



肺鳞状细胞癌特异性甲基化候选诊断标志物的识别*

王学栋 尚文慧 李晓琴** 常 宇

(北京工业大学生命科学与生物工程学院, 北京 100124)

摘要 甲基化异常是肿瘤早期的频发事件, DNA 甲基化随着时间的推移相对稳定, 并且可以在血液中非侵入性地检测到, 因此 DNA 甲基化具有成为癌症早期诊断生物标志物的巨大潜力. 为了找到肺鳞状细胞癌 (LUSC) 潜在的诊断标志物, 本文提出了一种 LUSC 特异性候选诊断标志物的识别方法, 使用癌症基因组图谱数据库 (TCGA) 的 LUSC 的甲基化数据集, 通过比较 LUSC 与正常肺组织和其他癌症类型, 得到了 6 个 LUSC 特异性甲基化位点, 使用支持向量机建立诊断模型, 采用六折交叉划分数据集, 验证特异性标志物的有效性. 6 个标志物的组合在预测 LUSC 方面达到约 93%~99% 的灵敏度, 在排除正常组织时达到 100% 的特异性, 在排除其他癌症时达到约 99% 的特异性. 我们的研究为 LUSC 的早期诊断提供了潜在的生物标志物.

关键词 DNA 甲基化, 特异性标志物, 肺鳞状细胞癌, 早期诊断

中图分类号 Q7, Q8

DOI: 10.16476/j.pibb.2019.0050

肺癌作为最常见和死亡率最高的原发性恶性肿瘤严重危害人类健康. 肺癌主要分为小细胞肺癌和非小细胞肺癌, 在非小细胞肺癌中 30% 以上为鳞状细胞癌 (LUSC). 肺鳞状细胞癌的 5 年生存期仅为 15%^[1], 70% 的患者确诊时已是局部晚期或远处转移 (IV 期), 错过了手术治疗的最佳机会; 如果患者可以在 I a 期被确诊, 5 年生存率可提高到 80%^[2]. 因此, 早期发现 LUSC 对于增加有效治疗的机会和提高存活率是非常重要的.

血清肿瘤标志物的检测是一种广泛用于筛查和诊断肺癌的方法. 肺鳞癌特异性较好的标志物有鳞癌细胞抗原 (SCC) 和细胞角蛋白 19 片段 (CYFRA21-1). 鳞癌细胞抗原 (SCC) 是肺鳞癌较特异的标志物, 肺鳞癌患者中 SCC 阳性率为 40%~55%^[3]. 细胞角蛋白 19 片段 (CYFRA21-1) 是 NSCLC 最灵敏的肿瘤标志物, 尤其对肺鳞癌更是如此. CYFRA21-1 对肺鳞癌的敏感性高达 70% 以上, 然而 CYFRA21-1 依赖肿瘤阶段, 晚期肿瘤 (T3 和 T4) 表现出较高的血清 CYFRA 21-1 水平 ($P < 0.05$)^[4]. 因此需要更有效的生物标志物来提

高肺鳞状细胞癌早期诊断的准确性.

已经在各种癌症中观察到 DNA 甲基化异常, 并且认为其是癌症发生的主要原因. 启动子区高甲基化和基因组整体低甲基化普遍存在于多种肿瘤细胞^[5], 基因组整体低甲基化常见于高度和中度重复的 DNA 序列, 并在染色体不稳定中起着关键作用^[6-8]. 在基因的启动子区域的高甲基化, 通常与基因沉默有关^[9]. 一些 DNA 甲基化涉及癌发生的早期阶段, 如卵巢癌中的 RASSF1A^[10]. 此外, DNA 甲基化随着时间的推移相对稳定并且可以在血液中非侵入性地检测到^[11]. 因此, DNA 甲基化具有成为癌症的早期诊断生物标志物的巨大潜力. 已经开发了越来越多的基于甲基化的生物标志物以帮助癌症的早期诊断^[12]. Sharma 等^[13] 使用 p16INK4A、p14ARF、Cyclin D2 及 Slit2 对乳腺癌

* 重点研发计划 (2017YFC0111104)、国家自然科学基金 (11572014) 和智能化生理测量与临床转化北京市国际科技合作基地资助项目.

** 通讯联系人.

Tel: 15313254516, E-mail: lxq0811@bjut.edu.cn

收稿日期: 2019-03-20, 接受日期: 2019-05-16

进行联合检测, 特异性达到 100%, 灵敏度达到 83%, 且肿瘤和血清情况符合. Anglim 等^[14]使用 GDNF、MTHFR、OPCML、TNFRSF25、TCF21、PAX8、PTPRN2 和 PITX2 可以在 LUSC 患者中区分肿瘤和正常组织, 敏感性和特异性均达到 95.6%. Cheng 等^[15]使用 6 个 CpGs 实现了 HCC 的特异性诊断, 6 个位点的组合在预测 HCC 方面达到约 92% 的灵敏度, 在排除正常肝脏时达到约 98% 的特异性, 并且在排除其他癌症方面具有约 98% 的特异性. 以上研究让我们看到了甲基化对于癌症早期诊断和癌症特异性诊断的可行性.

