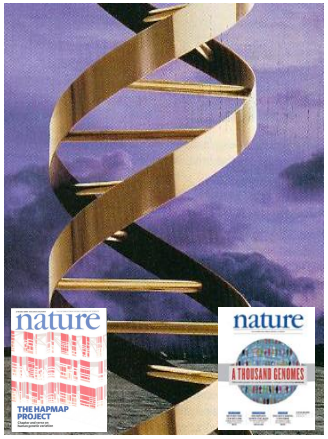
HGP, Genomics

& its Relevance to Medicine

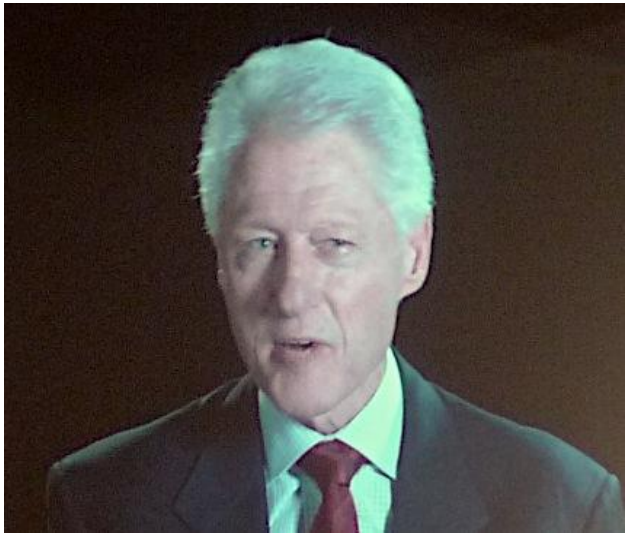


Huanming Yang, Ph.D.

BGI-China, Shenzhen

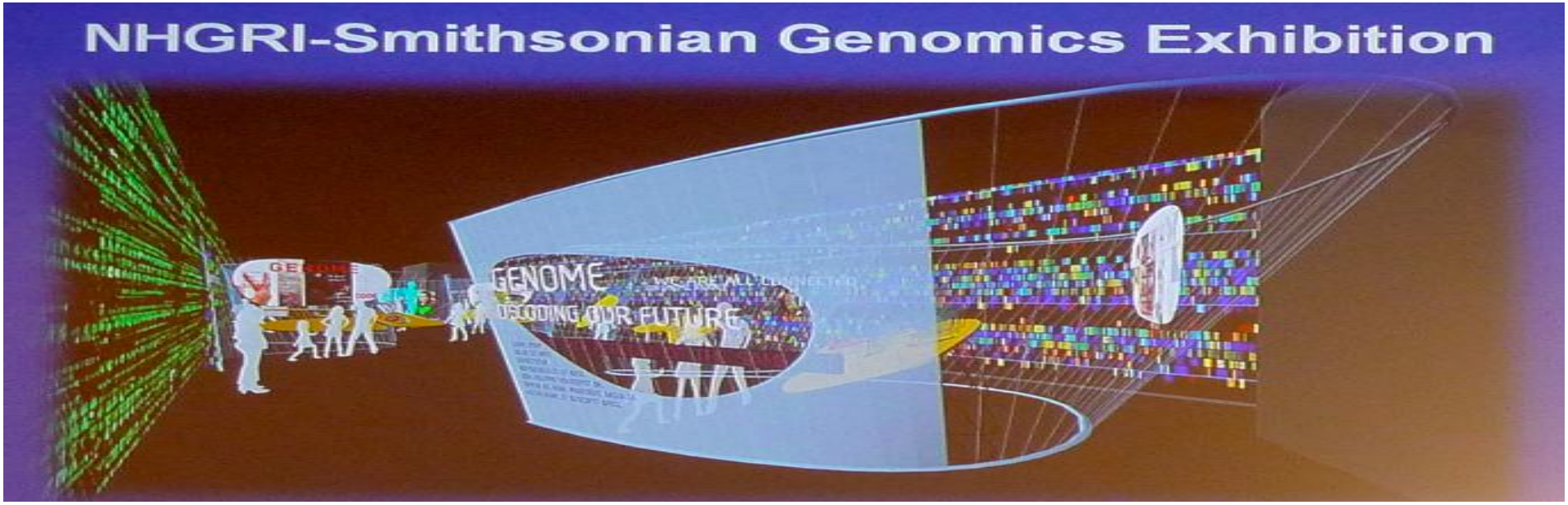
Outline:

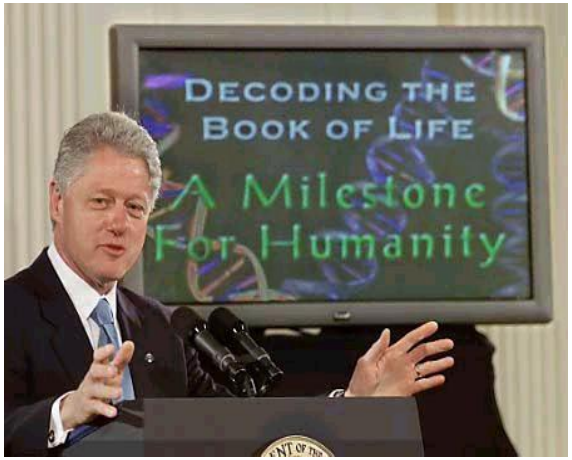
1. A later comer of a revolution
2. Two pillars of genomics
3. Three impacts of the HGP
4. Four applications to medicine
5. Five techs changing the world,

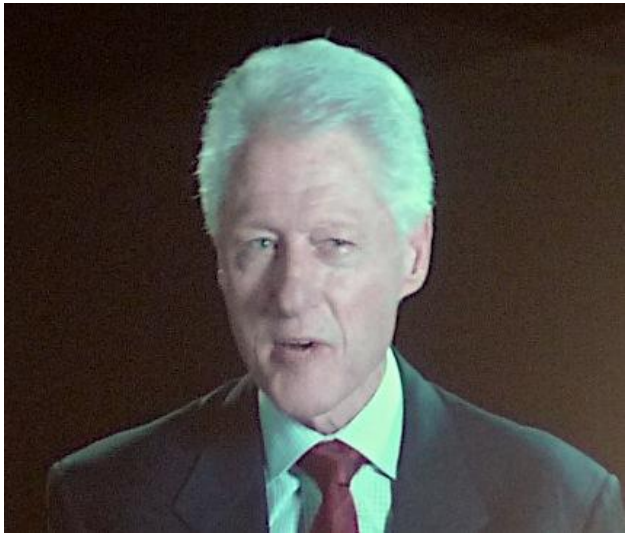


GENOME
UNLOCKING
LIFE'S
CODE

June 14, 2013 at Smithsonian







White Hourse Science Event **10:19 am EST / June 26 , 2000**

~15 Years Ago

THE PRESIDENT: Good morning. I want to first of all acknowledge Prime Minister Blair, who will join us by satellite in just a moment from London. I want to welcome the Ambassadors from the United Kingdom, Japan, Germany, France. And I would like to acknowledge the contributions not only that their scientists, but also scientists from China, made to the vast international consortium that is the Human Genome Project.

克林顿总统在“白宫科学庆典上的讲话（2000年6月26日，白宫东厅）：

“解读生命的天书，人类进步的里程碑”

……我还要感谢他们国家（美、日、德、法）的科学家，不仅是他们国家的，还有中国的科学家，对广泛国际合作的人类基因组计划所做的贡献！

江泽民就完成人类基因组“框架图”发表讲话

我向我国参与这一工作并作出杰出贡献的科学家和技术人员表示衷心的感谢……

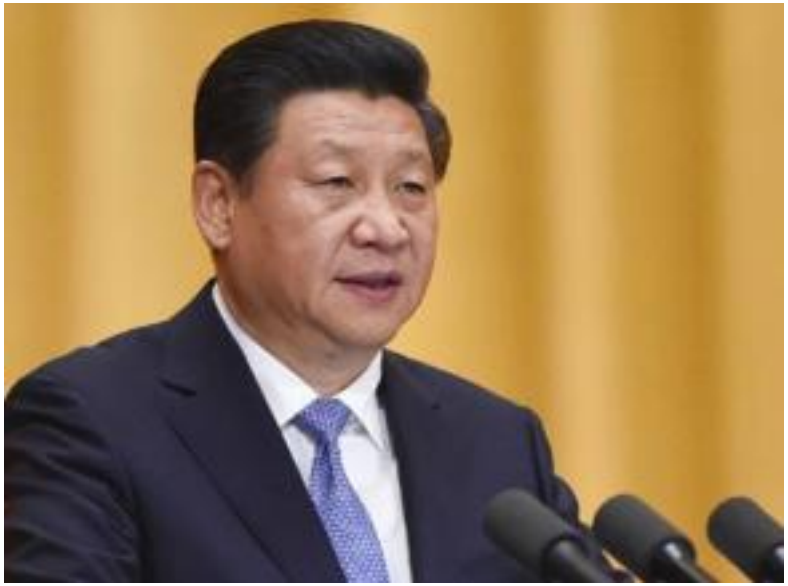
President lauds new map of human life

[Xinhua 06/29/2000 Beijing] In a spee
the recent completion of the first working draft of t
Project “a great scientific project ”in the scientific
development of life sciences，medicine and pharm
milestone in the overall process of the Human Ge
teamwork by scientists all over the
the common heritage of mankind
all human beings, Jiang said. Jiang con
thanked the participating Chinese scientists for th
make further contributions to the completion of
research as well.

2001年8月28日江泽民主席接见人类基因组计划负责人

“今天我们特地来看望大家。同志们曾在完成国际人类基因组计划‘中国卷’和绘制水稻基因组精细图等方面作出了贡献，现在，又在防治非典型肺炎中取得重大科技成果。我代表党中央、国务院，向同志们表示衷心的感谢和崇高的敬意！”

胡锦涛　2003年4月20日



“两弹一星”、… 人工合成牛胰岛素等成就，… 人类基因组测序等基础科学突破，超级杂交水稻、… 高速铁路、航空母舰等工程技术成果，为我国经济社会发展提供了坚强支撑，为国防安全作出了历史性贡献，也为我国作为一个有世界影响的大国奠定了重要基础。

在中国科学院第十七次院士大会、
中国工程院第十二次院士大会上的讲话
（2014年6月9日）

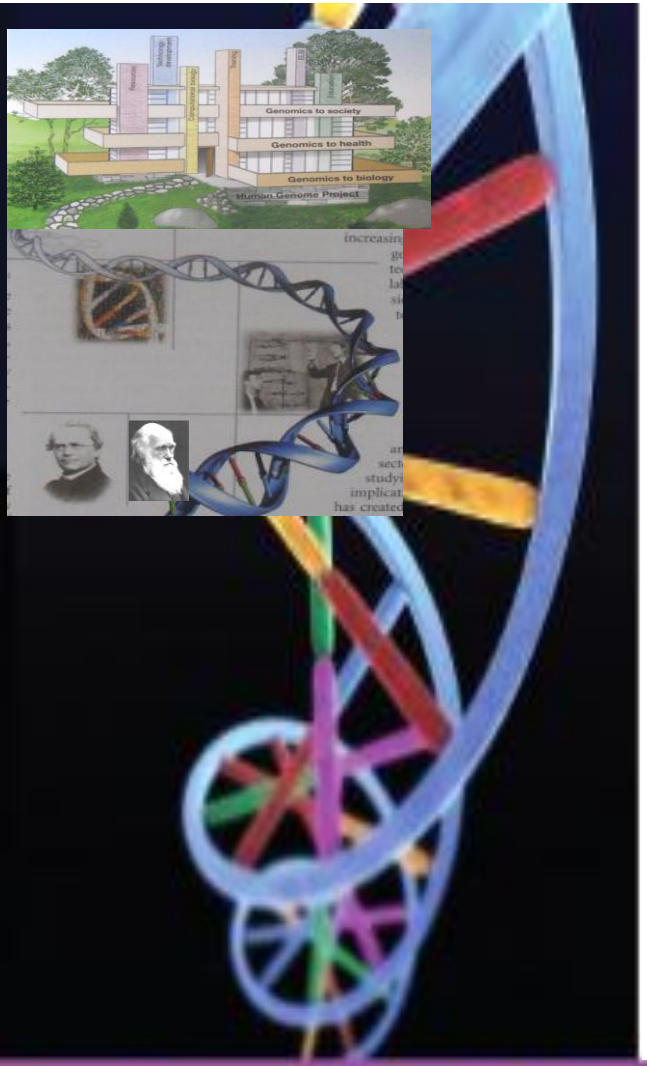
Three “BIG” projects in the 20th century

二十世纪的
“三大计划”

1945 USA
China1964

1969 USA
? China

2000 USA, UK, J, F, G,
& China



Economic Impact of the Human Genome Project

38 亿美元投入

How a $3.8 billion investment drove $796 billion in economic impact

created 310,000 jobs and launched the genomic revolution

(7,960 亿美元回报, 31 万就业)

Prepared by Battelle Technology Partnership Practice

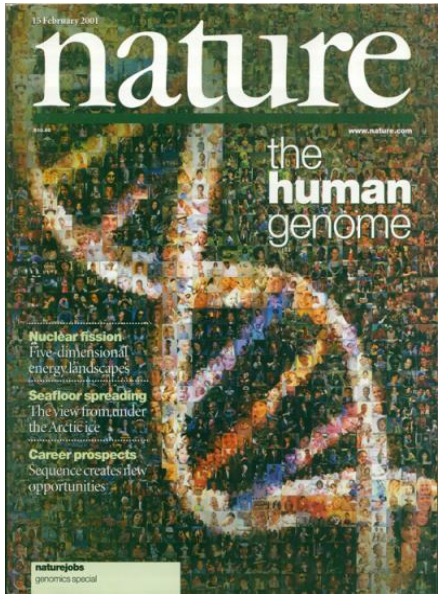
May 2011

HGP: *The 2nd Revolution in Life Science*

After giving a brief history of the advances in life sciences from the "first revolution," the discovery of the structure of DNA, through *the "second revolution," the sequencing of the human genome,* he asked, "What is the next revolution in life sciences?"

"生物学已经经历了两次革命，第一次革命是沃森和克里克发现 DNA 的双螺旋结构，*第二次革命是"人类基因组计划"*；生物学目前正经历第三次革命，即生物学在分子层面与物理学、工程学等领域的融合。"

Phillip Sharp (美国科学促进会（AAAS）主席、诺贝尔奖获得者), AAAS 2014 年会, 芝加哥, 13 Feb, 2014



"China has become the latest contributor to the worldwide sequencing effort alongside France, Germany, Japan, UK and USA."

The International Human Genome Sequencing Consortium

1 Sept. 1999

BGI
A later comer

Sept. 9, 2014

The 15th anniversary of BGI

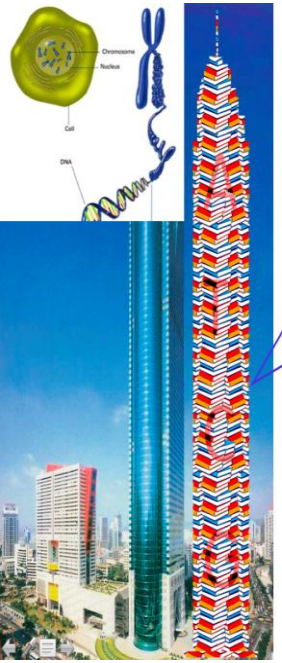
(9:09:09, 9/9/99)

BGI was born for the Chinese Chapter of the HGP

15 Years Ago

"The building had a nice double helix on the brick façade, but that was the only indication that this was a genome center as opposed to an empty warehouse. I really wondered if they could get the support to become an internationally competitive group".

>15 Years Ago

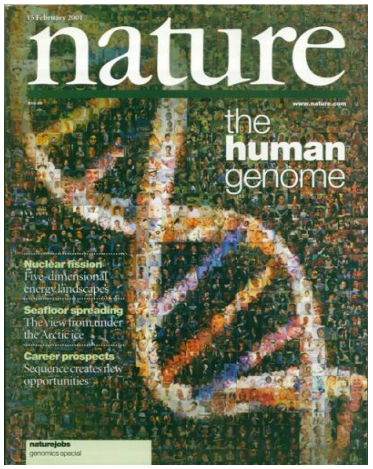


" With a diploid genome of around s
with no apparent regularities in base
methods in sight for breaking the DN
smaller pieces that can be amplified and sequenced, we shall never know the complete base sequence of the human genome. ..... Perhaps I need to modify this somewhat, never is a dangerous word considering the rapid advances in modern science. But I can promise you, it will not be known for the next 300 years!"

"Mission: Impossible!"

"不可完成的任务"

HGP – A revolution in life sciences



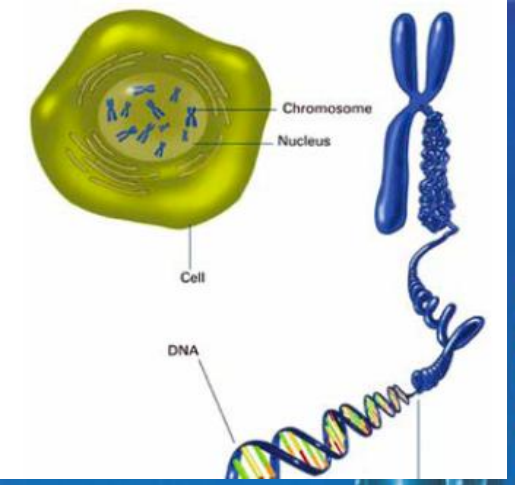
6 国 16 中心对 HGP 的贡献

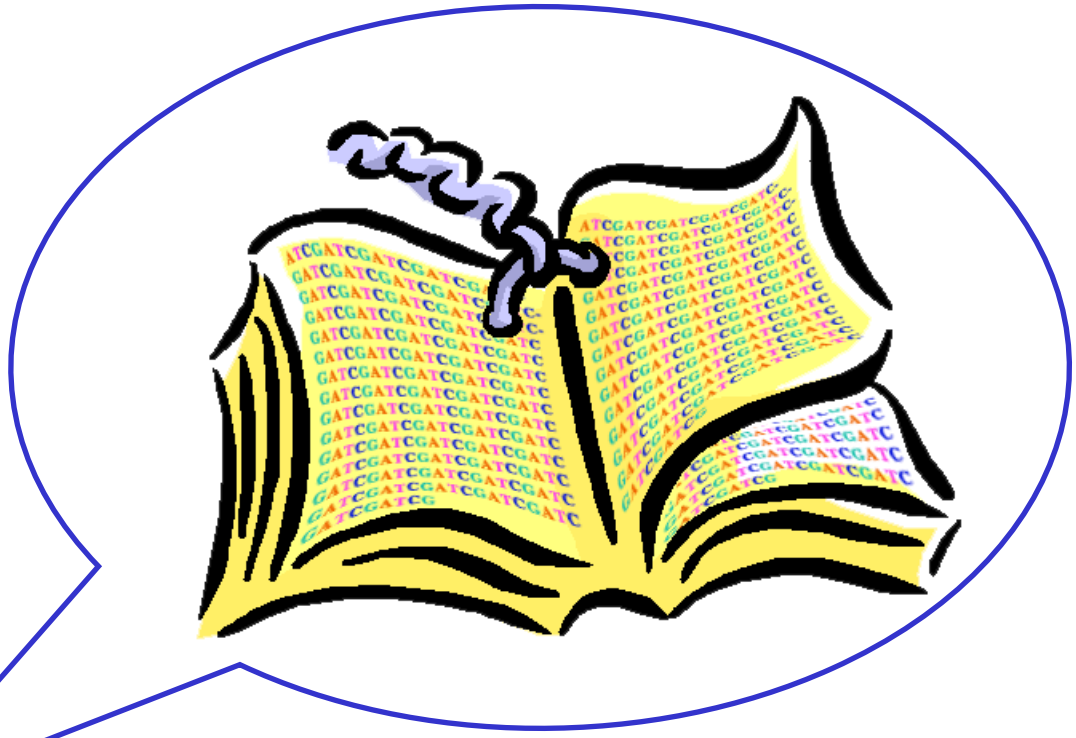
国别	中心	包干区域 (Mb)	贡献率 (%)	国家贡献率 (%)	
美国	Wash U + MIT	1050	~35.0	54.0	USA
	JGI – DOE	310	~10.3		
	Baylor	230	~8.0		
	GTC	50	1.7		
	Stanford	15	0.5		
	U Wash (H)	25	0.8		
	U Wash (O)	25	0.8		
英国	Sanger	1000	~33.0	33.0	UK
日本	Riken	180	6.0	7.0	Japan
	Keio	23	0.8		
法国	Genoscope	85	2.8	2.8	France
德国	IMB (Jena)	50	~1.7	2.2	Germany
	MPIMG	6.9	0.23		
	GBF (Gesellschaft)	6	0.2		
中国	Beijing	30	1.0	1.0	China

We have had it done!

We have had it done together

through the vast international collaboration!





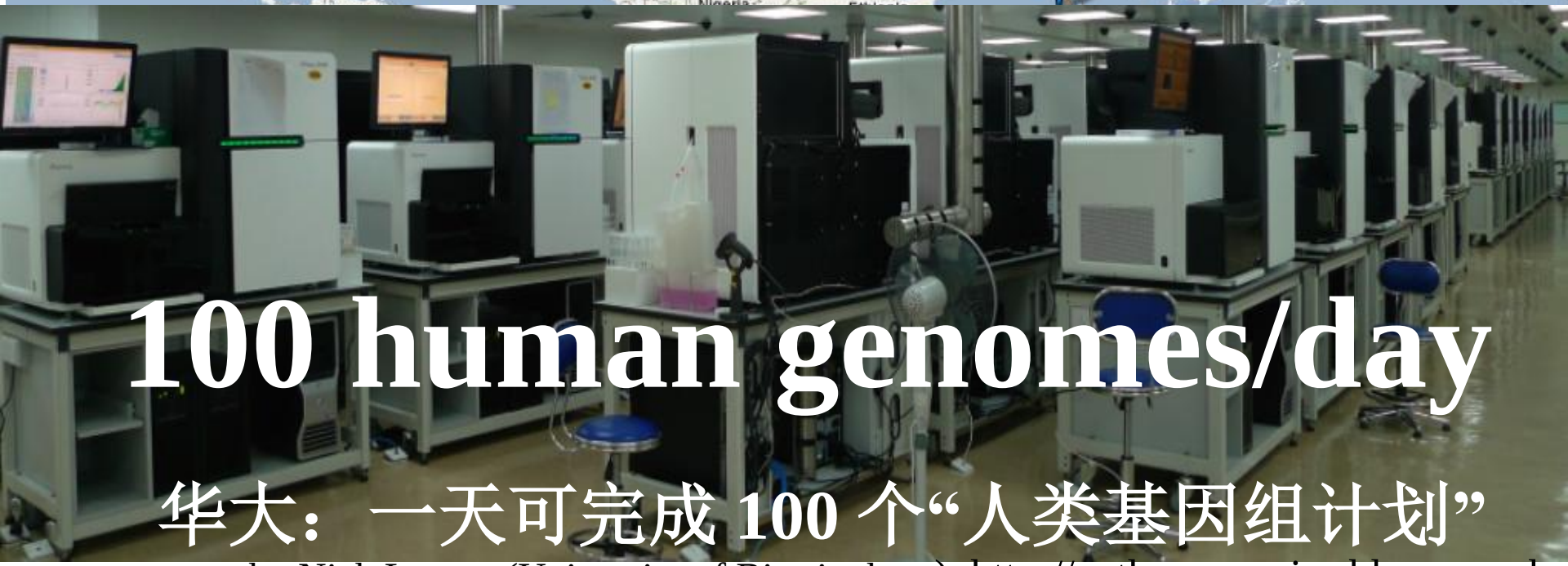
Only an individual genome

Too expensive (US$ 3b）

Too slow （13 yrs）

（～3000 staffs in 16 centers from 6 countries）

15 Years Later

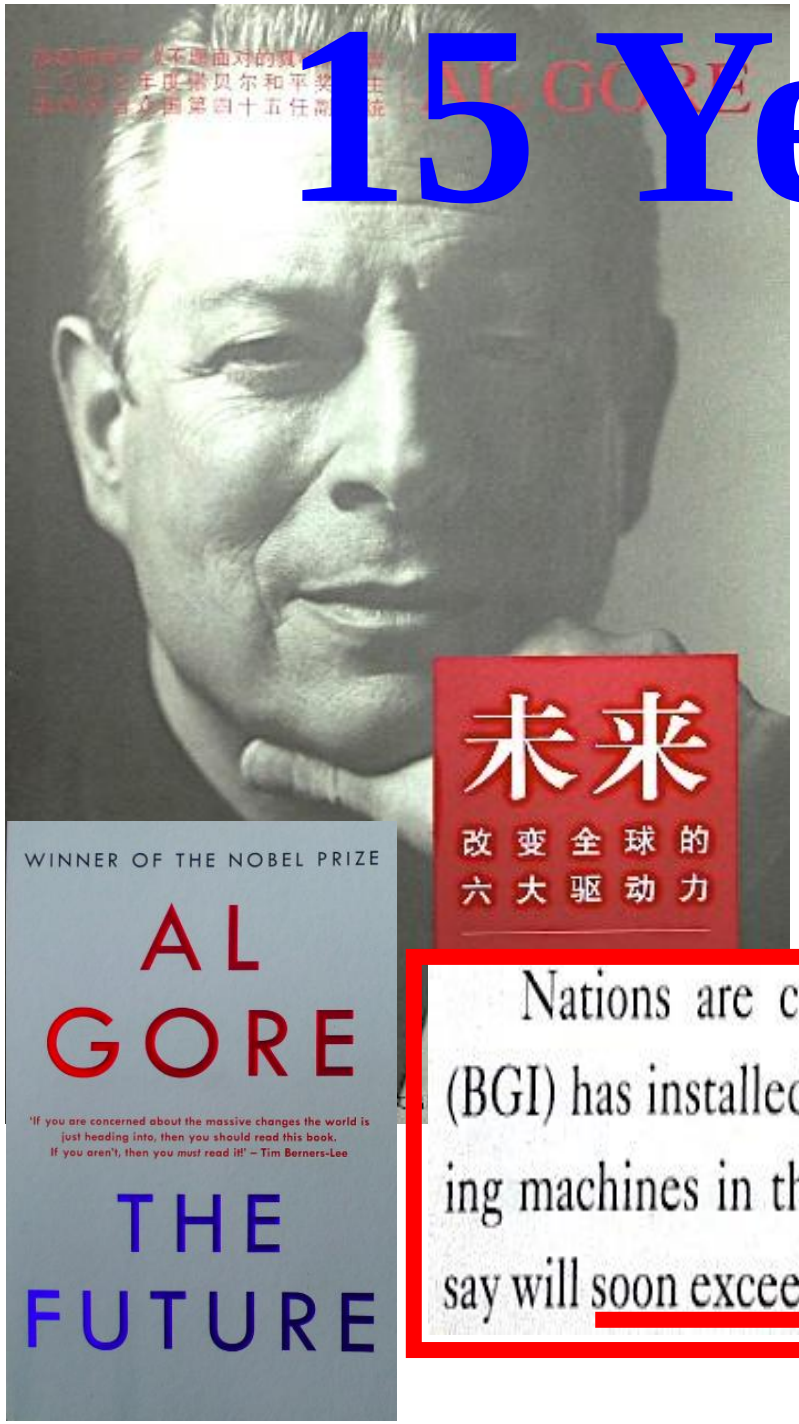


100 human genomes/day

华大：一天可完成 100 个“人类基因组计划”

by Nick Loman (University of Birmingham). http://pathogenomics.bham.ac.uk

15 Years Later



As Jun Wang, executive director of the Beijing Genomics Institute, put it, there is a strong network effect . . . the health profile and personal genetic information of one individual will, to a certain extent, provide clues to better understand others' genomes and their medical implications. In this sense, a personal genome is not only for one, but also for all humanity."

As noted earlier, China appears determined to become the world's superpower in the application of genetic and life science analysis. The Beijing Genomics Institute (BGI), which is leading China's commitment to genomic analysis, has already completed the full genomes of fifty animal and plant species, including silk worms, pandas, honeybees, rice, soybeans, and others—along with more than 1,000 species of bacteria.

Nations are competitive too. China's Beijing Genomic Institute (BGI) has installed 167 of the world's most powerful genomic sequencing machines in their Hong Kong and Shenzhen facilities that experts say will soon exceed the sequencing capacity of the entire United States.

15 Years Later

BGI：世界最大基因组和生物信息学中心

排名	研究机构	测序能力指数*	国家
1	BGI	179	中国
2	Broad Institute	114	美国
3	The Genome Center at Washington University	62	美国
4	Wellcome Trust Sanger Institute	35	英国
5	Canada's Michael Smith Genome Sciences Centre	27	加拿大
6	DOE Joint Genome Institute	21	美国
7	Yale Center for Genome Analysis	20	美国
8	Human Genome Sequencing Centre, Baylor College of Medicine	18	美国

* Sequencing Capacity Index

15 Years Later

- BGI (formerly Beijing Genomics Institute, a spin out of the CAS Beijing Institute of Genomics) has gone from accounting for around 1 per cent of the world's gene sequencing capabilities in 1999 to almost 50 per cent today. BGI works with more than 10,000 collaborators from universities and industry around the world.[15]

BGI

Founded 1999 • Headquarters: Shenzhen, China

Sequencing more genomes than anyone else and becoming a worldwide provider of genome services.

"BGI … went from accounting for about 1% of the genomics community's sequencing capacity at that time, to "more like '50%' today,…"

George Church, Harvard University

"…华大基因支持了HGP的1%，而到今天，华大对世界基因测序的贡献超出了50%。"

15 Years Later

MOST INNOVATIVE COMPANIES 2013

THE WORLD'S TOP 10 MOST INNOVATIVE COMPANIES IN CHINA

BY: FAST COMPANY STAFF

4_BGI

MOST INNOVATIVE COMPANIES

OUR ANNUAL GUIDE TO THE STATE OF INNOVATION IN OUR ECONOMY, FEATURING THE BUSINESSES WHOSE INNOVATIONS ARE HAVING THE GREATEST IMPACTS ACROSS THEIR INDUSTRIES AND OUR CULTURE AS A WHOLE. FOLLOW THE CONVERSATION ON TWITTER #FCMOSTINNOVATIVE

For advancing both the science and the business of genetic sequencing. Just before 2012 turned into 2013, the Committee on Foreign Investment in the United States approved the sale of Silicon Valley-based Complete Genomics--a struggling DNA-sequencing company, one of the first to offer sequencing as a service--to the model-efficient, Shenzhen-based genetic research institution BGI. Mathematicians are just as important as biologists in next-generation sequencing, and BGI has a staff of 1,000. Its Million Human, Plant and Animal, and Micro-Ecosystem Genome projects are ongoing.

中国的“世界十强”

MOST INNOVATIVE COMPANIES 2014

China

Industry TOP TENS

01. Xiaomi
02. BGI
03. China's Brands
04. Haier
05. Tencent
06. Geak
07. Phantom
08. Baidu
09. YY
10. Cootek

15 Years Later

"2013 颠覆世界的五十强"

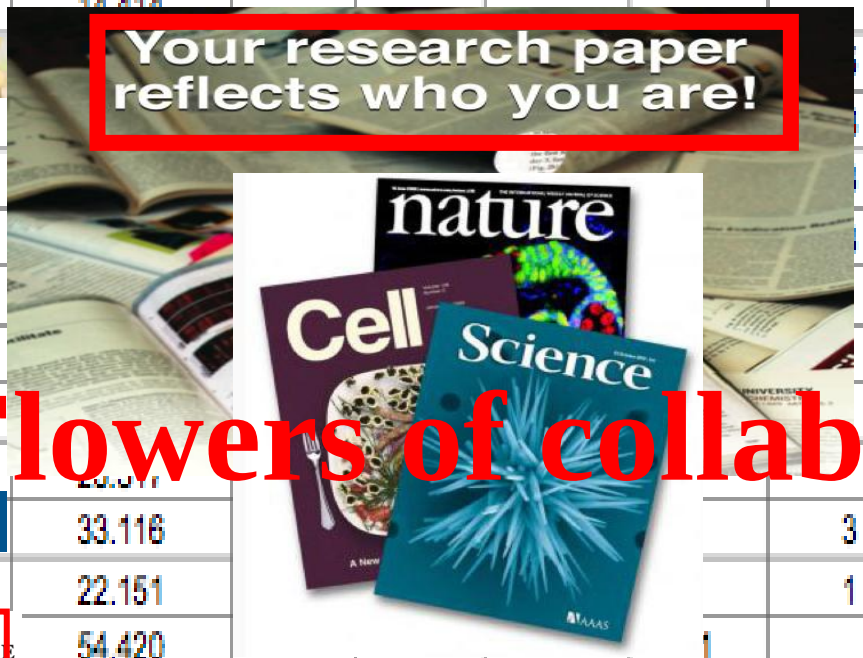
"2014全球最创新的五十强"

15 Years Later

Publications by BGI and its Partners

(Nov. 2008 – July 2014)

Journal	Impact Factor 2013	2008	2009	2010	2011	2012	2013	2014	Total
Science	31.477		1	3	3	4	2	3	16
Science Translational Medicine	14.414						1		1
Nature	[illegible]					[illegible]	4	4	25
Nature Genetics	[illegible]					[illegible]	10	6	38
Nature Communications	[illegible]					[illegible]	14	9	26
Nature Biotechnology	[illegible]					[illegible]	3	1	15
Nature Reviews Drug Discovery	[illegible]								1
Nature Methods	[illegible]					[illegible]	1	2	4
Nature Reviews Genetics	[illegible]								1
Nature Reviews Microbiology	[illegible]						1		1
Cell	33.116					3		1	4
Cell Stem Cell	22.151					1			1
NEJM	54.420				1				1
Sum		1	6	12	21	32	36	26	134



Flowers of collaboration!

nature PUBLISHING INDEX CHINA

Research Articles | Reviews | All

2013

Institution

1. Chinese Academy
2. University of Science a
3. Tsinghua University, Ch
4. Peking University, Chin
5. BGI, China
6. Nanjing University, Chir
7. Zhejiang University, Chi
8. Shanghai Jiao Tong Uni
9. Fudan University, China
10. The University of Hong Kong (HKU), China

CHINA TOP TEN

China has 65 institutions in the Asia-Pacific top 200

ASIA-PACIFIC RANK 2013	INSTITUTION	2013 NATIONAL RANK	2013 CORRECTED COUNT	2013 ARTICLES	2012 NATIONAL RANK	2012 CORRECTED COUNT	2012 ARTICLES	2009–13 NATIONAL RANK	2009–13 CORRECTED COUNT	2009–13 ARTICLES
1	Chinese Academy of Sciences (CAS)	1	63.15	165	1	36.38	93	1	147.70	399
9	University of Science and Technology of China (USTC)	2	15.11	37	2	9.46	17	2	39.67	87
10	Tsinghua University	3	13.83	39	3	8.99	34	3	39.39	116
14	Peking University	4	11.16	48	4	8.72	35	4	33.15	131
17	BGI	5	10.50	32	5	6.62	30	5	23.86	73
21	Nanjing University	6	7.89	19	6	6.27	20	6	18.87	101
35	Fudan University	7	5.72	25	7	4.70	15	7	17.79	49
36	Zhejiang University	8	5.70	25				8	15.64	64
37	Shanghai Jiao Tong University (SJTU)	9	5.63	35	9	3.90	13	9	15.01	57
39	The University of Hong Kong (HKU)	10	4.99	18	10	2.92	14	10	14.67	68

2010

RANK	INSTITUTION
1	Chinese Academy of Sciences
2	Tsinghua University
3	University of Science and Technology of China
4	BGI Shenzhen
5	Peking University
6	Nanjing University
7	The University of Hong Kong

2012

se Academy of Sciences (CAS), China
2. Peking University, China
3. University of Science and Technology of China, China
4. Tsinghua University, China
5. Beijing Genomics Institute (BGI), Shenzhen, China
6. Shanghai Jiao Tong Univ sity (SJTU), China
7. Fuda Un rsity, Chi
8. Hong on University o nd Technology HKU T), hina
9. Zhejiang University, China
10. National Institute of Biological Sciences, Beijing (NIBS, Beijing), China

中国前10名

ASIA-PACIFIC TOP 20 INSTITUTIONS By Nature

Research Articles | Reviews | All

2013

Institution
1. Chinese Academy of Sciences (CAS), China
2. The University of Tokyo, Japan
3. Kyoto University, Japan
4. National University of Singapore (NUS), Singapore
5. RIKEN, Japan
6. The University of Melbourne, Australia
7. Osaka University, Japan
8. Tohoku University, Japan
9. University of Science and Technology of China (USTC), China
10. Monash University, Australia
11. Australian National University (ANU), Australia
12. Tsinghua University, China
13. The University of Queensland (UQ), Australia
14. Nanyang Technological University (NTU), Singapore
15. Peking University, China
16. Nagoya University, Japan
17. Korea Advanced Institute of Science and Technology (KAIST)
18. BGI, China
19. Agency for Science, Technology and Research (A*STAR), Singapore
20. Hokkaido University, Japan

2012

RANK	INSTITUTION
1	The University of Tokyo
2	Chinese Academy of Sciences (CAS)
3	Kyoto University
4	RIKEN
5	Osaka University
6	The University of Melbourne
7	Nagoya University
8	University of Science & Technology of China
9	National University of Singapore (NUS)
10	Tohoku University
11	Tsinghua University
12	Peking University
13	Australian National University (ANU)
14	National Institute of Advanced Industrial Science & Technology (AIST)
15	The University of Queensland
16	Agency for Science, Technology & Research (A*STAR)
17	Shanghai Jiao Tong University (SJTU)
18	BGI
19	Kyushu University
20	Hokkaido University

2010

RANK	INSTITUTION
1	The Univ
2	RIKEN
3	Kyoto Un
4	Chinese
5	Osaka U
6	Tohoku
7	Tsinghua
8	Nagoya
9	Agency
10	Seoul Na
11	Korea Advanced Institute of Science and Technology
12	National Institute of Advanced Industrial Science and Technology
13	University of
14	The University of Queensland
15	National University of Singapore
16	University of Science and Technology of China
17	BGI
18	Peking University
19	Monash University
20	Nanjing University

亚太前 20 名

Global Top 100

BGI Ranked 87 of Global Top 100 in 2013

87.	BGI, China	10.76	34
88.	Nagoya University, Japan	10.68	33
89.	University of Copenhagen, Denmark	10.64	77
90.	University of Massachusetts Medical School (UMMS), USA	10.6[illegible]	30
91.	The University of Queensland (UQ), Australia	10.6[illegible]	50
92.	University of California Irvine (UCI), USA	10.6[illegible]	38
93.	Howard Hughes Medical Institute (HHMI), USA		
94.	Karolinska Institute, Sweden		
95.	University of Groningen, Netherlands		
96.	The Institute of Photonic Sciences (ICFO), Spa[illegible]		
97.	The Pennsylvania State University, USA		
98.	+ National Research Council (CNR), Italy		
99.	Aarhus University (AU), Denmark		
100.	University of Pittsburgh, USA		

全球前 100 名

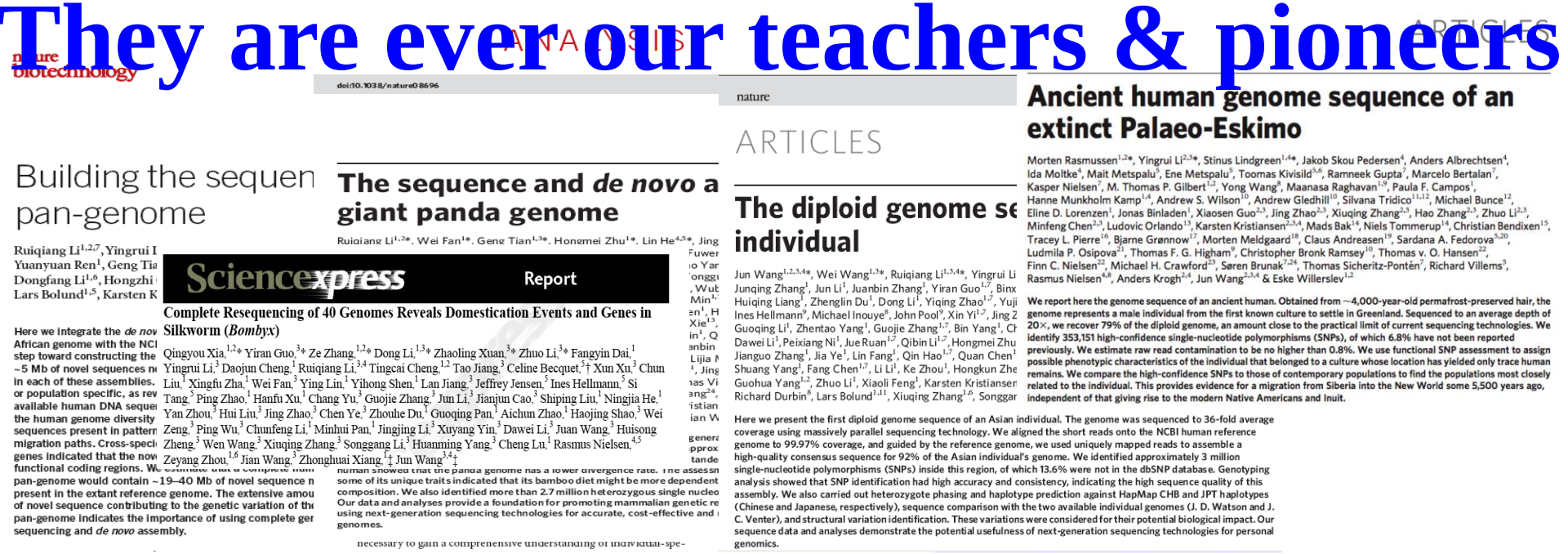
Global Top 200 (119) (2012)

119	BGI	China	6.27
120	University of Southampton	UK	6.25
121	Oregon Health & Science University (OHSU)	United States	6.21
122	University of Bonn	Germany	6.20
123	University of Tübingen	Germany	6.19
124	University of Basel	Switzerland	6.19
125	University of Hamburg	Germany	6.12
126	Catholic University of Leuven	Belgium	6.11
127	University of Virginia	United States	6.06
128	Aarhus University (AU)	Denmark	6.05
129	Mayo Clinic	United States	5.99
130	University of Dundee	UK	5.88
131	Wellcome Trust Sanger Institute	UK	5.87
132	University of London - King's College London	UK	5.86
133	Albert Einstein College of Medicine of Yeshiva University	United States	5.84
134	Kyushu University	Japan	5.81
135	University of Nottingham	UK	5.80
136	Karolinska Institute	Sweden	5.68

Never forget:

WE = All collaborators

They are ever our teachers & pioneers



nature biotechnology

Building the sequen
pan-genome

Ruiqiang Li[1,2,7], Yingrui I
Yuanyuan Ren[1], Geng Tia
Dongfang Li[1,6], Hongzhi
Lars Bolund[1,5], Karsten K

Here we integrate the *de nov*
African genome with the NCI
step toward constructing the
~5 Mb of novel sequences n
in each of these assemblies.
or population specific, as rev
available human DNA sequer
the human genome diversity
sequences present in pattern
migration paths. Cross-speci
genes indicated that the nov
functional coding regions. We
pan-genome would contain ~19–40 Mb of novel sequence n
present in the extant reference genome. The extensive amou
of novel sequence contributing to the genetic variation of the
pan-genome indicates the importance of using complete ger
sequencing and *de novo* assembly.

doi:10.1038/nature08696

ANALYSIS

**The sequence and *de novo* a
giant panda genome**

Ruiqiang Li[1,2]*, Wei Fan[1]*, Geng Tian[1,3]*, Hongmei Zhu[1]*, Lin He[4,5]*, Jing

Sciencexpress Report

Complete Resequencing of 40 Genomes Reveals Domestication Events and Genes in Silkworm (*Bombyx*)

Qingyou Xia,[1,2]* Yiran Guo,[3]* Ze Zhang,[1,2]* Dong Li,[1,3]* Zhaoling Xuan,[3]* Zhuo Li,[3]* Fangyin Dai,[1] Yingrui Li,[3] Daojun Cheng,[1] Ruiqiang Li,[3,4] Tingcai Cheng,[1,2] Tao Jiang,[3] Celine Becquet,[5]† Xun Xu,[3] Chun Liu,[1] Xingfu Zha,[1] Wei Fan,[3] Ying Lin,[1] Yihong Shen,[1] Lan Jiang,[3] Jeffrey Jensen,[5] Ines Hellmann,[5] Si Tang,[5] Ping Zhao,[1] Hanfu Xu,[1] Chang Yu,[3] Guojie Zhang,[3] Jun Li,[3] Jianjun Cao,[3] Shiping Liu,[1] Ningjia He,[1] Yan Zhou,[3] Hui Liu,[3] Jing Zhao,[3] Chen Ye,[3] Zhouhe Du,[1] Guoqing Pan,[1] Aichun Zhao,[1] Haojing Shao,[3] Wei Zeng,[3] Ping Wu,[3] Chunfeng Li,[1] Minhui Pan,[1] Jingjing Li,[3] Xuyang Yin,[3] Dawei Li,[3] Juan Wang,[3] Huisong Zheng,[3] Wen Wang,[3] Xiuqing Zhang,[3] Songgang Li,[3] Huanming Yang,[3] Cheng Lu,[1] Rasmus Nielsen,[4,5] Zeyang Zhou,[1,6] Jian Wang,[3] Zhonghuai Xiang,[1]‡ Jun Wang[3,4]‡

nature

ARTICLES

**The diploid genome se
individual**

Jun Wang[1,2,3,4]*, Wei Wang[1,3]*, Ruiqiang Li[1,3,4]*, Yingrui Li
Junqing Zhang[1], Jun Li[1], Juanbin Zhang[1], Yiran Guo[1,7], Binx
Huiqing Liang[1], Zhenglin Du[1], Dong Li[1], Yiqing Zhao[1,7], Yuji
Ines Hellmann[9], Michael Inouye[8], John Pool[9], Xin Yi[1,7], Jing Z
Guoqing Li[1], Zhentao Yang[1], Guojie Zhang[1,7], Bin Yang[1], Ch
Dawei Li[1], Peixiang Ni[1], Jue Ruan[1,7], Qibin Li[1,7], Hongmei Zhu
Jianguo Zhang[1], Jia Ye[1], Lin Fang[1], Qin Hao[1,7], Quan Chen[1
Shuang Yang[1], Fang Chen[1,7], Li Li[1], Ke Zhou[1], Hongkun Zhe
Guohua Yang[1,2], Zhuo Li[1], Xiaoli Feng[1], Karsten Kristiansen
Richard Durbin[8], Lars Bolund[1,11], Xiuqing Zhang[1,6], Songgar

Here we present the first diploid genome sequence of an Asian individual. The genome was sequenced to 36-fold average coverage using massively parallel sequencing technology. We aligned the short reads onto the NCBI human reference genome to 99.97% coverage, and guided by the reference genome, we used uniquely mapped reads to assemble a high-quality consensus sequence for 92% of the Asian individual's genome. We identified approximately 3 million single-nucleotide polymorphisms (SNPs) inside this region, of which 13.6% were not in the dbSNP database. Genotyping analysis showed that SNP identification had high accuracy and consistency, indicating the high sequence quality of this assembly. We also carried out heterozygote phasing and haplotype prediction against HapMap CHB and JPT haplotypes (Chinese and Japanese, respectively), sequence comparison with the two available individual genomes (J. D. Watson and J. C. Venter), and structural variation identification. These variations were considered for their potential biological impact. Our sequence data and analyses demonstrate the potential usefulness of next-generation sequencing technologies for personal genomics.

