

THE HUMAN GENOME

人类基因组

我们的DNA

C. 丹尼斯 R. 加拉格尔 编
J. D. 沃森 序

林侠 李彦 张秀清 译
张猛 包静月 等



科学出版社

www.sciencepress.com

人

THE HUMAN GENOME

类基因组

我们的DNA

C. 丹尼斯 R. 加拉格尔 编
J. D. 沃森 序

林 侠 李 彦 张秀清 译
张 猛 包静月 等



科学出版社
北京



0749400

内 容 简 介

人类基因组测序被公认为当代最具突破性的科学事件。它将有助于我们洞察自身的历史和个体。

本书引人入胜，装帧精美，内容贴近于社会公众。本书从对生物学和基因组测序技术的基础简介开始，以历史性的基因组序列研究论文结尾。书中记述了主要人类基因组计划的参与者和重要事件年表及关于该项工作的影响的各种观点，囊括了全球媒体的相关报道，及对人类基因组测序的伦理、法律和社会意义的评估。

本书是理解人类基因组的必备指南。

The Human Genome
Carina Dennis, Richard Gallagher
First published 2001 by
PALGRAVE

©Nature Publishing Group 2001

中文简体字翻译版权受国际国内法律保护。
限售于不包括香港的中国大陆地区。

图书在版编目(CIP)数据

人类基因组——我们的DNA/丹尼斯(Dennis, C.)等编著,林侠等译校.
—北京:科学出版社,2003.4
ISBN 7-03-010570-2

I.人… II.①丹… ②林… III.人类基因—基因组—研究—纪录 IV.Q987

中国版本图书馆CIP数据核字(2002)第050498号

责任编辑:马学海 盖宇 / 责任校对:曹锐军
责任印制:刘士平 / 封面设计:王浩

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>

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2003年4月第一版 开本:889×1194 1/16
2003年4月第一次印刷 印张:15
印数:1—10 000 字数:499 000

定价:48.00元

(如有印装质量问题,我社负责调换〈科印〉)

译者名单

(按姓氏笔画排序)

王劲强	邓亚军	邓素华
包其郁	包静月	李刚
李杰	李彦	李英华
李斐斐	刘红艳	刘建平
陈艳炯	陈聪	吴娴婷
张秀清	张猛	张清瑞
林侠	杨金娥	夏红爱
唐家斌	阎春霞	曾嵘

致谢

我们感谢 Barbara Cohen, Francis Collins, Elizabeth Dennis, Daniel Drell, Mark Guyer, Mark Hirst, Michael Hopkin, Ian Jones, Michael Kilborn, John MacFarlane, Joseph McInerney, Richard Nathan, Hemai Parthasarathy 和 Peter Wrobel 对本书编写的宝贵投入、建设性建议和广泛支持。我们还感谢 Jane Ades, Barbara Izdebska, Betty Mansfield, Majo Xeridat, 国家人类基因组研究所 (National Human Genome Research Institute, NHGRI), 能源部 (Department of Energy, DOE) 和 Wellcome Trust 基金会协助增设图文。

注

4~7 页的文本也见于 James D. Watson 著《激情 DNA: 基因、基因组和社会 (*A Passion for DNA: Genes, Genomes, and Society*)》，冷泉港实验室出版社 2001 年版。

目 录

序: 人类基因组揭密(J. D. 沃森).....	4
《自然(Nature)》杂志编辑评论	7
前言	8



1 人类基因组入门

人类基因组鸟瞰	9
聚焦基因组	11
DNA 的详细说明	13
生命的蓝图	13
DNA: 一个标志	14
从基因到蛋白质	18
个人的基因组	20
遗传变异与疾病	21



2 DNA测序的艺术

基因组测序概观	23
DNA 测序的方法	23
自动化——快, 更快, 再快	24
分子生物学家的工具箱	26
组装基因组的拼图游戏	27
图谱的构建	29
人类基因组计划	30
人类基因组组装	33
Celera Genomics 公司	33
基因组画廊	33
序列解读	35
人类基因组计划年表	36



3 我们基因组中的闪光点

基因在哪里?	38
基因组地形图	40
重复序列	43
围攻 Y 染色体	43
细菌的足迹	44
追踪疾病基因	45
合作、竞争和结论	45



4 新闻媒体报道的人类基因组

抓住概念	46
基因数目比预测的要少	47
消除种族歧视	48
对健康和社会的影响	48



5 序列与社会

面对遗传风险	50
ELSI 计划	51
儿童的遗传测试	52
遗传歧视和隐私	53
遗传学咨询的重要性	53
DNA 的“辩护”	54
DNA 档案	54
基因治疗	55
遗传改良	56
谁“占有”我们的基因?	57
了解我们的遗传本质	58
教育方面的呼吁	58



6 基因组内外

他山之石	59
探明功能	59
蛋白质特性研究	60
整合信息	60



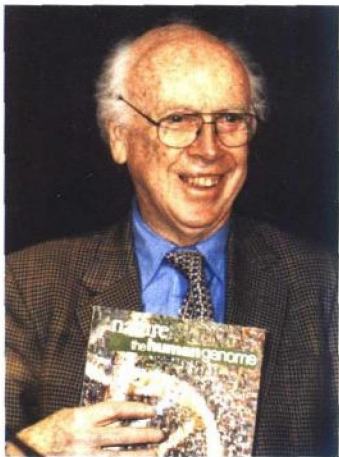
7 研究论文

《自然》新闻和观点选编 (英文)	62
人类基因组初步测序和分析(英文) ..	72
《自然》新闻和观点选编(译文)	135
人类基因组初步测序和分析(译文) .	145

基因组说文解字	201
主题索引	203
人名索引	215

人类基因组揭密

J. D. 沃森



National Human Genome Research Institute, NIH

读到“人类基因组计划”所绘制的序列草图，我感到极大的欣慰与满足。这张图的30亿个碱基序列，是严格地按照美国国家科学院（National Academy of Sciences, NAS）人类基因组制图和测序委员会十年前制订的路线而确定的。Bruce Alberts（现任国家科学院院长）是这个委员会的主席，而我以及其他13个人是它的成员。我们在1988年的那份报告中预计完成这项计划需要至少15年的时间和约30亿美元的资金，今天看来其精确性超出我们当时最大胆的估计。现在看来，完成后续的补洞和纠错工作还需要两年，到那时，我们就可以拿到这份几乎没什么错误的现代人类赖以存在的遗传学蓝图了。

