



普通高等教育“十二五”规划教材

Principles of Gene Engineering

基因工程原理

文铁桥 主编



科学出版社



普通高等教育“十二五”规划教材

基因工程原理

文铁桥 主编

科学出版社

北京

内 容 简 介

全书共8章,系统介绍了基因工程基本原理与操作技术,主要包括科学巨匠的贡献、常规分子技术、工具酶、表达系统、基因克隆、基因功能研究策略、动植物基因工程、基因信息分析技术等。注重基础知识与发展前沿的结合,紧扣基因工程发展脉络,体现了该领域的新技术、新成就与新动态。本书引用大量科学研究案例,简明扼要,深入浅出,注重科学性、先进性、实用性和启发性。

本书适用于高等院校生命科学领域本科生、研究生的教材,也可作为医学、环境科学等相关学科教学参考书,也可供相关研究人员参考。

图书在版编目(CIP)数据

基因工程原理/文铁桥主编. —北京:科学出版社,2014.8

普通高等教育“十二五”规划教材
ISBN 978-7-03-041558-5

I. ①基… II. ①文… III. ①基因工程—高等学校—教材 IV. ①Q78

中国版本图书馆CIP数据核字(2014)第174636号



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

南京展望文化发展有限公司排版

江苏省句容市排印厂印刷

科学出版社出版 各地新华书店经销

*

2014年8月第一版 开本:A4(890×1240)

2014年8月第一次印刷 印张:11

字数:310 000

定价:36.00元

《基因工程原理》编辑委员会

主 编 文铁桥

编 委 (按姓氏笔画排序)

王 凯 王 娇 开国银 文铁桥

冯瑞丽 吴春霞 陈 峰 查笑君

顾志敏 黄 海 龚一富 崔丽洁

前 言

20世纪70年代基因工程问世成为生命科学史继DNA双螺旋后的又一标志性里程碑。经过40余年的发展历程,基因工程技术已经成为生命科学与医学领域的主要研究工具,并渗透到了农业、食品、化工、环保等不同领域。借助基因工程技术,人们初步看到了一个从微观到宏观的生物世界,迈出从认识自然到改造自然的雄劲步伐。基因工程技术是现代生物技术的核心,是衡量一个国家科学与技术发展水平的重要标志之一。

由于基因工程技术的迅猛发展,新技术、新方法、新概念层出不穷,在这样一种背景下,本书编写中融入了新的知识,如ZFN、TALEN技术与原理、光遗传载体与调控等,这些都是目前国际前沿基因操纵技术;同样由于我们所处这个时代在不停地进步与发展,培养学习的能力就显得十分重要,甚至比掌握一门具体技术本身更重要,基于此,本书在一定程度上加强了对问题的分析与思考。考虑到基因工程与有关学科的交叉,本书还安排了基因信息分析技术一章,帮助读者对基因海量数据的把握和功能分析,在一定程度上是对传统基因工程的拓宽,论述明确,有指导作用。

自2002年以来,我一直承担着本科生基因工程原理课程的教学,从课程建设的角度,先后历经了上海市重点课程、上海市精品课程的建设与探索,并获得上海市优秀教学成果奖。与此同时,我也承担着包括国家“973”计划、国家自然科学基金、上海市重点基础研究项目等一些科学前沿研究项目,从科学研究的视角审视教学内涵。我深切体会到教学与科研并重对于培养学生“坚实基础、了解前沿、解决实际”的关键与重要,希望编写一本教材有助于提升教学内涵。由科学出版社组织,成立了本书编写委员会,作者全部来自教学科研一线教师,具有教学与科研实际的经历与经验,形成了本书科研融于教学的特色。本书由上海大学、上海交通大学、上海师范大学、山东师范大学、宁波大学、浙江师范大学等高校教师集体编写而成。全书共八章,第一章、第六章、第七章(转基因动物)由文铁桥编写,第二章由开国银、崔丽洁、龚一富编写,第三章、第七章(转基因植物)由吴春霞编写,第四章由查笑君编写,第五章由顾志敏、黄海编写,第八章由陈峰编写、王凯修改。冯瑞丽、王娇对全文进行了校稿、排版及文献编排工作。全书由文铁桥统筹与修改。

在本书编写过程中,上海大学分子神经生物学实验室研究生吴婕、顾知立、吴翌臻、李霄、蒋萌参与部分章节研究工作,张仕卿、郭健健、干林华、骆光洪参与部分章节校稿,在此一并致谢。

由于编者水平有限,内容上难免出现各种错误,由于不同作者语言表达方式、措辞风格不尽相同,虽经统筹修改,还是可能存在文字上的差异,真诚欢迎读者批评指正。

文铁桥

2014年3月于上海

目 录

第 1 章 绪论	1
1.1 基因工程的发展历程	1
1.1.1 基因概念提出	2
1.1.2 基因结构与功能	2
1.1.3 基因操作关键技术突破	4
1.1.4 基因工程问世	4
1.2 基因工程大科学与大科学	5
1.2.1 海量基因信息	6
1.2.2 复杂基因网络	7
1.2.3 基因工程系统	7
1.3 基因工程研究的意义	7
1.3.1 揭示生命现象规律	7
1.3.2 基因工程与疾病治疗	8
1.3.3 基因工程产业	8
思考题	10
第 2 章 基因工程基本技术原理	11
2.1 核酸的提取与鉴定	11
2.1.1 基因组 DNA 的提取	11
2.1.2 RNA 的提取	14
2.1.3 质粒 DNA 的提取	15
2.1.4 核酸的凝胶电泳	18
2.1.5 变性梯度凝胶电泳	20
2.1.6 核酸的定量与纯度的测定	21
2.2 PCR 技术	21
2.2.1 PCR 的原理	22
2.2.2 PCR 的操作	22
2.2.3 PCR 衍生技术	24
2.2.4 PCR 定点诱变	27
2.3 核酸序列测定	30
2.3.1 Sanger 双脱氧链终止法	31
2.3.2 Maxam - Gilbert 化学修饰法	32
2.3.3 Roche 454 测序	32
2.3.4 Illumina 测序	33
2.3.5 ABI SOLiD 测序	34
2.3.6 HeliScope 测序	36
2.3.7 单分子测序	36
2.3.8 纳米孔测序技术	37
2.4 分子杂交技术	37

