BMJ Open

A favourable prognostic marker for EGFR mutant non-small cell lung cancer: immunohistochemical analysis of MUC5B

Kouki Wakata,[1] Tomoshi Tsuchiya,[1] Koichi Tomoshige,[1] Katsunori Takagi,[1] Naoya Yamasaki,[1] Keitaro Matsumoto,[1] Takuro Miyazaki,[1] Atsushi Nanashima,[1] Jeffrey A Whitsett,[2] Yutaka Maeda,[2] Takeshi Nagayasu[1]

To cite: Wakata K, Tsuchiya T, Tomoshige K, *et al*. A favourable prognostic marker for EGFR mutant non-small cell lung cancer: immunohistochemical analysis of MUC5B. *BMJ Open* 2015;**5**:e008366. doi:10.1136/bmjopen-2015-008366

► Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2015-008366).

Received 30 March 2015
Revised 8 May 2015
Accepted 13 May 2015

[1]Division of Surgical Oncology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan
[2]Division of Neonatology, Perinatal and Pulmonary Biology, Perinatal Institute, Cincinnati Children's Hospital Medical Center and the University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

Correspondence to
Dr Yutaka Maeda; yutaka.maeda@cchmc.org;
Dr Takeshi Nagayasu; nagayasu@nagasaki-u.ac.jp

ABSTRACT

Objectives: To determine the use of the mucin proteins MUC5B and MUC5AC as prognosis markers for non-small cell lung cancer (NSCLC) carrying epidermal growth factor receptor (*EGFR*) mutations.

Setting: Patients who underwent surgical resection at Nagasaki University Hospital and related facilities in Japan between June 1996 and March 2013.

Participant: 159 Japanese patients (male: n=103; female: n=56) with NSCLC, who underwent surgical resection (*EGFR*-mutant type: n=78, *EGFR* wild type: n=81).

Results: Patients whose tumours expressed MUC5B had significantly longer overall survival and relapse-free survival compared to the MUC5B-negative patients with *EGFR* mutant NSCLC (p=0.0098 and p=0.0187, respectively). In patients with *EGFR* wild-type NSCLC, there was no association with MUC5B expression. MUC5AC expression was not different between *EGFR* mutant and wild-type NSCLC.

Conclusions: Present findings indicate that MUC5B, but not MUC5AC, is a novel prognostic biomarker for patients with NSCLC carrying *EGFR* mutations but not for patients with NSCLC carrying wild-type *EGFR*.

Strengths and limitations of this study

- A prognostic marker for each driver mutation in non-small cell lung cancer (NSCLC) has not yet been determined.
- MUC5B is a favourable postoperative prognostic marker for epidermal growth factor receptor (*EGFR*) mutant NSCLC.
- MUC5AC is not correlated with postoperative prognosis regardless of *EGFR* mutation status.
- The function of MUC5B in *EGFR* mutant NSCLC remains unknown.

INTRODUCTION

Lung cancer is the primary cause of cancer-related death in the USA and worldwide.[1] Non-small cell lung cancer (NSCLC) accounts for approximately 80–85% of all lung cancers.[1] Currently, targeted therapies for non-resectable NSCLC have progressed rapidly, based on the discovery of pharmacologically treatable driver mutations in epidermal growth factor receptor (*EGFR*) and fusions of anaplastic lymphoma kinase (ALK).[2 3] These moleculary targeted therapies have revealed distinct and/or overlapping tumorigenic pathways associated with each driver mutation, especially regarding the mechanisms of tumour recurrence.[4] Genetic screening of driver mutations, including *EGFR* mutations and *ALK* fusions, is now common for metastatic NSCLC but not for surgically resected primary NSCLC.[5] In the ALCHEMIST lung cancer trials (http://www.cancer.gov/researchandfunding/areas/clinical-trials/nctn/alchemist), patients whose primary lung tumours carry *EGFR* mutations (*EGFR*-mutant patients) are being tested for adjuvant therapy of erlotinib targeting EGFR mutations. However, a favourable or poor prognostic biomarker associated with *EGFR* mutations is not known. Such biomarkers will be useful to determine *EGFR*-mutant patients who would benefit most from the adjuvant therapy of erlotinib and to avoid such unnecessary therapy after surgery in patients who would not benefit.

