

**Table 1. Genes differentially expressed between DLD-1 and DLD-1/FU cells**

NCBI Acc. No.	Gene symbol	Gene name	Cytoband	P-value	Fold change
<b>Genes highly expressed in DLD1/FU</b>					
NM_006283	TACC1	Transforming, acidic coiled-coil containing protein 1	8p11	1.00E-05	4.41
NM_001903	CTNNA1	Catenin, alpha 1, 102 kDa	5q31	1.46E-06	3.60
NM_017855	ODAM	Odontogenic, ameloblast associated	4q13.3	6.62E-06	3.48
NM_001945	HBEGF	Heparin-binding EGF-like growth factor	5q23	3.54E-06	3.18
NM_015570	AUTS2	Autism susceptibility candidate 2	7q11.22	3.28E-05	2.81
NM_005406	ROCK1	Rho-associated, coiled-coil containing protein kinase 1	18q11.1	9.22E-06	2.80
NM_004815	ARHGAP29	Rho GTPase activation protein 29	1q22.1	2.03E-05	2.78
NM_012482	ZNF281	Zinc finger protein 281	1q32.1	1.82E-06	2.72
NM_004848	C1orf38	Chromosome 1 open reading frame 38	1p35.3	2.61E-05	2.68
NM_033439	IL33	Interleukin 33	9p24.1	2.45E-05	2.51
NM_001165	BIRC3, cIAP2	Baculoviral IAP repeat-containing 3	11q22	1.99E-05	2.31
NM_024796	FLJ22639	Hypothetical protein FLJ22639	1p36.33	3.94E-05	2.23
NM_012328	DNAJB9	DnaJ homolog, subfamily B, member 9	14q24.2-q24.3	2.51E-05	2.22
NM_006281	STK3	Serin/threonine kinase 3	8q22.2	4.91E-05	2.12
NM_005433	YES1	V-yes-1 Yamaguchi sarcoma viral oncogene homolog 1	18p11.31-p11.21	3.29E-05	2.07
NM_007283	MGLL	Monoglyceride lipase	3q21.3	5.29E-06	2.00
<b>Genes with low expression in DLD1/FU</b>					
NM_000507	FBP1	Fructose-1,6-bisphosphatase 1	9q22.3	9.33E-07	0.16
NM_000873	ICAM2	Intercellular adhesion molecule 2	17q23-q25	8.38E-07	0.18
NM_000295	SERPINA1	Serpin peptidase inhibitor, clade A, member 1	14q32.1	1.99E-05	0.19
NM_000802	FOLR1	Folate receptor 1	11q13.3-q14.1	1.79E-05	0.22
NM_000300	PLA2G2A	Phospholipase A2, group IIA	1p35	4.37E-05	0.23
NM_030763	NSBP1	Nucleosomal binding protein 1	Xq13.3	2.86E-05	0.24
NM_000149	FUT3	Fucosyltransferase 3	19p13.3	1.27E-05	0.34
NM_018725	IL17RB	Interleukin 17 receptor B	3p21.1	5.57E-06	0.34
NM_002889	RARRES2	Retinoic acid receptor responder 2	7q36.1	2.51E-05	0.38
NM_005739	RASGRP1	RAS guanyl releasing protein 1	15q14	2.89E-05	0.45
NM_032355	MON1A	MON1 homolog A	3p21.31	2.87E-05	0.47
NM_000407	GP1BB	Glycoprotein 1b, beta polypeptide	22q11.21	8.30E-06	0.48
NM_003415	ZNF268	Zinc finger protein 268	12q24.33	1.20E-06	0.49
NM_002816	PSMD12	Proteasome 26S subunit, non-ATPase, 12	17q24.2	4.10E-05	0.49
NM_003360	UGT8	UDP glycosyltransferase 8	4q26	3.93E-05	0.50
NM_005116	SCL23A2	Solute carrier family 23, member 2	20p13	9.12E-06	0.50

The expression level of DLD-1/FU was compared to that of DLD-1 with fold change. Listed are genes with  $P$ -value  $< 5.0E-05$  and fold change  $> 2$  or  $< 0.5$ . NCBI Acc. No., National Center for Biotechnology Information accession number.

ratios of which varied most diversely with SDs of  $> 1.25$  (Fig. 2A). In the hierarchical clustering of microarray data of DLD-1 and DLD-1/FU combined with those of another 21 colon cancer cell lines, DLD-1 and DLD-1/FU were most closely clustered. Clone A, which is a subclone isolated from the DLD-1 cell line with an assay using a Boyden chamber,<sup>(27)</sup> was also clustered together with DLD-1 and DLD-1/FU. The population of genes differentially expressed between DLD-1 and DLD-1/FU cell lines with more than two-fold change was limited to 3.3% (448/13 472 genes; Fig. 2C). On the other hand, in the other 21 colon cancer cell lines, the population of genes differentially expressed with more than a two-fold change in comparison to DLD-1 which was higher than that between DLD-1 and DLD-1/FU, ranging from 4.8% to 24.0% (Fig. 2C). In Clone A and MIP101, which were most closely clustered to DLD-1 and DLD-1/FU with the hierarchical clustering (Fig. 2A), the population of genes differentially expressed with more than a two-fold change in comparison to DLD-1, which was 7.5% and 4.8%, respectively (Fig. 2c). On the other hand, in SW480, SW948, HCT-8, COLO320, SW620, LoVo and HT-29, the population of genes differentially expressed with more than a two-fold change in comparison to DLD-1 which was higher than 15% (Fig. 2C).

To list genes differentially expressed between DLD-1 and DLD-1/FU, two-sample  $t$ -test was employed. The expression levels of genes in DLD-1/FU were compared to those of DLD-1 with fold change and the genes with  $P$ -values  $< 5.0E-05$  and with fold change  $> 2$  or  $< 0.5$  are listed in Table 1. Genes highly expressed in DLD-

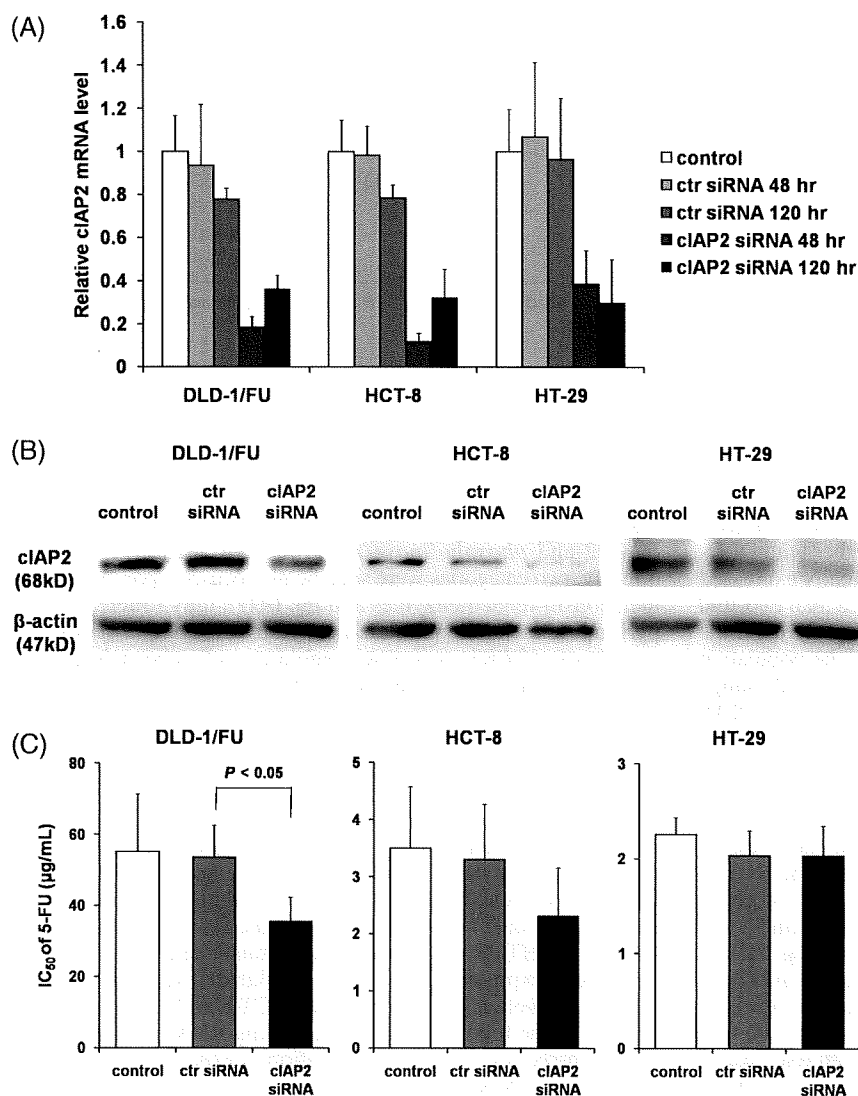
1/FU which had  $P$ -values  $< 5.0E-05$  and fold change  $> 2$  were 16 genes, including *catenin alpha 1 (CTNNA1)*, *heparin-binding EGF-like growth factor (HBEGF)*, and *cIAP2* (also known as *BIRC3*). There were 16 genes with a low expression in DLD-1/FU which had  $P$ -values  $< 5.0E-05$  and a fold change  $< 0.5$ , including *FBP1*, *ICAM2* and *phospholipase A2 group IIA (PLA2G2A)*. Subsequently, two categories were specifically focused upon: pyrimidine metabolism-related enzymes and the IAP family (Table 2). In the pyrimidine metabolism-related enzymes, *orotate phosphoribosyl transferase (OPRT)* was down-regulated as had been described in previous reports.<sup>(23)</sup> In seven genes of IAP family, *cIAP2* was most up-regulated in DLD-1/FU (2.31 times). Using real-time RT-PCR and Western blotting, we reconfirmed the results of cDNA microarray on cIAP2 expression in DLD-1 and DLD-1/FU cells (data not shown).

**Down-regulation of cIAP2 by siRNA.** The role of cIAP2 on 5-FU resistance in colon cancer cells was further investigated using RNAi. With the data from the microarray and RT-PCR, HCT-8, HT-29, SW948 and WiDr-TC cells showed high expression of the *cIAP2* gene (data not shown). Among the four types of cell lines, HCT-8 and HT-29 were successfully transfected with siRNA and the cIAP2 expression was down-regulated (Fig. 3A). For these reasons, as well as DLD-1/FU cells, HCT-8 and HT-29 cells were chosen for the following experiments using RNAi. At 48 and 120 h after transfection, the down-regulation of cIAP2 was assessed by real-time RT-PCR and 72 h after transfection by Western blotting

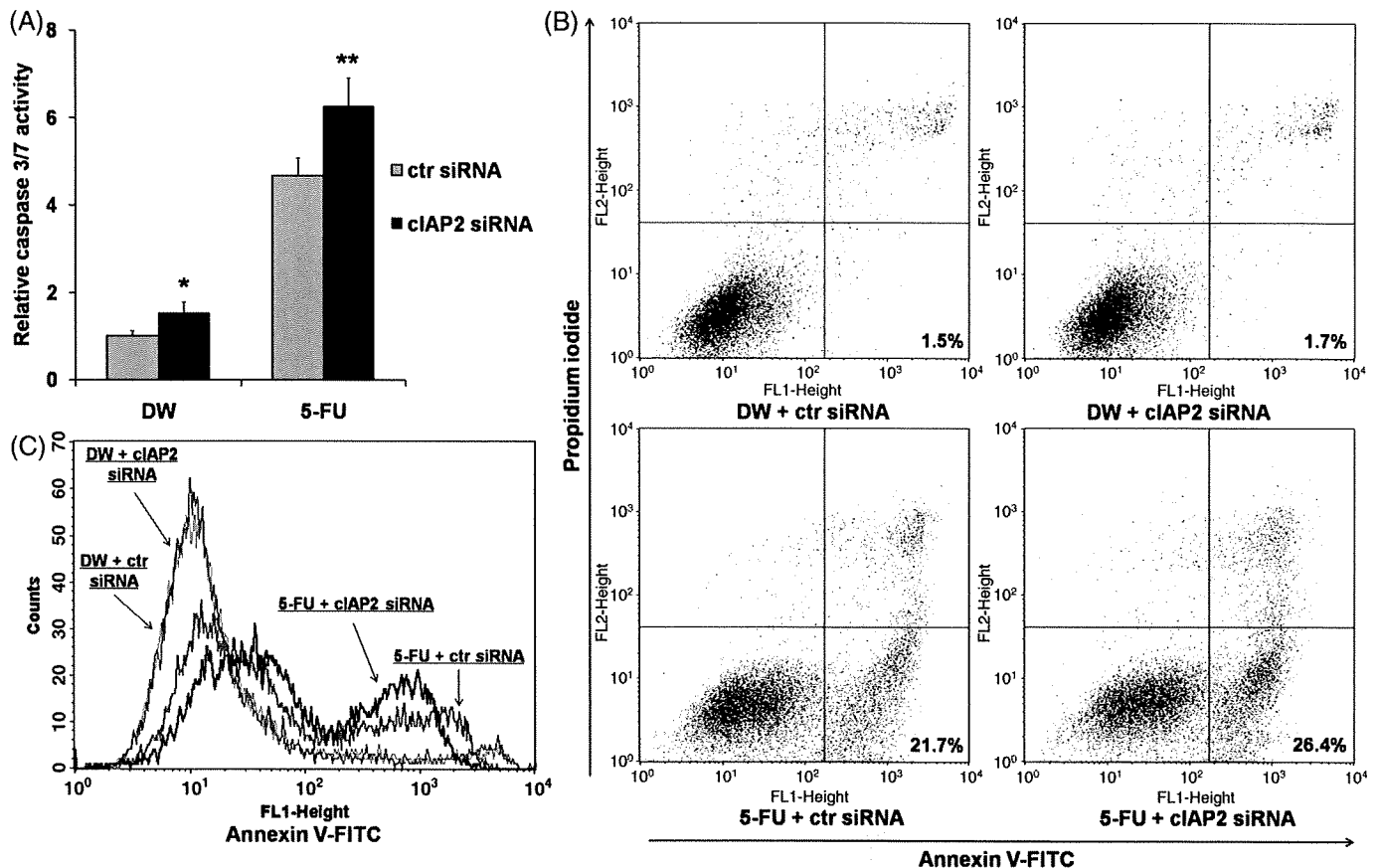
**Table 2. Genes differentially expressed between DLD-1 and DLD-1/FU cells involved in pyrimidine metabolism-related enzymes and IAP family**

