面向21世纪精品课程教材
全国高等医药教育规划教材

临床实用分子病理学

蒋仲荪 龚永生 王博浩 编著



ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

需要全本请访问 www.ertongbook.com

面向21世纪精品课程教材
全国高等医药教育规划教材

临床实用分子病理学

蒋仲荪　龚永生　王博浩　编著

ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

图书在版编目(CIP)数据

临床实用分子病理学/蒋仲苏,龚永生,王博浩编著. —杭州:浙江大学出版社,2011.5
ISBN 978-7-308-08563-2

Ⅰ.①临… Ⅱ.①蒋…②龚…③王… Ⅲ.①分子生物学:病理学 Ⅳ.①R36

中国版本图书馆 CIP 数据核字(2011)第 059821 号

临床实用分子病理学
蒋仲苏　龚永生　王博浩　编著

责任编辑　阮海潮(ruanhc@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　杭州大漠照排印刷有限公司
印　　刷　富阳市育才印刷有限公司
开　　本　710mm×1000mm　1/16
印　　张　8.5
字　　数　185 千
版 印 次　2011 年 5 月第 1 版　2011 年 5 月第 1 次印刷
书　　号　ISBN 978-7-308-08563-2
定　　价　26.00 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

前　言

医学是受益于分子生物学飞速发展的主要领域之一。现代生物医学的成就已经充分说明，人类的一切疾病发生，从根本上说，无不与自身遗传物质的改变有关，即由自身遗传物质的内在缺陷所致或是和外部环境与人类自身遗传物质的相互作用所致的改变有关。简言之，人类生理、病理的核心是基因问题。

近年来，随着人类基因组计划的完成和基因沉默机制及 RNA 干扰现象的发现，分子生物学已进入后基因时代。前者全面提供了人类 3×10^9 碱基对组成的基因组 DNA 的序列结构；后者，为人类基因组工作语言——“遗传语言”的破译提供了新的基因调控研究途径。后基因时代的主要任务是基因 DNA 序列的识别和其正常功能，基因变异与人类疾病研究。需生命科学工作者与临床工作者通力合作，才能完成。对临床工作者而言，主要是应用分子生物学基本理论、基本概念和某些实验技术，评估基因与疾病发生的风险，进行致病基因的识别与鉴定，寻找基因治疗的新途径。当前，许多感染性疾病和复杂性疾病，如老年性痴呆、糖尿病、原发性高血压及其并发症、恶性肿瘤等，均已进入分子生物学研究时代，而我们对分子生物学知识却知之不多，平时又忙于日常的诊疗工作，对于分子生物学的研究进展关注较少，这样势必影响临床工作者研究课题的深度与广度，为此，在医学继续教育中开设临床应用分子病理学课程就显得非常有必要。本书的编写不同于经典的分子生物学理论教材，在介绍分子生物学两个重要进展的基础上，重点叙述某些疾病发病的分子机制及其诊断治疗的新进展，使读者体会到如何运用分子生物学理论和概念去思考、分析他们所要解决的医学问题，并设计出解决问题的研究方案。此种分子生物学基础理论与临床研究相结合的编写方式，是我们的初步探索和尝试，经我们在浙江省知识更新班四期教学效果来看，能在有限的教学时数条件下，有助于临床医师应用分子生物学知识去思考、分析、解决医学领域内较具体的问题。

