Omics for ALL

2025.01.31

BGI China & BKGI Thailand

Wang Jian

华大BGI

Jensen Huang NVIDIA

Digital Biology will be the next revolution

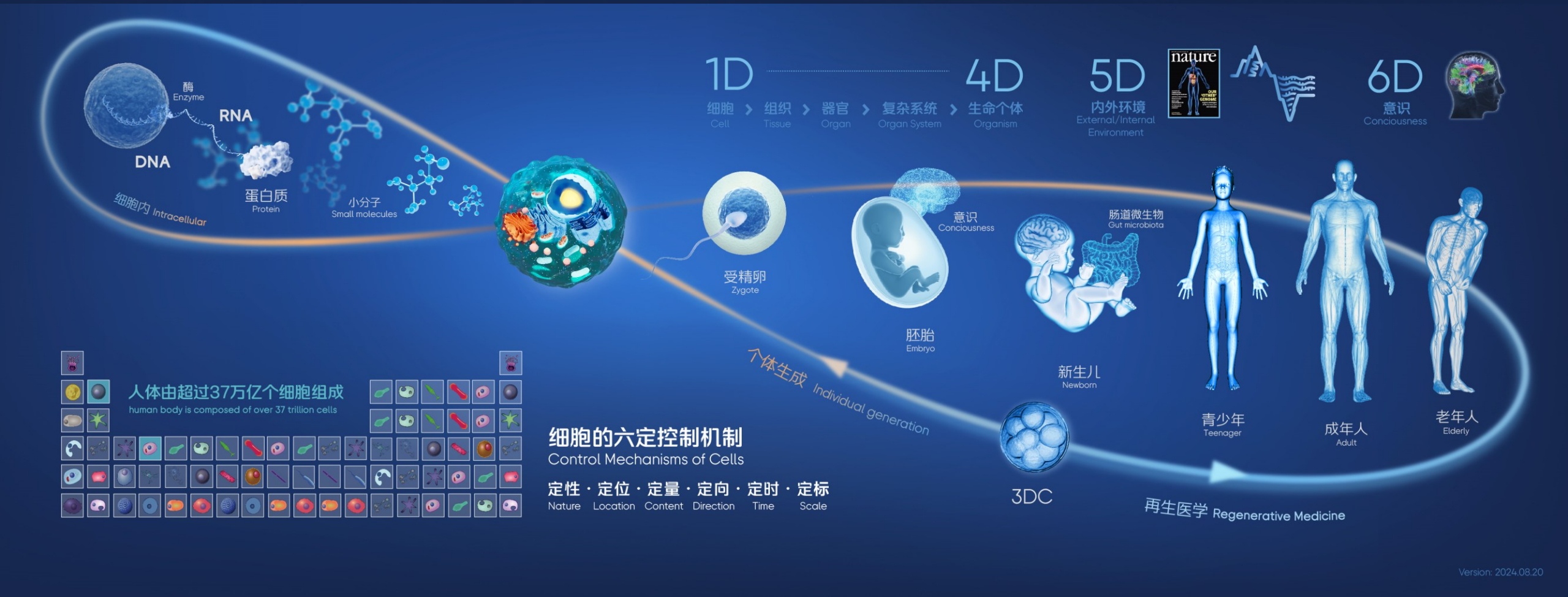
21st Century, Data is the most valuable

OpenAI Ilya Sutskever

Data is the fossil fuel of A.I., we've achieved peak data and there will be no more.

From Central Dogma to Spatial Temporal Dogma

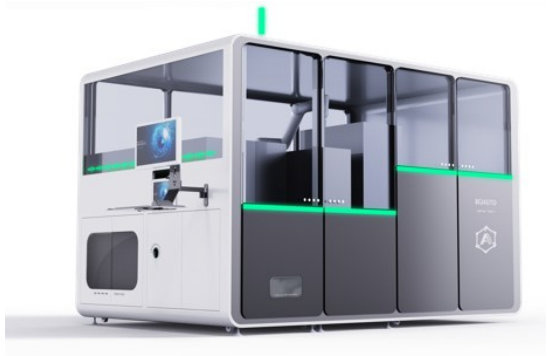
Genomics is the baseline

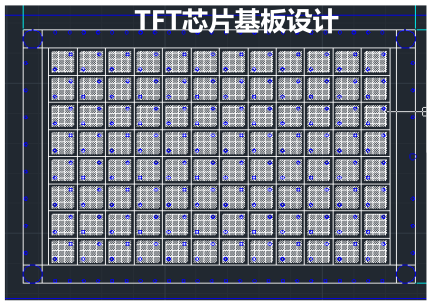


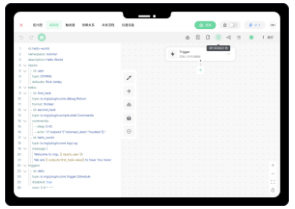
Version: 2024.08.20

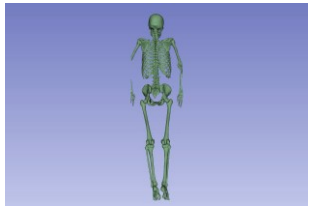
Prep ALL + Seq ALL + MS ALL + Image ALL + AI ALL=All Omics

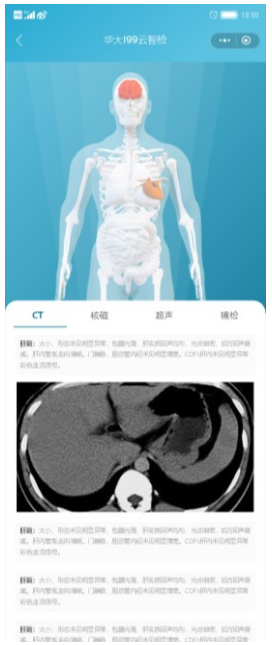
All Omics Equipment & Reagent made in BGI:







