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,
White House 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 our other Ambassadors from the United Kingdom, Japan, Germany, France, and China. I'd 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！
President lauds new map of human life

[ Xinhua 06/29/2000 Beijing ]  In a speech yesterday in Beijing, President Jiang Zemin applauded the recent completion of the first working draft of the Human Genome Project “a great scientific project” in the scientific development of life sciences, medicine and pharmaceutical research. Jiang called the Human Genome Project “a great scientific project” in the scientific history of human beings and one of vital importance to the development of life sciences, medicine and pharmaceutical study.

The completion of the working draft is a milestone in the overall process of the Human Genome Project, said Jiang, adding that it results from teamwork by scientists all over the world. The human genome sequence is the common heritage of mankind and should bring wealth and benefit to all human beings, Jiang said. Jiang congratulated the scientists and technical staff involved and thanked the participating Chinese scientists for their hard work. Jiang said he expects Chinese scientists to make further contributions to the completion of human genome mapping and China’s scientific genome research as well.
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
China

1964

1969 USA

? China

2000 USA, UK, J, F, G, & China
Economic Impact of the Human Genome Project

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

Science Feb. 16, 2001
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

“We began with nothing!”

9”9’9h 9/9/1999

“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.”
With a diploid genome of around 3,000 megabase pairs, with no apparent regularities in base composition, methods in sight for breaking the DNA into 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!”
**HGP – A revolution in life sciences**

*We have had it done! We have had it done together through the vast international collaboration!*

<table>
<thead>
<tr>
<th>国别</th>
<th>中心</th>
<th>包含区域(Mb)</th>
<th>贡献率 (%)</th>
<th>国家贡献率 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>美国</td>
<td>Wash U + MIT</td>
<td>1050</td>
<td>~35.0</td>
<td>.54.0</td>
</tr>
<tr>
<td></td>
<td>JGI – DOE</td>
<td>310</td>
<td>~10.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Baylor</td>
<td>230</td>
<td>~8.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GTC</td>
<td>50</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stanford</td>
<td>15</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U Wash (H)</td>
<td>25</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U Wash (O)</td>
<td>25</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>英国</td>
<td>Sanger</td>
<td>1000</td>
<td>~33.0</td>
<td>33.0</td>
</tr>
<tr>
<td>日本</td>
<td>Riken</td>
<td>180</td>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Keio</td>
<td>23</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>法国</td>
<td>Genoscope</td>
<td>85</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IMB (Jena)</td>
<td>50</td>
<td>~1.7</td>
<td></td>
</tr>
<tr>
<td>德国</td>
<td>MPIMG</td>
<td>69</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GFF (Gesellschaft)</td>
<td>6</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>中国</td>
<td>Beijing</td>
<td>30</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

USA  UK  Japan  France  Germany  China
Only an individual genome

Too expensive (US$ 3b)

Too slow (13 yrs)

(~3000 staffs in 16 centers from 6 countries)
15 Years Later

HiSeq2000 in the world (June 15, 2011)

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
As Jun Wang, executive director of the Beijing Genomics Institute, put it, there is “strong network effect... the health profile and personal genome 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

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

* Sequencing Capacity Index
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.

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

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”
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)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Science</td>
<td>31.477</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Science Translational Medicine</td>
<td>14.414</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Nature</td>
<td></td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td>1</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Nature Genetics</td>
<td></td>
<td>10</td>
<td>6</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Nature Communications</td>
<td></td>
<td>14</td>
<td>9</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Nature Biotechnology</td>
<td></td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Nature Reviews Drug Discovery</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Nature Methods</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Nature Reviews Genetics</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Nature Reviews Microbiology</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Cell</td>
<td>33.116</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Cell Stem Cell</td>
<td>22.151</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>NEJM</td>
<td>54.420</td>
<td>32</td>
<td>36</td>
<td>26</td>
<td>134</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Your research paper reflects who you are!

