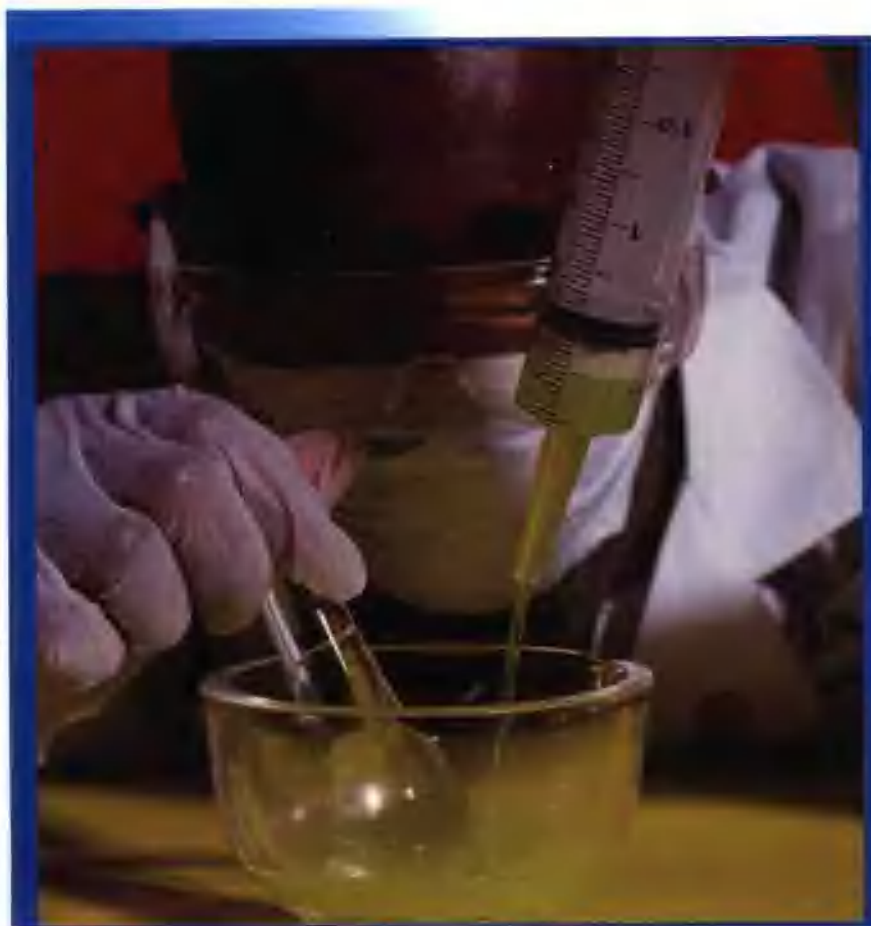




21世纪全国成人高等医药院校规划教材

医学遗传学基础



张国明 主编



中国科学技术出版社

21 世纪全国成人高等医药院校规划教材

医学遗传学基础

主 编	张国明				
编 委	翠 明	陈高云	王正群	高 珊	
	徐秀兰	杨永柱	杜建利	高新英	
	王乃友	张齐坤	朱贵良	王 尚	

中国科学技术出版社

· 北京 ·

21 世纪全国成人高等医药院校规划教材 丛书编委会

专 家 组：刘家权 郑伟清 杨绍珍 魏 玲 龚启梅 蔡 珍
梁观林 陈莉延 李明华 文 忠 宋燕丰 郭 祝
李 立 廖少玲 颜文贞 李春燕 邱锡坚 姜文平
韩晓杰 修 霞 于铁夫 聂亚玲 许堂林 万桃香

秘 书 处：陈露晓

责任编辑：周晓慧 高立波

封面设计：张 磊

责任校对：刘红岩

责任印制：王 沛

图书在版编目 (CIP) 数据

医学遗传学基础/张国明主编. —北京：中国科学技术出版社，2007.7

21 世纪全国成人高等医药院校规划教材

ISBN 978 - 7 - 5046 - 4751 - 1

I. 医... II. 张... III. 医学遗传学—成人教育：高等教育—教材 IV. R394

中国版本图书馆 CIP 数据核字 (2007) 第 109546 号

自 2006 年 4 月起本社图书封面均贴有防伪标志，未贴防伪标志的为盗版图书。

出版发行：中国科学技术出版社

社 址：北京市海淀区中关村南大街 16 号

邮 编：100081

电 话：010 - 62103210 传真：010 - 62183872

印 刷：广州市锐先印刷有限公司

开 本：787mm × 1092mm 1/16

印 张：7.25 字数：210 千字

版 次：2007 年 7 月第 1 版

印 次：2007 年 7 月第 1 次印刷

书 号：ISBN 978 - 7 - 5046 - 4751 - 1/R · 1265

定 价：11.00 元

前 言

近年来,医学各学科在人类基因组计划的推动下正在飞速地发展,医学遗传学成为横跨基础医学与临床医学的桥梁课程。鉴于医学遗传学在医学院校教育中日益重要的地位,以及发达国家医学专业的课程设置现状与趋势,我们特意组织了一批专家和一线教师在充分调研现行各版本教材使用情况的基础上,编写了该书。

