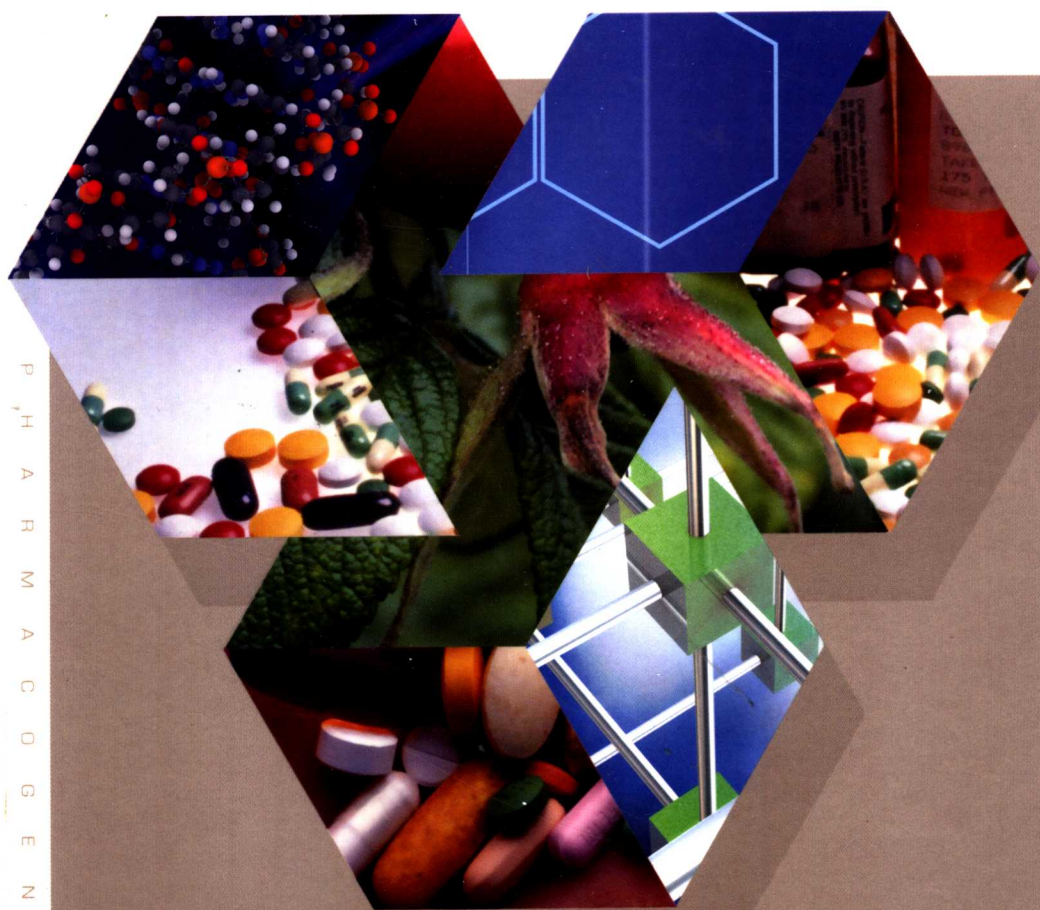


药物基因组学

PHARMACOGENOMICS

主编 / 姜远英



P
H
A
R
M
A
C
O
G
E
N
O
M
I
C
S

人民卫生出版社
People's Medical Publishing House

药物基因组学

PHARMACOGENOMICS

主 编 / 姜远英

副主编 / 郭葆玉

主 审 / 芮耀诚 龙 焜

编者(按姓氏笔画排列)

王 栋	毛俊琴	冯 皓	厉建中	刘 宏
肖振宇	芮耀诚	李铁军	邱 磊	张俊平
张 冉	苗 红	周 斌	姜远英	徐 铮
殷 明	郭葆玉	郭满盈	高平挥	黄 矛
曹永兵	曹颖英			

人 民 卫 生 出 版 社

图书在版编目(CIP)数据

药物基因组学/姜远英主编. —北京:人民卫生出版社,
2006. 4

ISBN 7-117-07390-X

I. 药… II. 姜… III. 药物-应用-基因组-基
因治疗 IV. R394

中国版本图书馆 CIP 数据核字(2005)第 157525 号

药物基因组学

主 编:姜 远 英

出版发行:人民卫生出版社(中继线 67616688)