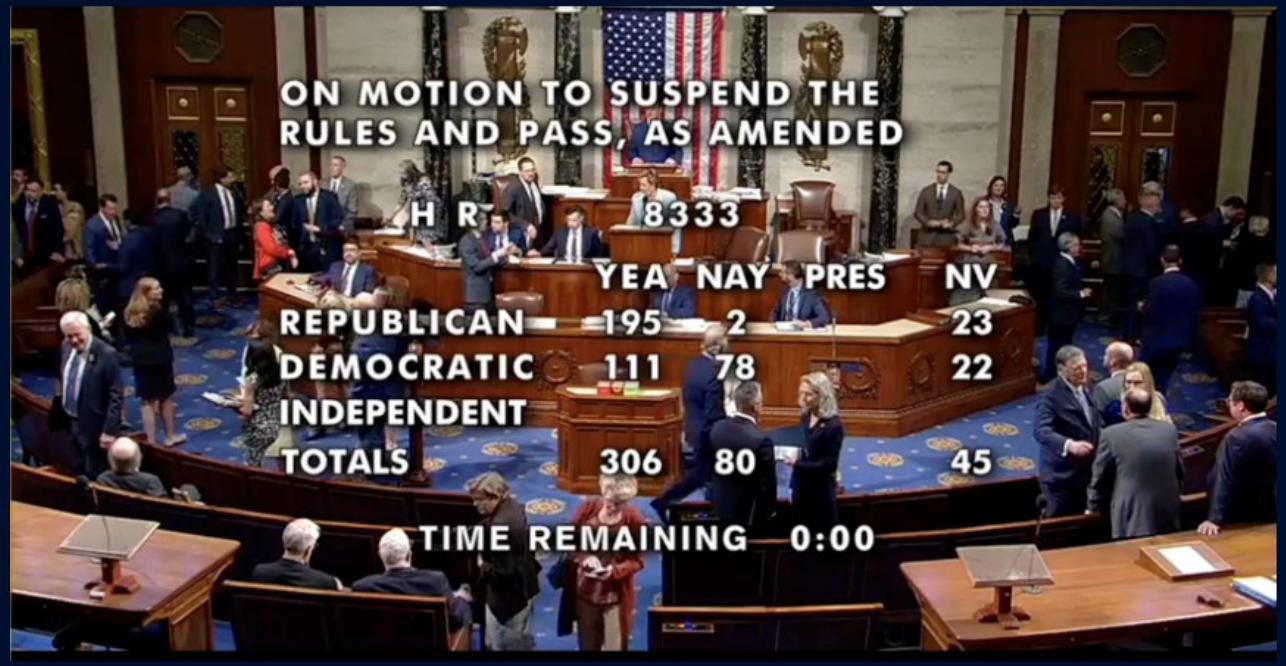




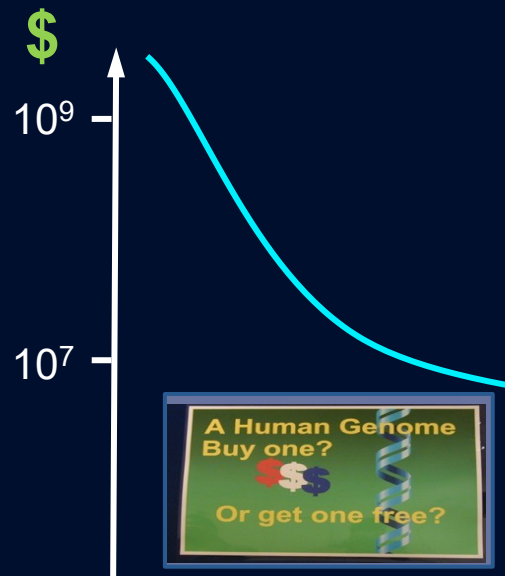
Big Data/ Multi-modality/ Whole Life Cycle/ Health management and service

Omics for AI, AI for Omics

Ban from Dipseq to Deepseek



From the 1% HGP to the al
From papers to textbooks



Cell Research
www.cell-research.com

EDITORIAL OPEN

Check for updates

The 1% gift to humanity: The Human Genome Project II

Cell Research (2024) 0:1–4; https://doi.org/10.1038/s41422-024-01026-y

Upon announcing the completion of the Human Genome Project (HGP) in 2003, scientific leaders envisaged the transformative promise of the human genome for health and society.[1] They anticipated genomics breakthroughs that could touch all of humanity and empower people to have longer healthier lives, ultimately creating a more prosperous society that also wields the human genome responsibly. Two decades on, human genomics now underpins the prospects of precision medicine and precision public health (Box 1). However, as further impactful advances increasingly depend on world-wide representation and collaboration, global-scale challenges that shackle progress must now be addressed. These are mainly economic, organizational, infrastructural, scientific, and Ethical, Legal and Social Implications (ELSI) challenges. Existing efforts lack coordination at the scale and scope required to overcome these challenges. In a time of heightened geopolitical uncertainty and legitimate citizen concern over the conscientious use of data and data-analysis methodologies, which include genomics big data and artificial intelligence (AI), we must draw inspiration from precedents where bold goals organized humanity to responsibly employ the best technologies and collaborative science toward their solution.[2] In an era where intense corporate competition and geopolitical agendas threaten global collaboration, we must galvanize efforts to make precision medicine a universal gift — available and accessible to all of humanity for the generations to come.

Box 1. Definitions of important terms.

Precision medicine is an approach to healthcare that takes into account inter-individual variability of genomic, environmental, and lifestyle characteristics. It aims to tailor medical treatment and prevention strategies to the specific characteristics of an individual or a group of individuals. This approach contrasts with the traditional one-size-fits-all model, where treatments and prevention strategies are developed based on population averages.

Precision public health applies the principles of precision medicine to the field of public health. It involves using advanced technologies, data analytics, and interventions to prevent disease, promote health, and improve healthcare delivery at the population level. Precision public health aims to identify and target interventions to specific vulnerable/disadvantaged groups and/or individuals who are at increased risk of disease or poor health outcomes, thereby maximizing the effectiveness and efficiency of public health efforts.

Health span refers to the period of a person's life during which they are generally healthy and free from serious illness or chronic diseases. Unlike lifespan, which simply refers to the total length of an individual's life, health span focuses specifically on the duration of time that a person enjoys good health and well-being. Maintaining and extending health span is a primary goal of public health efforts, medical interventions, and lifestyle choices aimed at promoting healthy aging and reducing the burden of chronic diseases and disability in older populations. Improving health span allows individuals to live longer, healthier, and more fulfilling lives.

The 20th century saw several remarkable research and development achievements brought about by a new way of doing big science.[3] These pioneering megaprojects effectively synchronized efforts across government, industry, science, and engineering, rapidly achieving world-changing breakthroughs that brought humanity into the nuclear age and the space age. In the final quarter of the 20th century, such an integrated approach was adopted to launch the HGP. By the 1980s there was growing consensus that a global view of the human genome was necessary to unravel genetic links to disease, but the scale of the HGP was beyond the technological capabilities of the time.[2] It thus took the first internationalized megaproject led by the International Human Genome Sequencing Consortium (IHGSC), mobilizing 20 centers across 6 countries, to invent and subsequently use the necessary technologies to complete the first draft of the human genome in 2003. Government, industry, science, and engineering collaborated effectively to usher in the age of genomics.