Flowers of collaboration!
BGI, China

5. Beijing Genomics Institute (BGI), Shenzhen, China
### ASIA-PACIFIC TOP 20 INSTITUTIONS

2013

#### Research Articles

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

#### 2012

<table>
<thead>
<tr>
<th>Rank</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The University of Tokyo</td>
</tr>
<tr>
<td>2</td>
<td>Chinese Academy of Sciences (CAS)</td>
</tr>
<tr>
<td>3</td>
<td>Kyoto University</td>
</tr>
<tr>
<td>4</td>
<td>RIKEN</td>
</tr>
<tr>
<td>5</td>
<td>Osaka University</td>
</tr>
<tr>
<td>6</td>
<td>The University of Melbourne</td>
</tr>
<tr>
<td>7</td>
<td>Nagoya University</td>
</tr>
<tr>
<td>8</td>
<td>University of Science &amp; Technology of China</td>
</tr>
<tr>
<td>9</td>
<td>National University of Singapore</td>
</tr>
<tr>
<td>10</td>
<td>Tohoku University</td>
</tr>
<tr>
<td>11</td>
<td>Tsinghua University</td>
</tr>
<tr>
<td>12</td>
<td>Chinese Academy of Sciences (CAS)</td>
</tr>
<tr>
<td>13</td>
<td>Osaka University</td>
</tr>
<tr>
<td>14</td>
<td>Tohoku University</td>
</tr>
<tr>
<td>15</td>
<td>Agency for Science, Technology &amp; Research (A*STAR)</td>
</tr>
<tr>
<td>16</td>
<td>Shanghai Jiao Tong University</td>
</tr>
<tr>
<td>17</td>
<td>BGI</td>
</tr>
<tr>
<td>18</td>
<td>Peking University</td>
</tr>
<tr>
<td>19</td>
<td>Monash University</td>
</tr>
<tr>
<td>20</td>
<td>Nanjing University</td>
</tr>
</tbody>
</table>

Asia-Pacific 前20名
<table>
<thead>
<tr>
<th>Rank</th>
<th>Institution</th>
<th>Country</th>
<th>Score</th>
<th>Guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>87</td>
<td>BGI, China</td>
<td>China</td>
<td>10.76</td>
<td>34</td>
</tr>
<tr>
<td>88</td>
<td>Nagoya University, Japan</td>
<td>Japan</td>
<td>10.68</td>
<td>33</td>
</tr>
<tr>
<td>89</td>
<td>University of Copenhagen, Denmark</td>
<td>Denmark</td>
<td>10.64</td>
<td>77</td>
</tr>
<tr>
<td>90</td>
<td>University of Massachusetts Medical School (UMMS), USA</td>
<td>USA</td>
<td>10.61</td>
<td>30</td>
</tr>
<tr>
<td>91</td>
<td>The University of Queensland (UQ), Australia</td>
<td>Australia</td>
<td>10.60</td>
<td>50</td>
</tr>
<tr>
<td>92</td>
<td>University of California Irvine (UCI), USA</td>
<td>USA</td>
<td>10.60</td>
<td>38</td>
</tr>
<tr>
<td>93</td>
<td>Howard Hughes Medical Institute (HHMI), USA</td>
<td>USA</td>
<td>10.60</td>
<td>45</td>
</tr>
<tr>
<td>94</td>
<td>Karolinska Institute, Sweden</td>
<td>Sweden</td>
<td>10.57</td>
<td>22</td>
</tr>
<tr>
<td>95</td>
<td>University of Groningen, Netherlands</td>
<td>Netherlands</td>
<td>10.57</td>
<td>15</td>
</tr>
<tr>
<td>96</td>
<td>The Institute of Photonic Sciences (ICFO), Spain</td>
<td>Spain</td>
<td>10.54</td>
<td>40</td>
</tr>
<tr>
<td>97</td>
<td>The Pennsylvania State University, USA</td>
<td>USA</td>
<td>10.54</td>
<td>32</td>
</tr>
<tr>
<td>98</td>
<td>National Research Council (CNR), Italy</td>
<td>Italy</td>
<td>10.51</td>
<td>28</td>
</tr>
<tr>
<td>99</td>
<td>Aarhus University (AU), Denmark</td>
<td>Denmark</td>
<td>10.49</td>
<td>58</td>
</tr>
<tr>
<td>100</td>
<td>University of Pittsburgh, USA</td>
<td>USA</td>
<td>10.49</td>
<td>35</td>
</tr>
</tbody>
</table>

