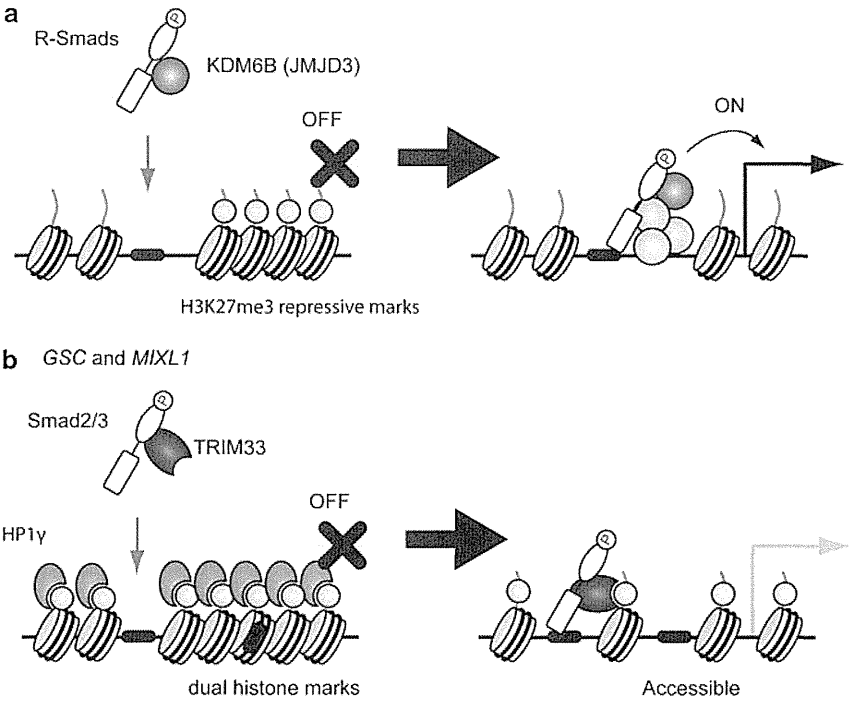

Figure 3. Smad proteins and histone modification marks. Smad proteins have been reported to induce local chromatin remodeling and modification at their binding sites. Several models are described in ES cells, where early developmental genes are poised and ready to be activated in response to extracellular signals, such as nodal. (**a**) R-Smads physically interact with a histone demethylase, KDM6B (also known as JMJD3), and recruit it to their target sites, followed by the loss of the H3K27me3 repressive mark (light green). (**b**) Xi *et al.*[70] reported that nodal signaling triggered TRIM33–Smad2/3 complex formation. The TRIM33–Smad2/3 complex recognizes and binds to H3K9me3-K18ac dual histone marks (light blue) and displaces the chromatin-compacting factor HP1γ (heterochromatin protein 1γ) in the *GSC* and *MIXL1* promoters, resulting in the remodeling of the local chromatin structure to make Smad-binding region(s) (red) accessible. A full colour version of this figure is available at the *Oncogene* journal online.

it possible to study long-distance regulation of genes by enhancers through formation of chromatin loops (reviewed in Simonis *et al.*[76]). Application of these technologies will help to identify the functional relationship between Smad-binding sites and genes implicated in cancer progression.

It is also possible that for many sites, binding of Smads is not sufficient for transcriptional regulation, but additional stimuli are required to drive the expression of the target genes. For example, costimulation with tumor necrosis factor-α, which induces the transcriptional repressor ATF-3, affects the expression regulation of the *Id1* gene and cellular response.[34,35] Sometimes, ligand stimulation itself induces these cofactors and makes a feed-forward circuit, like in myotube differentiation. The myogenic transcription factor MyoD directly regulates genes expressed during skeletal muscle differentiation together with other transcription factors such as MEF2[77] and Zfp238 (also known as RP58).[78] These transcription factors are also induced by MyoD, and MEF2 functions with MyoD in a positive feed-forward circuit,[77] while Zfp238 participates in a negative feed-forward circuit.[78] Comparison of MyoD-binding patterns of mouse C2C12 myoblasts and differentiated myotubes has revealed that most binding events in myoblasts are not directly associated with gene regulation. However, MyoD binding increases during myogenic differentiation at many of the regulatory regions associated with genes expressed in skeletal muscle. Intriguingly, the myotube-increased binding sites are enriched for MEF2-like motifs, while the myotube-decreased peaks are enriched for Zfp238-like motifs,[59] consistent with the fact that MEF2 positively and Zfp238 negatively cooperate with MyoD. It is possible that TGF-β stimulation induces certain transcription factors, which take part in feed-forward regulatory loops and cooperatively regulate gene expression especially at late time points.

IDENTIFICATION OF A TGF-β GENE SIGNATURE

The notion of 'gene signature' comes from the early work on cancer classification and prognosis prediction using genome-wide gene expression profiles obtained from microarray analyses of cancer patients.[79] Identification of a group of genes that reflect the activity of a common function, pathway or other property in a specific context, are sometimes more revealing compared with the analysis of single genes. Gene expression signatures obtained in experimental conditions has proved to subcategorize patients and predict their prognosis. Concerning TGF-β, Coulouarn *et al.*[80] reported that TGF-β-responsive genes at late time points, or a late TGF-β signature, which were identified in mouse primary hepatocytes, successfully discriminate distinct subgroups of hepatocellular carcinoma and possess a predictive value for hepatocellular carcinoma patients.

Combination of ChIP-chip/ChIP-seq and genome-wide transcriptome analyses provides an accurate prediction of target genes of Smad proteins. TGF-β family members regulate a variety of target genes both directly and indirectly, and modulate many biological processes. The chromatin-binding landscape of Smad proteins, obtained by ChIP-chip/ChIP-seq, will help to identify specific genes that are directly regulated by Smad proteins. It will also help to dissect a specific cellular program regulated by TGF-β family members, for example, the growth inhibitory and apoptosis programs of TGF-β. So far, many groups have identified groups of direct TGF-β target genes by using this strategy. Importantly, the TGF-β/Smad4 target gene signature identified in an ovarian cancer cell line predicts patient survival, based on *in silico* mining of publically available patient data bases.[21] Since TGF-β functions as a tumor suppressor in low-grade carcinoma cells, while it promotes metastasis in advanced carcinoma cells, a direct

comparison of the Smad-binding sites of these two stages of tumorigenesis, obtained from experimental models or from cancer patients, may reveal specific gene signatures of TGF-β correlating to its tumor suppressive and tumor-promoting roles, respectively. This may provide us more novel predictive indicators and biomarkers for TGF-β targeting treatments.

CONCLUSIONS AND PERSPECTIVES

The signaling pathways of TGF-β family members are key players in tumorigenesis and cancer progression. TGF-β can function both as a tumor-suppressing and a tumor-promoting factor during cancer progression. BMP signaling has been reported to play critical roles in oncogene-induced senescence, which is part of the tumorigenesis barrier and blocks cellular proliferation by inducing irreversible growth arrest.[66] Interestingly, BMP signaling induces differentiation of certain cancer-initiating cells, such as glioma-initiating cells,[81] while TGF-β/activin signaling maintains their stem cell-like properties.[9,10] Since Smad proteins are central mediators of the signal transduction, studies on global and genome-wide binding sites of Smad proteins may reveal important insights into their complex biological functions.

Identification of an appropriate antibody is the first and most important step for ChIP-chip and ChIP-seq analyses, because the quality of ChIP data depends crucially on the quality of the antibody used.[16] Since MH1 and MH2 domains are conserved among R-Smads, several specific antibodies for Smad proteins recognize their linker region. However, linker regions are targets of posttranslational modification and protein interactions, as discussed above. It is possible that such changes may attenuate the affinities of antibodies under specific conditions. Although ChIP-grade antibodies for Smad proteins have been established (Supplementary Table 2), careful interpretation of the results will be required.

In summary, genome-wide analysis of the binding sites of Smad proteins have led to important discoveries of their cell-type-specific and context-dependent functions. Application of genome-wide techniques to experimental models and human samples derived from cancer patients, will help to clarify their complex mechanisms during cancer progression, and may also provide potential prognostic biomarkers for future cancer therapy.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by a grant from Swedish Cancer Society (Grant number 100452); KAKENHI (grants-in-aid for scientific research on Innovative Area (Integrative Research on Cancer Microenvironment Network; Grant number 22112002)) and for Young Scientists (B) (Grant numbers 22790750 (DK)); the Global Center of Excellence Program (Integrative Life Science Based on the Study of Biosignaling Mechanisms) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan; a Research Grant from the Takeda Science Foundation and a Grant-in-Aid for Cancer Research for the Third-Term Comprehensive 10-Year Strategy for Cancer Control (H22-013) from the Ministry of Health, Labour and Welfare of Japan.

REFERENCES

1 Massague J. TGFbeta in cancer. *Cell* 2008; **134**: 215–230.
2 Ikushima H, Miyazono K. TGFbeta signalling: a complex web in cancer progression. *Nat Rev Cancer* 2010; **10**: 415–424.
3 Feng XH, Derynck R. Specificity and versatility in TGF-beta signaling through Smads. *Annu Rev Cell Dev Biol* 2005; **21**: 659–693.
4 Levy L, Hill CS. Alterations in components of the TGF-beta superfamily signaling pathways in human cancer. *Cytokine Growth Factor Rev* 2006; **17**: 41–58.
5 Yoshimura A, Wakabayashi Y, Mori T. Cellular and molecular basis for the regulation of inflammation by TGF-beta. *J Biochem* 2010; **147**: 781–792.
6 Roberts AB, Wakefield LM. The two faces of transforming growth factor beta in carcinogenesis. *Proc Natl Acad Sci USA* 2003; **100**: 8621–8623.
7 Moustakas A, Heldin CH. Non-Smad TGF-beta signals. *J Cell Sci* 2005; **118**: 3573–3584.
8 Miyazono K, Ehata S, Koinuma D. Tumor-promoting functions of transforming growth factor-beta in progression of cancer. *Ups J Med Sci* 2012; **117**: 143–152.
9 Ikushima H, Todo T, Ino Y, Takahashi M, Miyazawa K, Miyazono K. Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. *Cell Stem Cell* 2009; **5**: 504–514.
10 Penuelas S, Anido J, Prieto-Sanchez RM, Folch G, Barba I, Cuartas I *et al.* TGF-beta increases glioma-initiating cell self-renewal through the induction of LIF in human glioblastoma. *Cancer Cell* 2009; **15**: 315–327.
11 Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 2008; **133**: 704–715.
12 Lonardo E, Hermann PC, Mueller MT, Huber S, Balic A, Miranda-Lorenzo I *et al.* Nodal/Activin signaling drives self-renewal and tumorigenicity of pancreatic cancer stem cells and provides a target for combined drug therapy. *Cell Stem Cell* 2011; **9**: 433–446.
13 Naka K, Hoshii T, Muraguchi T, Tadokoro Y, Ooshio T, Kondo Y *et al.* TGF-beta-FOXO signalling maintains leukaemia-initiating cells in chronic myeloid leukaemia. *Nature* 2010; **463**: 676–680.
14 Heldin CH, Miyazono K, ten Dijke P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* 1997; **390**: 465–471.
15 Shi Y, Massague J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* 2003; **113**: 685–700.
16 Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009; **10**: 669–680.
17 Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. *Nat Methods* 2009; **6**: S22–S32.
18 Mullen AC, Orlando DA, Newman JJ, Loven J, Kumar RM, Bilodeau S *et al.* Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* 2011; **147**: 565–576.
19 Trompouki E, Bowman TV, Lawton LN, Fan ZP, Wu DC, DiBiase A *et al.* Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* 2011; **147**: 577–589.
20 Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 2008; **133**: 1106–1117.
21 Kennedy BA, Deatherage DE, Gu F, Tang B, Chan MW, Nephew KP *et al.* ChIP-seq defined genome-wide map of TGFbeta/SMAD4 targets: implications with clinical outcome of ovarian cancer. *PLoS One* 2011; **6**: e22606.
22 Miyazono K, Kamiya Y, Morikawa M. Bone morphogenetic protein receptors and signal transduction. *J Biochem* 2010; **147**: 35–51.
23 Zhang Y, Feng XH, Smad3 Derynck R. and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. *Nature* 1998; **394**: 909–913.
24 Koinuma D, Tsutsumi S, Kamimura N, Taniguchi H, Miyazawa K, Sunamura M *et al.* Chromatin immunoprecipitation on microarray analysis of Smad2/3 binding sites reveals roles of ETS1 and TFAP2A in transforming growth factor beta signaling. *Mol Cell Biol* 2009; **29**: 172–186.
25 Koinuma D, Tsutsumi S, Kamimura N, Imamura T, Aburatani H, Miyazono K. Promoter-wide analysis of Smad4 binding sites in human epithelial cells. *Cancer Sci* 2009; **100**: 2133–2142.
26 Yoon SJ, Wills AE, Chuong E, Gupta R, HEB Baker JC. and E2A function as SMAD/FOXH1 cofactors. *Genes Dev* 2011; **25**: 1654–1661.
27 Ikushima H, Komuro A, Isogaya K, Shinozaki M, Hellman U, Miyazawa K *et al.* An Id-like molecule, HHM, is a synexpression group-restricted regulator of TGF-beta signalling. *EMBO J* 2008; **27**: 2955–2965.
28 Gomis RR, Alarcon C, Nadal C, Van Poznak C, Massague J. C/EBPbeta at the core of the TGFbeta cytostatic response and its evasion in metastatic breast cancer cells. *Cancer Cell* 2006; **10**: 203–214.
29 Silvestri C, Narimatsu M, von Both I, Liu Y, Tan NB, Izzi L *et al.* Genome-wide identification of Smad/Foxh1 targets reveals a role for Foxh1 in retinoic acid regulation and forebrain development. *Dev Cell* 2008; **14**: 411–423.
30 Kim SW, Yoon SJ, Chuong E, Oyolu C, Wills AE, Gupta R *et al.* Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev Biol* 2011; **357**: 492–504.
31 Gomis RR, Alarcon C, He W, Wang Q, Seoane J, Lash A *et al.* A FoxO-Smad synexpression group in human keratinocytes. *Proc Natl Acad Sci USA* 2006; **103**: 12747–12752.
32 Janknecht R, Wells NJ, Hunter T. TGF-beta-stimulated cooperation of smad proteins with the coactivators CBP/p300. *Genes Dev* 1998; **12**: 2114–2119.
33 Feng XH, Zhang Y, Wu RY, Derynck R. The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for smad3 in TGF-beta-induced transcriptional activation. *Genes Dev* 1998; **12**: 2153–2163.

