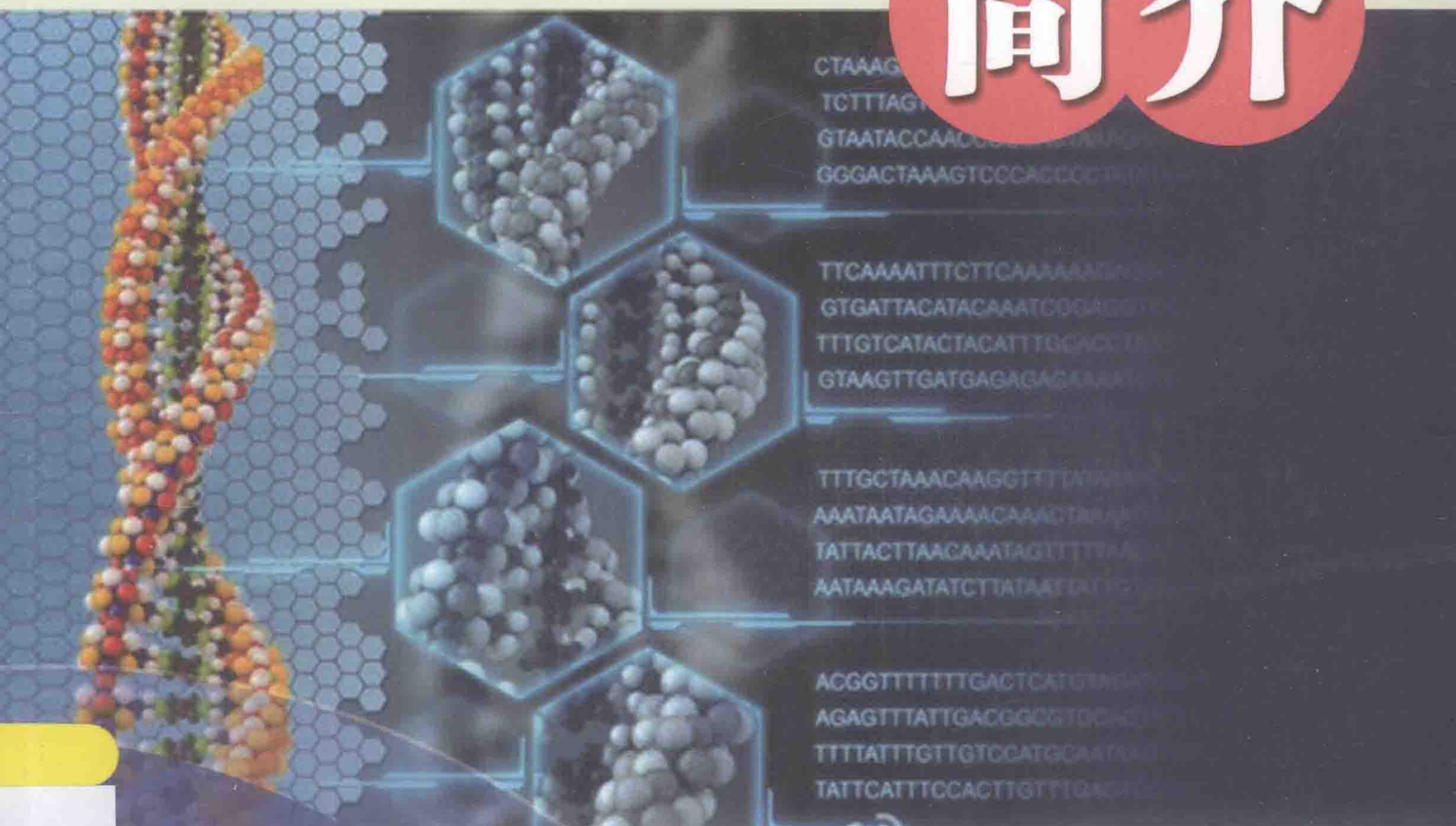


邱幼祥 邱京颖 编译

人类基因组

The Brief Introduction of the Human Genome

简介



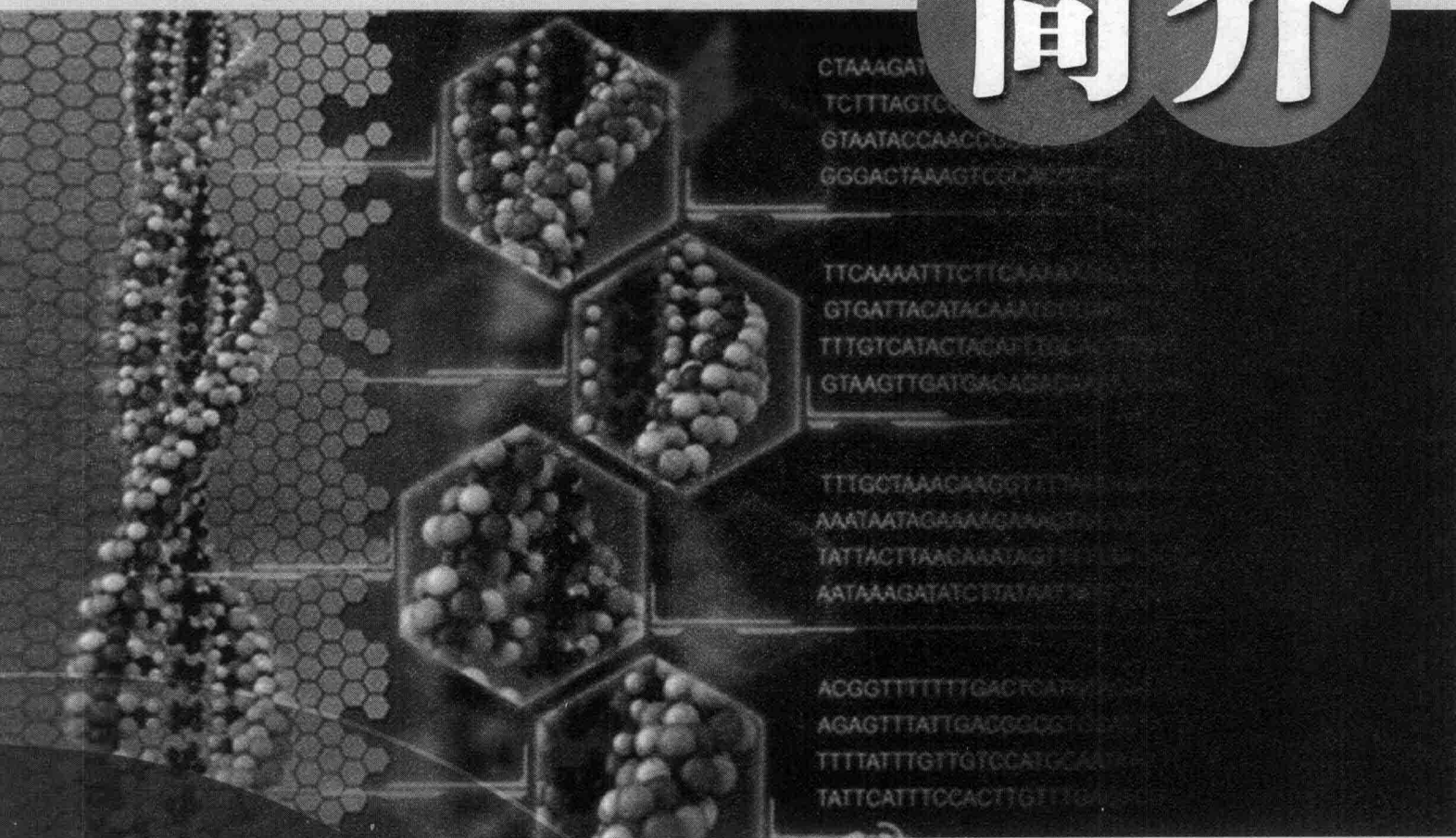
贵州出版集团
Guizhou Publishing Group
贵州科技出版社

邱幼祥 邱京颖 编著

人类基因组

The Brief Introduction of the Human Genome

简介



人类基因组计划国际科学计划

人类基因组计划国际科学计划

人类基因组计划国际科学计划

人类基因组计划国际科学计划

人类基因组计划国际科学计划

人类基因组计划国际科学计划

人类基因组计划国际科学计划



贵州出版集团
Guizhou Publishing Group
贵州科技出版社

图书在版编目(CIP)数据

人类基因组简介/邱幼祥,邱京颖编译. —贵阳:
贵州科技出版社,2014.6

ISBN 978-7-5532-0182-5

I. ①人… II. ①邱… III. ①人类基因组计
划—介绍 IV. ①Q78

中国版本图书馆 CIP 数据核字(2014)第 104923 号

出版发行	贵州出版集团 贵州科技出版社
地 址	贵阳市中华北路 289 号(邮政编码:550004)
网 址	http://www.gzstph.com http://www.gzkj.com.cn
经 销	全国各地新华书店
印 刷	贵阳科海印务有限公司
版 次	2014 年 6 月第 1 版
印 次	2014 年 6 月第 1 次
字 数	500 千字
印 张	19
开 本	889 mm×1 194 mm 1/16
书 号	ISBN 978-7-5532-0182-5
定 价	35.00 元

1990年，美国正式启动了人类基因组计划。随着2004年人类全基因组高精度序列图的完成，各国科学家都在延伸人类基因组计划。

本书介绍了遗传学、基因组以及与之相关的历史事件；提供了与遗传学、人类基因组相关的术语词汇，帮助读者更好地理解人类基因组的工作。本书也介绍了遗传物质的载体——染色体，提供了人类染色体与一些疾病有关的重要基因、由于基因突变所造成的若干种遗传病以及基因定位的最新知识。

本书的特点是知识新颖，重点突出，基本概念清晰，文字简练。本书对于学习有关基因组、人类基因组、遗传学、医学遗传学等生物学学者，医学中诸多学科的本科生、研究生以及从事相关领域的科学工作者、临床医生、医学科研人员等来说都是一本了解人类基因组，掌握当今本学科进展的科学书籍。

第1章 概 述.....	1
