利用公共数据深入挖掘食管鳞状细胞癌候选肿瘤抑制基因

孙玉琳[1],陈万青[2],赵晓航[1,3]
(1.中国医学科学院 北京协和医学院 肿瘤研究所 分子肿瘤学国家重点实验室,北京 100021;
2.中国医学科学院 北京协和医学院 全国肿瘤防治研究办公室,北京 100021;
3.海军总医院,北京 100048)

摘　要:[目的]自20世纪90年代开始,人们利用杂合性缺失(loss of heterozygosity,LOH)技术寻找引起食管鳞状细胞癌(esophageal squamous cell carcinoma,ESCC)癌变的肿瘤抑制基因。但除了TP53、CDKN2A等以外,目前很多高频LOH区域的抑癌基因仍未发现。随着近年来生物学信息公共数据库的建设发展,有必要重新筛选分析ESCC中的潜在抑癌基因。[方法]利用Pubmed检索1991~2012年间的食管癌LOH文章,整理信息,综合考虑LOH频率和文献重复鉴定次数后,利用Circos软件绘制并确认了ESCC中高度可信LOH热点区域。对每一区域,利用人类基因组数据库、肿瘤抑制基因数据库和人类蛋白质图集数据库分析其中的候选抑癌基因。[结果]基于85篇针对ESCC的LOH相关文献,共确定10个ESCC高度可信LOH热点区域,主要累及3p、5q、9p、13q、17p、17q等染色体臂。基于3个公共数据库的分析,在每个区域中均确定了一个以上的候选抑癌基因,其中很多基因从未在ESCC中受到关注。[结论]综合考虑文献中对LOH区域的可重复鉴定性可避免原始文献质量和样本选择偏倚等的干扰。利用公共数据库深入挖掘ESCC中的候选抑癌基因,为后续选择合适的靶基因开展深入的表达和功能分析提供了理论依据。

关键词:食管癌;肿瘤抑制基因;公共数据库
中图分类号:R735.1　　**文献标识码**:A　　**文章编号**:1004-0242(2013)06-0481-12

Deciphering the Potential Tumor Suppressor Genes in Esophageal Squamous Cell Carcinoma Based on Public Database

SUN Yu-lin[1],CHEN Wan-qing[2],ZHAO Xiao-hang[1,3]
(1.State Key Laboratory of Molecular Oncology,Cancer Institute & Hospital,Chinese Academy of Medical Sciences & Peking Union Medical College,Beijing 100021,China;2. National Office for Cancer Prevention and Control & National Central Cancer Center,Chinese Academy of Medical Sciences & Peking Union Medical College,Beijing 100021,China;3. Navy General Hospital,Beijing 100048,China)

Abstract: [Purpose] Since the early years of 1990s,people used the loss of heterozygosity (LOH) technology to find the tumor suppressor genes which led to tumorigenesis of esophageal squamous cell carcinoma(ESCC). However,the tumor suppressor genes in several high-frequency LOH regions remained to be discovered,except for TP53 and CDKN2A etc. With the development of public databases in recent years,it is necessary to re-analyze the potential tumor suppressor genes in ESCC. [Methods] Pubmed retrieved using the key phrase 'esophageal squamous cell carcinoma AND Loss of Heterozygosity'. Literature published between 1991 and 2012 were obtained. After considering the LOH frequency and repeatability in literature,the high-confidence LOH hot regions in ESCC were showed and confirmed using the Circos software. For each region,the candidate tumor suppressor genes were analyzed using three public databases,including the human genome database,the tumor suppressor gene database and the human protein atlas database. [Results] Based on 85 papers,a total of 10 high-confidence LOH hot regions in ESCC were identified,which mainly involved in 3p, 5q,9p,13q,17p,17q chromosomes. After the analysis based on three public databases,one or more candidates tumor suppressor genes were determined,many of which had never been concerned in ESCC. [Conclusion] Considering the repeatable identification of LOH regions in the literature can avoid the interference of the poor papers quality and bias on sample selection. Deciphering the candidate tumor suppressor genes in ESCC using public databases can provide a theoretical basis for the in-depth genes expression and functional analysis.

Key words: esophageal cancer;tumor suppressor gene;public database

收稿日期:2013-03-14
基金项目:国家863高技术研究发展计划(2012AA020206);国家自然科学基金(81071789)
通讯作者:赵晓航,E-mail: zhaoxh@cicams.ac.cn

食管癌是世界常见恶性肿瘤，位居癌症死因第6位，每年新增患者数超过48万例，主要高发于东亚、南部和东部非洲等地[1]。食管腺癌主要高发于英国、美国、澳大利亚等发达国家，而食管癌高发区中主要以食管鳞状细胞癌（esophageal squamous cell carcinoma，ESCC）为主，特别是在东亚地区，包括中国在内，ESCC占全部食管恶性肿瘤的90%以上[2]。根据国际癌症研究机构（International Agency for Research on Cancer，IARC）的估计，我国食管癌的发病率和死亡率均排在全部恶性肿瘤的第4位，年死亡人数超过21万人。因此，深入探讨我国食管鳞状细胞癌的分子遗传基础，对于食管癌的早期预警、早期诊断和临床治疗等将产生重大影响。

杂合性缺失（loss of heterozygosity，LOH）是指与同一个体的正常组织相比，肿瘤组织丢失一个亲代等位基因，由杂合子变成纯合子状态的现象[3,4]。在肿瘤形成过程中，LOH通常是肿瘤抑制基因（tumor suppressor genes，TSGs）失活的主要机制。LOH形成主要来自该染色体区域拷贝数缺失事件，比如半合子缺失等，此外，还有相当一部分LOH来自拷贝数中性变化事件，比如由有丝分裂重组或不分离形成的染色体复制等[4]。

从20世纪90年代起，人们开始探讨LOH在ESCC中的分布，揭示了ESCC患者在多条染色体臂上存在高频（30%以上）的LOH热点，如1p、3p、5q、8p、9p、9q、13q、16q、17p、17q、18q等[5]。但是，除了17p13区域的*TP53*基因、3p14区域的*FHIT*基因以及9p21区域的*CDKN2A*基因以外，其余LOH热点区对应的候选TSGs目前还不十分明确。这主要表现在部分热点区缺乏已知的TSGs，或已知的TSGs在食管癌中的表达或突变情况不符合人们对TSGs的传统认识等。此外，近年来，几十个非编码RNAs（non-coding RNAs）被人们认定为新的TSGs，而食管癌中的LOH是否与这些非编码RNAs有关，目前还不清楚。本研究基于目前的85篇ESCC相关的LOH文章，从LOH频率和研究报道次数两个角度分析ESCC中高度可信的LOH热点区域，而后基于文献调研、蛋白表达公共数据库等信息，筛选了各个区域中可能的候选TSGs，为深入揭示这些分子在ESCC癌变中的分子机制奠定了基础。

