

浙江省高校重点建设教材
宁波大学教材建设项目资助

人体基因学

Human Genetics

◎ 主编 郭俊明 ◎ 副主编 肖丙秀



ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

013069651

Q987

11

浙江省高校重点建设教材
宁波大学教材建设项目资助

人体基因学



主 编 郭俊明

副主编 肖丙秀

编 者 (按姓氏笔画为序)

习 阳 肖丙秀 季林丹

郭俊明 蒋孝明



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社



北航

C1677744

Q987
11

内 容 提 要

人体基因学是研究人体基因的结构与功能的新兴学科。随着人类基因组计划的实施,人们对自身遗传物质的认识进入了新时代。基因与人的生、老、病、死密切相关。基因的研究成果已广泛渗透到政治、经济和社会生活的各个方面。本书内容包括:绪论、基因与基因组、基因的功能、基因与疾病、基因克隆、基因与社会等六章。

本书可用作高等院校公共选修课和大学生素质教育的教材,也可供医学、生命科学相关专业的科研人员、教师、研究生和本科生参考,还可作为大众普及基因知识的参考资料。

图书在版编目(CIP)数据

人体基因学/郭俊明主编. —杭州:浙江大学出版社, 2013.8

ISBN 978-7-308-12169-9

I. ①人… II. ①郭… III. ①人类基因—高等学校—教材 IV. ①R394

中国版本图书馆CIP数据核字(2013)第200753号

人体基因学

郭俊明 主编 肖丙秀 副主编

责任编辑 张凌静 (zjl@zju.edu.cn)

封面设计 刘依群

出版发行 浙江大学出版社

(杭州市天目山路148号 邮政编码310007)

(网址: <http://www.zjupress.com>)

排 版 杭州立飞图文制作有限公司

印 刷 杭州日报报业集团盛元印务有限公司

开 本 787mm×1092mm 1/16

印 张 10.5

字 数 263千

版 次 2013年8月第1版 2013年8月第1次印刷

书 号 978-7-308-12169-9

定 价 23.00元

版权所有 翻印必究 印装差错 负责调换

浙江大学出版社发行部邮购电话(0571)88925591; <http://zjdxcs.tmall.com>

前 言

人体基因学是研究人体基因的结构与功能的新兴学科。随着人类基因组计划的实施,人类对自身遗传物质的认识进入了新时代。基因与人的生、老、病、死密切相关。基因的研究成果已广泛渗透到政治、经济和社会生活的各个方面。基因诊断和基因治疗是医学的重大进步,基因工程产品改变了人们的食品结构,克隆技术的产生引出了伦理学的新课题,大量 DNA 序列的测定促成了生物信息学的兴起和发展,生物芯片技术是多种高科技(激光扫描、荧光标记、计算机和微点阵等)融合的典范,有关基因隐私权是法律、道德、医保、招聘和婚姻问题专家讨论的热点,等等。

在这样的背景下,笔者总结了国内外与人体基因有关的研究成果,结合自己在国家自然科学基金和其他基金资助下的研究结果,重新组织编写了《人体基因学》这本教材。本书包括绪论、基因与基因组、基因的功能、基因与疾病、基因克隆和基因与社会等六章。

本书可用作高等院校公共选修课和大学生素质教育的教材,也可供医学、生命科学相关专业的科研人员、教师、研究生和本科生参考,还可作为大众普及基因知识的参考资料。

本书的编者们的长期从事人体基因学研究的年青学者,绝大多数有海外留学经历,能把握国内外相关研究的最新进展。第一、二、三章由郭俊明编写,第四章由季林丹编写,第五章由习阳编写,第六章由肖丙秀编写,附录由蒋孝明编写。

本书的出版,不仅得到宁波大学教务处和医学院领导的大力支持、浙江大学出版社的关心和帮助,而且受到“国家卓越医生教育培养计划”、“国家双语教学示范课程”、“浙江省高校重点建设教材”、“浙江省精品课程”和“宁波大学教材建设项目”等资助,在此表示衷心的感谢。

时间和水平有限,加之按本体系来编写人体基因学是一种新的尝试,因此,本书必有不少不妥甚至错误之处,恳请广大读者惠予斧正。

郭俊明

2013年6月于宁波大学

目 录

第一章 绪 论	1
第二章 基因与基因组	3
第一节 核酸的化学组成	3
第二节 核酸的一级结构	6
第三节 核酸的高级结构	11
第四节 核酸的理化性质	17
第五节 基因和基因组的概念与特点	20
第三章 基因的功能	24
第一节 生命的遗传物质	24
第二节 DNA 的功能	25
第三节 RNA 的功能	25
第四节 遗传信息的传递	35
第四章 基因与疾病	45
第一节 基因与遗传性疾病概述	46
第二节 单基因遗传病	48
第三节 多基因遗传病	53
第四节 染色体病	61
第五节 基因与肿瘤	65
第六节 线粒体病	76
第五章 基因克隆	78
第一节 基因克隆技术	78
第二节 基因相关技术	85
第三节 人类基因组计划	102
第四节 生物信息学	109