**Global Top 200 (119) (2012)**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Institution</th>
<th>Country</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>119</td>
<td>BGI</td>
<td>China</td>
<td>6.27</td>
</tr>
<tr>
<td>120</td>
<td>University of Southampton</td>
<td>UK</td>
<td>6.25</td>
</tr>
<tr>
<td>121</td>
<td>Oregon Health &amp; Science University (OHSU)</td>
<td>USA</td>
<td>6.21</td>
</tr>
<tr>
<td>122</td>
<td>University of Bonn</td>
<td>Germany</td>
<td>6.20</td>
</tr>
<tr>
<td>123</td>
<td>University of Texas, Dallas</td>
<td>USA</td>
<td>6.19</td>
</tr>
<tr>
<td>124</td>
<td>University of Basel</td>
<td>Switzerland</td>
<td>6.19</td>
</tr>
<tr>
<td>125</td>
<td>University of Hamburg</td>
<td>Germany</td>
<td>6.12</td>
</tr>
<tr>
<td>126</td>
<td>Catholic University of Leuven</td>
<td>Belgium</td>
<td>6.11</td>
</tr>
<tr>
<td>127</td>
<td>University of Virginia</td>
<td>USA</td>
<td>6.06</td>
</tr>
<tr>
<td>128</td>
<td>Aarhus University (AU)</td>
<td>Denmark</td>
<td>6.05</td>
</tr>
<tr>
<td>129</td>
<td>Mayo Clinic</td>
<td>USA</td>
<td>6.04</td>
</tr>
<tr>
<td>130</td>
<td>University of Dundee</td>
<td>UK</td>
<td>5.99</td>
</tr>
<tr>
<td>131</td>
<td>Wellcome Trust Sanger Institute</td>
<td>UK</td>
<td>5.88</td>
</tr>
<tr>
<td>132</td>
<td>University of London - King's College London</td>
<td>UK</td>
<td>5.87</td>
</tr>
<tr>
<td>133</td>
<td>Albert Einstein College of Medicine of Yeshiva University</td>
<td>USA</td>
<td>5.84</td>
</tr>
<tr>
<td>134</td>
<td>Kyushu University</td>
<td>Japan</td>
<td>5.81</td>
</tr>
<tr>
<td>135</td>
<td>University of Nottingham</td>
<td>UK</td>
<td>5.80</td>
</tr>
<tr>
<td>136</td>
<td>Karolinska Institute</td>
<td>Sweden</td>
<td>5.68</td>
</tr>
</tbody>
</table>
They are ever our teachers & pioneers

Building the sequer pan-genome

Nature's sequence and de novo a giant panda genome

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

Ancient human genome sequence of an extinct Palaeo-Eskimo

The diploid genome sequence of an individual

We never forget:

All collaborators

WE = All collaborators

Building the sequer pan-genome

Nature's sequence and de novo a giant panda genome

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

Ancient human genome sequence of an extinct Palaeo-Eskimo

The diploid genome sequence of an individual

We never forget:

All collaborators

Building the sequer pan-genome

Nature's sequence and de novo a giant panda genome

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

Ancient human genome sequence of an extinct Palaeo-Eskimo

The diploid genome sequence of an individual

We never forget:

All collaborators

Building the sequer pan-genome

Nature's sequence and de novo a giant panda genome

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

Ancient human genome sequence of an extinct Palaeo-Eskimo

The diploid genome sequence of an individual

We never forget:

All collaborators

Building the sequer pan-genome

Nature's sequence and de novo a giant panda genome

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

Ancient human genome sequence of an extinct Palaeo-Eskimo

The diploid genome sequence of an individual

We never forget:

All collaborators
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.¹³
Never forget:

Nothing would have been possible without all those pioneers, our partners, teachers and friends.
Joe Biden is wrong. China does innovate

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,
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
“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

A major part of Big Data
“Your DNA enters the digital age!”

Francis Collins
USA Today, May, 2014
Selection & Analysis
Data – Information – Knowledge - Laws

Safety & Security
Data, privacy & society

Sustainability
IP Right, Equal Access

Big Data for International Scientific Programmes:
Challenges and Opportunities

A Statement of Recommendations and Actions
Beijing 9 June 2014
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
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)!"
“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.
“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
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 Scientists (June, 2014)**

Members of the 3000 Rice Genomes Project last month (May 28) delivered on their promise to make public the genomic sequences of 3000 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贫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
Where once there was the genome, now there are thousands of ‘omes.
“-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.”

