An Embryonic Stem Cell–Like Signature Identifies Poorly Differentiated Lung Adenocarcinoma but not Squamous Cell Carcinoma

Khaled A. Hassan,1 Guoan Chen,2 Gregory P. Kalemkerian,1 Max S. Wicha,1 and David G. Beer2

Abstract

Purpose: An embryonic stem cell (ESC) profile correlates with poorly differentiated breast, bladder, and glioma cancers. In this article, we assess the correlation between the ESC profile and clinical variables in lung cancer.

Experimental Design: Microarray gene expression analysis was done using Affymetrix Human Genome U133A on 443 samples of human lung adenocarcinoma and 130 samples of squamous cell carcinoma (SCC). To identify gene set enrichment patterns, we used the Genomica software.

Results: Our analysis showed that an increased expression of the ESC gene set and a decreased expression of the Polycomb target gene set identified poorly differentiated lung adenocarcinoma. In addition, this gene expression signature was associated with markers of poor prognosis and worse overall survival in lung adenocarcinoma. However, there was no correlation between this ESC gene signature and any histologic or clinical variable assessed in lung SCC.

Conclusions: This work suggests that not all poorly differentiated non–small cell lung cancers exhibit a gene expression profile similar to that of ESC, and that other characteristics may play a more important role in the determination of differentiation and survival in SCC of the lung. (Clin Cancer Res 2009;15(20):OF1–5)

The cancer stem cell theory postulates the existence of a distinct population of undifferentiated cells responsible for tumor initiation and maintenance (1). In a seminal article, Kim et al. described a rare population of bronchioalveolar stem cells in adult mice. This population possesses the ability of self-renewal and multipotent differentiation and is crucial in lung repair after injury (2). The bronchioalveolar stem cell population was found in the precursor lesions of a mouse model of adenocarcinoma (3). In human lung cancer, several studies have shown the presence of clonogenic populations that possess cancer stem cell properties, using different markers including Hoechst 33342, urokinase-type plasminogen receptor, CD133, and aldehyde dehydrogenase (4–7). Cancer stem cells have the capacity for self-renewal, multipotency, and unlimited proliferation. These traits also characterize embryonic stem cells (ESC), thus suggesting probable overlap in the molecular signature between ESC and cancer stem cells.

ESC lines were first identified in 1998 and their molecular profiles have been determined in various studies (8). A meta-analysis identified 38 original studies analyzing the transcriptome of human ESC lines derived from human blastocysts (9). Genes that were consistently overexpressed or underexpressed in ESC as compared with differentiated cells were identified. Twenty ESC gene lists were collected from these studies, and 380 genes were found to be commonly overexpressed in five of them. Furthermore, Polycomb (10), Nanog (11), Oct4 (12), Sox2 (13), and their target genes play a major role in controlling ESC and seem to be involved in different cancer types. The expression of these genes and the possible correlation with differentiation status and outcome were assessed by Ben-Porath et al. (14) in various human tumors. They showed that an increase in the expression of the ESC gene set and a decrease in the expression of the Polycomb target gene set identified poorly differentiated breast cancer, glioma, and bladder cancer. In addition, patients whose tumors possessed such an expression profile had worse overall survival as compared with others. This was intriguing, as ESC regulatory genes seem to be crucial in determining differentiation and prognosis in multiple cancers. In this work, we attempted to establish whether these findings can be generalized to other cancers, namely, the adenocarcinoma and squamous cell carcinoma (SCC) subtypes of non–small cell lung cancer.

Authors’ Affiliations: 1Department of Internal Medicine and 2Section of Thoracic Surgery, Department of Surgery, Comprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan.

Received 5/1/09; revised 7/13/09; accepted 7/20/09; published OnlineFirst 10/6/09.

Grant support: Supported in part by the University of Michigan Cancer Center NIH support grant SPO3 CA146592.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

K.A. Hassan and G. Chen contributed equally to this work.

Requests for reprints: Khaled A. Hassan, C363 Med Inn Building, 1500 East Medical Center Drive, Ann Arbor, MI 48109-0848. Phone: 734-615-1623; Fax: 734-615-2109; E-mail: khaledh@med.umich.edu.

