

内科学中的基因研究

Studies on Genes in Internal Medicine

方福德

(中国医学科学院基础医学研究所研究员,中国协和医科大学教授)

在过去十余年中,生物医学所发生的革命性变化犹如本世纪初的物理学,是激动人心的。这是因为分子生物学,特别是基因工程技术的广泛深入发展和应用,一下子把过去人们想知道而不可知的基因一个个呈现在面前,并越来越多地知晓其结构和功能。这样,人类才开始真正步入分子层次去认识生物和医学问题。内科疾患涉及面广,故基因研究在内科学中硕果累累。

基因的概念最初源自遗传学,表示体现一定遗传特性的功能单元。现已进一步具体化,指脱氧核糖核酸(DNA)分子上一个能表达功能产物的片段,这便是基因的分子生物学概念。基因—环境—疾病作为人体与内外环境相互作用的模式,可用以解释许多疾病的发生机理,成为分子病理学研究的基础。

基因工程技术使我们能够构建人类基因组文库和 cDNA 文库,从中分离和确定各种基因,以测定其结构和功能,研究其与某种疾病的关系。值得一提的是,过去研究基因的思路和办法是“从表到里”,即从表型、蛋白质一级结构入手,最后研究相应的基因。现在则有新办法,是“由里及外”,即通过连锁分析和染色体步移或跳跃等方法先找到基因,再研究其产物的生物功能。这一办法即为反向生物学或反向遗传学,它是生物学思维模式的重大改进。这意味着,在现在只要人们认为值得研究和想得到的基因,都可以将其分离出来,按现已开始实施的人类基因组计划,15年后将作出全部基因图和核苷酸序列。届时人类对自身基因和疾病的认识必将更加深入,对待疾病和健康问题就更有针对性和把握。

在内科学领域中基因研究应用最早、最快、最直接和成果最多的分支学科是临床遗传学,很多遗传病的分子病理学已经搞清,并已可对致病基因进行检测,以诊断携带者和患病胎儿,在保障优生,提高人口质量方面发挥着越来越大的作用。反向遗传

学为越来越多的病因未知的遗传病找到了致病基因,这除了具有基因诊断(第四代诊断方法)的价值外,还为基因治疗提供了可能性和依据。

许多常见病如变性血管病、糖尿病、精神性疾病、癌症和风湿病等系多基因病(即有多种基因参与发病)。基因研究将可完全改变一直以流行病学和症候学为主要内容的工作面貌。对多基因病候选基因功能的了解不仅有助于搞清其分子病理学,还可进行遗传易感性预测。例如,Ⅰ型糖尿病的易感性是由 HLA-DR 系统Ⅱ类(ClassⅡ)基因点突变所致。以 HLA 系统基因为靶子可以确定对环境因子敏感的易感性基因。精神神经性疾病如老年性痴呆和双相型情感性精神病,通过基因分析证明,与基因单一突变密切相关。由此可以更好地了解其神经病理学和神经化学。基因研究也可了解肿瘤异质性的分子基础,以及细菌、病毒和寄生虫等外源性入侵基因的存在与否等,以为分型、预后和诊断所用。现已知道抗药性机理取决于抗药性基因的状况,通过对抗药性基因的控制或改造,有望解决此一棘手的医学问题。

基因作用的最后结果是由细胞应答和行为来体现的。其中涉及细胞膜、受体、生长因子和调节蛋白等对基因的作用,特别是对癌基因和抑癌基因的作用。这些方面都是目前医学分子生物学中的热点研究课题。

总之,基因研究对内科学的影响将越来越大。临床医生过细的专业分隔将在基因分子生物学的媒介下得到调节,因为他们可用同样的基因研究方法研究不同的疾病,彼此将有越来越多的共同语言。

