新一代DNA测序技术的原理及其应用

New generation DNA sequencing: mechanisms and applications

陆祖宏

北京大学 工学院 生物医学工程系



一、为什么要发展DNA测序技术

- 二、第一代测序技术原理与技术
- 三、第二代测序技术的原理与技术
- 四、第三代测序技术原理与技术
- 五、新一代测序技术的产业前景





一、为什么要发展DNA测序技术

- 二、第一代测序技术原理与技术
- 三、第二代测序技术的原理与技术
- 四、第三代测序技术原理与技术
- 五、新一代测序技术的产业前景





1. Fertilized Egg

2. 4-Cell Embryo



3. 8-Cell Embryo



4. Blastocyst



5. Hatching Blastocyst

复杂多细胞生 命体由单个受 精卵细胞发育 而成的。







Small molecule subunits Macromolecule

U.S. DEPARTMENT OF ENERGY

2014-4-3

生命的中心法则 Central dogma





Non-coding RNAs

DNA: 生命系统的软件

■ DNA序列是编码所有生命系统的分子程序。■ 解析DNA序列是生命科学研究的重要内容。





з



×

Y

Mouse chromosome key





基因的表达调控







基因转录调控





人类基因组多态性的分类

基因组结构多 态性大于 1000bp的差异

序列多态性小 于1000bp的差 异

染色体数量的变化 大片段的重复、拷贝数变化 移位 颠换 序列重复 转座 短序列删除和插入 微卫星 单核苷酸的插入和删除 单核苷酸多态性 突变







新一代基因组测序原理和应用

单核苷酸多态性 Single Nucleotide Polymorphisms (SNPs)



2014-4-3





一、为什么要发展DNA测序技术

- 二、第一代测序技术原理与技术
- 三、第二代测序技术的原理与技术
- 四、第三代测序技术原理与技术
- 五、新一代测序技术的产业前景



1977年开始的DNA测序技术

- Allan Maxam and Walter Gilbert at Harvard University (PNAS, Feb, 1977).
- Frederick Sanger at the U.K. Medical Research Council (MRC) (*PNAS*, Dec, 1977).
- 1980年获诺贝尔化学奖。

2014-4-3





MAXAM-GILBERT PROCEDURE



Figure 5-12 Two restriction maps for a segment of DNA. X and Y are simply labels for fragments generated by cuts at sites indi-

cated by arrows. The color indicates a set of sites for a particular restriction enzyme.



Sanger Sequencing





Sequencing Reaction



Collection of fragments of newly synthesized DNA: They all end in ddA at locations of complementary T bases in the template

ATGC



The first automated DNA sequencing machine in1986



Leroy Hood and Lloyd Smith of the California Institute of Technology



Multiple-sheathflow capillary array DNA analyser Kambara, Hideki Nature. Vol. 361, no. 6412, pp. 565-6. 11 Feb. 1993







Megabace 测序仪

3700 测序仪

2014-4-3

新一代基因组测序原理和应用

23

人类基因组计划1990-2001





一、为什么要发展DNA测序技术

- 二、第一代测序技术原理与技术
- 三、第二代测序技术的原理与技术
- 四、第三代测序技术原理与技术
- 五、新一代测序技术的产业前景





2005年 ABI 3730完成一个人类基因组的测序需要 100台仪器工作6个月,花费1000万美元

1000美元人类全基因组测序

文特尔在2003年9月首次 提出1000美元人类基因 组全测序的目标。 J. Craig Venter Science Foundation 设 立了\$500,000 Technology Prize



2004年: **Dramatic reductions in** sequencing cost will lead very different approach of biomedical researches, and eventually, will revolutionize the practice of medicine.



Francis S Collins the head of NIH



- Whole human genome re-sequencing
 - Present Cost: about 10, 000, 000 USD
 - Near Goal: 100, 000 USD
 - Ultimately: 1, 000 USD

DNA sequencing costs have fallen more than 100-fold over the past decade, fueled in large part by tools, technologies and process improvements developed as part of the successful effort to sequence the human genome. However, it still costs at least \$10 million to sequence 3 billion base pairs the amount of DNA found in the genomes of humans and other mammals.

NHGRI's near-term goal is to lower the cost of sequencing a mammalian-sized genome to \$100,000, which would enable researchers to sequence the genomes of hundreds or even thousands of people as part of studies to identify genes that contribute to cancer, diabetes and other common diseases. Ultimately, NHGRI's vision is to cut the cost of whole-genome sequencing to \$1,000 or less, which would enable the sequencing of individual genomes as part of medical care. The ability to sequence each person's

2006:



X Prize 提出了一千万美元的奖励, 以鼓励人们发展\$1000以下的人类全 基因组测序技术 非营利组织X Prize基金会举办了一个 被称为"基因组学Archon X Prize"的竞 赛,第一个能够在10天内为超过100 个人完成DNA测序工作的小组将获得 1,000万美元奖金。



2014-4-3

新一代基因组测序原理和应用

2013:





