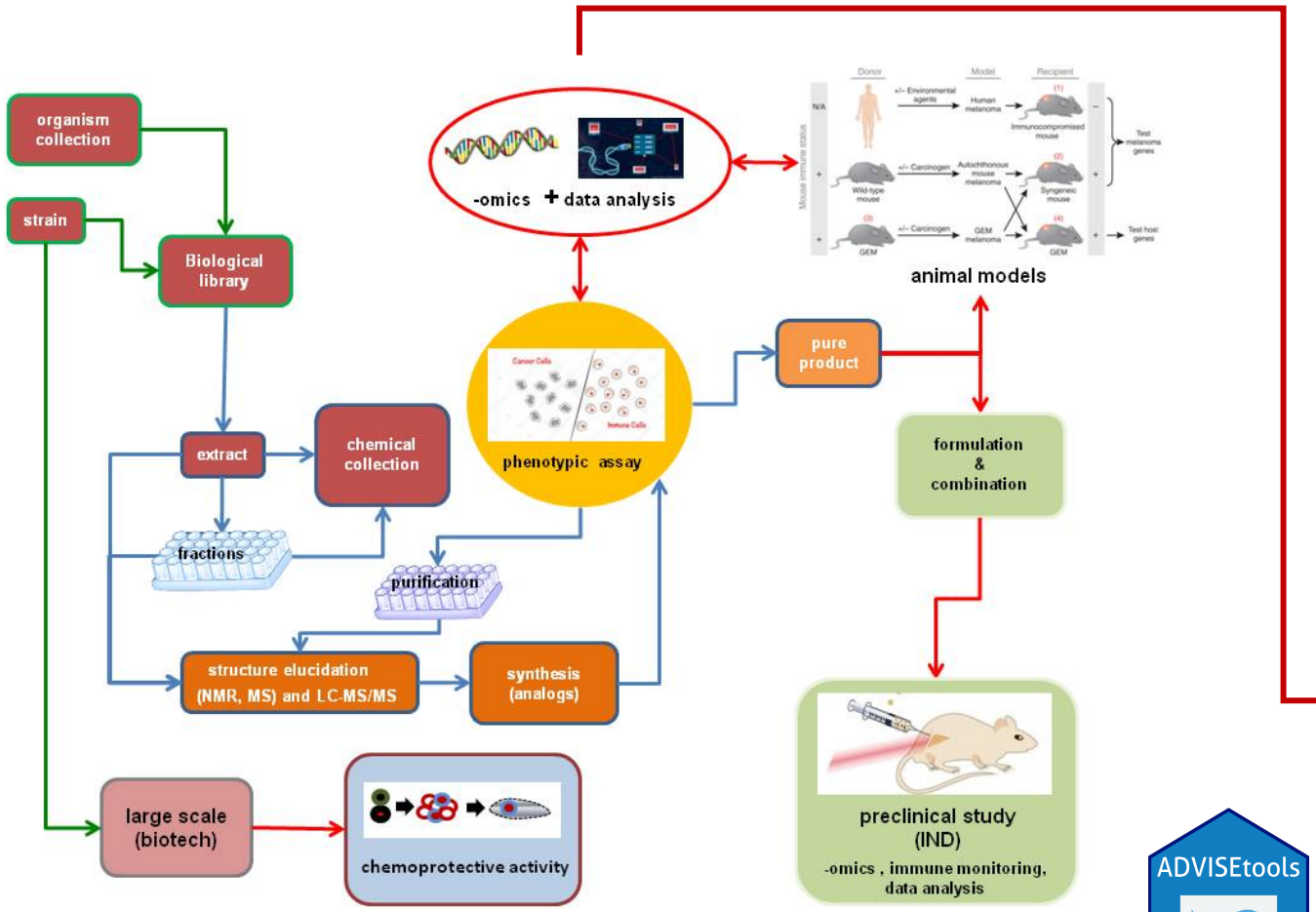
SICER-DF

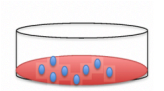
Differential Enrichment (DER)

(FDR- 1e-4; FDRC - 1e-2)

Data Integration

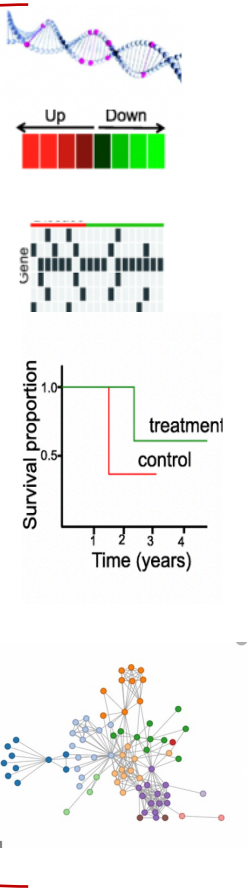
Snap-shot of some ongoing projects





Cell behaviour after challenge with the compound(s) *in vitro*

Tumor-Host interactions *in vivo*



Candidate Targets and Possible Biomarkers

Survival Analysis

COSMONET

Pathway analysis

ADVISEtools

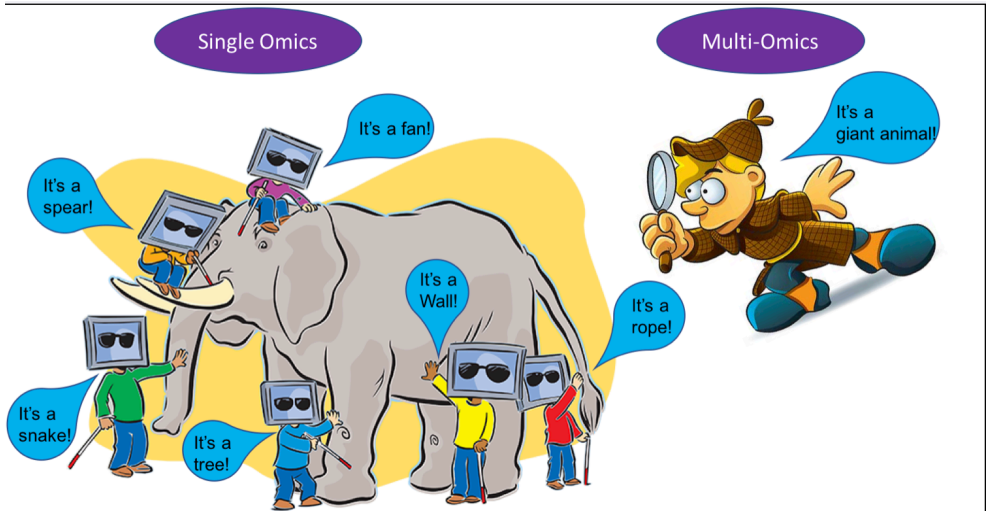
ADVISEarray

Antitumor **D**rugs and **V**acc**i**nes from the **SE**a

(FARMACI E VACCINI ANTITUMORALI DAL MARE)

