Decoding critical long non-coding RNA in ovarian cancer epithelial-to-mesenchymal transition

Ramkrishna Mitra1, Xi Chen1, Evan J. Greenawalt1, Ujjwal Maulik2, Wei Jiang3, Zhongming Zhao4 & Christine M. Eischen1

Long non-coding RNA (IncRNA) are emerging as contributors to malignancies. Little is understood about the contribution of IncRNA to epithelial-to-mesenchymal transition (EMT), which correlates with metastasis. Ovarian cancer is usually diagnosed after metastasis. Here we report an integrated analysis of >700 ovarian cancer molecular profiles, including genomic data sets, from four patient cohorts identifying IncRNA DNM3OS, MEG3, and MIAT overexpression and their reproducible gene regulation in ovarian cancer EMT. Genome-wide mapping shows 73% of MEG3-regulated EMT-linked pathway genes contain MEG3 binding sites. DNM3OS overexpression, but not MEG3 or MIAT, significantly correlates to worse overall patient survival. DNM3OS knockdown results in altered EMT-linked genes/pathways, mesenchymal-to-epithelial transition, and reduced cell migration and invasion. Proteotranscriptomic characterization further supports the DNM3OS and ovarian cancer EMT connection. TWIST1 overexpression and DNM3OS amplification provides an explanation for increased DNM3OS levels. Therefore, our results elucidate IncRNA that regulate EMT and demonstrate DNM3OS specifically contributes to EMT in ovarian cancer.
Ovarian cancer is the most lethal gynecologic malignancy in the United States with ~14,100 deaths and 22,500 new cases estimated for 2017\(^1\). The high mortality of ovarian cancer is primarily due to the high rate of therapy resistance and the diagnosis of the disease after it has metastasized, which occurs in ~80% of women\(^1,2,3\). Evidence indicates the ability of ovarian cancer cells to invade and metastasize is enhanced through the loss of epithelial features and the gain of a mesenchymal phenotype known as epithelial-to-mesenchymal transition (EMT)\(^4,5\). EMT leads to reversible reprogramming of cells, which is defined by fundamental changes initiated and maintained by critical genes, their regulatory circuits, and signaling pathways\(^6\).

EMT is regulated at the transcriptional level by EMT-inducing transcription factors and at the post-transcriptional level by non-

---

**Figure a:**

- IncRNA, mRNA, CNA (matched samples)
- Epi or Mes characterization
- Methylation
- miRNA

**Figure b:**

- Epi vs Mes characterization
- DNA methylation
- Copy number changes
- IncRNA expression changes

**Figure c:**

- Multivariate linear regression model
- Output
- Features for predicting gene expression changes

**Figure d:**

- Enriched association with DE genes (–log10 adj. P)
- Known IncRNA
- DNM3OS
- MIAT
- DIO3OS
- MEG3

**Figure e:**

- IncRNA (absolute log2 fold change)
- PhastCons score (primates)

**Figure f:**

- Focal adhesion
- ECM-receptor interaction
- TGF-β signaling pathway
- MAPK signaling pathway
- WNT signaling pathway
- Calcium signaling pathway
- Gap junction
- p53 signaling pathway

---
coding RNA, such as microRNA (miRNA)\(^7\). Recently, reports suggest that long non-coding RNA (lncRNA), non-coding RNA longer than 200 nucleotides, can also modulate gene expression through several not fully characterized mechanisms\(^8\). lncRNA function in a broad range of cellular processes including cell proliferation, apoptosis, and reprogramming of cell pluripotency\(^9\)–\(^11\). Aberrant expression of 15 lncRNA was detected in ovarian cancer tissue, but their functions were not investigated\(^12\). Elevated expression of one lncRNA, ANRIL, was reported to increase proliferation, migration, and invasion of serous ovarian cancer cells\(^13\)–\(^14\). Additionally, 11 lncRNA have recently been shown to function in migration, EMT, and metastasis in multiple cancers; however, their role in ovarian cancer is unknown\(^15\). Therefore, much about lncRNA and ovarian cancer remains to be investigated.

In this study, we systematically identified lncRNA and their regulation of ovarian cancer EMT. Leveraging large-scale ovarian cancer molecular profiles and genomics from multiple independent patient cohorts revealed reproducible regulation of EMT by lncRNA in ovarian cancer. Verification, including evaluation of gene expression, levels of EMT proteins, and ovarian cancer cell migration and invasion after knockdown and proteotranscriptomic characterization, supported the link between EMT and lncRNA. This study increases understanding of lncRNA-mediated regulatory mechanism of ovarian cancer EMT and increases knowledge of the processes involved in ovarian cancer metastasis that may be targetable.

**Results**

**Integrated analysis identifies EMT-associated lncRNA.** To determine whether lncRNA contribute to EMT in ovarian cancer, data from high-grade serous ovarian cancer patient samples were obtained from The Cancer Genome Atlas (TCGA) and stratified into 231 epithelial and 89 mesenchymal subtypes as previously defined\(^16\) (patient subtype information in Supplementary Data 1). For this large cohort of patients, matched DNA methylation, gene copy number, and expression profiles of mRNA, lncRNA, and miRNA available in TCGA were utilized (Methods section; summarized in Fig. 1a, Supplementary Table 1). To predict EMT-linked lncRNA, we constructed a computational framework (Fig. 1). First we employed a multivariate linear regression model considering mRNA and lncRNA expression, copy number, and methylation profiling data from the matched patient samples of TCGA cohort and determined differentially expressed genes whose expression was significantly correlated with lncRNA expression (Benjamini–Hochberg (BH)\(^16\) adjusted regression \(P < 10^{-6}; \text{Fig. 1b, c}\)). The regression model took into account the biases in estimating gene expression changes due to the corresponding copy number and DNA methylation changes (Methods section). In the spectrum of 386 protein coding genes that were significantly differentially expressed (twofold change; edgeR determined BH adjusted \(P < 10^{-3}\)) in the mesenchymal subtype compared with the epithelial subtype, and 2959 non-differentially expressed genes (BH adjusted \(P > 0.25\), as defined in ref. \(^5\)) as background data, we inferred 25 lncRNA that had significantly enriched association (BH adjusted hypergeometric test \(P < 10^{-5}\)) with the differentially expressed genes (Fig. 1d). Of note, 44 differentially expressed genes known for inducing mesenchymal features including, EMT inducing transcription factors (TWIST1, SNAI2, ZEB1, and ZEB2), matrix metalloproteinases, BMP family genes, and collagen family genes (Supplementary Table 2), were identified indicating the regression model was built on meaningful data in the context of EMT\(^18\)–\(^22\).

