Unit 1.1 资料总结

# What is Genomics?

## DNA

### 定义

* + - * + Is a long molecule contained within the nucleus of all of ourselves
        + It is made up of two strands that wrap around each other to form a double elig shape双螺旋结构

strands

Connected by molecules called nucleotides核苷酸连接

### Sequence

* + - * + Determines our unique genetic code
        + Unit

Nucleotides（核苷酸）

Deoxyribonucleic Acid脱氧核糖核酸

Adenine

Thymine

Cytosine

Guanine

核糖核酸（RNA）

A

腺嘌呤

U

尿嘧啶

C

腺嘧啶

G

鸟嘌呤

### Mutation(突变)

* + - * + Function

Missing letter

Sequence is incorrect

* + - * + Result

health disorders

Sickle cell Anemia镰状细胞性贫血

Cystic Fibrosis囊性纤维化

Cancer

### How does DNA make us who we are?

## GENES

### Definition

* + - * + An instruction for making protein molecules

It is the code for proteins

### Humans have about 20000 to 23000 genes

### Organising chromosomes

## CHROMOSOMES染色体

### Spaghetti-like structures incoming pairs(像意大利面条一样一对一对的)

### Humans have 23 pairs of chromosomes

* + - * + One side from the mother
        + One side from the father

### It is made up of a very long strand of dna

* + - * + Containing 500 - 4000 genes

## Genome基因组

### What is Genome?

* + - * + A genome is a whole set of an organism's chromosomes and therefore genes

### To understand the Cancerous Mutation

### The genomics is the study of the genomes

## 为了把DNA分离要提高温度还是减少温度

# What am I working on now?

## 原因

### 在不该分裂的时候分裂并扩散到周围的脑袋中

## Result

### Work on brain cancer genomics

## 分析DNA的方法 DNA-based technologies

### 1. In sit approach 原位方法

* + - * + 分析单个细胞内的DNA

### 2. 从细胞提取DNA Extract DNA from the cells

## What is the process for the genomic intervention in a brain cancer patient 通过基因手段对脑癌患者进行治疗干预的过程是什么

### We can take the tumour and extract the DNA

### Get other pieces of information

* + - * + 1. how long they are likely to survive

# 我们通过这门课能学到的东西

## 脑机接口

## Game-changing new test that diagnoses cause of deadly sepsis in HOURS not days to be rolled out across NHS

# DNA Sequencing

## A nano technology perspective

## 2.27有考试

## Composed of 4 units called nucleotides

### ATCG

### A binds T

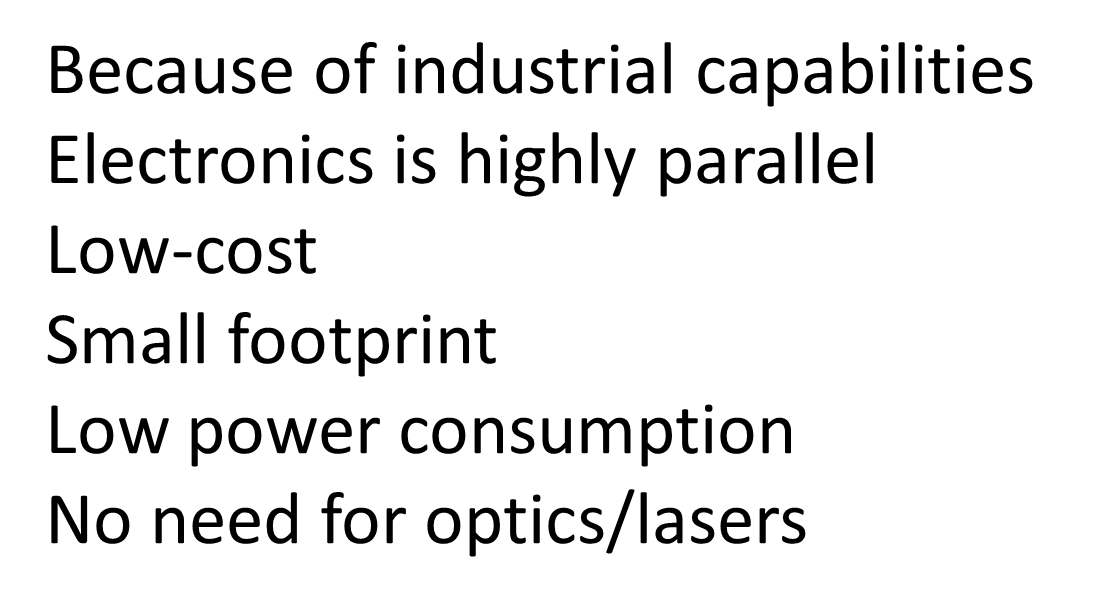
### C binds G

## Molecule that carries the information for the growth development, functioning and reproduction of all known living organisms

