Differential gene expression profiling in blood from patients with digestive system cancers

メタデータ	言語: eng 出版者: 公開日: 2017-10-03 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	http://hdl.handle.net/2297/25265

*Manuscript
Click here to view linked References

Differential Gene Expression Profiling in Blood from Patients with Digestive System Cancers

Running title: Gene expression profiling in patients with digestive system cancers

Masao Honda[1) 2)], Yoshio Sakai[1)], Taro Yamashita[1)], Tatsuya Yamashita[1)], Akito Sakai[1)], Eishiro Mizukoshi[1)], Yasunari Nakamoto[1)], Isamu Tatsumi[1)], Yoshitaka Miyazaki[1)], Hiroshi Tanno[1)], Shuichi Kaneko[1)] and Hokuriku Liver Study Group*

1) Department of Gastroenterology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan

2) Department of Advanced Medical Technology, Kanazawa University Graduate School of Health Medicine, Kanazawa, Japan

*Participating investigators are listed in Appendix A.

All the authors declare that no potential competing interests exist.

Abbreviations: AUC, Area under the curve; BMI, body mass index; CA 19-9, carbohydrate antigen (CA) 19-9; CEA, carcinoembryonic antigen; HCC, hepatocellular carcinoma; HSC, hematopoietic stem cell; IFN, interferon; NPV, negative predictive value; PBMC, peripheral blood mononuclear cell; PPV, positive predictive value; ROC, receiver operating characteristic; SVM, support vector machine

Corresponding author:

Shuchi kaneko, M.D., Ph.D.

Department of Gastroenterology, Graduate School of Medicine, Kanazawa University,

Takara-Machi 13-1, Kanazawa 920-8641, Japan

Tel: +81-76-265-2235; Fax: +81-76-234-4250

E-mail: skaneko@m-kanazawa.jp

Abstract

To develop a non-invasive and sensitive diagnostic test for cancer using peripheral blood, we evaluated gene expression profiling of blood obtained from patients with cancer of the digestive system and normal subjects. The expression profiles of blood-derived total RNA obtained from 39 cancer patients (11 colon cancer, 14 gastric cancer, and 14 pancreatic cancer) was clearly different from those obtained from 15 normal subjects. By comparing the gene expression profiles of cancer patients and normal subjects, 25 cancer-differentiating genes ($p < 5.0 \times 10^{-6}$ and fold differences >3) were identified and an “expression index” deduced from the expression values of these genes differentiated the validation cohort (11 colon cancer, 8 gastric cancer, 18 pancreatic cancer, and 15 normal subjects) into cancer patients and normal subjects with 100% (37/37) and 87% (13/15) accuracy, respectively. Although, the expression profiles were not clearly different between the cancer patients, some characteristic genes were identified according to the stage and species of the cancer. Interestingly, many immune-related genes such as antigen presenting, cell cycle accelerating, and apoptosis- and stress-inducing genes were up-regulated in cancer patients, reflecting the active turnover of immune regulatory cells in cancer patients. These results showed the potential relevance of peripheral blood gene expression profiling for the development of new diagnostic examination tools for cancer patients.

Key words: gene expression, digestive system cancer, peripheral blood, diagnosis, pathway

Introduction

Cancer of the digestive system is one of the most common forms of cancer worldwide. The early detection of cancer enables the administration of therapy and the subsequent prolongation of overall survival; however, the detection of early-stage cancer is difficult, and patients with general symptoms are likely to have advanced-stage cancer. Particularly, in pancreatic cancer [1, 2], early diagnosis is extremely difficult despite the development of modern imaging technology such as ultrasonography or computed tomography. Even though the recent development of chemotherapy combined with molecular target drugs has improved the survival rate of patients with advanced cancer, the therapeutic benefit of this treatment is limited [1].

Peripheral blood in patients includes a variety of immune regulatory cells such as leukocytes and lymphocytes that are essential players in the host immune defense system. These cells respond to various abnormal conditions such as viral infection, metabolic disease, and cancer [3-12]. We previously reported that the expression profiles of peripheral blood mononuclear cells (PBMCs) from patients with hepatocellular carcinoma (HCC) differed significantly from those of patients without HCC ($P < 0.0005$) [8]. The results also suggest that the gene expression profile of blood may be useful as a clinical surrogate biomarker for HCC assessment.

In this study, we extended our previous findings to the diagnosis of cancer of the digestive system, including gastric cancer, colorectal cancer, and pancreatic cancer. We identified clear differences in the gene expression profiles of cancer patients and normal subjects, suggesting the potential diagnostic relevance of gene expression signatures from blood samples for cancer of the digestive system.

Material and Methods

Patients and blood samples

We enrolled 76 patients with cancer of the digestive system, including 22 patients

with colon cancer, 22 patients with gastric cancer, and 32 patients with pancreatic cancer at the Graduate School of Medicine, Kanazawa University Hospital and its related hospitals, Japan from 2008–2009 (Table 1). Blood samples were obtained from patients following their diagnosis with cancer of the digestive system. The age- and sex-matched control samples were obtained from 30 healthy volunteers who received health screening examinations (Table 1). Informed consent was obtained from all patients, and ethics approval for this study was obtained from the Ethics Committee for Human Genome/Gene Analysis Research at Kanazawa University Graduate School of Medical Science. The cancer patients and normal subjects were randomly divided into the training (n = 54) and validation (n = 52) cohorts according to their entry number. There were no significant differences in age, sex, body mass index (BMI), and habits between the cancer patients and normal subjects (Table 1).

RNA extraction from blood

Blood samples collected in PAXgene Blood RNA tubes (BD, NJ, USA) were incubated and stored according to the manufacturer's instructions. Total RNA was isolated after thawing the samples at room temperature using the PAXgene Blood RNA System kit (Qiagen, CA, USA) following the manufacturer's instructions. The quality of purified RNA was analyzed using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). RNA concentration was determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, DE, USA).