第2章 遗传学中的术语和概念.....	12
第3章 染色体.....	73
第4章 具有正常功能的基因.....	90
第5章 遗传病.....	194
参考文献.....	292
后 记.....	294

第 1 章

概 述

Chapter 1

自从人类出现在地球上以来,人类的基因组就和我们相伴在一起,共同走过了很漫长的岁月。人类基因组存在于我们每个人当中,如影随形。然而,就我们人类对于人类基因组的了解而言,虽然其发展十分迅速,但是这段历史还比较短暂。从人类的历史来看,在绝大部分时间里,人类基因组一直都是作为我们生命中不被了解的未知数存在的。事实上,从20世纪30年代起,基因组这个词才渐渐浮出水面,开始了它的应用。在过去10多年间,才启动了构建人类基因组的步伐。而今天,基因组这个词无处不在,如在报纸的大标题上,在杂志中,在最畅销书籍中,在网络的信息中。例如在谷歌搜索中搜索“基因组”这个词,计算机在不到1秒的时间内就有将近800万个条目。仅中国一个国家,现在就有将近6亿多人上网。所以,不难理解为什么今天人们对于现代生物学的发展生命信息——基因组学的认识早已大大超过了人们的想象。

我们知道,生命活动的本质在于生命信息的流动。生命的重要特点之一,其生命信息有它的储存单位及记忆功能。实际上生命信息的储存单位,就是我们说的基因。在绝大部分的生命体中,生命信息的载体是脱氧核糖核酸(DNA),而它的执行单位,总的来说主要是蛋白质。在这里它所使用信息语言是不一样的,前一个是

核酸的语言,而后一个是氨基酸的语言。在生命体中,信息的流动是需要调控机制的。大家知道,调控的第一步就是转录。在转录中,生命信息的语言没有发生变化,都是核酸的语言,只是载体发生了变化,从DNA到了mRNA上面。随后,语言要发生变化,发生转换,就要进行翻译了。而翻译的过程,即mRNA上面的生命语言,变成蛋白质的生命语言的过程。这就是我们所说的“遗传的中心法则”。在单细胞生物体中,细胞的不同部分互相作用,细胞核和细胞质互相作用。在多细胞生物体中,不同的细胞之间、由不同的细胞所组成的组织之间、由不同的组织所形成的器官之间都在发生信息的流动。那么,基因组是什么意思?基因组就是一个生命体的遗传信息的总和。所以在这里我们指的就不是单个基因,而是所有的基因。

“基因组(genome)”是由词汇“基因(gene)”(遗传的一个单位)和“染色体(chromosome)”(绝大多数基因位于其上的物理性结构)两个词结合在一起而衍生出来的专门词汇。我们用基因组这个词来形容一个具有生命活动的有机体的全部的遗传物质。人们也常常用诸如“蓝图”“处方”和“生命天书”这样的词语来描述基因组并且解释它的功能意义,因为基因组相当于文字化地概括了所有使它的主人作为