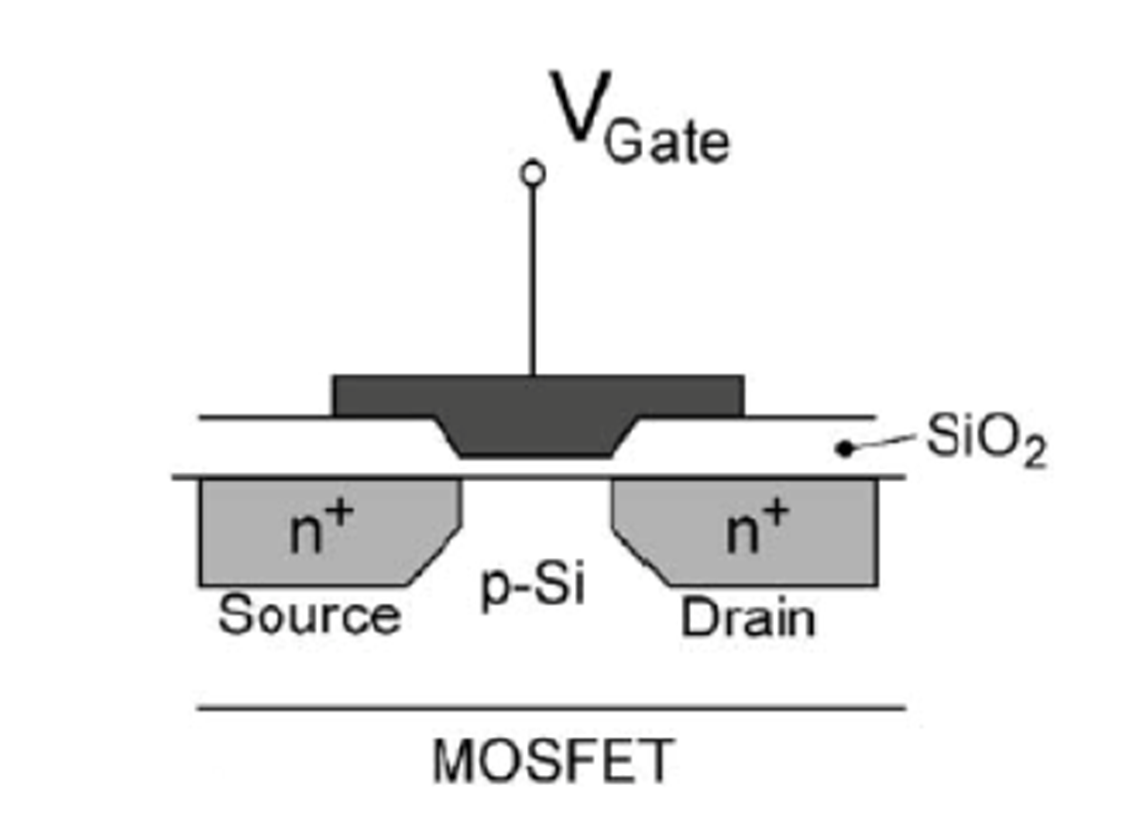
## DNA is double helix

## Can we use the electronic to measure the sequence of the DNA?

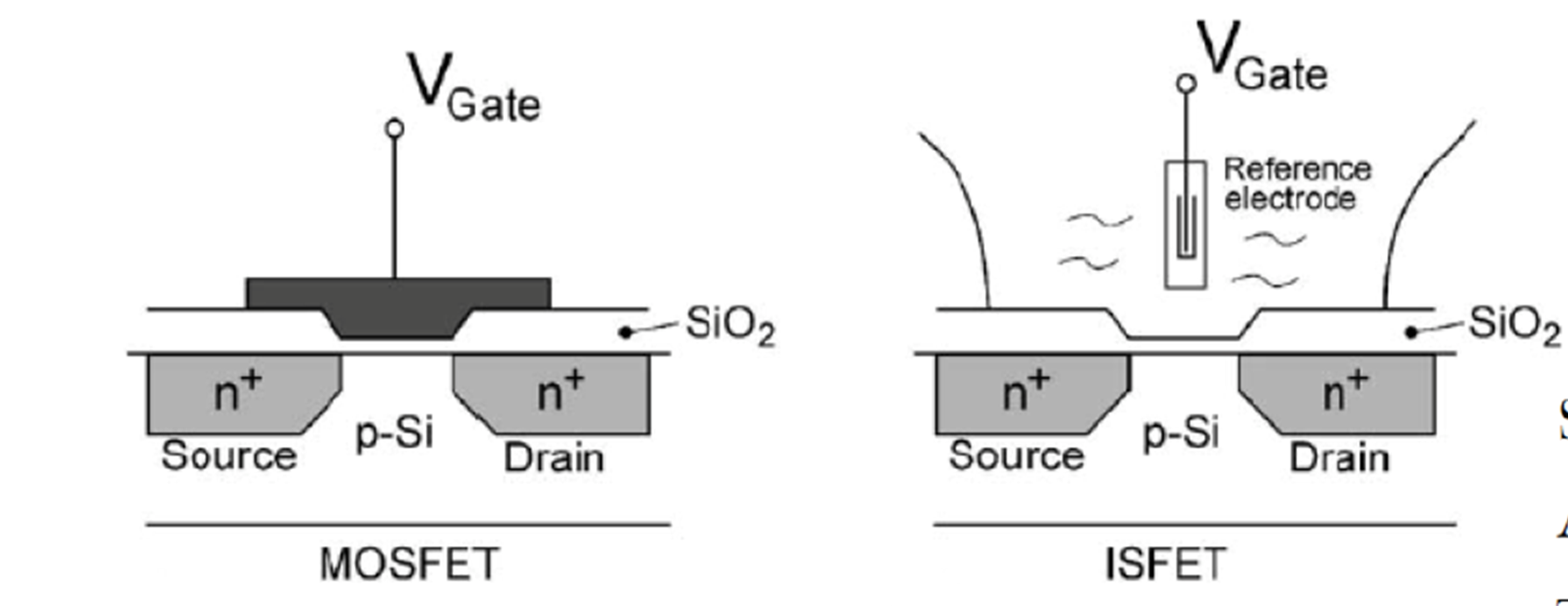
### DNA sequence needs water while the electronic hates water



### MOSFET



* + - * + 金属-氧化物-半导体场效应晶体管 Metal-Oxide-Semiconductor Field-Effect Transistor
        + Source
        + Gate
        + Drain
        + ISFET:离子敏感场效应晶体管（Ion-Sensitive Field-Effect Transistor）



工作原理

SIO2对PH敏感，可以测量PH

场效应晶体管(FET)：其中氢离子敏感层被用来替换传统FET的金属闸极。当氢离子浓度发生变化时，氢离子敏感层(通常是金属氧化物)的电位会发生变化，进而改变晶体管的导电性