34 Kang Y, Chen CR, Massague J. A self-enabling TGFbeta response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. *Mol Cell* 2003; **11**: 915–926.
35 Anido J, Saez-Borderias A, Gonzalez-Junca A, Rodon L, Folch G, Carmona MA *et al.* TGF-beta receptor inhibitors target the CD44(high)/Id1(high) glioma-initiating cell population in human glioblastoma. *Cancer Cell* 2010; **18**: 655–668.
36 Alarcon C, Zaromytidou AI, Xi Q, Gao S, Yu J, Fujisawa S *et al.* Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways. *Cell* 2009; **139**: 757–769.
37 Sapkota G, Alarcon C, Spagnoli FM, Brivanlou AH, Massague J. Balancing BMP signaling through integrated inputs into the Smad1 linker. *Mol Cell* 2007; **25**: 441–454.
38 Kretzschmar M, Doody J, Massague J. Opposing BMP and EGF signalling pathways converge on the TGF-beta family mediator Smad1. *Nature* 1997; **389**: 618–622.
39 Aubin J, Davy A, Soriano P. *In vivo* convergence of BMP and MAPK signaling pathways: impact of differential Smad1 phosphorylation on development and homeostasis. *Genes Dev* 2004; **18**: 1482–1494.
40 Fuentealba LC, Eivers E, Ikeda A, Hurtado C, Kuroda H, Pera EM *et al.* Integrating patterning signals: Wnt/GSK3 regulates the duration of the BMP/Smad1 signal. *Cell* 2007; **131**: 980–993.
41 Aragon E, Goerner N, Zaromytidou AI, Xi Q, Escobedo A, Massague J *et al.* A Smad action turnover switch operated by WW domain readers of a phosphoserine code. *Genes Dev* 2011; **25**: 1275–1288.
42 Zhu H, Kavsak P, Abdollah S, Wrana JL, Thomsen GHA. SMAD ubiquitin ligase targets the BMP pathway and affects embryonic pattern formation. *Nature* 1999; **400**: 687–693.
43 Kuratomi G, Komuro A, Goto K, Shinozaki M, Miyazawa K, Miyazono K *et al.* NEDD4-2 (neural precursor cell expressed, developmentally down-regulated 4-2) negatively regulates TGF-beta (transforming growth factor-beta) signalling by inducing ubiquitin-mediated degradation of Smad2 and TGF-beta type I receptor. *Biochem J* 2005; **386**: 461–470.
44 Gao S, Alarcon C, Sapkota G, Rahman S, Chen PY, Goerner N *et al.* Ubiquitin ligase Nedd4L targets activated Smad2/3 to limit TGF-beta signaling. *Mol Cell* 2009; **36**: 457–468.
45 Dupont S, Mamidi A, Cordenonsi M, Montagner M, Zacchigna L, Adorno M *et al.* FAM/USP9x, a deubiquitinating enzyme essential for TGFbeta signaling, controls Smad4 monoubiquitination. *Cell* 2009; **136**: 123–135.
46 Inui M, Manfrin A, Mamidi A, Martello G, Morsut L, Soligo S *et al.* USP15 is a deubiquitylating enzyme for receptor-activated SMADs. *Nat Cell Biol* 2011; **13**: 1368–1375.
47 Lonn P, van der Heide LP, Dahl M, Hellman U, Heldin CH, Moustakas A. PARP-1 attenuates Smad-mediated transcription. *Mol Cell* 2010; **40**: 521–532.
48 Huang D, Wang Y, Wang L, Zhang F, Deng S, Wang R *et al.* Poly(ADP-ribose) polymerase 1 is indispensable for transforming growth factor-beta Induced Smad3 activation in vascular smooth muscle cell. *PLoS One* 2011; **6**: e27123.
49 Zawel L, Dai JL, Buckhaults P, Zhou S, Kinzler KW, Vogelstein B *et al.* Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol Cell* 1998; **1**: 611–617.
50 Yagi K, Goto D, Hamamoto T, Takenoshita S, Kato M, Miyazono K. Alternatively spliced variant of Smad2 lacking exon 3. Comparison with wild-type Smad2 and Smad3. *J Biol Chem* 1999; **274**: 703–709.
51 Chai J, Wu JW, Yan N, Massague J, Pavletich NP, Shi Y. Features of a Smad3 MH1-DNA complex. Roles of water and zinc in DNA binding. *J Biol Chem* 2003; **278**: 20327–20331.
52 Shi Y, Wang YF, Jayaraman L, Yang H, Massague J, Pavletich NP. Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. *Cell* 1998; **94**: 585–594.
53 BabuRajendran N, Palasingam P, Narasimhan K, Sun W, Prabhakar S, Jauch R *et al.* Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF-beta effectors. *Nucleic Acids Res* 2010; **38**: 3477–3488.
54 Lee KL, Lim SK, Orlov YL, Yit le Y, Yang H, Ang LT *et al.* Graded Nodal/Activin signaling titrates conversion of quantitative phospho-Smad2 levels into qualitative embryonic stem cell fate decisions. *PLoS Genet* 2011; **7**: e1002130.
55 Zhang Y, Handley D, Kaplan T, Yu H, Bais AS, Richards T *et al.* High throughput determination of TGFbeta1/SMAD3 targets in A549 lung epithelial cells. *PLoS One* 2011; **6**: e20319.
56 Kim J, Johnson K, Chen HJ, Carroll S, Laughon A. Drosophila Mad binds to DNA and directly mediates activation of vestigial by Decapentaplegic. *Nature* 1997; **388**: 304–308.
57 Morikawa M, Koinuma D, Tsutsumi S, Vasilaki E, Kanki Y, Heldin CH *et al.* ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. *Nucleic Acids Res* 2011; **39**: 8712–8727.
58 Mizutani A, Koinuma D, Tsutsumi S, Kamimura N, Morikawa M, Suzuki HI *et al.* Cell type-specific target selection by combinatorial binding of Smad2/3 proteins and hepatocyte nuclear factor 4alpha in HepG2 cells. *J Biol Chem* 2011; **286**: 29848–29860.
59 Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH *et al.* Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev Cell* 2010; **18**: 662–674.
60 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010; **38**: 576–589.
61 John S, Sabo PJ, Thurman RE, Sung MH, Biddie SC, Johnson TA *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* 2011; **43**: 264–268.
62 Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 2011; **25**: 2227–2241.
63 Harvey SA, Smith JC. Visualisation and quantification of morphogen gradient formation in the zebrafish. *PLoS Biol* 2009; **7**: e1000101.
64 Green JB, New HV, Smith JC. Responses of embryonic Xenopus cells to activin and FGF are separated by multiple dose thresholds and correspond to distinct axes of the mesoderm. *Cell* 1992; **71**: 731–739.
65 Fei T, Xia K, Li Z, Zhou B, Zhu S, Chen H *et al.* Genome-wide mapping of SMAD target genes reveals the role of BMP signaling in embryonic stem cell fate determination. *Genome Res* 2010; **20**: 36–44.
66 Kaneda A, Fujita T, Anai M, Yamamoto S, Nagae G, Morikawa M *et al.* Activation of Bmp2-Smad1 signal and its regulation by coordinated alteration of H3K27 trimethylation in Ras-induced senescence. *PLoS Genet* 2011; **7**: e1002359.
67 Akizu N, Estaras C, Guerrero L, Marti E, Martinez-Balbas MA. H3K27me3 regulates BMP activity in developing spinal cord. *Development* 2010; **137**: 2915–2925.
68 Dahle O, Kumar A, Kuehn MR. Nodal signaling recruits the histone demethylase Jmjd3 to counteract polycomb-mediated repression at target genes. *Sci Signal* 2010; **3**: ra48.
69 He W, Dorn DC, Erdjument-Bromage H, Tempst P, Moore MA, Massague J. Hematopoiesis controlled by distinct TIF1gamma and Smad4 branches of the TGFbeta pathway. *Cell* 2006; **125**: 929–941.
70 Xi Q, Wang Z, Zaromytidou AI, Zhang XH, Chow-Tsang LF, Liu JX *et al.* A poised chromatin platform for TGF-beta access to master regulators. *Cell* 2011; **147**: 1511–1524.
71 Agricola E, Randall RA, Gaarenstroom T, Dupont S, Hill CS. Recruitment of TIF1gamma to chromatin via its PHD finger-bromodomain activates its ubiquitin ligase and transcriptional repressor activities. *Mol Cell* 2011; **43**: 85–96.
72 Dupont S, Zacchigna L, Cordenonsi M, Soligo S, Adorno M, Rugge M *et al.* Germ-layer specification and control of cell growth by Ectodermin, a Smad4 ubiquitin ligase. *Cell* 2005; **121**: 87–99.
73 Farnham PJ. Insights from genomic profiling of transcription factors. *Nat Rev Genet* 2009; **10**: 605–616.
74 Landry JR, Bonadies N, Kinston S, Knezevic K, Wilson NK, Oram SH *et al.* Expression of the leukemia oncogene Lmo2 is controlled by an array of tissue-specific elements dispersed over 100 kb and bound by Tal1/Lmo2, Ets, and Gata factors. *Blood* 2009; **113**: 5783–5792.
75 Pimanda JE, Donaldson IJ, de Bruijn MF, Kinston S, Knezevic K, Huckle L *et al.* The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. *Proc Natl Acad Sci USA* 2007; **104**: 840–845.
76 Simonis M, Kooren J, de Laat W. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 2007; **4**: 895–901.
77 Penn BH, Bergstrom DA, Dilworth FJ, Bengal E, Tapscott SJ. A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. *Genes Dev* 2004; **18**: 2348–2353.
78 Yokoyama S, Ito Y, Ueno-Kudoh H, Shimizu H, Uchibe K, Albini S *et al.* A systems approach reveals that the myogenesis genome network is regulated by the transcriptional repressor RP58. *Dev Cell* 2009; **17**: 836–848.
79 van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; **347**: 1999–2009.
80 Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. *Hepatology* 2008; **47**: 2059–2067.
81 Piccirillo SG, Reynolds BA, Zanetti N, Lamorte G, Binda E, Broggi G *et al.* Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells. *Nature* 2006; **444**: 761–765.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc)

Published online 28 December 2012

Nucleic Acids Research, 2013, Vol. 41, No. 5 ***e62***
doi:10.1093/nar/gks1439

Widespread inference of weighted microRNA-mediated gene regulation in cancer transcriptome analysis

Hiroshi I. Suzuki[1,*], Hajime Mihira[1], Tetsuro Watabe[1,2], Koichi Sugimoto[3,4] and Kohei Miyazono[1,*]

[1]Department of Molecular Pathology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, [2]PRESTO, Japan Science Technology Agency, 4-1-8 Honcho, Kawaguchi-shi, Saitama 332-0012, Japan, [3]Division of Hematology, Department of Internal Medicine, Juntendo University School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan and [4]Department of Hematology and Oncology, JR Tokyo General Hospital, 2-1-3, Yoyogi, Shibuya-ku, Tokyo 151-8528, Japan

Received August 28, 2012; Revised December 5, 2012; Accepted December 10, 2012

ABSTRACT

MicroRNAs (miRNAs) comprise a gene-regulatory network through sequence complementarity with target mRNAs. Previous studies have shown that mammalian miRNAs decrease many target mRNA levels and reduce protein production predominantly by target mRNA destabilization. However, it has not yet been fully assessed whether this scheme is widely applicable to more realistic conditions with multiple miRNA fluctuations. By combining two analytical frameworks for detecting the enrichment of gene sets, Gene Set Enrichment Analysis (GSEA) and Functional Assignment of miRNAs via Enrichment (FAME), we developed GSEA–FAME analysis (GFA), which enables the prediction of miRNA activities from mRNA expression data using rank-based enrichment analysis and weighted evaluation of miRNA–mRNA interactions. This cooperative approach delineated a better widespread correlation between miRNA expression levels and predicted miRNA activities in cancer transcriptomes, thereby providing proof-of-concept of the mRNA-destabilization scenario. In an integrative analysis of The Cancer Genome Atlas (TCGA) multidimensional data including profiles of both mRNA and miRNA, we also showed that GFA-based inference of miRNA activity could be used for the selection of prognostic miRNAs in the development of cancer survival prediction models. This approach proposes a next-generation strategy for the interpretation of miRNA function and identification of target miRNAs as biomarkers and therapeutic targets.

INTRODUCTION

MicroRNAs (miRNAs) are endogenous small non-coding RNAs that play important roles in various cellular functions and biological phenomena. In general, the regulatory mode of miRNA-mediated gene regulation is post-transcriptional gene silencing by miRNA–mRNA interactions on the basis of sequence complementarity between miRNAs and the 3′-untranslated region (3′UTR) of their target mRNAs (1). In animals, the pairing to target mRNAs required for gene silencing is less extensive than in plants, and the seed region (nucleotides 2–7) is important for target recognition (2). Details of this seed-mediated gene repression without Argonaute-catalysed RNA cleavage are less clear. Although a possibility was previously raised that this type of repression was not strictly associated with a reduction in mRNA levels, several microarray-based evaluations showed that mammalian miRNAs decrease many target mRNA levels (3,4). A recent study further revealed that mammalian miRNAs reduce protein production predominantly by destabilization of target mRNA, indicating the 'mRNA-destabilization (suppression)' scenario (5). In accordance with these findings, the accumulation of a few tissue-specific miRNAs has been associated with the suppression of a large number of their target mRNAs during differentiation, suggesting that several tissue-specific miRNAs contribute to the organization of tissue-specific transcript profiles through the suppression of target mRNAs (6).

*To whom correspondence should be addressed. Tel: +81 3 5841 3345; Fax: +81 3 5841 3354; Email: hisuzuki-tky@umin.ac.jp
Correspondence may also be addressed to Kohei Miyazono. Tel: +81 3 5841 3345; Fax: +81 3 5841 3354; Email: miyazono@m.u-tokyo.ac.jp

It has also been reported that genes preferentially expressed in the same tissue and at the same time as such tissue-specific miRNAs have evolutionarily avoided potential target sites matching these miRNAs to preserve their expression (6).

