Prognostic significance of GAD1 overexpression in patients with resected lung adenocarcinoma

Mitsuhiro Tsuboi MD[1], Kazuya Kondo, MD, PhD[2], Kiyoshi Masuda, MD, PhD[3,4], Shoichiro Tange, PhD[3], Koichiro Kajiura, MD, PhD[1], Tomohiro Kohmoto[3], Hiromitsu Takizawa, MD, PhD[1], Issei Imoto, MD, PhD[3, 5, 6], and Akira Tangoku, MD, PhD[1]

[1]Department of Thoracic, Endocrine Surgery and Oncology, Institute of Biomedical Sciences, Tokushima University Graduate School, Tokushima, Japan

[2]Department of Oncological Medical Services, Institute of Biomedical Sciences, Tokushima University Graduate School, Tokushima, Japan

[3]Department of Human Genetics, Institute of Biomedical Sciences, Tokushima University Graduate School, Tokushima, Japan

[4]Kawasaki Medical School, Kurashiki, Japan

[5]Division of Molecular Genetics, Aichi Cancer Center Research Institute, Nagoya, Japan

[6]Department of Cancer Genetics, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan

Corresponding author

Kazuya Kondo, MD, PhD

Department of Oncological Medical Services, Institute of Biomedical Sciences, Tokushima University Graduate School, Tokushima, Japan

3-18-15 Kuramotocho, Tokushima City, Tokushima Pref. 770-8503, Japan

E-mail: kzykondo@tokushima-u.ac.jp

ABSTRACT

Background and Objectives: In a previous genome-wide screening, we identified hypermethylated CpG islands around *glutamate decarboxylase 1* (*GAD1*) in lung adenocarcinoma (LADC). In this study, we aimed to investigate the methylation and expression status of GAD1 and its prognostic value in patients with LADC.

Methods: *GAD1* methylation and mRNA expression status were analyzed using 33 tumorous and paired non-tumorous LADC samples and publicly available datasets. The prognostic value of GAD1 overexpression was investigated using publicly available datasets of mRNA levels and 162 cases of LADC by immunohistochemistry.

Results: The methylation and mRNA expression levels of GAD1, each having a positive correlation, were significantly higher in LADC tumors than in paired non-tumorous tissues. LADC patients with higher GAD1 mRNA expression showed significantly poorer prognosis for overall survival in publicly available datasets. Higher immunoreactivity of GAD1 was significantly associated with the pathological stage, pleural invasion, lymph vessel invasion, and poorer prognosis for cancer-specific and disease-free survival. Multivariate analysis revealed that GAD1 protein overexpression is an independent prognosticator for disease-free survival.

Conclusions: GAD1 mRNA and protein expression levels were significant prognostic factors in LADC, suggesting that they might be useful biomarkers to stratify patients with worse clinical outcome after resection.

Keywords: GAD1, lung adenocarcinoma, expression, prognosis, DNA methylation

1. INTRODUCTION

Lung adenocarcinoma (LADC) is the predominant histological subtype of lung cancer and has the highest mortality rate worldwide [1, 2]. Although progress in the treatment of LADC has improved short-term survival, the impacts on long-term survival remain modest [3]. Therefore, a better understanding of the mechanisms of LADC tumor progression is needed and useful prognostic molecular markers for accurately predicting the clinical outcomes of LADC are of great clinical significance.

To identify genes in the tumor that are specifically methylated at an early-stage of LADC, we had previously performed a genome-wide screening of aberrantly methylated CpG islands (CGIs) using paired tumorous and non-tumorous tissues of early-stage LADC, and identified *TRIM58* as a novel candidate tumor-suppressor gene for this disease [4]. Through this screening, the glutamate decarboxylase 1 gene (*GAD1*) was found to be nearby hypermethylated CGIs in LADC. Because paradoxical hypermethylation-associated overexpression of *GAD1* was reported recently in colorectal and liver cancers [5] and *GAD1* overexpression has been reported in various neoplastic tissues, such as oral, nasopharyngeal, colorectal, liver, and gastric cancers [5-9], we focused on *GAD1* as a potential LADC-related gene in the present study. Moreover, the methylation and expression status and clinicopathological significance of *GAD1* in LADC tumorigenesis have also not been examined previously.

Therefore, in the present study, we investigated the DNA methylation and mRNA and protein expression status of GAD1 in resected LADC tumors. Moreover, we assessed the prognostic significance of GAD1 expression in LADC using our tumor panel and publicly available datasets.

2. MATERIALS AND METHODS

2.1. Selection of candidate CGI

Previously obtained Human Methylation 450K array-based methylation screening data of 12 paired tumorous/non-tumorous stage-I LADC sample sets from patients (6 smokers and 6 never-smokers) who underwent surgery at Tokushima University Hospital (Tokushima, Japan) between April 1999 and March 2015 were reevaluated (Supplementary Table S1) [4].

2.2. Patients and tissue samples

We included tumors and non-tumorous tissues of LADC that were surgically resected at Tokushima University Hospital between April 1999 and November 2013 for additional analyses. No patients had been administered preoperative radiation, chemotherapy, or immunotherapy. For pyrosequencing-based methylation analysis and real-time PCR-based expression analysis, 33 LADC samples were used (Supplementary Table S2). For immunohistochemical staining, 162 LADC samples were used (Supplementary Table S3). The mean follow-up duration for the 162 patients with LADC was 48 months (range, 0.6–147 months), with 45 recurrences (27.8%) and 34 deaths (21.0%) among the patients. Tumor staging was determined based on the seventh tumor-node-metastasis (TNM) classification for lung cancer [10]. The tumors were classified according to the predominant histological subtype, as proposed by the 2015 WHO classification [11].

This study was performed in accordance with the principles outlined in the Declaration of Helsinki. The ethics committee of Tokushima University Hospital approved the study (approval number 3048), and formal written consent was obtained from all patients or their representatives.

2.3. DNA and RNA preparation and bisulfite conversion of genomic DNA

DNA and RNA were extracted using standard methods. Bisulfite conversion of DNA was conducted using the EpiTect Bisulfite Kit (QIAGEN GmbH, Hilden, Germany) following the manufacturer's instructions.

2.4. Bisulfite pyrosequencing

Bisulfite-treated genomic DNA was amplified using a set of primers designed with PyroMark Assay Design Software version 2.0.01.15 (QIAGEN GmbH, Supplementary Table S4). The target region for sequencing began 10 nucleotides (nt) before and ended 26 nt after cg15126544. PCR product pyrosequencing and methylation quantification were performed with sequencing primers using the PyroMark 24 Pyrosequencing System, version 2.0.6 (QIAGEN GmbH), according to the manufacturer's instructions.

2.5. Real-time quantitative reverse-transcription polymerase chain reaction (rqRT-PCR)

Complementary DNA was generated from isolated total RNA using the PrimeScript II 1st strand cDNA Synthesis Kit (TaKaRa, Shiga, Japan). rqRT-PCR was performed using KAPA PROBE FAST qPCR Kits (Kapa Biosystems, Wilmington, MA, USA) and TaqMan Gene Expression Assays (Thermo Fisher Scientific, Waltham, MA, USA; Supplementary Table S4) according to the manufacturers' instructions. *GAPDH* mRNA levels were used as internal controls for normalization. Relative expression of *GAD1* mRNA was calculated using Human Lung Total RNA (TaKaRa) as a normal lung control.

2.6. Data mining in bioinformatics

Available RNA sequencing data (IlluminaHiSeq_RNASeqV2 Level 3) containing 488 tumor and 58 non-tumor samples and Infinium Human Methylation 450K data (Level 3) containing 473 tumor and 32 non-tumorous samples of LADC cases with clinical annotations were downloaded from The Cancer Genome Atlas (TCGA) Research Network (http://cancergenome.nih.gov). mRNA expression data and DNA methylation data were available for 36 and 29 paired tumorous/non-tumorous sample sets, respectively; both types of data were available for 18 sets. Tumorous samples with mRNA expression data and survival

data were available for 423 cases. Survival analyses were conducted on patients with normalized mRNA expression and overall survival (OS) profiles. Patients were divided into low- and high-expression groups according to the median *GAD1* mRNA expression value.

Kaplan-Meier Plotter (KM plotter, http://kmplot.com/analysis/), a publicly available online database of published microarray datasets for primary tumors with clinical information [12], was also used to generate OS curves in 9 studies from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/, Supplementary Table S5) by setting the auto-selected best value of *GAD1* mRNA expression as the cutoff. All other parameters were left at default settings.

2.7. Immunohistochemical staining

Paraffin sections (4 μm thick) were subjected to immunohistochemical staining using the Envision system (ChemMate Envision kit; Dako, Glostrup, Denmark) according to the manufacturer's instructions. Antigen retrieval was performed by heating the dewaxed and dehydrated sections in Dako Real Target Retrieval Solution, pH 9 (Dako), using a 2100 retriever (Aptum Biologics, Ltd., Southampton, UK). A mouse anti-GAD67 monoclonal antibody (Sigma-Aldrich, St. Louis, MO, USA; G5419), diluted to 1:200 with antibody diluents (Dako), was used as the primary antibody. The proportion and intensity of GAD1 staining in the LADC samples were scored (Supplementary Table S6A) independently by two different researchers.