NCBI Acc. No.	Gene symbol	Gene name	Cytoband	P-value	Fold change
<b>Pyrimidine metabolism-related enzymes</b>					
NM_001071	TYMS, TS	Thymidylate synthetase	18p11.32	7.34E-03	0.80
NM_000373	UMPS, OPRT	Uridine monophosphate synthetase	3q13	3.41E-04	0.57
NM_000110	DPYD, DPD	Dihydropyrimidine dehydrogenase	1p22	–	–
NM_001953	TYMP, TP	Thymidine phosphorylase	22q13.33	–	–
NM_003258	TK1	Thymidine kinase 1, soluble	17q23.2-q25.3	7.51E-03	0.78
NM_004614	TK2	Thymidine kinase 2, mitochondrial	16q22-q23.1	–	–
NM_001033	RRM1	Ribonucleotide reductase M1	11p15.5	N.S.	1.21
NM_001034	RRM2	Ribonucleotide reductase M2	2p25-p24	N.S.	0.83
NM_003364	UPP1	Uridine phosphorylase 1	7p12.3	N.S.	1.08
<b>IAP family</b>					
NM_004536	NAIP, BIRC1	NLR family, apoptosis inhibitory protein	5q13.1	–	–
NM_001166	BIRC2, cIAP1	Baculoviral IAP repeat-containing 2	11q22	N.S.	1.25
NM_001165	BIRC3, cIAP2	Baculoviral IAP repeat-containing 3	11q22	1.99E-05	2.31
NM_001167	XIAP, BIRC4	X-linked inhibitor of apoptosis	Xq25	–	–
NM_001168	BIRC5, SURVIVIN	Baculoviral IAP repeat-containing 5	17q25	N.S.	0.97
NM_016252	BIRC6, APOLLON	Baculoviral IAP repeat-containing 6	2p22-p21	4.82E-02	1.23
NM_022161	BIRC7, LIVIN	Baculoviral IAP repeat-containing 7	20q13.3	N.S.	1.02

The expression level of DLD-1/FU was compared to that of DLD-1 with fold change. NCBI Acc. No., National Center for Biotechnology Information accession number; IAP, apoptosis inhibitory protein; N.S., not significant; –, not reliable for low expressions.



**Fig. 3. Down-regulation of cellular inhibitor of apoptosis 2 (cIAP2) by small interfering RNA (siRNA).** The messenger RNA (mRNA) and protein levels of cIAP2 were assessed by (A) real-time reverse transcription – polymerase chain reaction or (B) Western blotting in DLD-1/FU, HCT-8 and HT-29 under transfection of cIAP2 siRNA. As controls, untransfected cells (control) and cells transfected with control siRNA (ctr siRNA) were used. mRNA were isolated 48 or 120 h after transfection (Fig. 3a). Cell lysates were collected 72 h after transfection (Fig. 3b). (C) Effect of the down-regulation of cIAP2 on 5-FU sensitivity in human colon cancer cell lines. *In vitro* cytotoxicity assay was performed using MTS assay. At 48 h after transfection with cIAP2 siRNA or control siRNA, 5-FU was added at various concentrations. 72 h after treatment with 5-FU, MTS assay was performed to determine the 50% inhibitory concentration. The results represent the means ± SD of three independent experiments.



**Fig. 4.** (A) Down-regulation of cellular inhibitor of apoptosis 2 (cIAP2) induced caspase 3/7 activation. At 48 h after cIAP2 small interfering RNA (siRNA) transfection, cells were exposed to distilled water (DW) or 5-fluorouracil (5-FU) for 48 h and measured for caspase 3/7 activity. As controls, the cells transfected with control siRNA (ctr siRNA) were used. The results represent the ratios to caspase 3/7 activity of untransfected cells with common culture. \* $P < 0.05$  versus cells treated with ctr siRNA + DW. \*\* $P < 0.05$  versus cells treated with each of the other treatments. (B) Incidence of apoptosis in DLD-1/FU revealed by flow cytometric analysis stained with annexin V-fluorescein isothiocyanate (V-FITC) and counterstained with propidium iodide (PI). In cIAP2 siRNA-transfected cells with exposure to 5-FU, the percentage of annexin V-positive/PI-negative cells (early apoptotic cells) shown in the right lower quadrant was high in comparison to the control siRNA-transfected cells. (C) The histogram represents the increase of annexin V-positive fraction in cIAP2 siRNA-transfected cells with exposure to 5-FU. A representative assay out of three independent assays was provided.

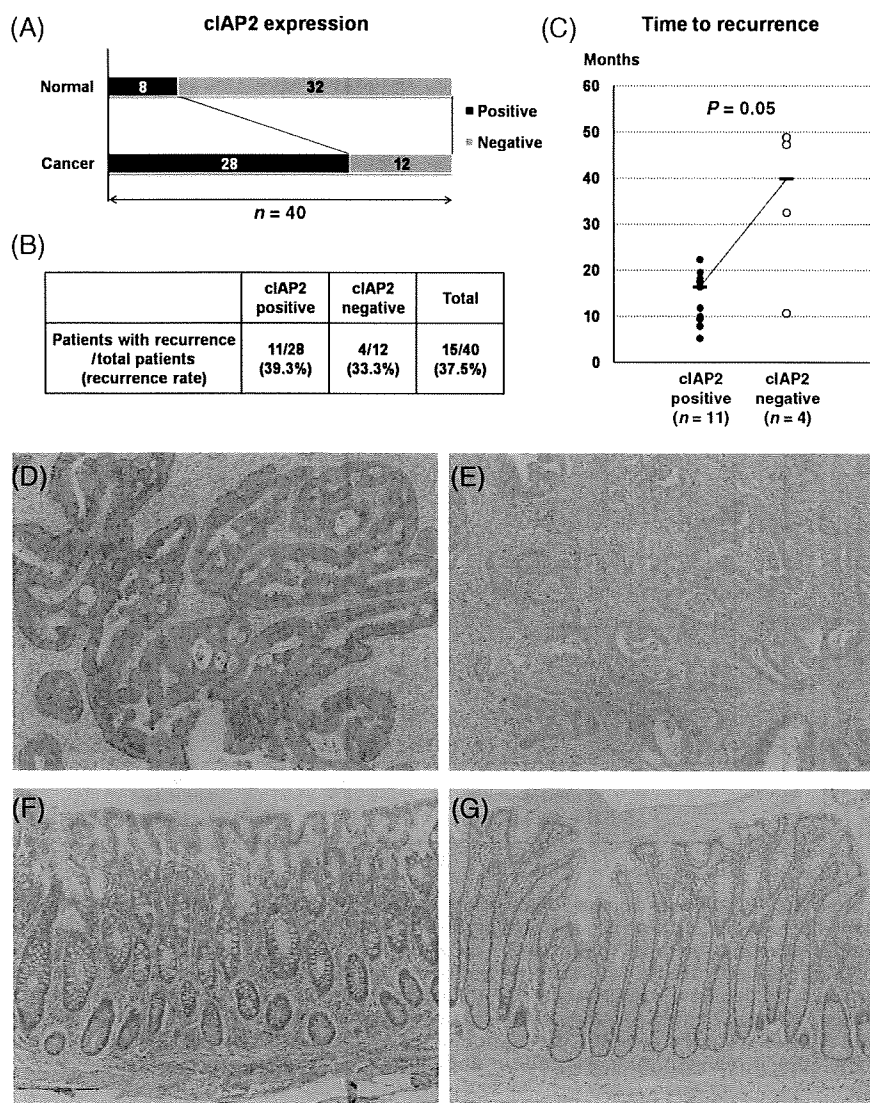
(Fig. 3A,B). The expression of cIAP2 was always down-regulated with higher than 50% efficiency, although it could vary depending on the cell types or time points (Fig. 3A,B).

**Enhanced 5-FU sensitivity by cIAP2 siRNA.** At 48 h after transfection with cIAP2 siRNA or non-targeting siRNA as control siRNA ('ctr siRNA' in Figs 3 and 4), DLD-1/FU, HCT-8 and HT-29 cells were cultured in medium with 10 dilutions of 5-FU and then the  $IC_{50}$  was determined by MTS assay. In all of the cell types, treatment with control siRNA did not have a cytotoxic effect on the  $IC_{50}$  (Fig. 3C). DLD-1/FU cells transfected with cIAP2 siRNA in the presence of 5-FU showed a significant decrease of the  $IC_{50}$  in comparison to those transfected with the control siRNA ( $P < 0.05$ ; Fig. 3C). These results indicate that the down-regulation of cIAP2 in DLD-1/FU cells enhanced the cytotoxicity with 5-FU. The  $IC_{50}$  of 5-FU in HCT-8 cells changed in the same manner as in DLD-1/FU ( $P = 0.15$ ); but the  $IC_{50}$  of 5-FU in HT-29 cells did not change under the down-regulation of cIAP2 (Fig. 3C).

**Activation of caspase 3/7 and induction of apoptosis in DLD-1/FU by cIAP2 siRNA.** At 48 h after transfection with cIAP2 siRNA or control siRNA, 5-FU (500  $\mu$ g/mL) or DW was added to DLD-1/FU cells and the cells were maintained for another 48 h. Then, caspase 3/7 activity was measured and the induction of apoptosis was also analyzed with flow cytometry. As shown in Fig. 4(A), the induction of caspase 3/7 activity was examined in DLD-1/FU cells under the down-regulation of cIAP2 with or without exposure to 5-FU. Without exposure to 5-FU, the caspase 3/7 activity under the down-

regulation of cIAP2 was enhanced in comparison to that with control siRNA (with arbitrary units 1.53 vs. 0.99,  $P = 0.03$ ; Fig. 4A). After exposure to 5-FU, again the caspase 3/7 activity under the down-regulation of cIAP2 was enhanced in comparison to that with control siRNA (6.26 vs. 4.67,  $P = 0.02$ ). Next, as shown in Fig. 4(B), the induction of apoptosis with annexin V and PI staining were examined in DLD-1/FU cells under the down-regulation of cIAP2 with or without exposure to 5-FU. Without exposure to 5-FU, induction of early or late apoptotic cells was not observed under the down-regulation of cIAP2 (subpopulated in the right lower portion or the right upper portion in each graph of Fig. 4B, respectively); but with exposure to 5-FU, the induction of early apoptotic cells under the down-regulation of cIAP2 was seen (26.4%). As shown in Fig. 4(C), with exposure to 5-FU, the fraction of early and late apoptotic cells in DLD-1/FU cells was enhanced under the down-regulation of cIAP2 with two-peaks curves, although it was not observed without exposure to 5-FU.

**cIAP2 expression in human colorectal cancer and normal tissues with immunohistochemistry and its effect on the prognosis.** Immunohistochemistry for cIAP2 was analyzed on colorectal cancer and corresponding normal tissues from 40 patients who underwent curative operations followed by fluorouracil-based adjuvant chemotherapies. cIAP2 was detected exclusively in the cytoplasm of cancer and normal epithelial cells. Seventy percent of cancer tissues were positive for cIAP2 (28/40 cases; Fig. 5A,D,E), whereas only 20% of the normal tissues were positive (8/40 cases; Fig. 5A,F,G).



