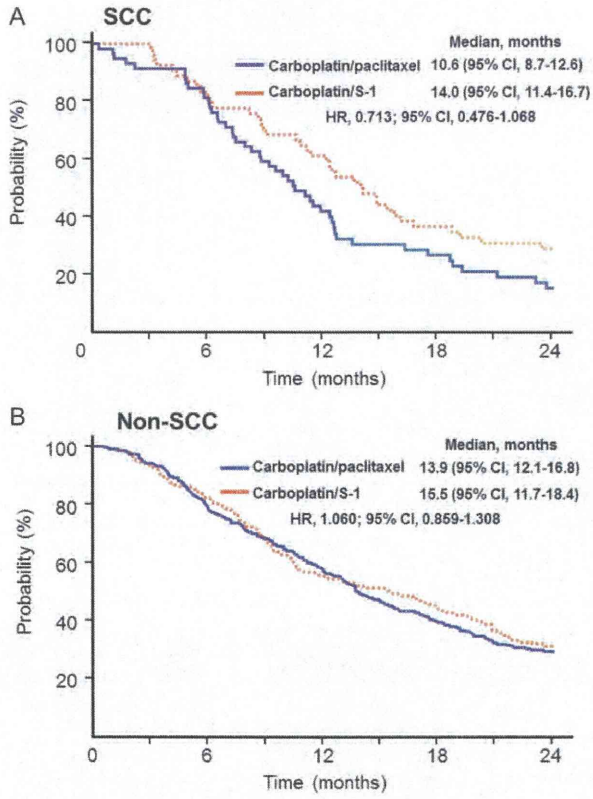

Figure 1. Kaplan–Meier curves for OS according to histological subtype of NSCLC. (A) SCC and (B) Non-SCC.

patients with non-SCC, the corresponding adjusted mean scores were 21.1 and 21.5 for carboplatin–S-1 and 21.3 and 21.3 for carboplatin–paclitaxel ($P = 0.702$). FACT/GOG-Ntx scores differed significantly between treatment arms regardless of histology. For SCC, the adjusted means were 41.1 and 41.5 at 6 and 9 weeks, respectively, for carboplatin–S-1 and 36.9 and 35.4 for carboplatin–paclitaxel ($P < 0.001$). For non-SCC, the adjusted means were 41.2 and 40.9 for carboplatin–S-1 and 38.6 and 37.6 for carboplatin–paclitaxel ($P < 0.001$).

post-study treatment based on histology

There were no major differences in post-study treatment between the two arms regardless of histological subtype (Table 4). The percentage of patients with SCC who received docetaxel as second-line treatment, however, was significantly higher for the carboplatin–S-1 arm than for the carboplatin–paclitaxel arm (58.2% versus 30.5%; $P = 0.003$, chi-square test).

discussion

The present updated analysis confirmed the noninferiority of carboplatin and S-1 compared with carboplatin and paclitaxel for the treatment of advanced NSCLC in terms of OS after completion of 2 years of follow-up and the occurrence of an adequate number of events, as planned in the original protocol. First-line treatment with carboplatin and S-1 showed a

Table 3. Treatment-related adverse events according to histological subtype of NSCLC

Event	Squamous						Nonsquamous					
	CBDCA/S-1 (*N* = 55)			CBDCA/PTX (*N* = 59)			CBDCA/S-1 (*N* = 224)			CBDCA/PTX (*N* = 221)		
	All	G3	G4	All	G3	G4	All	G3	G4	All	G3	G4
Hematologic (%)												
Leukopenia	55	2	0	85	24	7	55	6	1	86	31	2
Neutropenia	56	18	6	85	19	49	59	18	2	91	35	43
Anemia	96	13	6	85	19	3	84	16	3	82	13	2
Thrombocytopenia	91	27	16	76	12	3	86	17	13	59	6	2
Nonhematologic (%)												
Febrile neutropenia	4	4	0	19	17	2	1	1	0	4	4	0
Nausea	64	2	0	44	2	0	62	2	0	50	2	0
Vomiting	38	0	0	24	0	0	33	2	0	24	1	0
Diarrhea	40	2	0	17	0	0	31	4	0	22	1	0
Neuropathy: sensory	16	0	0	81	5	0	16	1	0	81	3	0
Arthralgia	9	0	0	59	0	0	8	0	0	69	3	0
Alopecia	11	0	0	73	0	0	9	0	0	78	0	0

favorable risk-benefit profile regardless of NSCLC histology compared with carboplatin and paclitaxel. As a first-line treatment of patients with SCC, carboplatin and S-1 showed a tendency to improve OS, with a 3.4-month increase in median OS, compared with carboplatin and paclitaxel (14.0 versus 10.6 months; HR 0.713; 95% CI 0.476–1.068). This outcome is of particular interest because of the limited therapeutic options for this patient population compared with patients with non-SCC. The current National Comprehensive Cancer Network (NCCN) guidelines highlight only cisplatin–gemcitabine and cisplatin–cetuximab–vinorelbine as treatment options for recurrence and distant metastases in patients with SCC [2, 16, 17]. Treatment of patients with SCC with gemcitabine–cisplatin versus pemetrexed–cisplatin yielded a median OS of 10.8 versus 9.4 months [2]. In the First-Line Erbitux in Lung Cancer (FLEX) trial, cetuximab–platinum-based chemotherapy was associated with a longer median OS in patients with SCC (10.2 versus 8.9 months) compared with chemotherapy alone [17]. The survival results for SCC patients treated with carboplatin and paclitaxel in our phase III trial are thus similar to those of recent previous studies. In this regard, given the historical context of NSCLC studies focusing on SCC, the survival advantage observed with carboplatin and S-1 in SCC patients is promising and warrants the performance of additional phase III studies for confirmation.