To Characterize & Rectify Biological Systems
From Genomics to Trans-omics

Big Data
Big Science
Big Platform
Big Applications
Big Collaboration
Big Sharing

Genome
Transcriptome
Proteome
Metabolome
Microbiome
Epigenome
Exposome
Social graph
Biosensors
Imaging
Phenomes
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

I. Sanger method published
   - Gel-based Sanger sequencing
   - Kilobases of DNA per day per machine

II. Capillary gel electrophoresis developed
   - Massively parallel ‘next generation’ sequencing

III. III

A breakthrough, an opportunity!
Three breakthroughs in sequencing techs

3. From Capillary- to Massively Parallel Seq
The Biggest Investment

“The biggest voice”
The biggest bill ever in this sector

by Nick Loman (University of Birmingham)

http://pathogenomics.bham.ac.uk
The Biggest Investment

The Best Software

“The biggest voice”

ARTICLES

The diploid genome sequence of an Asian individual

Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome de novo assembly
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
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 deserves all appreciation for advocating sequencing globally.
The Biggest Investment

“The biggest voice”

The Best Software

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

Yingrui Li1,10, Hancheng Zheng1,10, Ruthbang Luo1,3,10, Honglong Wu4,1,10, Hongmei Zhu1, Ruijiang Li1, Hongzhi Cao1,4, Boxin Wu1, Shuita Huang1,3, Haojing Shao1,3, Hanzhao Ma1,2, Fan Zhang1,2, Shuilian Feng1, Wei Zhang1, Hongli Du1, Geng Tian1, Jingxian Li1, Xiuqing Zhang1, Songgang Li1, Lars Bolund1,5, Karsten Kristiansen1,4, Adam J de Smith7, Alexandra I F Blakemore7, Lachlan J M Cohn7, Huaming Yang1, Jian Wang1 & Jun Wang1,8,9
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

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

Bradnam, UC Davis, CSHL, 12 May, 2010
The Next Breakthroughs …

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.

“A landmark paper should be read by all plant biologists.”
Animals

Vertebrates
- cynomolgus & Chinese rhesus
- naked mole rat

Invertebrates
- Ascaris suum
- Moth
- Oyster
- Yak

Online publication dates:
- July 1, 2012
- Sept. 19, 2012
- Jan 13, 2013
- Aug. 2010
- 23, 2011
- 735-741, 2011
“The Genome 10K Project”

<table>
<thead>
<tr>
<th>Group</th>
<th>Known Species</th>
<th>10K Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birds</td>
<td>9723</td>
<td>5074</td>
</tr>
<tr>
<td>Reptiles</td>
<td>9002</td>
<td>3297</td>
</tr>
<tr>
<td>Mammals</td>
<td>5416</td>
<td>1826</td>
</tr>
<tr>
<td>Amphibians</td>
<td>6570</td>
<td>1760</td>
</tr>
<tr>
<td>Fish</td>
<td>31564</td>
<td>4246</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>62275</strong></td>
<td><strong>16203</strong></td>
</tr>
</tbody>
</table>
Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle.
To sequence most, if not all, stains of a microorganism.
Most, if not all, of the plague 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.
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

5,329 cases (969 health care workers) and 334 deaths in 14 provinces

China is being attacked!
On 2 June, Chinese scientists announced that they had deciphered the microbe's entire 5.2-million-base-pair genome and immediately made the DNA sequence available for researchers to download.

Science 332 (6035): 1249  10 June 2011
“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)

For the health care of the people
BGI Plans to Sequence the World

The Global DNA Sequencing Powerhouse Based in Shenzhen

By Paul Diehl, About.com Guide

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

14种粮食作物

70种园艺作物

39种禽畜牧渔
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

Precision, Prediction, Prevention, Participation, Personalization, Integrated, Targeted
5P+IT Medicine in the BioCentury

- Precision
- Prediction
- Prevention
- Participation
- Personalization
- Integrated
- Targeted
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

Individualized genomic medicine
From prewomb to tomb

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.
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,
   ancient DNA/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
Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants.
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.