第六章 基因与社会	124
第一节 基因技术与医学	124
第二节 基因伦理学	131
第三节 基因技术对社会的影响	135
第四节 基因研究与诺贝尔奖	137
附 录	143
附录 1 核苷三磷酸的物理常数	143
附录 2 几种生物单倍体 DNA 的含量	143
附录 3 希腊字母表	144
附录 4 罗马数字表	145
附录 5 DNA 和 RNA 的符号	145
附录 6 密码子在人类中的使用频率	146
附录 7 科学单位	147
附录 8 核酸检测方法的灵敏度	147
附录 9 常用核酸换算数据	147
附录 10 核苷酸序列资料库	148
附录 11 常用基因相关网站	148
附录 12 基因及基因组研究大事记	149
参考文献	159

第一章 绪 论

基因的研究与遗传物质的发现有密切关系。早在 1869 年瑞士外科医生 F. Miescher 就首次从脓细胞核中分离出一种不同于蛋白质的酸性物质,并称之为“核素(nuclein)”;后来,他又从鲑鱼精子中分离出类似的物质,并指出它是由一种碱性蛋白质与一种酸性物质组成的复合物(此酸性物质即是现在所称的“核酸”)。又经过十多年的努力,“核素”中的四种基本碱基——腺嘌呤(adenine, A)、鸟嘌呤(guanine, G)、胞嘧啶(cytosine, C)和胸腺嘧啶(thymine, T)先后被发现。1889 年, R. Altamann 将“核素”命名为核酸(nucleic acid)。1909 年, 美籍俄裔生物化学家 P. A. T. Levene 发现酵母的核酸中含有核糖, 随后又发现动物的核酸中含有脱氧核糖, 于是他把核酸分为两大类: 核糖核酸(ribonucleic acid, RNA)和脱氧核糖核酸(deoxyribonucleic acid, DNA)。1909 年, W. L. Johannsen 用“基因(gene)”取代“遗传因子”, 而 1910 年 T. H. Morgan 首次用实验证明了基因位于染色体上, 并于 1917 年提出“基因是生物遗传的基本单位”。1926 年, Morgan 的专著《基因论》出版, 标志着经典遗传学理论体系的建立, Morgan 被称为“分子生物学之父”, 他也是第一位获得诺贝尔生理学和医学奖的遗传学家。1941 年, 美国遗传学家 G. W. Beadle 和生物化学家 E. L. Tatum 分离出链孢霉的生化突变型, 提出了“一个基因一个酶”的假说, 说明了基因的一些基本功能。1944 年, O. T. Avery 等科学家发现致病性肺炎双球菌的 DNA 可使非致病性的肺炎双球菌转变为具有致病性的细菌, 这就是著名的细菌转化实验, 通过这一实验证实了 DNA 是遗传物质。1953 年, J. D. Watson 和 F. Crick 共同提出的 DNA 双螺旋结构模型发表在当年 4 月 25 日的 *Nature* 杂志上, 为揭示遗传信息的传递规律奠定了基础, 标志着现代分子生物学的开始。20 世纪 70 年代, 基因克隆技术的发展和兴起, 使人们有可能按照自己的愿望改造基因, 为人类服务。由核酸研究而产生的分子生物学, 特别是基因工程技术已渗透到医学、药学、遗传学、农学等各个学科, 人类对生命本质的认识也进入了一个崭新的时代。

目前, 随着科学的进步, 分子生物学已经从研究单个基因发展到对生物体整个基因组结构与功能的研究。1990 年开始实施的人类基因组计划(human genome project, HGP)是生命科学领域有史以来最庞大的全球性科学研究计划, 确定了人体 23 对染色体 DNA 中总共约 30 亿个核苷酸的排列顺序。由这些碱基对组成的人类基因组中蕴藏着生命的奥秘。经过 13 年的努力, 人类在揭示生命奥秘、认识自我的漫漫长路上又迈出了重要的一步。2003 年 4 月 15 日, 中、美、英、日、法、德等国政府首脑联名发表了《六国政府首脑关于完成人类基因组序列图的联合声明》, 标志着人类基因组计划顺利完成。

近些年来, 新一代高通量基因测序技术和大规模转录组学研究的成果层出不穷, 为我

们揭示基因的奥妙提供了新思路。目前,以“DNA-RNA-蛋白质”为中心的遗传学法则和传统基因表达调控模式等经典理论受到前所未有的挑战。人们发现,在人类基因组中约93%的序列可被转录,其中具有蛋白质编码功能的转录物不超过2%,而占转录组90%以上的是无蛋白质编码功能的非编码RNA(noncoding RNA, ncRNA)。ncRNA尽管没有蛋白质编码功能,但它们广泛参与生命现象的各个环节,包括生长、分化、发育、免疫、凋亡等。

