MOLECULAR CARCINOGENESIS 50:301–309 (2011)

Proteasomal Non-catalytic Subunit PSMD2 as a Potential Therapeutic Target in Association With Various Clinicopathologic Features in Lung Adenocarcinomas

Yasushi Matsuyama,[1,2] Motoshi Suzuki,[1] Chinatsu Arima,[1] Qin Miao Huang,[1] Shuta Tomida,[1] Toshiyuki Takeuchi,[1] Ryoji Sugiyama,[1] Yasutomo Itoh,[3] Yasushi Yatabe,[4] Hidemi Goto,[2] and Takashi Takahashi[1]*

[1]*Division of Molecular Carcinogenesis, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan*
[2]*Department of Gastroenterology, Nagoya University Graduate School of Medicine, Nagoya, Japan*
[3]*Division of Medical Research Engineering, Nagoya University Graduate School of Medicine, Nagoya, Japan*
[4]*Department of Pathology and Molecular Diagnosis, Aichi Cancer Center, Nagoya, Japan*

We previously identified *PSMD2*, a subunit of the 19S regulatory complex of proteasomes, as a constituent of a signature associated with the acquisition of metastatic phenotype and poor prognosis in lung cancers. In the present study, we found that knockdown of *PSMD2* decreased proteasome activity, and induced growth inhibition and apoptosis in lung cancer cell lines. These effects of siRNA-mediated *PSMD2* inhibition were associated with changes in the balance between phosphorylated AKT and p38, as well as with induction of p21. In addition, patients with higher *PSMD2* expression had poorer prognosis and a small fraction of lung cancer specimens carried increased copies of *PSMD2*. Notably, our findings clearly illustrate that lung adenocarcinomas can be divided into two groups; those with and without general upregulation of proteasome pathway genes including *PSMD2*. This general upregulation was significantly more prevalent in the non-terminal respiratory unit (non-TRU)-type, a recently proposed genetically and clinicopathologically relevant expression profile-defined classification of adenocarcinomas ($P < 0.001$ by Fisher's exact test). Patients with adenocarcinomas with general upregulation had significantly shorter survival after potentially curative resection ($P = 0.0001$ by log-rank test) independent of disease stage, as shown by multivariate Cox regression analysis. Our results suggest that *PSMD2* may be a good molecular target candidate and that other co-regulated proteasome pathway genes and/or their common regulator(s) might also be potential targets, warranting future study including elucidation of the underlying common regulatory mechanism.

Key words: lung cancer; apoptosis; gene expression profile

INTRODUCTION

Lung cancer is the leading cause of cancer-related death in Japan, as well as in many other economically well-developed countries [1]. A better understanding of the molecular pathogenesis of this disease is thus urgently required for preventive or therapeutic breakthroughs to drastically reduce the unacceptable number of deaths. Non-small cell lung cancers (NSCLCs) are mainly comprised of adenocarcinomas, squamous cell carcinomas, and large cell carcinomas, of which adenocarcinomas are known to exhibit the highest degree of morphologic and clinical diversities.

Proteasomes are crucially involved in the execution of a number of cellular functions, such as apoptosis and cell-cycle progression in both normal and malignant cells. The 26S proteasome is an abundant multi-enzyme complex that provides the main pathway for degradation of intracellular proteins in eukaryotic cells [2]. It is found in the cytoplasm and nucleus, and consists of a 20S core catalytic cylindrical complex capped at both ends by a 19S regulatory complex [3,4]. The 20S core catalytic complex harbors the proteolytically active β1 (caspase-like), β2 (trypsin-like), and β5 (chymotrypsin-like) subunits. Despite the premise that it would be difficult to use as a target for chemotherapy while maintaining a tolerable therapeutic index,

Additional Supporting Information may be found in the online version of this article.

Abbreviations: NSCLC, non-small cell lung cancer; siRNA, small interfering RNA; TUNEL, Terminal Transferase dUTP Nick End Labeling; TRU, terminal respiratory unit.

*Correspondence to: Division of Molecular Carcinogenesis, Center for Neurological Disease and Cancer, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan.

Received 12 August 2009; Revised 1 February 2010; Accepted 12 February 2010

DOI 10.1002/mc.20632

Published online in Wiley Online Library (wileyonlinelibrary.com).

many types of actively proliferating malignant cells are more sensitive to proteasome blockade than non-cancerous cells. The β5 subunit in the 20S proteasome complex is the specific target of bortezomib, the first proteasome inhibitor to enter clinical development for treatment of multiple myelomas and other malignant disorders, including lung cancers [5–7]. The present study was initiated based on our previous finding that *PSMD2*, a subunit of the 19S regulatory complex, constituted an expression signature that was associated with the acquisition of a metastatic phenotype in the NCI-H460-LNM35 lung cancer cell line as well as with poor prognosis in surgically operated lung cancer patients [8]. Herein, we present evidence of apoptosis by small interfering RNA (siRNA)-mediated inhibition of *PSMD2* in lung cancer cell lines and the presence of increased copies of the *PSMD2* gene in a fraction of lung cancer tissue specimens. In addition, we report that not only *PSMD2*, but also multiple genes of the 26S proteasome and those involved in their assembly were co-upregulated, showing a significant association with shorter survival in surgically treated adenocarcinoma patients.

MATERIALS AND METHODS

Cell Lines

We used ACC-LC-94, ACC-LC-319, A549, and A427 cells derived from human lung adenocarcinoma cell lines, as well as Calu1 cells from a human lung squamous cell carcinoma cell line. NCI-H460-LNM35 cells [9] (hereafter termed LNM35), a highly metastatic subline of the NCI-H460 human lung large cell carcinoma cell line, were also used. Culture conditions and derivations of the panel of lung cancer cell lines were as described previously [10]. Human dermal fibroblasts, TIG112 cells, were also utilized and cultured in Dulbecco's Modified Eagle Medium (DMEM) with 20% fetal calf serum [11].

Antibodies and siRNAs

The following antibodies were used for Western blot analysis: anti-PSMD2 (ABR PA1-964) and anti-β-actin (Sigma–Aldrich, St. Louis, MO); anti-lamin B, anti-cyclin A, and anti-cyclinB1 (Santa Cruz Biotechnology, Santa Cruz, CA); anti-caspase-7, anti-p53, anti-phospho-p38 MAPK (Thr180/Tyr182), anti-phospho-AKT (Ser473), anti p53, and anti-AKT (Cell Signaling Technology, Danvers, MA); and anti-BCL2, anti-BAD, anti-p38α, and anti-p21 (BD Transduction Laboratories, San Jose, CA). The sequences of siRNA against *PSMD2* (si*PSMD2*) used were as follow: sense strand, 5′-CACACUAUGGCAAACUGAATT; and anti-sense strand, 5′-UUCAGUUUGCCAUAGUGUGTT. si*CTRL* refers to the AM4611 *Silencer*® Negative Control #1 siRNA obtained from Applied Biosystems (Foster City, CA). Silencing efficiency was evaluated by both Western blotting and TaqMan-based real-time reverse transcription-PCR analyses (TaqMan gene expression, Hs 01092070_g1, Applied Biosystems) (data not shown for the latter).

3-(4,5-Dimethylthiazol-2-yl)-2, 5-Diphenyltetrazolium Bromide (MTT) Assay

Cells were plated at a density of 8×10^4 in 0.5 mL of culture medium on six-well plates. The next day, either si*PSMD2* or si*CTRL* was transfected at a final concentration of 1 or 10 nM using RNAi MAX (Invitrogen, Carlsbad, CA) (day 0). Viable cells were measured in triplicate using TetraColor One (Seikagaku, Tokyo, Japan) in reference to the viability of non-treated cells on various days after transfection.

Proteasome Activity Assay

Proteasome activity was measured using a 20S proteasome activity assay kit (Chemicon, Temecula, CA), according to the manufacturer's instructions. In brief, control or siRNA-treated cells were lysed in a lysis buffer [150 mM NaCl, 20 mM Tris–HCl (pH 7.2), 1% Triton X-100, 1 mM DTT] without protease inhibitors. Total cell lysates (1.6 μg) prepared from LNM35 cells treated with either si*PSMD2* or si*CTRL* were incubated with 20 μmol/L of the fluorogenic substrate Suc-Leu-Leu-Val-Tyr-AMC for determining proteasomal chymotrypsin-like activity at 37°C in 100 μL of assay buffer [250 mM HEPES-KOH (pH 7.5), 5 mM EDTA, 0.5% NP40, and 0.01% SDS]. Free AMC liberated by the substrate hydrolysis was quantified using a microtiter plate fluorometer (ARVO™ MX 1420 multilabel counter, PerkinElmer, Kanagawa, Japan) with excitation at 355 nm and emission at 460 nm.

Terminal Transferase dUTP Nick End Labeling (TUNEL) Assay

Cells were plated at a density of 8×10^4 in 0.5 mL of culture medium on six-well plates and cultured for 24 h, then transfected with siRNA, as described above (day 0). Two days later, they were stained using an *in situ* cell death detection kit (Roche Diagnostics, Indianapolis, IN) according to the supplier's protocol.

Flow Cytometric Analysis

Cells were plated at a density of 1.5×10^5 in 4 mL of culture medium on 6-cm dishes. The next day, si*PSMD2* or si*CTRL* (10 nM) was transfected as described above. After 48 h of continuous exposure, the cells were treated with a phosphate-buffered saline solution containing 0.5% NP-40 and 50 μg/mL of propidium iodide (Sigma–Aldrich). Cellular DNA contents were measured using a FACSCalibur flow cytometer equipped with the CELLQuest program (BD Biosciences), as described previously [12].

Patients and Statistical Analysis

Our previous dataset comprising 149 patients with NSCLC, which included 90 patients with adenocar-

cinomas, was used in the present study [13] (GSE11969). Adenocarcinoma cases were divided into high and low expression groups by comparison with the median expression level of PSMD2 in the 90 adenocarcinoma specimens. Kaplan–Meier survival curves were used to estimate survival rate as a function of time and survival differences were analyzed by a log-rank test, as described previously [1314]. Approval for this study was obtained from the institutional review boards of both Nagoya University and Aichi Cancer Center.

TaqMan-based Copy Number Analysis

Gene copies of *PSMD2* and *PIK3CA* in each tumor were analyzed using TaqMan Universal PCR Master Mix (Applied Biosystems) with the following PCR primers and TaqMan probes: Hs03470293_cn and Hs04761440_cn for *PSMD2* and *PIK3CA*, respectively. Quantification of the control *ALB* has been described [15]. The fold increase in copy number in a given cancer tissue was calculated as the ratio of either *PSMD2* or *PIK3CA* signals to the *ALB* signal, followed by further normalization to that of the mean value obtained from 46 normal lung DNA samples.

RESULTS

Reduction of Lung Cancer Cell Proliferation by Inhibition of *PSMD2* Expression

First, we examined two human lung adenocarcinoma cell lines, LNM35 and ACC-LC-94, as well as a human normal fibroblast cell line, TIG112, for their response to siRNA-mediated inhibition of *PSMD2*. Western blot analysis showed a significant reduction of PSMD2 expression in all three cell lines treated with si*PSMD2* (Figure 1A). Furthermore, an MTT assay conducted on day 2 after the start of siRNA treatment revealed that si*PSMD2* treatment significantly reduced proliferation in both LNM35 and ACC-LC-94, but not in TIG112, when compared to treatment with the negative control si*CTRL* (Figure 1B). ACC-LC-94 and LNM35 showed gradual and further incremental decreases in proliferating cells over time, whereas TIG112 did not (Figure 1C).