地 址:(100078)北京市丰台区方庄芳群园 3 区 3 号楼

网 址:[http://www. pmph. com](http://www.pmph.com)

E - mail: pmph @ pmph. com

邮购电话:010-67605754

印 刷:北京人卫印刷厂

经 销:新华书店

开 本:787×1092 1/16 印张:26.25

字 数:620 千字

版 次:2006 年 4 月第 1 版 2006 年 4 月第 1 版第 1 次印刷

标准书号:ISBN 7-117-07390-X/R·7391

定 价:58.00 元

著作权所有,请勿擅自用本书制作各类出版物,违者必究
(凡属印装质量问题请与本社销售部联系退换)

前 言

药物基因组学研究人类基因组信息与药物反应之间的关系，利用基因组学信息解答不同个体对同一药物反应上存在差异的原因。药物基因组学是药物治疗学向个体化合理用药发展的结果，也是药理学在分子基因水平上阐明药物作用机制、药物代谢转运机制、药物不良反应发生机制的结果。但是由于功能基因组学的研究还刚起步，众多基因的功能及他们之间的网络关系还远没有阐明，因此，药物基因组学的知识体系还很不完善，要完成药物基因组学的目标还有很长的路要走。相信，随着科学技术的快速发展，药物基因组学的知识体系会很快完善起来，并将对人类产生巨大的贡献。作者本着作颗铺路石的心愿，在书中主要介绍了药物基因组学的基本概念、基本方法、基本知识，试图探索药物基因组学与相关学科（如遗传药理学、生物信息学、毒理基因组学等）的交叉关系，使读者了解药物基因组学的基本涵义和重要意义，引起人们对药物基因组学研究的兴趣和关注。本书在编写过程中，得到了第二军医大学同行的大力支持，在此表示感谢。另外要感谢龙焜教授、郑钦岳教授提出的宝贵意见，感谢上海长海医院孙华君副主任药师为本书编写中英文索引。由于编者水平所限，书中缺点错误在所难免，敬请广大读者批评指正。

姜 远 英

2006 年 1 月

目 录

第一篇 总 论

第一章 药物基因组学概论	姜远英 (1)
一、基本概念	(1)
二、发展历史背景	(6)
三、研究内容和任务	(8)
四、有关的法律、伦理问题	(9)
五、药物治疗学	(11)
第二章 人类基因组学的基本概念	姜远英 (14)
一、基因的基本结构	(15)
二、基因的基本特征	(18)
三、人类基因组结构	(30)
四、遗传多态性	(34)
五、基因芯片	(39)
第三章 蛋白质组学概论	徐 锋 姜远英 (43)
一、蛋白质组学研究的背景与概念	(43)
二、蛋白质组学的研究方法	(45)
三、药物蛋白质组学的研究趋势和展望	(58)
四、蛋白质的立体结构测定和转换	(59)
第四章 药物基因组学与新药研究	曹永兵 姜远英 (71)
第一节 新药研究的发展趋势	(71)
一、新药研究的历史	(71)
二、新药发现途径	(72)
三、新药研究过程	(77)
四、新药研究现状分析	(80)
五、新药研究面临的挑战	(84)
第二节 药物基因组学与新药研究	(86)

一、药物基因组学为新药研究提供新策略	(86)
二、药物效应基因及其分类	(89)
三、发现药物效应基因的研究方法	(91)
四、药物基因组学在新药研究中的应用	(94)
五、后基因组时代的新药研究	(97)

第二篇 药物基因组学的研究方法

第五章 基因表达连续分析	冯 皓 郭葆玉 (103)
一、概述	(103)
二、SAGE 步骤过程	(104)
三、SAGE 的生物信息学	(106)
四、SAGE 应用	(108)
五、总结	(110)
第六章 单核苷酸多态性的分析技术	郭葆玉 邱 磊 (111)
一、前言	(111)
二、SNPs 分析技术: 生物化学、读取及其平台	(111)
三、分析生物化学	(112)
四、结论	(114)
第七章 多重荧光微测序标记药物代谢酶基因	郭葆玉 王 栋 (116)
一、引言	(116)
二、多重荧光微测序的原理	(117)
三、CYP2D6 和 NAT2 多态性分析	(117)
四、基因型分析能力	(119)
五、分离方法的选择	(119)
六、展望	(120)
第八章 复合基因型的特异质谱法测定	郭葆玉 张 冉 (121)
一、概述	(121)
二、应用 MALDI-TOF MS 分析 DNA	(122)
三、高效的 SNP 分析模式	(124)
四、未来的方向	(126)
第九章 疾病基因定位克隆与鉴定	郭葆玉 郭满盈 (129)
一、前言	(129)
二、候选基因方法	(130)
三、定位克隆法	(130)

