

# 假基因及其生物学功能概述

付春鹏 (山东省潍坊科技学院 寿光 262700)

**摘要** 假基因是功能基因有缺陷的拷贝,指具有与功能基因相似的序列,但序列中发生了众多突变而失去了原有功能的基因。假基因曾被认为是典型的非编码“垃圾序列”,而如今越来越多的研究证实假基因具有重要的生物学功能。本文概述假基因的发现、生成机制、鉴定方法、作用机制和生物学功能诸方面研究的进展。

**关键词** 假基因 基因组 调控 生物学功能

自人类基因组计划完成以来,科学家相继完成了多个生物基因组的测序工作,庞大的数据让科学家对基因组有了全面的认识。研究表明,人类的基因组大约有 30 多亿个碱基对,而其中仅有不到 2% 的 DNA 序列编码蛋白质,其余 98% 以上的序列则一度被认为是近中性选择条件下的进化过程产生的、没有功能的“垃圾序列”。近年发现这些所谓的“垃圾序列”并不是“一无是处”,假基因(pseudogene)就是其中之一。相关研究证实,有些假基因不但能够转录生成 RNA,进一步翻译生成蛋白质,而且还在多个水平调控基因表达,进而参与有机体的多个生理、生化过程。假基因已成为生命科学研究的热点领域之一。

## 1 假基因

1.1 假基因的发现 假基因是指具有与功能基因相似的序列,但由于发生众多突变导致其失去了原有的功能的 DNA 序列。1977 年,Jaçq 等<sup>[1]</sup>在非洲爪蟾基因组中克隆到 1 个与 5s rRNA 相类似的基因序列。通过与 5s rRNA 基因比对后发现,该基因在 5'端存在 16 bp 的缺失和 14 bp 的错配,而在非洲爪蟾体内没有检测到该基因的 mRNA 序列,表明它没有表达活性,于是就将这个 5s rRNA 的同源物描述为假基因。随着基因组测序技术的发展,大量不同家族的假基因被陆续发现,假基因也被定义为功能基因有缺陷的拷贝<sup>[2]</sup>。

1.2 假基因的产生机制 假基因主要通过两种途径产生:①细胞在分裂之前复制整个基因组时,DNA 复制或染色体联会交换过程中功能基因的编码区或调控区发生的各种突变(碱基的插入、缺失、置换或移码),均会导致复制后的基因无法进行编码,从而丧失正常功能而成为假基因,这种假基因称为重复假基因<sup>[3]</sup>;②DNA 转录为 mRNA 后再逆转录为 cDNA 并重新整合进入基因组(很可能发生在生殖细胞中),在此过程中因为插入位点不合适或序列发生突变而失去正常功能,这样形成的假基因称为加工假基因或返座假基因<sup>[4]</sup>。

1.3 假基因的鉴定 由于假基因与亲本基因存在高度的序列相似性,所以正确区分假基因和亲本基因就成为假基因研究的关键环节。这方面最早的工作是通过计算核酸的非同义替换与同义替换的比率(Dn/Ds)

来识别假基因的<sup>[5]</sup>。理论上分析,由于假基因不编码蛋白质,不受正选择或者纯化选择影响,所以推测在近中性选择条件下假基因的 Dn/Ds 值应该等于或接近 1。通过该方法已经在多个物种中鉴别出众多假基因,例如在人类基因组序列中就发现了 8000 个假基因。随着测序技术发展,大量假基因的鉴定逐渐变得主要由生物信息学来完成,PseudoPipe、RetroFinder 和 PseudoFinder 是 3 种常用的分析软件<sup>[6]</sup>,但它们都需要物种的基因组、转录组以及蛋白组信息,因此上述软件不适用于非模式生物的假基因鉴定。鉴于此,Molineris 等<sup>[7]</sup>提出适用于多数物种的假基因鉴定方法——REGEXP。REGEXP 方法的流程为:将编码基因和假基因进行比对从而找到高匹配对(high score pairs, HSPs),进而寻找 HSPs 附近的基因簇。由于加工假基因仅含有原始基因外显子,且假基因对应的 HSPs 互非非常接近,故正常功能基因所对应的 HSPs 反而被内含子分割。据此建立数据库,并与 Ensemble、VEGA 以及 Pseudogene.org 等假基因数据比对,最终确定假基因<sup>[8]</sup>。

