

单细胞组学算法的研究热点和挑战

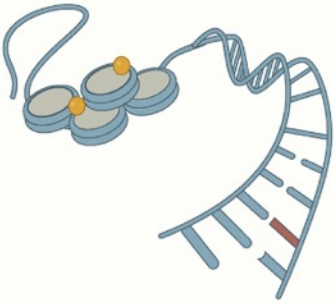
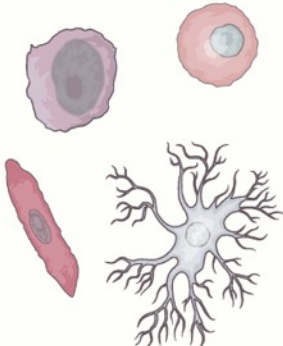
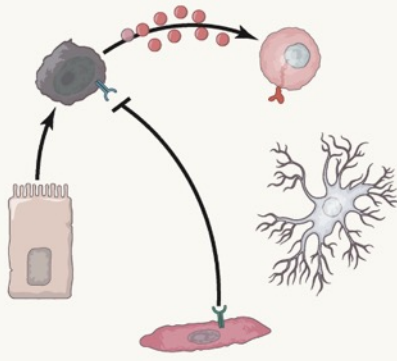
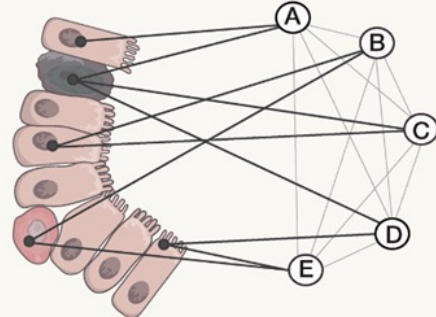

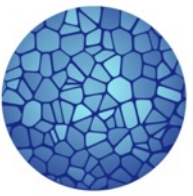

白寅琪

杭州华大生命科学研究院

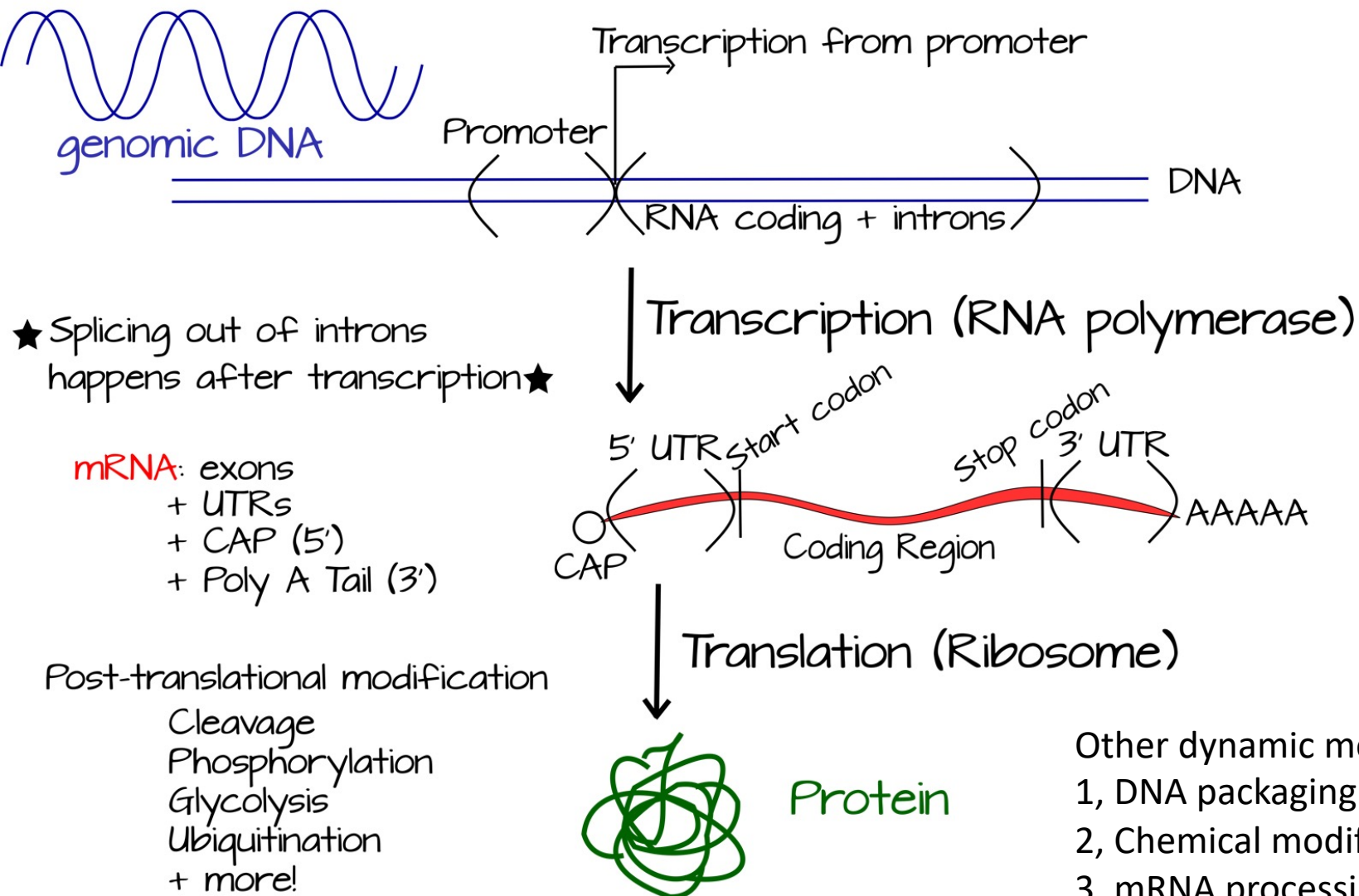
baiyinqi@genomics.cn

2023/10/24

同人类基因组计划一样，人类单细胞组学计划具有重大战略意义

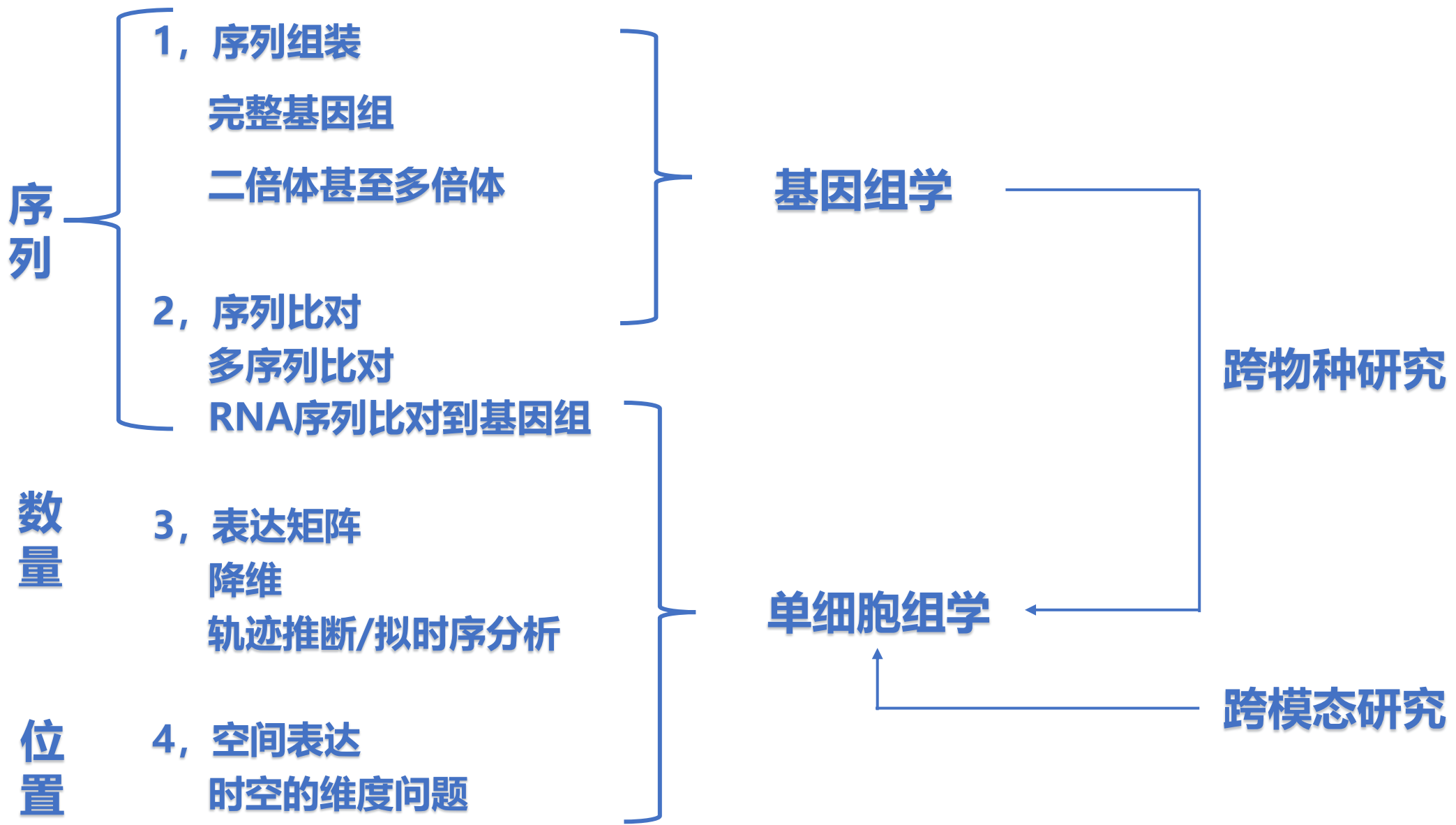
| 技术突破 提升解析维度 | 测序技术 | 单细胞技术 | 时空组学技术 | |
|------------------|---|--|---|---|
| | 核酸 DNA | 细胞 Cell | 空间 Spatial | |
| 科学问题 不断深入/复杂 | 基因结构/突变 | 细胞组成/类型 | 细胞互作/发育调控 | 结构与功能/器官形成 |
| |  |  |  |  |
| 大科学计划 系统化/工程化 |  1990-2003 HGP (美国) |  HUMAN CELL ATLAS 2017- HCA (美国) |  SPATIO TEMPORAL OMICS CONSORTIUM 2022- STOC (国际) | |

