Multiple cancer type classification by small RNA expression profiles with plasma samples from multiple facilities

Kuno Suzuki1 | Hideyoshi Igata2 | Motoki Abe2 | Yusuke Yamamoto3 | small RNA based cancer classification project

Original Article

Abstract

Liquid biopsy is expected to be a promising cancer screening method because of its low invasiveness and the possibility of detecting multiple types in a single test. In the last decade, many studies on cancer detection using small RNAs in blood have been reported. To put small RNA tests into practical use as a multiple cancer type screening test, it is necessary to develop a method that can be applied to multiple facilities. We collected samples of eight cancer types and healthy controls from 20 facilities to evaluate the performance of cancer type classification. A total of 2,475 cancer samples and 496 healthy control samples were collected using a standardized protocol. After obtaining a small RNA expression profile, we constructed a classification model and evaluated its performance. First, we investigated the classification performance using samples from five single facilities. Each model showed areas under the receiver curve (AUC) ranging from 0.67 to 0.89. Second, we performed principal component analysis (PCA) to examine the characteristics of the facilities. The degree of hemolysis and the data acquisition period affected the expression profiles. Finally, we constructed the classification model by reducing the influence of these factors, and its performance had an AUC of 0.76. The results reveal that small RNA can be used for the classification of cancer types in samples from a single facility. However, interfacility biases will affect the classification of samples from multiple facilities. These findings will provide important insights to improve the performance of multiple cancer type classifications using small RNA expression profiles acquired from multiple facilities.

Keywords
liquid biopsy, machine learning, multiple cancer type classification, multiple facilities, NGS, small RNA
1 | INTRODUCTION

Globally, the number of cancer patients has been increasing. It is important to detect and treat cancer at an early stage because the survival rate decreases significantly as the disease progresses. Various methods are used for cancer screening, including computed tomography, endoscopy, X-ray, and PET. These screening tests are effective in reducing cancer deaths. However, undergoing numerous screening tests is physically hard, time-consuming, and costly for patients.

As a solution to these problems, liquid biopsy is expected to be a promising cancer-screening method because of its low invasiveness and ability to detect multiple cancer types simultaneously. Circulating tumor DNA (ctDNA) and small RNAs (microRNA [miRNA] and PIWI-interacting RNA [piRNA]) are the focus of cancer screening tests. In a study of multiple cancer classification using ctDNA, eight types of cancers were successfully classified, with a median sensitivity of 70% and a specificity of over 99%. However, the classification performance of stage 1 was lower than that of stages 2 and 3. The performance of cancer classification by miRNAs has been reported. In gastric cancer, the sensitivity and specificity were 73% and 89%; in pancreatic and biliary tract cancers, they were 82% and 97%; in breast cancer, they were 97% and 83%; in cervical cancer, they were 84% and 90%; and in prostate cancer, they were 95% and 87%. In addition, several studies have reported that miRNAs can detect even early-stage cancers with high sensitivity, and it is expected that miRNAs will be useful as an early cancer screening test.

However, issues remain for miRNA-based multiple cancer type screening tests. First, the possibility of identifying cancer types by collecting samples from multiple facilities has not been determined. Second, there have been reports that small RNAs specific to cancer types have been identified, but the variety of small RNAs is not consistent among reports. To overcome these issues, we collected cancer and healthy control samples from multiple facilities with a standardized protocol. Then, the classification accuracy was evaluated by machine learning models constructed with the small RNA expression profiles.

2 | MATERIALS AND METHODS

2.1 | Clinical samples

In this study, 2475 cancer samples and 496 healthy control samples were collected from 20 facilities: BE, Kobe University; CH, Chiba University; DA, Saitama Medical University, International Medical Center; EH, National Cancer Center Hospital East; GM, Center Hospital of the National Center for Global Health and Medicine; HC, National Cancer Center Hospital, Cancer Screening Center; HS, Harasanshin Hospital; JD, Juntendo University; KB, Kyoto University; KC, Kanagawa Cancer Center; KH, Kurosawa Hospital; NC, National Cancer Center Hospital; NY, Nagoya University; OC, Osaka International Cancer Institute; OY, Okayama University; SG, Sagara Hospital; SU, Showa University; TA, Osaka City University; TM, Tokyo Medical University; and WG, Saitama Medical University, Saitama Medical Center. Esophageal cancer (ESO), gastric cancer (GAS), colorectal cancer (COL), liver and biliary tract cancer (LBI), pancreatic cancer (PAN), lung cancer (LUN), breast cancer (BRE), and prostate cancer (PRO) samples from stage 0 to 4 were obtained. The stage classification was based on the 8th edition of the TNM classification published by the Union for International Cancer Control. The ICD codes for each cancer type were ESO (C15), GAS (C16), COL (C18, 19, 20), LBI (C22, 23, 24), PAN (C25), LUN (C34), BRE (C50), and PRO (C61). All cancer samples were collected before the start of treatments. Healthy control (CTR) samples were obtained according to the criteria, with no obvious suspicion based on a comprehensive evaluation of imaging studies, blood tests, and clinical findings. Samples were collected by KB from 2013 to 2016; HC from 2016 to 2018; and NC, EH, OC, KC, JD, TM, HS, SG, GM, KH, WG, DA, OY, BE, CH, TA, and SU from 2020 to 2021. The details are described in Table S1. Informed consent was obtained from the participants. This research was approved by the National Cancer Center Hospital Institutional Review Board (approval number: 2018–200) and was conducted according to the guidelines of the Declaration of Helsinki.

2.2 | Plasma sample collection

Blood samples were collected in vacuum tubes with EDTA-2Na and refrigerated within 30 min. The blood samples were centrifuged only once at 1,500–1,900 g for 10 min within 12 h of venipuncture, and plasma fractions were collected. The plasma samples were stored at −80°C until further analysis, and more than two freeze–thaw cycles were avoided.

2.3 | RNA purification

Small RNAs were automatically extracted from 300 μL of plasma using the Maxwell RSC miRNA Plasma and Serum Kit (Promega) with the Equipment Maxprep Liquid Handler (Promega) and Maxwell RSC (Promega). The purifications were performed according to the manufacturer’s instruction manual.

2.4 | Small RNA sequence data acquisition

cDNA libraries were synthesized from 5 μL of RNA eluate with 22 cycles of PCR using the QIAseq miRNA Library Kit (QIAGEN) with the Biomek i5 Automated Liquid Handling Workstation (Beckman Coulter). The concentration of each amplification product was measured with a QuantiFluor ONE ds DNA System (Promega). cDNA libraries were merged in one tube at a final concentration of 1.0–1.5 pmol/L. Small RNA sequence data were obtained using NextSeq 550Dx with the NextSeq 500/550 High Output Kit v2.5, 75 cycles.
(Illumina) and the NextSeq Phix Control Kit (Illumina). An automated program was set up for each instrument according to the manufacturer’s instruction manual, and all the samples were processed using the same program.

