



doi:10.7659/j.issn.1005-6947.2018.09.007
http://dx.doi.org/10.7659/j.issn.1005-6947.2018.09.007
Chinese Journal of General Surgery, 2018, 27(9):1126-1134.

· 基础研究 ·

基于生物信息学的胰腺导管腺癌预后风险长链非编码 RNA 筛选

张志鹏, 孙维佳, 陈泓西, 夏华, 陆晔斌

(中南大学湘雅医院 胆胰外科, 湖南长沙 410008)

摘要

目的: 应用生物信息学方法筛选胰腺导管腺癌的预后风险长链非编码 RNA (lncRNA)。

方法: 从癌症和肿瘤基因图谱 (TCGA) 数据库下载胰腺导管腺癌患者的 RNA-seq Level 2 数据及其临床信息。根据 NCBI Gene 数据库及 GENCODE v7 数据库的基因注释信息, 将下载的 mRNA 和 lncRNA 测序数据进行重新注释。然后应用 R 软件的 edgeR 包和 limma 包筛选差异表达的 mRNA 和 lncRNA, 并对它们进行相关性分析, 进而获得 lncRNA-mRNA 有统计学意义的共表达关系对, 并将其中的 mRNA 定义为 lncRNA 的靶基因。通过 R 软件 clusterProfiler 包对 lncRNA 的靶基因进行功能富集分析, 以推测 lncRNA 的生物学功能。最后, 绘制差异表达 lncRNA 的 Kaplan-Meier 曲线, 筛选出与胰腺导管腺癌预后风险相关的 lncRNA。

结果: 经过基因重新注释, 共得到 19 791 个 mRNA 和 1 623 个 lncRNA 的测序数据, 随后共筛选得到 260 个差异表达的 mRNA 和 15 个差异表达的 lncRNA。经过相关性分析, 共得到包括 5 个 lncRNA 和 24 个 mRNA 在内的 24 个有统计学意义的共表达关系对。其中 lncRNA LINC00857 有 6 个靶基因, 分别为 C1orf116、ESRP1、GPC5A、LIPH、MAL2 和 PLS1。根据 lncRNA 靶基因的功能富集分析, 推测 LINC00857 主要富集于磷脂酶活性、细胞骨架结构组成和脂肪酶活性。生存分析发现, CASC8 和 LINC00857 与胰腺导管腺癌的预后风险明显有关 ($P=0.0052$ 、 $P=0.027$)。

结论: CASC8 和 LINC00857 可能是胰腺导管腺癌的预后风险 lncRNA, 有望在今后的研究中成为胰腺导管腺癌新的预后监测指标。

关键词

胰腺肿瘤; 癌, 导管; RNA, 长链非编码; 生物标记, 肿瘤; 计算生物学
中图分类号: R735.9

Identification of prognostic risk long noncoding RNAs for pancreatic ductal adenocarcinoma by bioinformatics analysis

ZHANG Zhipeng, SUN Weijia, CHEN Hongxi, XIA Hua, LU Yebin

(Department of Pancreatobiliary Surgery, Xiangya Hospital, Central South University, Changsha 410008, China)

Abstract

Objective: To screen the prognostic risk long noncoding RNAs (lncRNAs) for pancreatic ductal adenocarcinoma (PDAC) by bioinformatics approaches.

Methods: The RNA-Seq Level 2 data and clinical information of PDAC patients were downloaded from The Cancer Genome Atlas (TCGA) database. The sequencing data of the downloaded mRNAs and lncRNAs were re-

基金项目: 湖南省自然科学基金资助项目 (2017JJ3508)。

收稿日期: 2018-07-03; 修订日期: 2018-08-17。

作者简介: 张志鹏, 中南大学湘雅医院住院医师, 主要从事胰腺、胆道疾病临床与基础方面的研究。

通信作者: 陆晔斌, Email: luyebin6@sina.com

annotated according to the gene annotation data from NCBI Gene database and GENCODE v7 database. The differentially expressed mRNAs and lncRNAs were screened by using edgeR and limma packages in R. Then, the significantly co-expressed pairs between mRNAs and lncRNAs were obtained by correlation analysis, in which, the mRNAs were considered as target genes of the lncRNAs. After that, the functional modules of the lncRNAs were predicted by functional enrichment analysis of their target mRNAs with the clusterProfiler package in R. Finally, the significant prognostic risk lncRNAs for PDAC were determined by drawing Kaplan-Meier curves of the differentially expressed lncRNAs.

Results: After gene re-annotation, the sequencing data of a total of 19 791 mRNAs and 1 623 lncRNAs were obtained, and then, 260 differentially expressed mRNAs and 15 differentially expressed lncRNAs were picked up. From the correlation analysis, 24 significantly co-expressed pairs comprised of 24 mRNAs and 5 lncRNAs were identified. Of them, LINC00857 had 6 target genes that were C1orf116, ESRP1, GPRC5A, LIPH, MAL2 and PLS1, respectively. According to the functional enrichment analysis, the target genes of lncRNA LINC00857 were mainly enriched in phospholipase activity, structural constituent of cytoskeleton, and lipase activity. The results of survival analysis revealed that lncRNA CASC8 and LINC00857 were significantly associated with prognostic risk of PDAC ($P=0.0052$, $P=0.027$).

Conclusion: CASC8 and LINC00857 are potential prognostic risk lncRNAs for PDAC, and may probably become the novel indicators for prognosis of PDAC in the future.