**Fig. 5.** (A) Cellular inhibitor of apoptosis 2 (ciAP2) expression in human colorectal cancer and the corresponding normal tissues by immunohistochemistry. ciAP2 was more frequently expressed in cancer tissues than in normal tissues. (B) The incidence of cancer recurrence after curative operations in ciAP2-positive and ciAP2-negative patients. (C) The time to recurrence in ciAP2-positive and ciAP2-negative patients. (D–G) Representative image of immunohistochemical staining for ciAP2 in human colorectal cancer and normal tissues (D, cancer tissue/ciAP2-positive; E, cancer tissue/ciAP2-negative; F, normal tissue/ciAP2-positive; G, normal tissue/ciAP2-negative; original magnification:  $\times 100$ ).

Moreover, we analyzed the 40 colorectal cancer patients on the association between ciAP2 expression in the cancer tissues and cancer recurrence. Of the 40 patients, 15 patients (37.5%) experienced cancer recurrence within 5 years after surgery (Fig. 5B). Although we did not find a difference on incidence of cancer recurrence between ciAP2-positive and ciAP2-negative cancer patients (recurrence rate 39.3% vs. 33.3%; Fig. 5B), ciAP2-positive cancer patients had a trend to recur earlier after operations than ciAP2-negative cancer patients. Although the number of the cases was limited, we compared the time to recurrence between the two subgroups of the 15 patients: ciAP2-positive patients ( $n = 11$ ) and ciAP2-negative patients ( $n = 4$ ). The time to recurrence in ciAP2-positive patients and ciAP2-negative patients was 16.3 and 39.9 months, respectively ( $P = 0.05$ ; Fig. 5C).

## Discussion

Single-drug chemotherapy using 5-FU or its derivatives or combined cytotoxic chemotherapies including 5-FU have been chosen to treat colon cancer. Although anticancer agents and their regimens have been developed, their efficacies have not yet reached a satisfactory level. As Scheithauer *et al.* reported in 1993, the randomized comparison in patients with metastatic colorectal cancers revealed that the median survival time of patients without the use of any form of treatment is in the range of 5 months.<sup>(28)</sup> According to the

Cochrane database,<sup>(29)</sup> when either fluorouracil alone or fluorouracil combined with leucovorin is administered, the median survival time of patients with metastatic colorectal cancers is extended to 10–12 months. As Salts *et al.*<sup>(30)</sup> and de Gramont *et al.*<sup>(31)</sup> reported, the median survival time is prolonged to 14–16 months when either oxaliplatin or irinotecan is added to 5-FU-based chemotherapy. Furthermore, the median survival is prolonged to longer than 20 months when a targeted therapy including bevacizumab or cetuximab is added to the cytotoxic chemotherapies including FOLFOX (folinic acid, FU and oxaliplatin) or FOLFIRI (folinic acid, FU and irinotecan).<sup>(4)</sup> Currently 5-FU plays a central role in chemotherapies for colorectal cancers and thus it is important to elucidate the mechanisms defining the sensitivity to 5-FU for designing successful chemotherapies. Improvement of 5-FU sensitivity, the selection of agents combined with 5-FU-based chemotherapies and the selection of regimens for individual patients are especially important to achieve better prognosis in advanced colorectal cancer cases.

In the current study, focusing specifically on improving the sensitivity to 5-FU and the selection of types of therapy for each individual patient, to identify novel genes which define the sensitivity or tolerance to 5-FU in colorectal cancers, the expression profiles of the parental DLD-1 cell line was compared to its 5-FU-resistant subclone DLD-1/FU. First, a hierarchical clustering was performed with microarray data of DLD-1 and DLD-1/FU combined with those of 21 other colon cancer cell lines. As shown

in Fig. 2(A), DLD-1 and DLD-1/FU were most closely clustered and Clone A and MIP101 were also closely clustered. The expression profiles of genes originating from the entire genome will represent the characteristics of cells and this strategy will be good for prediction of the origin of cell types or derived organs; but this strategy will not be good for more precise analyses, including the prediction of sensitivity to some drugs.<sup>(32)</sup> For such a purpose, it will be necessary to identify a special set of genes. In Fig. 2(A) 289 genes were selected as representatives from the whole set of genes. The expression profiles of DLD-1 and DLD-1/FU cells were quite similar in comparison to the other cell lines, even after the acquisition of 5-FU resistance in DLD-1/FU cells as shown in Fig. 2(B). These results suggest that a change of expression profiles in a limited population of genes contributes to the acquisition of 5-FU resistance in DLD-1/FU cells. Figure 2(C) demonstrates that 3.3% of the genes (448/13 472 genes) were differentially expressed between DLD-1 and DLD-1/FU with more than a two-fold change. In the other 21 colon cancer cell lines, the population of genes differentially expressed with more than a two-fold change in comparison to DLD-1 was higher than that between DLD-1 and DLD-1/FU, ranging from 4.8% to 24.0% (Fig. 2C). This revealed that, as expected, the difference of 5-FU sensitivity was derived from a change of expressions in a limited member of genes. After the acquisition of 5-FU resistance, the secondary or tertiary effects may have affected the changes in expressions in some of the 448 genes; and therefore the number of genes essential for determining the difference in 5-FU sensitivity may be much smaller than 448 genes. Schmidt *et al.* analyzed the change of expression profiles during the establishment of a 5-FU-resistant subclone from the human colon cancer cell line SW620 using a cDNA microarray and the number of genes which showed expression changes in their analysis was limited to 4.8% among the genes they used (330/6888 genes).<sup>(33)</sup> Their result is consistent with the current findings, both of which suggested that the limited number of genes regulate the sensitivity or resistance to 5-FU in colorectal cancer cells. On the other hand, comparing the hierarchical clustering (Fig. 2A) to the  $IC_{50}$  of 5-FU (Fig. 2B) or the population of genes differentially expressed among the 23 colon cancer cell lines (Fig. 2C), it is difficult to identify a correlation. These results show that 5-FU sensitivity is regulated by diverse mechanisms or pathways. As shown in Fig. 3(C), the down-regulation of cIAP2 induced the HCT-8 cell line to become more sensitive to 5-FU, but the 5-FU sensitivity of HT-29 did not change. This may suggest the possibility that the sensitivity to anticancer agents including 5-FU is regulated by a variety of pathways in cancer cells. Harlin *et al.* reported the compensating pathways of IAP functions using a XIAP knock-out mouse.<sup>(34)</sup> In the current experiments using HT-29, such compensating pathways may therefore have functioned.

In regard to pyrimidine metabolism-related enzymes (Table 2), a comparison between DLD-1 and DLD-1/FU showed that the OPRT gene was down-regulated in DLD-1/FU. The correlation between the expression of OPRT and 5-FU sensitivity was previously described,<sup>(35,36)</sup> and the current result was consistent with those reports. However, our results on TS and DPD did not support the results reported previously, which may suggest the difficulty of such analyses.

To identify novel genes that regulate 5-FU sensitivity, the expression profiles of DLD-1 and DLD-1/FU were compared and the genes differentially expressed between DLD-1 and DLD-1/FU were listed (Table 1). Among the genes with low expression in DLD-1/FU, the *PLA2G2A* gene, which encodes a group IIA-secreted phospholipase A2, is of particular interest. The expression of *PLA2G2A* has been reported to suppress the progression or metastasis of human gastric cancers.<sup>(37)</sup> Recently, Ganesan *et al.* also reported that the anti-invasive effect of *PLA2G2A* occurs through its ability to inhibit the *S100A4* gene, which is a metastasis mediator gene in gastric cancer cells.<sup>(38)</sup> These results may suggest that the expression of the *PLA2G2A* gene may affect the character

of cancer cells, including their responses to anticancer drugs such as 5-FU.

In genes highly expressed in DLD-1/FU, some of these genes are of interest from the standpoint of 5-FU resistance. HBEGF, which binds to epidermal growth factor receptor (EGFR)/ErbB1 and ErbB4, is a member of the EGF family. The overexpression of HBEGF has been reported in several types of cancers including colon cancer and pancreatic cancer.<sup>(39,40)</sup> Suganuma *et al.* identified *HBEGF* as a 5-FU- and cisplatin-resistance-related gene in gastric cancer using cDNA microarray,<sup>(41)</sup> and Wang *et al.* further investigated the potential role of HBEGF which regulated the chemotherapy-induced EGFR activation,<sup>(42)</sup> and our results in Table 1 support their findings. Next, *CTNNA1* encodes  $\alpha$ -catenin whose interaction with actin-based cytoskeleton is required for the strong state of E-cadherin-based cell adhesion activity.<sup>(43)</sup> Matsubara *et al.* reported  $\alpha$ -catenin expression increases the resistance to apoptosis induced by sphingosine in the colon cancer cell line and  $\alpha$ -catenin mediated transduction of signals from the cadherin-catenin complex to regulate the apoptotic cascade.<sup>(44)</sup> *CTNNA1* is therefore a gene that can possibly regulate drug resistance through the apoptotic pathway.

Among the genes highly expressed in DLD-1/FU, the *cIAP2* gene has drawn a lot of attention and the role of the *cIAP2* gene on 5-FU sensitivity was further analyzed in colon cancer cell lines. In our analysis using RNAi, the down-regulation of *cIAP2* changed the phenotype of DLD-1/FU to be more sensitive to 5-FU (Fig. 3C). Lopes *et al.* reported that the expression of cIAP2 protein affected the sensitivity to cisplatin and doxorubicin in pancreatic cancer cell lines.<sup>(21)</sup> In addition, the expression of IAP family genes, such as survivin and *XIAP*, affect the sensitivity to several types of anticancer drugs in cancers of some organs such as the lung and ovary.<sup>(20-22,45-47)</sup> However, the association between drug sensitivity and the IAP family in colon cancer has not been reported so far. The current result raises the possibility for novel strategies of cancer chemotherapy in colorectal cancer.

After the down-regulation of *cIAP2* by siRNA, the expression level of *cIAP2* in DLD-1/FU was almost equivalent to that in its parental cell DLD-1 (Table 1, Fig. 3A); however, the  $IC_{50}$  of 5-FU in DLD-1/FU (35.5  $\mu$ g/mL) was still much higher than that in DLD-1 (0.488  $\mu$ g/mL). This result suggests that some other genes, such as those listed in Table 1, may also play an important role in regulating 5-FU resistance as well as *cIAP2*.

Although the association between IAPs and drug resistance has been previously discussed, there have been few reports discussing overexpression of IAPs. Although we have not shown the association between *cIAP2* overexpression and 5-FU sensitivity in our report, Pratt *et al.* have analyzed survivin overexpression in breast cancer cell line and showed that unexpectedly the increased survivin did not promote paclitaxel resistance.<sup>(47)</sup> Their results suggest the difficulty in elucidating the association between *cIAP2* induction and drug resistance.

With the down-regulation of *cIAP2* without 5-FU exposure, the activity of caspase 3/7 was induced, but apoptosis was not induced (Fig. 4). However, with the down-regulation of *cIAP2* with 5-FU exposure, apoptosis was efficiently induced (Fig. 4). The results indicate that the activation of the caspase 3/7 pathway is essential for the induction of apoptosis in DLD-1/FU cells; and exposure to 5-FU was required for the induction of the apoptosis in addition to the activation of the caspase 3/7 pathway. As Johnstone *et al.* reported previously,<sup>(12)</sup> the sensitivity or resistance to anticancer drugs is regulated by interactions between drugs and intracellular targets of the drugs (drug-target interaction); however, depending on the severity of the initial insult, drug-induced damage will also result in catastrophic death or it may initiate a series of secondary effects mediated by various stress signaling pathways leading to cell cycle arrest or cell death.<sup>(12)</sup> Apoptosis is one of the secondary effects that plays an important role and the caspase cascade is included in the pathways of apoptosis. Recently IAP has been

considered as a novel target of cancer therapy and the efficacy of antisense oligonucleotides or small-molecule IAP antagonists has been reported.<sup>448-50</sup> In colorectal cancer, a target therapy against IAP genes may play an important role in the future.

To further identify the role of cIAP2 and their association with 5-FU resistance in human colorectal cancer, immunohistochemistry for cIAP2 was performed on colorectal cancer and normal tissues from 40 patients who underwent curative operations followed by fluorouracil-based adjuvant chemotherapies. As shown in Fig. 5, cIAP2 was more frequently expressed in cancer tissues than in normal tissues. Moreover cIAP2-positive cancer patients had a trend toward early recurrence after fluorouracil-based adjuvant chemotherapies and this result sufficiently supported our results *in vitro*. It suggested that cIAP2 performed a crucial function and could be a potential therapeutic target in human colorectal cancer.

In conclusion, this study indicated that differences of 5-FU sensitivity could thus be defined by a limited number of genes.

