



Single-cell genome sequencing: current state of the science

单细胞基因组测序：现状及趋势

生信1301

第一组

2016.3.8




小组成员：

陈天哲：论文翻译 ppt制作

于志超：论文翻译 课堂讲解

范炳亮：论文翻译 ppt制作





单细胞测序技术是指在单个水平上对基因组进行扩增与测序的一项新技术。

单细胞测序主要涉及单细胞基因组测序和转录组测序两方面，分别针对单细胞的DNA和RNA进行序列分析和比较，进而揭示基因组和转录组的变化

单细胞全基因组测序是对选定的目的细胞的全部基因组序列进行非选择性、均匀扩增，随后利用外显子捕获技术进行高通量测序。

单细胞转录组测序是利用高通量测序技术进行cDNA测序，从而获取特定器官或组织在某一状态下的几乎全部转录本



一.技术挑战

☆高效的进行单细胞分离

☆单细胞基因组扩增

☆测定基因组时控制成本来测试研究中假说

☆误差控制与分析

1.单细胞分离

首先利用机械或酶的手段从原始的样本产生一份单细胞悬浮液，在悬浮液中，已经开发了几种方法来隔离单细胞。

手工操作的方法，如连续稀释，微量吸管，微孔稀释，**光钳**

分离完整细胞或细胞核的荧光激活细胞分选
([FACS](#))

自动化操作方法，利用液滴或微阀的微流控装置

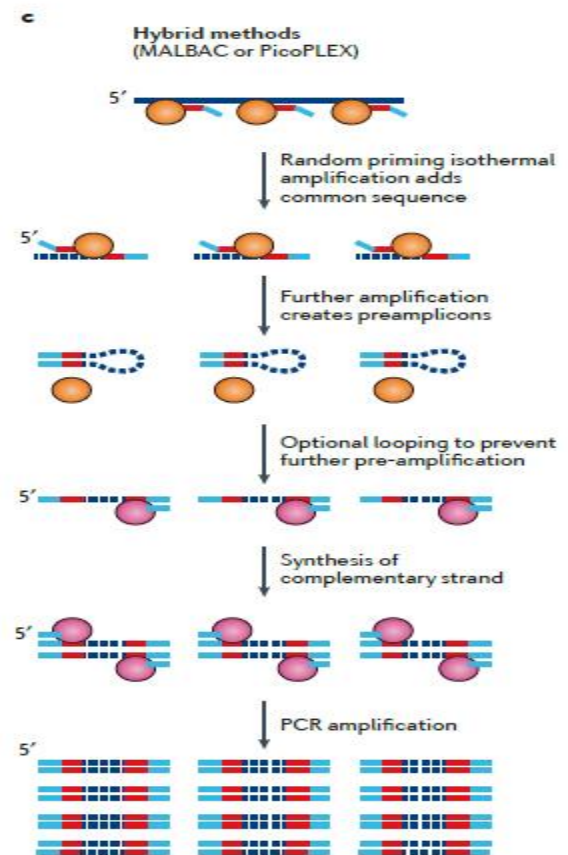
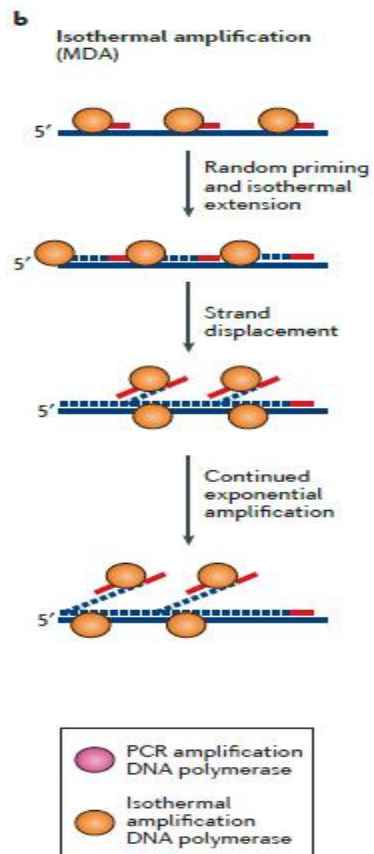
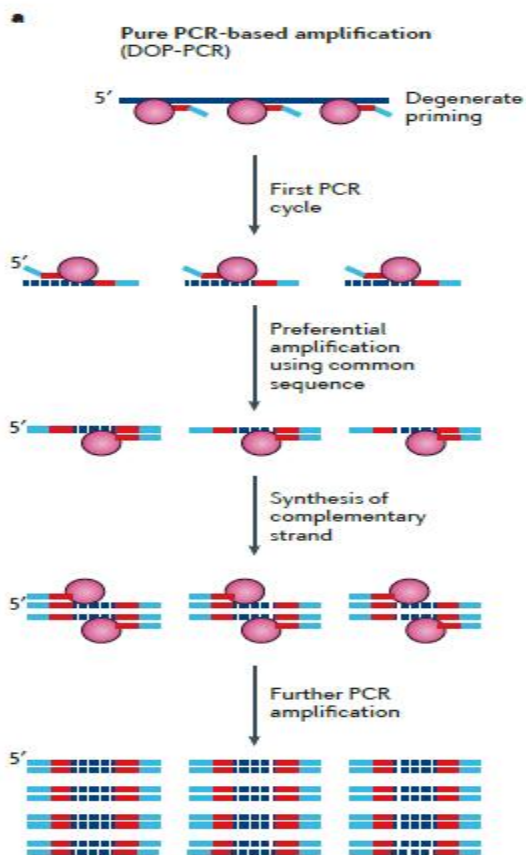
确定一个单一的细胞已被物理隔离

