



新生物学丛书

基因组科学的甲子“羽化”之路

—— 从人类基因组测序到精准医学

Mapping and Sequencing the Human Genome
Toward Precision Medicine



〔美〕美国科学院研究理事会 编

于 军 任鲁风 杨 宇 王绪敏

殷金龙 叶建伟 崔 丽 主译



科学出版社

新生物学丛书

基因组科学的甲子“羽化”之路 ——从人类基因组测序到精准医学

〔美〕美国科学院研究理事会 编

于 军 任鲁风 杨 宇 王绪敏

殷金龙 叶建伟 崔 丽 主译

科学出版社

北 京

图字: 01-2015-7179 号

内 容 简 介

“人类基因组计划”是一项里程碑式的伟大科学计划,该计划的完成带动了生物医学的迅猛发展。精准医疗则是目前国际生物医学研究的热点,已经有包括美英在内的 11 个国家启动精准医学计划,该计划的实施将实现个体水平的精准医疗。本书分上下两篇,分别介绍两个伟大计划的路线图。上篇阐述了人类基因组计划的必要性和重要意义,涵盖了图谱定位、基因组测序的基础原理、技术发展和应用趋势,同时包括了信息和材料的收集、数据库的发布,以及该计划的实施与管理。下篇分析了精准医学的迫切需求和充分条件,描述了疾病知识网络和新型分类法,列举了实现目标的多个经典模型和研究实例,如百万人基因组计划、代谢组模式等,据此可以展望基因组学基础研究与医学的有效结合,实现个体健康状况的精准咨询、预防和诊疗。

本书堪称基因组计划与精准医学的框架导航图,可供广大生命科学工作者和基础医学、检验医学工作者以及对精准医学感兴趣的研究人员参阅。

This is the translation of *Mapping and Sequencing the Human Genome* by Committee on Mapping and Sequencing the Human Genome; Commission on Life Sciences; Division on Earth and Life Studies and National Research Council © 1988; and *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* by Committee on Framework for Development a New Taxonomy of Disease; Board on Life Sciences; Division on Earth and Life Studies and National Research Council © 2011. First Published in English by the National Academies Press. All rights reserved. This edition published under agreement with the National Academy of Sciences.

图书在版编目(CIP)数据

基因组科学的甲子“羽化”之路:从人类基因组测序到精准医学/美国科学院研究理事会编;于军等译. —北京:科学出版社,2016.3

(新生物学丛书)

ISBN 978-7-03-047326-4

I. ①基… II. ①美… ②于… III. ①人类基因-医学-普及读物 IV. ①R394.49

中国版本图书馆 CIP 数据核字(2016)第 026742 号

责任编辑:罗 静 刘 晶 / 责任校对:郑金红

责任印制:肖 兴 / 封面设计:刘新新

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

新科印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 3 月第 一 版 开本:720×1000 1/16

2016 年 3 月第一次印刷 印张:14 1/2

字数:290 000

定价:79.00 元

(如有印装质量问题,我社负责调换)

《新生物学丛书》专家委员会

主 任：蒲慕明

副 主 任：吴家睿

专家委员会成员 (按姓氏汉语拼音排序)：

昌增益	陈洛南	陈晔光	邓兴旺	高 福
韩忠朝	贺福初	黄大昉	蒋华良	金 力
康 乐	李家洋	林其谁	马克平	孟安明
裴 钢	饶 毅	饶子和	施一公	舒红兵
王 琛	王梅祥	王小宁	吴仲义	徐安龙
许智宏	薛红卫	詹启敏	张先恩	赵国屏
赵立平	钟 扬	周 琪	周忠和	朱 祯

本书翻译人员名单

主译：于 军 任鲁风 杨 宇 王绪敏 殷金龙

叶建伟 崔 丽

译者：（按姓氏汉语拼音排序）

蔡亦梅 曹丹丹 陈 艳 陈雅萍 董运梓 付令元

胡 阳 黄大维 贾善刚 李 霞 刘 侃 刘万飞

宋利璞 孙博韬 王国良 吴 刚 吴 浩 禹海英

袁丽娜 张广宇 张航晓 周媛媛

《新生物学丛书》丛书序

当前，一场新的生物学革命正在展开。为此，美国国家科学院研究理事会于2009年发布了一份战略研究报告，提出一个“新生物学”(New Biology)时代即将来临。这个“新生物学”，一方面是生物学内部各种分支学科的重组与融合，另一方面是化学、物理、信息科学、材料科学等众多非生命学科与生物学的紧密交叉与整合。

在这样一个全球生命科学发展变革的时代，我国的生命科学研究也正在高速发展，并进入了一个充满机遇和挑战的黄金期。在这个时期，将会产生许多具有影响力、推动力的科研成果。因此，有必要通过系统性集成和出版相关主题的国内外优秀图书，为后人留下一笔宝贵的“新生物学”时代精神财富。

科学出版社联合国内一批有志于推进生命科学发展的专家与学者，联合打造了一个21世纪中国生命科学的传播平台——《新生物学丛书》。希望通过这套丛书的出版，记录生命科学的进步，传递对生物技术发展的梦想。

《新生物学丛书》下设三个子系列：科学风向标，着重收集科学发展战略和态势分析报告，为科学管理者和科研人员展示科学的最新动向；科学百家园，重点收录国内外专家与学者的科研专著，为专业工作者提供新思想和新方法；科学新视窗，主要发表高级科普著作，为不同领域的研究人员和科学爱好者普及生命科学的前沿知识。