由于编者的学识和经验有限，又是以临床应用为中心的编写与教学方式，书中的错误和疏漏在所难免，敬请各位同仁和读者批评指正，并提出宝贵意见以便进一步改进。

蒋仲荪　龚永生　王博浩

2011 年 4 月于温州医学院

目　　录

第一章　人类 DNA 序列及其变异的分析方法 …… 1

第一节　基因组 DNA 序列(全基因 DNA 碱基序列) …… 1

一、基因组概念 …… 1

二、核酸组成与结构 …… 2

三、人类基因组图谱 …… 3

四、基因组 DNA 序列提供哪些信息,有助于我们对分子生物学的研究 …… 7

第二节　人类基因组 DNA 序列的变异 …… 8

一、人类基因组 DNA 序列分类 …… 9

二、人类基因组 DNA 序列的变异 …… 9

第三节　DNA 多态分析在医学中的应用——致病基因识别与鉴定 …… 12

一、基因突变检测 …… 12

二、基因遗传连锁分析 …… 14

三、疾病的关联分析 …… 15

四、基因表达谱——mRNA 表达谱(mRNA 差异显示法) …… 17

五、复杂疾病的基因定位 …… 21

六、药物基因组学 …… 23

第二章　小分子 RNA 在基因沉默中的意义 …… 27

第一节　非编码区信息结构分析——重要组分微小分子 RNA(miRNA)和 RNA 干扰(siRNA)的发现 …… 27

第二节　内源性 miRNA …… 28

一、miRNA 形成 …… 29

需要全本请访问 www.ertongbook.com

二、miRNA 抑制基因表达——基因沉默机制 …… 29
三、miRNA 具有组织特异性或发育阶段特异性表达特征 …… 30
第三节 外源性 siRNA(小分子干扰 RNA) …… 30
一、siRNA(小分子干扰 RNA)形成 …… 30
二、siRNA 沉默目标基因的机制(转录后基因沉默通路 PTGS)
…… 31
三、研究 siRNA 的意义 …… 32

第三章 阿尔茨海默病(AD)的分子病理学 …… 35

第一节 阿尔茨海默病有关基因 …… 35
一、与 AD 有关的淀粉样蛋白聚集和 APP 基因改变 …… 35
二、AD 的神经元纤维缠结及其分子生物学改变 …… 37
三、载脂蛋白 E 的 ε4 等位基因与 AD …… 38
第二节 阿尔茨海默病的病理生理 …… 38
一、Aβ 聚集的神经毒机制 …… 39
二、神经元纤维缠结病理生理 …… 41
第三节 阿尔茨海默病分子机制研究进展对治疗学的启示 …… 45
一、对 AD 传统的治疗药物—胆碱能药物疗效评估 …… 45
二、阻止蛋白质聚集物形成或加速去除聚集物 …… 46
三、神经营养因子(NTF)应用——神经生长因子(NGF) …… 47
第四节 miRNA(微小分子 RNA)在 AD 发病学中意义的研究进展 …… 48
一、研究 miRNA 在 AD 发病机制中意义的背景 …… 48
二、调控细胞凋亡的 miRNA …… 50
三、调控 AD 的 miRNA …… 51

第四章 原发性高血压病的分子病理学 …… 53

第一节 高血压相关基因的复杂性及其研究策略 …… 53
第二节 原发性高血压的相关基因 …… 58
一、候选基因法 …… 58

二、全基因组扫描法 …… 61
第三节　高血压基因治疗的实验研究 …… 64
一、高血压基因治疗的基本途径 …… 64
二、高血压基因治疗的意义及存在的问题 …… 71

第五章　糖尿病分子病理学 …… 73

第一节　胰岛素代谢信号通路分子机制 …… 73
一、胰岛素生理功能 …… 74
二、胰岛素代谢通路分子机制 …… 74
第二节　1 型糖尿病(1-DM)分子机制 …… 77
一、1 型糖尿病作为自身免疫疾患的实验和临床依据 …… 77
二、1 型糖尿病病因学 …… 79
三、自身免疫对胰岛 β 细胞损伤的分子机制 …… 79
第三节　2 型糖尿病(2-DM)分子机制 …… 81
一、胰岛素抵抗 …… 81
二、高血糖症及蛋白非酶糖化在糖尿病慢性并发症中的意义 …… 82
第四节　miRNA 与胰岛 β 细胞功能 …… 85
一、开展 miRNA 与糖尿病胰岛 β 细胞功能关系研究背景 …… 85
二、胰岛素分泌调节机制 …… 86
三、miRNA 对胰岛素分泌调节作用 …… 86

第六章　恶性肿瘤的分子病理学 …… 89

第一节　原癌基因激活与细胞癌变 …… 90
一、癌基因的概念(原癌基因) …… 90
二、原癌基因 *c-onc* 及其与病毒癌基因 *v-onc* 的关系 …… 90
三、原癌基因被异常激活的分子机制与肿瘤的发生 …… 91
四、癌基因在肿瘤发生中的作用 …… 94
第二节　抑癌基因失活与细胞癌变 …… 96
一、抑癌基因致癌特点及其杂合性丢失(LDH)检测方法 …… 96

