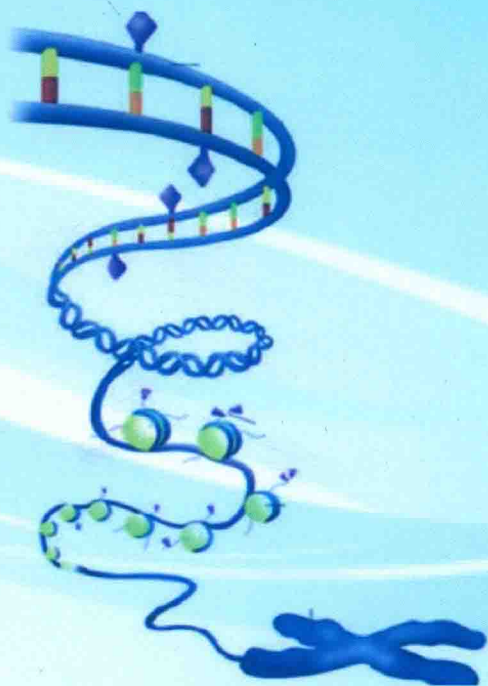





高通量测序技术

HIGH-THROUGHPUT SEQUENCING

主 编 李金明



 科学出版社

国家科学技术学术著作出版基金资助出版

高通量测序技术

主 编 李金明

科学出版社

北 京

内 容 简 介

本书从临床实验室一线应用的角度,系统阐述了高通量测序技术的基础理论及临床应用。全书共18章,内容包括测序技术的发展历程及趋势,文库构建原理及特点,高通量测序原理及特点,生物信息学分析原理及特点,数据库,高通量测序实验室设计,高通量测序仪及其发展,高通量测序临床应用的质量保证,以及高通量测序在染色体非整倍体无创产前筛查、胚胎植入前遗传学检测、单基因遗传病检测、遗传性肿瘤诊断和治疗、肿瘤靶向治疗、临床药物基因检测、病原微生物检测、表观遗传检测、免疫组库和转录组学等领域的应用等。

本书资料翔实、新颖,知识性、科学性和实用性强,特别适合作为临床实验室一线人员、各级临床医师和医学院校师生的参考书,也可作为推广高通量测序技术临床应用的培训教材。

图书在版编目(CIP)数据

高通量测序技术 / 李金明主编. —北京:科学出版社, 2018.12
ISBN 978-7-03-059306-1

I. ①高… II. ①李… III. ①基因组-序列-测试-研究
IV. ①Q343.1

中国版本图书馆CIP数据核字(2018)第244389号

责任编辑:杨卫华 / 责任校对:张小霞
责任印制:赵博 / 封面设计:龙岩

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京画中画印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2018年12月第一版 开本:787×1092 1/16

2019年1月第二次印刷 印张:40 插页:12

字数:908 000

定价:158.00元

(如有印装质量问题,我社负责调换)

《高通量测序技术》编写人员

主 编 李金明

副主编 张 瑞

编 者 (以姓氏笔画为序)

王 萌 李子阳 李文丽 李金明

杨 玲 杨 新 汪 维 张 括

张 瑞 张 燕 张红云 张嘉威

林贵高 易 浪 高 芄 黄 杰

彭绒雪 韩彦熙 楚玉星 戴平平

前 言

高通量测序 (high-throughput sequencing, HTS) 即下一代测序 (next generation sequencing, NGS), 又称为大规模平行测序 (massively parallel sequencing, MPS), 最早是从焦磷酸测序的原理上发展起来的, 目前较为成熟的应用平台主要有 Ion Torrent、Illumina 和 Complete Genomics (CG) 的各种型号。自 10 余年前其首次出现以来, 技术日趋成熟。因为高通量测序一次可检测大量靶基因及其变异位点, 检测灵敏度和特异性高, 兼具定性和定量检测, 并且检测费用与同样数量的基因和位点检测相比较低, 因而在无创产前筛查 (noninvasive prenatal screening, NIPS)、肿瘤基因突变、遗传病、胚胎植入前遗传学筛查 (preimplantation genetic screening, PGS) 和胚胎植入前遗传学诊断 (preimplantation genetic diagnosis, PGD)、病原微生物及宏基因组学 (metagenomics) 等领域展现了极为广阔的临床及科研应用前景, 成为目前 DNA 和 RNA 序列分析最高效的工具, 也是精准医学时代研究和临床疾病诊疗的支撑技术, 使得基因组学和转录组学迅速由概念走向临床疾病诊疗实践, 预示着疾病诊疗分子时代的到来。

高通量测序技术的特点之一是操作步骤多、程序复杂, 既包括实验室内的标本预处理、核酸提取及其片段化 (基因组检测时)、建库、扩增、靶序列富集、混样 (pooling)、测序前准备及测序, 即湿桌实验过程 (wet bench process); 也包括测序后的数据质量分析、比对、变异识别、注释和结果报告与解释等生物信息学分析流程 (bioinformatics pipeline), 即干桌实验过程 (dry bench process)。上述任一环节出现问题, 均会影响检测结果的准确性, 进而影响临床决策。同样, 我国精准医学的研究计划也涉及大量对特定疾病和人群生物样本的高通量测序分析研究, 若上述环节出现问题, 将导致研究失去可靠价值。