Central Dogma of Molecular Biology



| 类型 | 信号 | 维度 | 分辨率 |
|----|----|----|-----|
| 序列 | 连续 | 零 | 亚细胞 |
| 数量 | | ↓ | ↓ |
| 位置 | 离散 | 高 | 个体 |

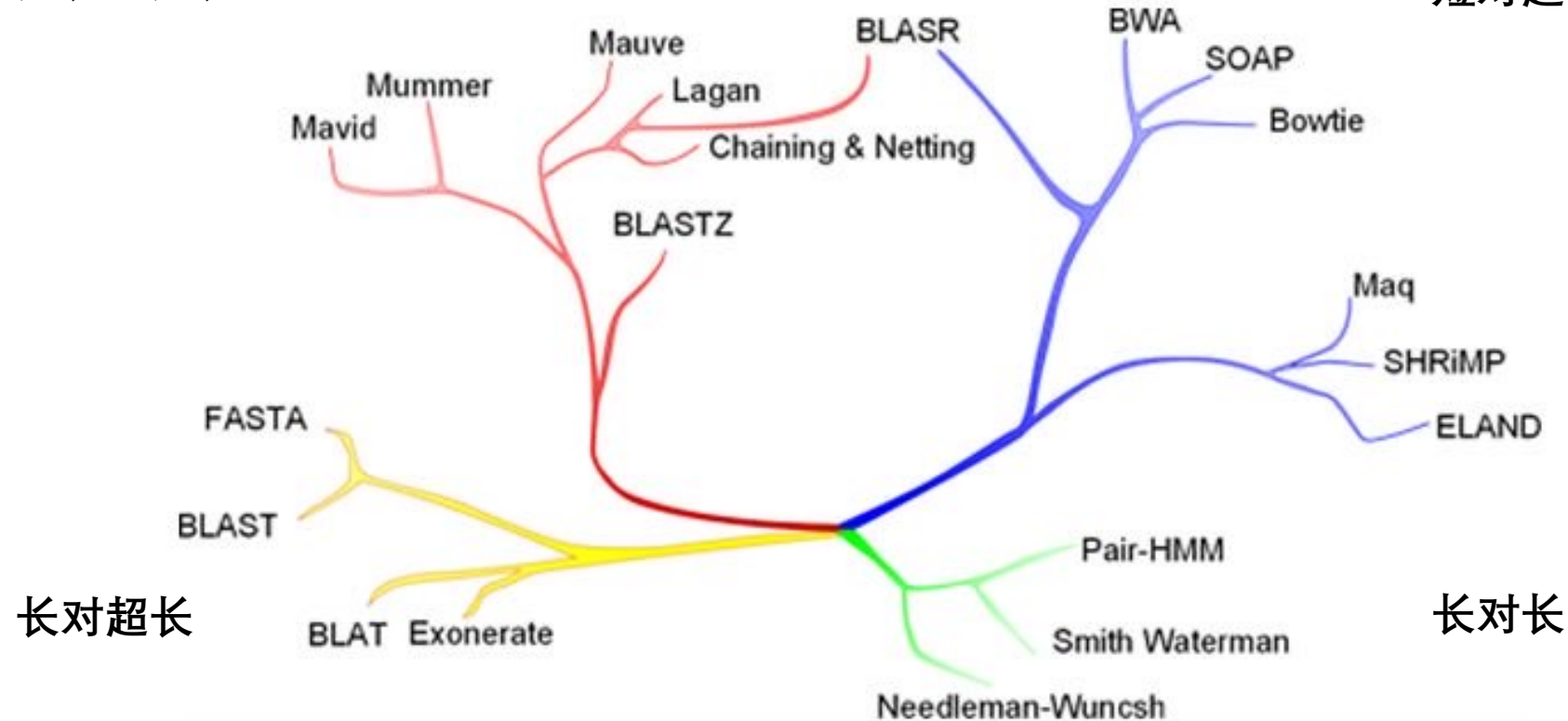
- Other dynamic modulations:
- 1, DNA packaging
 - 2, Chemical modifications
 - 3, mRNA processing



Aligner phylogeny

超长对超长

短对超长



超长：基因组 (~Mb/Gb)

长：基因 (~kb)

