筑　波　大　学

博 士 （ 医 学 ） 学 位 論 文

Identification of DNA methylation alteration during the course of lung adenocarcinogenesis

(肺線癌発生過程における DNA メチル化変化の解析)

2０１８

筑波大学大学院博士課程人間総合科学研究科

Ryan Edbert Husni

Table of contents

Abstract

Chapter 1: Introduction

Chapter 2: DNMT3a expression pattern and its prognostic value in lung adenocarcinoma

Chapter 3: Identification of differentially methylated genes in lung adenocarcinoma

List of figures

List of tables

Abstract

Lung cancer is still the leading cause of cancer death with varied survival rate. Among the histological subtypes, adenocarcinoma is the most frequent. Lung adenocarcinoma has been known to show a stepwise progression from precancerous lesion, adenocarcinoma *in situ* (AIS), to invasive adenocarcinoma, and also closely correlated with gene aberrations. While many new treatments targeted specific genetic aberrations have been developed for advanced stage adenocarcinoma, patients treated with them eventually acquire resistance against them, failing to decrease mortality rate. Although studies on early stage lung adenocarcinoma are still scarce, previous studies in our group have shown that overexpression of particular genes are correlated with progression of lung adenocarcinoma and most of them seems not to have genetic aberrations, which means that the abnormality of their expression might be epigenetically induced.

DNA methylation status has been reported to be correlated with the progression of adenocarcinoma. In previous study, our group has demonstrated that overexpression of Stratifin (SFN, 14-3-3 sigma) in invasive adenocarcinoma is triggered by DNA demethylation at SFN promoter region. Based on this finding, I determined to focus on DNA methylation.

My first study is about DNA Methyltransferase 3 alpha (DNMT3a), which is a key enzyme of the methylation pathway. I used IHC to demonstrate DNMT3a expression pattern in lung adenocarcinoma, and also Western blot to confirmed the specificity of the antibody used for IHC. Consequently, I found that low DNMT3a expression is associated with histologically invasive type and a poor prognosis.

Based on the fact of DNMT3a dysfunction, I expected that there would be many oncogenes that turn to demethylated status and facilitate tumor progression besides SFN. Thus the aim of my second study is to find another differentially methylated genes that have the same tendency as SFN, which is correlated with early stage adenocarcinogenesis.

For that purpose, 3 samples of invasive adenocarcinoma, 3 samples of AIS, and 2 samples of normal lung tissue were subjected to Infinium methylation array to screen extensive DNA methylation profiles across the whole genome. From the results, I found that 583 CpG sites showed more than 10% higher methylation rate in invasive adenocarcinoma compared to AIS and normal lung. On the other hands, only 23 CpG sites including SFN locus showed more than 10% lower methylation rate in invasive adenocarcinoma relative to AIS and normal lung. Among the later, we finally selected 6 CpG sites located in SFN, GORASP2, CD1D, ZYG11A, LOC10099657, and Mir656 and validated the result using 21 cases of lung adenocarcinoma by pyrosequencing. As a result, SFN, GORASP2, and ZYG11A showed stepwise demethylation tendency from normal lung, AIS to invasive adenocarcinoma as I expected. Moreover, its methylation rate is conversely correlated with the protein expression, suggesting that hypomethylation at those sites might lead to their overexpression. I next performed immunohistochemistry of GORASP2 and ZYG11A to clarify their clinicopathological implication for lung adenocarcinoma. I demonstrated that GORASP2 and ZYG11A show high expression in lung adenocarcinoma and were associated with histologically invasive subtype, and a poor prognosis. I also found that GORASP2 and ZYG11A are independent prognostic factors for lung adenocarcinoma.

In this study, I made the status of methylation alteration in lung adenocarcinoma clear. DNMT3a expression decreased in invasive adenocarcinoma and methylated genes have been widely changed in the course of malignant progression. Even though the number of hypomethylated genes in invasive adenocarcinoma was limited compared to those of hypermethylated genes, I found overexpression of GORASP2 and ZYG11A induced by DNA demethylation other than SFN. The epigenetic changes of the two genes may have an important function in the progression of lung adenocarcinoma. GORASP2 and ZYG11A might be clinically applicable as an indicator of prognosis and potential novel target molecule for drug development.

Chapter 1: Introduction

1.1 Prologue to lung adenocarcinoma

Lung cancer is not only the leading cause of cancer death in man but also the second cause of cancer death in woman worldwide. Age, sex, ethnicity, socioeconomic status, and smoking patterns make variations in lung cancer rates and trends. Globally, although the rates are plateauing or decreasing in the high-income country, it still increasing in the low-income countries and more than 50% of lung cancer deaths each year come from these countries [1, 2]. Thus, lung cancer will still be a major problem in the first half of this century with 5-years survival rate varied between 2-30% due to country differences [3].

Among the histological subtypes, adenocarcinoma is the most frequent and its incidence is still increasing [4-6] (Fig.1). The 5-year survival rate even from lung adenocarcinoma stage IA is 68.6%, which is still unsatisfactory [7]. However, Noguchi *et al*. have shown that very early stage of lung adenocarcinoma (type A and B of Noguchi classification) has a very favorable prognosis with a 5-years survival rate of 100% [8, 9]. In 2011, Travis *et al*. introduced a new concept of early stage lung adenocarcinoma into adenocarcinoma *in situ* (AIS) and minimally invasive adenocarcinoma (MIA) [10], which is closely related to Noguchi classification. This concept was accepted in the latest edition WHO classification of lung tumors [11]. Lung adenocarcinogenesis shows a multi step progression starting from pre-cancerous lesion adenomatous atypical hyperplasia (AAH) to AIS, MIA, and finally invasive lung adenocarcinoma, which has a poorer survival rate [12].

1.2 Genetic aberrations of lung adenocarcinoma and its recent treatment

Based on the comprehensive molecular study, lung adenocarcinoma has been known for its various genetic aberrations. While the frequency of gene aberrations shown difference among populations, they occur in a mutually exclusive manner. Well-known driver oncogene aberrations include EGFR, KRAS, BRAF, and HER2 driver mutations, ALK, RET, ROS1 fusion, and skipping of exon 14 of MET genes [13, 14] (Fig.2). Stepwise progression of lung adenocarcinoma is also closely associated with genetic alteration. Progression from AAH to AIS involves EGFR or KRAS mutation with p16 inactivation. Furthermore, p53 mutation and allelic imbalances of various chromosomes lead progression from AIS to invasive adenocarcinoma [15] (Fig.3).

Adding to current treatments for lung adenocarcinoma other than conventional surgery, chemotherapy, and radiotherapy, recent treatment such as targeted therapy and immunotherapy have been researched and reported extensively. EGFR targeted drugs such as erlotinib and gefitinib, EML4-ALK and other ALK fusion targeted drug such as crizotinib, anti-PD-1 agent such as nivolumab, anti-PD-L1 agent such as MK-3475, and an anti-CTLA-4 agent such as ipilimumab have been investigated [16, 17].

Even though many targeted drugs have already been approved and show efficacy on patients with specific genetic aberrations, most of the drug developed resistance especially in the advanced stage lung adenocarcinoma and have failed to decrease mortality rate [18-20]. I speculated that this is because advanced stage of such tumors harbors numerous genetic abnormalities, namely high tumor mutation burden. On the other hand, the number of genetic abbreviation in early stage lung adenocarcinoma is limited, and epigenetic alteration is appeared to be predominant rather than genetic

ones in this stage [21]. However, even though there are so many studies about advanced stage lung adenocarcinoma, study about early stage lung adenocarcinoma is still scarce. Thus, discovering epigenetic and genetic aberrations in the early stage lung adenocarcinoma such as AIS will be useful for early diagnosis and targeted therapy.

In previous studies of our group, we have shown several genes that show overexpression in early invasive lung adenocarcinoma such as Stratifin (SFN) [22], Immunoglobulin binding protein 1 (IGBP1) [23], Ovarian carcinoma immunoreactive antigen domain 2 (OCIAD2) [24], and Dimethylarginine dimethylaminohydrolase 2 (DDAH2) [25]. We expect that overexpression of all aforementioned genes are not associated with genome alteration and thought to be associated with epigenetic abnormality.

1.3 Epigenetic studies in lung adenocarcinoma

By recent advancement of technologies, it is now well known that cancer cells not only harbor genomic abnormalities but also epigenetic alterations [26]. Epigenetics define as a study about changes in the function of a gene that are heritable without a change in the sequence of DNA [27]. In the present time, the most studied epigenetic abnormalities are histone protein structure alteration, gene regulation by microRNA, and DNA methylation [21].

Acetylation and deacetylation, which is mediated by histone acetlytransferase (HAT) and histone deacetylase (HDAC) respectively has been well-studied mechanism of epigenetic alteration. Hyperacetylation has been associated with transcriptionally active gene and can activate proto-oncogenes, while on the other hand hypoacetylation correlate with decrease transcriptional action and might leading to silencing of tumor suppressor genes [21, 28]. In the recent study, one of the major families of HAT called GCN5 via interaction with transcription factor E2F1 can promote non-small cell lung cancer growth [29].

Another epigenetic alteration is raised by miRNAs. Generally, miRNAs are involved in regulating apoptosis, differentiation, the progression of cell cycle, and other important processes in the cell, and alteration of them via epigenetic pathways are shown in numerous cancers [30]. In the recent study, up-regulation of mIR-9 by TGF-β1 contributes to TGF-β1-induced non-small cell lung cancer cell adhesion and invasion [31]. Our group first showed that overexpression of IGBP1 in invasive adenocarcinoma is due to down-regulation of miR3941 and has anti-apoptotic function [23].

Abnormal DNA methylation is also important part of epigenetic changes and it has been reported to be correlated with the progression of adenocarcinoma [32, 33]. Hypermethylation can silence tumor suppressor genes function and result in the progression of the tumor [34]. On the other hand, global hypomethylation of genomic DNA can lead to genetic alteration [35] and also another phenomenon such as overexpression of proto-oncogenes and growth factors [36], resulting in tumorigenesis. Our group has shown that overexpression of SFN is correlated with DNA methylation status. Although the promoter region of SFN is totally methylated in normal lung and AIS by demethylation of the promoter region, most invasive adenocarcinomas show its overexpression [37].

Based on the SFN finding and the importance of epigenetic alteration study in lung adenocarcinoma, I determined to focus on abnormality of DNA methylation in early stage lung adenocarcinomas.

Chapter 2: DNMT3a expression pattern and its prognostic value in lung adenocarcinoma

2.1 Introduction

One of the important components correlated with the methylation pathway is the DNA methyltransferases (DNMTs), which transfer a methyl group to the C-5 position on the cytosine ring of DNA [38]. DNMT1, 3a, and 3b have catalytic activity in the context of methylation status. DNMT1 acts mainly as maintenance methyltransferase, while DNMT3a and 3b play roles in *de novo* methylation [39].

Studies of DNMT expression and its clinicopathological correlations have been conducted on various human cancers, including ovarian cancer, retinoblastoma, and gastroenteropancreatic neuroendocrine tumors. It has been reported that DNMT3b cooperates with HDAC in controlling the progression of ovarian cancer, while overexpression of DNMT1, 3a, and 3b correlates with tumorigenesis in retinoblastoma and gastroenteropancreatic neuroendocrine tumors [40-42]. Although DNMT1 and 3b have been reported to play a specific role in tumorigenesis by silencing tumor suppressors [43-46], the role of DNMT3a in cancer pathophysiology has remained unclear. While it has been reported that DNMT3a expression is significantly associated with the prognosis of gastric and hepatocellular carcinomas [47, 48], its correlation with the histological features or prognosis of lung adenocarcinoma has not yet been clarified.

The purpose of the present study was to evaluate the expression pattern of DNMT3a in lung adenocarcinoma and examine its implication for lung adenocarcinogenesis and malignant progression.

2.2 Materials and methods

2.2.1 Patients and sample selection

One hundred thirty five lung adenocarcinomas surgically resected at University of Tsukuba Hospital (Ibaraki, Japan) between 2002 and 2012 were selected and examined for this study. The patients comprised 71 males and 65 females ranging in age from 34 to 85 years with mean age of 67.04 years. Follow-up information obtained from medical records was available for all of the selected patients. Informed consent for this study had been obtained from all of the patients. Formalin-fixed and paraffin-embedded samples were used for immunohistochemistry (IHC).

2.2.2 Cell lines and culture conditions

The cell lines HepG2 and A549 were purchased from RIKEN BRC (Ibaraki, Japan). PL16T cells were established in our laboratory from human lung atypical adenomatous hyperplasia (AAH) [49]. HepG2 was maintained in RPMI 1640 (Life Technologies, Carlsbad, CA) supplemented with 5% fetal bovine serum (FBS) (Corning Inc., Corning, NY), and A549 was maintained in D-MEM/F12 (Life Technologies) supplemented with 10% FBS. PL16T was maintained in MCDB153HAA (Wako, Osaka, Japan) supplemented with 2% FBS. All cells were cultured in a 5% CO_2 incubator at 37°C. All were used for Western blotting.

2.2.3 Immunohistochemistry (IHC)

17

For IHC, we used 4-micrometer-thick sections cut from formalin-fixed and paraffin-embedded blocks. The sections were deparaffinized and rehydrated. For antigen retrieval, I autoclaved the sections in 10 mmol/L Tris-EDTA buffer (pH 9.0) at 121°C for 10 min. Further steps were performed on an automated stainer, Histostainer 48a (Nichirei Biosciences, Tokyo, Japan). Endogenous peroxidase was blocked with 3% hydrogen peroxide. The slides were incubated in a 1:100 dilution of rabbit polyclonal DNMT3a antibody (ABGENT, San Diego, CA) at RT for 1 hour, and subsequently incubated in the secondary antibody at RT for 1 hour. The signal was detected using DAB (Dako REAL Envision Detection System; Dako, Glostrup, Denmark). Hematoxylin was used for counterstaining.

I evaluated all cases without prior knowledge of the clinicopathological data. One thousands tumor cells were evaluated for the most intense nuclear staining ("hot spot") in accordance with previous report [50] with minor modification. Surrounding non-neoplastic bronchial epithelium served as an internal negative control. DNMT3a IHC score was determined by counting the number of stained nuclei. The ROC curve was drawn to determine the best cut-off point of the score. (Fig.4) Since the aim of the study is to clarify the correlation between DNMT3a expression and the prognosis of the patient, I used DNMT3a expression and the outcome of the patient as variables to draw ROC curve. After drawing the curve we draw diagonal line from bottom right into top left corner and the cut off point is the coordinates where the diagonal line cross over the curve. Here, the coordinate falls in 57.5 (0.622 sensitivity and 0.351 (1-specificity)), thus it was adopted as the cut off point in this study. Counts below 57.5 were judged as weak expression, and those above 57.5 as strong expression.

2.2.4 Transfection using siRNA

A549 cells were seeded at a density of 6.0 x 10^6/mL and incubated overnight in antibiotics-free D-MEM supplemented with 10% FBS. On the next day, the cells were washed with phosphate-buffered saline (PBS), and then OPTI- MEM reduced serum medium (Life Technologies) was added to the cells. Three sets of DNMT3a-specific siRNA (Life Technologies) (Table 1) and a nucleic acid transferring agent, lipofectamine RNAiMAX (Life Technologies), were incubated together in OPTI- MEM reduced serum medium for 20 min at room temperature to form an siRNA-lipofectamine complex. The medium containing the siRNA-lipofectamine complex was added to the cells to give a final siRNA concentration of 20 nM. The cells were incubated at 37^oC in a CO_2 incubator for 48h, and then western blotting was carried out with Stealth RNAi Negative Control Medium GC Duplex #2 (Life Technologies) was used as a negative control.

2.2.5 Western blotting

Total cell lysates were prepared on ice with M-PER (Life Technologies) containing protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO) and phosphatase inhibitor cocktail (Sigma-Aldrich). The lysates were centrifuged for 10 min at 4°C, and the insoluble fractions were discarded. The total protein in the soluble lysates was measured using a BCA protein assay kit (Life Technologies). Twenty micrograms of protein was mixed with sample buffer and denatured at 95°C for 5 min, followed by electrophoresis on a 7.5% Tris-HCl gel (Bio-Rad, Hercules, CA). An iBlot gel transfer system (Life technologies) was used to transfer the protein. Rabbit polyclonal DNMT3a antibody (1:4000, ABGENT) and mouse monoclonal

anti-beta actin antibody (1:5000, Sigma-Aldrich) were used as primary antibodies. To visualize the protein bands, Super Signal West Femto Maximum Sensitivity Substrate (Life Technologies) and care stream Kodak Biomax Light Film (Sigma-Aldrich) were used.

2.2.6 Statistical analysis

For all statistical analyses, SPSS 22 (SPSS, Chicago, IL) was used. For determining the cut-off point for IHC scoring, the ROC curve method was used. Correlation of clinicopathological features with DNMT3a expression was analyzed by chi-squared test. The Kaplan-Meier method was used for calculation of survival curves, and log-rank test was performed for comparisons. Multivariate analysis was done using the Cox proportional hazards model. Differences were considered statistically significant at $p \leq 0.05$.

2.3 Results

2.3.1 Western blotting of DNMT3a

In accordance with a previous report [51], I confirmed a single band of DNMT3a in HepG2 used as a positive control. Also, I detected specific bands of DNMT3a in the lung adenocarcinoma cell line A549 and the AAH cell line PL16T (Fig.5a). To further analyze the specificity of the antibody, I carried out transfection using specific siRNA for DNMT3a (siDNMT3a). As a result, all siDNMT3a shown suppression on the protein expression thus demonstrating the specificity of the anti-DNMT3a antibody used for IHC (Fig.5b).

2.3.2 Immunohistochemistry for DNMT3a

In lung adenocarcinoma, DNMT3a was stained not only in the nucleus of tumor cells but also in both the cytoplasm and nucleus (Fig. 6). In most cases, stronger positive staining was observed in nuclei. Since DNMT3a mainly functions in the nucleus, I adopted nuclear immunoreactivity for evaluation. Surrounding non-neoplastic bronchial epithelium was used as an internal negative control. Among the 135 samples, weak staining was found in 56 and strong staining in 79 (Table 2, Fig. 7).