ARTICLES

Ancient human genome sequence of an extinct Palaeo-Eskimo

Morten Rasmussen[1,2]*, Yingrui Li[2,3]*, Stinus Lindgreen[1,4]*, Jakob Skou Pedersen[4], Anders Albrechtsen[4], Ida Moltke[4], Mait Metspalu[5], Ene Metspalu[5], Toomas Kivisild[5,6], Ramneek Gupta[7], Marcelo Bertalan[7], Kasper Nielsen[7], M. Thomas P. Gilbert[1,2], Yong Wang[8], Maanasa Raghavan[1,9], Paula F. Campos[1], Hanne Munkholm Kamp[1,4], Andrew S. Wilson[10], Andrew Gledhill[10], Silvana Tridico[11,12], Michael Bunce[12], Eline D. Lorenzen[1], Jonas Binladen[1], Xiaosen Guo[2,3], Jing Zhao[2,3], Xiuqing Zhang[2,3], Hao Zhang[2,3], Zhuo Li[2,3], Minfeng Chen[2,3], Ludovic Orlando[13], Karsten Kristiansen[2,3,4], Mads Bak[14], Niels Tommerup[14], Christian Bendixen[15], Tracey L. Pierre[16], Bjarne Grønnow[17], Morten Meldgaard[18], Claus Andreasen[19], Sardana A. Fedorova[5,20], Ludmila P. Osipova[21], Thomas F. G. Higham[9], Christopher Bronk Ramsey[10], Thomas v. O. Hansen[22], Finn C. Nielsen[22], Michael H. Crawford[23], Søren Brunak[7,24], Thomas Sicheritz-Pontén[7], Richard Villems[5], Rasmus Nielsen[4,8], Anders Krogh[2,4], Jun Wang[2,3,4] & Eske Willerslev[1,2]

We report here the genome sequence of an ancient human. Obtained from ~4,000-year-old permafrost-preserved hair, the genome represents a male individual from the first known culture to settle in Greenland. Sequenced to an average depth of 20×, we recover 79% of the diploid genome, an amount close to the practical limit of current sequencing technologies. We identify 353,151 high-confidence single-nucleotide polymorphisms (SNPs), of which 6.8% have not been reported previously. We estimate raw read contamination to be no higher than 0.8%. We use functional SNP assessment to assign possible phenotypic characteristics of the individual that belonged to a culture whose location has yielded only trace human remains. We compare the high-confidence SNPs to those of contemporary populations to find the populations most closely related to the individual. This provides evidence for a migration from Siberia into the New World some 5,500 years ago, independent of that giving rise to the modern Native Americans and Inuit.

- BGI (formerly Beijing Genomics Institute, a spin out of the CAS Beijing Institute of Genomics) has gone from accounting for around 1 per cent of the world's gene sequencing capabilities in 1999 to almost 50 per cent today. BGI works with more than 10,000 collaborators from universities and industry around the world.[152]

Never forget:

Nothing would have been possible without all those pioneers, our partners, teachers and friends.

Joe Biden is wrong. China does innovate

By Charles Riley @CRrileyCNN May 29, 2014: 7:09 PM ET

Joe Biden described China this week as a nation incapable of producing innovative products and ideas. "I challenge you: Name me one innovative project, one innovative change, one innovative product that has come out of China," the vice president dared cadets at Wednesday's Air Force Academy graduation.

Biden wants 1 China innovation. Here's 4 : (Smartphone maker Xiaomi, Tech company Tencent, Network maker Huawei)

Biotech firm BGI: This company sequences more DNA than any other institution on earth -- more than Harvard, more than the National Institutes of Health. It's an industry leader with nearly limitless ambition.

今年5月，美国副总统乔·拜登 (Joe Biden) 曾在美国空军学院毕业典礼(Air Force Academy)上说，中国是一个缺乏创新产品和能力的国家，“我谅你们也说不出一个来自中国的创新项目、创新改变或创新产品。”

CNN网站很快发文回应，认为拜登低估了中国，说中国已经培养出一些全世界最具创新能力的企业，并举了华大基因为例。

CNN说，华大对DNA的排序工作做得比全球其他任何公司都要多，其数量超过了哈佛大学，也超过了美国NIH。在基因领域，华大基因是行业领导者，其追求永无止境。

Outline:

1. A later comer of a revolution
2. **Two pillars of genomics**
3. Three impacts of the HGP
4. Four applications to medicine
5. Five techs changing the world,

DNA: *The 1st Revolution in Life Science*

After giving a brief history of the advances in life sciences from the "first revolution," the discovery of the structure of DNA, through *the "second revolution," the sequencing of the human genome,* he asked, "What is the next revolution in life sciences?"

“生物学已经经历了两次革命，第一次革命是沃森和克里克发现 DNA 的双螺旋结构，*第二次革命是“人类基因组计划”*；生物学目前正经历第三次革命，即生物学在分子层面与物理学、工程学等领域的融合。”

Phillip Sharp (美国科学促进会（AAAS）主席、诺贝尔奖获得者), AAAS 2014 年会, 芝加哥, 13 Feb, 2014

~ 60 Years ago

April 25, 2013

Cell

1953: When Genes Became "Information"

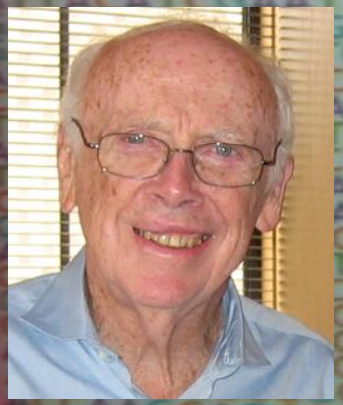
Matthew Cobb[1,*]
[1]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK

"In 1953, Watson and Crick not only described the double-helix structure of DNA, but also embraced the idea that genes contained a code that expresses information and thereby changed our view of life."

Two Pillars of Genomics

“Life is of sequence”

生命是序列的



“Life is digital!”

生命是数据的

Genomics/Sequencing

Digitalization of Life

TECHNOL

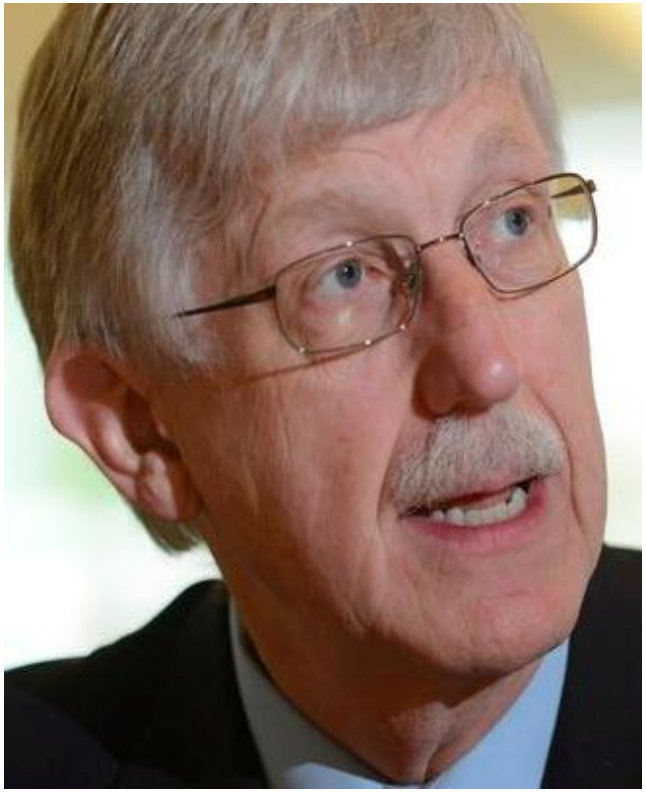
THE BIG CHALLENGES OF BIG DATA

As they grapple with increasingly large data sets, biologists and computer scientists uncork new bottlenecks.

A major part of Big Data

Nature, 13 June, 2013

data traffic jams.



“Your DNA enters the digital age!”

Francis Collins

USA Today, May, 2014

Big Data for International Scientific Programmes: Challenges and Opportunities

A Statement of Recommendations and Actions

Beijing 9 June 2014

3S's

Selection & Analysis

Data – Information – Knowledge - Laws

Safety & Security

Data, privacy & society

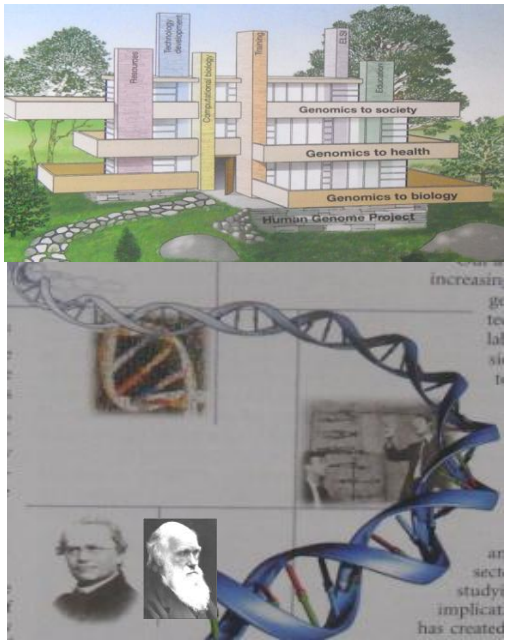
Sustainability

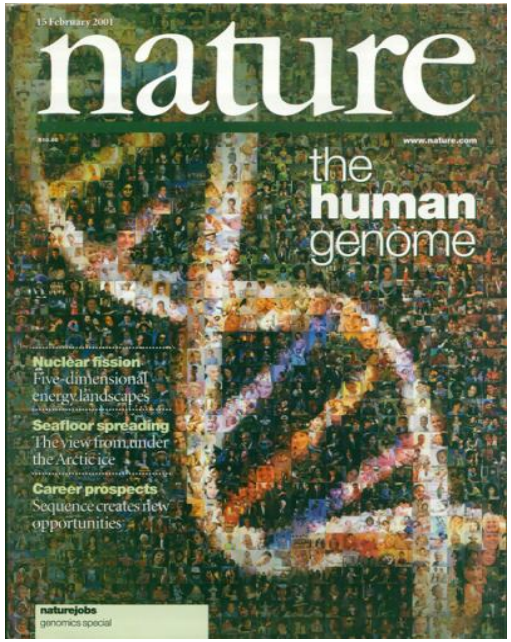
IP Right, Equal Access

Outline:

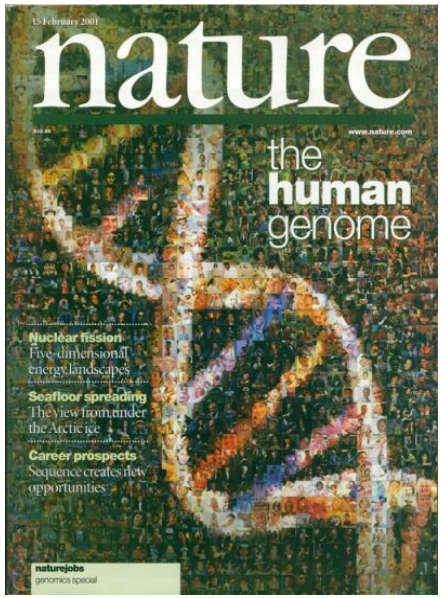
1. A later comer of a revolution
2. Two pillars of genomics
3. **Three impacts of the HGP**
4. Four applications to medicine
5. Five techs changing the world

Three Impacts of the HGP





1. **A Culture of Collaboration**
2. A New Field of Science
3. A New Tool for Biomedicine



六国首脑签署“国际人类基因组计划”宣言

温家宝(China)

希拉克(France)

施罗德(Germany)

小泉纯一郎(Japan)

布莱尔(UK)

布什(USA)

Remarkable advances in genetic science and technology have been made in the five decades since the landmark discovery of the double-helix structure of DNA in April 1953. Now, in the very month and year of the 50th anniversary of that important discovery by Watson and Crick, the International Human Genome Sequencing Consortium has completed decoding all the chapters of the instruction book of human life. This information is now freely available to the world without constraints via public databases on the World Wide Web.

Proclamation on the HGP

人类“**生命天**书”**全部章**节的解读，适逢 **DNA 双螺旋**结构发表五十周年。五十年前的这个月，**沃森与克里克**这一里程碑的发现， **使基因科学与生物技**术取得了举世瞩目的进展； **五十年后的**这一天，“**国**际人类基因组测序协作组” **公布了人**类基因组**序列信息, 全世界都可以通**过国际互联**网从公共数据**库中**自由分享**， **免**费使用而不受任何限制。

“人类基因组联合宣言”

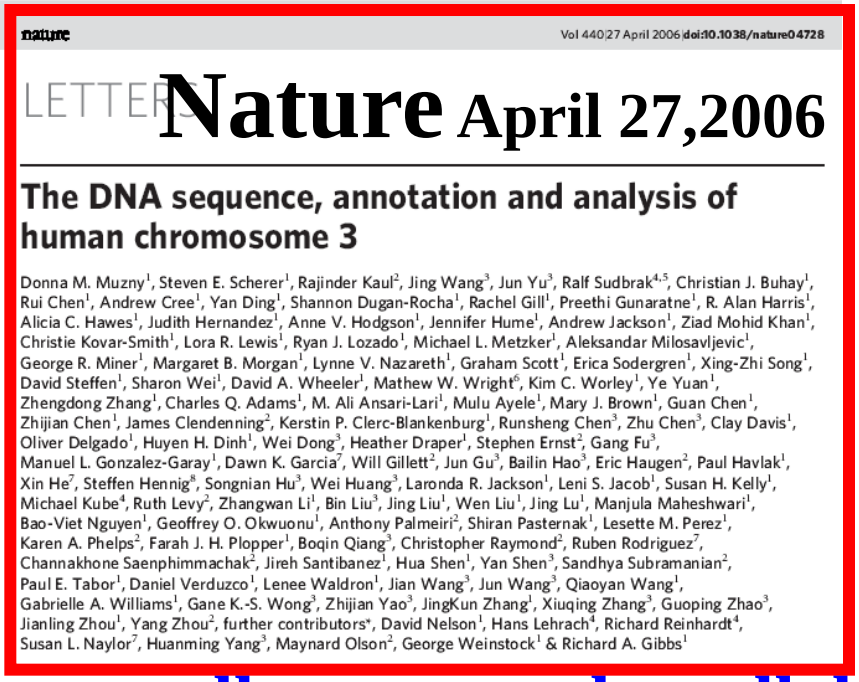
“The HGP Spirit”

“Needed by All (共需),

Owned by All (共有),

Done by All (共为),

Shared by All (共享)!”



nature Vol 440|27 April 2006 doi:10.1038/nature04728

LETTERS **Nature April 27,2006**

The DNA sequence, annotation and analysis of human chromosome 3

Donna M. Muzny[1], Steven E. Scherer[1], Rajinder Kaul[2], Jing Wang[3], Jun Yu[3], Ralf Sudbrak[4,5], Christian J. Buhay[1], Rui Chen[1], Andrew Cree[1], Yan Ding[1], Shannon Dugan-Rocha[1], Rachel Gill[1], Preethi Gunaratne[1], R. Alan Harris[1], Alicia C. Hawes[1], Judith Hernandez[1], Anne V. Hodgson[1], Jennifer Hume[1], Andrew Jackson[1], Ziad Mohid Khan[1], Christie Kovar-Smith[1], Lora R. Lewis[1], Ryan J. Lozado[1], Michael L. Metzker[1], Aleksandar Milosavljevic[1], George R. Miner[1], Margaret B. Morgan[1], Lynne V. Nazareth[1], Graham Scott[1], Erica Sodergren[1], Xing-Zhi Song[1], David Steffen[1], Sharon Wei[1], David A. Wheeler[1], Mathew W. Wright[6], Kim C. Worley[1], Ye Yuan[1], Zhengdong Zhang[1], Charles Q. Adams[1], M. Ali Ansari-Lari[1], Mulu Ayele[1], Mary J. Brown[1], Guan Chen[1], Zhijian Chen[1], James Clendenning[2], Kerstin P. Clerc-Blankenburg[1], Runsheng Chen[3], Zhu Chen[3], Clay Davis[1], Oliver Delgado[1], Huyen H. Dinh[1], Wei Dong[3], Heather Draper[1], Stephen Ernst[2], Gang Fu[3], Manuel L. Gonzalez-Garay[1], Dawn K. Garcia[7], Will Gillett[2], Jun Gu[3], Bailin Hao[3], Eric Haugen[2], Paul Havlak[1], Xin He[7], Steffen Hennig[8], Songnian Hu[3], Wei Huang[3], Laronda R. Jackson[1], Leni S. Jacob[1], Susan H. Kelly[1], Michael Kube[4], Ruth Levy[2], Zhangwan Li[1], Bin Liu[3], Jing Liu[1], Wen Liu[1], Jing Lu[1], Manjula Maheshwari[1], Bao-Viet Nguyen[1], Geoffrey O. Okwuonu[1], Anthony Palmeiri[2], Shiran Pasternak[1], Lesette M. Perez[1], Karen A. Phelps[2], Farah J. H. Plopper[1], Boqin Qiang[3], Christopher Raymond[2], Ruben Rodriguez[7], Channakhone Saenphimmachak[2], Jireh Santibanez[1], Hua Shen[1], Yan Shen[3], Sandhya Subramanian[2], Paul E. Tabor[1], Daniel Verduzco[1], Lenee Waldron[1], Jian Wang[3], Jun Wang[3], Qiaoyan Wang[1], Gabrielle A. Williams[1], Gane K.-S. Wong[3], Zhijian Yao[3], JingKun Zhang[1], Xiuqing Zhang[3], Guoping Zhao[3], Jianling Zhou[1], Yang Zhou[2], further contributors*, David Nelson[1], Hans Lehrach[4], Richard Reinhardt[4], Susan L. Naylor[7], Huanming Yang[3], Maynard Olson[2], George Weinstock[1] & Richard A. Gibbs[1]

"Human genome sequencing presented a unique opportunity for China to join the international community. I salute all our friends and colleagues at the collaborating institutions for their contributions to this task and for their support of free data-sharing under the spirit of the Human Genome Project that is 'owned by all, done by all and shared by all,'…" said Yang.

News Release for Completion of Chromosome 3

April 27, 2006

“The HGP Spirit”

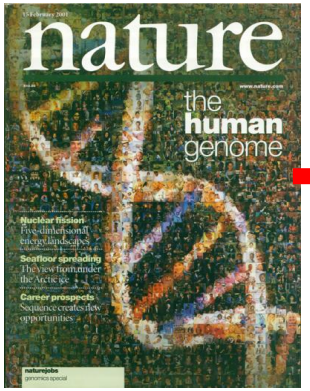
“Needed by All (共需),

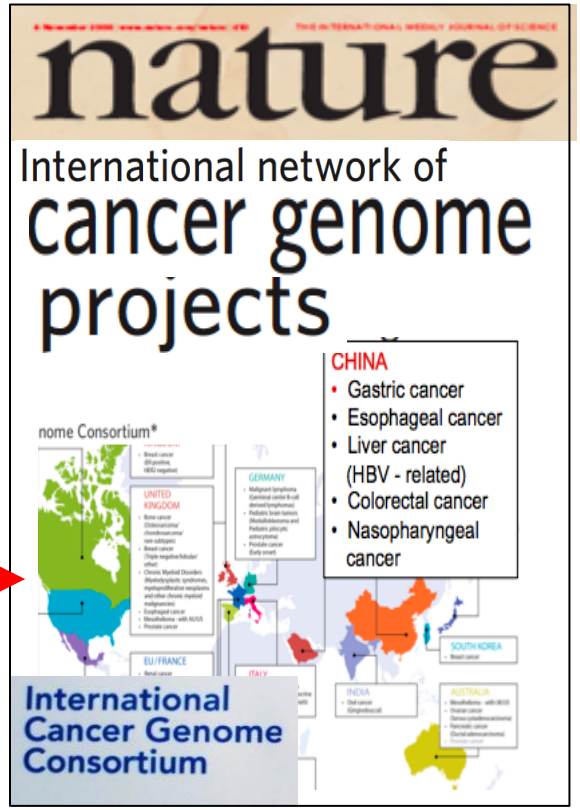
Owned by All (共有),

Done by All (共为),

Shared by All (共享)!”

A Culture/Tradition of Collaboration





Milestones in Hum/Med Genomics

Global Alliance
for Genomics & Health

June, 2013

KEY MESSAGE FOR CHALLENGING QUESTIONS

Challenge: The revolution in biomedicine made possible by widespread sequencing of the human genome and integration with clinical information will raise difficult and important questions about ethics, patient consent, technology, and regulation.

Our values: In forming this international partnership that brings together ethics, privacy, medicine, research, and technology under one tent, we aim to confront those questions from the most informed and responsible position.

The facts: The nearly 70 organizations that signed the Letter of Intent have committed to promoting the highest possible ethical standards and to a founding principle of establishing a framework so that participants will have the right to share genomic and clinical information to advance human health as broadly or narrowly as they are comfortable with, including not at all.

Bottom line: Just because the challenges are real and formidable is no reason to shy away from the opportunity before us—it just means we need to pool our intellectual resources so that we can confront difficult issues head on and in a unified way.

To continue the "tradition"

“The HGP Spirit”

Free sharing is the rule!

“免费已成‘王道’!”

3000 Rice Genome Sequences Made Publicly Available on World Hunger Day

3000株水稻基因组序列于“世界饥饿日”公开发布

（2014-05-28）

The open-access, open-data journal *GigaScience* (published by BGI) announces today the publication of the genome sequences of 3000 rice strains along with the release of this entire dataset. The publication and release of this enormous data set (which quadruples the current amount of publicly available rice sequence data) coincides with World Hunger Day to highlight one of the primary goals of this project— to develop resources that will aid in improving global food security, especially in the poorest areas of the world.

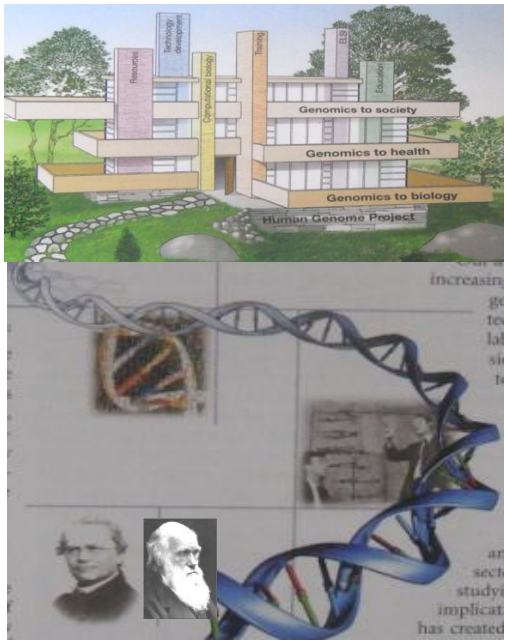
This work is the completion of stage one of the 3000 Rice Genomes Project, a collaborative effort made up of the Chinese Academy of Agricultural Sciences (CAAS), the International Rice Research Institute (IRRI), and BGI, and is funded by the Bill and Melinda Gates Foundation and the Chinese Ministry of Science and Technology.

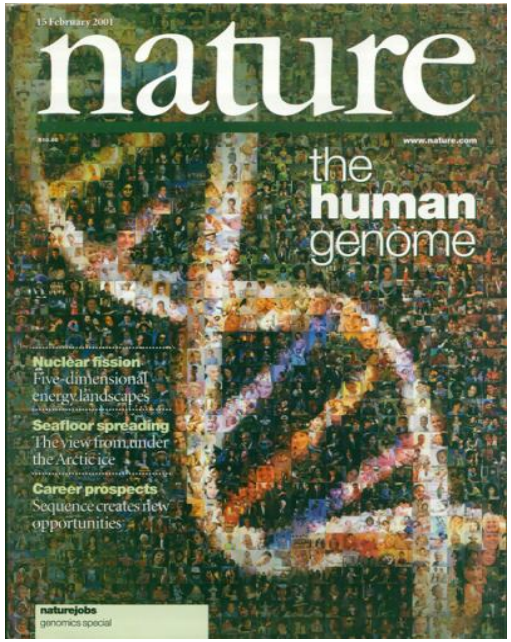
The Sciensts (June, 2014)

Members of the 3,000 Rice Genomes Project last month (May 28) delivered on their promise to make public thegenomic sequences of 3,000 rice varieties from 89 countries. Their initial analysis of the monumental dataset was published in*GigaScience.*

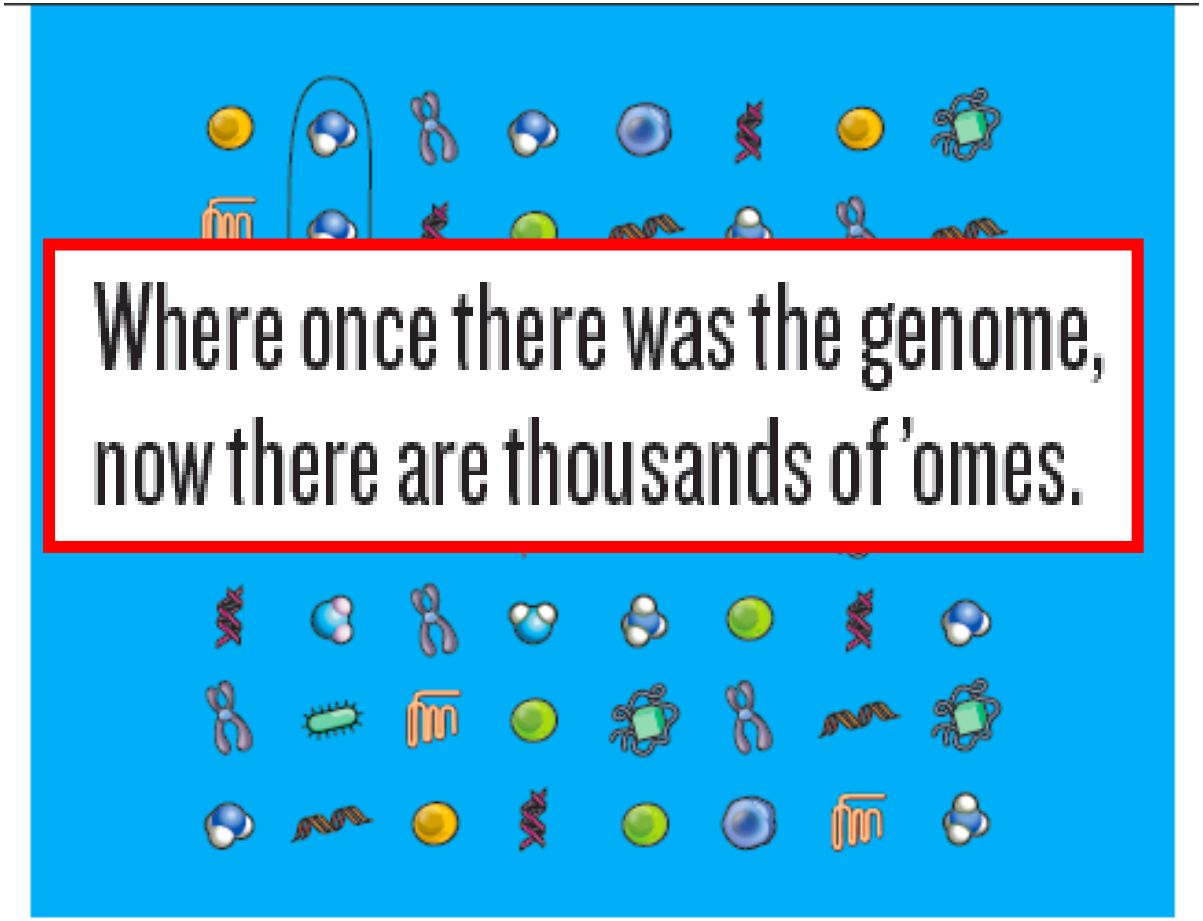
2014年5月28日，中国农业科学院、国际水稻研究所、华大基因联合开展的“3000株水稻基因组项目”在 GigaScience 上正式发表 3000株水稻基因组序列，所有数据以可引用形式在该杂志的数据库 GigaDB 中公开。该项目产生的数据是目前已公开水稻序列数据量的四倍。在“世界饥饿日”这天发布并公开这庞大的数据集，是为了体现该项目的最主要目标之一，为全球研究人员提供海量的水稻基因序列资源，为改善全球尤其是最穷困地区的粮食安全提供育种资源。

Three Impacts of the HGP





1. A Culture of Collaboration
2. **A New Field of Science**
3. A New Tool for Biomedicine



THE 'OMES PUZZLE

Where once there was the genome,
now there are thousands of 'omes.

Nature
28 Feb., 2013

“-omicsization” – HGP’s Impacts

“-Ome” & “-Omics”

“- 组”、“- 组学” 和 “组学化”

Transcriptome – Transcriptomics
Proteome – Proteomics
Methylome – Methylomics
Metabolome -- Metabolomics

Cancerome – Canceromics

……

“Today, we’ve gotten to the point where almost no biological phenomenon can escape “omicsization,” and within the next 25 years, omics will be the biggest, if not the only, game in town.”

Stephen Friend Oct. 1, 2011, The Scientist

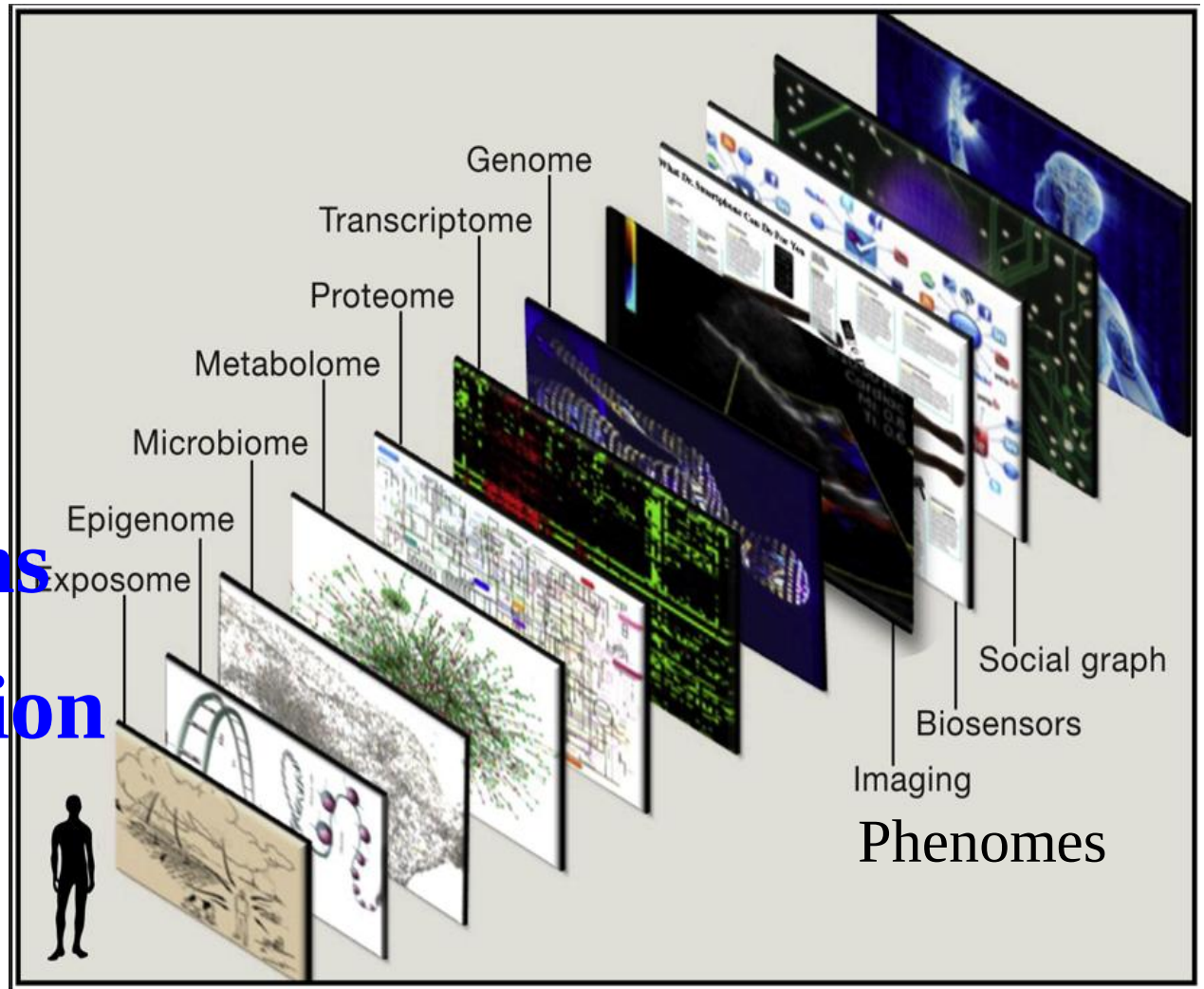
To Characterize & Rectify Biological Systems

From Genomics to Trans-omics

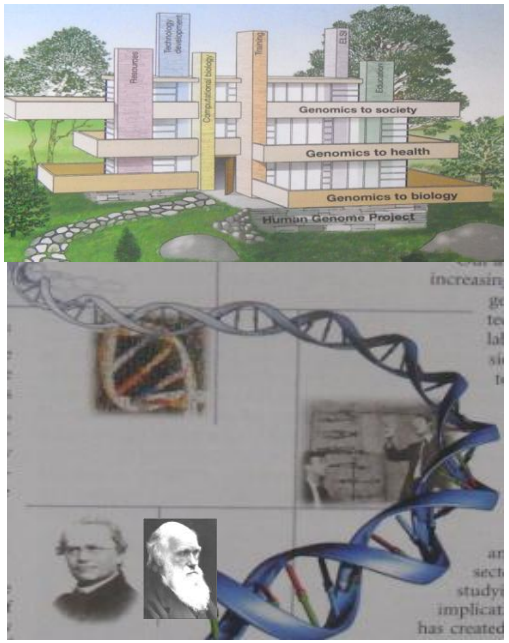
Big Data
Big Science

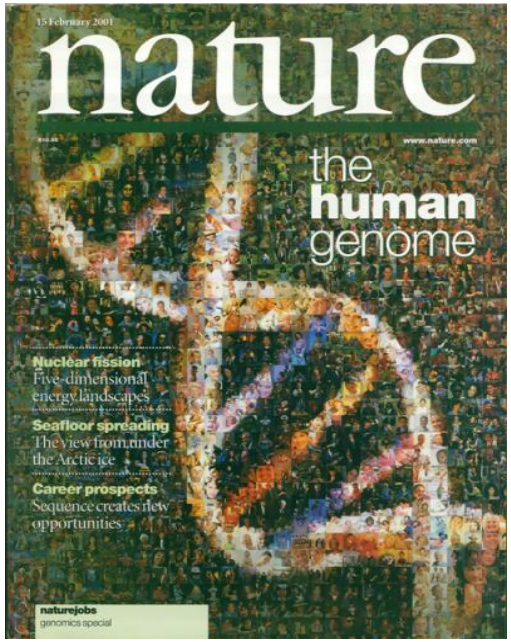
Big Platform
Big Applications

Big Collaboration
Big Sharing



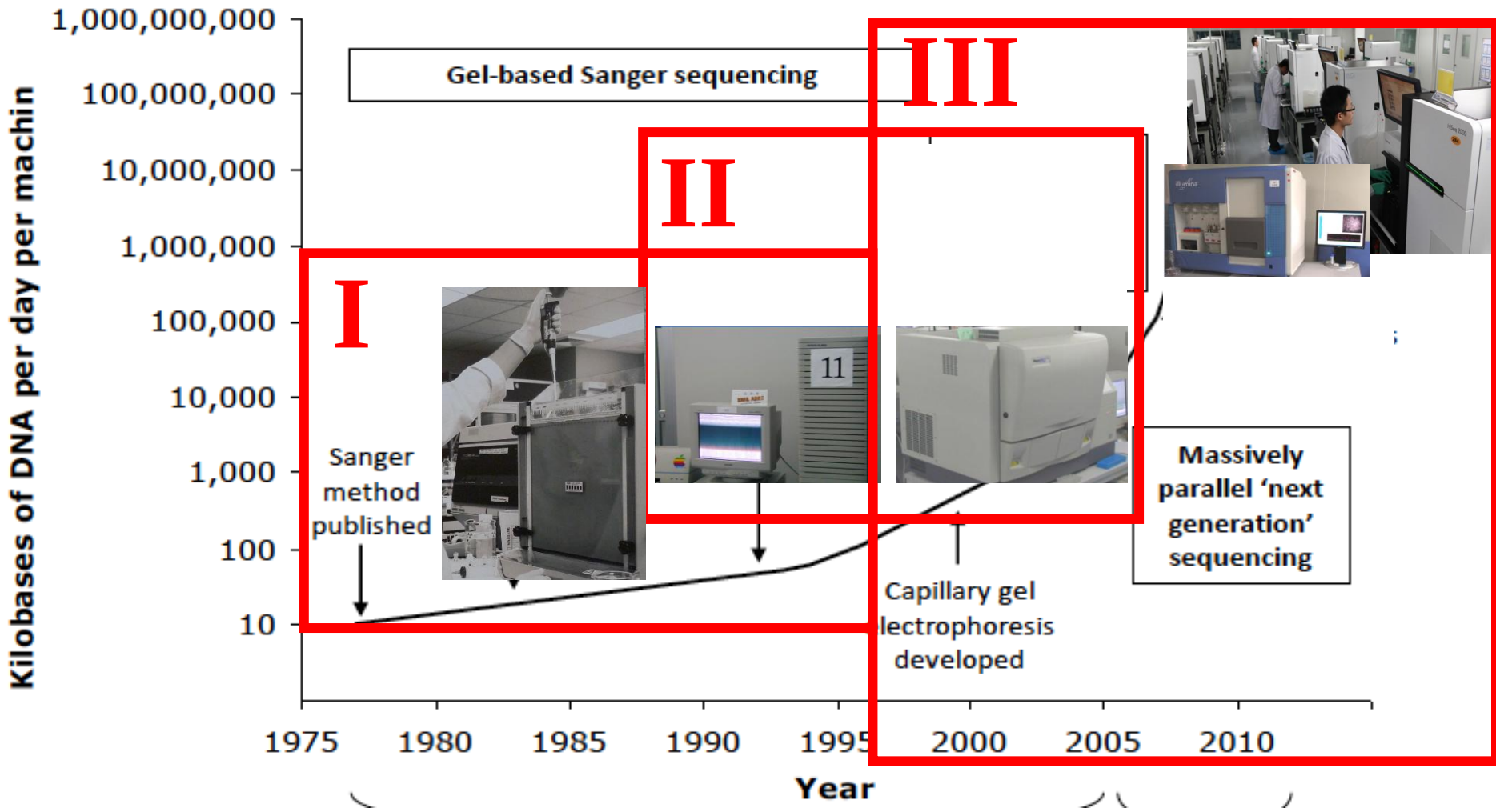
Three Impacts of the HGP





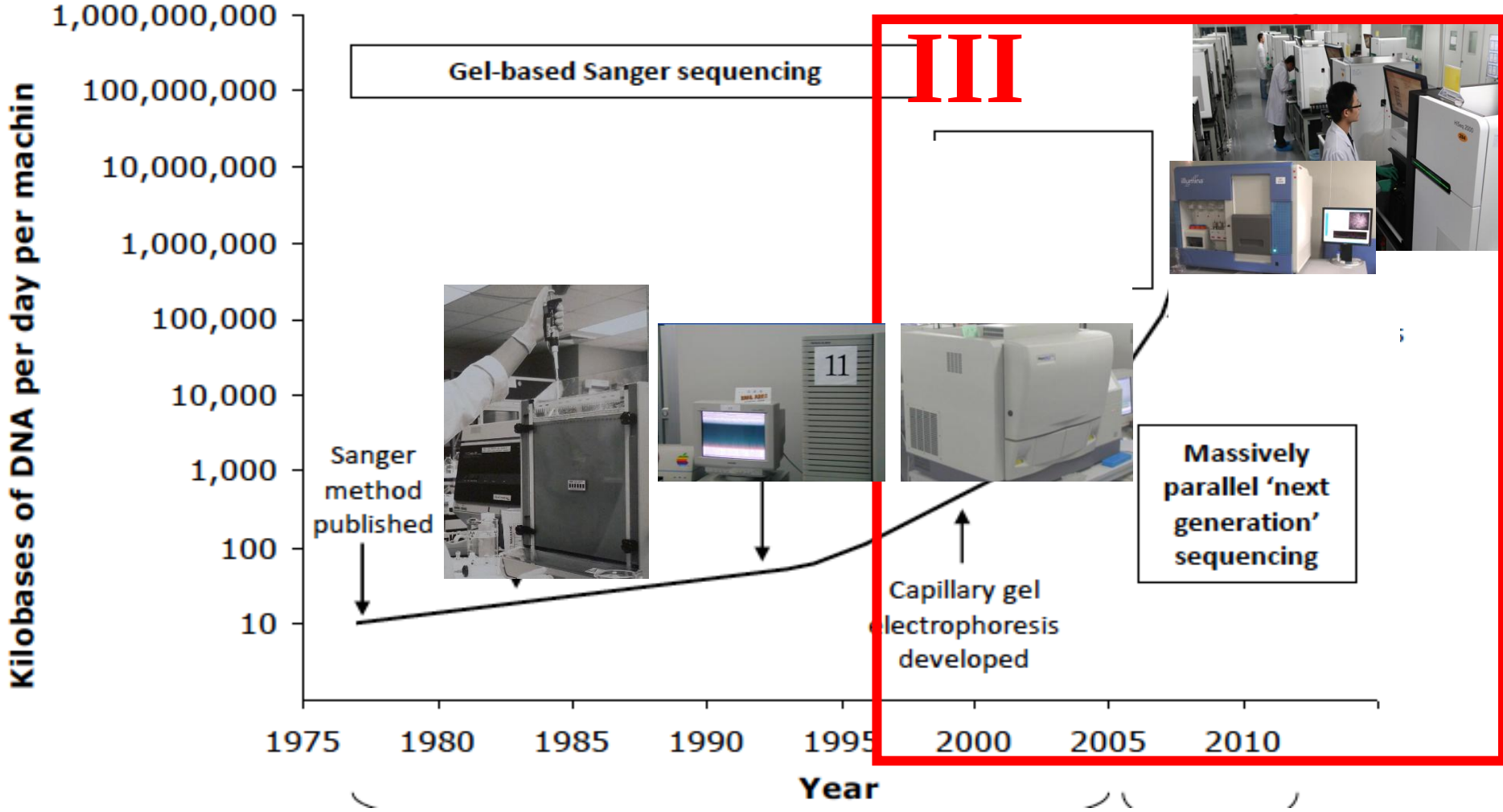
1. A Culture of Collaboration
2. A New Field of Science
3. **A New Tool for Biomedicine**

Three breakthroughs in sequencing techs



A breakthrough, an opportunity!

