


生命科学专论

功能基因组学

徐子勤 编著

 科学出版社
www.sciencep.com

功能基因组学

徐子勤 编著

科学出版社

北京

内 容 简 介

本书全面总结了功能基因组学的基本原理以及国内外在基因功能研究领域的主流技术和最新发展趋势,共13章,分为四个部分。第一部分(1~3章)概述核酸操作的主要工具和方法。第二部分(4~6章)系统地介绍了基因的克隆、定位和表达体系。第三部分(7~9章)归纳了功能基因组学研究的主要方法,涉及动物模型、蛋白质相互作用、定点突变、基因表达谱以及RNA干扰等多个方面。第四部分(10~13章)具体分析突变体在植物基因功能研究中的重要作用。

本书可供高等学校和科研机构相关教师和研究人員使用,也可以作为生物科学和生物技术专业高年级本科生和研究生教材。

图书在版编目(CIP)数据

功能基因组学/徐子勤编著. —北京:科学出版社,2007

ISBN 978-7-03-019052-9

I. 功… II. 徐… III. 基因组-研究 IV. Q343.1

中国版本图书馆CIP数据核字(2007)第080400号

责任编辑:李秀伟 王 静 赵 满 张 弛/责任校对:陈丽珠

责任印制:钱玉芬/封面设计:福瑞来书装

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>

源海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2007年6月第一版 开本:787×1092 1/16

2007年6月第一次印刷 印张:38

印数:1—3 000 字数:883 000

定价:88.00元

(如有印装质量问题,我社负责调换〈环伟〉)

前 言

自孟德尔的遗传学理论于1900年被重新发现以后,经过50年的时间,人们对遗传信息的了解从基因这样一个模糊的概念发展到了分子水平,1953年DNA双螺旋模型的建立标志着分子生物学时代的开始。20世纪五六十年代是核酸研究的黄金时期,很多分子遗传学的基本理论都是在这一时期建立起来的,其中最关键的工作是遗传密码的破译。20世纪70年代,分子生物学进入了实际应用阶段,1972~1973年重组DNA技术的建立标志着现代遗传工程即基因工程的开始,在迄今短短的30多年时间里已经产生了大量的转基因有机体和重组蛋白质药物。

20世纪90年代以来,基因操作技术逐渐被用于基因组研究,基因组全序列的测定在技术上已经相当成熟,并已完成大肠杆菌、酵母、线虫、果蝇、人类、拟南芥和水稻等十余种生物的基因组全序列测定,2008年小鼠的基因组全序列测定也将完成。随着人类和一些模式生物基因组计划的完成,诞生了一门新的学科——基因组学。基因组学研究的主要内容是从整体水平分析基因组如何发挥作用,注重基因在整个基因组中所扮演的角色和功能,而不是孤立地考虑基因的结构与表达。基因组计划最初的目标是确定一个物种基因组的碱基排列顺序,现在已经可以在几年的时间内完成一个高等生物的基因组测序任务,但得到的大量DNA序列的生物学功能并不明确。

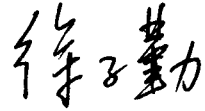
近年来,基因组学研究的焦点已经转移到大规模分析基因组序列的功能方面,并产生了功能基因组学这一概念。目前,功能基因组学已经成为基因组学乃至整个生命科学的核心研究内容。功能基因组学是确定基因组所有基因及其产物的生物学功能的科学,它的主要任务还包括分析各种基因组序列的功能。人类后基因组计划的主要研究内容就是功能基因组学,其中医学基因组学最受人们关注。人类的遗传和生理特征是由基因组决定的,功能基因组学研究将为人类健康做出很大贡献。

但是,迄今还没有一本专门论述功能基因组学和基因功能研究的学术专著,很多基因功能的研究方法散见于各种有关基因操作技术的专著和教材,亟待进行系统的整理使之上升到理论阶段,为建立高通量和规模化基因功能研究体系提供参考。另外,近年来还出现了一些新的功能基因组学研究方法,也需要进行发掘和整理。为了全面深入地介绍功能基因组学研究的主流技术和发展趋势,促进基因组学研究的发展,作者经过多年的资料积累,编写了本书。在编写过程中尽量做到以下三个方面:①以基因功能研究方法为主线,循序展开、逐步深入;②原理与技术相结合,在保证具有较强理论性的同时,注重实用性;③深入浅出,通俗易懂。本书在内容上分为四个部分,其中第一部分(1~3章)和第二部分(4~6章)主要介绍功能基因组学研究的理论和技术基础;第三部分(7~9章)介绍功能基因组学研究的主要方法,是本书的主体;第四部分(10~13章)主要讨论突变体在植物基因功能研究中的应用。在编写过程中,作者查阅了大量的最新文献,在介绍基因操作技术原理的同时,重点讨论了基因功能的研究方法,并分析了功能基因组学研究领域的热点问题和发展趋势。另外,作者结合自己多年从事植