2.5 | Sequence data acquisition

To avoid bias toward certain cancer types in a particular experimental batch, data acquisition was conducted under the following conditions. First, a batch consisted of at least three categories of samples. Second, the number of samples from a single category should be < 50% of the total.

2.6 | Small RNA expression analysis

The raw sequence data were processed using fastp (version 0.19.6)\(^\text{26}\) to remove the amplification adapter sequence. The unique molecular index (UMI) sequences and the probe adapter were trimmed using umi_tools (version 1.0.1).\(^\text{27}\) Mapping was performed using Bowtie (1.3.0) with reference to the human genome reference (hg19). The following options were applied: -n0, -v0, -i18, -k1, and -best. After sorting with SAMtool (version 1.7),\(^\text{28}\) an index was created and deduplicated with umi_tools. Then, using featureCounts (version 2.0.0),\(^\text{29}\) 2,576 miRNAs and 27,683 piRNAs were identified, and the expression levels of each gene were calculated using the following annotation files: miRBase (version 20)\(^\text{30}\) with QIAseq Spikeins and piRNAdb (version 1.7.5).\(^\text{31}\) The average total count of miRNAs and piRNAs was 4.27 ± 1.09 × 10⁶ (mean ± SD; min: 440,695, max: 8,890,949). Samples with a total count of < 1,000,000 miRNAs and piRNAs were omitted from subsequent analyses. For each of 2,576 miRNAs and 27,683 piRNAs, DESeq2\(^\text{32,33}\) was used for normalization. Features were removed if they met the following criteria: (i) genes with a minimum expression level of 50 or less in all samples, and (ii) Pearson correlation coefficient with miR-369-3p of |r| > 0.5 to exclude miRNAs that are under the control of genomic imprinting.\(^\text{34}\) The remaining 534 mi/piRNAs were normalized by DESeq2 and used for subsequent analyses.

2.7 | Machine learning for multiple cancer classification

XGBoost was used to classify the cancer types.\(^\text{35}\) Each experiment was conducted 50 times with different random seeds, and the dataset was randomly divided into 90% and 10% partitions for training and testing, respectively. Hyperparameters were optimized by Optuna\(^\text{36}\) to maximize the fivefold cross-validation accuracy or the area under the receiver operating characteristic (ROC) curve (AUC) score. The search ranges of the parameters were as follows: learning_rate [0.01, 0.4], max_depth [3, 10], min_child_weight [0.1, 10], gamma [1e-8, 1.0], and colsample_bytree [0.6, 0.95]. For the constructing classifier, the samples were randomly selected so that each class would have the same number of samples in each experiment. The SHAP value, and the contribution of all small RNAs was quantified for each cancer type.\(^\text{37}\) of the cancer type classification model was calculated

2.8 | Statistical analysis

The correlation analysis, principal component analysis (PCA), Cohen’s d value and box plots were performed using the statistical analysis software R (version 3.6.3).\(^\text{38}\) The effect size was calculated from Cohen’s d value.

2.9 | Calculation of the hemolysis index

The hemolysis index was calculated from the small RNA expression counts of each sample using miR-23a-3p for the miR-451a expression levels.\(^\text{37,39,40}\)

The following formula was used: Hemolysis index = \(\log_2(\text{miR-451a}) - \log_2(\text{miR-23a-3p})\).

2.10 | Feature selection based on principal component analysis loading factors

For the first and second principal components in the PCA, the factor loadings of each miRNA were calculated, the threshold values were set, and the small RNAs were removed from the data according to the threshold values. The small RNA lists for 10%, 15%, 20%, and 25% removal are shown in Table S2.

2.11 | One-dimensional nearest neighbor matching

The x-axis (PC1) and y-axis (PC2) of each sample in the PCA were extracted, and the matching rate was defined as the probability that both sides of an arbitrary point projected on the x-axis or y-axis were in the same class.

3 | RESULTS

3.1 | Study design

Cancer and healthy control plasma samples were collected from multiple facilities to construct and validate the cancer type classification model. A total of 2,475 cancer samples and 496 healthy control samples were collected. Information on the age, sex, and stage of the collected samples is summarized in Table 1.

For these 2,971 samples, the expression levels of small RNAs (miRNAs and piRNAs) were measured by next-generation
sequencing (NGS), and 583 types of small RNAs were detected. Thirty-four imprinted genes located on chromosome 14q32.31 were removed from our analysis because these genes were differentially expressed in healthy controls. A total of 534 small RNAs were used to construct the cancer classification models. First, samples of the multiple cancer types were collected from a single facility, and the

### Table 1 Characteristics of the participants

| Category | Number of samples | Age Mean SD | Sex (%) | Stage (%) |
|----------|-------------------|-------------|---------|-----------|
|          |                   |             | Male    | Female    |
|          |                   |             | 0 1 2 3 4 |           |
| ESO      | 251               | 65.7 8.9    | 182 (73) | 69 (27)   |
|          |                   |             | 10 (4) 78 (31) | 51 (20) 65 (26) 47 (19) |
| GAS      | 282               | 65.3 12.8   | 171 (61) | 111 (39)  |
|          |                   |             | 0 144 (51) 43 (15) | 41 (15) 54 (19) |
| COL      | 335               | 64.7 11.8   | 181 (54) | 154 (46)  |
|          |                   |             | 11 (3) 86 (26) 83 (25) | 96 (29) 59 (18) |
| LBI      | 321               | 69.9 9.2    | 214 (67) | 107 (33)  |
|          |                   |             | 3 (1) 69 (21) 101 (31) | 72 (22) 76 (24) |
| PAN      | 315               | 66.9 10.5   | 174 (55) | 141 (45)  |
|          |                   |             | 5 (2) 32 (10) 75 (24) | 56 (18) 147 (47) |
| LUN      | 335               | 67.9 10.8   | 216 (64) | 119 (36)  |
|          |                   |             | 9 (3) 129 (39) 36 (11) | 60 (18) 101 (30) |
| BRE      | 350               | 56.0 12.5   | 0       | 350 (100) |
|          |                   |             | 42 (12) 161 (46) 111 (32) | 26 (7) 10 (3) |
| PRO      | 286               | 68.6 7.0    | 286 (100) | 0         |
|          |                   |             | 0 45 (15) 192 (67) | 35 (12) 14 (5) |
| CTR      | 496               | 49.6 10.9   | 252 (51) | 244 (49)  |

**Figure 1** Schematic of cancer type classification model establishment. Plasma samples were obtained from 2971 subjects, including 496 healthy controls (CTR) and 2475 cases. The small RNA sequences were matched to a database, and 583 genes were identified. Imprinted genes and genes with low expression were removed, and 534 small RNAs were retained as features for the machine learning models. Cancer classification Models 1–5 were constructed using samples obtained from five facilities, and Model 6 was constructed using samples from all facilities.
performance of the cancer type classification in each of the five different facilities was examined using Models 1–5. Next, samples collected from multiple facilities were used to evaluate the performance of the classification model (Model 6) for multiple cancer types to determine whether the same results could be obtained across facilities (Figure 1).