在这项研究中, 我们通过比较 LUSC 与正常肺组织和其他癌症类型, 确定了 6 种 LUSC 特异性甲基化候选生物标志物. 6 个特异性标志物对于 LUSC 的预测达到约 99% 的敏感性以及 100% 的特异性, 在排除其他癌症时达到约 93% 的敏感性和 99% 的特异性, 并且在排除多种正常组织样本方面达到约 98% 的敏感性和 100% 的特异性.

1 材料与方法

1.1 数据准备

从癌症基因组图谱 (TCGA) 项目 (<https://portal.gdc.cancer.gov/>) 收集乳腺癌 (BRCA, 95 正常人, 126 癌症)、肾透明细胞癌 (KIRC, 154 正常人, 152 癌症)、肾乳头状细胞癌 (KIRP, 45 正常人, 168 癌症)、肝细胞癌 (LIHC, 49 正常人, 174 癌症)、肺鳞状细胞癌 (LUSC, 42 正常人, 172 癌症) 和甲状腺癌 (THCA, 56 正常人, 287 癌症) 的甲基化数据和临床数据, 其中癌症样本为癌症 I 期样本. 此外, 从 Gene Expression Omnibus (GEO) 数据库收集了 184 名年轻芬兰人的血液样本数据 (GSE69270). 甲基化阵列平台是 HumanMethylation450 BeadChip (GPL13534)^[16], CpG 注释从 ENCODE project (<http://genome.ucsc.edu/ENCODE/downloads.html>) 下载^[17].

1.2 差异甲基化筛选

删除所有样本中 “NA” 数目大于 10% 的位点, 剩余 “NA” 采用 K 近邻插补法进行插补^[18], 然后使用单因素方差分析鉴定两类样本之间的差异 CpG. 使用错误发现率 (FDR) 方法调整 P 值, FDR 小于 0.05 并且 β 差异的绝对值大于 0.2 的 CpG 被认为是差异甲基化 CpG.

1.3 鉴定 LUSC 特异性候选诊断标志物

本文通过比较 LUSC 与正常肺组织和其他癌症

类型样本, 提出一种 LUSC 特异性诊断标志物的识别流程, 如图 1 所示.

具体操作流程如下:

第一步, 鉴定在正常和 LUSC 癌症 I 期样本之间的差异 CpG;

第二步, 使用置信区间筛选^[19]在癌症和正常样本中甲基化分布有明显差异的 CpG 位点, 置信区间筛选的计算公式为:

$$[\mu - i \times \sigma, \mu + i \times \sigma] \quad (1)$$

其中 μ 为基因在癌旁或 I 期中的甲基化均值, σ 为标准差, 为保证区间的有效性, i 取值为 1, 分别计算每一个基因在癌旁和 I 期样本的置信区间, 比较两区间是否有重合来筛选对分类结果有差异性的位点;

第三步, 根据弹性网络选择^[20]对分类具有显著贡献的 CpG;

第四步, 配对正常和癌症样本, 删除在配对样本中无差异的 CpG;

第五步, 将所有的癌症 I 期样本与正常人群的血液样本进行比较, 删除在两类样本中无差异的 CpG^[15];

第六步, 删除在其他癌症类型中正常样本和癌症样本甲基化水平均值差异大于 0.1 的 CpG;

第七步, 比较不同癌症类型正常样本的甲基化均值, 删除在不同癌症类型中正常样本甲基化均值差异大于 0.1 的 CpG;

第八步, 使用弹性网络筛选对 LUSC 癌症样本与其他癌症样本的分类具有显著贡献的 CpG;

第九步, 提取弹性网络系数排名靠前的 CpG, 作为最终的候选诊断生物标志物.

1.4 评估候选诊断标志物

基于候选诊断生物标志物的甲基化水平构建支持向量机模型. 该模型用于预测肿瘤和正常样品, 计算灵敏度和特异性以评估预测模型的准确性, 使用六折交叉验证优化模型. 灵敏性和特异性计算公式如下:

$$SEN = \frac{TP}{TP + FN} \quad (2)$$

$$SPE = \frac{TN}{FP + TN} \quad (3)$$

正样本是 LUSC 癌症 I 期样本, 负样本是非 LUSC 样本, TP 表示正确分类的正样本的数量, 即被正确识别为 LUSC 的样本数目, TN 表示正确分类负样本的数量, 即被正确识别为非 LUSC 的样本