一个生物而生活在世界上所需要的全部指令。

基因组具有种属特异性。我们谈论小鼠、果蝇和人类基因组就是代表这些有机体的一个复合的有代表性的基因组。尽管在同一个物种中的每个有机体的基因组中，有着这样或者那样的变异，但是每个物种中基因的结构和排列顺序基本相同。举例来说，虽然世界上70亿人中彼此存在着明显的个体差异，但是所有的人都具有23对染色体。在个体水平上，我们能够看到有许多不同的特征，因此每一个人都是唯一的。而这个唯一性就反映在每个个体的基因序列排列的差异上。与此同时，也可能根据在某一种群中，一些成员具有一定的指标而另一些成员却不具有这样的指标而划分成不同的组别。例如，我们选择性别作为指标的话，就会看到在每一组的成员当中有着一些明确可辨的相似特点，同时在这两个组别的成员中会有着同样明显的差异。这个具有明确可辨的差的基因组或遗传基础就是性染色体——所有的女性都具有一对X染色体，而所有的男性都有一条X染色体和一条Y染色体。在人类中，除了上述性别所造成的这种差异以外，所有的人，无论男人还是女人，形成了生物界中单独的一个组别——人种（*Homo sapiens*），和其他的组别如黑猩猩（chimpanzees）或大猩猩（gorillas）有着明显的区别，每一种类型的生物个体彼此之间会有明显的区别的标记，而且在这个类型中的所有个体中都表现一致。每一个个体彼此之间会在一个主要的指标上表现出这样的特点：具有一定大小和大致相同结构的基因组。如果我们再扩大范围向前走一步，黑猩猩就可能和人划成一个组别，而人与鸟类则相距甚远。当我们看得更远，看到整个生命世界中越来越多的成员时，我们的“相似的”和“不同的”组别的比较范围就变得更大了。例如，我们可以把差异明显的植物与动物划分成两大组，也可以把同样是多细胞构成的植物与动物两者合在一起与单细胞

的微生物相比较，划分成两大组，同时它们彼此也有明显的差异。而这种差异也同样反映在基因组水平上。

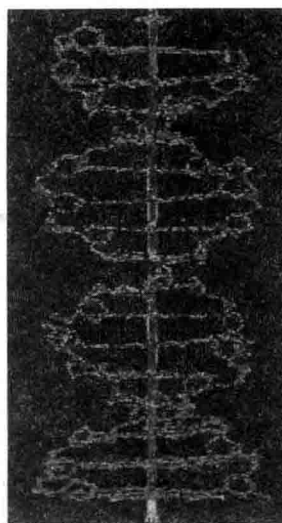


图1.1 DNA双螺旋结构
DNA双螺旋结构的数字模型（本书中所引用的照片和插图，除已标明引用出处以外，其他均引自Acharya, Tara和Neeraja Sankaran, 2005）

从物理学角度看，除了人类基因组是全部由染色体所组成，如同在一个有机体中没有任何额外的遗传物质，所以人们很难认为单独基因组是一个可以鉴别的结构。除去一些病毒以外，所有具有生命活动的有机体，其遗传物质的化学本质是脱氧核糖核酸（DNA）。

沃森（James Watson, 1928 ~ ）和克里克（Francis Crick, 1916 ~ 2004）在1953年应用DNA的X射线衍射成像技术，阐明了DNA分子的双螺旋结构（见图1.1），使得人类对于生命本质的认识大大地提高了一步，使得人们对于生命的研究迈入了分子水平，开启了生命科学的新时代。

DNA是一个复杂的、以一个已知为脱氧核苷酸的多聚体（由许多较小分子所构建的区段构成的一个较大分子）形式为单位所构建的螺旋形的结构。许多脱氧核苷酸相互连在一起，这些脱氧核苷酸构建的区段所形成的序列，就决定了这个DNA所携带的指令。这就好像是一个个词汇连在一起组成句子那样。对DNA来说，没有固定的大小，而且从理论上说，这个多聚体的大小没有尽

头。实际上, DNA分子常常高度地螺旋化, 而且彼此缠绕, 紧密成束。如果DNA分子伸展开来, 每一个单独的细胞中的DNA加在一起, 其长度将超过2 m, 但是它紧密地装配起来, 却能够位于一个细胞内, 而一个细胞要比这个句子结尾的句号小1 000倍之多。从一个物种到另一个物种, 染色体束的数量彼此不同。例如, 细菌和酵母仅仅具有一个较大的染色体, 小鼠基因组有20对染色体, 果蝇有4对, 狗有39对, 而人类具有23对染色体。在植物界中, 花生基因组中有20对染色体, 水稻有12对, 而甘蔗则有40对。

除了位于细胞核这种细胞特殊空间结构中的染色体以外, 绝大多数有机体的基因组还常常包含有一些位于细胞核外的DNA小片段。例如, 人类细胞核中包含有这个基因组99%以上的部分, 但是在线粒体中还有一些很小却很重要的线粒体DNA, 这是一些存在于我们细胞内的发力点或能量发生器。除了红细胞以外, 在我们身体上万亿细胞中的每一个细胞都含有这样的DNA。

那么, DNA和人类基因组是怎样工作的? 绝大多数具有生命活力的有机体都是依靠同样的化学分子(DNA)来传输生命的指令, 既然如此, 为什么它们的外在形态和能力却如此的多样化? 这个问题的回答取决于装载在这个基因组内的指令内容, 也就是DNA构建区段的序列。每一个有机体的基因组由许多称之为“基因”的单位构成, 而这些基因包含着制造携带有生命活动的生物化学分子的各项指令。我们知道, 大多数基因的产物是蛋白质, 几乎我们所有的细胞功能如呼吸、食物的消化, 甚至把DNA密码转录成核糖核酸(RNA)以及翻译成蛋白质, 这一切都依赖于不同蛋白质的活动, 而这些蛋白质是由一些特异性基因编码的。甚至于那些与蛋白质没有明显联系的行为以及其他的特性, 其实也是不同蛋白质彼此之间相互作用的产物。

所有的基因产物合在一起的共同作用决定了

一个有机体长成什么样、有什么样的行为。在整个生命世界中有一些基因很普遍, 例如用来帮助有机体保存这个基因组的基因产物, 去存取它的信息并且把这些信息传递给它的下一代的那些基因。而另一些基因是保证一个物种, 甚至是一个个体具有独特特性的制造者。一个人大约有30 000个基因, 其中有许多基因仍然有待于鉴别并且进一步了解它们的功能。在人类基因组中, 一项令人感兴趣的事情就是这些基因(除去它们的较大的数量以外)只占人类DNA的3%左右, 其余的大多是一些“非编码的”DNA, 这些DNA大大增加了人类基因组的复杂性。我们的确不知道那些非编码DNA的功能是什么。一些科学家猜测当这些非编码的DNA产生作用且产生它们的产物时, 它们就作为基因的“分子开关”, 可能会帮助协调那些基因的活性。