### 3.2 Verification of the reproducibility of data acquisition

To identify cancer type-specific expression profiles, it is necessary to have high reproducibility in data acquisition. To investigate the reproducibility of this study, six different reference samples were examined. For these reference samples, freezing and thawing were performed in the same cycle numbers to avoid unexpected sample alterations.

The correlation coefficients of the small RNA expression profiles between the experiments were calculated. Among the six reference samples, the highest mean value of the correlation coefficient was 0.993 (minimum 0.990, maximum 0.995), as shown in Figure 2A,C, and the lowest mean value was 0.981 (minimum 0.969, maximum 0.992), as shown in Figure 2B,D. In the series of experiments, the correlation coefficients ranged from 0.969 to 0.995 (Figure S1). These results verified that the data acquisition was highly reproducible.

### 3.3 Classification performance using each single-facility sample

The small RNA expression profiles of samples from multiple cancer types that were collected from five single facilities were obtained to evaluate the classification performance (Models 1–5). The classification models were constructed by downsampling with the fewest number of samples to equalize the cancer types. The facilities, cancer types, and number of samples used to construct the models are summarized in Table 2. The cancer type classification performance outcomes of Models 1–5 are shown in Figure 3. The AUCs of each model are shown in Figure S2. The sensitivities of Model 1 using COL, ESO, LBI, PAN, GAS, and CTR samples collected at KB were 69%, 58%, 47%, 52%, 45%, and 73%, respectively. These were higher than the sensitivity of 17% for six-class random classification. The AUC for the KB samples was 0.89 (95% confidence interval [CI], 0.88–0.90), as shown in Figure 3A. The sensitivities of Model 2 using BRE, COL, LUN, PAN, and GAS samples collected at OC were 83%, 30%, 38%, 34%, and 50%, respectively. These were higher than the sensitivity of 20% for five-class random classification. The AUC for the OC samples was 0.77 (95% CI, 0.75–0.78), as shown in Figure 3B. The sensitivities of Model 3 using BRE, LUN, PRO, and GAS samples collected at KC were 56%, 42%, 52%, and 56%, respectively. These were higher than the sensitivity of 25% for four-class random classification. The AUC was 0.76 (95% CI, 0.74–0.78), as shown in Figure 3C. The sensitivities of Model 4 using COL, ESO, PAN, and GAS samples collected at NC were 31%, 52%, 38%, and 47%, respectively. These were higher than the sensitivity of 25% for four-class random classification. The AUC was 0.67 (95% CI, 0.63–0.70), as shown in Figure 3D. The sensitivities of Model 5 using BRE, COL, LBI, and LUN samples collected at JD were 59%, 40%, 48%, and 52%, respectively. These were higher than the sensitivity or 25% for four-class random classification. The AUC was 0.77 (95% CI, 0.74–0.79), as shown in Figure 3E. Based on the above classification performance outcomes of the five facilities, Models 1–5 using samples from a single facility showed higher classification performance outcomes than random classification. These data suggest that a machine learning model with small RNA profiles can classify multiple cancer types at least within a single facility.

Next, we evaluated the influence of stage on performance in cancer classification. Stages 0, 1, and 2 were grouped together as early stages, and stages 3 and 4 were grouped together as advanced stages. We selected KB and OC based on sample numbers and compared their performance outcomes (Table S3). KB samples showed the following sensitivities: COL (early 58%, advanced 44%), ESO (early 42%, advanced 67%), GAS (early 71%, advanced 51%), PAN (early 53%, advanced 65%), and LBI (early 56%, advanced 54%), as shown in Figure S3A. OC samples showed the following sensitivities: COL (early 57%, advanced 57%), LUN (early 67%, advanced 31%), and PAN (early 21%, advanced 53%), as shown in Figure S3B. Early stages showed equal to or higher sensitivities than advanced stages in five out of eight classifications.

### 3.4 Principal component analysis using multiple-facility samples

To explore facility-specific biases, we performed PCA using multiple-facility samples. The color-coded PCA by cancer type, facility, and data acquisition period are shown in Figure 4A,B,C. No clear clusters of cancer types were observed. The data acquisition facilities and the first principal component (PC1) were related, and clusters were formed in some facilities. Regarding the data acquisition period, clusters were formed in relation to the second principal component (PC2).

To further clarify the reason for the clustering in some facilities, the characteristics of the top 10 miRNAs of the factor loadings of PC1 of each small RNA were investigated. Hsa-miR-144-3p, hsa-miR-32-5p, and hsa-miR-96-5p were included in the top 10 miRNAs. These small RNAs are known to be abundant in red blood cells (RBCs). Furthermore, other RBCs-derived miRNAs, such as hsa-miR-451a, hsa-miR-486-5p, hsa-miR-4732-3p, and hsa-miR-363-3p, were enriched in higher rankings in PC1 factor loadings (Table S4). The formation of clusters at different facilities was thought to be due to differences in the degree of hemolysis in the samples. Therefore, the hemolysis index of each sample was colored in the plot. The results showed that the increase or decrease in PC1 was linked to the hemolysis index (Figure S5A). Then, the distributions of the hemolysis index at collection facilities were compared. The distribution range of
The hemolytic index was different among facilities (Figure 5B). In fact, the distribution of the hemolytic index in KC was different from that in KH, and the clusters of both in PCA were also clearly separated (Figure 5C). The distribution of the hemolytic index in KB was near that in KC, and the clusters of both in PCA were also close (Figure 5D). These results suggest that PC1 reflects the effect of hemolysis and that there were differences in the hemolysis index among the facilities, which led to the formation of clusters.