<table>
<thead>
<tr>
<th>Inheritance Pattern</th>
<th>January 2007</th>
<th>July 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autosomal</td>
<td>1,851</td>
<td>3,525</td>
</tr>
<tr>
<td>X Linked</td>
<td>169</td>
<td>277</td>
</tr>
<tr>
<td>Y Linked</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Mitochondrial</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td>2,048</td>
<td>3,834</td>
</tr>
</tbody>
</table>
Whole-Genome Sequencing in Autism Identifies Hot Spots for De Novo Germline Mutation

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

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

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

Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation.
Identification of genomic alterations in oesophageal squamous cell cancer

Yongmei Song*, Lin Li*, Yunwei Ou†, Zhibo Gao†, Ermin Li†, Xiangchun Li†, Weimin Zhang†, Haqian Wang†, Liyan Xu†, Yong Zhou†, Xiaojuan Ma†, Lingyan Liu†, Zhongzhao†, Xuanlin Huang†, Jing Fan†, Liya Dong‡, Gang Chen‡, Liying Ma‡, Jie Yang‡, Longyun Chen*, Minghui He*, Maol Li*, Xuehan Zhuang§, Kai Huang, Kunlong Qu*, Quangliang Yin, Quangwu Guo, Qiang Peng, Peishan Chen, Zhiyong Wu, Jianyi Wu, Ling Ma, Jinyang Zhao, Longhai Luo, Ming Pu, Raiman Xu, Bo Chen, Yingrui Li, Tongli, Mingrong Wang, Zhuhua Liu, Dongxin Lin, Xueqin Zhang, Huanming Yang, Jun Wang and Qimin Zhao.

Oesophageal cancer is one of the most aggressive cancers and is the sixth leading cause of cancer death worldwide. 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%) [1]. 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 genome-wide genotyping of ESCC cases and the results 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, P53CA, NOTCH1, NF2), and two have not previously been described in ESCC (ADAM29 and FAM135B). Notably, FAM135B is identified as a novel cancer-associated gene as assayed for its ability to promote malignancy of ESCC cells. Additionally, MIR548K, a microRNA encoded in the amplified 1q13.3-13.4 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 (MEL2, also called KMT2D, ASH2L, MLL3, 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 common pathogenetic 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
   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
A human gut microbial gene catalogue established by metagenomic sequencing

March 4, 2010

Enterotypes of the human gut microbiome

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 fecal 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.
Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci.
A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes


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 years1. 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 (β = 3.8 mmol L⁻¹, \( P = 2.5 \times 10^{-35} \)) and serum insulin (β = 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 (β = -0.18 mmol L⁻¹, \( P = 1.1 \times 10^{-6} \)) and fasting serum insulin (β = -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 (β = 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 consistent 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 traits2-4 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.
A metagenome-wide association study of gut microbiota in type 2 diabetes

Junjie Qin¹*, Yingrui Li¹*, Zhiming Cai²*, Shenghui Li¹*, Jianfeng Zhu¹*, Fan Zhang³*, Suisha Liang¹, Wenwei Zhang¹, Yuanlin Guan¹, Dongqian Shen¹, Yangping Peng¹, Dongdong Zhang¹, Zhuye Li¹, Wenjuan Wu¹, Youwen Qin¹, Wenbin Xue¹, Junhua Li¹, Lingchuan Han³, Donghui Liu², Peitao Wu³, Xiaoming Li², Xiaojun Guan², Aifa Tang², Shilong Zhong², Xiaoping Li¹, Weineng Chen¹, Ran Xu¹, Minlan Huang¹, Jiangang¹, Meihua Gong¹, Zuxiu Yu¹, Yanyan Zhang¹, Ming Zhang¹, Torben Hansen⁵, Gaston Sanchez⁶, Jeroen van den Bogaard⁷, Gwen Rosindell⁸, Shujiro Oida⁹, Marcia Almeida⁹, Emmanuelle LeChatelier⁹, Pierre Renaud⁹, Nicolas Pons⁹, Jean-Michel Batto⁹, Zhaoxi Zhang¹, Hua Chen¹, Ruifu Yang¹,¹⁰, Weimou Zheng¹, Songgang Li¹, Huanming Yang¹, Jian Wang¹, S. Dusko Ehrlich⁹, Rasmus Nielsen⁶, Oluf Pedersen⁵,¹¹,¹², Karsten Kristiansen¹,¹³ & Jun Wang¹,⁵,¹³

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.
Richness of human gut microbiome correlates with metabolic markers.