Key words

Pancreatic Neoplasms; Carcinoma, Ductal; RNA, Long Noncoding; Biomarkers, Tumor; Computational Biology
CLC number: R735.9

胰腺导管腺癌是胰腺癌中最常见的病理类型, 约占全部胰腺癌的90%以上^[1], 因此通常所说的胰腺癌主要指的是胰腺导管腺癌。胰腺导管腺癌的总生存率仅为6%, 在所有实体肿瘤中是预后最差的, 而且这一数字在近25年间没有大的变化^[2]。据最新的统计资料显示, 预计2018年全美胰腺癌新发患者为55 440例, 死亡44 330例, 病死率在男女性别中均排第4位^[3]。手术是治愈本病的唯一手段。不幸的是, 绝大多数胰腺癌患者处于局部进展期或已发生转移, 这些患者的中位生存期仅为6~10个月、3~6个月^[4]。只有15%~20%的患者可能接受手术治疗, 然而即使完成了根治性切除, 其预后依然较差。对于接受了胰十二指肠切除术的胰腺癌患者, 切缘阴性的5年生存率为25%~30%, 切缘阳性的5年生存率仅为10%^[5]。虽然近年来在胰腺癌的研究及患者护理方面(如更多的联合化疗方案、术前治疗的新策略、更全面的围手术期护理及手术安全系数的提高)取得了长足的进步, 但胰腺癌患者的总体预后并没有发生显著性的改善。因此, 探索治疗及改善胰腺癌预后的新策略显得十分迫切。

长链非编码RNA(long noncoding RNA,

lncRNA)是一类转录本长度超过200个核苷酸的RNA分子, 在哺乳动物基因组中被普遍转录。lncRNA是目前分子生物学领域的一个全新课题, 它以RNA的形式在表观遗传调控、转录调控以及转录后调控3个层面发挥调控基因表达的作用。lncRNA起初被认为是基因组转录的“噪音”, 是RNA聚合酶II转录的副产物, 不具有任何生物学功能^[6]。但近年来的研究^[7]显示, lncRNA在细胞正常生理及病理过程中都发挥着重要作用。截止目前, 已发现了数万个lncRNA, 与恶性肿瘤的发生、发展密切相关, 并可作为很好的生物标志物^[8]。如AFAP1-AS1在胰腺导管腺癌组织中表达上调, 其高表达与淋巴结转移、周围神经侵犯和患者的预后不良明显相关, 可作为胰腺导管腺癌潜在的预后风险标志物及治疗靶点; HOTTIP在胰腺导管腺癌组织中高表达, 它通过作用于HOXA13促进肿瘤的增殖、侵袭及化疗耐药; 而ENST00000480739在胰腺导管腺癌组织中表达下调, 通过靶向调控低氧诱导因子1 α (hypoxia-inducible factor 1 α , HIF-1 α)影响肿瘤的进展及转移, 可作为抑制胰腺癌转移的治疗靶点^[9-11]。综上所述可以看出lncRNA与胰腺癌的发生发展、侵

袭转移密切相关。lncRNA有望成为胰腺癌早期诊断、预后判断的监测指标,以lncRNA为靶点的基因治疗将为胰腺癌的治疗开辟新天地。本研究从癌症和肿瘤基因图谱(The Cancer Genome Atlas, TCGA)数据库下载胰腺导管腺癌患者的mRNA、lncRNA测序数据及临床信息,通过R软件的edgeR包和limma包进一步筛选出差异表达的lncRNA,并对其进行功能预测。最后,绘制生存曲线,筛选出与胰腺导管腺癌预后风险明显相关的lncRNA,为后续的深入研究指明方向。

1 资料与方法

1.1 数据来源及注释

从TCGA数据库下载胰腺癌患者的RNA-seq Level 2数据及其临床信息。总计183例样本,其中胰腺导管腺癌组织146例,其他类型癌组织33例,癌旁正常胰腺组织4例;胰腺导管腺癌又可再分为IA期3例、IB期9例、IIA期23例、IIB期105例、III期3例、IV期3例。在这146例样本中,有61例在研究结束时仍存活,因此将他们的最后一次随访时间定义为生存时间。

TCGA数据库提供了RNA-seq Level 2 mRNA的原始计数(raw count)和Ensembl ID,结合NCBI Gene数据库的基因注释信息,重新注释得到mRNA的Gene symbol。GENCODE v7数据库收录并分析了目前最完整的人类lncRNA的注释,其收录的lncRNA数据集较其他数据库更全面^[12]。根据TCGA数据库提供的exon的染色体位置、原始计数和每1百万个map上的reads中map到外显子的每千个碱基上的reads数(reads per kilobase of exon per million mapped reads, RPKM)信息,与GENCODE v7数据库中lncRNA的染色体位置进行比对,若染色体位置信息一致,则将该exon定义为lncRNA。

1.2 筛选差异表达的 mRNA 和 lncRNA

首先利用R软件edgeR包中的截尾均值化M值(trimmed mean of M values, TMM)方法归一化预处理mRNA和lncRNA的原始计数;再采用该软件limma包中的voom方法将预处理的数据转化为log₂-count per million(log₂CPM),然后通过limma包分别计算mRNA和lncRNA在癌组织与正常组

织中的表达差异,并采用校正t检验计算P值^[13-15]。对于每个差异表达的mRNA或lncRNA,要求 $P < 0.05$,log₂-fold change的绝对值不小于1。最后,应用VennPlex工具提取4个分期中差异表达mRNA和lncRNA的交集^[16]。

1.3 筛选差异表达 lncRNA 的靶基因

根据所有差异表达lncRNA和mRNA在样本中的表达值,结合Pearson correlation coefficient,计算lncRNA与mRNA之间的共表达系数及Z-score,进而得出P值(根据Z-score计算),并利用Benjamini Hochberg校正P值^[17]。然后,选择lncRNA-mRNA共表达系数绝对值>0.7且 $P < 0.05$ 的关系对作为有统计学意义的共表达关系对,并将其中的mRNA定义为该lncRNA的靶基因。

1.4 差异表达 lncRNA 的功能预测

根据上述得到的lncRNA及其靶基因,借助R软件clusterProfiler包对差异表达lncRNA的靶基因进行功能富集分析,以推测lncRNA的生物学功能^[18]。 $P < 0.05$ 认为具有统计学意义。

