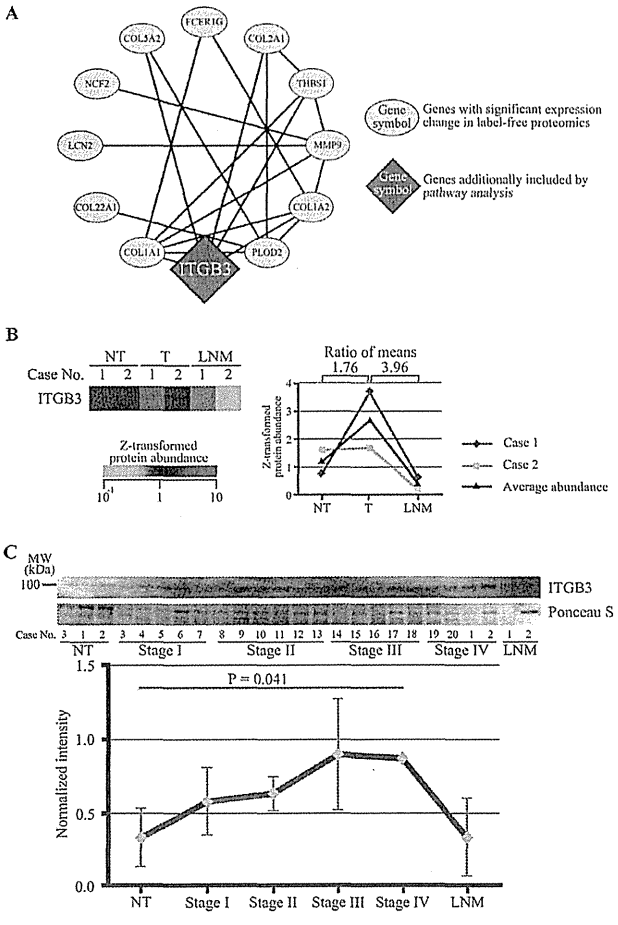


Fig. 4 – Beta-3 integrin (ITGB3) pathway and association of the ITGB3 protein expression with progression of gastric cancer. Case numbers corresponded to those in Table 1. Functional interaction network analysis by Cytoscape with the Reactome FI plug-in is demonstrated (A). The ITGB3 protein expression level quantified by a label-free method (B). ITGB3 protein expression level was z-transformed across all six samples examined and was represented as a heat-map format and line chart. ITGB3 protein expression in 20 gastric cancer cases, evaluated by Western blotting (C).

Investigation of the molecular background of tumor metastasis through "omics" studies revealed that multiple genes aberrations were contributed to the tumor metastasis [7,8]. Therefore, we tried to examine the overall features of the expressed proteins, and identified 109 aberrantly expressed proteins. Next, we identified the ITGB3 pathway as the most enriched protein interaction network thorough the pathway enrichment analysis in the 109 proteins (Fig. 4A). In addition, we monitored the ITGB3 protein expression using the created protein expression profile. Integrins are heterodimeric cell surface glycoproteins with alpha and beta subunits. ITGB3 expression is mainly associated with the migration and invasion of tumor cells [25], and overexpression of ITGB3 is related to the metastatic potential of melanoma [26], breast cancer [27], colorectal cancer [28], and bone metastasis in breast cancer [29]. However, in ovarian cancer, ITGB3 reduces the metastatic potential of tumor cells [30–32]. In hepatocellular carcinoma cells, ITGB3 has been shown to have a proapoptotic function, and downregulation of ITGB3 is related to aggressive tumor growth [33]. Therefore, the association of ITGB3 expression with tumor metastasis and progression depends on the cancer type. In the current study, ITGB3 expression was increased as cancer stage advanced, thereby linking it to gastric cancer progression. These observations are supported by previous findings [34]. However, downregulation of ITGB3 in the LNM tissues has not been reported until our study. Neoplastic invading cells must overcome the integrin-mediated death (IMD) induced by the interaction of ITGB3 with the extracellular matrix of host tissues to establish metastases [25]. When a tumor cell migrates through a microenvironment where the extracellular matrix (ECM) does not contain a suitable ligand for ITGB3, the integrin cytoplasmic tail recruits caspase-8 to the cell membrane, and apoptosis is induced. Stupack et al. [35] showed that ITGB3-expressing melanoma cells undergo apoptosis in collagen gels that lack ITGB3 ligands. ITGB3 specifically binds a wide range of ECM molecules, including fibronectin, fibrinogen, von Willebrand factor, and vitronectin [36]. The present proteomic analysis demonstrated that several collagen-related genes were upregulated in LNM tissues, while fibronectin, fibrinogen, and vitronectin were not identified as upregulated proteins in LNM tissues (Fig. 2 and Supplementary Table 3). Therefore, ITGB3 downregulation in gastric cancer cells that showed LNM might reflect a survival mechanism related to the prevention of IMD in metastatic tissues.

The problems of the proteomic biomarker studies are lack of the validation studies in the independent clinical samples from the multi-institutions. Selection of biomarker candidates from the differentially expressed proteins and acquisition of the independent clinical samples are critical steps in multi-institutional biomarker validation studies. Transcriptome data are deposited in public databases with clinicopathological data of the samples, and are freely available for independent validation studies; as clinical materials are generally not very accessible, such infrastructure is highly desirable for proteomic biomarker studies. However, presently, there is no analogous proteomic database. For proteins whose levels demonstrate concordance with mRNA expression, a transcriptome database should be a useful tool for biomarker selection and validation. Several studies have compared mRNA and protein levels in tissue culture cells. Chen et al. [37] compared mRNA and protein expression between two gastric cancer cell lines and estimated an overall correlation coefficient of 0.29. Previous studies comparing mRNA and protein expression in human and mouse cell lines concluded that the mRNA levels explain approximately 40% of the variability in protein levels [38]. Although these studies suggested possible benefits of using transcriptome databases for validation studies, the correlation between protein and mRNA expression in surgically resected tissues remain unclear. We analyzed the mRNA levels for the 109 differentially expressed proteins by using a public mRNA expression database. Of these 109 proteins, 67 were listed in the gastric cancer dataset of the GEO database. We found that 26 of 67 genes (38.8%) showed a concordant expression pattern for protein and mRNA. These concordant genes were enriched in GO terms related to

Table 2 – Pathway enrichment analysis for the identified protein network consisting of 11 proteins.

Pathway[a]	Source[b]	FDR[c]	Number of proteins[d]	Included proteins[e]
Beta3 integrin cell surface interactions	NCI-PID	1.00E–04	3	COL1A2, COL1A1, THBS1
Amoebiasis	KEGG	1.11E–04	4	COL1A2, COL2A1, COL1A1, COL5A2
Integrin cell surface interactions	Reactome	1.25E–04	4	COL1A2, COL2A1, COL1A1, THBS1
Protein digestion and absorption	KEGG	1.43E–04	4	COL1A2, COL2A1, COL1A1, COL5A2
Signaling by PDGF	Reactome	1.67E–04	5	COL1A2, COL2A1, COL1A1, THBS1, COL5A2
Integrin signaling pathway	Panther	1.82E–04	4	COL1A2, COL2A1, COL1A1, COL5A2
Focal adhesion	KEGG	2.00E–04	5	COL1A2, COL2A1, COL1A1, THBS1, COL5A2
Platelet adhesion to exposed collagen	Reactome	2.50E–04	3	COL1A2, FCER1G, COL1A1
ECM-receptor interaction	KEGG	3.33E–04	5	COL1A2, COL2A1, COL1A1, THBS1, COL5A2
Beta1 integrin cell surface interactions	NCI-PID	5.00E–04	5	COL1A2, COL2A1, COL1A1, THBS1, COL5A2
Extracellular matrix organization	Reactome	1.00E–03	7	PLOD2, MMP9, COL22A1, COL1A2, COL2A1, COL1A1, COL5A2
Cell surface interactions at the vascular wall	Reactome	1.31E–03	3	COL1A2, FCER1G, COL1A1
Axon guidance	Reactome	1.42E–03	4	COL1A2, COL2A1, COL1A1, COL5A2
Platelet activation, signaling and aggregation	Reactome	1.93E–03	4	COL1A2, FCER1G, COL1A1, THBS1
VEGFR3 signaling in lymphatic endothelium	NCI-PID	4.00E–03	2	COL1A2, COL1A1
Syndecan-4-mediated signaling events	NCI-PID	6.06E–03	2	MMP9, THBS1
Validated transcriptional targets of AP1 family members Fra1 and Fra2	NCI-PID	7.76E–03	2	MMP9, COL1A2
Bladder cancer	KEGG	9.39E–03	2	MMP9, THBS1

[a] Pathway enrichment analysis was performed using Cytoscape with Reactome FI plug-in [http://wiki.reactome.org/index.php/Reactome_FI_Cytoscape_Plugin].
[b] Reactome FI data set unites interactions from Reactome and those derived from other pathway databases, including KEGG, NCI-PID, BioCyc, Panther and The Cancer Cell Map.
[c] FDR was calculated by Reactome FI plug-in, and the significant threshold was set at less than 0.05.
[d] The numbers of proteins which were examined in this study involved in the pathway.
[e] Gene names which were identified to be significantly up- or down-regulated by proteomic experiment were shown.

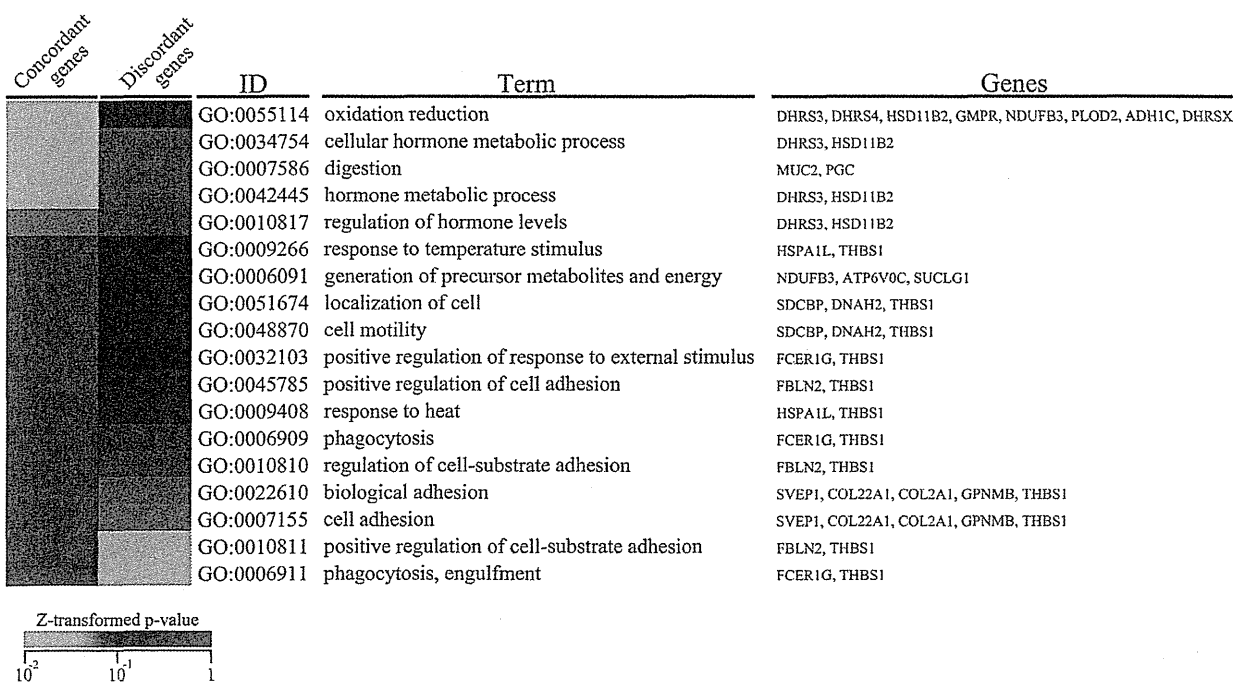


Fig. 5 – Functional characteristics of genes with concordant and discordant protein and mRNA expression. Genes were grouped according to their protein and mRNA expression patterns and analyzed for enriched gene ontology terms by DAVID, version 6.7 (http://david.abcc.ncifcrf.gov). GO terms with enrichment of $P < 0.20$ were listed, and P-values were z-transformed and have been presented in a heatmap format.

Table 3 – List of 26 genes with correlation between protein and mRNA expression.

Gene symbol[a]	Gene ID[b]	Gene description	Protein			mRNA			
			Accession number[c]	Ratio (T/NT)[d]	Ratio (LNM/T)[e]	Probe ID[f]	Accession number[g]	Ratio (T/NT)[d]	Ratio (LNM/T)[e]
Up-regulated genes at both protein and mRNA level									
MMP9	4318	Matrix metalloproteinase-9	P14780	3.46	2.63	10805	T64837	2.56	1.03
AGMAT	79814	Agmatinase, mitochondrial	Q9BSE5	2.91	2.51	15950	AA934764	1.11	1.24
CTSH	1512	Cathepsin H	P09668	2.78	3.33	953	AA487231	1.04	1.05
TGOLN2	10618	Trans-Golgi network integral membrane protein 2	O43493	2.84	3.73	15417	T81338	1.21	1.05
SLC26A2	1836	Sulfate transporter	P50443	2.42	2.07	10913	W15263	1.13	1.02
XRN2	22803	5′–3′ exoribonuclease 2	Q9H0D6	13.38	2.36	8162	AA028164	1.41	1.13
FAM103A1	83640	Protein FAM103A1	Q9BTL3	2.09	2.79	12871	AA432100	1.14	1.01
TAPBP	6892	Tapasin	O15533	102.06	726.07	18291	AA704775	2.20	1.09
TCL1A	8115	T-cell leukemia/lymphoma protein 1A	P56279	88.57	3.65	12786	R97095	1.07	2.62
GDAP1	54332	Ganglioside-induced differentiation-associated protein 1	Q8TB36	3.99	3.09	17310	H15302	1.45	1.06
NCF2	4688	Neutrophil cytosol factor 2	P19878	4.97	6.98	6201	AA872098	3.26	1.42
CHD4	1108	Chromodomain-helicase-DNA-binding protein 4	Q14839	2.07	2.49	5588	N34372	1.16	1.20
Down-regulated genes at both protein and mRNA level									
CA2	760	Carbonic anhydrase 2	P00918	4.61E−01	2.99E−01	15331	H23187	1.21E−01	9.92E−01
GKN2	200504	Gastrokine-2	Q86XP6	3.76E−01	4.97E−01	8090	AI732254	1.45E−02	3.96E−01
MUC2	4583	Mucin-2	Q02817	4.04E−01	3.47E−01	25161	AA534503	4.62E−01	5.30E−01
CLDN18	51208	Claudin-18	P56856	3.91E−01	2.98E−01	11694	AI820565	3.45E−01	6.81E−01
DHRS4	10901	Dehydrogenase/reductase SDR family member 4	Q9BTZ2	4.87E−01	2.76E−01	18212	AA429946	6.46E−01	9.40E−01
DHRS3	9249	Short-chain dehydrogenase/reductase 3	O75911	3.98E−01	1.68E−01	1133	AA171606	7.77E−01	8.00E−01
HSD11B2	3291	Corticosteroid 11-beta-dehydrogenase isozyme 2	P80365	3.70E−01	3.27E−01	174	W95082	6.95E−01	8.90E−01
PGC	5225	Gastricsin	P20142	4.34E−01	3.00E−01	19526	AI674972	3.04E−02	3.25E−01
REG4	83998	Regenerating islet-derived protein 4	Q9BYZ8	4.92E−01	4.95E−01	3562	AA535703	6.40E−01	7.55E−01
PRR15	222171	Protein PRR15	Q8IV56	4.33E−01	4.87E−01	17960	AA515032	8.49E−01	6.52E−01
GMPR	2766	GMP reductase 1	P36959	4.07E−01	3.27E−01	19	AA406242	5.21E−01	9.33E−01
CLMN	79789	Calmin	Q96JQ2	2.91E−01	4.71E−01	23849	AA775028	7.31E−01	8.14E−01
ABCC3	8714	Canalicular multispecific organic anion transporter 2	O15438	4.72E−01	2.75E−01	15480	AA429895	7.51E−01	6.54E−01