基于基因组研究成果的基因工程药物、基因治疗、生物芯片、基因诊断等,有着极其广阔的应用前景;基于基因组知识的药物设计,尤其是基于药物基因组学的个体化药物治疗,将在大大提高药物治疗效果的同时最大限度地降低药物的毒副作用。在此基础上,功能基因组计划将确定人体内的所有基因的功能及其调控方式。这些研究成果必将进一步加深人们对人类自身的认识,促进医学及社会的进步。

(郭俊明)

第二章 基因与基因组

生物体的遗传物质是核酸。核酸 (nucleic acid) 是以核苷酸为基本构成单位的生物大分子。天然的核酸有两大类: 核糖核酸 (ribonucleic acid, RNA) 和脱氧核糖核酸 (deoxyribonucleic acid, DNA)。绝大多数生物体的遗传物质是 DNA, 有些低等生物的遗传物质是 RNA。

人体的遗传物质是 DNA。人体 DNA 携带人类的遗传信息, 存在于细胞核和线粒体中, 决定细胞和个体的基因型 (genotype)。人类 RNA 主要存在于细胞核内和细胞质中, 功能性 RNA 主要参与蛋白质的合成过程, 把 DNA 中的遗传信息表达成蛋白质中的氨基酸排列顺序; 或者在多个层面 (DNA 复制、转录、转录后、翻译、翻译后、表观遗传等) 参与基因表达的调控过程。通过基因表达 (转录和翻译两个过程), 人类的各种遗传性状 (表现型, phenotype) 体现出来。

第一节 核酸的化学组成

通过水解作用, 两类核酸 (DNA 和 RNA) 均可生成其基本的构件单位——核苷酸 (nucleotide), 而核苷酸彻底水解的产物是磷酸 (phosphate)、戊糖 (pentose) 和碱基 (base) 等三类物质 (见图 2-1)。

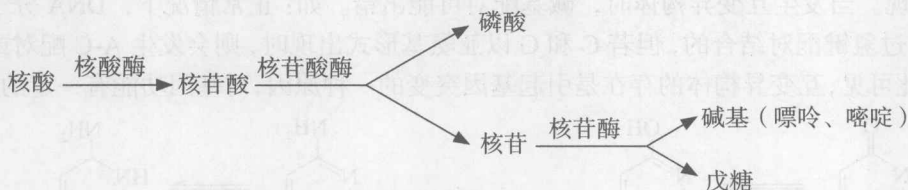


图 2-1 核酸的水解产物

一、核酸的元素组成

组成核酸的元素有碳 (C)、氢 (H)、氧 (O)、氮 (N)、磷 (P) 等, 其中磷元素的含量比较恒定, 约占 9%~10%。因此, 可以通过测定磷的含量来推算样品中核酸的含量。这种核酸定量分析的方法称为定磷法。这是一种比较实用且简便的测定核酸含量的传统方法。

二、核酸的化学组成

(一) 碱基

组成核苷酸的碱基 (base) 均为含氮杂环化合物。碱基可分嘌呤 (purine) 和嘧啶 (pyrimidine) 两大类 (见图 2-2)。嘌呤碱主要有腺嘌呤 (adenine, A) 和鸟嘌呤 (guanine, G) 等两种, 而嘧啶碱主要有胞嘧啶 (cytosine, C)、尿嘧啶 (uracil, U) 和胸腺嘧啶 (thymine, T) 等三种。DNA 和 RNA 分子中都含有腺嘌呤 (A)、鸟嘌呤 (G) 和胞嘧啶 (C); 胸腺嘧啶 (T) 一般只存在于 DNA 分子中, 而尿嘧啶 (U) 只存在于 RNA 分子中。



图 2-2 核酸分子中主要碱基的结构

从碱基的结构中可以看出, 嘌呤环和嘧啶环中都含有共轭双键。这种结构对波长在 260nm 左右的紫外光吸收较强。因此, 可以利用这一特性对碱基、核苷、核苷酸和核酸进行定性和定量分析。这种利用核酸紫外吸收性质进行比色分析的方法是一种最为常用的核酸分析方法。

上述五种碱基中的酮基和氨基, 均位于碱基环上氮原子的邻位, 这种结构有利于酮式—烯醇式和氨基—亚氨基互变异构体的形成 (见图 2-3)。生理条件下, 这些碱基主要以氨基和酮式出现。当发生互变异构体时, 碱基配对可能出错。如: 正常情况下, DNA 分子中 C 是与 G 通过氢键配对结合的, 但若 C 和 G 以亚氨基形式出现时, 则会发生 A-C 配对或 G-T 配对。由此可见, 互变异构体的存在是引起基因突变的一种原因, 对基因功能有一定的影响。

