通过生物信息学分析鉴定与非小细胞肺癌相关的关键差异表达基因

文章作者及期刊介绍

该文章由西南大学的李学刚教授带领其学生完成，于 2018 年发表于《分子医学报告》。

李学刚，西南大学药学院教授[1]，博士生导师[2]，1964 年 6 月生于四川省邻水县。1980～1987 年于哈尔滨工业大学，分别获得工学学士和硕士学位；1987～1990 年于北京大学，获得理学博士学位；1990～1992 年于中国科学院生物物理所从事分子酶学专业博士后研究工作。主要从事天然产物化学、药物化学、生物化学、食品化学等方面的研究工作。

《分子医学报告》是一份 SCI 期刊，影响因子 1.851，经同行评审，有印刷版和网络版，其中包括专门研究分子医学的文章，强调的方面包括药理学、病理学、遗传学、神经科学、传染病、分子心脏病学和分子外科学。关于各种疾病机制的实验模型系统的体内外研究为研究人员提供了必要的工具和知识，以帮助诊断和治疗人类疾病。

文章背景

肺癌占全世界所有癌症病例的十分之一以上，是最常见的恶性肿瘤之一，可能也是全世界与癌症相关的死亡的第一个原因。在中国，肺癌是目前最常见的癌症，2015 年有超过 733,300 个新病例和 610,200 人死亡,根据组织学亚分类，非小细胞肺癌和小细胞肺癌约 85%的肺癌患者被诊断为非小细胞肺癌（NSCLC）。迄今为止，手术切除仍是 NSCLC 唯一的治疗选择。然而，由于 NCLSC 早期无特异性症状，且缺乏准确、便捷的诊断方法进行早期检测，大多数患者在该疾病的晚期被确诊，手术切除已不再可行。

随着分子生物学的发展，分子靶向治疗 NSCLC 也取得了显著进展。例如，大约 10-20%的北美和西欧患者和 30-50%的亚洲 NSCLC 患者显示出 EGFR 基因突变。但是，使用 EGFR 酪氨酸激酶抑制剂(TKIs)，如厄洛替尼和吉非替尼，这些患者的反应良好。还发现 3-7%的 NSCLC 患者有活化的 ALK 基因,使用克里唑蒂尼治疗

的反应率升高 10 到 55%,同时增加了 6 个月无进展生存率 72%。然而,这种情况下只能代表一群非小细胞肺癌患者,由于耐药性,这些靶向治疗药物通常是暂时的。此外，尽管 NSCLC 靶向治疗取得了显著的成功，但其确切的分子机制仍未完全阐明。因此，为更有效的治疗策略，寻找更多的潜在靶点是非常重要的。

这篇文章旨在更好地了解关键差异基因（DEGs）对 SCLC 分子发病机制的影响

研究数据

系列微阵列数据的矩阵文件:GSE21933 GSE33532, GSE44077 和 GSE74706，在每个 GSE 文件中只选择非小细胞肺癌样本及其正常样本,获得 GSE21933 肿瘤样本 21 例，正常样本 21 例;GSE33532 肿瘤样本 80 例，正常样本 20 例;GSE44077 肿瘤样本 65 例，正常样本 65 例;GSE74706 肿瘤样本 18 例，正常样本 18 例。

研究方法

1.数据预处理和差异表达分析。

使用 R 软件和 R 包对每个系列的原始阵列数据进行分析。首先，使用 raffy 包中 RMA 算法对原始数据进行背景校正和分位数归一化。将肿瘤与正常组样本间的差异蛋白(DEGs)进行配对 t 检验。使用 Benjamini-Hochberg 法进行多重检验，得到调整后的 p 值。以|log2(差异倍数)|>1 且 P<0.05 作为显著差异表达基因筛选的阈值。为了进一步提高生物信息学分析的可靠性，使用 FunRich 来识别所有 4 个 GSE 文件中共存的重叠 DEGs。

2.GO 功能富集分析

对重叠 DEGs 进行 GO 功能注释和富集分析，确定基因集富集术语

3.PPI 网络构建和 hub 基因分析。

采用 STRING（一个有用的在线工具，专门用于评估 PPI 的相互作用信息）数据库及 CytoScape 构建 PPI 网络。利用 Cytoscape 中 cytoHubba 插件从 PPI 网络中筛选 hub 基因，只选择度值不小于 19 的 DEGs 作为 hub 基因。

4.中心基因与患者预后的关系。

使用在线生存分析工具 Kaplan Meier plotter，根据基因的中位表达值，将

患者样本分为高表达组和低表达组。

研究结果

DEGs 的识别。

分别在 GSE21933、GSE33532、GSE44077 和 GSE74706 中得到 1437、2127、963 和 4147 个 DEGs。如图 1A 所示，其中 GSE21933、GSE33532、GSE44077、GSE74706 表现上调分别有 676、865、274、1797 个，表现下调分别有 761、1262、689、2350 个。为了获得最可靠的 DEGs，分离所有四个数据集中的 DEGs，最终得到 195 个 DEGs，包括 57 个上调的 DEGs(图 1B)和 138 个下调的 DEGs(图 1C)。

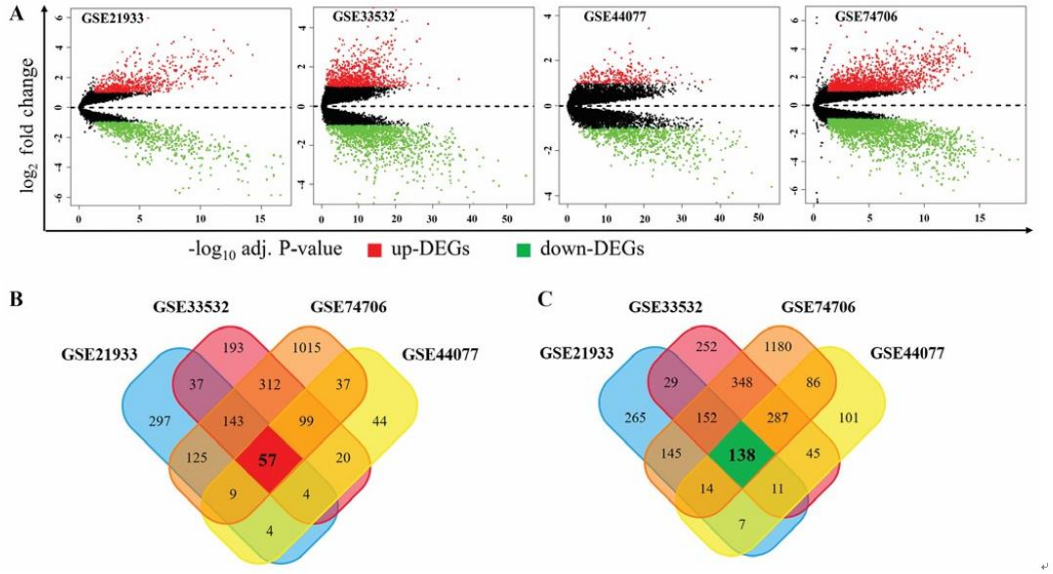

图 1 所示。4 个 NSCLC 数据集中表达基因的表达水平及分布。(A)每个微阵列数据的火山图。y 轴代表 log2fold-change (NSCLC 与正常样本)。(B)蛋白编码基因上调和(C)蛋白编码基因下调的文氏图显示。

GO 功能富集分析

为了找出与这 195 种 DEGS 相关的潜在生物学功能，使用在线软件 DAVID 来识别富集的 GO 类别。经过 GO 功能富集分析，如表 I 所示，上调的 DEGs 在 18 个生物过程(BP)项中显著富集，包括有丝分裂、细胞周期、蛋白- dna 复合物组装、微管基础过程和细胞增殖。下调的 DEGs 在 16 个 BP 项中也显著富集，如细胞对激素刺激的反应、血管生成、药物反应和细胞粘附。这些项都与肿瘤的发生发展密切相关。

