

# Geekbang>

极客邦科技

全球领先的技术人学习和交流平台

扫我，码上开启新世界



# Geekbang>

InfoQ | EGO NETWORKS | StuQ

## InfoQ

专注中高端技术人员  
的社区媒体

## EGO NETWORKS

EXTRA GEEKS' ORGANIZATION  
高端技术人员  
学习型社交网络

## StuQ

实践驱动的IT职业  
学习和服务平台

促进软件开发领域知识与创新的传播

InfoQ<sup>new</sup>

**QCon**  
全球软件开发大会

[ 上海 ] 2015年10月15-17日

**ArchSummit**  
全球架构师峰会

[ 北京 ] 2015年12月18日-19日



关注InfoQ官方微信  
及时获取ArchSummit演讲视频信息

# ArchSummit全球架构师峰会 深圳站2015

## 生命科学中的大数据



# 人人都是 超级 计算机



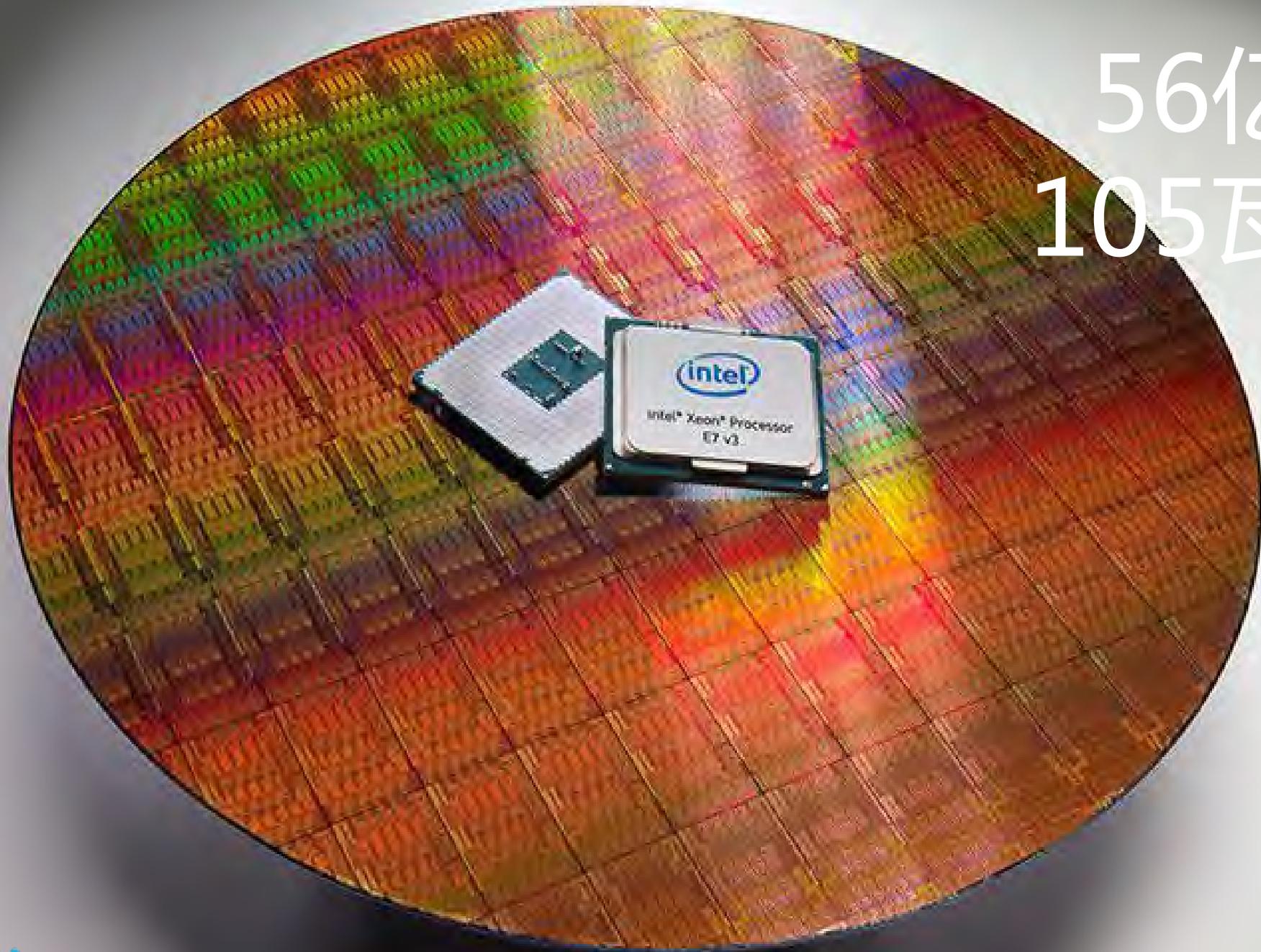
脑细胞：140亿





脑能耗：20瓦

56亿  
105瓦



# 脑存储：140PB



100 T messages/life cycle  
600 US National Libraries

# 脑计算：3.5EFlops

86G message/day



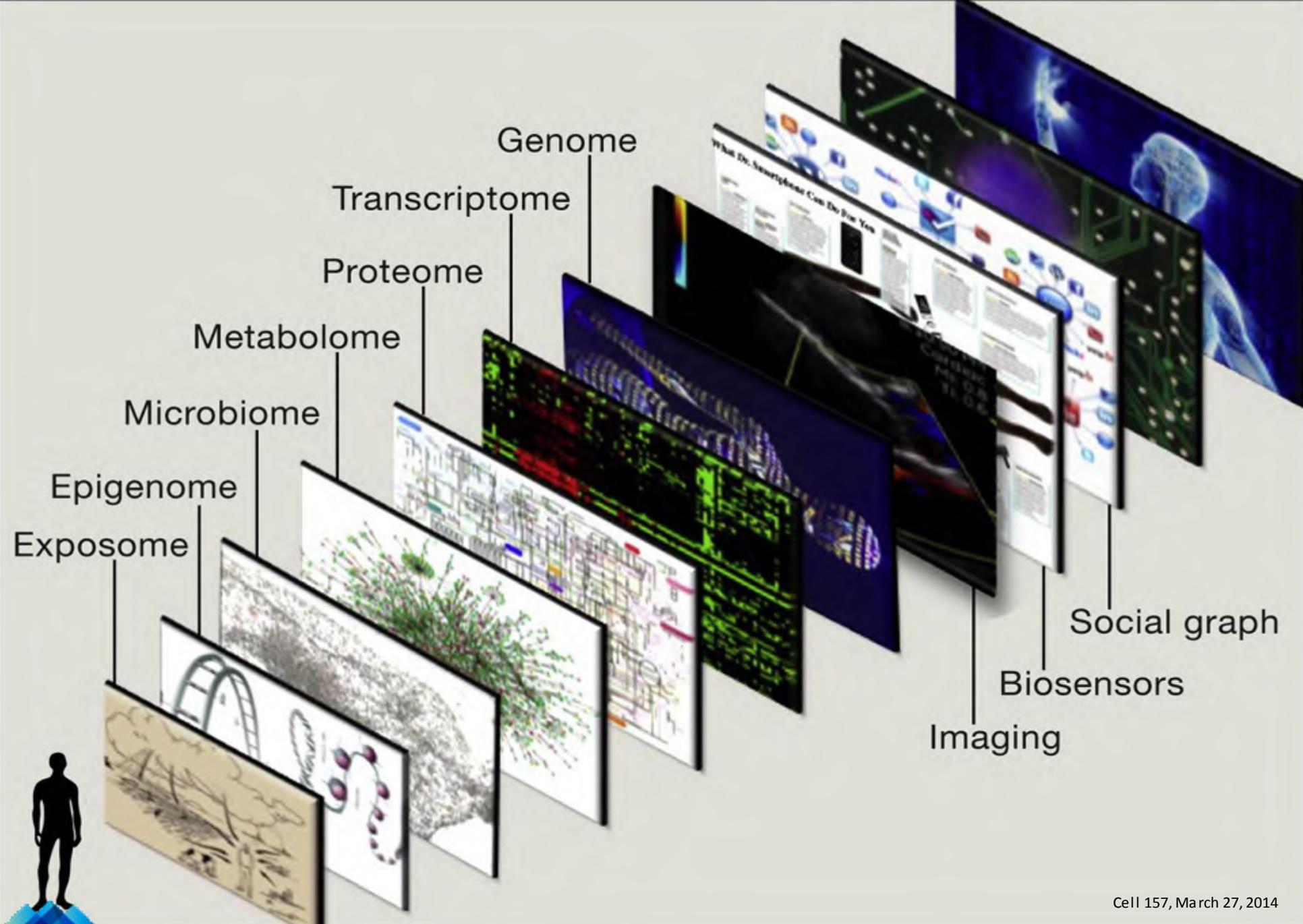
大数据：数量 / 大小