We also identified the lncRNA that were differentially expressed in epithelial and mesenchymal lineage commitment and that are evolutionary conserved (Fig. 1e). We determined these 25 lncRNA had remarkably more significant differential expression (BH adjusted two-tailed test) in the mesenchymal subtype compared with the epithelial subtype than the remaining lncRNA (Supplementary Fig. 1a). Notably, seven lncRNA had at least twofold change in differential expression (DNM3OS, MIAT, MEG3, DIO3OS, HCG14, UC141, and HCG144; Supplementary Fig. 1b). The distribution of the expression of these seven lncRNA showed both in epithelial and mesenchymal subtypes that they were all well above the detection level that was previously determined for lncRNA expression\(^23\) (Supplementary Fig. 2). Among the seven lncRNA, UC141 was not present in the 25 lncRNA list and lncRNA HCG14 and HAC1 were poorly conserved across the primate species. The remaining four lncRNA (DNM3OS, MIAT, MEG3, and DIO3OS) had significantly elevated expression (BH adjusted two-tailed t-test) in the mesenchymal subtype and they were highly conserved (PhastCons conservation score = 0.94 – 0.98) across the primate species (Fig. 1e, Supplementary Fig. 3a; Methods section). To determine whether the four lncRNA are present in normal ovarian cells, we extracted the expression of the four lncRNA in normal human tissues from the LncRNA2Function database\(^24\). All four lncRNA are expressed in normal ovary with MEG3 having the highest and MIAT the lowest expression level (Supplementary Fig. 3b).

Differentially expressed genes, which were predicted to be regulated by one of the four lncRNA (DNM3OS, MEG3, MIAT, and DIO3OS) had significantly enriched association (BH adjusted hypergeometric test \(P < 0.05\)) with the EMT-linked canonical pathways, including focal adhesion, ECM-receptor interaction, and gap junction (Fig. 1f; Methods section). Because genes in the same pathway are typically co-expressed, we employed first-order partial correlation statistic to examine if the co-expression of the EMT-linked pathway genes was potentially induced by the inferred EMT-linked lncRNA\(^25\) (Methods section). By definition, first-order partial correlation measures gene co-expression with removing the effect of a controlling variable; here the controlling variable is one lncRNA\(^26\)–\(^27\). We observed a significant reduction of gene co-expression after removing the effect of mutually associated lncRNA (\(P < 10^{-10}\), Wilcoxon rank-sum test; Supplementary Fig. 4), which indicate lncRNA-mediated regulation of

---

**Fig. 1** Identifying critical lncRNA in ovarian cancer EMT. a Ovarian cancer patients (\(n = 320\)) with genomic and molecular profiling data that classified into epithelial (Epi; \(n = 231\)) or mesenchymal (Mes; \(n = 89\)) subtypes were selected for analysis. b Heatmap of 386 genes that were differentially expressed in the mesenchymal subtype compared with the epithelial subtype. c Inferring deregulatory programs from ovarian cancer profiling data. Change in mRNA expression is modeled as linear function of the gene’s DNA methylation, copy number, and lncRNA expression. d, e Systematic prediction of EMT-linked lncRNA from the lncRNA-gene association information obtained from the linear model. f The lncRNA that had significantly enriched association with the differentially expressed genes (\(n = 25\), red dots; top 5 lncRNA labeled) were inferred as EMT associated. Remaining lncRNA were represented by gray dots. The X-axis with four different colors represent major annotation classes of the selected lncRNA (\(n = 120\)). The Y-axis denotes which lncRNA had enriched association with the differentially expressed genes compared with non-differentially expressed genes. e Filtering of high confidence EMT-linked lncRNA (\(n = 4\); blue dots with labels) based on their aberrant expression (X and Y-axis) in EMT and conservation score (Z-axis). Gray dots represent remaining lncRNA. f Heatmap shows significantly enriched association of the inferred lncRNA with EMT-linked pathways. For d and f, P-values determined by BH adjusted hypergeometric test.
EMT-linked pathway genes. Collectively, the data suggest the inferred lncRNA may have important roles in ovarian cancer EMT.

Independent ovarian cancer data reproduce lncRNA regulation. Reproducible regulation provides added confidence in the accuracy of the predictions and may reflect genuine molecular events. Therefore, we examined if the results obtained from TCGA data were consistent in another high-grade serous ovarian cancer patient cohort (Gene Expression Omnibus (GEO) accession ID: GSE9891; Table 1). This data set was stratified into 136 epithelial and 97 mesenchymal subtypes, as defined in Yang et al.5 (Table 1, Supplementary Data 2). TCGA and this independent data set showed that a similar number of genes had significant differential expression (>2-fold-change with BH adjusted P < 0.05; 386 genes in TCGA and 346 genes in GSE9891 determined by edgeR and two-tailed t-test, respectively) in the mesenchymal subtype compared to the epithelial subtype. Additionally, the expression fold changes of the genes in these two data sets were strongly correlated (Spearman ρ = 0.7; correlation P < 2.2 × 10^{-16}) in the spectrum of whole transcriptome data (Fig. 2a). Reannotation of microarray probe sets showed that DIO3OS, DN3MOS, MIAT, and MEG3 lncRNA were detected in the two ovarian cancer subtypes at levels similar to known protein coding EMT-linked genes (Supplementary Fig. 5). Except for DIO3OS, the other three lncRNA (DN3MOS, MIAT, and MEG3) had elevated expression in mesenchymal subtype compared to the epithelial subtype (Fig. 2b). These three lncRNA were strongly co-expressed (absolute Spearman ρ > 0.3; BH adjusted correlation P < 10^{-4}) preferentially with the genes that were differentially expressed in the two subtypes compared to the non-differentially expressed genes, but DIO3OS did not (Fig. 2c). Subsequent pathway analysis revealed that DN3MOS, MIAT, and MEG3-associated differentially expressed genes were significantly enriched in the EMT-linked pathways (Fig. 2d; BH adjusted hypergeometric test P < 0.05). These data are consistent with the results obtained from TCGA.