新一代DNA测序技术是近十年来生物医学领域中发展最快、 影响最大、竞争最为激烈的高技术研究领域之一。





个人基因组的测 序成本已经降到 1000美元以下, 测序时间降到1 周。



HiSeq X Ten测序平台



2014-4-3

(1) 微乳法在磁珠上生长DNA,微孔阵列并行的生物发光合成测序





Pyrosequencing reaction

新一代基因组测序原理和应用

Pyrosequencing Biochemistry



2014-4-3
Try Times Reader Free				Science				
WORLD	U.S.	N.Y. / REGION	BUSINESS	TECHNOLOGY	SCIENCE	HEALTH	SPORTS	OPINION
ENVIRONMENT SPACE & COSMOS								
📥 DELTA 👳								► HEALTH

CHANGE IS: A NEW EXPERIENCE

We're revolutionizing the way you travel

Genome of DNA Discoverer Is Deciphered

By NICHOLAS WADE Published: June 1, 2007

The full genome of James D. Watson, who jointly discovered the structure of DNA in 1953, has been deciphered, marking what some scientists believe is the gateway to an impending era of personalized genomic medicine.

Enlarge This Image



Richard Carson/Reuters

James D. Watson, co-discoverer of the DNA helix and father of the Human Genome Project, prepares to autograph his book at the Baylor College of Medicine's Human Genome Sequencing Cente 2014-4-3

DVDs, was presented to Dr. Watson Richard A. Gibbs, director of the

OR SAVE THIS SINGLE PAGE REPRINTS + SHARE ARTICLE TOOLS SPONSORED BY

SIGN IN TO E-MAIL

DELTA

SEE MO

A copy of his genome, recorded on two yesterday in a ceremony in Houston by

Human Genome Sequencing Center at the Baylor College of Medicine, and by Jonathan M. Rothberg, founder of the company 454 Life Sciences.

"I am thrilled to see my genome," Dr. Watson said.

Dr. Rothberg's company makes an innovative DNA sequencing machine, the latest version of which proved capable of decoding Dr. Watson's genome in two months at a cost of less than \$1 million, said Michael Egholm, 454's vice president for research. The sequence was verified and analyzed by Dr. Gibbs's center in Houston. It was Dr. Gibbs who proposed the idea of sequencing Dr. Watson's genome.

(2) 在基片上生长DNA, 四色荧光的合成测序:



Amplify to form clusters



100um

Random array of clusters

1000 molecules per ~ 1 um cluster 1000 clusters per 100 um square 40 million clusters per experiment

Illumina Solexa Sequencing





First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell. After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.



Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

新一代基因组测序原理和应用



2014-4-3

新一代基因组测序原理和应用

2008年:

A 218

个中国人基因组图谱

Diploid lengm+

6 November 2008 | www.nature.com/nature | £10 THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

0

LASS WHEN READY

nre

the human denome

- Individual genomes from Africa and China
- Acute myeloid leukaemia genome
- Designer nucleases for gene therapy
- Tracing gene flow across Europe

JR LIFE IN JR HANDS

is for the personal genome age



一代基因组测序原理和应用



(3) 微乳法在磁珠上生长 DNA, 微珠阵列并行的荧光连接测序



Reagent handling

ABI SOLiD Workflow



2014-4-3

(4) 微乳法在磁珠上生长DNA, 微孔半导体pH-ISFET阵列的合成测序



每一个碱基配对事件的效果:

■释放一个质子

■在传感单元上方的小槽中产生瞬间pH值的变化

■造成ISFET表面势的变化,进而导致MOS管阈值电压的移动

■进一步导致ISFET的源漏电流的变化

■相间有重复碱基时,电流变化加大





每一次输入4个碱基之一的时间通常是10秒左右,在此时间 段内,捕获整个信号变化的过程 两次输入之间一个0.1秒左右的清洗过程

全自动化的核酸导入、探测和分析



(5) 微乳法在磁珠上生长DNA, 微珠阵列并 行的简并碱基组合荧光合成测序

对待测核酸序列三种不同模式的采样测序,如采用两核苷酸交替采样方法,则可以分成三组对同一模板进行三次测序:

每组测序由包含四个标记的核苷酸dATP、dGTP、dCTP、dUTP,按照每个核苷酸在一个循环中只使用一次的方式,进行两次由两个不同标记核苷酸同时合成测序反应的循环,每进行一次测序反应得到由核苷酸(碱基) 片段构成的一个编码,若干次测序反应后得到由一组若干编码构成的核酸序列信息;

当该组测序反应完成后,通过变性将测序引物延伸链清除,重新杂交测 序引物,进行下一组测序反应,最后将三组测序反应获得的三组编码信 息,通过解码转化成对应的三组核苷酸(碱基)片段信息,并通过比较 三组核苷酸(碱基)信息组装出待测核酸序列的具体碱基信息。

组循环合成测序步骤参与合成测序碱基



5'-ATTAGTCCAGGGTAAAACCGGAT-3'

