



医学分子遗传学

第四讲 多基因病分子遗传学

洪贤慷 俞民澍

(复旦大学遗传所, 上海)

多基因病是一种异质性疾病, 是遗传因子和环境因子相互作用的结果。它的性质相当复杂, 其中某些疾病仅仅是由单个基因座位与个别环境因子相互作用所致, 而有些则是由多个基因座位的微小效应叠加在一起相互作用引起。尽管多基因病的遗传因子较为复杂, 但在不少多基因病中往往有一个或少数几个遗传因子的作用比较明显。这就为多基因病的分子遗传学研究提供了重要途径。下面着重介绍有关高脂血症和糖尿病分子遗传学研究内容, 从部分基因座位着手来揭示多基因病的分子病理机制。

高脂血症

高脂血症是一组常见的代谢性疾病, 表现为人体血浆中不同脂质异常堆积, 主要的脂质包括胆固醇和甘油三脂, 它们通过血浆脂蛋白在组织间转运。血浆脂蛋白在脂质代谢中起重要作用。在脂蛋白中脱辅基脂蛋白是很重要的成份。目前至少发现存在着 8 种脱辅基脂蛋白, 其中 6 种的氨基酸顺序已测定清楚。它们的主要功能是在脂蛋白颗粒的形成和稳定中起结构作用, 并能调节代谢中的酶。

高脂血症可分为 5 种类型, 每种类型主要表现为某一种脂蛋白增加, 但从临床、遗传学和病理生理学研究证明, 每一类高脂血症都是不同质的、各自内部都有更复杂的原因。下面从分子遗传学角度简述一下各个类型。

I 型高脂血症表现为在禁食条件的高乳糜微粒血症。它存在两种主要的分子缺陷, 其一是缺乏脂蛋白脂酶, 尤其是缺乏肝外脂蛋白脂酶; 其二是缺乏脂蛋白脂酶的辅因子——脱辅基蛋白质 C-II。一般来说, I 型高脂血症呈常染色体隐性遗传, 但脂蛋白异常也可继发于其它全身性疾病如系统性红斑狼疮、巨球蛋白血症等。

II 型高脂血症主要表现为极低密度脂蛋白(VLDL)、低密度脂蛋白(LDL)胆固醇和甘油三酯水平增高, 或是 LDL 水平单独增高, 它同三种主要的高胆固醇血症的遗传类型有关。提示这一类型可能属于多基因

遗传型, 有两个或两个以上的缺陷基因参与这一类型的异常表达。

III 型高脂血症表现为高胆固醇血症和高甘油三脂血症, 它与某一种脱辅基蛋白质 E (简称 apoE) 多态变异型的存在有关。现已经知道有 3 种 apoE 变异型, 分别为 E₂、E₃ 和 E₄, 用等电聚焦电泳可以区分。这 3 种变异型都是由于 apoE 肽链中氨基酸替换所引起的。存在于高脂血症 III⁻ 型病人中的主要变异型是 E₂/E₂。这种变异型可能是功能缺陷型, 它可以导致残留乳糜微粒代谢障碍和中间密度脂蛋白(IDL) 在血浆内堆积。然而许多研究表明, E₂/E₂ 表型本身并不足以引起 III 型高脂血症, 只有遗传缺陷和环境因子共同存在才能发生该病。

IV 型高脂血症表现为血浆甘油三脂和 VLDL 水平增高。在这一类型中, 肝细胞内 VLDL 生产速率显著增加, 而 VLDL 的廓清作用则受到阻碍, 其分子机制尚不清楚, 可能 IV 型也存在几种不同的遗传变异型。此外, 这一类型的表达明显受环境作用的影响。

V 型高脂血症的突出特征为由于 VLDL 和乳糜微粒增加引起血浆中甘油三脂增高。血浆胆固醇呈轻度增高, HDL-胆固醇水平通常正常或偏低, V 型病人的甘油三脂水平显著高于 IV 型病人, 但 V 型往往是暂时的。在这一类型中必然存在周围脂解机制的部分缺陷, 结果导致乳糜微粒和 VLDL 廓清延迟。但其具体的分子机制仍不清楚。