物基因工程研究和教学的经验，从提高读者动手能力的角度出发，在重点介绍基因功能研究方法和原理的同时，充分注意了与其他学科的交叉以及与实验教学的衔接问题，使得该书在内容上能够与遗传学、分子生物学等基础课程互相对应，希望能对从事该领域研究的科研和教学人员以及研究生和本科生有所帮助。

为了能更好地说明功能基因组学研究的方法和原理，本书引用了多位学者公开发表的研究结果和学术观点，在此谨致衷心的感谢。西北大学贾敬芬教授、中国科学院植物研究所孙敬三研究员和中国科学院遗传与发育生物学研究所杨维才研究员对本书的编写提出了许多宝贵的建议，科学出版社王静女士和李秀伟女士为本书的出版付出了辛勤的劳动，西北大学冯淑珍同志做了大量资料整理工作，在此一并致以衷心的感谢。

本书的编写经历了 8 年的时间。在这 8 年时间里，作者曾为本科生和研究生讲授过基因工程课程，并在西北大学重点课程建设项目资助下于 2002 年编写过一本《基因工程讲义》。在教学和科研过程中，作者深感编写和出版有关功能基因组学研究的专著和教材是十分必要的。功能基因组学是一个发展十分迅速的研究领域，涉及的知识面非常广泛，对于本书中的错漏之处，恳请同行和读者批评指正。



2006 年 8 月 10 日

目 录

前言

1 绪论	1
1.1 经典遗传学 (1857~1943 年)	1
1.2 分子遗传学 (1944~1972 年)	4
1.3 基因工程技术 (1973 年至今)	6
1.4 功能基因组学	13
1.5 基因工程与生物安全	16
2 基因操作的主要方法和工具	20
2.1 分子生物学基本知识	20
2.1.1 遗传信息	20
2.1.2 基因和基因组	23
2.1.3 可转座元件	36
2.2 核酸操作	43
2.2.1 核酸的分离和检测	43
2.2.2 核酸的序列测定	48
2.3 限制性内切核酸酶	55
2.3.1 限制性内切核酸酶的类型	55
2.3.2 II 型限制性内切核酸酶	58
2.3.3 限制性作图	63
2.4 连接酶和聚合酶	65
2.4.1 DNA 和 RNA 连接酶	65
2.4.2 DNA 聚合酶	66
2.4.3 RNA 聚合酶	68
2.5 DNA 修饰酶	70
3 受体细胞和载体	71
3.1 受体细胞类型	71
3.2 大肠杆菌质粒载体	72
3.2.1 克隆性质粒载体	73
3.2.2 改进的质粒载体	75
3.3 大肠杆菌噬菌体载体	76
3.3.1 λ 噬菌体载体	77
3.3.2 M13 噬菌体载体	83
3.3.3 质粒与噬菌体杂合性载体	85
3.4 植物基因工程载体	87

3.4.1	Ti质粒	87
3.4.2	Ri质粒	90
3.4.3	植物病毒载体	91
3.4.4	质体基因工程载体	92
3.5	动物基因工程载体	94
3.5.1	反转录病毒载体	94
3.5.2	腺病毒载体	97
3.5.3	SV40病毒载体	98
3.5.4	其他哺乳动物病毒载体	99
3.5.5	昆虫细胞的杆状病毒载体	100
3.6	酵母基因工程载体	101
3.6.1	酿酒酵母载体	101
3.6.2	毕赤酵母载体	102
3.7	基因组载体	103
3.7.1	酵母人工染色体	103
3.7.2	BAC载体和F黏粒	105
3.7.3	P1噬菌体载体和PAC载体	106
3.8	DNA向宿主细胞的导入方法	107
4	基因的克隆与鉴定方法	110
4.1	基因的化学合成	110
4.2	目的基因的克隆策略	111
4.2.1	根据已知碱基序列或氨基酸序列克隆基因	113
4.2.2	根据表型变异克隆基因	114
4.2.3	基于图谱的基因克隆方法	119
4.3	文库的构建与目的基因克隆	123
4.3.1	cDNA文库的构建	124
4.3.2	基因组文库的构建	131
4.3.3	目的基因克隆的鉴别与分析	133
4.4	聚合酶链反应与基因克隆	143
4.4.1	PCR的基本特征	144
4.4.2	改进的PCR技术	147
4.4.3	实时PCR	151
4.4.4	PCR与目的基因序列的克隆	154
4.5	通路克隆系统	155
4.5.1	LR反应	155
4.5.2	BP反应	156
5	基因定位和基因组作图	158
5.1	遗传作图与基因定位	160
5.1.1	RFLP与基因定位	161