These findings propose attractive possibilities where changes in transcriptomes measured by mRNA arrays or mRNA-seq may reflect the summation of miRNA-mediated gene repression, which may correspond to changes in miRNome. It may be also envisaged that this scheme could be used for the interpretation of miRNA function by analysing transcriptome data. Based on this assumption, some procedures have been developed to infer miRNA activities by assessing microarray expression data. In conjunction with the sequence-based prediction of miRNA targets, the log-likelihood test and hypergeometric test were often used for the inference of changes in miRNA activity in transcriptome data and assessment of the relationship between miRNA target genes and biological function (7–9). Rank-based approaches such as Gene Set Enrichment Analysis (GSEA) and related analysis were also applied to evaluate the enrichment of miRNA target genes (10,11). In addition, Ulitsky *et al.* (12) recently introduced a new permutation-based statistical method, FAME (functional assignment of miRNAs via enrichment), for these purposes.

However, this 'mRNA-destabilization (suppression)' scenario has been mainly based on results with strong experimental perturbation of target miRNA levels, such as ectopic overexpression and depletion by knockdown or gene targeting, and analyses focused on a few miRNAs with tissue-specific expression patterns (4–6). Therefore, it is unclear whether the strength of the 'mRNA-suppression' scenario is sufficient for an RNA expression-based interpretation of miRNA function. In addition, it has not yet been fully assessed whether this concept is widely applicable to more practical conditions with multiple miRNA fluctuations, such as disease pathogenesis conditions that deviate from evolutionarily conserved processes, besides tissue-specific transcriptome formation. In fact, altered miRNA activities have been shown to have a substantial impact on the modification of gene regulatory networks in cancer, but the extent and consequences of their contribution to cancer transcriptomes have not been investigated in detail (13–15).

Here we devised GSEA–FAME analysis (GFA), which enables the prediction of miRNA activity from mRNA expression data, using rank-based enrichment analysis and weighted prediction of miRNA–mRNA interactions, to address these issues. For this purpose, we combined two analytical pipelines, GSEA and FAME, to detect weaker expression changes caused by miRNAs and evaluate variations in the degree of miRNA–mRNA connections. We observed that this cooperative approach delineated a better widespread correlation between miRNA expression levels and predicted miRNA activities in cancer transcriptome analysis, providing proof-of-concept of the 'mRNA-destabilization' scenario. We also discuss the use of this second-generation miRNA activity inference procedure for the identification of target miRNAs as biomarkers and therapeutic targets.

MATERIALS AND METHODS

Outline of GFA

GFA is composed of three steps: (i) division of miRNA target gene sets into two classes by GSEA, (ii) collection of miRNA target genes contributing to the division of miRNA target gene sets into two classes in step 1 and (iii) FAME for the collection of miRNA target genes in step 2.

Gene sets

The Molecular Signature Database (MSigDB, http://www.broadinstitute.org/gsea/) was used with GSEA software as a collection of annotated gene sets. It consists of five types of gene sets: genomic positional gene sets (C1), biologically curated gene sets (C2), motif gene sets (C3), cancer-related computational gene sets (C4) and Gene Ontology (GO) gene sets (C5). C3 motif gene sets include two collections of potential miRNA targets (C3MIR) and transcription factor targets (C3TFT). C3MIR miRNA target gene sets contain potential target genes sharing a 3′UTR miRNA-binding motif, and C3TFT transcription factor target gene sets contain genes sharing a transcription factor-binding site around the transcription start site. C4 computational gene sets contain C4 cancer module (C4-CM) gene sets, which representatively change in a variety of cancer conditions (16). We performed GSEA using these collections of gene sets.

Gene Set Enrichment Analysis (GSEA)

In step 1, GSEA was performed using C3MIR gene sets or other gene sets to evaluate the enrichment of each gene set in group A or group B. GSEA was performed with GSEA software available from the Broad Institute (http://www.broadinstitute.org/gsea/) using default parameters (17). In step 2, leading-edge subsets were collected for each gene set and assembled for groups A and B, respectively. To reduce the influence of changes in transcriptional factor activity, similar procedures as steps 1 and 2 were performed using C3TFT gene sets that contain genes sharing a transcription factor-binding site, and this C3TFT target gene collection was subtracted from the C3MIR target gene collection in step 2 (option: step 2').

Functional Assignment of miRNAs via Enrichment (FAME)

In step 3, the collection of leading-edge subsets in C3MIR target gene sets in step 2 was analysed by FAME according to a previous report (12). FAME was performed in the Expander 6.0 microarray data analysis suite (http://acgt.cs.tau.ac.il/expander/) (18). FAME executes a permutation-based statistical test to evaluate significant over- or under-representation of miRNA targets in a target gene set, using TargetScan miRNA target predictions (12). We ran FAME for the collection of leading-edge subsets assembled for groups A and B in both directions of enrichment, i.e. over-representation (enrichment) and under-representation (depletion). For *P*-value calculations, 10 000 random iterations were performed throughout this study.

Data sets

Microarray data for miR-1 and miR-124 transfection experiments (GSE2075) and miRNA and mRNA expression profiling of DLBCL patient samples (GSE21849) were obtained from NCBI's Gene Expression Omnibus (3,19). miRNA expression, mRNA expression and clinical data for glioblastoma patient samples were downloaded from The Cancer Genome Atlas (TCGA) Data Portal (https://tcga-data.nci.nih.gov/tcga/) (20). For analysis of miRNA profiling, we used level 3 pre-interpreted data provided by TCGA.

Rank–rank hypergeometric overlap

Rank–rank hypergeometric overlap (RRHO) analysis was performed according to a previous report (21).

Survival prediction model

Level 3-processed data of paired miRNA and mRNA expression profiling for 478 glioblastoma patients were downloaded together with clinical data from TCGA Data Portal in March 2012 (20). After removal of viral miRNAs, profiling data for the remaining 470 miRNAs were mean centred, and the standard deviation was normalized to one per array. The 478 samples were randomly assigned to a training set ($n = 239$) or a testing set ($n = 239$). For all miRNAs, regression coefficients and P-values were estimated using a univariate Cox regression model in a training set. We also performed simple GFA for a training set by dividing samples into a poor prognosis group (death < 1 year) and a good prognosis group (others) using C3MIR gene sets. In this analysis, we adopted GFA results using the larger C3MIR target gene collection among the two C3MIR target gene collections ('poor < good' and 'good < poor'). In the 'expression level/GFA-based' strategy, miRNA showing $P < 0.1$ in univariate Cox regression analysis and $P < 0.12$ in the same direction (poor or good prognostic) in GFA were selected as prognostic miRNAs. In the 'expression level-based' strategy, the same number of miRNAs as miRNAs selected in the 'expression level/GFA-based' strategy was selected according to the order of P-values in univariate Cox regression analysis. Risk scores were evaluated by a linear combination of the expression levels of prognostic miRNAs weighted by their respective Cox regression coefficients according to a previous report (22) by the following formula: 'risk score' = Σ (regression coefficient) × (expression value of each prognostic miRNA). Risk scores were calculated for patients in both training and test sets, and patients with a risk score greater than 0 and those with a risk score less than 0 were assigned to 'High risk' and 'Low risk' groups, respectively. Survival analysis and validation of the fitness and accuracy of these survival prediction models were performed using the survival, rms, and survAUC packages of R. We repeated these analyses for the pair of a training set and test set after five randomizations.

RESULTS

GSEA–FAME analysis (GFA)

We first made two assumptions for building a procedure to infer miRNA activities from mRNA expression data. First, the correlation between fluctuations in multiple endogenous miRNA levels and alterations in target mRNAs may be weak, even if the 'mRNA-destabilization' scenario is operational, and may be better recognized by examining target genes as a population of target genes, not individual genes. Second, the efficacy of individual miRNA target sites was influenced by multiple features of site context, including AU-rich nucleotide composition near the site, relative distance from the stop codon and positioning near the ends of long UTRs (4). Thus, the strength of correlations between a miRNA and its target genes is highly variable and should be weighted to the final assessment of each miRNA activity.

We combined two analytical approaches to satisfy these two assumptions (Figure 1). To detect a weak connection between miRNA target genes and phenotypes in the first assumption, we took a 'gene set versus ranked list' approach through the GSEA algorithm, which is able to detect a weak one-sided inclination among two phenotypes (groups A and B) (17). We performed this analysis using the complete collection of miRNA target gene sets (C3MIR) implemented in GSEA, which contains potential target genes sharing a 3′UTR miRNA-binding motif, and roughly divided these gene sets into two groups that are enriched in group A or in group B (step 1). Although GSEA itself can rank these miRNA target gene sets according to each enrichment score, we used GSEA for the extraction of the core of miRNA target genes contributing to differential miRNA activities between two phenotypes. To this end, we made a collection of leading-edge gene subsets of individual gene sets accounting for GSEA enrichment signals to reconsider the second assumption (step 2).

FAME, a new permutation-based statistical method, was recently developed to test for over- or under-representation of miRNA targets in a designated gene set (12). In contrast to previous methods and conventional statistical tests, FAME uses weighted prediction for miRNA-target pairs according to 3′UTR contexts (i.e., context scores), provided by TargetScan (2,4), and evaluates the significance of the total weight of miRNA-target pairs for each miRNA (12). We applied this FAME algorithm to the collection of leading-edge gene subsets of miRNA target gene sets to integrate the second assumption in relative estimation of various miRNA activities (step 3). We performed FAME in both directions of enrichment, i.e. over-representation (enrichment) and under-representation (depletion). In addition, we also considered an optional procedure to reduce the potential influence of changes in transcriptional factor activity on the deregulation of genes modulated by both miRNAs and transcriptional factors. For this purpose, we performed similar procedures in steps 1 and 2 using transcription factor target gene sets (C3TFT), subtracted this gene collection from the C3MIR target gene collection (step 2') and then ran FAME analysis. We call this approach 'GFA'.

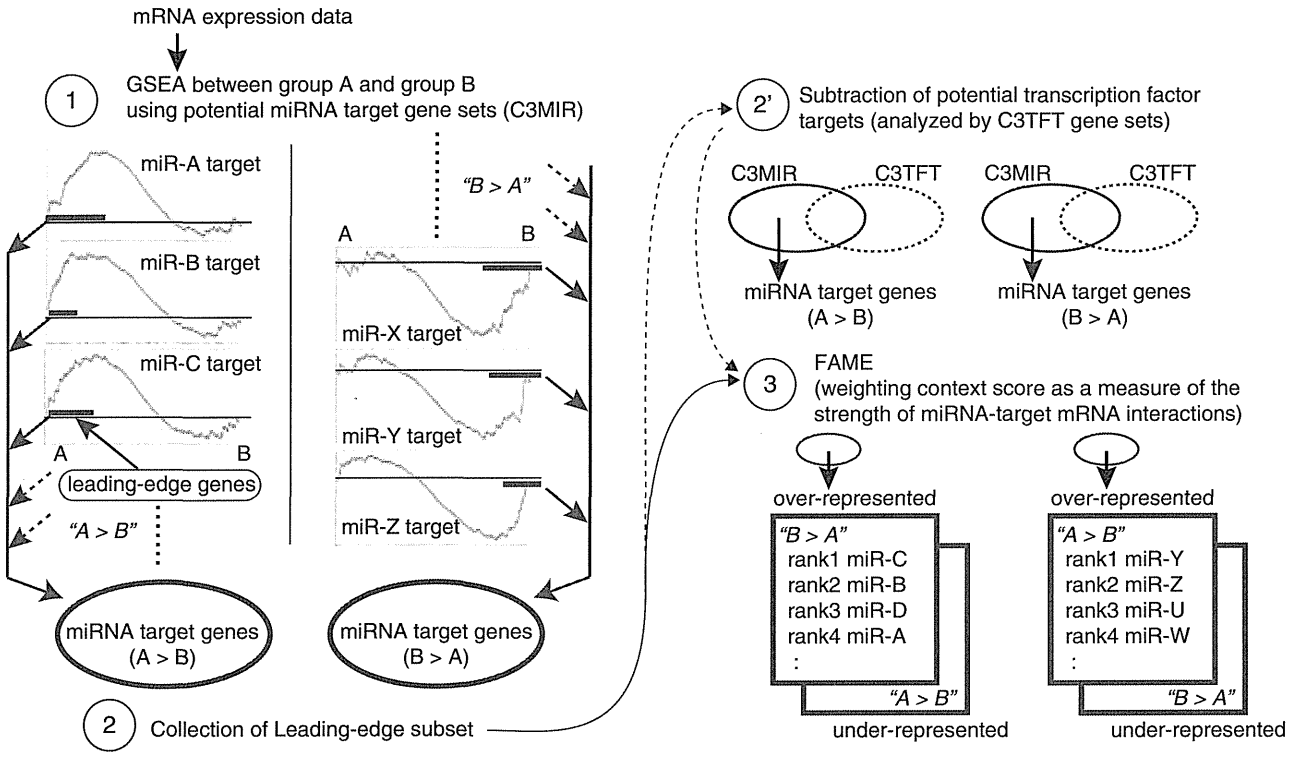

Figure 1. Outline of GFA. In step 1, GSEA is performed using C3 miRNA target gene sets (C3MIR), which contain potential target genes sharing a 3′UTR miRNA-binding motif for each miRNA, to assess whether each miRNA target gene set is enriched in group A or in group B. In step 2, leading-edge subsets, part of the members of each miRNA target gene set, which accounts for the enrichment of corresponding gene sets in group A or B in GSEA analysis, are collected for each gene set and assembled for groups A and B, respectively, to make the collection of overall miRNA target genes enriched in group A and B. As an option, similar procedures to steps 1 and 2 are performed using C3 transcription factor target gene sets (C3TFT) that contain genes sharing a transcription factor-binding site and the C3TFT target gene collection is subtracted from the C3MIR target gene collection subjected to FAME (step 2'). In step 3, FAME is applied to the C3MIR target gene collection in step 2, resulting in a ranked list of each miRNA activity and corresponding target genes accounting for this activity.