当我和Francis Crick在1953年确立DNA(deoxyribonucleic acid, 脱氧核糖核酸)双螺旋结构时，不论我还是他都无论如何想不到能在我们的有生之年里看到这张人类的基因图。那时候，我们觉得明白细胞是如何从这张图中读取遗传指令的就已经是一种奢求了。然而，令人欣慰的是，从1966年我们知道DNA三联体密码如何确定生命舞台的主角——蛋白质的氨基酸序列以来，这方面的研究进展比我们预期的要快很多；而随着1973年Stanley Cohen和Herb Boyer将DNA重组技术推入研究热点后，这方面的研究速度就更加令人咋舌了，对基因的克隆和操作因而由梦想变成了科学生活中的现实。在如此令人激动的浪潮鼓舞下，Fred Sanger和Walter Gilbert分别创造性地发展出了沿着DNA分子逐个确定碱基顺序的强有力的方法和途径，而这正表明了人类也可以像细胞一样去读取基因的信息，也从此打开了通往确定全部遗传学指令的大门，为解析任何生物的全部遗传密码也就是基因组序列开辟了道路，

当然这也要取决于资金、人力的和技术因素。

从只有几千个碱基组成的病毒基因组入手，这项艰巨的工作启动了。在20世纪80年代早期，已有含量超过100 000个碱基的病毒基因组被测序，而含量超过一百万个碱基的细菌基因组也已经可以操作了。这些基因组工作的完成可以帮助我们了解大量不同类型的蛋白质与细菌生存的相互关系。虽然那个时候我个人仍然认为几十亿个碱基长度的人类基因组对于测序而言是太长了些。但不久我就已经转变成了国际人类基因组计划强有力的支持者，坚信大规模制图和测序产生的雄厚资源将会加快重大疾病遗传机理的发现。

国家科学院委员会只用了很短的时间来确定是否要进行人类基因组计划，而实际上我们的精力更集中在如何组织这项计划以及筹集资金的问题上。当时看起来似乎最佳的办法是让一切先启动起来，再逐渐增大测序的力度，希望最终找到低耗费的测序方法后，再全力开展这项工作。基于人类与类似面包酵母或者果蝇这些模式生物在基因上有许多同源性的考虑，委员会全体成员都一致认为第一个大的基因组测序工作不从人的基因组而是从一些遗传学上经典的模式生物开始。

我们提出了一项15年工作计划，因为我们当时认为这些项目做准备的起始工作也是整体计划的一部分。Richard Gibbs, Eric Lander, Maynard Olson, John Sulston, Bob Waterson和Jean Weissenbach都坚持这条路线，纷纷致力于主持大型百万级碱基测序开发实验室的工作。委员会成员中，不在科研岗位上的只有三年前因白血病去世的70岁的Dan Nathans，Dan的离开是令人极为悲痛的。在委员会的商议中，没有人提出比15年更短的时间框架，因为需要革新和

研究的东西实在太多了。我只是在后来才了解到，国会更愿意批准为期在十年以内的大型项目，因为这样最初的支持者们还能够参加最后的庆功会。然而幸运的是，来自衣阿华州的民主党参议员 Tom Harkin 先生，又成为了三届任期的国会议员，那么，他就能和新墨西哥州的共和党参议员 Pete Domenici 先生一起看到人类基因组计划完成的整个过程。

人类基因组计划顺利完成的所有技术革新都按照计划进行了。其中大部分都是对以前技术和方法的修改，这与获取诺贝尔奖金所需要的技术上的创新和飞跃不同。目前的 DNA 测序仪和我们的大型测序实验室的工作量大约是最初 Mike Hunkapiller 和 Lloyd Smith 在 Lee Hood 的 Caltech 实验室的原始测序仪工作量的 1000 倍以上。而且，现在的计算机和相关运用软件对于获得的 DNA 原始数据分析的能力也是人类基因组计划开始时的 1000 倍以上。

完成人类基因组正确拼接的主要障碍是大量重复 DNA 的存在（重复约为 50%），因此人类基因组实验室决定先从已知的染色体位点开始测序。这种基于图谱的方案在 1998 年 5 月时受到了由 Craig Venter 领导的 Celera 公司的异议和挑战。Celera 提出了另一种可选的方案：将基因组随机分解成片段并测序，然后以一种简单的而不是建图的方式重新组装起来，这就是著名的“全基因组霰弹法”。这种方法的实施关键在于 300 台新的高容量的毛细管测序仪的使用，这种机器很快就上市了，而且配套的供高容量计算机使用的霰弹法拼接软件也相应地上市了。有了这样的装备条件，Celera 才可能仅用两年的时间做出人类基因组的第一张草图。

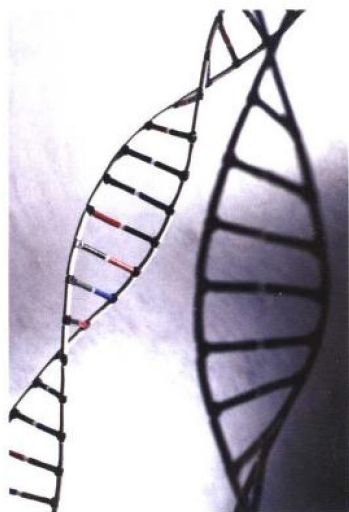
我是在与我的前任助手 Richard Roberts 打电话时第一次听说 Celera 的，Rich 告诉我 Celera 要把国际联盟赶出基因组计划并邀请我以科学顾问的身份加盟。由于考虑到可以从 Celera 的计划里了解到更多的东西以备即将举行的春季基因组会议，于是，我向国家卫生研究院（National Institutes of Health, NIH）和 Wellcome Trust 基金会打了报告，说明 Celera 想让

他们出局的想法。随后的那个礼拜里，Craig Venter 访问了国家卫生研究院并告诉 Harold Varmus 和 Francis Collins，他认为人类基因组计划对未来所做的努力更适合用到大鼠的测序中去。

从与 Rich Roberts 打电话时，我就认为一家私营公司仅通过一些关键性技术就能对整个人类基因组计划进行有效地掌握和控制的事实是不可信的。因为，这场在基因层面的交锋必然要与其巨额的资金费用相匹敌才行。因此，当看到 Wellcome Trust 基金会对此的反应是给 Sanger 中心追加了一倍的拨款时，我感到轻松了许多。尽管这两种方案的优点有待于进一步验证，但是 Celera 的“超级霰弹法”显然引起了一系列新闻媒体的关注，而他们报道说人类基因组计划没有按照预定的路线进行。实际上，在 1996 年春季的百慕大会议上，人类基因组计划的领导者们讨论了 Jim Weber 的提案，提案是关于用低要求的全基因组霰弹法取代高质量的作图方案。同时，Phil Green 的即会计算（后来经过整理后发表了）证明人 DNA 的重复性太高，单纯的霰弹方案并不能完全覆盖整个基因组。