四、物理作图	(134)
五、定位克隆使用计量算法和数据库法	(136)
六、结论	(137)
第十章 疾病图谱	郭葆玉 苗 红 邱 磊 (138)
一、前言	(138)
二、图谱绘制策略、DNA 多态性和种群参数	(139)
三、连锁图谱	(141)
四、连锁图谱(连接不平衡)绘制	(144)
五、疾病的动物模型	(146)
六、决定疾病的基因组成	(146)
七、展望	(147)
第十一章 生物芯片在药物基因组学研究中的应用	郭葆玉 曹颖璞 厉建中 (150)
前言	(150)
第一节 生物芯片技术	(151)
一、微阵列制备技术	(151)
二、生物芯片与微阵列技术的类型	(152)
第二节 微阵列技术与生物传感器技术的联合应用	(153)
一、DNA 杂交生物感应器芯片	(153)
二、PNA 生物反应器	(154)
三、细胞与基因生物感应器	(154)
四、光学生物传感器	(154)
五、表面离子共振生物感应器芯片	(154)
第三节 微阵列技术在药物基因组学中的应用	(155)
一、DNA 芯片技术在人类基因组图谱上的应用	(155)
二、微阵列技术在基因表达及发现中的应用	(156)
三、寡核苷酸探针在基因表达及发现中的应用	(156)
四、突变及其多态性的检测	(156)
五、基因文库图谱	(156)
六、微阵列技术在药物研发中的应用	(156)
七、DNA 芯片在分子肿瘤学中的应用	(158)
八、基因表达在炎性疾病中的应用	(158)
九、DNA 芯片技术在药物安全性方面的应用	(158)
第四节 结束语及专家观点	(158)
一、DNA 芯片与凝胶方法	(158)
二、微阵列与生物信息学	(158)
三、微阵列分析技术的标准化	(159)
四、微阵列技术的成就	(159)

五、生物芯片技术的缺陷	(159)
六、生物芯片技术的商业潜力	(159)
七、微阵列技术在药物基因组学及人类健康方面的应用前景	(160)