如何避免假阳性结果是所有高灵敏检测方法需要高度关注的问题。严格有序的实验室各功能区的物理分隔及持续的通风换气是避免因“基因或核酸”气溶胶导致实验室和标本间交叉污染的前提硬件条件; 在日常工作中, 严格遵循各功能区物品专用、各功能区域的单一工作流向、及时的实验室清洁, 以及生物信息学分析流程中适当的数据库和滤过策略的使用等, 是防止假阳性的软件要素。

假阴性结果通常容易被忽视, 但在精准医学尤其是精准肿瘤学的临床实践中, 假阴性结果与假阳性结果对临床疾病诊疗决策具有同等的危害性。临床标本的采集、运送和保存过程中处理不当, 核酸提取过程中靶核酸的丢失和提取试剂中可抑制后续检测过程 (如扩增) 的试剂 (如有机溶剂) 残留, 因仪器设备维护和校准不到位所致的加样不准、光路不正、温度不均一等, 均有可能造成假阴性结果。因此, 涉及上述各个关键环节的

具有可操作性的标准操作程序（standard operation procedure, SOP）的制订，以及工作人员在日常工作中的严格遵循，是避免假阴性结果产生的必要条件。

本书主要针对目前在国内外临床实验室使用最广泛及最具临床和科研应用前景的高通量测序技术，从测序技术的历史、发展和趋势，高通量测序的建库方法、数据库、生物信息学分析，各种高通量测序技术的原理及其特点，临床高通量测序实验室的设计及各区相应仪器设备的配备，各类高通量测序仪器的特点，临床高通量测序的质量保证，以及主要临床高通量测序现有或未来可能有应用前景领域的特点、临床意义及检测过程中质量保证应关注的细节等进行了论述，以期为从事临床和科研高通量测序的实验室技术人员提供参考性建议。

从事基础和临床科学研究的人员也可从本书得到相应的启发，因为一个成功的科研与实验过程中严格的质量控制是分不开的，尤其是精准医学科学研究通常离不开高通量测序的应用。试想一个采用高通量测序进行的基础或临床科学研究工作，如果从标本采集、运送和保存，到防污染的实验室功能分区，测序过程（湿桌实验和干桌实验）的方法学性能确认，仪器设备的操作、维护和校准，人员培训，SOP的可操作性及实际遵循，质量控制（室内质量控制和室间质量评价）等都处于不规范、不标准的状况，有可能得到真实的科研结果吗？结果也一定是不能重复和再现的，也就失去了其科学价值。大量证据表明，如果没有严格的质量控制，科研人员将会在实验室中得到的“假阳性”或“假阴性”结果，作为一个重大发现去发表，不但浪费了宝贵的时间和经费等资源，还会误导同行，并且可能会造成难以挽回的后果。有些涉及仲裁、具有重大影响的检测部门，则会因为“假阳性”或“假阴性”结果，对国家、社会和人民造成严重损失。上述方面国内外均有典型案例。

在编写本书3年余时间中，我国从事高通量测序的同道给了我们一些很好的启发，也是业内同行的建议及对高通量测序规范化和标准化的渴求，促使我们开始尝试并努力完成本书的撰写，并力求以最容易理解的方式来撰写，力求做到最好。

感谢本书所有编者的辛苦付出！尤其要感谢在本书编写过程中，丁健生、陈坤、周利、史吉平、胡丐锋、周琼、刘云青、鲁添、韩东升等博士研究生、硕士研究生，临床分子与免疫室的谢洁红老师，在文字、体例和插图校对、编辑过程中所做的大量认真细致的工作！感谢国家科学技术学术著作出版基金为本书出版提供保障；感谢北京医院学术专著出版基金对本书的资助！

由于编者水平有限，书中难免存在不足之处，敬请同道批评指正，以便再版时更正。

李金明

2018年7月8日

目 录

第一章 测序技术的发展历程及趋势	1
第一节 概述	1
第二节 第一代测序技术	3
第三节 第二代测序技术	11
第四节 第三代测序技术	20
第五节 测序技术的应用	23
第六节 测序技术的发展及展望	25
第二章 文库构建原理及特点	29
第一节 文库构建原理	30
第二节 文库构建的分类及特点	37
第三节 文库制备的质量评价及影响因素	57
第三章 高通量测序原理及特点	62
第一节 Illumina 测序	62
第二节 Ion Torrent 半导体测序	66
第三节 Complete Genomics 测序平台	71
第四节 单分子测序	73
第五节 纳米孔单分子测序	76
第四章 高通量测序的生物信息学分析原理及特点	84
第一节 概述	84
第二节 生物信息学数据存储格式	86
第三节 高通量测序生物信息学分析环境搭建	92
第四节 高通量测序检测数据分析原理和基本流程	109
第五节 生物信息学分析流程的性能确认	124
第五章 数据库	136
第一节 概述	136
第二节 序列比对类数据库	139
第三节 突变过滤类数据库	141
第四节 突变注释与解读类数据库	145
第五节 实验室自建数据库	170