1.5 筛选预后风险 lncRNA

根据差异表达lncRNA在胰腺导管腺癌组织中的表达水平及样本的临床信息,整理每个样本的生存时间及存活/死亡状态。采用Graphpad Prism 7统计学软件绘制lncRNA的Kaplan-Meier曲线,并通过Log-rank test确定两组患者之间有无统计学差异, $P < 0.05$ 认为具有统计学意义。

2 结果

2.1 基因注释

本研究共包含146例胰腺导管腺癌组织样本,并有4例癌旁正常胰腺组织作为对照(用于差异表达分析)。经过基因重新注释,最终得到19 791个mRNA及1 623个lncRNA的原始计数。

2.2 筛选差异表达的 mRNA 和 lncRNA

经过数据处理及统计分析,分别筛选出I、II、III、IV期中差异表达的mRNA和lncRNA。4个分期中共同差异表达的mRNA和lncRNA分别为260、15个,其中有161个mRNA和8个lncRNA表达上调,99个mRNA和7个lncRNA表达下调(图1)。15个共同差异表达的lncRNA在癌组织与正常组织中的表达情况见图2。

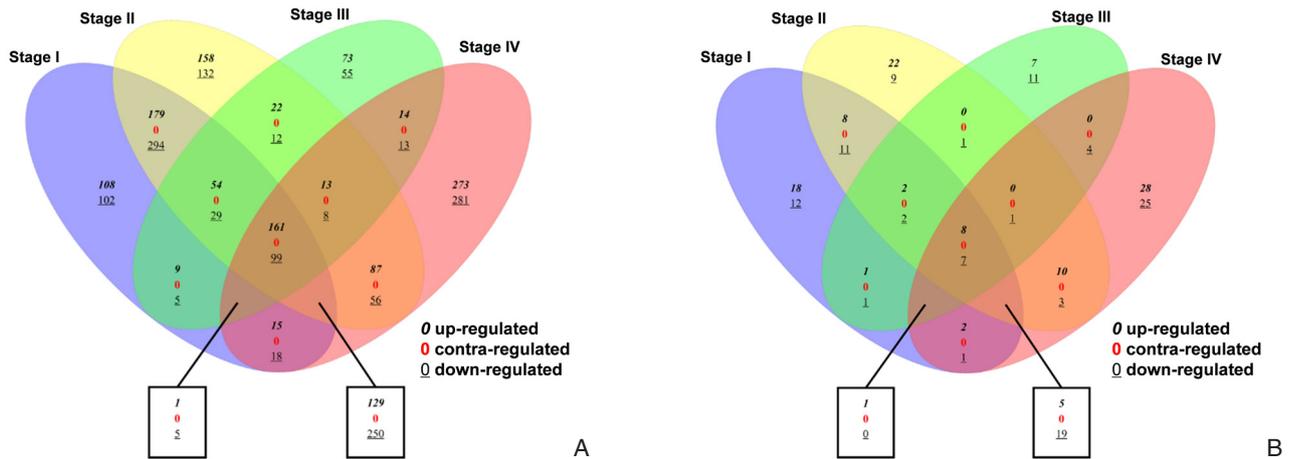


图 1 不同分期胰腺导管腺癌组织中差异表达 mRNA 和 lncRNA 的韦恩图 (蓝色、黄色、绿色和红色椭圆形分别代表 I、II、III、IV 期胰腺导管腺癌组织中差异表达的 mRNA 或 lncRNA; 斜体、红色及带下划线的数字分别为表达上调、相反和下调的 mRNA 或 lncRNA 的数目) A: 差异表达的 mRNA; B: 差异表达的 lncRNA

Figure 1 Venn diagram of differentially expressed mRNAs and lncRNA in different stages of PDAC (The blue, yellow, green, and red ellipses representing the differentially expressed mRNAs or lncRNAs in stage I, II, III and IV respectively; the italic, red and underlined number showing the count of up-regulated, contra-regulated and down-regulated mRNA or lncRNA respectively) A: Differentially expressed mRNAs; B: Differentially expressed lncRNAs

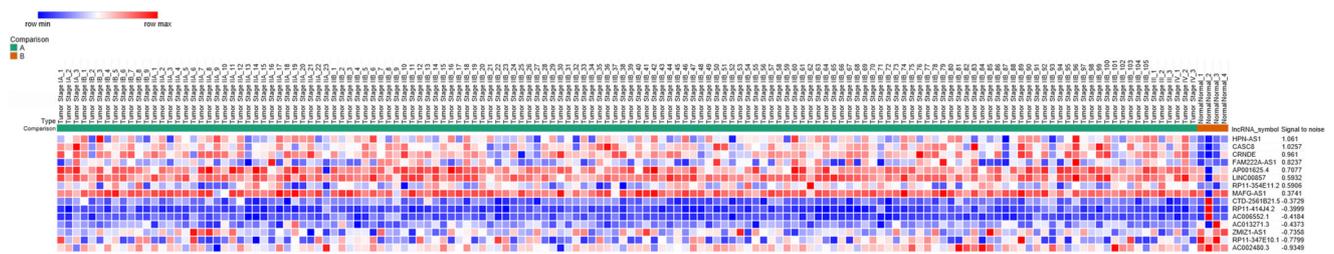


图 2 15 个共同差异表达 lncRNA 在胰腺导管腺癌及癌旁正常组织中的热图 (热图上方为胰腺导管腺癌及癌旁正常组织样本, 热图右侧为 15 个共同差异表达 lncRNA 的基因名; 红色为 lncRNA 在胰腺导管腺癌组织中表达上调, 蓝色为 lncRNA 在胰腺导管腺癌组织中表达下调)

Figure 2 Heat map of 15 common differentially expressed lncRNAs in PDAC and cancer adjacent pancreatic tissue (The samples of cancer and normal tissue listed above the heat map, and the gene symbols of 15 lncRNAs displayed on the right side of the map; the red color presenting the up-regulated lncRNAs in PDAC tissue and the blue presenting the down-regulated lncRNAs in PDAC tissue)