- 3: 5'-(AATT)GT(CC)A(GGG)(AAAAT)(CCGG)(AT)-3'
- 2: 5'-A(TT)A(GT)(ACC)(GGGT)(AAAACC)(GG)AT-3'
- 1: 5'-A(TT)(AG)(CCT)(AGGG)T(AAAA)(CC)(AGG)T-3'



解码测序方法的优势

采用二种碱基同时延伸,试剂合成简单,生化反应 效率高,测序速度快;

采用荧光标记,可简化光学系统,测序速度快;

在解码过程中具有自校正方案,可以提高测序的准确性;

测序精度高,目前所有测序方法的读出误差均与单 个碱基反应误差成正比,而唯一解码测序方法为与 单个碱基反应误差的平方成正比;

AG-100型高通量DNA测序仪



2011年3月二台AG-100A测序仪 在深圳华大基因稳定运行





新一代基因组测序原理和应用



▶ 已完成AG100系列产品样机4台。



2014-4-3



一、为什么要发展DNA测序技术

- 二、第一代测序技术原理与技术
- 三、第二代测序技术的原理与技术
- 四、第三代测序技术原理与技术
- 五、新一代测序技术的产业前景

第三代DNA测序技术的主要特征

▶ 第二代测序技术采用扩增的DNA测序文库 ,主要基于生化反应原理的测序方法(包 括基于DNA聚合酶作用的合成测序法和基 于连接酶作用的连接测序法)。

▶ 第三代测序技术采用单个DNA分子作为测 序文库,除了合成测序法之外,还有纳米 孔测序等物理原理的测序方法。

(1) 合成测序,单荧光分子的检测



Helicos Bio





2014-4-3

(2) 适时合成测序,零波导单荧光分子的检测





2014-4-3







Figure 2 | **Single-channel recordings comparing permanent and transient adapters. a**, The WT-(M113R/N139Q)₇ αHL pore showing transient adapter binding (40 μM am₂βCD) and nucleotide detection. Histogram of the residual current of nucleotide binding events. **b**, Corresponding data for the WT-(M113R/N139Q)₆(M113R/N139Q/L135C)₁ mutant with a covalently attached am₆amPDP₁βCD, allowing continuous nucleotide detection and enhanced nucleotide discrimination (see histogram). The traces were recorded in 800 mM KCl, 25 mM Tris HCl, pH 7.5, at +160 mV in the presence of 10 μM dG MP, 10 μM dTMP, 10 μM dAMP and 10 μM dCMP.



Figure 3 | Nucleotide event distributions with the permanent adapter. a, Single-channel recording from the WT-(M113R/N139Q)₆(M113R/N139Q/L135C)₁am₆amDP₁βCD pore showing dGMP, dTMP, dAMP and dCMP discrimination, with coloured bands (three standard deviations from the centre of the individual Gaussian fits) added to represent the residual current distribution for each nucleotide. **b**, Corresponding residual current histogram of nucleotide binding events including Gaussian fits. Data acquired in 400 mM KCl, 25 mM Tris HCl, pH 7.5, at +180 mV in the presence of 10 μM dGMP, 10 μM dTMP, 10 μM dAMP and 10 μM dCMP.

Oxford Nanopore Technologies Co.



MinION

2014-4-3

1211121122723



Oxford Nanopore Technologies Co.

(4) Solid-state nanopore



A. 纳米孔器件承载膜太厚 B. DNA分子穿孔太快



9 September 2010 THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

DNA SEQUENCING Enter the graphene nanopore

BIOLOGICAL NOISE Taking the good with the bad

THE NEW TAXONOMY Automation rules

DINOSAUR MORPHOLOGY

The hunchback of Las Hoyas NATUREJOBS Project management

TGCATTGCTAGCAATGCATTGC ATGCATGGCTAGCAAAAAGCATGGT AATGCACCCTGCTAAAGGGATACAG















Letter

pubs.acs.org/NanoLett

Pressure-Controlled Motion of Single Polymers through Solid-State Nanopores

Bo Lu,^{†,§} David P. Hoogerheide,^{‡,§} Qing Zhao,^{*,†} Hengbin Zhang,[†] Zhipeng Tang,[†] Dapeng Yu,^{*,†} and Jene A. Golovchenko^{*,‡}

[†]State Key Laboratory for Mesoscopic Physics and Electron Microscopy Laboratory, School of Physics, Peking University, Beijing 100871, People's Republic of China

[‡]Department of Physics, Harvard University, Cambridge, Massachusetts 02138, United States







www.MaterialsViews.com

Boron Nitride Nanopores: Highly Sensitive DNA Single-Molecule Detectors

Song Liu, Bo Lu, Qing Zhao,* Ji Li, Teng Gao, Yubin Chen, Yanfeng Zhang, Zhongfan Liu, Zhongchao Fan, Fuhua Yang, Liping You, and Dapeng Yu*

