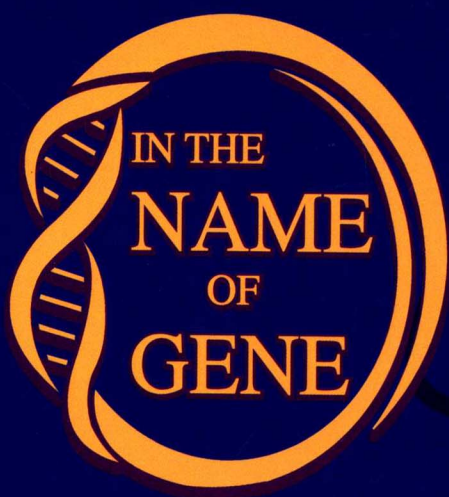




前沿零距离
Walking at Frontiers



黄捷·著

中国科学院院士
杨焕明

作序并推荐

基因的名义

高等教育出版社



前沿零距离
Walking at Frontiers

基因的名义

剑桥博士纵论基因的前世今生

黄捷◎著

高等教育出版社·北京

内容简介

随着基因领域各项技术的快速发展,其一部分研究成果已由“遥不可及”的实验室研究转变成成为“身边的科技”——从无创产前检测和癌症患者的精准用药到制药巨头通过基因技术的快速药物筛选,从社会名人通过基因寻祖认宗到国家公安部门开始组建打拐基因库,基因在新的时代将让更多的百姓实实在在地受益。本书结合基因领域最受关注的新闻、最前沿的科研成果,深入浅出地讲解了基因科学相关知识,包括基因检测、基因解读、基因编辑等,适合对基因领域感兴趣的各类人群阅读。

图书在版编目(CIP)数据

基因的名义:剑桥博士纵论基因的前世今生 / 黄捷
著. — 北京:高等教育出版社, 2018.7
ISBN 978-7-04-049950-6

I. ①基… II. ①黄… III. ①基因—普及读物 IV.
① Q343.1-49

中国版本图书馆 CIP 数据核字 (2018) 第 132364 号

策划编辑 黄慧靖 责任编辑 张 冉 封面设计 顾 斌 责任印制 尤 静

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
印 刷 涿州市星河印刷有限公司
开 本 787mm × 1092mm 1/16
印 张 16
字 数 218 千字
购书热线 010-58581118
咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.hepmall.com.cn>
<http://www.hepmall.com>
<http://www.hepmall.cn>
版 次 2018 年 7 月第 1 版
印 次 2018 年 7 月第 1 次印刷
定 价 58.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换
版权所有 侵权必究
物 料 号 49950-00

《基因的名义》， 开启基因科普大门的钥匙

这是一本难得的好书。

本书作者基于自己丰富的个人经历和深厚的科学功底，把一本科普书写得妙趣横生。书中针对诸多大众关注的话题，通过具体事例，向所有基因的“粉丝”介绍了基因学、统计学、计算机和大数据分析等方面的基本原理与相关知识，并通过介绍专业研究和新近进展，对基因进行了科学解读。对于基因行业的资深人士，对于其他行业内对生命科学特别是基因学有兴趣的专家和学者，以及广大的关注自身健康和想学习基因学基础知识的读者，这都不失为一本好书。

基因已经从实验室走进了社会、家庭以及我们的健康和生活。本书开篇的“基因亮剑”，通过作者母校剑桥大学（和其校友）首创的基因思想、基因科学和实用的基因技术，将基因领域的核心知识清晰地呈现给读者。接下来，该书聚焦了目前三个有关健康的热点（祖源和血缘分析、癌症预防治疗、出生缺陷预测），让读者能够得到较为全面且准确的科学信息。然后，该书科学地“揭秘”了目前大众非常关心的消费级基因。最后，该书系统地回顾

