A Patient-Derived, Pan-Cancer EMT Signature Identifies Global Molecular Alterations and Immune Target Enrichment Following Epithelial-to-Mesenchymal Transition

Milena P. Mak1,2, Pan Tong3, Lixia Diao3, Robert J. Cardnell1, Don L. Gibbons1,4, William N. Williams1, Ferdinandos Skoulidis1, Edwin R. Parra5, Jaime Rodriguez-Canales5, Ignacio I. Wistuba5, John V. Heymach1, John N. Weinstein3, Kevin R. Coombes6, Jing Wang3, and Lauren Averett Byers1

Abstract

Purpose: We previously demonstrated the association between epithelial-to-mesenchymal transition (EMT) and drug response in lung cancer using an EMT signature derived in cancer cell lines. Given the contribution of tumor microenvironments to EMT, we extended our investigation of EMT to patient tumors from 11 cancer types to develop a pan-cancer EMT signature.

Experimental Design: Using the pan-cancer EMT signature, we conducted an integrated, global analysis of genomic and proteomic profiles associated with EMT across 1,934 tumors including breast, lung, colon, ovarian, and bladder cancers. Differences in outcome and in vitro drug response corresponding to expression of the pan-cancer EMT signature were also investigated.

Results: Compared with the lung cancer EMT signature, the patient-derived, pan-cancer EMT signature encompasses a set of core EMT genes that correlate even more strongly with known EMT markers across diverse tumor types and identifies differences in drug sensitivity and global molecular alterations at the DNA, RNA, and protein levels. Among those changes associated with EMT, pathway analysis revealed a strong correlation between EMT and immune activation. Further supervised analysis demonstrated high expression of immune checkpoints and other druggable immune targets, such as PD1, PD-L1, CTLA4, OX40L, and PD-L2, in tumors with the highest mesenchymal EMT scores. Elevated PD-L1 protein expression in mesenchymal tumors was confirmed by IHC in an independent lung cancer cohort.

Conclusions: This new signature provides a novel, patient-based, histology-independent tool for the investigation of EMT and offers insights into potential novel therapeutic targets for mesenchymal tumors, independent of cancer type, including immune checkpoints.

Introduction

Over the past decade, multiple lines of evidence have suggested that epithelial cancers can transform into a more mesenchymal phenotype, a process known as “epithelial-to-mesenchymal transition” (EMT). Studies have shown that EMT plays an important biologic role in cancer progression, metastasis, and drug resistance (1–3). Tools that facilitate the study of EMT, therefore, will provide new insights into molecular regulation and evolution during oncogenesis and may help to improve treatments for mesenchymal cancers. The development of EMT signatures or other molecular markers to identify whether a cancer has undergone EMT is an area of active research. Most studies in this area, however, have focused on a single tumor type and/or on preclinical models (3–7).

We previously developed a robust, platform-independent EMT signature based on a set of 54 lung cancer cell lines (hereafter called the lung cancer EMT signature). In vitro, this lung cancer EMT signature predicted resistance to EGFR and PI3K/Akt inhibitors and identified AXL as a potential therapeutic target for overcoming resistance to EGFR inhibitors [a common treatment for non–small cell lung cancer (NSCLC)]. We and others observed significantly greater resistance to EGFR inhibitors in lung cancers that had undergone EMT, as determined by either their baseline EMT signature (3) or the expression of specific EMT markers after the development of acquired EGFR inhibitor resistance (8).

Although it is appealing to apply the lung cancer EMT signature to diverse tumor types, this approach may be limited by tumor type-specific differences in EMT or discrepancies between cell line
Translational Relevance

Epithelial-to-mesenchymal transition (EMT) is associated with resistance to many approved drugs and with tumor progression. Here, we sought to characterize the common biology of EMT across multiple tumor types and identify potential therapeutic vulnerabilities in mesenchymal tumors. A patient-derived, pan-cancer EMT signature was developed using 11 distinct tumor types from The Cancer Genome Atlas. Mesenchymal tumors had similar patterns of gene, protein, and miRNA expression independent of cancer type. Tumors with mesenchymal EMT scores not only had a higher expression of the receptor tyrosine kinase Axl (previously implicated with EMT and associated with response to Axl inhibitors), but also expressed high levels of multiple immune checkpoints including PD-L1, PD1, CTLA4, OX40L, and PD-L2. This novel finding, which was validated in an independent patient cohort, highlights the possibility of utilizing EMT status— independent of cancer type—as an additional selection tool to select patients who may benefit from immune checkpoint blockade.

Materials and Methods

Datasets

We downloaded cell line drug sensitivity databases from the CCLE (10) and GDSC (11). This included gene expression data (Affymetrix U133 Plus 2 array from CCLE and Affymetrix HT HG U133A array from GDSC) and drug sensitivity data (IC50 values). The CCLE data included 1,035 cell lines and the GDSC data included 654 cell lines. In total, 425 cell lines were present in both drug sensitivity databases. Twenty-four targeted drugs were profiled in CCLE and 138 drugs (both targeted and cytotoxic) were profiled in GDSC with a cell viability assay. The GDSC confirmed identity of cell lines by short tandem repeat (STR) analysis and the CCLE by SNP fingerprinting at multiple steps, profiles were compared with existing profiles (10, 11). We used level-3 TCGA pan-cancer data (12, 13), including RNAseqV2, reverse phase protein array (RPPA), miRNA, copy number, mutation, and clinical data. A summary of the TCGA data can be found in Supplementary Table S1. The Profiling of Resistance Patterns and Oncogenic Signaling Pathways in Evaluation of Cancers of the Thorax and Therapeutic Target Identification (PROSPECT) dataset of surgically resected NSCLC has been previously described (9). Array-based expression profiling of PROSPECT tumors was performed using the Illumina Human WG-6 v3 BeadChip according to the manufacturer’s protocol and gene expression data have been previously deposited in the GEO repository (GSE42127).

