Table 2 Adverse events of chemotherapy according to NCI-CTC version 2

	Preoperative S-1 + CDDP (n = 51)		Postoperative S-1 (n = 34)	
	Grade 3/4	%	Grade 3/4	%
Leukocytopenia	6	12	1	3
Neutropenia	13	25	4	12
Thrombocytopenia	2	4	0	0
Anemia	6	12	1	3
Fatigue	5	10	2	6
Anorexia	11	22	3	9
Nausea	6	12	2	6
Vomiting	4	8	0	0
Diarrhea	1	2	2	6
DIC	1	2	0	0
Dehydration	1	2	0	0
Dizziness	0	0	1	3

NCI-CTC National Cancer Institute common toxicity criteria, *CDDP* cisplatin, *DIC* disseminated intravascular coagulation

Table 3 Relationship between pre-stage IV factors and R0 resection

	No.	R0	%R0	MST (95% CI) (months)
Total	51	26	51.0	19.2 (15.4–43.5)
Multiple pre-stage IV factors	27	7	25.9	11.9 (9.0–17.5)
Single pre-stage IV factor	24	19	79.2	56.3 (19.8–NE)
Pre-Cy1	12	12	100	NE (56.3–NE)
Pre-P1	5	2	40	43.5 (7.9–NE)
Pre-N3	5	4	80	31.6 (17.3–NE)
Pre-H1	2	1	50	NE (15.9–NE)

Pre- pretreatment clinical tumor status at registration, *MST* median survival time, *CI* confidence interval, *NE* not estimated

they underwent surgery. R0 resection after induction chemotherapy was accomplished in 26 patients. Non-curative factors in R1/2 cases (n = 18) included 3 cases of unresectable primary tumor, 13 cases of post-P1 and/or Cy1, 1 case of residual nodal metastasis, and 1 case of multiple liver metastasis. The relationship between stage IV status at registration (pre-stage IV) and R0 resection is summarized in Table 3. Twelve patients showed pre-Cy1 without other c-stage IV factors (Cy1 alone). In these 12 patients, eradication of free peritoneal gastric cancer cells (FPGCs) with preoperative chemotherapy was observed and this led to R0 resection. Of 15 patients with pre-Cy0, two patients became post-Cy1 after preoperative chemotherapy.

Postoperative chemotherapy and recurrence

Of the 26 patients that underwent R0 resection, three patients showed no recurrence without postoperative chemotherapy. Reasons for not performing postoperative chemotherapy in these three patients were that 1 patient was ineligible for the starting criteria of the postoperative chemotherapy, 1 patient had prolonged pneumonia, and 1 patient refused. Postoperative chemotherapy was started in 34 patients, including 23 patients with R0 resection, and

Table 4 Postoperative chemotherapy and recurrence in patients with R0 resection (n = 23)

	Start	Completion	Rec.-free (%)
Total	23	11	9 (39)
Multiple pre-stage IV factors	7	2	0 (0)
Single pre-stage IV factor	16	9	9 (56)
Pre-Cy1	11	8	7
Pre-P1	1	0	0
Pre-N3	3	0	1
Pre-H1	1	1	1

Pre- pretreatment clinical tumor status at registration, *Rec.-free* recurrence free

was completed in 11. No patients with noncurative resection completed 1 year of postoperative chemotherapy (n = 11). The relationship between postoperative chemotherapy and recurrence in R0 resection patients (n = 23) is summarized in Table 4. The adverse effects of the postoperative chemotherapy are presented in Table 2. Grade 4 toxic effects of neutropenia, fatigue, and dizziness were observed in one patient. Throughout the treatment period, severe adverse events were recorded in three patients (5.9%). No treatment-related death occurred. Of the 26 patients with R0 resection, 14 experienced recurrence. The

most frequent recurrence site was the peritoneum ($n = 8$), followed by the lymph nodes ($n = 3$).

Discussion

The results of the present study indicate that the multimodal treatment used was safe and well tolerated by patients with stage IV GC. The clinical response rate for the preoperative chemotherapy was compatible with that reported in a previous study [6]. Surgical morbidity was acceptable. No treatment-related death occurred. With a mature follow-up, the 2-year OS rate of this treatment was 43% (95% CI 29.4–56.1). This appeared to be better than the 2-year OS of 20–25% that was estimated from previous studies of stage IV GC patients treated with chemotherapy alone or with surgery and postoperative chemotherapy [2, 3, 6, 22]. However, our study failed to show a survival benefit, because our threshold 2-year survival probability was estimated to be 35%.

Although most of the eligible patients in our study had microscopic or macroscopic peritoneal dissemination, R0 resection was achieved in 51% of the patients after preoperative chemotherapy with S-1 plus cisplatin. R0 resection is reported to be one of the most reliable prognostic indicators for patients after preoperative chemotherapy [10, 17]. In line with these reports, the R0 resection group in the present study exhibited a significantly better prognosis. It is well known that the prognosis in patients with GC with multiple stage IV factors is miserable [23–25]. By virtue of staging laparoscopy, we frequently confirmed overt peritoneal carcinomatosis (pre-P1) or FPGCs (pre-Cy1). Other synchronous pre-stage IV factors were identified in 80% of pre-P1 patients and 66% of pre-Cy1 patients. The R0 resection rate was low in these patients, mainly due to remaining peritoneal carcinomatosis or FPGCs. Preoperative S-1 plus cisplatin might be inadequate for treating GC with multiple c-stage IV factors. In this situation, preoperative staging laparoscopy after the induction chemotherapy could have been useful to prevent futile laparotomy.

As for GC with a single pre-stage IV factor, the R0 resection rate in our study was as high as 79%. R0 resection due to eradication of FPGCs with preoperative chemotherapy was recorded in all patients with pre-Cy1 alone, leading to an improved prognosis. Generally, the 2-year overall survival (OS) rate of GC with pre-Cy1 alone (around 25–30%) is poor, even when conventional postoperative chemotherapy is administered [22, 26, 27]. In general, preoperative chemotherapy using a conventional regimen is insufficient for GC with pre-Cy1 [28]. Recently, improved prognosis was reported in patients with pre-Cy1 alone as a result of the eradication of FPGCs with preoperative cisplatin and folic acid plus fluorouracil. However, the eradication rate was low (37%), and a shift from negative to positive cytology findings during preoperative chemotherapy was observed in 24% of patients [29]. As for adjuvant chemotherapy for GC with post-Cy1 alone, the 2-year OS rate of these patients who underwent radical surgery followed by S-1 alone was reported to be 47% [30]. In contrast, in the present study, the eradication rate of FPGCs was 100%, leading to a 2-year OS rate of 75.0% (95% CI 40.8–91.2) in the patients with pre-Cy1 alone, and two patients (3.9%) had a shift from negative to positive cytology findings. This result was compatible with findings in our preliminary reports [7, 31]. S-1 shows a high rate of transfer into the peritoneal cavity [32]. Compared with S-1 alone, S-1 plus cisplatin improves the prognosis of patients with peritoneal metastasis [6]. These features of S-1 plus cisplatin might have contributed to the high eradication rate of FPGCs in the present study. Also, due to the high eradication rate of FPGCs, pretreatment staging laparoscopy is strongly recommended to distinguish advanced GC with pre-Cy1 from that with pre-Cy0.

Low compliance is often reported for postoperative chemotherapy, because the toxicity of combination regimens cannot be tolerated by patients after gastrectomy [8, 9]. Postoperative S-1 plus cisplatin has also been reported to be too toxic after gastrectomy [33]. To improve compliance with postoperative chemotherapy, S-1 alone was adopted as a postoperative regimen [12]. Consequently, the toxic profile of postoperative S-1 alone was milder than that of preoperative S-1 plus cisplatin. In the present study, the low completion rate of the postoperative chemotherapy was due to tumor progression in patients with multiple stage IV factors. Despite the postoperative chemotherapy, all patients with multiple pre-stage IV factors experienced recurrence after R0 resection. On the other hand, the recurrence-free rate of patients with a single pre-stage IV factor was 56% after postoperative chemotherapy. In particular, due to the low frequency of recurrence, patients with pre-Cy1 alone showed a high postoperative chemotherapy completion rate. Postoperative chemotherapy with S-1 alone might be effective in treating stage IV GC if the latent tumor burden is minimal after R0 resection.

In conclusion, the administration of preoperative S-1 plus cisplatin, followed by surgery and postoperative S-1, is safe and feasible for stage IV GC. However, this multimodal approach failed to show a beneficial effect for GC with multiple stage IV factors. Although the sample size in our study was small and 2 of 15 pre-Cy0 patients became post-Cy1, it seems that patients with pre-Cy1 with no additional non-curative factors might be good candidates for the present approach. Further clinical investigations for this subset should be undertaken using promising perioperative chemotherapy.

Acknowledgments This study was supported by the Foundation for Biomedical and Innovation, Japan.

Conflict of interest The authors indicated no potential conflicts of interest.

