**Table 3**
Correlations between HER2 mutations and IHC, FISH, and DISH.

| | IHC | | | FISH | | | DISH | | |
|---|---|---|---|---|---|---|---|---|---|
| | (3+) | (0, 1+, 2+) | *P* value | Amplified | Not amplified | *P* value | Amplified | Not amplified | *P* value |
| HER2 mut. | | | | | | | | | |
| Positive | 1 | 5 | 0.392 | 2 | 4 | <0.001 | 2 | 4 | <0.001 |
| Negative | 5 | 209 | | 3 | 211 | | 7 | 207 | |

*Abbreviation*: mut., mutation.

found four cases with amplification among eight cases with mutations [31]. Recently Mazieres et al. [34] found three cases with amplification out of 34 cases with mutations. In their study, of 22 patients with *HER2* mutations who were treated with new HER2-targeted agents, 7 and 11 patients experienced disease stabilization and partial response, respectively. However, they did not investigate whether FISH-positive cases responded to these novel agents. Since *HER2* amplification could be involved in the response to anti-HER2 agents in the study, it is necessary to determine the frequency with which mutation-positive, but IHC- and FISH/DISH-negative cases respond to these drugs.

FISH remains the "gold standard" to demonstrate the presence of *HER2* gene amplifications in breast cancer. In lung cancer, *HER2* amplification is observed by FISH in approximately 2%–22.8% of NSCLCs [25,28,37,38]. DISH was recently approved by the FDA for use in determining the *HER2* gene status in patients with breast cancer. Although both FISH and DISH use formalin-fixed, paraffin-embedded human cancer tissue specimens, DISH has some advantages over FISH: ease of handling, similar to IHC, and rapid performance; visualization via DISH uses light microscopy; and the specimens can be archived and retrieved for later analysis. Additionally, DISH is more advantageous because heterogeneity within the same specimen is quite easy to detect. In our study, the overall concordance between *HER2* amplification determined by FISH and DISH was 96.7%. Additionally, DISH resulted in higher *HER2*/centromere 17 ratios than FISH, as illustrated by regression analysis. This result indicated that DISH could have an advantage over FISH when heterogeneity of *HER2*-amplified cells is present in the tumors. Recently Grob et al. [32] reported that heterogeneity of *HER2* amplification was found in four of ten tumors, as assessed by FISH, and they concluded that heterogeneity of *HER2* amplification does indeed occur, but may not be that common. Therefore, for selection of patients indicated for treatment with novel anti-HER2 agents based on *HER2* amplification, we speculate that DISH, which is less susceptible to intratumoral heterogeneity of the amplification, may be better for determination of HER2 status.

## 5. Conclusion

In summary, we reported the performances of IHC, DISH, FISH, and mutation analyses in determining HER2 status in a cohort of patients with lung adenocarcinomas. We found that *HER2* amplification tended to be consistent with mutations in the *HER2* gene. Although indications for the use of novel anti-HER2 agents are generally based on the status of *HER2* mutations, we suggest that these criteria should be revalidated to determine whether specific populations of patients (i.e., those with mutations only or those with both mutations and amplifications) respond better to these agents. Moreover, if it is necessary to determine the *HER2* amplification status for appropriate patient selection, we recommend DISH rather than FISH because DISH is easier to perform and less susceptible to intratumoral heterogeneity.

## Conflict of interest

The authors declare no conflicts of interest.

## Acknowledgements

The authors thank Mizuki Koyama, Shouri Ishikawa, Yasuyo Shimojo, and Kazuya Murata for technical assistance. This research was not supported by any funds.

## References

[1] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69–90.
[2] Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497–500.
[3] Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004;350:2129–39.
[4] Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 2007;448:561–6.
[5] Pao W, Girard N. New driver mutations in non-small-cell lung cancer. Lancet Oncol 2011;12:175–80.
[6] Yaziji H, Goldstein LC, Barry TS, Werling R, Hwang H, Ellis GK, et al. HER-2 testing in breast cancer using parallel tissue-based methods. JAMA 2004;291:1972–7.
[7] Bang YJ, Van Cutsem E, Feyereislova A, Chung HC, Shen L, Sawaki A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet 2010;376:687–97.
[8] Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 2001;344:783–92.
[9] Romond EH, Perez EA, Bryant J, Suman VJ, Geyer Jr CE, Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 2005;353:1673–84.
[10] Clamon G, Herndon J, Kern J, Govindan R, Garst J, Watson D, et al. Lack of trastuzumab activity in nonsmall cell lung carcinoma with overexpression of erb-B2: 39810: a phase II trial of Cancer and Leukemia Group B. Cancer 2005;103:1670–5.
[11] Gatzemeier U, Groth G, Butts C, Van Zandwijk N, Shepherd F, Ardizzoni A, et al. Randomized phase II trial of gemcitabine-cisplatin with or without trastuzumab in HER2-positive non-small-cell lung cancer. Ann Oncol 2004;15:19–27.
[12] Lara Jr PN, Laptalo L, Longmate J, Lau DH, Gandour-Edwards R, Gumerlock PH, et al. Trastuzumab plus docetaxel in HER2/neu-positive non-small-cell lung cancer: a California Cancer Consortium screening and phase II trial. Clin Lung Cancer 2004;5:231–6.
[13] Langer CJ, Stephenson P, Thor A, Vangel M, Johnson DH. Trastuzumab in the treatment of advanced non-small-cell lung cancer: is there a role? Focus on Eastern Cooperative Oncology Group study 2598. J Clin Oncol 2004;22:1180–7.
[14] Oxnard GR, Binder A, Janne PA. New targetable oncogenes in non-small-cell lung cancer. J Clin Oncol 2013;31:1097–104.
[15] Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol 2011;6:244–85.
[16] Sobin LH, Gospodarowicz MK, Wittekind C. TNM classification of malignant tumours. 7th ed. New York: John Wiley & Sons; 2009.
[17] Fujimoto M, Yoshizawa A, Sumiyoshi S, Sonobe M, Kobayashi M, Koyanagi I, et al. Stromal plasma cells expressing immunoglobulin G4 subclass in non-small cell lung cancer. Hum Pathol 2013;44:1569–76.
[18] Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Arch Pathol Lab Med 2007;131:18–43.
[19] Sonobe M, Manabe T, Wada H, Tanaka F. Mutations in the epidermal growth factor receptor gene are linked to smoking-independent, lung adenocarcinoma. Br J Cancer 2005;93:355–63.
[20] Sonobe M, Manabe T, Wada H, Tanaka F. Lung adenocarcinoma harboring mutations in the ERBB2 kinase domain. J Mol Diagn 2006;8:351–6.

[21] Takahashi T, Sonobe M, Kobayashi M, Yoshizawa A, Menju T, Nakayama E, et al. Clinicopathologic features of non-small-cell lung cancer with EML4-ALK fusion gene. Ann Surg Oncol 2010;17:889–97.
[22] Hatzaki A, Razi E, Anagnostopoulou K, Iliadis K, Kodaxis A, Papaioannou D, et al. A modified mutagenic PCR-RFLP method for K-ras codon 12 and 13 mutations detection in NSCLC patients. Mol Cell Probes 2001;15:243–7.
[23] Kakegawa S, Shimizu K, Sugano M, Miyamae Y, Kaira K, Araki T, et al. Clinicopathological features of lung adenocarcinoma with KRAS mutations. Cancer 2011;117:4257–66.
[24] Au NH, Cheang M, Huntsman DG, Yorida E, Coldman A, Elliott WM, et al. Evaluation of immunohistochemical markers in non-small cell lung cancer by unsupervised hierarchical clustering analysis: a tissue microarray study of 284 cases and 18 markers. J Pathol 2004;204:101–9.
[25] Hirsch FR, Varella-Garcia M, Franklin WA, Veve R, Chen L, Helfrich B, et al. Evaluation of HER-2/neu gene amplification and protein expression in non-small cell lung carcinomas. Br J Cancer 2002;86:1449–56.
[26] Kwiatkowski DJ, Harpole Jr DH, Godleski J, Herndon 2nd JE, Shieh DB, Richards W, et al. Molecular pathologic substaging in 244 stage I non-small-cell lung cancer patients: clinical implications. J Clin Oncol 1998;16:2468–77.
[27] Nakamura H, Saji H, Ogata A, Hosaka M, Hagiwara M, Kawasaki N, et al. Correlation between encoded protein overexpression and copy number of the HER2 gene with survival in non-small cell lung cancer. Int J Cancer 2003;103:61–6.
[28] Heinmoller P, Gross C, Beyser K, Schmidtgen C, Maass G, Pedrocchi M, et al. HER2 status in non-small cell lung cancer: results from patient screening for enrollment to a phase II study of herceptin. Clin Cancer Res 2003;9:5238–43.
[29] Zinner RG, Glisson BS, Fossella FV, Pisters KM, Kies MS, Lee PM, et al. Trastuzumab in combination with cisplatin and gemcitabine in patients with Her2-overexpressing, untreated, advanced non-small cell lung cancer: report of a phase II trial and findings regarding optimal identification of patients with Her2-overexpressing disease. Lung Cancer 2004;44:99–110.
[30] Soh J, Toyooka S, Ichihara S, Fujiwara Y, Hotta K, Suehisa H, et al. Impact of HER2 and EGFR gene status on gefitinib-treated patients with nonsmall-cell lung cancer. Int J Cancer 2007;121:1162–7.
[31] Li C, Sun Y, Fang R, Han X, Luo X, Wang R, et al. Lung adenocarcinomas with HER2-activating mutations are associated with distinct clinical features and HER2/EGFR copy number gains. J Thorac Oncol 2012;7:85–9.
[32] Grob TJ, Kannengiesser I, Tsourlakis MC, Atanackovic D, Koenig AM, Vashist YK, et al. Heterogeneity of ERBB2 amplification in adenocarcinoma, squamous cell carcinoma and large cell undifferentiated carcinoma of the lung. Mod Pathol 2012;25:1566–73.
[33] Arcila ME, Chaft JE, Nafa K, Roy-Chowdhuri S, Lau C, Zaidinski M, et al. Prevalence, clinicopathologic associations, and molecular spectrum of ERBB2 (HER2) tyrosine kinase mutations in lung adenocarcinomas. Clin Cancer Res 2012;18:4910–8.
[34] Mazieres J, Peters S, Lepage B, Cortot AB, Barlesi F, Beau-Faller M, et al. Lung cancer that harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives. J Clin Oncol 2013;31:1997–2003.
[35] Buttitta F, Barassi F, Fresu G, Felicioni L, Chella A, Paolizzi D, et al. Mutational analysis of the HER2 gene in lung tumors from Caucasian patients: mutations are mainly present in adenocarcinomas with bronchioloalveolar features. Int J Cancer 2006;119:2586–91.
[36] Cappuzzo F, Ligorio C, Janne PA, Toschi L, Rossi E, Trisolini R, et al. Prospective study of gefitinib in epidermal growth factor receptor fluorescence in situ hybridization-positive/phospho-Akt-positive or never smoker patients with advanced non-small-cell lung cancer: the ONCOBELL trial. J Clin Oncol 2007;25:2248–55.
[37] Cappuzzo F, Varella-Garcia M, Shigematsu H, Domenichini I, Bartolini S, Ceresoli GL, et al. Increased HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. J Clin Oncol 2005;23:5007–18.
[38] Pelosi G, Del Curto B, Dell'Orto P, Pasini F, Veronesi G, Spaggiari L, et al. Lack of prognostic implications of HER-2/neu abnormalities in 345 stage I non-small cell carcinomas (NSCLC) and 207 stage I-III neuroendocrine tumours (NET) of the lung. Int J Cancer 2005;113:101–8.

Ishikawa *et al. World Journal of Surgical Oncology* 2014, **12**:253
http://www.wjso.com/content/12/1/253

WORLD JOURNAL OF SURGICAL ONCOLOGY

RESEARCH Open Access

# Expression of the *GLI* family genes is associated with tumor progression in advanced lung adenocarcinoma

Masashi Ishikawa*, Makoto Sonobe, Naoto Imamura, Terumasa Sowa, Kei Shikuma and Hiroshi Date

## Abstract

**Background:** The hedgehog (Hh) signaling pathway is aberrantly activated in various cancers. Expression of the GLI family of genes, which encode for transcriptional factors of the Hh pathway, has not been fully assessed in clinical samples of advanced lung adenocarcinoma. In this study, we retrospectively evaluated the expression of the GLI family of genes in advanced stage lung adenocarcinoma samples and determined their relation to patient survival.

**Methods:** The levels of *GLI1*, *GLI2*, and *GLI3* mRNA expression were measured by quantitative real-time polymerase chain reaction in surgically obtained tissue samples from stage II-IV lung adenocarcinoma patients (n = 102). Pairwise comparisons between all three *GLI* mRNA expression were performed, and after dichotomizing the patients into low and high expression groups according to each *GLI* mRNA expression level, survival curves were calculated and multivariate analyses were conducted.