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.
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.
An integrated catalog of reference genes in the human gut microbiome

Junhua Li1,3,19, Huijie Jia1,19, Xianghang Cai1,19, Huanzi Zhong1,19, Qiang Feng1,4,19, Shinichi Sunagawa5, Manimozhiyan Arumugam1,5,6, Jens Røat Kultima5, Edi Prifti7, Trine Nielsen6, Agnieszka Sierakowska Juncker8, Chaysavanh Manichanh9, Bing Chen1, Wenwei Zhang1, Florence Levenez7, Juan Wang1, Xun Xu1, Liang Xiao1, Suisha Liang1, Dongya Zhang1, Zhaoxi Zhang1, Weineng Chen1, Hailong Zhao1, Jumana Yousuf Al-Aama10,11, Sherif Edris11,12, Huanming Yang1,11,13, Jian Wang1,13, Torben Hansen6, Henrik Bjørn Nielsen8, Søren Brunak8, Karsten Kristiansen4, Francisco Guarner9, Oluf Pedersen6, Joel Dore14, S Dusko Ehrlich7,15, MetaHIT Consortium16, Peer Bork5,17 & Jun Wang1,4,6,11,18

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.
Transforming Fat to Thin

How much does the microbiota influence the host’s phenotype? Ridaura et al. (p. 1079; see the Perspective by Walker and Parkhill) obtained uncultured fecal microbiota from twin pairs discordant for body mass and transplanted them into adult germ-free mice. It was discovered that adiposity is transmissible from human to mouse and that it was associated with changes in serum levels of branched-chain amino acids. Moreover, obese-phenotype mice were invaded by members of the Bacteroidales from the lean mice, but, happily, the lean animals resisted invasion by the obese microbiota.

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

Once considered radical, fecal microbiota transplants have produced promising clinical results in patients suffering from infection with the nasty stomach bug Clostridium difficile. In one 2013 study, 81% of patients recovered from their diarrhea after just one duodenal infusion of donor fecal flora.

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.
“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.
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*
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

Single-Cell Exome Sequencing and Monoclonal Evolution of a JAK2-Negative Myeloproliferative Neoplasm
Single Cell Sequencing
Current challenges in the bioinformatics of single cell genomics

Luwen Ning¹, Geng Liu², Guibo Li², Yong Hou², Yin Tong¹ and Jiankui He¹ *

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

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*
An early Middle Pleistocene horse (560-780 kyr BP)

微量DNA基因组分析
(Ancient DNA Analysis)

BGI-Shenzhen enhances its reputation as world’s largest sequencing center, deciphering an anti-Eskimo, the human methylome, and a gene catalog of the human gut microbiome.
Early diagnosis/typing of cancers/diseases by sequencing trace DNA fragments (cfDNA) in body fluids (blood, urine, etc)
NIPT (Non-Invasive Prenatal Testing)

Discovery of Fetal DNA in Mother’s Blood

Large Scale High Throughout Sequencing
Non-Invasive Prenatal Testing by Whole Genome/Exome Sequencing

Collaborations

Non-invasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing

Prenatal detection of aneuploidy and imbalance chromosomal arrangements by massively parallel sequencing

Abstract

Fetal chromosomal abnormalities and several molecular genotyping methods are highly relevant to the human genome and resolution allows detection as the tool for prenatal diagnosis of imbalanced chromosomal system, with average abnormalities by z-test, microduplication (ranged algorithm). Our work demonstrates chromosomal arrangements.
Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal plasma DNA sequencing

Tze Kin Lau*, Fang Chen*, Xiaoyu Pan†, Ritsuko K. Pooh*, Fuman Jiang*, Yihan Li‡, Hui Jiang‡, Xuchao Li‡,
Shengpei Chen§ & Xiuxing Zhang‡

*Department of Obstetrics and Gynecology, The Chinese University of Hong Kong, Hong Kong, †Guangdong Province Key Laboratory of Genome, BGI-Suzhou, Suzhou, China, ‡School of Laboratory Medicine and Healthcare, Guangzhou University of Traditional Chinese Medicine, Guangzhou, China, §CRIM Clinical Research Institute.