References

1. Parkin DM, Bray FI, Devesa SS. Cancer burden in the year 2000. The global picture. Eur J Cancer. 2001;37(Suppl 8):S4–66.
2. Japanese Gastric Cancer Association Registration Committee, Maruyama K, Kaminishi M, Hayashi K, Isobe Y, Honda I, et al. Gastric cancer treated in 1991 in Japan: data analysis of nationwide registry. Gastric Cancer. 2006;9:51–66.
3. Nio Y, Tsubono M, Kawabata K, Masai Y, Hayashi H, Meyer C, et al. Comparison of survival curves of gastric cancer patients after surgery according to the UICC stage classification and the General Rules for Gastric Cancer Study by the Japanese Research Society for gastric cancer. Ann Surg. 1993;218:47–53.
4. Van Cutsem E, Moiseyenko VM, Tjulandin S, Majlis A, Constenla M, Boni C, et al. V325 Study Group. Phase III study of docetaxel and cisplatin plus fluorouracil compared with cisplatin and fluorouracil as first-line therapy for advanced gastric cancer: a report of the V325 Study Group. J Clin Oncol. 2006;24:4991–7.
5. Ross P, Nicolson M, Cunningham D, Valle J, Seymour M, Harper P, et al. Prospective randomized trial comparing mitomycin, cisplatin, and protracted venous-infusion fluorouracil (PVI 5-FU) With epirubicin, cisplatin, and PVI 5-FU in advanced esophagogastric cancer. J Clin Oncol. 2002;20:1996–2004.
6. Koizumi W, Narahara H, Hara T, Takagane A, Akiya T, Takagi M, et al. S-1 plus cisplatin versus S-1 alone for first-line treatment of advanced gastric cancer (SPIRITS trial): a phase III trial. Lancet Oncol. 2008;9:215–21.
7. Okabe H, Ueda S, Obama K, Hosogi H, Sakai Y. Induction chemotherapy with S-1 plus cisplatin followed by surgery for treatment of gastric cancer with peritoneal dissemination. Ann Surg Oncol. 2009;16:3227–36.
8. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, et al. MAGIC Trial Participants. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. N Engl J Med. 2006;355:11–20.
9. Boige V, Pignon J, Saint-Aubert B, Lasser P, Conroy T, Bouche O. Final results of a randomized trial comparing preoperative 5-fluorouracil (F)/cisplatin (P) to surgery alone in adenocarcinoma of stomach and lower esophagus (ASLE): FNLCC ACCORD07-FFCD 9703 trial. J Clin Oncol. 2007;25:4510.
10. Brenner B, Shah MA, Karpeh MS, Gonen M, Brennan MF, Coit DG, et al. A phase II trial of neoadjuvant cisplatin-fluorouracil followed by postoperative intraperitoneal floxuridine-leucovorin in patients with locally advanced gastric cancer. Ann Oncol. 2006;17:1404–11.
11. Yoshikawa T, Sasako M, Yamamoto S, Sano T, Imamura H, Fujitani K, et al. Phase II study of neoadjuvant chemotherapy and extended surgery for locally advanced gastric cancer. Br J Surg. 2009;96:1015–22.
12. Sakuramoto S, Sasako M, Yamaguchi T, Kinoshita T, Fujii M, Nashimoto A, et al. ACTS-GC Group. Adjuvant chemotherapy for gastric cancer with S-1, an oral fluoropyrimidine. N Engl J Med. 2007;357:1810–20.
13. National Cancer Institute. Common Toxicity Criteria version 2.0 (CTC). http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm [accessed December, 2010].
14. Japanese Gastric Cancer Association. Japanese classification of gastric carcinoma, 2nd edition. Gastric Cancer. 1998;1:10–24.
15. Fukuya T, Honda H, Hayashi T, Kaneko K, Tateshi Y, Ro T, et al. Lymph-node metastases: efficacy for detection with helical CT in patients with gastric cancer. Radiology. 1995;197:705–11.
16. D'Elia F, Zingarelli A, Palli D, Grani M. Hydro-dynamic CT preoperative staging of gastric cancer: correlation with pathological findings. A prospective study of 107 cases. Eur Radiol. 2000;10:1877–85.
17. Biondi A, Persiani R, Cananzi F, Zoccali M, Vigorita V, Tufo A, et al. R0 resection in the treatment of gastric cancer: room for improvement. World J Gastroenterol. 2010;16:3358–70.
18. D'Ugo DM, Pende V, Persiani R, Rausei S, Picciocchi A. Laparoscopic staging of gastric cancer: an overview. J Am Coll Surg. 2003;196:965–74.
19. Ozmen MM, Zulfikaroglu B, Ozalp N, Ziraman I, Hengirmen S, Sahin B. Staging laparoscopy for gastric cancer. Surg Laparosc Endosc Percutan Tech. 2003;13:241–4.
20. Blackshaw GR, Barry JD, Edwards P, Allison MC, Thomas GV, Lewis WG. Laparoscopy significantly improves the perceived preoperative stage of gastric cancer. Gastric Cancer. 2003;6:225–9.
21. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst. 2000;92:205–16.
22. Fukagawa T, Katai H, Saka M, Morita S, Sasajima Y, Taniguchi H, et al. Significance of lavage cytology in advanced gastric cancer patients. World J Surg. 2010;34:563–8.
23. Korenaga D, Tsujitani S, Haraguchi M, Okamura T, Tamada R, Sugimachi K, et al. Long-term survival in Japanese patients with far advanced carcinoma of the stomach. World J Surg. 1988;12:236–40.
24. Kikuchi S, Tsukamoto H, Mieno H, Sato K, Kobayashi N, Shimao H, et al. Results of resection of gastric cancer with distant metastases. Hepatogastroenterology. 1998;45:592–6.
25. Maekawa S, Saku M, Maehara Y, Sadanaga N, Ikejiri K, Anai H, et al. Surgical treatment for advanced gastric cancer. Hepatogastroenterology. 1996;43:178–86.
26. Miyashiro I, Takachi K, Doki Y, Ishikawa O, Ohigashi H, Murata K, et al. When is curative gastrectomy justified for gastric cancer with positive peritoneal lavage cytology but negative macroscopic peritoneal implant? World J Surg. 2005;29:1131–4.
27. Mezhir JJ, Shah MA, Jacks LM, Brennan MF, Coit DG, Strong VE. Positive peritoneal cytology in patients with gastric cancer: natural history and outcome of 291 patients. Ann Surg Oncol. 2010;17:3173–80.
28. Badgwell B, Cormier JN, Krishnan S, Yao J, Staerkel GA, Lupo PJ, et al. Does neoadjuvant treatment for gastric cancer patients with positive peritoneal cytology at staging laparoscopy improve survival? Ann Surg Oncol. 2008;15:2684–91.
29. Lorenzen S, Panzram B, Rosenberg R, Nekarda H, Becker K, Schenk U, et al. Prognostic significance of free peritoneal tumor cells in the peritoneal cavity before and after neoadjuvant chemotherapy in patients with gastric carcinoma undergoing potentially curative resection. Ann Surg Oncol. 2010;17:2733–9.
30. Kodera Y, Ito S, Mochizuki Y, Kondo K, Koshikawa K, Suzuki N, et al. A phase II study of radical surgery followed by postoperative chemotherapy with S-1 for gastric carcinoma with free cancer cells in the peritoneal cavity (CCOG0301 study). Eur J Surg Oncol. 2009;35:1158–63.
31. Satoh S, Hasegawa S, Ozaki N, Okabe H, Watanabe G, Nagayama S, et al. Retrospective analysis of 45 consecutive patients with advanced gastric cancer treated with neoadjuvant

chemotherapy using an S-1/CDDP combination. Gastric Cancer. 2006;9:129–35.
32. Mori T, Fujiwara Y, Yano M, Tamura S, Yasuda T, Takiguchi S, et al. Prevention of peritoneal metastasis of human gastric cancer cells in nude mice by S-1, a novel oral derivative of 5-fluorouracil. Oncology. 2003;64:176–82.
33. Kodera Y, Ishiyama A, Yoshikawa T, Kinoshita T, Ito S, Yokoyama H, et al. A feasibility study of postoperative chemotherapy with S-1 and cisplatin (CDDP) for gastric carcinoma (CCOG0703). Gastric Cancer. 2010;13:197–203.

D924–D929 *Nucleic Acids Research, 2012, Vol. 40, Database issue* *Published online 2 December 2011*
doi:10.1093/nar/gkr1188

HGPD: Human Gene and Protein Database, 2012 update

Yukio Maruyama[1], Yoshifumi Kawamura[2], Tetsuo Nishikawa[3], Takao Isogai[4], Nobuo Nomura[5] and Naoki Goshima[1,*]

[1]National Institute of Advanced Industrial Science and Technology (AIST), [2]Japan Biological Informatics Consortium (JBIC), Aomi, Koto-ku, Tokyo 135-0064, [3]Life Science Research Laboratory, Central Research Laboratory, Hitachi Ltd, Kokubunji, Tokyo 185-8601, [4]Graduate School of Pharmaceutical Sciences, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033 and [5]Department of Human Studies, Musashino University, Nishi-Tokyo, Tokyo 202-8585, Japan

Received September 8, 2011; Revised November 14, 2011; Accepted November 15, 2011

ABSTRACT

The Human Gene and Protein Database (HGPD; http://www.HGPD.jp/) is a unique database that stores information on a set of human Gateway entry clones in addition to protein expression and protein synthesis data. The HGPD was launched in November 2008, and 33 275 human Gateway entry clones have been constructed from the open reading frames (ORFs) of full-length cDNA, thus representing the largest collection in the world. Recently, research objectives have focused on the development of new medicines and the establishment of novel diagnostic methods and medical treatments. And, studies using proteins and protein information, which are closely related to gene function, have been undertaken. For this update, we constructed an additional 9974 human Gateway entry clones, giving a total of 43 249. This set of human Gateway entry clones was named the Human Proteome Expression Resource, known as the 'HuPEX'. In addition, we also classified the clones into 10 groups according to protein function. Moreover, *in vivo* cellular localization data of proteins for 32 651 human Gateway entry clones were included for retrieval from the HGPD. In 'Information Overview', which presents the search results, the ORF region of each cDNA is now displayed allowing the Gateway entry clones to be searched more easily.

INTRODUCTION

In the post-genomic period, one of the most essential areas of research involves the functional and structural analysis of gene products (proteins). The key element in functional genomics studies is the acquisition of full-length cDNA clones. For this purpose, a number of projects such as the Japanese FLJ project, supported by New Energy and Industrial Technology Development Organization (NEDO) (1–3), and the Kazusa long cDNA project (4,5) have been implemented for the isolation of as many full-length cDNAs as possible and at the highest quality (6,7). To build an infrastructure that facilitates the systematic and comprehensive expression of human proteins, in addition to the availability of full-length cDNA clones, it is vital to have a versatile system for making use of these clones. The Gateway cloning system (Invitrogen, CA, USA) is based on such versatile expression vectors (8); therefore, we adopted this system and subsequently constructed 33 275 human Gateway entry clones from full-length cDNA (9). Accordingly, the Human Gene and Protein Database (HGPD; http://www.HGPD.jp/) was launched in November 2008 (10). Sequence information and protein expression data for the Gateway entry clones can also be retrieved from the HGPD, making it a unique database.

During the post-genomic period, research objectives have focused on the development of new medicines and the establishment of novel diagnostic methods and medical treatments. In addition, a large number of studies using proteins and protein information, which are closely related to gene function, have been undertaken. Thus, in further investigations, the combination of resources and protein information has an important role in accelerating research (11). The HGPD is a unique resource that stores information on a set of human genes, proteins and Gateway entry clones in addition to protein expression and synthesis data.

In this update of the HGPD, we improved sequence information and protein expression data for the human Gateway entry clones. Moreover, we generated new data

*To whom correspondence should be addressed. Tel: +81 3 3599 8137; Fax: +81 3 3599 8141; Email: n-goshima@aist.go.jp

for the functional classification and subcellular localization of proteins produced by the human Gateway entry clones. In the database interface, the open reading frame (ORF) of each cDNA sequence is visualized in 'Information Overview', thereby, allowing users to search the entry clones more easily. In addition, we also redesigned the entire web interface.

DATA IMPROVEMENT

In this update, we improved the sequence information and protein expression data for the Gateway entry clones. In the HGPD, biological data such as the *in vitro* expression data of human proteins are presented on the frame of cDNA clusters. To build the basic frame, cDNA sequences from FLJ and other public databases (Human ESTs, RefSeq, Ensembl, MGC and so on) are assembled onto the genome sequences. The sequence information of the Gateway entry clones is presented with the source cDNA.

Gateway entry clones

To facilitate the use of full-length cDNA clones, we adopted the versatile Gateway expression system that offers high-throughput gene transfer technology for functional gene analysis and protein expression. For conversion to entry clones, we selected an ORF region in each cDNA that meets one of the following criteria: (i) ORF encoding a product >150 amino acids (although the longest ORF starting with an AUG codon has the highest priority, the selected ORF is finally determined by taking into consideration homology search results of shorter ORFs with BLASTX(nr) and BLASTP against the SwissProt and RefSeq databases); (ii) both 149 amino acids > ORF > 100 amino acids and an ORF with an ATGpr value (12) >0.4 and (iii) both 100 amino acids > ORF and a known gene. The ORF regions were then PCR-amplified with the attB1 and attB2 sequences of the Gateway system at both ends and recombined with the attP1 and attP2 sequences of the Gateway pDONR201 donor vector (Invitrogen) (for details, see http://riodb.ibase.aist.go.jp/hgpd/sys_info/help.html#w120_gw). In this update, we constructed an additional 9974 Gateway entry clones (corresponding to approximately 2000 loci), giving a total of 43 249 clones (Table 1). A sequence summary and the nucleotide acid and amino acid sequence data for these clones (v 2.0) can be downloaded from http://riodb.ibase.aist.go.jp/hgpd/sys_info/download.html. This resource of human Gateway entry clones was named HuPEX. If you inquire our Gateway entry clones, you can order them by using Gateway entry clone IDs which are green character in 'Protein Info'.