**Results:** Significant positive correlation was found between *GLI1* and *GLI3* mRNA expression ($P$ <0.001). Tumors with higher expression (upper 15%) of *GLI1* or *GLI3* mRNA were associated with poor survival in stage II-IV (5-year overall survival rates: *GLI1* mRNA low, 41.7% *vs.* high, 20.0%, $P = 0.0074$; *GLI3* mRNA low, 43.1% *vs.* high, 13.3%, $P = 0.0062$) and stage III-IV (5-year overall survival rates: *GLI1* mRNA low, 34.0% *vs.* high, 0%, $P = 0.0012$; *GLI3* mRNA low, 33.4% *vs.* high, 7.7%, $P = 0.057$) lung adenocarcinoma patients. *GLI2* mRNA expression did not appear to have great clinical significance. Multivariate analysis revealed higher *GLI1* mRNA expression as an independent factor for unfavorable patient survival ($P = 0.0030$, hazard ratio = 3.1, 95% confidence interval = 1.5-6.2), as well as tumor differentiation and stage.

**Conclusions:** Expression of *GLI1* and *GLI3* mRNA was strongly correlated, and their overexpression, especially that of *GLI1*, was found to be predictive of aggressive tumor behavior. This study indicates that the Hh pathway may be a key oncogenic signaling network in tumor pathogenesis and, thus, a potential therapeutic target in advanced lung adenocarcinoma.

**Keywords:** GLI, Hedgehog signaling, Lung adenocarcinoma, Advanced stage, Expression, Prognosis

## Background

Non-small cell lung cancer (NSCLC) has been the leading cause of cancer-related deaths world-wide, and adenocarcinoma accounts for the majority of all NSCLC cases [1,2]. Although any possibility of a cure rests almost entirely with surgical resection with or without chemo- or radiotherapy, postoperative recurrence rate is still high and survival rate remains low compared with other types of cancers. In particular, advanced NSCLC can be exceptionally lethal even with aggressive anticancer therapies, despite the dramatic treatment shift of some targeted therapeutic agents for lung adenocarcinoma [3,4]. Developing and deploying new treatment approaches are vital to improve patient prognoses.

The hedgehog (Hh) signaling pathway, which consists of three distinct homologues (sonic, indian, and desert Hh) is constitutively activated in a variety of human tumor entities [5]. Normally, Hh signaling plays an important role during vertebral embryonic development and tissue maintenance [6]. The Hh proteins are secreted molecules, and their receptor, Patched, usually functions as a downstream

* Correspondence: mishi@kuhp.kyoto-u.ac.jp
Department of Thoracic Surgery, Faculty of Medicine, Kyoto University, 54 Shogoin-kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan

BioMed Central

© 2014 Ishikawa et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

inhibitor in the absence of Hh. Hh signaling is activated through binding to the Patched receptor, which unleashes the transmembrane protein, Smoothened, and ultimately upregulates GLI [5]. GLI proteins are zinc finger transcription factors which exert their functions by nuclear localization and promoting target gene transcription, and in humans, at least three distinct *GLI* genes, *GLI1*, *GLI2*, and *GLI3*, have been identified [7]. As the GLI family of proteins are the final mediators of Hh signaling, their expression, especially that of GLI1, is deemed to be the best reflection of the net effect of Hh activation, thus, survival studies on GLI may indicate the potential involvement of Hh in tumor pathogenesis [8].

In addition to the tumorigenesis based on genetic alterations of the Hh pathway predominantly seen in basal cell carcinoma and medulloblastoma, extensive involvement of ligand-dependent Hh activation has been suggested in various cancer entities [9,10]. With regard to lung cancer, while the involvement of Hh signaling in small cell lung cancer has been shown to be evident [11-13], it remains poorly understood in lung adenocarcinoma, especially in the advanced stages.

In this study, we evaluated the expression levels of the three *GLI* family genes (*GLI1*, *GLI2*, and *GLI3*) in surgical samples from advanced lung adenocarcinoma patients (stage II-IV), and revealed the possible involvement of the Hh signaling pathway in tumor advancement by analysis for correlations between GLI expression and patient prognosis. As we found considerable impact of *GLI1* and *GLI3* gene expression on patient survival, this study demonstrates that inhibition of Hh signaling may be a promising drug target for the treatment of advanced lung adenocarcinoma.

## Methods

### Patients, tissue samples, and tumor information

Tissue samples were obtained from patients who were diagnosed as stage II-IV lung adenocarcinoma after surgical resection at the Kyoto University Hospital between July 2001 and December 2007. Continuous surgical cases with sufficient tissue material for evaluations were included in this study (n = 102). All of these tumors were histologically confirmed as lung adenocarcinoma, and their tumor staging and degree of differentiation were all assessed by board-certified pathologists in the Department of Pathology of the Kyoto University Hospital. Tumor stage was determined by the latest tumor-node-metastasis classification system [14]. Histological type and grade of cell differentiation were determined according to the WHO classification system [15]. Their demographics and baseline characteristics are presented in Table 1. Pre- or postoperative chemotherapy was mainly composed of orally administered tegafur-uracil and/or conventional intravenous chemotherapeutic agents (carboplatin, paclitaxel, docetaxel, gemcitabine, vinorelbine) for NSCLC. Only one patient (stage II) received radiotherapy prior to surgery (concurrently with chemotherapy), and postoperative radiotherapy (mediastinum or chest wall) was performed in four patients (stage III). Complete tumor resection was achieved in 77 patients (75.5%), whereas in 25 patients (24.5%) macro- or microscopic residual tumors were suspected after surgery (for example, pleural dissemination, malignant pleural effusion, positive stumps, or distant metastasis prior to surgery). Informed consent for participation in this study was obtained from all patients prior to their surgeries, and this study was reviewed and approved by the Ethics Committee of the Graduate School and Faculty of Medicine at the Kyoto University.

### Preparation of tissue mRNA

For sample collection, tumor tissue samples were dissected immediately after surgical resection and soaked in RNAlater TissueProtect Tubes (Qiagen, Tokyo, Japan) for more than 48 h before storage at -80°C until use. Total RNA was isolated from tissue samples using RNeasy Plus Mini Kit (Qiagen), and reverse transcription of total RNA was conducted using the Ready-To-Go You-Prime First-Strand Beads (Amersham Biosciences, Uppsala, Sweden) to obtain cDNA.

### Quantification of *GLI* mRNA

To quantify *GLI* mRNA expression levels of each sample, quantitative real-time PCR was performed using the LightCycler thermal cycler system (Roche Diagnostics Japan, Tokyo, Japan). The PCR primers used for the quantitative amplification of *GLI1* mRNA were forward: 5′- CTCCCGAAGGACAGGTATGTAAC -3′ and reverse: 5′- CCCTACTCTTTAGGCACTAGAGTTG -3′, those of *GLI2* mRNA were forward: 5′- AGAAGCAGCGCAAT GACGTG -3′ and reverse: 5′- GTCATCCAGTGCCGT CAGGT -3′, and those of *GLI3* mRNA were forward: 5′- AAACCCCAATCATGGACTCAAC -3′ and reverse: 5′- TACGTGCTCCATCCATTTGGT-3′. The primers for *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) mRNA, used as an internal control, were forward: 5′-ACAACAGCCTCAAGATCATCAG-3′ and reverse: 5′-TCTTCTGGGTGGCAGTGATG-3′. After a 20-μL reaction mixture containing 0.5 μM forward/reverse primers and 0.03 μg cDNA in QuantiTect SYBR Green PCR Master Mix (Qiagen) was prepared, PCR amplification was initiated by preincubation for 15 min at 95°C for initial activation, followed by 40 cycles of the following protocol: denaturation at 94°C for 15 s, annealing at 59°C for 15 s, and elongation at 72°C for 15 s with detection of fluorescence products. The quantitative data were analyzed with LightCycler analysis software version 5.03 (Roche Diagnostics Japan). The expression levels of *GLI* genes were calculated as the ratio of *GLI* mRNA value to

**Table 1 Characteristics of the patients included in the study**

| | | Numbers (n) | Percent (%) |
|---|---|---|---|
| Age (years) | | | |
| | Mean ± SD (range) | 64.8 | ±9.8 (31-81) |
| Gender | | | |
| | Male | 68 | 66.7 |
| | Female | 34 | 33.3 |
| Cigarette smoking | | | |
| | Never smoked | 34 | 33.3 |
| | Ex-smoker | 20 | 19.6 |
| | Current smoker | 48 | 47.1 |
| | Pack-year/Mean ± SD (range) | 36.6 | ±36.8 (0-180) |
| Tumor differentiation | | | |
| | Well | 16 | 15.7 |
| | Moderate | 53 | 52.0 |
| | Poor | 32 | 31.4 |
| | Unknown | 1 | |
| Tumor stage | | | |
| | II | 31 | 30.4 |
| | III | 60 | 58.8 |
| | IV | 11 | 10.8 |
| Operation method | | | |
| | Pneumonectomy | 1 | 1.0 |
| | Bilobectomy | 1 | 1.0 |
| | Lobectomy | 83 | 81.4 |
| | Segmentectomy | 9 | 8.8 |
| | Partial resection | 8 | 7.8 |
| Complete resection | | | |
| | Yes | 77 | 75.5 |
| | No | 25 | 24.5 |
| Preoperative chemotherapy | | | |
| | Yes | 9 | 8.8 |
| | No | 93 | 91.2 |
| Postoperative chemotherapy | | | |
| | Yes | 74 | 72.5 |
| | No | 28 | 27.5 |
| Radiotherapy to thorax (chest wall or mediastinum)[a] | | | |
| | Yes (Pre/Post) | 5 (1/4) | 4.9 |
| | No | 97 | 95.1 |
| *EGFR* gene mutation status | | | |
| | Wild-type | 46 | 45.1 |
| | Mutation (+) | 33 | 32.4 |
| | Unknown | 23 | 22.5 |
| Recurrence[b] | | | |
| | Yes | 82 | 80.4 |
| | No | 20 | 19.6 |

**Table 1 Characteristics of the patients included in the study** *(Continued)*

| | | | |
|---|---|---|---|
| Prognosis | | | |
| | Alive | 33 | 34.4 |
| | Dead | 63 | 65.6 |

[a]Not including palliative radiotherapy.
[b]Including residual tumor after surgery.
SD: standard deviation.

*GAPDH* mRNA value, and are expressed as median and mean ± standard deviation.

### Statistical analysis

Relationships between each pair of mRNA expression were estimated by calculating the Spearman's rank correlation coefficient ($r_s$). Pairwise comparisons of postoperative survival of each GLI low and high groups were compared using the Mann-Whitney U test. Survival curves were evaluated by the Kaplan-Meier method, and cumulative survival data were compared using the log-rank test. Multivariate analysis of prognostic factors was performed by the Cox proportional hazard model. Differences were considered significant when $P$ <0.05. All statistical analyses were performed using StatMate IV software version 4.01 (ATMS, Tokyo, Japan) and JMP software version 8 (SAS Institute Japan, Tokyo, Japan).

## Results

### *GLI1, GLI2,* and *GLI3* mRNA were expressed in almost all advanced lung adenocarcinoma tissues

*GLI1* and *GLI3* mRNA expression was detected in all 102 samples, whereas *GLI2* mRNA was detected in 99 samples. The level of *GLI1* mRNA (normalized and expressed as a ratio to *GAPDH* mRNA) ranged from $2.35 \times 10^{-6}$ to $1.03 \times 10^{-1}$ (median, $3.03 \times 10^{-4}$; mean ± standard deviation, $3.97 \times 10^{-3} \pm 1.51 \times 10^{-2}$); *GLI2* mRNA level was 0 to $1.01 \times 10^{-2}$ ($1.25 \times 10^{-3}$, $1.78 \times 10^{-3} \pm 1.86 \times 10^{-3}$); and *GLI3* mRNA level was $7.68 \times 10^{-5}$ to $2.42 \times 10^{-1}$ ($1.07 \times 10^{-2}$, $2.23 \times 10^{-2} \pm 3.93 \times 10^{-2}$). The three *GLI* mRNA expression levels in all tissue samples are shown in Figure 1. Wide variations were observed in *GLI1* and *GLI3* mRNA levels in the primary tumor cohort, and based on these, we defined the upper 15% of patients as high *GLI* mRNA expression for each *GLI* gene (cutoff line; *GLI1* mRNA: $2.49 \times 10^{-3}$, *GLI2* mRNA: $3.71 \times 10^{-3}$, *GLI3* mRNA: $3.46 \times 10^{-2}$).

### *GLI1* and *GLI3* mRNA expression levels were remarkably correlated

By comparing *GLI1, GLI2,* and *GLI3* mRNA expression levels, strong positive correlation was observed between *GLI1* and *GLI3* mRNA expression ($r_s = 0.46$, $P$ <0.001) (Figure 2), whereas no such relations were found between *GLI2* and *GLI1* or *GLI3* mRNA expression levels. Even with exclusion of the four observed outliers (n = 98), strong positive correlation was also found between *GLI1* and *GLI3* mRNA expression ($r_s = 0.40$, $P$ <0.001).

