

Upregulation of RNA Processing Factors in Poorly Differentiated Lung Cancer Cells

Kenneth G. Geles^{*,1}, Wenyan Zhong^{*,1},
Siobhan K. O'Brien^{*,2}, Michelle Baxter^{*,3},
Christine Loreth^{*}, Diego Pallares[†] and Marc Damelin^{*}

^{*}Pfizer Inc., Oncology-Rinat Research & Development, 401 N. Middletown Rd., Pearl River, NY 10965 USA

[†]Diaxonhit, 65 Boulevard Massena, 75013 Paris France

Abstract

Intratumoral heterogeneity in non-small cell lung cancer (NSCLC) has been appreciated at the histological and cellular levels, but the association of less differentiated pathology with poor clinical outcome is not understood at the molecular level. Gene expression profiling of intact human tumors fails to reveal the molecular nature of functionally distinct epithelial cell subpopulations, in particular the tumor cells that fuel tumor growth, metastasis, and disease relapse. We generated primary serum-free cultures of NSCLC and then exposed them to conditions known to promote differentiation: the air-liquid interface (ALI) and serum. The transcriptional network of the primary cultures was associated with stem cells, indicating a poorly differentiated state, and worse overall survival of NSCLC patients. Strikingly, the overexpression of RNA splicing and processing factors was a prominent feature of the poorly differentiated cells and was also observed in clinical datasets. A genome-wide analysis of splice isoform expression revealed many alternative splicing events that were specific to the differentiation state of the cells, including an unexpectedly high frequency of events on chromosome 19. The poorly differentiated cells exhibited alternative splicing in many genes associated with tumor progression, as exemplified by the preferential expression of the short isoform of telomeric repeat-binding factor 1 (TERF1), also known as Pin2. Our findings demonstrate the utility of the ALI method for probing the molecular mechanisms that underlie NSCLC pathogenesis and provide novel insight into posttranscriptional mechanisms in poorly differentiated lung cancer cells.

Translational Oncology (2016) 9, 89–98

Introduction

Non-small cell lung cancer (NSCLC) remains the leading cause of cancer-related deaths despite the successful development of many therapies. Less differentiated NSCLC tumors are associated with increased risk of death and increased risk of recurrence following tumor resection [1]. Poorly differentiated and dedifferentiated cancer cells can exhibit stem cell-like properties, including multipotency, tumor-initiating capacity, and the expression of stem cell factors.

The isolation and characterization of tumor cell subpopulations by fluorescence-activated cell sorting are commonly used to generate gene expression profiles of more aggressive populations, but the liabilities associated with this approach include the need for dissociation of tissue to single cells and the use of immune-compromised mice to define the tumorigenic populations [2] — not by differentiation status, which is the basis of the clinical analyses. As a complementary approach that circumvents both of these conditions,

we characterized lung cancer cells with distinct differentiation states by establishing primary serum-free NSCLC cultures and then inducing differentiation by various methods.

Address all correspondence to: Marc Damelin, Ph.D., Pfizer, Inc., 401 N. Middletown Rd., 200-4611, Pearl River, NY 10965, USA.

E-mail: marc.damelin@pfizer.com

¹ Equal contributions.

² Present address: Prolong Therapeutics, South Plainfield, NJ.

³ Present address: Regeneron, Inc., Tarrytown NY.

Received 5 November 2015; Revised 11 January 2016; Accepted 19 January 2016

© 2016 The Authors. Published by Elsevier Inc. on behalf of Neoplasia Press, Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1936-5233/16

<http://dx.doi.org/10.1016/j.tranon.2016.01.006>

Faithful preservation of the undifferentiated state in culture without the acquisition of genetic or epigenetic alterations requires sophisticated techniques such as defined serum-free media and three-dimensional (3D) matrices [3,4]. A relevant culture method to study the developmental hierarchy of normal and cancerous lung cells is the air-liquid interface (ALI), in which cells are induced to differentiate in 3D growth by exposure to the air. The ALI method has been used to investigate gene networks, differentiation potential, and *in vivo* relevance of normal human bronchial epithelial cells (HBECs) [5,6] and subsequently adapted to study the cellular hierarchy of a primary NSCLC culture [7]. However, the direct comparison of ALI to conventional methods such as serum has not been conducted in NSCLC but would inform the relevant approaches to characterize the gene expression and mRNA splicing profiles of the poorly differentiated state.

The dysregulation of mRNA splicing has been implicated in tumorigenesis and therapeutic resistance [8–10], but genome-wide analysis of mRNA splicing has not been broadly applied to the study of poorly differentiated cancer cells. The differential expression of mRNA splice isoforms of various genes regulates differentiation and pluripotency in normal tissue stem cells [11,12], but its prevalence and functions in poorly differentiated cancer cells are not well characterized.

We demonstrate that poorly differentiated NSCLC cells overexpress RNA processing factors and exhibit a distinct profile of alternatively spliced mRNAs. As a striking example, the short isoform of the telomere factor *TERF1* is enriched in the poorly differentiated state. The clinical relevance of the ALI method for studying NSCLC was revealed by comparison to patient datasets and generation of gene signatures that predict survival outcomes. These results provide insight at the molecular level into the correlation between poorly differentiated tumor histology and worse clinical outcome.

Materials and Methods

Serum-Free Culture of Human NSCLC Cells

Human lung tumor tissue was obtained from Asterand in accordance with appropriate consent procedures, and cell cultures were established on Matrigel (BD Biosciences)-coated flasks in serum-free bronchial epithelial cell growth medium (BEGM; Lonza) as described [7]. Matrigel was diluted in PBS to 0.4 mg/ml, transferred to culture flasks, and allowed to solidify overnight at room temperature before adding cells. TUM449 cell line was generated from patient tumor sample 87449A1, an NSCLC adenocarcinoma, and TUM110 cell line was generated from patient sample 60110A1, an NSCLC squamous cell carcinoma.

Air-Liquid Interface Cultures

Millicell 1- μ m PET hanging cell culture inserts (Millipore) lacking Matrigel were seeded with 1.5×10^5 cells in BEGM. After 2 to 3 days, the upper and lower chambers were rinsed with PBS, and CnT23 medium (CELLnTec) containing 50 nmol/l of retinoic acid, 1 mmol/l of CaCl_2 , or BEGM was added only to the lower chamber, leaving cells exposed to the air (5% CO_2 in a tissue culture incubator). Lifted cultures were fed every other day with fresh medium. For serum differentiation, cells were seeded in BEGM onto multiwell plates lacking Matrigel and switched to Ham's F-12K (Kaighn's) medium containing 10% fetal bovine serum (FBS).

Alternative Splicing

SpliceArray™ microarrays (Diagonhit) contain body probes that monitor exon bodies and are designated F, T, and B, and junction probes that monitor exon-exon junctions are designated C, D, and E. Preprocessing steps including robust multi-array average background correction, quantile normalization, and mean probe summarization were performed on the CEL files using Partek software (Partek Inc., St. Louis, MO). Microarray processing methods have been published [13] and are described in Supplemental Information. Gene level expression from SpliceArray™ was calculated as the geometric mean of all F and T probes, whereas gene level expression from Affymetrix platform was summarized as median of all probe sets. Correlation analysis of gene level expression change between differentiated and undifferentiated condition (\log_2 ratio) was performed using Pearson method. Event level change for each alternatively spliced event was calculated as the ratio of geometric means between probes that target the variant transcript to the reference transcript (B, C, D, and E). Differential event level change between the differentiated and undifferentiated conditions was analyzed by analysis of variance method using Partek software (Partek Inc). Significantly altered events are defined as those that with a fold-change ≥ 2 and false discovery rate (FDR) ≤ 0.05 in both cultures. The RNAseq method was not feasible in this scenario because of limited quantities of RNA.