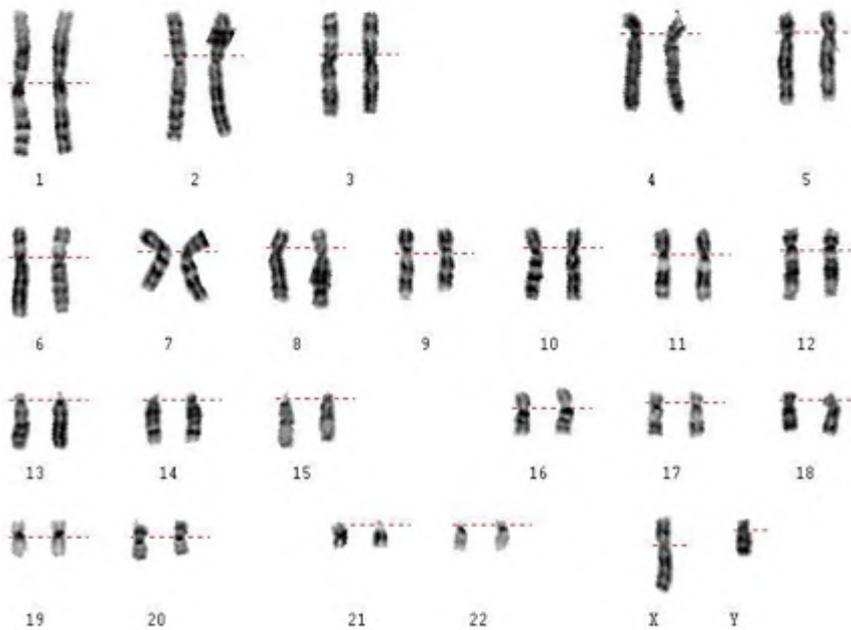
# 一个人的大数据

一个人：  
30亿个碱基（G）  
100万亿个细胞（T）  
1千万亿个细菌（P）



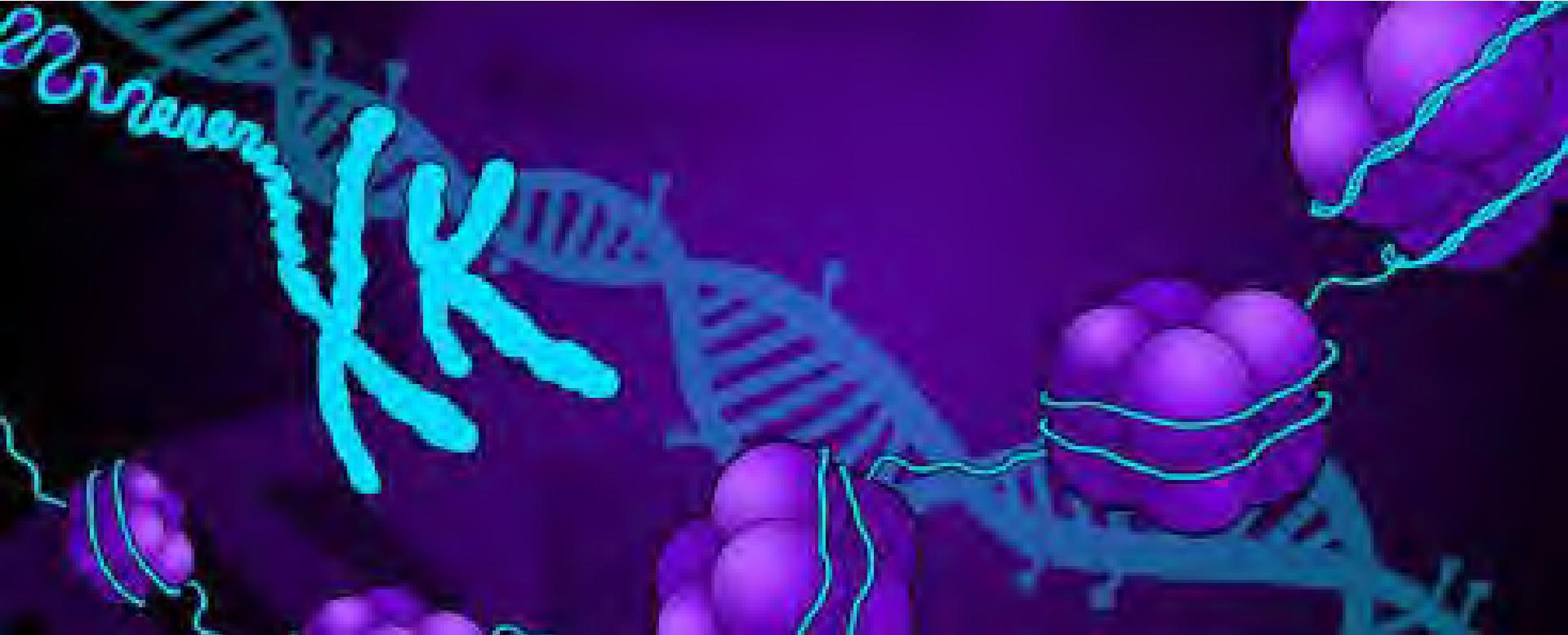


50%



50%



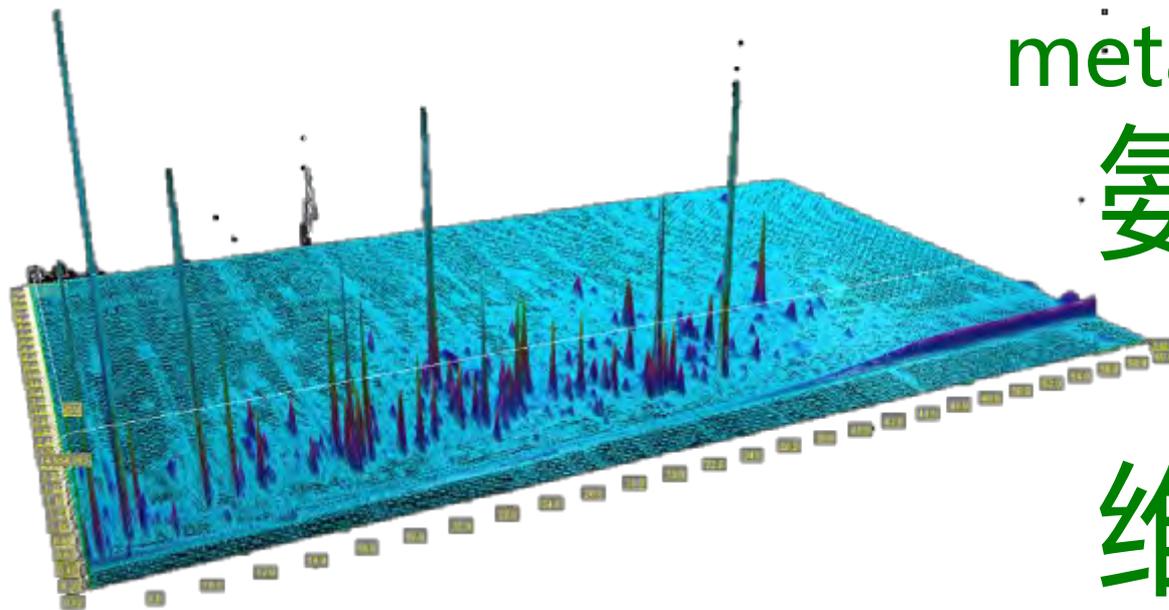


表观组  
epigenome  
28890/56%  
2TB

宏基因组  
microbiome



1PB  
2kg  
1ZB



代谢组  
metabolome

氨基酸

糖

维他命

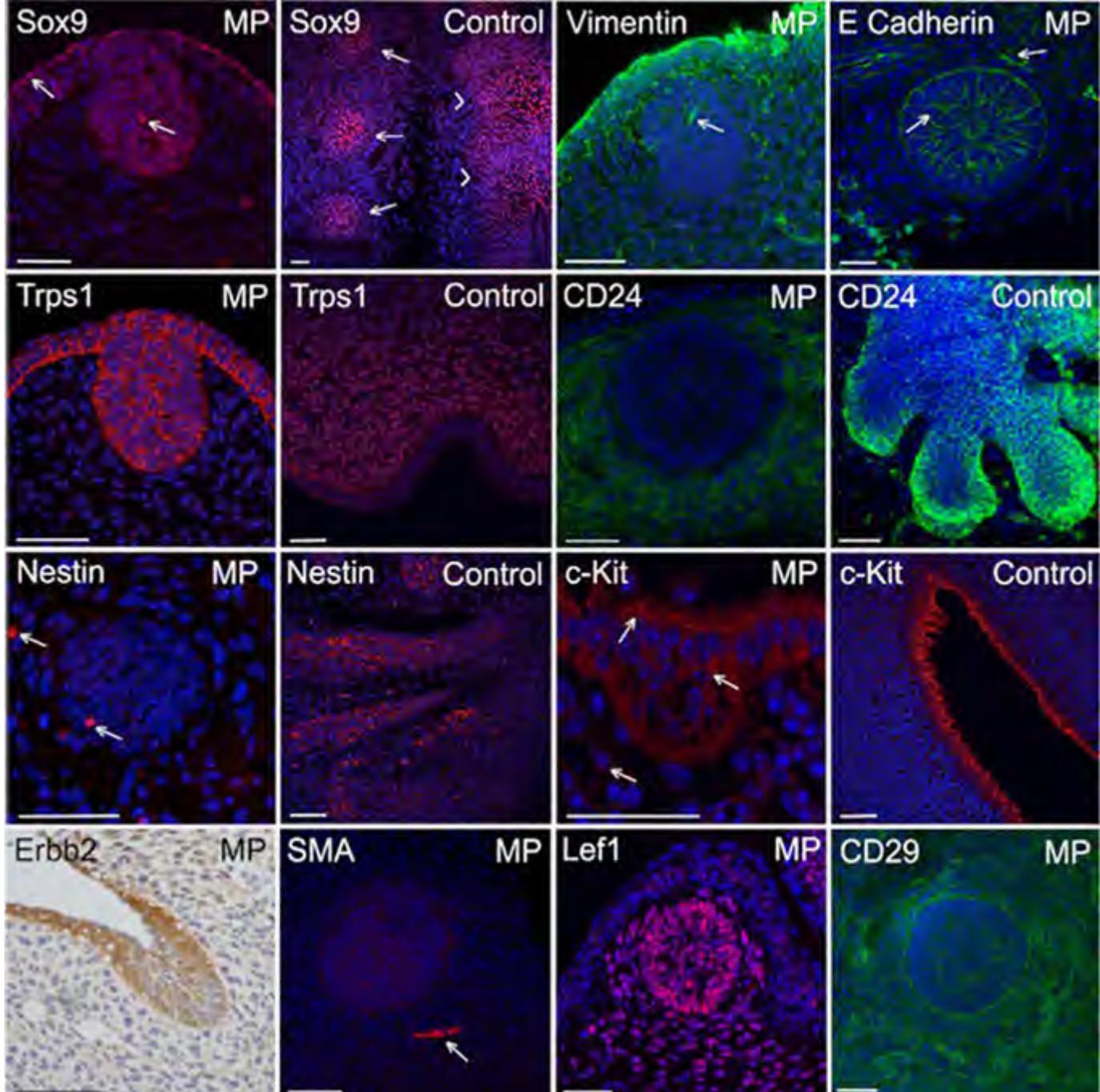
.....