Three breakthroughs in sequencing techs



3. From Capillary- to Massively Parallel Seq

The Biggest Investment

The Best Software

"The biggest voice"

ARTICLES

nature

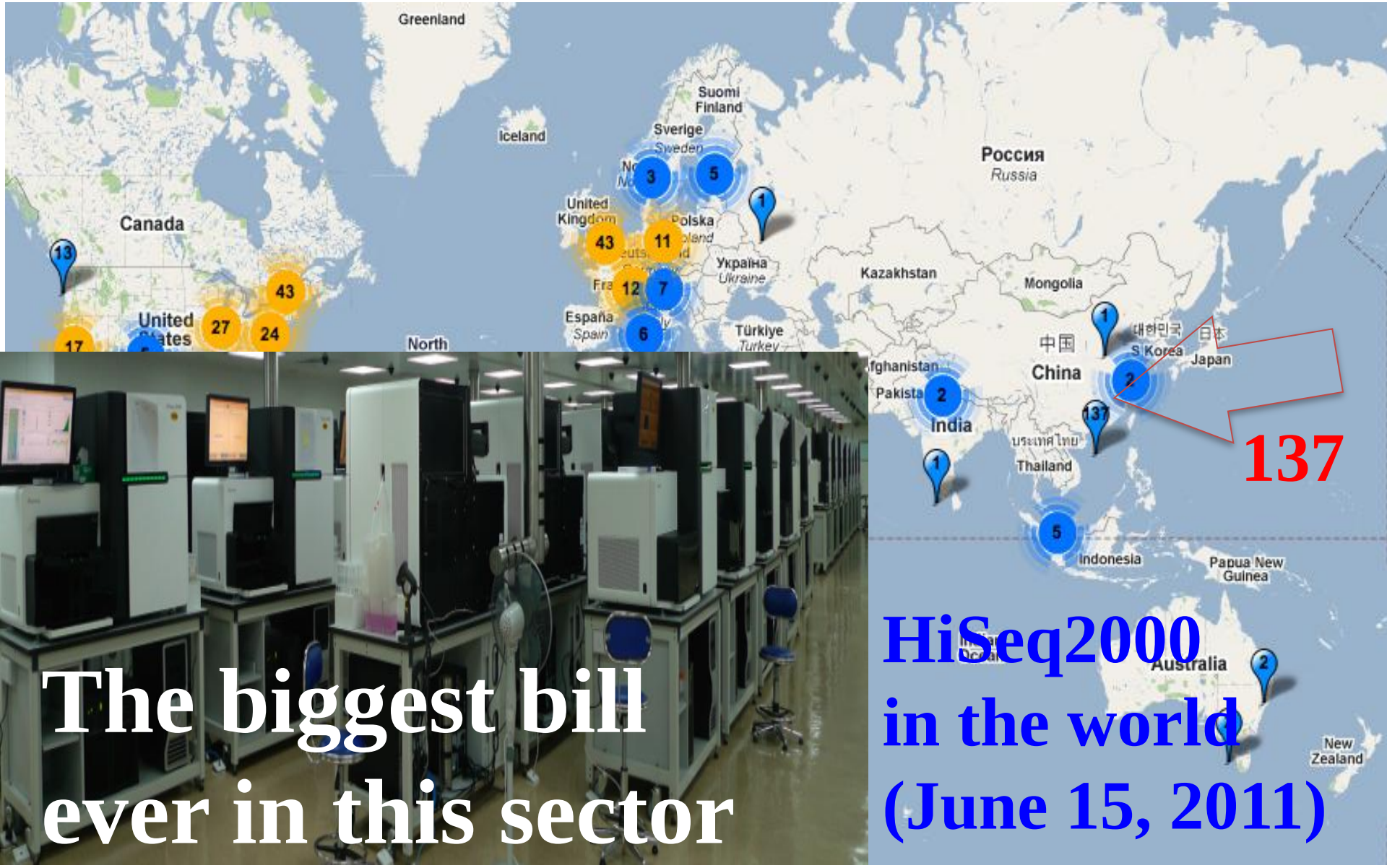
The diploid genome sequence of an Asian individual

Jun Wang[1,2,3,4*], Wei Wang[1,3*], Ruiqiang Li[1,3,4*], Yingrui Li[1,3,6*], Geng Tian[1,7], Laurie Goodman[1], Wei Fan[1], Junqing Zhang[1], Jun Li[1], Juanbin Zhang[1], Yiran Guo[1,7], Binxiao Feng[1], Heng Li[1,8], Yao Lu[1], Xiaodong Fang[1], Huiqing Liang[1], Zhenglin Du[1], Dong Li[1], Yiqing Zhao[1,7], Yujie Hu[1,7], Zhenzhen Yang[1], Hancheng Zheng[1], Ines Hellmann[9], Michael Inouye[8], John Pool[9], Xin Yi[1,7], Jing Zhao[1], Jinjie Duan[1], Yan Zhou[1], Junjie Qin[1,7], Lijia Ma[1,7], Guoqing Li[1], Zhentao Yang[1], Guojie Zhang[1,7], Bin Yang[1], Chang Yu[1], Fang Liang[1,7], Wenjie Li[1], Shaochuan Li[1], Dawei Li[1], Peixiang Ni[1], Jue Ruan[1,7], Qibin Li[1,7], Hongmei Zhu[1], Dongyuan Liu[1], Zhike Lu[1], Ning Li[1,7], Guangwu Guo[1,7], Jianguo Zhang[1], Jia Ye[1], Lin Fang[1], Qin Hao[1,5], Quan Chen[1,5], Yu Liang[1,7], Yeyang Su[1,7], A. san[1,7], Cuo Ping[1,7], Shuang Yang[1], Fang Chen[1,7], Li Li[1], Ke Zhou[1], Hongkun Zheng[1,4], Yuanyuan Ren[1], Ling Yang[1], Yang Gao[1,2], Guohua Yang[1,2], Zhuo Li[1], Xiaoli Feng[1], Karsten Kristiansen[4], Gane Ka-Shu Wong[1,10], Rasmus Nielsen[9], Richard Durbin[8], Lars Bolund[1,11], Xiuqing Zhang[1,6], Songgang Li[1,2,5], Huanming Yang[1,2,3] & Jian Wang[1,2,3]

_computational BIOLOGY

Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome *de novo* assembly

Yingrui Li[1,10], Hancheng Zheng[1,10], Ruibang Luo[1–3,10], Honglong Wu[1,4,10], Hongmei Zhu[1], Ruiqiang Li[1], Hongzhi Cao[1,4], Boxin Wu[1], Shujia Huang[1,2], Haojing Shao[1,2], Hanzhou Ma[1,2], Fan Zhang[1,2], Shuijian Feng[1], Wei Zhang[1], Hongli Du[2], Geng Tian[1], Jingxiang Li[1], Xiuqing Zhang[1], Songgang Li[1], Lars Bolund[1,5], Karsten Kristiansen[1,6], Adam J de Smith[7], Alexandra I F Blakemore[7], Lachlan J M Coin[8], Huanming Yang[1], Jian Wang[1] & Jun Wang[1,6,9]



by Nick Loman (University of Birmingham)

http://pathogenomics.bham.ac.uk

The Biggest Investment

The Best Software

"The biggest voice"

ARTICLES

nature

The diploid genome sequence of an Asian individual

Jun Wang[1,2,3,4*], Wei Wang[1,3*], Ruiqiang Li[1,3,4*], Yingrui Li[1,3,6*], Geng Tian[1,7], Laurie Goodman[1], Wei Fan[1], Junqing Zhang[1], Jun Li[1], Juanbin Zhang[1], Yiran Guo[1,7], Binxiao Feng[1], Heng Li[1,8], Yao Lu[1], Xiaodong Fang[1], Huiqing Liang[1], Zhenglin Du[1], Dong Li[1], Yiqing Zhao[1,7], Yujie Hu[1,7], Zhenzhen Yang[1], Hancheng Zheng[1], Ines Hellmann[9], Michael Inouye[8], John Pool[9], Xin Yi[1,7], Jing Zhao[1], Jinjie Duan[1], Yan Zhou[1], Junjie Qin[1,7], Lijia Ma[1,7], Guoqing Li[1], Zhentao Yang[1], Guojie Zhang[1,7], Bin Yang[1], Chang Yu[1], Fang Liang[1,7], Wenjie Li[1], Shaochuan Li[1], Dawei Li[1], Peixiang Ni[1], Jue Ruan[1,7], Qibin Li[1,7], Hongmei Zhu[1], Dongyuan Liu[1], Zhike Lu[1], Ning Li[1,7], Guangwu Guo[1,7], Jianguo Zhang[1], Jia Ye[1], Lin Fang[1], Qin Hao[1,5], Quan Chen[1,5], Yu Liang[1,7], Yeyang Su[1,7], A. san[1,7], Cuo Ping[1,7], Shuang Yang[1], Fang Chen[1,7], Li Li[1], Ke Zhou[1], Hongkun Zheng[1,4], Yuanyuan Ren[1], Ling Yang[1], Yang Gao[1,2], Guohua Yang[1,2], Zhuo Li[1], Xiaoli Feng[1], Karsten Kristiansen[4], Gane Ka-Shu Wong[1,10], Rasmus Nielsen[9], Richard Durbin[8], Lars Bolund[1,11], Xiuqing Zhang[1,6], Songgang Li[1,2,5], Huanming Yang[1,2,3] & Jian Wang[1,2,3]

_computational BIOLOGY

Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome *de novo* assembly

Yingrui Li[1,10], Hancheng Zheng[1,10], Ruibang Luo[1–3,10], Honglong Wu[1,4,10], Hongmei Zhu[1], Ruiqiang Li[1], Hongzhi Cao[1,4], Boxin Wu[1], Shujia Huang[1,2], Haojing Shao[1,2], Hanzhou Ma[1,2], Fan Zhang[1,2], Shuijian Feng[1], Wei Zhang[1], Hongli Du[2], Geng Tian[1], Jingxiang Li[1], Xiuqing Zhang[1], Songgang Li[1], Lars Bolund[1,5], Karsten Kristiansen[1,6], Adam J de Smith[7], Alexandra I F Blakemore[7], Lachlan J M Coin[8], Huanming Yang[1], Jian Wang[1] & Jun Wang[1,6,9]

We have been so stubborn/persistent in sequencing in the past 15 years.

“Sequencing, sequencing, and sequencing!”

– J. Wang

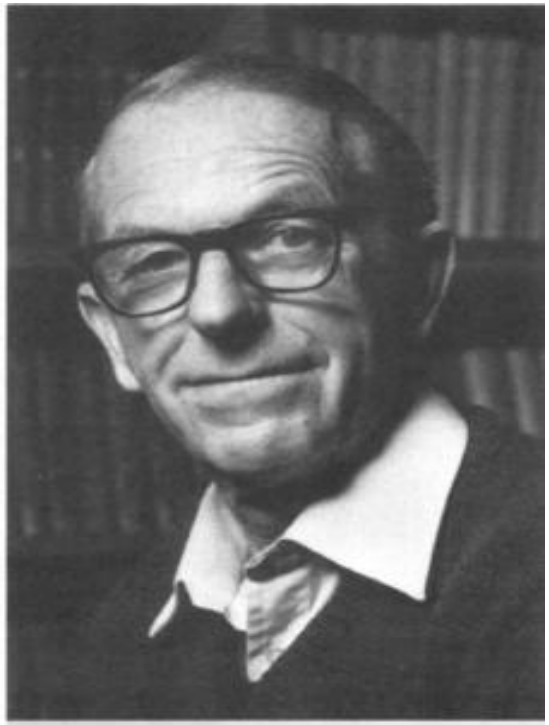
“测序，测序，再测序!” ***---王俊***

Ann. Rev. Biochem. 1988. 57:1–28

SEQUENCES, SEQUENCES, AND SEQUENCES

“序列，序列，还是序列!”

Frederick Sanger



“Father of Genomics”

His contribution to DNA sequencing technology has transformed our understanding of life on earth by *making life digital*

“Sequencing, sequencing, and sequencing!”

— J. Wang

“测序，测序，再测序!”---王俊

BGI SHENZHEN nature

TAKING THE WORLD BY STORM

CORRECTED COUNT: 3.572 ARTICLES: 9

华大旋风席卷全球

One particular thing that stands out in the Nature Publishing Index 2010 is the number of articles published in the Nature journals by BGI Shenzhen — the premier genomics institute in China. In 2010, the former Beijing Genomics Institute contributed nine articles (CC 3.57) to Nature journals, with the majority of these in genetics and biotechnology. BGI Shenzhen is now ranked second and third in the Asia-Pacific

走遍千山万水
走进千家万户
说尽千言万语
历经千辛万苦

BGI deserves all appreciation for advocating sequencing globally.

The Biggest Investment

The Best Software

"The biggest voice"

ARTICLES

nature

The diploid genome sequence of an Asian individual

Jun Wang[1,2,3,4*], Wei Wang[1,3*], Ruiqiang Li[1,3,4*], Yingrui Li[1,3,6*], Geng Tian[1,7], Laurie Goodman[1], Wei Fan[1], Junqing Zhang[1], Jun Li[1], Juanbin Zhang[1], Yiran Guo[1,7], Binxiao Feng[1], Heng Li[1,8], Yao Lu[1], Xiaodong Fang[1], Huiqing Liang[1], Zhenglin Du[1], Dong Li[1], Yiqing Zhao[1,7], Yujie Hu[1,7], Zhenzhen Yang[1], Hancheng Zheng[1], Ines Hellmann[9], Michael Inouye[8], John Pool[9], Xin Yi[1,7], Jing Zhao[1], Jinjie Duan[1], Yan Zhou[1], Junjie Qin[1,7], Lijia Ma[1,7], Guoqing Li[1], Zhentao Yang[1], Guojie Zhang[1,7], Bin Yang[1], Chang Yu[1], Fang Liang[1,7], Wenjie Li[1], Shaochuan Li[1], Dawei Li[1], Peixiang Ni[1], Jue Ruan[1,7], Qibin Li[1,7], Hongmei Zhu[1], Dongyuan Liu[1], Zhike Lu[1], Ning Li[1,7], Guangwu Guo[1,7], Jianguo Zhang[1], Jia Ye[1], Lin Fang[1], Qin Hao[1,5], Quan Chen[1,5], Yu Liang[1,7], Yeyang Su[1,7], A. san[1,7], Cuo Ping[1,7], Shuang Yang[1], Fang Chen[1,7], Li Li[1], Ke Zhou[1], Hongkun Zheng[1,4], Yuanyuan Ren[1], Ling Yang[1], Yang Gao[1,2], Guohua Yang[1,2], Zhuo Li[1], Xiaoli Feng[1], Karsten Kristiansen[4], Gane Ka-Shu Wong[1,10], Rasmus Nielsen[9], Richard Durbin[8], Lars Bolund[1,11], Xiuqing Zhang[1,6], Songgang Li[1,2,5], Huanming Yang[1,2,3] & Jian Wang[1,2,3]

computational BIOLOGY

Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome *de novo* assembly

Yingrui Li[1,10], Hancheng Zheng[1,10], Ruibang Luo[1–3,10], Honglong Wu[1,4,10], Hongmei Zhu[1], Ruiqiang Li[1], Hongzhi Cao[1,4], Boxin Wu[1], Shujia Huang[1,2], Haojing Shao[1,2], Hanzhou Ma[1,2], Fan Zhang[1,2], Shuijian Feng[1], Wei Zhang[1], Hongli Du[2], Geng Tian[1], Jingxiang Li[1], Xiuqing Zhang[1], Songgang Li[1], Lars Bolund[1,5], Karsten Kristiansen[1,6], Adam J de Smith[7], Alexandra I F Blakemore[7], Lachlan J M Coin[8], Huanming Yang[1], Jian Wang[1] & Jun Wang[1,6,9]

Best of the best (for synthetic genomes)

Team	Assembler	Affiliation
P	SOAPdenovo	BGI
Q	ALLPATHS	Broad Institute
D	SGA	

Bradnam, UC Davis, CSHL, 12 May, 2010

Must Read

F1000 Factor 6.0

The sequence and *de novo* assembly of the giant panda genome

No "genetic map"

No "physical map"

No prior knowledge of repeats

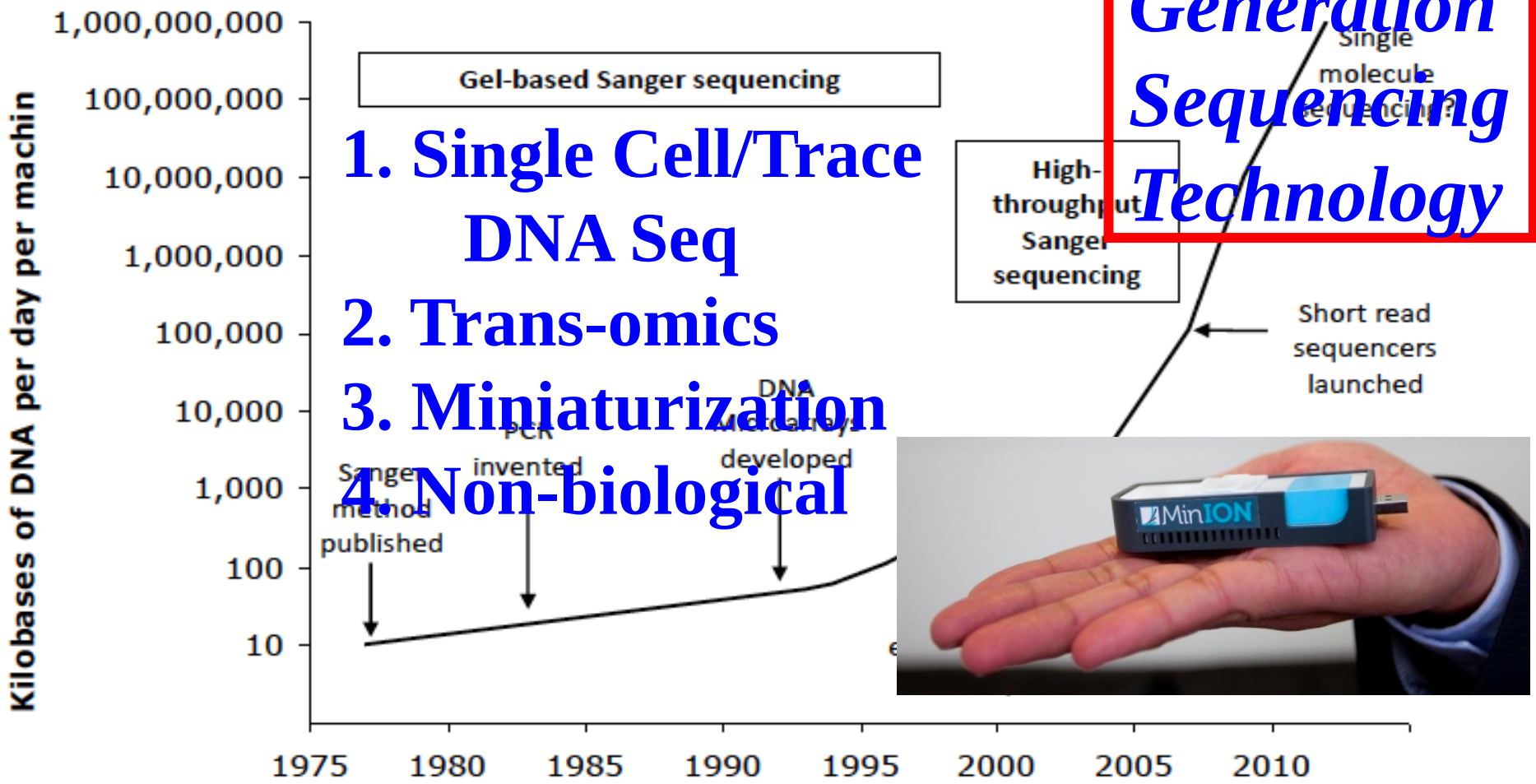
Making wide applications of the now-generation sequencing technology possible

NEWS & VIEW

The giant-panda genome is the first reported *de novo* assembly of a large mammalian genome achieved using next-generation sequencing methods.

The Next Breakthroughs …

The 4th, 5th Generation Sequencing Technology



1. Single Cell/Trace DNA Seq
2. Trans-omics
3. Miniaturization
4. Non-biological

What do we expect from the Next-, Next-Next Sequencing Technology?

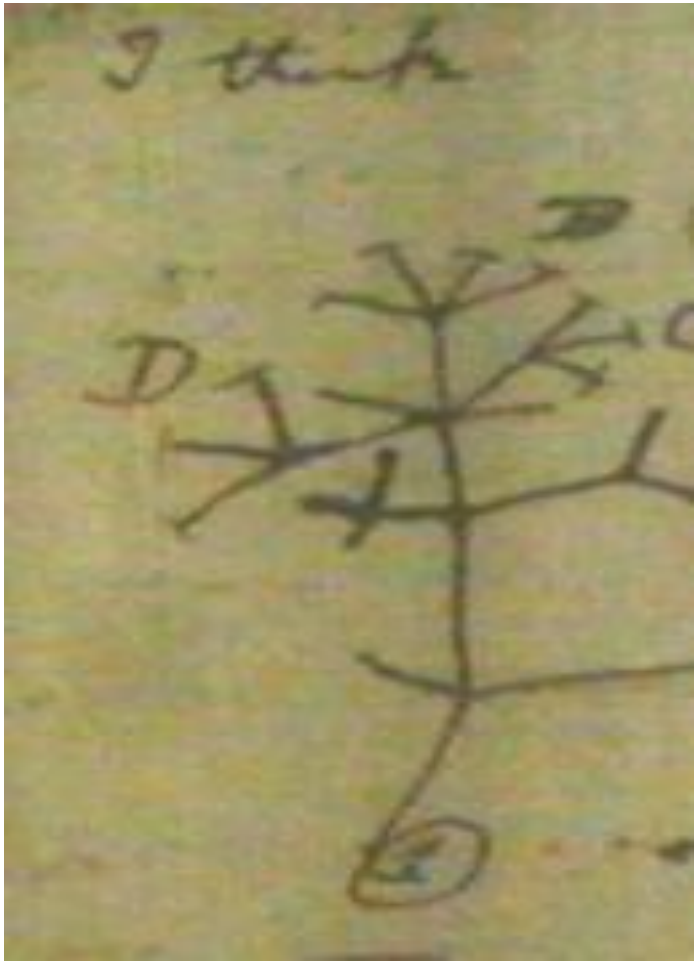
Applications of the DNA Data

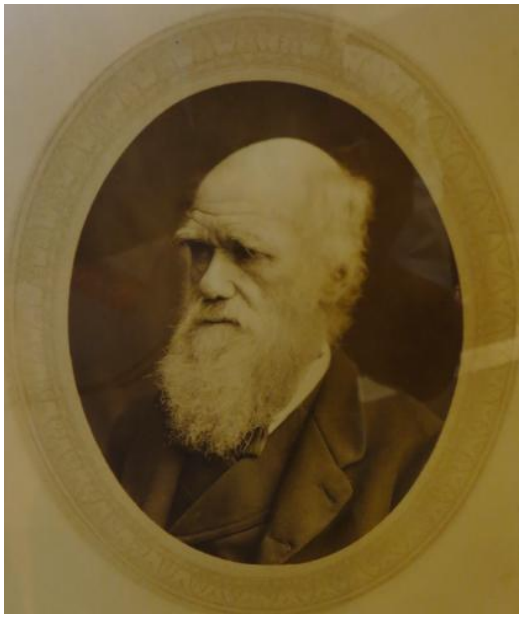
Three "to improve"

1. **To improve knowledge of life**
2. **To improve breeding**
3. **To improve health**

1. To Improve Knowledge of Life

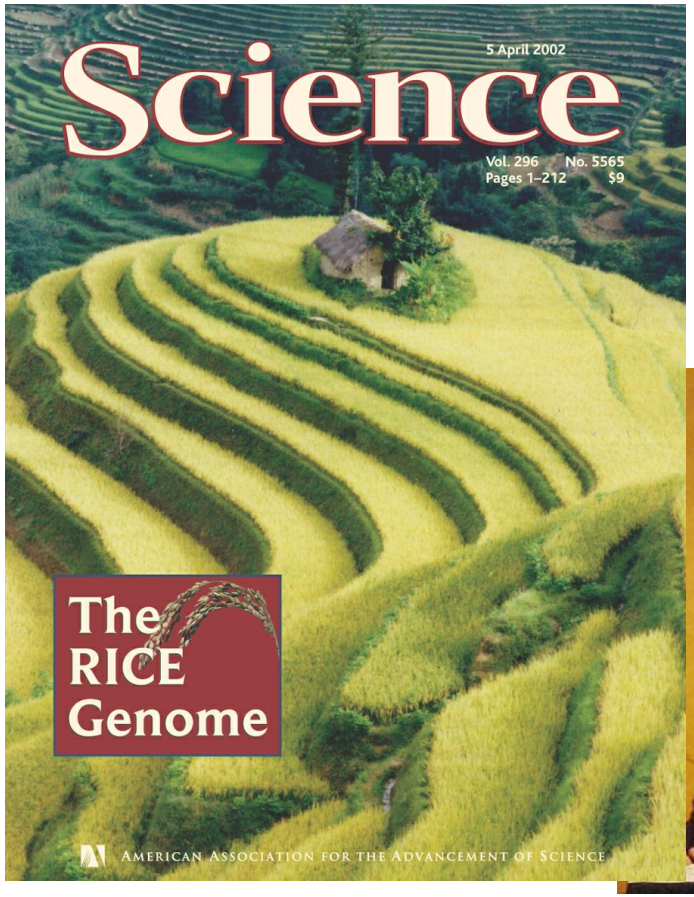
To redraw a sequence-based/ digitalized "Tree of Life" by *sequencing more species* (a reference genome for each species)

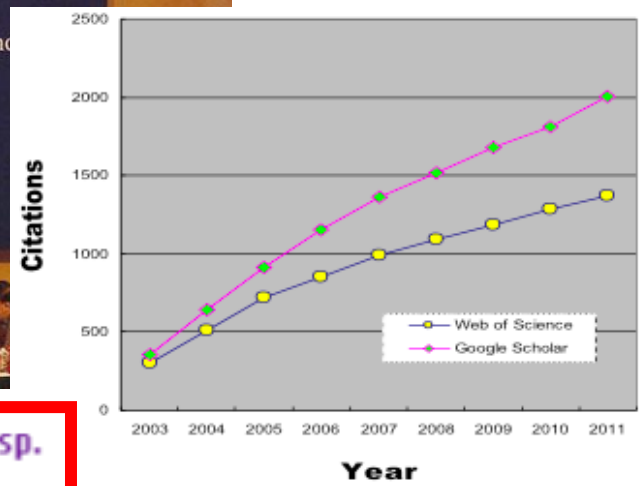




构建序列为基础的、数据化的"生命之树"

Rice: The first and biggest genome sequenced & assembled fully using WGS approach.





Exceptional
F1000 Factor 9.8

A draft sequence of the rice genome (Oryza sativa L. ssp. indica).

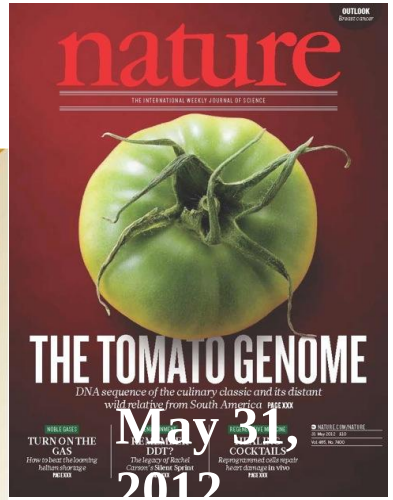
Yu J, Hu S, ..., Yuan L, Yang H

Science 2002 Apr 5 **296**(5565):79-92

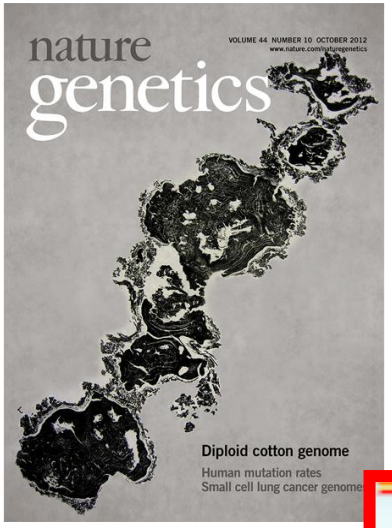
"A landmark paper should be read by all plant biologists."

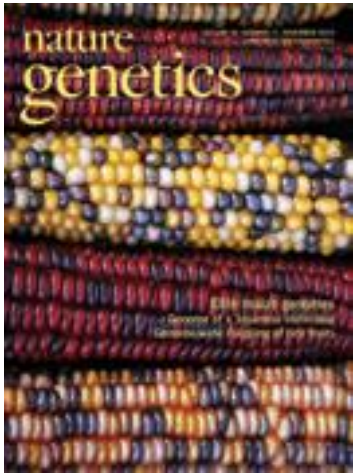
Plants
植 物

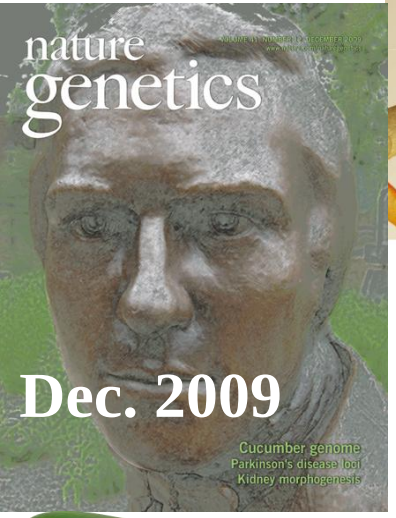
July, 2011



May 31, 2012







Dec. 2009

Cucumber

nature genetics

Brassica rapa

(大白菜)

Aug. 2011

nature biotechnology

pigeonpea

(*Cajanus cajan*)

30:83-89, 2011

nature genetics

Watermelon

(西瓜)

Nov. 26, 2012

nature COMMUNICATIONS

P. murne

(梅花)

Dec. 27, 2012

Animals
动物

Verte-brates

29:1019-23,2011

cynomolgus & Chinese rhesus

339:456, 2013

-227, 2011

naked mole rat

online July 1, 2012

Yak

Aug. 2010

online Sept.19, 2012

Oyster

online Jan 13, 2013

Moth

479:529-533, 2011

Ascaris suum

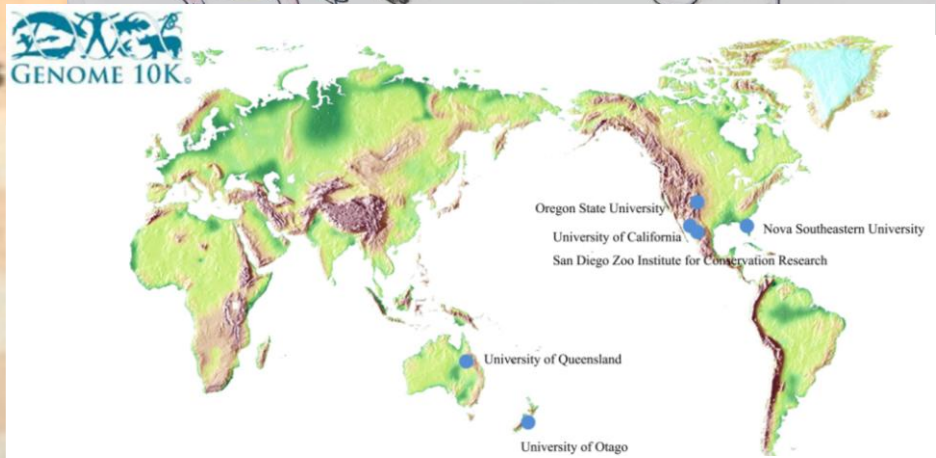
Inverte-brates

"The Genome 10K Project"

SPECIES ON THE SEQUENCING DOCKET

Group	Known Species	10K Species
Birds	9723	5074
Reptiles	9002	3297
Mammals	5416	1826
Amphibians	6570	1760
Fish	31564	4246
Total	62275	16203

26



Science
326: 794 Nov. 6, 2009

10000 Avian Genomes Project

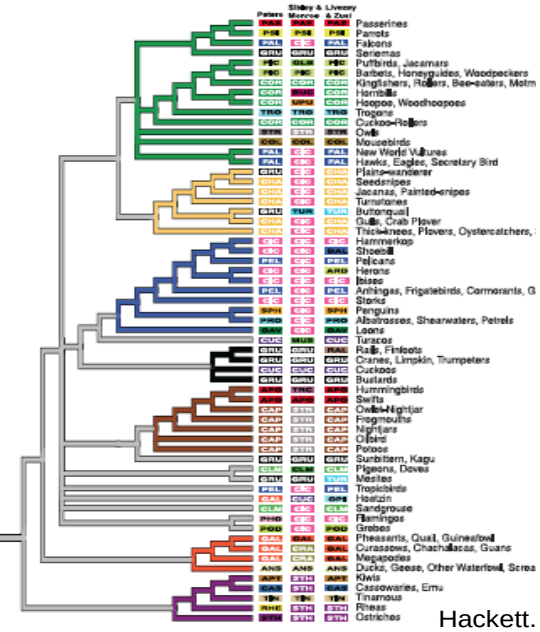
Genomic Diversity and Evolution of the Head Crest in the Rock Pigeon

Science 31 Jan. 2013

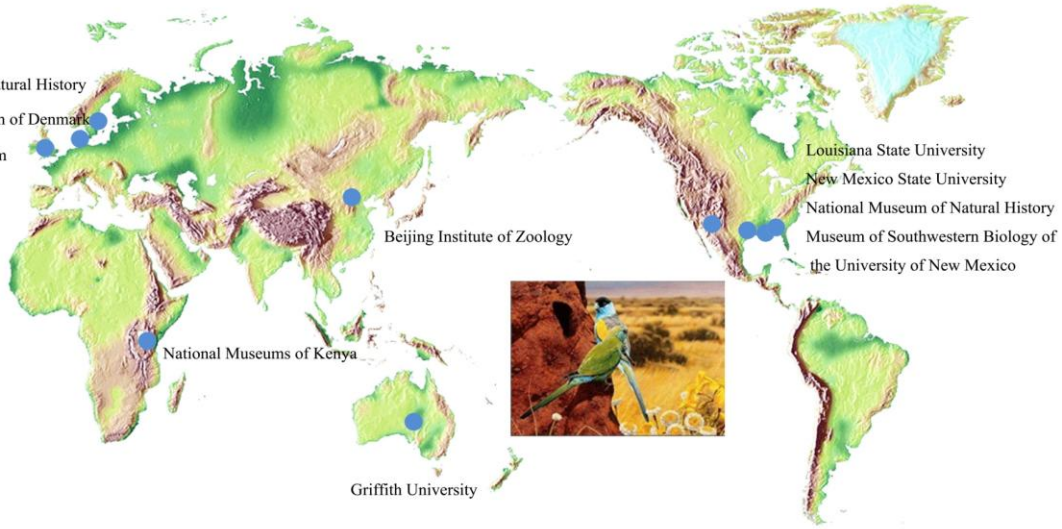
Darwin's finch

Peregrine and saker falcon genome sequences provid insights into evolution of a predatory lifestyle

nature genetics March 23, 2013



Hackett.



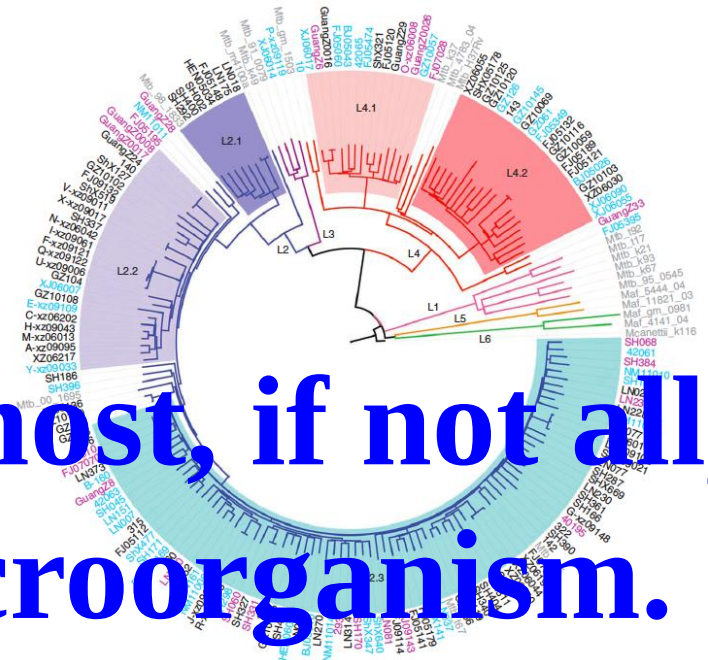
nature genetics
online
Sept 1, 2013

结核杆菌

Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance

ning[1,14], Nan Lin[7], Ting Wang[1], Zhangyi Liu[3],
Yuanfang Zhu[3], Yunrong Gao[1], Tong Wang[3],
Zhou[10], Tao Chen[10], Jie Zhou[11], Ruifu Yang[3],
n Wan[5,6], Jun Wang[3], Xian-En Zhang[2,9] & Lijun Bi[1]

The worldwide emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) tuberculosis threatens to make this disease incurable[1,2]. Drug resistance mechanisms are only partially understood[3–5], and whether the current understanding of the genetic basis of drug resistance in *M. tuberculosis* is sufficiently comprehensive remains unclear. Here we sequenced and analyzed 161 isolates with a range of drug resistance profiles, discovering 72 new genes, 28 intergenic regions (IGRs), 11 nonsynonymous SNPs and 10 IGR SNPs with strong, consistent associations with drug resistance. On the basis of our examination of the dN/dS ratios of nonsynonymous to synonymous SNPs among the isolates[6–8], we suggest that the drug resistance–associated genes identified here likely contain essentially all the nonsynonymous SNPs that have arisen as a result of drug pressure in these isolates and should thus represent a near-complete set of drug resistance–associated genes for these isolates and antibiotics. Our work indicates that the genetic basis of drug resistance is more complex than previously anticipated and provides a strong foundation for elucidating unknown drug resistance mechanisms.



To sequence most, if not all, stains of a microorganism.

Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*

鼠疫杆菌

The genetic diversity of *Yersinia pestis*, the etiologic agent of plague, is extremely limited because of its recent origin coupled with a slow clock rate. Here we identified 2,326 SNPs from 133 genomes of *Y. pestis* strains that were isolated in China and elsewhere. These SNPs

t com-
mon a
is that
happe
listrib-
uted
seven
gene
P, sug-
gestir
al pro-
cesse
ection.
Howe
eneal-
ogy: t
es was
highly
omies,
one o
Black
Death
speed
of eve
natural
select
ly dur-
ing in

Most, if not all, of the plaque pathogen strains have been collected and sequenced.

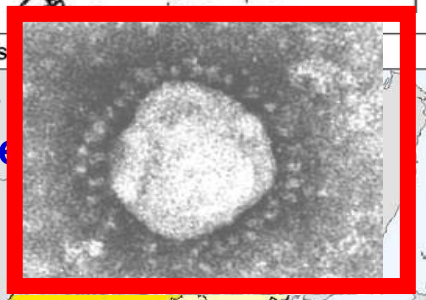
The 107 Chinese isolates represent the diversity revealed by molecular genotyping of > 900 strains, which in turn reflect the geographic and temporal diversity of > 5,000 isolates collected during annual surveillance of all sylvatic plague foci in China. Thus, these genomes should provide a good representation of the geographic and genetic diversity of Y. pestis in China.

Of the 118 genomes, 103 are from a variety of rodent genera and other non-primate mammals, 14 are from human patients, and 1 is from a human flea.

approved November 29, 2012

SARS: Number of Current Probable Cases as of 4 June 2003, 14:00 GMT+2

8,402 cases (5,746 recovered) and 772 deaths in 29 countries on 6 continents



(969 health care

ths in 14 province

China is being attacked!

The presentation of material on the maps contained herein does not imply the on the part of the World Health Organization concerning the legal status of any authorities, or concerning the delimitation of its frontiers or boundaries.

The presentation of material on the maps contained herein does not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or areas or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Data Source: China Ministry of Health
DoH Hong Kong / DoH Macao / CDC Taiwan
Map Production: Public Health Mapping Team
Communicable Diseases (CDS)

The NEW ENGLAND JOURNAL of MEDICINE July 28, 2011

Microbes

微生物 Open-Source Genomic Analysis of Shiga-Toxin–Producing *E. coli* O104:H4

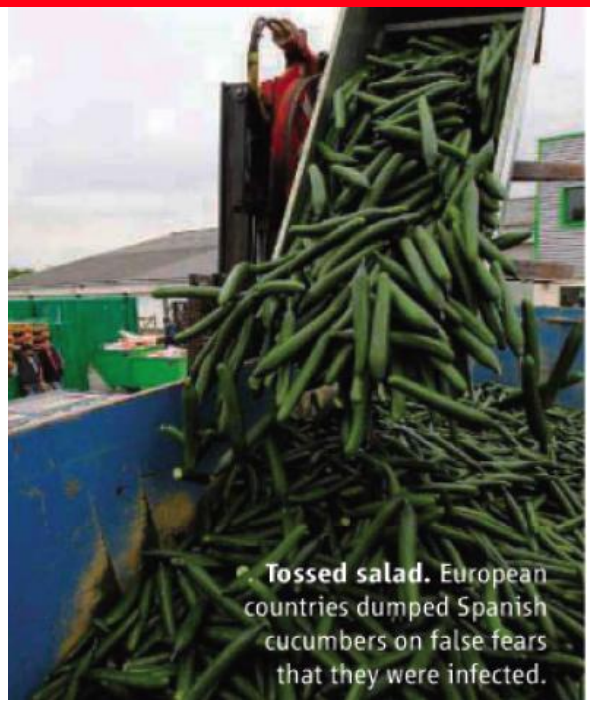
GERMANY

Scientists Rush to Study Genome of Lethal *E. coli*

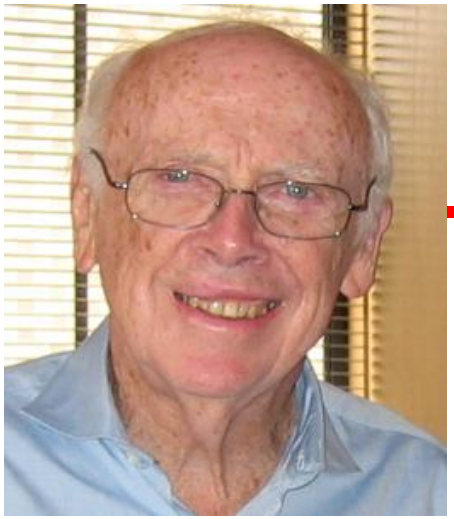
On 2 June, Chinese scientists announced that they had deciphered the microbe's entire 5.2-millionbase-pair genome and immediately made the DNA sequence available for researchers to download.

Science 332 (6035): 1249 10 June 2011

Science



Tossed salad. European countries dumped Spanish cucumbers on false fears that they were infected.



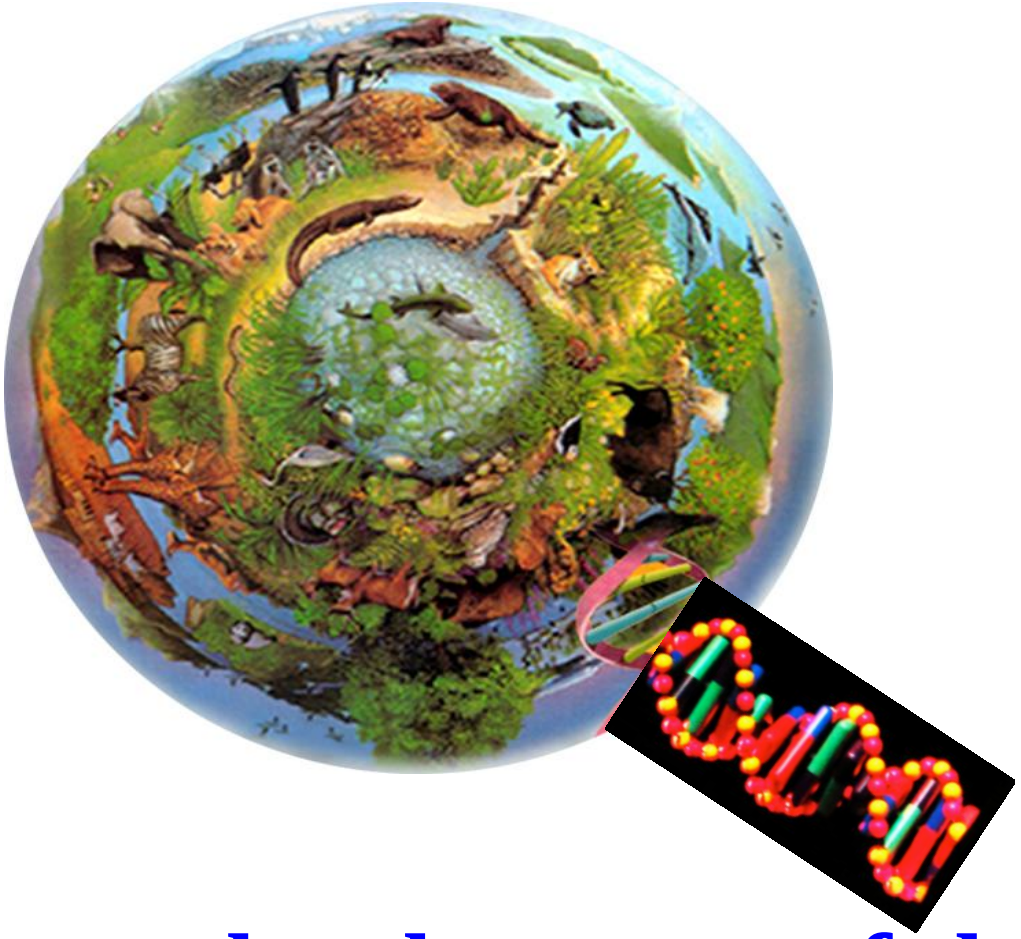
“No sequence, no biology”!