TERF1 Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

Short- and long-form *TERF1* transcripts were detected by PCR with splice variant specific primer sets previously developed [14]. PCRs were run on an Applied Biosystems Veriti thermocycler. PCR products were resolved on 20% polyacrylamide gels, stained with Sybr Green (Life Technologies), and imaged with the ChemiDoc XRS. Band intensities were quantitated using ImageJ software.

Results

Gene Signatures of Poorly Differentiated NSCLC Cells

To extend our initial study of NSCLC at the ALI [7], we established and maintained two additional NSCLC cultures, TUM110 and TUM449, in serum-free BEGM and induced the cells to differentiate in three conditions (Figure 1A). The TUM110 culture was established from a female patient diagnosed with poorly differentiated squamous cell carcinoma of the lung (Union for International Cancer Control stage T2N0MX), and the TUM449 culture was from a female patient diagnosed with moderately differentiated adenocarcinoma of the lung (Union for International Cancer Control stage T2N0MX). In one differentiation condition, “BEGM-ALI,” the cells were exposed to the ALI in standard BEGM medium, which is similar to the medium used for ALI cultures of HBECs [6,15]. In another condition, “CnT23-ALI,” which was used in the original study [7], cells were exposed to the ALI in protein-free CnT23 medium, which contains 50 nM retinoic acid and 1 mM CaCl_2 . In the third condition, “Serum,” cells were cultured in standard tissue-culture plates submerged in medium that contains 10% FBS; this condition has long been known to induce differentiation in HBECs [16].

Upon exposure to the ALI, the tumor cells formed a 3D multilayered epithelium (Figure 1B), consistent with previous reports [7,17]. ALI samples were harvested on day 16 for gene expression profiling based on our previous studies and similarity to published

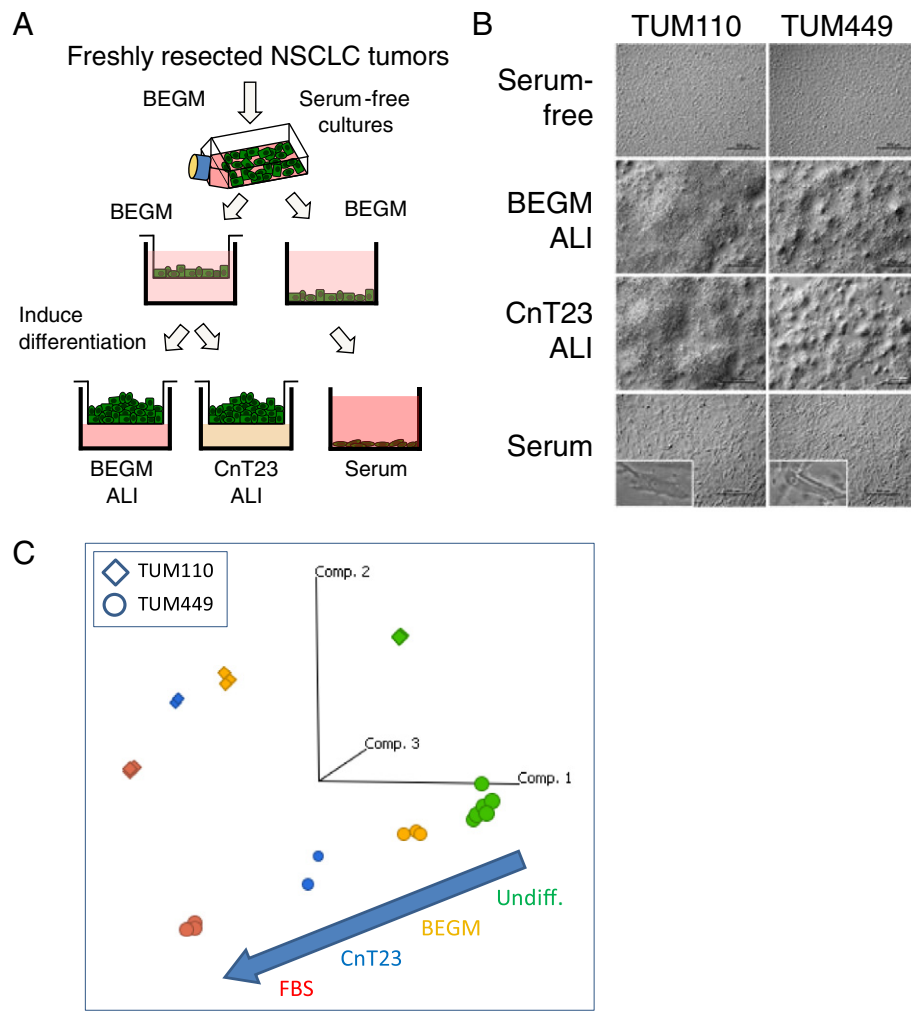


Figure 1. Induction of differentiation of primary NSCLC cultures. (A) Serum-free NSCLC cultures were induced to differentiate at the ALI in serum-free BEGM or protein-free CnT23 media for 16 days, or by submersion in 10% serum-containing media for 4 days. (B) Relief phase contrast images of cultures before and after differentiation. Cells exposed to the ALI formed 3D structures, whereas cells exposed to serum remained a monolayer. Scale bar = 500 μm. Insets, high-magnification phase contrast images of cells after serum differentiation. (C) Principal component analysis of gene expression profiles indicated the parallel nature of TUM110 and TUM449 across the four conditions (arrow) and highlighted the extreme effects of serum. Each dot or diamond corresponds to one sample replicate.

HBEC studies that support a lack of significant changes in gene expression after 8 to 10 days [4,6,7]. Cells that were exposed to serum remained in monolayer but exhibited significant morphological changes, such as a spindly morphology (Figure 1B). Serum samples were harvested on day 4 because they still appear to look healthy on that day and time courses had shown that the cells do not survive 16 days in serum. Gene expression profiles of three biological replicates per condition were generated with the Affymetrix U133Plus microarray platform. Principal component analysis revealed prominent patterns (Figure 1C). In both cultures, the BEGM-ALI profile was located closest to that of the poorly differentiated profile, followed by CnT23-ALI and then serum. The arrangement of the samples indicated that differentiation induced similar overall changes in transcription in TUM110 and TUM449. Similar patterns were observed by hierarchical clustering (Supplemental Figure S14). The analysis suggested distinct molecular outcomes of the three conditions, especially serum; these differences were subsequently explored by various bioinformatic analyses.

A gene signature termed GeneSig2 was defined for each condition as the group of probe sets with at least two-fold lower expression (FDR ≤ 0.01) in the differentiated state in both cultures (Supplemental Table S1). Interestingly, each GeneSig2 contained several genes that encode RNA processing factors; subsequent analysis demonstrated the highly significant enrichment of these factors in the poorly differentiated state (see below). We compared each GeneSig2 with clinical patient-matched gene expression and survival data. Patient samples from the NSCLC Directors' Challenge [18] were ordered according to their expression of each GeneSig2, and the survival data of the patients in the top and bottom quartiles were subjected to Kaplan-Meier analysis. High expression of the genes in GeneSig2 from each differentiation condition was associated with worse overall survival (Figure 2A). This result indicated that our approach could provide relevant molecular insights into the histological observation that less differentiated NSCLC tumors are associated with increased risk of death [1].