1.4 假基因的作用机制 假基因的作用序列具有专一性,只影响与假基因本身相似的一些序列。由假基因介导的调控机制主要发生于转录水平和转录后水平。

1.4.1 转录水平的调控机理 当 mRNA 逆转录生成加工假基因时,没有正常编码功能的加工假基因整合到基因组中作为一个新位点,并与其亲本基因相互作用而调控细胞的不同生化过程。高迁移率族蛋白 A1 (high mobility group protein A1, HMGA1) 是基因激活的特定辅助因子,与肿瘤和糖尿病的发生密切相关。其假基因 HMGA1-P 的转录物可降解功能基因 HMGA1 的转录物,原因在于 HMGA1-P 的转录物和亲本基因 HMGA1 的转录本在 3'非翻译区上共同分享一个重要的转录调控元件<sup>[9]</sup>。

1.4.2 转录后水平的调控机理 假基因的转录物可通过作为亲本基因的反义 RNA、竞争性结合亲本基因的微小 RNA(microRNA, miRNA) 和产生内源性小干扰 RNA(small interfering RNA, siRNA) 来降解亲本基因等几种方式在转录后水平调控亲本基因的表达。其中 siRNA 的生成机制研究较为透彻:假基因主要通过假

基因—亲本基因以及假基因—假基因两种配对方法来产生相关的 siRNA。前一种配对中的亲本基因作为产生 siRNA 的正义链,反义链则来自于互补的假基因片段;后一种配对情况则通过插入重复片段形成发夹结构而生成 siRNA。

## 2 假基因的生物学功能

2.1 基因表达 人们起初认为假基因不具有编码蛋白质功能,但是近年来相关研究表明某些假基因可以编码比亲本功能基因稍短的蛋白质。例如,磷酸甘油酸变位酶 3 基因( phosphoglycerate mutase family 3, *PGAM3*) 是第一个被发现的具有编码能力的假基因<sup>[10]</sup>; 2 个含脓素结构域的 NLR 家族假基因( NLR family, pyrin domain containing 2 pseudogene, *NLRP2P*) 是高等灵长目特有的加工型假基因;最近人们发现其具有与亲本基因脓素蛋白基因( pyrin-only protein 2, *POP2*) 类似的功能。*NLRP2P* 具有完整的开放阅读框,编码 45 个氨基酸,这些氨基酸可以形成类似的脓素结构域( pyrin-domain)。而且 *NLRP2P* 的编码区与功能基因 *POP2* 相似度达 80% 以上。进一步的研究表明,该假基因具有调控细胞因子生成、细胞周期和细胞死亡的功能<sup>[11]</sup>。上述结果表明,假基因并非没有编码能力,有些假基因在选择压力的作用下日趋进化,逐渐地具备了各种各样的功能。因此现在的观点倾向于认为:假基因是生物体的基因贮备库。

2.2 调控基因表达 假基因对基因表达调控主要通过反义链、siRNA 和 miRNA 等几种方式来进行。例如, Korneev 等<sup>[12]</sup> 在蜗牛中发现一氧化氮合酶的假基因( nitric oxide synthase, *NOS*) 转录本,能作为反义 RNA 与亲本基因的转录本形成 RNA 双链,在蜗牛记忆形成的关键时期抑制一氧化氮合成。此结果说明假基因转录的反义 RNA 在细胞中的重要作用。同源性磷酸酶张力蛋白基因( phosphatase and tensin homolog, *PTEN*) 是迄今发现的第一个具有磷酸酶活性的抑癌基因,维持细胞内正常的 *PTEN* 蛋白水平可以抑制肿瘤的发生。其假基因 *PTENP1* 的 3' 端非编码区序列与 *PTEN* 高度相似, *PTENP1* 假基因竞争性结合 miRNA, 导致细胞内 miRNA 的浓度降低,致使 *PTEN* 逃离 miRNA 介导的抑制作用,从而调控肿瘤的发生<sup>[13]</sup>。

2.3 基因重排 动物体存在多种免疫球蛋白,而基因重排是产生免疫球蛋白多样化的主要机制。基因重排机制首先在鸡中被发现,鸡免疫球蛋白重链和轻链的可变区(V 区) 基因上游存在大量的假基因,这些假基因插入并置换重排 V 区的同源序列,从而使免疫球蛋白呈现多样化<sup>[14]</sup>。类似的功能基因和假基因间的基因重排导致的抗体多样性例子在其他脊椎动物中都能发现。