2.4.1	Southern 杂交	37
2.4.2	Northern 杂交	39
2.4.3	Western 杂交	40
2.4.4	菌落原位杂交	41
2.4.5	荧光原位杂交	42
	思考题	43
第 3 章 基因工程工具酶		44
3.1	限制性核酸内切酶	44
3.1.1	限制和修饰现象	44
3.1.2	限制性核酸内切酶的类型	45
3.1.3	II 型限制性核酸内切酶的命名	48
3.1.4	影响限制性核酸内切酶活性的因素	48
3.2	DNA 聚合酶	50
3.2.1	大肠杆菌 DNA 聚合酶	50
3.2.2	大肠杆菌 DNA 聚合酶 I 的 Klenow 片段	51
3.2.3	T4 DNA 聚合酶	51
3.2.4	T7 DNA 聚合酶(修饰的 T7 DNA 聚合酶)	52
3.2.5	实验室常用的耐热 DNA 聚合酶	52
3.2.6	逆转录酶	53
3.3	DNA 连接酶	53
3.3.1	DNA 连接酶的类型	53
3.3.2	DNA 连接酶的特点	54
3.3.3	连接反应的机制	55
3.3.4	影响连接反应的因素	55
3.4	核酸修饰酶	57
3.4.1	末端脱氧核苷酸转移酶	57
3.4.2	T4 多核苷酸激酶	57
3.4.3	碱性磷酸酶	58
3.4.4	核酸外切酶	59
3.4.5	单链 DNA 内切酶	60
3.4.6	甲基化酶	61
	思考题	63
第 4 章 基因工程载体与表达系统		64
4.1	载体概述	64
4.1.1	载体研究背景	64
4.1.2	载体共性特征	65
4.2	克隆载体类型	65
4.2.1	质粒载体	65
4.2.2	噬菌体载体	70
4.2.3	噬菌粒载体	73
4.2.4	病毒载体	73
4.2.5	人工染色体克隆载体	74
4.3	原核表达系统	74

4.3.1	原核表达载体高效表达的元件	74
4.3.2	大肠杆菌表达系统	77
4.4	真核生物表达系统	82
4.4.1	酿酒酵母表达系统	82
4.4.2	巴斯德毕赤氏酵母表达系统	82
4.4.3	酵母表达系统的应用	83
	思考题	83
第5章	目的基因分离与鉴定	84
5.1	基因文库法获取目的基因	84
5.1.1	基因文库的构建	84
5.1.2	cDNA文库的构建	86
5.1.3	筛选基因文库获得目的基因	87
5.2	PCR技术分离目的基因	87
5.2.1	已知基因全长序列	87
5.2.2	已知基因部分序列	88
5.3	图位克隆获得目的基因	89
5.4	转座子标签法获得目的基因	90
5.5	电子克隆	91
5.6	目的基因的鉴定	91
5.6.1	表型检测法	92
5.6.2	核酸探针杂交法	93
5.6.3	PCR筛选和序列鉴定法	94
5.6.4	物理特性检测法	94
	思考题	95
第6章	基因功能研究技术	96
6.1	常规分析技术	96
6.1.1	基因沉默	96
6.1.2	基因敲除	104
6.2	高通量分析技术	108
6.2.1	基因芯片	108
6.2.2	蛋白芯片	109
6.3	大分子相互作用分析技术	109
6.3.1	免疫共沉淀	110
6.3.2	染色质免疫沉淀技术	110
6.3.3	RNA免疫沉淀法	111
6.3.4	酵母双杂交	111
6.3.5	光遗传	114
	思考题	115
第7章	动植物基因工程	116
7.1	转基因植物	116
7.1.1	植物基因工程的常用方法	116
7.1.2	转基因植物的筛选与鉴定	122

7.1.3	提高外源基因在植物中表达水平的策略	125
7.1.4	转基因植物的应用	128
7.2	转基因动物	131
7.2.1	动物基因工程的常用方法	131
7.2.2	转基因动物的筛选与鉴定	136
7.2.3	提高外源基因在动物中表达水平的策略	137
7.2.4	转基因动物的应用	139
	思考题	146
第8章	基因信息分析技术	147
8.1	常用生物分子数据库	147
8.1.1	核酸数据库	147
8.1.2	蛋白质序列数据库	151
8.1.3	其他常用生物分子数据库	152
8.2	软件分析应用	155
8.2.1	用 BLAST 进行局部序列比对	155
8.2.2	用 CLUSTAL 进行多序列比对	156
8.2.3	用 Primer Premier 进行引物设计	156
8.2.4	限制性内切酶分析	157
8.2.5	用 JASPAR 查找转录因子结合位点	158
8.2.6	EMBOSS 工具包	158
8.2.7	其他常用生物信息学软件	160
	思考题	160
	主要参考文献	162