2.3.3 Expression pattern of DNMT3a in lung adenocarcinoma

Among invasive lung adenocarcinomas, strong expression of DNMT3a was detected in 41/53 (77%) lepidic, 8/20 (40%) papillary, 13/31 (42%) acinar, and 2/15 (13%) solid adenocarcinomas. Among non-invasive lung adenocarcinomas, strong expression was detected in 13/14 (92.9%) AIS and 2/2 (100%) MIA (Table 2, Fig. 7).

I also assessed the correlation between DNMT3a expression and clinicopathological features of the patients using the chi-squared method. Although DNMT3a expression showed no significant correlation with age, sex, or lymphatic permeation it was significantly correlated with the pathological stage, lymph node status, pleural invasion, vascular invasion, and pathological subtype of lung adenocarcinoma (Table 2). DNMT3a showed significantly higher expression in lepidic adenocarcinomas (41/53 cases, 77%) than in other adenocarcinomas (39/83 cases, 47%).

2.3.4 Correlation of DNMT3a expression with postoperative overall survival

The Kaplan-Meier curves indicated that the patients in the strong expression group had a significantly more favorable outcome than those in the weak expression group (Fig. 8, p<0.001).

2.3.5 Multivariate analysis using the Cox proportional hazards model

After adjustment for gender, age, pathological stage, pleural invasion, vascular invasion, lymphatic permeation, and DNMT3a expression, patients with strong expression of DNMT3a showed a significantly lower risk of lung cancer-related death than those with weak expression (HR: 0.72, 95%CI: 0.529-0.974, P: 0.033).

Multivariate analysis also indicated that lymphatic permeation, pleural invasion, and DNMT3a expression were independent prognostic factors indicative of poor survival in patients with lung adenocarcinoma (Table 3).

2.4 Discussions

Alteration of the methylation state is involved in tumor initiation and progression. DNMT3a is one of the molecules best known to play a crucial role in the regulation of methylation status. It has been suggested that DNMT3a expression in tumor cells is associated with the prognosis of some cancers including gastric and hepatocellular carcinoma [47, 48]. Here, I found that nearly all non-invasive lung adenocarcinomas showed significantly strong expression of DNMT3a. Also, among invasive lung adenocarcinomas, the lepidic subtype showed significantly higher DNMT3a expression than other histological subtypes. Moreover, I identified a significant correlation between DNMT3a expression and prognosis. Patients who showed strong DNMT3a expression had a significantly better outcome. These findings reflected the reports that non-invasive lung adenocarcinoma and the lepidic subtype have a relatively more favorable prognosis than the other subtypes [52, 53]. Thus it is suggested that low DNMT3a expression in lung adenocarcinoma might be associated with poor prognosis.

In the present study, I only evaluated the nuclear immunoreactivity of DNMT3a on the basis of previous reports [44, 47]. However, the cytoplasm in tumor cells also showed DNMT3a expression. Since the function of DNMT3a in the cytoplasm is unclear, further analysis is required to clarify the relationship of cytoplasmic expression with DNMT3a function.

Alteration of tumor DNA methylation status involves two phenomena: hypermethylation and hypomethylation of DNA [33, 54]. In many tumor tissues, widespread global hypomethylation and locus-specific or regional hypermethylation, particularly in tumor suppressor genes, have been simultaneously observed. It is generally said that reduced methylation at repetitive elements, such as short

interspersed elements (SINEs) and long interspersed elements (LINEs), which collectively constitute about 33% of the human genome [55], significantly contributes to global hypomethylation in cancer cells. In addition to global hypomethylation, gene-specific hypomethylation has also been reported for several loci, including MAGEA, TKTL1, BORIS, DDR1, TMSB10, and stratifin (SFN) in lung adenocarcinoma [37, 56-62]. This hypomethylation of certain genes leads to a change in gene expression and alteration of cell proliferation, angiogenesis, and cell adhesion, thus promoting tumor progression [37, 63]. Although further analysis is required to clarify the mechanism in detail, DNMT3a is thought to be involved in this pathway. My data suggest that low expression of DNMT3a in tumor tissue might lead to demethylation of oncogenes and subsequent cancer progression, resulting in poor prognosis. Although previous reports indicating that DNMT3a expression is an indicator of poor prognosis in gastric cancer have cast doubt on this possibility [47], Gao *et al.* have demonstrated that deletion of DNMT3a promotes lung tumor progression in a mouse model [63]. Their data suggested that DNMT3a might act as a tumor suppressor gene, and that suppression of DNMT3a expression might lead to progression of lung adenocarcinoma, in accordance with my data. Furthermore, although DNMT3a mutation is frequently detected in hematological tumors, comprehensive profiling of lung adenocarcinoma has shown that it has no DNMT3a mutation [64]. The loss of DNMT3a expression demonstrated in the present study is considered not to be associated with somatic mutation. Further study will be required in order to clarify the molecular mechanism underlying the functions of DNMT3a in tumors.

In conclusion, my results indicate that DNMT3a expression in lung adenocarcinoma is associated with the histologically non-invasive type and lepidic

subtype, and a favorable prognosis. I also showed that DNMT3a expression is an independent prognostic marker in lung adenocarcinoma. Since lack of DNMT3a is thought to facilitate tumor progression, DNMT3a might be clinically applicable as an indicator of favorable prognosis.

Chapter 3: Identification of differentially methylated genes in lung adenocarcinoma

3.1 Introduction

Research on DNA methylation in human cancer has several clinical importance. It can provide biomarkers to aid conventional diagnosis or prognostic indicators that can predict patient's outcome or efficacy of the anti-cancer therapy, and also we might find a new target molecule for drugs development[65-67].

Based on the result of DNMT3a study, we conclude that DNMT3a is correlated with progression of lung adenocarcinoma [68]. Because of DNMT3a dysfunction, I expected that there would be many oncogenes that turn to demethylated status and thus overexpressed, leading to tumor progression. In support of this idea, previous study of our group has indicated that overexpression of SFN in invasive adenocarcinoma was induced by DNA demethylation.

To fully understand the role of DNA methylation, we need a development of tools that can acquire extensive DNA methylation profiles simultaneously across the whole genome and it must be interchangeable and easy to collect and interpret the results. The gold standard for DNA methylation analysis is whole-genome bisulfite sequencing (WGBS) [69]. WGBS use sodium bisulfite to convert unmethylated cytosine to uracil while the methylated cytosine will be as it is and the procedures then follow with whole-genome sequencing [70]. Even though WGBS has been successfully implemented, the expensive cost and the requirement of highly skilled expertise to generating, processing, collecting, and interpreting WGBS data, make it difficult and not feasible enough for DNA methylation studies in general.

Recently, the Illumina Infinium BeadChips have provided a more accessible method for the exhaustive methylation analysis. The concept is similar to WGBS using bisulfite conversion, but the single base resolution genotyping of targeted CpG sites is using probes on a microarray. The advantages of the Infinium method are their good cost performance, and not only relatively easy to use but also show a similar result from other methods [71]. The newest bead chip from Illumina is called Infinium MethylationEPIC Bead Chip microarray that covers >850.000 CpG sites which include >90% of the old Infinium Human Methylation450 Bead Chip and adds 333.265 CpGs located in enhancer regions identified by the ENCODE [72] and FANTOM5 [73] projects. The results from EPIC bead chip have already been proven to be reliable, has significant improvement high coverage compared to the old 450 bead chip, but still maintaining the good cost performance and the easy to analyze data which is the advantage of Infinium methylation array compare to the other DNA methylation analysis methods [74, 75].

The purpose of the present study was to find another differentially methylated genes that have the same tendency as SFN, which correlated with early stage lung adenocarcinogenesis.

3.2 Materials and Methods

3.2.1 Sample selection

For the Infinium methylation array, totally 8 patients' samples were collected from University of Tsukuba Hospital (Ibaraki, Japan) in 2017. Three of them were AIS, another three were invasive adenocarcinoma, and the rest were normal lung samples collected from surgically resected pulmonary bulla. All six lung adenocarcinoma samples were collected by scratching directly from the cut surface of freshly resected tumor. I checked the samples' papanicolaou smear slide under microscope and confirmed major component of the samples used were tumor cells. Genomic DNA was extracted from all the samples using DNeasy Blood and Tissue kit (QIAGEN, Hilden, Germany).

For pyrosequencing, 21 patients samples were selected from surgically resected specimens at University of Tsukuba Hospital between 2015 and 2017. Ten of the samples were AIS and the rest were invasive adenocarcinoma. I collected the tumor cells from cryosection (10μm thick) using laser microdissection system, LMD-6000 (Leica, Wetzlar, Germany).

For Immunohistochemistry (IHC), we selected 14 samples from the pyrosequencing samples that have the formalin-fixed and paraffin-embedded (FFPE) block. For clinicopathological analysis, we also add another 184 samples of lung adenocarcinoma surgically resected at University of Tsukuba Hospital between 1999 and 2007 and tissue microarray (TMA) blocks were constructed using those samples. Informed consent for this study had been obtained from all of the patients.

3.2.2 Infinium Methylation Array

Infinium methylation array analyses were performed using extracted genomic DNA using Infinium MethylationEPIC Bead Chip micro array (Illumina, San Diego, CA). Genomic DNA (1 μg) was treated with sodium bisulfite using a Zymo EZ DNA methylation kit (Zymo Research, Irvine, CA) and the bisulfite-modified DNA was amplified prior to hybridization array. The array was scanned with an iScan System (Illumina).

To process the raw data obtained from Infinium methylation array, I used Methylation Array Computing Navigation (MACON) web tool. MACON then process the probe filtering, beta mixture quantile (BMIQ) normalization, genomic block analysis, annotation, and finally produced a data subset for further analysis [76].

3.2.3 Pyrosequencing

Bisulfite conversions were done prior to pyrosequencing by using Epitect Bisulfite Kit (QIAGEN).

Primers used for PCR and pyrosequencing were designed using QIAGEN software, PyroMark Assay Design 2.0 (QIAGEN). The sequencing process used CpG assay and performed using PyroMarkQ24 (QIAGEN) with suitable reagents from QIAGEN.

Pyrosequencing primers

SFN-Sequence primer   AGGTGTTAGTGTAGG

SFN-Forward primer   TAGTTTGGAGTTTGGAAAGGTGTTAGTGTA

SFN-Reverse primer   CTACCTCCCTCTCCTCTCT-Biotin

GORASP2-Sequence primer   TTTGGGTTAGAATATTTTATGG

GORASP2-Forward primer   GAGGTAAGTTAATTATTTTGGGTTAGAA

GORASP2-Reverse primer   TCCCTCTAAAACCCCTAATTACT-Biotin

ZYG11A-Sequence primer   TGGTAGATATTTGTTAATATTTTGT

ZYG11A-Forward primer   GGGTTTTTTTTTGGAGTAGGTTAT

ZYG11A-Reverse primer   ATTTCAAAAACTTTACCCCTAAAAATC-Biotin

LOC10099657-Sequence primer GGTGTGGAGTTAATATTTATTT

LOC10099657-Forward primer AGTGGTTAGGTGTGGAGTTAATAT

LOC10099657-Reverse primer AAACACAACACCAACCATTTTATCA-Biotin

CD1D-Sequence primer   AGGTAGATATTAGGGTTAGA

CD1D-Forward primer   GGGGTGTGAGGTGATGTT

CD1D-Reverse primer   CCCTCTCCCTCACTCTTTTATC-Biotin

3.2.4 Immunohistochemistry

For IHC, we used 4-micrometer-thick sections cut from FFPE blocks and FFPE TMA blocks. The sections were deparaffinized and rehydrated. For antigen retrieval,

we autoclaved the sections in 10 mmol/L Citrate buffer (pH 6.0) at 105°C for 15 min. Further steps were performed on an automated stainer, Histostainer 48a (Nichirei Biosciences, Tokyo, Japan). Endogenous peroxidase was blocked with 3% hydrogen peroxide. The slides were incubated in a 1:100 dilution of rabbit polyclonal GORASP2 antibody (Atlas Antibodies, Bromma, Sweden), in a 1:50 dilution of rabbit polyclonal ZYG11A antibody (Atlas Antibodies), and in a 1:800 dilution of mouse monoclonal SFN antibody (Sigma-Aldrich, St. Louis, MO) at RT for 1 hour, and subsequently incubated in the secondary antibody at RT for 1 hour. The signal was detected using DAB (Dako REAL Envision Detection System; Dako, Glostrup, Denmark). Hematoxylin was used for counterstaining. I evaluated all cases without prior knowledge of the clinicopathological data.

The evaluation of the GORASP2, ZYG11A, and SFN staining was according to cytoplasmic staining. The results were evaluated by the H-score system [77]. For survival analysis, the ROC curve was drawn to determine the best cut-off point of the score. (Fig.9, 10) Since the aim of the study is to clarify the correlation between GORASP2 and ZYG11A expression with the prognosis of the patient, I used GORASP2 and ZYG11A expression and the outcome of the patient as variables to draw ROC curve. After drawing the curve we draw diagonal line from bottom right into top left corner and the cut off point is the coordinates where the diagonal line cross over the curve. Here, for the GORASP2 the coordinate falls in 185 (0.746 sensitivity and 0.287 (1-specificity)), and for ZYG11A the coordinate falls in 135 (0.762 sensitivity and 0.241 (1-specificity)) thus it was adopted as the cut off point in this study. Counts below the cut off point were judged as weak expression, and those above as strong expression.

3.2.5 Statistical Analysis

33

For all statistical analyses, SPSS 22 (SPSS, Chicago, IL) and CLC Genomics

Work Bench ver.11 (QIAGEN, Hilden, Germany) were used. For determining the cut-

off point for IHC scoring, the ROC curve method was used. Correlation of

clinicopathological features with GORASP2 and ZYG11A expression was analyzed

by chi-squared test. The Kaplan-Meier method was used for calculation of survival

curves, and log-rank test was performed for comparisons. Multivariate analysis was

done using the Cox proportional hazards model. Differences were considered

statistically significant at $p \leq 0.05$.

3.3 Result

3.3.1 Infinium methylation array and candidate genes selection

Methylation array using Infinium MethylationEPIC Bead Chip micro array (Illumina) analyzed more than 850.000 CpG sites and to cover 99% of the Refseq genes [78]. For further processing we used MACON web tool. We determined the methylation rate in terms of β value [79].

It has long been known that DNA methylation at CpG island in gene promoter region regulates expression of downstream gene. Thus, to narrow down and select the gene candidates we made several conditions for hypomethylated gene screening: the CpG sites located between transcription-starting site (TSS) and 2kb upstream from TSS; more than 10% methylation difference between AIS and invasive adenocarcinoma [80]; highest methylation rate in normal lung and stepwisely decreased to AIS and then invasive adenocarcinoma. Opposite conditions were used for hypermethylated gene screening. From the given conditions we selected 23 CpG sites as hypomethylated genes and 583 CpG sites as hypermethylated genes, which were differentially methylated region (DMR) between AIS and invasive adenocarcinoma. (Fig.11) Although the number of hypomethylated genes were relatively limited compared to those of hypermethylated genes, we considered that the genes showing hyper expression by DNA demethylation might be included in important oncogenes and focused on 23 CpG sites which showed hypomethylation.

For further selection, we performed statistical analysis between AIS vs. invasive adenocarcinoma and selected 3 CpG sites, ZYG-11 family member A (ZYG11A),

LOC10099657, and Mir656 (p<0.05) as genes, which show significantly, lower methylation rate in invasive adenocarcinoma compared to AIS. However, because I planned to validate the results using IHC, I supposed that microRNAs are not an appropriate candidate to investigate, and thus excluded Mir656 from my candidates. On the other hand, we previously conducted cDNA microarray analysis and obtained expression profiles of AIS and invasive adenocarcinoma. Here, I compared methylation profiles with RNA expression profiles to identify genes whose RNA expression is higher in invasive adenocarcinoma than AIS [81]. Consequently, 3 CpG sites, SFN, Golgi reassembly-stacking protein 2 (GORASP2), and Cluster of differentiation 1 (CD1D) were selected. (Table 4)

3.3.2 Pyrosequencing

I implemented pyrosequencing to validate the genes we selected based on the methylation array results [82]. I calculated the mean methylation rate from the entire 5 genes candidate. While I found difference in methylation rate between AIS and invasive adenocarcinoma (Normal>AIS>Invasive) on SFN, GORASP2, and ZYG11A, no difference can be seen on CD1D and LOC10099657 (Fig.12).

3.3.3 Association of methylation rate with protein expression

I carried out IHC to validate and find correlation between methylation status and

protein expression for ZYG11A, GORASP2, and SFN [83]. H-score was calculated and then compared with the methylation status percentage obtained from pyrosequencing.

GORASP2 showed statistically significant negative linear relationship between H-score and the methylation rate (r: -0.878, p<0.001). Mean H-score also showed the lowest in normal lung, followed by AIS then invasive adenocarcinoma (61<162<208) (Fig.13, 14).

ZYG11A also showed statistically significant negative linear relationship between H-score and the methylation rate (r: -0.623, p<0.001). Mean H-score also showed the lowest in normal lung, followed by AIS then invasive adenocarcinoma (33<94<131) (Fig.15, 16).

SFN also showed statistically significant negative linear relationship between H-score and the methylation rate (r: -0.642, p<0.001). Mean H-score also showed the lowest in normal lung, followed by AIS then invasive adenocarcinoma (10<154<184) (Fig.17, 18).

Those results suggest that protein expression of GORASP2, ZYG11A, and SFN are regulated by DNA methylation.

3.3.4 Expression pattern of GORASP2 and ZYG11A in lung adenocarcinoma

Next, to investigate expression pattern and clinicopathological implication of GORASP2 and ZYG11A, I performed IHC using 171 cases of lung adenocarcinoma. For SFN, our group has already reported elsewhere[84].

Among invasive adenocarcinomas, strong expression of GORASP2 was detected in 18/46 (39%) lepidic, 22/26 (84%) papillary, 11/19 (57%) acinar, and 25/32 (78%) solid adenocarcinomas. Among non-invasive lung adenocarcinomas, strong expression was detected in 0/19 (0%) AIS and 2/27 (7%) MIA. (Table 5).