Detection of the 'mRNA-destabilization' scenario under experimental miRNA perturbation by GFA

As the first application of our approach, miRNA transfection data that provides a basis for the 'mRNA-suppression' scenario was analysed by GSEA and GFA (Figure 2). In this data, HeLa cells were transfected with two miRNAs (miR-1 and miR-124), two mutant miRNAs (124mut5-6 and 124mut9-10) and two chimeric miRNAs (chimiR-124/1 and chimiR-1/124) and were subjected to microarray analysis 12 and 24 h after transfection (3). As shown in Figure 2A, GSEA ranked four wild-type-seed miRNAs (miR-1, miR-124, 124mut9-10 and chimiR-1/124) as top ones among all miRNAs and confirmed a functional compromise by seed mutation (124mut5-6). On the other hand, GSEA failed to detect chimiR-124/1 activity in 12-h data, although this miRNA has been shown to function in a similar manner to miR-124 albeit to a lesser extent (3,11). GFA demonstrated similar performance with GSEA and further succeeded in the detection of chimiR-124/1 activity in 12-h data, suggesting that FAME contributes to the better ranking of miRNA activities (Figure 2B). In addition, we observed that GFA showed detection ability for transfected miRNAs in two analytical directions of enrichment, over-representation (enrichment) assessment for down-regulated genes relative to controls and under-representation (depletion) assessment for up-regulated genes. In Supplementary Figure S1, the procedure reducing the influences of transcriptional factors (Figure 1, step 2') worsened GFA performance in under-representation assessment for up-regulated genes: it may have been because this experiment is a simple design with miRNA transfection. These results suggest that GFA is able to detect the 'mRNA-destabilization' scenario under experimental miRNA perturbation as well as other previous procedures to infer miRNA activities.

Widespread correlation of miRNA expression levels and miRNA activities in cancer transcriptomes: DLBCL study

The 'mRNA-destabilization (suppression)' scenario has been previously discussed in the setting of strong experimental perturbation of target miRNA levels and tissue-specific transcriptome organization. Therefore, the next important question is whether the 'mRNA-destabilization' scenario is widely applicable to more realistic settings with multiple miRNA fluctuations such as disease pathogenesis conditions. A microarray is a powerful widely used platform to study genome-wide gene expression and gene regulatory networks. In the field of cancer research, this technology has provided numerous advances in the understanding of disease pathogenesis, disease

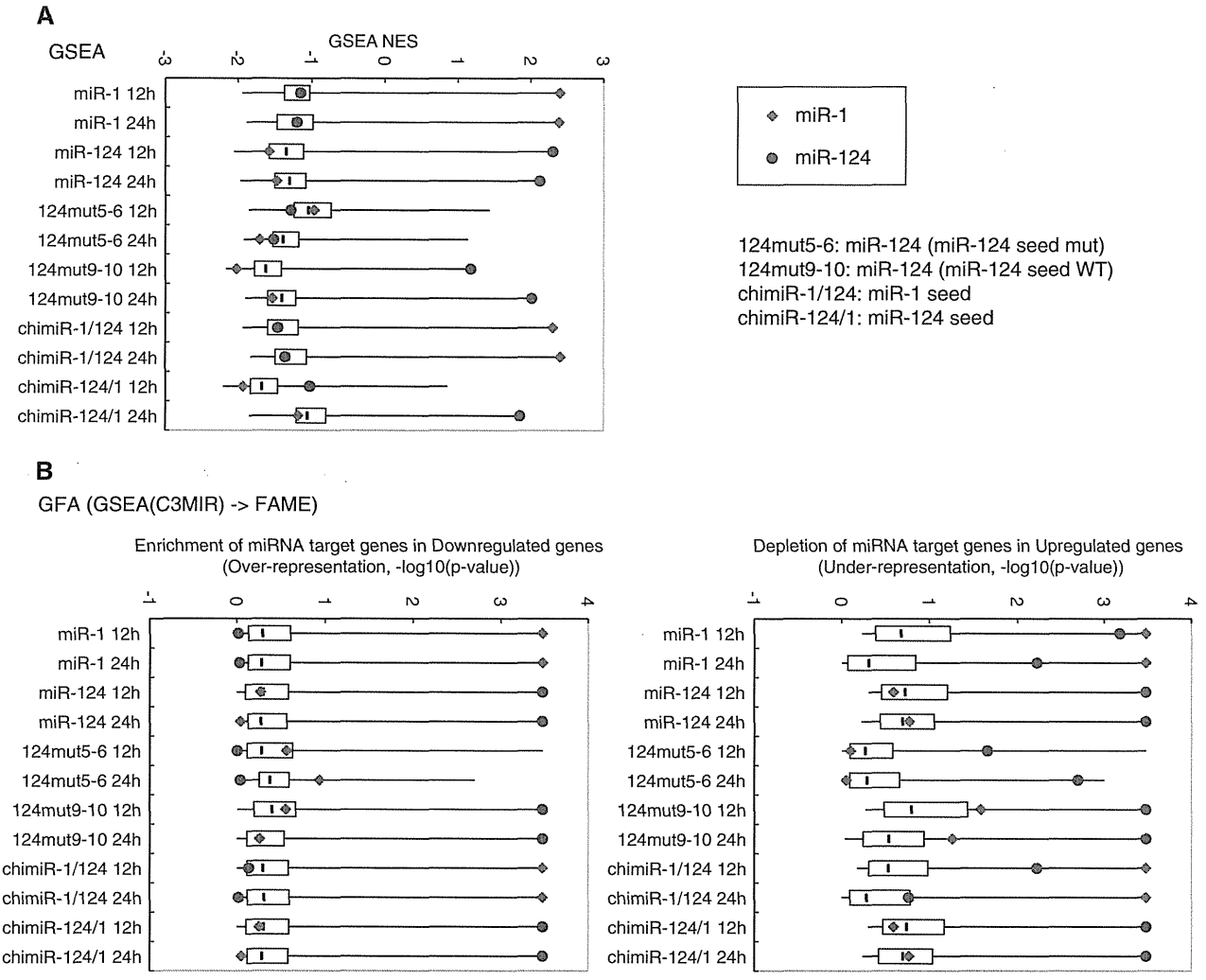

Figure 2. Comparison of GSEA and GFA for miR-1 and miR-124 transfection data. GSEA (**A**) and GFA (**B**) were performed for the microarray data of HeLa cells transfected with miR-1, miR-124, mutant miR-124 (124mut5-6 and 124mut9-10) and chimeric miRNAs (chimiR-1/124 and chimiR-124/1) for 12 or 24 h. In GFA, we ran FAME for the two collection of leading-edge subsets, which were down-regulated (left panel) and up-regulated (right panel) in the cells transfected with miRNA relative to controls, in an enrichment direction of over-representation (enrichment, left panel) and under-representation (depletion, right panel), respectively. The distribution of GSEA normalized enrichment scores (NES) and GFA ranking [−log10(*P*-value)] are shown. GSEA NES and GFA ranking for miR-1 and miR-124 are indicated by red diamonds and blue circles, respectively.

classification and development of survival prediction models (23). As an early example, microarray analysis showed that diffuse large B cell lymphoma (DLBCL), the most common form of human malignant lymphoma, can be divided into two major molecular subtypes: germinal centre (GC) B cell-like DLBCL (GC-DLBCL) and activated B cell-like DLBCL (ABC-DLBCL), according to the resemblance of the gene expression pattern to normal GC B cells or activated B cells, respectively (24). This study exhibited the usefulness of this method in detecting the gene deregulations responsible for biological heterogeneity (24). Using DLBCL as a disease model, we next examined whether the impact of endogenous miRNAs on target mRNA levels in cancer transcriptomes can be deduced by GFA.

We used matched miRNA and mRNA expression data of human clinical DLBCL samples comprising 11 GC subtypes and 18 ABC subtypes (19) (Figure 3A), and performed GSEA and GFA between GC and ABC subtypes. In addition, we compared the performance of the conventional FAME procedure in combination with a standard statistical test (*t*-test, $P < 0.05$) to extract differentially expressed genes between the two subtypes. As shown in Figure 3B, GFA strikingly yielded better evidence of a widespread correlation of miRNA expression levels and miRNA activities for differentially expressed miRNAs, compared with GSEA and the combination of the *t*-test and FAME. Evaluation with Spearman's rank correlation coefficients revealed the outperformance of GFA over the simple use of GSEA and FAME (Figure 3C). Comparisons of the similarity between miRNA expression ranks and miRNA activity ranks using the RRHO method (21) also demonstrated the better performance of GFA (Figure 3D). In addition, we found that the option

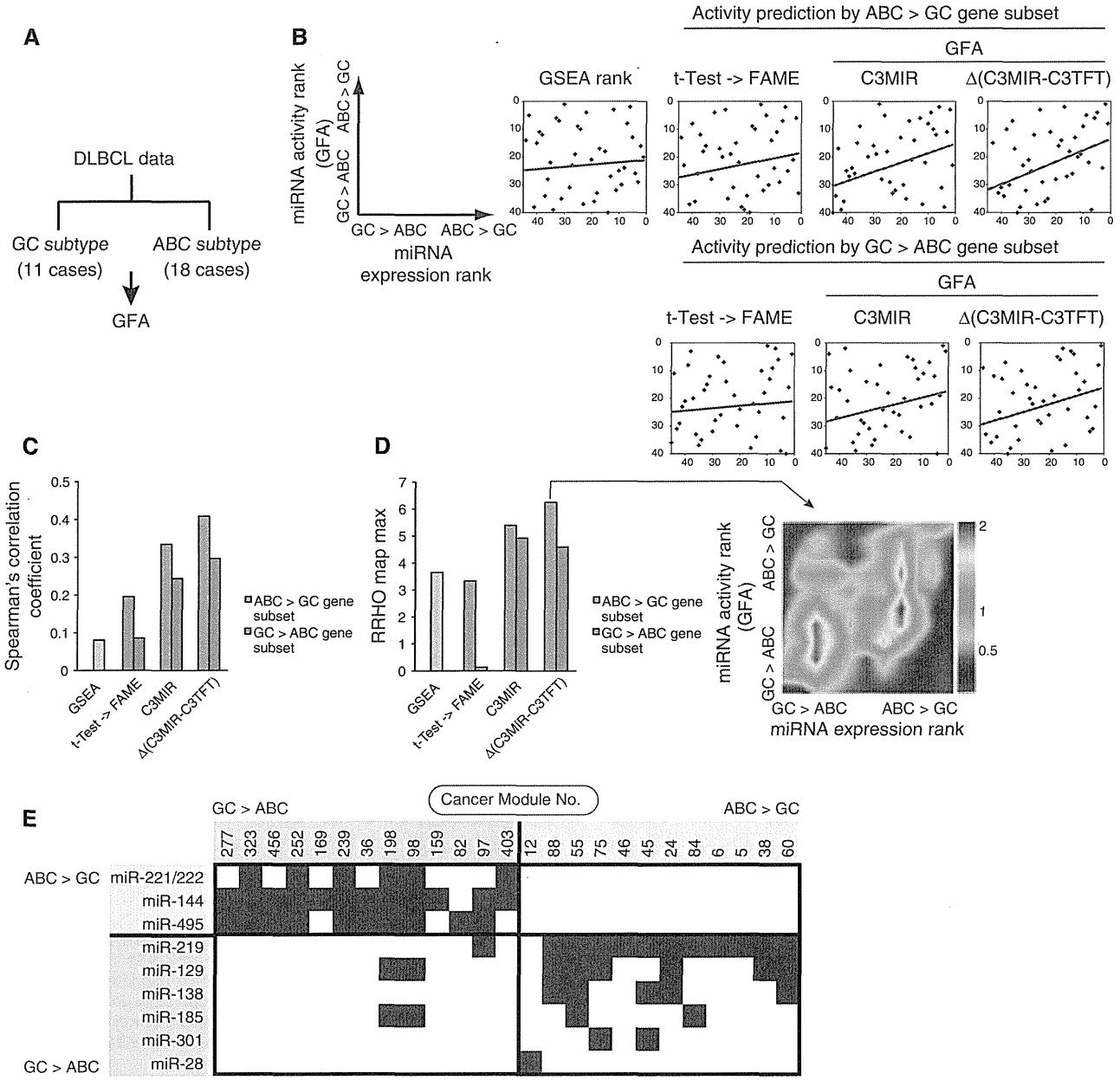

Figure 3. Widespread correlation of miRNA expression levels and miRNA activities assessed by GFA: DLBCL study. (**A**) Classification of DLBCL cases into ABC and GC subtypes. (**B**) Correlation between miRNA expression ranks and miRNA activity ranks analysed by GSEA, FAME [after the extraction of differentially expressed genes by the *t*-test ($P < 0.05$)] and GFA (using C3MIR and C3TFT gene sets) for the top 45 miRNAs with differential expression between ABC and GC subtypes. Differential miRNA expression was evaluated by $-\log10(P\text{-value})$. (**C**) Spearman's rank correlation coefficients for the correlation between miRNA expression ranks and miRNA activity ranks in (B). (**D**) RRHO analysis for the correlation between miRNA expression ranks and miRNA activity ranks in (B). Maximums of the Benjamini–Yekutieli-corrected RRHO map (left) and the representative RRHO heatmap (right) are shown. (**E**) Overlap between target genes of GFA-supported differentially expressed miRNAs and gene subsets of C4 cancer module (C4-CM) gene sets analysed by GSEA. In addition to GFA, GSEA was performed among ABC and GC subtypes using C4-CM gene sets. Red boxes represent overlap between target genes of GFA-supported differentially expressed miRNAs and leading-edge subsets of C4-CM gene sets enriched in either DLBCL subtypes.

process (Figure 1, step 2') further improved the fit of GFA to the correlation relative to results using C3MIR gene sets in this DLBCL study, while this process could sometimes decrease a large proportion of the collection of leading-edge gene subsets in the current form. This result suggests that the effort to consider the coexisting influences of transcriptional factor changes is substantially useful for a better assessment of miRNA activities in some cases. In an initial analysis, we also performed GFA using other gene sets (C1: genomic positional gene sets, C2: biologically curated gene sets, C4: cancer-related computational gene sets and C5: GO gene sets) and confirmed the general superiority of using C3MIR gene sets (Supplementary Figure S2A–S2C). Furthermore, GFA detected a better widespread correlation for all miRNAs (Supplementary Figure S2D). Collectively, these results

suggest that the 'mRNA-suppression' scenario could be taken into account for the interpretation of miRNA function in diverse conditions with physiological fluctuations in multiple miRNAs, and that GFA is a better method for detecting this mode of gene regulation events.