“All biology in the future will start with the knowledge of genomes and proceed hopefully.”

J. D. Watson, 2003

“未来所有生物学只有以基因组知识(重新)开始才有希望发展”

Dream I: To sequence every(living) thing on the planet（万物基因组）



For the green development of the world

Dream II. To sequence everybody in the world (“人人基因组”)

BGI Plans to Sequence the World

The Global DNA Sequencing Powerhouse Based in Shenzhen

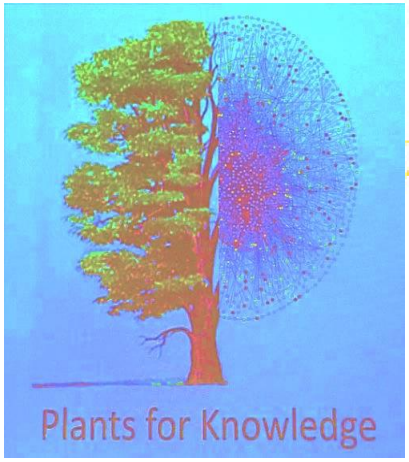
By Paul Diehl, About.com Guide

华大要测序
整个世界

During his presentation at the 2013 ICG-Americas Conference, Huaming Yang, Chairman and co-founder of BGI, briefly mentioned their 3 Million Genomes project. The idea is to sequence all the DNA of 1 million people, 1 million microorganisms, and 1 million plants and animals—3 million genomes.

2. To Improve Breeding

by *sequencing more strains/cultivars*



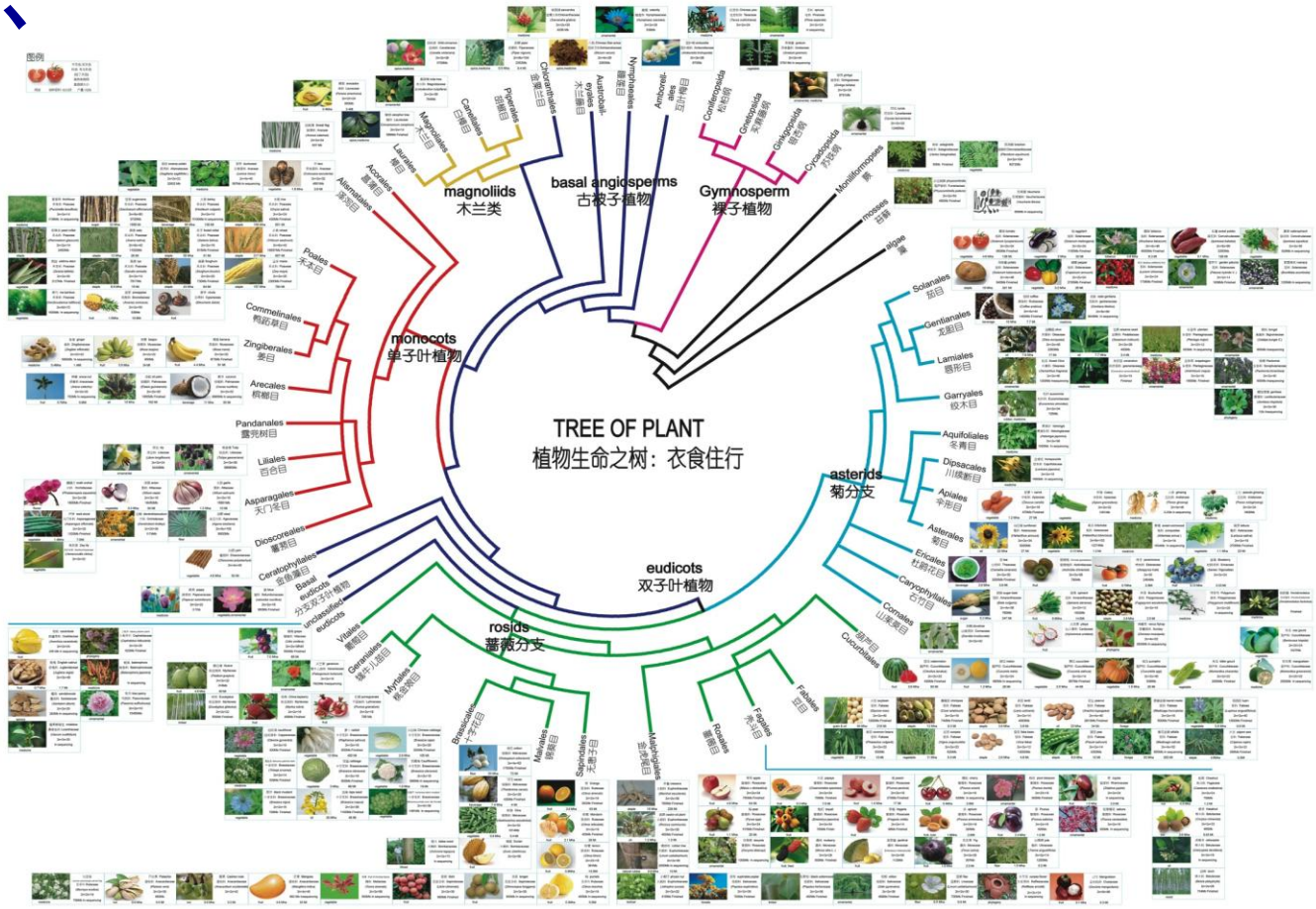
Omics-based digital breeding of crops & livestock

农业资源

14种粮食作物

70种园艺作物

39种禽畜牧渔



BGI & its partners contribute ~ 80% of sequences for agro germ plasma resources

3. To Improve Health Care

by *sequencing more individuals*

To sequence everybody in the world

“人人基因组”

The 1 M Chinese Genomes Project

5P+IT Medicine in the BioCentury

Prediction

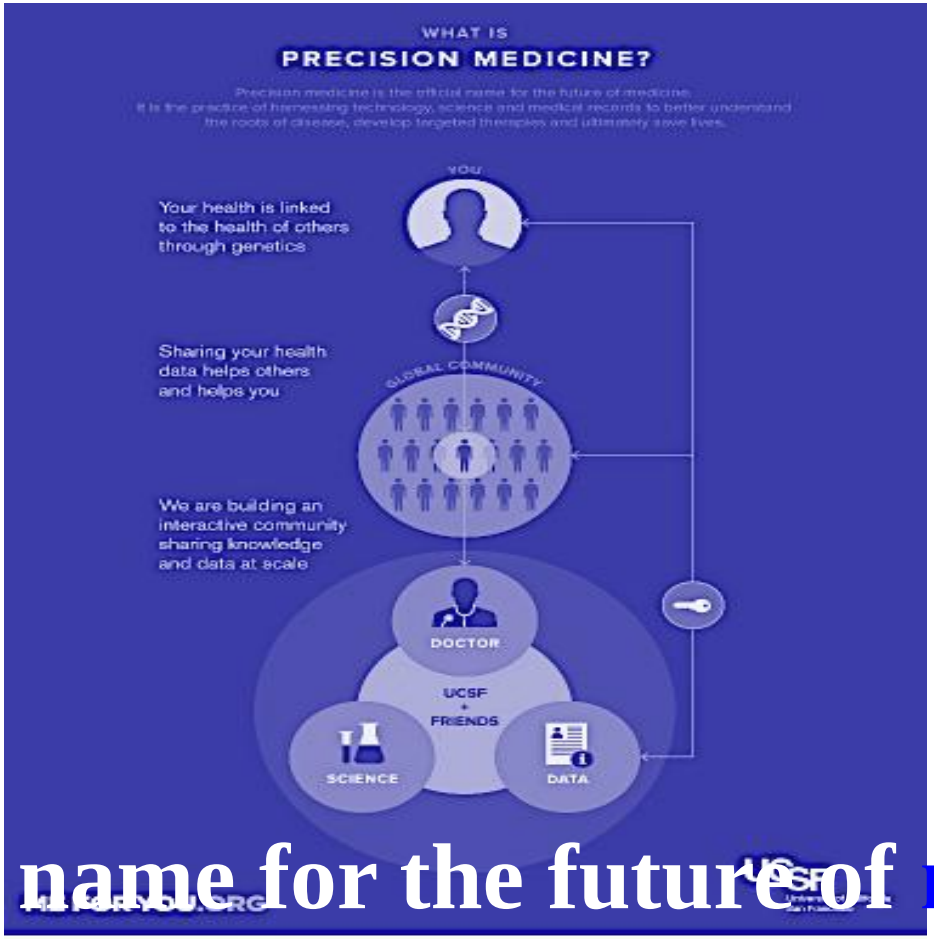
Prevention

Participation

Personalization

5P+IT Medicine in the BioCentury

Precision



"The official name for the future of medicine"

5P+IT Medicine in the BioCentury

Integrated

5P+IT Medicine in the BioCentury

Targeted

5P+IT Medicine in the BioCentury

***P*recision**	精准
Prediction	预测
Prevention	预防
Participation	参与
Personalization	个性化
Integrated	整合
Targeted	靶

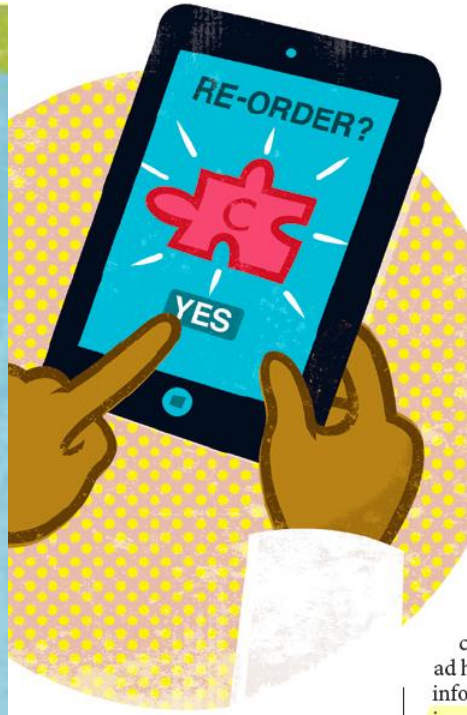
5P+IT Medicine in the BioCentury

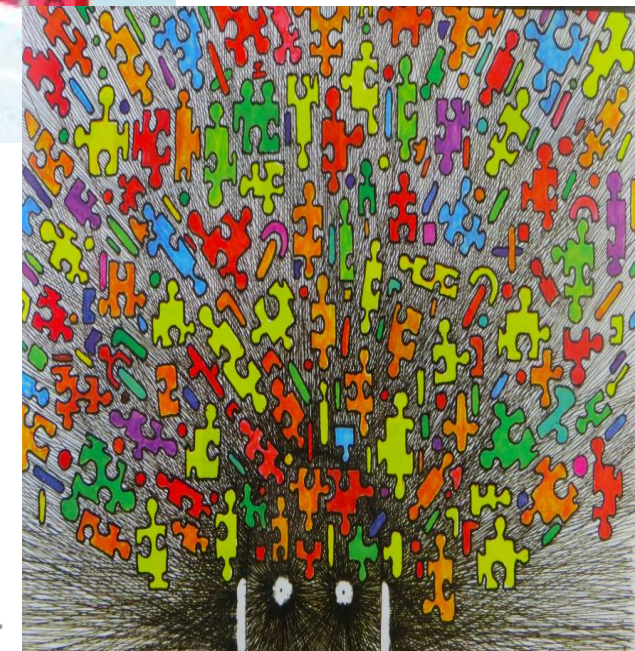
Precision	精准
Prediction	预测
Prevention	预防
Participation	参与
Personalization	个性化
Integrated	整合
Targeted	靶

5P+IT Medicine in the BioCentury

Personalized Medicine
is
Omics/Data-Based

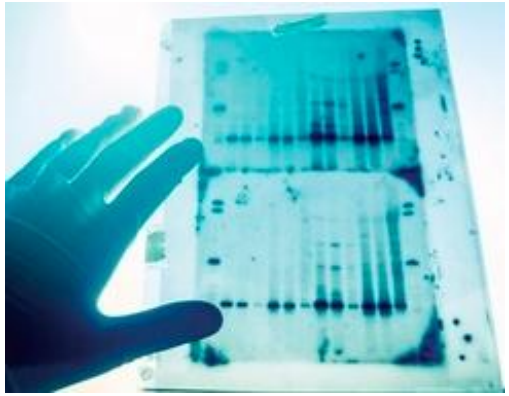
Personalization 个性化





Gather and use genetic data in health care

Research into how genetic variants can guide successful treatments must become part of routine medical practice and records, says **Geoffrey Ginsburg**.

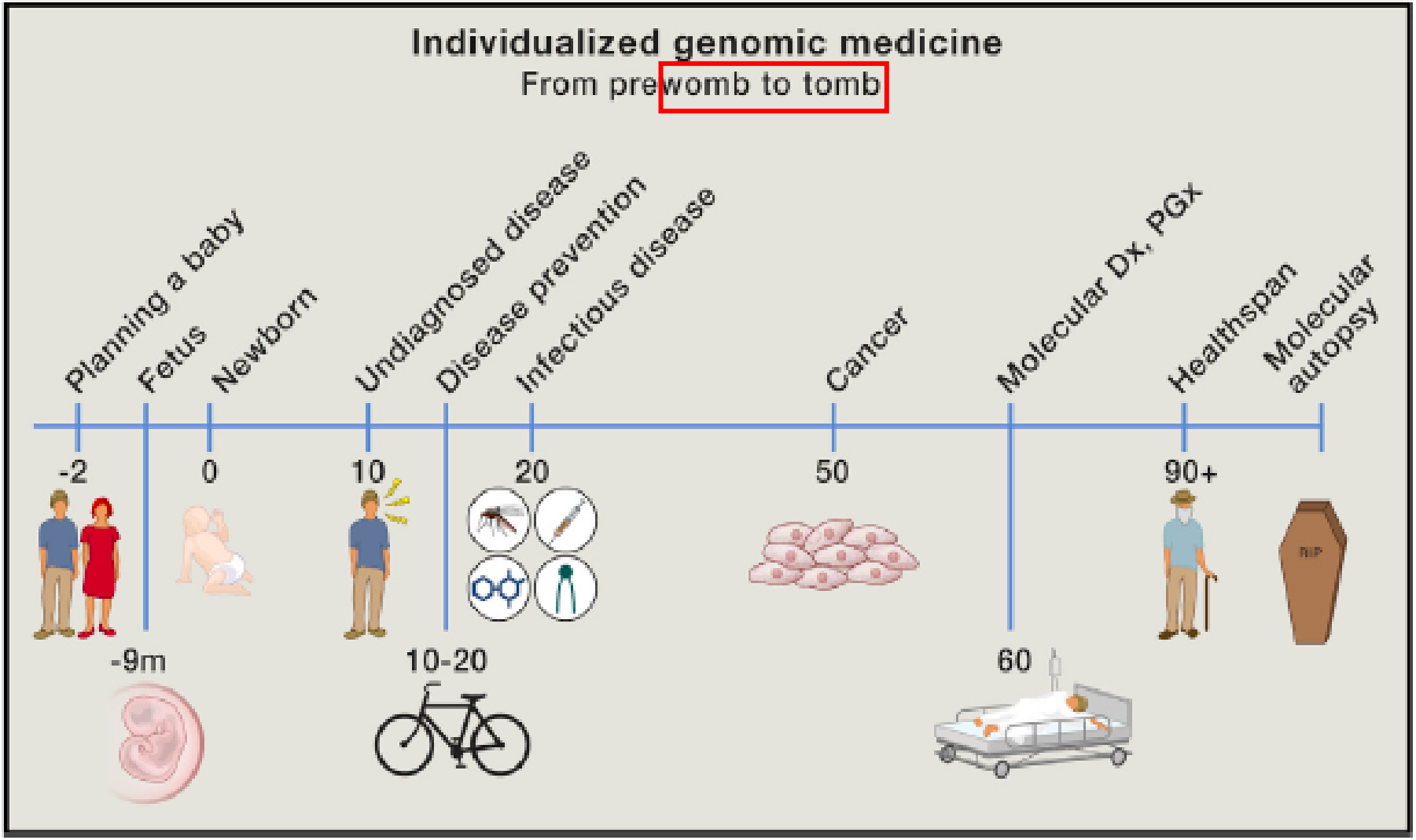


“In not too many years, most genetic diagnosis services will be done by sequencing the genome.”

Heidi Rehm, chief laboratory director at the Partners Laboratory for Molecular Medicine, USA

U.S. spending on genetic testing and molecular diagnostics, including about 1,000 targeted genetic tests, is expected to reach between $15 billion and $25 billion by 2021, compared with about $5 billion in 2010, according to a report released last year by UnitedHealth Group.

Omics in the Big Data Era

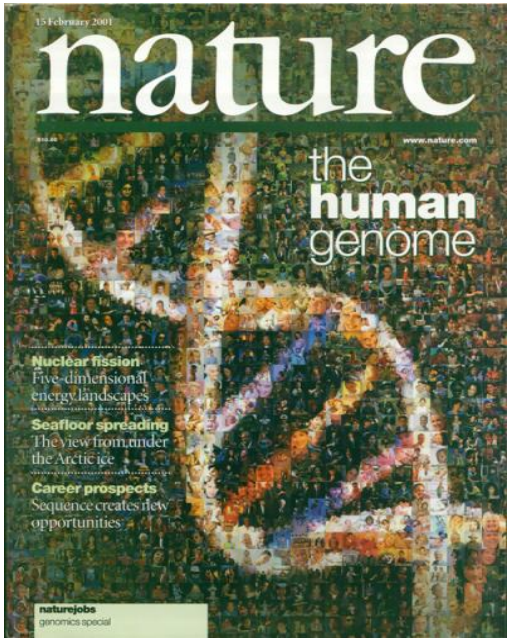


From womb to tomb …

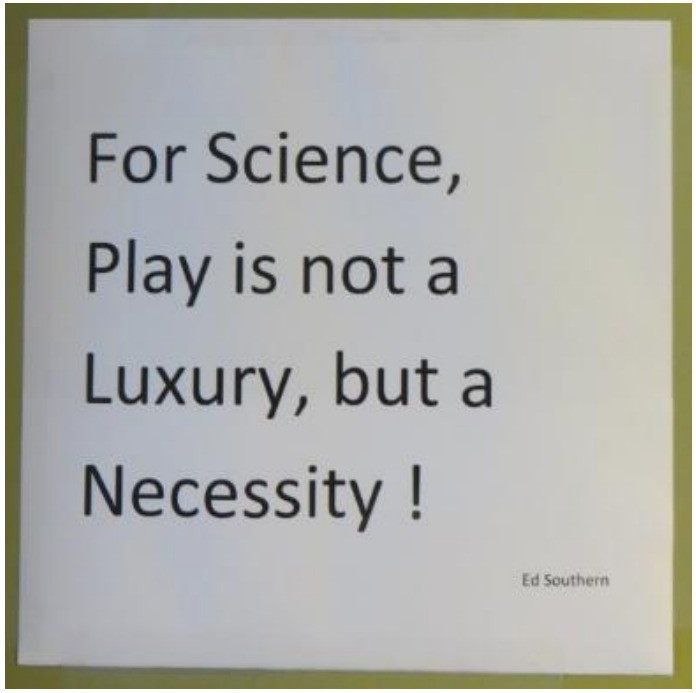
Outline:

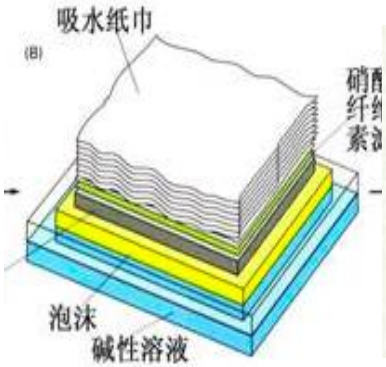
1. A later comer of a revolution
2. Two pillars of genomics
3. Three impacts of the HGP
4. **Four applications to medicine**
5. Five techs changing the world

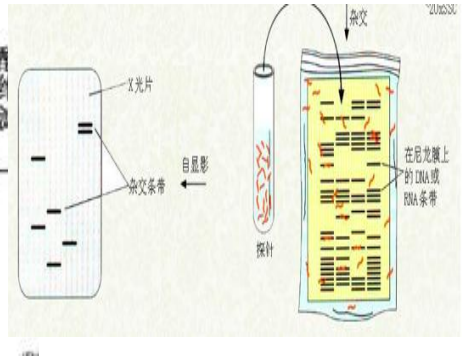
Four “Directions” of Sequencing



Southern Sequencing (DNA-Seq)
Northern Sequencing (RNA-Seq)
Western Sequencing
Eastern Sequencing

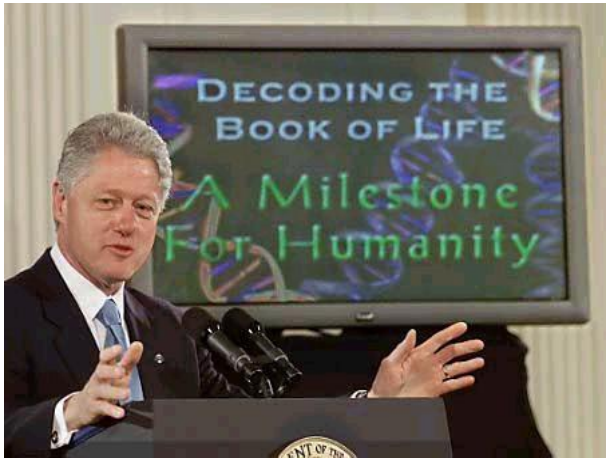




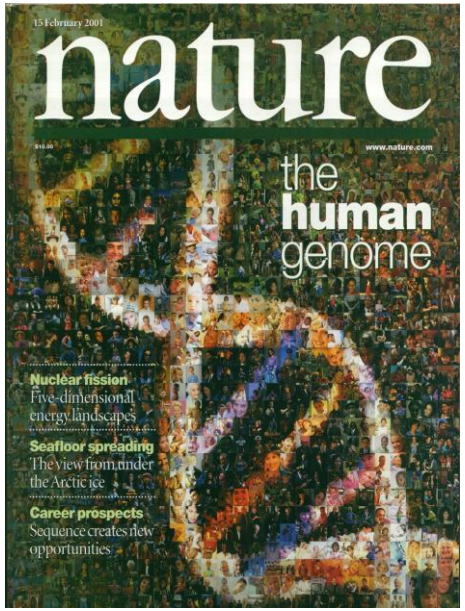


To acknowledge great contributions by

Prof. E. Southern



"A Decade Later, Genetic Map Yields Few New Cures," said a New York Times headline in June 2010. ***It declared the failure of the $3 billion Human Genome Project and claimed that medicine had seen none of the benefits*** that Bill Clinton had promised in announcing the first draft of the human-genome sequence in 2000.



4 Years Ago

4/15 Years Later

The New York Times judged the project too soon.

“The triumph of genomic medicine is just beginning.”

Washington Post March 13, 2014

Four More Applications

1. **WE/TR Sequencing** for monogenic diseases
   WG for cancers and other complex diseases

2. **Metagenome Sequencing**
   for metabolic/infectious diseases,
   pathogen/host, microbiology, & ecology

3. **Single Cell Sequencing**
   for cancers, meta-, neurology, & development

4. **Trace DNA sequencing**
   for NIPT, PIT, & other early diagnosis,
   ancientDNA/evolution

Four More Applications

1. **WE/TR Sequencing for monogenic diseases**
   WG for cancers and other complex diseases

2. **Metagenome Sequencing**
   for metabolic/infectious diseases,
   pathogen/host, microbiology, & ecology

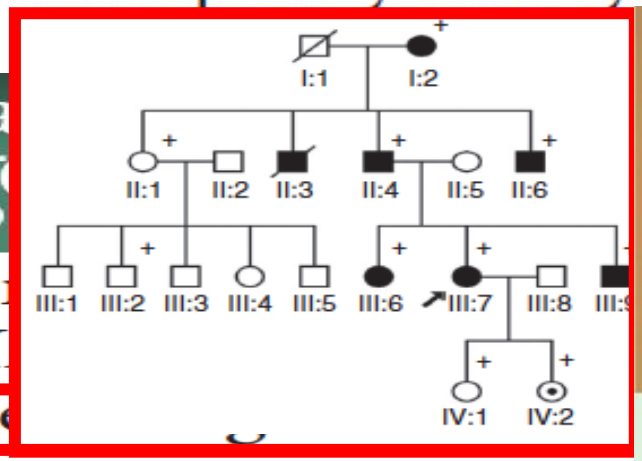
3. **Single Cell Sequencing**
   for cancers, meta-, neurology, & development

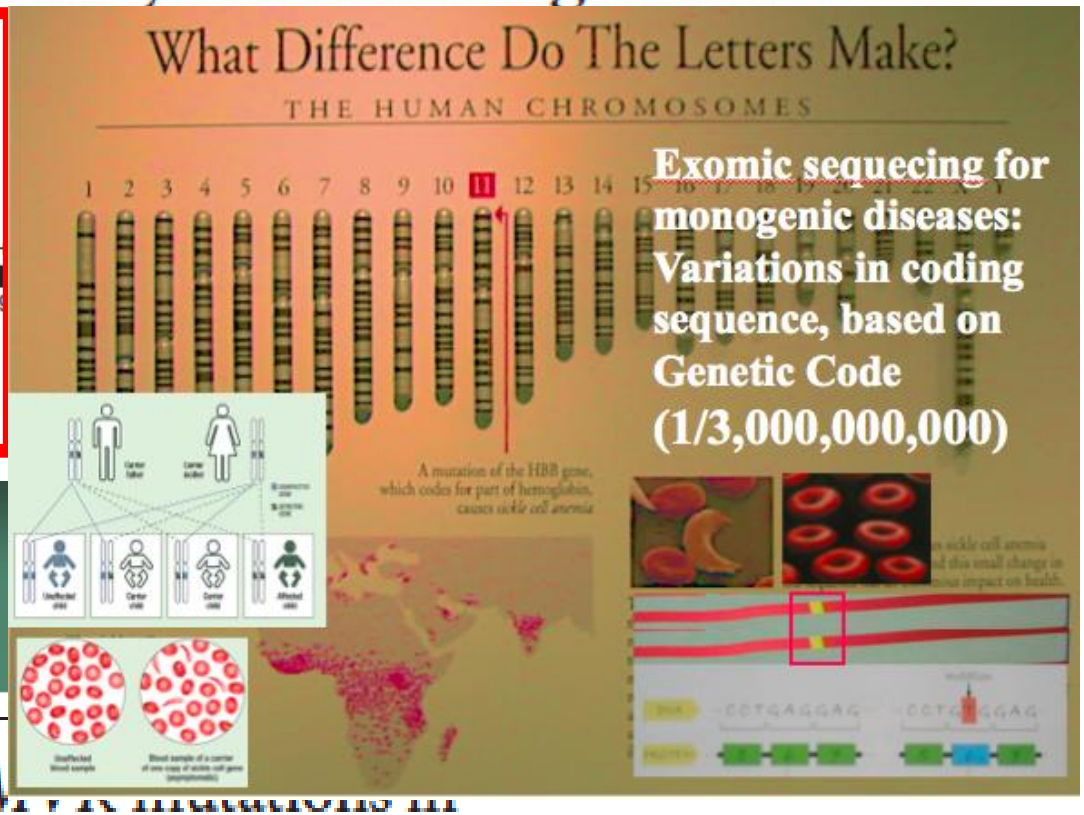
4. **Trace DNA sequencing**
   for NIPT, PIT, & other early diagnosis,
   ancientDNA/evolution

nature genetics

online, Oct. 3, 2010

Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants





nature genetics

Exome sequencing identifies *MYH* mutations in disseminated superficial actinic porokeratosis

A single pedigree

（已开展0336种单基因病临床检测服务；另1500种单基因病检测项目正在开发）

单基因病检测

Genetic testing of monogenic diseases

单基因病

Monogenic Disease

人体因单个基因缺陷所引发的疾病，缺陷能遗传；如地中海贫血、苯丙酮尿症、血友病等。

Congenital diseases that result from the abnormality of a single body. e.g. thalassemia, phenylketonuria, hemophilia etc.

（已完成05631人份病例检测）

幸福“两广”——地中海贫血基因检测

Genetic Testing of Thalassemia

地中海贫血是一种溶血性可遗传单基因病。目前，造血干细胞移植是目前根治本病的唯一方法。

Thalassemia is a hemolytic and inheritable monogenic disease. At the moment, stem cell transplantation is the only radical treatment to this disease.

、广西、海南省发病率最高。

g thalassemia gene, with a high incidence

nan Province of China.

有1人携带地贫基因，每年出生约

in Guangdong Province, with an average

thalassemia are born every year.

Table 1. OMIM Phenotypes for which the Molecular Basis Is Known, 2007 and 2013

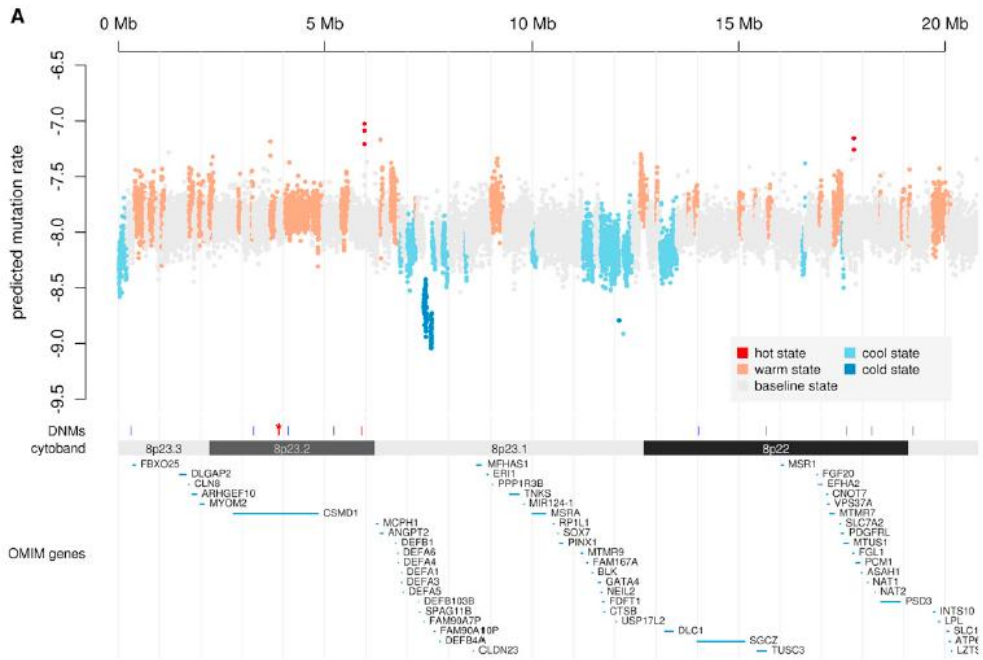
Inheritance Pattern	January 2007	July 2013
Autosomal	1,851	3,525
X Linked	169	277
Y Linked	2	4
Mitochondrial	26	28
Total	2,048	3,834

Whole-Genome Sequencing in Autism Identifies Hot Spots for De Novo Germline Mutation

Cell

151: 1431, 2012

Jacob J. Michaelson,[1,2,14] Yujian Shi,[5,14] Madhusudan Gujral,[1,2,14] Har
Xin Jin,[5,6,14] Minghan Jian,[5] Guangming Liu,[7,8] Douglas Greer,[1,2] Abhis
Roser Corominas,[2] Áine Peoples,[1,2,9] Amnon Koren,[10] Athurva Gore,[4]
Therese Gadomski,[2] Balvindar Singh,[1,2] Kun Zhang,[4] Natacha Akshoc
Lilia M. Iakoucheva,[2] Yingrui Li,[5] Jun Wang,[5,12,13,*] and Jonathan Seb



SUMMARY

De novo mutation plays an important role in autism spectrum disorders (ASDs). Notably, pathogenic copy number variants (CNVs) are characterized by high mutation rates. We hypothesize that hypermutability is a property of ASD genes and may also include nucleotide-substitution hot spots. We investigated global patterns of germline mutation by whole-genome sequencing of monozygotic twins concordant for ASD and their parents. Mutation rates varied widely throughout the genome (by 100-fold) and could be explained by intrinsic characteristics of DNA sequence and chromatin structure. Dense clusters of mutations within individual genomes were attributable to compound mutation or gene conversion. Hypermutability was a characteristic of genes involved in ASD and other diseases. In addition, genes impacted by mutations in this study were associated with ASD in independent exome-sequencing data sets. Our findings suggest that regional hypermutation is a significant factor shaping patterns of genetic variation and disease risk in humans.

nature genetics online, May 27, 2011

Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma

nature genetics 44:17-19, 2012

Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma

nature genetics online Dec 23, 2012

Exome sequencing identifies mutation in *CNOT3* and ribosomal genes *RPL5* and *RPL10* in T-cell acute lymphoblastic leukemia

nature genetics 43:875-878., 2011

Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder

nature genetics Oct. 13, 2013

Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation

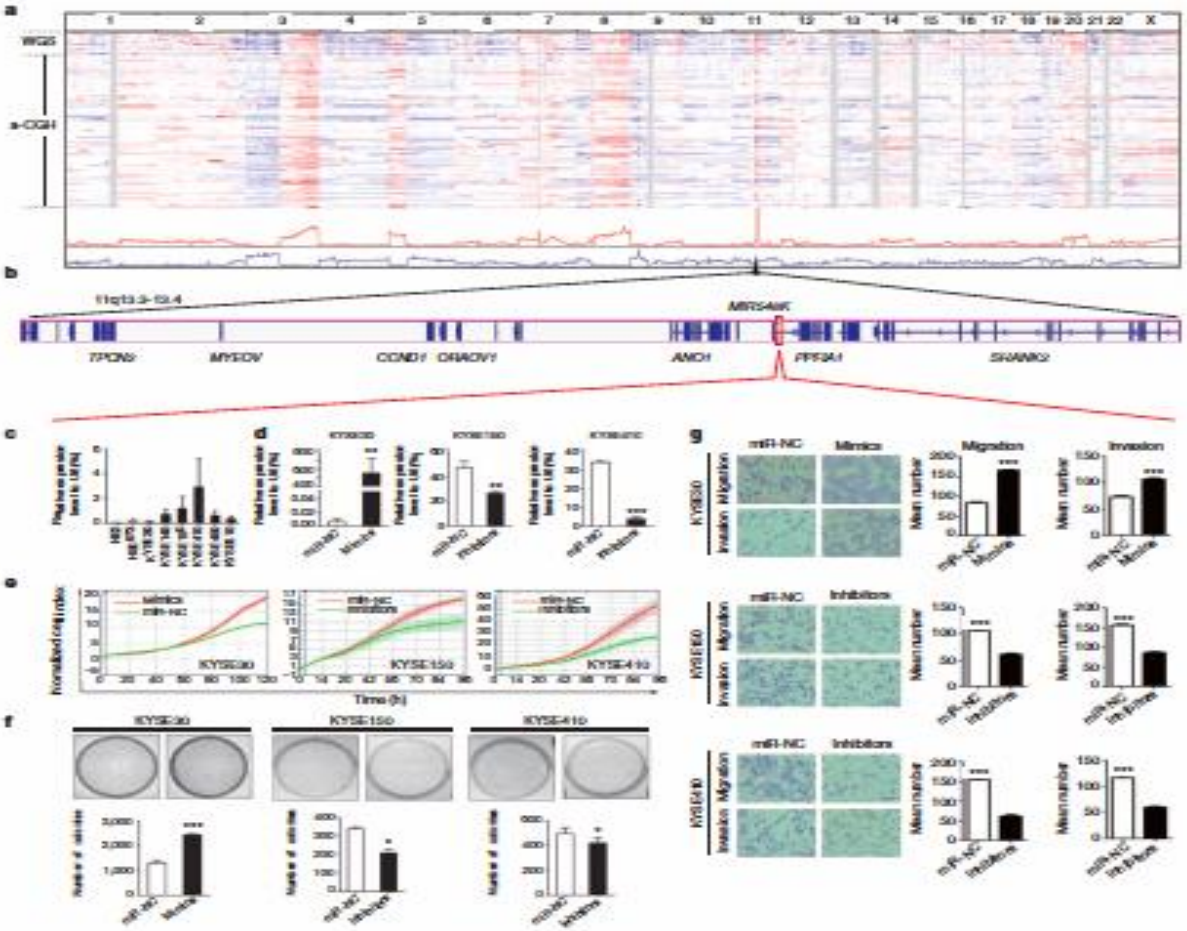
LETTER Online 16 March, 2014

nature

Identification of genomic alterations in oesophageal squamous cell cancer

Yongmei Song[1*], Lin Li[2*], Yunwei Ou[1,3*], Zhibo Gao[2*], Enmin Li[4*], Xiangchun Li[2*], Weimin Zhang[1], Jiaqian Wang[2], Liyan Xu[5], Yong Zhou[2], Xiaojuan Ma[1], Lingyan Liu[1], Zitong Zhao[1], Xuanlin Huang[2], Jing Fan[1], Lijia Dong[1], Gang Chen[2], Liying Ma[1], Jie Yang[2], Longyun Chen[2], Minghui He[2], Miao Li[2], Xuehan Zhuang[2], Kai Huang[2], Kunlong Qiu[2], Guangliang Yin[2], Guangwu Guo[2], Qiang Feng[2], Peishan Chen[2], Zhiyong Wu[6], Jianyi Wu[4], Ling Ma[1], Jinyang Zhao[2], Longhai Luo[2], Ming Pu[1], Bainan Xu[3], Bo Chen[5], Yingrui Li[2], Tong Tong[1], Mingrong Wang[1], Zhihua Liu[1], Dongxin Lin[1], Xiuqing Zhang[2], Huanming Yang[2], Jun Wang[2] & Qimin Zhan[1]

Oesophageal cancer is one of the most aggressive cancers and is the sixth leading cause of cancer death worldwide[1]. Approximately 70% of global oesophageal cancer cases occur in China, with oesophageal squamous cell carcinoma (ESCC) being the histopathological form in the vast majority of cases (>90%)[2,3]. Currently, there are limited clinical approaches for the early diagnosis and treatment of ESCC, resulting in a 10% five-year survival rate for patients. However, the full repertoire of genomic events leading to the pathogenesis of ESCC remains unclear. Here we describe a comprehensive genomic analysis of 158 ESCC cases, as part of the International Cancer Genome Consortium research project. We conducted whole-genome sequencing in 17 ESCC cases and whole-exome sequencing in 71 cases, of which 53 cases, plus an additional 70 ESCC cases not used in the whole-genome and whole-exome sequencing, were subjected to array comparative genomic hybridization analysis. We identified eight significantly mutated genes, of which six are well known tumour-associated genes (*TP53*, *RB1*, *CDKN2A*, *PIK3CA*, *NOTCH1*, *NFE2L2*), and two have not previously been described in ESCC (*ADAM29* and *FAM135B*). Notably, *FAM135B* is identified as a novel cancer-implicated gene as assayed for its ability to promote malignancy of ESCC cells. Additionally, *MIR548K*, a microRNA encoded in the amplified 11q13.3-13.4 region, is characterized as a novel oncogene, and functional assays demonstrate that *MIR548K* enhances malignant phenotypes of ESCC cells. Moreover, we have found that several important histone regulator genes (*MLL2* (also called *KMT2D*), *ASH1L*, *MLL3* (*KMT2C*), *SETD1B*, *CREBBP* and *EP300*) are frequently altered in ESCC. Pathway assessment reveals that somatic aberrations are mainly involved in the Wnt, cell cycle and Notch pathways. Genomic analyses suggest that ESCC and head and neck squamous cell carcinoma share some common pathogenic mechanisms, and ESCC development is associated with alcohol drinking. This study has explored novel biological markers and tumorigenic pathways that would greatly improve therapeutic strategies for ESCC.



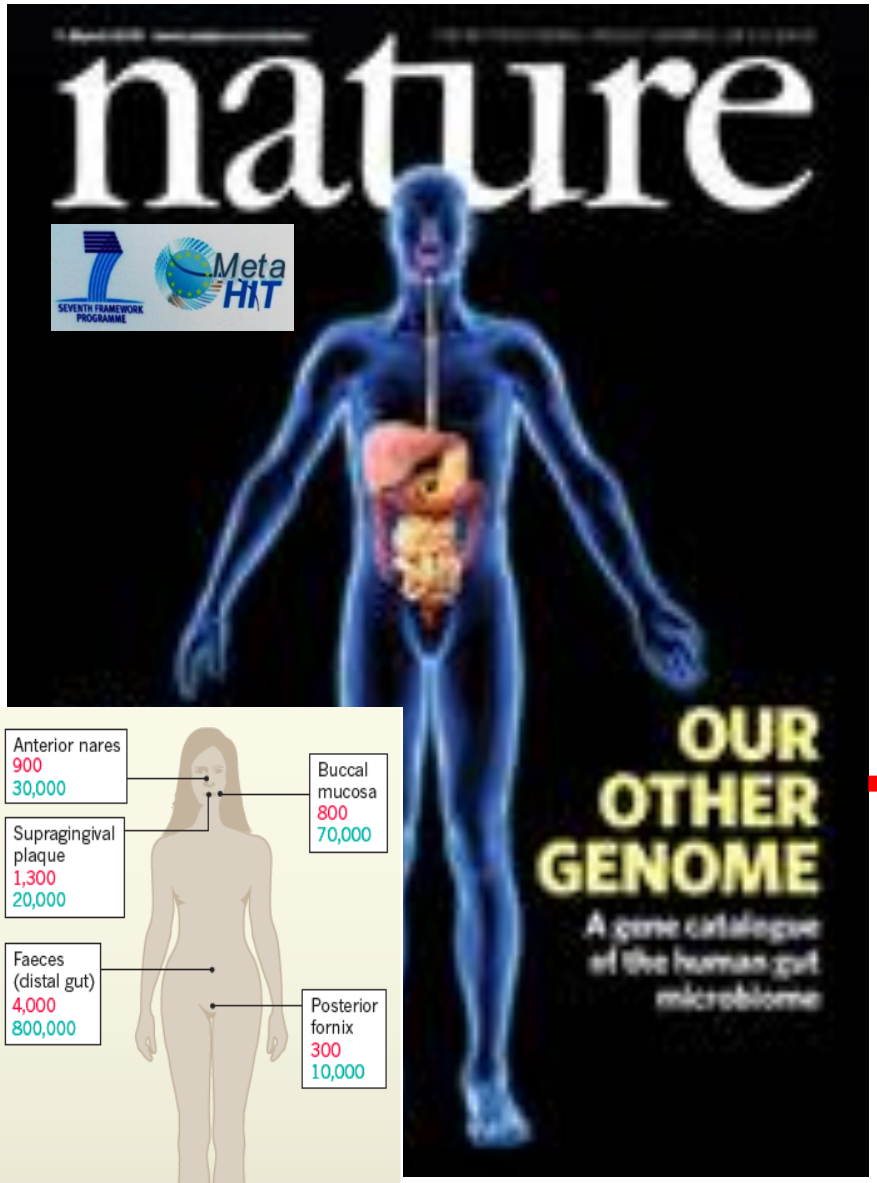
Four More Applications

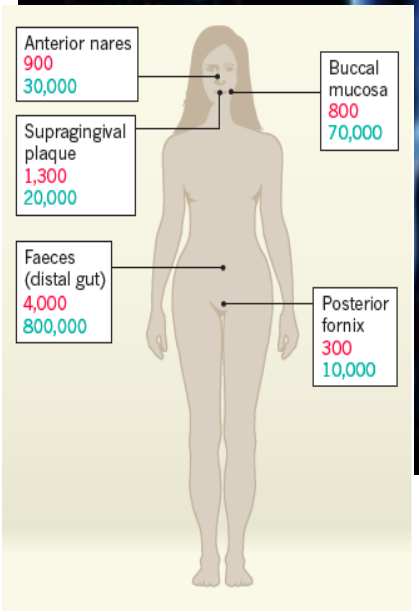
1. **WE/TR Sequencing** for monogenic diseases
   WG for cancers and other complex diseases

2. **Metagenome Sequencing**
   for metabolic/infectious diseases,
   pathogen/host, microbiology, & ecology

3. **Single Cell Sequencing**
   for cancers, meta-, neurology, & development

4. **Trace DNA sequencing**
   for NIPT, PIT, & other early diagnosis,
   ancientDNA/evolution



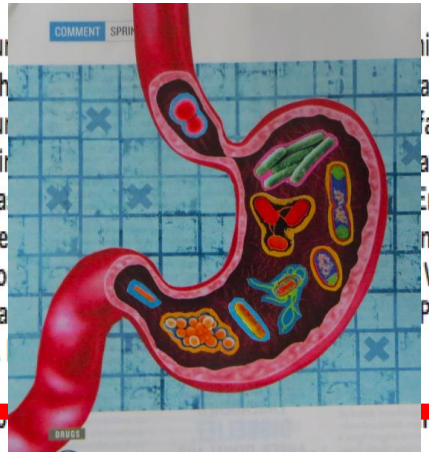


Human Metagenomes

ARTICLE

March 4, 2010

A human gut microbial gene catalogue established by metagenomic sequencing



ARTICLE

473: 174, 2011

doi:10.1038/nature09944

Enterotypes of the human gut microbiome

Manimozhiyan Arumugam[1]*, Jeroen Raes[1,2]*, Eric Pelletier[3,4,5], Denis Le Paslier[3,4,5], Takuji Yamada[1], Daniel R. Mende[1], Gabriel R. Fernandes[1,6], Julien Tap[1,7], Thomas Bruls[3,4,5], Jean-Michel Batto[7], Marcelo Bertalan[8], Natalia Borruel[9], Francesc Casellas[9], Leyden Fernandez[10], Laurent Gautier[8], Torben Hansen[11,12], Masahira Hattori[13], Tetsuya Hayashi[14], Michiel Kleerebezem[15], Ken Kurokawa[16], Marion Leclerc[7], Florence Levenez[7], Chaysavanh Manichanh[9], H. Bjørn Nielsen[8], Trine Nielsen[11], Nicolas Pons[7], Julie Poulain[3], Junjie Qin[17], Thomas Sicheritz-Ponten[8,18], Sebastian Tims[15], David Torrents[10,19], Edgardo Ugarte[3], Erwin G. Zoetendal[15], Jun Wang[17,20], Francisco Guarner[9], Oluf Pedersen[11,21,22,23], Willem M. de Vos[15,24], Søren Brunak[8], Joel Doré[7], MetaHIT Consortium†, Jean Weissenbach[3,4,5], S. Dusko Ehrlich[7] & Peer Bork[1,25]

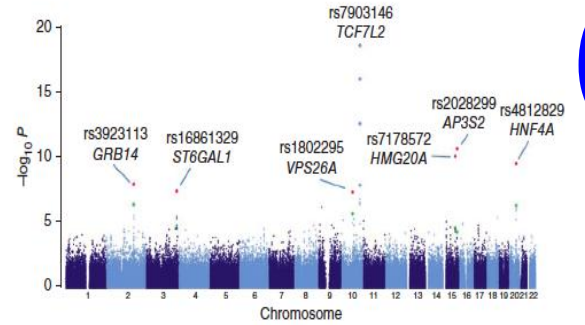
Our knowledge of species and functional composition of the human gut microbiome is rapidly increasing, but it is still based on very few cohorts and little is known about variation across the world. By combining 22 newly sequenced faecal metagenomes of individuals from four countries with previously published data sets, here we identify three robust clusters (referred to as enterotypes hereafter) that are not nation or continent specific. We also confirmed the enterotypes in two published, larger cohorts, indicating that intestinal microbiota variation is generally stratified, not continuous. This indicates further the existence of a limited number of well-balanced host–microbial symbiotic states that might respond differently to diet and drug intake. The enterotypes are mostly driven by species composition, but abundant molecular functions are not necessarily provided by abundant species, highlighting the importance of a functional analysis to understand microbial communities. Although individual host properties such as body mass index, age, or gender cannot explain the observed enterotypes, data-driven marker genes or functional modules can be identified for each of these host properties. For example, twelve genes significantly correlate with age and three functional modules with the body mass index, hinting at a diagnostic potential of microbial markers.