5.1.2	RFLP 与遗传作图	163
5.1.3	其他分子标记	167
5.2	人类基因的定位方法	170
5.2.1	家系分析法	170
5.2.2	细胞融合与基因定位	172
5.2.3	异常染色体与基因定位	174
5.2.4	分子杂交与基因定位	175
5.3	物理图谱的构建	176
5.4	基因组计划	179
5.4.1	克隆重叠群	180
5.4.2	基因组序列的测定与组装	181
5.4.3	人类基因组计划	182
5.5	基因序列的分析	187
5.5.1	基因序列的分辨和验证	187
5.5.2	基因的结构与表达分析	189
6	原核与真核细胞表达体系	193
6.1	原核基因表达调控	193
6.1.1	原核基因表达的基本特征	193
6.1.2	操纵子学说	197
6.1.3	大肠杆菌细胞中的突变基因	199
6.2	原核细胞表达体系	201
6.2.1	大肠杆菌表达体系	201
6.2.2	其他原核表达体系	213
6.3	真核基因表达调控	214
6.3.1	真核基因的转录与翻译	214
6.3.2	真核细胞周期的调控基因	222
6.4	真核细胞表达体系	228
6.4.1	酵母表达体系	229
6.4.2	昆虫细胞杆状病毒表达体系	237
6.4.3	哺乳动物细胞表达体系	238
7	基因功能研究与基因工程药物	243
7.1	重组蛋白质药物	243
7.1.1	蛋白质工程	243
7.1.2	重组治疗性药物蛋白	244
7.1.3	新型抗体分子	246
7.1.4	基因工程疫苗	254
7.2	细胞因子	255
7.2.1	细胞因子的一般生物学特征	256
7.2.2	介导和调节先天性免疫的细胞因子	258

7.2.3	介导和调节适应性免疫的细胞因子	266
7.2.4	与T细胞分化和造血功能有关的细胞因子	272
7.2.5	细胞因子与信号转导	275
7.3	基因治疗	282
7.3.1	人类疾病的基因诊断	282
7.3.2	基因治疗	285
7.3.3	遗传指纹	296
8	基因功能研究的主要方法	300
8.1	通过动物模型研究人类基因的功能	300
8.1.1	转基因动物与克隆动物	300
8.1.2	动物乳腺生物反应器	306
8.1.3	动物模型和基因敲除技术	308
8.2	通过蛋白质相互作用研究基因功能	310
8.2.1	酵母双杂交技术	311
8.2.2	表面展示技术	320
8.2.3	免疫共沉淀技术	323
8.2.4	蛋白质亲和层析技术	324
8.2.5	交联技术和荧光共振能量转移技术	326
8.2.6	抗原表位的鉴定与分析方法	327
8.2.7	免疫磁珠和荧光微球技术	332
8.3	通过定点突变技术研究基因功能	336
8.3.1	早期的定点突变技术	336
8.3.2	改进的定点突变技术	337
8.3.3	PCR定点突变技术	339
8.3.4	利用定点突变技术研究核孔复合体转运机制	343
8.4	通过基因表达谱和表型变化研究基因功能	348
8.4.1	基因表达的差异分析	349
8.4.2	功能补偿分析和突变体库技术	358
8.4.3	数据库搜索与基因功能验证	363
8.4.4	基因芯片技术和基因表达谱分析	365
8.4.5	基因传感器	370
9	RNA干扰与基因功能研究	372
9.1	RNA干扰的一般生物学特征	372
9.1.1	RNA干扰的基本步骤	373
9.1.2	siRNA、stRNA和smRNA	375
9.1.3	RNA干扰技术	377
9.1.4	RNA干扰与表观遗传学	379
9.2	参与RNA干扰过程的主要蛋白质	380
9.2.1	RdRP	380