Optical tweezers

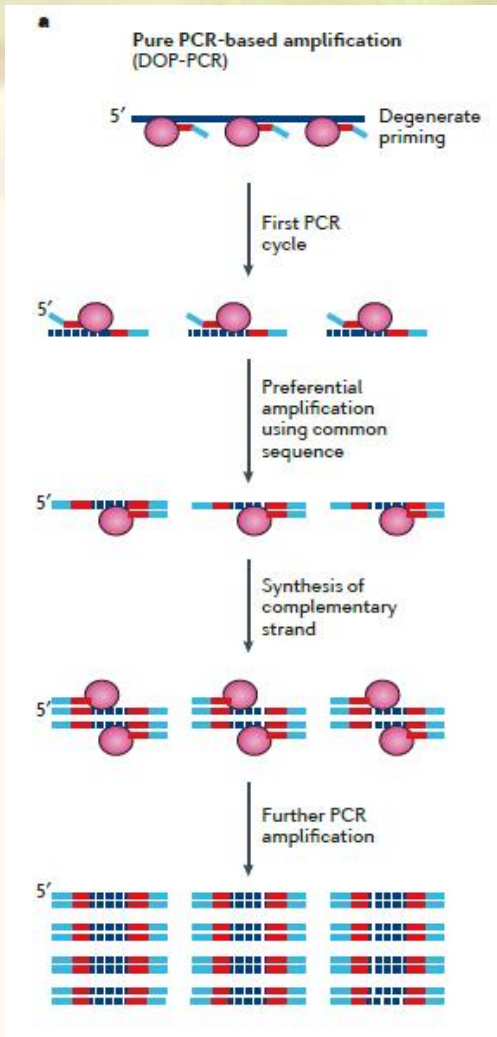
Devices that use a laser to manipulate submicron particles, such as bacterial cells or cellular macromolecules



2.基因组扩增



DOP-PCR



Pure PCR methods such as degenerate oligonucleotide primed PCR (DOP-PCR) use random priming followed by PCR amplification, which preferentially amplifies specific sites in the genome. This results in low physical coverage of the genome, but better uniformity of amplification