We also assessed the correlation between GORASP2 expression and clinicopathological features of the patients using the chi-squared method. Although GORASP2 expression showed no significant correlation with age, it was significantly correlated with the pathological stage, lymphatic permeation, vascular invasion, and pathological subtype of lung adenocarcinoma (Table 5). GORASP2 showed significantly higher expression in invasive lung adenocarcinoma (76/123 cases, 61%) than in non-invasive lung adenocarcinomas (2/48 cases, 4%).

On the other hand, strong expression of ZYG11A was detected in 17/29 (58%) lepidic, 16/26 (61%) papillary, 13/19 (68%) acinar, and 16/16 (50%) solid adenocarcinomas. Among non-invasive lung adenocarcinomas, strong expression was detected in 0/19 (0%) AIS and 11/18 (61%) MIA. (Table 6).

We also assessed the correlation between ZYG11A expression and clinicopathological features of the patients using the chi-squared method. Although ZYG11A expression showed no significant correlation with age and sex, it was significantly correlated with the pathological stage, lymphatic permeation, vascular invasion, and pathological subtype of lung adenocarcinoma (Table 6). ZYG11A showed significantly higher expression in invasive lung adenocarcinoma (62/123 cases, 50%) than in non-invasive lung adenocarcinomas (11/48 cases, 22%).

3.3.5 Correlation of GORASP2 and ZYG11A expression with postoperative overall survival

The Kaplan-Meier curves indicated that the patients in the strong GORASP2 or ZYG11A expression group had a significantly poorer outcome than those in the weak expression group (Fig.19, 20). Correlation between SFN expression with postoperative overall survival has already been evaluated in our previous study [84].

3.3.6 Multivariate analysis using the Cox proportional hazards model

After adjustment for gender, age, pathological stage, vascular invasion, lymphatic permeation, GORASP2 expression, ZYG11A expression, and SFN expression, patients with strong expression of GORASP2 and ZYG11A showed a significantly higher risk of lung cancer-related death than those with weak expression {(HR: 0.332, 95%CI: 0.179-0.613, P: <0.001) and (HR: 0.286, 95%CI: 0.155-0.527, P: <0.001)}. Multivariate analysis also indicated that lymphatic permeation, vascular invasion, pathological stage, expression of GORASP2 or ZYG11A were independent prognostic factors indicative of poor survival in patients with lung adenocarcinoma (Table 7).

3.4 Discussion

DNA methylation is one of the most important epigenetic alterations that is correlated with tumor initiation and progression. There are two phenomenon related to DNA methylation which are hypermethylation and hypomethylation [33, 54]. Many tumor tissues show global hypomethylation and regional hypermethylation particularly in tumor suppressor genes.

Based on my DNMT3a study, I showed that weak expression of DNMT3a correlated with poor prognosis of the patients, which might indicate that DNA hypomethylation is an important event in lung adenocarcinogenesis. However, here I selected 583 CpG sites, which showed hypermethylation in invasive adenocarcinoma relative to AIS and normal lung, but only 23 CpG sites showed hypomethylation in invasive adenocarcinoma. This can be explained because in this study I mainly focused on the limited areas such as promoter CpG islands and shore regions, which known to have hypermethylated tendency compare to the other areas or repetitive elements (LINE, ALU) that have hypomethylated tendency. This suggested that DNA demethylation at promoter regions might be a relatively rare event compared to DNA hypermethylation.

Since we previously demonstrated that SFN promoter region showed DNA hypomethylation in invasive adenocarcinoma leading to its overexpression, this fact motivated me to search another gene, which show hypomethylation at their promoter regions. One of the advantages to find genes overexpressed in tumor is that they have possibility to serve novel prognostic indicators or new target molecule for drug development.

Here, I selected 2 genes GORASP2 and ZYG11A, which show hypomethylation and overexpression in invasive adenocarcinoma, implying to have important functions in tumor cells.

I found that expression of GORASP2, ZYG11A, and SFN has a statistically significant negative linear relationship with the methylation status. Moreover, I found that invasive adenocarcinoma showed significantly higher expression of both genes than in AIS. These result suggested that these two genes has the same tendency as SFN, whose expression were conversely correlated with methylation rate and were strongly expressed in invasive adenocarcinoma compared to AIS. Moreover, I found that strong expression of both genes is associated with patient's poor prognosis. Finally, I revealed that both genes are independent prognostic factors for lung adenocarcinoma.

Golgi apparatus is an organelle, whose main function is glycosylation, packaging and transport of the synthesized protein from endoplasmic reticulum. Golgi then distributes the secretory vesicle to the plasma membrane, lysosome, or extracellular space thus its importance in the transportation and secretory pathway [85]. GORASP2 gene encodes membrane protein golgi reassembly-stacking protein of 55 kda (GRASP55) also known as GORASP2 protein whose function is to stack and link the golgi cisternae in the formation and membrane dynamics of golgi apparatus [86-88]. Recent study has shown the importance of golgi apparatus for a solution to overcome acquired resistance in lung adenocarcinoma toward EGFR tyrosine kinase inhibitors. Disrupting the golgi apparatus via inhibition of ADP ribosylation factor 1 (ARF-1) showed anti-tumor activity toward EGFR-activated tumor cells [89]. Similarly, inhibition of GORASP2 might disrupt the membrane dynamics of golgi apparatus and substantially lead to anti-tumor effect.

Study about ZYG11 gene is still very limited in human, but in *Caernohabditis elegans* the human homologue ZYG11 gene is important for cell division and embryonic development [90]. Further study shows that ZYG11 homologue expresses in human sperm cell and has function in cell division during meiosis [91]. In human, ZYG11 gene family has three homologues, which are ZYG11A, ZYG11B, and ZER-1 [92]. Recently, one study has shown oncogenic function of ZYG11A genes in non-small cell lung cancer (NSCLC). Their data indicated that ZYG11A showed overexpression in NSCLC and is correlated with poor prognosis, in accordance with my data. Further, it also stated that depletion of ZYG11A suppress the Cyclin E1 (CCNE1) and thus suppress cell cycle progression [93].

Although further analysis are required to understand and elucidate the function of both genes in the lung adenocarcinogenesis, my data suggest that these two genes are possibly new prognostic indicators and might have potential for new target molecule for lung adenocarcinoma.

In conclusion, my results indicate that GORASP2 and ZYG11A are correlated with methylation and show strong expression in lung adenocarcinoma is associated with the histologically invasive type, and a poor prognosis. I also showed that GORASP2 and ZYG11A expression is an independent prognostic marker in lung adenocarcinoma. Since overexpression of both genes are thought to facilitate tumor progression, GORASP2 and ZYG11A might be clinically applicable as an indicator of prognosis and potential target molecule for drug development.

Chapter 4: Perspectives

Even though DNA demethylation triggered overexpression of particular oncogene is relatively rare event in tumor cells, we previously found that overexpression of SFN was induced by DNA demethylation in invasive adenocarcinoma whereas AIS or normal lung showed 100% methylation in SFN promoter region.[37] Recently our group reported that SFN stabilizes receptor tyrosine kinases including EGFR or MET by interacting with ubiquitin-specific protease 8 (USP8), resulting in up-regulation of cell proliferation and inhibition of apoptosis [84]. These facts indicate that there might be several potent oncogenes whose expression is epigenetically elevated by DNA demethylation. Consistent with this, here I newly selected two oncogenes, GORASP2 and ZYG11A, which showed hypomethylation and overexpression in invasive adenocarcinoma similar to SFN.

Although we need further investigation to elucidate their function or molecular mechanism, they might have oncogenic function to facilitate tumor initiation or progression of lung adenocarcinoma, thus suggesting the possibility to be therapeutic targets for such tumors.

Additionally, the mechanism responsible for abnormal demethylation of SFN, GORASP2, or ZYG11A at the very early stage of lung adenocarcinogenesis has not been elucidated. This demethylation in lung adenocarcinoma is considered to be an appropriate model for investigating the mechanism underlying epigenetic modification in cancer. One of the hypothesis is age-related change in DNA methylation, which have been reported before and may be responsible for change in gene expression. In human samples, age-related changes in DNA methylation have been detected in blood sample [94], brain [95], saliva [96] and many other tissue

[97]. Since the mechanism underlying age-related change in DNA methylation is believed to be tissue and cell-specific [98], it might be worth studying more in lung tissue.

Chapter 5: Conclusion

Although research on hypermethylation in tumor suppressor genes has been extensively conducted worldwide, hypomethylation in particular oncogenes are less studied. In the present study, I demonstrated that overexpression of GORASP2 and ZYG11A is conversely associated with methylation rate at their promoter region, and significantly correlated with tumor malignancy and patient's poorer outcome in lung adenocarcinoma. Since both genes showed strong expression in invasive adenocarcinoma compared to AIS, they might be novel prognostic markers or potent therapeutic targets for this tumor.
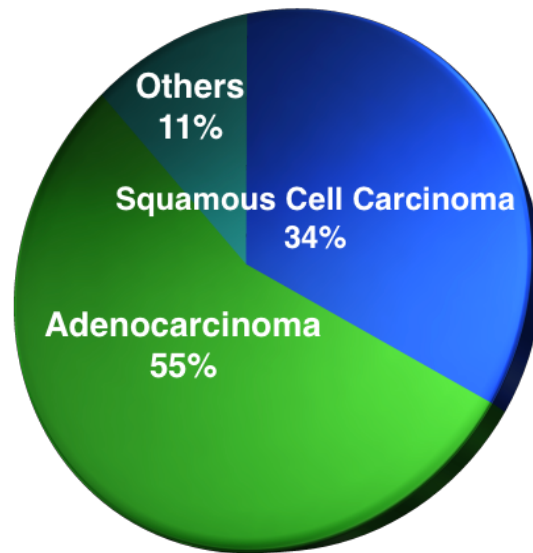
Fig. 1. Histological subtypes of lung cancer. Figure modified from Li T et. al. J Clin Oncol 2013;31: 1039-1049
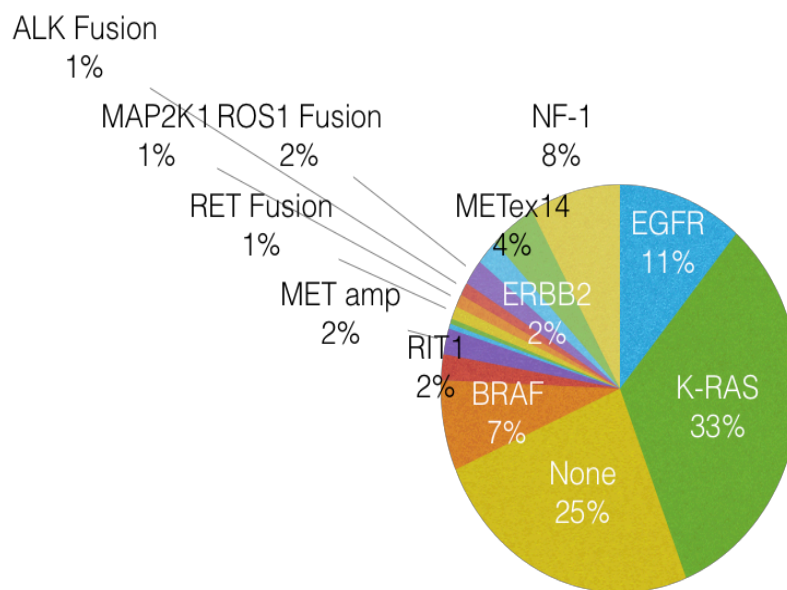
Fig. 2. Common driver oncogenes aberrations in lung adenocarcinoma.
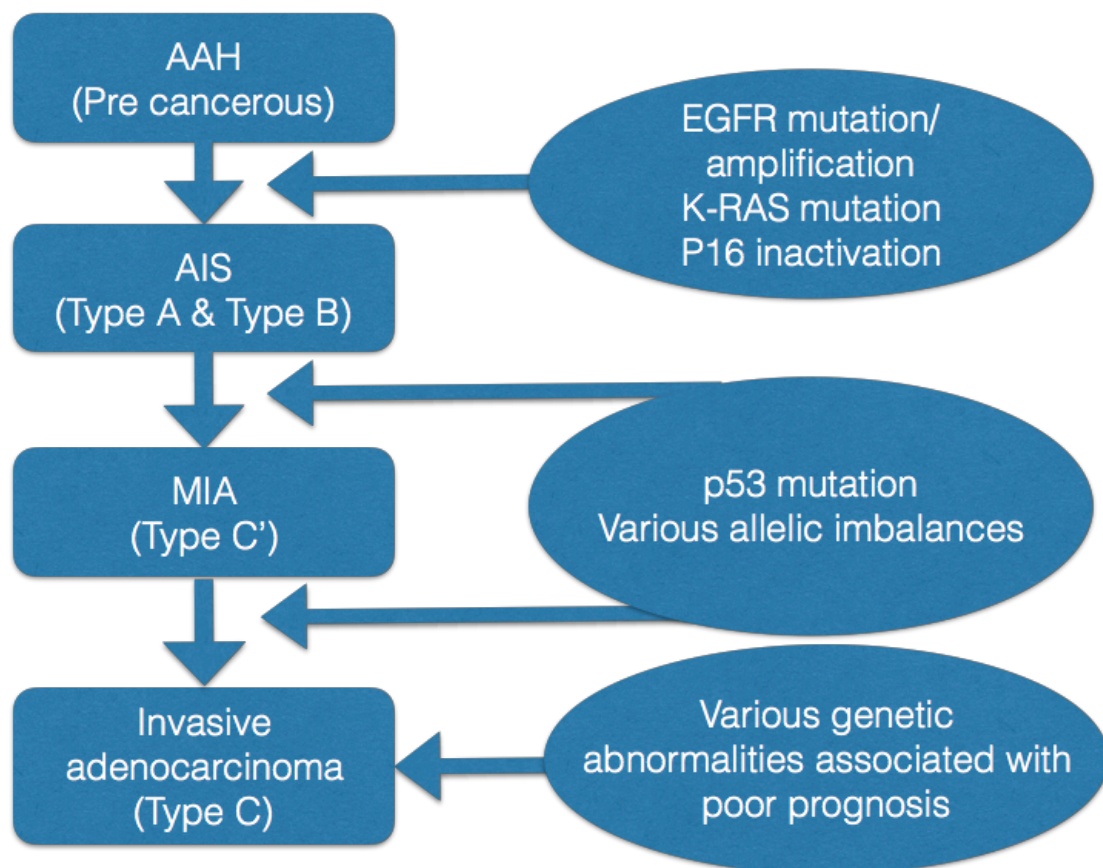Figure modified from the Cancer Genome Atlas Research N. Nature 2014;511: 543.

Fig. 3. Multi step progression of pulmonary adenocarcinoma with histology and molecular alterations. Lung adenocarcinogenesis shows a multi step progression starting from pre-cancerous lesion adenomatous atypical hyperplasia (AAH) to AIS, MIA, and finally invasive lung adenocarcinoma, which are closely related to Noguchi classification subcategorized into type A to C. Not only histologically but also related to genetic aberrations in each step. Figure modified from Nakanishi H et. al. ,Cancer Res 2009;69: 1615-1623
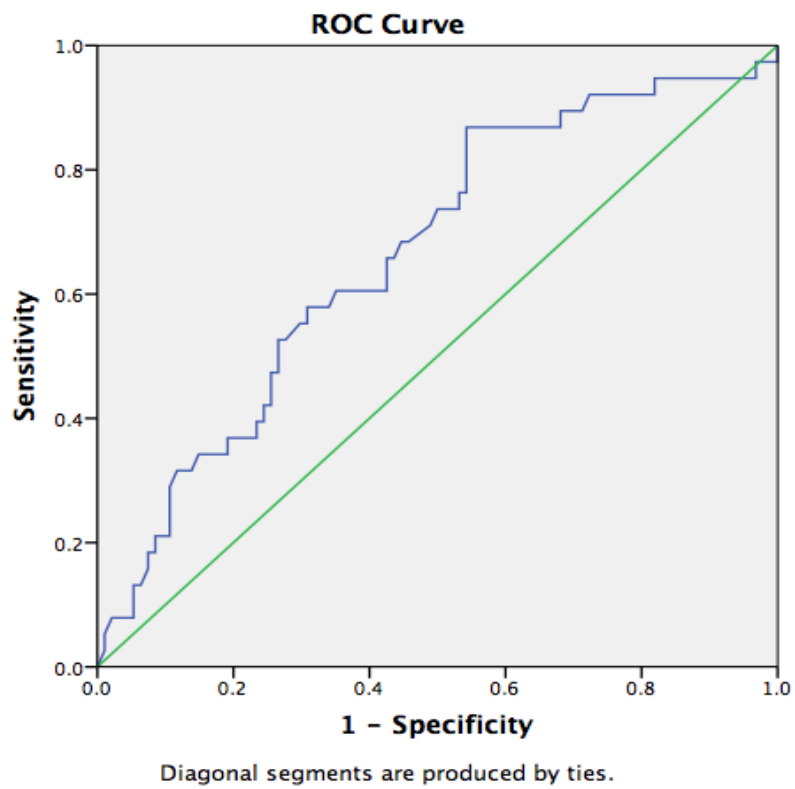
Fig.4 DNMT3a ROC Curve. ROC curve to determine IHC score cut-off point. 57.5 was adopted as the cut-off point for the IHC score.