从医学科学的发展来看,未来的医学工作者应该具有利用医学遗传学原理研究疾病的发生机制,并从医学遗传学的角度探讨疾病的诊断、治疗和预防的基本思路和基本手段。因此本书没有局限于某些遗传病及其相关的医学遗传学知识的介绍,而是更多地着力于疾病与遗传关系的介绍,将理论与实践相联系。随着经济全球化步伐的加快,医学教育与医学临床必定要接受来自国内外的挑战,全外文医学教育、双语医学教育要求已经摆在了医学教育工作者面前,所以教材中含有相关的英语名词,以作参考。

本书可作为全国高等医药院校专科、成人教育、高职高专学生的教材使用。

由于医学专业课程体系的建设尚在完善之中,医学遗传学又是一个不断发展的学科,其教学内容、实现形式都需不断探索,因此本次教材的编写肯定存在一些不足,也只能作为今后改版的基础。同时由于编者水平有限,诚恳希望广大使用本教材的师生提出批评和改进意见,在此表示谢意。

编 者

2007年5月

目 录

第一章 医学遗传学概述	(1)
第一节 医学遗传学研究的对象	(1)
第二节 医学遗传学在现代医学中的地位	(2)
第三节 医学遗传学研究的技术与方法	(3)
第四节 遗传性疾病概述	(5)
第二章 基因与基因突变	(8)
第一节 核基因组的序列组织	(8)
第二节 真核生物结构基因的结构	(9)
第三节 基因的表达与调控	(11)
第四节 基因突变	(14)
第三章 单基因遗传与单基因遗传病	(19)
第一节 遗传的基本规律	(20)
第二节 常染色体显性遗传	(21)
第三节 常染色体隐性遗传	(24)
第四节 性连锁遗传	(25)
第五节 影响单基因遗传病发病因素	(28)
第四章 多基因遗传与多基因遗传病	(30)
第一节 多基因遗传	(30)
第二节 多基因遗传病	(31)
第五章 人类染色体与染色体病	(38)
第一节 人类染色体	(38)
第二节 染色体畸变	(43)
第三节 染色体病	(53)
第六章 线粒体遗传病	(60)
第一节 线粒体基因组	(60)
第二节 线粒体基因突变与疾病	(62)
第七章 分子病与遗传性酶病	(64)
第一节 血红蛋白病	(64)
第二节 血浆蛋白病	(69)
第三节 遗传性酶病	(70)
第四节 受体蛋白病	(74)
第五节 膜转运载体蛋白病	(75)

第八章 群体中的基因	(76)
第一节 遗传平衡定律	(76)
第二节 影响群体遗传平衡的因素	(79)
第三节 近婚系数	(83)
第九章 药物反应的遗传基础	(87)
第一节 药物代谢的遗传控制	(87)
第二节 异常药物反应的遗传基础	(88)
第三节 毒物反应的遗传基础	(90)
第十章 肿瘤遗传	(92)
第一节 遗传性恶性肿瘤和癌前病变	(92)
第二节 染色体异常与肿瘤	(93)
第三节 癌基因与肿瘤抑制基因	(95)
第四节 肿瘤发生的遗传学说	(97)
第十一章 遗传病的诊断、预防和治疗	(99)
第一节 遗传病的诊断	(99)
第二节 遗传病的治疗与预防	(103)
第三节 遗传咨询	(106)

第一章 医学遗传学概述

第一节 医学遗传学研究的对象

医学遗传学(medical genetics)是遗传学与医学相结合的一门边缘学科,它所研究的对象是有关人类遗传的疾病,即遗传病(genetic disease)。

具体来说,医学遗传学研究人类遗传性疾病的发生机制、传递方式、发展规律,并为遗传病的预防诊断、治疗提供科学的依据和手段,从而控制遗传病在家庭或者群体中的发生,提高人类的健康水平和人口质量。

人类遗传学与医学遗传学是整体与部分的关系。人类遗传学(human genetics)是研究人类正常性状与病理性状的遗传现象及其物质基础的科学。主要研究其病理性状。所以,医学遗传学是在人类遗传学的基础上,从医学角度来研究人类疾病和遗传的关系,具有重要的临床意义。随着科学的发展和各个学科的相互渗透,医学遗传学与生物化学、生理学、免疫学、病理学、药理学、组织胚胎学、社会医学等学科的关系越来越密切,逐步形成了临床遗传学(clinical genetics),为遗传病的遗传咨询、预防与治疗、临床诊断等,提供了更便利的服务和更科学的方法理论。

医学遗传学在其发展过程中,已经引伸出许多分支学科,其中主要有:

细胞遗传学(cytogenetics)研究人类染色体数目异常和结构畸变的类型、发生频率、产生机制以及其与疾病的关系。至今,人们已经认识到 100 多种染色体异常综合征和 10000 多种罕见的异常核型。

生化遗传学(biochemical genetics)是用生物化学的方法研究遗传病中的蛋白质或酶以及核酸的相应改变。从而加深了人们对分子病(molecular disease)和遗传性代谢病(genetic metabolic disease)的认识和了解

分子遗传学(molecular genetics)是用现代新技术从基因的结构、突变、表达、调控等方面研究遗传病的分子变化,为遗传病的基因诊断、基因治疗等提供了新的策略和手段。

群体遗传学(population genetics)是研究群体的遗传结构及其遗传变化规律。医学群体遗传学或者遗传流行病学(genetic epidemiology)则是研究人群中遗传病的发病率、遗传方式、基因频率、基因型频率、携带者频率以及其影响因素,包括突变、选择、迁移、隔离、婚配等,以便控制遗传病在群体中的流行。