短：~ 100bp reads

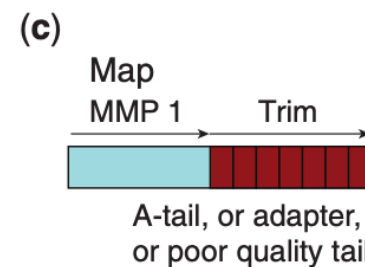
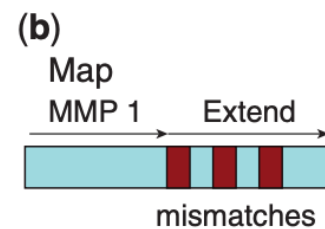
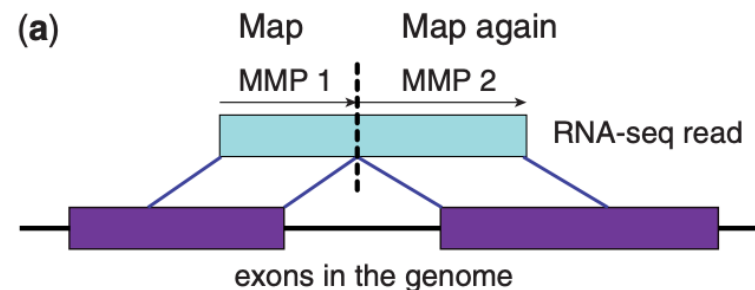
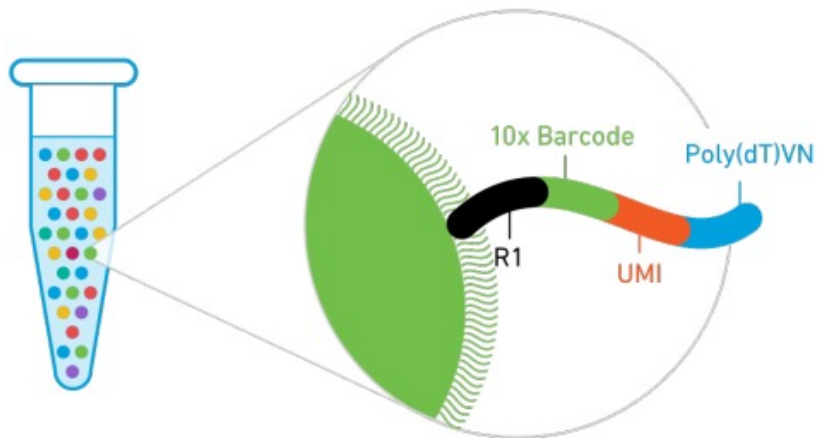
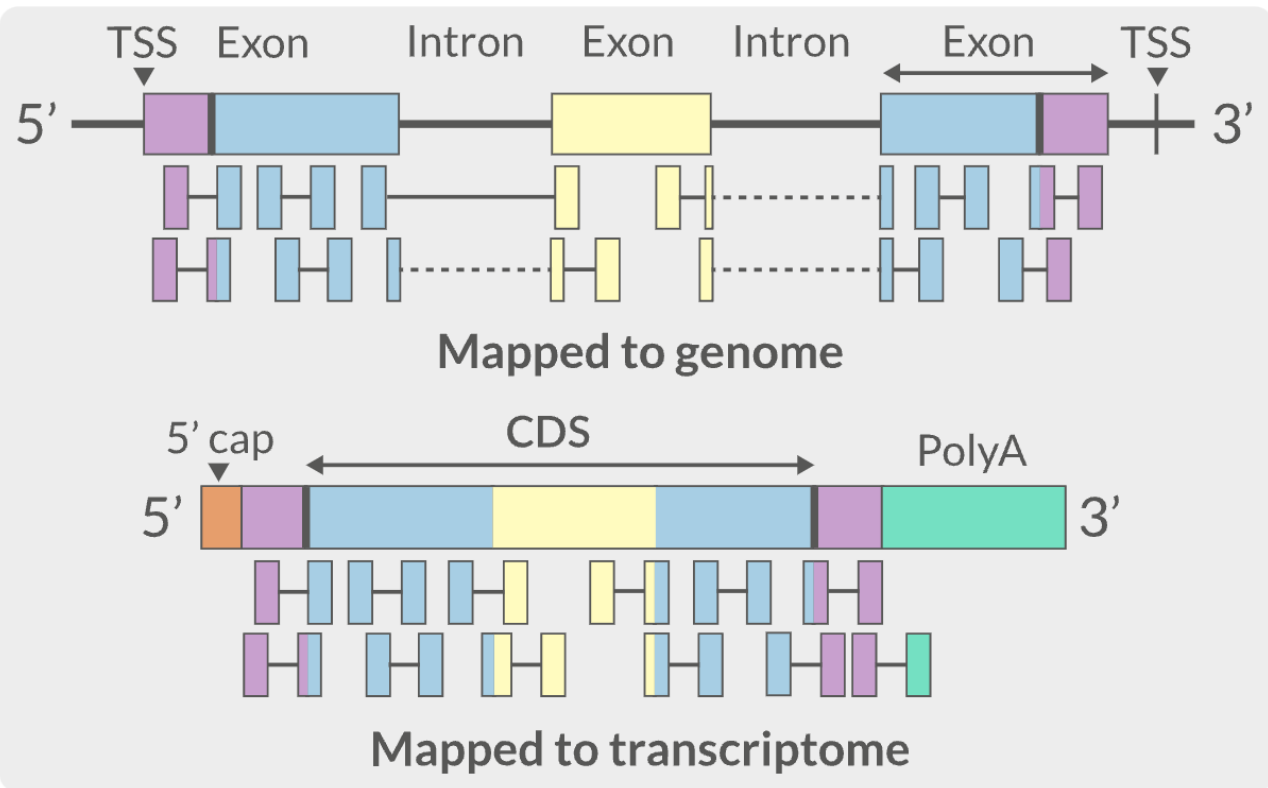
长对超长

长对长

Whole genome
Pairwise heuristic

Short read
Sensitive global aligners

单细胞序列比对的迷思



Reads

FastQ

FastQ

FastQ



Mapping

STAR

HiSat2

[Kallisto/
Salmon]



Quantification

featureCounts

StringTie



Differential
gene expression

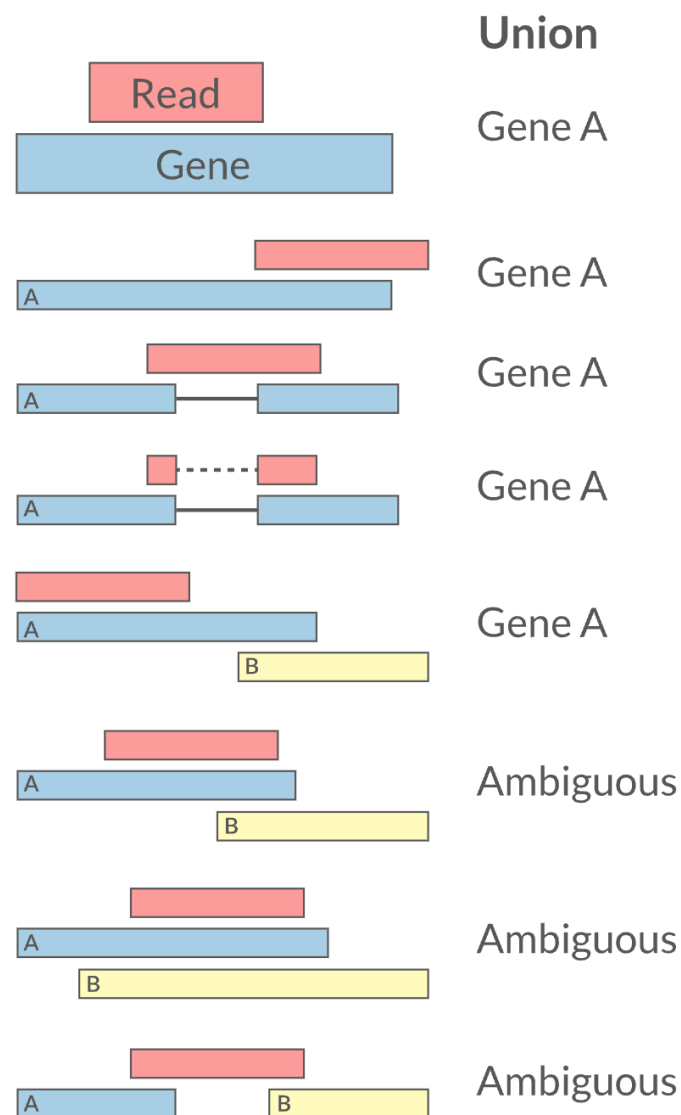
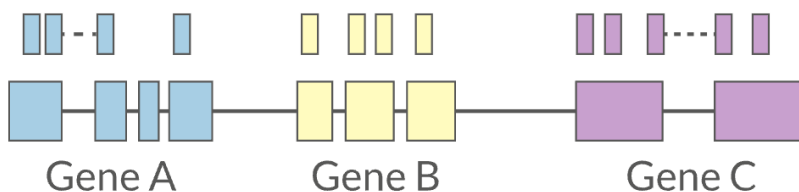
DESeq2/
edgeR/
Limma

Ballgown

Sleuth

RNA定量与参考序列的优化

- Read counts = gene expression
- Reads can be quantified on any feature (gene, transcript, exon etc)
- Intersection on gene models
- Gene/Transcript level



nature methods

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature methods](#) > [articles](#) > [article](#)

Article | [Published: 11 September 2023](#)

Recovery of missing single-cell RNA-sequencing data with optimized transcriptomic references

[Allan-Hermann Pool](#) , [Helen Poldsam](#), [Sisi Chen](#), [Matt Thomson](#) & [Yuki Oka](#) 

[Nature Methods](#) (2023) | [Cite this article](#)