**TABLE 2 Number of samples used in each classification model**

| Model | Facility | Category and number of samples |
|-------|----------|--------------------------------|
|       | ESO      | GAS | COL | LBI | PAN | LUN | BRE | PRO | CTR |
| 1     | KB       | 100 | 100 | 100 | 134 | 100 | 100 |
| 2     | OC       | 66  | 80  | 66  | 72  | 104 |
| 3     | KC       | 48  | 115 | 92  | 49  |
| 4     | NC       | 46  | 33  | 45  | 40  |
| 5     | JD       | 37  | 31  | 28  | 49  |
It was suggested that the hemolysis index and data acquisition period could affect the cancer classification performance. Therefore, we attempted to reduce these effects by removing the small RNAs that have large contributions to the PC1 and PC2 scores. PCA was performed again using the data from which 10%, 15%, 20%, and 25% of small RNAs in both positive and negative directions ranked by factor loadings were removed. The numbers of remaining features were

FIGURE 3 Classification performance outcomes of models constructed from single-facility samples. Classification models of multiple cancer types were constructed only for cancer types with a sample size of 25 or more. The number of samples for each cancer type was downsampled to 100 in Model 1 for KB (A), 66 in Model 2 for OC (B), 48 in Model 3 for KC (C), 33 in Model 4 for NC (D), and 28 in Model 5 for JD (E)

FIGURE 4 PCA using samples from multiple facilities. The results of PCA using 2967 samples from all facilities. The results are color-coded to indicate the categories (A), facilities with a sample size of 150 or more (B), and data acquisition periods (C)

3.5 Classification performance using multiple-facility samples

It was suggested that the hemolysis index and data acquisition period could affect the cancer classification performance. Therefore,
534 for no removal (Figure 6A), 324 for 10% removal (Figure 6B), 232 for 15% removal (Figure 6C), 160 for 20% removal (Figure 6D), and 98 for 25% removal (Figure 6E). The retained small RNAs are listed in Table S2. As the removal rate increased, these clusters converged into a group (Figure 6E).

To quantitate the change in cluster formation with/without the removal process, the probability that the nearest neighbor points of an arbitrary point were in the same class was calculated. In PC1, the matching rates focusing on cancer type were almost the same with/without small RNA removal, but the matching rates focusing on facility were reduced: 5.0% (without removal), 6.2% (10% removal), 5.7% (15% removal), 3.6% (20% removal), and 2.6% (25% removal). The matching rates of the data acquisition period were reduced to 5.5% (without removal), 3.0% (10% removal), 3.0% (15% removal), 2.4% (20% removal), and 2.2% (25% removal), as shown in Figure 6F. The reducing effect was largest under 25% removal. In PC2, the matching rates of cancer types and facilities were almost the same with/without small RNA removal, but the matching rates of the data acquisition period were reduced to 7.6% (without removal), 13.0% (10% removal), 11.0% (15% removal), 4.1% (20% removal), and 4.3% (25% removal), as shown in Figure 6G. The reducing effect was largest under 25% removal. These results indicate that the removal of small RNAs, which have a large contribution to the PC1 and PC2 scores, can reduce the impact on the hemolysis index and the data acquisition period.

Therefore, we evaluated the sensitivities of the multiple cancer type classification model constructed using samples from multiple facilities after the removal of small RNAs. The sensitivities before the removal of small RNAs were as follows: CTR 67%, BRE 59%, COL 32%, ESO 44%, LBI 53%, LUN 27%, PAN 29%, PRO 54%, and GAS 34% (Figure 7A). The AUC was 0.83 (95% CI, 0.83–0.84). The sensitivities after the removal of 25% small RNAs were as follows (Model 6): CTR 58%, BRE 47%, COL 19%, ESO 41%, LBI 44%, LUN 21%, PAN 23%, PRO 39%, and GAS 23% (Figure 7B). The AUC was 0.76 (95% CI, 0.76–0.76). The random classification sensitivity for nine classes was 11%. The results showed that the sensitivity of the classification of multiple cancer type samples from multiple facilities was higher than that of random classification.
In this study, the performance of the cancer classification models was evaluated by collecting samples of multiple cancer types from multiple facilities. As a result, the performance of the models was higher than that of random classification in both the single-facility and multiple-facility models. This finding suggests the existence of miRNAs that could identify each cancer type. However, the data

**FIGURE 6** Improvement of the clustering results by feature selection. (A) The results of PCA using 2967 samples from all facilities without any feature selection. The PCA result was shown that excluding the top 10%, 15%, 20%, and 25% of small RNAs from PC1 and PC2 factor loadings in both positive and negative directions. The number of retained features was 324 small RNAs (B), 232 small RNAs (C), 160 small RNAs (D), and 98 small RNAs (E), respectively. Each category of the matching rate for PC1 was shown in (F), and that for PC2 is shown in (G).

**FIGURE 7** Classification performance of the selected feature model from multiple-facility samples. The confusion matrix for the XGBoost classifier with all 534 features is shown in (A), and the confusion matrix with 98 selected features is shown in (B).

**4 | DISCUSSION**

In this study, the performance of the cancer classification models was evaluated by collecting samples of multiple cancer types from multiple facilities. As a result, the performance of the models was higher than that of random classification in both the single-facility and multiple-facility models. This finding suggests the existence of miRNAs that could identify each cancer type. However, the data
acquisition period and the degree of hemolysis in each facility could affect the cancer type classification performance.

One of the possible disturbing factors for the cancer type classification is batch effects. The KB confusion matrix showed that the six-class classification results were divided into roughly two groups (Figure 3A). Depending on the data acquisition periods, there were differences in the reagent lots, including polymerases and ligases. This suggested that an amplification bias of small RNA may have occurred. Small RNAs with high variability between the two groups were extracted by effect sizes (see Materials and Methods 2.8) from the single reference sample. The top five were hsa-miR-21-5p, hsa-miR-425-5p, hsa-piR-33057, hsa-piR-32993, and hsa-piR-32871; the effect sizes were 5.00, 2.60, 2.26, 2.06, and 2.04; and the fold changes were 1.64, 1.19, 1.77, 2.02, and 2.07, respectively. Although hsa-miR-21-5p is known as a representative cancer-associated miRNA, it showed an effect size of 0.59 and a fold change of 1.07 in GAS versus CTR and 0.47 and 1.07 in PAN versus CTR when comparing the same data acquisition periods. It was one-half to one-third smaller than that reported in similar studies.47-50 These results indicate that amplification bias occurred due to differences in the data acquisition period, which was thought to be the cause of the division into two groups in the KB confusion matrix. The change in small RNA expression of each cancer type in the same acquisition condition was smaller compared to the amplification bias. This situation could have affected the performance of the cancer type classification.