第六章 高通量测序实验室设计	174
第一节 高通量测序实验室分区设计	174
第二节 高通量测序实验室各功能区主要功能及基本仪器设备配置	188
第七章 高通量测序仪及其发展	197
第一节 高通量测序仪的发展历史	197
第二节 高通量测序仪简介	202
第三节 测序仪的选择	214
第四节 高通量测序仪的使用与维护	216
第八章 高通量测序临床应用的质量保证	221
第一节 高通量测序临床应用的管理	222
第二节 高通量测序检测分析前质量控制	231
第三节 高通量测序检测分析中质量控制	246
第四节 高通量测序检测分析后质量控制	319
第九章 高通量测序在染色体非整倍体无创产前筛查中的应用	342
第一节 概述	342
第二节 染色体非整倍体无创产前筛查高通量测序检测的临床有效性	344
第三节 染色体非整倍体无创产前筛查高通量测序检测原理	345
第四节 染色体非整倍体无创产前筛查高通量测序检测的质量保证	351
第十章 高通量测序在胚胎植入前遗传学检测中的应用	369
第一节 胚胎植入前遗传学诊断/筛查技术发展简史	369
第二节 胚胎植入前遗传学诊断/筛查临床实践现状	371
第三节 胚胎植入前遗传学诊断/筛查相关指南	373
第四节 胚胎植入前遗传学诊断产品质量评价参考品制备	377
第五节 胚胎植入前染色体非整倍体检测试剂盒(高通量测序法)质量控制 评价技术指南中诊断试剂的设计开发要求	385
第六节 胚胎植入前染色体非整倍体检测试剂盒(高通量测序法)质量控制 评价技术指南中诊断试剂设计开发流程的质量控制要求	388
第七节 胚胎植入前遗传学诊断/筛查技术的局限性	393
第八节 胚胎植入前遗传学诊断/筛查技术的发展方向	395
第十一章 高通量测序在单基因遗传病检测中的应用	401
第一节 概述	401
第二节 单基因遗传病高通量测序检测流程	402
第三节 无创单基因病检测	421
第四节 展望与挑战	427
第十二章 高通量测序在遗传性肿瘤诊断和治疗中的应用	431
第一节 概述	431

第二节	遗传性肿瘤的遗传咨询和检测	436
第三节	遗传性肿瘤相关检测指南和数据库	451
第四节	数据共享：突变型分类和结果解释	453
第十三章	高通量测序在肿瘤靶向治疗中的应用	458
第一节	概述	458
第二节	高通量测序进行肿瘤基因突变检测的技术策略和分析原理	459
第三节	肿瘤基因突变高通量测序检测的质量保证	470
第四节	高通量测序在肿瘤靶向治疗中的应用	479
第十四章	高通量测序在临床药物基因检测中的应用	486
第一节	概述	486
第二节	药物相关基因及其临床检测的意义	488
第三节	临床药物基因高通量测序检测	491
第四节	临床药物基因高通量测序检测的性能确认和质量控制	507
第五节	国内外临床药物基因检测指南	508
第六节	药物基因检测的临床推广应用	510
第七节	展望	511
第十五章	高通量测序在病原微生物检测中的应用	515
第一节	概述	515
第二节	基于高通量测序的病原微生物检测原理	518
第三节	基于高通量测序的病原微生物检测流程	519
第四节	病原微生物高通量测序的检测性能和质量控制	523
第五节	病原微生物高通量测序面临的挑战	526
第十六章	高通量测序在表观遗传检测中的应用	530
第一节	表观遗传学概述	530
第二节	高通量测序在 DNA 甲基化检测中的应用	532
第三节	高通量测序在组蛋白修饰检测中的应用	543
第四节	基于高通量测序的非编码 RNA 检测	548
第十七章	高通量测序在免疫组库测序中的应用	562
第一节	概述	562
第二节	免疫组库高通量测序分析	567
第十八章	高通量测序在转录组学中的应用	577
第一节	概述	577
第二节	高通量转录组测序	580
第三节	转录组测序数据分析流程	590
第四节	转录组测序的临床应用	597
附录		601

附录 1	国家卫生计生委临床检验中心 2016 年全国肿瘤游离 DNA (ctDNA) 基因突变检测室间质量评价调查活动安排及注意事项	601
附录 2	国家卫生计生委临床检验中心 2016 年全国肿瘤游离 DNA (ctDNA) 基因突变高通量测序检测室间质量评价调查活动回报表	602
附录 3	国家卫生计生委临床检验中心 2016 年全国肿瘤游离 DNA (ctDNA) 基因突变 ARMS 和数字 PCR 检测室间质量评价调查活动回报表	606
附录 4	国家卫生计生委临床检验中心 2016 年全国肿瘤游离 DNA (ctDNA) 基因突变检测室间质量评价调查活动总结	608
附录 5	单基因病样品采集须知	613
中英文名词对照		616
彩图		