提 要

基因是存在于 DNA 上的核苷酸序列,具有特定的结构、功能以及遗传特征。从基因到基因工程的一个多世纪的研究结果,充分展示了人类从认识自然到改造自然的过程。基因工程技术在造福于人类的同时,也给人类无限遐想的空间,基因工程的迅猛发展将人类带入实现梦幻的年代。

基因工程(gene engineering),又称遗传工程,一般指利用分子生物学的手段,将不同生物的遗传物质按人们的意志或需要,进行基因的定向改造、突变与重组,从而使生物体的遗传性状发生变异。基因工程操作涉及 DNA 特异位点酶切、载体连接、不同 DNA 片段重组、转入受体细胞、重组子筛选等系列过程。基因工程作为一门学科已被广泛地接受与认同。基因工程最显著的特征是能够打破种属的界限,在基因水平上改变生物遗传性,并通过工程化手段为人类提供有用的产品及服务。

1.1 基因工程的发展历程

生命科学发展的历史长河中,经历了从描述生物学、实验生物学到创造生物学三个发展阶段。描述生物学阶段可以追溯穴居的远古人类,通过观察与文字记载,积累形成对生命现象的描述;到了 18~19 世纪,人们对动物、植物和微生物做了大量的形态结构的观察、描述和分类的工作。从 19 世纪中叶至 20 世纪中叶,由于显微镜技术的迅速发展,人们可以看到细胞和细胞器的精细结构,可以通过设计实验来研究生命现象的规律,不再仅仅满足于生命现象的描述。这一世纪,可以说是生物学飞速发展的百年,最重要的里程碑有两个:1944 年 Avery 首次证实了 DNA 是遗传物质;1953 年 Watson 和 Crick 的 DNA 双螺旋模型提出,完成了对 DNA 功能与结构最本质的认识,标志着实验生物学阶段达到顶峰。人类花了近 20 年时间探索对 DNA 分子的操作技术,终于于 20 世纪 70 年代,人们在体外实现了 DNA 分子重组,意味着生物学进入了一个可以按照人们的意志对遗传物质操作的年代,这就是创造生物学阶段。20 世纪末,人类意识到生命科学众多、重大的科学问题有望在 21 世纪解决,一个“21 世纪是生命科学的世纪”在人们的期盼中到来。10 多年过去了,一个面临大科学时代的大学科产生了,这就是系统生物学,它利用数理、计算机科学原理和基因工程技术,从系统、整体层面回答生命科学问题,这种不同学科高度交叉与融合,将生命科学推到一个前所未有的高度。



(Wilhelm Johannsen, 1857~1927)

1.1.1 基因概念提出

基因(gene)是丹麦科学家 Wilhelm Johannsen 在 1909 年提出来的。Wilhelm Johannsen(1857~1927)是丹麦的植物学家、植物生理学家和遗传学家。他首先提出了 phenotype(表型)和 genotype(基因型)概念。1909 年,当这部书翻译成德文(*Elemente der exakten Erblchkeitslehre*)时, Johannsen 用了“gene”一词。此时的基因仅是一个推理的概念,没有确凿实验证据。

1.1.2 基因结构与功能

在遗传学发展的早期阶段,人们意识到基因的重要性,但没有证实基因的物质结构。19 世纪 60 年代初, Mendel 以豌豆为实验材料表明生物的某种性状是由遗传因子负责传递的,遗传下来的不是具体的性状,而是遗传因子即基因,从而提出了遗传的分离定律和自由组合定律。这是具有历史意义的发现,可惜在当时并未获得认同,直到 1900 年,他的研究被荷兰科学家 Hugo de Vries、德国科学家 Carl Correns 和奥地利科学家 Erich von Tschermak 各自再次发现。



(Thomas Hunt Morgan, 1866~1945)

最先从结构上赋予基因内涵的是 Thomas Hunt Morgan (1866~1945),美国遗传学家、现代遗传学之父。以果蝇为实验材料,通过遗传突变,确立了基因位于染色体上。发现位于同一染色体上的基因之间的连锁遗传特性,将多种突变基因定位在染色体上,制成染色体图谱,即基因的连锁图。1933 年获诺贝尔生理学或医学奖。Morgan 出版书籍 22 部,发表科学论文 370 篇,正是他的研究使果蝇成为当时主要的模式生物,时至现代,果蝇仍然是几类主要模式生物之一。Morgan 对基因研究的贡献在于将代表特定性状的特定基因与某一条特定染色体上的特定位置联系起来,使基因不再是抽象的符号,而是在染色体上占有一定空间的实体,从而赋予基因以物质的内涵。中国遗传学家谈家桢院士就读于美国加州理工学院,在 Morgan 实验室攻读研究生,获博士学位,后来成为中国现代遗传学的奠基人之一。



(Frederick Griffith, 1879~1941)

20 世纪 40 年代以前,遗传信息被认为存在于细胞的蛋白质中,究竟什么是遗传物质不能定论。1927 年 Frederick Griffith 首次完成细菌转化实验,证实肺炎链球菌之间可以转化,但原因并不知道,猜测可能存在一种转化因子。肺炎病菌有两种,一种是光滑型肺炎双球菌,有荚膜、菌落光滑且有毒,这种菌通常外包有一层黏性发光的多糖荚膜,它是细菌致病性的必要成分,引起肺炎;另一种是粗糙型肺炎双球菌,无荚膜、菌落粗糙且无毒。将光滑型肺炎双球菌注入小鼠体内,使小鼠致死;将粗糙型肺炎双球菌注入小鼠体内,小鼠存活;将光滑型肺炎双球菌加热杀死后,再注入小鼠体内,小鼠不会死亡;但将加热杀死的光滑型肺炎双球菌与粗糙型肺炎双球菌一起注入小鼠体内,小鼠死亡。这说明粗糙型肺炎双球菌转变成了致死的光滑型肺炎双球菌。