Recently, we reported that decreased expression of *Nkx2–1* (also known as TTF-1) in a mouse model of *EGFR* mutant NSCLC reduced the number and size of lung tumours,[6] and extended the survival of the mice (see online supplementary figure S1). Unexpectedly, the decreased *Nkx2–1* induced the expression of a mucin protein MUC5B but not MUC5AC in *EGFR*-mutant

lung tumours in the mice,[6] suggesting that MUC5B may serve as a favourable prognostic marker associated with *EGFR* mutant NSCLC in humans. In the present study, we assessed whether the expression of MUC5B in the primarily resected *EGFR* mutant or wild-type lung tumours is linked to survival of the patients after surgery. Our study provides a novel approach to assess prognosis for patients whose primarily resected lung tumours carry *EGFR* mutations.

METHODS

Study population

Among the patients who underwent surgical resection at Nagasaki University Hospital and related facilities between June 1996 and March 2013, patients who were tested for the presence or absence of *EGFR* mutations were selected for this study. The *EGFR* mutations were confirmed internally or externally (LSI Medience Corporation, Japan). We further selected the patients whose clinicopathological characteristics were retrieved from the patients' charts and whose prognosis was followed at our institution and related facilities. We enrolled 159 patients (*EGFR*-mutant type: n=78, *EGFR* wild type: n=81) for this study (table 1). All investigations were approved by our institution and related facilities' review boards, and informed consent was obtained from all participants prior to the study.

Clinicopathological evaluation

Histological classification of NSCLC was designated as three types—well, moderately and poorly differentiated—based on the predominant features according to the WHO classification.[7] The patients remained for a median follow-up period of 1680 days, ranging from 55 to 4503 days. For all patients, periodic inspection with chest X-ray, CT scan and tumour marker assays was performed at least every 6 months to confirm the presence or absence of recurrence, even if patients experienced no problems or no symptoms.

Antibody information

For immunohistochemical staining, primary antibodies were used at the following concentrations: rabbit polyclonal anti-MUC5B (1:200; sc-20 119, Santa Cruz Biotechnology) and rabbit polyclonal anti-MUC5AC (1:50; sc-20 118, Santa Cruz Biotechnology).

Sample preparation, selection and immunohistochemistry

The 5 µm thick formalin-fixed paraffin-embedded (FFPE) lung sections were deparaffinised in dimethylbenzene and dehydrated through a graded alcohol series. For antigen retrieval, the FFPE lung sections were incubated in 10 mM citric acid (pH 6) at 121°C for 15 min and then washed in phosphate-buffered saline (PBS). Next, the lung sections were immersed in 3% H_2O_2 solution for 30 min to block the endogenous peroxidase followed by incubation with each primary antibody at 4°C overnight. After washing in PBS, the lung sections were incubated with the peroxidase-conjugated secondary antibodies (Simple Stain MAX-PO kit, Nichirei, Tokyo, Japan) for 30 min at room temperature. For immunohistochemistry (IHC) staining, the lung sections were visualised with a diaminobenzidine (DAB: brown) kit (Histofine, Nichirei) and counterstained with H&E. The lung sections visualised with

Table 1 Baseline characteristics of the 159 patients with NSCLC

Number of patients	
Median age (range, years)	67.5 (32–90)
Gender	
Male	103 (65%)
Female	56 (35%)
Smoking status	
Non-smoker	54 (34%)
Smoker	105 (66%)
Histological type	
Adenocarcinoma: bronchoalveolar	33 (21%)
Adenocarcinoma	91 (57%)
Squamous cell carcinoma	32 (20%)
Adenosquamous carcinoma	2 (1%)
Other*	1 (1%)
Median tumour size (range, mm)	32.4 (8–120)
Degree of differentiation	
Well	55 (35%)
Moderately	69 (43%)
Poorly	28 (18%)
MD	7 (4%)
Stage	
IA/IB	81 (51%)
IIA/IIB	40 (25%)
IIIA/IIIB	38 (24%)
Tumour status	
T1–2	136 (86%)
T3–4	23 (14%)
Nodal status	
N0	103 (65%)
N1–3	56 (35%)
Lymphatic invasion	
Negative	56 (35%)
Positive	102 (64%)
MD	1 (1%)
Venous invasion	
Negative	76 (48%)
Positive	82 (52%)
MD	1 (1%)
Adjuvant chemotherapy	
Yes	85 (53%)
No	74 (47%)
EGFR	
Wild type	81 (51%)
Mutant type	78 (49%)

Data are median (range) or number (%) unless otherwise stated.
*Other mean NSCLC neuroendocrine.
EGFR, epidermal growth factor receptor; MD, missing data; NSCLC, non-small cell lung cancer.