Microarray and data analysis

Cy-3-labeled cRNA was synthesized from 300 ng of total RNA using the Quick Amp Labeling kit, One-color (Agilent Technologies, CA, USA) and purified using an RNeasy column (Qiagen). After checking the quality of the RNA using an Agilent 2100 Bioanalyzer, the RNA was hybridized to 4 × 44K Whole Human Genome Microarray (Agilent Technologies, CA,

USA). The microarray slide was incubated in a hybridization oven at 65°C for 17 h, washed, and then scanned using a DNA Microarray Scanner, Model G2505B (Agilent Technologies, CA, USA). All procedures from the labeling to the scanning were performed according to the manufacturer's instructions (Agilent Technologies, CA, USA). The scanned data of each slide were extracted using Feature Extraction software (Agilent Technologies).

Gene expression analysis was carried out using GeneSpring GX software (Agilent Technologies). Each measurement was divided by the 75th percentile of all measurements in that sample at per chip normalization. Hierarchical clustering was generated using the Pearson correlation similarity metric and the average or complete linkage clustering algorithm. Welch's t-test with Benjamini and Hochberg's false discovery rate were used to identify the genes that were differentially expressed in the patients of each category.

Class prediction analysis and calculation of the expression index

Building and running prediction models were performed using GeneSpring GX software (Agilent Technologies). Models were generated for the statistically extracted genes from the training cohort using a support vector machine (SVM) algorithm.

In addition to the supervised learning methods, we calculated an "expression index" that was used for class prediction analysis. Logistic regression analysis to predict cancer patients and normal subjects was performed using the individual gene expression values. The gene expression cut-off values were determined using a receiver operating characteristic (ROC) curve. If the expression value of a gene exceeded the cut-off value, the index was scored as "1," and if the expression value of a gene was not beyond the cut-off value, then the index was scored as "0." The total index was calculated and designated as the "expression index." The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the expression index for the prediction of cancer patients and normal subjects were further evaluated using the training and variation cohorts.

Pathway analysis of the expression data

The pathway analysis of the differentially expressed genes was performed using the MetaCore software suite (GeneGo, MI, USA), a unique, curated database of human protein-protein and protein-DNA interactions, transcription factors, and signaling, metabolic, and bioactive molecules. Differentially expressed genes were analyzed by GeneGo annotation, and categories of differentially expressed genes are shown by their frequency; moreover, possible networks of differentially expressed genes were created according to the direct interaction relation program of MetaCore.

Statistical analysis

The Mann-Whitney U-test was used to analyze continuous variables. Chi-squared and Fisher's exact tests were used to analyze categorical data. Multivariate logistic analysis was performed using a stepwise logistic regression model. A p-value of less than 0.05 was considered significant. Statistical analyses were performed using JMP8 for Windows (SAS Institute, NC, USA).

Results

Clinical characteristic of patients enrolled in this study

The clinical characteristics of the patients enrolled in this study are shown in Table 1. The training cohort included 39 patients with cancer of the digestive system (11 colon cancer, 14 gastric cancer, and 14 pancreatic cancer) and 15 normal subjects. The validation cohort included 37 patients with cancer of the digestive system (11 colon cancer, 8 gastric cancer, and 18 pancreatic cancer) and 15 normal subjects. There were no statistical differences in age, gender, habits, BMI (kg/m^2), and blood cell count between the cancer patients and the normal subjects in the training and validation cohorts. The majority of the advanced stage cancer was observed in the pancreatic cancer patients in both cohorts. The serum levels of CA 19-9 were significantly higher in patients with pancreatic cancer than in those with gastric or colon cancers in the validation cohort (Table 2).

Hierarchical clustering analysis

The results from the unsupervised hierarchical cluster analysis of the training cohort using an average linkage clustering algorithm based on the 23278 expressed genes are shown in Figure 1A. Interestingly, the expression profiles in the blood obtained from cancer patients and normal subjects were clearly different, except in one normal subject. There was no clear clustering within the cancer patients; however, patients with pancreatic cancer or advanced-stage cancer associated with distant metastasis or vascular invasion were likely to be clustered together (Figure 1A). We performed class prediction analysis using a supervised learning method based on the SVM algorithm to confirm these findings. Using the statistical values ($p < 0.05$) and fold differences (>2) as filtering criteria, 1348 genes were identified that differentiated cancer patients from normal subjects (cancer-differentiating genes) (Table 2). Similarly, 45 genes were identified that differentiated patients with advanced-stage cancer (stages III–IV) from early-stage cancer (stages 0–II) (stage-differentiating genes)

(Supplemental Table 2), and 44 genes were identified that differentiated patients with gastric or colon cancers from those with pancreatic cancer (GI tract/pancreas-differentiating genes) (Table 2) (Supplemental Table 3). No significant differences were identified in gene expression between the patients with different ages (≥ 65 yr. and < 65 yr), and between patients with gastric or colon cancers (Table 2). We observed a high prediction capacity for the cancer-differentiating genes (87–97% accuracy), while the predictive value of stage- and GI tract/pancreas-differentiating genes was not sufficient (71–89% accuracy) (Table 2). Hierarchical clustering using more strict selection criteria ($p < 5.0 \times 10^{-6}$ and fold differences > 3) identified 25 cancer-differentiating genes (Figure 1B), confirming the clear differentiation of cancer patients and normal subjects. Hierarchical clustering using 45 stage- and 44 GI tract/pancreas-differentiating genes is shown in Figure 1C. Within the cancer patients, gastric or colon cancer was differentiated from pancreatic cancer, and advanced-stage cancer associated with metastasis or vascular invasion was roughly differentiated from early-stage cancer (Figure 1C).