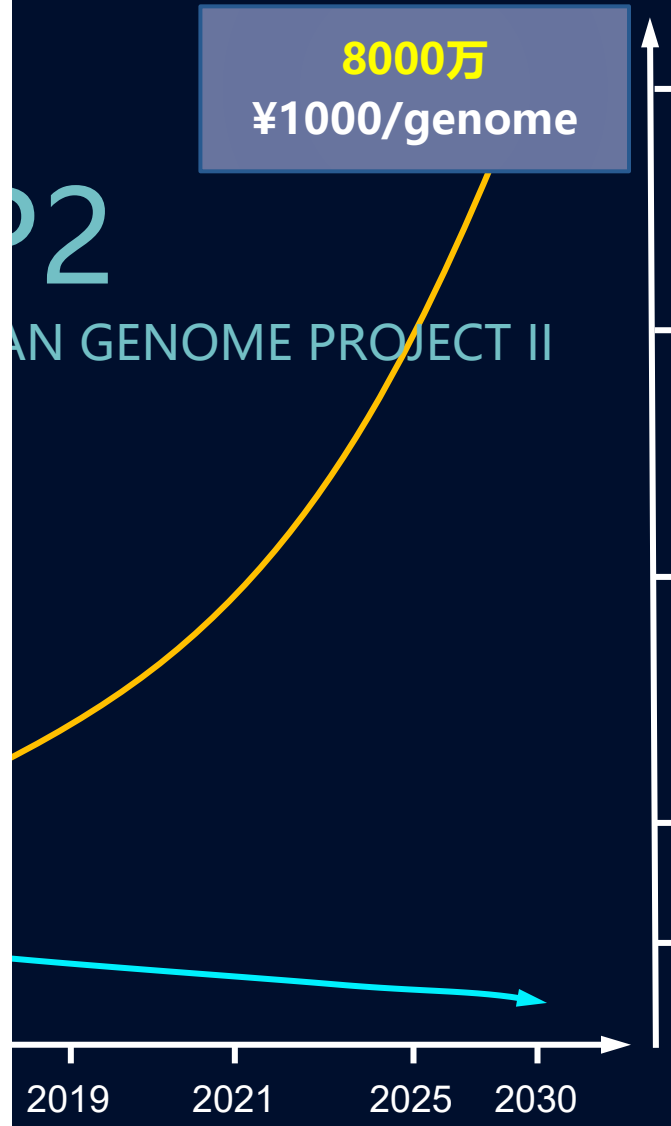
The exciting vistas the completed HGP offered quickly met a bigger dose of challenges. The first drafts of the human genome were limited by technologies of the time and were not truly complete; many sequences and features were still missing, as was the dimension of genetic diversity. Cultivating and scaling the fledgling field of genomics required not only improved sequencing technologies, but also cost falling by orders of magnitude. Thanks to grants announced by the US National Human Genome Research Institute (NHGRI) in 2004 to tackle these challenges, a burgeoning of massively parallel sequencing (MPS) technologies cut diploid genome sequencing costs faster than Moore's Law predictions.[4] With subsequent development of long-read technologies, greater dimensions of the human genome could be accurately and economically resolved. These advances empowered scientific consortia to sequence the first gap-less phased diploid genome to truly complete the HGP,[5] assemble human pangenome references to better represent the breadth of genetic diversity,[6] venture beyond single-nucleotide polymorphisms (SNPs) to catalog human genome structural variations (SVs), generate population genome baselines to discover the rarest alleles and employ multiple datasets to validate the clinical relevance of "variants of uncertain significance" (VUS).[7] Genomics and biology blossomed, owing to these advanced sequencing technologies and their dramatic cost reduction. Government, industry, science, and engineering again tackled the problem together and created success.

Thanks to these bold integrative efforts, we could begin dissecting the interplay between our genes, environment, and lifestyle. Thus rose the field of precision medicine, in turn paving a path to precision public health. If the ultimate measure of impact for the human genome is the improvement of the global population health span (Box 1), then precision medicine and precision public health would be the means. Recent multi-disciplinary developments have further facilitated precision medicine and precision public health.[8] As genome sequencing costs start falling below $100, personal genomes are fast becoming widely affordable. As multi-omics technologies improve, researchers are moving beyond genome-wide association studies to multi-ome-wide association studies. As data science

Received: 9 August 2024 Accepted: 26 August 2024

80,000,000

8000万
¥1000/genome



HGP2 RaDiAnce (Rare Disease Alliance)

The GeneT large model for genetic testing for precise diagnosis of rare diseases.

Human Genome Project 2 Rare Disease Alliance

HGP2 RaDiAnce

Twenty years ago, the Human Genome Project (HGP) promised to decode the blueprint of life, offering to solve the mysteries of human health. It was hailed as a silver bullet for every health challenge—heralding a new age of omics. Yet, two decades on, despite great scientific advances granted by the HGP, the precision health we envisioned is still out of reach for humanity, hindered by barriers of affordability, accessibility, and global collaboration.

The Human Genome Project 2 (HGP2)[1] emerges not just as a continuation of the first, but as a rallying cry to unite the scientific community and society in addressing the complexities that go beyond decoding DNA, striving to make precision public health a reality. In an era of geopolitical uncertainty, HGP2 proposes to bridge gaps in equity, technology, and collaboration to ensure that the benefits of genomics extend to all. It aims to harness the power of collaborative science by uniting fragmented databases and scaling efforts to sequence >1% of the world's population. Through this, we lay the foundation for equitable precision health, where everyone, especially those from underrepresented populations, can benefit from the promise of genomics.

At the heart of HGP2 is the Rare Disease Alliance (RaDiAnce), an initiative aimed at revolutionizing our approach to rare diseases (RDs). With over 70% of RD cases linked to genetic factors, these conditions represent an urgent and actionable priority within HGP2, offering a clear and direct path to significant public health and clinical impact using genomic technologies. RaDiAnce is dedicated to addressing the complex challenges of RDs through a comprehensive framework encompassing policy development, cutting-edge research, optimized clinical workflows, robust patient advocacy, and the establishment of advanced infrastructure and platforms.

Rare Diseases, Common Challenges

Rare diseases (RDs) affect approximately 400,000,000 people globally, making up 5% of the population[2]. The burden of these conditions falls heavily on children, who makes up 75% of RD patients, with 30% of them dying before the age of 5[3]. Those who survive often face chronic disabilities, severe health complications, and reduced quality of life. A study by Chiesi Global Rare Diseases found that RDs are approximately 10x more costly to treat per patient compared to common diseases, with a total cost to society of US$125 billion [4], this is exacerbated by indirect costs, such as lost productivity, caregiving, and long-term healthcare support. Despite the significant health and economic burden, policies supporting these patients remain inadequate and often fail to align with their specific needs. Additionally, clinical workflows for diagnosis and treatment are frequently absent, largely due to lack of comprehensive medical data. While the siloed and unstandardized nature of existing data further complicates efforts to develop common guidelines and approaches.

With about 50% of the world's new-borns—497 out of every 1,000 births—born in Asia[5], the health impact of RDs in this region is particularly pronounced. RDs often go undiagnosed or misdiagnosed in this region due to **inadequate healthcare infrastructure, limited awareness, and genetic testing capabilities**. Furthermore, RD databases are predominantly Caucasian-centric, hindering accurate diagnosis and tailored treatments here. Condition such as Thalassemia is 5-6 times more common in

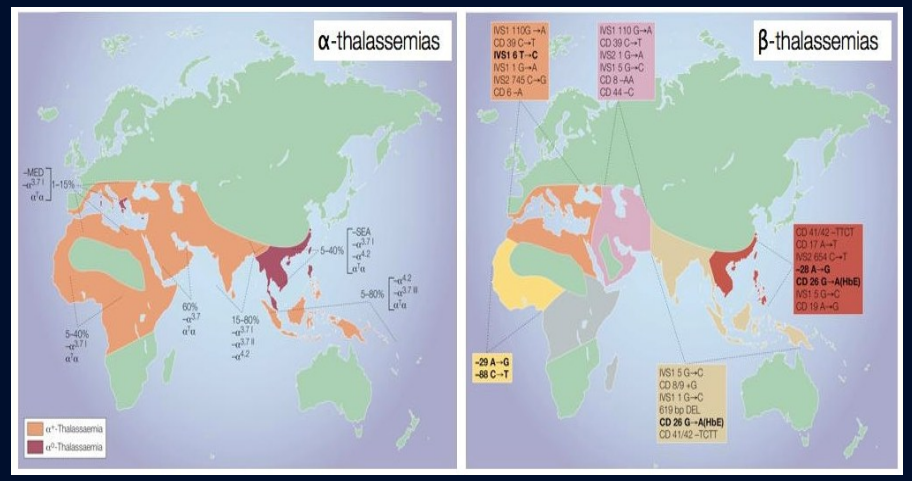
Thalassemia: Screen → diagnosis → treatment

Lower the cost, save the life!