The consistency of cancer type-associated small RNAs between our classification models and previously reported models was investigated (Table S5). To avoid interfacility variability, we extracted cancer type-associated small RNAs from Models 1–5 by quantifying their contribution. ESO, GAS, COL, LBI, PAN, and BRE were selected because these cancer types were tested in more than two facilities. We narrowed down to the common miRNAs in multiple facilities and compared these common miRNAs with previously reported miRNAs. Two or three miRNAs in our common miRNAs were also reported in previous studies of cancer type-associated miRNAs. Our results were consistent with previous reports to some extent.

Regarding the effect of hemolysis on the classification performance, we collected samples with a standardized sample collection protocol (see Materials and Methods 2.2). Therefore, we expected that the interfacility variation in the hemolysis index would not exceed the intrafacility variation. However, it was found that the interfacility variability exceeded the intrafacility variability (Figure 5B). Thus, the differences in the hemolysis index among the facilities became noise, and the accuracy of cancer type classification decreased. At the same time, the differences between facilities could be misinterpreted as differences in cancer types, resulting in a higher accuracy than the values with minimized hemolysis influences (Figure 7A,B). As a result, the difference in the hemolysis index among facilities not only affected the performance of the cancer classification but also suggested the possibility that the differences among the facilities could be confounded by the differences in the cancer types.

Although the classification performance was higher than random classification in multiple cancer type classifications, further improvement is needed when considering the application as a cancer screening test. The expression levels of miRNAs in plasma and serum are reportedly similar,40 and we decided to use plasma samples to avoid differences in clot formation between samples when preparing serum. To improve the performance, two essential issues exist. First, the expression changes of small RNAs characterizing cancer types are not significant enough. To overcome this issue, it is necessary to efficiently extract cancer-associated small RNAs to increase expression changes. Second, improvement of the sample preparation method is necessary to reduce the differences in the degree of hemolysis among different facilities. A practical approach for efficiently extracting cancer-associated small RNAs is to purify and analyze small RNAs within exosomes after preventing the contamination of red blood cells in plasma samples using the double centrifugation method51,52 or prefiltering.53 Although the ultracentrifugation protocol54 is used as a standard method for exosome purification, it has poor throughput and is not suitable for automated sample preparation. Therefore, for practical use, the use of size exclusion chromatography55,56 or chemical precipitation57 would be helpful. As a sample preparation method to reduce the degree of hemolysis, it should be effective to collect blood samples using a blood collection tube58 that contains a reagent to inhibit the hemolysis of red blood cells. It is expected that these approaches will further improve the classification performance of multiple cancer types, thus establishing a cancer screening test with practical performance.

In conclusion, to develop a robust diagnosis test for the cancer type screening using small RNAs, we found that small RNAs can be used for the classification of different cancer types in samples from a single facility; however, interfacility noise affects the classification of samples from multiple facilities. This finding provides insight into the importance of enhancing the signal-to-noise ratio to overcome interfacility biases for the classification of multiple cancer types from samples derived from multiple facilities.

ACKNOWLEDGMENTS

We would like to thank the participants in this study for their involvement. We would like to thank Hiroyuki Hamada and Yuko Yamakage for NGS data acquisition; Hiroyuki Mano (National Cancer Center), Kosei Hasegawa (Saitama Medical University), Motohide Uemura (Osaka University), Nobuyuki Ota (Preferred Networks America), Shinya Fukumoto (Osaka City University), Shoji Nagao (Hyogo Cancer Center), and Tamotsu Sudo (Hyogo Cancer Center), for research advice; and Kiyohide Ishikura, who is a liaison officer in medical facility relations.

DISCLOSURE

KS, MK, MT, TN and TY are employees of PFDeNA. HI and MA are employees of Preferred Networks. All other authors have no conflicts of interest to declare.
REFERENCES

1. Leung WK, Wu M-S, Kakugawa Y, et al. Screening for gastric cancer in Asia: current evidence and practice. Lancet Oncol. 2008;9:279-287.

2. Japan VS Cancer registry and statistics. Cancer information service, national cancer center. Japan. Google Scholar.

3. Pastorino U, Silva M, Sestini S, et al. Prolonged lung cancer screening: the miR-Test Case. Front Genet. 2019;20:71-88.

4. Nakamura Y, Taniguchi H, Ikeda M, et al. Clinical utility of circulating tumor DNA sequencing in advanced gastrointestinal cancer: SCRUM-Japan GI-SCREEN and GOZILA studies. Nat Med. 2020;26:1859-1864.

5. Heitzer E, Haque IS, Roberts CES, et al. Current and future perspectives of liquid biopsies in genomes-driven oncology. Nat Rev Genet. 2019;20:71-88.

6. Shiotani A, Murao T, Kimura Y, et al. Identification of serum microRNAs as novel non-invasive biomarkers for detection of high risk for early gastric cancer. Br J Cancer. 2013;109:2323-2330.

7. Montani F, Marzi MJ, Dezi F, et al. miR-Test: a blood test for lung cancer early detection. J Natl Cancer Inst. 2015;107:dv063.

8. Huang Z, Huang D, Ni S, et al. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. Int J Cancer. 2010;127:118-126.

9. Chen S, Zhou Y, Chen Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:1864-1890.

10. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923-930.

11. Shimomura A, Shiino S, Kawauchi J, et al. Novel combination of miRNA profiling for ovarian cancer screening. Breast Cancer Res Treat. 2018;9:1879.

12. Montani F, Marzi MJ, Dezi F, et al. miR-Test: a blood test for lung cancer early detection. J Natl Cancer Inst. 2015;107:dv063.

13. Malnou EC, Umlauf D, Mouysset M, et al. Imprinted MicroRNA gene clusters in the evolution, development, and functions of mammalian placenta. Front Genet. 2018;9:4319.

14. Usuba W, Urabe F, Yamamoto Y, et al. Circulating miRNA panels for specific and early detection in bladder cancer. Cancer Sci. 2019;110:408-419.

15. Hennessey PT, Sanford T, Choudhary A, et al. Serum microRNA biomarkers for detection of non-small cell lung cancer. PLoS One. 2012;7:e32307.

16. Fang R, Zhu Y, Hu L, et al. Plasma MicroRNA pair panels as novel biomarkers for detection of early stage breast cancer. Front Physiol. 2018;9:1879.

17. Shah JS, Soon PS, Marsh DJ. Comparison of methodologies to detect low levels of hemolysis in serum for accurate assessment of serum microRNAs. PLoS One. 2016;11:e0153200.