(5) 单分子成像测序法

HALCYON MOLECULAR








2. Stretch DNA onto imaging substrate



Platination



新



Direct platinator: 21 labels made







Osbipy



HM62, Ir_4 platinator



Au₁₁, electrostatic binding



Uniform ssDN

2014-4-3

+ + + +

+

+

and a second

ן גיודו פותפייקטריפיש טין

0



Modified Nion UltraSTEM 100

- <1 Å resolution @ 100 kV
- CFEG: high brightness
- Ultra-stable sample stage: ~0.1 Å/min
- UHV: 1e-9 @ sample
- Simultaneous detector combinations
- 4 additional Hitachi scanning transmission EMs

(6) 蛋白质晶体管测序器件



DNA sequencing using electrical conductance measurements of a DNA polymerase

Yu-Shiun Chen12, Chia-Hui Lee3, Meng-Yen Hung3, Hsu-An Pan23, Jin-Chern Chiou24 and G. Steven Huang^{2,3*}

The development of personalized medicine-in which medical treatment is customized to an individual on the basis of genetic information-requires techniques that can sequence DNA quickly and cheaply. Single-molecule sequencing technologies, such as nanopores, can potentially be used to sequence long strands of DNA without labels or amplification, but a viable technique has yet to be established. Here, we show that single DNA molecules can be sequenced by monitoring the electrical conductance of a phi29 DNA polymerase as it incorporates unlabelled nucleotides into a template strand of DNA. The conductance of the polymerase is measured by attaching it to a protein transistor that consists of an antibody molecule (immunoglobulin G) bound to two gold nanoparticles, which are in turn connected to source and drain electrodes. The electrical conductance of the DNA polymerase exhibits well-separated plateaux that are \sim 3 pA in height. Each plateau corresponds to an individual base and is formed at a rate of ~22 nucleotides per second. Additional spikes appear on top of the plateaux and can be used to discriminate between the four different nucleotides. We also show that the sequencing platform works with a variety of DNA polymerases and can sequence difficult templates such as homopolymers.

ccurate DNA sequencing is critical to personalized medicine^{1,2} and requires a high-throughput technique that can decode genomes at an affordable price and within a reasonable timeframe³⁻⁶. In the past decade, next-generation sequencing technologies have been developed that are based on arrayed reactions that sequence amplified DNA targets7-10. Compared with first-generation Sanger sequencing, this approach significantly reduces the time required to completely sequence a human genome, but the short read length and high error rate limit its further application to unknown genomes2,11,12

Third-generation sequencing (single-molecule sequencing technology) does not require amplification, ligation or cloning, and is expected to provide single-molecule resolution, a long read length and negligible error rate, together with a reduction in cost¹¹⁻²¹. Such methods typically involve cyclic reactions using fluorescent substrates that are monitored by optical imaging^{11,12,14}, and have, for example, been used to sequence the M13 viral genome14.

An alternative third-generation approach is nanopore sequencing, which identifies a molecule by measuring the modulations in the ionic current across a synthetic or biological pore as a DNA molecule is driven through it under an applied potential¹⁶⁻¹⁸. This approach has been used to read DNA at single-nucleotide resolution by using a phi29 DNA polymerase (Φ 29) to control the rate of DNA translocation through a MspA nanopore¹⁷. Oxford Nanopore and the enzyme-substrate complex could become more conductive Technologies has also reportedly used a prototype nanopore device to decode a viral genome in a single pass of a complete DNA strand^{19,20}.

However, the performance of commercialized third-generation technology is currently only comparable to next-generation sequencing methods13,22 and the problems of short read lengths and high

sequencing method that can provide high-quality sequencing data with long read lengths and low error rates is still required.

DNA polymerase is an enzyme that catalyses the synthesis of DNA when provided with the four deoxynucleoside-5'-triphosphates (dATP, dTTP, dCTP, dGTP; abbreviated dNTP), a template strand and a primer with a free 3'-OH end25. During the reaction, a complementary dNTP is chosen based on template base-pairing, which forms a phosphodiester bond to the 3'-OH of the primer and releases pyrophosphate. The chain elongates as DNA polymerase proceeds along the template strand, before dissociating from the template. The interaction between a dNTP and DNA polymerase exhibits a classical Michaelis-Menten mechanism consisting of substrate-binding (base-pairing) and bond-formation steps26

Electrical conductivity is a useful tool for monitoring singlemolecule dynamics27. The sensitivity, reproducibility, reversibility, convenience and dynamic response of conductance measurements should also be beneficial in monitoring enzyme dynamics. Therefore, by tracking the electrical conductance of DNA polymerase during replication, it should be possible to identify sequencespecific nucleotide incorporation due to differences in base pairing and chemical composition. In particular, the electrostatic reorganization and non-covalent interactions between the enzyme and the substrate may enhance the electron transfer environment28, during catalysis, particularly during the reorientation and redistribution of charges in the transition state.