遗传毒理学(genetic toxicology)是研究环境因素对遗传物质的损伤机制和原理,及这些致病因素(诱变剂、致畸剂、致癌剂)的检测方法和评价手段。

免疫遗传学(immunogenetics)是研究免疫反应的遗传基础和遗传调控。例如抗原的遗传控制、抗体多样性产生的遗传机制、补体的遗传基础等,为控制免疫过程、阐明免疫缺陷病提供

手段。

药物遗传学(pharmacogenetics)是研究药物代谢的遗传差异和同一药物对不同个体产生不同反应的遗传基础,从而指导临床医生因人施药以及为新药物的开发提供科学依据

肿瘤遗传学(cancer genetics)是研究肿瘤发生、生长的遗传基础,其中包括染色体的改变、癌基因、抗癌基因的作用、突变与修复等,阐明肿瘤发生发展的遗传机制,对肿瘤的预防、诊断和治疗具有积极而重大的意义。

发育遗传学(developmental genetics)是研究胚胎发育过程中,双亲基因组的影响、同源框、基因表达的时序等,对阐明发育过程的遗传控制有重要作用。

行为遗传学(behavior genetics)是运用各种遗传学方法研究人类行为的遗传学科,尤其是异常行为,例如癫痫、躁狂抑郁病、精神分裂症、Alzheimer 病等的遗传基础,以控制其发生。

从上述医学遗传学的分科来看,它的研究领域非常广泛,而且与临床医学紧密相关。所以,医学遗传学在医学教育中占有一席之地。

第二节 医学遗传学在现代医学中的地位

医学遗传学近年来迅猛发展,俨然已成为现代医学中的领军学科之一。它与基础医学、临床医学、预防医学关系密切,并相互促进影响。

一、在基础医学研究中的地位

医学遗传学的发展,不仅促进了细胞培养技术、生化检验技术和免疫学实验技术等基础医学研究的基本技术的发展和完善,而且拓宽了基础医学研究的内涵,使基础医学的研究不断深入。例如,DNA 的双螺旋结构被发现使人们能够用分子生物学语言来诠释生物界千变万化的遗传现象。而今,分子生物学的理论和技术已高速、高效地渗入医学遗传学领域,使其扩展和加深了医学遗传学基本理论和技术,并在疾病的发生机制、病原学检测、新药开发等基础医学研究中发挥了巨大作用。特别是 20 世纪 80 年代启动的人类基因组计划,使人们可以从分子水平上揭示疾病的本质,并预示其最后的完成,这必将使人们对疾病的认识及治疗发生根本性的转变。

二、在临床遗传病研究中的地位

1984 年诺贝尔奖金获得者保罗·伯克说:“几乎所有疾病都和遗传有关,遗传学的研究是治疗所有疾病的关键。”虽然现代医学技术的迅猛发展,使得危害人类健康与生命的传染病已得到基本有效的控制,但是遗传病无论是其相对发病率还是绝对发病率都存在着明显增加的趋势,其对人类健康的威胁已不容忽视。因此,医学遗传学的基本原理和技术在临床遗传病研究中占有着不可替代的地位。

三、在优生研究中的地位

据世界卫生组织(WHO)1973 年技术报告,在加拿大 Montreal 儿童医院 1969 ~ 1970 年的 1146 名住院儿童中,患有与遗传有关的疾病占 29.4%。根据 1989 年的普查,我国新生儿中约有 1.3% 有严重出生缺陷(birth defect)或先天畸形(congenital malformation),据统计,其中有 70% ~ 80% 涉及遗传因素。因此,我们可以粗略估计出我国每年出生的 1500 万个孩子中,患

有遗传因素所致的严重出生缺陷或先天畸形竟约有 13 万 ~ 15 万。此外,出生后由于所携带的致病基因的表达也可能出现某种遗传病,例如,假肥大型肌营养不良、血友病 A、家族性多发性结肠息肉等。再则,智力低下(mental retardation, MR)在我国人群中的发生率为 2.2%,对我国人口质量产生重大的原因,多于 1/3 的智力低下者是由于多基因、单基因或者当染色体遗传物质改变而引起的。

对于一个人口大国,人口问题始终是制约国民经济发展的的重要因素。为了有效地解决人口问题,国家提出了“控制人口数量,提高人口素质”的基本国策。而只有依靠医学遗传学原理指导人类生育,才能提高人类的体质和智力,真正实现优生的目的。

四、在卫生保健研究工作中的作用

据估计群体中每个人都携带有 5 ~ 6 个隐性有害基因,这些致病基因携带者(carrier)虽然不患病,但却可以通过有性生殖的方式将致病基因传给后代,而使后代患病。也就是说,不为遗传病所累的正常人,也可能携带有某种隐性致病基因。此外,一些严重危害人类健康的常见病已经证明与遗传因素有关,诸如动脉粥样硬化、冠心病、高血压、糖尿病、精神分裂症等。所以说,医学遗传学的研究有助于人类的卫生保健。

综上所述,随着科学的进步,医学遗传学的发展,先进研究技术和方法的应用,人们一定会逐步了解危害人类健康的遗传性疾病产生的原因,最终有效地控制这些疾病的发生,找到行之有效的治疗方法,从而提高人口质量和人类健康水平。

