

基于多组学数据构建肺腺癌预后相关风险预测模型

李鑫¹, 魏铎沛¹, 吕章艳¹, 冯小双¹, 李江¹, 陈宏达¹, 温艳¹, 高禹舜², 薛奇², 高树庚², 谭锋维^{2*}

¹国家癌症中心/国家肿瘤临床医学研究中心/中国医学科学院北京协和医学院肿瘤医院癌症早诊早治办公室, ²胸外科, 北京100021

[摘要] **目的:**通过整合肺腺癌(lung adenocarcinoma, LUAD)临床信息组、基因组和转录组的多组学信息,构建预后相关风险预测模型,为肺腺癌患者的预后风险预测提供依据。**方法:**通过肿瘤基因组图谱(the cancer genome atlas, TCGA)网站下载522例肺腺癌患者数据,完成肺腺癌患者多组学数据库构建。对经过质量控制的数据库,采用Cox比例风险模型评价多组学数据对LUAD预后的影响,构建基于临床信息、基因分型和基因表达多组学信息的预后风险预测模型,并使用受试者工作特征(receiver operating characteristic, ROC)曲线和曲线下面积(area under curve, AUC)评估预后模型风险识别的能力。**结果:**共386例肺腺癌患者纳入多组学数据库,分别筛选出1个临床信息组、5个基因组单核苷酸多态性(single nucleotide polymorphisms, SNP)和4个转录组mRNA预后相关特征因素与肺腺癌的预后相关。与肺腺癌低风险组相比,高风险组患者较低风险组的预后更差(HR=3.67, 95%CI=2.46~5.49, $P=1.96\times 10^{-10}$),联合基因组、转录组与临床信息构建多组学预后指数后,其ROC曲线下面积为0.694(95%CI=0.633~0.754)。**结论:**采用联合临床信息组、基因组和转录组特征的多组学信息构建的预后风险预测指数可有效区分肺腺癌患者预后好坏,其研究结果将为肺腺癌患者的精准治疗提供有效支撑。

[关键词] 肺腺癌; SNP; mRNA; 预后; TCGA

[中图分类号] R734.2

[文献标志码] A

[文章编号] 1007-4368(2018)12-1819-07

doi:10.7655/NYDXBNS20181243

Prediction model of prognosis risk of lung adenocarcinoma based on multi genomic data

Li Xin¹, Wei Luopei¹, Lü Zhangyan¹, Feng Xiaoshuang¹, Li Jiang¹, Chen Hongda¹, Wen Yan¹, Gao Yushun², Xue Qi², Gao Shugeng², Tan Fengwei^{2*}

¹Office of Cancer Screening, ²Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

[Abstract] **Objective:** By integrating the clinical information group, genome and transcriptome of lung adenocarcinoma, a prognostic risk prediction model was constructed to provide evidence of lung adenocarcinoma patients. **Methods:** The data of 522 patients with lung adenocarcinoma were downloaded from TCGA website, and the multi group database of lung adenocarcinoma was completed. Cox proportional hazard model was used to evaluate the effect of multigroup data on the prognosis of LUAD. A prognostic risk prediction model based on clinical information, genotyping and multigroup information of gene expression was constructed. ROC curves and AUC was used to assess the ability of risk recognition in prognostic models. **Results:** A total of 386 patients with lung adenocarcinoma were included in the multi group database. One clinical information group, five genomes and four transcription groups for prognostic factors were detected to associate with the prognosis of lung adenocarcinoma. HR (95% CI) of prognosis of lung adenocarcinoma for high-risk group was 3.67 (95% CI=2.46~5.49), as compared with low-risk group. After combining genome, transcriptome and clinical information, the ROC curve is 0.694 (95% CI=0.633~0.754). **Conclusion:** Prognostic risk prediction index combining clinical

[基金项目] 国家重点研发计划(2018YFC1315000, 2017YFC0907900, 2016YFC1302500, 2016YFC0905400); 国家自然科学基金(81871885, 81673265); 中央保健专项资金(W2017BJ39); 北京协和医学院协和青年科研基金(2017320013); 北京市科技计划首都临床特色应用研究(Z181100001718212), 中国医学科学院医学与健康科技创新工程(2017-12M-1-005); 中国医学科学院肿瘤医院院所科研课题(LC2017A01, LC2017D01); 北京市优秀人才培养资助(2017000021223TD05)

*通信作者(Corresponding author), E-mail: tanfengwei@126.com

information group, genome and transcriptome features can effectively distinguish the prognosis of lung adenocarcinoma patients. The results will provide effective support for accurate treatment of lung adenocarcinoma patients.