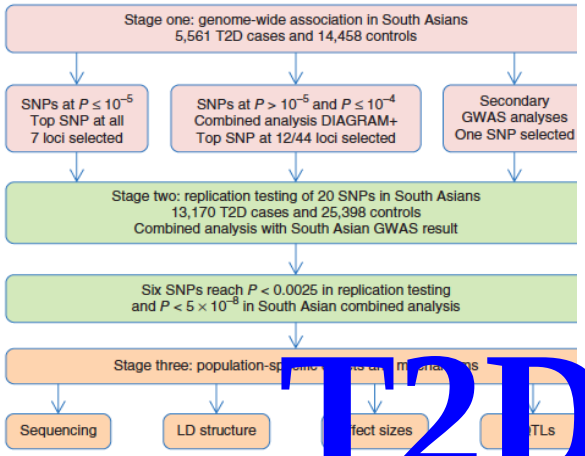
nature

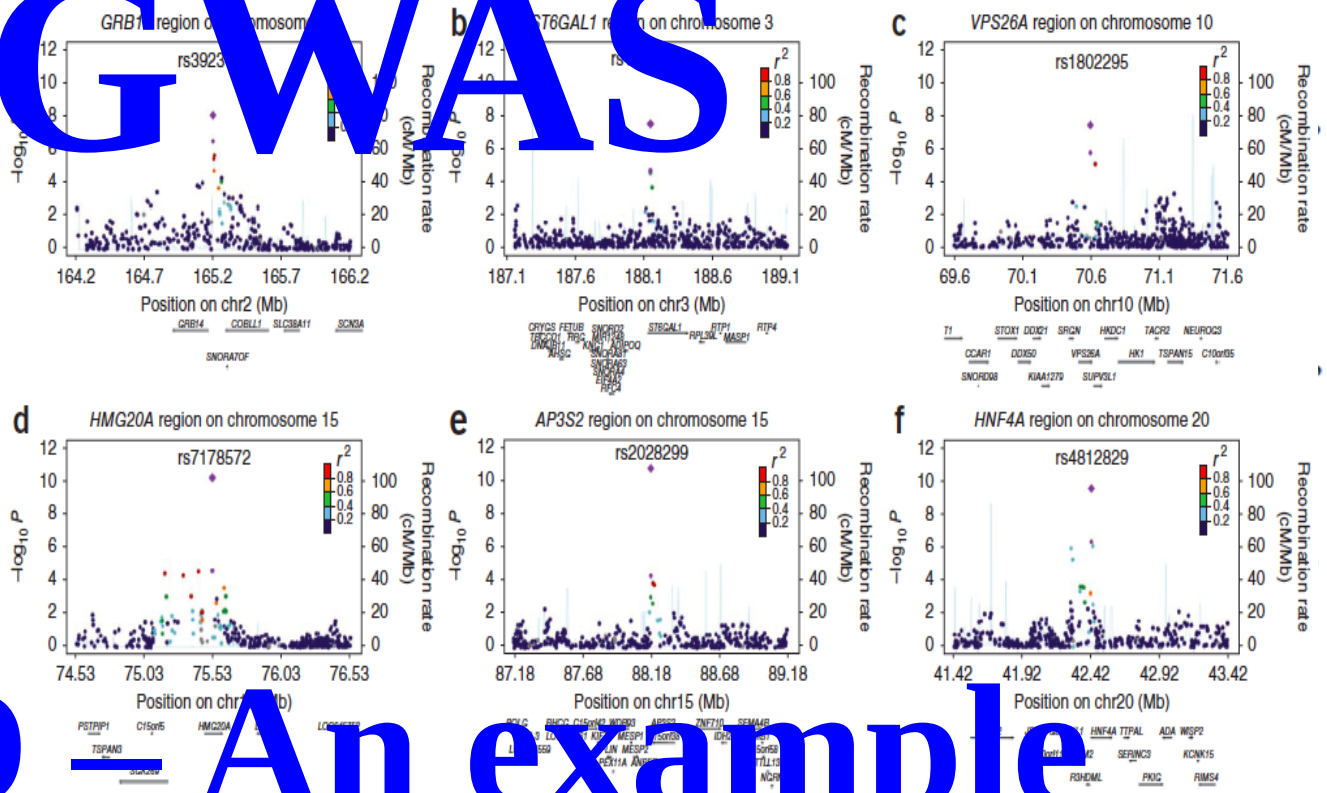
online Aug. 28, 2011

Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci

GWAS







T2D – An example

June, 2014

nature

A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes

Ida Moltke[1,2]*, Niels Grarup[3]*, Marit E. Jørgensen[4], Peter Bjerregaard[5], Jonas T. Treebak[6], Matteo Fumagalli[7], Thorfinn S. Korneliussen[8], Marianne A. Andersen[6], Thomas S. Nielsen[6], Nikolaj T. Krarup[3], Anette P. Gjesing[3], Juleen R. Zierath[6,9], Allan Linneberg[10], Xueli Wu[11], Guangqing Sun[11], Xin Jin[11], Jumana Al-Aama[11,12], Jun Wang[3,11,12,13,14], Knut Borch-Johnsen[15], Oluf Pedersen[3], Rasmus Nielsen[7,16], Anders Albrechtsen[1] & Torben Hansen[3,17]

The Greenlandic population, a small and historically isolated founder population comprising about 57,000 inhabitants, has experienced a dramatic increase in type 2 diabetes (T2D) prevalence during the past 25 years[1]. Motivated by this, we performed association mapping of T2D-related quantitative traits in up to 2,575 Greenlandic individuals without known diabetes. Using array-based genotyping and exome sequencing, we discovered a nonsense p.Arg684Ter variant (in which arginine is replaced by a termination codon) in the gene *TBC1D4* with an allele frequency of 17%. Here we show that homozygous carriers of this variant have markedly higher concentrations of plasma glucose ($\beta = 3.8$ mmol l^{-1}, $P = 2.5 \times 10^{-35}$) and serum insulin ($\beta = 165$ pmol l^{-1}, $P = 1.5 \times 10^{-20}$) 2 hours after an oral glucose load compared with individuals with other genotypes (both non-carriers and heterozygous carriers). Furthermore, homozygous carriers have marginally lower concentrations of fasting plasma glucose ($\beta = -0.18$ mmol l^{-1}, $P = 1.1 \times 10^{-6}$) and fasting serum insulin ($\beta = -8.3$ pmol l^{-1}, $P = 0.0014$), and their T2D risk is markedly increased (odds ratio (OR) = 10.3, $P = 1.6 \times 10^{-24}$). Heterozygous carriers have a moderately higher plasma glucose concentration 2 hours after an oral glucose load than non-carriers ($\beta = 0.43$ mmol l^{-1}, $P = 5.3 \times 10^{-5}$). Analyses of skeletal muscle biopsies showed lower messenger RNA and protein levels of the long isoform of TBC1D4, and lower muscle protein levels of the glucose transporter GLUT4, with increasing number of p.Arg684Ter alleles. These findings are concomitant with a severely decreased insulin-stimulated glucose uptake in muscle, leading to postprandial hyperglycaemia, impaired glucose tolerance and T2D. The observed effect sizes are several times larger than any previous findings in large-scale genome-wide association studies of these traits[2–4] and constitute further proof of the value of conducting genetic association studies outside the traditional setting of large homogeneous populations.

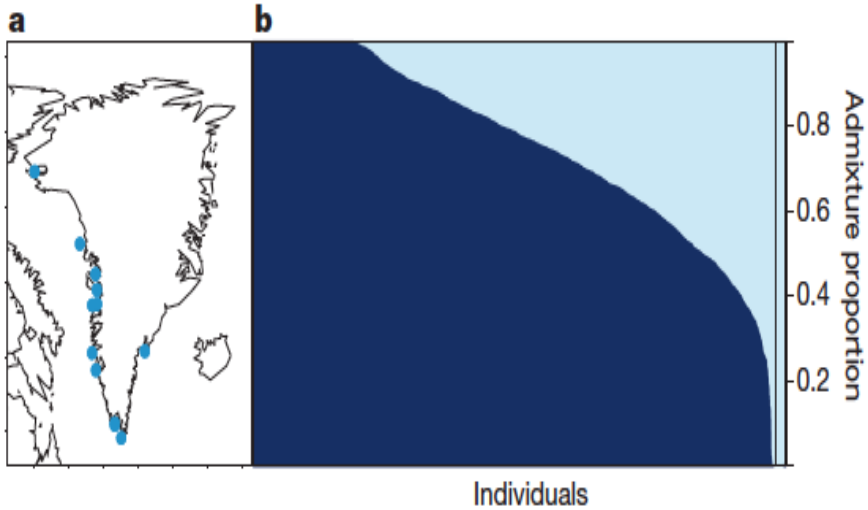


Figure 1 | Greenlandic study population. a, Sampling locations in Greenland. b, Estimated admixture proportions of Inuit and European ancestry. The admixture proportions were estimated assuming two source populations ($K = 2$). The estimates are both for the 2,733 individuals in the Greenlandic sample (IHIT), depicted to the left of the vertical line, and for 50 Danes, to the right of the vertical line.

online Sept.26, 2012

nature

A metagenome-wide association study of gut microbiota in type 2 diabetes

MWAS

Junjie Qin[1]*, Yingrui Li[1]*, Zhiming Cai[2]*, Shenghui Li[1]*, Jianfeng Zhu[1]*, Fan Zhang[3]*, Suisha Liang[1], Wenwei Zhang[1], Yuanlin Guan[1], Dongqian Shen[1], Yang[illegible] Peng[illegible] Zhang[1], Zhuye [illegible] Wen[illegible] Wu[1], Youwen Qin[1], Wenbin Xue[1], Junhua Li[1], Lingchuan Han[3], Donghui [illegible] Pei[illegible] Wu[illegible] Ali [illegible], Xiaoju[illegible] [illegible]n[2], Z[illegible]ng Li[2], Aifa Tang[2], Shilong Zhong[4], Xiaoping Li[1], Weineng Chen[1], Ran Xu[1], M[illegible] Wang [illegible] [illegible]g[1], [illegible] Gong, [illegible] Yu[1], Yanyan Zhang[1], Ming Zhang[1], Torben Hansen[5], Gaston Sanchez[6], Jer[illegible]en [illegible]es[7], [illegible] Gwen [illegible]ony [illegible] Shu[illegible]iro O[illegible]da[illegible], Ma[illegible]ieu Almeida[9], Emmanuelle LeChatelier[9], Pierre Renault[9], Nicolas Pons[9], Jean-Michel Batto[9], Zhaoxi Zhang[1], Hua Chen[1], Ruifu Yang[1,10], Weimou Zheng[1], Songgang Li[1], Huanming Yang[1], Jian Wang[1], S. Dusko Ehrlich[9], Rasmus Nielsen[6], Oluf Pedersen[5,11,12], Karsten Kristiansen[1,13] & Jun Wang[1,5,13]

Assessment and characterization of gut microbiota has become a major research area in human disease, including type 2 diabetes, the most prevalent endocrine disease worldwide. To carry out analysis on gut microbial content in patients with type 2 diabetes, we developed a protocol for a metagenome-wide association study (MGWAS) and undertook a two-stage MGWAS based on deep shotgun sequencing of the gut microbial DNA from 345 Chinese individuals. We identified and validated approximately 60,000 type-2-diabetes-associated markers and established the concept of a metagenomic linkage group, enabling taxonomic species-level analyses. MGWAS analysis showed that patients with type 2 diabetes were characterized by a moderate degree of gut microbial dysbiosis, a decrease in the abundance of some universal butyrate-producing bacteria and an increase in various opportunistic pathogens, as well as an enrichment of other microbial functions conferring sulphate reduction and oxidative stress resistance. An analysis of 23 additional individuals demonstrated that these gut microbial markers might be useful for classifying type 2 diabetes.

ARTICLE Aug. 29, 2013 nature

Richness of human gut microbiome correlates with metabolic markers

Emmanuelle Le Chatelier[1*], Trine Nielsen[2*], Junjie Qin[3*], Edi Prifti[1*], Falk Hildebrand[4,5], Gwen Falony[4,5], Mathieu Almeida[1], Manimozhiyan Arumugam[2,3,6], Jean-Michel Batto[1], Sean Kennedy[1], Pierre Leonard[1], Junhua Li[3,7], Kristoffer Burgdorf[2], Niels Grarup[2], Torben Jørgensen[8,9,10], Ivan Brandslund[11,12], Henrik Bjørn Nielsen[13], Agnieszka S. Juncker[13], Marcelo Bertalan[13], Florence Levenez[1], Nicolas Pons[1], Simon Rasmussen[13], Shinichi Sunagawa[6], Julien Tap[1,6], Sebastian Tims[14], Erwin G. Zoetendal[14], Søren Brunak[13], Karine Clément[15,16,17], Joël Doré[1,18], Michiel Kleerebezem[14], Karsten Kristiansen[19], Pierre Renault[18], Thomas Sicheritz-Ponten[13], Willem M. de Vos[14,20], Jean-Daniel Zucker[15,16,21], Jeroen Raes[4,5], Torben Hansen[2,22], MetaHIT consortium†, Peer Bork[6], Jun Wang[3,19,23,24,25], S. Dusko Ehrlich[1] & Oluf Pedersen[2,26,27,28]

We are facing a global metabolic health crisis provoked by an obesity epidemic. Here we report the human gut microbial composition in a population sample of 123 non-obese and 169 obese Danish individuals. We find two groups of individuals that differ by the number of gut microbial genes and thus gut bacterial richness. They contain known and previously unknown bacterial species at different proportions; individuals with a low bacterial richness (23% of the population) are characterized by more marked overall adiposity, insulin resistance and dyslipidaemia and a more pronounced inflammatory phenotype when compared with high bacterial richness individuals. The obese individuals among the lower bacterial richness group also gain more weight over time. Only a few bacterial species are sufficient to distinguish between individuals with high and low bacterial richness, and even between lean and obese participants. Our classifications based on variation in the gut microbiome identify subsets of individuals in the general white adult population who may be at increased risk of progressing to adiposity-associated co-morbidities.

nature COMMUNICATIONS

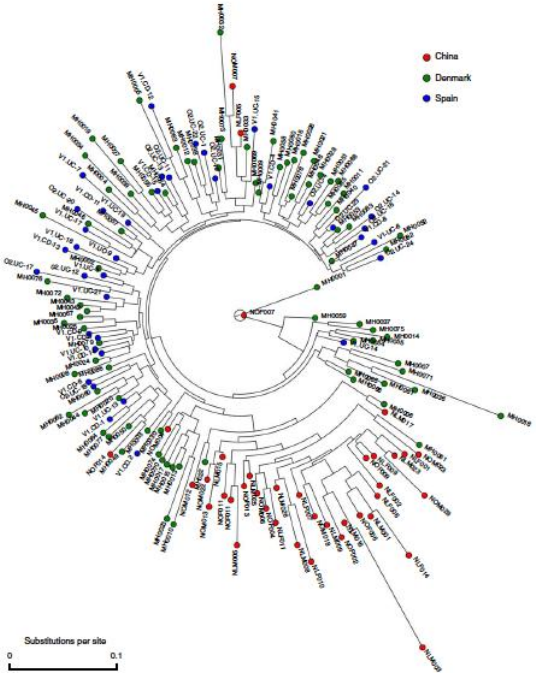
ARTICLE

Received 21 Feb 2013 | Accepted 13 Jun 2013 | Published 23 Jul 2013

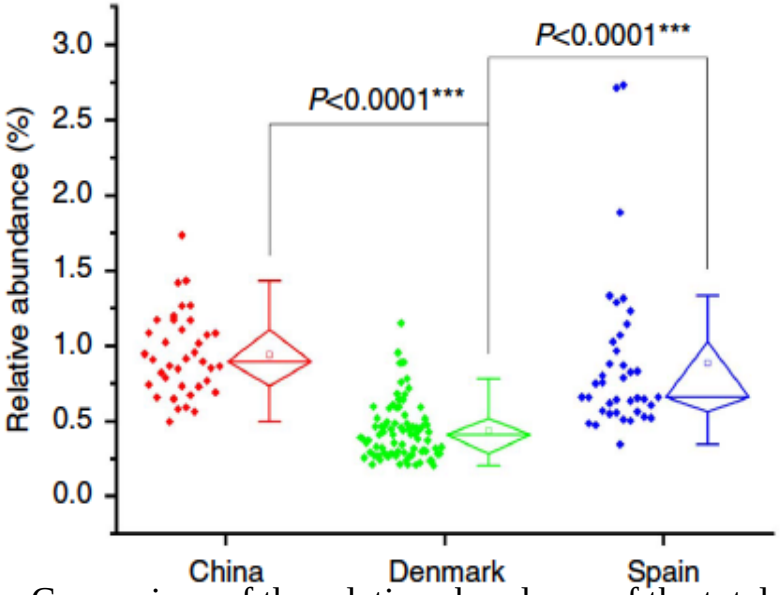
DOI: 10.1038/ncomms3151

Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota

Yongfei Hu[1,*], Xi Yang[1,*], Junjie Qin[2], Na Lu[1], Gong Cheng[1], Na Wu[1], Yuanlong Pan[1], Jing Li[1], Liying Zhu[3], Xin Wang[3], Zhiqi Meng[3], Fangqing Zhao[4], Di Liu[1], Juncai Ma[1], Nan Qin[5], Chunsheng Xiang[5], Yonghong Xiao[5], Lanjuan Li[5], Huanming Yang[2], Jian Wang[2], Ruifu Yang[6], George F. Gao[1,7], Jun Wang[2] & Baoli Zhu[1]



The human gut microbiota is a reservoir of antibiotic resistance genes, but little is known about their diversity and richness within the gut. Here we analyse the antibiotic resistance genes of gut microbiota from 162 individuals. We identify a total of 1,093 antibiotic resistance genes and find that Chinese individuals harbour the highest number and abundance of antibiotic resistance genes, followed by Danish and Spanish individuals. Single-nucleotide polymorphism-based analysis indicates that antibiotic resistance genes from the two European populations are more closely related while the Chinese ones are clustered separately. We also confirm high abundance of tetracycline resistance genes with this large cohort study. Our study provides a broad view of antibiotic resistance genes in the human gut Microbiota.



Comparison of the relative abundance of the total antibiotic resistance genes in each individual.

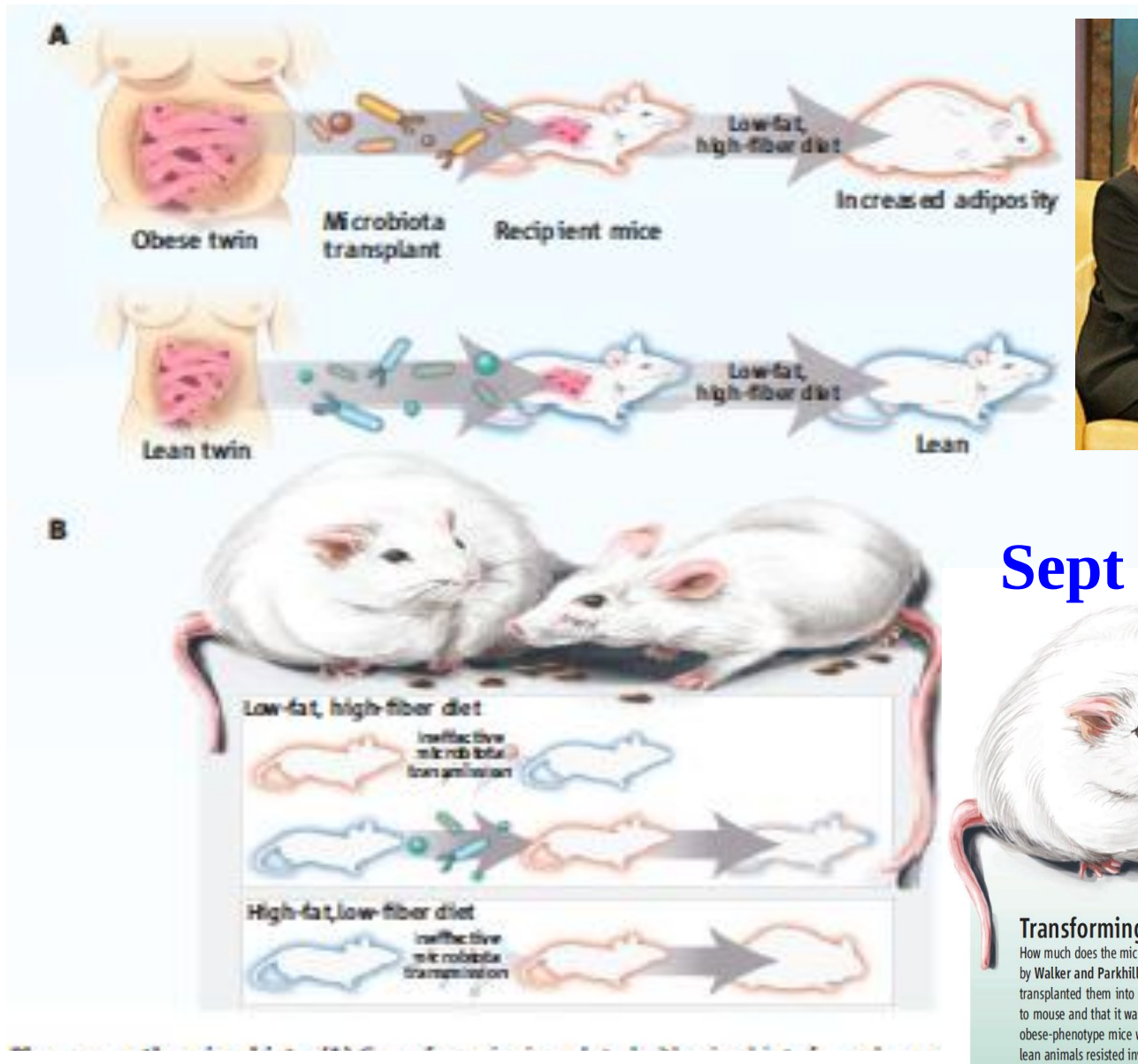
online 6 July 2014

nature biotechnology

An integrated catalog of reference genes in the human gut microbiome

Junhua Li[1–3,19], Huijue Jia[1,19], Xianghang Cai[1,19], Huanzi Zhong[1,19], Qiang Feng[1,4,19], Shinichi Sunagawa[5], Manimozhiyan Arumugam[1,5,6], Jens Roat Kultima[5], Edi Prifti[7], Trine Nielsen[6], Agnieszka Sierakowska Juncker[8], Chaysavanh Manichanh[9], Bing Chen[1], Wenwei Zhang[1], Florence Levenez[7], Juan Wang[1], Xun Xu[1], Liang Xiao[1], Suisha Liang[1], Dongya Zhang[1], Zhaoxi Zhang[1], Weineng Chen[1], Hailong Zhao[1], Jumana Yousuf Al-Aama[10,11], Sherif Edris[11,12], Huanming Yang[1,11,13], Jian Wang[1,13], Torben Hansen[6], Henrik Bjørn Nielsen[8], Søren Brunak[8], Karsten Kristiansen[4], Francisco Guarner[9], Oluf Pedersen[6], Joel Doré[7,14], S Dusko Ehrlich[7,15], MetaHIT Consortium[16], Peer Bork[5,17] & Jun Wang[1,4,6,11,18]

Many analyses of the human gut microbiome depend on a catalog of reference genes. Existing catalogs for the human gut microbiome are based on samples from single cohorts or on reference genomes or protein sequences, which limits coverage of global microbiome diversity. Here we combined 249 newly sequenced samples of the Metagenomics of the Human Intestinal Tract (MetaHit) project with 1,018 previously sequenced samples to create a cohort from three continents that is at least threefold larger than cohorts used for previous gene catalogs. From this we established the integrated gene catalog (IGC) comprising 9,879,896 genes. The catalog includes close-to-complete sets of genes for most gut microbes, which are also of considerably higher quality than in previous catalogs. Analyses of a group of samples from Chinese and Danish individuals using the catalog revealed country-specific gut microbial signatures. This expanded catalog should facilitate quantitative characterization of metagenomic, metatranscriptomic and metaproteomic data from the gut microbiome to understand its variation across populations in human health and disease.



Sept 5, 2013

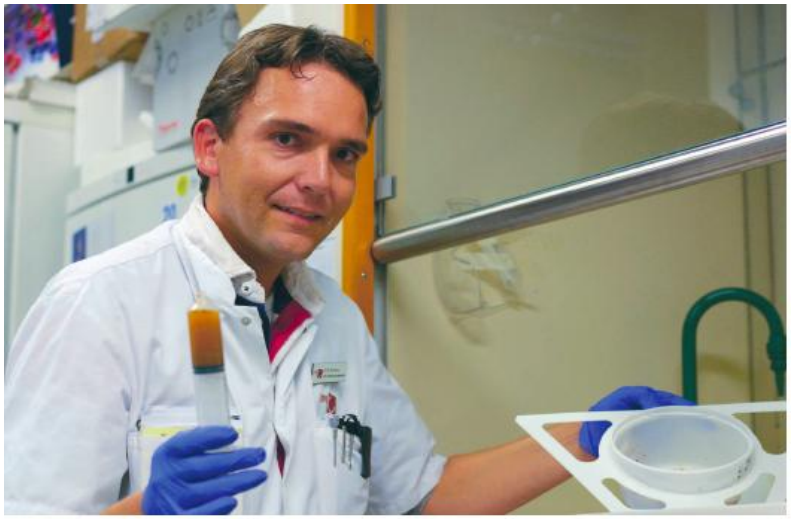
MICROBIOLOGY

Fighting Obesity with Bacteria

Intestinal bacteria from lean humans can confer protection against fat gain in experimental mice.

NATURE MEDICINE 12 DEC. 2013

Drugs that made headlines in 2013



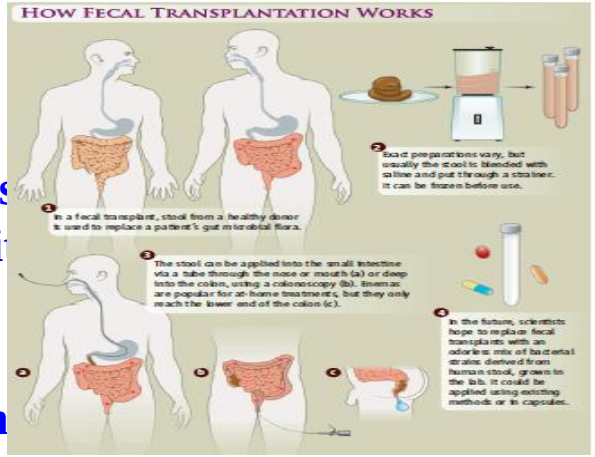
MEDICAL RESEARCH

The Promise of Poop

Fecal transplants offer hope for treating many diseases. But they need to be studied more scientifically, says one of the treatment's pioneers

fully, the *C. difficile* population, and replace it with the healthy bacterial flora from a donor, in this case her son. To do so, they would mix the son's feces with saline in a blender and squirt it straight into the patient's duodenum, the upper part of her intestine, via a thin plastic tube inserted through her nose.

Fecal transplants



iota transplants

om infection wi

overed from th

At the beginning of the summer, FDA officials declared fecal transplants to be a biologic, requiring doctors to have an experimental drug permit to perform the procedure, but

in July the agency (FDA) issued a guidance lifting that restriction.

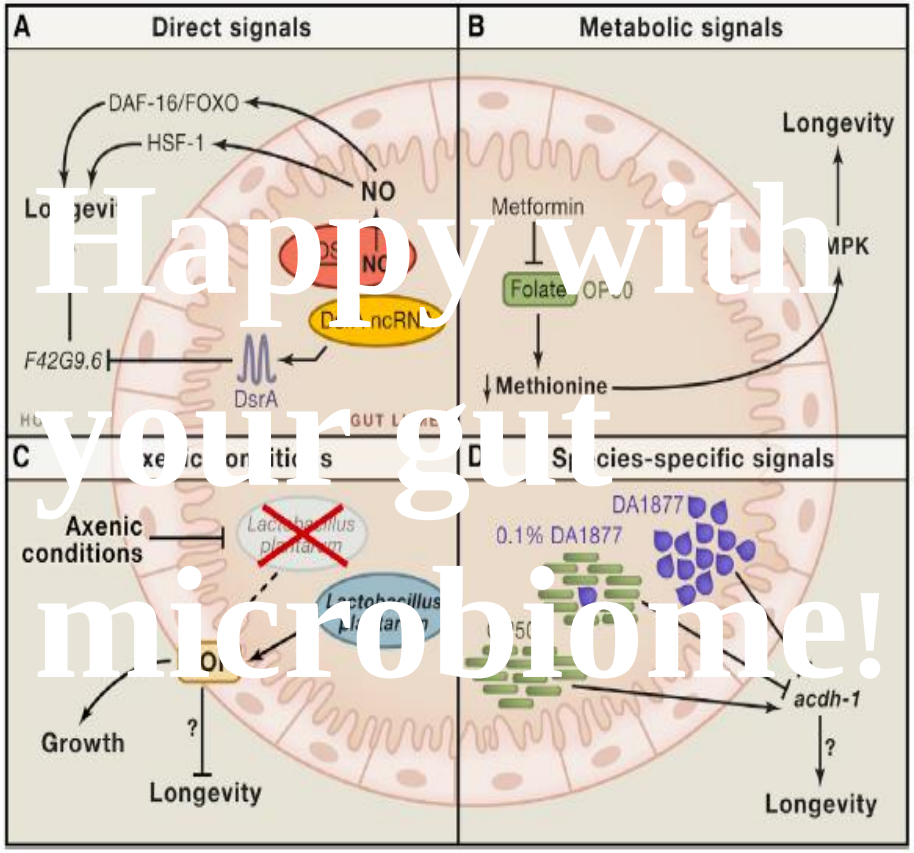
Cell 156: 408, Jan 30, 2014

Leading Edge
Minireview

You Are What You Host

"You are what you eat!"

"Despite coevolving in the presence of this "microbiome" for 500 million years, only recently have advances in sequencing technology allowed us to appreciate the complexities of this relationship and the manner by which genomes within metaorganisms interact and affect one another."



Happy with your gut microbiome!

Bring microbial sequencing to hospitals

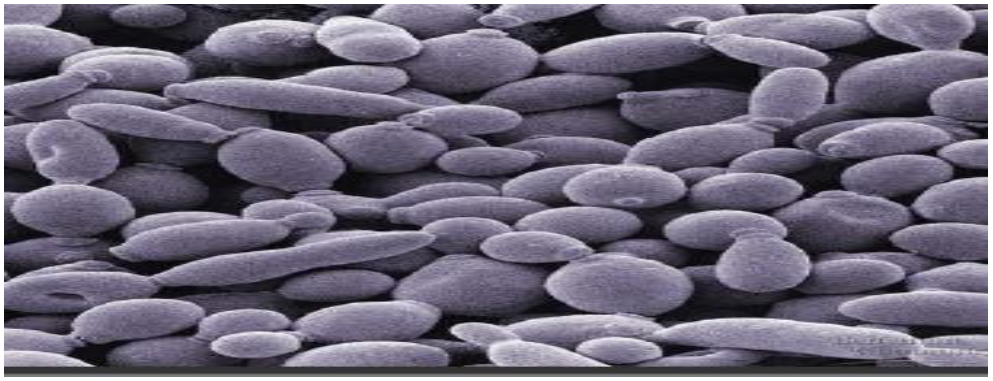
Analysing bacterial and viral DNA can help doctors to pick effective drugs quickly, says **Sharon Peacock**.

Nature, 5 0 9: 5 5 7, 2 9 M ay 2 0 1 4

Metagenomics

The most important breakthrough since microscopes and immunoassays in microbiology

Great contribution to Medicine, Ecology, & Ancient DNA



Four More Applications

1. **WE/TR Sequencing** for monogenic diseases
   WG for cancers and other complex diseases

2. **Metagenome Sequencing**
   for metabolic/infectious diseases,
   pathogen/host, microbiology, & ecology

3. **Single Cell Sequencing**
   for cancers, meta-, neurology, & development

4. **Trace DNA sequencing**
   for NIPT, PIT, & other early diagnosis,
   ancientDNA/evolution

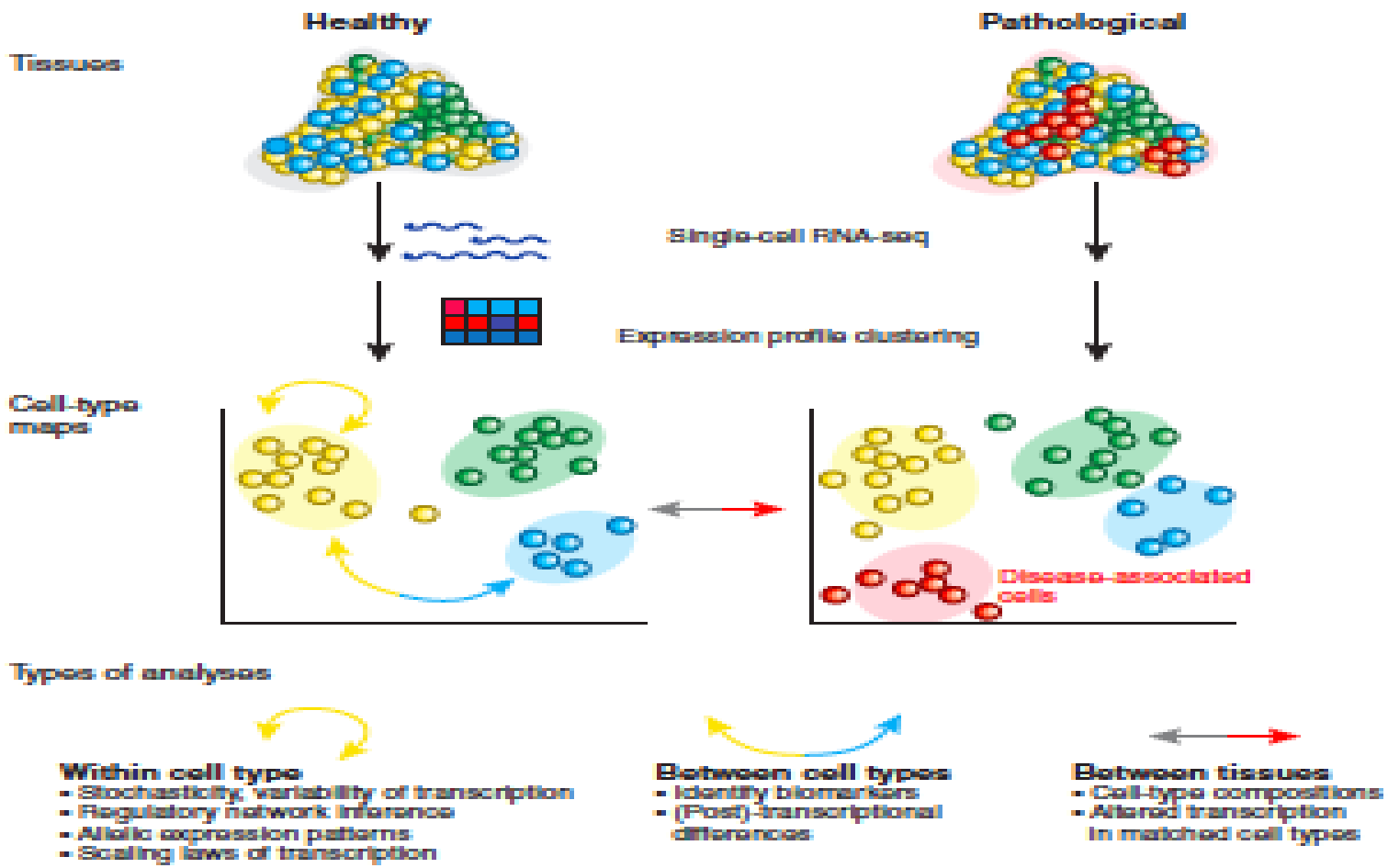
NATURE METHODS

EDITORIAL

VOL.11 NO.1 | JANUARY 2014 | 1

Method of the Year 2013

Methods to sequence the DNA and RNA of single cells are poised to transform many areas of biology and medicine.



Single-Cell Exome Sequencing Reveals Single-Nucleotide Mutation Characteristics of a Kidney Tumor

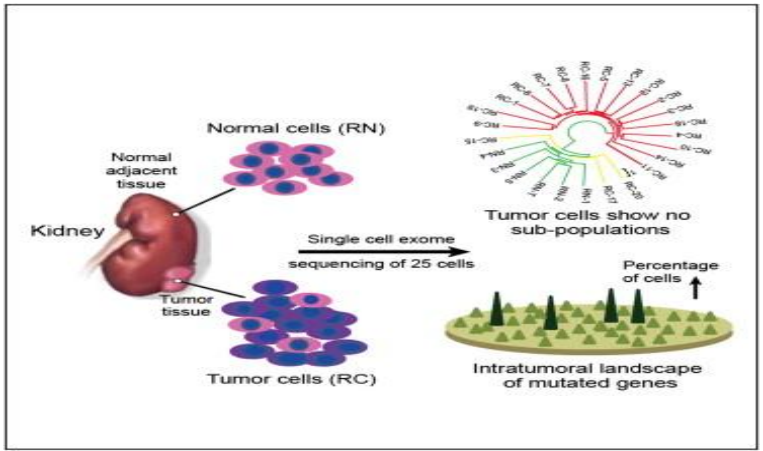
Cell

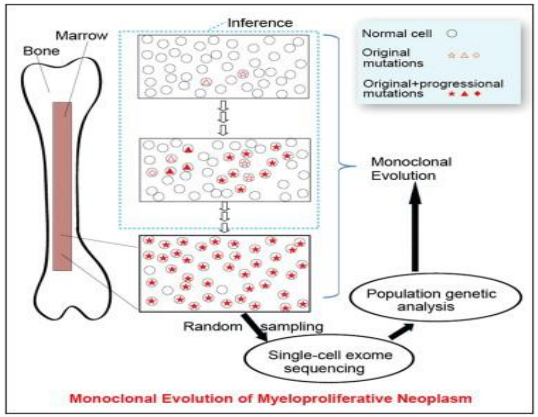
148: 886, 2012

Single-Cell Exome Sequencing and Monoclonal Evolution of a *JAK2*-Negative Myeloproliferative Neoplasm

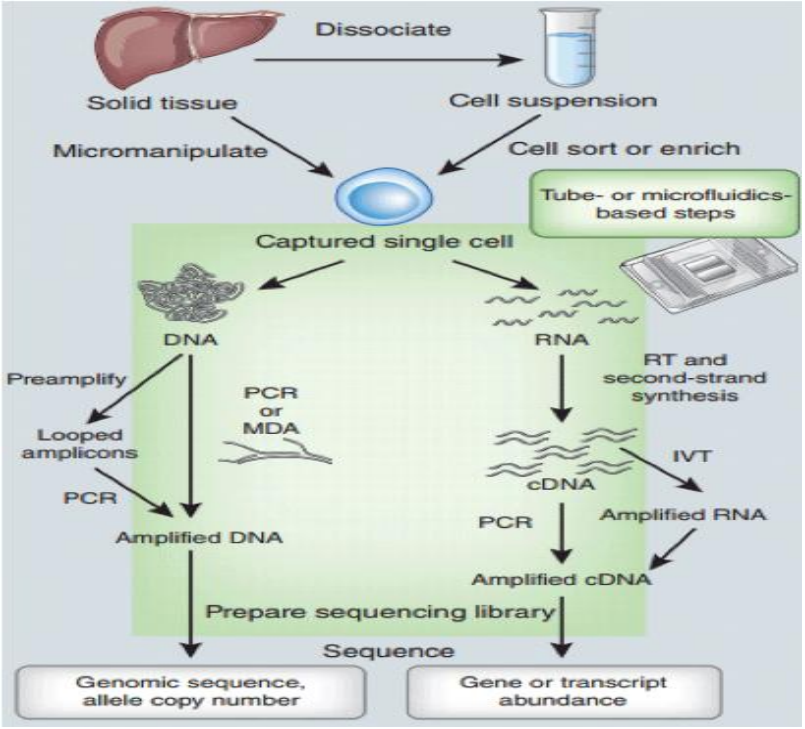
Cell

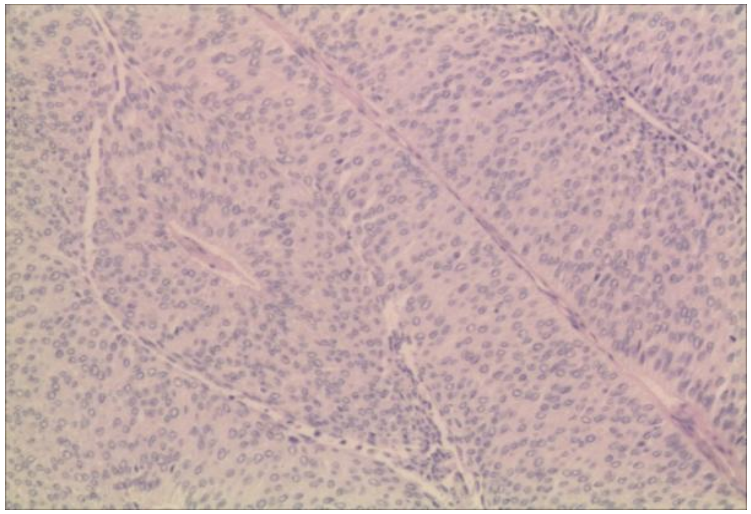
148: 873, 2012

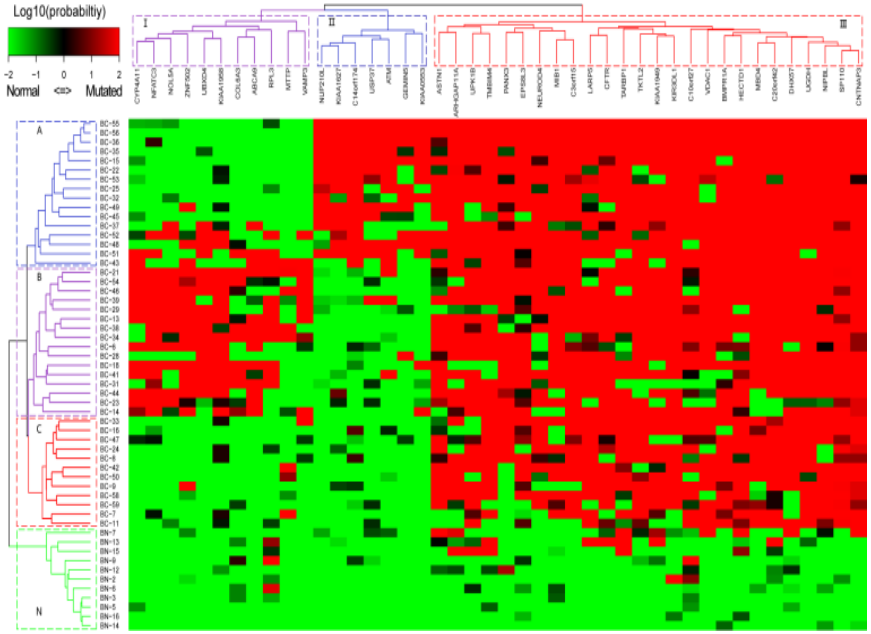


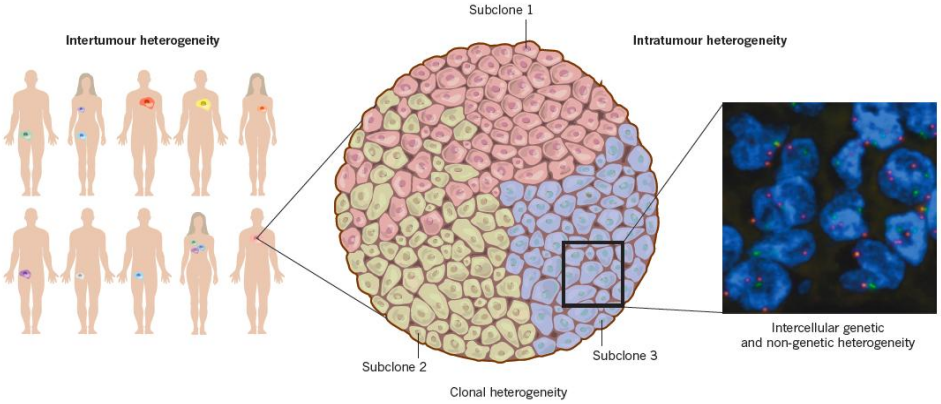


Single Cell Sequencing









frontiers in ONCOLOGY

Jan, 2014

REVIEW ARTICLE
published: 27 January 2014
doi: 10.3389/fonc.2014.00007

Current challenges in the bioinformatics of single cell genomics

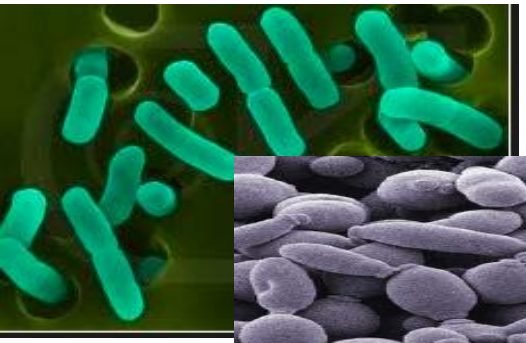
*Luwen Ning[1], Geng Liu[2], Guibo Li[2], Yong Hou[2], Yin Tong[1] and Jiankui He[1]**

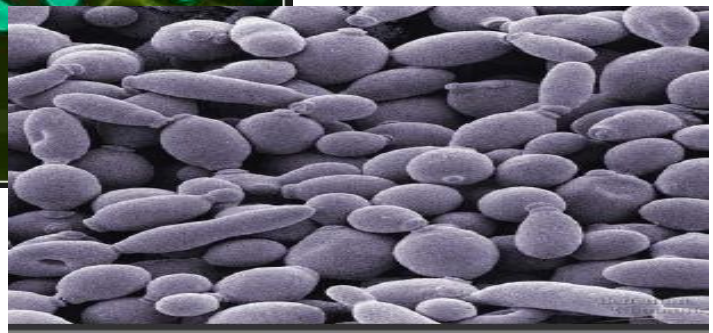
[1] Department of Biology, South University of Science and Technology of China, Shenzhen, China
[2] BGI-Shenzhen, Shenzhen, China

Single cell genomics is a rapidly growing field with many new techniques emerging in the past few years. However, few bioinformatics tools specific for single cell genomics analysis are available. Single cell DNA/RNA sequencing data usually have low genome coverage and high amplification bias, which makes bioinformatics analysis challenging. Many current bioinformatics tools developed for bulk cell sequencing do not work well with single cell sequencing data. Here, we summarize current challenges in the bioinformatics analysis of single cell genomic DNA sequencing and single cell transcriptomes. These challenges include calling copy number variations, identifying mutated genes in tumor samples, reconstructing cell lineages, recovering low abundant transcripts, and improving the accuracy of quantitative analysis of transcripts. Development in single cell genomics bioinformatics analysis will promote the application of this technology to basic biology and medical research.

