DANGDAIKEXUECONGSHU

人类基因组

赵寿元 编著

当代科学丛书

RENLEIJIYINZU

上海科技教育出版社

人类基因组

赵寿元 编著

上海科技教育出版社

(沪)新登字116号

人类基因组

赵寿元 编著

上海科技教育出版社出版发行

(上海冠生园路393号)

各地新华书店经销 商务印书馆上海印刷厂印刷

开本 787×1092 1/32 印张 4.125 字数 94,000

1993第4次印1刷 1993年4月第1次印刷

印数 1—2,000

ISBN 7-5428-0710-2

G·667

定价：2.35元

导　言

自古以来，人类一直在企求认识自身，力图穷究生、老、病、死、感觉、思维、行为和意识的底蕴。最初是从人体解剖开始，随着生命科学特别是医学和分子生物学的建立和发展，人们已找到了揭示这些奥秘的途径。

自从15世纪以来，解剖人体结构时所作的描述和图解，不断对人类认识自身作出贡献。16世纪时比利时医生、意大利巴丢阿大学教授维萨里(Andreas Vesalius, 1514～1564) 从事人的尸体解剖，详尽地描述了人体各器官的结构，积累了翔实的资料，纠正了当时的一些错误记载，积极地推动了人体解剖学的发展，奠定了近代人体解剖学的基础，促进了医学科学的研究。与此同时，医学和外科学的进展，又常常使人们认识到过去认为是纯学术性的解剖学细节具有重大的实用价值。应用科学和基础研究总是相辅相成，互相促进。麻醉学发展之前，不可能作长时间的外科手术，因而对肺的详细的局部解剖学知识是不充分的。胸腔外科的发展则促进了对肺叶及其血管构造的研究，从而为肺的局部病变的外科治疗提供了基础。与此相似，当心脏手术和外科移植成为可能时，对心脏发育及其解剖学的研究成果就立刻得到了应用。成人耳朵发育和解剖学的描述，对于现代内耳外科改善耳聋则是必不可少的。

随着科学技术的日益昌明，人体解剖学的研究也逐步从大体解剖发展到局部解剖，从器官深入到组织、细胞、细胞器，一直到蛋白质、核酸等大分子的水平，即从宏观的解剖深入到

微观的解剖。在比较了健康人和病人的解剖学资料后，在人体的不同结构层次上认识到造成不同生理功能的原因，进而确定了各种疾病的诊断指标，提出有效的预防和治疗的措施，极大地造福于人类。

心血管系统疾病和癌症是危害人类最严重的两类常见多发病。对患者与正常人的比较解剖已深入到代谢的差别。例如主要表现为心肌梗塞和脑栓塞的动脉粥样硬化症，脂质和脂蛋白代谢异常是发病的关键。由胆固醇与细胞增生及结缔组织相结合形成粥样硬化的损伤部位，引起动脉床的关键区域逐渐变窄以致最后堵塞。血浆脂蛋白是球形微粒，包括一个由疏水性胆固醇酯和甘油三酯组成的内核，四周表皮层是亲水性磷脂，它含有胆固醇和脱辅基脂蛋白。血浆中主要类型的脂蛋白包括乳糜粒、极低密度脂蛋白(VLDL)、低密度脂蛋白(LDL)、中密度脂蛋白(IDL)及高密度脂蛋白(HDL)。每一类脂蛋白分别进入不同的分解代谢途径，有的可能形成动脉粥样硬化的微粒。当LDL浓度高而HDL浓度低时，表明脂蛋白的产生是遗传和环境的调控的结果，在动脉管壁细胞内造成脂质聚集。脱辅基脂蛋白($A_{po}L$)是脂蛋白聚集、分泌、分解、消除的主要决定因子。上述有关动脉粥样硬化症的发病过程、物质代谢过程等的描述，看来既复杂又繁琐，其实还没有弄清楚其真正的发病机制。因为，这涉及编码产生各种脂蛋白的基因表达和调控的机制、参与细胞内各种脂蛋白代谢路线的有关的酶及其编码基因，所有这一切都还有待进一步认识。至于癌症的致病原因可以有多种多样，但所有癌症有一点是共同的，即癌变细胞内的DNA都发生了改变，或是表现为染色体畸变，或是检出癌基因突变等，最终导致细胞的失控生长分裂。说到遗传病那更是如此，现在已知的将近4000种遗