© 2009 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-09-1105
Translational Relevance

Our study shows that overexpression of the embryonic stem cell (ESC) profile correlates with various poor clinical features in adenocarcinoma of the lung, including smoking, lymph node involvement, and advanced stage. We have also shown that over-expression of this profile is an independent poor prognostic factor in adenocarcinoma, which can be used clinically as a prognostic tool. Furthermore, the ESC pathways that control self-renewal, multipotency, and unlimited proliferation ability represent components that could be targeted with specifically tailored treatments. In addition, this work highlights the difference in the ESC gene expression profile between adenocarcinoma and squamous cell carcinoma of the lung and raises an important issue about similar treatment approaches in these lung cancer subtypes.

Materials and Methods

Specimens and gene sets. Details of the adenocarcinoma specimens, criteria for inclusion, mRNA processing and hybridization procedures, and pathologic and clinical data are all available from ref. 15. Similarly, the SCC details are available from ref. 16. A summary of the clinical variables in 443 adenocarcinomas and 130 squamous cell lung cancers used in this study is provided in Supplementary Table S1. In addition, the correlation of clinical variables with survival is provided in Supplementary Table S2. The original gene sets of embryonic stem (ES) cell; Polycomb (PRC) targets; Nanog, Oct4, and Sox2 (NOS) targets; and Myc targets were obtained from Ben-Porath et al. (14). We matched the original gene name to the Affymetrix Human Genome U133A gene name, and we focused on gene sets ES exp1, PRC2 targets, NOS targets, and Myc targets. The gene list is provided in Supplementary Table S3.

Gene expression data and analysis of gene set enrichment. Microarray gene expression data on 443 human lung adenocarcinomas (15) and 130 squamous cell lung cancers (16) were downloaded from the websites described in the original articles. Raw data were processed by log 2 transformation of the expression values, and the mean center expression level for each gene across all samples was determined. The processed expression data are provided as Supplementary Tables S4 and S5. To identify gene set enrichment patterns, we used the Genomica software used by Ben-Porath et al. (16), which was downloaded. 3 In brief, we identified genes that were overexpressed or underepressed in each sample, determined genes whose expression was at least 2-fold above or below the mean expression level, and calculated a P value. A threshold of P < 0.05 was used as a cutoff for significant enrichment. We determined the gene set to which each differentially expressed gene in a specific sample belonged. Then, for all samples showing enrichment for a particular gene set, we determined the correlation between the samples and each clinical variable annotation, and assigned a P value according to the hypergeometric distribution. We used a more stringent threshold of P < 0.01 for this calculation.

Real-time reverse transcription-PCR. To validate the ES cell gene expression of the microarray data, we performed real-time PCR experiments using Custom TaqMan Low Density Arrays (Applied Biosystems) on 47 lung cancers. A total of 109 genes were randomly picked from ES, PRC2, and other gene lists used in this study. A standard reverse transcription-PCR technique was run on the Applied Biosystems 7900HT Fast Real-Time PCR System. For detailed information on TaqMan arrays as well as card setup and data analysis, refer to the TaqMan Low Density Array Getting Started Guide (P/N 4319399), which can be downloaded from the ABI website. 4

Statistical analysis. Statistical analyses were done using the R package. 5 Individual tumors enriched for overexpression of the ES exp1 set were considered to have an ES signature. P values were calculated using the log-rank test and Kaplan-Meier survival curves comparing the group of individuals with tumors showing the ES signature to all other individuals. Survival-related genes were selected by Cox regression model, and differentiation-related genes were obtained using t test by comparing well-differentiated with poorly differentiated lung tumors. Spearman correlation was used for the correlation analysis of ES genes between real-time PCR and microarray data.