如果说科学出版社是一个“支点”，这套丛书就像一根“杠杆”，那么读者就能够借助这根“杠杆”成为撬动“地球”的人。编委会相信，不同类型的读者都能够从这套丛书中得到新的知识信息，获得思考与启迪。

《新生物学丛书》专家委员会

主任：蒲慕明

副主任：吴家睿

2012年3月

中译本序

在给再版的来自美国国家研究委员会（NRC）的两个报告——于 1988 年和 2011 年分别发表的《图解与测定人类基因组序列》和《迈向精准医疗：构建生物医学研究的知识网络和新型疾病分类法》（后简称为《迈向精准医疗》）两个报告——作序之前，我想先谈谈撰写此序的目的和将所面对的读者。在这里，我并不打算对这两份报告的内容进行概括与总结，因为它们本身既短小精悍，又言简意赅。如果读者想要快速了解报告内容，我建议从《摘要》开始，特别是它的“结论和建议”部分，因为大部分决策者都习惯这样的阅读方式。比较报告本身，此部分内容往往会吸引更多读者，尤其是有决策权的读者。因此，NCR 通常都会慎之又慎地编写这部分内容。关于读者，我觉得这本书一定能够吸引一批热衷于探索美国决策机构运行机制的科学家和决策者，借此了解生物学与医学领域中这两大关键性举措的诞生历程。

我虽已退休，却是唯一一位参加了“图解与测定人类基因组序列委员会”和“迈向精准医疗委员会”两个委员会的科研工作者。作为一个科研工作者，我的科研生涯恰巧跨越了这个科技发展史上无比辉煌的四分之一世纪。在此期间，我们测定了第一个人类基因组的 DNA 序列，为日后其在主流医学中的应用奠定了丰富的知识基础。正因为面临了这个绝好时机，我才有幸成为“图解与测定人类基因组序列委员会”的入门成员以及“迈向精准医疗委员会”的资深成员。

在此次合并出版中，我希望加入一些从内部工作人员的视角来看待这些委员会如何应运而生以及发挥功能的看法，这仅代表一个科研工作者的观点。科学政策制定过程中会涉及许多利益相关者，包括科学家、行政官员、议会议员、国家政要、特殊利益集团以及社会公众。但通常都会有很大一批活跃的科学家志愿投身于这些工作，这也是 NRC 决策过程中一个标志性的特点。而这正是现代经济政治体系中经过反复论证能有效解决科技问题的典型性程序。我认为，在政策制定初期阶段科学家的积极参与是该政策体系能够成功的最关键因素。当然，还包括一些其他的原因，如 NCR 严格的审核文化。对于 NRC 委员会成员的筛选，往往都要经过许多资深顾问数次讨论与反复磋商，在持有不同科学态度、不同专业观点，以及拥有不同科研经验的众多科学家中谨慎地做出选择。被筛选出的科学家们虽然意见不同但彼此尊重，并就某一科学问题经过激烈讨论后总能形成共识。但过分强调共识有时候又会产生一定的风险，例如，委员会会议上只能达成一系列平淡无奇的建议，好在 NRC 总是能够提出一些影响后续发展的创新性建议。希望我在此序言中所陈述的历史展望、科研趣事以及观

点看法能够给科学家和决策者应对政策挑战带来一些启示。

相较于正在从事于基因组学研究的年轻科学家而言,图解与测定人类基因组序列委员会的成立时间已经很久远了,甚至比一些人的出生时间还要早。所以,为了能够理性理解《图解与测定人类基因组序列》,我觉得有必要对当时的时代背景做些简单介绍。记得最开始研究酵母基因组的时候,我的研究工具只有限制酶 *EcoR* I 和琼脂糖凝胶电泳。我至今还记得 1974 年第一次看到酵母基因组 *EcoR* I 酶切电泳图时的情景,那是我整个科研生涯中最激动人心的时刻之一。酶切产生的成千上万条 DNA 片段组成了一条清晰连续的条带,其中至少包含了大部分单拷贝基因组 DNA。如果酵母基因组中含有上千个 *EcoR* I 位点,那么此刻每个拷贝中相应位置上的 *EcoR* I 位点都在相同的位置上。激动之余,我就想立刻开始绘制限制性酶切图谱,但当时我并不知道该如何操作。即便如此,我仍然坚信这一定是可行的。这个时候,我突然想起了英国生物化学家 John Kendrew 首次看到血红蛋白(hemoglobin)晶体 X 射线衍射图谱时的情景描述。当观察到晶体中结构相同的分子整齐排列时,他就知道一定能够计算出所有原子的精确位置。当然,这个研究花费了近 25 年的时间才得以完成^[1]。虽然酵母基因组绘图并没有那么困难,但从 1979 年开始着手到绘制出第一个相对完整的基因组图谱,我也足足花费了 10 年的时间。

当时的科研圈中已经出现了一些酵母相关研究,但非常奇怪的是我几乎没有感受到任何同业竞争的氛围,因为大家都在闷头研究自己的那一小部分基因组。唯一称得上具有竞争性的项目只有 John Sulston 和 Alan Coulson 的线虫基因组图谱绘制项目。但我们之间是友好的竞争伙伴关系。时不时地,John、Alan 和我会约在剑桥附近的一个酒吧见见面,喝点儿英国啤酒,交流交流想法。并且,我们先后于 1986 年在《国家科学研究院学报》发表了我们的第一篇研究论文^[2,3]。