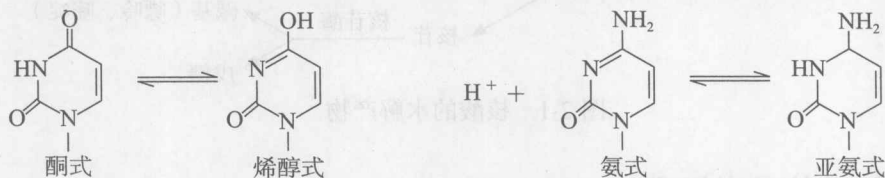


图 2-3 碱基的互变异构体

(二) 戊糖

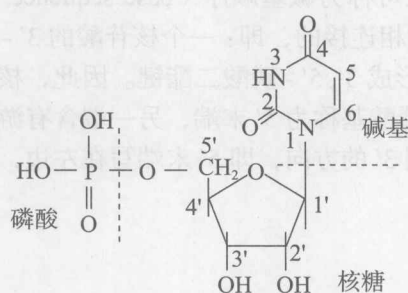
核酸中的戊糖有核糖 (ribose) 和脱氧核糖 (deoxyribose) 两种, 分别存在于 RNA 和 DNA 中。为了与碱基原子编号相区别, 通常将戊糖的碳原子编号都加上 “'”。

(三) 核苷

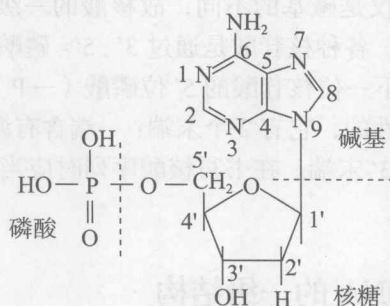
戊糖与嘧啶或嘌呤等碱基以糖苷键连接成的化合物称为核苷 (ribonucleoside)，一般是戊糖的 C-1' 与嘧啶碱的 N-1 或嘌呤碱的 N-9 相连接。

(四) 核苷酸

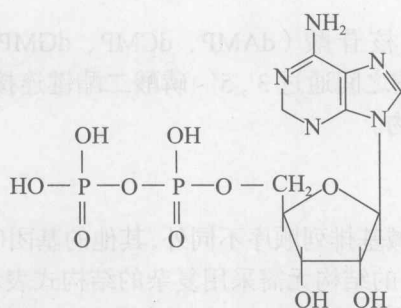
核苷中戊糖的羟基与磷酸以磷酸酯键连接而成的化合物称为核苷酸。生物体内的核苷酸大多数是核糖或脱氧核糖的 C-5' 上羟基被磷酸酯化，形成 5'-核苷酸。含有一个磷酸基团的核苷酸称为核苷一磷酸 (nucleoside monophosphate, NMP)。NMP 的磷酸基团再进一步与磷酸通过磷酸酯键连接即生成核苷二磷酸 (nucleoside diphosphate, NDP)。同样，NDP 又可以生成核苷三磷酸 (nucleoside triphosphate, NTP)。如：核糖腺苷酸就有腺苷一磷酸 (AMP)、腺苷二磷酸 (ADP) 和腺苷三磷酸 (ATP) 这三种形式。脱氧核苷酸的表示方法是在核苷酸的前面加小写字母 “d” (deoxy, 脱氧) 即可，如：脱氧腺苷一磷酸用 dAMP 表示 (见图 2-4)。



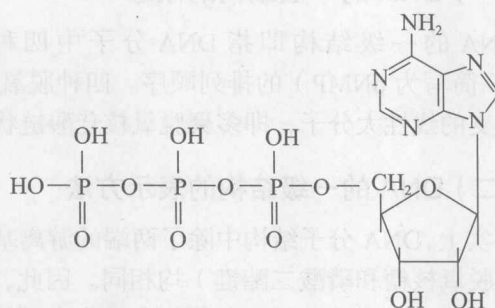
尿苷酸 (UMP)



脱氧腺苷酸 (dAMP)



腺苷二磷酸 (ADP)



腺苷三磷酸 (ATP)

图 2-4 几种核苷酸的结构

核苷酸是核酸的基本单位，核酸就是由许多单核苷酸聚合而形成的多聚核苷酸 (polynucleotide)。

为了方便书写，人们常用缩写符号来表示碱基和核苷酸。现把构成 DNA 和 RNA 分子的主要碱基和核苷酸的缩写符号总结于表 2-1 中。

表 2-1 构成 DNA 和 RNA 分子的主要碱基和核苷酸的缩写符号

DNA					RNA			
碱 基	A	T	C	G	A	U	C	G
核苷酸	dAMP	dTMP	dCMP	dGMP	AMP	UMP	CMP	GMP

第二节 核酸的一级结构

核酸的结构是其具有各种生物学功能的基础。为了更好地理解基因的功能，有必要先了解一下核酸的结构。从简单到复杂，核酸的结构可人为地划分为初级结构和高级结构。初级结构是基本的结构，又称为一级结构（primary structure）；高级结构还可再分为二级结构（secondary structure）和三级结构（tertiary structure）。