Calculation of the expression index

To apply these findings to clinical and practical settings, we calculated the expression index in individual cases. Logistic regression analysis of cancer patients and normal subjects was performed using the individual expression values of the 25 cancer-differentiating genes. The cut-off value of gene expression was determined from the ROC curve. The individual distribution of the expression values of the 25 genes in the training cohort patients is shown in Supplemental Figure 1. Eleven genes were up-regulated in cancer patients, while 14 genes were down-regulated. We standardized each expression value using the following approach: if the expression value exceeded the cut-off value, the expression value was counted “1,” and if the expression value was less than the cut-off value, the expression value was counted as

"0." The hierarchical clustering of the training cohort patients using the standardized expression values is shown in Figure 2; we observed clearer clustering of the cancer patients and normal subjects for these values. For statistical evaluation, a total expression score was calculated and designated as the "expression index," where high expression index values could indicate patients with cancer. The cut-off value of the expression index was determined by an ROC curve, and the sensitivity, specificity, PPV, and NPV of the expression index are shown in Table 3. The distribution of the expression index in patients is shown in Figure 3A. The results demonstrated the high sensitivity, specificity, PPV, and NPV of the expression index for predicting cancer patients and normal subjects in the training and validation cohorts. The predictive values of the 44 stage- and 45 GI tract/pancreas-differentiating genes in the training cohort were fair (70-100%); however, they were not sufficient in the validation cohort (59-84%) (Table 3, Figure 3B and 3C).

Pathway analysis

To examine which signaling pathways were differentially expressed in blood from cancer patients, we performed pathway analysis of the 841 differentially expressed genes ($p < 5.0 \times 10^{-5}$ and fold differences > 1.7) using MetaCore software (GeneGo). Interestingly, many of the immune-related genes, such as antigen presenting, cell cycle accelerating, and apoptosis- and stress-inducing genes, were up-regulated in cancer patients, while development-related genes, such as tissue remodeling and hedgehog signaling, were down-regulated (Supplemental Figure 2). We generated the possible network processes of the differentially expressed genes according to the direct interaction algorithm (Supplemental Figure 3). Interestingly, many p53 target genes were up-regulated in association with the induction of caspase-3, suggesting the presence of cell cycle regulation and the induction of apoptosis. Interestingly, stem cell-related and differentiation genes such as Oct-3/4 and Oct-1 were down-regulated, suggesting the impaired differentiation of immune regulatory cells.

Therefore, the expression profile may reflect the active immune reaction and the decreased pluripotency or repertoire of immune regulatory cells in cancer patients.

With regard to the stage-differentiating genes, it is interesting to note that a larger number of interferon-stimulated genes were up-regulated in advanced-stage cancer than in early-stage cancer (Supplemental Table 2). With regard to the GI tract/pancreas-differentiating genes, a larger number of G-protein-related genes were up-regulated in pancreatic cancer patients (Supplemental Table 3). These differences may reflect the possible interaction between tumor cells and tumor-infiltrating lymphocytes.

Discussion

Detection of cancer of the digestive system using peripheral blood is an attractive diagnostic method because of its simplicity and non-invasive nature. For the detection of early-stage cancer of the digestive system, endoscopic examinations of the stomach and colon or imaging studies, such as abdominal ultrasonography or computed tomography, should be performed periodically; however, these examinations are expensive and the patients suffer from high levels of stress during these examination. Although serological tumor markers such as CEA and CA 19-9 have been utilized for the diagnosis of cancer of the digestive system, these tumor markers have a low sensitivity and specificity [13, 14].

Peripheral blood in patients includes a variety of immune regulatory cells that respond to various abnormal conditions such as viral infection, metabolic disease, and cancer. Recent emerging reports including ours [5, 6, 8] support the possibility that the gene expression profiling of peripheral blood could be a useful surrogate biomarker [3, 4, 7, 9-12].

In this study, we evaluated gene expression profiling of blood obtained from patients with cancer of various digestive system including gastric cancer, colon cancer and pancreatic cancer that have not been characterized systematically. To our knowledge, this is the first report to find a common gene set for the diagnosis of cancer with the digestive system. The identified gene set could be useful for the screening of patients with cancer of the digestive system. The gene expression profiles of peripheral blood from cancer patients were clearly different from those in normal subjects (Figure 1A). We identified 1348 cancer-differentiating genes using the filtering criteria of $p < 0.05$ and fold differences > 2, and 25 more strictly selected genes using the filtering criteria of $p < 5.0 \times 10^{-6}$ and fold differences > 3 in the training cohort. Hierarchical clustering using the unsupervised learning method clearly differentiated cancer patients and normal subjects in the validation cohort using the 25 selected genes (data not shown). The supervised learning method based on the SVM using the 25 cancer-differentiating genes predicted the cancer patients in the training cohort with an

accuracy of 93–100% (Table 2) and 77–100% in the validation cohort (data not shown). Thus, unsupervised and supervised learning methods successfully identified the cancer patients in the training and validation cohorts with a high accuracy. Importantly, there were no obvious differences in the clinical backgrounds of the cancer patients and normal subjects in the training and validation cohorts, except for the serum CA 19-9 levels in the pancreatic cancer patients of the validation cohort (Table 1).

For more practical and clinical usage, we calculated the conventional “expression index” and utilized it for the prediction of cancer patients. The expression index was based on the individual expression values (see Material and Methods, and Results) and the cut-off value was determined by the ROC curve generated from the logistic regression analysis. The sensitivity, specificity, PPV (%), and NPV (%) of the expression index for the 25 cancer-differentiating genes were well tolerated for the prediction of cancer patients and normal subjects in the training and validation cohorts (Table 3, Figure 3). Multivariate analysis using the expression index, CA-19-9, CEA, age, and sex in the validation cohort indicated that the expression index was the only independent variable associated with cancer patients ($p < 0.001$, Odds = 3.0×10^5/score). Thus, the expression index is practically useful for the identification of cancer patients with digestive system.

By using the same strategy, we identified 45 stage-differentiating genes (Supplemental Table 2) and 44 GI tract/pancreas-differentiating genes (Supplemental Table 3). Although the predictive performance of these genes was less efficient, the results suggest that the expression profiles may be different according to the stage and species of the cancer.