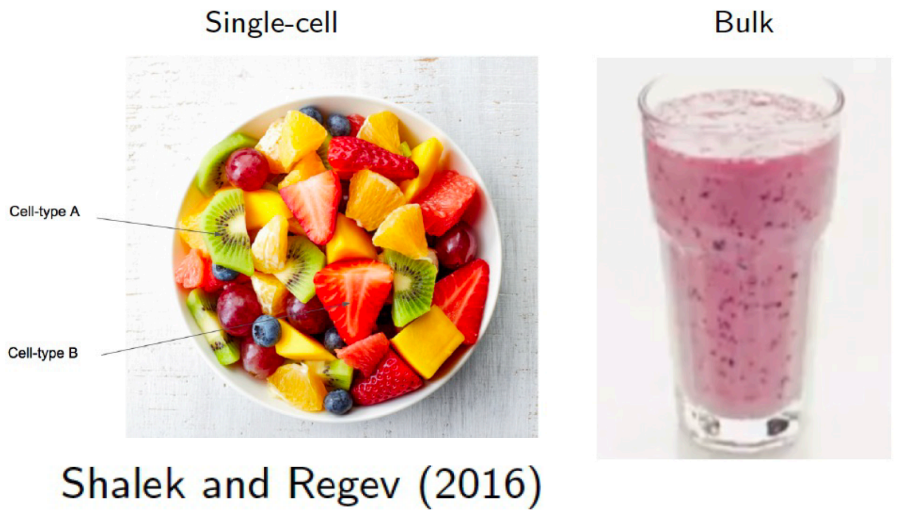
The lesson learned so far….

- Each type of **omics data** provides only a **partial view of cellular mechanisms**.
- Different omics data types requires different **computational methodology**
- **Integration of different omics** data types can better elucidate the potential mechanisms that lead to disease, or the treatment targets
- **Data integration** is an extremely challenging and emerging novel research area that is shifting from **naïve data integration** approaches to **advanced mathematical methods**

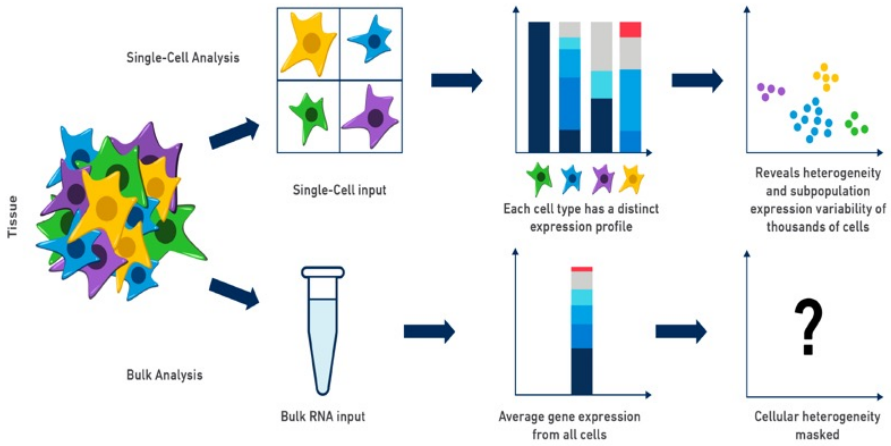


Omics data are multi-omics

Moreover, Omics data goes toward single cell



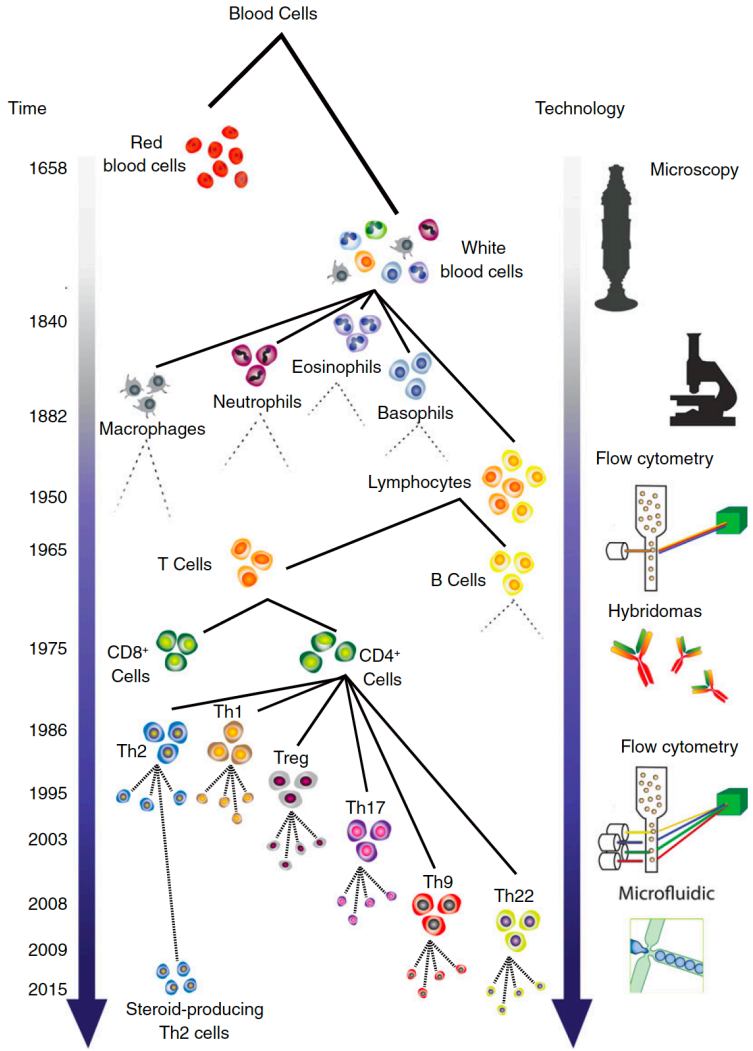
Shalek and Regev (2016)



Example: scRNA-seq experiments

- Human body - trillions of specialized cells with different functions
- **Cells are heterogeneous**… different type or subtypes
- Cell of the same type can be at different stages
- **Classical (Bulk) RNA-seq** data analysis requires hundred of thousand or millions of cells
- **scRNA-seq** allows to investigate individual cell gene expression

What can I investigate with scRNA-seq?



Proserpio & Mahata (2016)

Most relevant scRNA-seq **applications** include

- Study cellular **heterogeneity**
- Discover **novel** cell populations
- Predict **cell fate** differentiation

By-cell analysis

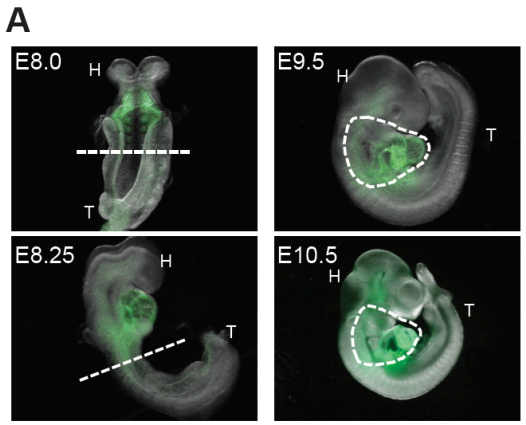
- Detect cell-type specific **differentially** expressed genes
- Understand cell-type specific **regulatory** networks