基因组  
genome  
3GB  
19,000  
2%

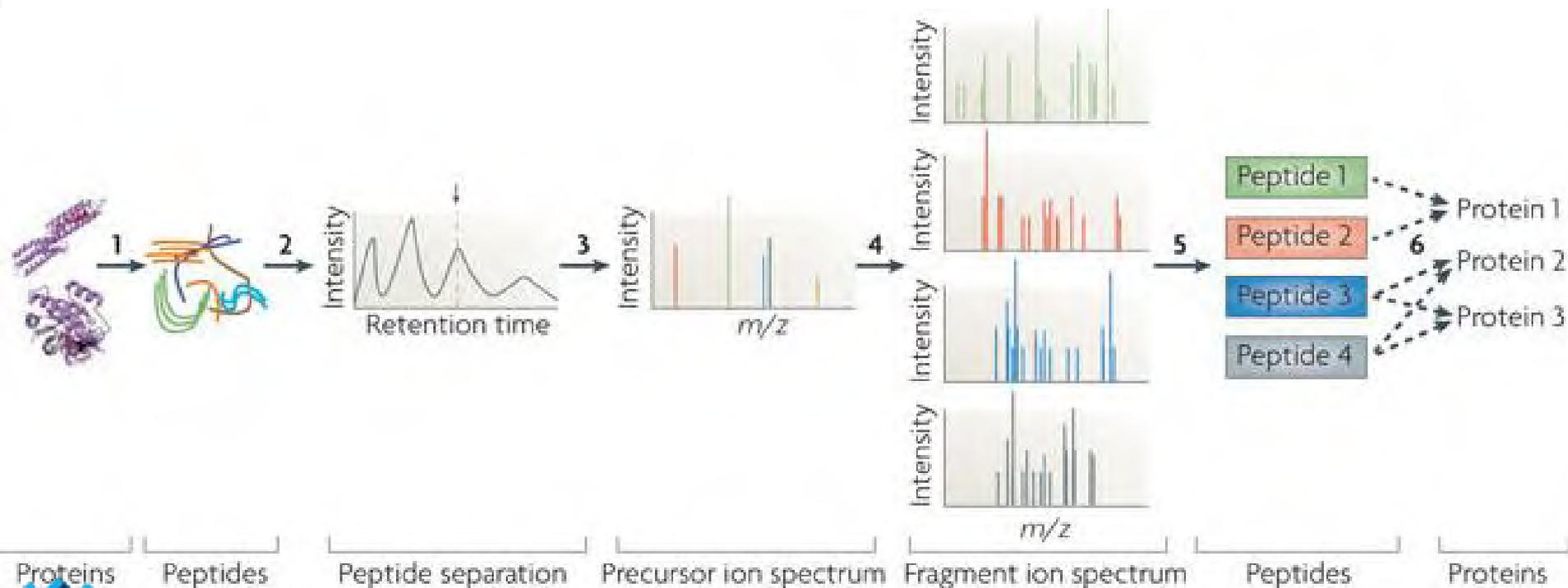


一个人的基因组：100GB  
一百万人：100TB

转录组  
transcriptome  
19,000  
80%  
60,000



# 蛋白组 proteome 0.2-2M 6788

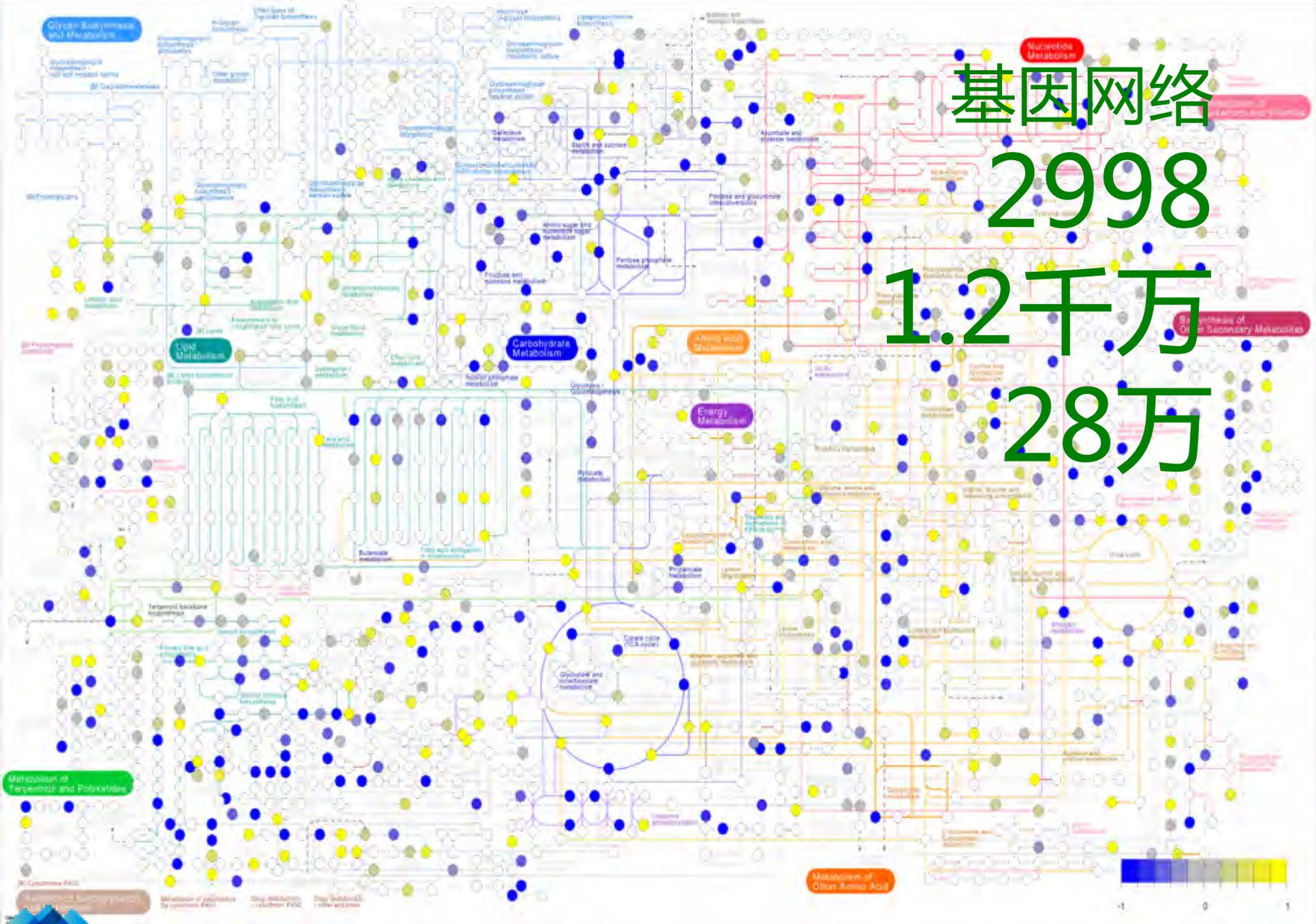


基因网络

2998

1.2千万

28万





0.5% & 6-100 SNPs



卷发  
LCE3E  
TCHH  
WNT10A

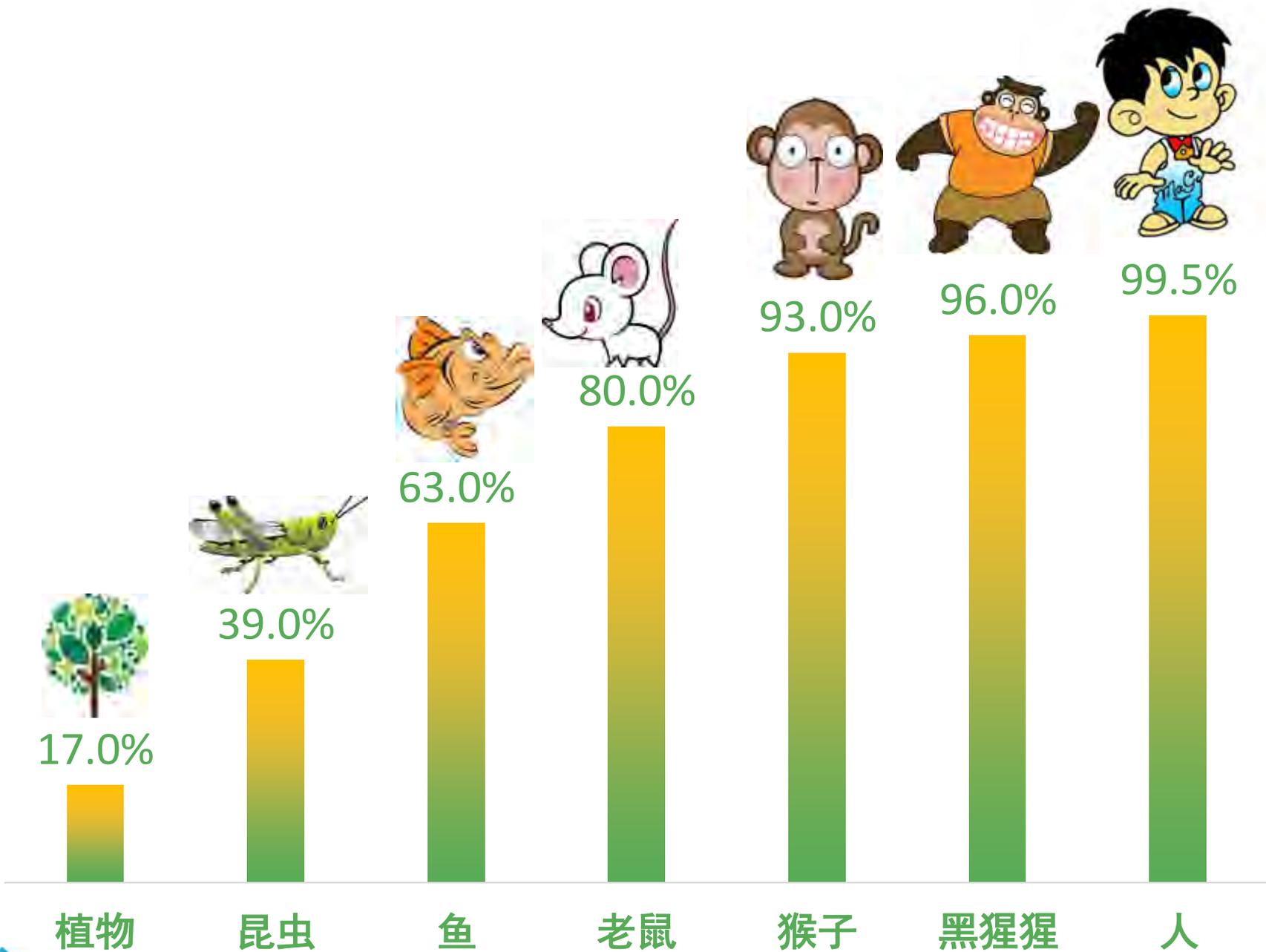
拇指

双眼皮

卷舌

秃顶  
PAX1

脸红  
ALDH2



类型	生物	学名	基因组大小 (bp)	注
病毒	猪圆环病毒I型		1,759	已知最小的基因组
病毒	猿猴病毒SV40		5,224	
病毒	噬菌体Φ-X174		5,386	最早完成测序的DNA基因组
病毒	人类免疫缺陷病毒HIV		9,749	
病毒	噬菌体λ		48,502	常作为重组DNA的克隆载体
细菌	大肠杆菌	Escherichia coli	4.6Mb	
真菌	酿酒酵母	Saccharomyces cerevisiae	12.1Mb	第一个测序的真核生物基因组，完成于1996年
线虫	咖啡短体线虫	Pratylenchus coffeae	20Mb	已知最小的动物基因组
线虫	秀丽隐杆线虫	Caenorhabditis elegans	100Mb	第一个测序的多细胞生物基因组，完成于1998年12月
昆虫	黑腹果蝇	Drosophila melanogaster	130Mb	
鱼类	金娃娃（一种河豚）	Tetraodon nigroviridis	385Mb	已知最小的脊椎动物基因组约为340Mb-385Mb
哺乳动物	小家鼠	Mus musculus	2.7Gb	
哺乳动物	人	Homo sapiens	3.2Gb	
植物	贝母属一种	Fritillary assyriaca	130Gb	
鱼类	石花肺鱼	Protopterus aethiopicus	130Gb	已知最大的脊椎动物基因组
变形虫	无恒变形虫	Amoeba dubia	670Gb	已知的最大基因组 <sup>[19]</sup> （但有争议）



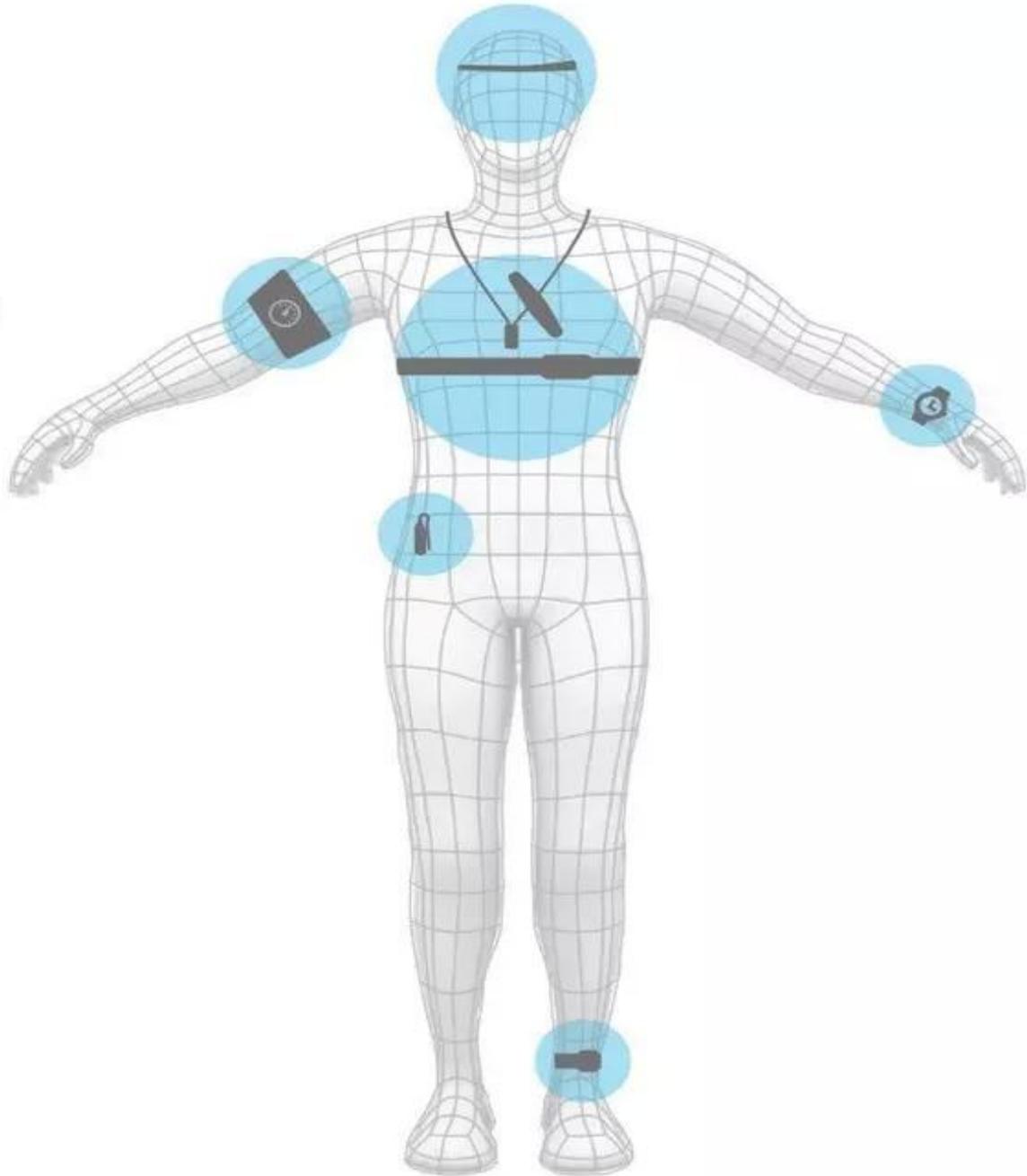
500亿吨  
10亿集装箱  
 $10^{21}$  超级计算机

3D MRI 150MB  
3D CT 1GB  
胸透120MB  
X Ray 30MB



36.6M 病人/年  
665TB/年  
80% 非结构化数据  
24~40% 增长率

- 身体姿势**
  - Zephyr
  - Jins Meme
- 肌肉活动**
  - Athos
- 血压**
  - iHealth
  - Withings
- 皮肤导电情况**
  - Basis
  - Body Media
  - Empatica
  - Neumitra
- 运动**
  - Fitbit