It is unclear whether the possible survival benefit conferred by carboplatin and S-1 in SCC patients is due to an intrinsic superiority of this drug combination compared with carboplatin and paclitaxel, to a reduced toxicity, or to other factors. Carboplatin–S-1 was as effective as carboplatin–paclitaxel in terms of response rate and PFS in patients with SCC. For such patients, carboplatin–S-1 was associated with a significantly lower rate of febrile neutropenia compared with carboplatin–paclitaxel (4% versus 19%, respectively; $P = 0.017$, chi-square test) as well as with a lower rate of neuropathy. SCC patients in the carboplatin–S-1 arm received docetaxel more

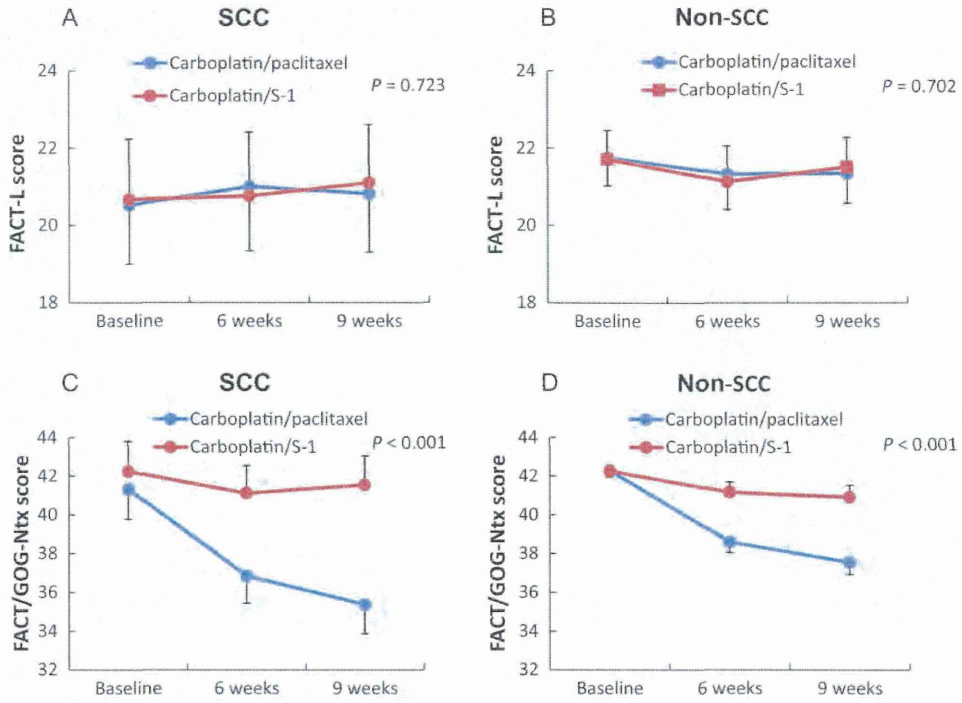

Figure 2. QOL assessments according to histological subtype of NSCLC. Assessments were carried out with the seven-item FACT-L (A and B) and 11-item FACT/GOG-Ntx (C and D) subscales for patients with SCC (A and C) or with non-SCC (B and D). Data are presented as least-square means and 95% CIs. Higher scores indicate a better QOL. *P* values were determined by analysis of variance.

Table 4. Post-treatment rate according to histological subtype of NSCLC

	Squamous			Nonsquamous		
	CBDCA–S-1 (*N* = 55)	CBDCA–PTX (*N* = 59)	*P*	CBDCA–S-1 (*N* = 227)	CBDCA–PTX (*N* = 223)	*P*
Second-line, *N* (%)	43 (78.2)	39 (66.1)	0.15	168 (74.0)	156 (70.0)	0.34
Docetaxel, *N* (%)	32 (58.2)	18 (30.5)	0.003	107 (47.1)	99 (44.4)	0.56
EGFR-TKI, *N* (%)	7 (12.7)	6 (10.2)	0.67	122 (53.7)	102 (45.7)	0.09

P values were determined by the chi-square test.

frequently as a second-line treatment than did those in the carboplatin–paclitaxel arm (58.2% versus 30.5%, respectively, $P = 0.003$), possibly because the former patients were in better condition as a result of a better tolerated first-line regimen. The reduced toxicity of carboplatin–S-1, especially with regard to neuropathy and neutropenia, may thus have allowed for more frequent application of second-line treatment with docetaxel, which has been shown to improve survival over best supportive care for the second-line setting in phase III trials [18]. Kaplan–Meier survival curves for the patients with SCC began to diverge shortly after the end of the study treatment, suggesting that the higher percentage of active second-line treatment in the carboplatin–S-1 arm of the SCC cohort may have contributed to the improved survival outcome. Given the increasing number of active drugs available for second-line treatment, subsequent therapies instituted after disease progression can have a substantial impact on OS in advanced NSCLC [19]. If multiple drugs with no large differences in effectiveness are indicated for NSCLC, treatment strategies should take into account the overall treatment plan envisioned for a given patient, including second-line and subsequent therapies as well as first-line chemotherapy.

In conclusion, we have presented the results of updated survival analysis and subgroup analysis by histology for the first phase III study of the combination of carboplatin and S-1 for the treatment of chemotherapy-naïve patients with advanced NSCLC. This regimen is therapeutically beneficial and well tolerated in such patients with either SCC or non-SCC histology. Given its efficacy and favorable toxicity profile, the combination of carboplatin and S-1 is a feasible platinum-based option to which molecularly targeted agents can be added. We are currently conducting a phase II trial of carboplatin and S-1 in combination with bevacizumab for patients with previously untreated advanced non-SCC NSCLC [20]. Furthermore, on the basis of the promising results showing a survival advantage for

SCC patients, carboplatin and S-1 should be considered among first-line treatment options for NSCLC patients with SCC.

acknowledgements

We thank all participating sites and investigators for their support with the clinical study.

funding

This work was supported by West Japan Oncology Group.

disclosure

The authors have declared no conflicts of interest.

