Characterizing the Clinical Relevance of an Embryonic Stem Cell Phenotype in Lung Adenocarcinoma

Marvaretta Stevenson,† William Mostertz,2 Chaitanya Acharya,2 William Kim,2 Kelli Walters,2 William Barry,3 Kristin Higgins,3 Sascha A. Tuchman,1 Jeffrey Crawford,1 Gordana Vlahovic,1 Neal Ready,1 Mark Onaitis,5 and Anil Potti1,2

Abstract

Purpose: Cancer cells possess traits reminiscent of those ascribed to normal stem cells. It is unclear whether these phenotypic similarities are the result of a common biological phenotype, such as regulatory pathways.

Experimental Design: Lung cancer cell lines with corresponding gene expression data and genes associated with an embryonic stem cell identity were used to develop a signature of embryonic stemness (ES) activity specific to lung adenocarcinoma. Biological characteristics were elucidated as a function of cancer biology/oncogenic pathway dysregulation. The ES signature was applied to three independent early-stage (I-IIIa) lung adenocarcinoma data sets with clinically annotated gene expression data. The relationship between the ES phenotype and cisplatin (current standard of care) sensitivity was evaluated.

Results: Pathway analysis identified specific regulatory networks [Ras (P = 0.0005), Myc (P = 0.0224), wound healing (P < 0.0001), chromosomal instability (P < 0.0001), and invasiveness (P < 0.0001)] associated with the ES phenotype. The prognostic relevance of the ES signature, as related to patient survival, was characterized in three cohorts [CALGB 9761 (n = 82; P = 0.0001), National Cancer Institute Director’s Challenge Consortium (n = 442; P = 0.0002), and Duke (n = 45; P = 0.006)]. The ES signature was not prognostic in prostate, breast, or ovarian adenocarcinomas. Lung tumors (n = 569) and adenocarcinoma cell lines (n = 31) expressing the ES phenotype were more likely to be resistant to cisplatin (P < 0.0001 and P = 0.045, respectively).

Conclusions: Lung adenocarcinomas that share a common gene expression pattern with normal human embryonic stem cells were associated with decreased survival, increased biological complexity, and increased likelihood of resistance to cisplatin. This indicates the aggressiveness of these tumors. (Clin Cancer Res 2009;15(24):7553-61)

Lung cancer remains the leading cause of cancer-related deaths in the United States with only a 15% five-year overall survival rate (1, 2). In addition, many patients that are diagnosed early and treated with standard cisplatin-based therapy still go on to have disease relapses (3). Research efforts continue in lung cancer treatment to improve the clinical outcomes of these patients.

Recently, attention has turned toward stem cells as a new frontier in the battle against cancer. The hallmark traits of stem cells—self-renewal and differentiation capacity—are mirrored by the high proliferative ability and phenotypic plasticity of cancer cells. There is also emerging evidence to suggest that signaling pathways that are critical to regulating the function of normal embryonic stem cells may also be active in tumors (4–7). It is unclear, however, whether these phenotypic similarities between normal stem cells and tumor cells are the result of a common biological phenotype(s), that is, changes in specific molecular pathways and regulatory networks. One approach to addressing this important question is to use genome-wide expression data to better dissect the biology driving the “stemness” of a tumor. Such studies may eventually lead to improved understanding of molecular mechanisms of disease progression and novel therapeutic options.

Seminal work from two laboratories has recently described specific gene sets representative of “embryonic stemness” (ES) in normal stem cells (8, 9). Ben-Porath et al. (8) identified gene sets associated with an embryonic stem cell identity that were overexpressed in poorly differentiated tumors and associated
Translational Relevance

Cancer cells and normal stem cells share certain qualities, such as self-renewal and multilineage differentiation, which would allow these cancer cells to be able to maintain tumor growth. We developed a 100-gene signature that would identify embryonic stem cell characteristics in lung adenocarcinoma that has prognostic and therapeutic implications. Lung adenocarcinomas that share a common gene expression pattern with normal embryonic stem cells were associated with decreased survival, increased activation of oncogenic pathways, and increased likelihood of resistance to cisplatin, indicating the aggressiveness of these tumors. Future steps should include evaluating normal stem cell and lung cancer stem cell cultures to further validate and manipulate the cancer biology/oncogenic pathways found to be activated in tumors with embryonic stemness. This may lead to new and effective therapeutic strategies.