18. Sartori V, Soares R, Terzi E, Karypis G. Comparison of methodologies to detect low levels of hemolysis in serum for accurate assessment of serum microRNAs. PLoS One. 2016;11:e0153200.

19. McDonald JS, Milosevic D, Reddi HV, et al. Analysis of circulating microRNA: preanalytical and analytical challenges. Clin Chem. 2011;57:833-840.
43. MacLellan SA, MacAulay C, Lam S, et al. Pre-profiling factors influencing serum microRNA levels. BMC Clin Pathol. 2014;14:27.

44. Pizzamiglio S, Zanutto S, Ciniselli CM, et al. A methodological procedure for evaluating the impact of hemolysis on circulating microRNAs. Oncof Lett. 2017;13:315-320.

45. Sun L, Yu Y, Niu B, et al. Red blood cells as potential repositories of MicroRNAs in the circulatory system. Front Genet. 2020;11:442.

46. Wu CW, Cao X, Berger CK, et al. Novel approach to fecal occult blood testing by assay of erythrocyte-specific microRNA markers. Dig Dis Sci. 2017;62:1985-1994.

47. Sierzega M, Kaczor M, Kolodziejczyk P, et al. Evaluation of serum microRNA biomarkers for gastric cancer based on blood and tissue pools profiling: the importance of miR-21 and miR-331. Br J Cancer. 2017;117:266-273.

48. Wang J, Chen J, Chang P, et al. MicroRNAs in plasma of pancreatic ductal adenocarcinoma patients as novel blood-based biomarkers of disease. Cancer Prev Res. 2009;2:807-813.

49. Matamed M, Hashemzadeh Chaleshtori M, Ghasemi S, et al. Plasma level of miR-21 And miR-451 in primary and recurrent breast cancer patients. Breast Cancer. 2019;11:293-301.

50. Porzyczki P, Ciszkwoczewicz E, Semik M, et al. Combination of three microRNA (miR-141, miR-21, and miR-375) as potential diagnostic tool for prostate cancer recognition. Int Urol Nephrol. 2018;50:1619-1626.

51. Cheng HH, Yi HS, Kim Y, et al. Plasma processing conditions substantially influence circulating microRNA biomarker levels. PLoS One. 2013;8:e64795.

52. Mitchell AJ, Gray WD, Hayek SS, et al. Platelets confound the measurement of extracellular microRNA in archived plasma. Sci Rep. 2016:6:32651.

53. Nik Mohamed Kamal NNSB. Shahidan WNS. Non-exosomal and exosomal circulatory MicroRNAs: which are more valid as biomarkers? Front Pharmacol. 2019;10:1500.

54. Théry C, Amigorena S, Raposo G, et al. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr Protoc Cell Biol. 2006.Chapter 3:Unit 3.22.1-29.

55. Baranyai T, Herczeg K, Onodi Z, et al. Isolation of exosomes from blood plasma: qualitative and quantitative comparison of ultracentrifugation and size exclusion chromatography methods. PLoS One. 2015;10:e0145686.

56. Sidhom K, Obi PO, Saleem A. A review of exosomal isolation methods: is size exclusion chromatography the best option? Int J Mol Sci; 2020;21(18):6466. doi: 10.3390/ijms21186466.

57. Rekker K, Saare M, Roost AM, et al. Comparison of serum exosome isolation methods for microRNA profiling. Clin Biochem. 2014;47:135-138.

58. Qin J, Alt JR, Hunsley BA, et al. Stabilization of circulating tumor cells in blood using a collection device with a preservative reagent. Cancer Cell Int. 2014;14:23.

APPENDIX 1

Author names: Kuno Suzuki, Hideyoshi Igata, Terunao Iwanaga, Hiroaki Kanzaki, Naoya Kato, Nobuko Tanaka, Kenji Kawasaki, Kazuyuki Matsushita, Ryu Samashima, Keiji Tsukino, Akira Yokomizo, Yosuke Miyashita, Issei Sumiyoshi, Kazuhsia Takahashi, Nobuko Serizawa, Ko Tomishima, Akihito Nagahara, Yumiko Ishizuka, Yoshiha Horimoto, Masayoshi Nagata, Keisuke Ishikawa, Shigeo Horie, Shuichiro Shinia, Motomi Nasu, Takashi Hashimoto, Shinji Mine, Shingo Kawano, Kiichi Sugimoto, Kazuhiro Sakamoto, Hiroyuki Takemura, Mitsuru Wakita, Yoko Tabe, Shunsuke Kato, Yohei Miyagi, Hiroyuki Adachi, Tetsuya Isaka, Hiroyuki Ito, Takashi Yamakana, Tatsuya Yoshida, Toshinari Yamashita, Takashi Ogata, Takanobu Yamada, Takashi Oshima, Naoto Yamamoto, Masaaki Murakawa, Soichiro Morinaga, Satoshi Kobayashi, Shun Tezuka, Makoto Ueno, Mitsuuyuki Koizumi, Kimito Osaka, Takeshi Kishida, Sumito Sato, Yo Mikayama, Manabu Shiozawa, Yasuhiro Inokuchi, Mitsuhiro Furuta, Nozomu Machida, Shinya Sato, Yosihiko Yano, Atsushi Miwa, Kazuto Ito, Isao Kurosawa, Osamu Kikuchi, Hiromitsu Tazawa, Manabu Muto, Takashi Honda, Hiroki Kawashima, Masatoshi Ishigami, Yutaka Saito, Hiroyuki Diako, Takashi Yoshida, Yukihide Kanemitsu, Ken Kato, Minoru Esaki, Takui Oksaka, Hiromi Sakamoto, Teruhiko Yoshida, Takahiro Ochiya, Mitsuhito Sasaki, Masafumi Ikeda, Masashi Kudo, Naoto Gotohda, Shuichi Mitsunaga, Takeshi Kowata, Takashi Kojima, Tatsuro Murano, Tomonori Yano, Taiki Yamaji, Takahisa Matsuda, Shuichiyo Tsugane, Kazuki Hashimoto, Kazuhiro Yamada, Nobuyuki Takemura, Kyoji Ito, Fuminori Mihara, Akihiro Shimomura, Kentoshi Shigeyasu, Kazuhiro Noma, Toshiyoshi Fujiwara, Hideki Yamamoto, Mizuki Morita, Shinichi Toyooka, Akihiro Tamori, Yasuhiro Akabori, Kenji Ikezawa, Kazuyoshi Ohtawa, Kei Kuniyama, Kazumi Nishino, Toru Kumagai, Toshihiro Kudo, Naotoshi Sugimoto, Masayoshi Yasui, Takeshi Omori, Hiroshi Miyata, Toru Kimura, Tomohiro Maniwa, Jiro Okami, Hiroki Kusama, Nobuyoshi Kitkata, Takahiro Nakayama, Masashi Nakayama, Yasutomo Nakai, Kazuo Nishimura, Shoji Yotsui, Takashi Yamamoto, Tomoyuki Yamasaki, Emi Yamashita, Kazune Saito, Keichi Yoshida, Masayuki Ohue, Masakazu Kohda, Tatsuya Yamaguchi, Masami Tanaka, Takashi Nishizawa, Tetsuhiko Taira, Junko Kawano, Yasuaki Sagara, Yosuke Horita, Yoshiaki Mihara, Tetsuya Hamaguchi, Okide Suzuki, Yoichi Kumagai, Hideyuki Ishida, Motoki Yamagishi, Hideaki Shimoyama, Haruaki Sasaki, Takehiko Nakasato, Takeshi Shichijo, Takashi Fukagai, Kota Nishimura, Kidai Hirayama, Masashi Morita, Yujin Kudo, Susumu Takeuchi, Norihiko Ikeda, Naohiro Kamoda, Kazunori Namiki, Yoshiho Ohno, Tomohiro Umezuz, Yoshihi Murakami, Masahiko Kuroda, Motoki Abe, Yusuke Yamamoto.

