

国家自然科学基金青年基金项目(编号:81501151)

国家自然科学基金地区项目(编号:81660234)

精神分裂症与抑郁症的遗传学进展与研究

贺宽军 郭 闯 著

苏州大学出版社

前 言

人人渴望拥有健康的心理,然而由于生活、工作节奏日益加快,社会竞争日益加剧,精神疾病患者的数量却与日俱增。精神疾病是指在遗传和环境因素共同作用下大脑的功能失调,进而导致认知、情感、意志、行为等精神活动出现不同程度障碍的疾病。这种疾病给患者带来了巨大的痛苦,给家庭造成了巨大的经济负担,也成为社会不稳定因素。

据2001年世界卫生组织报道,全球罹患某种精神疾病的人数高达4.5亿。我国人口众多,且正处于各种复杂的社会、经济、文化因素的大变革时期,人们的生活方式、行为、心理情感均受到了冲击与影响,因此精神疾病发病率不断攀升。进入21世纪后,我国的各类精神问题更加突出,精神障碍疾病超过了心脑血管、呼吸系统及恶性肿瘤,在疾病总负担中位列第一。

精神疾病病因复杂,到目前为止尚不清楚。多项研究证实,精神疾病的发生与遗传因素关系很大。除遗传因素外,环境因素也是精神疾病发病的主要原因。环境因素有自然环境和社会环境两方面,其中社会环境因素是精神疾病发病的主导因素。尽管精神疾病的遗传学和神经影像学研究取得了令人瞩目的成就,但这些研究均不能解释清楚精神疾病的所有病理现象。关于精神疾病的发病原因,存在各种假说。这些假说分别试图从社会的、生物心理学的、遗传和环境的角度对精神疾病的发病原因做出解释。

本书在详细介绍精神疾病的分类、定义、临床症状、诊断标准及亚型的基础上,对两种主要的精神疾病——精神分裂症和抑郁症,从流行病学、病因学、发病机理、遗传学、表观遗传学等方面取得的研究进行提炼与概括。本书重点介绍精神分裂症和抑郁症在遗传学和表观遗传学领域取得的研究进展。最后本书以汉族人群的 *CACNA1C*、*ITIHs* 家族基因与精神分裂症、重度抑郁症的遗传学关联研究和作为精

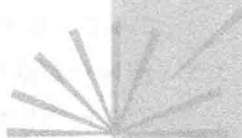
精神分裂症的诊断标志物的血清 miRNA 鉴定两方面的研究进行详细介绍。

在充分挖掘、认真梳理研究文献的基础上,笔者采用宽视域、多视角的研究方法深入探究精神疾病的致病机理。全书视野开阔,新意迭出,创新之处体现在以下四个方面:

(1) 在系统介绍精神疾病的分类、定义、临床症状、诊断标准及亚型的基础上,从流行病学、病因学、发病机理三个方面提炼并概括精神分裂症和抑郁症的研究进展。重点聚焦精神分裂症和抑郁症在遗传学领域和表观遗传学领域取得的研究进展,为相关研究人员提供背景知识与借鉴。

(2) 对汉族人群中 *CACNA1C*、*ITIH3* 家族基因与精神分裂症、重度抑郁症的遗传学关联研究进行探究,从而为精神疾病生物学研究奠定了基础,也为精神疾病的诊断及病因学分析提供了借鉴和参考。

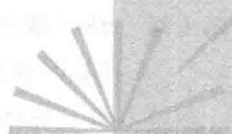
(3) 专著探讨了将血清 miRNA 作为精神分裂症诊断标志物的过程与可行性,并对精神分裂症表观遗传学研究方法进行了具体阐述。



