Published in final edited form as:

Ann Thorac Surg. 2011 April ; 91(4): 1046–1050. doi:10.1016/j.athoracsur.2010.12.054.

A 10-Gene Progenitor Cell Signature Predicts Poor Prognosis in Lung Adenocarcinoma

Mark Onaitis, MD¹, Thomas A. D'Amico, MD¹, Cheryl P. Clark, BS², Justin Guinney, PhD², David H. Harpole, MD¹, and Emma L. Rawlins, PhD³

¹Duke University Department of Surgery, Seattle, WA

²Duke University Department of Cell Biology, Seattle, WA

³Fred Hutchinson Cancer Center, Seattle, WA

⁴Gurdon Institute, University of Cambridge

Abstract

Purpose—One aspect of the cancer stem cell hypothesis is that patients with tumors that exhibit stem-like phenotypes have poor prognoses. We sought to create a lung-specific progenitor cell signature for possible prognosis prediction in human lung cancer. Distal epithelial progenitors from lungs early in development demonstrate both self-renewal and potential to differentiate into all bronchial and alveolar epithelial cell types. By contrast, late progenitors are only able to produce alveolar cells. A “stem-like” signature produced by comparing the transcriptome of these early and late progenitors was then applied to microarray data from resected lung adenocarcinomas.

Methods—A transgenic mouse was created in which embryonic distal epithelial progenitor cells express green fluorescent protein when tamoxifen is administered. Lung progenitor cells were harvested after tamoxifen injection at either embryonic day 11.5 (E11.5) or 16.5 (E16.5). RNA extracted from these cells was hybridized to Affymetrix 430.2 mouse chips. A “stem-like” signature was created by comparing the cell types using L1 logistic regression and applied to transcriptome datasets of resected patients from our tumor bank and the NIH Director's Challenge Consortium.

Results—When a 10-gene “stem-like” signature was applied to resected human adenocarcinoma datasets, tumors which were transcriptionally similar to the early progenitors (more “stem-like”), had a significantly worse prognosis than those similar to the late progenitors (top $p=.009$; bottom $p<.0001$ by logrank). Using a Cox model in which age and stage were included, the predicted score from the logistic regression model was an independent predictor of survival.

Conclusions—A genomic signature of “stem-like” phenotypes predicts poor prognosis in lung adenocarcinoma. Modulation of these genes or their signaling pathways may be effective therapeutic strategies in the future.

Keywords

Progenitor cell; lung cancer

Corresponding author: Mark Onaitis MD, DUMC Box 3305, Durham, NC 27710, Phone (919) 684-6974, Fax (919) 684-8508, mark.onaitis@duke.edu.

To be presented at Society of Thoracic Surgeons annual meeting January 22-27, 2010

Introduction

Generally-accepted characteristics of stem cells include the ability to self-renew and the capacity to differentiate into multiple lineages. Other phenotypes of these cells often include quiescence, mutation resistance, and the ability to perform divide asymmetrically. Progenitor cells display many of the same characteristics as stem cells, but they have a more limited ability to self-renew. Progenitor cells can either be the immediate daughters of adult stem cells, or cells which act as stem cells during organogenesis, but are not retained in the adult. Normal tissue stem cells and cancer cells share many phenotypic characteristics including the ability to self-renew and differentiate into multiple types of progeny. The cancer stem cell hypothesis states that tumors arise from cells that possess these properties and are thus able to initiate new tumors from a single cell[1]. A corollary of this hypothesis is that tumors that contain high percentages of these cells will be most aggressive. However, the existence of such a cancer stem cell population in non-small cell lung cancer has not been proven. Recently, several groups have applied the genomic signatures of stem and progenitor populations to human lung cancer microarray data for prognosis prediction[2, 3]. These studies have used embryonic stem cell signatures that are not lung-specific.

Lung development proceeds through four histologic stages: pseudoglandular, canalicular, terminal sac, and alveolar. Recently, the *Inhibitor of differentiation (Id2)*-expressing distal tip epithelial cells of the developing mouse lung have been shown to be both multipotent (able to contribute descendants to all of the different lung epithelial cell lineages) and to self-renew. Using a transgenic mouse in which distal tip epithelial cells expressing *Id2* were induced to express fluorescent markers, E11.5 distal tip cells from the pseudoglandular stage were demonstrated to give rise to all bronchiolar and alveolar epithelial cells, whereas E17.5 distal tip cells from the canalicular stage could only give rise to alveolar epithelial cells[4]. These *Id2*⁺ distal tip cells are a lung-specific embryonic progenitor population.

We undertook the present study to assess whether microarray information from lung-specific embryonic progenitor cells would provide prognostic information for human lung cancer patients. We hypothesized that a “signature” comprised of genes that are significantly differently expressed between multipotent pseudoglandular distal tip cells and oligopotent canalicular distal tip cells would distinguish between groups of human lung cancer patients with different prognoses.