1 资料与方法

1.1 数据收集

在Pubmed中，利用关键词“("Esophageal Neoplasms" [Mesh] AND "Loss of Heterozygosity" [Mesh])”检索，得到了257篇1991~2012年的食管癌LOH相关文献。阅读文献，收集整理文献中报道的肿瘤类型（食管腺癌或ESCC）、研究人群、样本量、检测方法、LOH区域、频率等基本信息。筛选保留85篇针对ESCC的LOH研究文章。

在85篇ESCC的LOH研究文章中，基于中国人群的研究有44篇，占51.8%；基于日本人群的研究有21篇，占24.7%；基于其他人群的研究有20篇，占23.5%。这些研究涉及样本例数最高为132例，最低只有3例，中位值为36例。

1.2 数据处理

利用生物信息学软件Circos v0.62绘制ESCC中各条染色体上LOH区域的位置、频率和该区域被文章重复鉴定次数等信息。将文献中报道的LOH频率超过50%，且文献重复鉴定次数超过7次的区域，认定为ESCC中高度可信的LOH热点区域。

1.3 ESCC候选TSGs的筛选

利用人类基因组数据库NCBI查找高度可信LOH热点区域的全部编码基因，利用肿瘤抑制基因数据库[6]筛选这些区域中已报道过的包括食管癌在内的恶性肿瘤中的TSGs，利用文献检索查找这些编码基因的主要功能、在ESCC中的表达、突变和启动子甲基化情况。最后，利用人类蛋白质图集数据库（human protein atlas，HPA）查找在正常食管上皮和一半以上的消化道器官中呈中—高表达（moderate-strong），而在70%以上的常见肿瘤组织中阴性~中度表达（negative-weak-moderate）的全部蛋白，将它们对应到ESCC高度可信LOH热点区域的编码基因上，从蛋白表达的角度评价哪些基因有成为潜在TSGs的可能。

2 结 果

2.1 LOH在ESCC中的分布

在ESCC中，LOH在各条染色体上的分布并不均衡（Figure 1），12号和22号染色体完全没有LOH

的相关报道，而 20 号和 21 号染色体仅有 1 篇和 3 篇在染色体臂层次上的报道，1 号、6 号、7 号、10 号、14 号、15 号、16 号和 19 号染色体上仅 2~4 篇相关报道，虽然很多区域报道的 LOH 频率高达 60%以上甚至 100%，但不同报道之间缺乏重复性，提示这些染色体上的 LOH 可能与研究使用样本的差异有关。

2.2　ESCC 中高可信度 LOH 热点区域

综合考虑每一染色体区段上的 LOH 频率和文献重复鉴定次数，取 LOH 频率>50%，且文献重复鉴定次数超过 7 次的区域，认定为 ESCC 中高度可信的 LOH 热点区域，包括 3p24.3、3p14、5q21~q22.1、9p21、13q11~q12.1、13q12.3~q14.3、13q32~q34、17p13、17p11.2、17q25.1 等，主要累及 3p、5q、9p、13q、17p、17q 等染色体臂(Figure 1)。

2.2.1　3p24.3

共有 10 篇文献在 ESCC 中鉴定到了 3p24.3 区域的 LOH，频率在 67%~86%之间。这一区域全长 7.5Mb，编码 35 个已知基因，包括 10 个蛋白编码基因，11 个非编码 RNA 基因和 14 个假基因。这一区域没有已报道过的人类 TSGs，这些基因也没有在食管癌中存在异常表达的相关报道(Table 1)。利用 HPA 数据库检索发现，这一区域中只有 1 个蛋白编码基因在食管上皮中高表达，且在常见肿瘤中表达降低，即细胞核蛋白 PP2D1(Figure 2)。

2.2.2　3p14

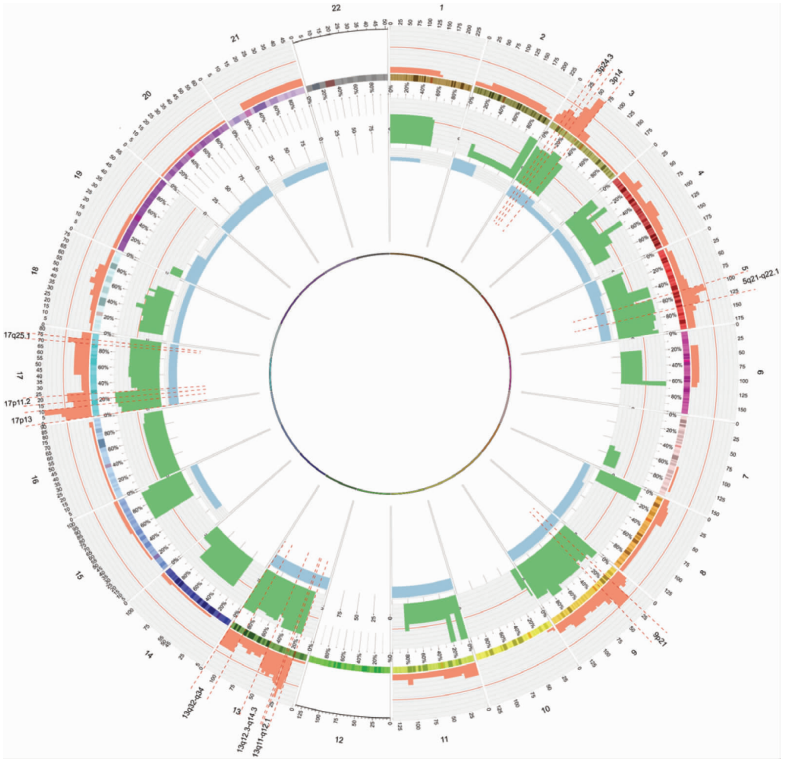
共有 19 篇文献在 ESCC 中鉴定到了 3p14 全长及部分区域的 LOH，频率在 57%~100%之间。这一区域全长 15.4Mb，编码 85 个已知基因，包括 43 个蛋白编码基因，19 个非编码 RNA 基因和 23 个假基因。这一区域包含 3 个已报道过的人类 TSGs，分别是 FHIT、ADAMTS9 和 FOXP1。在 ESCC 中的研究发现，FHIT 和 ADAMTS9 蛋白在肿瘤组织中表达下调，肿瘤组织和肿瘤细胞系中可检测到 2 个基因启动子区的高甲基化(Table 1)，提示这两个基因都可能是食管癌中该区域的 TSGs(Figure 2)。

