

# 从“人类”基因组计划 到“千人”基因组计划

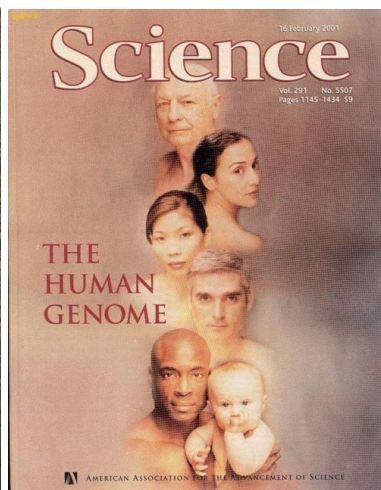
撰文 田埂 苏夜阳



“国际人类基因组计划”被誉为生物学的“曼哈顿原子弹计划”和“阿波罗登月计划”，在2000年，其绘制的人类基因组草图也正式发表。国际顶尖的两大科学杂志《自然》和《科学》同日发表了多国基因组学家联合发表的学术论文，揭开了探索人类基因组奥秘之旅。美国和英国是国际人类基因组计划的主要资助国，他们的领导人也同日向全世界公告了这一科学界新兴取得的成就。中国基因组学科学家们积极投入了人类基因组图谱的绘制工作，高效、高质地提前完成了人类3号染色体上约3000万个碱基对的测序工作（约合人类基因组1%）。在人类基因组图谱揭幕之时，我国领导人也随即代表发展中国家向这一学界盛事表达了祝贺。

国际人类基因组计划绘制的一张属于全人类的基因组图谱，“绘制”的方法是于20世纪80年代发明并完善的桑格DNA测序法。研究者在测序时稳扎稳打，小到一个碱基序列，大到一个染色体都研究得分外严格、精确。当时的测序成本折合美元约为1个碱基对1美元，人类基因组30亿碱基对的测定工作共花费30亿美元，由美、英、法、德、中、日等国联合承担。2003年人类基因组的完全图发表，给为期13年的绘制人类基因组图谱的国际科学计划划下了圆满句点。

揭示人类共有基因组图谱之后，



国际顶尖的两大科学杂志《自然》和《科学》在2000年同日发表了有关人类基因组草图成果的文章。

遗传学家们将目标锁定在深度挖掘这些序列的生物学意义方面。通过研究人类基因组中人与人的不同，理解序列多态性被确定为下一步作战策略。2002年，美、加、中、日、英、尼日利亚等国科学家发起了“国际人类基因组单体型图谱计划（HapMap）”，将研究目标锁定在揭示群体水平上人类基因组遗传多态性中单核苷酸多态性的信息上，以用于绘制供全球生命科学家使用的人类基因组在群体水平上的高密度单体型图谱。

在HapMap第一阶段，来自非洲尼日利亚、欧洲高加索地区、东亚地区的群体由于其人群代表性，被确认为主要研究族群。在其第二和第三阶段，又陆续加入来自（或祖先来自）亚、非、欧和拉丁美洲的其他族群。通过绘制单核苷酸多态性在不同人群中的常见分布和遗传模式的图谱，遗传学家们揭示了人类群体的遗传机制，使人们对自身基因组多态性的认识有了革命性的飞跃，更为发现

复杂性疾病的相关易感基因提供了基础数据和全新的研究思路。

HapMap计划一至三期数据的发表，促进了全基因组芯片的开发。整个人类基因组大约存在300万个多态性位点区域，目前最先进的全基因组芯片已能覆盖其中的200多万个位点区域。运用这些芯片，遗传学家进一步开发了以“病例-对照”研究为基础的全基因组疾病关联分析方法，近年来广泛运用于定位复杂疾病易感基因的研究中，并取得了巨大突破。其研究所挖掘出的遗传信息包括威胁人类健康的常见疾病，如糖尿病、冠心病、前列腺、乳腺癌、风湿性关节炎和肠炎疾病以及老年性黄斑病变等相关的易感基因。然而，芯片技术存在一定的局限性：如遗传标记分辨率不足、无法覆盖低频突变、不同群体存在遗传背景差异等，而这从根本上阻碍了特定族群高发疾病的研究。