第三篇 常见疾病的药物基因组学

第十二章 心血管疾病的药物基因组学	芮耀诚 (163)
-------------------------	-----------

第一节 心衰的药物基因组学	(163)
---------------------	-------

一、扩张型心肌病的家族性和单基因形式	(163)
--------------------------	-------

二、非单基因形式的收缩功能紊乱的遗传因素	(166)
----------------------------	-------

三、药物基因组学	(167)
----------------	-------

第二节 冠心病和心肌梗死的遗传变异和药物基因组学	(169)
--------------------------------	-------

一、流行病学及遗传学	(169)
------------------	-------

二、脂质代谢	(170)
--------------	-------

三、凝血和纤维蛋白溶解	(172)
-------------------	-------

四、循环和血管自稳平衡	(173)
-------------------	-------

五、代谢因素	(174)
--------------	-------

六、与环境因素的相互作用	(174)
--------------------	-------

七、与年龄和性别因素的相互作用	(174)
-----------------------	-------

第三节 高血压基因研究与药物基因组学	(175)
--------------------------	-------

一、高血压基因研究	(175)
-----------------	-------

二、高血压的药物基因组学	(176)
--------------------	-------

第四节 心律失常基因研究与药物基因组学	(178)
---------------------------	-------

一、先天性长 QT 综合征	(178)
---------------------	-------

二、致心律失常性右心室心肌病	(179)
----------------------	-------

第十三章 神经系统疾病	殷 明 (181)
-------------------	-----------

一、精神分裂症的药物基因组学	(182)
----------------------	-------

二、情感性疾病的药物基因组学	(186)
----------------------	-------

三、阿尔茨海默病的药物基因组学	(187)
-----------------------	-------

四、偏头痛的药物基因组学	(188)
--------------------	-------

五、神经离子通道病的药物基因组学	(189)
------------------------	-------

六、帕金森病的药物基因组学	(189)
---------------------	-------

七、亨廷顿病的药物基因组学	(190)
---------------------	-------

八、肌萎缩性侧索硬化症的药物基因组学	(190)
--------------------------	-------

九、以后的研究	(190)
---------------	-------

第十四章 内分泌系统疾病	李铁军 毛俊琴 徐添颖 (193)
--------------------	-------------------

第一节 糖尿病相关基因的研究	(193)
----------------------	-------



一、引言	(194)
二、胰岛素受体结构	(195)
三、胰岛素受体的亚型	(196)
四、胰岛素/IGF- I 杂合受体	(197)
五、胰岛素受体多态性	(199)
六、胰岛素受体突变和 II 型糖尿病的突变体	(199)
七、结论	(202)
八、专家意见	(202)
第二节 雄激素类和前列腺疾病的药物基因组学	(203)
一、引言	(203)
二、类固醇 5A-还原酶 (SRD5A2)	(204)
三、雄激素受体	(205)
四、专家意见	(208)
第三节 骨质疏松症相关基因的研究	(208)
一、骨质疏松症的流行病学	(208)
二、骨质疏松症的遗传学	(208)
三、骨质疏松症的药物基因组学	(209)
四、其他疾病	(210)
第十五章 呼吸、消化系统和口腔疾病的药物基因组学	刘 宏 周 斌 (211)
第一节 哮喘	(211)
一、糖皮质激素	(212)
二、 β_2 AR 激动剂	(212)
三、5-LOX 抑制剂和 LTs 受体拮抗剂	(214)
四、其他哮喘治疗药物	(214)
第二节 消化性溃疡	(215)
一、基因多态性	(215)
二、药物基因组学	(216)
第三节 牙周病	(217)
一、白细胞介素-1 基因簇遗传多态性	(217)
二、其他相关基因遗传多态性	(218)
三、遗传多态性对治疗的影响	(219)
第十六章 肿瘤	张俊平 (222)
一、宿主基因组和肿瘤组织基因组对药物反应的影响	(222)
二、抗肿瘤药物与基因多态性	(224)
三、靶向抗肿瘤药物	(236)
四、遗传标志的分析方法	(238)
五、结语	(240)

第十七章 血液系统疾病	肖振宇 (241)
第一节 白血病的药物基因组学	(241)
一、CYP3A4 基因与表鬼臼毒素	(241)
二、TPMT 与 6-巯基嘌呤	(243)
三、新技术在白血病药物基因组学中的应用	(246)
第二节 恶性淋巴瘤的药物基因组学	(247)
一、化疗药物及其相关蛋白和基因	(247)
二、新技术在恶性淋巴瘤药物基因组学中的应用	(249)
第三节 抗凝血药物的药物基因组学	(250)
一、华法林的临床与代谢	(250)
二、CYP2C9 的遗传多态性	(250)
三、华法林与 CYP2C9	(251)
四、小结	(251)
第四节 药物基因组学在血液系统疾病中的研究现状和前景	(251)

第四篇 遗传药理学

第十八章 遗传药理学概论	黄 矛 姜远英 (253)
一、遗传药理学的起源和发展	(253)
二、遗传药理学的基本概念	(255)
三、遗传药理学的任务和研究内容	(256)
四、遗传药理学研究方法	(258)
第十九章 遗传变异的普遍性	黄 矛 姜远英 (261)
一、表型和基因型	(261)
二、人类基因的多样性	(262)
三、展望	(266)
第二十章 遗传变异与药动学	黄 矛 姜远英 (268)
一、遗传变异与药物吸收	(268)
二、遗传变异与药物分布	(268)
三、遗传变异与药物代谢	(274)
四、遗传变异与药物排泄	(283)
第二十一章 遗传变异与药效学	黄 矛 姜远英 (284)
第一节 受体的遗传多态性	(284)
一、受体的分类和作用机制	(284)
二、受体变异影响药物效应的机制	(286)
三、肾上腺素受体的遗传多态性	(288)