在一个有机体中, 对于某一个性状来说, 一定会存在着某一个基因, 但是, 这并不足以决定一个性状的外在形态。而一个有机体要表现出某一个性状, 这个基因就一定要表达出来, 这就意味着要把它翻译成为对这个有机体有用的形式。虽然我们的每一个细胞中都包含有完整的基因组, 但是, 在某一个时间, 在单独一个细胞中只有一定的基因得到表达。在一个有机体生命的不同发育阶段, 是由在不同细胞中的这个发育基因的表达, 才可以制造出不同于脚的耳朵或者能从一粒种子中发育出一棵蚕豆植株来。显而易见, 这个精细又具调整功能的分化是由单独一整套遗传指令的发布而获得的, 就是这种分化使我们变成完整的、复杂的有机体, 同时又具有特异性的生命功能的组织和器官。

迄今为止, 我们所谈到的绝大部分内容都是关于在一个有机体的个体中基因组在做什么, 而这些只代表了它的一半功能。今天来看看基因组的另一半功能, 即把遗传信息成功传输到后

代中（也就是遗传）。在人们开展遗传学研究之前，对于生物的有关性状是如何从父母向他们的后代传递的，有着不同的理论。几个世纪以来，人们一直把血液当作是遗传的基础。在18世纪，有关遗传的流行理论是先成论，其提出了“在一个世代的生殖细胞中包含了一个已经预先形成的小型化的有机体个体，而且它们可以传递给下一代”的观点，所谓发育就是这些小型化个体生长的简单化。18世纪后期至19世纪前期，有许多人相信每一代生物都遗传那些由它们的上代通过对环境的应答而发展起来的性状。人们甚至这样解释，（在遥远的过去）因为离开地面比较近的树叶已经吃光了，一代又一代的长颈鹿为了能够吃到树叶而不得不拼命地向上伸长脖子，这样年复一年，一代又一代，长颈鹿就有了长长的脖子。这个用进废退的观点，就是拉马克（Jean Baptiste Lamarck, 1744 ~ 1882）学说。到了19世纪后期，达尔文（Charles Darwin, 1809 ~ 1882）的物竞天择、自然选择的理论才渐渐替代了拉马克学说，达尔文理论认为进化来自作用于所遗传的性状的自然选择的结果。到了20世纪，随着遗传学的深入发展，遗传学开启了新的时代。1953年，沃森（James Watson）和克里克（Francis Crick）成功地应用DNA的X射线衍射成像技术，阐明了DNA分子的双螺旋结构，并提出了著名的“DNA分子的双螺旋模型”。从此人们把研究生命现象的焦点移到了分子水平，达尔文的理论也被翻译成基因的语言。每一代生物的基因组DNA都被重复地传递到后代中，当序列发生改变或者突变，并且传递到下一代时，就获得了新的特点。人类基因组的复制和传递在每一个个体都保持着唯一性的同时，也保证了一个物种的繁衍。

我们现在已经了解到，控制遗传过程的关键在DNA的双螺旋结构中。DNA中碱基的序列保存着功能的信息，并且需要一代代地传递下去。双螺旋结构的一条链完美地与另一条链互补，这样当两条

链分开时，细胞的机制就能够通过被称为复制的过程来复制原始DNA的每一条链。这个能进行自我复制的非同凡响的能力，就保证了几乎所有具有生命活动的有机体进行遗传的能力。

1. 历史的回顾：从遗传到人类基因组计划

绝大多数的遗传学课本都把现代遗传学的诞生定在1866年，这是因为在这一年，奥地利神父葛力格·孟德尔（Gregor Mendel, 1822 ~ 1884）发表了他使用豌豆进行研究的论文《植物杂交试验（Experiments on Plant Hybridization）》。在这项试验中，他选用了农作物豌豆来研究，诸如豌豆豆皮的颜色等一些性状从上一代传递到下一代的方式。孟德尔的这篇文献的重要意义在于，这是第一个系统观察一个有机体的遗传形式的记录。当然，也有人提出不同的观点，因为遗传学的复杂性以及遗传学在各个不同领域中的影响，所以不能把遗传学的诞生归结为时间上的某一点。除了应用在植物学（或者更多的是在植物育种）以外，遗传学在各种各样而且几乎毫无联系的学科，如医学、昆虫学及免疫学等学科上都颇有贡献。此外，尽管遗传学在20世纪才被作为一个独立科学，人类却早已以多种多样的方式推测了遗传的本质——物理特性和行为性状两者是从亲代到子代的传递。

孟德尔的学说在他的那个时代获得了人们的认同，直到今天，其崇高的地位仍然是不可撼动的。当年（1866），他的这篇《植物杂交试验》论文是在一个相当不知名的地方杂志[《Versuche über Pflanzenhybriden》（《布吕恩自然科学研究协会会刊》）]上发表的，由于印量少、发行量小，所以影响不大，以至于这篇文章发表后的几年中也没有引起别人的注意。连孟德尔他自己都没有活着看到他的这个

理论获得成功。到了1900年,欧洲3位自然科学家 Carl Correns (德国植物学家,第一代遗传学家,1864~1933)、Hugo de Vries (荷兰植物学家,第一代遗传学家,1848~1935)和Erich von Tschermak (奥地利农艺学家,1871~1962)彼此独立进行了在遗传特性方面地探讨,并且先后在同一刊物《德国植物学会杂志》上发表了他们的实验结果。尽管他们并没有根据孟德尔的设计和实验,但是都与30年前孟德尔所做出的结论几乎一致。在孟德尔的豌豆实验及其相关论文“沉睡”35年之后,终于在1900年得到了科学界的承认和肯定,开创了经典遗传学的时代。虽然在这个时候,孟德尔早已撒手人寰,但是所有科学家都给予了孟德尔应得的荣誉,并认为这3位科学家的发现证实了孟德尔定律。1900年的这个“再发现”标志着遗传特性研究辉煌时代的开始。

与此同时,医学对于遗传的理论已经做出了显著的贡献。在18世纪,一个名叫Pierre de Maupertuis (1698~1759)的法国外科医生注意到在一些家庭成员中有多指症的倾向——在每个手或脚上有多于5个手指或脚趾的情况发生。人们也会注意到像血友病和色盲这样的病症,会以一个非常特异性的性别方式遗传。在血友病和色盲这两种情况中,男孩患病的频率要比女孩高许多,内科医生注意到这些男孩来自这样一些家庭:其父亲或是外公,或是叔伯早已受此之累。Archibald Garrod爵士(1857~1936)是世界上第一个在人类性状上应用孟德尔定律来加以分析的人。这位英国内科医生,在1908年提出具有特异性酶缺失的尿黑酸症(alkaptonuria)是可以遗传给下一代的看法。他使用了“新陈代谢中的先天错误”来描述像白化病、戊糖尿症(pentosuria)、胱氨酸尿症(cystinuria)等同类型遗传病。