Table.1 DNMT3a specific siRNA

siDNMT3a	Forward sequence (5' to 3')
	Reverse sequence (5' to 3')
siDNMT3a-1	CCACGUACAACAAGCAGCCCAUGUA
	UACAUGGGCUGCUUGUUGUACGUGG
siDNMT3a-2	CAUCCGGGUGCUGUCUCUCUUUGAU
	AUCAAAGAGAGACAGCACCCGGAUG
siDNMT3a-3	GAGGACAUCUUAUGGUGCACUGAAA
	UUUCAGUGCACCAUAAGAUGUCCUC

A

B

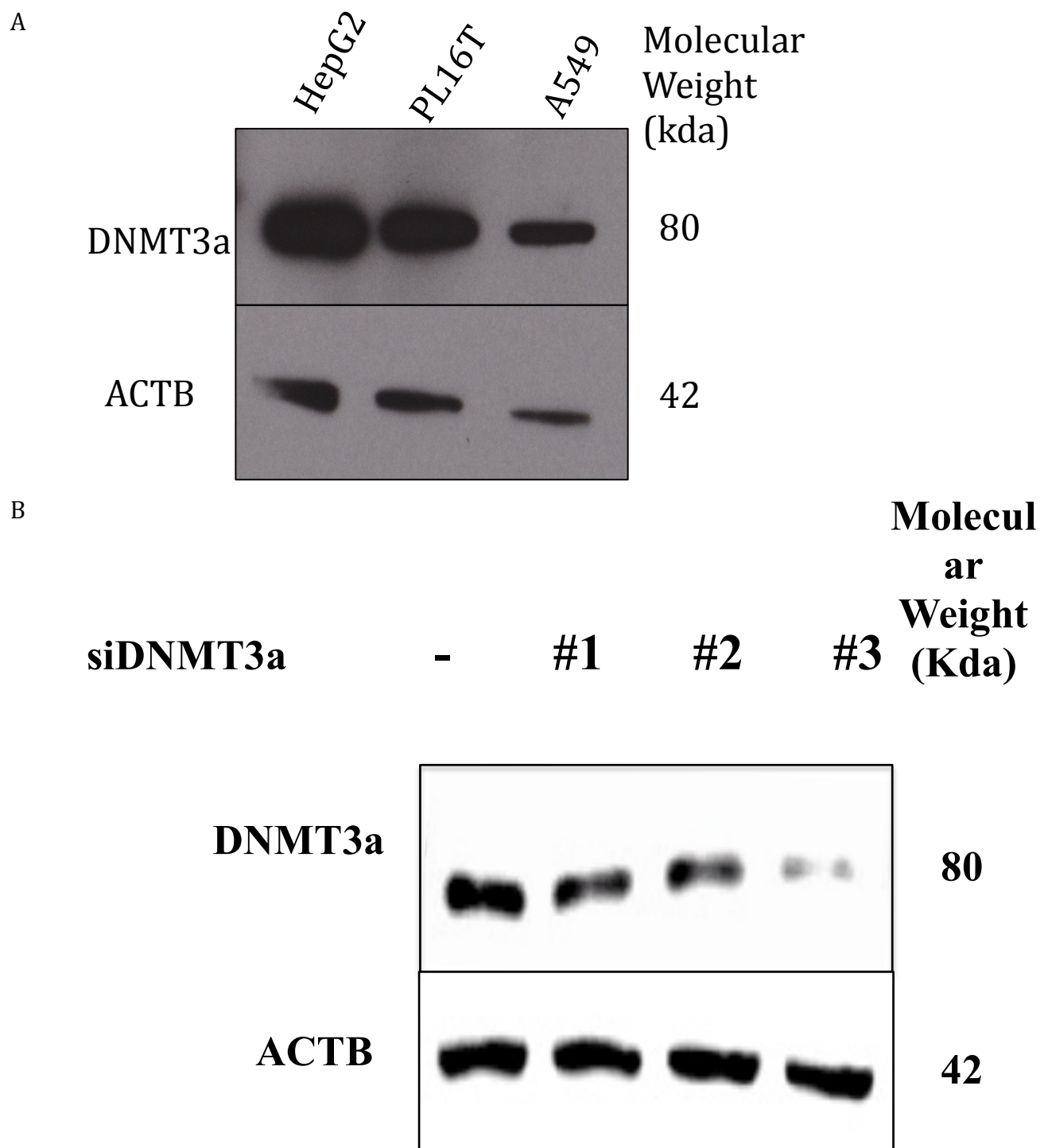siDNMT3a   -   #1   #2   #3   Molecular Weight (Kda)

Fig.5. DNMT3a western blot result. (A) Western blot of DNMT3a in lung. HepG2 was used as a positive control. A549 and PL16T cell lines showed a single band specific for DNMT3a (80 kDa). **(B)** Western blot of DNMT3a using A549 transfected with siDNMT3a. Scrambled siRNA was used as a negative control. NT indicates A549 not treated with siRNA.
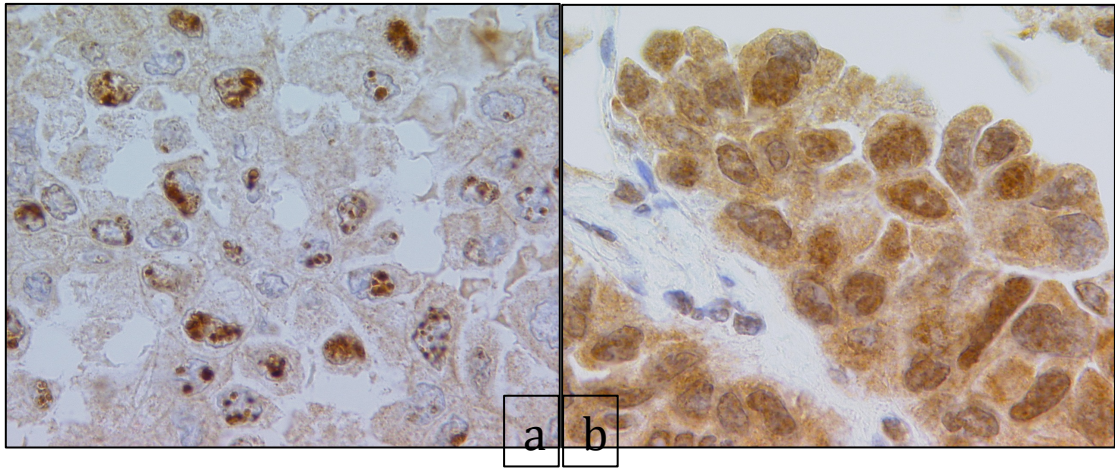
Fig.6 Immunohistochemistry of DNMT3a in lung adenocarcinoma cells.
DNMT3a was stained not only in the nucleus of the tumor cells (a) but also in both the cytoplasm and nucleus (b).

Table.2 DNMT3a and Clinicohistopathological features. Stage I includes IA and IB, stage II includes IIA and IIB, stage III includes IIIA and IIIB. Correlation between DNMT3a expression and clinicopathological features was analyzed using chi-squared test.

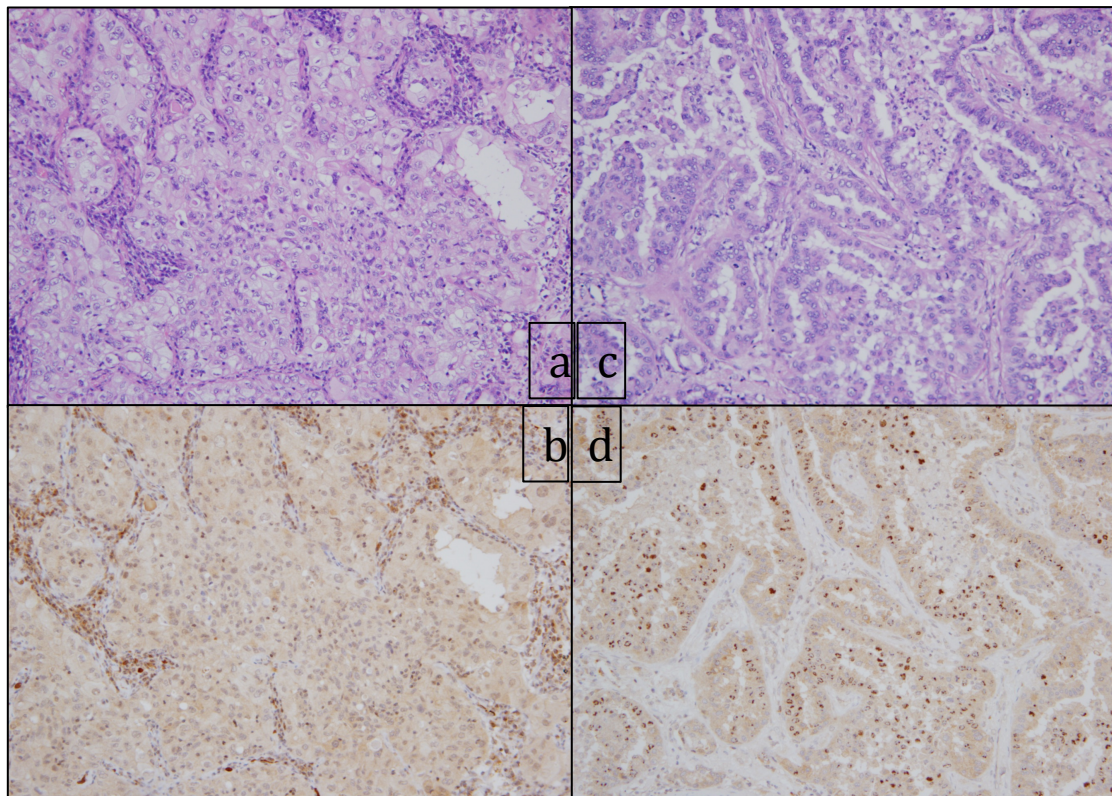| Clinicopathological Features	Total Patients	DNMT3a Expression		P value
		Weak	Strong	
Total Patients	135	56	79	
Age (years)				0.323
≤60	35	17	18	
>60	100	39	61	
Sex				0.097
Male	71	32	39	
Female	64	22	42	
Pathological Stage				0.001*
Stage I	85	25	60	
Stage II	21	11	10	
Stage III	27	18	9	
Stage IV	2	2	0	
Lymph Node Status				0.001*
N0/Nx	95	31	64	
N1 and N2	40	25	15	
Pleural Invasion				<0.001*
pI0	84	25	59	
pI1	27	18	9	
pI2	15	5	10	
pI3	9	8	1	
Vascular Invasion				<0.001*
-	67	12	55	
+	68	44	24	
Lymphatic Permeation				0.051
-	83	29	54	
+	52	27	25	
Pathological Subtype				<0.001*
AIS	14	1	13	
MIA	2	0	2	
Lepidic	53	12	41	
Acinar	31	18	13	
Papillary	20	12	8	
Solid	15	13	2	

Fig.7 Immunohistochemistry of DNMT3a in lung adenocarcinoma. (a) and (c) HE staining of the solid and lepidic subtype, (b) negative expression of DNMT3a, (d) strong expression of DNMT3a.
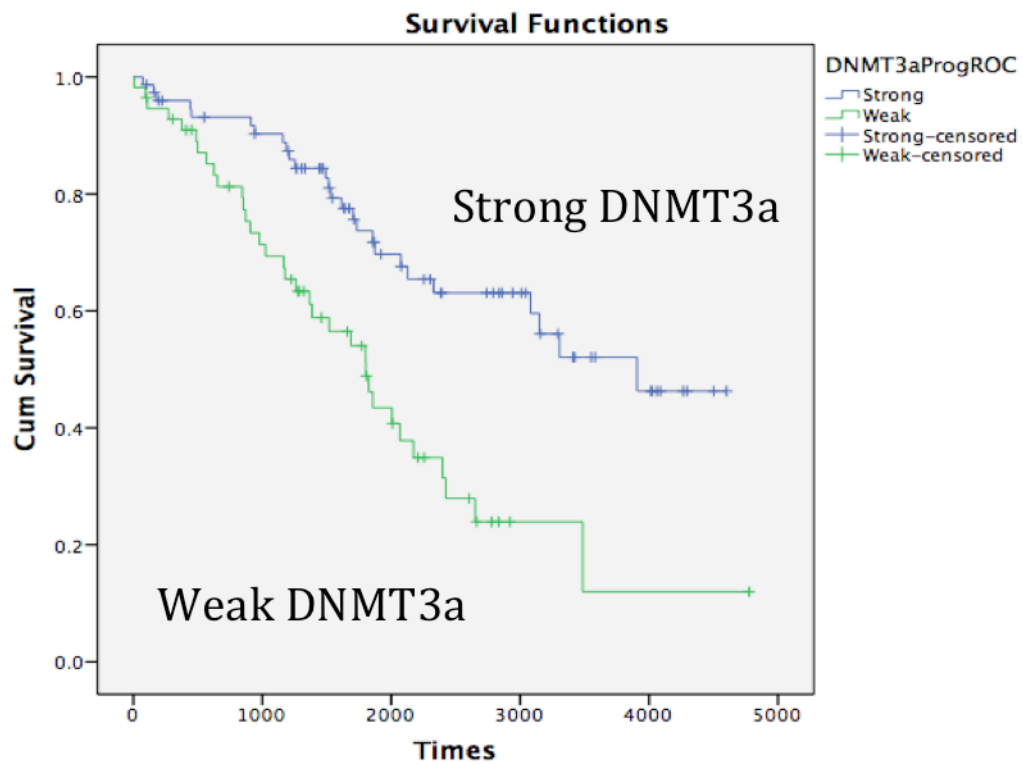
Fig.8 DNMT3a KM curve. Disease-free survival depicted as Kaplan-Meier curves shows the correlation between DNMT3a expression and outcome. Weak expression of DNMT3a was associated with poor prognosis relative to strong expression (*p<0.001*)

Table.3 DNMT3a multivariate analysis. Adjusted for gender, age, DNMT3a expression, pleural invasion, vascular invasion, lymphatic permeation, and pathological stage.

Clin.Features	Univariate Analysis			Multivariate Analysis		
	HR	95% CI	P value	HR	95% CI	Pvalue
Gender	0.91	0.541-1.518	0.708			
Age	1.025	0.998-1.052	0.069			
Pleural Invasion (0,1 vs other)	1.67	1.200-2.151	0.001	1.46	1.062-1.993	0.02
Vascular Invasion (- vs +)	0.512	0.382-0.687	<0.001	0.77	0.543-1.093	0.14
Lymphatic Permeation (- vs +)	0.53	0.401-0.688	<0.001	0.6	0.446-0.812	0.001
DNMT3a Exp (Weak vs Strg)	0.61	0.472-0.797	<0.001	0.72	0.529-0.974	0.033
Pathological Stg (1 vs other)	0.65	0.501-0.843	0.001	1.03	0.763-1.395	0.84

ROC Curve

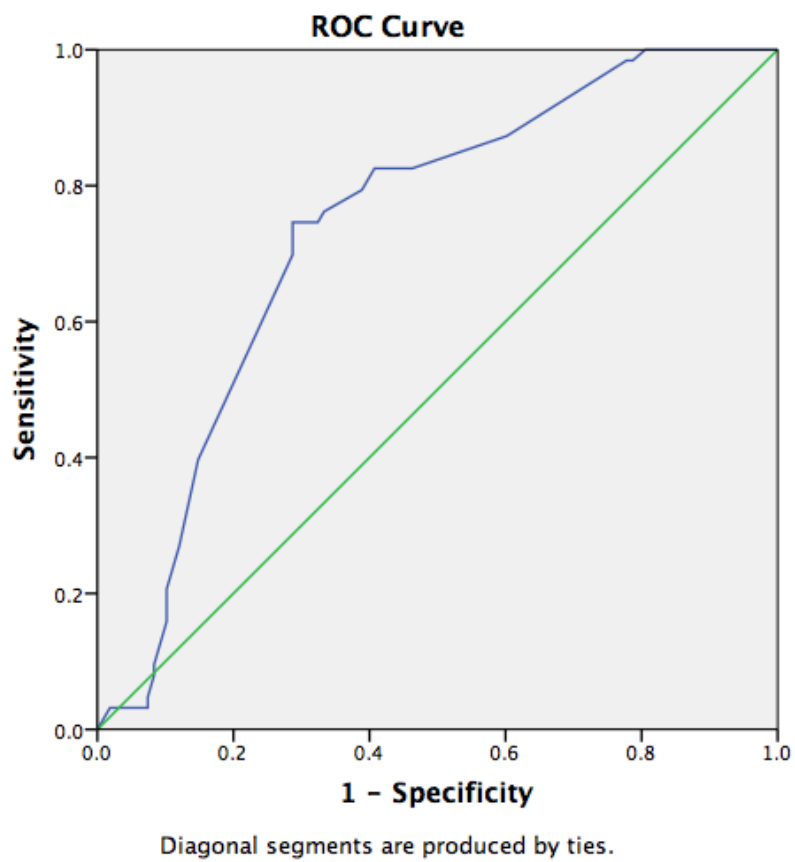Diagonal segments are produced by ties.