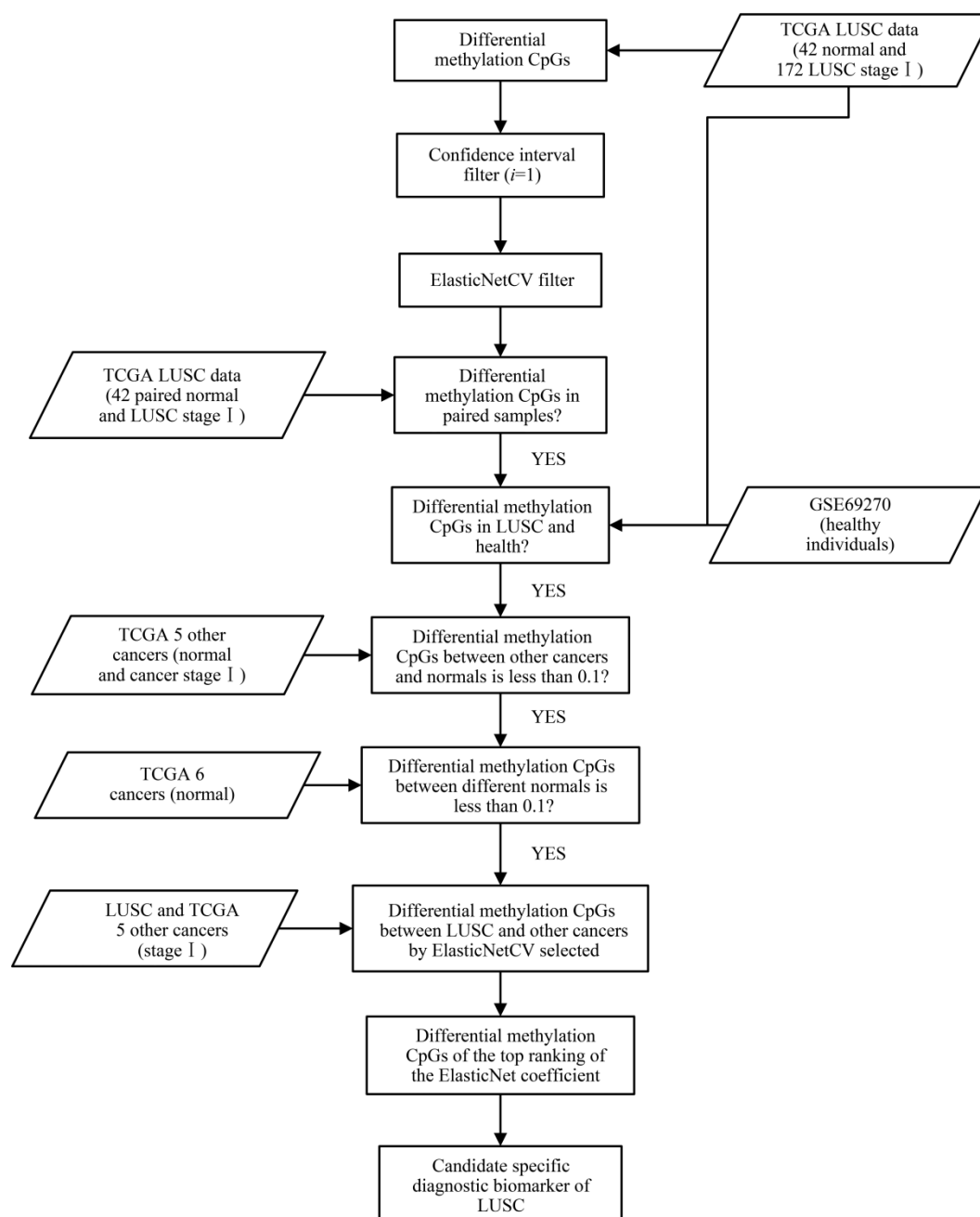


Fig. 1 Protocol for finding candidate diagnostic biomarkers for LUSC

数目, *FP* 表示非 LUSC 样本被误判为 LUSC 的样本数目, *FN* 表示 LUSC 被误判为非 LUSC 的样本数目。

2 结 果

2.1 差异甲基化分析

本节分析了 6 种实体瘤全基因组 (WGS)、启动子区域 (Promoter) 以及启动子区 CpG 岛

(Pro&CGI) 的差异甲基化的分布, 如图 2 所示。

启动子区高甲基化和基因组整体低甲基化普遍存在于多种肿瘤细胞。由图 2 可以看出, 肾透明细胞癌 (KIRC)、肝细胞癌 (LIHC)、肺鳞状细胞癌 (LUSC) 以及甲状腺癌 (THCA) 在基因启动子区域则表现出广泛的低甲基化, 同时肾透明细胞癌和甲状腺癌在启动子区域 CpG 岛上也存在广泛的低甲基化, 这与前面提到的肿瘤细胞中甲基化的分布

