

剪接蛋白 U2AF65 基因多态性和吸烟交互作用与胰腺癌风险的关联

田晶 朱贝贝 田瑶 钟荣 缪小平 王丽

【摘要】 目的 探讨U2小核糖核蛋白体辅助分子U2AF35和U2AF65功能基因多态性及其与吸烟交互作用与胰腺癌风险关联。方法 采用两阶段病例对照研究,筛查阶段采用Openarray高通量分型技术对筛选出的U2AF35和U2AF65基因4个潜在功能标签单核苷酸多态性(SNP)位点在研究对象(武汉地区298例原发性胰腺癌病例和525例非肿瘤对照)中进行基因分型;对筛查阶段的阳性位点进一步在验证人群(北京地区413例原发性胰腺癌病例和557例非肿瘤对照人群)中采用TaqMan基因分型技术进行验证。并采用加法交互模型分析基因和吸烟交互作用对胰腺癌发生风险的影响。结果 筛查阶段U2AF65基因的rs310441和rs310445与胰腺癌发病风险均存在显著关联。对rs310445进行验证后发现,以携带TT基因型者为参照,携带C等位基因者的胰腺癌发病风险增加,合并两阶段人群的OR值为1.31(95%CI:1.07~1.60, $P=0.010$)。合并人群观察到吸烟和rs310445 C等位基因之间的协同作用可增加胰腺癌的发生风险,协同指数为2.08(95%CI:1.37~2.78)。未发现U2AF35位点多态与胰腺癌发生风险相关。结论 U2AF65 rs310445 C等位基因与吸烟协同作用可能增加胰腺癌发生风险,但仍需大样本研究验证。

【关键词】 吸烟;胰腺癌;U2小核糖核蛋白体辅助分子;单核苷酸多态性

Association between pancreatic cancer risk and the interaction of U2AF65 gene polymorphisms and smoking

Tian Jing¹, Zhu Beibei², Tian Yao¹, Zhong Rong², Miao Xiaoping², Wang Li¹.
1 Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing 100005, China; 2 Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology

Corresponding authors: Wang Li, Email: liwang@ibms.pumc.edu.cn; Miao Xiaoping, Email: miaoxp@gmail.com

This work was supported by a grant from the National Natural Science Foundation of China (No. 81041079).

【Abstract】 Objective To determine the association between U2 small nuclear ribonucleoprotein auxiliary factor 35/65 (U2AF35 and U2AF65) and pancreatic cancer (PC). **Methods** A two-stage analysis case-control study was conducted. Four candidate tag single nucleotide polymorphisms (tagSNPs) were genotyped by Taqman Openarray assay in a screening population living in Central China (298 PC cases and 525 controls). Thereafter, rs310445 in U2AF65 was genotyped by TaqMan real-time polymerase chain reaction (RT-PCR) in a validation Chinese Han population from Beijing (413 cases and 557 controls). **Results** rs310445 in U2AF65 gene was significantly associated with PC in both screened population and combined population. Subjects with C allele had a higher risk of PC compared to those with the TT genotype, with OR of 1.31 (95% CI: 1.07–1.60, $P=0.010$) for the combined population. A synergic effect of smoking and C allele of rs310445 was also observed in the combined population, with Synergic Index of 2.08 (95% CI: 1.37–2.78) in the combined population. **Conclusion** Our findings suggested the interaction between smoking and U2AF65 might play a role in PC. These findings should be confirmed by further independently large-scale population studies.

【Key words】 Smoking; Pancreatic cancer; U2 small nuclear ribonucleoprotein auxiliary factor; Single nucleotide polymorphism

DOI:10.3760/cma.j.issn.0254-6450.2014.06.021

基金项目:国家自然科学基金(81041079)

作者单位:100005 北京,中国医学科学院基础医学研究所 北京协和医学院基础学院流行病学和卫生统计学系(田晶、田瑶、王丽);华中科技大学同济医学院公共卫生学院流行病和卫生统计学系(朱贝贝、钟荣、缪小平)