第一章

测序技术的发展历程及趋势

自从 DNA 双螺旋结构被解析以来，人们一直致力于探究基因组的复杂性和差异性。测序技术面世至今的几十年发生了诸多技术革新和大规模增长，经历了从以 Sanger 测序为代表的第一代测序、大规模平行测序 [又称高通量测序 (high-throughput sequencing, HTS)] 和第三代测序技术的发展。迄今为止，高通量测序 [也称下一代测序 (next generation sequencing, NGS)] 技术日趋成熟，正进入临床疾病的诊疗应用阶段，尤以 Illumina、Ion Torrent 和 CG 三大平台为代表。本章就整个测序技术的发展历程、主要原理、应用领域及发展前景进行介绍。

第一节 概 述

核酸 (nucleic acid) 是生命体遗传信息的主要载体。1953 年 Watson 和 Crick 关于 DNA 双螺旋结构的发现表明，几乎所有基因的三维结构基本一致，A、T、C、G (RNA 中为 A、U、C、G) 4 种碱基数量和排列顺序的变化造就了生命体的多样性。因此 DNA 和 RNA 序列被称为遗传密码，是分析基因结构、功能及其相互关系的基础，也是临床疾病分子诊断最精确的判定依据。基因测序技术正是解读这些生命密码的基本手段。

1977 年 Sanger 和 Gilbert 分别提出双脱氧链终止法 (dideoxy chain-termination method / Sanger sequencing) 和化学降解法 (chemical degradation method / Maxam-Gilbert method) ，标志着第一代测序技术的诞生。由于对核酸测序技术做出的重大贡献，Sanger 和 Gilbert，以及发现 DNA 重组技术的 Berg 分享了 1980 年的诺贝尔化学奖。此后，随着研究工作的日益深入，自动化测序、焦磷酸测序、高通量测序、甲基化修饰测序等技术应运而生。2005 ~ 2007 年，以基于边合成边测序原理的 Roche 公司的 454 技术和 Illumina 公司的 Solexa 技术，以及基于边连接边测序原理的 Life Technologies 公司的 SOLiD 技术为标志的第二代高通量测序技术诞生，测序的规模化进程有了突破性的进展。尽管高通量测序技术很大程度上增加了测序通量、降低了测序成本，并且极大地推进了相关研究的进展，但是模板扩增和序列读长 (即读段长度，指在单次测序反应中产生的碱基数) 短始终是其难以克服的技术瓶颈。2008 年至今，随着物理、化学、材料和生命科学的不断发展和融合，以单分子测序和长读长为标志的第三代测序技术应运而生。第三代测序技术读长

