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

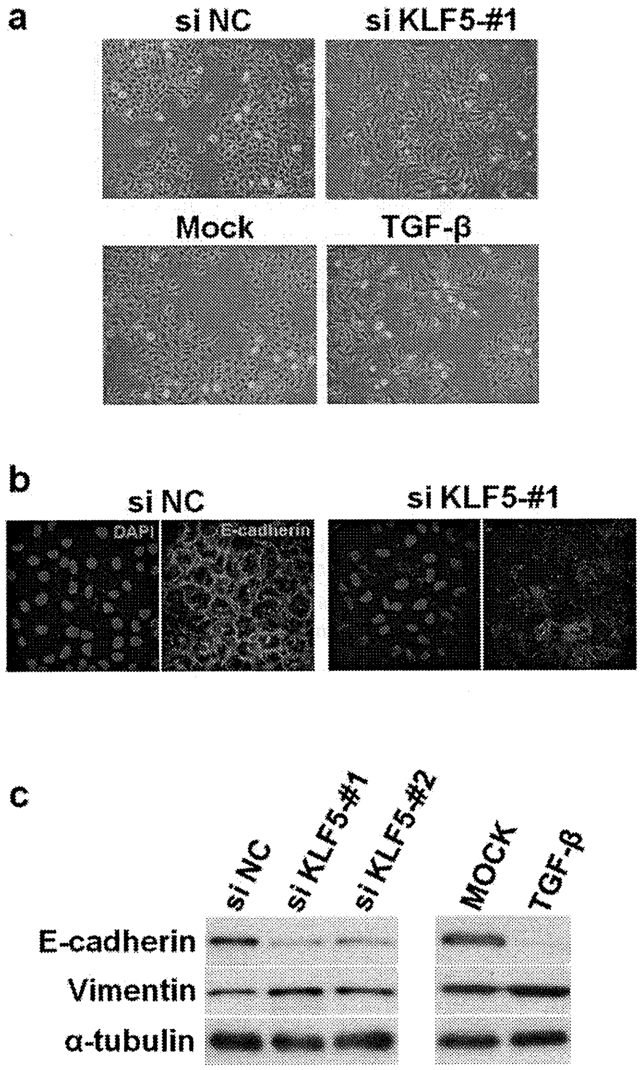

Figure 7. Induction of EMT by KLF5 knockdown in A549 NSCLC cell line. (a) Phase contrast images of A549 cells 72 hours after siRNA transfection, showing a fibroblast-like morphology in siKLF5 treated cells. TGF-β treatment serves as a positive control for EMT induction in A549 cells. (b) Representative 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

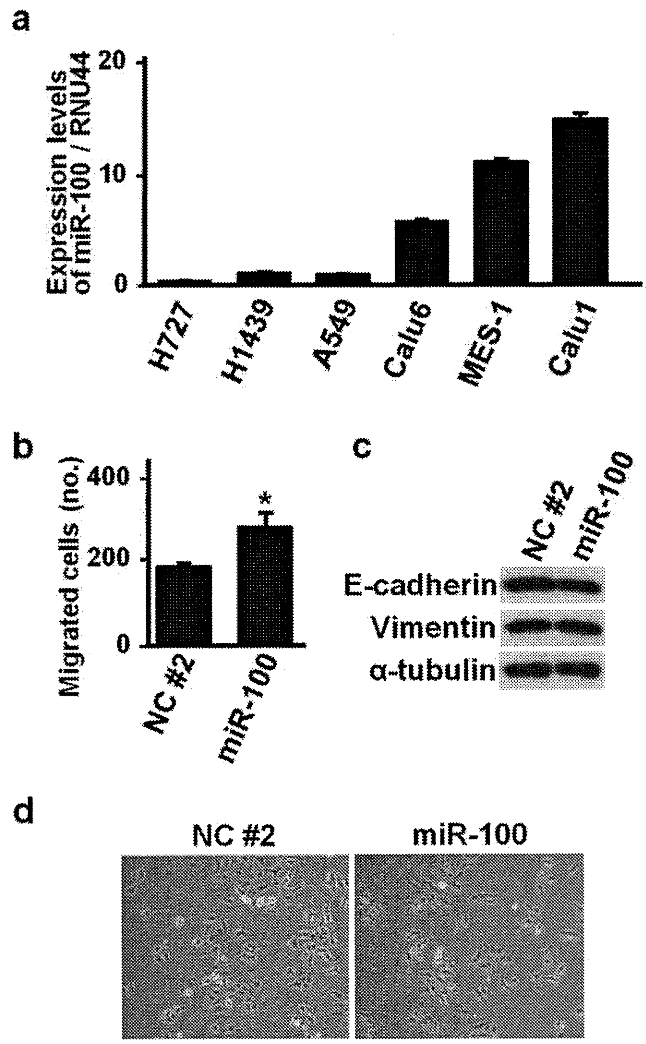

Figure 8. miR-100-induced changes in biologic characteristics in A549 NSCLC cell line. (a) Quantitative real-time RT-PCR analysis of miR-100 in six NSCLC cell lines, showing low miR-100 expression in A549, NCI-H727 and NCI-H1437. (b) Motility assay showing increased migration in miR-100-transfected A549 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 $= 2.45 \times 10^{-8}$) and with metastasis (q-value $= 2.03 \times 10^{-6}$) 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 $= 3.27 \times 10^{-9}$) and invasion (q-value $= 1.01 \times 10^{-4}$) 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 $= 7.17 \times 10^{-4}$ and 3.15×10^{-5}) and migration (q-value $= 1.37 \times 10^{-6}$ 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 $= 1.60 \times 10^{-24}$). 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 $= 1.42 \times 10^{4}$) 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 traditional SEM approach, the performance of NetworkProfiler was superior for identifying regulators of E-cadherin during the EMT. We also showed that NetworkProfiler can reveal regulatory changes of E-cadherin during the EMT. In particular, our results suggested that decreased expression of miR-141 disrupts the negative feedback loop between miR-141 and ZEB1, which would allow ZEB1 to decrease the expression of E-cadherin.

Furthermore, we also identified putative relationships between regulators and EMT-dependent functional gene sets, some of which had published evidence. Based on the significance of the enrichment of downstream target genes for the regulator on the 5 functional gene sets, we identified 45 putative master regulators for the EMT. We found that 17 regulators were downstream targets of TGFB1 that is a master switch of the EMT. We then showed that NetworkProfiler can not only predict the relationships between these regulators and functions that were supported by many published evidence, but also produce new hypotheses that some of them might enhance EMT-related functions or induce EMT.

Finally, it is of note that we were able to validate the *in silico* predictions obtained by NetworkProfiler in our *in vitro* experiments. KLF5, a newly identified candidate regulator of EMT, was experimentally shown to affect E-cadherin expression as well as morphologic characteristics related to EMT, validating the NetworkProfiler-based prediction that KLF5 is a negative regulator of EMT. We also conducted *in vitro* experiments of another regulator, miR-100, for which NetworkProfiler predicted its association with some EMT-associated functions. As a result, we found that the predicted miR-100 functions conformed to the results of *in vitro* experiments. Thus, we conclude that the effectiveness of the proposed method was validated not only from published literature but also from new *in vitro* validation experiments.

We anticipate several possible applications and extensions of NetworkProfiler. In this study, we only focused on the system changes that are associated with the EMT. NetworkProfiler also could be used to infer system changes and reconstruct modulator-dependent gene networks for other well-defined modulators, such as drug sensitivity and prognosis risk. Currently, a significant limitation of NetworkProfiler is that the modulator must be one-dimensional. However, cancer development is a multivariate process. It may be possible to use multivariate kernel functions in NetworkProfiler to overcome this limitation.

During the past decade, cancer therapy has become increasingly personalized [2,3]. Unlike the traditional "one-size-fits-all" approach to cancer therapy, patient-specific cancer therapy reduces the side effects of chemotherapy and predicts the odds of cancer recurrence more accurately by tailoring cancer treatment to specific genetic defects in the tumors of individual patients. However, this goal is not an easy task since cancer is an extremely complex and heterogeneous disease. We believe that NetworkProfiler will help elucidate the systems biology of cancer and facilitate personalized chemotherapy.

Materials and Methods

Cell lines and reagents

Human non-small cell lung cancer (NSCLC) cell lines, A549, NCI-H1437 and NCI-H727, were purchased from American Tissue Culture Collection, while other NSCLC cell lines, Calu1, Calu6 and SK-MES1, were generously provided by Dr. L. J. Old (Memorial Sloan-Kettering Cancer Center). Cells were maintained in RPMI 1640 supplemented with 10% fetal bovine serum. The anti-E-cadherin antibody was purchased from BD Transduction Laboratories, anti-vimentin from Santa Cruz Biotechnology, anti-α-tublin from Sigma Aldrich, and anti-mouse IgG from Cell Signaling Technology. The Alexa-conjugated anti-mouse IgG was purchased from Molecular Probes. siRNAs against KLF5 (siKLF5 #1 and #2) and a negative control (siNC) were purchased from Sigma Genosys. Pre-miR has-miR-100 and negative control #2 were purchased from Ambion. Human TGF-β was purchased from R&D Systems, Inc.

Immunostaining, western blot analysis and in vitro motility assay

2×10^4 cells in 6-well plates were transiently transfected with either 20 nM siRNA or 10 nM Pre-miR molecules using Lipofectamine RNAiMAX (Invitrogen), as previously described [65]. Immunofluorescence staining was carried out after fixation with 3.7% formaldehyde and postfixing with 0.1% Triton X-100 each for 10 min at RT. Photographs were taken 72 hr after transfection. Cells were harvested 48 hr after transfection for western blot analysis. In vitro motility assay based on Transwell-chamber culture systems was performed, as previously described [66].

Quantitative real-time reverse transcription (RT)-PCR analysis

Quantitative real-time RT-PCR analysis of KLF5 was performed using Power SYBR Green (Applied Biosystems) and the following PCR primers:

5′-CCCTTGCACATACACAATGC-3′ and 5′-GGATGGAGGTGGGGTTAAAT-3′. Quantitative real-time RT-PCR analysis of miR-100 and RNU44 was performed using TaqMan probes and 7500 Fast Real-Time PCR system (Applied Biosystems), essentially as previously described [67].

NetworkProfiler

NetworkProfiler employed a varying-coefficient structural equation model (SEM) to represent the modulator-dependent conditional independence between gene transcripts. Let there be q possible regulators, $R_1,\ldots,R_q$, that may control the transcription of the k-th target gene T_k when the modulator $M=m$. Then the varying-coefficient structural equation model for T_k is

$$T_k=\sum_{j=0}^{q}\beta_{jk}(m)\cdot R_j+\varepsilon_k,$$

where $\beta_{jk}(m)$ is the coefficient function that represents the effect of R_j on T_k, $R_0=1$, and ε_k is a noise term. If $T_k=R_l$, then the term $\beta_{lk}(m)\cdot R_l$ can be omitted from the model, i.e., $\beta_{lk}(m)=0$ for all m. By estimating the parameters $\beta_{jk}(m)$, we obtain the transcriptional regulatory gene network at $M=m$.

We used a kernel-based method to estimate these parameters. Let there be n sets of gene expression profiles. Then, the SEM for the α-th sample can be rewritten as

$$t_{\alpha k}=\sum_{j=0}^{q}\beta_{jk\alpha}\cdot r_{\alpha j}+\varepsilon_{\alpha k},\alpha=1,\ldots,n,$$

where $t_{\alpha k}$, $r_{\alpha j}$, and m_α are the values of the k-th target gene, the j-th regulator, and the modulator for the α-th sample, respectively; $r_{0k}=1$, and $\beta_{jk\alpha}=\beta_{jk}(m_\alpha)$. For n samples, we obtain n modulator-dependent gene regulatory networks, i.e., the regulatory effects of R_j $(j=1,\ldots,q)$ on T_k $(k=1,\ldots,p)$ are determined by $\hat{\beta}_{111},\ldots,\hat{\beta}_{qpn}$, where $\hat{\beta}_{jk\alpha}$ is the estimate of $\beta_{jk\alpha}$.

We assumed that the values of the coefficients are almost constant for the neighborhood samples of the α-th sample with respect to the modulator m, that is, $\beta_{jki}\approx c$ for the i-th sample that satisfies $|m_i - m_\alpha| < \delta$ for some constant c and small δ. Then, we estimated the parameters $\beta_{jk\alpha}$ for fixed α by minimizing a regularized kernel-based weighted residual sum of squares

$$L_k(\beta_{1k\alpha},\ldots,\beta_{qk\alpha}|h_k) = \frac{1}{2}\sum_{i=1}^{n}\{t_{ik} - \sum_{j=1}^{q}\beta_{jk\alpha}\cdot r_{ij}\}^2 K(m_i - m_\alpha|h_k)$$

$$+\lambda_{k\alpha}\sum_{j=1}^{q} w_{jk\alpha}\cdot|\beta_{jk\alpha}| + \frac{\gamma_{k\alpha}}{2}\sum_{j=1}^{q}\beta_{jk\alpha}^2, \quad (1)$$

where $K(m_i - m_\alpha|h_k)$ is a Gaussian kernel function defined by

$$K(m_i - m_\alpha|h_k) = \exp\left\{-\frac{1}{h_k}(m_i - m_\alpha)^2\right\},$$

and $\lambda_{k\alpha}$ and $\gamma_{k\alpha}$ are hyperparameters that control the L_1 (lasso [68]) and L_2 (ridge [69]) penalties, respectively. In addition, $w_{jk\alpha}$ is an importance weight for $\beta_{jk\alpha}$, and h_k is the bandwidth of the Gaussian kernel. The kernel function $K(m_i - m_\alpha|h_k)$ defines the neighborhood around the α-th sample in terms of M; a large value of $K(m_i - m_\alpha|h_k)$ means that the i-th sample is in the neighborhood of the α-th sample. By fixing $\lambda_{k\alpha}$, $\gamma_{k\alpha}$, $w_{jk\alpha}$, and h_k, we obtain the estimates

$$\{\hat{\beta}_{1k\alpha},\ldots,\hat{\beta}_{qk\alpha}\} = \arg\min_{\beta_{jk\alpha}} L_k(\beta_{1k\alpha},\ldots,\beta_{qk\alpha}).$$

This parameter estimation method is a weighted version of the elastic net [22]. The L_1 penalty zeroes some coefficients [68], which produces a sparse network structure. In contrast, the L_2 penalty stabilizes the solution by a grouping effect that promotes the collective inclusion or exclusion of highly correlated variables in the model [22]. The importance weights $w_{jk\alpha}$ allow tuning parameters to take on different values for different coefficients $\beta_{jk\alpha}$. For example, if $w_{jk\alpha}$ has a large value, then an estimator $\hat{\beta}_{jk\alpha}$ tends to be zero. In contrast, if $w_{jk\alpha}$ has a small value that is nearly equal to zero, $\hat{\beta}_{jk\alpha}$ tends to be non-zero. These weights create a sparser network structure than the lasso and elastic net methods. The parameters $\beta_{jk\alpha}$ were estimated by using a recursive procedure, and the weights $w_{jk\alpha}$ were updated by $w_{jk\alpha} = 1/(\tilde{\beta}_{jk\alpha} + \xi)$ [70], where $\tilde{\beta}_{jk\alpha}$ is the estimate from the previous step and $\xi = 10^{-5}$ to avoid dividing by zero. Then, the modulator-dependent networks for n samples can be derived from the estimates of $\hat{\beta}_{jk\alpha}$ ($j=1,\ldots,q$, $k=1,\ldots,p$, and $\alpha=1,\ldots,n$).