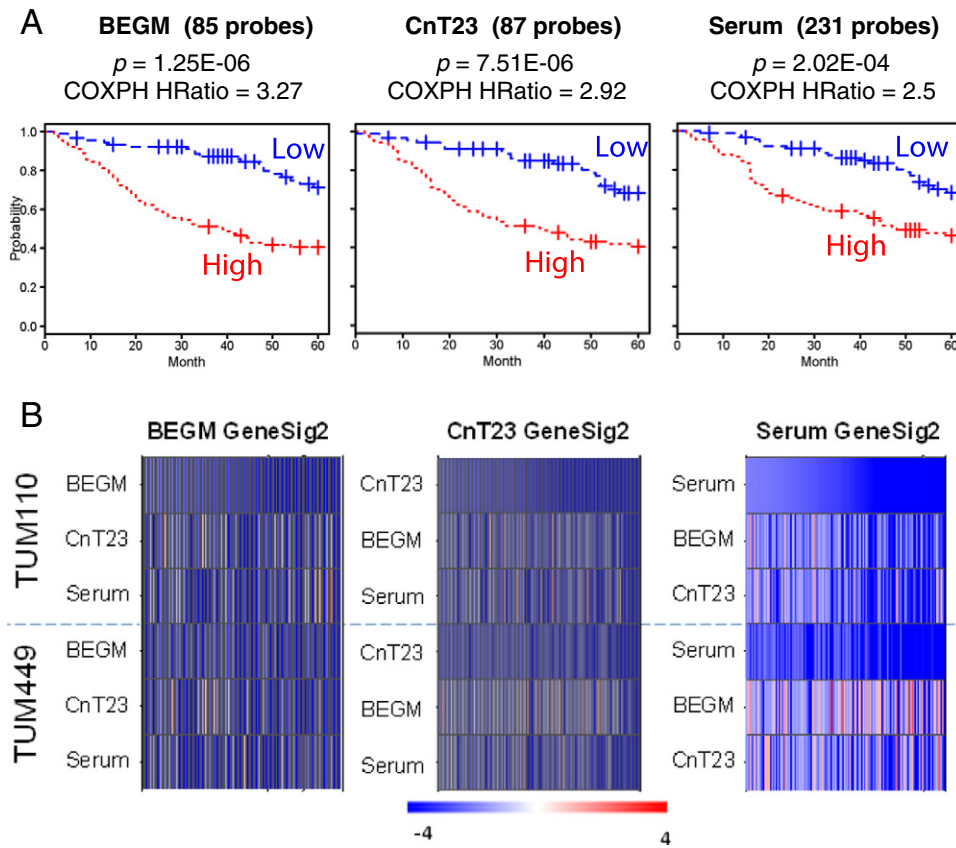


Figure 2. Gene signatures of the poorly differentiated state are associated with shorter survival in NSCLC. (A) The gene signature Genesig2 was defined as the group of probe sets with at least two-fold higher expression ($FDR \leq 0.01$) in the poorly differentiated state in both cultures. Patient samples were ordered by median expression of the genes in GeneSig2, and survival data for patients in the top quartile (“High”; 90 patients) and bottom quartile (“Low”; 90 patients) were subjected to Kaplan-Meier analysis. Log-rank test P value and Cox proportional hazards (COXPH) ratio (HRatio) was calculated using R survival package. (B) Comparison of the gene signatures from three conditions. Changes (fold-change between differentiated state and undifferentiated state) in gene expression were relatively consistent across tumor cultures and conditions. However, in TUM449, approximately one fourth of genes in FBS GeneSig2 changed in the opposite direction in both ALI conditions.

The BEGM-ALI GeneSig2 had the strongest hazard ratio (HRatio), whereas the serum signature yielded the weakest values (BEGM: 3.27; Cnt23: 2.92; Serum: 2.5; **Figure 2A**) despite the fact that the serum signature contained more than twice as many probe sets as the other signatures. Six probes defined the intersection of the three GeneSig2 sets (Supplemental Figure S1B), and strikingly, this small gene set was correlated with worse overall survival (Hazard Ratio [HRatio] = 2.02, $P = .0023$). Similarly, significant differences in patient survival based on Genesig2 were also observed when the entire patient cohort was bisected into top and bottom halves (BEGM: $P = 2.84E-05$, HRatio = 1.96; Cnt23: $P = .000237$, HRatio = 1.8; Serum: $P = .000486$, HRatio = 1.75).

However, there were also significant differences among the GeneSig2 from the three conditions. The changes in gene expression were generally consistent across tumor cultures and conditions; most genes exhibited changes that trended in the same direction even if the changes did not achieve the thresholds used to generate GeneSig2 (**Figure 2B**). Interestingly, in TUM449, approximately 25% of the probe sets in the serum signature changed in opposite directions in serum versus both ALI conditions. Overall, the GeneSig2 analysis demonstrated the clinical relevance of the experimental approach to characterize the undifferentiated state of NSCLC at the molecular level. Although clinical relevance was

demonstrated for all three differentiation conditions, the results indicated certain molecular consequences of both ALI and serum, which were followed up in additional experiments below.

The converse set of gene signatures—genes with higher expression in the differentiated state—did not exhibit any correlation with patient survival data and thus was not pursued. This observation could be due to the heterogeneity of differentiation as opposed to the relatively few deviations within the poorly differentiated serum-free culture. Additional gene signatures were generated that were comprised of genes that increased *or* decreased during differentiation; patient cohorts that were clustered by these compound gene signatures exhibited a significant survival difference (e.g. log rank P value .013 for BEGM-ALI), but the results were not as robust as those from GeneSig2. Together, these results demonstrate that the signatures defined by genes overexpressed in the poorly differentiated state (GeneSig2) exhibited the most clinical relevance and overall robustness.

Overexpression of mRNA Processing Factors in Poorly Differentiated Cells

Additional bioinformatic analyses were performed with the full gene expression profiles and thus were independent of GeneSig2. Gene Set Enrichment Analysis (GSEA; Broad Institute) demonstrated

that the expression profiles of the poorly differentiated state were significantly enriched for gene sets of poor clinical prognosis (Figure 3A), most notably one from the Directors' Challenge [18] which is the same dataset used above in the independent GeneSig2 analysis. In addition, the identification of stem cell factors demonstrates that the poorly differentiated state of the culture was associated with metastasis factors (Figure 3A, Supplemental Figure S2A), which is consistent with a large body of data that supports the links between metastatic stem cells and poor survival [19,20]. Notably, the gene expression profiles in the poorly differentiated state were enriched for at least five gene sets containing targets of the Myc oncogene (Supplemental Figure S2A).

Strikingly, GSEA identified many RNA processing signatures that were enriched in the poorly differentiated state (Figure 3A). Specifically, the signatures converged on posttranscriptional

modification, especially mRNA splicing. Notably, there were not many significant differences in the expression of ribosomal protein genes and transcription factors in the poorly differentiated versus differentiated states (Supplemental Figure S3). Thus, these results likely reflected the overexpression of these factors independent of any differences in total RNA synthesis or metabolism.

Conversely, GSEA of the profiles from the differentiated state identified gene sets associated with better patient survival and downregulation in stem cells and metastasis (Figure 3B, Supplemental Figure S2B). Differentiation in serum but not ALI was enriched for apoptosis and cell death pathways (Figure 3B), which represented a striking contrast between the serum and ALI. The serum condition also yielded weaker statistics or lack of significance for several gene sets that were enriched in the ALI conditions.

