遗传 HEREDITAS (Beijing) 12 (2): 46—48 1990

讲座

医学分子遗传学

第八讲 重组 DNA 技术与基因定位

薛京伦 俞民澍

(复旦大学遗传学研究所，上海)

人体基因组是一个十分复杂的结构，每单倍体基因组大约有 3×10^9bp 组成，分成 22 条常染色体和 1 条性染色体。据估计，在人体基因组上存在着大约 50,000—100,000 个结构基因，平均每分摩 (cM) 染色体上分布着 20 个结构基因。人体基因组大约有 33 cM，由此推算出人体基因组上含有约 66,000 个结构基因。现在已经明确地鉴定出了 1,600 多个人体基因，其中约有 1,000 多个基因定位在特定的染色体上。基因定位是研究遗传性疾病和肿瘤发病机制及进行基因诊断的基础。因此，基因定位研究的迅速发展，对于揭示遗传病和肿瘤发病机制的本质，对于预防和治疗遗传性疾病和肿瘤都将起到重要的促进和推动作用。

用体细胞遗传学技术进行基因定位的方法都需要通过检测细胞内的基因产物，但有相当多的基因，如调节基因、某些疾病基因等并没有发现基因产物。另外，在应用杂种细胞进行基因定位时，一般采用的人体亲本细胞都是成纤维细胞或淋巴细胞，但是有许多基因在这两种细胞中并不能表达。由于这些因素的影响，许多人体基因在单纯应用体细胞遗传学技术的条件下不能得到定位。

随着重组 DNA 技术的建立和发展，克隆了许多人体基因，并且基因的结构和功能得到详细研究，发现了一大批存在于人体基因组上的 DNA 随机片段和限制性多态位点。由于这些研究成果的引入，使基因定位研究不再依赖于培养细胞内基因产物的检测，而只需要通过分子杂交方法直接在染色体 DNA 水平上进行定位。但是，必须强调的是，应用重组 DNA 技术进行基因定位仍然需要以体细胞遗传学技术作为基础。因此，正是由于这两项技术的有机结合物巧妙应用，使定位基因的数目、定位的精确程度、准确性和效率都大大提高。

克隆基因的定位 克隆基因定位法就是采用已被克隆基因的 cDNA 探针与保留在杂种细胞内的人染色体 DNA 顺序根据两者的同源性进行分子杂交，以确定克隆基因究竟位于哪条染色体上。但是，必须注意一点，即：用这种方法进行定位时，所采用的杂交探针必须是高度专一性的，即只能与人体 DNA 中的相应基因顺序杂交，而不能与杂种细胞内的小鼠或中国仓鼠 DNA 顺序杂交。用这一方法进行基因定位时，首先是选择合适的限制性内切酶，将人-鼠杂种细胞内的基因组顺序切割成不同大小的片段。由于绝大多数动物基因的 DNA 顺序都具有不同程度的保守性，因此必须设法将人和鼠的 DNA 区分开来。一般都是选择一种特定的限制性内切酶，用这一酶切割后的人体细胞 DNA 和小鼠或中国仓鼠细胞 DNA 再与特定的克隆基因 cDNA 探针杂交，它们各自所产生的酶切带型大小是不同的。用同一酶切割杂种细胞 DNA 顺序，与克隆基因探针进行杂交，凡出现杂交带型的杂种细胞，表明不仅含有鼠的基因顺序，而且含有人的基因顺序，即证明在这一杂种细胞内，必然存在带有该基因的特定人体染色体。相反，凡不出现人体基因杂交带型的杂种细胞，即表明不含有带有这一基因的特定染色体。

如需要定位人体清蛋白基因，需应用人体清蛋白基因 cDNA 作为探针，先分别与经 Hind III 酶切后的人体细胞和 CHO 细胞（均为杂种细胞的亲本）杂交。杂交后的人体细胞 DNA 显示 6.8kb 带型。CHO 细胞 DNA 则显示 3.5 kb 带型。在含有人体 4 号染色体的人-CHO 杂种细胞中，6.8 kb 和 3.5 kb 带型都出现，这一杂种细胞即称为阳性杂种细胞(+)，而不含有人体 4 号染色体的杂种细胞中，只显示 3.5 kb 带型，即为阴性杂种细胞(−)。相似的结果也可见于用 EcoRI 酶切后的结果。由于人体清蛋白基因 cDNA 探针的特异性 Hind III 或 EcoRI 杂交带只出现在含有人体 4 号染色体的杂种细胞，所以将这一基因定位在人的 4 号染色体上。