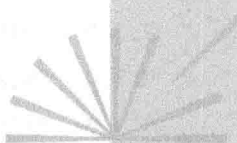
目 录

CONTENTS

- 第1章 分子遗传学基础 / 001
 - 1.1 人类基因和基因组 / 001
 - 1.1.1 基因的概念 / 001
 - 1.1.2 基因的化学本质 / 002
 - 1.1.3 人类基因和基因组介绍 / 003
 - 1.1.4 人类基因组计划 / 005
 - 1.2 基因的复制与表达 / 007
 - 1.2.1 基因的自我复制 / 007
 - 1.2.2 基因的表达 / 009
 - 1.3 基因突变 / 011
- 第2章 精神病学概述 / 015
 - 2.1 精神疾病与精神病学概述 / 015
 - 2.2 脑与精神活动 / 015
 - 2.2.1 脑结构与精神活动 / 015
 - 2.2.2 脑神经化学与精神活动 / 016
 - 2.2.3 脑可塑性与精神活动 / 018
 - 2.3 精神疾病与病因 / 018
 - 2.3.1 精神疾病与遗传因素 / 018
 - 2.3.2 环境因素与精神疾病 / 018
 - 2.4 器质性精神疾病 / 020
 - 2.4.1 脑器质性精神疾病 / 020
 - 2.4.2 躯体疾病所致精神疾病 / 021
 - 2.5 精神疾病分类与诊断 / 021
 - 2.5.1 常用的精神疾病分类系统 / 022
 - 2.5.2 精神疾病的诊断标准 / 023
- 第3章 精神分裂症概述 / 025
 - 3.1 精神分裂症患者的临床症状 / 025
 - 3.2 精神分裂症的诊断和亚型 / 027
 - 3.3 精神分裂症的流行病学研究 / 029
 - 3.4 精神分裂症的病因学研究 / 030
 - 3.4.1 遗传因素 / 030



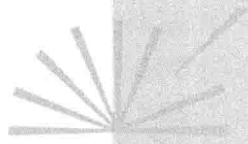
- 3.4.2 环境因素 / 031
- 3.5 精神分裂症的发病机理 / 033
 - 3.5.1 神经递质假说 / 034
 - 3.5.2 神经发育假说 / 038
- 第4章 抑郁症概述 / 049
 - 4.1 抑郁症的定义与症状 / 049
 - 4.2 抑郁症的诊断与分类 / 049
 - 4.3 抑郁症的流行病学研究 / 050
 - 4.4 抑郁症的病因学研究 / 051
 - 4.5 抑郁症的致病机理 / 052
- 第5章 精神疾病遗传学研究方法 / 056
 - 5.1 遗传性疾病简介 / 056
 - 5.1.1 遗传病的分类 / 056
 - 5.1.2 遗传病的研究方法 / 057
 - 5.2 精神疾病的遗传学研究方法 / 059
 - 5.2.1 遗传标记 / 059
 - 5.2.2 连锁分析 / 060
 - 5.2.3 关联分析 / 061
 - 5.2.4 全基因组关联分析 / 063
 - 5.2.5 拷贝数变异的检测 / 066
 - 5.2.6 小RNA测序 / 068
 - 5.2.7 PCR技术 / 070
 - 5.2.8 RT-PCR技术 / 073
 - 5.2.9 实时定量PCR / 074
- 第6章 精神分裂症和抑郁症的遗传学研究进展 / 080
 - 6.1 精神分裂症的遗传学研究进展 / 080
 - 6.1.1 遗传发现 / 082
 - 6.1.2 基因组测序研究 / 088
 - 6.1.3 转化医学的影响 / 089
 - 6.1.4 诊断学 / 090

- 
- 6.1.5 药物发现 / 091
- 6.1.6 个性化治疗 / 092
- 6.1.7 小结 / 094
- 6.2 抑郁症的遗传学研究进展 / 094
- 6.2.1 单胺类假说 / 096
- 6.2.2 压力:抑郁症的另外一个病因 / 098
- 6.2.3 神经发生和神经可塑性的紊乱 / 100
- 6.2.4 细胞因子理论 / 101
- 6.2.5 昼夜节律理论 / 103
- 6.2.6 其他候选基因 / 104
- 6.2.7 抑郁症的全基因组关联分析 / 108
- 6.3 常见精神疾病致病基因的遗传共享 / 118
- 第7章 精神分裂症和抑郁症的表现遗传学研究进展 / 146**
- 7.1 表观遗传学介绍 / 146
- 7.1.1 DNA 甲基化 / 147
- 7.1.2 组蛋白修饰 / 147
- 7.1.3 基因组印记 / 151
- 7.1.4 非编码 RNA 在表观遗传学中的作用 / 151
- 7.1.5 染色质重塑 / 154
- 7.1.6 X 染色体失活 / 154
- 7.2 精神分裂症表观遗传学研究进展 / 154
- 7.2.1 DNA 甲基化与精神分裂症 / 160
- 7.2.2 组蛋白修饰与精神分裂症 / 161
- 7.2.3 miRNA 与精神分裂症 / 165
- 7.2.4 exoRNA 与精神分裂症 / 170
- 7.3 抑郁症表观遗传学研究进展 / 174
- 7.3.1 重度抑郁及相关模型中表观遗传谱的全基因组分析 / 178
- 7.3.2 压力反应性基因 / 180