第三节 医学遗传学研究的技术与方法

由于医学遗传学是一门交叉学科,其研究手段中广泛地包含了细胞学、生物化学、免疫学、生物统计学的研究技术和方法。而针对不同的研究对象和目的时,又具有它独特的研究方法,这些方法的运用往往是用来确定某种疾病与遗传因素是否有关。

一、系谱分析法

临床上判断单基因病的遗传方法常用系谱分析法,系谱分析的基本程序是先对某遗传病患者家庭成员的发病情况进行详细调查、再用一定方式将调查结果绘成系谱,然后根据孟德尔定律对成员的表现型和基因型进行分析。通过系谱分析来判断某种遗传病是单基因还是多基因病,以及确定单基因病的遗传方式和遗传异质性的存在。

二、群体筛查法

群体筛查法是采用一种或者几种高效、简便且较为准确的方法对一般人群或者特定人群(如患者亲属)进行某种遗传病或性状的普查。主要目的:①了解某种遗传病在群体中的发病率及其基因型频率;②筛查某种遗传病,尤其是隐性遗传病杂合子携带者;③与家系调查相结合判断某种疾病是否与遗传因素有关。

三、双生儿分析法

双生是指一次娩出两个胎儿,俗称双胞胎,双生分为一卵双生和二卵双生。从发生学角度来看,一卵双生是指由同一受精卵卵裂发育而成的两个个体,他们之间的遗传物质是相同的,如血型、性别等性状是一致的,若存在差异,则是由环境因素作用的结果。二卵双生是由两个

受精卵发育形成的两个个体,由于他们之间的遗传物质不完全相同,在各种性状上,存在着较大差异。

双生儿分析法则是通过比较一卵双生儿和二卵双生儿表型特征的一致性和不一致性,来分析遗传因素和环境因素,在生理和病理性状发生中各自作用的程度,判断某种性状(或者疾病)是否与遗传因素有关。两种双生子可以从外貌特征、皮肤纹理、血型、同工酶谱、DNA 多态等方法加以鉴定。一般用发病的一致率来表示:

$$\text{发病一致率}(\%) = \frac{\text{同病双生子(单卵和双卵)对数}}{\text{总双生子(单卵或双卵)对数}} \times 100\%$$

如果某一疾病在两种双生子中的发病一致率没有显著差异,说明环境因素是该病的主要作用因素,如果某一疾病在两种双生子中的发病一致率存在十分显著差异,则说明该病的发生与遗传因素有关。

表 1-1 列举了双生子法研究的几种与遗传因素有关的疾病。

表 1-1 几种疾病单卵双生子与双卵双生子发病一致的比较

疾病名称	单卵双生病一致率(%)	双卵双生发病一致率(%)
先天愚型	89	7
精神分裂症	80	13
结核病	74	28
糖尿病	84	37
原发性癫痫	72	15
十二指肠溃疡	50	14
麻疹	95	87

从表 1-1 中可以看出,传染病中的麻疹在单卵双生子与双卵双生子之间的发病率差异最小,说明麻疹与遗传因素的关系最小。先天愚型在单卵双生子与双卵双生子之间的发病率差异最大,说明先天愚型与遗传因素最为密切。

四、实验室检查法

实验室检测是用实验模拟的方式确定某种疾病是否为遗传病,这是遗传病判断的最可靠方法。例如染色体畸变所导致的染色体病,必须依靠染色体核型分析;代谢病与分子病则必须依据生物化学的检查和基因分析。

五、种族差异比较法

种族是在地理、文化、风俗习惯及宗教信仰上相对隔离的群体,因而在繁殖上也存在隔离的群体。世界上主要有高加索人(白种人)、黑人、亚洲黄种人、美洲的印第安人等六种类型,每个种族的基因库彼此不同,它们之间不但在外表、肤色、虹膜、颧骨等方面显示出遗传学差异,而且在血型、HLA 类型、血清型、同工酶的基因频率等方面也各不相同。因此,如果某种疾病在不同的种族具有不同的发病率、发病年龄不同,甚至临床表现不同,就应该考虑这种疾病与遗传因素相关。例如美籍华人的鼻咽癌发病率高出美国白人 34 倍,且中国人的鼻咽癌发病率位居全球之首。因此,可以推断鼻咽癌的发生与遗传因素有关。但是不同的种族因为生活环境、气候条件、饮食习惯、社会经济状况等的差异,也可能造成某种疾病的发病率的差异,在

调查不同种族的发病率差异时,应该严格排除这些环境因素对疾病的影响。

六、分子生物学方法

分子生物学方法是研究疾病在分子水平的发生机制,主要运用基因克隆、基因定位等手段,寻找已经确定与遗传因素有关的疾病的基因,然后将基因定位于染色体的具体位点,并克隆出与疾病相关基因,从而为遗传病的预防、诊断、治疗提供新的方法。

重组 DNA 技术(recombinant DNA technique)是分子生物学方法中的核心技术,即将一种生物的 DNA 片段或人工合成的基因,通过载体在体外重组,转移并插入到另一生物的基因组中,使这种生物表达新的目的性状或用于进一步地深入研究,从而鉴别和治疗人类遗传病。

七、疾病组分分析法

疾病组分分析法主要用于研究一些复杂且发病机制不清的疾病的遗传基础。具体操作是先将这种病分解为若干环节(组分),对各组分进行单独的遗传学研究,当某些组分被确定为受遗传控制的,则可以认为这种病是有遗传基础的。

