

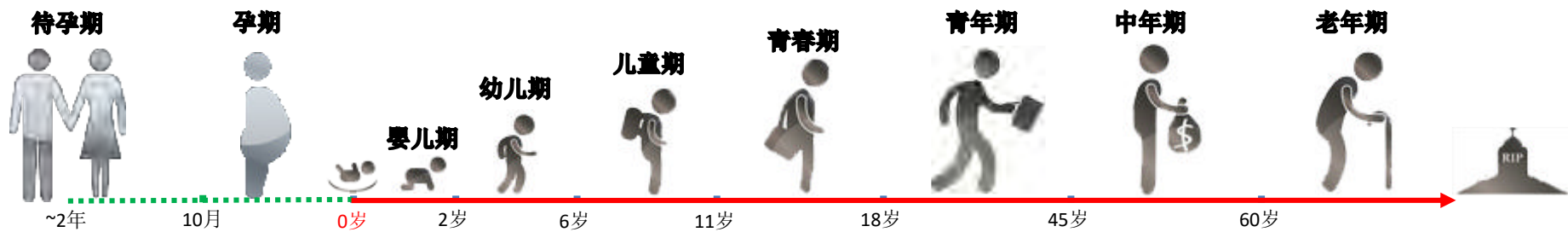


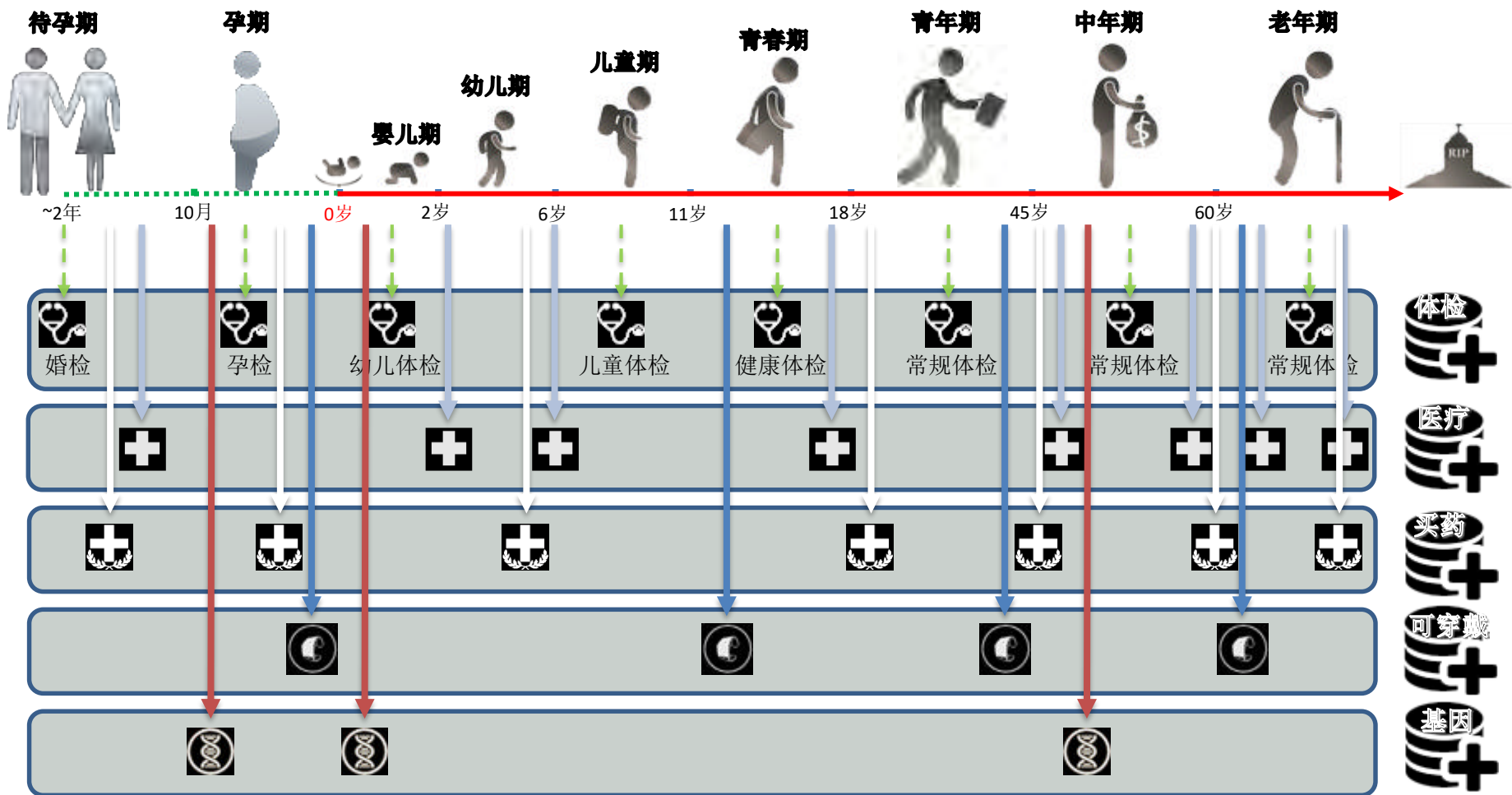
基因大数据产业发展

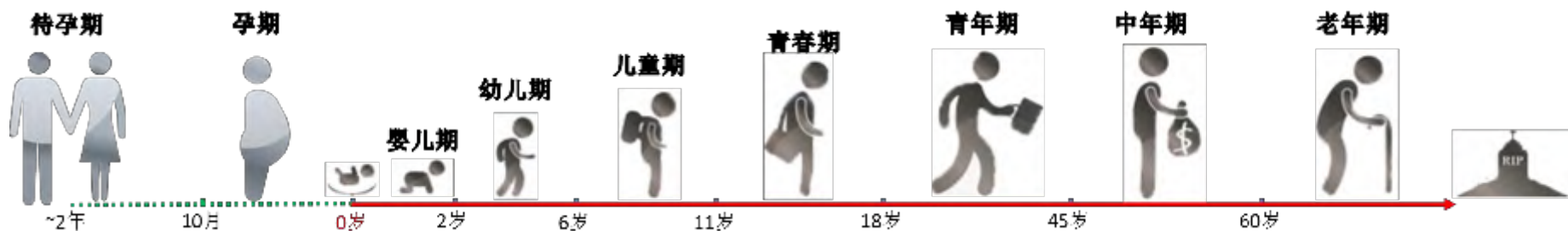
上海大数据联盟

时炜

2016年7月





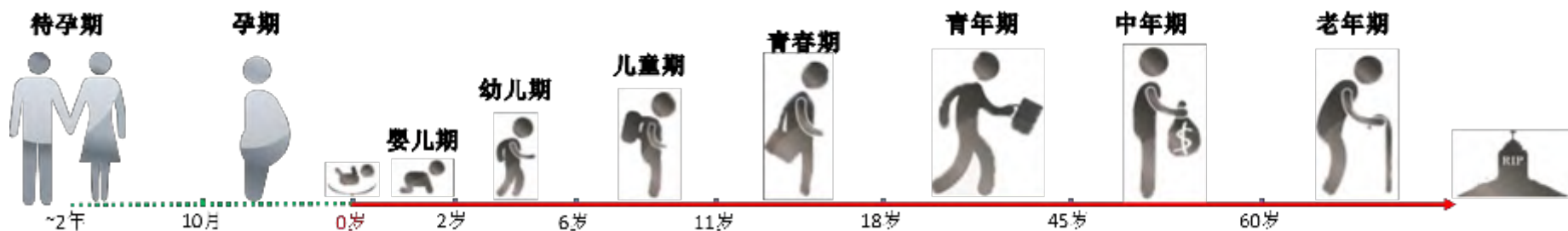


居民档案



健康医疗数据与诊疗信息与个人隐私、健康状况乃至家族遗传联系紧密，是健康大数据隐私性与医学伦理的根源所在

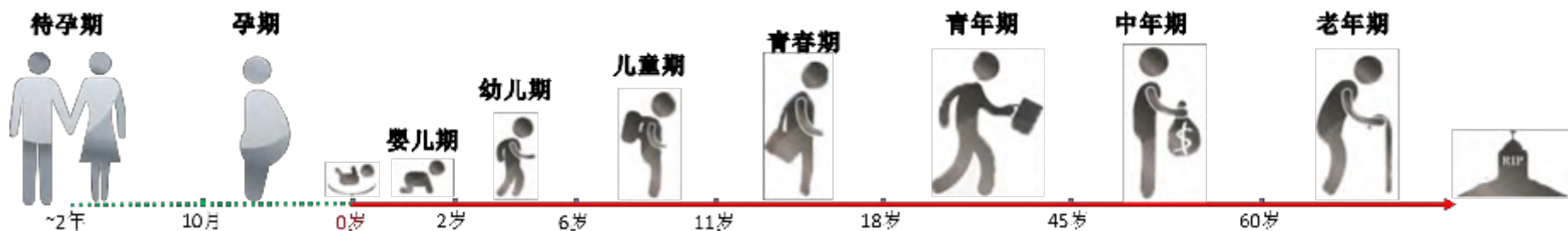




体检
 医疗
 买药
 可穿戴
 基因

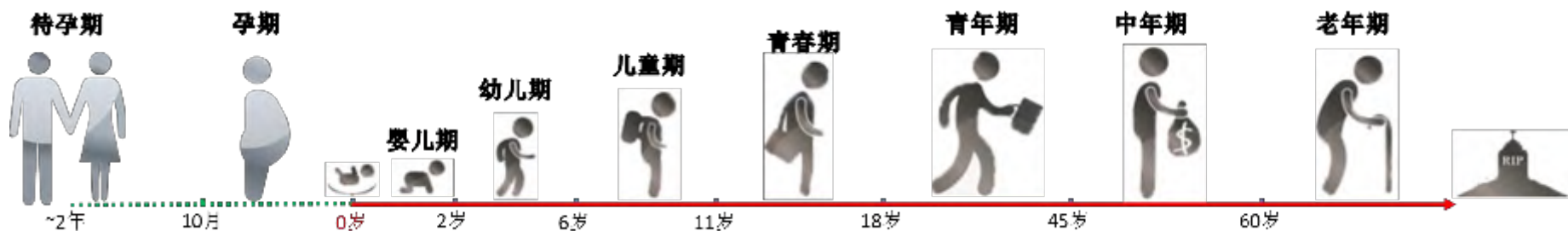
从个人全生命周期来看，只有当散落在各个数据库中的信息按照时间轴串联起来的时候，才能够较为准确的监控并预测一个人的健康状况。