- 7.3.3 神经营养信号基因 / 181
- 7.3.4 单胺能转运蛋白基因 / 182
- 7.3.5 全甲基化组的关联研究 / 182
- 7.3.6 DNA 甲基化和抗抑郁处理 / 185
- 7.3.7 讨论与前景展望 / 185
- 7.3.8 非编码 RNA 与抑郁症 / 186
- 7.3.9 作用于单胺类的 miRNA 研究 / 188
- 7.3.10 miRNA 作用于谷氨酸能系统的研究 / 191
- 7.3.11 抑郁症患者总的 miRNA 变化 / 191

第 8 章 汉族人群中 *CACNA1C*、*ITIHs* 家族基因与精神分裂症、重度抑郁症的关联研究 / 219

- 8.1 汉族人群中 *CACNA1C* 与精神分裂症、重度抑郁症的遗传易感性分析 / 219
 - 8.1.1 *CACNA1C* 与精神疾病 / 219
 - 8.1.2 研究位点的选取 / 225
 - 8.1.3 研究材料与方法 / 226
 - 8.1.4 统计分析 / 234
 - 8.1.5 实验结果 / 239
 - 8.1.6 讨论 / 252
- 8.2 *ITIHs* 家族基因与精神分裂症和重度抑郁症的遗传易感性分析 / 254
 - 8.2.1 *ITIHs* 家族基因与精神疾病 / 254
 - 8.2.2 研究位点的选取 / 255
 - 8.2.3 研究材料与方法 / 257
 - 8.2.4 统计分析 / 258
 - 8.2.5 实验结果 / 258
 - 8.2.6 讨论 / 266
- 8.3 *CACNA1C* 和 *ITIHs* 家族蛋白相互作用关系网络及其功能富集的生物信息学分析 / 268
 - 8.3.1 分析方法 / 268



- 8.3.2 结果 / 269
- 8.3.3 讨论 / 274
- 8.4 总结与展望 / 294
 - 8.4.1 总结 / 294
 - 8.4.2 展望 / 296
- 第9章 血清 miRNA 作为精神分裂症诊断标志物的鉴定 / 305**
 - 9.1 背景 / 305
 - 9.2 材料与方法 / 307
 - 9.2.1 研究对象 / 307
 - 9.2.2 血液的采集 / 307
 - 9.2.3 miRNA 的选择 / 307
 - 9.2.4 总 RNA 抽提和 qRT-PCR 定量分析 / 307
 - 9.2.5 数据分析 / 309
 - 9.3 结果 / 309
 - 9.4 讨论 / 313

第1章 分子遗传学基础

1.1 人类基因和基因组

基因是构成生物遗传物质的结构和功能的基本单位,储存着生命的生长、发育、衰老、凋亡等过程的全部信息。基因组是指以核苷酸序列形式存储的细胞内所有遗传信息。在细胞或生物体中,基因组即为一套完整的单倍体的遗传物质的总和。