洛伊斯效应（Loyds effect）：该效应描述了电解质溶液导电性随氢离子浓度变化的特性。

* + - * + 当DNA聚合酶DNA polymerase 填充一个核苷酸的时候会释放一个氢离子

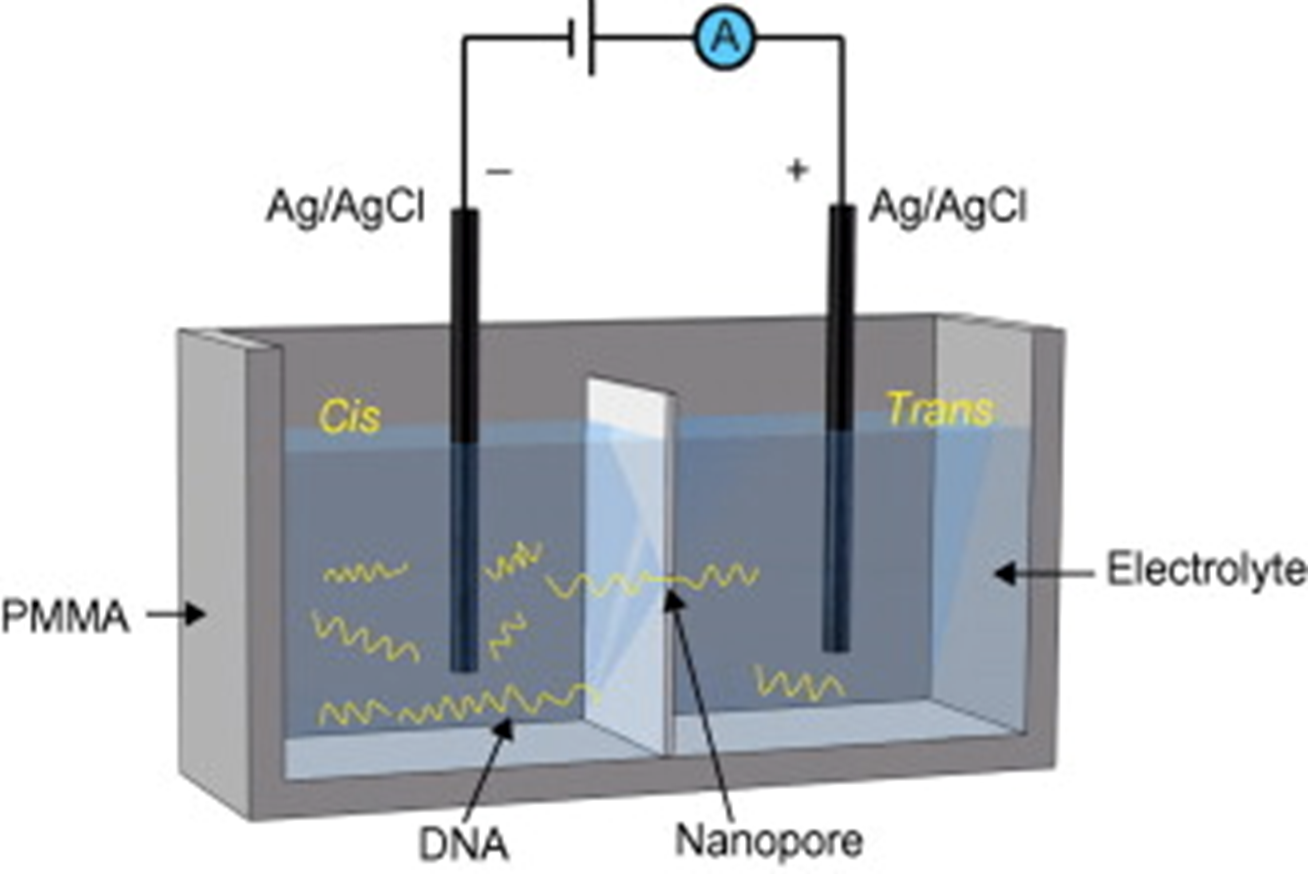
## 第三代测序

### Nanopore Sequencing (Oxford Nanopores) 纳米孔测序(牛津纳米孔)

* + - * + Longer reads
        + Portable (Oxford Nanopore)
        + Speed

## Width of single stranded DNA is 1.5 nm. How do you make nanopores this small?

### Nanopore Sequencing：纳米孔测序



* + - * + DNA带负电，所以会跑到正极

### 它是一种基因组测序技术，利用纳米孔（nanopore）来进行DNA或RNA的测序。这项技术的基本原理是通过一个微小的孔，将待测序的核酸拉伸通过，通过监测核酸分子通过孔道时电流的变化来实现序列的读取。

### 优点

* + - * + 无需PCR
        + 直接测序RNA
        + 更长测序长度

### 制造纳米孔的技术

* + - * + 纳米加工技术： 先进的纳米加工技术，如电子束或离子束刻蚀，可以用于制备精确的孔洞。这些技术允许在材料表面准确地雕刻出纳米尺度的结构，包括纳米孔。
        + 自组装技术： 自组装技术是一种利用自然界中分子自发排列的原理来构建结构的方法。通过合理设计分子或聚合物，可以使它们在特定条件下自组装成所需的形状，包括纳米孔。
        + 纳米粒子模板法： 利用纳米颗粒作为模板，将所需的物质沉积在颗粒表面，然后去除颗粒，留下所需的孔洞。这种方法可通过控制颗粒的尺寸和形状来实现对孔洞尺寸的控制。
        + DNA Origami技术： 利用DNA Origami技术，可以设计并自组装DNA分子形成特定形状，包括纳米孔。这种方法通过合成DNA链并使其自组装成所需的结构，可以实现精确的纳米孔制备。

## DNA travels through the nanopore really fast (less than 10 𝛍s) and causes a modulation in the current of few tens of pA How do you read single bases?

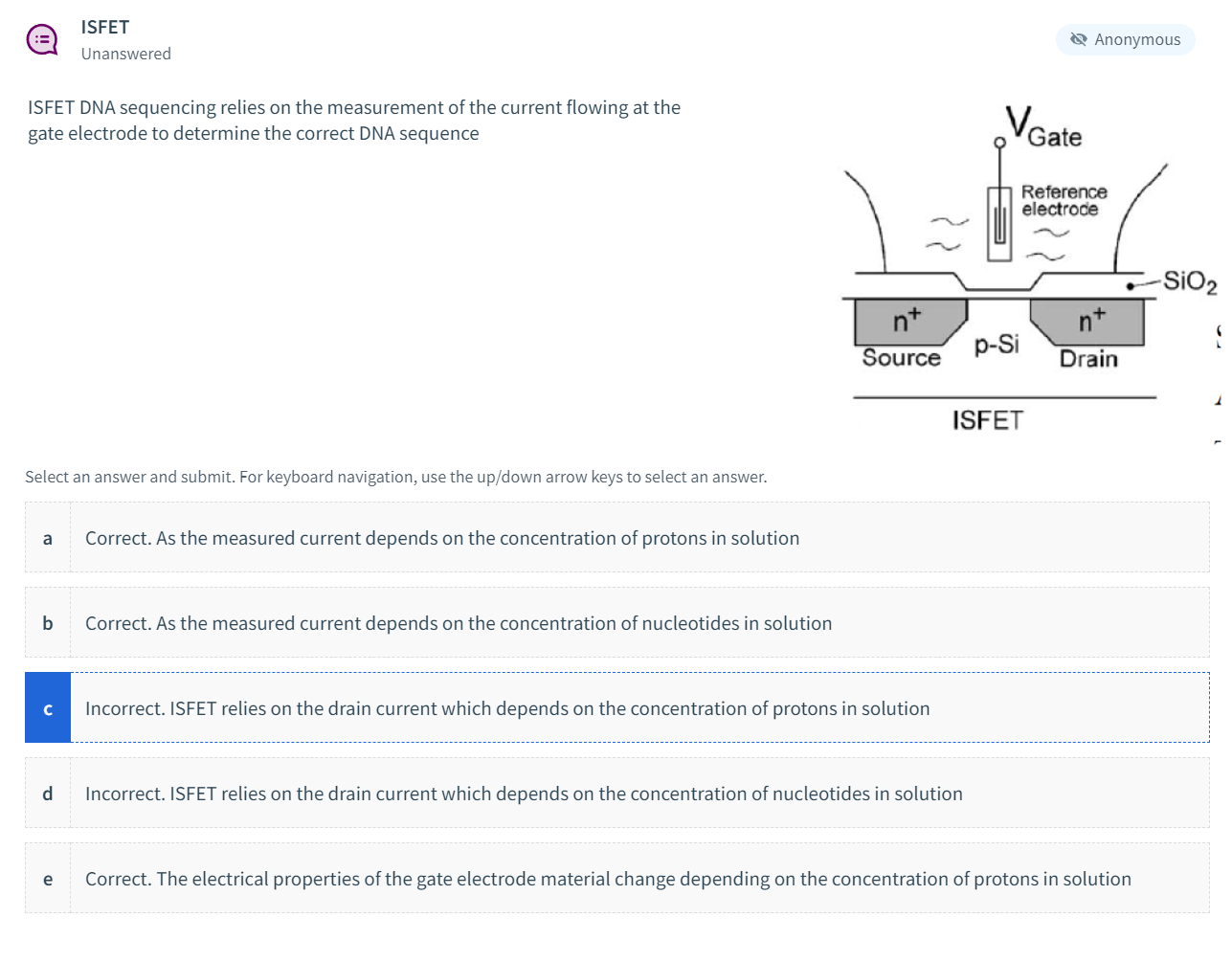
### 在纳米孔测序中，DNA经过纳米孔的速度非常快，通常在微秒级别（小于10微秒），并且会引起几十皮安（pA）电流的调制。具体如何读取单个碱基，涉及到对电流变化的高分辨率检测和解读。