2.2.3　5q21~q22.1

共有 14 篇文献在 ESCC 中鉴定到 5q21~q22.1 全长及部分区域的 LOH，频率在 29%~91%之间。这一区域全长 13.3Mb，编码 51 个已知基因，包括 20 个蛋白编码基因，9 个非编码 RNA 基因和 22 个假基因。这一区域包含 2 个已报道过的人类 TSGs，分别是 EFNA5 和 APC。在 ESCC 中的研究发现，APC 蛋白在肿瘤组织中表达下调，肿瘤组织中可检测到 APC 基因启动子区的高甲基化(Table 1)，提示 APC 可能是食管癌中该区域的 TSGs。

2.2.4　9p21

共有 13 篇文献在 ESCC 中鉴定到了 9p21 全长及部分区域的 LOH，频率在 17%~83%之间。这一区域全长 13.3Mb，编码 78 个已知基因，包括 33 个蛋



All of human chromosomes were highlighted by different colors. The outer histogram represents the repeatability in literature and marks orange; the inner histogram(marks green) represents the LOH frequency, and the innermost histogram(marks blue) represents the LOH frequency reported in whole chromosome arms. Ten high-confidence LOH hot regions in ESCC were labeled by red dotted line.

Figure 1　Visualization of LOH region reported in ESCC on human chromosomes

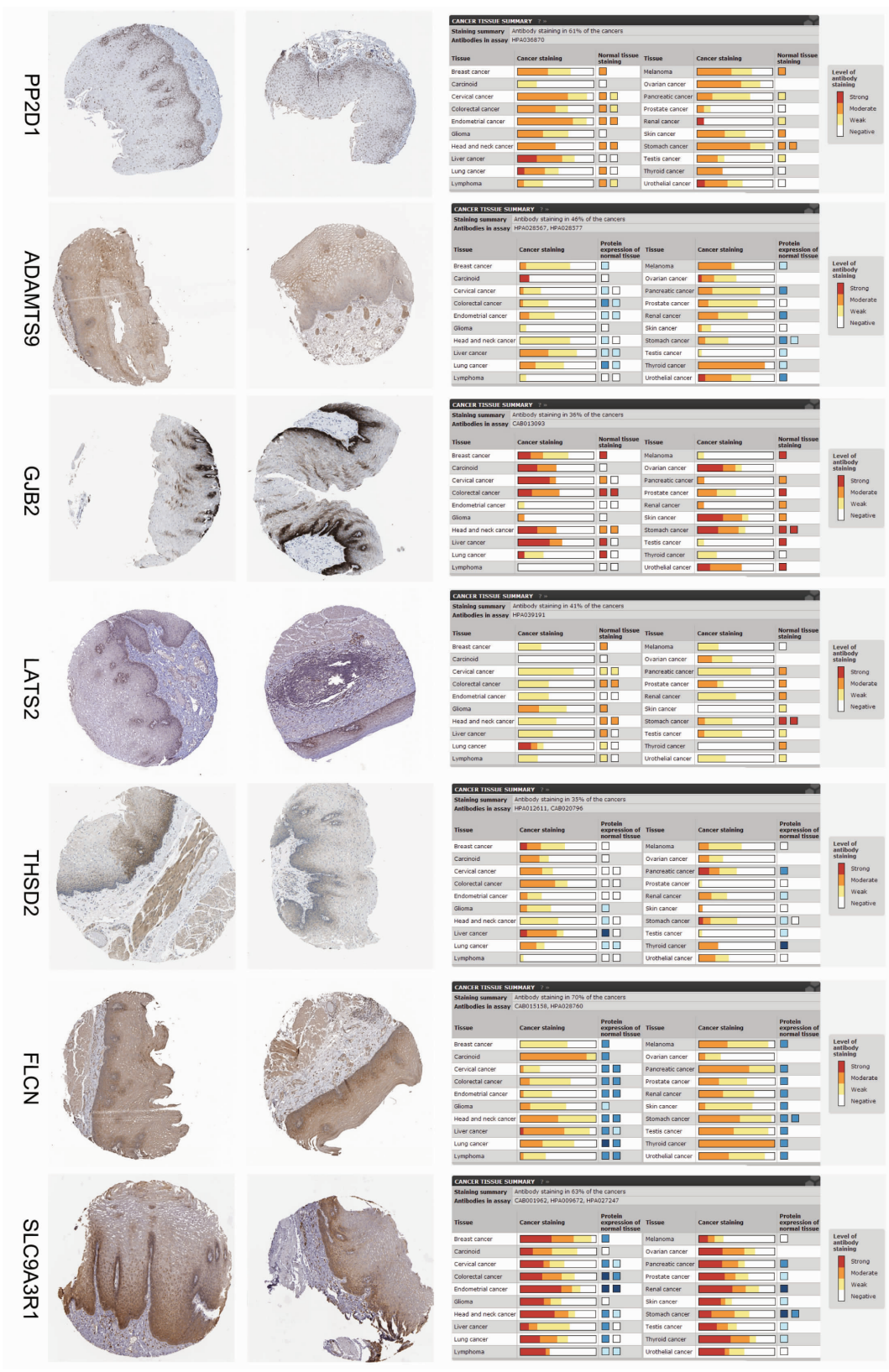


Figure 2　The protein expression of candidate tumor suppressor genes of each high-confidence LOH hot regions in tumor tissues and the normal esophagus mucosa.(Data are from Human Protein Atlas database)

Table 1 The analysis of candidate tumor suppressor genes in 10 high-confidence LOH hot regions in ESCC