1.1.1 基因的概念

1866年,奥地利学者孟德尔在他的豌豆杂交实验论文中,用大写字母代表显性性状,用小写字母代表隐性性状,并提出了生物的性状是由“遗传因子”决定的,并总结出了遗传因子传递的分离定律(Law of Segregation)和自由组合定律(Law of Independent Assortment)。尽管当时孟德尔并没有严格区分所观察到的性状和控制这些性状的遗传因子,但是从孟德尔用这些符号所表示的杂交结果来看,这些符号正是代表着形式上的基因,而且为了方便起见,至今在遗传学的分析中仍沿用这些符号来代表基因。20世纪初,丹麦遗传学家维尔赫姆·路德维希·约翰逊在其出版的《遗传学原理》(*Elements of Heredity*)一书中将遗传因子更名为基因,并一直沿用至今。随后,美国遗传学家托马斯·亨特·摩尔根及其学生艾尔弗雷德·斯特蒂文特、卡尔文·布里奇斯、赫尔曼·约瑟夫·穆勒等通过果蝇杂交实验证实,基因在染色体上呈直线排列,并提出了基因的连锁与互换定律(Law of Linkage and Crossing-Over)。

目前,科学家将分离定律、自由组合定律和连锁与互换定律统称为遗传学三大定律。1927年,Muller等人证明了采用人为干扰因素如X线等可使基因发生突变,而在此之前英国内科医生阿

奇·博尔德·加罗德认为人类的尿黑酸尿症是由于人体内缺乏某种生化反应酶而导致的一种遗传病,于是人们将基因和酶联系起来。1941年,乔治·韦尔斯·比德尔等人基于果蝇眼型和粗糙链孢霉的研究结果,提出了“一个基因决定一种酶”的假说。然而,后来随着研究的深入,人们认识到基因除了决定酶之外,其他蛋白质也是由基因决定的,于是出现了“一个基因一种蛋白质”的假说。但当人们又发现蛋白质由几条多肽链组成时,这种提法又被修正为“一个基因一条多肽链”。后来随着中心法则的提出、RNA编辑的发现,人们认识到,由于RNA剪切加工过程的不同,一个基因可以产生多种mRNA,进而编码多条多肽链,即又出现了“一个基因多条多肽链”的提法。就这样反反复复,历经100余年的研究,基因的概念也逐渐清晰起来。到20世纪50年代,人们对基因的物质基础和理化本质的认识依然模糊,对于基因的认识仅仅停留在逻辑概念阶段。1928年,格里菲斯发现当毒性光滑型肺炎球菌死株和无毒性粗糙型肺炎球菌活菌株混合后,部分粗糙型菌株转化为光滑型。在此基础上,1944年,美国的埃弗雷、麦克利奥特等科学家证明转化试验中的转化因子不是蛋白质,而是DNA。这首次用实验的方法直接证明了生物的遗传物质是DNA,而不是蛋白质。但这个发现没有得到广泛的认可,理由是人们怀疑当时的技术不能除净蛋白质,残留的蛋白质起到转化的作用。1952年,噬菌体小组主要成员美国的艾尔弗雷德·赫尔希和他的学生蔡斯完成了“噬菌体侵染细菌实验”,实验利用了放射性同位素标记的新技术标记蛋白质和核酸。实验证明,侵染细菌细胞的是噬菌体的核酸。这表明细菌携带遗传信息的是核酸而不是蛋白质,这一结果立即为学术界所接受。在前人的工作基础上,1953年沃森和克里克应用了X射线晶体衍射结构分析,提出了著名的DNA分子双螺旋结构模型,这为分析基因复杂功能奠定了结构基础。此后,遗传学的研究开始了突飞猛进、日新月异的发展。人们认识到,基因决定了细胞内核酸和蛋白质的合成,其中具有特定遗传效应的DNA片段会决定生物遗传性状。

1.1.2 基因的化学本质

在整个生物界中,对于细胞生物来说,基因的化学本质是

DNA,生物的遗传物质也是DNA。但对于非细胞生物来说(如病毒),其仅含有RNA和蛋白质,其遗传物质是RNA,基因化学本质也是RNA。所以说,DNA是生物的主要遗传物质。