表 I.富集差异表达基因上调和下调的 GO 类别

A, Upregulated

Term	Gene function	Count	P-value
GO:0007067	Mitosis	14	$6.11x10^{-13}$
GO:0000280	Nuclear division	14	$6.11x10^{-13}$
GO:0000087	M phase of mitotic cell cycle	14	$7.70x10^{-13}$
GO:0048285	Organelle fission	14	$1.02x10^{-12}$
GO:0000279	M phase	15	$6.02x10^{-12}$
GO:0022403	Cell cycle phase	16	$9.16x10^{-12}$
GO:0000278	Mitotic cell cycle	15	$2.91x10^{-11}$
GO:0022402	Cell cycle process	16	$7.21x10^{-10}$
GO:0007049	Cell cycle	16	$5.28x10^{-08}$
GO:0051301	Cell division	11	$8.09x10^{-08}$
GO:0065004	Protein-DNA complex assembly	5	$3.22x10^{-04}$
GO:0007018	Microtubule-based movement	5	$7.32x10^{-04}$
GO:0007017	Microtubule-based process	6	$2.14x10^{-03}$
GO:0034622	Cellular macromolecular complex assembly	6	$5.69x10^{-03}$
GO:0051726	Regulation of cell cycle	6	$6.72x10^{-03}$
GO:0034621	Cellular macromolecular complex subunit organization	6	$9.16x10^{-03}$
GO:0008283	Cell proliferation	6	$2.02x10^{-02}$

B, Downregulated

Term	Gene function	Count	P-value
GO:0032870	Cellular response to hormone stimulus	6	$1.73x10^{-05}$
GO:0001525	Angiogenesis	10	$3.41x10^{-05}$
GO:0050900	Leukocyte migration	7	$2.53x10^{-04}$
GO:0030336	Negative regulation of cell migration	6	$6.18x10^{-04}$
GO:0042493	Response to drug	9	$1.62x10^{-03}$
GO:0007165	Signal transduction	18	$3.98x10^{-03}$
GO:0007155	Cell adhesion	10	$5.99x10^{-03}$
GO:0016337	Single organismal cell-cell adhesion	5	$6.12x10^{-03}$
GO:0002576	Platelet degranulation	5	$6.56x10^{-03}$
GO:0006898	Receptor-mediated endocytosis	6	$1.12x10^{-02}$
GO:0043065	Positive regulation of apoptotic process	7	$2.15x10^{-02}$
GO:0018108	Peptidyl-tyrosine phosphorylation	5	$2.48x10^{-02}$
GO:0045893	Positive regulation of transcription, DNA-templated	9	$3.28x10^{-02}$
GO:0045087	Innate immune response	8	$3.61x10^{-02}$
GO:0006810	Transport	7	$4.02x10^{-02}$

Count, the number of enriched genes in each term.

构建 PPI 网络筛选 hub 基因

为了进一步研究这 195 个 DEGs 之间的相互作用，使用 STRING 数据库和 Cytoscape 构建相应 PPI 网络。如图 2A 所示，去除游离蛋白对的影响后，得到了一个含有 38 个上调 DEGs、66 个下调 DEGs 和 424 条边的网络。此外，在 PPI 网络中，与其他许多基因有强烈相互作用的基因被称为“hub 基因”。由于它们在 PPI 网络中的关键位置，hub 基因是疾病状态的潜在驱动因素。为了鉴定 NSCLC 的关键致瘤基因，使用 cyhubba 插件筛选所有 DEGs 中的 hub 基因。如图 2B 和表 II 所示，获得了 25 个 hub 基因(红色或橙色的节点)，这些节点的得分特别高

(≥19)，而且所有这 25 个中心基因都是上调的 DEGs。基因 G2/有丝分裂特异性的 cyclinB1 (CCNB1)、CyclinA2 (CCNA2)、55 kDa 的中心体蛋白(CEP55)、淋巴激酶激活的杀伤 t 细胞源性蛋白激酶(PBK)和透明质酸介导的运动受体(HMMR)被选为连接度最高的前 5 位枢纽基因。CCNB1，也被称为 cyclin B1，被观察到表现出最高程度的连通性，如表二所示。

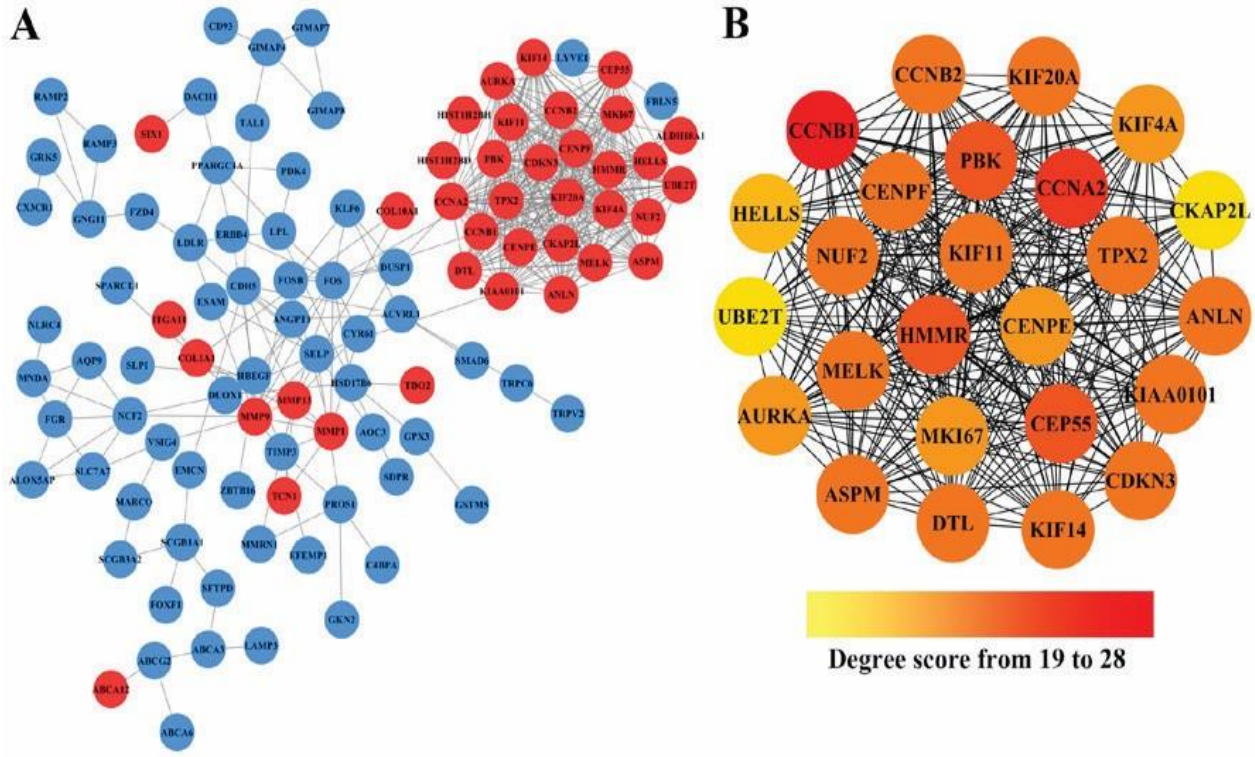

图 2。构建的 DEGs 的 PPI 网络。(A)利用 195 个重叠 DEGs 作为种子基因，构建基于 STRING 数据库的 PPI 子网络。红色节点表示上调的 DEG，而蓝色节点表示下调的 DEG。节点之间的边表示它们之间的相互作用。(B)筛选的 hub 基因子网。颜色随度值增加从橙色到红色变化，所有节点代表上调的基因。

Hub 基因生存分析

为了深入了解 hub 基因与 NSCLC 之间的关系，使用 Kaplan Meier 在线工具进行生存分析，该工具包含大量肺癌微阵列数据集。如图 3 所示，TOP5 hub 基因在 NSCLC 患者中高表达组和低表达组的生存时间有显著性差异(P<0.05)，这 5 个 hub 基因的低表达组在 NSCLC 中具有良好的预后效果。此外，其余 20 个 hub 基因也呈现出相同的趋势(数据未显示)。这说明这 25 个 hub 基因的表达水平与 NSCLC 的临床预后密切相关，可能在 NSCLC 的进展中发挥重要作用。

表Ⅱ.PPI 网络中的连通度统计结果。PPI，protein-protein_interaction.（蛋白质的相互作用）

Gene	Degree	Gene	Degree
CCNB1	28	KIF11	24
CCNA2	27	KIF20A	24
CEP55	25	CCNB2	24
PBK	25	CENPF	24
HMMR	25	TPX2	24
DTL	24	MKI67	23
KIAA0101	24	KIF4A	23
MELK	24	CENPE	23
ANLN	24	AURKA	23
KIF14	24	HELLS	21
CDKN3	24	CKAP2L	19
ASPM	24	UBE2T	19
NUF2	24		