### *GLI1* and *GLI3* mRNA expression reflected postoperative patient survival

There were statistically significant differences in postoperative survival between *GLI1/GLI3* low and high mRNA expression groups (*GLI1* mRNA: $P$ <0.001; *GLI3* mRNA: $P$ <0.001). Survival curves for each high/low *GLI* mRNA expression groups are presented in Figure 3. Five-year overall survival (OAS) rates in the *GLI1* mRNA low and high groups were 41.7% and 20.0%, and those in the *GLI3* mRNA low and high groups were 43.1% and 13.3%, respectively. Higher *GLI1* and *GLI3* mRNA expressing tumors in advanced lung adenocarcinoma patients were correlated with significantly worse overall survival rates than lower expression groups (*GLI1* mRNA: $P = 0.0074$, hazard ratio (HR) = 2.2 (95% confidence interval: 1.4-6.9); *GLI3* mRNA: $P = 0.0062$, HR = 2.2 (1.4-6.6)). Similarly, 5-year disease-free survival (DFS) rates in the *GLI1* mRNA low and high groups were 22.9% and 7.5% ($P = 0.086$; HR = 1.7 (0.91-3.9)), and those in the *GLI3* mRNA low and high groups were 23.2% and 6.7% ($P = 0.0027$; HR = 2.3 (1.5-7.1)), respectively. With regard to *GLI2* mRNA expression, no statistical significance was observed between high and low expression groups (OAS: $P = 0.32$; DFS: $P = 0.93$).

### Higher *GLI1* mRNA expression was correlated with significantly worse prognosis especially in stage III-IV lung adenocarcinoma patients

Subset analysis of patients with stage III-IV lung adenocarcinoma (n = 71) showed significantly worse clinical outcomes in both the higher *GLI1* and *GLI3* mRNA expression groups. In this cohort, 5-year OAS rates in the *GLI1* mRNA low and high groups were 34.0% and 0%, respectively ($P = 0.0012$, HR = 2.8 (1.9-13.5)), and 5-year DFS rates in these groups were 18.3% and 0%, respectively ($P = 0.027$, HR = 2.1 (1.1-7.2)). Although the difference in OAS between the *GLI3* mRNA expression groups was marginal (low, 33.4% *vs.* high, 7.7%, $P = 0.057$), 5-year DFS rates were statistically different between *GLI3* mRNA low and high groups (low, 17.6% *vs.* high, 7.7%, $P = 0.030$, HR = 1.9 (1.1-5.3)). Survival curves are presented in Figure 4.

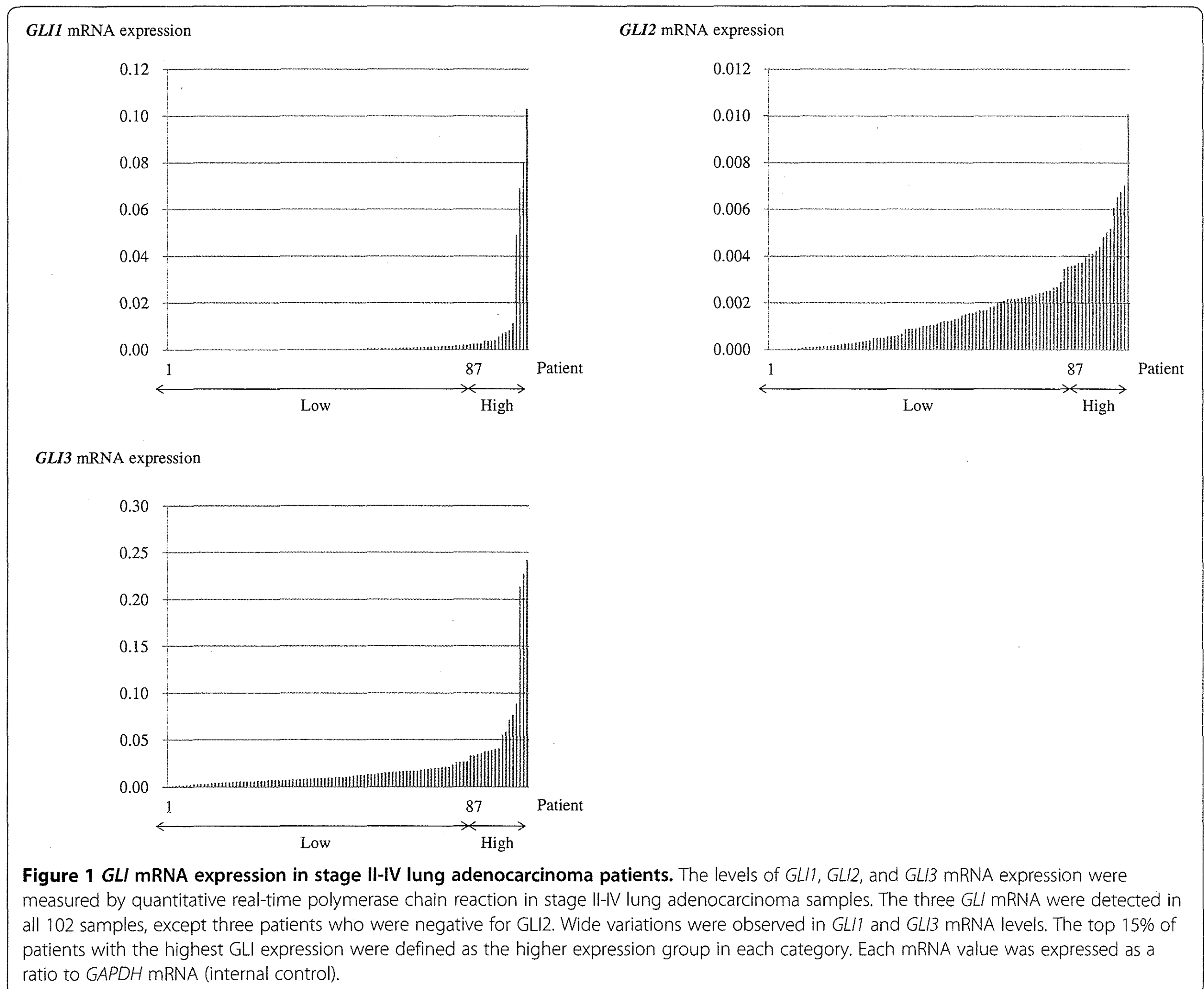


**Figure 1** ***GLI*** **mRNA expression in stage II-IV lung adenocarcinoma patients.** The levels of *GLI1*, *GLI2*, and *GLI3* mRNA expression were measured by quantitative real-time polymerase chain reaction in stage II-IV lung adenocarcinoma samples. The three *GLI* mRNA were detected in all 102 samples, except three patients who were negative for GLI2. Wide variations were observed in *GLI1* and *GLI3* mRNA levels. The top 15% of patients with the highest GLI expression were defined as the higher expression group in each category. Each mRNA value was expressed as a ratio to *GAPDH* mRNA (internal control).

### Higher *GLI1* mRNA expression was an independent prognostic factor in advanced lung adenocarcinoma

The results of multivariate analyses for each clinicopathological parameter are presented in Table 2. Higher *GLI1* mRNA expression was demonstrated to be an independent prognostic factor for poor patient prognosis ($P = 0.0030$, HR = 3.1 (1.5-6.2)) as well as tumor differentiation (well *vs.* poor) and tumor stage. No correlations were found for *GLI2* and *GLI3* mRNA expression with other clinicopathological features.

## Discussion

Previous reports on Hh molecules have already provided us with convincing information on its etiologic contribution to NSCLC [16]. A large percentage of primary NSCLC cell lines have been shown to express Hh target genes including GLI1, indicating constitutive activation of the Hh pathway in NSCLC cell lines [17]. It was recently shown that disrupting the Hh pathway inhibited NSCLC proliferation *in vitro* and abrogated tumor growth *in vivo* [18]. In clinical samples, immunohistochemical analysis of tissue microarray have revealed significant correlation among the Hh molecules in early-stage NSCLC and of lymph node metastasis with nuclear GLI1 immunolocalization in lung adenocarcinoma [19,20]. However, the impact of Hh signaling on patient survival has not been previously reported in NSCLC.

In this study, expression of the three *GLI* genes was demonstrated in almost all cases in this cohort, although their expression levels varied widely. Interestingly, some lung adenocarcinoma samples showed relatively high *GLI1* and *GLI3* mRNA expression, and their expression patterns were considerably correlated. Although GLI1 is considered to be the major effector in Hh signaling, the role of GLI3 is still controversial. Whereas GLI3 acts mainly as a repressor in normal Hh signaling [21], some reports have suggested its activator capacity in oncogenesis [22-23]. The present study also provides evidence of

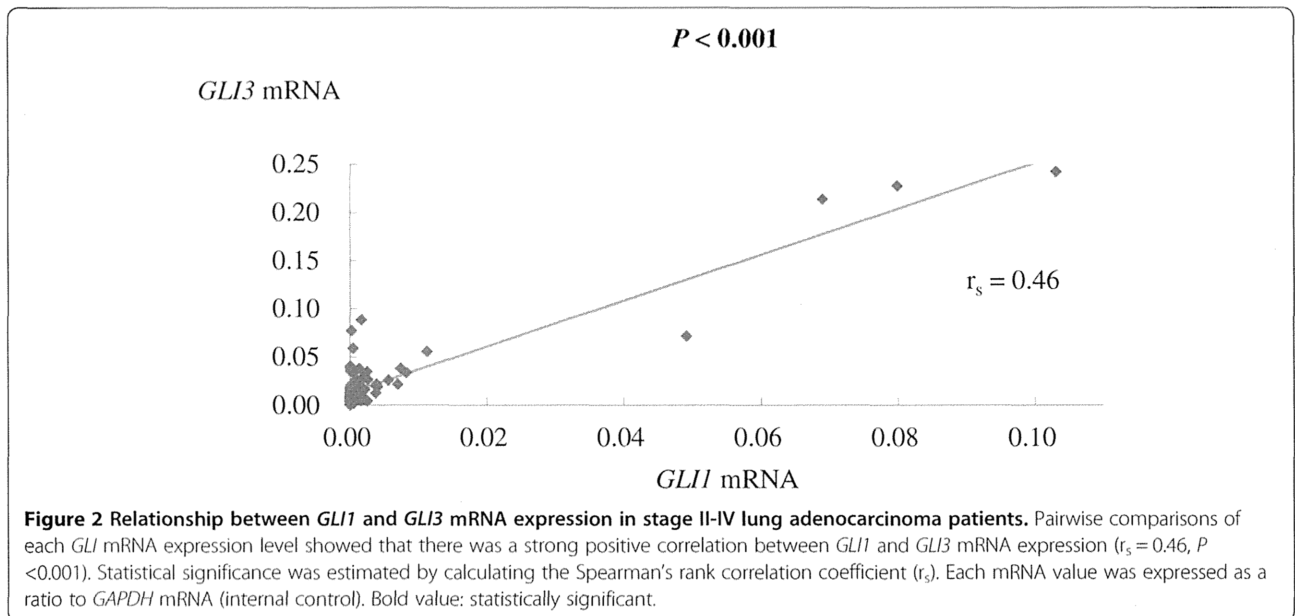


**Figure 2 Relationship between *GLI1* and *GLI3* mRNA expression in stage II-IV lung adenocarcinoma patients.** Pairwise comparisons of each *GLI* mRNA expression level showed that there was a strong positive correlation between *GLI1* and *GLI3* mRNA expression ($r_s = 0.46$, $P < 0.001$). Statistical significance was estimated by calculating the Spearman's rank correlation coefficient ($r_s$). Each mRNA value was expressed as a ratio to *GAPDH* mRNA (internal control). Bold value: statistically significant.