Fig.9 GORASP2 ROC curve. ROC curve to determine IHC score cut-off point. 185 was adopted as the cut-off point for the IHC score.
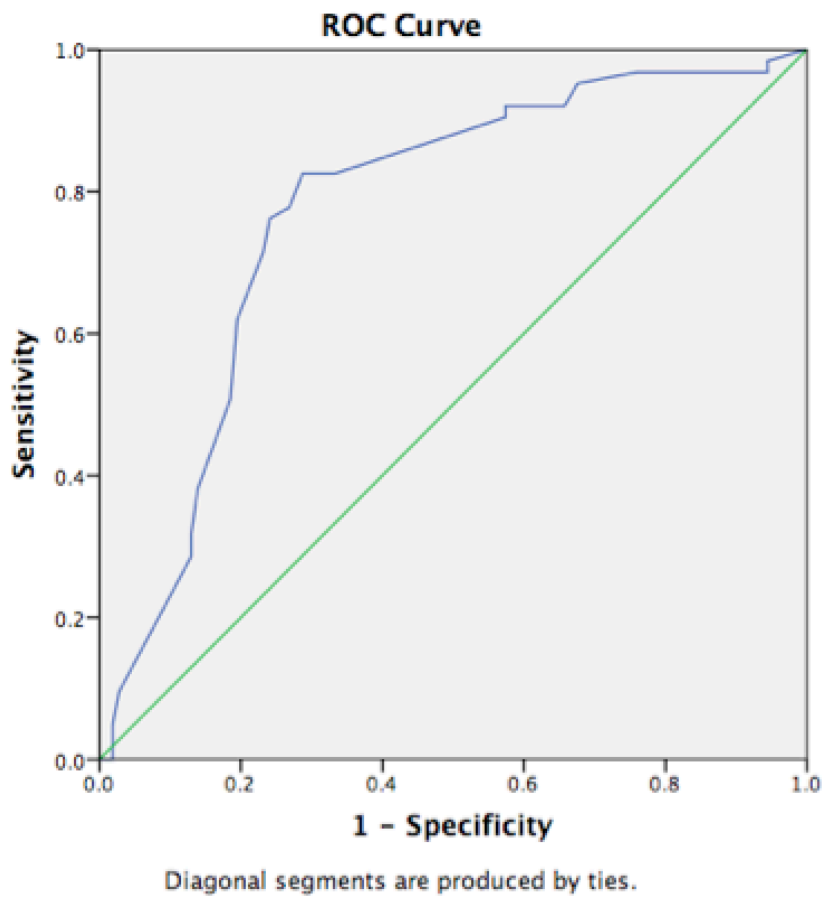
Fig.10 ZYG11A ROC curve. ROC curve to determine IHC score cut-off point. 135 was adopted as the cut-off point for the IHC score.
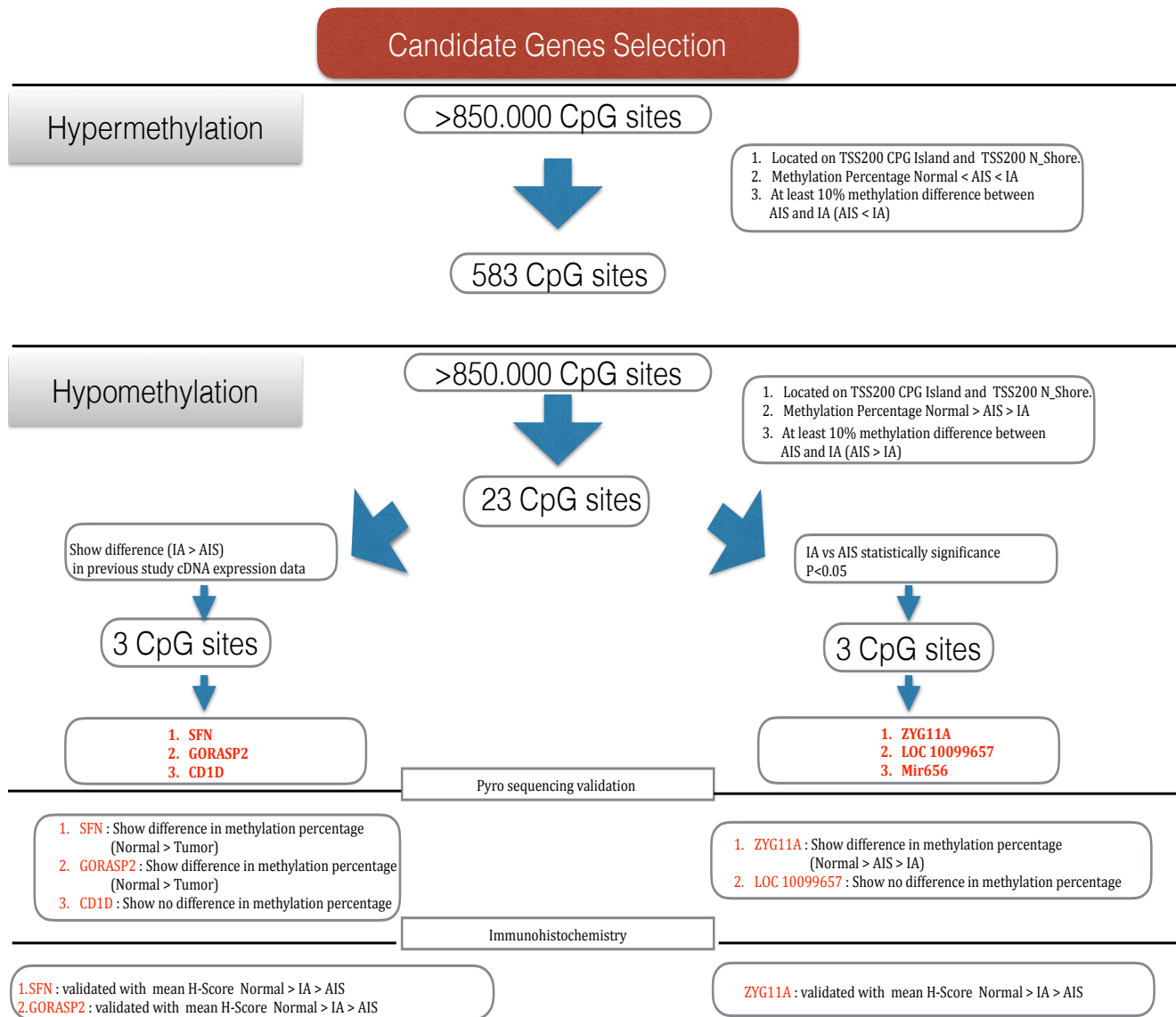
Fig.11 Candidate Genes selection. Adenocarcinoma *in situ* (AIS) and Invasive Adenocarcinoma (IA)**.** Several conditions were used to narrow down from >850.000 CpG sites, which are: the CpG sites located at the transcription-starting site (TSS) 200 island and N_shore; at least 10% methylation difference between non-invasive lung adenocarcinoma and invasive lung adenocarcinoma; highest methylation percentage in normal then non-invasive lung adenocarcinoma and then invasive lung adenocarcinoma for hypomethylation group and vice versa for hypermethylation group. From the given conditions we narrowed down to 23 CpG sites for hypomethylation group and 583 CpG sites for hypermethylation group, which are differentially methylated region (DMR) between groups.

Table.4 Six candidate genes array result. Beta value represents the methylation percentage in the scale of 0-1. All 6 genes show at least 10% methylation difference between AIS and invasive adenocarcinoma.

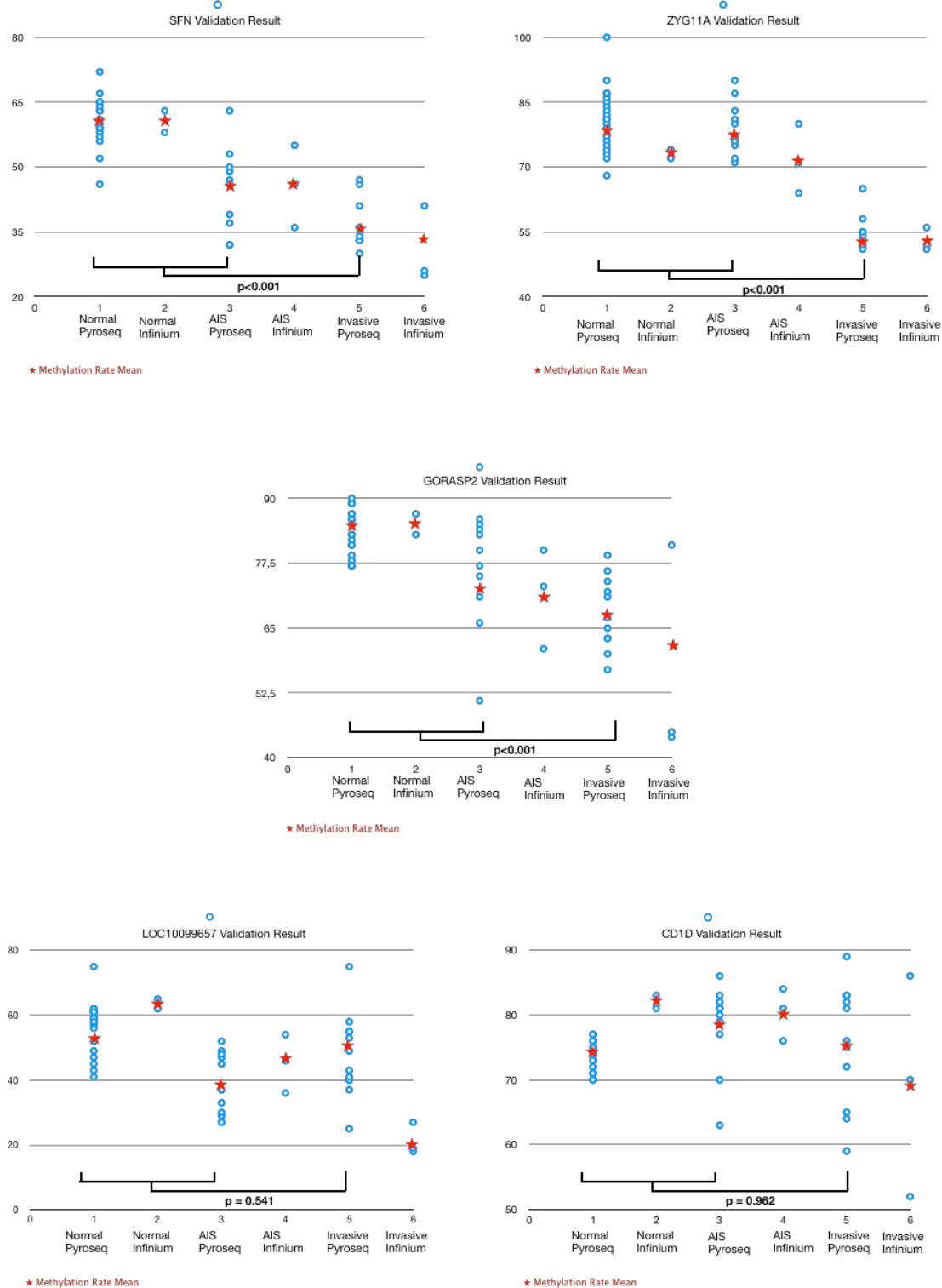| Gene Name	AIS			Invasive			Normal		Mean AIS	Mean Invasive	Meth Diff
	1	2	3	1	2	3	1	2			
ZYG11A	0.8	0.71	0.64	0.52	0.51	0.56	0.72	0.74	0.72	0.53	0.19
SFN	0.36	0.55	0.46	0.25	0.27	0.41	0.59	0.63	0.46	0.31	0.15
CD1D	0.84	0.81	0.76	0.86	0.71	0.52	0.81	0.84	0.8	0.70	0.11
LOC10099657	0.36	0.54	0.46	0.18	0.27	0.19	0.63	0.66	0.45	0.21	0.24
GORASP2	0.8	0.77	0.61	0.82	0.44	0.44	0.87	0.83	0.73	0.57	0.16
MIR656	0.76	0.74	0.75	0.71	0.62	0.63	0.77	0.8	0.75	0.65	0.1

Beta Value

Fig.12 Pyrosequencing validation graph. DNA methylation rates of SFN, GORASP2, ZYG11A, LOC10099657, and CD1D were validated using pyrosequencing with another 11 frozen samples of invasive adenocarcinoma, 10 samples of AIS, and 21 samples of normal lung tissue. While LOC10099657 (p=0.541) and CD1D (p=0.962) showed no significant difference of methylation rate among invasive adenocarcinoma, AIS, and normal lung, the other 3 genes showed significantly lower methylation rate in invasive adecocarcinoma relative to AIS or normal lung (p<0.001).
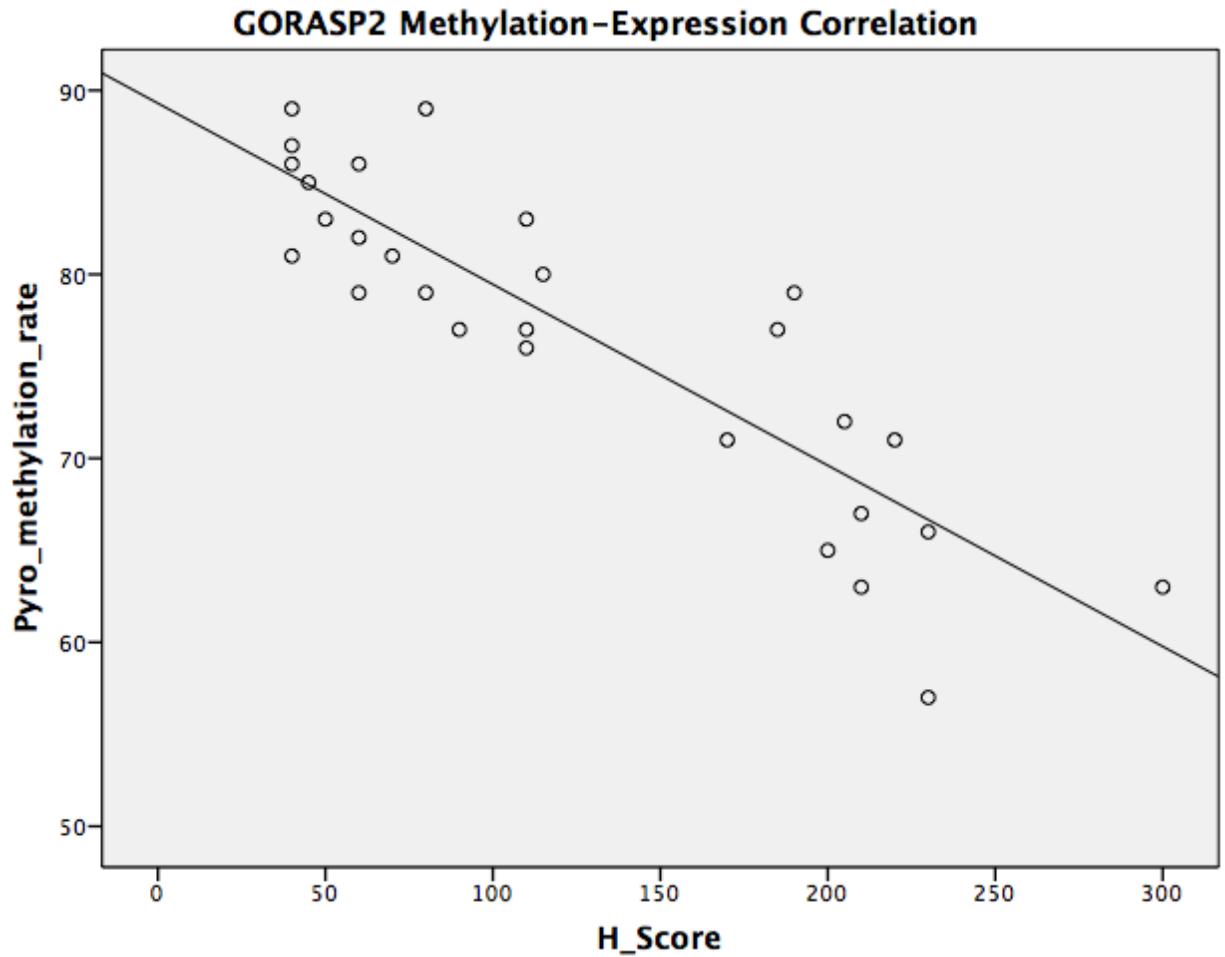
60