For convenience of subsequent explanations, we introduce the following notations:

$$\boldsymbol{t}_{k\alpha}(h_k) = \begin{pmatrix} \kappa_{1\alpha}(h_k)\cdot t_{1k} \\ \vdots \\ \kappa_{n\alpha}(h_k)\cdot t_{nk} \end{pmatrix}, and$$

$$\boldsymbol{R}_\alpha(h_k) = \begin{pmatrix} \kappa_{1\alpha}(h_k)\cdot r_{11} & \cdots & \kappa_{1\alpha}(h_k)\cdot r_{1q} \\ \vdots & \ddots & \vdots \\ \kappa_{n\alpha}(h_k)\cdot r_{n1} & \cdots & \kappa_{n\alpha}(h_k)\cdot r_{nq} \end{pmatrix},$$

where $\kappa_{i\alpha}(h_k) = \sqrt{K(m_i - m_\alpha|h_k)}$.

In these expressions, $\boldsymbol{t}_{k\alpha}(h_k)$ and $\boldsymbol{R}_\alpha(h_k)$ were normalized so that the means and variances for $\boldsymbol{t}_{k\alpha}(h_k)$ and each column of $\boldsymbol{R}_\alpha(h_k)$ were 0 and 1, respectively. As a result, the intercept $\beta_{0k\alpha}$ was not included in the loss function (1). For fixed h_k, the loss function (1) can be minimized by using a kernel-based weighted version of the recursive elastic net [70]. The tuning parameters $\lambda_{k\alpha}$ and $\gamma_{k\alpha}$ were selected by minimizing a modified version of the bias-corrected weighted Akaike information criterion (AIC) [71]:

$$\mathrm{mWAICc}_{k\alpha} = (n_\alpha(h_k)+1)\cdot\log(2\pi\hat{\sigma}_{k\alpha}^2) + \frac{2n_\alpha(h_k)(\hat{\mathrm{df}}_{k\alpha}+1)}{n_\alpha(h_k) - \hat{\mathrm{df}}_{k\alpha} - 2},$$

where $n_\alpha(h_k) = \sum_{i=1}^{n}\kappa_{i\alpha}(h_k)$, and $\hat{\sigma}_{k\alpha}^2$ is estimated by

$$\hat{\sigma}_{k\alpha}^2 = \frac{1}{n_\alpha(h_k)}\|\boldsymbol{t}_{k\alpha}(h_k) - \boldsymbol{R}_\alpha(h_k)\hat{\beta}_{k\alpha}\|_2^2,$$

with $\hat{\beta}_{k\alpha} = (\hat{\beta}_{1k\alpha},\ldots,\hat{\beta}_{qk\alpha})'$. In addition, $\hat{\mathrm{df}}_{k\alpha}$ is the unbiased estimate of the degrees of freedom given by

$$\hat{\mathrm{df}}_{k\alpha} = tr\left[(\tilde{\boldsymbol{R}}(h_k)'\tilde{\boldsymbol{R}}(h_k) + \gamma_{k\alpha}\boldsymbol{I})^{-1}\tilde{\boldsymbol{R}}(h_k)'\tilde{\boldsymbol{R}}(h_k)\right],$$

where $\boldsymbol{I}$ is the identify matrix and $\tilde{\boldsymbol{R}}(h_k)$ is the submatrix of $\boldsymbol{R}(h_k)$, which has columns that correspond to the nonzero coefficients, respectively.

The NetworkProfiler algorithm is shown below:

Algorithm: NetworkProfiler.

$\tilde{w}_{jk\alpha} \leftarrow 1$ ($j=1,\ldots,q$)
iter $\leftarrow$ 1
for $\gamma_{k\alpha} = \gamma[r]$ ($r=1,\ldots,G$) **do**
repeat
Calculate $\hat{\beta}_{k\alpha}[l,r]$ and $\mathrm{mWAICc}_{k\alpha}[l,r]$ corresponding to $\lambda_{k\alpha} = \lambda_k[l]$ ($l=1,\ldots,L$).
$z_r[\text{iter}] \leftarrow \min\{\mathrm{mWAICc}_{k\alpha}(l,r); l=1,\ldots,L\}$
$l^* \leftarrow \arg\min_l\{\mathrm{mWAICc}_{k\alpha}(l,r); l=1,\ldots,L\}$
if $z_r[\text{iter}] - z_r[\text{iter}-1] > 0$ **then**
Exit loop
else
$z^*[r] \leftarrow z_r[\text{iter}]$
$\hat{\beta}_{k\alpha}[r] \leftarrow \hat{\beta}_{k\alpha}[l^*,r]$
$\tilde{w}_{jk\alpha} \leftarrow 1/(|\hat{\beta}_{jk\alpha}(r)| + \xi)$ ($j=1,\ldots,q$)
iter $\leftarrow$ iter $+1$
end if
untill iter reaches to M.
end for
$r^* \leftarrow \arg\min_r\{z^*[r]; r=1,\ldots,G\}$
Return the coefficient vector $\hat{\beta}_{k\alpha} = \hat{\beta}_{k\alpha}[r^*]$.

The results from NetworkProfiler, which are the estimates of q coefficients $\hat{\beta}_{jk\alpha}$ ($j=1,\ldots,q$) for the k-th target gene of the α-th patient, depend on the values of h_k. We used cross-validation to select an optimal value of h_k and estimate $q \times n$ coefficients, $\beta_{1k1},\ldots,\beta_{qkn}$ by minimizing the cross-validation error:

$$\mathrm{CV}_k = \sum_{\alpha\in\mathbb{S}}(t_{\alpha k} - \sum_{j=0}^{q}\hat{\beta}_{jk\alpha}^{(-\alpha)}\cdot r_{\alpha j})^2, \quad (2)$$

where $\mathbb{S}$ is a randomly selected set of samples and $\hat{\beta}_{1k\alpha}^{(-\alpha)},\ldots,\hat{\beta}_{qk\alpha}^{(-\alpha)}$ are estimated from the remaining samples by minimizing:

$$L_k^{-\alpha}(\beta_{1k\alpha},\dots,\beta_{qk\alpha}|h_k)=\frac{1}{2}\sum_{i\notin\mathbb{S}}\{t_{ik}-\sum_{j=0}^{q}\beta_{jk\alpha}\cdot r_{ij}\}^2K(m_i-m_\alpha|h_k)$$

$$+\lambda_{k\alpha}\sum_{j=1}^{q}w_{jk\alpha}\cdot|\beta_{jk\alpha}|+\frac{\gamma_{k\alpha}}{2}\sum_{j=1}^{q}\beta_{jk\alpha}^2. \quad (3)$$

The algorithm in NetworkProfiler for minimizing this loss function (3) is shown below:

Algorithm: Conditional optimization with cross-validation.

1: **for** $h_k=h_l$ $(l=1,\dots,H)$ **do**
2: **for all** α such that $\alpha\in\mathbb{S}$ **do**
3: Calculate $\hat\beta_{1k\alpha}^{(-\alpha)}[h_l],\dots,\hat\beta_{qk\alpha}^{(-\alpha)}[h_l]$ with NetworkProfiler.
4: **end for**
5: Calculate $\mathrm{CV}_k[h_l]$.
6: **end for**
7: $h_k^*\leftarrow\mathrm{argmin}_{h_l}\{\mathrm{CV}_k[h_l];l=1,\dots,H\}$
8: **for** $\alpha=1,\dots,n$ **do**
9: Calculate $\hat\beta_{1k\alpha}[h_k^*],\dots,\hat\beta_{qk\alpha}[h_k^*]$ with NetworkProfiler.
10: **end for**
11: Return a sequence of the coefficient vectors $\hat\beta_{k1}(h_k^*),\dots,\hat\beta_{kn}(h_k^*)$.

Subsequently, the modulator-dependent gene networks for n samples are determined from the coefficient vectors $\hat\beta_{k1}(\hat h_k),\dots,\hat\beta_{kn}(\hat h_k)$ $(k=1,\dots,p)$ by applying the above algorithm for all $k=1,\dots,p$. The computational cost of this algorithm rapidly increases as the number of samples and genes increase. Thus, for computers that only have a single central processing unit (CPU), this algorithm is only practical for medium-sized networks with up to several genes. However, since this algorithm can be executed in parallel for every k, it can be run on a stand-alone workstation with multi-core CPUs and computer clusters. Figure S4 represents the histogram of computational times based on 12 core CPUs (Intel Xeon Processor E5450 (# of cores = 4, clock speed = 3.0GHz) × 3) for calculating 762 cancer cell line-specific gene networks from 13,508 × 762 gene expression data through 100,000 iterations when 100 target genes were randomly selected among 13,508 genes and the number of regulators was not restricted, i.e., 1732 regulators were used. The average computational time was about 9 days. In this situation, we can find putative master regulators of the focused target genes related with a modulator of interest. Of course, for calculating gene networks of 762 samples for a large number of target genes, a supercomputer is required. In this study, we used the Super Computer System at the Human Genome Center, Institute of Medical Science, University of Tokyo, Japan, to analyze 762 gene networks with 13,508 target genes.

Signature-based hidden modulator extraction

When the modulator was a variable that is difficult to observe, we used a signature-based hidden modulator extraction algorithm to estimate the value of the modulator for each sample. This algorithm takes seed genes that are related to the modulator and computes the underlying latent variable of the modulator by using principal components and extraction of expression modules (EEM) [7]. Let $\mathbb{M}$ be a gene set that is related to the modulator and let $X_{\mathbb{M}}$ be an $n\times|\mathbb{M}|$ matrix of n expression levels of $\mathbb{M}$. Then, a linear model, which is a special case of the single factor model [72], relates $\mathbb{M}^*$, a subset of $\mathbb{M}$, to an underlying latent variable U as follows:

$$X_j=\alpha_{0j}+\alpha_{1j}U+\varepsilon'_j,\ j\in\mathbb{M}^*\subseteq\mathbb{M}, \quad (4)$$

where X_j is the expression level of the j-th gene in $\mathbb{M}^*$, α_{0j} is the y-intercept, α_{1j} is a coefficient, and ε'_j is a noise term. We assumed that other genes that do not include $\mathbb{M}^*$ $(\{X_j;j\notin\mathbb{M}^*\})$ are independent of U.

The values of U for n samples, u_i $(i=1,\dots,n)$, can be estimated by the following procedure:

Algorithm: signature-based hidden modulator extraction.

1: For a given set $\mathbb{M}$, find a subset $\mathbb{M}^*$ based on the expression coherence with the EEM algorithm [7].
2: Given $\mathbb{M}^*$, singular value decomposition of the data matrix $X_{\mathbb{M}^*}$ estimates u_i by the largest principal component.
3: Return the values u_i $(i=1,\dots,n)$.

In the first step, we estimate $\mathbb{M}^*$. In the second step, we assume that the noise terms ε'_j have Gaussian distributions with equal variances. As a result, the singular value decomposition generates maximum likelihood estimates of u_i for the single factor model [72].

Regulatory effect

To identify upstream regulators that had strong effects on the expression of a target gene of interest in the constructed modulator-dependent gene networks, we defined a measure, called the regulatory effect, of the effect of the j-th regulator on the k-th target gene in the α-th sample as

$$\mathrm{RE}_{jk\alpha}=\sum_{l\in\pi_{jk\alpha}}\hat\beta_l^{(j\to k)}(m_\alpha)\cdot r_{\alpha j}, \quad (5)$$

where $\pi_{jk\alpha}$ is the set of all possible paths from R_j to T_k, and $\hat\beta_l^{(j\to k)}(m_\alpha)$ is the product of the estimated coefficients on the l-th path that includes $\pi_{jk\alpha}$. For example, given all the possible paths from R_1 to T_2 in the α-th sample (Figure S5), the set $\pi_{12\alpha}$ is

$$\pi_{12\alpha}=\{R_1\to T_2,R_1\to R_3\to T_2,R_1\to R_3\to R_4\to T_2\}, \quad (6)$$

and the regulatory effect $\mathrm{RE}_{12\alpha}$ is

$$\mathrm{RE}_{12\alpha}=(\hat\beta_{12\alpha}+\hat\beta_{13\alpha}\cdot\hat\beta_{32\alpha}+\hat\beta_{13\alpha}\cdot\hat\beta_{34\alpha}\cdot\hat\beta_{42\alpha})\cdot r_{\alpha j}. \quad (7)$$

In our analysis, the length of the paths from R_j to T_k is restricted to either 1 or 2.

To determine how the modulator affects the regulatory effect $\mathrm{RE}_{jk\alpha}$, we also defined the change in the regulatory effect of the j-th regulator on the k-th target as

$$\mathrm{REC}_{jk}=\max\{\mathrm{RE}_{jk\alpha};\alpha=1,\dots,n\}-\min\{\mathrm{RE}_{jk\alpha};\alpha=1,\dots,n\}. \quad (8)$$

In addition to this definition, it is also possible to use percentiles instead of max and min to achieve more robust results. However, in our analysis, we used max and min to increase the power of the method. It should be noted that the change in the regulatory effect REC_{jk} does not explain the mode of action for the modulator with respect to the regulator-target relationship. File S5 (http://bonsai.hgc.jp/~shima/NetworkProfiler) is provided to determine the modulator mode of action by statistical test.

Gene set analysis of downstream genes for a regulator

To identify regulators that enhanced the functions of their targets, we calculated the statistical significance of the enrichment of targets for a given regulator in each sample. To test the enrichment, we use the degree of independence between the two properties:

$\mathcal{A}_{j\alpha}$:gene is in the list of targets for the j-th regulator in the α – th sample

$\mathcal{B}_u$:gene is a member of the u-th priori set

Testing the association between the properties $\mathcal{A}_{j\alpha}$ and $\mathcal{B}_u$ corresponds to Fisher's exact test. The p-value calculated by this test, $P_{ju\alpha}$, indicates the probability of observing at least the same amount of enrichment when downstream genes are randomly selected out of all genes. Thus, a very small p-value gives strong evidence for an association between $\mathcal{A}_{j\alpha}$ and $\mathcal{B}_u$ for the j-th regulator in the α-th sample. To correct for multiple hypotheses testing, Benjamini-Hochberg (BH)-corrected p-values (q-values) [73], $Q_{ju\alpha}$, were calculated.