Table 1. Improvement of entries of HGPD

Data set	V. 1.0	V. 2.0 (in this update)
Gateway entry clones	33 275 (14 590 loci[a]) (N-type: 12 754) (F-type: 20 521)	43 249 (16 484 loci[a]) (N-type: 17 802) (F-type: 25 447)
In vitro protein expression (SDS–PAGE) patterns	13 364	17 821
In vivo subcellular localization image	–	32 651

[a]Numbers of locus in HGPD are represented.

SDS–PAGE patterns of human proteins synthesized *in vitro*

The Gateway system is a versatile expression vector system that can adequately handle large numbers of clones. For the expression of human proteins, we adopted the wheat germ cell-free protein synthesis system (13). In this update, we additionally expressed 4457 human proteins with a C-terminal V5 or His tag and analyzed them using SDS–PAGE (Table 1). The expression data from a total of 17 821 human proteins are stored in the HGPD. These expression patterns are displayed on the 'PE: Protein Expression' page (Figure 1; for details, see http://riodb.ibase.aist.go.jp/hgpd/sys_info/help.html#w120_pe).

RECENT DEVELOPMENTS

In this update of the database, we provided new classification information for the protein function and subcellular localization of the Gateway entry clones.

Classification under protein function

To facilitate searching of the Gateway entry clones, we considered the addition of useful annotation within the HGPD. In the HGPD, certain information on the cDNA clones and Gateway entry clones can be accessed on the 'Information Overview' page. We previously stored sequence information and homology search results (i.e. BLAST, Pfam, PROSITE, SignalP, SOSUI and GO) for each cDNA clone. Then, we focused on the protein function of the Gateway entry clones, and matched the entry clones to RefSeq clones using a homology search (BLAST). On the basis of the homology search results, we were able to match the Gateway entry clones to the annotations in public databases (i.e. NCBI Entrez Gene, HPRD, Swiss Prot, OMIM and GO) and published data. In this update, we classified the entry clones into 10 groups according to protein function; these groups were included in the HGPD (Table 2). For example, 407 gene symbols (genes) were classified as 'Protein kinase', representing 1322 Gateway entry clones within HuPEX. Thus, a total of 4000 gene symbols (genes) were classified, representing 10 684 Gateway entry clones within HuPEX. These classifications can be retrieved from 'Category Search' under 'Advanced Search' in the HGPD (Figure 2).

Subcellular localization of human proteins fused with fluorescent proteins

To construct image data for the subcellular localization of proteins, 32 651 entry clones were used for the overexpression of proteins with fluorescent protein's tag at the N- or C-terminus in HeLa cells. In this subcellular localization study, there are some differences between the

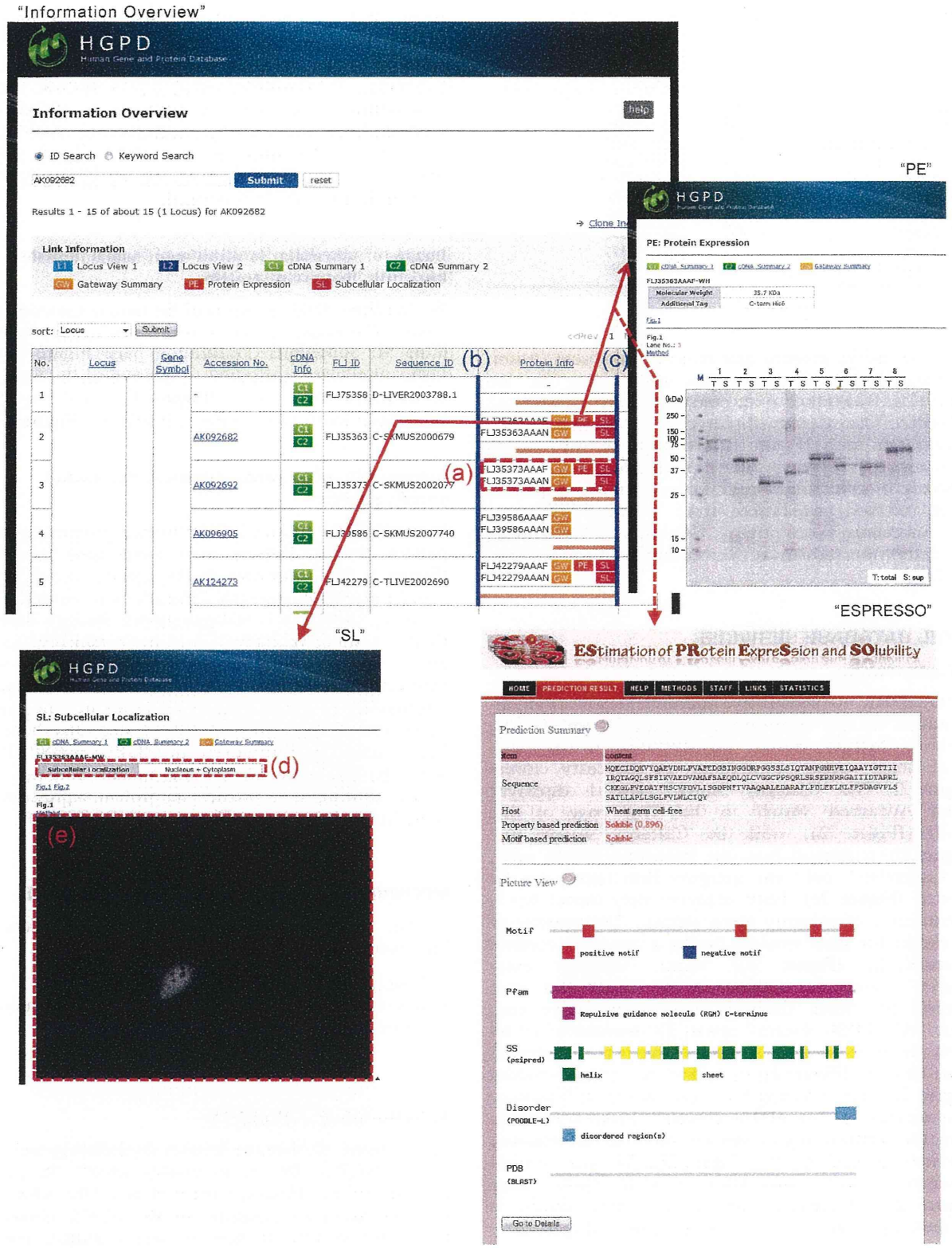


Figure 1. Improved display of the HGPD. In 'Information Overview', (**a**) open reading frames (ORFs) of each cDNA clone can be visualized. In 'Protein Info', the left endpoint represents the N-terminus of each 'Locus' (**b**) and the right endpoint represents the C-terminus (**c**). 'SL', the subcellular localization of human proteins, is linked by the 'SL' button in 'Protein Info'. (**d**) Protein subcellular localization by visual observation of the images. (**e**) The subcellular localization image. 'PE' or 'ESPRESSO' information is linked by the 'PE' button in 'Protein Info'.

Table 2. Classification under protein function

Category	Numbers of gene symbol	Numbers of Gateway entry clone
Protein kinase	385	1322
Protein phosphatase	84	291
G-protein coupled receptor	208	405
Small GTPase	134	255
Transcription factor	1704	4319
Protease	399	1104
Gricans	159	405
Ion channel	177	395
Epigenetics	171	519
Mitochondrial protein	637	1669

results of native proteins and those of fluorescent fusion proteins (14,15). Therefore, the image data for the subcellular localization of ectopic proteins are constructed in HGPD. These data are displayed in the 'SL: Subcellular Localization' page (Figure 1; for details, see http://riodb.ibase.aist.go.jp/hgpd/sys_info/help.html#w120_sl). The protocols of subcellular localization analysis are explained in the method page linked from 'Method' button at 'SL: Subcellular Localization'. The kinds of subcellular localization are shown at 'SL: Subcellular Localization' on 'Help'.

NOVEL DATABASE SERVICES

Category search

In this update, we constructed additional classification information for the Gateway entry clones (Table 2). The interface 'Category Search' was developed so that users could obtain useful information on the entry clones (Figure 2). The 'Category Search' view is displayed within 'Advanced Search' in the 'Top' page of the HGPD (Figure 2a), while the 'Category Search' tab (Figure 2b) is selected from the 'Advanced Search' view. In 'Categories', only one category item (term) can be selected (Figure 2c). Each category item (term) has a second level of category items (terms). The number of hit results for gene symbols during a search is counted automatically (Figure 2d), where 'Gateway entry clone(s)' counts the number of hit results for gene symbols for which Gateway entry clones have been made. 'All cDNA clone(s)' counts the number of all hit results for gene symbols. A results table is then displayed under 'Results' (Figure 2f) by clicking the 'Submit' button (Figure 2e) after selecting the radio button of 'Gateway entry clone(s)' or 'All cDNA clone(s)'. From this results table, the details of each result are linked to 'Information Overview' by clicking 'IO' (Figure 2h). The gene symbol provided for each result links to NCBI Entrez Gene (Figure 2g). Moreover, the list of search results is obtained by clicking the 'Download' button (Figure 2i).

Visualization of ORFs

In 'Information Overview', the ORF of each cDNA clone is visualized in the 'Protein Info' column as an orange line (Figure 1a). For the displayed ORF, the left endpoint (Figure 1b) represents the N-terminus of each 'Locus' and the right endpoint (Figure 1c) represents the C-terminus. It should be noted that if multiple 'Locus' information is indicated by a search in 'Information Overview', the left and right endpoints of each 'Locus' are different. By visualizing the ORF of each cDNA sequence, a rough homology among the cDNA sequences at each 'Locus' can be confirmed.

Images of subcellular localization of human proteins fused with fluorescent proteins

We localized 32 651 proteins of the human Gateway entry clones. The localized images of the proteins are displayed in the 'SL: Subcellular Localization' page (Figure 1; 'SL'). Each subcellular localization is determined by visual observation of the images (Figure 1d), and the actual subcellular localization image is displayed (Figure 1e).

Linkage with other public databases that predict protein expression

In the HGPD, a total of 17 821 human protein expression patterns for the Gateway entry clones have been stored (Figure 1; 'PE'); however, in this update, not all human protein expression patterns have yet been entered in the database. ESPRESSO (Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, http://mbs.cbrc.jp/ESPRESSO/), which predicts protein expression and solubility from sequence information, has been launched. To predict the protein expression of the Gateway entry clones that have not been characterized, we linked the HGPD to ESPRESSO using the ORF sequences of the human Gateway entry clones, allowing us to predict the protein expression and solubility of these clones (Figure 1; 'ESPRESSO').

IMPROVEMENT OF THE DATABASE SERVICE

In this update, the following database services are improvement.

- Redesigned the entire web interface.
- Addition of the function of sort for 'Gene Symbol' in 'Information Overview'.

FUTURE DEVELOPMENTS

In the future, we hope to develop the following two areas of the HGPD. One is to further classify the protein function of the Gateway entry clones. The other is to link the accession numbers of the cDNA clones and their gene symbol to their protein accession numbers (e.g. NCBI reference sequence NP and XP numbers, Swiss-Prot entry names and accession numbers and so on), and to improve the 'ID search' tool by including protein accession numbers.