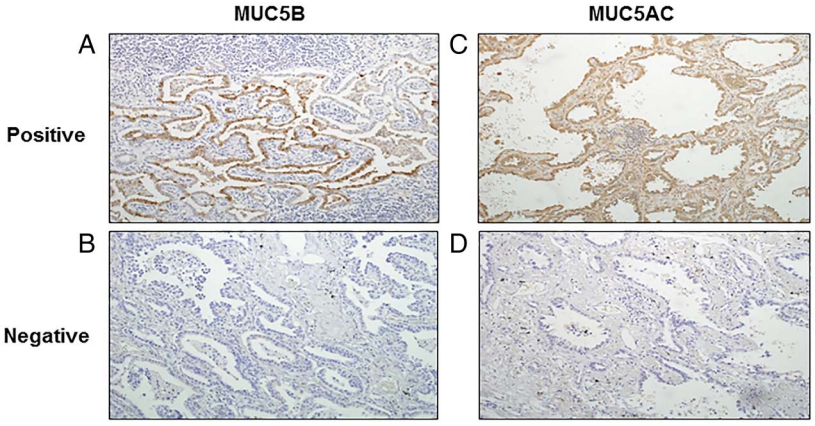

Figure 1 Immunohistochemical staining for MUC5B and MUC5AC expression in NSCLC. Representative images of immune-positive staining for MUC5B in non-small cell lung cancer (A), negative staining (B), positive staining for MUC5AC (C) and negative staining (D).

DAB were dehydrated with alcohol and dimethylbenzene and mounted in a conventional fashion.

Normal bronchial tissue specimens that moderately expressed MUC5B were prepared as positive controls in all cases. Normal gastric mucosa tissue specimens that moderately expressed MUC5AC were prepared as positive controls in all cases. Negative controls were also prepared in all cases. MUC5B and MUC5AC staining was evaluated by IHC by two independent trained observers (KW and TT). The pathological criteria were determined by reference to guideline for human epidermal growth factor receptor 2 (Her2/neu) testing in breast cancer (score 0, no staining observed or incomplete faint/barely perceptible cytoplasmic staining of <10% of tumour cells; score 1, incomplete faint/barely perceptible cytoplasmic staining of >10% of tumour cells; score 2, incomplete weak/moderate cytoplasmic staining of >10% of tumour cells; score 3, complete and intense cytoplasmic staining of >30% of tumour cells).[8] Scores 0 and 1 were further categorised as negative, and scores 2 and 3 as positive.

Statistical analysis

For univariate analysis, categorical data were analysed by the χ^2 test, Fisher's exact test or the Cochran-Armitage test. Continuous data were expressed as a mean using the Mann-Whitney U test or the Kruskal-Wallis test. The overall survival (OS) and relapse-free survival (RFS) were calculated according to the Kaplan-Meier method, and differences between groups were tested for significance using the log-rank test. Participants who neither died nor had recurrence were censored at the time of their last follow-up. The prognostic relevance of a single factor was determined by multivariate Cox regression analysis. A p value of 0.05 or less was considered significant. SPSS V.17 software (SPSS Japan, Tokyo, Japan) was used for the analysis.

RESULTS

Expression of MUC5B and MUC5AC in human NSCLC

Since MUC5B is an abundant cytoplasmic and secreted protein, we assessed whether MUC5B could be used as a prognostic marker for patients with NSCLC carrying *EGFR* mutations in primary resected human lung tumours. Primary resected NSCLC tumours were tested immunohistochemically for the presence of MUC5B. MUC5B staining was detected in the cytoplasm of NSCLC cells in 27 of the 78 samples with *EGFR* mutations and 29 of the 81 samples with wild-type *EGFR* (figure 1A, B). The NSCLC samples were also tested using MUC5AC antibody, detecting expression of MUC5AC in the cytoplasm of NSCLC cells in 20 of the 73 samples with *EGFR* mutations and 24 of the 79 samples with wild-type *EGFR* (figure 1C, D). These results indicate that both MUC5B and MUC5AC are expressed in a portion of human NSCLC.

Prognostic association of MUC5B or MUC5AC with *EGFR* mutant or wild-type NSCLC

Expression of MUC5B in NSCLC tumours carrying *EGFR* mutations was not correlated with clinicopathological parameters, including age, gender, smoking status, histological type, tumour size, degree of differentiation, stage, tumour status, nodal status, lymphatic invasion, venous invasion or adjuvant chemotherapy (table 2). Expression of MUC5B in NSCLC tumours with wild-type *EGFR* was not correlated with all of the clinicopathological parameters but was correlated with the histological type (see online supplementary table S1).