具有不一致性, 究其原因可能是癌症发生早期低甲基化起重要作用, 而低甲基化与炎症和免疫反应相

关, 所以炎症反应和免疫相关功能的缺失或异常可能是上述癌症早期发生的主要原因^[21-23]。

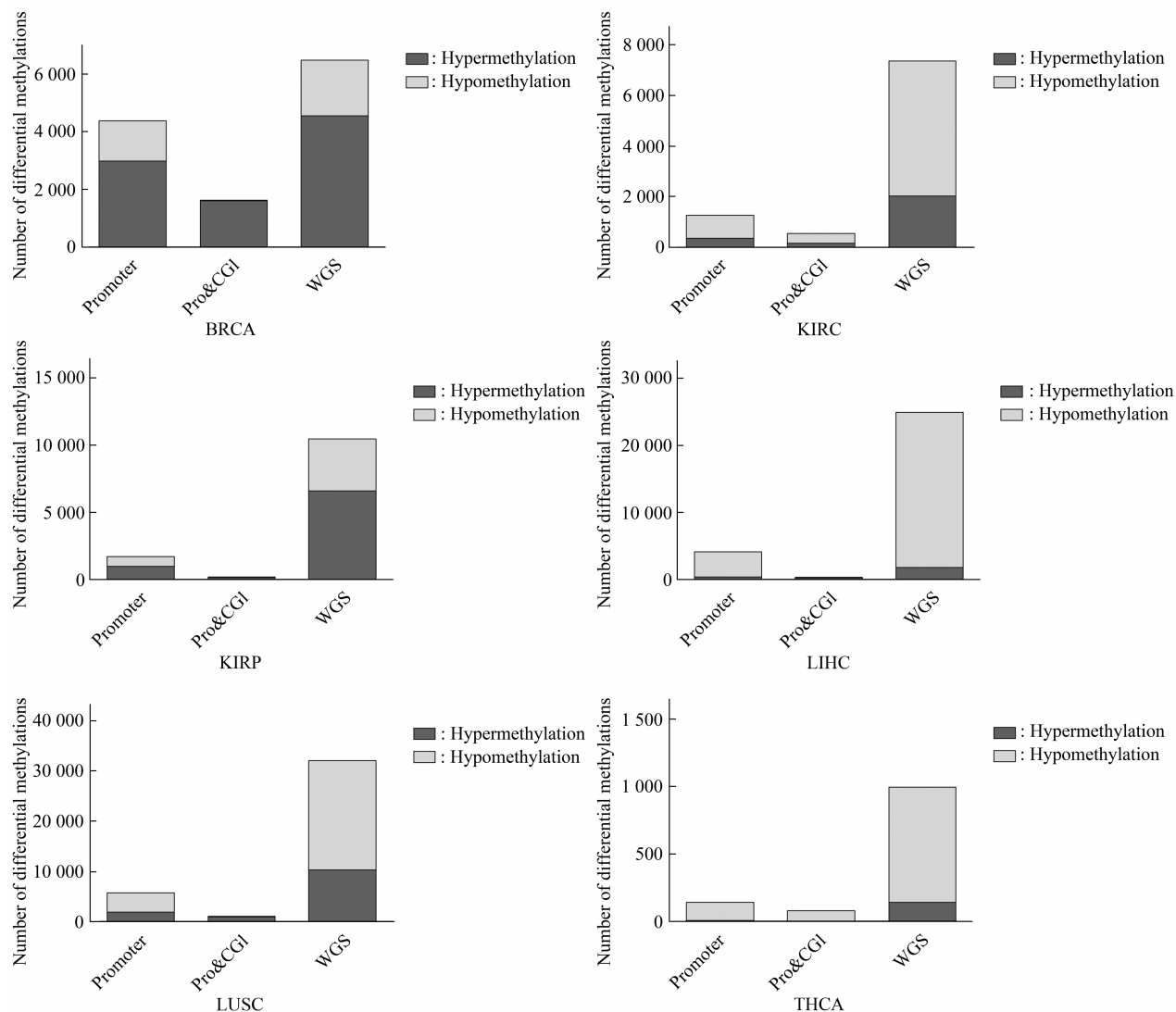


Fig. 2 Distribution of differential methylation CpG

2.2 LUSC特异性候选诊断标志物

为了鉴定LUSC特异性甲基化诊断标志物, 我们设计了一个工作流程并从LUSC的全基因组CpG中筛选得到6个CpG (cg21026460、cg08445080、cg16628135、cg16476940、cg22098115、cg00532449), 其中4个为低甲基化, 2个为高甲基化, 图3显示了6个CpG在LUSC癌旁和癌症I期样本中的甲基化水平分布。

如图3所示, cg21026460、cg08445080、cg16628135、cg16476940在正常样本中甲基化水平接近1.0, 其在癌症I期的样本中甲基化水平则

低于0.8, cg22098115、cg00532449在正常样本中甲基化水平远小于0.1, 在癌症I期样本中则高于0.3, 这6个CpG在正常和癌症I期样本中具有显著差异。