这一方法的基本原理与用体细胞遗传学进行染色体定位的方法相似。某一人体基因的杂交带型与杂种细胞中所保留的某一人体染色体的一致出现率最高，这一基因就必然位于该条染色体上。如采用缺失 4 号染色体的杂种细胞，我们将 ADH 基因 I 型定位在 4q 21 位置上。

人体肿瘤基因的定位大多采用基因转染技术获得细胞转化基因。克隆后制成专性探针，然后按与上述类似的办法，定位在特定的人染色体上。如 c-sis, c-fes, c-myb, c-mos, c-abl, c-src, c-kl-ras, c-Ha-ras

和 c-myc 等致癌基因，均系采用这一方法定位出的。

随机 DNA 片段的定位　在人体基因上存在着大量的随机片段，这些片段虽然是一些未知功能的片段，但它们都来自于基因组内不同的特定位置上，可用来作为人体基因组内特定区域的遗传标记。因此，分离和鉴定大量的 DNA 片段，并将它们定位在特定的染色体上，对于进行遗传学分析和基因定位具有很大用途。

这一定位方法通常是先用限制性内切酶将来自人体细胞的全部 DNA 切割成大约 15—20 kb 大小的片段，然后插入 λ 噬菌体中，构建成人体基因组文库。由于在人体基因组中存在着大量散在分布的重复顺序。因此在人体基因文库中插入 λ 噬菌体内的随机片段既包括单一顺序又包括重复顺序。而要将随机 DNA 片段作为遗传标记，必须是单一顺序。故应从插入片段中去除重复顺序部分，只采用单一顺序。现已知道，在人体基因组文库中，只有大约 1% 的重组噬菌体中含有单一顺序。这批噬菌体的鉴定方法一般都采用人的重复顺序作为检测探针，凡不能与这一探针杂交的重组子即含有单一顺序插入片段。在获得单一顺序片段后，按照克隆基因定位相同的方法，利用杂种细胞，将它们分别定位在特定的人染色体上。如果能获得合适的具有染色体缺失的杂种细胞，还能将这些片段定位在染色体的具体位置上。

专性染色体 DNA 片段的定位　除了能够从人体基因组文库中分离出随机 DNA 片段定位在特定染色体上外，还可以应用分子克隆的方法，从某一特定的染色体上分离特定的片段，然后进行定位。这种定位方法称为专性染色体 DNA 片段的定位。专性染色体 DNA 片段的获得有两种方法：(1)首先构建只含有单条人染色体的杂种细胞，从中提取出 DNA，将人和鼠的 DNA 区分开来，最后分离出单一顺序片段；(2)采用流式细胞分类仪分离出单条人体染色体，再从中分离出单一顺序片段。从专性染色体中分离得到 DNA 片段后，可采用缺失杂种细胞系，将这些片段定位在这一染色体的特定区域内。

重复顺序与染色体精细结构定位　在人类基因组中含有 70%以上的重复顺序，其中大部分分布于整个基因组的单一顺序之间。在这些重复顺序中，有的重复顺序达到几十万甚至几百万个拷贝，如 Alu 重复顺序家族就由 300,000 个拷贝组成。而另一些类型的重复顺序仅含有几千个或更少的拷贝。当这类几千个拷贝的重复顺序与仅含一条人染色体的杂种细胞DNA进行杂交时，可显示出一些清楚的带型。这些带型代表着染色体的不同区域。因此，这种重复顺序探针可以用来作为特定染色体内不同位点的遗传标记。现已分离出两个重复顺序，可用来作为人体第 12 号和 11 号染色体内不同位点的遗传标记。第一个重复顺序是从人类第 12 号染色体 DNA 文库中分离并克隆的 2.2 kb 重复顺序。这一 2.2 kb 重复顺序在人类基因组中有几千个拷贝。廖英华等证实，当该重复顺序探针与来自含有一条人类第 12 号染色体的杂种细胞 DNA 进行杂交时，可显示出多条带型。当这一重复顺序与其他单一人类染色体杂交时，同样也可发现有特征性的多重带型。用 Rvu II 切割 2.2 kb 片段，可将这一片段分成 1.2，0.6 和 0.4 kb 三个亚片段。如将其中的一个亚片段作为探针，与只含 12 号染色体的杂种细胞 DNA 杂交，可显示出几条较小的带型。应用含 12 号染色体不同缺失的杂种细胞以及 2.2 Kb 重复顺序的不同探针，已经有 5 条带定位在第 12 号染色体上的特定区域内。