Fig.13 IHC pearson correlation result GORASP2. GORASP2 showed statistically significant negative linear relationship between H-score and the methylation rate (r: -0.878, p<0.001). Mean H-score also showed low score in the normal samples, followed by AIS then invasive lung adenocarcinoma samples (61<162<208)
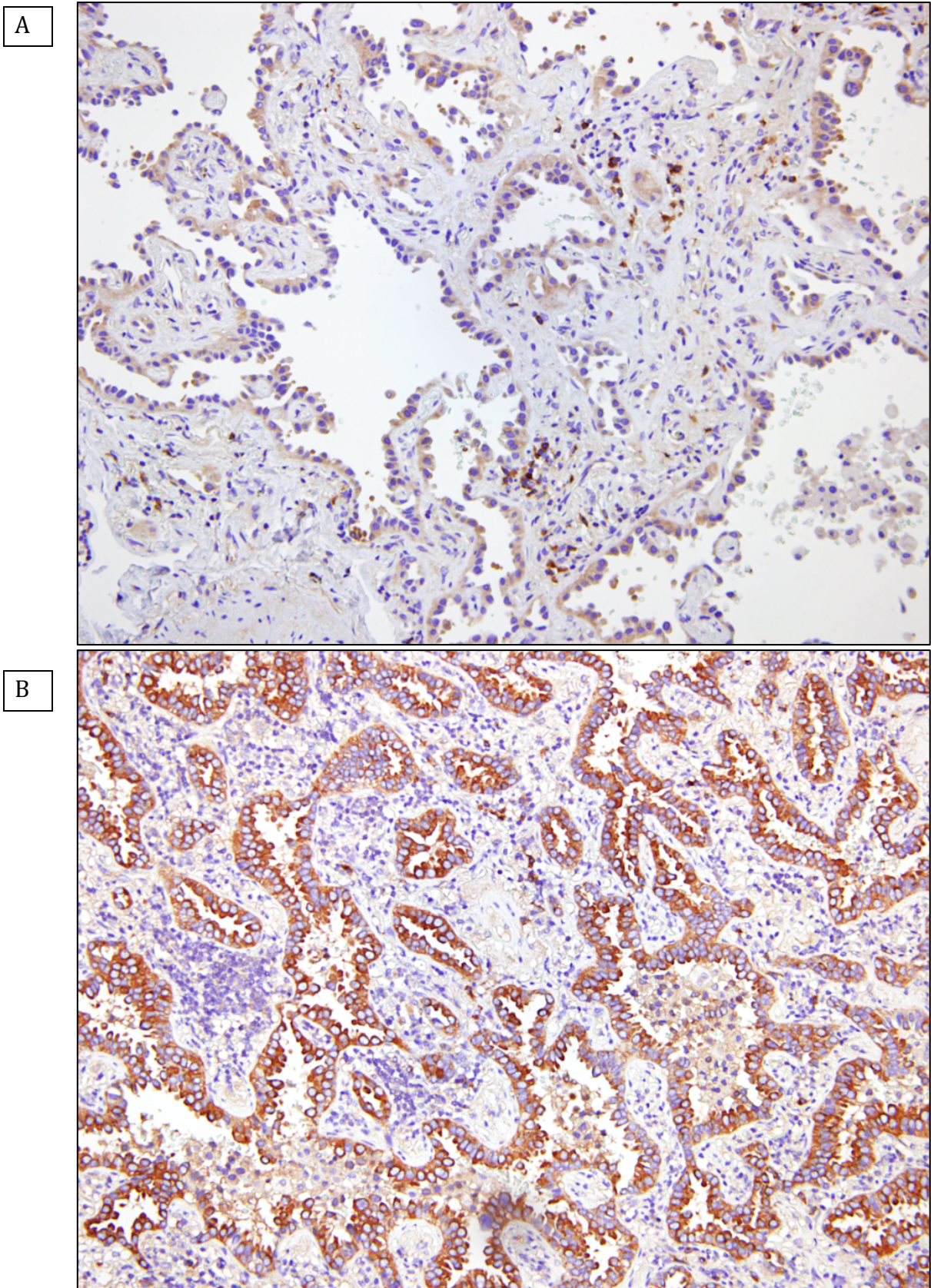
Fig.14 GORASP2 staining pattern. Positive in Cytoplasm. (A) Type-A (H-score 115) (80% methylated); (B) Type-C (H-score 300) (66% methylated).
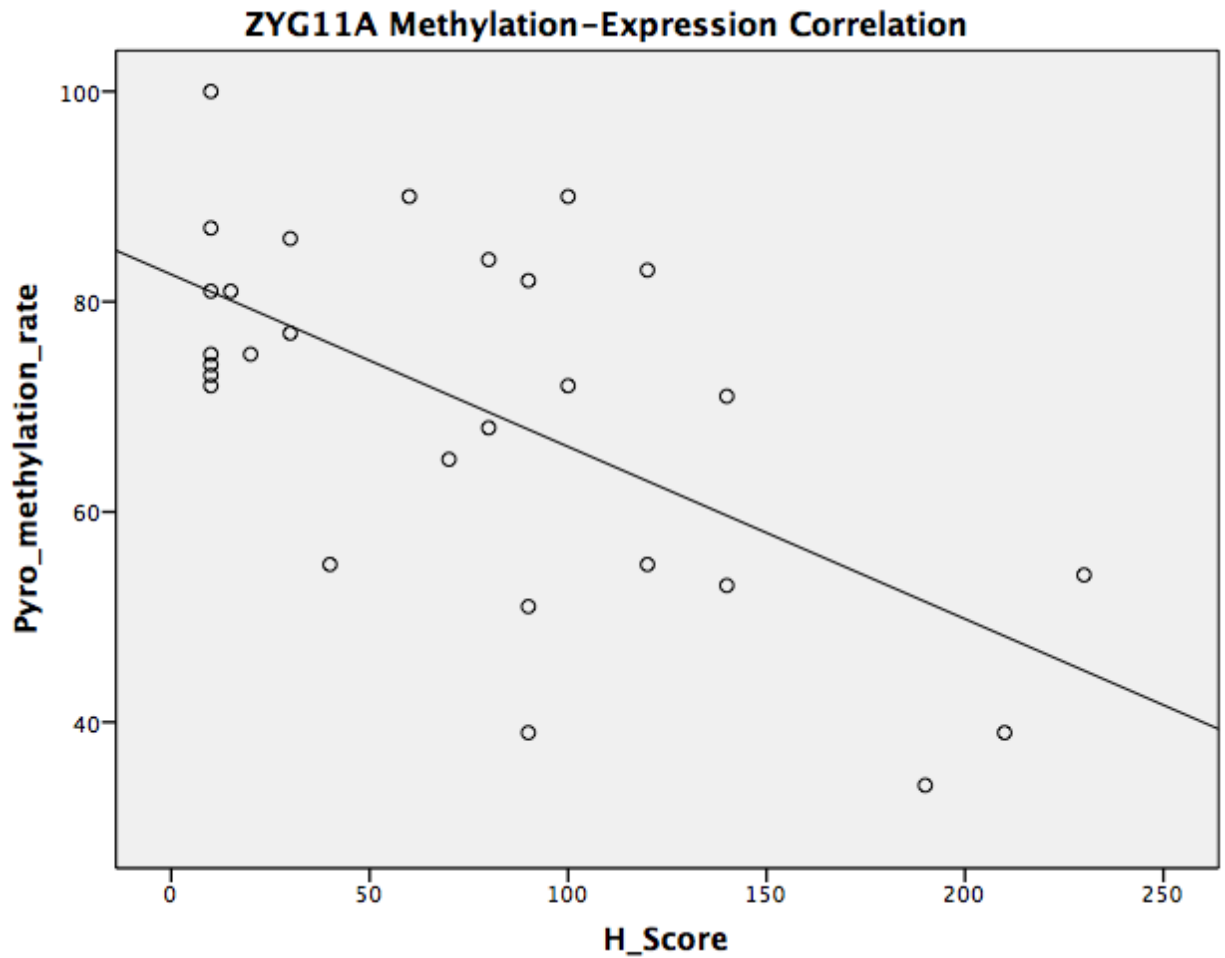
Fig.15 IHC pearson correlation result ZYG11A. ZYG11A showed statistically significant negative linear relationship between H-score and the methylation rate (r: -0.623, p<0.001). Mean H-score also showed low score in the normal samples, followed by AIS then invasive lung adenocarcinoma samples (33<94<131)

Fig.16 ZYG11A staining pattern. Positive in Cytoplasm. (A) Type-A (H-score 100) (85% methylated); (B) Type-C (H-score 230) (66% methylated).
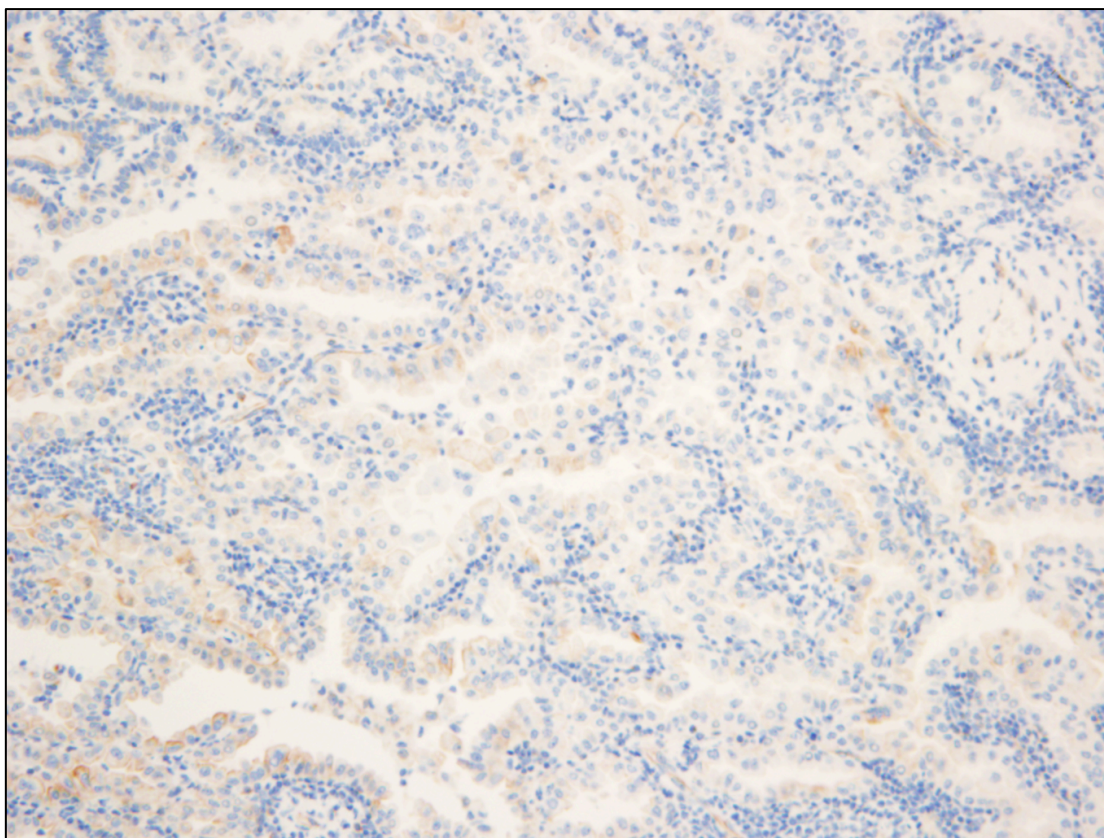
Fig.17 IHC pearson correlation result SFN. SFN showed statistically significant negative linear relationship between H-score and the methylation rate (r: -0.793, p<0.001). Mean H-score also showed low score in the normal samples, followed by AIS then invasive lung adenocarcinoma samples (10<154<184)

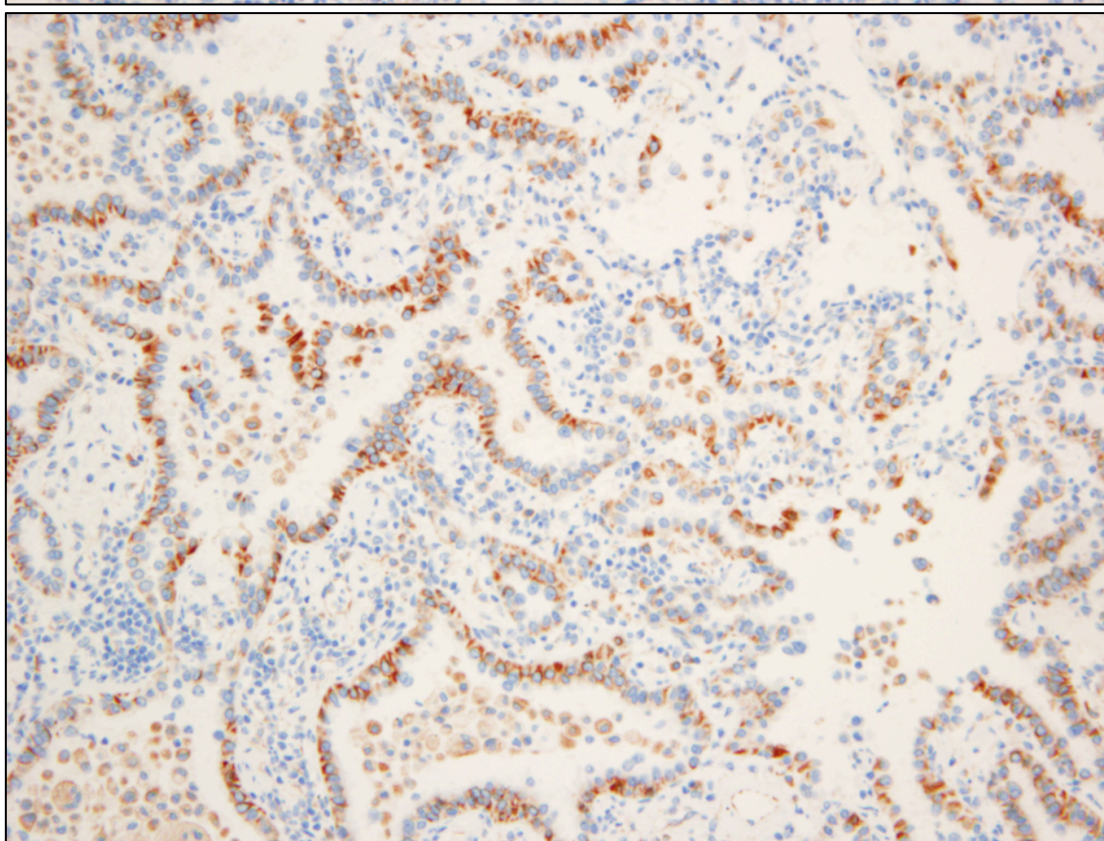**Fig.18 SFN staining pattern.** Positive in Cytoplasm. (A) Type-A (H-score 130) (50% methylated); (B) Type-C (H-score 220) (33% methylated).

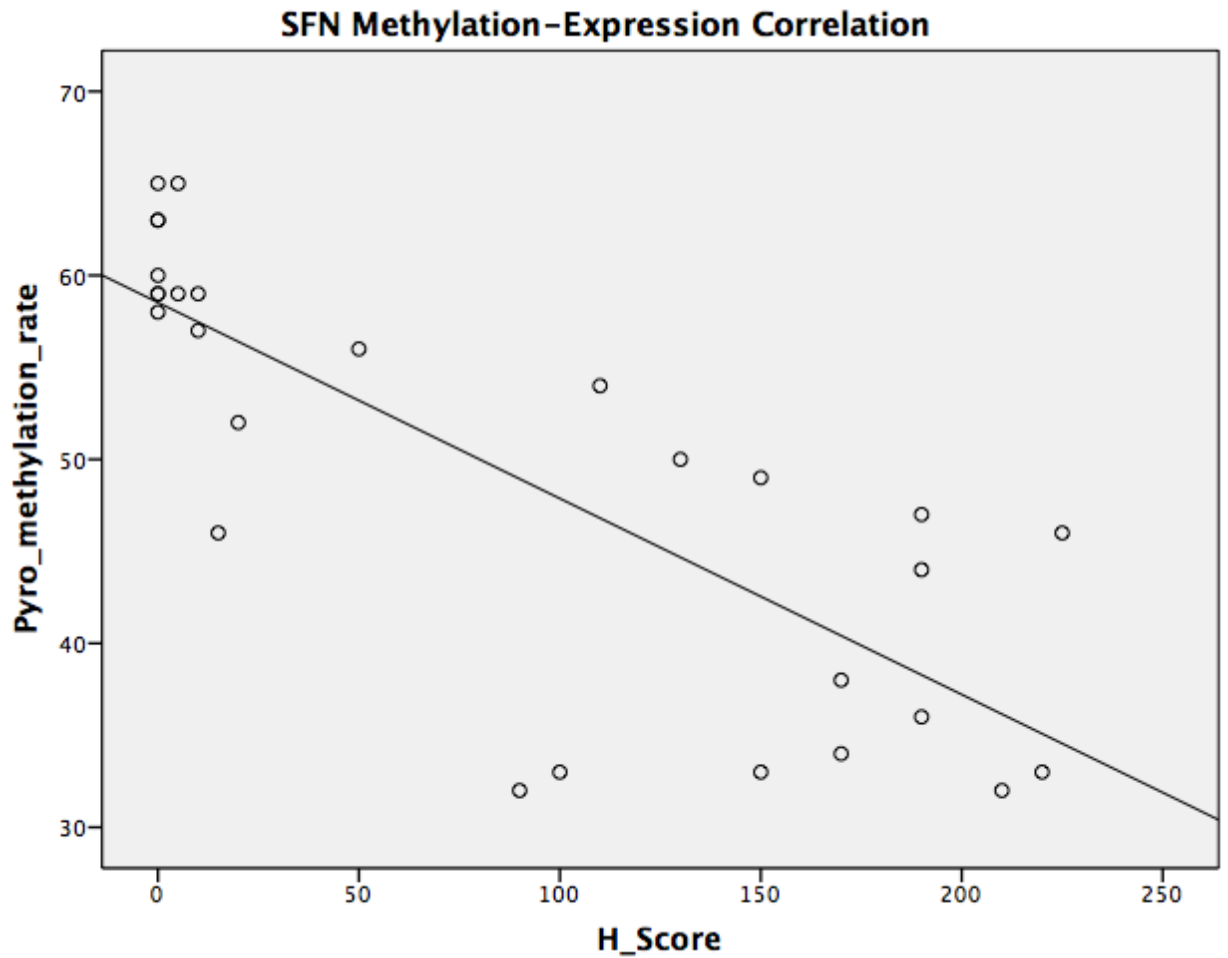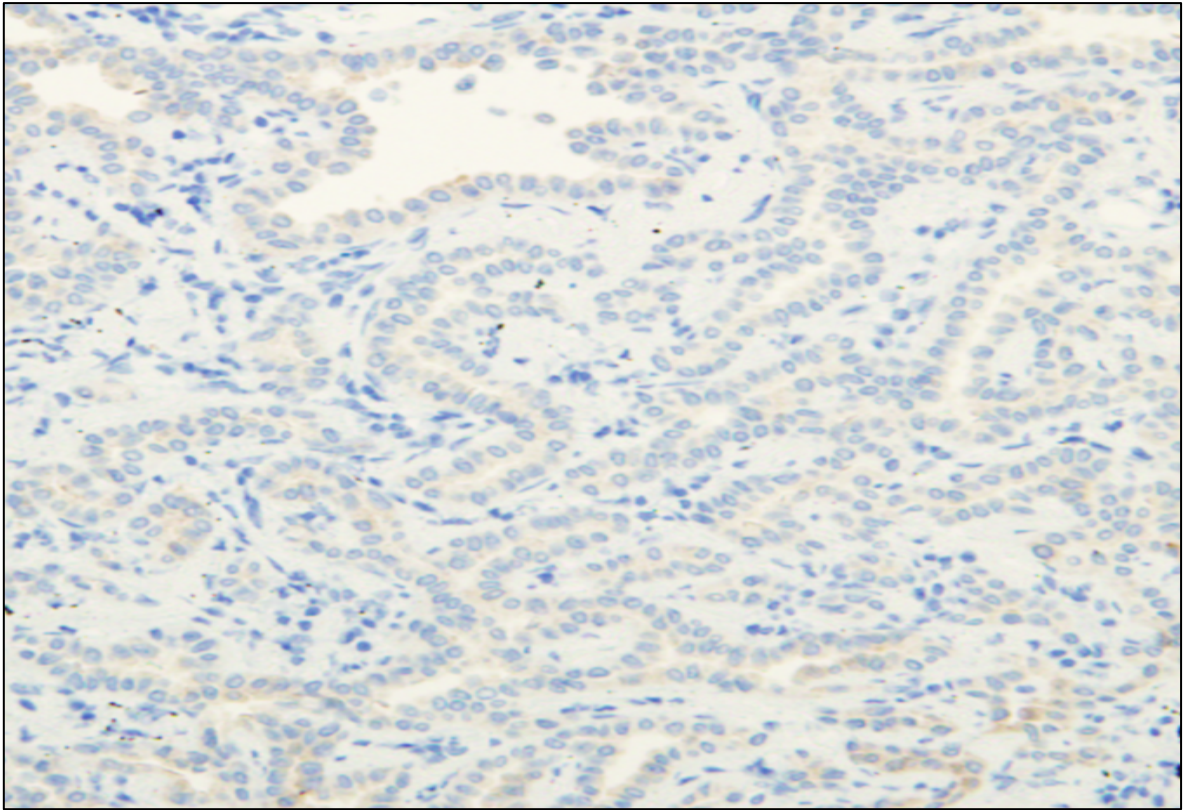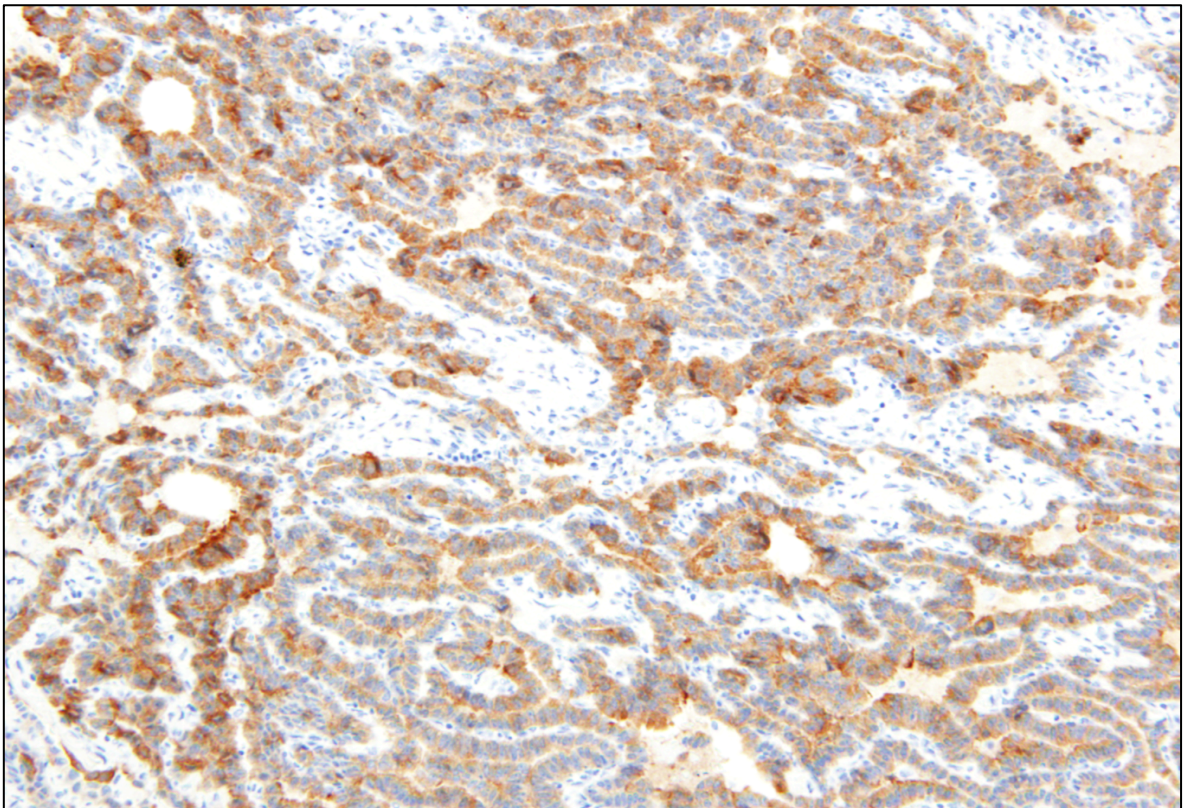**Table.5 GORASP2 and clinicohistopathological features.** Stage I includes IA and IB, stage II includes IIA and IIB, stage III includes IIIA and IIIB. Correlation between GORASP2 expression and clinicopathological features was analyzed using chi-squared test.