To determine how the modulator affects the functions of downstream genes for a regulator, we defined the enrichment score, ES_{ju}, as a change in the statistical significance of the enrichment of targets for the j-th regulator on the u-th function:

$$ES_{ju} = \log(\max\{Q_{ju\alpha}; \alpha=1,\ldots,n\} / \min\{Q_{ju\alpha}; \alpha=1,\ldots,n\}). \quad (9)$$

Thus, a very large ES_{ju} indicates that the modulator causes a significant change of the enrichment of the targets for the j-th regulator on the u-th function.

To identify putative master regulators that control more functional gene sets than other regulators, we also calculated the total enrichment score, TES_j, by combining independent enrichment scores, $ES_{j1},\ldots,ES_{jU}$, where U is the number of functional gene sets:

$$TES_j = 2\sum_{u=1}^{U} ES_{ju}. \quad (10)$$

The total enrichment score is equivalent to the difference of the Fisher's statistic $-2\sum_{i=1}^{k} \log P_k$ [74] which was used to combine independent tests obtained from k studies based on the p-values, $P_1,\ldots,P_k$. The Fisher's method is based on the fact that the statistic $-2\sum_{i=1}^{k} \log P_i$ follows a chi-square distribution with $2k$ degrees of freedom under the global null hypothesis that all null hypotheses are true. A small integral p-value for the hypothesis indicates that the j-th regulator controlled at least one or more functional gene sets during the change of the modulator.

Supporting Information

Figure S1 Quantitative real-time RT-PCR analysis of KLF5 in siKLF5-treated A549 cells.
(PDF)

Figure S2 Expression profiles of miR-100 in order of ascending the EMT-related modulator values.
(PDF)

Figure S3 miR-100-induced changes in biologic characteristics in NCI-H1437 and NCI-H727 NSCLC cell lines. (a) Representative phase contrast microscopic images showing negligible changes in morphology by miR-100 introduction in both NSCLC cells lines. NC#2, negative control. (b) Motility assay showing increased migration by introduction of miR-100 in both NSCLC cell lines. *, $P<0.05$.
(PDF)

Figure S4 Histogram of computational times for inferring cancer cell line-specific gene networks running on 12 core CPUs. The 762 cancer cell line-specific gene networks related with the EMT were calculated from 13,508 × 762 gene expression data when 100 target genes were randomly selected among 13,508 genes and the number of regulators was not restricted, i.e., 1,732 regulators were used. The comptational times were based on 12 core CPUs (Intel Xeon Processor E5450 (# of cores = 4, clock speed = 3.0 GHz) × 3). The histogram was calculated by 100,000 iterations.
(PDF)

Figure S5 Example of paths among four genes, R_1, T_2, R_3 , and R_4.
(PDF)

Table S1 List of candidate regulators mapped to 1183 transcription factors and 47 nuclear receptors.
(XLS)

Table S2 List of candidate regulators mapped to 502 human microRNAs.
(XLS)

Table S3 List of coherent genes (p-value $<10^{-5}$) related to EMT calculated by extraction of expression module (EEM).
(XLS)

Table S4 EMT-related modulator values of 762 cancer cell lines calculated by signature-based hidden modulator extraction.
(XLS)

Table S5 List of 370 putative master regulators of E-cadherin during the EMT which were estimated by NetworkProfiler.
(XLS)

Table S6 List of 627 putative master regulators of E-cadherin which were estimated by a structual equation model (SEM) with the elastic net.
(XLS)

Table S7 Regulator function matrix between 1732 regulators and 5 functions. The row and column indicate regulator and functional gene set, respectively. The (i,j)-th element represents the change during the EMT in the statistical significance (-$\log_{10}$(q-value)) for the enrichment of target genes of the i-th regulator on the j-th function. The last column indicate the integral q-value of each row regulator which were used to determine which regulator strongly affected the functional gene sets.
(XLS)

Table S8 List of 17 putative master regulators (integral q-value $<10^{-10}$) which correlated at least one or more EMT-related functions and were known to be downstream targets of TGFB1 with published evidence from Ingenuity Knowledge Base (http://www.ingenuity.com).
(XLS)

Table S9 List of the changes in the regulatory effects from 1732 regulators to E-cadherin and vimentin during the EMT.
(XLS)

Acknowledgments

The supercomputing resource was provided by Human Genome Center (University of Tokyo).

Author Contributions

Conceived and designed the experiments: TT. Performed the experiments: YS YH. Analyzed the data: TS AN. Wrote the paper: TS. Organized the project: SM. Provided statistical expertise: SI RY. Provided computational expertize: AN MN. Provided experimental expertise: TT. Provided manuscript review: SI AN RY TT.

References

1. Wang E (2010) Cancer systems biology CRC Press.
2. Schisky RL (2010) Personalized medicine in oncology: the future is now. Nat Rev Drug Discov 9(5): 363–6.
3. Gonzalez-Angulo AM, Hennessy BT, Mills GB (2010) Future of personalized medicine in oncology: a systems biology approach. J Clin Oncol 28(16): 2777–83.
4. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D (2007) How to infer gene networks from expression profiles. Mol Syst Biol 3: 78.
5. Thiery JP, Acloque H, Huang RY, Nieto MA (2009) Epithelial-mesenchymal transitions in development and disease. Cell 139(5): 871–90.
6. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102(43): 15545–50.
7. Niida A, Smith AD, Imoto S, Aburatani H, Zhang MQ, et al. (2009) Gene set-based module discovery in the breast cancer transcriptome. BMC Bioinformatics 10: 71.
8. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, et al. (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol 10(5): 593–601.
9. Comijn J, Berx G, Vermassen P, Verschueren K, van Grunsven L, et al. (2001) The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. Mol Cell 7(6): 1267–78.
10. Yori JL, Johnson E, Zhou G, Jain MK, Keri RA (2010) Kruppel-like factor 4 inhibits epithelial-tomesenchymal transition through regulation of E-cadherin gene expression. J Biol Chem 285(22): 16854–63.
11. Song Y, Washington MK, Crawford HC (2010) Loss of FOXA1/2 is essential for the epithelialto-mesenchymal transition in pancreatic cancer. Cancer Res 70(5): 2115–25.
12. Sobrado VR, Moreno-Bueno G, Cubillo E, Holt LJ, Nieto MA, et al. (2009) The class I bHLH factors E2-2A and E2-2B regulate EMT. J Cell Sci 122(Pt 7): 1014–24.
13. Kato M, Zhang J, Wang M, Lanting L, Yuan H, et al. (2007) MicroRNA-192 in diabetic kidney glomeruli and its function in TGF-beta-induced collagen expression via inhibition of E-box repressors. Proc Natl Acad Sci U S A 104(9): 3432–7.
14. Wang B, Herman-Edelstein M, Koh P, Burns W, Jandeleit-Dahm K, et al. (2010) E-cadherin expression is regulated by miR-192/215 by a mechanism that is independent of the profibrotic effects of transforming growth factor-beta. Diabetes 59(7): 1794–802.
15. Eger A, Aigner K, Sonderegger S, Dampier B, Oehler S, et al. (2005) DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. Oncogene 24(14): 2375–85.
16. Hajra KM, Chen DY, Fearon ER (2002) The SLUG zinc-finger protein represses E-cadherin in breast cancer. Cancer Res 62(6): 1613–8.
17. Yang J, Mani SA, Donaher JL, Ramaswamy S, Itzykson RA, et al. (2004) Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell 117(7): 927–39.
18. Perez-Moreno MA, Locascio A, Rodrigo I, Dhondt G, Portillo F, et al. (2001) A new role for E12/E47 in the repression of E-cadherin expression and epithelial-mesenchymal transitions. J Biol Chem 276(29): 27424–31.
19. Almeida MS, Bray SJ (2005) Regulation of post-embryonic neuroblasts by Drosophila Grainyhead. Mech Dev 122(12): 1282–93.
20. Cowger JJ, Zhao Q, Isovic M, Torchia J (2007) Biochemical characterization of the zinc-finger protein 217 transcriptional repressor complex: identification of a ZNF217 consensus recognition sequence. Oncogene 26(23): 3378–86.
21. Yang Y, Goldstein BG, Chao HH, Katz JP (2005) KLF4 and KLF5 regulate proliferation, apoptosis and invasion in esophageal cancer cells. Cancer Biol Ther 4(11): 1216–21.
22. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J Roy Statist Soc Ser B 67: 301–20.
23. Bracken CP, Gregory PA, Kolesnikoff N, Bert AG, Wang J, et al. (2008) A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. Cancer Res 68(19): 7846–54.
24. Willis BC, Borok Z (2007) TGF-beta-induced EMT: mechanisms and implications for fibrotic lung disease. Am J Physiol Lung Cell Mol Physiol 293(3): L525–34.
25. Chen H, Zhu G, Li Y, Padia RN, Dong Z, et al. (2009) Extracellular signal-regulated kinase signaling pathway regulates breast cancer cell migration by maintaining slug expression. Cancer Res 69(24): 9228–35.
26. Adiseshaiah P, Lindner DJ, Kalvakolanu DV, Reddy SP (2007) FRA-1 proto-oncogene induces lung epithelial cell invasion and anchorage-independent growth in vitro, but is insufficient to promote tumor growth in vivo. Cancer Res 67(13): 6204–11.
27. Cowden Dahl KD, Robertson SE, Weaver VM, Simon MC (2005) Hypoxia-inducible Factor Regulates alphavbeta3 Integrin Cell Surface Expression. Mol Biol Cell 16(4): 1901–12.
28. Imtiyaz HZ, Williams EP, Hickey MM, Patel SA, Durham AC, et al. (2010) Hypoxia-inducible factor 2alpha regulates macrophage function in mouse models of acute and tumor inflammation. J Clin Invest 120(8): 2699–714.
29. Yang Y, Tetreault MP, Yermolina YA, Goldstein BG, Katz JP (2008) Kruppel-like Factor 5Controls Keratinocyte Migration via the Integrin-linked Kinase. J Biol Chem 283(27): 18812–20.
30. Marlowe JL, Puga A (2005) Aryl hydrocarbon receptor, cell cycle regulation, toxicity, and tumorigenesis. J Cell Biochem 96(6): 1174–84.
31. Nakagawa H, Liyanarachchi S, Davuluri RV, Auer H, Martin EW, et al. (2004) Role of cancerassociated stromal fibroblasts in metastatic colon cancer to the liver and their expression profiles. Oncogene 23(44): 7366–77.
32. Malin D, Kim IM, Boetticher E, Kalin TV, Ramakrishna S, et al. (2007) Forkhead box f1 is essential for migration of mesenchymal cells and directly induces integrin-beta3 expression. Mol Cell Biol 27(7): 2486–98.
33. Buchwalter G, Gross C, Wasylyk B (2005) The Ternary Complex Factor Net Regulates Cell Migration through Inhibition of PAI-1 Expression. Mol Cell Biol 25(24): 10853–62.
34. Hayes SA, Huang X, Kambhampati S, Platanias LC, Bergan RC (2003) p38 MAP kinase modulatesSmad-dependent changes in human prostate cell adhesion. Oncogene 22(31): 4841–50.
35. Matsuzaki K, Kitano C, Murata M, Sekimoto G, Yoshida K, et al. (2009) Smad2 and Smad3 phosphorylated at both linker and COOH-terminal regions transmit malignant TGF-beta signal in later stages of human colorectal cancer. Cancer Res 69(13): 5321–30.
36. Sekimoto G, Matsuzaki K, Yoshida K, Mori S, Murata M, et al. (2007) Reversible Smad-Dependent Signaling between Tumor Suppression and Oncogenesis. Cancer Res 67(11): 5090–6.
37. Sato M, Muragaki Y, Saika S, Roberts AB, Ooshima A (2003) Targeted disruption of TGFbeta1/Smad3 signaling protects against renal tubulointerstitial fibrosis induced by unilateral ureteral obstruction. J Clin Invest 112(10): 1486–94.
38. Chan SW, Lim CJ, Guo K, Ng CP, Lee I, et al. (2008) A role for TAZ in migration, invasion, and tumorigenesis of breast cancer cells. Cancer Res 68(8): 2592–8.
39. Sachdeva M, Mo YY (2010) MicroRNA-145 suppresses cell invasion and metastasis by directly targeting mucin 1. Cancer Res 70(1): 378–87.
40. Matsuya M, Sasaki H, Aoto H, Mitaka T, Nagura K, et al. (1998) Cell adhesion kinase beta forms a complex with a new member, Hic-5, of proteins localized at focal adhesions. J Biol Chem 273(2): 1003–14.
41. Lu X, Yan CH, Yuan M, Wei Y, Hu G, et al. (2010) In vivo dynamics and distinct functions of hypoxia in primary tumor growth and organotropic metastasis of breast cancer. Cancer Res 70(10): 3905–14.
42. Okuyama H, Krishnamachary B, Zhou YF, Nagasawa H, Bosch-Marce M, et al. (2006) Expressionof vascular endothelial growth factor receptor 1 in bone marrow-derived mesenchymal cells is dependent on hypoxia-inducible factor 1. J Biol Chem 281(22): 15554–63.
43. Kim KS, Sengupta S, Berk M, Kwak YG, Escobar PF, et al. (2006) Hypoxia Enhances Lysophosphatidic Acid Responsiveness in Ovarian Cancer Cells and Lysophosphatidic Acid Induces Ovarian Tumor Metastasis In vivo. Cancer Res 66(16): 7983–90.
44. Schedin PJ, Eckel-Mahan KL, McDaniel SM, Prescott JD, Brodsky KS, et al. (2004) ESX induces transformation and functional epithelial to mesenchymal transition in MCF-12A mammary epithelial cells. Oncogene 23(9): 1766–79.
45. Panda DK, Miao D, Lefebvre V, Hendy GN, Goltzman D (2001) The transcription factor SOX9 regulates cell cycle and differentiation genes in chondrocytic CFK2 cells. J Biol Chem 276(44): 41229–36.
46. Mori-Akiyama Y, Akiyama H, Rowitch DH, de Crombrugghe B (2003) Sox9 is required for determination of the chondrogenic cell lineage in the cranial neural crest. Proc Natl Acad Sci U S A 100(16): 9360–5.
47. Tomioka N, Osumi N, Sato Y, Inoue T, Nakamura S, et al. (2000) Neocortical origin and tangential migration of guidepost neurons in the lateral olfactory tract. J Neurosci 20(15): 5802–12.
48. Jean C, Blanc A, Prade-Houdellier N, Ysebaert L, Hernandez-Pigeon H, et al. (2009) Epidermal growth factor receptor/beta-catenin/T-cell factor 4/matrix metalloproteinase 1: a new pathway for regulating keratinocyte invasiveness after UVA irradiation. Cancer Res 69(8): 3291–9.
49. Minami T, Miura M, AirdWC, Kodama T (2006) Thrombin-induced Autoinhibitory Factor, Down Syndrome Critical Region-1, Attenuates NFAT-dependent Vascular Cell Adhesion Molecule-1 Expression and Inflammation in the Endothelium. J Biol Chem 281(29): 20503–20.

