中国人群肝癌的易感基因研究

周钢桥*, 贺福初, 张红星

军事医学科学院放射与辐射医学研究所蛋白质组学国家重点实验室, 北京 100850

* 联系人, E-mail: zhougq114@gmail.com

收稿日期: 2011-07-20; 接受日期: 2011-08-27

国家科技重大专项(批准号: 2008ZX10002-016)、国家自然科学基金(批准号: 30900819, 30901231 和 30901707)资助项目

doi: 10.1360/052011-581

摘要 以单核苷酸多态性(single nucleotide polymorphism, SNP)为遗传标记的遗传关联研究是近年来鉴定复杂疾病易感基因的主要策略之一. 尤其是新近发展成熟的全基因组关联研究(genome-wide association study, GWAS), 已被公认为是行之有效的系统搜寻重大疾病易感基因的研究方法. 军事医学科学院与国内同行合作开展的 HBV 相关肝癌 GWAS 结果表明, 1p36.22 的 *UBE4B-KIF1B-PGD* 区域是一个全新的肝癌易感基因区域, 证明了遗传易感性在肝癌发生发展中的病因学意义. 肝癌易感基因的发现, 不仅为深入阐明肝癌的发生机制开辟了新的研究方向, 而且为肝癌的风险预测和早期预警研究提供了理论依据; 同时, 也为后续开发新型的治疗药物奠定了基础.

关键词

肝癌
易感基因
全基因组关联研究
单核苷酸多态性

1 肝癌的临床学、流行病学和遗传学

原发性肝癌是最常见的恶性肿瘤之一, 90%以上的原发性肝癌为肝细胞癌(hepatocellular carcinoma, HCC; 以下简称肝癌). 在人类常见的恶性肿瘤中, 肝癌的发病率位居第 5 位, 死亡率位居第 3 位^[1]. 我国是肝癌高发国家, 每年新发病例约 30 万, 约占全球肝癌新发病例的半数以上^[2]. 肝癌恶性程度高、病程进展迅速、易复发和转移, 确诊患者多以中、晚期为主, 5 年生存率仅在 4%左右. 即使采取手术切除以及术后的免疫治疗、中药治疗等手段来预防肝癌切除后的复发转移, 但总体术后 5 年复发率仍高达 60%以上. 因此肝癌仍是目前临床上治疗效果最差的肿瘤之一.

流行病学和实验研究均表明, 肝癌属于多因素复杂疾病, 是病毒因素、环境因素和机体的遗传因

素交互作用, 经过多个阶段的发展而产生的^[3]. 目前已初步揭示的病毒和环境因素包括乙型肝炎病毒(hepatitis B virus, HBV)或丙型肝炎病毒(hepatitis C virus, HCV)感染、黄曲霉毒素的摄入、饮水污染(如微囊毒素)、嗜酒、长期吸烟和食物中铁过量等. 我国的肝癌类型与欧美和日本等国家不同, HBV 感染仍是我国肝癌发生的最主要的原因^[2], 由此, HBV 相关的肝癌是我国肝癌研究的重点. 在我国, 虽然超过 80%的肝癌与慢性乙型肝炎相关, 但在乙肝患者中, 只有一小部分个体最终罹患肝癌; 此外, 肝癌的发生存在家族聚集现象^[4], 因此, 肝癌的发生可能与机体的易感基因密切相关. 发现肝癌的易感基因, 不仅有助于深入解析肝癌的发病机制, 而且能为个体罹患肝癌的风险预测、早期预防、个体化医疗及新型高效药物的筛选提供理论依据和生物靶标^[5].

引用格式: 周钢桥, 贺福初, 张红星. 中国人群肝癌的易感基因研究. 中国科学: 生命科学, 2011, 41: 785-789, doi: 10.1360/052011-581