  - Nike Fuelband
  - Jawbone Up Band
  - Garmin
  - Samsung MC10
  - Zephyr
  - Withings
  - Spire
  - iHealth
  - Jins Meme
  - Proteus
  - Neumitra
  - Body Media
  - Empatica
  - Owlet
- 携氧能力**
  - iHealth
  - Withings
  - Owlet
- 含水量**
  - Corventis
  - MC10
- 体温**
  - Tempdrop
  - MC10
  - Empatica
  - BodyMedia
  - Basis
  - Owlet



- 大脑活动**
  - NeuroSky
  - Melon (acquired by DAQRI)
  - Emotiv
- 血糖**
  - Google
  - Dexcom
  - Glysens Incorporated
- 眼动跟踪**
  - Jins Meme
- 睡眠**
  - FitBit
  - Rest Devices
  - Garmin
  - Nike
  - Amiigo
  - BodyMedia
  - Withings
  - Samsung
  - Misfit
  - Jawbone
  - iHealth
  - Basis
  - Owlet
- 呼吸**
  - Spire
  - Zephyr
  - Rest Devices
- 消化情况**
  - Proteus
- 心脏监护**
  - Zephyr
  - Withings
  - Sprouting
  - Proteus
  - iHealth
  - Basis
  - Corventis
  - AliveCor
  - Samsung
  - Garmin
  - Empatica
  - Owlet

# Friendship and natural selection

Nicholas A. Christakis<sup>1,2,3,4</sup> and James H. Fowler<sup>4,5,†</sup>

<sup>1</sup>Department of Psychology, <sup>2</sup>Ecology and Evolutionary Biology, and <sup>3</sup>Medicine, Yale University, New Haven, CT 06520, and <sup>4</sup>Medical Genetics Division and <sup>5</sup>Political Science Department, University of California, San Diego, San Diego, CA 92103

Edited by Francisco J. Ayala, University of California, Irvine, CA, and approved March 7, 2014 (received for review January 15, 2014)