2.3 筛选差异表达 lncRNA 的靶基因

经过共表达分析, 共发现包括 5 个 lncRNA 和 24 个 mRNA 在内的 24 个有统计学意义的共表达关系对, 并将与某个 lncRNA 有统计学意义共表达的 mRNA 作为该 lncRNA 的靶基因 (表 1)。如 LINC00857 与 GPRC5A、ESRP1、C1orf116、LIPH、MAL2、PLS1 存在明显共表达, 则这 6 个 mRNA 被认为是 LINC00857 的靶基因。然后利用 Cytoscape 软件重建 lncRNA-mRNA 共表达网络, 该网络由 29 个节点 (代表 5 个 lncRNA 和 24 个 mRNA) 和 24 条边 (代表 24 个有统计学意义的共表

达关系对) 组成, 其中红色/蓝色菱形节点分别代表上调/下调表达的 lncRNA, 椭圆形节点代表差异表达的 mRNA。根据 lncRNA-mRNA 共表达系数决定 mRNA 节点及边的颜色, 即共表达系数绝对值越大, mRNA 节点及边的绿色越深; 反之, 共表达系数绝对值越小, mRNA 节点及边的黄色越深 (图 3)。

2.4 差异表达 lncRNA 的功能预测

基于上述得到的 lncRNA 及其靶基因, 对其靶基因进行功能富集分析, 以推测 lncRNA 的生物学功能 (表 2)。

表 1 差异表达 lncRNA 与其靶基因

Table 1 The differentially expressed lncRNAs and their target mRNAs

lncRNA	靶 mRNA
CRNDE (12)	PITX1、TMPRSS4、FXVD3、IRX5、C19orf33、MLPH、SFN、NQO1、FLJ23867、S100A6、KCNK1、S100A10
AP001625.4 (1)	RSPH1
LINC00857 (6)	GPRC5A、ESRP1、C1orf116、LIPH、MAL2、PLS1
AC006552.1 (3)	REXO1LIP、RNU4ATAC、OR5AK2
RP11-414J4.2 (2)	LMAN1L、CPLX3

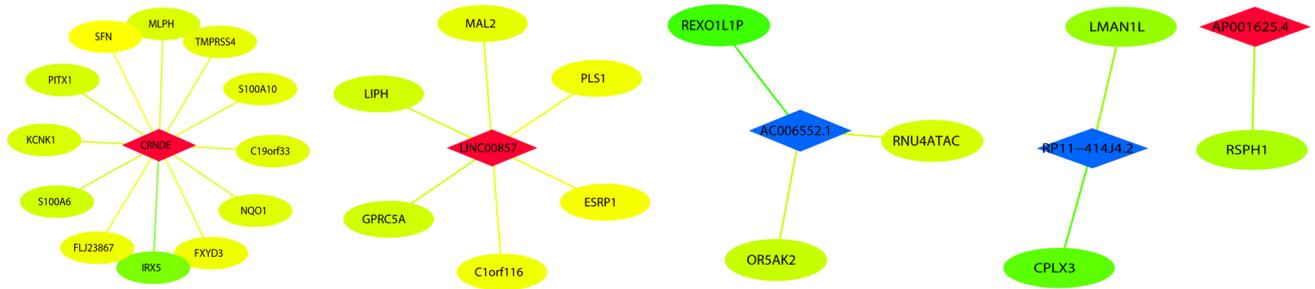


图 3 差异表达 lncRNA 和 mRNA 的共表达网络 (红色、蓝色菱形节点分别代表上调、下调表达的 lncRNA, 椭圆形节点代表差异表达的 mRNA; 根据 lncRNA-mRNA 共表达系数决定 mRNA 节点及边的颜色, 即共表达系数绝对值越大, mRNA 节点及边的绿色越深, 反之, 共表达系数绝对值越小, mRNA 节点及边的黄色越深)

Figure 3 The co-expression network of differentially expressed lncRNAs and mRNAs (The red or blue diamond node representing upregulated or downregulated lncRNAs, and the ellipse node representing differentially expressed mRNAs; based on co-expression coefficient of lncRNA-mRNA, the greater the absolute value, the deeper the green color of the ellipse node and the edge, otherwise, the smaller the absolute value, the deeper the yellow color of the ellipse node and the edge)

表 2 差异表达 lncRNA 的功能富集分析

Table 2 Functional enrichment analysis of differentially expressed lncRNAs

lncRNA	ID/Term	靶 mRNA	校正 P 值
AC006552.1	GO: 0004527/exonuclease activity (核酸外切酶活性)	REXO1LIP	0.008 94
	GO: 0005549/odorant binding (气味结合)	OR5AK2	0.011
	GO: 0004518/nuclease activity (核酸酶活性)	REXO1LIP	0.023 4
CRNDE	GO: 0044548/S100 protein binding S100 (蛋白结合)	S100A6	0.007 74
	GO: 0005523/tropomyosin binding (原肌球蛋白结合)	S100A6	0.008 33
	GO: 0022841/potassium ion leak channel activity (钾离子渗漏通道活性)	KCNK1	0.009 52
	hsa00130/ubiquinone and other terpenoid-quinone biosynthesis (泛醌及其他萜醌生物合成)	NQO1	0.004 6
	hsa04960/aldosterone-regulated sodium reabsorption (醛固酮调节的钠从吸收)	SFN	0.016 3
	hsa04115/p53 signaling pathway (p53 信号通路)	SFN	0.016 3
LINC00857	GO: 0004620/phospholipase activity (磷脂酶活性)	LIPH	0.031 1
	GO: 0005200/structural constituent of cytoskeleton (细胞骨架结构组成)	PLS1	0.035 7
	GO: 0016298/lipase activity (脂肪酶活性)	LIPH	0.036
RP11-414J4.2	GO: 0005326/neurotransmitter transporter activity (神经递质转运蛋白活性)	CPLX3	0.002 87
	GO: 0019905/syntaxin binding (突触融合蛋白结合)	CPLX3	0.010 2
	GO: 0000149/SNARE binding (SNARE 结合)	CPLX3	0.014 3
	hsa04721/synaptic vesicle cycle (突触囊泡循环)	CPLX3	0.017 5
	hsa04141/protein processing in endoplasmic reticulum (内质网蛋白加工)	LMAN1L	0.045 8