Developing the pan-cancer EMT signature

We adopted an approach similar to that used by Byers and colleagues (3) to derive the pan-cancer EMT signature. We used four established EMT markers, namely, CDH1 (epithelial marker, E type), CDH2 (mesenchymal marker, M type), VIM (M type), and FN1 (M type) as seeds to derive the pan-cancer EMT signature on the basis of TCGA pan-cancer RNAseq data. In particular, we computed correlations (Pearson correlation, r) between all mRNAs in the RNAseq data and each of the established EMT markers for each individual tumor type.

Identification of EMT-associated genomic features

We computed Pearson correlation between the EMT score and individual genomic features because these features are normally distributed. The genomic features per gene included the mRNA expression, RPPA expression (either total protein or phosphorylated protein), and miRNA expression levels (5p and 3p mature strands).

Association between EMT score and other covariates

We applied the log-rank test to assess the association between the dichotomized EMT score (with a cutoff of 0) and overall survival. We used the ANOVA test to assess the association between EMT score and tumor grade.

Pathway analysis

We performed a functional annotation of the pan-cancer EMT signature as well as pathway enrichment analysis using QIAGEN’s Ingenuity Pathway Analysis (14).

Quantitative IHC

Tissue sections (4 μM thick) were cut from formalin-fixed, paraffin-embedded blocks containing representative tumor and processed for IHC, using an automated staining system (Leica Bond Max, Leica Microsystems). For assessment of PD-L1 expression, the PD-L1 (E1L3N) XP rabbit monoclonal antibody (Cell Signaling Technology) was applied at 1:100 dilution followed by detection using the Leica Bond Polymer Refine detection kit (Leica Microsystems). DAB staining, and hematoxylin counterstaining. For quantification, stained slides were digitally scanned using the Aperio ScanScope Turbo slide scanner (Leica Microsystems). Captured images (200× magnification) were visualized using ImageScope software (Leica Microsystems,) and analyzed using Aperio Image Toolbox (Aperio, Leica Microsystems). For PD-L1 analysis in tumor and nontumor cells, five randomly selected square regions (1 mm²) in the core of each tumor were evaluated.
Analysis of PD-L1 expression specifically in tumor cells was based on assessment of the whole section. The cell membrane staining algorithm was used to obtain the PD-L1 Histo-score (H-score; 0–300), which is computed on the basis of both extent and intensity of PD-L1 staining.

Association with drug sensitivity data

For each drug sensitivity database, we calculated the EMT scores for the cell lines and associated them with IC_{50} values using the Spearman rank correlation. Cell lines were obtained from commercial vendors and authentication was confirmed by STR analysis.

Results

Development and evaluation of the pan-cancer EMT signature

We derived the pan-cancer EMT signature using an approach similar to that for the lung cancer EMT signature, as shown in Fig. 1A (3). We selected candidate genes from the mRNAs that
best correlated with four established EMT markers "the seed," E-cadherin (CDH1), vimentin (VIM), fibronectin (FN1), and N-cadherin (CDH2), across nine distinct tumor types (see Fig. 1; Supplementary Table S1 for TCGA tumor types and sample sizes). For this initial process, we excluded kidney renal clear cell carcinoma (KIRC) and rectal adenocarcinoma (READ) tumors because they contained predominantly mesenchymal (KIRC) or epithelial (READ) tumors (i.e., low dynamic range of EMT; see Materials and Methods). Using this approach, we identified a set of 77 unique genes as the pan-cancer EMT signature (see Materials and Methods and Supplementary Table S2). Figure 1B shows the correlation between these 77 signature genes and the seed component (one of the four established markers) that identified them. We observed strong and consistent correlations with the seed across different tumor types. This suggests that the pan-cancer EMT signature encompasses core EMT markers that function across different tumors.

We observed a wide dynamic range of EMT scores calculated from the pan-cancer EMT signature across the 1,934 tumors that represent 11 distinct tumor types (Fig. 1C). As expected, EMT scores (calculated from the resulting pan-cancer EMT signature) successfully identified KIRC as mostly mesenchymal and READ as mostly epithelial. With the exception of KIRC (90.1% mesenchymal), READ (87.2% epithelial), and colon adenocarcinoma (COAD; 94.4% epithelial), each tumor type included a significant number of both epithelial (EMT score >0) and mesenchymal (EMT score ≤0) samples.