当碎片化的数据进行聚合时，必然涉及到不同数据库之间的数据标准、数据的质量控制，以及在数据流动过程中的安全管理。后台则对应着数据共享与开放过程的商业模式与利益分成机制，这是整个健康大数据的难点所在。



体检
医疗
买药
可穿戴
基因

单一的监控数据能够反映一个人的生理指标变化趋势，但无法确认是否健康。相应的标准是从千百万人的生理数据中提炼而来，具有统计性特征，采样的样本数越多，参考价值越大。



健康，除了与家族遗传、个人体质有关外，受周围环境的影响巨大，环境质量监控数据、食品流通及安全数据，以及与生活方式相关的各类应用及由此产生的数据，都属于大数据的广义范畴。

体检
医疗
买药
可穿戴
基因



水质质量



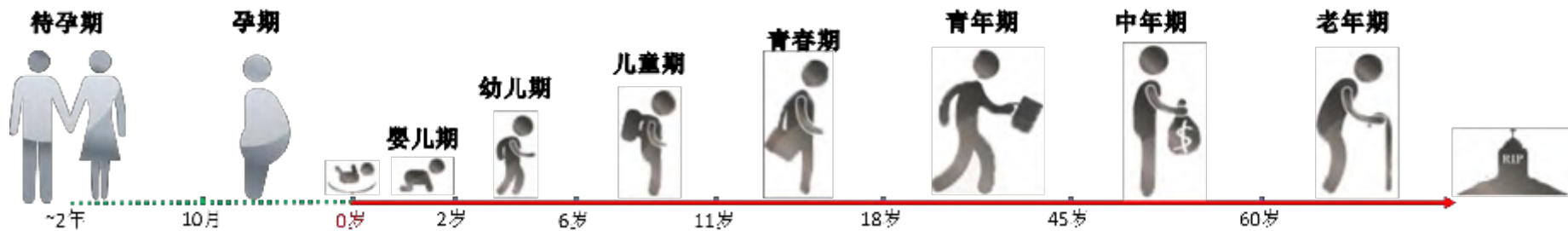
大气质量



食品



生活方式



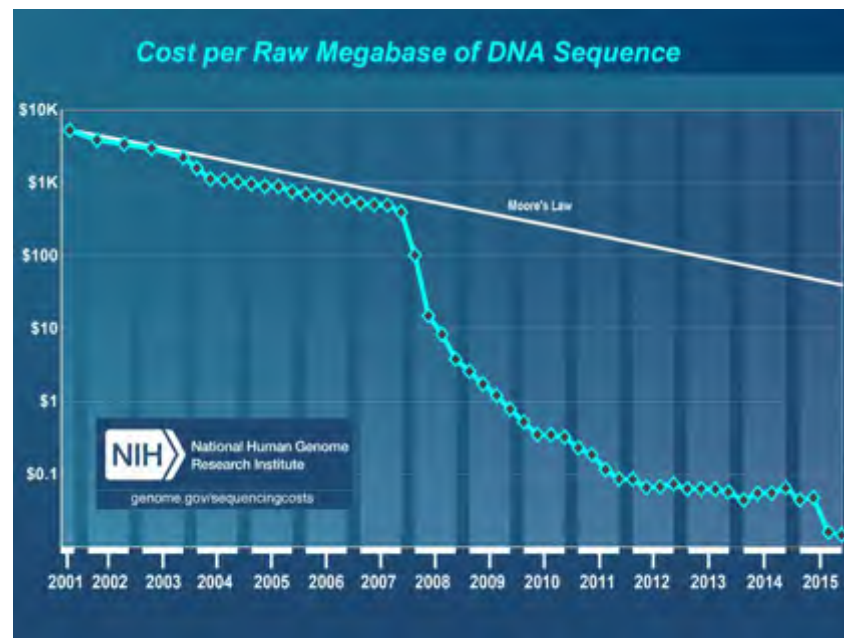
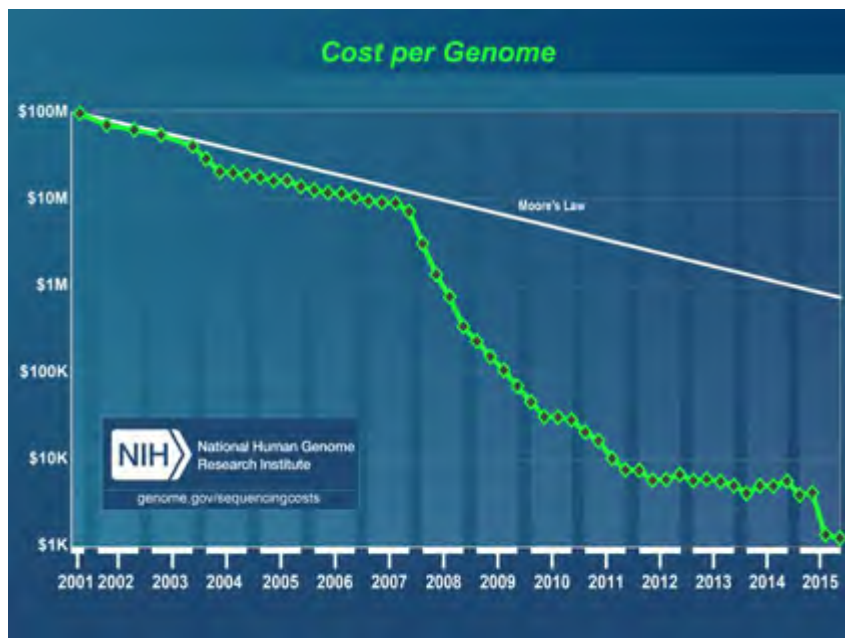
基因是精准医疗的核心，是健康大数据最为重要的组成部分。