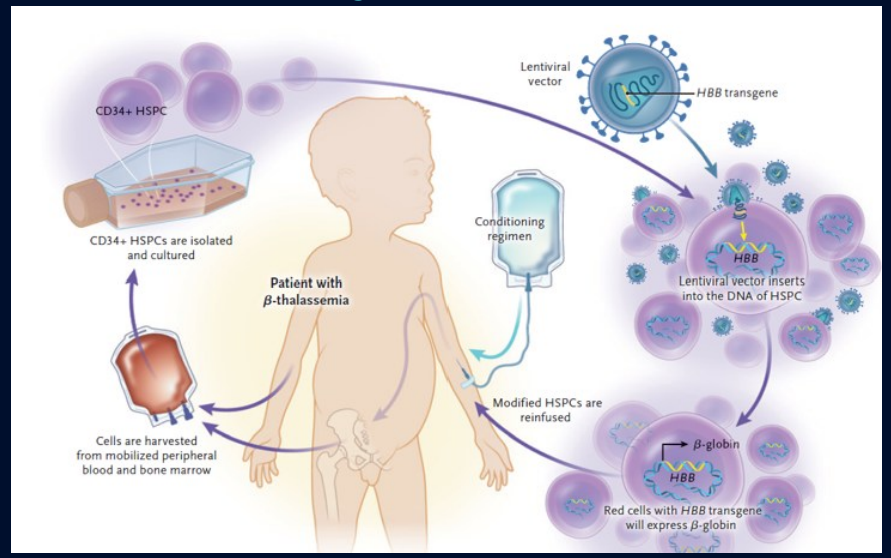
华大BGI

BGI signed a MOU with Faculty of Medicine Siriraj hospital, BKKGI, and HGI to jointly carry out Cell and Gene therapy projects on the prevention and treatment of thalassemia.

Benefit: Cost 72:1



Gene Therapy



From precision medicine to early diagnosis/ prevention/ controlling:

华大BGI

All of US vs UK Biobank:

All of Us℠ | The Precision Medicine Initiative®
THE FUTURE OF HEALTH BEGINS WITH YOU

Large-scale national cohort of at least 1 million U.S. residents initiated in 2015

Partial Data Available for Sharing (But Restricted for Countries of Concern)

biobank uk
Improving the health of future generations

National cohort of 500,000 using electronic health records linked with genetic data.

Submit Research Proposal for Review

Complete Data Use Agreement and Make Payment for Different Tiers of Use

Haerbin Lab:

CKB Dashboard:

ChinaHEART:

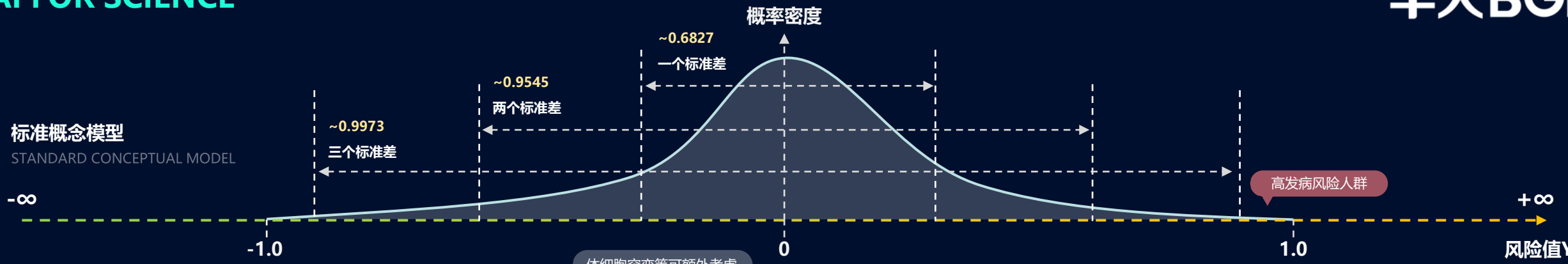
Cohort Profile: ChinaHEART (Health Evaluation And risk Reduction through nationwide Teamwork) Cohort

Runsi Wang,[1†] Yang Yang,[1†] Jiapeng Lu,[1] Jianlan Cui,[1] Wei Xu,[1] Lijuan Song,[1] Chaoqun Wu,[1] Xiaoyan Zhang,[1] Hao Dai,[1] Hui Zhong,[1] Binbin Jin,[1] Wenyan He,[1] Yan Zhang,[1] Hao Yang,[1] Yunfeng Wang,[1] Xingyi Zhang,[1] Xi Li ,[1,2,*] and Shengshou Hu[1‡]

AI FOR SCIENCE

华大BGI

标准概念模型
STANDARD CONCEPTUAL MODEL



体细胞突变等可额外考虑

$$\text{风险值 Y} = Xg W_1 + X_t W_2 + X_m W_3 + X_i W_4 + X_{me} W_5 \cdots + X_e W_n + \varepsilon$$

X 为特征值
W 为权重值

基因组数据 | 转录组数据 | 代谢组数据 | 影像组数据 | 宏基因组 | 环境因素 | 误差

$$\left| \text{风险值 Y}_{\text{真实}} - \text{Y}_{\text{预测}} \right|_{\min}$$

基础保费A * 风险值 Y * 比例系数 β

可进一步细化拆分

早期 低频变化数据 | 中期 高频变化数据 | 晚期 高频变化数据 | 优化目标 – 最小化差异 | 风险值的意外可能性 – 保险保费计算

各时期 W 权重不同

出生前 BEFORE BIRTH → 青年 YOUTH → 老年 ELDERLY → 疾病发生 或 死亡 ILLNESS OR DEATH

预测性能： 较低慢病/负责疾病准确性，AUC约0.6-0.7
配合策略： 早期干预和主动健康
应用场景： 早期妇幼保险，极前瞻性疾病保险保障

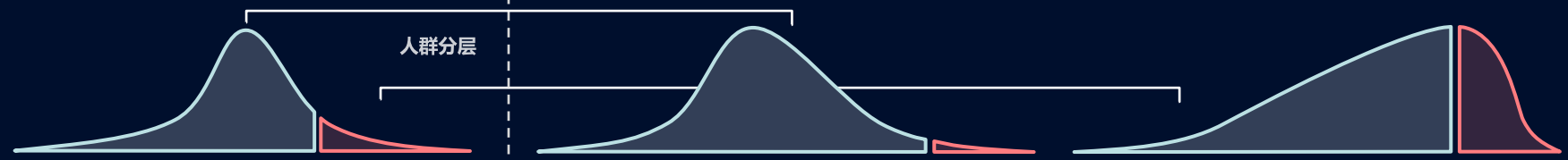
此阶段尚未出现任何病灶，有效信息仅仅为遗传背景数据，预测能力优先，但可以在人群水平进行初步分层