八、关联分析法

关联(association)是指两种遗传上独立的性状非随机、非连锁地同时出现。如果其中一种性状与遗传因素有关,就可作为遗传标记(genetic marker),来检测另一种性状与之是否关联。如果确证有关联,则表明后一性状也有遗传基础。

第四节 遗传性疾病概述

一、遗传病的概念

遗传病是由于遗传物质改变所造成的疾病,并且可以在上、下代之间按一定的方式垂直传递。其基本特征是根本上由于遗传物质改变而发生的,但也受环境因素的作用。例如,某些遗传性疾病的发病是环境因素在不同程度上的诱发而发病的。

有人认为家族性疾病(familial disease)就是遗传病。家族性聚集发病只是遗传病的一个特点,因共同的致病基因继承而表现,但这并不意味着家族性疾病就等于遗传病。首先,一些常染色体隐性遗传病通常不是家族性发病而是散发病例(sporadic case);再则,一些环境因素所致的疾病中,由于同一家族的不同成员均生活在相同的环境中,而表现出发病的家族聚集性。例如,在某些缺碘地区,甲状腺功能低下所致的痴呆病就有发病的家族聚集,但这并不是遗传病。

还有一种误解,认为先天性疾病(congenital disease)就是遗传病。所谓先天性疾病(congenital disease)是指个体一出生时就具有的疾病。例如,先天畸形是指个体一出生就表现出机体或者某些器官系统的结构形态异常。这种畸形可以是遗传的,也可能是由胚胎发育过程中的环境因素所致。20 世纪 70 年代某些国家用“反应停”治疗妊娠反应,导致多例胎儿畸形。再例如孕妇怀孕期间感染风疹病毒可以导致胎儿先天性心脏病。这些疾病不是遗传病。而唇裂、先天性巨结肠、脊柱裂等则多是由遗传因素所致且属于多基因遗传病,是可遗传的。

此外,遗传病不一定出生时就表现出疾病的症状,有的是在出生后的漫长生命过程中逐步表现出来的,因此遗传病不代表先天性。例如甲型血友病一般在儿童期发病;成年型多囊肾在

中年后发病;先天性家族性多发性结肠息肉一般在青壮年期发病。

二、疾病发生中的遗传因素与环境因素

遗传(heredity)是生物体的基本生命现象,表现为性状在亲代与子代之间的相似性和连续性。虽然人类的一切正常或异常的性状,综合起来看都是由遗传与环境共同作用的,但它们在每一具体性状上的表现程度却不尽相同(图1-1)。

1. 完全由遗传因素决定发病

这类疾病的发生并非与环境因素无关,是目前没有发现其发病必须具有何种特定的环境因素,例如单基因遗传病中的先天性成骨不全症、白化病、血友病A以及一些染色体病。

2. 主要由遗传因素决定发病

这类疾病发病需要依靠环境因素的诱发。例如苯丙酮尿症是常染色体隐性遗传病,但必须在进乳成食用含苯丙氨酸量高的食物情况下才能发生。蚕豆病除具有X连锁致病基因所致葡萄糖-6-磷酸脱氢酶(G-6PD)缺乏的遗传基础外,还必须在进食蚕豆后才会诱发溶血性贫血。

3. 遗传因素和环境因素对发病都有作用

在不同的疾病中遗传因素对发病作用的影响程度是不同的,也就是说,其遗传度各不相同。例如在唇裂、腭裂、先天性幽门狭窄等畸形中,遗传度都在70%以上,说明虽然环境因素也有一定作用。遗传因素对这些疾病的发生较为重要,另一些疾病,例如先天性心脏病、十二指肠溃疡、某些糖尿病等的发生,环境因素占较大比重,而遗传因素的作用较小,遗传度不足40%,但是只有必须存在这个遗传基础。才能发病,还有一些疾病如脊柱裂、无脑儿、高血压、冠心病等的发病,遗传因素和环境因素相等重要,遗传度约50%~60%左右。