DNA即脱氧核糖核酸(Deoxyribonucleic Acid)的英文缩写。DNA是由脱氧核苷酸重复排列组成的长链聚合物,每个脱氧核苷酸包括三个部分:碱基、戊糖和磷酸。构成DNA的核苷酸有四种:腺嘌呤脱氧核苷酸(dAMP)、胸腺嘧啶脱氧核苷酸(dTMP)、胞嘧啶脱氧核苷酸(dCMP)、鸟嘌呤脱氧核苷酸(dGMP)。脱氧多核苷酸长链是由这四种脱氧核苷酸按照一定顺序排列起来构成的,而排列在外侧的组成其长链骨架是两个相邻的脱氧核苷酸之间由脱氧核糖(五碳糖)与磷酸分子借由酯键相连,四种碱基排列在内侧。四种碱基里的任何一种都会与一个糖分子相连,这些碱基沿着DNA长链进行排列而成的序列,可组成遗传密码,指导蛋白质的合成。

DNA双螺旋结构的分子模型阐明了DNA分子的空间结构,后来被誉为20世纪以来生物学方面最伟大的发现,标志着分子生物学的诞生。构成DNA双螺旋结构模型的要点如下:(1)两条反向平行的多核苷酸链围绕同一条中心轴相互盘曲而成。(2)链的外侧是核糖与磷酸,内侧是碱基,碱基平面与螺旋轴垂直。(3)螺旋的两条链具有互补序列,其中G与C配对,A与T配对,两条链由碱基对间的氢键加以稳定。(4)螺旋的直径约为2 nm;沿螺旋轴方向每一圈有10个碱基对,相邻两个碱基对间的夹角为 36° ,双螺旋螺距为3.4 nm。(5)双螺旋表面有大沟和小沟之分,一般大沟较宽,而小沟较窄。(6)双螺旋结构在不同条件下可以表现出不同形式,如B-DNA,A-DNA,Z-DNA。

1.1.3 人类基因和基因组介绍

随着人类基因组测序的完成,科学家发现人体内估计有20 000到25 000个蛋白质的编码基因,占整个基因组序列的2%,远远低于原本估计的数目。虽然与某些较为原始的生物相比,人类的基因数量较少,但是在人类细胞中使用了大量的选择性剪接,将转录产物中穿插在内含子中的外显子以选择性的方式进行剪切和保留,形成不同的RNA剪切产物,这使得一个基因能够制造出多种

不同的蛋白质。也就是说,基因的编码效率在长期的进化过程中得到了加强,使得编码效率更高。大多数人类基因拥有许多的外显子,但人类的内含子比位在其两端的外显子更长。这些基因参差不齐地分布在染色体中,每一个染色体皆含有一些基因较多的区段与基因较少的区段。除了蛋白质编码基因之外,人类的基因组还包含了用来转录 tRNA(转运 RNA)、rRNA(核糖体 RNA)与 mRNA(信使 RNA)的数千个 RNA 基因。

人类基因或人类基因组中的功能序列可分为 4 大类:单一基因、基因家族、假基因和串联重复基因(序列)。(1)单一基因:指人的基因中 25%~50% 蛋白质基因在单倍体基因组中只有一份,称为单一基因或单一序列。(2)基因家族:从已克隆的许多基因中发现它们不完全是单拷贝的,有的是重复的多拷贝,这一部分基因属于两个或更多个相似基因的家族,它们在结构和功能上具有明显的相似性,编码相似的蛋白质产物,同一家族基因可以紧密排列在一起,形成一个基因簇,但多数时候它们是分散在不同的染色体上,或者存在于同一染色体的不同位置上,各自具有不同的表达调控模式。(3)假基因:假基因也叫伪基因,是指基因家族在进化过程中形成的无功能的残留物。它与正常基因相似,但丧失正常功能的 DNA 序列,往往存在于真核生物的多基因家族中。假基因可视为基因组中与编码基因序列非常相似的非功能性基因组 DNA 拷贝,一般情况都不被转录,且没有明确生理意义。(4)串联重复基因(序列):串联重复序列是指以相对恒定的短序列为重复单位、首尾相接、串联连接形成的重复序列,又称卫星 DNA(satellite DNA)。在人类基因组中,串联重复序列约占 10%,主要分布在非编码区,少数位于编码区。编码区中的串联重复序列与功能有关,非编码区串联重复序列多分布在间隔 DNA 或内含子中,重复单位短的仅 2 bp,长度可达数十碱基对,重复次数少则几次,多则几百次。重复序列的重复次数不同,是形成 DNA 长度多态性的基础。人类基因的编码序列往往被非编码序列所分割,呈现断裂状的结构,编码序列称为外显子,间隔的非编码序列称为内含子。内含子在转录后的加工中会被切除。除外显子、内含子外,人类基因还包括与外显子、内含子相邻的 DNA 序列组件,它们对基因转录的起