3722 Accesses | 89 Altmetric | [Metrics](#)

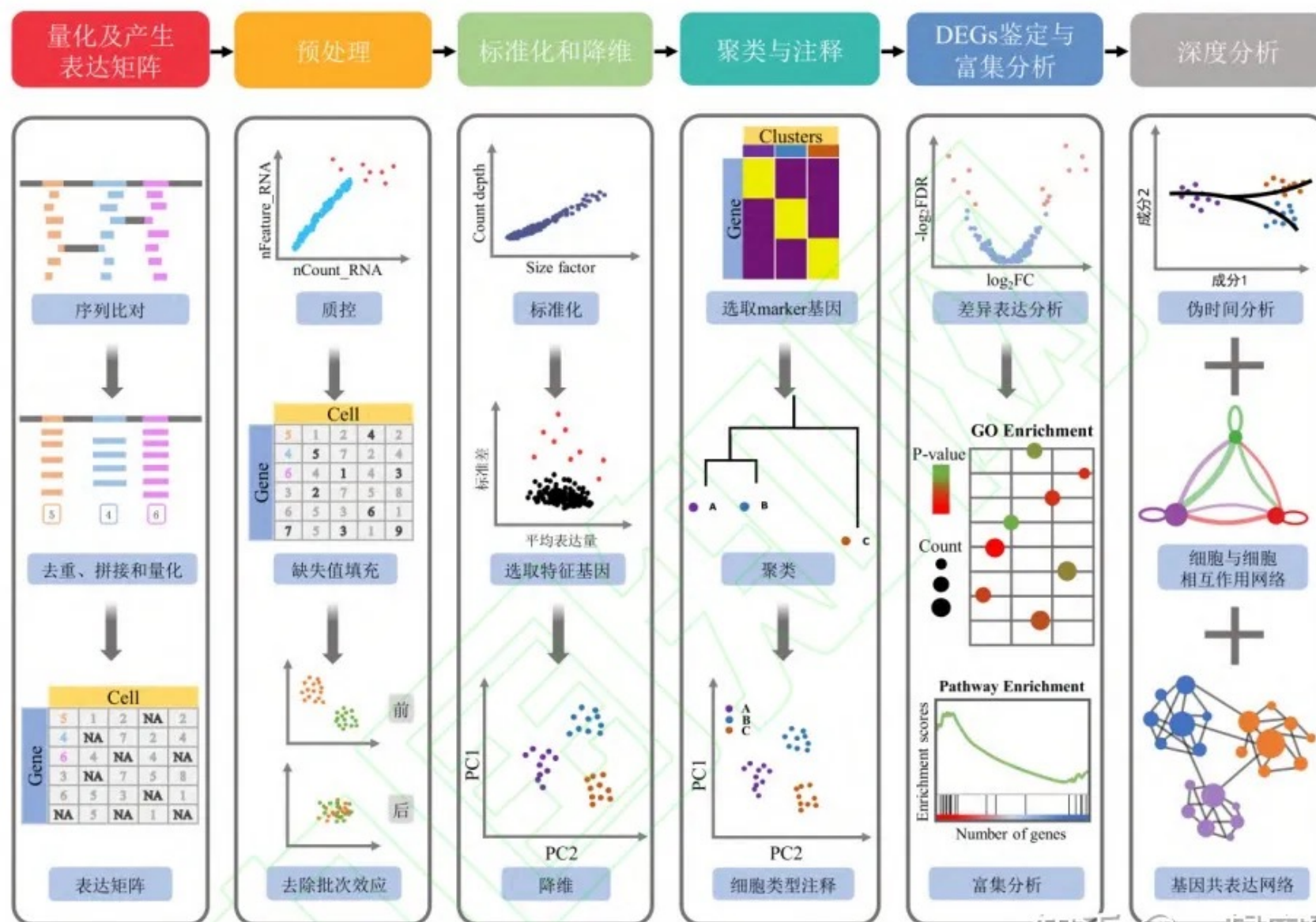
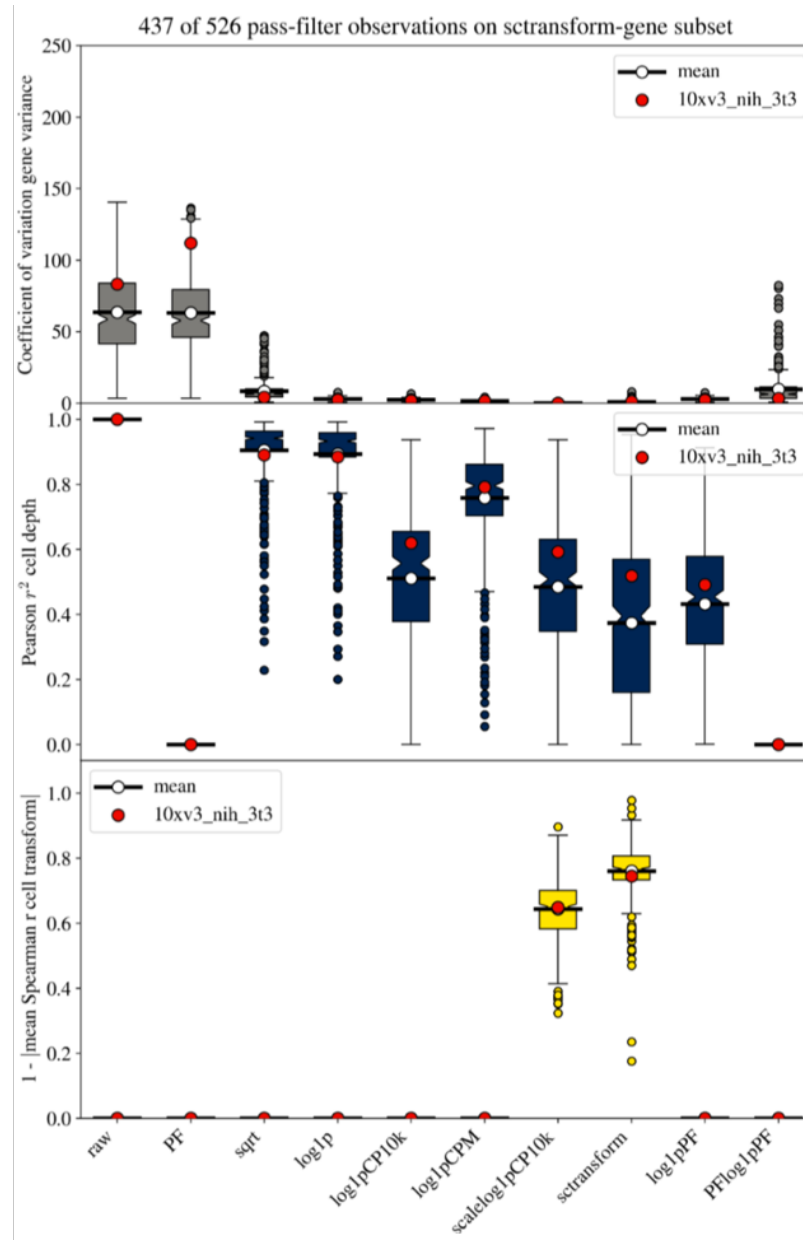
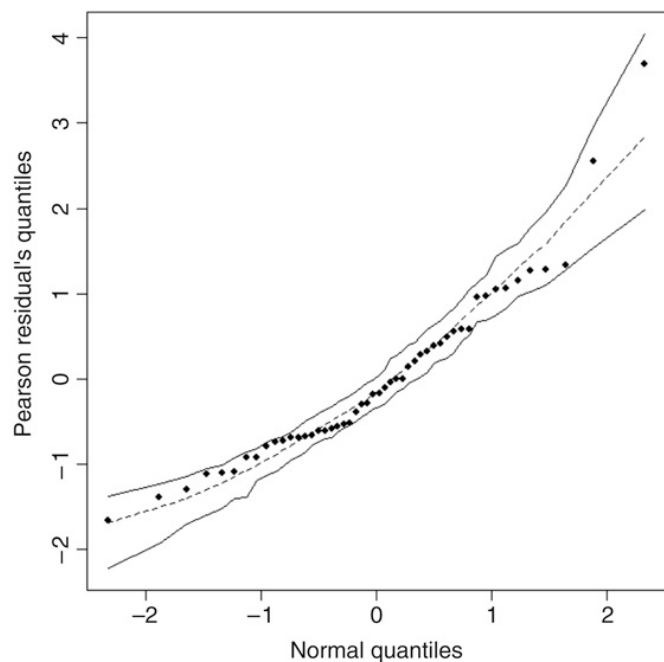
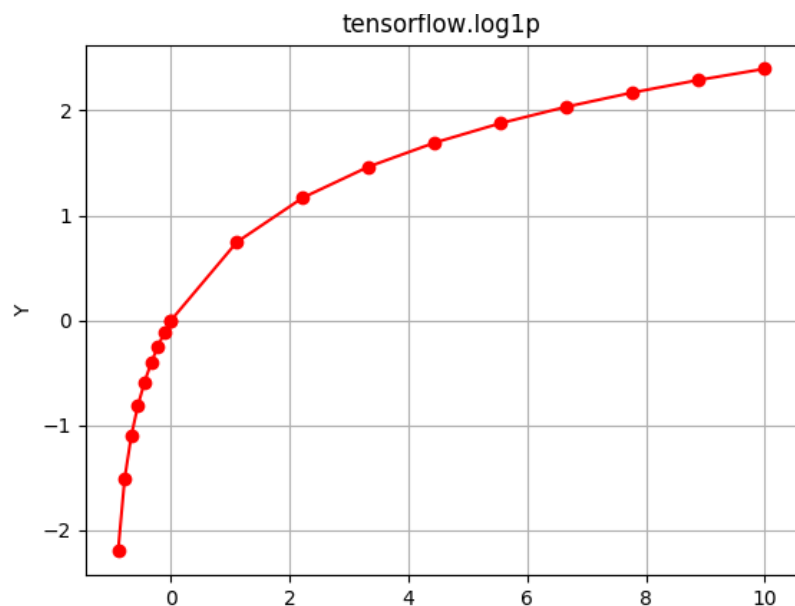