一、基因的结构和功能

基因排列在 DNA 分子上。人 DNA 含有 3×10^9 个碱基对(bp),以双螺旋结构模式存在,一个 DNA

分子拉直后约2米长。人DNA中含有5~10万个有蛋白功能的基因,每个基因的长度平均约为1000bp,故有功能基因只占基因组DNA的5~10%。这些基因中只有少数基因表达,此种表达具有严格的时空性。那些在所有细胞中都有低水平表达的基因叫看家基因(housekeeping gene),另外一些基因则只能在特定的组织和细胞类型中表达。基因组中基因之外的结构部位为非编码序列,约占DNA分子的90%,其功能现尚不详,有可能与基因调控和分子进化有关,不过现在已经知道在那些部位存在很多特殊的序列,其中30~40%含有重复序列,分散于整个基因组中。有一种叫卫星DNA的序列,由重复序列串联而成,每个重复序列中有一个短的核心序列。有意思的是,在每个人DNA中这种串联重复序列的长度都不相同(因串联的重复序列的数目不一样),从而使所在区域成为高度可变区,若用合适的限制性内切酶将这些区域切下来,在凝胶电泳上可显示出具有个体特征的图谱。这种图谱象人的指纹一样,彼此不同,故叫DNA指纹(DNA fingerprint),在医学中用途很多。

人类基因在结构上的特点是不连续性,结构基因(即编码序列)被非编码序列隔开,所以一个基因可在基因组中很宽范围内分布。严格地讲,一个基因应包括五个区域:其一,5'非翻译区,原先有人叫前导区;其二,外显子,即编码区;其三,内含子,即外显子之间的非编码序列;其四,3'非翻译区,也叫尾部区;其五,调控元件,即在基因5'或3'旁侧序列中含有的对基因转录起调控作用的元件,如启动子、增强子和沉默子等,通常在5'旁侧区域有两个保守的序列,叫TATA盒和CAT盒。此外,还存在CPG岛,其中胞嘧啶(C)的甲基化程度对调节基因活性起重要作用。一般而言,甲基化程度越高,基因表达程度越低,反之亦然。人基因的大小和复杂性各异,迄今已发现的最大基因是抗肌营养不良蛋白基因,有14kb长的编码序列,它们被分为至少60个外显子,在基因组中分布范围为2000kb。该基因发生缺陷时会引起杜兴氏肌营养不良症。

基因的功能主要有两方面,遗传与变异。为了保证基因遗传的稳定性,细胞除了具有一套准确的DNA复制机制外,还具有严密无误的转录、翻译和修复机制。基因表达主要指转录和翻译过程,即遗传信息的传递过程。基因变异有好的一面,生物进化需要这种变异。淋巴细胞中免疫球蛋白基因的重排也可认为是变异,这种变异对于产生人体各种各样抗体是绝对必要的。但是,基因变异在更多情况下是有害的。基因突变包括点突变、缺失、插入、重排、重复等形式,其影响范围很广,如点突变可引起密码改变,破坏启动子,改变起始密码,产生终止密码或改变切接位点等等;缺失则可改变密码的读

框架。

二、研究基因的方法和技术

在临床上,研究基因的目的有三:其一,找出致病基因,如果是多基因病则需确定候选基因和易感基因,并在此基础上阐明发病机理;其二,诊断(包括疾病分型及预后);其三,指导治疗和预防。

寻找致病基因的方法主要有两种。一种是根据表型上特征蛋白质的结构,依遗传密码规则合成其相应的cDNA或寡核苷酸作为分子探针,从基因文库中筛选出基因。用免疫筛选法也能达到同样目的,与上不同的是基因文库中目的基因是与表达性载体重组,以便在宿主细菌中表达特征蛋白质,而后用它的抗体筛选之。构建含基因的重组体必须用“基因剪”限制性内切酶先将DNA分子剪成不同大小片段,再用连结酶把它们与基因载体连结起来。为了筛选准确,重组体还应在宿主菌中形成集落或噬菌斑,故须将重组体转移至宿主菌,以上过程叫转化。另一种方法就是反向遗传学方法,它特别适用于那些无表型特征可找的未知致病基因的寻找。其步骤是采用大量现在已知的随意DNA探针(即遗传标志)进行限制酶片段长度的多态性(RFLP)分析,根据最大优势对数值的计算推测致病基因是否与已知探针连锁以及彼此间的大概距离。一旦确定了连锁关系,即可进一步采用染色体步移或跳跃等技术逐步缩小寻找范围,最终找出致病基因片段,经基因序列测定便可详细了解其核苷酸排列及在基因组中的组织(Organization)。这种方法对于寻找由单一基因引起的疾病的致病基因是很适合的,对于多基因病,也可帮助筛选候选基因。