在 Frederick Griffith 的研究基础上, Oswald Theodore Avery(1877~1955)找到了细菌转化的原因。Avery 是最早的分子生物学家之一、免疫化学先驱。1944 年, Avery 等在 *Journal*

of *Experimental Medicine* 发表了“脱氧核糖型的核酸是Ⅲ型肺炎球菌转化要素的基本单位”即 DNA 是细菌的转化因子,第一次证明了 DNA 是遗传物质。从加热杀死的光滑型肺炎双球菌中提取 DNA,与粗糙型肺炎双球菌混合后,再注入小鼠体内,小鼠死掉。Avery 的实验说明带有有毒性的和能使荚膜形成的信息分子 DNA 整合进了无毒菌种的染色体里,使它转化成了有毒的菌种。从一个供体菌得到 DNA,通过一定的途径,授予另一个细菌,从而使后者的遗传特性发生改变,Avery 的进一步实验证明了这种转化因子就是 DNA。

1953 年 Watson 和 Crick 在 *Nature* 发表了 DNA 分子结构模型,从结构上解释了 DNA 作为遗传物质的遗传信息传递机制。James Dewey Watson(1928~),美国分子生物学家,与 Crick 一同利用 X 射线衍射的数据,建构了 DNA 模型。1962 年, Watson 与 Crick、Wilkins 因为对 DNA 结构的研究而共同获得诺贝尔生理学或医学奖。1965 年出版了划时代的教科书《基因的分子生物学》(*Molecular Biology of the Gene*),1968 年出版《双螺旋》(*Double Helix*),同年开始兼任位于纽约长岛的冷泉港实验室(Cold Spring Harbor Laboratory)的总管,并将研究方向转移到癌症。

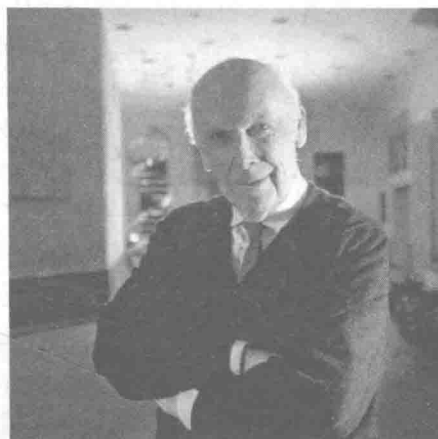
Francis Harry Compton Crick(1916~2004)是英国生物学家、物理学家及神经科学家。Crick 最重要的成就是 1953 年在剑桥大学卡文迪许实验室与 Watson 共同发现了脱氧核糖核酸(DNA)的双螺旋结构。此外,他提出了众所周知的“中心法则”(central dogma)概念,认为在细胞中的遗传信息的流动基本上是单向的,即从 DNA、RNA 到蛋白质,尽管随后的研究表明 RNA 也可逆向到 DNA,但“中心法则”的提出对生命科学研究起到了推动作用。1961 年 Crick 又提出了三联体遗传密码,这样将 DNA 分子的结构与生物学功能有机地统一起来,也为揭示基因的本质奠定了分子基础。他最后的研究集中在神经生物学,试图探索人类的意识。2004 年因大肠癌病逝于美国加州大学圣地亚哥分校,临死前还在修改研究论文,将终身献给了科学事业。

到 1959 年,关于对 DNA 的几个重要的认识已经完成,如 DNA 作为遗传物质的证实,DNA 双螺旋结构的提出,DNA 半保留复制机制。但 DNA 如何指令蛋白质的表达,RNA 在此过程有什么作用还不知道。Nirenberg 和他的学生 Matthaei 建立和完善了大肠杆菌的无细胞翻译体系,回答了这些问题,完成了遗传密码的破译。

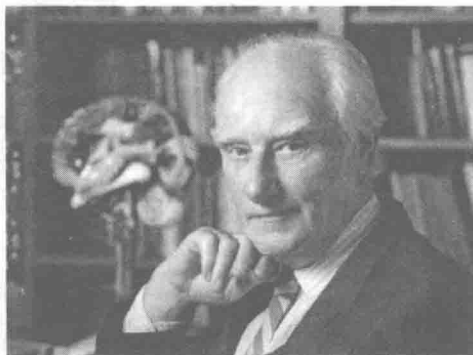
Marshall Warren Nirenberg(1927~2010)是一位犹太裔美国生物化学家与遗传学家。他们合成了一条完全由尿嘧啶核苷酸组成的 RNA,将此放入来自大肠杆菌制备的无细胞体系,再加入 20 种氨基酸,其中一种是放射性标记的,在不同的实验中依次改变标记的氨基酸,结果发现合成的带有放射性标记的蛋白质为苯丙氨酸所组成,因此知道了 RNA 中 UUU 可以编码出的氨基酸为苯丙氨酸,首次破译了遗传密码。1961 年 8 月,国际生物化学大会在莫斯科举行,



(Oswald Theodore Avery, 1877~1955)



(James Dewey Watson, 1928~)



(Francis Harry Compton Crick, 1916~2004)