9.2.2	Dicer 酶	384
9.2.3	Argonaute 蛋白	385
9.3	染色质结构与 RNA 干扰	388
9.3.1	染色质结构与表观基因沉默	388
9.3.2	RNA 干扰与可转座元件的异染色质化	392
9.4	RNA 干扰与植物基因功能研究	396
9.4.1	植物中的基因共抑制现象	396
9.4.2	转录后基因沉默	398
9.4.3	双链 RNA 对基因组的调节作用	401
9.4.4	通过 RNAi 技术确定植物基因功能	404
9.4.5	沉默载体向植物细胞的导入方法	411
9.5	RNA 干扰与动物基因功能研究	412
9.5.1	RNA 干扰与线虫基因功能研究	413
9.5.2	RNA 干扰与果蝇基因功能研究	418
9.5.3	RNA 干扰与哺乳动物基因功能研究	423
10	植物基因功能研究的特殊方法	431
10.1	植物基因组	431
10.1.1	植物基因组的基本特征	431
10.1.2	表观遗传与基因表达	434
10.1.3	植物细胞周期的调节特征	436
10.2	植物基因工程	438
10.2.1	植物遗传转化	438
10.2.2	植物抗性基因工程	440
10.2.3	高等植物的质体基因工程	444
10.2.4	植物基因工程技术在医学中的应用	445
10.3	植物基因表达的调控方式	447
10.3.1	植物基因的启动子	447
10.3.2	RNA 加工和 RNA 编辑	452
10.3.3	转录因子和顺式元件	454
10.3.4	转移基因的表达调控	456
10.4	植物遗传转化的选择体系	463
10.4.1	常用的植物标记基因	463
10.4.2	植物遗传转化的正选择体系	466
10.5	植物诱导表达体系和定点重组系统	470
10.5.1	植物基因表达的诱导体系	470
10.5.2	T7 RNA 聚合酶——植物耦联表达体系	472
10.5.3	定点重组系统	472
10.6	植物细胞结构相关蛋白的功能研究	475
10.6.1	植物细胞中膜蛋白的功能研究	475

10.6.2	细胞壁蛋白的功能研究	480
10.6.3	细胞骨架蛋白的功能研究	485
10.6.4	液泡蛋白的功能研究	487
11	突变体与植物代谢基因功能研究	491
11.1	氨基酸合成代谢基因的功能研究	491
11.1.1	芳香族氨基酸合成	491
11.1.2	脯氨酸合成代谢途径	496
11.1.3	其他氨基酸合成代谢突变体	497
11.1.4	固氮作用	499
11.1.5	植物细胞对氨的同化过程	500
11.1.6	次生产物合成代谢基因的功能研究	501
11.1.7	含硫氨基酸及其衍生物的功能	503
11.2	脂类合成代谢基因的功能研究	505
11.2.1	植物脂类代谢突变体	505
11.2.2	脂类基因工程	509
11.2.3	聚-3-羟基链烷酸	512
11.3	光合作用和淀粉合成代谢基因的功能研究	513
11.3.1	突变体与光合作用研究	513
11.3.2	类胡萝卜素的生物合成	515
11.3.3	淀粉合成代谢	515
11.4	物质转运突变体	519
11.4.1	钾离子转运	519
11.4.2	磷的转运	520
11.4.3	微量元素转运蛋白	521
12	突变体与植物形态发生和开花基因功能研究	523
12.1	控制分生组织和信号转导的基因	523
12.1.1	芽和根的发生	523
12.1.2	突变体与植物信号转导研究	525
12.2	植物激素合成代谢基因的功能	527
12.2.1	赤霉素合成代谢突变体	527
12.2.2	脱落酸合成代谢突变体	529
12.2.3	生长素与细胞分裂素合成基因	532
12.2.4	其他植物激素合成基因的功能研究	535
12.3	植物生殖发育突变体与基因功能研究	537
12.3.1	花分生组织的形成	537
12.3.2	控制开花的基因	539
12.3.3	花器官身份确定与花器官发育基因	541
12.3.4	雄配子体与雌配子体发育基因	547
12.3.5	雄性不育基因	550

12.3.6	与种子形成有关的基因	552
12.4	植物的衰老与程序性死亡基因	556
12.4.1	与植物组织衰老有关的基因	556
12.4.2	植物组织衰老的调节因素	558
13	植物抗性基因的功能研究	560
13.1	植物对病原体的应答特征	560
13.1.1	植物对病原体应答的一般特征	560
13.1.2	基因对基因假说	562
13.1.3	R 基因	564
13.2	植物对病原体的防御机制	571
13.2.1	超敏反应和损伤模拟突变体	571
13.2.2	植物防御体系	573
13.3	基因工程与植物病害的控制	577
13.3.1	R 基因转移体系	577
13.3.2	防御性基因与植物对病原体的抗性	579
13.4	植物对逆境胁迫的抗性	580
13.4.1	植物对水分胁迫和盐胁迫的应答机制	580
13.4.2	水分胁迫抗性基因工程	585
13.4.3	低温与其他逆境胁迫	586
13.4.4	植物应答非生物胁迫的关键性转录因子	588
	参考文献	591

1 绪 论

1.1 经典遗传学 (1857~1943 年)