了基因行业的创业，为有志于推动基因科技向前发展的创业者提供了行业发展方向的有益参考。

基因，蕴藏着很多我们还不知道的秘密。对于钟情于基因、有志于生命探索的研究者来说，《基因的名义》可以说是一把开启这扇大门的钥匙。

中国科学院院士 杨焕明

2018年5月20日

峰回路转，寻梦基因

黄捷同学让我为他的新著《基因的名义》作序，实在令我惶恐。作为一个天文学家，与基因学所属的学科领域相距甚远，实在不能从专业角度对此进行评价与解读。但是作为相识近 20 年的老友，十分期待这本书的出版，也相信这是一本用心之作，值得一读。

天文学家研究的是起源，物质的起源，宇宙的起源，地球的起源，以及未来的世界。而生物学家特别是基因学家其实很相似，他们研究的是生命的起源、人类的未来。美国国家宇航局（NASA）首席科学家 Kathie Olsen 是一位资深的生物学家。在一次演讲中，她提到，天文学家的重要任务之一是发现外星生命，所以让一个生物学家当此重任没有什么惊奇的。英国现代遗传学和统计学之父 Ronald Fisher 的剑桥大学本科学位是天文学学士。天文学研究与生物基因研究看来有着不解之缘。

在我眼中，黄捷是一个追梦人，一直坚持追求自己的梦想。我认识黄捷已近 20 年，1999 年，我从北大去密歇根大学读天文学博士，他从北医去密歇根大学读体育科学。美丽的密歇根安娜堡博大而包容，学风扎实，文理兼蓄，密歇根的中国学子也非常活跃。黄捷是个思维活跃又颇具勇气的人，更何况他还“倨傲不逊”，不走寻常路，后来他放弃了在密歇根攻读博士学位，转而去学习更能点燃他激情的 IT 技术，他基本从 $1 + 1 = 2$ 学起。由于放弃了读博士的全额奖学金，黄捷一边兼职打工一边学习一个全新的领域，而我则一直在最热爱的大文学领域探索努力。我们在美国的第一站，虽然有着完

全不同的专业和心路历程，但在坚持追求梦想的道路上，我们一直保持着真诚的友谊。

2005年，我博士毕业后到哈佛大学做博士后，后来又在哈佛大学史密松天体物理中心（CFA）做PI研究员。2008年，黄捷也来到了哈佛大学医学院做博士后。从那个时候开始，黄捷的“倨傲不逊”和那颗“不安分的心”算是有了一个归宿。他所在的麻省总医院人类基因研究中心是全美业内第一，并位于世界最前沿。相信他在此期间收获颇多。黄捷在大西洋彼岸的剑桥大学做研究，在基因研究的探索之路上也取得丰硕成果，他的研究成果登上了国际权威的《自然》杂志。

这些年来，虽然我先行回国，但我们之间一直保持着密切联系，每年也总能见上几面，黄捷每每会向我谈起他在基因学领域的研究成果与趣事。在北京这片热土，当年哈佛、密歇根的同学朋友越来越多，祖国的发展进步令人振奋。我想这本《基因的名义》是黄捷对自己在海外学习18年的一个知识总结，相信此书会令读者受益。同时期望黄捷同学不忘初心，实现梦想，报效祖国！

刘继峰

中国科学院国家天文台首席科学家

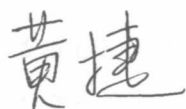
2018年5月于北京

前言

自从地球上有了人类，科学家和哲学家就对人类自身的遗传现象和本质，对地球上的动物和植物的遗传现象和本质，展开了思考和探索。自从达尔文（Charles Darwin）在 1859 年出版《物种起源》，提出进化论，科学家们对于遗传和基因的研究越来越热烈，并已经持续了近 160 年。从达尔文提出的泛生论到孟德尔（Gregor Mendel）和摩尔根（Thomas Morgan）奠定的遗传三大定律，从沃森（James Dewey Watson）与克里克（Francis Harry Compton Crick）发现 DNA 双螺旋结构，到麻省理工学院华裔科学家张锋领衔创建的基因编辑技术，长江后浪推前浪，其间的过程可谓传奇。仅仅 100 余年的时间里，令一众高能的科学家们在基因研究领域收获了 40 项诺贝尔奖，不可谓不惊人与“高大上”。如今，基因已经从纯粹的研究领域进入了百姓的生活。当基因检测技术用于司法鉴定、亲子鉴定，说是你的就是你的。正是这一份精准，让基因在走入大众视野后由神秘变成神奇。近年来随着基因研究领域各项技术的快速发展，相关的研究与检测费用也大幅下降，使这一原本“远在天边”的科技被越来越广泛地应用到了大众的生活中，不仅仅可以用来鉴定血缘关系和遗传疾病，甚至还被用来预测天赋、预测颜值。