Note

Note



(Marshall Warren Nirenberg,
1927~2010)



(François Jacob, 1920~)

Crick 说服了大会主持人,邀请 Nirenberg 报告他们的研究,Nirenberg 的报告震惊了科学界。Nirenberg 很快便受到科学界极大关注,在接下来的几年,他们分别破译了赖氨酸密码(AAA)、脯氨酸密码(CCC)。从 1961 年到 1962 年的两年,被称为“译码竞赛”年代,Nirenberg 在(National Institutes of Health, NIH)工作,面临纽约大学医学院、诺贝尔奖得主 Severo Ochoa 大型团队的竞争。因遗传密码的破译,Nirenberg 与 Har Gobind Khorana 和 Robert W. Holley 共同获得 1968 年诺贝尔生理学或医学奖。Nirenberg 后来聚焦于神经科学、神经发育的研究,2010 年因癌症去世。

较早阐明基因结构、功能及调控的是 François Jacob,他是一位犹太裔法国生物学家,他与 Jacques Monod 通过不同的大肠杆菌乳糖代谢突变体来研究基因的作用,发现了乳糖利用过程中基因的调控角色,即乳糖操纵子。细菌和其他细胞可以通过调节关键代谢酶的水平,或这些酶的活性,以适应或应对外部条件。例如,如果细菌在含有乳糖而不含葡萄糖的培养基中,它必须适应性地做出调整,吸收乳糖,并能将乳糖切割成为葡萄糖和半乳糖,再将半乳糖转换成葡萄糖。执行这些步骤需要酶来完成,当所形成的产物足够时,酶的结构构象发生改变,停止酶的产生,起到开关的作用。1961 年,Jacob 和 Monod 认为细胞中酶的表达可能存在反馈抑制,以使酶的表达不至于过量。他们发现有阻遏物(repressor)能直接与 DNA 结合,从而控制该基因的表达。在有乳糖存在条件下,阻遏物与乳糖结合,而再也不能与 DNA 结合,转录抑制解除。用这种方式,构建一个反馈环路,允许消化乳糖的蛋白质进行生产。由于发现乳糖操纵子调节机制,Jacob 和 Monod 与 André Lwoff 共同获得了 1965 年的诺贝尔生理学或医学奖。

1.1.3 基因操作关键技术突破

1968 年,H. O. Smith 等分离到一种酶,能够打断 DNA 分子,而且总是在 GTTAAC 序列中切开双链 DNA,表明此种酶具有酶切位点识别的专一性,也称为限制性,这就为剪断 DNA 分子提供了“手术刀”。由于这种酶是从流感嗜血杆菌(*Haemophilus influenzae*)分离得到的,所以称为 HindII,这是第一个分离到的限制性核酸内切酶。一个基因执行功能需要转录成 RNA,翻译成蛋白质。在原核生物中,DNA 转录出可以直接翻译成蛋白质的 RNA,但真核生物由于内含子的存在,RNA 需要剪接加工,才能成为成熟的 RNA。RNA 不稳定,很容易被降解,要研究一个真核生物基因功能非常困难。1963 年,Baltimore 发现肉瘤病毒 RNA 能够逆转录成 DNA,给了分子生物学家新的启迪,分离真核生物 RNA 逆转录成 DNA,就可以用限制酶进行选择切割,表达所获得的真核基因,然而所获的 DNA 存取成为问题。1970 年代,Stanley Cohen 首次将外源基因放入质粒,便可进行表达,这样质粒可以作为基因的运载体,携带人们希望表达的基因。上述三个重要技术的突破,奠定了基因操作技术基础。

1.1.4 基因工程问世

千百年来,人类进行基因改造主要通过人工选择,在人类生活的早期是自然选择,随着人类文明进步和科学技术的发展,出现了诱变选择。自 20 世纪 70 年代以来,人类可以通过直接操

纵 DNA 分子进行物种改造,出现了真正工程意义学科——基因工程。“基因工程”一词首次出现于 1951 年出版的 Jack Williamson 的科幻小说“龙岛”,比 Watson 和 Crick 发现 DNA 分子双螺旋结构还早两年。1972 年 Paul Berg 首次完成了第一个重组 DNA 分子,将猴病毒 SV40 与 λ 病毒的 DNA 相结合,因此与 Gilbert 和 Sanger 分享了 1980 年诺贝尔化学奖。1973 年 Herbert Boyer 和 Stanley Cohen 将抗生素抗性基因插入质粒转化大肠杆菌产生了第一个转基因生物。一年以后,Rudolf Jaenisch 将外源 DNA 注入早期小鼠胚胎,产生了世界上第一例转基因小鼠,成为国际上公认的转基因科学的开拓者,此后 Jaenisch 聚焦于转基因小鼠研究肿瘤和神经疾病。1976 年 Herbert Boyer 和 Robert Swanson 成立了世界上首家基因工程公司(Genentech),次年该公司利用大肠杆菌生产出人类蛋白质(生长抑素)。1978 年 Genentech 公司宣布生产基因工程人胰岛素。1980 年,美国最高法院裁定基因工程产品可以申请专利,因此,由细菌生产的人胰岛素于 1982 年获得美国 FDA 批准。1986 年第一例植物基因工程产品——除草剂耐受烟草问世。1992 年,中国将转基因植物商业化。1994 年,具有较长货架期的基因工程番茄获准上市,同年欧盟批准抗除草剂(溴苯腈)的基因工程烟草商业化,在欧洲出现了第一个商业化的转基因植物。到 2009 年,11 个转基因作物商业化种植于 25 个国家和地区,种植面积依次是美国、巴西、阿根廷、印度、加拿大、中国、巴拉圭和南非。