Hotspot	Involved genes	Known TSG	Gene name	Type	Mutation	Inactivated mechanisms	Major function	Aberrant exp in ESCC
3p24.3	35							
3p14	85	FHIT	Fragile histidine triad gene	Protein-coding		Hypermethylation [12-15]	Hydrolase involved in purine metabolism.	Down[16, 17]
		ADAMTS9	ADAM metallopeptidase with thrombospondin type 1 motif, 9	Protein-coding		Hypermethylation in ESCC cell lines [18]	Cleaves the large aggregating proteoglycans, aggrecan and versican.	Down[18]
		FOXP1	Forkhead box P1	Protein-coding			Transcription factor	
5q21~q22.1	51	EFNA5	Ephrin-A5	Protein-coding			Cell surface GPI-bound ligand for Eph receptors, a family of receptor tyrosine kinases.	
		APC	Adenomatous polyposis coli	Protein-coding		Hypermethylation [19, 20]	Promotes rapid degradation of CTNNB1 and participates in Wnt signaling.	Down[21]
9p21	78	CDKN2A	Cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	Protein-coding	Frequent[22, 23]	Hypermethylation [12, 24-28]	Inducing cell cycle arrest in G1 and G2 phases.	Down[12, 27, 29, 30]
		CDKN2B	Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	Protein-coding		Homozygous deletions [31~33] Hypermethylation [34]	Cyclin-dependent kinase inhibitor	
		MTAP	Methylthioadenosine phosphorylase	Protein-coding		Homozygous deletions with CDKN2A[35]	Enzyme that plays a major role in polyamine metabolism and is important for the salvage of both adenine and methionine.	
		PTENP1	Phosphatase and tensin homolog pseudogene 1	lncRNA			Exerts a tumor suppressive function by acting as a decoy for PTEN-targeting miRNAs.	
		TOPORS	Topoisomerase Ⅰ binding, arginine/serine-rich, E3 ubiquitin protein ligase	Protein-coding			E3 ubiquitin-protein ligase.	
13q11~q12.1	163	IFT88	Intraflagellar transport 88 homolog (chlamydomonas)	Protein-coding			Involved in primary cilium biogenesis (by similarity).	
		PDX1	Pancreatic and duodenal homeobox 1	Protein-coding			Transcription factor	Up[36]

Continuous Table 1 The analysis of candidate tumor suppressor genes in 10 high-confidence LOH hot regions in ESCC

Hotspot	Involved genes	Known TSG	Gene name	Type	Mutation	Inactivated mechanisms	Major function	Aberrant exp in ESCC
		GJB2	Gap junction protein, beta 2, 26kDa	Protein-coding			Gap junction protein family member	Up[37], Down[38]
		LATS2	LATS, large tumor suppressor, homolog 2 (Drosophila)	Protein-coding	Rare[39]		Serine/threonine-protein kinase	
13q12.3~q14.3	314	BRCA2	Breast cancer 2, early onset	Protein-coding	Frequent[40]; Infreqent [41, 42]		Involved in double-strand break repair and/or homologous recombination, S phase checkpoint activation.	Up[43]
		STARD13	StAR-related lipid transfer (START) domain containing 13	Protein-coding			Involved in regulation of cytoskeletal reorganization, cell proliferation, and cell motility.	
		PDS5B	PDS5, regulator of cohesion maintenance, homolog B (S. cerevisiae)	Protein-coding	Rare[44]		Regulator of sister chromatid cohesion in mitosis which may stabilize cohesin complex association with chromatin.	Down[45]
		DLEU2	Deleted in lymphocytic leukemia 2 (non-protein coding)	lncRNA			May act as a tumor suppressor	
		TRIM13	Tripartite motif containing 13	Protein-coding			E3 ubiquitin-protein ligase	
		TSC22D1	TSC22 domain family, member 1	Protein-coding			Transcription repressors	
		FOXO1	Forkhead box O1	Protein-coding			Transcription factor	
		KCNRG	Potassium channel regulator	Protein-coding			Inhibits potassium fluxes in cells.	
		ARL11	ADP-ribosylation factor-like 11	Protein-coding			Play a role in apoptosis in a caspase-dependent manner.	
		THSD1	Thrombospondin, type I, domain containing 1	Protein-coding		Hypermethylation [46]	Involved in the complement pathway, as well as in extracellular matrix proteins.	Down[46]
		MIR15A	microRNA 15a	miRNA				
		MIR16-1	microRNA 16-1	miRNA				
		RB1	Retinoblastoma 1	Protein-coding		Hypermethylation [34]	Key regulator of entry into cell division that acts as a tumor suppressor. Directly involved in heterochromatin formation.	Down[47, 48]
		DLEU1	Deleted in lymphocytic leukemia 1 (non-protein coding)	lncRNA			May act as a tumor suppressor	
		INTS6	Integrator complex subunit 6	Protein-coding	Infrequent[49]		Component of the Integrator complex, a complex involved in the small nuclear RNAs (snRNA) U1 and U2 transcription and in their 3′-box-dependent processing.	

Continuous Table 1 The analysis of candidate tumor suppressor genes in 10 high-confidence LOH hot regions in ESCC

Hotspot	Involved genes	Known TSG	Gene name	Type	Mutation	Inactivated mechanisms	Major function	Aberrant exp in ESCC
13q32~q34	189	CUL4A	Cullin 4A	Protein-coding			Core component of multiple cullin-RING-based E3 ubiquitin-protein ligase complexes	
		ING1	Inhibitor of growth family, member 1	Protein-coding	Frequent[50]		Nuclear protein that physically interacts with TP53, induce cell growth arrest and apoptosis.	Down[50, 51]
17p13	280	GABARAP	GABA(A) receptor-associated protein	Protein-coding			Ligand-gated chloride channels that mediate inhibitory neurotransmission.	
		MIR195	microRNA 195	miRNA				
		TP53	Tumor protein p53	Protein-coding	High[52-55]	Rare hypermethylation [34]	Transcription factor	Up[25, 56-60]
		KCTD11	Potassium channel tetramerisation domain containing 11	Protein-coding			Induces apoptosis, growth arrest and the expression of CDKN1B. Acts as an E3 ubiquitin-protein ligase towards HDAC1.Functions as antagonist of the Hedgehog pathway on cell proliferation and differentiation.	
		PFN1	Profilin 1	Protein-coding			Binds to actin and affects the structure of the cytoskeleton.	Up[61, 62]
		XAF1	XIAP associated factor 1	Protein-coding		Hypermethylation [63]	Negative regulator of members of the IAP (inhibitor of apoptosis protein) family	Down[63]
		DPH1	DPH1 homolog (S. cerevisiae)	Protein-coding			Required for the first step in the synthesis of diphthamide, a post-translational modification of histidine which occurs in translation elongation factor 2.	
		HIC1	Hypermethylated in cancer 1	Protein-coding			Transcription factor	
		MIR22	MicroRNA 22	miRNA				
		MYBBP1A	MYB binding protein (P160) 1a	Protein-coding			Transcription factor	

Continuous Table 1 The analysis of candidate tumor suppressor genes in 10 high-confidence LOH hot regions in ESCC