2 国内外肝癌易感基因的研究进展

过去 30 多年间, 研究人员已发展了多种方法和策略用以鉴定肝癌等复杂疾病的易感/致病基因, 包括定位克隆法(如遗传连锁研究)、模式动物疾病模型(如基因敲除小鼠等)、基因表达谱、甲基化、小 RNA (microRNA, miRNA)、基因组杂合缺失(loss of heterozygosity, LOH)、基因组重排和基于连锁不平衡(linkage disequilibrium, LD)原理的遗传关联研究(genetic association study)等。其中, 以单核苷酸多态性(single nucleotide polymorphism, SNP)为遗传标记的遗传关联研究, 因其具有诸多优势而成为近年来鉴定复杂疾病易感基因的主要策略之一^[6]。利用该研究策略迄今已分析了近百个基因的 SNP 与肝癌发生发展的遗传相关性, 发现了 10 多个肝癌的候选易感基因。例如, 我们对来自我国肝癌高发区广西扶绥的肝癌病例-对照人群中的研究发现, 雌激素受体 α (estrogen receptor α , *ESR1*) 基因启动子区的(TA)_n 串联重复多态性位点、外显子 1 第 10 密码子处的同义 SNP 位点 T29C 和内含子 1 中的 *Pvu* II 多态性位点与肝癌的遗传易感性显著关联。以 T29C 位点为标记, 运用实时定量 PCR 检测显示, 在杂合子细胞中 T29C 位点的风险等位型 29C 与 *ESR1* 基因 mRNA 的高表达显著相关。由此, 上述研究在群体和分子水平两个层面初步证明了 *ESR1* 基因是中国人群肝癌的遗传易感基因。除 *ESR1* 基因外, 肝癌缺失基因 1 (deleted in liver cancer 1, *DLC1*)、细胞周期蛋白 D1 (cyclin D1, *CCND1*)、环氧化物水解酶(epoxide hydrolase, *EPHX*)、乙酰基转移酶(*N*-acetyltransferase, *NAT*)、细胞色素氧化酶 P450 (cytochrome oxidase P450, *CYP*)、雄激素受体 (androgen receptor, *AR*)、亚甲基四氢叶酸还原酶 (methylenetetrahydrofolate reductase, *MTHFR*)、胸苷酸合成酶 (thymidylate synthetase, *TYMS*)、肿瘤坏死因子 α (tumor necrosis factor α , *TNFA*)、鼠双微体同源基因 2 (murine double minute 2, *MDM2*)、肿瘤蛋白 p53 (tumor protein 53, *TP53*)、转化生长因子 β 1 (transforming growth factor α 1, *TGFBI*)、甘露糖结合凝集素(mannose-binding lectin, *MBL2*)和白介素 1B (interleukin 1B, *IL1B*)等基因的多态性也相继被发现与肝癌的发生具有显著的遗传学关联, 提示这些基因也有可能是肝癌的遗传易感基因^[7,8]。

上述研究主要是基于候选基因的遗传关联分析

方法, 采用的是“假设驱动(hypothesis driven)”的策略, 其中候选基因的选择通常建立在先验的生物学假设基础之上, 或者基于以往的家系连锁研究(linkage analysis)所确定的基因组“易感区域”之上。该策略能够鉴定遗传易感基因中的一小部分, 但不足以全面揭示疾病的遗传学基础^[9]。因此, 需要采用新的、更高效的研究方法或策略, 在全基因组水平系统筛选肝癌的易感基因。

3 肝癌的全基因组关联研究

近年来, 人类基因组计划(Human Genome Project, HGP)、单倍型图谱计划(Haplotype Mapping Project, HapMap; 包括 1~3 期)和千人基因组计划(The 1000 Genomes Project)的相继实施揭示了人类基因组中海量的 SNP 信息, 基本阐明了人类基因组的 LD 模式^[10~12]。另外, 近年来以 Affymetrix 和 Illumina 公司的基因芯片为代表的高通量 SNP 分型技术得到了飞速的发展, 为实现全基因组海量 SNP 的基因型扫描分析奠定了技术基础。在上述两大学术和技术突破的基础上, 复杂疾病的全基因组关联研究(genome-wide association study, GWAS)策略应运而生^[13,14]。该策略是一种不基于生物学假设的、全面的策略, 有望发现全新的疾病相关基因或通路。截至 2011 年 8 月, 研究人员已经对 400 余种复杂疾病和性状开展了 1000 多项 GWAS 研究, 发现了一大批易感基因或位点, 其中近 300 项研究成果发表在 *New England Journal of Medicine*, *Science*, *Nature* 和 *Nature Genetics* 等国际顶级学术刊物^[15]。因此, GWAS 已被科学界公认是行之有效的系统搜寻重大疾病易感基因的研究方法, 曾列 *Science* 评选的 2007 年度 10 大科学进展的第一位。2011 年, *Nature* 在评论重大科学热点时, 因 GWAS 策略在定位、发现复杂疾病易感基因中的巨大贡献而将之列为第二位, 并认为利用 GWAS 发现了大量有价值的疾病易感基因/位点, 后续开展易感基因的功能研究将对阐明肥胖、糖尿病等重大疾病的发病机制具有重要意义^[16]。在肝病研究领域, 迄今已有近 20 项 GWAS 的报道, 涉及的疾病包括慢性 HBV 感染、慢性 HCV 感染的治疗反应、胆石症、肝纤维化、非酒精性脂肪肝和药物诱发的肝损伤等^[17]。

