# 非小细胞肺癌的五基因特征和临床结果

## 摘要

## 背景

目前的分期方法不足以预测非小细胞肺癌（NSCLC）的治疗结果。我们开发了一种与NSCLC患者生存密切相关的五基因特征。

## 方法

我们使用计算机生成的随机数分配185个冷冻标本用于微阵列分析，实时逆转录酶聚合酶链反应（RT-PCR）分析或两者。我们研究了125例随机选择的经历过NSCLC手术切除的患者的肺癌组织冷冻标本中的基因表达，并评估了表达水平与存活率之间的关系。我们使用风险评分和决策树分析来开发基因表达模型，用于预测NSCLC的治疗结果。我们使用来自其他60名患者的随机分配的标本用于验证。

## 结果

通过分析微阵列数据和风险评分，确定了与NSCLC患者存活率相关的16个基因。我们选择了5个基因（*DUSP6，MMD，STAT1，ERBB3*和*LCK*）进行RT-PCR和决策树分析。五基因特征是无复发和总体存活的独立预测因子。我们使用来自60名NSCLC患者的独立队列的数据以及来自86名NSCLC患者的一组公布的微阵列数据验证了该模型。

## 结论

我们的五基因特征与NSCLC患者的无复发和总体生存密切相关。

肺癌 - 主要是非小细胞肺癌（NSCLC） - 是全球癌症死亡的最常见原因。[1](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)早期NSCLC患者的复发率在潜在治愈性治疗后5年内为40％。[2](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)目前NSCLC的分期系统不足以预测治疗结果。

通过微阵列[3,4](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)和逆转录酶聚合酶链反应（RT-PCR）[5,6](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)进行基因表达谱分析（见术语表），可用于对各种类型癌症患者的肿瘤进行分类和制定预后，[7-9](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)包括肺癌。[10-16](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)使用在临床实践中的微阵列是有限的，然而，通过大量的基因谱分析，使用的基因的[17](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)需要复杂的方法，而且缺乏既再现性和独立验证。在肺癌研究中选择用于分析的基因有很大差异; 只有少数基因一直被包括在内。[10-13](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)此外，基因表达谱可根据微阵列平台和所用的分析策略而变化。[6](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)

RT-PCR方法可应用于石蜡包埋的病理标本，并且可重复并适用于临床实践。然而，RT-PCR可用于仅分析少量基因。[17](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)在先前的研究中，我们的组，执行从侵入性NSCLC的标本来源的细胞系的微阵列分析，鉴定与侵入性活性相关的672个基因。[18个](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)我们还确定了基因（*CRMP-1*和*HLJ1*是与NSCLC患者的临床结果相关联）。[19,20](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)最近的一项研究表明，8个基因的RT-PCR分析结果与肺腺癌患者的预后相关。[五](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)

在目前的研究中，我们使用微阵列和实时RT-PCR检查了125个NSCLC手术标本中的基因表达，以便鉴定与临床结果相关的基因标记。

## 方法

## 患者和组织标本

我们使用计算机生成的随机数来分配来自185名连续患者的标本进行微阵列分析。我们研究了1999年12月至2003年12月期间在台中退伍军人总医院接受手术切除NSCLC的125名随机选择患者的肺癌组织冷冻标本。其中60例为腺癌，52例为鳞状细胞癌， 13是其他类型的癌症。我们使用独立队列对60个基因风险预测模型进行了验证，这些患者在1999年11月至2003年12月期间在台中退伍军人总医院接受了NSCLC手术切除的60名随机选择的患者。患者未接受辅助化疗。该研究得到了医院的机构审查委员会的批准。

## 互补DNA的微阵列分析

在我们组的先前研究中鉴定的与侵入活性相关的672个基因[18](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)在尼龙膜上一式两份地重印。我们从每个样本中分离出4μg总RNA，使用扩增试剂盒（Ambion）对其进行扩增，并在逆转录期间用地高辛标记。[21](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)之前已经描述了靶制备，杂交，显色，图像分析和斑点定量的细节。[18,21,22](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)