此外许多方面的研究还表明, 脂蛋白异常是引起冠状动脉心脏病和成熟前动脉粥样硬化症的主要原因。大量的流行病学和实验材料强烈提示, LDL 具有致动脉粥样硬化性质。但它的机制尚不清楚。高甘油三脂血症与成熟前动脉粥样硬化心脏病和外周血管疾病的发生有关, 但二者之间究竟是因果关系, 还是继发于其它脂蛋白的代谢紊乱, 则有待进一步深入研究。

既然脂蛋白异常能导致高脂血症和一系列心血管病变, 而它的主要成份是脱辅基脂蛋白, 故有必要详细研究编码这组蛋白质的几个基因。

血浆高密度脂蛋白(HDL)的主肽是受 apoA-i 基

因控制的,它的基因位点多态性与过早发生动脉粥样硬化有关。另外,也有人发现邻近 apoA-I 基因的 DNA 多态性与高甘油三脂血症的发生有关。apoA-I 基因结构主要有 4 个特点:(1) 整个 apoA-I 基因有 3 个插入顺序,分别具有 197bp、185bp 和 588bp,且 5' 端均为 GT,3' 端是 AG;(2) 编码成熟 apoA-I mRNA 的区域有 804 个核苷酸,其初级翻译产物由 267 个氨基酸组成,包括成熟的 apoA-I 蛋白 243 个氨基酸和前片段原 24 个氨基酸;(3) 它的 3' 端非编码区由 55 个核苷酸组成, RNA 裂解信号 AAUAAA 位于终止密码子下游 36bp 处, poly(A) 尾位于 55bp 处;(4) apoA-I 启动子内无 TATA box 保守顺序。apoA-I 基因位于第 11 号染色体长臂 1 区 3 带至末端区域内,它与另一个 apoC-III 基因紧密连锁。apoC-III 基因位于 apoA-IB' 端下游 2.6kb 处, III 可以同时转录,暗示这两个基因的表达可能有同一调控机制。

在对两名严重冠心病患者的检查中发现了 apoA-I 基因有限制性片段长度多态性存在,进一步研究表明, apoA-I 基因位点多态性是由 DNA 片段插入所致,插入位置在 apoA-I 基因第三插入顺序上游 100bp 到下游 312bp 区域内。估计插入片段的大小至少要大于 6.5kb。这一 apoA-I 基因插入 DNA 片段形成的变异型是在严重缺乏 apoA-I 蛋白和 apoC-III 蛋白的病人中发现的。根据对 apoA-I 基因结构本身的了解可以推测, DNA 片段的插入不但可抑制 apoA-I 基因本身的表达,还可抑制邻近基因的表达。

有人用 *Ssi*I 酶解高甘油三脂血症病人和正常人的 DNA,发现了邻近 apoA-I 基因 3' 末端 DNA 顺序也具有多态性,进一步深入研究发现,这一多态性实际上存在于与 apoA-I 基因紧密连锁的 apoC-III 基因中。因此可以推测,这两个基因是由一个共同的祖先基因经过重复产生的,形成了一个基因复合物,因而能一起进行转录,还可能受共同的基因表达调控机制的调控。

除了 apoA-I 和 apoC-III 基因外,已发现了其它一些脱辅基脂蛋白基因的变异型。如 apoE 分子的 112 位和 158 位上可发生半胱氨酸/精氨酸替换,从而产生 apoE 多态变异型。这显然是由于 apoE 基因中的碱基突变造成的。这一突变与 III 型高脂蛋白血症的表达密切相关。有如 apoC-II 基因 3' 端 2kb 处的一个单碱基突变造成 apoC-II 基因位点的多态性。虽然这种多态性并不是直接构成高脂血症的遗传基础,但并不能排除 apoC-II 基因本身可能是产生高甘油三脂血症的一个主要因素。

另一个同高脂血症相关的 LDL 受体基因研究得也较为深入。LDL 受体基因约有 50kb,位于 19p13.1~13.3 处,共含有 18 个外显子,17 个内含子,转录下来的 mRNA 为 5.3kb,其中 2.5kb 为编码