50. Huang S, Pettaway CA, Uehara H, Bucana CD, Fidler IJ (2001) Blockade of NF-kappaB activity in human prostate cancer cells is associated with suppression of angiogenesis, invasion, and metastasis. Oncogene 20(31): 4188–97.
51. Shair KH, Schnegg CI, Raab-Traub N (2008) EBV Latent Membrane Protein 1 Effects on Plakoglobin, Cell Growth, and Migration. Cancer Res 68(17): 6997–7005.
52. Gakidis MA, Cullere X, Olson T, Wilsbacher JL, Zhang B, et al. (2004) Vav GEFs are required for beta2 integrin-dependent functions of neutrophils. J Cell Biol 166(2): 273–82.
53. Schymeinsky J, Sindrilaru A, Frommhold D, Sperandio M, Gerstl R, et al. (2006) The Vav binding site of the non-receptor tyrosine kinase Syk at Tyr 348 is critical for beta2 integrin (CD11/CD18)- mediated neutrophil migration. Blood 108(12): 3919–27.
54. Hong IK, Jin YJ, Byun HJ, Jeoung DI, Kim YM, et al. (2006) Homophilic Interactions of Tetraspanin CD151 Up-regulate Motility and Matrix Metalloproteinase-9 Expression of Human Melanoma Cells through Adhesion-dependent c-Jun Activation Signaling Pathways. J Biol Chem 281(34): 24279–92.
55. Janulis M, Silberman S, Ambegaokar A, Gutkind JS, Schultz RM (1999) Role of mitogen-activated protein kinases and c-Jun/AP-1 trans-activating activity in the regulation of protease mRNAs and the malignant phenotype in NIH 3T3 fibroblasts. J Biol Chem 274(2): 801–13.
56. Cai C, Hsieh CL, Omwancha J, Zheng Z, Chen SY, et al. (2007) ETV1 Is a Novel Androgen Receptor-Regulated Gene that Mediates Prostate Cancer Cell Invasion. Mol Endocrinol 21(8): 1835–46.
57. Bauer K, Kratzer M, Otte M, de Quintana KL, Hagmann J, et al. (2000) Human CLP36, a PDZdomain and LIM-domain protein, binds to alpha-actinin-1 and associates with actin filaments and stress fibers in activated platelets and endothelial cells. Blood 96(13): 4236–45.
58. Zent CS, Mathieu C, Claxton DF, Zhang DE, Tenen DG, et al. (1996) The chimeric genes AML1/MDS1 and AML1/EAP inhibit AML1B activation at the CSF1R promoter, but only AML1/MDS1 has tumor-promoter properties. Proc Natl Acad Sci U S A 93(3): 1044–8.
59. Perry C, Sklan EH, Birikh K, Shapira M, Trejo L, et al. (2002) Complex regulation of acetylcholinesterase gene expression in human brain tumors. Oncogene 21(55): 8428–41.
60. Zhang X, Milton CC, Humbert PO, Harvey KF (2009) Transcriptional output of the Salvador/ warts/hippo pathway is controlled in distinct fashions in Drosophila melanogaster and mammalian cell lines. Cancer Res 69(15): 6033–41.
61. Aung CS, Hill MM, Bastiani M, Parton RG, Parat MO (2010) PTRF-cavin-1 expression decreases the migration of PC3 prostate cancer cells: role of matrix metalloprotease 9. Eur J Cell Biol, 2010 Aug 21. [Epub ahead of print].
62. Hurst DR, Edmonds MD, Scott GK, Benz CC, Vaidya KS, et al. (2009) Breast cancer metastasis suppressor 1 up-regulates miR-146, which suppresses breast cancer metastasis. Cancer Res 69(4): 1279–83.
63. Cordes KR, Sheehy NT, White MP, Berry EC, Morton SU, et al. (2009) miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. Nature 460(7256): 705–10.
64. Kasai H, Allen JT, Mason RM, Kamimura T, Zhang Z (2005) TGF-β1 induces human alveolar epithelial to mesenchymal cell transition (EMT). Respir Res 6: 56.
65. Taguchi A, Yanagisawa K, Tanaka M, Cao K, Matsuyama Y, et al. (2008) Identification of hypoxiainducible factor-1 alpha as a novel target for miR-17-92 microRNA cluster. Cancer Res 68(14): 5540–5.
66. Kozaki K, Miyaishi O, Tsukamoto T, Tatematsu Y, Hida T, et al. (2000) 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 60(9): 2535–40.
67. Tokumaru S, Suzuki M, Yamada H, Nagino M, Takahashi T (2008) let-7 regulates Dicer expression and constitutes a negative feedback loop. Carcinogenesis 29(11): 2073–7.
68. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Royal Statist Soc B 58(1): 267–88.
69. Hoerl AE, Kennard R (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12: 55–67.
70. Shimamura T, Imoto S, Yamaguchi R, Fujita A, Nagasaki M, et al. (2009) Recursive regularization for inferring gene networks from time-course gene expression profiles. BMC Syst Biol 3: 41.
71. Shimamura T, Imoto S, Yamaguchi R, Nagasaki M, Miyano S (2010) Inferring dynamic gene networks under varying conditions for transcriptomic network comparison. Bioinformatics 26(8): 1064–72.
72. Mardia K, Kent J, Bibby J (1979) Multivariate Analysis Academic Press.
73. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Statist Soc Ser B 57(1): 289–300.
74. Hartung J, Knapp G, Sinha BK (2008) Statistical meta-analysis with applications. Wiley.

Cancer Science
The official journal of the Japanese Cancer Association

Review Article

let-7 and miR-17-92: Small-sized major players in lung cancer development

Hirotaka Osada[1,2] and Takashi Takahashi[3,4]

[1]Division of Molecular Oncology, Aichi Cancer Center Research Institute, [2]Department of Cancer Genetics, [3]Division of Molecular Carcinogenesis, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan

(Received April 8, 2010/Revised July 21, 2010/Accepted August 2, 2010/Accepted manuscript online August 5, 2010/Article first published online August 25, 2010)

MicroRNA (miRNA)-encoding small non-coding RNA have been recognized as important regulators of a number of biological processes that inhibit the expression of hundreds of genes. Accumulating evidence also indicates the involvement of miRNA alterations in various types of human cancer, including lung cancer, which has long been the leading cause of cancer death in economically well-developed countries, including Japan. We previously found that downregulation of members of the tumor-suppressive *let-7* miRNA family and overexpression of the oncogenic *miR-17-92* miRNA cluster frequently occur in lung cancers, and molecular insight into how these miRNA alterations may contribute to tumor development has been rapidly accumulating. The present review summarizes recent advances in elucidation of the molecular functions of these miRNA in relation to their roles in the pathogenesis of lung cancer. Given the crucial roles of miRNA alterations, additional studies are expected to provide a better understanding of the underlying molecular mechanisms of disease development, as well as a foundation for novel strategies for cancer diagnosis and treatment of this devastating disease. (*Cancer Sci* 2011; 102: 9–17)

Lung cancer, the number one killer

Lung cancer is the leading cause of cancer death in most economically developed countries, including Japan. Solid evidence indicates that the disease develops from accumulations of various genetic and epigenetic alterations[1,2] resulting in alterations of gene expression profiles, which are tightly associated with the clinicopathological features of lung cancer. MicroRNA (miRNA) in the human genome were only recently discovered[3] and accumulating evidence clearly indicates their roles in various crucial aspects of gene expression regulation. We initiated a search for miRNA that are dysregulated in lung cancer, which resulted in the discovery of major representative miRNA involved in lung cancer development with either tumor suppressive or oncogenic roles. These miRNA are members of the *let-7* miRNA family and among the most representative type of tumor suppressor miRNA,[4] along with the *miR-17-92* miRNA cluster, which is recognized as a typical oncogene-type miRNA.[5] There is a number of high-quality review articles dealing with the general roles of miRNA alterations in carcinogenesis,[6–9] thus in the present review we specifically focus on recent advances related to *let-7* and *miR-17-92*, with special emphasis on their relationships to lung carcinogenesis.

Discovery of miRNA in lower eukaryotes and humans

miRNA are evolutionally conserved approximately 22 nucleotide-long short non-coding RNA molecules. As of March 2010, 721 hairpin miRNA precursors and 1007 mature miRNA in the human genome had been deposited into the primary database (miRBase: http://www.mirbase.org/index.shtml). The genes first recognized to encode miRNA were *lin-4* and *let-7*, both of which were originally identified as heterochronic mutant genes and affect the progression of larval stages during the development of *C. elegans*.[10–12] As *C. elegans* develops, *lin-4* is upregulated in the first larval (L1) stage and suppresses expression of lin-14, thus promoting progression from the L1 to L2 stage. Subsequent downregulation of a second *lin-4* target, lin-28, is required for execution of the L3 larval stage. In contrast, *let-7* is expressed later in development and required for execution of the larval to adult (L/A) switch, which occurs at the end of the L4 stage. Mutations of *lin-4* and *let-7* have effects on the differentiation of seam stem cells, resulting in reiteration of the larval stages.[10–13]

miRNA are generated from long precursor transcripts and have an imperfectly matched stem-loop structure. These primary transcripts (pri-miRNA) are first processed into hairpin RNA (pre-miRNA) by a nuclear ribonuclease, Drosha, then transported to the cytoplasm and processed by a second ribonuclease, Dicer. Subsequently, the single stranded miRNA (mature miRNA) are incorporated into a miRNA-induced silencing complex (miRISC) and interact with ''seed'' sequence-matched recognition sites at the 3′ UTR of target mRNA. These miRNA–mRNA interactions result in inhibition of expression of the target genes at the post-transcriptional level through translational inhibition and mRNA destabilization.[14,15] Each miRNA directly represses, albeit mildly in general, hundreds of target genes, most of which contain conserved seed sequence(s) at the 3′ UTR. Because a large number of miRNA is present in the human genome, more than 60% of human protein-coding genes are targeted by miRNA,[16] suggesting that miRNA abnormalities may cause a wide spectrum of alterations in gene expressions.

let-7 alterations in lung cancer

In 2004, we reported that expression levels of the *let-7* family members are generally reduced in lung cancer when compared with those in normal lung tissues, indicating an association with poor prognosis in surgically treated patients who have tumors with low levels of *let-7* expression.[4] That study was the first report of *let-7* alterations in any type of cancer, as well as of the relationship of cancer patient prognosis with any type of miRNA alteration. Perhaps more importantly, our experimental finding that the introduction of *let-7* into a lung cancer cell line with a low level of *let-7* expression significantly inhibited the growth of lung cancer cells was the first direct indication that the miRNA expression level has an effect on the biological behavior

[4]To whom correspondence should be addressed.
E-mail: tak@med.nagoya-u.ac.jp

doi: 10.1111/j.1349-7006.2010.01707.x

of cancer cells. Subsequently, Slack's group identified K-ras as a target of *let-7* and showed that antisense-mediated inhibition of *let-7* increased cancer cell division, whereas overexpression of *let-7* induced cell-cycle arrest in cancer cell lines.[17] Together, these findings observed in human lung cancer cells appear to be consistent with the roles of *let-7* in *C. elegans*, as seam cells in *let-7* mutants fail to exit the cell cycle and reiterate the larval stage, showing dysregulation of the cell cycle and cell growth. The significance of reduced *let-7* expression in lung carcinogenesis was further supported in studies of genetically engineered mice. Jacks' group showed that *let-7* suppressed tumor initiation in an autochthonous non-small cell lung cancer (NSCLC) model of K-RasG12D transgenic mice, which was effectively rescued by ectopic expression of K-RasG12D lacking the 3′ UTR.[18] *let-7* also inhibited *in vitro* and *in vivo* growth of K-RasG12D-expressing murine lung cancer cells and human lung cancer xenografts.[19] Inhibitory effects of *let-7* against human lung cancer development have also been supported by circumstantial evidence reported by Chin *et al.*,[20] who sequenced *let-7* complementary sites (LCS) in the KRAS 3′ UTR from NSCLC cases and found that the single nucleotide polymorphism (SNP) at LCS6 was significantly associated with NSCLC patients who smoked <40 pack-years. Interestingly, they also showed that this SNP results in KRAS overexpression *in vitro*.

Each miRNA is thought to regulate tens or hundreds of protein coding genes, thus it is reasonable to speculate that *let-7* downregulates other growth-promoting genes, such as oncogenes (Fig. 1). Indeed, HMGA2, which encodes a non-histone chromosomal high-mobility group protein with a putative oncogenic function, has been shown to be under the control of *let-7*.[21] In several types of malignancy, the HMGA2 gene locus is disrupted by chromosomal translocation and loses its 3′ UTR that harbors multiple *let-7* recognition sites, while HMGA2 promotes anchorage-independent growth.[22] In mice, Hmga2 promotes self-renewal of fetal and young-adult neural stem cells, partly by decreasing p16Ink4a/p19Arf expression, while *let-7* expression, which increases with age, negatively affects Hmga2 expression and self-renewal capacity.[23] Other targets of *let-7* include various cell-cycle-related genes such as cyclin D2, CDK6 and CDC25A,[24] as well as various oncofetal genes, including insulin-like growth factor 2 mRNA binding protein 1 (IGF2BP1, also called IMP-1/CRD-BP) and IGF2BP2/IMP-2,[25] which are known to bind various mRNA and regulate their translation, leading to stabilization of crucial mRNA such as Myc.[25]

Shell *et al.*[26] also reported the importance of *let-7* in cancer classification. Cancer cell lines can be divided into two groups, epithelial type (II) and mesenchymal type (I), suggesting a progression from type II to type I through epithelial–mesenchymal transition (EMT).[27] Shell *et al.*[26] found that cancer cell lines that exhibit epithelial-type characteristics express higher levels of *let-7* than those with mesenchymal features, and suggested that loss of *let-7* expression may be a marker for less differentiated and advanced cancer. Also, a joint study conducted by the laboratories of Croce and Harris reported associations of miRNA profiles with survival of patients with lung adenocarcinomas, and showed that high expression of *miR-155* and low expression of *let-7a-2* were strongly associated with poor survival.[28] An association of reduced *let-7a* level with unfavorable postoperative prognosis in patients with NSCLC was also reported by Yu *et al.* using quantitative RT-PCR-based analysis, in which a poorer prognosis was shown to be associated with reduced *let-7* and *miR-221* expression, as well as with increased levels of *miR-137*, *miR-372* and *miR-182**.[29] Interestingly, a search for miRNA differentially altered between lung cancer patients who never smoked and those who were smokers showed that downregulation of *let-7c* and *miR-138* was preferentially present in the never-smoked patients.[30]