图2 单细胞 RNA 测序数据分析的流程和环节

单细胞算法的挑战：数据归一化

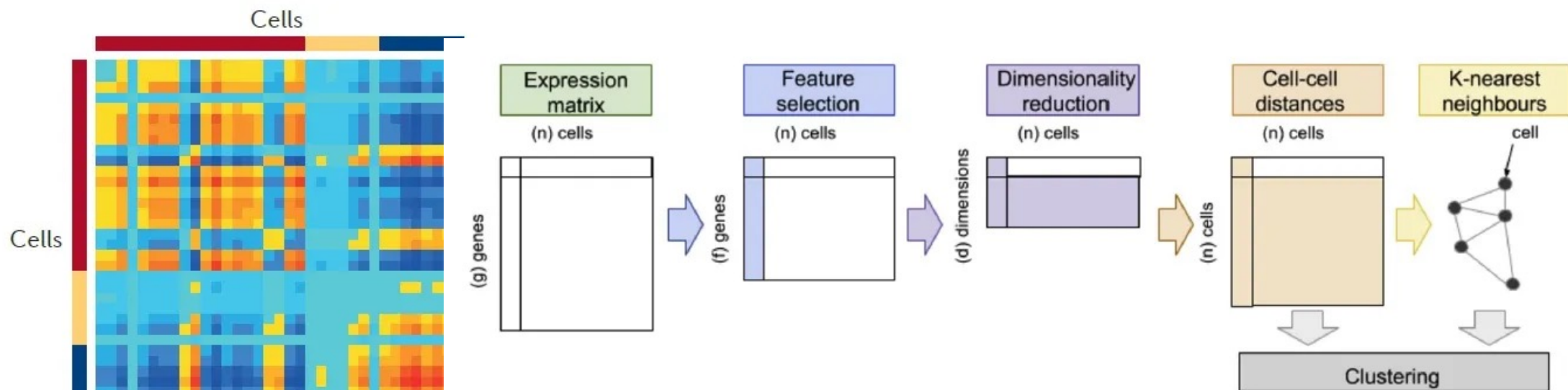


Variance stabilization

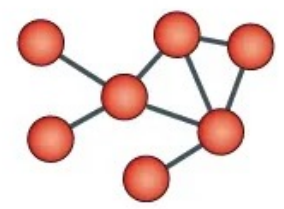
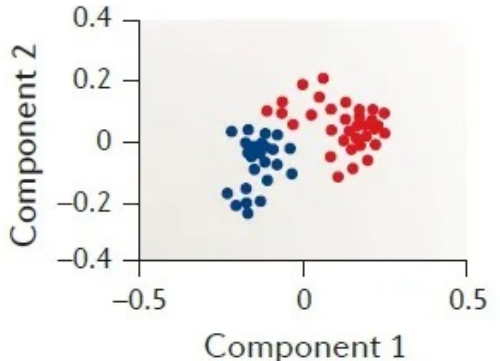
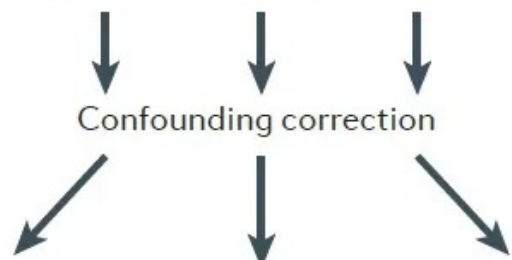
Depth normalization

Monotonicity

单细胞组学算法的挑战：降维



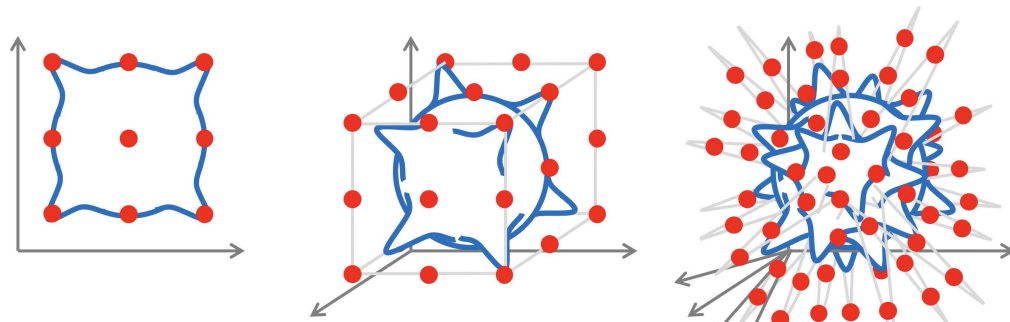
- Main methods:
- HVG, M3Drop, Spike-ins
 - PCA, tSNE, Diffusion maps
 - K-means, Hierarchical, DBSCAN, Louvain, Infomap, SNN-Cliq



单细胞组学算法的挑战：轨迹推断

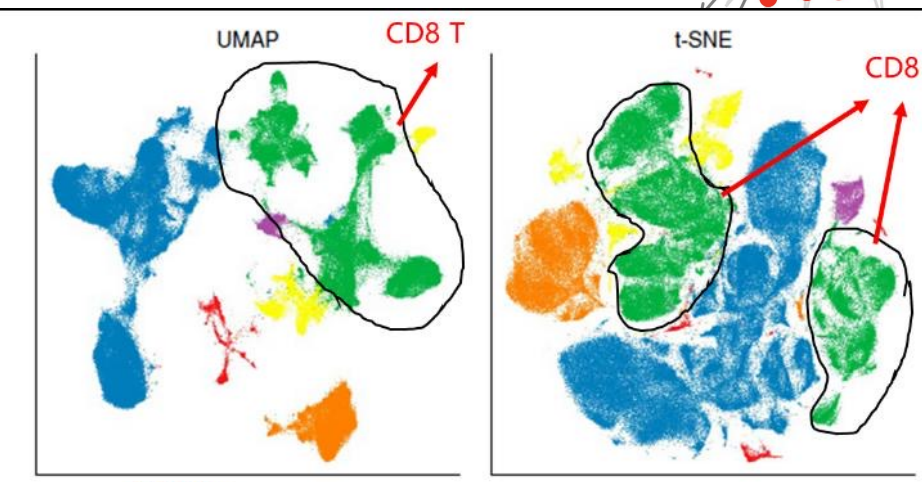
单细胞数据的基本概念

- 1) 超高维度 -> 维度灾难 (样本点不够, 点容易在角落)
- 2) 数据高度稀疏 (所有点都在角落)
- 2) 降维 -> 带来损失 -> 再降到二维可视化