Fourteen genes overlapped between the original lung cancer EMT signature and the new pan-cancer EMT signature (Supplementary Fig. S1A). EMT scores computed from the pan-cancer EMT signature were highly correlated with scores calculated from the lung cancer EMT signature for individual tumor types (r = 0.76–0.95; overall r = 0.82; Fig. 1D). The most notable outlier when comparing the two signatures was KIRC, which had the lowest correlation (r = 0.759), possibly due to a relatively narrow range of EMT scores due to its mesenchymal tissue of origin.

To investigate whether the pan-cancer EMT signature performed better than the original lung cancer EMT signature, we correlated the EMT scores computed from each signature with six putative EMT markers. The markers included two proteins (E-cadherin and fibronectin) quantified by RPPA, two miRNAs [miR-200a, miR-200b; well established as regulators of EMT (15)] as measured by RNAseq, and two EMT-associated transcription factors (TWIST1 and TWIST2) represented by mRNA expression levels (Fig. 1D). These six putative markers were selected on the basis of their established roles in EMT (1, 2, 16). They served as an independent validation set because they (1) include data from different (nonsignature) platforms (e.g., protein, miRNA) and/or (2) were not components of either mRNA-based EMT signature. Figure 1D shows that correlations with the three putative mesenchymal markers, fibronectin, TWIST1, and TWIST2 (positive correlation with EMT scores), were significantly stronger for the pan-cancer EMT signature (one-sided t test, P < 0.01); whereas correlations with the three epithelial markers, E-cadherin, miR-200a, and miR-200b, were similar for both signatures (one-sided t test, P > 0.05; see also Supplementary Fig. S1B).

We then performed a pathway analysis to investigate the biologic function of the genes identified in the new pan-cancer EMT signature (Fig. 2A). Genes in the pan-cancer EMT signature were significantly associated with several molecular functions, such as cellular movement, morphology, growth, and proliferation, cell-to-cell signaling, and cellular assembly and organization. Another enriched canonical pathway was leukocyte extravasation signaling, suggesting immune activation.

**An integrated molecular analysis of EMT across tumor types:**

**gene expression and potential therapeutic targets**

Using the pan-cancer EMT signature, we evaluated the impact of EMT on the mutational landscape, copy number alterations (CNA), mRNA, miRNA, and protein expression to identify events associated with EMT in patient tumors. To assess how EMT impacts the overall transcriptome, we correlated EMT scores with mRNA expression data for each tumor type (see Supplementary File S1). Genes expressed at higher levels in mesenchymal tumors were highly conserved across multiple tumor types, indicating greater post-EMT biologic homogeneity; whereas those expressed at higher levels in epithelial tumors tended to be more specific to individual tumor types (Supplementary Fig. S2A–S2C).

Given our previously reported association of EMT with drug resistance (3, 4), individual genes or pathways overexpressed in mesenchymal tumors may have clinical utility as therapeutic targets. For example, we previously showed that the receptor tyrosine kinase AXL is highly expressed in lung cancers that have undergone EMT and, therefore, is a top candidate drug target in mesenchymal lung cancers (3). In the pan-cancer EMT signature derived here, AXL was again identified as a signature gene based on its strong correlation with established EMT markers. AXL, therefore, represent a hallmark gene that is positively correlated with EMT scores across all tumor types (r = 0.53–0.93, median = 0.674; Supplementary Fig. S2D). Other potential therapeutic targets that were highly expressed in mesenchymal tumors include the α-type platelet-derived growth factor receptor (PDGFRα; r = 0.22–0.78, median = 0.65), PDGFRβ (r = 0.59–0.92, median = 0.81), and discoidin domain-containing receptor-2 (DDR2; r = 0.29–0.84, median = 0.63), inhibitors of which are in clinical studies or have been approved for cancer treatment (17–20).

**Immune checkpoint expression in mesenchymal tumors**

To better understand pathways globally dysregulated in the setting of EMT, we next performed a pathway analysis of genes that correlated with the pan-cancer EMT signature in all 11 tumor datasets, using a cutoff of an absolute r > 0.3. In addition to functions related to EMT regulation (e.g., cellular movement, cell–cell signaling), the top activated pathways were related to immune cell signaling (Fig. 2B). Given recently published data from our group that suggests a role of EMT in regulating immune escape in lung cancer (9) and the strong therapeutic potential of emerging immunotherapies, we performed a more focused analysis to investigate the association between EMT and key genes involved in the immune response across the 11 cancer types.

The expression of 20 potentially targetable immune checkpoint genes [based on current drug inhibitors that are in preclinical development, clinical trials, or which have been approved by the FDA for specific cancer types (Table 1)] were correlated with the EMT scores for each cancer type. We observed a consistent and strong positive correlation between EMT score and the expression levels of the immune checkpoint genes across all cancer types, with more mesenchymal tumors expressing higher levels of these immune targets (Fig. 2C). Notably, high expression of programmed cell death 1 (PDI), cytotoxic T-lymphocyte-associated protein 4 (CTLA4), and TNF receptor superfamily, member 4...
(OX40L, also known as CD134) were associated with high EMT score regardless of tumor type (PD1 median r = 0.33, median P = 2.1 \times 10^{-4}; CTLA4 median r = 0.36, median P = 1.3 \times 10^{-5}; OX40L median r = 0.65, median P = 2.4 \times 10^{-11}). It has been reported that OX40L regulates T-cell response, which has led to the study of OX40L inhibition in conjunction with other checkpoint blockers (21, 22).