二、几种抑癌基因失活的致癌机制 …… 97
第三节 肿瘤逃避免疫攻击的分子机制 …… 103
一、肿瘤细胞膜分子表达异常 …… 103
二、肿瘤细胞分泌免疫抑制性细胞因子 …… 106
三、与肿瘤免疫有关的细胞因子缺乏 …… 107
第四节 miRNA 在恶性肿瘤发病学中的意义 …… 107
一、恶性肿瘤 miRNA 表达谱 …… 108
二、miRNA 对肿瘤细胞生物程序的调控 …… 110
三、miRNA 检测在肿瘤诊断与治疗中的意义 …… 113
第五节 癌症基因治疗策略研究进展 …… 115
一、免疫疗法 …… 116
二、原癌基因反义核酸和小分子 RNA 的肿瘤基因治疗 …… 119
三、抑癌基因的肿瘤基因治疗 …… 119
四、导入药敏自杀基因(病毒介导的酶解药物前体治疗) …… 121

主要参考文献 …… 125

第一章

人类DNA序列及其变异的分析方法

人类基因组计划是20世纪80年代在全球范围内广泛参与和合作的一项研究计划，目标是全面、透彻地了解人类基因组的正常结构、功能及其变异与人类疾病，核心内容是绘制人类遗传图、物理图和序列图，其中最核心、最本质的就是序列图（基因组DNA）的绘制。目前，基因组DNA测序已基本完成，那么人类基因组的DNA序列及其变异，对临床工作者有什么意义呢？基因组序列变异又有哪些分析方法有助于对疾病分子机制的认识，进而为诊治疾病拓展新的研究思路与途径？此为我们所关注的几个问题。

第一节　基因组DNA序列（全基因DNA碱基序列）

一、基因组概念

基因是指细胞中所有遗传信息的总和，即包含全部遗传信息的核酸量。全部基因组合在一起称基因组。人类基因组分为复杂的核基因组和较简单的线粒体基因组。

线粒体基因组是全长约16569bp环状DNA，其核酸序列早在1981年全部测定完毕。人类绝大多数基因位于核基因组，占所有基因的99.98%，故一般所说的基因组是指核基因组，它位于核内染色体，包含于一个全长DNA分子（约3×10^9碱基对）中。在全长DNA分子中，已知编码序列仅占全部核基因组的3%，即除具有编码功能的外显子外，存在大量非编码序列，以及大量的功能尚不十分清楚的DNA序列（内含子序列、基因与基因之间的间隔序列、调控序列等）。

基因组 DNA 系高分子化合物，基本组成单位是单核苷酸，因而有必要简单复习一下核酸的一些基本知识。

二、核酸组成与结构

（一）单核苷酸组成

核苷酸是由磷酸基团、戊糖、碱基通过一定结合方式而构成。

戊糖有核糖和脱氧核糖两种，据此核苷酸分为核糖核酸（RNA）和脱氧核糖核酸（DNA）。构成核酸的碱基主要有五种：腺嘌呤（A）、鸟嘌呤（G）、胞嘧啶（C）、胸腺嘧啶（T）、尿嘧啶（V）。其中胸腺嘧啶只存在于脱氧核糖核酸中；而尿嘧啶只出现在核糖核酸中。

（二）核酸的结构

1. 核酸一级结构

核酸由成百上千乃至上百万个结构单位——核苷酸通过四种核苷酸的 3′，5′羟基之间形成磷酸二酯键，彼此连接起来的线性顺序，称之为 DNA 一级结构，即 DNA 分子中脱氧核糖核酸排列顺序。由于在 DNA 分子中脱氧核糖和磷酸组成相同，而碱基不同，故碱基排列顺序就可称为 DNA 一级结构。

一级结构不仅蕴藏着极为丰富的转换蛋白质的遗传信息（结构基因），而且，一些特定序列及其大量信息决定着 DNA 空间结构及基因间的相互作用和调控功能。在一定意义上可以说，生命过程都是根据储存在 DNA 内的程序进行的。已完成的人类基因组测序就是对基因组 3×10^9bp 的测序，再给予划分、标记，得到人类基因组 DNA 序列碱基的排列全貌。