今天,我们会很自然而然地把像“遗传学”

和“基因”这样的术语与遗传联系起来,但是在20世纪之初,英文词汇中根本就没有这些术语。英国生物学家William Bateson (1861~1926)在他所著的《Mendel's Principles of Heredity》(1902)一书中提出了用名词“遗传学”作为遗传的学科名,或者作为应用在性状起源的术语,还提出了有关等位基因、杂合子和纯合子等名词或结构。William Bateson的主要实验贡献是在1905年,他发现了一定性状可以在一起遗传的事实(这个特点,被称为遗传连锁,后来证明是基因物理定位的重要工具)。Bateson是研究植物遗传的第一批科学家之一。同时,“基因”这个词来自希腊文“生产、产仔”这个意思,是德国人Wilhelm Johannsen (1857~1927)在他的“Johannsen, W. (1909) Elemente der exakten Erblchkeitslehre. Gustav Fischer, Jena.”这篇文章中,第一次把基因(gene)这个词作为遗传的单位来使用,将其用于遗传学。在那个时候,“基因”这个词还是一个相当抽象的概念,只是把孟德尔在早期称为“遗传因子”的词汇简化,而没有什么实际的物理意义。而对基因的物理和化学性质的研究都扎根于科学的其他领域,并且在几十年后才开始进入到孟德尔研究的主要工作之中,才使得基因一词成为遗传学中极为重要的概念。

1866年除了作为孟德尔的重要论文发表的年份之外,也是遗传学历史上十分重要的一年,因为在这一年,德国生物学家、哲学家Ernest Haeckel (1834~1919)推出了这样一个假说:“一个细胞中的遗传信息是通过它的细胞核来传递的”。10多年后,Oskar Hertwig (1849~1922)于1875年确认了这个假说,他指出细胞核对于细胞分裂的重要性。而在下一个10多年间,Walter Flemming (1843~1905)用一系列实验发现了细胞核内的染色体,并且在他同事的帮助下,描述了染色体在细胞分裂中

的行为。

最初把有关遗传的研究和染色体结合在一起科学家是美国遗传学家摩尔根(Thomas Hunt Morgan, 1866~1945)。从1909年开始,摩尔根与一些科学家一起,在哥伦比亚大学里一间后来被称为“苍蝇室”的实验室中,用小型的果蝇(*Drosophila melanogaster*)作为模式生物来研究遗传机制。今天我们所了解的关于遗传学及其机制中的许多成果,都来自摩尔根的研究。其中,基因在染色体中呈线性排列这个很重要的理念,就出自这个研究。

摩尔根在果蝇的科研工作中,获得了许多关于基因的物理信息以及基因在许多性状的遗传方面的作用。因为基因包含许多指令或信息,而且被假设可以由它们形成蛋白质。而在当时,基因被认为是具有可以储存信息的唯一的分子。对于基因的化学性质,或者说是遗传分子的化学性质的发现当推Oswald Avery(1877~1955)、Colin Macleod(1909~1972)和Maclyn McCarty(1911~2005)这些洛克菲勒大学(后来的洛克菲勒医学研究中心)的医学细菌学家,他们当时正在研究已知可引起肺炎的细菌的转化过程。在1944年他们发表了一篇论文,有效地说明了遗传信息的携带者是DNA,而第一个鉴别DNA的是Friedrich Miescher(1844~1895),其早在1869年就已经完成了DNA的鉴别,当时他用“核素”命名。从一开始他们的观点就受到了抵制,但是在此后10年间,绝大多数科学界人士逐渐接受了这个观点,并且都不约而同地把焦点集中到如何了解及认识DNA的结构,他们察觉到这是在分子水平上认识遗传学的关键所在。在1953年,James Watson(1928~)和Francis Crick(1916~2004)使用了由他们的同事Rosalind Franklin(1920~1958)和Maurice Wilkins(1916~2004)所提供的DNA的X射线衍射成像技术,阐明了DNA分子的双螺旋结构。

在接下来的时间里,综合认识了不同研究领域信息,这为有关基因和基因组的研究提供了更为美妙的画卷。人们开始认识生命不同分子(DNA和蛋白质)之间相互作用的本质特点。1958年,Crick提出用“中心法则”来描述在生命系统中的信息流。他指出生命的遗传信息是在细胞内的生物大分子(DNA、RNA和蛋白质)之间转移、流动的。他揭示了遗传信息的流动方向和传递规律。在1966年,Marshall Nirenberg(1927~2010)、Har Gobind Khorana(1922~2011)和Robert William Holley(1922~1993)决定研究在DNA上的核苷酸碱基排列和在蛋白质上的氨基酸排列的相互关系,就这样成功地揭示了“遗传密码”。20世纪70年代至80年代,科学家首次得到了能够接近人类基因组的新技术。在这个方面的贡献特别重要的是,由Friderick Sanger(1918~)和Walter Gilbert(1932~)发展的用来测定DNA序列的技术,以及发现了能够把不同来源的DNA片段切开和修补的酶。接下来,Kary Bank Mullis(1944~)在1983年成功发现了聚合酶链式反应(PCR)技术,能让科学家在几个小时内从小的样本中获得数十亿DNA分子,通过几个级别的放大作用,测序方法的自动化加速了大量遗传信息的获得,与快速的数据加工和改进技术相结合,使科学家获得了能够在单个基因水平上进行操作的能力。这些技术大大地武装了科学家。与此同时,认识并医治患病病人的希望来临了,在其他的星球上改进人类生活质量的预言也将随之来临。

1986年,意大利裔美国人Dulbecco在《科学》杂志上发表了一篇短文《肿瘤研究的转折点——人类基因组研究》。Dulbecco在这篇文章中提出,如果我们想更多地了解肿瘤,那么我们从现在开始必须关注细胞的基因组。在遗传学知识得到巨大进步的影响下,美国能源部和美国国