[Key words] Lung Adenocarcinoma, SNP, mRNA, Prognosis, TCGA

[Acta Univ Med Nanjing, 2018, 38(12): 1819-1822]

肺腺癌是非小细胞肺癌(non-small cell lung cancer, NSCLC)中的一种常见亚型,大约占肺原发肿瘤的40%,其恶性程度虽较低,但由于其易复发,仍有超过30%的患者面临着术后复发的风险^[1]。另外由于肿瘤异质性的存在,包括患者的临床特征、分子表型等多种因素的差异,使得肺腺癌的个性化治疗面临巨大的挑战^[2]。近年来,多项研究证实了基因组信息和转录组信息在肿瘤预后中的预测作用,综合分析基因组和转录组的信息有助于对肿瘤发展及治疗过程的理解,并对肿瘤患者的预后具有指导意义^[3-7]。

随着大规模高通量测序技术与基因组芯片技术的快速发展,TCGA计划已成功绘制了人类肿瘤的基因组图谱^[8]。该数据库涵盖了33种肿瘤及其丰富的遗传信息和详尽的临床信息,包括基因组单核苷酸变异数据、基因组拷贝数变异数据、体细胞突变数据、转录组数据、DNA甲基化数据、蛋白表达数据以及临床相关数据,其数据可通过公共平台(<https://portal.gdc.cancer.gov/>)免费下载,为各肿瘤研究人员的数据挖掘和分析研究提供了便捷的多组学数据资源^[9-11]。尽管基于TCGA数据库的肺腺癌转录组预后分析已得到了广泛的研究,包括miRNA^[12]、lncRNA^[13]、mRNA^[7]等,但这些研究均限于单个组学数据,多组学信息联合用于肺腺癌预后预测尚缺乏数据支撑。

本研究基于TCGA公共数据库中肺腺癌的基因组、转录组和临床信息组的多组学信息进行分析,利用Cox回归和LASSO回归筛选与肺腺癌预后相关的临床特征、SNPs和基因构建风险模型,并将其应用于肺腺癌的预后预测。

1 资料和方法

1.1 资料

研究涉及的数据资源下载于TCGA在线公共数据库(<https://portal.gdc.cancer.gov/>),包括522例肺腺癌患者的基因分型数据、mRNA表达数据和临床基

本信息及预后数据。数据库中患者的基因分型数据由Affymetrix SNP Array 6.0芯片完成分型,mRNA表达数据由Illumina HiSeq和Illumina Genome Analyzer平台完成检测,临床基本信息涵盖年龄、性别、吸烟状态、TNM分期、生存状态和生存时间。

1.2 方法

本研究通过对临床信息组、基因组和转录组信息进行质控和筛选,得到了预后相关的临床特征、SNPs和基因,用于计算总预后指数。

1.2.1 质量控制和基因型填补

对于由Affymetrix SNP Array 6.0芯片获得的基因分型数据,采用SHAPEIT^[14]和Impute2^[15]软件进行了基因型填补,使用千人基因组计划(1000 Genomes Project)Phase III(发布时间:2013年5月2日)的数据作为填补的参考基因组。此外,为避免患者死于手术并发症等其他非肿瘤因素而造成的偏倚,本研究所纳入患者的生存时间均超过30 d。