Clinicopathological Features	GORASP2 Expression			
	Total Patients	Weak	Strong	P value
Total Patients	171	93	78	
Age (years)				0.955
<60	42	23	19	
≥60	129	70	59	
Sex				0.009*
Male	100	46	54	
Female	71	47	24	
Pathological Stage				<0.001*
Stage I	114	74	40	
Stage II	29	13	16	
Stage III	26	6	20	
Stage IV	2	0	2	
Vascular Invasion				<0.001*
-	100	71	29	
+	71	22	49	
Lymphatic Permeation				<0.001*
-	108	73	35	
+	63	20	43	
Pathological Subtype				<0.001*
AIS	19	19	0	
MIA	29	27	2	
Lepidic	46	28	18	
Acinar	19	8	11	
Papillary	26	4	22	
Solid	32	7	25	

Table.6 ZYG11A and clinicopathological features. Stage I includes IA and IB, stage II includes IIA and IIB, stage III includes IIIA and IIIB. Correlation between ZYG11A expression and clinicopathological features was analyzed using chi-squared test.

Clinicopathological Features	ZYG11A Expression			P value
	Total Patients	Weak	Strong	
Total Patients	171	98	73	
Age (years)				0.293
<60	42	27	15	
≥60	129	71	58	
Sex				0.096
Male	100	52	48	
Female	71	46	25	
Pathological Stage				0.003*
Stage I	114	75	39	
Stage II	29	15	14	
Stage III	26	8	18	
Stage IV	2	0	2	
Vascular Invasion				<0.001*
-	100	69	31	
+	71	29	42	
Lymphatic Permeation				<0.001*
-	108	73	35	
+	63	25	38	
Pathological Subtype				<0.001*
AIS	19	19	0	
MIA	29	18	11	
Lepidic	46	29	17	
Acinar	19	6	13	
Papillary	26	10	16	
Solid	32	16	16	

Fig.19 GORASP2 KM curve. Disease-free survival depicted as Kaplan-Meier curves

shows the correlation between GORASP2 expression and outcome. Strong expression

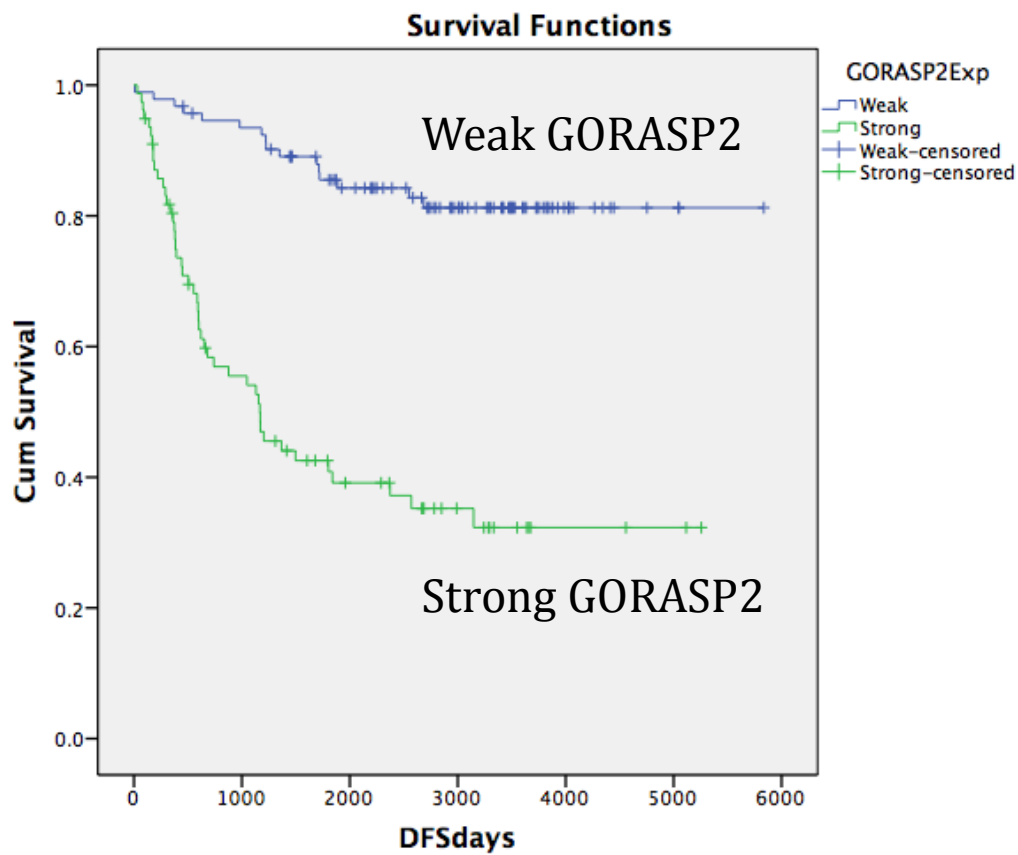of GORASP2 was associated with poor prognosis relative to weak expression
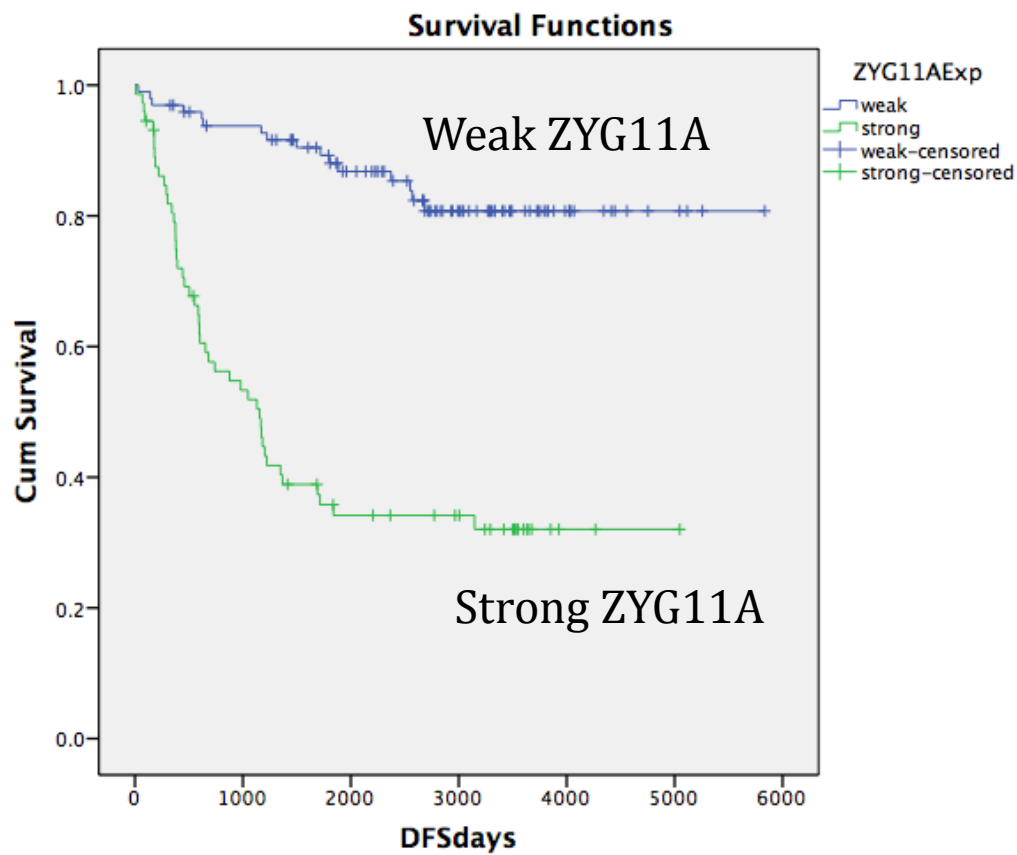
(*p<0.001*)

Fig.20 ZYG11A KM curve. Disease-free survival depicted as Kaplan-Meier curves shows the correlation between ZYG11A expression and outcome. Strong expression of ZYG11A was associated with poor prognosis relative to weak expression (*p<0.001*).

Table.**7 Multivariate analysis.** Adjusted for gender, age, GORASP2 expression, ZYG11A expression, SFN expression, vascular invasion, lymphatic permeation, and pathological stage.

Clin.Features	Univariate Analysis			Multivariate Analysis		
	HR	95% CI	P value	HR	95% CI	Pvalue
Gender	0.394	0.220-0.704	0.002	0.791	0.425-1.473	0.460
Age	1.091	0.625-1.903	0.760			
Vascular Invasion (- vs +)	0.13	0.072-0.234	<0.001	0.272	0.133-0.555	<0.001*
Lymphatic Permeation (- vs +)	0.194	0.115-0.326	<0.001	0.420	0.236-0.747	0.003*
GORASP2 Exp (Weak vs Strg)	0.173	0.098-0.306	<0.001	0.332	0.179-0.613	<0.001*
ZYG11A Exp (Weak vs Strg)	0.161	0.091-0.285	<0.001	0.286	0.155-0.527	<0.001*
SFN Exp (- vs +)	0.555	0.328-0.938	0.028	1.427	0.819-2.487	0.209
Pathological Stg (1 vs other)	0.219	0.132-0.364	<0.001	0.534	0.316-0.903	0.019*

References

1.   Torre LA, Siegel RL, Jemal A. Lung Cancer Statistics. Adv Exp Med Biol 2016;893: 1-19.
2.   Torre LA, Siegel RL, Ward EM, Jemal A. Global Cancer Incidence and Mortality Rates and Trends--An Update. Cancer Epidemiol Biomarkers Prev 2016;25: 16-27.
3.   Cheng TY, Cramb SM, Baade PD, Youlden DR, Nwogu C, Reid ME. The International Epidemiology of Lung Cancer: Latest Trends, Disparities, and Tumor Characteristics. J Thorac Oncol 2016;11: 1653-1671.
4.   Lortet-Tieulent J, Soerjomataram I, Ferlay J, Rutherford M, Weiderpass E, Bray F. International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women. Lung Cancer 2014;84: 13-22.
5.   Nakamura H, Saji H. A worldwide trend of increasing primary adenocarcinoma of the lung. Surgery Today 2014;44: 1004-1012.
6.   Li T, Kung HJ, Mack PC, Gandara DR. Genotyping and genomic profiling of non-small-cell lung cancer: implications for current and future therapies. J Clin Oncol 2013;31: 1039-1049.
7.   Urer HN, Kocaturk CI, Gunluoglu MZ, Arda N, Bedirhan MA, Fener N, Dincer SI. Relationship between lung adenocarcinoma histological subtype and patient prognosis. Ann Thorac Cardiovasc Surg 2014;20: 12-18.
8.   Noguchi M, Morikawa A, Kawasaki M, Matsuno Y, Yamada T, Hirohashi S, Kondo H, Shimosato Y. Small adenocarcinoma of the lung. Histologic characteristics and prognosis. Cancer 1995;75: 2844-2852.
9.   Minami Y, Matsuno Y, Iijima T, Morishita Y, Onizuka M, Sakakibara Y, Noguchi M. Prognostication of small-sized primary pulmonary adenocarcinomas by histopathological and karyometric analysis. Lung Cancer 2005;48: 339-348.
10.   Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, Beer DG, Powell CA, Riely GJ, Van Schil PE, Garg K, Austin JH, Asamura H, Rusch VW, Hirsch FR, Scagliotti G, Mitsudomi T, Huber RM, Ishikawa Y, Jett J, Sanchez-Cespedes M, Sculier JP, Takahashi T, Tsuboi M, Vansteenkiste J, Wistuba I, Yang PC, Aberle D, Brambilla C, Flieder D, Franklin W, Gazdar A, Gould M, Hasleton P, Henderson D, Johnson B, Johnson D, Kerr K, Kuriyama K, Lee JS, Miller VA, Petersen I, Roggli V, Rosell R, Saijo N, Thunnissen E, Tsao M, Yankelewitz D. 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: 244-285.
11.   Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, Chirieac LR, Dacic S, Duhig E, Flieder DB, Geisinger K, Hirsch FR, Ishikawa Y, Kerr KM, Noguchi M, Pelosi G, Powell CA, Tsao MS, Wistuba I. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. Journal of Thoracic Oncology 2015;10: 1243-1260.

12. Noguchi M. Stepwise progression of pulmonary adenocarcinoma—clinical and molecular implications. Cancer and Metastasis Reviews 2010;29: 15-21.

13. Saito M, Shiraishi K, Kunitoh H, Takenoshita S, Yokota J, Kohno T. Gene aberrations for precision medicine against lung adenocarcinoma. Cancer Sci 2016;107: 713-720.

14. The Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511: 543.

15. Nakanishi H, Matsumoto S, Iwakawa R, Kohno T, Suzuki K, Tsuta K, Matsuno Y, Noguchi M, Shimizu E, Yokota J. Whole genome comparison of allelic imbalance between noninvasive and invasive small-sized lung adenocarcinomas. Cancer Res 2009;69: 1615-1623.

16. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res 2016;5: 288-300.

17. Valentino F, Borra G, Allione P, Rossi L. Emerging targets in advanced non-small-cell lung cancer. Future Oncology 2018;14: 61-72.

18. Yu HA, Arcila ME, Rekhtman N, Sima CS, Zakowski MF, Pao W, Kris MG, Miller VA, Ladanyi M, Riely GJ. Analysis of Tumor Specimens at the Time of Acquired Resistance to EGFR TKI therapy in 155 patients with EGFR mutant Lung Cancers. Clinical cancer research : an official journal of the American Association for Cancer Research 2013;19: 2240-2247.

19. Lovly CM, Horn L. Strategies for Overcoming EGFR Resistance in the Treatment of Advanced-Stage NSCLC. Current Treatment Options in Oncology 2012;13: 516-526.

20. Shaw AT, Engelman JA. ALK in Lung Cancer: Past, Present, and Future. Journal of Clinical Oncology 2013;31: 1105-1111.

21. Biswas S, Rao CM. Epigenetics in cancer: Fundamentals and Beyond. Pharmacology & Therapeutics 2017;173: 118-134.

22. Shiba-Ishii A, Kim Y, Shiozawa T, Iyama S, Satomi K, Kano J, Sakashita S, Morishita Y, Noguchi M. Stratifin accelerates progression of lung adenocarcinoma at an early stage. Molecular Cancer 2015;14: 142.

23. Sato T, Shiba‐Ishii A, Kim Y, Dai T, Husni RE, Hong JM, Kano J, Sakashita S, Iijima T, Noguchi M. miR‐3941: A novel microRNA that controls IGBP1 expression and is associated with malignant progression of lung adenocarcinoma. Cancer Sci 2017;108: 536-542.

24. Itoguchi N, Nakagawa T, Murata Y, Li D, Shiba-Ishii A, Minami Y, Noguchi M. Immunocytochemical staining for stratifin and OCIAD2 in bronchial washing specimens increases sensitivity for diagnosis of lung cancer. Cytopathology 2015;26: 354-361.