references

1. Jemal A, Siegel R, Ward E et al. Cancer statistics, 2009. CA Cancer J Clin 2009; 59: 225–249.
2. Scagliotti GV, Parikh P, von Pawel J et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naïve patients with advanced-stage non-small-cell lung cancer. J Clin Oncol 2008; 26: 3543–3551.
3. Ferrara N, Hillan KJ, Gerber HP et al. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. Nat Rev Drug Discov 2004; 3: 391–400.
4. Johnson DH, Fehrenbacher L, Novotny WF et al. Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. J Clin Oncol 2004; 22: 2184–2191.
5. Sandler A, Gray R, Perry MC et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. N Engl J Med 2006; 355: 2542–2550.
6. Mitsudomi T, Morita S, Yatabe Y et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomized phase 3 trial. Lancet Oncol 2010; 11: 121–128.
7. Maemondo M, Inoue A, Kobayashi K et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. N Engl J Med 2010; 362: 2380–2388.
8. Zhou C, Wu YL, Chen G et al. Erlotinib versus chemotherapy as first line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011; 12: 735–742.
9. Rosell R, Carcereny E, Gervais R et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012; 13: 239–246.
10. Kwak EL, Bang YJ, Camidge R et al. Anaplastic lymphoma kinase inhibitor in non-small-cell lung cancer. N Engl J Med 2010; 363: 1693–1703.
11. Shirasaka T, Nakano K, Takechi T et al. Antitumor activity of 1 M tegafur-0.4 M 5-chloro-2,4-dihydroxypyridine-1 M potassium oxonate (S-1) against human colon carcinoma orthotopically implanted into nude rats. Cancer Res 1996; 56: 2602–2606.
12. Okamoto I, Fukuoka M. S-1: a new oral fluoropyrimidine in the treatment of patients with advanced non-small-cell lung cancer. Clin Lung Cancer 2009; 10: 290–294.
13. Okamoto I, Yoshioka H, Morita S et al. Phase III trial comparing oral S-1 plus carboplatin with paclitaxel plus carboplatin in chemotherapy-naïve patients with advanced non-small-cell lung cancer: results of a West Japan Oncology Group study. J Clin Oncol 2010; 28: 5240–5246.
14. Cella DF, Bonomi AE, Lloyd SR et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Lung (FACT-L) quality of life instrument. Lung Cancer 1995; 12: 199–220.
15. Cella D, Peterman A, Hudgens S et al. Measuring the side effects of taxane therapy in oncology: the functional assessment of cancer therapy-taxane (FACT-taxane). Cancer 2003; 98: 822–831.
16. National Comprehensive Cancer Network Inc. NCCN Practice Guidelines in Oncology; Non-Small Cell Lung Cancer-v.3., 2012; http://www.nccn.org/professionals/physician_gls/PDF/nscl.pdf (20 August 2012, data last accessed).
17. Pirker R, Pereira JR, Szczesna A et al. Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. Lancet 2009; 373: 1525–1531.
18. Shepherd FA, Pereira JR, Ciuleanu T et al. Erlotinib in previously treated non-small-cell lung cancer. New Engl J Med 2005; 353: 123–132.
19. Hayashi H, Okamoto I, Morita S et al. Postprogression survival for first-line chemotherapy of patients with advanced non-small-cell lung cancer. Ann Oncol 2012; 23: 1537–1541.
20. Kaneda H, Urata Y, Okamoto I et al. Oral S-1 and carboplatin plus bevacizumab with maintenance S-1 and bevacizumab for chemo-naïve patients with advanced non-squamous non-small cell lung cancer (NSCLC). J Clin Oncol 2012; 30: e18060.

OPEN ACCESS Freely available online

PLOS ONE

A Microarray-Based Gene Expression Analysis to Identify Diagnostic Biomarkers for Unknown Primary Cancer

Issei Kurahashi[1], Yoshihiko Fujita[2], Tokuzo Arao[2], Takayasu Kurata[3], Yasuhiro Koh[4], Kazuko Sakai[2], Koji Matsumoto[5], Maki Tanioka[5], Koji Takeda[6], Yuichi Takiguchi[7], Nobuyuki Yamamoto[8], Asuka Tsuya[6], Nobuaki Matsubara[9], Hirofumi Mukai[9], Hironobu Minami[10], Naoko Chayahara[10], Yasuhiro Yamanaka[11], Keisuke Miwa[12], Shin Takahashi[13], Shunji Takahashi[14], Kazuhiko Nakagawa[3], Kazuto Nishio[2]*

1 Department of Planning, Information, and Management, University of Tokyo, Tokyo, Japan, **2** Department of Genome Biology, Kinki University School of Medicine, Osaka-Sayama, Japan, **3** Department of Medical Oncology, Kinki University School of Medicine, Osaka-Sayama, Japan, **4** Division of Drug Discovery and Development, Shizuoka Cancer Center Research Institute, Shizuoka, Japan, **5** Medical Oncology Division, Hyogo Cancer Center, Akashi, Japan, **6** Department of Clinical Oncology, Osaka City General Hospital, Osaka, Japan, **7** Department of Medical Oncology, Graduate School of Medicine, Chiba University, Chiba, Japan, **8** Division of Thoracic Oncology, Shizuoka Cancer Center, Shizuoka, Japan, **9** Division of Oncology and Hematology, National Cancer Center Hospital East, Chiba, Japan, **10** Division of Medical Oncology/Hematology, Kobe University Graduate School of Medicine, Kobe, Japan, **11** Department of Medical Oncology, Tochigi Cancer Center, Utsunomiya, Japan, **12** Department of Medical Oncology, International Medical Center-Comprehensive Cancer Center, Saitama Medical University, Saitama, Japan, **13** Department of Clinical Oncology, Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan, **14** Division of Medical Oncology, Cancer Institute Hospital, Tokyo, Japan

Abstract

Background: The biological basis for cancer of unknown primary (CUP) at the molecular level remains largely unknown, with no evidence of whether a common biological entity exists. Here, we assessed the possibility of identifying a common diagnostic biomarker for CUP using a microarray gene expression analysis.

Methods: Tumor mRNA samples from 60 patients with CUP were analyzed using the Affymetrix U133A Plus 2.0 GeneChip and were normalized by asinh (hyperbolic arc sine) transformation to construct a mean gene-expression profile specific to CUP. A gene-expression profile specific to non-CUP group was constructed using publicly available raw microarray datasets. The t-tests were performed to compare the CUP with non-CUP groups and the top 59 CUP specific genes with the highest fold change were selected (p-value$<$0.001).

Results: Among the 44 genes that were up-regulated in the CUP group, 6 genes for ribosomal proteins were identified. Two of these genes (*RPS7* and *RPL11*) are known to be involved in the Mdm2–p53 pathway. We also identified several genes related to metastasis and apoptosis, suggesting a biological attribute of CUP.

Conclusions: The protein products of the up-regulated and down-regulated genes identified in this study may be clinically useful as unique biomarkers for CUP.

Citation: Kurahashi I, Fujita Y, Arao T, Kurata T, Koh Y, et al. (2013) A Microarray-Based Gene Expression Analysis to Identify Diagnostic Biomarkers for Unknown Primary Cancer. PLoS ONE 8(5): e63249. doi:10.1371/journal.pone.0063249

Editor: Ming Tat Ling, Queensland University of Technology, Australia

Received November 27, 2012; **Accepted** April 1, 2013; **Published** May 9, 2013

Funding: This work was supported by the grant-in-aid for Scientific Research from the Ministry of Health, Labour and Welfare. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: knishio@med.kindai.ac.jp

Introduction

Patients with cancer of unknown primary (CUP) present with metastatic disease for which the primary site cannot be found, despite extensive standard investigation. The prognosis of patients with CUP is usually poor for those receiving empiric treatments. The median survival period is 3–9 months even when newer combination treatment regimens are administered [1–5]. The survival of patients with CUP can be improved if the primary site can be identified and a site-specific therapy can be applied [6,7].