Results

ESC and Polycomb gene set expression correlate with differentiation status in lung adenocarcinoma. We performed microarray gene expression analysis using Affymetrix Human Genome U133A on 443 samples of human lung adenocarcinoma (15). Using the Genomica software as described by Ben-Porath et al., we analyzed the expression of the ESC, NOS, Myc, and Polycomb gene sets according to various clinical features. Increased ESC gene set expression (P = 1 × 10^-10) and decreased Polycomb gene set expression (P = 6.3 × 10^-8) were detected in histologically poorly differentiated tumors (Fig. 1A). This association was independent of proliferation and remained significant even after eliminating proliferation-related genes from both ESC (P = 1.2 × 10^-5) and Polycomb (P = 0.01) gene sets. This indicates that poorly differentiated tumors express genes that are related to those of ESC, and that such tumors may include a more robust cancer stem cell population.

ESC gene set expression associates with poor clinical variables. Patients with advanced stage disease (T2, T3, and T4) had increased expression of the ESC gene set as compared with patients with T1 disease, who had a decreased expression (Fig. 1B). Similarly, patients with lymph node involvement (N1 and N2) had increased expression of the ESC gene set as compared with patients with no lymph node involvement (N0). Current smokers also had increased expression of the ESC gene set (Fig. 1B). Clinically, current smokers and patients with advanced stage disease or lymph node involvement have poor outcome. This suggests that ESC gene set expression correlates with markers of poor prognosis in lung adenocarcinoma.

Poor prognosis is associated with ESC gene set expression. To determine whether the ESC gene set expression correlates with poor prognosis, we performed Kaplan-Meier and log-rank test analyses of overall survival. The analyses showed that patients whose tumors had increased expression of the ESC gene set had a worse 5-year overall survival than patients with decreased expression (P = 0.005; Fig. 1C). Kaplan-Meier analysis of overall survival based on differentiation showed a non-significant trend toward worse 5-year overall survival in patients with poorly differentiated tumors as compared with patients with moderately differentiated or well-differentiated (P = 0.06) tumors (Fig. 1D). This analysis shows that poorly differentiated lung adenocarcinomas possess a molecular signature that is similar to the ESC.
profile, and that patients with such a profile have a poor prognosis. This may also indicate that such tumors possess a larger cancer stem cell population as compared with well-differentiated or moderately differentiated tumors.

**ESC gene set expression in squamous cell lung cancer.** To assess whether these findings apply to squamous cell lung cancer, we further analyzed the expression of ESC and Polycomb target gene sets in 130 samples of lung SCC (16). There was no correlation between the expression of these gene sets and any histologic or clinical variable assessed, including differentiation and survival (Fig. 2A). In an attempt to understand these unexpected results, we performed a Cox regression model or t test–based analysis of Polycomb, NOS, and Myc target genes for survival and differentiation in the lung adenocarcinoma and SCC samples, and these analyses detected no significant difference (results not shown). Further, the percentage of relation-related genes expressed in the ESC gene set was 28.6% in adenocarcinoma as compared with 5.9% in SCC, and the percentage of poor-differentiation–related genes expressed in the ESC gene set was 44.4% in adenocarcinoma as compared with 3.6% in SCC (Fig. 2B). The variation in expression of these genes in SCC samples (Fig. 2C), despite being statistically significant, was less compared with the variation seen in the adenocarcinoma samples (Fig. 2D). This implies that the ESC and Polycomb target gene sets do not correlate with the genes that determine differentiation or survival in SCC of the lung. This is in contrast to other tumor types, including adenocarcinoma of the lung.

**Discussion**

Cancer stem/progenitor cells were initially identified in acute myelogenous leukemia (17) and recently have been identified in several solid tumors, including melanoma and breast, brain, prostate, pancreatic, and colon carcinomas (18–24). The capacity for self-renewal, multipotency, and unlimited proliferation is shared between cancer stem cells and ESC. This suggests that pathways controlling such biological processes might be shared between ESC and cancer stem cells. In an effort to establish the gene expression profile of ESC, Ben-Porath et al. identified 380 genes, designated gene set ES exp1, which were commonly overexpressed in ESC (14). Furthermore, a Polycomb target gene set representing overlapping genes bound to Polycomb repressive complex 2 (PRC2) in human ESC was designated as PRC2 targets. Overlapping Nanog, Oct4, and Sox2 target genes