RFLP分析的应用很广泛,它的原理是利用基因旁侧序列较少保守性,容易发生中性突变(即无害突变)而使DNA呈现多态性的特点。如果中性突变发生在限制酶识别位点上,则该酶切位点会消失。在一定情况下,某些中性突变会产生出新的酶切位点。不论哪种情况,其结果都使酶切片段的长度发生明显改变,这很容易用放射自显影图谱显示出来。

利用基因的特点对疾病进行诊断叫基因诊断。基因诊断的物质对象是基因本身或基因的转录产物mRNA,采用的主要方法和技术是核酸分子杂交、基因探针和聚合酶链反应(PCR)法。核酸分子杂交是基于基因的两条互补链在热或碱性条件下变性(解链),而在缓慢冷却时恢复原状(称为复性),以及碱基配对原则(即腺嘌呤A与胸腺嘧啶T配对,鸟嘌呤G与胞嘧啶C配对)。这就是说,两个基因片段只要它们是相同的,则两条链双双互补,经变性—复性等操作后就能形成杂交体;若它们不

相同,则不能形成杂交体。检测杂交体的方法是把其中一个基因用放射性同位素标记,用自显影图谱来显示。常用的杂交方式有斑点杂交、印迹杂交和原位杂交等。前二者是将样品直接点在或印迹在硝酸纤维膜上与标记探针杂交,后者可在细胞涂片、组织切片或染色体上加标记探针杂交,以便定位。

对基因变异的检测,自1985年PCR技术问世以来,有了很大发展。有人估计,今后几年内PCR技术将在大多数情况下取代DNA探针技术。PCR技术灵敏、快速、简便,需样品量少,对DNA质量要求低,在临床上有多方面用途。

三、内科疾病的基因缺陷和分子病理学

采用上述基因研究方法及其它一些方法业已确定不少内科疾病的基因缺陷,也推动了分子病理学的发展。最早开展基因诊断的镰刀状细胞贫血症,是因为搞清了其病因为 β 珠蛋白基因第6密码子点突变(GAG $\rightarrow$ GTG)。肺气肿与 α_1 抗胰蛋白酶(α_1 -AT)基因的点突变有关。肺中性白细胞在吞噬和膜损伤过程中释放出弹性蛋白酶(NE)以消化肺泡壁内的弹性蛋白等结缔组织,使其失去弹性,导致肺气肿。正常情况下,肺内NE活性与 α_1 -AT对其抑制活性处于平衡状态,如果 α_1 -AT基因发生点突变则破坏这种平衡,使 α_1 -AT产生量和活性大大降低, α_1 -AT基因由3个非编码外显子和4个编码外显子及6个内含子构成,位于第14号染色体长臂,人群中已发现10多种正常的等位基因,当两个等位基因都发生缺陷时就会发病。这些缺陷基因叫危险基因,其中以Z基因为害最甚。Z纯合子引起严重的肺气肿和肝硬化。其机理是:位于 α_1 -AT基因第5个外显子的第342位密码子发生GAG $\rightarrow$ AAG点突变,这一突变并不影响肺巨噬细胞内 α_1 -AT的产生,却使产生的 α_1 -AT不能分泌出来起作用。

有的疾病所发生的基因变异类型不止一种,而是多种,这表明不同类型的基因变异可产生同一种疾病。最典型的例子是 α 地中海贫血。该病可因 α 珠蛋白基因缺失和点突变而引致 α 珠蛋白合成缺如或减少。现已知,在 α 地贫2中有两种基因缺陷类型,在 α 地贫1中有4种基因缺陷类型。在亚洲人群中还存在较多非缺失型 α 地贫,它们的基因变异主要有4种情况:其一, α_2 基因第一内含子5'切接位点处有5个核苷酸缺失,这使得mRNA前体不能正常切接,从而不能产生正常 α 珠蛋白;其二, α 珠蛋白基因终止密码点突变;其三, α 珠蛋白基因编码区点突变,如第125密码子发生CTG $\rightarrow$ CCG点突变,从而使原来编码的亮氨酸变成脯氨酸,由于这一氨基酸位置正好处于 α_1 与 β_1 珠蛋白接触时

