

Spatial reconstruction of single-cell
gene expression data

单细胞基因表达数据的空间重构

Group 3:
胡琪、赵伊、刘倩



Some questions



- ① 什么是单细胞基因表达数据的空间重构?
- ② 为什么要进行单细胞基因表达数据的空间重构?
- ③ 怎么来利用单细胞基因表达数据进行空间重构?

背景

- Spatial localization is a key determinant of cellular fate and behavior. when cell fate is being decided on the basis of inputs from several morphogens whose gradients originate from different regions of the embryo.

空间定位对细胞的重要性不言而喻，正如胚胎干细胞的分化也受到其所处空间位置因素的影响。

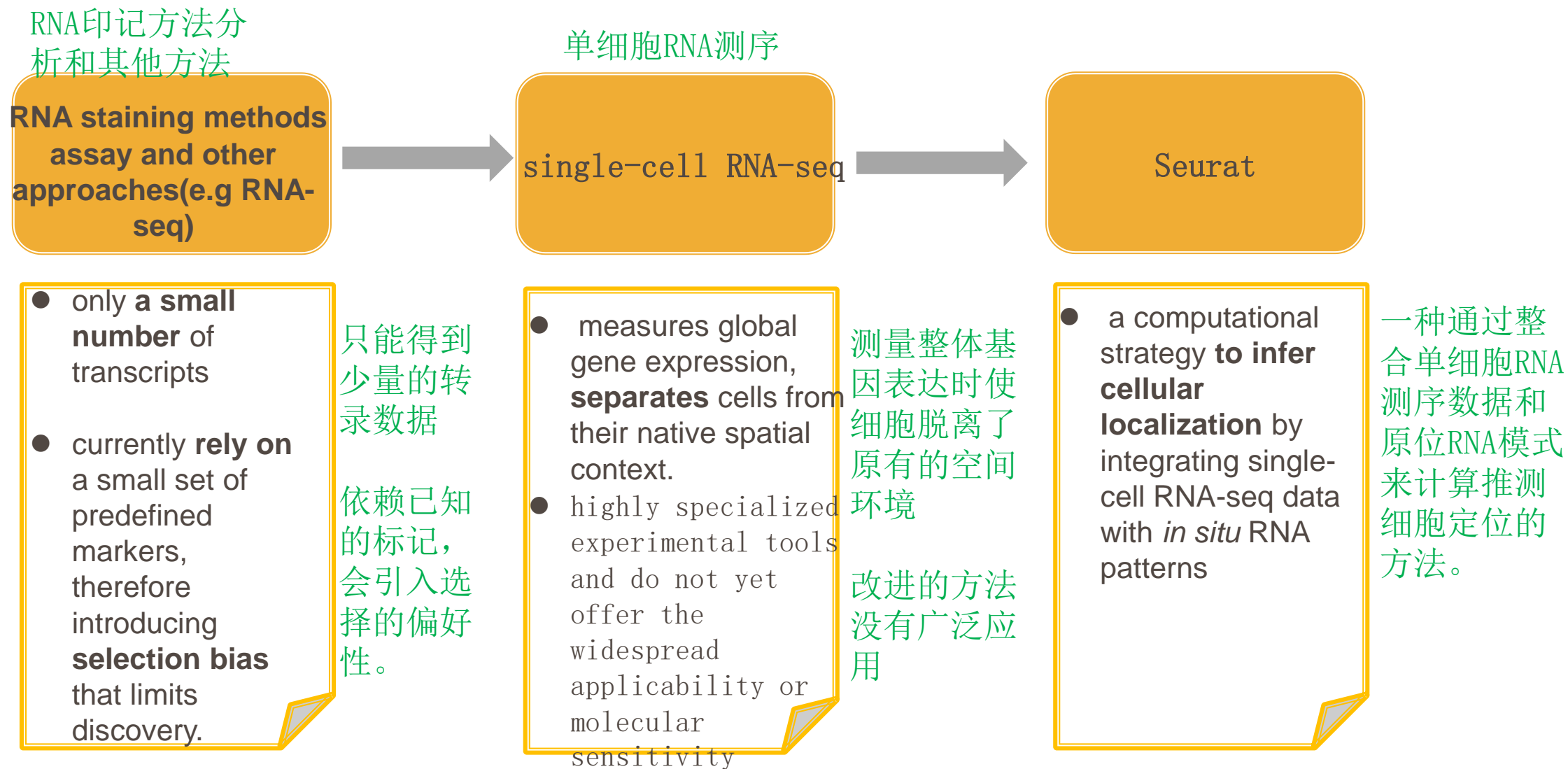
- Spatial heterogeneity in developing organisms has typically been studied by RNA hybridization.