More than any other species, humans form social ties to individuals who are neither kin nor mates, and these ties tend to be with similar people. Here, we show that this similarity extends to genotypes. Across the whole genome, friends' genotypes at the single nucleotide polymorphism level tend to be positively correlated (homophilic). In fact, the increase in similarity relative to strangers is at the level of fourth cousins. However, certain genotypes are also negatively correlated (heterophilic) in friends. And the degree of correlation in genotypes can be used to create a "friendship score" that predicts the existence of friendship ties in a hold-out sample. A focused gene-set analysis indicates that some of the overall correlation in genotypes can be explained by specific systems; for example, an olfactory gene set is homophilic and an immune system gene set is heterophilic, suggesting that these systems may play a role in the formation or maintenance of friendship ties. Friends may be a kind of "functional kin." Finally, homophilic genotypes exhibit significantly higher measures of positive selection, suggesting that, on average, they may yield a synergistic fitness advantage that has been helping to drive recent human evolution.

genetics | social networks | friendship detection

Human social interactions, and the networks they give rise to, show striking structural regularities (1, 2), even when comparing modernized networks with those in hunter-gatherer societies (3). Indeed, friendship is a fundamental characteristic of human beings (3–5), and genes are known to play a role in the formation (6), attributes (7), and structures (8) of friendship ties. Social ties also evince homophily, the tendency of people to form connections with phenotypically similar others (9). Evolutionary models suggest that homophily can evolve under a wide range of conditions if there is a fitness advantage to same-type interactions (10, 11). And candidate gene studies (12, 13) have recently identified one gene variant that exhibits positive correlation or similarity between friends (homophily) and another variant that exhibits negative correlation or dissimilarity (heterophily). It remains unclear, however, whether this phenomenon extends to multiple genotypes across the whole genome, and it is not known what role genotypic correlation may have played in human evolution.

There are (at least) four possible reasons that friends may exhibit homophily in their genotypes (12). First, correlation in genotypes may be a trivial by-product of the tendency of people to make friends with geographically proximate or ethnically similar individuals who also tend to share the same ancestry. Thus, it is important to use strict controls for population stratification in tests of genetic correlation (below, we rely on the widely used principal-components method to control for ancestry). Second, people may actively choose and retain friends of a similar genotype or they may avoid or terminate friendships with people who have different genotypes ("birds of a feather flock together"). This process may take place via a variety of mechanisms; for example, although it is unlikely that people would observe the actual genotypes of others around them, they can observe and prefer certain phenotypes, and these may obviously be influenced by specific genotypes. It is well known that people prefer to associate with others they resemble phenotypically

(9), but what is not known is how this observation translates to the single-nucleotide polymorphism (SNP) level. Third, people may actively choose particular environments, and in those environments, they may be more likely to encounter people with similar phenotypes influenced by specific genotypes. If people then choose friends from within these environments (even at random), it would tend to generate correlated genotypes. Fourth, people may be chosen by third parties or otherwise selected into environments or circumstances where they then come into contact with similar people. These four reasons are not mutually exclusive, of course, and they may operate in parallel: two people may become friends through both active choice of each other and active or passive choice of a consocial environment.

In contrast, there are fewer reasons that friends may exhibit heterophily in their genotypes (12). For example, heterophily is not likely to arise by population stratification, nor by a simple process of people choosing, or being drawn to, the same environment for the same reason. Instead, there are two other processes that might be at work. First, certain environments might foster interactions between individuals with dissimilar traits. Second, people may actively choose to befriend people of a different type ("opposites attract"). In fact, such a phenomenon has been proposed for reproductive relationships, and some experiments suggest that men and women may choose partners with opposite immune system genotypes (14, 15).

Importantly, all of these processes may be at work simultaneously, and humans may select friends and environments based on a wide variety of traits, some of which result in advantages when homophily is present (cooper) and others of which may yield advantages to heterophily (complementarity or specialization) (3, 11). The people to whom we are connected provide important capabilities, from the ability to ward off infections, to the ability to transmit or exploit useful information, to the ability to reciprocate cooperative exchanges. Consequently, the fitness advantage of some gene variants might be influenced by their parallel presence or absence in other individuals to whom a person is connected.

Evolutionary models show that preferences for both homophily and heterophily can evolve depending on the relative fitness advantages of genotype similarity and dissimilarity on given traits (10). However, these models also show that homophily evolves under a much wider variety of conditions than heterophily—even when the fitness advantage to dissimilarity exceeds the fitness advantage to similarity (10). The reason is that it is less costly to find and successfully interact with a similar partner

## 社交/婚姻与基因

# 1%

# DDR2

# CYP2A6

This paper results from the Author-McNair-Cockburn of the National Academy of Sciences, "In the Light of Evolution III: Darwinian Thinking in the Social Sciences," held January 10–13, 2014, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and audio files of most presentations are available on the NAE website at [www.nae.edu/eng/8102/Author-McNair-Cockburn](http://www.nae.edu/eng/8102/Author-McNair-Cockburn).

Author contributions: N.A.C. and J.H.F. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

Supplemental materials are available for this article.

†To whom correspondence should be addressed. Email: [jhfowler@ucsd.edu](mailto:jhfowler@ucsd.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1408111111/-/DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1408111111/-/DCSupplemental).

一个人  
一生数据

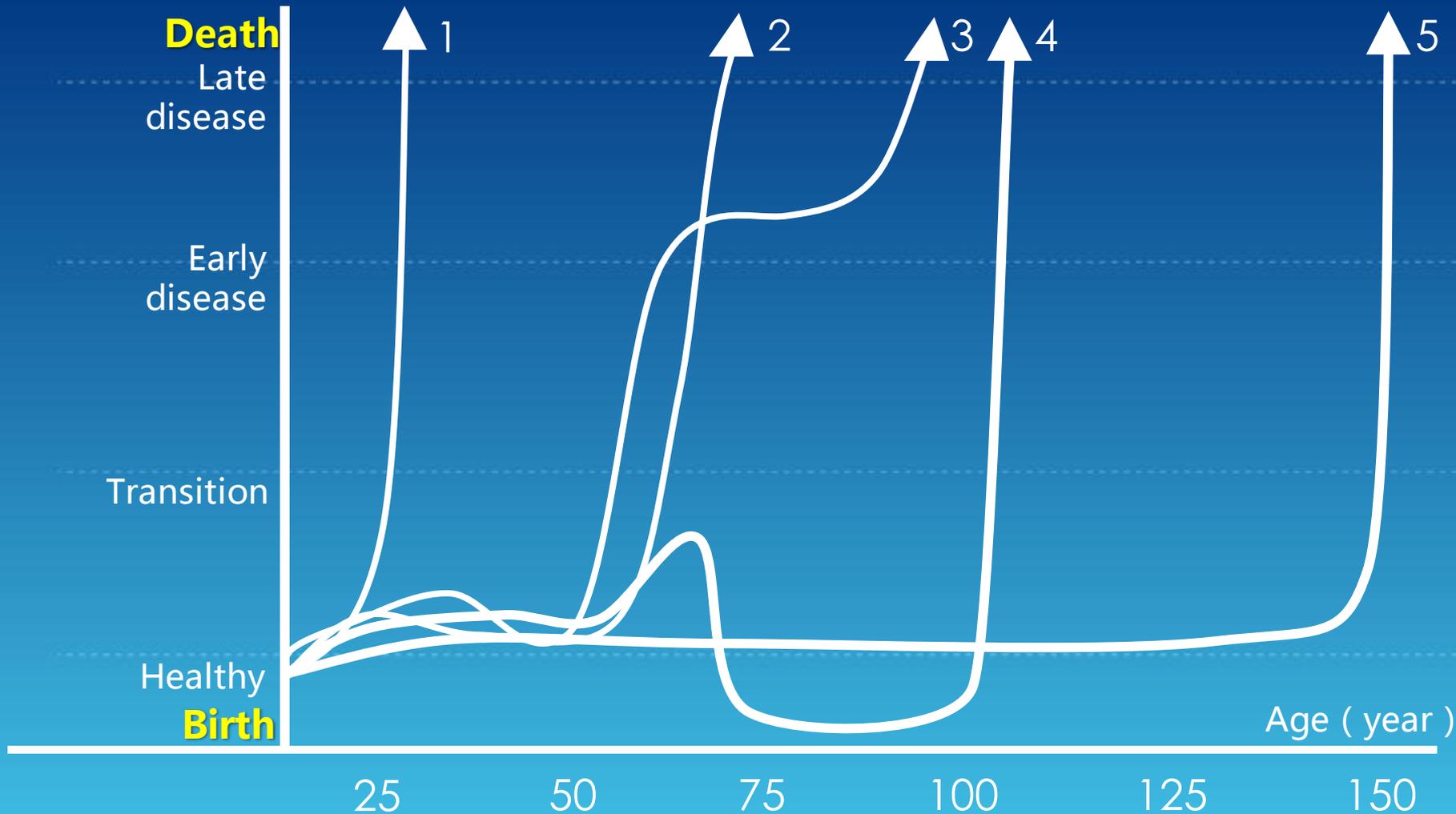
10TB = 1TB + 0.7TB + 2TB + 3TB + X TB  
基因组 转录组 表观组 宏基因组 其他