四、阿片受体的遗传多态性	(294)
五、多巴胺受体的遗传多态性	(296)
六、其他受体的遗传多态性	(297)
第二节 常见药物反应的遗传变异	(299)
一、5-氟尿嘧啶反应的遗传变异	(299)
二、香豆素耐受	(301)
三、胰岛素抵抗	(302)
四、葡萄糖-6-磷酸脱氢酶缺陷	(302)
五、恶性高热	(303)
六、血管加压素抵抗	(304)
 第二十二章 遗传变异与药物转运	姜远英 (306)
一、前言	(306)
二、ABC 转运蛋白	(306)
三、有机阴离子转运蛋白	(320)
四、有机阳离子转运蛋白	(321)
五、前景与展望	(323)
 第五篇 毒理基因组学	
第二十三章 毒理基因组学概述	肖振宇 姜远英 (325)
一、毒理基因组学的任务	(325)
二、毒理基因组学和毒理学芯片	(326)
三、毒理基因组学和生物学技术	(327)
四、毒理基因组学和生物信息学	(328)
第二十四章 毒理基因组学在基础研究中的应用	肖振宇 姜远英 (329)
一、基因芯片技术与毒理基因组学	(329)
二、生物学技术在毒理基因组学中的联合应用	(331)
三、毒理基因组学研究中的热点问题	(333)
四、毒理基因组学在毒理机制和预测研究中的挑战和限制	(334)
第二十五章 毒理基因组学与药物不良反应	肖振宇 姜远英 (336)
一、历史回顾	(336)
二、现状	(337)
三、前景与挑战	(339)
第二十六章 毒理基因组学与药物开发	肖振宇 姜远英 (341)
一、毒理基因组学与药物开发	(341)

二、毒理基因组学的具体应用	(342)
三、毒理基因组学的产业化前景	(344)

第六篇 生物信息学

第二十七章 互联网与生物信息学	高平挥 (349)
一、带宽	(349)
二、生物信息学	(349)
三、浏览器	(350)
四、CGI (公共网关接口)	(350)
五、CGI 书写	(351)
六、客户	(351)
七、编译器	(351)
八、计算机操作系统	(351)
九、计算机程序语言	(351)
十、HTML	(351)
十一、互联网连接	(351)
十二、Java 应用程序及程序	(351)
十三、Perl	(352)
十四、服务器	(352)
十五、网站、网页、主页	(352)
十六、TCP/IP	(352)
十七、FTP (文件传输协议)	(352)
第二十八章 生物信息数据库	高平挥 (353)
一、基因组数据库	(353)
二、核酸序列数据库	(356)
三、蛋白质序列数据库	(361)
四、结构数据库	(363)
五、功能数据库	(366)
六、其他数据库	(367)
第二十九章 生物信息软件	高平挥 (368)
一、序列比对程序	(368)
二、多序列比对程序	(370)
三、构建系统树的软件	(371)
四、核酸与蛋白质的结构和功能预测软件	(372)
五、其他用途的生物信息学软件	(378)

第三十章 总结与展望	姜远英 (380)
一、概述	(380)
二、社会问题	(381)
三、伦理问题	(382)
药物基因组学常用词汇对照表	(384)
中文关键词索引	(395)
英文关键词索引	(400)

第一篇

总 论

第一章

药物基因组学概论

一、基本概念

药物基因组学 (pharmacogenomics) 研究人类基因组信息和药物反应的关系, 利用基因组学信息解答不同个体对同一药物反应存在差异的原因。人们早就注意到药物反应存在个体差异, 并关心这种差异的遗传学基础, 但是用基因组信息来解释药物反应的个体差异只是近几年的事; 过去多关注单基因变异对药物反应的影响, 现在更关注多基因变异或基因组变异对药物反应的影响。科学家不但关注人类对药物反应差异的基因组学基础, 还关注地球上所有生物包括低等和高等生物对药物和环境中化学物质反应差异的基因组学基础。人类个体和其他生物体不仅对药物和化学物质, 对其他生物特别是病原体反应的差异也有内在的基因组学基础。