为确定 HBV 相关肝癌的遗传易感基因, 军事医

学科学院与南京医科大学、中国医学科学院肿瘤研究所、第二军医大学等 10 多家医院和研究所合作, 招募了中国大陆 5 个肝癌高发区的 4500 多名肝癌病例和无肝癌对照个体(均为慢性 HBV 携带者). 在招募自肝癌高发区广西扶绥的 355 个肝癌病例和 360 个对照者中, 我们采用 Affymetrix SNP 5.0 芯片技术, 共分型了 440794 个 SNP 位点. 经过质量控制后, 在另外 4 个分别招募自广东、上海、江苏和北京的肝癌病例-对照人群(总的肝癌病例数达 1962 例, 对照 1430 例)和 1 个招募自广西的核心家系人群(共 159 个)中对最

显著关联的 45 个 SNPs 进行了重复验证, 结果发现, 1p36.22 区域的一个 SNP 位点(rs17401966)在上述 5 个样本集中均得到验证, 联合 P 值达 1.7×10^{-18} (联合 OR 值为 0.61). 进一步的 LD 作图分析显示, 该易感基因区域包含了泛素化因子 E4B (ubiquitination factor E4B, *UBE4B*)、驱动蛋白家族成员 1B(kinesin family member 1B, *KIF1B*)和磷酸葡萄糖醛酸脱氢酶(phosphogluconate dehydrogenase, *PGD*)3 个基因(图 1).

1p36 是一重要的 LOH 区域, 迄今已在血液细

图 1 1p36.22 区域的关联图

横坐标是基因组的位置(NCBI 版本号 36), 纵坐标是 SNP 的 P 值($-\log_{10}P$). rs17401966 在荟萃分析中的 P 值以蓝色菱形显示, 在 GWAS 阶段的 P 值以大的红色菱形表示. 其他位点的颜色代表其与 rs17401966 之间的连锁不平衡(LD)程度: 红色表示 $r^2 \geq 0.8$, 橙色表示 $0.5 \leq r^2 < 0.8$, 黄色表示 $0.2 \leq r^2 < 0.5$, 白色表示 $r^2 < 0.2$. 浅蓝色线表示重组率(来自 HapMap 计划的数据). 从 Santa Cruz Genome Browser (<http://genome.ucsc.edu/>)基因组浏览器中得出基因的位置. rs17401966 位于 *KIF1B* 基因的内含子 24. 图的下部显示了 rs17401966 (红色箭头所示)前后共约 1 Mb 范围内的 LD 结构, 该 LD 结构根据广西病例对照人群的基因型数据计算得出. SNP 之间的 LD 值以 r^2 表示, 红色的程度表示 r^2 值的大小, 最深的红色表示 $r^2=1$. 重组率和 LD 结构均清晰地指出 rs17401966 所在区域的边界. 该区域长约 244 kb, 包括了 *KIF1B*, *PGD* 和 *UBE4B* 基因的 3'端

胞、神经细胞和上皮细胞源性的 20 余种肿瘤中频繁观察到 1p36 LOH 的存在, 其中也包括肝癌^[18,19]。前人研究认为, 1p36 可能存在 3 个包含抑癌基因的微缺失区(microdeletion region, mcr)。2007 年, 位于其中一个 mcr 的抑癌基因——人色素域解旋酶 DNA 结合蛋白 5 (chromodomain helicase DNA binding protein 5, *CHD5*)得以鉴定^[20]; 而我们此次发现的 1p36.22 区域则正好位于另一个 mcr 中。1p36.22 *UBE4B-KIF1B-PGD* 区域位于一个长约 244 kb 的 LD 模块中, 包含了 *UBE4B*, *KIF1B* 和 *PGD* 三个基因, 已有研究报道了这些基因可在神经母细胞瘤和膀胱癌等肿瘤的发生中起作用, 因而提示其在肝癌的发生中可能也会发挥类似的作用^[21,22]。