体检
医疗
买药
可穿戴
基因

测序技术	原理	公司及仪器	测序通量	准确率	读长	优缺点	应用状况
第一代	链终止法	ABI: 3730XL	0.2MB	>99%	400-900	高读长、高精度、一次性达标率高、通量相对较低	成本高、速度慢、应用少
第二代	边合成边测序，可逆终止法	Illumina: Sloexa Roche: 454 (停产) ABI: Solid	400Mb-1.8T	>99%	50-300	高通量、低成本，但存在模板扩展和序列读长的缺陷	目前应用最为广泛的技术
第三代	单分子合成测序	PacBio: Heloscope; PacBio: SMRT	0.2-30Gb	<90%	>1000	高通量、高读长、低成本，但准确度不高	研发阶段，未真正商业应用
第四代	纳米孔外切酶测序	Oxford: MiniION、 GridION	5-50Gb	>90%	>1000	高通量、高读长、低成本、小型化	研发阶段，未真正商业应用

来源：经公开资料整理

基因测序成本以超“摩尔定律”的速度下降，不仅为基因测序技术的广泛使用提供良好的经济技术，推动着测序产业的迅猛发展，同时也产生了海量的基因数据。



来源：National Human Genome Research Institute

测序技术的不断进步和测序成本的进一步下降，使得测序数据呈现爆发式增长趋势。各国所推出的基因组图谱计划，更是推动了这一增长态势，保守估计2015年生物医学行业内所产生的数据量已经超过了500PB，这其中绝大多数来自于基因数据。

- 千人基因组计划（英国，美国，中国）
- 癌症基因组图谱（TCGA，美国）
- 国际癌症基因组协会（ICGC，欧洲，美国，中国）
- 个人基因组计划（美国）
- 临床测序计划（ClinSeq，美国）
- 23andMe（美国）
- 云健康（中国）