Single Cell Genomics/Biology

1. **Somatic mutations**
   - Heterogeneity & lineage structure of cancers
   - Cell differentiations & development (neurons)
2. **ctDNA (Circulating Tumors' DNA)**
3. **Pre-natal/implantational diagnosis**
4. **Metagenomics**





对体外受精卵分裂产生的极体进行基因组测序能够为临床医生们进行胚胎植入前的遗传学诊断和筛查提供非常有价值的帮助。

Four More Applications

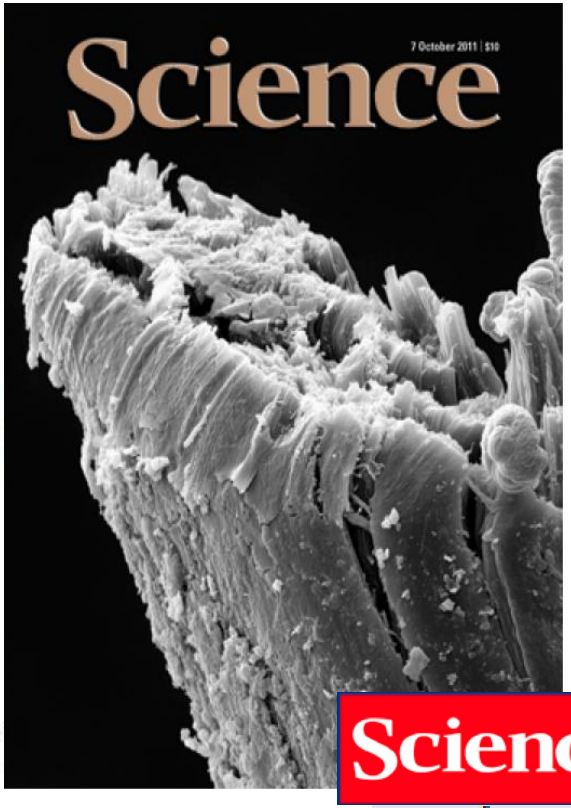
1. **WE/TR Sequencing** for monogenic diseases
   WG for cancers and other complex diseases

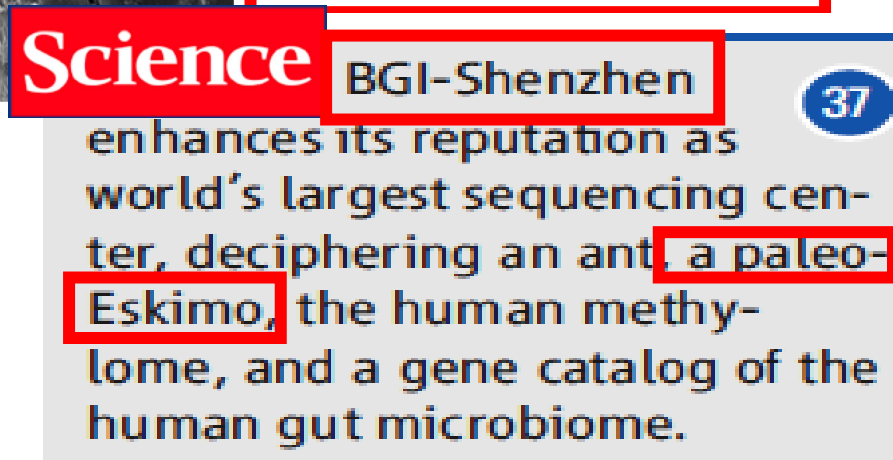
2. **Metagenome Sequencing**
   for metabolic/infectious diseases,
   pathogen/host, microbiology, & ecology

3. **Single Cell Sequencing**
   for cancers, meta-, neurology, & development

4. **Trace DNA sequencing**
   for NIPT, PIT, & other early diagnosis,
   ancientDNA/evolution

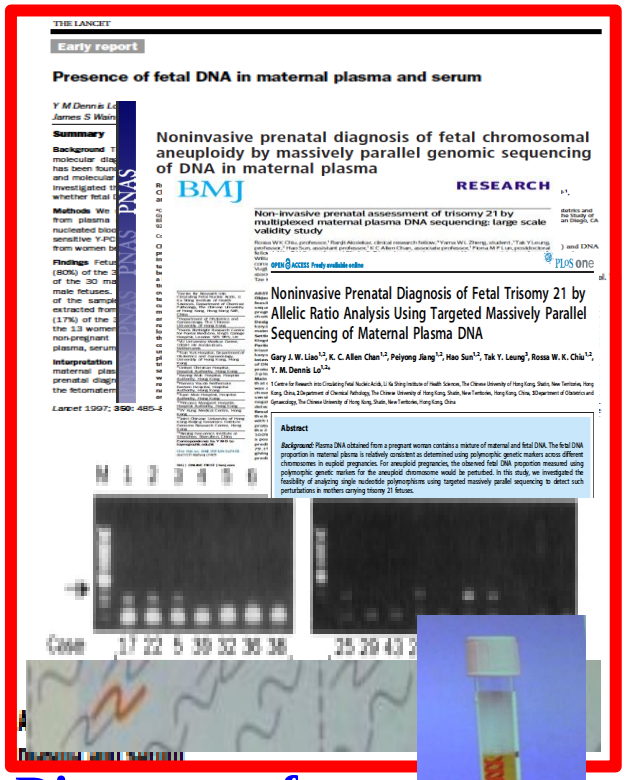
Trace DNA Sequencing





微量DNA基因组分析

(Ancient DNA Analysis)

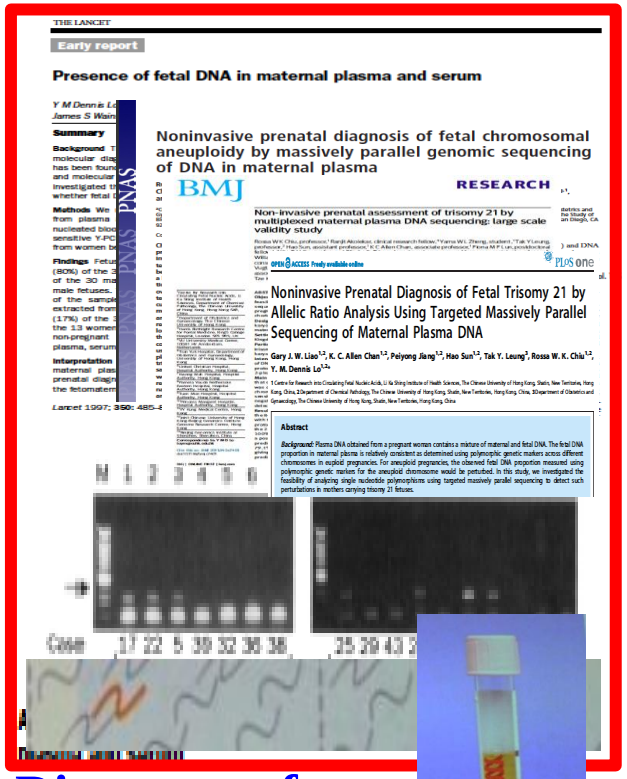


Discovery of DNA fragments in body fluids

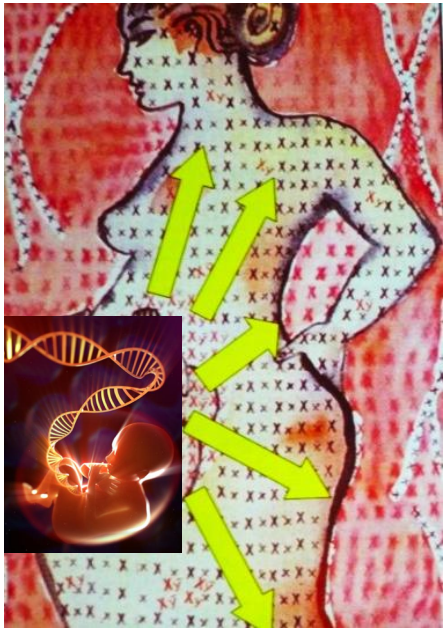
+

Large Scale High Throughout Sequencing

Early diagnosis/typing of cancers/diseases by sequencing trace DNA fragments (cfDNA) in body fluids (blood, urine, etc)



Discovery of Fetal DNA in Mother's Blood



Large Scale High Throughout Sequencing

NIPT

(Non-Invasive Prenatal Testing)

Collaborations

OPEN ACCESS Freely available online PLoS one

Noninvasive Prenatal Diagnosis of Fetal Trisomy 18 and Trisomy 13 by Maternal Plasma DNA Sequencing

Eric Z. Chen[1,2], Rossa W. K. C
Peiyong Jiang[1,2], Yama W. L
Go[5], Elizabeth T. Lau[6], Willia
Helena Lam[11], Yu Y. Kung[12],
Tang[6], Jun Wang[13,14], Cees B
1 Centre for Research into Circulating Fetal N
2 Department of Chemical Pathology, The Chi
of Hong Kong, Hong Kong SAR, China, 4 Hann
Center, Amsterdam, The Netherlands, 6 Tsua
Christian Hospital, Hong Kong SAR, China, 8 U
Mun Hospital, Hong Kong SAR, China, 11 Pri
13 Joint Chinese University of Hong Kong-Bei
Genomics Institute at Shenzhen, Shenzhen,

Abstract

Massively parallel sequencing of
noninvasive prenatal detection
noninvasive prenatal diagnosis
392 pregnancies, among which
sequencing. By using our prev
identify 36.0% and 73.0% of tri
detection of trisomy 13 and 18
increase the number of aligned
content bias in the sequencing
98.9% (261 out of 264 non-trisom
of 252 non-trisomy 18 cases)
diagnosis of trisomy 13 and tri

OPEN ACCESS Freely available online

Prenatal Detection of Aneuploidy and Imbalanc Chromosomal Arrangements by Massively Paral Sequencing

Shan Dan[1*9], Fang Cher
Wang[2], Shengpei Chen[2]
Zhang[1], Xiuqing Zhang[1]
1 Department of Perinatal Medicine, B
Shenzhen, China, 3 Fetal Medicine Unit,

Abstract

Fetal chromosomal abnorm
and several molecular gen
testing methods are highly
on the human genome at
resolution allows detectio
as the tool for prenatal d
imbalanced chromosomal
platform and with average
abnormalities by z-test.
microduplication (ranged
algorithm. Our work der
chromosomal arrangemen

BMJ

Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study

Rossa W K Chiu, professor,[1] Ranjit Akolekar, clinical research fellow,[3] Yama W L Zheng, student,[1] Tak Y Leung, professor,[2] Hao Sun, assistant professor,[1] K C Allen Chan, associate professor,[1] Fiona M F Lun, postdoctoral fellow,[1] Attie T J I Go, professor,[4] Elizabeth T Lau, department manager and honorary assistant professor,[5] William W K To, consultant,[6] Wing C Leung, consultant,[7] Rebecca Y K Tang, consultant,[8] Sidney K C Au-Yeung, consultant,[9] Helena Lam, consultant,[10] Yu Y Kung, obstetrician,[11] Xiuqing Zhang, manager,[12,13] John M G van Vugt, professor,[4] Ryoko Minekawa, postdoctoral fellow,[3] Mary H Y Tang, consultant and honorary clinical associate professor,[9] Jun Wang, professor,[12] associate director,[13] Cees B M Oudejans, associate professor,[4] Tze K Lau, professor,[2] Kypros H Nicolaides, professor,[3] Y M Dennis Lo, professor[1,12]

[1]Centre for Research into Circulating Fetal Nucleic Acids, Li Ka Shing Institute of Health Sciences, Department of Chemical Pathology, The Chinese University of Hong Kong, Hong Kong SAR, China
[2]Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong
[3]Harris Birthright Research Centre for Foetal Medicine, Kings College Hospital, London SE5 9RS, UK
[4]VU University Medical Center, 1081 HV Amsterdam, Netherlands
[5]Tsan Yuk Hospital, Department of Obstetrics and Gynaecology, University of Hong Kong, Hong Kong
[6]United Christian Hospital, Hospital Authority, Hong Kong
[7]Kwong Wah Hospital, Hospital Authority, Hong Kong
[8]Pamela Youde Nethersole Eastern Hospital, Hospital Authority, Hong Kong
[9]Tuen Mun Hospital, Hospital Authority, Hong Kong
[10]Princess Margaret Hospital, Hospital Authority, Hong Kong
[11]Yu Kung Medical Centre, Hong Kong
[12]Beijing Genomics Institute at Shenzhen, Shenzhen, China
Correspondence to: Y M D Lo lo@cuhk.edu.hk

ABSTRACT

Objectives To validate the clinical efficacy and practical feasibility of massively parallel maternal plasma DNA sequencing to screen for fetal trisomy 21 among high risk pregnancies clinically indicated for amniocentesis or chorionic villus sampling.

Design Diagnostic accuracy validated against full karyotyping, using prospectively collected or archived maternal plasma samples.

Setting Prenatal diagnostic units in Hong Kong, United Kingdom, and the Netherlands.

Participants 753 pregnant women at high risk for fetal trisomy 21 who underwent definitive diagnosis by full karyotyping, of whom 86 had a fetus with trisomy 21.

Intervention Multiplexed massively parallel sequencing of DNA molecules in maternal plasma according to two protocols with different levels of sample throughput: 2-plex and 8-plex sequencing.

Main outcome measures Proportion of DNA molecules that originated from chromosome 21. A trisomy 21 fetus was diagnosed when the z score for the proportion of chromosome 21 DNA molecules was >3. Diagnostic sensitivity, specificity, positive predictive value, and negative predictive value were calculated for trisomy 21 detection.

Results [illegible] available from 753 pregnancies with [illegible] protocol and from 314 pregnancies [illegible] The performance of the 2-plex protocol was superior to that of the 8-plex protocol. With the 2-plex protocol, trisomy 21 fetuses were detected at 100% sensitivity and 97.9% specificity, which resulted in a positive predictive value of 96.6% and negative predictive value of 100%. The 8-plex protocol detected [illegible]

Conclusion Multiplexed maternal plasma DNA sequencing analysis could be used to rule out fetal trisomy 21 among high risk pregnancies. If referrals for amniocentesis or chorionic villus sampling were based on the sequencing test results, about 98% of the invasive

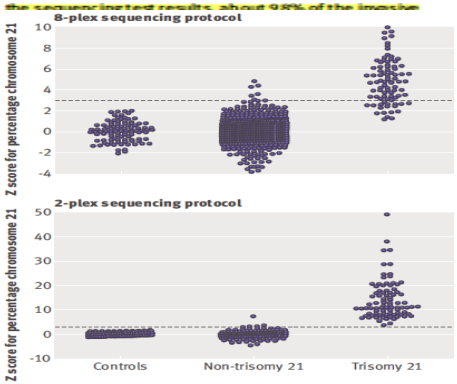


Fig 2 | Z scores of percentage chromosome 21 (proportion of sequenced plasma DNA molecules originating from chromosome 21) determined by the 8-plex and 2-plex sequencing protocols. Broken lines indicate the z score cut-off value of 3

Non-Invasive Prenatal Testing by Whole Genome/Exome Sequencing

Collaborations

Journal of Maternal-Fetal and Neonatal Medicine, 2012; Early Online: 1–5
Copyright © 2012 Informa UK, Ltd.
ISSN 1476-7058 print/ISSN 1476-4954 online
DOI: 10.3109/14767058.2011.635730

informa healthcare

Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal plasma DNA sequencing

Tze Kin Lau[1*], Fang Chen[2*], Xiaoyu Pan[2,3], Ritsuko K. Pooh[4], Fuman Jiang[2], Yihan Li[2], Hui Jiang[2], Xuchao Li[2], Shengpei Chen[2] & Xiuqing Zhang[2]

[1]Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong, [2]Guangdong Province Key Laboratory of Genome, BGI-Shenzhen, Shenzhen, China, [3]School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China, and [4]CRIFM Clinical Research

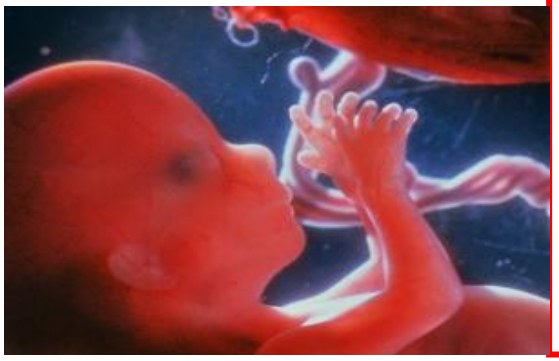
Objective: To develop a new bioinformatic metho
vasive prenatal identification of common fetal an
using massively parallel sequencing on maternal
Methods: Massively parallel sequencing was perf
plasma DNA samples from 108 pregnant women
gestation: 12+5 week) immediately before chorio
sampling (CVS) or amniocentesis. Data were anal
a novel z-score method with internal reference ch
The diagnostic accuracies of the fetal karyotypin
compared against two previously reported z-score
without adjustment and the other with GC corre
A total of 32 cases with fetal aneuploidy were co
conventional karyotyping, including 11 cases of
10 cases of Trisomy 18, 2 cases of Trisomy 13, 8 ca
syndrome (45, XO) and one case of Klinefelter syn
XXY). Using the z-score method without referenc
the detection rate for Trisomy 21, Trisomy 18, Tris
syndrome, and Klinefelter's syndrome is 100%, 4
and 0% respectively. Using the z-score method w
tion, the detection rate increased to 100% for Tris
for Trisomy 18, 100% for Trisomy 13. By using the
with internal reference, the detection rate increa
for all aneuploidies. The false positive rate was 0
methods. *Conclusion*: This massively parallel sequ
approach, combined with the improved z-score t
ology, enables the prenatal diagnosis of most co
dies with a high degree of accuracy, even in the
pregnancy.

OPEN ACCESS Freely available online

Noninvasive Prenatal Diagnosis of Fe
Trisomy 13 by Maternal Plasma DNA Sequencing

Eric Z. Chen[1,2], Rossa W. K. Chiu[1,2], Hao Sun[1,2], Ranjit Akolekar[4], K. C. Allen Chan[1,2], Tak Y. Leung[3], Peiyong Jiang[1,2], Yama W. L. Zheng[1,2], Fiona M. F. Lun[1,2], Lisa Y. S. Chan[1,2], Yongjie Jin[1,2], Attie T. J. I. Go[5], Elizabeth T. Lau[6], William W. K. To[7], Wing C. Leung[8], Rebecca Y. K. Tang[9], Sidney K. C. Au-Yeung[10], Helena Lam[11], Yu Y. Kung[12], Xiuqing Zhang[13,14], John M. G. van Vugt[5], Ryoko Minekawa[4], Mary H. Y. Tang[6], Jun Wang[13,14], Cees B. M. Oudejans[5], Tze K. Lau[3], Kypros H. Nicolaides[4], Y. M. Dennis Lo[1,2,13,*]

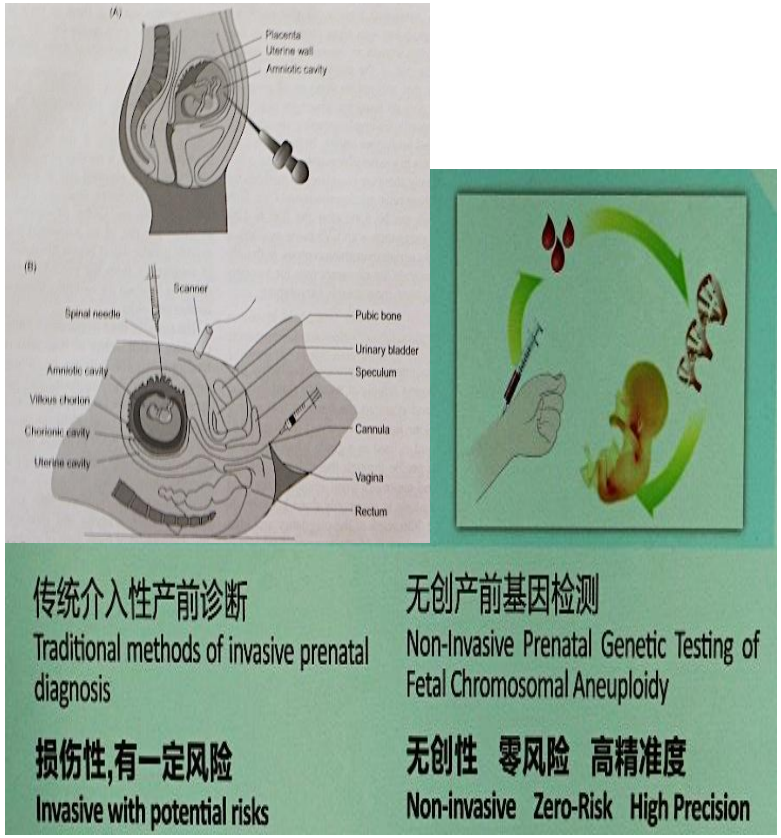
1 Centre for Research into Circulating Fetal Nucleic Acids, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China, 2 Department of Chemical Pathology, The Chinese University of Hong Kong, Hong Kong SAR, China, 3 Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong SAR, China, 4 Harris Birthright Research Centre for Fetal Medicine, King's College Hospital, London, United Kingdom, 5 VU University Medical Center, Amsterdam, The Netherlands, 6 Tsan Yuk Hospital, Department of Obstetrics and Gynaecology, University of Hong Kong, Hong Kong SAR, China, 7 United Christian Hospital, Hong Kong SAR, China, 8 Kwong Wah Hospital, Hong Kong SAR, China, 9 Pamela Youde Nethersole Eastern Hospital, Hong Kong SAR, China, 10 Tuen Mun Hospital, Hong Kong SAR, China, 11 Princess Margaret Hospital, Hospital Authority, Hong Kong SAR, China, 12 YY Kung Medical Centre, Hong Kong SAR, China, 13 Joint Chinese University of Hong Kong-Beijing Genomics Institute Genome Research Centre, Li Ka Shing Institute of Health Sciences, Hong Kong SAR, China, 14 Beijing Genomics Institute at Shenzhen, Shenzhen, China

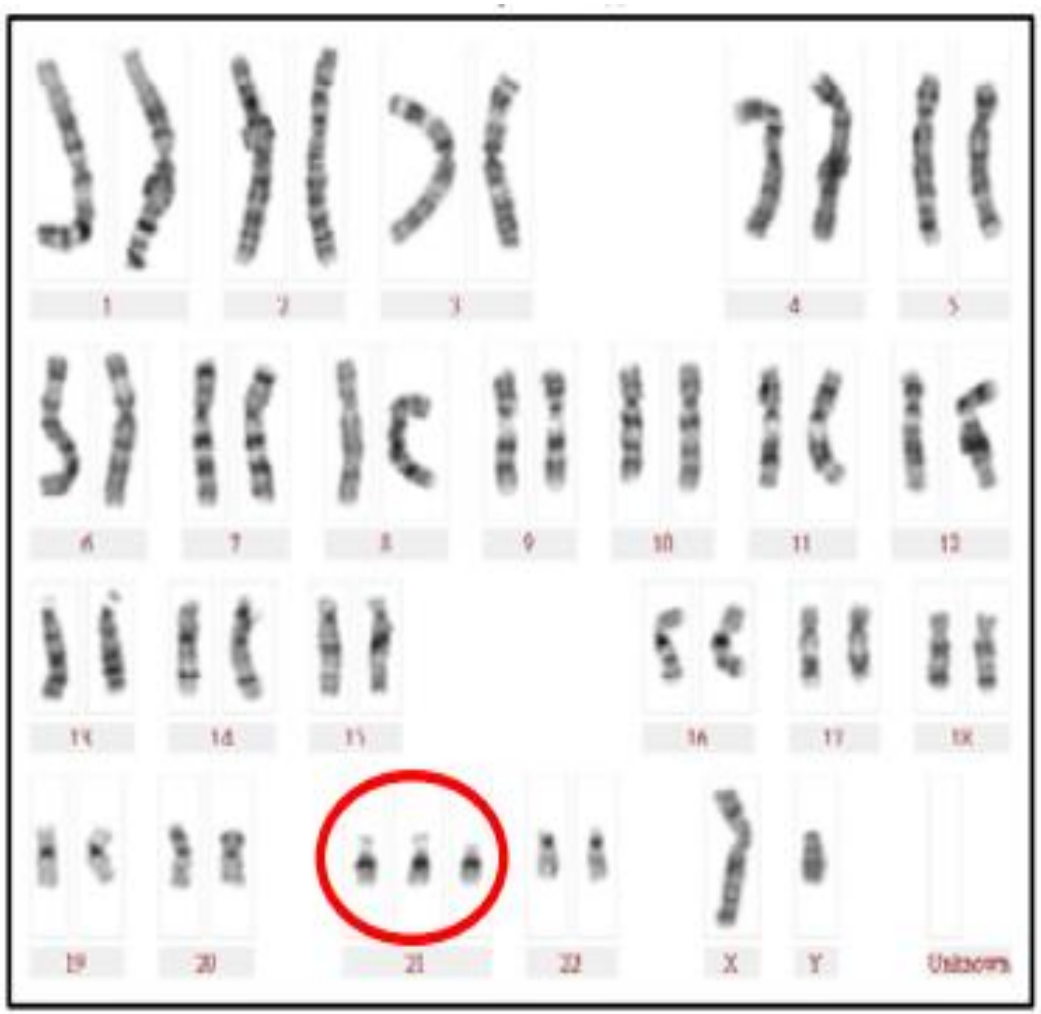


~ 400,000 testings conducted,
> 2500 cases of Tri 21 detected
as of Aug. 30, 2014

Non-Invasive Prenatal Testings (NIPT)

By sequencing

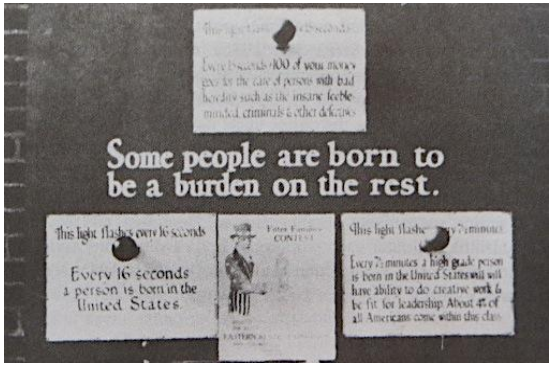




Down Syndrome

To test or not to test This is for your choice.

Raise the Banner of Science & Humanity!



Genomics has never provided any evidences for Eugenics!

We are all equal!

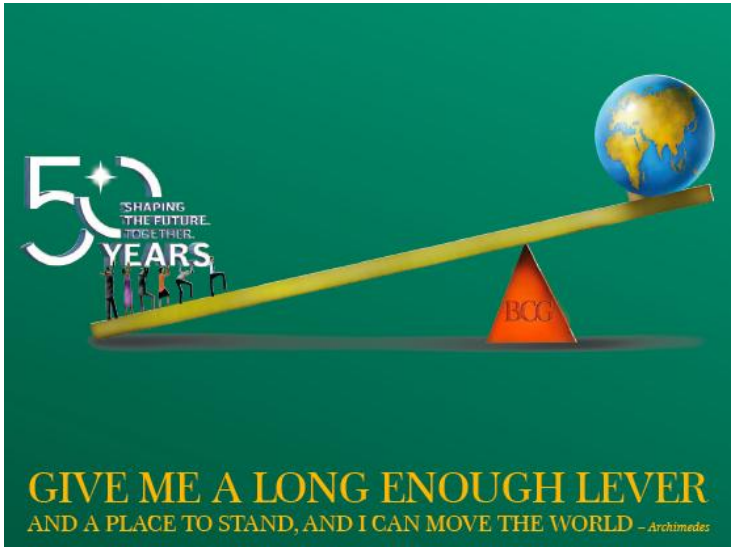
Outline:

1. A later comer of a revolution
2. Two pillars of genomics
3. Three impacts of the HGP
4. Four “to improve”
5. **Five techs changing the world**


In partnership with
YouGov
Let's
Talk
about
The
Future
Full of Opportunitie

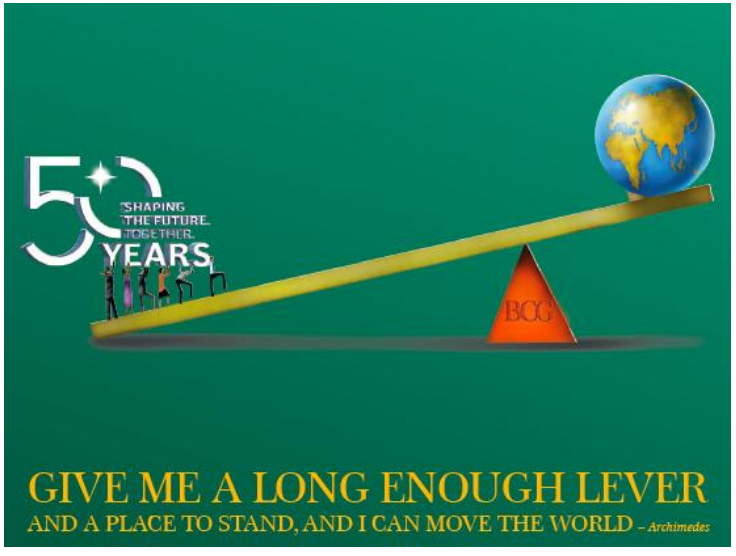
Five Technologies Changing the World

1. Molecular Technologies
2. SCs & iPS
3. Synthetic Genomics
4. Animal Cloning
5. BioBanking



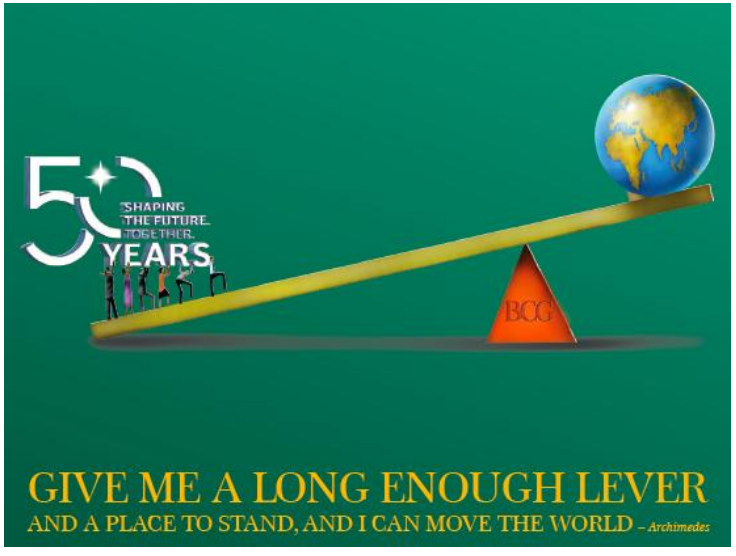
Five Technologies Changing the World

1. Molecular Technologies
2. SCs & iPS
3. **Synthetic Genomics**
4. Animal Cloning
5. BioBanking



Five Technologies Changing the World

1. Molecular Technologies
2. SCs & iPS
3. **Synthetic Genomics**
4. Animal Cloning
5. BioBanking

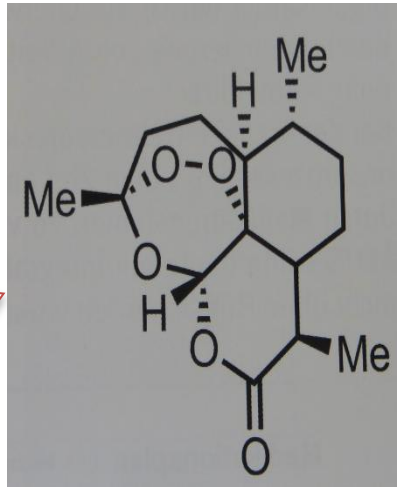


Artemisinin: An anti-malaria drug

Wild weed

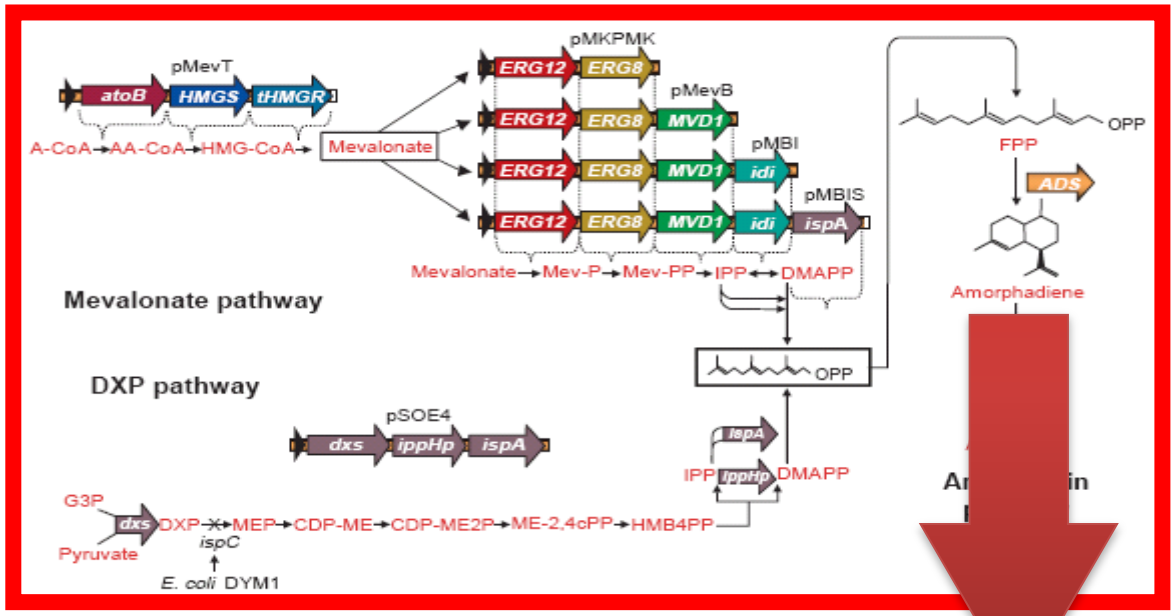
Farmed weeds

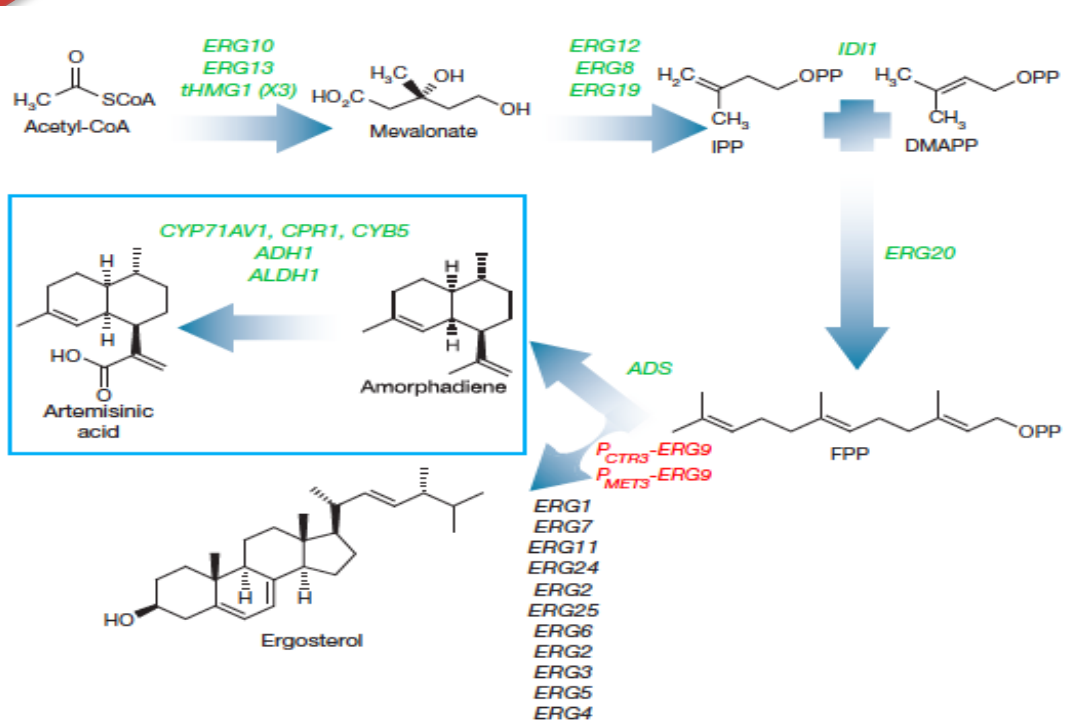
Artemisia apiacea

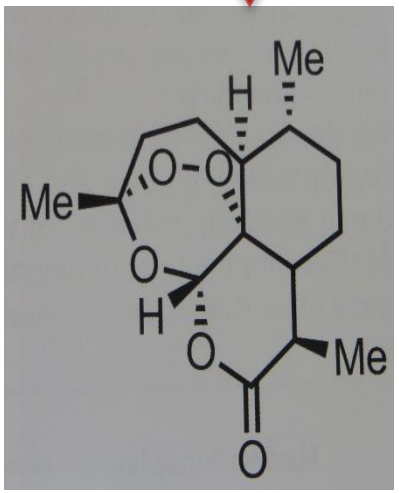


Artemisinin

Genes involved in the metabolic pathway to synthesize artenisinin

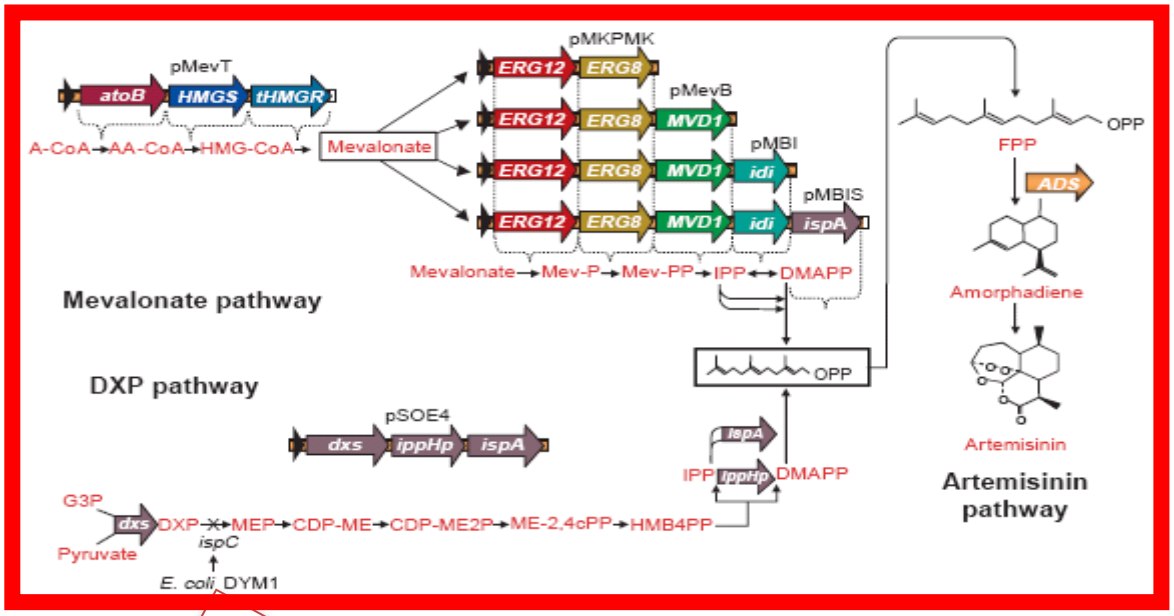


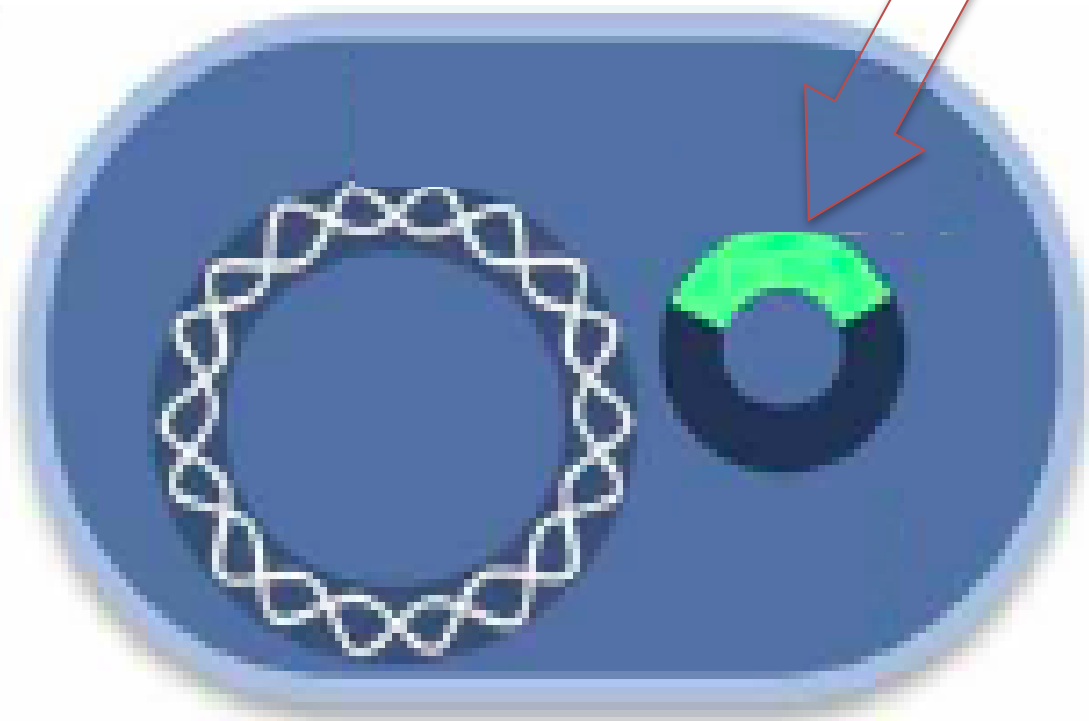


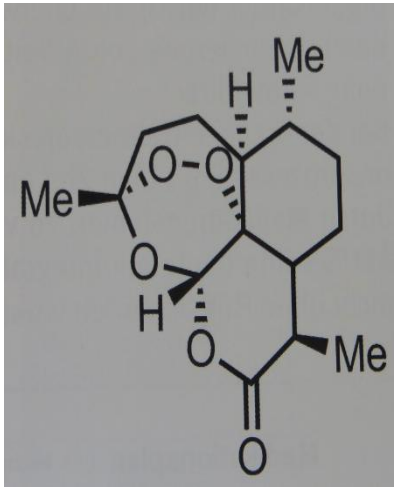


Artemisinin

Genes involved in the metabolic pathway to synthesize artenisinin

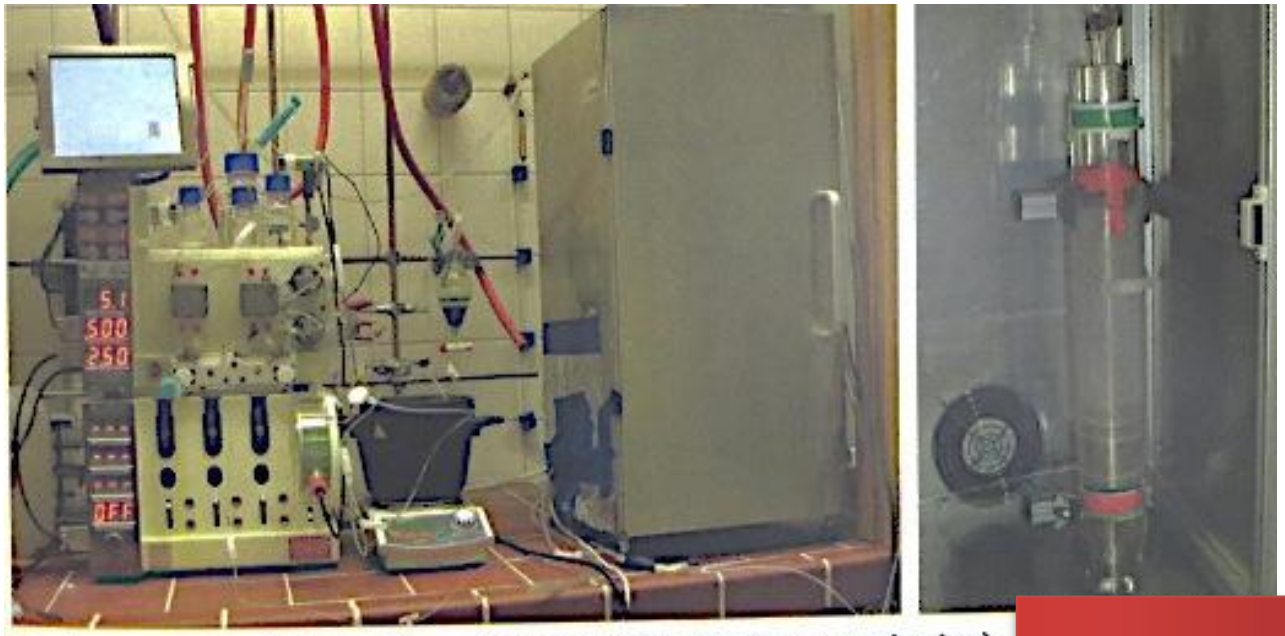




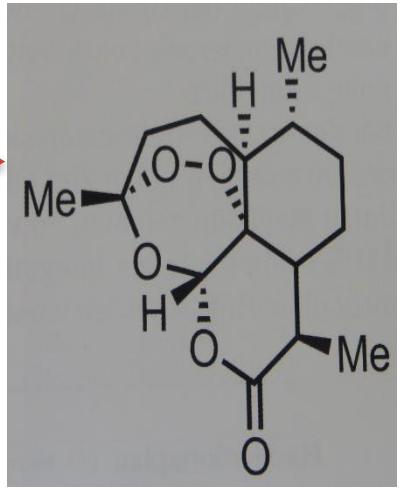


Artemisinin

Industrial Production of Artemisinin

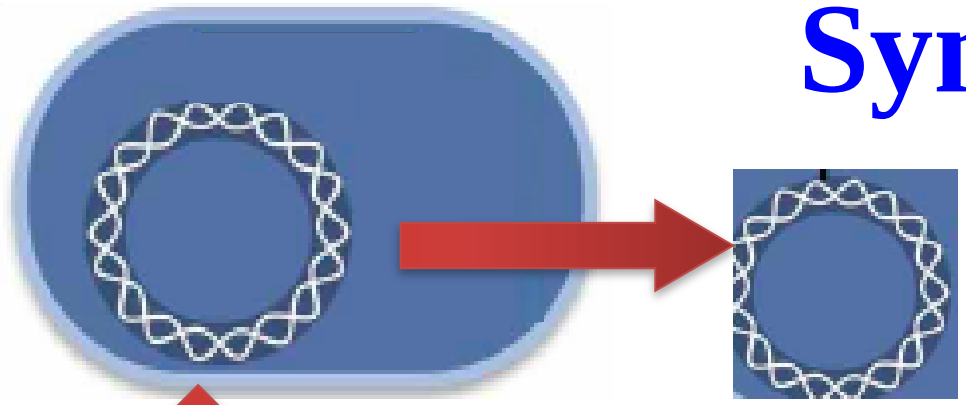


Reaktor zur Artemisininsynthese (links) mit dem Photoreaktor (rechts)