通信作者:王丽, Email: liwang@ibms.pumc.edu.cn; 缪小平, Email: miaoxp@gmail.com

mRNA 选择性剪接与癌症的关系在近几年的研究中备受关注^[1]。2011 年以来, *Nature*、*Cell* 等杂志先后报道了剪接体相关基因的突变与肿瘤发生之间的关系, 尤其是参与 U2 型剪切位点识别相关的基因^[2-4]。U2 核糖核蛋白体辅助因子(U2AF)是 U2 型剪接体的主要组成蛋白, 是协助 U2 核糖核蛋白体结合 pre-mRNA 分子并识别 pre-mRNA 上 3' 剪接位点的重要功能性蛋白。U2AF35 和 U2AF65 编码的蛋白是构成 U2AF 分子二聚体的 2 个亚基, 其主要功能为: 在 U2 型剪接体识别内含子的过程中, 分别与 pre-mRNA 上 3' 剪接位点的 AG 序列以及分支点下游的多聚嘧啶区结合, 共同介导 pre-mRNA 上 3' 剪接位点的识别^[2]。研究显示, 它们的基因突变与肺癌、慢性髓性白血病等多种恶性肿瘤的发生密切相关^[3,4], 但其与胰腺癌的风险关联尚未见报道。本研究采用两阶段病例对照研究, 探讨 U2AF35 和 U2AF65 的基因多态性及其与环境因素(吸烟)的交互作用对胰腺癌发病风险的影响。

对象与方法

1. 研究对象: 采用两阶段病例对照研究, 筛查阶段共纳入 298 名病例和 525 名对照, 病例为 2008 年 1 月至 2012 年 9 月在武汉市同济医院经病理组织确诊的新发胰腺癌患者, 对照来自同期武汉同济医院参加体检的非肿瘤健康人群。验证阶段共纳入 413 名胰腺癌病例和 557 名对照, 病例为 2008 年 1 月至 2012 年 12 月在北京协和医院经病理组织确诊的新发胰腺癌患者, 对照来自同期同医院体检的非肿瘤人群。两阶段病例和对照人群均按性别和年龄(± 5 岁)进行频数匹配。所有研究对象均在入院未做任何治疗之前收集 2 ml 抗凝血标本, 同时以问卷调查的方式收集其人口学以及吸烟、饮酒等资料。其中规律吸烟定义为每日至少吸 1 支烟且持续 ≥ 1 年; 规律饮酒定义为每天至少饮酒 1 次且持续 ≥ 3 个月。所有研究对象均获知情同意书。本研究获得中国医学科学院基础医学研究所伦理学委员会同意和批准。

2. 候选 SNP 选择: 首先从 HapMap Data Rel 24/Phase II Nov 08 (<http://hapmap.ncbi.nlm.nih.gov/>) 中下载中国北京地区汉族人群 U2AF35 和 U2AF65 分型数据; 并运用 Haploview4.2 (<http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview>) 软件挑选标签 SNP, 入选标准为 $R^2 > 0.8$ 且低频等位基因频率 > 0.05 。

3. 基因分型: 筛查阶段采用美国 Applied Biosystems 公司的 Taqman Openarray 基因分型系统对上述 SNP 进行分型; 对阳性位点在验证人群中使用时 TaqMan RT-PCR(美国 Applied Biosystems 公司) 进行分型验证。抽取 5% 样本来验证分型的准确性, 结果显示重复样本分型一致率 100%。

4. 统计学分析: 采用拟合优度 χ^2 检验判定对照组的 SNP 位点分型结果是否符合 Hardy-Weinberg 平衡(HWE)。采用非条件 logistic 回归分别估计两阶段人群在调整性别、年龄、吸烟和饮酒因素后 SNP 位点与胰腺癌风险关联的 OR 值及其 95% CI。采用 Breslow-Day 检验评价两阶段结果的层间一致性, 对符合层间一致性者进行数据合并, 在合并的同时校正研究阶段^[5]。采用加法交互模型探讨吸烟与 SNP 之间的交互作用, 具体指标包括协同指数(SI), 归因交互百分比(AP)和交互作用的超额危险度(RERI)^[6]。除加法交互采用 R 语言 3.0.0 (<http://www.r-project.org/>) 进行分析外, 其他所有分析均采用 SAS 9.2 软件完成。

结 果

1. 人口学特征: 两阶段共纳入 711 例胰腺癌病例和 1 082 例对照(表 1)。每阶段人群中病例组和对照组年龄、性别分布及饮酒史差异均无统计学意义($P > 0.05$); 筛查和验证阶段病例组吸烟者所占比例分别为 35.6% 和 41.7%, 均高于对照组吸烟者所占比例(分别为 23.2% 和 34.5%), 差异有统计学意义($P < 0.05$)。

2. 筛查人群 SNP 基因型分布及其与胰腺癌关联分析: 共挑选出 4 个具有潜在功能的标签 SNP, 包括 U2AF35 基因的 rs1789956、rs4920040 和 U2AF65 基因的 rs310441、rs310445。对上述 4 个 SNP 在对照人群中的基因型频率进行 HWE 检验, 除 rs4920040 (U2AF35) 不符合 HWE 之外($P < 0.001$), 其余 3 个 SNP 基因型分布均符合 HWE 平衡, P 值分别为 0.848、0.779 和 0.217, 纳入进一步分析中。