OS and RFS for patients with NSCLC carrying *EGFR* mutations or wild-type *EGFR* were assessed. In a cohort of patients whose resected NSCLC tumours carried *EGFR* mutations, univariate analysis showed significant differences ($p<0.05$) in OS in expression of MUC5B, tumour size, histological type, degree of differentiation, stage, lymphatic invasion and venous invasion, and in RFS in expression of MUC5B, tumour size, degree of differentiation, stage, lymphatic invasion and venous invasion (table 3). Patients whose tumours expressed MUC5B (MUC5B-positive patients) survived significantly longer than patients whose tumours did not express MUC5B (MUC5B-negative patients) in both OS (5-year OS; 95.8% vs 65.1%, p=0.0098; figure 2A) and RFS (5-year RFS; 69.9% vs 44%, p=0.0187; figure 2B).

Table 2 Association with clinicopathological data and the expression of MUC5B of patients with EGFR-mutant NSCLC

Parameters	Total (n=78)	MUC5B Negative (−) (n=51)	MUC5B Positive (+) (n=27)	p Value
Median age (range, years)	66.9 (41–85)	66.7 (42–83)	67.3 (41–85)	0.674
Gender				
Male	35 (45%)	22 (28%)	13 (17%)	
Female	43 (55%)	29 (37%)	14 (18%)	0.6721
Smoking status				
Non-smoker	42 (54%)	28 (36%)	14 (18%)	
Smoker	36 (46%)	23 (29%)	13 (17%)	0.7971
Histological type				
Adenocarcinoma: bronchoalveolar	27 (35%)	16 (21%)	11 (14%)	
Adenocarcinoma	43 (55%)	29 (37%)	14 (18%)	
Squamous cell carcinoma	8 (10%)	6 (8%)	2 (3%)	0.6522
Median tumour size (range, mm)	25.7 (8–60)	26.1 (8–60)	24.9 (8–50)	0.8771
Degree of differentiation				
Well	35 (45%)	22 (28%)	13 (17%)	
Moderately	28 (36%)	19 (24%)	9 (12%)	
Poorly	11 (14%)	6 (8%)	5 (6%)	0.7348
MD	4 (5%)	4 (5%)	0 (0%)	
Stage				
IA/IB	46 (59%)	29 (37%)	17 (22%)	
IIA/IIB	14 (18%)	8 (10%)	6 (8%)	
IIIA/IIIB	18 (23%)	14 (18%)	4 (5%)	0.4162
Tumour status				
T1–2	68 (87%)	42 (54%)	26 (33%)	
T3–4	10 (13%)	9 (12%)	1 (1%)	0.0797
Nodal status				
N0	53 (68%)	35 (45%)	18 (23%)	
N1–3	25 (32%)	16 (21%)	9 (12%)	0.8599
Lymphatic invasion				
Negative	35 (45%)	26 (33%)	9 (12%)	
Positive	42 (54%)	24 (31%)	18 (23%)	0.1165
MD	1 (1%)	1 (1%)	0 (0%)	
Venous invasion				
Negative	45 (58%)	30 (38%)	15 (19%)	
Positive	32 (41%)	20 (26%)	12 (15%)	0.7057
MD	1 (1%)	1 (1%)	0 (0%)	
Adjuvant chemotherapy				
Yes	33 (42%)	20 (26%)	13 (17%)	
No	45 (58%)	31 (40%)	14 (18%)	0.4475

Data are median (range) or number (%) unless otherwise stated.
EGFR, epidermal growth factor receptor; MD, missing data; NSCLC, non-small cell lung cancer.

Multivariate Cox regression analysis using the variables that were p<0.05 in univariate analysis showed that the expression of MUC5B was independently associated with better OS and RFS (p<0.05; table 4). In a cohort of patients whose resected NSCLC tumours had wild-type *EGFR*, univariate analysis showed significant differences (p<0.05) in OS in smoking status, stage, venous invasion and adjuvant chemotherapy, and in RFS, lymphatic invasion and venous invasion (see online supplementary table S2). There was no significant difference between the MUC5B-positive patients and MUC5B-negative patients in OS and RFS (5-year OS; 59.5% vs 63.6%, 5-year RFS; 36% vs 48.5%, respectively, figure 2C, D). Expression of MUC5AC in NSCLC was not associated with OS and RFS regardless of *EGFR* mutation status (figure 3). These results indicate that MUC5B is a favourable prognostic marker for postoperative patients whose resected NSCLC tumours carry *EGFR* mutation but not for those with wild type-*EGFR*.