以上三类疾病的发病都需要遗传基础,所以都是遗传病。

4. 完全由环境因素决定发病

发病完全取决于环境因素,与遗传因素基本无关。例如烧伤、烫伤等外伤的发生与遗传因素无关。有人认为,这类疾病损伤的修复和恢复依赖于个体的遗传类型。

三、遗传病的分类

现代医学将遗传病分为基因病、染色体病、体细胞遗传病等(图1-2),其遗传方式各不相同。

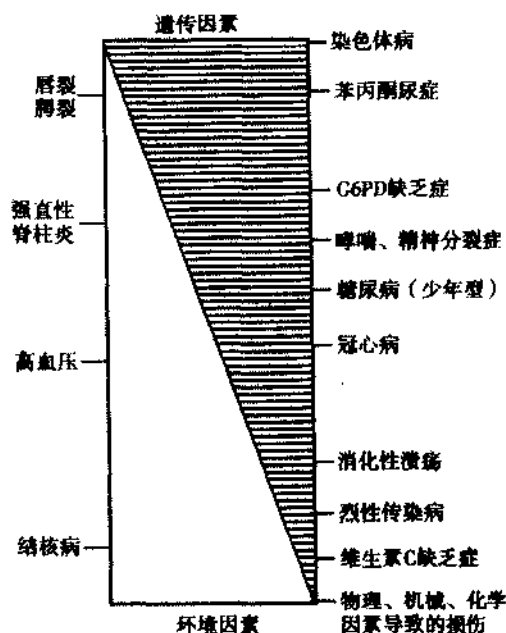


图1-1 疾病发生中的相互作用

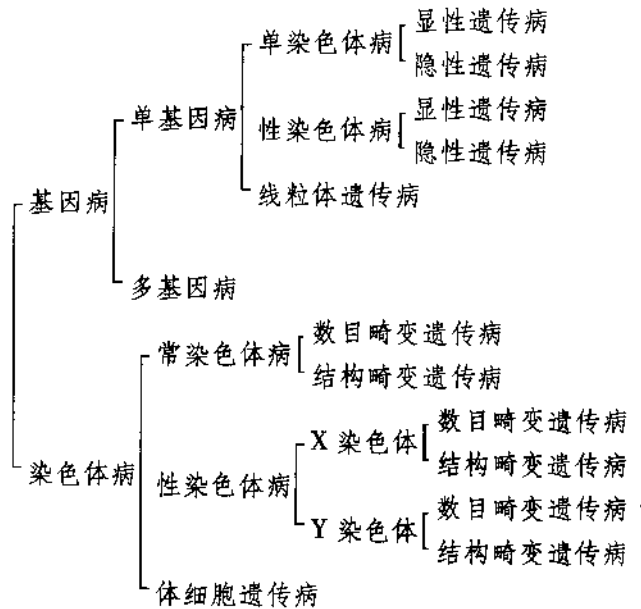


图 1-2 遗传病的分类

1. 单基因病

单基因病由单基因突变所致。突变发生在两条染色体中的一条,则引起常染色体(或性染色体)显性遗传;突变作用在两条染色体上,而引起常染色体(或性染色体)隐性遗传。这类单基因病较少见,发生率较高时也仅为 1/500,但由于其遗传性,危害也较大。

2. 多基因病

多基因病又称为多因子病(multifactorial disease, MF),具有复杂的病因,既涉及到遗传基础,又需要环境因素的作用才发病。其遗传基础不仅仅是一对基因,而是涉及到许多对基因,具有一定家族史,但没有单基因性状遗传中所见到的系谱特征。

3. 染色体病

人的体细胞中有 23 对染色体,1~22 号为常染色体,X 和 Y 为性染色体。染色体数目或结构的改变所致的疾病称为染色体病。染色体病亦分为常染色体病和性染色体病。由于染色体上共有 3.5 万基因,所以染色体病往往涉及许多基因,临床上常表现为复杂的综合征(syndrome)。

4. 体细胞遗传病

体细胞中遗传物质的改变所致的疾病称为体细胞遗传病。肿瘤起源于体细胞遗传物质的突变,尽管这种突变不会传给后代,但是在体内可以随着细胞的分裂而不断传给新产生的子代细胞。各种肿瘤的发生都涉及特定的组织细胞中的染色体、癌基因、抑癌基因的改变。有的先天畸形是在发育过程中某种细胞的遗传物质改变而引起的,所以也属于体细胞遗传病,如孕期感染风疹病毒导致的先天性心脏病。

第二章 基因与基因突变

基因最初被称为遗传因子,1909年,Johannsen将遗传因子改称为基因。摩尔根在1926年发表的《基因论》提出,基因是位于染色体上呈直线排列的遗传单位。1941年,Beadle和Tatum在生化反应的遗传控制研究中,通过对红色链孢霉的营养缺陷型进行分析,证明基因的功能是调控酶的合成,从而提出了“一个基因一种酶”学说。1944年Avery等对肺炎双球菌的研究证明脱氧核糖核酸(DNA)是遗传物质。1953年,Watson和Crick提出了DNA分子双螺旋结构模型,阐明了DNA自我复制的机制,蕴藏在DNA分子碱基顺序中的遗传信息,以及转录和翻译控制蛋白质的合成过程。1977年Brooker和Sharp等人对真核生物基因结构的研究中发现了断裂基因,并证实大部分真核生物的结构基因是断裂基因。到此为止,基因的本质、功能及结构才逐渐为人们所认识。

遗传学的研究始终是以基因的结构和功能为中心展开的。20世纪中叶人们就已证实,生物体的遗传信息不仅蕴藏在DNA中,而在某些病毒中,RNA也是遗传物质。自此,根据目前的认识,在分子生物学水平上的研究,基因又重新定义为遗传的功能单位,是能够表达和产生基因产物(蛋白质或RNA)的核酸(DNA或RNA)序列。基因有三个基本特征:①基因可以自我复制。基因的复制实际上是DNA的复制,通过复制,使遗传的连续性得到保持;②基因决定性状。基因通过转录和翻译决定多肽链的氨基酸顺序,从而决定某种酶或蛋白质的性质,最终表达为某一性状;③基因可以产生突变。基因虽很稳定,但也会发生突变,新突变的基因一旦形成,可通过自我复制在随后的细胞分裂中保留下来。

基因组(genome)是指细胞或生物体的全套遗传信息。人类基因组包括核基因组与线粒体基因组。但在无特殊证明的情况下一般所说的人类基因组是指核基因组。

第一节 核基因组的序列组织

一、单一序列和重复序列

根据某种DNA片段在基因组拷贝数的不同,将基因组的DNA分为单一序列和重复序列。重复序列又可根据重复次数的不同分为高度重复序列和中度重复序列。

单一序列(unique sequence)是指在一个基因组中只有一个拷贝或很少几个拷贝的DNA序列。单一序列约占基因组DNA的50%~60%。其包括编码蛋白质的序列以及间隔序列,但前者只占很少一部分。单一序列常被重复序列隔开。