In this work, we aim to develop a single-molecule sequencing technology based on the measurement of the electrical conductance of DNA polymerase as nucleotides are incorporated into the growing DNA strand. A protein transistor, which provides stable error rates have yet to be solved^{23,24}. Accordingly, a third-generation conductance readings, was designed to hold a DNA polymerase

NATURE NANOTECHNOLOGY | ADVANCE ONLINE PUBLICATION | www.nature.com/naturenanotechnology

nature nanotechnology

A protein transistor made of an antibody molecule and two gold nanoparticles

Yu-Shiun Chen¹, Meng-Yen Hong² and G. Steven Huang^{2*}

A major challenge in molecular electronics is to attach electrodes to single molecules in a reproducible manner to make molecular junctions that can be operated as transistors. Several attempts have been made to attach electrodes to proteins, but these devices have been unstable. Here, we show that self-assembly can be used to fabricate, in a highly reproducible manner, molecular junctions in which an antibody molecule (immunoglobulin G) binds to two gold nanoparticles, which in turn are connected to source and drain electrodes. We also demonstrate effective gating of the devices with an applied voltage, and show that the charge transport characteristics of these protein transistors are caused by conformational changes in the antibody. Moreover, by attaching CdSe quantum dots to the antibody, we show that the protein transistor can also be gated by an applied optical field. This approach offers a versatile platform for investigations of singlemolecule-based biological functions and might also lead to the large-scale manufacture of integrated biolectronic circuits.

he electronic properties of an isolated protein molecule depend on its orientation1-3, and because most proteins maintain their native structures only in physiological media, measurements are only meaningful when the protein has a welldefined orientation and is in a hydrated form4.5. Scanning tunnelling microscopes (STMs)467 and atomic force microscopes (AFMs)58-12 have been used to measure the electrical conductance of selfassembled monolayers or single molecules of proteins. For example, the STM break-junction approach13 has been used to probe electron transport in a variety of organic molecules3,14-17, An STM can also be used to measure the electrical properties of molecules attached to a metal substrate18-21, and this method has been used to study proteins such as azurin46,12,22, bacteriorhodopsin^{2,22}, yeast cytochrome c (ref. 9) and ferritins^{8,11}.

Based on analyses of electronic coupling strengths, it has been suggested that the efficiency of long-range electron transfer in proteins depends on their secondary structure. In particular, structures called sheets appear to mediate coupling more efficiently than helical structures (even though hydrogen bonds have a critical role in both structures18). These studies suggest that it might be possible to modulate the charge transfer properties of proteins-and thus make a protein field-effect transistor-by manipulating their secondary structure.

Attempts have been made to integrate protein monolayers into electronic devices23-27. For example, photosynthetic protein complexes have been integrated into organic self-assembled monolayers on gold surfaces in solid-state electronic devices²⁴, and the orientation of the proteins is controlled by specific binding of polyhistidine-tag-Ni2+ to the monolayers. Internal quantum efficiencies of ~12% have been achieved for photodetectors and photovoltaic cells. A protein thin-film transistor consisting of a 100 nm nanogap coated with an azurin monolayer25 has also been constructed, but the currentvoltage (I-V) performance deteriorates over time, probably due to an unstable molecular junction between the electrode and protein. A vertical-type molecular transistor made with a 4 nm channel and a bovine serum albumin monolayer between the source and drain electrodes has shown high gate sensitivity27.

Attempts to construct a single-molecule protein device have led to the fabrication of a large-scale nanojunction array that serves as a framework for azurin molecule adaptation26. The nanojunction, which has a nanogap of ~5 nm and azurin immobilized on the bottom electrode via a disulphide bridge, is claimed to be able to integrate a single protein molecule between the electrodes.

It is now technically possible to achieve electrical contact between both sides of a protein, but, in the absence of specific binding, forcing an electrode or AFM tip into a protein can cause unpredictable denaturation of the protein that, in many instances, results in inconsistent electron transfer. Furthermore, to achieve large-scale fabrication of molecular devices, the ability to reliably self-assemble the molecular junction is an important consideration.

In this work we use self-assembly between an antibody and antigen to make molecular junctions. We have previously isolated an immunoglobulin G (IgG) antibody that recognizes 5-nm-diameter gold nanoparticles28,29. We allowed this anti-nanoparticle IgG to bind to two gold nanoparticles to form a junction. IgG is a Y-shaped molecule that contains three separate domains two Fab fragments (the 'arms' of the molecule) and an Fc fragment (the stalk of the Y shape). The Fab is connected to each Fc by a flexible hinge of 12-19 amino acids, which allows the IgG molecules to bind to a broad range of antigens. This means that it is possible to form a stable protein-to-metal junction between the IgG and the electrodes, which could be particularly useful in the creation of protein transistors.