2010 年,J. Craig Venter 研究所的科学家宣布,他们已经创建了第一个合成细菌基因组,并把它添加到一个不含有 DNA 的细胞。由此产生的细菌,名为 Synthia,这种微生物能够进行正常的生命活动,是世界上第一个人造生命形式。

1.2 基因工程大学科与大科学

基因工程问世以后,受到生命科学、医学等领域的高度关注,基因工程技术很快应用于动物、植物、微生物基础与应用的研究,形成了以基因工程为基础的大学科,不同学科科学家首次在基因层面达成共同语言,阐释基因的本质,呈现了前所未有的大科学时代。这个时代具有两个突出特征:跨学科和跨国界。随着科学研究的迅猛发展,生命科学与其他学科如数学、物理、化学、计算机科学在研究兴趣点上形成聚焦。淡化专业,科学家突破自己所学专业的壁垒,不再受限于狭小的专业之间的隔离,充分利用各个学科的优势,解决所面临的重大科学问题。理工结合,文理交叉,大学科格局在科学研究需要的期盼中形成。从 SCI 发表论文统计来看(图 1-1),近 20 年来,与生命科学交叉发表论文数量急剧增加,尤其是近十年,2012 年发表的论文大概是 20 世纪 90 年代的 7 倍,每年引文数急速攀升,显示学科交叉的高度关注。这些论文主要分布在工程类、计算机科学、环境科学领域。从统计数据看,美国几乎占到总数的一半,

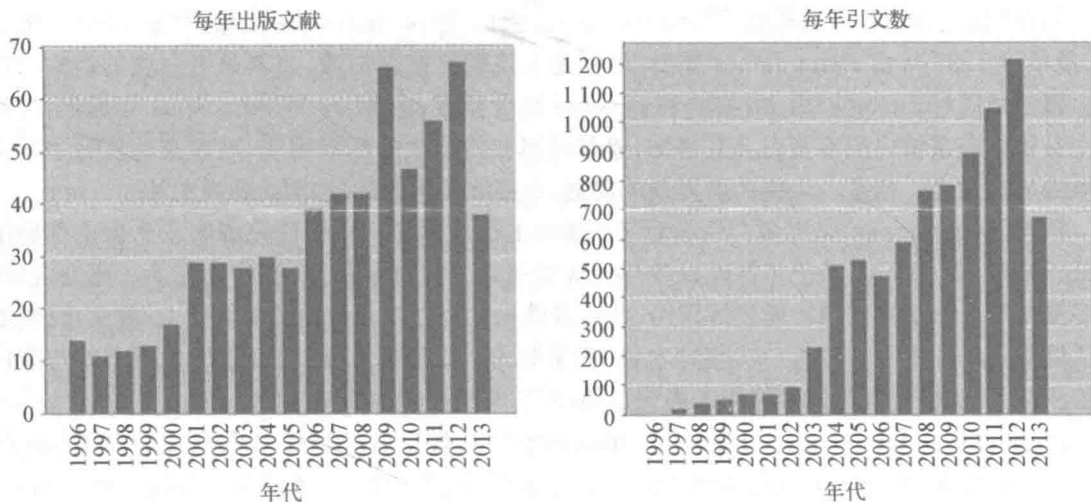


图 1-1 跨学科发表的 SCI 论文统计

Note

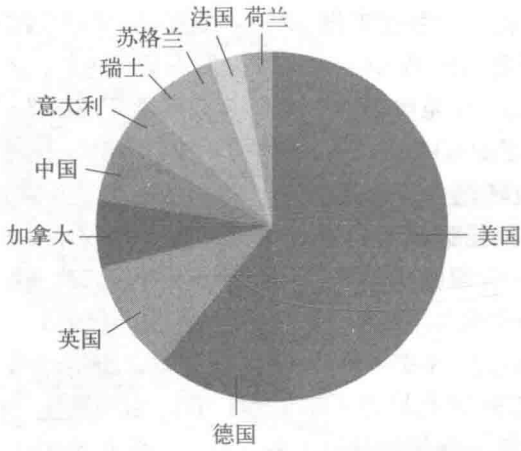


图 1-2 2012 年发表生命科学交叉论文前十国家排名

其次为德国、英国、加拿大、中国(图 1-2)。

1.2.1 海量基因信息

20 世纪 80 年代, Sanger 等创建了 DNA 序列测定方法, 有力推动了人类对基因序列的认识。随后毛细管电泳测序法, 采用的是与 Sanger 生物化学测序方法完全不同的原理, 与荧光成像技术相结合, 使测序能力提高到一个新的水平。进入 21 世纪, 随着人类基因组计划的完成, 获得一大批包括人类基因组在内的数据。2005 年, 新一代测序技术问世, 分析速度骤增, 基因数据海量释放, 科学家的传统基因信息概念受到挑战。庞大的数据信息, 使得用户无法将初始图像数据进行分类存档, 而必须交给相关公司, 利用软件对数据

进行读取, 然后才能对数据进行保存。此外, 研究人员还需要拥有一个足够强大的计算机平台, 以便将测序得到的许多短小基因片段进行组合, 拼接成基因组外显子。要处理这些海量数据, 研究人员需要学习、掌握甚至开发能够应用的生物学信息分析软件, 从而完成从基因信息获取到基因信息分析与研究的过程。