核酸的一级结构是指核酸分子中核苷酸的排列顺序。由于同类核酸分子中核苷酸之间的差异仅仅是碱基的不同，故核酸的一级结构又可称为碱基顺序（base sequence）。在核酸分子中，各种核苷酸是通过 3', 5'-磷酸二酯键相连接的，即：一个核苷酸的 3'-羟基（—OH）与下一位核苷酸的 5' 位磷酸（—P）之间形成 3', 5'-磷酸二酯键。因此，核酸分子是有方向性的，它有 2 个末端：一端含有游离的磷酸基称为 5' 末端，另一端含有游离的 3' 羟基称为 3' 末端。在书写核酸序列时应当按 5' 到 3' 的方向，即 5' 末端写在左边，3' 末端写在右边。

一、DNA 的一级结构

（一）DNA 的一级结构的概念

DNA 的一级结构即指 DNA 分子中四种脱氧核苷酸（dAMP、dCMP、dGMP 和 dTMP，简称为 dNMP）的排列顺序。四种脱氧核苷酸之间通过 3', 5'-磷酸二酯键连接成为不分支的线性大分子，即多聚脱氧核苷酸链状化合物。

（二）DNA 的一级结构的表示方法

事实上，DNA 分子结构中除了两端的游离基团和碱基排列顺序不同外，其他的基团（如磷酸、脱氧核糖和磷酸二酯键）均相同。因此，DNA 的结构无需采用复杂的结构式表示，一般采用简写方式表示，即用英文字母代表碱基或相应的核苷酸。以下的例子即为文献中最常见的表示一段 DNA 一级结构的方法：

GCGGTAATGCA；

或者

5' - GCGGTAATGCA-3'

这些字母即表示这段 DNA 序列的 5' 磷酸末端在左边，3' 羟基末端在右边；由 3', 5'-磷酸二酯键把 11 个脱氧核苷酸按该顺序连接起来。人类基因组计划最初的目的就是要把人类 DNA 中的所有排列顺序（即一级结构）测定出来。

二、RNA 的一级结构

(一) RNA 的一级结构的概念

在一级结构形式上, RNA 与 DNA 基本相似, 也是一种多聚核苷酸链状化合物。RNA 的一级结构是指 RNA 分子中四种核苷酸 (AMP、CMP、GMP 和 UMP, 简称为 NMP) 的排列顺序。

与 DNA 相比, RNA 的一级结构有如下三个方面的不同: ①组成 RNA 的戊糖是核糖而不是脱氧核糖; ②在 RNA 分子的碱基中有尿嘧啶 (U) 而没有胸腺嘧啶 (T)。③在 RNA 分子的碱基中有较多的稀有碱基。所谓稀有碱基 (rare base) 是指在核酸中除腺嘌呤 (A)、胞嘧啶 (C)、鸟嘌呤 (G)、胸腺嘧啶 (T) 和尿嘧啶 (U) 以外的含量较少的碱基。这些碱基大多是在基本嘌呤碱或嘧啶碱的不同部位进行化学修饰而形成的衍生物。如: 5,6-二氢尿嘧啶 (5,6-dihydroxyuridine, DHU), 假尿嘧啶 (pseudouridine, Ψ), 甲基化的嘌呤 (mA, mG) 和次黄嘌呤 (inosine, I) 等。

(二) RNA 的分类

根据基因表达的最终产物是否为蛋白质, 可以把 RNA 分为两大类: 编码 RNA (coding RNA) 和非编码 RNA (noncoding RNA, ncRNA)。顾名思义, 编码 RNA 是指能指导蛋白质合成的 RNA, 即信使 RNA (messenger RNA, mRNA); 而 ncRNA 则是指不编码蛋白质的 RNA (见图 2-5)。