1.2.2 预后相关基因组特征筛选

首先对填补后的分型数据进行了质控,质控标准包括:①去除填补质量指标INFO<0.9的位点;②去除不符合哈迪-温伯格平衡($P < 1 \times 10^{-5}$)的位点;③去除最小的等位基因频率(Minor Allele Frequency, MAF)<0.05的位点。通过质量控制,填补后的分型数据共保留了4 631 327个位点。随后,我们对这些位点进行了功能注释,筛选了682个潜在的功能性位点,功能注释方法包括:①基于整合的独立于注释的消减算法评分(combined annotation dependent depletion, CADD)>20;②位于启动子区、增强子区或DNA酶高敏位点(DNase Hypersensitive Site)区域。功能注释数据来源于“DNA元件百科全书”计划(encyclopedia of DNA elements, ENCODE)、哺乳动物基因组功能注释计划(functional annotation of the mammalian genome, FANTOM)和CADD网站(<https://cadd.gs.washington.edu/>)。对于筛选得到的潜在功能性位点,使用Cox比例风险模型评估与肺腺癌预后的关联,并将与肺腺癌预后显著相关的47个

SNPs纳入后续的分析,再使用LASSO降维,最终获得用于肺腺癌预后模型的5个SNPs。

1.2.3 预后相关转录组特征筛选

对于TCGA数据库下载的60 483个基因的表达数据,我们筛选了459个由IntOGen平台(<https://www.intogen.org/search>)获得的肿瘤驱动基因,并剔除了在不到一半(386/2=193)的样本中其标化reads数 ≥ 5 的基因,共得到284个基因。对于筛选得到的基因,使用Cox比例风险模型评估与肺腺癌预后的关联,并将与肺腺癌预后显著相关的51个基因纳入后续的分析,再使用LASSO降维,最终获得用于肺腺癌预后模型的4个基因。

1.2.4 预后风险模型构建

基于筛选得到的SNPs和基因,构建了遗传预后指数(genetic prognostic index, GPI),该指数综合了预后相关SNPs的分型信息和预后相关基因的mRNA表达信息,用于对肺腺癌的预后状态进行预测,计算公式如下:

$$GPI = \sum_{i=1}^4 \beta_i SNP_i + \sum_{j=1}^3 \beta_j Gene_j$$

其中 β 为Cox模型中各SNPs或基因的回归系数,SNP为各位点的基因分型,Gene为各基因的mRNA表达量。根据公式计算出每位患者的遗传预后指数,并以其中位数作为界值,将患者分为高风险组和低风险组,采用Kaplan-Meier生存区间描述高、低风险组患者的总体生存率,用Log-rank检验比较组间总体生存率的差异,并采用ROC曲线及AUC评估该指数对于结局变量的预测性能。

1.3 统计学方法

采用Cox回归模型的风险比(hazard ratio, HR)及其95%可信区间(confidence interval, CI)评价SNPs及基因mRNA表达与预后的关联,并调整年龄,性别,吸烟和TNM分期。采用LASSO回归^[16]对候选的预后相关SNPs及基因降维并选择最终的预测变量。采用ROC曲线及AUC评估遗传预测指数的辨别能力。以 $P < 0.05$ 为统计学检验水准,所有的数据分析及图形绘制均使用R 3.5.1软件。

2 结果

2.1 肺腺癌患者的临床基本情况

本研究纳入了TCGA项目的386例肺腺癌患者,平均年龄为(65.48 \pm 9.97)岁,中位生存时间为23.53个月。其中男177例(45.9%),女209例(54.1%);非吸烟者49例(13.1%),吸烟者326例(86.9%);肺腺癌TNM分期I-II期300例(77.7%),III-IV期86例(22.3%)。(表1)

2.2 预后相关的基因组特征

通过Cox比例风险模型评估了682个潜在功能性SNPs与肺腺癌预后的关联,共有47个与肺腺癌预后显著相关的SNPs纳入后续的分析。经过LASSO降维,最终得到了5个与肺腺癌预后相关的SNPs(rs1371906, rs9384411, rs10982846, rs2013335, rs8078328),各SNPs与肺腺癌预后的关联见表2。