当异质性的细胞群彼此紧邻交织在一起时，基因组学的研究方法不容易应用到三维组织中去，所以这时要换个角度看问题，用RNA的杂交结果来显示细胞空间的异质性。

- Developmental biology: lack crucial information on cells' environments and locations.

发育生物学：缺少细胞所处的环境和位置这方面的重要信息。——和系统生物学的联系

发展历程



Seurat 的技术支持

- **in situ hybridization (原位杂交)**：原位杂交是指将特定标记的已知顺序核酸为探针与细胞或组织切片中核酸进行杂交，再以放射自显影或免疫细胞化学方法对标记探针进行探测，从而在保持组织与细胞的结构完整情况下对特定核酸顺序进行精确定量定位的过程。

RNA in situ hybridization

Fluorescent in situ hybridization. (荧光原位杂交)

Colorogenic in situ hybridization.

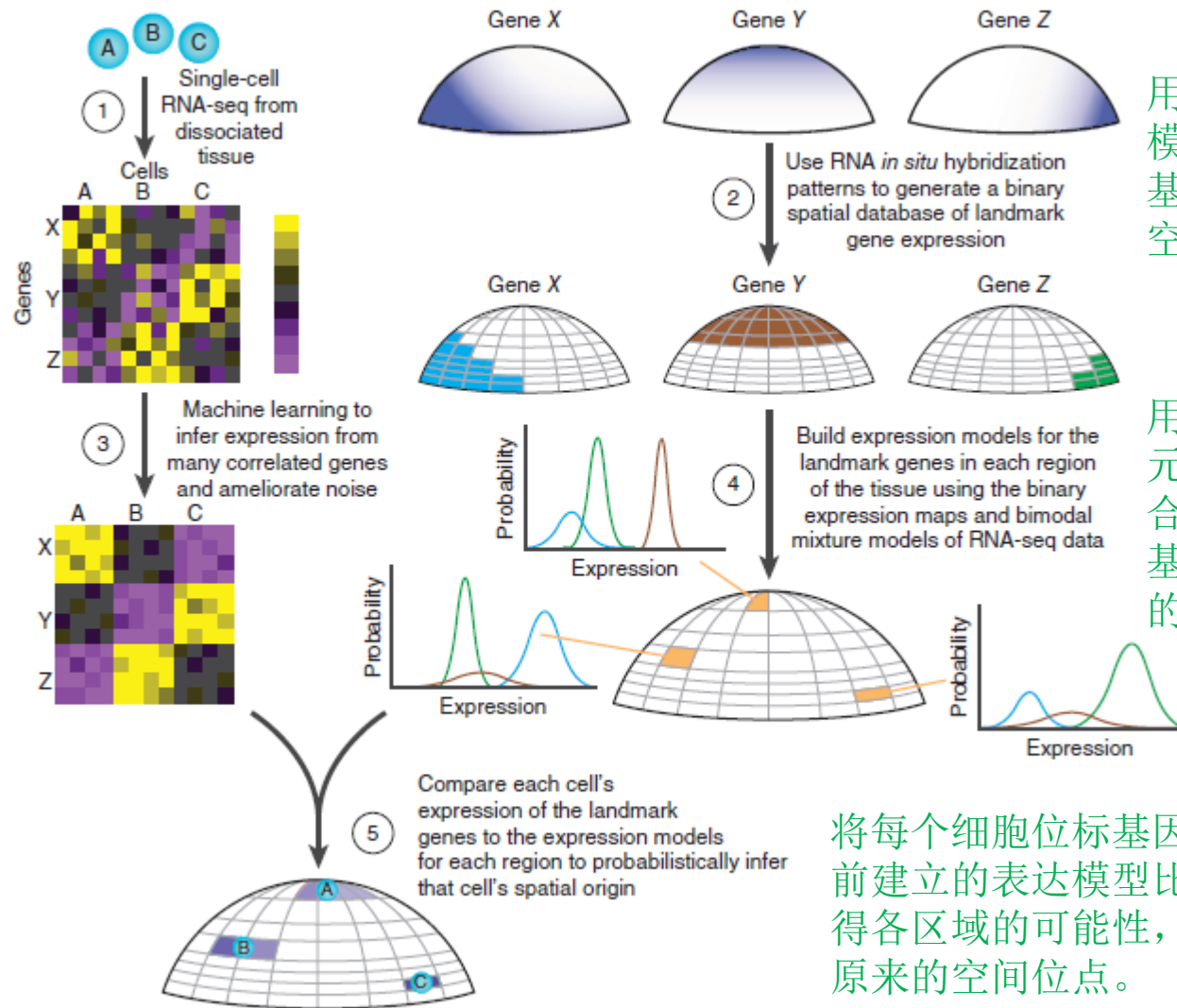
- **RNA-seq (转录组测序技术)**：把mRNA, small RNA, and non-coding RNA等或者其中一些用高通量测序技术把它们的序列测出来。反映出它们的表达水平。
 - (i) Single-cell RNA-seq: reverse transcription. (逆转录)
 - (ii) Single-cell RNA-seq: PCR pre-amplification. (PCR前/预扩增)
 - (iii) Single-cell RNA-seq: library preparation. (文库的制备)
 - (iv) Single-cell RNA-seq: read processing, alignment and gene quantification. (read加工, 比对和基因量化)