2.8. Statistical analysis

Student's *t*-test or Fischer's exact test was used for comparisons between two groups. The paired *t*-test was used for comparisons between paired samples. The relationship between continuous variables was investigated by calculating the Spearman's correlation coefficient. For survival analysis, Kaplan–Meier survival curves were constructed for groups based on univariate predictors, and differences among groups were tested with the log-rank test.

Univariate and multivariate survival analyses were performed using the likelihood ratio test of the stratified Cox proportional hazard regression analysis. Differences were assessed using two-sided tests and were considered significant at a *P*-value of < 0.05. Statistical analyses were performed using IBM SPSS version 24 (IBM Corporation, Armonk, NY, USA) or the Survival package for R (https://cran.r-project.org).

3. RESULTS

3.1. Methylation status of CGIs and each CpG site within CGIs around GAD1

In a previous array-based, genome-wide methylation screening of 12 paired tumorous/non-tumorous LADC sample sets [4], CGI-3 around *GAD1* was ranked 14th as a hypermethylated CGI with a high *P*-value (Supplementary Table S1). Because hypermethylation-associated overexpression of *GAD1* was reported in colorectal and liver cancers [5], we re-evaluated the results of the array-based methylation status of each CpG site within CGI-1–4 (Fig. 1A) around *GAD1* (Fig. 1B). The methylation levels of all CpG sites determined by array-based analysis within CGI-3 and in tumors were significantly higher than those in paired non-tumorous tissues. Although the methylation levels in tumors were higher in CpG sites within CGI-3 than in those within CGI-4, the average β-value in non-tumor tissues was extremely and specifically low at cg15126544 and showed the largest difference of average β-value between tumors and non-tumor tissues at this site (Fig. 1B and Supplementary Table S7), which is localized within the CCCTC-binding factor (CTCF)-binding site of *GAD1*. Similar results were observed in the Level 3 Infinium Human Methylation 450K data of 29 LADC tumors and paired non-tumor tissues from TCGA dataset (Supplementary Fig. S1). Because hypermethylation around this CTCF-binding site has been reported as a possible cause of *GAD1* overexpression [5], we further assessed the methylation status of cg15126544 and *GAD1* mRNA expression levels.

3.2. Correlation between GAD1 expression and CGI methylation in LADC clinical cases

The DNA methylation status and mRNA expression status were investigated in our panel of LADC tumorous and paired non-tumorous tissues (Supplementary Table S2) using pyrosequence-based methylation assays and rqRT-PCR-based expression analysis, respectively. Of the 33 sample sets, 26 (78.8%) demonstrated significantly higher methylation levels in tumor samples than in non-tumorous tissues (Fig. 1C). In the same cases, the mean *GAD1* mRNA expression levels in the tumors were significantly higher than those in the paired non-tumorous tissues (Fig. 1D). There was a slightly positive ($\rho = 0.251$) but significant correlation between methylation levels at cg15126544 and *GAD1* mRNA expression (Fig. 1E). The LADC sample set containing 18-paired samples obtained from TCGA demonstrated similar results both in methylation levels at cg15126544 and *GAD1* mRNA expression (Fig. 1F, 1G and Supplementary Fig. S1). A significant and highly positive correlation between them was also observed in TCGA dataset ($\rho = 0.706$, Fig. 1H). Because the gene expression status of cancer cells directly affects their phenotypes, including malignant features, we focused on GAD1 expression in tumors to further assess its prognostic significance in patients with LADC.

3.3. Association of GAD1 mRNA expression levels with prognosis in LADC tumors

In our LADC cohort, a sufficient number of cases with high-quality RNA suitable for expression analysis was not available for survival analysis. Therefore, to test the association between *GAD1* mRNA expression levels in tumors and patients' prognosis, we first performed survival analysis of 423 patients with LADC using data obtained from TCGA dataset. The OS rate of patients with LADC with higher *GAD1* mRNA expression in tumors was significantly poorer than that of patients with lower *GAD1* mRNA expression in tumors (Fig. 2A). Univariate Cox regression analysis using data obtained from TCGA dataset confirmed that high *GAD1* mRNA expression was associated with a worse prognostic significance for OS (Table 1). In multivariate Cox regression analysis, high *GAD1* mRNA expression was also significantly

associated with a poorer OS rate, suggesting that *GAD1* mRNA expression is an independent prognostic factor for OS ($P = 0.036$, Table 1).

To validate this result, we performed survival analysis by drawing Kaplan-Meier survival curves using KM plotter (Fig. 2B). A total of 9 studies from the GEO dataset were included (Supplementary Table S5). In a total of 720 patients with LADC from 9 cohorts, high *GAD1* mRNA expression also significantly correlated with worse OS. In subgroup analysis of OS using datasets of KM plotter, heterogeneous results were obtained among different cohorts. Larger cohorts such as GSE31210 and GSE50081 consistently showed that higher *GAD1* mRNA expression was a poor prognostic factor, whereas cohorts with a smaller number of cases showed varying results (Supplementary Fig. S2). The results of univariate Cox regression analysis confirmed these results (Fig. 2C).

3.4. Immunohistochemical staining pattern of GAD1 and its association with prognosis in LADC tumors

To further validate the prognostic significance of GAD1 expression status, we further examined the correlation between GAD1 protein expression and clinicopathological features including prognosis in patients with LADC. We performed immunohistochemical staining of GAD1 in tissue samples from our cohort of 162 patients with LADC (Supplementary Table S3). Cytoplasmic GAD1 staining was observed in LADC tumor cells with higher mRNA expression, whereas nearly no staining was observed in normal lung epithelial cells and either tumorous or non-tumorous epithelial cells in LADC with lower mRNA expression (Fig. 3A). According to the staining score (Supplementary Table S6B), 112 patients (69.1%) were classified into the group with tumors showing GAD1 protein overexpression (positive GAD1 immunoreactivity). Among the various clinicopathological factors, the pathological stage, pleural invasion, and lymph vessel invasion were identified as factors significantly and positively associated with

positive GAD1 immunoreactivity (Table 2). Lymph node metastasis also tended to be more frequently observed in the positive GAD1 immunoreactivity group.

According to the GAD1 protein expression status of LADC tumors, Kaplan-Meier curves of estimated OS, disease-free survival (DFS), and cancer-specific survival (CSS) were generated. Patients with GAD1 protein-overexpressing tumors showed significantly poorer DFS ($P < 0.001$, log-rank test) and CSS ($P = 0.031$, log-rank test) than those without GAD1 protein overexpressing tumors. Patients with GAD1 protein-overexpressing tumors tended to show poorer OS, although the difference between groups was not significant (Fig. 3B). Univariate Cox regression analysis confirmed that positive GAD1 immunoreactivity was significantly associated with a worse prognostic significance for DFS (Table 3). Multivariate Cox regression analysis in 162 patients revealed that GAD1 immunoreactivity was an independent prognostic factor for DFS ($P = 0.011$, hazard ratio = 6.424, Table 3), but not for OS and CSS (Supplementary Table S8 and S9).

4. DISCUSSION

In the present study, we focused on *GAD1* as a hypermethylated gene at specific CpG sites in LADC tumors and demonstrated its overexpression in tumor-specific and methylation level-associated manners in LADC. We also demonstrated the prognostic significance of GAD1 mRNA and protein expression levels in resected LADC tumors using various independent publicly available datasets and our cohort, respectively. Our study suggested that GAD1 overexpression may be a useful biomarker for predicting the prognosis of patients with LADC.

GAD1 is known to catalyze the production of γ-aminobutyric acid (GABA) from L-glutamic acid, the principal inhibitory neurotransmitter in the brain [13, 14]. GAD1 overexpression has been reported in various neoplastic tissues, but not in LADC. Moreover, the associations between clinicopathological characteristics and GAD1 expression have not been well-established. The most striking finding in this study is the prognostic significance of GAD1

mRNA and protein expression in patients with LADC. Although a sufficient number of RNA samples suitable for expression analysis was not available in our cohort for survival analyses, we used various publicly available data and demonstrated that *GAD1* mRNA overexpression in tumors was significantly associated with poor prognosis (OS) in independent TCGA and GEO datasets of LADC cases. In immunohistochemical analysis using our cohort, a positive cytoplasmic GAD1 staining pattern in tumor cells was significantly associated with poor prognosis, particularly DFS but not OS, in patients with LADC. Although the difference in the association between GAD1 expression and OS among datasets remains unclear, it may be explained by (1) variations in GAD1 mRNA and protein expression, (2) the smaller size of the cohort for immunohistochemical analysis compared to those of cohorts used for mRNA analysis used in our study, and (3) variations in GAD1 expression level and/or pattern among different ethnicities.

Our study also demonstrated that GAD1 protein expression in LADC was significantly associated with pleural invasion and lymph vessel invasion. These findings suggest that GAD1 overexpression might be closely associated with cellular invasion. This hypothesis is supported by previous reports of another cancers. Kimura et al. [6] demonstrated that *GAD1* promotes the cancer cell invasion and metastasis of oral cancer by inducing the nuclear translocation of β-catenin and secretion of MMP7 [15-20], although the regulatory mechanisms of GAD1 in β-catenin translocation remain unclear. In a brain metastasis model, it was reported that the metastatic activity of tumor cells depends on the GAD1-GABA synthesis pathway [21]. Further studies are needed to clarify the tumor-promoting activity of overexpressed GAD1.