至于为什么当时基因组学领域的研究那么少?原因有很多。在这里我只强调两点。首先这是由当时的分子生物学发展史所决定的。随着 1953 年 DNA 双螺旋结构的提出,分子生物学首次以一个确切的定义出现在科学界。一时间,绝大多数的科研人员都开始专注于 DNA 各项机制的研究。例如,DNA 是如何复制的?蛋白质是如何生成的?基因表达又是如何调控的?不久之后,他们就通过假说-演绎法找到了这些问题的答案。在此期间,面对一个未知的全新领域,一些较小的研究团队或者个人研究者经常会遭到各种质疑,但久经考验之后他们总是能够证实自己提出的假说。在当时的科研条件下,一些实验方法还是可行的,他们总能在数天、数周、最多几个月内找到问题的答案。碰到难题,他们会及时调整目标以继续验证自己的想法。当时确实有很多未知需要人们去探索,但时间也是非常宝贵的。这些科学家们通过广泛的咨询,制定了合理的实验体系和决定性的实验设计,最终取得了巨大的科研进展,极大推动了分子生物学的快速发展。之前没有哪一门科学是以这样的方式发展起来的,但在那段辉煌灿烂的日子里,分子生物学做到了。

随着 20 世纪 70 年代早期重组 DNA 技术的出现, 科学界迎来了一次科学技术研究大暴发。各种生物体的基因突然间变得唾手可得。新的技术手段开始逐步进入人们的视野, 如拼接技术和 RNA 编辑。人们的注意力逐渐开始从相对简单的 DNA 机制研究, 如大肠杆菌乳糖操纵子的调控机制, 转移到了更加复杂的研究项目当中。但是研究思路并没有发生改变。到处都是一片欣欣向荣的景象, 没有什么能够阻挡成功的到来, 胜利在即。虽然新兴技术异常受欢迎, 但只有分子生物学家敏锐地感觉到他们已经找到了推动科学发展的所有秘密。在当时的这种科研氛围中, 分子生物学领域还没有基因组计划发展的空间。

导致基因组学发展缓慢的第二个原因涉及哲学层面的问题。分子生物学早期阶段人们坚持的分治策略是彻底的还原论/简化论。科学工作者们理解细胞的主要方法是解剖细胞, 之后鉴定分子组成, 研究分子间作用, 最后产生功能效应。简单地认为所有功能的总和就能概括细胞和生物体的运行全貌。这就是还原论者的信条。我对此表示怀疑, 分子生物学需要解决的问题是获取更多的机械论细节, 如转录、翻译和拼接。我的怀疑源于我之前受到的化学研究历练。我认为分子生物学家普遍欠缺一种化学方面的意识。在我看来, 完全不能相信被分解零碎后的细胞还能够重新聚合起来。对我来说, 化整为零后一堆化学物质看起来更加复杂, 更加难以分析。

Jim Watson 改变了我的想法。认识 Jim 是在我参加人类基因组计划后才发生的事。之前我并不认识他, 但深知他的著作《基因的分子生物学》^[4], 并且, 这一著作在数年前已经深深影响了我的科研观念。在这本优秀作品的早期版本中, 有一个题为“A Chemist’s Look at the Bacterial Cell”的章节。当读到这章的时候, 我强烈感觉到这章就是专门为我而写的。章节以代谢基础概述开始写起, 接下来是细胞中小、中、大分子合成方式的描述, 代谢通路实例展示。最后, Watson 以一个简单的代谢图表对本章内容进行了总结。读完整章内容之后, 我满腹狐疑。Watson 究竟想如何说服我相信所描述的仅仅是他研究项目的冰山一角? 很显然, 他早就预料到了我会有这样的疑问, 正如本章的结束语所说: “It is easy for the sophisticated pure chemist to look at this metabolic chart with initial skepticism. (一位有经验的专业化学家最初看到这张代谢图表时很容易产生疑惑。)” 不出所料, 我没有感到迷惑, 但我确实开始怀疑了。Watson 继续写道: “The question arises whether this figure, by its simplification, complete misses the point of metabolism in *E. coli*. (问题来了, 这张简单的图表能否完整概括大肠杆菌的代谢网络?)” 毫无疑问, 这是非常恰当的问题, 但是他将会如何回答这个问题呢? 在题为“The Significance of a Finite Amount of DNA (特定 DNA 数量的意义)”的部分他给我做出了精彩的解答。通过对大肠杆菌基因组、大肠杆菌蛋白平均分子量以及典型代谢途径步骤进行简单的估算, Watson 估算出了大肠杆菌化学复杂度的上限, 这一估算至少到目前为止仍然是正确的。至此, 我已经目瞪口呆了, 他用这样一段话结束了此章节:

“Therefore even a cautious chemist, when properly informed, need not look at a bacterial

cell as a hopelessly complex object. Instead he might easily adopt an almost joyous enthusiasm, for it is clear that he, unlike his nineteenth-century equivalent, at last possesses the tools to describe completely the essential features of life. (因此, 谨慎的化学家们不必将细菌细胞想的过于复杂。相反, 大家都应该感到兴奋。因为相较于同处 19 世纪的其他人来说, 至少有一点很明确, 就是我们掌握了完整描述生命体本质特征的工具。)”