Inhibition of Proteasome Activity in Cells Treated with siRNA Against *PSMD2*

PSMD2 is an ortholog of the proteasome subunit of *Rpn1/NAS1* in *Saccharomyces cerevisiae* as well as *Mts4* in *Schizosaccharomyces pombe* [16,17]. We examined the proteolytic cleavage activity of PSMD2 using a fluorogenic peptide as a substrate for various times to confirm that PSMD2 functions as a proteasome subunit in human lung cancer cells. Cell lysates were prepared from LNM35 cells that had been treated with either si*PSMD2* or si*RNA* for 48 h, and subjected to proteasome activity analysis. si*PSMD2* treatment clearly suppressed the proteasome activity (Figure 2),

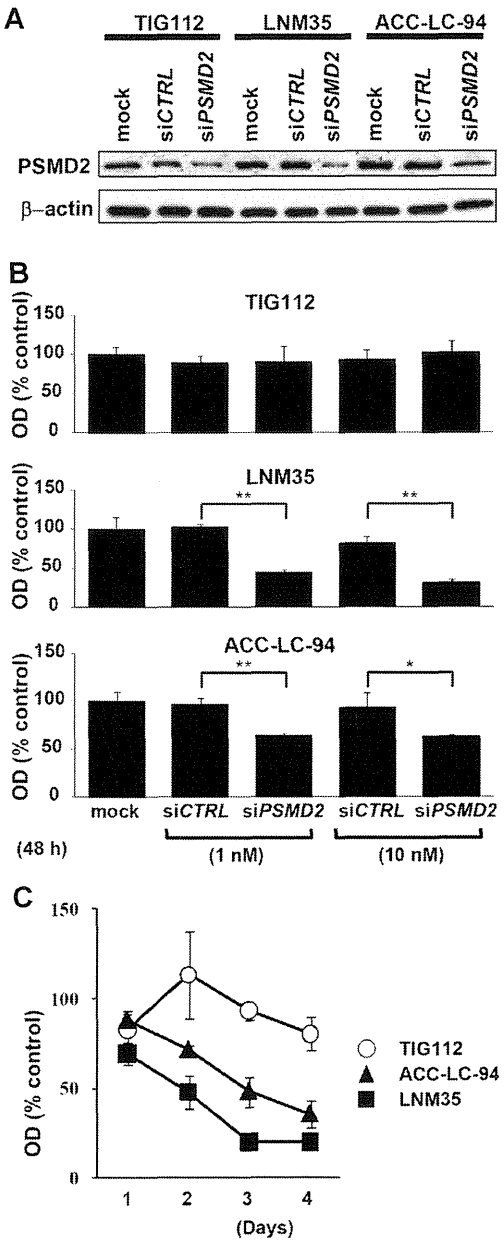

Figure 1. Reduction of cell growth by *PSMD2* knockdown in lung cancer cell lines (LNM35 and ACC-LC-94), but not in a normal fibroblast cell line (TIG112). (A) Western blot analysis of PSMD2 at 48 h after siRNA treatment. β-actin was used as a loading control. (B) MTT assay of cells treated with either 1 or 10 nM siRNA for 48 h. * and ** indicate the significant growth suppression with the *P* values of $P < 0.05$ and $P < 0.01$, respectively. (C) Results of MTT assay that measured cell growth for up to 4 days after treatment with 10 nM siRNA. OD values with si*CTRL* in (B) and (C) were regarded as 100%. si*PSMD2*, siRNA against *PSMD2*; si*CNTL*, negative control siRNA.

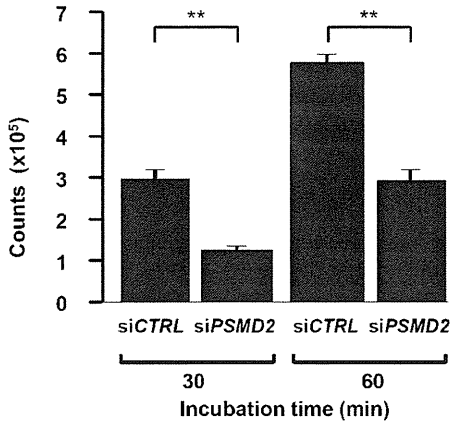

Figure 2. Reduction of proteasome activity by *PSMD2* knockdown. Intracellular proteasome activity was measured using cell lysates of LNM35 treated with either 10 nM si*CTRL* or si*PSMD2* for 2 days. Data shown were obtained after incubation for 30 and 60 min, during which linear increases in proteasome activity were observed. **Indicates the significant reduction in proteasome activity with $P < 0.01$.

showing that *PSMD2* is required for 26S proteasome functions in human lung cancer cells.

Growth Inhibitory Effects of *PSMD2* Knockdown in Panel of Lung Adenocarcinoma Cell Lines

A panel of human lung adenocarcinoma specimens was examined for growth inhibition in response to *PSMD2* knockdown using an MTT assay. Western blot analysis showed apparent reduction of PSMD2 expression in all cell lines treated with si*PSMD2* (Figure 3A). An MTT assay was performed on day 2 after initiating siRNA treatment, which showed significantly reduced proliferation in LNM35, ACC-LC-94, ACC-LC-319, and A549 (Figure 3B). In contrast, growth inhibition was not observed in A427, Calu1, or TIG112.

Induction of Apoptosis by *PSMD2* Knockdown in Lung Cancer Cell Lines

Next, we performed TUNEL and flow cytometric assays to determine whether *PSMD2* knockdown induced apoptosis. The TUNEL assay results clearly showed induction of apoptosis by si*PSMD2* treatment in LNM35, but not in TIG112 (Figure 4A). The results of flow cytometric analysis also indicated apoptosis occurrence in LNM35 cells treated with si*PSMD2*, with a significant increase in the subG1 fraction (Figure 4B). The induction of apoptosis in si*PSMD2*-treated LNM35 cells was further substantiated by Western blot analysis, which detected cleaved forms of an effector caspase, caspase-7, and lamin B, indicating activation of the apoptotic pathway (Figure 4C). Furthermore, we found that a phosphorylated form of p38 was increased as was a phosphorylated form of AKT (Figure 4D), indicating

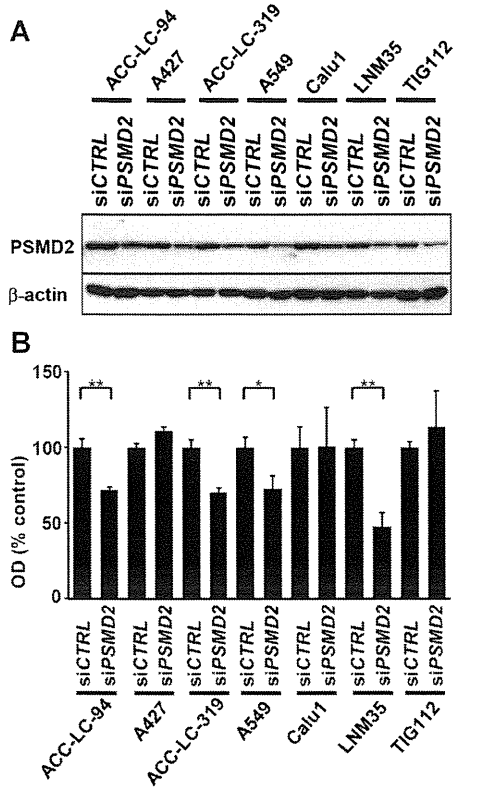

Figure 3. Reduction of cell growth by *PSMD2* knockdown in panel of lung cancer cell lines. (A) Western blot analysis of PSMD2 in cells treated with 10 nM siRNA for 48 h. β-actin was used as a loading control. (B) Results of MTT assay showing varying degrees of cell growth inhibition by si*PSMD2* treatment. OD values with si*CTRL* were regarded as 100%. * and ** indicate the significant growth suppression with the P values of $P < 0.05$ and $P < 0.01$, respectively.

that PSMD2 inhibition affects the crucial balance between active forms of a pro-apoptotic MAPK, p38, and a survival kinase, AKT [18,19]. We also observed an increase in G1 in si*PSMD2*-treated LNM35 cells (Figure 4E), which appeared to be consistent with marked induction of p21, a negative regulator of cell cycle and known target for proteasomes (Figure 4D). In contrast, we did not observe any effects of si*PSMD2* treatment in LNM35 cells on the expression of BCL2, BCL-x, Bad, cyclin A, or cyclin B, whereas p53 was modestly affected by si*PSMD2* treatment (Supplementary Figure 1), consistent with a previous report that p53 is degraded in a PSMD2/S2 dependent manner [20].

Increased Gene Copies of *PSMD2* and Relationship of *PSMD2* Expression with Postoperative Survival in Lung Adenocarcinomas

The present findings that *PSMD2* expression was involved in lung adenocarcinoma cell growth prompted us to analyze the relationship between the

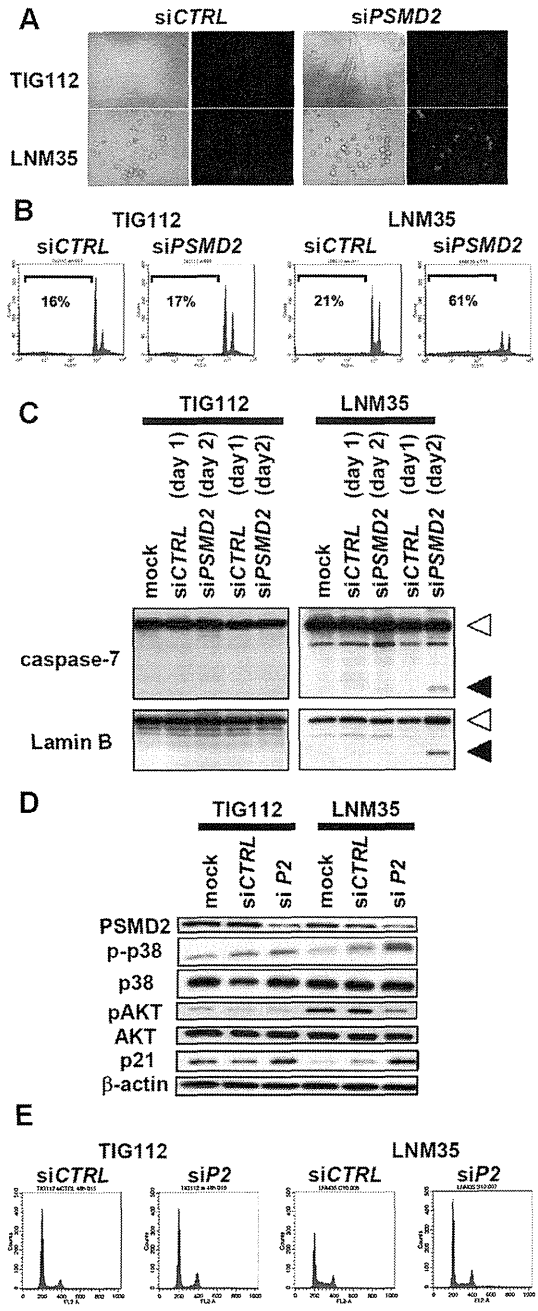

Figure 4. Apoptosis induction and G1 phase accumulation in cells treated with *siPSMD2*. (A) TUNEL assay results showing apoptosis induction in cells treated with siRNA for 48 h. (B) Flow cytometric analysis showing increased subG1 fraction in LNM35 treated with *siPSMD2*. (C) Western blot analysis of proteolytic cleavage of caspase 7 and lamin B in cells treated with siRNAs for the indicated periods. Closed and open arrowheads indicate cleaved and intact forms, respectively. (D) Western blot analysis of various apoptosis- and cell cycle-related proteins in cells after 48 h of siRNA treatment. (E) Flow cytometric analysis of cell cycle progression in cells after 48 h of siRNA treatment.

level of *PSMD2* expression and postoperative survival in surgically treated adenocarcinoma patients using our previously published dataset obtained by microarray analysis. Kaplan–Meier survival curves generated according to expression of *PSMD2* either higher or lower than the median value in the 90 adenocarcinomas revealed poorer prognosis in patients with higher expression ($P = 0.047$ by log-rank test; Figure 5A).

Several studies have shown that genetic mutations that directly activate the PI3K signaling pathway are common in human cancers. Beside the loss of the tumor suppressor PTEN, somatic activating mutations and amplification of *PIK3CA* and *AKT* are occasionally observed in epithelial cancers [21]. Among the PI3K signaling genes, *PIK3CA* resides at 3q26.3, which is close to *PSMD2* at 3q27.1, and locates in the area where gains uniquely related to lung cancer at 3q24–q29 [22–24]. For this reason, we investigated the relationship between the expression levels of *PSMD2* and *PIK3CA*. Interestingly, we found that the *PSMD2* expression was significantly associated with that of *PIK3CA* in the 90 adenocarcinoma cases (Figure 5B). Next, we examined 220 NSCLC cases including 148 with adenocarcinomas for the presence of increased gene copies of *PSMD2* and *PIK3CA* using a TaqMan-based PCR analysis. Our results showed that 12 cases (5.5%) carried more than a 2.5-fold gain at the *PSMD2* locus, while 9 (4.1%) had similar levels of gain at the *PIK3CA* locus (Figure 5C). Modest gains between 1.5- and 2.5-fold for both *PSMD2* and *PIK3CA* were also observed in 38 (17.3%) and 15 (6.8%) cases, respectively. Notably, a single case exhibited 4.7-fold increase only at the *PSMD2* locus without any appreciable increase at *PIK3CA*.

