DOI:10.16016/j.1000-5404.2003.04.033
第25卷第4期 第三军医大学学报 Vol. 25, No. 4
2003年2月 ACTA ACADEMIAE MEDICINAE MILITARIS TERTIAE Feb. 2003

文章编号:1000-5404(2003)04-0349-03

综述

人类疾病相关基因定位研究的策略和方法

Strategies and methods for the study of human disease-related gene position

黄 文,朱佩芳,王正国 (第三军医大学附属大坪医院野战外科研究所第四研究室,重庆 400042)

认识自我,揭开人类的奥秘,追求健康,战胜疾病,是人类有史以来最重要的行为之一。美国科学家于1985年率先提出"人类基因组计划(Human genome project, HGP)"[1],旨在阐明人类基因组3×10^9 bp的序列,发现所有人类基因并阐明其在染色体上的位置,破译人类全部遗传信息,使得人类第一次在分子水平上全面认识自我。而基因鉴定最能反映人类基因组计划的科学、社会与经济意义。通过基因鉴定,弄清各种疾病中遗传因素(即基因)所处的地位和作用不但有利于揭示疾病的病因、理解非遗传的环境因素对人类健康的影响,而且将对诊断学、治疗学及预防医学的领域带来深刻的革命。因此,定位和克隆人类各种"基因病"的致病或相关基因便成为人类基因组学(Genomics)研究的主流方向[2,3]。

1 疾病与基因

基因[4](Gene)一词最早在1909年Wilhelm Johansen采用来描述传递和表达特定的生物形状的可遗传因子,并未涉及任何特殊的遗传理论。基因的遗传学概念是生物的遗传物质及遗传的基本单位,现代分子生物学将基因定义为:合成有功能的蛋白质多肽链或RNA所必需的全部核酸系列(指DNA序列)[5],,及基因是染色体上最小的功能转录单位。人类基因组织人的23对染色体(22对常染色体和X,Y两条性染色体)上全部DNA所携带的遗传信息的总和,约30亿个核苷酸碱基对,约3~4万个基因[6,7],它携带了人类任何一个个体有关生、长、育、老、病、死的全部遗传信息[8,9]。因此,在认识上,"人类疾病都直接或间接的与基因有关",或者说"基因是大多数(不是所有的)人类疾病的主要病因"的观点现已被越来越多的人们所接受。换言之,在"与基因相关"这一认识上,人类各种疾病都可以视为"基因病"。据此广义的概念,目前可将"基因病"分为四大类[10]:①经典的单基因病:这类疾病目前至少已发现6 000余种,其主要病因是一个基因位点上的缺陷等位基因,如:红绿色盲、血友病等。②复杂性状的多基因病:这类疾病的发生涉及两个以上的基因位点,这些基因也许作用近乎等同,即"多因微效(Multi minor genes)",或者可能有主次之分。而各基因之间的相互作用可能调控这些基因表达的环境因子,使其更加复杂多样。人类的大多数疾病都属这一范畴,如:高血压、糖尿病等。③染色体病:人的体细胞中共有23对染色体,每条染色体上均载有基因。因染色体数目或结构的改变所致的疾病称为染色体病,由于染色体病涉及许多基因,故这类疾病常表现为复杂的综合征(Syndrome),如:先天愚型等。④获得性基因病:如病原微生物感染引起的传染病。此类疾病当然不符合原来的"世代传递(遗传)"方式,但大多有一定的遗产基础,或者是病原微生物的基因组和人类基因组相互作用的结果,因此涉及到基因结构与表达的改变。HIV病毒的研究已在一定程度上证明了这一点。

2 疾病相关基因的策略和方法

对"基因病"进行遗传分析的"灵魂"在于紧紧抓住了"表象(表型)"与"本质(基因)"之间存在着的实质性的必然联系[11],即:①有表型(遗传性状或疾病)就可以肯定在基因组中有它的基因;②有基因则在基因组中必有其位置(位点);③该位点与基因组中另一位点(如多态性的遗传标志)之间必有某种联系(相距远则重组率高,近则重组率低或不重组),进而相信可用遗传连锁分析得到二者之间的遗传距离并将其定位。因此,通过疾病的表型来定位、克隆及鉴定疾病相关或易感基因,是最客观、最有效的"解读天书"的手段之一。迄今,以各种技术克隆到的人类基因已达1千多个,这些技术按所采取的策略分为以下三类:定位克隆(Positional cloning)、候选克隆(Candidate cloning)与功能克隆(Functional cloning)。