To begin to evaluate the results obtained from our bioinformatics approach, we first focused on MEG3, which was reported to regulate EMT in lung cancer.29 We examined genome-wide mapping of MEG3 binding sites, which were previously determined.30 The data indicate that MEG3 potentially modulates the expression of 30 genes that are members of the EMT-linked pathways of which 22 genes had MEG3 binding sites at their proximal or distal regulatory regions. This is a 2.6-fold enrichment (73.3% genes) compared with the total MEG3 bound genes in genome-wide scale (28.1%) (Fig. 3a, b; Methods section). Therefore, this MEG3 binding information verified the reliability of our prediction results and suggests direct regulation of EMT-linked genes by MEG3. Taken together, we observed highly reproducible lncRNA regulation in two independent patient cohorts, indicating the lncRNA MEG3, DN3MOS, and MIAT likely have important roles in ovarian cancer cell EMT.

### Table 1 Demographics and clinical information of ovarian cancer patient cohorts

| Category (Number of samples) | TCGA<sup>a,b,c</sup> (320) | GSE9891<sup>b,c,d</sup> (233) | GSE18520<sup>b,c</sup> (53) | GSE26193<sup>b,c</sup> (100) | CPTAC<sup>c</sup> (103) |
|------------------------------|-----------------------------|-------------------------------|-------------------------------|-----------------------------|-----------------------------|
| Subtype                      | 231                         | 136                           | NA                           | NA                          | 71                           |
| Epithelial                   | 89                          | 97                            | NA                           | NA                          | 32                           |
| Mesenchymal                  | 320                         | 233                           | 53                           | 75                          | 103                          |
| Histology                    |                             |                               |                               |                             |                              |
| Serous                       | 0                           | 0                             | 0                             | 0                           | 0                            |
| Other                        | 40                          | 88                            | 1                             | 1                           | 1                            |
| Tumor grade                  |                             |                               |                               |                             |                              |
| I                            | 0                           | 0                             | 0                             | 0                           | 0                            |
| II                           | 7                            | 8                             | 33                           | 16                          |                              |
| III                          | 1                            | 1                             | 67                           | 86                          |                              |
| IV                           | 1                            | 0                             | 0                             | 0                           | 0                            |
| Undetermined                 | 5                            | 0                             | 0                             | 0                           | 1                            |
| Tumor stage                  |                             |                               |                               |                             |                              |
| I                            | 0                           | 10                            | 0                             | 17                          | 0                            |
| II                           | 18                          | 9                             | 0                             | 9                           | 7                            |
| III                          | 252                         | 193                           | 58                           | 78                          |                              |
| IV                           | 47                          | 21                            | 16                           | 18                          |                              |
| Undetermined                 | 3                            | 0                             | 3                             | 0                           | 0                            |
| Age at initial pathologic diagnosis | 30 - 87 | 23 - 80 | NA | NA | 34 - 87 |

<sup>a</sup>Discovery data  
<sup>b</sup>Data used for survival analysis  
<sup>c</sup>Data used for meta-analysis  
<sup>d</sup>Independent validation data

DNM3OS overexpression correlates with worse survival. Given that overexpression of the three identified lncRNA potentially induces mesenchymal features, which contribute to metastasis, we questioned whether their overexpression would correlate with patient survival. To address this, we evaluated four independent ovarian cancer data sets (Table 1) and performed a 5-year survival analysis for each lncRNA separately. Patient samples were stratified based on the median expression of the specific lncRNA into high or low. There was no significant correlation of MEG3 or MIAT overexpression with overall patient survival (Supplementary Fig. 6). However, three of the four patient cohorts showed that patients with higher DN3MOS expression had significantly worse overall survival than those with lower DN3MOS expression (Fig. 4; P = 0.041, P = 0.033, and P = 0.054, log-rank test for GSE9891, GSE18520, and GSE26193, respectively). There was a loss of 10, 17, and 16 months, respectively, in median survival for those patients with increased levels of DN3MOS in the three data sets. Evaluation of genes associated with EMT showed that E-CADHERIN, N-CADHERIN, and SNAIL expression were not correlated with overall survival, but SLUG and TWIST1 expression were (Supplementary Fig. 7). We also assessed whether the expression of DN3MOS, MEG3, and MIAT were linearly anticorrelated with survival time. DN3MOS, but not MEG3 and MIAT, had significant negative association (Spearman correlation coefficient ranging from -0.2 to -0.4) with survival in the three data sets. These results suggest that overexpression of DN3MOS, MEG3, and MIAT may predict worse survival in ovarian cancer patients, and their expression might be potential biomarkers for personalized therapy.
Fig. 2 Reproducible lncRNA regulation in independent validation data. a Correlation of mRNA expression fold-changes between TCGA and GSE9891. In each dataset, fold changes were measured in the mesenchymal (Mes) subtype compared to the epithelial (Epi) subtype. b Expression differences of lncRNA in mesenchymal subtype (n = 97) compared to epithelial subtype (n = 136) in GSE9891. ***P = 3.04 × 10^{-18}; **P = 3.95 × 10^{-11}; *P = 0.03; NS = no significant difference, one-tailed t-test; mean ± SEM. c Correlation expression of the lncRNA indicated with differentially expressed (DE) or non-differentially expressed (non-DE) genes in GSE9891. d Pathway enrichment analysis for the differentially expressed genes that were co-expressed with lncRNA. X- and Y-axes represent enrichment P-value of EMT-linked pathway for TCGA and GSE9891, respectively. Dashed lines denote P-value threshold of 0.05 (i.e., 1.301 in log_{10} scale). P-values determined by BH adjusted hypergeometric test.

