

U2依赖型mRNA剪接体与肝癌的相关性研究

宋伟 朱贝贝 田瑶 钟荣 田晶 缪小平 王丽

【摘要】 目的 探讨U2依赖型mRNA剪接复合体相关基因多态性与肝癌的风险关联。**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U2依赖型mRNA剪接复合体的关键基因(U2AF1、U2AF2、SRSF1、SRSF2、SF3B1、SF3A1、SF1和PRPF40)的21个潜在功能标签SNP位点在筛查人群(武汉地区378例新发肝癌病例和461例非肿瘤对照)中进行基因分型;对筛查阶段得到的阳性位点进一步在验证人群(北京地区428例新发肝癌病例和647例非肿瘤对照)中采用TaqMan基因分型技术进行验证。采用加法交互模型分析吸烟/饮酒及遗传因素的交互作用。**结果** 两阶段人群中均观察到SF3A1基因的rs5994293与肝癌发病风险存在显著关联。合并分析显示,以携带TT基因型者为参照,携带G等位基因者的肝癌发病风险降低30% ($OR=0.70, 95\%CI: 0.58 \sim 0.84, FDR-P=0.0005$);合并阶段在HBsAg阴性者中观察到吸烟合并饮酒和rs5994293 TT基因型之间存在加法交互作用。**结论** 携带SF3A1 rs5994293 T等位基因可能增加肝癌发生风险,但仍需大样本研究及功能研究的验证。

【关键词】 肝癌; U2依赖型mRNA剪接复合体; 吸烟; 饮酒; 单核苷酸多态性

Research on the association between U2-dependent spliceosome gene and hepatocellular cancer Song Wei¹, Zhu Beibei², Tian Yao¹, Zhong Rong², Tian Jing¹, 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 grants from the Research Fund for the Doctoral Program of Higher Education (No. 20111106120020) and National Science and Technology Major Project of China (No. 2012ZX10004904, No. 2013ZX10002002006002).

【Abstract】 Objective To determine the association between U2-dependent spliceosome related 8 key genes and hepatocellular cancer (HCC). **Methods** A two-stage case-control study was conducted. Twenty-two candidate tag single nucleotide polymorphisms (tagSNPs) were genotyped by TaqMan Openarray assay in a screened population that living in Central China (378 HCC incident cases and 461 controls). Frequencies of 4 SNPs (rs2074733, rs9608886, rs7288947 and rs5994293) showed significant difference between cases and controls in the screened population and then genotyped by TaqMan real-time polymerase chain reaction in the validation Chinese Han population from Beijing (428 cases and 647 controls). **Results** The rs5994293 in SF3A1 gene showed a significant association with HCC in both screened population and combined population. Subjects with G allele had a lower risk of HCC, compared to those with the TT genotype. OR appeared to be 0.70 (95% $CI: 0.58-0.84$, false discovery rate adjusted $P=0.0005$) for the combined population. An additive interaction between smoking, drinking alcohol and rs5994293 TT was observed in HBsAg negative subjects of the combined populations. **Conclusion** Our results showed an association existing between SF3A1 rs5994293 and HCC. These findings should be confirmed by further independently large-scale population studies and functional analysis.

【Key words】 Hepatocellular cancer; U2 small nuclear ribonucleoproteins; Smoking; Drinking; Single nucleotide polymorphism

DOI: 10.3760/cma.j.issn.0254-6450.2015.06.021

基金项目:教育部博士点基金(20111106120020); 国家科技重大专项(2012ZX10004904, 2013ZX10002002006002)

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(宋伟、田瑶、田晶、王丽); 华中科技大学同济医学院公共卫生学院流行病学和卫生统计学系(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