2006至2007年间，以焦磷酸测序、边合成边测序技术为代表

的新一代测序技术陆续通过先期研究的检验,逐步投入到基因组学的研究中。新一代测序技术从根本上提高了测序效率,为便捷、高效地开展人类基因组的重测序和数据的深度挖掘工作打下了牢靠的技术基础。为进一步揭示人类个体间基因组差异的生物学意义,通过近一年的论证和筹备工作,于2008年1月22日,英国桑格研究所、中国华大基因和美国国立人类基因组研究所联合发表声明,正式启动了人类基因组计划的三期工程——“千人基因组计划”。

号称研究“千人”基因组的这项雄心勃勃的国际科学合作实际上已将招募志愿者的数量定位在2500人左右。这些志愿者将来自世界各地的27个族群,参与研究工作的科学家也遍布全球。通过对这些志愿者捐献的遗传样本进行全基因组测序,来自英国、中国、美国等地的科学家们正积极努力地绘制一份迄今为止分辨率

最高、最有医学应用价值的人类基因组遗传多态性图谱。这张图谱将揭示所有在群体水平上发生频率高于1%的人类基因组多态性位点,并将发现疾病易感基因的可能性提高至少10倍。

自项目启动以来,千人基因组计划第一阶段总计共完成了1200个人类个体的全基因组测序(覆盖了世界各地区14个不同的人类族群),成功绘制了一张覆盖亚欧非各色人种的全基因组图谱。第一阶段测序和分析任务由世界9大测序中心共同协作完成;其中,华大基因承担并已完成6个族群中251个样品的测序及相应的数据分析任务。其快速、准确、低成本的测序工作和高质量的数据分析得到了国际同仁的广泛赞誉和一致认可,体现了中国基因组学研究的雄厚实力。目前,千人基因组计划已产生高达50TB(5万GB)、约合8万亿个DNA碱基对的测序数据,并将人类参考基因组的数据集合更新至1500万

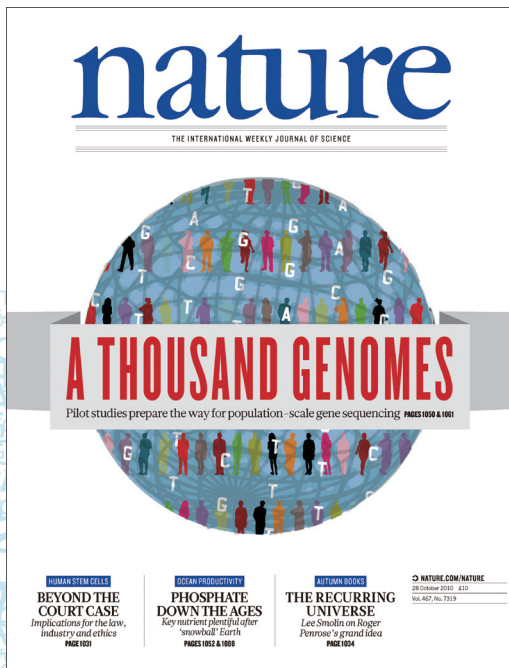
个单核苷酸多态性位点、100万个插入/缺失变异和2万个结构变异。这些遗传变异绝大多数是之前未被发现的。研究还揭示,每个人平均携带约250到300个失去功能(loss-of-function)的突变,其中,50到100个突变与遗传病有关。

2010年10月28日,来自世界各地的参与千人基因组计划的科学家们在《自然》科学杂志上联合发表了题为《通过群体水平测序揭示的人类基因组多态性图谱》的文章,系统报告了千人基因组计划第一阶段的部分研究成果。即千人基因组计划的三个先导项目的研究成果:对两个家系(父母-成年孩子)的高覆盖深度测序,对来自4个人类族群的179个人的低覆盖深度全基因组测序,以及对来自7个人类族群中697个人的全基因组外显子测序。