来源：公开资料整理

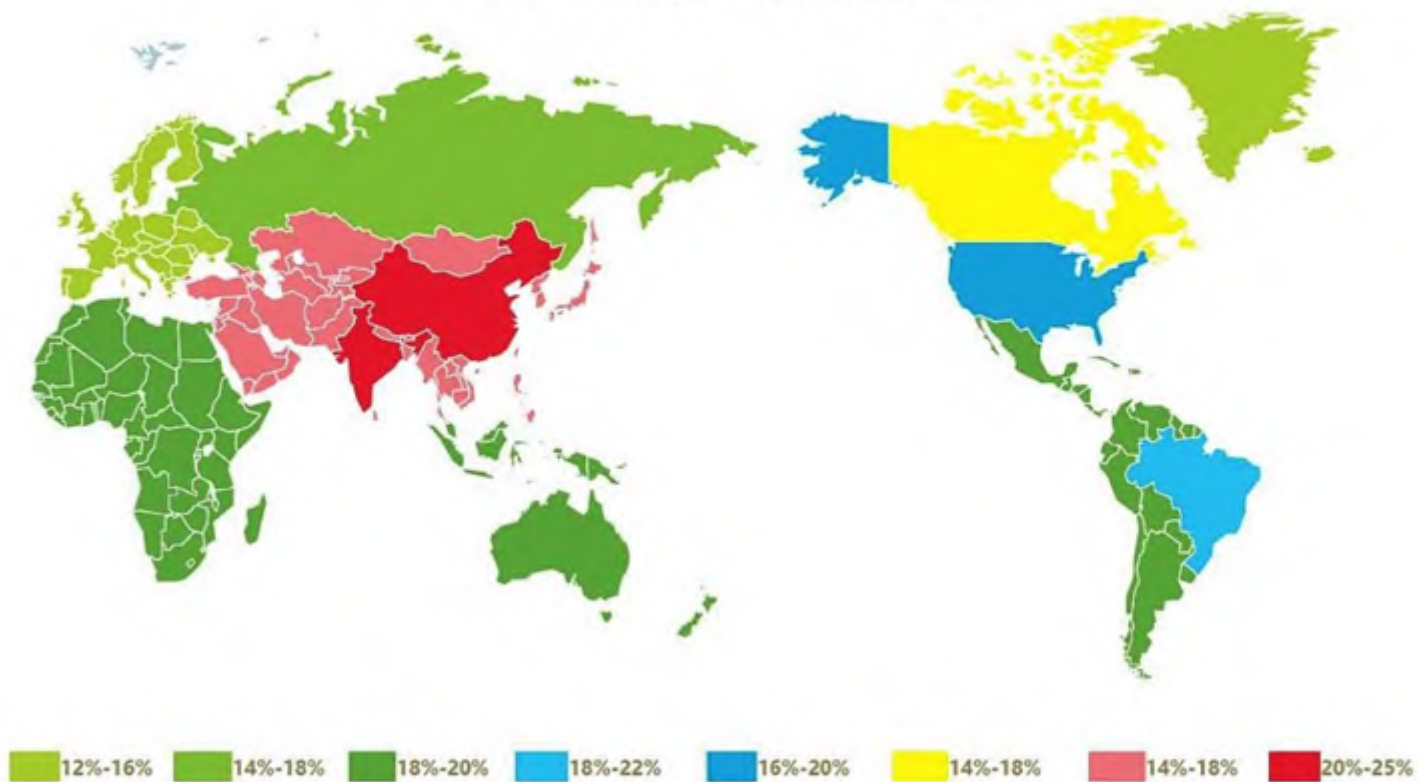


2007-2022全球基因测序行业市场规模



来源：艾瑞咨询《全球二代测序行业投研报告》

2012-2017年全球基因测序市场增长情况

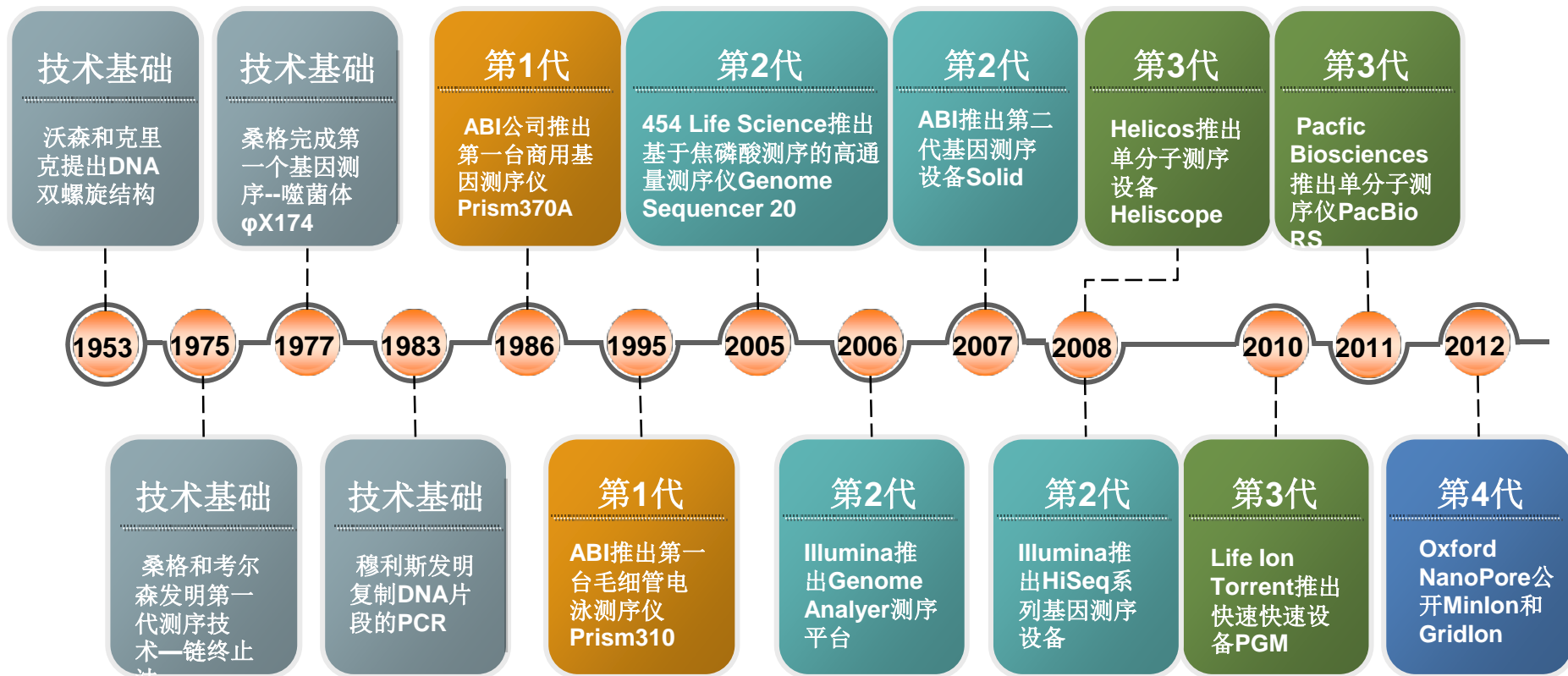


来源：艾瑞咨询《全球二代测序行业投研报告》

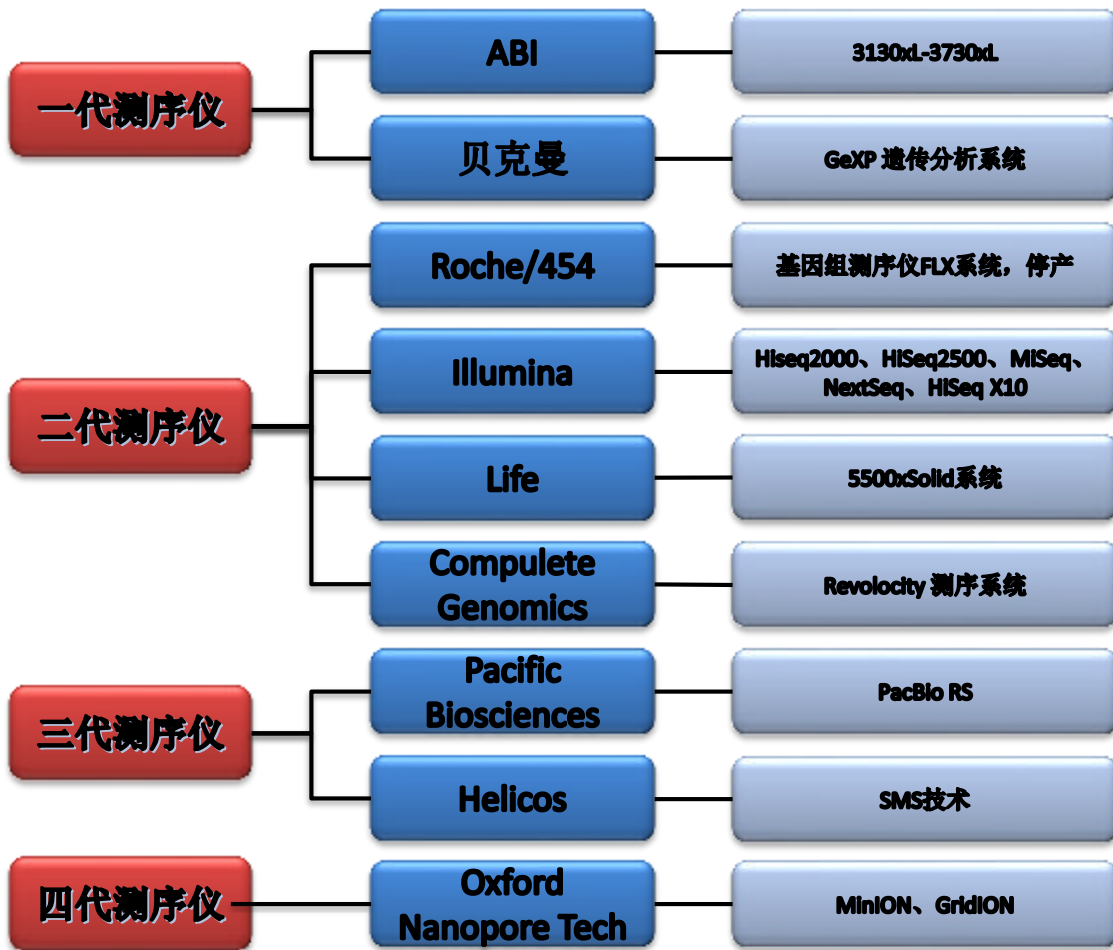
换一个角度看基因产业



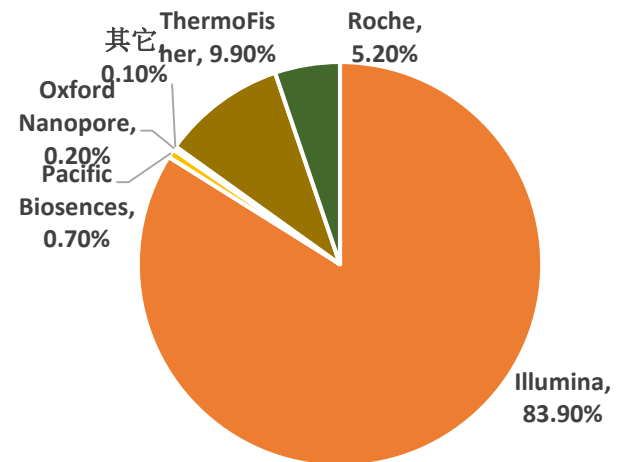
基因测序技术发展历程



来源：艾瑞咨询《全球二代测序行业投研报告》



全球目前共有7389台测序设备，其中ILLUMINA测序平台占比83.9%，Thermo Fisher测序平台占比9.9%，Roche测序平台占比5.5%，CR3超过99%，属于绝对垄断行业。