2.3 预后相关的转录组特征

通过Cox比例风险模型评估了284个候选基因与肺腺癌预后的关联,共有51个与肺腺癌预后显著

表1 肺腺癌患者的基本临床病理资料情况

Table 1 Basic clinical and pathological data of patients with lung adenocarcinoma

临床变量	存活(%)N=276	死亡(%)N=110	中位生存时间(月)	HR(95% CI)*	P值*
年龄(岁)					
≤65	131(47.46)	49(44.55)	24.37	1	
>65	145(52.54)	61(55.45)	21.77	1.31(0.88~1.95)	1.78×10 ⁻¹
性别					
男性	127(46.01)	50(45.45)	21.53	1	
女性	149(53.99)	60(54.55)	26.78	1.08(0.71~1.63)	7.21×10 ⁻¹
吸烟状态					
不吸烟	37(13.65)	12(11.54)	26.37	1	
吸烟	234(86.35)	92(88.46)	23.53	0.90(0.47~1.72)	7.58×10 ⁻¹
TNM分期					
I-II期	232(84.06)	68(61.82)	29.87	1	
III-IV期	44(15.94)	42(38.18)	13.32	2.84(1.89~4.26)	4.42×10 ⁻⁷

*调整年龄、性别、吸烟状态和TNM分期。

相关的基因纳入后续的分析。经过LASSO降维,我们最终得到了4个与肺腺癌预后相关的基因(COL1A1, LDHA, KLF4, EPHA2),各基因的mRNA表达与肺腺癌预后的关联见表2。

2.4 遗传预后指数的构建与评价

基于上述与肺腺癌患者预后相关的5个SNPs和4个基因,构建了遗传预后指数, $GPI = (-0.380 \times rs1371906) + (0.565 \times rs9384411) + (-0.379 \times$

表2 模型中SNP位点及基因mRNA表达与肺腺癌预后的关联

Table 2 Association between SNPs and mRNA expression in prognosis of lung adenocarcinoma

遗传特征	样本分布频率 ^a		MAF (等位基因) ^b	HR(95% CI) ^c	P值 ^d
	死亡	存活			
rs1371906	52/50/8	124/115/37	0.330(G)	0.69(0.50~0.95)	2.40×10^{-2}
rs9384411	30/55/25	99/133/44	0.422(C)	1.46(1.09~1.97)	1.18×10^{-2}
rs10982846	64/41/5	141/107/28	0.277(G)	0.68(0.48~0.96)	2.88×10^{-2}
rs2013335	33/46/31	80/133/63	0.475(A)	1.50(1.12~1.99)	5.93×10^{-3}
rs8078328	28/53/29	93/125/58	0.456(A)	1.40(1.08~1.83)	1.22×10^{-2}
COL1A1	26/28/23/33	71/68/73/64	-	1.15(1.02~1.30)	2.55×10^{-2}
LDHA	22/26/19/43	75/70/77/54	-	1.75(1.33~2.31)	6.51×10^{-5}
KLF4	24/26/29/31	73/70/67/66	-	1.29(1.08~1.54)	5.52×10^{-3}
EPHA2	20/31/29/30	77/65/67/67	-	1.32(1.10~1.58)	3.35×10^{-3}

^a在SNP位点中分别对应野生纯合型、杂合型及突变纯合型,在mRNA表达中分别对应0%~25%、25%~50%、50%~75%、75%~100%;^bMAF:最小等位基因频率;^c调整年龄、性别、吸烟状态和TNM分期。

$rs10982846) + (0.311 \times rs2013335) + (0.294 \times rs8078328) + (0.067 \times COL1A1) + (0.497 \times LDHA) + (0.188 \times KLF4) + (0.157 \times EPHA2)$ 。根据该公式,每位患者均获得一个预后指数,取中位数(5.71)作为界值,所有的肺腺癌患者可划分为高风险组($n=193$)和低风险组($n=193$)。如图2中的Kaplan-Meier生存曲线所示,高风险组患者的总体生存率较低,且Log-rank检验提示高、低风险组患者间生存率的差异具有统计学意义($P=1.42 \times 10^{-11}$)。Cox回归的结果也提示高风险组患者的死亡风险更高($HR=3.67$, $95\% CI=2.46 \sim 5.49$, $P=1.96 \times 10^{-10}$),且能与临床指数结合共同提示患者的死亡风险(表3)。采用ROC曲