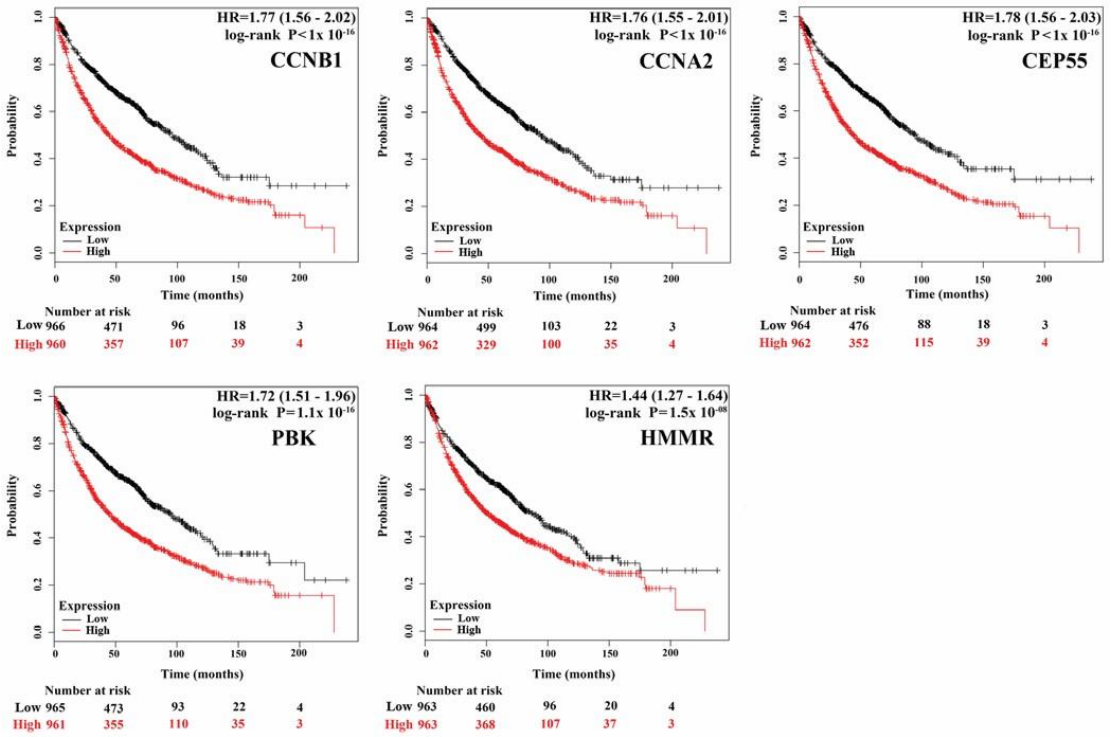

图 3.TOP5hub 基因表达的预后价值。生存数据采用 Kaplan - Meier 绘图仪分析，根据所有 NSCLC 患者(n=1,926)的数据绘制生存曲线，红线代表中位以上表达的患者，黑线代表中位以下表达的患者。HP 表示危险比。

结果讨论

与其他癌症一样，肺癌的发生、发展和转移被认为是一个非常复杂的过程，因为它涉及多个基因和细胞通路的畸变。某些 DEGs 与其他 DEGs 存在多种相互作用，可能是促进肿瘤发生的核心功能基因。发现这些异常基因并了解其在非小细胞肺癌分子机制中的作用对于提高非小细胞肺癌的诊治水平至关重要。

本研究利用 GSE21933、GSE33532、GSE44077、GSE74706 的基因表达谱筛选 NSCLC 与正常样本共表达的 DEGs，共 195 个 DEGs，其中上调的 57 个，下调的 138

个。

观察到上调的 DEGs 在与有丝分裂、细胞周期和细胞增殖相关的 BP 方面得到了丰富，这与先前的认识一致，即细胞周期功能缺陷和细胞增殖调节因子是导致肿瘤发生和发展的主要原因。下调的 DEGs 在血管生成、细胞黏附等方面有丰富的表达。血管生成是肿瘤生长和转移的重要生物学过程，其在肿瘤组织中的阻断作用被认为是抑制肿瘤生长的一种有效策略。有研究表明细胞粘附分子表达的改变可能影响细胞的粘附库、细胞的信号转导状态、细胞与环境的相互作用，并在肿瘤的进展、侵袭和转移中起重要作用。我们的结果显示这些上调和下调的 DEGs 参与了这些 BP，可能在 NSCLC 的进展中发挥重要作用。

基于 PPI 网络，本研究识别了一系列 hub 基因，前 5 位 hub 基因分别为 CCNB1、CCNA2、CEP55、PBK 和 HMMR。

CCNB1 是参与有丝分裂的关键调控蛋白，是细胞周期中 G2/M 转化的关键蛋白。有研究称其能抑制人肿瘤细胞的增殖，诱导细胞凋亡。CCNB1 的表达在非小细胞肺癌中占显著比例，但并非所有类型的非小细胞肺癌中都有表达。研究发现，NSCLC 的不同亚型不仅在生物学上存在差异，而且在 CCNB1 表达上也存在差异。在所有的组织学亚型中，CCNB1 的过表达在鳞状细胞癌(SCC)亚型中更为常见。据研究，这种过表达也会影响患者的生存时间，可能是 SCC 亚型 NSCLC 患者的不良预后标志物。

第二个 hub 基因 CCNA2，编码细胞周期蛋白既控制细胞周期的 G1/S，也控制细胞周期的 G2/M 过渡。其蛋白表达在多种肿瘤中均有升高，是预测生存或早期复发的预后标志物。

第三个 hub 基因 CEP55(也被称为 c10orf3 和 FLJ10540)是细胞分裂的关键组成部分，它通过与 PI3K 催化亚基的相互作用与 PI3K/AKT 通路的激活有关。CEP55 过表达与包括肺癌、乳腺癌、肝癌和结肠癌在内的多种癌症类型的癌变高度相关，并已被确定为几种癌症预后基因标记的成员之一。有研究称，CEP55 的异位表达可在裸鼠体内诱导肿瘤发生，而其在胃癌细胞中的下调通过诱导 G2/M 期阻滞来抑制细胞增殖，显示其具有抗肿瘤的潜力。

PBK 也被称为 TOPK，被发现在肿瘤的形成和发展中起着重要的作用。其高表达水平在大多数肺癌组织和细胞系中均可检测到，但在正常组织中未检测到。据

研究，PBK 过表达可促进 PI3K/ akt 依赖的细胞迁移。PBK 表达与 Ki67、p53 表达呈正相关，可作为 NSCLC 的独立预后因子。

NSCLC 中另一个高表达的 hub 基因是 HMMR，这是一种多功能致癌基因。它的高水平表达也在乳腺癌中被发现，通常导致疾病恶化。最近的研究已经解释了 HMMR 可以通过与 CD44(一种著名的癌症干细胞标记物)和透明质酸(大多数恶性肿瘤微环境的关键成分)形成复合物来调节细胞增殖、存活和迁移。在本研究中，HMMR 与前 5 位的枢纽基因均表现出很强的相互作用，这表明了它们在 NSCLC 中的联合作用，尽管还需要进一步的研究来阐明 HMMR 与 NSCLC 之间潜在的生物学联系。

CCNB1、CCNA2、CEP55、PBK 和 HMMR 主要通过影响细胞周期和有丝分裂参与恶性肿瘤的发病过程。本研究的 Kaplan Meier 生存分析显示，这 5 个 hub 基因以及其余 20 个 hub 基因的 mRNA 表达水平与肺癌的临床预后显著相关。虽然这些可能暗示了它们在 NSCLC 进展中的作用，从而使它们成为 NSCLC 诊断和治疗的潜在靶点，但对每个 hub 基因及其亚型的临床意义还需要进一步的实验验证。

文章评价

该文章对 NSCLC 与正常肺组织之间的关键基因差异进行研究，得到了五个 hub 基因并对这些基因的功能进行了讨论，并认为这些基因可能有助于全面了解 NSCLC 的分子机制，可作为 NSCLC 诊断和预后的生物标志物以及治疗 NSCLC 的分子靶点。