1998 年九月，我返回华盛顿，为了说服主要的高层国会人员扩大联邦政府公共基金对测序的支持，因为这对于防止人类遗传信息的垄断是十分必要的。许多大型制药公司对人类基因组计划的鼎力支持也是出于这样的考虑，即认为 Celera 的数据只有通过与公布序列的比对后才能有效的观点。我同样高兴地看到，国会极大地提高了对测序资金的公共资助。受此鼓舞，人类基因组计划也像 Celera 那样宣布将在 2000 年春季完成一个人类基因组的草图，但是这个图与 Celera 的图不同，它将力图完成一张高精度图。

2001 年 2 月由人类基因组计划和 Celera 共同公布的人类基因组草图毫无疑问是人类历史上的一块里程碑，它揭示了人类遗传学图谱的基本特点。这个图保证了我们可以从中鉴定绝大多数与人类生存和生活相关的基因。然后，再通过遗传密码将这些信息翻译成蛋白质，我们就能够获得对于组成我们身体的大分子



物质的完整、深刻的认识和理解。同时，我们也意识到这些分子基石与构成其他形式生物体的基石是相似的。达尔文进化论可以在遗传学图谱这一增加的信息基础上得到进一步的阐释。

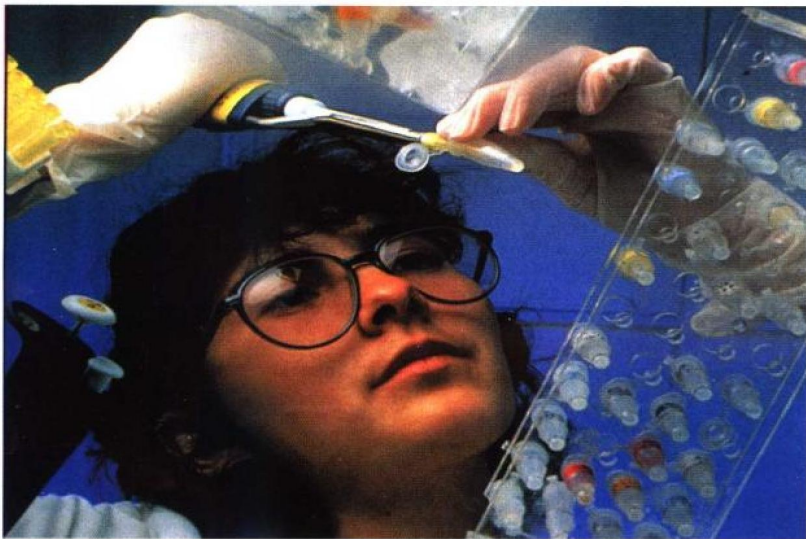
然而，草图的精确度是否能达到由它获取正确的蛋白质信息还不能完全肯定。实际上，基于这两张草图的蛋白质组预测存在着严重的误导。由于只有一张完全无错的人类DNA的“金图”才能将我们正确地引入蛋白质的世界，所以，还有很多的测序工作需要做，尽管如此，并不能减损我们对那些完成这项艰巨工作双方团队的尊敬之情。

因数量之间建立必然联系的原因不得而知。可能部分原因在于RNA切割作用可以保证从单个基因产生多个蛋白质产物吧；脊椎动物基因比无脊椎动物基因会产生更多的这种切割；但是，它们神经系统的发达程度与上述的相应数值应该是相关的。线虫比果蝇来说相对迟钝，就需要更多的特异蛋白质（相应更多的基因数）来对环境或天敌做出反应——即果蝇的相对发达的神经系统可以让它通过飞行而避开许多潜在的敌人或环境压力。同样的道理，更迟钝些的植物，当然需要不断地进化出新基因来应付不断出现的新敌人和各种天气环境的变化。

我们如果想知道造成明显进化变化的新基因的产生频率，就必须对更多的脊椎动物基因组进行测序。我们还需要搞清楚为什么脊椎动物比无脊椎动物具有更多的重复基因序列。人基因组中大多数的重复序列增加似乎是转座遗传因子产生或移动的结果。这种推测是有根据的，因为在脊椎动物进化过程中，许多突变体就是由于转座子移动到调控区，改变了基因表达方式而产生的。两栖动物和肺鱼的高重复DNA序列应该来源于适应它们不断改变的环境的快速进化之要求。

在脊椎动物进化过程中，考察占主导地位的调控因子的变化比考察蛋白编码因子变化的想法更好。例如，具有形态学差异的不同种类的狗的序列信息就是相当有用的，而且我认为获取同一生物不同种类的基因组草图也较有希望拿到资金资助。还不清楚还有多久我们才能比较我们人类基因组与猩猩基因组之间的差异。但如果能够进行这样的比较的话，我们显然更希望了解引起人类大脑比猩猩的相对较大和较发达的遗传学变化在哪儿。

从人类基因组草图中产生的许多新事实中，我认为最令人兴奋的是在4个成簇的同源框基因中并没有找到重复序列。与大多数相关的功能基因不同，同源框基因在染色体上的顺序正好反映了它们在胚胎发育过程中的时序表达模式。从这个方面来讲，它们与细菌基因中的能转录为单链mRNA的操纵子类似；位于细菌操纵子起始位置的基因先被RNA聚合



National Human Genome Research Institute, NIH

当我们看到作为多细胞生物赖以存在的基础的DNA文本时，会很自然地想到不断增加的生物体的复杂度会导致相应的基因数量的增加。因此，我和我的同事们都因为去年的一个与此截然相反的结果而深感吃惊，那就是果蝇（*Drosophila melanogaster*）的复杂度比线虫（*Caenorhabditis elegans*）高，但它的基因总数（13 500）却比线虫的（18 500）要少很多。更让人吃惊的是，最近有关小型植物拟南芥的基因组数据表明，它所含的基因数比线虫多出约28 000个之多。现在，又出现了再一次令人震惊的事实，一年前我个人估计人类基因组中应涵盖约7万到10万个基因，然而事实上却只有不足3万个。

生物体的复杂度为什么并没有与其基

酶转录，并且接着转录各自的DNA。可以想像，人的许多早期发育时序是由RNA聚合酶分子转录同源框基因中的那些较长内含子的所需的时间来反映的。假如果真如此的话，在胚胎发育时，在同源框基因中插入较长的转座序列可能会导致致命时序性关键事件错置。

对于参与草图计划的乃至更广泛范围内的生物学家们而言，在阅读人类基因组草图时必然会出现许许多多难以预料的观测结果和假说。然而，即使那些最热心的人士也会发现如果看得太多也会难以承受。在不久的将来，最大的研究成果将很有可能出现在人类与一种或多种模式生物已知功能的同源性基因的研究领域。在此之后，重要的研究成果会在研究人