## RT-PCR分析

为了验证在微阵列分析中发现的基因的表达水平，使用特异性TaqMan探针和引物组对16个基因和TATA盒结合蛋白（*TBP*）的对照基因进行RT-PCR 。用试剂（TaqMan One-Step RT-PCR Master Mix Reagent，Applied Biosystems）和序列检测系统（ABI Prism 7900HT，Applied Biosystems）扩增转录物。使用序列检测器软件和相对定量方法（Applied Biosystems）定量基因表达与*TBP*的表达相关（详情请参阅[补充附录](https://www.nejm.org/doi/suppl/10.1056/NEJMoa060096/suppl_file/nejm_chen_11sa1.pdf)的方法部分，可在www的全文中找到） .nejm.org）。我们选择了*TBP*作为实时RT-PCR的内部对照，因为它在临床癌症标本中是不变的。[23](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)

## 统计分析

将125个样本随机分配到训练集或测试集（参见[补充附录的](https://www.nejm.org/doi/suppl/10.1056/NEJMoa060096/suppl_file/nejm_chen_11sa1.pdf)表1 ）。评估微阵列中每个基因的平均强度。为了减少微阵列之间的变化，通过分位数归一化方法重新调整每个微阵列中样品的强度值。[24](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)为了减少背景噪音，小于3000的背景的强度值被分配的3000值[22](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)然后将每个强度值对数转换为基数2标度。变异系数小于3％的基因被排除在进一步分析之外。最后，根据125名患者中485基因中基因表达水平的排序（60,625个观察结果），将基因表达强度值转换为序数编码值。强度值编码为1，表达水平排在总基因表达的25％或以下，2表示高于25和50或低于50百分位数，3表示高于50和75或低于75百分位数，以及高于第75百分位数的4。

单因素Cox回归分析的危险比用于确定哪些基因与癌症的任何原因或复发有关。保护性基因定义为与死亡风险比小于1相关的基因; 风险基因定义为与死亡风险比相关的风险基因超过1.我们使用单变量Cox比例风险回归分析来评估生存与微阵列分析中每个基因表达水平之间的关联。[25](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)对于已经显著与存活相关的基因，我们所用的基因表达编码由回归系数加权来计算风险分数为每个病人的值的线性组合。[6,10](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)

### *16基因签名*

计算16个基因的风险评分。患者的风险评分计算为每个基因的表达水平之和，通过微阵列分析测量，乘以相应的回归系数（参见[补充附录](https://www.nejm.org/doi/suppl/10.1056/NEJMoa060096/suppl_file/nejm_chen_11sa1.pdf)的方法部分）。患者被分类为具有高风险基因特征或低风险基因特征，风险评分的第50百分位数（中位数）为阈值（中位数，4.9;范围，1.3至21.9）。选择中位风险评分作为阈值，以反映近半数早期NSCLC患者在潜在治愈性手术后5年内复发的事实[2](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)并且通过确保在高风险和低风险组中有相同数量的患者来消除训练队列中极端值的影响。来自训练组的风险评分和阈值未被重新估计，而是直接应用于测试组。

### *五基因签名*

通过RT-PCR确认16种基因的表达水平，并通过Spearman的秩相关性检验进行索引。[26](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)这16个基因，我们进一步确定了将显著与生存相关的五个基因。通过RT-PCR测量的这五种基因的表达水平用于构建递归分区决策树。然后使用[27,28](https://www.nejm.org/doi/full/10.1056/NEJMoa060096) Avadis软件[29,30](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)（Strand Genomic）将患者分类为具有高风险基因特征或基于决策树的低风险基因特征。