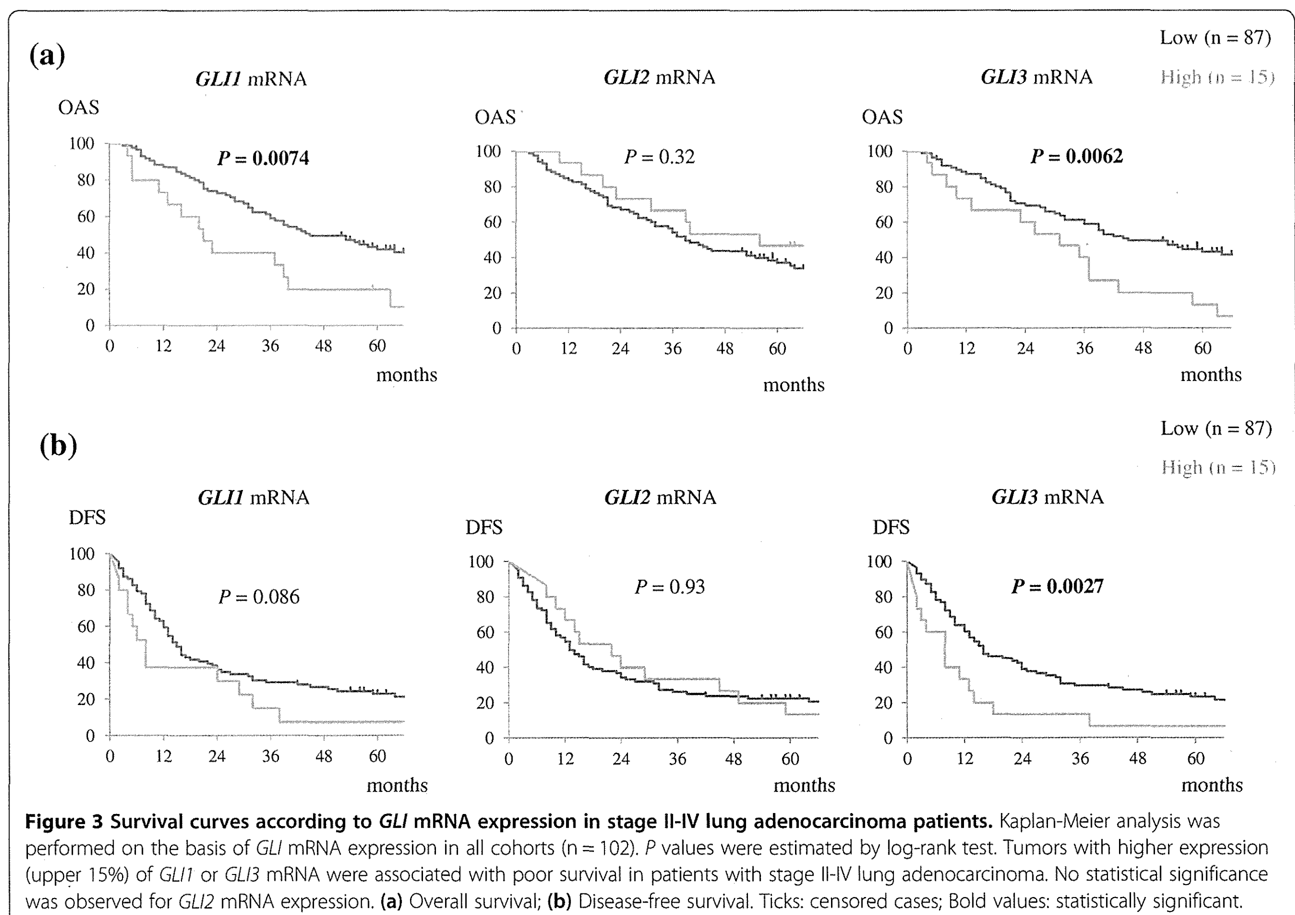


**Figure 3 Survival curves according to *GLI* mRNA expression in stage II-IV lung adenocarcinoma patients.** Kaplan-Meier analysis was performed on the basis of *GLI* mRNA expression in all cohorts (n = 102). *P* values were estimated by log-rank test. Tumors with higher expression (upper 15%) of *GLI1* or *GLI3* mRNA were associated with poor survival in patients with stage II-IV lung adenocarcinoma. No statistical significance was observed for *GLI2* mRNA expression. **(a)** Overall survival; **(b)** Disease-free survival. Ticks: censored cases; Bold values: statistically significant.

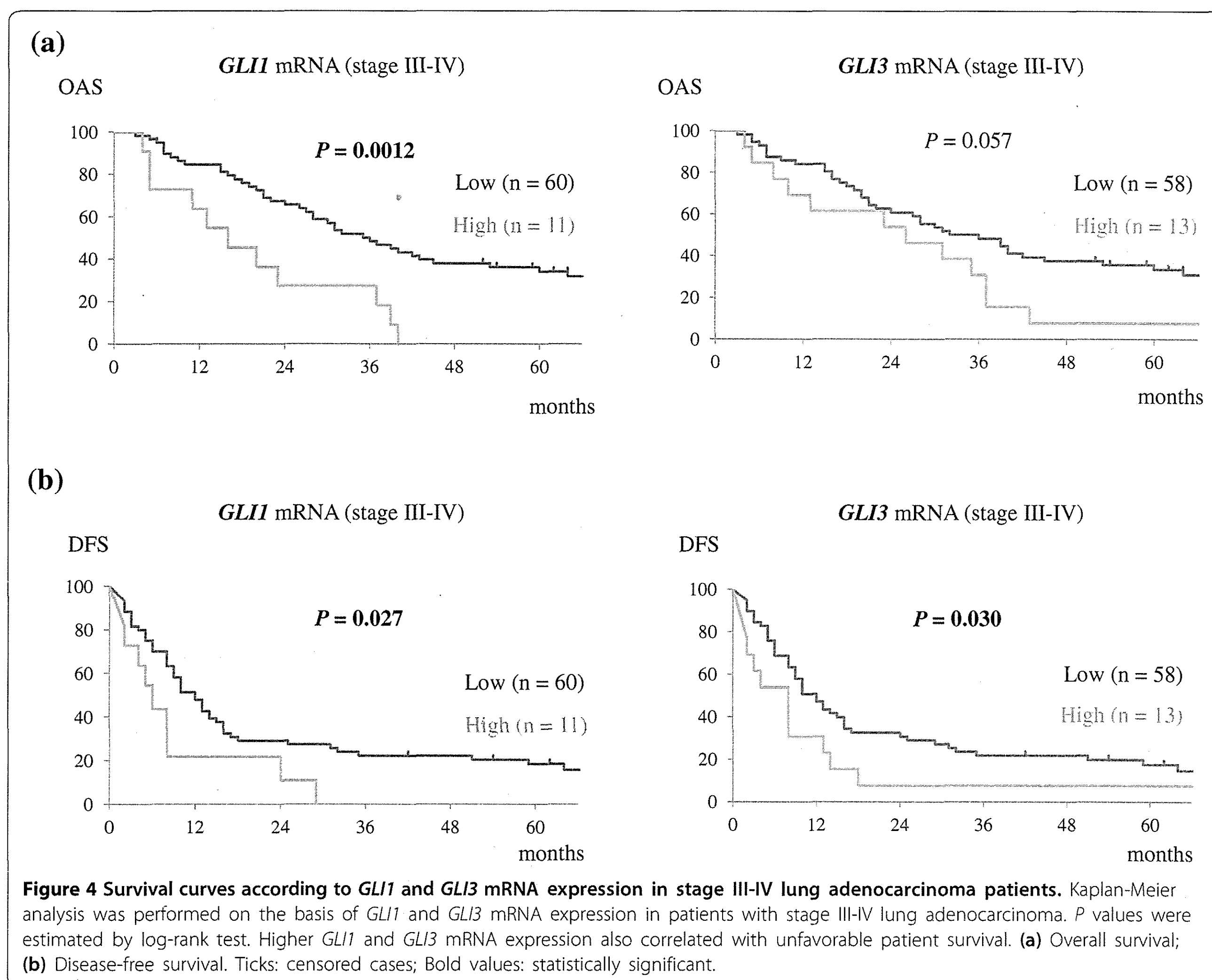


**Figure 4 Survival curves according to *GLI1* and *GLI3* mRNA expression in stage III-IV lung adenocarcinoma patients.** Kaplan-Meier analysis was performed on the basis of *GLI1* and *GLI3* mRNA expression in patients with stage III-IV lung adenocarcinoma. *P* values were estimated by log-rank test. Higher *GLI1* and *GLI3* mRNA expression also correlated with unfavorable patient survival. **(a)** Overall survival; **(b)** Disease-free survival. Ticks: censored cases; Bold values: statistically significant.

the capacity of GLI3 to activate Hh signaling in concert with GLI1 in advanced lung adenocarcinoma. With regard to GLI2, very little is characterized about its expression in NSCLC, and only correlation between nuclear GLI2 with sonic Hh expression in lung squamous cell carcinoma has been reported [19]. From our research, no relation between *GLI2* mRNA expression and *GLI1* mRNA level, *GLI3* mRNA level, or patient survival was detected, thus, the role of GLI2 in the pathogenesis of lung adenocarcinoma is still unclear.

While GLI1 is mainly regulated at the transcriptional level, GLI2 is considered as a latent transcriptional regulator that can be activated by Hh signaling. Modification of the N-terminus of GLI2 is a critical step in the regulation of the transcriptional activity of the protein [8]. This as well as the difference in cancer types may explain the discrepancies in GLI2 expression and patient outcomes between our results and previous ones [24]. Our research is based on the genetic expression of GLI, whereas for GLI2, it may be necessary to assess its functional protein level instead.

The mechanisms by which Hh signaling is involved in cancer pathogenesis are not uniform, and there are several different models of Hh signal transduction. Besides the most evident pattern of ligand-independent mutational Hh activation seen in basal cell carcinoma and medulloblastoma, ligand-dependent autocrine Hh activation in tumor cells has also been shown to be common in many cancers including NSCLC. In addition, ligand-dependent paracrine Hh signaling that involves the tumor microenvironment can lead to tumor development [25]. Crosstalk with other developmental cascades such as Wnt or Notch is also frequently observed [26]. This diversity of Hh signaling in cancer development in part explains the differences across cancer types [27,28].

Previous reports on expression profiles of Hh-related molecules in NSCLC specimens had focused mainly on the estimation of protein levels by immunohistochemistry [19,20], and their gene expression has not been thoroughly investigated. This study is unique in that thorough investigation was performed for the gene

**Table 2 Multivariate analysis of *GLI1* mRNA expression and other clinicopathological variables for patient survival**

| Variables | | Multivariate analysis | | |
|---|---|---|---|---|
| | | Hazard ratio | 95% CI | *P*-value[a] |
| *GLI1* mRNA expression | | | | |
| | Low | 1 | | **0.0030** |
| | High | 3.1 | 1.5-6.2 | |
| Age (years) | | | | |
| | ≤65 | 1 | | 0.23 |
| | >65 | 1.3 | 0.83-2.2 | |
| Gender | | | | |
| | Male | 1 | | 0.86 |
| | Female | 0.92 | 0.37-2.2 | |
| Cigarette smoking | | | | |
| | Never smoked | 1 | | 0.69 |
| | Smoker (ex or current) | 1.2 | 0.48-3.1 | |
| Tumor differentiation | | | | |
| | Well | 1 | | |
| | Moderately | 1.7 | 0.83-3.6 | 0.16 |
| | Poorly | 2.5 | 1.1-5.8 | **0.030** |
| Tumor stage | | | | |
| | II | 1 | | |
| | III | 2.5 | 1.5-4.5 | **<0.001** |
| | IV | 4.2 | 1.7-9.8 | **0.0027** |
| *EGFR* gene mutation status | | | | |
| | Wild-type | 1 | | |
| | Mutation (+) | 1.1 | 0.58-2.0 | 0.77 |
| Pre/postoperative chemo- and/or radiotherapy | | | | |
| | No | 1 | | |
| | Yes | 0.67 | 0.33-1.4 | 0.29 |

*Cox's proportional hazard model.
CI: confidence interval.
Bold values are statistically significant.

expression of all three GLI family members, which enabled us to directly compare their levels in clinical samples and to reveal the unexpected correlation between *GLI1* and *GLI3* mRNA expression. Interestingly, some GLI1/GLI3 highly expressing tumors (in this cohort, defined as the upper 15th percentile of patients for each mRNA expression) showed significantly poorer patient prognosis in advanced lung adenocarcinoma. Although a statistically significant negative correlation between survival and GLI1 expression has been shown in breast [29], esophageal [30], colorectal [31], and bladder cancers [32], this is the first study to demonstrate a clinical impact of GLI1 on patient survival in lung adenocarcinoma. Our speculation was further supported by the multivariate analysis, which showed that *GLI1* mRNA expression is an indicator for unfavorable prognosis. Although the independency of Hh molecule expression on patient survival has been previously shown in bladder and prostate cancers [32,33], we are the first to reveal that of GLI1 in lung adenocarcinoma, which is indicative of the involvement of Hh signaling in a subset of NSCLC. Investigation on GLI3 expression in tumor samples has been rare, and the possible clinical impact of GLI3 expression raised by our study needs further corroboration by additional research studies.

Irrespective of these novel observations, some limitations should be addressed in the present study. First, the validity of the cutoff of mRNA values was not ensured. Unfortunately, for our data, dichotomizing patients by either median or mean value of mRNA expression provided little information on survival difference, thus, we defined the upper 15% expression level as the cutoff. Further, patient backgrounds were diverse. Due to the retrospective nature of analysis of these advanced stage cancer cases, pre- or postoperative therapies were not

standardized and the completeness of resection was not fully warranted (for example, in stage IV patients, operations were mainly performed for salvage or diagnosis). We believe that these limitations can be overcome by expanding patient numbers and/or careful selection of patients with similar clinical backgrounds. Additionally, analyzing the expression profiles of GLI proteins as well as that of other Hh molecules in this cohort would further facilitate the deepening of our understanding.

With the advent of small molecule Hh inhibitors, such as vismodegib [34,35], recent interest has focused on the clinical application of Hh inhibitors in various types of tumors harboring aberrantly activated Hh signaling [25,36]. Based on our results, Hh inhibition may be a good candidate treatment for advanced lung adenocarcinoma. Because a similar study on stage I-II lung adenocarcinoma did not show any significance on patient survival [19], this signaling pathway may be involved more in tumor progression rather than tumorigenesis, and disrupting Hh may be more efficacious in the advanced stages of lung adenocarcinoma. Further studies including analysis of the expression of other Hh molecules or on other histological types of NSCLC are necessary to provide extended support of our findings.

## Conclusions

We found that *GLI1*, *GLI2*, and *GLI3* mRNA were expressed in almost all advanced stage lung adenocarcinoma tissue samples, and strong correlation was observed between *GLI1* and *GLI3* mRNA expression. While *GLI2* mRNA expression level was of little clinical significance, a subset of patients showing relatively high *GLI1*/*GLI3* expression showed poorer patient survival. In particular, higher *GLI1* mRNA expression was an independent prognostic factor for unfavorable prognosis. This is the first report to describe the clinical involvement of the GLI family of genes in advanced lung adenocarcinoma, and these results suggest that Hh signal inhibition is a good candidate therapeutic target for advanced stage lung adenocarcinoma, which provide an impetus for further investigations on the Hh signaling in NSCLC.