Identification of Co-regulated Expression of Proteasome Pathway Genes, and Association With Expression Profile-defined Subtype and Prognosis in Lung Adenocarcinomas

It is possible that general co-upregulation of proteasome pathway genes is advantageous for cancer development and associated with clinicopathologic features such as postoperative survival. Therefore, we selected a gene set comprised of various 19S and 20S subunits of the 26S proteasome, as well as those functioning in proteasome assembly [25], and performed an unsupervised hierarchical clustering analysis of the 90 adenocarcinoma cases (Figure 6A). We found that the cases could be clearly divided into two groups; those with and without a general high expression of genes involved in proteasome-mediated protein degradation. Cluster 1, consisting of cases with general high expression, showed a significant association with adenocarcinomas with non-terminal respiratory unit (non-TRU)-type histology ($P < 0.001$ by Fisher's exact test), a recently proposed expression Profile-defined subtype, as well as a lack of bronchiole alveolar features ($P = 0.005$;

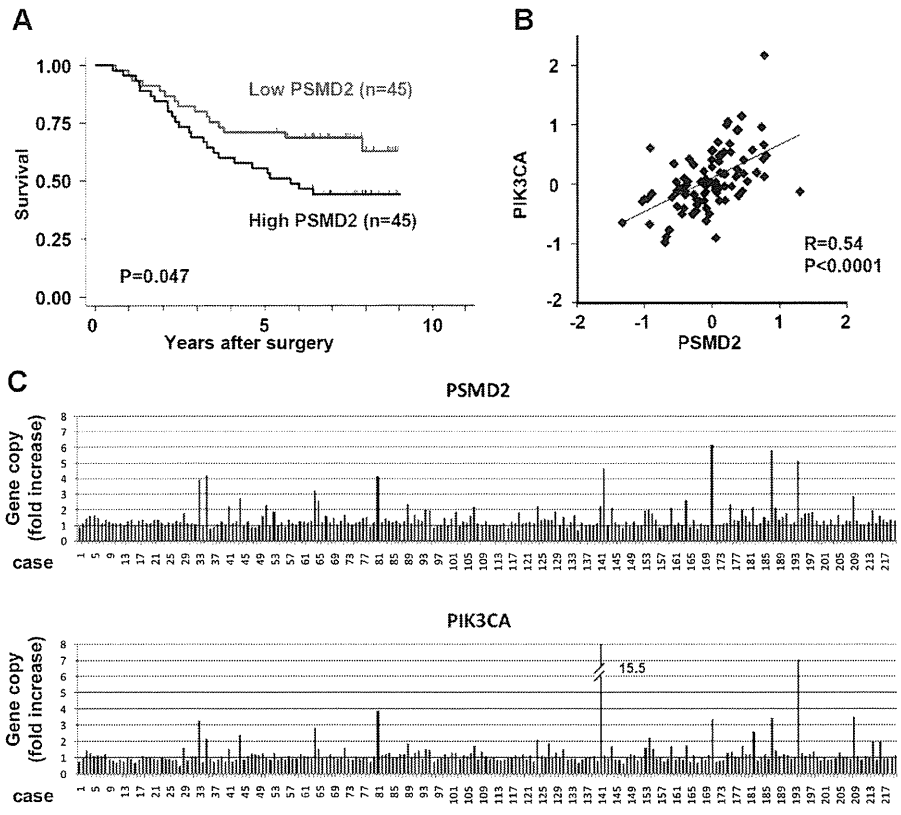

Figure 5. Relationship between *PSMD2* expression and postoperative survival in patients following potentially curative resection. (A) Kaplan–Meier survival curves according to expression levels of *PSMD2* in 90 adenocarcinoma cases [13]. The cases were divided into high and low expression groups based on the median value of *PSMD2* expression. (B) Scatter plot analysis showing a positive correlation between expression levels of *PSMD2* and *PIK3CA* ($R = 0.54$, $P < 0.0001$). (C) TaqMan-based gene dosage analysis of PSMD2 and PIK3CA showing modest increases in a small fraction of NSCLC tissues. ALB signals were used as an internal control for loading. The mean value obtained from 48 normal lung tissues was set as 1.

Table 1). In addition, we observed significant associations of cluster 1 with past smoking history ($P = 0.020$) and higher histological grade ($P = 0.031$). Kaplan–Meier survival curves revealed that patients with adenocarcinomas belonging to cluster 1 had a significantly shorter survival than that of those in cluster 2 ($P = 0.0001$ by log-rank test; Figure 6B). Multivariate Cox regression analysis also showed that cluster distinction based on the proteasome-mediated protein degradation pathway was a significant predictor of postoperative prognosis (Hazard ratio = 3.372; $P = 0.001$), independent of p-Stage (Hazard ratio = 3.135; $P = 0.001$; Table 2).

DISCUSSION

It has been suggested that PSMD2 is a functional equivalent to yeast proteins of RPN1/NAS1 in *S. cerevisiae* and Mts4 in *S. pombe*, based on findings that enforced expression of *PSMD2* suppressed growth defects in both *rpn1/nas1* and *mts4* disrupted yeast cells [16,17]. In the 19S regulatory complex, RPN1/NAS1 is a component of the base complex, which is required for substrate translocation and gating of the proteolytic channel, and possibly serves as a docking site for a substrate-recruitment factor in the 19S complex [26]. In the present study, we found that siRNA-mediated inhibition of *PSMD2* in human lung cancer cells resulted in a reduction of proteasome activity, in association with induction of apoptosis and increase in number of cells in the G1 phase. A decrease in phosphorylated AKT and increase in phosphorylated p38 in si*PSMD2*-treated cells, as well as induction of p21 appears to be consistent with the treatment effects. A future study to elucidate how such molecular consequences are mediated in *PSMD2*-inhibited lung cancer cells is warranted.

Interestingly, *PSMD2* has been reported to constitute cancer-associated signatures including those related to undifferentiated cancers, metastatic phenotypes, and prognosis in breast cancer [8,27–29]. Those previous studies identified either *PSMD2* alone, or *PSMD2* and *PSMD7* among various

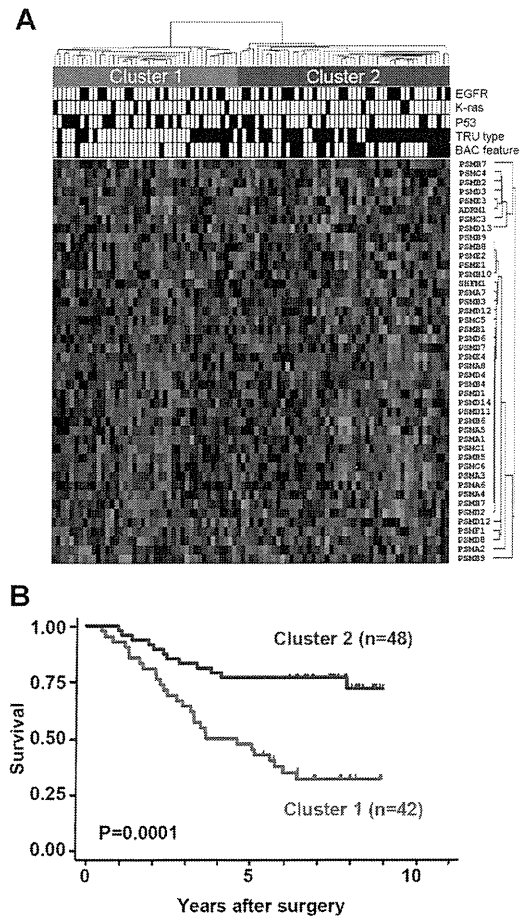

Figure 6. Relationships among expressions of proteasome pathway genes and various clinicopathologic features. (A) Hierarchical clustering analysis of 90 adenocarcinoma cases showing co-upregulation of a majority of the proteasome pathway genes. (B) Kaplan–Meier survival curves according to the two major clusters in (A) showing a significant difference in postoperative survival ($P = 0.0001$ by log-rank test).

Table 1. Associations Among Two Major Clusters and Clinicopathologic Features

Variables	Cluster 1	Cluster 2	*P*-Value*
Age (yr)			
>61	18	22	0.834
≤61	24	26	
Sex			
Male	26	21	0.096
Female	16	27	
Smoking status			
Current/former	27	18	0.020
Never	15	30	
Histologic grade			
Poor/moderate	21	13	0.031
Well	21	35	
p-Stage			
II or III	20	18	0.395
I	22	30	
EGFR			
Mutant	13	19	0.509
Wild-type	29	29	
K-ras			
Mutant	6	4	0.505
Wild-type	36	44	
p53			
Mutant	18	11	0.070
Wild-type	24	37	
EP-defined subtype[a]			
TRU	15	38	<0.001
Non-TRU	27	10	
BAC features			
+	6	20	0.005
–	36	28	

*Fisher's exact test.
[a]EP-defined subtype, expression profile-defined subtype of adenocarcinomas (13).

members of the *PSMD* family. In addition, Wan et al. [30] conducted a large-scale screening study and reported that introduction of *PSMD2* caused stimulated colony formation of NIH3T3 cells by more than six-fold, suggesting that *PSMD2* offers a growth advantage to cells under certain conditions. In the present study, though *PSMD2* and *PIK3CA* expressions were significantly correlated and often had the same increased gene copies, there was an exception of lung cancer specimen with increased gene copies only at the *PSMD2* locus, suggesting that *PSMD2* may not be a mere bystander of *PIK3CA*, a well established cancer associated gene.

Our results clearly demonstrated that genes encoding components of the 26S proteasome and those involved in proteasome assembly are co-regulated in lung adenocarcinomas, showing a significant association between general high expression of the proteasome pathway genes and shorter postoperative survival. These suggest that such co-regulated upregulation may confer greater advantage to cancer cell growth during cancer development. Further, they indicated the possibility of a common transcriptional regulatory mechanism(s) among the proteasome-related genes, which may be an effective target(s) for diminishing proteasome activity for cancer treatment.

It is also interesting to note that a general high expression of proteasome-related genes was highly significantly associated with the non-TRU type of adenocarcinomas. We previously proposed a TRU/non-TRU classification as a genetically and clinically relevant grouping of adenocarcinoma patients [31,32], which was further substantiated by a report of their marked distinctions in expression profiles [13]. Similar observations were subsequently reported by other groups, confirming the existence of adenocarcinoma subtypes with distinct expression profiles similarly associated with various

Table 2. Results of Multivariate Cox Regression Analysis

Variables	Unfavorable/favorable	Hazard ratio	95% CI*	*P*-Value
Age (yr)	>61/≤61	1.158	0.565–2.373	0.689
Sex	Male/female	0.960	0.406–2.273	0.926
Smoking status	Current or former/never	0.660	0.252–1.732	0.399
Histologic grade	Poor or moderate/well	1.830	0.898–3.727	0.096
p-Stage	II or III/I	3.135	1.591–6.177	0.001
EP-defined subtype[a]	Non-TRU/TRU	1.376	0.637–2.973	0.416
Cluster	Cluster 1/cluster 2	3.372	1.643–6.921	0.001

*95% CI, 95% confidence interval.
EP-defined subtype, expression profile-defined subtype of adenocarcinomas (13).

clinicopathologic features [33–35]. Following our previous report of a significant association of TRU-type with EGFR mutations [13,31], the present findings add another layer of molecular distinctions to the TRU/non-TRU classification of adenocarcinomas, that is, prevalent high expression of proteasome pathway genes specifically in the non-TRU type, which suggests that proteasome inhibitors may be more effective for this adenocarcinoma subtype.

Taken together, the present study results show that PSMD2 may be a good candidate for development of novel therapy for lung cancer treatment. In addition, the findings of our expression profile analysis indicate that other co-regulated proteasome pathway genes and/or common regulator(s) of their expression may be relevant potential target(s).

ACKNOWLEDGMENTS

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (C), and a grant from the Novartis Foundation (Japan) for the Promotion of Science.