Watson 的话让我豁然开朗, 基因组就是所有问题的关键, 它是解释生物复杂性的唯一约束! 此时, 我已毫不犹豫地迷上了基因组学研究。

在此, 我不会详述 1979 年到 1989 年间我绘制酵母基因组限制性图谱的科研奋斗史, 因为这段历史实在是太漫长了。在那期间, 我确实做了很多工作, 而当时只有一两个技术人员从旁协助。更加不幸的是, 对于成千上万的碱基对来说, 重组 DNA 技术并不是一个很适用的工具。当然, 当时也没有生物信息学, 没有互联网, 更没有个人电脑。为方便研究, 我和当地的一个晶体绘制员做了一个交易, 可以获得他的迷你电脑的共享访问权限。但是, 我们必须在两个房屋之间的公用粮仓中接通网线, 将我的终端连接到他的电脑上。当我安装好终端后, 好多人都来围观。一个同事好奇地问我: “你要用这个做什么?” “这是一个遗传学研究所!” 在发表第一篇文章之前, 我一直从事酵母基因组绘图研究, 一做就是 7 年。而这篇文章也只是证实了我们正在进行的研究很可能是可行的, 前提是我们要坚持做下去。

其间, 卓有远见的业界领袖人物 Charles DeLisi 和 Robert Sinsheimer 提出了一个足以轰动整个研究界的疯狂想法——人类基因组测序。而当时的我只是华盛顿大学圣路易斯分校的普通教师, 可以说我们之间完全没有任何交集。所以, 对于发生在 1985 年和 1986 年间的那场批判性讨论^[5,6], 我并没有什么要添加和补充的。他们最终成立了 NRC 图解与测定人类基因组序列委员会。

NRC 是美国国家科学院、工程院和国家医学院的一个业务部门, 其职责是就国家利益相关事务做出高科技含量的专业报告。作为程序繁琐的官僚机构, NRC 运行效率低, 是典型的保守派, 但其出具的各项报告却具有很高的权威性。之所以如此, 是因为 NRC 具有一个非常强大的员工队伍, 他们会雇佣一些优秀的委员会成员, 一经任命, 这些忠于委员会自治的成员便会无偿递交一些严谨、经过层层审核的高质量报告。在人类基因组研究项目中, NRC 聘请了 Bruce Alberts 作为委员会主席。除此之外, 还任命了 Jim Watson、Sydney Brenner、Lee Hood、Dan Nathans、Wally Gilbert 等, 在此我就不详细罗列委员会成员名单了。1987 年, 为了调查当下还有哪些基因组相关研究, 委员会发现了我在圣路易斯分校的这个项目, 并邀请我于 1 月份向委员会做一个项目展示。这让我很吃惊, 因为之前从没见过这些牛人, 我对他们充满了敬畏。即便如此, 我还是尽全力做了最好的陈述, 并强调了重组 DNA 技术正在经历改进与完善, 同时也指出了目前的科技水平与人类基因组测序这一美好愿景之间仍然存在很大的差距。

之后一个月，委员会成员之一 Wally Gilbert 决定去一家基因测序相关的私营机构任职，由于存在利益冲突，不得不辞去了委员会职务。不久，我便接到了 NRC 项目负责人 John Burris 的电话，问我是否愿意接替 Wally Gilbert 的职务。这与我在这个大名鼎鼎的委员会上展示一些幻灯片，讲完就可以飞回家完全不是一回事，这次我真是碰到了一个大难题。委员会需要我加入 NCR 的这个撰写组，并参与写出一个划时代的科技政策文件。久经思虑，虽然心存担忧，但我还是接受了。

在 1987 年举办的几次会议中，我们逐渐认识了彼此。在 Bruce Alberts 的杰出领导之下，整个团队所洋溢出的诚挚友情深深打动了我。与我专业最相近的同行伙伴是 Shirley Tilghman，她后来做了美国普林斯顿大学备受尊重的校长。虽然 Shirley 和我们都处在相似的事业阶段，但她比我更有名望，在资深委员会成员中更受欢迎。幸好这一切都不影响我们之间的关系，我们会经常分享彼此的科研经验。在一次茶歇时，大家在谈到谁曾经测序过 1000 个碱基对以上的 DNA 序列这个话题时产生了争论，毫无疑问我们俩都做过。当时，David Botstein 声称他也测过，我和 Shirley Tilghman 都对此表示怀疑。David 拿出的证据是当时已经发表的酵母 *URA3* 基因，这个基因长度虽然仅有 1170bp，并且文章中除他之外还有两个作者。不过，David 所说确实是真的。这只是我们工作中的一个小插曲。我们三个的小争执确实为整个议题注入了一剂清醒剂。因为我们正在讨论的内容就是该如何测序未经绘图的数以亿计的碱基对。而当时整个委员会所有成员加起来的经验也仅仅局限于对生物材料的数千个碱基对绘制图谱而已。而且，即使是这些经验还是以数不清的汗水、泪水和暴露于 ^{32}P 发射出的 β 粒子为代价才换来的。额外加一句，基于荧光原理的技术仍然具有远大的前景。