#### Abbreviations

CI: Confidence interval; DFS: Disease-free survival; GAPDH: Glyceraldehyde-3-phosphate dehydrogenase; Hh: Hedgehog; HR: Hazard ratio; NSCLC: Non-small cell lung cancer; OAS: Overall survival; $r_s$: Spearman's rank correlation coefficient; SD: Standard deviation.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

MI designed the study, carried out the experiments, performed the statistical analysis, and wrote the manuscript. NI, TS, and KS made substantial contributions to the collection of clinical samples and the preparation of the materials. SM provided the clinical data and expertise in experimental techniques. HD supervised the entire study and gave final approval of the version to be published. All authors have read and approved the final manuscript.

#### Acknowledgements

We would like to thank Ms. Seiko Sakai and Ms. Yuuka Taketani for their excellent secretarial assistance. No outside funding was used for this project.

Received: 13 August 2013 Accepted: 20 July 2014
Published: 8 August 2014

#### References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D: **Global cancer statistics.** *CA Cancer J Clin* 2011, **61**:69–90.
2. Alberg AJ, Samet JM: **Epidemiology of lung cancer.** *Chest* 2003, **123**(1 Suppl):21S–49S.
3. William WN Jr, Lin HY, Lee JJ, Lippman SM, Roth JA, Kim ES: **Revisiting stage IIIB and IV non-small cell lung cancer: analysis of the surveillance, epidemiology, and end results data.** *Chest* 2009, **136**:701–709.
4. Shaw AT, Yeap BY, Solomon BJ, Riely GJ, Gainor J, Engelman JA, Shapiro GI, Costa DB, Ou SH, Butaney M, Salgia R, Maki RG, Varella-Garcia M, Doebele RC, Bang YJ, Kulig K, Selaru P, Tang Y, Wilner KD, Kwak EL, Clark JW, Iafrate AJ, Camidge DR: **Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: a retrospective analysis.** *Lancet Oncol* 2011, **12**:1004–1012.
5. Ruiz I, Altaba A, Sanchez P, Dahmane N: **Gli and hedgehog in cancer: tumours, embryos and stem cells.** *Nat Rev Cancer* 2002, **2**:361–372.
6. Ingham PW, McMahon AP: **Hedgehog signaling in animal development: paradigms and principles.** *Genes Dev* 2001, **15**:3059–3087.
7. Matise MP, Joyner AL: **Gli genes in development and cancer.** *Oncogene* 1999, **18**:7852–7859.
8. Kasper M, Regl G, Frischauf AM, Aberger F: **GLI transcription factors: mediators of oncogenic Hedgehog signalling.** *Eur J Cancer* 2006, **42**:437–445.
9. Yang L, Xie G, Fan Q, Xie J: **Activation of the hedgehog-signaling pathway in human cancer and the clinical implications.** *Oncogene* 2010, **29**:469–481.
10. Teglund S, Toftgard R: **Hedgehog beyond medulloblastoma and basal cell carcinoma.** *Biochim Biophys Acta* 2010, **1805**:181–208.
11. Watkins DN, Berman DM, Baylin SB: **Hedgehog signaling: progenitor phenotype in small-cell lung cancer.** *Cell Cycle* 2003, **2**:196–198.
12. Park KS, Martelotto LG, Peifer M, Sos ML, Karnezis AN, Mahjoub MR, Bernard K, Conklin JF, Szczepny A, Yuan J, Guo R, Ospina B, Falzon J, Bennett S, Brown TJ, Markovic A, Devereux WL, Ocasio CA, Chen JK, Stearns T, Thomas RK, Dorsch M, Buonamici S, Watkins DN, Peacock CD, Sage J: **A crucial requirement for hedgehog signaling in small cell lung cancer.** *Nat Med* 2011, **17**:1504–1508.
13. Vestergaard J, Pedersen MW, Pedersen N, Ensinger C, Tumer Z, Tommerup N, Poulsen HS, Larsen LA: **Hedgehog signaling in small-cell lung cancer: frequent in vivo but a rare event in vitro.** *Lung Cancer* 2006, **52**:281–290.
14. Sobin L, Gospodarowicz M, Wittekind C: *TNM Classification of Malignant Tumours (UICC International Union Against Cancer).* 7th edition. Oxford: Wiley-Blackwell Co.; 2009.
15. Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC: *Pathology & Genetics of Tumours of the Lung, Pleura, Thymus and Heart (World Health Organization Classification of Tumours)*, Volume Volume 10. Lyon: IARC Press; 2004.
16. Velcheti V, Govindan R: **Hedgehog signaling pathway and lung cancer.** *J Thorac Oncol* 2007, **2**:7–10.
17. Yuan Z, Goetz JA, Singh S, Ogden SK, Petty WJ, Black CC, Memoli VA, Dmitrovsky E, Robbins DJ: **Frequent requirement of hedgehog signaling in non-small cell lung carcinoma.** *Oncogene* 2007, **26**:1046–1055.
18. Rodriguez-Blanco J, Schilling NS, Tokhunts R, Giambelli C, Long J, Liang Fei D, Singh S, Black KE, Wang Z, Galimberti F, Bejarano PA, Elliot S, Glassberg MK, Nguyen DM, Lockwood WW, Lam WL, Dmitrovsky E, Capobianco AJ, Robbins DJ: **The hedgehog processing pathway is required for NSCLC growth and survival.** *Oncogene* 2013, **32**:2335–2345.
19. Raz G, Allen KE, Kingsley C, Cherni I, Arora S, Watanabe A, Lorenzo CD, Edwards VD, Sridhar S, Hostetter G, Weiss GJ: **Hedgehog signaling pathway molecules and ALDH1A1 expression in early-stage non-small cell lung cancer.** *Lung Cancer* 2012, **76**:191–196.

20. Gialmanidis IP, Bravou V, Amanetopoulou SG, Varakis J, Kourea H, Papadaki H: **Overexpression of hedgehog pathway molecules and FOXM1 in non-small cell lung carcinomas.** *Lung Cancer* 2009, **66**:64–74.
21. Wang B, Fallon JF, Beachy PA: **Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb.** *Cell* 2000, **100**:423–434.
22. Iwasaki H, Nakano K, Shinkai K, Kunisawa Y, Hirahashi M, Oda Y, Onishi H, Katano M: **Hedgehog Gli3 activator signal augments tumorigenicity of colorectal cancer via upregulation of adherence-related genes.** *Cancer Sci* 2013, **104**:328–336.
23. Kang HN, Oh SC, Kim JS, Yoo YA: **Abrogation of Gli3 expression suppresses the growth of colon cancer cells via activation of p53.** *Exp Cell Res* 2012, **318**:539–549.
24. Im S, Choi HJ, Yoo C, Jung JH, Jeon YW, Suh YJ, Kang CS: **Hedgehog related protein expression in breast cancer: gli-2 is associated with poor overall survival.** *Korean J Pathol* 2013, **47**:116–123.
25. Gupta S, Takebe N, Lorusso P: **Targeting the hedgehog pathway in cancer.** *Ther Adv Med Oncol* 2010, **2**:237–250.
26. Geissler K, Zach O: **Pathways involved in drosophila and human cancer development: the notch, hedgehog, wingless, runt, and trithorax pathway.** *Ann Hematol* 2012, **91**:645–669.
27. Bertrand FE, Angus CW, Partis WJ, Sigounas G: **Developmental pathways in colon cancer: crosstalk between WNT, BMP, hedgehog and notch.** *Cell Cycle* 2012, **11**:4344–4351.
28. Santoni M, Burattini L, Nabissi M, Morelli MB, Berardi R, Santoni G, Cascinu S: **Essential role of Gli proteins in glioblastoma multiforme.** *Curr Protein Pept Sci* 2013, **14**:133–140.
29. ten Haaf A, Bektas N, von Serenyi S, Losen I, Arweiler EC, Hartmann A, Knuchel R, Dahl E: **Expression of the glioma-associated oncogene homolog (GLI) 1 in human breast cancer is associated with unfavourable overall survival.** *BMC Cancer* 2009, **9**:298.
30. Mori Y, Okumura T, Tsunoda S, Sakai Y, Shimada Y: **Gli-1 expression is associated with lymph node metastasis and tumor progression in esophageal squamous cell carcinoma.** *Oncology* 2006, **70**:378–389.
31. Ding YL, Zhou Y, Xiang L, Ji ZP, Luo ZH: **Expression of glioma-associated oncogene homolog 1 is associated with invasion and postoperative liver metastasis in colon cancer.** *International journal of medical sciences* 2012, **9**:334–338.
32. He HC, Chen JH, Chen XB, Qin GQ, Cai C, Liang YX, Han ZD, Dai QS, Chen YR, Zeng GH, Zhu JG, Jiang FN, Zhong WD: **Expression of hedgehog pathway components is associated with bladder cancer progression and clinical outcome.** *Pathol Oncol Res* 2012, **18**:349–355.
33. Azoulay S, Terry S, Chimingqi M, Sirab N, Faucon H, Gil Diez De Medina S, Moutereau S, Maille P, Soyeux P, Abbou C, Salomon L, Vacherot F, de la Taille A, Loric S, Allory Y: **Comparative expression of hedgehog ligands at different stages of prostate carcinoma progression.** *J Pathol* 2008, **216**:460–470.
34. Henkin RI: **Vismodegib in advanced basal-cell carcinoma.** *N Engl J Med* 2012, **367**:970. Author reply 970-971.
35. Berlin J, Bendell JC, Hart LL, Firdaus I, Gore I, Hermann RC, Mulcahy MF, Zalupski MM, Mackey HM, Yauch RL, Graham RA, Bray GL, Low JA: **A randomized phase II trial of vismodegib versus placebo with FOLFOX or FOLFIRI and bevacizumab in patients with previously untreated metastatic colorectal cancer.** *Clin Cancer Res* 2013, **19**:258–267.
36. Neal JW, Sequist LV: **Exciting new targets in lung cancer therapy: ALK, IGF-1R, HDAC, and Hh.** *Curr Treat Options in Oncol* 2010, **11**:36–44.

doi:10.1186/1477-7819-12-253
**Cite this article as:** Ishikawa *et al.*: Expression of the *GLI* family genes is associated with tumor progression in advanced lung adenocarcinoma. *World Journal of Surgical Oncology* 2014 **12**:253.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

Lin et al. BMC Genomics 2014, 15:1079
http://www.biomedcentral.com/1471-2164/15/1079

BMC Genomics

RESEARCH ARTICLE Open Access

# Genes suppressed by DNA methylation in non-small cell lung cancer reveal the epigenetics of epithelial–mesenchymal transition

Steven H Lin[1*], Jing Wang[2*], Pierre Saintigny[3], Chia-Chin Wu[4], Uma Giri[5], Jing Zhang[1], Toshi Menju[5], Lixia Diao[1], Lauren Byers[5], John N Weinstein[2], Kevin R Coombes[6], Luc Girard[7], Ritsuko Komaki[1], Ignacio I Wistuba[8], Hiroshi Date[9], John D Minna[7] and John V Heymach[5*]

## Abstract

**Background:** DNA methylation is associated with aberrant gene expression in cancer, and has been shown to correlate with therapeutic response and disease prognosis in some types of cancer. We sought to investigate the biological significance of DNA methylation in lung cancer.

**Results:** We integrated the gene expression profiles and data of gene promoter methylation for a large panel of non-small cell lung cancer cell lines, and identified 578 candidate genes with expression levels that were inversely correlated to the degree of DNA methylation. We found these candidate genes to be differentially methylated in normal lung tissue versus non-small cell lung cancer tumors, and segregated by histologic and tumor subtypes. We used gene set enrichment analysis of the genes ranked by the degree of correlation between gene expression and DNA methylation to identify gene sets involved in cellular migration and metastasis. Our unsupervised hierarchical clustering of the candidate genes segregated cell lines according to the epithelial-to-mesenchymal transition phenotype. Genes related to the epithelial-to-mesenchymal transition, such as *AXL*, *ESRP1*, *HoxB4*, and *SPINT1/2*, were among the nearly 20% of the candidate genes that were differentially methylated between epithelial and mesenchymal cells. Greater numbers of genes were methylated in the mesenchymal cells and their expressions were upregulated by 5-azacytidine treatment. Methylation of the candidate genes was associated with erlotinib resistance in wild-type *EGFR* cell lines. The expression profiles of the candidate genes were associated with 8-week disease control in patients with wild-type *EGFR* who had unresectable non-small cell lung cancer treated with erlotinib, but not in patients treated with sorafenib.

**Conclusions:** Our results demonstrate that the underlying biology of genes regulated by DNA methylation may have predictive value in lung cancer that can be exploited therapeutically.

**Keywords:** DNA methylation, Epithelial-mesenchymal transition, Erlotinib, Lung cancer

* Correspondence: shlin@mdanderson.org; jingwang@mdanderson.org; jheymach@mdanderson.org
[1]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Unit 097, Houston, TX 77030, USA
[2]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Unit 097, Houston, TX 77030, USA
[5]Department of Thoracic/Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Unit 097, Houston, TX 77030, USA
Full list of author information is available at the end of the article

BioMed Central

© 2014 Lin et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

## Background

The methylation of DNA is involved in the control of chromatin folding, protein complex assembly, and gene expression. It is a normal process of cellular development, aging, paternal/maternal genetic imprinting, and X-chromosome inactivation [1]. DNA methylation is the only known heritable modification of DNA, and is aberrant in cancer, involving both excessive and insufficient methylation of DNA regions of cytosine–guanine bonds (CpG), which then alters the expression of certain genes [2]. The aberrant hypermethylation of dense CpG sites in many cancers occurs at tumor suppressor genes in a non-random, tumor-specific pattern [3].