Hotspot	Involved genes	Known TSG	Gene name	Type	Mutation	Inactivated mechanisms	Major function	Aberrant exp in ESCC
		PAFAH1B1	Platelet-activating factor acetylhydrolase 1b, regulatory subunit 1 (45kDa)	Protein-coding			Required for proper activation of Rho GTPases and actin polymerization at the leading edge of locomoting cerebellar neurons and postmigratory hippocampal neurons.	
		SMYD4	SET and MYND domain containing 4	Protein-coding				
		VPS53	Vacuolar protein sorting 53 homolog (S. cerevisiae)	Protein-coding			May be involved in retrograde transport of early and late endosomes to the late Golgi.	
17p11.2	152	FLCN	Folliculin	Protein-coding			May be involved in energy and/or nutrient sensing through the AMPK and mTOR signaling pathways.	
17q25.1	59	RAD51C	RAD51 homolog C (S. cerevisiae)	Protein-coding			Essential for the homologous recombination (HR) pathway of DNA repair.	
		SLC9A3R1	Solute carrier family 9 (sodium/hydrogen exchanger), member 3 regulator 1	Protein-coding			Scaffold protein that connects plasma membrane proteins with members of the ezrin/moesin/radixin family and thereby helps to link them to the actin cytoskeleton and to regulate their surface expression.	

白编码基因,10 个非编码 RNA 基因和 35 个假基因。这一区域包含 5 个已报道过的人类 TSGs,分别是 CDKN2A、CDKN2B、MTAP、PTENP1 和 TOPORS。在 ESCC 中的研究发现,CDKN2A 蛋白在肿瘤组织中表达下调,肿瘤组织中可检测到 *CDKN2A* 和 *CDKN2B* 基因启动子区的高甲基化,CDKN2B 和 MTAP 基因在染色体上因为紧邻 *CDKN2A* 基因,通常表现为与 *CDKN2A* 基因一起发生纯和缺失 (Table 1),提示至少 CDKN2A 基因是食管癌中该区域中的 TSGs。

2.2.5 13q11~q12.1

共有 12 篇文献在 ESCC 中鉴定到了 13q11~q12.1 全长及部分区域的 LOH, 频率在 40%~100%之间。这一区域全长 9.9Mb, 编码 163 个已知基因, 包括 35 个蛋白编码基因,34 个非编码 RNA 基因和 94 个假基因。这一区域包含 4 个已报道过的人类 TSGs,分别是 IFT88、PDX1、GJB2 和 LATS2。在 ESCC 中的研究发现,PDX1 蛋白在肿瘤组织中表达上调,而基因表达芯片和免疫组织化学染色对 GJB2 在食管癌中的表达情况报道不一致,还有 1 篇文献报道 LATS2 在食管癌中很少发生基因突变(Table 1)。目前对这一区域几个候选 TSGs 的研究还较少,HPA 数据显示,GJB2 和 LATS2 均在食管上皮中高表达, 而在常见肿瘤中低表达 (Figure 2)。

2.2.6 13q12.3~q14.3

共有 21 篇文献在 ESCC 中鉴定到了 13q12.3~q14.3 全长及部分区域的 LOH, 频率在 46%~100% 之间。这一区域全长

26.4Mb,编码 101 个已知基因,包括 35 个蛋白编码基因,95 个非编码 RNA 基因和 118 个假基因。这一区域包含 15 个已报道过的人类 TSGs，分别是 BRCA2、STARD13、PDS5B、DLEU2、TRIM13、TSC22D1、FOXO1、KCNRG、ARL11、THSD1、MIR15A、MIR16-1、RB1、DLEU1 和 INTS6。在 ESCC 中的研究发现，BRCA2 蛋白在肿瘤组织中表达上调，而 *PDX5B*、*THSD1* 和 *RB1* 基因在肿瘤组织中表达下调,肿瘤组织中存在 *THSD1* 和 *RB1* 基因启动子区的高甲基化，而突变分析显示,PDX5B 和 INTS6 在食管癌中的突变比较罕见,BRCA2 的突变率在不同的研究中报道不一致(Table 1)。由于这一区域距离较长,可能存在多个 TSGs 均在食管癌癌变中起作用,BRCA2、PDS5B、THSD1 和 RB1 可能都是这一区域中 ESCC TSGs 的有力候选(Figure 2)。

2.2.7　13q32~q34

共有 11 篇文献在 ESCC 中鉴定到了 13q32~q34 全长及部分区域的 LOH，频率在 41%~100%之间。这一区域全长 20.17Mb,编码 189 个已知基因,包括 61 个蛋白编码基因,72 个非编码 RNA 基因和 56 个假基因。这一区域包含 2 个已报道过的人类 TSGs,分别是 CUL4A 和 ING1。在 ESCC 中的研究发现,ING1 蛋白在肿瘤组织中表达下调，而且突变分析显示,其在食管癌中的突变比较常见(Table 1),提示至少 *ING1* 基因是食管癌该区域中的 TSGs 候选。

2.2.8　17p13

共有 17 篇文献在 ESCC 中鉴定到了 17p13 全长及部分区域的 LOH,频率在 36%~95%之间。这一区域全长 10.7Mb，编码 280 个已知基因，包括 214 个蛋白编码基因,31 个非编码 RNA 基因和 35 个假基因。这一区域包含 13 个已报道过的人类 TSGs,分别是 GABARAP、MIR195、TP53、KCTD11、PFN1、XAF1、DPH1、HIC1、MIR22、MYBBP1A、PAFAH1B1、SMYD4 和 VPS53。在 ESCC 中的研究发现,XAF1 蛋白在肿瘤组织中表达下调,TP53 和 PFN1 蛋白在肿瘤中表达上调，突变分析显示,TP53 在食管癌中的突变率非常高,而启动子甲基化罕见,*XAF1* 基因启动子在肿瘤组织中呈高甲基化状态 (Table 1)。位于 17p13.1 区域中的 TP53 已被认为是食管癌中最重要的 TSGs，但也不排除位于 17p13.2 区域的 XAF1 和 KCTD11 基因是食管癌相关 TSGs 的可能性。

2.2.9　17p11.2

共有 7 篇文献在 ESCC 中鉴定到了 17p11.2 区域的 LOH,频率在 72%~100%之间。这一区域全长 6.2Mb,编码 152 个已知基因,包括 76 个蛋白编码基因,18 个非编码 RNA 基因和 58 个假基因。这一区域包含 1 个已报道过的人类 TSGs,即 FLCN。目前还没有 FLCN 在 ESCC 中研究的相关报道,HPA 数据显示,FLCN 在食管上皮中呈中—高强度表达,而在部分常见肿瘤,如乳腺癌、结肠癌、宫颈癌、皮肤癌、脑胶质瘤等中明显低表达(Figure 2)。