Recently, Yan *et al*. [5] reported hypermethylation-associated *GAD1* overexpression in colorectal and liver cancers and found that this paradoxical effect was caused by the hypermethylation of the CTCF-binding site within *GAD1*, which may prevent CTCF binding, inhibit CTCF-mediated repressive Polycomb repressive complex 2 (PRC2) complex recruitment to the *GAD1* promoter, inhibit PRC2-induced trimethylation of histone H3 lysine 27

(H3K27m3), and eliminate the blocking activity H3K27m3 for *GAD1* transcription [22, 23]. These observations are contradictory to the well-established paradigm that promoter DNA methylation represses transcription by inhibiting transcription factor binding and/or chromatin structure modification [24-26]. In this study, we also detected hypermethylation at cg15126544 within the CTCF-binding site in LADC tumors, and tumor-specific *GAD1* overexpression was positively associated with hypermethylation at cg15126544 in our cohort and the TCGA dataset. Therefore, methylation of CTCF-binding sites may regulate *GAD1* expression in LADC as well. However, it remains unknown whether the methylation of CGI or each CpG site around *GAD1*, particularly cg15126544, is the only mechanism underlying the regulation of its transcription. Interestingly, in brain metastatic tumor cells, it was reported that the downregulation of the DNA methyltransferase DNMT1 induced by the brain microenvironment-derived clusterin resulted in decreased GAD1 promoter methylation and subsequent upregulation of GAD1 expression [21]. Therefore, even the effect of methylation levels of CpG sites around *GAD1* on its expression level may vary under different conditions or in different cell lineages. Indeed, MethSurv, a web tool for multivariable survival analysis using DNA methylation data obtained from TCGA datasets (https://biit.cs.ut.ee/methsurv/), failed to show the prognostic significance of CpG sites around *GAD1*, including cg15126544 for OS (data not shown). Therefore, the methylation status of some CpG sites around *GAD1* may contribute to its gene expression at some stages of LADC development, but not to the progression of this tumor. The *GAD1* mRNA expression level data in normal lung tissues available in public databases, such as the NIH Genotype-Tissue Expression Project (https://www.gtexportal.org/), as well as our immunohistochemical staining results revealed no or low GAD1 expression in normal lung tissue, suggesting that GAD1 is specifically expressed in tumor cells and contributes to the progression of tumors in LADC. Because the gene expression status appears to more directly contribute to the establishment of clinicopathological phenotypes in tumor cells, it is necessary

to investigate the detailed regulatory mechanisms of GAD1 expression in LADC cells at each developmental stage of the tumor.

There are some limitations to this study. First, we demonstrated the prognostic impact of GAD1 mRNA and protein statuses mainly in Caucasian and Japanese (Asian) populations, respectively, but no data are available to directly compare GAD1 mRNA and protein expression levels among different ethnicities. Because it has been reported that the frequency of acquired alterations, such as epidermal growth factor receptor mutation, in lung tumors can vary across different ethnicities [27-29], it is possible that the GAD1 expression pattern and/or levels differ between Caucasian and Asian populations. However, the prognostic significance of the *GAD1* mRNA expression status in Japanese cases with LADC was demonstrated by GSE31210 in GEO datasets (Fig. 2C and Supplementary Fig. S2). Meta-analysis using 9 GEO datasets, including GSE31210 and 8 other studies from western countries (Supplementary Table S5) also revealed the prognostic significance of the GAD1 mRNA expression status (Fig. 2C), suggesting that GAD1 overexpression is a common prognostic factor in various populations. Second, our patient cohort was relatively small even for immunohistochemical analysis, and a sufficient number of samples was not available for mRNA expression analysis to perform survival analysis. Prospective multi-institutional studies are needed to further validate the prognostic value of GAD1 overexpression in patients with LADC.

5. CONCLUSION

GAD1 overexpression appears to be a significant and independent prognostic indicator in patients with resected LADC at both the mRNA and protein levels. This information may be helpful for identifying patients at high risk of recurrence and overall survival after tumor resection of LADC.

REFERENCES

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *Cancer J Clin*. 2015;65(2):87–108.

2. Travis WD, Brambilla E, Noguchi M, 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(2):244–285.

3. Johnson DH, Schiller JH. Recent clinical advances in lung cancer management. *J Clin Oncol*. 2014;32(10):973–982.

4. Kajiura K, Masuda K, Naruto T, et al. Frequent silencing of the candidate tumor suppressor TRIM58 by promoter methylation in early-stage lung adenocarcinoma. *Oncotarget*. 2017;10(2):2890–2905.

5. Yan H, Tang G, Wang H, et al. DNA methylation reactivates GAD1 expression in cancer by preventing CTCF-mediated polycomb repressive complex 2 recruitment. *Oncogene*. 2016;35(30):3995–4008.

6. Kimura R, Kasamatsu A, Koyama T, et al. Glutamate acid decarboxylase 1 promotes metastasis of human oral cancer by beta-catenin translocation and MMP7 activation. *BMC Cancer*. 2013;13:555.

7. Lee YY, Chao TB, Sheu MJ, et al. Glutamate decarboxylase 1 overexpression as a poor prognostic factor in patients with nasopharyngeal carcinoma. *J Cancer*. 2016;7(12):1716–1723.

8. Maemura K, Yamauchi H, Hayasaki H, et al. Gamma-amino-butyric acid immunoreactivity in intramucosal colonic tumors. *J Gastroenterol Hepatol*. 2003;18(9):1089–1094.

9. Matuszek M, Jesipowicz M, Kleinrok Z. GABA content and GAD activity in gastric cancer. *Med Sci Monit*. 2001;7(3):377-381.

10. Goldstraw P, Crowley J, Chansky K, et al. International Association for the Study of Lung Cancer International Staging Committee, and Participating Institutions. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. *J Thorac Oncol* 2007;2(8):706–714.

11. Travis WD, Brambilla E, Burke AP, Marx A, Nicholson, AG. WHO classification of tumours of the lung, pleura, thymus and heart. Fourth edition. Lyon: International Agency for Research on Cancer; 2015.

12. Győrffy B, Surowiak P, Budczies J, Lánczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One*. 2013;18:e82241.

13. Erlander MG, Tobin AJ. The structural and functional heterogeneity of glutamic acid decarboxylase: a review. *Neurochem Res*. 1991;16(3):215–226.

14. Young SZ, Bordey A. GABA's control of stem and cancer cell proliferation in adult neural and peripheral niches. *Physiology (Bethesda)*. 2009;24:171–185.

15. Kruck S, Eyrich C, Scharpf M, et al. Impact of an altered Wnt1/β-catenin expression on clinicopathology and prognosis in clear cell renal cell carcinoma. *Int J Mol Sci*. 2013;14(6):10944–10957.

16. Sawada G, Ueo H, Matsumura T, et al. CHD8 is an independent prognostic indicator that regulates Wnt/β-catenin signaling and the cell cycle in gastric cancer. *Oncol Rep*. 2013;30(3):1137–1142.

17. Abd El-Rehim D, Ali MM. Aberrant expression of β-catenin in invasive ductal breast carcinomas. *J Egypt Natl Canc Inst*. 2009;21(2):185–195.

18. Ishikawa T, Kimura Y, Hirano H, Higashi S. Matrix metalloproteinase-7 induces homotypic tumor cell aggregation via proteolytic cleavage of the membrane-bound Kunitz-type inhibitor HAI-1. *J Biol Chem*. 2017;292(50):20769–20784.

19. Mylona E, Kapranou A, Mavrommatis J, et al. The multifunctional role of the immunohistochemical expression of MMP-7 in invasive breast cancer. *APMIS*. 2005;113(4):246–255.

20. Wang FQ, So J, Reierstad S, Fishman DA. Matrilysin (MMP-7) promotes invasion of ovarian cancer cells by activation of progelatinase. *Int J Cancer*. 2005;114(1):19–31.

21. Schnepp PM, Lee DD, Guldner IH, et al. GAD1 upregulation programs aggressive features of cancer cell metabolism in the brain metastatic microenvironment. *Cancer Res*. 2017;77(11):2844-2856.

22. Schuettengruber B, Chourrout D, Vervoort M, et al. Genome regulation by polycomb and trithorax proteins. *Cell*. 2007;128(4):735–745.

23. Li T, Hu JF, Qiu X, et al. CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol Cell Biol*. 2008;28(20):6473–6482.

24. Razin A, Riggs AD. DNA methylation and gene function. *Science*. 1980;210(4470):604–610.

25. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. *Science*. 1975;187(4173):226–232.

26. Bird A. Perceptions of epigenetics. *Nature*. 2007;447(7143):396–398.

27. Arrieta O, Cardona AF, Martín C, et al. Updated frequency of EGFR and KRAS mutations in non-small-cell lung cancer in Latin America: The Latin-American consortium for the investigation of lung cancer (CLICaP). *J Thorac Oncol*. 2015;10(5):838–843.