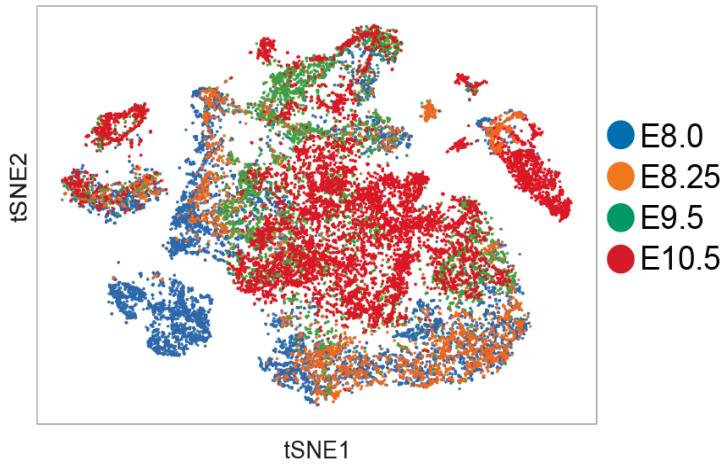
By-gene analysis

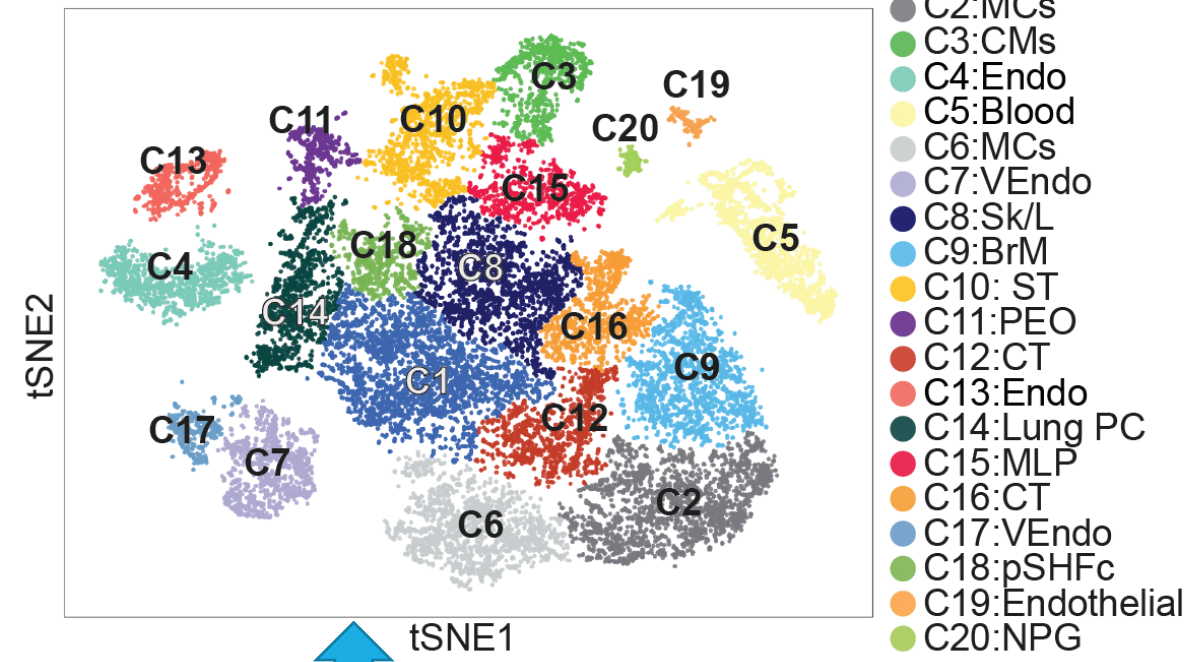
- Etc..

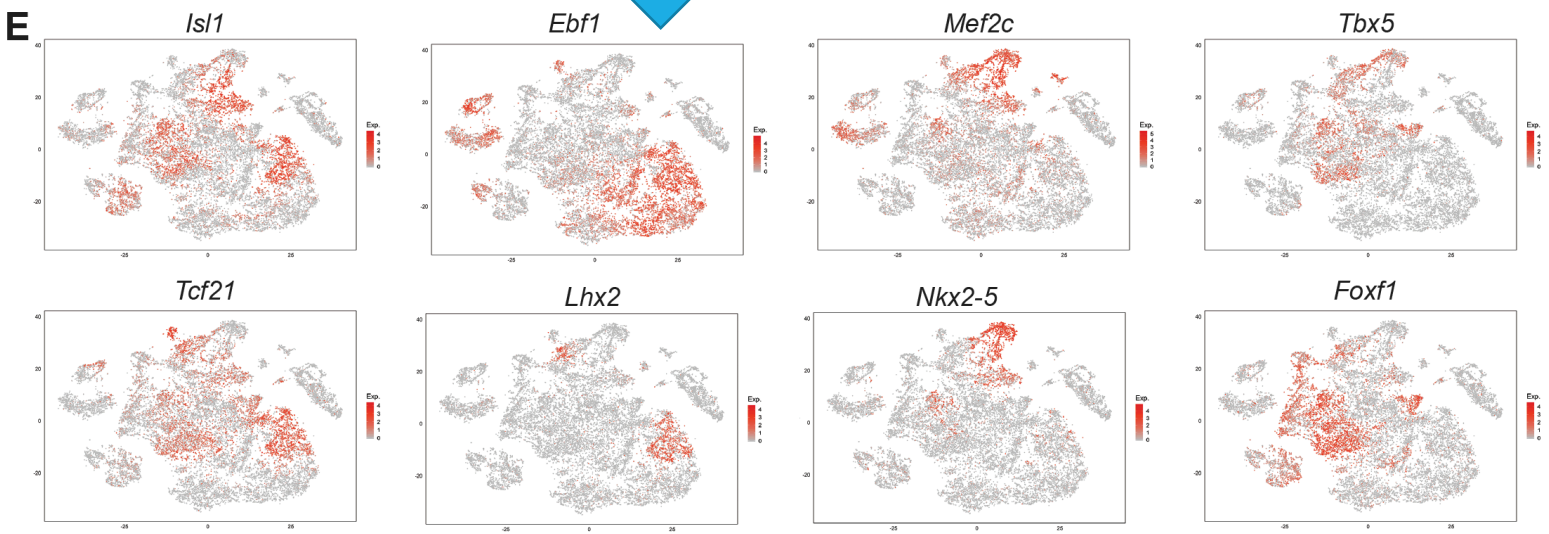
Other applications include
Cell Atlas: Human Cell Atlas, Mouse Cell Atlas,….

Naïve Example









Can scRNA-seq help immunology research ?