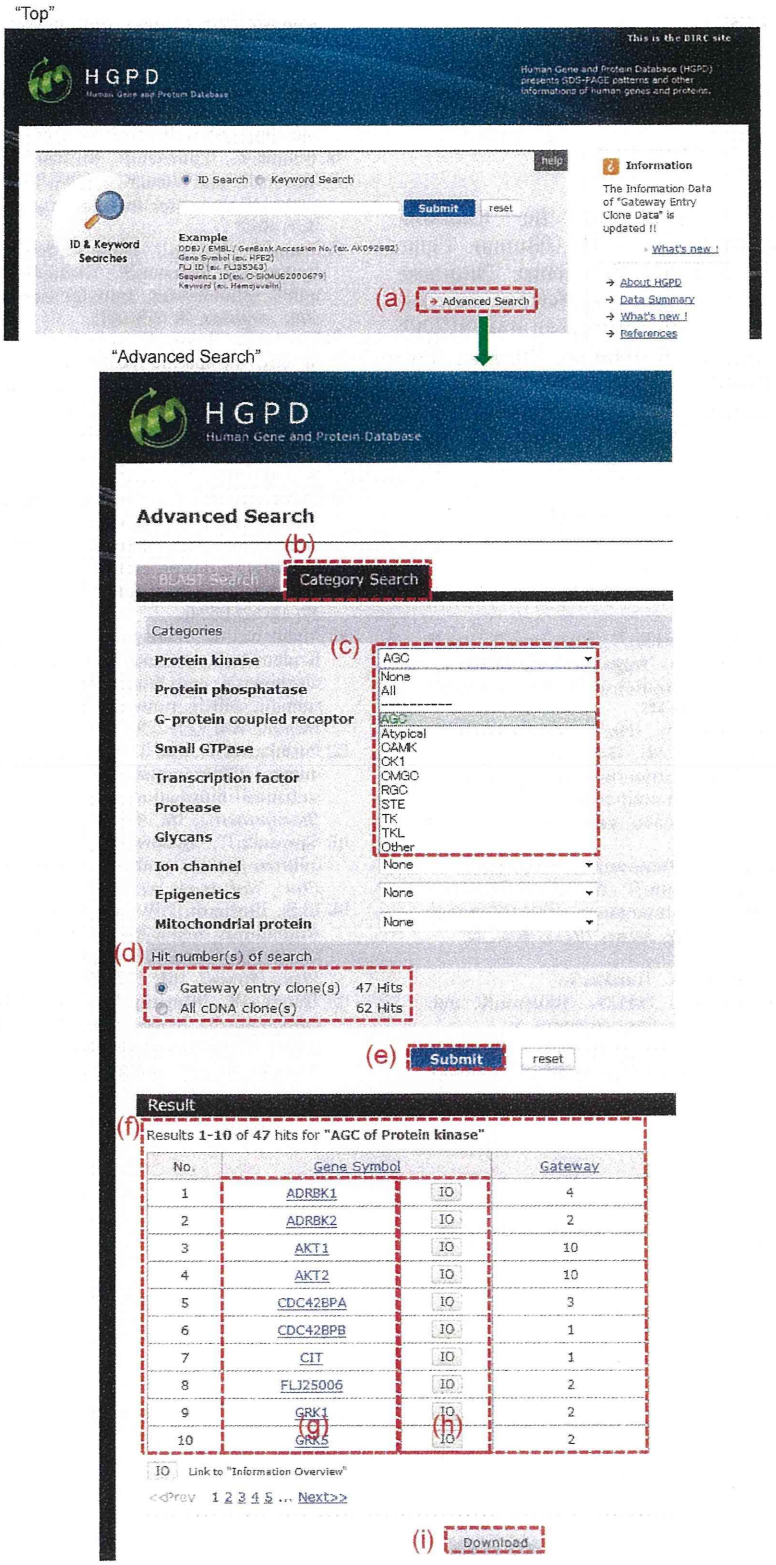


Figure 2. Category Search. This interface is accessed through 'Advanced Search' under 'Top View' (**a**). (**b**) A category search is initiated by clicking the 'Category Search' tab. (**c**) In each category, only one item (term) can be selected. (**d**) The number of hits from a selected item (term) is then counted, and (**f**) displayed by clicking the 'Submit' button (**e**). (**h**) The results are then linked to 'Information Overview' by clicking 'IO'. (**g**) The 'Gene Symbol' interface is linked to NCBI Entrez Gene. (**i**) The list of search results is obtained by clicking the 'Download' button.

ACKNOWLEDGEMENTS

We thank the Helix Research Institute and the Research Association for Biotechnology for the FLJ cDNA clones.

FUNDING

Research Information Database of the Tsukuba Advanced Computing Center of AIST (Human Gene and Protein Database); National Bioscience Database Center and Database Center for Life Science (mirror site of Human Gene and Protein Database) and NEDO project, 'Functional Analysis of Human Proteins and its Application (2001–2005)' and 'Development of Basic Technology to Control Biological Systems Using Chemical Compounds (2006–2010)'. Funding for open access charge: AIST.

Conflict of interest statement. None declared.

REFERENCES

1. Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R., Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K. *et al.* (2004) Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat. Genet.*, **36**, 40–45.
2. Kimura,K., Wakamatsu,A., Suzuki,Y., Ota,T., Nishikawa,T., Yamashita,R., Yamamoto,J., Sekine,M., Tsuritani,K., Wakaguri,H. *et al.* (2006) Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res.*, **16**, 55–65.
3. Imanishi,T., Itoh,T., Suzuki,Y., O'Donovan,C., Fukuchi,S., Koyanagi,O.K., Barrero,A.R., Tamura,T., Yamaguchi-Kabata,Y., Tanino,M. *et al.* (2004) Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol.*, **2**, 0001–0020.
4. Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y., Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S. (1994) Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA Res.*, **1**, 27–35.
5. Ohara,O., Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N. and Nomura,N. (1997) Construction and characterization of human brain cDNA libraries suitable for analysis of clones encoding relatively large proteins. *DNA Res.*, **4**, 53–59.
6. Temple,G., Lamesch,P., Milstein,S., Hill,D.E., Wagner,L., Moore,T. and Vidal,M. (2006) Proteome: developing expression clone resources for the human genome. *Hum. Mol. Genet.*, **15**, R31–R43.
7. Yang,X., Boehm,J., Yang,X., Salehi-Ashtiani,K., Hao,T., Shen,Y., Lubonja,R., Thomas,S., Alkan,O., Bhimdi,T. *et al.* (2011) A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods*, **8**, 659–661.
8. Hartley,J., Temple,G. and Brasch,M. (2000) DNA cloning using in vitro site-specific recombination. *Genome Res.*, **10**, 1788–7895.
9. Goshima,N., Kawamura,Y., Fukumoto,A., Miura,A., Honma,R., Sato,R., Wakamatsu,A., Yamamoto,J., Kimura,K., Nishikawa,T. *et al.* (2008) Human protein factory for converting the transcriptome into an in vitro-expressed proteome. *Nat. Methods*, **5**, 1011–1017.
10. Maruyama,Y., Wakamatsu,A., Kawamura,Y., Kimura,K., Yamamoto,J., Nishikawa,T., Sugano,S., Goshima,N., Isogai,T. and Nomura,N. (2009) Human Gene and Protein Database (HGPD): a novel database presenting a large quantity of experiment-based results in human proteomics. *Nucleic Acid Res.*, **37**, D762–D766.
11. Maekawa,M., Yamaguchi,K., Nakamura,T., Shibukawa,R., Kodanaka,I., Ichisaka,T., Kawamura,Y., Mochizuki,H., Goshima,N. and Yamanaka,S. (2011) Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. *Nature*, **474**, 225–229.
12. Nishikawa,T., Ota,T. and Isogai,T. (2000) Prediction whether a human cDNA sequence contains initiation codon by combining statistical information and similarity with protein sequences. *Bioinformatics*, **16**, 960–967.
13. Sawasaki,T., Ogasawara,T., Morishita,R. and Endo,Y. (2002) A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl Acad. Sci. USA*, **99**, 14652–14657.
14. Li,S., Ehrhardt,D.W. and Rhee,S.Y. (2006) Systematic analysis of Arabidopsis organelles and a protein localization database for facilitating fluorescent tagging of full-length arabidopsis proteins. *Plant Physiol.*, **141**, 527–539.
15. Tian,G.-W., Mohanty,A., Chary,S.N., Li,S., Paap,B., Drakakaki,G., Kopec,C.D., Li,J., Ehrhardt,D., Jackson,D. *et al.* (2004) High-throughput fluorescent tagging of full-length Arabidopsis gene products in planta. *Plant Physiol.*, **135**, 25–38.

OPEN ACCESS Freely available online

PLoS one

CAXII Is a Sero-Diagnostic Marker for Lung Cancer

Makoto Kobayashi[1,2], Toshihide Matsumoto[1,4], Shinichiro Ryuge[6], Kengo Yanagita[1,2], Ryo Nagashio[1,2], Yoshitaka Kawakami[3], Naoki Goshima[3], Shi-Xu Jiang[4], Makoto Saegusa[4], Akira Iyoda[5], Yukitoshi Satoh[5], Noriyuki Masuda[6], Yuichi Sato[1,2]*

1 Department of Applied Tumor Pathology, Graduate School of Medical Sciences, Kitasato University, Kanagawa, Japan, **2** Department of Molecular Diagnostics, School of Allied Health Sciences, Kitasato University, Kanagawa, Japan, **3** Biomedicinal Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan, **4** Department of Pathology, School of Medicine, Kitasato University, Kanagawa, Japan, **5** Department of Thoracic and Cardiovascular Surgery, School of Medicine, Kitasato University, Kanagawa, Japan, **6** Department of Respiratory Medicine, School of Medicine, Kitasato University, Kanagawa, Japan

Abstract

To develop sero-diagnostic markers for lung cancer, we generated monoclonal antibodies using pulmonary adenocarcinoma (AD)-derived A549 cells as antigens by employing the random immunization method. Hybridoma supernatants were immunohistochemically screened for antibodies with AMeX-fixed and paraffin-embedded A549 cell preparations. Positive clones were monocloned twice through limiting dilutions. From the obtained monoclonal antibodies, we selected an antibody designated as KU-Lu-5 which showed intense membrane staining of A549 cells. Based on immunoprecipitation and MADLI TOF/TOF-MS analysis, this antibody was recognized as carbonic anhydrase XII (CAXII). To evaluate the utility of this antibody as a sero-diagnostic marker for lung cancer, we performed dot blot analysis with a training set consisting of sera from 70 lung cancer patients and 30 healthy controls. The CAXII expression levels were significantly higher in lung cancer patients than in healthy controls in the training set ($P<0.0001$), and the area under the curve of ROC was 0.794, with 70.0% specificity and 82.9% sensitivity. In lung cancers, expression levels of CAXII were significantly higher in patients with squamous cell carcinoma (SCC) than with AD ($P=0.035$). Furthermore, CAXII was significantly higher in well- and moderately differentiated SCCs than in poorly differentiated ones ($P=0.027$). To further confirm the utility of serum CAXII levels as a sero-diagnostic marker, an additional set consisting of sera from 26 lung cancer patients and 30 healthy controls was also investigated by dot blot analysis as a validation study. Serum CAXII levels were also significantly higher in lung cancer patients than in healthy controls in the validation set ($P=0.030$). Thus, the serum CAXII levels should be applicable markers discriminating lung cancer patients from healthy controls. To our knowledge, this is the first report providing evidence that CAXII may be a novel sero-diagnostic marker for lung cancer.