第二个重复顺序来源于人体 11 号染色体上的 β-珠蛋白基因复合体。当这一重复顺序探针与内切酶消化后的 11 号染色体 DNA 杂交时，可显示出 24 条带型。用一系列含 11 号染色体不同末端缺失的杂种细胞和这一重复顺序探针杂交，已将 19 条带定位于 11 号染色体的特定区域内。

这些研究证实，特定的重复顺序探针可用于鉴定染色体内的多重位点。选用不同的限制性内切酶和其他重复顺序探针，可鉴定和标记人体基因组各染色体上的更多位点。这一方法对于制作人类染色体详细的基因图，对于检测肿瘤或其他伴有染色体恒定异常遗传病中 DNA 顺序的可能改变，以及对于建立染色体特殊限制性片段详细结构图，都是极为有用的。

原位杂交与基因定位　这是目前应用比较广泛的一种基因定位方法。它的基础是必须分离和克隆获得特定的基因或随机 DNA 片段。将这些克隆片段用同位素 ^{3}H 标记作为探针，直接在玻片上和中期染色体杂交，经过放射自显影就可显示出这一基因在染色体上的杂交位置。统计、分析、汇总各中期细胞内染色体的杂交结果，选出杂交最集中的染色体区域，便是该基因在染色体上的具体位置。采用这一方法，基因定位一直可以达到染色体亚带 (subband) 水平。

基因定位的最终目的是为了构建完整的人类基因图。所谓基因图是指将 24 条人染色体上所有的基因按其所在具体位置、次序和间隔距离详细地排列而成的图，故又称染色体连锁图。构建详细的人类基因图对于揭示基因结构和组织以及对于遗传病进行产前诊断都具有重要意义。我们前面所介绍的基因定位方法只是确定了每个人体基因在染色体上的具体位置，但作为一个完整的基因图，还必须确定各个基因在染色体上的先后次序和间隔距离。

用 DNA 标记建立人类基因图　应用限制性内切酶直接检测 DNA 顺序多态性的遗传标记系统，使人们有可能建立起详细的人类基因图。用克隆化的 DNA 片段所确定的新遗传座位代表着已知的特定基

因或功能未知的座位。应用一般的基因定位方法可将这些片段定位在染色体的特定区域，而通过家系研究方法，则能估计出它们在染色体上的连锁距离和排列次序。据估计，人类基因组的大小为 33 M，如果标记座位的间距为 0.4 M，新座位就可被确定。说明为了标记整个人类基因组、并高度可靠地定位任何新标记，可能需要 80—100 个均匀分布的遗传标记座位。迄今已报道了三百多个多态性 DNA 标记，虽然所报道的大多数标记仅有两个等位片段，但还是有一些复等位片段，这是由于单个碱基改变影响了限制性位点所致。由于频率较高的杂合性为连锁研究所必需，因此，多态性座位所具有的等位片段的数量和频率决定了它作为遗传标记的有用程度。