25. Shiozawa T, Iyama S, Toshima S, Sakata A, Usui S, Minami Y, Sato Y, Hizawa N, Noguchi M. Dimethylarginine dimethylaminohydrolase 2 promotes tumor angiogenesis in lung adenocarcinoma. Virchows Archiv 2016;468: 179-190.

26. Sadikovic B, Al-Romaih K, Squire JA, Zielenska M. Cause and Consequences of Genetic and Epigenetic Alterations in Human Cancer. Current Genomics 2008;9: 394-408.

27. Dupont C, Armant DR, Brenner CA. Epigenetics: Definition, Mechanisms and Clinical Perspective. Seminars in reproductive medicine 2009;27: 351-357.

28. Damaskos C, Tomos I, Garmpis N, Karakatsani A, Dimitroulis D, Garmpi A, Spartalis E, Kampolis CF, Tsagkari E, Loukeri AA, Margonis GA, Spartalis M, Andreatos N, Schizas D, Kokkineli S, Antoniou EA, Nonni A, Tsourouflis G, Markatos K, Kontzoglou K, Kostakis A, Tomos P. Histone Deacetylase Inhibitors as a Novel Targeted Therapy Against Non-small Cell Lung Cancer: Where Are We Now and What Should We Expect? Anticancer Res 2018;38: 37-43.

29. Chen L, Wei T, Si X, Wang Q, Li Y, Leng Y, Deng A, Chen J, Wang G, Zhu S, Kang J. Lysine acetyltransferase GCN5 potentiates the growth of non-small cell lung cancer via promotion of E2F1, cyclin D1, and cyclin E1 expression. J Biol Chem 2013;288: 14510-14521.

30. Kunej T, Godnic I, Ferdin J, Horvat S, Dovc P, Calin GA. Epigenetic regulation of microRNAs in cancer: An integrated review of literature. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 2011;717: 77-84.

31. Han L, Wang W, Ding W, Zhang L. MiR-9 is involved in TGF-beta1-induced lung cancer cell invasion and adhesion by targeting SOX7. J Cell Mol Med 2017;21: 2000-2008.

32. Belinsky SA. Gene-promoter hypermethylation as a biomarker in lung cancer. Nat Rev Cancer 2004;4: 707-717.

33. Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007;128: 683-692.

34. Baylin SB. DNA methylation and gene silencing in cancer. Nat Clin Pract Oncol 2005;2 Suppl 1: S4-11.

35. Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, Jaenisch R. Induction of tumors in mice by genomic hypomethylation. Science 2003;300: 489-492.

36. Szyf M, Pakneshan P, Rabbani SA. DNA demethylation and cancer: therapeutic implications. Cancer Letters 2004;211: 133-143.

37. Shiba-Ishii A, Noguchi M. Aberrant stratifin overexpression is regulated by tumor-associated CpG demethylation in lung adenocarcinoma. Am J Pathol 2012;180: 1653-1662.

38. Robertson KD. DNA methylation and human disease. Nat Rev Genet 2005;6: 597-610.

39. Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? Genes Cancer 2011;2: 607-617.

40. Gu Y, Yang P, Shao Q, Liu X, Xia S, Zhang M, Xu H, Shao Q. Investigation of the expression patterns and correlation of DNA methyltransferases and class I histone deacetylases in ovarian cancer tissues. Oncol Lett 2013;5: 452-458.

41. Qu Y, Mu G, Wu Y, Dai X, Zhou F, Xu X, Wang Y, Wei F. Overexpression of DNA methyltransferases 1, 3a, and 3b significantly correlates with retinoblastoma tumorigenesis. Am J Clin Pathol 2010;134: 826-834.

42. Rahman MM, Qian ZR, Wang EL, Yoshimoto K, Nakasono M, Sultana R, Yoshida T, Hayashi T, Haba R, Ishida M, Okabe H, Sano T. DNA methyltransferases 1, 3a, and 3b overexpression and clinical significance in gastroenteropancreatic neuroendocrine tumors. Hum Pathol 2010;41: 1069-1078.

43. Lin H, Yamada Y, Nguyen S, Linhart H, Jackson-Grusby L, Meissner A, Meletis K, Lo G, Jaenisch R. Suppression of intestinal neoplasia by deletion of Dnmt3b. Mol Cell Biol 2006;26: 2976-2983.
44. Lin RK, Hsu HS, Chang JW, Chen CY, Chen JT, Wang YC. Alteration of DNA methyltransferases contributes to 5'CpG methylation and poor prognosis in lung cancer. Lung Cancer 2007;55: 205-213.
45. Linhart HG, Lin H, Yamada Y, Moran E, Steine EJ, Gokhale S, Lo G, Cantu E, Ehrich M, He T, Meissner A, Jaenisch R. Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing. Genes Dev 2007;21: 3110-3122.
46. Rhee I, Bachman KE, Park BH, Jair KW, Yen RW, Schuebel KE, Cui H, Feinberg AP, Lengauer C, Kinzler KW, Baylin SB, Vogelstein B. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature 2002;416: 552-556.
47. Cao XY, Ma HX, Shang YH, Jin MS, Kong F, Jia ZF, Cao DH, Wang YP, Suo J, Jiang J. DNA methyltransferase3a expression is an independent poor prognostic indicator in gastric cancer. World J Gastroenterol 2014;20: 8201-8208.
48. Oh BK, Kim H, Park HJ, Shim YH, Choi J, Park C, Park YN. DNA methyltransferase expression and DNA methylation in human hepatocellular carcinoma and their clinicopathological correlation. Int J Mol Med 2007;20: 65-73.
49. Shimada A, Kano J, Ishiyama T, Okubo C, Iijima T, Morishita Y, Minami Y, Inadome Y, Shu Y, Sugita S, Takeuchi T, Noguchi M. Establishment of an immortalized cell line from a precancerous lesion of lung adenocarcinoma, and genes highly expressed in the early stages of lung adenocarcinoma development. Cancer Sci 2005;96: 668-675.
50. Aleskandarany MA, Rakha EA, Macmillan RD, Powe DG, Ellis IO, Green AR. MIB1/Ki-67 labelling index can classify grade 2 breast cancer into two clinically distinct subgroups. Breast Cancer Research and Treatment 2010;127: 591-599.
51. Chuturgoon AA, Phulukdaree A, Moodley D. Fumonisin B1 modulates expression of human cytochrome P450 1b1 in human hepatoma (Hepg2) cells by repressing Mir-27b. Toxicology Letters 2014;227: 50-55.
52. Russell PA, Wainer Z, Wright GM, Daniels M, Conron M, Williams RA. Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification. J Thorac Oncol 2011;6: 1496-1504.
53. Weichert W, Warth A. Early lung cancer with lepidic pattern: adenocarcinoma in situ, minimally invasive adenocarcinoma, and lepidic predominant adenocarcinoma. Curr Opin Pulm Med 2014;20: 309-316.
54. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis 2010;31: 27-36.
55. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti

M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, et al. Initial sequencing and analysis of the human genome. Nature 2001;409: 860-921.

56. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983;301: 89-92.

57. Glazer CA, Smith IM, Ochs MF, Begum S, Westra W, Chang SS, Sun W, Bhan S, Khan Z, Ahrendt S, Califano JA. Integrative Discovery of Epigenetically Derepressed Cancer Testis Antigens in NSCLC. PLoS ONE 2009;4: e8189.

58. Gu Y, Wang C, Wang Y, Qiu X, Wang E. Expression of thymosin β 10 and its role in non–small cell lung cancer. Human Pathology 2009;40: 117-124.

59. Hong JA, Kang Y, Abdullaev Z, Flanagan PT, Pack SD, Fischette MR, Adnani MT, Loukinov DI, Vatolin S, Risinger JI, Custer M, Chen GA, Zhao M, Nguyen DM, Barrett JC, Lobanenkov VV, Schrump DS. Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells. Cancer Res 2005;65: 7763-7774.

60. Kim S, Lee S, Lee C, Lee M, Kim Y, Shin D, Choi K, Kim J, Park D, Sol M. Expression of Cancer-Testis Antigens MAGE-A3/6 and NY-ESO-1 in Non-Small-Cell Lung Carcinomas and Their Relationship with Immune Cell Infiltration. Lung 2009;187: 401-411.

61. Nelson HH, Marsit CJ, Christensen BC, Houseman EA, Kontic M, Wiemels JL, Karagas MR, Wrensch MR, Zheng S, Wiencke JK, Kelsey KT. Key epigenetic changes associated with lung cancer development. Epigenetics 2012;7: 559-566.

62. Renaud S, Pugacheva EM, Delgado MD, Braunschweig R, Abdullaev Z, Loukinov D, Benhattar J, Lobanenkov V. Expression of the CTCF-paralogous cancer-testis gene, brother of the regulator of imprinted sites (BORIS), is regulated by three alternative promoters modulated by CpG methylation and by CTCF and p53 transcription factors. Nucleic Acids Research 2007;35: 7372-7388.

63. Gao Q, Steine EJ, Barrasa MI, Hockemeyer D, Pawlak M, Fu D, Reddy S, Bell GW, Jaenisch R. Deletion of the de novo DNA methyltransferase Dnmt3a promotes lung tumor progression. Proc Natl Acad Sci U S A 2011;108: 18061-18066.

64. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511: 543-550.

65. Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, Carmona FJ, Sayols S, Moran S, Heyn H, Vizoso M, Gomez A, Sanchez-Cespedes M, Assenov Y, Muller F, Bock C, Taron M, Mora J, Muscarella LA, Liloglou T, Davies M,

Pollan M, Pajares MJ, Torre W, Montuenga LM, Brambilla E, Field JK, Roz L, Lo Iacono M, Scagliotti GV, Rosell R, Beer DG, Esteller M. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. J Clin Oncol 2013;31: 4140-4147.

66. Gyparaki M-T, Basdra EK, Papavassiliou AG. DNA methylation biomarkers as diagnostic and prognostic tools in colorectal cancer. Journal of Molecular Medicine 2013;91: 1249-1256.

67. Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. Nature Reviews Genetics 2012;13: 679.

68. Husni RE, Shiba-Ishii A, Iiyama S, Shiozawa T, Kim Y, Nakagawa T, Sato T, Kano J, Minami Y, Noguchi M. DNMT3a expression pattern and its prognostic value in lung adenocarcinoma. Lung Cancer 2016;97: 59-65.

69. Stirzaker C, Taberlay PC, Statham AL, Clark SJ. Mining cancer methylomes: prospects and challenges. Trends in Genetics 2014;30: 75-84.

70. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 1994;22: 2990-2997.

71. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL. Genome-wide DNA methylation profiling using Infinium(R) assay. Epigenomics 2009;1: 177-200.

72. Siggens L, Ekwall K. Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. J Intern Med 2014;276: 201-214.

73. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, Abugessaisa I, Fukuda S, Hori F, Ishikawa-Kato S, Mungall CJ, Arner E, Baillie JK, Bertin N, Bono H, de Hoon M, Diehl AD, Dimont E, Freeman TC, Fujieda K, Hide W, Kaliyaperumal R, Katayama T, Lassmann T, Meehan TF, Nishikata K, Ono H, Rehli M, Sandelin A, Schultes EA, t Hoen PA, Tatum Z, Thompson M, Toyoda T, Wright DW, Daub CO, Itoh M, Carninci P, Hayashizaki Y, Forrest AR, Kawaji H. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol 2015;16: 22.

74. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics 2016;8: 389-399.

75. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Djik S, Muhlhausler B, Stirzaker C, Clark SJ. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biology 2016;17: 208.

76. Iida N, Okuda Y, Ogasawara O, Yamashita S, Takeshima H, Ushijima T. MACON: a web tool for computing DNA methylation data obtained by the Illumina Infinium Human DNA methylation BeadArray. Epigenomics 2018;10: 249-258.

77. Detre S, Saclani Jotti G, Dowsett M. A "quickscore" method for immunohistochemical semiquantitation: validation for oestrogen receptor in breast carcinomas. J Clin Pathol 1995;48: 876-878.

78. Tahara S, Tahara T, Horiguchi N, Kato T, Shinkai Y, Yamashita H, Yamada H, Kawamura T, Terada T, Okubo M, Nagasaka M, Nakagawa Y, Shibata T, Yamada S, Urano M, Tsukamoto T, Kurahashi H, Kuroda M, Ohmiya N. DNA methylation accumulation in gastric mucosa adjacent to cancer after Helicobacter pylori eradication. Int J Cancer 2018.

79. Fujii S, Yamashita S, Yamaguchi T, Takahashi M, Hozumi Y, Ushijima T, Mukai H. Pathological complete response of HER2-positive breast cancer to trastuzumab and chemotherapy can be predicted by HSD17B4 methylation. Oncotarget 2017;8: 19039-19048.

80. Shiomi H. SY, Ushijima T. Standard Protocols on Epigenetics. ed.: Yodosha; 2017.

81. Shiba-Ishii A, Kano J, Morishita Y, Sato Y, Minami Y, Noguchi M. High expression of stratifin is a universal abnormality during the course of malignant progression of early-stage lung adenocarcinoma. Int J Cancer 2011;129: 2445-2453.

82. Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, Lehmann U. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. BMC Res Notes 2012;5: 210.

83. Murcia O, Jover R, Egoavil C, Perez-Carbonell L, Juárez M, Hernández-Illán E, Rojas E, Alenda C, Balaguer F, Andreu M, Llor X, Castells A, Boland CR, Goel A. Methylation and Expression Status Does Not Predict Response to 5-FU-based Chemotherapy in Colorectal Cancer. Clinical Cancer Research 2018;24: 2820.

84. Kim Y, Shiba-Ishii A, Nakagawa T, Iemura S-i, Natsume T, Nakano N, Matsuoka R, Sakashita S, Lee S, Kawaguchi A, Sato Y, Noguchi M. Stratifin regulates stabilization of receptor tyrosine kinases via interaction with ubiquitin-specific protease 8 in lung adenocarcinoma. Oncogene 2018.

85. Xu D, Esko JD. A Golgi-on-a-chip for glycan synthesis. Nature Chemical Biology 2009;5: 612.

86. Shorter J, Watson R, Giannakou ME, Clarke M, Warren G, Barr FA. GRASP55, a second mammalian GRASP protein involved in the stacking of Golgi cisternae in a cell-free system. The EMBO Journal 1999;18: 4949-4960.

87. Lin W, Zhu C, Hong J, Zhao L, Jilg N, Fusco DN, Schaefer EA, Brisac C, Liu X, Peng LF, Xu Q, Chung RT. The spliceosome factor SART1 exerts its anti-HCV action through mRNA splicing. Journal of hepatology 2015;62: 1024-1032.

88. Huang S, Wang Y. Golgi structure formation, function, and post-translational modifications in mammalian cells. F1000Research 2017;6: 2050.

89. Ohashi Y, Okamura M, Katayama R, Fang S, Tsutsui S, Akatsuka A, Shan M, Choi H-W, Fujita N, Yoshimatsu K, Shiina I, Yamori T, Dan S. Targeting the Golgi apparatus to overcome acquired resistance of non-small cell lung cancer cells to EGFR tyrosine kinase inhibitors. Oncotarget 2018;9: 1641-1655.

90. Sonneville R, Gonczy P. Zyg-11 and cul-2 regulate progression through meiosis II and polarity establishment in C. elegans. Development 2004;131: 3527-3543.

91. Féral C, Wu Y-Q, Pawlak A, Guellaën G. Meiotic human sperm cells express a leucine-rich homologue of Caenorhabditis elegans early embryogenesis gene, Zyg-11. Molecular Human Reproduction 2001;7: 1115-1122.

92. Vasudevan S, Starostina NG, Kipreos ET. The Caenorhabditis elegans cell-cycle regulator ZYG-11 defines a conserved family of CUL-2 complex components. EMBO Reports 2007;8: 279-286.

93. Wang X, Sun Q, Chen C, Yin R, Huang X, Wang X, Shi R, Xu L, Ren B. ZYG11A serves as an oncogene in non-small cell lung cancer and influences CCNE1 expression. Oncotarget 2016;7: 8029-8042.

94. Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, Teschendorff AE, Butcher LM, Beck S, Muller M. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. Age (Dordr) 2014;36: 9648.

95. Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S, Moore M, Longo DL, Cookson MR, Traynor BJ, Singleton AB. Distinct DNA methylation changes highly correlated with chronological age in the human brain. Hum Mol Genet 2011;20: 1164-1172.

96. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, Vilain E. Epigenetic Predictor of Age. PLoS ONE 2011;6: e14821.

97. Koch CM, Wagner W. Epigenetic-aging-signature to determine age in different tissues. Aging (Albany NY) 2011;3: 1018-1027.

98. D'Aquila P, Rose G, Bellizzi D, Passarino G. Epigenetics and aging. Maturitas 2013;74: 130-136.

Acknowledgements

First and foremost I would like to thank my main advisor Professor Masayuki Noguchi. It has been an honor to be his student. Thanks to him, I had the opportunity to come to Tsukuba and join his laboratory. His passion in research has opened my eyes to the interesting research world despite my background as clinical medical doctor. Under his guidance I overcame many difficulties and learned a lot. His valuable advice, criticism, and extensive discussions made my PhD experience productive and stimulating. I am very grateful for his patience, motivation, enthusiasm, and vast scientific knowledge.

I would also like to thank my second advisor Doctor Aya Shiba for her endless help and careful guidance from the beginning of my study. During these past 5 years, she has taught and always been available to advise me. She helped me a lot during my first day in the laboratory and taught me everything from the very basic until the end. She was always willing to discuss with me and helped me to solve problems when I was in trouble. Her patience guidance really helped me a lot mentally and technically through my PhD journey. I very much appreciate all her contributions of ideas and time to make my publications to be accepted and finished my dissertation on time.

For this dissertation, I also would like to thank to all the members of my PhD committee for their helpful advice and constructive comments.

I want to thank all present and past members of the Noguchi lab for their friendship and support, especially Doctor Shinji Iyama for his constant support in my private and academicals life throughout my journey. You were all very kind and very supportive toward my PhD journey. I will never forget the wonderful time and for every party

we've had together.

<div align="right">

Ryan Edbert Husni

University of Tsukuba

September 2018

</div>