原发性肝癌病死率高,目前居于全球癌症死因第三位^[1],其发生是遗传和环境因素共同作用的结果^[2]。mRNA 选择性剪接影响着人类90%以上的基因表达,一旦发生异常,则可能通过影响包括细胞增殖、发育与分化等在内的多个基因的正常剪接过程,在肿瘤发生中扮演着重要角色^[3]。2011年以来, *Nature*、*Cell*等杂志先后报道剪接体基因,尤其是U2核糖核蛋白体依赖型剪接体相关基因(U2AF35、U2AF65、SF1、SRSF1、SF3A1、SF3B1和PRPF40B)的突变与血液肿瘤如慢性髓细胞性白血病以及实体瘤如胰腺癌、乳腺癌以及黑色素瘤等发生有关^[4]。本课题组前期研究也发现U2AF65基因多态性与胰腺癌发生有关^[5]。本研究采用两阶段病例对照研究方法,探讨U2核糖核蛋白体依赖型剪接体相关基因(U2AF1、U2AF2、SRSF1、SRSF2、SF3B1、SF3A1、SF1和PRPF40B)与肝癌易感性的关系。

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378例肝癌患者和461例对照,病例为2009年5月至2012年6月湖北省武汉市同济医院经病理组织确诊的新发肝癌患者。对照来自同期武汉市同济医院参加体检的非肿瘤个体。验证阶段共纳入468例肝癌患者和707例对照,病例为2007年9月至2009年9月北京市佑安医院经病理组织确诊的新发肝癌患者。对照来源于同期在北京市佑安医院参加体检的非肿瘤个体。上述两阶段研究人群病例和对照均按照年龄(± 5 岁)和性别进行匹配。问卷收集性别、年龄、吸烟和饮酒状况等资料。将每天至少吸烟1支,连续或累计吸烟 ≥ 6 个月定义为吸烟;将每天至少饮酒1次,连续饮酒 ≥ 3 个月定义为饮酒。HBsAg携带状况从医院电子病历系统中获得。所有病例和对照均为汉族且无亲缘关系,在资料收集和血样采集前获得所有研究对象签署的知情同意书。

2. 候选SNP挑选:在HapMap中下载中国北京汉族人群(Han Chinese in Beijing, CHB)关于剪接体关键基因U2AF1、U2AF2、SF1、SRSF1、SRSF2、SF3A1、SF3B1和PRPF40B的数据(HapMap Data Release 27 phase I + II + III, <http://www.Hapmap.org>)后采用HaploView 4.2软件,以参数 $R^2 > 0.8$ 、MAF > 0.05 的标准在整个基因上(包括基因上下游各2 kb)筛选能代表这个基因的标签SNP(tagSNPs)82个,然后采用功能预测软件SNP info挑选具有潜

在功能的21个SNP。

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

4. 统计学分析:采用 χ^2 检验比较病例组和对照组间性别、年龄分组、吸烟、饮酒和HBsAg携带状况的差异,对两阶段人群对照组中各位点基因型频率分布进行拟合优度 χ^2 检验判断分型结果是否符合Hardy-Weinberg遗传平衡。按照显性模型并采用非条件logistic回归模型校正吸烟、饮酒和HBsAg因素后分析SNP与肝癌发生的关联。采用错误发现率(false discovery rate, FDR)对所有研究结果进行校正。利用SAS 9.3软件完成各种统计学分析,所有检验均为双侧,以 $P < 0.05$ 为检验水准;运用R语言V3.1.2评价加法交互作用,具体指标包括交互作用的超额危险度(RERI)、归因交互百分比(AP)和协同指数(SI)^[6]。

结果

1. 一般人口学特征:两阶段共纳入肝癌患者806例,对照1108例(表1)。筛查阶段共纳入378例肝癌患者和461例对照,病例组平均年龄(55.7 ± 12.5)岁,男女性别比为3.4:1;对照组平均年龄(55.7 ± 14.0)岁,男女性别比为2.5:1。验证阶段共纳入428例肝癌患者和647例对照,病例组平均年龄(51.7 ± 12.0)岁,男女性别比为5.1:1;对照组平均年龄(51.3 ± 14.6)岁,男女性别比为5.3:1。两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($P > 0.05$);筛查阶段病例组吸烟者所占比例(45.5%)高于对照组(32.1%),差异有统计学意义($P < 0.0001$),而验证阶段两者差异无统计学意义。筛查、验证两阶段均观察到病例组HBsAg阳性者所占比例(分别为69.3%和76.4%)高于对照组(分别为10.0%和2.9%),差异有统计学意义($P < 0.0001$)。