Patients and Methods

Mouse Cells and Microarrays

Pregnant female *Id2-Cre-estrogen receptor (Id2CreER)-Rosa26-lox-stop-lox farnesylated enhanced green fluorescent protein* mice were injected with tamoxifen at gestational days 11.5 and 16.5 as described in Rawlins et al [4, 5]. Lungs from the embryos were isolated, dissected and dissociated one day after tamoxifen injection. The distal tip cells at embryonic day 17.5 were isolated by fluorescence activated cell sorting (FACS). The distal tip cells at embryonic day 11.5 were isolated by microdissection. Because the E11.5 cells were collected by microdissection, the cell population probably also included small amounts of blood vessels. In each case the cells were lysed for isolation of RNA using the Qiagen RNEasy Mini Kit. The RNA was hybridized to Affymetrix mouse 430.2 microarray chips (5 chips per condition) and the resulting CEL files were normalized by robust multichip averaging. The mouse probesets IDs were translated to human probeset IDs using ChipComparer software (<http://chipcomparer.genome.duke.edu>).

Human Lung Cancer Microarray Data

Two human lung cancer microarray datasets were created by importing the CEL files from GEO corresponding to the data from the Potti et al[6] and Shedden et al[7] manuscripts. These datasets were also normalized using robust multichip averaging.

Statistical Analysis

Logistic regression with L1 regularization was used to identify a sparse set of genes capable of differentiating between the early and late distal tip progenitor cells. This procedure uses a statistical algorithm to simplify a large amount of data into the most predictive variables. Due to the small number of samples, the regularization parameter was chosen in advance to select a sparse model to avoid over-fitting and to provide focus on strong candidate genes. Applying this model to the mouse expression data resulted in selection of a ten-gene logistic model. Kaplan-Meier survival curves were created for each human dataset with patients stratified by the median of the predicted score (~.5); the Mantel-Cox (logrank) test was used to assess survival differences. To assess significance of the genomic predictor adjusting for tumor stage and patient age, a Cox regression model of survival was also applied to the human lung data.

Results

Generation of a ten-gene signature for late distal tip epithelial progenitor cells

A schematic for the experiment is depicted in Figure 1. After isolation of early (E11.5) and late (E17.5) Id2-positive distal tip progenitor cells, isolation of RNA, and hybridization to Affymetrix mouse microarray chips, the gene expression data was analyzed using L1 lasso logistic regression. The top ten genes that are differentially expressed between the early and late progenitor cells are listed in Table 1. Interestingly, all ten genes had higher expression in the late progenitors.

Application of the ten-gene signature to human resected lung cancer datasets

The ten-gene signature was then applied first to a dataset of patients whose tumors were resected at Duke University. This dataset contains 91 patients, of whom 45 were histologically squamous cell carcinoma and 46 were histologically adenocarcinoma. As Figure 2A demonstrates, those patients with tumors transcriptionally similar to the late progenitors enjoy a statistically-significant survival advantage. The ten-gene signature was then applied to the large Director's Challenge dataset of resected adenocarcinoma patients. As Figure 2B reveals, patients with tumors transcriptionally similar to late progenitors have statistically-significantly improved survival when compared to those with tumors similar to early progenitors.

Multivariable Survival Analysis

In order to ascertain whether the ten-gene signature prediction score independently predicts lung cancer patient survival, a multivariable Cox model was performed using the prediction score, age, and pathologic stage. As Table 2 demonstrates, even when controlling for age and stage the genomic predictor continues to predict survival in these Director's Challenge patients.

Comment

The current study tests whether a genomic "signature" of lung progenitor cells at more differentiated versus less differentiated stages derived from transgenic mouse cells predicts prognosis among resected non-small cell lung cancer patients. We found that the patients

with tumors more similar to the late (E17.5) progenitors had a survival advantage. We hypothesize that the tumors of the other patients were more similar to the early (E11.5) progenitors, and therefore more ‘stem-like’, and had greater ability to self-renew. However, we cannot rule out that the ten genes contribute to other phenotypes, or to differentiation itself. If this is the case, our signature may identify tumors that are better differentiated. Including pathologic tumor differentiation status is planned for future multivariable survival models.

Prognosis prediction is not the main goal of this sort of analysis. The purpose of this type of approach is to identify potential genes that contribute to phenotypes shared by both tumors and progenitor cells. Therapeutic approaches that target these genes and pathways may then ultimately be developed and employed.

Direct experiments testing the cancer stem cell hypothesis have used complex cell-sorting techniques to identify populations of tumor cells that can recapitulate the tumor when transplanted into immunodeficient mice[8-15]. However, several groups have identified problems with this type of study. These include mathematical arguments identifying inconsistencies between percentages of cancer stem cells in primary tumors, percentages of cancer stem cells obtained via flow cytometry, and expected numbers of cells and sizes of recurrent tumors[16]. Also, studies performed by transferring tumor cells from mouse tumors to other mice reveal no growth advantage for any subpopulation of cells, raising the possibility that cross-species effects explain the experimental findings[17].

Stem and progenitor cell populations are only beginning to be understood on a genetic level. As important genes and pathways contributing to stem-like phenotypes are dissected further, possible further links to cancer phenotypes may be identified. Our study serves as an initial proof-of-principle that identification of precise mouse phenotypes followed by microarray analysis, translation of probeset IDs to human, and application to human cancers for prognosis prediction is possible.