Materials and Methods

Below is a brief description of the methods. Complete details are available in Supplementary Methods. Supplementary data files can be found online.6

**NSCLC cell lines.** Cells were grown according to media recommendations by the commercial vendor (American Type Culture Collection) with minimal modification. Total RNA was extracted from 56 cell lines using Qiagen RNeasy kits. Gene expression was measured using Affymetrix Human Genome U133A GeneChips.7

**Signature development.** Two thousand eighty-one genes were extracted from the Wong et al. data (courtesy of Dr. Howard Chang) and 7,602 genes were extracted from the Ben-Porath et al. data (courtesy of Dr. Ittai Ben-Porath). Duplicate genes were removed from each gene list. Gene expression values corresponding to the two sets of gene lists were extracted separately from the microarray data of the NSCLC cell lines to develop individual gene expression signatures. Gene filtering was done to isolate genes suspected of having the most effect. Intensity filtering was used to filter out genes with consistently low intensity values or low variance across samples. Interquartile range filtering was used to decrease the total number of genes by including genes with intensity values from the 25th percentile to the 75th percentile.

Briefly, the statistical approach in signature development first involves the identification of cell lines with two distinct classes/pheno-
types (ES and “no ES”). Gene expression data are obtained from the two classes of cell lines and using a Bayesian regression analysis, a predictor of ES is generated. The Bayesian model uses singular value decomposition and probit regression to build a gene signature from a training data set. First, the most dominant genes are selected, which show the greatest amount of correlation to the observed phenotype, as determined by Student’s t tests, and “metagenes” are defined as the top orthogonal factors (principal components) identified by singular value decomposition to capture the greatest amount of variance in gene expression within the collection of samples. This identifies the final genes to be included in a signature, or metagene, model. A Bayesian probit function is then applied to determine the probability that each sample has the ES phenotype. The model is then validated using leave-one-out cross-validation where the differentially expressed features that contribute to the metagene are properly reselected before fitting the Bayesian model. Subsequently, the final model is used to predict the probability of the ES phenotype in an independent data set. Further details are available in Supplementary Methods.

After analyzing the 56 NSCLC cell lines, independent signatures were developed, composed of data from 28 and 22 NSCLC cell lines for the Wong et al. and Ben-Porath et al. data, respectively. These smaller groups of cell lines, identified through unsupervised hierarchical clustering, showed biphasic gene expression patterns that allowed the genes from the Ben-Porath et al. and Wong et al. data to be used to divide the cells into those that showed ES and “no ES.”

The resulting signatures were individually evaluated for their ability to separate lung cancer data sets into patient samples that exhibited ES and “no ES.” Using independent samples to validate the model, CALEG 9761 (n = 82; n = 79; ref. 13), breast adenocarcinoma data set, these two signatures were combined. Cell lines with discordant gene expression patterns were removed. The remaining cell lines from both signatures underwent independent hierarchical clustering again to identify cell lines with expression patterns representing the extremes of the ES phenotype.

The dominant 100 genes in the ES signature (the “metagene”) were used to validate the signature. Three early-stage (I-IIIA) lung adenocarcinoma data sets were used: National Cancer Institute (NCI) Director’s Challenge Consortium (n = 442; ref. 11), Duke (GSE3151; n = 45; ref. 12), and CALGB 9761 (GSE3593; n = 82; ref. 10). A small cohort of patient samples from CALGB 9761 may overlap with samples from the NCI Director’s Challenge Consortium. A data set including 25 advanced-stage (IIb/IV) lung adenocarcinomas was also evaluated. The majority of these lung cancer patients were smokers. Three non-lung adenocarcinoma data sets were evaluated, including prostate adenocarcinoma (n = 79; ref. 13), breast adenocarcinoma (GSE2034; n = 286; ref. 14), and ovarian adenocarcinoma (n = 119; ref. 15). Normal lung epithelium from smokers was evaluated from Spira et al. (GSE4115; n = 90; ref. 16). Duke data set samples were obtained from patients enrolled in institutional review board-approved clinical trials through the Duke Lung Cancer Prognostic Laboratory.

The dominant 100 genes in the ES signature (the “metagene”) were applied to each validation data set using Bayesian regression methods (17, 18). This enabled the assignment of a probability value to individual samples reflecting an estimation (range, 0-1) of a given sample’s likelihood of having an ES phenotype. A predicted probability of ≥0.5 was chosen as a predetermined cutoff to identify those samples with the ES phenotype. The ES phenotype was then correlated with clinical survival data for patients using Kaplan-Meier curves to determine clinical relevance.