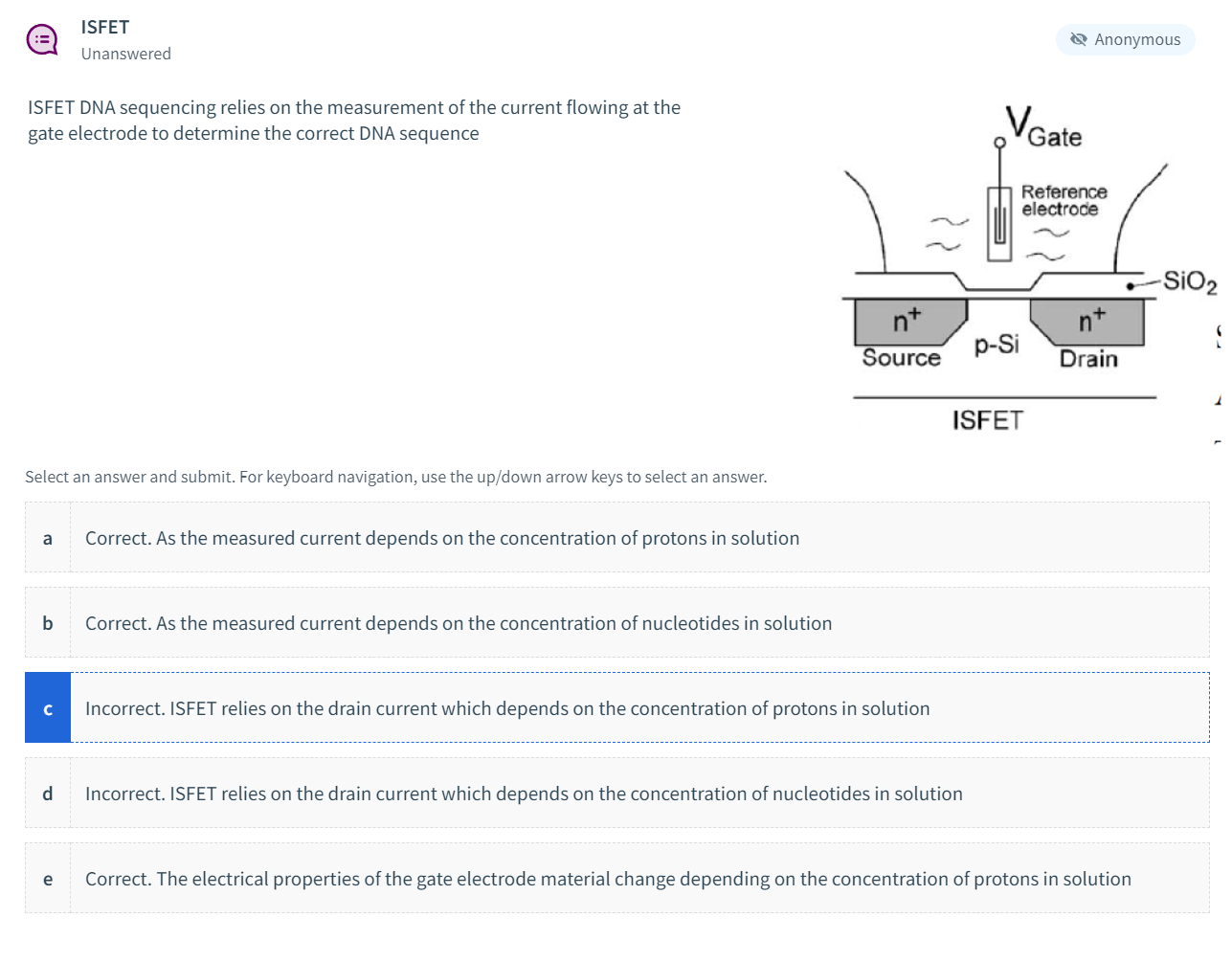
### 电流测量： 通过测量DNA经过纳米孔时电流的微小变化，可以推断出DNA的碱基序列。每个碱基的化学结构和大小都不同，因此当DNA经过纳米孔时，不同碱基引起的电流变化也是不同的。

### 信号解码算法： 采用高度复杂的信号解码算法，通过分析电流信号的模式和变化，可以将这些模式映射到特定的碱基。这可能涉及使用机器学习、神经网络或其他先进的算法来处理大量的电流数据。

### 引入核酸标记物： 有时，在DNA链上引入特殊的标记物，如异质性基团或化学修饰，以便更容易地识别不同碱基。这些标记物可以在纳米孔测序中引起更大的电流变化，提高分辨率。

* + - * + 答案是C





### 控制DNA的速度： 通过调整DNA的传输速度，可以使电流信号更容易解读。这可能涉及到使用外部力或电场来控制DNA的传输速度，以便更好地捕捉和分析每个碱基的电流信号。

* + - * + 使用motor protein

A motor protein sits on top of the nanopore and ratchets DNA down the nanopore with a controlled speed 马达蛋白位于纳米孔的顶部，以可控的速度将DNA沿着纳米孔棘轮移动

# Why we need DNA sequencing

## research the drugs, genomic contribution to disease

### cancer

### Specialised drugs and treatments for indivicuals

## To discover new virus like SARS-COV-2

### design the vaccine that injected

### 30000 bases long while human genome is 3 billion

## To study virus evolution

## Agricultural important species

### Food

## Environmental

## Sequencing of many animal species compared with human

### study the structure and function of the human genome

### study the genome evolution

# 基因测序价格

## 2001

### 100 000 000 dollars

* + - * + 克林顿时期 美国和英国带头提出人类基因组计划
        + 下一代基因组计划

## 2008

### 明显下降

### Next-Generation Sequencing

* + - * + Masssively Parallel
        + Higher throughput
        + Lower cost
        + Faster
        + More accurate
        + You can learn more about this in the Labster lab

## 2021

### 1000 dollars

# Sequencing by Synthesis

## DNA is sheared into 200 bp pieces

## The shared DNA is attached to the surface of a flow cell

## The attached DNA is amplified via PCR

## Sequencing by synthesis using fluorescently labelled nucleotides

# comment thread

# Leroy Hood

# 要读的论文

## An integrated semiconductor device enabling non-optical genome sequencing, Nature, 2011

### 一种集成半导体器件（使用ION芯片），可实现非光学基因组测序

* + - * + 非光学基因组测序：不需要依赖光学设备的DNA测序技术

原理：通过使用集成芯片测序，感知DNA聚合中释放的氢离子变化确定DNA核苷酸顺序

### 实现原理

* + - * + 利用集成电路中的离子敏感器ion-sensitive sensors 直接感知DNA聚合酶polymerase合成synthesis产生的离子，从而获取DNA序列信息。

### 特点

* + - * + 1. 利用半导体技术实现低成本\大规模生产
        + 2. 高准确性和稳健性，适用于细菌基因组测序和全基因组测序
        + 3. 对短序列细菌基因组准确性高
        + 4. 能够识别SNP，实现高质量的全基因组测序结果