Fabrication and characterization of a protein transistor

Electron microscopy provided structural information on the gold nanoparticle-immunoglobulin complex (NP-IgG). Because the protein components could not be observed easily, we incorporated CdSe quantum dot-conjugated goat anti-mouse IgG (QD-IgG') to assist in visualizing the NP-IgG complex. Binding complexes of NP-IgG and OD-IgG' were examined using electron microscopy (Fig. 1a). The nanoparticles and quantum dots appear as solid spheres with diameters of 5 nm and 3 nm, respectively, and the protein components appear as a blurred mass. IgG can bind two nanoparticles, forming a NP-IgG-NP dimer. Each complex consists

Biomedical Electronics Translational Research Center, National Chiao Tung University, 1001 University Road, Hsinchu, Taiwan, ROC, ²Department of Materials Science and Engineering, National Chiao Tung University, 1001 University Road, Hsinchu, Taiwan, ROC. e-mail: gstevehuang@mail.nctu.edu.tw

NATURE NANOTECHNOLOGY | VOL 7 | MARCH 2012 | www.nature.com/naturenanotechnology

¹Department of Biological Science and Technology, National Chiao Tung University, 1001 University Road, Hsinchu, Taiwan, China, ²Biomedical Electronics Translational Research Center, National Chiao Tung University, 1001 University Road, Hsinchu, Taiwan, China, ³Department of Materials Science and Engineering, National Chiao Tung University, 1001 University Road, Hsinchu, Taiwan, China, "Institute of Electrical and Control Engineering, National Chiao Tung University 1001 University Road, Hsinchu, Taiwan, China. *e-mail: gstevehuang@mail.nctu.edu.tw





2014-4-3

Table 1 | Heights and widths of the plateaux sequenced by various polymerases.

	Hr* (pA)			r _o (ms)				
	G	Ť	A	c	G	T	A	C
429 [†]	31±0.41	301±0.41	10±0.43	31±0.42	223±24	295±22	203±21	30.2±2.3
T4	32 ± 0.42	32 ± 0.58	31±0.43	33±054	222 ± 25	292±24	201+23	305±23
77	33 ± 0.42	36±0.43	3.1±0.4	37±0.41	231±23	264±23	192±21	28.3+2.2
ы	34±0.8	37±0.68	3.2±072	3.8±0.67	201±2.4	254±24	262±23	332±26

"It, height of the plateaux, og, width of the plateaux.

Symbols: \$29, ph/29 DNA polymeraes; T4, T4 DNA polymeraes; T7, T7 DNA polymeraes; Pol 1.E. od/ DNA polymeraes L













主要质疑有:

- Meni Wanunu, an assistant professor in the departments of physics and chemistry as well as chemical biology at Northeastern University
 - The study's authors say that they applied a voltage across the electrodes of up to 9.0 volts. However, at that high of a voltage, water would become hydrolyzed, generating hydrogen and oxygen gases. If that happened, it would be difficult to measure signals that were a property of the enzyme.
 - In order to not hydrolyze water, he said the applied voltage would have to be below around 1.5 volts.
- Stuart Lindsay, director of the Center for Single Molecule Biophysics at Arizona State University's Biodesign Institute
 - The study describes using superconducting materials at the interface of the protein transistor and probes to reduce signal decay. None of us know of a superconducting material that works at the same temperature as a polymerase.
 - Superconducting materials operate at well below freezing with even so-called high temperature superconductors operating below -135 degrees Celsius. Polymerases, on the other hand, generally operate around room temperature
 - Polymerase incorporation of nucleotides depends on diffusion of the nucleotide into the appropriate binding site. This diffusion is stochastic. However, the authors describe incorporation events occurring at very regular intervals. "None of us understand how this would be achieved."
- Monica Heger tracks trends in next-generation sequencing for research and clinical applications for GenomeWeb's In Sequence and Clinical Sequencing News. E-mail Monica Heger or follow her 2014-4-3



一、为什么要发展DNA测序技术

- 二、第一代测序技术原理与技术
- 三、第二代测序技术的原理与技术
- 四、第三代测序技术原理与技术
- 五、新一代测序技术的产业前景

新一代测序技术的主要指标

- (1) 测序通量:单次运行的碱基数在100万-4000亿个碱 基范围,可达到1万亿个碱基。
- (2) 测序速度:运行一个完整测序流程所需要的时间。
- (3) 测序读长:测序仪所能读出的单个DNA片段序列长 度。
- (4) 测序准确率: 指测序仪读出碱基正确率。
- (5)"测序深度",测序深度是针对样本而言,是指对于 特定的样本新一代测序仪测得的碱基数与该样本实际碱 基数的比率。

Next-Generation Sequencing

TABLE 1: (a) Advantage and mechanism of sequencers. (b) Components and cost of sequencers. (c) Application of sequencers.

		(a)		
Sequencer	454 GS FLX	HiSeq 2000	SOLiDv4	Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Ligation and two-base coding	Dideoxy chain termination
Read length	700 bp	50SE, 50PE, 101PE	50 + 35 bp or 50 + 50 bp	400~900 bp
Accuracy	99.9%*	98%, (100PE)	99.94% *raw data	99.999%
Reads	1 M	3 G	1200~1400 M	
Output data/run	0.7 Gb	600 Gb	120 Gb	1.9~84 Kb
Time/run	24 Hours	3~10 Days	7 Days for SE 14 Days for PE	20 Mins~3 Hours
Advantage	Read length, fast	High throughput	Accuracy	High quality, long read length
Disadvantage	Error rate with polybase more than 6, high cost, low throughput	Short read assembly	Short read assembly	High cost low throughput
		(b)		
Sequencers	454 GS FLX	HiSeq 2000	SOLiDv4	3730xl
Instrument price Instrument \$500,000, \$7000 per run		Instrument \$690,000, \$6000/(30x) human genome	Instrument \$495,000, \$15,000/100 Gb	Instrument \$95,000, about \$4 per 800 bp reaction
CPU	2* Intel Xeon X5675	2* Intel Xeon X5560	8* processor 2.0 GHz	Pentium IV 3.0 GHz
Memory	48 GB	48 GB	16 GB	1 GB
Hard disk	1.1 TB	3 TB	10 TB	280 GB
Automation in library preparation	Yes	Yes	Yes	No
Other required device	REM e system	cBot system	EZ beads system	No
Cost/million bases	\$10	\$0.07	\$0.13	\$2400