2. 二级结构——双螺旋结构模型

DNA 由两条反向平行的多核苷酸链缠绕同一中心轴，构成双螺旋结构。其特点是磷酸基团与戊糖在链的外侧，两条链上碱基以 A═T、C≡G 相连，构成一平面，碱基平面垂直于螺旋中心轴，位于内侧。维持 DNA 双螺旋结构的作用力是一条链上碱基与另一条链上碱基严格依据碱基互补配对原则，形成氢键、碱基堆砌力以及离子键。所谓碱基互补配对原则即 A—T、G—C 配对。其中 A—T 之间形成两个氢键，G—C 之间形成三个氢键，因此 G 与 C 之间配对更稳定。

氢键是一低能键，其特性是易断、易建，从而构成 DNA 变性、复性的特性，此系分子杂交检测的重要理论基础之一。

3. 三级结构（超螺旋结构）

DNA 的三级结构是指双螺旋结构基础上的卷曲。其普遍形式是环形 DNA 双

螺旋链进一步盘绕形成的超螺旋。

超螺旋无疑有重要的生物学意义：① 使 DNA 体积变得更小，将人类细胞核中 46 条染色体长达 1.7m 的 DNA 压缩成 200nm 左右；② DNA 的线性序列携带有遗传信息，而紧缩的超螺旋态储存着必需的生物学过程的能量和信息，是一种重要的功能态。

三、人类基因组图谱

单倍体人类基因组由大约 30 亿个碱基对组成，这些碱基对分布在 22 条常染色体和一对性染色体中的一条上。随着人类基因组计划开展以来，检测技术的进步，奠定了全基因组自动化测序和基因鉴定的细胞遗传学图谱、遗传连锁图、物理图谱、转录图谱、序列图谱，它们从不同角度为基因组测序和基因鉴定做出了贡献。这些图谱首先解决了基因在染色体上定位，随后基因组被分割成 DNA 序列片段越来越小，最初的长距离图谱到可以克隆的 DNA 序列图，分辨率从最初的1～5Mb 提高到 50～1000kb，因而为基因鉴定提供了理论与技术基础。

（一）细胞遗传学图谱

人类染色体应用显带技术，在光镜下可区分出染色体的 400 余条带，而且人类基因又分布于 22 条染色体各自特定的染色体之中。后者则通过基因定位的体细胞杂交技术寻找基因特定的染色体，前者则应用荧光原位杂交（FISH），通过荧光标记的基因组克隆，可有效地在细胞遗传学图谱上定位基因或其他 DNA 片段。

1. 基因定位的体细胞杂交

大多数体细胞杂交是用人分离培养的体细胞和小鼠、大鼠或仓鼠的体细胞进行杂交，形成融合细胞（杂种细胞）。此杂种细胞有两个重要特点：

（1）此杂种细胞在繁殖传代过程中出现保留齿类一方染色体，而人类染色体则逐渐丢失，最后只剩一条或几条。

（2）杂种细胞保留或丢失人类染色体时有重叠现象。

应用上述两个特点而设计的克隆嵌板法来判别人类 22 对常染色体和一对性染色体是一种简便、有用的基因染色体定位方法（表 1-1）。

表 1-1　克隆嵌板示意表

杂种克隆	保留的人染色体							
	1	2	3	4	5	6	7	8
A	＋	＋	＋	＋	－	－	－	－
B	＋	＋	－	－	＋	＋	－	－
C	＋	－	＋	－	＋	－	＋	－

需要全本请访问 www.ertongbook.com

如表 1-1 所示，选择三个保留有 4 条人染色体的杂种细胞克隆，但染色体各不相同。如果某一表型特征只出现在克隆 A、C 而不见于 B，就可确定该特征基因只在第 3 号染色体上，因为 A、C 克隆都在第 3 号染色体而 B 克隆则无，其他亦可同理判别。这样三个克隆可对 8 条（$2^3=8$）染色体作出判断；如果有五个克隆（$2^5=32$），就可对人类 22 对常染色体和 1 对性染色体作出判断。