Given the possible role of aberrant DNA methylation in cancer initiation and progression, a significant effort has been directed to identify DNA methylation biomarkers in cancer and to use such markers to predict therapeutic responses and disease prognosis [4]. A notable example is the association between the silencing of the *O6-methylguanine-DNA methyltransferase* (*MGMT*) gene and the improved response to temozolomide and radiation in patients with glioblastoma [5]. Others include the association of the methylation of *PTEN*, *IGFBP-3*, and various DNA repair enzymes (hMLH1, MRN, BRCA1, ATM, and the *FANC* genes) with either resistance (*PTEN*, *IGFBP-3*) or sensitivity (DNA repair enzymes) to chemotherapy or radiation [6-10]. For lung cancer, DNA methylation of genes may be useful for assessing cancer risk based on the analysis of the sputum of smokers [11], and may be prognostic in early-stage lung cancer [12,13].

To better understand the DNA methylation changes associated with gene expression in lung cancer, we determined the methylation status of more than 27,000 CpG sites across 14,000 genes at selective promoter regions in 73 non-small cell lung cancer (NSCLC) cell lines using the Illumina HumanMethylation27 BeadChip. We integrated this dataset with the gene expression profiles of the same panel of cell lines to identify the genes with expression levels that were most affected by DNA promoter methylation. We sought to determine the biologic significance of genes that were significantly repressed in association with methylation in silico, as well as experimentally, both *in vitro* and *in vivo*.

## Results and discussion

### Identifying genes aberrantly associated with DNA methylation in NSCLC Cell Lines

For each of the 73 NSCLC cell lines examined, we analyzed the relationship between the degree of methylation (beta value 0 to 1, with 1 being completed methylated) of a particular gene and the expression level of that gene. We sought to identify genes that were significantly repressed in association with DNA methylation, which we subsequently refer to as SRAMs, a term first used in the study of breast cancer [14]. We computed the Spearman correlation of the relationship between methylation and gene expression for all the cell lines, assuming that genes negatively regulated by DNA promoter methylation have a negative Spearman rho value. This approach would only identify differentially regulated genes and not genes that are fully methylated or unmethylated in these cell lines. We first tested our approach by looking at the Spearman correlation for *CDKN2A*, a gene well known to be frequently downregulated by promoter methylation in NSCLC [15]. In the Illumina assay, four probes assessed four CpG sites within the *CDKN2A* promoter. We found that *CDKN2A* was differentially methylated in the various cell lines, and the degree of methylation was significantly and inversely correlated with gene expression (Figure 1A). This relationship was preserved in each of the four probes examined (Additional file 1: Figure S1).

We expanded our analysis to the rest of the probes. We looked at the distribution of rho values for probes located within or outside of CpG islands (CpGi). Rho values of CpGi probes had a mean distribution that was significantly shifted to the left of 0 compared to that of non-CpGi probes ($p < 0.00001$, Figure 1B). For both sets of probes, an excess was seen in the experimental dataset at the rho cut-off of −0.5 that was more than expected when compared to a random distribution ($p < 0.00001$, Figure 1B). This was not observed for probes that had a positive correlation with gene expression. There were also no differences regarding whether the probes corresponded to CpGi or non-CpGi genes (3.3% vs. 3.5%, $P = 0.48$). Greater than 80% of these probes lie within 500 bp of the transcription starting site (TSS), corresponding to sites under transcriptional control by DNA methylation (Figure 1C). We found 750 probes that correspond to 578 unique SRAMs, genes with expression levels that were inversely correlated with the degree of DNA methylation (Additional file 2: Table S1). We found the same pattern as that of *CDKN2A* for other genes known to be regulated by methylation (*CDKN2B*, *MGMT*), as well as for many novel genes not previously known to be regulated by methylation (Additional file 3: Figure S2). We validated a set of SRAMs using pyrosequencing and quantitative real-time polymerase chain reaction (PCR) in 26 cell lines (Figure 2, Additional file 4: Figure S3).

### SRAMs differentiate tumor from normal lung and distinguish tumor subtypes

As aberrant DNA methylation is one of the defining characteristics of cancer, SRAMs should reflect alterations that discriminate cancer from normal tissue and reflect the underlying tumor biology. We evaluated the methylation status of SRAMs in a set of 110 lung tumors from a Japanese cohort, hereafter termed the Kyoto tumor set, which included 20 unmatched normal lung tissues arrayed using the Illumina HumanMethylation450 BeadChip.

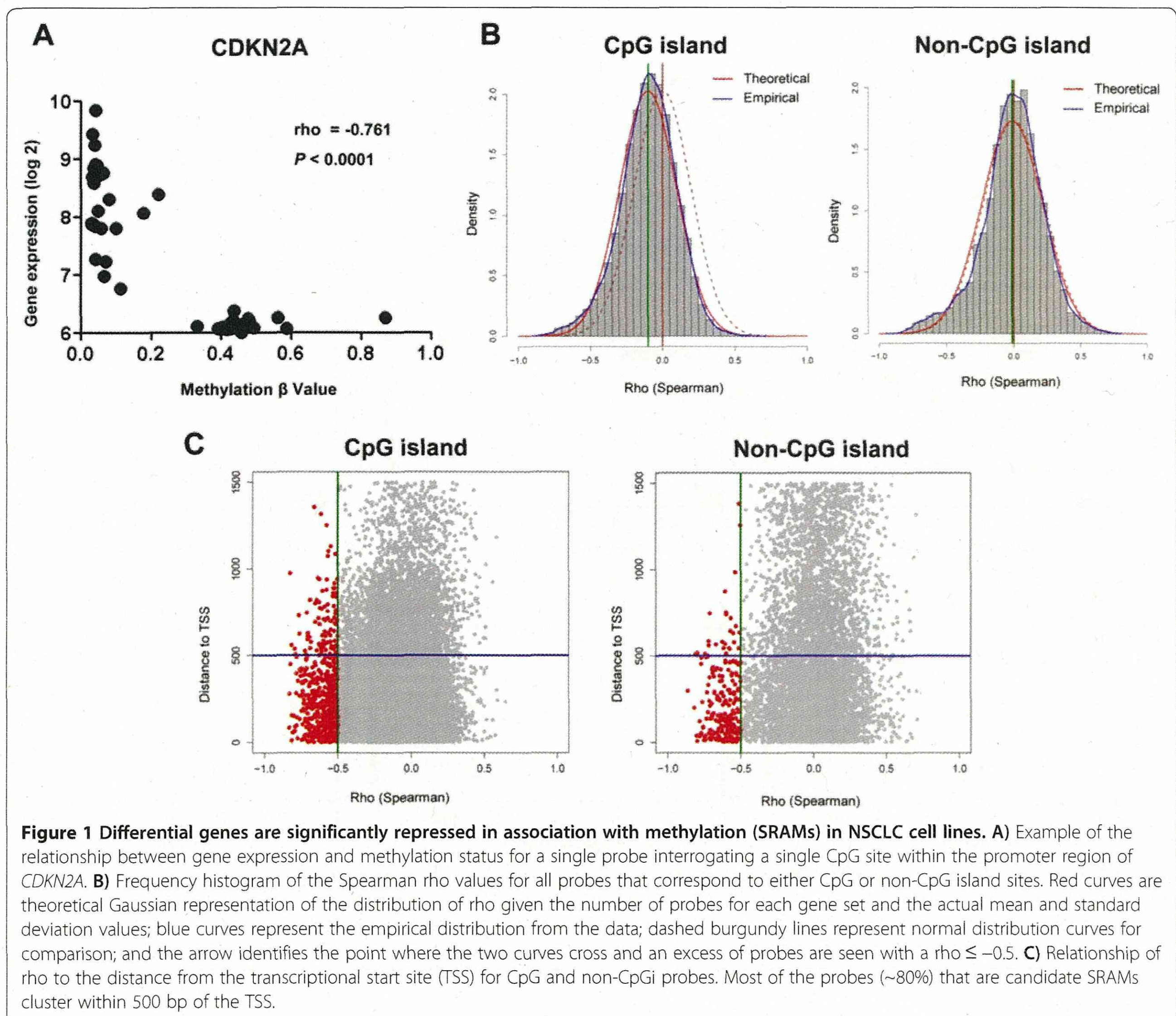


**Figure 1 Differential genes are significantly repressed in association with methylation (SRAMs) in NSCLC cell lines. A)** Example of the relationship between gene expression and methylation status for a single probe interrogating a single CpG site within the promoter region of *CDKN2A*. **B)** Frequency histogram of the Spearman rho values for all probes that correspond to either CpG or non-CpG island sites. Red curves are theoretical Gaussian representation of the distribution of rho given the number of probes for each gene set and the actual mean and standard deviation values; blue curves represent the empirical distribution from the data; dashed burgundy lines represent normal distribution curves for comparison; and the arrow identifies the point where the two curves cross and an excess of probes are seen with a rho ≤ –0.5. **C)** Relationship of rho to the distance from the transcriptional start site (TSS) for CpG and non-CpGi probes. Most of the probes (~80%) that are candidate SRAMs cluster within 500 bp of the TSS.

Since the cell line SRAM dataset was generated from the HumanMethylation27 platform, we took 637 probes (520 unique genes) that overlapped between the two Illumina platforms for the analysis. We found that SRAMs discriminated NSCLC from normal lung tissue, with clusters of probes that were either predominantly methylated in normal tissue or methylated in tumors (Figure 3A). SRAMs also distinguished squamous carcinoma from adenocarcinoma, the two major histological subtypes of NSCLC, as well as lung adenocarcinomas with or without an *EGFR* mutation (Figure 3B). These observations were corroborated using SRAM expression profiles of lung adenocarcinomas included in the Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma [16]. SRAMs preferentially separated *EGFR* mutants from wild-type tumors (Figure 3C and D). In the Kyoto tumor set, patients that separated into these clusters had significant survival differences, similar to that seen for the Director's Challenge tumor set (Figure 3E).

## SRAMs and the epithelial-to-mesenchymal transition (EMT) phenotype

To further define the functional significance of genes negatively regulated by DNA promoter methylation, we performed a gene set enrichment analysis (GSEA) using the Spearman correlation rho value as the ranking variable [17]. The top gene sets that included genes negatively regulated by DNA promoter methylation had been obtained from diverse cancer datasets and were involved in various biologic functions, such as cell migration in bladder cancer and resistance to gefitinib in NSCLC, and included genes downregulated with E-cadherin knockdown in human breast mammary epithelial (HBME) cell lines, genes differentially expressed in metastatic melanoma, genes

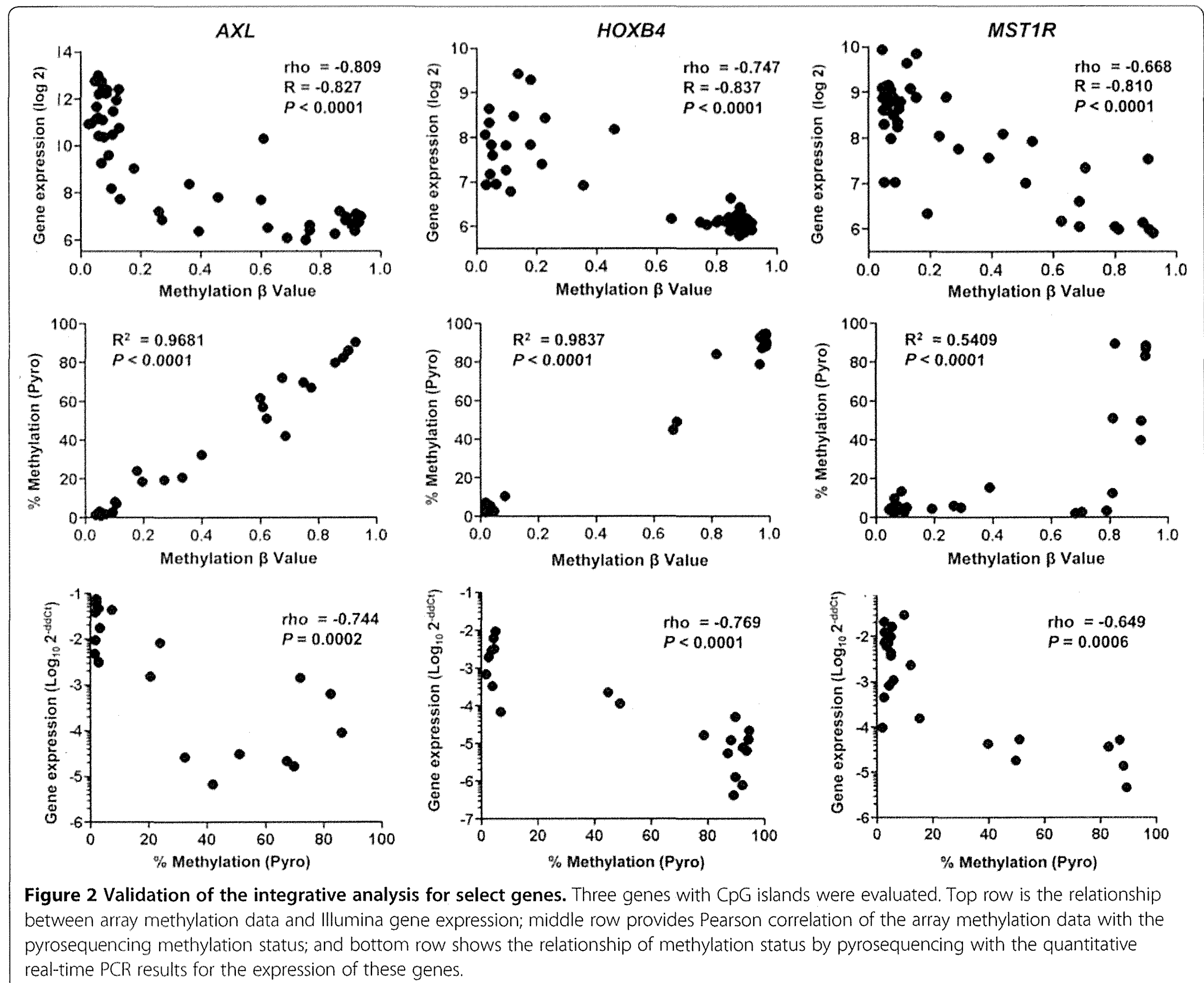


**Figure 2 Validation of the integrative analysis for select genes.** Three genes with CpG islands were evaluated. Top row is the relationship between array methylation data and Illumina gene expression; middle row provides Pearson correlation of the array methylation data with the pyrosequencing methylation status; and bottom row shows the relationship of methylation status by pyrosequencing with the quantitative real-time PCR results for the expression of these genes.

methylated in glioblastoma and pancreatic cancer, and genes differentially regulated in luminal vs. basal/mesenchymal breast cancer cells (Additional file 5: Table S2).