[a] Gene symbols were derived from UniGene.
[b] Gene IDs were derived from Entrez Gene database.
[c] Accession numbers of proteins were derived from SWISS-PROT and NCBI nonredundant databases.
[d] Ratios were calculated by dividing the mean expression value of primary tumor samples (T) by that of non-tumor samples (NT).
[e] Ratios were calculated by dividing the mean expression value of lymph node metastasis samples (LNM) by that of primary tumor samples (T).
[f] Probe IDs were derived from NCBI GEO platform (GPL1283).
[g] Accession numbers of genes were derived from GenBank database.

oxidation reduction, or hormone metabolism, while the genes with discordant expression pattern were enriched in GO terms related to phagocytosis and cell adhesion. These results are consistent with the findings of Schwänhausser et al. [38], who reported that genes related to oxidation reduction, and metabolism were enriched in the group with stable and concordant mRNA and protein expression. They also found that the GO terms of cell adhesion were enriched in the group of genes that was characterized by stable mRNA but unstable protein and was expected to have discordant expression [38]. Unfortunately, the reliability of our analysis did not reach their one because the number of samples in our proteomic study was limited and mRNA expression data was obtained from different patient cohorts. Further investigation of additional samples will be required to generalize our observations. The use of public transcriptome databases will solve the problems of proteomic biomarker studies, and therefore, these investigations should be continued.

In our study, highly comprehensive and reproducible proteomic analysis performed using a label-free quantification method showed downregulation of ITGB3 gene expression in the LNM tissues. Downregulation of ITGB3 represents a prosurvival response for overcoming apoptotic IMD at the metastatic site. Further studies on the prognostic and biological significance of ITGB3 may lead to novel risk-stratification approaches for gastric cancer.

Acknowledgements

We would greatly appreciate Dr. M Sasagawa (Nanbugo General Hospital), Dr. N Shimakage (Nagaoka Red Cross Hospital), Dr. T Suda (Nippon Dental University Medical Hospital), Dr. N Katayanagi (Niigata City General Hospital), Dr. T Tada (Tachikawa General Hospital), Dr. S Shimoda (Shibata Hospital), Dr. N Musha (Saiseikai Niigata Daini Hospital), Dr. H Tomita (Sakamachi Hospital) and K Ueki (Kashiwazaki General Hospital and Medical Center) to collect the clinical samples and the clinicopathological data. This work was supported by the National Cancer Center Research Core Facility and the National Cancer Center Research and Development Fund (23-A-8 and 23-A-10). Hiroshi Ichikawa was awardee of Research Resident Fellowship from the Foundation for Promotion of Cancer Research (Japan) for the 3rd Term Comprehensive 10-Year Strategy for Cancer Control.

REFERENCES

[1] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010;127:2893–917.
[2] Isobe Y, Nashimoto A, Akazawa K, Oda I, Hayashi K, Miyashiro I, et al. Gastric cancer treatment in Japan: 2008 annual report of the JGCA nationwide registry. Gastric Cancer 2011;14:301–16.
[3] Wang W, Li YF, Sun XW, Chen YB, Li W, Xu DZ, et al. Prognosis of 980 patients with gastric cancer after surgical resection. Chin J Cancer 2010;29:923–30.
[4] Park SR, Kim MJ, Ryu KW, Lee JH, Lee JS, Nam BH, et al. Prognostic value of preoperative clinical staging assessed by computed tomography in resectable gastric cancer patients: a viewpoint in the era of preoperative treatment. Ann Surg 2010;251:428–35.
[5] Deng J, Liang H, Sun D, Zhang R, Zhan H, Wang X. Prognosis of gastric cancer patients with node-negative metastasis following curative resection: outcomes of the survival and recurrence. Can J Gastroenterol 2008;22:835–9.
[6] Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. Science 2011;331:1559–64.
[7] Xie HL, Li ZY, Gan RL, Li XJ, Zhang QL, Hui M, et al. Differential gene and protein expression in primary gastric carcinomas and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. J Dig Dis 2010;11:167–75.
[8] Liu X-P, Li D-Y, Liu X-L, Xu J-D, Furuya T, Kawauchi S, et al. Comparison of chromosomal aberrations between primary tumors and their synchronous lymph-node metastases in intestinal-type gastric carcinoma. Pathol Res Pract 2009;205:105–11.
[9] Silvestri A, Calvert V, Belluco C, Lipsky M, De Maria R, Deng J, et al. Protein pathway activation mapping of colorectal metastatic progression reveals metastasis-specific network alterations. Clin Exp Metastasis 2013;30:309–16.
[10] Rusch VW, Rice TW, Crowley J, Blackstone EH, Rami-Porta R, Goldstraw P. The seventh edition of the American Joint Committee on Cancer/International Union Against Cancer Staging Manuals: the new era of data-driven revisions. J Thorac Cardiovasc Surg 2010;139:819–21.
[11] Kondo T, Hirohashi S. Application of highly sensitive fluorescent dyes (CyDye DIGE Fluor saturation dyes) to laser microdissection and two-dimensional difference gel electrophoresis (2D-DIGE) for cancer proteomics. Nat Protoc 2007;1:2940–56.
[12] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[13] Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 2011;39:D691–7.
[14] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.
[15] Xie LQ, Zhao C, Cai SJ, Xu Y, Huang LY, Bian JS, et al. Novel proteomic strategy reveal combined alpha1 antitrypsin and cathepsin D as biomarkers for colorectal cancer early screening. J Proteome Res 2010;9:4701–9.
[16] Thakur D, Rejtar T, Wang D, Bones J, Cha S, Clodfelder-Miller B, et al. Microproteomic analysis of 10,000 laser captured microdissected breast tumor cells using short-range sodium dodecyl sulfate-polyacrylamide gel electrophoresis and porous layer open tubular liquid chromatography tandem mass spectrometry. J Chromatogr A 2011;1218:8168–74.
[17] Zhang Y, Xu B, Liu Y, Yao H, Lu N, Li B, et al. The ovarian cancer-derived secretory/releasing proteome: a repertoire of tumor markers. Proteomics 2012;12:1883–91.
[18] Pernemalm M, Orre LM, Lengqvist J, Wikstrom P, Lewensohn R, Lehtio J. Evaluation of three principally different intact protein prefractionation methods for plasma biomarker discovery. J Proteome Res 2008;7:2712–22.
[19] Jafari M, Primo V, Smejkal GB, Moskovets EV, Kuo WP, Ivanov AR. Comparison of in-gel protein separation techniques commonly used for fractionation in mass spectrometry-based proteomic profiling. Electrophoresis 2012;33:2516–26.
[20] Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 1999;17:994–9.

[21] Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 2004;3:1154–69.
[22] Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 2002;1:376–86.
[23] Li Z, Adams RM, Chourey K, Hurst GB, Hettich RL, Pan C. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. J Proteome Res 2012;11:1582–90.
[24] Merl J, Ueffing M, Hauck SM, von Toerne C. Direct comparison of MS-based label-free and SILAC quantitative proteome profiling strategies in primary retinal Muller cells. Proteomics 2012;12:1902–11.
[25] Hood JD, Cheresh DA. Role of integrins in cell invasion and migration. Nat Rev Cancer 2002;2:91–100.
[26] Filardo EJ, Brooks PC, Deming SL, Damsky C, Cheresh DA. Requirement of the NPXY motif in the integrin beta 3 subunit cytoplasmic tail for melanoma cell migration in vitro and in vivo. J Cell Biol 1995;130:441–50.
[27] Galliher AJ, Schiemann WP. Beta3 integrin and Src facilitate transforming growth factor-beta mediated induction of epithelial-mesenchymal transition in mammary epithelial cells. Breast Cancer Res 2006;8:R42.
[28] Lei Y, Huang K, Gao C, Lau QC, Pan H, Xie K, et al. Proteomics identification of ITGB3 as a key regulator in reactive oxygen species-induced migration and invasion of colorectal cancer cells. Mol Cell Proteomics 2011;10. M110.005397.
[29] Liapis H, Flath A, Kitazawa S. Integrin alpha V beta 3 expression by bone-residing breast cancer metastases. Diagn Mol Pathol 1996;5:127–35.
[30] Chen J, Zhang J, Zhao Y, Li J, Fu M. Integrin beta3 down-regulates invasive features of ovarian cancer cells in SKOV3 cell subclones. J Cancer Res Clin Oncol 2009;135:909–17.
[31] Partheen K, Levan K, Osterberg L, Claesson I, Fallenius G, Sundfeldt K, et al. Four potential biomarkers as prognostic factors in stage III serous ovarian adenocarcinomas. Int J Cancer 2008;123:2130–7.
[32] Partheen K, Levan K, Osterberg L, Claesson I, Sundfeldt K, Horvath G. External validation suggests Integrin beta 3 as prognostic biomarker in serous ovarian adenocarcinomas. BMC Cancer 2009;9:336.
[33] Wu Y, Zuo J, Ji G, Saiyin H, Liu X, Yin F, et al. Proapoptotic function of integrin beta(3) in human hepatocellular carcinoma cells. Clin Cancer Res 2009;15:60–9.
[34] Chu YQ, Ye ZY, Tao HQ, Wang YY, Zhao ZS. Relationship between cell adhesion molecules expression and the biological behavior of gastric carcinoma. World J Gastroenterol 2008;14:1990–6.
[35] Stupack DG, Puente XS, Boutsaboualoy S, Storgard CM, Cheresh DA. Apoptosis of adherent cells by recruitment of caspase-8 to unligated integrins. J Cell Biol 2001;155: 459–70.
[36] van der Flier A, Sonnenberg A. Function and interactions of integrins. Cell Tissue Res 2001;305:285–98.
[37] Chen YR, Juan HF, Huang HC, Huang HH, Lee YJ, Liao MY, et al. Quantitative proteomic and genomic profiling reveals metastasis-related protein expression patterns in gastric cancer cells. J Proteome Res 2006;5:2727–42.
[38] Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature 2011;473:337–42.

FEBS Letters 588 (2014) 1630–1636

ELSEVIER

FEBS Letters

journal homepage: www.FEBSLetters.org

1964–2014

7SK small nuclear ribonucleoprotein complex is recruited to the HIV-1 promoter via short viral transcripts

Taketoshi Mizutani [a,b,c,*], Aya Ishizaka [a,b], Yutaka Suzuki [d], Hideo Iba [a]

[a] Division of Host-Parasite Interaction, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

[b] Laboratory of Basic Science, Institute of Microbial Chemistry;BIKAKEN, 3-14-23 Kamiosaki, Shinagawa-ku, Tokyo 141-0021, Japan

[c] RNA and Biofunctions, Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency (JST), 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

[d] Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan

ARTICLE INFO

Article history:
Received 9 December 2013
Revised 13 January 2014
Accepted 14 January 2014
Available online 6 March 2014

Edited by Ivan Sadowski

Keywords:
Virology
Viral transcription
RNA virus
Chromatin regulation
Molecular biology

ABSTRACT

In this study, we demonstrate that the 7SK small nuclear ribonucleoprotein (snRNP) complex is recruited to the HIV-1 promoter via newly-synthesized HIV-1 nascent transcripts (short transcripts) in an hnRNP A1-dependent manner and negatively regulates viral transcript elongation. Our deep-sequence analysis showed these short transcripts were mainly arrested at approximately +50 to +70 nucleotides from the transcriptional start site in the U1 cells, an HIV-1 latent model. TNF-α treatment promptly disrupted the 7SK snRNP complex on the nascent transcripts and viral elongated transcripts were increased. This report provides insight into how 7SK snRNP complex is recruited to HIV-1 promoter in the absence of Tat.

1. Introduction

Under antiretroviral therapy (ART), latent infections of HIV-1 are often caused by repression of transcriptional initiation of the viral genome and elongation of nascent transcripts. A useful clue in the study of latent infection is the presence of abortive transcripts, approximately 60 nucleotides long, in resting T cells of patients undergoing ART, even though latently infected T cells may not produce virions and are difficult to distinguish from uninfected cells [1,2]. Generally, these observations are not viral-specific, as genome-wide studies have suggested that cellular RNA polymerase II (RNAPII) generates many abortive transcripts [3,4].

Transcription of the HIV-1 provirus is characterized by early Tat-independent and late Tat-dependent phases. In the early Tat-independent phase, HIV-1 transcription depends upon the interaction of host transcription factors with cis-regulatory DNA elements within the viral 5′ long terminal repeat (LTR) as well as the assembly of the transcription apparatus, including RNAPII, on these sequences. In the steady state, Tat-independent HIV-1 basal promoter activity is weak, because some host factors, which are reported to function as negative regulators of viral LTRs, are known to restrict HIV-1 basal promoter activity [5]. One of these host factors, the 7SK small nuclear ribonucleoprotein complex (snRNP), functions as a negative regulator of HIV-1 transcription by interacting with the positive elongation factor b (p-TEFb), which is composed of cyclin dependent kinase 9 (CDK9) and Cyclin T1 or Cyclin T2, to mask RNAPII-directed kinase activity. Main components of the 7SK snRNP complex are 7SK snRNA, hexamethylene bis-acetamide inducible 1 (HEXIM1), La-related protein (LARP7) and methylphosphate capping enzyme (MePCE). LARP7 is a crucial component of the 7SK snRNP complex and acts as an integral molecule in 7SK-mediated negative regulation [6,7]. However the viral accessory protein, Tat interacts with viral promoter proximal transactivation responsive element (TAR) RNA element and

Abbreviations: ART, antiretroviral therapy; RNAPII, RNA polymerase II; LTR, long terminal repeat; TAR, transactivation responsive element; snRNP, small nuclear ribonucleoprotein complex; p-TEFb, positive elongation factor b; CDK9, cyclin dependent kinase 9; LARP7, La-related protein; HEXIM1, hexamethylene bis-acetamide inducible 1; MePCE, methylphosphate capping enzyme; hnRNP A1, heterogeneous nuclear ribonucleoprotein A1

* Corresponding author at: Laboratory of Basic Science, Institute of Microbial Chemistry, 3-14-23 Kamiosaki, Shinagawa-ku, Tokyo 141-0021, Japan. Fax: +81 3 3441 7589.