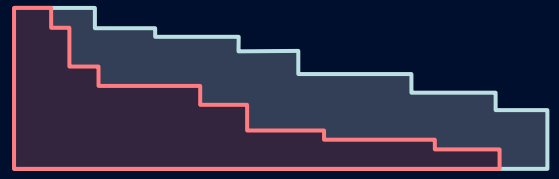
预测性能： 较高慢病/负责疾病准确性，AUC约0.8-0.95
配合策略： 风险分层，条件概率的预测修正。用于精准诊断。MOHA，BOHA...
应用场景： 疾病的预测预警，早期半强制的主动健康管理

此阶段已经慢慢出现早期疾病中下游信号特征，遗传背景的权重降低，中期，晚期特征数据开始权重升高，在早期分层情况下，人群风险分布出现偏态分布情况

预测性能： 极高，可转为预后预测
配合策略： 用药，饮食，疾病史等外部环境 + 遗传背景
应用场景： 精准治疗，写方式干预，疾病机制研究等

此阶段疾病预测可以作为查缺补漏，尽可能提早治疗时间。着重在干预治疗，用药影响等领域进行指导和研究。





我的健康我做主
My life in my hand

2002.10

Beijing Hospital

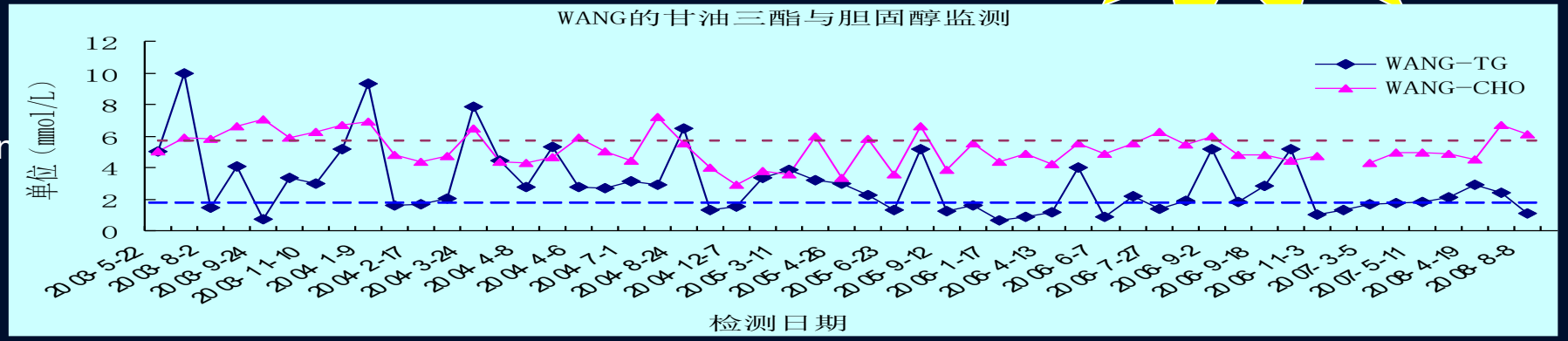
- Genes related to sugar and lipid and metabolic syndrome
  - ApoA5 pathogenic mutation: hyperlipidemia!
- Genes related to cardiovascular and sudden death
  - HRG heterozygous mutation: thrombophilia?
- Alzheimer's disease gene
  - ApoE3/E4 heterozygous mutation: Alzheimer's disease?
- Drug-sensitive gene
- Tumor-related gene
- Alcoholism and smoking ger
- Genetic disease 8:30

Hypertension

Hyperlipidemia

Coronary Heart Disease

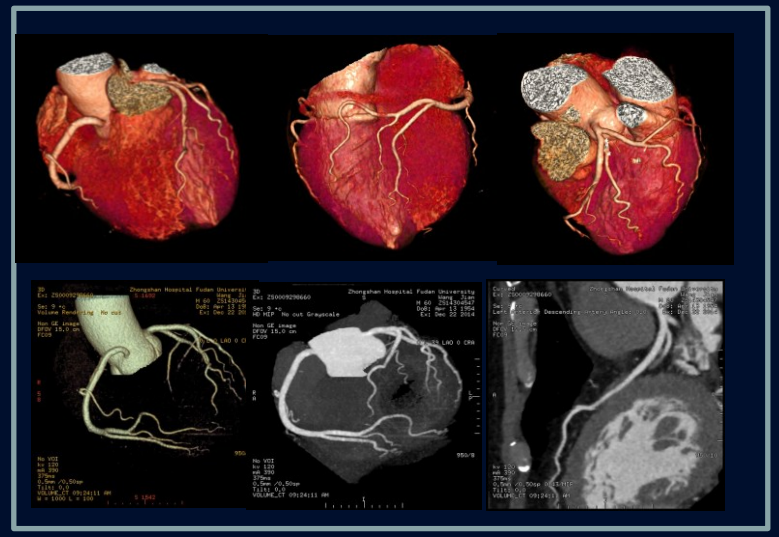
150/100 mmHg

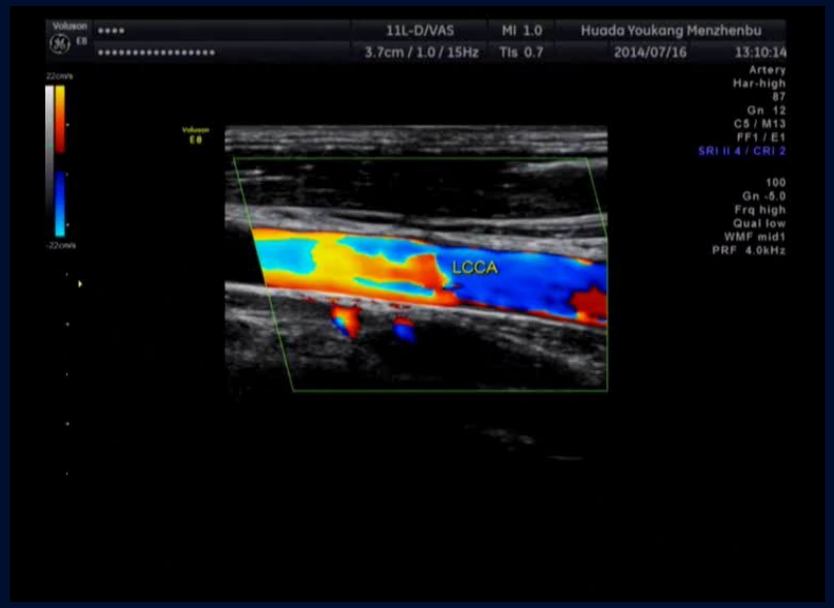


Cardio-cerebrovascular Disease

华大BGI

Plaque on blood vessel wall reduced by 54%/vascular cavity enlarged by 25%





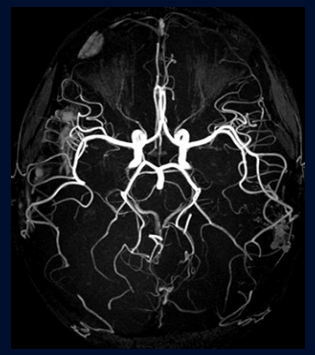
GM-CSF and M-CSF Expression Is Associated with Macrophage Proliferation in Progressing and Regressing Rabbit Atheromatous Lesions