经典遗传学的创始人是奥地利的伟大科学家孟德尔 (G. J. Mendel)。1857~1864 年, 孟德尔以豌豆 (*Pisum sativum*) 为实验材料进行了大量系统的杂交实验, 提出了特殊因子 (particular factor; 也称为遗传因子, hereditary factor) 这一概念, 将性状 (表型, phenotype) 与可遗传因子 (基因型, genotype) 联系到了一起。孟德尔的豌豆实验一共研究了 8 对简单性状, 分别为种子形状 (圆形和皱形)、种子颜色 (黄色和绿色)、种皮颜色 (灰色和白色)、花的颜色 (紫色和白色)、种荚形状 (膨胀和皱缩)、种荚颜色 (绿色和黄色)、植株高度 (长秆和矮化) 和花在植株上的位置 (花和豆荚沿轴排列, 花和豆荚位于末端)。他采用的主要方法是在花药释放花粉之前剪去未成熟花药, 再将另一个亲本的花粉涂抹到第一个亲本的柱头上, 然后对杂种第一代和第二代进行观察。基于大量的统计学分析, 孟德尔提出了遗传学的两个基本规律, 分离规律 (the principle of segregation) 和自由组合规律 (the principle of independent assortment)。分离规律指一个基因的两种等位形式在配子形成过程中会发生分离, 一种配子得到一个等位基因, 另一种配子得到另一个等位基因。自由组合规律认为与不同性状有关的基因在配子形成过程中是独立的, 等位基因以随机方式进行组合。孟德尔认为生物的每一种性状都是由特殊的遗传因子控制的, 两个不同的个体杂交产生的下一代可能与亲本之一完全相同, 他把这一现象称为统一律, 并认为遗传因子可以代代相传; 在体细胞内, 遗传因子是成对存在的, 其中一个来自父本, 另一个来自母本; 在形成配子时成对的遗传因子彼此分开, 单独存在。他还认为, 有些遗传因子以显性 (dominant) 形式存在, 即在任何杂种一代个体中都能够表达; 而有些遗传因子呈隐性 (recessive) 状态, 只有当父本和母本同时含有这一因子时, 才能在一些子代中表达。但是, 从 1865 年孟德尔发表“植物杂交试验”一文到 1884 年逝世, 科学界几乎无人理睬他的巨大贡献。

孟德尔的研究结果在 1900 年整理德国植物学家 Carl von Nägeli 资料中的一些信件时重新被发现。这些信件显示孟德尔也在其他植物上进行了试验, 他在一些以无性方式繁殖的植物中进行的杂交实验是失败的。孟德尔当时也认为豌豆中表现出的遗传规律的适用范围可能是有限的。信件还显示孟德尔后来又进行了玉米杂交实验, 并进一步证实了他在豌豆实验中得到的结果。

孟德尔提出的分离规律和自由组合规律适用于所有非连锁基因, 他的实验能够成功取决于以下几个因素: 第一, 豌豆为自花受精 (self-fertilized) 植物, 可以控制授粉过程, 便于人工杂交; 第二, 他挑选的性状属于简单性状, 每一种性状有两种不同的形式, 杂交一代中只显示其中一种; 第三, 孟德尔在每次杂交实验中挑选的性状不多于两个, 能够在可控的遗传背景下跟踪每个性状的传递规律; 第四, 他没有观察连锁性状, 这一点十分幸运; 第五, 他分析的样本很大, 可通过统计学方法进行分析。豌豆种子的

皱皮表型是孟德尔遗传学实验中讨论最多的性状，现在已从分子水平揭示了该表型的形成机理，并克隆了与之有关的基因。皱皮表型由一个基因突变引起，该基因编码淀粉分支酶 SBEI (starch-branching enzyme isoform)，这种突变降低了种子中淀粉的含量，增加了蔗糖浓度。蔗糖可以作为渗透剂，细胞内蔗糖浓度升高时，水分会进入细胞，使种子体积变大，鲜重增加。当成熟种子干燥后，突变种子大量失水，所以种皮会变皱。

早期的遗传学、细胞学和生物化学研究是相互独立的。直到 1903 年，W. Sutton 和 T. Boveri 通过对染色体的细胞学观察提出了遗传的染色体理论 (chromosome theory of heredity)。染色体在光学显微镜下是可见的，并且可以一个个加以区分。当时已发现染色体总是始终如一地从上一代传递到下一代，与性状的传递规律相似。由于染色体的行为与遗传因子的行为十分相似，所以 Sutton 和 Boveri 推论孟德尔因子 (Mendelian factor) 存在于染色体上，染色体可能是遗传物质的载体。1909 年，W. L. Johannsen 提出“基因” (gene) 一词，用来代替孟德尔提出的“遗传因子”概念。

1905 年，W. Bateson、E. R. Saunders 和 P. C. Punnett 在进行香豌豆 (*Lathyrus odoratus*) 杂交试验时注意到基因对的自由组合并不是绝对的，他们通过杂交实验发现染色体片段可以交换。按照我们熟知的自由组合规律，当两个拥有两种不同性状的亲本杂交时， F_1 代会显示出两种性状中的显性形式， F_1 代自交产生的 F_2 代将表现出 9 : 3 : 3 : 1 的分离比率。但是，Bateson 和他的同事却得到了不同的结论。他们所用的两个亲本中，一个具长花粉和紫花表型，另一个具圆花粉和红花表型。杂交之后， F_1 代显示出期望的显性表型 (紫花长花粉)。但 F_1 代自交后，绝大部分 F_2 代植株产生紫花长花粉或红花圆花粉，说明 F_2 代类似于亲本中的任何一个。另外，这一实验也产生了少量的非亲本表型，紫花圆花粉和红花长花粉。这意味着在配子形成过程中并不是所有的基因都遵循孟德尔的自由组合规律。