基因无处不在，我们从医院收到的化验单就有可能有基因检测的项目，我们从银行和保险公司收到的赠品里面就有可能有基因检测的产品，我们想知道我们的长辈是不是有阿尔茨海默病的基因，我们想知道我们的晚辈是不是有某一方面的天赋基因。有人说基因是“自私”的，在生物体中的基因总

是有利于基因自身的传播，哪怕那样造成的结果不一定是有利于生物体自身生存的。而恰恰是这种“自私”，解释了很多自然界中尤其是亲属之间的利他行为：当一个生物体愿意冒着很大风险，为了保护近亲的安全而去战斗的时候，它保护了与它基因相近的更多的基因去繁衍；当一只雄蜘蛛冒着被吃掉的生命危险去寻找雌蜘蛛交配的时候，它把蜘蛛的基因遗传给了下一代。基因是科学的，基因是客观的；基因是我们每个人都应该读懂也能读懂的。我写这本书的目的，就是要传播基因的知识，特别是非常实用的知识，期待每一位读者真正地懂得基因，健康一生！



2018年5月于北京

目录

001 / 基因简史 160 年（从 1859 年到 2018 年）

009 / 第一篇 基因亮“剑”，来自剑桥大学的基因传奇

思考基因，从剑桥大学校友达尔文提出的进化论说起……………011

观测基因，从剑桥大学发现的双螺旋结构说起……………023

检测基因，从剑桥大学发明的桑格测序法说起……………033

调控基因，从剑桥大学首创的克隆技术说起……………058

我的基因故事：与克隆之父面对面……………070

071 / 第二篇 基因不再“高冷”，走进大众生活

知古——基因寻祖……………073

通今——亲子鉴定、亲人鉴定……………083

预测未来——科学的风险评估不是“算命”……………091

改变基因——从转基因到基因编辑……………105

我的基因故事：远道而来寻亲的舅舅……………111

113 / 第三篇 一切的癌症，都是基因突变惹的祸

世上本没有癌，基因突变多了，也便成了癌……………115

癌症的传染性……………121

液态活检，全球十大新兴技术榜首……………127

癌症的靶向治疗……………130

我的基因故事：外甥女和乳腺癌……………147

149 / 第四篇 预防出生缺陷的三道基因防线

一级预防，婚前和孕前基因检测……………151

二级预防，产前基因检测……………154

三级预防，新生儿基因检测……………159

普京的警告，比核弹更可怕的“超人类战士”……………165

我的基因故事：10年后破译堂姐的多囊肾基因……………167

169 / 第五篇 消费级基因，先看懂再买单

相亲：一颗诚恳的心，和永恒的基因……………171

寻找基因“高富帅”：端粒高，线粒体富，Y染色体帅……………175

还有待于进一步研究的天赋基因……………178

基因与“美好生活”和“长命百岁”……………181

我的基因故事：遗憾没能留住母亲的基因……………186

187 / 第六篇 揭秘基因解读，不确定不等于不靠谱

基因数据长这个模样……………189

基因扫码，测没测，准不准，一扫便知……………191

“精准”的定义和衡量标准……………193

十大“靠谱”的基因解读示例……………196

“说三道四”的基因解读示例……………210

人工智能，让基因解读进入新时代……………213

221 / 第七篇 基因创业，下一个首富出自基因领域？

拥有 GENE.COM 域名的 Genentech 公司，在 2009 年市值就达一千亿美元…223

II 基因的名义

宜曼达，垄断产业链上游的龙头老大·····	224
消费基因解读，“谷歌夫人”醉翁之意不在酒·····	225
基因存储，让大数据变成“小蛋糕”·····	227
共享基因，让我们一起去颠覆·····	230