文章的研究方法中规中矩，并没有太多创新内容，而且研究结果只是对他人文章结果的整合及验证，整体意义不是很大。该文章没有对不同的 NSCLC 亚型进行细致研究讨论，所以得出的研究结果是不完全准确的，而且在使用 FunRich 对差异基因进行筛选时只选取了一部分基因，因此，许多可能重要的基因可能没有被讨论到。

文章对研究方法和研究结果的说明通俗易懂，但在阐述的内容上存在大量的冗余内容，可能是因为研究内容比较简单，这样简单的研究更适合将方法和结果放在一起说明。

Identification of key differentially expressed genes associated with non-small cell lung cancer by bioinformatics analyses

YUBO XIAO[1,2], MIN FENG[3], HAIYING RAN[4], XIAO HAN[5] and XUEGANG LI[1]

[1]Chongqing Productivity Promotion Center for The Modernization of Chinese Traditional Medicine, School of Pharmaceutical Sciences, Southwest University, Chongqing 400715; [2]Department of Clinical Laboratory, Hunan University of Medicine, Huaihua, Hunan 418000; [3]Chongqing Key Laboratory of Catalysis and Functional Organic Molecules, College of Environment and Resource, Chongqing Technology and Business University, Chongqing 400067; [4]Biomedical Analysis Center, Third Military Medical University; [5]Department of Respiratory Medicine, Southwest Hospital, Third Military Medical University, Chongqing 400715, P.R. China

Received November 9, 2017; Accepted February 27, 2018

DOI: 10.3892/mmr.2018.8726

Abstract. Increasing evidence has indicated that the abnormal expressions of certain genes serve important roles in tumorigenesis, progression and metastasis. The aim of the present study was to explore the key differentially expressed genes (DEGs) between non-small cell lung cancer (NSCLC) and matched normal lung tissues by analyzing 4 different mRNA microarray datasets downloaded from the Gene Expression Omnibus (GEO) database. In improving the reliability of the bioinformatics analysis, the DEGs in each dataset that met the cut-off criteria (adjust P-value <0.05 and $|\log_2$fold-change (FC)|>1) were intersected with each other, from which 195 were identified (consisting of 57 upregulated and 138 downregulated DEGs). The GO analysis results revealed that the upregulated DEGs were significantly enriched in various biological processes (BP), including cell cycle, mitosis and cell proliferation while the downregulated DEGs were significantly enriched in angiogenesis and response to drug and cell adhesion. The hub genes, including CCNB1, CCNA2, CEP55, PBK and HMMR, were identified based on the protein-protein interaction (PPI) network. The Kaplan-Meier survival analysis indicated that the high expression level of each of these hub genes correlates with poorer overall survival in all patients with NSCLC, which indicates that they may serve important roles in the progression of NSCLC. In conclusion, the DEGs and hub genes identified in the present study may contribute to the comprehensive understanding of the molecular mechanisms of NSCLC and may be used as diagnostic and prognostic biomarkers as well as molecular targets for the treatment of NSCLC.

Correspondence to: Professor Xuegang Li, Chongqing Productivity Promotion Center for The Modernization of Chinese Traditional Medicine, School of Pharmaceutical Sciences, Southwest University, 2 Tiansheng Road, Beibei, Chongqing 400715, P.R. China
E-mail: xuegangli@swu.edu.cn

Key words: non-small cell lung cancer, microarray, differentially expressed genes, bioinformatics analysis, hub genes

Introduction

Lung cancer, which accounts for more than one-tenth of all cancer cases worldwide, is one of the most common types of malignancies and perhaps, the first cause of cancer-related deaths worldwide (1). In China, lung cancer is currently the first most common cancer with more than 733,300 new cases and 610,200 deaths in 2015 (2). Based on histological sub-classification, there are non-small cell lung cancer (NSCLC) and SCLC with approximately 85% of lung cancer patients diagnosed with NSCLC (3).

To date, surgical resection remains the only treatment option for NSCLC (4,5). However, with no specific symptoms associated with the early stage of NCLSC and the lack of accurate and convenient diagnostic methods for its early detection, most patients are diagnosed at the advanced stage of this disease when surgical resection is no longer feasible (6).

With the recent advances in the molecular biology, molecule targeted therapy for NSCLC has equally made significant progress. For example, approximately 10-20% of North American and Western European patients and 30-50% of Asian patients with NSCLC were shown to have mutation of EGFR gene (7). But, with the use of EGFR tyrosine kinase inhibitors (TKIs), such as erlotinib and gefitinib, these patients were found to response well (with a response rate of 10%, and an estimated 2-month prolongation in median survival over placebo) (8,9). It was also found that 3-7% of NSCLC patients had an activated ALK gene, but treatment with crizotinib (an ALK inhibitor) elevated their response rate from 10 to 55% (compared to conventional chemotherapy), and increased their 6-month progression-free survival rate to 72% (8). Nevertheless, these situations can only represent a cohort of NSCLC patients and due to drug resistance these targeted therapy agents are often transient.

Moreover, even with the remarkable successes and withdraw that have been obtained in the targeted therapy of

NSCLC, the precise molecular mechanisms of this disease are still far from being fully understood. Thus, the need to find more potential targets for more effective therapeutic strategies is very paramount.

In order to better understand the DEGs influence on molecular pathogenesis of NSCLC, in this study, we downloaded 4 NSCLC related mRNA datasets from Gene Expression Omnibus (GEO) and identified the DEGs between the NSCLC patients and healthy control samples. Subsequently, we performed Gene Ontology (GO) and protein-protein interaction (PPI) networks analysis so as to study and identify potential key DEGs for the diagnosis and treatment of NSCLC. Further studies on these identified key DEGs can aid a comprehensive understanding of the molecular development of NSCLC and be explored as potential biomarkers for its diagnosis, prognosis, and drug targets.

Materials and methods

Affymetrix microarray data. Series matrix files of the microarray data: GSE21933, GSE33532, GSE44077 and GSE74706 were downloaded from the public GEO database (http://www.ncbi.nlm.nih.gov/geo/), and the probe names based on platforms were annotated to gene symbols according to their corresponding new version of annotation files, such as GPL6254, GPL570, GPL6244 and GPL13497. In each GSE file, we only choose the NSCLC samples and their matched normal samples, and from which we obtained 21 tumor samples and 21 normal samples for GSE21933 (10), 80 tumor samples and 20 normal samples for GSE33532 (11), 65 tumor samples and 65 normal samples for GSE44077 (12) and 18 tumor samples and 18 normal samples for GSE74706 (13) to analyze downstream.

Data pre-processing and differential expression analysis. The original array data of each series were analyzed separately using R software (version 3.4.0) and R packages. In brief: Firstly, background correction and quantile normalization was performed on the raw data using the robust multi-array average (RMA) algorithm in R affy package (14). The differentially expressed proteins (DEGs) between the tumor and normal group samples were then analyzed by the paired t-test based on the limma package in R. Multiple testing was corrected by the Benjamini-Hochberg method to obtain the adjusted P-value (15). Finally, the Genes with adjusted $P<0.05$ and $|\log_2$fold-change (FC)$|>1$ were considered to be significant. To further enhance the reliability of the bioinformatics analysis, the overlapping DEGs co-existed in all 4 GSE files were identified using FunRich software version 2.1.2 (http://www.funrich.org) (16).

GO enrichment analysis of DEGs. GO analysis is widely used to identify characteristic biological processes (BP) for a large number of genes in microarray analysis (17). DAVID (Database for Annotation, Visualization and Integrated Discovery, https://david.ncifcrf.gov/), as a comprehensive set of functional annotation tools (18), was used to investigate enriched BP terms for all the overlapping DEGs. $P<0.05$ and Count ≥5 were considered statistically significant.

PPI network construction and hub gene analysis. STRING (Search Tool for the Retrieval of Interacting Genes, https://string-db.org/) database is a useful online tool dedicated to evaluate the PPI interaction information (19). STRING (version 10.5) was used to explore the interactive relationships of the overlapping DEGs, and only the experimentally validated interactions with a combined score >0.4 were selected as significant. Then, the PPI network was built by Cytoscape software (version 3.5.1) (20). The plug-in cytoHubba in Cytoscape was used to screen the hub genes from the PPI network and only DEGs with a degree score ≥19 was selected as hub genes.