## References

- Giacchetti S, Perpoint B, Zidani R *et al*. Phase III multicenter randomized trial of oxaliplatin added to chronomodulated fluorouracil-leucovorin as first-line treatment of metastatic colorectal cancer. *J Clin Oncol* 2000; **18**: 136-47.
- Douillard JY, Cunningham D, Roth AD *et al*. Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. *Lancet* 2000; **355**: 1041-7.
- Cunningham D, Humblet Y, Siena S *et al*. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med* 2004; **351**: 337-45.
- Hurwitz H, Fehrenbacher L, Novotny W *et al*. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med* 2004; **350**: 2335-42.
- Hinoda Y, Sasaki S, Ishida T, Imai K. Monoclonal antibodies as effective therapeutic agents for solid tumors. *Cancer Sci* 2004; **95**: 621-5.
- Wohlhueter RM, McIvor RS, Plagemann PG. Facilitated transport of uracil and 5-fluorouracil, and permeation of orotic acid into cultured mammalian cells. *J Cell Physiol* 1980; **104**: 309-19.
- Longley DB, Harkin DP, Johnston PG. 5-Fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer* 2003; **3**: 330-8.
- Pojo T, Bates S. Strategies for reversing drug resistance. *Oncogene* 2003; **22**: 7512-23.
- Zembutsu H, Ohnishi Y, Tsunoda T *et al*. Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs. *Cancer Res* 2002; **62**: 518-27.
- Maxwell PJ, Longley DB, Latif T *et al*. Identification of 5-fluorouracil-inducible target genes using cDNA microarray profiling. *Cancer Res* 2003; **63**: 4602-6.
- Lowe SW, Lin AW. Apoptosis in cancer. *Carcinogenesis* 2000; **21**: 485-95.
- Johnstone RW, Ruefli AA, Lowe SW. Apoptosis: a link between cancer genetics and chemotherapy. *Cell* 2002; **108**: 153-64.
- Brown JM, Wouters BG. Apoptosis, p53, and tumor cell sensitivity to anticancer agents. *Cancer Res* 1999; **59**: 1391-9.
- Chang HY, Yang X. Proteases for cell suicide: functions and regulation of caspases. *Microbiol Mol Biol Rev* 2000; **64**: 821-46.
- Fuentes-Prior P, Salvesen GS. The protein structures that shape caspase activity, specificity, activation and inhibition. *Biochem J* 2004; **384**: 201-32.
- Ambrosini G, Adida C, Altieri DC. A novel anti-apoptosis gene, survivin, expressed in cancer and lymphoma. *Nat Med* 1997; **3**: 917-21.
- Tamm I, Kornblau SM, Segall H *et al*. Expression and prognostic significance of IAP-family genes in human cancers and myeloid leukemias. *Clin Cancer Res* 2000; **6**: 1796-803.
- Schimmer AD. Inhibitor of apoptosis proteins: translating basic knowledge into clinical practice. *Cancer Res* 2004; **64**: 7183-90.
- Hunter AM, LaCasse EC, Korneluk RG. The inhibitors of apoptosis (IAPs) as cancer targets. *Apoptosis* 2007; **12**: 1543-68.
- Yonesaka K, Tamura K, Kurata T *et al*. Small interfering RNA targeting survivin sensitizes lung cancer cell with mutant p53 to adriamycin. *Int J Cancer* 2006; **118**: 812-20.
- Lopes RB, Gangeswaran R, McNeish IA, Wang Y, Lemoine NR. Expression of the IAP protein family is dysregulated in pancreatic cancer cells and is important for resistance to chemotherapy. *Int J Cancer* 2007; **120**: 2344-52.
- Bilim V, Yuuki K, Itoi T *et al*. Double inhibition of XIAP and Bcl-2 axis is beneficial for retrieving sensitivity of renal cell cancer to apoptosis. *Br J Cancer* 2008; **98**: 941-9.
- Murakami Y, Kazuno H, Emura T, Tsujimoto H, Suzuki N, Fukushima M. Different mechanisms of acquired resistance to fluorinated pyrimidines in human colorectal cancer cells. *Int J Oncol* 2000; **17**: 277-83.
- Cory AH, Owen TC, Barltrop JA, Cory JG. Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. *Cancer Commun* 1991; **3**: 207-12.
- de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics* 2004; **20**: 1453-4.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; **95**: 14863-8.
- Ghoda LY, Savarese TM, Dexter DL, Parks RE Jr, Trackman PC, Abeles RH. Characterization of a defect in the pathway for converting 5'-deoxy-5'-methylthioadenosine to methionine in a subline of a cultured heterogeneous human colon carcinoma. *J Biol Chem* 1984; **259**: 6715-19.
- Scheithauer W, Rosen H, Kornek GV, Sebesta C, Depisch D. Randomised comparison of combination chemotherapy plus supportive care with supportive care alone in patients with metastatic colorectal cancer. *Br Med J* 1993; **306**: 752-5.
- Colorectal Meta-analysis Collaboration. Palliative chemotherapy for advanced or metastatic colorectal cancer. *Cochrane Database Syst Rev* 2000; CD001545.
- Saltz LB, Cox JV, Blanke C *et al*. Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. *Irinotecan Study Group N Engl J Med* 2000; **343**: 905-14.
- de Gramont A, Figuer A, Seymour M *et al*. Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. *J Clin Oncol* 2000; **18**: 2938-47.
- Ross DT, Scherf U, Eisen MB *et al*. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000; **24**: 227-35.
- Schmidt WM, Kalipciyan M, Dornstauder E *et al*. Dissecting progressive stages of 5-fluorouracil resistance in vitro using RNA expression profiling. *Int J Cancer* 2004; **112**: 200-12.
- Harlin H, Reffey SB, Duckett CS, Lindsten T, Thompson CB. Characterization of XIAP-deficient mice. *Mol Cell Biol* 2001; **21**: 3604-8.
- Inaba M, Mitsuhashi J, Sawada H *et al*. Reduced activity of anabolizing enzymes in 5-fluorouracil-resistant human stomach cancer cells. *Jpn J Cancer Res* 1996; **87**: 212-20.
- Tokunaga Y, Sasaki H, Saito T. Clinical role of orotate phosphoribosyl transferase and dihydropyrimidine dehydrogenase in colorectal cancer treated with postoperative fluoropyrimidine. *Surgery* 2007; **141**: 346-53.
- Leung SY, Chen X, Chu KM *et al*. Phospholipase A2 group IIA expression in gastric adenocarcinoma is associated with prolonged survival and less frequent metastasis. *Proc Natl Acad Sci USA* 2002; **99**: 16203-8.
- Ganesan K, Ivanova T, Wu Y *et al*. Inhibition of gastric cancer invasion and metastasis by PLA2G2A, a novel beta-catenin/TCF target gene. *Cancer Res* 2008; **68**: 4277-86.
- Ito Y, Higashiyama S, Takeda T, Yamamoto Y, Wakasa KI, Matsuura N. Expression of heparin-binding epidermal growth factor-like growth factor in pancreatic adenocarcinoma. *Int J Pancreatol* 2001; **29**: 47-52.
- Itoh Y, Joh T, Tanida S *et al*. IL-8 promotes cell proliferation and migration through metalloproteinase-cleavage proHB-EGF in human colon carcinoma cells. *Cytokine* 2005; **29**: 275-82.
- Suganuma K, Kubota T, Saikawa Y *et al*. Possible chemoresistance-related genes for gastric cancer detected by cDNA microarray. *Cancer Sci* 2003; **94**:

355–9.

- 42 Wang F, Liu R, Lee SW, Sloss CM, Couget J, Cusack JC. Heparin-binding EGF-like growth factor is an early response gene to chemotherapy and contributes to chemotherapy resistance. *Oncogene* 2007; **26**: 2006–16.
- 43 Imamura Y, Itoh M, Maeno Y, Tsukita S, Nagafuchi A. Functional domains of alpha-catenin required for the strong state of cadherin-based cell adhesion. *J Cell Biol* 1999; **144**: 1311–22.
- 44 Matsubara S, Ozawa M. Expression of alpha-catenin in alpha-catenin-deficient cells increases resistance to sphingosine-induced apoptosis. *J Cell Biol* 2001; **154**: 573–84.
- 45 Sasaki H, Sheng Y, Kotsuji F, Tsang BK. Down-regulation of X-linked inhibitor of apoptosis protein induces apoptosis in chemoresistant human ovarian cancer cells. *Cancer Res* 2000; **60**: 5659–66.
- 46 Nomura T, Yamasaki M, Nomura Y, Mimata H. Expression of the inhibitors of apoptosis proteins in cisplatin-resistant prostate cancer cells. *Oncol Rep* 2005; **14**: 993–7.
- 47 Pratt MA, Niu MY, Renart LI. Regulation of survivin by retinoic acid and its role in paclitaxel-mediated cytotoxicity in MCF-7 breast cancer cells. *Apoptosis* 2006; **11**: 589–605.
- 48 LaCasse EC, Cherton-Horvat GG, Hewitt KE *et al.* Preclinical characterization of AEG35156/GEM 640, a second-generation antisense oligonucleotide targeting X-linked inhibitor of apoptosis. *Clin Cancer Res* 2006; **12**: 5231–41.
- 49 Varfolomeev E, Blankenship JW, Wayson SM *et al.* IAP antagonists induce autoubiquitination of c-IAPs, NF-kappaB activation, and TNFalpha-dependent apoptosis. *Cell* 2007; **131**: 669–81.
- 50 Vince JE, Wong WW, Khan N *et al.* IAP antagonists target cIAP1 to induce TNFalpha-dependent apoptosis. *Cell* 2007; **131**: 682–93.

## 正則化正準相関解析を用いた抗がん剤の影響 による共通パスウェイ解析

金蕙鈴<sup>1,2</sup>, 加藤毅<sup>3,2</sup>, 茂櫛薫<sup>1</sup>,  
田中博<sup>1</sup>, 藤沢航<sup>2,1</sup>

患者の QOL(Quality Of Life)における重要性にも関わらず、薬物に由来する作用メカニズムは不明な点が多く、これまででシステマティックなアプローチが極端に少なかった。我々は、汎化能力の高いリッジパラメータを導入した正則化正準相関解析を行い、遺伝子活性とパスウェイの相関関係を解析した。その結果、細胞周期関連パスウェイにおいて薬理的または化学的分類の異なる抗がん剤同士の関連性を見出すことができた。

### Detection of common pathways activated by anticancer drugs using regularized canonical correlation analysis

Hyeryung Kim<sup>1,2</sup>, Tsuyoshi Kato<sup>3,2</sup>, Kaoru Mogushi<sup>1</sup>,  
Hiroshi Tanaka<sup>1</sup> and Wataru Fujibuchi<sup>2,1</sup>

In spite of the importance to improve the quality of life (QOL) of patients, few systematic approaches have been available to uncover molecular mechanisms of drug-action. We adopted a new approach to solve this problem using the regularized canonical correlation analysis which is expected to enhance generalizing capability of finding correlations between gene activities and pathway information. As a result, we found a novel relationship between various anti-cancer drugs that are classified in different pharmacological or chemical groups.

### 1. はじめに

近年のゲノム分野における様々な技術の発展によって、膨大な生物データが蓄積されてきた。DNA チップを用いた遺伝子発現データをはじめ、代謝およびシグナル伝達に関するパスウェイデータ、ChIP-chip データ等がその例である。これらのデータは生体分子に関する情報であることは共通であるものの、転写産物の量、代謝する化合物または遺伝子間相互作用の有無、または DNA と転写因子との相互作用を表すなど、データの内容やプラットフォームは様々である。

異質データ同士のデータマイニング法として正準相関解析法(CCA, *canonical correlation analysis*)が知られている[1]。これは2つのデータセットにおいて互いの相関を最大にする成分を抽出する方法である。近年では典型的な CCA を改良した手法がいくつかが報告されている。例えば、スパース正準相関解析を用いてパスウェイの遺伝子を予測した研究[2]、カーネル正準相関解析による大腸菌のオペロン予測問題に適用した例[3]など、いずれもプラットフォームの異なる複数のデータのセットの相関関係に注目している。特に、正準相関解析にカーネル法を適用した方法[4, 5]では、カーネルの選択によって多変量間の非線形的な関係を特徴空間に射影して線形関係を解析することができた。また正則化項が導入されており高次元データで起こる不良設定問題が解決されている。そのため、高次元かつ複雑な構造を持つ生物データを解析する方法として極めて有効であると考えられる。

一方、J. Lambら(2006)は、様々な薬物による影響を比較した遺伝子発現データベースを発表した[6]。この研究では薬の種類、投与量、細胞の種類などの条件の類似関係を比較している。しかし、薬の作用機序を理解し、副作用の原因を把握するためには、条件の比較に加えてそれに関わる詳細な分子メカニズムの同定が必要である。また、作用機序の不明な薬については分子レベルでの潜在的な影響を調べることが要求されている。

そこで本研究では、カーネル正準相関解析法という優れたアプローチによって、抗がん剤の暴露によって活性化される共通パスウェイを検出することを目的とした。

<sup>1</sup> 東京医科歯科大学 大学院生命情報科学教育部  
<sup>2</sup> 産業技術総合研究所 生命情報工学研究センター  
<sup>3</sup> お茶の水女子大学 生命情報科学教育部