Liu et al. Journal of Biomedicine and Biotechnology Volume 2012 (2012), Article ID 251364, 11 pages **2014-4-3** doi:10.1155/2012/251364 新一代基因组测序原理和应用



Cost per Genome.

Adapted from the National Human Genome Research Institute. Collins, et al. NEJM,2013

2014-4-3

Sequencing Cost

Date	Cost per Mb	Cost per Genome
Sep-01	\$5,292.39	\$95,263,072
Sep-02	\$3,413.80	\$61,448,422
Oct-03	\$2,230.98	\$40,157,554
Oct-04	\$1,028.85	\$18,519,312
Oct-05	\$766.73	\$13,801,124
Oct-06	\$581.92	\$10,474,556
Oct-07	\$397.09	\$7,147,571
Oct-08	\$3.81	\$342,502
Oct-09	\$0.78	\$70,333
Oct-10	\$0.32	\$29,092
Oct-11	\$0.09	\$7,743
Oct-12	\$0.07	\$6,618
Jan-13	\$0.06	\$5,671

Source - NHGRI: http://www.蘇nofte基网組測标原理和茲用

(1) NGS正在改变现代生命科学的研究方法

- > 系统地获取生命的遗传信息,从整体上认识细胞内极为 复杂的基因调控过程。如,RNA-Seq、ChIP-Seq和 MeDIP-Seq等基于测序技术的新技术层出不穷;
- ➢ 产生大量的生物分子数据,催生基于数据驱动的生物学研究模式(data driven science)。







OMMI Phenotypes for which genetic is known 2007-2013: 2,048->3,834 Koboldt et al. Cell, 2013

Until 2011.9 more than 1000 GWAS papers;
Report about 5000loci and related to 200 diseases or traits



开展个人基因组的必要性: 大量群体分析研究将治病因素指向罕见突变



New SNPs

Pilot project have identified 15M SNPs including 8.5M novel SNPs (right)
Most novel SNPs are population specific



1000 Genome

A Deep Catalog of Human Genetic Variation

•Evidences on higher-impact of rarer variations based on 200 Danish exomes (Y. Li, et.al. 2010. *Nature Genetics*)

•There is a stronger excess of low frequency deleterious amino acid changing mutations in the human genome than previously thought.

•This indicates that heritable diseases might be affected by many rare mutations rather than a few common mutations.

(2) 基因测序与个体化医疗



For about \$6,000, this test uncovers all the genetic mutations that lead to a person's tumor. It's helping to usher in a new era of "personalized medicine" where doctors choose cancer treatments based on genetic knowledge.

Jobs spent some \$100,000 to have this kind of test done, according to Walter Isaacson's biography, and in the end, it obviously didn't save him.

"我相信我会成为这家公司治愈的首批癌症患者之一,抑或是最后一批因为癌症而死亡的患者之一。" 耗费 10 万美元进行了一次癌症全基因组测序的治疗尝试后,乔布斯对基础医学公司(Foundation Medicine)带来的价值坚信不疑。

But he believed deeply in the value of the attempt, saying "I'm either going to be one of the first to be able to outrun a cancer like this, or I'm going to be one of the last to die from it," reports Antonio Regalado, MIT Technology Review.









好莱坞女星安吉丽娜·朱 莉高调宣布切乳腺

美媒:不要接受安吉丽娜·朱莉的乳腺癌建议

2013年12月26日 14:23:57 来源:新华国际 🛨 分享到: 🚺 🔁 🚮 😰 🖊 🦲



新华网消息 据海外网报道,《新闻周刊》(Newsweek)12月24日发表题为《当谈及乳腺癌时,不要 接受安吉丽娜·朱莉的建议》的文章表示,据癌症预防方面的美国专家表示,大多数的女性都不应该接 受朱莉所推荐的乳腺癌基因测试。



2014-4-3



The NEW ENGLAND JOURNAL of MEDICINE

HOME	ARTICLES & MULTIMEDIA *	ISSUES *	SPECIALTIES & TOPICS *	FOR AUTHORS *	CME »



Perspective

First FDA Authorization for Next-Generation Sequencer

Francis S. Collins, M.D., Ph.D., and Margaret A. Hamburg, M.D. November 19, 2013 | DOI: 10.1056/NEJMp1314561

Francis Collins (Director of NIH) :