- **Principal components analysis (主成分分析又称主分量分析)**：将多个变量通过线性变换以选出较少个数重要变量的一种多元统计分析方法。是模型处理中常用的一种方法，因为它能够有效地提取出模型的主要成分和结构，一定程度上去除噪音和干扰。

Seurat 的数据输入

分离组织的
单细胞RNA测序数据

机器学习来
推断一些关联基因的表达，并减少噪声。



用RNA的原位杂交模式来产生位标基因表达的二元空间数据库

用RNA测序数据的二元表达图谱和双峰混合模型，来建立位标基因在组织各个区域的表达模型。

将每个细胞位标基因的表达和先前建立的表达模型比对，通过所得各区域的可能性，来推断细胞原来的空间位点。

Seurat 的具体流程

建立模型

计算结果

数据搜集

(i) It uses co-expression patterns across cells in the single-cell RNA-seq profiles to impute the expression of each landmark gene in each cell.

(i) 利用细胞在单细胞RNA测序中得到的共表达模式来归纳出每个位标基因在相应细胞的表达。

这样做能减少因为在单细胞RNA测序中的技术限制，而导致的单个细胞中特定转录的观测误差。

将数据间建立联系

(ii) It relates the continuous imputed RNA-seq expression levels of each landmark gene to the binary spatial expression values using a mixture model constrained by the proportion of cells expressing the gene in the reference map.

(ii) 通过细胞在参考图谱中表达基因的比例来限制得到混合模型，并用这个模型来将连续估算的RNA序列每个位标基因的表达水平，和二元的空间表达值相联系起来。

(iii) For each bin, it constructs a multivariate normal model for the joint expression of the landmark genes based on these mixture models, the binary spatial reference map and an optional quantitative refinement step that estimates covariance parameters between all pairs of

(iii) 对于每一个接收器，构建一个关于位标基因共表达的多元正态模型，基于这些混合模型（评估每对基因协方差参数的二元空间参考图谱和一个可选定量的提炼步骤）

(iv) Given these models, it infers the spatial origin of each profiled cell by calculating a posterior probability for each cell-bin pair, allowing determination of the cell's likely position(s) and confidence