2.4 标记物种亲缘关系和进化距离 细胞核线粒体假基因( nuclear mitochondrial pseudogenes, *Numts*) 是线粒体 DNA 转移到核 DNA 中的片段,是研究细胞核内外两套遗传物质的桥梁。因为 *Numts* 的进化速率慢于线粒体基因,常被看做是线粒体 DNA 的“分子化石”,可以根据 *Numts* 中保留的原始基因信息确定物种之间的亲缘关系和进化距离。例如,在人类起源问题上, Zischler 等<sup>[15]</sup> 使用核基因组中的一段 D-loop 区的 *Numts* 序列作为外群,证实了现代人类起源于非洲的假说,结束了很久以来关于现代人类起源地的争论。

## 3 展望

长期以来一直认为假基因是看似正常、却没有功能的“死亡基因”,是基因组进化历程中的“化石记录”,然而事实却远非如此。如上所述,假基因在基因表达、基因调控、产生基因多样性等方面都扮演着极为重要的角色。因此,人们重新定义了假基因的概念:即假基因是指源于功能基因,却不能够表达相同类型产物(如蛋白质、tRNA 及 rRNA 等)的基因组序列<sup>[16]</sup>。

生物体犹如一架精密的仪器,每个生化反应都要经历严谨而复杂的调控过程。在生物进化的历程中,细胞基因组中大约有 98% 的“垃圾序列”会被“束之高阁”吗? 答案显然是否定的,而物尽其用应是其更合理的解释。例如,当下受到研究者广泛关注的长链非编码 RNA( long non-coding RNA, lncRNA) 就是所谓的“垃圾序列”的一部分,原先被认为是转录过程中产生的“噪音”,现在被证明在生物体中发挥着重要的生物学功能。假基因也不例外,相信在不久的将来,随着分子生物学技术的发展,科研人员必定对假基因有一个全面正确的认识,还假基因本来之真面目。

( 基金项目: 山东省自然科学基金面上项目, No. ZR2016CM12; 山东省高等学校科技计划项目, No. J17KB112、No. J16LE59; 潍坊科技学院博士基金, No. 2017BS03)

## 主要参考文献

- [ 1 ] JACQ C, MILLER JR, BROWNLEE GG. A pseudogene structure in 5S DNA of *Xenopus laevis* Cell [J]. Cell, 1977, 12( 1): 109-120.
- [ 2 ] 赵云航, 康相涛, 孙桂荣, 等. 假基因的研究现状及展望[J]. 中国畜牧兽医, 2008, 35( 7): 70-72.
- [ 3 ] MILLSON A, LEWIS T, PESARAN T, et al. Processed pseudogene confounding deletion/duplication assays for SMAD4 [J]. The Journal of Molecular Diagnostics, 2015, 17( 5): 576-582.
- [ 4 ] NOLL A, RAABE CA, CHURAKOV G, et al. Ancient traces of tailless retropseudogenes in therian genomes [J]. Genome Biology and Evolution, 2015, 7( 3): 889-900.
- [ 5 ] LI WH, GOJOBORI T, NEI M. Pseudogenes as a paradigm of neutral evolution [J]. Nature, 1981, 292( 5820): 237-239.

# 人乳头瘤病毒的致癌机制及预防疫苗研究概述

肖 雷 于紫荆 姚菁华\* (中国矿业大学化工学院生物工程系 徐州 221116)

摘 要 人乳头瘤病毒可导致宫颈癌、外阴癌和肛门癌等严重疾病,其中宫颈癌的发病率和死亡率较高,而且发病有年轻化的趋势,其预防和治疗受到广泛关注。本文概述人乳头瘤病毒的结构、分类和染病机制,并介绍预防和治疗该病毒感染的疫苗研究现状。

关键词 人乳头瘤病毒 宫颈癌 疫苗 治疗策略

人乳头瘤病毒(human papillomavirus, HPV)可致宫颈癌、外阴癌和肛门癌等严重疾病,其中宫颈癌不仅是世界上女性高发的三大恶性肿瘤之一,也是死亡率仅次于乳腺癌的女性恶性肿瘤。据统计,发展中国家患宫颈癌的人口占全世界的 2/3<sup>[1]</sup>,我国宫颈癌死亡率在所有癌症死亡率中排第四位。宫颈癌死亡率较高,高发于中年妇女,近年发现该病的发生有年轻化的趋势<sup>[2]</sup>。因此,HPV 感染的预防与治疗研究对保障妇女身心健康具有十分重要的意义。本文概述 HPV 的结构、分类和染病机制,并介绍预防和治疗该病毒感染的疫苗研究现状。