邻界的H螺旋区,妨碍了 $\alpha_1\beta_1$ 二聚体的形成,进而影响四聚体形成;其四,poly A信号突变,使mRNA不能成熟。甲型血友病是因VIII因子缺陷所致,其基因变异也有多种类型(缺失和点突变)。VIII因子基因含有26个外显子,在第7、14、18、22、23、24和26外显子中各存在一个Taq I位点(其识别序列为TCGA),CG中的C可因甲基化而使Taq I位点消失,并均可引致本病。

有些疾病虽然基因变异的类型很多,但在同一种族中只有一种或少数几种主要类型,如 β 地中海贫血。国外已发现的 β 地贫在 β 珠蛋白基因上的突变类型达40多种,我国发现的有10多种,但在撒丁人中主要的突变类型是第39密码子发生点突变(CAG $\rightarrow$ TAG),中国人中的主要突变类型则为第17密码子点突变(AAG $\rightarrow$ TAG),这些均属无意义突变,使肽链合成提前中断。

最近有些文献报道表明,在某些基因编码序列内或非编码序列中发生特殊序列的重复可导致疾病产生。如表现为智力低下的脆性X综合症,现已从胎儿脑cDNA文库分离出致病基因FMR-1,在该基因5'侧和靠近5'端的编码序列中分别存在28个和30个CGG重复序列。患者的CG中C的甲基化作用发生频率高,造成脆性位点区域长度变异区,正常人则无此现象。在一种成人型运动神经元病——脊肌萎缩症患者中,在靠近雄激素受体基因(X连锁)5'端的地方,CAG重复序列比正常人增加了一倍,并与疾病表型紧密连锁,被认为是病理性突变。

线粒体是细胞内重要的细胞器,是向机体提供能源的部门。线粒体DNA也是除人体基因组DNA之外所存在的唯一DNA,在其上分布着数目不多的基因(主要与能量代谢有关)。这些基因的突变会引起很多疾病。现在已知道的是与某些神经、肌肉系统疾病的发生具有因果关系。由于线粒体DNA分子比较小,研究起来较方便,故线粒体DNA与疾病关系的研究是一个很好的模型,现已受到瞩目,很可能会成为一个新的研究热点。

很多常见病如冠心病、糖尿病、癫痫、高血压、甲状腺疾病、牛皮癣和精神分裂症等,其发病过程涉及多种环境因素和多种基因,属多基因病。对于这些疾病,首先应搞清楚相关的基因。一般来说,做到这一点困难很大,但不是不可能。主要的工作路线有三条:

其一,选择候选基因。根据表型或代谢异常的线索,确定与这些异常相关的基因是否与各该疾病相关。如狂躁—抑郁型精神病患者通常都有异常的多巴胺代谢,而酪氨酸羟化酶是多巴胺生物合成的关键酶,故可推测该酶基因的变异可能与该病的发病有关。有了这一线索,再经过家系中变异等位基

因与疾病共分离(cosegregation)分析即可确定该基因是否与本病相关。

其二,利用动物模型找出一些相关基因,再以此为线索验证这些基因与人的同种疾病是否相关。如在小鼠模型中已证明HLA DR3和DR4基因复合体、MHC复合体(H-2)和THY-1基因复合体与胰岛素依赖性糖尿病的发病有关。也找到一个与调节血浆高密度脂蛋白浓度相关的基因,它可能与动脉粥样硬化相关。

其三,由分子流行病学调查提供线索。如apo E和apo B基因与血浆胆固醇浓度有关联。血纤维蛋白原基因与血浆血纤维蛋白原浓度改变有关,而后者是冠心病的另一个危险因素。