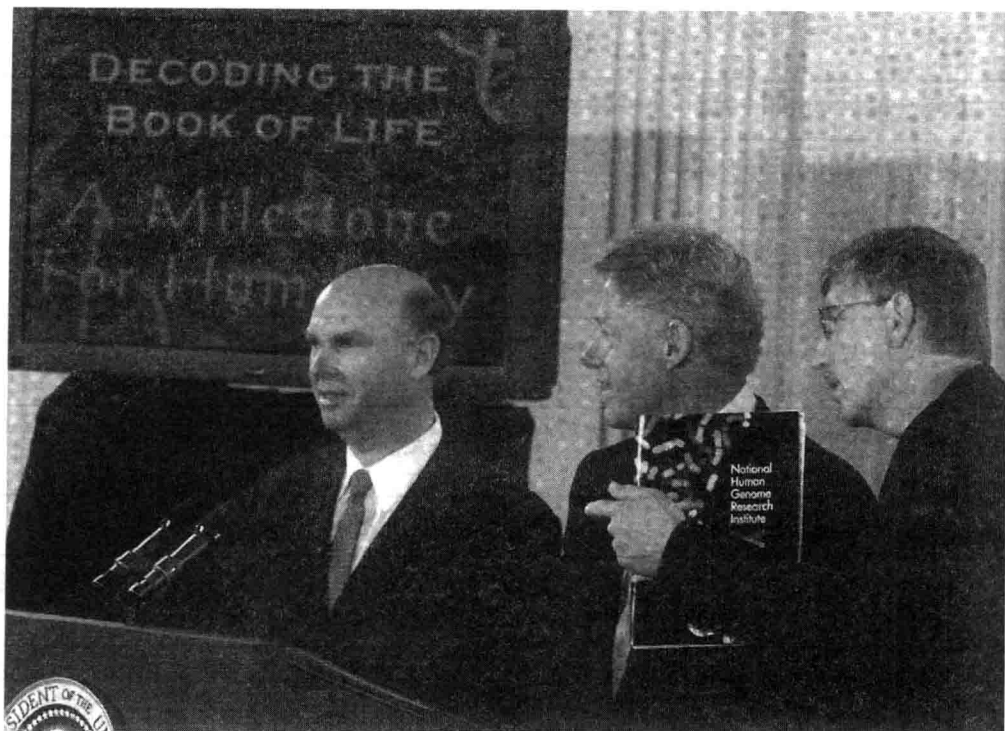


图1.2 人类基因组序列草图宣布完成

2000年6月26日Francis Collins、Craig Venter和时任美国总统比尔·克林顿（Bill Clinton）在白宫纪念会上宣布完成了人类基因组序列草图。

立卫生研究院确立了为高达31亿碱基对的人类基因组进行测序的宏大计划。1990年，美国国会通过并正式启动了这个计划。这个计划雄心勃勃，计划从1990年起，共耗时15年，到2005年完成DNA全序列的测定。当时计算费用的依据是测一个碱基对大概需要1美元，共投资30亿美元。人类基因组计划正式启动，基因组的革命也随之来临。乘着信息技术革命的东风（如果没有强有力的计算机和自动测序仪的支持，那么测序大概要用几十年才能完成），人类基因组计划以令人瞩目的速度飞速地发展。2000年4月，公众资金支持的人类基因组计划和私人企业塞雷拉基因组公司一起宣布了人类基因组序列草图正式完成（见图1.2）。美国国立卫生研究院人类基因组研究中心的前负责人Francis Collins自豪地说，这项计划“所用时间比预期的要少，而且所用费用低于计划”。这是在人类历史上一个实实在在的创举和里程碑，是科学家为基因组下一个时期的到来

而大声呼唤，也是获得令人激动不已的关于人类存在的新认知。

表1.1揭示了在遗传学和人类基因组中一些主要的理论概念和实验性研究的重大事件，也是给读者提供有关在这个领域，特别是20世纪60年代后迅速发展的一些资料。

2.基因组和未来：社会和伦理学方面的思考

人类（智人，*Homo sapiens*）基因组是人类遗传信息的完整集合，储存在细胞核中的23对染色体上的DNA以及线粒体内的小DNA分子中。单倍体的人类基因组（包括卵子和精子细胞）含有30亿个DNA碱基对，而二倍体基因组（在体细胞中所发现的）是单倍体DNA含量的2倍。虽然人类基因组已经通过DNA测序并被完全地读出，但尚未充分地理解清楚。许多基因虽然可能已经被确定下来，但仍有许多工作要做，其中包括完

表1.1 遗传学和基因组学的重大事件

1866	孟德尔根据用农作物豌豆进行的杂交实验而提出了遗传学基本理论
1868	Friedrich Miescher从脓细胞的细胞核中分离出新物质，为其取名叫核素（后来称之为DNA）
1882	Walter Flemmings在显微镜下观察蝶螈的幼体，并且认识到在细胞的细胞核中小小的丝状结构就是染色体
1900	Carl Corren、Hugo de Vries和Erich von Tschermak发现了孟德尔定律，创立了现代遗传学的新纪元
1908	Archibald Garrod爵士发表了《新陈代谢的先天错误》一文，并且提出在白化病中因酶的缺失所造成的隐性遗传
1905	Nettie Stevens和Edmund Wilson各自独立地描述了性染色体的行为
1910	Thomas Hunt Morgan用果蝇进行研究，发现了一些遗传决定的性状是性连锁的，而且还发现并且确认了基因位于染色体上
1918~1926	Herman Muller通过X射线来诱导突变，他认为基因通过产生它自己的内部改变来构成进化的基础
1941	George Beadle和Edward Tatum提供了酶和基因之间联系的实验证据
1944	Oswald Avery、Colin Macleod和Maclyn McCarty认为DNA是引起肺炎球菌转化的遗传物质
194~1949	Linus Pauling及其同事一起追踪引起镰刀形红细胞贫血症的原因，找到了异常的血红蛋白（把这个疾病归为“第一个分子病”）。此后，又提出这个异常与在血红蛋白基因上的一个特异性突变有关
1953	由James Watson和Francis Crick共同提出了DNA的双螺旋结构
1956	Joe Hin Tjio（蒋有兴）和Albert Levan确立了人类染色体是46条的正确数目
1958	Meselson和Srahl的实验确认了DNA的重要特点是半保留复制
1958	Arthur Kornberg从大肠杆菌（ <i>E. coli</i> ）中纯化出DNA聚合酶I（形成DNA的酶）
1960	Francis Crick提出了中心法则，并用中心法则来描述生命系统中的信息流
1961~1966	Marshall Nirenberg、Gobind Khorna和Robert Holly提出了“遗传密码”（即核苷酸三联体密码子和氨基酸之间的关系）
1970	Peter Duesberg和Peter Vogt发现了在病毒中的第一个癌基因（诱导癌变的基因）
1972	Hugh McDevitt找到了在外来物质中控制免疫应答的基因，他提出有一些基因可能会引起一些疾病的易感性
1973	Bruce Ames提出了一个鉴别破坏DNA的化学物质的实验，这个实验对于致癌物的鉴别有广泛应用
1975~1977	由Walter Gilbert和Frwderick Sanger各自独立并使用不同方法发展了DNA测序技术
1976	J. Michael Bishop和Harold Varmus鉴别了动物染色体上的癌基因，并说明它们在表达上的变化导致癌的发生