高度重复序列(highly repetitive sequence)是指在一个基因组中存在大量拷贝的DNA序列。这些序列通常很短,其长度为6~200bp,拷贝数为 $10^6 \sim 10^8$,散在的分布于基因组中,占基因组DNA的10%~30%。高度重复序列大多数集中在异染色质区,目前认为其主要功能是参与维持染色体的结构,以及协助减数分裂时染色体配对。

中度重复序列(intermediate repetitive sequence)是以不同的量重复分布于整个基因组的不同部位。这些序列可短至 100 ~ 500bp,称为短分散元件(short interspersed element),也可长达 6000 ~ 7000bp,称为长分散元件(long interspersed element),占整个基因组的 25% ~ 40%。

Alu 家族(Alu family)是短分散元件典型的例子,是人类基因组含量最丰富的中度重复序列,长达 300bp,在一个基因组中重复 30 万 ~ 50 万次。Alu 序列内含有一个限制性内切酶 Alu I 的特异性识别位点 AGCT,因此它可被 Alu I 酶裂解为 170bp 和 130bp 的两个片段。研究表明神经母细胞纤维瘤发生是由于 NF1 抑癌基因突变所致,一个新的 Alu 序列插入患者的 1 对 NF1 等位基因之一,使这一 NF1 基因失活,当另一个 NF1 基因发生同样突变后,就造成肿瘤的发生。

Kpn I 家族(Kpn I family)是长分散元件的典型例子,是中度重复顺序中仅次于 Alu 家族的第二大家族,形成 6.5kb 的中度重复顺序,拷贝数为 3000 ~ 4800 个,约占人体基因组的 3% ~ 6%。用限制性内切酶 Kpn I 消化,可分解成 4 个长度不等的片断,分别是 1.2kb、1.5kb、1.8kb 和 1.9kb,其具体功能不清。

二、多基因家族

多基因家族(multigene family)是指由一个祖先基因(ancestral gene)经过重复和变异所产生的一组来源相同,结构相似,功能相关的基因。多基因家族是真核基因组中重要的结构之一。多基因家族根据其在染色体上的分布,大致可分为两种类型:一种是高度同源,成簇地排列于同一条染色体上,形成一个基因簇,同时发挥作用,编码某些蛋白质;另一种类型是一个基因家族中的不同成员成簇地分布在不同的染色体上,编码一组关系密切的蛋白质,其碱基序列可能不尽相同。人类组蛋白基因家族和珠蛋白基因家族均为多基因家族的典型例证。

三、假基因

假基因是指在多基因家族中,不产生有功能的基因产物的基因。假基因的核苷酸顺序与相应的活性基因极为相似,它们可能由于发生了突变,失去了原有的活性,从而不能表达的无功能基因。例如, α 珠蛋白基因族中的假基因 $\psi\alpha$ 与 α 基因相比,只是失去了内含子。这种无内含子的假基因可能是由相应活性基因的 mRNA 经过逆转录成为 cDNA,再整合到基因组而形成的。假基因可以与有功能的基因相连,也可以通过染色体易位或作为转座子的一部分,从一个部位转移到另一个部位。

第二节 真核生物结构基因的结构

编码蛋白的基因称为结构基因。真核生物(包括人类)与原核生物的结构基因有所不同:在基因的数量和大小上,原核生物的基因较少,往往只有一个 DNA 分子,基因大小一般为 1kb;真核生物的结构基因数量较多,基因彼此间的大小相差较大。在基因结构上,原核生物为连续基因即编码蛋白质的基因核苷酸序列是连续的;真核生物结构基因的核苷酸序列包括编码序列和非编码序列两部分,非编码序列将编码序列隔开,而在 DNA 分子不连续分布称为断裂基因。另外,原核生物有较多的重叠基因,即两个或两个以上的基因共用一段 DNA 序列;而真核生物基因组中仅有极少数的重叠基因。

真核生物的结构基因由外显子和内含子组成的编码区及其两侧的侧翼序列组成。(图 2-1)。

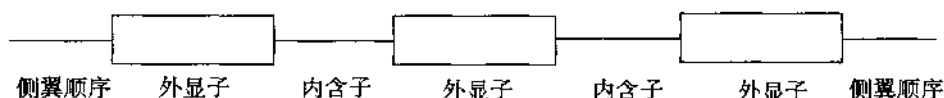


图 2-1 结构基因示意图

真核细胞的结构基因中含有的编码顺序,称为外显子(exon)。两个外显子之间的无编码功能的基因,称为内含子(intron)。不同结构基因所含有的内含子数目和大小也不同。例如,人血红蛋白 β 珠蛋白基因有三个外显子和两个内含子,全长约 1700 个碱基对,编码 146 个氨基酸(图 2-2)。人的假肥大型进行性肌营养不良(DMD)基因有 75 个外显子和 74 个内含子,全长 2300kb,编码 3685 个氨基酸。在每个外显子和内含子的接头区都存在有高度保守的一致顺序,称为外显子—内含子接头,即在每个内含子的 5'端开始的两个核苷酸为 GT,3'端结束的两个核苷酸是 AG,这种接头形式通常称为 GT—AG 法则。