Clinically, CUPs exhibit common characteristics, such as rapid progression, early dissemination and a silent primary tumor, with signs and symptoms related to the metastatic site(s) [8]. The primary tumor may either have a slow growth pattern or may become involuted and undetectable. Existence of such common properties prompts us to hypothesize that there may be potential biological markers that elucidate CUP as a whole. Gene expression analysis is one of the means by which to identify genes characteristic to CUP.

Several studies using gene expression microarrays have demonstrated that the expression levels of thousands of genes can be used as a "molecular fingerprint" to classify a multitude of tumor types [9–15]. We are presently involved in a multicenter clinical study to predict the primary site of CUP based on the analysis of gene expression patterns. The analysis interprets the expression of ~22,000 genes in each specimen by applying normalization and classification algorithms to gene expression data from a microarray. The similarity of each tumor specimen's gene expression

pattern is then compared to the patterns for tumors from 24 known primary sites covered by the test. This study enabled the identification of genes that exhibited a unique expression pattern in CUP. Here, we present several genes encoding metastasis-and apoptosis-related proteins thus identified that may biologically characterize CUP.

Materials and Methods

Ethic Statement

All the patients provided written informed consent. Study approval was obtained from independent ethics committees of Kinki University, Shizuoka Cancer Center, Hyogo Cancer Center, Osaka City General Hospital, Chiba University, National Cancer Center Hospital East, Kobe University, Tochigi Cancer Center, Saitama Medical University, Tohoku University, and Cancer Institute Hospital. The study was undertaken in accordance with the Declaration of Helsinki.

Study Design

This study originated from currently ongoing multicenter, randomized, phase 2 prospective trial for the treatment of untreated CUP based on prediction of the primary site using data from a DNA chip. The patients had been diagnosed as having CUP between November 2008 and November 2010 at one of 13 centers of the West Japan Oncology Group (WJOG), a Japanese non-profit organization for conducting oncological clinical trials. The laboratory analyses were performed at 2 centers in Japan (Kinki University, Osaka-Sayama and Mitsubishi Chemical Medience Corporation, Tokyo).

Patients

All eligible patients had undergone a standard investigation for CUP. They were categorized into unfavorable subsets of CUP. Diagnoses of histologically or cytologically confirmed adenocarcinoma, poorly differentiated carcinoma, or squamous cell carcinoma were permitted. In each of the patients, a primary site had not been identified after a complete medical history, physical examination, chemistry profile, computed tomography (CT) scan of the chest, abdomen, and pelvis, mammography in women, measurements of the prostate-specific antigen (PSA) level in men, and a directed workup of any symptomatic areas. Patients in the following categories were excluded: women with adenocarcinoma involving only the axillary lymph nodes or the peritoneal cavity, patients with squamous cell carcinoma involving only cervical lymph nodes or inguinal lymph nodes, patients with poorly differentiated carcinoma consistent with a germ cell tumor (isolated midline structures, multiple pulmonary nodules, or elevated levels of β-human chorionic gonadotropin or α-human chorionic gonadotropin-fetoprotein), men with an elevated plasma PSA level or PSA-positive staining in a tumor, patients with a single, small, potentially resectable tumor, and patients with neuroendocrine carcinomas.

Sample Collection

Fresh frozen samples obtained from 60 patients with CUP were used for the analysis. All the samples were tested without knowledge of either the clinical characteristics or the subsequent response to treatment, except for the sex of the patient and the site of biopsy (mostly lymph nodes or ascites fluid).

Assay Procedure

RNA was extracted from the samples using an Isogene kit (Nippon Gene, Toyama, Japan). Spectrophotometry was used to assess whether an adequate total RNA concentration and purity was present. In general, the protocol for processing the RNA, amplifying and labeling fragments, hybridizing material on the microarray, and scanning was similar to the standard Affymetrix protocol for GeneChip® expression analysis. Affymetrix GeneChip® Human Genome U133 Plus 2.0 was used on an Affymetrix 3000 or 3000Dx GeneChip instrument (fluidics station and scanner) running Gene-Chip operating software to generate gene expression data (.CEL files).

Database Submission of Microarray Data

The microarray data were deposited in the Gene Expression Omnibus (GEO) database: http://www.ncbi.nlm.nih.gov/geo/. The GEO accession number for the platform is GSE42392, samples GSM1038716-GSM 1038775.

Data Analysis

All the microarray data were normalized using asinh (hyperbolic arc sine) transformation, which is a modified version of Huber's normalization with variance stabilization [16,17], and also a part of generalized log transformation (glog) [18]. Interinstitutional and array-to-array biases were corrected by subtracting their specific effects that were estimated by the mixed model [19]. The equation for asinh transformation is Igk/I.k, where I represents the expression value, g represents the gene, k represents the array, and the dot indicates the mean. The resulting asinh-transformed values, representing the relative expression of each gene, were used in further analyses.

The raw microarray datasets for 2,364 cancers of several primary types and 10 normal lymph nodes were obtained from the Gene Expression Omnibus (GEO) (Table 1). These datasets were normalized and used to construct gene-expression profiles specific to each type of cancer (n = 24) as well as an overall profile for cancer with known primary (CKP). The normal lymph node dataset was used as a reference. The data quality of CUP samples was monitored to ensure that data analysis of CUP samples was comparable to that of samples of CKP collected from GEO. Only the samples whose GAPDH, a housekeeping control gene, at 5′-terminal region (AFFX-HUMGAPDH/M33197_5_at) showed a minimum expression>500, and with the ratio of expression intensity (GAPDH at the 3′-region/5′-region)<3 were chosen.

The gene-expression profile specific to CUP was constructed using 30 CUP samples as training data and another 30 samples as test data (odd and even numbered cases, respectively). Of the 22,215 genes that were measured using both CUP samples (this work) and CKP samples (publicly accessed), a total of 5,645 genes with a present call for every sample were selected for further analysis. To identify CUP specific genes, the gene-expression profiles specific to CUP (training datasets) and normal lymph node were compared using t-tests. A histogram of the p-values is shown in Figure 1. The p-values for most of the genes were less than 0.001; when we selected the top 100 genes according to their p-values, the false discovery rate (FDR) was 4.56×10^{-12} [20]. To validate whether the genes identified using the CUP training datasets were significantly specific to CUP, the linear discriminant analysis (LDA) using these genes was performed for the CUP test datasets and the accuracy was estimated as described [21]. Heatmaps and a cluster dendrogram were then constructed using the Ward method [22].