Fine tuning of *let-7* expression level and cancer

In addition to the cancer-related genes described above, *let-7* appears to have another interesting target (Fig. 1). We found that *let-7* directly downregulates Dicer through binding sites at the 3′ UTR.[31] Dicer is an essential endonuclease required at the final processing step in miRNA biogenesis that includes *let-7*. Overexpression of *let-7* reduces the expression of Dicer as well as that of a large number of other mature miRNA, whereas antisense-mediated inhibition of *let-7* leads to upregulation of Dicer expression associated with increased expression levels of mature miRNA.[31] The existence of three conserved *let-7* target sites within the open reading frame in Dicer was also reported,[32] although they appear to be less efficient than the 3′ UTR binding sites (Tokumaru S and Takahashi T, unpublished observation 2008). Therefore, the existence of *let-7*-mediated negative regulation of Dicer may provide a basis for the tightly regulated equilibration of expression levels of Dicer and *let-7*, as well as of other miRNA. Interestingly, *let-7* appears to be a constituent of another regulatory loop within the miRNA processing steps (Figs 1,2). Lin28 was shown to be a direct target for *let-7*-mediated inhibition, while it in turn inhibits Drosha- and/or Dicer-mediated processing of *let-7*.[33,34] Both Lin28 and a homologue, Lin28B, are overexpressed in approximately 15% of primary human tumor samples in association with reduced expression of the entire *let-7* family, as well as with a poor clinical prognosis.[35] Furthermore, negative regulation of the *let-7* family by Lin28 and Lin28B involves induction of uridylation of the *pre-let-7* 3′-terminus.[36] In addition, Lin28 proteins may directly recruit the uridylating enzyme TUTase4 (TUT4),[37] also known as zinc finger, CCHC domain containing 11 (Zcchc11),[38] to *pre-let-7*. The terminal uridylation of *pre-let-7* blocks Dicer processing and also promotes its decay, while a tetra-nucleotide sequence motif (GGAG) in the terminal loop is recognized by Lin28. Thus, other miRNA with the same loop sequence motif may also be regulated via the same mechanism. Indeed, Zcchc11 has been shown to uridylate *miR-26a* targeting IL-6 and stabilize IL-6 transcripts.[39]

It is notable that reduced Dicer expression appears to be involved in tumor development. We previously reported an association of reduced expression of Dicer with poor prognosis in lung cancer patients.[40] Consistent with that finding in human lung cancer, Jacks' group reported that knockdown of Dicer1 accelerated the tumorigenicity and invasiveness of a mouse lung adenocarcinoma cell line, while conditional deletion of Dicer1 enhanced tumor development in a K-Ras-induced mouse model of lung cancer.[41]

Maintenance of stemness in relation to *let-7* expression

Oct4, Sox2 and Nanog, core regulators of ES cell differentiation, co-occupy the promoters of differentiation-related transcriptional factors and also several miRNA, suggesting miRNA plays a role in regulation of differentiation.[42] A number of mature miRNA are not expressed in ES or P19 EC cells, whereas they are expressed at the late embryonic stage. Lin-28 binds conserved nucleotides in the loop region of *let-7* precursors,[43] and effectively blocks their cleavage by the Drosha-DGCR8 microprocessor in the nucleus[33,43] and by Dicer in the cytoplasm[34] of embryonic stem cells (Fig. 1). In neuronal stem (NS) cells, which are more differentiated than ES cells, Lin-28 is downregulated by *mir-125* (*lin-4* homologue) and *let-7*, which allows *pre-let-7* processing to proceed. Suppression of *let-7* or *mir-125* activity in NS cells leads to upregulation of Lin-28 and loss of *pre-let-7* processing activity, suggesting that *let-7*, *mir-125* and Lin-28 participate in an autoregulatory circuit that controls miRNA processing during NS cell commitment.[34] Thus, Lin28 functions as a negative regulator of miRNA biogenesis, and may play a central role in blocking

doi: 10.1111/j.1349-7006.2010.01707.x

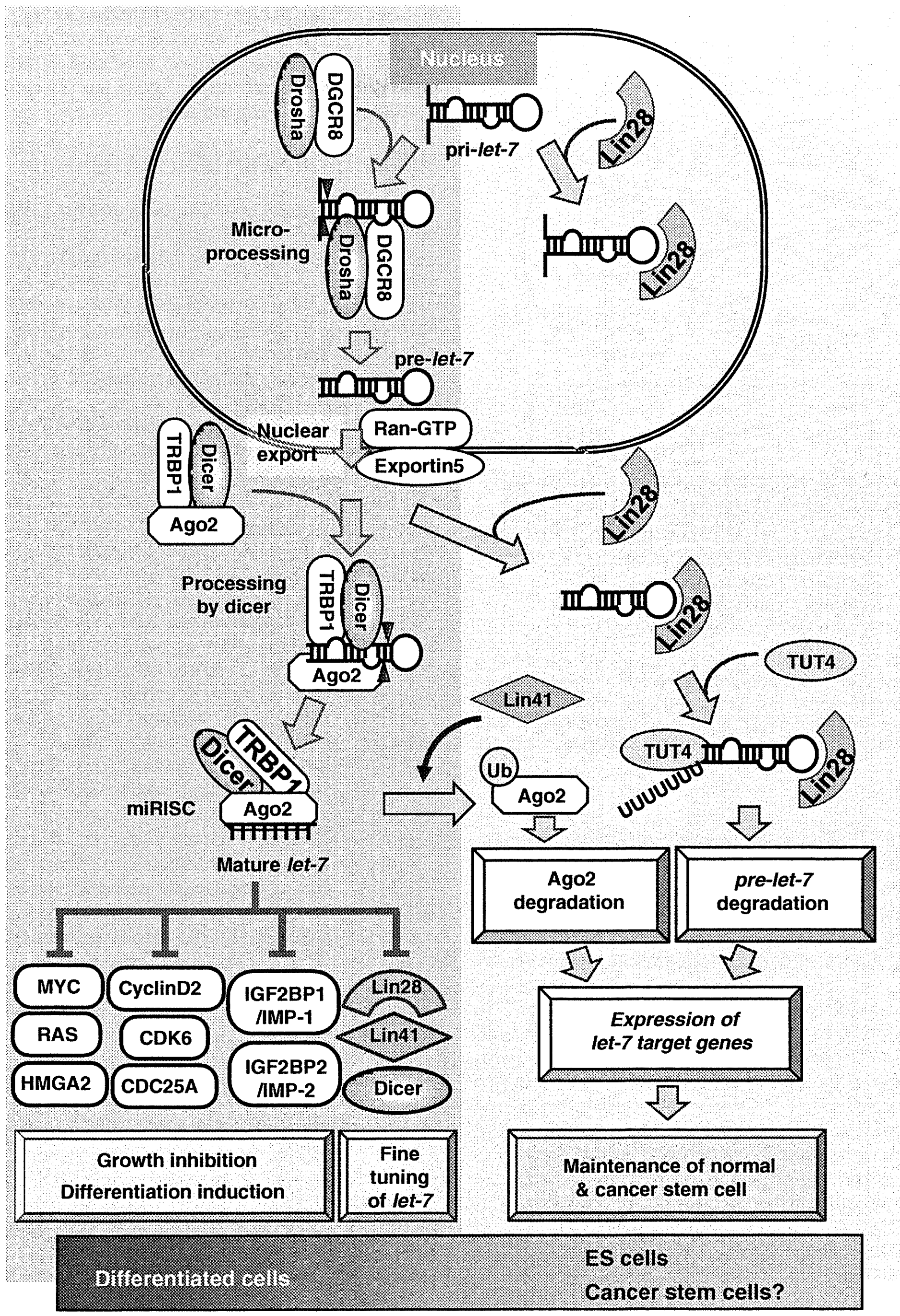

Fig. 1. Regulation of biogenesis of *let-7*. MicroRNA (miRNA) are transcribed by RNA polymerase II as a long transcript, pri-miRNA, and then processed sequentially in the nucleus and cytoplasm. The biogenesis and functions of *let-7* are regulated by Lin41, Lin28 and TUT4 in a complex manner, while *let-7* plays the role of tumor suppressor by inhibiting the expression of tumor-promoting genes such as RAS and HMGA2.

miRNA-mediated differentiation in stem cells and certain cancers (Figs 1,2).

In addition to Lin-28, the zinc finger protein Lin41 is also a target of *let-7* and involved in the regulatory network that controls pluripotency.[44] Lin41 interacts with Dicer and the Ago family at P-body, and acts as an E3 ubiquitin ligase, mediating the ubiquitylation of Ago2.[44] Therefore, Lin41 negatively regulates *let-7* activity and co-operates with Lin28 in stem cells (Figs 1,2). These findings indicate the importance of Lin-28 to maintain pluripotency and are consistent with the finding that Lin-28 is included in a cocktail of reprogramming factors (Oct3/4, Sox-2, Nanog, Lin28) to create induced pluripotent stem (iPS) cells from adult human fibroblasts.[45] In addition, Myc directly associates with the Lin28B promoter to induce Lin-28B expression, resulting in *let-7* repression. Accordingly, Lin-28B loss-of-function significantly impairs Myc-dependent

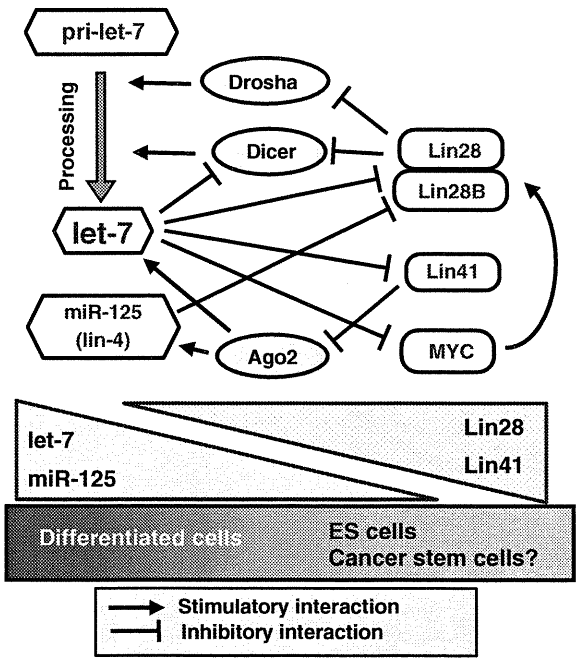

Fig. 2. Fine tuning of the expression level of mature *let-7* balancing stemness and differentiation. Lin28/Lin28B and Lin41 are conserved targets of the *let-7* family, while the *let-7* family is conversely under negative regulation via inhibition of Drosha and Dicer byLin28/Lin28B and Ago2 by Lin41, thus implementing possible positive feedback regulations.

cellular proliferation.(46) The self-renewing progenitor population in mouse mammary epithelial cells shows a unique miRNA signature of high expression levels of *miR-205* and *miR-22*, and depletion of *let-7* and *miR-93*, while enforced *let-7* expression was shown to induce loss of the self-renewing population, suggesting negative regulation of tissue progenitor maintenance by *let-7*.(47)

Other lines of evidence suggest the involvement of *let-7* in carcinogenesis in relation to its function to regulate differentiation. For example, *let-7* expression is markedly reduced in mammospheres/tumor-initiating cells of breast cancer and increased along with cell differentiation.(48) Conversely, forced expression of *let-7* has been shown to reduce cellular proliferation and mammosphere formation, as well as *in vivo* tumor formation and metastasis. Interestingly, silencing of H-RAS reduced the self-renewal of mammospheres but had no effect on differentiation, while that of HMGA2 enhanced differentiation but did not affect self-renewal, suggesting that both H-RAS and HMGA2 are major target genes of *let-7* and de-repression of both is involved separately in tumorigenesis.(48) In addition, those findings indicate an important role for *let-7* and its regulation in the regulation of pluripotency. In contrast to *let-7*, several miRNA such as the members of the *miR-290* family are expressed specifically in ES cells(49) and positively regulate ES cell self-renewal.(50,51) Dgcr8-deficient ES cells are unable to suppress self-renewal, because of defective biogenesis of miRNA. However, introduction of *let-7* can suppress self-renewal and induce differentiation, whereas *miR-294*, an ES cell-specific miRNA, blocks the suppression of self-renewal by *let-7*, suggesting that *let-7* and ES cell-specific miRNA alternatively regulate ES cell fate, that is, self-renewal *versus* differentiation.(51) Our recent miRNA microarray analysis findings showed that lung adenocarcinomas are grouped into four major clusters with distinct miRNA expression profiles. Along the same line, it is interesting that one of the clusters with characteristically low *let-7* and high *miR-17-92* expression levels was related to a significantly worse prognosis, and those patients exhibited significantly higher dysregulation of ES cell-related gene sets (Arima C and Takahashi T, manuscript in preparation).

miR-17-92 overexpression in lung cancer

Our initial discovery of frequent downregulation of *let-7* and its biological and clinicopathological involvement in lung cancer prompted us to search for miRNA conversely overexpressed in lung cancers.(4) Consequently, we found frequent and marked overexpression, with occasional gene amplification, of clustered miRNA (*miR-17-92*) within intron 3 of the C13orf25 gene at 13q31.3 in lung cancer samples, especially those with a small cell lung cancer (SCLC) histology.(5) Stimulatory activity by this miRNA cluster toward lung cancer cell growth was observed, while antisense-mediated inhibition of *miR-17-5p* and *miR-20a*, constituents of *miR-17-92*, induced apoptosis in *miR-17-92*-overexpressing lung cancer cell lines, suggesting an addiction to continued overexpression of *miR-17-92* for cancer development. In contrast to our approach, Hammond *et al* initiated a study based on evidence suggesting the involvement of the C13orf25 genomic region in B-cell lymphomas, as previously reported by Ota *et al.*(52) in the results of detailed array CGH analysis. Consequently, they identified overexpression of *miR-17-92* in occasional association with gene amplification in B-cell lymphomas(53) and showed that introduction of *miR-17-92* into hematopoietic stem cells in Eμ-myc transgenic mice significantly accelerated formation of lymphoid malignancies. MYC transactivates expression of the *miR-17-92* miRNA cluster,(54) while members of the myc gene family have been shown to be frequently amplified and/or overexpressed in SCLC.(55) Interestingly, our previous studies of *miR-17-92* and the myc gene family in lung cancers suggested the existence of two potential mechanisms that lead to overexpression of the *miR-17-92* cluster, that is, gene amplification of the miRNA cluster itself and increased expression of the myc gene family, with or without gene amplification. It is also important to note that a significant role of the *miR-17-92* cluster in tumorigenesis is also supported by frequent retrovirus integration-mediated activations of mouse *miR-17-92*(56) and paralogous *miR-160a-363*(57) in mouse tumors.