**Gene set analysis.** Analysis of Sample Set Enrichment Scores and Gene Set Enrichment Analysis were used to identify gene sets/pathways associated specifically with the 100 genes constituting the final ES signature. Analysis of Sample Set Enrichment Scores8 is a statistical method used to measure the degree of oncogenic pathway variation or gene

6 http://data.genome.duke.edu/Stemness
7 http://www.affymetrix.com/products
8 http://people.genome.duke.edu/~jh9g/assess/index.shtml


Downloaded from clincancerres.aacrjournals.org on October 31, 2017. © 2009 American Association for Cancer Research.
set enrichment for each sample in a gene expression data set (19, 20). It expands on Gene Set Enrichment Analysis9 (version 2.0), which is an analytic tool for relating gene expression data to gene sets to identify unifying biological themes (21, 22).

**Oncogenic pathway analysis.** Signatures of oncogenic pathway activation and tumor biology/microenvironment status were applied to each of the adenocarcinoma lung cancer tumor samples (10, 12, 23–26). Bayesian binary regression methodologies described previously (17, 18) were applied to estimate relative probabilities of pathway deregulation/activation or expression.

**Cisplatin sensitivity in lung tumors and cell lines.** A cisplatin sensitivity predictor was applied to each of the adenocarcinoma lung cancer tumor samples (17, 27). Bayesian binary regression methodologies described previously (17, 18) were applied to estimate relative probabilities of cisplatin sensitivity.

EC50 data for cisplatin was collected on 31 NSCLC cell lines (from American Type Culture Collection). Cells were grown in the presence of cisplatin for 96 h, and sensitivity to cisplatin was determined by quantifying the percent reduction in growth (versus DMSO controls) at 96 h using a standard MTT colorimetric assay (CellTiter 96 Aqueous One Solution Cell Proliferation Assay Kit; Promega; refs. 28, 29).

**Results**

**Development of an ES signature specific to lung adenocarcinoma.** Genes were extracted from the Wong et al. (ref. 9; courtesy of Dr. Howard Chang) and the Ben-Porath et al. (ref. 8; courtesy of Dr. Ittai Ben-Porath) data to develop a signature specific to lung adenocarcinoma (Fig. 1). These gene sets were then used to analyze 56 NSCLC (adenocarcinoma) cell lines (Fig. 2). First, we tested the robustness of the Wong et al. and Ben-Porath et al. gene sets individually; thus, gene expression signatures were developed independently from each gene set (17, 18) that were able to divide the lung cancer cell lines into two distinct groups: cell lines with and without ES. Further details on the signature development are included in Materials and Methods. The prognostic ability of the two signatures was evaluated in a pilot cohort of...
early-stage (I-IIIa) lung adenocarcinoma samples obtained from a multicenter cooperative group trial (CALGB 9761; n = 82; ref. 10). It is important to note that because our classifiers were built from only lung adenocarcinoma cell lines, we tested the clinical relevance in a patient cohort that was exclusive to lung adenocarcinoma. It was found that patients with tumors with an ES designation had a poorer survival compared with those without ES (Ben-Porath et al. survival $P = 0.03$ and Wong et al. survival $P = 0.06$; Supplementary Fig. S1).

With an underlying hypothesis that optimizing the strengths of the Wong et al. and the Ben-Porath et al. approaches into a unified signature would lead to a more robust clinically valid representation of the ES phenotype, we proceeded to generate a gene expression signature representative of ES using a meta-gene approach (10, 17). Importantly, the individual signatures were combined after removing cell lines with discordant gene expression patterns (Fig. 2). A 100-gene ES signature derived from 20 lung adenocarcinoma cell lines was created, and the intrinsic robustness was verified in a leave-one-out cross-validation. This final “hybrid” signature was used as the classifier for further analysis.