REFERENCES

1. Osada H, Takahashi T. Genetic alterations of multiple tumor suppressors and oncogenes in the carcinogenesis and progression of lung cancer. Oncogene 2002;21:7421–7434.
2. Adams J. The proteasome: A suitable antineoplastic target. Nat Rev Cancer 2004;4:349–360.
3. Peters JM, Cejka Z, Harris JR, Kleinschmidt JA, Baumeister W. Structural features of the 26 S proteasome complex. J Mol Biol 1993;234:932–937.
4. Brooks P, Fuertes G, Murray RZ, et al. Subcellular localization of proteasomes and their regulatory complexes in mammalian cells. Biochem J 2000;346:155–161.
5. Davies AM, Lara PN, Jr., Mack PC, Gandara DR. Incorporating bortezomib into the treatment of lung cancer. Clin Cancer Res 2007;13:s4647–s4651.
6. Orlowski RZ, Stinchcombe TE, Mitchell BS, et al. Phase I trial of the proteasome inhibitor PS-341 in patients with refractory hematologic malignancies. J Clin Oncol 2002; 20:4420–4427.
7. Richardson PG, Barlogie B, Berenson J, et al. A phase 2 study of bortezomib in relapsed, refractory myeloma. N Engl J Med 2003;348:2609–2617.
8. Tomida S, Yanagisawa K, Koshikawa K, et al. Identification of a metastasis signature and the DLX4 homeobox protein as a regulator of metastasis by combined transcriptome approach. Oncogene 2007;26:4600–4608.
9. Kozaki K, Miyaishi O, Tsukamoto T, et al. Establishment and characterization of a human lung cancer cell line NCI-H460-LNM35 with consistent lymphogenous metastasis via both subcutaneous and orthotopic propagation. Cancer Res 2000;60:2535–2540.
10. Nakagawa T, Hayashita Y, Maeno K, et al. Identification of decatenation G2 checkpoint impairment independently of DNA damage G2 checkpoint in human lung cancer cell lines. Cancer Res 2004;64:4826–4832.
11. Tokumaru S, Suzuki M, Yamada H, Nagino M, Takahashi T. let-7 regulates Dicer expression and constitutes a negative feedback loop. Carcinogenesis 2008;29:2073–2077.
12. Matsubara H, Takeuchi T, Nishikawa E, et al. Apoptosis induction by antisense oligonucleotides against miR-17-5p and miR-20a in lung cancers overexpressing miR-17-92. Oncogene 2007;26:6099–6105.
13. Takeuchi T, Tomida S, Yatabe Y, et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol 2006;24:1679–1688.
14. Tomida S, Takeuchi T, Shimada Y, et al. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. J Clin Oncol 2009;27: 2793–2799.
15. Tanaka H, Yanagisawa K, Shinjo K, et al. Lineage-specific dependency of lung adenocarcinomas on the lung development regulator TTF-1. Cancer Res 2007;67:6007–6011.
16. Tsurumi C, Shimizu Y, Saeki M, et al. cDNA cloning and functional analysis of the p97 subunit of the 26S proteasome, a polypeptide identical to the type-1 tumor-necrosis-factor-receptor-associated protein-2/55.11. Eur J Biochem 1996;239:912–921.
17. Wilkinson CR, Wallace M, Seeger M, Dubiel W, Gordon C. Mts4, a non-ATPase subunit of the 26 S protease in fission yeast is essential for mitosis and interacts directly with the ATPase subunit Mts2. J Biol Chem 1997;272:25768–25777.
18. Wagner EF, Nebreda AR. Signal integration by JNK and p38 MAPK pathways in cancer development. Nat Rev Cancer 2009;9:537–549.
19. Lowe SW, Cepero E, Evan G. Intrinsic tumour suppression. Nature 2004;432:307–315.
20. Zhang X, Turnell AS, Gorbea C, Mymryk JS, Gallimore PH, Grand RJ. The targeting of the proteasomal regulatory subunit S2 by adenovirus E1A causes inhibition

of proteasomal activity and increased p53 expression. J Biol Chem 2004;279:25122–25133.
21. Engelman JA, Luo J, Cantley LC. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. Nat Rev Genet 2006;7:606–619.
22. Lo KC, Stein LC, Panzarella JA, Cowell JK, Hawthorn L. Identification of genes involved in squamous cell carcinoma of the lung using synchronized data from DNA copy number and transcript expression profiling analysis. Lung Cancer 2008;59:315–331.
23. Pei J, Balsara BR, Li W, et al. Genomic imbalances in human lung adenocarcinomas and squamous cell carcinomas. Genes Chromosomes Cancer 2001;31:282–287.
24. Petersen I, Bujard M, Petersen S, et al. Patterns of chromosomal imbalances in adenocarcinoma and squamous cell carcinoma of the lung. Cancer Res 1997;57:2331–2335.
25. Tanaka K. The proteasome: Overview of structure and functions. Proc Jpn Acad Ser B Phys Biol Sci 2009;85:12–36.
26. Elsasser S, Gali RR, Schwickart M, et al. Proteasome subunit Rpn1 binds ubiquitin-like protein domains. Nat Cell Biol 2002;4:725–730.
27. Naderi A, Teschendorff AE, Barbosa-Morais NL, et al. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene 2007;26:1507–1516.
28. Rhodes DR, Yu J, Shanker K, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci USA 2004;101:9309–9314.
29. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–536.
30. Wan D, Gong Y, Qin W, et al. Large-scale cDNA transfection screening for genes related to cancer development and progression. Proc Natl Acad Sci USA 2004;101:15724–15729.
31. Yatabe Y, Kosaka T, Takahashi T, Mitsudomi T. EGFR mutation is specific for terminal respiratory unit type adenocarcinoma. Am J Surg Pathol 2005;29:633–639.
32. Yatabe Y, Mitsudomi T, Takahashi T. TTF-1 expression in pulmonary adenocarcinomas. Am J Surg Pathol 2002;26:767–773.
33. Shibata T, Hanada S, Kokubu A, et al. Gene expression profiling of epidermal growth factor receptor/KRAS pathway activation in lung adenocarcinoma. Cancer Res 2007;98:985–991.
34. Hayes DN, Monti S, Parmigiani G, et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. J Clin Oncol 2006;24:5079–5090.
35. Motoi N, Szoke J, Riely GJ, et al. Lung adenocarcinoma: Modification of the 2004 WHO mixed subtype to include the major histologic subtype suggests correlations between papillary and micropapillary adenocarcinoma subtypes, EGFR mutations and gene expression analysis. Am J Surg Pathol 2008;32:810–827.

OPEN ACCESS Freely available online

PLoS one

A Novel Network Profiling Analysis Reveals System Changes in Epithelial-Mesenchymal Transition

Teppei Shimamura[1]*, Seiya Imoto[1], Yukako Shimada[2], Yasuyuki Hosono[2], Atsushi Niida[1], Masao Nagasaki[1], Rui Yamaguchi[1], Takashi Takahashi[2], Satoru Miyano[1]

1 Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo, Japan, **2** Nagoya University Graduate School of Medicine, Showa-ku, Nagoya, Japan

Abstract

Patient-specific analysis of molecular networks is a promising strategy for making individual risk predictions and treatment decisions in cancer therapy. Although systems biology allows the gene network of a cell to be reconstructed from clinical gene expression data, traditional methods, such as Bayesian networks, only provide an averaged network for all samples. Therefore, these methods cannot reveal patient-specific differences in molecular networks during cancer progression. In this study, we developed a novel statistical method called NetworkProfiler, which infers patient-specific gene regulatory networks for a specific clinical characteristic, such as cancer progression, from gene expression data of cancer patients. We applied NetworkProfiler to microarray gene expression data from 762 cancer cell lines and extracted the system changes that were related to the epithelial-mesenchymal transition (EMT). Out of 1732 possible regulators of E-cadherin, a cell adhesion molecule that modulates the EMT, NetworkProfiler, identified 25 candidate regulators, of which about half have been experimentally verified in the literature. In addition, we used NetworkProfiler to predict EMT-dependent master regulators that enhanced cell adhesion, migration, invasion, and metastasis. In order to further evaluate the performance of NetworkProfiler, we selected Krueppel-like factor 5 (KLF5) from a list of the remaining candidate regulators of E-cadherin and conducted *in vitro* validation experiments. As a result, we found that knockdown of KLF5 by siRNA significantly decreased E-cadherin expression and induced morphological changes characteristic of EMT. In addition, *in vitro* experiments of a novel candidate EMT-related microRNA, miR-100, confirmed the involvement of miR-100 in several EMT-related aspects, which was consistent with the predictions obtained by NetworkProfiler.

Citation: Shimamura T, Imoto S, Shimada Y, Hosono Y, Niida A, et al. (2011) A Novel Network Profiling Analysis Reveals System Changes in Epithelial-Mesenchymal Transition. PLoS ONE 6(6): e20804. doi:10.1371/journal.pone.0020804

Editor: Eric J. Bernhard, National Cancer Institute, United States of America

Received November 2, 2010; **Accepted** May 13, 2011; **Published** June 7, 2011

Funding: This research was supported by "Research and Development of the Next-Generation Integrated Simulation of Living Matter" (a part of the Development and Use of the Next-Generation Supercomputer Project of MEXT) and "Integrative Systems Understanding of Cancer for Advanced Diagnosis, Therapy and Prevention" (Grant-in-Aid for Scientific Research on Innovative Areas from MEXT, Japan). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: shima@ims.u-tokyo.ac.jp

Introduction

Currently, several large-scale omics projects, such as the National Cancer Institute's Cancer Genome Atlas (http://cancergenome.nih.gov/) and the Sanger Institute's Cancer Genome Project (http://www.sanger.ac.uk/genetics/CGP/), produce large amounts of data, including genomic, epigenomic, and transcriptomic information, about cancer patients or cell lines. Two challenges in omics are to construct and analyze patient-specific molecular networks to develop a comprehensive understanding of the molecular mechanisms of tumorigenesis and to identify molecules that are critical for tumor proliferation and progression [1]. If these challenges can be overcome, it may be possible to personalize cancer therapy, improve its efficacy, and reduce its toxicity and cost [2,3].

Systems biology integrates various types of omics data and computational tools to represent and analyze complex biological systems. For example, gene network estimation that is based on Bayesian networks or mutual information networks can reconstruct biological systems from gene expression data [4]. However, most traditional gene network estimation methods construct a static network by using gene expression data from different cellular conditions. As a result, these methods only produce an averaged network for all patients and cannot reveal patient-specific molecular mechanisms of cancer. In addition, it is very difficult to infer a patient-specific gene network from only a few gene expression profiles of the patient without making any assumptions about the network.

In this study, we developed a novel statistical method called NetworkProfiler, which infers patient-specific gene regulatory networks from a dataset of cancer gene expression profiles. NetworkProfiler is based on a statistical graphical model with varying coefficients and a kernel-based data integration method with elastic net regularization for parameter estimation. A key feature of NetworkProfiler is that the strengths of the relationships between genes are allowed to vary depending on cancer characteristics, such as cancer progression, metastasis, disease-free survival, and drug sensitivity. NetworkProfiler groups samples according to the specific cancer characteristics so that neighboring samples have common gene regulatory systems. Then, by integrating the gene expression profiles of neighboring samples with a kernel method, NetworkProfiler produces a gene regulatory network for each sample. Finally, we analyzed 2 post-analysis to discover upstream regulatory genes and downstream target genes for specific cancer characteristics. Network-

Profiler is the first algorithm for constructing patient-specific gene regulatory networks from clinical cancer gene expression data to elucidate cancer heterogeneity.

We applied NetworkProfiler to gene expression microarray data from 762 cancer cell lines to determine system changes related to the epithelial-mesenchymal transition (EMT). The epithelial-mesenchymal transition (EMT) is a process that changes proliferating cells from an aplanetic state to a motile state [5], which allows cancer cells to leave the primary tumor and metastasize. The loss of E-cadherin, a cell adhesion molecule, is a biomarker of EMT [5]. NetworkProfiler identified 25 key regulators of E-cadherin, of which half have been previously described and the other half were novel candidates. NetworkProfiler also revealed regulatory changes in *miR-141*, *ZEB1*, and E-cadherin. Specifically, our results suggested that decreased expression of *miR-141* in mesenchymal cells disrupts the negative feedback loop between *miR-141* and *ZEB1*, which would allow *ZEB1* to decrease the expression of E-cadherin during the EMT. In addition, we predicted 45 EMT-dependent putative master regulators that control sets of genes involved in cell adhesion, migration, invasion and metastasis, namely, 17 of which are downstream targets of TGFB1, a master switch of the EMT. To further validate the performance of NetworkProfiler, we experimentally evaluated *in silico* predictions obtained by NetworkProfiler. We consequently found that knockdown of KLF5, a new candidate regulator of E-cadherin, decreased E-cadherin expression and induced morphological changes characteristic of EMT. In addition, the functional involvement of miR-100 was validated in some EMT-related aspects, which was consistent with the predictions obtained by Network Profiler.

Results

Overview of NetworkProfiler

Here, we provide an overview of NetworkProfiler; please refer to the Methods section for a complete description. NetworkProfiler is a modulator-dependent graphical model because it includes a modulator (M) variable in addition to regulator (R) and target (T) variables (genes). R controls the transcription of T and M is a cofactor that modulates the interaction between R and T. In this study, we defined M as a biological or a clinical feature that is related to cancer, such as drug response, survival risk, or a molecule or pathway that is related to cancer initiation, progression, or metastasis. The relationships between R, T, and M are illustrated in Figure 1a. As shown in Figure 1b, the strength of the relationship between R and T varies depending on the value of M. Thus, M does not affect R and T directly; instead, it influences the strength of the relationship between R and T. In contrast, existing graphical models, such as Bayesian networks and mutual information networks [4], do not consider the effect of M (Figure 1c), so the strength of the relationship between R and T remains constant for all values of M (Figure 1d).

In addition, NetworkProfiler can infer the relationships between R and T, given a value of M. As a result, we could use NetworkProfiler to construct patient-specific networks with varying R-T relationships that reflect changes in the feature of interest in cancer patients. A simple example with synthetic data for R, T, and M is shown in Figure 2a. In this example, we assume that R regulates T only with a high value of M (Figure 2b). In this case, most existing methods that only consider R and T in all of the samples (Figure 2c) and ignore M would conclude that R does not regulate T. In contrast, NetworkProfiler attempts to quantify the strength of the relationship between R and T for a specific value m of M by reweighting the data according to the value of M to identify the neighborhood of samples with values of M that are close to m. Then, NetworkProfiler measures the dependency between R and T on the basis of these neighboring samples. The optimization of the size of the neighborhood is explained in the Method section.