断裂基因中的内含子和外显子的关系是不完全固定的。即在同一条 DNA 分子上的某一段 DNA 顺序,在编码某一条多肽链时是外显子,而当它作为编码另一条多肽链的基因时却是内含子,这是由于 mRNA 剪接加工的方式不同所致。结果使同一个基因(确切地说是同一段 DNA 顺序)产生两条或者两条以上不同的 mRNA 链。这使真核生物基因的表达,产生多样性,从而构成了断裂基因结构上一个重要特点。

侧翼序列是位于每个结构基因在第一个和最后一个外显子外侧,一段不编码氨基酸的非编码区,起着对基因表达调控的作用。

侧翼序列包含有启动子、增强子和终止子。

一、启动子(promoter)

启动子是通常位于基因转录起始点上游 100bp 的范围内的---段特异 DNA 序列,是 RNA 聚合酶的结合部位,能启动并促进转录过程。启动子包括以下几种重要的结构序列(图 2-2)。

(1)TATA 框(TATA box),位于基因转录起始点上游大约 -19 ~ -27bp 处,高度保守,由 7 个碱基组成,顺序为 TATAAAA 或 TATATAT,该序列只有两个碱基(A/T, A/T)可以变换,周围富含 GC 的顺序。TATA 框通过与转录因子 TF II 结合,再与 RNA 聚合酶 II 形成复合物,从而能够准确识别转录起始点。

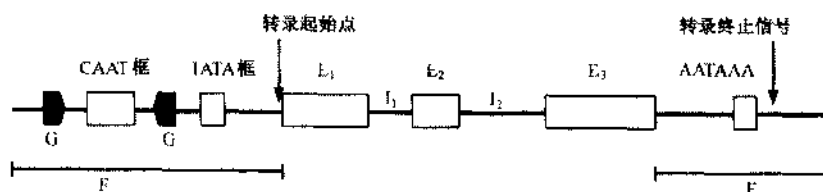


图 2-2 人血红蛋白 β 珠蛋白基因

E:外显子; I:内含子; F:侧翼序列; G:GC 框

(2) CAAT 框(CAAT box), 位于转录起始点上游约 -70 ~ -80bp 处, 也是保守序列, 由 9 个碱基组成, 其顺序为 GGCCAATCT 或 GGCTAATCT, 其中只有一个碱基(T/C)可以变化。转录因子 CTF 可以识别 CAAT 框并与之结合, 提高转录效率。

(3) GC 框(GCbox)有两个拷贝, 其序列为 GGCGGG, 分别位于 CAAT 框的两侧, 能与转录因子 SP1 结合, 起到增强转录效率的作用。

二、增强子(enhancer)

增强子是位于启动子上游或下游的一段 DNA 序列, 它可以增强启动子的转录能力, 提高基因转录的活性。当它被基因活化蛋白识别并结合后, 通过此蛋白与结合于启动子上的蛋白质之间的相互作用而起到增强转录作用。增强子作用无明显的方向性, 可以是 5'→3' 方向, 也可以是 3'→5' 方向, 且位置不固定。例如人的 β 珠蛋白基因的增强子是由两个相同顺序的 72bp 串联重复顺序组成, 可位于转录起始点上游 -1400bp 或下游 3300bp 处, 当它被激活时, 能使转录活性增强 200 倍。

三、终止子(terminator)

终止子是位于 3' 端非编码区下游的一段碱基序列, 由一段反向重复序列(inverted repeat sequence)以及特定的序列 5' - AATAAA - 3' 组成, 在转录中提供转录终止信号。AATAAA 的反向重复序列是 RNA 聚合酶停止工作的信号, 该序列转录后, 可以形成发卡式结构, 从而阻碍了 RNA 聚合酶的移动, 并由于其末尾的一串 U 与模板中的 A 结合不稳定, 从而使 mRNA 从模板上脱落, 转录终止。

第三节 基因的表达与调控

一、基因的表达

基因的表达(expression)是指通过转录和翻译将 DNA 分子中所蕴藏的遗传信息转变成蛋白质分子的过程。在原核生物中, 转录和翻译两个过程是同步进行的, 随着 mRNA 的转录, 核糖体即与之结合, 并进行翻译产生蛋白质。在真核生物中, 结构基因的转录是在细胞核中进行的, 而翻译是在细胞质中进行的。

1. 转录(transcription)

转录是在 RNA 聚合酶的催化下, 以 DNA 的反编码链为模板, 按照碱基互补的方式合成 mRNA 分子的过程, 是基因表达的第一步。真核生物基因的转录是在细胞核中进行的。对于任何一个特定的基因来说, DNA 双链分子中只有一条链带有遗传信息, 即编码链, 而另一条为其互补顺序的链, 称为反编码链。首先转录成的 RNA 分子由于大小长短形态不一称为核内异质 RNA(hnRNA), 它包括外显子、内含子和部分侧翼顺序。hnRNA 要经过戴帽、剪接、加尾等加工过程才能形成成熟的 mRNA。

(1) 剪接: 剪接(splicing)是指在酶的作用下, 按 GU - AG 法则将 hnRNA 中的内含子转录序列切掉, 然后将各个外显子转录的序列按照顺序连接起来的过程。

(2) 戴帽: 戴帽(capping)是指在剪切的 mRNA 分子的 5' 端的第一个核苷酸(通常为 G)的前方连接上一个 7 - 甲基鸟嘌呤核苷酸, 类似一个帽子结构。这个“帽”使 mRNA 移入细胞质