2.5 筛选预后风险 lncRNA

通过对这 15 个差异表达的 lncRNA 绘制 Kaplan-Meier 生存曲线, 发现有 2 个 lncRNA 与胰腺导管腺癌的预后风险明显有关, 分别为 CASC8

($P=0.0052$) 和 LINC00857 ($P=0.027$), 它们在癌组织中均为高表达, 生存分析提示低表达预后较好 (图 4)。

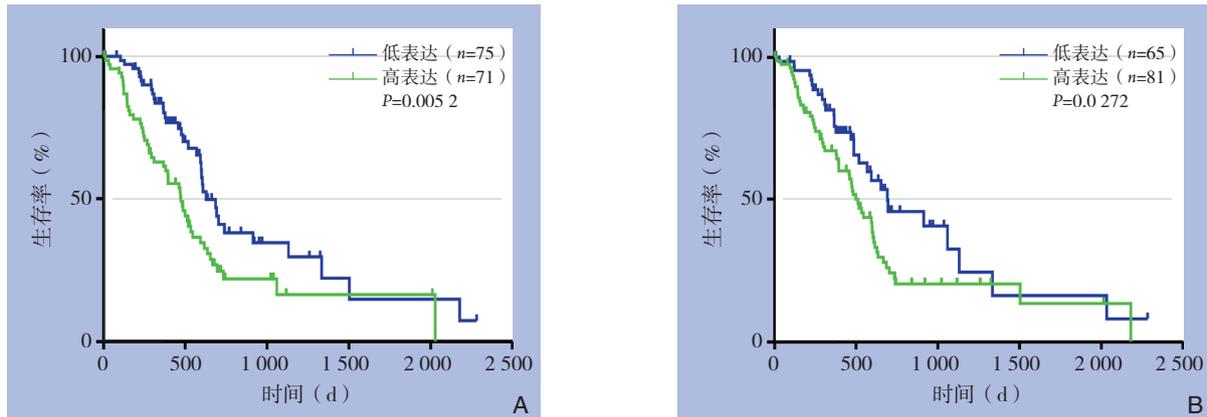


图4 CASC8 和 LINC00857 在 146 例胰腺导管腺癌患者中的 Kaplan-Meier 分析 (基于 CASC8 和 LINC00857 的表达平均值, 将所有患者分为低表达组和高表达组) A: CASC8; B: LINC00857

Figure 4 Kaplan-Meier analysis of CASC8 and LINC00857 in the 146 PDAC patients (Patients divided into low and high expression group according to the mean expression value of lncRNA) A: CASC8; B: LINC00857

此外, 本研究发现 LINC00857 的靶基因为 C1orf116 (chromosome 1 open reading frame 116)、ESRP1 (epithelial splicing regulatory protein 1)、GPRC5A (G protein-coupled receptor class C group 5 member A)、LIPH (lipase H)、MAL2 (mal T-cell differentiation protein 2)、PLS1 (plastin 1) (图5), 未发现 CASC8 的靶基因。LINC00857 靶基因的功能富集分析详见表3。

表3 LINC00857 靶基因的功能富集分析

Table 3 Function enrichment analysis of target genes of LINC00857

lncRNA	ID/Term	校正 P 值
LINC00857	GO:0004620/phospholipase activity (磷脂酶活性)	0.031 1
	GO:0005200/structural constituent of cytoskeleton (细胞骨架结构组成)	0.035 7
	GO:0016298/lipase activity (脂肪酶活性)	0.036
	GO:0051015/actin filament binding (肌动蛋白丝结合)	0.040 9
	GO:0052689/ carboxylic ester hydrolase activity (羧酸酯水解酶活性)	0.042 6

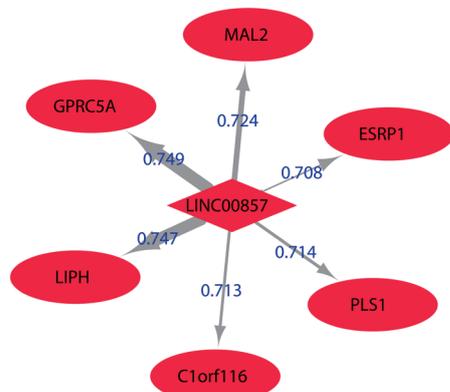


图5 LINC00857 与其靶基因的调控网络 (菱形节点代表 LINC00857, 椭圆形节点代表 LINC00857 的靶基因; 箭头上的数字为 LINC00857 与其靶基因的共表达系数, 共表达系数越大, 箭头越粗; 红色代表在胰腺导管腺癌组织中表达上调)

Figure 5 The regulatory network of LINC00857 and target mRNAs (The diamond node representing lncRNA LINC00857, and the ellipse node representing the target mRNAs of LINC00857; the blue number above the arrow showing the co-expression coefficient between LINC00857 and target mRNAs, and the greater the value, the thicker the arrow; the red color presenting up-regulation in PDAC tissue)

3 讨论

随着基因芯片及二代测序技术的进步, 在肿瘤研究过程中产生了大量的数据, 这些数据可能蕴含着肿瘤发生发展的重要信息。TCGA项目是迄今为止世界范围内最大、最成功的癌症基因组研究项目, 它试图通过基因组分析技术, 特别是采用大规模的基因组测序, 将人类全部癌症的基因组变异图谱绘制出来, 并进行系统分析, 旨在找到所有致癌和抑癌基因的微小变异, 了解癌细胞发生、发展的机制, 并在此基础上探索新的诊断和治疗方法, 最后总结出新型的“癌症预防策略”。目前, TCGA数据库已收录了11 000例患者的基因组序列、表达水平、甲基化和拷贝数变化在内的30多种癌症的数据及临床信息, 并且允许科研人员免费下载^[19]。lncRNA作为分子生物学一