DISCUSSION

In the present study, we demonstrate that expression of MUC5B in primary *EGFR* mutant NSCLC is associated with longer survival in patients with NSCLC. MUC5B, but not MUC5AC, is a favourable prognostic biomarker for NSCLC in humans carrying *EGFR* mutations. Our study also indicates that adjuvant chemotherapy is not

Table 3 Univariate analysis for OS and RFS in EGFR-mutant patients with NSCLC

Parameters	N	OS		RFS	
		Survival (%)	p Value	Survival (%)	p Value
Age (years)					
<70	39	75.5		63.6	
≥70	39	76.3	0.7311	41.5	0.1207
Gender					
Male	35	69		54.4	
Female	43	81.5	0.3166	51.8	0.813
Smoking status					
Non-smoker	43	75.9		49.1	
Smoker	36	73.3	0.9754	55.8	0.8508
MUC5B expression					
Positive	27	95.8		69.9	
Negative	51	65.1	0.0098	44	0.0187
Tumour size					
<20 mm	25	95		85.7	
≥20 mm	44	64.4	0.0058	31.2	0.0001
Histological type					
Adenocarcinoma	69	78.8		54	
Squamous cell carcinoma	8	46.9	0.0245	50	0.6369
Degree of differentiation					
Well	35	90		72.6	
Moderately	28	62.3		34.5	
Poorly	11	50.6	0.0172	38.1	0.0137
Stage					
I	45	87.8		62.8	
II/III	33	59.3	0.0095	38.8	0.0308
Lymphatic invasion					
Negative	35	92.3		74.2	
Positive	42	62.7	0.0075	35.8	0.0011
Venous invasion					
Negative	45	89		66.5	
Positive	32	55.7	0.0045	25.9	0.0021
Adjuvant chemotherapy					
Yes	33	62		37.5	
No	45	82.8	0.2569	63.4	0.0183

Data are p values by Kaplan-Meier analysis.
EGFR, epidermal growth factor receptor; NSCLC, non-small cell lung cancer; OS, overall survival; RFS, relapse-free survival.

effective for EGFR-mutant patients, suggesting that mutant EGFR-targeting drugs, including gefitinib or erlotinib, should be used as an adjuvant therapy mainly for MUC5B-negative EGFR-mutant patients who have a poorer prognosis than MUC5B-positive EGFR-mutant patients. Our results using MUC5B as a prognosis biomarker for EGFR-mutant patients should be integrated into the ALCHEMIST lung cancer trials to determine patients who would benefit most from the adjuvant therapy.

MUC5B has been assessed as a prognostic biomarker for multiple cancers in several studies, using reverse transcription-PCR, microarray analysis and IHC (see online supplementary table S3).[9–14] Messenger RNA (mRNA) data assessing MUC5B as a prognostic biomarker is available at PrognoScan, a database for meta-analysis of the prognostic value of genes using microarray data deposited to the public domain.[15] The prognostic impact of MUC5B expression differed among cancer types. In lung cancer, six microarray studies analysed by PrognoScan did not indicate MUC5B as either a good or a poor prognostic biomarker.[16–21] Immunohistochemical analysis indicated MUC5B as a poor prognosis biomarker (see online supplementary table S3), [9 14] a finding contradicting our present study. Previous mRNA microarray and immunostaining were based on all NSCLCs independent of driver mutation-based classification, which differs from our analysis, which was based on classification by *EGFR* mutations. The utility of MUC5B as a prognostic factor differed in the two breast cancer studies, depending on the molecular basis of the tumours. In all breast cancers, PrognoScan indicated that MUC5B was associated with poor prognosis;[22 23] however, in (estrogen receptor-) positive breast cancers, MUC5B was associated with favourable prognosis,[24] indicating the potential