一百万人  
一生数据  
10EB

# 目标？

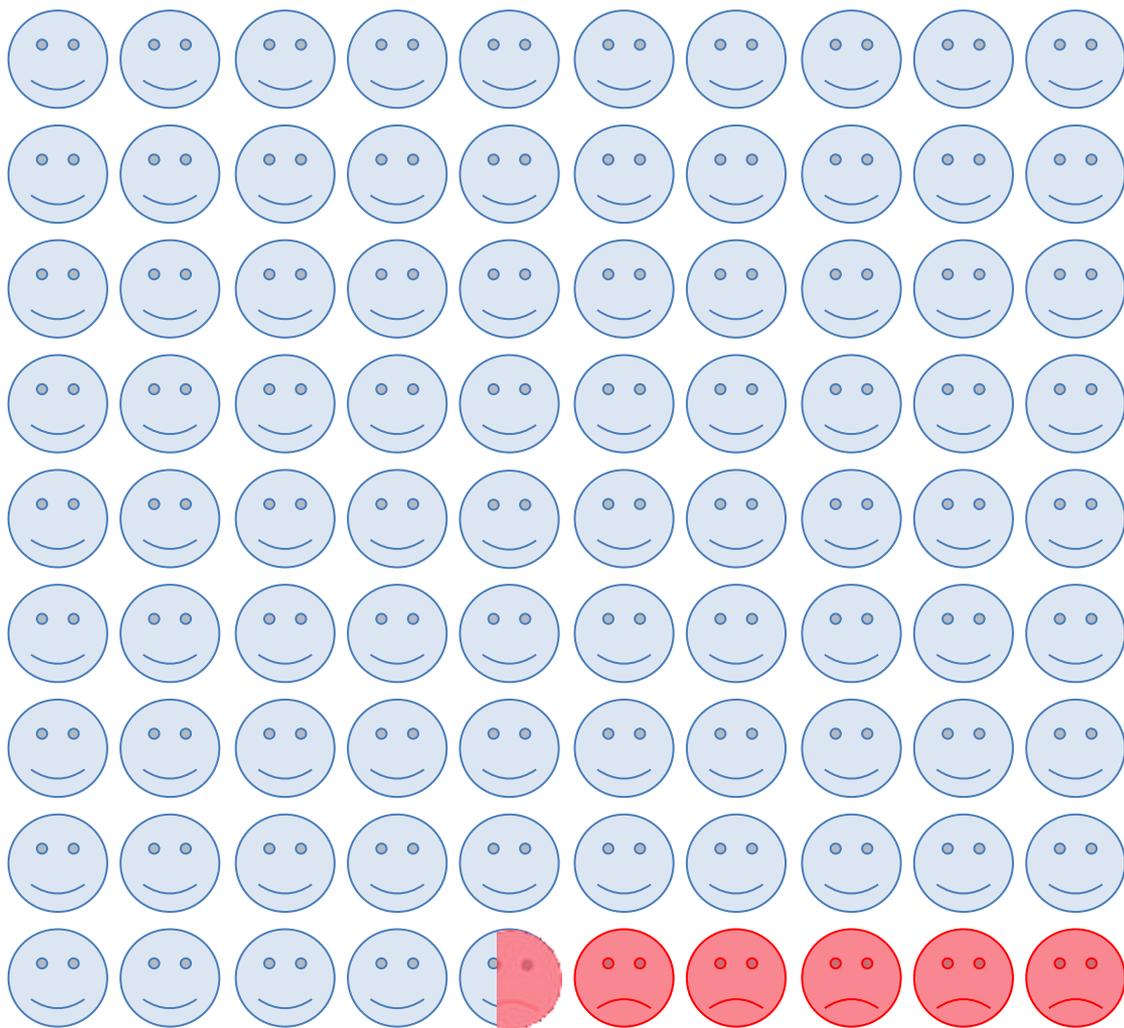


15% 不孕不育  
5% 出生缺陷

代谢病  
癌症  
传染病

30% 代谢病  
心血管疾病  
癌症

50% 心血管疾病  
癌症  
老年病



出生缺陷

5.6%

遗传疾病

30%

# 85M/5.6%

视力: 12.6M  
听力: 27.8M  
语言: 1.3M  
肢体: 24.7M  
智力: 5.7M  
精神: 6.3M  
猝死: 550K  
癌症: 3.2M  
老年痴呆: 20M  
.....