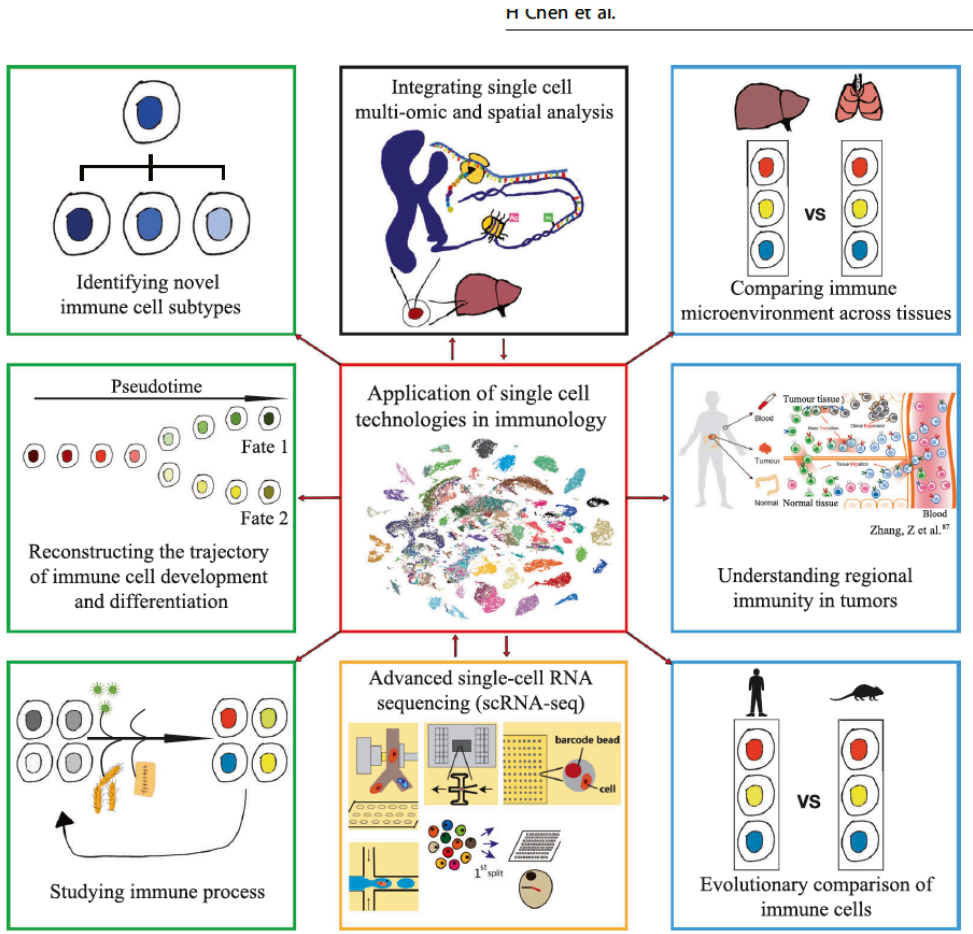


Fig. 1 Mapping the immune cell atlas by single-cell RNA sequencing (scRNA-seq). The advanced technologies in scRNA-seq allows construction of an immune cell atlas at the single-cell level. The immune cell atlas contains the detailed cellular and molecular signatures of immune cells from different physiological as well as pathological contexts, tissues, individuals, and species. scRNA-seq can also be combined with single-cell multi-omic analysis, and spatial gene expression analysis to promote our understanding of the immune system

REVIEW ARTICLE

Revolutionizing immunology with single-cell RNA sequencing

Haide Chen[1,2,3], Fang Ye[1] and Guoji Guo[1,2,3,4,5]

Single-Cell Genomics: A Stepping Stone for Future Immunology Discoveries

Amir Giladi[1] and Ido Amit[1,*]
[1]Department of Immunology, Weizmann Institute, Rehovot 76100, Israel
*Correspondence: ido.amit@weizmann.ac.il
https://doi.org/10.1016/j.cell.2017.11.011

Opinion

Immunology Driven by Large-Scale Single-Cell Sequencing

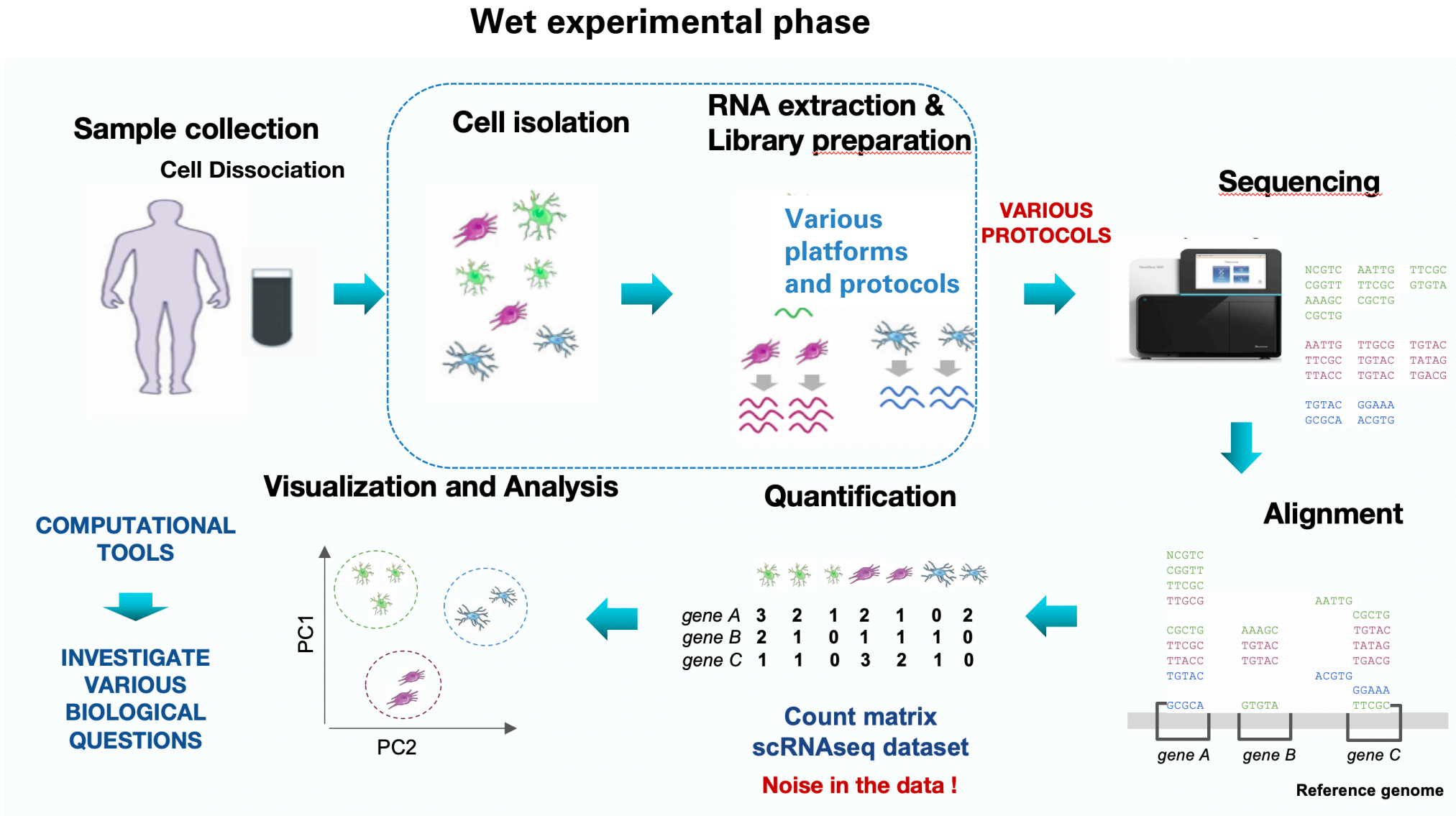
Tomás Gomes,[1] Sarah A. Teichmann,[1,2,3,*] and Carlos Talavera-López[1,2]