Looking at individual tumor types, samples with higher EMT scores expressed the immune checkpoint genes in a coordinated way, even in predominantly epithelial tumors, such as READ (Fig. 3A). This enrichment of immune target expression in tumors with more mesenchymal characteristics corroborated our recent findings in lung cancer, namely, that lung adenocarcinomas with higher lung cancer EMT signature scores expressed higher levels of PD-L1, which is a target of miR-200 (a suppressor of EMT and metastasis; ref. 9). Clinically, our findings are consistent with results of immunotherapy trials that have shown the greatest activity of immune checkpoint inhibitors in cancer types with the strongest mesenchymal phenotypes, such as melanoma (23–25).

Finally, we validated the association between the mesenchymal phenotype and elevated expression of PD-L1 immunohistochemically in chemotherapy-naïve lung adenocarcinomas and squamous lung cancers that were surgically resected at the University of Texas MD Anderson Cancer Center (PROSPECT dataset). For this analysis, we compared tumors in the top and bottom tertiles for EMT score, using a PD-L1 H-score ranging from 0 to 300. In agreement with their gene expression profiles that confirmed enrichment for immune-related genes (Fig. 3B and C, top panels), mesenchymal adenocarcinomas from PROSPECT exhibited higher PD-L1 H-score than epithelial tumors (P = 0.003, Mann–Whitney U test; Fig. 3B, bottom left). Significantly, this finding was upheld when the analysis was limited to tumor cells (P = 0.002, Mann–Whitney U test; Fig. 3B, bottom right). We further confirmed the significant, positive correlation between the EMT score and PD-L1 H-score in the entire cohort of adenocarcinomas from PROSPECT, using Spearman rank correlation coefficient (P = 0.002 for tumor and nontumor cells and P = 0.011 when the analysis was limited to tumor cells).
Table 1. Clinical development of enriched immune targets in mesenchymal tumors