来自人类的连锁数据通常都用最大似然分析法估算。该法认为，两个座位间重组距离 r 的估计值就是给出了所得数据之最大概率的数值。人类遗传学的经典问题之一就是确定两个座位是否连锁，连锁的定量表达为数据取最大似然值 r 的概率与 $r = 0.5$（标记座位不连锁）的概率之比。这个比值通常用对数来表示，即连锁的 Lod 指标。Lod 指标≥3 时，表示连锁与非连锁的相对可能性之比≥1,000:1，从而认为两个座位是连锁的。多世代的家系是研究人类连锁的最有效结构。这种多世代家系常能提供双亲等位基因的分布情况。应用单个家系还能使所涉及的基因为一个以上座位的概率达到最小。这对遗传病研究来说是很重要的。收集来自大量家系的数据资料，能够提供更多的染色体信息，以便确定父母和祖父母的基因型，并为双亲等位基因分布的概率估计提供有力的基础。White 等用连锁分析的方法确定了一个家系（K 1085）内 206 个个体的 4 个座位的基因型。这 4 个基因座位包括 β-血红蛋白基因座位（用 B 表示），$D_{11}S_{12}$ 座位（用 A 表示），胰岛素基因座位（用 I 表示）和 Harvey-ras-1 癌基因座位（用 H 表示）。他们进一步研究了许多完整的三代家庭，获得了有关这 4 个标记座位更详细的基因型资料，并分析了这 4 个座位的基因次序。最后用最大似然法综合分析了这 4 个标记座位相互之间的重组值及其 Lod 值。根据重组值可知这 4 个基因座位的次序为 B-A-I-H。随着基因型资料的不断积累，这些基因座位就可用于新标记座位的染色体初步定位以及更为精密的制图。但是，用大量的双因子分析来对新座位进行初步定位的速度太慢，故新座位的初步定位可通过比较新座位的祖父母起源和其后代染色体上其他已知座位来完成。其具体思路是：对于后代染色体上的每一个等位基因座位确定其祖父母起源，来自祖母记为 0，来自祖父记为 1。这样，每个座位就由一长串 0 和 1 来表示，称为染色体分布样式（chromosomal distribution pattern, CDP），紧密连锁的座位有着大致相同的 CDP，其差异频率精确地代表了重组频率。将新座位的 CDP 与其他座位的已知 CDP 相比较，就可很容易地确定新座位的位置。上述 4 个基因座位均存在于第 11 号染色体短臂上，它们的 CDP 非常接近，表明它们是紧密连锁的。

遗传病基因的制图　遗传病基因的制图对于遗传病分子机制及其诊断的研究具有重要意义。囊性纤维化病是西方最常见的隐性遗传病之一，它在新生儿中的发病率约为 1/2,000，其基因携带者频率高达 1/20。在 1985 年对此基因还一无所知，只发现 CF 基因与邻近的 PON 基因之间有连锁迹象，其 Lod 记分为 3.38，而 Lod 记分＞3.0 就被认为是连锁。1986 年的数据表明，CF 基因和 PON 基因位于同一条染色体上，且相距 10 cM。如完成 PON 基因的克隆化，就能应用体细胞遗传学方法或原位杂交技术确定 CF 基因所在的染色体位置。然后再筛选染色体特异的文库，以期获得紧密连锁的探针。这应比从 PON 到 CF 进行染色体巡查的方法能更快地检出 CF 基因。因为 10 cM 的遗传距离很长，相当于 10^7bp，单应用染色体巡查的方法效率仍不是很高。1987 年的资料表明，CF 基因位于原癌基因 met 和 D7S8 片段之间，位于 met 基因的上游，位于染色体 7q21.3—q22。目前离克隆 CF 基因的日子已不远了，已有好几个探针可用于囊性纤维瘤的产前诊断和 RFLP 研究，已经找到了一个 4×10^6bp 的区段，其中包含着 CF 位点。成人多囊性肾病（APCKD）是另一种较为多见的遗传病，其基因频率为 1/1,000，呈显性遗传，故其发病率等于基因频率。现已发现，$APCKD_1$ 与 α-珠蛋白基因之间紧密连锁。α-珠蛋白基因位于 16 号染色体短臂上，因此 $APCKD_1$ 基因也必定位于其附近。已证明两者之间的重组距离为 5 cM，相当于 5×10^6bp，Lod 记分高达 25.85。同时发现，$APCKD_1$ 基因与编码磷酸羟基醋酸磷酯酶的基因连锁也非常紧密。

利用这种连锁分析的方法，可以使许多疾病基因得到定位，并能展示它们之间的连锁关系。这些关系的确立对于遗传病的分析和产前诊断是极为有用的。如经过许多人研究，得到了 19 号染色体上若干基因的详细连锁资料，这些基因的顺序是：家族性高胆固醇血症、肽酶 D、强直性肌营养不良、分泌腺、闭止蛋白 C 和 E. Lutheran 血型。应用遗传学分析和分子生物学技术，基因定位的水平已达到 10^5—10^8bp 水平，相当于 0.1—1 cM 的精细程度。