为了验证这一假设, 我们运用免疫组化实验检测了 20 例肝癌的癌组织及其对应的癌旁组织, 发现 *UBE4B* 的表达没有异常, 但 *KIF1B* 和 *PGD* 在癌组织中的表达显著减少。进一步通过实时定量 PCR 实验显示, 只有 *KIF1B* 的表达与关联的 SNP 位点 rs17401966 的基因型具有相关性, 其高表达与 rs17401966 的保护性等位型 G 显著相关, 从而提示 *KIF1B* 有可能是真正的肝癌抑制基因。PGD 是位于戊糖代谢途径的一种酶, 前人对其生理、生化功能已有深入研究, 但回溯 30 余年的文献未见其与肿瘤关系密切的报道。*KIF1B* 是驱动蛋白超家族(kinesin superfamily)成员之一, 是细胞内小泡运输和细胞器转运的重要分子^[23]。*KIF1B* 在正常成人和胎儿的心、脑、脾、肺、肝、肾和结肠等有广泛表达, 尤以脑和神经组织的表达最为丰富。与在肝癌的研究相一致, 许多肿瘤组织中也存在 *KIF1B* 的表达缺失。而且, 已在多种肿瘤中发现该基因存在种系细胞和体细胞功能缺失(loss-of-function)突变^[24,25]。尤其值得关注的是, 最近两项独立研究发现, 抑制神经瘤细胞中 *KIF1B* 的表达能加速细胞增殖、并在裸鼠中促进肿瘤的生长, 而过表达 *KIF1B* 则可诱导瘤细胞发生凋亡, 从而提示 *KIF1B* 可能是潜在的神经母细胞瘤的抑癌基因^[21,22]。我们最近一系列的体外实验初步表明, *KIF1B* 也能显著影响肝癌细胞的增殖、侵袭和转移能力, 提示 *KIF1B* 相关通路参

与了肝癌的发病机制。我们的研究同时表明, 如果该基因组区域的基因发生变异, 会使人更容易罹患肝癌。通过计算群体归因分数显示, 肝癌发生的 24.1% 可归因于 1p36.22 *UBE4B-KIF1B-PGD* 区域的 rs17401966 位点, 同时, 该位点增加的肝癌家族相对风险度约为 3%。

4 肝癌全基因组关联研究及易感基因研究的生物学和医学意义

肝癌全基因组关联研究是国际上第一项关于 HBV 相关肝癌的 GWAS, 首次实现了从整体基因组水平系统搜寻肝癌的易感基因并取得成功, 发现了 1p36.22 *UBE4B-KIF1B-PGD* 是一个全新的肝癌易感基因区域, 为肝癌发生的遗传病因和分子机制提供了新的线索。对这些基因及其通路进行深入的功能研究, 有望解析肝癌发生发展的分子机制, 为新药的研发提供理论依据和候选生物靶标。

同时, 肝癌易感基因的发现也为肝癌的一级预防——风险预测和早期预警奠定了理论依据。疾病的风险预测已成为新的疾病防治模式, 它将使包括肝癌在内的诸多重大疾病的防治策略发生战略性的变革, 从以前对“终末状态”的救治转变为对疾病的早期预警和先期干预^[26,27]。当更多的肝癌遗传易感基因得到鉴定, 它们组合起来, 并与传统的风险因素(包括性别、年龄、饮酒史、肝癌家族史、肝硬化和 HBV 因素等)相结合, 将有望准确预测个体罹患肝癌的发生风险^[28-30]。

我们的研究证明了肝癌的易感基因能够通过系统的、不基于假设的、无偏倚的全基因组关联分析得到鉴定。然而, 与其他复杂疾病类似, 目前利用全基因组关联分析发现的疾病易感基因位点仅能解释很小一部分病例, 更多的疾病易感基因尚有待发现。可以预见, 在未来几年里, 应用新一代测序技术进行肝癌的全外显子和全基因组测序, 同时分析基因组的稀有变异和常见变异, 将会进一步促进这一研究领域的发展^[31,32]。

参考文献

- Gomaa A I, Khan S A, Toledano M B, et al. Hepatocellular carcinoma: Epidemiology, risk factors and pathogenesis. *World J Gastroenterol*, 2008, 14: 4300-4308
- Nguyen V T, Law M G, Dore G J. Hepatitis B-related hepatocellular carcinoma: Epidemiological characteristics and disease burden. *J Viral*