A schematic representation of the entire analytical process of NetworkProfiler is shown in Figure 3. NetworkProfiler used 2 inputs: (1) gene expression data and (2) the modulator for each sample (Figure 3a). The gene expression data was represented as a $p \times n$ matrix, where p is the number of genes and n is the number

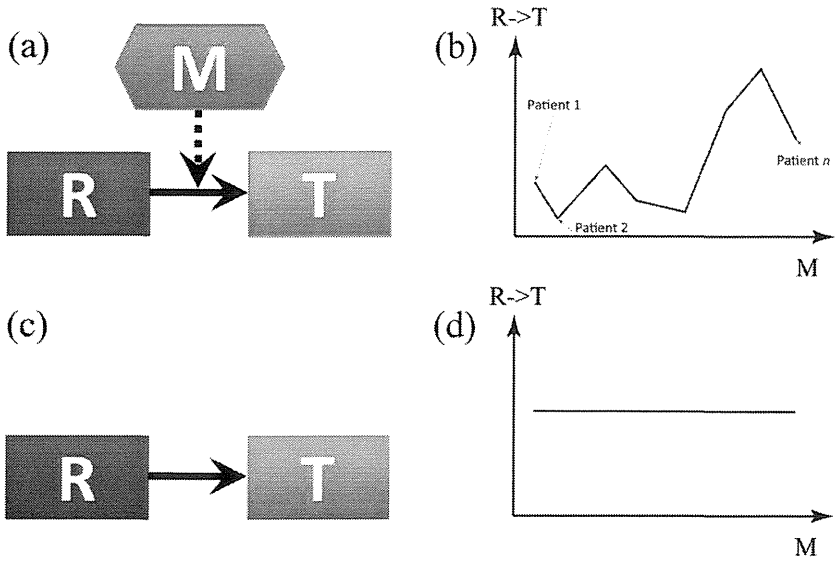

Figure 1. The relationships between a regulator (R), a target(T), and a modulator (M) in NetworkProfiler and existing graphical models. (a). The relationships between R, T and M in NetworkProfiler. The directed solid-line edge from R to T represents "R regulates the transcript of T". The directed dot-line edge from M to the edge between R and T describes "M controls the strength of the relationship between R and T". (b). The strength of the relationship between R and T in NetworkProfiler that varies depending on the value of M. (c). The relationships between R and T in existing graphical models that do not consider the effect of M. (d). The strength of the relationship between R and T in existing graphical models that remains constant for all values of M.
doi:10.1371/journal.pone.0020804.g001

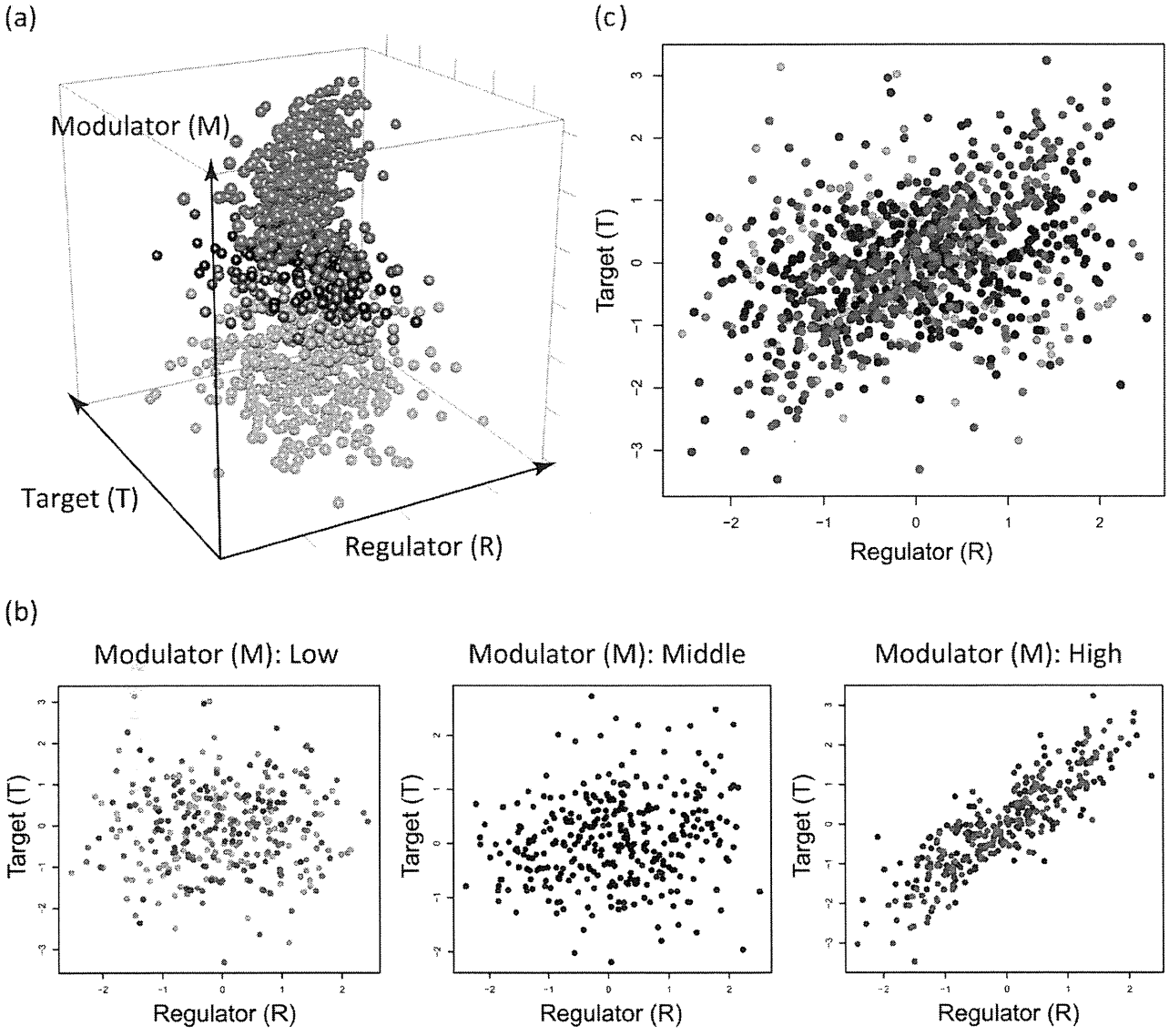

Figure 2. A regulatory change between a regulator (R) and a target (T) depending on the value of a modulator M. (a). A simple example with synthetic data from 1000 samples for R, T, and M where x-, y-, and z-axises correspond to the expressions of R and T, and the values of M, respectively. (b). The 3 scatter plots of R and T that are conditioned on the value of M. The left, middle, and right figures represent the scatter plots from 1-st sample to 333-th sample, from 334-th sample to 666-th sample, and from 667-th sample to 1000-th sample in order of ascending M, respectively. (c). The scatter plot of R and T that are not conditioned on the value of M.
doi:10.1371/journal.pone.0020804.g002

of samples (patients). If the modulator was an observable variable, then we directly applied NetworkProfiler to these inputs. However, if the modulator was a variable that is difficult to observe, then we used a signature-based hidden modulator extraction algorithm to estimate the value of the modulator. The output of NetworkProfiler is a set of gene networks for every value of M (i.e., sample-specific gene networks) shown in Figure 3b.

Afterwards, we used 2 post-analysis techniques to extract biological information from the networks. The first technique identified upstream regulators of a target gene of interest in the constructed modulator-dependent gene networks. To evaluate the modulator-dependent strength of a regulator for the target gene, we created a measure called the regulatory effect. The regulatory effect profiles of the upstream regulators for specific target genes are shown in Figure 3c. The second technique discovered putative master regulators that control downstream target gene sets with previously curated functions. To evaluate the enrichment of the target genes on a functional gene set, we created measure called the enrichment score. The resulting regulator-function matrix (Figure 3d) illustrates the candidate regulators (rows) of functions (columns) that are enhanced in the target genes.

Identification of system changes in the epithelial-mesenchymal transition

To identify system changes during the EMT, we applied NetworkProfiler to gene expression profiles of 762 cancer cell lines from the Sanger Cell Line Project (http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi). This dataset included the expression profiles of 22,777 probes, which correspond to 13,006 mRNAs in these cancer cell lines from the Affymetrix GeneChip

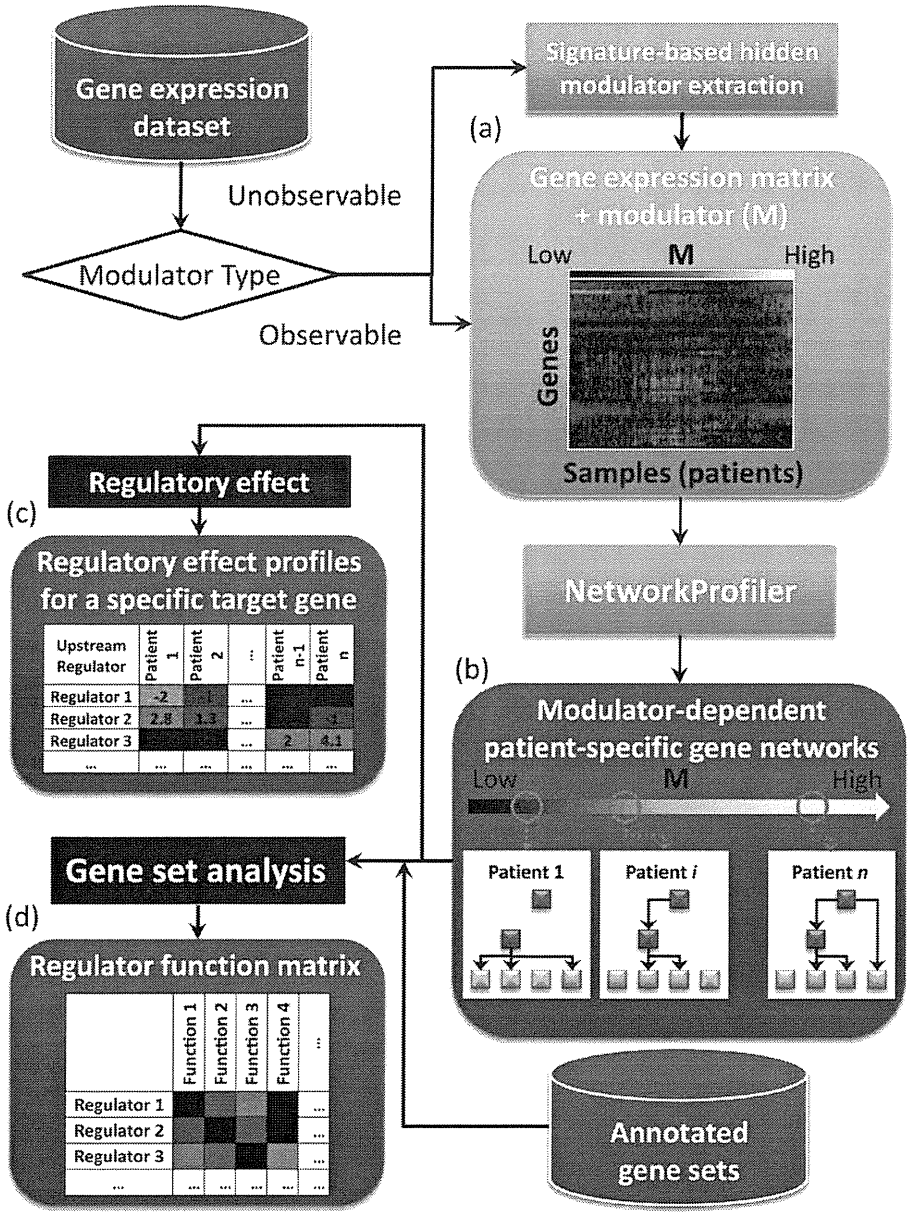

Figure 3. A schematic representation of the entire analytical process of NetworkProfiler. (a). Inputs of NetworkProfiler: gene expression data matrix and the modulator for each sample. (b). Outputs of NetworkProfiler: a set of gene networks for every value of M (i.e., sample-specific gene networks). (c). The regulatory effect profiles of the upstream regulators for a specific target gene. (d). The resulting regulator function matrix whose columns are the candidate regulators and rows are functions that are enhanced in the target genes.
doi:10.1371/journal.pone.0020804.g003

Human Genome U133 Array Set (HG-U133A) and the expression profiles of 502 human microRNAs from bead-based oligonucleotide arrays. The MAS5-normalized mRNA dataset was further transformed to the log scale and quantile-normalized. During the mapping of the probes to genes, we selected 1 probe for each gene that had the largest variance, which produced a final 13,508 (genes) × 762 (cancer cell lines) gene expression matrix.