2.1 定位克隆

定位克隆又称位置克隆,完全是伴随"遗传图"的建立而发展起来的基因克隆的新策略。定位克隆的起点是定位,即建立表现型(包括疾病)与基因组中的一位点(或某一区域)之间的联系,然后根据这一位置信息,应用"物理图"的物理标记经典遗传学信息转变为"物理标记"所代表的明确的基因组区域,再以相关区域的"相连片段群(Contig)"(或建立更为可靠的"相连片段群")来筛选可表达的结构基因,即建立该区域的"转录图"。然后,比较病人与正常人中这些结构基因的区别以确定与疾病相关的改变,最后确定疾病相关基因。定位克隆的最独特的优点,正是抓住了如前所述的遗传分析的灵魂,抓住了"表现型"与"基因"之间的联系。而以遗传分析确定表现型在基因组中定位这一策略,应归功于人类基因组计划所建立的"遗传图"。该策略的具体步骤如下[12]:1)疾病家系的选择与评估:一般来说,疾病家系越大越好。据一般的经验公式(Basic rule of thumb),一个有40多个能提供信息(Informative)减数分裂事件的三代以上的家系,基本上可得到"肯定连锁"的证据(如果使用200个遗传标记,平均距离为20 cm)。大家系的重要优点是可以在很大程度上避免遗传异质性(Heterogeneity)。不管是大家

作者简介:黄 文(1972—),男,重庆市垫江县人,博士研究生,主治医师,主要从事创伤反应差异性遗传机制方面的研究。电话:(023)68757444

收稿日期:2002-05-09;修回日期:2002-10-17

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

系还是小家系，对每一个体的诊断是至关重要的。由于我们只能假定疾病上有两个等位基因，即正常与“有病”两种状况，一个误诊将使整个分析的可靠性受到影响。特别是高血压等“拟数量性状”、要据统一的临床标准详细进行鉴别诊断，还要注意发病的时间性、治疗效应和发病年龄、发病史。临床诊断上的异质性，则要根据已知的鉴别诊断标准尽可能加以区分，特别是小的家系，当然最终肯定“临床异质性”与“位点异质性”，还有待致病定位的最后鉴定。2）遗传标记物的选择：遗传标记的种类现在基本采用简短衔接重复多态性（Short tandem repeat polymorphism，STRP）标记，其突出优点是多态性高，能以PCR大规模操作，DNA用量小。遗传标记的选用可遵循以下步骤[13]：①据现有的同一位点报道区域。②据同一疾病的不同位点报道区域。③模式生物的“同源区域”。④据可能的候选基因。⑤据已知的基因密集区。⑥已知连锁群的代表标记（人类基因组至少有70个连锁群，即相互遗传距离为50 cm的标记）⑦已发现远距离连锁线索的区域中的标记。⑧更为密集的遗传标记进行全基因组扫描。有关这些标记的数据，都可以通过Internet从GDB（Genome data base）上获得。3）靶区域的基因组分析：HGP所提供全基因组物理图（STS图）、部分区域的转录图（EST图）与序列图已使一部分工作简化到前所未有的程度。我们做的工作只是把遗传学分析或细胞遗传学信息转化为物理信息。大多数短串联重复序列（Short tandem repeat，STR）之间的距离都可从GDB网络上查到。在得到候选区域的物理大小后，基因组分析的第二步，便是基因组contig的构建。以YAC构建的contig可以从GDB上查到，可从YAC文库服务中心直接拿到contig中的所有克隆。部分区域的BAC或PI contig也已建好，都可直接向资源服务中心索取或购买。如没有合适的现成的contig，则需自己以STS或其他区段通过筛选对应片段进行构建，或以YAC来筛选BAC克隆。值得注意的是，现有的contig不完全可靠，须逐一验证每个克隆。4）编码序列的筛选：现在已知的蛋白质都是由RNA聚合酶指导合成的带3′多聚A的mRNA的外显子中的编码序列编码的。因此，从候选区域中筛选编码序列是当前位置克隆的主要途径。现在的常规是以候选区域中基因片段来筛选可转录序列。经典的cDNA文库直接筛选的方法简单易行，但较大的DNA片段，如YAC、BAC等用于大规模筛选cDNA常带来严重假阳性及后来的复选问题。在这基础上采用PCR技术建立起来的cDNA selection是目前使用较多的方法。5）致病突变的筛选：致病突变的筛选实质上就是致病基因的筛选。一般可用部分或全长的cDNA来进行。筛选疾病相关突变的技术可采用突变检测的所有技术：如酶切杂交技术、杂质双链技术等等。还要注意检测mRNA的表达水平与大小的变化。突变检测还要注意与“正常”的多态性的区别。因此，要选择多个无亲缘关系的正常个体与患者个体进行比较。可能的突变方式要加以较全面的分析，除了通常的突变类型，如表达与否、点突变与一般重排以外，还要考虑短核苷酸重复、大规模的基因组重排。定位克隆目前广泛的用于许多疾病的遗传研究。如：1999年，Krushkal等[14,15]采用美国Rochester心脏病家系研究的样本，对一组55个家系的血压不一致性同胞对427名成员，用359个微卫星引物进行22对常染色体扫描。将与人群中收缩压变异有关的基因成功定位到染色体上。