我们使用基于RT-PCR而不是微阵列分析的决策树的基本原理是实用性。RT-PCR使用少量基因来捕获相关的协变量结构，尤其是基因表达水平的复杂相互作用和非线性。[28](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)在我们的单变量分裂树中，五个基因中只有一个用于在每个中间节点做出分裂决策。为了避免过度拟合，我们使用了一种称为最小误差的修剪方法（参见“ [补充附录”](https://www.nejm.org/doi/suppl/10.1056/NEJMoa060096/suppl_file/nejm_chen_11sa1.pdf)的“方法”部分和图1 ）。

Kaplan-Meier方法用于估计总体存活率和无复发存活率。使用对数秩检验分析高风险组和低风险组之间的存活差异。采用逐步选择的多变量Cox比例风险回归分析评估与生存相关的独立预后因素，并将五基因特征，年龄，性别，肿瘤分期和组织学特征用作协变量。AP值小于0.05被认为表示统计学显着性，并且所有测试都是双尾的。

我们还在1999年11月至2003年12月期间在台中退伍军人总医院对60名接受NSCLC手术切除的患者进行了独立研究。该队列用于验证我们的五基因风险预测模型。

为进一步验证我们的模型，我们将其应用于86例NSCLC患者的微阵列数据，Beer等报道。[10](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)（可从[http://dot.ped.med.umich.edu:2000/ourimage/pub/Lung/index.html获取](http://dot.ped.med.umich.edu:2000/ourimage/pub/Lung/index.html)）。五个基因（及其相应的Affymetrix探针组）是*DUSP6*（X93920\_at），*MMD*（X85750\_at），*STAT1*（M97936\_at），*ERBB3*（S61953\_at）和*LCK*（M26692\_s\_at）; 对照基因是*TBP*（X54993\_s\_at）。为了使来自微阵列和来自RT-PCR的基因表达水平相当，我们在将值1.1赋予强度值小于1.1之后将微阵列数据对数转换为基数2。对数转换后，将5个基因的表达水平除以对照基因*TBP*的表达水平，以计算相对表达水平。我们使用来自86名NSCLC患者的数据将决策树模型应用于这些相对表达水平。[10](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)由于我们研究中生存分析的最大随访时间为62个月，因此我们使用了86例患者的5年生存率数据。

## 结果

## 16基因签名和生存

在对来自125名患者的肿瘤的微阵列分析中，672个基因中的485个具有大于3％的变异系数，因此包括在分析中。单变量Cox回归分析的危险比表明，16种基因的表达水平与任何原因的死亡相关：4种是保护性基因（风险比小于1），12种是风险基因（与风险比相关）超过1（[表1](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。

[补充附录](https://www.nejm.org/doi/suppl/10.1056/NEJMoa060096/suppl_file/nejm_chen_11sa1.pdf)表1 列出了第一次分析中125名患者的特征。在训练组中的63名患者中，具有高风险评分的肿瘤表达风险基因，而具有低风险评分的肿瘤表达保护性基因（[图1A](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。具有高风险16基因特征的患者的中位总生存期低于具有低风险16基因特征的患者（20个月与未达到）（[图1B](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。具有高风险基因特征的肿瘤与具有低风险基因特征的肿瘤（12个月与未达到）相比具有较低的中位无复发存活率（[图1B](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。培训队列的中位随访时间为20个月。

测试组中的结果与训练组中的结果相似。在62名患者中，具有高风险评分的肿瘤表达风险基因，而具有低风险评分的肿瘤表达保护性基因（[图1A](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。具有高风险16基因特征的患者的中位总生存期低于具有低风险基因特征的患者（[图1C](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。具有高风险基因特征的肿瘤与具有低风险基因特征的肿瘤（18个月与未达到）相比具有较低的中位无复发存活率（[图1C](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。检测队列的中位随访时间为18个月。我们的整个微阵列数据集可在线获取（[www.ncbi.nlm.nih.gov/projects/geo/）数据系列登记号为GSE4882](http://www.ncbi.nlm.nih.gov/projects/geo/）数据系列登记号为GSE4882)。