费尽周折，报告草案终于完成了。最开始的时候，一些委员会成员对“图解与测定人类基因组序列计划”的提议充满了怀疑。但现在大家的观念改变了，人们开始担心如此大的研究项目将会耗光大部分传统研究的科研经费，还有一些成员不喜欢将“big science”定位到生物学领域。此外，还有人在担心这个项目的可行性，即便大量事实已经证明了项目的可行性。

最终，大家也没有就此计划达成一个统一的观点。不过，我想指出几个在此过程中大家总结出来的发挥重要作用的观点。

(1) 慎用措辞。委员会中的一些资深成员，比如 Dan Nathans，认识到我们现在好比正在进行一场外交谈判，而在国际外交中措辞是非常重要的。例如，出现在《摘要》的第一个建议中的“special effort”这个词汇就是非常关键。在“merits a special effort that should be organized and funded specifically for this purpose（为达到获取人类基因组图谱、序列及更多认识这一目的，委员会应该在组织与资金方面做出巨大的努力）”这句话中我们没有使用“big project”或者“crash program”，而是“special effort”。在这里使用这个词汇就会给外界释放出这样一种信号：我们不认为“图解与测定人类基因组序列计划”是一个偶然性的普通研究课程，我们也不是在寻求对传统研究的突破。

试想谁能反对这样一种“special effort”!

(2) 避免炒作。炒作这一行为很容易成为评论家攻击的目标,其预期效果也许并非一些科学家所设想的那么好。同样是在第一个建议中,我们用这样一句话来表述我们的承诺:“a special effort in the next two decades will greatly enhance progress in human biology and medicine. (未来二十年间的巨大努力将极大地推进人类生物学和医学的发展与进步。)”换言之,我们并没有承诺一定可以治愈癌症,或者一定可以开发出许多新的药物,甚至彻底变革医疗卫生体系,而是通过这些项目数据在很大程度上推进人类生物学和医学的发展与进步。

(3) 在项目难易程度的衡量方面必须找到平衡点。在报告实施过程中,委员会关于项目难或易的观点总是反复出现。如果说这个项目很难,那就意味着我们所作出的努力都是不成熟的;如果说这个项目很容易实现,那就说明我们并不需要作出什么努力,顺其自然发展就好。在这场辩论中我起到了积极的作用,因为我比其他委员会成员更加明白当前技术的优缺点。仅仅依靠大力改进 1987 年的重组 DNA 技术来开展这个测序项目是注定要失败的,并且,目前有一些更加先进的技术手段已经初露端倪。因此,我对大规模数据采集将带来大量的技术改进保持乐观的态度。之后,在与 John Burris 的通话过程中,我向他陈述了我的观点。当然,在那个年代还没有电子邮件,所以我只能电话告知。当时的交流细节将在《摘要》的第二个建议中逐字呈现:“Although the needed capabilities do not yet exist, the broad outlines of how they could be developed are clear. Prospects are therefore good that the required advanced DNA technologies would emerge from a focused effort that emphasizes pilot projects and technological development. (虽然目前我们还不具备这样的能力,但如何开发这种技术的框架却是显而易见的。因此,乐观的坚信我们所需的先进 DNA 技术将会从大量的试验项目和技术改进中出现。)”

(4) 关于模式生物。这是一个全新的政策主张,因为之前的讨论焦点全部集中在人类基因组。当读到下面这段描述时,大家都顿住了:“To succeed, ... this project must not be restricted to the human genome; rather it must include an extensive sequence analysis of the genomes of selected other species. (为了取得成功,.....项目不能仅仅局限于人类基因组,还必须针对其他物种的基因组开展大量的序列分析。)”随后,我们重点列出了一些模式生物名单,包括细菌、酵母、果蝇和秀丽隐杆线虫。我非常赞成这种做法,因为我的研究课题就是依托酵母展开的,并且深知生物学研究的生机与活力与特定生物体的研究唇齿相依。这些生物体具有独特的价值体系和强大的社会网络,它们所衍生出来的技术、观念以及科研问题对于所有生物体研究具有典型代表性。如果基因组学研究有利于人体生物学发展,那么它一定有益于模式生物研究。而且,模式生物基因组大小还不到人类基因组的十分之一,测序这些生物体所带来的额外风险并不会影响到“图解与测定人类基因组序列计划”的成败,反而会大大增加科学界对此计划的支持,并将吸引更多的顶级科学家加入到这个研究队伍。就比如 Bob

Waterston 和 John Sulston, 如果没有之前对他们所挚爱的线虫基因组进行测序这段宝贵的科研经历, 他们很可能就不会提出“图解与测定人类基因组序列”这一主张。关于模式生物更为“精妙”的一点是, 虽然我们强烈推荐列入模式生物基因组研究, 但仍然将此报告命名为“图解与测定人类基因组序列”。就这样, 模式生物测序成了《图解与测定人类基因组序列》的重要内容, 而不是报告的附加资料, 供我们偶尔翻翻。一些模式生物研究界的同事没有敏感的察觉出这一“精妙”的策略, 还在抱怨委员会目光短浅, 没有政策远见。即便如此, 他们最终还是从中获益颇丰。在“图解与测定人类基因组序列计划”实施之前, 很难向决策机构解释清楚研究酵母的重要性; 但在此之后, 酵母研究和医学研究之间的紧密联系已经不言而喻。在此需要申明一点, 并不是说酵母研究不能独立证明自己的重要性, 而是在对有限资源不可避免的竞争局面中很难解释清楚。