Characterization of DNM3OS as a regulator of EMT genes. We first selected DNM3OS-associated 256 protein coding genes that were differentially expressed in the mesenchymal subtype compared with the epithelial subtype in TCGA cohort (Supplementary Table 4). We extracted physical relationships among these genes at their protein level from the curated human protein interaction database PINA v2.031. Eighty-eight genes (34%) had at least one interaction with another DNM3OS-associated gene and in total, 132 interactions were determined (Supplementary Table 5). The networks had significantly enriched association (BH adjusted hypergeometric test P < 0.05) with several EMT-linked Gene Ontology biological processes and canonical signaling pathways (Supplementary Tables 6, 7). The results suggest the differential gene expression in the EMT-linked protein interaction networks may be due to changes in DNM3OS expression. Second, we extracted DNM3OS-associated EMT-linked pathway genes identified in the TCGA cohort along with three additional EMT marker genes E-CADHERIN (CDH1), N-CADHERIN (CDH2), and SNAI1 (SNAI1) and predicted their binding affinity with DNM3OS32. We observed that the distribution of minimum interaction energy between DNM3OS and the EMT-linked genes is significantly lower (P = 7.43 × 10^{-06}; Kolmogorov–Smirnov test) compared with genome-wide DNM3OS-RNA interactions (Fig. 5a, Supplementary Fig. 8). To gain additional insight into DNM3OS regulation of EMT genes and determine whether DNM3OS has the potential to regulate the expression of EMT genes, we evaluated where DNM3OS resided in ovarian cancer cells. Cellular fractionation revealed DNM3OS is localized to the nucleus and not to the cytoplasm of ovarian cancer cells (Fig. 5b). Collectively, these results provide further support for DNM3OS regulating genes that mediate EMT.
Loss of DNM3OS induces mesenchymal-to-epithelial transition. To further elucidate the contribution of DNM3OS in EMT in ovarian cancer and to experimentally validate our bioinformatics data, we evaluated knockdown of DNM3OS in ovarian cancer cells through multiple approaches. First, we performed whole transcriptome RNA-sequencing expression profiling after siRNA-mediated knockdown of DNM3OS in SKOV3 cells compared to non-targeting siRNA control (Fig. 6a). Gene set enrichment analysis (GSEA) based on Kyoto Encyclopedia of Genes and Genome (KEGG) database indicated DNM3OS knockdown results in deregulation of several EMT-linked pathways, such as regulation of actin cytoskeleton, focal adhesion, and WNT signaling pathways (Fig. 6b and Methods section). GSEA Hallmark data also showed deregulation of EMT process, Notch signaling and TGFβ signaling pathways in DNM3OS knockdown cells compared with the controls. Genes downregulated in DNM3OS knockdown cells (edgeR; at least twofold change with BH adjusted P < 0.05) were significantly enriched (BH adjusted hypergeometric test P < 0.05) with several EMT-linked pathways including focal adhesion, regulation of cytoskeleton, adherens, gap and tight junction, ECM-receptor interaction, and calcium and MAPK signaling pathways (Fig. 6c). These data indicate that these EMT pathways were preferentially deregulated with DNM3OS loss.

As a second approach, we performed western blot analysis of SKOV3 ovarian cancer cells after knockdown of DNM3OS. There were elevated protein levels of the epithelial marker E-CADHERIN, and reduced levels of the mesenchymal protein N-CADHERIN in the DNM3OS knockdown cells compared to control (Fig. 6d). Additionally, two transcription factors, SNAIL and SLUG that repress E-CADHERIN expression showed reduced levels in the DNM3OS knockdown cells (Fig. 6d). Therefore, our data indicate that loss of DNM3OS results in a mesenchymal-to-epithelial transition and alterations in these linked pathways.

Because mesenchymal cells typically have a greater capacity to metastasize, we next evaluated with transwell assays whether loss of DNM3OS would impact the ability of ovarian cancer cells to migrate and/or invade. DNM3OS knockdown in SKOV3 ovarian cancer cells resulted in significantly reduced numbers of cells migrating (Fig. 6e) and invading (Fig. 6f) as compared to cells with non-targeting control (two-tailed t-test). Because changes in proliferation rates can impact cell migration and invasion, we also assessed proliferation after DNM3OS knockdown. Rates of proliferation were analogous between the cells with DNM3OS knockdown and non-targeting control (Supplementary Fig. 9), indicating DNM3OS does not influence ovarian cancer cell growth. Therefore, DNM3OS regulates ovarian cancer cell movement, which is a critical contributor to metastasis.

We then performed proteotranscriptomic characterization of EMT-linked genes by conducting a meta-analysis to infer if the modulation of their expression related to the changes of DNM3OS expression. We assessed the siRNA-mediated DNM3OS knockdown data, three microarray gene expression profiling data sets available in the GEO database, RNA-sequencing data from TCGA, and protein expression profiling data from the Cancer Proteomic Tumor Analysis Consortium (CPTAC) (Table 1). For patient data, samples were stratified based on the highest or lowest quartile expression of DNM3OS. Gene expression changes were evaluated in patients that had the lowest quartile expression of DNM3OS compared with the patients that had the highest quartile expression of DNM3OS. We determined that genes involved in EMT were consistently deregulated in multiple data sets. Specifically, there was elevated expression of the epithelial marker gene E-CADHERIN, and decreased expression of mesenchymal marker genes N-CADHERIN, SNAIL, SLUG, VERSICAN, VIMENTIN in the samples with reduced DNM3OS (Fig. 6g). There were also reductions in the genes that induce or contribute to mesenchymal features such as collagen family genes, matrix metalloproteinase genes, and TGFβ pathway genes. Differential gene expression P-values, obtained from the multiple independent data sets, were summarized by computing the Fisher’s combined probability test by using the R/Bioconductor package Survcomp. The P-values represent overall significance of the differential expression of these important genes inferred from multiple independent data sets. Collectively, the results indicate that DNM3OS expression modulates ovarian cancer EMT by regulating the expression of several EMT-linked genes and their associated pathways.
**Regulation of DNM3OS expression.** Among the 320 TCGA patients, 16 patients (5%) showed copy number amplification of DNM3OS, indicating that for these patients increased DNM3OS was presumably due to genomic amplification. However, the 304 patients that did not have DNM3OS amplification suggest that elevated DNM3OS expression in these patients is more likely to be the consequence of transcription factor dysregulation. The TWIST1 transcription factor was previously reported to regulate the DNM3OS locus. Specifically, in Twist knockout mice, the DNM3OS locus was downregulated. Moreover, overexpression of TWIST1 in neuroblastoma cells resulted in increased DNM3OS levels, and luciferase reporter assays showed that the TWIST1-binding domain in the promoter of the DNM3OS locus was necessary for expression. TWIST1 is overexpressed in ovarian cancer and a known ovarian cancer EMT marker. In TCGA and independent validation (GSE9891) patient cohorts both TWIST1 (Fig. 7a) and DNM3OS (Figs. 1c, 2b) were significantly overexpressed (>twofold change with BH adjusted P < 0.05; TWIST1 in TGCA was determined by edgeR and remaining data were determined by two-tailed t-test) in the mesenchymal subtype compared with the epithelial subtype of ovarian cancer. We observed that knockdown of TWIST1 resulted in significantly reduced levels of DNM3OS in SKOV3 ovarian cancer cells (Fig. 7b; two-tailed t-test P < 0.008). The shRNA that was the most effective in knocking down TWIST1 resulted in the largest decrease in DNM3OS. Analysis of TCGA and an independent microarray mRNA expression profiling data set (GSE9891) both showed a strong positive correlation between DNM3OS and TWIST1 (0.526 and 0.449, respectively; Spearman’s ρ) (Fig. 7c). Our results together with previously published data indicate TWIST1 regulates the expression of DNM3OS in ovarian cancer cells.