(一) **药物反应** 药物是指用于治疗、预防和诊断人类疾病的化学物质。药物可以是来自于自然界的天然产物, 可以用化学方法制备的化合物, 也可以是用生物工程技术获得的蛋白质或多肽等大分子。现在大多数药物是结构已知的单一物质, 也有许多药物特别是“中成药”是成分复杂的混合物, 具有多靶点复合作用的特点。随着科技的不断发展, 药物的品种正在迅速增多, 这为人们防病治病提供了有利的条件, 但同时也给医药工作者掌握迅速膨胀的药学知识带来了困难。

药物作用于有机体呈现的效应是直接作用与间接作用复合的结果。直接作用是指药物作用于特异性靶点 (包括受体、酶、离子通道等) 所产生的直接效应; 间接作用是指药物对机体产生生理生化效应后及机体对药物转运和代谢过程中, 机体为维持自身内环境的稳定所产生的反馈性、代偿性变化。这种生理生化功能的代偿在整体可能是不同系统 (如生理系统、不同生化代谢通路) 之间的协同变化, 如使用某些降血压药时可引起水钠潴留等不良反应, 长期大剂量使用糖皮质激素可引起肾上腺皮质萎缩等不良反应。在个别细胞可能表现为基因表达的改变甚至可能导致基因突变, 如使用胰岛素治疗糖尿病时, 可引起胰岛素受体的数目下调、胰岛素受体对胰岛素的敏感性下降等不良反

应；长期使用氟康唑治疗深部真菌感染，可引起真菌细胞膜上的“多药抗药蛋白”表达增多，使真菌细胞对氟康唑产生抗药性。有些化疗药物除能对抗病原微生物外，还能对宿主产生影响，这种对宿主的影响也是间接作用。有些药物的作用靶点可能直接就是某些转录因子或基因。药物可作用于基因组，反过来基因组差异也能显著影响药物的作用。

药物反应是药物产生的治疗效应、不良反应和机体转运代谢药物所产生的效应的总和。药物反应不仅与机体的生理状态、病理状态有关，还与机体的遗传特征即基因组多态性有关。遗传变异与药物反应的关系十分复杂，有些遗传变异可能不影响药物的反应；有些遗传变异若发生在药物作用靶点、药物代谢酶或药物转运载体，则可能对药物反应产生显著的影响。有些遗传变异可能使药物效应增强，有些可能使药物效应减弱；有些遗传变异可能使不良反应增多或加重，有些可能使不良反应减少或减轻；有些遗传变异通过影响药物的转运或代谢来改变药物的疗效或不良反应。总之，机体表现出来的是所有遗传变异复合作用的结果。基因组多态性在全基因组水平上考察个体的遗传特征，就是用基因组学的整体观和生物信息学强大的信息处理能力来预测或解释个体可能产生的药物反应，因此产生了新的药物基因组学。

(二) 基因组信息 基因组 (genome) 是指生物体单倍体细胞中一套完整的遗传物质，包括所有的基因和基因间区域 (即所有的编码区和非编码区)。所有生物体的基因组都是由核苷酸构成的，核苷酸序列反映基因组的真实面貌。基因组序列的测定需首先建立一个基因组图 (genome map)，基因组图能通过显示基因以及其他明显特征序列在染色体上的位置为测序提供指导。基因组作图的方法分为两类。

一类是遗传作图 (genetic mapping)，即应用遗传学技术构建能显示基因以及其他特征序列在基因组中位置的图。遗传学技术包括杂交育种实验，对人类则是检查家族史。基因是用于遗传作图非常有用的标记，但不是很理想，因为在人类基因组中，多数基因是散在分布的，而且基因内部有许多大的内含子将其断开。除基因外用于作图的 DNA 特征序列称为 DNA 标记，有三种特征序列可作为 DNA 标记，分别是限制性片段长度多态性 (restriction fragment length polymorphism, RFLP)、简单序列 (包括小卫星、微卫星或称简单串联重复) 长度多态性 (simple sequence length polymorphism, SSLP) 和单核苷酸多态性 (single nucleotide polymorphism, SNP)。遗传作图的方法主要是以遗传连锁 (genetic linkage) 分析为基础。如果一对基因位于同一条染色体上，那么它们就物理地连接在一起，在极少数情况下它们总是一起遗传，这对基因就表现为完全连锁；在多数情况下，位于同一条染色体上的成对基因或者像不同染色体上的基因那样独立遗传 (完全不连锁)，或者只表现为部分连锁 (partial linkage)，有时一起遗传，有时不一起遗传。研究表明，染色体上相邻近的两个基因在减数分裂中由于交换 (或称重组, recombination) 而分开的几率将低于距离较远的两个基因，换句话说，基因间因交换而去连锁的几率与它们在染色体上的距离成正比。因而重组频率 (recombination frequency) 就成为衡量两个基因间距离的单位，计算出不同基因对间的重组频率，就可以构建出显示基因在染色体上相对位置的图。