同一时代的摩尔根和他的同事在果蝇 (*Drosophila melanogaster*) 中也做着同样的工作，他们选择的性状为果蝇复眼的颜色和翅的大小，所获结果也不同于自由组合规律。摩尔根及其同事推论，如果在一条染色体上控制不同性状的孟德尔因子相互靠近时，它们可能会一起遗传，亦即它们是连锁的，一组连锁基因构成了一个连锁群 (linkage group)。少数非亲本型可用两个遗传因子之间的连锁在配子形成过程中被随机打断进行解释，并由此产生新的孟德尔因子组合，这一过程被称为重组。遗传因子之间距离越近，重组的机会就越低。进一步的工作将这种现象与同源染色体之间的交换 (crossing-over) 联系到了一起，重组率 (recombinant frequency) 在后来被当作同一染色体上两个不同基因之间距离的相对指标。

1910 年，摩尔根提出基因的连锁和交换规律，创立了遗传的染色体理论，将决定性状的基因同染色体联系到了一起。他所选用的最重要的果蝇性状是复眼的颜色，野生型果蝇为红眼，突变体果蝇为白眼。复眼的颜色与 X 染色体连锁， F_2 代所有的白眼果蝇都是雄性个体。当摩尔根认识到部分连锁可以通过减数分裂过程中发生的交换现象给予解释时，他开始考虑如何设计一种方法来确定基因在染色体上的位置。他的研究生 Arthur Sturtevant 取得了关键性的突破，发现因交换使两个连锁基因分开的频率同它们在染色体上所处位置之间的距离成正比，因此重组率可作为测量基因之间距离的相对

尺度。为了纪念摩尔根对遗传学的重大贡献，1%重组率被定义为一个 cM（厘摩尔根）。在遗传学实验中，只要获得不同基因之间的重组率，就可以绘制一张关于不同基因在染色体上相对位置的图谱，这就是遗传图谱。例如，*A1* 是玉米 3 号染色体上与种子花青素颜色有关的基因（紫色种子），*Etched* 基因与种子形状有关（蚀刻形种子），这两个基因之间的重组率为 12%，也即图距（map unit）为 12cM。*Mdh3* 基因编码苹果酸脱氢酶（malate dehydrogenase），*Sh2*（shrunken kernel）基因与皱缩种子有关。*Mdh3* 与 *A1* 之间的重组率为 3%，*Sh2* 与 *A1* 之间的重组率为 0.2%。20 世纪 80 年代产生的限制性片段长度多态性（RFLP）标记也可以当作表型性状用于遗传作图，*cdo455*、*umc#63*、*csu305* 为玉米 3 号染色体上的几个 RFLP。现在的遗传图谱上既有基因，也包括很多分子标记。

1916 年，C. B. Bridges 将果蝇 X 染色体的不分离现象与特殊的遗传性状联系起来，发现这一现象发生在减数分裂过程中同源染色体不分离的情况下，一些子代细胞拥有某一条染色体的两个拷贝，而另外一些子代细胞丢掉了某一条染色体。之后在研究玉米的细胞遗传学特征时很多出色的工作也将染色体和基因联系到了一起。玉米有 10 对染色体，大小不一，1 号最长，10 号最短。20 世纪 30 年代早期，B. McClintock 研究了三体玉米，发现在杂交过程中产生三体表型的比率与带有特殊基因群体的额外染色体有关，利用三体特征和表型性状可将一个基因定位在某一条染色体上。B. McClintock 通过这种方法将玉米三体染色体和一个影响糊粉层颜色的基因（*R*, red）联系到一起，她发现回交后代的表型比例决定于该基因在亲本中的数量。