数据来源: <http://omicsmaps.com>

数据来源: 根据公开资料整理

国产测序仪公司一览

批准名称	批号	公司	研发方式	合作方
二代测序仪	国械注准20153400460	杭州贝瑞和康基因诊断技术有限公司	合作开发	Illumina
二代测序仪	国械注准20143402171	深圳华因康基因科技有限公司	自主研发	
二代测序仪	国械注准20143401961	中山大学达安基因股份有限公司	合作开发	ThermoFisher
二代测序仪	国械注准20153400309	博奥生物集团有限公司	合作开发	ThermoFisher
二代测序仪	国食药监械(准)字2014第2401126	武汉华大基因生物医学工程有限公司	合作研发	ThermoFisher
二代测序仪	国食药监械(准)字2014第2401127	武汉华大基因生物医学工程有限公司	收购	CG
二代测序仪	研发中	中科紫馨科技有限公司	自主开发	
三代测序仪	GeneCare	深圳瀚海基因生物科技有限公司	自主开发	
二代测序仪	研发中	中源协和基因工程股份有限公司	合作开发	北京旷博生物(李同恩)
二代测序仪	研发中	深圳市共进电子股份有限公司	合作开发	上海小海龟(吴东平)

数据来源：根据公开资料整理



换一个角度看基因产业



地区	测序服务机构												
北京	诺禾致源	安诺优达	贝瑞和康	圣谷同创	博奥生物	百迈克生物	康普森	嘉宝仁和	源宜基因	博淼生物	圣庭集团	中美泰和	斯科尔基因
	华牛生物	微旋基因	基云惠康	爱普益	迈基诺基因	量化健康	诺赛基因	毅新兴业	博恒生物	百麦华康	华生恒业	路思达	鑫诺美迪
	中科紫馨	海克维尔	瑞德百奥	英木和	溯源精微	华奕生物	奥维森	布斯坦	信诺佰世	银河基因	华诺时代	泛生子	
上海	药明康德	云健康	派森诺生物	晶能生物	美吉生物	宝腾生物	凡迪生物	佰真生物	南方基因	烈冰科技	生工生物	鼎晶圣物	锐羿基因
	欧易生物	翰宇生物	泛亚基因	尤妮曼	联合基因	吉玛生物	康成生物	赛安生物	吉凯基因	上海敏芯	阿趣生物	博苑生物	丰核信息
	生咨生物	英拜生物	凌科生物	泉麦生物	卓立生物	达迈生物	基因科技	基龙生物	源奇生物	赛优生物	希匹吉生物	百世嘉	派航生物
	伯豪生物	祥音生物	锐羿生物	惠研生物	嘉因生物	允英医疗	虹舜生物	捷易生物	思路迪-埃提斯	illumina中国			TAAG Genetics
深圳	华大基因	千年基因	博大维康	易基因	海普洛斯	裕策生物	蓝图基因	普元科技	早知道科技	英马诺生物	恒创基因	锐澳康晨	华因康基因
广州	达安基因	托普基因	锐博生物	燃石科技	基迪奥生物	永诺生物	坤图生物	英格生物	金域检验	瑞科基因	赛哲生物	洪祥生物	
杭州	贝达药业	谷禾生物	浙江天科	中翰金诺	杭州英睿	壹基金	然纳生物	晶佰生物	美格生物	联川生物	奥拓生物	艾迪康医学	迪安诊断
	博圣生物												
武汉	菲沙基因	康圣环球	贝纳基因	生命之美	数桥科技	锦奥生物	大众源生						
苏州	帕诺米克生物	金唯智生物	贝斯派生物	天昊生物	赛业生物	Qiagen	苏州生物医药创新中心						
其它	中宜金大(江苏宜兴)		亿康基因(江苏泰州)		所罗门兄弟医学(江苏盐城)		世和基因(江苏南京)	迪康金诺(江苏南京)		广而生物(江苏南京)			
	苏博生物(江苏宿迁)		锐创生物(江苏绍兴)		健海生物(河北石家庄)		盘古基因(天津)	天津生物芯片(天津)		国信凯尔(山西太原)			
	先导药物(四川成都)		湘雅医学检验所(湖南长沙)		博川基因(湖南长沙)								

测序仪设备被国外厂商垄断，且进入门槛较高，国内企业在消费级测序服务方面发展迅猛，目前约有150家测序服务企业。

来源：银河证券《基因测序的发展趋势与商业模式——探讨精准医疗系列报告之（一）》



中国大陆地区Illumina X10 分布

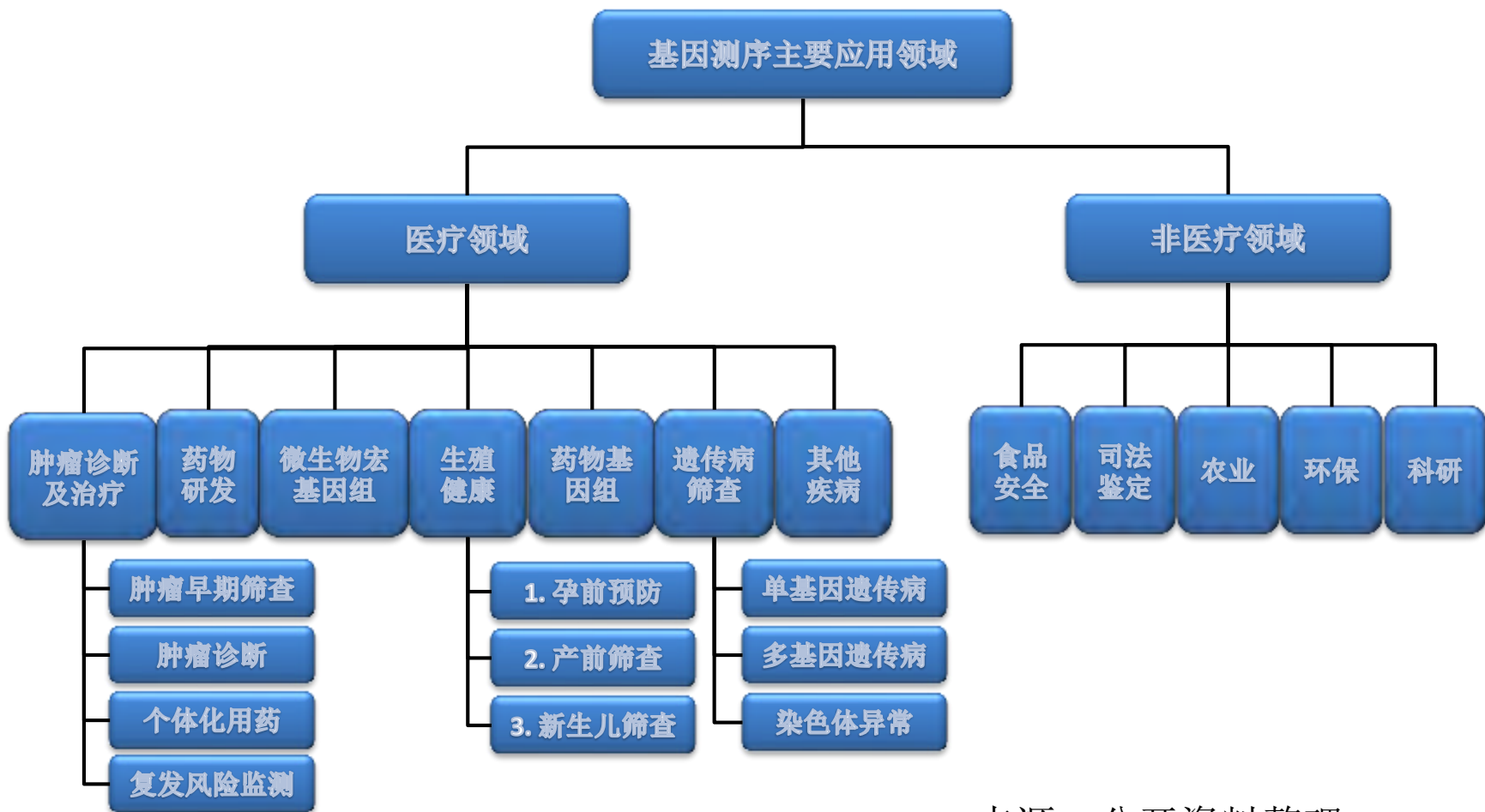
地区	企业或机构	引入时间	套数
上海	药明康德	2014年3月	1
上海	云健康	2014年10月	1
北京	诺禾致源	2014年3月	2
		2016年4月	
北京	安诺优达	2016年4月	1
北京	泛生子	2016年4月	1
苏州	金唯智	2016年4月	1
杭州	贝达药业	-	1
哈尔滨	黑龙江省临床医学精准检验检测中心	-	1