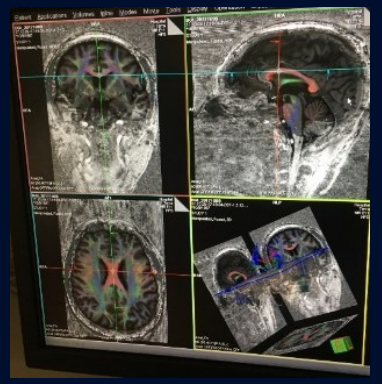
JIAN WANG,[1] SHUHUA WANG,[2] YUCAI LU,[2] YUCHUN WENG,[2] AND ALLEN M. GOWN[1]

[1]*Department of Pathology, University of Washington, Seattle, Washington 98195; and* [2]*Beijing College of Chinese Medicine, Beijing, People's Republic of China*

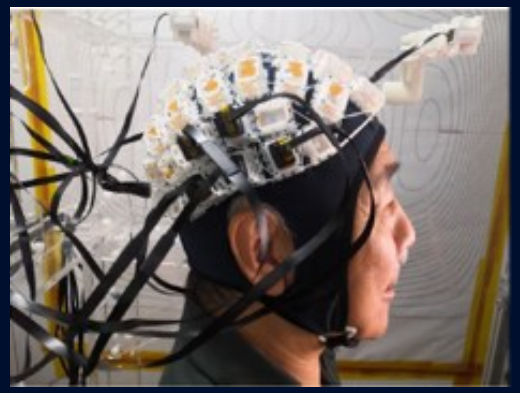
Received May 18, 1994, and in revised form August 23, 1994