2. 筛查阶段SNP位点与肝癌发病风险的关联分析:对照组21候选个位点中17个位点(表2)符合Hardy-Weinberg遗传平衡,纳入下一步分析。仅观察到SF3A1的4个位点(rs2074733、rs5994293、rs7288947和rs9608886)的基因型分布在病例组和

表1 两阶段肝癌患者和对照组人口学特征及吸烟饮酒状况

变 量	筛查阶段				验证阶段			
	病例组(n=378)	对照组 (n=461)	χ^2 值	P值	病例组 (n=428)	对照组 (n=647)	χ^2 值	P值
年龄组(岁)			5.11	0.164			2.18	0.536
<45	82(21.7)	108(23.4)			123(28.7)	179(27.7)		
45~	91(24.1)	103(22.4)			135(31.6)	185(28.6)		
55~	110(29.1)	109(23.6)			108(25.2)	172(26.6)		
≥65	95(25.1)	141(30.6)			62(14.5)	111(17.1)		
性别			3.23	0.073			0.07	0.796
男	292(77.3)	331(71.8)			358(83.6)	545(84.2)		
女	86(22.7)	130(28.2)			70(16.4)	102(15.8)		
吸烟			15.80	<0.000 1			0.56	0.456
否	206(54.5)	313(67.9)			213(49.8)	337(52.1)		
是	172(45.5)	148(32.1)			215(50.2)	310(47.9)		
饮酒			2.28	0.131			3.80	0.051
否	237(62.7)	312(67.7)			256(59.8)	348(53.8)		
是	141(37.3)	149(32.3)			172(40.2)	299(46.2)		
HBsAg			321.47	<0.000 1			637.44	<0.000 1
阴性	115(30.4)	375(81.3)			100(23.4)	626(96.8)		
阳性	262(69.3)	46(10.0)			327(76.4)	19(2.9)		
不详	1(0.3)	40(8.7)			1(0.2)	2(0.3)		

注:括号外数据为人数,括号内数据为百分比(%)

表2 筛查阶段17个SNPs位点基因型频率在肝癌患者与对照组中的分布差异及其与肝癌发生风险关联

基因	SNP	基因型	病例组 (n=378)	对照组 (n=461)	OR值(95%CI) ^a	P值 ^a	FDR-P值 ^b
PRPF40B	rs4073998	AG+AA/GG ^c	59/319	59/402	1.39(0.84~2.29)	0.198	0.410
PRPF40B	rs8626	AG+GG/AA ^c	57/321	63/398	1.35(0.83~2.22)	0.231	0.428
SF1	rs474707	CT+TT/CC ^c	243/135	276/185	1.20(0.83~1.74)	0.321	0.513
SF3A1	rs12484880	TG+GG/TT ^c	68/310	73/388	1.14(0.72~1.82)	0.579	0.730
	rs2074733	CT+TT/CC ^c	173/205	313/148	0.36(0.25~0.53)	1.33E-10	<0.000 1
	rs5753081	CT+TT/CC ^c	26/352	22/439	1.21(0.57~2.57)	0.626	0.756
	rs5994293	TG+GG/TT ^c	132/246	234/227	0.42(0.29~0.61)	5.19E-06	<0.000 1
	rs7288947	CT+TT/CC ^c	90/288	165/296	0.54(0.41~0.72)	2.12E-04	0.001
	rs8141656	TC+CC/TT ^c	206/172	241/220	1.23(0.86~1.76)	0.251	0.428
	rs9608886	TG+GG/TT ^c	29/349	90/371	0.17(0.10~0.31)	8.08E-07	<0.000 1
SF3B1	rs11683572	CG+GG/CC ^c	154/224	211/250	0.94(0.66~1.35)	0.737	0.792
SRSF1	rs2233911	AG+GG/AA ^c	162/216	164/297	1.13(0.78~1.62)	0.523	0.717
	rs8819	CT+TT/CC ^c	24/354	24/437	1.15(0.53~2.46)	0.725	0.792
SRSF2	rs237059	CT+TT/CC ^c	10/368	5/456	2.75(0.66~11.54)	0.166	0.401
U2AF1	rs1789956	GA+AA/GG ^c	36/342	51/410	0.67(0.38~1.21)	0.185	0.410
U2AF2	rs310441	TG+GG/TT ^c	202/176	249/212	1.07(0.75~1.52)	0.730	0.792
	rs310445	TC+CC/TT ^c	132/246	161/300	1.18(0.81~1.70)	0.387	0.561