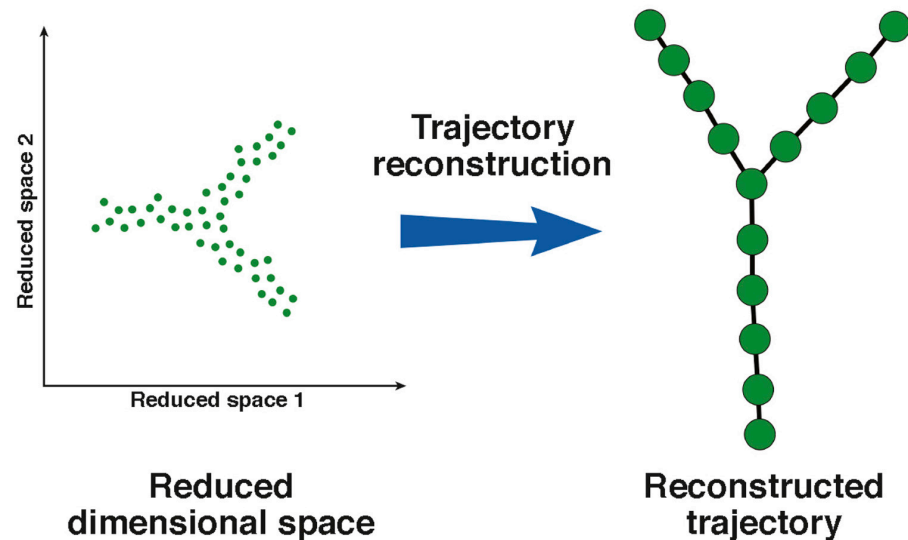
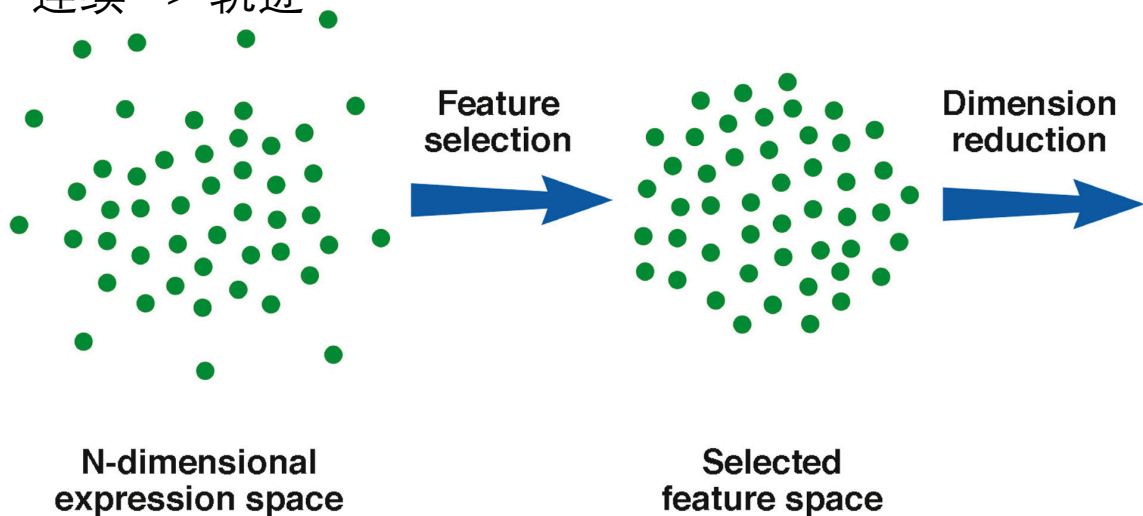


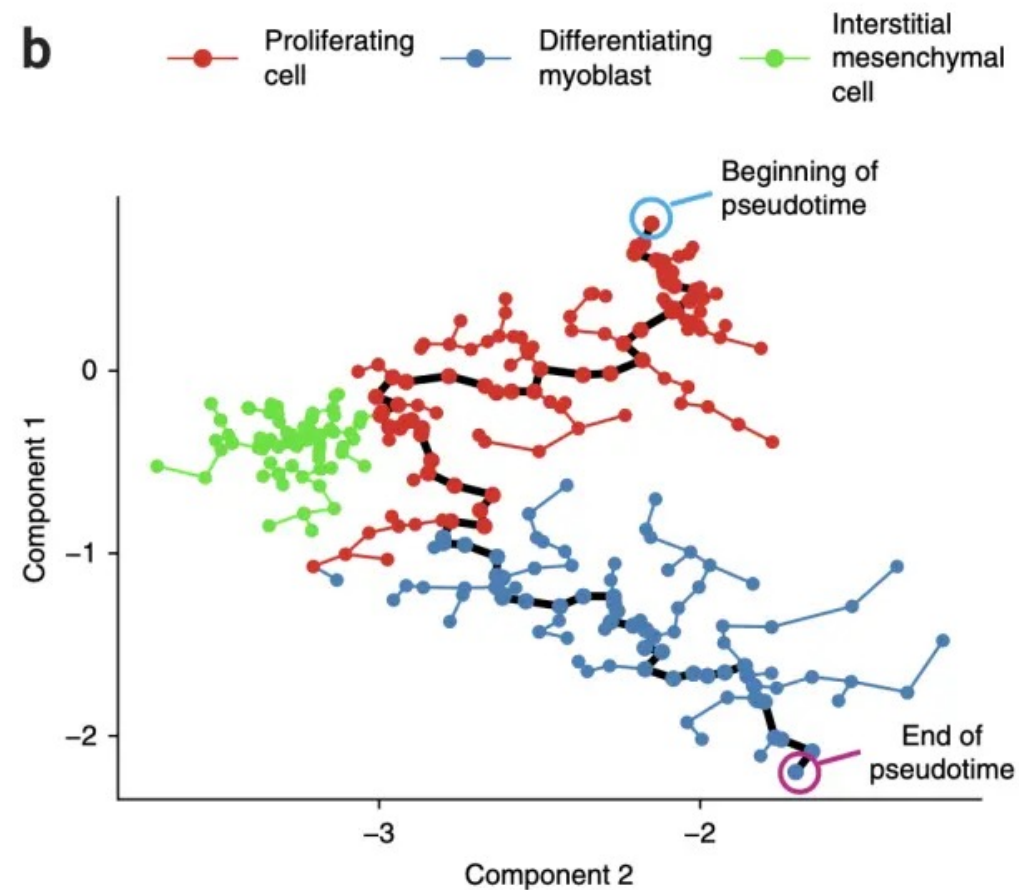
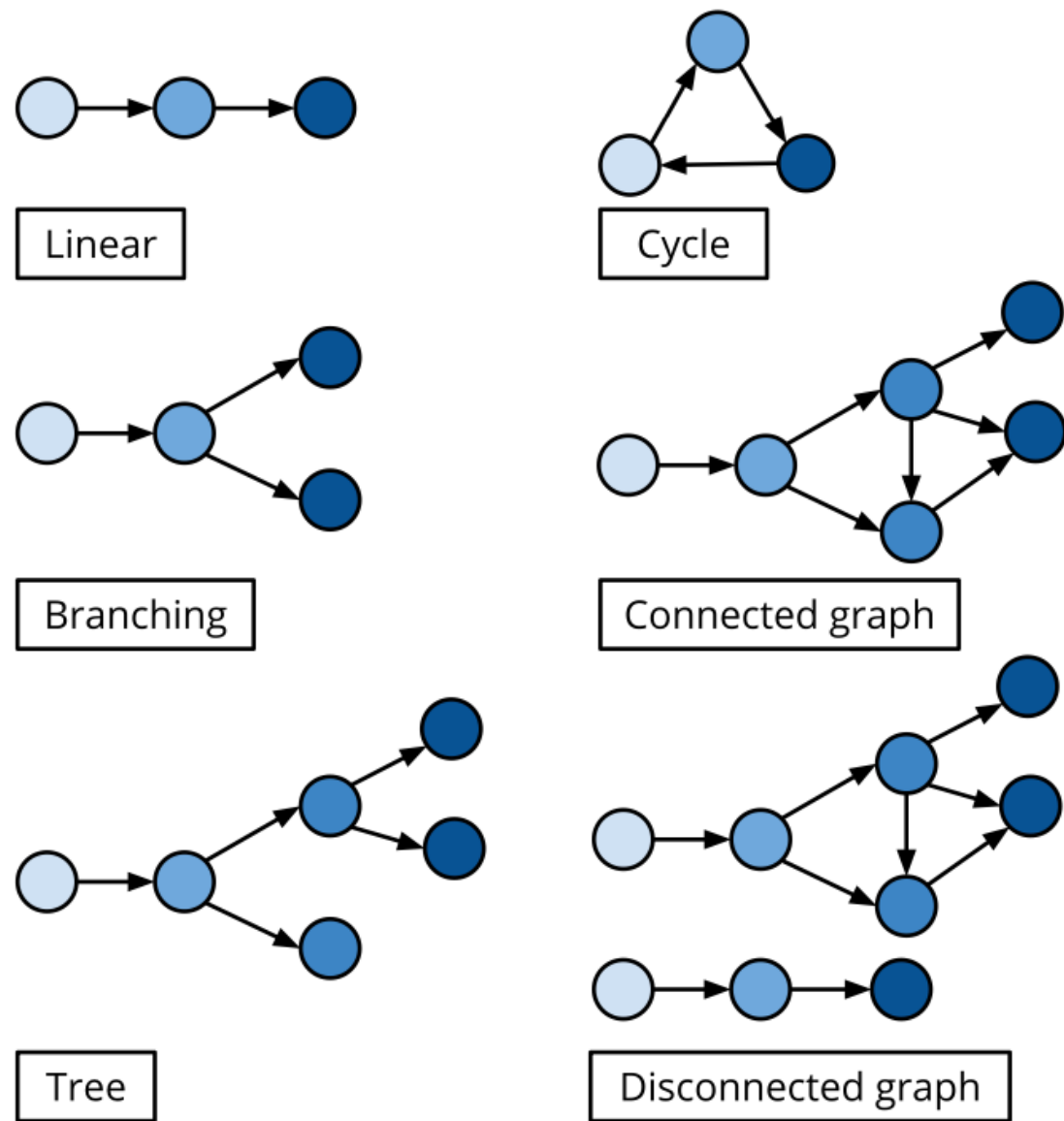
轨迹分析 (拟时序分析) 的基本概念