A Upregulated in Undifferentiated

Category	Gene Set (enriched in undifferentiated)	BEGM	CnT23	Serum
Clinical	SHEDDEN_LUNG_CANCER_POOR_SURVIVAL_A6	<0.001	<0.001	<0.001
	KAUFFMANN_MELANOMA_RELAPSE_UP	<0.001	<0.001	<0.001
	POOLA_INVASIVE_BREAST_CANCER_UP	<0.001	0.061	0.057
	RHODES_UNDIFFERENTIATED_CANCER	<0.001	<0.001	<0.001
Stem Cell	WONG_EMBRYONIC_STEM_CELL_CORE	<0.001	<0.001	<0.001
	BENPORATH_ES_1	<0.001	<0.001	0.001
	MUELLER_PLURINET	<0.001	<0.001	<0.001
	ZHANG_BREAST_CANCER_PROGENITORS_UP	<0.001	<0.001	<0.001
	BHATTACHARYA_EMBRYONIC_STEM_CELL	<0.001	0.001	0.026
	RAMALHO_STEMNESS_UP	<0.001	0.007	0.023
mRNA Splicing and Processing	KEGG_SPLICEOSOME	<0.001	<0.001	0.002
	REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA	<0.001	<0.001	<0.001
	REACTOME_ELONGATION_AND_PROCESSING_OF_CAPPED_TRANSCRIPTS	<0.001	<0.001	<0.001
	REACTOME_MRNA_SPLICING	<0.001	<0.001	0.001
	REACTOME_FORMATION_AND_MATURATION_OF_MRNA_TRANSCRIPT	<0.001	<0.001	0.002
	REACTOME_METABOLISM_OF_RNA	<0.001	<0.001	<0.001
	REACTOME_TRANSPORT_OF_MATURE_MRNA_DERIVED_FROM_AN_INTRON_CONTAINING_TRANSCRIPT	<0.001	<0.001	0.001
	REACTOME_MRNA_3_END_PROCESSING	<0.001	<0.001	0.006

B Upregulated in Differentiated

Category	Gene Set (enriched in differentiated)	BEGM	CnT23	Serum
Cell Death	KEGG_APOPTOSIS	0.207	0.045	0.008
	BIOCARTA_DEATH_PATHWAY	0.407	0.191	0.006
Clinical	SHEDDEN_LUNG_CANCER_GOOD_SURVIVAL_A4	0.001	0.001	0.027
Stem Cell	JAATINEN_HEMATOPOIETIC_STEM_CELL_DN	0.003	<0.001	0.016
	BEIER_GLIOMA_STEM_CELL_DN	0.032	0.001	0.013

Figure 3. GSEA of genes that are overexpressed in the undifferentiated state (A) or the differentiated state (B). Green boxes denote high significance (FDR ≤ 0.01); yellow denotes moderate significance (0.01 ≤ FDR ≤ 0.05); red denotes lack of significance. Additional gene sets are shown in Supplemental Figure S2.

The enrichment of mRNA processing factors in the poorly differentiated cells was independently revealed by Ingenuity Pathway Analysis (IPA). The two pathways that were uniquely enriched in the undifferentiated state were RNA posttranscriptional modification and RNA trafficking (Figure 4A). Notably, these pathways were enriched only in the ALI condition, not in serum. The pathways uniquely enriched in the differentiated state were protein degradation, free radical scavenging, and antigen presentation (Supplemental Figure S4).

An examination of individual genes in these pathways demonstrated higher expression of RNA processing factors in the

poorly differentiated state on the microarrays (Figure 4B, Supplemental Table S2). In many cases, the change in expression was more pronounced in ALI than in serum, which likely explains why the IPA pathways were enriched only in ALI.

Splicing Factor Gene Alterations and Overexpression in NSCLC

Given these results, we hypothesized that mRNA processing factors could be dysregulated across NSCLC. The expression of the genes in IPA's RNA posttranscriptional modification pathway was queried in The Cancer Genome Atlas (TCGA). Indeed, almost all of the factors

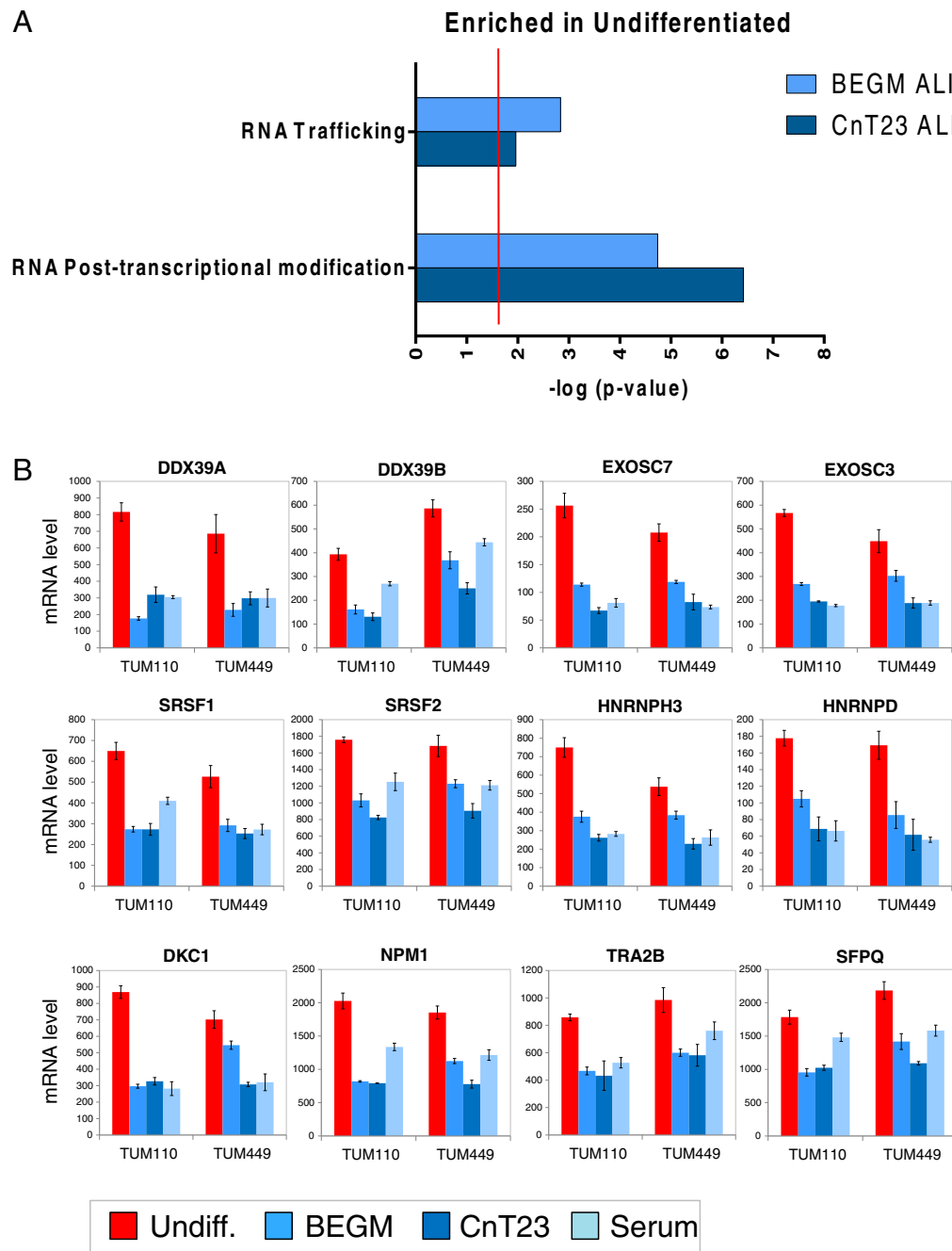


Figure 4. RNA factors are overexpressed in poorly differentiated NSCLC cells. (A) IPA was performed for each condition with the set of genes that changed in both cultures by at least 1.5-fold with $FDR \leq 0.01$. The serum condition is not shown because there was no enrichment of these pathways. (B) Gene expression of mRNA splicing and processing factors is higher in poorly differentiated cells. The corresponding P values are provided in Supplemental Table S2. The mRNA values represent fluorescence intensity after corrections described in Methods.

in the panel were found to be overexpressed in NSCLC compared with normal tissue; the results were observed in both squamous cell carcinoma and adenocarcinoma (Figure 5).