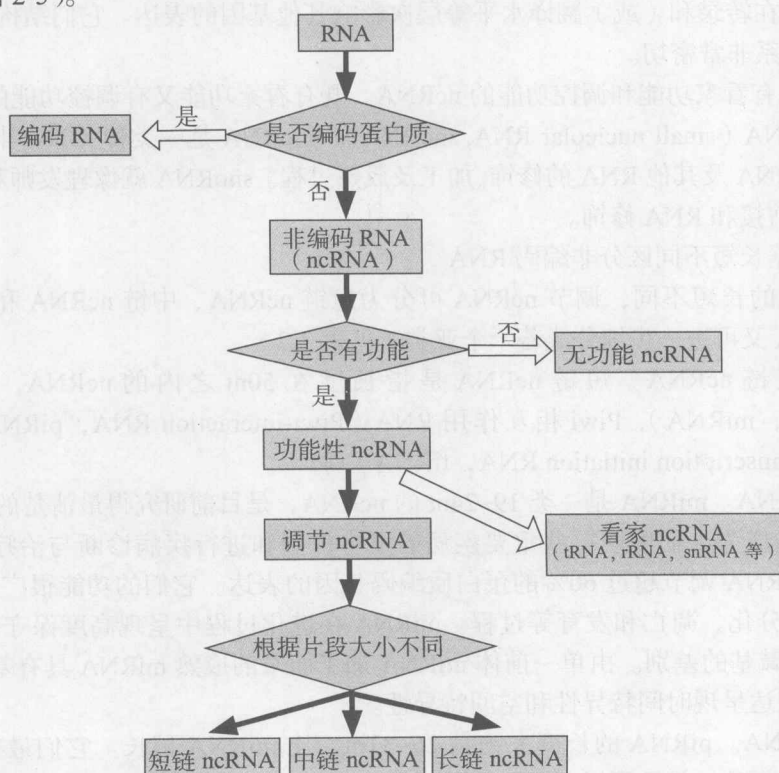


图 2-5 RNA 的分类

与 mRNA 相似, ncRNA 也是单链的 RNA。长期以来, 这些 ncRNA 及其所对应的 DNA 被认为是垃圾或“暗物质”。随着人类基因组计划的完成, 科学家们发现人类基因组中能编码蛋白质的 DNA 只占整个基因组序列的极少数, 而 ncRNA 及其所对应的 DNA 的数量远远多于编码蛋白质的 mRNA 及其所对应的 DNA 的数量。现在研究发现, ncRNA 不仅不是垃圾, 而且广泛参与生命现象的各个环节, 包括生长、分化、发育、免疫等, 甚至在肿瘤的形成中也具有重要的调控作用。

1. 根据功能不同区分 ncRNA

ncRNA 种类繁多, 分类方法有多种。可根据其是否具有功能分成无功能 ncRNA 和功能性 ncRNA。功能性 ncRNA 又可以根据功能不同分为看家 ncRNA (house-keeping noncoding RNA) 和调节 ncRNA (regulatory noncoding RNA) 两大类。

(1) 看家 ncRNA。这是组成性表达的 RNA, 它们具有细胞存活至关重要的一系列功能。它们通常表达稳定, 其表达水平受到体内环境变化的影响不明显。看家 ncRNA 包括: 转运 RNA (transfer RNA, tRNA)、核糖体 RNA (ribosomal RNA, rRNA) 和核小 RNA (small nuclear RNA, snRNA) 等。这类 ncRNA 大多数与蛋白质的合成过程有关, 如: 参与蛋白质生物合成的 ncRNA 有 tRNA 和 rRNA; 与转录后加工有关的 RNA 有 snRNA 等。现在发现有 5 种 snRNA, 其长度约为 100~215 个核苷酸 (nucleotide, nt)。snRNA 一直存在于细胞核中, 与 40 种左右的核内蛋白质共同组成 RNA 剪接体, 在 RNA 转录后加工中起重要作用。另外, 看家 ncRNA 还有端粒酶 RNA (telomerase RNA), 它与染色体末端的复制有关, 参与维持真核生物染色体的完整性。

(2) 调节 ncRNA。调节 ncRNA 在细胞分化和器官发育的特定阶段或应对外界刺激时表达。它们在转录和 (或) 翻译水平等层次影响其他基因的表达。它们结构或功能的异常与疾病的关系非常密切。

(3) 兼有看家功能和调控功能的 ncRNA。既有看家功能又有调控功能的 ncRNA 主要有核仁小 RNA (small nucleolar RNA, snoRNA)。snoRNA 是一类种类非常丰富的 ncRNA, 主要参与 rRNA 及其他 RNA 的修饰、加工及成熟过程。snoRNA 就像理发师和美容师一样, 参与 RNA 剪接和 RNA 修饰。

2. 根据长短不同区分非编码 RNA

根据链的长短不同, 调节 ncRNA 可分为短链 ncRNA、中链 ncRNA 和长链 ncRNA。这些 ncRNA 又可进一步细分成若干个亚类 (见表 2-2)。

(1) 短链 ncRNA。短链 ncRNA 是指长度在 50nt 之内的 ncRNA, 包括微 RNA (microRNA, miRNA)、Piwi 相互作用 RNA (Piwi-interaction RNA, piRNA) 和转录起始 RNA (transcription initiation RNA, tiRNA) 等。

1) miRNA。miRNA 是一类 19~24nt 的 ncRNA, 是目前研究得最清楚的一类 ncRNA。miRNA 已经成为人们阐明一些重要疾病的发生机制和进行疾病诊断与治疗的新切入点。据估计, miRNA 调节超过 60% 的蛋白质编码基因的表达。它们的功能很广泛, 包括调节细胞增殖、分化、凋亡和发育等过程。miRNA 在进化过程中呈现高度保守性, 保守片段只有 1~2 个碱基的差别。由单一前体 miRNA 加工而来的成熟 miRNA 具有基因簇集现象。miRNA 的表达呈现时间特异性和空间特异性。

2) piRNA。piRNA 的长度大约是 26~31nt, 比 miRNA 稍长。它们最初在哺乳动物的睾丸中被发现, 并且可以与 Piwi 蛋白结合形成 piRNA 复合物, 然后发挥作用。目前