Citation: Kobayashi M, Matsumoto T, Ryuge S, Yanagita K, Nagashio R, et al. (2012) CAXII Is a Sero-Diagnostic Marker for Lung Cancer. PLoS ONE 7(3): e33952. doi:10.1371/journal.pone.0033952

Editor: Vladimir Brusic, Dana-Farber Cancer Institute, United States of America

Received September 23, 2011; **Accepted** February 20, 2012; **Published** March 16, 2012

Funding: This study was supported in part by a Grant-in-Aid for Scientific Research Council (23590414) from the Japan Society for the Promotion of Science, grants from the Third Term Comprehensive Control Research for Cancer conducted by the Ministry of Health, Labour and Welfare of Japan, as well as from the Research Project (No. 2011-1006) of the School of Allied Health Sciences, Kitasato University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding received for this study.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: yuichi@med.kitasato-u.ac.jp

Introduction

Lung cancer is the leading cause of cancer death, comprising 13% (1.6 million) of the total cancer cases and 18% (1.4 million) of the cancer deaths in the world in 2008 [1,2].

Tumor markers have been detected in sera, urine, and tissues from patients with malignant tumors, and can be used for an exact diagnosis, discrimination of benign or malignant tumors, follow-up after therapies, and prediction of the patient's outcome. At present, some sero-diagnostic markers are used for lung cancer, such as carcinoembryonic antigen (CEA) and sialyl Lewis X antigen (SLX) for adenocarcinoma (AD), and cytokeratin 19 fragment (CYFRA) and squamous cell carcinoma antigen (SCCa) for squamous cell carcinoma (SCC) [3]. The positive rates of CEA, SLX, CYFRA, and SCCa are reportedly 57, 40~50, 50~60, and 60~80%, respectively. However, it has been reported that these markers do not show sufficient tumor or organ specificities; for example, SLX can show false-positive results in the presence of pulmonary tuberculosis and pulmonary fibrosis, and CYFRA can elevate with interstitial pneumonia and renal failure.

Antibodies are usually developed using purified proteins or synthetic peptides. We have exhaustively generated monoclonal antibodies (MoAbs) against various tumor-associated proteins using the pulmonary AD-derived A549 cell as an antigen with the random immunization method [4], and over 1,000 MoAbs have been obtained [5]. This method is expected to generate antibodies against proteins with tumor-specific post-translational modifications, which are difficult to obtain by conventional immunization methods.

Carbonic anhydrase XII is a transmembrane zinc metalloenzyme that catalyzes the reversible hydration of carbon dioxide to form bicarbonate ($H_2O+CO_2 \Leftrightarrow H^+ + HCO_3^-$), and is a member of the alpha carbonic anhydrase (CA) family. CAXII has been proposed to be involved in the acidification of the extracellular microenvironment, which is suitable for rapid tumor growth. CAXII overexpression was initially detected in renal cell carcinoma, and subsequent studies confirmed its expression in

various human cancers, such as diffuse astrocytoma, breast, pancreatic, and ovarian carcinoma, as well as in non-small cell lung cancer (NSCLC) [6–11]. Its expression was influenced both by factors related to differentiation and hypoxia in breast cancer *in vivo*, and was associated with a more favorable prognosis in invasive breast carcinoma patients [12]. Higher CAXII expression was also correlated with a better overall and disease-specific survival in patients with resectable NSCLC [13]. However, no study has clarified CAXII in sera and its clinical utility as a sero-diagnostic marker for patients with malignant tumors.

In this study, the specificity of the obtained anti-CAXII antibody was confirmed by immunohistochemistry (IHC) and immunoblotting with lung cancer cell lines and lung cancer tissues. To further confirm its utility as a sero-diagnostic marker, CAXII levels in sera from patients with lung cancer were studied by dot blot analysis.

Materials and Methods

1. Cell lines

The A549 and LC-2/ad cells derived from lung AD were purchased from the Japanese Cancer Research Resources Bank (Tokyo, Japan) and RIKEN BioResource Center (Ibaraki, Japan), respectively. The RERF-LC-AI cells derived from lung SCC was purchased from the RIKEN BioResource Center. The N231 cells derived from SCLC were purchased from the American Type Culture Collection (Rockville, MD, USA). LCN1, a large cell neuroendocrine carcinoma (LCNEC) line, was established in our laboratory [14]. These cells were grown in RPMI-1640 medium (SIGMA, Steinheim, Germany) supplemented with 10% fetal bovine serum (FBS; Biowest, Miami, FL, USA), 100 units/ml of penicillin, and 100 μg/ml of streptomycin (GIBCO, Auckland, New Zealand). After harvesting and washing twice with phosphate-buffered saline without divalent ions (PBS-), sub-confluent cells were stored at −80°C for proteomics analysis or fixed in 10% formalin and embedded in paraffin for immunohistochemistry. A549 cells were also AMeX-fixed [15] for immunohistochemical screening. The SP2/O-Ag14 cells derived from a mouse myeloma were purchased from the RIKEN BioResource Center, and were grown in RPMI-1640 medium supplemented with 1× 8-azaguanine (50× Hybri-Max, SIGMA), 10% FBS, penicillin, and streptomycin.

Table 1. Clinicopathological characteristics of the patients.

Characteristics		Training set (N = 70)	Validation set (N = 26)
Age	<70	40 (57.1%)	19 (73.1%)
	≥70	30 (42.9%)	7 (26.9%)
Gender	Male	52 (74.3%)	16 (61.5%)
	Female	18 (25.7%)	10 (38.5%)
Stage	I	19 (27.2%)	17 (65.4%)
	II	11 (15.7%)	2 (7.7%)
	III	26 (37.1%)	4 (15.4%)
	IV	14 (20.0%)	3 (11.5%)
Tumor differentiation	Well	7 (13.2%)	11 (52.4%)
(NSCLC)	Moderate	10 (18.9%)	5 (23.8%)
	Poor	18 (34.0%)	4 (19.0%)
	Unknown	18 (34.0%)	1 (4.8%)
Tumor size	<3 cm	24 (34.3%)	15 (57.7%)
	≥3 cm	45 (64.3%)	6 (23.1)
	Unknown	1 (1.4%)	5 (19.2)
Nodal status	N0	23 (32.9%)	18 (69.3%)
	N1	12 (17.1%)	1 (3.8%)
	N2	23 (32.9%)	5 (19.2%)
	N3	12 (17.1%)	2 (7.7%)
Distant metastasis	M0	56 (80.0%)	23 (88.5%)
	M1	14 (20.0%)	3 (11.5%)
Histological type	AD[a]	29 (41.4%)	20 (77.0%)
	SCC[b]	21 (30.0%)	0 (0.0%)
	SCLC[c]	17 (24.3%)	5 (19.2%)
	LCNEC[d]	3 (4.3%)	1 (3.8%)

[a]Adenocarcinoma.
[b]Squamous cell carcinoma.
[c]Small cell lung carcinoma.
[d]Large cell neuroendocrine carcinoma.
doi:10.1371/journal.pone.0033952.t001

2. Ethics statement

All samples were collected in accordance with the ethical guidelines and written consent mandated, and this study was approved by the Ethics Committee of Kitasato University School of Medicine. All patients and healthy controls were approached based on approved ethical guidelines, and those who agreed to participate in this study were required to sign consent forms. Patients could refuse entry and discontinue participation at any time. All participants provided written consent.

2.1. Sera. Sera from 70 patients with lung cancer (AD: 29, SCC: 21, SCLC: 17, and LCNEC: 3) and 30 healthy controls were used in the training set. In addition, a validation set consisting of sera from 26 patients with lung cancer (AD: 20, SCLC: 5, and LCNEC: 1) and 30 healthy controls was also studied. The clinicopathological characteristics of the patients data are summarized in Table 1.

Patient sera were collected at Kitasato University Hospital, and healthy control sera were provided by Kyowa Medex Co., Ltd. (Tokyo, Japan) and kept at −80°C until use.

3. Generation of monoclonal antibodies

A549 cell lysate was prepared with PBS(-) using an ultra-sonic homogenizer (UH-50; SMT Company, Tokyo, Japan). Five-week-old female BALB/c mice were immunized intra-peritoneally with 50 mg wet- weight of A549 cell lysate in 500 μl of PBS(-) 3 times with a two-week interval. The antibody titer was tested by IHC using 100-times diluted sera from the immunized mice as the first antibody on AMeX-fixed A549 cells. Three days prior to cell fusion, the animal with the highest titer was intra-peritoneally boosted by the same amount of A549 lysate. Hybridoma preparation and IHC screening with AMeX-fixed A549 cells were previously described [4,5].

4. Proteomics analysis

4.1. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Proteins were extracted from each of A549, LC-2/ad, RERF-LC-AI, N231, and LCN1 cells with detergent lysis buffer [16] using an ultra-sonic homogenizer. Ten μg each of extracted proteins were boiled and separated by SDS-PAGE with 10% polyacrylamide gel at a constant current of 20 mA. After SDS-PAGE, proteins in gels were transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore Corp., Billerica, MA, USA) for immunoblotting.

4.2. Immunoblotting. Blotting membranes were blocked with 0.5% casein from bovine milk (Sigma, St. Louis, MO, USA) for 30 min at RT. The membranes were then reacted with non-

diluted hybridoma supernatant for 1 hr at RT, followed by incubation with 1,000-times diluted horseradish peroxidase-conjugated rabbit anti-mouse IgG polyclonal antibody (Dako, Glostrup, Denmark) with 0.025% Casein for 45 min at RT. Finally, signals were developed using Immobilon Western HRP reagent (Millipore Corp.).

4.3. Determination of antibody isotype. To determine the isotype of the established KU-Lu-5 antibody, we used the IsoStrip™ Mouse Monoclonal Antibody Isotyping Kit (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions.

4.4. Immunoprecipitation. The immunoprecipitation method used in this study was previously described [17]. In brief, A549 cells were washed with PBS (-) and treated with radioimmunoprecipitation assay (RIPA) buffer containing Complete-mini EDTA-free (Roche Diagnostics) on ice for 30 min. After centrifugation at 15,000 rpm for 30 min at 4°C, the supernatant was collected and precleared with protein G sepharose (50% slurry) (GE Healthcare Bio-Sciences Corp., Piscataway, NJ, USA) at 4°C overnight. To conjugate the primary antibody, 250 µL of primary antibody (KU-Lu-5 hybridoma supernatant) and 20 µL of protein G sepharose beads suspended in RIPA buffer were incubated with mixing at 4°C overnight. After centrifugation, the antibody-sepharose conjugate and 500 µg of total cellular protein from the precleared supernatant were incubated with mixing at 4°C for 4 hrs. The immunoprecipitates were collected by centrifugation at 15,000 rpm for 5 min at 4°C. After washing four times with RIPA buffer, the supernatant was carefully removed and the pellets were resuspended in 15 µL of 1×Laemmili's buffer. Then, 15 µL of samples were boiled and separated by SDS-PAGE with 10% polyacrylamide gel. After SDS-PAGE, gels were Zn-stained with the Negative Gel Stain MS kit (Wako Pure Chemical, Tokyo, Japan) according to the manufacturer's instructions.

4.5. Identification of antigen protein. *4.5.1. In-gel digestion.* The protein spot was excised from the SDS-PAGE gel and minced to 1 mm^3, destained with destaining solution (Wako Pure Chemical), dehydrated with 100% (v/v) ACN, and dried under vacuum conditions. Tryptic digestion was performed with a minimal volume of digestion solution which contained 20 ng/µl of trypsin (Trypsin Gold, Mass Spectrometry Grade, Promega, Madison, WI, USA) and 25 mM NH_4HCO_3 for 24 hrs at 37°C. After incubation, digested protein fragments eluted in solution were collected, and gels were washed once in 5% (v/v) trifuloroacetic acid /50% (v/v) ACN and collected in the same tube.