我们正处于生物信息爆炸的时代, 各种生物基因组测序先后完成, 随后的蛋白组学、糖组学、转录组学、代谢组学全面展开, 为推进各个相关领域的研究提供了庞大的数据资源。现在已经完成的基因组测序有 人类基因组、酵母基因组、线虫基因组等, 存储于世界各地的生命科学、医学、农业相关的信息中心、研究机构和大学, 供研究人员查阅分析。此外, 小鼠、河豚、水稻、果蝇、大肠杆菌等各种模式生物基因组数据库或基因组信息资源都可以在网上找到。在英国 Roslin 研究所的 ArkDB 包括了猪、牛、绵羊、山羊、马等家畜以及鹿、狗、鸡等动物的基因组数据库, 斑马鱼等鱼类基因组数据库以及玉米、大麦、高粱、菜豆等农作物基因组数据库可以分别在美国、英国、日本有关机构查找。

蛋白质序列数据库的雏形可以追溯到 20 世纪 60 年代, 这显然早于基因数据库, 因为当时已经可以测定蛋白序列, 所以有蛋白序列的积累, 只是没能称为“库”。从 20 世纪 60 年代到 80 年代, 美国国家生物医学研究基金会将搜集到的蛋白质序列和结构信息编辑成“蛋白质序列和结构图集”(atlas of protein sequence and structure), 主要用来研究蛋白质的进化关系。1984 年, 蛋白质序列数据库(protein information resource, PIR) 诞生, 1988 年美国、日本、德国合作成立了国际蛋白质信息中心(PIR-International)。

与基因组学相似, 蛋白质组学(proteomics)是蛋白质(protein)与基因组学(genomics)的组合, 最早于 1997 年由 Peter James 提出, 目标是大规模研究蛋白质, 尤其是蛋白质的结构与功能。蛋白质组(proteome)由 protein 和 genome 组合而成, 由 Marc Wilkins 1994 年提出, 指在一个有机体或系统内所有蛋白及其修饰, 随着时间和特定生命活动需要, 或在某种压力之下蛋白质的表达改变。这是一个庞大而复杂的信息, 直接涉及生命活动, 因此备受关注。

糖组学(glycomics)是继蛋白组学后又一重要的研究领域。主要研究细胞或个体全部糖链的分子结构、表达调控、功能多样性以及与疾病的关系, 与蛋白组学研究密切相关。研究表明, 很多蛋白需要糖基化修饰才能发挥作用, 意味着糖基化位点就是蛋白的功能位点, 因此, 阻断或激活糖基化的发生意味着蛋白功能的改变, 是理想的药物靶标。此外, 也可借此研究糖与蛋白质的相互关系, 探讨蛋白质糖修饰的表观遗传问题, 解析可能的疾病机制。

代谢组学(metabolomics)由英国 Jeremy Nicholson 教授首次提出, 主要研究生命体接收外界刺激或病理生理变化、基因突变而产生的体内各种代谢物变化, 反映出生物体代谢活动结果。通过代谢组学研究既可以发现生物体在受到各种内外环境扰动后的应答不同, 也可以区分同种

不同个体之间的表型差异,广泛应用于疾病诊断、医药研制开发、营养食品科学、毒理学、环境学、植物学等与人类健康护理密切相关的领域。常用核磁共振波谱和质谱两大分析技术表征代谢物,如与色谱联用,其方法的灵敏度和准确度可以进一步提高。与基因组学和蛋白质组学相比,代谢组学的研究侧重于对代谢组分的共性、特性和规律的认识,从最终引起代谢产物发生相应的改变,提示机体病理生理过程变化,这些代谢物异常,则是疾病的标志,因此可以用来作为疾病的判断依据。

1.2.2 复杂基因网络

基因行使其功能可以在不同水平上,有些是单基因,有些是多基因,从生命现象的复杂性出发则更多是基因的相互作用,以及由此构成的基因网络。从目前来看,建立基因网络,阐明它们之间的关系,最有效的方法之一就是数学理论与计算机技术相结合,将基因间相互作用整合成网络,通过建立模型来研究基因上下游关系、关键调控点以及动力学性质。最初随机网络模型被用来描述基因网络结构,随后在规则网络与随机网络相结合的基础上,建立了小世界(small-world)网络模型。1999年,Barabase等发现了真实网络的无标度性质,与真实基因网络具有很大程度的一致性。可以从基因数据库如KEGG查询一些基因网络以及调控转录因子调控信息,结合实验数据,或通过实验对基因网络进行验证,完成理论网络的修正,为认识网络的复杂性提供切入点,也为药物开发提供借鉴。

1.2.3 基因工程系统

随着基因研究的深入开展,单个基因层面的研究已经不能满足科学研究的需要,在很多情况下需要从系统的层面解释基因或基因间的复杂关系。如一个基因的突变,可能引发多个基因表达异常,继而导致整个基因网络的紊乱、基因表达失调。事实上,多基因的突变,可能更为常见,更为复杂。即使不是基因突变,由于环境因素的影响,也会使得基因之间、蛋白质之间相互作用系统异常。这种失衡的结果,可能导致不同层面上错误的结局,出现mRNA降解,蛋白翻译障碍,蛋白网络系统不能工作等。因此,现代的基因工程具有系统学特征,已经明显区别于早年(问世年代)的基因工程,这种长足的发展,使得基因工程更具活力和魅力,对研究人员知识储备、背景、能力的要求更高。