退火，启动

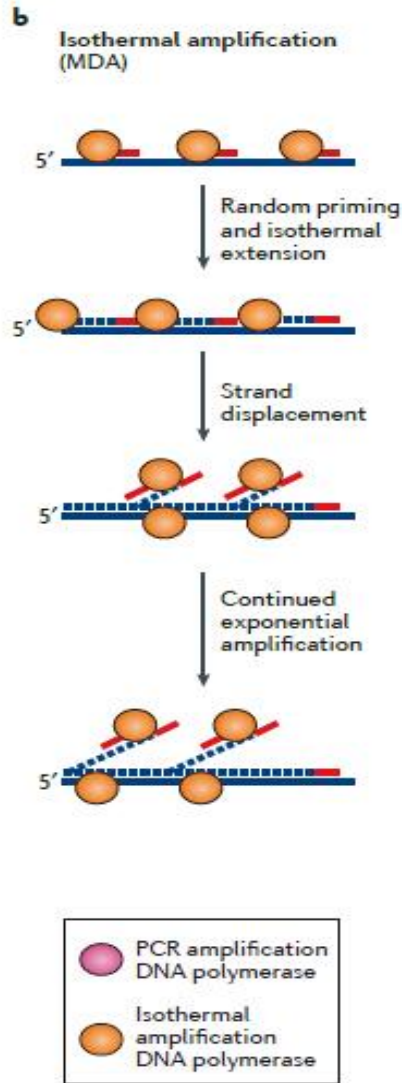
—>第一次PCR反应

—>优先放大有共同序列的基因片段

—>合成互补链

—>扩增

MDA



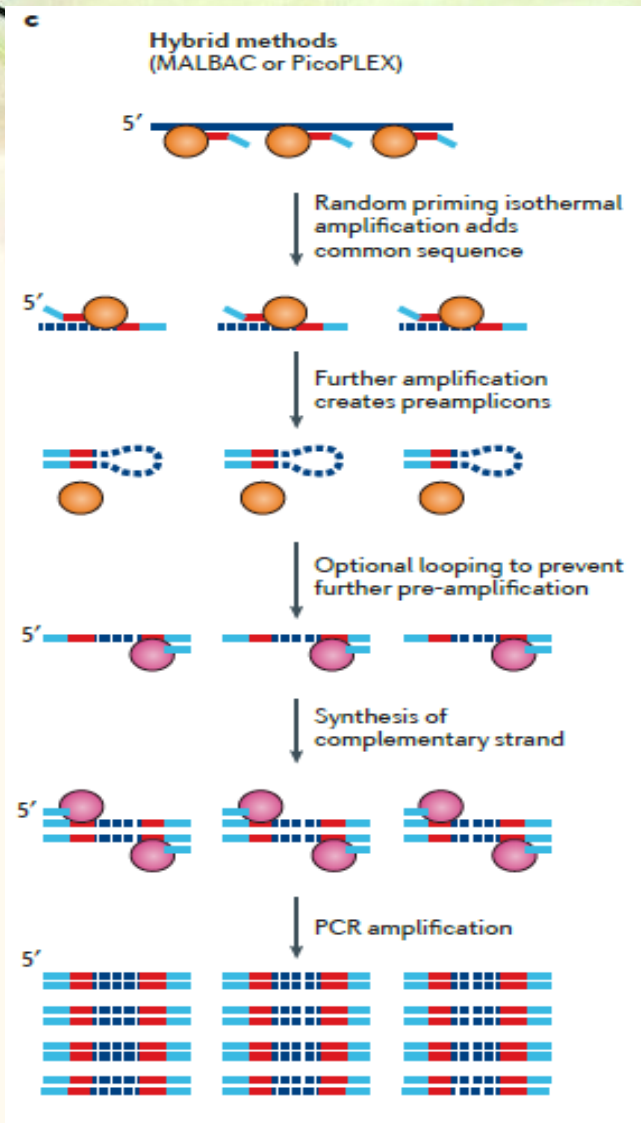
Isothermal methods such as multiple displacement amplification (MDA) use random priming followed by isothermal exponential amplification using a polymerase with high processivity and strand displacement activity. These methods can cover most of the genome, but have much less uniformity.

随机启动等温延伸

—>链位移

—>指数扩增

MALBAC or PicoPLEX



Hybrid methods such as multiple annealing and looping based amplification cycles (MALBAC) and PicoPLEX have an initial isothermal preamplification, in which common sequences are added, followed by PCR amplification using those sequences. These methods have intermediate coverage and uniformity when compared with pure PCR and isothermal methods.