始和终止起调控作用。

人类基因组包括两个相对独立而又相互关联的基因组:核基因组和线粒体基因组。它们是人体所有遗传信息的总和。如不特别说明,人类基因组通常是指核基因组。人类基因组是由23对染色体组成,其中包括22对体常染色体、1对性染色体(2条X染色体或1条X染色体和1条Y染色体)。科学家将这些染色体分为A组(1—3号)、B组(4—5号)、C组(6—12号,X)、E组(16—18号)、F组(19—20号)、G组(21—22号,Y)共6个组。一般对于人类来说,每个细胞核内只有两套染色体。

构成人类基因组的DNA序列既可以是单拷贝序列,也可以是重复频率不等的多拷贝序列。(1)单拷贝序列:又称非重复序列,在单倍体基因组中只出现一次,在基因组中占50%~80%,如人类基因组中大约有60%~65%的序列属于这一类。单拷贝序列中储存了巨大的遗传信息,这些序列可编码各种不同功能的蛋白质。(2)重复多拷贝序列:重复多拷贝序列是指染色体上存在的大量无转录活性的重复DNA序列。其组织形式有两种:串联重复序列和分散重复序列。重复序列又可以分为高度重复序列、中度重复序列和轻度重复序列。高度重复序列一般是由少于10个核苷酸残基组成的短片段,重复几百到几百万次,如异染色质上的卫星DNA。重复次数为十次到几百次,如rRNA基因、tRNA基因和某些蛋白质的基因属于中度重复序列。重复2—10次的一些序列属于轻度重复序列,如tRNA基因和一些组蛋白基因。

1.1.4 人类基因组计划

1985年,美国科学家率先提出人类基因组计划(Human Genome Project, HGP)。HGP于1990年正式启动,旨在阐明人类基因组DNA上 3.2×10^9 个核苷酸的序列,发现所有人类基因并阐明其在染色体上的位置,破译人类全部遗传信息。HGP由美国、英国、法国、德国、日本和中国科学家共同参与,计划投资30亿美元,其与曼哈顿原子弹计划和阿波罗登月计划并称为三大科学计划,被誉为生命科学的“登月计划”。HGP的整体目标是阐明人类遗传信息的组成和表达;为人类遗传多样性的研究提供基本数据;揭示人类单基因异常和严重危害人类健康的多基因病的致病基因

或者疾病易感基因;建立对各种基因病新的诊治方法,从而推动整个生命科学和医学领域发展。其基本任务是建立人类基因组的结构图谱,即遗传图、物理图、转录图与序列图,并在“制图—测序”的基础上鉴定人类的基因,绘出人类的基因图。

(1) 人类基因组测序遗传图谱的构建:人类的遗传图谱是利用家系分析法,在对 8 个家系的 134 个成员的分析中,主要根据 5 264 个 STR 标记绘制而成的。利用这些家系的资料绘制第 1 至 22 号染色体图谱。对于 X 染色体图谱,还利用了来自另外 12 个家系、170 个成员的资料绘制而成。将 5 264 个标记定位在 2 335 个位点,据此构建的人类基因组计划遗传图谱的密度为每个标记 599 kb。

(2) 人类基因组测序物理图谱的构建:由于遗传图谱的分辨率有限、精确性不高,所以还要构建物理图谱。基因组物理图谱的构建主要有三种途径:①限制性酶图谱;②荧光原位杂交技术(FISH);③序列标签位点(STS),如来自 cDNA 的表达序列标签(EST)。

(3) 人类基因组测序转录图谱的构建:人类基因组计划要完成的另外一个图谱就是转录图。各种生命的现象都是通过蛋白质来表现和实现自己的功能,人类基因中只有少数是用来编码蛋白质的。因此,转录图抓住了这些能编码蛋白质的 DNA,就大致抓住了人类的基因,这就是转录图所要做的事情。