Table 1. Number of cases for each cancer type and GEO series used for gene expression profiles.

Cancer type	n	GEO Series
Bladder	80	GSE2109, GSE3167, GSE7476
Brain	106	GSE2109, GSE3185, GSE4271
Breast (Basal)	25	GSE1456
Breast (ERBB2)	15	GSE1456
Breast (Inflammatory)	49	GSE1456
Breast (Luminal A)	39	GSE1456
Breast (Luminal B)	23	GSE1456
breast (No subtype)	20	GSE1456
Breast (Normal-like)	37	GSE1456
Cervical	89	GSE2109, GSE5787, GSE6791
Colon	365	GSE2109, GSE2509, GSE2742, GSE5486, MEXP101, MEXP170
Corpus_uteri	205	GSE2109
Gallbladder	2	GSE2109
Germ cell	101	GSE3218
Head (oral squamous cell)	42	GSE6791
Kidney	270	GSE2109, GSE6357, GSE781
Liver	13	GSE2109
Lung adenocarcinoma	61	GSE4127, MEXP231
Lymphoma	18	GSE2109, GSE4176
Ovarian	420	GSE2109, GSE3149
Panreas	56	GSE2109
Prostate	229	GSE2109, GSE3325, GSE7930, GSE8218
Stomach	42	GSE2109
Thyroid	57	GSE2109, MEXP97
Normal lymph node	10	GSE2665
CUP (This work)	60	GSE42392
Total	**2434**	

doi:10.1371/journal.pone.0063249.t001

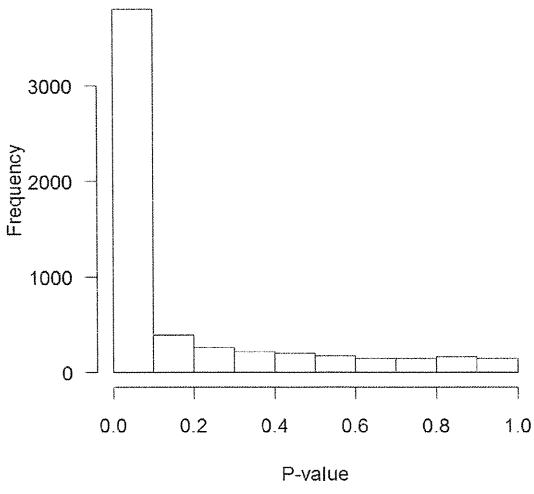

Figure 1. Frequency histogram of *p*-values.
doi:10.1371/journal.pone.0063249.g001

Results

Gene Expression Profile of CUP and Known Primary Cancers

A total of 237 genes were found to be either up-regulated or down-regulated by more than 2-fold between the normal lymph node and 30 CUP samples (training datasets). Of these, 59 genes with more than a 2.5-fold change (44 up-regulated and 15 down-regulated genes) are listed in Table 2. We designated the gene sets consisting of these CUP associated genes with >2 fold and >2.5 fold up-regulation or down-regulation as $M_{CUP}(2.0)$ and $M_{CUP}(2.5)$, respectively. Using these probe sets in $M_{CUP}(2.5)$, linear discriminant analysis (LDA) was performed for the CUP training datasets together with 2,364 cancers of various known types and 10 normal lymph nodes. As expected, all 2,404 samples were correctly discriminated. When the remaining 30 CUP samples (test datasets) were assessed using LDA that was modeled with the training datasets, 26 out of the 30 CUP samples were assigned correctly to "CUP", while only the 4 samples were predicted as "the other cancer". Thus, the accuracy of CUP was validated to be 86.7%, indicating that the 59 genes selected were of statistically significance as having biological attributes of CUP.

Figure 2 shows the supervised clustering of all 60 CUP samples performed together with 2,364 cancers of various known types and

Table 2. Genes identified as being up-regulatred or down-regulated in CUP.