<sup>1</sup> School of Biomedical Sciences, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan

<sup>2</sup> National Institute of Advanced Industrial Science and Technology (AIST), Computational Biology Research Center, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

<sup>3</sup> Ochanomizu University, Center for Informational Biology, 2-1-1 Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan



2. 正則化正準相関解析

(1) 正準相関解析

正準相関解析とは対象となる2組の変数群において、それぞれの合成変数(=正準変数)を作り、その合成変数間の相関係数が最大になるような重み係数を求める多変量解析手法である。

ここで、2つのデータ行列  $X_a \in \mathbb{R}^{N \times r}$ ,  $X_b \in \mathbb{R}^{N \times (p-r)}$  があるとし、このデータに対する正準変数を、

$$f_a = (X_a)^T w_a, \quad f_b = (X_b)^T w_b$$

とする。但し、 $w_a \in \mathbb{R}^r$  と  $w_b \in \mathbb{R}^{p-r}$  は正準相関解析のパラメータである。ここで、2つの合成変数間の相関係数は、

$$\rho(f_a, f_b) = \frac{\langle f_a, f_b \rangle}{\|f_a\| \|f_b\|}$$

と表される。但し、 $\langle f_a, f_b \rangle$  は内積を意味する。

この  $\rho(f_a, f_b)$  を、

$$w_a^T X_a X_a^T w_a = w_b^T X_b X_b^T w_b = 1$$

のもとで最大にする  $w_a \in \mathbb{R}^r$  と  $w_b \in \mathbb{R}^{p-r}$  を求める。但し、最大の相関係数に対する正準変数を第1正準変数、次に大きい相関係数に対する正準変数を第2正準変数という。

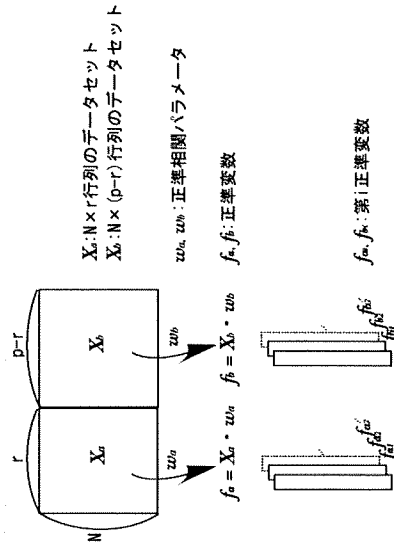


図1 正準相関解析の概念図

(2) リッジ正準相関解析

リッジ正準相関解析では、2つの正則化パラメータ  $\gamma_a, \gamma_b$  を用い、正則化相関係数、

$$\rho(f_a, f_b) = \frac{\langle f_a, f_b \rangle}{\sqrt{\|f_a\|^2 + \gamma_a w_a^T w_a} \sqrt{\|f_b\|^2 + \gamma_b w_b^T w_b}}$$

を最大する  $w_a \in \mathbb{R}^r$  と  $w_b \in \mathbb{R}^{p-r}$  を求める方法である。但し、 $w_a^T X_a X_a^T w_a + \gamma_a w_a^T w_a = w_b^T X_b X_b^T w_b + \gamma_b w_b^T w_b = 1$  とする。正則化パラメータは、所与のデータへの過剰なフィッティングを防ぐ。

(3) カーネル正準相関解析

カーネル正準相関解析は、基本的にリッジ正準相関解析にカーネル法を適用したものである。カーネル法では、ある写像  $\phi_a, \phi_b$  が存在すると仮定して、

$$K_a(x_a, x_a) = \langle \phi_a(x_a), \phi_a(x_a) \rangle, \quad K_b(x_b, x_b) = \langle \phi_b(x_b), \phi_b(x_b) \rangle$$

を満たすような関数  $K_a, K_b$  を用いる。このような関数はカーネル関数と呼ばれる。カーネル関数によって非線形な成分の抽出が可能となる。このカーネル関数を用いると正準変数は、

$$f_a = \phi_a(X_a)^T w_a = K_a \alpha_a, \quad f_b = \phi_b(X_b)^T w_b = K_b \alpha_b$$

と表される。但し、 $\alpha_a$  および  $\alpha_b$  はカーネル正準相関解析のパラメータである。正則化相関係数は、

$$\rho(f_a, f_b) = \frac{\alpha_a^T K_a K_b \alpha_b}{\sqrt{\alpha_a^T K_a^2 \alpha_a + \gamma_a \alpha_a^T \alpha_a} \sqrt{\alpha_b^T K_b^2 \alpha_b + \gamma_b \alpha_b^T \alpha_b}}$$

と表される。但し、 $\alpha_a^T K_a^2 \alpha_a + \gamma_a \alpha_a^T \alpha_a = \alpha_b^T K_b^2 \alpha_b + \gamma_b \alpha_b^T \alpha_b = 1$  であるとする。

正確には、分母の  $\alpha_a^T K_a \alpha_a$  は  $\alpha_a^T K_a \alpha_a$ ,  $\alpha_b^T K_b \alpha_b$  は  $\alpha_b^T K_b \alpha_b$  となればリッジ正準相関解析をカーネル化したものにならない。しかし、そうすると正則化相関係数を最大化するときに特異行列が現れてしまい数値計算との問題が生じる。故に、正則化項として  $\alpha_a^T \alpha_a$  および  $\alpha_b^T \alpha_b$  に正則化パラメータをかけたものを用いた[5]。

### 3. 実験データ

#### 3.1 遺伝子発現データ

##### (1) データの収集

攪乱物質をヒト培養細胞に暴露した遺伝子発現データを The Connectivity Map [6] のデータベースである <http://www.broadinstitute.org/cmap/> よりダウンロードした。また、その化合物が抗がん剤かどうかの判別は、STITCH [7] のデータベースにある化合物と ATC (Anatomical Therapeutic Chemical Classification System) 分類の対応表を用いて行った。ATC 分類とは、解剖治療化学分類法と呼ばれ医薬品の分類に用いられており、特に L01 で始まるコードの医薬品は抗悪性腫瘍の治療に使用されている。具体的な抗がん剤の名前や ATC コードおよび薬理学的分類等を付録 A に表した。

- ◆ 暴露物質：抗がん剤 16 種
- ◆ 暴露対象：ヒト培養細胞 MCF7 (乳癌培養細胞), PC3 (前立腺癌培養細胞) および HL60 (骨髄性白血病) の 3 種類
- ◆ マイクロアレイの種類：GeneChip® HT Human Genome U133 Array (Affymetrix)
- ◆ プローブセットの数：22277 個
- ◆ 薬物処理したサンプル数：66 トリートメント (薬物の種類、濃度、培養細胞の種類による分類)
- ◆ コントロールサンプルの数：249 コントロール (一つのトリートメントに対して細胞の種類が同一な 1~6 つの溶媒コントロールが存在)
- ◆ 暴露時間：6 時間

##### (2) 前処理

遺伝子発現データの前処理は、RMA (Robust Multichip Average) 法によって行った [10]。その後、溶媒コントロールに対するトリートメントの発現量の比 (=fold-change) を式(1)より求めた。

$$\text{fold-change} = \log_2 (V_i / V_0) \dots\dots\dots (1)$$

( $V_i$ : トリートメントの発現値,  $V_0$ : コントロールの発現値)

但し、一つのトリートメントに対して複数のコントロールが存在した場合は、コントロール同士の発現量の幾何平均値を用いて比を計算した。fold-change を 66 対のトリートメントおよびコントロールの 22277 個の遺伝子に対してそれぞれ計算し、データセット  $X_i$  とした。また、遺伝子の ID としては Entrez Gene ID を用いた。Entrez Gene ID は 13524 種類あり、複数のプローブセット ID が対応した場合は四分位数範囲 (IQR) を

計算して、サンプル間においてばらつきの大いプローブセット ID のみを対応づけした。

#### 3.2 遺伝子パスウェイ

遺伝子のパスウェイ情報として KEGG パスウェイデータを用いた。現在分類されているヒトのパスウェイおよび遺伝子の種類はそれぞれ 197 個, 5185 個であった。これをもとにパスウェイに対してその遺伝子が含まれている場合は 1, そうでない場合は 0 とバイナリデータとして作成し、データセット  $X_0$  とした。遺伝子 ID としては 3.1(2)と同様に Entrez Gene ID を用いた。最終的に  $X_0$  および  $X_i$  において共通な Entrez Gene ID は 4287 個あった。

### 4. 実験

#### 4.1 カーネル正準相関解析

上記の方法によって、作られた  $X_0$  および  $X_i$  の最終的なデータセットの詳細を表 1 に示した。

表 1 データセットの内容

	$X_0$	$X_i$
データの種類	パスウェイ	遺伝子発現プロファイル
行の情報 (行数)	遺伝子 (4287)	遺伝子 (4287)
列の情報 (列数)	パスウェイの ID (197)	トリートメント (66)

#### 4.2 パラメータの調整

##### (1) 正則化パラメータ

正則化パラメータ  $\gamma_0$  および  $\gamma_i$  の値を決定するために、10-fold クロスバリデーション法を用いた。候補の値として  $\gamma_0 = 0.01, 0.1, 1, 10, 100, 1000, 10000$  および  $\gamma_i = 0.01, 0.1, 1, 10, 100, 1000, 10000$  を選び、すべての組み合わせに対して次の手順を実行した。

データセットを 10 グループに分割し、9 グループを訓練用、1 グループを評価用とした。訓練用データから  $\alpha_0$  および  $\alpha_i$  を求め、そのパラメータを使って得られる評価用データの正準変量間の相関係数を計算した。これをローテーションして 10 回繰り返し、10 個の相関係数の平均を算出した。これを ( $\gamma_0, \gamma_i$ ) のすべての値の組み合わせに対して実行し、最も最大の固有値に対する正準変数 (第 1 正準変数) の平均相関係数が高かった  $\gamma_0 = 0.1, \gamma_i = 1$  を正則化パラメータとして採用した。

(2) **RBF カーネルパラメータ**

本研究では, RBF(Radial basis function; 動径基底関数)関数をカーネル関数として用いた. また, この RBF カーネルのパラメータ  $\sigma_a$  および  $\sigma_b$  は,  $X_a$  および  $X_b$  のそれぞれのデータセットにおける遺伝子間のユークリッド距離を表す距離行列  $D_a \in R^{237 \times 237}$  および  $D_b \in R^{4287 \times 4287}$  を作成し, それらの要素の平均値である  $\sigma_a = 2.19$ ,  $\sigma_b = 2.95$  をパラメータとして使用した[9].

**5. 結果**

**5.1 第1正準変数と相関の高い変数(r=0.98)**

第1正準変数に対する評価用データセットの変数同士の相関係数は, 0.98 であった. さらに, 求められた正準変数と最初の変数群(パスウェイ変数=197種類, トリートメント=66種類)とのピアソンの積率相関係数(=構造係数, S-score)を計算し, 互いの相関係数の傾向と有意にプラスの相関またはマイナスの相関している変数を求めた. その結果, Ascorbate and alda rate metabolism と Pentose and glucur onate interconversions パスウェイは最も S-score が高く, 強い相関関係にあることを示唆した. また, このパスウェイと含め, Porphyrin and chloro phyll m etabolism や Androgen and estro gen metabolism などのパスウェイ変数と, ニトロソ尿素系のアルキル化剤である carmustine, semustine が正の相関関係にあった. 一方, Streptozotocin, Altretam ine 等とは負の相関を示した.

表 2 第1正準変数と相関の高いパスウェイ変数(S-score  $\geq$  0.20)

パスウェイ ID†	パスウェイ名	S-score
00053	Ascorbate and aldarate metabolism	-0.84
00040	Pentose and glucuronate interconversions	-0.80
00860	Porphyrin and chlorophyll metabolism	-0.64
00150	Androgen and estrogen metabolism	-0.60
00500	Starch and sucrose metabolism	-0.58
00983	Drug metabolism - other enzymes	-0.58
00830	Retinol metabolism	-0.53
00980	Metabolism of xenobiotics by cytochrome P450	-0.51
00982	Drug metabolism - cytochrome P450	-0.49

†パスウェイ ID は KEGG パスウェイにおけるマップ ID を表す

表 3 第1正準変数と相関の高いトリートメント変数(S-score 上位 10 個)

トリートメント ID†	S-score	化合物名
6888	-0.18	carmustine
7487	-0.16	semustine
7540	-0.15	semustine
6098	0.12	Streptozotocin
4627	0.11	Altretamine
5571	0.10	retinoic acid
6681	0.10	Etoposide
6914	-0.10	carmustine
7089	-0.10	lomustine
7050	-0.095	lomustine

†トリートメント ID として Connectivity map のデータベースの認識コードを用いた

**5.2 第2正準変数と相関の高い変数(r=0.44)**

第2正準変数に対して高い相関を示しているパスウェイ変数としては, 細胞周期, DNA 複製などが得られた. これらに対し, 正の相関を示したのは Streptozotocin であった. 負の相関を示したのは paclitaxel, daunorubicin HCl 等であった.