注:^a校正年龄、性别、吸烟、饮酒;^b经FDR多重校正后的P值(校正17次);^c参照基因型

对照组中差异有统计学意义。校正性别、年龄、吸烟、饮酒和HBsAg因素后,携带SF3A1 rs2074733 T等位基因的个体较携带野生纯合型CC基因型的个体发生肝癌风险降低64%(OR=0.36, 95%CI: 0.25~0.53, FDR-P<0.000 1)。携带SF3A1 rs5994293 G等位基因者较携带野生纯合型TT基因型者发生肝癌风险降低58%(OR=0.42, 95%CI: 0.29~0.61, FDR-P<0.000 1)。携带SF3A1 rs7288947 T等位基

因者较CC基因型者发生肝癌风险降低46%(OR=0.54, 95%CI: 0.41~0.72, FDR-P=0.001)。携带SF3A1 rs9608886 G等位基因者较TT基因型者发病风险降低83%(OR=0.17, 95%CI: 0.10~0.31, FDR-P<0.000 1)。

3. 验证阶段单位点与肝癌发病风险的关联分析: 4个位点rs2074733、rs5994293、rs7288947和rs9608886在验证阶段人群中进行验证。其在对照

组中的基因型频率分布均满足 Hardy-Weinberg 遗传平衡。校正性别、年龄、吸烟、饮酒和 HBsAg 因素后,携带 rs5994293 G 等位基因者较 TT 基因型者发生肝癌风险降低 34% ($OR=0.66, 95\% CI: 0.44 \sim 0.99, P=0.042$), 与筛查阶段的结果一致。但经 FDR 校正后差异无统计学意义 ($FDR-P=0.116$)。其余 3 个位点均未发现与肝癌发病风险存在关联(表 3)。

4. 合并两阶段人群 SF3A1 rs5994293 与肝癌发病风险的关联分析:将两阶段人群合并分析,校正性别、年龄、吸烟、饮酒和 HBsAg 因素后,携带 rs5994293 G 等位基因者较 TT 基因型者发生肝癌风险降低 30% ($OR=0.70, 95\% CI: 0.58 \sim 0.84, FDR-P=0.0005$)。

5. 合并阶段按照 HBsAg 因素分层后 SF3A1 rs5994293 与合并吸烟饮酒的交互作用:合并阶段按照 HBsAg 因素分层后研究结果显示,HBsAg 阴性者中合并吸烟饮酒与 SF3A1 rs5994293 TT 基因型之间存在协同效应,加法交互协同指数 SI 为 2.27 ($95\% CI: 1.43 \sim 3.12$)。但未观察到 HBsAg 阳性者中合并吸烟饮酒与上述基因型之间存在加法交互作用(表 4)。

讨 论

本研究发现 U2 依赖型 mRNA 剪接复合体相关基因中 SF3A1 rs5994293 可能与肝癌发病相关,携带 G 等位基因的个体较携带 TT 基因型者发病风险降