Symbol	Gene description (Gene up-regulated in CUP)	Probe_set_ID	Log-fold change*	Fold change
RPL18A	Ribosomal protein L18A	200869_at	1.974	7.2
S100A4	S100 calcium binding protein A4	203186_s_at	1.587	4.9
PRG1	Proteoglycan 1, secretory granule	201858_s_at	1.539	4.7
SUB1	SUB1 homolog (S. cerevisiae)	214512_s_at	1.535	4.6
S100A6	S100 calcium binding protein A6	217728_at	1.523	4.6
RPS7	Ribosomal protein S7	200082_s_at	1.369	3.9
RPL11	Ribosomal protein L11	200010_at	1.245	3.5
PFN1	Profilin 1	200634_at	1.229	3.4
LOC23117	KIAA0220-like proteinKIAA0220	211996_s_at	1.212	3.4
TYROBP	TYRO protein tyrosine kinase binding protein	204122_at	1.196	3.3
TIMP1	TIMP metallopeptidase inhibitor 1	201666_at	1.178	3.2
SERF2	Small EDRK-rich factor 2	217756_x_at	1.173	3.2
YWHAZ	14-3-3 protein, zeta polypeptide	200641_s_at	1.169	3.2
LSM7	LSM7 homolog, U6 small nuclear RNA associated (S. cerevisiae)	204559_s_at	1.151	3.2
GSTP1	Glutathione S-transferase pi	200824_at	1.141	3.1
YWHAH	14-3-3 protein, eta polypeptide	201020_at	1.102	3.0
LAPTM5	Lysosomal associated multispanning membrane protein 5	201721_s_at	1.095	3.0
SNRPD2	Small nuclear ribonucleoprotein D2 polypeptide 16.5 kDa	200826_at	1.087	3.0
LOC392501	similar to 60 S ribosomal protein L26	222229_x_at	1.076	2.9
OAZ1	Ornithine decarboxylase antizyme 1	215952_s_at	1.073	2.9
POLR2J	Polymerase (RNA) II (DNA directed) polypeptide J, 13.3kDa	212782_x_at	1.062	2.9
EIF5A	Eukaryotic translation initiation factor 5A	201123_s_at	1.028	2.8
ATP5H	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	210149_s_at	1.023	2.8
APOC1	Apolipoprotein C-I	213553_x_at	1.018	2.8
LGALS1	Lectin, galactoside-binding, soluble, 1 (galectin 1)	201105_at	1.013	2.8
S100A11	S100 calcium binding protein A11	200660_at	1.010	2.7
SH3BGRL3	SH3 domain binding glutamic acid-rich protein like 3	221269_s_at	0.996	2.7
C1QB	complement component 1, q subcomponent, B chain	202953_s_at	0.984	2.7
RPS10	Ribosomal protein S10	216505_x_at	0.984	2.7
HSPA8	Heat shock 70 kDa protein 8	210338_s_at	0.972	2.6
NUTF2	Nuclear transport factor 2	202397_at	0.972	2.6
PRKDC	Protein kinase, DNA-activated, catalytic polypeptide	208694_at	0.967	2.6
NOLA3	Nucleolar protein family A, member 3 (H/ACA small nucleolar RNPs)	217962_at	0.957	2.6
TCEB2	Transcription elongation factor B (SIII), polypeptide 2 (18 kDa, elongin B)	200085_s_at	0.953	2.6
LOC442171	similar to ribosomal protein L10	217379_at	0.952	2.6
NEDD8	Neural precursor cell expressed, developmentally down-regulated 8	201840_at	0.944	2.6
LOC646417	similar to 60 S ribosomal protein L29 (P23)	216570_x_at	0.939	2.6
RPL36	Ribosomal protein L36	219762_s_at	0.937	2.6
VIM	Vimentin	201426_s_at	0.924	2.5
STK17A	Serine/threonine kinase 17a (apoptosis-inducing)	202693_s_at	0.922	2.5
NDUFS8	NADH dehydrogenase (ubiquinone) Fe-S protein 8, 23 kDa	203189_s_at	0.911	2.5
SELT	Selenoprotein T	217811_at	0.908	2.5
CST3	Cystatin C (amyloid angiopathy and cerebral hemorrhage)	201360_at	0.906	2.5
RPLP2	Ribosomal protein, large, P2	200909_s_at	0.901	2.5
Symbol	Gene description (Gene down-regulated in CUP)	Probe_set_ID	Log-fold change*	Fold change
ATP1B1	ATPase, Na+/K+ transporting, beta 1 polypeptide	201242_s_at	−0.891	0.4
NGFRAP1	Nerve growth factor receptor (TNFRSF16) associated protein 1	217963_s_at	−0.968	0.4
FOXJ3	Forkhead box J3	206015_s_at	−0.978	0.4
GABARAPL1	GABA(A) receptor-associated protein like 1	211458_s_at	−0.984	0.4

Table 2. Cont.

Symbol	Gene description (Gene up-regulated in CUP)	Probe_set_ID	Log-fold change*	Fold change
CD24	CD24 molecule	216379_x_at	−0.995	0.4
IVNS1ABP	Influenza virus NS1A binding protein	206245_s_at	−1.000	0.4
SCAMP1	Secretory carrier membrane protein 1	212417_at	−1.037	0.4
SEC22B	SEC22 vesicle trafficking protein homolog B (S. cerevisiae)	214257_s_at	−1.047	0.4
ITM2B	Integral membrane protein 2B	217731_s_at	−1.071	0.3
PDIA3	Protein disulfide isomerase family A, member 3	208612_at	−1.071	0.3
PIN4	Protein (peptidylprolyl cis/trans isomerase) NIMA-interacting, 4 (parvulin)	214224_s_at	−1.087	0.3
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	214352_s_at	−1.175	0.3
DICER1	Dicer1, Dcr-1 homolog (Drosophila)	213229_at	−1.264	0.3
SWAP70	SWAP-70 protein	209306_s_at	−1.342	0.3
VAPA	VAMP (vesicle-associated membrane protein)-associated protein A, 33 kDa	208780_x_at	−2.720	0.1

Each of the gene symbols, description, probe set in HG-U133 plus 2.0, log fold change and fold change are given in the table.
*Natural logarithm of fold change (CUP/normal lymph node).
doi:10.1371/journal.pone.0063249.t002