Brain Health





7T 磁共振 MRI

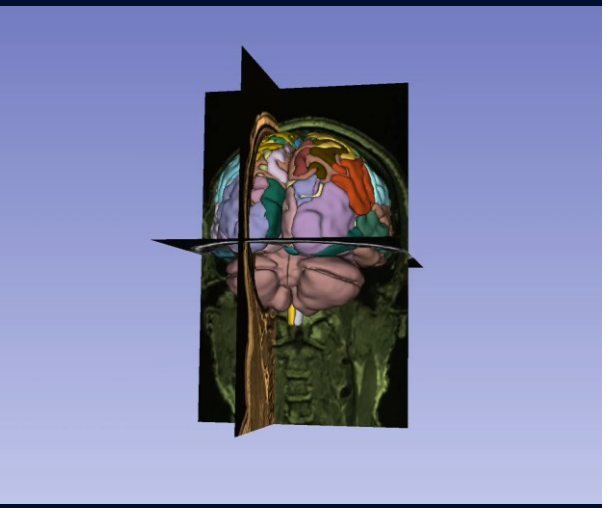


便携式脑磁图MEG-OPM

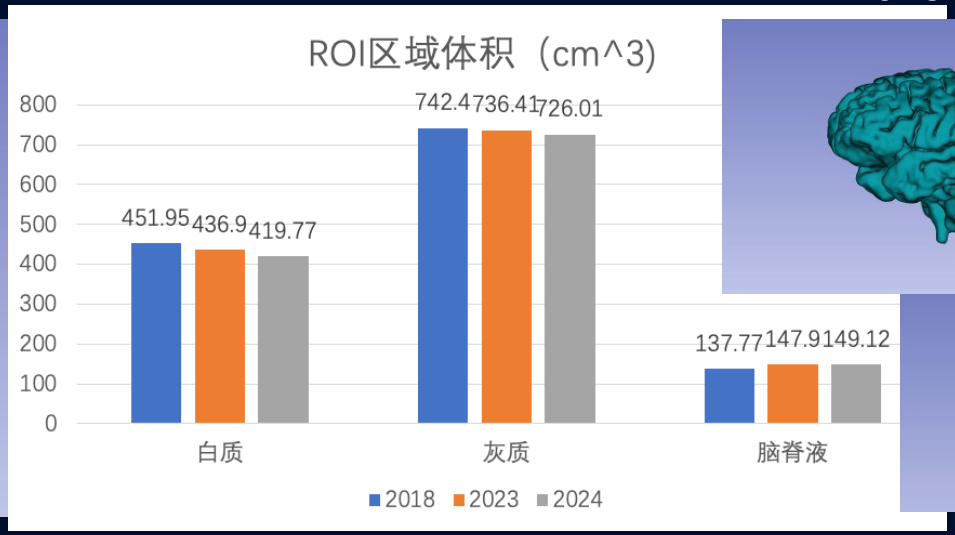
脑电图EEG

运动便携式EEG

脑部 Brain：

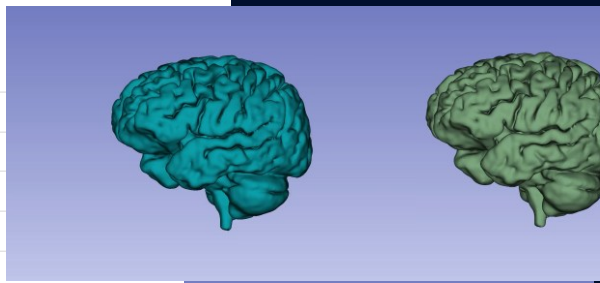


Brain Volume Changes

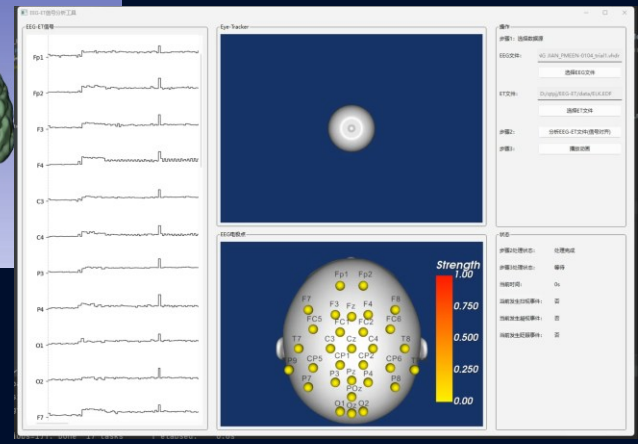


2018

2023

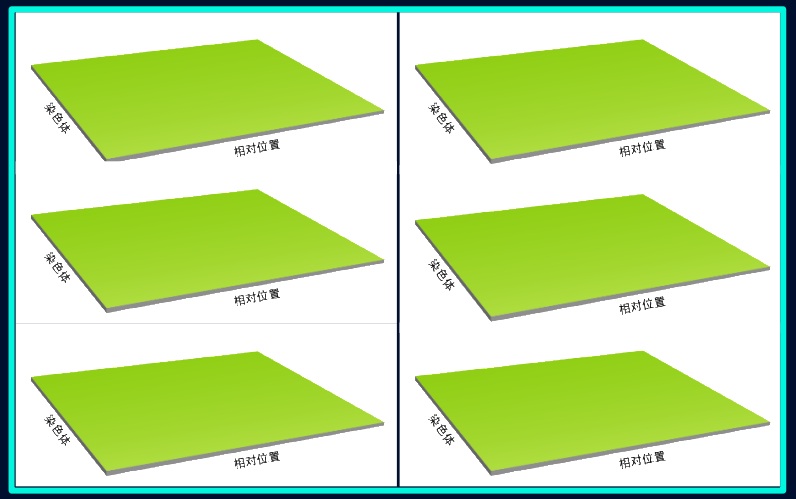


2024

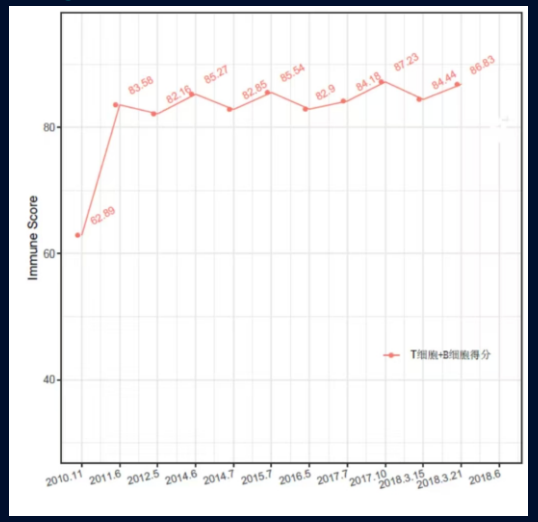


Combined EEG and eye movement analysis

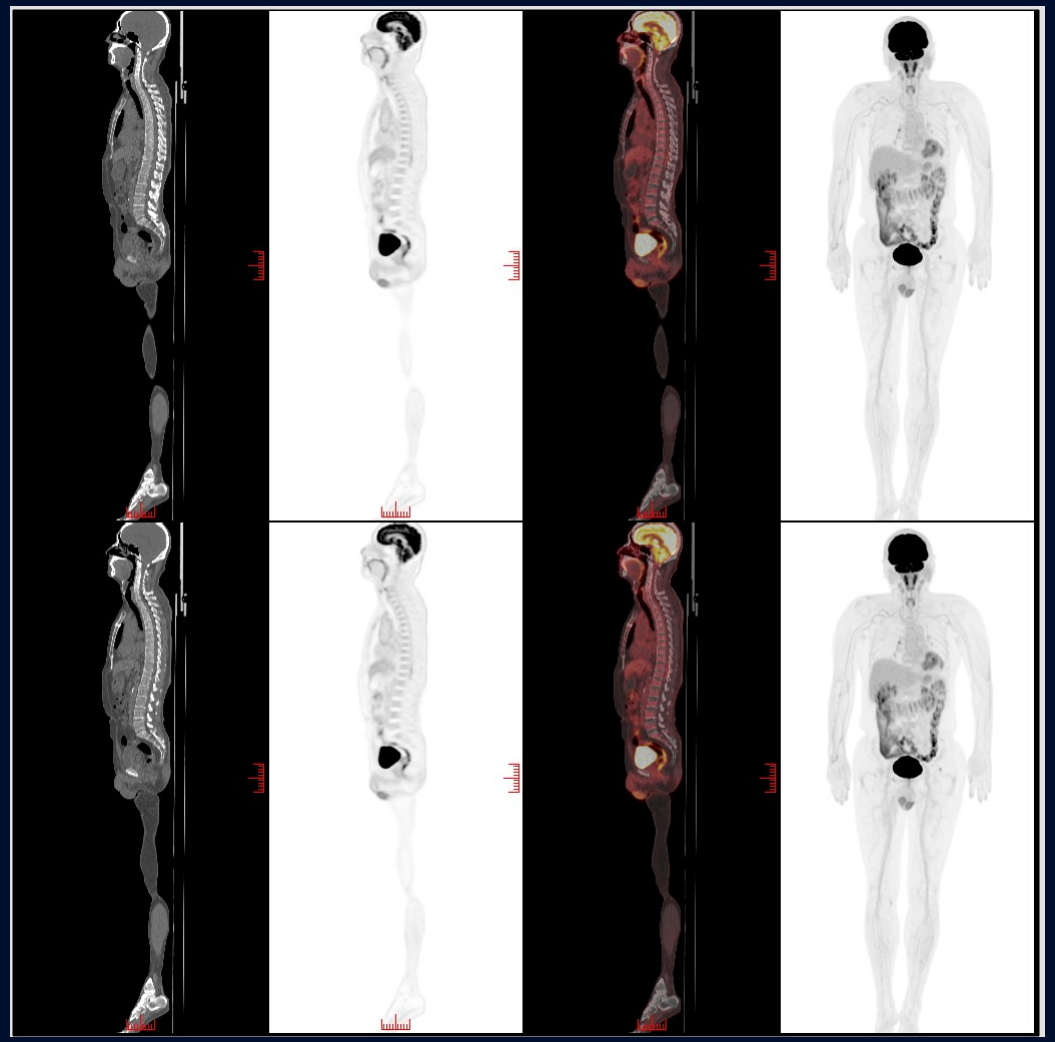
Tumor early monitoring & 5 modality

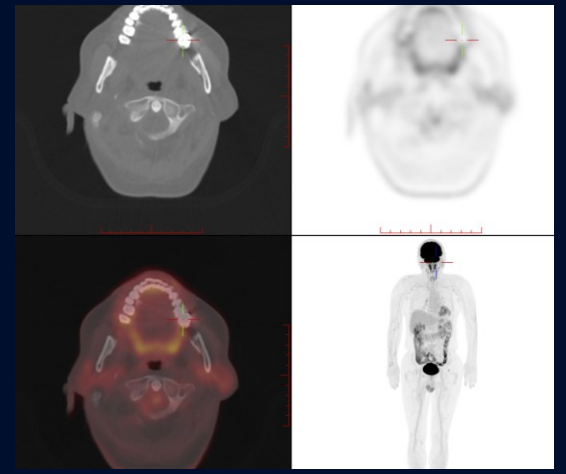


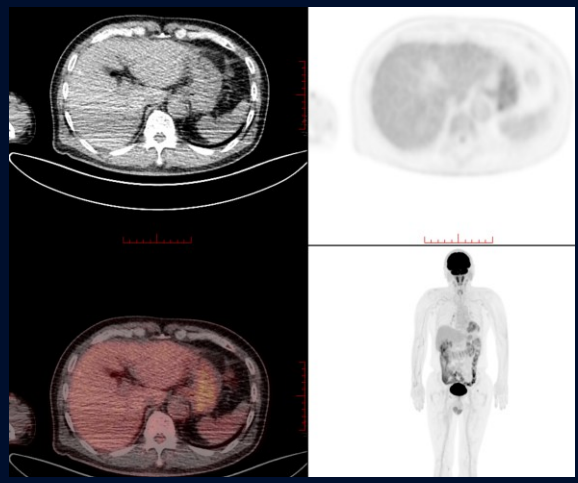
cfDNA tumor early screening results are all negative