233 / 附录

附录一 基因学领域诺贝尔生理学或医学奖获奖者概览·····	235
附录二 与基因相关的获得诺贝尔化学奖的科学家概览·····	241

基因简史 160 年（从 1859 年到 2018 年）

基因的英文单词为 **gene**，来源于古希腊语 **génos**，是种族和后代的意思。基因支持着生命的基本构造和性能，储存着生命的孕育——生长——凋亡过程的全部信息。可以毫不夸张地说，生物体的生、老、病、死等一切生命现象都与基因有关。基因有两个特点，一是复制自己，二是传宗接代。研究基因的学科叫基因学（**genetics**），最开始在中文里翻译成遗传学。现在的基因学研究已经超越了遗传的范畴，比如，由于核辐射导致的基因突变和癌症，跟传统的遗传学关系不大。但是，本书中，还是延用“遗传学”这个词。公认的遗传学鼻祖是 1822 年出生于当时的奥地利帝国（现在的捷克共和国布尔诺市）的一位修道士——孟德尔，他在 19 世纪 60 年代就发现了遗传的基本定律，并提出了基因的概念。1909 年，丹麦植物学家约翰森（**Wilhelm Johannsen**）首次正式提出“**gene**”这个名词，而基因学（**genetics**）是由英国剑桥大学生物学家贝特森（**William Bateson**）在 1905 年正式提出的。中文里的“基因”一词是我国著名的遗传学家谈家桢教授翻译过来的，这个翻译不仅发音相似，意思也很符合，“基因”就是携带和传递生命密码的“基”本“因”子！

一说到基因，我们首先想到的可能是它的功能，比如说基因可以从父母传给子女、基因可以影响一个人的外貌和身体健康状况等。那么，基因到底是由什么组成的呢？基因的基本组成成分是核酸（**nucleic acid**），或许有的

读者从某些公司或媒体的宣传上听说过核酸，比如核酸饮料、核酸保健品等。核酸是由许多核苷酸聚合成的生物大分子化合物。如果把核酸看作是一个摩天大楼的话，我们可以把核苷酸理解成组成该大楼的每一户。一个核苷酸分子由磷酸根、核糖和含氮碱基三部分构成，我们可以想象为每一户都有客厅、卧室、厨房三部分。核糖有两种：一种是普通的核糖，就好比是连着阳台的客厅；另一种称为脱氧核糖，我们可以把“脱氧”想象成没有阳台的客厅。正如居民楼一般都有阳台，而办公楼一般都没有阳台一样，由脱氧核糖组成的核酸称为“脱氧核糖核酸”（deoxyribonucleic acid, DNA）。而由普通的核糖组成的核酸称为“核糖核酸”（ribonucleic acid, RNA）。一个一个核苷酸连接成为核酸，正如一层一层、一户一户组成高的大楼。除某些病毒的基因是由 RNA 构成外，多数生物包括人类的基因由 DNA 构成。RNA 与 DNA 不仅仅是在核糖上有差别（DNA 是脱氧核糖，RNA 是核糖），在含氮碱基上，DNA 有腺嘌呤（adenine，缩写为 A）、胞嘧啶（cytosine，缩写为 C）、鸟嘌呤（guanine，缩写为 G）、胸腺嘧啶（thymine，缩写为 T），而 RNA 的胸腺嘧啶（T）被尿嘧啶（uridine，缩写为 U）替代，就好比写字楼的会议室被居民楼的厨房替代一样。我再用写字楼和居民楼的比喻来说一下 DNA 和 RNA 的最后一个差别：DNA 是两条长链缠绕在一起，而 RNA 是一条单链。这就好比高档写字楼有时候喜欢建个双子塔，而居民楼一般是独立的一栋一栋。由于本书的重点是说人的基因，所以我主要就介绍 DNA。通常，DNA 的序列就用碱基的四个字母 A、C、G、T 的排列来表示。我们人类的每一个细胞（除了生殖细胞）里的 DNA 总共有两套大约 30 亿对碱基组成的基因组，每一套基因组上面有大概 2 万多个可生成蛋白质或功能 RNA 的基因。组成基因的 DNA 只占 30 亿对碱基的大约 2%。我们现在说的基因，通常可以泛指全部的 DNA（基因组）。虽然 DNA 在细胞中很早就被发现了，但证明其为遗传物质的决定性实验是 1944 年美国细菌学家艾弗里（Oswald Avery）的肺炎双球菌转化实验。