How to cite this article: Suzuki K, Igata H, Abe M, Yamamoto Y, small RNA based cancer classification project. Multiple cancer type classification by small RNA expression profiles with plasma samples from multiple facilities. Cancer Sci. 2022;113:2144–2166. doi:10.1111/cas.15309

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher’s website.
| First author | Corresponding author | Author name | ORCID | Affiliation | Contribution 1 | Contribution 2 | Contribution 3 |
|--------------|----------------------|-------------|-------|-------------|----------------|----------------|---------------|
| Co-first     | Co-corresponding     | Kuno Suzuki | 0000-0001-9776-9008 | Healthcare Business Department, PFDeNA, Inc., Tokyo, Japan | Sample selection | Data analysis | Project administration |
| Co-first     |                      | Hideyoshi Igata | 0000-0001-8342-0614 | Preferred Networks, Inc., Tokyo, Japan | Data analysis | Project administration |
|              |                      | Terunao Iwanaga |          | Department of Gastroenterology, Graduate School of Medicine, Chiba University, Chiba, Japan | Sample selection | Sample collection |
|              |                      | Hiroaki Kanzaki |          | Department of Gastroenterology, Graduate School of Medicine, Chiba University, Chiba, Japan | Sample selection | Sample collection | Project administration |
|              |                      | Naoya Kato |          | Department of Gastroenterology, Graduate School of Medicine, Chiba University, Chiba, Japan | | Supervisor |
|              |                      | Nobuko Tanaka |          | Department of Laboratory Medicine, Chiba University Hospital, Chiba, Japan | | Sample collection |
|              |                      | Kenji Kawasaki |          | Department of Laboratory Medicine, Chiba University Hospital, Chiba, Japan | Sample selection | Sample collection |
|              |                      | Kazuyuki Matsushita | 0000-0001-5812-2818 | Department of Laboratory Medicine & Division of Clinical Genetics and Proteomics/Center for Cancer Genomes/Center for Ultrasound, Chiba University Hospital | | Supervisor |
|              |                      | Ryu Samashima |          | Department of Urology, Harasanshin Hospital, Fukuoka, Japan | Sample selection | Sample collection |
|              |                      | Keiji Tsukino |          | Department of Urology, Harasanshin Hospital, Fukuoka, Japan | Sample selection | Sample collection |
|              |                      | Akira Yokomizo | 0000-0002-5741-6280 | Department of Urology, Harasanshin Hospital, Fukuoka, Japan | | Supervisor |
|              |                      | Yosuke Miyashita |          | Department of Respiratory Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection | Sample collection |
|              |                      | Issei Sumiyoshi |          | Department of Respiratory Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection | Sample collection |
|              |                      | Kazuhisa Takahashi |          | Department of Respiratory Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan | | Supervisor |
|              |                      | Nobuko Serizawa |          | Department of Gastroenterology, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection |
| First author  | Corresponding author | Author name               | ORCID                | Affiliation                                                                 | Contribution 1                      | Contribution 2                      | Contribution 3                      |
|--------------|----------------------|---------------------------|----------------------|------------------------------------------------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Ko Tomishima |                      | Akihito Nagahara          | Department of Gastroenterology, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection                  | Sample collection                   |                                     |
|              |                      | Yumiko Ishizuka           | Department of Breast and Endocrine Surgery, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection                  | Sample collection                   |                                     |
|              |                      | Yoshiya Horimoto          | Department of Breast and Endocrine Surgery, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection                  |                                     | Project administration             |
|              |                      | Masayoshi Nagata          | Department of Urology, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection                  |                                     |                                     |
|              |                      | Keisuke Ishikawa          | Department of Urology, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection                  | Sample collection                   |                                     |
|              |                      | Shigeo Horie              | Department of Urology, Juntendo University Graduate School of Medicine, Tokyo, Japan |                                   |                                     | Supervisor                           |
|              |                      | Shuichiro Shiina          | Department of Gastroenterological Imaging and Interventional Oncology, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection                  | Sample collection                   | Project administration             |
|              |                      | Motomi Nasu               | Department of Esophageal and Gastroenterological Surgery, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection                  | Sample collection                   |                                     |
|              |                      | Takashi Hashimoto         | Department of Esophageal and Gastroenterological Surgery, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection                  |                                     |                                     |
|              |                      | Shinji Mine               | Department of Esophageal and Gastroenterological Surgery, Juntendo University Graduate School of Medicine, Tokyo, Japan |                                   |                                     | Supervisor                           |
|              |                      | Shingo Kawano             | Department of Coloproctological Surgery, Juntendo University Faculty of Medicine, Tokyo, Japan | Sample selection                  | Sample collection                   |                                     |
|              |                      | Kiichi Sugimoto           | Department of Coloproctological Surgery, Juntendo University Faculty of Medicine, Tokyo, Japan | Sample selection                  |                                     |                                     |
| First author | Corresponding author | Author name | ORCID | Affiliation | Contribution 1 | Contribution 2 | Contribution 3 |
|--------------|----------------------|-------------|-------|-------------|----------------|----------------|----------------|
| Kazuhiro Sakamoto | | 0000-0002-7087-1667 | Department of Coloproctological Surgery, Juntendo University Faculty of Medicine, Tokyo, Japan | Supervisor |
| Hiroyuki Takemura | | | Department of Clinical Laboratory, Juntendo University Hospital, Tokyo, Japan | Sample selection Sample collection |
| Mitsuru Wakita | | | Department of Clinical Laboratory, Juntendo University Hospital, Tokyo, Japan | Sample selection Sample collection |