In addition to transcriptionally inducing the expression of DNM3OS, TWIST1 also induces the miR-199/214 cluster, which resides within the human DNM3OS gene locus. Analysis of miRNA-seq expression profiling data from the same TCGA patient cohort showed that the miRNA within the miR-199/214 cluster (miR-214-5p, -214-3p, 199a-5p, and 199a-3p) were also significantly upregulated (>twofold change; BH adjusted P < 10⁻¹⁸ determined by edgeR) in the mesenchymal subtype compared with the epithelial subtype (Fig. 7d). Analysis of TCGA matched miRNA-seq and mRNA-seq expression data revealed strong positive correlations for all four mature miRNA with DNM3OS (0.602 ~ 0.682; Spearman’s ρ) and TWIST1 (0.322 ~ 0.457; Spearman’s ρ) (Fig. 7e). Therefore, overall, our results indicate lncRNA participate in ovarian cancer cell EMT, and specifically, increased DNM3OS expression by amplification or by TWIST1 overexpression contributes to EMT in ovarian cancer.

**Discussion**

Ovarian cancer is a deadly disease, and EMT is believed to be a significant contributor to its aggressiveness. EMT is a complicated process that remains incompletely resolved, making it difficult to target therapeutically. Therefore, it is important to identify critical molecules that regulate EMT. In this study, we leveraged large-scale multidimensional TCGA genomic and protein expression data as well as multiple independent molecular profiling data for high-grade serous ovarian cancer to infer active lncRNA and their regulation potential in ovarian cancer EMT. Our comprehensive study identified three novel lncRNA (DNM3OS, MEG3, and MIAT) associated with ovarian cancer EMT. Genes predicted to be regulated by these lncRNA had significantly enriched association with the EMT-linked pathways. Several of these genes are known epithelial or mesenchymal markers whose reduced or elevated mRNA expression were strongly associated with expression changes of the inferred lncRNA in both TCGA and independent validation data. Additionally, genome-wide mapping of MEG3 binding sites revealed that 73% of EMT-linked pathway genes that were deregulated in EMT in TCGA cohort are bound by MEG3, suggesting MEG3 is likely involved in EMT in ovarian cancer. Previously, it was reported that MEG3 regulated EMT in lung cancer. MIAT had not been previously linked to EMT, but was shown to be upregulated in chronic lymphocytic leukemia and neuroendocrine prostate cancer. Our experimental data showed alterations in DNM3OS expression were linked to EMT in ovarian cancer through changes in cell migration and invasion and EMT-linked RNA and protein levels, and ovarian cancer patient survival. Therefore, these specific lncRNA regulate EMT in ovarian cancer and likely contribute to metastasis and the high mortality of this disease.

One main issue in identifying EMT-linked lncRNA in large-scale data is to minimize false-positives. To achieve this goal, we started from the analysis of only ‘known lncRNA’ that are most reliable and well annotated in leading databases. Second, we applied stringent thresholds to infer key lncRNA and their regulations. Finally, we required the lncRNA to be conserved across the primate species, which is an important filtering step since EMT is an evolutionary conserved process. More importantly, with the use of completely independent high-quality validation data, we highlighted lncRNA-mediated reproducible regulations in EMT. Reproducible results are expected to more likely reflect the true biological regulations in cellular system. Due to the rapid growth of high-throughput genomic data, our integrated computational framework can be applied to other complex diseases for the purpose of deciphering their regulatory systems and identifying critical biomolecules.
**DNM3OS** was the top ranked deregulated lncRNA in ovarian cancer EMT, as well as the top ranked IncRNA among the IncRNAs that had enriched association with the deregulated genes. **DNM3OS** has been reported to be a transcriptional target of TWIST1, which regulates ovarian cancer EMT, and was overexpressed in the mesenchymal subtype of ovarian cancer. Expression data also indicate a significant positive correlation between **DNM3OS** and TWIST1 expression in ovarian cancer. We also showed that knockdown of TWIST1 in ovarian cancer cells led to reduced **DNM3OS**. Additionally, the miR-199/-214 cluster encoded within the human **DNM3OS** gene locus was also overexpressed in the mesenchymal subtype. Of note, elevated **DNM3OS** overexpressed in the mesenchymal subtype of ovarian cancer. **DNM3OS** is a potential regulator of ovarian cancer EMT genes. **Fig. 5**

**ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-01781-0**

**Methods**

**Analysis of TCGA genomic and transcriptomic data.** Gene copy number and DNA methylation profiles, IncRNA, mRNA, and miRNA expression profiles of high-grade serous ovarian cancer patient samples were extracted from TCGA. Normalized IncRNA and mRNA (RNA-seq) expression profiling data in terms of reads per kilobase per million mapped reads (RPKM), and gene copy number data are available in Akrami et al. Additionaly, we extracted and processed TCGA RNA-seq and miRNA-seq raw read counts to identify differentially expressed genes and miRNA, respectively. Level 4 TCGA DNA methylation data were extracted from the UCSC cancer browser. Data from Illumina Infinium HumanMethylation450 platform were used since larger set of DNA methylation profiles was available for this platform compared to Illumina Infinium HumanMethylation450 platform at the time of analysis. Relative DNA methylation levels were measured as β-values ranging from 0 to 1 that represent the ratio of the intensity of the methylated bead type to the combined locus intensity. Here the β-values were offset by −0.5 to shift the whole data set to values between −0.5 and +0.5, as explained in ref. If multiple methylation probes mapped to a promoter region of a gene, a representative probe was selected that showed strongest negative correlation of methylation β-value and mRNA expression of that gene. Genes commonly present in DNA methylation, gene copy number, and mRNA expression profiling data were selected for the analysis.