REVIEW

Single-cell transcriptomics to explore the immune system in health and disease

Michael J. T. Stubbington,[1]* Orit Rozenblatt-Rosen,[2]*
Aviv Regev,[2,3]†‡ Sarah A. Teichmann[1,4]†‡

ScRNA-seq overview

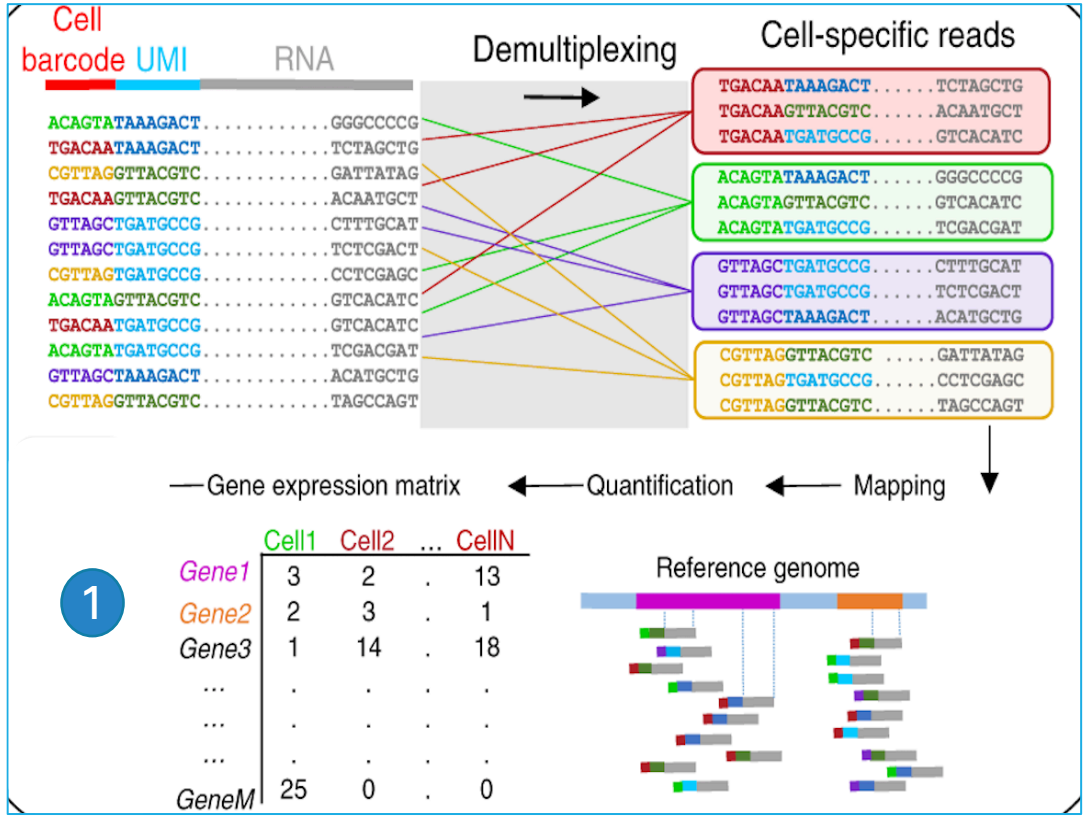


scRNA-seq data analysis overview

- **Read alignment**
- **Count matrix**
- **Quality control** and **filtering**

1

1 **<u>Low-level preprocessing</u>**: from sequence to raw-count matrices

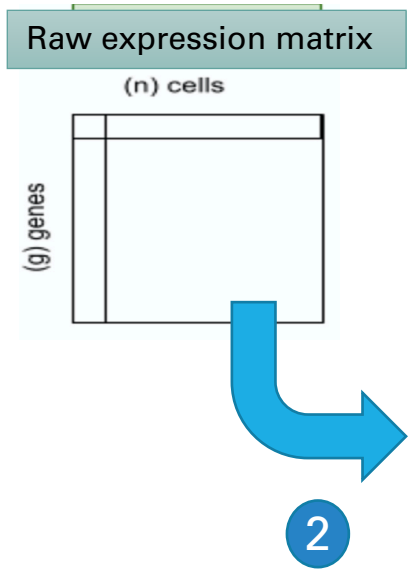


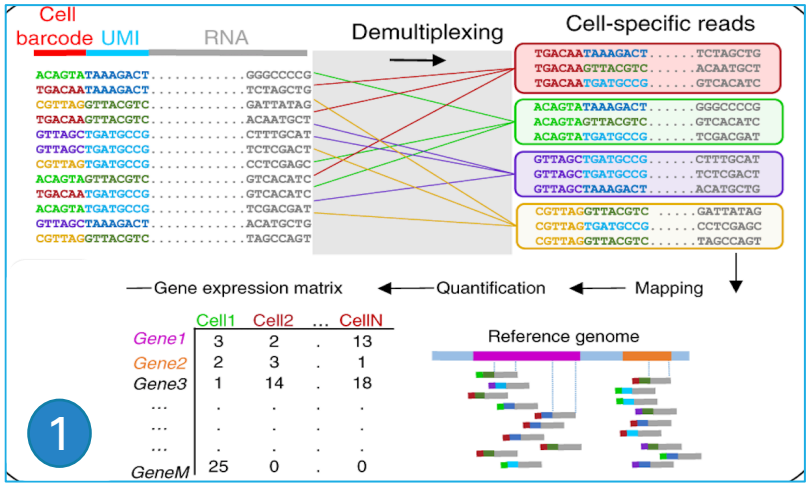
scRNA-seq data analysis overview

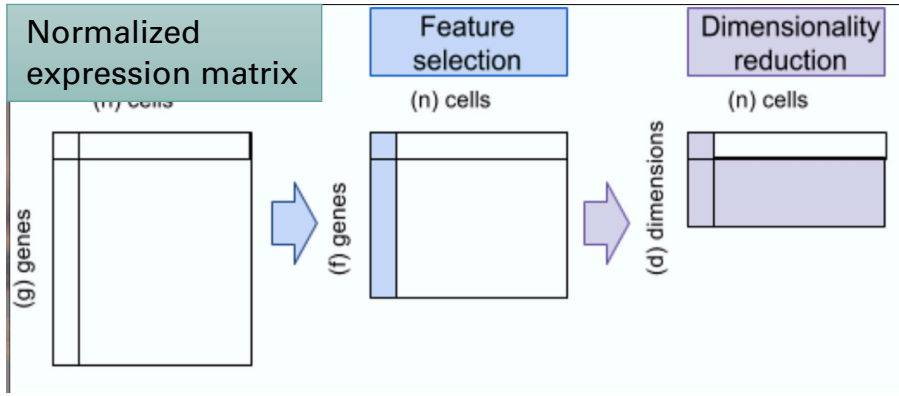
- Read alignment
- Count matrix
- Quality control and filtering
- **Normalization** and **Batch effect removal**
- **Feature selection** and **dimension reduction**