(iv) 给定这些模型，通过计算每对细胞接收器的后验概率来推测出每个细胞的原始空间。给出细胞的可能位置和在图谱中的置信度。

验证

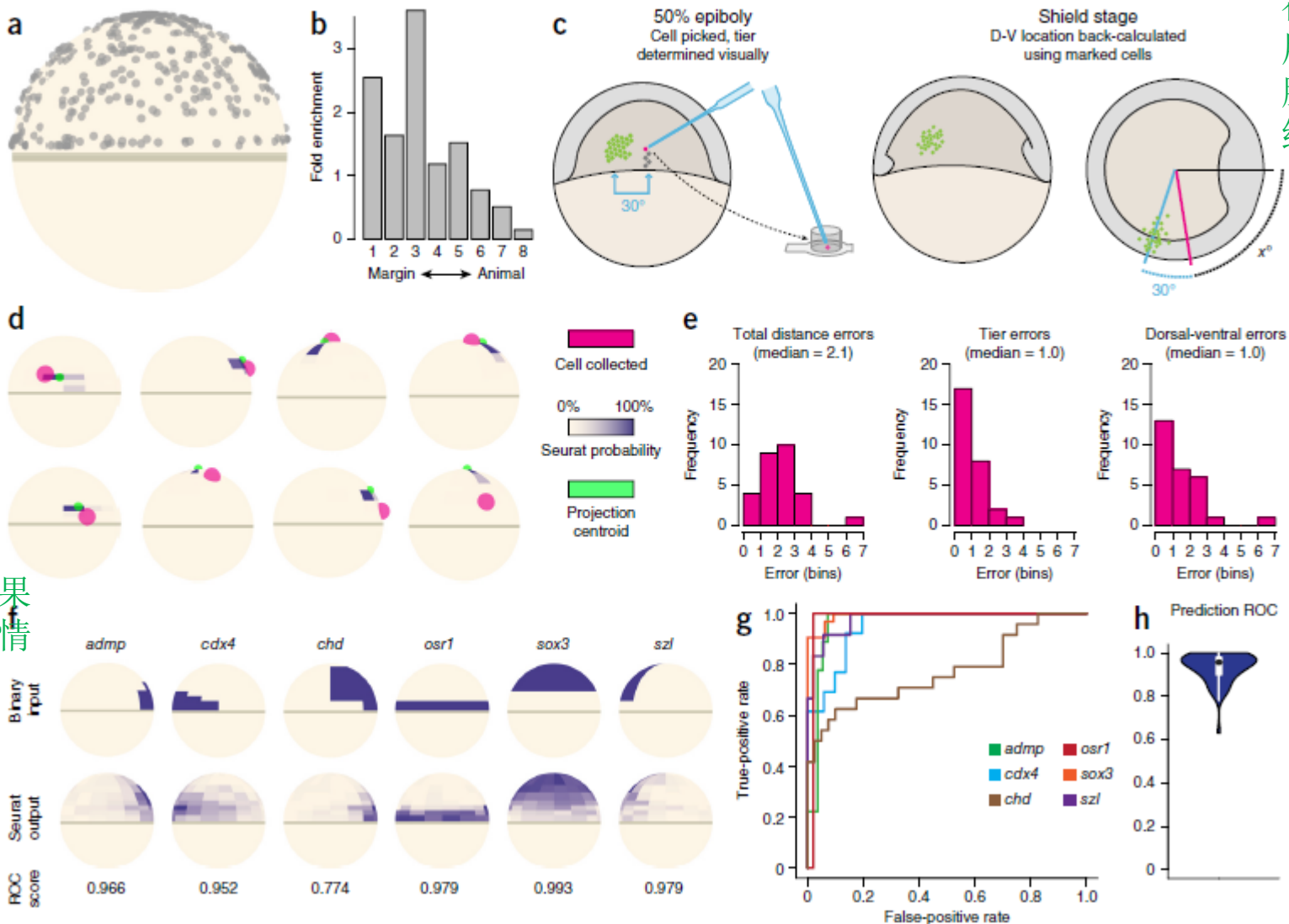
胚胎符合
组织的随
机解离

一小部分参考细胞
是手动在解剖显微
镜下取的，它们的
位置已知。

材料: late
blastula stage
(50% epiboly)
zebrafish
embryos

斑马鱼位于
囊胚后期
(50%包覆)
的胚胎

预测结果
与真实情况



在“盾阶段”
用标记的细胞回推背-腹
细胞的位置。

分析

之后还进行了一些扩大样本集，减小训练集的方法来测试Seurat的准确度，结果都还不错，大家有兴趣可以下去看一下，这里就不做详细介绍了。Seurat有以下三个方面的应用：

- Seurat correctly localizes rare cell populations（正确定位稀有细胞群）
- Seurat discovers markers of rare subpopulations（发现稀有亚细胞群的标记）
- Seurat identifies dispersed, rare cell populations（辨别出分散的，稀有的细胞群）

Seurat解决了以下技术挑战：

- the representation of in situ hybridizations for algorithmic input（对于原位杂交算法输入的表示）
- handling stochastic noise in RNA-seq data for landmark genes and finding a correspondence between the two data types（解决了RNA测序数据中位标基因的随机噪音并找到两种数据类型之间的关系）

研究趋势

Seurat重点研究推断空间起源，结合转录组学数据与RNA-FISH，CyTOF质谱仪或流式细胞仪数据可以帮助确定细胞的发育状态或疾病表型，并将其与大量已发表的研究结合，**Seurat**与这些潜在的方法共享挑战和目标。

Seurat package (Seurat包, R中的一个开源软件包, an R toolkit for single cell genomics) .

<https://github.com/satijalab/seurat>

THANKS