From a practical standpoint, GFA produces a list of the more likely target genes for each miRNA and thus presents a convenient platform for further experimental validation of miRNA–mRNA interactions. In the DLBCL study, we performed GSEA using C4 cancer module (C4-CM) gene sets, which representatively change in a variety of cancer conditions (16), and observed that the above-mentioned gene list for individual miRNAs overlapped with part of the deregulated C4-CM gene sets in a mutually exclusive manner (Figure 3E), suggesting that these miRNAs are potentially responsible for the deregulation of these modules. These analyses suggested the presence of several potential miRNA–mRNA interactions such as miR-144-BCL6, miR-219-TGFBR2, miR-219-PDGFRA, miR-138-EIF4BP1 and miR-144/223/495-PDE4D in association with DLBCL-related genes (Supplementary Table S1). Among them, miR-219-PDGFRA was previously validated through an experimental evaluation (25).

Integration of GFA-based functional assessment into the development of a cancer survival prediction model using miRNA signature: glioblastoma study

Using matched miRNA and mRNA profiling data, one may infer that GFA can be used for better identifying miRNAs, mRNAs and miRNA–mRNA pathways as biomarkers and therapeutic targets in various disease conditions beyond the use of one-sided data and simple application of statistical procedures. Although microarray analysis has indeed enabled the development of a gene signature-based disease classification and stratification strategy especially in the cancer field, this type of strategy frequently suffers from several problems, such as multicollinearity and overfitting. In the latter case, pursuit of better performance in the training set may actually lead to worse performance in the test set and future cases (23). On the basis of these considerations, we finally investigated the usefulness of the GFA approach as a method of feature selection in the construction of a survival prediction model for cancer patients.

In this study, we used matched miRNA and mRNA expression data of glioblastoma patients derived from TCGA database, a comprehensive collection of genomic and expression profiling of various cancer patients, together with clinical data (20) (Figure 4A). We randomly divided 478 glioblastoma samples into a training set and a test set. We subjected the miRNA expression data in the training set to univariate Cox proportional hazard regression analysis to identify miRNAs whose expression status was significantly correlated with patient survival. We also performed GFA in the training data set by dividing samples into a poor prognosis group (death < 1 year) and a good prognosis group (others) to identify miRNAs whose activity status was correlated with patient prognosis. In a combinational strategy using miRNA expression information and GFA-based assessment of miRNA activity (the 'expression level/GFA-based' strategy), we defined miRNAs that showed a correlation with patient prognosis both in Cox regression analysis and in GFA as prognostic miRNAs, and calculated risk scores using a linear combination of these miRNAs (22) (Figure 4A). On the other hand, in the 'expression level-based' strategy, we simply used the results of Cox regression analysis and extracted the same number of miRNAs as miRNAs selected in the 'expression level/GFA-based' strategy.

As a result, we repeated randomization and subsequent analyses several times and observed that the 'expression level-based' strategy produces probable overfitting models that function poorly in test sets, although these models function well in training sets (Figure 4B). In contrast, models developed by the 'expression level/GFA-based' strategy showed a tendency to work better than the former in test sets (Figure 4B). We compared the likelihood ratio of both models and noticed that models in the 'expression level-based' strategy showed a strong decline in a fit to the survival status between training sets and test sets, and that models in the 'expression level/GFA-based' strategy retained a fit even in test sets (Figure 4C). We also estimated the time-dependent prediction error rate of both models and found that the 'expression level/GFA-based' strategy exhibited better performance in spite of the selection of miRNAs with lower significance in Cox analysis than that of the simple 'expression level-based' strategy (Figure 4D). Taken together, these results suggest that GFA-based assessment of miRNA function can provide an attractive option for the more rational design of an identification framework for biomarkers and therapeutic targets using large expression data sets, although further optimization should be required for practical use in combination with unsupervised or supervised procedures and various learning methods.

DISCUSSION

In the present study, we devised a novel procedure GFA to infer miRNA activities from mRNA expression data, taking advantage of two analytical pipelines, GSEA and FAME. Despite advances in sequence-based approaches for miRNA target predictions, it is still a major challenge to deduce miRNA function from the list of hundreds of putative target genes. So far, certain *in silico* methods including FAME and the recent CoMeTa method, which uses coexpression patterns of target genes for each miRNA, have been developed as the next approach for this problem (11,12,26). These methods are based on the concept that miRNAs must stamp their footprints on overall mRNA expression in accordance with the 'mRNA-suppression' scenario. However, it has not been fully assessed whether the scenario is widely applicable to diverse conditions with multiple miRNA fluctuations and whether endogenous miRNA-mediated mRNA repression has enough strength to predict miRNA function and activity from mRNA expression data. By examining this issue, we showed here that the 'mRNA-suppression'

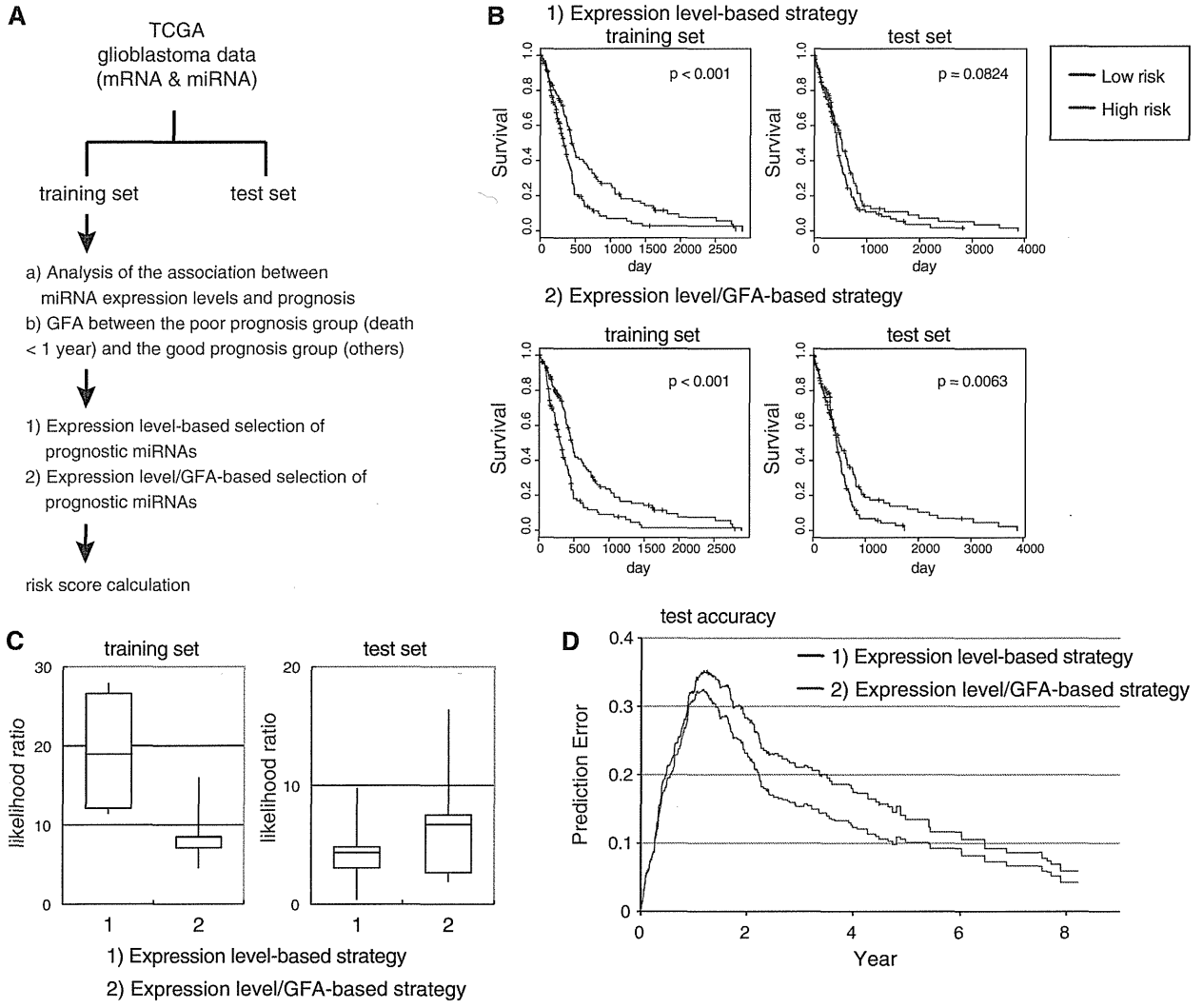

Figure 4. Availability of GFA for the selection of prognostic miRNAs in the development of cancer survival prediction models: TCGA glioblastoma study. (**A**) Analytical outline of survival prediction using TCGA glioblastoma data sets. (**B**) Examples of Kaplan–Meier plots representing survival probabilities according to low or high levels of risk scores developed by (i) the 'expression level-based' strategy and (ii) the 'expression level/GFA-based' strategy in a training set and a test set. (**C**) Likelihood ratio of survival prediction models developed by (i) the 'expression level-based' strategy and (ii) the 'expression level/GFA-based' strategy in training sets and test sets. Positive numbers indicate that the model fits the data better. Results with five randomizations are shown. (**D**) Time-dependent prediction errors of survival prediction models developed by (i) the 'expression level-based' strategy and (ii) the 'expression level/GFA-based' strategy in training sets and test sets. Average results with five randomizations are shown.

scenario is certainly realized as a widespread correlation between miRNA expression levels and miRNA activities predicted by mRNA expression in cancer transcriptome analysis. Therefore, these results provided proof-of-concept of the 'mRNA-destabilization' scenario. In addition, our findings may encourage an RNA expression-based approach to infer miRNA function for comprehensive understanding of the miRNA–mRNA network in diverse fields.

FAME can be used in combination with several clustering methods and standard statistical tests including the *t*-test to extract differentially expressed genes (12). Although it was previously shown that this type of usage of FAME gives a better correlation between miRNA expression and miRNA activity than the hypergeometric test in the analysis of cell type-specific formation of transcriptome profiles (12), we have developed a GFA with a better and steadfast performance than that of the combination with the *t*-test and FAME in the DLBCL study (Figure 3). We postulate that the better performance of GFA is attained by the merits of GSEA and FAME: the ability of GSEA to detect weaker expression changes and the characteristics of FAME using a weighting scheme for miRNA–mRNA pairs. In addition, the use of GSEA confers a threshold-free approach, thereby providing practical convenience. This may also reduce the chance of missing weak, but biologically relevant, expression changes due to choosing a threshold that is either too stringent or too lenient. Furthermore, we observed that GFA was informative in the detection of activities of transfected miRNAs not only in over-representation (enrichment) assessment for down-regulated genes relative to controls, but also in under-representation (depletion) assessment for up-regulated genes (Figure 2). This suggests

that GFA can aid the detection of relative low activities of miRNAs in addition to the detection of active miRNAs. Regarding that FAME is informative in both analytical directions, over-representation (enrichment) assessment and under-representation (depletion) assessment (12,18), FAME may contribute to the detection of relative low activities of transfected miRNAs in control samples in this case. Under-representation evaluations may be also more informative than over-representation evaluations in certain conditions where a group of genes have evolved to avoid miRNA targeting during development (6,18).

On the other hand, we observed that some differentially expressed miRNAs did not appear to mark their effects on mRNA expression data. It may be interpreted by several modification mechanisms such as a mask of miRNA function by other RNA binding protein(s), non-canonical seed-independent gene regulation and targeting of non-3′UTR regions by miRNAs (27,28). Therefore, further mechanistic understanding of miRNA-mediated gene regulation may improve the ability of first-generation miRNA target prediction procedures (i.e. TargetScan and so on) and second-generation miRNA activity inference procedures (i.e. GFA and so on).

Transcription factors are key regulators of gene expression as well as miRNAs. Recent progress in the ENCODE project demonstrated that most transcription factors involved in miRNA regulation tend to be enriched at the top of the network hierarchy of transcription factors and to either largely regulate miRNAs or be regulated by miRNAs (29). A similar pattern can also be seen for miRNAs, suggesting the presence of a few high-degree connections between transcription factors and miRNAs, either transcription factor-to-miRNA regulatory interactions or miRNA-to-transcription factor regulatory interactions, with balanced regulation at the top of the gene regulation network (29). On the other hand, various modes of relationships between transcription factors and miRNAs converging on the same target genes have been considered (30,31). In the differentiation of embryonic stem cells, it has been shown that let-7 miRNAs modulate the transcriptional network through dual effects on transcription factors: direct inhibition of Myc activity and collateral suppression of miRNA target genes regulated overlappingly by the pluripotency transcription factors Oct4, Sox2, Nanog and Tcf3 (32). To consider the potential confounding mechanisms seen in the latter case, we prepared an optional procedure to reduce the potential influence of changes in transcriptional factor activity (Figure 1, step 2') and observed that this procedure may gain in the detection of the 'mRNA-suppression' scenario in the DLBCL study (Figure 3). Further improvements in processing relationships between miRNAs and transcription factors may provide more valuable insights into understanding the impact of miRNAs on gene regulation.

In addition, our study presents some implications of GFA for the field of conventional microarray-based approaches in biomarker and therapeutic target identification. We showed that inference of miRNA activity by GFA could be used for the selection of prognostic miRNAs in the development of cancer survival prediction models (Figure 4). Considering the presence of an extensive interplay between miRNA layers and transcriptome layers, the paired profiling of miRNA expression and mRNA expression can provide a robust platform to select biologically relevant features, together with an analytical framework for understanding RNA regulatory networks, such as GFA. This approach may allow us to avoid problems associated with the use of one-sided data and the simple application of statistical procedures, including overfitting (23). This beneficial effect may also be obtained in the mRNA signature-based strategy, in which mRNA expression data could be affected more by a 'large P small N' problem. Furthermore, GFA-supported miRNAs may be good candidates for therapeutic targeting to correct the pathological deregulation of gene expression and modify disease processes. In conclusion, this analytical approach may offer a basis for a next-generation strategy to interpret miRNA function and identify target miRNAs as biomarkers and therapeutic targets.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1 and 2.