28. Shigematsu H, Lin L, Takahashi T, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst*. 2005;97(5):339–346.

29. Campbell JD, Lathan C, Sholl L, et al. Comparison of Prevalence and Types of Mutations in Lung Cancers Among Black and White Populations. *JAMA Oncol*. 2017;3(6):801–809.

436 Table 1. Cox proportional hazard regression analysis of overall survival in 400 patients with LADC in TCGA dataset
437

Factor (number)	Univariate			Multivariate		
	Hazard ratio	95% confidence interval	*P*-value	Hazard ratio	95% confidence interval	*P*-value
Sex Male (n = 184) vs. Female (n = 216)	1.048	0.704 – 1.560	0.818	1.087	0.705 – 1.675	0.706
Age (years) >67 (n = 210) vs. ≤67 (n = 190)	1.348	0.897 – 2.025	0.151	1.639	1.079 – 2.490	**0.021**
Smoking history Positive (n = 339) vs. Negative (n = 61)	1.069	0.569 – 2.006	0.836	1.521	0.766 – 3.020	0.230
Pathological stage II, III, IV (n = 184) vs. I (n=216)	2.620	1.725 – 3.979	**6.21E-6**	-	-	-
Tumor size pT2-4 (n = 272) vs. pT1 (n = 128)	1.631	0.978 – 2.720	0.0609	1.565	0.922 – 2.658	0.097
N stage (pN) pN1-3 (n = 136) vs. pN0 (n = 264)	2.475	1.662 – 3.688	**8.32E-6**	2.487	1.649 – 3.750	**1.38E-5**
M stage (pM) pM1 (n 19) vs. pM0 (n = 381)	1.539	0.773 - 3.066	0.220	1.528	0.752 – 3.103	0.241
GAD1 mRNA expression High (n = 217) vs. Low (n = 183)	1.749	1.165 – 2.626	**6.97E-3**	1.573	1.029 – 2.404	**0.036**

438 Statistically significant values are in boldface type.
439 The analysis was performed in 400 patients with complete clinical information in the TCGA dataset.
440 The population was divided using the auto-selected best value of *GAD1* mRNA expression as the cutoff.
441

Table 2. Correlation between GAD1 immunoreactivity and clinicopathological factors in 162 patients with LADC

Factor	GAD1 immunoreactivity (n = 162)		*P*-value[a]
	Negative (n = 50)	Positive (n = 112)	
Male / Female	26 / 24	55 / 57	0.865
Age[b]	69.0 ± 9.6	67.4 ± 9.0	0.386
Smoking history[b, c] (+/-)	22 / 27	55 / 56	0.603
Brinkman index[b, c]	406.5 ± 536.4	485.0 ± 622.7	0.461
Tumor size[b, c]	23.5 ± 14.3	26.1 ± 13.4	0.226
pStage (I/II+III)	39 / 11	65 / 47	**0.021**
Lymph node metastasis (+/-)	8 / 42	36 / 76	0.054
Pleural invasion[c] (+/-)	5 / 40	35 / 72	**0.005**
Vascular invasion[c] (+/-)	5 / 40	22 / 79	0.284
Lymph vessel invasion[c] (+/-)	6 / 39	31 / 66	**0.023**
EGFR mutation[c] (+/-)	10/ 6	30/ 29	0.573
Predominant histologic subtype (lepidic / papillary / acinar/ solid/ enteric)	23 / 18 / 4 / 4 / 1	36 / 47 / 24 / 5 / 0	0.068

[a]*P*-values were calculated using Fischer's exact test for gender, smoking history, lymph node metastasis, pleural invasion, lymph vessel invasion, and vascular invasion, *EGFR* mutation, and using Student's *t*-test for age, Brinkman index, and tumor size and using $\chi 2$ test for trend for predominant histologic subtype. Statistically significant values ($P < 0.05$) are in boldface type.

[b]Age, Brinkman index, and tumor size are expressed as the mean ± standard deviation.

[c]Data of these factors were not available for all patients.

Table 3. Cox proportional hazard regression analysis for disease-free survival in 162 patients with LADC

Factor	Univariate			Multivariate		
	Hazard ratio	95% confidence interval	*P*-value	Hazard ratio	95% confidence interval	*P*-value
Sex Male (n=81) vs. Female (n=81)	1.202	0.666 - 2.170	0.541	2.459	0.520 - 11.624	0.256
Age (years) >67 (n=87) vs. ≤67 (n=75)	1.048	0.582 - 1.887	0.875	0.995	0.515 - 1.922	0.988
Smoking history[a] Positive (n=77) vs. Negative (n=83)	1.302	0.724 - 2.344	0.378	0.324	0.068 - 1.546	0.158
Pathological stage II, III (n=58) vs. I (n=104)	7.466	3.769 - 14.789	< **0.001**	-	-	-
Tumor size[a] pT2-4 (n=39) vs. pT1 (n=115)	2.309	1.241 - 4.296	**0.008**	2.033	0.961 - 4.303	0.070
N stage (pN) pN1-3 (n=44) vs. pN0 (n=118)	7.100	3.837 - 13.140	< **0.001**	2.507	1.057 - 5.949	**0.037**
Pleural invasion[a] Positive (n=40) vs. Negative (n=112)	4.926	2.637 - 9.202	< **0.001**	2.091	0.977 - 4.478	0.058
Vascular invasion[a] Positive (n=27) vs. Negative (n=119)	4.706	2.529 – 8.757	< **0.001**	1.139	0.389 - 3.341	0.812
Lymph vessel invasion[a] Positive (n=37) vs. Negative (n=105)	5.346	2.809 - 10.175	< **0.001**	1.355	0.478 - 3.847	0.568
Adjuvant chemotherapy[a] With (n=47) vs. Without (n=106)	2.972	1.614 - 5.470	< **0.001**	-	-	-
EGFR mutation[a] Negative (n=35) vs. Positive (n=40)	1.285	0.678 - 2.433	0.442	-	-	-
Predominant subtype Non-lepidic (n=103) vs. Lepidic (n=59)	6.711	2.392- 18.868	< **0.001**	2.725	0.861 - 8.621	0.088
GAD1 immunoreactivity Positive (n=112) vs. Negative (n=50)	9.341	2.248 - 38.824	**0.002**	6.424	1.522 - 27.108	**0.011**

Statistically significant values are in boldface type.

[a]Data of these factors were not available for all patients.

Supplementary Table S1. The top 14 CpG islands significantly hypermethylated in tumorous tissues of 12 stage-I LADC cases[4]

No.	CpG island	Adjusted *P*-value[a]	β-difference[b]	Gene name
1	chr7:153583317-153585666	0.000495704	0.277562652	*DPP6*
2	chr19:52390841-52391368	0.000671077	0.291495741	*ZNF577*
3	chr11:125774292-125774584	0.000839411	0.271871389	*DDX25*
4	chr3:62355315-62355534	0.001348289	0.25890625	*FEZF2*
5	chr1:156863415-156863711	0.001564128	0.369176667	*PEAR1*
6	chr15:37390175-37390380	0.002225792	0.324917222	*MEIS2*
7	chr1:248020330-248021252	0.00443318	0.270335	*TRIM58*
8	chr12:103696090-103696418	0.006552561	0.318931667	*C12orf42*
9	chr7:158110569-158110881	0.008975336	0.270233333	*PTPRN2*
10	chr6:50810642-50810994	0.010799023	0.30752125	*TFAP2B*
11	chr5:134363092-134365146	0.011483039	0.262798796	*PITX1*
12	chr19:58545115-58545897	0.011599292	0.27722213	*ZSCAN1*
13	chr6:50791110-50791573	0.012733507	0.331794167	*TFAP2B*
14	**chr2:171676552-171676980**	**0.017464759**	**0.251871944**	***GAD1***

The row corresponding to *GAD1* is in boldface type.

[a]Differences between methylation levels (β-values) of CpG islands in tumors and paired non-tumorous tissues were assessed by paired *t*-test. *P*-values were adjusted with the Benjamini–Hochberg correction (false discovery rate, FDR). CpG islands were sorted by the adjusted *P*-value.

[b]β-differences (differential methylation levels) represent the average of [(β-value of tumorous tissue) - (β-value of paired non-tumorous tissue)] in 12 stage-I LADC cases.

Supplementary Table S2. Clinicopathological characteristics of 33 patients with LADC analyzed by qPCR and pyrosequencing

Characteristics	Number
Gender	
Male	18
Female	15
Age (years)	62.9 ± 9.6
Stage	
Ia, Ib	16
IIa, IIb	8
IIIa, IIIb	9
Smoking History	
+	15
-	18
Brinkman Index	616.7 ± 745.4

Age and Brinkman index are expressed as the mean ± standard deviation.