表 2 列出符合 HWE 平衡的 3 个 SNP 在筛查阶段病例对照人群中的分布。结果显示, U2AF65 基因的 2 个 SNP 位点均与胰腺癌风险存在统计学关联。经年龄、性别、吸烟和饮酒状态调整后, 与携带 rs310441 TT 基因型相比, 携带 G 等位基因者胰腺癌风险均增加, OR 值为 1.42 (95% CI: 1.06 ~ 1.90, $P = 0.020$)。而与携带 rs310445 TT 基因型者相比, 携带 C 等位基因者发生胰腺癌风险也增加, 校正后 OR 值

表 1 两阶段两组人群人口学特征

特 征	筛查阶段		χ^2 值	P 值	验证阶段		χ^2 值	P 值
	对照组(n=525)	病例组(n=298)			对照组(n=557)	病例组(n=413)		
年龄($\bar{x} \pm s$, 岁)	58.9 ± 12.9	60.3 ± 12.8		0.132	59.0 ± 13.3	58.6 ± 13.1		0.630
性别			2.311	0.129			0.008	0.929
男	260(49.5)	164(55.0)			341(61.2)	254(61.5)		
女	265(50.5)	134(45.0)			216(38.8)	159(38.5)		
吸烟			14.435	<0.001			5.209	0.023
否	403(76.8)	192(64.4)			365(65.5)	241(58.4)		
是	122(23.2)	106(35.6)			192(34.5)	172(41.7)		
饮酒			2.642	0.104			0.091	0.763
否	412(78.5)	219(73.5)			376(67.5)	275(66.6)		
是	113(21.5)	79(26.5)			181(32.5)	138(33.4)		

注: 括号外数据为人数, 括号内数据为构成比(%)

表 2 筛查阶段 U2AF35 和 U2AF65 候选 SNPs 在两组人群中的基因型分布及其与胰腺癌风险关联

基因	SNP	基因型	对照组(n=525)	病例组(n=298)	P 值	OR 值(95%CI) ^a	P 值 ^a
U2AF35	rs1789956	GG	460(87.6)	266(89.3)	0.779		
		AG	62(11.8)	31(10.4)			
		AA	3(0.6)	1(0.3)			
		GG ^b /AA+AG	460(87.6)/65(12.4)	266(89.3)/32(10.7)	0.482	0.83(0.53 ~ 1.31)	0.424
U2AF65	rs310441	TT	242(46.1)	113(37.9)	0.072		
		GT	234(44.6)	151(50.7)			
		GG	49(9.3)	34(11.4)			
		TT ^b /GG+GT	242(46.1)/283(53.9)	113(37.9)/185(62.1)	0.023	1.42(1.06 ~ 1.90)	0.020
	rs310445	TT	337(64.2)	159(53.4)	0.008		
CT		159(30.3)	121(40.6)				
CC		29(5.5)	18(6.0)				
	TT ^b /CC+CT	337(64.2)/188(35.8)	159(53.4)/139(46.6)	0.002	1.59(1.19 ~ 2.13)	0.002	

注: 同表 1; ^a调整年龄、性别、吸烟及饮酒状态; ^b基因型为 OR 值计算时的参考基因型

为 1.59 (95% CI: 1.19 ~ 2.13, P=0.002)。未观察到 U2AF35 rs1789956 在病例组和对照组中的基因型分布差异。

连锁不平衡分析的结果显示, rs310441、rs310445 两个位点之间存在强连锁不平衡性(D' = 0.871, LOD=102.46, R²=0.443)。故仅纳入 rs310445 进入第二阶段验证。

3. 验证人群 rs310445 基因型分布及与胰腺癌关联: 第二阶段验证研究在北京地区的病例对照人群中进行。对照组 TT、CT 和 CC 三种基因型的频率分别为 73.4%、23.7% 和 2.9%, 符合 HWE 平衡(P=0.185)。病例组中携带 C 等位基因者频率略高于对照组, 差异无统计学意义(P=0.828), 见表 3。

4. 合并人群 rs310445 基因型分布及其与胰腺癌关联: Breslow-Day 检验结果显示, 筛查和验证阶段 rs310445 基因型分布符合层间一致(P=0.072), 故将两阶段人群进行合并。校正年龄、性别、吸烟、饮酒状况和研究阶段后, 结果显示, 携带 C 等位基因者与携带 TT 基因型者相比, 发生胰腺癌风险增加, OR 值为 1.31(95% CI: 1.07 ~ 1.60, P=0.010), 见表 3。