E-mail address: mizutanit@bikaken.or.jp (T. Mizutani).

http://dx.doi.org/10.1016/j.febslet.2014.01.067

recruits p-TEFb from the nucleoplasm or 7SK snRNP complex to promote elongation of RNAPII. After loss of p-TEFb, the 7SK snRNP complex is known to release HEXIM1 by changing its conformation [8]. A previous report showed that 7SK snRNP complex, with or without Tat, are recruited to the HIV-1 *cis* regulatory enhancer element (SP1) before transcription initiation, although it is unclear how HIV-1 transcription is selectively negative regulated by the 7SK snRNP complex [9].

Heterogeneous nuclear ribonucleoprotein A1 (HnRNP) A1 is a known as an RNA-binding protein that associates with pre-mRNA, and functions in viral and host mRNA splicing or metabolism, including the splicing machinery involved in Tat expression in HIV-1 [10,11]. HnRNP A1 shuttles from the nucleus to cytoplasm and is involved in HIV-1 mRNA transport [12]. HnRNP A1 also plays a role in stabilizing the 7SK snRNP complex through interacting with 7SK snRNA lacking p-TEFb and HEXIM1 [8].

In this study, we report that hnRNP A1 facilitates recruitment of the 7SK snRNP complex to the HIV-1 promoter region through short viral transcripts generated by promoter-paused RNAPII.

2. Materials and methods

2.1. Plasmid construction and cell culture

A 0.7-kb PCR fragment of the 5′ LTR region of the NL43 HIV-1 molecular clone was inserted into a pWLG plasmid [13] to generate pHIV-LTR-GFP. The region from the 5′ LTR to the splicing acceptor region (0.9 kb) of pHIV-LTR-GFP was inserted into the *ClaI-BamHI* minimal promoter region of the pNF-κB-MinP-Luc plasmid to generate pLTR-Luc-pA (HLpA6), which has been previously described in detail [14]. The transiently transfected shRNA expressing plasmid (pmU6) produces shRNAs from the mouse U6 RNA polymerase III promoter. Control shGFP and sh-hnRNP A1 transduced U1 derived cells were generated by stably infecting with pSSSP (SIN type-shRNA expressing SV40-puro) retrovirus vectors and puromycin drug selection for 2 weeks. The shRNA sequence used in this study are listed in Supplemental Table S1.

2.2. Quantitation of HIV-1 viral production by TNF-α stimulation

U1-derived cells (1×10^6/24-well plate) were stimulated by TNF-α (10 ng/ml; R&D Systems), and culture supernatants were then collected by centrifugation at 0, 0.5, 1, 3, 6, 12, or 24 h after stimulation. Viral transcript and particle production were monitored by qRT-PCR and HIV-1 p24 antigen ELISA (ZeptoMetrix) according to the manufacturer's instructions.

2.3. Antibodies

Antibodies used in the experiments were as follows: HEXIM1 [ab25388] (Abcam); CDK9 [C12F7]; hnRNP A1 [R196] (Cell Signaling Technology); LARP7 [A303-723A] (Bethyl Laboratories); MePCE [14917-1-AP] (proteintech); CyclinT1 [SC10750]; GAPDH [SC25778] (Santa Cruz) (BD Transduction Laboratories); β-actin [013-24553] (Wako).

2.4. Chromatin immunoprecipitation (ChIP) assay

Cells were cross-linked with 1% formaldehyde. The lysates were then sonicated on ice to shear the DNA into fragments with an average length of less than 0.5 kb by using ELESTAIN035SD (ELECON Science. Corp). The lysates were incubated overnight on a rotating platform at 4 °C with the respective antibodies (5 mg each), which were previously bound to Dynabeads Protein G (Invitrogen). After washing, the DNA was purified and quantified according to the same protocol described in the RT-PCR section.

2.5. RNA preparation and quantitative RT-PCR

Total RNA was prepared from cells using the Isogen II isolation kit (Wako). For short transcripts, the small RNA fraction (<200 nt) was purified using the same kit. The isolated RNA was treated with Turbo DNase (Ambion) in accordance with the manufacturer's instructions. For long transcripts (>200 nt), cDNA was synthesized from total RNA by first-strand cDNA synthesis using the PrimeScript RT Master Mix (Takara Bio). Quantitative PCR was performed with a 7300 Real-Time PCR System (Applied Biosystems) using Premix Ex Taq (Probe qPCR) or SYBR Premix Ex Taq (Takara Bio). For short transcripts, cDNA was synthesized using the miScript Reverse Transcription Kit (Qiagen). Quantitative PCR was performed with a 7300 Real-Time PCR System (Applied Biosystems) using Premix Ex Taq (Probe qPCR; Takara Bio). The elongated transcript was amplified as the region between the 5′ LTR and splicing donor site. Threshold values (*Ct*) were calculated, and all reactions were run in triplicate. The specific primer pairs and probes used in this study are listed in Supplemental Table S1.

2.6. RNA immunoprecipitation (RIP) assay

RIP assays were performed and modified using the RiboCluster Profiler™/RIP-Assay Kit (MBL), according to the manufacturer's instructions. RNA was purified from the precipitates with Isogen II (Nippon Gene) reagent. These isolated RNAs were treated with Turbo DNase (Ambion) in accordance with the manufacturer's instructions. The cDNA was synthesized and quantified according to the same protocol described in the RT-PCR section.

2.7. Deep-sequence analysis of viral short transcripts

We prepared a small RNA fraction (<200 nt) from U1 cells and amplified viral short transcripts. Deep sequencing was performed using the Illumina GAIIx genome analyzer (Illumina) according to the manufacturer's instructions. Generated sequences were mapped to the reference HIV proviral genomic sequence (HIV-NL43; GenBank accession No. M19921-2). Sequence data that met the criteria of a quality value greater than 30 (100000 sequences) were analyzed. Only sequences that perfectly matched the first 20 nt (GGU CUC UCU GGU CCA UAG GA) of the reference HIV genomic sequence were used, and we then analyzed the following sequence of the short transcripts (total 82000 reads). All sequence data are deposited at the DNA Databank of Japan (DDBJ) (http://www.ddbj.nig.ac.jp) (accession number DRA000547).

2.8. Statistical analyses

We performed *t* test. All statistical tests were two-sided. We considered *P* values less than 0.05 to be statistically significant.

3. Results

3.1. HnRNP A1-knockdown releases suppression of HIV production in U1 cells

To elucidate the role of hnRNP A1 in the transcriptional regulation of HIV-1 proviral expression, retroviral vectors mediating either hnRNP A1 shRNA or control (GFP) shRNA expression vectors were stably transduced into U1 cells. This cell line harbors 2 copies of intact latent HIV-1 with defective Tat genes, and are often used as a cell model of latent HIV-1-infection [15]. The expression level of hnRNP A1 was reduced in U1 cells transduced with the hnRNP A1 shRNA vector (Fig. 1A). Interestingly, we found that, when hnRNP A1 was knocked down, albeit only partially, the basal viral

transcription level was increased 1.4-fold in comparison with the control (Fig. 1B and C). Since it has been shown that viral transcriptional initiation and elongation is regulated by host factors, we thought that hnRNP A1 may play a role in regulation of HIV-1 transcriptional initiation or elongation.

After 96 h TNF-α stimulation, the level of elongated viral transcription was upregulated 8-fold in cells in which hnRNP A1 was knocked down relative to the control (Fig. 1D). Furthermore, a p24 ELISA demonstrated that, after TNF-α stimulation, production of HIV-1 particles from the hnRNP A1 shRNA-transduced U1 cells was significantly increased compared with that of U1 cells expressing control shRNA (Fig. 1E). In contrast some pTEFb-dependent cellular genes (Myc and Fos) were not affected by knockdown of hnRNP A1 in U1 cells (Fig. 1C) [16]. Together, these results indicated that hnRNP A1 functions as a negative regulator of proviral transcription in U1 cells.

3.2. Knockdown of hnRNP A1 releases LARP7 from the HIV-1 promoter

Previous research on HIV-1 transcription has demonstrated that the 7SK snRNP complex negatively regulates viral elongation. To elucidate whether this complex may be involved in hnRNP A1-

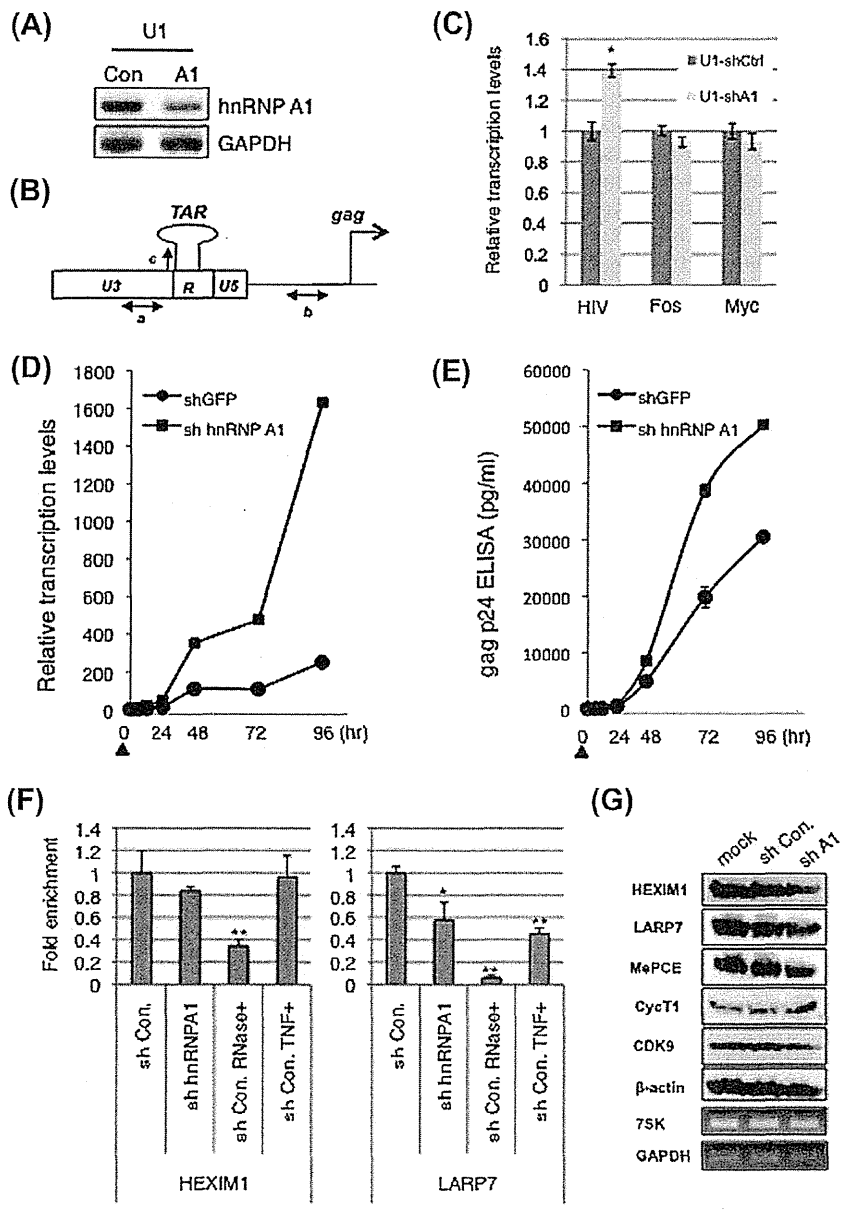


Fig. 1. hnRNP A1 negatively regulates viral transcript and viral production. (A) Expression of hnRNP A1 and GAPDH protein in U1 cells stably transduced with a retroviral vector expressing sh-hnRNP A1 (A1) or control GFP shRNA (Con). (B) Schematic representation of the primer positions in this study. Arrow "a" (ChIP), arrow "b" (qRT-PCR; elongated transcript) and arrow "c" (qRT-PCR; short transcript). (C) Expression levels of HIV-1 elongated transcripts, Myc, and Fos of U1 cells stably transduced with a retroviral vector expressing hnRNP A1 (A1) or control shRNA (Con). Asterisks denote values determined to be significantly different from sh-controls [$^{*}P < 0.05$]. (D and E) Kinetics of HIV-1 transcripts, viral production and p24 protein of U1 cells stably transduced with a retroviral vector expressing hnRNP A1 or control shRNA (GFP). Culture media and cell lysates were collected by centrifugation at 0, 3, 12, 24, 48, 72, or 96 h after TNF-α (10 ng/mL) stimulation (arrow head). Viral mRNA was quantified by qRT-PCR (D), and p24 gag antigen in supernatants was detected by ELISA in order to quantify HIV-1 particles (E). (F) ChIP assay of relative HEXIM1 and LARP7 abundance around the HIV-1 promoter site in hnRNP A1 shRNA- or control (GFP) shRNA-transduced U1 cells. RNase treatment was performed before immunoprecipitation in RNase + control (GFP) shRNA-transduced U1 cells. TNF-α (10 ng/mL) stimulation (1 h) was performed in TNF-α + control U1 shRNA-transduced cells. (G) Expression of 7SK snRNPs, control protein (β-actin) and control RNA (GAPDH) in U1 (mock) and U1 derived cells which are stably infected with shRNA against GFP (shCon) or hnRNP A1 (shA1).

dependent viral-selective transcriptional repression, ChIP assays were performed, using lysates of U1 cells stably expressing control or hnRNP A1 shRNA (Fig. 1F). HEXIM1 and LARP7 were recruited to the HIV-1 LTR in control shRNA-transduced U1 cells. Interestingly, LARP7, but not HEXIM1, was depleted from the HIV-1 promoter in hnRNP A1-knockdown cells, but expression of the 7SK snRNP components (HEXIM1, LARP7, MePCE, CycT1, CDK9 and 7SK RNA) was not changed in U1 cells (Fig. 1G). After TNF-α treatment, recruitment of LARP7 to the HIV-1 LTR was reduced even further, although recruitment of HEXIM1 remained unchanged (Fig. 1F). These results implied that LARP7 is most likely recruited to the HIV-1 promoter in the presence of hnRNP A1 and is released from the LTR after TNF-α treatment in U1 cells.