线对遗传预后指数的预测性进行评价,如图3所示,遗传指数与临床指数结合后能够提高整体的预测性能,其AUC可达到0.694(95% CI=0.633~0.754)。

3 讨论

本研究通过对386例肺腺癌患者的基因组和转录组数据进行筛选,共得到与肺腺癌患者的预后存在显著关联的5个SNPs和4个基因,并用于预测模型的构建。利用模型计算得到的遗传预后指数,将肺腺癌患者分为高风险组和低风险组,生存分析结果显示,组间的生存率有显著差异,高风险组与低风险组相比,其死亡的风险明显增加。此外,整合

表3 三种预后指数与肺腺癌预后的关联

Table 3 Correlation between three prognostic indices and prognosis of lung adenocarcinoma

预后指数	存活(%)N=276	死亡(%)N=110	中位生存时间(月)	HR(95% CI)	P值
遗传指数					
低风险	154(55.80)	39(35.45)	38.90	1.00	
高风险	122(44.20)	71(64.55)	15.33	3.67(2.46~5.49)	1.96×10^{-10}
临床指数					
低风险	176(63.77)	36(32.73)	38.37	1.00	
高风险	100(36.23)	74(67.27)	19.27	3.46(2.31~5.19)	1.83×10^{-9}
遗传+临床					
低风险	158(57.25)	35(31.82)	37.83	1.00	
高风险	118(42.75)	75(68.18)	19.40	3.49(2.31~5.26)	2.47×10^{-9}

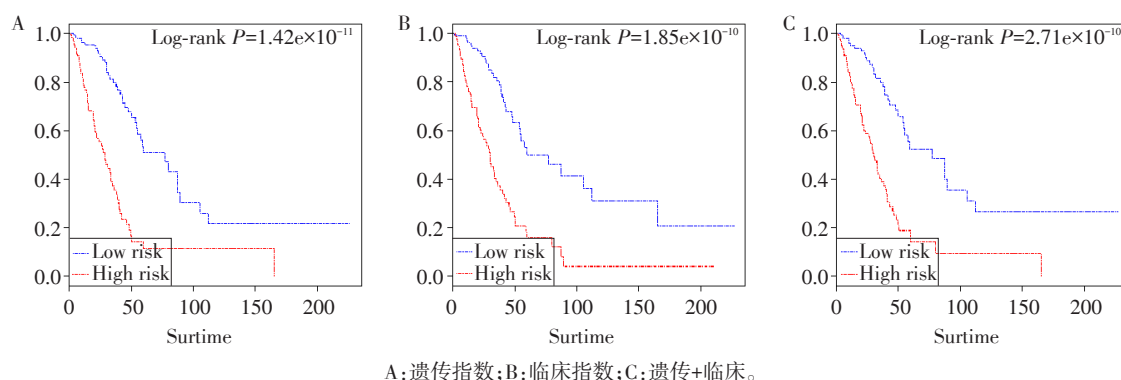
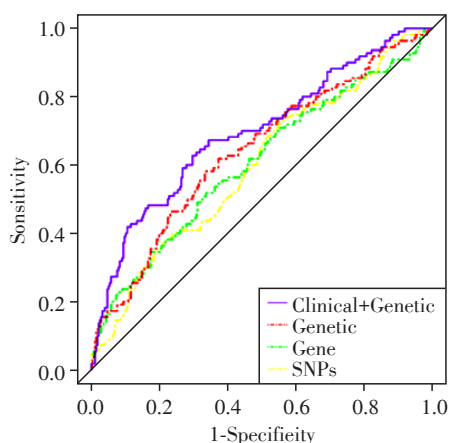