4.5.2. Protein identification. The collected peptide fragments were analyzed using autoflex III matrix-associated laser desorption/ionization-time of flight/time of flight mass spectrometry (MALDI-TOF/TOF MS; Bruker Daltonik, Bremen, Germany). A disposable plate, spotted α-cyano-4-hydroxycinnamic acid matrix for samples, and PAC Peptide Calibstandard for calibration (Prespotted AnchorChip 96 set for Proteomics, Bruker Daltonik) were used. Peptide mass fingerprints (PMF) were measured, and then a few peaks obtained from PMF were further measured for their tandem mass spectra as parent masses. MASCOT (http://www.matrixscience.com) using the IPI Human database (93,289 sequences; 36,994,704 residues), released on 3 May, 2011 (http://www.matrixscience.com), was used to determine proteins from PMF and tandem mass data.

5. Immunoblot analysis with recombinant CAXII protein

Recombinant CAXII protein and Venus protein as a negative control with GST-tag were prepared using a wheat germ cell-free system [18]. Fourteen µg each of recombinant CAXII and Venus proteins were boiled and separated with SDS-PAGE, followed by immunoblotting with KU-Lu-5 antibody, as mentioned in 2.4.1.

6. Immunohistochemical staining

Three-µm-thick sections, made from 10% formalin-fixed and paraffin-embedded lung cancer cell lines and 37 surgically resected lung cancers (AD: 28, SCC: 9) were deparaffinized in xylene, rehydrated in a descending ethanol series, and then treated with 3% hydrogen peroxide for 20 min. After the antigen was retrieved by autoclaving in 0.01 mol/L citrate buffer (pH 6.0) with 0.1% Tween 20 at 121°C for 10 min, the sections were reacted with non-diluted KU-Lu-5 hybridoma supernatant for 16–18 hrs at room temperature (RT). After rinsing in TBS three times for 5 min each, the sections were reacted with ChemMate Envision reagent (Dako) for 30 min at RT. Finally, the sections were visualized with Stable DAB solution (Invitrogen Corp.) and counterstained with Mayer's hematoxylin.

7. Dot blot analysis

7.1. Sample preparation. *7.1.1. Removal of albumin and IgG from serum samples.* The removal of albumin and IgG from sera was performed using a ProteoExtract Albumin/IgG Removal kit (Merck, Darmstadt, Germany) according to the manufacturer's instructions. A 60-µL sample of each sera was diluted with 540 µL of binding buffer, and allowed to pass the column by gravity flow. The flow-through fraction was collected in a collection tube. To wash the column, binding buffer was allowed to pass the column by gravity flow. The flow-through fraction was collected in the same collection tube.

7.1.2. Desalting and concentration by ultrafiltration. The albumin- and IgG-depleted samples were buffer-exchanged and concentrated using 10-kDa molecular-weight cut-off ultra-filtration VIVASPIN 2 (Sartorius, Gottingen, Germany). The samples were centrifuged at 6,000×g at 4°C until less than 100 µL, and then the buffer was exchanged for PBS (-) with concentration at 6,000×g at 4°C until concentrated to less than 50 µL. The concentrated samples were adjusted to a final volume of 60 µL with PBS (-).

7.2. Dot blot analysis. One µl each of albumin- and IgG-depleted samples diluted to 1:20 with PBS(-) and mouse IgG (purified in our laboratory) for a positive control were spotted on a PVDF membrane (Millipore Corp.) using the automatic dot blot system with a 256-solid pin configuration (Kakengeneqs Inc., Chiba, Japan). Two sheets of membrane were prepared for one set of experiment. Spotted membranes were washed in TBS for 10 min, and blocked with 0.5% casein (Sigma) for 1 hr at RT. One membrane was then reacted with non-diluted KU-Lu-5 hybridoma supernatant, and the other membrane was reacted with antibody diluting solution [20-times diluted 0.5% casein with 0.1% Tween 20 added TBS (TBS-T)] for 30 min at RT. After rinsing in TBS-T 3 times for 5 min each, membranes were incubated with 1,000-times diluted horseradish peroxidase-conjugated rabbit anti-mouse IgG polyclonal antibody (Dako) for 30 min at RT. Finally, signals were developed with Immobilon Western reagent (Millipore Corp.). The data were analyzed using DotBlotChipSystem Ver. 4.0 (Dynacom Co., Ltd., Chiba, Japan). Each normalized signal was presented as the ratio of the positive intensity versus the negative background intensity.

8. Statistical analysis

Serum CAXII levels in patients with lung cancer and healthy controls were statistically analyzed using the Mann-Whitney *U*-test. Sensitivity, specificity, and predictive values were calculated with the SPBS software package (Ver. 9.42 for Windows) for each variable at a corresponding cut-off. Discriminant function analysis

was performed to classify patients in the "lung cancer" vs. "healthy control" group, according to the status of the biomarkers, using the SPBS software package. The area under the curve (AUC) and best cut-off point were calculated employing receiver operating characteristic (ROC) analysis. Results were considered significant when $P<0.05$.

Results

1. Confirmation of antibody titer in mouse sera

The antibody titer was tested by IHC with 1,000-times diluted sera of immunized mice as the first antibody on AMeX-fixed A549 cells. As a result, the sera from immunized mice contained antibodies that reacted with various components of A549 cells.

Using AMeX-fixed A549 cell preparations for the immunohistochemical screening of hybridomas, we finally established 188 MoAbs in total and a further study was performed with the KU-Lu-5 clone, which showed intense staining in A549 cells (Fig. 1 A).

2. Identification of antigen protein

In order to identify the antigen protein recognized by the KU-Lu-5 antibody, we performed IP with lysate from A549 cells. The results of IP are shown in Fig. 1 B, C. The antigenic protein was observed at roughly 40 kDa. To determine the antigenic protein recognized by KU-Lu-5 antibody, we excised and collected the spot from the Zn-stained gel, and proceeded with in-gel digestion. After analysis employing a MALDI-TOF/TOF MS and a MASCOT search, the protein was determined as isoform 2 of carbonic anhydrase XII (CAXII, accession: IPI00221392), which is composed of 343 amino acids with a predicted M.W. of 38,384 Da. The result was confirmed by immunoblot analysis with recombinant CAXII protein using KU-Lu-5 hybridoma supernatant as the first antibody (Fig. 1 D). The immunoglobulin isotype of KU-Lu-5 antibody was determined as IgG_1, κ.

3. Immunoblot analysis

Expression of CAXII was detected only in A549 cells as a roughly 40-kDa protein, and no clear band was detected in other cells used in this study (Fig. 2 A).

4. Immunohistochemical staining for CAXII

Immunohistochemically, membranous expression of CAXII was observed only in A549 cells (Fig. 2 B). Membranous staining was detected in 2 of the 28 ADs (7.1%) and in 2 of the 9 SCCs (22.2%) (Fig. 2 C, D).

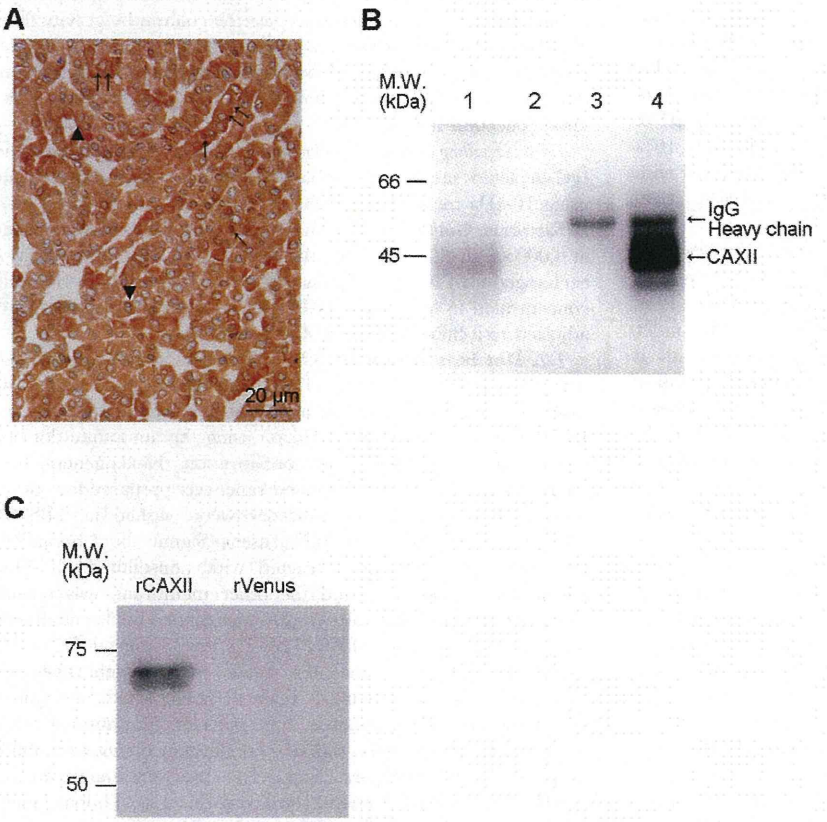


Figure 1. Production of anti-CAXII monoclonal antibody and its antigen identification. (A) The antibody titer was tested immunohistochemically using 1,000-times diluted sera of immunized mice as the first antibody on AMeX-fixed A549 cells, which were used as an immunogen. The sera of immunized mice contained antibodies that reacted with various cell components, such as the nucleus (↑), plasma membrane (▲), and cytoplasm (↑ ↑). (B) Immunoprecipitation with KU-Lu-5 antibody. Immunoblot analysis using KU-Lu-5 hybridoma supernatant as the first antibody [lane 1: A549 lysate, lane 2: A549 lysate combined with protein G, lane 3: KU-Lu-5 antibody combined with protein G, lane 4: A549 lysate combined with KU-Lu-5 antibody]. Lanes 2 to 3 are negative controls, and immunoprecipitated product with KU-Lu-5 antibody was detected in lane 4 (↑). (C) Confirmation of identified antigen protein. KU-Lu-5 antibody reacted with recombinant CAXII protein (64 kDa), but not with recombinant Venus protein.