(4) 人类基因组测序序列图谱的构建:前面的物理图、转录图和遗传图都是序列前计划,这些图的绘制都是为人类基因组的序列图做准备,只有在序列图完成的基础上,才能将人群内序列的差异作为密度最高的遗传标记来完善遗传图,因此序列图是人类基因组计划中的最重要部分。归根到底,就是要把人体内约 2.5 万个基因的密码全部解开,同时绘制出人类基因的图谱。换句话说,就是要揭开组成人体 2.5 万个基因的 30 亿个碱基对的秘密。

2000 年 6 月 26 日,参加人类基因组工程项目的美国、英国、法国、德国、日本和中国的六国科学家共同宣布人类基因组草图的绘制工作已经完成。最终序列图于 2003 年完成,比预计提前两年,其序列错误率低于万分之一。人类基因组计划测定了组成人类基

因组的 30 亿个核苷酸序列,从而阐明人类基因组及所有基因的结构与功能,解读人类的全部遗传信息,揭开了人体奥秘的基础,是生命科学研究的大工程。人类基因组计划对生命科学的研究和生物产业的发展具有非常重要的意义。首先,获得人类全部基因序列将有助于人类认识许多遗传疾病、癌症等的致病机理,为分子诊断、基因治疗等新方法提供理论依据。其次,破译生命密码的人类基因组计划有助于人们对基因的表达调控有更深入的了解。对进化的研究不再建立在假说的基础上,而是利用比较基因组学,通过研究古代 DNA,来揭示生命进化的奥秘以及古今生物的联系,帮助人们更好地认识人类在自然界中的地位。人类基因组序列图不仅奠定了人类认识自我的基石,推动了生命与医学科学的革命性进展,而且为全人类的健康带来了福音,使我们向着更加幸福的未来迈出了意义非凡的一步^[1]。

随着 HGP 的完成,即遗传图、物理图、转录图、序列图等结构基因组学任务的基本完成,科学家们又启动了后基因组计划。以研究基因组的功能为主要任务的后基因组计划,也被称为功能基因组学,它是以提示基因组的功能及调控机制为目标,核心科学问题包括:基因组的多样性、基因组的表达调控与蛋白质产物的功能、模式生物基因组研究等。功能基因组学的研究将为人们深入理解以下几个方面提供共同的科学基础:人类基因组遗传语言的逻辑构架,基因结构与功能的关系,个体发育、生长、衰老和死亡机理,神经活动和脑功能表现机理,细胞增殖、分化和凋亡机理,信息传递和作用机理,疾病发生、发展的基因及机理(如发病机理、病理过程),以及各种生命科学问题。

1.2 基因的复制与表达

基因的基本功能和特性是由 DNA 分子中碱基对的排列顺序所蕴藏着的遗传信息决定的,即可以进行复制与表达。

1.2.1 基因的自我复制

能够自我复制是基因的一个重要特性。基因的自我复制也称为 DNA 的复制,其在细胞分裂周期的 S 期发生,以自身 DNA 分子为模板来合成新的 DNA 分子。正常复制的结果由一条双链变成

两条一样的双链,每条双链都与原来的双链一样。这个过程是通过半保留复制的机制来得以顺利完成的。

DNA 复制是一个连续的过程,为了便于理解,可将其划分为起始、延伸和终止三个阶段,现以大肠杆菌(*E. coli*)为例对该过程进行介绍。

(1) 起始:大肠杆菌的环形 DNA 分子上有一个固定的复制起始点(origin C),多种特定的蛋白因子和酶能准确识别并结合在复制起始点上,从而形成起始复合物。解旋酶能将 DNA 双链局部解开,形成 Y 形复制叉(replication fork);然后在 RNA 聚合酶的作用下,以 DNA 链为模板,沿 5'→3'方向合成一小段 RNA 引物以引导复制。复制往往是双向的复制(bidirectional replication),即从复制起点开始同时向两个相反方向进行。