In this study, we considered transcription factors, nuclear receptors, and microRNAs to be potential regulators. To identify transcription factors and nuclear receptors, we selected human genes that were annotated as a "transcription regulator" or "ligand-dependent nuclear receptor" from the Ingenuity Knowledge Base (IKB; http://www.ingenuity.com). We also included some transcription factors that were not annotated in the IKB but were annotated in the Biobase Knowledge Library (BKL; http://biobase-international.com/). We mapped a total of 1230 genes in the HG-U133A microarray gene set to 1183 transcription factors and 47 nuclear receptors (Table S1). In addition, we included 502 human miRNA probes (Table S2).

To calculate the modulator values for the EMT in the 762 cancer cell lines, we applied a signature-based hidden modulator extraction algorithm (see Methods for details) to the expression data. First, we selected 122 genes labeled "EMT_UP", "EMT_DN", "JECHLINGER_EMT_UP", and "JECHLINGER_EMT_DN" from Molecular Signatures Database v2.5 ([6]; http://www.broadinstitute.org/gsea/msigdb/index.jsp). Then, this algorithm narrowed the set to

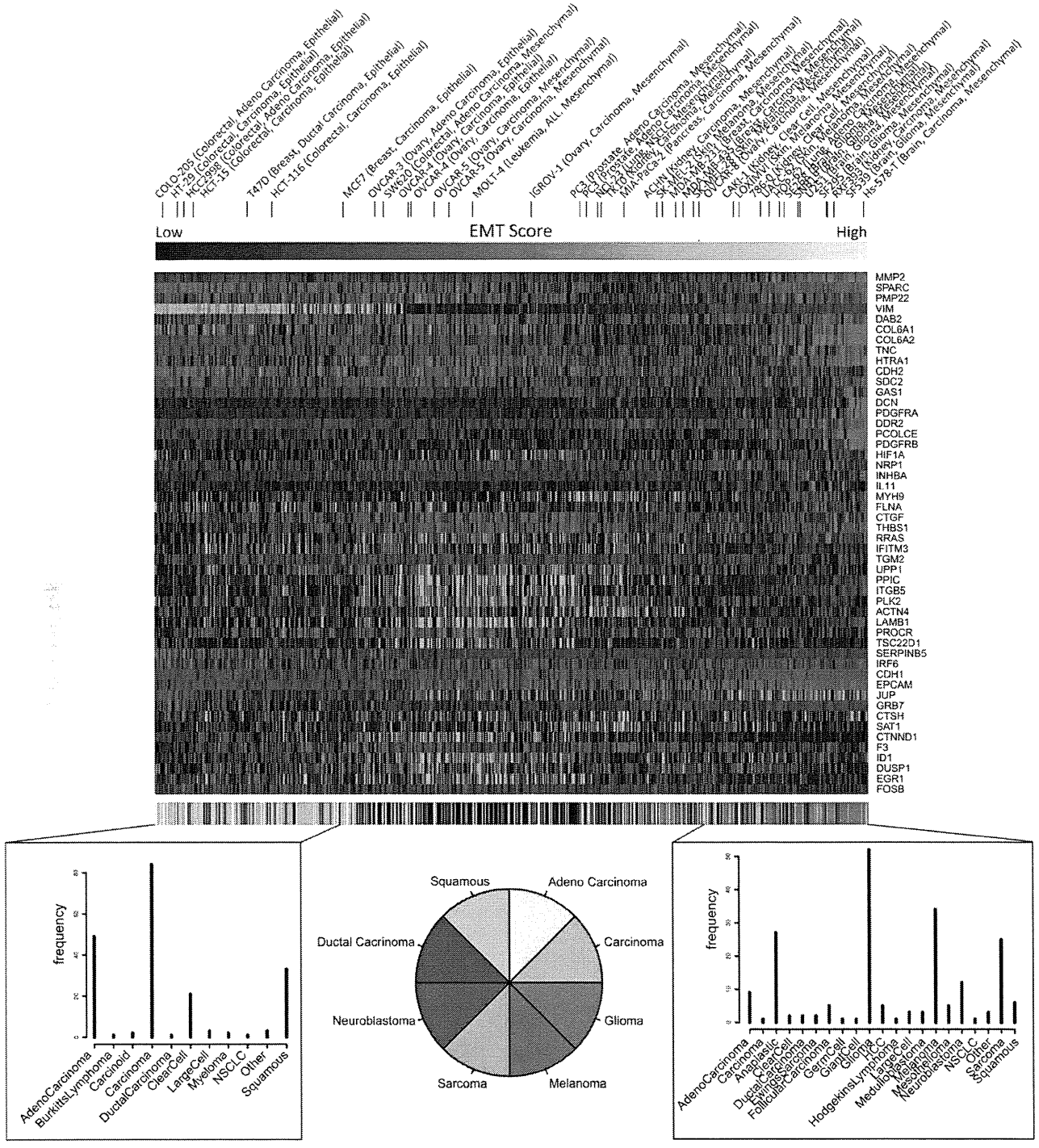

Figure 4. Expression profiles of the 50 functionally coherent genes in ascending order of the EMT-related modulator values. The heatmap represents normalized expression profiles so that the mean and variance for each gene are 0 and 1, respectively. The red color represents positive expressions and the green color represents negative expressions. The upper strings indicate cell line names which are known to be epithelial or mesenchymal. The upper horizontal color bar represents the values of the EMT-related modulator with the signature-based hidden modulator extraction algorithm. The bottom horizontal color bar shows primary histories of 762 cancer cell lines whose color corresponds to one of the eight primary histories or the other histories (black). The bottom histograms represent frequencies of the primary histories between samples with the 200 lowest and 200 highest values of the EMT-related modulator, respectively.
doi:10.1371/journal.pone.0020804.g004

50 functionally coherent genes with $p<10^{-5}$ by using the extraction of expression module (EEM) [7] (Table S3) and computed the first principal component of these 50 genes as hidden values of the EMT-related modulator (Table S4). Since the direction of the first principal component did not always correspond to that of the EMT, we changed the sign of the modulator values by

multiplying either plus or minus one so that epithelial-like cells have lower modulator values than mesenchymal-like cells.

Figure 4 shows the expression profiles of the 50 functionally coherent genes in ascending order of the EMT-related modulator values. These modulator values clearly discriminated cell lines that were epithelial-like or mesenchymal-like. Specifically, cells with smaller or larger modulator values had more epithelial or mesenchymal phenotypes, respectively. Furthermore, many carcinomas and squamous tumors had low modulator values, while many gliomas and melanomas had high values. By using these EMT-related modulator values, NetworkProfiler constructed 762 regulatory gene networks that are related to the EMT. The list of the estimated edges in each of these networks can be downloaded from the supporting web site (Files S1, S2, and S3; http://bonsai.hgc.jp/~shima/NetworkProfiler).

Identification of regulators of E-cadherin that induce the epithelial-mesenchymal transition

To identify possible regulators that might control the expression of E-cadherin during the EMT, we calculated the regulatory effects of the upstream regulators of E-cadherin. Out of 1732 potential regulators, NetworkProfiler inferred that 370 of them may control the expression of E-cadherin in any of the 762 cancer cell lines (Table S5). These putative regulators were ranked according to the change in their regulatory effect during the EMT. Although we did not include any information on known E-cadherin regulators, about half of the 25 highest ranked regulators were previously reported in the literature (Table 1). For example, 2 zinc finger transcription factors, ZEB1 and ZEB2, are direct repressors of E-cadherin and are involved in the EMT [9,15]. In addition, the miR-200 family indirectly suppresses the EMT by inhibiting the translation of ZEB1 and ZEB2 mRNAs [8]. Similarly, miR-192 inhibits the translation of ZEB2 [13,14]. In addition, SNAI2, a member of the Snail superfamily of zinc finger transcription factors, also is involved in the EMT [16]. Likewise, TCF4 (also known as E2-2), a class I bHLH transcription factor, is an EMT regulator; its isoforms induce the EMT in MDCK kidney epithelial cells [12]. In contrast, FOXA1 and FOXA2 are positive regulators of E-cadherin, which suppress the EMT in pancreatic ductal adenocarcinoma [11]. KLF4 also inhibits the EMT by regulating E-cadherin expression [10]. NetworkProfiler also identified several other known direct repressors of E-cadherin, such as TWIST1 [17] and TCF3 (also known as E47) [18]; however, these regulators were ranked 38th and 84th, respectively.

The other half of the 25 highest ranked regulators has not yet been reported and may be novel EMT-dependent regulators of E-cadherin. For example, although the relationship between GRHL2 and EMT is not known, GRHL2 is required for morphogenesis of epidermal and tracheal cells and plays an important role in regulating the expression levels of E-cadherin in *Drosophila* post-embryonic neuroblasts [19]. ZNF217 binds the E-cadherin promoter [20], which suggests that ZNF217 might be a transcription factor for E-cadherin.

Next, we compared the performance of NetworkProfiler with that of a structural equation model (SEM) of E-cadherin that was inferred by the elastic net [22]. This model was equivalent to a regression model where the response variable is the expression of E-cadherin and the explanatory variables are the 1732 regulator expressions. The significance of each regulator was evaluated based on the number of non-zero regression coefficients in 1000 bootstrapped datasets. The SEM inferred 627 putative regulators (Table S6). Among these putative regulators, there were only 6 regulators, namely, *ZEB1*, *miR-141*, *ZEB2*, *TCF3*, *miR-200b*, and *miR-200c*, in the 25 highest ranked regulators that were previously reported in the literature. This result suggested that NetworkProfiler was superior to the traditional gene network estimation methods to identify regulators of E-cadherin that are involved in the EMT. Moreover, NetworkProfiler can reveal regulatory changes among genes during the EMT. Figures 5a and 5b show the regulatory profiles of putative regulators of E-cadherin when the lengths of the paths from the regulators to E-cadherin is 1 and 2, respectively.

Table 1. 25 top-ranked regulators of E-cadherin for the change in the regulatory effect change among the EMT with published evidence.

regulator	type	regulatory effect change	Evidence
IRF6	A	101.04	
miR-141	A	87.58	[8]
GRHL2	A	64.13	
ZEB1 (SIP1)	I	50.72	[9]
LSR	I	46.89	
miR-200b	A	31.55	[8]
KLF4	A	26.28	[10]
OVOL2	A	22.08	
miR-200a	A	17.70	[8]
FOXA2	A	17.26	[11]
TCF4 (E2.2)	I	14.15	[12]
ELF3	A	13.58	
ZNF217	A	13.53	
MYB	A	12.50	
KLF5	A	12.42	
miR-192	A	12.30	[13, 14]
FOXA1	A	11.69	[11]
ZNF165	A	11.39	
NKX2-1	A	11.21	
HNF1B	A	11.08	
TFE3	A	11.01	
ZEB2 (δEF)	I	10.66	[15]
TRIM29	I	9.87	
SNAI2	I	9.74	[16]

The labels "A" and "I" indicate 2 types of the regulator: activator (A) and inhibitor (I). See Table S5 for the complete table of the 370 putative regulators for E-cadherin.

doi:10.1371/journal.pone.0020804.t001

NetworkProfiler can also predict mechanistic interpretations of published experiments. For example, it is known that ZEB1 and ZEB2 induce EMT by repressing E-cadherin transcription and that ectopic expression of the miR-200 family (miR-200a, miR-200b, miR-200c, and miR-141) or miR-205 leads to downregulation of ZEB1 and ZEB2, upregulation of E-cadherin, and mesenchymal-epithelial transition (MET) in cells [8]. As the relationships between these genes, the prediction of NetworkProfiler provides the following results. As shown in Figures 6c and 6d, although the expression of miR-141 had a strong positive effect on that of E-cadherin in epithelial-like cells, this effect decreases during the EMT. In contrast, although the expression of ZEB1 had a weak negative effect on that of E-cadherin in epithelial-like cells, this effect increased during the EMT. Interestingly, miR-141 and ZEB1 had a strong, direct