续表

1980	美国高等法院制定了生命形式能够取得专利的法律
1982	制造重组人类胰岛素的基因技术是重组DNA技术，而重组人胰岛素也是将此技术推向市场的第一个产物
1983	Kary Mullis和在Cetus合作的其他人发明了聚合酶链式反应（PCR）技术，这项技术可以在几个小时内产生数十亿DNA拷贝
1987	美国国立卫生研究院确定建立GenBank，这是用于存储核酸序列的国家级计算机数据库
1989	发现了在人类7号染色体上的基因，并由Francis Collins和Lap Chee Tsui（徐立之）完成测序，它编码引起囊性纤维化跨膜传导调节因子（CFTR），如果该基因突变，会导致囊性纤维化
1990	人类基因组计划开始进行人类基因组的定位
1992	发表了第一个关于21号染色体和Y染色体的物理图
1993	发现了遗传性结肠癌、亨廷顿病、多动症（hyperactivity）、Lou Gehrig病、Alzheimer病、糖尿病基因
1995	发表了3号、11号、12号、16号、19号和22号染色体的物理图
1997	成功地克隆了第一个哺乳动物——克隆羊“多莉”
2000	人类基因组计划协作组和塞雷拉基因组公司宣布了人类基因组的序列草图

全阐明它们的蛋白质以及RNA产物的生物学功能。

最近的研究结果表明，大量的非编码DNA与在基因组中相关的生化活动，包括基因表达的调控、染色体结构的构成和控制表观遗传的信号有关。

人类基因组计划产生了第一个完整的人类个体的基因组序列。截至2012年，成千上万的人类基因组序列被完全定位测序，还有许多序列在比较低的分辨率下被定位。由此所得到的数据可以在世界范围内应用到生物医学科学、人类学、法医学及其他科学分支。人们还有一个广泛持有的期望，希望基因组研究能促进疾病的诊断和治疗的进步，并能在生物（包括人类）进化上推出新的见解。

单倍体人类基因组包含有差不多20 000个编码蛋白质的基因，明显少于预期。蛋白质编码序列只占人类基因组很小的零头（大约1.5%），其余部分都与非编码的RNA分子、调节DNA分子以及内含子有关，没有功能的序列到现在还没有完

全搞清楚。

像任何一个新的发现或者新的技术一样，基因组学的迅速发展对于人类来说是一把双刃剑。一方面，不同基因的发现可以救治病人，改进农作物产量，改进人类甚至是其他星球的生命的质量。另一方面，科学的迅速发展又可能让人们对于回想过去而担惊受怕，同时也让人们对于未来的一些不确定因素充满忧患和担心。尽管有些情况不过是许多科学虚构，而其他的一些还处在知识的襁褓之中，仅仅是初试啼声。这样相当数量的相伴而生的关切，应该由致力于基因组科学的科学家和广大的社会公众加以考虑和讨论。知识带来了巨大的力量，这是事实，但是必须要把力量放在负责任的人手中。我们常常有可能失掉某些洞察力，但是我们必须小心牢记，不能掉以轻心，特别是在基因组学时代，这一点是重中之重。

新的基因组科学所暴露出来的最大的危险之

一是一些被称为“遗传决定”的可能性，也就是说，基因决定了我们的一切。遗传决定论认为认知和发展是由先天的遗传基因所决定的。人的发展过程只不过是这些内在的遗传因素自我展开的过程，环境的作用仅在于引发、促进或延缓这种过程。其观点创始人是高尔登，他的典型论调是“一个人的能力是由遗传得来的，它受遗传决定的程度，正如一切有机体的形态及躯体组织受遗传决定一样”。这一论点是需商榷的，因为基因在我们自己的鉴别中起了很大的作用，甚至在生命当中起了主要作用，但它们并不是唯一的作用。我们周围的环境以及我们的基因与环境的相互作用，在决定一个给定的基因的最终表达结果中起了相当重要的作用。

对于人类基因组计划的许多顾虑是来自“关于遗传信息到底能不能应用”这一点，这些遗传信息已经接近于我们人类存在的核心部分，并且能够暴露出我们的基本弱点。而且，基因组中隐藏着“麦卡锡主义”幽灵一直存在并潜伏着。对于一个社会来讲，健康护理和健康保险正在给一些利益集团施以压力，人们较多关心的一个问题是保险公司可能利用遗传信息来否定对特殊疾病的保险，而不顾个人存在遗传易患病的体质的可能性。同样，当失业率正在上升时，有一些雇主会根据一些人的遗传学外形特点歧视工作申请人。在美国和一些地方的社会活动家组织已经把关切上升到并高于发展中国家对基因组和遗传技术的错误应用的可能性方面。例如，已经有人建议跨国公司在人种学组别上使用门诊检验来了解在药物应答和遗传组成之间的相互作用，而不提供给参加者以适当的补偿，或者不提供给这些实验组分享那些来自研究而出现的潜在的利益，哪怕这些利益只是一点点。这样的游戏规则是真正可能的，还是仅仅口头上虚假的恐吓呢？当一个人考虑到像人种或种族外形等歧视现象时，也会情不自禁地发出种种疑问。对于许多人来说，声名狼藉的Tuskegee梅毒实验是美国在1932年至

1972年间以免费治疗梅毒为名，在亚拉巴马州对399名非洲裔男子进行的一系列人体实验。美国总统克林顿在1997年5月14日正式代表美国政府对受害者作出道歉以及赔偿。这是对当年非洲裔美国人中的梅毒患者所采取的“治疗措施”的一个道歉，为了研究这个疾病在少数族裔的发病过程，以科学的名义而滥用权利的道歉。那么，人们不禁要问，人类基因组计划会不会对我们的基因再现过去的这种固执和歧视呢？

值得注意的是，有人认为基因治疗也是一把双刃剑。当1990年开始进行基因治疗的第一个临床试验时，人们对于基因治疗的潜在作为，对于各种各样的遗传病提供全新的奇迹般的治疗方式寄予厚望。然而，从那以后，只有很小的进步。

1999年，杰西·基辛格（Jesse Gelsinger，1981~1999）是一个18岁的男孩，他是公开进行基因治疗临床试验后确定死亡的第一人。他患有一种以鸟氨酸转氨甲酰（基）转移酶（OTC）的缺失为特征的罕见遗传病，这个病是肝上的X连锁遗传病，主要表现为蛋白质裂解的产物——氨的代谢的丧失。他没有遗传疾病，患有此病只是一个基因突变的结果，因此并不严重。他的一些细胞是正常的，所以他只要限制饮食并且服用特殊药物就可以生存。后来，他加入到宾夕法尼亚大学的临床试验，旨在开发一种用于治疗有严重疾病的婴儿的药物。1999年9月13日，基辛格被注射携带有纠正基因的腺病毒载体。1999年9月17日下午14点30分，他死了。显然，他遭受了用来运送正常的OTC基因到他的细胞的腺病毒载体大规模免疫反应，引发并导致多器官功能衰竭、脑死亡。另一个事件发生在2002年，对一个由于具有X连锁的严重免疫缺陷疾病发展成类似白血病症状而前往就医的小孩，同样是在使用一种法国基因治疗试验时，令人遗憾的偶然事件又发生了。美国食品药品监督管理局对这些令人不愉快的结果几乎瞬间就做出了相应的反应，暂停了基