可达上百 kb，无须 PCR 扩增即可实现对 DNA 分子的实时检测。

自 1952 年 Alfred D. Hershey 和 Martha Chase 证实 DNA 是遗传物质以来，整个测序技术的发展历程见图 1-1。

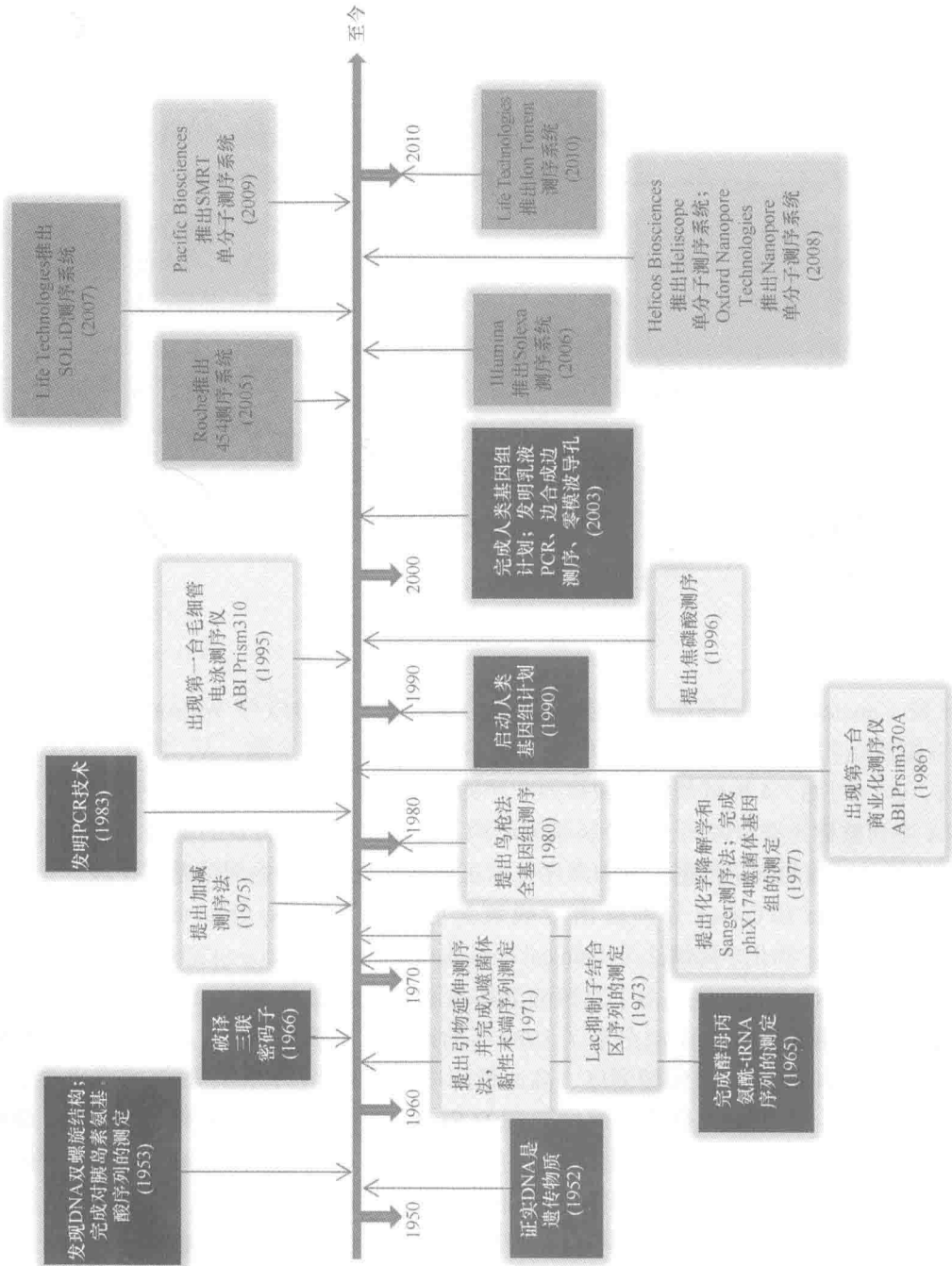


图 1-1 核酸测序技术的发展历程

第二节 第一代测序技术

核酸测序技术的萌芽期最早可追溯到 20 世纪 50 年代, 即 Whitfield 等使用化学降解法测定无支链的 RNA 序列^[1]。该方法利用磷酸单酯酶的脱磷酸作用和高碘酸盐的氧化作用从多聚核苷酸链末端逐一解离寡核糖核酸, 并用色谱法测定其种类。但是由于操作复杂, 并未被广泛应用。20 世纪 60 年代中期, Robert 等利用小片段重叠法, 耗时 7 年首次完成对酵母丙氨酸-tRNA 序列 76 个核苷酸序列的测定^[2]。70 年代初, 华裔分子生物学家吴瑞提出位置特异性引物延伸策略, 并于 1971 年首次成功测定了 λ 噬菌体 12 个碱基的黏性末端序列^[3], 这是文献记录的最早的 DNA 序列分析方法^[4], 但仅限于测定 DNA 短序列。1973 年, Gilbert 和 Maxam 利用化学降解法测定出 Lac 抑制子结合区 24 个碱基的 DNA 序列^[5]。Sanger 紧随其后于 1975 年报道了更为简易的加减测序法 (plus-minus sequencing)^[6]。1977 年 Sanger 在加减测序法的基础上创建了双脱氧链终止测序法^[7]。同年, Gilbert 和 Maxam 在原有方法的基础上合创了化学降解测序法^[8]。上述两种测序方法原理虽大相径庭, 但是都生成了相互独立的若干组带放射性标记的寡核苷酸混合物, 这些混合物有共同的起点, 随机终止于一种或多种特定的碱基。通过对各组寡核苷酸混合物进行聚丙烯酰胺凝胶电泳 (polyacrylamide gel electrophoresis, PAGE) 即可从放射自显影片上直接读出 DNA 核苷酸顺序。

由此, 人类获得了探究生命体遗传信息的能力, 并以此为开端步入基因组学时代。随着现代分子生物学技术的不断发展, 经典的 Sanger 测序法不断改进和优化, 并发展为自动化测序, 为人类基因组计划 (Human Genome Project, HGP) 做出了重大贡献。这一时期还出现了如鸟枪法 (shotgun method)、杂交测序 (sequencing by hybridization, SBH) 法等新一代测序方法, 为 DNA 序列分析提供了强有力的支持。

一、Maxam-Gilbert 化学降解测序法

Maxam-Gilbert 化学降解测序法, 即先对 DNA 片段的 5' 或 3' 末端进行放射性标记, 再采用特异性化学试剂修饰和裂解特定的碱基位点, 从而得到一系列有着共同放射起点但长度不一的 DNA 片段混合物。这些以特定碱基结尾的 DNA 片段依据断裂点的位置决定其片段长度及聚丙烯酰胺凝胶电泳的排布, 最后经放射自显影技术检测末端标记的分子, 自下而上直接识读 DNA 碱基序列 (图 1-2)。

(一) 待测 DNA 末端放射性标记及纯化

待测样本既可以为单链 DNA 也可以为双链 DNA。一般使用 ^{32}P 标记链的 5' 或 3' 末端^[8]。放射性核素标记的方法通常有 3 种: ①用 T4 噬菌体多核苷酸激酶和 $\gamma\text{-}^{32}\text{P}\text{-ATP}$ 标记待测样本的 5' 末端; ②用末端转移酶和 $\alpha\text{-}^{32}\text{P}\text{-dNTP}$ 标记待测样本的 3' 末端; ③用 DNA 聚合酶 Klenow 片段和 $\alpha\text{-}^{32}\text{P}\text{-dNTP}$ 标记有 5' 突出端的待测 DNA 样本的 3' 末端, 这种方法能直接制备单侧末端标记的 DNA。