To assess genetic alterations in these factors, TCGA was queried with cBioPortal [21] for the same panel of genes. Genomic alterations were observed in 71% (351/493) of squamous cell carcinoma cases and 37% (206/563) of adenocarcinoma cases (Supplemental Figure S5, A and B). Most of these alterations were copy number gain, defined as at least 2.8 copies, and some were mutations; very few of the alterations were copy number loss. The high frequency (58%) of *TRA2B* copy number gain in squamous cell carcinoma accounted for most of the difference in alteration frequencies between the subtypes. The *TRA2B* amplification was not focal. Notably, the mRNA expression level of *TRA2B* was highly correlated with copy number (Pearson $r = 0.75$; Supplemental Figure S5C).

Alternative Splicing in Poorly Differentiated NSCLC Cells

Based on the above observations, we hypothesized that poorly differentiated NSCLC cells might exhibit a distinct alternative splicing profile from their differentiated counterparts. The SpliceArray™ platform enables the measurement of 281,209 splicing events,

both documented and theoretical, and has identified novel alternative splice isoforms in several studies [22,23]. The coordinated measurement of many probes at each event provides an internal control and allows for normalization. To generate a global alternative splicing profile, mRNA samples from TUM110 and TUM449 in the poorly differentiated and BEGM-ALI states (three biological replicates per condition) were evaluated on the SpliceArray™; the BEGM-ALI condition was chosen because it produced the most robust results in the above analyses. Importantly, gene expression changes (measured independently of alternative splicing events) were highly correlated on the SpliceArray™ and Affymetrix platforms, with Pearson r values of 0.798 and 0.772 in TUM110 and TUM449, respectively (Supplemental Figure S6).

We identified 264 alternative splice events within 137 unique genes that changed significantly between poorly differentiated and differentiated cells in both TUM110 and TUM449 (Table 1; Supplemental Table S3). Most of the events with significant changes (67%) were classified as exon skipping. A total of 86.7% of the events mapped to the coding sequence, which represented a significant enrichment over the corresponding proportion of events on SpliceArray (81%; $P = .03$ by the Fisher exact test). The remainder mapped to the 5' or 3' untranslated region (Supplemental Table S4).

Global analysis of the alternative splicing events that changed between poorly differentiated and differentiated cells revealed an unexpectedly high proportion on chromosome 19, as normalized to the SpliceArray events per chromosome and thus independent of gene density (Figure 6A; $P < .0001$, chi-square goodness-of-fit test). The chromosome 19 genes with splicing event changes were highly enriched within ontology categories of extracellular factors ($P = 1.7E-11$), many of which have been implicated in tumorigenesis, such as *CEACAM1*, *GDF15*, *ICAM1*, *C3*, and *MUC16* and the kallikrein-related peptidases *KLK6*, *KLK7*, *KLK8*, and *KLK10*.

In addition, among the 137 genes were factors that have been associated with NSCLC (e.g., *ANLN*, *C3*, *CLDN4*, *CRYAB*, *EIF4A2*, *ELF3*, *S100P*, *SAA*, *SOD2*), factors with documented function in undifferentiated cells (e.g., *ALDH3A1*, *CEACAM3*, *CEACAM5*, *DNMT1*, *GTF2I*, *ICAM1*, *L1CAM*, *TERF1*, *TOP2A*), and factors that mediate interactions with the tumor microenvironment (e.g.,

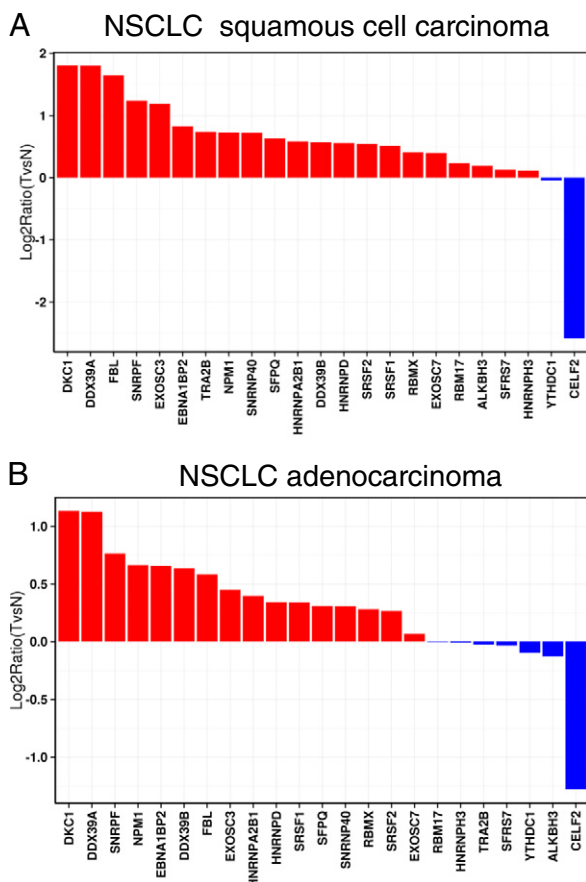


Figure 5. Upregulation of RNA processing factors in NSCLC primary tumors. (A) mRNA expression levels of splicing factors in NSCLC squamous cell carcinomas, expressed as a ratio of expression in tumor (T) to normal tissue (N). TCGA RNAseq data were analyzed with DESeq R package as described in the supplement. (B) mRNA expression levels of splicing factors in NSCLC adenocarcinomas, expressed as a ratio of expression in tumor (T) to normal tissue (N).

Table 1. Overview of Alternatively Spliced Events That Depended on the Differentiation State

Event Type	Number of Events	Number of Genes	Selected Examples [‡]
Alternative splice acceptor	14 (5.3%)	13	<i>ANLN</i> , <i>C3</i> , <i>DLG7</i> , <i>GPX3</i> , <i>PRC1</i> , <i>SLITRK6</i>
Alternative splice donor	14 (5.3%)	14	<i>C3</i> , <i>EIF4A2</i> , <i>INSIG1</i> , <i>S100P</i> , <i>SAAI</i> , <i>SOD2</i> , <i>TLR2</i>
Exon skipped	177 (67%)	92	<i>ALDH3A1</i> , <i>ANLN</i> , <i>C3</i> , <i>CEACAM5</i> , <i>DDX58</i> , <i>DLG7</i> , <i>DNMT1</i> , <i>DPP4</i> , <i>ELF3</i> , <i>ICAM1</i> , <i>IL8</i> , <i>IL32</i> , <i>L1CAM</i> , <i>MUC1</i> , <i>MUC16</i> , <i>TERF1</i> [†] , <i>TOP2A</i>
Intron retention	3 (1.1%)	3	<i>MUC20</i> , <i>CCL2</i>
Novel exon	43 (16.3%)	35	<i>CEACAM3</i> , <i>CRYAB</i> , <i>DNER</i> , <i>EIF4A2</i> , <i>GPX3</i> , <i>GTF2I</i> , <i>MUC1</i> , <i>SAAI</i> , <i>SAA2</i> , <i>SFRS3</i>
Novel intron	13 (4.9%)	10	<i>CLDN4</i> , <i>DDX39</i> , <i>GDF15</i> , <i>HMMR</i> , <i>IL32</i>
Total	264 (100%)	137	

[†] For *TERF1* isoforms, fold-change in the isoforms between the two conditions was 4.26 (FDR = 0.024) in TUM110 and 3.92 (FDR = 0.067) in TUM449.