传病，都是遗传物质发生改变而导致的病变。有关这方面的材料，将在以后各章节中作详细介绍。

总之，人体解剖正逐步深入到基因组解剖；随着今后医学的发展，一些今天看来是疑难杂症或者甚至是不治之症的疾病的发病原因，可望随着分子生物学特别是分子遗传学的发展而得到阐明，进而在此认识基础上提出有效的预防和治疗的措施；使癌症、艾滋病、心血管疾病等像上个世纪肆虐人间的天花、霍乱、鼠疫、伤寒等一样地处于人类的控制之下，逐步地使之灭迹。在这一过程中，人基因组的研究成果将起十分积极的作用和做出重大的贡献。

现代遗传学的奠基人是奥地利神父格里戈尔·约翰·孟德尔(Gregor Johann Mendel, 1822～1884)。他从1857年开始从事豌豆杂交试验，整整花了8年时间，反复研究了豌豆花的颜色、种子的形状、子叶的颜色、成熟豆荚是否分节、未成熟豆荚的颜色、花在主茎轴上的位置、植株的高矮等7对相对性状的遗传规律，用统计方法分析了实验数据，发现了生物性状的分离规律，指出生物的每一个性状是由一个遗传因子负责传递，但遗传下来的并不是具体的、直观的性状，而是遗传因子。这一被后人称为孟德尔定律的重大科学发现，在遭淹没35年后，于1900年才被3位科学家的实验所证实而重新被发现。在此以后，遗传学建立在科学基础上而成为一门真正的学科，并以前所未有的速度迅猛地成长和发展。这是因为遗传学是生命科学领域中各个学科的共同的基础理论，同时遗传学本身又在实际应用领域中有很大价值的缘故。

美国遗传学家托马斯·亨特·摩尔根 (Thomas Hunt Morgan, 1866～1945)是又一位对遗传学的发展作出杰出贡献的科学家。他用果蝇作为实验材料，证实了孟德尔时代所说

的遗传因子也就是基因，基因负责传递与其相应的性状，基因位于细胞核内的染色体上，染色体是在亲代和子代之间传递基因的物质载体。他将遗传因子的假设证实为确凿无疑的科学事实。本世纪40年代的研究结果令人信服地确证了基因确如摩尔根所预言的是一种化学实体，即一种被称为脱氧核糖核酸的化学大分子。1953年，华生(James D. Watson)和克里克(Francis H. C. Crick)提出了 DNA 双螺旋结构模型，为在分子水平上阐明遗传物质——基因的复制、分配和决定蛋白质分子结构的机制，奠定了坚实的理论基础，从而揭开了分子遗传学的历史新篇章。

分子遗传学研究取得了巨大成果，使人类对自身的认识进入了一个全新的境地。人体从宏观的解剖可分成呼吸、循环、消化、神经、肌肉、骨骼、内分泌、泌尿、生殖和感觉等器官或生理系统，每一种器官都有特别的生理功能。每一种器官又分别由许多种特定类型的组织和细胞所构成。一个成年人体估计有一百万亿(10^{14})个细胞，至少可分出100多种类型；细胞里又包含着许多种分别具有不同功能的更为精细的结构。数量如此巨大和功能如此复杂但又如此精确协调地发挥作用的细胞器、细胞、组织和器官，全都来自一个细胞，即来自包含父母双方遗传信息的受精卵，由它分裂和分化而成。换句话说，个体的分化、发育和成长，正是细胞所携带的遗传信息依照一个极其精确的时空程序逐步表达实现的综合结果。当遗传信息本身或表达程序出现差错时，就会导致细胞的结构和功能发生异常，导致个体性状的改变、发生疾病乃至造成死亡。因此，人体解剖学从大体解剖进入局部解剖、显微解剖后，如今正面临着进入微观的遗传信息的解剖分析的挑战，即由此开创了人体基因组解剖的新时代。