## Piet Bergveld - 40 years of ISFET technology: From neuronal sensing to DNA sequencing, Electronics Letter, 2011

### ISFET（Ion-Sensitive Field Effect Transistor）的发明过程

### MOSFET的硅氧化层作为传感层

### 工作原理

* + - * + 常规MOSFET

使用金属作为控制源电极和漏电极间电流的门

* + - * + ISFET

这个门被敏感层代替

可对离子浓度变化进行有效响应

非常薄，纳米级别

### 特点

* + - * + 勿需用到脆弱的玻璃膜电极，避免发生电极损伤的问题
        + 体积小，重量轻，便携
        + 响应速度快，能够实现快速准确地测量

### 优点

* + - * + 抗冲击性强：因为无需脆弱的玻璃膜电极
        + 响应时间短：相较玻璃膜电极
        + 微型化的可能：体积小巧

因为ISFET的基本构造类似于常规的金属氧化物半导体场效应管（MOSFET）,是一种固态设备，由半导体材料制成

### 缺点

* + - * + 长期稳定性欠佳
        + 生产成本高
        + 对环境有要求：有机溶剂对ISFET有一定的腐蚀性

### EMOSFET与ISFET的设备工作原理区别

* + - * + EMOSFET的工作原理是基于闸性材料的性质变化

### 应用场景

* + - * + 油漆工业的传感器

监控ph值

* + - * + 人工心脏起搏器项目
        + 用于测量血液的PH值的导管系统

### 题目

* + - * + ISFET的全称是什么？ A. Inverse-Sensitive Field Effect Transistor B. Ion-Sensitive Field Effect Transistor C. Inert-Sensitive Field Effect Transistor D. Intrinsic-Sensitive Field Effect Transistor

答案是 B. Ion-Sensitive Field Effect Transistor

* + - * + 下列哪一项不是ISFET的优点？ A. ISFET的响应时间短 B. ISFET抗冲击性强，不易损坏 C. ISFET的长期稳定性优于传统的玻璃膜电极 D. ISFET便于实现微型化

答案是 C. ISFET的长期稳定性优于传统的玻璃膜电极

* + - * + 下列哪项技术不适用于ISFET? A. 油漆工业的传感器 B. 人工心脏起搏器项目 C. 基因测序 D. 用于测量血液的pH值的导管系统

答案是 C. 基因测序

* + - * + 以下哪个选项是ISFET的劣势？ A. 生产成本较高 B. 抗冲击性强 C. 响应时间短 D. 便于在微型设备中使用

答案是 A. 生产成本较高

* + - * + EMOSFET是一种不同于传统ISFET的设备，它的工作原理是？ A. 基于测量表面电势 B. 基于闸极材料的性质变化 C. 基于电流与电压的转换 D. 与传统的ISFET工作原理相同

答案是 B. 基于闸极材料的性质变化

## Nanopore Reading

### Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision, Nature Biotech, 2012

* + - * + 内容

主要介绍了一个新的DNA测序技术

利用蛋白质纳米孔或固态纳米孔来分析单个DNA链

DNA通过纳米孔控制前进和后退，平均每秒可以处理2.5-40个核苷酸

介绍了一种新的纳米孔策略需要整合六个功能

1. 自动捕获和处理基因组DNA模板

2. 系统的空间控制

3. 时间控制

4. 避免使用复杂活性电压控制的策略

5. 能确定单个碱基的传感器

6. 能够识别同源聚合体区域中核苷酸之间转换的计数器

减慢DNA模板移动速度的方式

通过配对纳米孔和酶引擎来降低速度的策略

论文中通过DNA聚合酶phi29 DNAP的控制功能来降低DNA在纳米孔中的运动速度

可通过单个纳米孔按照顺序处理多大500个DNA分子

可能发生的错误（导致相对于正确序列的碱基插入或删除）

1. 链位移速度超过数据获取速度导致

2. 链反复滑动使得给定位置被读取多次

概率

插入错误率

5-10.5%

删除错误率

5-15%

综合错误率

10-24.5%

题目

在实验中，DNA strands被通过孔拉链并检查的平均速率是多少？ A. 0.5-4 nucleotides per second B. 2.5-10 nucleotides per second C. 10-20 nucleotides per second D. 2.5–40 nucleotides per second

答案是 D. 2.5–40 nucleotides per second。在文章的摘要部分提到："DNA strands were ratcheted through the pore at median rates of 2.5–40 nucleotides per second..."

在没有优化的情况下，一次通过模板链的过程中的注册错误（插入或删除）的概率是多少？ A. 5% to 15% B. 10% to 25% C. 20% to 40% D. 30% to 50%

答案是 B. 10% to 25%。在文章的摘要部分提到："The probability of a registry error (an insertion or deletion) at individual positions during one pass along the template strand ranged from 10% to 24.5% without optimization."

以下哪项被确定为新的纳米孔策略将需要的六个功能之一？ A. 防止复杂活性电压控制 B. 检测DNA strand的活性部分 C. DNA strand的逐个处理 D. 能确定单个碱基的传感器

答案是 D. 能确定单个碱基的传感器。在文章内容中一开始的部分提到："A promising new nanopore-based strategy will require integration of six features..." 其中包括了答案 D.

以下哪种错误没有在文章中被报告？ A. 链位移速度超过数据获取速度引起的错误 B. 链反复滑动引起的错误 C. DNA strand反复折断引致的错误 D. 碱基插入或删除的错误

答案是 C. DNA strand反复折断引致的错误。在文章的内容中，"A promising new nanopore-based strategy will require integration of six features..." 提到了弹簧型状错误，和三种可能引起注册错误的因素，但并未提及DNA strand反复折断引致的错误。

### Three decades of nanopore sequencing, Nature Biotech, 2016

* + - * + 主要内容

介绍了纳米孔测序的关键

依赖于通过DNA分子的单独核苷酸移动速度来探测DNA序列

DNA分子通过电厂驱动过纳米孔，所以需要精确读取每个核苷酸

DNA分子移动过程涉及的两个关键因素

1. DNA和纳米孔之间的作用力

电场力来调控

2. 纳米孔的几何形状

生物工程或物理方法进行优化设计

确定纳米孔测序是否能区分嘌呤和嘧啶

* + - * + 突破点

串行化方法控制DNA移动过程

利用特殊的酶将DNA序列进行串行化

通过纳米孔进行单个核苷酸测序

特点

大大提高测序速度和精确性

需要解决酶活性和稳定性问题

可以通过其对a-溶血素离子电导的特征来区分由所有胞嘧啶构成的聚-d（C）链中的一种腺苷核苷酸

* + - * + 生产了MinION设备

特点

优点

1. 便携

2. 实时数据分析

3. 长读长

获取更加全面的基因信息

4. 无需特殊设施

连USB

缺点

1. 错误率高

5-15%

2. 需要高性能电脑

3. 实时性导致的问题

设备随着测序的进行表现会下降

* + - * + 题目

MinION设备的尺寸是？ a) 与笔记本电脑一样大 b) 与手机一样大 c) 与USB设备一样大 d) 与书本一样大 以下哪项是纳米孔测序的挑战？ a) 确保DNA分子单独穿过纳米孔 b) 处理背景噪音和误码率 c) 确保酶的活性和稳定性 d) 所有以上选项 以下哪项不是MinION设备的特点？ a) 便携 b) 高成本 c) 能在实时产生测序数据 d) 能长读长 phi29 DNA聚合酶在纳米孔测序中的主要作用是什么？ a) 序列串联化 b) 提高测序速度 c) 防止DNA分子错误穿过纳米孔 d) 增加测序成本 MinION设备的优势是什么？ a) 高精度 b) 高通量 c) 实时产生测序数据和长读长 d) 极低的成本 答案： 1. c) 与USB设备一样大 2. d) 所有以上选项 3. b) 高成本 4. a) 序列串联化 5. c) 实时产生测序数据和长读长