What causes these differences in the expression profiles of blood from cancer patients? Previously, we examined the gene expression profiles of PBMCs obtained from patients with or without HCC and showed that the expression profiles of PBMC from patients with HCC differed significantly from those of patients without HCC [8]. Interestingly, the gene expression profiles of the redox status, cell cycle, and proteasome system, along with

immunologic genes were up-regulated in PBMCs from patients with HCC, suggesting the regulation of anticancer immunity. Importantly, these genes were also up-regulated in HCC-infiltrating mononuclear inflammatory cells, implying that local anticancer immunity may be reflected in the peripheral gene expression signature. In this study, it was also found that many immune-related genes, such as antigen presenting, cell cycle accelerating, and apoptosis- and stress-inducing genes, were up-regulated in cancer patients, reflecting the presence of an active immune reaction in cancer patients. Interestingly, the expression of many differentiation-related genes such as Oct-3/4 and Oct-1 was down-regulated, suggesting that the differentiation of immune cells was impaired. These may represent a characteristic immune feature of cancer and reflect the impaired immune system of cancer patients. Although we did not analyze regional tumor-infiltrating mononuclear inflammatory cells in this study, a similar reaction may occur in the local tumor lesion.

In addition to the cancer-differentiating genes, there could be characteristic genes that reflect the stage and species of the cancer. It is interesting to note that more interferon-stimulated genes were up-regulated in advanced-stage cancers. A recent study reported that interferon (INF)-α activated dormant hematopoietic stem cells (HSCs) and sensitize these cells to 5-fluoro-uracil exposure. In contrast, HSCs chronically activated by INF-α are functionally compromised. Therefore, the up-regulation of IFN signaling in advanced-stage cancer reflects the refractory state of the differentiation of immune regulatory cells. Although, the specificity of these genes was not sufficient, the detailed diagnosis of cancer of the digestive system may be possible by generating a decision tree (Supplemental Figure 4).

In conclusion, we demonstrated a distinct gene expression profile of blood from cancer patients of the digestive system compared to healthy individuals, and showed the potential diagnostic values of these differences for clinical usage. Further studies should be performed to validate these findings in detail and identify the fundamental mechanisms

underlying this phenomenon.

Appendix A

The Hokuriku Liver Study Group (HLSG) is composed of the following members:
Drs. Takashi Kagaya, Kuniaki Arai, Kaheita Kakinoki, Kazunori Kawaguchi, Kazuya Kitamura, Hajime Takatori, Hajime Sunakosaka (Department of Gastroenterology, Kanazawa University Graduate School of Medicine, Kanazawa); Drs. Touru Nakahama, Shinji Kamiyamamoto, (Kurobe City Hospital, Kurobe, Toyama); Dr. Yasuhiro Takemori (Toyama Rosai Hospital, Uozu, Toyama); Dr. Hikaru Oguri (Koseiren Namerikawa Hospital, Namerikawa, Toyama); Drs. Yatsugi Noda, Hidero Ogino (Toyama Prefectural Central Hospital, Toyama, Toyama); Drs. Yoshinobu Hinoue, Keiji Minouchi (Toyama City Hospital, Toyama, Toyama); Dr. Nobuyuki Hirai (Koseiren Takaoka Hospital, Takaoka, Toyama); Drs. Tatsuho Sugimoto, Koji Adachi (Tonami General Hospital, Tonami, Toyama); Dr. Yuichi Nakamura (Noto General Hospital, Nanao, Ishikawa); Drs. Masashi Unoura, Ryuhei Nishino (Public Hakui Hospital, Hakui, Ishikawa); Drs. Hideo Morimoto, Hajime Ohta (National Hospital Organization Kanazawa Medical Center, Kanazawa, Ishikawa); Dr. Hirokazu Tsuji (Kanazawa Municipal Hospital, Kanazawa, Ishikawa); Drs. Akira Iwata, Shuichi Terasaki (Kanazawa Red Cross Hospital, Kanazawa, Ishikawa); Drs. Tokio Wakabayashi, Yukihiro Shirota (Saiseikai Kanazawa Hospital, Kanazawa, Ishikawa); Drs. Takeshi Urabe, Hiroshi Kawai (Public Central Hospital of Matto Ishikawa, Hakusan, Ishikawa); Dr. Yasutsugu Mizuno (Nomi Municipal Hospital, Nomi, Ishikawa); Dr. Shoni Kameda (Komatsu Municipal Hospital, Komatsu, Kanazawa); Drs. Hirotoshi Miyamori, Uichiro Fuchizaki (Keiju Medical Center, Nanao, Ishikawa); Dr. Haruhiko Shyugo (Kanazawa Arimatsu Hospital, Kanazawa, Ishikawa); Dr. Hideki Osaka (Yawata Medical Center, Komatsu, Ishikawa); Dr. Eiki Matsushita (Kahoku Central Hospital, Tsubata, Ishikawa); Dr. Yasuhiro Katou (Katou Hospital, Komatsu, Ishikawa); Drs. Nobuyoshi Tanaka, Kazuo Notsumata (Fukuiken Saiseikai Hospital, Fukuil, Fukui); Dr. Mikio Kumagai (Kumagai Clinic, Tsuruga, Fukui); Dr. Manabu Yoneshima (Municipal Tsuruga Hospital, Tsuruga, Fukui)

References

[1] X. Liu, J. Li, Y. Tian, P. Xu, X. Chen, K. Xie, Z. Qiu, Y. Wang, D. Zhang, F. Wolf, C. Li, and Q. Huang, Enhanced pancreatic cancer gene therapy by combination of adenoviral vector expressing c-erb-B2 (Her-2/neu)-targeted immunotoxin with a replication-competent adenovirus or etoposide, Hum Gene Ther 21 (2004) 157-170.

[2] G. Schneider, J. T. Siveke, F. Eckel, and R. M. Schmid, Pancreatic cancer: basic and clinical aspects, Gastroenterology 128 (2005) 1606-1625.

[3] M. E. Burczynski, N. C. Twine, G. Dukart, B. Marshall, M. Hidalgo, W. M. Stadler, T. Logan, J. Dutcher, G. Hudes, W. L. Trepicchio, A. Strahs, F. Immermann, D. K. Slonim, and A. J. Dorner, Transcriptional profiles in peripheral blood mononuclear cells prognostic of clinical outcomes in patients with advanced renal cell carcinoma, Clin Cancer Res 11 (2005) 1181-1189.