顺序。它在结构上的一个重要特点是这一基因为外显子镶嵌物。外显子可由 LDL 受体基因移动到另一个基因上,也可反之。因此, LDL 受体基因中某些突变可能是由于这一机制所引起的。该基因的突变类型有 3 种:(1) 无用等位基因突变,这一类型不编码 LDL 受体蛋白;(2) 结合缺陷等位基因突变,这一突变型由第 4 和第 5 内含子重复顺序之间的重组引起,缺失 850bp,它能导致受体结合 LDL 的功能障碍;(3) 内吞缺陷突变型是一种 LDL 受体不能将被结合的 LDL 颗粒运入细胞,也是由重复顺序重组引起的。现已克隆到两种这类突变基因,分别为 FH274 和 FH781。FH 274 缺失了 5.5kb,包括 16 到 18 外显子区域; FH781 缺失了 LDL 受体基因 3' 末端 8kb 顺序。

糖 尿 病

糖尿病也是一种异质性疾病,表现为血糖慢性增高、尿糖丢失和脂肪分解过度。严重的还可出现酮症酸中毒。糖尿病分原发性和继发型两种。原发性又可分为胰岛素依赖型 (IDDM) 和非胰岛素依赖型 (NIDDM)。糖尿病是多个遗传因子与环境因子共同作用的结果。大量研究表明,胰岛素基因和其它一些基因可能是糖尿病多基因遗传的主要因子。这里将着重讨论有关胰岛素基因的研究。

1980 年 Bell 第一次分离和克隆了胰岛素基因,并测定了整个顺序,以后用大鼠胰岛素基因的 cDNA 探针从人体胰岛肿瘤中分离出人体胰岛素 mRNA,然后在反转录酶作用下获得了人体胰岛素基因 cDNA 克隆,并以此为探针,从人体基因组文库中分离出了人体胰岛素基因。人体胰岛素基因位于 11 号染色体短臂,长 1355bp,包括 3 个外显子和两个内含子。第一外显子编码成熟 mRNA 上的核糖体结合位点,第二外显子包括起始密码子 ATG 以及编码信号肽、B 链和 C 肽的 DNA 顺序,第三外显子编码 A 链顺序,位于 3' 端距离终止密码子 74 核苷酸处,是 poly(A) 尾。胰岛素基因的 5' 侧翼区是调节顺序,启动子位于第一外显子上游 25bp 处,第一外显子上游 168—258bp 之间的顺序为增强子。

胰岛素是最为保守的生物分子之一。在人类胰岛素基因内存在着一定量的等位性差异,在 4 个已被测定顺序的人体胰岛素基因中,各自有 4 个核苷酸不同。这些多态性的位置在第一内含子中有一个,第二内含子中有一个,3' 端非翻译区中有两个,但均没有发现它们具有病理效应。

在胰岛素基因 5' 端侧翼区,距胰岛素 mRNA 合成起始上游 363bp 处,存在一个高度多态区。导致这一区域产生高度多态性的原因是这一区域内插入了不同长度的 DNA 顺序,根据插入片段的长度,可将该区域分成三个等位基因:(1) I 类等位基因,插入 DNA 顺

序为 0—600bp; (2) II 类等位基因, 插入 DNA 顺序约 600~1.600bp; (3) III 类等位基因, 插入 DNA 顺序为 1.600~2.000bp。尽管三类等位基因插入片段大小悬殊, 但基本结构相似, 都是由一段 14bp 的保守顺序串联重复组成, 顺序为: ACAGGGGTGTGGGG。由于人是二倍体, 从父母亲中获得的胰岛素等位基因型就会出现以下几种: 1/1, 1/2, 1/3, 2/2, 2/3, 3/3。因此这一高度多态区域能用来作为双亲胰岛素基因的遗传标记。值得注意的是, 这一多态区域的顺序在人体基因组中是特有的, 它与增强子位置非常靠近。据推测, 在该区域内存在不止一个遗传标记, 而且该区域可能是胰岛素基因转录的重要功能调节区。