个全新领域,可通过多种途径调控基因的表达,如在细胞核中以顺式或反式作用方式调控基因转录^[20-21],作为蛋白质诱饵进行间接调控^[22];在细胞质中作用于mRNA影响其稳定性及翻译过程^[23],作为竞争性内源RNA参与microRNA调控^[24],与转录因子结合影响其功能发挥^[25]等,为肿瘤的研究提供了新的视角。越来越多的研究表明,lncRNA参与了细胞正常生理及病理过程,包括肿瘤的发生和进展。但是,关于lncRNA在胰腺导管腺癌中的作用研究仍然较少,它们在胰腺导管腺癌中的表达模式及作用机制还有待进一步的探索。因此,本研究从TCGA数据库下载胰腺导管腺癌患者的mRNA、lncRNA测序数据及临床信息。通过R软件的edgeR包和limma包筛选出15个差异表达的lncRNA和260个差异表达的mRNA;经过共表达分析,筛选出24个lncRNA-mRNA显著共表达关系对,将其中的mRNA定义为lncRNA的靶基因,并通过靶基因进行功能富集分析以预测lncRNA的功能。最后,对差异表达lncRNA进行生存分析,筛选出两个与胰腺导管腺癌预后风险明显相关的lncRNA,为CASC8和LINC00857,为接下来的功能研究指明方向。

本研究发现了2个新的潜在的预后风险lncRNA,其中1个是CASC8,它在胰腺导管腺癌组织中高表达,低表达组有较好的预后。CASC8位于8q24.21。众多研究发现,CASC8基因的单核苷酸多态性与许多肿瘤的发生和化疗抵抗相关。Yao等^[26]进行的一项Meta分析显示,CASC8基因上rs7837328和rs6983267位点的多态性是欧洲及亚洲人患结直肠癌的风险因素,而rs7014346和rs10505477只与欧洲人患结直肠癌的风险相关。Ma等^[27]发现携带CASC8 rs10505477 GG基因型的中国胃癌患者有更长的生存期,在肿瘤≤5 cm、弥漫性胃癌、淋巴结转移、III及IV期的患者中尤为明显,从而推测rs10505477多态性可作为潜在的预后风险标志。此外,还有研究发现CASC8 rs10505477多态性不仅与卵巢癌的进展有关,还可被用于肺癌的诊断、确定铂类化疗药物的反应和毒性等^[28-29]。目前对CASC8的研究主要集中在其基因位点的单核苷酸多态性上,但尚缺乏在胰腺导管腺癌中的研究。

此外,另一个潜在的预后风险lncRNA LINC00857也被发现在胰腺导管腺癌组织中高表达,并与其预后不良密切相关,而目前尚无

LINC00857在胰腺导管腺癌中的研究报道。来自MiTranscriptome的6 220个肿瘤RNA-Seq结果显示,LINC00857在膀胱癌、胃癌、肺腺癌、宫颈癌和胰腺癌中高表达($\log_2 \text{FRKM} > 0.5$)^[30],与本研究结果一致。此外,Wang等^[31]应用二代测序技术检测肺腺癌和正常肺组织中lncRNA的表达水平,发现LINC00857在肺腺癌组织中高表达,并与肺腺癌较差的预后明显相关。而且,干扰LINC00857的表达会抑制肺癌细胞的增殖、迁移与侵袭能力;反之,LINC00857的过表达会促进肺癌细胞的增殖和侵袭,并在动物实验中得到了进一步证实。研究^[31]还发现LINC00857是通过调控细胞周期而发挥促癌作用的,与本研究中LINC00857的表达模式相符。不同的是,本研究发现LINC00857的作用机制主要是影响酶的活性、细胞骨架构成和结合肌动蛋白等。Zhang等^[32]首次通过全基因组lncRNA测序分析胃癌患者血浆及癌组织标本中lncRNA的表达,发现LINC00857在术前的血浆及胃癌组织中均为高表达,并可稳定存在于不同温度、pH值、核糖核酸酶A的环境中。通过与另外4个lncRNA联合应用,在对胃癌的诊断和预后监测上,具有较癌组织抗原更高的敏感性和特异性。因此,本研究结果对于LINC00857在胰腺导管腺癌诊断和治疗上的探索,具有非常重要的指导意义。

在本研究中,LINC00857被发现6个靶基因,分别为C1orf116、ESRP1、GPCR5A、LIPH、MAL2、PLS1。C1orf116是一种蛋白编码基因,它的别名是SARG (specifically androgen-regulated gene)。SARG基因是在前列腺细胞亚系LNCaP-1F5中被首次发现的,该细胞系能够表达内源性的雄激素受体(AR)和糖皮质激素受体(GR)^[33]。GeneNote数据库([http://bioinformatics/weizmann.ac.il/cards](http://bioinformatics.weizmann.ac.il/cards))中的表达谱数据显示SARG仅仅在前列腺和肺组织中高表达。目前尚无SARG在胰腺中的相关报道。Ueda等^[34]报道ESRP1作为一种RNA结合蛋白,能调控成纤维细胞生长因子受体2(fibroblast growth factor receptor 2, FGFR-2)亚型的表达模式,减少上皮细胞向间质转化,从而抑制胰腺导管腺癌细胞的生长、迁移与侵袭能力,且其高表达与PDAC较长的生存期相关。该研究提示ESRP1可作为PDAC的肿瘤抑制因子,与本研究结果不相符。Zhou等^[35]发现无论在胰腺癌细胞系或者胰腺癌组织中,GPCR5A均为高表达,并且GPCR5A的表达上调能