# Covid结构

## Spike Glycoprotein

### Spike glycoprotein（尖刺糖蛋白）是一种存在于冠状病毒（包括SARS-CoV-2）表面的蛋白质。

## RNA and N protein

### 复制和组件的关键

### RNA指的是病毒的遗传物质，即核糖核酸（Ribonucleic acid）。SARS-CoV-2是一种RNA病毒，它的遗传信息存储在单链的RNA分子中，而不是DNA中。

### N蛋白，或称核衣壳蛋白，是冠状病毒中的一种蛋白质，它与病毒的RNA遗传物质紧密结合，形成了一个复合结构，称为核衣壳。

## Envelope

### "Envelope"（包膜）在病毒学中是指一种包围着病毒核衣壳的脂质双层膜。这种脂质双层膜主要由宿主细胞的膜蛋白质和病毒自身的蛋白质组成。包膜的存在赋予了病毒某些特定的属性和功能。

## Hemagglutinin-esterase dimer (HE)

### Hemagglutinin-esterase dimer (HE)是一种存在于某些病毒表面的糖蛋白

# What is an antibody?

## Y-shaped protein used by the immune system to neutralize pathogens 免疫系统用来中和病原体的y形蛋白质

## How do you detect proteins?

### ELISA

* + - * + ELISA是一种常用的实验室技术，用于检测和量化抗体、抗原、蛋白质和激素等物质的存在和浓度。这种方法依赖于特异性抗体与抗原之间的相互作用，以及酶标记的使用，这些酶能够催化产生可检测的信号（通常是颜色变化）。

### Lateral flow assay

* + - * + 侧向流动测定（Lateral Flow Assay，LFA），也常被称为快速检测条或金标快速诊断试剂条，是一种用于快速、简便地检测特定抗体或抗原存在的技术。
        + 不需要专用设备，30分钟内回答定性(是或否)
        + 早早孕试纸：如果刺突蛋白-抗体复合物形成，那么它将在测试线上被针对刺突蛋白的第二组抗体捕获，并且由于金纳米颗粒的存在，测试线上将变成红色

# Experiment

## EXP1: PCR

### What is PCR

* + - * + Make the DNA copy so we can analyse it conveniently

### Learn PCR in the Labster lab check it out

### PCR准备

* + - * + DNA has to be isolated from the cells

### What is the function of primers in a PCR reaction?

* + - * + 子主题 1

### DNA synthesis合成

* + - * + create a new DNA molecules

### DNA replication复制

* + - * + using a DNA template to replication

### DNA polymerases聚合酶

* + - * + It is the enzymes酶 responsible for DNA synthesis in living organisms

### Taq polymerase聚合酶

* + - * + It is active at 95°

### Primers引物

* + - * + Short fragments of DNA or RNA used to start DNA synthesis by a DNA polymerase
        + Will bind to regions of the DNA that match their own sequence.

### nucleotides核苷酸

* + - * + The units on which nucleic acids are built
        + Include ATCG
        + Made up of a sugar molecule

DNA is deoxyribose脱氧核糖

RNA is ribose 核糖

### Template

* + - * + DNA is the template for the PCR

### A new sterile pipette tip should be used every time.

* + - * + To avoid cross contamination
        + A pipette can be used to collect small and precise volumes

### 95° DNA解链

* + - * + 3‘端 结束端

退火Annealing：降低5-10°

降低温度到72°聚合酶开始起作用 开始合成

The polymerase will add nucleotides to the 3' end of the primers

DNA两段的5‘和3’位置是互换的

suoyi

* + - * + 5’端 开始端

### The placement of the primers defines the fragment length

### Electrophoresis is employed in biotechnology for the size-based separation of macromolecules such as proteins and nucleic acids

* + - * + anode+

DNA and RNA are negatively charged due to their phosphate backbone

* + - * + cathode-

### 至少需要13个不同区域的样本才能准确判断DNA

### DNA ladder

* + - * + Used as a scale for determining the lengths of unknown nucleic acid fragments when performing gel electrophoresis experiments
        + Determing the length of the DNA

## EXP2:Isolate DNA

### Electrophoresis电泳

* + - * + It is a way to separate molecules include DNA by size.

It is a technique to separate charged macromolecules based on their size using an electric current.这是一种利用电流根据带电大分子的大小分离带电大分子的技术。

* + - * + Agarose gel琼脂糖凝胶

常常被用来做凝胶电泳的工具，用来分离DNA\RNA或者蛋白质

* + - * + 负极到正极

因为DNA和RNA是负电由于它们的磷酸盐主键(Phosphate Backbone)

### What is the Nucleic acid核酸 ladder or Molecular Weight Standard?

* + - * + It is a mix of DNA or RNA fragments with known lengths.
        + "Nucleic acid ladder"（核酸阶梯）是实验室中使用的一种标准或参照物，用于核酸电泳过程中比较和确定DNA或RNA样品的大小。

### Buffer

* + - * + Provide ions that ensure a current flow from anode to cathode.提供确保电流从阳极流向阴极的离子。

### Gel

* + - * + Acts as a Molecular sieve分子滤网,降低大分子DNA片段的移动速度

所以大分子DNA会停留在顶端移动不了

## EXP3: Next Generation Sequencing(NGS)

### It is a cutting edge technique that enables sequencing of many DNA molecules at the same time.

* + - * + 并行测序

### 测序之后有什么用？

* + - * + 可以通过SNP分析确定他的物理特征

SNP（Single Nucleotide Polymorphism，单核苷酸多态性）分析是一种遗传学和分子生物学技术，用于检测基因组中个体间单个核苷酸的变异。SNP是基因组中最常见的变异形式，它们是由于单个核苷酸（例如，A、T、C或G）的替换、插入或删除而引起的。

SNP鉴定：通过基因组测序、SNP芯片或其他高通量方法，对基因组进行扫描，识别SNP位点。

SNP基因型鉴定：确定特定个体在SNP位点上的基因型，即其所携带的等位基因。

SNP频率分析：评估特定SNP在群体中的频率，以了解其在人群或物种中的分布情况。

关联分析：将SNP与特定性状或疾病之间的关联进行分析，以探索其在遗传疾病、个体特征或药物反应等方面的潜在作用。

* + - * + 获得基因表达谱 Gene expression profiling
        + 检查遗传畸变Detecting genetic aberration

### 步骤Steps

* + - * + 1. 提取DNA，DNA Fragmented into small pieces DNA分成小块

Often generates overhangs, also known as sticky ends

* + - * + 2. Repair the sticky ends, making them into blunt ends.