[‡] The full list of genes and details on each alternative splice event is provided in Supplemental Table S3.

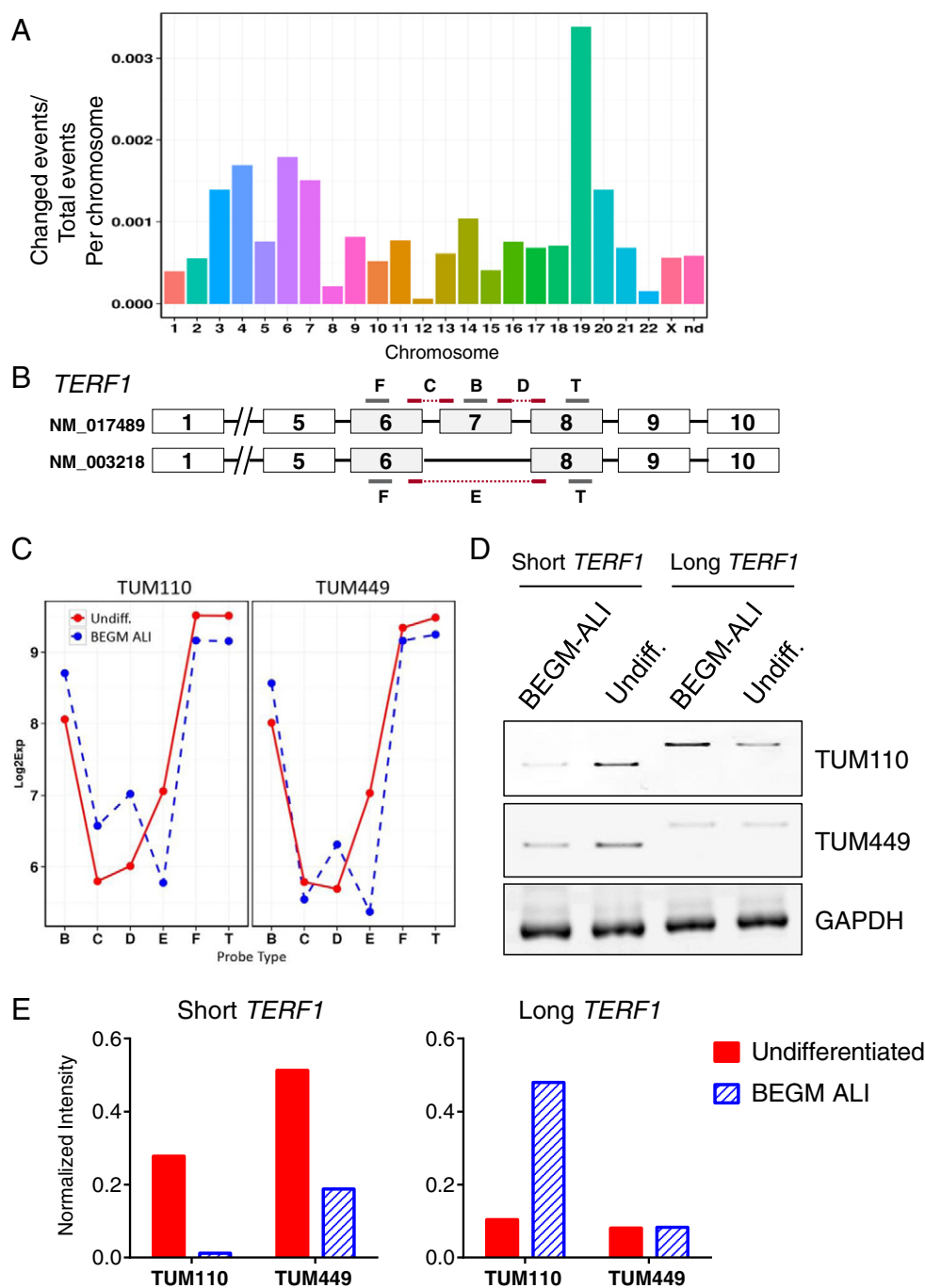


Figure 6. Alternative splicing in poorly differentiated NSCLC cells. (A) Chromosomal distribution of splicing events that changed between undifferentiated and differentiated cells, normalized by the total number of SpliceArray events on each chromosome. The enrichment on chromosome 19 was highly significant ($P < .0001$, chi-square goodness-of-fit test). nd, event not mapped. (B) Schematic representation of *TERF1* alternatively spliced isoforms and the SpliceArray probes (labeled B, C, D, E, F, T). (C) Differential splicing of *TERF1* isoforms detected by Splicearray in TUM110 and TUM449. (D) Polyacrylamide gel electrophoresis was used to visualize RT-PCR products amplified by primer pairs specific to the short and long isoforms of *TERF1*. (E) Quantitative image analysis of *TERF1* RT-PCR products normalized to GAPDH.

CCL2, *IL8*, *IL32*, *GDF15*). The type of alternative splicing event identified for each of these genes is provided in Table 1.

As a notable example of affected genes, the analysis revealed the alternative splicing of *TERF1*, a gene that encodes a telomeric protein and has two characterized splice isoforms [24–26]. Higher expression of the short form was observed in the poorly differentiated state in both cultures; expression relative to the differentiated state was

4.26-fold (FDR = 0.024) in TUM110 and 3.92-fold (FDR = 0.067) in TUM449 (Figure 6, B and C). *TERF1* was selected for confirmation of the SpliceArray data by RT-PCR because of prior characterization of its two isoforms, the documented function of *TERF1* in stem cells [27], and the association between telomeres and tumorigenesis [28,29]. We performed RT-PCR with two probe sets: one that specifically amplifies the short form and the other that

specifically amplifies the long form. The results were consistent with the SpliceArray data (Figure 6, C–E). In fact, a subtle distinction was observed in both array and RT-PCR: the short form was upregulated in the poorly differentiated cells in both cultures; the long form was upregulated in the differentiated cells in TUM110 but unchanged during differentiation in TUM449. These results demonstrate that the alternative splicing of *TERF1* depends on the differentiation state of NSCLC cells.

Discussion

The novel strategy employed in this study to characterize the differentiation hierarchy in NSCLC revealed the upregulation of RNA processing factors and a distinct mRNA splicing profile in poorly differentiated cells, including the preferential expression of the short form of *TERF1*. Our strategy circumvents specific liabilities of other methods such as cell sorting and intact tissue analysis and thus is complementary to such approaches. Our analysis of samples that represent the two major subtypes of NSCLC suggests that the results may apply broadly to NSCLC. Indeed, the overexpression of RNA processing factors that we identified in this study was also observed in the larger datasets that represent hundreds of human NSCLC samples.