<table>
<thead>
<tr>
<th>Target</th>
<th>Basal</th>
<th>BLCA</th>
<th>BRCA</th>
<th>COAD</th>
<th>HNSC</th>
<th>KIRC</th>
<th>LUAD</th>
<th>LUSC</th>
<th>OVCA</th>
<th>READ</th>
<th>UCEC</th>
<th>Clinical development</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADORA2A</td>
<td>0.27</td>
<td>0.006</td>
<td>0.54</td>
<td>&lt;0.001</td>
<td>0.27</td>
<td>&lt;0.001</td>
<td>0.45</td>
<td>&lt;0.001</td>
<td>0.36</td>
<td>&lt;0.001</td>
<td>0.16</td>
<td>0.002</td>
</tr>
<tr>
<td>CCL2</td>
<td>0.24</td>
<td>0.016</td>
<td>0.63</td>
<td>&lt;0.001</td>
<td>0.42</td>
<td>&lt;0.001</td>
<td>0.54</td>
<td>&lt;0.001</td>
<td>0.48</td>
<td>&lt;0.001</td>
<td>-0.03</td>
<td>0.625</td>
</tr>
<tr>
<td>CD274 (PD-L1)</td>
<td>0.12</td>
<td>0.228</td>
<td>0.38</td>
<td>&lt;0.001</td>
<td>0.28</td>
<td>&lt;0.001</td>
<td>0.45</td>
<td>&lt;0.001</td>
<td>0.24</td>
<td>&lt;0.001</td>
<td>-0.16</td>
<td>0.002</td>
</tr>
<tr>
<td>CD276 (B7-H3)</td>
<td>0.23</td>
<td>0.021</td>
<td>0.41</td>
<td>&lt;0.001</td>
<td>0.41</td>
<td>&lt;0.001</td>
<td>0.27</td>
<td>0.010</td>
<td>0.55</td>
<td>&lt;0.001</td>
<td>0.66</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CD4</td>
<td>0.34</td>
<td>&lt;0.001</td>
<td>0.57</td>
<td>&lt;0.001</td>
<td>0.45</td>
<td>&lt;0.001</td>
<td>0.56</td>
<td>&lt;0.001</td>
<td>0.48</td>
<td>&lt;0.001</td>
<td>0.29</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CXCR4</td>
<td>0.17</td>
<td>0.083</td>
<td>0.57</td>
<td>&lt;0.001</td>
<td>0.34</td>
<td>&lt;0.001</td>
<td>0.43</td>
<td>&lt;0.001</td>
<td>0.50</td>
<td>&lt;0.001</td>
<td>0.53</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CTLA4</td>
<td>0.25</td>
<td>0.010</td>
<td>0.56</td>
<td>&lt;0.001</td>
<td>0.26</td>
<td>&lt;0.001</td>
<td>0.41</td>
<td>&lt;0.001</td>
<td>0.35</td>
<td>&lt;0.001</td>
<td>0.15</td>
<td>0.005</td>
</tr>
<tr>
<td>HAVCR2 (TIM3)</td>
<td>0.37</td>
<td>&lt;0.001</td>
<td>0.72</td>
<td>&lt;0.001</td>
<td>0.50</td>
<td>&lt;0.001</td>
<td>0.65</td>
<td>&lt;0.001</td>
<td>0.52</td>
<td>&lt;0.001</td>
<td>-0.13</td>
<td>0.014</td>
</tr>
<tr>
<td>ICOS</td>
<td>0.20</td>
<td>0.048</td>
<td>0.57</td>
<td>&lt;0.001</td>
<td>0.24</td>
<td>&lt;0.001</td>
<td>0.31</td>
<td>0.003</td>
<td>0.39</td>
<td>&lt;0.001</td>
<td>0.11</td>
<td>0.038</td>
</tr>
<tr>
<td>ICOSLG</td>
<td>0.05</td>
<td>0.609</td>
<td>0.28</td>
<td>0.008</td>
<td>0.12</td>
<td>0.006</td>
<td>0.18</td>
<td>0.094</td>
<td>0.31</td>
<td>&lt;0.001</td>
<td>-0.01</td>
<td>0.912</td>
</tr>
<tr>
<td>IL1A</td>
<td>-0.02</td>
<td>0.834</td>
<td>0.13</td>
<td>0.236</td>
<td>0.28</td>
<td>&lt;0.001</td>
<td>-0.03</td>
<td>0.756</td>
<td>-0.13</td>
<td>0.663</td>
<td>0.18</td>
<td>0.001</td>
</tr>
<tr>
<td>IL6</td>
<td>0.16</td>
<td>0.097</td>
<td>0.67</td>
<td>&lt;0.001</td>
<td>0.30</td>
<td>&lt;0.001</td>
<td>0.43</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>&lt;0.001</td>
<td>0.56</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IL10</td>
<td>0.32</td>
<td>0.001</td>
<td>0.71</td>
<td>&lt;0.001</td>
<td>0.34</td>
<td>&lt;0.001</td>
<td>0.51</td>
<td>&lt;0.001</td>
<td>0.53</td>
<td>&lt;0.001</td>
<td>0.27</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>LAG3</td>
<td>0.09</td>
<td>0.380</td>
<td>0.59</td>
<td>&lt;0.001</td>
<td>0.14</td>
<td>0.001</td>
<td>0.40</td>
<td>&lt;0.001</td>
<td>0.23</td>
<td>&lt;0.001</td>
<td>0.16</td>
<td>0.003</td>
</tr>
<tr>
<td>PDCD1 (PD1)</td>
<td>0.12</td>
<td>0.226</td>
<td>0.59</td>
<td>&lt;0.001</td>
<td>0.23</td>
<td>&lt;0.001</td>
<td>0.34</td>
<td>&lt;0.001</td>
<td>0.32</td>
<td>&lt;0.001</td>
<td>0.20</td>
<td>0.002</td>
</tr>
<tr>
<td>PDCD1LG2 (PD-L2)</td>
<td>0.30</td>
<td>0.002</td>
<td>0.66</td>
<td>&lt;0.001</td>
<td>0.51</td>
<td>&lt;0.001</td>
<td>0.64</td>
<td>&lt;0.001</td>
<td>0.47</td>
<td>&lt;0.001</td>
<td>0.26</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TGFBI</td>
<td>0.67</td>
<td>&lt;0.001</td>
<td>0.32</td>
<td>0.002</td>
<td>0.64</td>
<td>&lt;0.001</td>
<td>0.74</td>
<td>&lt;0.001</td>
<td>0.20</td>
<td>0.005</td>
<td>0.66</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TNFRSF4 (OX40)</td>
<td>0.33</td>
<td>0.001</td>
<td>0.60</td>
<td>&lt;0.001</td>
<td>0.44</td>
<td>&lt;0.001</td>
<td>0.49</td>
<td>&lt;0.001</td>
<td>0.45</td>
<td>&lt;0.001</td>
<td>0.41</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TNFRSF4 (OX40L)</td>
<td>0.59</td>
<td>&lt;0.001</td>
<td>0.65</td>
<td>&lt;0.001</td>
<td>0.58</td>
<td>&lt;0.001</td>
<td>0.68</td>
<td>&lt;0.001</td>
<td>0.65</td>
<td>&lt;0.001</td>
<td>0.35</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TNFRSF9 (CD137)</td>
<td>0.28</td>
<td>0.004</td>
<td>0.63</td>
<td>&lt;0.001</td>
<td>0.41</td>
<td>&lt;0.001</td>
<td>0.53</td>
<td>&lt;0.001</td>
<td>0.49</td>
<td>&lt;0.001</td>
<td>0.03</td>
<td>0.564</td>
</tr>
</tbody>
</table>

Abbreviations: MoAb, monoclonal antibody; r, Pearson correlation.

aApproved by the FDA.
bIndirect targeting.
Pan-Cancer EMT Molecular and Immune Alterations

A

Basal

Lung adenocarcinoma

Lung squamous

HNSC

KIRC

LUAD

LUSC

OVCA

READ

UCEC

B

Expression (mRNA)