To elucidate the effect of negative regulation by hnRNP A1 on viral transcription in greater detail, 293FT-HLpA6 cells stably expressing firefly luciferase under the control of the HIV LTR promoter were generated. We then performed ChIP assays, using lysates of the 293FT-HLpA6 cells transiently expressing control or hnRNP A1 shRNA (Fig. 2A). With knockdown of hnRNP A1, elongated viral transcription from the LTR was increased 1.8-fold relative to the control (Fig. 2B). Also LARP7, but not HEXIM1, was depleted from the HIV-1 promoter in 293FT-HLpA6 cells (Figs. 1B and 2C). Since LARP7 is an integral molecule of the 7SK snRNP complex and maintains pTEF-b in an inactive state in 7SK snRNP complex [6,7], LARP7 dissociation was examined to determine whether it affected HIV-1 transcriptional elongation. After transduction with siRNA directed against LARP7 (Fig. 2D), HIV-1 transcriptional elongation was increased 1.6-fold in parental 293FT-HLpA6 cells relative to the control (Fig. 2E). Together, these results indicated that the recruitment of LARP7 to the LTR by hnRNP A1 is one of the causes for the observed negative regulation of viral elongation.

To facilitate further examination of the effect of transfection with hnRNP A1 shRNA on HIV-1 LTR promoter transcription levels, we performed limiting dilution analysis and selected 10 luciferase-expressing cellular clones from this parental reporter cell population. With knockdown of hnRNP A1, the changes in elongated viral transcription from the LTR in these clones could be divided into distinct subgroups based on the levels of viral transcription (Fig. 2F and G): [1] up-regulated (>2-fold), [2] marginally up-regulated (<2-fold), and [3] unchanged (<1.2-fold). Knockdown of hnRNP A1 affected viral transcription in one-third of the clones; we found that the basal transcriptional levels of these up-regulated clones were lower than those of clones in the unchanged category (Fig. 2H). These results suggested that hnRNP A1 is involved in the restriction of viral transcription and its effect is closely related to the viral integration site.

3.3. HEXIM1 interacts with the HIV-1 trans-acting responsive (TAR) RNA stem-loop in vivo

Next, we investigated the mechanism by which HEXIM1 and LARP7 are recruited to the HIV-1 LTR. Previous studies have shown that HEXIM1 binds tightly to HIV-1 TAR RNA and recruits and inhibits p-TEFb activity *in vitro* [17,18]. We observed that the recruitment of HEXIM1 and LARP7 to the HIV-1 LTR in control cells was eliminated by treatment with RNase before immunoprecipitation (Fig. 1F). Although this observation was possibly due to cleavage of 7SK RNA by RNase treatment, it also suggested that the 7SK snRNP complex is likely to be present near the HIV-1 promoter, because of an interaction with the viral TAR RNA stem-loops. In U1 cells, there are few elongated transcripts, and only abortive short transcripts (about 60 nt long) are detected [2]. To ascertain whether interaction occurs between the 7SK snRNP complex and the TAR RNA stem-loop region, we identified the paused or arrested sites of short viral transcripts in U1 cells by using deep-sequence analysis. HIV-1 abortive short transcripts have been suggested to pause at +62U, and stop at +65U and +66U from the TSS, based on *in vitro* transcription assays [19]. A small-RNA fraction (<200 nucleotides) was prepared from U1 cells, and the short viral transcripts were amplified. The short transcripts were mainly arrested at approximately +50 to +70 nucleotides from the TSS, but no "hot spot" of transcriptional pause sites was observed at the nucleotide level (Fig. 3). We next performed an RNA immunoprecipitation (RIP) assay by transiently transducing 293FT cells with a 62 nt long TAR RNA stem-loop expressing plasmid (pmU6-ST) and using antibodies directed against HEXIM1 (Fig. 4A and B). As predicted, the exogenous TAR RNA stem-loop was efficiently precipitated with endogenous HEXIM1, as well as with LARP7 and TAR RNA-binding protein (TRBP; as a positive control; Fig. 4C). These data indicated the possibility that the 7SK snRNP complex is recruited to the HIV-1 LTR via interaction with the TAR RNA stem-loop *in vivo*.

3.4. The 7SK snRNP complex is released from short transcripts after TNF-α stimulation

To examine whether the 7SK snRNP complex is recruited to endogenous nascent early transcripts and released from transcripts after TNF-α stimulation, the production kinetics of the abortive short transcripts were compared with those of elongated viral transcripts after TNF-α stimulation in U1 cells. The production of short transcripts increased up to 3 h after TNF-α stimulation and then decreased moderately (Fig. 4D). In contrast, elongated viral transcripts gradually increased within the initial 6 h after stimulation, and rapidly increased after 6 h. These results imply that a dynamic shift in viral transcription occurred over the 3 h period from the commencement of transcription.

Next, RIP assays were performed on U1 cell extracts before and 3 h after TNF-α treatment (Fig. 4E). Nascent short viral transcripts were efficiently precipitated with the endogenous 7SK snRNP negative-regulatory transcriptional complex (inhibitory 7SK snRNP complex, consisting of HEXIM1, MePCE, LARP7, and cyclin T1) prior to TNF-α treatment. LARP7 and MePCE were released from short viral transcripts after TNF-α treatment, but the other components were not. These observations indicated that the 7SK snRNP complex is promptly recruited to the nascent viral transcripts and TNF-α treatment disrupted the 7SK snRNP complex on the early viral transcripts. Our observations were congruent with previous work that indicated that exogenous stimulation disrupted the inhibitory 7SK snRNP complex [20]. Interestingly, HEXIM1 interaction with short viral transcripts increased after TNF-α treatment (Fig. 4E). Taken together with the ChIP assay results showing that HEXIM1 binds to the HIV-1 promoter throughout TNF-α treatment (Fig. 1F), these results suggested that HEXIM1 is constantly present at the HIV promoter via interaction with viral short transcripts in U1 cells.

4. Discussion

Since hnRNP A1 has multiple functions in cellular and viral RNA splicing and metabolism [10,21], knockdown of hnRNP A1 would influence the expression of cellular factors affecting the regulation of 7SK snRNP complex recruitment to the HIV-1 promoter region. Notably, the inhibitory 7SK snRNP complex is disrupted on the HIV-1 short viral transcripts at 3 h after TNF-α stimulation (Fig. 4E). In agreement with these observations, knockdown of LARP7 by RNA interference increased the number of elongated viral transcripts (Fig. 2E). These results showed that LARP7 is one of the molecules involved in negative regulation of viral expression

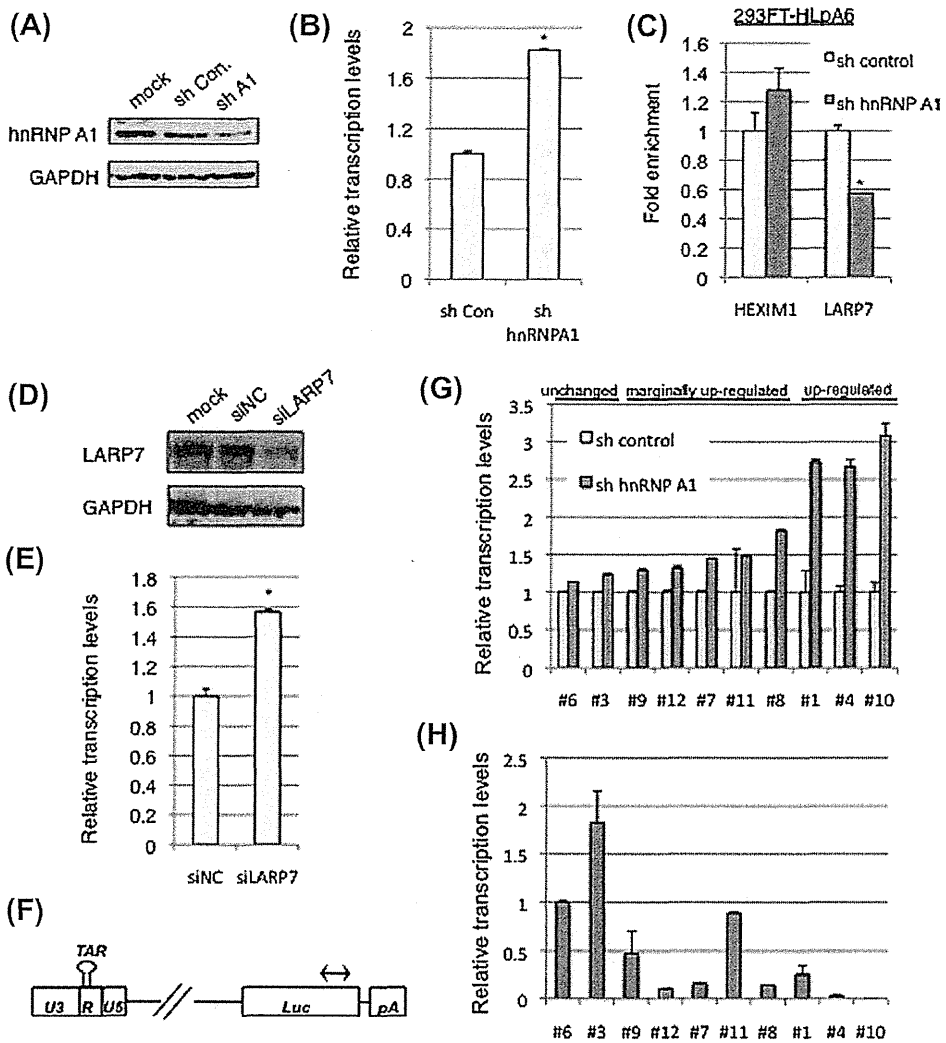


Fig. 2. LARP7 is recruited to the HIV-1 promoter in an hnRNP A1-dependent manner. (A) Expression of hnRNP A1 protein and GAPDH protein (loading control) in the HLpA6-293FT cells after 48-h transient transfection of with LacZ (Con) or hnRNP A1 (A1) shRNA-expressing vector. (B) Expression of HIV-1 elongated mRNA in HLpA6-293FT cells transiently transfected with hnRNP A1 shRNA (sh hnRNP A1) or LacZ shRNA (sh Con). (C) ChIP assay of relative LARP7 abundance around the HIV-1 promoter site after 48-h transient transfection of HLpA6-293FT cells with hnRNP A1 shRNA (sh hnRNP A1) or LacZ shRNA (sh control). (D and E) Expression of LARP7 protein and GAPDH protein (loading control) (D) and HIV-1 elongated mRNA (E) in HLpA6-293FT cells transiently transfected with LARP7 siRNA (siLARP7) or vimentin siRNA (siNC). (F and G) qRT-PCR analysis of HIV-1 elongated mRNA after 48-h transient transfection of 10 HLpA6-293FT clones with hnRNP A1 shRNA or LacZ shRNA (control). The region expanded in qRT-PCR is denoted with an arrow in (F). *Y*-Axis represents "Relative transcription levels" based on the value of control LacZ shRNA transduced HLpA6-293FT clones (G). (H) Expression levels of HIV-1 transcripts of HLpA6-293FT clones. *Y*-Axis represents "Relative transcription levels" based on the value of HLpA6-293FT#6. Asterisks denote values determined to be significantly different from si- or sh-controls [$^{*}P < 0.05$, $^{**}P < 0.01$].

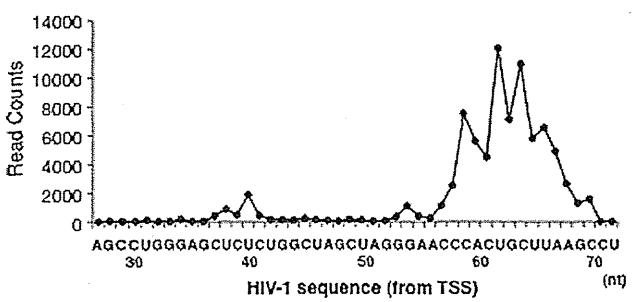


Fig. 3. Pausing sites of HIV-1 short transcripts in U1 cells by deep-sequence analysis. The 3′ end of each sequence is mapped to the HIV-1 mRNA (HIVNL43; GenBank accession No. M19921-2), and the number of reads at each sequence is plotted on the *Y*-axis. TSS; transcriptional start site.

by hnRNP A1, although it is necessary to elucidate how LARP7 is recruited to the short transcripts in an hnRNP A1-dependent manner.

Our data demonstrates that the 7SK snRNP complex can interact with newly synthesized TAR RNA. In RIP and ChIP assays, HEXIM1 maintains interaction with TAR RNA after TNF-α treatment and is permanently present at the LTR even when there is no nascent viral RNA (Fig. 1F). LARP7 and MePCE are recruited to the viral TAR RNA; however, these proteins dissociate from short transcripts after TNF-α treatment and LARP7 moves away from the area of the LTR (Figs. 1F and 4E). These results indicate that some of the components of the 7SK snRNP complex could be recruited independently to TAR RNA. Previous work has indicated that exogenous stimulation disrupts the inhibitory 7SK snRNP complex [20]. In our results, it is possible that HEXIM1 and LARP7 assemble into the 7SK snRNP complex on the TAR RNA.