随机启动等温扩增增加共有序列
—>进一步放大，创建扩增单元
—>可选的循环，防止进一步预放大
—>合成互补链
—>PCR扩增

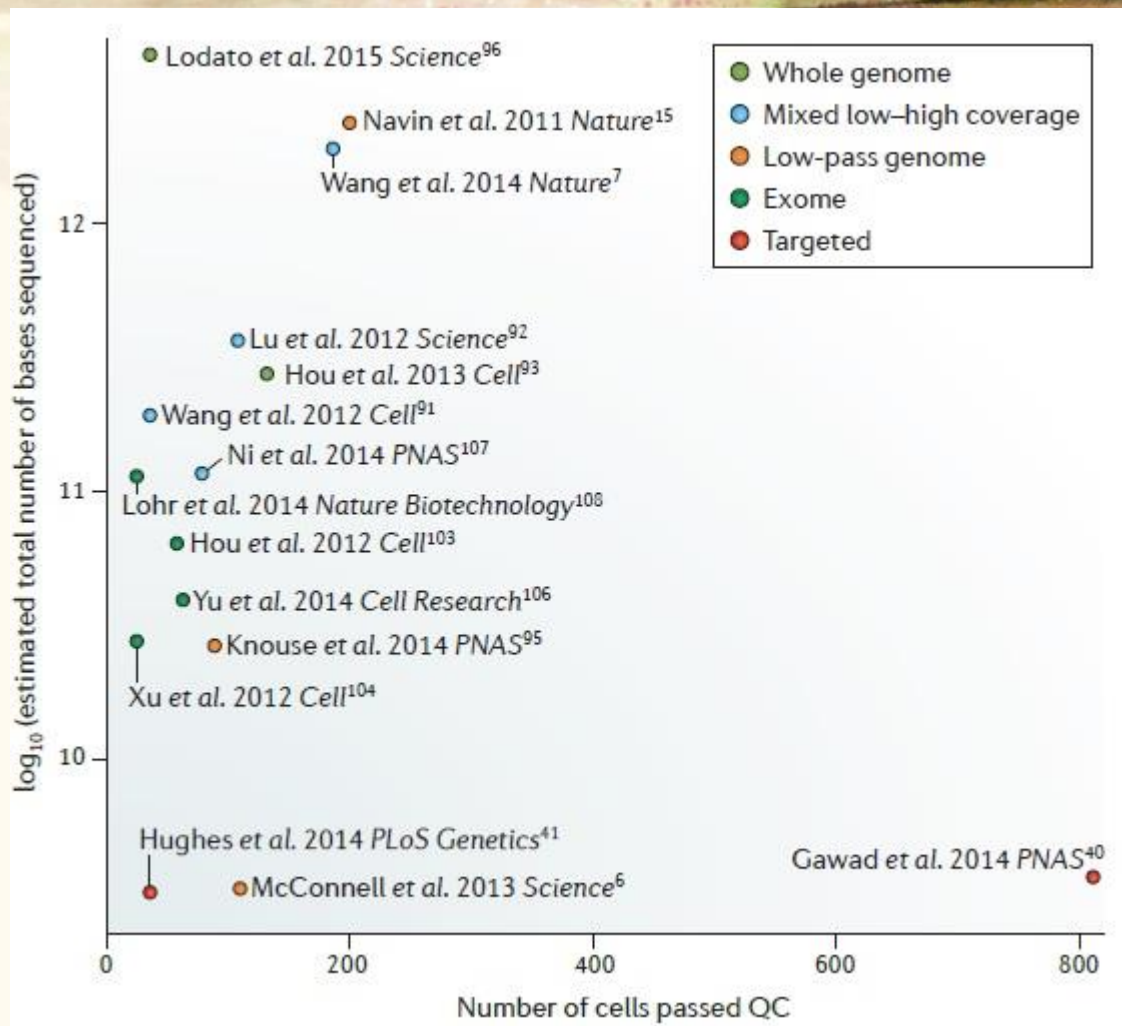
三种方法的比较

	PCR-based (DOP-PCR)	Isothermal(MDA)	Hybrid(MALBAC or PicoPLEX)
False-negative rate (coverage and allelic dropout)	High	Low	Intermediate
Non-uniformity	Low	High	Low
False-positive rate (amplification error rate)	High	Low	Intermediate

二.成本效益控制

对于复杂的真核基因组像人的，可以选择测定特殊的位点（代表性序列<1MB），蛋白转录区域（外显子，30-60MB）或者全基因组（3GB）把单细胞基因组特定位点作为靶点能帮助聚焦区域，在系统研究中有巨大的生物学贡献，可以减少测序的花费和错误突变的发现。小的靶点区域可能包含较少的错误，像通过第一次小轮的WGA（全基因组扩增）可能会导致基因突变的错误识别（被称为假阳性突变）。使用大规模样品单细胞的全基因组也可以被检测。这个也是要权衡是否增加错误突变发现和增加基因组检测范围的花费。