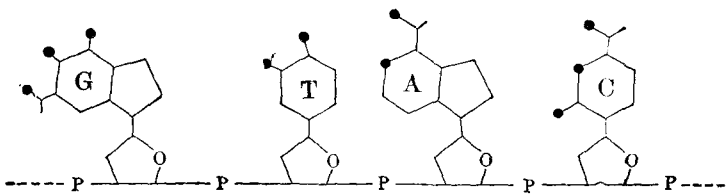
人体双倍体细胞里有22对常染色体和1对性染色体，共46条染色体：23条是来自父亲，23条是来自母亲。人体的单倍体细胞即生殖细胞——精子或卵子里只含单倍体基因组，有22条常染色体和1条性染色体(X染色体或Y染色体)。单倍体基因组由30亿个(3×10^9)核苷酸对所组成，每一条染色体就是一个双链DNA分子。解剖人体基因组，说到底就是要分析测定这30亿个核苷酸依次排列的顺序。美国开展了自1991年开始的为期15年、计划拨款30亿美元的"人类基因组作图和测序"的研究计划。这是一个跨世纪的宏大的科学研究项目，国际上已有许多国家都确定了相应的研究内容，组织了研究队伍，成立了专门的研究机构，拨出专款开始了实验工作。我国也有了一些初步的设想与安排。国际生物学界的一些知名学者认为，人类基因组的解剖分析是十分重要的，其意义堪与制造原子弹的"曼哈顿"(Manhatto)计划、阿波罗(Appolo)飞船载人登月的壮举相媲美；可是相比之下，所耗费用也许只及前者的十分之一，而其影响及对人类的利益将更为重大和深远。

赵寿元，1931年5月出生于苏州市。1964年复旦大学遗传研究所研究生毕业。80年代中，两度在美国耶鲁大学访问工作3年。现为教授、博士生导师，任复旦大学遗传学研究所所长，中国遗传学会副理事长，国务院学位委员会学科评议组成员国家自然科学基金委员会委员，国家教委科技委委员，国家教委理科教学指导委员会委员等。已在国内外学术刊物上发表论文[illegible]多篇，出版专著和译著14部，曾获国家科委、部、市科技进步奖4次。

目　录

一、人基因组的组成

包括人类在内的所有生物的基因组都是由DNA分子，即一种极长的双链结构的化学多聚体所组成。每条DNA链由四种称为脱氧核糖核苷酸的结构单元所组成。脱氧核糖核苷酸端端相接连成长链，用A、G、C和T来代表这四种核苷酸的碱基，它们分别是腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶。具体地说，碱基与脱氧核糖以及磷酸构成了核苷酸，这些核苷酸通过磷酸与糖连接在一起，连续重复的糖-磷酸-糖-磷酸的长链好像是DNA分子的脊梁骨：



碱基的角形环代表碳和氮原子联系的环，黑点则代表嘌呤和嘧啶碱基之间形成氢键的部位。DNA中的每一个碱基都附在一个糖分子和一个磷酸分子上。糖是由一个氧和一些碳原子组成的环，碱基直接附在糖分子上。碱基-糖-磷酸组成了核苷酸。DNA链长得出奇，一个病毒的DNA要由20万个核苷酸组成，一个细菌的DNA则多达200万个核苷酸。如果把人的一个细胞里所有的DNA连成一条直线，大约长91厘米；

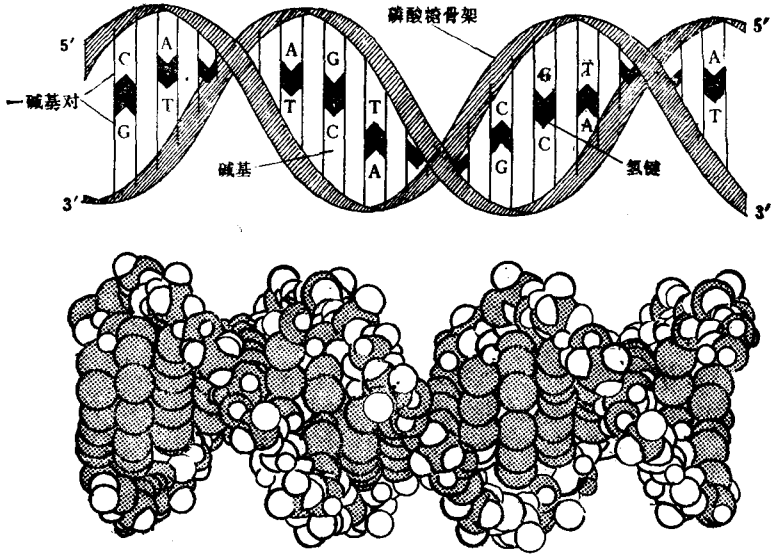


图1　两种表示DNA双螺旋的方法

图示每条染色体中DNA分子的极小一部分。下图为主体分子模型。人的基因组所含的DNA量较所示的大2亿倍。DNA双螺旋的两条链走向相反，彼此以互补核苷酸对间专一配合的方式配对(取自 Alberts等,1983)