类本身上产生巨大收益，以及弄清在自然和营养（nature and nurture）棘手的相互关系的难题上。人类的生命和生活，显然多于DNA文本、RNA和执行其指令的蛋白质以及它们之间的相互作用，而且，我们对此的各种思考，基于文化的事件和传统，由我们的大脑从一代传至下一代的知识同样也必将影响我们的生活。

因此，我们的基因组也许永远不能准确地预言我们的未来。但是，我们如果不能充分地运用这些信息的话，就更愚不可及了。毫无疑问，我们现在正在完成的这幅人类遗传图将是我们所读过的书中最为重要的一本。

《自然(Nature)》杂志编辑评论

Philip Campbell

只有极偶然的情况，科学上的突破才在某个特定的领域内引起轰动，喜悦和兴奋由此波及到其他研究领域并最后渗透到普通大众之中。人类基因组的第一张草图的发表和分析就是这样的一个绝好例证。它让我们得以对组成我们身体的亿万个细胞中的每一个细胞的核心，即组成我们染色体的DNA分子和遗传物质，有了第一次有意义的分析和观察。这本书为我们提供了一个机会，可以去体会人类基因组生物建构艺术、赞赏它们已经告诉我们的关于我们自身的一切，以及在生命王国中了解我们的“亲戚”和那些先我们而去的前辈。

像往常一样，科学界最大最新的突破性事件总会出现于《自然》杂志中。为什么研究者们会选择在《自然》杂志上发表他们的发现，而且这样做又有什么意义呢？还有，科学性期刊的作用究竟何在呢？它们的作用有以下几个方面：科学期刊提供了一个让科学家们呈现和讨论其发现的平台，并且为科学研究进步的成果提供了永久存录的地方。他们还会为努力得来的成果打上评价或评估的标

记，类似“同行评议”，就是完全、正式鉴定的标记。那些作鉴定的专家都是由杂志的编辑们选出的，而且他们的评价是在不知情或没有经济回报的条件下做出的。正如所有在《自然》杂志上发表的文章一样，人类基因组的论文被许多科学界的泰斗们阅读并极度推崇。我为此而向他们致敬。我们也同样为我们自己而感到自豪，因为我们能参与筛选科学界最引人注目的事件，为权威们做出评论而协调，同时，还能参与补充性评论的撰写。

到目前为止，基因组序列的数据已经能在网上查询了，为此应该感谢人类基因组计划的政策和及时数据发布。同时，所有文献，特别是在本书中重现的那篇精彩的有关序列的文章，使得所有有关这些C、A、G、T们的故事对于大多数的读者而言变得有意义了。我们也很乐意继承人类基因组计划的精神而把这些文章免费地与所有读者在网上共享。能接触到公共资助的工作，并且能帮助将其重要性尽可能清楚地展示给公众，是不可多得的荣誉。

前 言

有史以来，我们第一次能阅读构建人类的一套完整的指令。

这一意义深远的进展不可避免地引起了赞成和反对两方面强烈的反应，有时即使是同一个人对它也是毁誉各半。一方面，这是一个渴望了解自身及周围世界的社会取得的自我认识的又一高峰；并且这也是一项令人震惊的技术成就。它承诺了超过医学科学以往任何进展的实际利益：一个疾病治疗、诊断和预防的新时代。另一方面，通过植物、动物和人类的遗传改变“扰乱自然”的可能性引起了许多人本能的反应。而且，关于这种新知识对社会究竟会造成怎样的影响，对“做人究竟意味着什么”这一根本问题可能产生什么样的影响，人们普遍地、深深地感到不安。

在英国首相托尼·布莱尔 (Tony Blair) 的致词中曾提到，前面的路是“致力于各种可能性，并努力发展，然后直面各种艰难的道德、伦理问题，这些问题是这样一项不同寻常的科学发现不可避免地要引起的。”要做到这些，就需要对基因组和基因组学（对基因组进行测序、分析并得出结论的科学）有一个基本的理解。撰写本书的目的就是为了有助于这种基本了解。我们不想写得包罗万象，相反，我们希望本书能成为人类基因组探索的介绍或者是续篇。我们有意避开了介入基因组计划政治和个人的因素，而是集中于基因组本身。

本书的主要部分是一篇科学论文，这篇论文描述了人类遗传密码的首次拼接和分析，是全世界数千名科学工作者努力十多年的结果。和这篇论文一起的是一系列由各主要研究者写的短文，这些

文章提供了这一研究的科学背景，表达了他们的兴奋之情，以及对这一工作所做的确切评价。这些短文和那篇科学论文都来自《自然》杂志。

为了让更多的读者了解“人类基因组计划”，我们写了一系列介绍性的章节，概述了一些科学概念和支持这一计划的技术革新。本书首先对基因组作了定义，描述了其结构单元和基因组序列中包含的信息，以及从总体上讲我们每个人如何是同一主题的精细变体。接着是测序指导，描述了使科学家们得以测定基因组序列和构建那些基因组探秘所需要的图谱技术。然后是一个时间表，从这个计划最初灵感的产生一直追踪到它发展成为一个国际大项目。还有一篇关于序列分析中产生的主要发现的概述，作为该研究论文的阅读指南。在最后的几章中，我们讨论了遗传信息对个人和社会的含义，以及媒体关注的焦点。

我们目前得到的只是人类基因组的概貌，即所谓的“工作草图”。在这个图中，还有很多间隙要补，有一些模棱两可的序列要重新验证，这些工作足以让科学家们忙到至少2003年。很多很多年后，我们可能还是不能解开人类基因组的全部秘密。但是，我们已经到达了一个里程碑，我们第一次看到了人类基因组的全貌，这本身就是一个值得我们庆祝的成就。

C.丹尼斯 (Carina Dennis)
R.加拉格尔 (Richard Gallagher)
分别于华盛顿和伦敦
2001年8月

人类基因组入门

人类基因组序列被描述成人们所能想象得到的最珍贵的信息集合。但是，基因组究竟是什么？它有什么用？它究竟独特到什么程度？在揭示了基因组序列后，我们学到了什么？你——作为一个独一无二的基因组版本的主人——对这些问题可能很好奇，很想知道答案。在这开篇第一章，我们从生物学角度介绍基因组，并解释几个概念，这些概念是理解“人类基因组计划（HGP）”中的发现将产生的影响所必需的。

人类基因组鸟瞰

你的基因组就是你的遗传组成。它是你从父母那里遗传得到的信息，部分地指导你的生命过程。人类基因组由 30 亿（3 000 000 000）个数据，以脱氧核糖核酸（DNA）的形式组成。每个数据称为核苷酸或碱基，这就是 DNA 的组成单位。核苷酸连在一起成为极长的链，这些核苷酸沿着链的排列顺序就是 DNA 序列。