ACKNOWLEDGEMENTS

We thank Matsuyama H for valuable discussions and all members of Department of Molecular Pathology, the University of Tokyo.

FUNDING

KAKENHI [Grant-in-Aid for Young Scientists (A) (No. 24689018) and for Scientific Research on Innovative Areas 'RNA regulation' (No. 23112702) and 'Integrative research on cancer microenvironment network' (No. 22112002)]; Global Center of Excellence Program for 'Integrative Life Science Based on the Study of Biosignaling Mechanisms' from the Ministry of Education, Culture, Sports, Science, and Technology of Japan; Cell Science Research Foundation. Funding for open access charge: Grant-in-Aid for Scientific Research on Innovative Areas 'Integrative research on cancer microenvironment network' (No. 22112002).

Conflict of interest statement. None declared.

REFERENCES

1. Ambros,V. and Chen,X. (2007) The regulation of genes and genomes by small RNAs. *Development*, **134**, 1635–1641.
2. Lewis,B.P., Burge,C.B. and Bartel,D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
3. Lim,L.P., Lau,N.C., Garrett-Engele,P., Grimson,A., Schelter,J.M., Castle,J., Bartel,D.P., Linsley,P.S. and Johnson,J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.

4. Grimson,A., Farh,K.K., Johnston,W.K., Garrett-Engele,P., Lim,L.P. and Bartel,D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
5. Guo,H., Ingolia,N.T., Weissman,J.S. and Bartel,D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.
6. Farh,K.K., Grimson,A., Jan,C., Lewis,B.P., Johnston,W.K., Lim,L.P., Burge,C.B. and Bartel,D.P. (2005) The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science*, **310**, 1817–1821.
7. Gaidatzis,D., van Nimwegen,E., Hausser,J. and Zavolan,M. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*, **8**, 69.
8. Nam,S., Kim,B., Shin,S. and Lee,S. (2008) miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res.*, **36**, D159–D164.
9. Creighton,C.J., Nagaraja,A.K., Hanash,S.M., Matzuk,M.M. and Gunaratne,P.H. (2008) A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions. *RNA*, **14**, 2290–2296.
10. Viswanathan,S.R., Powers,J.T., Einhorn,W., Hoshida,Y., Ng,T.L., Toffanin,S., O'Sullivan,M., Lu,J., Phillips,L.A., Lockhart,V.L. *et al.* (2009) Lin28 promotes transformation and is associated with advanced human malignancies. *Nat. Genet.*, **41**, 843–848.
11. Cheng,C. and Li,L.M. (2008) Inferring microRNA activities by combining gene expression with microRNA target prediction. *PLoS One*, **3**, e1989.
12. Ulitsky,I., Laurent,L.C. and Shamir,R. (2010) Towards computational prediction of microRNA function and activity. *Nucleic Acids Res.*, **38**, e160.
13. Suzuki,H.I., Yamagata,K., Sugimoto,K., Iwamoto,T., Kato,S. and Miyazono,K. (2009) Modulation of microRNA processing by p53. *Nature*, **460**, 529–533.
14. Suzuki,H.I. and Miyazono,K. (2010) Dynamics of microRNA biogenesis: crosstalk between p53 network and microRNA processing pathway. *J. Mol. Med.*, **88**, 1085–1094.
15. Suzuki,H.I., Arase,M., Matsuyama,H., Choi,Y.L., Ueno,T., Mano,H., Sugimoto,K. and Miyazono,K. (2011) MCPIP1 ribonuclease antagonizes dicer and terminates microRNA biogenesis through precursor microRNA degradation. *Mol. Cell*, **44**, 424–436.
16. Segal,E., Friedman,N., Koller,D. and Regev,A. (2004) A module map showing conditional activity of expression modules in cancer. *Nat. Genet.*, **36**, 1090–1098.
17. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.
18. Ulitsky,I., Maron-Katz,A., Shavit,S., Sagir,D., Linhart,C., Elkon,R., Tanay,A., Sharan,R., Shiloh,Y. and Shamir,R. (2010) Expander: from expression microarrays to networks and functions. *Nat. Protoc.*, **5**, 303–322.
19. Montes-Moreno,S., Martinez,N., Sanchez-Espiridion,B., Diaz Uriarte,R., Rodriguez,M.E., Saez,A., Montalban,C., Gomez,G., Pisano,D.G., Garcia,J.F. *et al.* (2011) miRNA expression in diffuse large B-cell lymphoma treated with chemoimmunotherapy. *Blood*, **118**, 1034–1040.
20. The Cancer Genome Atlas Research Network. (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**, 1061–1068.
21. Plaisier,S.B., Taschereau,R., Wong,J.A. and Graeber,T.G. (2010) Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.*, **38**, e169.
22. Lossos,I.S., Czerwinski,D.K., Alizadeh,A.A., Wechser,M.A., Tibshirani,R., Botstein,D. and Levy,R. (2004) Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N. Engl. J. Med.*, **350**, 1828–1837.
23. Allison,D.B., Cui,X., Page,G.P. and Sabripour,M. (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat. Rev. Genet.*, **7**, 55–65.
24. Alizadeh,A.A., Eisen,M.B., Davis,R.E., Ma,C., Lossos,I.S., Rosenwald,A., Boldrick,J.C., Sabet,H., Tran,T., Yu,X. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503–511.
25. Dugas,J.C., Cuellar,T.L., Scholze,A., Ason,B., Ibrahim,A., Emery,B., Zamanian,J.L., Foo,L.C., McManus,M.T. and Barres,B.A. (2010) Dicer1 and miR-219 Are required for normal oligodendrocyte differentiation and myelination. *Neuron*, **65**, 597–611.
26. Gennarino,V.A., D'Angelo,G., Dharmalingam,G., Fernandez,S., Russolillo,G., Sanges,R., Mutarelli,M., Belcastro,V., Ballabio,A., Verde,P. *et al.* (2012) Identification of microRNA-regulated gene networks by expression analysis of target genes. *Genome Res.*, **22**, 1163–1172.
27. van Kouwenhove,M., Kedde,M. and Agami,R. (2011) MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat. Rev. Cancer*, **11**, 644–656.
28. Chi,S.W., Hannon,G.J. and Darnell,R.B. (2012) An alternative mode of microRNA target recognition. *Nat. Struct. Mol. Biol.*, **19**, 321–327.
29. Gerstein,M.B., Kundaje,A., Hariharan,M., Landt,S.G., Yan,K.K., Cheng,C., Mu,X.J., Khurana,E., Rozowsky,J., Alexander,R. *et al.* (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature*, **489**, 91–100.
30. Ooi,C.H., Oh,H.K., Wang,H.Z., Tan,A.L., Wu,J., Lee,M., Rha,S.Y., Chung,H.C., Virshup,D.M. and Tan,P. (2011) A densely interconnected genome-wide network of microRNAs and oncogenic pathways revealed using gene expression signatures. *PLoS Genet.*, **7**, e1002415.
31. Sun,J., Gong,X., Purow,B. and Zhao,Z. (2012) Uncovering microRNA and transcription factor mediated regulatory networks in glioblastoma. *PLoS Comput. Biol.*, **8**, e1002488.
32. Melton,C., Judson,R.L. and Blelloch,R. (2010) Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature*, **463**, 621–626.

Upsala Journal of Medical Sciences. 2012; 117: 143–152

informa
healthcare

REVIEW ARTICLE

Tumor-promoting functions of transforming growth factor-β in progression of cancer

KOHEI MIYAZONO[1,2], SHOGO EHATA[1] & DAIZO KOINUMA[1]

[1]*Department of Molecular Pathology, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan, and* [2]*Ludwig Institute for Cancer Research, Box 595, SE 751-24 Uppsala, Sweden*

Abstract

Transforming growth factor-β (TGF-β) elicits both tumor-suppressive and tumor-promoting functions during cancer progression. Here, we describe the tumor-promoting functions of TGF-β and how these functions play a role in cancer progression. Normal epithelial cells undergo epithelial-mesenchymal transition (EMT) through the action of TGF-β, while treatment with TGF-β and fibroblast growth factor (FGF)-2 results in transdifferentiation into activated fibroblastic cells that are highly migratory, thereby facilitating cancer invasion and metastasis. TGF-β also induces EMT in tumor cells, which can be regulated by oncogenic and anti-oncogenic signals. In addition to EMT promotion, invasion and metastasis of cancer are facilitated by TGF-β through other mechanisms, such as regulation of cell survival, angiogenesis, and vascular integrity, and interaction with the tumor microenvironment. TGF-β also plays a critical role in regulating the cancer-initiating properties of certain types of cells, including glioma-initiating cells. These findings thus may be useful for establishing treatment strategies for advanced cancer by inhibiting TGF-β signaling.

Key words: *Angiogenesis, cancer-initiating cell, EMT, invasion, metastasis, TGF-β*

Introduction

Transforming growth factor-β (TGF-β) is a multifunctional regulator of cell growth, apoptosis, differentiation, and migration. TGF-β1 was originally discovered as a secreted protein that induces anchorage-independent growth in normal rat kidney NRK49F fibroblasts in the presence of epidermal growth factor (EGF) (1). TGF-β was shown to potently inhibit the proliferation of most cell types, including epithelial cells, endothelial cells, hematopoietic cells, and lymphocytes, and is widely known as a tumor suppressor. Studies investigating TGF-β signaling have revealed that perturbations of the TGF-β signaling pathway, such as mutations of TGF-β receptors or Smad proteins, lead to cancer progression and are related to poor prognosis of certain types of cancer. However, recent findings have shown that cancer cells become resistant to the growth inhibitory activity of TGF-β and that TGF-β facilitates invasion and metastasis of these cells both *in vitro* and *in vivo*.

Accumulating evidence has revealed that TGF-β plays a bidirectional role in cancer progression (2,3). TGF-β acts as a tumor suppressor by inhibiting cell growth through suppressing c-Myc expression and stimulating certain cyclin-dependent kinase inhibitors, including $p21^{WAF1}$ and $p15^{Ink4b}$, and by inducing cellular apoptosis through inducing DAP kinase, GADD45β, and Bim (4). Conversely, TGF-β functions as a tumor-promoting factor by stimulating extracellular matrix deposition and tissue fibrosis, perturbing immune and inflammatory function, stimulating angiogenesis, and promoting epithelial-mesenchymal transition (EMT).

In this review article, we discuss the tumor-promoting functions of TGF-β, particularly on EMT, on the basis of recent findings in our

Correspondence: Kohei Miyazono, Department of Molecular Pathology, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. Fax: +81-3-5841-3354. E-mail: miyazono@m.u-tokyo.ac.jp

(Received 9 September 2011; accepted 3 November 2011)

ISSN 0300-9734 print/ISSN 2000-1967 online
DOI: 10.3109/03009734.2011.638729

laboratory. We also describe the function of TGF-β in some cancer-initiating cells and discuss how inhibition of TGF-β signaling can be used for treating different types of cancer.

TGF-β family signaling

TGF-β binds to two different serine/threonine kinase receptors, TβRII and TβRI (5). Betaglycan, also known as the TGF-β type III receptor, facilitates binding of TGF-β (particularly TGF-β2 among the three isoforms of TGF-β) to TβRII. TβRII activates TβRI through phosphorylation of the Gly-Ser-rich (GS) domain of TβRI, which in turn phosphorylates and activates Smad2 and Smad3, receptor-regulated Smads (R-Smads) specific for TGF-β and activin signaling (Figure 1). Bone morphogenetic proteins (BMPs) activate another set of R-Smads, including Smad1, Smad5, and Smad8 (6). Activated Smad2 and Smad3 form complexes with Smad4, common partner Smad (co-Smad), and translocate into the nucleus. R-Smad/co-Smad complexes associate with various transcription factors (AP-2, Ets, and HNF-4α (7–9)) and transcriptional co-activators (p300, CBP, and GCN5) or co-repressors (p107, Ski, and SnoN) in the nucleus and regulate transcription of a wide spectrum of TGF-β target genes. Smad7, an inhibitory Smad (I-Smad), represses TGF-β signaling through multiple mechanisms; among these mechanisms, binding to activated type I receptors and competition with R-Smads for receptor binding play a major role in regulation of TGF-β signaling (10). c-Ski (also known as SKI) and the related SnoN (also known as SKIL) bind directly to Smad2/3 and Smad4 and function as transcriptional co-repressors by recruiting histone deacetylases and competing for binding with p300/CBP. C-Ski also disrupts formation of the R-Smads and