细胞中的 DNA 以染色体的形式存在。染色体的英文是 chromosome，由于它容易被碱性染料染成深色而得名。染色体是由直径仅 $100\text{ \AA}$ ($1\text{ \AA}=0.1\text{ nm}$) 的 DNA 和组蛋白高度螺旋化的纤维所组成，每一条染色单体都可看作是一条双螺旋的 DNA 分子。细胞分裂期的染色体呈圆柱状或杆状。细胞在分裂间期时，染色体解螺旋而形成极其伸展的细丝，此时不易为染料所着色，光镜下呈无定形物质，称之为染色质。由染色体形态结构或数量引起的疾病称为染色体病。目前已发现的染色体病有 100 余种，最常见的是由于第 21 号染色体多出一条而导致的唐氏综合征 (Down syndrome)。

1842 年，瑞士植物学家内格里 (Karl Wilhelm von Nägeli) 发现了染色体。1879 年，德国生物学家弗莱明 (Walter Flemming) 发现细胞核中的丝状和粒状的物质可被碱性染料染色，因此将其命名为染色质 (chromatin)。染色质和染色体的主要成分都是 DNA 和蛋白质，它们是同一物质在细胞分裂间期和分裂期的不同形态表现而已。1902 年，美国生物学家萨顿 (Walter Sutton) 与鲍维里 (Theodor Boveri) 一起观察并发现了细胞的减数分裂时染色体与基因具有明显的平行关系，并据此推测基因位于染色体上。1928 年，美国遗传学家摩尔根通过果蝇杂交实验证实了染色体是基因的载体，并因此在 1933 年获得了诺贝尔生理学或医学奖。1956 年，美籍华裔遗传学家蒋有兴 (Joe Hin Tjio) 首次发现人的体细胞的染色体数目为 46 条。这 46 条染色体呈 23 对，分别来自父母，第 23 对染色体决定性别，被称为性染色体，XX 为女性，XY 为男性，其他的 22 对被称为常染色体。

一个细胞全部的 DNA 加起来，我们称之为基因组 (genome)。除了基因组，还有转录组 (transcriptome)、蛋白组 (proteome)、代谢组 (metabolome)、调控组 (regulome)，等等。这些以 “ome” 结尾的学科这几年蓬勃兴起，统称为 “组学” (omics)。或许我们觉得既然人是最高等的动物，那我们的染色体数目应该是最多的，基因组应该是最长的。其实不然，比如黑猩猩 (*Pan troglodytes*) 有 48 条染色体，总长 3.3 Gb；小麦 (*Triticum aestivum*) 也有