---

**Fig. 1.** Poorly differentiated lung adenocarcinomas possess an ESC expression pattern that correlates with poor prognosis. A, expression pattern of gene sets (rows) in 443 lung adenocarcinoma samples. Red and green, overexpressed or underexpressed gene sets, respectively. Brown bars (bottom) indicate each sample annotation for grade: 1, well-differentiated tumor; 2, moderately differentiated tumor; and 3, poorly differentiated tumor. Gene set expression (right) with or without proliferation genes in lung adenocarcinoma samples stratified by tumor grade. B, gene set expression stratified by smoking status, tumor-node-metastasis classification, and stage. C, Kaplan-Meier curve analysis of overall survival in patients with lung adenocarcinoma with overexpressed or underexpressed ESC gene set. D, Kaplan-Meier curve analysis of overall survival in lung adenocarcinoma patients based on tumor grade, with P value indicating significance of survival difference between grade 1 and grade 3 tumors.
were designated as NOS targets, and genes affected by Myc were designated as Myc targets.

Using these gene sets and Genomica software, Ben-Porath et al. showed an inverse relationship between differentiation and outcome in breast carcinoma, glioblastoma, and bladder carcinoma. The enrichment of an ESC-like gene set signature was identified by an overexpression of the ESC gene set and a decreased expression of the PRC2 target gene set. In this study, we applied the same gene sets and software used by Ben-Porath et al. to lung cancer samples, and our results confirm that an ESC-like gene expression profile is preferentially detected in histologically poorly differentiated lung adenocarcinoma, independent of cell proliferation. In addition, advanced stage disease, lymph node involvement, and current smoker status correlated with the ESC-like gene expression profile, and overall survival was worse in patients who expressed this profile. These findings clearly suggest that ESC genes are involved in both differentiation and prognosis of lung adenocarcinoma. Because the lung cancer stem cell has not yet been definitively identified, a direct correlation between the ESC and lung cancer stem cell expression profiles cannot be done. To confirm the microarray findings, real-time quantitative PCR was done on 47 samples for 109 genes. The Spearman correlation analysis shows that 88.1% (96 of 109) of the genes have good correlation to microarray data (R > 0.5; Supplementary Fig. S1).

Interestingly, these findings did not apply to lung SCC. No correlation between the expression of these gene sets and any histologic or clinical variable assessed was detected in SCC. Specifically, overexpression of ESC genes had no effect on differentiation or survival. This could be explained by the fact that adenocarcinoma had a higher percentage of survival-related and poor-differentiation–related genes expressed in the ESC gene set as compared with SCC. This implies that the ESC and Polycomb gene sets do not correlate with the genes driving differentiation or affecting survival in SCC, a finding that is in direct contrast to adenocarcinoma.

Several studies have used gene signature profiles to predict patient outcome (25–27). Data from these profiles vary, and there is a lack of consistency among published studies. Attempts to compare profiles and evaluate whether the results could be integrated were inconsistent, but a common gene profile that is a significant predictor of survival could be identified (28). In addition, similarity in gene sets that are prognostic for both adenocarcinoma and SCC has been identified (16). This article is the first to use ESC profiling in lung cancer with demonstration of differences among subtypes of lung cancer.

In conclusion, these studies suggest that although many poorly differentiated tumors of different tissue origins exhibit a gene expression profile similar to ESC, it is not a universal phenomenon, and other characteristics play a major role in some cancers.

**Disclosure of Potential Conflicts of Interest**

M.S. Wicha holds equity in and is a scientific consultant for OncoMed Pharmaceuticals. The other authors disclosed no potential conflicts of interest.
References

1. Pardal R, Clarke MF, Morrison SJ. Applying the principles of stem-cell biology to cancer. Nat Rev Cancer 2003;3:895–902.

2. Kim CF, Jackson EL, Woolfenden AE, et al. Identification of bronchioalveolar stem cells in normal lung and lung cancer. Cell 2005;121:823–35.