Association of hub genes and patient prognosis. An online survival analysis tool Kaplan Meier plotter (http://kmplot.com/analysis/) (21), which includes both clinical data and gene expression data of lung, breast, gastric and ovarian cancers, was used to evaluate the prognostic significance of each hub gene in NSCLC. According to the median expression value of a gene, the patient samples were divided into high and low expression groups. In this study, the analysis was carried out under the default parameters, which is in brief, no subtypes restriction, 'univariate' for Cox regression and 'exclude biased arrays' for array quality control and in each survival plot, the log rank P-value and hazard ratio (HR, 95% confidence intervals) were calculated and displayed on the main plot.

Results

Identification of DEGs. In order to identify the DEGs, we downloaded 4 public RNA-seq datasets of NSCLC from the GEO database for our analysis: GSE21933, GSE33532, GSE44077 and GSE74706. We defined DEGs between tumor samples and matched normal samples according to the threshold criterion (logFC >1 or logFC <-1 and adjusted $P<0.05$), and we got 1437, 2127, 963 and 4147 DEGs in GSE21933, GSE33532, GSE44077 and GSE74706, respectively. As shown in Fig. 1A, among these DEGs, 676, 865, 274 and 1797 genes were upregulated while 761, 1262, 689, 2350 ones were downregulated in GSE21933, GSE33532, GSE44077 and GSE74706, respectively. In order to obtain the most reliable DEGs, we isolated the DEGs presented in all four datasets and finally got a total of 195 DEGs, consisting of 57 upregulated (Fig. 1B) and 138 downregulated DEGs (Fig. 1C).

GO term enrichment analysis. In an attempt to find the potential biological functions associated with these 195 DEGs, the online software DAVID was used to identify the overrepresented GO categories. After GO functional enrichment analysis, as shown in Table I, the upregulated DEGs were significantly enriched in 18 biological process (BP) terms, which includes mitosis, cell cycle, protein-DNA complex assembly, microtubule-based process and cell proliferation. The downregulated DEGs were also significantly enriched in 16 BP terms, such as cellular response to hormone stimulus, angiogenesis, response to drug and cell adhesion. All these terms are closely related to the tumorigenesis and development.

Hub genes screening from the PPI network. To further investigate the interaction between these 195 DEGs, the

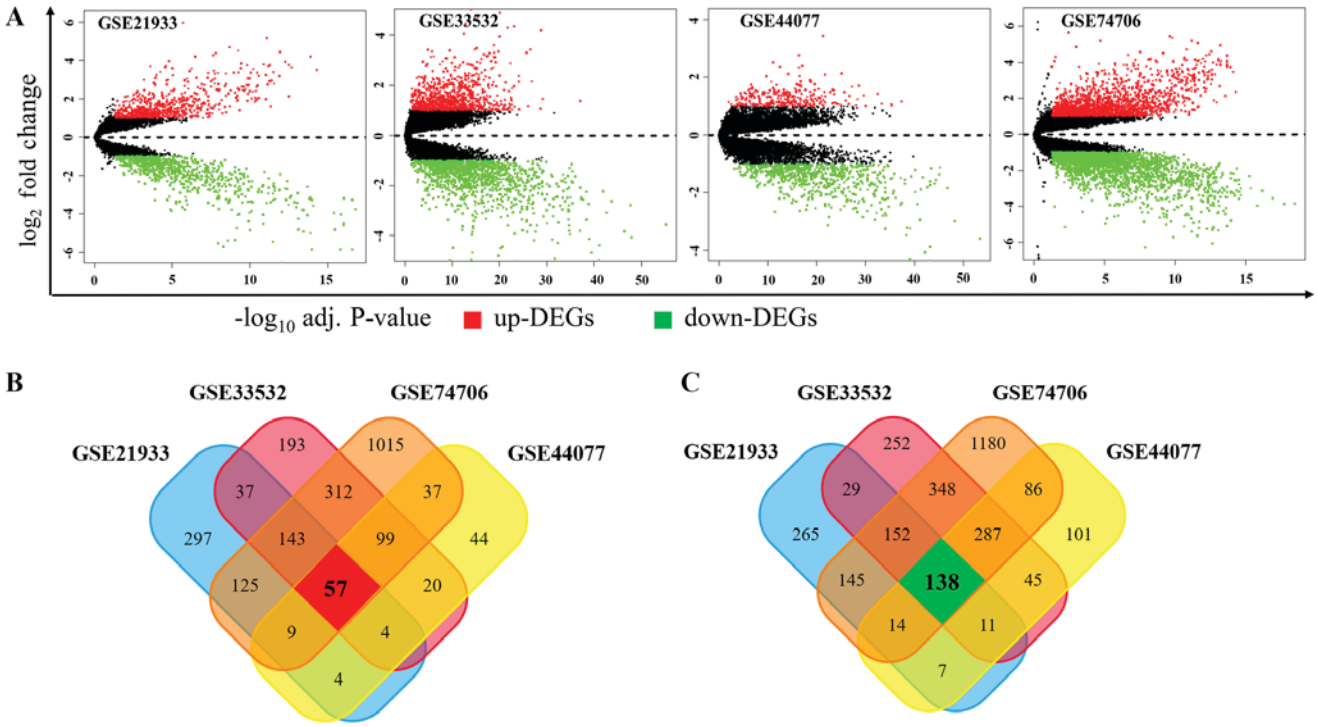

Figure 1. Expression levels of expressed genes and distributions in four NSCLC datasets. (A) Volcano plots of each microarray data. Y-axis represents the log_2fold-change (NSCLC vs. normal samples). Venn plots of (B) upregulated and (C) downregulated protein-coding genes were indicated. Adj. P-value, adjusted P-value; NSCLC, non-small cell lung cancer.

STRING database was used to analysis their PPI networks, and the resulted PPI networks were constructed by Cytoscape. As shown in Fig. 2A, after removing the effect of free protein pairs, we obtained a network with 38 upregulated DEGs, 66 downregulated DEGs and 424 edges.

In addition, in the PPI networks, genes that have strong interactions with many other genes are called 'hub genes'. Due to their key positions in the PPI network, the hub genes are potential drivers of the diseases status. In order to identify the key tumorigenic genes of NSCLC, the cytoHubba plugin for Cytoscape was used to screen the hub genes among all the DEGs. As shown in Fig. 2B and Table II, we obtained 25 hub genes (the nodes with red or orange colors) that exhibited especially high degree scores (≥19). To our surprise, all these 25 hub genes were upregulated DEGs. The genes G2/mitotic-specific cyclinB1 (CCNB1), CyclinA2 (CCNA2), Centrosomal protein of 55 kDa (CEP55), lymphokine-activated killer T-cell-originated protein kinase (PBK), and hyaluronan mediated motility receptor (HMMR) were selected as the top five hub genes with high connectivity degree. CCNB1, which is also known as cyclin B1, was observed to exhibit the highest degree of connectivity as shown in Table II.

Survival analysis of each hub genes using online tool Kaplan-Meier Plotter. To gain insight into the associations between the hub genes and NSCLC, we did a survival analysis through an online tool Kaplan Meier Plotter, which contains a large number of microarray datasets of lung cancer. As shown in Fig. 3, the survival time of the top five hub genes was significantly separated between the high expression groups and low expression groups in NSCLC patients ($P<0.05$), with the low expression groups of these 5 hub genes exhibiting a good prognostic effect in NSCLC. Moreover, the remaining 20 hub genes also showed the same trend (data not shown). These imply that the expression levels of the 25 hub genes are significantly associated with clinical prognosis of NSCLC and they may play important roles in the progression of NSCLC.

Discussion

The tumorigenesis, progression and metastasis of lung cancer, like in any other cancer is considered as a very complex process as it involves aberrations of multiple genes and cellular pathways (22). Some DEGs that have multi-interactions with other DEGs are probably core functional genes in promoting the carcinogenesis (23,24). To improve diagnosis and treatment of NSCLC, it is vitally important to find these abnormal genes and understand their roles in the molecular mechanism of NSCLC. With the development of microarray and high throughput technologies, we have been able to detect the cancer etiology by examining aberrations in whole-genome level as these technologies have been widely used to predict the potential therapeutic targets for cancers (25).