Objective: To develop a new bioinformatic method for noninvasive prenatal identification of common fetal aneuploidies using massively parallel sequencing of maternal DNA. This method was applied to test plasma DNA samples from 108 pregnant women. The detection rate for Trisomy 21 was 100% (14/14) in 14 case-control studies. The detection rates for Trisomy 18, 9% (1/11) in 11 case-control studies, and Trisomy 13, 0% (0/11) in 11 case-control studies. The false positive rate was 0% in all studies. Conclusion: This massive parallel sequencing approach, combined with the improved z-score technology, enables the prenatal diagnosis of most common fetal aneuploidies with a high degree of accuracy even in the first trimester of pregnancy.

~ 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
Let's talk about The Future Full of Opportunities
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 artemisinin
Genes involved in the metabolic pathway to synthesize artemisinin.
Industrial Production of Artemisinin

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

Artemisinin
“DNA is the software of the cell, and when we change the software we change the species.”

J. C. Venter
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 collaboration in 8-10 years

in 3 – 5 years
Reinventing Chromosome 3

- **Natural**
  - 2 telomeres
  - 10 rRNA genes
  - 21 retrotransposons
  - 316,667 bases

- **Sc2.0**
  - 2 synthetic telomeres
  - 98 loxP sites for CRISPRi and CRISPRa
  - Changed G to A in 43 stop codons
  - 272,871 bases
  - PCR tags—each synthetic chromosome has multiple sequences that distinguish it from natural counterpart

---

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
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

Chromosome II
~770kb

MegaChunk A-Y
~30kb

Chunks
~10kb

2kb MiniChunks
~10kb

Strategy of Assembly

- 40bp overlap between each MiniChunk fragment
- Maximum 7 fragments
- All chunk assembly been done for ChrII
Sequencing of *semi*Syn ChrII

<table>
<thead>
<tr>
<th>Colony ID</th>
<th>Clean read base(M)</th>
<th># of Reads</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>synII A-I _1</td>
<td>251.7</td>
<td>1,327,610</td>
<td></td>
</tr>
<tr>
<td>synIIA-I _2</td>
<td>280.2</td>
<td>1,520,249</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variation type</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deletion</td>
<td>3</td>
</tr>
<tr>
<td>Insertion</td>
<td>1</td>
</tr>
<tr>
<td>SNV</td>
<td>22</td>
</tr>
</tbody>
</table>

*Indel*
Growth Curves after each “Replacements”
**ChrVII** (1,015kb)

### Design summary

<table>
<thead>
<tr>
<th>Design Operation</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add PCR tags</td>
<td>706 pairs</td>
</tr>
<tr>
<td>TAG -&gt; TAA</td>
<td>128/129</td>
</tr>
<tr>
<td>Insert loxpsym sites</td>
<td>316</td>
</tr>
<tr>
<td>Delete tRNAs</td>
<td>36</td>
</tr>
<tr>
<td>Delete LTR/transposons</td>
<td>44 LTR; 16 transposon relevant</td>
</tr>
<tr>
<td>Delete Introns</td>
<td>14/26</td>
</tr>
<tr>
<td>Replace telomeres</td>
<td>2</td>
</tr>
</tbody>
</table>
Timeline of Sc2.0 in BGI

2013
- SynChrII initiated
- SynChrII finished

2014
- Finish SynChrII design
- SynChrII finished

2015
- Finish SynChrVII design
- Finish SynChrX design
- Finish SynChrXIII design

2016
- Finish SynChrX
- Finish SynChrXIII
Design Yeas
What will be the next milestone in SynGenome?

A multicellular eukaryotic genome!?
“Small” Genomes Sequenced

- 1982 Bacteriophage ФХ174 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”

Gene modifications
Man-made cells
Cell cultures from donor
SCs & iPS

CRISPR
Gene manipulations

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

（英）P. 安德森主编
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!
A long march towards the brilliant future requires generations’ effort!

Mutual trust & personal friendship are more important than ever!
China's Sequencing Powerhouse Comes of Age

by Dennis Normile

Three Generations Leaders of BGI

“三代同堂”
The 4th Generation is already there!
Education is for the day after tomorrow!

“Gene should also be taught to the kids!”

I’m coming!

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.”
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.”

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.
Confident of you all younger generation!
“When you drink from the well, don’t forget who helped dig it!”

“饮水思源”  helped dig it!”

Gratitude speech at the reception by Scientific American
Salute all collaborators & supporters
all teachers & friends
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