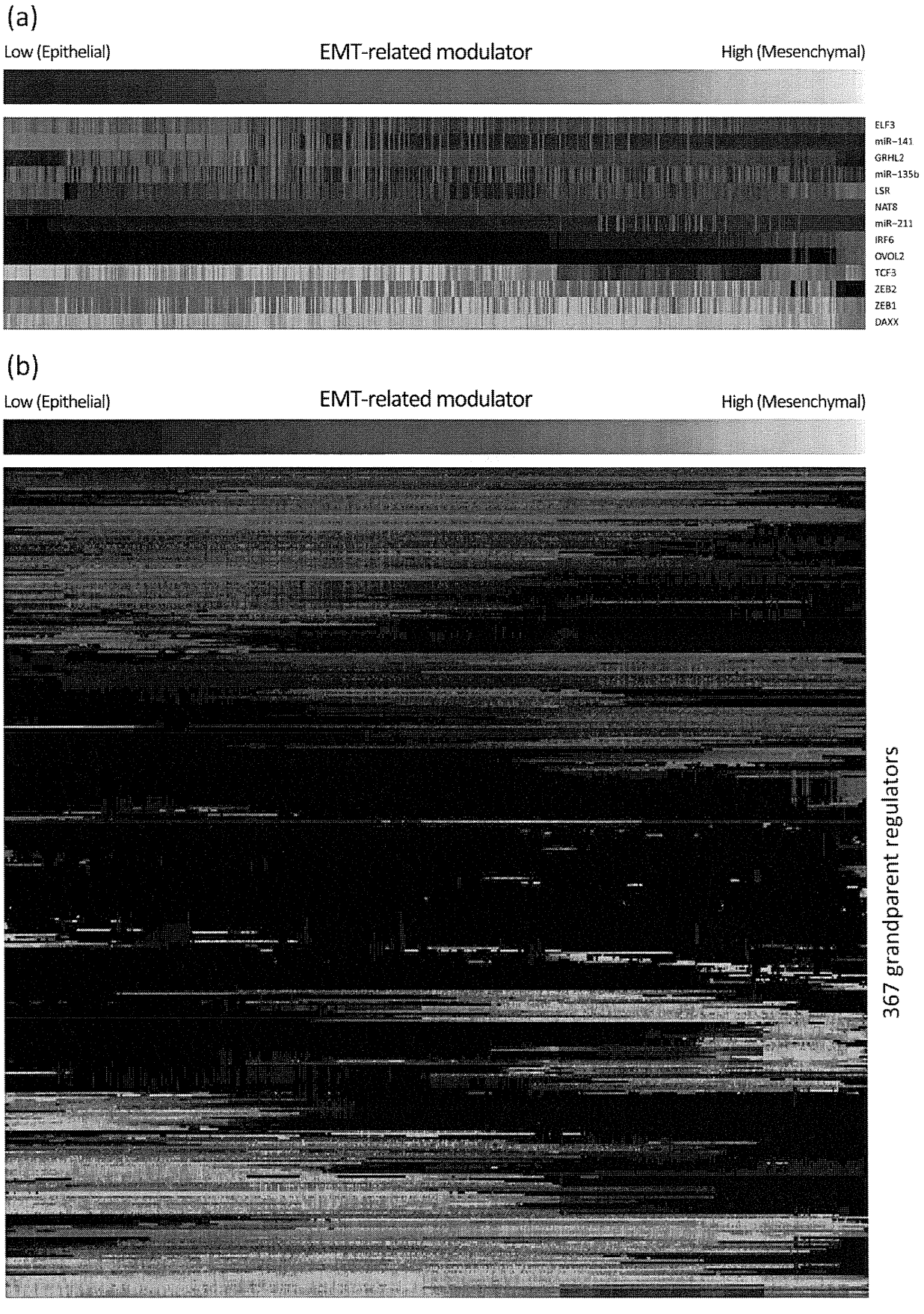
(a)
Low (Epithelial)
EMT-related modulator
High (Mesenchymal)
ELF3
miR-141
GRHL2
miR-135b
LSR
NAT8
miR-211
IRF6
OVOL2
TCF3
ZEB2
ZEB1
DAXX
(b)
Low (Epithelial)
EMT-related modulator
High (Mesenchymal)
367 grandparent regulators

Figure 5. Regulatory effect profiles of the putative regulators of E-cadherin among the EMT. (a). The regulatory effect profiles of the 13 putative regulators among the EMT when the length of the paths from the regulators to E-cadherin is 1 where rows indicate the putative regulators of E-cadherin and columns indicate samples (cancer cell lines). The positive (red) and negative (green) regulatory effect indicate that the parent regulator controls the transcript of E-cadherin positively and negatively, respectively. (b). The regulatory effect profiles of the 13 putative regulators among the EMT when the length of the paths from the regulators to E-cadherin is 2.
doi:10.1371/journal.pone.0020804.g005

negative effect on each other only when the EMT-related modulator values were low. This implied that there is a negative feedback loop between miR-141 and ZEB1 in epithelial-like cells, which is consistent with a previous study [23]. Furthermore, during the EMT, the expression levels of miR-141 and E-cadherin decreased, while the expression level of ZEB1

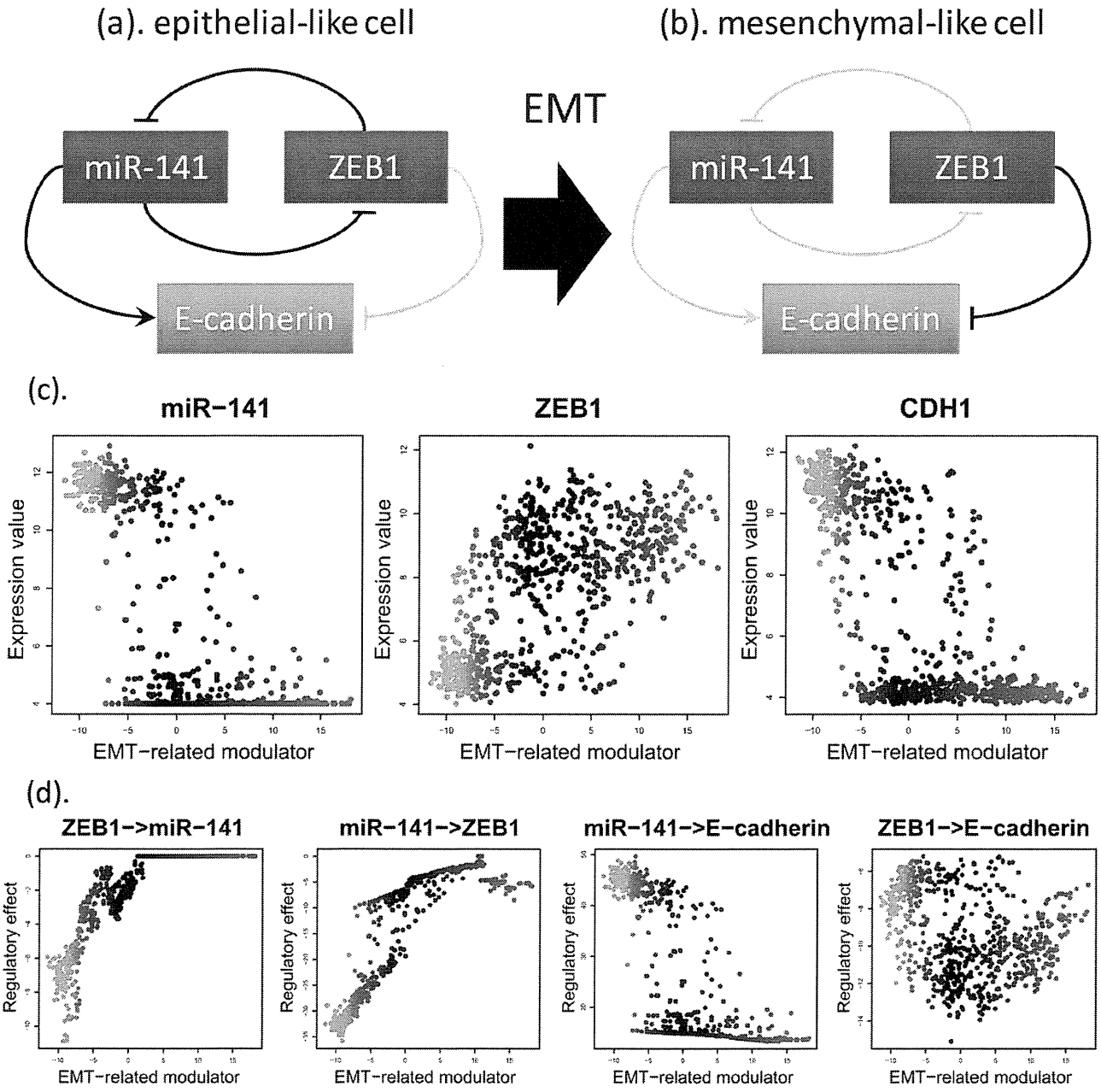

Figure 6. Regulatory changes among miR-141, ZEB1, and E-cadherin among the EMT. (a). The relationship among miR-141, ZEB1, and E-cadherin in epithelial-like cells. (b). The relationship among miR-141, ZEB1, and E-cadherin in mesenchymal-like cells. (c). The expression profiles of miR-141 (left), ZEB1 (middle), and E-cadherin (right) in order of ascending the EMT-related modulator values. The green and red colors indicate epithelial- and mesenchymal-like cells, respectively. (d). The regulatory effects from ZEB1 to miR-141, from miR-141 to ZEB1, from miR-141 to E-cadherin, and from ZEB1 to E-cadherin when the length of the paths is 1.
doi:10.1371/journal.pone.0020804.g006

Table 2. Selected relationships between the 47 putative master regulators and the 5 functional categories with published evidence.

regulator	function	-$\log_{10}$(q-value)	mode of action (E⇒M)					evidence
			A⇑	A⇓	I⇑	I⇓	–	
FOSL1	migration	9.82	29	2	42	3	41	[25]
	invasion	8.42	14	2	24	3	22	[26]
EPAS1	adhesion	5.90	26	1	10	0	16	[27]
	migration	7.66	32	1	14	0	24	[28]
KLF5	migration	5.93	28	2	27	5	25	[29]
AHR	metastasis	3.67	12	0	11	0	9	[30]
FOXF1	metastasis	6.10	24	0	9	0	8	[31]
	migration	6.09	29	0	17	0	14	[32]
ELK3	migration	6.23	41	8	17	7	19	[33]
SMAD3	adhesion	4.57	9	3	23	0	10	[34]
	metastasis	3.12	5	1	12	1	9	[35]
	migration	5.24	14	5	26	1	21	[36]
	EMT	2.47	1	1	2	0	0	[37]
WWTR1	migration	5.08	32	0	17	3	16	[38]
	invasion	3.48	17	0	8	2	5	[38]
hsa-miR-145	invasion	2.52	13	0	8	3	17	[39]
CEBPD	metastasis	4.88	17	2	10	0	9	[31]
TGFB1I1	adhesion	5.12	25	2	23	5	11	[40]
HIF1A	adhesion	3.84	10	0	25	3	10	[27]
	metastasis	4.45	14	1	14	0	8	[41]
	migration	5.00	18	3	25	4	21	[42]
	invasion	3.65	12	0	9	3	10	[43]
SNAI2	migration	3.45	36	2	25	14	25	[25]
ELF3	adhesion	7.87	24	4	24	11	14	[44]
	invasion	4.45	9	3	18	6	21	[44]
SOX9	adhesion	6.80	18	2	19	0	26	[45]
	migration	5.46	28	2	15	1	23	[46]
GLI3	migration	4.53	24	7	24	7	26	[47]
TCF7L2	migration	4.52	19	10	18	1	27	[48]
NFKBIA	adhesion	2.73	12	2	14	3	12	[49]
	metastasis	2.39	5	0	5	3	9	[50]
	migration	3.98	18	2	18	7	23	[51]
	invasion	2.69	9	2	5	2	12	[50]
VAV1	adhesion	5.51	3	5	15	3	14	[52]
	migration	5.10	7	10	16	5	16	[53]
JUN	adhesion	3.03	15	4	6	5	6	[54]
	migration	3.31	19	2	7	7	14	[25]
	invasion	2.07	8	2	7	2	5	[55]
ETV1	invasion	2.50	13	1	13	5	7	[56]
PDLIM1	adhesion	4.27	16	6	17	6	29	[57]
MAFB	metastasis	4.41	9	0	3	8	6	[31]
GATA6	metastasis	3.25	11	3	4	1	4	[31]
RUNX1	adhesion	6.27	15	5	16	12	14	[58]
	migration	2.46	23	7	20	7	20	[59]
YAP1	migration	3.30	7	2	20	0	9	[60]

The labels "A⇑", "A⇓", "I⇑", and "I⇓", and "–" indicate the number of the five modulator modes of action for the relationship between a regulator and its target included in the functional gene set: "the activation of a regulator on the expressions of its target genes with the functional category was increased by the modulator", "inhibition increased", "activation decreased", "inhibition decreased", and "the modulator mode of action is not determined", respectively.

doi:10.1371/journal.pone.0020804.t002

increased. These results suggested that reduced expression of miR-141 disrupts the negative feedback loop between miR-141 and ZEB1 (Figures 6a and 6b), which would allow ZEB1 to decrease the expression of E-cadherin, as illustrated in Figure 6c. It should be noted that these results cannot be predicted by traditional graphical models which infer a static gene network structure.

Identification of relationships between regulators and epithelial-mesenchymal transition-related functional gene sets

The EMT-dependent relationships between downstream target genes for each regulator and previously curated functional gene sets in each sample were analyzed by applying gene set analysis (see Methods for details) to the constructed gene networks for 762 cancer cell lines. We tested five curated gene sets included in Ingenuity Knowledge Base (IKB; http://www.ingenuity.com). These gene sets were related with *adhesion*, *migration*, *invasion*, and *metastasis* which were hallmarks of EMT [5], and EMT itself. By using gene set analysis, the statistical significances (q-values) for the enrichments of downstream genes for the 1732 regulators on the five functional gene sets were calculated in each of the 762 cell lines. These results can be downloaded from the supporting web site (File S4; http://bonsai.hgc.jp/~shima/NetworkProfiler).