**TCGA ovarian cancer patient samples subtype classification.** A consensus clustering analysis of mRNA expression profiles of TCGA ovarian cancer patients classified four subtypes: differentiated, immunoreactive, proliferative, and mesenchymal. Subsequently, considering miRNA-associated genes, Yang et al. reclassified the data into epithelial and mesenchymal subtype. A fraction of the samples denoted as mesenchymal subtype by Yang et al. was classified as non-mesenchymal in original TCGA analysis, and these samples were excluded to retain high-quality data.

**Detection of IncRNA, mRNA, and miRNA in TCGA cohort.** To detect IncRNA, we applied two filters as mentioned in ref. First, IncRNA were eliminated if 50th percentile RPKM value = 0; second, IncRNA were selected if 90th percentile RPKM value > 0.1. Subsequently, 120 IncRNA with status known IncRNA in GENCODE database (v21) were selected for downstream analyses. We selected mRNA or miRNA if at least 75% of the samples had a normalized expression value ≥ 1.

**Differentially expressed mRNA, IncRNA, and miRNA in TCGA.** Both RNA-seq and RNA-seq expression profiles of 231 epithelial and 89 mesenchymal patients were analyzed for identifying differentially expressed mRNAs and miRNA, respectively. We employed the R/Bioconductor package edgeR, which was designed to analyze digital gene expression data. Read counts were imported into edger for differential expression analysis. The data were normalized based on negative binomial distribution. Differential expression of mRNA or miRNA between mesenchymal and epithelial subtypes was assessed by estimating an exact test P-value, which is similar to the Fisher’s exact test. The results were further adjusted using the Benjiamini-Hochberg (BH) multiple testing correction method. We quantified differential expression of IncRNA using t-test. Similar to mRNA or miRNA, P-values were adjusted using the BH method.

**Inferring IncRNA-associated mRNA in TCGA cohort.** The IncRNA-mRNA association strength was estimated by the corresponding coefficient score and P-values in a multivariate linear regression model. The model estimated mRNA expression changes due to copy number, DNA methylation, and IncRNA expression changes. Let \( Y_{i,t} \) and \( s_{i,t}^\text{IncRNA} \) denote the expression of gene \( i \) and IncRNA \( k \), respectively in sample \( t \),

\[
Y_{i,t} = \beta_0 + \beta_1^\text{DM} X_{i,t} + \beta_2^\text{CN} X_{i,t} + \beta_3^\text{IncRNA} s_{i,t}^\text{IncRNA},
\]

where \( \beta_0 \) is the bias, \( \beta_1^\text{DM} \) and \( \beta_2^\text{CN} \) are the offsets accounting for the gene’s expression changes at RNA level due to DNA methylation and gene copy number
changes, respectively. The coefficient of interest $p_{\text{lncRNA}}$ indicates the association strength between lncRNA k and gene i. Regression P-values were adjusted using multiple test correction method (FDR) and significantly associated lncRNA and mRNA pairs were selected after applying the strict criteria of regression coefficient cutoff $\pm 0.3$ and BH adjusted $P < 10^{-6}$.

Inferring the impact of lncRNA in gene co-expression. Let us assume that, at the transcript level, a co-expressed gene pair has two genes, $g_x$ and $g_y$. Let us also assume that one lncRNA is mutually co-expressed with $g_x$ and $g_y$. The Spearman’s $\rho$ is denoted by $r_{\text{g}x\text{g}y}$. The first-order partial correlation between $g_x$ and $g_y$ conditioning on lncRNA is:

$$r_{g_xg_y|\text{lncRNA}} = \frac{r_{g_xg_y} - r_{g_x\text{lncRNA}}r_{g_y\text{lncRNA}}}{\sqrt{(1 - r_{g_x\text{lncRNA}}^2)(1 - r_{g_y\text{lncRNA}}^2)}}$$

Genome-wide MEG3 binding sites inferred by ChOP-Seq. We obtained genome-wide mapping of MEG3 binding sites determined by a modified chromatin oligo affinity precipitation (ChOP) method from the author30. As explained in the original study, we identified genes having MEG3 bound genomic regions using the tool GREAT52.
Expression in two data sets (experiments performed); error bars are ± SEM. 

**Fig. 7** TWIST1 regulates DNM3OS expression in ovarian cancer cells. a Expression of TWIST1 in mesenchymal (Mes; n = 89 for TCGA and n = 97 for GSE9891) subtype compared with the epithelial (Epi; n = 231 for TCGA and n = 136 for GSE9891) subtype in two independent patient cohorts. BH adjusted *P < 10^-19*. Determined by edgeR for TCGA and two-tailed t-test for GSE9891; mean ± SEM. b SKOV3 ovarian cancer cells expressing one of three TWIST1 shRNA or control non-targeting shRNA were Western blotted and DNM3OS levels were evaluated by qRT-PCR (data presented are one of two independent experiments performed); error bars are ± SEM *P < 0.008; two tailed t-test. c Spearman’s correlation coefficient (y-axis) between TWIST1 and DNM3OS expression in two data sets (x-axis); *P = 4.16 × 10^-13, **P < 2.2 × 10^-16. d Volcano plot of miRNA in TCGA mesenchymal subtype compared with epithelial subtype that are at least twofold upregulated (red dots) or downregulated (blue dots). Gray dots represent remaining miRNA. miRNA in the miR-199/-214 cluster are labeled. X-axis is miRNA expression fold-change, and Y-axis is differential expression P-values determined by edgeR. e Expression correlation between DNM3OS and miRNA or TWIST1 and miRNA in TCGA cohort. X-axis represents DNM3OS or TWIST1 and indicated miRNA pair. Y-axis represents Spearman’s rank correlation coefficient (Spearman’s ρ) for each pair. *P < 3.6 × 10^-09; **P = 1.41 × 10^-15, ***P < 2.2 × 10^-16.

Detection of conserved lncRNA. A sliding window of 200 nucleotides was chosen along the lncRNA transcript and conservation levels were measured using the average phastCons score across the primate species. The maximally conserved 200 nucleotides sliding window was selected as the representative conservation score.