目前发现的胰岛素基因突变型基本上都是由点突变所引起的。按突变后产生的功能可分为两大类, 一类为裂解缺陷型, 另一类为胰岛素受体结合缺陷型。

在一个高胰岛素原血症家系中鉴定出一例胰岛素基因突变型。这一突变是由于胰岛素原中的一个氨基酸——精氨酸被替换, 从而阻止了 β 链与 C 肽的裂解。这一突变呈常染色体显性遗传, 但受累个体并不表现为糖尿病。在另一个日本家系中发现了另一个突变型, 也是由于氨基酸替换导致 A 链与 C 肽之间的裂解不能正常进行, 然而具有该突变型的病人, 不但表现出高胰岛素原血症, 而且还患有 NIDDM。

胰岛素受体缺陷型有两类, 均位于 β 链上, 一个位于 β 链第 24 位, 另一个在第 25 位。突变的形式均为苯丙氨酸为亮氨酸所替换。据认为, 该位置是胰岛素受体的结合位点, 具有重要功能。这一位置上的突变可导致胰岛素受体结合障碍, 出现 NIDDM 一系列临床表现。最近对这类突变型用限制性内切酶 *Mbo*II 和胰岛素基因 cDNA 探针分析, 发现这一突变使胰岛素基因中的一个 *Mbo*II 限制性位点丢失。*Mbo*II 识别顺序为 TCTTC, 而编码 β 链第 24 位和 25 位苯丙氨酸的密码子为 TTC—TTC。进一步作顺序分析证明, 这一突变是由于这两个密码子中的 C 转换为 G 所致, 而 TTG 正好是编码亮氨酸的密码子。

群体研究是揭示邻近胰岛素基因的多态区与糖尿病相互关系的一个重要方法。从基因型频率的调查中已经知道, 在高加索人群中 I 类和 III 类等位基因杂合子 (1/3) 占被检个体的 50%, 1/1 占 42%, 3/3 则不

到 10%。这样就可根据 Hardy-Weinberg 原理来检查 I 类和 III 类等位基因频率。设 I 类等位基因频率为 p , III 类等位基因频率为 q , 且 $p + q = 1$ 。那么, 根据公式就可从等位基因频率计算出基因频率: p^2 : $2pq = q^2$ 。在应用这一公式时的假定基础是: (1) 被检群体为随机婚配; (2) 被检基因按孟德尔规律遗传; (3) 没有突变、选择和漂变 (如移民) 等显著干扰。将正常群体和 NIDDM 型糖尿病病人的有关基因频率进行比较, 发现在某些人群中, 3/3 纯合子基因型频率在 NIDDM 型病人中显著增高, 疾病的相对发生率为 5, 在另一部分人群中甚至达到 10.5, 并且合并有高甘油三脂血症, 说明 III 类等位基因与 NIDDM 型糖尿病的发生有密切关系。

家系分析是研究疾病相关性的另一种有效方法, 应用这一方法可以检查多态插入顺序与糖尿病是否存在连锁关系。1983 年 Owerbach 曾对一例具有青年成熟期发作性糖尿病患者的大家系进行了调查, 没有发现胰岛素基因侧翼的 DNA 插入顺序与糖尿病之间的连锁关系, 其他有关家系的研究也证明, 不存在这种连锁关系。

在高加索人群中, 将 I、III 类等位基因的基因频率用于 IDDM 型糖尿病病人的研究发现, 1/1 基因型频率在 IDDM 型糖尿病病人中显著增高, 表明 IDDM 型糖尿病的发生与 I 类等位基因呈高度相关, 而且这一相关关系的显著性甚至高于 3/3 基因型与 NIDDM 型糖尿病之间的关系。但是到目前为止, 尚未见到有关 IDDM 型糖尿病的家系研究报告。

参 考 文 献

- [1] Henderson, H. E. et al.: 1987. *Hum. Genet.*, 75(1): 62—65.
- [2] Karathanasis, S. K. et al.: 1983. *Nature*, 301: 718—720.
- [3] Karathanasis, S. K. et al.: 1983. *Nature*, 304: 371—373.
- [4] Karathanasis, S. K. et al.: 1983. *Nature*, 305: 823—825.
- [5] Sudhof, T. C. et al.: 1985. *Science*, 228: 815—820.
- [6] Torfs, C. P. et al.: 1986. *Am. J. Hum. Genetics*, 38(3): 170—187.