2. 荧光原位杂交（FISH）

应用上述克隆嵌板法基因定位于某一染色体后，尚需进一步了解待测基因的该染色体上的区带定位，此时则要进行荧光原位杂交（FISH）。荧光原位杂交是一种用特殊荧光标记核酸探针，在显微镜载玻片上，对中期染色体标本进行杂交（原位即指标本上 DNA 原位变性）。通过荧光强弱来确定标记探针（已知基因序列）在染色体上位置（基因定位或 DNA 片段）。现已用不同荧光染料同时进行多重原位杂交。用不同的荧光染料标记多个探针，进行多个基因片段的染色体定位，获得大量基因组 DNA 克隆，为 DNA 序列直接测序提供了理想材料。

原位杂交是一种分子杂交，其基本原理是碱基的互补配对原则，同源的 DNA-DNA 双链或 DNA-RNA 链在一定条件下能结合成双链，因而必须在已知探针（基因序列）条件下方可进行，而在未知致病基因时，则无法进行基因在染色体上的区带定位。

（二）遗传连锁图

应用原位杂交对基因在染色体上的区带进行定位其前提是要了解该基因的碱基序列作为探针，但大多数的情况该基因序列是未知的。

遗传学上有两个重要规律，即基因连锁遗传和基因分离与重组。遗传连锁分析就是据此两规律而设计的一种分析方法。应用与待测基因相邻近的多态性遗传标记（可测序列），以遗传连锁特征和连锁遗传标记之间重组频率作为遗传性距离，观察与疾病风险关系，进行基因在染色体区带定位。

所谓遗传连锁图是以遗传多态性的遗传标记为路标与遗传性距离（图距）的基因图。通过计算由于重组使两个遗传标记的等位基因之间产生分离重组频率，可以估计出两个遗传位点之间的距离（图距），一般用厘摩（cM），即每次减数分裂的重组率为 1% 来表示。一般控制在 0.7cM 之内，否则由于重组分离而发生偏差。

建立精细遗传图的关键是要获得足够的高度多态性遗传标记。最初是应用限制性酶切长度多态性（RFIP）作为遗传标记，到 20 世纪 90 年代已被微卫星（MS，又称 STR）所取代。微卫星标记是散在分布整个基因组中的短串联重复的二核苷酸、三核苷酸或四核苷酸序列，其中最常见的是 CA/GT 串联重复。在某一染色体特定位点，同源染色体间微卫星标记的重复次数有很大差异，使用微卫星重复序列的侧翼序列作为引物进行聚合酶链反应（PCR）即可检出微卫星片段长度差别。不

同的扩增产物(等位片段),可用聚合烯酸胺凝胶电泳方法分离,通过银染或放射自显影、荧光标记等方法加以检测。

总之,由于微卫星分布广、多态性、易检测等特点,已被作为遗传标记广泛采用。近年来第三代的多态性标记——单核苷酸多态性 SNP 标记开始应用于遗传连锁分析。单核苷酸多态性是稳定的单碱基取代的两个等位基因(二态性),以较高密度遍布于整个基因组中(将两个基因组相比较,约 1000 个 bp 出现 1 个 SNP)。这些特性使之成为适合于高通量自动化基因型分析方法的遗传标记。

(三) 物理图

1. 物理图概况

物理图谱是指 DNA 序列上两点之间实际距离。应用以已定位的 DNA 序列 STS(序列标记位点)作为路标(与遗传图不同,遗传图是以高度多态的遗传标记如简单串联重复序列 STR,MS 为路标),以 DNA 实际长度[即 bp、kb、Mb(碱基对、千碱基对、百万碱基对)]为图距的基因组图谱(遗传图以重组率为图距),本质上是一种以特异的 DNA 序列为标志所展示的染色体图。

物理图谱的连续性是大规模测序的基本保障,那么,怎样把染色体划分成可测定形式呢? 因而物理图谱中要解决两个问题:① 哪些 DNA 序列可作为序列标记位点(STS)。② DNA 克隆(选择性扩增所需 DNA 片段),用于 DNA 测序。

2. 哪些 DNA 序列作为序列标记位点(STS)