来源：据公开资料整理

2015年底中国大陆已有的3台HiSeq X10接近满负荷运转，而16年后中国大陆的测序能力将达到16.2万人/年，20PB/年的数据产出能力，整个市场的NGS基因组测序处于供大于求的状态，竞争将加剧。

换一个角度看基因产业





来源：公开资料整理

肿瘤诊断及治疗

- 预防——肿瘤基因检测
 - ✓ 肿瘤易感性相关基因
 - ✓ 肿瘤发生相关基因
- 诊治——个性化用药指导
 - ✓ 靶向用药相关基因
 - ✓ 化疗用药相关基因
- 预后——复发检测和药效评估

肿瘤诊断及治疗



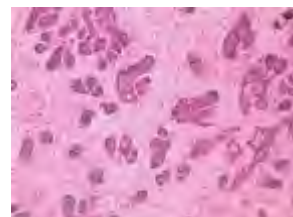
正常DNA



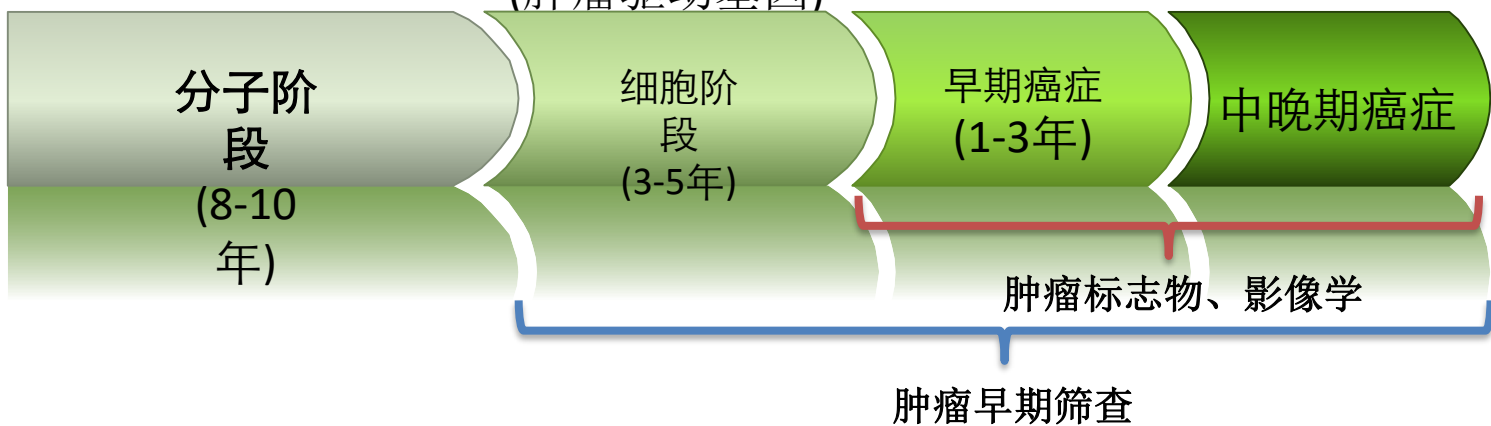
DNA突变
(肿瘤驱动基因)