2.2.10　17q25.1

共有 8 篇文献在 ESCC 中鉴定到了 17q25.1 区域的 LOH，频率在 25%~69%之间。这一区域全长 3.9Mb,编码 59 个已知基因,包括 41 个蛋白编码基因,11 个非编码 RNA 基因和 7 个假基因。这一区域包含 2 个已报道过的人类 TSGs,分别是 RAD51C 和 SLC9A3R1。目前还没有 RAD51C 和 SLC9A3R1 在 ESCC 中研究的相关报道,HPA 数据未涵盖 RAD51C,而 SLC9A3R1 蛋白在食管上皮中强表达,而在常见肿瘤中低表达，提示 *SLC9A3R1* 基因有可能是食管癌该区域中的 TSGs 候选(Figure 2)。

3　讨　论

肿瘤抑制基因通常是指一类维持基因组稳定性的癌症基因，目前认为 TSGs 在各种肿瘤的发生发展中发挥重要作用。在正常细胞中,TSGs 作为“细胞卫士”，在建立细胞周期检测点、DNA 损伤修复、蛋白泛素化降解、有丝分裂信号、细胞转化、分化、迁移、诱导凋亡和代谢调节中发挥关键性的作用[7]。TSGs 遗传失活或功能缺失通常被认为是使肿瘤细胞获得生长优势的驱动力量，其具体机制主要有以下几种：①点突变，如 *TP53* 基因的体细胞突变在 ESCC 患者中的突变率可达 48%~90%[8]；②等位基因缺失，即基因组中拷贝数的异常缺失，又包括 LOH 和纯合性缺失,后者指与同一个体的正常组织相比,肿瘤组织中两个亲代等位基因全部丢失。肿瘤细胞中的 LOH 已被证明是候选抑癌基因的有效指示剂;③在表观基因组水平,启动子区的高甲基化是许多 TSGs 失活的重要机制。在这些机制的作用下，

TSGs 在肿瘤组织中通常表达显著降低，但也有部分突变型 TSGs 的表达量反而显著增加，如突变型的 TP53 由于空间构象的改变，失去了对其负调控因子 MDM2 的转录激活作用，从而导致了突变蛋白在肿瘤细胞核内的累积[9]。此外，最近几年研究发现，非编码 RNA，如 miRNAs、长链非编码 RNA（lncRNA）等，也可以在肿瘤发生过程中作为 TSGs，在转录后水平上调节细胞增殖、凋亡过程。

本研究通过收集整理文献报道的 ESCC 中的 LOH 区域和频率等，利用生物信息学工具绘制了这些 LOH 区域在人类染色体上的分布图。与以往只关注 LOH 区域的频率有所不同，我们特别关注了这些区域的频率和重复鉴定次数，将频率高、重复鉴定次数多的区域列为 ESCC 中高度可信的 LOH 热点区域，这样共得到 10 个染色体区带，主要累及第 3、5、9、13、17 号染色体。以往还认为染色体 1p36、8p22、16q23.3~q24.1、18q21.1 和 18q23.3，在 ESCC 中也存在高频的 LOH，可能累及某些候选 TSGs[5]。但是，这些染色体上的 LOH 研究报道重复性较低，阳性结果可能与使用样本的样本量较少和选择偏倚有关。比如，仅有 2 篇文献报道了 1p36 区域的 LOH 现象，其中日本人群研究使用样本 48 例，报道频率为 8%[10]；而中国人群研究使用样本仅 14 例，报道频率 64%[11]。再比如，有 3 篇文献报道了中国人群 8p22 区域的 LOH，2 篇样本例数≤15 例，检测频率在 50%以上，而还有 1 篇的样本例数为 18 例，频率只有 33%。由此可见，除了 LOH 频率以外，该区域的可重复鉴定性也应该是参与食管癌变 TSGs 的重要判断标准。

在得到高度可信的 LOH 热点区域以后，我们利用 3 个公共数据库，人类基因组数据库 NCBI、肿瘤抑制基因数据库和人类蛋白质图集数据库，查找了这些区域中全部的编码基因、已知的 TSGs 以及在食管上皮和常见肿瘤中存在异常表达的蛋白质，结合文献调研，分析了每个区域中可能参与食管癌变的候选 TSGs。其中，除了几个已知的 ESCC TSGs，如 FHIT、CDKN2A、TP53、BRCA2、RB1、ING1 等以外，在每一区域中均鉴定到至少一个候选的 TSGs。HPA 数据显示，这些候选 TSGs 在食管上皮，尤其是增殖活跃的基底层细胞中明显高表达，而在多种常见恶性肿瘤，特别是鳞状细胞起源的宫颈癌、头颈鳞癌等中表达下调（Figure 2）。值得注意的是，在 9p21、13q12.3~q14.3、17p13 等区域中，包括多个已知在其他肿瘤中具有 TSGs 活性的长非编码 RNA（long non-coding RNA，lncRNA）和 miRNA。目前在食管癌中未见这些非编码 RNA 存在异常表达的相关报道。因此它们是否在食管癌变中发挥 TSGs 的作用，还需进一步的实验证实。

综上所述，随着近年来生物信息学可视化工具和公共数据库的不断发展，非常有必要重新审视 ESCC 中的潜在抑癌基因。本研究基于文献获取了 ESCC 中高度可信 LOH 热点区域，研究结果对后续深入的表达和功能分析提供了充分的理论依据。

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] Hongo M，Nagasaki Y，Shoji T. Epidemiology of esophageal cancer: orient to Occident. Effects of chronology，geography and ethnicity[J]. J Gastroenterol Hepatol，2009,24(5):729–735.

[3] Knudson AG. Two genetic hits (more or less) to cancer[J]. Nat Rev Cancer，2001,1(2):157–162.

[4] Beroukhim R，Lin M，Park Y，et al. Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays [J]. PLoS Comput Biol，2006,2(5):e41.

[5] Li J，Liu ZH. Loss of heterozygosity in esophageal carcinoma[J]. World Chin J Digestol，2003，11(11):1777–1781.[李洁，刘芝华. 食管癌中的等位基因缺失[J]. 世界华人消化杂志，2003,11(11):1777–1781.]

[6] Zhao M，Sun J，Zhao Z. TSGene: a web resource for tumor suppressor genes[J]. Nucleic Acids Res，2013,41(Database issue):D970–976.

[7] Sherr CJ. Principles of tumor suppression[J]. Cell，2004,116(2):235–246.

[8] Bellini MF，Cadamuro AC，Succi M，et al. Alterations of the TP53 gene in gastric and esophageal carcinogenesis[J]. J Biomed Biotechnol，2012,2012:891961.