另一类是物理作图 (physical mapping)，即应用杂交分析和 PCR 等分子生物学技术来直接检测 DNA 分子，从而构建能显示包括基因在内的、有显著特征的序列在染色体

上的位置图,所有基因组序列测定都要以物理图谱作为基础,物理作图能弥补遗传作图的局限性。最重要的物理作图方法有以下三类。①限制酶作图(restriction mapping):即在DNA分子上定位出限制性内切核酸酶酶切点的相对位置,该作图法快速、简便,能提供详细定位信息,但它不能用于大基因组,是对小基因组进行物理作图的一种有效方法。②荧光原位杂交(fluorescent in situ hybridization, FISH):是将DNA分子用荧光标记后与完整染色体杂交来确定标记DNA分子在染色体上或伸展的DNA分子上的位置, FISH主要用于正在分裂的染色体或大的基因组作图,过去由于染色体高度凝缩,只能进行低分辨率作图,两个标记至少分隔1Mb才能作为分开的杂交信号被分辨出来。这种分辨率不足以构建有效的染色体图谱,只能用于初步确定新标记在染色体的大概位置,为其他精细作图做准备。现在使用改进的方法如选用较为伸展的染色体,可以得到高分辨率的图谱,但FISH难以操作,数据积累慢,一次实验仅能获得不超过3~4个标记的图谱位点。③序列标志位点(sequence tagged sites, STS)作图:是目前最有用的基因组染色体物理图,STS在染色体上的排列顺序就构成了染色体的STS图谱。STS是一段短的DNA序列,通常长度在100~500bp,易于识别,仅存在于待研究的染色体或基因组中。一个DNA序列要成为STS,必须满足两个前提。首先它的序列必须是已知的,以便使用PCR方法检测STS在不同DNA片段中是否存在。其次STS必须在待研究的染色体上有唯一的定位,或STS在整个基因组中具有唯一的定位位点。STS作图过程中所必须的另一要素是有一群数量足够多的能覆盖全部待研究染色体或基因组的DNA片段,这样的片段群有时也称为作图试剂(mapping reagent)。获得一份遗传标记密度至少是每100kb含有一个标记的遗传和物理综合图谱是人类基因组计划进行大规模测序的前提,人类基因组计划中的物理作图阶段于1996年完成。

无论测序还是物理作图都需将基因组或分离的染色体断裂成DNA片段,并将每个片段克隆到高容量载体中,高容量载体能够容纳大约平均长度为1~2Mb的DNA片段,这样可以产生一个克隆文库。单独的克隆可提供测序的DNA,其中的STS数据可用来构建物理图谱,也可用来建立克隆重叠群(clone contig,即含有重叠DNA片段的克隆群)。组装好的重叠克隆可作为长的连续DNA测序的基本材料,借助于已确定的基因组中特征序列(如STS)的位置(物理图谱),我们可以分析和解释从克隆重叠群中获得的序列资料。最精细的基因组图是核苷酸顺序图,位标之间的距离或图距以碱基对(base pair, bp)表示,最粗略的基因组图是染色体组图。人类基因组计划的核心内容是绘制人类基因组的遗传图、物理图、序列图、转录图甚至变异图。其中最核心最本质的就是序列图的绘制。

完成基因组作图后,可以开始对DNA分子进行测序。测序方法虽然很重要,但有局限性,即使用最精细的技术也很难在一次测序实验中获得一个大于750bp的序列,这意味着一个长的DNA分子序列只能通过将一系列短序列拼接起来而得到。常用的一种方法为霰弹法(shotgun method),即将长DNA分子打断为片段,测得每个片段的序列,然后用计算机搜索重叠区并构建主序列。定向霰弹法是用基因组图上的显著特征作为界标,指导将应用霰弹法获得的大量短序列拼接成主序列。基因组图还可用于检查对重复区域序列的拼接有无错误。从理论上说,整个人类基因组都可以通过这种方法来进行序列测定。