人体基因组 DNA 全顺序测定的设想　利用上述介绍的方法构建完整的人类基因图虽然可以确定每个人体基因在染色体上的位置，相互间的距离以及排列次序，但这仍然属于是在基因整体水平上的制图。一个最为理想的人类基因图应当是建立在核苷酸水平上，这就意味着要测定整个人体的单倍体基因组中 23

（下转第35页）

图1 左为大Y染色体，右为F组染色体

图2 右为正常9号染色体，左为同源9号染色体臂间倒位（C带）

Y长度有种族差异性[4]，本文受检者均为汉族，可排除种族因素的影响。从各组大Y染色体例数分析：实验组18例占72%，对照组1为1例占4.1%。正常男性人群也有大Y，据报道为1.4%—18.6%[3]。本文正常男性组为8.3%，与前人报道相似。由此可见只有实验组大Y例数明显高于正常。再从Y/E指数分析：Unnerus对30例正常男性测得Y/E指数为0.55—0.88，平均0.73[5]。本文48例正常男性为0.58—0.93，平均0.75，对照组1为0.74，两组均与前者一致。但实验组却高于正常而为0.84(0.73—1.0)，与前两组对比有显著差异。根据上述分析我们初步认为暴力行为男性精分患者有大Y倾向。就此提示临床对大Y男性病人多采取预防性措施，防止暴力行为发生可能具有实际意义。至于暴力行为与大Y的关系，有的学者提示暴力行为与大Y有关，但有的学者持否定态度，因此二者关系尚难定论。

（二）暴力行为精分患者四倍体频率高

实验组四倍体例数明显高于对照组(15/25对1/72)，而且15例均为大Y。Amice曾发现携带异染色质变异体的个体表现出有丝分裂不分离的频率增高[2]。我们推测本文四倍体现象是否由于大Y的异染色质区变异所产生的影响。

（三）精分病人9号臂间倒位频率高

两组精分病人中10例有9号臂间倒位，其频率（20.4%）与洪氏等报道的20.8%频率相仿[1]，这进一步说明此种改变对本病发病很可能作为一种遗传因素而起作用。

参考文献

[1] 洪美玲等：1986。中华神经和精神病杂志，19(3)：188—191。

[2] Amice, V.et al.: 1983. *La. Presse Med.*, 12:889.

[3] Soudek, D. et al.: 1973. *Humangenetic*, 18: 285—290.

[4] Grouchy, J. D. et al.: 1977. *Clinical Atlas of Human chromosomes*, John Wiley, New York, pp. 224.

[5] Unnerus, V. et al.: 1967. *Cytogenetics*, 6: 213—227.

（上接第48页）

条染色体的全部核苷酸顺序。这无疑是一项非常艰巨而又浩大的工程。我们知道，人类基因组约为33—35 M，这也就是说整个人体基因组大约有33—35亿碱基对。过去人们一直认为，要完成这样巨大的工程完全是不可能的。最近，一些美国科学家提出，在适当长的时间内是可以完成这项工作的。他们设计了一个研究方案，进行这项研究不需要特定的克隆或限制性图谱分析，而是从随机DNA开始进行排序工作。采用这种方法不能区分重复顺序，但是所有的单一顺序都会被组合起来。

为了表明这项工程的可能性，科学家们还提出了一个具体设想：在全美国建立工作站来完成这一工程，每一个工作站有两个技术人员和4台自动排序机，4台机器有40个通道，而每个通道可以记录750 bp，工作站按每天工作5天，每天两班，每班8小时计算，750 bp×40个通道×4台机器×2个工作班×5＝120万bp/每周。这样每年可测定约6000万bp。为了确保整个基因组都不遗漏地测到，需要将人体基因组的总碱基对数乘以10。运用这一设想，580个工作站在一年内就可完成该项工作。如果这一设想能付诸实现，人类就可以彻底弄清自身基因组结构的全部顺序。这将是人类认识自身的一个重大突破，同时也为揭开生命之谜，为诊断、预防和治疗遗传性疾病与肿瘤铺平了道路。