序列标记位点(STS)作为基因组染色体位标,即每个 STS 在基因组内有特定而单一的同源位置,应用其作为 DNA 标志,在染色体上定位。根据这些特性凡是能用简单、快速的 PCR 方法测定的 DNA 标志均可作为 STS。

STS 组成分析:STS 中有些来自非编码区的单拷贝 DNA 顺序只起位标作用,有些还有多态性,这部分 STS 的引物顺序仍然具有单拷贝性质,但扩增产物中通常含有短串联重复序列(STR),它们可作为基因连锁分析也可作为染色体位标应用。还有一些 STS 本身就是基因,它们中有些可能原本是从 cDNA 发展而来的,称之为 EST(人类基因表达序列标记)。

总之,在基因组内有其特定而单一的同源位置的 DNA 顺序包括 STR 和 EST,均可作为序列标记位点(STS)。

3. 构建覆盖每条染色体的大片段 YAC 和/或 BAC 连续克隆系列

对人类基因组 DNA 的测序,必须从 DNA 克隆开始,对基因组 DNA 的每个基因的研究也需要首先得到 DNA 克隆。DNA 克隆可选择性地扩增所需 DNA 片段,用于 DNA 测序。此扩增的片段是通过合适的载体才能获得。载体有酵母人工染色体(YAC)、细菌人工染色体(BAC)等。

(1) YAC 连续克隆系(YAC-STS 物理图谱)(人工酵母染色体携带的人类

DNA 文库）

人工酵母染色体是利用酵母染色体的结构和复制系统来复制 DNA 大分子，可以用来克隆 200kb 到 2000kb 的 DNA 片段。

YAC 载体可以装载大小超过百万碱基对（Mb）的人类 DNA 片段。绝大多数人类 YAC 作图是使用 PCR 扩增标记（序列标记位点 STS_S）来检测 YAC 载体中是否存在某一特定基因座。当然也可使用 DNA-DNA 杂交技术。如果使用足够多的 STS_S 或杂交探针，就可构建出覆盖人类基因组上较大区域的克隆重叠群（把大片段 DNA 克隆，依其在染色体上所在位置排序，得到的相应重叠的一系列克隆称为克隆重叠群），选取有关的克隆进行 DNA 测序，可以较为容易地组装成全染色体或基因组 DNA 顺序。

YAC 的克隆容量大是其主要优点，但 YAC 载体有两个严重缺陷。首先是将人类 DNA 克隆进 YAC 载体过程中会发生因缺失而产生 DNA 重排，产生嵌合体。嵌合体是由不同染色体来源的 DNA 片段，连接在一起构成的克隆，使原本两个在人类基因组中物理上无关区域被连接在一起而装载到同一个 YAC 载体中，因而不能把这种克隆放在任何染色体的任何位置，它不能被用来进行 DNA 测序或基因的克隆研究。

其次是 YAC 克隆 DNA 分离困难。YAC 的大小和性质与酵母染色体无明显差异，因而，酵母染色体中分离出来的人工染色体进行鸟枪法克隆和测序是一个费时又低效的过程。

为了给基因组测序提供一定容量，不会发生大量嵌合体，而且又易 DNA 分离纯化的载体，有人构建了细菌人工染色体 BAC 载体。

（2）BAC（细菌人工染色体）连续克隆系

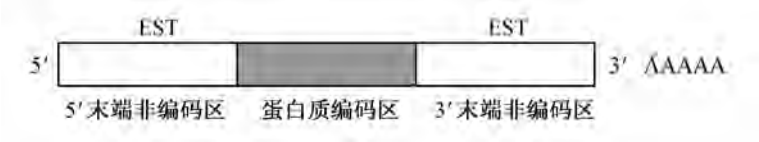
BAC 是以大肠杆菌 F 质粒为基础构建的（即利用 F 质粒的单拷贝控制系统来设计的）。BAC 是一种环状分子，在大肠杆菌细胞内非常稳定，不会发生缺失或重排，而且易分离纯化，但其克隆容量与 YAC 载体相比不及 YAC，其克隆容量为 100～300kb 基因组 DNA。