469 Supplementary Table S3. Characteristics of 162 patients with LADC analyzed by immunohistochemistry
470

Characteristics	N = 162 (%)
Gender	
Male	81 (50.0%)
Female	81 (50.0%)
Age (years)	67.0 ± 9.2
Stage	
Ia, Ib	104 (64.2%)
IIa, IIb	26 (16.0%)
IIIa, IIIb	32 (19.8%)
EGFR mutation	
Positive	40 (24.7%)
Negative	35 (21.6%)
Unknown	87 (53.7%)
Predominant histologic subtype	
lepidic	59 (36.4%)
papillary	65 (40.1%)
acinar	28 (17.3%)
solid	9 (5.6%)
enteric	1 (0.6%)
Adjuvant chemotherapy	
With	47 (29.0%)
Without	106 (65.4%)
Unknown	9 (5.6%)
Smoking History	
+	78 (48.1%)
-	82 (50.6%)
Unknown	2 (1.2%)
Brinkman Index	461.0 ± 597.0

471 Age and Brinkman index are expressed as the mean ± standard deviation.

Supplementary Table S4. List of primer sets used in qPCR and pyrosequencing

Gene/primer name		Sequence/ID
TaqMan gene expression assay		
GAD1	FAM	Hs01065893_m1
GAPDH	FAM	Hs02758991_g1
Pyrosequencing of *GAD1*		
cg15126544	Forward	5'-TGGTTTTTAGGGGTTTTTTTTTTGGA-3'
	Reverse	5'-ACAAATACACCCCCTTTAATCTACTCTCC-3'
	Sequence	5'-GTAGAAGAGGGAGGAA-3'

Supplementary Table S5. List of GEO data sets

GEO accession	Survival period	Submission date	Number of patients	Country	Race	Platform
GSE14814	from date of random assignment to death from disease or treatment complication	12-Feb-9	27	Canada USA Germany	NA	HG-U133A,
GSE19188	NA	25-Nov-9	41	Netherlands	Mostly Caucasian	HG-U133_Plus_2
GSE3141	NA	16-Aug-5	58	USA	NA	HG-U133_Plus_2
GSE50081	NA	21-Aug-13	127	Canada	NA	HG-U133_Plus_2
GSE31908	NA	6-Sep-11	20	USA	Mostly Caucasian	HG-U133A HG-U133B HG-U133_Plus_2
GSE37745	NA	3-May-12	106	Sweden	NA	HG-U133_Plus_2
GSE29013	from the date of surgery to death or the last follow-up contact.	2-May-11	30	USA	Mostly Caucasian	HG-U133_Plus_2
GSE30219	NA	26-Jun-11	85	France USA	NA	HG-U133_Plus_2
GSE31210	NA	4-Aug-11	226	Japan	Asian	HG-U133_Plus_2

Supplementary Table S6. Evaluation criteria for GAD1 immunohistochemistry

A. Proportion and intensity scores for GAD1 staining in immunohistochemical analysis

Proportion score (PS)		Intensity score (IS)	
Score	Observation	Score	Observation
1	< 25%	0	None
2	26 - 50%	1	Weak
3	51 - 75%	2	Intermediate
4	76% ≤	3	Strong

B. Evaluation of GAD1 immunoreactivity using PS and IS

GAD1 immunoreactivity	Sum of PS and IS	Number of cases
Negative	1	0
	2	14
	3	10
	4	26
Positive	5	44
	6	43
	7	25

The staining score is defined as the sum of the proportion and intensity scores.

A staining score ≥ 5 indicated overexpression of the GAD1 protein (positive GAD1 immunoreactivity).

Supplementary Table S7. The methylation levels of each CpG site of *GAD1* in tumorous and non-tumorous samples

CpG site	β-value (average[a])		β-value (SD[b])		*P*-value[c]	β-difference[d]
	Tumor	Non-tumor	Tumor	Non-tumor		
cg09404592	0.109475	0.0853917	0.04619235	0.02516931	0.1810443145	0.024083333
cg03443455	0.462661667	0.3074692	0.09759064	0.04014665	0.0004807189	0.1551925
cg00782607	0.106084167	0.0710583	0.04960498	0.03644913	0.0389929681	0.035025833
cg13612847	0.133873333	0.1472975	0.03817365	0.03253079	0.0972520412	-0.013424167
cg03448612	0.083106667	0.0758158	0.02049695	0.03057835	0.3611596712	0.007290833
cg09742688	0.019295	0.014535	0.00964859	0.00580472	0.1767899083	0.00476
cg23221504	0.100365	0.1095158	0.02678416	0.03834864	0.3202119878	-0.009150833
cg00915206	0.067003333	0.0671983	0.01743706	0.03309326	0.9850457390	-0.000195
cg11582100	0.05327	0.0542092	0.01019424	0.02677933	0.8871008599	-0.000939167
cg15306595	0.0394325	0.0438158	0.01045799	0.01931569	0.4524979803	-0.004383333
cg19538089	0.104192727	0.0801767	0.05266865	0.02819045	0.1330389745	0.024016061
cg26391350	0.086990833	0.0757083	0.02871752	0.03569546	0.2703867579	0.0112825
cg16911423	0.179124167	0.13477	0.05271661	0.02415412	0.0169907853	0.044354167
cg01763173	0.085408333	0.0639708	0.03334425	0.02293707	0.0705795610	0.0214375
cg11281641	0.154460833	0.0533167	0.0959479	0.02516992	0.0046353604	0.101144167
cg07420274	0.536216667	0.3580283	0.06408349	0.05326207	0.0001252001	0.178188333
cg01089249	0.529403333	0.2895042	0.06934644	0.02343744	0.0000006700	0.239899167
cg01089319	0.505844167	0.256055	0.06506377	0.03497551	0.0000011299	0.249789167
cg14005211	0.539773333	0.2738458	0.0827698	0.05432622	0.0000013853	0.2659275
cg14486905	0.46974	0.2449983	0.12168293	0.04163571	0.0002223438	0.224741667
cg09144707	0.494621667	0.2901625	0.10770392	0.03041366	0.0000452368	0.204459167
cg02723395	0.411985	0.1740225	0.15358608	0.0351129	0.0005880955	0.2379625
cg15126544	**0.363693333**	**0.0397042**	**0.14160994**	**0.02161663**	**0.0000079194**	**0.323989167**
cg04105250	0.337811667	0.1510408	0.11281548	0.02977583	0.0001733147	0.186770833
cg00729049	0.2934125	0.1690383	0.08912118	0.02842007	0.0014668716	0.124374167
cg15753746	0.363454167	0.1337517	0.17759626	0.03602908	0.0009859990	0.2297025
cg21535772	0.4300025	0.2728233	0.09775366	0.03462184	0.0004699895	0.157179167
cg19846314	0.445076667	0.2232308	0.17797695	0.07571523	0.0024014465	0.221845833
cg08863440	0.403660833	0.2232942	0.15063177	0.06307915	0.0033596932	0.180366667
cg07620853	0.5666125	0.5479283	0.19697479	0.14942316	0.7081756486	0.018684167

The row corresponding to cg15126544 is in **boldface type**.

[a]The average methylation level of 12 LADC samples.

[b]The standard deviation (SD) of methylation levels of 12 LADC samples.

[c]Differences between methylation levels (β-values) of CpG islands in tumors and paired non-tumorous tissues were assessed by paired *t*-test.

[d]β-differences (differential methylation levels) represent the average of [(β-value of tumorous tissue) - (β-value of paired non-tumorous tissue)] in 12 stage-I LADC cases.

Supplementary Table S8. Cox proportional hazard regression analysis of overall survival in 162 patients with LADC

Factor	Univariate			Multivariate		
	Hazard ratio	95% confidence interval	*P*-value	Hazard ratio	95% confidence interval	*P*-value
Sex Male (n=81) vs. Female (n=81)	3.219	1.452 - 7.138	**0.004**	1.311	0.220 - 7.802	1.311
Age (years) >67 (n=87) vs. ≤67 (n=75)	2.471	1.169 - 5.224	**0.018**	2.562	1.073 - 6.120	**0.034**
Smoking history[a] Positive (n=77) vs. Negative (n=83)	4.177	1.817 - 9.602	**0.001**	2.166	0.341 - 13.759	0.413
Pathological stage II, III (n=58) vs. I (n=104)	4.328	1.999 - 9.372	< **0.001**	-	-	-
Tumor size[a] pT2-4 (n=39) vs. pT1 (n=115)	2.262	1.119 - 4.573	**0.023**	2.466	1.116 - 5.447	**0.026**
N stage (pN) pN1-3 (n=44) vs. pN0 (n=118)	3.577	1.789 - 7.151	< **0.001**	0.909	0.343 - 2.410	0.848
Pleural invasion[a] Positive (n=40) vs. Negative (n=112)	2.051	0.987 - 4.264	0.054	1.635	0.626 - 4.267	0.315
Vascular invasion[a] Positive (n=27) vs. Negative (n=119)	2.735	1.284 – 5.826	**0.009**	0.487	0.157 - 1.512	0.213
Lymph vessel invasion[a] Positive (n=37) vs. Negative (n=105)	4.700	2.203 - 10.027	< **0.001**	3.897	1.311 - 11.580	**0.014**
Adjuvant chemotherapy[a] With (n=47) vs. Without (n=106)	0.996	0.472 - 2.101	0.991	-	-	-
EGFR mutation[a] Negative (n=35) vs. Positive (n=40)	2.882	1.151 - 2.564	**0.024**	-	-	-
Predominant subtype Non-lepidic (n=103) vs. Lepidic (n=59)	3.311	1.156- 9.524	**0.026**	2.841	0.590 - 13.699	0.193
GAD1 immunoreactivity Positive (n=112) vs. Negative (n=50)	2.315	0.895 - 5.992	0.084	1.216	0.366 - 4.042	0.750

Statistically significant values are in boldface type.