[4] I. Osman, D. F. Bajorin, T. T. Sun, H. Zhong, D. Douglas, J. Scattergood, R. Zheng, M. Han, K. W. Marshall, and C. C. Liew, Novel blood biomarkers of human urinary bladder cancer, Clin Cancer Res 12 (2006) 3374-3380.

[5] M. Tateno, M. Honda, T. Kawamura, H. Honda, and S. Kaneko, Expression profiling of peripheral-blood mononuclear cells from patients with chronic hepatitis C undergoing interferon therapy, J Infect Dis 195 (2007) 255-267.

[6] T. Takamura, M. Honda, Y. Sakai, H. Ando, A. Shimizu, T. Ota, M. Sakurai, H. Misu, S. Kurita, N. Matsuzawa-Nagata, M. Uchikata, S. Nakamura, R. Matoba, M. Tanino, K. Matsubara, and S. Kaneko, Gene expression profiles in peripheral blood mononuclear cells reflect the pathophysiology of type 2 diabetes, Biochem Biophys Res Commun 361 (2007) 379-384.

[7] M. Han, C. T. Liew, H. W. Zhang, S. Chao, R. Zheng, K. T. Yip, Z. Y. Song, H. M. Li, X. P. Geng, L. X. Zhu, J. J. Lin, K. W. Marshall, and C. C. Liew, Novel blood-based, five-gene biomarker set for the detection of colorectal cancer, Clin Cancer Res 14 (2008) 455-460.

[8] Y. Sakai, M. Honda, H. Fujinaga, I. Tatsumi, E. Mizukoshi, Y. Nakamoto, and S. Kaneko, Common transcriptional signature of tumor-infiltrating mononuclear inflammatory cells and peripheral blood mononuclear cells in hepatocellular carcinoma patients, Cancer Res 68 (2008) 10267-10279.

[9] M. K. Showe, A. Vachani, A. V. Kossenkov, M. Yousef, C. Nichols, E. V. Nikonova, C. Chang, J. Kucharczuk, B. Tran, E. Wakeam, T. A. Yie, D. Speicher, W. N. Rom, S. Albelda, and L. C. Showe, Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease, Cancer Res 69 (2009) 9202-9210.

[10] H. Huang, X. Dong, M. X. Kang, B. Xu, Y. Chen, B. Zhang, J. Chen, Q. P. Xie, and Y. L. Wu, Novel blood biomarkers of pancreatic cancer-associated diabetes mellitus identified by peripheral blood-based gene expression profiles, Am J Gastroenterol 105 (2010) 1661-1669.

[11] B. Mesko, S. Poliska, A. Szegedi, Z. Szekanecz, K. Palatka, M. Papp, and L. Nagy, Peripheral blood gene expression patterns discriminate among chronic inflammatory diseases and healthy controls and identify novel targets, BMC Med Genomics 3 (2010) 15.

[12] M. X. Pham, J. J. Teuteberg, A. G. Kfoury, R. C. Starling, M. C. Deng, T. P. Cappola, A. Kao, A. S. Anderson, W. G. Cotts, G. A. Ewald, D. A. Baran, R. C. Bogaev, B. Elashoff, H. Baron, J. Yee, and H. A. Valantine, Gene-expression profiling for rejection surveillance after cardiac transplantation, N Engl J Med 362 (2010) 1890-1900.

[13] K. Satake, T. Takeuchi, T. Homma, and H. Ozaki, CA19-9 as a screening and diagnostic tool in symptomatic patients: the Japanese experience, Pancreas 9 (1994) 703-706.

[14] J. E. Kim, K. T. Lee, J. K. Lee, S. W. Paik, J. C. Rhee, and K. W. Choi, Clinical usefulness of carbohydrate antigen 19-9 as a screening test for pancreatic cancer in an asymptomatic population, J Gastroenterol Hepatol 19 (2004) 182-186.

Figure Legends

Figure 1

A: Hierarchical clustering analysis of 54 training cohort samples based on the expression levels of 23278 genes.

B: Hierarchical clustering analysis of 54 training cohort samples based on the expression levels of 25 cancer-differentiating genes ($p < 5.0 \times 10^{-6}$ and fold differences > 3).

C: Hierarchical clustering analysis of 542 training cohort samples based on the expression levels of 45 stage- and 44 GI tract/pancreas-differentiating genes ($p < 0.05$ and fold differences > 2).

Figure 2

Hierarchical clustering analysis of 54 training cohort samples based on the standardized expression level (0 or 1).

Figure 3

A: Calculated expression index of 25 cancer-differentiating genes in the training and validation cohorts.

B: Calculated expression index of 44 GI tract/pancreas-differentiating genes in the training and validation cohorts.

C: Calculated expression index of 45 stage-differentiating genes in the training and validation cohorts.

Supplemental Figure 1

Distribution of the individual expression values of 25 cancer-differentiating genes.

Supplemental Figure 2

Pathway analysis of differentially expressed genes in the blood from cancer patients and normal subjects using MetaCore. The frequently observed pathway processes are listed in their order of significance.

Supplemental Figure 3

Gene network of differentially expressed genes in the blood from cancer patients compared with normal subjects. A direct interaction algorithm was utilized to connect individual genes using MetaCore. A red mark implies up-regulation and a blue mark implies down-regulation in cancer patients compared with normal subjects.

Supplemental Figure 4

A putative decision tree for the diagnosis of cancer of the digestive system.