低 30%, 且 SF3A1 的基因多态性及其与吸烟的协同作用可能与肝癌发生风险相关。

SF3A1 是剪接体催化核心 U2-snRNP 的重要组成部分,后者能识别 3' 端剪接位上游的分支位点,定位 mRNA 上被剪切除去的内含子,从而决定内含子的切除位置正确与否^[7]。SF3A1 主要功能是与 U2-snRNP 结合,共同参与 3' 剪接位点的识别,其发生突变将会影响 U2-snRNP 与分支点结合,从而最终抑制剪接过程^[7]。目前慢性淋巴细胞性白血病和骨髓增生异常综合征以及实体瘤包括肺癌、乳腺癌以及胰腺癌患者中均发现 SF3A1 基因突变^[8-9]。SF3A1 的表达研究显示人类非小细胞、小细胞肺癌细胞中 SF3A1 蛋白表达量较人正常气管上皮细胞高 4.7 ~ 11.0 倍^[10],Chin 等^[11]在头颈部癌以及 Rimkus 等^[12]在晚期直肠癌中也发现类似现象。功能分析显示,rs5994293 为转录因子结合位点,其基因多态可能会影响到 SF3A1 的表达,而 SF3A1 的异常表达可能会改变剪切位点的选择^[13],从而最终影响原癌基因和抑癌基因的正常剪接,最终直接或间接地导致肿瘤发生。有意思的是,SF3A1 基因位于染色体 22q.12.2 上,Hu 等^[14]对中国汉族 GWAS 研究中发现 22q.12.2 这一区段 2 个 SNP 位点 rs17728461 与 rs36600 与肺癌发生密切相关。结合本研究结果,提示 22q.12.2 这一区段可能与肿瘤发生密切相关。

本研究在合并阶段 HBsAg 阴性者中观察到合并吸烟饮酒与 SF3A1 rs5994293 有协同作用。目前对于交互作用的解释尚不清楚。大量流行病学研究

表 3 验证阶段 SF3A1 4 个 SNPs 位点基因型频率在肝癌患者与对照组中的分布差异及其与肝癌发生风险的关联

SNP	基因型	病例组(n=428)	对照组(n=647)	OR 值(95%CI)	P 值 ^a	FDR-P 值 ^b
rs2074733	CT+TT/CC	302/126	437/210	1.14(0.75 ~ 1.73)	0.544	0.717
rs5994293	TG+GG/TT	178/250	289/358	0.66(0.44 ~ 0.99)	0.042	0.116
rs7288947	CT+TT/CC	105/323	157/490	1.04(0.66 ~ 1.64)	0.866	0.897
rs9608886	TG+GG/TT	20/408	26/621	1.53(0.65 ~ 3.62)	0.336	0.513

注:^a 同表 2; ^b 经 FDR 多重校正后的 P 值(校正 21 次); ^c 野生基因型

表 4 合并两阶段人群吸烟、饮酒与 rs5994293 位点的相加交互作用与肝癌发生风险的关联

合并吸烟饮酒	rs5994293	HBsAg(+)(n=654)		HBsAg(-)(n=1216)	
		病例/对照	OR 值(95%CI) ^a	病例/对照	OR 值(95%CI) ^a
否	GG+GT	160/26	1.00	47/338	1.00
是	GG+GT	90/4	3.36(1.12 ~ 10.12)	13/132	1.14(0.58 ~ 2.25)
否	TT	228/28	1.37(0.77 ~ 2.44)	113/397	2.07(1.41 ~ 3.03)
是	TT	111/7	2.59(1.06 ~ 6.33)	42/134	3.75(2.26 ~ 6.22)
交互作用分析					
	RERI 值(95%CI)		-1.13(-5.18 ~ 2.92)		1.54(0.05 ~ 3.03) ^b
	AP 值(95%CI)		-0.44(-2.16 ~ 1.29)		0.41(0.13 ~ 0.69) ^b
	SI 值(95%CI)		-0.59(-1.18 ~ 2.35)		2.27(1.43 ~ 3.12) ^b