Myc-E2F-*miR-17-92* network

Overexpression of E2F1 induces inappropriate entry into the S-phase, resulting in apoptosis induction.(58) MYC and E2F1 positively regulate each other, while MYC-induced *miR-17-92* negatively regulates E2F1,(54) suggesting possible fine tuning of the E2F1 expression level for correct regulation of S-phase entry. In addition to Myc, the E2F family also transactivates *miR-17-92*, which exerts a negative feedback loop, resulting in suppression of E2F family expression.(59,60) Therefore, the expression levels of MYC, the E2F family and *miR-17-92* are finely regulated by each other, suggesting their crucial roles in cell-cycle regulation (Fig. 3). *miR-17-92* is preferentially overexpressed in lung cancers with neuroendocrine features, especially in SCLC, which is known to exhibit overexpression of members of the MYC gene family with frequent gene amplification. We reported that survival of lung cancer cell lines with *miR-17-92* overexpression relies on the continued expression of *miR-17-92*.(61) Interestingly, we also found frequent accumulation of constitutively phosphorylated H2AX (γ-H2AX), which reflects persistent DNA damage, preferentially in SCLC. Small cell lung cancers almost invariably carry inactivated retinoblastoma (RB) and p53, which conceivably contributes to elicit dysregulated cell-cycle progress, leading to replication-dependent

doi: 10.1111/j.1349-7006.2010.01707.x

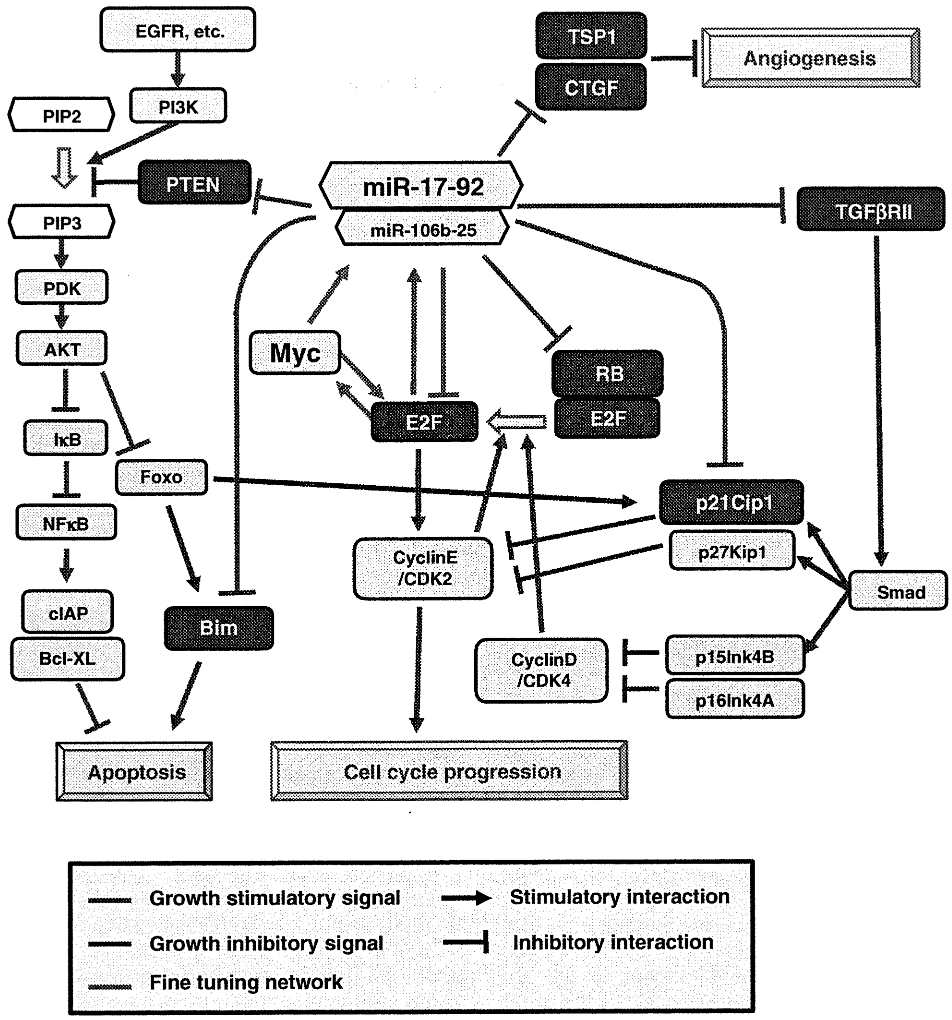

Fig. 3. Tumor growth stimulatory and apoptosis inhibitory regulations by *miR-17-92* and its paralogous microRNA (miRNA) clusters via inhibition of their target genes (marked with white letters on a blue background). The finely tuned network involving *miR-17-92* Myc, E2F and Rb is indicated by purple lines. EGFR; epidermal growth factor receptor.

DNA double-strand breaks. In fact, in NSCLC cells with wild-type RB, knockdown of RB induced γ-H2AX foci formation and growth inhibition in NSCLC cells with wild-type RB, which was canceled by overexpression of *miR-20a*. In addition, suppression of *miR-20a* with antisense-oligonucleotides further induced γ-H2AX foci formation in a *miR-17-92*-overexpressing SCLC cell line.[62] RB disruption also induces ROS, which are negatively regulated by *miR-20a*. Therefore, *miR-17-92* overexpression may serve as a fine-tuning influence to counterbalance the generation of DNA damage in RB-inactivated SCLC cells, thus reducing excessive DNA damage to a tolerable level and consequently leading to genetic instability (Fig. 4).[62] These findings are consistent with the report by Pickering *et al.*,[63] who showed the role of *miR-17/miR-20a* in the cell-cycle regulation of fibroblasts. Inhibition of *miR-17/miR-20a* leads to G1 checkpoint activation due to an accumulation of DNA double-strand breaks, resulting from premature temporal accumulation of the E2F1 transcription factor. Thus, Myc-regulated *miR-17/miR-20a* appears to play a role in controlling the precise timing of E2F1expression and circumventing the G1 checkpoint caused by E2F1 accumulation, which is perturbed in cancer overexpressing *miR-17-92*. It is also important to note that the consequences of coupling between the E2F/Myc positive feedback and E2F/Myc/*miR-17-92* negative feedback loops have been analyzed using a mathematical model, which predicted that *miR-17-92* plays a critical role in regulating the position of the on–off switch related to E2F/Myc protein levels.[64] Cyclin D1 may also be involved in this *miR-17/miR-20a* negative feedback loop in breast cancer.[65]

Other targets of *miR-17-92* related to cancer

Each miRNA may potentially influence more than 100 target mRNA. Accordingly, a search for targets of *miR-17-92*, which are actually affected in immortalized lung epithelial cells by the components of this miRNA cluster, was conducted through global expression profiling using differential tagging with iTRAQ™ reagent, followed by multidimensional liquid chromatography and tandem mass spectrometry analysis, which resulted in identification of HIF-1α as a target for *miR-17-92* (Fig. 4).[66] Interestingly, an intricate and finely tuned circuit involving c-myc, HIF-1α and *miR-17-92* exists and plays a role in cancer cell proliferation under normoxia in a cellular context-dependent manner without interfering with the robust induction of HIF-1α for cellular adaptation to hypoxia. Yan *et al.*[67] recently reported that hypoxia reduced *miR-17-92* expression in colon cancer cells

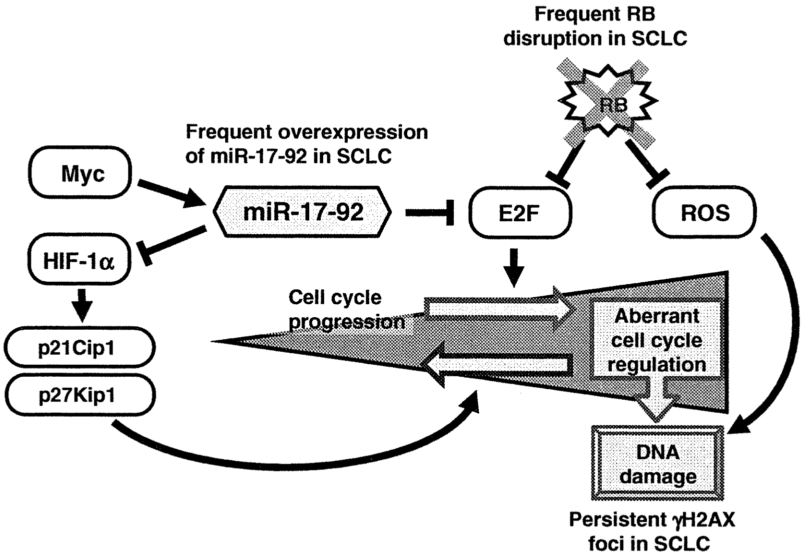

Fig. 4. Counterbalance between RB inactivation and *miR-17-92* overexpression in SCLC. SCLC tumors almost invariably carry RB inactivation and frequently exhibit *miR-17-92* overexpression, which potentially attenuates the aberrant cell-cycle progression and consequential excessive DNA damage in cells with RB inactivation.

through p53-mediated repression by its direct binding to the promoter of *miR-17-92* and consequential competition with the TATA-binding protein (TBP). They also showed that forced expression of *miR-17-92* markedly inhibited hypoxia-induced apoptosis, whereas antisense-mediated inhibition of *miR-17-5p* and *miR-20a* sensitized the cells to hypoxia-induced apoptosis, indicating that p53-mediated repression of *miR-17-92* expression is likely to have an important function in hypoxia-induced apoptosis. In contrast, we did not detect a readily noticeable change in *miR-17-92* expression under hypoxia in an immortalized normal bronchial epithelial cell line,[66] suggesting that there might be different effects depending on the cellular contexts.

Additional targets for *miR-17-92* have been reported in studies that used various systems (Fig. 3). Transgenic mice carrying the *miR-17-92* transgene conditionally active in lymphocytes showed increased proliferation and reduced activation-induced cell death of lymphocytes, resulting in lethal lymphoproliferative and autoimmune diseases.[68] That study also found that *miR-17-92* miRNA suppressed the expression of Pten and Bim, both of which contribute to the phenotype. BIM, a proapoptotic BCL2 family member, functionally inhibits anti-apoptotic BCL2 family members through physical interaction and plays an essential role in apoptosis induction during lymphocyte differentiation. PTEN encoding phosphatidylinositol-3,4,5-trisphosphate (PIP3) 3-phosphatase inhibits activation of the PDK1–AKT signaling pathway through inhibition of PIP3 generation and is frequently inactivated by mutations in several cancer types. Disruption of both genes induces lymphoproliferative and autoimmune diseases, suggesting that the lethal phenotype is attributable mainly to repression of PTEN and Bim by *miR-17-92*.[68] Meanwhile, disruption of *miR-17-92* leads to upregulation of Bim and inhibits B cell development during the transition from pro-B to pre-B.[69] Another tumor suppressor gene, CDKN1A (p21Waf1/Cip1), is also repressed by *miR-17*, *miR-20a* and *miR-106b*.[62,70,71]

It has also been shown that *miR-17-92* is involved in regulation of angiogenesis. Although K-Ras-transformed colonocytes were shown to form indolent and poorly vascularized tumors, transduction of the Myc gene caused upregulation of *miR-17-92* in K-Ras-colonocytes and neovascularization in related tumors, in association with downregulation of anti-angiogenic thrombospondin-1 (Tsp1) and a related protein, connective tissue growth factor (CTGF).[72] In addition, antisense-mediated suppression of *miR-17-92* expression partly restored Tsp1 and CTGF expressions, while transduction of *miR-17-92* reduced Tsp1 and CTGF levels, resulting in larger, better-perfused tumors.[72] Similarly, vascular endothelial growth factor (VEGF) induced expression of *miR-17-92* in endothelial cells, which was shown to be via *miR-18a*-mediated inhibition of Tsp1expression.[73] These results suggest a possible role of *miR-17-92* overexpression in tumor angiogenesis in lung cancer. In contrast to these reports on the proangiogenic effects of *miR-17-92*, forced overexpression of *miR-92a* in endothelial cells was shown to block angiogenesis both *in vitro* and *in vivo* through repression of several proangiogenic proteins (integrin α-subunits, etc.).[74] In contrast, a different network was observed in chronic lymphocytic leukemia (CLL), as upregulation of *miR-92* was found to contribute to repression of von Hippel-Lindau tumor suppressor gene (VHL) expression under a normoxic condition in CLL cells, which led to reduced ubiquitination and degradation of HIF-1α, and consequential autocrine stimulation of VEGF secretion.[75] Such overexpression of *miR-17-92* observed in lung cancer may contribute to angiogenesis. Therefore, *miR-17-92* may regulate the angiogenic network positively or negatively in a cellular context-dependent manner.

Paralogous clusters of *miR-17-92*

In the mammalian genome, there are three paralogous miRNA clusters; *miR-17-92*, *miR-106a-363* and *miR-106b-25*, among which the *miR-17-92* and *miR-106b-25* clusters have similar expression patterns in adult mice, while the expression level of the *miR-106a-363* cluster is generally undetectably low.[69] *miR-106b-25* is localized in an intron of the MCM7 gene, which is involved in licensing of DNA replication, and is transcriptionally regulated by E2F1 and MYC, similar to *miR-17-92*. *miR-106b-25* was reported to be overexpressed in gastric cancer,[76] while it has also been shown that overexpression of *miR-106b-25* modulates transforming growth factor (TGF)-β-induced cell-cycle arrest and apoptosis through inhibition of CDKN1A (p21Waf1/Cip1) and BIM, respectively.[76] However, a large body of evidence points to crucial involvement of *miR-17-92* in tumor development among these three paralogous miRNA clusters. Along this line, it is interesting that only the *miR-17-92* cluster contains *miR-18* and *miR-19*, which are absent in other miRNA clusters, while *miR-19* was suggested to be crucially involved as a key oncogenic miRNA in mice models of lymphoma development through inhibition of PTEN expression and consequent activation of AKT-mTOR and apoptosis repression.[77,78]

doi: 10.1111/j.1349-7006.2010.01707.x

miR-17-92 in lung development

There are several lines of evidence that support the notion that miRNA are crucially involved in lung development.[69,79–81] Dicer deficiency induces branching arrests without epithelial growth arrest, resulting in a few large epithelial pouches. Therefore, miRNA processed by Dicer appear to play important roles in regulating lung epithelial morphogenesis.[79] miRNA expression profiling analysis has also shown that *miR-17-92* clusters are abundantly expressed at the early stages of lung development, while the expression level declines as development proceeds. In contrast, the *let-7* miRNA family has an inverse expression pattern and becomes predominant at the late stage.[80,82] Since the expression pattern suggests a physiological role of *miR-17-92* in the early development of the lung, SPC-*miR-17-92* transgenic mice were produced, which demonstrated expansion of the distal epithelial progenitors and increases in neuroendocrine cell clusters, indicating that *miR-17-92* promotes a high level of proliferation and an undifferentiated phenotype of normal lung epithelial progenitors. Meanwhile, disruption of *miR-17-92* clusters was shown to cause lethal abnormalities, including lung hypoplasia, ventricular septal defects and inhibition of B cell development.[69] In contrast, ablation of either *miR-106b-25* or *miR-106a-363* had no obvious phenotypic consequences. Interestingly, combined disruption of both *miR-106b-25* and *miR-17-92* resulted in a more severely lethal phenotype,[69] suggesting an additive effect of *miR-106b-25*. Crucial roles of *miR-17*, *miR-20a* and *miR-106b*, all of which are highly expressed at the pseudo-glandular stage of embryonic lung development, were also reported by Carraro *et al.*[81] In that study, expression of these *miR-17* family members was suppressed in explants of isolated lung epithelium, and experimental results showed that these miRNA modulate FGF10-induced budding morphology by specifically targeting the signal transducer and activator of transcription 3 (STAT3), as well as mitogen-activated protein kinase 14 (MAPK14), which are FGF10-FGFR2β downstream signal mediators.[81] These results indicate a tight relationship between oncogenic properties and physiological functions of *miR-17-92* in the lung.