P = 0.003

P = 0.002

P = 0.019

P = 0.110

C

Lung adenocarcinoma

Lung squamous

P = 0.05

P = 0.019

Figure 3. Immune target enrichment in mesenchymal tumors. A, expression of immune checkpoint genes was found in the pan-cancer tumor types. Tumor samples within each cancer type are ordered by EMT score. Mesenchymal tumors have higher expression than epithelial tumors of immune checkpoint genes. B and C, heatmap representation of mRNA expression levels of selected immune checkpoint genes in 151 lung adenocarcinomas (B) and 57 lung squamous carcinomas (C) from the PROSPECT study, rank ordered on the basis of their EMT score (top). Comparison of PD-L1 H-score in epithelial (E) and mesenchymal (M) tumors that were included in the PROSPECT tumor microarray. Automated quantification of extent and intensity of PD-L1 staining was performed in tumor and nontumor cells or limited to tumor cells (B and C, bottom). Lung adenocarcinomas and squamous carcinomas with EMT score in the top/bottom tertile were included in this analysis. Statistical comparisons are based on the Mann-Whitney U test. *, P ≤ 0.05; **, P ≤ 0.01.

www.aacrjournals.org
miRNA expression patterns in mesenchymal tumors

To better understand potential factors contributing to the EMT phenotype, we then investigated changes in other data types (miRNA, protein, mutation, and copy number) that corresponded with differences in EMT score across the pan-cancer set. Because several miRNAs are known to regulate EMT, we identified miRNAs that were significantly correlated with EMT scores, either positively or negatively, across the pan-cancer sets. Figure 4A and B show the miRNAs that strongly correlate with the EMT score (absolute r > 0.3 in four or more tumor types). As expected, high expression of miR-200 family members (miR-200a/b/c, miR-141, etc.) in four or more tumor types. As expected, high expression of miR-200 family members (miR-200a/b/c, miR-141, etc.).
and miR-429), which are known to suppress ZEB1 and thereby maintain E-cadherin expression and other hallmarks of an epithelial phenotype, was observed among tumors with the lowest (most epithelial) EMT scores (Fig. 4A and B). We observed a positive correlation between the EMT score and numerous other miRNAs, with the miR-199 family (miR-199a/b) having the greatest expression in mesenchymal tumors (Supplementary Fig. S3A). The miR-199 family has been shown to regulate skeletal development and to play a role in melanoma invasion and metastasis, but has been rarely studied in EMT (26, 27).

Despite the tumor-specific association observed between the expression of some miRNAs and the EMT score, positively correlated miRNAs (i.e., those expressed at higher levels in mesenchymal tumors) were more frequent and homogenous in this analysis than negatively correlated miRNAs (those higher in epithelial tumors; Fig. 4A and B; Supplementary Fig. S3A), suggesting that miRNA expression may play an important role in regulating EMT.

To investigate the functional roles of miRNAs identified by this analysis, we next investigated whether expression of EMT-associated miRNA target genes also correlated with EMT scores. For the EMT-associated miRNAs in Fig. 4A and B, we downloaded their experimentally validated targets using miRWalk (28) and observed a significant overall association between target expression levels and EMT scores (P < 0.001; Fig. 4C).

Taking the miR-200 family as an example, Supplementary Fig. S3B shows the strong associations between the miRNA targets and the EMT score. Collectively, these findings support an intricate regulation of EMT that is conserved across multiple tumor types by miRNAs beyond the miR-200 family and which merit further investigation.

Protein expression and EMT

Next, we evaluated the correlation between the EMT score and protein expression using RPPA data. RPPA is a highly quantitative platform for the measurement of key total and phosphorylated proteins. In this analysis, we included 181 proteins for which data were available across all 11 tumor types. Notably, some proteins relevant to EMT, such as AXL, are excluded from the analysis because they were evaluated in only a subset of the TCGA pancreas cancer tumor types included here.

We found higher levels of several protein markers such as E-cadherin, claudin-7, and Her-3 in epithelial tumors across most tumor types (≥8); with higher levels of PAI1 and p21 in mesenchymal tumors (Fig. 4D). Fibronectin-1 was the only marker with consistently higher expression in all 11 cancer types, possibly due to the use of fibronectin (FN1) mRNAs levels in deriving the pancancer signature (one of the seed genes). In contrast, other proteins such as protein kinase C-α (PKCα) were associated with EMT in a subset of tumor types (4 of 11), reflecting tumor-specific EMT differences.

Mutational landscape and CNAs

To determine the mutational landscape of mesenchymal tumors, we investigated whether a mutation was associated with the EMT score independent of tumor origin. Despite a significant association, the results can still be primarily attributed to mutations specific to certain tumor types (Supplementary Fig. S4A). For example, a VHL mutation primarily observed in KIRC, was strongly associated with higher EMT score. Therefore, after adjusting for tumor type and multiple testing, we performed another analysis (see Materials and Methods), and found no significant association between mutation and EMT score (Supplementary Fig. S4A). We also explored whether mesenchymal status would influence tumor mutation and CNAs burden by quantifying the mutation rate and percentage of CNAs in each sample per tumor type, as described by Cirillo and colleagues (29). Using this approach, we found no correlation between CNAs versus mutation distribution and EMT score (Supplementary Fig. S4C), indicating that enrichment of CNAs or mutations is also primarily driven by the tumor tissue of origin.

EMT score and clinical outcomes

In some cancers, EMT may be associated with worse clinical outcomes. Therefore, we assessed whether pan-cancer EMT scores were associated with clinical covariates such as overall survival and tumor grade. We did not observe a significant association between EMT score and overall survival, although we found a trend toward shorter survival times in patients with mesenchymal ovarian carcinomas [HR, 1.32; 95% confidence interval (CI), 0.96–1.81; P = 0.08] and uterine corpus endometrial carcinomas (HR, 2.07; 95% CI, 0.93–4.63; P = 0.09). We also found higher EMT scores in KIRC and head and neck squamous cell carcinomas with higher tumor grade (P < 0.0001 and P = 0.01, respectively; Supplementary Fig. S5).