那么，基因组是干什么的呢？设想人体是一台复杂的生化机器，充满了各种相互作用的化学物质、分子和大分子，那么，基因组就是这台机器的蓝图。基因组包含组装各大分子组分的指令，这些指令就是基因，基因编码的功能成分主要是蛋白质。基因影响我们的身体特征，如眼睛的颜色、身高以及头发的颜色，还影响我们对某些疾病的易感性。

然而，基因组并不仅仅是一张各组成部分的列表，它还对何时、何地产生这些组分的信息进行整合。比如胚胎发

育早期的神奇过程，从一个单细胞一直到一个多细胞的生物体，使这一过程得以发生的种种变化需要一整套基因的高度协调的表达。甚至那些比较简单的活动，如对抗感染的过程，也需要复杂的基因间相互作用的模式来组织合适的抵抗反应。因此，基因组并不是各个单独起作用的基因的集合，正相反，它包含着对信息的全局性、高度协同的控制，以执行一系列细胞功能。知道基因组序列的一大好处就是我们能开始理解细胞功能的调控。

你自己的基因组和这个星球上的任何其他人类至少有 99.9% 是一致的，如果拿它同与你具有亲缘关系的人的基因组相比，这个数字会更高。0.1% 的差异看上去并不是一个给人深刻印象的关于独特性的数量，但它意味着在我们的遗传密码中包含大约 300 万个差异，这为个体特性提供了足够的余地。一些差异没有明显的效应，而另外一些则会影响我们的外表、行为、对疾病的易感性以及对药物治疗的反应。本质上，我们全都由同样的遗传底稿描画而成，但我们各自间的差别发挥了很大的作用。

“在过去 40 年中，我已见过许多令人振奋的生物学事件的发生。但是，当我第一次读到这篇描述我们基因组概貌的草图时，我仍然感到震撼。”

David Baltimore, 加州理工大学



DNA 晶体。液晶状态的 DNA 形成一种像扇子一样的精致的圆锥形结构。[Courtesy of Michael W. Davidson, Florida State University, Tallahassee]

“第一次看到一条人类染色体的全部DNA排列顺序，就像看惯了小舢板的人，第一次看见一艘大海轮破雾出现一样。”

Francis Collins，国家人类基因组研究所

也就是说，各个个体并不能简化成他们的遗传特征，我们并不仅仅是我们的基因组的产物。我们共有什么，又怎样不同，我们的基因定义什么，又不定义什么，这些都是复杂的问题。联合国科教文组织的《关于人类基因组和人权的全球宣言（The Universal Declaration on the Human Genome and Human Rights）》对这一概念作了精彩的表述。

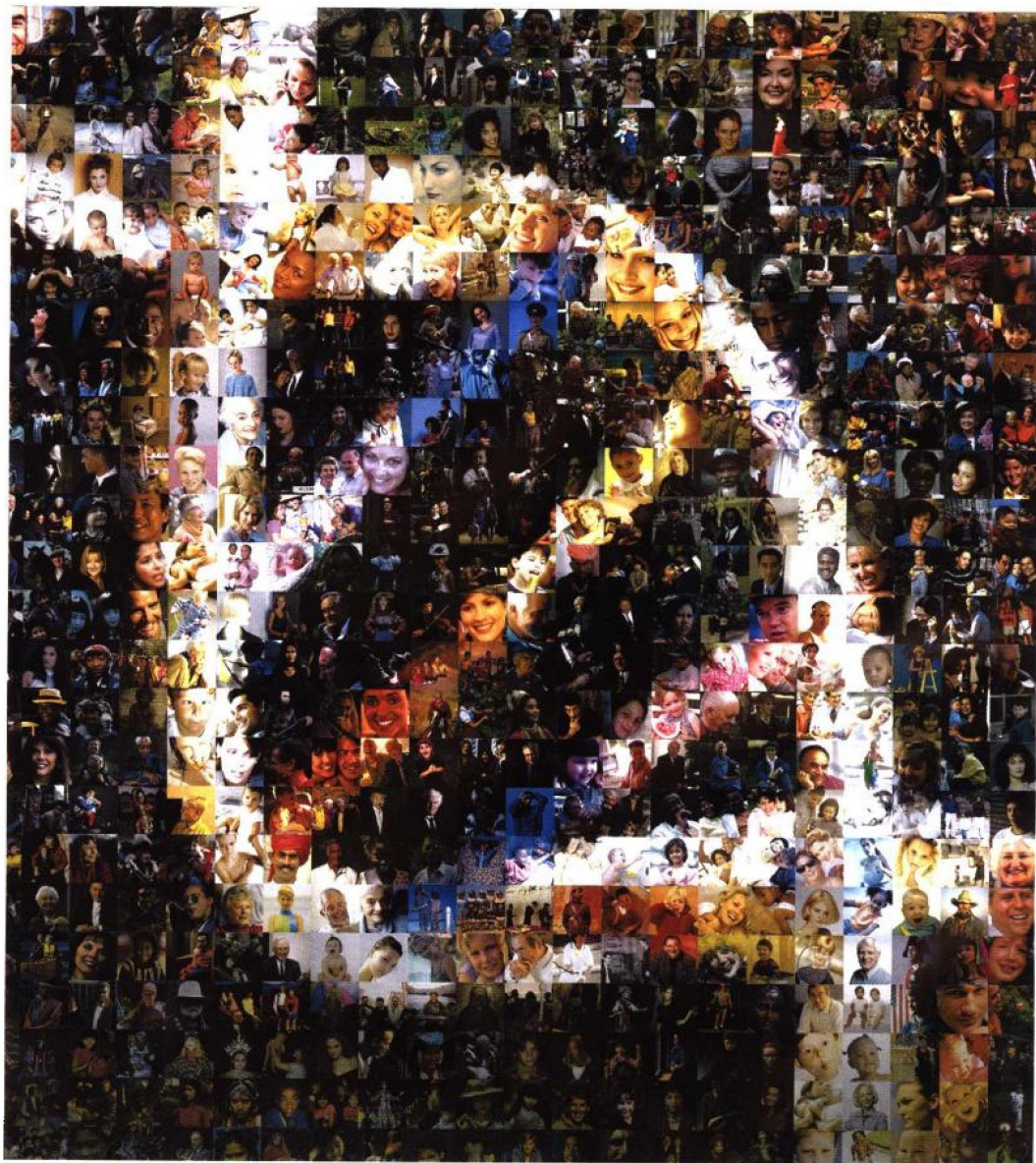
为了号召对人类独特性和多样性的尊重，该宣言指出：“人类基因组是人类大家庭所有成员基本一致的基础，同时也是认识其遗传珍贵性和多样性的基础。从象征意义上来说，它是人类的遗产。”

确定人类基因组序列有什么实际的利益呢？对于科学家们来说，利益是丰厚的、立竿见影的。事实上，生物学的所有领域，从生物化学到行为心理学，从古生物学到寄生虫学，一直到癌症的研究，都将获得新的洞见。该序列的应用之一就是为我们人类历史的研究打开了一个窗口，为探索我们的祖先如何在全世界散布和定居提供了生物学的视窗。

人类：人类是两足灵长类哺乳动物，特征是有一个高度进化的大脑和直立行走的姿势，从而将双手解放出来用于制造工具。根据统计结果，成人平均体重为36~95公斤，平均身高1.7米，平均年龄为75岁。

[The cover image, by Eric Lander, was created by Runaway Technology, Inc.