Pathway enrichment analysis. Differentially expressed genes significantly co-expressed with a specific lncRNA were selected to conduct pathway enrichment analysis using the canonical KEGG pathway database, which is embedded into the software WebGestalt. Pathways showing significantly enriched (BH adjusted hypergeometric test P < 0.05) number of differentially expressed genes were recognized as the lncRNA associated pathways.

Analysis of microarray mRNA and lncRNA expression data. Three independent microarray gene expression profiling data (accession numbers GSE9891, GSE18520, and GSE26193) for ovarian cancer patients were downloaded from the GEO database. The data were generated on the Affymetrix Human Genome U133 Plus 2.0 platform following standard Affymetrix protocol. The data were normalized using the robust multi-array average (RMA) algorithm in the Affy package for R. We mapped the probe set IDs to the NetAffx annotation file to extract lncRNA expression. Probes were averaged to calculate a single-expression intensity measure per gene/lncRNA per array. Differential expression was measured using unpaired t-tests. lncRNA and mRNA association was established by Spearman rank correlation.

siRNA, shRNA, and western blotting. Human ovarian cancer SKOV3 cell lines (mycoplasma free, authenticated) from Dr. Dino Khabele were cultured as previously described in RPMI1640 media, 10% fetal bovine serum, penicillin, and streptomycin. SKOV3 cells were transfected with 50 nM of non-targeting control or DNM3OS siRNA Smartpool (Dharmacon) containing 4 siRNA using Lipofectamine 2000 (Thermo Fisher Scientific). Lysate from SKOV3 cells grown in ≤5% fetal bovine serum was prepared as previously reported with RIPA-buffer containing protease inhibitors. For TWIST1 knockdown, SKOV3 cells were infected with TWIST1-specific shRNA encoded lentivirus or non-targeting control (obtained from Dr. Andrew Aplin, Thomas Jefferson University); 3 days after puromycin selection, cells were lysed as indicated above. Proteins were subjected to SDS-PAGE, transferred to nitrocellulose, and Western blotted with the following antibodies: E-CADHERIN, N-CADHERIN, SNAIL, and Slug (1:1000; Cell Signaling), TWIST1 (1:500; Santa Cruz), and β-Actin (1:2000; Sigma). Unscropped scans of the blots are in Supplementary Fig. 10.

Cell migration and invasion analysis. SKOV3 ovarian cancer cells were transfected with DNM3OS siRNA or non-targeting control (described above), re-suspended in serum free RPMI1640. For migration assays, cells were placed into upper chamber inserts (70,000 cells per well, 8.0 µm pores, BioCoat insert). For invasion assays, cells were seeded in matrigel invasion chambers (150,000 cells per well, 8.0 µm pores, Corning BioCoat Matrigel invasion chamber). After 18 h (migration) or 24 h (invasion) incubation at 37 °C, migrated or invaded cells, respectively, were fixed and stained with Siemens Diff-Quik stain kit (Siemens Healthcare Diagnostic). Using a fixed grid, migrated and invaded cells were counted in 4 and 8 independent fields, respectively, using an Olympus CKX53 inverted microscope (>10), and images were taken by Cytation 5 cell imaging reader (>4 magnification).

qRT-PCR and RNA sequencing. Total RNA was isolated with Trizol (Sigma) from SKOV3 cells transfected with DNM3OS siRNA or non-targeting control and cDNA generated as we previously reported with Superscript III (Invitrogen) as per manufacturer’s instructions. qRT-PCR (triplicates) performed for DNM3OS and β-Actin as previously described with RT2 SYBR Green qPCR Mastermixes (Qiagen). DNM3OS forward: 5′-GTGCTCAAATCTGATGGGACTCT-3′ and reverse: 5′-ACTCAAGGGCTGTGATTTCC-3′. RNA sequencing profiles (48 h; triplicates) were generated using the Illumina TrueSeq platform. Maximum likelihood estimates of transcript read count for each sample were computed with Kallisto v0.43.0. Human transcriptome fasta file was downloaded from Kallisto (Ensembl version GRCch38) and used to create a Kallisto index. On paired fastq files, for each sample, the Kallisto quantification algorithm was run to estimate transcript abundance. Using tximport the transcript level abundances were summarized into gene level abundances. The read counts obtained from tximport were imported into edgeR for differential expression analysis. In edgeR, data were presented as log2-scale expression (10–4) = expression (10–4).
normalized based on negative binomial distribution. The differential expression of genes between DNMSOs knockdown and non-targeting control samples was assessed by estimating an exact test P-value.

IncrRNA cellular localization evaluation. The nuclear and cytoplasmic RNA from SKOV3 cells were separated as previously reported, except lysis buffer was 10 mM Tris (pH 8.0), 140 mM NaCl, 1.5 mM MgCl2, 0.5% Igepal, 3 μl/μl RNAseout. cDNA was synthesized and RT-PCR performed as described above. Primers included the DNMSO primers described above, 7SL primers previously reported, and 45S rRNA forward: 5′-GTCCAGGGTTCGTGCTG-3′ and reverse 5′-CACCACATGCATCAGGAA-3′ primers.

GSEA. We conducted GSEA33 using the KEGG suite34 and the Hallmark gene set embedded in GSEA to identify signaling pathways that were differentially activated in DNMSOs knockdown cells compared to non-targeting siRNA control cells. GSEA was run on pre-ranked list of the genes obtained from the edgeR analyses. Ranking was based on fold-change induction in DNMSOs knockdown cells compared to control cells. GSEA results were assessed as being statistically significant by permutation of 1000 samples.

Data availability. The RNA-sequencing data have been deposited in the Gene Expression Omnibus (GEO) and NCBI Sequence Read Archive (SRA) databases under the accession code GSE104295 and SRR118934, respectively. All other publicly available data referenced during the study can be retrieved from GEO (https://www.ncbi.nlm.nih.gov/gds/), TCGA (https://cancergenome.nih.gov/), or respective authors’ websites. The other data supporting the findings of this study are available within the article and its Supplementary Information and Supplementary Data files.