（四）转录图谱

由于要搞清人类基因组 3×10^9bp 全部序列绝非轻而易举的工作，而且人类基因组中大多数成分是重复序列和非编码序列，于是部分科学家重点研究一些可被转录的序列，这就是人类基因组计划中的 cDNA 计划，它是从研究基因转录产物 mRNA 开始的。由于 mRNA 不稳定，以及诺瑟思印迹杂交检测条件苛刻，一般应用逆转录酶把 mRNA 逆转录成 cDNA，故称之为 cDNA 计划。其主要目标是收集大量 cDNA 序列片段，这些序列被称为基因表达序列标志（EST）。

EST 指长度为 300～500bp 的一段 cDNA。自动化 DNA 测序仪的应用极大

地促进了人体组织中 mRNA 转录本的详尽测定。它通过 PolyT（cDNA 克隆末端为 PolyA，T 与 A 互补）或随机引物引导的细胞 mRNA 逆转录反应构建 cDNA。这些克隆末端可测得 3′末端非编码区或 5′末端非编码区的短序列，即表达序列标志也可获得这些基因表达转录本（蛋白质编码区）。



基因的不同表达序列标志 ESTs 示意图

目前已经构建了大规模的 EST 集合。原则上应对所有组织细胞以及其不同的发育阶段都进行测定，找到足够多的表达序列标志。截至 1997 年底，公共数据库中 EST 序列已近 90 万条。据推算现有 EST 可能包含 90% 以上人类基因编码序列。

（五）序列图谱

高分辨率的遗传图是制作人类基因组物理图的基础，而物理图谱的连续性是大规模测序的基本保障。

人类基因组序列测定，一般利用基因组 BAC 物理图为基础采用下述步骤：① 将待测克隆随机切成小片段（约 1500bp）；② 将小片段 DNA 克隆入测序载体；③ 对每千个碱基对的 DNA 进行 10～30 个亚克隆，使用荧光标记的双脱氧法和自动序列仪进行测序；④ 将相互重叠的读出序列组装成连续的多序列重叠线；⑤ 从质量最高的读出序列中取得最后的确认序列。

关于 cDNA 测序，其序列仅占总序列的 3%～5%，对这部分序列进行测定，将有可能直接导致编码蛋白质基因的发现，获得基因组中对医学和生物制药关系最密切的信息。当前 cDNA 测序的趋势是由对部分片段（即 EST）的随机测序转向全长 cDNA 的克隆和测序。因为 cDNA 测序属于编码序列，有潜在的巨大商业价值。

四、基因组 DNA 序列提供哪些信息，有助于我们对分子生物学的研究

（一）DNA 序列可推测出基因的结构

基因结构包括启动子、内含子、外显子等，而且多个外显子存在不同的拼接。cDNA 是与基因转录信使 RNA 互补的 DNA，代表了基因外显子的生物学信息，依靠表达序列标记技术（EST）推测基因外显子序列提供有力线索，从中有可能找到与疾病相关的新基因。

（二）DNA 序列中可找出很多与结构有关的序列

例如，端粒特有的序列—GGATTT；染色体与蛋白质之间相互作用，是通过 DNA 的一些特定序列来实现的，如转录因子与顺式作用元件。

（三）非编码大量重复序列和小分子 RNA 的发现

人类基因组序列分析发现，非编码区存在大量重复序列及与基因沉默有关的 miRNA 和 RNA 干扰现象，为基因调控提供了全新的视角（详见第二章 miRNA 和 siRNA）。本章主要讨论大量重复序列的意义。

（1）上述遗传连锁图中已谈到的短串联重复序列（微卫星）可作为遗传标记。

（2）人类 DNA 序列存在大量重复序列，如果将重复序列比作大海的话，那么基因则是大海中的岛屿，两者同存同消。Alu 家族是人类特有的一大类中度重复序列，其成员共有 30000 多个，平均 5～6kbp 就有一个这样的结构，每个长度约 300bp，它的 Alu 命名，是由于其序列中第 170 碱基处有一个限制性内切酶 AluI（AG/CT）的酶切位点而得名。Alu 序列分散在整个人类基因组中，具有种属特异性。中度重复序列中还有 KpnI 家族、Hinf 家族，也有类似效应。它们的共同特征是重复单位序列相似，但不完全一样，散在分布在基因组中，序列的长度和拷贝数很不均一，均具有种属特性，可制成探针，以鉴别不同种属的细胞 DNA 来源。值得一提的是，Alu 序列是人类特有的序列，有别于细菌与病毒，这就为感染性疾病的诊断提供了理论基础。此外，高度重复序列 GC，在原核生物 DNA 中 GC 含量均匀，而真核细胞大部分片段 GC 含量在 30%～50%，也可作为参考。