图2 高风险组和低风险组Kaplan-Meier生存曲线比较

Figure 2 Comparison of Kaplan-Meier survival curves between high and low risk group



注:SNPs,基因,遗传指数,遗传+临床指数ROC曲线

图3 遗传预后指数预测评价ROC曲线

Figure 3 Prediction of ROC curve by genetic prognostic index

遗传指数与临床指数的Cox回归分析和ROC曲线分析均显示,该遗传预后指数能够与临床指数联合共同对肺腺癌患者的预后进行预测。

风险模型中的5个SNPs,有2个SNPs对肺腺癌患者的预后起保护性作用。其中rs1371906位于IL20RB基因的下游(3q22.3),IL20RB编码白细胞介素-20(IL-20)受体的一个亚单位,有研究表明,在早期进展的乳腺癌患者中IL20RB的表达降低^[17],提示IL20RB的低表达可能与肿瘤的不良预后相关;由ENCODE的数据可知,rs1371906与增强子的活化相关,推测该位点可能通过调控IL20RB的转录,从而影响肺腺癌患者的预后。rs9384411位于miR-1202的下游(6q25.3),已有报道miR-1202能够抑制肿瘤细胞的迁移和侵袭^[18],并可以作为子宫内膜癌^[19]、乳腺癌^[20]、胃癌^[21]、肝癌等^[18]多种肿瘤的预后标志物。在增加肺腺癌患者死亡风险的3个SNPs中,

rs10982846位于DEC1的上游(9q33.1),且其所处区域在肺癌细胞系(A549细胞系)中有较强的增强子信号,提示其在肺癌细胞中具有转录调控活性。已有报道表明,DEC1的表达与食管鳞癌的淋巴结转移有关^[22],且其能够通过延缓细胞周期来调节乳腺癌细胞的增殖^[23],此外也有研究发现,DEC1能够通过促进细胞生长和侵袭来维持甲状腺癌的进展^[24]。rs2013335位于KRT78基因的第一外显子(12q13.13),是一个错义突变。KRT78是II型角蛋白基因家族的成员,编码具有中间丝状结构域的蛋白质,已被证明可以作为HPV+宫颈癌^[25]、转移性黑色素瘤^[26]和口腔鳞状细胞癌^[27]的标志物。rs8078328位于PIPOX基因的上游(17q11.2),Cha等^[28]的研究发现,在乳腺癌肺转移患者中,PIPOX表达与较短的总生存期相关。

筛选得到的4个基因均与肺腺癌患者的死亡风险增加有关,其中COL1A1已被报道与肾细胞癌^[29]、卵巢癌^[30]、胃癌^[31]、结肠癌^[32]等的不良预后有关,可作为相应肿瘤的预后标志物。LDHA与多种肿瘤相关,其过表达能促进胶质瘤细胞的增殖和侵袭^[33],并且可作为肾透明细胞癌不良预后的标志物^[34]。KLF4编码的蛋白质属于Kruppel转录因子家族,该蛋白通过介导p53来调控DNA损伤后细胞周期的转变。Halim等^[35]的研究发现,KLF4的激活与结直肠癌的组织重塑、髓细胞浸润和预后不良有关。EPHA2在细胞信号传导、细胞运动和存活中发挥作用,是Ewing肉瘤发生转移的关键因素^[36],此外,其高表达也与胃癌患者的不良预后相关^[37]。

本研究也存在一些局限性,本研究未考虑体细胞突变,结构变异,甲基化^[38]等组学水平的变化与肺腺癌的关系,将在后续文章中分析。总之,本研

究通过整合运用TCGA的基因组和转录组信息,筛选出了与肺腺癌预后相关的5个SNPs和4个基因。遗传与临床指数整合后,与单独作用相比,HR值没有明显的增加,提示遗传指数和临床指数间可能存在相互影响,因而整合多组学信息更有利于全面评估患者的预后风险,同时也为后续的分析研究提供了思路。基于基因组和转录组遗传信息构建的预测模型能够很好地识别预后不良高风险的肺腺癌患者,且能与临床预后因素共同对患者的预后进行预测,故能为肺腺癌患者的预后风险评估提供依据。