1

2







2 **<u>High-level preprocessing</u>**: from raw-data to "clean data signal"

scRNA-seq data analysis overview

- Read alignment
- Count matrix
- Quality control and filtering
- Normalization and Batch effect removal
- Feature selection and dimension reduction
- **Downstream analysis**

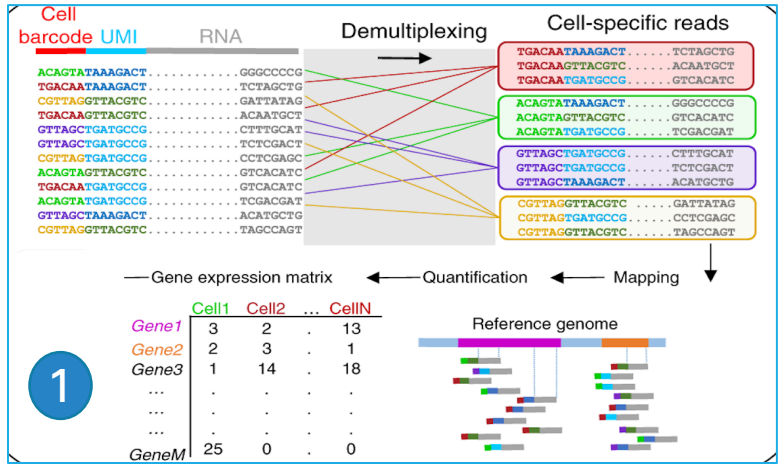
(clustering, trajectory inference, differential expression,…)

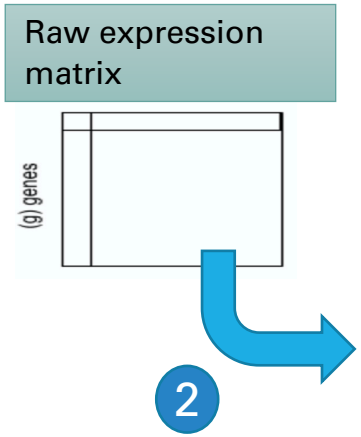
1

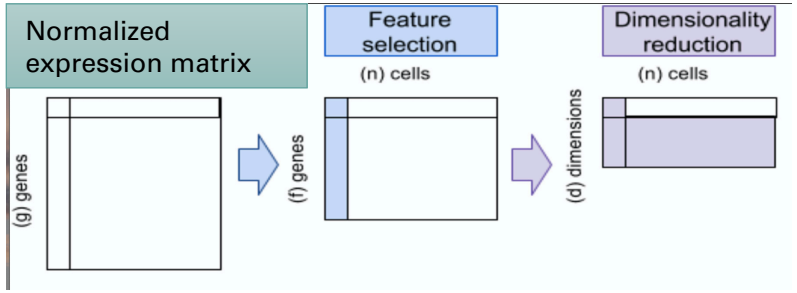
2

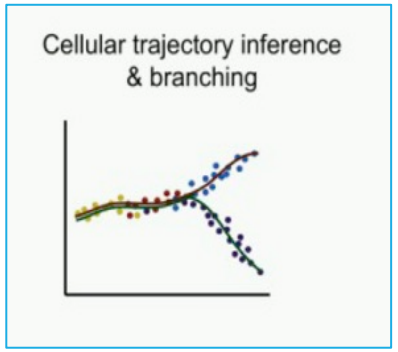
3

3 **Data analysis** : Extracting signal from data









scRNA-seq data analysis overview

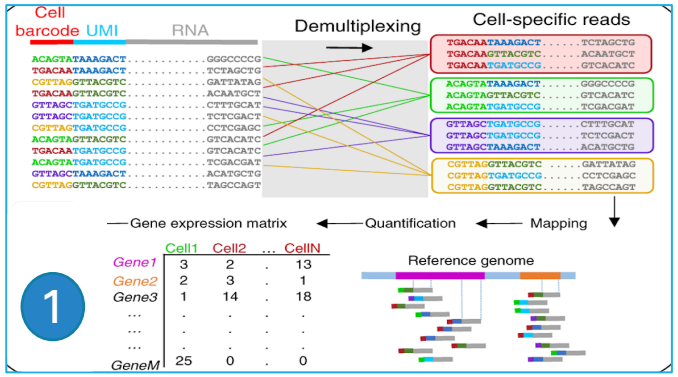
- Read alignment
- Count matrix
- Quality control and filtering
- Normalization and Batch effect removal
- Feature selection and dimension reduction
- Downstream analysis (clustering, trajectory inference, differential expression,…)
- **Interpretation** (cell type identification, marker detection, novel cell type functions, etc…)