基因组比较分析，将会提供基因进化史的相应证据，GC 分布比率的差异，重复序列的分布都将可能与基因分布有关，这些非编码区序列的意义有待进一步研究。

总之，基因组测序告诉我们，人类基因组是一个结构稳定的体系，不同民族、群体和个体都有 46 条染色体（22 对常染色体和一对性染色体），有相同数目的基因和基因分布，也有相同的核苷酸序列，从而保证了人类作为一个物种的共同性和稳定性。这就产生了一个重要问题，即基因组测序的了解，对疾病发生发展有什么关系呢？也就是说，不同族群和个体都有相同数目的基因和基因分布，也有相同的核苷酸序列，它是如何导致疾病发生的呢？这就涉及遗传与变异问题。下节将叙述人类基因组 DNA 序列的变异。

第二节　人类基因组 DNA 序列的变异

生物领域中的基本规律之一是遗传与变异。遗传决定了人类基因组是一个结构稳定体系，基因测序具有代表性；变异则导致了不同种族、群体和个体间基因组

序列的差异。在长期进化过程中，基因组的 DNA 序列不断发生变异，这种变异可能是有害的（导致疾病发生）、有益的（适应内外环境变化）或中性的。它们其中的一些被保留下来，导致个体、群体间基因组序列变异或多态性。这种变异或多态性有助于我们分析它与疾病表型、疾病风险之间的关系和建立多态性遗传标志，进行基因鉴定和定位，使我们对疾病预防、诊断、治疗置于坚实的遗传学基础之上。

一、人类基因组 DNA 序列分类

人类基因组 DNA 因序列重复性和序列相似性而构成两大类型。序列相似性是理解基因家族系列与基因超家族系列的基础；而序列重复性变异程度较大，又具有个体遗传特性，常作为遗传标记重点介绍。

人类基因组 DNA 序列的显著特点之一，是含有许多重复序列。所谓重复序列是指含有一个以上相同序列的拷贝数。重复序列的长短和次数有极显著差别，根据重复序列的重复程度（次数），分为高度重复序列（卫星 DNA）、中度重复序列和单拷贝序列。中度重复序列中又根据碱基对长度不同，串联重复序列中又有小卫星（15～65bp 的基本单位串联重复）和微卫星（基本序列 1～8bp 重复）。

人类基因组 DNA 中约有 3/4 是单拷贝的，即单核苷的序列只出现一次或很少几次，其余 1/4 则属重复序列。

二、人类基因组 DNA 序列的变异

遗传与变异是生物界的普遍现象，这样生物才能生存与发展。DNA 水平上变异可表现在重复序列拷贝数不同，也可表现为单拷贝序列的变异。除一卵双生子外没有两个体的基因组序列是完全相同的。个体的这种差异称为 DNA 多态性（多样性位点），其产生是由于两条染色体分别来自父和母任意两套染色体。绝大多数变异是无害的或者只引起蛋白质的微小变化，仅有很小部分变异会影响基因功能，此种变异称为突变，也就是说，突变通常是指那些可以或可能影响基因表达水平或基因功能的 DNA 序列变异，即与疾病相关的特定基因状态称为突变。在基因组 DNA 序列变异中，遗传多态性较多表现在重复序列和单核苷酸的多态性和突变。

（一）DNA 重复序列的多态性

人类基因组中遗传多态性较多地表现在重复序列，特别是短串联重复序列的小卫星 DNA 和微卫星 DNA。它们的多态性主要基于重复序列拷贝数的变异。

1. 小卫星 DNA

通常小卫星 DNA 全长在 20kb 左右，由 15～65bp 的基本单位串联重复而成，

需要全本请访问 www.ertongbook.com