doi:10.1371/journal.pone.0033952.g001

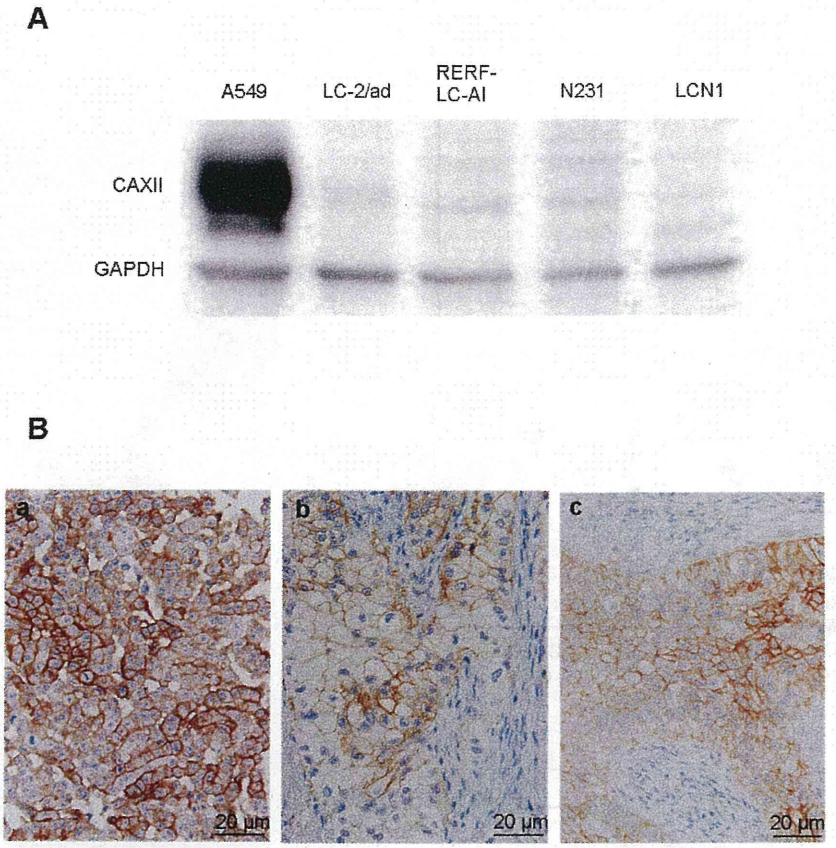


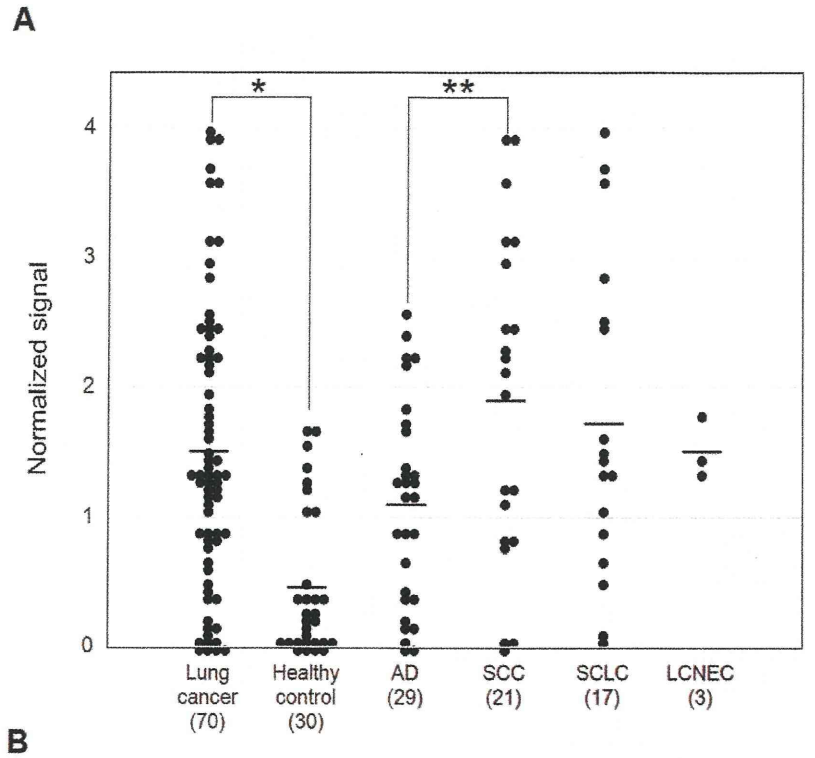
Figure 2. Expression of CAXII antibody in lung cancer cell lines and tissues. (A) Immunoblot analysis of CAXII in lung cancer cell lines. CAXII was detected as an approximately 40-kDa protein with A549 cells. (B) Immunostaining of CAXII in A549 cells (a), adenocarcinoma (b), and squamous cell carcinoma (c) of the lung, and each showed membranous staining of CAXII.
doi:10.1371/journal.pone.0033952.g002

5. Serum CAXII in patients with lung cancer

The serum CAXII levels were significantly higher in lung cancer patients than in healthy controls in the training set ($P<0.0001$). Relative values of serum CAXII levels ranged from 0.101 to 4.01 (median: 1.520) in lung cancer patients, but 0.006 to 1.679 (median: 0.290) in healthy controls (Fig. 3 A). In lung cancer, CAXII serum levels of SCC patients were significantly higher than those of AD patients ($P = 0.03$) (Fig. 3 A). The area under the ROC curve (AUC) between lung cancers and healthy controls was 0.794 (Fig. 3 B). When an optimal cut-off value of 0.387 for CAXII was applied, the diagnostic sensitivity and specificity for lung cancer were 82.9 and 70.0, respectively, and the negative and positive predictive values were 0.617 and 0.863, respectively. Furthermore, within SCCs, serum CAXII levels were significantly higher in patients with well- and moderately differentiated tumors than those with poorly differentiated ones ($P = 0.027$) (Fig. 4 A), and tended to be higher in patients with a tumor size of less than 3 cm rather than more than 3 cm ($P = 0.0538$). However, there was no difference in the smoking history of patients (Fig. 4 B). CAXII levels in stage I, II, and III ADs were 1.501, 0.704, and 1.001, respectively, and CAXII levels in stage I, II, and III SCCs were 1.764, 2.093, and 1.854, respectively. These data were summarized in Table 2. No relations between the CAXII serum levels and tumor stage or presence of metastasis were identified for either ADs or SCCs. To further confirm the utility of serum CAXII levels as a sero-diagnostic marker, 56 additional samples of sera were analyzed by dot blot analysis as a validation study. The serum CAXII levels were also significantly higher in lung cancer patients than in healthy controls in the validation set ($P = 0.030$). Relative values of serum CAXII levels ranged from 0.000 to 8.023 (median: 3.921) in lung cancer patients, but 0.000 to 8.331 (median: 2.806) in healthy controls (Fig. 5). When an optimal cut-off value of 3.086 for applied, the diagnostic sensitivity and specificity for lung cancer were 65.4 and 70.0, respectively.

Discussion

In this study, aiming to discover useful sero-diagnostic markers for lung cancer, we generated monoclonal antibodies using lung AD-derived A549 cells as antigens. From the obtained 188 antibodies, we focused on an antibody recognizing CAXII, and explored its clinical utility as a sero-diagnostic marker for lung cancer. This random immunization method is expected to yield antibodies against tumor-specific proteins with post-translational modifications, which are difficult to obtain by conventional immunization methods. Actually, several authors have reported that monoclonal antibodies generated by this method are useful as diagnostic and prognostic markers for cancers [5,17,19]. Battke *et al.* [20] established a 6A10 antibody recognizing CAXII using a similar immunization methodology. However, the obtained antibodies were limited to those only reacting with cell surface antigens because of using flow cytometry for the screening of hybridomas. In the present study, the hybridomas were immunohistochemically screened which facilitated the obtaining of



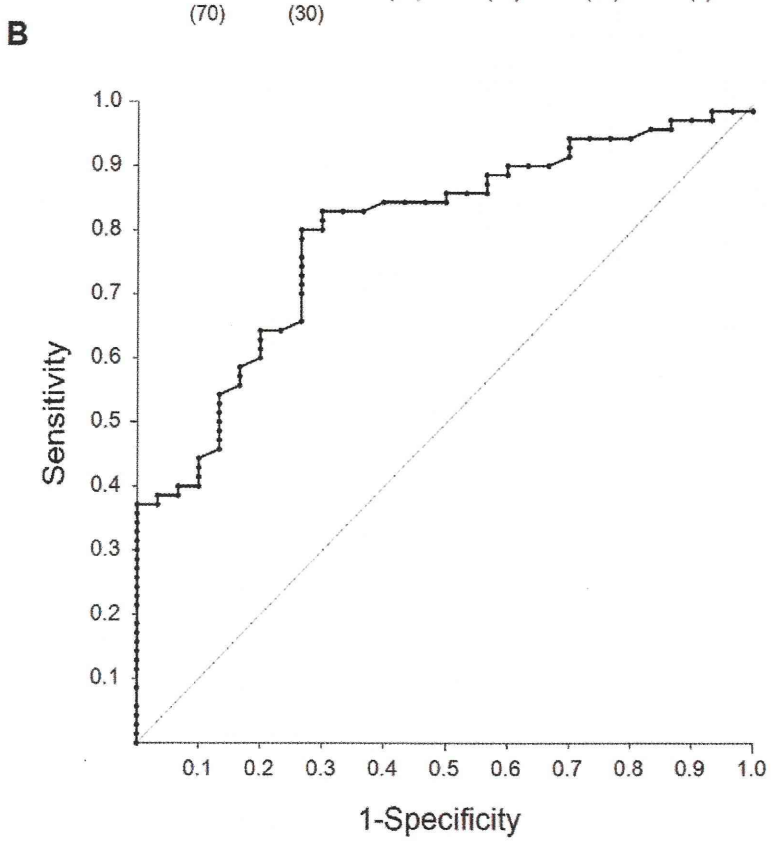


Figure 3. Serum CAXII levels in patients with lung cancer and healthy controls in the training set. Serum CAXII levels in patients with lung cancer and healthy controls. (A) The median CAXII level in the sera from healthy controls was 0.29, and that in sera from lung cancer patients was 1.52. Serum CAXII levels were significantly higher in lung cancer patients (*$P<0.001$). Furthermore, serum CAXII levels were higher in SCCs than ADs (**$P=0.0381$). (B) Receiver-operating characteristic curve analysis of CAXII as a serum marker for lung cancer. The corresponding areas under the curves were 0.794 for CAXII. With a 70.0% specificity, the sensitivity of CAXII for lung cancer was 82.9%, at a cut-off value corresponding to 0.387. doi:10.1371/journal.pone.0033952.g003

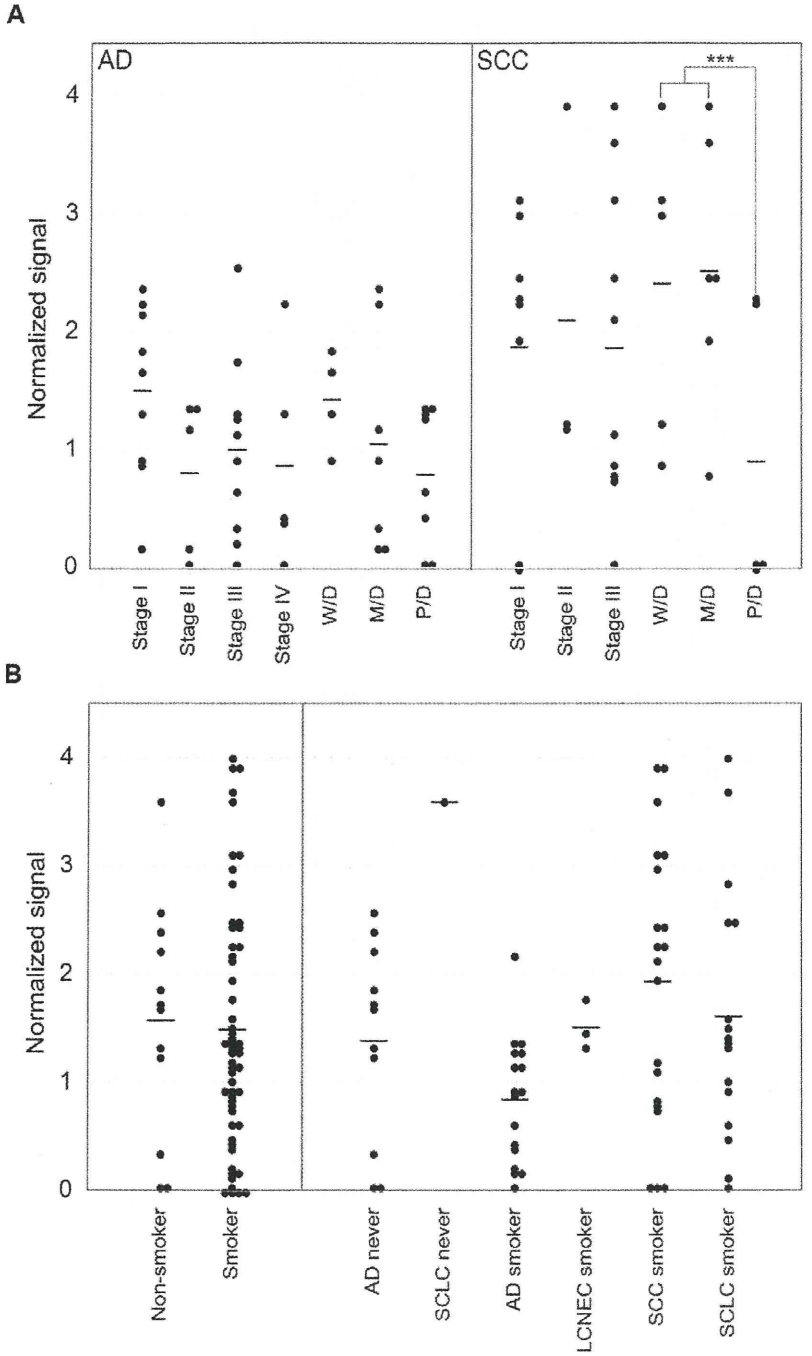


Figure 4. Correlation between serum CAXII levels and patients' clinicopathological characteristics. (A) CAXII levels in sera from patients with ADs and SCCs with a focus on the stage and differentiation. In SCCs, CAXII levels were significantly higher in well- and moderately differentiated tumors than in poorly differentiated ones (***P = 0.0272). In ADs, no significant difference based on the differentiation extent was detected. (B) Smoking history in lung cancer patients. The median CAXII level in the sera from non-smokers was 1.56, and that in smokers was 1.54, showing no significant difference.
doi:10.1371/journal.pone.0033952.g004