The ES signature is composed of several genes necessary for cell cycle regulation, cell proliferation, angiogenesis, apoptosis regulation, and gene transcription (Supplementary Table S1). Genes highly upregulated in the signature are known to be involved in tumor invasion and cell proliferation, such as RAB25 and MAPK13. These genes were compared with the genes included in other prognostic lung cancer signatures, such as Beer et al. (30), Chen et al. (31), Potti et al. (10), Bhattacharjee et al. (32), and Lu et al. (ref. 33; Supplementary Table S2). Although the
Fig. 3. **A**, Analysis of Sample Set Enrichment Scores and Gene Set Enrichment Analysis were used to identify gene sets/pathways associated with cells lines with and without ES. Certain biologically relevant processes are highlighted. NES, no ES. **B**, oncogenic pathway deregulation shown as a function of ES in lung adenocarcinoma samples (n = 569). The distribution of samples from each data set is shown above the heat map. Red on the heat map indicates pathway deregulation/activation and blue on the heat map indicates normal activity. Bottom, probability of oncogenic pathway deregulation. WH, wound healing; CIN, chromosomal instability; IGS, invasiveness; BCAT, β-catenin; TNF, tumor necrosis factor.
individual genes are different among the listed signatures, there is significant overlap at a functional/pathway level.

**Biology of ES in lung adenocarcinoma.** To deepen our investigation of the biological underpinnings of the ES phenotype, we applied previously validated approaches, including Gene Set Enrichment Analysis (21, 22) and Analysis of Sample Set Enrichment Scores (19, 20), to cell lines that constitute our 100-gene lung cancer–specific classifier. Elucidation of patterns of pathway activation specific to ES in lung adenocarcinoma revealed numerous biologically relevant features (Fig. 3A). For example, the most striking characteristics in the stemness cohort included gene sets/pathways associated with neoplastic transformation in normal tissues, cancer invasion, and metastasis (12, 34, 35). Specifically, the Ras pathway was associated with the ES phenotype.

In addition, previously described gene signatures that delineate the activity of several cancer cell biology and oncogenic pathways (12, 23–26) were applied to 569 samples from patients with early-stage lung adenocarcinoma. Oncogenic pathways and cellular processes that were significantly associated with increased deregulation in adenocarcinomas with ES (Fig. 3B) included Ras ($P = 0.0005$), Myc ($P = 0.0224$), wound healing ($P < 0.0001$), chromosomal instability ($P < 0.0001$), and invasiveness ($P < 0.0001$) gene signatures. Src, β-catenin, tumor necrosis factor, and E2F pathways were not significantly different. These findings are not completely surprising because the embryonic signature was found to be enriched with gene sets involved in cancer invasiveness, metastasis, and the Ras pathway. The findings from the Analysis of Sample Set Enrichment Scores/Gene Set Enrichment Analysis analyses and oncogenic pathway evaluation confirm the biological complexity associated with ES in cancer at a pathway level.

**ES and prognosis in lung adenocarcinoma.** As an initial measure of clinical relevance, the probability of having an ES phenotype, as determined by the 100-gene signature, was evaluated in smokers with normal lung epithelium ($n = 90$; ref. 16) compared with patients with early-stage (I, II, and IIIa; $n = 569$) and advanced-stage (IIIb and IV; $n = 25$) lung adenocarcinoma. This showed a significant difference in probability of ES for smokers without cancer when compared with patients with early- or advanced-stage lung adenocarcinoma in a stepwise fashion ($P < 0.0001$; Fig. 4A). In addition, patients with poorly ($n = 166$), moderately ($n = 209$), and well ($n = 60$) differentiated adenocarcinomas also showed significant differences in probability of ES ($P < 0.0001$; Fig. 4A), with poorly differentiated tumors having the highest probability of ES. This result provides preliminary evidence that ES may be associated with lung cancer invasion and progression.

To further characterize the clinical relevance of the ES phenotype, we analyzed patients with early-stage lung adenocarcinoma. Using survival as a measurable endpoint, the probability
of ES in patient tumor samples from these three data sets [CALGB 9761 (n = 82; ref. 10), NCI Director's Challenge Consortium (n = 442; ref. 11), and Duke (n = 45; ref. 12)] was plotted as a function of survival (Fig. 4B). These patients had not received adjuvant chemotherapy. Kaplan-Meier survival analysis showed that patient samples designated as having ES were associated with a poorer survival compared with those without ES in two of the data sets (CALGB 9761 P = 0.0001 and NCI Director's Challenge Consortium P = 0.0002), with a trend toward poorer survival in the third data set (Duke P = 0.06).