促进胰腺癌细胞的增殖、迁移和对吉西他滨的化疗抵抗,从而认为GPRC5A在胰腺癌中发挥的是促癌基因的作用,与本研究结果一致。Cui等^[36]研究报道LIPH在乳腺癌组织中呈高表达,并与肿瘤大小、病理分级、淋巴结和远处转移密切相关。而且,LIPH高表达的患者预后不良。相反,LIPH在食管癌和肺癌组织的高表达提示有更好的预后,推测可能与LIPH仅仅引起肿瘤的发生,而不影响肿瘤的转移有关^[37-38]。目前尚无LIPH在胰腺癌中的相关报道。Arumugam等^[39]研究发现MAL2作为上皮细胞的标志物,在胰腺癌中表达下调;当上皮间质转化被抑制时其表达增加,并可恢复胰腺癌对吉西他滨等化疗药物的敏感性。Protein kinase Ciota (PKCiota)是非小细胞肺癌的癌基因。Erdogan等^[40]通过Meta分析发现在肺腺癌中PLS1与PKCiota的表达相关,进一步的实验证实了它们在肺腺癌中均为高表达,且PLS1作为PKCiota的靶基因在肺腺癌细胞的增殖和侵袭中发挥重要作用。目前尚无PLS1在胰腺癌和其他癌症中研究报道。

通过以上对LINC00857靶基因的相关文献进行挖掘发现,LINC00857的靶基因ESRP1和MAL2均可作为上皮细胞的标志物,都参与了胰腺癌的上皮间质转化过程,由此推测LINC00857可能通过调控上皮间质转化而影响胰腺癌的发生发展。但是,这需要进一步的实验证实。

参考文献

- [1] Ying H, Dey P, Yao W, et al. Genetics and biology of pancreatic ductal adenocarcinoma[J]. *Genes Dev*, 2016, 30(4):355-385. doi: 10.1101/gad.275776.115.
- [2] Liang C, Shi S, Meng Q, et al. Complex roles of the stroma in the intrinsic resistance to gemcitabine in pancreatic cancer: where we are and where we are going[J]. *Exp Mol Med*, 2017, 49(12):e406. doi: 10.1038/emmm.2017.255.
- [3] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018[J]. *CA Cancer J Clin*, 2018, 68(1):7-30. doi: 10.3322/caac.21442.
- [4] Hruban RH, Canto M, Goggins M, et al. Update on familial pancreatic cancer[J]. *Adv Surg*, 2010, 44:293-311.
- [5] Fogel EL, Shahda S, Sandrasegaran K, et al. A Multidisciplinary Approach to Pancreas Cancer in 2016: A Review[J]. *Am J Gastroenterol*, 2017, 112(4):537-554. doi: 10.1038/ajg.2016.610.
- [6] Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs[J]. *Cell*, 2009, 136(4):629-641. doi: 10.1016/j.cell.2009.02.006.
- [7] Seton-Rogers S. Non-coding RNAs: the cancer X factor[J]. *Nat Rev Cancer*, 2013, 13(4):224-225. doi: 10.1038/nrc3489.
- [8] Hung T, Chang HY. Long noncoding RNA in genome regulation: prospects and mechanisms[J]. *RNA Biol*, 2010, 7(5):582-585.
- [9] Sun YW, Chen YF, Li J, et al. A novel long non-coding RNA ENST00000480739 suppresses tumour cell invasion by regulating OS-9 and HIF-1 α in pancreatic ductal adenocarcinoma[J]. *Br J Cancer*, 2014, 111(11):2131-2141. doi: 10.1038/bjc.2014.520.
- [10] Li Z, Zhao X, Zhou Y, et al. The long non-coding RNA HOTTIP promotes progression and gemcitabine resistance by regulating HOXA13 in pancreatic cancer[J]. *J Transl Med*, 2015, 13:84. doi: 10.1186/s12967-015-0442-z.
- [11] Ye Y, Chen J, Zhou Y, et al. High expression of AFAP1-AS1 is associated with poor survival and short-term recurrence in pancreatic ductal adenocarcinoma[J]. *J Transl Med*, 2015, 13:137. doi: 10.1186/s12967-015-0490-4.
- [12] Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression[J]. *Genome Res*, 2012, 22(9):1775-1789. doi: 10.1101/gr.132159.111.
- [13] Dai Z, Sheridan JM, Gearing LJ. edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens[J]. *Version 2. F1000Res*, 2014, 3:95. doi: 10.12688/f1000research.3928.2.
- [14] Law CW, Chen Y, Shi W, et al. voom: precision weights unlock linear model analysis tools for RNA-seq read counts[J]. *Genome Biol*, 2014, 15(2):R29. doi: 10.1186/gb-2014-15-2-r29.
- [15] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. *Nucleic Acids Res*, 2015, 43(7):e47. doi: 10.1093/nar/gkv007.
- [16] Cai H, Chen H, Yi T, et al. VennPlex-A Novel Venn Diagram Program for Comparing and Visualizing Datasets with Differentially Regulated Datapoints[J]. *PLoS One*, 2013, 8(1):e53388. doi: 10.1371/journal.pone.0053388.
- [17] Liang M, Zhang F, Jin G, et al. FastGCN: a GPU accelerated tool for fast gene co-expression networks[J]. *PLoS One*, 2015, 10(1):e0116776. doi: 10.1371/journal.pone.0116776.
- [18] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. *OMICS*, 2012, 16(5):284-287. doi: 10.1089/omi.2011.0118.
- [19] Wang Z, Jensen MA, Zenklusen JC. A Practical Guide to The Cancer Genome Atlas (TCGA)[J]. *Methods Mol Biol*, 2016, 1418:111-141. doi: 10.1007/978-1-4939-3578-9_6.
- [20] Chu C, Qu K, Zhong F, et al. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions[J]. *Mol Cell*, 2011, 44(4):667-678. doi: 10.1016/j.molcel.2011.08.027.