Table1

Table 1. Clinical characteristics of patients

Clinical category	Training cohort (n = 54)					Validation cohort (n = 52)				
	Patients with digestive cancer			Normal	p-value	Patients with digestive cancer			Normal	p-value
	Colon (n = 11)	Gastric (n = 14)	Pancreatic (n = 14)	(n = 15)		Colon (n = 11)	Gastric (n = 8)	Pancreatic (n = 18)	(n = 15)	
Age	68.8 ± 8.3	66.7 ± 12.7	68.2 ± 8.2	62.4 ± 4.8	N.S.	70.1 ± 9.3	68.9 ± 7.3	66.7 ± 13.8	62.2 ± 5.9	N.S.
Gender										
Male	10	7	9	7	N.S.	9	4	10	6	N.S.
Female	1	6	5	8		2	4	8	9	
BMI (> 25m^2/Kg)	19.9 ± 3.2	22.2 ± 3.3	19.5 ± 3.1	22.6 ± 2.2	N.S.	22.5 ± 5.0	24.0 ± 2.4	22.0 ± 4.1	22.5 ± 2.4	N.S.
Clinical Stage										
0 or I	3	6	0	-	C vs. P: 0.002 G vs. P: 0.002 (0–II vs. III–IV)	4	7	0	-	C vs. G: 0.009 C vs. P: 0.03 G vs. P: 0.001 (0–II vs III–IV)
II	2	0	0	-		2	1	3	-	
III	3	2	1	-		4	0	3	-	
IV	3	6	13	-		1	0	12	-	
Laboratory data										
WBC (x10^3)	6.62 ± 2.2	6.72 ± 2.6	6.77 ± 2.5	5.95 ± 1.9	N.S.	6.05 ± 1.7	6.60 ± 1.3	5.64 ± 1.9	5.85 ± 3.0	N.S.
RBC (x10^6)	393 ± 54	414 ± 50	417 ± 70	441 ± 37	N.S.	415 ± 76	411 ± 65	417 ± 69	451 ± 120	N.S.
Hb (g/dL)	11.1 ± 2.8	12.5 ± 2.8	12.9 ± 2.0	13.5 ± 1.4	N.S.	12.3 ± 3.2	12.1 ± 3.5	12.6 ± 2.3	13.1 ± 0.7	N.S.
Tumor marker										
CEA (> 5 ng/mL))										
mean ± SD	442 ± 1433	120 ± 450	98 ± 273	2 ± 0.8	N.S.	47 ± 124	10 ± 23	9 ± 15	2 ± 0.8	N.S.
CA19-9 (> 37 U/mL)										
mean ± SD	6011 ± 1988	1169 ± 4263	86867 ± 257340	2 ± 1.6	N.S.	47 ± 96	21 ± 30	1714 ± 2473	2.2 ± 1.6	P vs. N: 0.009 P vs. C: 0.02 P vs.
Habits										
Alcohol	0	1	1	0	N.S.	1	0	0	0	N.S.
Smoking	0	0	0	0	N.S.	0	1	2	0	N.S.

Alcohol: History of alcohol intake more than 60 g/day
Smoking: History of smoking more than 400 Brinkman index
Data are expressed as mean ± SD
C: Colon cancer
G: Gastric cancer
P: Pancreatic cancer
N.S: Not significant

Table2

Table 2. Class prediction analysis by supervised learning method based on the support vector machine (SVM)

Clinical category	Subgroup	Total no. of classes	No. of cases misclassified	Mean percent of correct classification	No. of differentially expressed genes	
					$p < 0.05$, Fold > 2	$p < 5.0 \times 10^{-6}$, Fold > 3
Normal vs. cancer	Normal	15	2 (1)*	87 (93)*	1348	23
	Cancer	39	1 (0)*	97 (100)*		
Age	≧65	30	-	-	0	0
	<65	24	-	-		
Stage	0–II	11	2	82	45	0
	III–IV	28	3	89		
Colon+Gastric vs. Pancreatic	Gastric+Colon	25	2	92	44	0
	Pancreatic	14	4	71		
Colon vs. Gastric	Gastric	14	-	-	0	0
	Colon	11	-	-		

( )*: No. of cases misclassified and mean percent of correct classification using 23 genes ($p < 5.0 \times 10^{-6}$, Fold > 3)

Table3

Table 3. Sensitivity, specificity, PPV, and NPV of the expression index

Prediction category	No.of genes	Expression index cut-off	Training/ Validation	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC
Cancer (vs. normal)	23	14	Training	100 (39/39)	100 (15/15)	100 (39/39)	100 (15/15)	1
			Validation	100 (37/37)	87 (13/15)	95 (37/39)	100 (13/13)	0.99
Stage III–IV (vs. 0–II)	45	18	Training	96 (27/28)	82 (9/11)	93 (27/29)	90 (9/10)	0.94
			Validation	80 (16/20)	59 (10/17)	70 (16/23)	71 (10/14)	0.69
Colon+Gastric (vs. Pancreatic)	44	28	Training	76 (19/25)	100 (14/14)	100 (19/19)	70 (14/20)	0.95
			Validation	84 (16/19)	65 (11/17)	73 (16/22)	79 (11/14)	0.78

PPV: Positive predictive value
NPV: Negative predictive value
AUC: Area under the curve