表 4 第2正準変数と相関の高いパスウェイ変数(S-score  $\geq$  0.20)

パスウェイ ID	パスウェイ名	S-score
04110	Cell cycle	0.32
03030	DNA replication	0.26
04080	Neuroactive ligand-receptor interaction	-0.25
04060	Cytokine-cytokine receptor interaction	-0.25
03040	Spliceosome	0.25
00240	Pyrimidine metabolism	0.22
03010	Ribosome	0.20
04120	Ubiquitin mediated proteolysis	0.20
03420	Nucleotide excision repair	0.20

表 5 第 2 正準変数と相関の高いトリートメント変数(S-score 上位 10 個)

トリートメント ID	化合物名	S-score
2535	Streptozotocin	0.30
7193	Streptozotocin	0.20
6720	Paclitaxel	-0.18
7050	lomustine	-0.17
7507	daunorubicin HCl	-0.17
3241	Etoposide	-0.15
5583	nordihydroguaiaretic acid	0.14
1636	retinoic acid	-0.14
5688	Altretamine	-0.13
5320	Paclitaxel	-0.13

期, DNA 複製などの細胞周期関連パスウェイが過剰に働いているのが, Streptozotocin と正の相関を示した原因として考えられる。

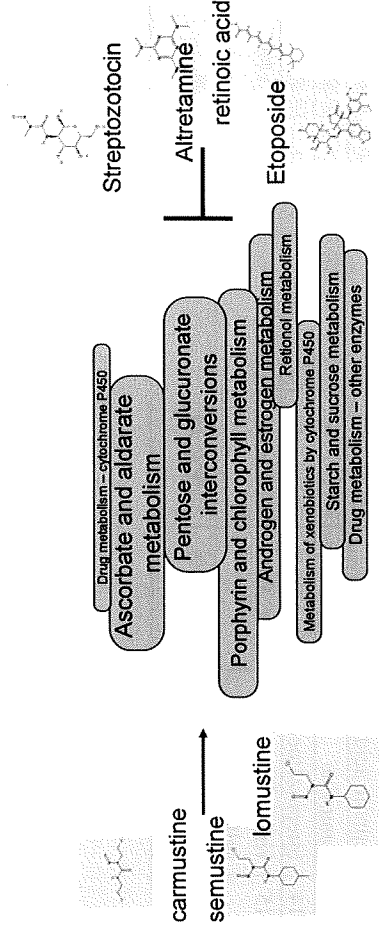


図 2 抗がん剤とパスウェイの相関関係の仮説図 (第 1 正準変数関連)

## 6. 考察

まず, 第 1 正準変数と相関の高いトリートメント変数を検討したところ, ニトロソ尿素系のアルキル化剤とそれ以外の薬剤との相違が見られた。ニトロソ尿素系のアルキル化剤である carmustine, semustine および lomustine ではビタミン代謝, ホルモン代謝, 糖代謝, 薬物代謝など生体全般的な代謝関連のパスウェイの促進作用が見られるが, 逆にそれ以外の Streptozotocin, Altretamine, retinoic acid および Etoposide では, これらのパスウェイが全体的に抑制されている。これらのパスウェイは, 細胞が取り入れた外部の物質を分解し, 必要なものを生成するために必要な生体反応である。したがって, 今回の解析に用いたニトロソ尿素系のアルキル化剤以外の抗がん剤においてはこれらのパスウェイの抑制が作用機構となっている可能性がある。

次に, 第 2 正準変数との相関の高い変数について考察する。それぞれの変数群において, 第 1 正準変数と比べて S-score は低いものの, 特徴的なパスウェイ群との相関関係が見られた。例えば, 細胞周期, DNA 複製, スプラインソーム, ペリミジン代謝, リボソーム, ユビキチン媒介蛋白質分解, スクレオチド除去修正は細胞増殖に関わるものであるが, いずれも第 2 正準変数と負に相関している。したがって, これらは薬剤の抗腫瘍活性に関わっていると考えられる。一方で, Streptozotocin はこれらのパスウェイと負の相関を示した。Streptozotocin は化合物分類ではニトロソ尿素系であるが, 薬理学的分類ではアルキル化剤ではなく抗生物質に含まれる。また, 日本ではこの薬剤が主に動物実験だけで使われており, 発がん物質として分類されるケースもあるなど非常に細胞毒性が強い物質である。そのため, 暴露された細胞が損傷され, 細胞周

## 7. おわりに

我々は, 正則化項およびカーネル法を導入した正準相関解析法を用いて, 抗がん剤の影響によって活性または抑制されるパスウェイを同定した。まだ, 正準相関解析は遺伝子発現などのオミックスデータ解析に応用された例が少なく, 結果の解釈, 性能の比較の方法が定着していない傾向がある。本研究では, 解釈において S-score を用い, 正準変数と各変数の相関係数を調べることにより, 変数同士がどのような正または負の相関関係とどの程度合いについて意味づけすることが出来た。今後の課題として, さらにカーネル関数の種類を検討し, それによって抽出される正準変数の特徴を比較していく方針である。

**謝辞** 薬物関連のデータベースにおいて助言をいただいた東京医科歯科大学の高井貴子先生, その他ご協力頂いた皆様に, 心から感謝の意を表します。

参考文献

- 1) W.T. Anderson, An introduction to multivariate statistical analysis, John Wiley & Sons. 1984.
- 2) S. Waajenberg and A.H.Z., *Sparse canonical correlation analysis for identifying, connecting and completing gene-expression networks*. BMC Bioinformatics, 2009, 10(315).
- 3) Y. Yamanishi, J.V., A. Nakaya and M. Kanehisa, *Extraction of correlated gene clusters from multiple genomic data by generalized kernel canonical correlation analysis*. Bioinformatics, 2003, 19.
- 4) S. Akaho, *A kernel method for canonical correlation analysis*. Workshop on information-based induction sciences (IBIS2000), 2000.
- 5) Kuss, M. and T. Graepel, *The Geometry of kernel canonical correlation analysis*. 2003.
- 6) J. Lamb et al., *The Connectivity Map: a new tool for biomedical research*. Nat Rev Cancer, 2007, 7(1): p. 54-60.
- 7) M. Kuhn, D.S., A. Franceschini, M. Campillos, C.V. Mering, L.J. Jensen, A.B. and P. Bork, *STITCH 2: an interaction network database for small molecules and proteins*. Nucleic Acids Research, 2009.
- 8) NCBI PUBCHEM:<http://pubchem.ncbi.nlm.nih.gov/>
- 9) T. Kato, H. Kashima., M. Sugiyama and K. Asai, *Conic Programming for Multi-Task Learning*. IEEE Transactions on Knowledge and Data Engineering. Accepted.
- 10) R. A. Irizarry, B.H., F. Collin, Y. D. Beazer-Barclay, K.J. Antonellis, U. Scherf and T. P. Speed, *Exploration, normalization, and summaries of high density oligonucleotide array probe level data* Biostatistics, 2003, 4(2): p. 249-264.

付録

付録 A 薬の名前と CID, ATC コード, 病理学のおよび化合物の分類

番号	化合物名	ATC 分類 コード [7]	病理学的分類[8]	化合物分類[8]
1	Altretamine	L01XX03	抗がん剤, アルキル化剤	複素環式化合物
2	carmustine	L01AD01	抗がん剤, アルキル化剤	ニトロソ尿素
3	Chlorambucil	L01AA02	抗がん剤, アルキル化剤	ナイトロジェンマ スタード
4	Dacarbazine	L01AX04	抗がん剤, アルキル化剤	複素環式化合物
5	daunorubicin HC1	L01DB02	抗がん剤, 抗生物質	多環式炭化水素, 芳香族
6	Etoposide	L01CB01	抗がん剤, 植物性	多環式炭化水素, 芳香族
7	Ifosfamide	L01AA06	抗がん剤, アルキル化剤	ナイトロジェンマ スタード
8	Isotretinoin	L01XX22	抗がん剤, 腫科用薬物	環状炭化水素
9	lomustine	L01AD02	抗がん剤, アルキル化剤	ニトロソ尿素
10	Methotrexate	L01BA01	代謝拮抗薬, 抗がん剤	複素環式化合物
11	Paclitaxel	L01CD01	抗がん剤, 植物性	環状炭化水素
12	retinoic acid	L01XX22	抗がん剤	環状炭化水素
13	semustine	L01AD03	抗がん剤, アルキル化剤	ニトロソ尿素
14	Streptozotocin	L01AD04	抗がん剤, 抗生物質	ニトロソ尿素
15	vinblastine sulfate	L01CA01	抗がん剤, 植物性	複素環式化合物
16	nordihydroguaiar etic acid	L01XX10	抗酸化剤	環状炭化水素

## 細胞分化解析を目指した 網羅的ヒト細胞データベース「CELLPEDIA」

幡野晶子<sup>1,2</sup>, Harry Amri Moesa<sup>1,3</sup>, 千葉啓和<sup>1</sup>,  
谷口丈晃<sup>4</sup>, 永家聖<sup>1,2</sup>, 山根木康嗣<sup>5</sup>, 藤渕航<sup>1</sup>

本研究では、細胞を人体の物理的な位置に基づいて 2000 種類以上に細分化し、細胞画像、形態値、文献、遺伝子発現データを収集し統合した、細胞の網羅的データベースを作成した。さらに成体の組織幹細胞、前駆細胞を約 60 種類に分類し、情報を収集した。本データベースを用いることにより、細胞の形態値、遺伝子発現データをもとに、細胞の個性を定量的に解析することが可能である。

### CELLPEDIA: Comprehensive human cell database toward cell differentiation analysis

Akiko Hatano<sup>1,2</sup>, Harry Amri Moesa<sup>1,3</sup>, Hirokazu Chiba<sup>1</sup>,  
Takeaki Taniguchi<sup>4</sup>, Satoshi Nagaie<sup>1,2</sup>, Koji Yamanegi<sup>5</sup>,  
Wataru Fujibuchi<sup>1</sup>

In this research, we construct a comprehensive database for human cells, in which various types of information including cell images, morphological measurements, related journal articles and gene expression data are collected and integrated. In this database, the human cells are subdivided into more than 2,000 cells based on their physical locations. Similarly, adult stem cells and progenitor cells are classified into about 60 kinds of cells, and their related information is collected. By using our database, cell characteristics can be quantitatively analyzed.

## 1. はじめに

ヒトには 200 種類以上、約 60 兆個の細胞が存在していると考えられ、その構造と機能を理解することは、人体の正常な機能、疾患メカニズムの解明において重要である。さらに同じ名称がついている細胞であっても、細胞の年齢、周囲の環境、人体における物理的存在位置等の要因により、性質が異なる可能性がある。このような細胞の個性を認識することは生体を理解するために必要不可欠である。そのため我々は成体に含まれる正常細胞、幹細胞を物理的位置に基づいて細分化するとともに、培養細胞を加えた 3 種類の細胞カテゴリーを設け、ヒト細胞に関する網羅的データベースを作成した。これにより、今後得られるであろう 1 細胞レベルのデータにも対応できる、細胞の電子辞書的役割を備えた細胞のデータベースの提供を可能とした。

## 2. CELLPEDIA の構成

CELLPEDIA は細胞分類表、細胞画像、文献、遺伝子発現データの 4 つの基本情報と、細胞画像に対してそれらを集約し、細胞画像の形態値情報、オントロジーを加えた統合ページから構成されている。さらに遺伝子発現データを用いて作成した遺伝子モジュール辞書、キーワードによる検索機能を付随している。トップページと統合ページをそれぞれ図 1 に示す。また構成内容に関して以下に順を追って説明する。

<sup>1</sup> 産業技術総合研究所 生命情報工学研究センター  
<sup>2</sup> 東京医科歯科大学大学院 歯学総合研究科  
<sup>3</sup> 東京大学医学部研究所 ヒトゲノム解析センター  
<sup>4</sup> (株)三菱総合研究所  
<sup>5</sup> 兵庫医科大学 病理学講座 機能病理部門  
<sup>1</sup> Computational Biology Research Center, Advanced Industrial Science and Technology  
<sup>2</sup> Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University  
<sup>3</sup> Human Genome Center, Institute of Medical Science, University of Tokyo  
<sup>4</sup> Mitsubishi Research Institute Inc  
<sup>5</sup> Department of Clinical Pathology, Hyogo College Of Medicine