[参考文献]

- [1] Goodgame B, Viswanathan A, Miller CR, et al. A clinical model to estimate recurrence risk in resected stage I non-small cell lung cancer[J]. *Am J Clin Oncol*, 2008, 31(1): 22-28
- [2] Yoshizawa A, Motoi N, Riely GJ, et al. Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases[J]. *Mod Pathol*, 2011, 24(5): 653-664
- [3] Sun Q, Zong L, Zhang H, et al. A 10microRNA prognosis scoring system in esophageal squamous cell carcinoma constructed using bioinformatic methods [J]. *Mol Med Rep*, 2018, 17(4): 5222-5228
- [4] Zhang C, Sun Q. Weighted gene co-expression network analysis of gene modules for the prognosis of esophageal cancer [J]. *J Huazhong Univ Sci Technolog Med Sci*, 2017, 37(3): 319-325
- [5] Shukla S, Evans JR, Malik R, et al. Development of a RNA-Seq Based Prognostic Signature in Lung Adenocarcinoma [J]. *J Natl Cancer Inst*, 2017, 109(1): djw200
- [6] Liu C, Wang X, Genchev GZ, et al. Multi-omics facilitated variable selection in Cox-regression model for cancer prognosis prediction[J]. *Methods*, 2017, 124: 100-107
- [7] Zhao Q, Shi X, Xie Y, et al. Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA [J]. *Brief Bioinform*, 2015, 16(2): 291-303
- [8] Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas(TCGA): an immeasurable source of knowledge[J]. *Contemp Oncol(Pozn)*, 2015, 19(1A): A68-77
- [9] Hoadley KA, Yau C, Hinoue T, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer [J]. *Cell*, 2018, 173(2): 291-304 e296
- [10] Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic signaling pathways in the cancer genome atlas [J]. *Cell*, 2018, 173(2): 321-337 e310
- [11] Ding L, Bailey MH, Porta-Pardo E, et al. Perspective on oncogenic processes at the end of the beginning of cancer genomics[J]. *Cell*, 2018, 173(2): 305-320 e310
- [12] Sui J, Yang RS, Xu SY, et al. Comprehensive analysis of aberrantly expressed microRNA profiles reveals potential biomarkers of human lung adenocarcinoma progression [J]. *Oncol Rep*, 2017, 38(4): 2453-2463
- [13] Zheng S, Zheng D, Dong C, et al. Development of a novel prognostic signature of long non-coding RNAs in lung adenocarcinoma[J]. *J Cancer Res Clin Oncol*, 2017, 143(9): 1649-1657
- [14] Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes [J]. *Nat Methods*, 2011, 9(2): 179-181
- [15] Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies [J]. *PLoS Genet*, 2009, 5(6): e1000529
- [16] Tibshirani R. The lasso method for variable selection in the Cox model[J]. *Stat Med*, 1997, 16(4): 385-395
- [17] Omarini C, Bettelli S, Caprera C, et al. Clinical and molecular predictors of long-term response in HER2 positive metastatic breast cancer patients [J]. *Cancer Biol Ther*, 2018, 19(10): 879-886
- [18] Du B, Zhang P, Tan Z, et al. MiR-1202 suppresses hepatocellular carcinoma cells migration and invasion by targeting cyclin dependent kinase 14 [J]. *Biomed Pharmacother*, 2017, 96: 1246-1252
- [19] Chen H, Fan Y, Xu W, et al. Exploration of miR-1202 and miR-196a in human endometrial cancer based on high throughput gene screening analysis[J]. *Oncol Rep*, 2017, 37(6): 3493-3501
- [20] Hamam R, Ali AM, Alsaleh KA, et al. microRNA expression profiling on individual breast cancer patients identifies novel panel of circulating microRNA for early detection[J]. *Sci Rep*, 2016, 6: 25997
- [21] Tokuhsa M, Ichikawa Y, Kosaka N, et al. Exosomal miRNAs from peritoneum lavage fluid as potential prognostic biomarkers of peritoneal metastasis in gastric cancer [J]. *PLoS One*, 2015, 10(7): e0130472
- [22] Qian Y, Jung YS, Chen X. Differentiated embryo-chondrocyte expressed gene 1 regulates p53-dependent cell survival versus cell death through macrophage inhibitory cytokine-1 [J]. *Proc Natl Acad Sci USA*, 2012, 109(28): 11300-11305
- [23] Bi H, Li S, Qu X, et al. DEC1 regulates breast cancer cell proliferation by stabilizing cyclin E protein and delays the progression of cell cycle S phase [J]. *Cell Death Dis*,