(5) 正确的时间期限和预算支出。上文中我已经讲述了这个计划是如何开始的, 但还远远不够^[7]。Watson 通过基础设施、劳动力、设备以及其他物资的预计成本估算出“图解与测定人类基因组序列计划”的年度预算为 2 亿美元。以 15 年为期, 完成该项目一共需要 30 亿美元, 折合每个碱基对的预算成本为 1 美元。其实, 我们并不知道测序一个碱基对的实际成本, 因为我们指望用来测序的技术手段还没有开发出来, 也许只有 Jim Watson 或许知道决定最终成本的细节因素。关于项目时间, 当时有两种声音, 一方建议 10 年, 一方建议 20 年, Jim Watson 提议折中为 15 年。他说, 在 10 年的时间内项目可能才刚刚开始, 到时候评论家们会说我们执行进度落后于预定计划; 如果定为 20 年的话, 应该没有哪个机构会资助持续时间这么长的项目。2 亿美元的年度预算则与美国政策体系有关, 我仍然记得当 Watson 作出以下声明时, 现场气氛陡然变得沉默: “The budget should be \$200 M/yr; anything bigger would be a fat target during annual budget reviews, while with anything smaller there will not be anything in Illinois. (项目年度预算为 2 亿美元! 在年度预算评审过程中, 所报预算太多的话很容易就会成为大家非议的靶标, 从而影响项目预算的批准; 但如果预算太少又什么也做不了。)”至于伊利诺斯州与人类基因组测序成本预算有何相关性? Watson 再次提出了颇具前瞻性的想法, 如果“图解与测定人类基因组序列计划”落地美国, 那么它将需要得到全国各地广泛的政策支持。到时候, 美国约三分之一的州都会堂而皇之地竞争如此大规模项目的主办权, 包括伊利诺斯州。由于一些较大的州往往具有许多综合学科研究中心, 它们往往会占据很大的资助份额。如果年度预算少于 2 亿美元, 许多其他的研究活动将由于得不到足够的资助而不能顺利进行。如果项目仅仅局限于某一地点, 如加利福尼亚、马萨诸塞州、纽约等, 相较于分散存在于多个州, 政策扶持力度会减小很多。

有些人会觉得上述观点过于强调科学政策制定过程中的外交策略和政治维度的重要性, 但考虑到 19 世纪 80 年代后期分子生物学的发展现状, 《图解与测定人类基

因组序列》这一报告确实为解释人类基因组测序的基本原理和方法步骤提供了可靠的基础支撑。而这些对于一份报告的公信度而言是至关重要的，同时，这也是 NRC 审议工作的一个重要特点。相较于技术细节的描述，整个报告信息的影响力往往更加广泛，而决定整体信息的关键往往就取决于外交和政治层面的细微差别之中。

整个事件中我们最显著的不足体现在对于知识产权的处理方面。我们低估了基因专利问题对于整个基因组学科学文化所带来的潜在威胁。在题为“社会影响”这一章的最后几页中，我们用一段话就这一主题进行了专门讨论，并提出了一些问题。我们建议应该设立独立的组织对专利事宜进行及时的研究。我们认为基因组序列是一种公共资源，不应该设立版权保护。我们将太多的注意力集中在了“版权”进而忽视了“专利”，更为严重的是，我们从此再也没有就这一问题展开讨论，以至于我们在报告中的简短声明对后续事项没有发挥任何作用。NRC 能够很好地解决科学、技术和狭义的政策问题，但对于一些存在利益纠纷的根深蒂固的社会问题，可能就无能为力了。那些社会问题属于政治层面的分内事儿，不是政治意识较为薄弱的科学家可以正确处理的。这里需要解释一下，关于基因组学的知识产权问题在美国可以说是不堪重负。美国专利局坚持一个主张，联邦机构例如国立卫生研究院却赞成相悖的政策，生物科技产业强烈赞成基因专利，而一些制药公司则更倾向于共享成果。最终的结果只能是产生了大量的基因序列专利，但几乎没有几个能带来价值。2013 年，美国最高法院毫无异议地拒绝了人类基因的可专利性，这距离《图解与测定人类基因组序列》报告发布已经整整 25 年。许多科学家包括我在内都非常赞成这一判决，但总觉得有些不妥。所以，不久的将来这个问题可能会再次浮出水面。但至少到目前为止，基因组序列、cDNA 序列、基因注释，以及其他具有增加值的派生物仍然是一种公共资源，就和 NRC 委员会一贯坚持的一样。

总体而言，《图解与测定人类基因组序列》是一个闪亮的成功之举。NRC 一年内就发行了数百份报告。虽然很多人并没有真正读完它，它也没有带来明显的政策改变，但确实促成了后续事情的发展：“人类基因组计划”得以按照既定的方式逐步形成，维持美国政府的政策支持，并抵挡私营机构的伺机挑战，获得了国际合作伙伴的援助，最终实现了目标。《图解与测定人类基因组序列》为所发生的这一切提供了可行的、高质量的工作思路，助力“人类基因组计划”继续其历史使命。