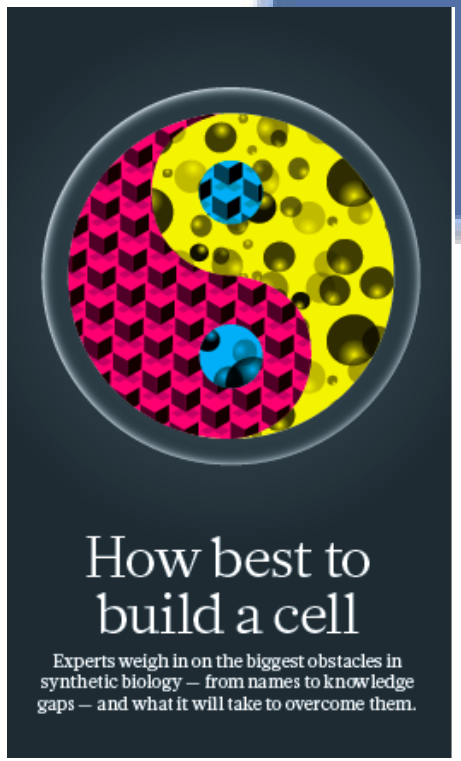


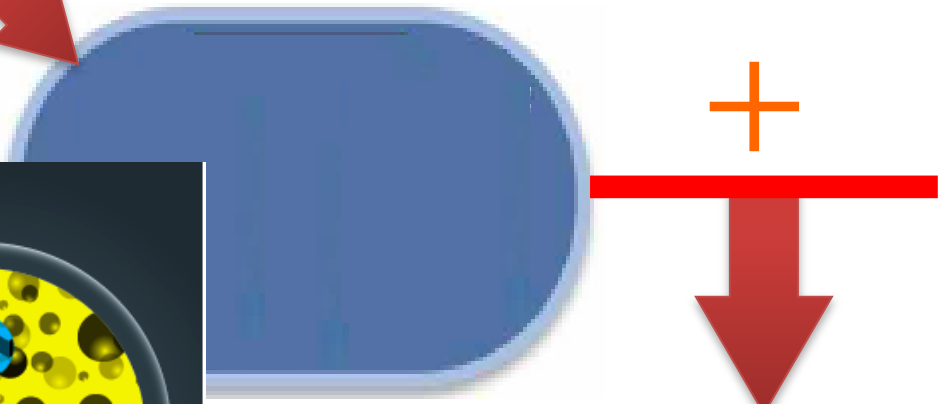
Artemisinin

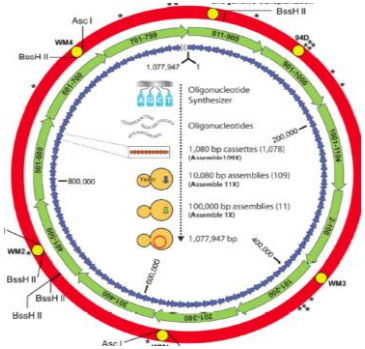
Synthetic Genomics

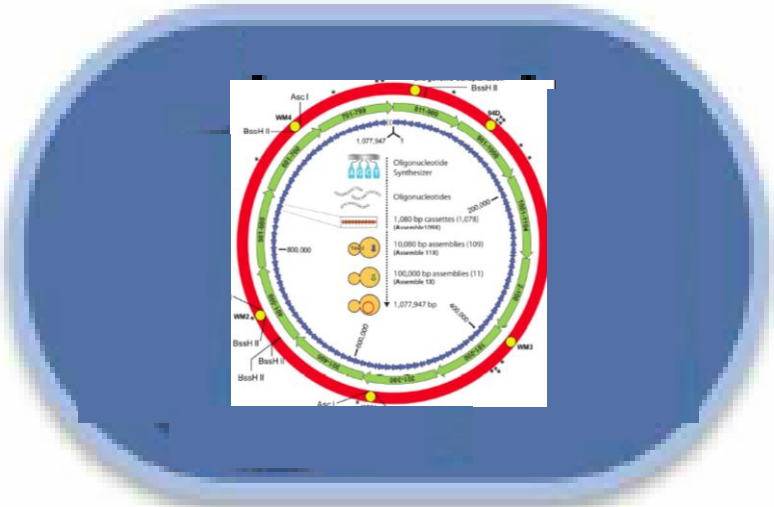


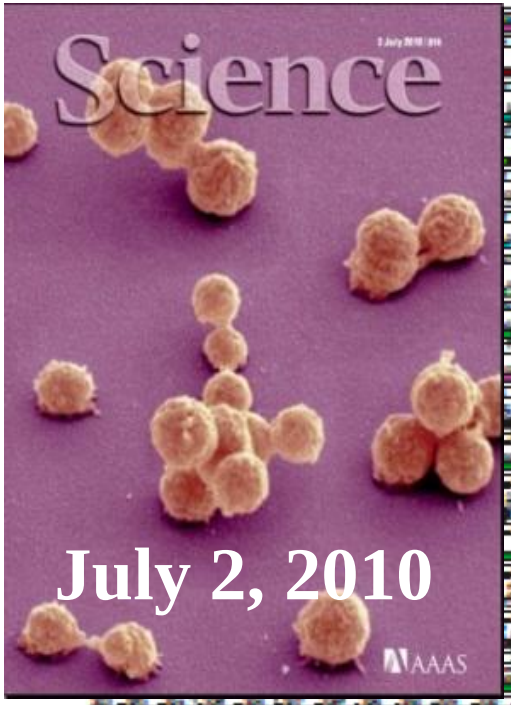
"DNA is the software of the cell, and when we change the software we change the species." J. C. Venter

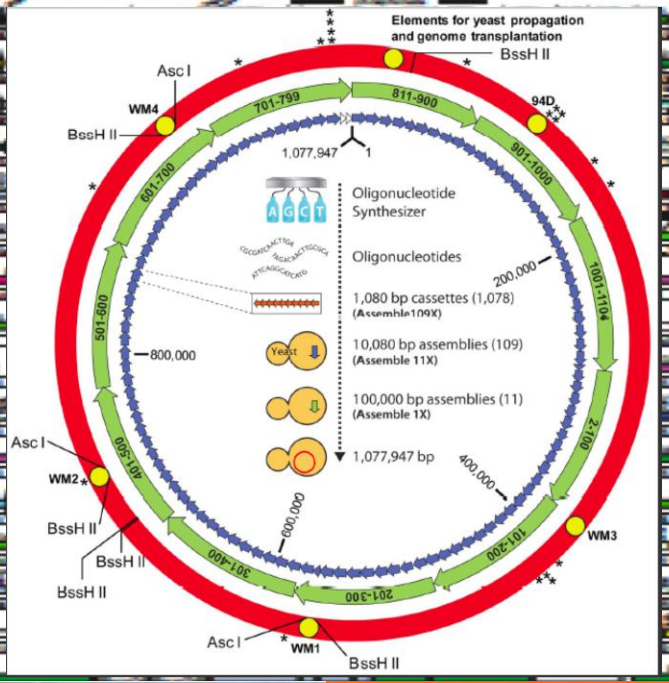


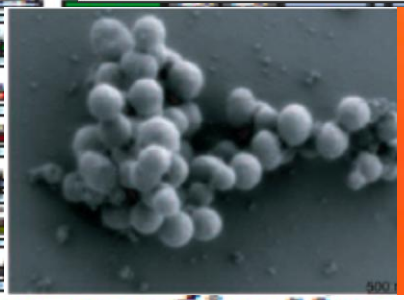












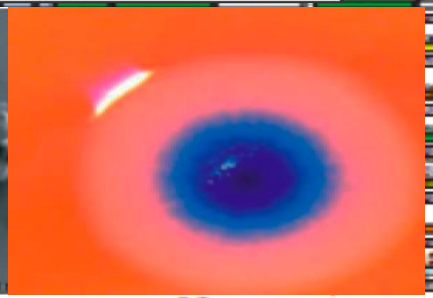
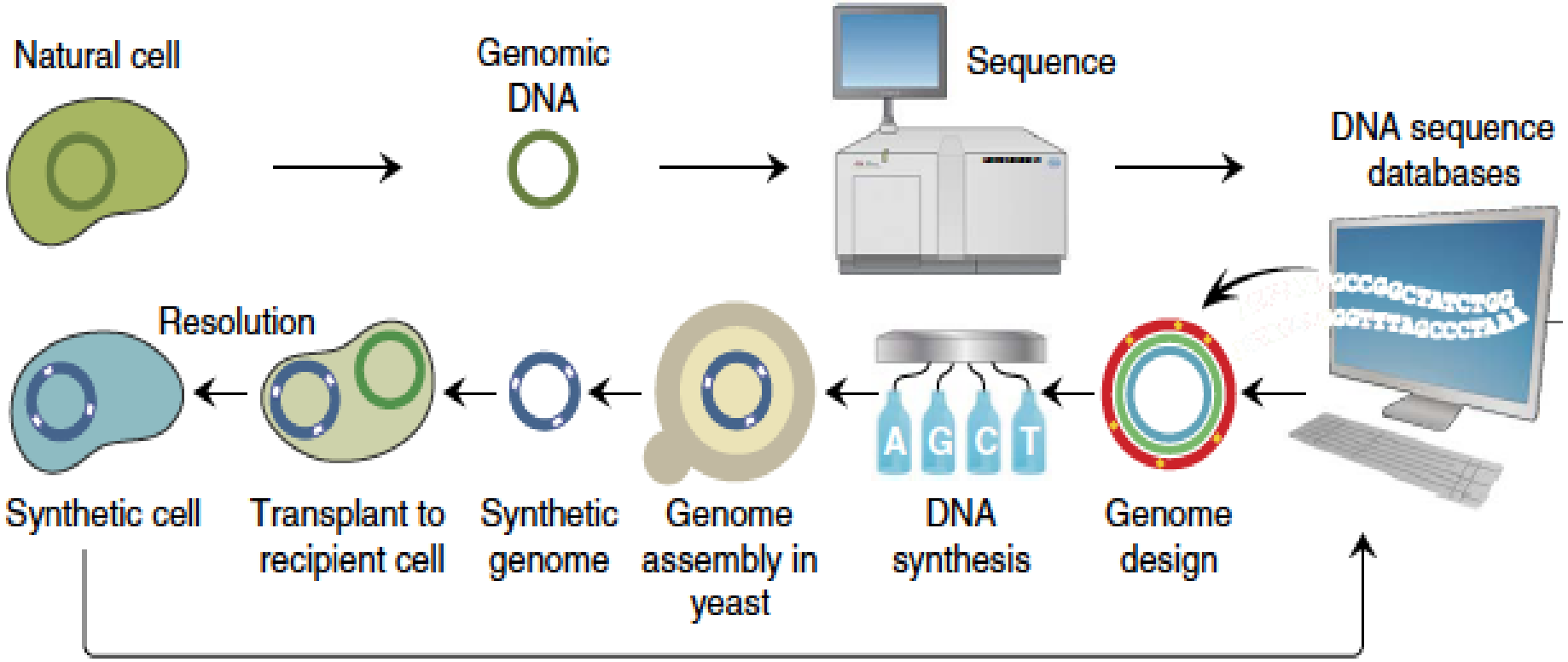


Figure S2

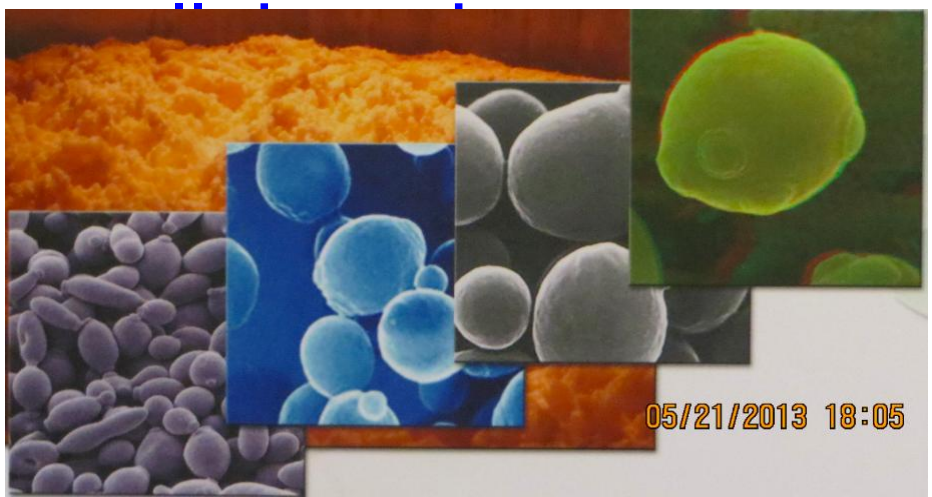
The first genome designed by computer.



Genomics: From Reading to Writing

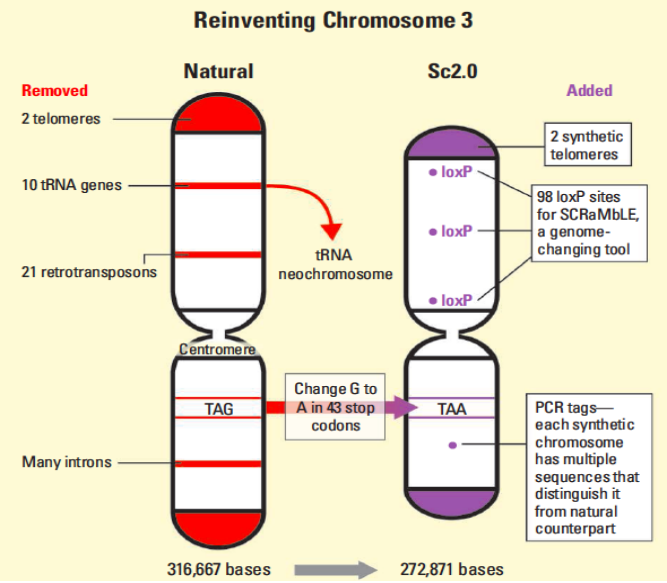
Another Milestone in Life Sciences

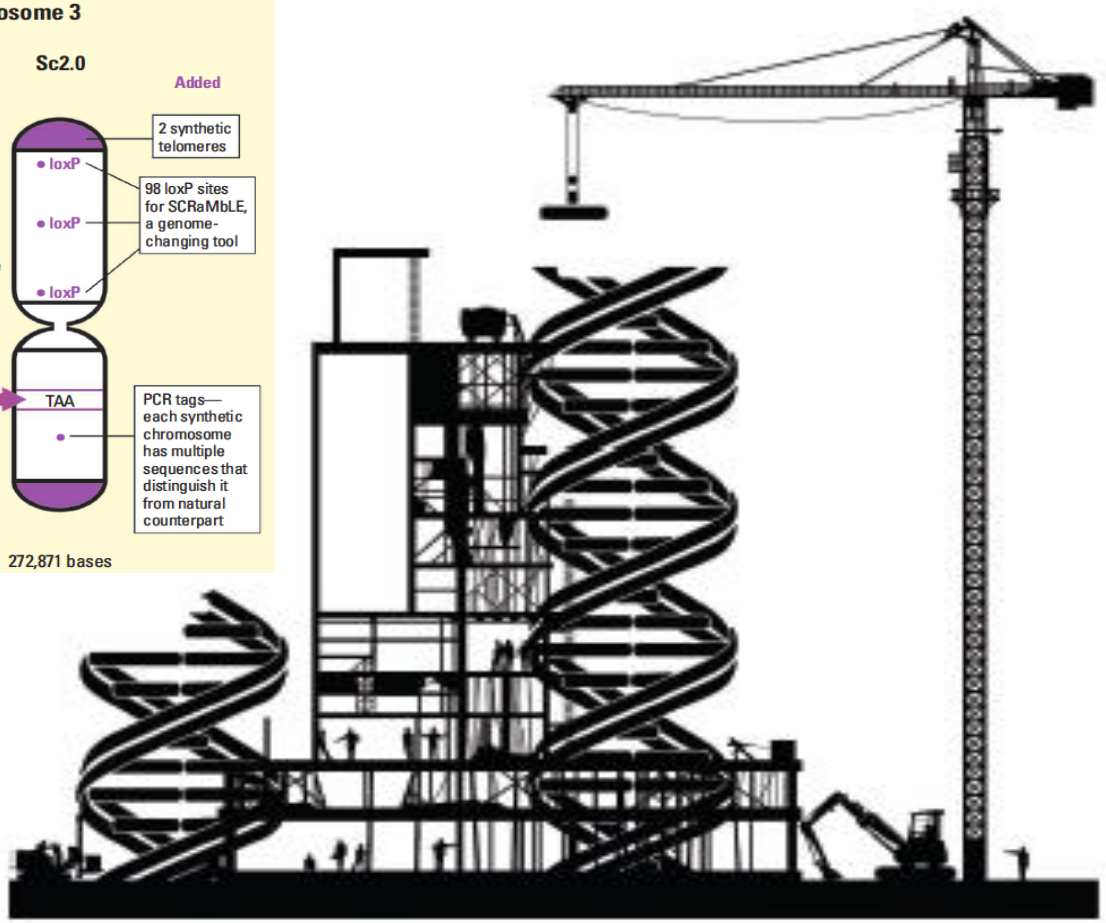
- ***M. laboratorium*** （1.08 M) —— C. Venter 2010
- *E. coli*（4.6M） —— G. Church (Harvard) in 3-5 years
- *S. cerevisiae*（15M） —— International ~~in 8-10 years~~ in 3 – 5 years



March 28, 2014

Science

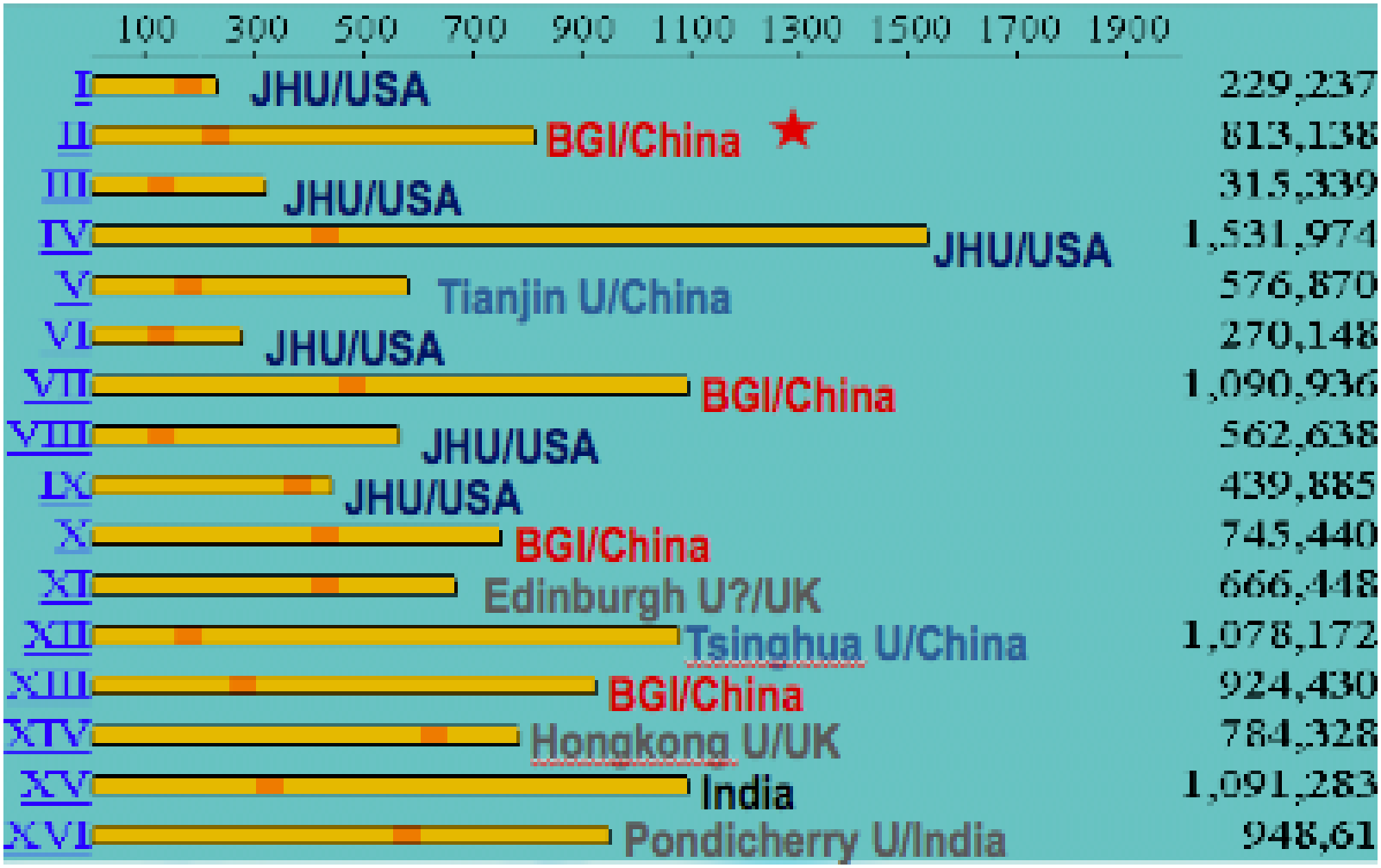




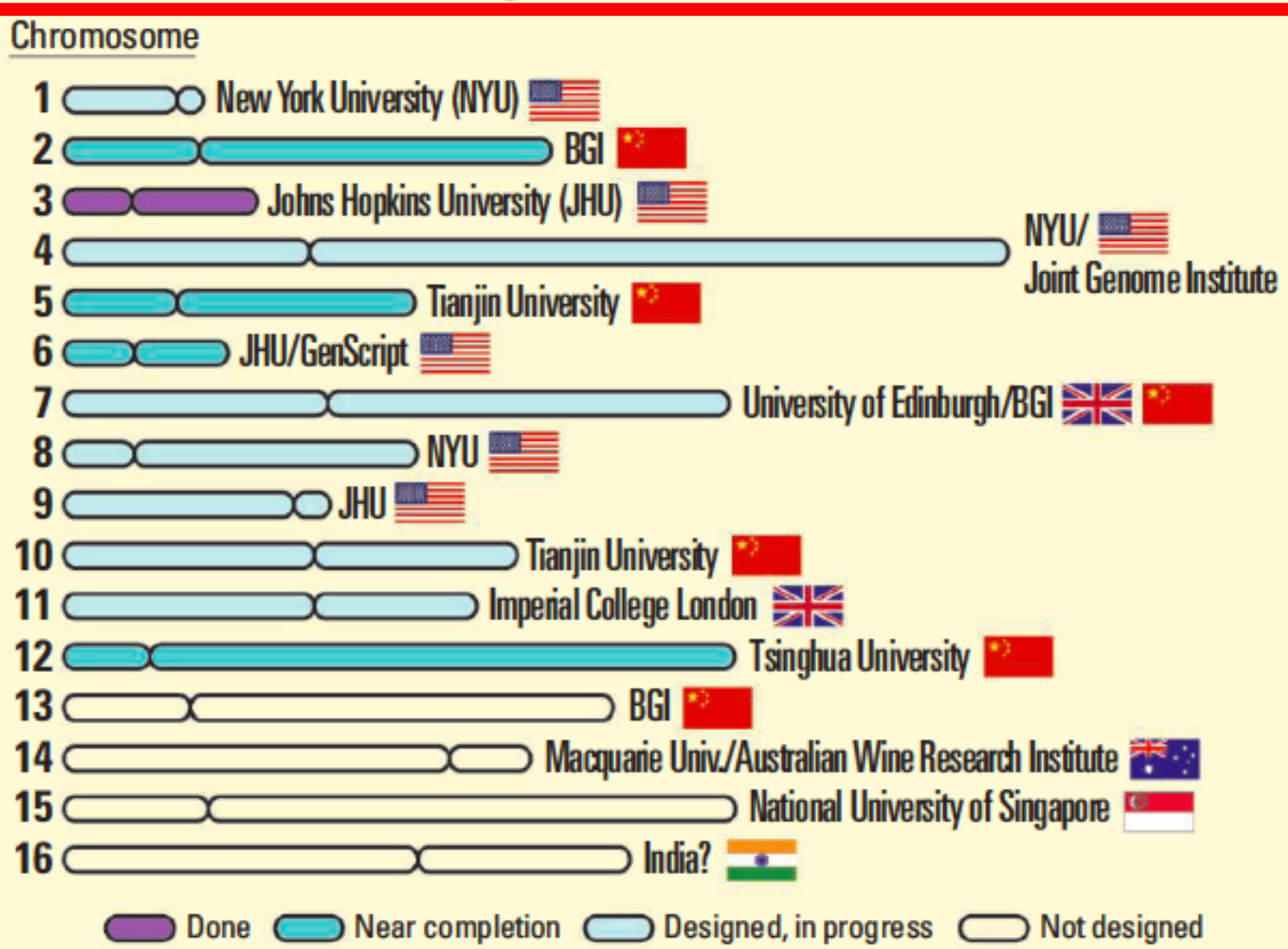
Building the Ultimate Yeast Genome

Chromosome by chromosome, a global army of researchers and students is putting together the first synthetic eukaryote genome

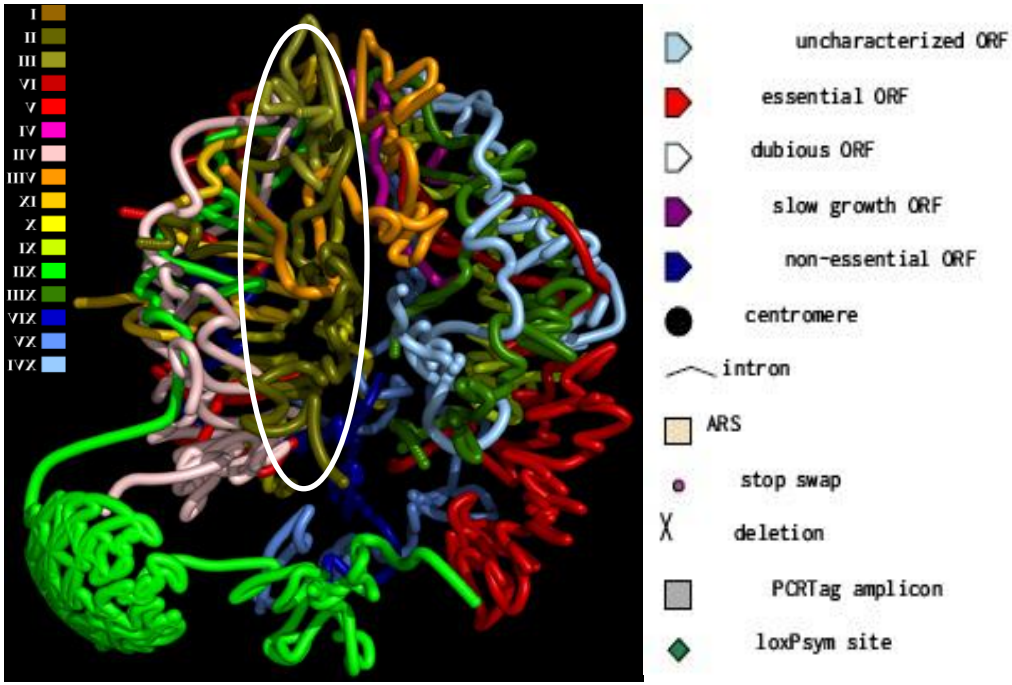
Design and Synthesis of the 1st Eukaryotic Genome



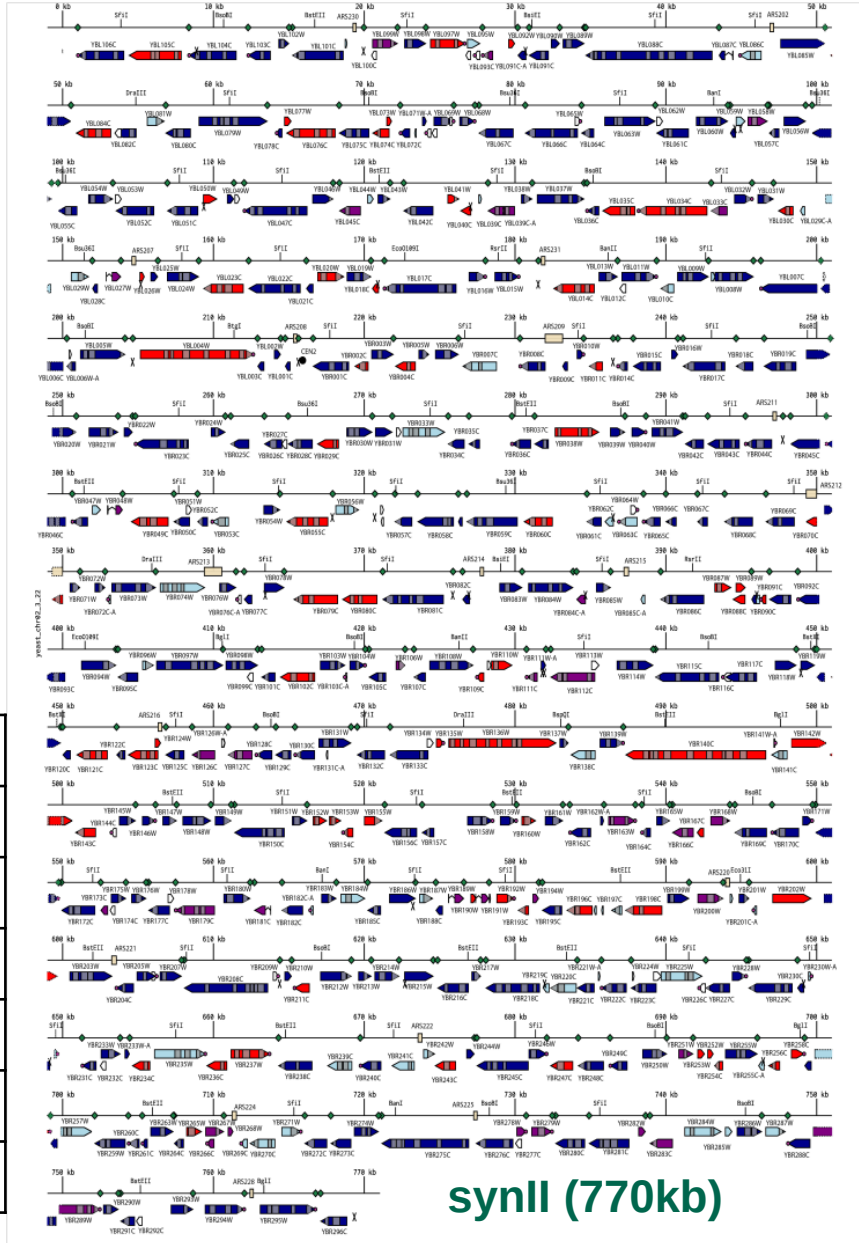
Status of Synthetic Yeast Genome



Re-designing of Yeast ChrII

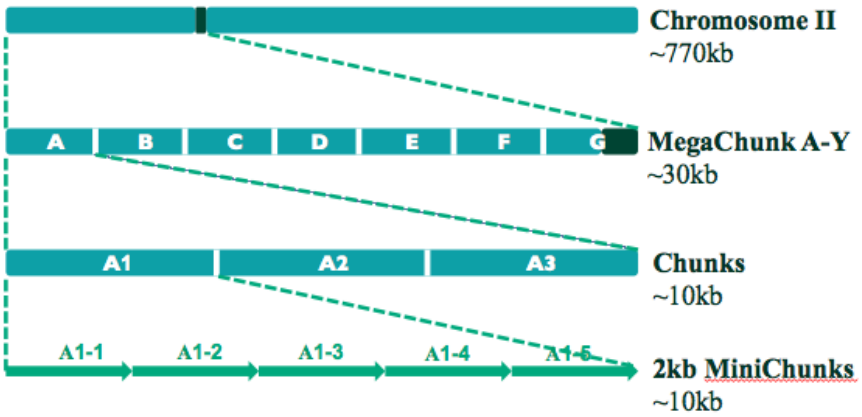


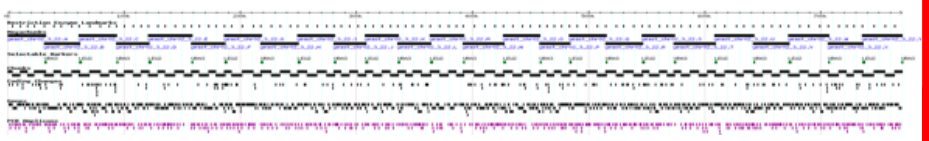
Protein coding gene number	437/456
PCR tags	532 pairs
TAG>TAA swap	90/98
LoxPsym sites added	367
tRNA deletion	13
LTR/transposons deleted	22/9
Introns deleted	22/30



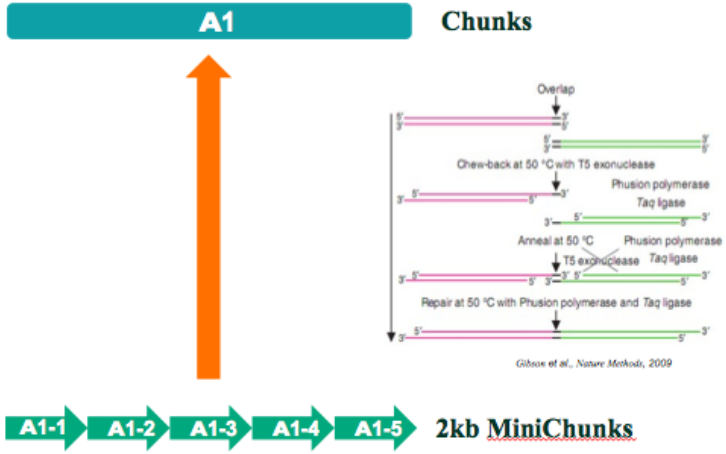
synII (770kb)

Strategy of Breaking Down





Strategy of Assembly



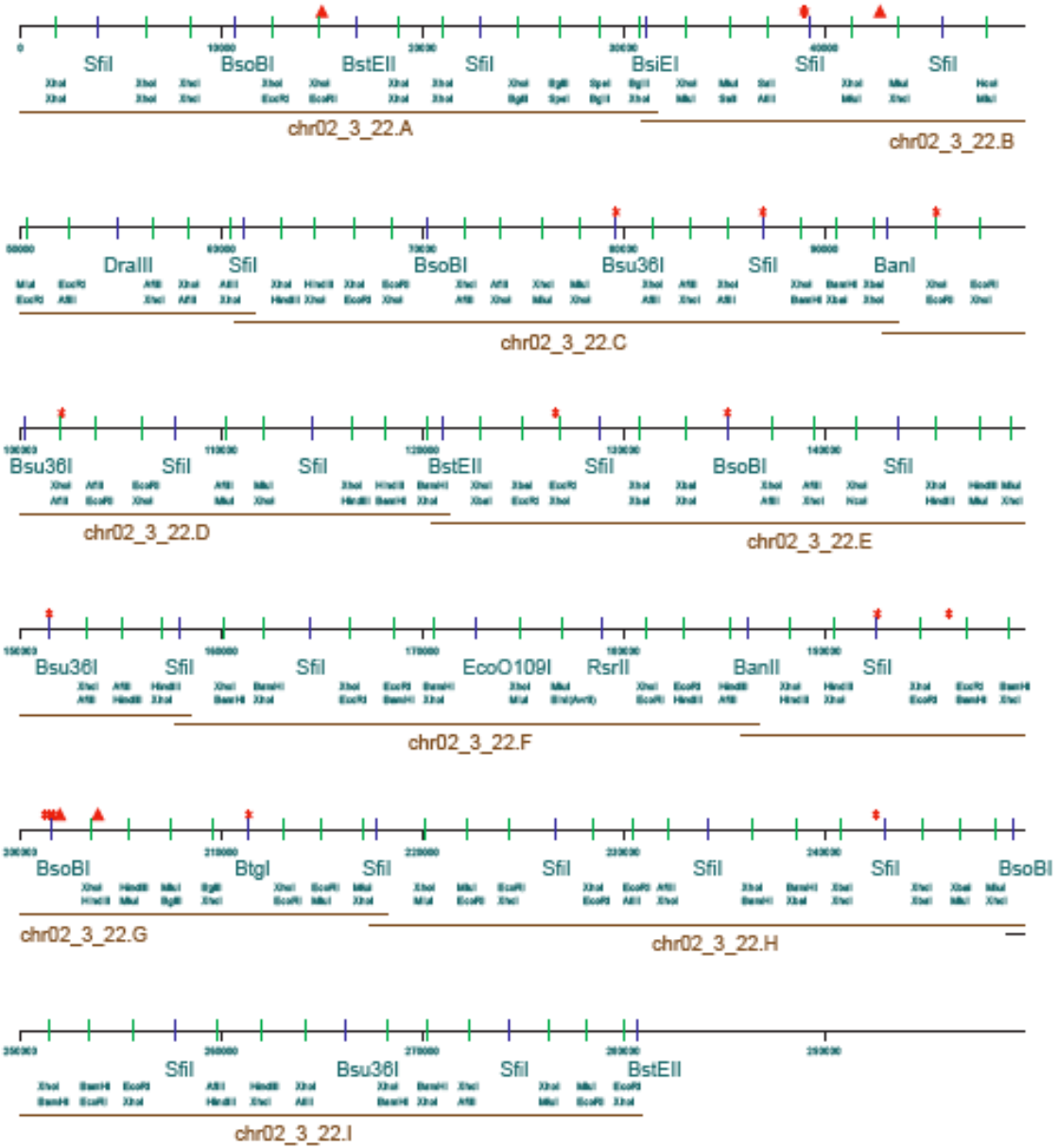
- 40bp overlap between each MiniChunk fragment
- Maximum 7 fragments
- All chunk assembly been done for ChrII