We postulated that biologic functions identified by the GSEA were related to the process of EMT, a developmental and adaptive cellular process that has been associated with resistance to cancer therapies and regulation of metastasis [18]. We performed a hierarchical cluster analysis of the cell lines using the SRAMs, and overlaid the relative protein expression of E-cadherin, a cell adhesion molecule that is downregulated during EMT and which plays a key role in the signaling and regulation of EMT [19], along with the expression of EMT-related genes *ZEB1, VIM, Twist1, FN1, CDH2,* and *CDH1*. We found that the identification of SRAMs clustered cells into two groups: "epithelial-like (E)" cells, which expressed high E-cadherin levels; and "mesenchymal-like (M)" cells, which expressed low E-cadherin levels (Figure 4A). To identify genes within the SRAMs that distinguish E cells from M cells, we performed a Wilcoxon rank sum test. Out of the 578 unique genes, 111 (19%) were differentially methylated between E and M cells, with a *P*-value < 0.001. We called these genes EMT-SRAMs. A majority of EMT-SRAMs (N = 88) were methylated in M cells compared to E cells (Figure 4B, Additional file 6: Table S3). A few of the EMT-SRAMs, including *CDH1*, were already known to be epigenetically regulated during the process of EMT [19,20]. However, the epigenetic regulation of the remaining EMT-SRAMs had not been established, including that of *SPINT1 (HAI-1)* [21], *SYK* [22], *MST1R(RON)* [22] and *ESRP1* [23]. Interestingly, one of these genes is *AXL*, which encodes a receptor tyrosine kinase that is associated with EMT and with resistance to EGFR tyrosine kinase inhibition [24,25]. We found that the expression of *AXL* was inversely correlated with promoter methylation (Figure 2), and its promoter methylation was positively correlated with

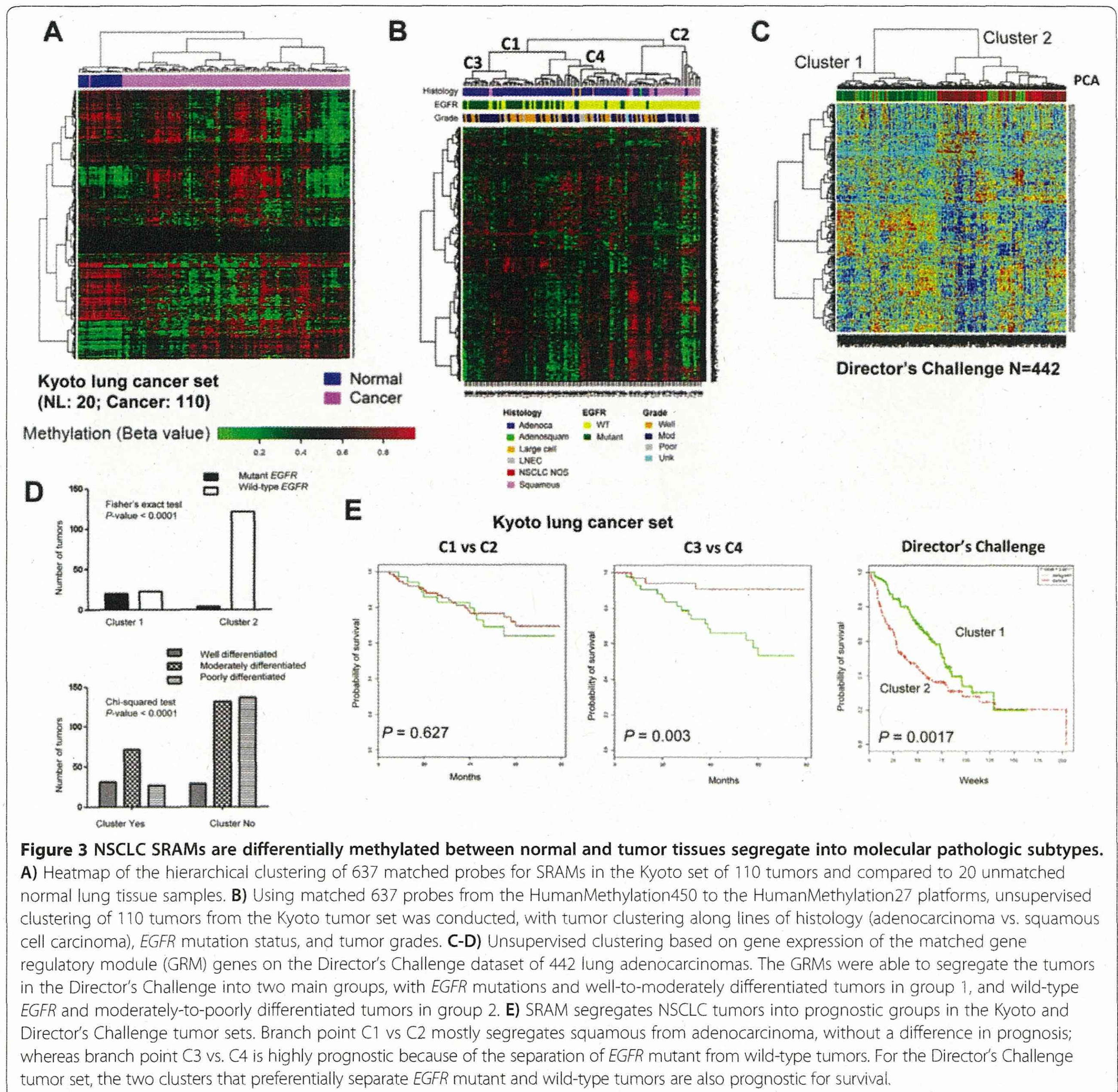


**Figure 3 NSCLC SRAMs are differentially methylated between normal and tumor tissues segregate into molecular pathologic subtypes.** **A)** Heatmap of the hierarchical clustering of 637 matched probes for SRAMs in the Kyoto set of 110 tumors and compared to 20 unmatched normal lung tissue samples. **B)** Using matched 637 probes from the HumanMethylation450 to the HumanMethylation27 platforms, unsupervised clustering of 110 tumors from the Kyoto tumor set was conducted, with tumor clustering along lines of histology (adenocarcinoma vs. squamous cell carcinoma), *EGFR* mutation status, and tumor grades. **C-D)** Unsupervised clustering based on gene expression of the matched gene regulatory module (GRM) genes on the Director's Challenge dataset of 442 lung adenocarcinomas. The GRMs were able to segregate the tumors in the Director's Challenge into two main groups, with *EGFR* mutations and well-to-moderately differentiated tumors in group 1, and wild-type *EGFR* and moderately-to-poorly differentiated tumors in group 2. **E)** SRAM segregates NSCLC tumors into prognostic groups in the Kyoto and Director's Challenge tumor sets. Branch point C1 vs C2 mostly segregates squamous from adenocarcinoma, without a difference in prognosis; whereas branch point C3 vs. C4 is highly prognostic because of the separation of *EGFR* mutant from wild-type tumors. For the Director's Challenge tumor set, the two clusters that preferentially separate *EGFR* mutant and wild-type tumors are also prognostic for survival.

E-cadherin expression (Additional file 7: Figure S4), indicating that *AXL* may be epigenetically regulated during EMT.

To further determine whether EMT-SRAMs define an epigenetically regulated group of genes relevant to EMT, we compared our gene list to a variety of publically available gene expression signatures from studies that experimentally manipulated cells to undergo EMT. Specifically, we examined datasets from prostate cancer cells transduced with *SNAI1* [26], a NSCLC epithelial cell line (H358) transduced with *SNAI1*, *ZEB1* or stimulated with TGF-beta [27], and HBME cells transduced with *Twist* [28]. We applied hypergeometric statistics to test the significance of overlap between the entire dataset of SRAMs (578 genes) and the EMT-SRAMs (111 genes). We found that the SRAMs significantly overlapped with the genes in the published datasets. On average, there was a 16% overlap between the SRAMs and published EMT gene signatures. EMT-SRAMs had greater overlap with the published EMT gene signatures, ranging from 18% in the breast cancer cell line signature to 48% in the H358 cells treated with TGF-beta (Additional file 8: Table S4). The majority of overlapping genes were downregulated, which is consistent with our observation of a greater proportion of genes being methylated in M cell lines. There appeared to be a greater overlap between genes from

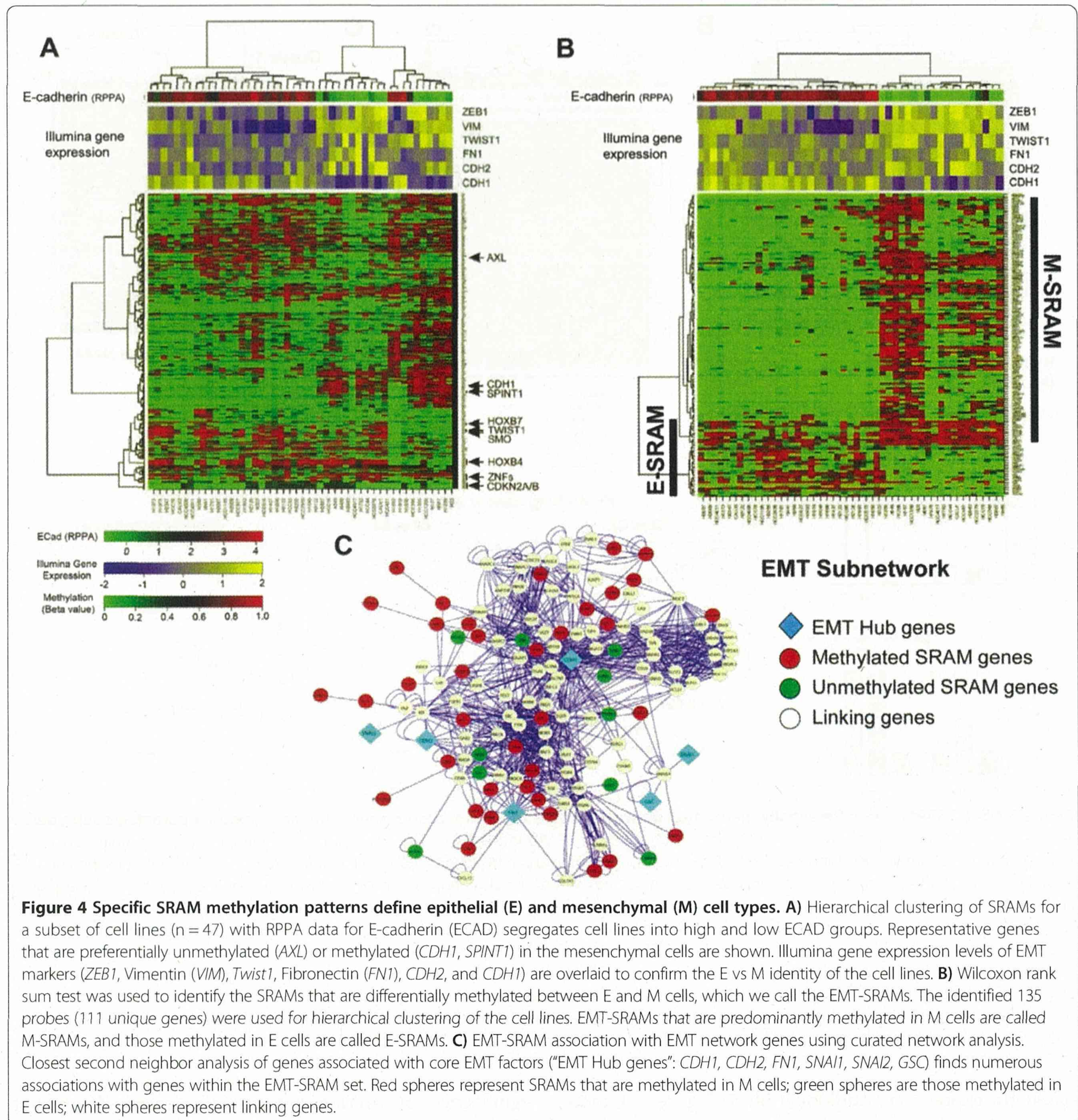


**Figure 4 Specific SRAM methylation patterns define epithelial (E) and mesenchymal (M) cell types. A)** Hierarchical clustering of SRAMs for a subset of cell lines (n = 47) with RPPA data for E-cadherin (ECAD) segregates cell lines into high and low ECAD groups. Representative genes that are preferentially unmethylated (*AXL*) or methylated (*CDH1, SPINT1*) in the mesenchymal cells are shown. Illumina gene expression levels of EMT markers (*ZEB1*, Vimentin (*VIM*), *Twist1*, Fibronectin (*FN1*), *CDH2*, and *CDH1*) are overlaid to confirm the E vs M identity of the cell lines. **B)** Wilcoxon rank sum test was used to identify the SRAMs that are differentially methylated between E and M cells, which we call the EMT-SRAMs. The identified 135 probes (111 unique genes) were used for hierarchical clustering of the cell lines. EMT-SRAMs that are predominantly methylated in M cells are called M-SRAMs, and those methylated in E cells are called E-SRAMs. **C)** EMT-SRAM association with EMT network genes using curated network analysis. Closest second neighbor analysis of genes associated with core EMT factors ("EMT Hub genes": *CDH1, CDH2, FN1, SNAI1, SNAI2, GSC*) finds numerous associations with genes within the EMT-SRAM set. Red spheres represent SRAMs that are methylated in M cells; green spheres are those methylated in E cells; white spheres represent linking genes.