Drug sensitivity in mesenchymal tumors

Finally, we tested the ability of the pan-cancer EMT signature to identify associations between EMT and patterns of drug response across 1,689 preclinical models, representing 18 tumor types (Supplementary Table S3). We curated those observed across patient tumors (Figs. 1B and 5B). For example (and as expected), bone sarcoma, glioblastoma multiforme, and melanoma were predominantly mesenchymal (Fig. 5B). To determine patterns of drug sensitivity in mesenchymal tumors, we correlated the EMT score from each cell line and IC50 values to compute an EMT score for each cell line, which we subsequently correlated with drug sensitivity (based on IC50 values). Among 1,035 cell lines in CCLE and 654 cell lines in GDSC with gene expression data, 425 cell lines were common to both sets, allowing us to compare the concordance of EMT scores derived from independently generated expression datasets. Despite the gene expression data having been generated from different versions of the Affymetrix array platforms, EMT scores calculated for individual cell lines were highly concordant, suggesting a robust performance of our approach across platforms (Fig. 5A).

Furthermore, the distribution of EMT scores for the cell line models of each disease type closely mirrored those observed across patient tumors (Figs. 1B and 5B). For example (and as expected), bone sarcoma, glioblastoma multiforme, and melanoma were predominantly mesenchymal (Fig. 5B). To determine patterns of drug sensitivity in mesenchymal tumors, we correlated the EMT score from each cell line and IC50 values for targeted drugs tested in the CCEL (24 drugs) and GDSC (138 drugs) databases. As previously demonstrated with the lung cancer EMT signature (3), we found increased resistance to EGFR inhibitors (e.g., erlotinib). Mesenchymal models also demonstrated greater resistance to other drugs that target the ErbB family, including the dual EGFR/VEGF inhibitor vandetanib (ZD-4674) and the EGFR/Her2 inhibitor lapatinib (Fig. 5C and D). In contrast to ErbB inhibitors, but consistent with higher expression of FGFR1 and PDGFR in mesenchymal cancers, we observed increased sensitivity to the FGFR1 and PDGFR multikinase inhibitor dovitinib (TKI258) in cell lines with higher EMT scores (Fig. 5C). To evaluate potential therapeutic targets in the GDSC dataset, we performed an exploratory analysis of the in vitro effect of drugs against the same targets. From this target analysis (Fig. 5D; Supplementary Fig. S6), we observed...
greater sensitivity to PDGFR inhibitors in mesenchymal tumors. Another interesting finding was increased sensitivity in the mesenchymal cell lines to the inhibition of the serine/threonine protein kinases GSK3 and TBK1 (30).

**Discussion**

EMT has long been recognized as playing a major role in cancer progression and metastasis (1, 2). Despite extensive research in the field, the best way to characterize this phenomenon, especially in the clinical setting, is still under debate (2). Here, through the use of a gene expression signature developed from multiple tumor types, we were able to evaluate global molecular alterations associated with EMT, as well as potential therapeutic targets. Comparing our original lung cancer EMT signature with the pan-cancer EMT signature, we observed a better correlation of the pan-cancer EMT signature with known markers of EMT across all tumor types evaluated, verifying the strength of this approach. Interestingly, commonly expressed genes correlated with the pan-cancer EMT signature in all 11 tumor types included a number of potential therapeutic targets, such as the receptor tyrosine kinase AXL, which was also an important target in mesenchymal lung cancers identified in our lung cancer EMT signature (3). With AXL inhibitors entering clinical trials, the finding of high AXL expression across multiple cancer types in the setting of EMT has important clinical implications (31). Another major finding of this analysis was the significant enrichment of multiple immune targets in cancers that have undergone EMT, which has potentially important implications for identifying patients with cancer who may benefit from immunotherapies. Specifically, we found a high correlation of expression of 20 druggable immune targets...
and mesenchymal status, as defined by the pan-cancer EMT signature and were able to validate these findings by mRNA expression analysis and IHC in an independent lung cancer cohort.

Several reports have shown the importance of miRNA regulation in EMT (2, 15, 32). As expected, the mir-200 family were highly expressed in epithelial tumors. However, a number of novel miRNAs were found to be highly expressed in mesenchymal tumors, regardless of cancer type, including the mir-199 family. Previously implicated in skeletal development and fibrosis (27, 33), mir-199a seems to have a dual role in cancer, both as a tumor suppressor and a promoter, depending on the tissue of origin (27). Although its role in EMT is still unknown, the correlation we found between mir-199 and EMT warrants further investigation.