5. U2AF65 基因的 rs310445 与吸烟交互作用:

吸烟为目前确定的胰腺癌环境危险因素, 进一步分析吸烟与 rs310445 的交互作用。结果发现, 验证阶段和合并阶段均观察到吸烟和 rs310445 C 等位基因之间的协同作用可增加胰腺癌的发生风险, 协同指数 SI 分别为 6.79(95% CI: 3.68 ~ 9.89) 和 2.08 (95% CI: 1.37 ~ 2.78), 但筛查阶段未观察到二者之间存在加法交互, 见表 4。

讨 论

胰腺癌是目前已知的恶性度最高的肿瘤之一, 探讨其遗传易感性及其与环境因素的交互作用用于确定胰腺癌高危人群, 并进行有针对性的预防是目前降低胰腺癌发生风险的有效措施。本研究采用两阶段病例对照研究设计, 结果显示, U2AF65 蛋白的基因多态性及其与吸烟协同作用可能与胰腺癌发生风险相关。

U2AF65 作为 U2 型剪接体的重要组成蛋白^[7], 已在血液病肿瘤如骨髓增生异常综合征^[3]以及肺癌中发现了基因突变^[4]。本研究中发现的 rs310445 位点位于第 19 号染色体的 q13.4 区段, 处于 U2AF65 基因的内含子区。虽然该位点在北方人群验证时未发

表 3 验证及合并人群 U2AF65 rs310445 位点的基因型和等位基因在两组人群中的分布及其与胰腺癌风险关联

基因型	验证人群					合并人群				
	对照组 (n=557)	病例组 (n=413)	P 值	OR 值(95%CI) ^a	P 值 ^a	对照组 (n=1 082)	病例组 (n=711)	P 值	OR 值(95%CI) ^b	P 值 ^b
TT	409(73.4)	297(71.9)	0.828			746(68.9)	456(64.1)	0.101		
CT	132(23.7)	102(24.7)				291(57.0)	223(31.4)			
CC	16(2.9)	14(3.4)				45(4.2)	32(4.5)			
TT ^c /CC+CT	409(73.4)/ 148(26.6)	297(71.9)/ 116(28.1)	0.600	1.08(0.81 ~ 1.44)	0.585	746(68.9)/ 336(31.1)	456(64.1)/ 255(35.9)	0.034	1.31(1.07 ~ 1.60)	0.010

注: 同表 1; ^a调整年龄、性别、吸烟及饮酒状态; ^b调整年龄、性别、吸烟、饮酒及研究阶段; ^c标注的基因型为 OR 值计算时的参考基因型

表 4 U2AF65 基因 rs310445 的多态性与吸烟交互对胰腺癌风险的关联性分析

吸烟	基因型	筛查阶段(n=823)			验证阶段(n=970)			合并阶段(n=1 793)		
		对照组 (百分比, %)	病例组 (百分比, %)	OR 值(95%CI) ^a	对照组 (百分比, %)	病例组 (百分比, %)	OR 值(95%CI) ^a	对照组 (百分比, %)	病例组 (百分比, %)	OR 值(95%CI) ^a
否	TT	258(49.1)	100(33.6)	1	258(46.3)	178(43.1)	1	516(47.7)	278(39.1)	1
是	TT	79(15.1)	59(19.8)	1.99(1.20 ~ 3.30)	151(27.1)	119(28.8)	1.34(0.92 ~ 1.97)	230(21.3)	178(25.0)	1.56(1.15 ~ 2.11)
否	CC+CT	145(27.6)	92(30.9)	1.64(1.15 ~ 2.32)	107(19.2)	63(15.3)	0.85(0.59 ~ 1.23)	252(23.3)	155(21.8)	1.20(0.93 ~ 1.54)
是	CC+CT	43(8.2)	47(15.8)	3.27(1.84 ~ 5.81)	41(7.4)	53(12.8)	2.17(1.27 ~ 3.73)	84(7.8)	100(14.1)	2.65(1.79 ~ 3.93)
SI(95%CI)		1.20(0.49 ~ 1.92)			6.79(3.68 ~ 9.89)			2.08(1.37 ~ 2.78)		
AP(95%CI)		0.11(-0.30 ~ 0.52)			0.47(0.21 ~ 0.73)			0.31(0.08 ~ 0.54)		
RERI(95%CI)		0.33(-0.99 ~ 1.65)			1.05(0.11 ~ 1.99)			0.79(0.03 ~ 1.54)		