6个LUSC特异CpG定位于6个基因: ARHGEF4 (cg22098115)、PLCH2 (cg00532449)、SH2B2 (cg16476940)、BAIAP3 (cg16628135)、HDAC11 (cg21026460)、RP11-681L4.2 (cg08445080)。除RP11-681L4.2外, 其余5个基因均是蛋白质编码基因, 其功能主要涉及细胞内信号转导、细胞凋亡调控、能量代谢过程的调节、转录

调控以及免疫相关调控（细胞因子介导的信号通路、Ras 蛋白信号转导调控、抗原受体介导的信号通路），上述功能均与癌症发生密切相关。

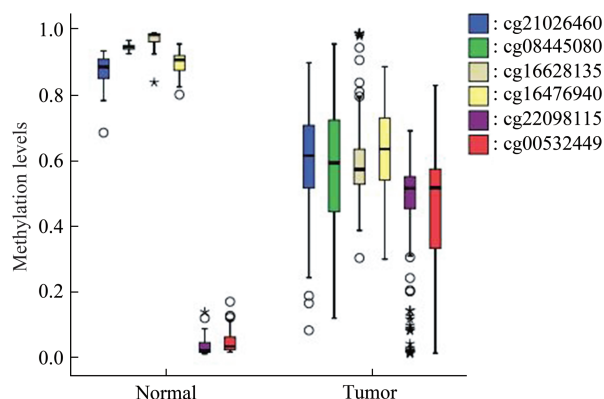


Fig 3 Box chart of methylation levels of six LUSC-specific CpGs in normal and stage I

2.3 评估LUSC的特异性甲基化候选标志物

2.3.1 对LUSC和正常肺组织的识别能力

使用支持向量机建立肿瘤分类模型，同时使用六折交叉验证优化模型。使用6个LUSC特异性的CpG对正常肺组织和LUSC肿瘤I期样本的分类敏感性和特异性分别为99.41%和100%，图4显示了LUSC癌旁和I期样本的分类情况。由图可以看出使用6个LUSC特异性标志物，所有的正常样本全部预测为正常，即误诊率为0，与此同时癌症样本中仅有1例样本被错误地预测为正常，即漏诊率仅为0.58%，说明该6个LUSC特异性候选标志物对癌症具有极高的识别度。

2.3.2 识别LUSC和其余癌症的能力

接下来我们验证了LUSC特异性标志物区分LUSC和其他癌症的能力。使用支持向量机建立肿瘤分类模型，同时使用六折交叉验证优化模型，表1

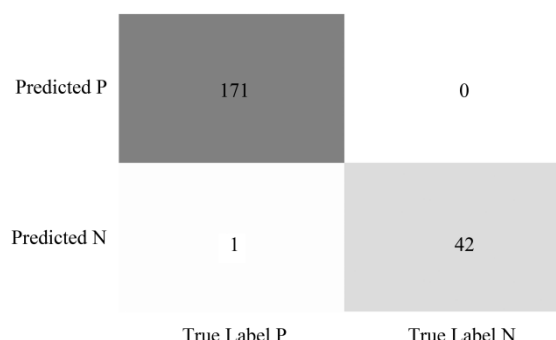


Fig. 4 Prediction matrix of LUSC samples

True Label P, The cancer samples; True Label N, The normal samples; Predicted P, Predicted to be cancer samples; Predicted N, Predicted to be normal samples.

显示了6个LUSC特异性CpG对于LUSC和其余癌症样本以及LUSC和正常人群的模式识别结果。

Table 1 Pattern recognition results of LUSC-specific CpG

Samples	SEN	SPE
LUSC--Normal	98.25%	100%
LUSC--Other cancer	93.02%	99.44%

由表1可知，LUSC特异性CpG在排除正常样本时达到100%的特异性和98%的敏感性，同时在排除其余癌症时达到99.44%的特异性和93%的敏感性，表明使用6个CpG既能够对LUSC样本有极高的识别度，同时也能完成对LUSC的特异性识别。

图5显示了6个LUSC特异性候选标志物在不同癌症类型中的平均甲基化水平。由图5可以看出LUSC特异性候选标志物在其余5种癌症中均未发生差异甲基化，且在除LUSC样本外，其余所有样本的甲基化水平几乎一致，表明该6个CpG对LUSC具有特异性的差异甲基化。

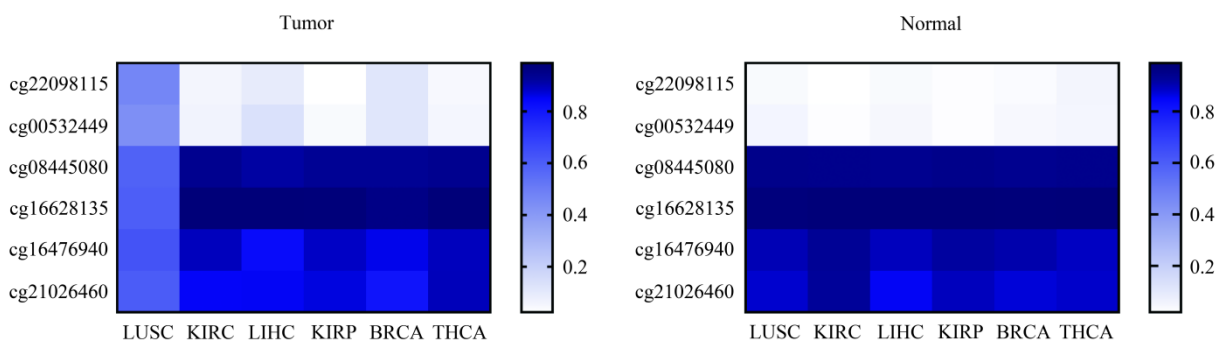


Fig. 5 The average methylation level of six LUSC-specific CpGs in LUSC and other cancers