(www.photomosaic.com) using PhotoMosaic by Robert Silvers from original artwork by Darryl Leja. It is used courtesy of the Whitehead Institute for Biomedical Research.]



基因组序列带来的益处不仅仅局限在科学界，它也将导致医药革命。我们的遗传组成影响我们对疾病的易感性。大多数常见的疾病，如心血管病、风湿性关节炎和糖尿病，与多个基因有关，还受环境因素的影响。这些复杂的疾病，也称为“多因素”或“多源性”疾病，其遗传基础难以解释清楚。基因组学将帮什么忙呢？除了鉴定所有的基因外，研究者一直在收集基因组中的变异，并查明个体间的遗传差异的位置和类型。这将有助于识别在那些复杂的疾病中占优势的遗传构造。这不仅使疾病得以早发现、早治疗，还将把我们推向预防疾病的新时代，常言道，预防优于治疗。知道了基因组的序列及其在个体间如何变化，还将帮我们弄清楚为什么有些人对某种病的药物治疗的反应比其他人要好。这将为药物的开发揭示新的药靶。

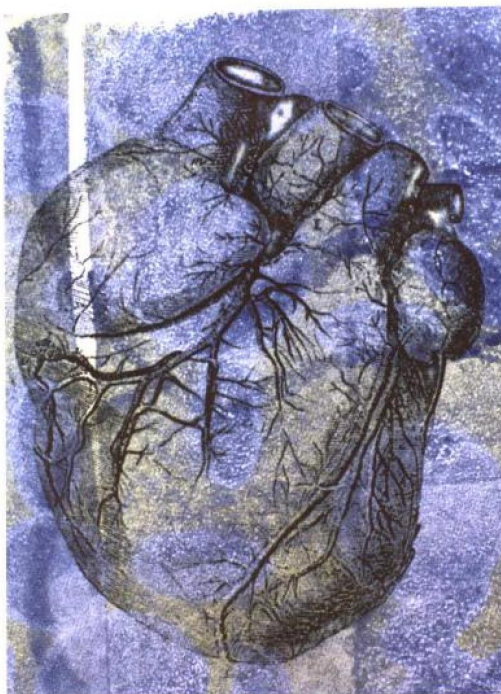
基因组测序只不过是我们要全面认识人的生物学本质的漫漫长路上的第一步。然而，尽管还有很多东西要去探求，我们已经获得了一些关于基因组性质和使我们有别于其他生物的原因的伟大洞见。

聚焦基因组

在深入介绍基因组序列的细节之前，先有生物体水平的知识，会让我们有所帮助。让我们从人体开始，并逐渐把它分解成小一级的构成成分。

第一个划分水平是构成人体的十个相互依赖的系统。它们是骨骼系统、肌肉系统、循环系统、神经系统、呼吸系统、消化系统、分泌系统、内分泌系统、生殖系统和免疫系统。每个系统都由许多相互作用的组织和器官构成一个网络，每一种组织和器官都有其独特的细胞类型的组合。

“获得基因组序列仅仅是个开始。科学家们更想知道的是基因以及它们在疾病的预防、诊断和治疗中能起到的作用。”

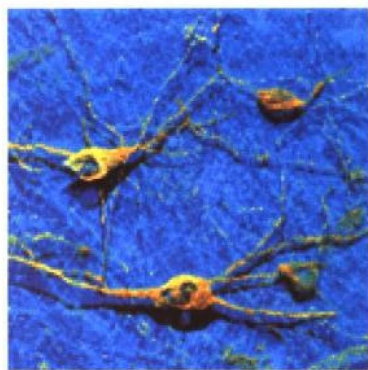


心脏是循环系统的泵。像其他器官一样，心脏有着独特的结构和功能。它包含有多种类型的细胞。

“毫无疑问，我们正在见证医药科学领域的伟大革命，其重要性远远超过了抗生素的发现，这是21世纪第一个伟大的技术进步。”

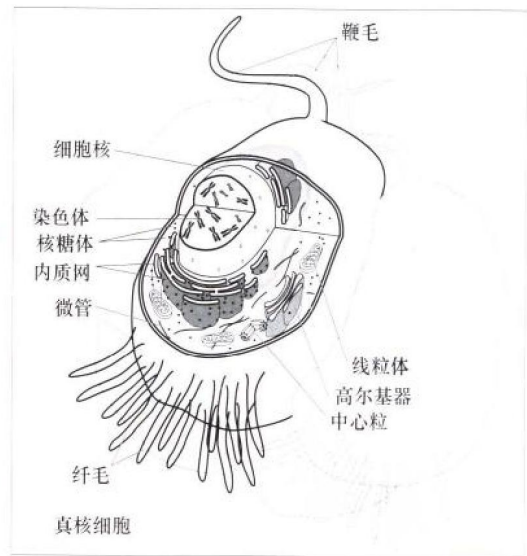
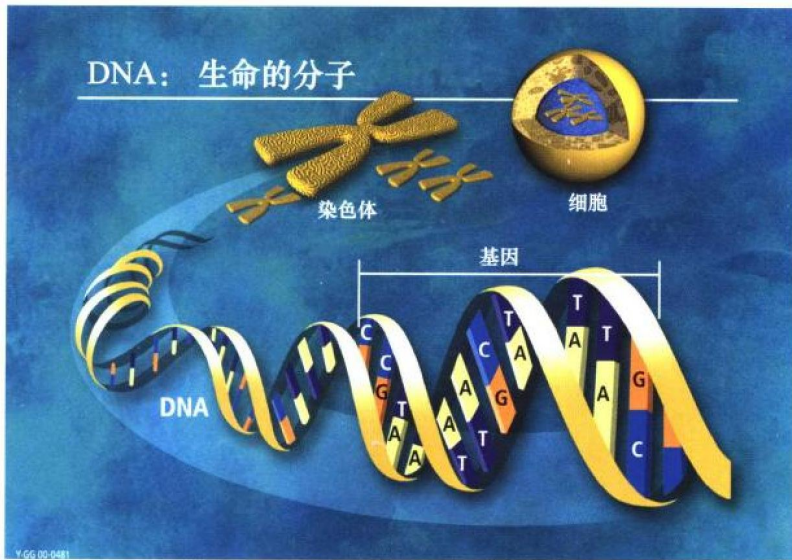
英国首相托尼·布莱尔