References

1. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature*. 2001; 414:105–11. [PubMed: 11689955]
2. Hassan KA, Chen G, Kalemkerian GP, Wicha MS, Beer DG. An embryonic stem cell-like signature identifies poorly differentiated lung adenocarcinoma but not squamous cell carcinoma. *Clin Cancer Res*. 2009; 15(20):6386–90. [PubMed: 19808871]
3. Stevenson M, Mostertz W, Acharya C, et al. Characterizing the Clinical Relevance of an Embryonic Stem Cell Phenotype in Lung Adenocarcinoma. *Clin Cancer Res*. 2009; 15(24):7553–7561. [PubMed: 19996213]
4. Rawlins EL, Clark CP, Xue Y, Hogan BL. The Id2+ distal tip lung epithelium contains individual multipotent embryonic progenitor cells. *Development*. 2009; 136(22):3741–5. [PubMed: 19855016]
5. Rawlins EL, Okubo T, Xue Y, et al. The role of Scgb1a1+ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell*. 2009; 4(6):425–34.
6. Potti A, Mukherjee S, Petersen R, et al. A genomic strategy to refine prognosis in early-stage small cell lung cancer. *New England Journal of Medicine*. 2006; 355(6):570–80. [PubMed: 16899777]
7. Director’s Challenge Consortium for the Molecular Classification of Lung, A. Shedden K, Taylor JM, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nature Medicine*. 2008; 14(8):822–7. see comment.
8. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *PNAS*. 2003; 100(7):3983–8. [PubMed: 12629218]
9. Bao S, Wu Q, McLendon RE, et al. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature*. 2006; 444(7):756–60. [PubMed: 17051156]

10. Li C, Heidt DG, Dalerba P, et al. Identification of pancreatic cancer stem cells. *Cancer Res.* 2007; 67(3):1030–7. [PubMed: 17283135]
11. Prince ME, Sivanandan R, Kaczorowski A, et al. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *PNAS.* 2007; 104(3):973–8. [PubMed: 17210912]
12. Ricci-Vitiani L, Lombardi DG, Pilozzi E, et al. Identification and expansion of human colon-cancer-initiating cells. *Nature.* 2007; 445(4):111–5. [PubMed: 17122771]
13. Singh SK, Hawkins C, Clarke ID, et al. Identification of human brain tumour initiating cells. *Nature.* 2004; 432:396–401. [PubMed: 15549107]
14. Wang JCY, Dick JE. Cancer stem cells: lessons from leukemia. *Trends in Cell Biology.* 2005; 15:494–501. [PubMed: 16084092]
15. Zucchi I, Sanzone S, Astigiano S, et al. The properties of a breast cancer stem cell. *PNAS.* 2007; 104(25):10476–81. [PubMed: 17566110]
16. Kern SE, Shibata D. The fuzzy math of solid tumor stem cells: a perspective. *Cancer Research.* 2007; 67(19):8985–8. [PubMed: 17908998]
17. Kelly PN, Dakic A, Adams JM, Nutt SL, Strasser A. Tumor growth need not be driven by rare cancer stem cells. *Science.* 2007; 317(5836):337. see comment. [PubMed: 17641192]

Figure 1. Schematic of the mouse cell isolation, RNA extraction, hybridization to Affymetrix 430.2 microarray chips, and statistical analysis.

Figure 2.

Kaplan-Meier survival curves of (A) 91 non-small cell lung cancer patients resected at Duke and (B) 469 patients from the Director's Challenge dataset described by Shedden et al. stratified by the 10-gene signature score. In each case, the patients with tumors transcriptionally similar to late progenitors have a significantly better prognosis than those patients with tumors transcriptionally similar to the early progenitors.

Table 1

The ten genes in the signature. Probeset IDs are human Affymetrix identifiers.

Probeset ID	Symbol	Gene name
202295_s_at	<i>CTSH</i>	Cathepsin H
205982_x_at	<i>SFTPC</i>	Surfactant protein C
204039_at	<i>CEBPA</i>	CCAAT.enhancer binding protein (C/EBP), alpha
202768_at	<i>FOSB</i>	Murine osteosarcoma viral oncogene homolog B
211663_x_at	<i>PTGDS</i>	Prostaglandin D2 synthase 21kD
209679_s_at	<i>LOC57228</i>	Small transmembrane and glycosylated protein
209619_at	<i>CD74</i>	CD74 molecule
219476_at	<i>C1orf116</i>	Chromosome 1 open reading frame 116
209788_s_at	<i>ARTS-1</i>	Type I tumor necrosis factor receptor shedding aminopeptidase
213281_at	<i>JUN</i>	V-jun sarcoma virus 17 oncogene homolog

Table 2

Results of a multivariable survival analysis in which age, stage, and the 10-gene genomic signature score were included. Each of these variables predicts patient prognosis independently of the others.

Variable	Odds Ratio	p Value
Genomic Predictor	3.55	0.0004
Age	4.83	<.0001
Stage II (vs. I)	5.71	<.0001
Stage III (vs. I)	9.65	<.0001