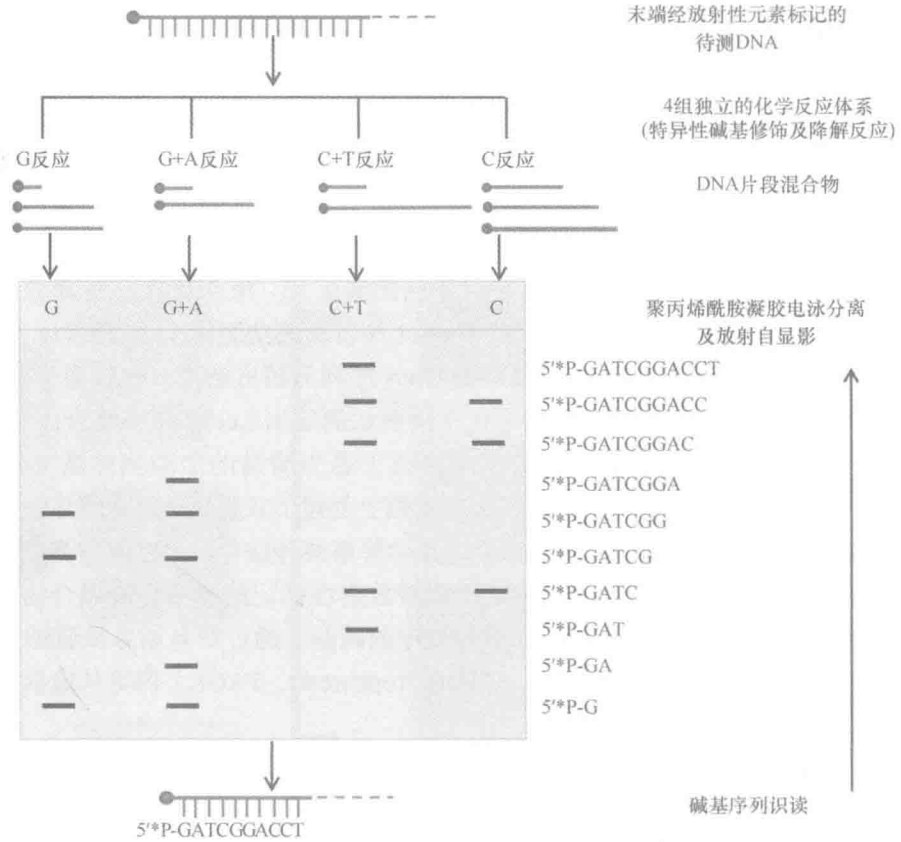


图 1-2 Maxam-Gilbert 化学降解测序法原理

使用前两种标记方法会使双链 DNA 的每一条链均有末端标记，进行碱基特异性化学降解后，产生两套互相重叠但有不同末端标记的 DNA 片段混合物。因此，必须分离两个标记的末端或互补双链，得到只有一端标记的 DNA 片段，才能进行测序反应。可以使用两种不同的限制性核酸内切酶切割，得到两个大小不同的末端 DNA 双链片段，然后用聚丙烯酰胺凝胶电泳进行分离，或者将双链 DNA 变性处理为单链后用聚丙烯酰胺凝胶电泳进行分离^[8]。

(二) 特异性碱基修饰及降解反应

将放射性核素标记后的待测 DNA 片段分成 4 个或 5 个 (A > C 反应可以不做) 独立的反应体系，进行特定碱基的化学修饰 (表 1-1)^[8,9]。被修饰的碱基容易与其糖基分离，邻近的磷酸二酯键容易断裂，从而产生各种特定大小的 DNA 片段，这个过程即为碱基特异性切割。

1. G 反应 硫酸二甲酯在 N7 位置甲基化 G，在 N3 位置甲基化 A。甲基化修饰后的糖苷键极易发生水解，中性环境下加热即可断裂，释放戊糖基团。通过碱催化的 β-消除反应，导致 DNA 链的断裂。由于 G 的 N7 原子甲基化速率比 A 的 N3 原子快 5 倍以上，中性条件下 G 的水解速率也比 A 快得多^[9]。现行的化学裂解反应均采用六氢吡啶与 DNA

中的甲基化碱基反应,这种反应对甲基化的 G 是特异的,因此获得的是含有 G 末端的 DNA 混合片段。

表 1-1 化学降解测序法碱基特异性化学修饰及反应

碱基体系	化学修饰试剂	修饰反应	主链断裂剂	断裂部位
G	硫酸二甲酯	甲基化鸟嘌呤	六氢吡啶	G
G+A	嘧啶甲酸	脱嘌呤	六氢吡啶	G 和 A
C+T	胂	断开嘧啶环	六氢吡啶	C 和 T
C	胂(高盐)	断开胞嘧啶环	六氢吡啶	C