To search for regulators that strongly affected the five EMT-related functional gene sets, the change in the enrichment score during the EMT and their integral q-value were calculated. The result was summarized by a regulator function matrix (Table S7). We focused on 45 regulators with the integral q-values less than 10^{-10} as putative master regulators that strongly enhanced the

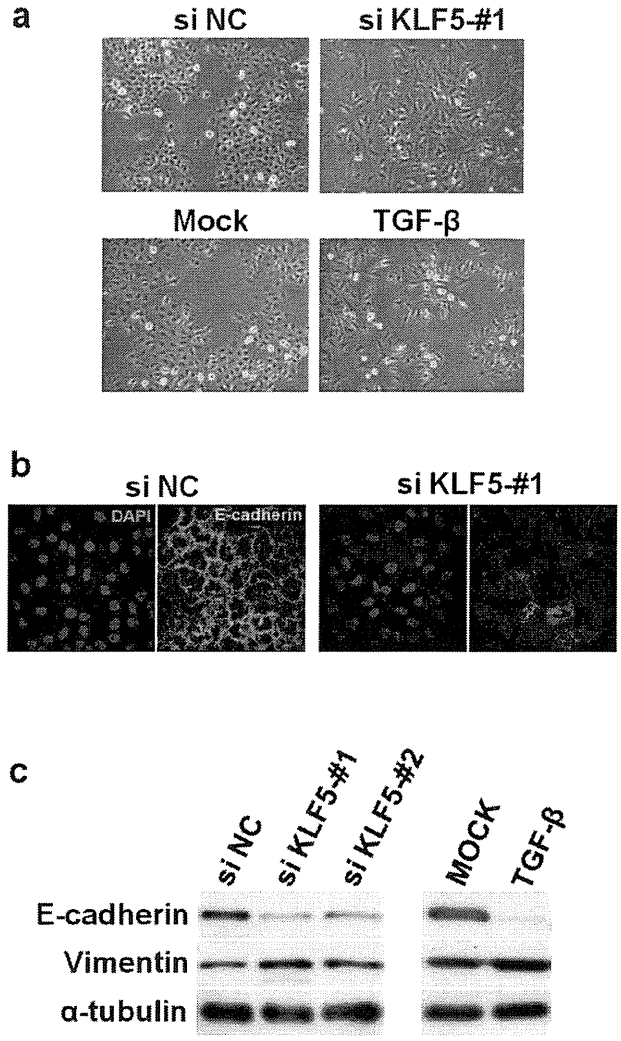

Figure 7. Induction of EMT by KLF5 knockdown in A549 NSCLC cell line. (a) Phase contrast images of A549 cells 72 hours after siRNA transfection, showing a fibroblast-like morphology in siKLF5 treated cells. TGF-β treatment serves as a positive control for EMT induction in A549 cells. (b) Representative immunofluorescence staining images, showing reduced E-cadherin expression in siKLF5-treated A549 cells. (c) Western blot analysis of E-cadherin and vimentin, showing EMT-related changes in their expression in A549 cells treated with two differenct siRNAs.
doi:10.1371/journal.pone.0020804.g007

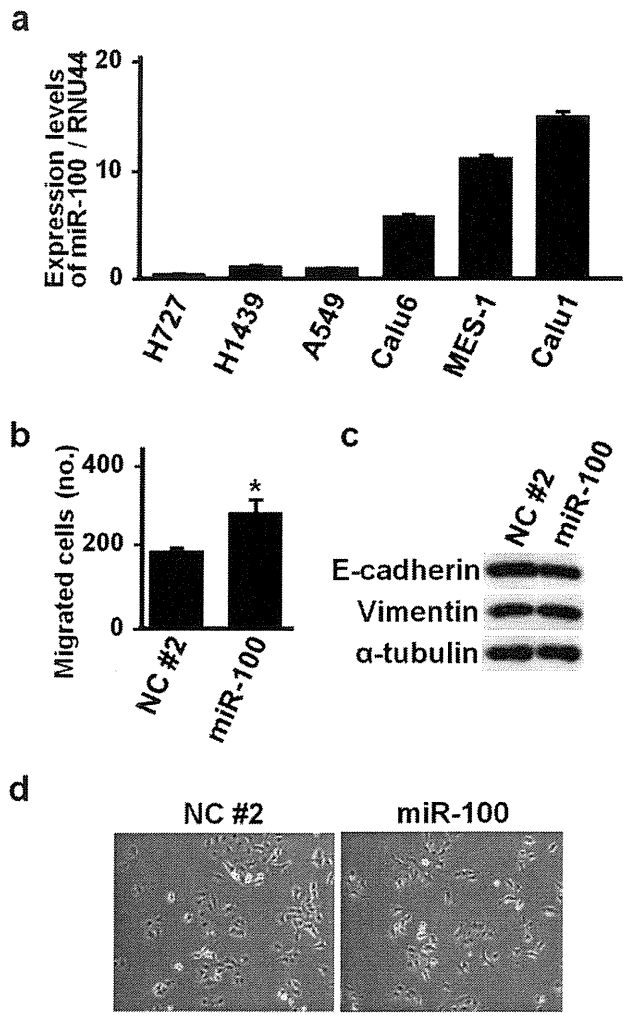

Figure 8. miR-100-induced changes in biologic characteristics in A549 NSCLC cell line. (a) Quantitative real-time RT-PCR analysis of miR-100 in six NSCLC cell lines, showing low miR-100 expression in A549, NCI-H727 and NCI-H1437. (b) Motility assay showing increased migration in miR-100-transfected A549 cells. Error bars indicate SE in three independent experiments (*, $p<0.05$). NC#2, negative control. (c) Western blot analysis of E-cadherin, vimentin and α-tubulin, showing lack of noticeable changes in miR-100-transfected A549 cells (d) Representative phase contrast microscopic images showing negligible changes in miR-100-trasfected A549 cells.
doi:10.1371/journal.pone.0020804.g008

functional gene sets related with the EMT. Interestingly, among the 45 regulators, 17 regulators were downstream targets of transforming growth factor β-1 (TGFB1), a master switch of EMT [24], with published evidence (Table S8). This result suggests that these regulators have crucial roles in TGFB1-induced EMT.

As a control, we tested how well the NetworkProfiler analysis identified known relationships between regulators and functional gene sets in the Ingenuity Knowledge Base. The known functional relationships of the 45 putative master regulators are shown in Table 2. For example, FOSL1 increases the migration of MDA-MB-436 cells [25] and the invasion of A549 cells [26]. SMAD3 increases the adhesion [34], the metastasis [35], and the migration [36] of cells, respectively. Similarly, HIF1A increases the adhesion of undifferentiated trophoblast stem cells [27], the metastasis of LM2 cells [41], the migration of HUVEC cells [42], and the invasion of Achn cells [43], respectively.

Although some of the 47 putative master regulators have not been reported to enhance the EMT-related functions in IKB, some predictions were supported by other resent works which were not included in IKB. For example, the prediction of NetworkProfiler suggested that PTRF regulates gene sets related with migration (q-value = 2.45×10^{-8}) and with metastasis (q-value = 2.03×10^{-6}) during the EMT. Consistent with the *in silico* result, PTRF expression inhibits migration and correlates with metastasis in PC3 prostate cancer cells [61]. Similarly, NetworkProfiler predicted that miR-146 contributes to migration (q-value = 3.27×10^{-9}) and invasion (q-value = 1.01×10^{-4}) during the EMT. This *in silico* result is comparable with the *in vitro* result that miR-146 inhibits invasion and migration, and acts as a metastasis suppressor [62]. In addition, some predictions between miRNAs and functions seem reasonable based on the known functions of the miRNA host genes. For example, the prediction of NetworkProfiler provided the hypothesis that miR-143 and miR-145 promotes metastasis (q-value = 7.17×10^{-4} and 3.15×10^{-5}) and migration (q-value = 1.37×10^{-6} and 6.10×10^{-8}), respectively. miR-143 and miR-145 cooperatively target a network of transcription factors, such as KLF4, to control smooth muscle phenotype switching [63]. Since KLF4 increases the migration of cells [29] and induces EMT [10], these miRNAs might be related with EMT-related functions or control EMT by targeting KLF4. Again, it should be noted that these relationships between regulators and functions cannot be predicted from one gene network constructed by traditional graphical models, and only the results of multiple network comparison between epithelial-like and mesenchymal-like cells based on NetworkProfiler enables us to support the recent biological knowledge and new hypotheses about unknown relationships.

Comparison between *in silico* predictions and *in vitro* results

To validate the performance of NetworkProfiler, *in silico* predictions obtained by NetworkProfiler were evaluated experimentally. We first conducted *in vitro* experiments of a new candidate regulator of E-cadherin listed in Table 1, KLF5, to investigate whether KLF5 affects E-cadherin expression and induces morphologic changes characteristic of EMT using A549 lung adenocarcinoma cell line, which is well known to exhibit EMT in response to TGF-β [64]. KLF5 knockdown markedly altered a cobblestone epithelial morphology of A549 cells and induced a more fibroblast-like morphology with reduced cell-cell contact, which was similar to that seen in TGF-β-treated A549 cells (Figure 7a and Figure S1). Immunofluorescence analysis showed significant reduction of E-cadherin expression in A549 cells knocked down for KLF5 (Figure 7b), which was also confirmed by western blot analysis (Figure 7c). Conversely, vimentin expression was shown to be modestly increased by siKLF5 treatment (Figure 7c). Consistent with the *in vitro* results, the prediction of NetworkProfiler suggested that KLF5 affects E-cadherin expression as well as Vimentin expression during the EMT, since the changes in the regulatory effects from KLF5 to E-cadherin and Vimentin were much larger compared with the other regulators (12.42 and 16.57, respectively) which was ranked 15-th and 10-th among the 1732 regulators (Table S9). The result of gene set analysis (Table S7) also suggested that KLF5 affects EMT (q-value = 1.60×10^{-24}). Thus, we consequently found that *in silico* predictions obtained by NetworkProfiler was confirmed with the results of *in vitro* experiments; KLF5, a newly identified candidate regulator of EMT, was shown to affect expressions of E-cadherin and Vimentin as well as morphologic characteristics related to EMT as a repressor of EMT.

We also conducted *in vitro* experiments to validate functional involvement of a novel candidate EMT-related microRNA, miR-100 whose expression was increased in 762 cancer cell lines during the EMT (Figure S2). miR-100 was found to be expressed at a low level in A549, NCI-H727 and NCI-H1439 NSCLC cell lines, which had low EMT-related modulator values among the 762 cell lines panel (Figure 8a). miR-100 was transiently introduced into A549 cells, resulting in a significant increase of cell migration activity (Figure 8b). However, overexpression of miR-100 did not affect expressions of an epithelial marker, E-cadherin, and a mesenchymal marker, vimentin (Figure 8c), and also did not influence cell morphology (Figure 8d). However, overexpression of miR-100 significantly increased cell migration without noticeably affecting morphology in NCI-H727 and NCI-H1437 cells (Figure S3). Consistent with the *in vitro* results, the prediction of NetworkProfiler suggested that miR-100 enhances migration (q-value = 1.42×10^{4}) but does not affect EMT itself (q-value = 0.24) from gene set analysis (Table S7). It also suggested that miR-100 does not affect the expressions of E-cadherin and Vimentin during the EMT, since E-cadherin and Vimentin were not target genes of miR-100 in all the 762 cell line-specific gene networks related with the EMT(Files S1, S2, and S3) and the changes in the regulatory effects from miR-100 to E-cadherin and Vimentin were much smaller compared with the other regulators (0 and 1.72, respectively), which were ranked 371-th and 151-th among the 1732 regulators (Table S9). Thus, we conclude that several hypotheses of miR-100 functions provided by NetworkProfiler are consistent with the results of *in vitro* experiments; NetworkProfiler has the potential to uncover novel biological mechanisms.

Discussion

We developed a novel algorithm called NetworkProfiler to infer patient-specific, modulator-dependent gene regulatory networks from gene expression data. Unlike traditional methods that infer a static network for a specific state of a cell or an averaged network for many patients, NetworkProfiler can be used to construct patient-specific gene networks for specific diseases, such as cancer. Subsequently, information about the regulatory effects of individual genes and functional gene sets can be extracted from these networks. In order to show the performance of NetworkProfiler, we applied NetworkProfiler to microarray gene expression data from 762 cancer cell lines to identify the system changes that were related to the EMT. As a result, we identified 25 EMT-dependent regulators of E-cadherin. Although some of these regulators have been reported in the literature, others may be novel master regulators of E-cadherin that induce the EMT. Moreover, in comparison to the