In the present study, using the gene expression profiles of GSE21933, GSE33532, GSE44077 and GSE74706 to screen the co-expressed DEGs between NSCLC and normal samples, a total of 195 DEGs, consisting of 57 upregulated and 138 downregulated DEGs were obtained. The observation that, upregulated DEGs were enriched in BP terms related to mitosis, cell cycle and cell proliferation is consistent with previous knowledge that the functional deficits of cell cycle and cell proliferation regulators are the main causes of tumorigenesis and progression (26). The downregulated DEGs were

Table I. Enriched Gene Ontology terms for upregulated and downregulated differentially expressed genes.

A, Upregulated

Term	Gene function	Count	P-value
GO:0007067	Mitosis	14	$6.11x10^{-13}$
GO:0000280	Nuclear division	14	$6.11x10^{-13}$
GO:0000087	M phase of mitotic cell cycle	14	$7.70x10^{-13}$
GO:0048285	Organelle fission	14	$1.02x10^{-12}$
GO:0000279	M phase	15	$6.02x10^{-12}$
GO:0022403	Cell cycle phase	16	$9.16x10^{-12}$
GO:0000278	Mitotic cell cycle	15	$2.91x10^{-11}$
GO:0022402	Cell cycle process	16	$7.21x10^{-10}$
GO:0007049	Cell cycle	16	$5.28x10^{-08}$
GO:0051301	Cell division	11	$8.09x10^{-08}$
GO:0065004	Protein-DNA complex assembly	5	$3.22x10^{-04}$
GO:0007018	Microtubule-based movement	5	$7.32x10^{-04}$
GO:0007017	Microtubule-based process	6	$2.14x10^{-03}$
GO:0034622	Cellular macromolecular complex assembly	6	$5.69x10^{-03}$
GO:0051726	Regulation of cell cycle	6	$6.72x10^{-03}$
GO:0034621	Cellular macromolecular complex subunit organization	6	$9.16x10^{-03}$
GO:0008283	Cell proliferation	6	$2.02x10^{-02}$

B, Downregulated

Term	Gene function	Count	P-value
GO:0032870	Cellular response to hormone stimulus	6	$1.73x10^{-05}$
GO:0001525	Angiogenesis	10	$3.41x10^{-05}$
GO:0050900	Leukocyte migration	7	$2.53x10^{-04}$
GO:0030336	Negative regulation of cell migration	6	$6.18x10^{-04}$
GO:0042493	Response to drug	9	$1.62x10^{-03}$
GO:0007165	Signal transduction	18	$3.98x10^{-03}$
GO:0007155	Cell adhesion	10	$5.99x10^{-03}$
GO:0016337	Single organismal cell-cell adhesion	5	$6.12x10^{-03}$
GO:0002576	Platelet degranulation	5	$6.56x10^{-03}$
GO:0006898	Receptor-mediated endocytosis	6	$1.12x10^{-02}$
GO:0043065	Positive regulation of apoptotic process	7	$2.15x10^{-02}$
GO:0018108	Peptidyl-tyrosine phosphorylation	5	$2.48x10^{-02}$
GO:0045893	Positive regulation of transcription, DNA-templated	9	$3.28x10^{-02}$
GO:0045087	Innate immune response	8	$3.61x10^{-02}$
GO:0006810	Transport	7	$4.02x10^{-02}$

Count, the number of enriched genes in each term.

enriched in BP terms such as angiogenesis and cell adhesion. Angiogenesis is an essential biological process for tumor growth and metastasis in that, its blockage in tumor tissue has been recognized as a charming strategy to inhibit tumor growth (27). According to Cavallaro and Christofori (28), changes in the expression of cell adhesion molecules could affect the adhesive repertoire of a cell, signal transduction status of cells, the interactions between cells and their environment, and play a crucial role in tumor progression, invasion and metastasis. Our results suggested that these up and downregulated DEGs are involved in these BP and may play important role in the progression of NSCLC.

Based on the PPI network, we identified a series of hub genes and among the list of the top five hub genes were CCNB1, CCNA2, CEP55, PBK and HMMR, respectively. CCNB1, as a key regulatory protein involved in mitosis, is essential for G2/M transition during the cell cycle (29). Yuan *et al* (30). reported that its depletion inhibited proliferation and induced apoptosis in human tumor cells. The work of Soria *et al* (31). established the expression of CCNB1 in a significant fraction

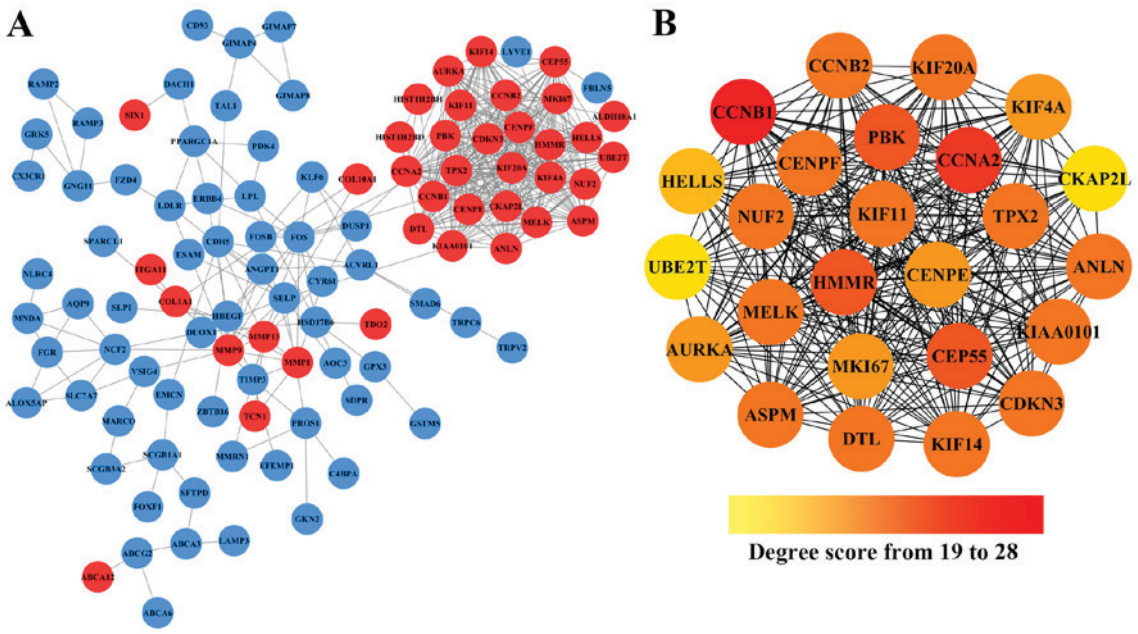

Figure 2. The constructed PPI network of DEGs. (A) A total of 195 overlapping DEGs were used as seed genes to build a PPI subnetwork based on the STRING database. Red nodes represent the upregulated DEGs, while bule nodes represent downregulated DEGs. Edges between the nodes represent their interation. (B) The screened subnetwork of hub genes. The color changes with the increase of degree score from orange to red, all nodes represent the upregulated genes.