42 条染色体，基因总数达到 95 000 个，比我们人类还多。

下面是基因科学和技术发展的重要时间表：

年份	发现
1859	达尔文发表《物种起源》（ <i>On the Origin of Sepecies</i> ）
1866	奥地利修道士孟德尔发表“植物杂交试验”（ <i>Experiments on plant hybridization</i> ）一文，提出了孟德尔遗传定律
1866	海克尔（Ernst Haeckel）提出遗传物质在细胞核
1866	英国医生唐（John Down）首次详细阐述唐氏综合征
1871	瑞士科学家米歇尔（Frederick Miescher）第一次从脓细胞中提取到一种富含磷元素的酸性化合物，因存在于细胞核中而将它命名为“核质”（nuclein）。但是，“核酸（nucleic acid）”这一名词 20 年后才被正式启用
1879	德国生物学家弗莱明（Walter Flemming）首次报道细胞的有丝分裂（mitosis），并发现染色质 [苏格兰医生弗莱明（Alexander Fleming），姓氏比他少了一个字母，于 1928 年首先发现抗生素，获得 1945 年的诺贝尔奖]
1879	德国生物化学家科塞尔（Albrecht Kossel）发现核素由蛋白质部分和非蛋白质部分组成，后来进一步证实核酸中的 A、C、G、T 组成成分，他因此在 1910 年获得了基因领域的第一个诺贝尔生理学或医学奖
1894	皮尔逊（Karl Pearson）提出了统计频率分布计算的方法
1900	孟德尔的遗传理论被三位科学家弗里斯（Hugo de Vries）、柯灵斯（Carl Correns）、契马克（Erich von Tschermak）证实
1902	美国生物学家萨顿和鲍维里观察并发现了细胞减数分裂时染色体与基因具有明显的平行关系，并据此推测基因位于染色体上
1905	贝特森（William Bateson）和庞尼特（R. C. Punnett），通过豌豆（pea）实验，发现了连锁（linkage）现象的存在
1908	哈代（Godfrey Harold Hardy）和温伯格（Wilhelm Weinberg）提出了遗传学上的平衡公式（Hardy-Weinberg 遗传平衡定律）
1909	丹麦植物学家和遗传学家约翰森首次提出“基因”这一名词，并且同时提出了“genotype”和“phenotype”
1910	美国遗传学家摩尔根阐述了连锁与互换定律（该定律与孟德尔的分离定律、自由组合定律被称为遗传学三大定律）
1913	斯特蒂文特（Alfred Sturtevant）构建了果蝇（ <i>Drosophila</i> ）的首张遗传图谱（genetic map），又称为连锁图谱（linkage map）

年份	发现
1924	德国细胞学家福尔根（Joachim Wilhelm Robert Feulgen）发现核酸中戊糖有两种：核糖与脱氧核糖。根据含糖的不同，核酸就分为核糖核酸（RNA）与脱氧核糖核酸（DNA）。同年，发明了福尔根染色（Feulgen stain）方法，该方法至今仍是 DNA 定量测定的主要染色方法之一
1927	美国遗传学家穆勒（Hermann Joseph Muller）在《科学》杂志发表“基因的人工诱变”（Artificial transmutation of the gene）一文，首次证实 X 射线在诱发突变中的作用。穆勒是辐射遗传学的创始人，荣获 1946 年诺贝尔生理学或医学奖
1937	杜布赞斯基（Theodore Dobzhansky）发表了遗传进化学上的经典著作《遗传学与物种起源》（ <i>Genetics and the Origin of Species</i> ）
1941	比德尔（George Wells Beadle）和塔特姆（Edward Lawrie Tatum）提出了“一个基因，一个酶”（one gene, one enzyme）的概念
1941	马瑟（K. Mather）提出了多基因学说（polygenes）
1944	埃弗里（Oswald Avery）发现核酸是遗传物质，而不是蛋白质
1946	第一台电子计算机诞生
1950	查加夫（Erwin Chargaff）证明所有的 DNA 中：A 和 T 的数目相等，C 和 G 的数目相等
1952	桑格（Frederick Sanger）完成胰岛素氨基酸序列的测定
1952	赫希（Alfred Hershey）和蔡斯（M. Chase）完成了经典的“混合实验”（blender experiment），噬菌体 DNA 进入细菌细胞并最终导致其破裂，该实验常常被错误地认为是证明 DNA（而不是蛋白质）是遗传物质的决定性实验
1953	沃森和克里克首次提出 DNA 的双螺旋结构模型
1954	伽莫夫（George Gamow）提出了 DNA 含有编码蛋白质的代码（code）
1955	奥乔亚（Severo Ochoa）发现了 RNA 聚合酶（RNA polymerase）
1955	莫顿（Newton Morton）提出了通过 LOD score 来进行连锁分析的方法
1956	科恩伯格（Arthur Kornberg）从大肠杆菌中提取了 DNA 聚合酶（DNA polymerase I）。对该酶的研究促进了后来 DNA 重组技术的创建
1956	华裔生物学家蒋有兴首次发现人的体细胞的染色体数目为 46 条。这 46 条染色体呈 23 对，分别来自父母，第 23 对染色体决定性别，被称为性染色体，XX 为女性，XY 为男性。其他的 22 对被称为常染色体