核酸外切酶Exonuclease enzyme has attached to the DNA molecule and will now remove the 3' overhangs

通过磁珠magnetic beads 绑定DNA

磁珠的表面可以修饰有各种吸附或结合特定生物分子的化学基团或生物配体（如抗体、亲和标签、寡核苷酸等）。当含有目标分子的样本添加到这些磁珠中时，目标分子会特异性地结合到磁珠上。随后，使用磁架或磁棒在外部施加磁场，使得载有目标分子的磁珠在溶液中聚集并靠拢到容器的一侧或底部。这时，非结合的杂质可以通过简单的洗涤步骤去除。最后，去除磁场后，可以通过适当的洗脱步骤释放并收集目标分子。

* + - * + 3. Adenine腺嘌呤 is added to the 3' end to create an A-overhang

To ensure that the adapters bind specifically

* + - * + 4. 加Adapter

Adapter是短的、人工合成的单链或双链DNA序列，它们在测序前被连接（连接）到DNA片段的两端。

For the PCR primers to bind to it

adapter序列中包含了用于PCR扩增的引物结合位点。

The DNA needs to be anchored to a plate surface where it will again amplified and then sequenced.DNA 需要锚定在板表面，在那里它将再次扩增并测序。

This plate surface is also called a flow cell and has a high density of short DNA strands atached to its surface.该板表面也称为流动池，其表面附着有高密度的短 DNA 链。

Because the sequences are complementary, 所以他们才能与DNA绑定

* + - * + 5.PCR

正常30次，这次15次是因为担心降解后的DNA复制次数过多导致出错

也叫做Bridge PCR

A polymerase will then attach to the purple molecule and amplify the DNA然后聚合酶会附着在紫色分子上并扩增 DNA

Each cluster has amplified 4000 DNA

去除一半DNA

Because maybe only DNA we can bound to the blue molecule, so we need to move the DNA which is linked to purple molecule

We need to reserve the 3' to 5' or 5' to 3' so that we can bound it.

Each one flow cell we have 200 million clusters

We can sequence it at the same time.

* + - * + 6. Generating clusters

目的是在固定的表面上（如流动槽的表面）通过桥式扩增（bridge amplification）产生数以百万计的DNA分子的密集复制品，每个复制品形成一个单独的、空间上分离的聚集体或簇。

Bind with the adapters

DNA片段与adapter结合后可以桥式扩增，形成相同的簇。

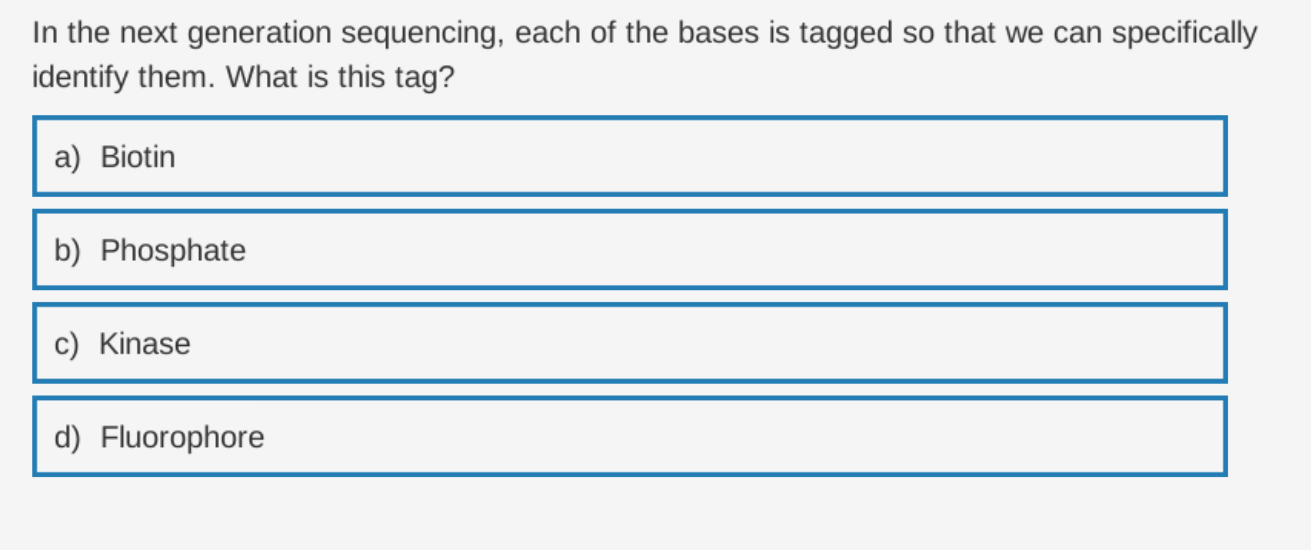
* + - * + 7. Sequenceing

Which enzyme is used in the sequencing process?

T4 DNA polymerase

Sequencing by synthesis

D 萤光团



Each of the bases is specifically tagged with a fluorophore

Each picture records only one specific color, so we need 4 pictures.

If we sequence only from one direction, it is called single-end sequencing

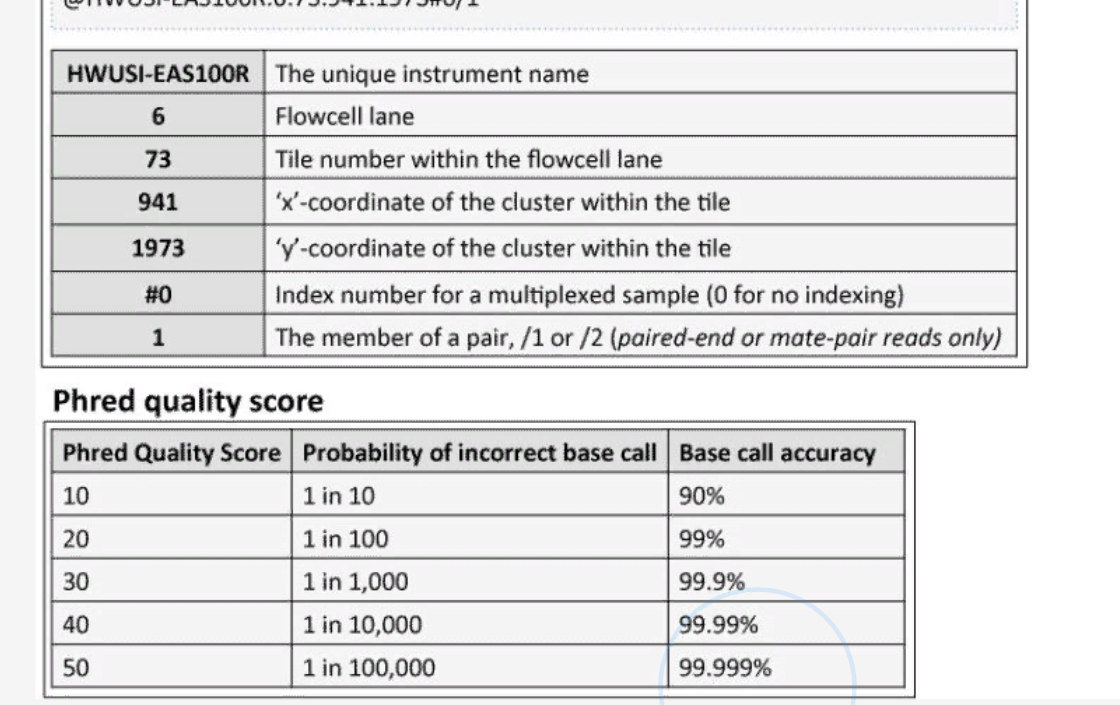
* + - * + 8. NGS Data analysis

Primary analysis

Includes all the steps required to call or identify each base

Stored as a FASTQ file, containing the sequence identifiers, ATCG and the associated Phred quality score