注: 同表 3

现其与胰腺癌的风险关联,但在武汉地区和合并人群中携带 C 等位基因者均与胰腺癌风险增加有关,提示该基因 rs310445 的多态性可能与胰腺癌发病风险存在一定的关联,未来的研究还需进一步扩大样本量进行验证。此外,本研究结果显示,该位点的基因多态性与吸烟存在有统计学意义的正交互作用。虽然对于交互作用的解释尚不清楚,但吸烟作为确认的胰腺癌环境危险因素^[8],研究发现吸烟产生的多环芳烃能诱导胰腺癌相关致癌基因如 MDM2 的剪接异构体产生^[9];而剪接体相关蛋白的基因多态性会导致其表达量的异常,从而最终影响其下游基因的正常剪接过程,若影响到肿瘤相关基因的正常剪接,产生的剪接异构体则可能与肿瘤的发生发展密切相关^[10]。提示吸烟可能影响某些肿瘤相关基因的正常剪接,与剪接体 U2AF65 基因多态性产生的效应相协同,从而增加了个体患胰腺癌的发病风险。

U2AF35 作为小亚基,与 U2AF65 共同构成 U2AF 二聚体,它在 U2 核糖核蛋白体依赖的主要剪接体中的功能为识别 pre-mRNA 上 3' 剪接位点的 AG 序列,同 U2AF65 共同介导 pre-mRNA 上 3' 剪接位点的识别^[7]。Ding 等^[11]发现相比于正常胰腺细胞,其在胰腺肿瘤细胞中下调表达,从而影响了胆囊收缩素β受体蛋白基因的正常剪接,错误的保留其第 4 内含子,产生的剪接异构体可能与诱发胰腺癌相关。同时 Je 等^[12]针对韩国人群的研究发现该蛋白的基因突变与血液病肿瘤的发生可能有着密切的联系,但其与实体肿瘤的关系尚无定论^[3, 12, 13]。而本研

究中也未观察到 U2AF35 的基因多态性与胰腺癌风险存在显著性联系,未来还需加大样本量,在多个人群中对该基因与实体肿瘤的关联进行深入地研究。由于病例对照研究的局限性,未来还需更多的重复研究以及生物功能性研究进行验证。

参 考 文 献

- [1] Garraway LA, Lander ES. Lessons from the cancer genome[J]. Cell, 2013, 153(1): 17-37.
- [2] Padgett RA. New connections between splicing and human disease [J]. Trends Genet, 2012, 28(4): 147-154.
- [3] Yoshida K, Sanada M, Shiraiishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia [J]. Nature, 2011, 478(7367): 64-69.
- [4] Imielinski M, Berger AH, Hammerman PS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing[J]. Cell, 2012, 150(6): 1107-1120.
- [5] Blettner M, Sauerbrei W, Schlehofer B, et al. Traditional reviews, meta-analyses and pooled analyses in epidemiology [J]. Int J Epidemiol, 1999, 28(1): 1-9.
- [6] Kallberg H, Ahlbom A, Alfredsson L. Calculating measures of biological interaction using R[J]. Eur J Epidemiol. 2006, 21(8): 571-573.
- [7] Will CL, Luhrmann R. Protein functions in pre-mRNA splicing [J]. Curr Opin Cell Biol, 1997, 9(3): 320-328.
- [8] Hidalgo M. Pancreatic cancer[J]. N Engl J Med, 2010, 362(17): 1605-1617.
- [9] Weng MW, Lai JC, Hsu CP, et al. Alternative splicing of MDM2 mRNA in lung carcinomas and lung cell lines [J]. Environ Mol Mutagen, 2005, 46(1): 1-11.
- [10] Grosso AR, Martins S, Carmo-Fonseca M. The emerging role of splicing factors in cancer[J]. EMBO Rep, 2008, 9(11): 1087-1093.
- [11] Ding WQ, Kuntz SM, Miller LJ. A misspliced form of the cholecystokinin-B/gastrin receptor in pancreatic carcinoma: role of reduced cellular U2AF35 and a suboptimal 3'-splicing site leading to retention of the fourth intron [J]. Cancer Res, 2002, 62(3): 947-952.
- [12] Je EM, Yoo NJ, Kim YJ, et al. Mutational analysis of splicing machinery genes SF3B1, U2AF1 and SRSF2 in myelodysplasia and other common tumors [J]. Int J Cancer, 2013, 133(1): 260-265.
- [13] Thol F, Kade S, Schlarmann C, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes [J]. Blood, 2012, 119(15): 3578-3584.

(收稿日期: 2014-01-10)
(本文编辑: 王岚)