(2) 延伸:以复制点一侧的 DNA 复制为例。在 RNA 引物的引导下,DNA 聚合酶催化子链沿 5'→3'方向延伸。在 3'→5'模板链上,DNA 新链按碱基互补原则沿 5'→3'方向连续复制,合成的子链称为前导链(leading strand)。而在 5'→3'模板链上,DNA 新链合成的方向与解链方向相反,需分段进行;这些先合成的、短的 DNA 片段称为冈崎片段(Okazaki fragment),冈崎片段经 DNA 连接酶连接形成的子链称为后随链(lagging strand)。前导链 DNA 的合成是连续的。而后随链则不连续,这种复制方式称为半不连续复制(semi-discontinuous replication)。

(3) 终止:RNA 酶水解 RNA 引物,并使新链继续延伸,填补引物水解后留下的空隙。最后在 DNA 连接酶作用下,将冈崎片段连接起来,完成 DNA 复制。复制完成后,一个亲代 DNA 分子生成 2 个子代分子,且每个子代分子均由一条亲代 DNA 链和一条子代 DNA 链组成,这种复制方式称为半保留复制(semi-conservative replication)。

真核细胞的 DNA 分子量大,通常与组蛋白结合形成核小体,并以染色质的形式存在于细胞核中,故真核细胞的 DNA 复制过程更为复杂,且速度较慢。真核生物的 DNA 复制除了拥有与原核生物相似的特点——双向复制、半保留复制和半不连续复制以外,还有其自身特点:①多起点复制,真核细胞的每个 DNA 分子有 100—

1 000 个复制起点,这些复制起点均能起始复制。②延伸速度慢,核小体的存在导致了 DNA 分子不易解链,从而影响了新 DNA 链的延伸。③RNA 引物和冈崎片段小,真核细胞 DNA 复制时合成的 RNA 引物约为 10 个核苷酸,原核细胞核苷酸则可达数十个;真核细胞的冈崎片段长 100—200 个核苷酸,而原核细胞的冈崎片段大约由 1 000—2 000 个核苷酸组成。④DNA 聚合酶不同,真核细胞有五种不同的 DNA 聚合酶,其中起主要作用的是聚合酶 α 及 δ ,聚合酶 α 的功能是起始复制,而 δ 则主要负责延伸子链。⑤复制完成前不能开始新一轮复制,真核细胞的 DNA 复制全部完成前,各复制起点不能开始新一轮的 DNA 复制,而快速生长的原核细胞则可在起始点处连续开始新一轮的复制。

复制形成新链的过程具有如下特点:①互补性:新链与复制前的双链 DNA 分子保持的结构完全一样,新链与模板链的碱基序列呈互补关系,这样就构成了一个完整的双链 DNA 分子。②半保留性:两条模板链分别成为子代 DNA 分子双链中的一条链,即在每个子代 DNA 分子的双链中,总是保留着一条母链。③反向平行性:DNA 分子的两条双链之间是反向平行的,一条是 $5' \rightarrow 3'$,另一条必然是 $3' \rightarrow 5'$;新复制出来的 DNA 分子的子链与母链也是反向平行的。④不对称性:DNA 的复制是不对称的。以 $3' \rightarrow 5'$ 母链作模板时,其子链合成是连续的;而以 $5' \rightarrow 3'$ 母链作模板时,子链的合成则是不连续的。⑤不连续性:真核细胞的 DNA 包含多个复制起点,其 DNA 的复制是以复制单位进行的。

1.2.2 基因的表达

基因的表达是所储存的遗传信息转变为由特定氨基酸种类和序列构成的多肽链,再由多肽链构成蛋白质,从而决定生物的各种性状(表型)的过程。基因表达包括两个步骤:①以 DNA 为模板转录合成 mRNA;②将遗传信息翻译成多肽链的过程。

(1) 转录

转录是遗传信息从 DNA 流向 RNA 的过程。即以双链 DNA 中的模板链为模板,以 ATP,CTP,GTP,UTP 四种核苷三磷酸为原料,在 RNA 聚合酶催化下合成 RNA 的过程。作为蛋白质生物合成的第一步,进行转录时,一个基因会被读取并被复制为 mRNA,即特