Although mesenchymal tumors had similar patterns of gene and miRNA expression, the overall mutational landscape and alterations in copy number seem to be mostly determined by tumor type. In contrast, protein expression and drug sensitivity patterns appear to be driven by both tumor type and EMT status. From this analysis, we identified several drugs as potentially more active in the mesenchymal subgroup of each tumor type. However, some limitations of this study include different numbers of cell lines for each drug tested within each tumor type and multiple targets (both inhibitory and promoter) for many of the drugs included in the analysis. Nevertheless, we identified drugs that may be more active in mesenchymal cancers, which are often highly resistant to established targeted therapies (e.g., ErbB2 family inhibitors used in breast, head and neck, and lung cancers). Further validation is necessary to determine the clinical relevance of these drugs in the context of EMT.

In summary, the new pan-cancer EMT signature introduced here identifies global molecular alterations across multiple cancer types that are associated with EMT. Furthermore, the identification of the association between pan-cancer EMT signature scores and expression of candidate drug targets suggests that EMT may be a clinically useful marker for selecting patients most likely to respond to specific cancer-targeting approaches. The finding that mesenchymal tumors express higher levels of immune checkpoint targets is of particular clinical relevance, given the significant clinical activity of immunotherapies, such as PD1 and PD-L1 inhibitors used in lung cancer (34, 35) and other malignancies, which currently lack validated biomarkers to select patients most likely to benefit from these targeted immunotherapies. The results described here suggest a possible role for using a tumor’s EMT phenotype to identify cancers that may be more sensitive to immune targeting strategies.

Except for one previous report (4), other existing EMT signatures were developed from a single tumor type (3, 5–7), which is likely to limit their application across diverse cancer types. Furthermore, many of these were developed in cell line models, which do not reflect the contribution of the tumor microenvironment and immune response, which play important roles in therapeutic response, especially with the emergence of the field of immunotherapy. By deriving the pan-cancer EMT signature from patients’ tumors, we overcome some of these limitations.

### Disclosure of Potential Conflicts of Interest

J.V. Heymach reports receiving commercial research support from AstraZeneca, Bayer, and GlaxoSmithKline; and is a consultant/advisory board member for AstraZeneca, Boehringer Ingelheim, Exelixis, Genentech, GlaxoSmithKline, Lilly, Novartis, and Synnta. No potential conflicts of interest were disclosed by the other authors.

### Authors’ Contributions

Conception and design: M.P. Mak, J.V. Heymach, J. Wang, L.A. Byers


Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): P. Tong, E.R. Parra, J. Rodriguez-Canales, I.I. Wistuba, L.A. Byers


**Grant Support**

This project was partially supported by the NIH through the TCGA (S1024CA143883-05), the Lung SPORE (PS0 CA070907), DoD PROSPECT grant W81XWH-07-1-0306, and the Cancer Center Support Grant (CCSG -CA016672), by the Mary K. Chapman Foundation, and through generous philanthropic contributions to The University of Texas MD Anderson Lung Cancer Moon Shot Program. D.L. Gibbons was supported by the University Foundation for Fighting Lung Cancer, The Stading Lung Cancer Research Fund, the National Institutes of Health (R08-CA151651), and the Cancer Prevention & Research Institute of Texas (RP150405). W.N. Williams was supported by a Conquer Cancer Foundation Career Development Award (4546); I.I. Wistuba was supported by MD Anderson’s Institutional Tissue Bank (ITB), Award Number 2P50CA16672 from the NIH National Cancer Institute; J.V. Heymach was supported by NHINCI (1 R01 CA168484-04) the LUNGevity foundation, The V Foundation for Cancer Research, the David Bruton, Jr. Endowed Chair and the Rexanna Foundation for Fighting Lung Cancer, The Stading Lung Cancer Research Fund, the National Institutes of Health (R08-CA151651), and the Cancer Prevention & Research Institute of Texas (RP150405); W.N. Williams was supported by a Conquer Cancer Foundation Career Development Award (4546); I.I. Wistuba was supported by MD Anderson’s Institutional Tissue Bank (ITB), Award Number 2P50CA16672 from the NIH National Cancer Institute; J.V. Heymach was supported by NHINCI (1 R01 CA168484-04) the LUNGevity foundation, The V Foundation for Cancer Research, the David Bruton, Jr. Endowed Chair and the Rexanna Foundation for Fighting Lung Cancer, J.N. Weinstein was supported by NIH/NICI (3, 2P50 CA010632-11, 2 U10-IRM00037106-A1, and 2 P30 CA016672 36), the Cancer Prevention & Research Institute of Texas (RP103057), and The Michael and Susan Dell Foundation; L.A. Byers was supported by an NCI Cancer Clinical Investigator Team leadership Award (P30 CA016672), The Sidney Kimmel Foundation for Cancer Research, and the Sheikh Khalifa Bin Zayed Al Nahyan Institute for the Personalized Cancer Therapy’s (IPCT’s) Center for Professional Education and Training, D.L. Gibbons and L.A. Byers are supported by MD Anderson Cancer Center Biostatistical Consultant Awards, R. Lee Clark Fellowships of The University of Texas MD Anderson Cancer Center, supported by the Jeanne F. Shelby Scholarship Fund, and the LUNGevity Foundation.

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.

Received April 13, 2015; revised August 11, 2015; accepted August 28, 2015; published OnlineFirst September 29, 2015.
A Patient-Derived, Pan-Cancer EMT Signature Identifies Global Molecular Alterations and Immune Target Enrichment Following Epithelial-to-Mesenchymal Transition

Milena P. Mak, Pan Tong, Lixia Diao, et al.