肿瘤细胞



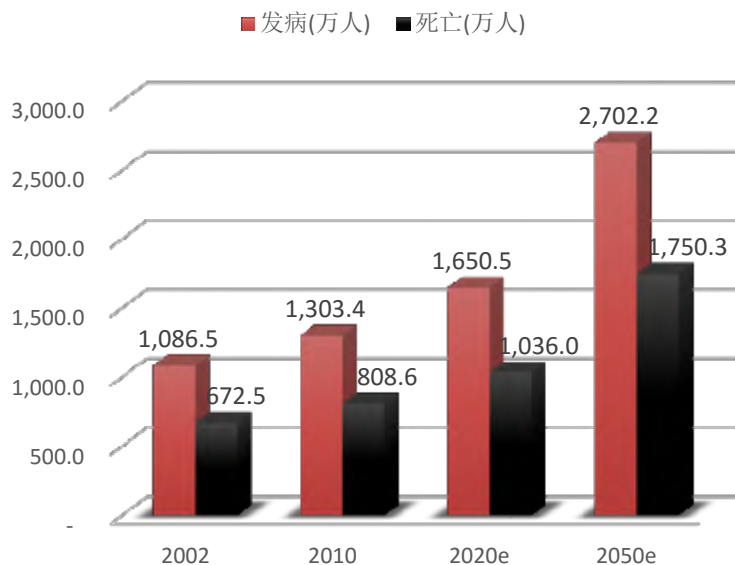
肿瘤



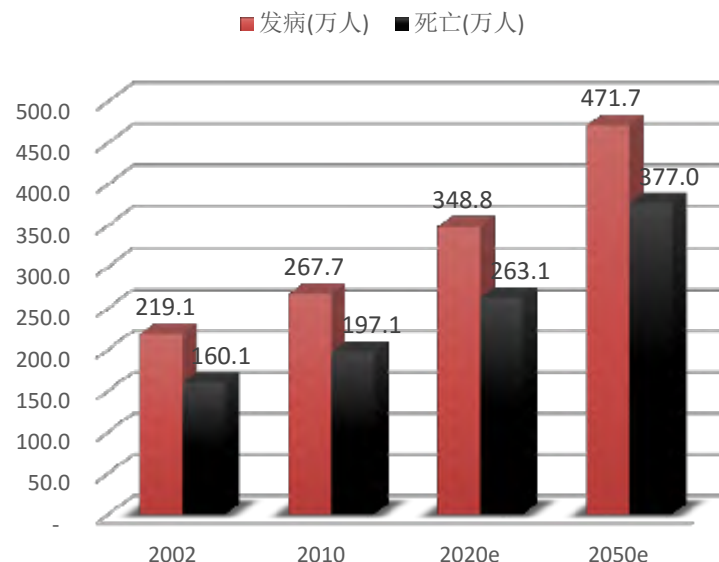
来源：公开资料整理

肿瘤诊断治疗和数据服务是基因测序行业最具投资价值的细分领域。

2002-2050年全球癌症发病和死亡预测



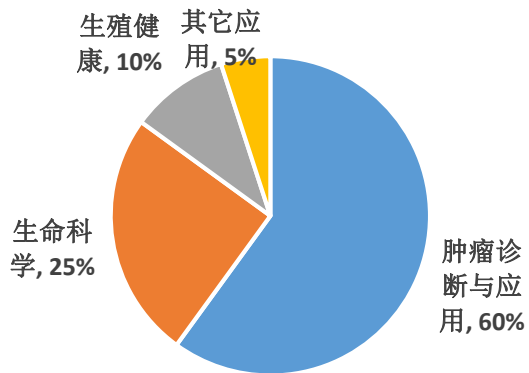
2002-2050年中国癌症发病和死亡预测



来源：艾瑞咨询《全球二代测序行业投研报告》

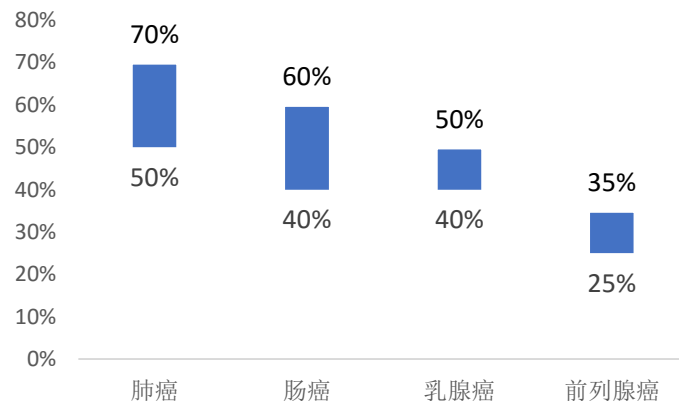
ILLUMINA公司预测基因测序全球市场总容量为200亿美元，肿瘤诊断与应用方向为120亿美元，占60%，是基因测序的最大应用市场。同时，麦肯锡预测基因测序在肺癌、肠癌、乳腺癌和前列腺癌等领域的渗透率超过20%。

基因测序各细分市场占比情况



来源：Illumina

2018年测序技术在癌症领域渗透率预测

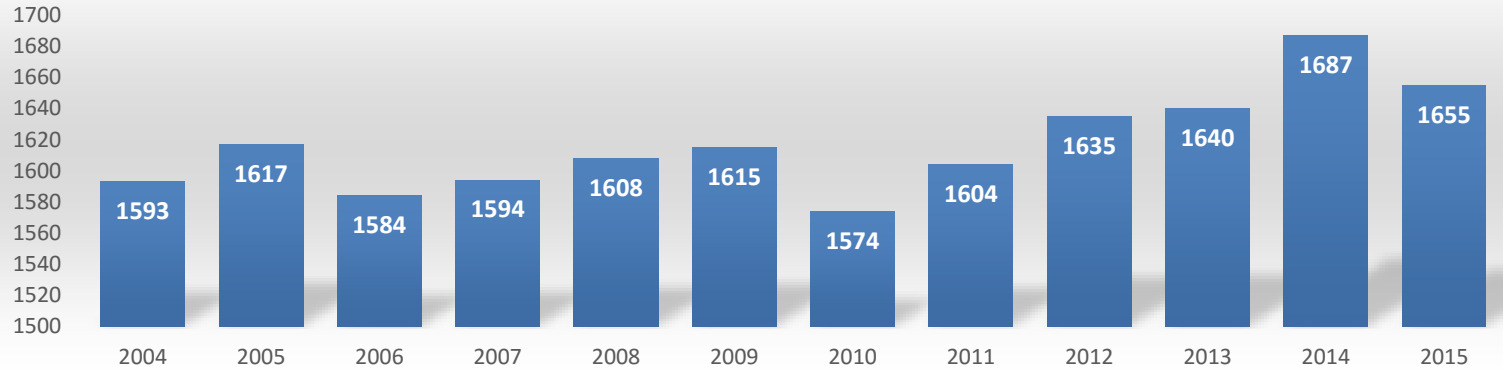


来源：麦肯锡

生殖健康

- 一级预防——孕前预防(PGD/PGS)
- 二级预防——产前诊断、筛查(NIPT)
- 三级预防——新生儿筛查

中国历年新出生人口数(万)



产品名称	项目	2015 年上半年	2014 年	2013 年	2012 年
生育健康类服务 (以样本数为统计口径)	产能(个)	360,500	573,000	403,000	210,000
	产量(个)	326,987	515,567	350,904	178,466
	销量(个)	326,985	515,552	350,896	178,460
	产能利用率	90.70%	89.98%	87.07%	84.98%
	产销率	100.00%	100.00%	100.00%	100.00%

来源：华大基因招股说明书

2016年中国胚胎植入前及新生儿基因测序市场规模预测

项目		假定情况及说明		
试管受孕人群 (万人)	124.35	1) 2014年末育龄妇女怀孕人数为1381.7万人; 2) 不孕不育夫妇占育龄夫妇比例15%; 3) 假设不孕不育人群中60%的人接受试管受孕		
假设胚胎植入前基因测序渗透率 (%)		20%	40%	60%
胚胎植入前基因测序渗透人群 (万人)		24.87	49.74	74.61
胚胎植入前基因测序市场规模 (亿元)		6.22	12.44	18.65
项目		假定情况及说明		
怀孕人数 (万人)	1381.7	2014年末育龄妇女怀孕人数1381.7万人, 即新生儿人数为1381.7万人		
假定新生儿基因疾病筛查渗透率 (%)		1%	5%	10%
新生儿基因疾病筛查渗透人群 (万人)		13.82	69.08	138.17
新生儿基因疾病筛查市场容量 (亿元)		2.76	13.82	27.63

2016年中国产前基因测序市场规模预测

项目		假定情况及说明		
高龄产妇 (万人)	26.50	2014年末35岁以上并首次怀孕人数为26.50万人;		
假设高龄产妇产前基因测序渗透率 (%)		50%	70%	90%
高龄产妇产前基因测序渗透人群 (万人)		13.25	18.55	23.85
高龄产妇产前基因测序市场规模 (亿元)		3.18	4.45	5.72
非高龄产妇 (万人)	1355.03	2014年末35岁以下产妇, 及35岁以上非首次怀孕产妇人数为1355.03万		
假定非高龄产妇产前基因测序渗透率 (%)		5%	10%	20%
非高龄产妇产前基因测序渗透人群 (万人)		67.75	135.50	271.01
非高龄产妇产前基因测序市场规模 (亿元)		16.26	32.52	65.04
无创产检市场规模 (亿元)		19.44	36.97	70.77
生殖健康基因测序市场规模 (亿元)		28.42	63.23	117.05

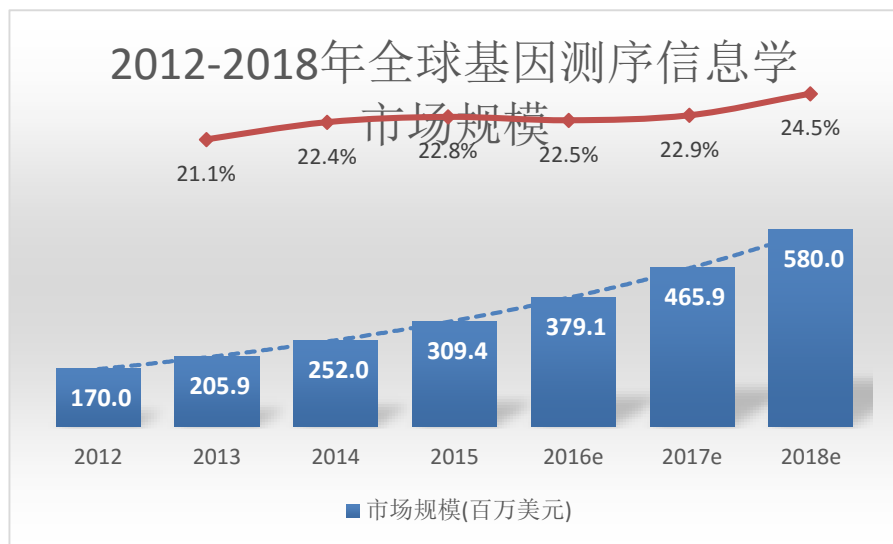
注释: 1) 假定育龄妇女人数稳定, 2016年育龄妇女人数与2014年育龄妇女人数相当;
2) PGS价格为2500/例, NIPT价格为2400元/例, 新生儿基因筛查价格为2500元/例。