3. Jackson EL, Willis N, Mercer K, et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. Genes Dev 2001;15:3243–8.

4. Ho MM, Ng AV, Lam S, et al. Side population in human lung cancer cell lines and tumors is enriched with stem-like cancer cells. Cancer Res 2007;67:4827–33.

5. Gutova M, Najbauer J, Gevorgyan A, et al. Identification of uPAR-positive chemoresistant cells in small cell lung cancer. PLoS ONE 2007;2:243.

6. Eramo A, Lotti F, Sette G, et al. Identification and expansion of the tumorigenic lung cancer stem cell population. Cell Death Differ 2008;15:504–14.

7. Jiang F, Qiu Q, Khanna A, Todd NW, et al. Aldehyde dehydrogenase 1 is a tumor stem cell-associated marker in lung cancer. Mol Cancer Res 2009;7:330–8.

8. Thomson JA, Itskovitz-Eldor J, Shapiro SS, et al. Embryonic stem cell lines derived from human blastocysts. Science 1998;282:1145–7.

9. Assou S, Le Carrou T, Tondeur S, et al. A meta-analysis of human embryonic stem cell transcriptome integrated into a web-based expression atlas. Stem Cells 2007;25:961–73.

10. O’Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA, Jenneu T. The polycomb-group gene Ezh2 is required for early mouse development. Mol Cell Biol 2001;21:4330–6.

11. Chambers I, Colby D, Robertson M, et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 2003;113:643–55.

12. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 2000;24:372–6.

13. Graham V, Khudyakov J, Ellis P, et al. Sox2 functions to maintain neural progenitor identity. Neuron 2003;39:749–65.

14. Ben-Porath I, Thomson MW, Carey VJ, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet 2008;40:499–507.

15. Shedden K, Taylor JM, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nat Med 2008;14:822–7.

16. Raponi M, Zhang Y, Yu J, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. Cancer Res 2006;66:7466–72.

17. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 1997;3:730–7.

18. Fang D, Nguyen TK, Leishear K, et al. A tumorigenic subpopulation with stem cell properties in melanomas. Cancer Res 2006;66:3326–37.

19. Al-Hajj M, Wicha MS, Benito-Hernandez A, et al. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 2003;100:3983–8.

20. Patrawala L, Calhoun T, Schneider-Broussard R, et al. Highly purified CD44+ prostate cancer cells from xenograft human tumors are enriched in tumorigenic and metastatic progenitor cells. Oncogene 2006;25:1696–708.

21. Singh SK, Clarke ID, Terasaki M, et al. Identification of a cancer stem cell in human brain tumors. Cancer Res 2003;63:5821–8.

22. Li C, Heidt DG, Dalerba P, et al. Identification of pancreatic cancer stem cells. Cancer Res 2007;67:1030–7.

23. Ricci-Vitiani L, Lombardi DG, Pilozi E, et al. Identification and expansion of human colon-cancer-initiating cells. Nature 2006;445:111–5.

24. O’Brien CA, Pollett A, Gallinger S, et al. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature 2006;445:106–10.

25. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816–24.

26. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A 2001;98:13780–5.

27. Guo L, Ma Y, Ward R, et al. Constructing molecular classifiers for the accurate prognosis of lung adenocarcinoma. Clin Cancer Res 2006;12:3344–54.

28. Parmigiani G, Garrett-Mayer ES, Anbazhagan R, et al. A cross-study comparison of gene expression studies for the molecular classification of lung cancer. Clin Cancer Res 2004;10:2922–7.
Clinical Cancer Research

An Embryonic Stem Cell–Like Signature Identifies Poorly Differentiated Lung Adenocarcinoma but not Squamous Cell Carcinoma

Khaled A. Hassan, Guoan Chen, Gregory P. Kalemkerian, et al.

Clin Cancer Res  Published OnlineFirst October 6, 2009.

Updated version  Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-09-1105

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, use this link  http://clincancerres.aacrjournals.org/content/early/2009/10/03/1078-0432.CCR-09-1105. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.