[a]Data of these factors were not available for all patients.

Supplementary Table S9. Cox proportional hazard regression analysis of cancer-specific survival in 162 patients with LADC

Factor	Univariate			Multivariate		
	Hazard ratio	95% confidence interval	*P*-value	Hazard ratio	95% confidence interval	*P*-value
Sex Male (n=81) vs. Female (n=81)	1.945	0.827 - 4.576	0.127	1.160	0.194 - 6.923	0.871
Age (years) >67 (n=87) vs. ≤67 (n=75)	1.735	0.749 - 4.021	0.199	1.932	0.748 - 4.987	0.173
Smoking history[a] Positive (n=77) vs. Negative (n=83)	2.599	1.076 - 6.279	**0.034**	1.398	0.221 - 8.839	0.722
Pathological stage II, III (n=58) vs. I (n=104)	7.706	2.606 - 22.791	< **0.001**	-	-	-
Tumor size[a] pT2-4 (n=39) vs. pT1 (n=115)	1.994	0.847 - 4.697	0.114	1.867	0.733 - 4.758	0.191
N stage (pN) pN1-3 (n=44) vs. pN0 (n=118)	6.066	2.488 - 14.789	< **0.001**	1.322	0.398 - 4.395	0.649
Pleural invasion[a] Positive (n=40) vs. Negative (n=112)	2.331	0.957 - 5.677	0.063	1.255	0.429 - 3.673	0.679
Vascular invasion[a] Positive (n=27) vs. Negative (n=119)	4.089	1.697 – 9.854	0.002	0.892	0.236 - 3.378	0.867
Lymph vessel invasion[a] Positive (n=37) vs. Negative (n=105)	5.610	2.239 - 14.055	< **0.001**	2.654	0.703 - 10.022	0.150
Adjuvant chemotherapy[a] With (n=47) vs. Without (n=106)	1.036	0.428 - 2.504	0.938	-	-	-
EGFR mutation[a] Negative (n=35) vs. Positive (n=40)	3.165	1.188 - 8.403	**0.021**	-	-	-
Predominant subtype Non-lepidic (n=103) vs. Lepidic (n=59)	9.804	1.311- 71.429	**0.026**	3.378	0.392 - 29.411	0.268
GAD1 immunoreactivity Positive (n=112) vs. Negative (n=50)	4.323	1.015 - 18.420	**0.048**	3.400	0.415- 27.827	0.254

Statistically significant values are in boldface type.

[a]Data of these factors were not available for all patients.

511 **Figure Legends**

512

513 **Figure 1.** DNA methylation and mRNA expression status of *GAD1* in patients with
514 LADC.
515 (A) A schematic diagram of the *GAD1* gene structure with CGIs around *GAD1*. The
516 arrow indicates the location of cg15126544.
517 (B) The average β-value (methylation level) of each CpG site targeted in the array-based
518 methylation experiment involving 12 LADC cases. **P* < 0.05 vs. paired non-tumorous
519 tissues.
520 (C) Linear plots of the average DNA methylation values (percentages) of cg15126544 in
521 33 LADC tumorous and paired non-tumorous tissues, as determined by quantitative
522 pyrosequencing. Samples from the same patient are linked with straight lines.
523 (D) Linear plots of expression levels of *GAD1* mRNA relative to those of the control
524 normal human lung in 33 LADC tumorous and paired non-tumorous tissues. Relative
525 expression of *GAD1* mRNA was calculated using Human Lung Total RNA as a normal
526 control.
527 (E) Correlation between the average methylation levels of cg15126544 (x-axis) and
528 relative *GAD1* mRNA expression levels (y-axis) in 33 LADC tumorous and paired non-
529 tumorous tissues.
530 (F) Linear plots of the methylation levels (β-values) of cg15126544 determined through
531 an array-based methylation experiment using HumanMethylation450K array in 18
532 paired LADC tumor and non-tumorous tissue samples obtained from the TCGA dataset
533 (http://cancergenome.nih.gov).

(G) Linear plots of mRNA expression of *GAD1* determined by RNA sequencing and quantified by RNA-Seq by Expectation Maximization (RSEM) in 18 paired LADC tumor and non-tumorous tissue samples obtained from the TCGA dataset.
(H) Correlation between the methylation levels (β-values) of cg15126544 (x-axis) and *GAD1* mRNA expression levels (y-axis) in 18 paired LADC tumor and non-tumorous tissue samples obtained from the TCGA dataset.

Figure 2. Publicly available datasets showing association between *GAD1* mRNA expression status and prognosis in patients with LADC.
(A) Kaplan-Meier curve for OS rate of 423 LADC patients according to *GAD1* mRNA expression levels using data obtained from the TCGA dataset. *P*-values were calculated using the log-rank test. Statistically significant *P*-values are in boldface type.
(B) Kaplan-Meier curve for OS rate of 720 LADC patients in cohorts GSE14814, GSE19188, GSE3141, GSE50081, GSE31908, GSE37745, GSE29013, GSE30219, and GSE31210 according to *GAD1* mRNA expression levels obtained from the online survival analysis software, Kaplan–Meier plotter (KM plotter; http://www.kmplot.com). *P*-values were calculated using the log-rank test. Statistically significant *P*-values are in boldface type.
(C) Subgroup analysis of KM plotter databases for *GAD1* mRNA expression in LADC. Hazard ratios (HR, center of the box) and 95% confidence intervals (CI, horizontal line) were calculated with Cox's regression models.

Figure 3. Association between GAD1 protein expression status and prognosis in patients with LADC.

(A) Representative images of immunohistochemically detected GAD1 protein in tumors and non-tumorous lesions of LADC samples and normal lung tissue. Scale bars, 200 μm. The relative GAD1 mRNA expression level of each sample as determined by rqRT-PCR is also shown.

(B) Kaplan-Meier curves for overall survival, disease-free survival, and cancer-specific survival rates of 162 LADC patients according to the immunoreactivity of GAD1. *P*-values were calculated using the log-rank test. Statistically significant *P*-values are in boldface type.

Supplementary Figure Legends

Supplementary Figure S1. The average β-value (methylation level) of each CpG site targeted in the Infinium HumanMethylation450K data (Level 3) of 29 paired LADC tumor and non-tumorous tissue samples downloaded from TCGA Research Network (http://cancergenome.nih.gov). Data of some CpG sites were missing in the TCGA dataset. *$P < 0.05$ vs. paired non-tumorous tissues.

Supplementary Figure S2. Kaplan-Meier overall survival curves of all selected datasets from KM plotter used in the present study (see Figure 2C). Hazard ratios (HR) and 95% confidence intervals in parentheses are shown for each dataset. *P*-values were calculated using the log-rank test.

ACKNOWLEDGMENT

The authors would like to thank Dr. Yoshimi Bando, Division of Pathology, Tokushima University Hospital for her contributing to this study. This work was supported by JSPS KAKENHI (grant number JP18H02894). The funding source was not involved in the design of the study, in the collection, analysis, and interpretation of results, in the preparation of the manuscript, or in the decision to submit the article for publication.

CONFLICT OF INTEREST

All authors declare no conflicts of interest associated with this manuscript.