染色体交换首先是在玉米中被观察到的。H. B. Creighton 和 B. McClintock 使用了一个带有非正常染色体的玉米株系，该株系 9 号染色体短臂上有一个大的染色结（knob），同时由于 8 号染色体片段的易位，9 号染色体长臂加长。这种突变的 9 号染色体包含有控制种子糊粉层颜色的 *C* 基因（*C* 为显性，无色为隐性，该基因靠近染色结）和种子胚乳蜡质基因 *wx*（显性为非蜡质，*wx* 为隐性，靠近 8 号染色体易位片段）。Creighton 和 McClintock 正是利用找到的两种细胞学标记（染色结和加长的染色体臂）和两种遗传标记（控制两种性状的基因），观察到了染色体的交换现象。用含正常 9 号染色体的亲本与含突变 9 号染色体的亲本杂交，可将染色体的物理特征与其包含的基因的遗传模式联系起来。每当他们观察到一种重组表型时，细胞学标记显示在 knob/*C* 位置和易位的 8 号染色体片段/*wx* 位置之间的某个位点就发生染色体物质的交换。带有 *c* 和 *Wx* 正常 9 号染色体的亲本与带有 *C* 和 *wx* 变异 9 号染色体的亲本杂交后，下一代在配子形成时由于交换可使遗传标记和细胞学标记发生重排；在 F_2 代产生的一种重组型染色体带有 *C* 和 *Wx*，有结；另一种带有 *c* 和 *wx*，具有 8 号染色体易位片段。

20 世纪前 40 年，遗传学领域的研究热点是遗传作图（genetic mapping）。遗传作图的目的是将基因标定在具体的染色体上，并用重组率显示出它们之间的距离，连锁群这一重要概念被用来描述存在于同一条染色体上的连锁基因。有些情况下，连锁群指相互连锁的多个基因或标记，但尚未确定它们位于哪一条具体的染色体。对于染色体组型还没有研究清楚的物种，连锁基因只能用连锁群来界定。在这一时期，另外一个重要进展是发现了细菌的转化现象。

从以上分析可以看出，遗传学早期的研究路线具有由表及里的特征，具体表现为从

性状到基因，然后将基因定位到染色体上。摩尔根学派将决定特定性状的基因同特定的染色体联系在一起，使科学界普遍认识到了染色体的重要性并接受了孟德尔的遗传学原理。

1.2 分子遗传学（1944~1972年）

从孟德尔提出遗传因子到基因被定位到染色体上，大约花了50年时间。尽管由于摩尔根及其学派的出色工作，使基因学说得到了普遍承认，但20世纪上半叶人们对于基因的理解仍然是抽象的、概念化的，缺乏具体内容，就像我们现在对“意识”的理解一样。当时还不能解释基因在细胞增殖过程中是如何准确地复制和遗传的，也不能解释位于细胞核中的染色体和基因是怎样控制显然发生在细胞质中的各种生化过程的。基因的结构特征和遗传信息的表达过程一直到20世纪五六十年代才初见端倪。

1869年，F. Miescher首次从莱茵河鲑鱼精子中分离到DNA。由于核酸在细胞中的含量很低，人们一直认为它并不重要，而将主要精力用于蛋白质、脂类和多糖的研究；所以在后来的很长时间里，核酸研究备受冷落，大多数人认为它只不过是一种能够在70%乙醇溶液中形成线状沉淀的酸性物质。直到20世纪40年代，人们终于明白DNA是传递遗传信息的分子，并发现染色体由蛋白质和DNA构成，DNA携带有遗传信息。因此，从基因被定位于染色体上到发现遗传信息的载体分子是DNA，又走过了大约50年的时间。作为传递遗传信息的分子，应该能够编码细胞生长分裂以及个体发育、结构形成和产生后代的信息；能够精确地进行自我复制，以保证子代与亲本含有相同的遗传信息；能够发生变异，以满足进化过程中适应环境的需要。DNA满足以上所有条件。1941年，G. W. Beadle和E. L. Tatum以脉孢菌为材料，发现遗传突变和生化代谢途径中所需特定酶的缺失之间存在对应关系，提出一个基因决定一个酶的假说，将蛋白质的合成与基因联系到了一起。

1928年，英国学者F. Griffith在利用肺炎双球菌（*Diplococcus pneumoniae*）感染家鼠时，把不产荚膜的无毒的粗糙型肺炎双球菌和加热杀死后的产荚膜的有毒的光滑型肺炎双球菌混合后注射小鼠，意外地发现小鼠被感染致死，而且可以从小鼠的血液中分离出活的产荚膜的肺炎双球菌，从而发现了细菌的转化现象。

1944年，美国化学家O. T. Avery等以肺炎双球菌为研究对象，证明DNA是遗传信息的携带者。肺炎双球菌可以引起肺炎，导致小鼠死亡。细菌的毒性取决于细胞表面荚膜中的多糖，荚膜多糖能保护细菌免受动物白细胞的攻击。S型肺炎双球菌带有荚膜多糖，外表光滑，能使小鼠发病。R型肺炎双球菌没有荚膜多糖，外表粗糙，没有致病能力。煮杀灭活的S型肺炎双球菌不具有致病能力，但是将煮杀的S型肺炎双球菌与R型肺炎双球菌混合后一起感染小鼠时，却可以引起小鼠死亡。解剖学实验证实死亡的小鼠体内存在大量的S型肺炎双球菌。Avery等对S型肺炎双球菌使R型转化为S型的转化物质进行分离和鉴定，从元素分析、酶学分析、血清学分析以及生物活性物质鉴定等方面证实，无细胞提取物中引起肺炎双球菌荚膜形成的转化因子是DNA，发现只有DNA成分能将R型细菌转化为S型，直接证明了DNA是遗传信息的载体。