H358 cells and the SRAMs or the EMT-SRAMs than either the prostate or breast cancer datasets. This suggests that while there could be a common core group of genes for both the SRAMs and the EMT-associated SRAMs, the tissue-specific SRAMs may reflect the underlying biology of the tissue of origin.

Using a curated network extracted from Pathway Commons [29], which consists of 11,570 genes and over a million biological interactions, we performed a network analysis using markers of EMT, including 8 transcription factors and major regulators of EMT (*TWIST, SNAIL, SLUG, GSC, FOXC1, FOXC2, ZEB1* and *ZEB2*), and three recognized markers of EMT (*CDH1, CDH2* and *FN*), and the EMT-SRAM gene list. We were able to map 73/111 EMT-SRAMs. The pair-wise shortest distance between markers of EMT with EMT-SRAMs was calculated using Dijkastra's algorithm [30]: 51/73 EMT-SRAMs (~70%) in the network had the shortest path lengths of 1 and 2, connecting at least one of the 11 markers of EMT (Additional file 9: Figure S5). This

implies that a large proportion of the EMT-SRAMs are functionally related to regulators of EMT. Using hypergeometric statistics, we created a core EMT subnetwork ("EMT hub genes") that accounted for all biological interactions between markers of EMT-linking genes (genes that are functionally related to EMT but are not core EMT factors or part of the EMT-SRAMs), and EMT-SRAMs (Figure 4C). While 2.8% of the genes in the entire curated network were cited in the literature as being related to EMT, 24.8% of the genes in the subnetwork were cited thusly ($P < 0.0001$, hypergeometric testing). This indicates that there was a significant enrichment for known markers of EMT in our subnetwork, as well as novel interactions.

Taken together, these findings demonstrate that a subset of NSCLC SRAMs is related to EMT, and that epigenetic regulation through DNA methylation may play an important role in the process of EMT.

### EMT-related genes are epigenetically regulated

To determine whether genes associated with EMT were epigenetically regulated, we created an EMT gene set based on the E-cadherin differential expression (high or low) in our panel of NSCLC cell lines. A total of 407 genes were ≥ 4-fold differentially expressed between these two groups and thus were included in the EMT gene set. We then identified the publicly available gene expression profiles of NSCLC cell lines before and after treatment with the hypomethylating agent 5-azacytidine (5AZA) from a single study [31]. We performed a GSEA using the A549 mesenchymal cell line. We found the EMT gene set to be negatively enriched when genes were ranked according to the rho value between gene expression and the average beta value in our integrative analysis (enrichment score −0.571, $P < 0.001$). We found the EMT gene set to be positively enriched when genes were ranked according to the $\log_2$ fold-change before and after treatment with 5AZA (enrichment score −0.612, $P < 0.001$) (Figure 5A). We found statistical significance ($P < 0.001$, hypergeometric testing) in the overlap between genes that were overexpressed in response to 5AZA and genes that were regulated by DNA promoter methylation and were represented in the EMT gene set enrichment.

Since the M cells have proportionally more SRAMs compared to E cells, we next determined whether 5AZA was upregulating a larger number of genes in M cells (A549, H157, H460, and H1299) compared to E cells (H1819, H1993, and H2347). We performed a GSEA that included the EMT gene set, with genes ranked according to the $\log_2$ fold-change before and after treatment with 5AZA, in all 7 NSCLC cell lines [31]. While the enrichment of the EMT gene set was significant for all but one E cell (H2347), the number of genes induced by 5AZA with a fold-change ≥ 2 and the enrichment scores tended to be higher in M cells (A549, H157, H460, H1299, Additional file 10: Figure S6A) compared to E cells (H1819, H1993, H2347, Additional file 10: Figure S6B) (Figure 5B). We confirmed these results with the dataset from Heller et al. [32] where two of the three cell lines (A549 and H1993) overlapped with that in the Shames et al. dataset. We performed GSEA analysis of these two lines and compared the 5AZA and 5AZA + trichostatin (TSA) treated cells compared to untreated cells. We found similar enrichment of the EMT gene set for the A549 cell line but not for the E cell type H1993 (Additional file 11: Table S5).

To further test the hypothesis that M-related SRAMs are more likely than E-related SRAMs to exhibit increased expression in M cells after 5AZA treatment, we used a Bayesian statistical method (Additional file 12: Text S1; Additional file 13: Table S6; Additional file 14: Table S7). We assigned each gene in each cell line to one of three categories (up, down, or constant), as defined by a change in expression that was at least a 2-fold change in response to 5AZA. We calculated the posterior probability that more genes increase (rather than decrease) expression after treatment with 5AZA, by the type of cell line, using either an uninformative prior or a conservative prior that assumes most genes do not change expression. We found that on average, the probability of induced gene expression in the M cells is high for the M-related SRAM category of genes, and is high for the E cells in the E-related SRAM category of genes (Figure 5C). These results corroborate the in silico data that indicated that more genes were methylated in the M cells and were inducible upon 5AZA treatment.

### Association between SRAM methylation and resistance to erlotinib in wild-type *EGFR/KRAS* NSCLC cell lines

Since the SRAMs segregate the cell lines into E and M cell types, and EMT has been implicated in erlotinib resistance [33,34], we determined whether the two clusters of NSCLC cell lines defined by SRAMs (methylated and unmethylated EMT-SRAMs) could be associated with erlotinib resistance. Since *EGFR* mutation is known to be associated with sensitivity to erlotinib, and *KRAS* mutations are associated with resistance, we focused on the group of wild-type *EGFR/KRAS* cell lines. We found that the half maximal inhibitory concentration ($IC_{50}$) for erlotinib was significantly higher in cell lines that segregated to clusters with methylated SRAMs compared to those that segregated to clusters with unmethylated SRAMs (Figures 6A and B).

### SRAM expression profiles in tumors segregate along E vs M expression patterns and erlotinib response

We next determined whether the SRAM profile was associated with clinical benefit in patients. Since we

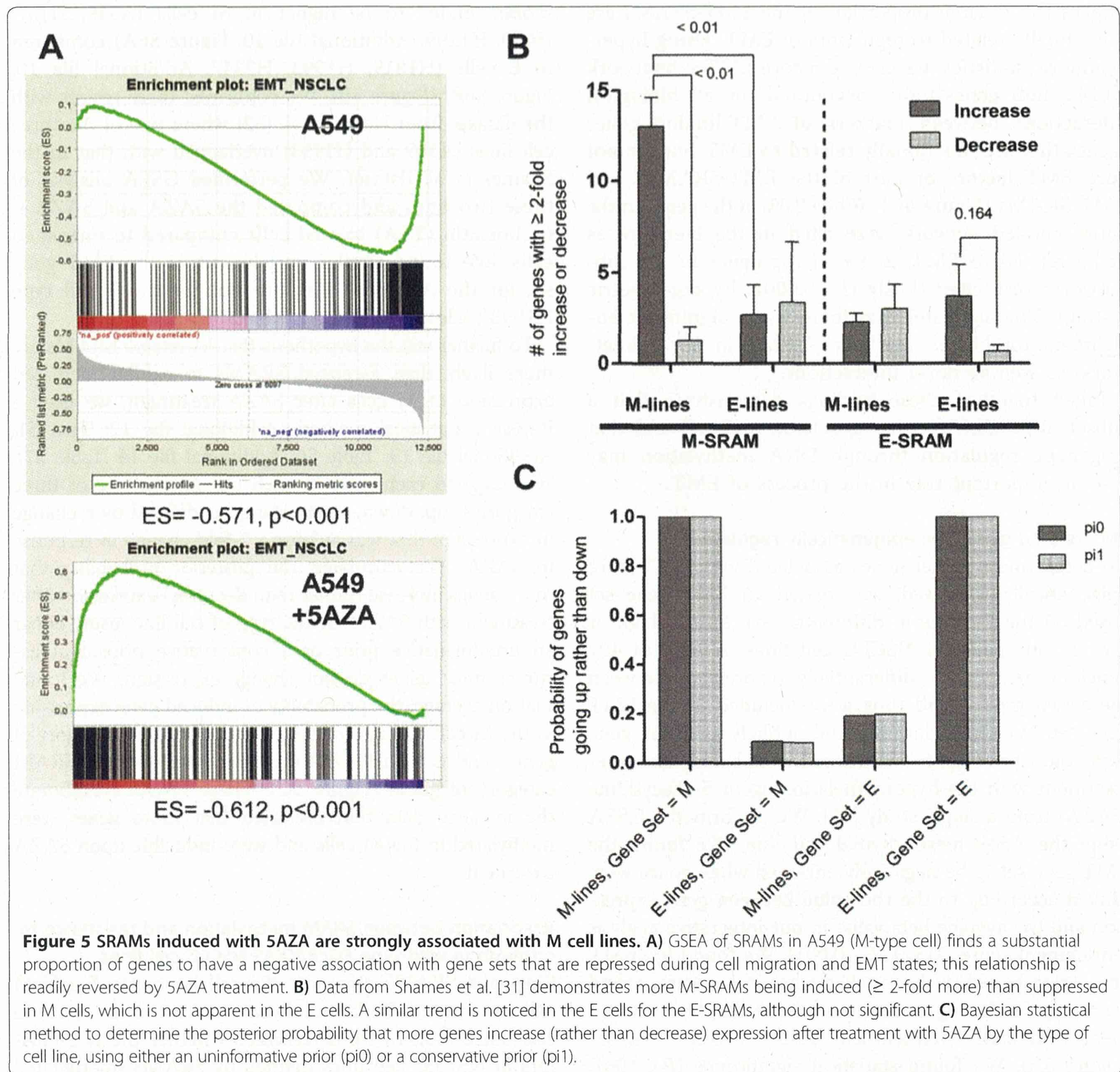


**Figure 5 SRAMs induced with 5AZA are strongly associated with M cell lines. A)** GSEA of SRAMs in A549 (M-type cell) finds a substantial proportion of genes to have a negative association with gene sets that are repressed during cell migration and EMT states; this relationship is readily reversed by 5AZA treatment. **B)** Data from Shames et al. [31] demonstrates more M-SRAMs being induced (≥ 2-fold more) than suppressed in M cells, which is not apparent in the E cells. A similar trend is noticed in the E cells for the E-SRAMs, although not significant. **C)** Bayesian statistical method to determine the posterior probability that more genes increase (rather than decrease) expression after treatment with 5AZA by the type of cell line, using either an uninformative prior (pi0) or a conservative prior (pi1).

found no available clinical datasets with DNA methylation profiling and erlotinib drug response information, we analyzed the gene expression profile for SRAMs in baseline tumor biopsies collected in patients with stage IV NSCLC that were included in the clinical trial Biomarker-Integrated Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE-1) [35,36]. Gene expression profiles were available for 27 patients treated with erlotinib (including 25 patients with wild-type *EGFR*), and for 47 patients treated with sorafenib (including 37 patients with wild-type *EGFR*), all of whom were evaluable for the primary endpoint of the trial, which was 8-week disease control. Interestingly, when SRAMs were used to cluster tumor samples collected at baseline in BATTLE-1, a similar pattern was observed *in vitro* (Figures 6C and D). In order to summarize the effect of the SRAMs, we computed the first principal component in each sample. Among the patients treated with erlotinib who had wild-type *EGFR*, the first principal component was lower in the patients who had a clinical status of disease control ($P = 0.05$) (Figure 6E). No difference was observed in patients treated with sorafenib ($P = 0.92$). When we further analyzed the 21 patients treated with erlotinib who had both *EGFR* and *KRAS* wild-type tumors, we