Immune Cell







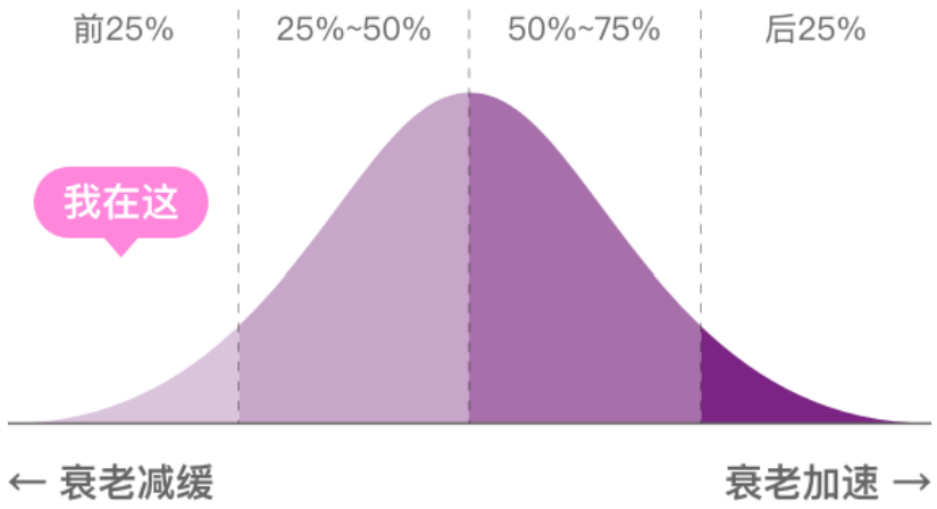
Biological age 61.7, Chronological age 70.7

您的生物学年龄为： 61.7岁，衰老速率为0.872

您的衰老速率位于同龄人中较慢的前25%，处于较慢水平

The rate of aging is in the top 25% of slower populations

衰老速率分布



检测时间	2025/1/23
实际年龄	70.7岁
生物学年龄	61.7岁

华大员工“三不”
“Three Rules” for BGIers

THREE RULES：

- No Birth Defects：0
- No Tumors Discovered Later than in Hospital: 76 (All received timely treatment)
- No Stents (Cardio-cerebrovascular disease)：1

器官生物学年龄和LIFE INDEX

华大生命科学研究院
BGI·Research

The first time in the world to construct the biological age of various organs of the human body.

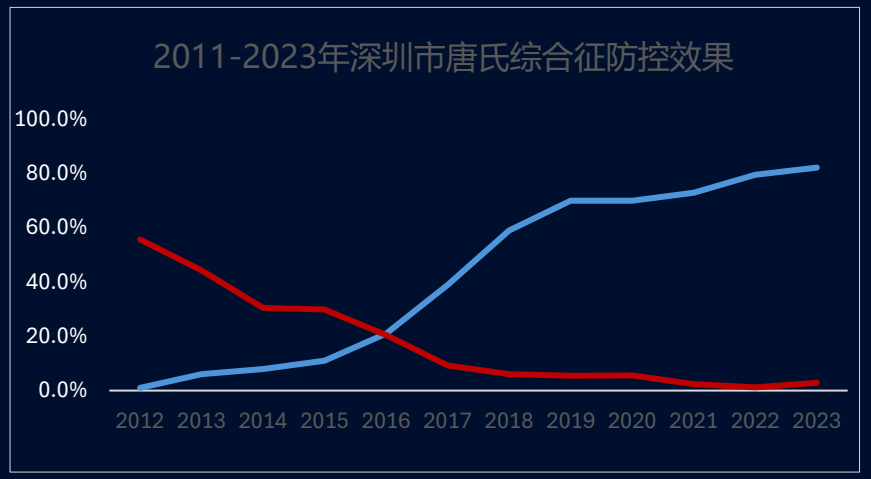
The data-driven method constructs **11 biological functional modules** (BFM) and health assessment algorithm (BFM-ash) to warn of **health status and guide precise intervention 。**

Utilize 1000+ multi-omics indicators to evaluate 9 major organs/systems of the human body
Aging quantification and health assessment model based on 2 algorithms

I can do it→ BGIers can do it → Everyone can do it

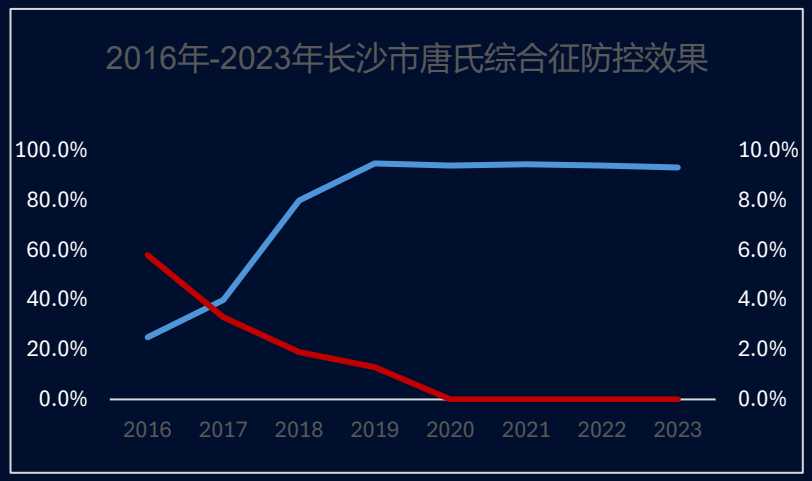
华大BGI

Shenzhen



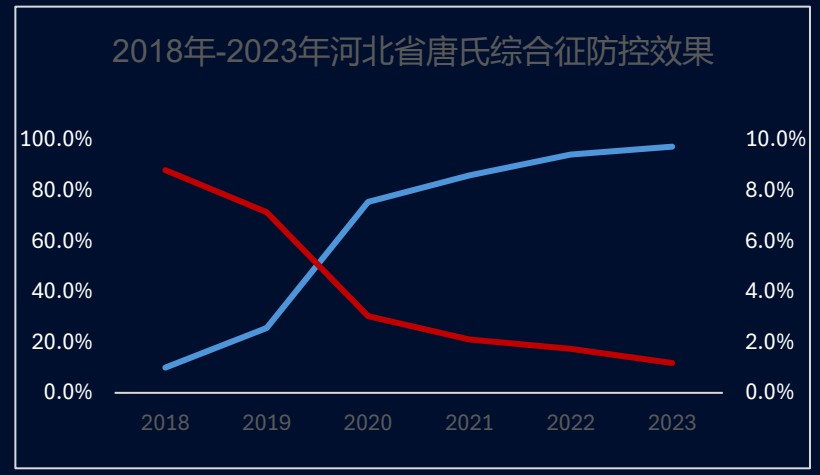
1,690,794 samples
1:10.47

Changsha



432,824 samples
1:13.55

Hebei



2,203.958 samples
1:16.73

Put the early testing and diagnosis first

- **Birth defect:** gene testing (cost-benefit ratio 1:16)
- **Cardio-cerebro vascular/stroke**: imaging + omics monitoring
- **Cancer:** early screen and detecting

Omics for all

US aim to be 1000 $/individual

BGI aim to be 1000 ¥ /individual

b 10^3b 10^6b 10^9b 10^{12}b 10^{15}b 10^{18}b 10^{21}b 10^{24}b

b Kb Mb Gb Tb Pb Eb Zb Yb

Does WHO take a lead?

Lower the cost, save the life!

Omics for All, Life-index for ALL!

Happy Chinese Spring Festival

สวัสดีครับ (swipe my card)