Ben-Porath et al. and Wong et al. showed previously that the genes they identified as associated with an ES phenotype were not predominately related to cell proliferation. To show that the ES signature goes beyond just genes involved in cell cycle and cell proliferation regulation, these genes were removed and the signature was reanalyzed. Repeat evaluation of the

Fig. 5. A, probability of sensitivity to cisplatin was analyzed as a function of ES in lung adenocarcinoma samples (n = 569). Red on the heat map indicates cisplatin sensitivity and blue on the heat map indicates cisplatin resistance. B, cisplatin sensitivity for NSCLC cell lines as a function of ES. Sensitivity to cisplatin was delineated from EC50 assays. Red on the heat map indicates cisplatin sensitivity and blue on the heat map indicates cisplatin resistance.
Numerous examples now can be found where gene expression profiles have been able to dissect lung cancer subtypes (30–32). The use of gene expression signatures as surrogate phenotypes has been particularly important, outlining the complexity of biological systems in a way that was not previously feasible (10, 37, 40, 41). However, in lung cancer, studies addressing the potential effect of specific stem cell regulatory networks in defining clinically relevant cancer phenotypes have been limited.

Cancer cells and normal stem cells share certain qualities, such as self-renewal and multilineage differentiation, which would allow these cancer cells to be able to maintain tumor growth. This has sparked interest in exploring whether these shared qualities are due to shared molecular pathways between cancer cells and stem cells versus due solely to a small population of cancer stem cells that drive the behavior of a tumor. Further investigation is thus warranted to better understand the biological and clinical relevance of “stem cell-like” characteristics in cancers, specifically lung cancer.

Our findings describe a 100-gene signature representative of ES specific to lung adenocarcinoma. Preliminary findings suggest that merely the presence of genes regulating cell proliferation do not portend clinical relevance or predict a stem cell phenotype at the gene expression level (8, 9). In fact, in our analysis, removing proliferation-associated genes from the ES signature did not affect its prognostic ability to identify patients at risk for poor survival. The finding of biological complexity at a pathway level, as evidenced by activation of Ras, Myc, and chromosomal instability in the stemness phenotype, might be expected considering the fact that this phenotype also has adverse prognostic and therapeutic implications shown by its association with poor survival, poorly differentiated cancers, and resistance to cisplatin-based chemotherapy.

It is, however, important to note here that the goals of our current study are not to develop yet another gene signature that predicts lung cancer recurrence and/or sensitivity to chemotherapy. There are several other predictors that are likely more robust in addressing these questions more directly (10, 11, 42, 43). We used disease recurrence/survival and sensitivity to cisplatin merely to depict the clinical relevance of studying stem–associated phenotypes relevant to lung adenocarcinoma.

Our results need to be considered in light of the fact that stem cells of all types (normal adult stem cells, embryonic stem cells, and cancer stem cells) are still poorly characterized. Our ability to isolate these cells and ensure cell purity is limited, specifically in lung cancer. As an alternative, the embryonic stem cell phenotype we describe here may carry meaningful clinical relevance and be representative of an “inherited” behavior in lung adenocarcinoma as it evolves in the process of differentiation.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References


Retraction: Characterizing the Clinical Relevance of an Embryonic Stem Cell Phenotype in Lung Adenocarcinoma

The authors wish to retract the article titled "Characterizing the Clinical Relevance of an Embryonic Stem Cell Phenotype in Lung Adenocarcinoma," which was published in the December 15, 2009, issue of Clinical Cancer Research (1).

It has recently come to our attention that clinical information from a data set CALGB 9761 (GSE3593), available at the time of the signature development, was incorrect. Since survival data from this data set were integral to the creation of this signature, this makes the findings of this manuscript inaccurate. Drs. Anil Potti and Marvaretta Stevenson take full responsibility for this error. Therefore, we would like to formally retract the above paper.

Marvaretta Stevenson
William Mostertz
Chaitanya R. Acharya
William Kim
Kelli Walters
William Barry
Kristin Higgins
Sascha A. Tuchman
Jeffrey Crawford
Gordana Vlahovic
Neal Ready
Mark Onaitis
Anil Potti

Duke University
Durham, North Carolina

Reference

Characterizing the Clinical Relevance of an Embryonic Stem Cell Phenotype in Lung Adenocarcinoma

Marvaretta Stevenson, William Mostertz, Chaitanya Acharya, et al.


Updated version

Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-09-1939

Supplementary Material

Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2009/12/21/1078-0432.CCR-09-1939.DC1

Cited articles

This article cites 42 articles, 14 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/15/24/7553.full#ref-list-1

Citing articles

This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/15/24/7553.full#related-urls

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, contact the AACR Publications Department at permissions@aacr.org.