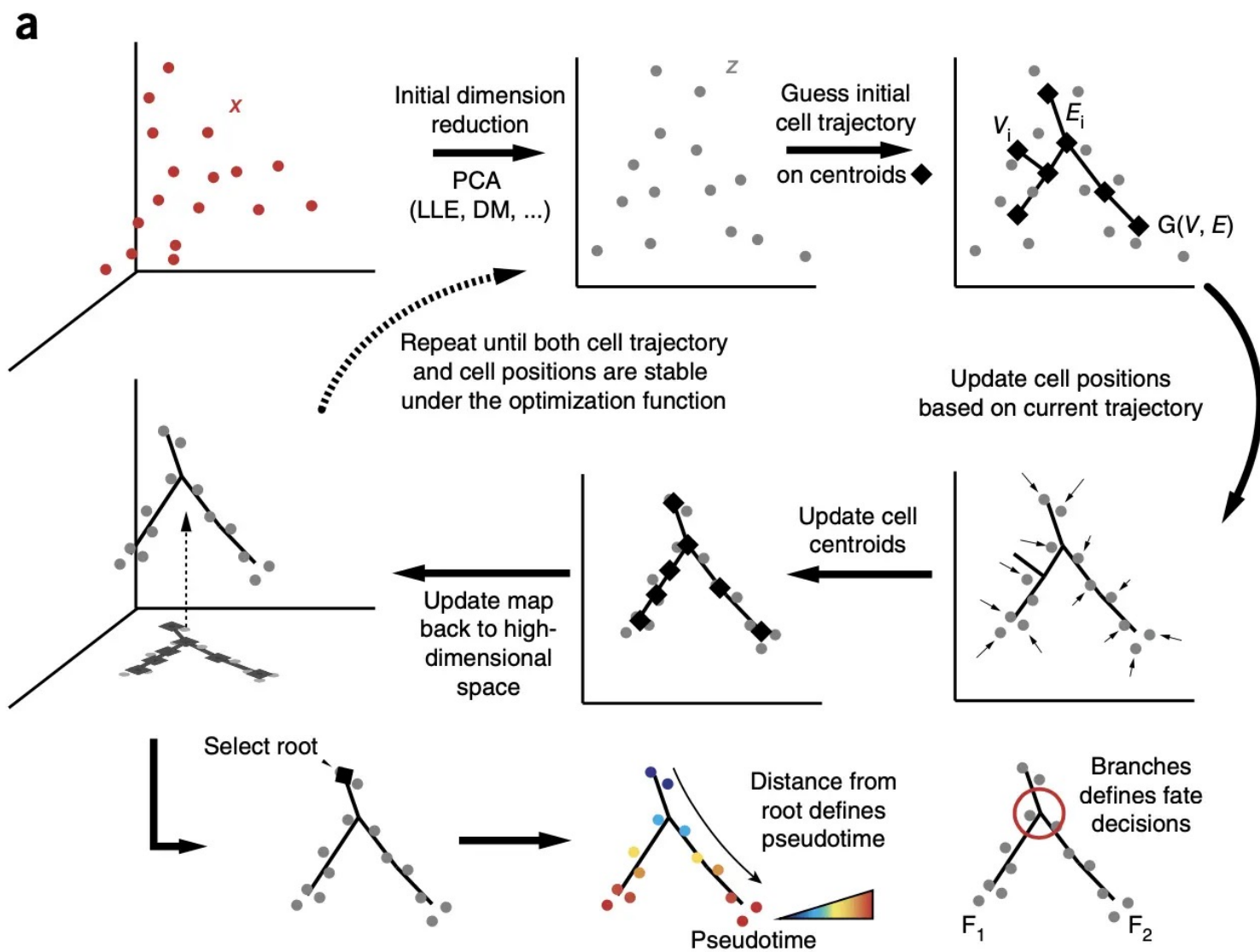
- 1) 数据采样是连续过程下的snapshot
- 2) 离散 -> 聚类 (分类)



连续 -> 轨迹







RNA velocity of single cells (Nature, 2018)

Gioele La Manno^{1,2}, Ruslan Soldatov³, Amit Zeisel^{1,2}, Emelie Braun^{1,2}, Hannah Hochgerner^{1,2}, Viktor Petukhov^{3,4}, Katja Lidschreiber⁵, Maria E. Kastriiti⁶, Peter Lönnerberg^{1,2}, Alessandro Furlan¹, Jean Fan³, Lars E. Borm^{1,2}, Zehua Liu³, David van Bruggen¹, Jimin Guo³, Xiaoling He⁷, Roger Barker⁷, Erik Sundström⁸, Gonçalo Castelo-Branco¹, Patrick Cramer^{5,9}, Igor Adameyko⁶, Sten Linnarsson^{1,2,*} & Peter V. Kharchenko^{3,10,*}

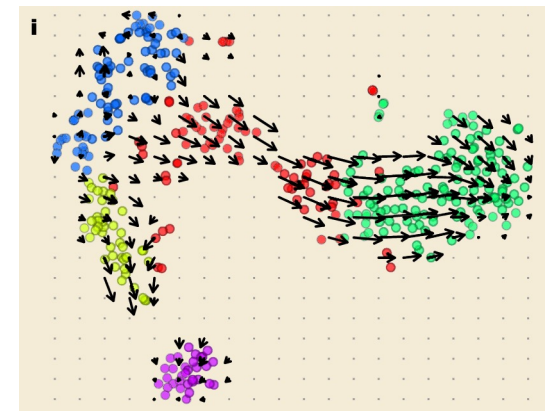
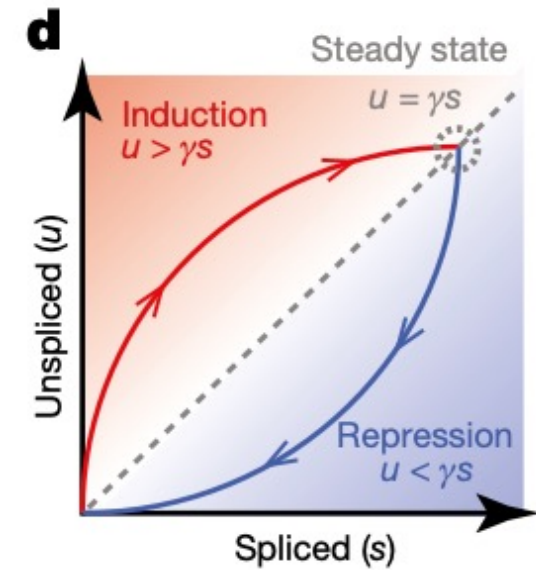
Generalizing RNA velocity to transient cell states through dynamical modeling (NBT, 2020)