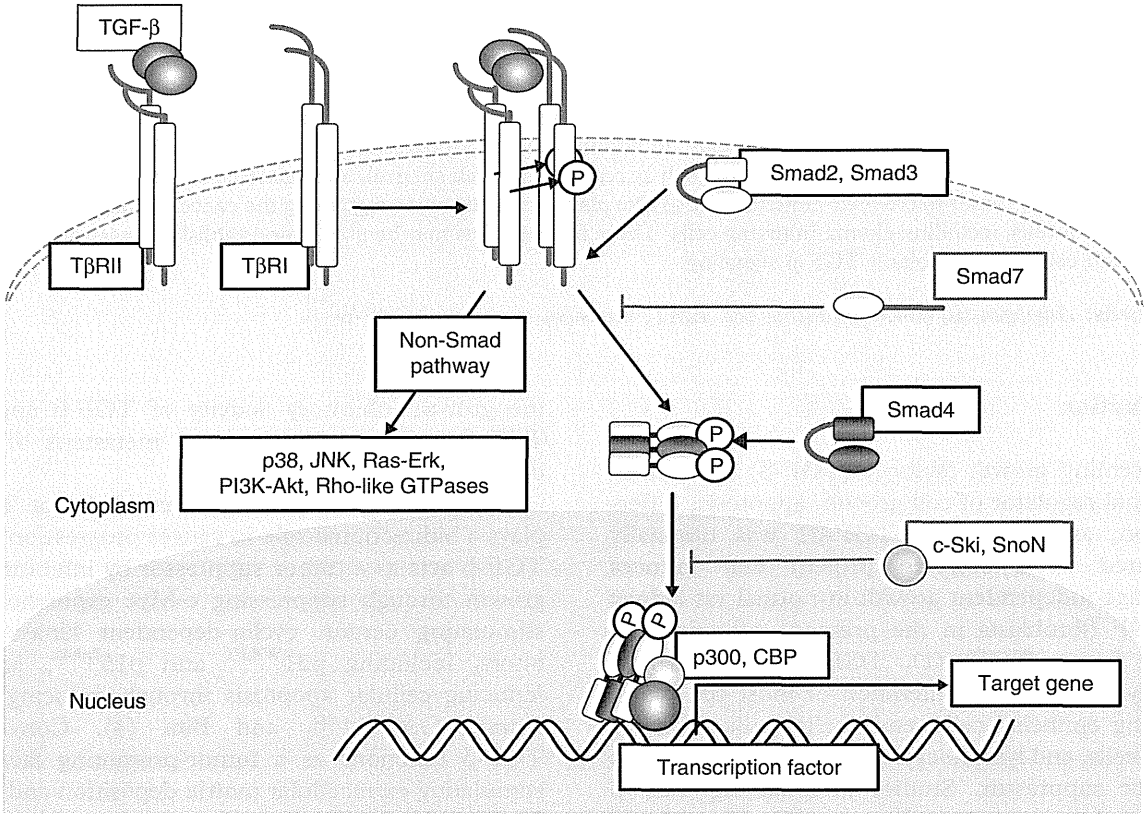

Figure 1. Schematic representation of TGF-β signal transduction pathways. TGF-β transduces signals through two different types of serine/threonine (and tyrosine) kinase receptors, termed TβRI and TβRII. Upon TGF-β binding, TβRI and TβRII form heterotetrameric complexes, and TβRII kinase transphosphorylates the juxtamembrane portion (GS domain) of the cytoplasmic region of TβRI. Phosphorylated TβRI transmits intracellular signaling through R-Smad phosphorylation. Smad2 and Smad3 are R-Smads phosphorylated by TβRI kinase and form heteromeric complexes with Smad4 (co-Smad). Smad complexes translocate into the nucleus and act as transcriptional regulators of target genes by interacting with other transcription factors and transcriptional regulators. Smad7 (I-Smad), which lacks the typical MH1 domain, interferes with the activation of R-Smads by interacting with TβRI and competitively prevents R-Smads from being phosphorylated by TβRI. TGF-β activates other intracellular signaling pathways in addition to Smads in order to regulate a wide array of cellular functions. These non-Smad pathways are activated by TGF-β receptors through phosphorylation or direct interaction.

co-Smad complex to inhibit TGF-β signaling (11). In addition to its involvement in Smad signaling pathways, TGF-β activates various non-Smad signaling pathways, including ERK, JNK, and p38 MAP kinases, phosphatidylinositol-3 kinase (PI3K)-Akt, and small GTPase pathways (12). TβRI functions as a dual-specificity kinase (tyrosine and serine/threonine kinase) and phosphorylates ShcA on tyrosine and serine residues to activate the MAP kinase pathway (13).

Induction of EMT

EMT is a differentiation switch through which epithelial cells differentiate into mesenchymal cells, and it occurs in the process of tissue morphogenesis during development, wound repair, and cancer progression in adult tissues (14,15). An early event of EMT includes disruption of tight junctions connecting epithelial cells and delocalization of tight junction proteins, such as ZO-1, claudin-1, and occludin. Early events of EMT also include disruption of adherence junctions, which contain E-cadherin and β-catenin, and reorganization of the actin cytoskeleton. Epithelial cells lose cell polarity and show spindle-like morphology with expression of various mesenchymal markers, including N-cadherin, fibronectin, and α-smooth muscle actin (α-SMA). Cell motility and invasive properties are enhanced in resulting mesenchymal cells.

EMT can be classified into three subtypes (16). Type 1 EMT occurs during development and includes the mesenchymal transition of primitive epithelial cells during gastrulation, generation of migrating neural crest cells from neuroepithelial cells, and formation of endocardial cushion tissue from cardiac endothelial cells. Type 2 EMT includes the transition of secondary epithelial (and endothelial) cells to tissue fibroblasts, which can be observed during the processes of wound healing, regeneration, and fibrosis in adult tissues. Type 3 EMT also occurs in adult tissues and involves the mesenchymal transition of epithelial carcinoma cells, leading to generation of metastatic tumor cells.

TGF-β is well known to induce EMT in various epithelial cells, including normal mouse epithelial NMuMG cells and A549 lung adenocarcinoma cells (17). Many transcription factors, including the two-handed zinc-finger factors δEF1 (also known as ZEB1) and SIP1 (ZEB2), the zinc-finger factors Snail (also known as SNAI1) and Slug (SNAI2), and the basic helix-loop-helix (bHLH) factors Twist and E12/E47, are induced by TGF-β signaling in a Smad-dependent fashion and play critical roles in EMT induction. Additionally, non-Smad signaling pathways activated by TGF-β and cross-talk with other signaling pathways, including fibroblast growth factor (FGF) and tumor necrosis factor-α (TNF-α) signaling, play important roles in EMT promotion.

Induction of EMT in tumor stromal cells by TGF-β

Epithelial cells in the tumor stroma undergo EMT (type 2 EMT) and play a critical role in cancer progression. We cocultured NMuMG cells with mouse mammary tumor JygMC(A) cells and found that NMuMG cells that have undergone EMT express α-SMA (18). The effect of the JygMC(A) cells was abolished by treatment with the TβRI inhibitor SB431542. Interestingly, when NMuMG cells were cocultured with the mouse mammary tumor cell line 4T1, NMuMG cells underwent EMT and produced mesenchymal cells with an activated fibroblastic phenotype, which lacked α-SMA expression. 4T1 cells produced TGF-β1 at a level comparable to that produced by JygMC(A) cells. When 4T1 cells were treated with FGF receptor 1 (FGFR1) inhibitor SU5402, α-SMA-positive NMuMG cells were detected, indicating that the loss of α-SMA expression is due to FGF(s) secreted from 4T1 cells. We have shown that treatment of NMuMG cells with TGF-β and FGF-2 prevents the production of mesenchymal cells expressing α-SMA and calponin by activating the MEK-ERK pathway. Interestingly, NMuMG cells that have undergone EMT following treatment with TGF-β and FGF-2 exhibit drastic morphological changes with marked actin reorganization, enhanced cell migration, and increased production of matrix metalloproteinases (MMPs), including MMP-9. Moreover, NMuMG cells treated with TGF-β and FGF-2 enhanced the invasion of cocultured breast cancer cells into collagen gels *in vitro*. Thus, TGF-β and FGF-2 co-operate with each other to produce 'activated' fibroblasts in the tumor microenvironment, and activated fibroblasts may in turn secrete substances such as MMPs to induce invasion and metastasis of adjacent cancer cells (Figure 2).

During EMT progression, TGF-β induces isoform switching of FGFRs. Of the 22 FGFs (19), epithelial cells respond to specific FGFs, including FGF-7 (also known as keratinocyte growth factor (KGF)), but not to FGF-2 (basic FGF) or FGF-4. However, cells that have undergone EMT become responsive to FGF-2 and FGF-4, but not to FGF-7 (18). We have shown that TGF-β-mediated EMT induces isoform switching of FGFRs through alternative splicing, following which expression of the IIIb isoform of FGFR decreased and that of the IIIc isoform increased.

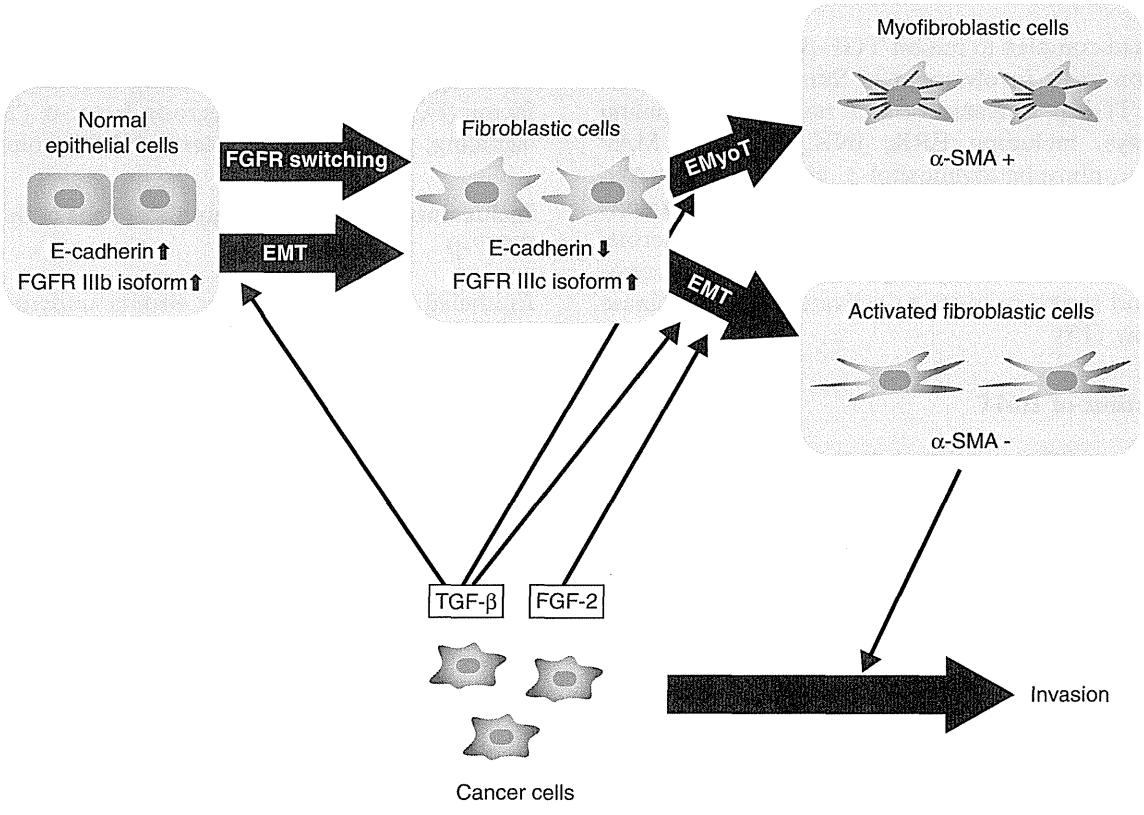

Figure 2. Schematic representation of EMT induction by TGF-β and FGF-2. 'Epithelial cells' differentiate into 'fibroblastic cells' through EMT induced by TGF-β and further differentiate into α-SMA-positive 'myofibroblastic cells' through epithelial-myofibroblastic transition (EMyoT). When FGF-2 is present in this process, FGF-2 induces differentiation of epithelial cells to 'activated fibroblastic cells'.

Exon array analysis showed that TGF-β alters a broad spectrum of splicing patterns by reducing the expression of epithelial splicing regulatory proteins (ESRPs) 1 and 2 (20). Warzecha et al. (21) recently reported that the ESRP-regulated splicing pathway is abrogated during EMT. We found that repression of the expression of ESRPs by TGF-β is mediated by up-regulation of the δEF1 family proteins δEF1 and SIP1, which suppress the transcription of ESRP(s) by binding to the ESRP promoter(s). Interestingly, the expression profiles of ESRPs were reciprocally correlated with those of δEF1 and SIP1 in human breast cancer cell lines as well as in tumor specimens. In addition to FGFRs, TGF-β induces alternative splicing of CD44, Mena, and CTNND1 (also known as δ-catenin or p120 catenin), which are reportedly involved in cancer progression. We have also shown that over-expression of ESRPs attenuates TGF-β-induced EMT and restores the expression of E-cadherin and some other epithelial phenotypes. Thus, ESRPs are downstream targets of TGF-β and serve as antagonists to EMT by regulating alternative splicing of specific genes involved in TGF-β-induced EMT.

Induction of EMT in cancer cells

EMT is observed in some transformed epithelial cells (type 3 EMT) to facilitate their invasive and metastatic properties. Type 3 EMT can be regulated by specific oncogenic and anti-oncogenic signals. We have shown that a zinc-finger transcription factor Snail is induced by TGF-β in pancreatic cancer Panc1 cells and plays a key role in EMT progression (22). Panc1 cells express active *K-ras*, and we found that induction of Snail by TGF-β is dependent on oncogenic Ras signals. Snail was strongly induced by TGF-β in Panc1 cells, but knock-down of Ras in Panc1 cells abolished Snail induction by TGF-β. Consequently, TGF-β failed to efficiently induce EMT in Panc1 cells in the absence of active Ras signaling. Exogenous expression of constitutively active Ras into HeLa cells resulted in marked induction of Snail by TGF-β, while induction of other direct targets of TGF-β, including Smad7 and PAI-1, was not enhanced by Ras signaling. MAP kinases have been reported to phosphorylate the linker region of Smad2 and Smad3, which both positively and negatively regulates TGF-β signaling

(23). However, MAP kinase signaling was not required for induction of Snail by TGF-β, and it is currently unknown which downstream signals of Ras co-operate with TGF-β signaling (22).