获得的基因组序列是稳定的、静态的。在生物个体的不同生长发育阶段,不同组织细胞的基因组是相同的,差别在于基因表达谱。基因表达谱是动态的、变化的。生物个体在不同的生理、病理、药理状态,不同的生长发育时期,不同的组织细胞,其基因表达谱有很大差异。在某一状态下,有些基因表达量高,有些基因表达量低,有些基因有表达,有些基因不表达。研究这些表达差异及其临床意义,有助于确定基因的功能,为新药创制提供新的靶标。这就是人类基因组计划绘制转录图的任务和目的。

基因组测序获得的只是一个“标准人”的全部 DNA 序列,其实不同个体的基因组并不完全相同,这在科学上叫做“基因组的多态性”。多态性表现在 DNA 序列上,统计表明,任意两个人之间的 DNA 核苷酸差异约占基因组的 0.1%。按人类基因组共有 30 亿对碱基计算,将有 300 万个核苷酸位点的不同。就是这基因组中千分之一的差异,决定了人类的遗传多样性。如有的高、有的矮、有的胖、有的瘦、有的是黄皮肤、有的是白皮肤等。在医学上人们更关注的问题是有的人容易生病,而有的人却对疾病的免疫能力特别高;有些药物,有的人用了有效,有的人用了却无效,有的人用了易产生不良反应,有的人用了不产生不良反应。只有从不同个体 DNA 序列的差异上阐明人类基因组的多态性,才能真正了解与疾病特别是多基因疾病及药物反应有关的遗传机制,同时深入准确地了解人类起源、进化和迁徙过程中的 DNA 序列变化。

DNA 或基因多态性可通过直接测序来检出,也有许多其他检测序列变异的方法在广泛应用。

限制性片段长度多态性 (RFLP): 各种限制性内切酶能识别特定的碱基序列,并能将其切开。碱基的变异可能导致识别序列消失,使内切酶不能切割或出现新的切点,而引起酶切后 DNA 片段长度的差异。这种由于内切核酸酶酶切位点变化所导致的 DNA 片段长度的差异称为 RFLP。RFLP 反映了个体间 DNA 序列的变异。传统上 RFLP 的检出需先通过电泳将长度不同的片段分开,并印迹于硝酸纤维素膜上,然后再与相应的探针杂交,这称为 DNA 印迹杂交。

小卫星 DNA (minisatellite DNA) 和微卫星 DNA (microsatellite DNA) 的多态性: 小卫星 DNA 通常不超过 20kb, 由长 15 ~ 65bp 的基本单位重复串联而成,其重复次数在人群中是高度变异的,这决定了小卫星 DNA 区长度的高度变异性;微卫星 DNA 也是由重复序列串联构成,但它们的基本单位只有 1 ~ 8bp, 如 $(TA)_n$ 、 $(CGG)_n$ 等,且通常只重复 10 ~ 60 次。已定位的微卫星位点多达 8000 个,重复次数在个体间呈高度变异性,因而在遗传分析和诊断中应用广泛。小卫星 DNA 和微卫星 DNA 为数众多,分布广泛,当用限制性内切酶切割它们的所在区域时,只要酶切位点不在重复区内,就可以得到各种长度不同的片段,通过电泳可以检测和分析这些片段。

单核苷酸多态性 (SNP): 在不同个体基因组的差异当中,大多数属于 DNA 链上的单个核苷酸的不同,科学家称之为 SNP。SNP 是广泛分布于人类基因组中的稳定多态位点,最大程度地代表了不同个体间的遗传差异。在不同个体的同一染色体的同一位置的核苷酸序列中,可能绝大多数核苷酸的排列顺序一致,只有一个核苷酸不同,这种现象就是 SNP,探讨以 SNP 为遗传标志的基因分型与药物反应的关系,是药物基因组学的重要研究内容。原则上任何用于检测单碱基突变或基因多态性的技术都可用于 SNP 的识别和检出,但当前在人类基因组中搜寻 SNP 普遍采用的策略是对已定位的 STS 和表