antibodies reacting with tumor-associated proteins localized in several intra-cellular compartments. The CAs constitute a family of ubiquitous enzymes with important roles in many physiological and pathological processes which reversely catalyse the conversion of CO_2+H_2O to HCO_3^- and H^+, contributing to regulation of the intracellular pH [6–11,19]. Several clinical studies have shown a clear relationship between high CAXII expression levels in tumor cells and a favorable prognosis.

Table 2. Serum CAXII levels in ADs and SCCs.

		Average value (AD)	Average value (SCC)
Stage	I	1.501	1.764
	II	0.704	2.093
	III	1.001	1.854
	IV	0.654	0.000
Tumor differentiation	Well	1.424	2.403
	Moderate	1.046	2.511
	Poor	0.727	0.742

doi:10.1371/journal.pone.0033952.t002

Watson *et al.* [12] reported that CAXII was expressed in 75% of invasive breast carcinoma cases, and was significantly associated with a lower histological grade (P = 0.001), positive estrogen receptor status (P<0.01), and negative epidermal growth factor receptor overexpression (P<0.001). Using univariate analysis, CAXII-positive tumors were associated with a lower relapse rate (P = 0.04) and a better OS (P = 0.01). On the other hand, although 98% of astrocytomas were immunohistochemically positive for CAXII, higher CAXII expression levels were correlated with a higher histological grade and a poorer patient outcome either by univariate (P = 0.010) or multivariate (P = 0.039) survival analysis [9].

An immunohistochemical study of the expression of CAXII in lung cancer was reported by Ilie *et al.* [13]. CAXII overexpression was observed in 105/555 cases, and was significantly associated with a better differentiation (P = 0.015) and SCC histological type (P<0.001). Furthermore, high CAXII expression was also significantly correlated with better overall and disease-specific survival. From our results, CAXII levels were higher in sera from SCC patients than ADs (Fig. 3 A). Also, they correlated more favorably with differentiation (Fig. 4 A). Integrating these results, CAXII may not only be a candidate tissue marker, but also a sero-diagnostic marker for lung cancer.

Although the association of CAXII expression and clinicopathologic factors and patient outcome in different tumors has been reported, to our knowledge, no study concerning the serum CAXII protein levels or its autoantibody levels in patients with tumors has been reported.

To confirm the possibility of CAXII as a sero-diagnostic marker, we measured its serum levels in patients with lung cancer and healthy controls. We demonstrated that the CAXII protein flowed out into the sera and its levels in patients with lung cancer were significantly higher than in healthy controls in both the training set (P<0.0001) and validation set (P = 0.030). It is possible that the gap in the P-value between the training set and validation set is caused by the fact that serum levels of CAXII of SCC patients were generally higher than those of patients with other

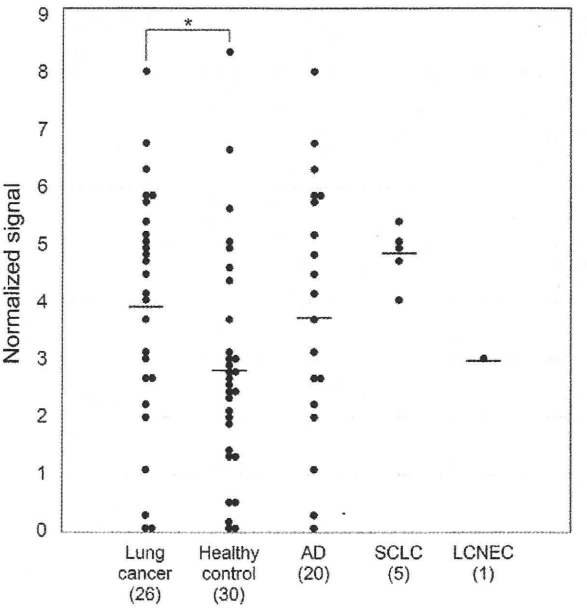


Figure 5. Serum CAXII levels in patients with lung cancer and healthy controls in the validation set. To confirm the utility of serum CAXII levels as a sero-diagnostic marker, 56 additional sera were analyzed by dot blot analysis as a validation study. The serum CAXII levels were also significantly higher in lung cancer patients than in healthy controls (P = 0.030). Relative values of serum CAXII levels ranged from 0.000 to 8.023 (median: 3.921) in lung cancer patients, but 0.000 to 8.331 (median: 2.806) in healthy controls.
doi:10.1371/journal.pone.0033952.g005

histologies, and the validation set included no SCC case. Taken together, the serum CAXII levels should be applicable markers discriminating lung cancer patients from healthy controls. Currently, CT scan or chest X-ray is the main method of lung cancer screening [21]. Mazzone et al. suggested that blood and breath tests should also be included for lung cancer screening, because they are both easy to perform and free of risks related to test administration [21,22]. In this study, we analyzed CAXII levels in sera from lung cancer patients and healthy controls using monoclonal antibody, and our results suggested that the serum CAXII level was a useful sero-diagnostic marker for lung cancer.

Author Contributions

Conceived and designed the experiments: MK TM Y. Sato. Performed the experiments: MK KY RN TM. Analyzed the data: MK TM SR. Contributed reagents/materials/analysis tools: SR YK NG SXJ MS AI Y. Satoh NM. Wrote the paper: MK TM RN Y. Sato.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, et al. (2011) Global Cancer Statistics. CA Cancer J Clin 61: 69–90.
2. Jemal A, Siegel R, Xu J, Ward E (2010) Cancer statistics. CA Cancer J Clin 60: 277–300.
3. Patel JL, Erickson JA, Roberts WI, Grenache DG (2010) Performance characteristics of an automated assay for the quantitation of CYFRA21-1 in human serum. Clin Biochem 43: 1449–1452.
4. Hirohashi S, Watanabe M, Shimosato Y, Sekine T (1984) Monoclonal antibody reactive with the sialyl-sugar residue of a high molecular weight glycoprotein in sera of cancer patients. Gann 75: 485–488.
5. Nagashio R, Sato Y, Matsumoto T, Kageyama T, Satoh Y, et al. (2010) Expression of RACK1 is a novel biomarker in pulmonary adenocarcinomas. Lung Cancer 69: 54–59.
6. Kivela A, Parkkila S, Saarnio J, Karttunen TJ, Kivela J, et al. (2000) Expression of a novel transmembrane carbonic anhydrase isozyme XII in normal human gut and colorectal tumors. Am J Pathol 156: 577–584.
7. Kivela AJ, Parkkila S, Saarnio J, Karttunen TJ, Kivela J, et al. (2005) Expression of von Hippel-Lindau tumor suppressor and tumor-associated carbonic anhydrases IX and XII in normal and neoplastic colorectal mucosa. World J Gastroenterol 11: 2616–2625.
8. Wykoff CC, Beasley N, Watson PH, Campo L, Chia SK, et al. (2001) Expression of the hypoxia-inducible and tumorassociated carbonic anhydrases in ductal carcinoma in situ of the breast. Am J Pathol 158: 1011–1019.
9. Haapasalo J, Hilvo M, Nordfors K, Haapasalo H, Parkkila S, et al. (2008) Identification of an alternatively spliced isoform of carbonic anhydrase XII in diffusely infiltrating astrocytic gliomas. Neuro Oncol 10: 131–138.

10. Hynninen P, Vaskivuo L, Saarnio J, Haapasalo H, Kivela J, et al. (2006) Expression of transmembrane carbonic anhydrases IX and XII in ovarian tumours. Histopathology 49: 594–602.
11. Parkkila S, Parkkila AK, Saarnio J, Kivela J, Karttunen TJ, et al. (2000) Expression of the membrane-associated carbonic anhydrase isozyme XII in the human kidney and renal tumors. J Histochem Cytochem 48: 1601–1608.
12. Watson PH, Chia SK, Wykoff CC, Han C, Leek RD, et al. (2003) Carbonic anhydrase XII is a marker of good prognosis in invasive breast carcinoma. Br J Cancer 88: 1065–1070.
13. Ilie MI, Hofman V, Ortholan C, Ammadi RE, Bonnetaud C, et al. (2010) Overexpression of carbonic anhydrase XII in tissues from resectable non-small cell lung cancers is a biomarker of good prognosis. Int J Cancer 128: 1614–1623.
14. Jiang SX, Kameya T, Asamura H, Umezawa A, Sato Y, et al. (2004) hASH1 expression is closely correlated with endocrine phenotype and differentiation extent in pulmonary neuroendocrine tumors. Mod Pathol 17: 222–229.
15. Sato Y, Mukai K, Watanbe S, Goto M, Shimosato Y (1986) The AMeX method. A simplified technique of tissue processing and paraffin embedding with improved preservation of antigens for immunostaining. Am J pathol 125: 431–435.
16. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685.
17. Nitori N, Ino Y, Nakanishi Y, Yamada T, Honda K, et al. (2005) Prognostic significance of tissue factor in pancreatic ductal adenocarcinoma. Clin Cancer Res 11: 2531–2539.
18. Goshima N, Kawamura Y, Fukumoto A, Miura A, Honma R, et al. (2008) Human protein factory for converting the transcriptome into an in vitro-expressed proteome. Nature Methods 5: 1011–1017.
19. Akishima-Fukusawa Y, Ino Y, Nakanishi Y, Miura A, Moriya Y, et al. (2010) Significance of PGP9.5 expression in cancer-associated fibroblasts for prognosis of colorectal carcinoma. Am J Clin Pathol 134: 71–79.
20. Battke C, Kremmer E, Mysliwietz J, Gondi G, Dumitru C, et al. (2011) Generation and characterization of the first inhibitory antibody targeting tumor-associated carbonic anhydrase XII. Cancer Immunol Immunother 60: 649–658.
21. Mazzone PJ (2010) Lung cancer screening: an update, discussion, and look ahead. Curr Oncol Rep 12: 226–234.
22. Mazzone PJ, Mekbail T (2007) Lung cancer screening. Curr Oncol Rep 9: 265–274.