《自然》杂志在刊发这篇封面文章时,对千人基因组计划的第一阶段工作给予了高度评价,将其描述为“人类基因组学研究史上里程碑式的科学突破”。该计划英方协调人、英国桑格研究所Richard M. Durbin(理查德杜宾)博士指出:“这是人类历史上首次成功地获得了自身的参考基因组数据集合。第一阶段的完成具有重大科学价值。科学家可据此在更大的人群范围内定位人群突变基因,鉴定特定遗传病人群中含有的罕见致病基因以及检测导致人类遗传疾病的相关基因。”换句话说,千人基因组计划第一阶段完成的群体遗传多态性图谱不仅将加速人们对常见疾病易感性基因的

人类基因组计划的实施还将带动医药业、农业、工业等相关行业的发展,产生极其巨大的经济效益和无法估量的社会效益。





千人基因组计划获得突破，在2010年10月的《自然》杂志上，以封面文章的形式发布了迄今最详尽的人类基因多态性图谱。

发现，还将加深人们对自身基因组结构差异的认识，同时也为各种疾病的关联分析提供迄今为止最详细的基础数据。

由于其产出数据量庞大，截至2010年底，千人基因组计划产出的大部分数据尚未在国际学术刊物上发表。为了第一时间给全球研究者提供这些有意义的数据，科学家们在千人基因组计划的官方网站上搭建了一个公共数据库，将所分析完成的全部数据及时上传到该IP地址：<http://www.1000genomes.org/data>。

这一公共数据资源库中存放着的研究数据为各种疾病的关联分析提供了无比详实的基础数据，为解释人类重大疾病发病机理，开展个性化预测、预防和治疗打下了基础，同时还有望推进遗传学家对人类群体遗传学和人类进化史的研究。可以说，作为国际人类基因组计划和国际人类基因组单体型图计划的延续和发展，

千人基因组计划完成了基因组科学从基础科学向应用科学过渡的关键战略转移，有效地推进了临床转化医学的兴起和发展。

在人类数百万年的进化过程中，由于自然环境和地域的不同，形成了各具特色的人类族群。这些独特的族群遗传背景使得每一族群在疾病易感性和药物敏感反应方面与其他族群存在一定的差异。以我国为例，我国是一个幅员辽阔的多民族国家，人口占全球人口1/5以上，而少数民族多达55个。过去，由于受制于技术手段而无法挖掘出族群特有的基因组序列。千人基因组计划证实人与人之间存在个体独有的基因序列，并成功地识别出在人群中罕见的低频变异，所构建的人类基因组多态性的新图谱能更快地锁定与疾病相关的遗传多态性位点，并有效地推动常见疾病的诊断、治疗和预防工作。通过设计、参与千人基因组计划，我国的基因组学家在了解中华民

族的基因组序列信息和遗传多态性特征的旅程上又迈出了关键性的一步，为研究中国高发遗传疾病和保护我国特有遗传资源的工作打下了扎实的基础。

十多年间，华大基因从承担“国际人类基因组计划”的1%工作，到完成“国际人类基因组单体型图计划”的10%，到独立完成第一个中国人基因组图谱“炎黄一号”的绘制，再到人类泛基因组概念的提出和联合启动、推进、开展千人基因组计划，中国基因组学的研究者们通过持之以恒的不懈努力，实现了在生命科学领域国际合作中由参与、到接轨、到同步、再到引领和定标的飞跃。十多年的历练也使华大基因成长为国际基因组学界的一朵奇葩，其研究工作对人类基因组研究的横向和纵深方向都取得长足的进步，这不仅使中国在基因组学与转化医学研究方面与国际前沿保持同步发展，更在近年来展示了后来者居上的研究势头，主动、积极地引领国际基因组学界和生命科学界的研究和跨学科讨论。作为人类基因组学和遗传学研究的核心研究团队，华大基因已作好准备，更深、更广地挖掘中华民族基因序列信息，以期尽快将基础研究的成果转化到临床运用方面，为提高我国各族人民的健康服务。■

作者简介

田埂，华大基因华北区第一负责人。  
苏夜阳，华大基因。

(责编 桑新华)