



The BIG Data Center: from deposition to integration to translation

Zhang Zhang On Behalf Of BIG Data Center Members

Professor in The CAS 100-Talent Program

Executive Director of BIG Data Center

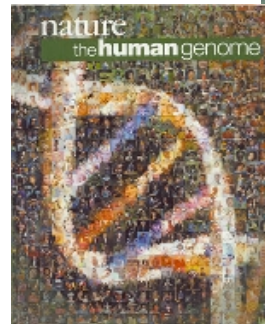
Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (CAS)



Beijing Institute of Genomics (BIG), CAS

- International Human Genome Project (China Part)
- National Genome Research Projects: Rice, Chicken, Silkworm, Common Carp, Cassava, Rubber Tree, etc.
- Sino-Saudi Genome Sequencing Project: Date Palm

Human Genome
Research Center
in 1999



2000

Officially
founded BIG
in 2003



2002



2004



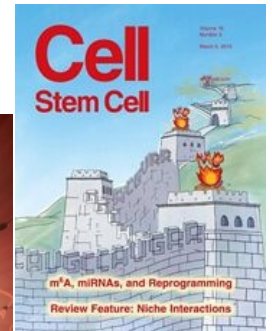
2004



2008



2013



2015

The First Chinese Population
Precision Medicine Project
Launched by BIG in Jan. 2016

Precision Medicine and Big Data

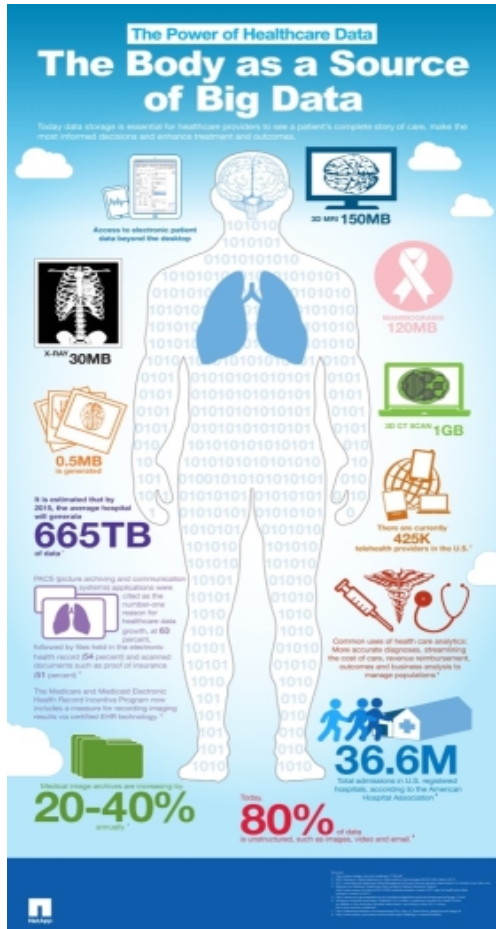
Year	Countries & Projects	Goal
2014	Lee Hood, 100K Wellness Project	Lee Hood, 100K
2014	US Human Longevity	Craig Venter, 40K-100K/yr
2015	Obama 1M Genome Project	1M
2012	Canada Personal Genome Project	100K
2011	The Genome of the Netherlands (GoNL)	250 Trio
2011	Iceland Genome project	100K
2012	Norwegian Cancer Genomics Consortium	25000 cancer patients/year
2012	UK 100K Genome Project	100K patients
2013	Japan 1000 Genome Project	1000
2013	Saudi Arabia 100K Genome Project	Rare diseases
2016	GenomeAsia 100K Genome Project	100K
2015	CAS Precision Medicine Project	CAS cohort
2016	China National Precision Medicine Projects	Cohort, diseases, 1M



Background

- **Big Data from Large-scale National Research Projects based on genome sequencing**
- **China National Big Data Strategy:** “...execute a national big data strategy and promote opening and sharing of data resources...”
 - Government funding for archiving domestic and international sequence data and establishing a national data center
 - Greater support from the scientific community
- **Big Data Submission & Transfer** across countries becomes increasingly challenging
 - Ever-growing data volume
 - Formidable bottleneck in network bandwidth
 - Lack of Chinese language service and communication for complex metadata description and input

Big Data for Everyone



Source: netapp.com

Clinical data, Imaging data, Electric health record, Phenotype data, Exposome data, Omics data

~10TB/person
 →



1M persons × 10TB/person = 10000PB
 (10M PCs for storage, 1TB per PC)