Recently, D'Orso et al. reported that the inhibitory 7SK snRNP complex is involved in transcriptional regulation by recruitment

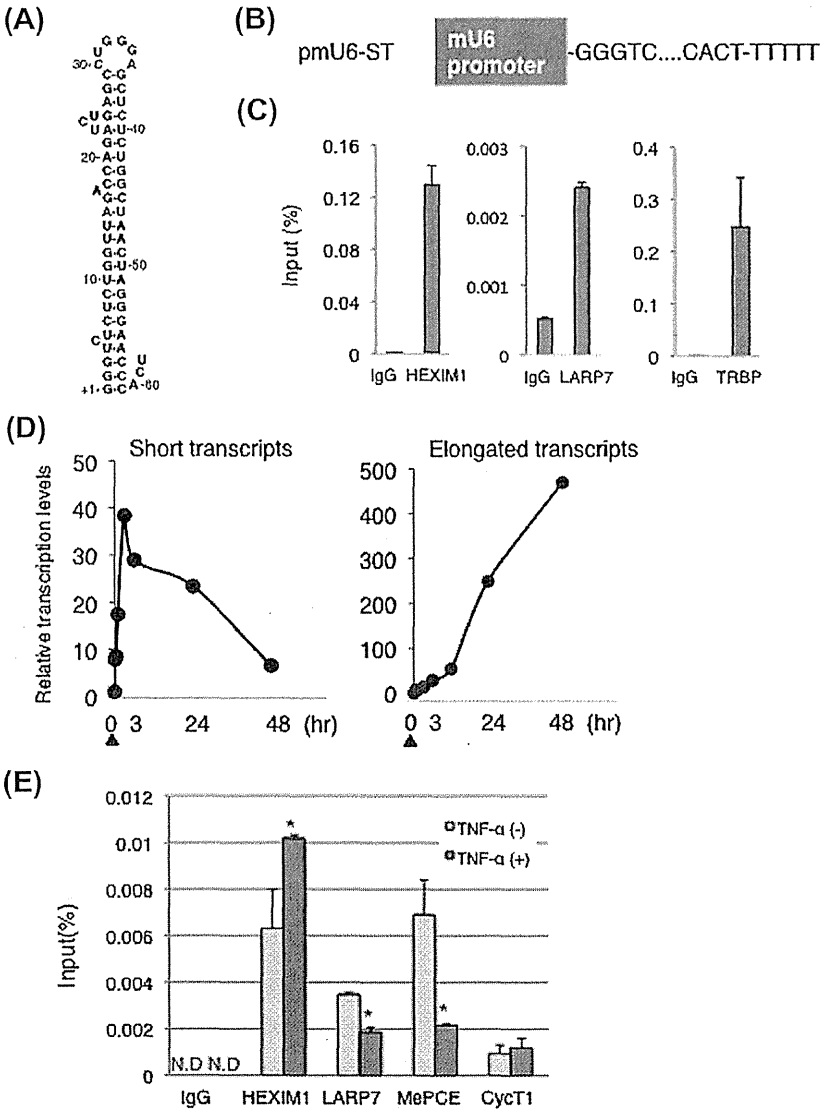


Fig. 4. 7SK snRNPs are promptly dissociated from viral short transcript by TNF-α stimulation. (A) A schematic representation of the 62 nt trans-acting responsive (TAR) stem region (nt +1 to +62) of the HIV-1 short transcript. (B) This figure pertains to the 62 nt of the TAR stem-loop region of HIV-1 transcripts driven by the mouse pol III-type U6 promoter. (C) pmU6-ST was transiently transfected into 293FT cells and, after 48 h, cells were analyzed by an RIP assay using anti-HEXIM1 antibodies, anti-LARP7 antibodies and anti-TRBP antibodies. Error bars indicate the S.D. ($n = 3$). (D) The kinetics of HIV-1 short transcripts and elongated viral transcripts at 0, 1/6, 1/2, 1, 3, 6, 24, and 48 h after TNF-α stimulation (arrow) in U1 cells. (E) RNA immunoprecipitation assay of relative HEXIM1, LARP7, MePCE and CycT1 abundance on HIV-1 viral short transcripts after TNF-α stimulation (3 h) in U1 cells. Asterisks denote values determined to be significantly different from TNF-α (−) [$^*P < 0.05$]. N.D means "not detected".

to the proviral LTR before TAR RNA is synthesized and can be involved in transcriptional regulation [9]. Prior to transcriptional initiation, the inhibitory p-TEFb-7SK snRNP complex with Tat is loaded into HIV-1 preinitiation complexes (PICs) in an SP1-dependent manner. In this model, upon transcriptional initiation, the newly synthesized TAR competes for Tat-P-TEFb with 7SK snRNP complex, which results in efficient viral elongation. Together, these observations suggest two independent pathways for the recruitment of 7SK snRNP complex to the HIV-1 promoter region. First, the 7SK snRNP complex with Tat is recruited prior to transcription via SP1 and PIC. Second, in our model, 7SK snRNP complex directly recruits TAR RNA immediately after being synthesized from paused RNAPII in the absence of Tat. The mechanism of transition from promoter-proximal pausing into transcriptional elongation with Tat has previously been demonstrated, whereas we have here uncovered a novel mechanism of transcriptional insufficiency of basal LTR activity in the absence of Tat. Considering both mechanisms together, it appears that the 7SK snRNP complex plays dual roles in transcriptional regulation of the HIV LTR: it is a positive regulator in the presence of Tat, and a negative regulator in the absence of Tat. These dual roles of 7SK snRNP complex may contribute to influencing the fate of transcriptional activation of the HIV LTR.

Conflict of interest

There is no conflict of interest.

Acknowledgments

We thank Eimi Horiuchi (Tokyo Univ.) for analyzing the short transcript sequence data. We thank Editage for providing editorial assistance. This research was supported by JST and PRESTO. TM is a recipient of PRESTO.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2014.01.067.

References

[1] Massanella, M., Martinez-Picado, J. and Blanco, J. (2013) Attacking the HIV reservoir from the immune and viral perspective. Current HIV/AIDS reports 10, 33–41.

[2] Adams, M., Sharmeen, L., Kimpton, J., Romeo, J.M., Garcia, J.V., Peterlin, B.M., Groudine, M. and Emerman, M. (1994) Cellular latency in human immunodeficiency virus-infected individuals with high CD4 levels can be detected by the presence of promoter-proximal transcripts. Proc. Nat. Acad. Sci. USA 91, 3862–3866.

[3] Core, L.J., Waterfall, J.J. and Lis, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322, 1845–1848.

[4] Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A. and Sharp, P.A. (2008) Divergent transcription from active promoters. Science 322, 1849–1851.

[5] Mbonye, U. and Karn, J. (2011) Control of HIV latency by epigenetic and non-epigenetic mechanisms. Curr. HIV Res. 9, 554–567.

[6] He, N., Jahchan, N.S., Hong, E., Li, Q., Bayfield, M.A., Maraia, R.J., Luo, K. and Zhou, Q. (2008) A La-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. Mol. Cell 29, 588–599.

[7] Markert, A., Grimm, M., Martinez, J., Wiesner, J., Meyerhans, A., Meyuhas, O., Sickmann, A. and Fischer, U. (2008) The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. EMBO Rep. 9, 569–575.

[8] Krueger, B.J., Varzavand, K., Cooper, J.J. and Price, D.H. (2010) The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK. PLoS ONE 5, e12335.

[9] D'Orso, I. and Frankel, A.D. (2010) RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. Nat. Struct. Mol. Biol. 17, 815–821.

[10] Chen, M. and Manley, J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat. Rev. Mol. Cell Biol. 10, 741–754.

[11] Okunola, H.L. and Krainer, A.R. (2009) Cooperative-binding and splicing-repressive properties of hnRNP A1. Mol. Cell. Biol. 29, 5620–5631.

[12] Monette, A., Ajamian, L., Lopez-Lastra, M. and Mouland, A.J. (2009) Human immunodeficiency virus type 1 (HIV-1) induces the cytoplasmic retention of heterogeneous nuclear ribonucleoprotein A1 by disrupting nuclear import: implications for HIV-1 gene expression. J. Biol. Chem. 284, 31350–31362.

[13] Mizutani, T., Ishizaka, A., Tomizawa, M., Okazaki, T., Yamamichi, N., Kawana-Tachikawa, A., Iwamoto, A. and Iba, H. (2009) Loss of the Brm-type SWI/SNF chromatin remodeling complex is a strong barrier to the Tat-independent transcriptional elongation of human immunodeficiency virus type 1 transcripts. J. Virol. 83, 11569–11580.

[14] Ishizaka, A., Mizutani, T., Kobayashi, K., Tando, T., Sakurai, K., Fujiwara, T. and Iba, H. (2012) Double plant homeodomain (PHD) finger proteins DPF3a and -3b are required as transcriptional co-activators in SWI/SNF complex-dependent activation of NF-kappaB RelA/p50 heterodimer. J. Biol. Chem. 287, 11924–11933.

[15] Emiliani, S., Fischle, W., Ott, M., Van Lint, C., Amella, C.A. and Verdin, E. (1998) Mutations in the tat gene are responsible for human immunodeficiency virus type 1 postintegration latency in the U1 cell line. J. Virol. 72, 1666–1670.

[16] Albert, T.K., Rigault, C., Eickhoff, J., Baumgart, K., Antrecht, C., Klebl, B., Mittler, G. and Meisterernst, M. (2014) Characterization of molecular and cellular functions of the cyclin-dependent kinase CDK9 using a novel specific inhibitor. Br. J. Pharmacol. 171, 55–68.

[17] Sedore, S.C., Byers, S.A., Biglione, S., Price, J.P., Maury, W.J. and Price, D.H. (2007) Manipulation of P-TEFb control machinery by HIV: recruitment of P-TEFb from the large form by Tat and binding of HEXIM1 to TAR. Nucleic Acids Res. 35, 4347–4358.

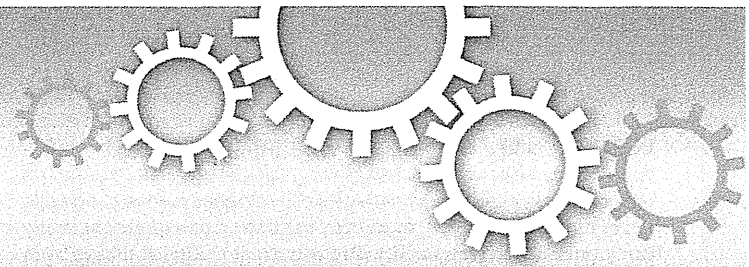
[18] Li, Q., Cooper, J.J., Altwerger, G.H., Feldkamp, M.D., Shea, M.A. and Price, D.H. (2007) HEXIM1 is a promiscuous double-stranded RNA-binding protein and interacts with RNAs in addition to 7SK in cultured cells. Nucleic Acids Res. 35, 2503–2512.

[19] Palangat, M., Meier, T.I., Keene, R.G. and Landick, R. (1998) Transcriptional pausing at +62 of the HIV-1 nascent RNA modulates formation of the TAR RNA structure. Mol. Cell 1, 1033–1042.

[20] Krueger, B.J., Jeronimo, C., Roy, B.B., Bouchard, A., Barrandon, C., Byers, S.A., Searcey, C.E., Cooper, J.J., Bensaude, O., Cohen, E.A., Coulombe, B. and Price, D.H. (2008) LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic Acids Res. 36, 2219–2229.

[21] He, Y. and Smith, R. (2009) Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. Cell. Mol. Life Sci.: CMLS 66, 1239–1256.

SCIENTIFIC REPORTS



OPEN

SUBJECT AREAS:
CHROMATIN REMODELLING
ONCOGENES
MIRNAS

Received
21 October 2014

Accepted
15 January 2015

Published
12 February 2015

Correspondence and requests for materials should be addressed to H.I. (iba@ims.u-tokyo.ac.jp)

The miR-199a/Brm/EGR1 axis is a determinant of anchorage-independent growth in epithelial tumor cell lines

Kazuyoshi Kobayashi[1], Kouhei Sakurai[1], Hiroaki Hiramatsu[1], Ken-ichi Inada[3], Kazuya Shiogama[3], Shinya Nakamura[1], Fumiko Suemasa[1], Kyosuke Kobayashi[1], Seiya Imoto[2], Takeshi Haraguchi[1], Hiroaki Ito[1], Aya Ishizaka[1], Yutaka Tsutsumi[3] & Hideo Iba[1]

[1]Division of Host-Parasite Interaction, Department of Microbiology and Immunology, University of Tokyo, Tokyo, Japan, [2]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan, [3]First Department of Pathology, Faculty of Medicine, Fujita Health University, Aichi, Japan.

In epithelial cells, miRNA-199a-5p/-3p and Brm, a catalytic subunit of the SWI/SNF complex were previously shown to form a double-negative feedback loop through EGR1, by which human cancer cell lines tend to fall into either of the steady states, types 1 [miR-199a(−)/Brm(+)/EGR1(−)] and 2 [miR-199a(+)/Brm (−)/EGR1(+)]. We show here, that type 2 cells, unlike type 1, failed to form colonies in soft agar, and that CD44, MET, CAV1 and CAV2 (miR-199a targets), all of which function as plasma membrane sensors and can co-localize in caveolae, are expressed specifically in type 1 cells. Single knockdown of any of them suppressed anchorage-independent growth of type 1 cells, indicating that the miR-199a/Brm/EGR1 axis is a determinant of anchorage-independent growth. Importantly, two coherent feedforward loops are integrated into this axis, supporting the robustness of type 1-specific gene expression and exemplifying how the miRNA-target gene relationship can be stably sustained in a variety of epithelial tumors.

Chromatin remodeling factors play vital roles in epigenetical regulation via genome-wide gene transcription[1]. On the other hand, microRNAs (miRNAs) are post-transcriptional regulatory molecules that are involved in diverse biological processes, including development, differentiation, and homeostasis[2]. Growing evidence indicates that the robustness of gene expression is often supported by coordinated transcriptional and miRNA-mediated regulatory networks[3,4]. In addition, improper use of these networks may lead to human diseases such as cancer. However, the interplay between chromatin remodeling factors and miRNA, as well as its biological outcome, is not fully understood in the context of gene regulatory networks common to a wide variety of cell lines.

The human SWI/SNF-A complex (also known as the BAF complex), a member of a family chromatin remodeling factors[5] composed of about 10 proteins, regulates gene transcription, either positively or negatively. The SWI/SNF complex contains a single molecule of either Brm or BRG1 as ATP-dependent catalytic subunits. Brm and BRG1 regulate target promoters that do not fully overlap and show clear differences in their biological activities[6–9]. This SWI/SNF complex interacts with various proteins, including transcriptional regulators, through many specific and varied associations with its several subunits. For example, the d4-family proteins DPF2 (REQ) and DPF3a/3b function as efficient adaptor proteins for RELB/p52[10] and RELA/p50[11] dimers to induce SWI/SNF-dependent NFκB target genes.

In terms of human cancers, we and other groups have reported that Brm is frequently undetectable in various cancer cell lines[12], and in primary tumors of the lung[13], stomach[14], and prostate[15]. We found in nuclear run-on transcription assays that a functional *Brm* gene was present and actively transcribed in all of the Brm-deficient cancer cell lines tested[12,16], indicating that Brm expression is largely suppressed by post-transcriptional gene silencing. Brm was later shown to be efficiently targeted by both miR-199a-5p and miR-199a-3p[17]. In addition, Brm acts as a potent negative regulator of endogenous *EGR1* gene expression. EGR1 activates the *miR-199a (2)* gene locus, which is mainly responsible for the biogenesis of mature miR-199a-5p and -3p in these cancer cell lines. Overall, these findings suggest that, in the cell lines examined, Brm and miR-199a form a robust double-negative feedback loop that includes EGR1[17]. By examining a panel of human cell lines that were derived from a wide variety of cancer tissues, we found that they tend to fall into either of the steady states, miR-199(−)/Brm(+)/

EGR1(−) cells and miR-199a(+)/Brm(−)/EGR1(+) cells[17], denoted hereafter as type 1 and type 2, respectively. These regulatory networks may explain why variable (either higher or lower) expression of miR-199a-5p/-3p[18] or EGR1[19] has been reported among many carcinomas when compared with the normal epithelial tissues from which they originated.