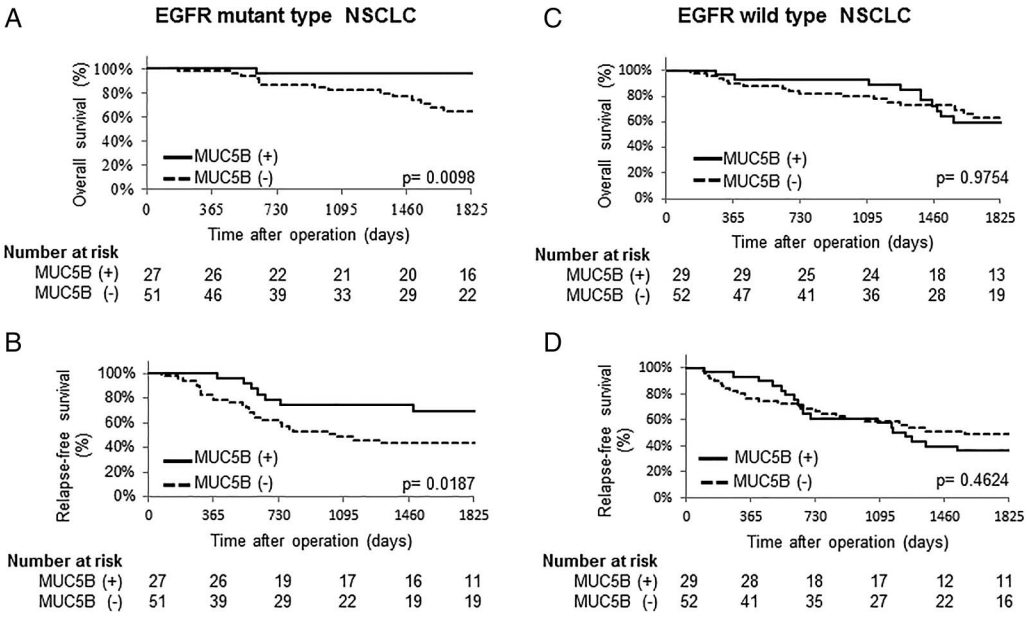

Figure 2 Survival curves for patients based on the expression of MUC5B in epidermal growth factor receptor (*EGFR*) mutant or wild-type non-small cell lung cancer (NSCLC). Overall and relapse-free survivals (OS and RFS) for patients with NSCLC carrying *EGFR* mutations or *EGFR* wild type. OS (A) RFS (B) in the patients with *EGF*- mutant type NSCLC and OS (C) and RFS (D) in the patients with EGFR wild-type NSCLC.

importance of tumour classification on a molecular basis. In the present study, we assessed MUC5B as a biomarker for NSCLC based on *EGFR* mutation status rather than on all NSCLCs, identifying MUC5B as a favourable prognosis biomarker for *EGFR* mutant NSCLC.

Regulation of *MUC5B* in *EGFR* mutant NSCLC is not well understood. *MUC5AC* and *MUC5B* genes are closely located at a locus on human chromosome 11. Both are evolutionally conserved gel-forming mucins secreted from airway epithelial cells in the lung. In normal lung, MUC5B is constitutively expressed at higher levels than MUC5AC.[25] In asthma and other inflammatory lung diseases, MUC5AC is highly induced in airway goblet cells.[25] In idiopathic pulmonary fibrosis (IPF), MUC5B but not MUC5AC is highly expressed in the airway goblet cells.[26 27] The single-nucleotide polymorphism (SNP) rs35705950 located at the *MUC5B* promoter is associated with induction of *MUC5B* mRNA in IPF;[27] however, we detected the SNP rs35705950 in only one of 27 cases in the *EGFR* mutant NSCLC expressing MUC5B (data not shown), indicating that the SNP is not associated with increased MUC5B in *EGFR* mutant NSCLC. Since MUC5B was induced in *EGFR* mutant lung tumours in *Nkx2–1* heterozygous mice ($EGFR^{L858R}$; $Nkx2\text{–}1^{+/-}$), *Muc5b* is suppressed by NKX2–1 in *EGFR* mutant NSCLC in mice.[6] Regulatory mechanisms controlling the *MUC5B* gene are not understood in *EGFR* mutant NSCLC in humans. The function of MUC5B in cancer has been analysed using a truncated MUC5B in MCF7 breast cancer cells, truncated MUC5B promoting tumourigenesis of MCF7 cells.[13] However, the use of truncated MUC5B may obscure the intrinsic role of full-length MUC5B since there is a possibility that the truncated MUC5B may function in a dominant-negative fashion. In lung, MUC5B is required for mucociliary

Table 4 Multivariate analysis for OS and RFS in EGFR-mutant patients with NSCLC

	HR	95% CI	p Value
OS			
MUC5B (positive vs negative)	0.053	0.0064 to 0.4402	0.0065
Differentiation (well vs moderately vs poorly)	0.3762	0.1357 to 1.0428	0.0602
Stage I vs stage II/III	0.5199	0.1687 to 1.6019	0.2545
Lymphatic invasion (positive vs negative)	2.9524	0.6572 to 13.2631	0.1578
RFS			
MUC5B (positive vs negative)	0.2913	0.1233 to 0.6886	0.005
Stage I vs stage II/III	0.5937	0.2735 to 1.2888	0.1874
Lymphatic invasion (positive vs negative)	3.7624	1.5294 to 9.2556	0.0039

EGFR, epidermal growth factor receptor; NSCLC, non-small cell lung cancer; OS, overall survival; RFS, relapse-free survival.