Figure 1

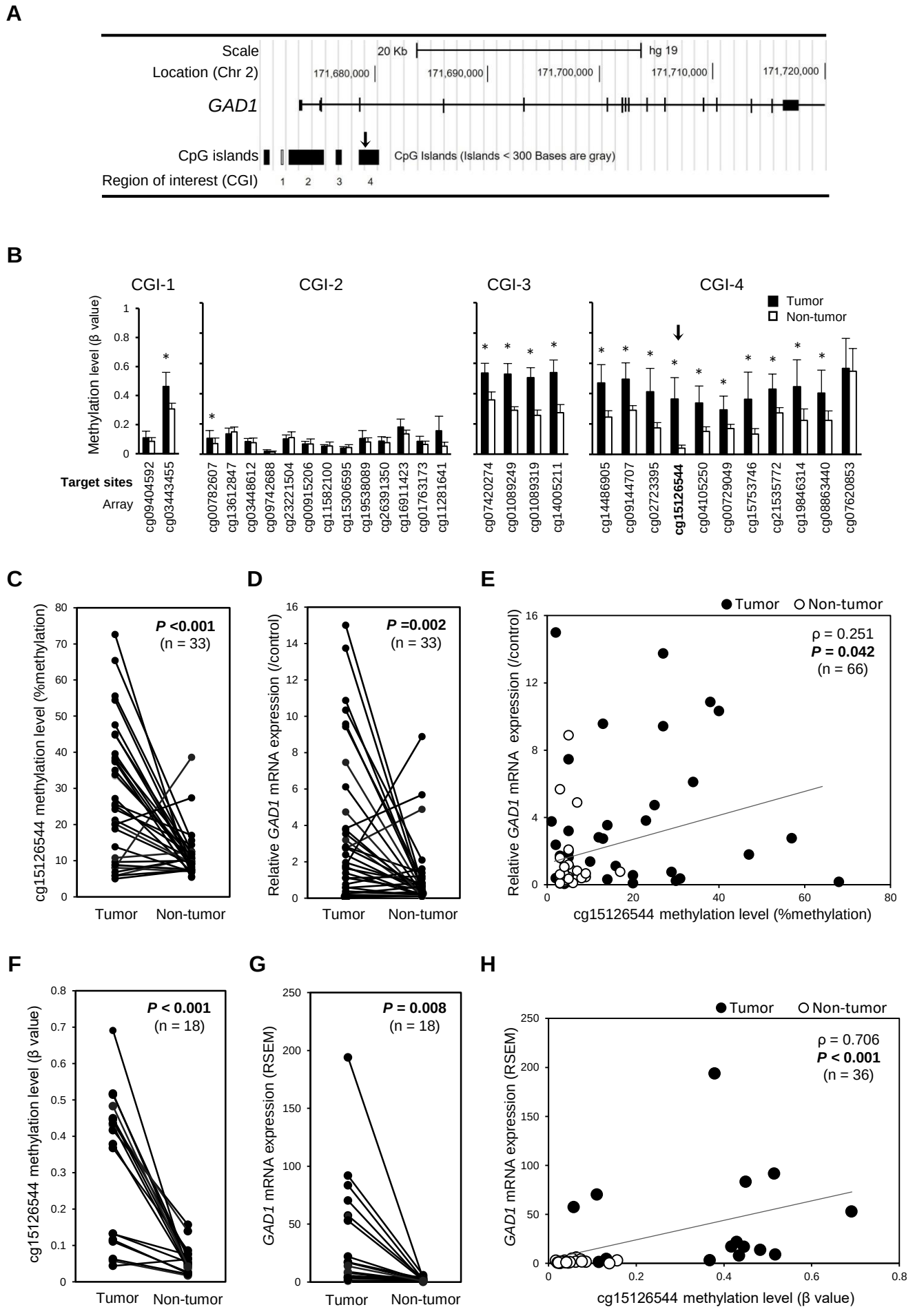
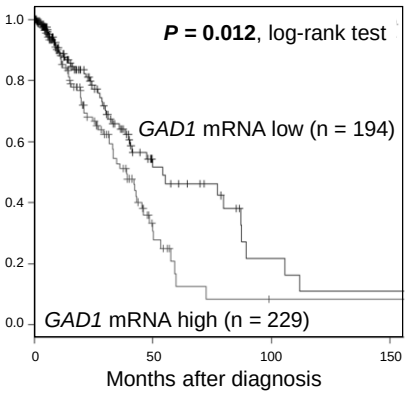
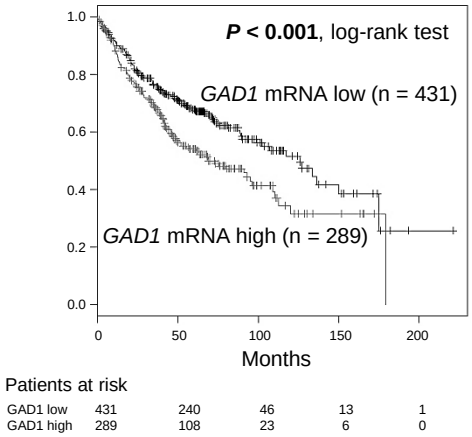

A
Scale
20 Kb
hg 19
Location (Chr 2)
171,680,000
171,690,000
171,700,000
171,710,000
171,720,000
GAD1
CpG islands
CpG Islands (Islands < 300 Bases are gray)
Region of interest (CGI)
1 2 3 4
B
CGI-1
CGI-2
CGI-3
CGI-4
Tumor
Non-tumor
Methylation level (β value)
Target sites
Array
cg09404592
cg03443455
cg00782607
cg13612847
cg03448612
cg09742688
cg23221504
cg00915206
cg11582100
cg15306595
cg19538089
cg26391350
cg16911423
cg01763173
cg11281641
cg07420274
cg01089249
cg01089319
cg14005211
cg14486905
cg09144707
cg02723395
cg15126544
cg04105250
cg00729049
cg15753746
cg21535772
cg19846314
cg08863440
cg07620853
C
P <0.001
(n = 33)
cg15126544 methylation level (%methylation)
Tumor
Non-tumor
D
P =0.002
(n = 33)
Relative GAD1 mRNA expression (/control)
Tumor
Non-tumor
E
Tumor
Non-tumor
ρ = 0.251
P = 0.042
(n = 66)
Relative GAD1 mRNA expression (/control)
cg15126544 methylation level (%methylation)
F
P < 0.001
(n = 18)
cg15126544 methylation level (β value)
Tumor
Non-tumor
G
P = 0.008
(n = 18)
GAD1 mRNA expression (RSEM)
Tumor
Non-tumor
H
Tumor
Non-tumor
ρ = 0.706
P < 0.001
(n = 36)
GAD1 mRNA expression (RSEM)
cg15126544 methylation level (β value)