Figure1

Fig 1

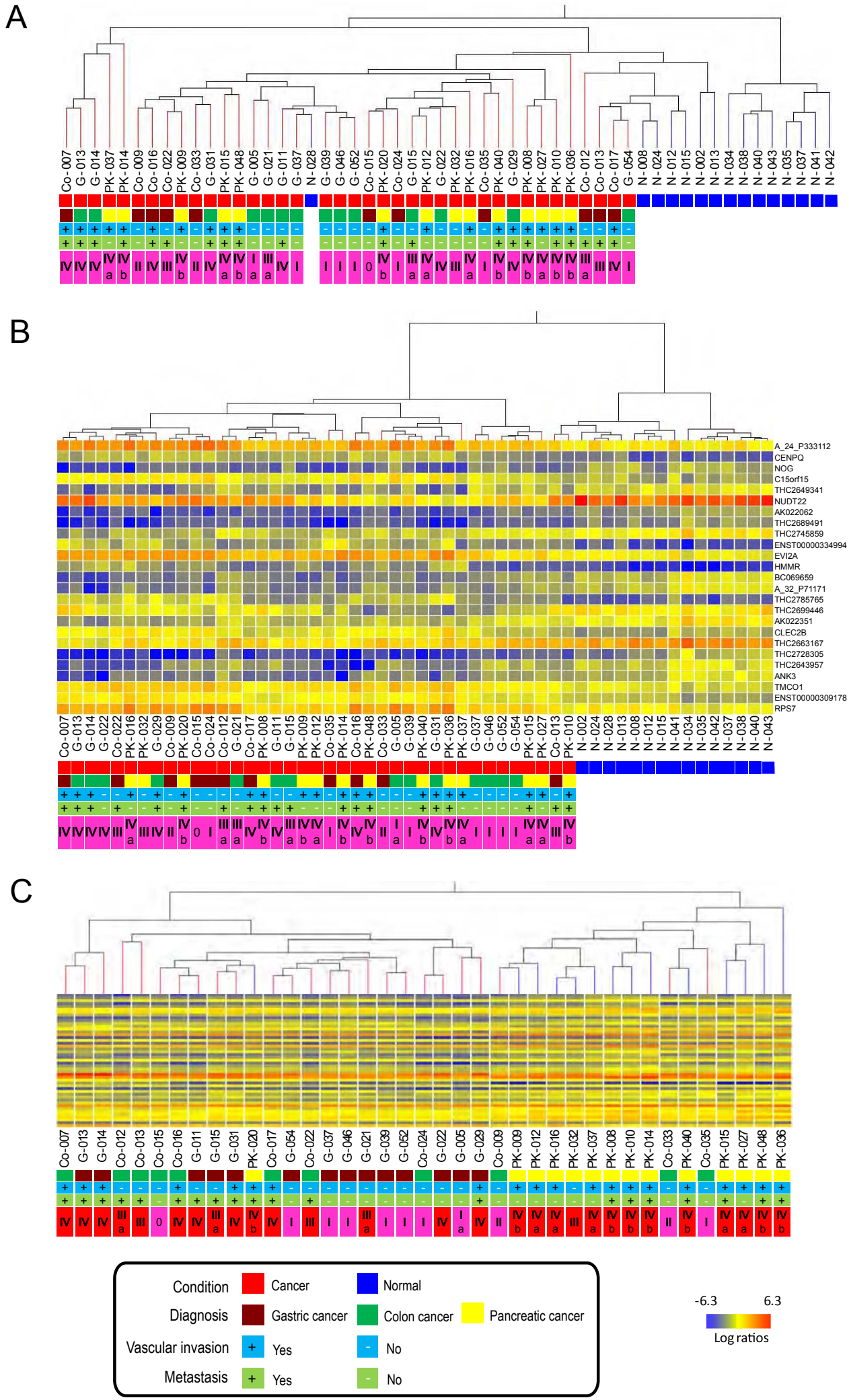


Fig 2

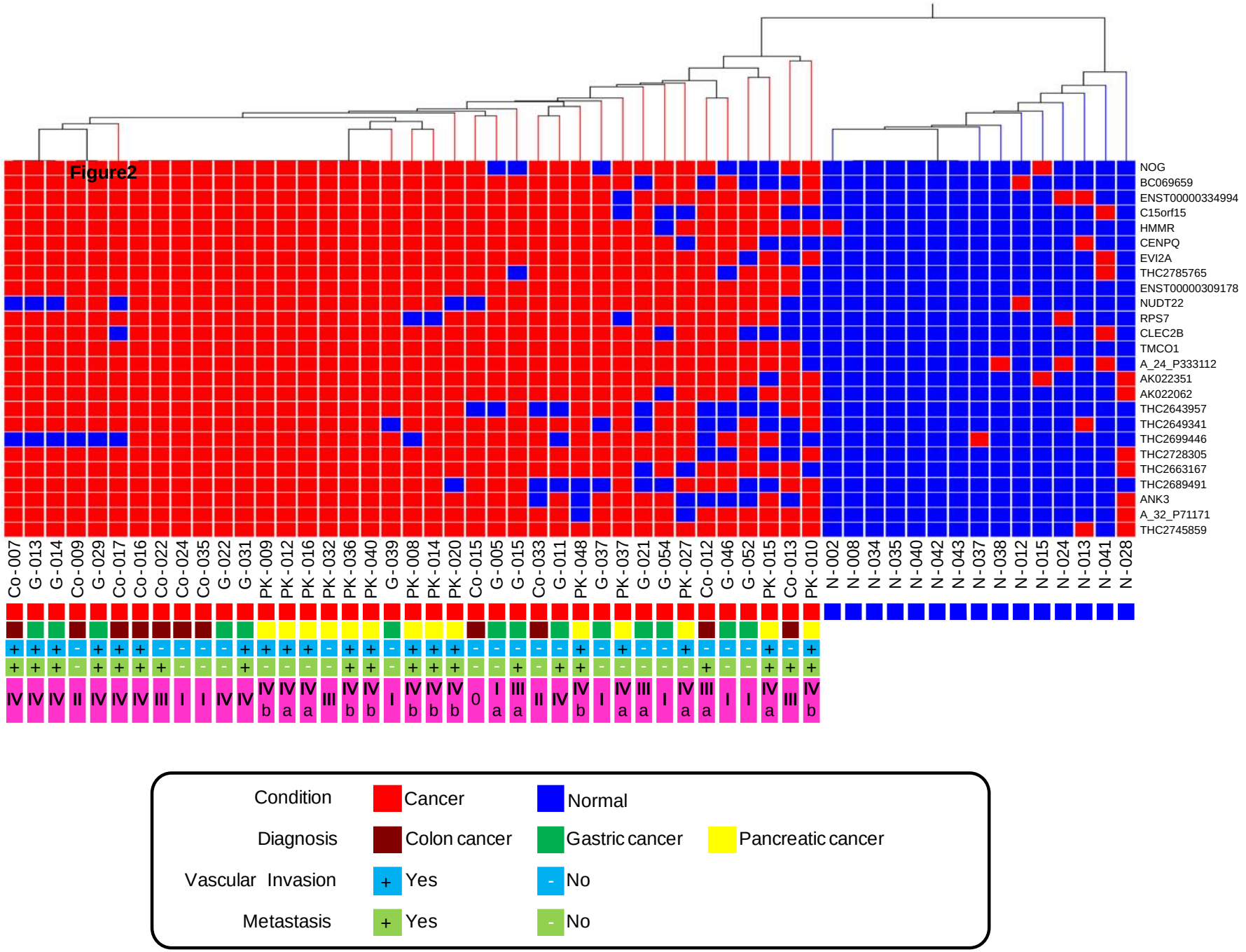
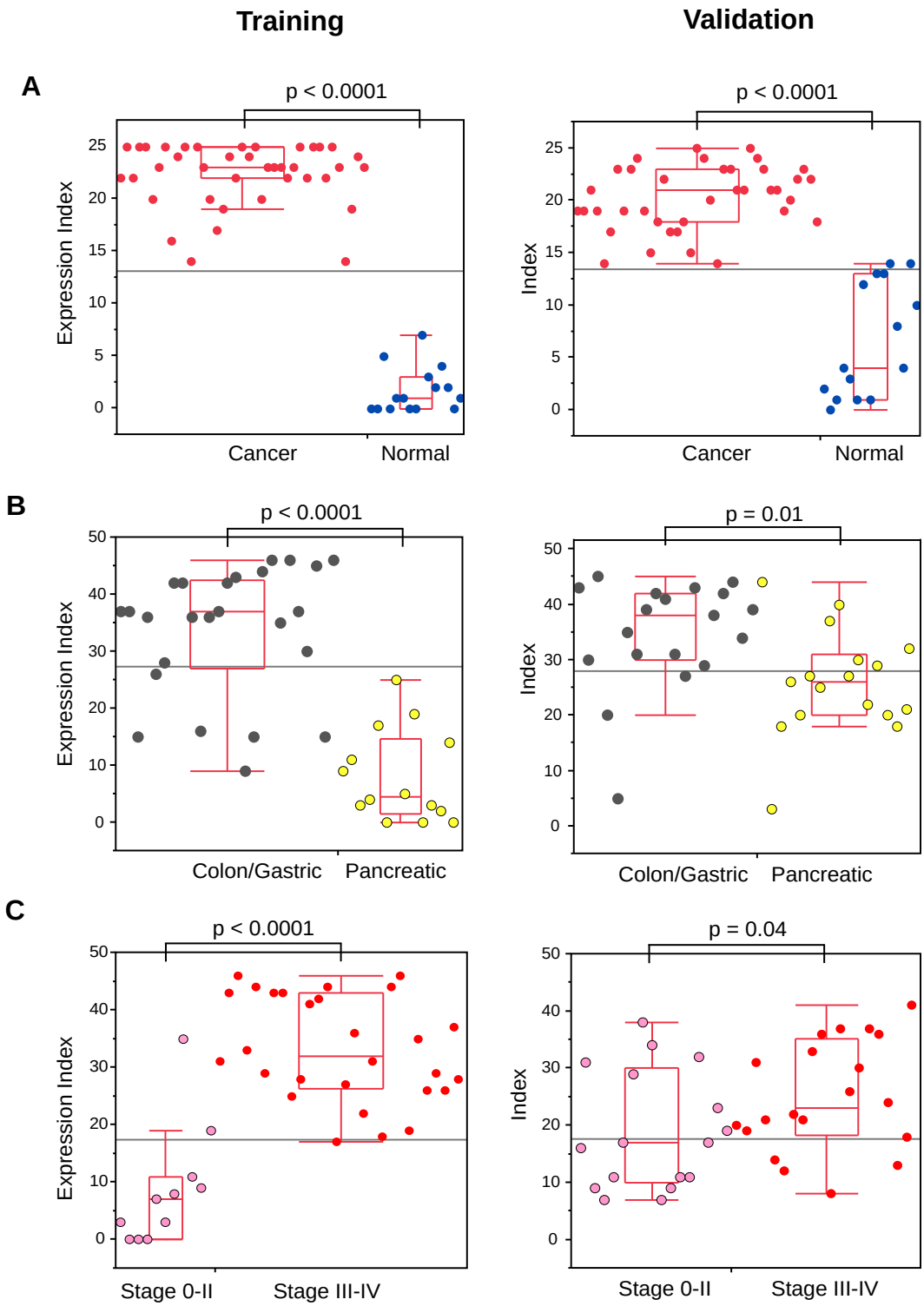

Figure2
NOG
BC069659
ENST00000334994
C15orf15
HMMR
CENPQ
EVI2A
THC2785765
ENST00000309178
NUDT22
RPS7
CLEC2B
TMCO1
A_24_P333112
AK022351
AK022062
THC2643957
THC2649341
THC2699446
THC2728305
THC2663167
THC2689491
ANK3
A_32_P71171
THC2745859
Co-007 G-013 G-014 Co-009 G-029 Co-017 Co-016 Co-022 Co-024 Co-035 G-022 G-031 PK-009 PK-012 PK-016 PK-032 PK-036 PK-040 G-039 PK-008 PK-014 PK-020 Co-015 G-005 G-015 Co-033 G-011 PK-048 G-037 PK-037 G-021 G-054 PK-027 Co-012 G-046 G-052 PK-015 Co-013 PK-010 N-002 N-008 N-034 N-035 N-040 N-042 N-043 N-037 N-038 N-012 N-015 N-024 N-013 N-041 N-028
IV IV IV II IV IV IV III I I IV IV IVb IVa IVa III IVb IVb I IVb IVb IVb 0 Ia IIIa II IV IVb I IVa IIIa I IVa IIIa I I IVa III IVb
Condition Cancer Normal
Diagnosis Colon cancer Gastric cancer Pancreatic cancer
Vascular Invasion + Yes - No
Metastasis + Yes - No

Figure3

Fig 3



SupplementaryTable1

Click here to download Electronic Supplementary Material (online publication only): Supple Table1.doc

SupplementaryTable2

Click here to download Electronic Supplementary Material (online publication only): Supple Table2.doc

SupplementaryTable3

Click here to download Electronic Supplementary Material (online publication only): Supple Table3.doc

SupplementaryFigure1

Click here to download Electronic Supplementary Material (online publication only): Fig Supple 1.pdf

SupplementaryFigure2

Click here to download Electronic Supplementary Material (online publication only): Fig supple 2.pdf

SupplementaryFigure3

Click here to download Electronic Supplementary Material (online publication only): Fig supple 3.pdf

SupplementaryFigure4

Click here to download Electronic Supplementary Material (online publication only): Fig supple 4.pdf