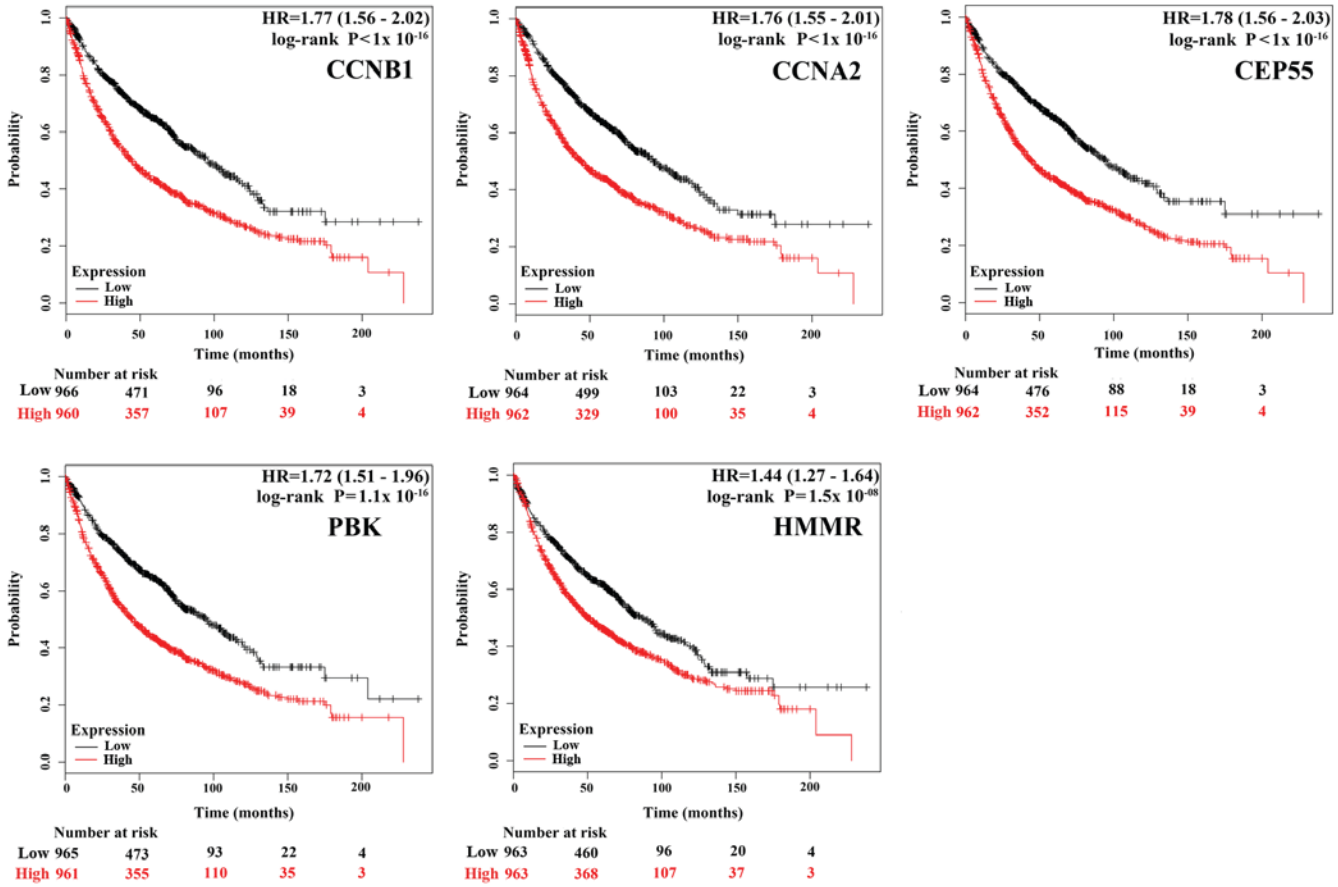

Figure 3. The prognostic value of the expression of the top five hub genes. The survival data was analyzed using Kaplan-Meier Plotter, and the survival curves were plotted based on data of all NSCLC patients (n=1,926), the red line represents patients with expression above the median, while black line represents expression below the median. HP represents hazard ratio.

but not all types of NSCLC. They showed that, different subtypes of NSCLC are not only biologically different but also showed variation in their CCNB1 expressions. Of all the histological subtypes examined, overexpression of CCNB1

Table II. Statistical results of connectivity degrees of the hub genes in the PPI network.

Gene	Degree	Gene	Degree
CCNB1	28	KIF11	24
CCNA2	27	KIF20A	24
CEP55	25	CCNB2	24
PBK	25	CENPF	24
HMMR	25	TPX2	24
DTL	24	MKI67	23
KIAA0101	24	KIF4A	23
MELK	24	CENPE	23
ANLN	24	AURKA	23
KIF14	24	HELLS	21
CDKN3	24	CKAP2L	19
ASPM	24	UBE2T	19
NUF2	24		

PPI, protein-protein interaction.

was more frequently observed in the squamous cell carcinoma (SCC) subtypes. This overexpression was also reported to affects patient survival time and might be an adverse prognostic marker for SCC-subtype NSCLC patients (31). While our work establishes CCNB1 as a top hub gene in NSCLC in agreement with these reports, it did not specify the subtypes, which could be a possible limitation to our analysis. The second hub gene CCNA2, encoding cyclins controls both the G1/S and the G2/M transition of the cell cycle (32). Its protein expression was found to be elevated in variety of tumors, including breast, liver, prostate and lung cancers, and appears to be a prognostic marker in prediction of survival or early relapse (33). The third hub gene CEP55 (also known as c10orf3 and FLJ10540), a pivotal component of cytokinesis, is associated with the PI3K/AKT pathway activation via an interaction with the catalytic subunit of PI3K (34). CEP55 overexpression is highly correlated with carcinogenesis across multiple cancer types including lung, breast, liver, and colon cancers and has been identified as a member of several prognostic 'gene signatures' for cancer (35). Tao *et al* (36). has reported that ectopic expression of CEP55 induces tumorigenesis in nude mice, while its knockdown in gastric cancer cells suppressed cellular proliferation by inducing G2/M phase arrest, indicating its potential as an anti-tumor target. PBK also known as TOPK, was found to play an important role in tumor formation and progression. Its high expression level was detectable in the majority of lung cancer tissues and cell lines, but not in normal tissues (37-40). Overexpression of PBK was reported to promote a PI3K/AKT-dependent cell migration (5). Lei *et al* (40). suggested that PBK expression was positively correlated with Ki67 and p53 expression, and could be used as an independent prognostic factor in NSCLC. Another hub gene which we found to be hyper-expressed in NSCLC is HMMR, a multifunctional oncogene. Its high expression level, which was also found in breast cancer, is associated with poor disease outcome (41). Recent studies have deciphered that HMMR could regulate cell proliferation, survival, and migration via forming a complex with CD44 (a well-known cancer stem cell marker) and hyaluronan (a key component of the microenvironment of most malignant tumors) (42). In this study, HMMR showed a strong interaction with each of the top five hub genes, which indicates their joint functions in NSCLC, although further investigations are still needed to clarify underlying biological associations between HMMR and NSCLC. All these reports indicate that CCNB1, CCNA2, CEP55, PBK, and HMMR are involved in the pathogenesis of malignant tumors mainly by affecting cell cycle and mitosis. This is consistent with our findings. Furthermore, the Kaplan Meier Plotter survival analysis from our study demonstrated that mRNA expression levels of these five hub genes, as well as the remaining 20 hub genes, were significantly associated with clinical prognosis of lung cancer. While these could imply their roles in the progression of NSCLC thereby making them a potential target for its diagnosis and treatment, further experimental verification of the clinical significance of each hub gene and its subtypes is still necessary.

In summary, using the bioinformatics, we have provided information on the DEGs and hub genes that may be involved in NSCLC. This analysis is, however, limited in that it does not the capture the expression of these genes in the different NSCLC subtypes. Another limitation of this work is that, in reducing the number of false positive DEGs, we obtained co-expressed DEGs among 4 different NSCLC associated microarray datasets. By this, many likely important genes might have been lost. While our work showed the benefit and usefulness of microarray analysis in bringing out the DEGs and hub genes that can be potential diagnostic and treatment targets of NSCLC, the need for improved analysis and prospective clinical studies is still imperative. Albeit, this study can contribute to the overall understanding of the underlying molecular mechanisms of NSCLC and serve as guide to subsequent experimental studies.

Acknowledgements

Not applicable.

Funding

This study was financially supported by Fundamental Research Funds for the Central Universities (grant no. XDJK2017D143), Fundamental Research Program of Hunan University of Medicine (grant no. 2014KY03) and Science and Technology Plan Projects of Huaihua (2014-17). Funding was also received from Chongqing Science and Technology Commission (grant no. cstc2015jcyjAoo61), Chongqing Educational Ministry (grant no. Kj1500637) and The National Key Research and Development Program of China for Traditional Chinese Medicine Modernization (grant nos. 2017YFC1702600 and 2017YFC1702605).