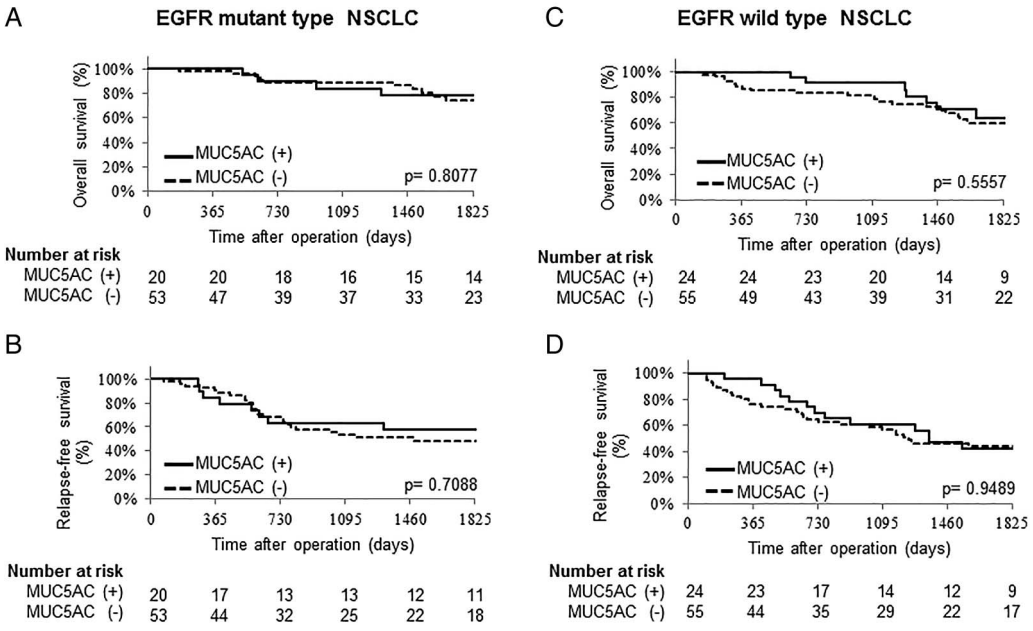

Figure 3 Survival curves for patients based on the expression of MUC5AC in epidermal growth factor receptor (***EGFR***) mutant or wild type non-small cell lung cancer (NSCLC). Overall and relapse-free survivals (OS and RFS) for patients with NSCLC carrying *EGFR* mutations or *EGFR* wild type. OS (A) RFS (B) in the patients with *EGFR*-mutant type NSCLC and OS (C) and RFS (D) in the patients with *EGFR* wild-type NSCLC.

clearance and innate immunity against bacterial infection.[28] The potential functions of MUC5B in lung cancer, including *EGFR* mutant lung cancer, are not known. The present study suggests that MUC5B or processes regulating *MUC5B* may influence the growth and metastasis of *EGFR* mutant NSCLC. MUC5B may serve as a surrogate biomarker influenced by a pathway involved in metastasis and recurrence associated with *EGFR* mutant NSCLC.

In conclusion, our data revealed the clinicopathological significance of MUC5B as a favourable prognostic factor in resected *EGFR* mutant NSCLC. Further studies are necessary to elucidate the gene regulatory mechanism and the function of MUC5B in *EGFR* mutant NSCLC.

Acknowledgements The authors would like to thank the patients, their families, all of the investigators who participated in the study, the central laboratories that carried out the *EGFR* mutation testing and Dr Mary Durbin for discussions.

Contributors KW, TT, JAW, YM and TN participated in conception and design. KW, TT, KTo, KTa, NY, KM, TM, AN, JAW and YM were involved in provision of study material, patients and data acquisition. KW, TT, KTo, KM, JAW, YM and TN were involved in data analysis and interpretation. KW, TT, JAW, YM and TN were responsible for drafting manuscript and intellectual content.