Mechanisms of dysregulation of *let-7* and *miR-17-92* in cancer

Elucidation of the molecular mechanisms of miRNA dysregulation is of immense interest and should help to better explain the global picture of the molecular pathogenesis of cancer, which would eventually lead to development of therapeutic strategies targeting miRNA abnormalities. Transcriptional repression, epigenetic silencing and genetic alteration may play roles in the reduced expression of *let-7*, as has been shown following down-regulation of protein-coding genes. Among the 11 *let-7* family members, six are localized within cancer-associated genomic regions or in fragile sites,[83] while there are also lung cancer cell lines that harbor homozygous deletions of the *let-7c* cluster at 21q11.2- q21.1[84]. Epigenetic silencing has been specifically reported in *let-7a-3*[85,86], although cancer-related epigenetic silencing has not been reported in other *let-7* family members. Furthermore, expression of the *let-7* family was reported to be under the influence of direct repression by c-MYC.[87]

Aberrations in miRNA processing also appear to be involved. *let-7* biogenesis is controlled by multiple layers of regulation, including negative regulation by LIN28, as discussed above. Along this line, c-MYC overexpression indirectly suppresses the expression level of mature *let-7* through induction of LIN28,[46] which inhibits the processing of *let-7* precursors. LIN28/LIN28B have also been shown to be induced by overexpression of c-MYC[46,88] as well as NF-κB activation,[89] both of which are known to be frequent in lung cancers.

Overexpression of *miR-17-92* appears to be caused by transcriptional activation and/or genetic amplification. The *miR-17-92* cluster is transactivated by c-MYC,[54] E2F1/E2F3[59,60] and STAT3,[90] each of which are frequently activated in cancer. In addition, a paralogous cluster, *miR-106b-25*, is transcriptionally upregulated together with a host gene, MCM7, by E2F1.[76] Inactivation of p53, which is frequently present in various types of cancer including lung cancer, may also be involved, since transcription of the *miR-17-92* cluster has been shown to be repressed by this tumor suppressor.[67] Furthermore, we previously reported occasional association of the gene amplification of *miR-17-92* with its overexpression in lung cancers,[5] while our preliminary analysis of a CGH dataset at Sanger Institute (http://www.sanger.ac.uk) showed an association of focal amplification/gain of the *miR-17-92* locus with SCLC histology and large cell carcinomas (data not shown), confirming our previous report.[5] Also, Re *et al.*[91] performed a genome-wide survey and reported the possible existence of feed-forward regulatory circuits involving microRNA and transcription factors, including those of *let-7* and *miR-17-92*. Given that computing power continues to increase, future detailed investigations of genome-wide mRNA–miRNA networks using high-powered computing methods will be of particular interest and should provide in-depth insight into the molecular mechanisms of dysregulation of miRNA in cancer.

Conclusion

Findings thus far reported clearly point to crucial roles for *let-7* and *miR-17-92* in the pathogenesis and progression of lung cancer, as they appear to affect the machinery of two key cellular functions, stemness maintenance and cell-cycle regulation. Several relevant targets for *let-7* and *miR-17-92* have been identified, and suggested to play roles in cancer development. However, we are far from gaining a complete picture of the dysregulation involved in the complex regulatory networks related to these miRNA. In addition, the world of non-coding RNA is rapidly expanding. Recent reports have demonstrated a miRNA-like function of snoRNA[92] and a novel RNA decoy function of miRNA.[93] Each miRNA is thought to regulate hundreds of target mRNA, which in turn regulate multiple genes, including protein-coding genes and miRNA, while tens of thousands of non-coding RNA other than miRNA are known to be transcribed from the human genome. Thus, it would be reasonable to predict the future necessity of a radically different approach to elucidate the resultant unbelievably complex regulatory networks present in cells in both normal and cancerous conditions. Along this line, a cancer systems biology approach with the aid of ever evolving computing power may help to show how these indispensably informative pieces of an as yet unresolved puzzle fit into a comprehensive understanding of lung cancer biology. Therapeutic application of such acquired knowledge of miRNA alterations in cancer remains a daunting challenge, although additional information should ultimately lead us to the answers we seek.

Acknowledgments

We thank all members of our laboratories at Nagoya University and Aichi Cancer Center for their invaluable contributions, including helpful discussions and critical comments. In addition, we apologize for the incompleteness of the referencing due to space constraints. Studies performed in our laboratories were supported in part by grants from the Ministry of Education, Science, Sports and Culture, Japan, and from the Ministry of Health, Labour, and Welfare, Japan, as well as from the Princess Takamatsu Cancer Research Fund and Uehara Memorial Foundation.

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**: 7421–34.

2 Takahashi T, Sidransky D. Biology of lung cancer. In: Mason R, Broaddus V, Murray J, Nadel J, eds. *Textbook of Respiratory Medicine*, 4th edn. Philadelphia: Elsevier Science, 2005; 1311–27.

3 Pasquinelli AE, Reinhart BJ, Slack F *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 2000; **408**: 86–9.

4 Takamizawa J, Konishi H, Yanagisawa K *et al.* Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* 2004; **64**: 3753–6.

5 Hayashita Y, Osada H, Tatematsu Y *et al.* A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res* 2005; **65**: 9628–32.

6 Osada H, Takahashi T. MicroRNAs in biological processes and carcinogenesis. *Carcinogenesis* 2007; **28**: 2–12.

7 Schickel R, Boyerinas B, Park SM, Peter ME. MicroRNAs: key players in the immune system, differentiation, tumorigenesis and cell death. *Oncogene* 2008; **27**: 5959–74.

8 Negrini M, Nicoloso MS, Calin GA. MicroRNAs and cancer–new paradigms in molecular oncology. *Curr Opin Cell Biol* 2009; **21**: 470–9.

9 Croce CM. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* 2009; **10**: 704–14.

10 Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993; **75**: 843–54.

11 Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 1993; **75**: 855–62.

12 Reinhart BJ, Slack FJ, Basson M *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* 2000; **403**: 901–6.

13 Nimmo RA, Slack FJ. An elegant miRror: microRNAs in stem cells, developmental timing and cancer. *Chromosoma* 2009; **118**: 405–18.

14 Selbach M, Schwanhausser B, Thierfelder N *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* 2008; **455**: 58–63.

15 Baek D, Villen J, Shin C *et al.* The impact of microRNAs on protein output. *Nature* 2008; **455**: 64–71.

16 Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009; **19**: 92–105.

17 Johnson SM, Grosshans H, Shingara J *et al.* RAS is regulated by the let-7 microRNA family. *Cell* 2005; **120**: 635–47.

18 Kumar MS, Erkeland SJ, Pester RE *et al.* Suppression of non-small cell lung tumor development by the let-7 microRNA family. *Proc Natl Acad Sci USA* 2008; **105**: 3903–8.

19 Trang P, Medina PP, Wiggins JF *et al.* Regression of murine lung tumors by the let-7 microRNA. *Oncogene* 2010; **29**: 1580–7.

20 Chin LJ, Ratner E, Leng S *et al.* A SNP in a let-7 microRNA complementary site in the KRAS 3′ untranslated region increases non-small cell lung cancer risk. *Cancer Res* 2008; **68**: 8535–40.

21 Lee YS, Dutta A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev* 2007; **21**: 1025–30.

22 Mayr C, Hemann MT, Bartel DP. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science* 2007; **315**: 1576–9.

23 Nishino J, Kim I, Chada K, Morrison SJ. Hmga2 promotes neural stem cell self-renewal in young but not old mice by reducing p16Ink4a and p19Arf Expression. *Cell* 2008; **135**: 227–39.

24 Johnson CD, Esquela-Kerscher A, Stefani G *et al.* The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res* 2007; **67**: 7713–22.

25 Boyerinas B, Park SM, Shomron N *et al.* Identification of let-7-regulated oncofetal genes. *Cancer Res* 2008; **68**: 2587–91.

26 Shell S, Park SM, Radjabi AR *et al.* Let-7 expression defines two differentiation stages of cancer. *Proc Natl Acad Sci USA* 2007; **104**: 11400–5.

27 Algeciras-Schimnich A, Pietras EM, Barnhart BC *et al.* Two CD95 tumor classes with different sensitivities to antitumor drugs. *Proc Natl Acad Sci USA* 2003; **100**: 11445–50.

28 Yanaihara N, Caplen N, Bowman E *et al.* Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 2006; **9**: 189–98.

29 Yu SL, Chen HY, Chang GC *et al.* MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell* 2008; **13**: 48–57.

30 Seike M, Goto A, Okano T *et al.* MiR-21 is an EGFR-regulated anti-apoptotic factor in lung cancer in never-smokers. *Proc Natl Acad Sci USA* 2009; **106**: 12085–90.

31 Tokumaru S, Suzuki M, Yamada H, Nagino M, Takahashi T. let-7 regulates Dicer expression and constitutes a negative feedback loop. *Carcinogenesis* 2008; **29**: 073–7.

32 Forman JJ, Legesse-Miller A, Coller HA. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci USA* 2008; **105**: 14879–84.

33 Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin28. *Science* 2008; **320**: 97–100.

34 Rybak A, Fuchs H, Smirnova L *et al.* A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat Cell Biol* 2008; **10**: 987–93.

35 Viswanathan SR, Powers JT, Einhorn W *et al.* Lin28 promotes transformation and is associated with advanced human malignancies. *Nat Genet* 2009; **41**: 843–8.

36 Heo I, Joo C, Cho J *et al.* Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* 2008; **32**: 276–84.

37 Heo I, Joo C, Kim YK *et al.* TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* 2009; **138**: 696–708.

38 Hagan JP, Piskounova E, Gregory RI. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol Biol* 2009; **16**: 1021–5.

39 Jones MR, Quinton LJ, Blahna MT *et al.* Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat Cell Biol* 2009; **11**: 1157–63.

40 Karube Y, Tanaka H, Osada H *et al.* Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer Sci* 2005; **96**: 111–15.

41 Kumar MS, Pester RE, Chen CY *et al.* Dicerl functions as a haploinsufficient tumor suppressor. *Genes Dev* 2009; **23**: 2700–4.

42 Boyer LA, Lee TI, Cole MF *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 2005; **122**: 947–56.

43 Newman MA, Thomson JM, Hammond SM. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* 2008; **14**: 1539–49.

44 Rybak A, Fuchs H, Hadian K *et al.* The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2. *Nat Cell Biol* 2009; **11**: 1411–20.

45 Yu J, Vodyanik MA, Smuga-Otto K *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007; **318**: 917–20.

46 Chang TC, Zeitels LR, Hwang HW *et al.* Lin-28B transactivation is necessary for Myc-mediated let-7 repression and proliferation. *Proc Natl Acad Sci USA* 2009; **106**: 3384–9.

47 Ibarra I, Erlich Y, Muthuswamy SK, Sachidanandam R, Hannon GJ. A role for microRNAs in maintenance of mouse mammary epithelial progenitor cells. *Genes Dev* 2007; **21**: 3238–43.

48 Yu F, Yao H, Zhu P *et al.* let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* 2007; **131**: 1109–23.

49 Houbaviy HB, Murray MF, Sharp PA. Embryonic stem cell-specific MicroRNAs. *Dev Cell* 2003; **5**: 351–8.

50 Wang Y, Baskerville S, Shenoy A *et al.* Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet* 2008; **40**: 1478–83.

51 Melton C, Judson RL, Blelloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* 2010; **463**: 621–6.

52 Ota A, Tagawa H, Karnan S *et al.* Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Res* 2004; **64**: 3087–95.

53 He L, Thomson JM, Hemann MT *et al.* A microRNA polycistron as a potential human oncogene. *Nature* 2005; **435**: 828–33.

54 O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 2005; **435**: 839–43.

55 Takahashi T, Obata Y, Sekido Y *et al.* Expression and amplification of myc gene family in small cell lung cancer and its relation to biological characteristics. *Cancer Res* 1989; **49**: 2683–8.

56 Wang CL, Wang BB, Bartha G *et al.* Activation of an oncogenic microRNA cistron by provirus integration. *Proc Natl Acad Sci USA* 2006; **103**: 18680–4.

57 Uren AG, Kool J, Matentzoglu K *et al.* Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks. *Cell* 2008; **133**: 727–41.

58 Polager S, Ginsberg D. p53 and E2f: partners in life and death. *Nat Rev Cancer* 2009; **9**: 738–48.

59 Sylvestre Y, De Guire V, Querido E *et al.* An E2F/miR-20a autoregulatory feedback loop. *J Biol Chem* 2007; **282**: 2135–43.

60 Woods K, Thomson JM, Hammond SM. Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors. *J Biol Chem* 2007; **282**: 2130–4.

61 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**: 6099–105.

62 Ebi H, Sato T, Sugito N *et al.* Counterbalance between RB inactivation and miR-17-92 overexpression in reactive oxygen species and DNA damage induction in lung cancers. *Oncogene* 2009; **28**: 3371–9.