In the early stage of our current study, we noticed clear differences in the biological properties between type 1 and type 2 cells: all of the type 1 cell lines tested (8 lines), but no type 2 cell lines (4 lines), could grow in soft agar, providing us with an unprecedented opportunity to unravel the robust regulatory networks involved in anchorage-independent growth common to these cancer cell lines. Of course, the gene expression patterns of each cancer cell line would be expected to be largely cell line-specific and dependent on a wide variety of factors, including the originating tissue type, mutated genes, and pathological properties, such as the tumor stage. However, in our current study, we speculated that epithelial tumors would share regulatory networks that control their basic biological activities. In addition, we hypothesized that several genes would be specifically expressed in type 1 cancer cells, but not in type 2, and, further, that some of them would be crucial for their anchorage independency. Here, we have identified several genes specifically expressed in type 1 cells and show that single knockdown of some of these genes is sufficient to suppress the colony-forming activity of type 1 cells in soft agar. We further examined the underlying molecular mechanisms of the all-or-none regulation of these type 1-specific genes in the two cell types, leading to the identification of two coherent feedforward loops associated with the miR199a/Brm/EGR1 axis. We finally present evidence that these type-specific gene expression patterns can be recapitulated in tumor some lesions of non-small-cell lung carcinomas (NSCLCs).

Results

Type 1, but not type 2, cells grow efficiently in soft agar. For type 1 and type 2 cells, we chose 12 cell lines (8 for type 1 and 4 for type 2) originating from various human epithelial tumors (Fig. 1a). To ensure that we only examined cell lines originating from epithelial tumors, PA-1 (originating from tridermic teratocarcinoma), MDA-MB435 (recently discovered to originate from melanoma), and HEK-293FT (originating from human embryonic kidney) were removed from the panel of cancer cell lines previously used for categorization[17]. In addition, cell lines from one pancreatic (Panc-1) and three colon cancers (DLD-1, HT29, and HCT116) were added to the current analysis.

The results of a series of quantitative reverse transcription-polymerase chain reaction (RT-PCR) experiments (Fig. 1a) confirmed

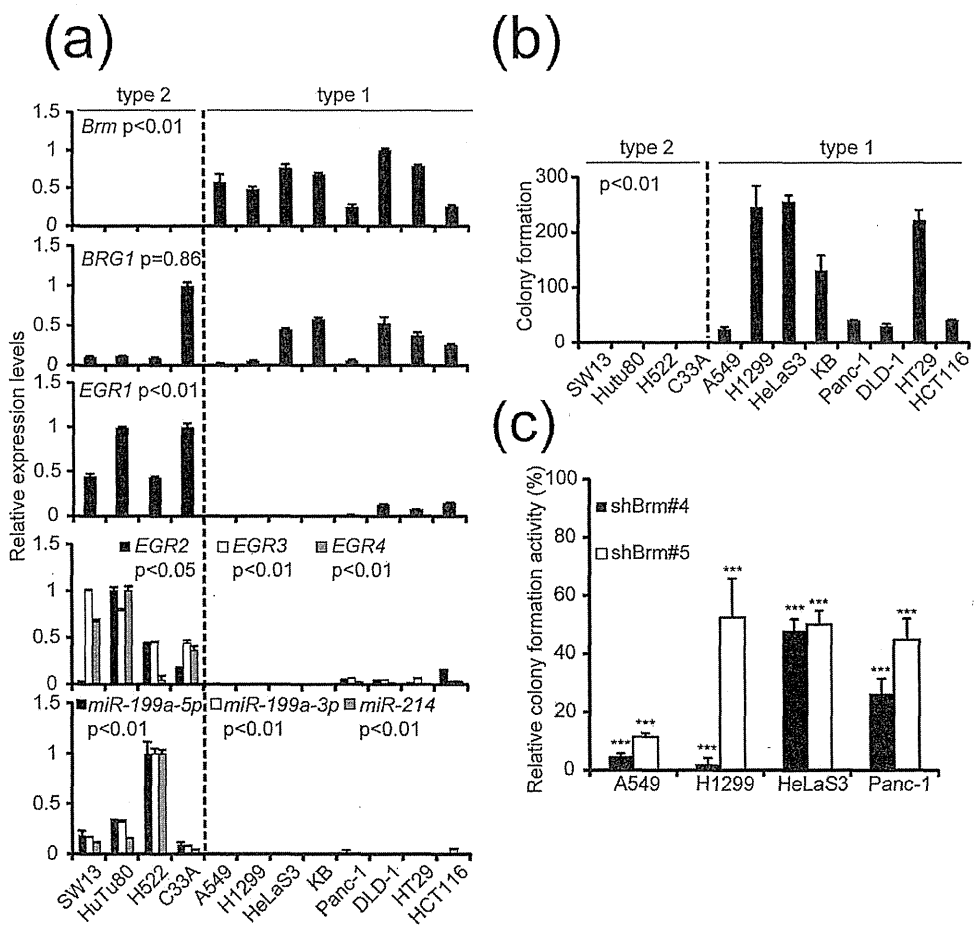


Figure 1 | Basic properties of 12 human cell lines originating from various epithelial tumors. (a) Relative expression levels of *Brm*, *BRG1*, *EGR1*, *EGR2*, *EGR3*, and *EGR4* mRNA and mature miR-199a-5p, -3p, and miR-214 were determined by quantitative RT-PCR. The data represent the means ± S.D. (n=3). (b) Numbers of colonies formed in soft agar by each cell line. Numbers of colonies (more than 150 μm in diameter) in soft agar were counted 21–28 days after 1,000 cells were seeded in 60-mm plates. The data represent the means ± S.D. ($n = 3$) In a–b, P values were determined using Mann-Whitney test. (c) Four type 1 cell lines transduced with retroviral vectors expressing shBrm (#4 or #5) or shCre#4 (negative control) were seeded as in b. Colony numbers of shBrm-expressing cells were compared with those of shCre#4-expressing cells and the ratio was shown as a percentage. The data represent the means ± S.D. ($n = 4$). Asterisks indicate P value, compared with those transduced with shCre#4. *P<0.05, **P<0.01, ***P<0.001.

our previous observations that epithelial tumor cell lines can be classified into two types according to the expression levels of Brm, EGR1, and miR-199a; type 1 cells specifically express *Brm* mRNA, whereas expressions of *EGR1* mRNA and miR-199a-5p and -3p, as well as miR-214, which is also generated from the *miR-199a (2)* gene locus, are restricted to type 2 cells. Notably, like *EGR1*, *EGR2*, *EGR3*, and *EGR4*, which are the other members of EGR family gene and which recognize the same DNA sequence[20], were shown to be type 2-specific (Fig. 1a) by the parallel analysis. This might indicate EGR2, EGR3, and EGR4 are involved in the miR-199a/Brm axis in a similar manner to EGR1.

After several preliminary comparative analyses between type 1 and type 2 cells, we noticed clear differences in terms of anchorage-independent growth. type 1 cells formed 25–300 colonies (more than 150 μm in diameter) in soft agar when 1,000 cells were seeded per 60-mm plate and kept for 21–28 days (Fig. 1b). None of the four type 2 cells formed clear colonies in the same conditions. Notably, all of the type 2 cancer cell lines tested—SW13[21], HuTu80[22], H522[23] and C33A[24]—have shown clear tumor-forming activity in mouse xenograft models.

To test whether the anchorage-independent growth of type 1 cells requires Brm, we performed Brm knockdown experiments in several type 1 cells—A549, H1229, HeLaS3, and Panc-1—using a set of retroviral vectors containing shBrm (#4 and #5). We confirmed that all shBrm vectors significantly suppressed the levels of *Brm* mRNA and its product (see below) and that cells transduced with these vectors reduced colony-forming activity in soft agar when compared with that of negative control cells transduced with the shCre#4 vector (Fig. 1c). These results reveal the pivotal role of Brm in anchorage-independent growth in type 1 cells. In several cases, strong suppression in colony formation in soft agar was observed by Brm knockdown, whereas the same culture grow normally when kept in monolayer culture (A549 cells expressing shBrm#5 and HeLaS3 and Panc-1 cells expressing either shBrm#4 or #5, Supplementary Fig. 1). In the case of H1299, however, we cannot exclude the possibility that reduction in the growth rate of monolayer culture partly contributed reduction in anchorage-independent growth. These findings provided us with an excellent opportunity to uncover the critical genes required for anchorage-independent growth of type 1 cells and suggested that these candidate genes would be expressed in a type 1-specific manner.

Several genes are preferentially expressed in type 1 cells. Whereas we know that suppression of the expression of a target protein by a certain miRNA is usually moderate and is not unconditionally retained in steady states, we selected the candidates of type 1-specific genes from various targets of miR-199a-5p (10 genes tested), miR-199a-3p (11 genes tested), and miR-214 (6 genes tested). These target genes were identified in previous reports or were predicted by target prediction algorithms as well as by our own analysis (Supplementary Table 1). Of 32 candidate genes tested by quantitative RT-PCR (Fig. 2a and Supplementary Figs. 2 and 3), *CAV1*[25] and *KRT80* (both miR-199a-5p target genes) and *CD44*[26], *MET*[27], and *CAV2*[28] (all miR-199a-3p targets) were expressed in most of the type 1 cells but in none of the type 2 cells (Fig. 2a). We also performed FACS analysis for CD44 and MET using

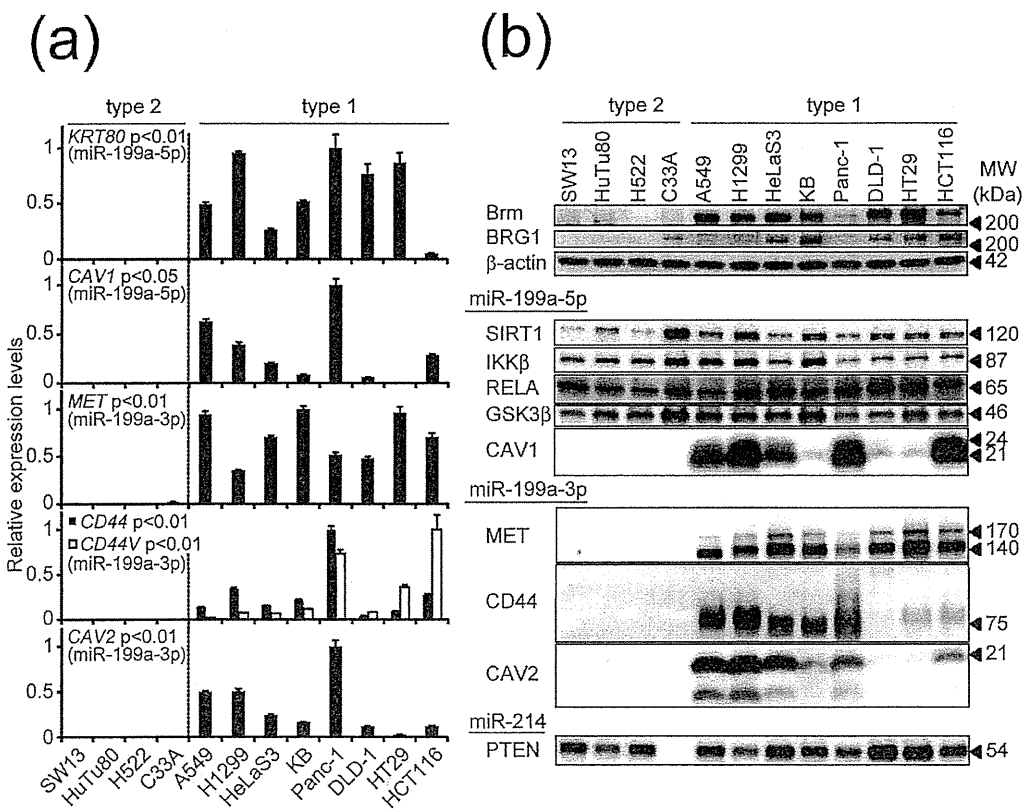


Figure 2 | Detection of five type 1-specific genes. (a) Relative expression levels of five type 1-specific genes—*CAV1*, *KRT80*, *MET*, *CD44* (all transcripts and variant types v8–v10), and *CAV2*—in 12 cancer cell lines were determined by quantitative RT-PCR. The data were normalized by taking the highest levels as 1.0. The data represent the means ± S.D. ($n = 3$). P values were determined using Mann-Whitney test. (b) Protein expression levels of Brm, BRG1, and four type 1-specific genes products as well as five type-independent gene products were analyzed by western blotting using 10 gels. The full-length blots were presented in the supplementary Figure 6. β-actin was used as the loading control for each gel. Analysis of KRT80 protein was not possible because of the lack of a specific antibody. Relative expression levels of each protein were quantified in Supplementary Figure 4.

H1299 cells (type 1), and found that the entire population showed high levels of both CD44 and MET (data not shorn). We thus designated them as type 1-specific genes. Interestingly, two other epithelial-type keratin genes[29,30], *KRT7*[31] (an miR-199a-3p target) and *KRT19* (an miR-199a-5p target), were not expressed in any of the type 2 cells but were expressed in some of the type 1 cells (Supplementary Fig. 2 and 3).

The protein expression profiles of most of these type 1-specific genes, as well as four genes whose mRNA was expressed in both types, were examined by western blotting (Fig. 2b) and relative amounts of each protein were quantified (Supplementary Fig. 4). The protein expression profiles were generally very similar to those of the corresponding mRNA (Fig. 2a). Mann-Whitney test further confirmed that these miR-199a target gene products (CD44, MET, CAV1 and CAV2) are specifically expressed in type 1 cells (Supplementary Fig. 4). KRT80 protein was not analyzed because of the lack of a specific antibody. PTEN protein, a well-known miR-214 target, was clearly expressed in all of the cell lines except C33A cells, because C33A have homozygous nonsense mutation of this gene. It is noteworthy that among these type 1-specific proteins, CD44, MET, CAV1 and CAV2 function as plasma membrane sensors and signaling platforms and can be colocalized in caveolae in several physiological conditions[32].

A single knockdown of CD44, MET, CAV1, or CAV2 is sufficient to suppress anchorage-independent growth of type 1 cell lines. To evaluate whether the type 1-specific genes identified above contribute to the anchorage-independent growth of this cell type, we developed pairs of shRNA constructs for *CD44*, *MET*, *CAV1*, and *CAV2* capable of efficiently suppressing their target gene products (Supplementary Fig. 5). Specific knockdown of KRT80 by short hairpin was not possible because there are too many conserved regions among the large *keratin* gene family paralogues. After A549, H1299, HeLaS3, and Panc-1 cells were transduced with the shRNA-expressing retroviral vectors, their colony-forming activity in soft agar was evaluated (Fig. 3). Single knockdown of *CD44*, *MET*, *CAV1*, or *CAV2* efficiently suppressed colony formation compared with negative control cells expressing shRNA for *Cre#4*. As an exception, the colony number of A549 cells expressing shCAV2#2 was slightly increased, probably due to off-target effects. Knockdown of *CD44*, *MET*, *CAV1*, and *CAV2* did not significantly affect cell growth. These results indicate that all of the four type 1-specific genes tested significantly contribute to the anchorage-independent growth of these type 1 cell lines.