1952年，A. D. Hershey和M. Chase通过实验证实T2噬菌体的遗传物质是DNA。

他们用³⁵S和³²P分别标记 T2 噬菌体的外壳蛋白和核心 DNA，然后用双标记的噬菌体感染大肠杆菌，剧烈振荡使附着在细菌表面的噬菌体外壳脱落，离心收集细菌进行培养，再分别提取外壳蛋白和 DNA 测定放射量，发现蛋白质外壳没有进入细菌，只有 DNA 进入了细菌细胞中。新的子代噬菌体 DNA 中只有少量的噬菌体 DNA 带有³²P标记（来自亲本），而外壳蛋白全部不带有³⁵S标记。

1953 年，F. Crick 和 J. Watson 在 R. Franklin 和 M. Wilkins 的 DNA X 射线晶体衍射分析数据以及 E. Chargoff 有关 DNA 组成的数量分析数据（A=T，G=C）的基础上，提出了 DNA 的双螺旋结构模型。DNA 的结构模型是 20 世纪最伟大的原创性发现之一，它标志着分子生物学时代的开始。英国科学家 Crick 第二次世界大战前的专业是物理学，第二次世界大战中从事防御鱼雷袭击的军事科学研究。第二次世界大战结束后，Crick 对“生命体和非生命体的分界”产生了兴趣，决定自学生物和化学，并于 1949 年进入剑桥大学凯文迪什实验室，从事多肽和蛋白质的 X 射线衍射分析研究。Crick 还与素有“分子医学之父”之称的弗农·英格拉姆一道发现了遗传物质在决定蛋白质特异性方面的功能，也就是 DNA 与遗传密码之间关系的中心法则。美国科学家 Watson 大学阶段的专业为动物学，后来从事 X 射线对噬菌体增殖的影响研究，1951 年到剑桥大学开始蛋白质和多肽晶体结构的分析研究。同一时期伦敦的两位晶体分析学家 M. Wilkins 和 R. Franklin 获得的研究结果对他们产生了极大的影响，直接导致了 DNA 旋转的梯子模型的建立。

之后不久，第二种核酸 RNA 被鉴定为烟草花叶病毒（tobacco mosaic virus, TMV）的遗传物质。A. Gierer 和 G. Schramm 在 1956 年发现纯化的 RNA 单独使用可导致烟草叶片被感染，说明这种 RNA 分子携带了新病毒产生所需要的所有遗传信息。第二年，H. Fraenkel-Conrat 和 B. Singer 证实了这一假说，他们使用两株不同的 TMV 杂合体进行实验，每一种杂合体由一株的 RNA 和另一株的蛋白质构成，结果发现在被感染的烟草叶片中杂合病毒的子代总是与所包含的 RNA 组分代表的类型相匹配。

20 世纪五六十年代，核酸研究进入高潮，取得的重大发现数不胜数。1957 年，A. Kornberg 在大肠杆菌中发现 DNA 聚合酶 I。1958 年，M. Meselson 和 F. W. Stahl 提出 DNA 半保留复制模型。1959~1960 年，S. Ochoa 发现 RNA 聚合酶和信使 RNA，并证明信使 RNA 携带的遗传信息决定蛋白质分子中的氨基酸顺序。1961 年，M. W. Nirenberg 等应用合成的 mRNA 分子多聚 U 破译出第一个遗传密码。1961 年，F. Jacob 和 J. Monod 提出了调节原核基因表达的操纵子模型。1964 年，C. Yanofsky 和 S. Brenner 等证明多肽链上的氨基酸顺序同其编码基因中的核苷酸顺序存在着共线性（colinear）关系。1965 年，S. W. Holley 完成了第一个酵母丙氨酸 tRNA 的核苷酸全序列测定。1966 年，M. W. Nirenberg、S. Ochoa 和 H. G. Khorana 共同破译了全部遗传密码。

在生命科学研究中，学科之间的交叉起着非常重要的作用，很多生物学的重大发现是采用物理学和化学方法完成的，其中最典型的例子就是 X 射线衍射技术。1895 年伦琴发现了 X 射线，1912 年德国人劳厄发现了 X 射线通过晶体时产生的衍射现象，导致了 X 射线衍射技术的诞生，它成为研究和测定晶体内部结构的重要手段。英国女晶体学家 R. Franklin 对 DNA 晶体进行了详细的 X 射线衍射分析，得到了 DNA 的电子密度