- Hepat, 2009, 16: 453–463
- 3 Farazi P A, DePinho R A. Hepatocellular carcinoma pathogenesis: From genes to environment. *Nat Rev Cancer*, 2006, 6: 674–687
 - 4 Yu M W, Chang H C, Liaw Y F, et al. Familial risk of hepatocellular carcinoma among chronic hepatitis B carriers and their relatives. *J Natl Cancer Inst*, 2000, 92: 1159–1164
 - 5 Krawczyk M, Müllenbach R, Weber S N, et al. Genome-wide association studies and genetic risk assessment of liver diseases. *Nat Rev Gastroenterol Hepatol*, 2010, 7: 669–681
 - 6 Altshuler D, Pollara V J, Cowles C R, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, 2000, 407: 513–516
 - 7 Kao J H, Chen P J, Chen D S. Recent advances in the research of hepatitis B virus-related hepatocellular carcinoma: Epidemiologic and molecular biological aspects. *Adv Cancer Res*, 2010, 108: 21–72
 - 8 Kim Y J, Lee H S. Single nucleotide polymorphisms associated with hepatocellular carcinoma in patients with chronic hepatitis B virus infection. *Intervirol*, 2005, 48: 10–15
 - 9 de Bakker P I, Yelensky R, Pe'er I, et al. Efficiency and power in genetic association studies. *Nat Genet*, 2005, 37: 1217–1223
 - 10 International HapMap Consortium. A haplotype map of the human genome. *Nature*, 2005, 437: 1299–1320
 - 11 International HapMap Consortium, Frazer K A, Ballinger D G, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 2007, 449: 851–861
 - 12 1000 Genomes Project Consortium, Durbin R M, Abecasis G R, et al. A map of human genome variation from population-scale sequencing. *Nature*, 2010, 467: 1061–1073
 - 13 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 2007, 447: 661–678
 - 14 Easton D F, Pooley K A, Dunning A M, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 2007, 447: 1087–1093
 - 15 <http://www.genome.gov/26525384>
 - 16 Van Noorden R, Ledford H, Mann A. New year, new science. *Nature*, 2011, 469: 12
 - 17 Krawczyk M, Müllenbach R, Weber S N, et al. Genome-wide association studies and genetic risk assessment of liver diseases. *Nat Rev Gastroenterol Hepatol*, 2010, 7: 669–681
 - 18 Bagchi A, Mills A A. The quest for the 1p36 tumor suppressor. *Cancer Res*, 2008, 68: 2551–2556
 - 19 Li S P, Wang H Y, Li J Q, et al. Genome-wide analyses on loss of heterozygosity in hepatocellular carcinoma in Southern China. *J Hepatol*, 2001, 34: 840–849
 - 20 Bagchi A, Papazoglu C, Wu Y, et al. CHD5 is a tumor suppressor at human 1p36. *Cell*, 2007, 128: 459–475
 - 21 Munirajan A K, Ando K, Mukai A, et al. KIF1Bbeta functions as a haploinsufficient tumor suppressor gene mapped to chromosome 1p36.2 by inducing apoptotic cell death. *J Biol Chem*, 2008, 283: 24426–24434
 - 22 Schlisio S, Kenchappa R S, Vredeveld L C, et al. The kinesin KIF1Bbeta acts downstream from EglN3 to induce apoptosis and is a potential 1p36 tumor suppressor. *Genes Dev*, 2008, 22: 884–893
 - 23 Hirokawa N, Takemura R. Biochemical and molecular characterization of diseases linked to motor proteins. *Trends Biochem Sci*, 2003, 28: 558–565
 - 24 Yang H W, Chen Y Z, Takita J, et al. Genomic structure and mutational analysis of the human *KIF1B* gene which is homozygously deleted in neuroblastoma at chromosome 1p36.2. *Oncogene*, 2001, 20: 5075–5083
 - 25 Yeh I T, Lenci R E, Qin Y, et al. A germline mutation of the *KIF1B* beta gene on 1p36 in a family with neural and nonneural tumors. *Hum Genet*, 2008, 124: 279–285
 - 26 Manolio T A. Genomewide association studies and assessment of the risk of disease. *N Engl J Med*, 2010, 363: 166–176
 - 27 Kraft P, Wacholder S, Cornelis M C, et al. Beyond odds ratios—communicating disease risk based on genetic profiles. *Nat Rev Genet*, 2009, 10: 264–269
 - 28 Yuen M F, Tanaka Y, Fong D Y, et al. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. *J Hepatol*, 2009, 50: 80–88
 - 29 Wong V W, Chan S L, Mo F, et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. *J Clin Oncol*, 2010, 28: 1660–1665
 - 30 Yang H I, Sherman M, Su J, et al. Nomograms for risk of hepatocellular carcinoma in patients with chronic hepatitis B virus infection. *J Clin Oncol*, 2010, 28: 2437–2444
 - 31 Kryukov G V, Pennacchio L A, Sunyaev S R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet*, 2007, 80: 727–739
 - 32 Biasecker L G. Exome sequencing makes medical genomics a reality. *Nat Genet*, 2010, 42: 13–14