来源: 艾瑞咨询《全球二代测序行业投研报告》

换一个角度看基因产业



据公开数据显示，全球基因测序数据分析市场规模发展迅速，2012年市场规模不超过2亿美元，预计2018年将达到6亿美元，复合增长率为22.7%。



数据量的增长，使得基因测序的工作重心从前期的人工测序转移到了数据分析上，原始测序数据并不能反映疾病信息，需要经过基因组的比对，数据过滤筛选等多个步骤才能得到基因组上的变异信息，才能为疾病的诊疗和治疗提供参考。

基因数据分析和解读，关系到下游基因测序的应用，是制约基因测序行业发展的关键环节。

基因数据的处理，既包括了对于测序仪下机数据处理和解读过程，也包括了这一过程中所涉及各类专业数据库。

数据处理



美国国家生物技术信息中心（NCBI），建立于1988年

癌症和肿瘤基因组图谱计划

(The Cancer Genome Atlas, TCGA)

TCGA以人类基因组计划的成果为基础，研究癌症中基因组的遗传变化与出生后细胞中后天的基因变化。TCGA是迄今为止世界上所进行的最大一项基因工程。项目2005年底启动，投入2亿美元，涉及50种以上的包括某些特定亚型在内的肿瘤，收集了数万种人的肿瘤及其对照样本，2007年开始对世界提供数据共享服务，产出的各类组学数据总量达到PB数量级。运用TCGA产出数据到2014年已经在国际高水平研究杂志上发表了超过1000篇论文。

• DDBJ与NCBI的GenBank和EBI的ENA数据库作为国际核酸序列数据库INSD的三大成员，相互共享数据源从C到B

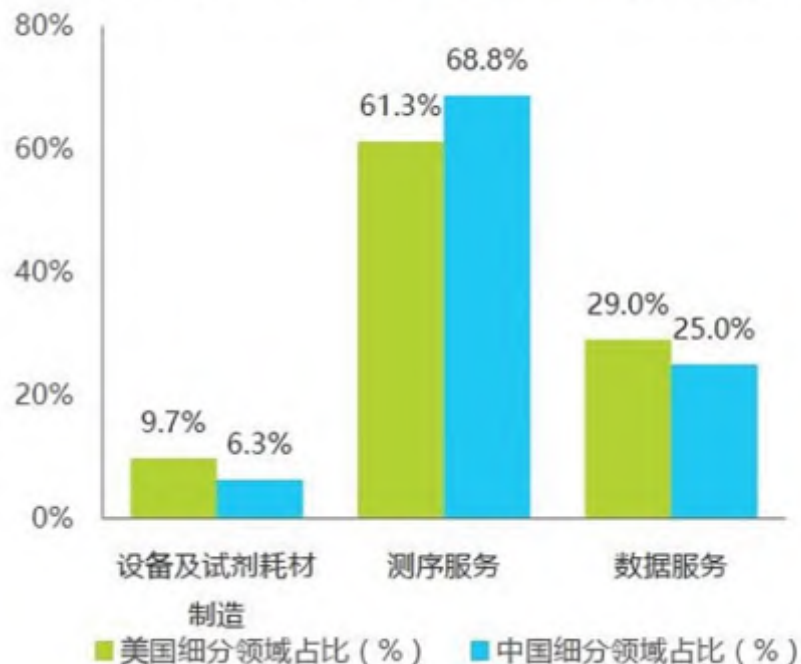
• 提供自有数据库和数据服务20余项

• 100员工，年均投入约5000万美元

2015年中美国基因测序行业细分领域投资数据对比



2015年中美基因测序行业细分领域投资笔数占比



来源：艾瑞咨询《全球二代测序行业投研报告》

总结

- 基因测序行业总体向好，中国市场正以超过20%的增速大步跃进；
- 生物信息学(数据加工)正在成为阻碍整个行业发展的瓶颈；
- 基于基因数据和临床病例，建立一个统一、科学、标准的数据库迫在眉睫；
- 肿瘤市场是基因检测的主阵地，但肿瘤机理过于复杂，基因测序和实施临床干预还不能顺利对接，操作层面和政策层面还有很多困难；
- 基因行业从业者应当建立属于自己的“护城河”；
- 基因大数据是健康大数据的一个有机组成部分，不同数据库之间的共享与打通将是整个行业的发展方向。

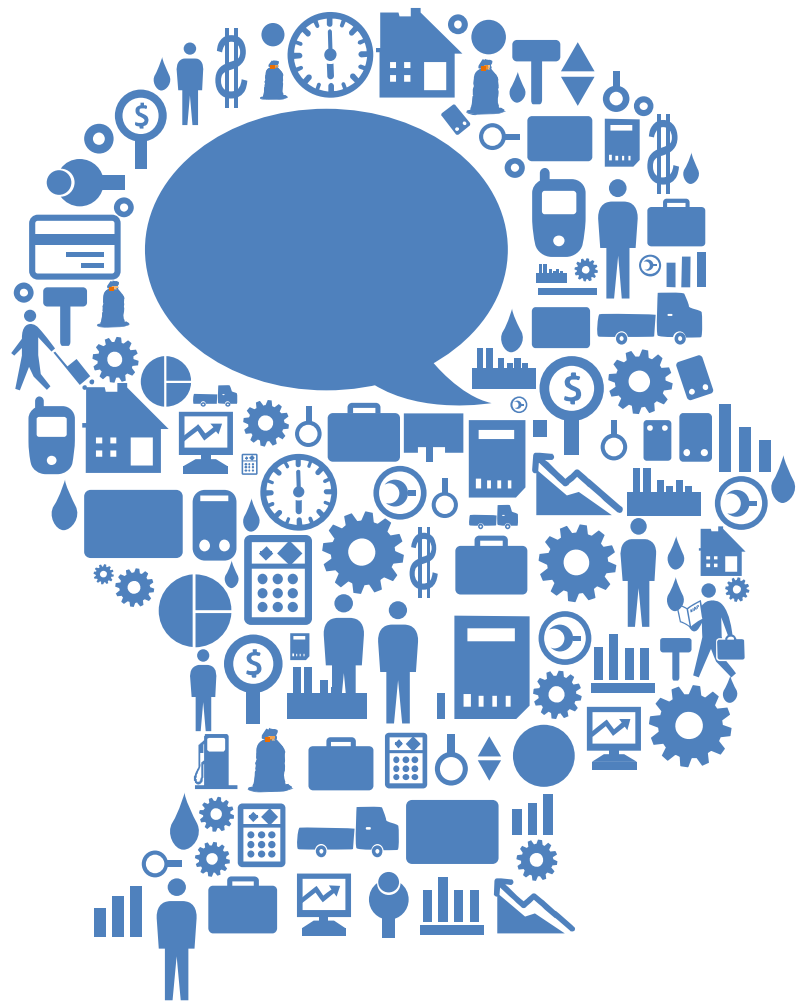
加入我们



欢迎加入我们
微信公众号扫一扫



上海大数据联盟
SHANGHAI BIG DATA ALLIANCE



THANK
YOU