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Authors' contributions

YBX and XGL conceived and designed the study. YBX, MF, HYR and XH performed the data analysis. YBX and HYR wrote the paper. YBX, XGL, MF, HYR and XH reviewed and edited the manuscript. All authors read and approved the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Semenova EA, Nagel R and Berns A: Origins, genetic landscape, and emerging therapies of small cell lung cancer. Genes Dev 29: 1447-1462, 2015.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ and He J: Cancer statistics in China, 2015. CA Cancer J Clin 66: 115-132, 2016.
3. Herbst RS, Heymach JV and Lippman SM: Lung cancer. N Engl J Med 359: 1367-1380, 2008.
4. Danesi R, Pasqualetti G, Giovannetti E, Crea F, Altavilla G, Del Tacca M and Rosell R: Pharmacogenomics in non-small-cell lung cancer chemotherapy. Adv Drug Deliv Rev 61: 408-417, 2009.
5. Shih MC, Chen JY, Wu YC, Jan YH, Yang BM, Lu PJ, Cheng HC, Huang MS, Yang CJ, Hsiao M and Lai JM: TOPK/PBK promotes cell migration via modulation of the PI3K/PTEN/AKT pathway and is associated with poor prognosis in lung cancer. Oncogene 31: 2389-2400, 2012.
6. Keith RL and Miller YE: Lung cancer chemoprevention: Current status and future prospects. Nat Rev Clin Oncol 10: 334-343, 2013.
7. Oxnard GR, Lo PC, Nishino M, Dahlberg SE, Lindeman NI, Butaney M, Jackman DM, Johnson BE and Jänne PA: Natural history and molecular characteristics of lung cancers harboring EGFR exon 20 insertions. J Thorac Oncol 8: 179-184, 2013.
8. Gerber DE and Minna JD: ALK inhibition for non-small cell lung cancer: From discovery to therapy in record time. Cancer cell 18: 548-551, 2010.
9. Janku F, Stewart DJ and Kurzrock R: Targeted therapy in non-small-cell lung cancer-is it becoming a reality? Nat Rev Clin Oncol 7: 401-414, 2010.
10. Lo FY, Chang JW, Chang IS, Chen YJ, Hsu HS, Huang SF, Tsai FY, Jiang SS, Kanteti R, Nandi S, *et al*: The database of chromosome imbalance regions and genes resided in lung cancer from Asian and Caucasian identified by array-comparative genomic hybridization. BMC Cancer 12: 235, 2012.
11. Meister M, Belousov A, Xu EC, Schnabel PA, Warth A, Hoffmann H, Dienemann H, Riedlinger J, Bodenmueller H, Zolg W, *et al*: Intra-tumor heterogeneity of gene expression profiles in early stage. Non-small cell lung cancer, 2014.
12. Kadara H, Fujimoto J, Yoo SY, Maki Y, Gower AC, Kabbout M, Garcia MM, Chow CW, Chu Z, Mendoza G, *et al*: Transcriptomic architecture of the adjacent airway field cancerization in non-small cell lung cancer. J Natl Cancer Inst 106: dju004, 2014.
13. Marwitz S, Depner S, Dvornikov D, Merkle R, Szczygieł M, Müller-Decker K, Lucarelli P, Wäsch M, Mairbäurl H, Rabe KF, *et al*: Downregulation of the TGFβ Pseudoreceptor BAMBI in non-small cell lung cancer enhances TGFβ signaling and invasion. Cancer Res 76: 3785-3801, 2016.
14. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249-264, 2003.
15. Benjamini Y and Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Royal Statis Soc 57: 289-300, 1995.
16. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, Mouradov D, Sieber OM, Simpson RJ, Salim A, *et al*: FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics 15: 2597-2601, 2015.
17. Gene Ontology Consortium: The gene ontology (GO) project in 2006. Nucleic Acids Res 34 (Database Issue): D322-D326, 2006.
18. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC and Lempicki RA: DAVID: Database for annotation, visualization, and integrated discovery. Genome Biol 4: P3, 2003.
19. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, *et al*: The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 45 (D1): D362-D368, 2017.
20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504, 2003.
21. Szász AM, Lánczky A, Nagy Á, Förster S, Hark K, Green JE, Boussioutas A, Busuttil R, Szabó A and Győrffy B: Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. Oncotarget 7: 49322-49333, 2016.
22. Vogelstein B and Kinzler KW: Cancer genes and the pathways they control. Nat Med 10: 789-799, 2004.
23. Saito M, Shiraishi K, Kunitoh H, Takenoshita S, Yokota J and Kohno T: Gene aberrations for precision medicine against lung adenocarcinoma. Cancer Sci 107: 713-720, 2016.
24. Cardarella S and Johnson BE: The impact of genomic changes on treatment of lung cancer. Am J Respir Crit Care Med 188: 770-775, 2013.
25. Liang B, Li C and Zhao J: Identification of key pathways and genes in colorectal cancer using bioinformatics analysis. Med Oncol 33: 111, 2016.
26. Kastan MB and Bartek J: Cell-cycle checkpoints and cancer. Nature 432: 316-323, 2004.
27. Feng X, Ofstad W and Hawkins D: Antiangiogenesis therapy: A new strategy for cancer treatment. US Pharm 35 (Oncology Suppl): S4-S9, 2010.
28. Cavallaro U and Christofori G: Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. Nat Rev Cancer 4: 118-132, 2004.
29. Petri ET, Errico A, Escobedo L, Hunt T and Basavappa R: The crystal structure of human cyclin B. Cell Cycle 6: 1342-1349, 2007.
30. Yuan J, Yan R, Krämer A, Eckerdt F, Roller M, Kaufmann M and Strebhardt K: Cyclin B1 depletion inhibits proliferation and induces apoptosis in human tumor cells. Oncogene 23: 5843-5852, 2004.
31. Soria JC, Jang SJ, Khuri FR, Hassan K, Liu D, Hong WK and Mao L: Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication. Cancer Res 60: 4000-4004, 2000.
32. Pagano M, Pepperkok R, Verde F, Ansorge W and Draetta G: Cyclin A is required at two points in the human cell cycle. EMBO J 11: 961-971, 1992.
33. Yam CH, Fung TK and Poon RY: Cyclin A in cell cycle control and cancer. Cell Mol Life Sci 59: 1317-1326, 2002.
34. Chen CH, Lu PJ, Chen YC, Fu SL, Wu KJ, Tsou AP, Lee YC, Lin TC, Hsu SL, Lin WJ, *et al*: FLJ10540-elicited cell transformation is through the activation of PI3-kinase/AKT pathway. Oncogene 26: 4272-4283, 2007.
35. Jeffery J, Sinha D, Srihari S, Kalimutho M and Khanna KK: Beyond cytokinesis: The emerging roles of CEP55 in tumorigenesis. Oncogene 35: 683-690, 2016.
36. Tao J, Zhi X, Tian Y, Li Z, Zhu Y, Wang W, Xie K, Tang J, Zhang X, Wang L and Xu Z: CEP55 contributes to human gastric carcinoma by regulating cell proliferation. Tumour Biol 35: 4389-4399, 2014.
37. Zhu F, Zykova TA, Kang BS, Wang Z, Ebeling MC, Abe Y, Ma WY, Bode AM and Dong Z: Bidirectional signals transduced by TOPK-ERK interaction increase tumorigenesis of HCT116 colorectal cancer cells. Gastroenterology 133: 219-231, 2007.
38. Zykova TA, Zhu F, Lu C, Higgins L, Tatsumi Y, Abe Y, Bode AM and Dong Z: Lymphokine-activated killer T-cell-originated protein kinase phosphorylation of histone H2AX prevents arsenite-induced apoptosis in RPMI7951 melanoma cells. Clin Cancer Res 12: 6884-6893, 2006.

39. Zlobec I, Molinari F, Kovac M, Bihl MP, Altermatt HJ, Diebold J, Frick H, Germer M, Horcic M, Montani M, *et al*: Prognostic and predictive value of TOPK stratified by KRAS and BRAF gene alterations in sporadic, hereditary and metastatic colorectal cancer patients. Br J Cancer 102: 151-161, 2010.
40. Lei B, Liu S, Qi W, Zhao Y, Li Y, Lin N, Xu X, Zhi C, Mei J, Yan Z, *et al*: PBK/TOPK expression in non-small-cell lung cancer: Its correlation and prognostic significance with Ki67 and p53 expression. Histopathology 63: 696-703, 2013.
41. Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Ahn JS, Rennert G, Moreno V, Kirchhoff T, Gold B, *et al*: Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat Genet 39: 1338-1349, 2007.
42. Hamilton SR, Fard SF, Paiwand FF, Tolg C, Veiseh M, Wang C, McCarthy JB, Bissell MJ, Koropatnick J and Turley EA: The hyaluronan receptors CD44 and Rhamm (CD168) form complexes with ERK1,2 that sustain high basal motility in breast cancer cells. J Biol Chem 282: 16667-16680, 2007.