把人体所有细胞里的DNA都连成一线，就可从地球到太阳来回好几趟！

在细胞核里，DNA 分子与一些蛋白质复合构成了染色体。染色体因所包含的核苷酸数目的多少以及核苷酸排列次序的不同而彼此有差别；不止是形态上可以区分，而且功能也各异。人的最小一条和最大一条染色体所包含的核苷酸数目大约相差5倍，分别是 5000 万个和 25000 万个核苷酸对。

人是二倍体生物，含有两套遗传信息，分别来自母亲和父亲。23对染色体，除性染色体分X和Y染色体外，其余的22对常染色体都是成对出现的，成对的染色体几乎是彼此相同的，称为同源染色体。因此，人基因组分析只需分析单套染色体

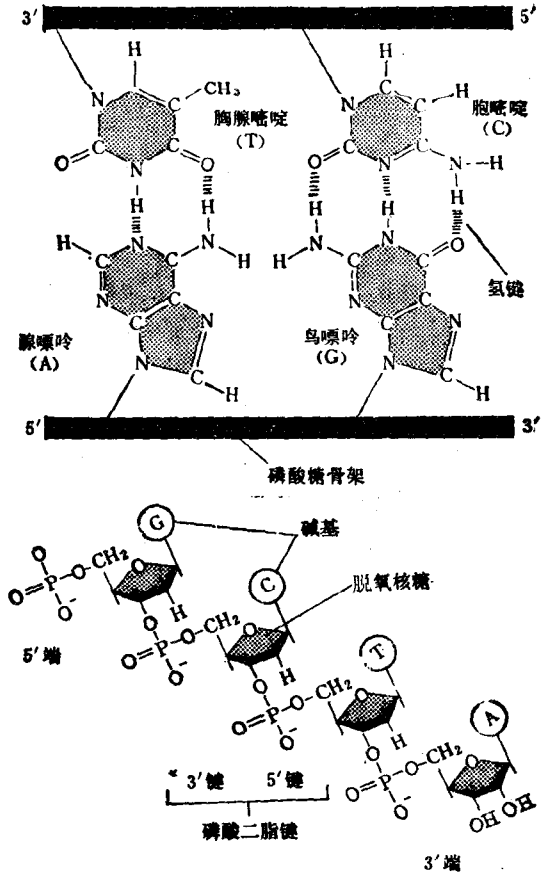


图2 构成DNA分子的核苷酸

上图：G和C、A、T 碱基之间专一的氢键相互作用产生互补核苷酸对(即G总是与C、A 总是与T结合)。一个单倍体的人基因组含有30亿对核苷酸。下图：一条DNA 链的化学结构。每一条链是由图示的 4 种核苷酸所组成的极长的链。每一核苷酸包括上图中所示4个碱基中的一个脱氧核糖磷酸残基

就够了，即只需分析30亿个核苷酸排列的次序，就可获得完整

的人基因组的全部信息。

如上所述，DNA 是双股螺旋的长链分子，一条链上的核苷酸与另一条链上的核苷酸通过碱基间的氢键而配对，即形成核苷酸对，或称为碱基对，通常用bp(碱基对，base pair)来表示。碱基间的配对也有一定的规律，腺嘌呤(A)与胸腺嘧啶(T)配对，彼此间形成两个氢键；鸟嘌呤(G)与胞嘧啶(C)配对，彼此间形成三个氢键。通过碱基之间配对而形成双螺旋的两条 DNA单链，彼此是互补的。测定了一条链上的核苷酸，就可确定与其互补的另一条链上的是那一种核苷酸。因此，原则上可根据 DNA 分子一条链上的核苷酸序列来确定另一条链上互补的核苷酸序列。但是，为了保证测序所得结果的准确性，有必要分别测定两条互补链的核苷酸序列，一条链上的序列可用来核对另一条链的测序结果。这样，完成单倍体人类基因组的30亿个核苷酸序列，实际上仍需测定60亿个核苷酸的序列。