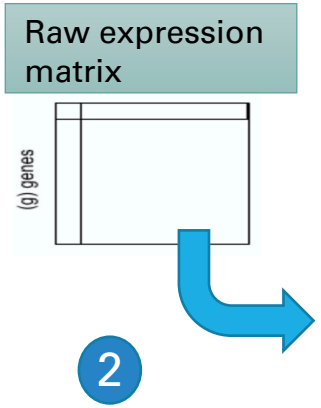
1

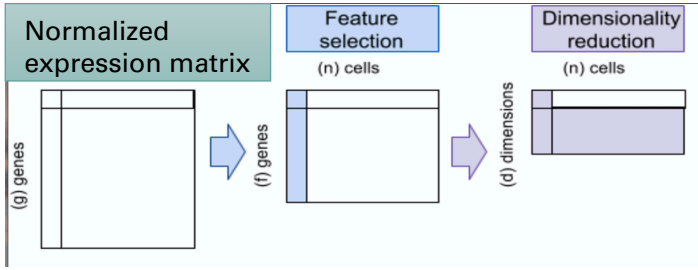
2

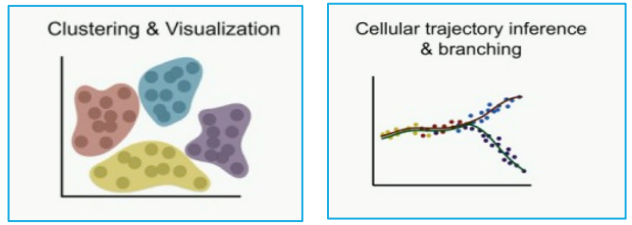
3

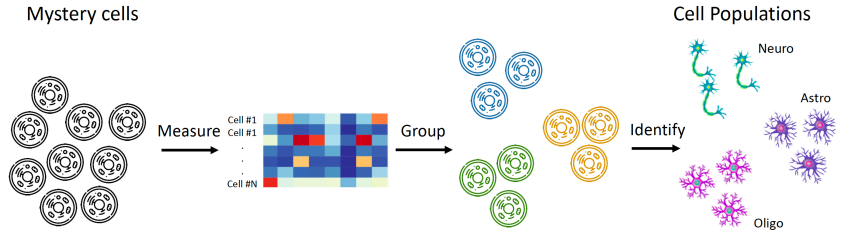
4





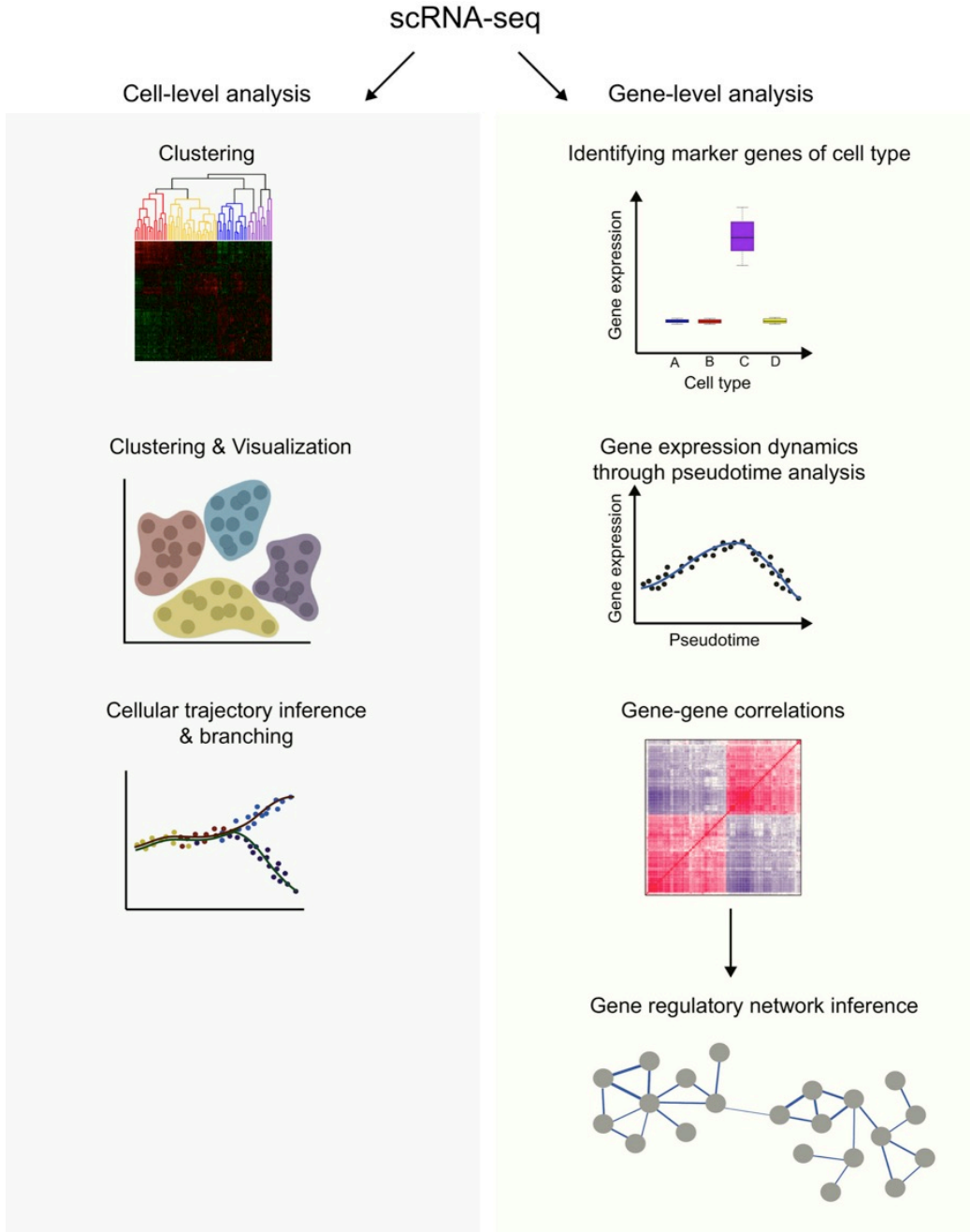




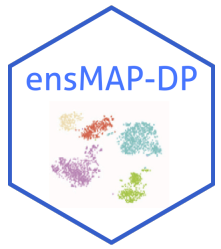


ScRNA-seq downstream analysis

- **Clustering**
- Cluster annotation
- Compositional analysis
- Trajectory inference and branching
- Cell marker identification
- Differential expression
- Gene regulatory networks
- Others… continuosly emerging

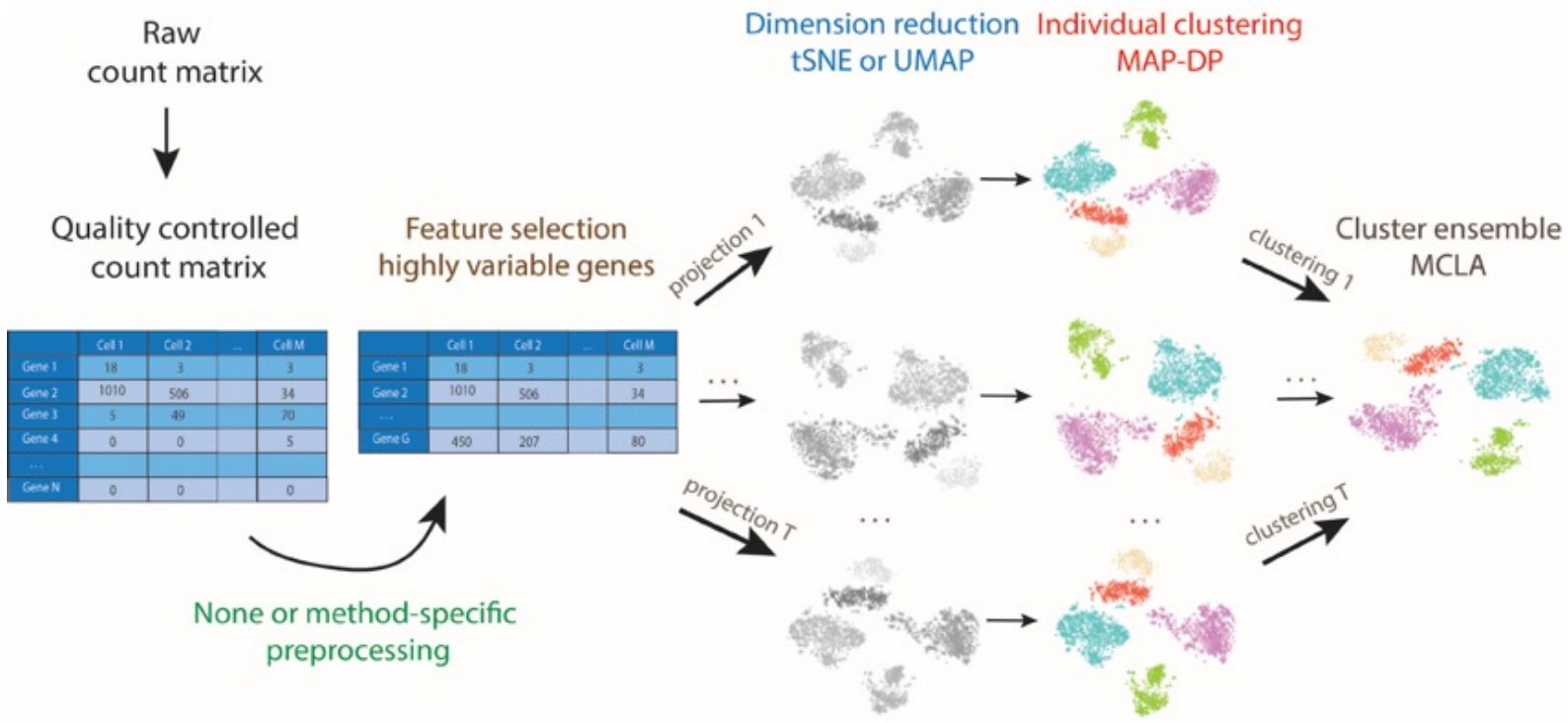


EnsMAP-DP algorithm



The algorithm is based on four main steps:

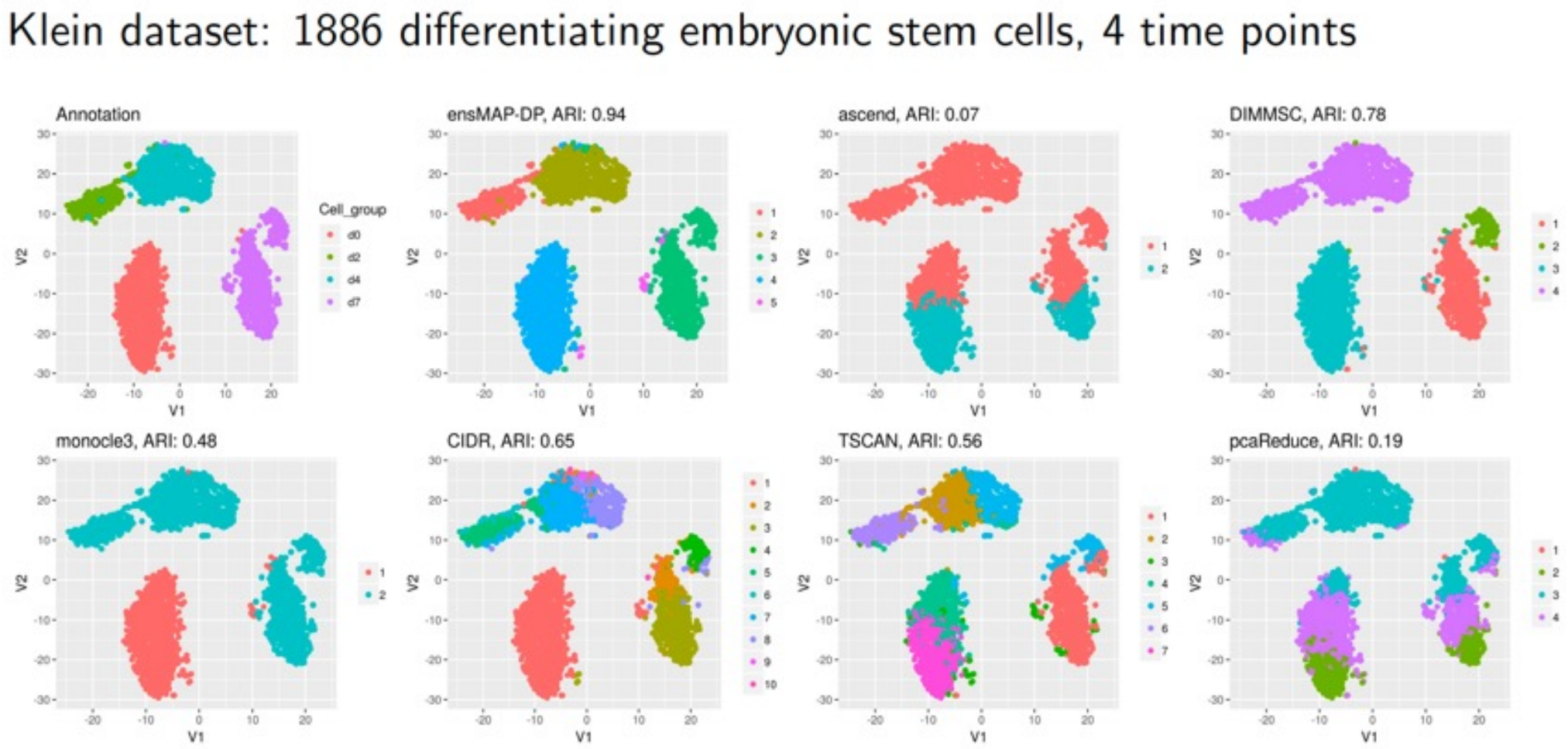
- ✓ Feature selection
- ✓ Dimension reduction
- ✓ Clustering
- ✓ Cluster ensemble



Based on the **Maximum a Posteriori Dirichlet Process Mixture Model** (MAP-DP)

- It assumes that data follows a *Gaussian Infinite Mixture* model
- It estimates both cluster allocation and the number of clusters (cell populations) through the posterior inference
- It uses a direct maximization formulation to speed-up the computation

EnsMAP-DP compared with other existing approaches



Conclusions

- scRNA-seq allows to investigate **cell heterogeneity** and **development**
- scRNA-seq can be very useful in **immunolgical studies,** cancer and cell development
- scRNA-seq data are by far noisier and sparser than bulk RNA-seq data, but there are many cells → **novel computational methods are required**
- Dataset size is increasing→ **scalability of methods** in terms of memory and running time
- **Novel applications** of scRNA-seq are continuolsy proposed (i.e., spatial single cell)
- And now…. Single cells are becoming **multi-omics**

Challenges and emerging directions in single-cell analysis

Guo-Cheng Yuan, Long Cai, Michael Elowitz, Tariq Enver, Guoping Fan, Guoji Guo, Rafael Irizarry, Peter Kharchenko, Junhyong Kim, Stuart Orkin, John Quackenbush, Assieh Saadatpour, Timm Schroeder, Ramesh Shivdasani & Itay Tirosh

A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications

Ashraful Haque, Jessica Engel, Sarah A. Teichmann & Tapio Lönnberg

Single-cell multiomics: technologies and data analysis methods

Jeongwoo Lee, Do Young Hyeon & Daehee Hwang

Experimental & Molecular Medicine **52**, 1428–1442(2020) | Cite this article

Integrative single-cell analysis

Tim Stuart & Rahul Satija

Nature Reviews Genetics **20**, 257–272(2019) | Cite this article

Resource

An Integrated Multi-omic Single-Cell Atlas of Human B Cell Identity

David R. Glass [1, 2, 7], Albert G. Tsai [2, 7], John Paul Oliveria [2, 3], Felix J. Hartmann [2], Samuel C. Kimmey [2, 4], Ariel A. Calderon [1, 2], Luciene Borges [2], Marla C. Glass [5], Lisa E. Wagar [6], Mark M. Davis [6], Sean C. Bendall [1, 2, 8]

Single-Cell Multiomics: Multiple Measurements from Single Cells

Iain C. Macaulay [1], Chris P. Ponting [2, 3], Thierry Voet [2, 4]


THANK
YOU!


QUESTIONS