2.2 候选克隆

候选克隆策略完全是在HGP的进展、已定位克隆的基因越来越多的情况下派生出来的一种新捷径。候选克隆又分为“定位候选克隆”、“功能克隆”和“定位、功能候选克隆”，即“定位克隆”分别甚至于同时与“定位克隆”和“功能克隆”相结合。“定位候选克隆”是在将疾病基因以连锁分析或染色体分析基本完成后，即从GDB中检索该标志附近与该区域之中所有已知的基因，包括与疾病已有联系或未建立联系的基因，已知的EST，已知的模式生物对应的“同源板块”上的已知基因与EST，用它们直接进行致病改变的筛选。“功能候选克隆”是根据致病基因的可能功能，检测GDB或Gen Bank中的基因功能区域，将含有接近功能域用于致病的改变检测。而“定位、功能候选克隆”途径则是在基本区域确定之后，将该区域中，已知或探测功能可能接近致病基因的基因作为候选者。这一策略可能得到事半功倍的效果。“候选克隆”的另一途径是将功能已知、但与疾病尚未建立联系的基因定位，再在定位区域附近检索已定位的疾病。其著名的成功例子便是α-丙种球蛋白血症[16,17]（α-Gammaglobinemia）。美国UCLA一个小组将早已发现的激酶（Kinase）用FISH定位于Xq21，再从GDB中查到α-Gammaglobinemia也已定位到这一区域，便以该基因筛选病人而确定了该kinase便是α-gammaglobinemia的基因。

2.3 功能克隆

HGP之前的人类基因基本上是按功能克隆的策略成功分离的。人类的性状（包括疾病）都是与蛋白质功能有关。功能克隆便是从蛋白质功能着手的基因分离策略。镰刀细胞贫血的基因鉴定是功能克隆的典型例子。功能克隆的另一常规方法是以某种蛋白质的抗体来筛选可表达的cDNA文库。现在，人们把通过比较mRNA的克隆策略也归类于功能克隆，其根据是蛋白质功能的差异首先反映在同一组织中mRNA的差异。这一归类显然不很全面。这类方法较经典的有“消减杂交（Subtractive hybridization）”，即将正常与异常的同一组织（或不同组织经诱导的不同细胞培养物等）的mRNA或cDNA文库以一定比例进行液相杂交，再以特定的方法克隆其相异的组分，从中筛选代表其相异的克隆。近几年被广泛采用的“差异显示（Differential display）”巧妙地采用了“锚定PCR（Anchor PCR）”技术，能很灵敏的扩增两个mRNA样品中的相异成分。其基本原理也与“消减杂交”的“克隆相异”类似。功能克隆的共同特点应该说是不依赖于基因组信息，其缺点主要是由于大多数症状的生理、生化变化很复杂，对与之有关的蛋白质了解甚少，对于其基因表达的主要组织也知之不多。而且不同组织以至于同一组织的蛋白质与mRNA的“正常”差异很难与特异性差异区别。这就是为什么在“定位克隆”问世之前，只有少数疾病相关基因被克隆。但功能克隆技术成熟、方法直接、费用较低，迄今仍然是克隆基因的首选策略。最近克隆的人的一种Disintegrin metalloproteinase［即TACE（TNF-α-converting enzyme）］[18~20]，便采取了这一途径，但是有了不少新的改进。

3 展望

HGP的发展，尤其遗传图和物理图的制作完成，为疾病相关基因定位、克隆和发展展示了诱人的前景。我国地大人稠、

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

民族众多，由于历史、地理、传统等原因，保存着许多相对隔离群体，这是我们开发人类遗传病相关基因研究一项不可多得的资源优势。充分利用我国丰富的家系资源，迅速开展多基因疾病全基因组扫描和分型的连锁分析研究，对推动我国遗传病研究和提高我国人类遗传学科研水平具有极为重要的现实意义。相信经过不懈的努力，中华民族将来一定能在人类基因组研究中占有自己应有的一席之地。

4 结语

综上所述，要掌握基因克隆技术，其知识需要了解疾病的病理、真核基因的结构、所有已知致病基因的改变方式，其技术则综合了遗传学、分子生物学的几乎所有技术。而其最重要的则是疾病家系的获得。人类基因组计划为基因克隆奠定了基础，而每一基因的成功克隆都进一步丰富了“遗传图”、“物理图”、“转录图”与“序列图”。不久的将来，我们将第一次从分子水平上全面认识自我，实现真正掌握自己命运的宏伟目标。