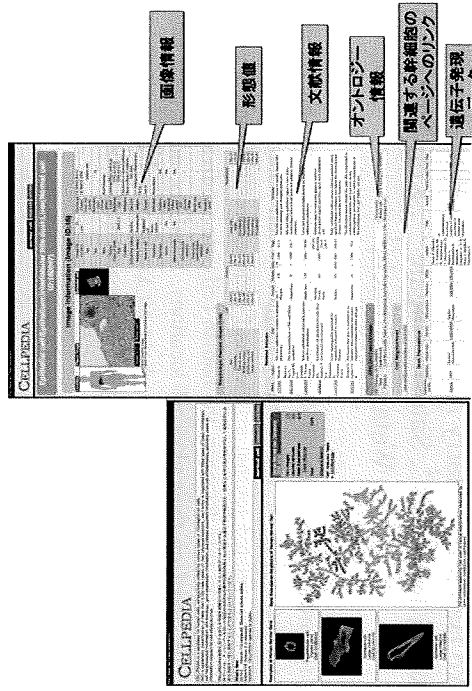


図 1 トップページ (左) と細胞情報統合ページ (右)

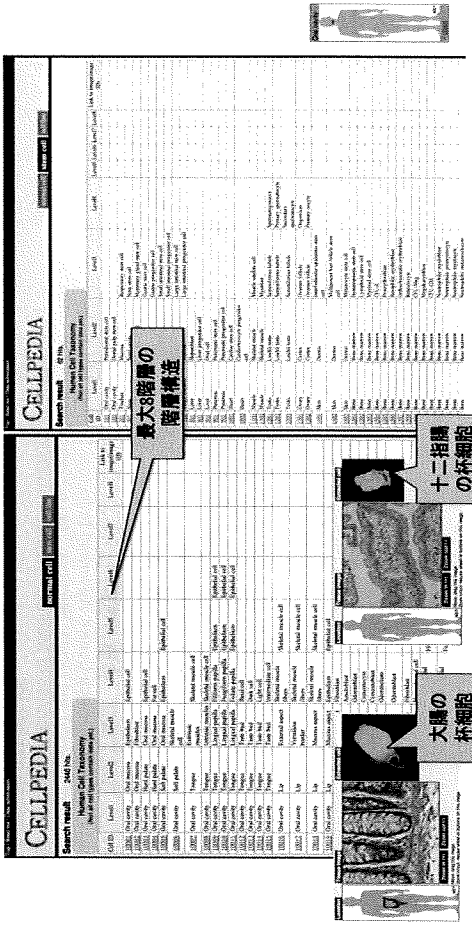


図 2 細胞の分類表、物理的存在位置の異なる杯細胞の例 (左) と幹細胞の分類表 (右)

## 2.1 細胞分類

現在分類されている細胞は、十二指腸や大腸に存在する杯細胞のように、存在する環境、すなわち物理的存在位置が異なっても、同じ名称がつけられている細胞も少なくない。そのような細胞と物理的存在位置の例を表 1 に示す。これらの細胞は遺伝子発現レベルでは、異なる性質を持つ可能性がある。そのため我々は、今後 1 細胞レベルでの細胞機能解析が可能となる時代に対応できる、詳細な細胞分類表を作成した。具体的には、人体の物理的存在位置に基づき、各器官または組織から細胞までを最大 8 階層まで階層化し、細胞を 2446 種類に細分化した。さらに幹細胞に関する 62 種類に分類し、同様に情報を収集した (図 2)。

表 1 細胞の物理的存在位置数

細胞名	物理的存在位置数
fibroblast	297
smooth muscle cell	264
epithelial cell	122
endothelial cell	118
nerve cell	44
goblet cell	31

## 2.2 細胞画像

本研究では、物理的存在位置に基づいて分類した、ヒト細胞カタログに対応する各細胞画像を収集し、細胞の電子辞書的作用を持つデータベース構築を目標としている。現在の画像登録数は、組織画像を含め正常細胞が 174 件、幹細胞が 30 件である。各画像に関して、画像登録者が提供した情報を記載した (図 3)。顕微鏡の倍率に関しては今後追加する予定である。

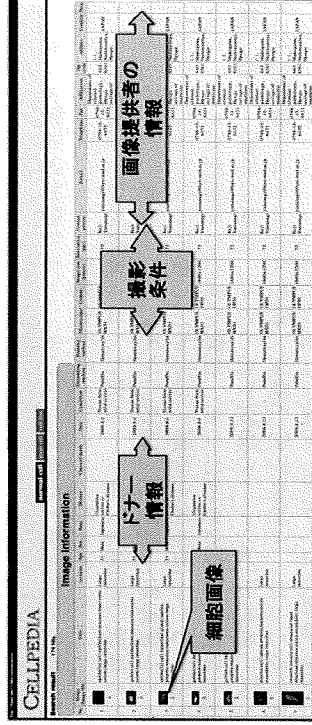


図 3 撮影装置や画像登録者情報を記載した画像情報のページ

### 2.3 文献データ

各細胞の生理学的機能、特異的に発現している遺伝子等の特徴を記載している論文を検索し、正常細胞に関して324件、幹細胞に関して76件の論文を収集した。雑誌名、著者名等の情報を表示するとともに、Pubmedや本データベース中の関連する細胞の統合ページにリンクを張ってある。さらに医学をバックグラウンドとする研究者が論文を読み、内容を要約して表示することにより、細胞の特徴を理解しやすくした(図4)。

No.	PMID	Author	Title	Journal	Volume	Issue	Year	Summary
1	13312	Harrod JR	The C cells of the thyroid gland and their modulatory thyroid hormones.	Am J Physiol	48	1	1977	The primary sensory neurons of the parathyroid gland in the parathyroid gland of the rat. The parathyroid gland is a specialized endocrine gland that secretes parathyroid hormone (PTH) which regulates calcium metabolism. The parathyroid gland is composed of parathyroid cells and parathyroid chief cells. The parathyroid gland is a specialized endocrine gland that secretes parathyroid hormone (PTH) which regulates calcium metabolism. The parathyroid gland is composed of parathyroid cells and parathyroid chief cells.
2	22213	Komuro H	Unusual marks in health and disease.	Dynamics	17	3	1979	Unusual marks in health and disease. The parathyroid gland is a specialized endocrine gland that secretes parathyroid hormone (PTH) which regulates calcium metabolism. The parathyroid gland is composed of parathyroid cells and parathyroid chief cells.
3	33320	Kameda Y, Ito A	Cell specificities in immunoreactive parathyroid cells.	Histochemistry	46	2	1979	Cell specificities in immunoreactive parathyroid cells. The parathyroid gland is a specialized endocrine gland that secretes parathyroid hormone (PTH) which regulates calcium metabolism. The parathyroid gland is composed of parathyroid cells and parathyroid chief cells.
4	22222	Morse WJ, Levy GJ et al.	Morphological observations on the parathyroid gland.	Am J Rec	192	1	1979	Morphological observations on the parathyroid gland. The parathyroid gland is a specialized endocrine gland that secretes parathyroid hormone (PTH) which regulates calcium metabolism. The parathyroid gland is composed of parathyroid cells and parathyroid chief cells.
5	22223	Morse WJ, Levy GJ et al.	Ultrastructure of the parathyroid gland.	Am J Anat (Hum)	96	1	1977	Ultrastructure of the parathyroid gland. The parathyroid gland is a specialized endocrine gland that secretes parathyroid hormone (PTH) which regulates calcium metabolism. The parathyroid gland is composed of parathyroid cells and parathyroid chief cells.

図4 文献情報

### 2.4 遺伝子発現データ

公共のデータベース GEO[1]、ArrayExpress[2]から正常細胞に関して878件、幹細胞に関して92件の遺伝子発現データを収集し、組織特徴的遺伝子、発現分布、ヒートマップと遺伝子リスト、論文や実験情報を表示した。現在1細胞レベルのデータが少ないため、組織あるいは複数の細胞から得られたデータがほとんどとなっている。1枚のアレイ上で発現量の多い上位10遺伝子、発現量の少ない下位10遺伝子を表示するとともに、発現量の分布をヒストグラムで表した。さらに全発現データを用いて、自己組織化マップ(SOM)を作成することにより、発現量のビジュアル化を可能とした(図5)。ヒストグラム、SOMの図をクリックするとその値を示す遺伝子リストが表示される。

図5 遺伝子発現データのページ

### 2.5 形態値

各組織画像から、画像中に含まれる1細胞を対象に、独自に開発したツールであるCytometorica[3]を用いて、細胞の形態値を測定した。多くがHE染色によるノイズの大きい画像であり、NIH-Image、CellProfilerのような他の測定技術では、細胞の輪郭を認識することが困難であった。しかしCytometoricaを用いることにより、ほとんどのHE染色画像に対して形態値を得ることができた。測定した形態値は、細胞と細胞核の面積、凸包面積、円形度、慣性モーメント、bending energy、楕円長軸長、細胞と細胞核の面積比率、細胞の主軸核の比、細胞内における核の偏在性である(図6)。

Morphologic Feature (Shape ID:127)		Cell	Nucleus	Nucleus:Cell
Area	8.88e-02	3.65e-02	4.11e-01	1.99e-02
Circularity	6.23e-01	8.58e-01	Angle	4.47e-02
Convex area	1.05e-03	4.08e-02	Eccentricity1	3.87e-02
Bending energy	2.00e-02	1.31e-00	Eccentricity2	7.19e-02
Elliptic long axis	3.66e-01	2.38e-01	Eccentricity3	5.94e-02
Inertia moment	1.29e-05	2.18e-04	Eccentricity4	8.33e-02
			Eccentricity5	

図6 Cytometoricaにより測定された細胞の形態値



### 2.6 OBO cell type ontology

細胞のオントロジー情報を提供している The open biomedical ontology [4]に掲載された細胞名と、本データベース中の細胞名との関連づけを行った。正常細胞に関して 87 件、幹細胞 42 件に対し、オントロジー情報を表示した。

### 2.7 遺伝子のモジュール辞書

我々が開発した遺伝子モジュール高速マイニングシステム、SAMURAI [5]を用い、収集した遺伝子発現データからマイクロアスタリダ法により、ヒト正常細胞間の遺伝子モジュールデータを作成した。遺伝子モジュールを抽出することにより、組織特異的、あるいは異なる組織で共通する遺伝子発現パターンの両方を探索することが可能である。各モジュールのページには KEGG パスウェイ解析、遺伝子オントロジーに関する情報を記載しており、細胞接着性、代謝パスウェイ等、モジュールが関連している生物学的機能を確認することができる。

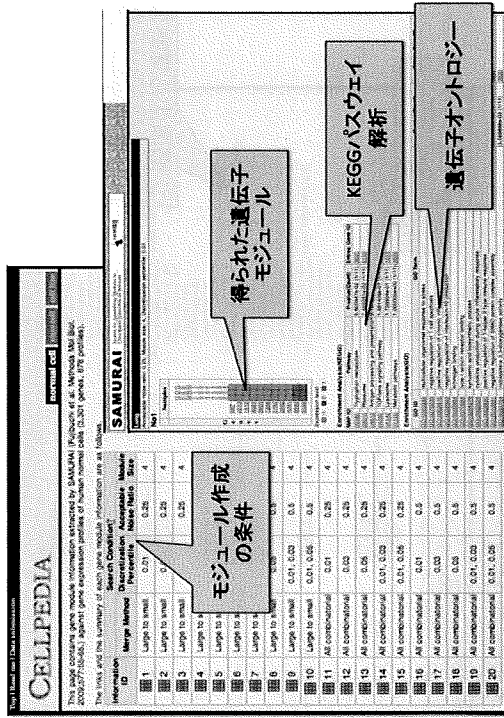


図7 遺伝子モジュールのリスト (左) と各モジュールの詳細ページ (右)

### 3. CELLPEDIA の応用 (細胞の形態値、遺伝子発現データを組み合わせた細胞の分化解析)

細胞は同じ組織中に存在する、同種の細胞であっても、細胞周囲の環境や、細胞の年齢等、様々な要因により、遺伝子発現や細胞の形態に違いが生じていると考えられる。従って本研究で測定した形態値は、個々の細胞の状態を反映した値と捉えること

ができる。形態値を用いた解析例の1つとして、形態値と遺伝子発現データを組み合わせ、細胞の形態に関連する遺伝子の同定が挙げられる。さらに応用として、細胞が成熟するまでの各段階の形態値を抽出し、整理することにより、形態値情報に基づいた細胞の分化状態の予測や、遺伝子モジュールとの組み合わせによる細胞の分化転換経路の探索等、細胞の分化解析を目指している。

### 4. おわりに

本研究では、画像、論文、遺伝子発現等、細胞に関する情報を網羅的に集め、これら膨大なデータを収容することができる、ヒト細胞統合データベース「CELLPEDIA」を構築した。今後は細胞画像の充実に加えて、細胞分化や分化転換情報、細胞レベルの遺伝子発現データを増やし、多くの研究者に利用される「電子細胞辞書」的役割を果たすデータベースを構築する予定である。さらに独自に抽出した形態値情報と収集した遺伝子発現データを組み合わせることにより、細胞分化解析を可能とするデータベースを目指している。

### 参考文献

- 1) Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>
- 2) ArrayExpress, <http://www.ebi.ac.uk/microarray-as/ae/>
- 3) 細胞輪郭抽出装置、細胞輪郭抽出方法およびプログラム 特願 2006-331416、2006
- 4) The Open Biomedical Ontology, <http://www.obofoundry.org/>
- 5) Fujibuchi et al. *Methods Mol Biol.* 2009;577:55-65.