因治疗在血液干细胞中使用反转录载体。在这些事件的阴影下，时至今日，美国食品药品监督管理局依然没有可以应用于医学的人类基因治疗任何切实可行的措施。

生殖技术是关于基因组潜在危险的另一个争论得十分激烈的领域。现今，许多人常常有这样的选择（如果他们能够接受的话），就是使用植入前的遗传学诊断来检查他们成功怀孕的胚胎中一些遗传病的遗传风险。尽管这可能是很好的意图，但是这项技术的确会破坏胚胎，而且对于其他的选择会产生带有一定遗传外形的“一个设计好的婴儿”。在2004年，有一对英国夫妇生下了一个小男孩，这个小男孩已经经过了遗传学方面的选择，他的骨髓与他需要进行骨髓移植的哥哥恰好准确无误的匹配。尽管许多人把这个技术当成是在科学上十分显著的进步，但其他人则把它当成是针对某些多余部分而人为设计婴儿的危险的第一步。因为这个选择过程只是根据对人类基因组上非常小的一个零头的认识来完成的。诸如前面所叙述的，在我们的基因组中只有少于3%的部分是编码序列，而其他部分，我们把它们称之为“垃圾DNA”，这是一些我们了解较少的物质。如果我们能够了解这一点的话，我们就能够保持冷静，包括有时我们基于很少信息作出的关于我们孩子的决定。

在克隆羊多莉成功诞生之后，这项克隆技术极大地刺激了人类社会，许多科学家开始尝试克隆各种动物，甚至是克隆人的胚胎。人类克隆的可能性吸引了许多医学方面的关注。这个可能性也给伦理带来了一波新的冲击。胚胎干细胞基因至今仍然是另一个颇受争议的话题，之所以有一些话题人们喜爱它，是由于它在诊断上的潜力，另一些人反对它是因为人类胚胎的必然死亡。所有的这些议题已经触动了家庭、科学研究工作者、政府、保险公司、制药公司、法律界和雇主等，各个方面都在进行着激烈的辩论。基因组计

划是离不开伦理学问题的，人类基因组DNA的序列是全人类的共同财富，应该造福于全人类。对基因组基础数据的垄断，将给人类利益和科学发展带来不良的后果。我们每一个人在一定的条件下都可能是弱者，所以反对基因歧视实际上是维护全人类每一个成员的共同利益。

重要的是，当这个科学跑在前头时，对于人类社会，伦理学和合法政策都要尽量把基因组学的利益最大化，而把它的风险最小化。人类基因组计划在1990年开始实施的时候，在美国也建立了一个用来研究它所带来的伦理、合法和社会意义（ELSI）的项目分支。这个叫做ELSI的机构每年在联邦资金上有1 000万美元开支的预算，是世界上对于生物伦理研究获得资助最多的单位。许多国家政府、学术中心和国际机构，如世界卫生组织，都已经确定了发掘基因组学潜在风险和利益的研究项目，并且制定了使利益最大化、风险最小化的政策。

不管有无潜在的缺陷和危险，人们毫不怀疑基因组这个领域已经在我们生活中产生了很大的实际效果，而且在未来仍将继续如此。它已经被完全地转化成科学家研究生物学的一个方式。现在，研究人员有了大量的遗传数据，并且有了许多新的研究这些信息的技术。今天，进行疾病遗传突变的鉴别比起过去更为简单，而且能节省更多时间，具有公众性。现代医疗可以使用强有力的软件程序来处理基因组数据，如DNA测序机器以及能够追踪在一个处理中成千上万基因的表达的DNA微阵列（芯片）技术。遗传信息这个财富正无一例外地被转变成健康利益，虽然其作为预防措施、诊断以及治疗非传染和传染性疾病的新的医疗工具，但在世界范围内改进人们的生命质量仍然需要再走上若干年。

第2章

遗传学中的术语和概念

Chapter 2

为了便于大家阅读和查找,在下面所介绍的遗传学的概念和术语中,都有它们的中文、英文名词,并且所有的名词术语都按其英语词汇的首字母顺序排列。

1. 等位基因 ALLELE

等位基因,又称对偶基因,是一对同源染色体上同一基因座上控制着相对性状的两个不同形式的基因。在细胞有丝分裂时一对同源染色体上同一位置的两个基因位点是一一对应排列的,细胞遗传学里称其为等位基因。任何一个基因,由于在它的DNA序列中,可能会出现变异,因而具有多种形式。这个由同样的基因所构成的不同形式,称为等位基因。

同一个基因座上的两个等位基因的差异可以小到一个核苷酸的碱基对,也可以大到较大的基因片段。当我们说“突变”基因时,事实上我们说的是一个等位基因,而这个等位基因的序列与其他正常基因有较多的差异,以至于它的功能都发生了改变。等位基因,除了它们的序列表现有微小差异以外,这些等位基因基本上正常。等位基因变异中最普通的(即在这个种群中分布最为广泛的基因版本)被称为野生型等位基因。因为人类染色体是以成对的形式存在的,每一个人在

染色体上任何一个给定的位置(已知为基因座)上的基因有两个等位基因的形式。等位基因中的一个位于由母亲遗传而来的染色体上,而另一个位于从父亲遗传而来的相应位置或相应基因座上。如果在单独一个个体的给定基因座上的两个等位基因是相同的,也就是说他的这个等位基因是纯合的,如果它们不一样,那么,这个个体就是杂合的。尽管一个个体能够在一个给定基因座最多具有两个等位基因,但在人类中,绝大多数的基因具有超过两个等位基因分布。人们常常追踪等位基因的变异及种群遗传学来认识疾病或其在各种群体中的历史性迁移形式。

2. 早现遗传 ANTICIPATION

早现是指在一个家庭的几代人发生的遗传病中所观察到的遗传现象,在这样的家庭中,这种综合征会在连续几个世代中较早出现,而且(在年轻时)常常表现较严重。通常认为这种现象可能与遗传的不稳定有关,也可能是由于在受影响的基因的重复核苷酸序列的组成上发生突变所致。早现最富有戏剧性的例子在肌强直性营养不良病例中,在亨廷顿病(Huntington's disease, HD)和弗里德赖希共济失调中可观察到较低程度的早现现象。这些疾病是由受影响的基因三核