## 1 HPV 的结构及分类

HPV 属于乳头多瘤空泡病毒科(Papovaviridae)的乳头瘤空泡病毒 A 属,是一种无包膜的正十二面体球形小 DNA 病毒,会感染人体的皮肤及黏膜上皮组织,不仅会引发宫颈癌,还可能造成扁桃体癌、口腔癌、喉癌、鼻腔内癌以及食道癌等各种严重疾病。

研究发现,HPV 的 DNA 分子含有约 8000 个碱基对,其全基因组根据不同功能分为非编码区、早期基因区和晚期基因区三个功能区<sup>[3]</sup>。这三个功能区在感染时分别有着不同的功能:非编码区又称长控制区或上游调控区,包括启动子及增强子等负责基因的转录调

控的基因序列;早期基因区具有控制表达 E1、E2、E3、E4、E5、E6 及 E7 蛋白的 6 个基因,负责病毒的复制、转录和细胞转化,会产生致癌的蛋白;晚期基因区,包括产生衣壳蛋白 L1 和 L2 的蛋白,负责在病毒转录后组装子代病毒,其中 L1 蛋白作为 HPV 的主要特异性抗原可以诱导免疫反应。早期区和晚期区的 8 个开放读码框在形成肿瘤的过程中起着关键的作用<sup>[4,5]</sup>。

目前已知的 HPV 大约有 150 个型,其中包括高危型和低危型 HPV。低危型包括 HPV6、11、30、39、42、43、44 等;高危型包括 HPV16、18、31、33、34、35、39、45、51、52 等。高危型 HPV 的持续感染是引发宫颈癌的主要病因。在我国,HPV16 和 18 为最常见的两种感染型,这两类 HPV 也导致了 70% 以上的宫颈病变,因此目前的研究主要围绕这两类病毒进行<sup>[6]</sup>。

## 2 HPV 感染的致病机制

HPV 病毒普遍存在于人群中,主要通过性接触传播,但也会通过皮肤接触传播(不仅会通过接触感染者病变部位而受感染,也会通过接触被污染物体而导致间接感染)。成年女性一生中感染 HPV 的风险高达 80% ~ 90%<sup>[7]</sup>,但 95% 的人群感染 HPV 病毒后可以自发清除,并不会引起恶性肿瘤,只有在高危 HPV 持续感染的情况下才会引发癌症。持续感染时,高危 HPV

[6] 汤静思,杨明耀,李 英. 假基因的功能及其在癌症疾病中的重要作用[J]. 遗传, 2015, 37(1): 8-16.

[7] MOLINERIS I, SALES G, BIANCHI F, et al. A new approach for the identification of processed pseudogenes [J]. Journal of Computational Biology, 2010, 17(5): 755-765.

[8] 刘 慧,邹 彬,林 凤. 假基因鉴定及其功能分析[J]. 生物工程学报, 2013, 29(5): 551-567.

[9] CHIEFARI E, IIRITANO S, PAONESSA F, et al. Pseudogene mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes [J]. Nature Communications, 2010(1): 40.

[10] BETRÁN E, WANG W, JIN L, et al. Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene [J]. Molecular Biology and Evolution, 2002, 19(5): 654-663.

[11] PORTER KA, DUFFY EB, NYLAND P, et al. The CLRX. 1/NOD24 (NLRP2P) pseudogene codes a functional negative

regulator of NF, pyrin-only protein 4 [J]. Genes and Immunity, 2014, 15(6): 392-403.

[12] KORNEEV SA, KEMENES I, BETTINI NL, et al. Axonal trafficking of an antisense RNA transcribed from a pseudogene is regulated by classical conditioning [J]. Scientific Reports, 2013(3): 1027.

[13] POLISENO L, SALMENA L, ZHANG J, et al. 2010. A coding independent function of gene and pseudogene mRNAs regulates tumour biology [J]. Nature, 2010, 465(7301): 1033-1038.

[14] 张 红,刘钧珂,张改平,等. 鸡免疫球蛋白基因结构与抗体多样性产生的分子机制 [J]. 河南农业科学, 2007(2): 102-105.

[15] ZISCHLER H, GEISERT H, VON HAESLER A, et al. A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans [J]. Nature, 1995, 378(6556): 489-492.

[16] 刘国庆,白音宝力高,邢永强. 假基因研究进展 [J]. 生物化学与生物物理进展, 2010(11): 1165-1174. ◊