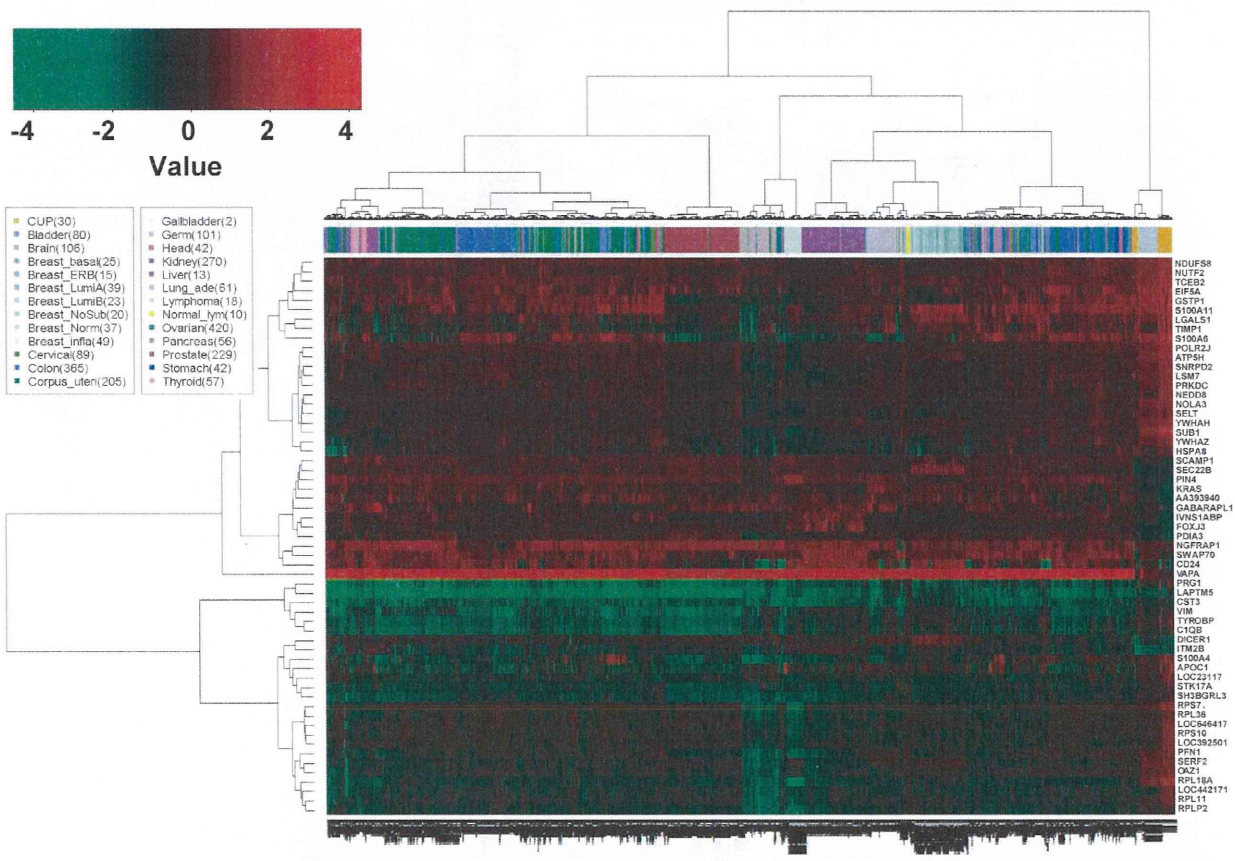

Figure 2. Heatmap representing the expressions of 59 genes with significant different expression in CUP compared with other cancer types or normal lymph nodes. Genes are indicated on the right. The colored bar above the heatmap represents the different cancer types, and the legend key is on the left. On the heatmap, red represents up-regulated genes and green represents down-regulated genes, relative to the expression levels in normal lymph nodes, with the scale shown in the upper left corner. The gene expression profiling datasets for normal lymph nodes and 24 known cancer types other than CUP were obtained from publicly available sources, as described in the Materials and Methods.
doi:10.1371/journal.pone.0063249.g002

10 normal lymph nodes using the 59 genes. The CUP samples were split into 2 groups with lung adenocarcinoma (LAC) clustered in between (right most part of the heat map). The larger group consisted of 42 samples, while the smaller consisted of 15 samples. Only 3 CUP samples were not included in any of these groups and instead were included in the clusters for normal lymphoma, brain tumors, and ovarian cancer, respectively. These were among the 4 samples that were predicted as "the other cancer" in the LDA. The *VAPA* gene, which was overexpressed in most of the cancer samples but not in CUP or LAC, revealed a striking contrast between CUP/LAC and other samples, which may have influenced the clustering analysis. When we re-analyzed the data after excluding the *VAPA* gene, the grouping for CUP was unchanged, but the smaller group with 15 samples was no longer clustered with LAC (Figure S1). The mean gene expression profiles (GEPs) for CUP, normal lymphoma, and 24 known cancer types were compared to create a dendrogram representing the quantified relations among CUP and the known cancer types, which again showed the clustering of CUP together with LAC (Figure S2).

Selection of CUP Associated Genes

Although the functions were diverse or unknown for the 44 up-regulated genes in the $M_{CUP}(2.5)$ datasets (Table 2), we found that

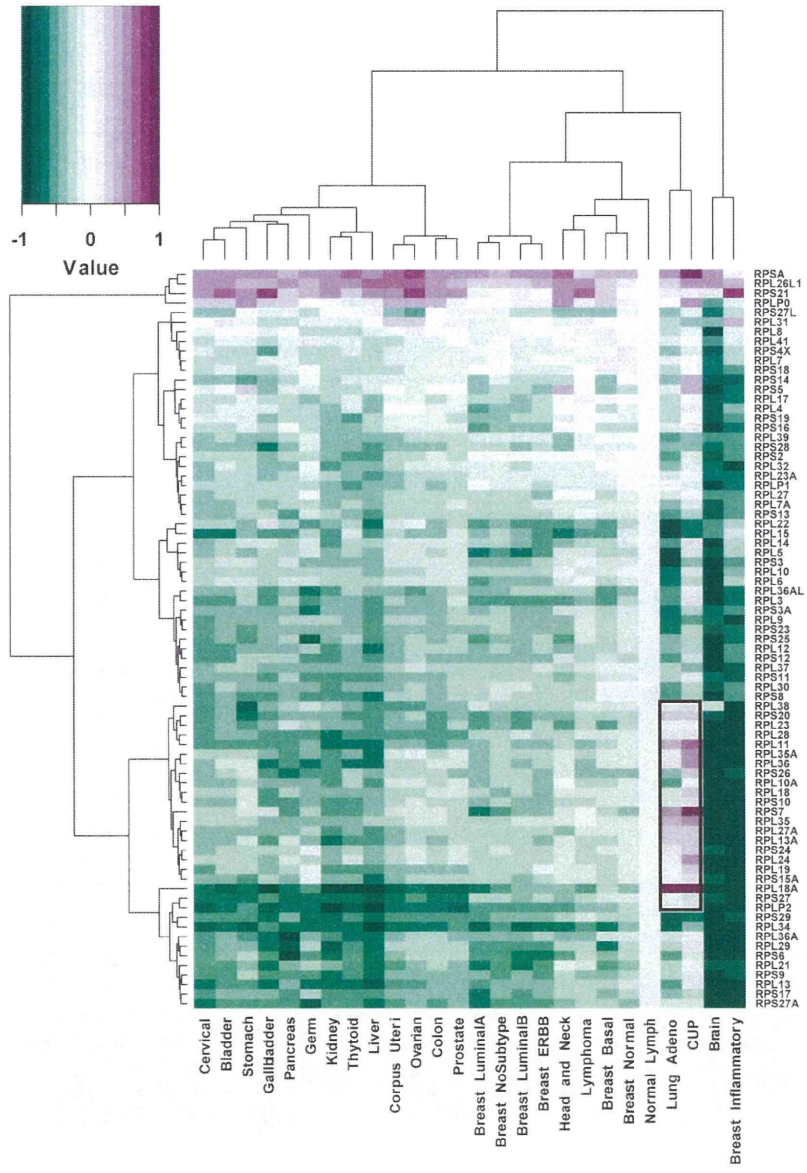

Figure 3. Heatmap representing the expression of 77 ribosomal protein genes in CUP, normal lymph nodes, and other cancer types. Ribosomal protein genes are indicated on the right. On the heatmap, purple represents up-regulated genes and green represents down-regulated genes, relative to the expression levels in normal lymph nodes, with the scale shown in the upper left. The genes that were exclusively overexpressed in CUP and lung adenocarcinoma are highlighted.
doi:10.1371/journal.pone.0063249.g003