相关基因的确定有可能找到遗传多态性和特别的等位基因,以便确定致病的易感性。如只在15%人群中存在的apo E4等位基因能增加血浆胆固醇浓度,与心肌梗塞发生有关。 α^2 球蛋白基因(维生素D结合因子位点)的多态性存在与易感艾滋病有关。胱硫醚合成酶基因的缺陷纯合子易感胶原畸形和血栓,其杂合子易感外因血管病和脑血管病。对疾病的易感性实际上是遗传变异对环境致病因素的敏感程度。肿瘤的形成涉及很多基因改变。一般认为,癌基因的激活和抑癌基因失活是重要的事件。目前最引人注目的抑癌基因P53,其失活易罹患肺癌、骨肉瘤、结直肠癌、乳癌、神经纤维肉瘤及横纹肌肉瘤等。多数肿瘤中P53因缺失和点突变共存而使其失活,通常是一个等位基因缺失,另一个等位基因发生点突变,点突变集中在P53基因的4个高度保守区域,即密码132~143、174~179、236~248和272~281。另一个抑癌基因RB失活引起视网膜母细胞瘤也是大家熟知的事实。此外,在癌基因H-ras3'端存在一可变串联重复序列,用Msp I/Hpa II进行RFLP分析时若出现一罕见片段,则肺癌发病率增高,所以这是肺癌易感性片段。在肿瘤化疗过程中癌细胞或早或迟会产生耐药性,这是因为多药抗药性基因(mdr 1)表达增强的结果,mdr 1基因产物通过不同途径将药物泵出细胞,使其不能有效地到达靶分子。

四、基因诊断

只要找到了致病基因的突变类型或与疾病相关的RFLP片段,即可在临床诊断中发挥作用。特别在其它诊断方法不能解决问题的情况下,基因诊断的优越性更加得到体现。基因诊断可从三个方面入手:其一,基因突变的检测。此种检测可了解大片段

缺失和插入,小片段缺失和插入,重排和重复,以及各种点突变等基因变异。其二,基因探查。主要采用RFLP结合家系分析,判断连锁的致病基因的遗传状况,以区分亲代染色体中哪条携带致病基因。其三,基因转录产物的检测。由于转录过程已将基因拷贝数扩增了很多倍并进行了加工步骤,切去了内含子,因此检测灵敏度比直接检测基因本身更高,特别对于很大的基因的检测更合适。

基因诊断不仅能诊断疾病,还能进行预后、疾病分型、检测外源性病原体,以及亲子鉴定、性别鉴定和法医学中的罪犯鉴定等。慢性粒细胞白血病(CML)95%病例发生9号与22号染色体重排,使位于前者的原癌基因C-abl与后者断裂点集中区域bcr发生易位嵌合,形成嵌合基因C-abl/bcr,用基因探针和PCR法很容易查出这种嵌合基因。实际检测结果表明,在一千多例标本中99%以上查到了bcr重排。若在缓解期查出bcr重排者,复发均快,无重排者病情稳定,缓解期延长。囊性淋巴瘤也有类似的染色体易位,用上述方法亦可查出。这些癌症的复发是由于残存癌细胞引起的,用PCR法将C-abl/bcr嵌合基因部分放大,并用含有嵌合结构的特异寡核苷酸探针检测之,可在 10^5 个细胞中检出一个残存癌细胞,这对于评价疗效和改进治疗方案有很大帮助。在乳腺癌中也发现原癌基因HER2/neu放大者预后极差。在疾病分类分型方面,最好的例子是根据N-myc基因表达谱对儿童神经母细胞瘤进行重新分类,解决了组织病理学很难鉴别诊断和分类的问题。

五、基因治疗

如果确定了致病基因,在理论上可采用基因治疗根治之。基因治疗的目标是通过不同途径纠正基因的缺陷或抑制致病基因的表达或提高机体自身的免疫力和抵抗力。国际上于1990年正式将基因治疗应用于临床。第1例经治的腺苷脱氨酶缺乏症(重症免疫缺陷)患者情况大有好转。今后将在遗传病和癌症这两个方面逐步扩大、深入研究和应用这一疗法。美国国家卫生研究院(NIH)上任不久的第一个女院长希利(B. Healy)被认为是“铁腕人物”,在她的科研战略中,主要突出了基因治疗和人类分子遗传学这两大领域,投入了大量资金。由此亦可见医学研究趋势之一斑。不可否认,基因治疗在目前尚存在一些技术难题,但逐步获得解决是有希望的。

(责任编辑 蔡德诚)