Figure 2

A



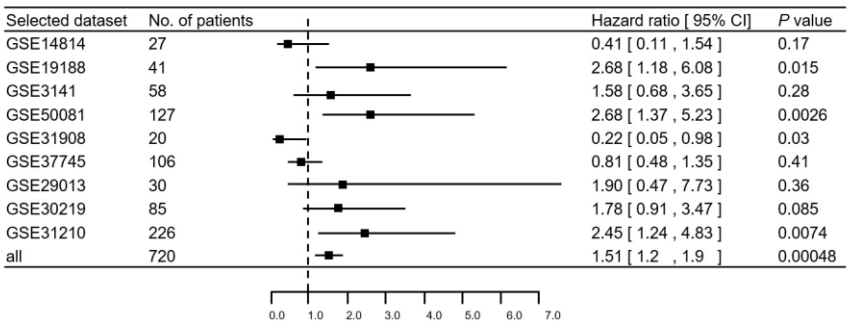
B



Patients at risk

GAD1 low	431	240	46	13	1
GAD1 high	289	108	23	6	0

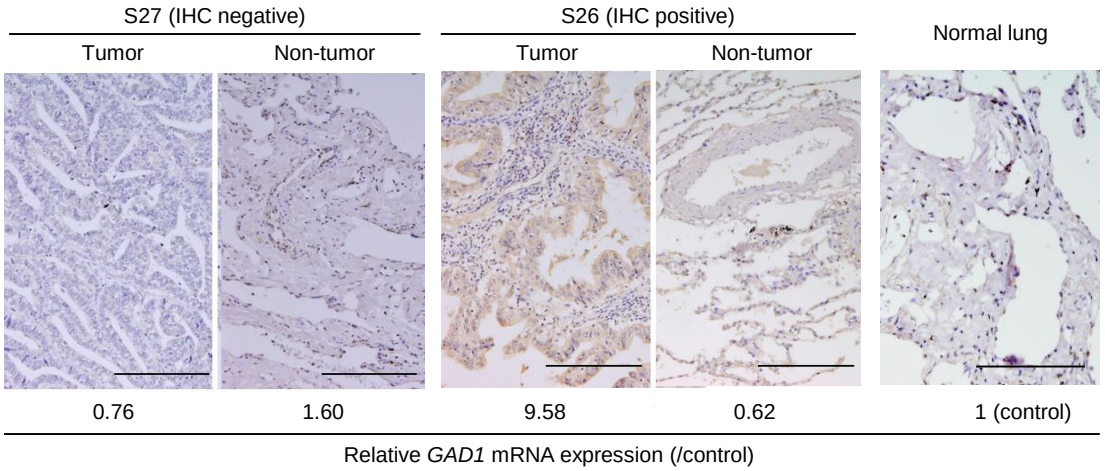
C



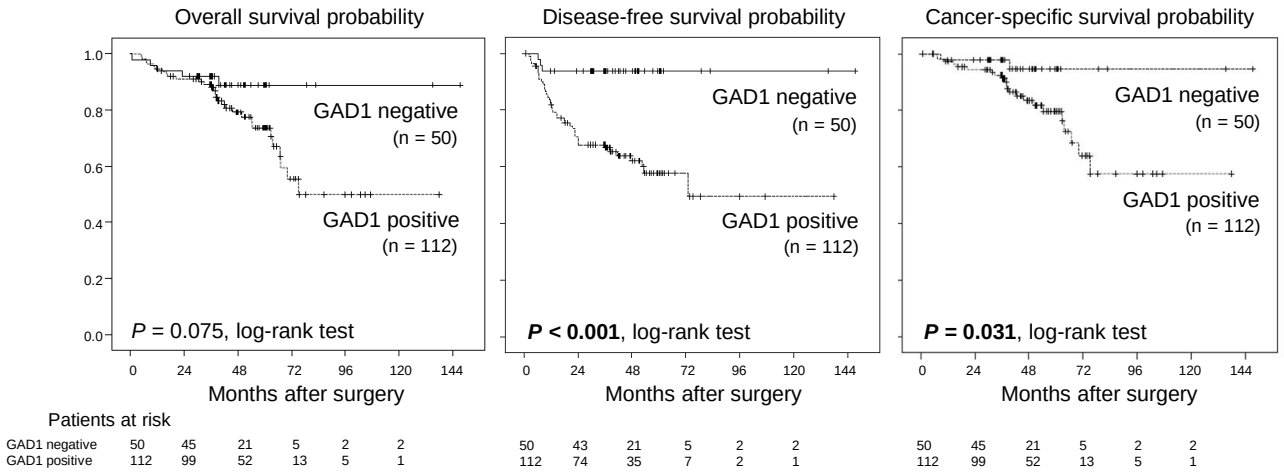
Selected dataset	No. of patients	Hazard ratio [95% CI]	*P* value
GSE14814	27	0.41 [0.11 , 1.54]	0.17
GSE19188	41	2.68 [1.18 , 6.08]	0.015
GSE3141	58	1.58 [0.68 , 3.65]	0.28
GSE50081	127	2.68 [1.37 , 5.23]	0.0026
GSE31908	20	0.22 [0.05 , 0.98]	0.03
GSE37745	106	0.81 [0.48 , 1.35]	0.41
GSE29013	30	1.90 [0.47 , 7.73]	0.36
GSE30219	85	1.78 [0.91 , 3.47]	0.085
GSE31210	226	2.45 [1.24 , 4.83]	0.0074
all	720	1.51 [1.2 , 1.9]	0.00048

Figure 3

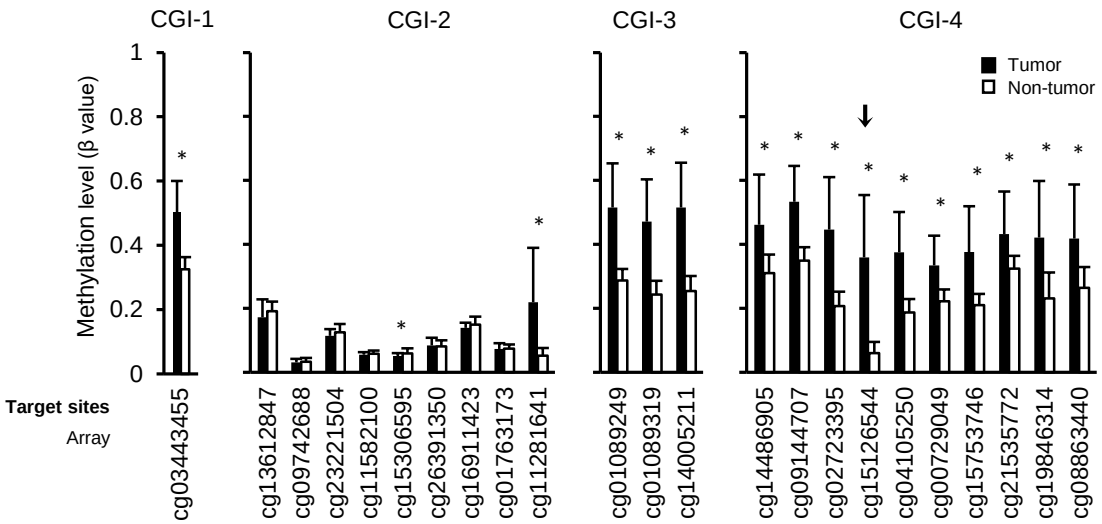
A


S27 (IHC negative)
S26 (IHC positive)
Normal lung
Tumor
Non-tumor
Tumor
Non-tumor
0.76
1.60
9.58
0.62
1 (control)
Relative *GAD1* mRNA expression (/control)

B

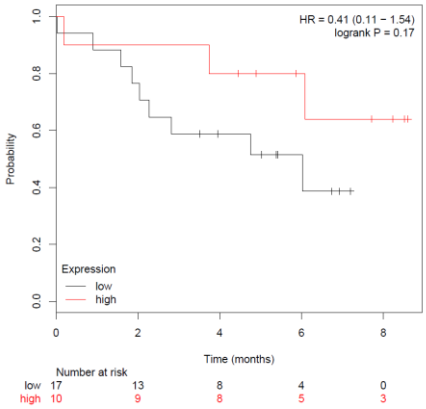

Overall survival probability
Disease-free survival probability
Cancer-specific survival probability
GAD1 negative
(n = 50)
GAD1 positive
(n = 112)
P = 0.075, log-rank test
***P* < 0.001**, log-rank test
***P* = 0.031**, log-rank test
Months after surgery
Patients at risk
GAD1 negative 50 45 21 5 2 2
GAD1 positive 112 99 52 13 5 1
50 43 21 5 2 2
112 74 35 7 2 1
50 45 21 5 2 2
112 99 52 13 5 1

Supplementary Figure S1

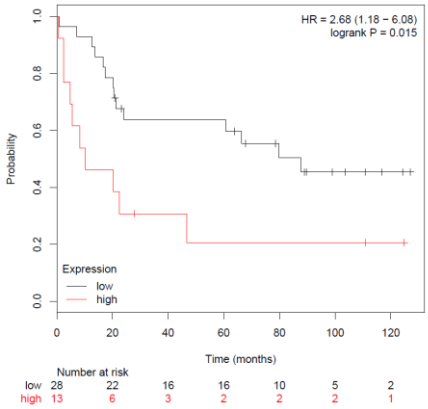


Supplementary Figure S2

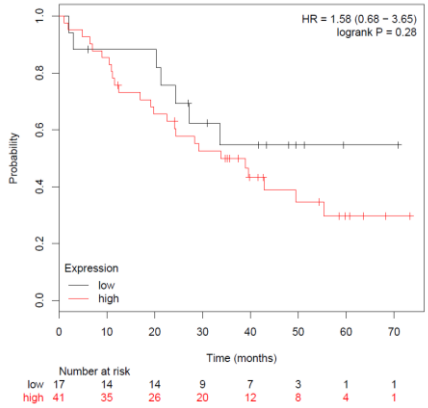
GSE14814



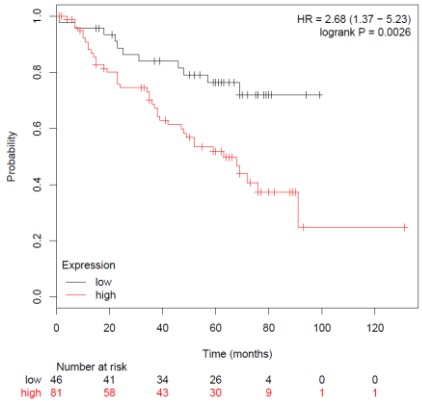
GSE19188



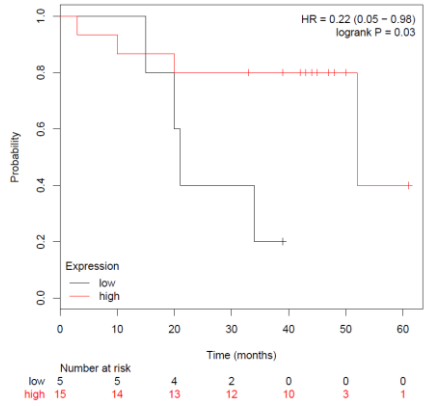
GSE3141



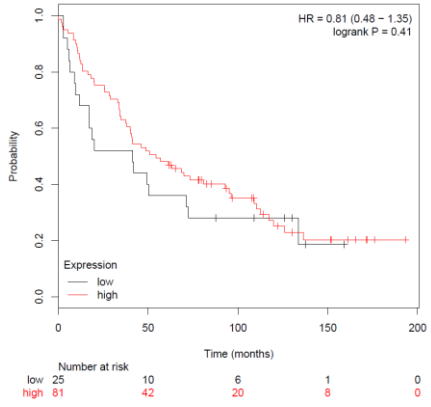
GSE50081



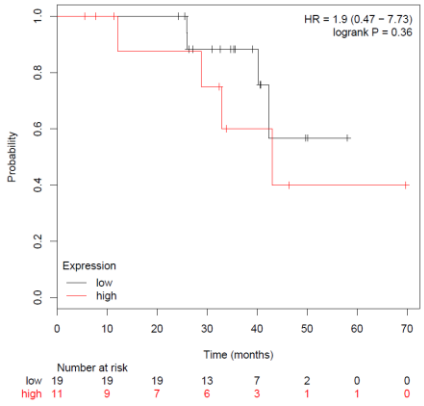
GSE31908



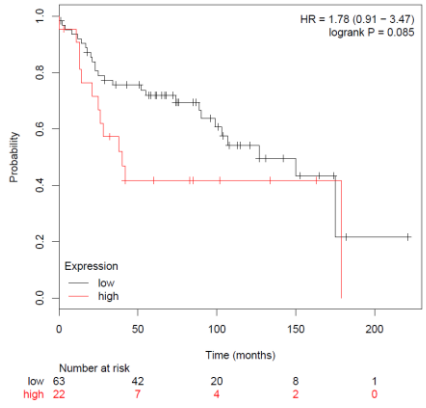
GSE37745



GSE29013



GSE30219



GSE31210