ALI Method to Study NSCLC

Inducing differentiation of stem cells by exposure to serum is a common method [3], but our results indicate that the ALI method constitutes a cleaner, more precise technique in NSCLC. As appreciated previously, the 3D structures formed by cells at the ALI resemble solid tissues more than the spindly low-density cultures of cells exposed to serum.

Consistent with the morphological observations, three independent lines of analysis demonstrated the greater precision and relevance of the ALI method, in particular BEGM-ALI. First, the BEGM-ALI gene signature showed the strongest HRatios when compared with clinical datasets, whereas the serum signature had the weakest values despite being comprised of a larger number of genes. Second, GSEA revealed that BEGM-ALI exhibited the strongest enrichment of clinical and stem cell gene signatures, whereas, in contrast, serum was the only condition to enrich for cell death. Third, the enrichment of RNA trafficking and RNA posttranscriptional modification in poorly differentiated cells was observed to a significant degree for BEGM-ALI and CnT23-ALI but not serum.

Although the ALI technique requires more time and expense and has lower throughput relative to serum, our analysis demonstrates its greater precision and clinical relevance. In serum differentiation, the more extreme changes in gene expression coupled with less statistical enrichments suggest that the method is prone to generate a substantial amount of false positives, as well as the induction of cell death. Our results also provide an explanation for the emergence in conventional serum-cultured cancer cell lines of grossly altered states that do not reflect the primary tumors [3].

Distinct mRNA Splicing Profile in Poorly Differentiated NSCLC

Analysis of the primary NSCLC cells cultured at the ALI revealed the overexpression of mRNA processing and splicing factors in the poorly differentiated state. The finding was based on the gene expression changes common to TUM110 and TUM449 and was identified by two independent analyses: GSEA and IPA. Because there were no differences in the overall expression levels of

transcription factors or ribosomal protein genes, the finding suggested that mRNA was being processed differently and/or more efficiently in the less differentiated cells. Consistent with our findings, analysis of mouse embryonic stem cell–like gene expression modules compared with those from differentiated cells revealed the upregulation of RNA processing factors [30]. Notably, the upregulation of mRNA processing and splicing factors was not observed in the poorly differentiated state of normal HBECs [6], which indicates that the dysregulation of these factors may distinguish cancerous from normal lung cells. Interestingly, we also identified the enrichment of Myc signatures in the poorly differentiated cells, and one mechanism of Myc action is to stimulate mRNA capping which in turn promotes downstream RNA processing and translation [31].

Based on the specific enrichment of RNA processing factors in the poorly differentiated cells, we hypothesized that a distinct mRNA splicing profile would be evident and might promote certain phenotypes. Indeed, we identified 264 splice events within 137 genes that differed significantly in the poorly differentiated cells of both TUM110 and TUM449. Although alternative splice isoforms and mutations in splicing factors have been observed widely in cancer, our results provide evidence that alternative splicing might be exploited differently by cancer cells within the same tumor, for example, depending on their differentiation state. We speculate that the disproportionately high frequency of alternative splicing events on chromosome 19 may be connected to the high gene density on the chromosome which could result in especially open or poised chromatin structure and accessibility to splicing factors in the poorly differentiated cells. These results are consistent with many other associations between chromosome 19 and NSCLC [32].

Poorly differentiated tumor cells likely employ some of the mechanisms that maintain normal stem cells, and the alternative splicing of *TERF1* may affect lifespan or unique DNA maintenance functions; some cell cycle checkpoints are distinct in stem cells versus nonstem cells [33,34]. In this respect, it is intriguing that the short isoform does not contain the serine residue that is phosphorylated by Aurora-A kinase and known to promote mitotic defects [35]. In addition, it is possible that the alternative splicing of *TERF1* affects its gene product's interaction with DNA, proteins in the shelterin complex, or proteins that modify *TERF1* such as tankyrase and F-box protein FBX4. Future studies will be needed to explore the functional consequences of differentiation state-dependent splicing of *TERF1* to validate their potential roles in telomere biology and tumorigenesis. Alternative splicing events in *TERF1* and other genes that are specific to tumors and/or undifferentiated tumor cells may offer novel targets for therapeutic intervention with less toxicity than general splicing inhibitors.

Additional downstream effects of the overexpression of RNA processing factors may help maintain poorly differentiated cells. For instance, mRNA transcripts might be processed more efficiently, which may also be coupled to more efficient packaging and export to the cytoplasm [36,37]. The faster production of translation-ready mRNAs could enable the cell to respond more rapidly to microenvironmental cues that regulate cell growth and differentiation. Separately but not exclusively, the overexpression of RNA splicing factors could provide a buffer against stochastic splicing events [38] that could have unpredictable but profound consequences in stem and progenitor cells. In the ALI model and in general, differentiated cells proliferate more slowly than the poorly differentiated cells, and it will be important to distinguish alternative splicing as a function of proliferation versus other aspects of differentiation.

Conclusion

This study has introduced a novel strategy for the characterization of poorly differentiated NSCLC cells and has provided insights into their transcriptional and posttranscriptional regulation as they relate to lung cancer pathology and clinical outcomes.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.tranon.2016.01.006>.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank Max Follettie and Veronica Diesl for support with the Affymetrix platform, Laurent Desire and Matthew Pando for support with the SpliceArray platform, and Jonathan Preall for comments on the manuscript.