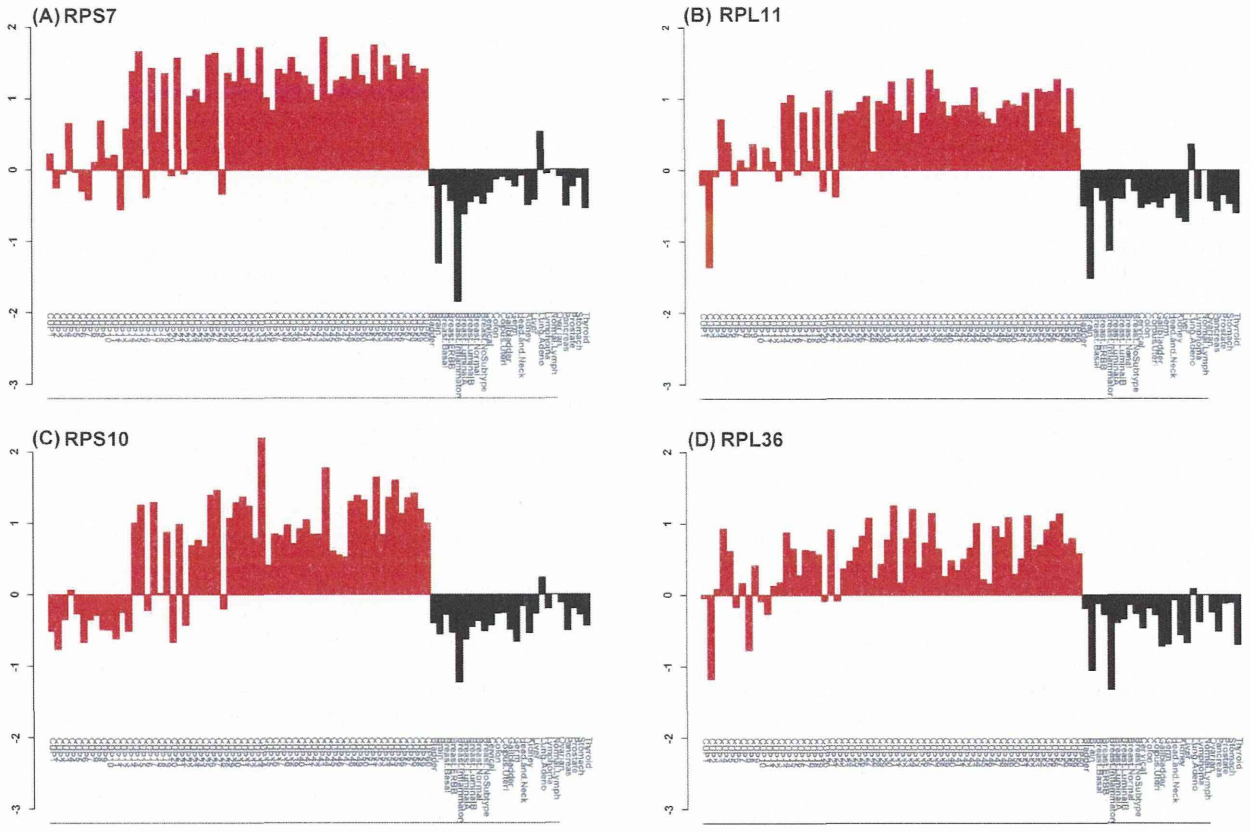

Figure 4. Relative expression levels for 4 ribosomal proteins. The relative expression levels of (A)RPS7, (B)RPL11, (C)RPS10, and (D)RPL36 were compared using individual CUP samples (n = 60), the mean expression levels of known cancer types, and a normal lymph node samples (n = 25). The asinh-transformed values for each gene were used for the calculations.
doi:10.1371/journal.pone.0063249.g004

14 genes (*S100A4*, *PRG1*, *S100A6*, *GSTP1*, *EIF5A*, *LGALS1*, *S100A11*, *PRKDC*, *VIM*, *CST3*, *TIMP1*, *YWHAZ*, *NEDD8*, *STK17A*) could be characterized after a search using the keywords "metastasis" and "apoptosis". Some of these genes were associated with the epithelial-to-mesenchymal transition (EMT), a function that has been increasingly recognized as a key step in cancer metastasis [23].

In the $M_{CUP}(2.5)$ dataset, 15 genes were down-regulated. Of these genes, we focused on *CD24*, *KRAS* and *DICER1*. The known functions of the above-mentioned up-regulated and down-regulated genes will be discussed in detail below.

Relative Expression of Up-Regulated Ribosomal Proteins

In the $M_{CUP}(2.5)$ dataset, we also identified 6 ribosomal proteins (*RPL18A*, *RPS7*, *RPL11*, *RPS10*, *RPL36*, *and RPLP2*). We found 11 more genes for ribosomal proteins (*RPL24*, *RPL35*, *RPL35A*, *RPS20*, *RPL13A*, *RPL28*, *RPS26*, *RPS14*, *RPL27A*, *RPL19*, *and RPL29*) in the $M_{CUP}(2.0)$ dataset. Ribosomal proteins are assembled into small and large ribosomal subunits. The small 40 S and large 60 S ribosomal subunits contain approximately 32 and 47 ribosomal proteins (known as RPS and RPL proteins), respectively [24]. The increased expression of ribosomal proteins has been associated with increased proliferation and growth; in some cases, however, increased expression has also been shown to suppress tumorigenesis [25,26].

To examine whether ribosomal protein genes can be used as biomarkers to discriminate CUP from other cancer types, the mean GEPs for a total of 77 ribosomal protein genes were compared using clustering for CUP, normal lymphoma, and 24 known cancer types (Figure 3). The ribosomal protein genes that were up-regulated in CUP were also up-regulated in LAC.

The relative mRNA expression levels of 4 ribosomal protein genes that were up-regulated in CUP (*RPS7*, *RPL11*, *RPS10*, and *RPL36*) were compared with the levels in normal lymphoma and 24 known cancer types (Figure 4). The 42 CUP samples that consistently contained large amounts of these mRNAs belonged to the larger CUP cluster, while the remaining 15 sample that showed relatively smaller amounts of these mRNAs belonged to the smaller cluster, as shown in Figure 2. As expected, the increased expressions of these mRNAs were also observed in LAC, but not in the other cancer types (Figure 4).

Discussion

Accumulating data sets from gene-expression microarray analyzed for various types of tumors have enabled the establishment of organ- and tumor-specific expression profiles that improve precise prediction of primary site of CUP [9,10,14,15]. Our official phase 2 study to corroborate the feasibility of CUP prediction using our algorithm is currently ongoing and will provide genes that exhibit unique expression pattern in CUP. A