2.4 筛选其他癌症的特异性候选诊断标志物

本文尝试将LUSC特异性标志物识别流程应用于识别LIHC特异性候选诊断标志物, 结果得到7个可用于在LIHC早期进行诊断的特异性候选诊断标志物 (cg15625324、cg01794405、cg06620541、cg14821923、cg08162372、cg26301389、21074827)。图6, 7分别显示了7个LIHC特异性CpG在LIHC和正常肝脏中的甲基化水平, 以及在LIHC和非LIHC样本中的平均甲基化水平。

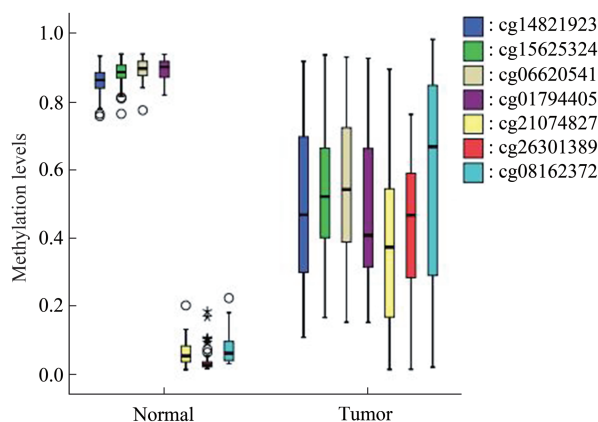


Fig. 6 Box chart of methylation levels of seven LIHC-specific CpGs in normal and stage I

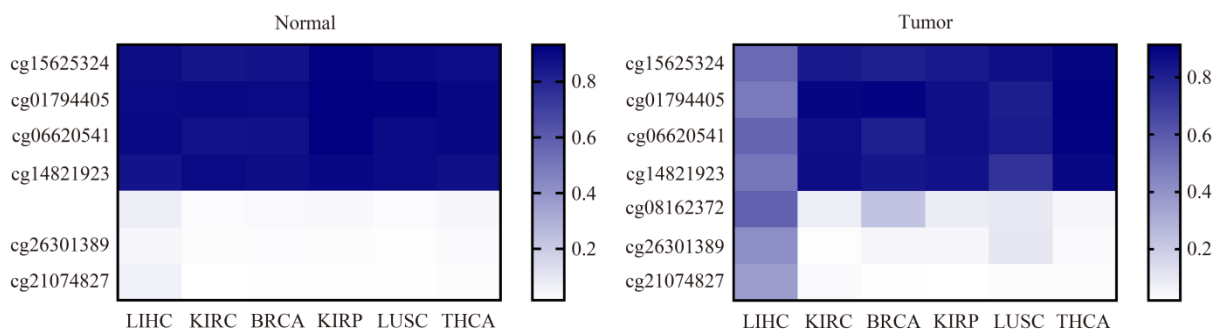


Fig. 7 The average methylation level of seven LIHC-specific CpGs in LIHC and other cancers

由图6可以看出7个CpG在正常肝组织和癌症I期样本中甲基化水平具有显著差异, 由图7可知7个CpG在非LIHC样本中甲基化水平均值几乎一致. 同时LIHC特异性候选标志物模式识别结果显示, 7个CpG对于LIHC与正常肝脏样本的分类敏感性为95%, 特异性为100%, 对LIHC与非LIHC癌症样本的分类敏感性为89%, 特异性为98%, 其对LIHC有较高的辨识能力, 但是其特异性识别能力相对较弱。

同时本节还对其余4种癌症进行特异性候选标志物的识别, 我们发现经过第七步筛选过后, 4种癌症候选CpG的数量明显减少, 其中肾乳头状细胞癌和甲状腺癌分别减少至2和4个. 经最后一步弹性网络筛选, 肾乳头状细胞癌没有得到相应的标志位点, 乳腺癌得到9个候选CpG, 肾透明细胞癌得到4个候选CpG, 甲状腺癌得到4个候选CpG. 除肾乳头状细胞癌外, 其余3种癌症候选CpG对相应癌症样本的特异性识别能力如表2所示。