2011 年，距离《图解与测定人类基因组序列》报告发布 23 年后，NRC 发行了《迈向精准医疗：构建生物医学研究的知识网络和新型疾病分类法》。从基本方面来讲，该报告是《图解与测定人类基因组序列》的续篇，有文为证：人类基因组序列会大力推进人类生物学和医学的发展进程。到 2011 年，基因序列研究已经整整十年了。期间，重测序技术得到了暴发式地改进和完善，人们对于如何利用个人基因组信息提高医疗服务的思路逐渐变得更为清晰。我觉得是时候来实现当初的预言了。

我所说的关于《图解与测定人类基因组序列》的评论已经在别处出版过了，也已

经被别人证实过了，至少很多科学家都熟知此事。但对于《迈向精准医疗》的评论还比较少。这份报告本身已经引发了很大的关注，尤其是在 2015 年 Obama 总统宣布了美国精准医疗计划“Precision Medicine”之后。当然了，Obama 所宣布的“Precision Medicine”以及其他词汇均出于 NRC 精准医学报告。从实质层面来讲，精准医学计划尚未启动，实际发展是否符合 NRC 报告中的建议尚待分晓。

关于撰写《迈向精准医疗》的最初想法是于 2009 年 10 月下旬在我和 David Walt、Alan Williamson 的反复讨论中逐步成型的。当时，在加利福尼亚州（California）圣地亚哥（San Diego）Illumina 科学咨询委员会会议上，我们意识到，数年后个人基因组大规模测序将会成为现实，但是目前并没有明确的计划来推动这一即将改善医疗现状的重大科技革命。我们有足够的兴趣和热情，但没有明确可行的计划。这一现状使我想起了 1987 年基因组学发展所面临的困境。那时候，逐渐有越来越多的科学家开始系统性分析细胞基因组，雄心勃勃地追求绘图与测序事业。但是，几乎所有的项目都是小规模研究，并且受地方利益驱动，所以很难发展壮大。David、Alan 和我就如何保护个人基因组测序、酵母基因组测序项目展开讨论。

Alan 具有丰富的制药行业相关经验，并且长期担任国家人类基因组研究所（the National Human Genome Research Institute）顾问，她提议确立一个贯穿整体的主题能够推动疾病分类法取得全新的发展。在一些领域，基因组测序技术已经能够对曾经混为一谈的疾病进行精确的区分，癌症就是一个最明显的例子。因为癌症的发展是由体细胞突变所驱动的。如果有足够多的患者可以接受 DNA 序列和其他辅助分析（例如，基因表达谱和表观遗传改变），那么对于其他疾病来说，也许疾病分类法同样可以取得重大改进。Alan 认为疾病分类法应扎根于分子病理学，基因组测序技术能够为实现主流医学改善健康格局提供最具潜力的方法。

我和 David Walt 非常认同 Alan 的观点，并马上开始讨论该如何实施。Alan 打算向 NIH 提议资助试点项目，但我认为这个想法有点儿不切实际。根据我任职于“图解与测定人类基因组序列委员会”和 2005 年 *Mathematics and 21st Century Biology* 委员会主席的经验来看，新的分类法可以从 NRC 开始着手研究。这样做最大的优势在于：NRC 可以对新的分类法给出明确定义，并提供技术路线图，推动联邦机构展开后续讨论，而对于一些重要决策的实施最终都是由联邦机构做出的。这一观点得到了 David 和 Alan 的赞同。于是我们商量着联系时任 NIH 主任的 Francis Collins，因为我们对他都非常了解，并就编撰 NRC 报告以探索大规模分子数据的潜在应用进而开发出新的疾病分类法这个问题向他咨询获得 NIH 资助的可能性。

在接下来的几个月中，为达成上述目标，我们与 Collins 及其员工进行了频繁的电子邮件和电话会议沟通。虽然他对此研究能否顺利转交给 NRC 还存在一些担忧，但他最终还是答应了支持我们的项目。通常情况下，NRC 研究费用都比较高，标准的预算范围在 50~100 万美元。相比较而言，NIH 主任则拥有充足的自有资金，可以对

符合机构利益的诸如此类研究提供及时的资助。经过不懈努力，在 2009 年年末 2010 年年初，NIH 和 NRC 终于就我们的研究项目展开了讨论。虽然我没有参加此次谈论，但很明显事情进展顺利。2010 年 3 月，我们得到了 NIH 和 NRC 准备启动研究项目的及时反馈。10 月，委员会招募工作正式启动。期间，NRC 成员问我是否愿意加入委员会，我毫不犹豫地答应了。12 月中旬我们便召开了第一次委员会会议。