[9] Kato S，Han SY，Liu W，et al. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis[J]. Proc Natl Acad Sci USA，2003,100(14):8424–8429.

[10] Nimura Y，Mihara M，Ichimiya S，et al. p73，a gene related to p53，is not mutated in esophageal carcinomas[J]. Int J Cancer，1998,78(4):437–440.

[11] Cai YC，Yang GY，Nie Y，et al. Molecular alterations of p73 in human esophageal squamous cell carcinomas: loss of heterozygosity occurs frequently; loss of imprinting and elevation of p73 expression may be related to defective

p53[J]. Carcinogenesis, 2000,21(4):683–689.
[12] Guo XQ, Wang SJ, Zhang LW, et al. DNA methylation and loss of protein expression in esophageal squamous cell carcinogenesis of high-risk area[J]. J Exp Clin Cancer Res, 2007,26(4):587–594.
[13] Lee EJ, Lee BB, Kim JW, et al. Aberrant methylation of Fragile Histidine Triad gene is associated with poor prognosis in early stage esophageal squamous cell carcinoma [J]. Eur J Cancer, 2006,42(7):972–980.
[14] Kuroki T, Trapasso F, Yendamuri S, et al. Allele loss and promoter hypermethylation of VHL, RAR-beta, RASSF1A, and FHIT tumor suppressor genes on chromosome 3p in esophageal squamous cell carcinoma[J]. Cancer Res 2003, 63(13):3724–3728.
[15] Noguchi T, Takeno S, Kimura Y, et al. FHIT expression and hypermethylation in esophageal squamous cell carcinoma[J]. Int J Mol Med, 2003,11(4):441–447.
[16] Shimada Y, Sato F, Watanabe G, et al. Loss of fragile histidine triad gene expression is associated with progression of esophageal squamous cell carcinoma, but not with the patient´s prognosis and smoking history[J]. Cancer, 2000, 89(1):5–11.
[17] Kitamura A, Yashima K, Okamoto E, et al. Reduced Fhit expression occurs in the early stage of esophageal tumorigenesis: no correlation with p53 expression and apoptosis [J]. Oncology, 2001,61(3):205–211.
[18] Lo PH, Leung AC, Kwok CY, et al. Identification of a tumor suppressive critical region mapping to 3p14.2 in esophageal squamous cell carcinoma and studies of a candidate tumor suppressor gene, ADAMTS9[J]. Oncogene, 2007,26(1):148–157.
[19] Wang CC, Mao WM, Ling ZQ. Expression of DNA methylation of APC in peripheral blood and tumor tissue in patients with esophageal squamous cell carcinoma[J]. Zhonghua Wei Chang Wai Ke Za Zhi, 2011,14(9):719–722.
[20] Ishii T, Murakami J, Notohara K, et al. Oesophageal squamous cell carcinoma may develop within a background of accumulating DNA methylation in normal and dysplastic mucosa[J]. Gut, 2007,56(1):13–19.
[21] Peng H, Zhong XY, Liu KP, et al. Expression and significance of adenomatous polyposis coli, beta-catenin, E-cadherin and cyclin D1 in esophageal squamous cell carcinoma assessed by tissue microarray[J]. Ai Zheng, 2009,28(1): 38–41.
[22] Hu N, Wang C, Su H, et al. High frequency of CDKN2A alterations in esophageal squamous cell carcinoma from a high-risk Chinese population[J]. Genes Chromosomes Cancer, 2004,39(3):205–216.
[23] Qureshi MA, Jan N, Dar NA, et al. A novel p16(INK4A) mutation associated with esophageal squamous cell carcinoma in a high risk population[J]. Biomarkers, 2012,17(6): 552–556.
[24] Chen J, Huang ZJ, Duan YQ, et al. Aberrant DNA methylation of P16, MGMT, and hMLH1 genes in combination with MTHFR C677T genetic polymorphism and folate intake in esophageal squamous cell carcinoma[J]. Asian Pac J Cancer Prev, 2012,13(10):5303–5306.
[25] Taghavi N, Biramijamal F, Sotoudeh M, et al. p16INK4a hypermethylation and p53, p16 and MDM2 protein expression in esophageal squamous cell carcinoma[J]. BMC Cancer, 2010,10:138.
[26] Salam I, Hussain S, Mir MM, et al. Aberrant promoter methylation and reduced expression of p16 gene in esophageal squamous cell carcinoma from Kashmir valley: a high-risk area[J]. Mol Cell Biochem, 2009,332(1–2):51–58.
[27] Fujiwara S, Noguchi T, Takeno S, et al. Hypermethylation of p16 gene promoter correlates with loss of p16 expression that results in poorer prognosis in esophageal squamous cell carcinomas[J]. Dis Esophagus, 2008,21(2):125–131.
[28] Fukuoka T, Hibi K, Nakao A. Aberrant methylation is frequently observed in advanced esophageal squamous cell carcinoma[J]. Anticancer Res, 2006,26(5A):3333–3335.
[29] Tokugawa T, Sugihara H, Tani T, et al. Modes of silencing of p16 in development of esophageal squamous cell carcinoma[J]. Cancer Res, 2002,62(17):4938–4944.
[30] Cheng TH, Hsu PK, Li AF, et al. Correlation of p53, MDM2 and p14(ARF) protein expression in human esophageal squamous cell carcinoma [J]. J Cancer Res Clin Oncol, 2009,135(11):1577–1582.
[31] Xing EP, Nie Y, Song Y, et al. Mechanisms of inactivation of p14ARF, p15INK4b, and p16INK4a genes in human esophageal squamous cell carcinoma[J]. Clin Cancer Res, 1999,5(10):2704–2713.
[32] Tanaka H, Shimada Y, Imamura M, et al. Multiple types of aberrations in the p16 (INK4a) and the p15(INK4b) genes in 30 esophageal squamous-cell-carcinoma cell lines [J]. Int J Cancer, 1997,70(4):437–442.
[33] Xing EP, Nie Y, Wang LD, et al. Aberrant methylation of p16INK4a and deletion of p15INK4b are frequent events in human esophageal cancer in Linxian, China[J]. Carcinogenesis, 1999,20(1):77–84.
[34] Ling Y, Huang G, Fan L, et al. CpG island methylator phenotype of cell-cycle regulators associated with TNM stage and poor prognosis in patients with oesophageal squamous cell carcinoma[J]. J Clin Pathol, 2011,64(3): 246–251.
[35] Kim DH, Muto M, Kuwahara Y, et al. Array-based comparative genomic hybridization of circulating esophageal tumor cells[J]. Oncol Rep, 2006,16(5):1053–1059.
[36] Takahashi O, Hamada J, Abe M, et al. Dysregulated expression of HOX and ParaHOX genes in human esophageal squamous cell carcinoma[J]. Oncol Rep, 2007,17(4):753–760.
[37] Inose T, Kato H, Kimura H, et al. Correlation between connexin 26 expression and poor prognosis of esophageal