## 2-way prediction 法による GPCR リガンドの結合予測

百石 弘 澄<sup>†1</sup> 杉原 稔<sup>†2</sup> 諏訪 牧 子<sup>†2,†1</sup>  
加藤 毅<sup>†3,†2</sup> 山名 早 人<sup>†1</sup> 藤 洸 航<sup>†2,†1</sup>

Gタンパク質共役型受容体 (GPCR) は創薬研究の中心となっており、GPCR と相互作用する未知のリガンドを予測することは重要な課題である。しかしながら、相互作用するリガンドが全く知られていない GPCR も多く存在し、これらの GPCR と相互作用するリガンドを予測することは学習用サンプルの不足から困難となる。そこで、我々はサポータベクターマシンを利用した 2-way prediction 法を提案する。この方法では、リガンド、GPCR、双方から予測を行い、相互作用するリガンドの情報と GPCR の情報が全くない場合の予測に対応した。実験により提案手法の有効性を示す。

## Prediction of GPCR ligands by 2-way prediction method

HIROTO HYAKKOKU,<sup>†1</sup> MINORU SUGIHARA,<sup>†2</sup>  
MAKIKO SUWA,<sup>†2,†1</sup> TSUYOSHI KATO,<sup>†3,†2</sup>  
HAYATO YAMANA<sup>†1</sup> and WATARU FUJIBUCHI<sup>†2,†1</sup>

G-protein coupled receptors (GPCRs) are important pharmacological targets and to predict unknown interactions between GPCRs and ligands is one of the most interesting topics in the current computational biology. However, ligands of many GPCRs are experimentally not identified yet and it is difficult to predict unknown ligands of GPCRs because of insufficiency of training data set. We have developed a 2-way prediction method based on the support vector machine. In this method, the prediction is performed by using both information of ligands and GPCRs and one can apply this method to the case where information of the interaction between ligands and GPCRs is completely missing. We discuss the validity of the 2-way prediction method.

## 1. はじめに

Gタンパク質共役型受容体 (GPCR) は細胞膜を貫通するタンパク質であり、細胞外からリガンドとの相互作用によって、細胞外からの信号を細胞内へ伝える役割を持っている。この外部の信号を内部に伝える役割は、嗅覚、視覚、味覚など感覚器でも重要な役割を果たしている。その重要性から創薬研究の中心になることが多く、近年、GPCR を計算機で同定する研究<sup>1)</sup>や、GPCR に相互作用するリガンドをコンピュータ上で予測する研究が盛んに行われている<sup>2)3)</sup>。しかし、数種類しか立体構造が分かっているものがないことや、相互作用するリガンドがほとんど分かっていない GPCR が多数存在するなど、予測の役に立つと思われる実験データが非常に少ない。そのため、GPCR に相互作用するリガンドを予測することは困難である。

実際にサポータベクターマシンを用いて、GPCR に相互作用するリガンドを予測する研究が行われている<sup>3)</sup>。リガンドと GPCR の特徴を組合せ、リガンド情報のない GPCR と相互作用するリガンドを予測しているが、その予測は難しく、予測精度はよい場合で 78% であった。

本研究では、相互作用する GPCR の情報がないリガンドの結合予測、リガンド情報のない GPCR に対しての予測、さらに、リガンドと GPCR 両者についての情報が全て欠けている状態での予測方法である 2-way prediction 法を提案する。図 1、図 2、図 3 にそれぞれの場合の例を示す。

本稿の構成は以下のようになっている。第 2 節ではデータセットについて述べる。第 3 節では提案する予測方法と予測に使う特徴を示す。第 4 節では実験方法、実験結果、考察について述べる。第 5 節で結論をまとめる。

## 2. データセット

データセットには、既存研究との比較を容易にするため、Jacob らと同じデータセットを用いた<sup>3)</sup>。Jacob らは GPCR とリガンドのデータベース GLIDA<sup>4)</sup> から、創薬ターゲット

†1 早稲田大学

Waseda University

†2 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology (AIST)

†3 お茶の水女子大学

Ochanomizu University

	G1	G2	G3	G4	G5	G6	G7	...	Gn
L1			○			○			
L2	○				○		○		
L3	○	○		○					○
L4			○		○				
L5	○	○			○	○			
L6									
L7			○						
L8			○						○
L9					○	○			
...									
Lm				○	○				

図 1 相互作用するGPCR が分かっていないリガンド結合予測の例。○は正の相互作用を表す。○記号のない部分は相互作用未知である。網掛け部分であるリガンド L6 の行は、相互作用既知の GPCR が存在しない。

Fig. 1 An example of predicting a ligand with unknown GPCRs. The ○ and blank represent positive and unknown interactions, respectively. The information in shaded area (the L6 row) is absent.

になり得る分子量 450 未満程度のリガンドを選び、選ばれたリガンドのいづれかと相互作用する GPCR のみを抽出した。表 1 に利用したデータの概要を示す。

データベース GLIDA には、相互作用する GPCR とリガンドの組合せは登録されているが、相互作用しない組合せは登録されていない。例として図 3 のように○のみとなっている。ここで、○は正の反応を表しており、記号の無い部分は相互作用未知であることを表している。相互作用しない負のデータについても、Jacob らと同様に以下の方法で選択する。ある GPCR に対して、リガンドの特徴を用いて相互作用するリガンドを予測する実験をする際は GPCR には次のように選択する。あるリガンドがある GPCR と相互作用するが、他の GPCR とは相互作用するという情報が GLIDA に含まれていない場合、そのリガンドはその GPCR とは相互作用しないと仮定する。ただし、この手順で正例と同数だけ負例を選択する。負例を選択した結果は例として図 4 のようになる。× は負の反応を表す。また、あるリガンドに対して、GPCR の特徴を用いて相互作用する GPCR を予測する実験をする際には、相互作用しない GPCR の負例を正例と同数だけ同様に選択する。

	G1	G2	G3	G4	G5	G6	G7	...	Gn
L1			○						
L2	○						○		
L3	○	○		○					○
L4			○		○				
L5	○	○			○	○			
L6									
L7			○						
L8			○						○
L9					○	○			
...									
Lm				○	○				

図 2 相互作用するリガンドが分かっていない GPCR 結合予測の例。網掛け部分である GPCR G5 の列は、相互作用既知のリガンドが存在しない。

Fig. 2 An example of predicting a GPCR with unknown ligands. The information in shaded area (the G5 column) is absent.

### 3. 手 法

今回、我々は以下の 3 つの方法によって GPCR とリガンドの相互作用の予測を行った。

- (1) 各 GPCR について、相互作用既知のリガンドの特徴を用いて、未知のリガンドを予測する方法。
- (2) 各リガンドについて、相互作用既知の GPCR の特徴を用いて、未知の GPCR を予測する方法。
- (3) 相互作用する GPCR、リガンド、の情報が全くない場合に、方法 (1) と (2) を組み合わせてリガンドと GPCR の結合を予測する 2-way prediction 法。

この節では、方法 (1) で利用するリガンドの特徴の取り方、方法 (2) で利用する GPCR

表 1 使用データの概要

Table 1 Summary of the data set.

リガンド数	GPCR 数	相互作用の組合せ数	GPCR あたりの相互作用の数
2445	80	4051	50.6

の特徴の取り方, 方法 (3) の 2-way prediction 法について述べる.

### 3.1 リガンドの特徴

化学構造式で表現されるリガンド物質を, 辺と頂点の集合で表現されるグラフで表す. 頂点にはそれぞれの原子, 辺には結合の多重度に応じた重みを対応づける. グラフで表現された全てのリガンドから頻出部分グラフを抽出する<sup>9)</sup>. 頻出部分グラフとは, 任意の閾値を  $s$  としたとき, ある部分グラフが  $s$  個以上のリガンドに含まれるグラフのことをいう. この  $s$  をサポート値と呼ぶことにする. 得られた頻出部分グラフの構造をリガンドが持つならば 1, 持たないならば 0, という 1 と 0 のみの値を持つ特徴ベクトルを予測に用いる.

### 3.2 GPCR の特徴

まず立体構造が分かっている GPCR4 種類 ( $\beta_1$  アドレナリン受容体<sup>6)</sup>,  $\beta_2$  アドレナリン受容体, <sup>7)8)</sup> アデノシン受容体<sup>9)</sup>, ロドプシン<sup>10)</sup>) について, リガンド結合部位を決定する. リガンド結合部位は, リガンドに最も近接するアミノ酸残基を選び, それらのアミノ酸をリガンド結合部位とした. 具体的には  $\beta_1$  アドレナリン受容体,  $\beta_2$  アドレナリン受容体, ロドプシンについてはリガンドから 4.5Å 以内にあるヘリックス中のアミノ酸を, アデノシン受容体については 5.0Å 以内にあるヘリックス中のアミノ酸を選んだ. 選んだアミノ酸を表 2 に示す.

立体構造既知の各 GPCR と予測に利用する GPCR のアミノ酸配列をアライメントし, 立体構造既知 GPCR のリガンド結合部位に対応する部分のアミノ酸のみを抽出し, 特徴とする.

各リガンド結合部位を 21 次元のベクトルで表現する. ここで 21 次元としたのは, アライメントの際に例外的にギャップを含む可能性があるため, 20 種類のアミノ酸に加えてギャップという情報の特徴を取り入れるためである. そして対応するアミノ酸の部分のみ 1, 他は 0 といった特徴ベクトルを生成する.

### 3.3 2-way prediction 法

2-way prediction 法では, ある GPCR とあるリガンドについて, その GPCR に相互作用するリガンドが全て未知であり, そのリガンドに反応する GPCR も全て未知である, という GPCR とリガンドの組合せが相互作用するかどうかを予測する. ここで, 相互作用するリガンドが全て未知の GPCR を orphan GPCR と呼び, 相互作用する GPCR が全て未知のリガンドを orphan リガンドと呼ぶことにする. 図 4 にこの例を示す. 網掛け部分は GPCR (G5) について全てのリガンドが未知の場合と, リガンド (L6) について全て GPCR が未知の場合を表している. G5 が orphan GPCR であり, L6 が orphan リガンド

である. この網掛け部分が交差する部分が 2-way prediction 法で予測する対象である. この交差する部分を double orphan と呼ぶことにする.

2-way prediction 法には, 各 GPCR について orphan リガンドが相互作用するかどうかを予測した後, 予測結果と GPCR の特徴から orphan GPCR と orphan リガンドの相互作用を予測する方法と, 各リガンドについて orphan GPCR が相互作用するかどうかを予測した後, 予測結果とリガンドの特徴から orphan GPCR と orphan リガンドの相互作用を予測する方法がある. 前者を 2-way prediction 法 (A) とし, 後者を 2-way prediction 法 (B) とする.

2-way prediction 法 (A) では, まず相互作用するリガンドを  $k$  個以上含む GPCR それぞれについて, リガンドの特徴から, orphan リガンドがそれぞれの GPCR に相互作用するかどうかを予測する. その結果, orphan リガンドに相互作用する複数の GPCR の情報が得られる. 得られた GPCR の特徴を用いて最終的に予測したい double orphan の相互作用を予測する. このように 2 段階の予測を行う. 例えば図 4 の場合には, G1 の GPCR について, L2, L3, L5, L7, L8 などの情報を用いて L6 と G1 が相互作用するかどうかを予測する. この予測を G2, G3 など他の GPCR についても行う. その結果図 5 のように空白であった L6 の行が  $\circ$  もしくは  $\times$  で埋められる. 次に L6 の行について, 予測された G1, G2, G3 などの GPCR の特徴を用いて G5 の部分を予測する.

逆に 2-way prediction 法 (B) では, 先に相互作用する GPCR を  $k$  個以上含むリガンドそれぞれについて, GPCR の特徴から orphan GPCR がそれぞれのリガンドに相互作用するかどうかを予測する. その結果を用いて, リガンドの特徴から最終的に予測したい double orphan の相互作用を予測する. 図 4 の場合には, 先に G5 の列の L6 以外の空白を予測し, 予測された L6 以外の L1, L2, L3 などの情報から L6 を予測する方法である.

## 4. 実験

### 4.1 実験方法

各 GPCR について, リガンドの特徴を用いて各 GPCR に相互作用する未知リガンドの予測をする方法 (1), 各リガンドについて, GPCR の特徴を用いて各リガンドに相互作用する未知 GPCR の予測をする方法 (2), 2-way prediction 法による未知 GPCR と未知リガンドの相互作用の予測をする方法 (3), の 3 通りの実験を行った.

- 方法 (1)

既知の正例と同数の負例を前節のようにランダムに生成する. データが少ないとうまく