成本效益控制



单细胞实验设计成本考虑

三.误差控制与分析

*单细胞测序错误概述

分析单细胞基因组测序的一个主要挑战是发展工具，区分工艺和噪声而推出的单细胞分离，WGA和正确的生物突变基因组。

WGA有许多错误，包括范围损失，覆盖范围均一性差，等位基因不平衡，ADO和基因组扩增错误一个要特别留意的是确定的细胞系和细胞类型可能不是双倍体；他们可能是高非整倍体或者多倍体，而去对于实验影响巨大。

***单细胞变体识别。**对于假阳性变体。首先，大样本可以被作为借鉴来减少错误发现率。第二，当只使用单细胞数据，两个或三个细胞能够在一样的位点有一样的变异，在人类3GB基因组单细胞WGA引入几千个突变，不可能有重现的机会。

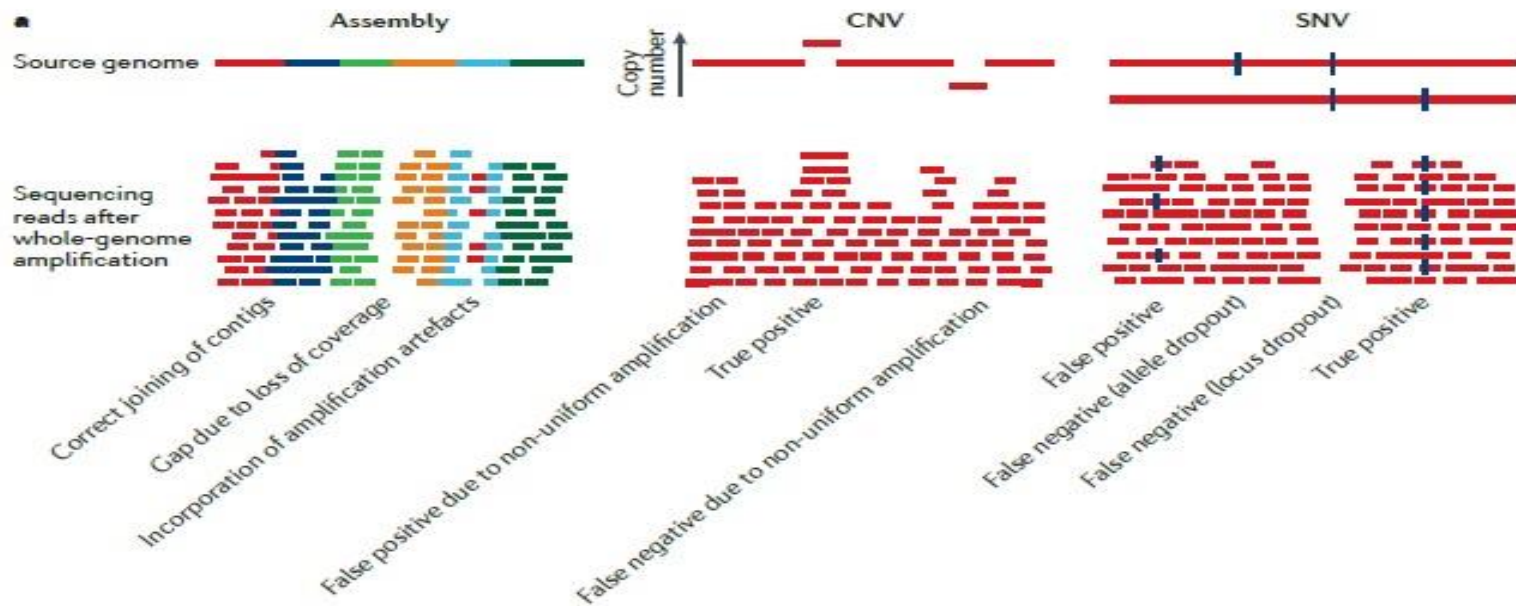
为克服等位基因不平衡。一个策略是需要所有的变体识别在对照试样中技术噪声等级之外，另一个途径是降低测序错误率，通过使用分子识别条码。