All-or-none expression patterns of some type 1-specific genes in the cell line panel are supported by two coherent feedforward loops that associate with the axis. Given that miRNA usually suppresses its target protein in a modest manner, it might be somewhat unexpected that some miR-199a targets were regulated in an all-or-none manner between type 1 and type 2 cells (Fig. 2a,b). We speculated that this all-or-none phenomenon could be reflecting regulation by the molecular switch through miR-199a/Brm/EGR1 axis, where Brm and miR-199a expressions manifest a mutually exclusive pattern. Therefore, we first tested whether these type-1 specific genes, *CD44*, *MET*, *CAV1*, *CAV2*, and *KRT80* genes are under the positive control of the Brm-type SWI/SNF complex, for which type 1 cells are competent.

To test whether type 2 cells can induce type 1-specific genes when Brm is exogenously introduced, we transfected SW13 cells (type 2) with Brm expression plasmid or empty plasmid (EV1). In these experiments, some parallel cultures were cotransfected with the expression plasmids for representative NFκB dimers—RELA/p50 (canonical pathway) and RELB/p52 (noncanonical pathway)—or

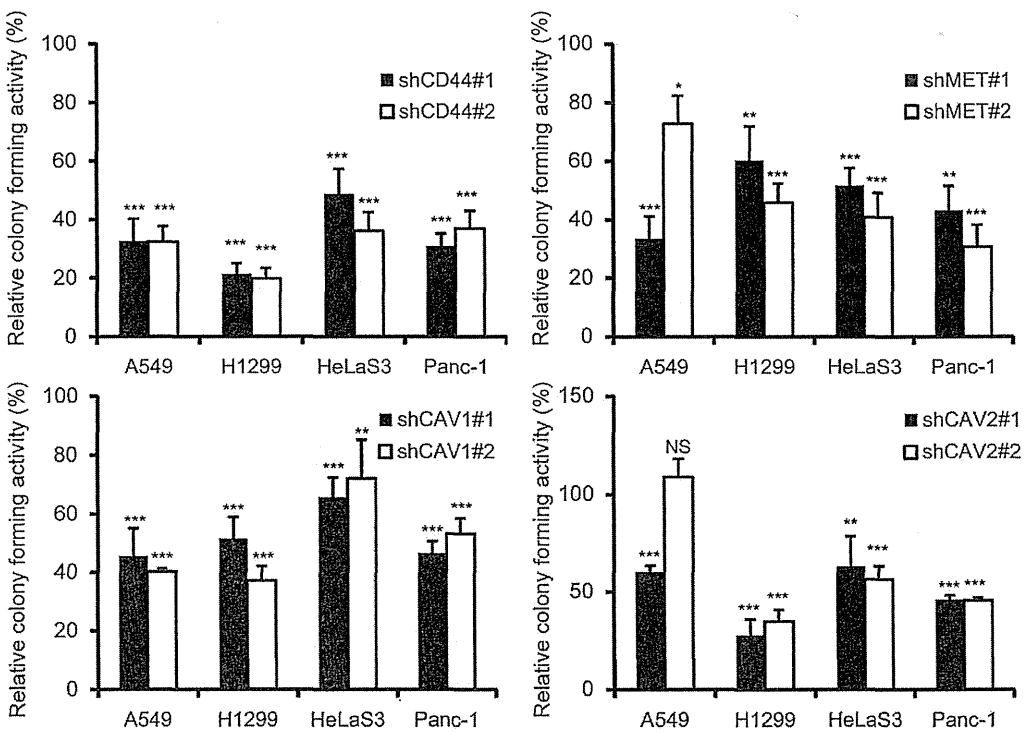


Figure 3 | Effect of single knockdown of several type 1-specific genes on anchorage-independent growth of four type1 cell lines. 300-1,000 cells of A549, H1299, HeLaS3 and Panc-1 transduced with retrovirus vectors expressing shCD44 (#1 or #2), shMET (#1 or #2), shCAV1 (#1 or #2), shCAV2 (#1 or #2) or sh Cre#4 (a negative control) were seeded in 60 mm plates. Colony numbers of these shRNA expressing cells were compared to those of shCre#4 expressing cells and the ratio was shown in percentage. The data represent the means ± S.D. ($n = 4$). Asterisks indicate P value, compared with those transduced with shCre#4. NS, not significant. *P<0.05, **P<0.01, ***P<0.001.

empty plasmid (EV2) to determine whether the activation is enhanced by NFκB dimers. As shown in Fig. 4a, *CD44*, *MET*, *CAV1*, and *KRT80* mRNA were induced by Brm, as judged by quantitative RT-PCR, although the Brm induction effect varied among the genes. *CD44* and *CAV1* mRNA levels were increased by cotransfection with NFκB dimers and Brm, whereas expression of *MET* gene was not NFκB dependent at all. *CD44* expression was more strongly dependent upon the noncanonical dimer RelB/p52 than RelA/p50, consistent with a recent report[33]. *CAV1* expression was further increased by cotransfection with RelA/p50 and Brm but was also significantly induced by RelA/p50 alone. Therefore, *CAV1* would only require Brm for its full expression. On the other hands, the

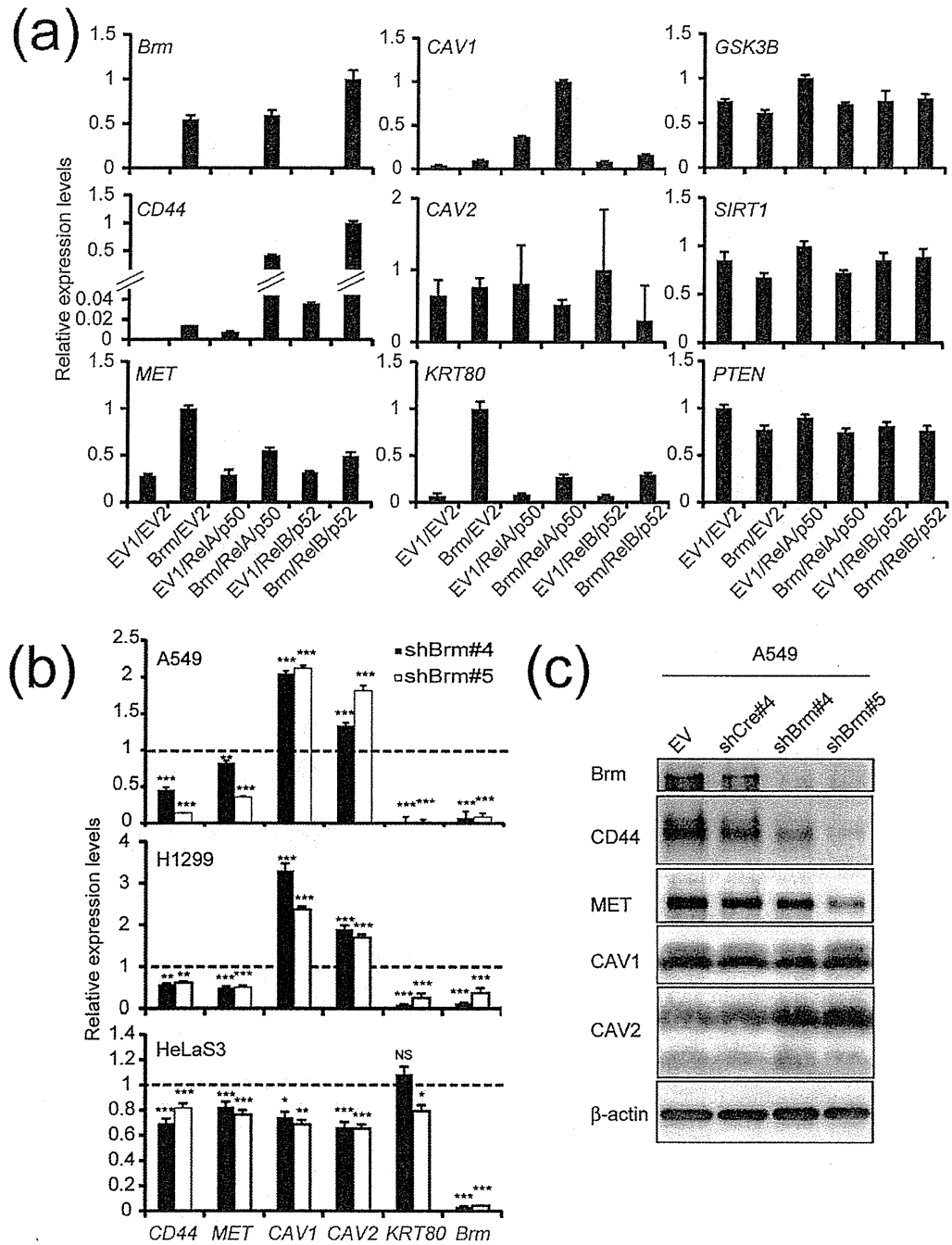


Figure 4 | *Brm* is required for the expression of some type 1-specific genes. (a) Expression of *Brm* and type 1-specific (*CD44*, MET, *CAV1*, *CAV2*, and *KRT80*) and non-type-specific (*GSK3β*, *SIRT1* and *PTEN*) genes in SW13 cells transfected with a Brm expression vector or an empty vector (EV1;pCAG-IG) with or without NFκB dimers (RelA/p50, RelB/p52) or another empty vector (EV2;pRK5). The highest expression level was taken as 1.0. The data represent the means ± S.D. ($n = 3$). (b) Relative expression levels of *CD44* (all transcripts), *MET*, *CAV1*, *CAV2*, and *KRT80* as well as *Brm* mRNA in three cell lines of type 1 cells transduced with shBrm-expressing retroviral vector. The expression levels of cells transduced with shCre#4-expressing vector was taken as 1.0. The data represent the means ± S.D. ($n = 3$). Asterisks indicate P value, compared with those transduced with shCre#4. NS, not significant. *P<0.05, **P<0.01, ***P<0.001 (c) Protein analysis of the parallel A549 cultures prepared as shown in b. β-actin was used as the internal control. The full-length blots were presented in the supplementary Figure 7.

expression levels of *GSK3β* and *Sirt1* (miR-199a-5p targets), and *PTEN* (a miR-214 target), which did not show type 1-specific expression patterns, were not affected by high levels of either Brm or NFκB in SW13 cells (Fig. 4a). Next, A549, H1299, and HeLaS3 cells (type 1) were transduced with retroviral vector encoding Brm shRNA, and the steady-state expression of *CD44*, *MET*, *CAV1*, *CAV2*, and *KRT80* was evaluated by quantitative RT-PCR (Fig. 4b). The levels of *CD44*, *MET*, and *KRT80* mRNA were suppressed to various extents by Brm knockdown. Western blot analysis of parallel A549 cultures also indicated that *CD44* and *MET*, but not *CAV1* and *CAV2*, required Brm for expression (Fig. 4c)[7]. This strong Brm-dependency of CD44 expression is consistent with the previous report that assorted tissues from Brm null/BRG1-positive mice lack CD44 expression. Overall, these findings indicated that genes that are suppressed by miR-199a and simultaneously require the Brm-type SWI/SNF complex for efficient expression show distinct expression patterns: expression in type 1 cells but no expression in type 2 cells. But *CAV1* and *CAV2* expression failed to show clear Brm dependency in A549 and H1299 cells.

We next tested whether type 1-specific genes are under the negative control of EGR1. When HeLaS3 and A549 cells were stably transduced with EGR1-expressing retrovirus, endogenous miR-199a-3p levels were elevated as expected from the axis (Fig. 5a). In HeLaS3, levels of *MET*, *CAV1* and *CAV2* mRNA (Fig. 5a) and their gene products (Fig. 5b and Supplementary Table 2) were reduced by exogenous EGR1 expression. In the case of A549 cells, slight reduction of *CAV1* and *CAV2* mRNA and reduction of MET, CAV1 and CAV2 proteins were observed (Fig. 5ab and Supplementary Table2). These results are consistent with previous reports indicating that *MET*[34] and *CAV1*[35] genes are negatively regulated by EGR1: the *MET* and *CAV1* promoters have one and three EGR/SP-1 binding sites, respectively. We also found EGR1 binding sites on the *CAV1*, *CAV2* and *MET* promoter regions (from -1600 to +500bp of TSS) by using ChIP-seq data obtained by ENCODE. These results suggest that *CAV1/2* and possibly *MET* are specifically expressed in type 1 cells by evading transcriptional suppression by EGR1 proteins and also post-transcriptional suppression by miR-199a-5p/3p.

Overall, these results suggest that there are at least two feedforward loops. One is composed of miR-199a-5p/3p, Brm and CD44, MET and KRT80 (Fig. 6a left) and another is composed of EGR1, miR-199a-5p/-3p and CAV1, and CAV2 (and possibly MET) (Fig. 6a right). Type 1-specific genes would be regulated in an all-or-none manner by either of these two feedforward loops that associate with the robust miR-199a/Brm/EGR1 axis that dictates cancer cell lines to either of the steady states, [miR-199(−)/Brm(+)/EGR1(−)] and [miR-199a(+)/Brm(−)/EGR(+)] (Fig. 6b).

The miR-199a/Brm/EGR1 axis persists in an extended panel of cell lines originating from epithelial tumors. Because our panel of cancer cell lines used for the development of the cell line typing was limited to 14 cell lines, we intended to increase the number of cell lines by directly performing quantitative RT-PCR of *Brm* mRNA (using totally 4 PCR primer pairs), *EGR1* mRNA (using totally 4 PCR primer pairs) and miR-199a-3p by adding 12 new cell lines using the same experimental protocol as used for Fig. 1a (Fig. 7). The 4 independent PCR primer pairs for *Brm* or *EGR1* gave essentially the same expression profile, respectively (Fig. 7a). Out of the total 26 cell lines, 23 could be categorized as either type 1 (17 lines) or type 2 (6 lines), according to the criteria shown in Supplementary Table 3. The remaining 3 lines, which were originated from gastric carcinomas and mammary tumors, cannot be categorized into either type 1 or type 2 (designated type 3). These results indicate that the miR-199a/Brm/EGR1 axis is largely retained in variety of epithelial tumor cell lines.