doi: 10.1111/j.1349-7006.2010.01707.x

63 Pickering MT, Stadler BM, Kowalik TF.miR-17 and miR-20a temper an E2F1-induced G1 checkpoint to regulate cell cycle progression. *Oncogene* 2009; **28**: 140–5.
64 Aguda BD, Kim Y, Piper-Hunter MG, Friedman A, Marsh CB. MicroRNA regulation of a cancer network: consequences of the feedback loops involving miR-17-92, E2F, and Myc. *Proc Natl Acad Sci USA* 2008; **105**: 19678–83.
65 Yu Z, Wang C, Wang M *et al.* A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation. *J Cell Biol* 2008; **182**: 509–17.
66 Taguchi A, Yanagisawa K, Tanaka M *et al.* Identification of hypoxia-inducible factor-1 alpha as a novel target for miR-17-92 microRNA cluster. *Cancer Res* 2008; **68**: 5540–5.
67 Yan HL, Xue G, Mei Q *et al.* Repression of the miR-17-92 cluster by p53 has an important function in hypoxia-induced apoptosis. *EMBO J* 2009; **28**: 2719–32.
68 Xiao C, Srinivasan L, Calado DP *et al.* Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. *Nat Immunol* 2008; **9**: 405–14.
69 Ventura A, Young AG, Winslow MM *et al.* Targeted deletion reveals essential and overlapping functions of the miR-17-92 family of miRNA clusters. *Cell* 2008; **132**: 875–86.
70 Ivanovska I, Ball AS, Diaz RL *et al.* MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression. *Mol Cell Biol* 2008; **28**: 2167–74.
71 Inomata M, Tagawa H, Guo YM *et al.* MicroRNA-17-92 down-regulates expression of distinct targets in different B-cell lymphoma subtypes. *Blood* 2009; **113**: 396–402.
72 Dews M, Homayouni A, Yu D *et al.* Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. *Nat Genet* 2006; **38**: 1060–5.
73 Suarez Y, Fernandez-Hernando C, Yu J *et al.* Dicer-dependent endothelial microRNAs are necessary for postnatal angiogenesis. *Proc Natl Acad Sci USA* 2008; **105**: 14082–7.
74 Bonauer A, Carmona G, Iwasaki M *et al.* MicroRNA-92a controls angiogenesis and functional recovery of ischemic tissues in mice. *Science* 2009; **324**: 1710–13.
75 Ghosh AK, Shanafelt TD, Cimmino A *et al.* Aberrant regulation of pVHL levels by microRNA promotes the HIF/VEGF axis in CLL B cells. *Blood* 2009; **113**: 5568–74.
76 Petrocca F, Visone R, Onelli MR *et al.* E2F1-regulated microRNAs impair TGFbeta-dependent cell-cycle arrest and apoptosis in gastric cancer. *Cancer Cell* 2008; **13**: 272–86.
77 Olive V, Bennett MJ, Walker JC *et al.* miR-19 is a key oncogenic component of mir-17-92. *Genes Dev* 2009; **23**: 2839–49.
78 Mu P, Han YC, Betel D *et al.* Genetic dissection of the miR-17∼92 cluster of microRNAs in Myc-induced B-cell lymphomas. *Genes Dev* 2009; **23**: 2806–11.
79 Harris KS, Zhang Z, McManus MT, Harfe BD, Sun X. Dicer function is essential for lung epithelium morphogenesis. *Proc Natl Acad Sci USA* 2006; **103**: 2208–13.
80 Lu Y, Thomson JM, Wong HY, Hammond SM, Hogan BL. Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells. *Dev Biol* 2007; **310**: 442–53.
81 Carraro G, El-Hashash A, Guidolin D *et al.* miR-17 family of microRNAs controls FGF10-mediated embryonic lung epithelial branching morphogenesis through MAPK14 and STAT3 regulation of E-Cadherin distribution. *Dev Biol* 2009; **333**: 238–50.
82 Lu Y, Okubo T, Rawlins E, Hogan BL. Epithelial progenitor cells of the embryonic lung and the role of microRNAs in their proliferation. *Proc Am Thorac Soc* 2008; **5**: 300–4.
83 Calin GA, Sevignani C, Dumitru CD *et al.* Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci USA* 2004; **101**: 2999–3004.
84 Yamada H, Yanagisawa K, Tokumaru S *et al.* Detailed characterization of a homozygously deleted region corresponding to a candidate tumor suppressor locus at 21q11-21 in human lung cancer. *Genes Chromosomes Cancer* 2008; **47**: 810–18.
85 Brueckner B, Stresemann C, Kuner R *et al.* The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. *Cancer Res* 2007; **67**: 1419–23.
86 Lu L, Katsaros D, de la Longrais IA, Sochirca O, Yu H. Hypermethylation of let-7a-3 in epithelial ovarian cancer is associated with low insulin-like growth factor-II expression and favorable prognosis. *Cancer Res* 2007; **67**: 10117–22.
87 Chang TC, Yu D, Lee YS *et al.* Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat Genet* 2008; **40**: 43–50.
88 Dangi-Garimella S, Yun J, Eves EM *et al.* Raf kinase inhibitory protein suppresses a metastasis signalling cascade involving LIN28 and let-7. *EMBO J* 2009; **28**: 347–58.
89 Iliopoulos D, Hirsch HA, Struhl K. An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. *Cell* 2009; **139**: 693–706.
90 Brock M, Trenkmann M, Gay RE *et al.* Interleukin-6 modulates the expression of the bone morphogenic protein receptor type II through a novel STAT3-microRNA cluster 17/92 pathway. *Circ Res* 2009; **104**: 1184–91.
91 Re A, Cora D, Taverna D, Caselle M. Genome-wide survey of microRNA-transcription factor feed-forward regulatory circuits in human. *Mol Biosyst* 2009; **5**: 854–67.
92 Ender C, Krek A, Friedlander MR *et al.* A human snoRNA with microRNA-like functions. *Mol Cell* 2008; **32**: 519–28.
93 Eiring AM, Harb JG, Neviani P *et al.* miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts. *Cell* 2010; **140**: 652–65.

Published OnlineFirst January 18, 2011; DOI:10.1158/0008-5472.CAN-10-2164

Cancer Research

LATS2 Is a Tumor Suppressor Gene of Malignant Mesothelioma

Hideki Murakami[1], Tetsuya Mizuno[1,4], Tetsuo Taniguchi[1,4], Makiko Fujii[1], Futoshi Ishiguro[1,4], Takayuki Fukui[2], Shinya Akatsuka[6], Yoshitsugu Horio[3], Toyoaki Hida[3], Yutaka Kondo[1], Shinya Toyokuni[6], Hirotaka Osada[1,5], and Yoshitaka Sekido[1,5]

Abstract

Malignant mesothelioma (MM) is an aggressive neoplasm associated with asbestos exposure. We carried out genome-wide array-based comparative genomic hybridization analysis with 14 MM cell lines. Three cell lines showed overlapping homozygous deletion at chromosome 13q12, which harbored the *LATS2* (*large tumor suppressor homolog 2*) gene. With 6 other MM cell lines and 25 MM tumors, we found 10 inactivating homozygous deletions or mutations of *LATS2* among 45 MMs. *LATS2* encodes a serine/threonine kinase, a component of the Hippo tumor-suppressive signaling pathway, and we transduced *LATS2* in MM cells with its mutation. Transduction of *LATS2* inactivated oncoprotein YAP, a transcriptional coactivator, via phosphorylation, and inhibited MM cell growth. We also analyzed LATS2 immunohistochemically and found that 13 of 45 MM tumors had low expression of *LATS2*. Because *NF2* is genetically mutated in 40% to 50% of MM, our data indicate that Hippo pathway dysregulation is frequent in MM cells with inactivation of *LATS2* or an upstream regulator of this pathway, Merlin, which is encoded by *NF2*. Thus, our results suggest that the inactivation of LATS2 is one of the key mechanisms for constitutive activation of YAP, which induces deregulation of MM cell proliferation.

Cancer Res; 71(3); 873–83. ©2011 AACR.

Introduction

Malignant mesothelioma (MM) is an aggressive neoplasm associated with asbestos (1–4). Because MM is usually diagnosed at advanced stages and is largely unresponsive to conventional therapy, the prognosis of patients with MM is very poor (5, 6). MM shows frequent mutation of $p16^{INK4a}$/$p14^{ARF}$ and *NF2* (*neurofibromatosis type 2*) tumor suppressor genes (TSG) and recent comprehensive analyses have identified other candidate cancer-associated genes responsible for MM development, progression, and poor outcome (7–10).

The *NF2* gene, which encodes Merlin, is inactivated in 40% to 50% of MMs (11–13). Transduction of *NF2* into MM cells was shown to inhibit cell proliferation and invasiveness of MM cells (14, 15). Mouse models with *nf2* allele loss have been shown to enhance mesothelioma development after asbestos exposure (16, 17). Mesothelioma also develops with a high incidence in *Nf2;Arf* conditional knockout mice (18). However, it remains unclear whether MM tumors without an *NF2* mutation express functional Merlin or the tumor-suppressive activity of Merlin is inactivated by other mechanisms. In this regard, possible involvement of the increased expression of CPI-17, a regulator of Merlin, or the upregulation of microRNA that might target *NF2* has been suggested (19, 20).

The mammalian Hippo cascade, which was initially identified via genetic studies in *Drosophila*, is one of the possible downstream signaling cascades of Merlin and Expanded (21–25). This pathway controls tissue growth by inhibiting cell proliferation and by promoting apoptosis. The components of this pathway include SAV1 (also called WW45), MST (*Drosophila* Hippo), and LATS family members, which ultimately phosphorylate and inactivate the YAP transcription coactivator. *YAP*, a candidate oncogene, was shown to be amplified in human cancers (26, 27). We previously reported amplification of the chromosomal 11q22 region including *YAP* in a subset of MM specimens and a positive role of YAP in MM cell proliferation (28).

In this study, we carried out array-based comparative genomic hybridization (CGH) and sequencing analyses and found that 10 of 45 MMs had an inactivating homozygous deletion or mutation of *LATS2*. Furthermore, we showed that transduction of *LATS2* induced phosphorylation of YAP and inhibited MM cell growth. Our results suggest that the Merlin-Hippo pathway is frequently inactivated in MM cells and that *LATS2* is a TSG of MM.

Authors' Affiliations: [1]Division of Molecular Oncology, Aichi Cancer Center Research Institute; Departments of [2]Thoracic Surgery and [3]Thoracic Oncology, Aichi Cancer Center Hospital; Departments of [4]Cardio-Thoracic Surgery and [5]Cancer Genetics, Program in Function Construction Medicine, and [6]Pathology and Biological Responses, Nagoya University Graduate School of Medicine, Nagoya, Japan

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Author: Yoshitaka Sekido, Division of Molecular Oncology, Aichi Cancer Center Research Institute, 1-1 Kanokoden, Chikusa-ku, Nagoya, Aichi 464-8681, Japan. Phone: 81-52-764-2983; Fax: 81-52-764-2993. E-mail: ysekido@aichi-cc.jp

doi: 10.1158/0008-5472.CAN-10-2164

Published OnlineFirst January 18, 2011; DOI:10.1158/0008-5472.CAN-10-2164

Materials and Methods

Cell lines and primary specimens of malignant mesothelioma

Fourteen Japanese MPM (malignant pleural mesothelioma) cell lines, including ACC-MESO-1, -4, Y-MESO-8D, -9, -12, -14, -21, -22, -25, -26B, -27, -28, -29, and -30, were established in our laboratory as reported previously and described elsewhere, and the cells at 10 to 15 passages were used for assays (29, 30). Four MPM cell lines, including NCI-H28, NCI-H2052, NCI-H2373, and MSTO-211H, and one immortalized mesothelial cell line, MeT-5A, were purchased from the American Type Culture Collection (ATCC) and cells at 3 to 5 passages were used after receiving from ATCC. NCI-H290 and NCI-H2452 were the kind gifts of Dr. Adi F. Gazdar. All MPM cell lines were cultured in RPMI 1640 medium supplemented with 10% fetal calf serum (FCS) and 1× antibiotic-antimycotic (Invitrogen) at 37°C in a humidified incubator with 5% CO_2. MeT-5A was cultured according to ATCC instructions. MM tissue samples from patients treated at Aichi Cancer Center Hospital, Nagoya University Hospital, Japanese Red Cross Nagoya First Hospital, Toyota Kosei Hospital, and Kasugai City Hospital were obtained according to the Institutional Review Board–approved protocol for each and the written informed consent from each patient. The human mesothelioma tissue array with 19 MM samples was also used (US Biomax Inc.).

Preparation of DNA and RNA

Genomic DNA was extracted using a standard phenol–chloroform method (31). Total RNA was prepared using RNeasy Plus RNA extraction kit (Qiagen K.K.) according to the manufacturer's protocol. Random-primed, first-strand cDNA was synthesized from 3 μg of total RNA, using Superscript II, according to the manufacturer's instructions (Invitrogen).

Oligonucleotide array CGH analysis

All microarrays used were Agilent 244K whole human genome microarrays, with an average distance of 6.4 kb between each probe (array G4411B sourced from the NCBI genome Build 36; Agilent Technologies). Comparison genomic DNA was obtained commercially (Promega) and matched for sex. The methods for labeling, hybridization, and scanning using a G2505B Agilent DNA microarray scanner (Agilent Technologies) were conducted according to the manufacturer's protocol. The scanned TIFF image data were processed with Agilent Feature Extraction software (version 9.5.3.1) by the CGH-v4_95_Feb07 protocol (Agilent Technologies). Extracted data were analyzed with Agilent DNA Analytics 4.0 software (version 4.0.81; Agilent Technologies), and the Aberration Detection Method 2 (ADM-2) algorithm was used to identify contiguous genomic regions that corresponded to chromosomal aberrations. The following parameters were used in this analysis: threshold of ADM-2: 5.0; centralization: ON (threshold: 5.0, bin size: 10); aberration filters: ON (minimum number of probes in region = 2 and minimum absolute average log ratio for region = 1.6 and maximum number of aberrant regions = 10,000 and % penetrance per feature = 0). At a minimum, 2 contiguous suprathreshold probes were required to define a change. To find an obvious homozygous deletion in cell line DNA, aberrant regions with a signal $\log_2$ ratio of less than -1.6 were searched. Genomic positions were based on the UCSC March 2006 human reference sequence (hg18; NCBI build 36.1 reference sequence). The accession number of array CGH analysis data to Gene Expression Omnibus is GSE22237. For tumor tissue DNA, regions of homozygous deletion or one allelic loss of the *LATS2* locus were defined as $\log_2$ ratio < -1.0 or $-1.0 < -0.4$ for at least 3 consecutive probes, respectively.

Mutation analysis

Mutation analysis of all coding exons of the *LATS2* and *SAV1* and *NF2* genes was carried out by direct sequencing after PCR amplification of genomic DNA. The primer sets of *LATS2* are described in Supplementary Materials and Methods. The primer sets of *NF2* were described previously (11, 29), and sequences of the primer sets of *SAV1* are available upon request.

Antibodies and reagents

Rabbit anti-LATS2 antibody (NB200-199) for Western blot analysis was purchased from Novus Biologicals, and mouse anti-YAP (clone 2F12, H00010413-M01) and anti-SAV1 (clone 3B2, H00060485-M02) antibodies were from Abnova. Rabbit anti-LATS2 antibody (ab70565) for immunohistochemistry and rabbit anti-YAP antibody (EP1674Y) were purchased from Abcam, and anti-NF2 (1C4, #9169) and anti-phospho-YAP (S127; #4911) antibodies were from Cell Signaling Technology. Anti-β-actin (clone AC74) and anti-Flag (M2) antibodies were from Sigma, and anti-V5 antibody was from Invitrogen. Rabbit anti-β-catenin (SC-7199) was from Santa Cruz Biotechnology.

Plasmid and lentiviral vector

The cDNA fragments of wild-type or mutant *LATS2* were amplified by PCR, using PrimeSTAR Max DNA polymerase (Takara Bio), and introduced into the pFLAG-CMV2 expression vector (Sigma) with an infusion cloning system (Clontech), thereby fusing these cDNAs with the FLAG sequence. The sequences of all constructs were confirmed. To generate *LATS2*-expressing lentiviral vector, cDNA coding for the human *LATS2* tagged with FLAG was amplified by PCR and cloned in the pLL3.7 lentiviral vector. *NF2* expression vectors were described previously (28). RNA interference vectors to generate lentiviruses that transcribe short hairpin (sh)-RNA were prepared as described previously (32). sh-LATS2-RNA interference vector (sh-LATS2) contained a target sequence of the hairpin loop of LATS2 (5′-GGACCTCACTGCATTAAA-3′). A control shRNA vector for luciferase (Sh-Luc), which contained a target sequence for luciferase (5′-CGTACGCGGAA-TACTTCGA-3′), was described previously (32).

Cell proliferation assays

A total of 1.0×10^4 and 2.0×10^5 cells were seeded onto flat-bottomed 24- and 12-well plates, respectively. Cells were transduced with lentiviral vectors at the multiplicity of infection of 5,