## 五基因签名与生存

对于125个肿瘤样本中的101个中的16个基因中的5个的基因表达数据，微阵列和RT-PCR分析的结果之间存在显着相关性（[表1](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。这五个基因用于双特异性磷酸酶6（*DUSP6*），单核细胞 - 巨噬细胞分化相关蛋白（*MMD*），信号转导和转录激活因子1（*STAT1*），v-erb-b2禽红细胞白血病病毒致癌基因同源3 （*ERBB3*）和淋巴细胞特异性蛋白酪氨酸激酶（*LCK*）。

根据用RT-PCR和决策树分析测量的基因表达，我们鉴定了59名具有高风险基因特征的患者和42名具有低风险基因特征的患者（参见[补充附录的](https://www.nejm.org/doi/suppl/10.1056/NEJMoa060096/suppl_file/nejm_chen_11sa1.pdf)图1 ）。决策树的结构基于五个基因中每一个的表达阈值，如根据递归分区算法自动确定的。使用该算法可以将具有高风险特征的患者与具有低风险特征的患者最准确地分离。[表2](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)根据他们的五基因特征，总结了101名患者的临床特征，以下称为原始队列。五基因特征与总体存活率密切相关（敏感性，98％;特异性，93％;阳性预测值，95％;阴性预测值，98％;总体准确性，96％）。

101名患者的中位随访时间为20个月。具有高风险基因特征的患者具有比具有低风险基因特征的患者更短的中位总生存期（20个月对40个月，通过对数秩检验P <0.001）（[图2A](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。高风险基因特征与13个月无中位无复发生存期相关，而低风险基因特征与29个月无中位无复发生存期相关（对数秩检验P = 0.002）（[图2B](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。

根据Cox多变量回归分析，高风险五基因特征，肿瘤分期III和老年人与101名患者中的任何原因的死亡显着相关（[表3](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)），以及高风险的五基因特征和肿瘤III期与癌症复发显着相关（高风险特征与低风险特征的风险比，1.92; 95％置信区间[CI]，1.06至3.46; P = 0.03;危险比为III期对I期或II期疾病，2.28; 95％CI，1.33至3.91; P = 0.003）。在对59名I期或II期疾病患者进行的亚组分析中，具有高风险基因特征的患者的总生存期较短，无复发生存率低于具有低风险基因特征的患者（分别为[图2C和2D](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)） ）。

## 验证五基因签名

验证队列中60名患者的临床特征列于[表2中](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)。中位随访时间为17个月。具有高风险基因特征的患者具有比具有低风险基因特征的患者更短的中位总生存期（21个月与未达到）（[图2E](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。根据Cox多变量回归分析，五基因特征与总体存活率显着相关（[表3](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。

我们分析了在患有I期或II期疾病的验证队列中的患者获得的肿瘤标本中的五基因特征。在患有I期或II期疾病的患者中，具有高风险基因特征的患者的总生存期比具有低风险基因特征的患者短（[图2F](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。在患有I期疾病的患者中，低风险基因特征与高风险基因特征相比具有更长的总体存活率（通过对数秩检验P = 0.02）。在患有II期疾病的患者中，高危患者和低风险基因签名患者的总体生存率没有显着差异，可能是由于患者数量较少。

我们还在来自西方NSCLC人群的86名患者的一组独立的微阵列数据中验证了五基因特征。[10](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)[补充附录的](https://www.nejm.org/doi/suppl/10.1056/NEJMoa060096/suppl_file/nejm_chen_11sa1.pdf)表2 根据其五基因特征列出了这86名患者的临床特征。具有高风险基因特征的患者的总生存期比具有低风险基因特征的患者短（[图2G](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）（对数秩检验P = 0.06）。根据Cox多变量回归分析，高风险五基因特征和肿瘤III期与任何原因的死亡显着相关（[表3](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)）。