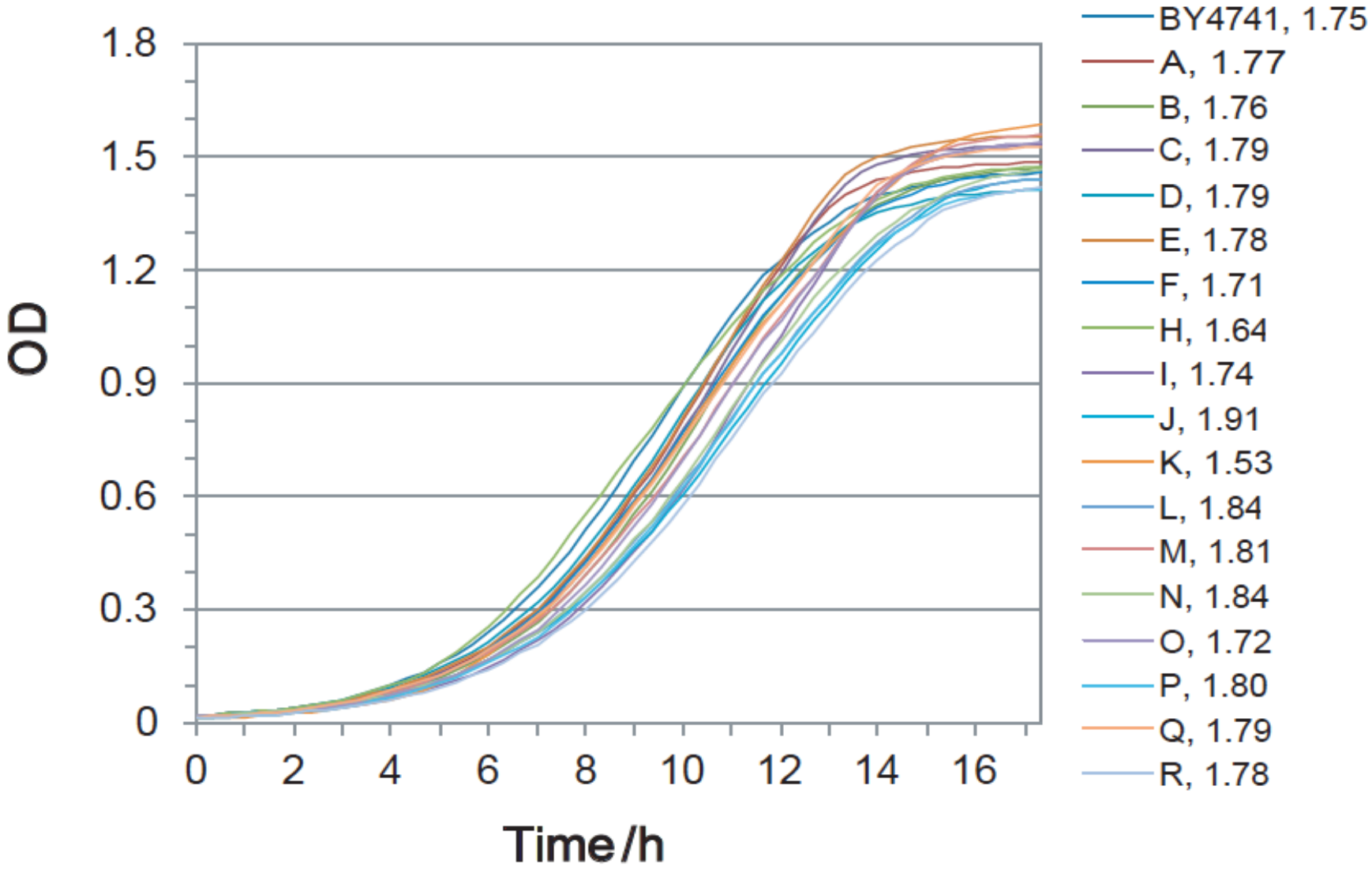
Sequencing of *semi*Syn ChrII

Colony ID	Clean read base(M)	# of Reads	D
synII A-I _1	251.7	1,327,610	
synIIA-I _2	280.2	1,520,249	

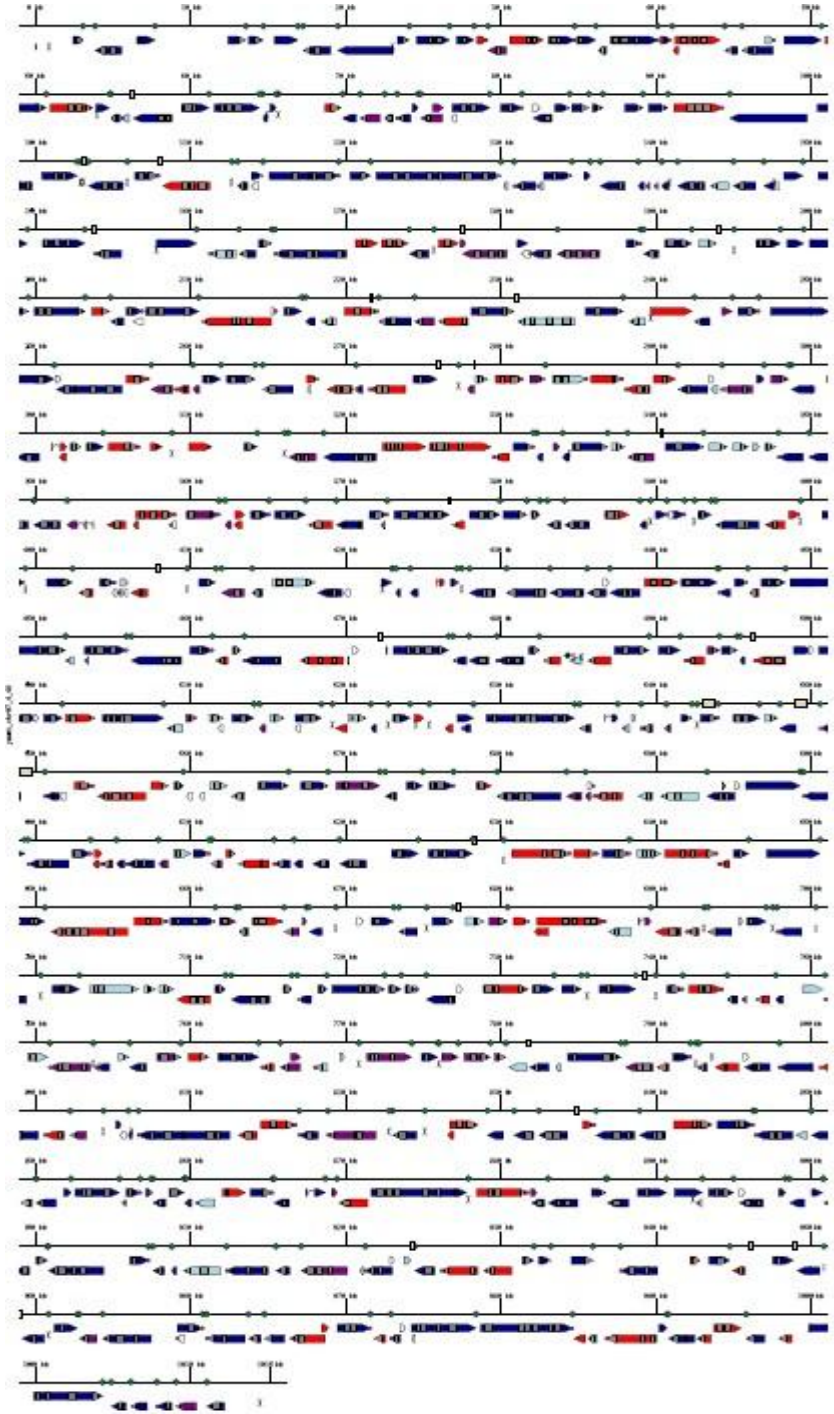
▲

Variation type		Number
*Indel	Deletion	3
	Insertion	1
SNV		22





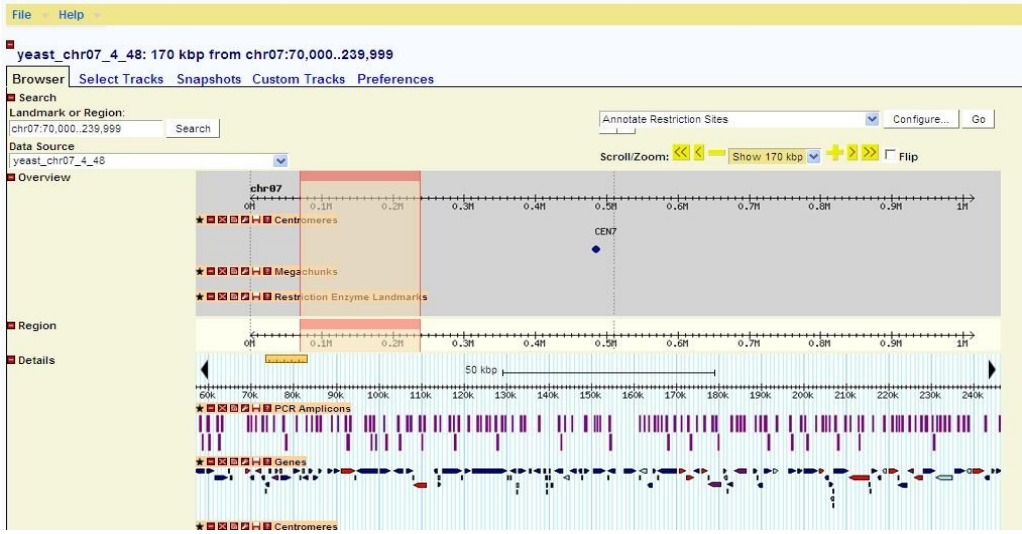
Growth Curves after each "Replacements"



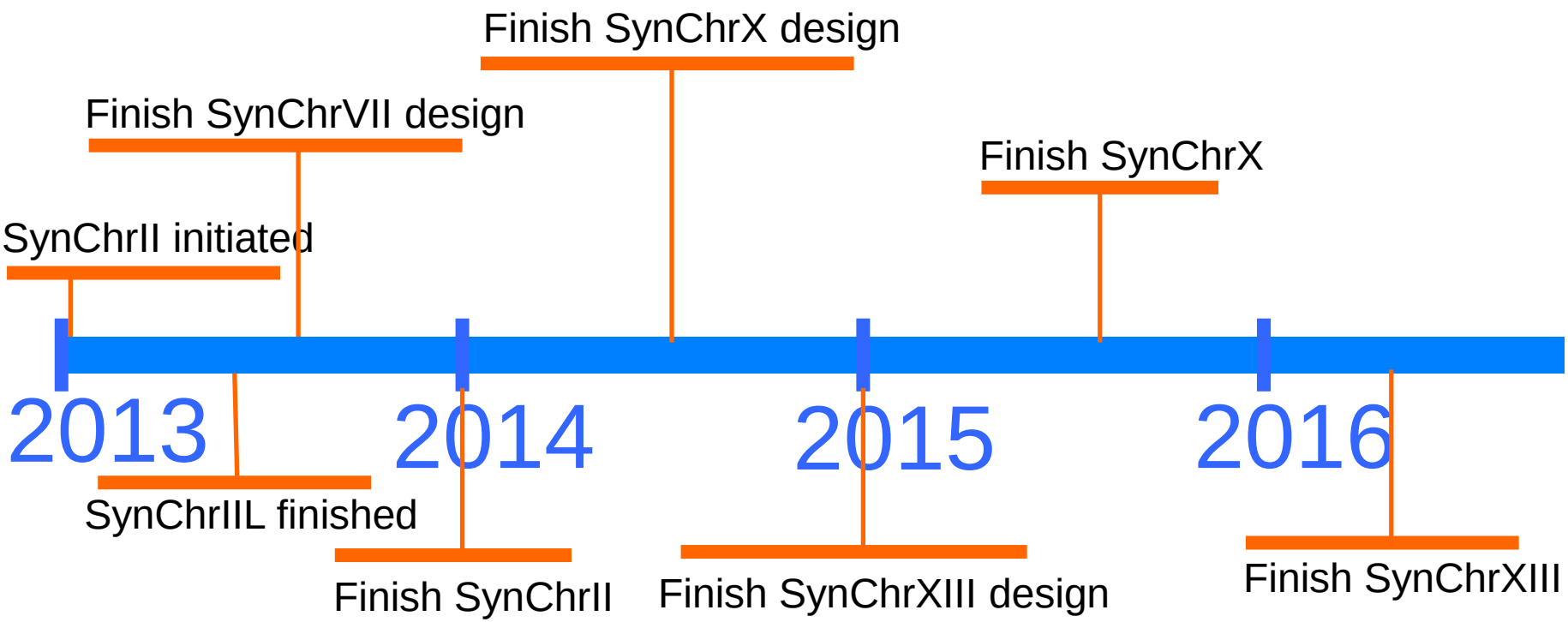
ChrVII (1,015kb)

Design summary

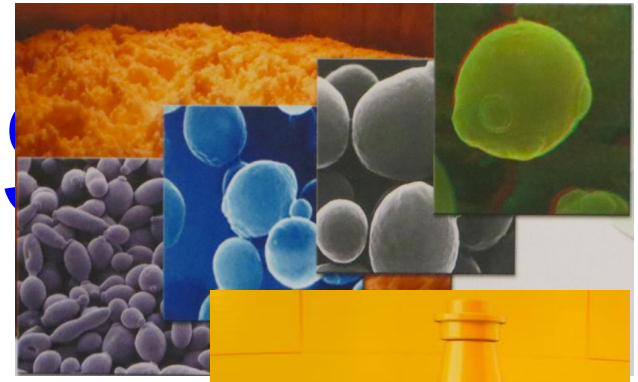
Add PCR tags	706 pairs
TAG -> TAA	128/129
Insert loxpsym sites	316
Delete tRNAs	36
Delete LTR/transposons	44 LTR; 16 transposon relevant
Delete Introns	14/26
Replace telomeres	2



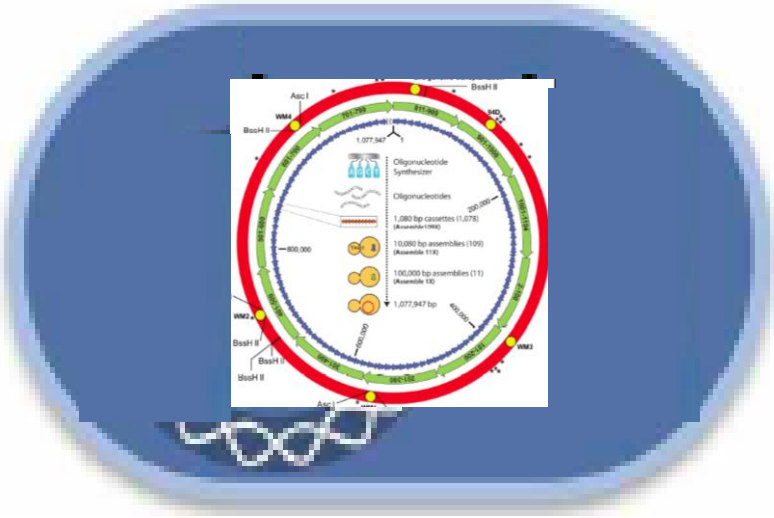
Timeline of Sc2.0 in BGI



Design Yeas



LEGO YEAS



What

will be the next milestone in SynGenome?

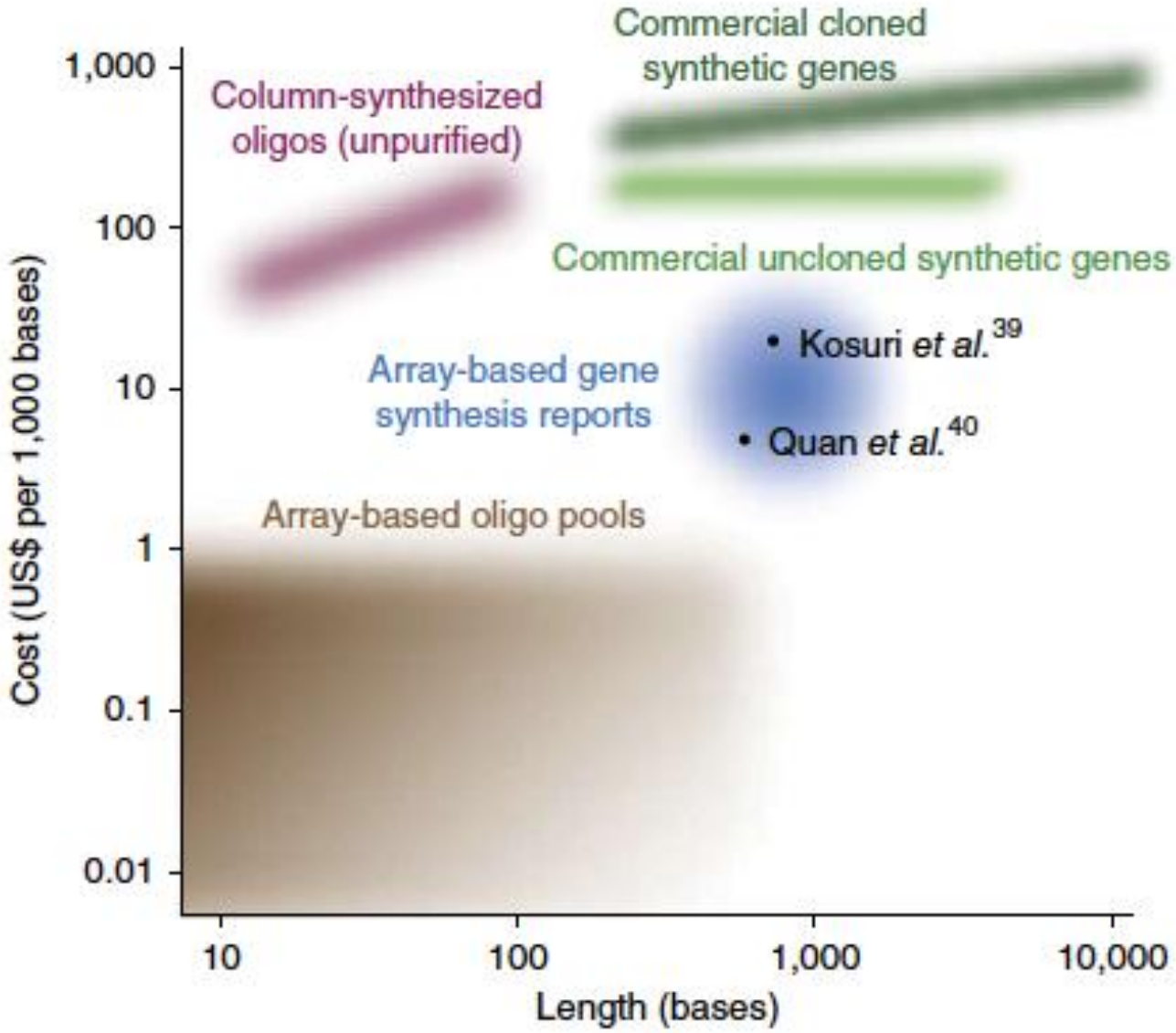
A multicellular eukaryotic genome!?

“Small” Genomes Sequenced

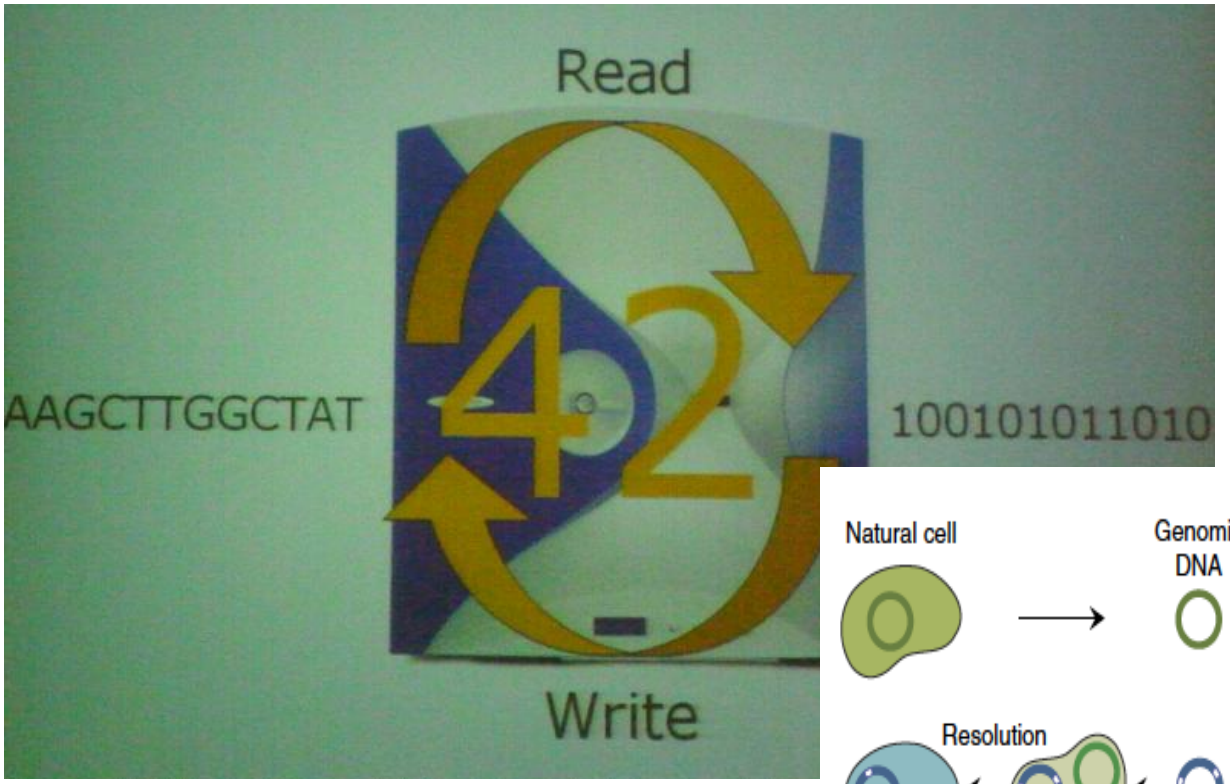
- 1982 Bacteriophage ΦX174 5 Kb
- 1995 *Haemophilus influenzae* 1,8 Mb
- 1996 *Saccharomyces cerevisiae* 15 Mb

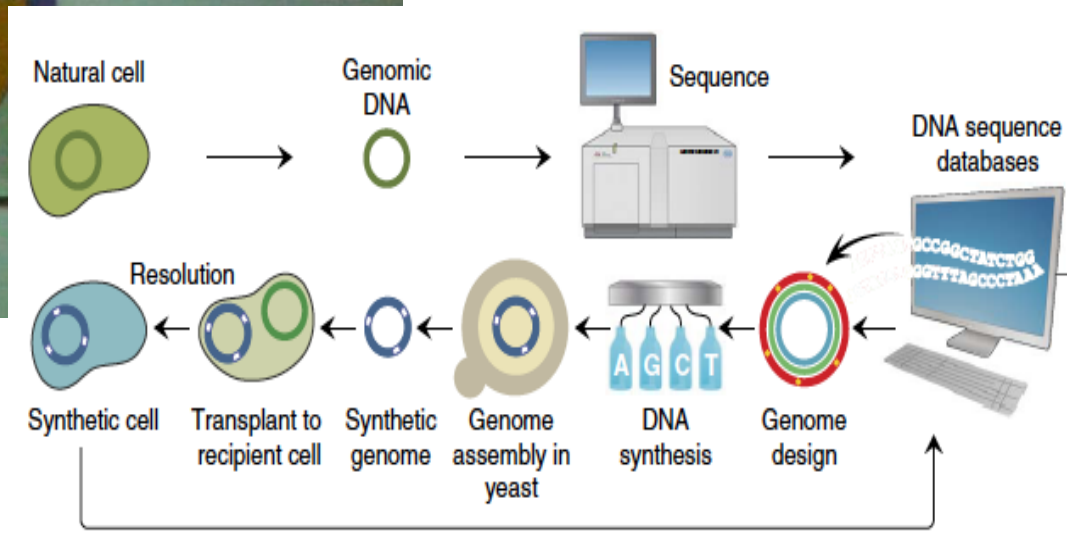
- **1998 *Caenorhabditis elegans* 95.5 Mb**
- **1999 *Arabidopsis thaliana* 117 Mb**
- **1999 *Drosophila melanogaster* 180 Mb**

- Fission yeast…
- A monocellular organism with differentiations



Lengths and costs of oligo/gene synthesis technologies





从“解读”到“编写”

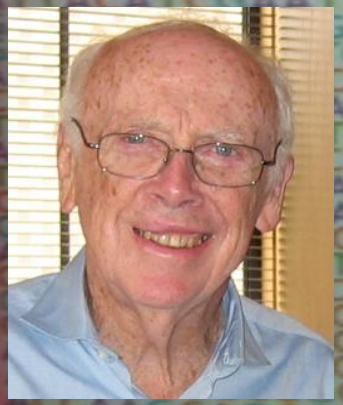
from READING *to* WRITING

A natural development and the highest stage of GENOMICS

Two Pillars of Genomics

“Life is of sequence”

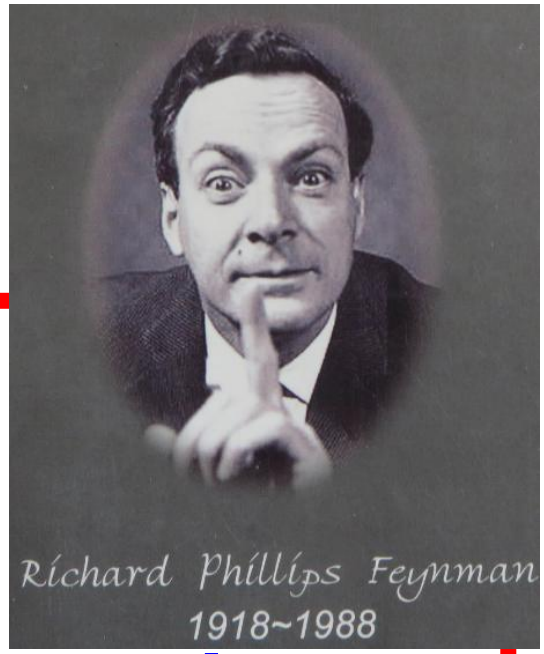
生命是序列的



“Life is digital!”

生命是数据的

To DIY or not to dIY, This is the question.

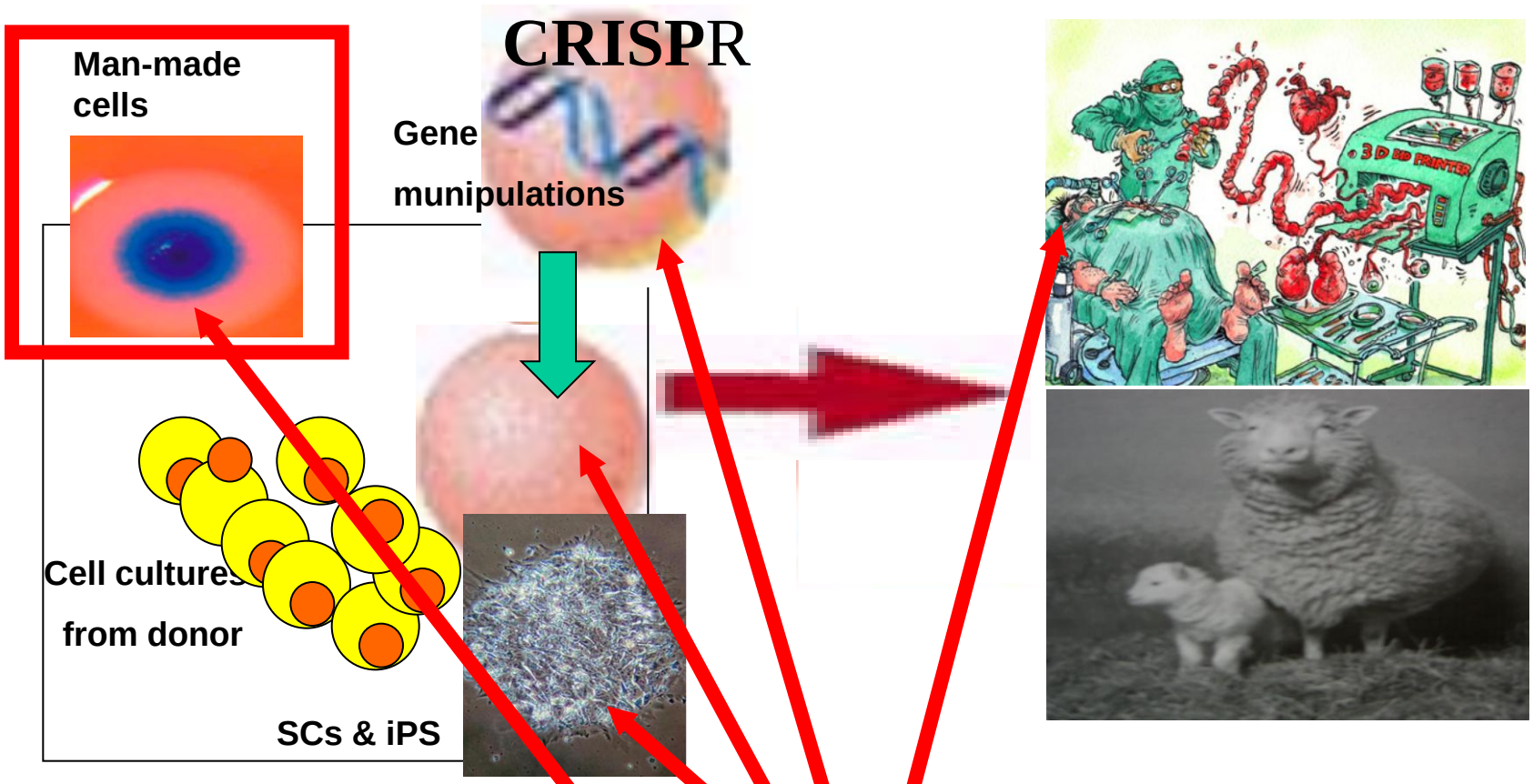


“What I cannot create, I do not understand.”

“不能创造，不算知道”。

Richard Feynman (US physicist)

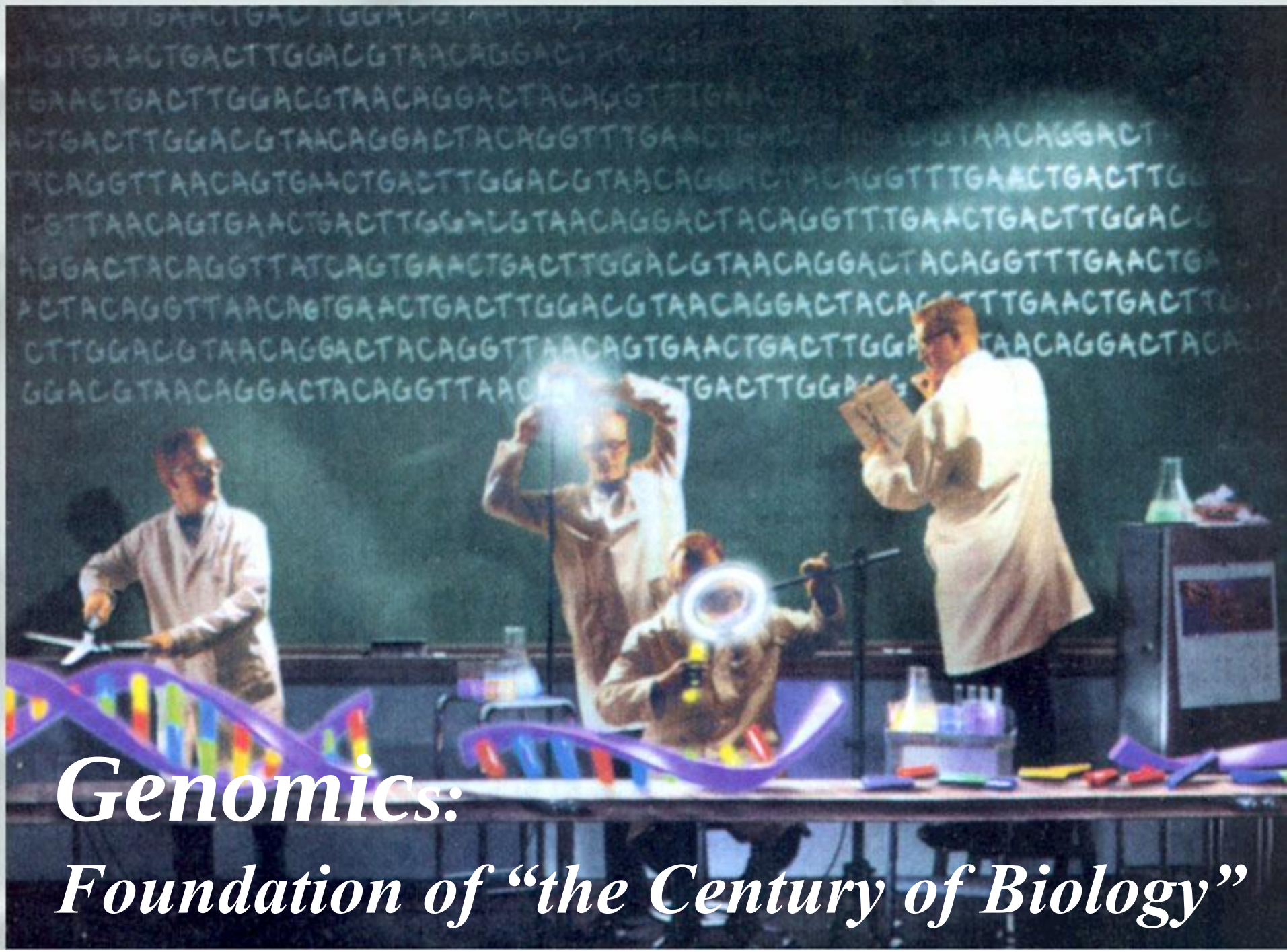
“The BioCentury”



Genomics

Foundation of Life Sciences and biotechs

by providing knowledge of genes, genomes and “Three Networks”

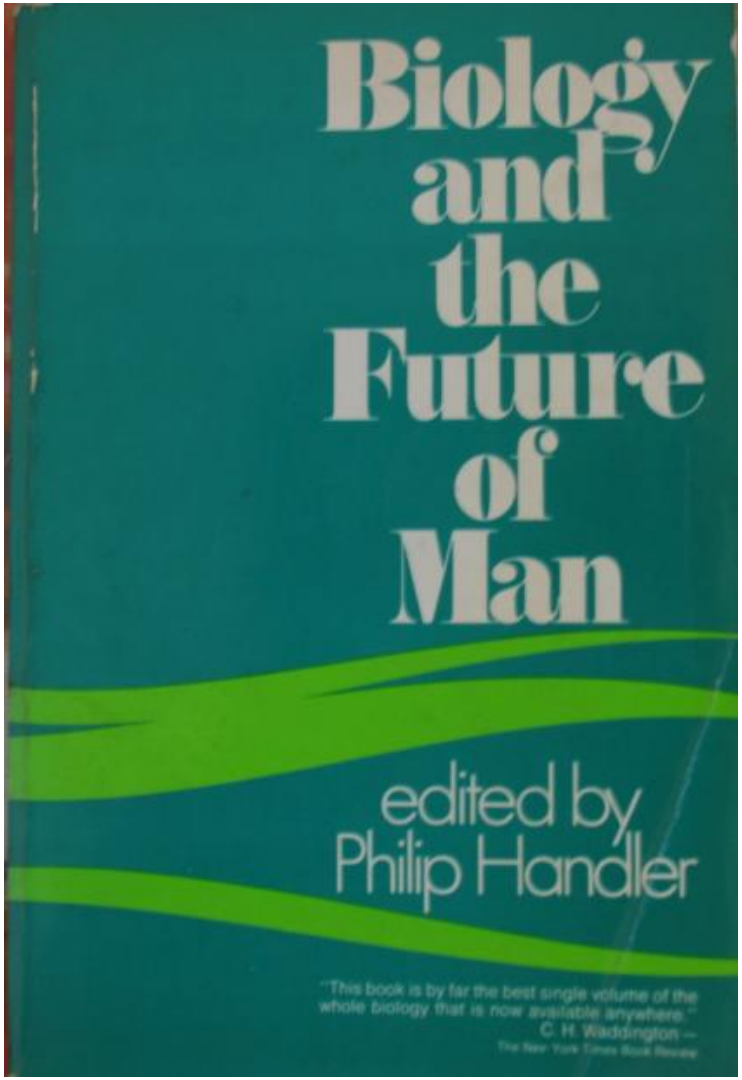


Genomics:
Foundation of "the Century of Biology"

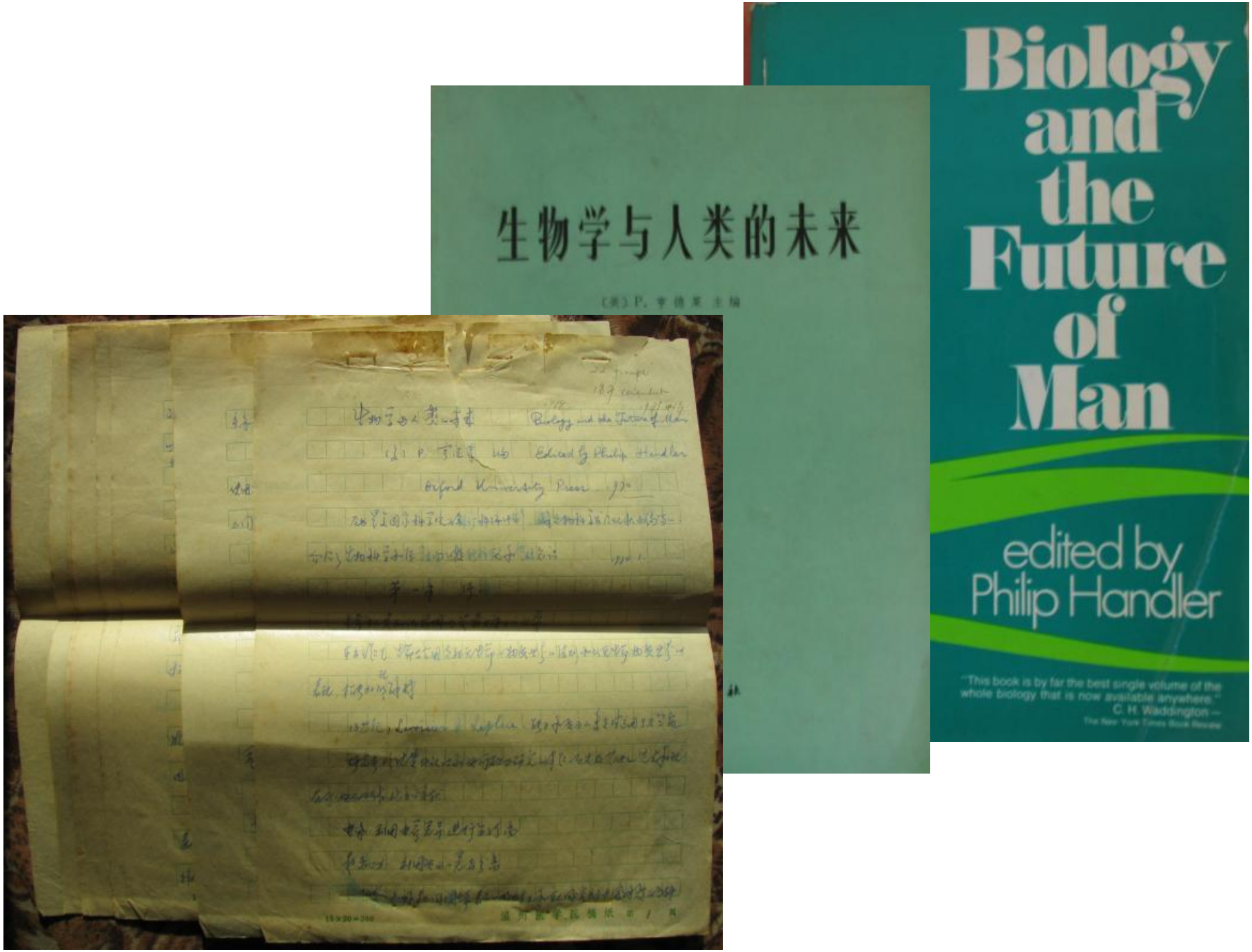
To Share is
to Share the Responsibilities!

To Share is
to Share the Future!

Biology and the Future of Man



1968


生物学与人类的未来
Biology and the Future of Man
edited by
Philip Handler
"This book is by far the best single volume of the whole biology that is now available anywhere."
C. H. Waddington

tion he is to strive.

Man's view of himself has undergone many changes. From a unique position in the universe, the Copernican revolution reduced him to an inhabitant

***Homo sapiens*, the creation of Nature, has transcended her. From a product of circumstances,**

he has risen to responsibility.

has reached beyond the hard regularities of physical phenomena. *Homo sapiens*, the creation of Nature, has transcended her. From a product of circumstances, he has risen to responsibility. At last, he is Man. May he behave so!

tion he is to strive.

Man's view of himself has undergone many changes. From a unique posi-

At last, he is Man.

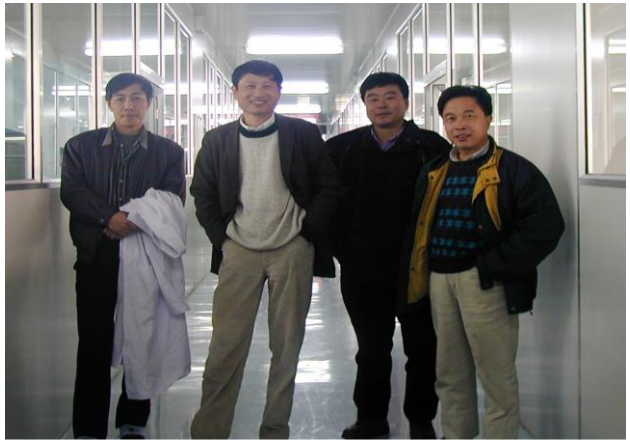
May he behave so!

sapiens, the creation of Nature, has transcended her. From a p

circumstances, he has risen to responsibility. At last, he is Man. May he behave so!

A long march towards the brilliant future requires generations' effort!

Mutual trust & personal friendship are more important than ever!



~ 5500 staffs

3 February 2012
335: 516-519, 2012

Science

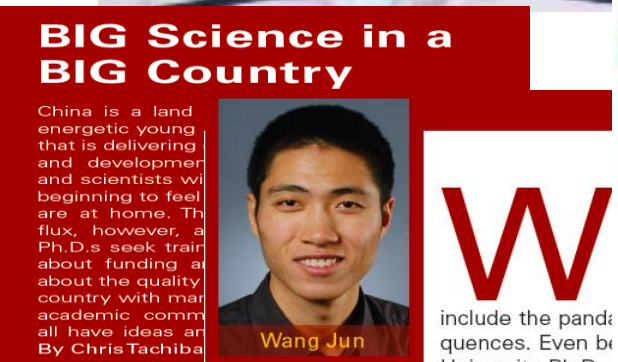
中国已经成为测序强国

China's Sequencing Powerhouse Comes of Age

by Dennis Normile

Three Generations Leaders of BGI

“三代同堂”


BIG Science in a BIG Country

China is a land energetic young that is delivering and developmen and scientists wi beginning to feel are at home. Th flux, however, a Ph.D.s seek train about funding a about the quality country with mar academic comm all have ideas ar

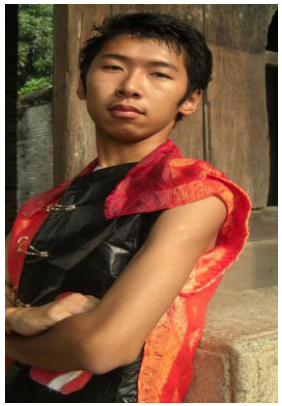
By Chris Tachiba

Wang Jun

W

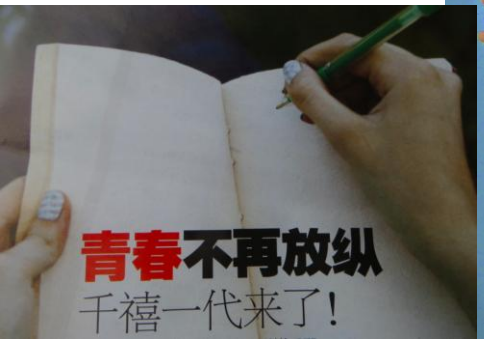
include the panda quences. Even be

The 4th Generation is already there!



Education is for the day after tomorrow!

“Gene should also be taught to the kids!”



The 5th generation is growing up!

Give me!

"Give me those who are younger

Give me those
who are more bravely thinking,

Give me those who are more devoted!

I am responsible for
all the troubles you would make,

I am obligated to
find right teachers for you."

LETTER FROM SHENZHEN

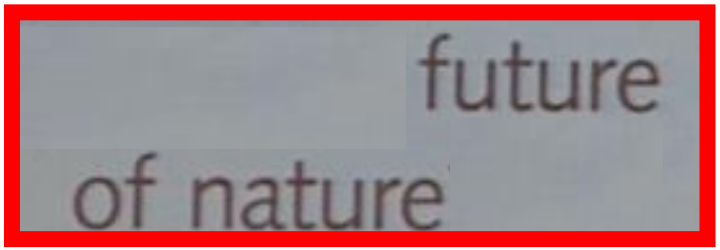
THE GENE FACTORY

B.G.I.'s leaders are aware of the perception that the company is little more than a biological data mill. The next afternoon, before leaving Boston, I attended a luncheon hosted by the company. Three hundred people filled a lecture hall that usually holds far fewer.

Most had come to hear Huanming Yang, the B.G.I. chairman, deliver a long, emotional presentation that included a PowerPoint display with ninety-one slides. Yang is warm and self-effacing, and he thanked a roster of American biologists for their help and "collaboration." In talking about the promise of genomics, he invoked Martin Luther King, Jr.,'s "I Have a Dream" speech and the Declaration of Independence. It was a "Kumbaya" moment in a field where the soul is rarely mentioned. Yang referred to his company as "an unruly adolescent," and ended his talk by saying, "Please do me a favor: Take the young B.G.I.'ers as your friends, as your students. To treat them as you treated me, to teach them as you taught me. I assure you it is very rewarding. It is not only for a successful project; it is also for the brilliant future of mankind."

Confident of the brilliant

&



Confident
of
you all younger generation!

Scientific American: Research Leader of the Year, 2002

“When you drink from the well, don’t forget who “饮水思源” helped dig it!”

Gratitude speech at the reception by *Scientific American*

Salute all collaborators & supporters
all teachers & friends

欢迎参观华大
Welcome to BGI!

期盼与君合作

Let’s collaborate!

Acknowledgment

Sponsors

Chinese Academy of Sciences
Ministry of Science and Technology
National Natural Science Foundation
Zhejiang Government
Chongqing Government
Beijing Municipal Government
Hangzhou Municipal Government
Yueqing Municipal Government

All our supporters, collaborators international advisors colleagues and friends and all my young staff

Thanks !

yanghm@genomics.cn