Funding This work was supported by the Ministry of Education, Science, and Culture of Japan, National Institute of Health, American Lung Association, the University of Cincinnati Postdoctoral Fellow Research Program and Cincinnati Children's Hospital Medical Center.

Competing interests None declared.

Patient consent Obtained.

Ethics approval Ethics Committee from every participant hospital.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data available.

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics for hispanics/latinos, 2012. *CA Cancer J Clin* 2012;62:283–98.
2. Lynch T, Bell D, Sordella R, *et al*. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129–39.
3. Paez JG, Jänne PA, Lee JC, *et al*. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–500.
4. Camidge DR, Pao W, Sequist LV. Acquired resistance to TKIs in solid tumours: learning from lung cancer. *Nat Rev Clin Oncol* 2014;11:473–81.
5. Shames DS, Wistuba II. The evolving genomic classification of lung cancer. *J Pathol* 2014;232:121–33.
6. Maeda Y, Tsuchiya T, Hao H, *et al*. KrasG12D and Nkx2-1 haploinsufficiency induce mucinous adenocarcinoma of the lung. *J Clin Invest* 2012;122:4388–400.
7. 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: executive summary. *Proc Am Thorac Soc* 2011;8:381–5.
8. Wolff AC, Hammond MEH, Hicks DG, *et al*. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol* 2013;31:3997–4013.
9. Yu CJ, Yang PC, Shun CT, *et al*. Overexpression of MUC5 genes is associated with early post-operative metastasis in non-small-cell lung cancer. *Int J Cancer* 1996;69:457–65.
10. Pinto-de-Sousa J, Reis CA, David L, *et al*. MUC5B expression in gastric carcinoma: relationship with clinico-pathological parameters

and with expression of mucins MUC1, MUC2, MUC5AC and MUC6. *Virchows Arch* 2004;444:224–30.
11. Varangot M, Barrios E, Sóñora C, *et al*. Clinical evaluation of a panel of mRNA markers in the detection of disseminated tumor cells in patients with operable breast cancer. *Oncol reports* 2005;14:537–45.
12. Partheen K, Levan K, Osterberg L, *et al*. Expression analysis of stage III serous ovarian adenocarcinoma distinguishes a sub-group of survivors. *Eur J Cancer* 2006;42:2846–54.
13. Valque H, Gouyer V, Gottrand F, *et al*. MUC5B leads to aggressive behavior of breast cancer MCF7 cells. *PLoS ONE* 2012;7:e46699.
14. Nagashio R, Ueda J, Ryuge S, *et al*. Diagnostic and prognostic significances of MUC5B and TTF-1 expressions in resected non-small cell lung cancer. *Sci Rep* 2015;5:8649.
15. Mizuno H, Kitada K, Nakai K, *et al*. PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics* 2009;2:18.
16. Shedden K, Taylor JMG, Enkemann SA, *et al*. Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008;14:822–7.
17. Okayama H, Kohno T, Ishii Y, *et al*. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* 2012;72:100–11.
18. Bild AH, Yao G, Chang JT, *et al*. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006;439:353–7.
19. Zhu CQ, Ding K, Strumpf D, *et al*. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol* 2010;28:4417–24.
20. Lee ES, Son DS, Kim SH, *et al*. Prediction of recurrence-free survival in postoperative non-small cell lung cancer patients by using an integrated model of clinical information and gene expression. *Clin Cancer Res* 2008;14:7397–404.
21. Raponi M, Zhang Y, Yu J, *et al*. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res* 2006;66:7466–72.
22. Pawitan Y, Bjöhle J, Amler L, *et al*. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 2005;7:R953–64.
23. Miller LD, Smeds J, George J, *et al*. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA* 2005;102:13550–5.
24. Chanrion M, Negre V, Fontaine H, *et al*. A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. *Clin Cancer Res* 2008;14:1744–52.
25. Evans C, Kim K, Tuvim M, *et al*. Mucus hypersecretion in asthma: causes and effects. *Curr Opin Pulm Med* 2009;15:4–11.
26. Plantier L, Crestani B, Wert SE, *et al*. Ectopic respiratory epithelial cell differentiation in bronchiolised distal airspaces in idiopathic pulmonary fibrosis. *Thorax* 2011;66:651–7.
27. Seibold M, Wise A, Speer M, *et al*. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011;364:1503–12.
28. Roy MG, Livraghi-Butrico A, Fletcher AA, *et al*. Muc5b is required for airway defence. *Nature* 2014;505:412–16.