References

- Sun Z, Aubry MC, Deschamps C, Marks RS, Okuno SH, and Williams BA, et al (2006). Histologic grade is an independent prognostic factor for survival in non-small cell lung cancer: an analysis of 5018 hospital- and 712 population-based cases. *J Thorac Cardiovasc Surg* **131**, 1014–1020. <http://dx.doi.org/10.1016/j.jtcvs.2005.12.057>.
- Zhou BB, Zhang H, Damelin M, Geles KG, Grindley JC, and Dirks PB (2009). Tumour-initiating cells: challenges and opportunities for anticancer drug discovery. *Nat Rev Drug Discov* **8**, 806–823. <http://dx.doi.org/10.1038/nrd2137>.
- Lee J, Kotliarova S, Kotliarov Y, Li A, Su Q, and Donin NM, et al (2006). Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* **9**, 391–403. <http://dx.doi.org/10.1016/j.ccr.2006.03.030>.
- Prytherch Z, Job C, Marshall H, Oreffo V, Foster M, and Berube K (2011). Tissue-specific stem cell differentiation in an in vitro airway model. *Macromol Biosci* **11**, 1467–1477. <http://dx.doi.org/10.1002/mabi.201100181>.
- Pezzulo AA, Starner TD, Scheetz TE, Traver GL, Tilley AE, and Harvey BG, et al (2011). The air-liquid interface and use of primary cell cultures are important to recapitulate the transcriptional profile of in vivo airway epithelia. *Am J Physiol Lung Cell Mol Physiol* **300**, L25–L31. <http://dx.doi.org/10.1152/ajplung.00256.2010>.
- Ross AJ, Dailey LA, Brighton LE, and Devlin RB (2007). Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. *Am J Respir Cell Mol Biol* **37**, 169–185. <http://dx.doi.org/10.1165/rcmb.2006-0466OC>.
- Damelin M, Geles KG, Follettie MT, Yuan P, Baxter M, and Golas J, et al (2011). Delineation of a cellular hierarchy in lung cancer reveals an oncofetal antigen expressed on tumor-initiating cells. *Cancer Res* **71**, 4236–4246. <http://dx.doi.org/10.1158/0008-5472.CAN-10-3919>.
- Kandath C, McLellan MD, Vandin F, Ye K, Niu B, and Lu C, et al (2013). Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339. <http://dx.doi.org/10.1038/nature12634>.
- Li Y, Chan SC, Brand LJ, Hwang TH, Silverstein KA, and Dehm SM (2013). Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines. *Cancer Res* **73**, 483–489. <http://dx.doi.org/10.1158/0008-5472.CAN-12-3630>.
- Zhang J and Manley JL (2013). Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov* **3**, 1228–1237. <http://dx.doi.org/10.1158/2159-8290.CD-13-0253>.
- Gabut M, Samavarchi-Tehrani P, Wang X, Slobodeniuc V, O'Hanlon D, and Sung HK, et al (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **147**, 132–146. <http://dx.doi.org/10.1016/j.cell.2011.08.023>.
- Chepelev I and Chen X (2013). Alternative splicing switching in stem cell lineages. *Front Biol* **8**, 50–59. <http://dx.doi.org/10.1007/s11515-012-1198-y>.
- Zhou W, Calciano MA, Jordan H, Brenner M, Johnson S, and Wu D, et al (2009). High resolution analysis of the human transcriptome: detection of extensive alternative splicing independent of transcriptional activity. *BMC Genet* **10**, 63. <http://dx.doi.org/10.1186/1471-2156-10-63>.
- Lages CS, Etienne O, Comte J, Gauthier LR, Granotier C, and Pennarun G, et al (2004). Identification of alternative transcripts of the TRF1/Pin2 gene. *J Cell Biochem* **93**, 968–979. <http://dx.doi.org/10.1002/jcb.20235>.
- Lin H, Li H, Cho HJ, Bian S, Roh HJ, and Lee MK, et al (2007). Air-liquid interface (ALI) culture of human bronchial epithelial cell monolayers as an in vitro model for airway drug transport studies. *J Pharm Sci* **96**, 341–350. <http://dx.doi.org/10.1002/jps.20803>.
- Lechner JF, Haugen A, McClendon IA, and Shamsuddin AM (1984). Induction of squamous differentiation of normal human bronchial epithelial cells by small amounts of serum. *Differentiation* **25**, 229–237.
- Vaughan MB, Ramirez RD, Wright WE, Minna JD, and Shay JW (2006). A three-dimensional model of differentiation of immortalized human bronchial epithelial cells. *Differentiation* **74**, 141–148. <http://dx.doi.org/10.1111/j.1432-0436.2006.00069.x>.
- Director's Challenge Consortium for the Molecular Classification of Lung AShedden K, Taylor JM, Enkemann SA, Tsao MS, and Yeatman TJ, et al (2008). Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**, 822–827. <http://dx.doi.org/10.1038/nm.1790>.
- Oskarsson T, Batlle E, and Massague J (2014). Metastatic stem cells: sources, niches, and vital pathways. *Cell Stem Cell* **14**, 306–321. <http://dx.doi.org/10.1016/j.stem.2014.02.002>.
- Pantel K, Alix-Panabieres C, and Riethdorf S (2009). Cancer micrometastases. *Nat Rev Clin Oncol* **6**, 339–351. <http://dx.doi.org/10.1038/nrclinonc.2009.44>.
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, and Aksoy BA, et al (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401–404. <http://dx.doi.org/10.1158/2159-8290.CD-12-0095>.
- Pando MP, Kotraiah V, McGowan K, Bracco L, and Einstein R (2006). Alternative isoform discrimination by the next generation of expression profiling microarrays. *Expert Opin Ther Targets* **10**, 613–625. <http://dx.doi.org/10.1517/14728222.10.4.613>.
- Zhang L, Gong C, Lau SL, Yang N, Wong OG, and Cheung AN, et al (2013). SpliceArray profiling of breast cancer reveals a novel variant of NCOR2/SMRT that is associated with tamoxifen resistance and control of ERalpha transcriptional activity. *Cancer Res* **73**, 246–255. <http://dx.doi.org/10.1158/0008-5472.CAN-12-2241>.
- Palm W and de Lange T (2008). How shelterin protects mammalian telomeres. *Annu Rev Genet* **42**, 301–334. <http://dx.doi.org/10.1146/annurev.genet.41.110306.130350>.
- Shen M, Haggblom C, Vogt M, Hunter T, and Lu KP (1997). Characterization and cell cycle regulation of the related human telomeric proteins Pin2 and TRF1 suggest a role in mitosis. *Proc Natl Acad Sci U S A* **94**, 13618–13623.
- van Steensel B and de Lange T (1997). Control of telomere length by the human telomeric protein TRF1. *Nature* **385**, 740–743. <http://dx.doi.org/10.1038/385740a0>.
- Schneider RP, Garrobo I, Foronda M, Palacios JA, Marion RM, and Flores I, et al (2013). TRF1 is a stem cell marker and is essential for the generation of induced pluripotent stem cells. *Nat Commun* **4**, 1946. <http://dx.doi.org/10.1038/ncomms2946>.
- Artandi SE, Chang S, Lee SL, Alson S, Gortlieb GJ, and Chin L, et al (2000). Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* **406**, 641–645. <http://dx.doi.org/10.1038/35020592>.
- Blasco MA, Lee HW, Hande MP, Samper E, Lansdorp PM, and DePinho RA, et al (1997). Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34.
- Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, and Chang HY (2008). Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* **2**, 333–344. <http://dx.doi.org/10.1016/j.stem.2008.02.009>.
- Dunn S and Cowling VH (2014). Myc and mRNAs capping. *Biochim Biophys Acta* . <http://dx.doi.org/10.1016/j.bbagr.2014.03.007>.
- Wang X, Zhang Y, Nilsson CL, Berven FS, Andren PE, and Carlsson E, et al (2015). Association of chromosome 19 to lung cancer genotypes and phenotypes. *Cancer Metastasis Rev* . <http://dx.doi.org/10.1007/s10555-015-9556-2>.
- Aladjem MI, Spike BT, Rodewald LW, Hope TJ, Klemm M, and Jaenisch R, et al (1998). ES cells do not activate p53-dependent stress responses and undergo p53-independent apoptosis in response to DNA damage. *Curr Biol* **8**, 145–155.
- Damelin M, Sun YE, Sodja VB, and Bestor TH (2005). Decatenation checkpoint deficiency in stem and progenitor cells. *Cancer Cell* **8**, 479–484. <http://dx.doi.org/10.1016/j.ccr.2005.11.004>.
- Ohishi T, Hirota T, Tsuruo T, and Seimiya H (2010). TRF1 mediates mitotic abnormalities induced by Aurora-A overexpression. *Cancer Res* **70**, 2041–2052. <http://dx.doi.org/10.1158/0008-5472.CAN-09-2008>.
- Lei EP and Silver PA (2002). Intron status and 3'-end formation control cotranscriptional export of mRNA. *Genes Dev* **16**, 2761–2766. <http://dx.doi.org/10.1101/gad.1032902>.
- Valencia P, Dias AP, and Reed R (2008). Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proc Natl Acad Sci U S A* **105**, 3386–3391. <http://dx.doi.org/10.1073/pnas.0800250105>.
- Miura SK, Martins A, Zhang KX, Graveley BR, and Zipursky SL (2013). Probabilistic splicing of Dscam1 establishes identity at the level of single neurons. *Cell* **155**, 1166–1177. <http://dx.doi.org/10.1016/j.cell.2013.10.018>.