Table 2 Pattern recognition results of other three cancers-specific CpG

Cancer type	SEN	SPE
BRCA	80.95%	98.99%
KIRC	82.89%	99.41%
THCA	82.98%	99.77%

由表2可以看出, 3种癌症的候选CpG对相应癌症样本的特异性的识别敏感性均低于83%, 对癌症样本的识别能力较低, 而对其余样本的识别准确率都远高于98%, 两者差距高于15%, 不利于临床诊断。

3 讨 论

该研究最重要的发现是鉴定了几种甲基化CpG作为LUSC特异性候选诊断生物标志物. 理想的诊断生物标志物应具有高灵敏度, 能够在早期检测LUSC; 应该特异于LUSC, 而不是在其他肿瘤

类型检测到;应该通过非侵入性和成本效益的技术来衡量;并且应该在不同的人群中进行验证.在这里,通过比较LUSC和正常样本以及LUSC和非LUSC癌症样本,将6种CpG鉴定为LUSC特异性诊断生物标志物.在预测LUSC时,这些CpG的敏感性为93%~99%,特异性为99%~100%.然而,我们尚未使用非侵入性生物样本验证其诊断能力,所以接下来的工作需要比较癌组织和血液之间甲基化的一致性,并通过测量血液中的DNA甲基化来验证候选生物标记物的预测能力.同时,由于肺鳞状细胞癌450K数据较少,无法完成LUSC特异性CpG对不同数据集鲁棒性的验证.

本流程同样应用于其他几种癌症中,我们发现,除了肝细胞癌,其余的4种癌症得到的候选CpG的特异性识别敏感性均低于83%,且模式识别的敏感性和特异性的差距都在15%以上,其中肾乳头状细胞癌甚至没有得到相应的候选标志物.在实验过程中我们发现,甲状腺癌、肾透明细胞癌和肾乳头状细胞癌的特异性差异甲基化位点(即仅在单一癌症中发生差异甲基化的位点)数目非常少,而乳腺癌的特异性差异甲基化位点数目相对较多,但是其对不同的癌症样本的识别能力也偏低,究其原因可能是乳腺癌的发病原因比较复杂,单纯的甲基化可能无法完成特异性的早期诊断.

4 结 论

本文通过分析172个LUSC样本和42个正常肺组织样本以及其他5种癌症类型的正常组织样本和癌症I期样本的全基因组甲基化数据,发现了6种基于甲基化的LUSC特异性候选诊断生物标志物.候选生物标志物对LUSC和正常组织样本的分类敏感性高于98%,特异性高达100%,同时对LUSC和非LUSC癌症样本的分类敏感性高于93%,特异性高于99%.然后,本文将LUSC的特异性标志物识别流程应用于LIHC的分析,得到7个LIHC特异性CpG,其对于LIHC和正常样本以及LIHC和非LIHC具有良好的识别能力.然而,我们尚未使用非侵入性生物样本验证其诊断能力,所以接下来的工作需要比较癌组织和血液之间甲基化的一致性,并通过测量血液中的DNA甲基化来验证候选生物标记物的预测能力,同时未来的转化研究将加速候选生物标志物的临床验证,并促进LUSC的早期检测.

参 考 文 献

- [1] Perez-Moreno P, Brambilla E, Thomas R, *et al.* Squamous cell carcinoma of the lung: molecular subtypes and therapeutic opportunities. *Clinical Cancer Research An Official Journal of the American Association for Cancer Research*, 2012, **18**(9): 2443-2451
- [2] Bach P B. Perilous potential: the chance to save lives, or lose them, through low dose computed tomography screening for lung cancer. *Journal of Surgical Oncology*, 2013, **108**(5): 287-288
- [3] Wang R, Wang G, Zhang N, *et al.* Clinical evaluation and cost-effectiveness analysis of serum tumor markers in lung cancer. *BioMed Research International*, 2013, **2013**: 195692
- [4] Moro D, Villemain D, Vuillez J P, *et al.* CEA, CYFRA21-1 and SCC in non-small cell lung cancer. *Lung Cancer*, 1995, **13**(2): 169-176
- [5] Baylin S B, Jones P A. A decade of exploring the cancer epigenome - biological and translational implications. *Nature Reviews Cancer*, 2011, **11**(10): 726-734
- [6] Karpf A R, Sei-Ichi M. Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. *Cancer Research*, 2005, **65**(19): 8635-8639
- [7] Baba Y, Huttenhower C, Nosh K, *et al.* Epigenomic diversity of colorectal cancer indicated by LINE-1 methylation in a database of 869 tumors. *Molecular Cancer*, 2010, **9**(1): 125
- [8] 董桂银, 钱叶本, 朱立新. DNA甲基化与肝细胞癌. *国际外科学杂志*, 2005, **32**(3): 161-165
Dong G Y, Qian Y B, Zhu L X. *Surgery Foreign Medical Sciences*, 2005, **32**(3): 161-165
- [9] Taby R, Issa J P. Cancer epigenetics. *Ca A Cancer Journal for Clinicians*, 2010, **60**(6): 376-392
- [10] Jin-Ge S, Yuan-Yuan S, Yan-Hua H, *et al.* Role of RASSF1A promoter methylation in the pathogenesis of ovarian cancer: a meta-analysis. *Genet Test Mol Biomarkers*, 2014, **18**(6): 394-402
- [11] Laird P W. Early detection: The power and the promise of DNA methylation markers. *Nature Reviews Cancer*, 2003, **3**(4): 253-266
- [12] Leygo C, Williams M, Hong C J, *et al.* DNA methylation as a noninvasive epigenetic biomarker for the detection of cancer. *Disease Markers*, 2017, **2017**(4): 1-13
- [13] Sharma G, Mirza S, Prasad C P, *et al.* Promoter hypermethylation of p16INK4A, p14ARF, CyclinD2 and Slit2 in serum and tumor DNA from breast cancer patients. *Life Sciences*, 2007, **80**(20): 1873-1881
- [14] Anglim P P, Galler J S, Koss M N, *et al.* Identification of a panel of sensitive and specific DNA methylation markers for squamous cell lung cancer. *Mol Cancer*, 2008, **7**: 62
- [15] Cheng J M, Wei D K, Yuan J, *et al.* Integrative analysis of DNA methylation and gene expression reveals hepatocellular carcinoma-specific diagnostic biomarkers. *Genome Medicine*, 2018, **10**(1): 42
- [16] Kananen L, Marttila S, Nevalainen T, *et al.* Aging-associated DNA methylation changes in middle-aged individuals: the Young Finns