在论述迈向精准医疗委员会的工作之前，我想就同时身兼 Illumina 科学咨询委员会（SAB）委员与精准医学委员会成员之间存在的显著冲突发表一下我的看法。作为全球领先的 DNA 测序仪及配套试剂制造商，Illumina 毫无疑问会从如此重要的联邦基金项目中获益，进一步扩大 DNA 测序技术在医学领域的应用。首先，在我看来，二者之间的冲突其实并没有那么严重。我和 Illumina 之间的合同一年一签，Illumina 会支付我数千美元作为参加其 SAB 年度会议的报酬，我并没有实际经营管理责任，也不持有其公司股票。虽然拥有一小部分产权信息，但也被一纸《保密协议》紧紧约束。其次，美国科学界通常会充分披露利益冲突的处理方式。当时是由诸如高等院校、国立卫生研究院、NRC 之类的机构来负责解决相关事宜。他们能够评判这个利益冲突是否合理，能够决定是否需要取消科学家的顾问资格。一开始我就向 Francis Collins 和 NCR 说明了我担任 Illumina SAB 顾问一事。Collins 只是稍微犹豫了一下，而 NRC 经过全面审查后认为这个冲突其实无关紧要。但是，在第一次委员会会议上，我还是当场表明了这个问题，以确保所有的委员会成员都能作出他们的判断。在美国，这些问题必须严肃对待。因为大家都意识到公众的信任是我们最宝贵的财富。冲突本身是普遍存在的，处理不当就会严重损害这份信任。这其实也是一件好事。将新兴科学领域的专家隔离成缺乏沟通与交流的小团体不符合任何人的利益，所以，高等院校、研究机构、政府和私营机构要密切合作，以推进科学进步，进而实现公共利益。

2010 年年底到 2011 年上半年，精准医学委员会共召开了四次会议。2011 年春天，我和委员会其他成员将大量的时间投入到了报告的写作当中。期间，我们进行了多次电话会议和电子邮件交流，但没有举办碰头会。我们有职业作家顾问，但仅限于报告草案的文体一致性，最终版的报告几乎全部由委员会成员完成的。

委员会审议有一个显著的特征：在审议过程中逐步扩大授权范围。报告附录部分的《任务声明》中并没有讲述任何有关精准医学的内容，甚至连医学相关内容都没有涉及，因为我们得到的授权是“explore the feasibility and need, and develop a potential framework for creating a ‘New Taxonomy’ of human diseases based on molecular biology. (探索精准医学的可行性与市场需求，搭建一种全新的基于分子生物学的人类疾病分类法的可行性框架。)”这一措辞色调与 Alan Williamson 的初始观念非常贴切。NRC 报告中的《任务声明》是经过报告发起人、NIH 和 NRC 三方反复磋商而确定的措辞极其谨慎的文件。一经确定，不可更改。NRC 成员有义务确保委员会始终坚持这一决定。所以，我们自始至终密切关注着 NRC 的审议过程，迫不及待地等待着审议结果。

我们没有采用分类法作为报告主题，主要原因在于分子数据能够理性影响主流医学发展之前，科学研究和临床应用是协同发展的。我们想要推荐的是一个进程，而不是项目。进程是对人类基因型、生物标记和表型研究进行循序渐进逐步改进的研究方式，其目的如《迈向精准医学》中图 1-3 所示。对于这幅图片的修改和完善成了委员会讨论最后阶段的核心问题，因此不断更新的“信息共享中心”也成为了图解最重要的特征。听着似乎显而易见，但仔细审查就会发现其中蕴含着一些激进的色彩。为什么？接下来我会强调一些原因：

(1) “信息共享中心”必须是一个公共资源。这也许不符合委员会的愿景，但任何人都会赞同普遍数据访问的权限。我们又回到了《图解与测定人类基因组序列》的老话题——公众信任。但现实是，国家利益、机构利益、商业利益和学术利益都在妨碍着这一公共资源的建立。如果不能果断地说服决策者，大家如诸侯分据般各自为政，那么，谁的利益都无法保证。

(2) “信息共享中心”必须包含丰富的个体表型资料。对于精准医学来讲，最重要的表型资料就是影响健康状况的信息。在谈论到数百万人的健康数据时，我们唯一来源是医疗卫生体系。但是目前，即使有也是极少数的国家会组织收集这些数据。美国卫生体系就非常抵触这一做法。

(3) “信息共享中心”必须不断更新。人们的健康状况不断的发生改变，看护标准也在随时改变，分子生物学技术同样在不断改变。我们的主要兴趣在于医疗干预后的健康结果。对于老龄化人群来说，慢性病是几年或几十年来疾病积累的表现形式。只有卫生保健体系才能密切观察追踪病患病情，使其安然度过这一时期。在到处充斥着移动电子产品的现代社会，我们在追踪个体患者健康状况方面做得并不到位，如在一年又一年的疫苗接种记录方面。

(4) “信息共享中心”必须能够索引到患者个人。这一点看起来很好实现，其实不然。我有权参与精准医学研究，并乐于此事，但目前为止，我还没有办法对信息共享中心某个病患的分子和表型数据进行综合分析。如果我的想法得以实现，患者会受益匪浅。但一些既得利益者并不希望改变这一现状，如研究人员、卫生保健机构、政府官员等。在这些强大的利益集团面前，患者毫无胜算。

虽然，人类基因组计划顺利进行并实现了目标，人们都在热烈庆祝这一伟大的成功。但精准医学计划不适于采取同样的方式。其目标在于医疗服务能够更加高效地利用分子数据，用较低的成本获得更好的健康服务。但这不是一劳永逸的事情，一次成功只能带来对更多成功的迫切需求。因此，迈向精准医疗委员会建议创造一个开放式、没有边界的研究项目。《迈向精准医疗》即是这样一份充满愿景性质的报告。

在讨论的最后阶段，我们纠结于该如何为这份报告命名。“Taxonomy”的范围似乎过于宽泛，过于强调技术性。而且，我们要做的是一个进程性报告，而不是项目导向型报告，改进诊断技术仅仅是报告内容的一部分而已。我们所讲述的是如何将基础