Received: 26 April 2017 Accepted: 16 October 2017
Published online: 17 November 2017

References
1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. Cancer J. Clin. 67, 7–30 (2017).
2. Cassinistra, S. A. Cancer of the ovary. N. Engl. J. Med. 351, 2519–2529 (2004).
3. Bast, R. C. Jr, Hennessy, B. & Mills, G. B. The biology of ovarian cancer: new insights. J. Nat. Cancer Inst. 97, 679–690 (2005).
4. Dhamija, S. & Diederichs, S. From junk to master regulators of invasion: microRNA evolution and powerful approach to multiple testing. J. Clin. Oncol. 32, 2505–2509 (2014).
5. Zhao, M., Kong, L., Liu, Y. & Qu, H. dEMT: an epithelial-mesenchymal transition associated gene resource. Sci. Rep. 5, 11459 (2015).
6. Li, C. W. et al. Epithelial-mesenchymal transition induced by TNF-alpha requires NF-kappaB-mediated transcriptional upregulation of Twist1. Cancer Res. 72, 1290–1300 (2012).
7. Heldin, C. H., Vanlandewijck, M. & Moustakas, A. Regulation of EMT by TGFbeta in cancer. FEBS Lett. 586, 1959–1970 (2012).
8. Radisky, E. S. & Radisky, D. C. Matrix metalloproteinase-induced epithelial-mesenchymal transition in breast cancer. J. Mammary Gland Biol. Neoplasia 15, 201–212 (2010).
9. Van, X. et al. Comprehensive genomic characterization of long non-coding RNAs across human cancers. Cancer Cell 28, 529–540 (2015).
10. Jiang, Q. et al. LncRNA2Function: a comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data. BMC Genom. 16 (Suppl 3): S2 (2015).
11. Williams, E. J. & Bowles, D. J. Coexpression of neighboring genes in the genome of Arabidopsis thaliana. Genome Res. 14, 1060–1067 (2004).
12. Crawley, M. J. Statistics: an introduction using R (John Wiley and Sons, 2005).
13. Chen, L. & Zheng, S. Studying alternative splicing regulatory networks through partial correlation analysis. Genome Biol. 10, R3 (2009).
14. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? PLoS Comput. Biol. 7, e1001057 (2011).
15. Terashima, M., Tanige, S., Ishimura, A. & Suzuki, T. MEG3 long noncoding RNA contributes to the epigenetic regulation of epithelial-mesenchymal transition in lung cancer cell lines. J. Biol. Chem. 292, 82–97 (2017).
16. Mondal, T. et al. MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. Nat. Commun. 6, 7743 (2015).
17. Cowley, M. J. et al. PNA v2.0: mixing interaction modules. Nucleic Acids Res. 40, D862–D865 (2012).
18. Terai, G., Iwakiri, J., Kameda, T., Hamada, M. & Asai, K. Comprehensive prediction of lncRNA-RNA interactions in human transcriptome. BMC Genom. 17(Suppl 1): 12 (2016).
19. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA 102, 15545–15550 (2005).
20. Kanesha, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30 (2000).
21. Tsai, J. H. & Yang, J. Epithelial-mesenchymal plasticity in carcinoma metastasis. Genes Dev. 27, 2192–2206 (2013).
22. Barrett, T. & Edgar, R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods Enzymol. 411, 352–369 (2006).
23. Schroder, M. S., Cullane, A. C., Quackenbush, J. & Haitjema-Kains, B. Survcomp an R/Bioconductor package for performance assessment and comparison of survival models. Bioinform. 27, 3206–3208 (2011).
24. Loebel, D. A., O’Rourke, M. P., Steiner, K. A., Banyer, J. & Tan, P. P. Isolation of differentially expressed genes from wild-type and Twist mutant mouse limb buds. Genesis 33, 103–113 (2002).
25. Lee, Y. B. et al. Twist-1 regulates the miR-199a/214 cluster during development. Nuclear Acids Res. 37, 123–128 (2009).
26. Nuti, S. V., Mor, G., Li, P. & Yin, G. TWIST and ovarian cancer stem cells: implications for chemoresistance and metastasis. Oncotarget 5, 7260–7271 (2014).
27. Yin, G. et al. TWISTing stemness, inflammation and proliferation of epithelial ovarian cancer cells through MIR199A2/214. Oncogene 29, 3545–3553 (2010).
28. Takai, M. et al. The EMT (epithelial-mesenchymal-transition)-related protein expression indicates the metastatic status and prognosis in patients with ovarian cancer. J. Ovarian Res. 7, 76 (2014).
29. Sattari, A. et al. Upregulation of long noncoding RNA MIAT in aggressive form of chronic lymphocytic leukemias. Oncotarget 7, 5417–54182 (2016).
30. Crea, F. et al. The role of epigenetics and long noncoding RNA MIAT in neuroendocrine prostate cancer. Epigenomics 8, 721–731 (2016).
31. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 22, 1760–1774 (2012).
32. Long, H. et al. MicroRNA-214 promotes epithelial-mesenchymal transition and metastasis in lung adenocarcinoma by targeting the suppressor-of-fused protein (Sufu). Oncotarget 6, 38705–38718 (2015).
33. Sato, T. et al. Comprehensive analysis of long non-coding RNAs in ovarian cancer reveals global patterns and targeted DNA amplification. PLoS ONE 8, e80306 (2013).
34. Cline, M. S. et al. Exploring TCGA pan-cancer data at the UCSC cancer genomics browser. Sci. Rep. 3, 2652 (2013).
35. Jacobsen, A. et al. Analysis of microRNA-target interactions across diverse cancer types. Nat. Struct. Mol. Biol. 20, 1325–1332 (2013).
Acknowledgments

We thank the MetaOmics facility supported by the NCI Cancer Center support grant P30CA056036 for RNA-seq. Drs. Andrew Aplin and Edward Hartsgough for TWIST1 shRNA. Dr. Mick Edmonds for advice on cell migration and invasion analysis, and members of the Eischen lab for helpful suggestions. Support for this study was provided by NCI R01CA177786 (C.M.E.), the Pellini Foundation Fund, and the Sidney Kimmel Cancer Center.

Author contributions

R.M. and C.M.E designed the study; R.M. performed the bioinformatic analyses; W.J. performed initial TCGA RNA-seq analyses; U.M. performed methylation analysis; X.C. performed the knockdown experiments; E.G. performed the transwell assays; Z.Z. supervised the initial TCGA RNA-seq analysis; R.M. and C.M.E. wrote the manuscript; and all authors read and approved the manuscript.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-01781-0.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2017