squamous cell carcinoma[J]. Ann Surg Oncol, 2009,16(6): 1704–1710.
[38] Uchikado Y, Inoue H, Haraguchi N, et al. Gene expression profiling of lymph node metastasis by oligomicroarray analysis using laser microdissection in esophageal squamous cell carcinoma[J]. Int J Oncol, 2006,29(6):1337–1347.
[39] Ishizaki K, Fujimoto J, Kumimoto H, et al. Frequent polymorphic changes but rare tumor specific mutations of the LATS2 gene on 13q11-12 in esophageal squamous cell carcinoma[J]. Int J Oncol, 2002,21(5):1053–1057.
[40] Akbari MR, Malekzadeh R, Lepage P, et al. Mutations in Fanconi anemia genes and the risk of esophageal cancer [J]. Hum Genet, 2011,129(5):573–582.
[41] Akbari MR, Malekzadeh R, Nasrollahzadeh D, et al. Germline BRCA2 mutations and the risk of esophageal squamous cell carcinoma[J]. Oncogene, 2008,27(9):1290–1296.
[42] Hu N, Li G, Li WJ, et al. Infrequent mutation in the BRCA2 gene in esophageal squamous cell carcinoma[J]. Clin Cancer Res, 2002,8(4):1121–1126.
[43] Hu N, Clifford RJ, Yang HH, et al. Genome wide analysis of DNA copy number neutral loss of heterozygosity (CNNLOH) and its relation to gene expression in esophageal squamous cell carcinoma [J]. BMC Genomics, 2010,11:576.
[44] Harada H, Uchida N, Shimada Y, et al. Polymorphism and allelic loss at the AS3 locus on 13q12-13 in esophageal squamous cell carcinoma[J]. Int J Oncol, 2001,18(5):1003–1007.
[45] Zhang Y, Huang X, Qi J, et al. Correlation of genomic and expression alterations of AS3 with esophageal squamous cell carcinoma[J]. J Genet Genomics, 2008,35(5):267–271.
[46] Ko JM, Chan PL, Yau WL, et al. Monochromosome transfer and microarray analysis identify a critical tumor-suppressive region mapping to chromosome 13q14 and THSD1 in esophageal carcinoma [J]. Mol Cancer Res, 2008, 6(4):592–603.
[47] Mathew R, Arora S, Khanna R, et al. Alterations in p53 and pRb pathways and their prognostic significance in oesophageal cancer[J]. Eur J Cancer, 2002,38(6):832–841.
[48] Contu SS, Contu PC, Damin DC, et al. pRB expression in esophageal mucosa of individuals at high risk for squamous cell carcinoma of the esophagus[J]. World J Gastroenterol, 2007,13(11):1728–1731.
[49] Li WJ, Hu N, Su H, et al. Allelic loss on chromosome 13q14 and mutation in deleted in cancer 1 gene in esophageal squamous cell carcinoma[J]. Oncogene, 2003, 22(2):314–318.
[50] Chen L, Matsubara N, Yoshino T, et al. Genetic alterations of candidate tumor suppressor ING1 in human esophageal squamous cell cancer[J]. Cancer Res, 2001,61(11):4345–4349.
[51] Zhu ZL, Yan BY, Zhang Y, et al. Cytoplasmic expression of p33(ING1b) is correlated with tumorigenesis and progression of human esophageal squamous cell carcinoma[J]. Oncol Lett, 2013,5(1):161–166.
[52] Abedi-Ardekani B, Kamangar F, Sotoudeh M, et al. Extremely high Tp53 mutation load in esophageal squamous cell carcinoma in Golestan Province, Iran[J]. PLoS One, 2011,6(12):e29488.
[53] Rossini A, de Almeida Simao T, Marques CB, et al. TP53 mutation profile of esophageal squamous cell carcinomas of patients from Southeastern Brazil[J]. Mutat Res, 2010, 696(1):10–15.
[54] Putz A, Hartmann AA, Fontes PR, et al. TP53 mutation pattern of esophageal squamous cell carcinomas in a high risk area (Southern Brazil): role of life style factors[J]. Int J Cancer, 2002,98(1):99–105.
[55] Hu N, Huang J, Emmert-Buck MR, et al. Frequent inactivation of the TP53 gene in esophageal squamous cell carcinoma from a high-risk population in China[J]. Clin Cancer Res, 2001,7(4):883–891.
[56] Egashira A, Morita M, Yoshida R, et al. Loss of p53 in esophageal squamous cell carcinoma and the correlation with survival: analyses of gene mutations, protein expression, and loss of heterozygosity in Japanese patients[J]. J Surg Oncol, 2011,104(2):169–175.
[57] Gu ZD, Shen LY, Wang H, et al. HOXA13 promotes cancer cell growth and predicts poor survival of patients with esophageal squamous cell carcinoma[J]. Cancer Res, 2009, 69(12):4969–4973.
[58] Matsumoto M, Furihata M, Kurabayashi A, et al. Prognostic significance of serine 392 phosphorylation in overexpressed p53 protein in human esophageal squamous cell carcinoma[J]. Oncology, 2004,67(2):143–150.
[59] Bergqvist AS, Bergqvist M, Brattstrom D, et al. Serum p53 autoantibodies as prognostic marker in patients with oesophageal carcinoma[J]. Anticancer Res, 2001,21(6A): 4141–4145.
[60] Taniere P, Martel-Planche G, Saurin JC, et al. TP53 mutations, amplification of P63 and expression of cell cycle proteins in squamous cell carcinoma of the oesophagus from a low incidence area in Western Europe[J]. Br J Cancer, 2001,85(5):721–726.
[61] Zhang J, Wang K, Zhang J, et al. Using proteomic approach to identify tumor-associated proteins as biomarkers in human esophageal squamous cell carcinoma[J]. J Proteome Res, 2011,10(6):2863–2872.
[62] Zhu X, Ding M, Yu ML, et al. Identification of galectin-7 as a potential biomarker for esophageal squamous cell carcinoma by proteomic analysis[J]. BMC Cancer, 2010,10: 290.
[63] Chen XY, He QY, Guo MZ. XAF1 is frequently methylated in human esophageal cancer[J]. World J Gastroenterol, 2012,18(22):2844–2849.