Volker Bergen^{1,2}, Marius Lange^{1,2}, Stefan Peidli^{1,2}, F. Alexander Wolf¹ and Fabian J. Theis^{1,2}

Theory

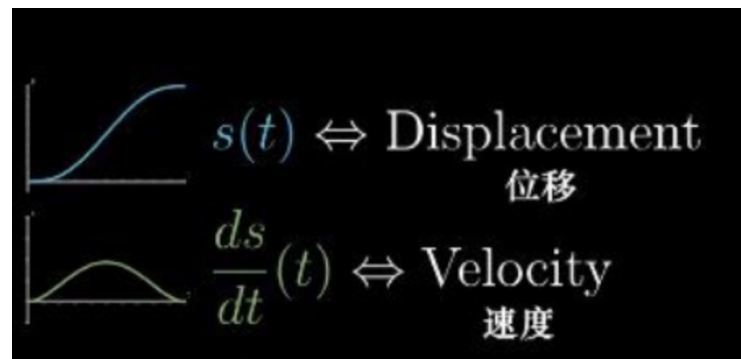
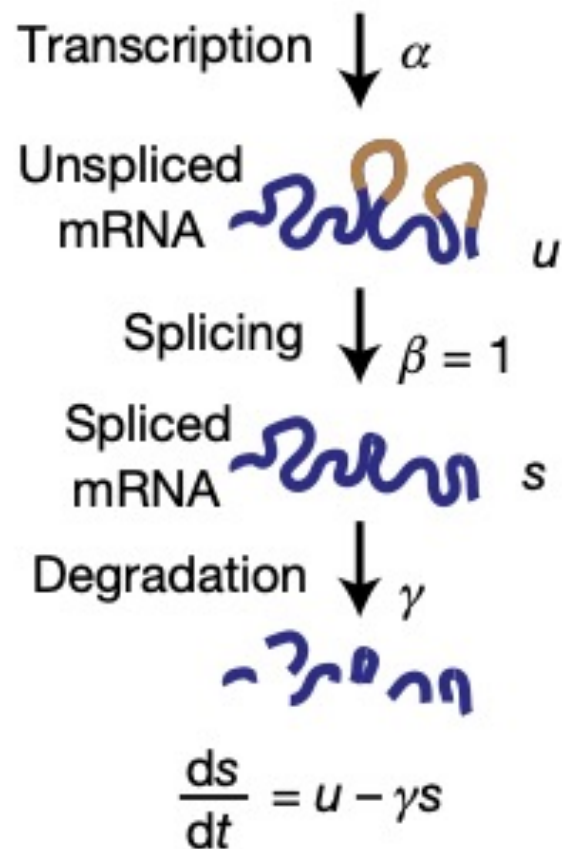
Mapping transcriptomic vector fields of single cells (Cell, 2022)

Xiaojie Qiu,^{1,2,22,*} Yan Zhang,^{3,4,22} Jorge D. Martin-Rufino,^{5,6} Chen Weng,^{1,2,6} Shayan Hosseinzadeh,⁷ Dian Yang,^{1,2} Angela N. Pogson,^{1,2} Marco Y. Hein,⁸ Kyung Hoi (Joseph) Min,^{1,2,9} Li Wang,¹⁰ Emanuelle I. Grody,⁵ Matthew J. Shurtleff,¹¹ Ruoshi Yuan,¹² Song Xu,¹³ Yian Ma,¹⁴ Joseph M. Replogle,^{1,2,15} Eric S. Lander,^{5,16,17} Spyros Darmanis,¹⁸ Ivet Bahar,^{3,4} Vijay G. Sankaran,^{5,6} Jianhua Xing,^{3,4,19,20,*} and Jonathan S. Weissman^{1,2,21,23,*}



常微分方程描述RNA转录过程

b



$$\frac{du}{dt} = \alpha(t) - \beta(t)u(t)$$

$$\frac{ds}{dt} = \beta(t)u(t) - \gamma(t)s(t)$$

$\frac{du}{dt} = \alpha(t) - \beta(t)u(t)$
 Assume $\alpha(t)$ is a constant, and set $\beta(t) = 1$.
 u is a function of t .
 $\Rightarrow \frac{du}{dt} = \alpha - u$
 $\Rightarrow \frac{d(u-\alpha)}{u-\alpha} = -dt$
 $\Rightarrow \ln(u-\alpha) = -t + C'$
 $\Rightarrow u-\alpha = c \cdot e^{-t}$ ($c = e^{C'}$).
 When $t=0$: $u_0 = \alpha + c \Rightarrow c = u_0 - \alpha$
 $\Rightarrow u(t) = u = \alpha + (u_0 - \alpha) \cdot e^{-t}$
 $= \alpha \cdot (1 - e^{-t}) + u_0 \cdot e^{-t}$

nature

[Explore content](#) [About the journal](#) [Publish with us](#)

[nature](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 03 May 2023](#)

Learnable latent embeddings for joint behavioural and neural analysis

[Steffen Schneider](#), [Jin Hwa Lee](#) & [Mackenzie Weygandt Mathis](#) [✉](#)

解析小鼠的大脑活动信号，从中重构出小鼠观看到的电影片段，几乎与电影原作一致。

nature neuroscience

[Explore content](#) [About the journal](#) [Publish with us](#) [Subscribe](#)

[nature](#) > [nature neuroscience](#) > [articles](#) > [article](#)

Article | [Published: 01 May 2023](#)

Semantic reconstruction of continuous language from non-invasive brain recordings

[Jerry Tang](#), [Amanda LeBel](#), [Shailee Jain](#) & [Alexander G. Huth](#) [✉](#)

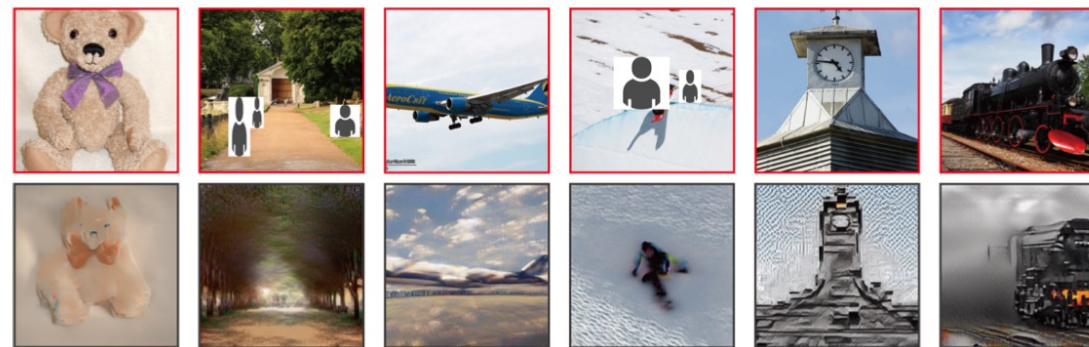
High-resolution image reconstruction with latent diffusion models from human brain activity

Yu Takagi^{1,2*} Shinji Nishimoto^{1,2}

¹Graduate School of Frontier Biosciences, Osaka University, Japan

²CiNet, NICT, Japan

{takagi.yuu.fbs,nishimoto.shinji.fbs}@osaka-u.ac.jp



功能性磁共振（fMRI）数据，（1）将一个人在听故事或默默想象讲述故事时的大脑活动翻译成连续的文本流（2）直接重建人脑所想高清图像

Thank you!