Phred quality score refers to the probability of an incorrect base calling.Phred 质量分数是指错误碱基识别的概率。



Secondary analysis

This step is aim to assemble all the short DNA sequences in order to interpret the sequence data.此步骤的目的是组装所有短 DNA 序列以解释序列数据。

The first step is to trimming out the adapters,去除掉我们不感兴趣的非基因片段

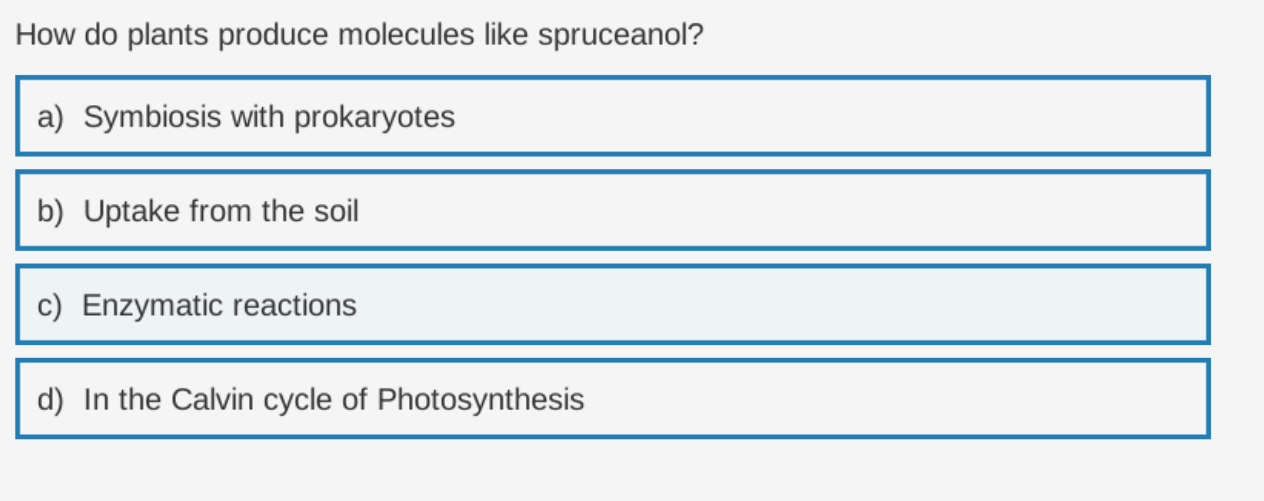
Alignment is performed when you have a reference genome to compare to.当您有可供比较的参考基因组时，就会执行比对。

Since the human genome has been sequenced already

Tertiary analysis

## EXP4: Bioinformatics: An introduction

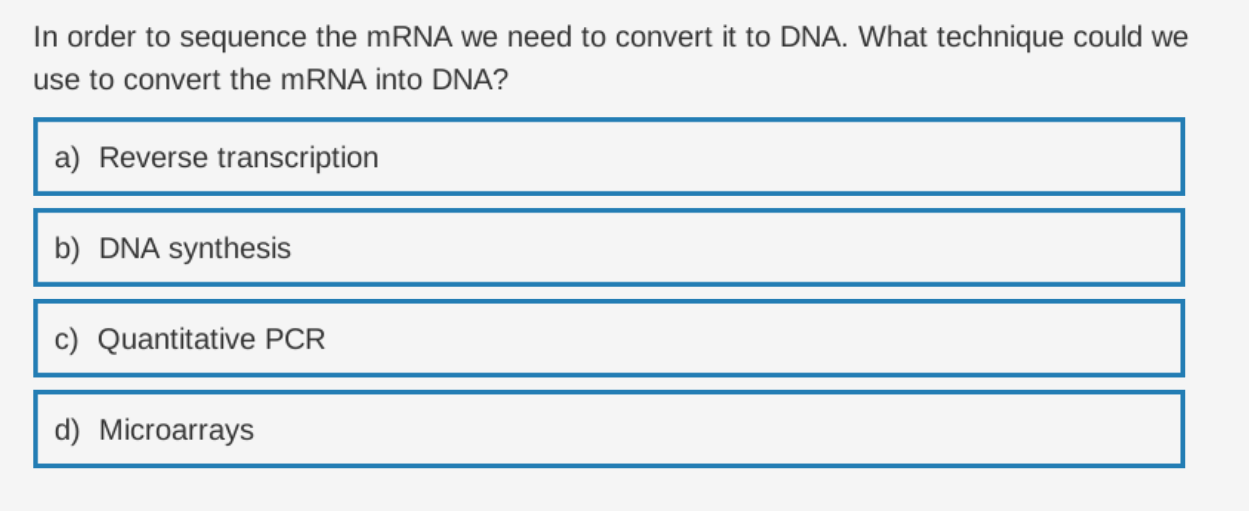
### spruceanol 云山醇



* + - * + It is produced via enzymatic reactions from a precursor molecule它是通过前体分子的酶促反应产生的
        + if we can identify and isolate the genes that encode the enzymes needed forspruceanol synthesis we can transform cells from a different species to producespruceanol in large quantities.如果我们能够识别并分离出编码云杉醇合成所需酶的基因，我们就可以转化不同物种的细胞以大量生产云杉醇。
        + We can measure the gene expression and the resulting protein levels byquantifying the mRNA. Genes in the same gene family have similar sequences. Hence we should be able toidentify the mRNA that encodes the Cytochrome P450 enzymes.

我们可以通过量化 mRNA 来测量基因表达和由此产生的蛋白质水平。 同一基因家族中的基因具有相似的序列。 因此，我们应该能够识别编码细胞色素 P450 酶的 mRNA。

* + - * + A



The introns are spliced out of the pre-mRNA before it is exported out of thenucleus as mature mRNA. This means our cDNA won't contain any introns.

在将前体 mRNA 作为成熟 mRNA 输出到细胞核外之前，将内含子从前体 mRNA 中剪接出来。 这意味着我们的 cDNA 不会包含任何内含子。

因为PCR不能对RNA起作用，所以需要把mRNA变成cDNA 然后进行PCR

### phylogenetic tree

* + - * + Show the evolutionary pathways and connections among organisms or groups of organisms