关键词：基因定位；疾病/病因学
中图法分类号：R394 文献标识码：A

参考文献：

[1] The U S. Human Genome Project: The First Years F Y. 1991 1995. by NTH and DOE. 21.
[2] Marshall E. The genome project′s conscience[J]. Science, 1996, 274(5287): 488—490.
[3] Kahn P. Coming to grips with genes and risk[J]. Science, 1996, 274(5287): 496—498.
[4] R. M 特怀曼[英] 著. 陈 淳, 徐 沁 译. 高级分子生物学要义[M]. 北京: 科学出版社, 2001. 95.
[5] 谷志远. 现代医学分子生物学[M]. 北京: 人民军医出版社, 1998. 159—161.
[6] International Human Genome Sequence Consortium. Initial sequencing and analysis of the human genome[J]. Nature, 2001, 409: 860—921.
[7] Venter J C, Adams M D, Myers E W, *et al*. The sequence of the human genome[J]. Science, 2001, 291(16): 1304—1351.
[8] Lander E S. The new genomics: global views of biology[J]. Science, 1996, 274(5287): 536—539.
[9] Goffeau A, Barrell B G, Bussey H, *et al*. Life with 6000 genes[J]. Science, 1996, 274(5287): 546—547.
[10] 罗会元. 人类遗传学问题与方法[M]. 北京: 人民卫生出版社, 1999. 257—258.
[11] 张安平, 张学军, 朱文元. 疾病相关基因定位的全基因组扫描策略与方法[J]. 疾病控制杂志, 2001, 5(2): 135—138.
[12] 况少青, 傅 刚, 陈 竺. 微卫星标记分析在人类基因组多样性研究中的应用[J]. 国外医学遗传学分册, 1997, 20(1): 5—7.
[13] William D, Hillary K, Stephen E, *et al*. A genetic map of the mouse suitable for typing intraspecific crosses[J]. Genetics, 1992, 131: 423—447.
[14] Krushkal J, Xinopg M, Ferrell R, *et al*. Linkage and association of adrenergic and dopamine receptor genes in the distal portion of the long arm of chromosome 5 with systolic blood pressure variation[J]. Hum Mol Genet, 1999, 7: 1379.
[15] June L D, Yoshihiko K, Simon T, *et al*. A genome-wide search for human type 1 diabetes susceptibility genes[J]. Nature, 1994, 371(8): 130—136.
[16] Pandey J, Chellani H, Garg M, *et al*. Congenital sulfhemoglobin and transient methemoglobinemia secondary to diarrhoea[J]. Indian J Pathol Microbiol, 1996, 39(3): 217—220.
[17] Lin Y C, Sun G H, Lee Y M, *et al*. Cloning and characterization of a complementary DNA encoding a human epididymis-associated disintegrin and metalloprotease 7 protein[J]. Biol Reprod, 2001, 65(3): 944—950.
[18] Bohm B B, Aigner T, Blobel C P, *et al*. Highly enhanced expression of the disintegrin metalloproteinase MDC15 (metargidin) in rheumatoid synovial tissue[J]. Arthritis Rheum, 2001, 44(9): 2046—2054.
[19] Cal S, Arguelles J M, Fernandez P L, *et al*. Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin with a complex structural organization involving multiple thrombospondin 1 repeats[J]. J Biol Chem, 2001, 276(21): 17932—17940.
[20] Gottschalk G, Malberg K, Arndt M, *et al*. Matrix metalloproteinases and TACE play a role in the pathogenesis of endometriosis[J]. Adv Exp Med Biol, 2000, 477: 483—486.

（编辑 郭建秀）

文章编号：1000-5404(2003)04-0351-03

综述

ICOS：一种重要的新型诱导共刺激分子

ICOS: a new important costimulatory molecule

许雪青 综述，白 云 审校 （第三军医大学基础医学部分子遗传学教研室，重庆 400038）

抗原特异性 T 淋巴细胞免疫应答的有效产生及维持需要 2 种信号——MHC-肽提供的特异性的抗原信号，以及由 APC 表面共刺激分子提供的共刺激信号（Costimulatory signal）。随着研究的深入，这一观点逐渐得到广泛认可[1,2]。在 T 细胞活化的不同时空阶段，多种共刺激分子发挥着不同的调控作用。CD28/CTLA-4—B7-1/B7-2 共刺激途径是研究最早最深入的共刺激途径之一，来自该途径的共刺激信号对 T 细胞的活化和耐受发挥着重要调节作用。但众多证据，例如 CD28 缺失的小鼠仍能产生一定的免疫应答[3]，表明 CD28 并不是惟一的共刺激分子。

作者简介：许雪青（1979—），男，湖北省京山县人，硕士，主要从事 T 细胞活化机制方面的研究。电话：(023)68752258

收稿日期：2002-02-03；修回日期：2002-07-09

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net