史蒂芬·霍金

物理學家

漸凍人



发病率

0.65‰ ~ 1‰



23andMe

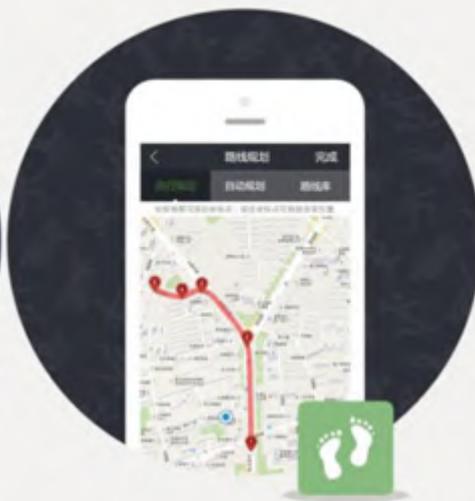


# 日常生活大数据



## 多终端云同步

为您的健康提供一站式服务



## 全新运动

独到规划每段新旅途



## 全新饮食

科学计算每日膳食营养均衡



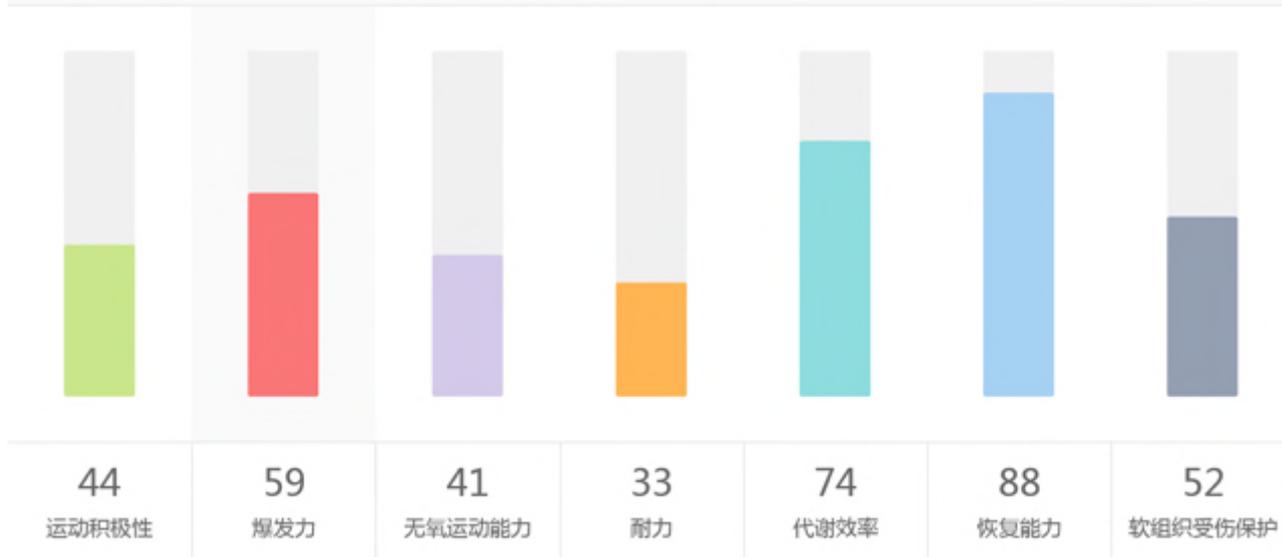
## 华大检测上线

整合检测报告预知疾病风险

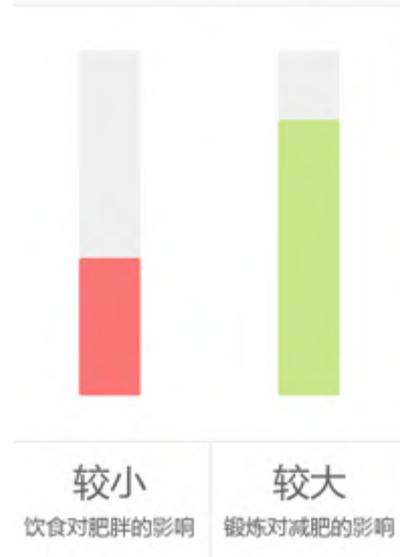


# 运动基因检测

## 运动基因综合结果



## 肥胖基因综合结果



## 爆发力

肌肉在短时间高强度运动中的表现，包括短跑，跳跃，和瞬间加速。

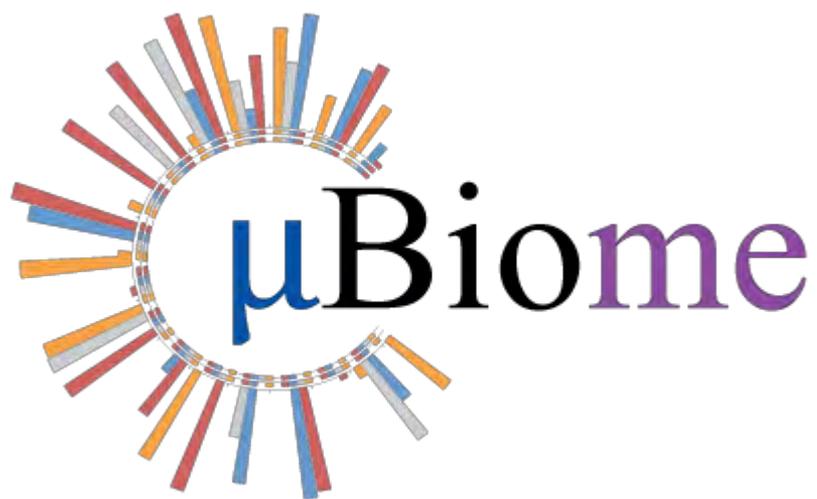




# 饮酒健康管理应用

结合基因检测数据，为您提供饮酒健康管理服务





厚壁菌

Firmicutes

拟杆菌

Bacteroidetes

变形菌

Proteobacteria

纤维菌

Fibrobacteres

放线菌

Actinobacteria



All



Gut



Mouth



Nose



Skin



Genitals

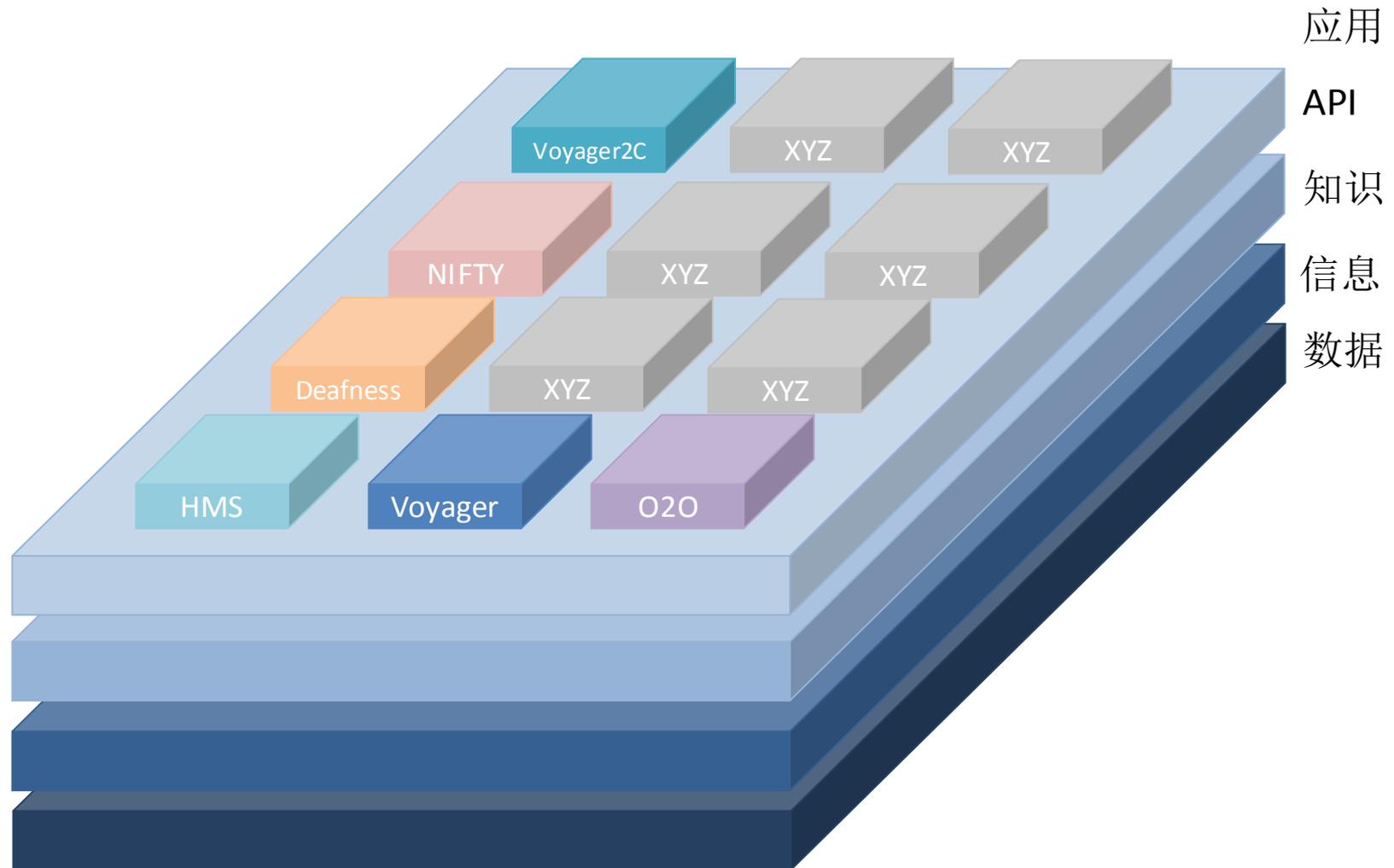


Other



Spare

# 生物医学大数据生态系统



**Thanks!**