Since microarray data of *Brm* and *EGR1* mRNA for 17 among these cell lines categorized as type 1 and type 2 were available from

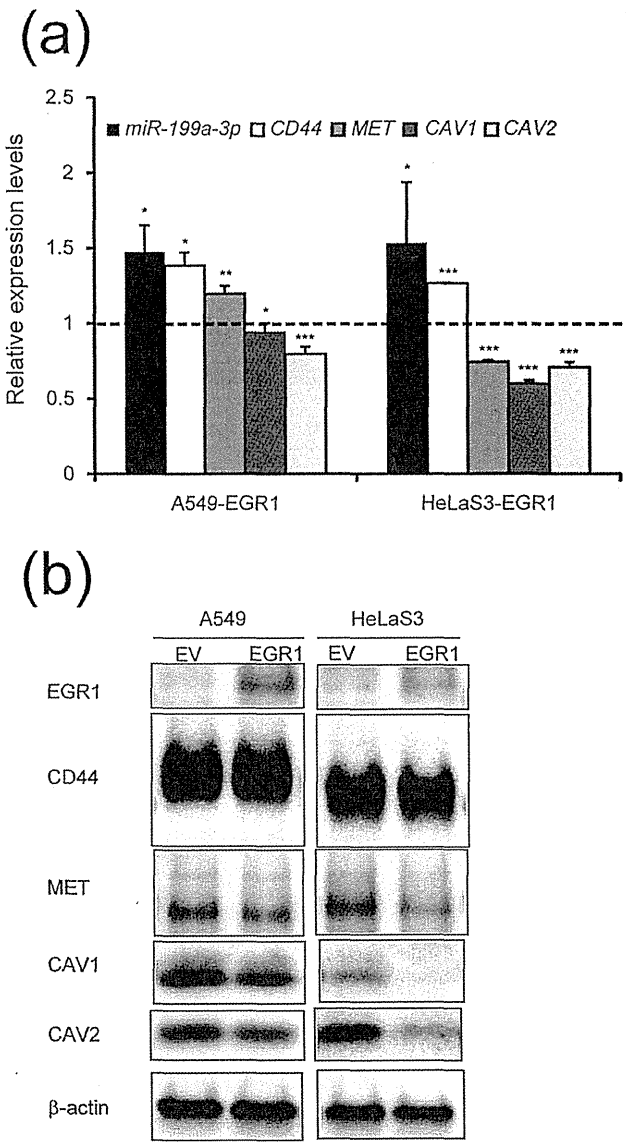


Figure 5 | Effects of exogenous EGR1 expression in type 1 cells. (a) Relative expression levels of miR-199a-3p and type 1-specific mRNAs were determined by quantitative RT-PCR in A549 and HeLaS3 cells which were transduced with retroviral vectors expressing EGR1. The expression levels of cells transduced with empty vector was taken as 1.0. The data represent the means ± S.D. ($n = 3$). Asterisks indicate P value, compared with those transduced with empty vector. *P<0.05, **P<0.01, ***P<0.001 (b) Analysis of type 1-specific gene products and EGR1 in the parallel cultures prepared in a by western blotting. β-actin was used as the loading control. The full-length blots were presented in the supplementary Figure 8 and relative expression levels of each protein including two additional sets of blots were quantified in Supplementary Table 2.

Sanger database (Genomics of Drug Sensitivity in Cancer http://www.cancerrxgene.org), and their expression profiles obtained from the database was compared with those of the qRT-PCR data shown in Fig. 7a (Supplementary Fig. 9). We found that the expression profiles of *Brm* and *EGR1* are not correlates well between them. Since *Brm* mRNA levels of even Brm-deficient cell lines such as SW13, H522, C33A, A427, and H23—previously reported by our[6,12] and other groups[7,36] by RT-PCR or Northern blotting analysis—were significantly high according to Sanger database, there would be limitations

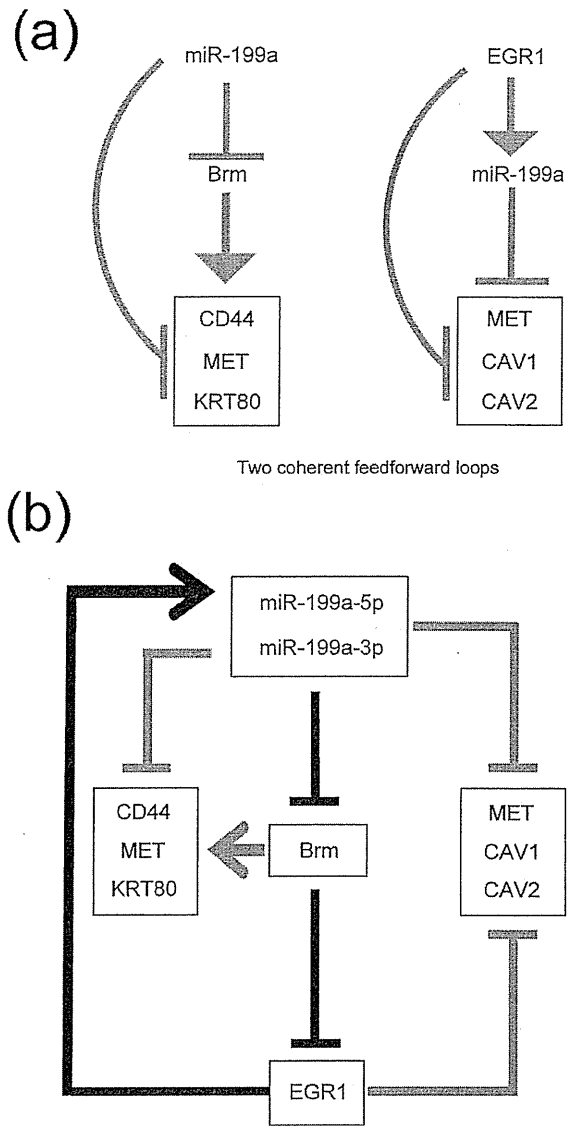


Figure 6 | Models of the regulatory networks operating in various epithelial tumor cell lines. (a) Two feedforward loops that function to establish type 1-specific gene expression in an all-or-none manner. (b) A double-negative feedback loop (indicated by black arrows) is integrated by the two feedforward loops (indicated by gray arrows) shown in a.

in microarray data to estimate mRNA levels of such transcriptional regulatory genes as *Brm* and *EGR1* accurately.

Since we found expression profiles of *CD44*, *MET*, *CAV1* and *CAV2* mRNA by our qRT-PCR and those obtained from Sanger Database are correlated well, we showed both of them in Fig. 7b. The expression levels of *CD44* and *MET* were high in type 1 cell lines, whereas they were mostly undetectable in the type 2 cell lines even in these extended panels, and specific *CD44* and *MET* expression in type 1 cells were statistically supported in both data of qRT-PCR and the Database. *CAV1* and *CAV2* expression was not detected in most type 2 cells with a clear exception of A427. Because of this, type 1 specific expression of *CAV1* and *CAV2* was not supported statistically. Relatively low *EGR1* expression in A427 among typel cells (Fig. 7a, Supplementary Table 3) might partly explain this exception.

Expression patterns observed in type 1 or type 2 cell lines are recapitulated in some cancer lesions of NSCLCs. We finally examined whether the distinct expression patterns observed between the two cell types are reflected in human primary tumors. Since the cell lines originating from NSCLCs in the cell line panel used here can be categorized as both type 1 and type 2 (Supplementary Table 3), we pathologically analyzed surgically resected, formalin-fixed, paraffin-embedded tissues from human cancer lesions of NSCLCs. Among NSCLCs, we especially focused upon squamous cell carcinoma (SCC), because in this type of cancer, we can easily understand activity of proliferation or the status/direction of differentiation two-dimensionally in the histological section. After preparing sequential thin sections of total 21 SCC cases, they were immunohistochemically stained with antibodies against Brm, CD44, MET, and CAV1 and also probed for miR-199a-5p by *in situ* hybridization and interrelationships among their expression patterns in the coincident area of the each section were analyzed by comparing lower and higher differentiation status.

In the area of lower differentiation status where cancer cells are crowded by the active proliferation and have increased nuclear/cytoplasmic ratio without keratinization, we clearly observed a Brm^+, $CD44^+$, MET^+, and $CAV1^+$ phenotype in almost all cases. In some of these areas, expression of miR-199a was undetectable as shown in Fig. 8 (surrounded by solid line), which recapitulates the expression patterns of type 1 cells. However, in the other areas expressing these 4 proteins, we detected also miR-199a expression, indicating expression heterogeneity in cancer legions.

As for the areas of highly differentiation status, we observed them in many so-called cancer pearls in 4 cases of SCC, where cancer cells are sparse with large cytoplasm. Even in cancer pearls, significant population at the periphery retains clearly recognized nuclei indicating that cells are still alive, but in the central regions, cells are gradually losing their nuclei on their process of keratinization. We detected a Brm^-, $CD44^-$, MET^-, and $CAV1^-$, and miR-199a$^+$ phenotype in all of the cancer pearls where the cell retained nuclei, which recapitulated that of type 2 cells (Fig. 8 within the broken line). In the area between solid line and the cancer pearl in Fig. 8, where tumor cells assumed intermediate differentiation status, these 4 proteins and miR-199a were weakly expressed with various extents. At least in these regions, tumor cells might be undergoing changes from the type 1 cells into the type 2 cells through the process of cellular differentiation.

Discussion

Using 12 cell lines that were strictly derived from human epithelial tumors, we can confirm the findings of our previous report that these cells can be classified into type 1 [mir-199a(−)/Brm(+)/EGR1(−)] (8 lines) or type 2 [miR-199a(+)/Brm(−)/EGR1(+)] (4 lines) cells. In our current study, we were able to efficiently identify the type 1-specific genes by setting the reported miR-199a and miR-214 target genes as the candidates. Some of the identified type 1-specific genes (*CD44*, *MET*, and *KRT80*) required *Brm*, whereas others (*CAV1*, *CAV2* and probably *MET*) required the absence of EGR1 for their efficient expression, indicating that two coherent feedforward loops are formed (Fig. 6a). These two feedforward loops are integrated into the robust double-negative feedback loop forming a regulatory network that functions as an efficient switch that determines the expression levels of these type 1-specific genes in an all-or-none manner (Fig. 6b). Thus, the current situation would be a good example of a network formed by multiple miRNA-mediated feedback and feedforward loops[37,38], which are commonly present in a wide variety of cell lines. Importantly, we observed regions whose expression patterns recapitulated those of type 1 or type 2 cells by pathological analysis of SCC legions of NSCLC tissues. It should be pointed out, however, that there are several lesions whose expression patterns do not belong to either of them. We speculate that in the process of cell line establishment from primary tumors, they would tend to fall into either of steady states, type 1 or type 2 cells.

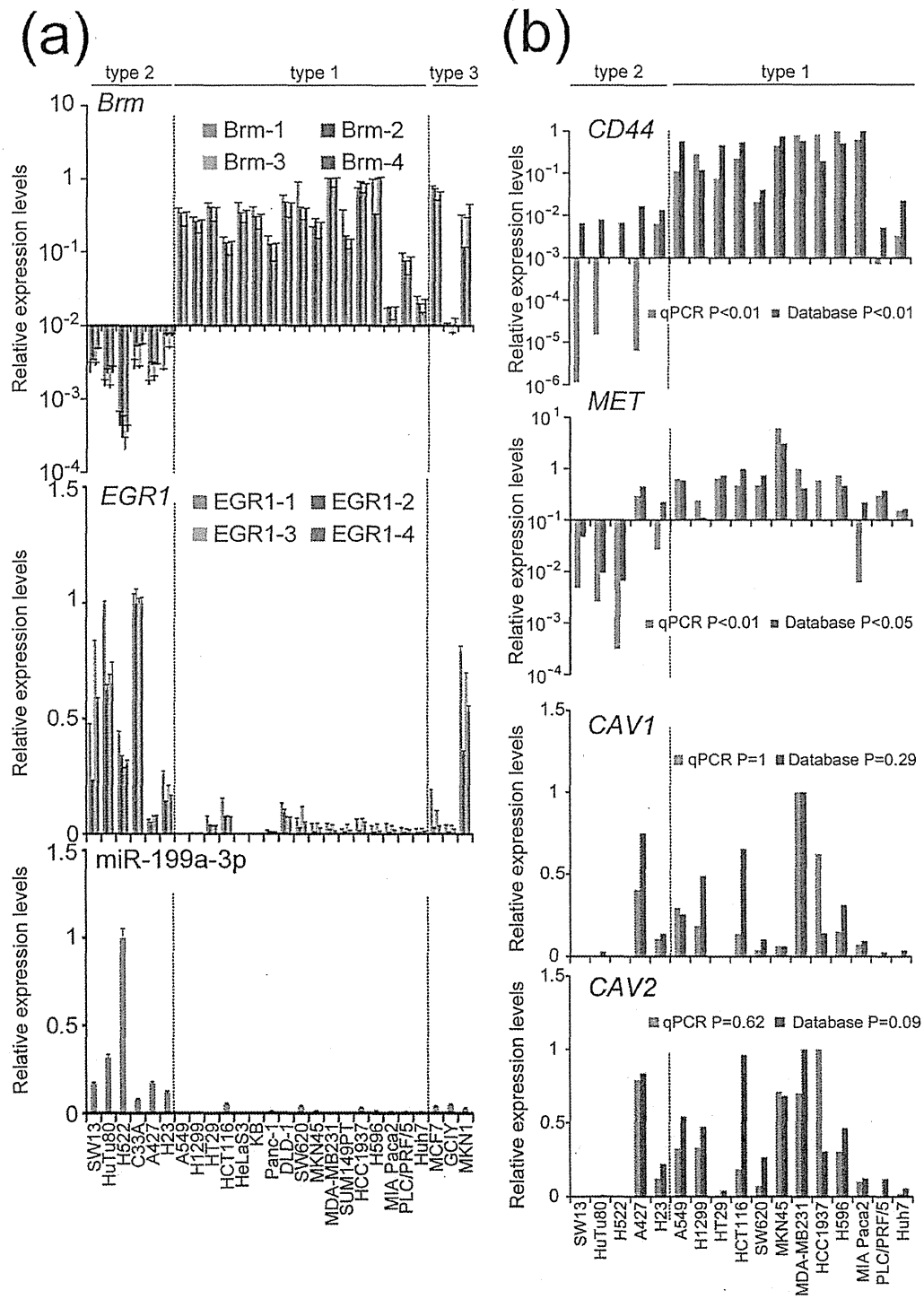


Figure 7 | Gene expression analysis on an extended epithelial tumor cell line panel. (a)Expression profiles of *Brm* and *EGR1* mRNA and mature miR-199a-3p RNA of 26 cell lines, as determined by quantitative RT-PCR. Three additional primer pairs were designed and for *Brm* and *EGR1* mRNA quantification other than used in Fig. 1 (Brm-1, EGR1-1). The relative expression levels are shown by taking the highest as 1.0. Two red vertical break lines indicate the boundary among each type cell lines. Detailed criteria for type1-3 cells were indicated in Supplementary Table 3. (b)Expression profiles of *CD44, MET, CAV1 and CAV2* mRNA of 17 cell lines, determined by quantitative RT-PCR (blue bars) or obtained from Sanger database (red bars).

Among the type 1-specific genes shown here, *CD44*, *MET*, *CAV1*, and *CAV2* alone significantly contributed to anchorage-independent growth of type 1 cells when tested in knockdown experiments using four type 1 cell lines (Fig. 3). Several previous reports indicated that CD44[39], MET[40], and CAV1[41,42] are potentially important for colony formation in some epithelial tumor cell lines. We observed that these four genes are all simultaneously suppressed to a marginal level in type 2 cells by the regulatory network shown here, ensuring the anchorage dependency of type 2 cells. Whereas CD44[43,44], MET[45], CAV1[46], and CAV2[47] have their own multiple downstream signaling