误差控制与分析

确定单细胞间亲缘关系。总体战略对于基因表达聚类和其他大数据集是依赖于距离函数提供一个定量测定一堆样本之间不同。在单细胞测序环境中，我们要求这些距离函数有鲁棒性对于假阳性变体测定的结果的缺失数据。我们要找到杰卡德距离是最适合于基因型数据，是自然界的二进制。然而，我们也观察到通常假阳性率影响一个样本中克隆数目统计测定。

误差控制与分析



b Magnitude of deleterious effects of specific genome amplification errors on single-cell applications

	Assembly	CNV detection	SNV detection
False negative rate (coverage and allelic dropout)	Large	Small	Large
Non-uniformity	Intermediate	Large	Intermediate
False positive rate (sequencing and amplification errors)	Intermediate	Small	Large

Figure 3 | Effects of various error types on specific single-cell sequencing applications. a

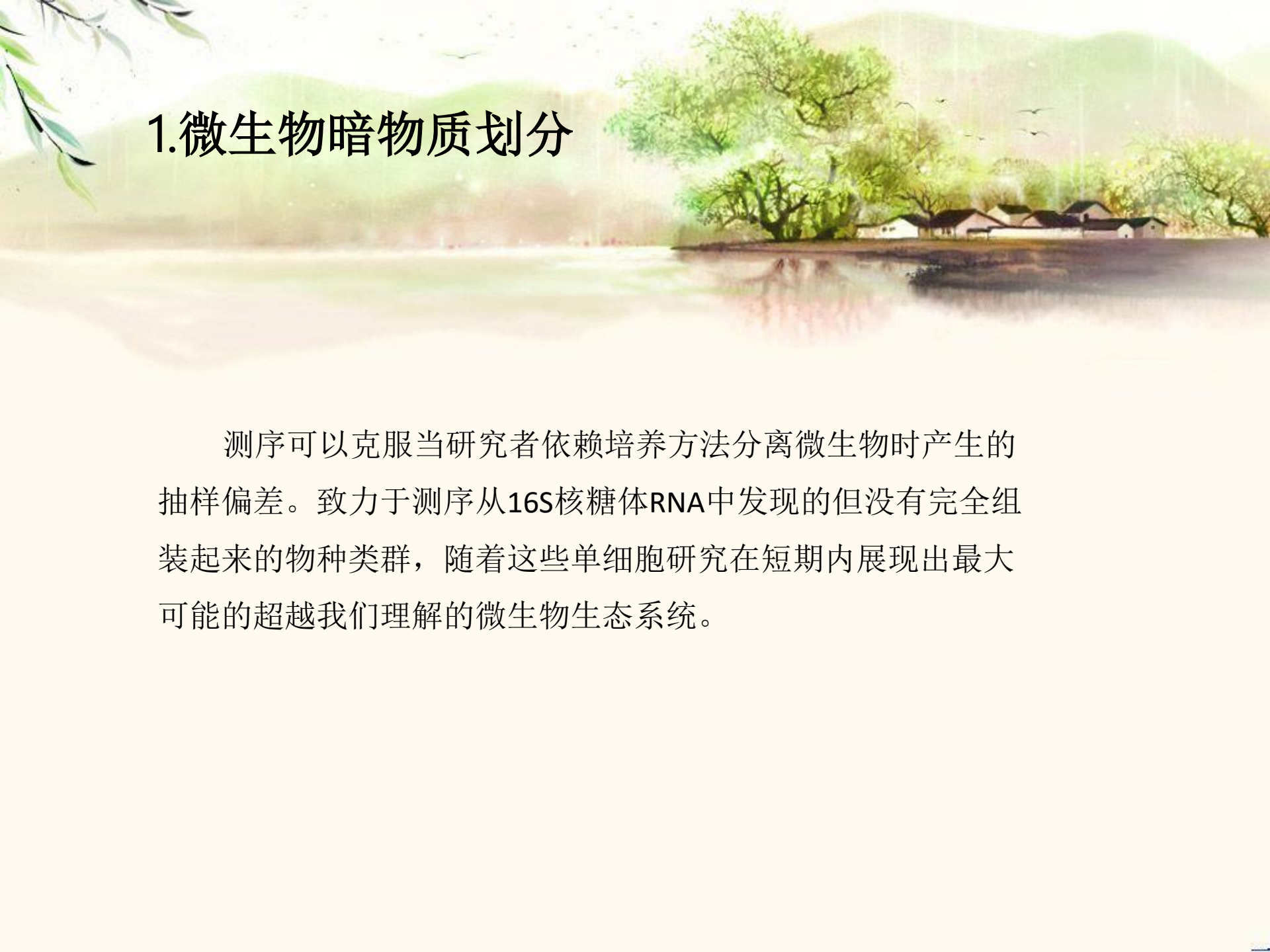


应用

1.微生物暗物质划分

2.多细胞生物遗传镶嵌性识别

3.癌症

A scenic landscape painting of a village by a lake. In the foreground, a calm body of water reflects the sky and the buildings. The middle ground features a cluster of traditional houses with dark roofs and white walls, surrounded by lush green trees. In the background, rolling green mountains rise under a soft, hazy sky. A few birds are visible in flight. The overall style is soft and painterly, with a warm, natural color palette.

1.微生物暗物质划分

测序可以克服当研究者依赖培养方法分离微生物时产生的抽样偏差。致力于测序从**16S核糖体RNA**中发现的但没有完全组装起来的物种类群，随着这些单细胞研究在短期内展现出最大可能的超越我们理解的微生物生态系统。



2.多细胞生物遗传镶嵌性识别

五十年代细胞遗传学发现在同一个体的细胞中可以有不同数量的染色体。最近，依据微矩阵和下一代测序发展来的突变体侦测方法已经可以识别许多由于SNVs或CNVs镶嵌导致的新疾病。

从一个人的不同位点抽取组织研究显示镶嵌CNV和SNV率比预期要高。然而，低水平遗传变异在易变位点和人类疾病的发病机理还有许多未去探索。

目前研究看是使用单细胞测序特化镶嵌遗传变异的人类样本。

3.癌症

最好的研究例如遗传镶嵌现象就是癌症。

大量持续进行癌症序列测序的目的是分类这些突变体来更好的理解肿瘤生物学。

单细胞测序的研究目前已经在单细胞分辨率下开始分析内部肿瘤遗传杂合率。