- 2015,6:e1891
- [24] Gallo C, Fragliasso V, Donati B, et al. The bHLH transcription factor DEC1 promotes thyroid cancer aggressiveness by the interplay with NOTCH1 [J]. *Cell Death Dis*, 2018,9(9):871
- [25] Varga N, Mozes J, Keegan H, et al. The Value of a Novel panel of cervical cancer biomarkers for triage of HPV positive patients and for detecting disease progression [J]. *Pathol Oncol Res*, 2017, 23(2):295-305
- [26] Wang LX, Li Y, Chen GZ. Network-based co-expression analysis for exploring the potential diagnostic biomarkers of metastatic melanoma [J]. *PLoS One*, 2018, 13 (1) : e0190447
- [27] Zhang X, Feng H, Li Z, et al. Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis[J]. *Onco Targets Ther*, 2018, 11 :6001-6021
- [28] Cha YJ, Kim DH, Jung WH, et al. Expression of sarcosine metabolism-related proteins according to metastatic site in breast cancer[J]. *Int J Clin Exp Pathol*, 2014, 7(11) : 7824-7833
- [29] Yamada Y, Sugawara S, Arai T, et al. Molecular pathogenesis of renal cell carcinoma: Impact of the anti-tumor miR-29 family on gene regulation [J]. *Int J Urol*, 2018, 25 (11):953-965
- [30] Yang L, Jing J, Sun L, et al. Exploring prognostic genes in ovarian cancer stage-related coexpression network modules[J]. *Medicine(Baltimore)*, 2018, 97(34):e11895
- [31] Liu X, Wu J, Zhang D, et al. Identification of potential key genes associated with the pathogenesis and prognosis of gastric cancer based on integrated bioinformatics analysis [J]. *Front Genet*, 2018, 9:265
- [32] Zhou XG, Huang XL, Liang SY, et al. Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis[J]. *Onco Targets Ther*, 2018, 11 :2815-2830
- [33] Hu S, Jiang Q, Luo D, et al. miR-200b is a key regulator of tumor progression and metabolism targeting lactate dehydrogenase A in human malignant glioma [J]. *Oncotarget*, 2016, 7(30):48423-48431
- [34] Girgis H, Masui O, White NM, et al. Lactate dehydrogenase A is a potential prognostic marker in clear cell renal cell carcinoma[J]. *Mol Cancer*, 2014, 13:101
- [35] Halim S, Markert EK, Vazquez A. Analysis of cell proliferation and tissue remodelling uncovers a KLF4 activity score associated with poor prognosis in colorectal cancer [J]. *Br J Cancer*, 2018, 119(7):855-863
- [36] Garcia - Monclus S, Lopez - Alemany R, Almacellas - Rabaiget O, et al. EphA2 receptor is a key player in the metastatic onset of Ewing sarcoma[J]. *Int J Cancer*, 2018, 143(5):1188-1201
- [37] Inokuchi M, Nakagawa M, Baogok N, et al. Prognostic significance of high EphA1-4 expression levels in locally advanced gastric cancer[J]. *Anticancer Res*, 2018, 38(3) : 1685-1693
- [38] 桂 珍,严 枫. UHRF1 在表观遗传调控和肿瘤诊治中的研究进展[J].*南京医科大学学报(自然科学版)*, 2016,36(2):129-134

[收稿日期] 2018-09-03