Landmark move that will help realize the promise of personalized medicine

(**3**) 基因测序与大数据: 创造新的产业群-生物信息产业

DNA测序技术正在催生一个新的产业集群, 未来具有市场容量巨大

- 全球个体基因组测序的市场容量为6万亿美元,我国可达到6万亿人民币(1000美元个体测序);
- 与疾病相关测序的市场容量将要增加数千倍;
- 带动巨大的生物信息服务市场需求。

Heritability

Psychological behaviors







2014-4-3





2014-4-3

单胺氧化酶 MAO-A

MAOA基因是编码神经递质代谢酶,降解去甲肾上腺素 (NE), 5-羟色胺(5-HT), 以及多巴胺 (DA)。





基因与环境对于三个孩子的智力影响











2014-4-3

新一代基因组测序原理和应用

107





Illumina MyGenome App for iPad[®]

Explore a real human genome. Learn about important genetic variations.

🕑 DOWNLOAD NOW

0

BaseSpace

Securely and easily analyze, store, and share MiSeq data with anyone, anywhere, anytime

典型文献

Detection of redundant fusion transcripts as biomarkers or targets in breast cancer

基因组服务网络

Whole-genome sequencing services without compromise

Agrigenomics

Discover, develop, and de genetic variation in any spe

23 and Me





Personal Genome Service™

Get to know your DNA. All it takes is a little bit of spit.

Here's what you do:











1. Order a kit from our online store.

 Claim your kit, spit into the tube, and send it to the lab.

 Our CLIA-certified lab analyzes your DNA in 6-8 weeks.

es 4. Log in and start exploring you genome.



Find relatives across continents or across the street.

Ever wonder who's related to you? You'll likely discover dozens or even hundreds of people who share DNA and ancestors. The matches you'll get can range from close family to distant cousins.



Build your family tree and enhance your experience.

Preserve your family's history by highlighting names, dates, events, and more. See if you can gather data from newly found relatives to fill in the missing branches of your tree. If you already use a family tree service, you can easily import your tree.



Share your knowledge. Watch it grow.

Make the most of your results by inspiring family members to join 23andMe. The more family members you have in the database, the more comprehensive the picture of your ancestry will be.



Best Inventions of 2008

From a genetic testing service to an invisibility cloak to an ingenious public bike system to the world's first moving skyscraper — here are TIME's picks for the top innovations of 2008

Select a Section



CORBIS

The Best Inventions of the Year

-By Jeremy Caplan, Kristina Dell, Andrea Dorfman, Laura Fitzpatrick, Justin Fox, Sean Gregory, Lev Grossman, Barbara Kiviat, Jeffrey Kluger, Richard Lacayo, Michael Lemonick, Lisa McLaughlin, Jay Newton-Small, Alice Park, Mark Thompson, Bryan Walsh and Rebecca Winters Keegan More »









羽

"Here's my sequence..."

基因伦理学问题





The PGP 10: The first 10 volunteers in the Personal Genome Project are currently having the coding regions of their genomes sequenced; a small piece of sequence is shown for those whose data is posted online. The sequence data will be stored in a public database, along with the volunteers' medical records and other information, such as their facial morphology (as measured by the forehead tapes). Scientists will use the database, which is expected eventually to include 100,000 people, to search for links between genes and diseases or other characteristics.

Credit: courtesy of personalgenomes.org

2014-4-3



新一代基因组测序原理和应用 personal genetics education

个体化基因组信息是否会影响人类统 一伦理或司法标准?



Guilty or not?

Division of Forensic Psychiatry

Gainesville, FL 32610-0256 Phone: 352-265-3284 Fax: 352-265-3285

Date of Evaluation: March 31, 2003 Date Report Typed: April 16, 2003 Date Report Revised/by:

PSYCHIATRIC EVALUATION

Date of Final Report:

NAME: Igor Stravinsky DATE OF BIRTH: May 26, 1982 AGE AT EVALUATION: 20 DATE OF EVALUATION: March 31, 2003 LOCATION OF EVALUATION: Marion County Jail, Ocala, FL. CASE NAME: State of Florida v Igor Stravinsky CASE NO: 02-540-CF-A-Z LEGAL CHARGES: Kidmappingport and a Battery and 1st Degree Murder EVALUATORS: Erable Management and Wage Composed MD

INTROD CTION: Psychiatric evaluation of this 20 -year-old male way requested by Mark Kney, a timey, at was specificantly requested that this evaluation address his competency to proceed to using and the defendent formula the time of the alleged events.

CONFIDENTIALITY: The nature, purpose and non-confidential aspects of this evaluation were explained to Mr. Stravinsky. He was informed we would be preparing a report based on our evaluation and that we might be called upon to testify about our evaluation. He acknowledged that he understood these elements and agreed to proceed.

SOURCES OF DATA

- 1. 105 minute clinical interview with Igor Stravinsky on March 31, 2003.
- 2. Medical records from Charles Shuman, M.D., University of Florida, College of Medicine, Department of Psychiatry
- 3. Records from the Children's' Evaluation and Rehabilitation Clinic of the

新一代基因组测序原理<mark>和应</mark>用