- [21] Engreitz JM, Pandya-Jones A, McDonel P, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome[J]. *Science*, 2013, 341(6147):1237973. doi: 10.1126/science.1237973.
- [22] Hung T, Wang Y, Lin MF, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters[J]. *Nat Genet*, 2011, 43(7):621–629. doi: 10.1038/ng.848.
- [23] Kretz M, Siprashvili Z, Chu C, et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR[J]. *Nature*, 2013, 493(7431):231–235. doi: 10.1038/nature11661.
- [24] Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?[J]. *Cell*, 2011, 146(3):353–358. doi: 10.1016/j.cell.2011.07.014.
- [25] Wang P, Xue Y, Han Y, et al. The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation[J]. *Science*, 2014, 344(6181):310–313. doi: 10.1126/science.1251456.
- [26] Yao K, Hua L, Wei L, et al. Correlation Between CASC8, SMAD7 Polymorphisms and the Susceptibility to Colorectal Cancer: An Updated Meta-Analysis Based on GWAS Results[J]. *Medicine (Baltimore)*, 2015, 94(46):e1884. doi: 10.1097/MD.0000000000001884.
- [27] Ma G, Gu D, Lv C, et al. Genetic variant in 8q24 is associated with prognosis for gastric cancer in a Chinese population[J]. *J Gastroenterol Hepatol*, 2015, 30(4):689–695. doi: 10.1111/jgh.12801.
- [28] Ghousaini M, Song H, Koessler T, et al. Multiple loci with different cancer specificities within the 8q24 gene desert[J]. *J Natl Cancer Inst*, 2008, 100(13):962–966. doi: 10.1093/jnci/djn190.
- [29] Hu L, Chen SH, Lv QL, et al. Clinical Significance of Long Non-Coding RNA CASC8 rs10505477 Polymorphism in Lung Cancer Susceptibility, Platinum-Based Chemotherapy Response, and Toxicity[J]. *Int J Environ Res Public Health*, 2016, 13(6):pii: E545. doi: 10.3390/ijerph13060545.
- [30] Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome[J]. *Nat Genet*, 2015, 47(3):199–208. doi: 10.1038/ng.3192.
- [31] Wang L, He Y, Liu W, et al. Non-coding RNA LINC00857 is predictive of poor patient survival and promotes tumor progression via cell cycle regulation in lung cancer[J]. *Oncotarget*, 2016, 7(10):11487–11499. doi: 10.18632/oncotarget.7203.
- [32] Zhang K, Shi H, Xi H, et al. Genome-Wide lncRNA Microarray Profiling Identifies Novel Circulating lncRNAs for Detection of Gastric Cancer[J]. *Theranostics*, 2017, 7(1):213–227. doi: 10.7150/thno.16044.
- [33] Cleutjens C, Steketee K, van EC, et al. Both androgen receptor and glucocorticoid receptor are able to induce prostate-specific antigen expression, but differ in their growth-stimulating properties of LNCaP cells[J]. *Endocrinology*, 1997, 138(12):5293–5300. doi: 10.1210/en.138.12.5293.
- [34] Ueda J, Matsuda Y, Yamahatsu K, et al. Epithelial Splicing Regulatory Protein 1 is a Favorable Prognostic Factor in Pancreatic Cancer that Attenuates Pancreatic Metastases[J]. *Oncogene*, 2014, 33(36):4485–4495. doi: 10.1038/onc.2013.392.
- [35] Zhou H, Telonis A, Jing Y, et al. GPRC5A is a potential oncogene in pancreatic ductal adenocarcinoma cells that is upregulated by gemcitabine with help from HuR[J]. *Cell Death Dis*, 2016, 7:e2294. doi: 10.1038/cddis.2016.169.
- [36] Cui M, Jin H, Shi X, et al. Lipase member H is a novel secreted protein associated with a poor prognosis for breast cancer patients[J]. *Tumour Biol*, 2014, 35(11):11461–11465. doi: 10.1007/s13277-014-2436-5.
- [37] Seki Y, Yoshida Y, Ishimine H, et al. Lipase member H is a novel secreted protein selectively upregulated in human lung adenocarcinomas and bronchioloalveolar carcinomas[J]. *Biochem Biophys Res Commun*, 2014, 443(4):1141–1147. doi: 10.1016/j.bbrc.2013.12.106.
- [38] Ishimine H, Zhou R, Sumitomo K, et al. Lipase member H frequently overexpressed in human esophageal adenocarcinomas[J]. *Tumour Biol*, 2016, 37(2):2075–2081. doi: 10.1007/s13277-015-3985-y.
- [39] Arumugam T, Ramachandran V, Fournier KF, et al. Epithelial to mesenchymal transition contributes to drug resistance in pancreatic cancer[J]. *Cancer Res*, 2009, 69(14):5820–5828. doi: 10.1158/0008-5472.CAN-08-2819.
- [40] Erdogan E, Klee EW, Thompson EA, et al. Meta-analysis of oncogenic protein kinase Ciota signaling in lung adenocarcinoma[J]. *Clin Cancer Res*, 2009, 15(5):1527–1533. doi: 10.1158/1078-0432.CCR-08-2459.

(本文编辑 宋涛)

本文引用格式: 张志鹏, 孙维佳, 陈泓西, 等. 基于生物信息学的胰腺导管腺癌预后风险长链非编码RNA筛选[J]. 中国普通外科杂志, 2018, 27(9):1126–1134. doi:10.7659/j.issn.1005-6947.2018.09.007
Cite this article as: Zhang ZP, Sun WJ, Chen HX, et al. Identification of prognostic risk long noncoding RNAs for pancreatic ductal adenocarcinoma by bioinformatics analysis[J]. *Chin J Gen Surg*, 2018, 27(9):1126–1134. doi:10.7659/j.issn.1005-6947.2018.09.007