循环肿瘤细胞可以被孤立且作为一个遗传学潜在窗口通过非入侵样本创造一个肿瘤。这种独特的技术挑战了孤立和分析单个CTCs基因组的关联性。而且，这项研究对确认和特化CTCs和疾病监管战略有成功的希望。

一个至今还未解决的基本问题是是否CTCs将提供典型抽样来代表肿瘤来源内部的基因多样性。



总结

在这篇文章中，他们做了关于目前单细胞测序领域前沿的一个综述。在最近几年里获取单细胞数据的能力取得了实质性进展，发现了许多不可能被标准基因组学发现的新的生物学现象。但是，还存在许多挑战。随着细胞分离技术通量的增加和基因组扩增技术的进步，假设逐一被验证，测序和计算方法发展，无疑地将更多研究者吸引到这一领域。此外，在同一细胞中单细胞基因组测序已经与RNA或蛋白质测量所耦合。基因型和细胞模块关联的能力和表型测量将使更多生物问题的解决变得可能。最后，合并细胞内和细胞内空间基因组测量的信息将使研究者可以把细胞模块聚合来给周围细胞提供环境。许多障碍依旧存在，但是我们相信单细胞基因组这一领域将迅速发展超越我们对微生物生态学，进化和人类疾病的理解。



谢谢！

FACS

The process begins by placing the cells into a flask and forcing the cells to enter a small nozzle one at a time (figure 1). The cells travel down the nozzle which is vibrated at an optimal frequency to produce drops at fixed distance from the nozzle. As the cells flow down the stream of liquid, they are scanned by a laser (blue light in figure 1). Some of the laser light is scattered (red cone emanating from the red cell) by the cells and this is used to count the cells. This scattered light can also be used to measure the size of the cells.

If you wanted to separate a subpopulation of cells, you could do so by tagging those of interest with an antibody linked to a fluorescent dye. The antibody is bound to a protein that is uniquely expressed in the cells you want to separate. The laser light excites the dye which emits a color of light that is detected by the photomultiplier tube, or light detector. By collecting the information from the light (scatter and fluorescence) a computer can determine which cells are to be separated and collected.

来源: <http://www.bb100.com/experiment/18-2511-1.html> 生物秀