- study. BMC Genomics, 2016, **17**(1): 103
- [17] Yang X F, Gao L, Zhang S H. Comparative pan-cancer DNA methylation analysis reveals cancer common and specific patterns. Briefings in Bioinformatics, 2016, **18**(5): 761-773
- [18] Wang D, Lv Y, Guo Z, *et al.* Effects of replacing the unreliable cDNA microarray measurements on the disease classification based on gene expression profiles and functional modules. Bioinformatics, 2006, **22**(23): 2883-2889
- [19] 温建鑫, 王学栋, 李晓琴, 等. 乳腺癌发生的特征基因筛选及模式识别. 生物化学与生物物理进展, 2017, **44**(11): 1016-1025
Wen J X, Wang X D, Li X Q, *et al.* Prog. Biochem. Biophys, 2017, **44**(11): 1016-1025
- [20] Wang X D, Shang W H, Chang Y, *et al.* Methylation signature genes identification of the lung squamous cell carcinoma occurrence and recognition research. Journal of Computational Biology : a Journal of Computational Molecular Cell Biology, 2018, **25**(10): 1161-1169
- [21] Grivennikov S I, Greten F R, Karin M. Immunity, inflammation, and cancer. Cell, 2010, **140**(6): 883-899
- [22] Kelly-Spratt K S, Pitteri S J, Gurley K E, *et al.* Plasma proteome profiles associated with inflammation, angiogenesis, and cancer. Plos One, 2011, **6**(5): e19721
- [23] Lawrence T. Inflammation and cancer: a failure of resolution?. Trends in Pharmacological Sciences, 2007, **28**(4): 162-165

Identification of Lung Squamous Cell Carcinoma-Specific Methylation Candidate Diagnostic Biomarkers*

WANG Xue-Dong, SHANG Wen-Hui, LI Xiao-Qin**, CHANG Yu

(School of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China)

Abstract DNA methylation abnormalities are frequent events in early tumors. Additionally, DNA methylation is relatively stable over time and can be non-invasively detected in blood. Therefore, DNA methylation has a great potential to become an early diagnostic biomarker of cancers. In order to find potential diagnostic markers for lung squamous cell carcinoma (LUSC), a method for identifying LUSC-specific candidate diagnostic markers was proposed. We screened 6 LUSC-specific CpGs by comparing the methylation profiles of 172 samples from LUSC patients, 42 normal lung samples, 184 normal blood samples, and 1 306 samples from patients with other cancers which was collected from TCGA (The Cancer GenomeAtlas) database. A supportvector machine model was constructed to distinguish LUSC patients from normal controls. The combination of six sites achieved 93%–99% sensitivity in predicting LUSC, 100% specificity in excluding all normal samples, and ~ 99% specificity in excluding other cancers. Overall, our study provides promising biomarkers for the diagnosis of LUSC.

Key words DNA methylation, specific biomarkers, lung squamous cell carcinoma, early diagnosis

DOI: 10.16476/j.pibb.2019.0050

* This study was supported by grants from Key Research and Development Program (2017YFC0111104), The National Natural Science Foundation of China (11572014) and Intelligent Physiological Measurement and Clinical Translation, Beijing International Base for Scientific and Technological Cooperation.

** Corresponding author.

Tel: 15313254516, E-mail: lxq0811@bjut.edu.cn

Received: March 20, 2019 Accepted: May 16, 2019