随着基因工程技术的飞速发展,基于基因工程系统的药物开发备受关注,人们试图从基因网络、蛋白网络、基因相互作用、蛋白相互作用、基因与蛋白相互作用、网络调控来寻找可能的候选药物,这种药物明显的优势是作用于关键调控点,因此药效强;由于是基于网络系统筛选,最大程度避免了药物不良反应。

1.3 基因工程研究的意义

1.3.1 揭示生命现象规律

基因工程是自然科学家的重要工具。从众多生物体来源的基因或遗传信息可以转化至细菌,形成基因工程菌株,用以基因存储和改造。细菌价格便宜,容易生长,繁殖很快,比较容易转换,几乎可以在 -80°C 无限期保存。一旦分离到一个基因,它可以存储在细菌内,为研究提供了一个无限量供应。

转入有特定基因的生物体表型的改变可以用来研究、发现这些基因的功能。常用实验手段是功能丧失或获取、基因跟踪和表达。通过基因敲除技术使生物体缺失一个或更多基因,从而判断这些被敲掉基因的功能,常称之为缺失基因功能(loss of function);与基因敲除相对应的是在生物体增加基因的拷贝数,增强基因功能(gain of function);为了追踪所关注的蛋白在生物体是如何作用的,往往将编码基因与报告基因(如GFP)相结合,形成融合蛋白,虽然

Note

这是一个很有用的技巧,但形成融合蛋白后可能影响到基因功能,因此这样的实验结果往往会被人质疑。

为了解决这一问题,在构建载体时添加一段可以与单克隆抗体结合的序列,通过免疫组化检测即可追踪蛋白去向。为了研究一个基因的时空表达,可以将该基因的启动子连接报告基因或能催化反应产生有色产物的酶类,以便观察该基因何时何处表达。

1.3.2 基因工程与疾病治疗

基因工程在医学上已被广泛应用,如用基因工程菌生产胰岛素、人类生长激素、人血白蛋白、单克隆抗体、抗血友病因子、疫苗以及其他许多药物。遗传工程病毒正在开发,此种病毒可以赋予机体免疫力,但失去传染性。通过基因工程鼠杂交瘤可产生人源化单克隆抗体,实施抗体治疗。

利用基因工程技术可以制作人类疾病的动物模型。最常见的基因工程动物模型是转基因小鼠,用于癌症、肥胖、心脏疾病、糖尿病、关节炎、焦虑和帕金森病等的研究。利用这些小鼠模型,可以测试潜在治疗药物;基因工程改造的猪可用于人体器官移植。

基因工程可以用于由于基因缺失、突变疾病的基因治疗,将正常的基因插入到基因组取代引起疾病的非正常基因,如果用于治疗的基因插入生殖细胞,患者后代将不再患此基因缺陷的遗传疾病。一个治疗的基因送入体内并准确的进入靶细胞需要运载工具,即载体。目前用于基因治疗的常用载体是病毒载体,主要有逆转录酶病毒、腺病毒、腺相关病毒、单纯疱疹病毒,这些载体已经被改造过,对人体没有毒害。此外,也有其他基因传送方式,如用脂质体包裹将治疗用的 DNA 直接穿膜送入靶细胞;将 DNA 用化学方法连接到能与细胞膜受体结合的分子上,通过细胞膜的内吞作用,将 DNA 引进细胞。

基因治疗存在的问题是安全性问题,到目前为止 FDA 尚未批准人类基因治疗产品上市,临床试验尚未证实技术水平的完全成熟。1990 年开始了第一例基因治疗的临床试验,然而在 1999 年,基因治疗面临了极大挫折。例如,美国患者 Jesse Gelsinger 患有鸟氨酸氨甲酰基转移酶缺陷症(ornithine transcarbamylase deficiency,OTCD),即尿素循环障碍,是一种伴性遗传病。该病往往在患者出生时就会致命,而 Gelsinger 所患不是遗传病,是基因突变的结果,病情并非十分严重,他有些细胞是正常的,因而使他能够在特制的食品、药物控制下存活。1999 年 9 月 13 日,Gelsinger 在美国宾夕法尼亚大学接受基因治疗,注射了携带纠正基因的腺病毒载体,四天之后引发了体内大规模的免疫反应,死于多器官功能衰竭,年仅 18 岁。

1.3.3 基因工程产业

1. 工业

基因工程工业是一个涵盖较广的概念,可以包括所有基因工程的产业。但习惯上,人们往往狭义的限于微生物基因工程,尤其是微生物发酵,以此与医药业分开。利用细菌作为表达系统可以生产蛋白质和酶等,对于不适于细菌表达的产品,可以用酵母菌来表达。此外利用基因工程还可以制造生物燃料,治理环境污染,分解有害有毒废物等。

由于遗传操作便利,酵母菌、大肠杆菌是广泛而主要的微生物细胞工厂,可以生产酶类和生化药物。由于杆菌属是一类非致病菌,不含内毒素和外毒素,被认为是安全菌株,是未来极具发展前途的产业平台。新建的 T7 枯草芽孢杆菌表达系统已用于产生大量重组蛋白,通过 IPTG(异丙基硫代半乳糖苷)诱导调控内源性和外源性酶的表达。由于 IPTG 价格昂贵且有毒性,不适合工业上酶的生产,人们很快发展了一种麦芽糖诱导表达系统和甘露醇诱导表达系统。

2. 农业与食品

早期的转基因作物希望能够抗虫、抗除草剂、抗真菌和抗病毒,并且已经商业化。随后的转基因作物主要在抗盐、抗冻、抗旱以及增加营养品质方面进行提升。基因工程为制药农作物诞