表 2-2 非编码 RNA 的分类

名称	大小 (nt)	分布	人类中的数量	主要功能
短链 ncRNA				
miRNA	19~24	广泛	>1424	靶向调控 mRNA 和其他分子
piRNA	26~31	成簇, 基因内	23439	抑制转座、DNA 甲基化
tiRNA	17~18	TSS 的下游	>5 000	调控转录
中链 ncRNA				
snoRNA	60~300	基因内	>300	rRNA 修饰
PASR	22~200	蛋白质编码基因的 5' 区	>10000	不详
TSSa-RNA	20~90	TSS 的-250 至 +50 bp 之间	>10000	维持转录
PROMPT	<200	TSS 的-205 bp 至-5 kb 之间	不详	激活转录
长链 ncRNA				
lincRNA	>200	广泛	>1000	DNA 支架—染色质复合物
T-UCR	>200	广泛	>350	调控 miRNA 和 mRNA 水平
其他 lincRNA	>200	广泛	>3000	X 染色体失活、端粒调节、基因印迹等

注: 在分布档中“-”和“+”分别表示转录起始点 (transcription start site, TSS) 上游和下游碱基对的位置。lincRNA: 长链基因间非编码 RNA (long intergenic noncoding RNA); lncRNA: 长链非编码 RNA (long noncoding RNA); miRNA: 微 RNA (microRNA); piRNA: Piwi 相互作用 RNA (Piwi-interacting RNA); PASR: 启动子相关小 RNA (promoter-associated small RNA); PROMPT: 启动子上游转录本 (promoter upstream transcript); snoRNA: 核仁小 RNA (small nucleolar RNA); tiRNA: 转录起始 RNA (transcription initiation RNA); TSSa-RNA: 转录起始点相关 RNA (TSS-associated RNA); T-UCR: 转录超保守区 (transcribed ultraconserved region)

已发现的 piRNA 主要存在于基因间隔区, 而很少存在于基因区和重复序列区。piRNA 与 miRNA 的主要区别在于: 第一, 形成过程中不依赖 Dicer 酶; 第二, 通过结合 Ago 蛋白的 Piwi 亚族发挥作用。这一特点也是它被命名的依据。其他物种中的 Piwi 同源蛋白则根据其物种名称的首字母来依次命名, 如人类的 Piwi 同源蛋白为 Hiwi。piRNA 在生殖干细胞分化、胚胎发育、维持 DNA 的完整性和表观遗传学调控等方面具有十分重要的生物学作用。

(2) 中链 ncRNA。中链 ncRNA 的长度一般为 50~200nt, 主要包括 snoRNA、PASR、TSSa-RNA 和 PROMPT。在这些 ncRNA 中研究得最多的是 snoRNA。

snoRNA 是真核细胞核仁中的小分子非编码 RNA, 链长为 60~300nt。它们的主要功能是参与细胞质中 rRNA 和其他 RNA 转录后的加工过程, 如假尿苷化和 2'-甲基化等。

根据结构元件的不同, 人们常把 snoRNA 分为三大类: C/D box snoRNA、H/ACA box snoRNA 和 MRP RNA。MRP RNA 是极为特殊的 snoRNA, 在数量和功能上都异于其他两类 snoRNA, 它们参与 5.8S rRNA 的加工和线粒体 DNA 的复制。细胞中主要的 snoRNA 是 C/D box snoRNA 和 H/ACA box snoRNA (见图 2-6)。

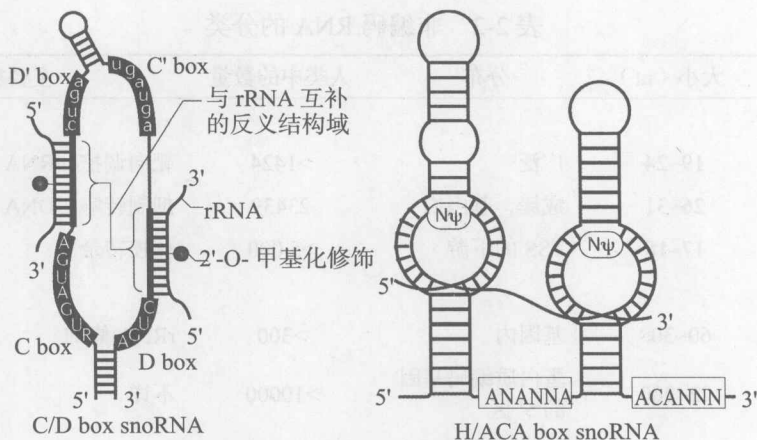


图 2-6 两类主要 snoRNA 的结构

C/D box snoRNA 包含两个短的特征性序列元件,即位于 5' 末端的 C box (RUGAUGA, R 代表嘌呤核苷酸) 和 3' 末端的 D box (CUGA)。大部分 C/D box snoRNA 分子的中部还具有类似于 C box 和 D box 的结构,分别被称为 C' box 和 D' box。C/D box snoRNA 通过碱基互补作用行使功能,即:参与 rRNA 特定位点的 2'-O-甲基化修饰。