2. G+A 反应 甲酸可使嘌呤环上的 N 原子质子化,糖苷键不稳定而发生脱嘌呤反应,得到一系列 A 和 G 混合的末端片段。

3. T+C 反应 胂在碱性条件下使 T 和 C 的嘧啶环断裂。在六氢吡啶的作用下,通过 β -消除反应,碱基两端的磷酸基团以磷酸分子的形式释放出来,从而导致该位置 DNA 链发生断裂,得到一系列 T 和 C 混合的末端片段。

4. C 反应 高盐(如 2 mol/L NaCl)条件抑制了 T 和胂的反应,只有 C 与胂特异性反应,得到一系列 C 末端片段。

(三) DNA 碱基序列的识读

聚丙烯酰胺凝胶电泳将 DNA 链按大小分开,经放射自显影显示末端标记的碱基。自胶片底部向顶部逐一读取条带。每组测序图谱为 4 条或 5 条(A > C 反应可以不做)泳道。因为化学降解反应并非绝对的碱基特异性反应,在识读时需要从 G+A 泳道中扣除 G 泳道的条带而推断 A 碱基,从 C+T 泳道中扣除 C 泳道的条带而推断 T 碱基,即如果 G+A 泳道中出现一条带,且 G 泳道中有相同大小的带,则为 G 碱基,如果 G 泳道中没有相同大小的条带则为 A 碱基。类似地,在 C+T 泳道中出现的条带,检查 C 泳道中有无同样大小的条带,如果有即为 C 碱基,无则为 T 碱基(见图 1-2)。

在测序的初始阶段,化学测序法由于仅需要简单的化学试剂和普通的实验条件,无须特异性寡核苷酸引物及高质量的大肠杆菌 DNA 聚合酶 I 大片段(Klenow 片段),并且其测序的重复性高,因此较 Sanger 测序法容易被普通的实验室和研究人员掌握。化学测序法的整个测序过程无须酶促反应,避免了合成时可能造成的错误。该测序法可以对 DNA 甲基化的修饰进行分析,对 G+C 含量较高的 DNA 片段的测序尤其适用。

但是,化学测序法对待测 DNA 纯度要求很高(盐浓度可干扰胂对 T 的修饰,也可抑制 A+G 反应中的去嘌呤过程)。由于碱基堆积力和空间位阻的作用影响双链中个别碱基的甲基化反应(如在 GGA 序列中,中间 G 的甲基化反应受到抑制)^[8]。受制于当时聚丙烯酰胺凝胶电泳的分辨能力,最多能识读 200 ~ 250 个核苷酸序列。除此之外,化学降解测序法自建立以来没有很大的改进,由于操作繁琐、化学试剂毒性大、放射性核素标记效率偏低等原因未能成为主流的测序技术,目前仅在分析某些特殊 DNA 序列,以及

DNA 和蛋白质相互作用中的 DNA 一级结构时才使用。

二、Sanger 双脱氧链终止法

20 世纪 70 年代, Sanger 将其注意力从 RNA 序列研究转向了 DNA, 并于 1975 年和 Coulson 一起提出加减测序法^[6]。利用该方法他们完成了第一个基因组——噬菌体 $\phi \times 1745$ 386 个碱基序列的测序工作^[10]。

加减测序法首次运用特异性引物, 以放射性核素标记的 dNTP 为原料, 在 DNA 聚合酶的作用下进行 DNA 链的延伸反应和碱基特异性链终止反应。合成的 DNA 产物一部分用于“加法”系统进行降解反应, 另一部分用于“减法”系统进行合成反应。用于“加法”系统的产物再分成 4 份, 每份中仅加入一种脱氧核苷三磷酸 (deoxynucleoside triphosphate, dNTP) (此处以 dTTP 为例)。由于 DNA 聚合酶具有 $3' \rightarrow 5'$ 的外切酶活性, 合成产物就从 $3' \rightarrow 5'$ 方向发生降解, 降解过程中如遇到加入的 dTTP 的位置, 降解反应即停止, 因而所有 DNA 片段均以 T 碱基结尾。同理, 可分别制备以 A、C、G 结尾的 DNA 片段。经聚丙烯酰胺凝胶电泳和放射自显影推断碱基序列。由于当时技术上的原因, 只用“加法”系统往往不能得到完全正确的结果, 因此又设计了“减法”系统。“减法”系统也是将产物分成 4 份, 每份中加入 3 种 dNTP, 缺少一种 (此处以缺少 dTTP 为例)。合成反应在遇到应该加入 dTTP 的位置时即停止。由此即可得到均以 T 前一个位置结尾的 DNA 片段混合物, 同样可以推断 T 的位置。尽管采取双系统相互验证的测序模式, 但结果依然不尽如人意。由于反应速度上的差异, 有些片段可能多一些, 有些片段可能少一些, 这就导致了重读和漏读现象。加减测序法的可能误差为 $1/50$ ^[6]。

基于加减测序法, Sanger 和 Coulson 在测序体系中引入双脱氧核苷三磷酸 (dideoxynucleoside triphosphate, ddNTP) 作为链终止剂, 并于 1977 年创建了更加快速且准确的双脱氧链末端终止测序法, 又称 Sanger 测序法或酶法。这是 DNA 测序进程中的重大突破。