Thyroid transcription factor-1 (TTF-1, the protein product of the *NKX2.1* gene) is expressed in normal lung tissues and acts as a master regulator of lung morphogenesis (24). TTF-1 is primarily expressed in type II pneumocytes and Clara cells and frequently expressed in lung cancer cells, including lung adenocarcinoma cells. Although the *TTF1* gene is amplified in some lung adenocarcinoma cells and may function as an oncogene (25), loss of TTF-1 expression is reportedly associated with poor prognosis of lung carcinoma. Recently, Winslow et al. (26) reported that TTF-1 controls differentiation of lung carcinoma cells and limits their metastatic potential in mice with active K-Ras and inactive p53. Interestingly, we found that TTF-1 functions as a tumor-suppressor during EMT induction. TTF-1 is highly expressed in certain types of lung adenocarcinoma cell lines, including H441 cells and LC-2/ad cells, but not in A549 cells (27). A549 cells show a spindle-like phenotype and grow rapidly, while H441 cells show tight cell–cell junctions with cobblestone-like morphology and grow much more slowly than A549 cells. A549 cells express low levels of TTF-1 and E-cadherin, while H441 cells express high levels of TTF-1 and E-cadherin. We have further shown that exogenous expression of TTF-1 in A549 cells inhibits TGF-β-induced EMT, decreases MMP-2 activity, cell migration, and cellular invasive capacity, and restores the epithelial phenotype through high E-cadherin expression. Conversely, TGF-β induces the expression of Snail and Slug in A549 cells, and silencing of TTF-1 in H441 cells enhances TGF-β-mediated EMT. TTF-1 has been reported to interact physically with Smad3 (28) and may inhibit Smad3 function. We have also shown that TGF-β down-regulates TTF-1 expression in A549 cells and that TTF-1 inhibits the expression of TGF-β2, which is expressed in epithelial cells at the tip of the distal airway during lung morphogenesis. Thus, TTF-1 may exert a tumor-suppressive effect through antagonizing the effect of TGF-β. These findings indicate a functionally inverse relationship between TTF-1 and TGF-β signaling in the progression of lung adenocarcinoma through regulation of EMT.

TGF-β signaling in vascular tissues and angiogenesis

New blood vessel formation in tumor tissues (tumor angiogenesis) is essential for the growth and metastasis of tumor cells. Although TGF-β potently inhibits the growth of endothelial cells *in vitro*, it functions as a pro-angiogenic factor and stimulates angiogenesis *in vivo*. Increased expression of TGF-β is correlated to increased vascular density in some types of tumors.

For induction of tumor angiogenesis, TGF-β induces the expression of angiogenic factors, including connective tissue growth factor (CTGF) and vascular endothelial growth factor (VEGF) (29). Additionally, TGF-β stimulates the synthesis of MMP-2 and MMP-9 and down-regulates the expression of tissue inhibitors of metalloproteinase (TIMPs) in tumor tissues. Increased MMP activity leads to stimulation of migration and invasion of vascular endothelial cells, resulting in accelerated tumor angiogenesis.

However, TGF-β suppresses angiogenesis in certain types of tumors through reduced expression of some angiogenic factors or increased expression of angiogenic inhibitors. In diffuse-type gastric carcinoma, TGF-β induces the production of some angiogenic inhibitors, including thrombospondin-1 and TIMP-2, and perturbations of TGF-β signaling may thus lead to induction of angiogenesis and tumor growth *in vivo* (30,31).

In addition to induction of tumor angiogenesis, TGF-β acts on vascular endothelial cells and may disrupt cell–cell junctions and support the colonization of tumor cells to establish metastasis. Using endothelial cells derived from mouse embryonic stem (ES) cells, we showed that TGF-β suppresses the expression of claudin-5 and disrupts sheet formation *in vitro* (32). We also showed that TGF-β induces differentiation of certain endothelial cells into mesenchymal cells, resulting in the loss of tight cell–cell contacts *in vitro* (33). Moreover, through disruption of endothelial cell–cell junctions by inducing angiopoietin-like 4 (Angptl4) expression, TGF-β has been shown to increase the permeability of blood vessels and stimulate the trans-endothelial movement of cancer cells (34).

Acceleration of cancer metastasis by TGF-β signaling

TGF-β facilitates metastasis of certain types of cancer in advanced stages, including breast cancer (35). Inhibition of TGF-β signaling may thus be a potential strategy for preventing metastasis of advanced cancers. Though not discussed in detail in this review, TGF-β regulates tumor development by regulating immune functions (36,37). Wakefield and colleagues reported that inhibition of TGF-β function prevents the progression of breast cancer by enhancing various immune functions (38).

We have shown that Smad7, an I-Smad that inhibits TGF-β and BMP signaling, efficiently inhibits lung and liver metastasis of mouse breast cancer JygMC(A) cells (39). We subcutaneously inoculated JygMC(A) cells, which spontaneously metastasize to the lung, liver, and other organs in 3 to 4 weeks, in nude mice. Ten days after subcutaneous inoculation, adenoviruses containing Smad7 or LacZ were intravenously administered to the mice once weekly. Mice bearing JygMC(A) tumors and treated with LacZ adenovirus developed numerous metastases to the lung and liver, and all mice died by 50 days (median survival time, 41 days) after inoculation of JygMC(A) cells. In contrast, mice treated with Smad7 adenovirus showed a significant decrease in metastases of tumors in both the lung and liver, and the median survival time of Smad7-treated mice was 55 days. JygMC(A) cells treated with Smad7 showed increased expression of components involved in adherence and tight junctions, including E-cadherin, and decreased expression of mesenchymal markers, including N-cadherin. Smad7 also inhibited the migration and invasion of cells, indicating that Smad7 leads to prevention of the EMT process. Interestingly, Smad6, which preferentially inhibits BMP signaling, failed to show significant effects on the metastasis of JygMC(A) cells in nude mice, whereas c-Ski adenovirus showed effects similar to Smad7. Thus, inhibiting TGF-β signaling using Smad7 or c-Ski prevents the EMT process and eventually inhibits lung and liver metastasis of JygMC(A) cells (Figure 3).

In addition to preventing EMT, TGF-β appears to inhibit metastasis of JygMC(A) cells by some other mechanisms. Although TGF-β induces apoptosis of many different types of cells by inducing specific genes, it stimulates survival of certain types of cells in a context-dependent manner through activation of the PI3K-Akt signaling pathway. We have identified Dec1 (differentially expressed in chondrocytes 1, also known as SHARP2 and Stra13) as a downstream target of TGF-β–Smad signaling by DNA microarray analysis (40). Dec1 is a bHLH transcription factor, which is widely expressed in many tissues and overexpressed in certain types of cancer cells. Dec1 prevented the apoptosis of JygMC(A) as well as 4T1 cells, and a dominant-negative mutant of Dec1 suppressed lung and liver metastases of JygMC(A) cells in nude mice (Figure 3). Dec1 has been reported to induce the expression of an anti-apoptotic protein, survivin, in certain types of cells (41); however, we failed to show induction of survivin by TGF-β in JygMC(A) cells. Mechanisms of Dec1 induction of cell survival in JygMC(A) cells should be examined in the future.

We also found that inhibiting endogenous TGF-β signaling by a TβRI inhibitor, SB431542, induces the expression of the BH3-only protein, Bim (also known as Bcl2-like 11), in JygMC(A) and stimulates apoptosis in these cells (42). We showed that suppression of

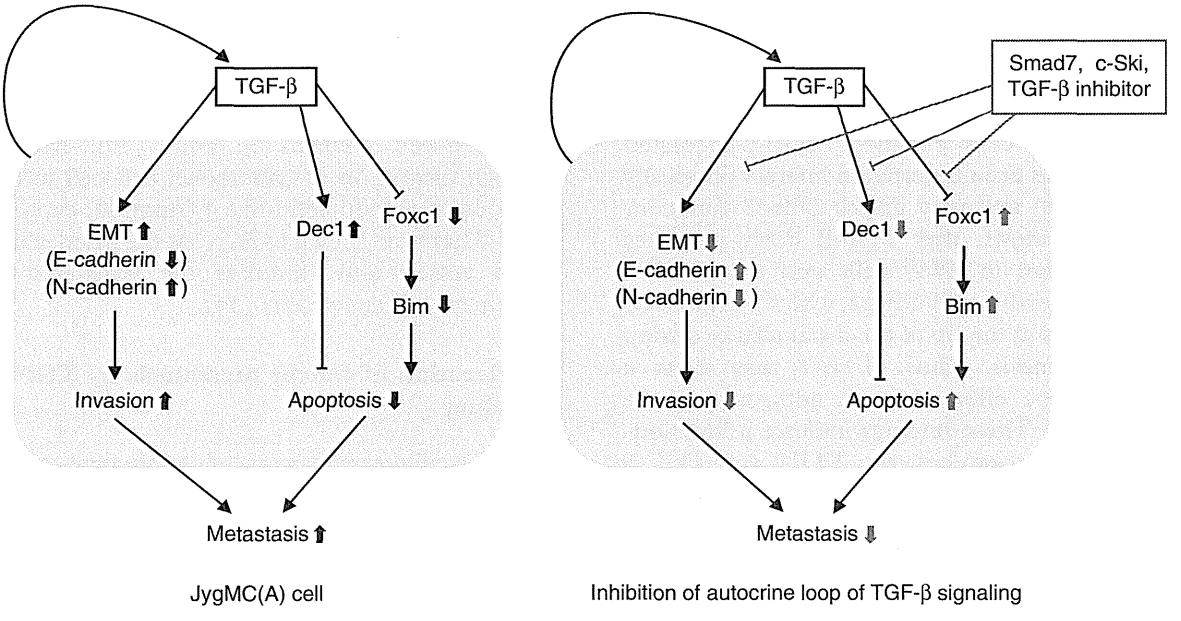

Figure 3. Mechanisms of TGF-β action on prevention of breast cancer metastasis using JygMC(A) cells. Endogenously activated autocrine loop of TGF-β regulates the expression of E-cadherin and N-cadherin by inducing EMT in JygMC(A) cells. Autocrine TGF-β also regulates the expression of various transcription factors, including Dec1 and Foxc1, and promotes the survival. Negative regulators of TGF-β signaling (Smad7, c-Ski, or TGF-β inhibitors) block these pathways and inhibit metastasis of JygMC(A) cells.

Bim expression by TGF-β is mediated by repression of a FOX family transcription factor, Foxc1, in JygMC(A) cells, thus suggesting an important role of the TGF-β–Foxc1–Bim axis in the survival of certain types of cells. Further studies are needed to determine whether the TGF-β–Foxc1–Bim axis is involved in lung and liver metastases of this type of cancer (Figure 3).

TGF-β also plays critical roles in bone metastasis, during which functional interaction between cancer cells and the bone microenvironment is important. *In-vivo* experimental models using intracardiac injection of cancer cells have been widely used to study the mechanisms of bone metastasis. Several studies revealed that TGF-β and its target molecules, such as parathyroid hormone-related protein (PTHrP) and interleukin-11 (IL-11), play critical roles in the development of bone metastasis of breast cancers (29,43), which occurs in a Smad-dependent fashion (44). PTHrP stimulates the expression of the RANK ligand (RANKL) in osteoblasts and induces differentiation of osteoclast precursors and resorption of bone. We studied the effects of a TβRI inhibitor, Ki26894, on bone metastasis in the human breast cancer cell line MDA-MB-231–5a-D (MDA-231-D), which is a highly metastatic variant of MDA-MB-231 cells. Ki26894 suppressed induction of PTHrP and IL-11 mRNA in MDA-231-D cells stimulated by TGF-β (45). When MDA-231-D cells were injected into the left ventricle of nude mice and treated with systemic administration of Ki26894 (treatment with Ki26894 was started 1 day before tumor cell inoculation), X-ray radiography showed that treatment with Ki26894 decreased bone metastasis of breast cancer cells and prolonged the survival of MDA-231-D-bearing mice compared to vehicle treatment. These findings suggest that inhibition of TGF-β signaling may be useful for preventing bone metastasis of advanced breast cancers.

TGF-β maintains stemness of certain cancer-initiating cells

Cancer-initiating cells show increased tumor-initiating ability and often exhibit stem cell-like properties such as self-renewal, multipotency, and expression of specific stem cell markers. The concept of cancer-initiating cells reveals a new strategy of therapy against intractable cancers, though it remains unclear how cancer-initiating cells can be specifically eradicated. It is important to investigate the differences between cancer-initiating cells and normal stem cells and to identify specific molecules to target cancer-initiating cells without affecting the function of normal stem cells. Recent studies have also revealed critical roles of TGF-β signaling in the maintenance of stem cell-like properties of certain cancer-initiating cells, including glioma-initiating cells (GICs) (46,47), breast cancer-initiating cells (48), and leukemia-initiating cells in chronic myeloid leukemia (CML) (49).

Glioma cells produce TGF-β1 and TGF-β2, and autocrine TGF-β signaling plays a pivotal role in maintaining the stem cell-like properties and tumorigenic activity of GICs (46,47). GICs obtained from patients with glioblastoma multiforme exhibit sphere-forming ability in a self-renewal medium containing EGF and FGF-2. Although TGF-β did not significantly affect the sphere-forming ability of GICs, a TβRI inhibitor, SB431542, efficiently reduced this ability in GICs. Moreover, SB431542 dramatically reduced the number of CD133-expressing cells and induced differentiation of GICs, leading to the appearance of cells expressing neural or glial cell markers. Analyses of TGF-β target genes using quantitative RT-PCR and by searching public datasets showed that TGF-β induces expression of the Sry-related HMG box (Sox) transcription factors Sox2 and Sox4. We showed that Sox4 is a direct target of Smad proteins activated by TGF-β and that it induces the expression of Sox2, which plays a critical role in the maintenance of GIC stemness. We also confirmed that in intracranial transplantation assays using immunocompromised mice, GICs pretreated with SB431542 showed decreased lethal potency. These results indicate that the TGF-β–Sox4–Sox2 pathway is essential for retaining the stemness of GICs, and inhibition of TGF-β signaling may be a potential method for treating glioma through targeting GICs (Figure 4).

Westermark and his colleagues (50) reported that another Sox family protein, Sox21, is expressed in glioma cells. Sox21 is an antagonizing partner of Sox2 and negatively regulates the expression of Sox2 in glioma cells. They showed that reduction in Sox2 expression using Sox2 siRNA or Sox21 over-expression reduced the cell number by inducing apoptosis.

In addition to the Sox4-Sox2 pathway, TGF-β also induces the expression of leukemia inhibitory factor (LIF) in a Smad-dependent fashion. LIF activates the downstream JAK-STAT pathway, leading to increased tumorigenesis of GICs (47). Anido et al. have shown that TGF-β inhibitors target GICs with high levels of CD44 and Id1 and that $CD44^{high}$/$Id1^{high}$ GICs are generally localized in the perivascular niche (51).

Conclusion and perspectives

As proposed by Roberts and Wakefield (2), it is now well known that TGF-β exhibits both positive and