H/ACA box snoRNA 具有保守的“发夹—铰链—发夹—尾 (hairpin-hinge-hairpin-tail)”的二级结构。H box (ANANNA, N 代表任一核苷酸) 位于单链形式的铰链区,而 ACA box 则一般位于 3' 末端上游 3 个核苷酸处。H box 和 ACA box 不仅是 snoRNA 正确行使功能的必需结构,而且与 snoRNA 的稳定性密切相关。H/ACA box snoRNA 的主要功能是参与 rRNA 的假尿嘧啶化修饰。

经 snoRNA 加工成熟的 rRNA 先在核仁中与核糖体蛋白结合,再经过复杂的进一步成熟过程和转运过程出核,最终在细胞质中形成功能成熟的核糖体。核糖体是蛋白质合成的场所,几乎控制着细胞内所有蛋白质的合成。

主要的 snoRNA 均需与特定蛋白质结合形成核仁小核糖核蛋白复合体 (small nucleolar ribonucleoprotein complexes, snoRNP) 后才能发挥作用。

(3) 长链 ncRNA。长链非编码 RNA 一般是指大于 200nt 的 ncRNA。这一类 ncRNA 在人类基因组中分布非常广泛,也是最近几年才发现并被重视的 ncRNA,其中最先发现的 lncRNA 为长链基因间非编码 RNA。

lncRNA 可分为正义 (sense)、反义 (antisense)、双向 (bidirectional)、基因内 (intronic)、基因间 (intergenic) 等五种类型 (见图 2-7)。

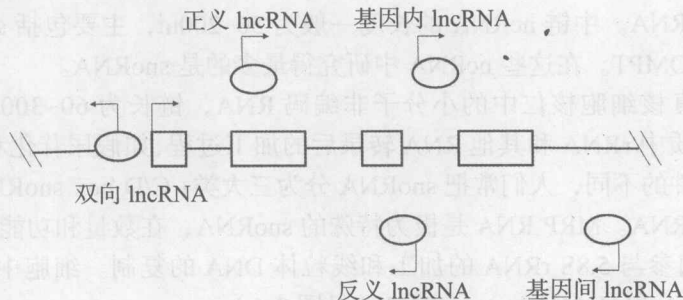


图 2-7 lncRNA 的主要类型

(三) mRNA 的一级结构

研究发现,人体 mRNA 在两个末端均具有特殊结构。

1. 5' 末端的帽子

帽子 (cap) 是指 mRNA 转录后在 5' 末端加上一个 7-甲基鸟苷,同时第一个核苷酸的 C-2' 甲基化,形成 m⁷GpppNm 的特殊结构。帽子结构具有重要功能,它具有增加 mRNA 的稳定性,在翻译过程中促进 mRNA 与小亚基结合等作用。

2. 3' 末端尾巴

尾巴是指 mRNA 3' 末端的 20~200 个左右的多聚腺苷酸 (polyA)。它和帽子结构一样也是在转录后加工形成的。polyA 的功能是增加 mRNA 的稳定性、有助于 mRNA 从核内转移到核外。

各种 mRNA 的长短有较大的差别,这也说明其表达产物——蛋白质的相对分子质量有较大差别。在所有 RNA 中 mRNA 的半寿期 (half life, 含量保留一半的时间) 最短。

(四) tRNA 的一级结构

人体 tRNA 具有以下三个特点。

(1) 相对分子质量比较小。tRNA 只由 70~90 个 nt 组成。

(2) 有较多稀有碱基。这是 tRNA 在组成上的最大特点。稀有碱基约占总碱基的 10%~20%。

(3) 3' 末端有 CCA—OH 结构。这一结构与其主要功能——携带氨基酸有密切关系。这从一个侧面说明了结构是功能的基础。

(五) rRNA 的一级结构

rRNA 是细胞内含量最多的 RNA,占 RNA 总量的 80% 以上。根据其相对分子质量的不同,人体的 rRNA 可分为 5S、5.8S、18S 和 28S 四种类型。此处,S 是指沉降系数 (sedimentation coefficient)。它是大分子物质在超速离心沉降中的一个物理学单位,可间接反映物质的相对分子质量的大小。应用超速离心法测定物质 (蛋白质和核酸) 的相对分子质量时,S 与相对分子质量呈正比关系。特别是物质的相对分子质量较大时,用 S 表示较方便。

第三节 核酸的高级结构

核酸的高级结构又可称为核酸的空间结构 (spatial structure)。核酸在一级结构的基础上通过分子之间的相互作用而形成具有空间构象的生物大分子,体现其作为遗传物质的各种功能。

一、DNA 的二级结构——双螺旋结构模型

Watson 和 Crick 在前人研究的基础上,于 1953 年 4 月 25 日在著名的 *Nature* 杂志上发表了题为《核酸的分子结构》的论文,提出了著名的 DNA 分子的双螺旋结构模型 (double