四种核苷酸不同的排列组合，具有不同的生物学功能，根据已知的实验资料，人基因组的核苷酸序列构成了不同的功能单位。一定的序列决定一定的功能。结构决定功能的关系，在基因组的组成分析中表现得十分明显而具体。这也反映出基因组分析的复杂性。

1. 结 构 基 因

结构基因是编码决定蛋白质分子结构的基因。根据蛋白质分子的平均大小可推算出基因的平均大小为1000bp，即 1 Kb(千碱基对)。据估计，人类基因组中的编码序列大约有1亿bp，即约10万个基因，这只占整个基因组的3%左右。大量的 DNA 位于基因之间，以及结构基因之中——内含子。非编码

的核苷酸序列对基因活动起调控作用，或在DNA组装成染色体及其复制中起重要作用。虽然大多数非编码序列的功能还不清楚，但据猜测其中有些可能是无功能的。

1-1 ' 外显子和内含子

结构基因规定特定的蛋白质的合成。基因是否进入活动状态是由作用于基因的调控区的信号所决定。基因活动时，整个基因的核苷酸序列被转录成信使RNA（mRNA）的前体——核不均一RNA（hnRNA），然后很快在称为RNA剪接（RNA Splicing）的过程中，将一些序列剪除，并将剩下的序列拼接起来成为信使RNA。在RNA剪接过程中被除去的序列，是基因中的内含子序列；在剪接后剩下的序列并出现在mRNA中的，则是外显子序列。按照遗传密码的规则，mRNA的每三个核苷酸组成一个密码子，规定蛋白质分子的一个基本单元——氨基酸。从mRNA到蛋白质合成的过程称为翻译。在一个基因中，内含子的总长度往往超过外显子的总长度。结构基因中的内含子不编码蛋白质，它的生物学功能以

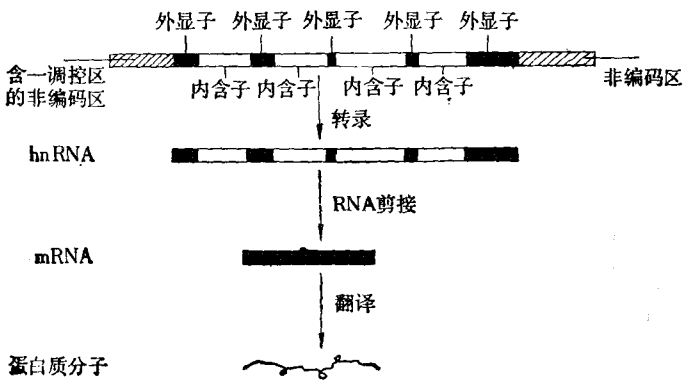


图3 人基因表达过程的示意图

表1 人的一些基因的大小

基 因	基因大小(kb)	mRNA大小(kb)	内含子数目
α球蛋白	0.8	0.5	2
β球蛋白	1.5	0.6	2
胰岛素	1.7	0.4	2
甲状旁腺素	4.2	1.0	2
蛋白激酶C	11	1.4	7
Ⅰ型胶原蛋白			
Pro-α-1(Ⅰ)	18	5	50
Pro-α-2(Ⅱ)	38	5	50
卵清蛋白	25	2.1	14
腺苷脱氨酶	32	1.5	11
凝血因子Ⅸ	34	2.8	7
过氧化氢酶	34	1.6	12
低密度脂蛋白受体	45	5.5	17
苯丙氨酸羟化酶	90	2.4	12
凝血因子Ⅷ	186	9	25
甲状腺球蛋白	300	8.7	36
杜兴氏肌营养不良	2800	～18	～50

及它在进化上的起源都还不太清楚，但有一些证据表明内含子与基因活性的调控可能有关。

随着分子遗传学研究的深入发展，已发现结构基因中的外显子序列也不一定全都被转录和翻译成蛋白质产物，有的翻译成信号肽，但在最终的蛋白质分子形成前被切除，有的则根本不被翻译成多肽。例如，人体产生的肿瘤坏死因子α(TNF-α)基因有4个外显子和3个内含子，初级转录产物 hnRNA 长2762个核苷酸，5′端和3′端各有一个非翻译区，分别长180个核