## 讨论

NSCLC是一种异质性疾病。即使在具有相似临床和病理特征的患者中，结果也各不相同：一些患者已治愈，而在另一些患者中，癌症复发。基于临床和病理结果的肺癌分期系统可能已达到预测结果的有效性限制，但分子方法增加了价值。使用微阵列[3,4](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)或PCR [5,6](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)进行的基因表达谱分析已经显示可以准确地估计肺癌患者的预后。[10-16](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)然而，在临床实践中使用微阵列受到分析中大量基因的限制，[17](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)复杂的方法，缺乏可重复性和结果的独立验证，以及对新鲜冷冻组织的需求。[17](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)涉及少量基因的RT-PCR可能是临床上更有用的方法。它允许从少量石蜡包埋的样本中获得的RNA的结果的准确和可重复的定量。[对于](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)从总共45种基因中选出的8种基因进行的RT-PCR结果最近显示与肺腺癌的结果相关。[五](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)

我们使用基于微阵列和来自NSCLC患者的125个冷冻肿瘤标本的决策树分析的风险评分*，*鉴定了基于RT-PCR的五基因标记（包括*DUSP6，MMD，STAT1，ERBB3*和*LCK*）。将样本随机分成训练组（63个样本）和测试集（62个样本）。NSCLC肿瘤中存在高风险的五基因特征与复发风险增加和总体存活率降低相关。

我们在微阵列训练集中选择的基因在微阵列测试集中得到验证，并且通过RT-PCR验证在微阵列分析中发现的基因表达模式。我们的结果也在一个由台中退伍军人总医院接受治疗的60名患者的独立队列中得到验证。我们的中国患者的这些结果也通过使用来自西方NSCLC患者的一组公布的NSCLC微阵列数据进行了验证。因此，我们认为使用五基因特征获得的数据是可靠的。

鉴定与NSCLC患者的结果密切相关的五个基因具有临床意义。基于顺铂的辅助化疗对一些NSCLC患者有效。[32](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)我们建议，谁拥有一个高风险的基因标记肿瘤患者可以从这种类型的辅助治疗中受益，而那些具有低风险的基因标记可以幸免了可能是不必要的治疗。为了测试这个想法，有必要进行前瞻性，大规模，多中心的研究。

可以预测NSCLC患者临床结果的五种基因的鉴定可以揭示肺癌治疗发展的目标。STAT1通过诱导p21 WAF1和caspase 的表达引起许多类型癌细胞的停滞生长和凋亡。[33,34](https://www.nejm.org/doi/full/10.1056/NEJMoa060096) MMD优先在成熟巨噬细胞中表达。[35](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)我们小组已经表明，巨噬细胞活化促进癌症转移，[22](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)虽然MMD蛋白质的功能是未知的。DUSP6灭活细胞外信号调节激酶2（ERK2）（也称为丝裂原活化蛋白激酶1 [MAPK1]），导致肿瘤抑制和凋亡。[36](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)ERBB3是酪氨酸激酶的表皮生长因子受体家族的成员，可以缩短细胞存活。[37](https://www.nejm.org/doi/full/10.1056/NEJMoa060096) LCK是Src蛋白酪氨酸激酶家族的成员，主要在T细胞中表达，是T细胞受体下游的第一个信号分子之一。它不仅在T细胞的分化和活化中起关键作用，而且在诱导细胞凋亡中起关键作用。[38](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)此外，LCK在许多癌症中表达并调节癌细胞的移动性。[39,40](https://www.nejm.org/doi/full/10.1056/NEJMoa060096)

总之，我们鉴定的五基因表达特征与手术切除的NSCLC患者的临床结果密切相关。根据疾病辅助治疗试验中的风险，该签名可用于对患者进行分层。

得到了中华人民共和国国家科学委员会国家基因组医学研究计划（NSC94-3112-B002-013-Y）和Advpharma的资助。

Terng博士报告称他是Advpharma的员工。没有报道与本文相关的其他潜在利益冲突。

博士。WJ Chen，JJW Chen和PC Yang对本文作出了同样的贡献。