细胞是所有生命形式的基本单位。生物体的细胞数量从一个到数万亿个不等。最简单的生命体是可以单独生活的单细胞生物如细菌，人体则大约由75万亿（75 000 000 000 000）个细胞组成。细胞的大小差别很大：典型的人体细胞直径约为20微米（0.000 02米），而有些神经细胞长达1米以上。人体所有的细胞尽管有相似的基本结构，但表现出极大的多样性。共有200多种在形态、寿命和功能上都不相同的人体细胞。但是，一个人的每一个细胞都含有相同的遗传信息。正是遗传信息的选择性表达，即基



所有有生命的东西都由细胞组成。这是一张脑组织的共焦显微镜图，黄色的为神经细胞。[Medical Microcopy Science/ Wellcome Photo Library]

Randy Scott, Incyte Genomics 的主席



细胞内部结构。右上图：部分细胞器，每一种都有自己独特的功能(NIH/NIH)。DNA 位于细胞核内，上图显示了 DNA 包装为染色体。[NHGRI,NIH]

因开启或关闭，导致细胞的多样性。我们可以把基因组比作一个储藏充分的厨房；就像原料的不同搭配可以烹饪出不同的菜肴，基因表达的不同模式可以生成许多不同的细胞类型。

每一个细胞可进一步细分成许多称为细胞器（意为“小器官”）的结构。细

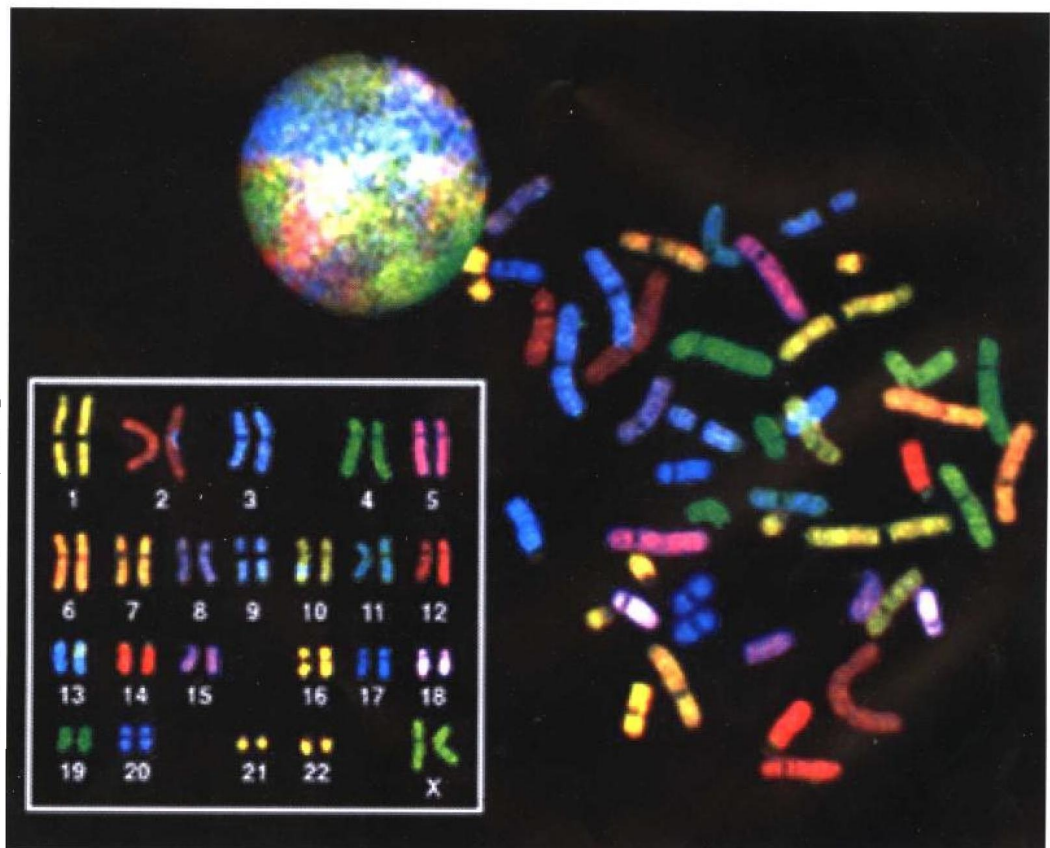
胞器在细胞内执行特定的功能；例如，线粒体是能量产生者，核糖体是蛋白质工厂，溶酶体是废物处理站，内质网则负责分子的标别、分类和运输。这里，我们所关心的细胞器是细胞核，可以把它看成细胞的总部，DNA 就在核内。

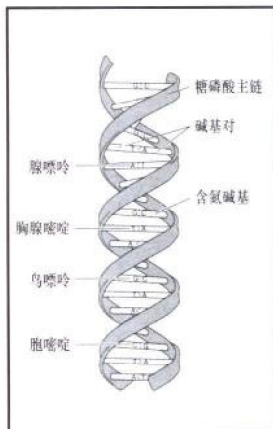
在细胞核内，细细长长的线状 DNA

人类染色体组型谱

染色体(意为可被染色的物质)可被某些染料染色。各染色体可由大小及染色条带区分开来。这一技术称作染色体组型分析。每一条染色体都染上了一定的颜色，这对辨别染色体异常很有帮助。人类很多疾病与染色体的异常有关，包括丢失、增加、断裂、易位。例如唐氏综合征(Down syndrome)就是由多了一条 21 号染色体引起。

[NHGRI,NIH]





紧密地盘绕成清晰可见的棒状结构，称为染色体。每一物种将其DNA分装成特定数目的染色体，每个染色体有特定的大小和形状。蚊子只有6条染色体，豌豆有14条，而某几种蕨类植物则多达1000条以上。猫和狗分别有38和78条，黑猩猩有48条。人类有46条染色体，排成23对，其中22对称为常染色体。常染色体按大小依次递减编为1~22号（21号染色体是个例外，它实际上比22号染色体小一些）。剩下的一对是性染色体。正常的男性有一条X染色体和一条Y染色体，正常的女性则有两条X染色体。Y染色体是人类基因组中最小的染色体，像一片森林中的一个小树桩。

DNA的详细说明

DNA的基本单位——核苷酸由一个糖分子、一个磷酸分子和一个含氮碱基构成。碱基是编码信息的成分，有4种不同的形式：腺嘌呤、胞嘧啶、鸟嘌呤和胸腺嘧啶。这些碱基分别简写为A、C、G、T，是遗传字母表中的四种字母。