注:^a 经年龄、性别、研究阶段调整; ^b 有统计学意义

表明,嗜酒是肝癌发生的危险因素,酒精在人体内一系列代谢产物能与肝脏DNA形成螯合物从而引起细胞DNA的损伤,如P53基因损伤;而现烟草中含有多种致癌化学物质,与肺癌、口腔癌、胰腺癌、食管癌、胃癌和咽喉癌等13种癌症发生有关。吸烟产生的多环芳烃能诱导某些原癌基因如MDM2的剪接异构体产生,而MDM2是调节P53通路的重要基因,且已证实P53基因突变是肝癌发生的原因之一^[15],提示吸烟、饮酒可能会导致肝癌相关的原癌基因或抑癌基因的选择性剪切,从而增加携带SF3A1 rs5994293 TT基因型者发生肝癌的风险。HBsAg阳性人群中未观察到合并吸烟饮酒与SNP的交互作用,其原因可能与HBV感染自身与肝癌发生效应太强($OR=7.59, 95\% CI: 6.13 \sim 9.40$)或样本量不够的缘故。有待加大样本量,在多个人群中对该基因与肝癌的发生风险进行深入地研究。

参 考 文 献

- [1] Jemal A, Bray F, Center MM, et al. Global cancer statistics [J]. *CA Cancer J Clin*, 2011, 61(2): 69–90.
- [2] Yu MC, Yuan JM. Environmental factors and risk for hepatocellular carcinoma [J]. *Gastroenterology*, 2004, 127(5 Suppl 1): S72–78.
- [3] Wang ET, Sandberg R, Luo SJ, et al. Alternative isoform regulation in human tissue transcriptomes [J]. *Nature*, 2008, 456(7221): 470–476.
- [4] Padgett RA. New connections between splicing and human disease [J]. *Trends Genet*, 2012, 28(4): 147–154.
- [5] Tian J, Zhu BB, Tian Y, et al. Association between pancreatic cancer risk and the interaction of U2AF65 gene polymorphisms and smoking [J]. *Chin J Epidemiol*, 2014, 35(6): 710–713. (in Chinese)
田晶,朱贝贝,田瑶,等.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[J]. *中华流行病学杂志*, 2014, 35(6): 710–713.
- [6] Källberg H, Ahlbom A, Alfredsson L. Calculating measures of biological interaction using R [J]. *Eur J Epidemiol*, 2006, 21(8): 571–573.
- [7] Tanackovic G, Krämer A. Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo [J]. *Mol Biol Cell*, 2005, 16(3): 1366–1377.
- [8] Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer [J]. *Nature*, 2013, 500(7463): 415–421.
- [9] Ellis MJ, Ding L, Shen D, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition [J]. *Nature*, 2012, 486(7403): 353–360.
- [10] Difilippantonio S, Chen Y, Pietas A, et al. Gene expression profiles in human non-small and small-cell lung cancers [J]. *Eur J Cancer*, 2003, 39(13): 1936–1947.
- [11] Chin D, Boyle GM, Williams RM, et al. Novel markers for poor prognosis in head and neck cancer [J]. *Int J Cancer*, 2005, 113(5): 789–797.
- [12] Rimkus C, Friederichs J, Boulesteix AL, et al. Microarray-based prediction of tumor response to neoadjuvant radiochemotherapy of patients with locally advanced rectal cancer [J]. *Clin Gastroenterol Hepatol*, 2008, 6(1): 53–61.
- [13] Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code [J]. *Nat Rev Mol Cell Biol*, 2005, 6(5): 386–398.
- [14] Hu ZB, Wu C, Shi YY, et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese [J]. *Nat Genet*, 2011, 43(8): 792–796.
- [15] Bressan B, Kew M, Wands J, et al. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa [J]. *Nature*, 1991, 350(6317): 429–431.

(收稿日期: 2014-12-15)

(本文编辑: 王岚)