其原理是以待测单链 DNA 为模板, 以 dNTP 为底物, 在寡核苷酸引物的引导下依据碱基互补配对原则, 利用 DNA 聚合酶催化 dNTP 的 $5'$ 磷酸基团与引物的 $3'$ -OH 末端生成 $3', 5'$ -磷酸二酯键。通过磷酸二酯键的形成, 新的互补 DNA 链得以从 $5' \rightarrow 3'$ 方向不断延伸。ddNTP 作为链终止剂通过 $5'$ 三磷酸基团渗入正在延伸的 DNA 链中, 由于较 dNTP 在 $3'$ 位置缺少一个羟基 (图 1-3), 而不能与后续的 dNTP 形成 $3', 5'$ -磷酸二酯键, 进而终止 DNA 链的延伸^[11]。最终产生 4 组分别终止于 $3'$ 末端每一个 A、T、C、G 位置上的 DNA 片段混合物。由于对引物进行了事先标记 (后来改进为对 ddNTP 进行荧光素标记), 可以实现 4 个测序反应在同一反应管中进行^[12], 所以通过高分辨变性聚丙烯酰胺凝胶电泳和放射自显影技术即可直接读出新合成链的 DNA 碱基顺序, 进而推断互补待测 DNA 链的碱基序列。Sanger 双脱氧链终止测序法的原理如图 1-4^[7] 所示。

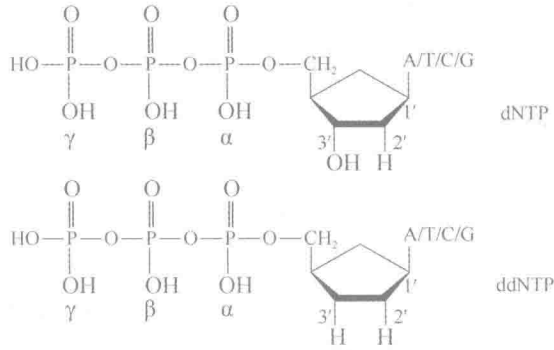


图 1-3 dNTP 和 ddNTP 分子结构式

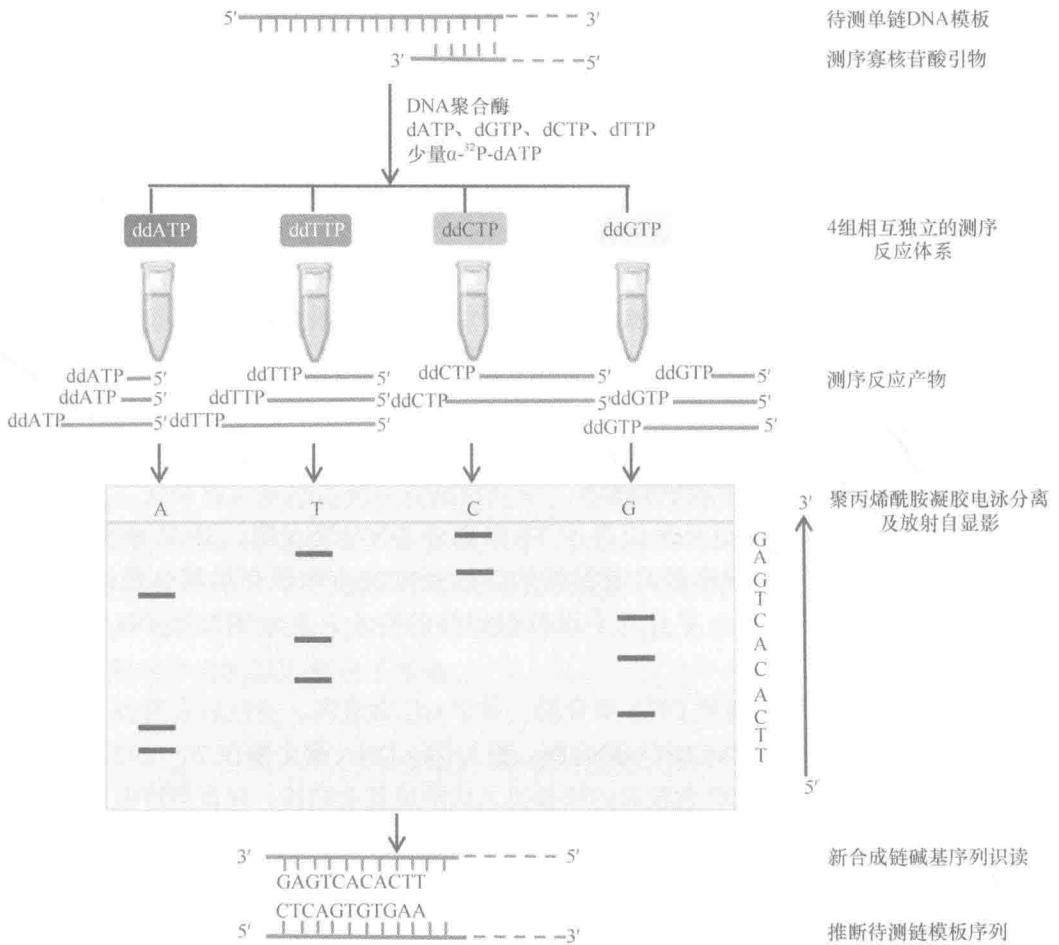


图 1-4 Sanger 双脱氧链终止测序法原理

测序体系

1. 待测 DNA 模板 纯化的单链 DNA 和经热变性或碱变性的双链 DNA 均可以作为测序模板。