核苷酸串在一起形成难以置信的DNA长链；如果把一个细胞中的DNA连起来不弄断，长度可达2米，而其宽度仅有50万亿分之一厘米。细胞内的DNA

分子由两条链组成，这两条链相互盘绕形成一个螺旋状的梯子，这就是非常著名的双螺旋结构。梯子的两个边是由糖和磷酸分子构成的链，“横档”是一对一对的碱基，碱基分别连在两条链上。DNA一个显著的特征是碱基配对的方式：A只与T配对连接，C只与G配对连接。这些配对称作“碱基对”。

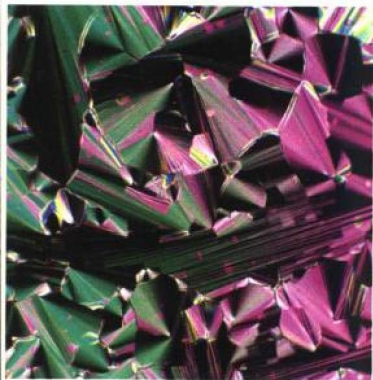
碱基沿着DNA链的线性排列次序就是DNA的序列。一种生物的基因组大小通常表示为碱基对总数；人类基因组约有30亿碱基对。生物的复杂性同其基因组的大小之间几乎没有相关性。虽然人类基因组至少是酵母基因组的200多倍，却和青蛙、鲨鱼的基因组相当，而和蜉蝣的相比就太小了，蜉蝣基因组有150亿碱基对；再同另一种单细胞微生物（*Amoeba dubia*，一种变形虫）相比，更是格外的小，它的基因组足有人类的200多倍。

生命的蓝图

人的一生中，细胞不断地替代更新。这或许是因为大多数细胞都能复制。细胞在分裂前，先复制其基因组，以便于相同的遗传物质能分配到每一个子细胞中。

DNA结构。左上图展示了DNA双螺旋结构(c. Pete Artymiuk/Welcome Picture Library)。上图展示了碱基对。两个核苷酸连接形成碱基对。胞嘧啶和鸟嘌呤由三个氢键相连，腺嘌呤和胸腺嘧啶由两个氢键相连。因此胞嘧啶和鸟嘌呤碱基的连接要比腺嘌呤和胸腺嘧啶的连接稳固一些。
[NHGRI,HIN]

DNA：一个标志



高密度液晶状态的 DNA 显微图。[Courtesy of Michael W. Davidson, Florida State University, Tallahassee]

人们如何发现DNA，认识其重要性以及揭示其结构的历史是非同寻常的。有很多书专门讲述这一历史，向人们详细描述那些迷人的发现，以及科学背后那些同样迷人的人物。我们将重述这段历史中的几个闪光点，以便于理解DNA是如何从一个“令人生厌的分子”转变成为一种大受欢迎的现代标志。

作为一个不算吉利的开端，DNA最早是从被丢弃的外科绷带上的脓细胞中分离出来的。那是在1869年，瑞士的生化学家Friedrich Miescher报道了一种不能被蛋白消化酶破坏的物质，也就是说“不属于迄今为止已知的任何一种蛋白质”。因为它是从细胞核中得到的，Miescher将它命名为核素（nuclein），但后来发现它是酸性的，就重命名为核酸（nucleic acid）。核酸分两种，根据含有的糖基（核糖或脱氧核糖）的不同，分别命名为核糖核酸（RNA）和脱氧核糖核酸（DNA）。

DNA可以传递信息的线索最早来源于人们对肺炎球菌的研究。肺炎球菌传播肺炎，是当时很普遍的致命疾病的细菌。20世纪20年代，Frederick Griffiths报道了两个肺炎球菌菌株：一个是有毒性的，一旦注射到小鼠体内会使小鼠死亡；而另一个是无毒性的。无论是无毒性的菌株，还是经热处理杀死的毒性菌株都不是致死的。但是，如果将经热处理杀死的毒性菌株和不经处理的无毒性菌株一起注射，会导致小鼠死亡。死去的毒性菌株不知怎么将一种致死因子传递到了无毒性的菌株上。1943年，Oswald Avery和同事们发现DNA是起作用的转变因子。他的研究组在用了大约20加仑（76升）的细菌后，终于得到了不含任何别的物质的纯化的DNA，并证明只有它在转化中起作用。

在当时，这项工作几乎没有引起注意。DNA被认为是一种过于简单的分子，尤其同蛋白质相比，以至于不能包含一种生物体的所有遗传信息。DNA是遗传物质的确切证据最终由Alfred Hershey和Martha Chase在1952年得到。他们的实验用的是噬

菌体（意为“吃细菌的生物”）。这是一类病毒，它能侵染并控制细菌并利用细菌的细胞机制繁殖更多的病毒颗粒。噬菌体由蛋白质和DNA组成，究竟是哪一個在起作用呢？Hershey和Chase将两种噬菌体的培养物做了标记：一种用放射性的磷标记DNA；另一种用放射性的硫标记蛋白质。然后，他们分别用这两种噬菌体感染不同的细菌克隆，并且追踪放射性物质的位置。结果，放射性的硫（也就是说蛋白质）留在细菌外，而放射性的磷（也就是说DNA）进入到细菌里。而且，新产生的噬菌体含有放射性的磷。这些结果有力地证明了DNA是遗传物质。

在接受了DNA作为遗传的化学基础后，许多科学家试图发现其结构。Rosalind Franklin就是其中之一，她是伦敦大学的化学家。她使用的是X射线结晶技术，就是把X射线打在待研究的物质的晶体上，产生的X射线的散射被用于构建结构图。Franklin及其同事Maurice Wilkins得到了关于DNA结构的重要认识，特别是其螺旋的性质。但真正揭示DNA那令人困扰的结构的人却是在英国剑桥工作的两名见解独特的科学家：詹姆斯·沃森（James Watson）和弗兰西斯·克里克（Francis Crick）。

正如沃森后来所说的那样，DNA结构的发现过程之所以成为一个广为流传的科学故事而不是被遗忘，是因为这是一个关于无限勇气、不服权威、蔑视传统的传奇故事。他们俩最初（在1951年底）揭示DNA结构的尝试是错误的；凭着对Franklin一次讲座的凌乱回忆，他们得到了一个骨架在内部的三链模型。尽管实验室负责人要求他们把DNA结构的问题交给其他人，沃森和克里克还是很勇敢地做了另一次尝试。1953年1月，沃森分析了Franklin和Wilkins的X射线数据，也就在这时他和克里克开始用大的三维模型疯狂地构建DNA的结构，这成为他们成功的关键之一。1953年3月7日，他们已经准备好向剑桥Eagle酒吧的年轻顾客们宣布他们已经发现了“生命的奥秘”。