| Yoko Tabe | | | Department of Clinical Laboratory Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan | Sample selection Sample collection |
| Shunsuke Kato | | 0000-0002-4860-1687 | Department of Clinical Oncology, Juntendo University Graduate School of Medicine, Tokyo, Japan | Supervisor Sample selection |
| Yohei Miyagi | | | Molecular Pathology and Genetics Division, Kanagawa Cancer Center Research Institute, Yokohama, Japan | Supervisor |
| Hiroyuki Adachi | | | Department of Thoracic Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| Tetsuya Isaka | | | Department of Thoracic Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| Hiroyuki Ito | | | Department of Thoracic Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| Takashi Yamanaka | | | Department of Breast and Endocrine Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| Tatsuya Yoshida | | | Department of Breast and Endocrine Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| Toshinari Yamashita | | | Department of Breast and Endocrine Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| Takashi Ogata | | | Department of Gastrointestinal Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| Takanobu Yamada | | | Department of Gastrointestinal Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| Takashi Oshima | | | Department of Gastrointestinal Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| Naoto Yamamoto | | | Department of Gastrointestinal Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| Masaaki Murakawa | | | Department of Gastrointestinal Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection Sample collection |
| First author | Affiliation | Contribution 1 | Contribution 2 | Contribution 3 |
|--------------|-------------|----------------|----------------|----------------|
| Soichiro Morinaga | Department of Gastrointestinal Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Satoshi Kobayashi | Department of Gastroenterology, Hepatobiliary and Pancreatic Medical Oncology Division, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Shun Tezuka | Department of Gastroenterology, Hepatobiliary and Pancreatic Medical Oncology Division, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Makoto Ueno | Department of Gastroenterology, Hepatobiliary and Pancreatic Medical Oncology Division, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Mitsuyuki Koizumi | Department of Urology, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Kimito Osaka | Department of Urology, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Takeshi Kishida | Department of Urology, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Sumio Sato | Department of Gastrointestinal Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Yo Mikayama | Department of Gastrointestinal Surgery, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Manabu Satozawa | Department of Gastroenterology, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Yasuhito Inokuchi | Department of Gastroenterology, Kanagawa Cancer Center, Yokohama, Japan | Sample selection | Sample collection | Sample collection |
| Mitsunori Funata | Molecular Pathology and Genetics Division, Kanagawa Cancer Center Research Institute, Yokohama, Japan | Sample selection | Project administration | Sample collection |
| Nozomu Machida | Division of Gastroenterology, Department of Internal Medicine, Kobe University Graduate School of Medicine, Kobe, Japan | Sample selection | Sample collection | Sample collection |
| Shinya Sato | Department of Preventive Medicine, Kurosawa Hospital, Takasaki, Japan | Sample selection | Project administration | Sample collection |
| Yoshihiko Yano | Institute for Preventive Medicine, Kurosawa Hospital, Takasaki, Japan | Sample selection | Project administration | Sample collection |
| First author | Corresponding author | Author name | ORCID | Affiliation | Contribution 1 | Contribution 2 | Contribution 3 |
|--------------|----------------------|-------------|-------|-------------|----------------|----------------|----------------|
| Kazuto Ito   |                      | 0000-0002-  |       | Department of Urology, Kurosawa Hospital, Takasaki, Japan | Supervisor     |                |                |
|              |                      | 3735-2769   |       |             |                |                |                |
| Isao Kurosawa|                      |             |       | Department of Urology, Kurosawa Hospital, Takasaki, Japan | Supervisor     |                |                |
| Osamu Kikuchi|                      |             |       | Department of Therapeutic Oncology, Kyoto University Graduate School of Medicine, Kyoto, Japan | Project administration |                |                |
| Hiromitsu Tazawa |              |             |       | Clinical bioresource Center, Kyoto University Hospital, Kyoto, Japan | Sample selection | Sample collection |                |
| Manabu Muto  |                      |             |       | Department of Clinical Oncology, Kyoto University Hospital, Kyoto, Japan | Supervisor     |                |                |
| Takashi Honda|                      |             |       | Department of Gastroenterology and Hepatology, Nagoya University Graduate School of Medicine, Nagoya, Japan | Project administration | Sample selection | Sample collection |
| Hiroki Kawashima |                |             |       | Department of Gastroenterology and Hepatology, Nagoya University Graduate School of Medicine, Nagoya, Japan | Sample selection | Sample collection | Supervisor     |
| Masatoshi Ishigami |            |             |       | Department of Gastroenterology and Hepatology, Nagoya University Graduate School of Medicine, Nagoya, Japan | Sample selection | Sample collection | Supervisor     |
| Yutaka Saito |                      |             |       | Department of Endoscopy, National Cancer Center Hospital, Tokyo, Japan | Sample selection | Project administration |                |
| Hiroyuki Daiko |                    |             |       | Department of Esophageal Surgery, National Cancer Center Hospital, Tokyo, Japan | Sample selection | Project administration |                |
| Takaki Yoshikawa |                  |             |       | Department of Gastric Surgery, National Cancer Center Hospital, Tokyo, Japan | Sample selection | Project administration |                |
| Yukihide Kanemitsu |                |             |       | Department of Colorectal Surgery, National Cancer Center Hospital, Tokyo, Japan | Sample selection | Project administration |                |
| Ken Kato     |                      | 0000-0002-  |       | Department of Head and Neck, Esophageal Medical Oncology / Department of Gastrointestinal Medical Oncology, National Cancer Center Hospital, Tokyo, Japan | Sample selection | Project administration |                |
|              |                      | 1733-5072   |       |             |                |                |                |
| Minoru Esaki |                      |             |       | Department of Hepatobiliary and Pancreatic Surgery, National Cancer Center Hospital, Tokyo, Japan | Sample selection | Sample collection |                |
| Takuji Okusaka |                    |             |       | Department of Hepatobiliary and Pancreatic Oncology, National Cancer Center Hospital, Tokyo, Japan | Sample selection | Sample collection |                |
| First author | Corresponding author | Author name | ORCID | Affiliation                                                                 | Contribution 1                           | Contribution 2                           | Contribution 3                       |
|--------------|----------------------|-------------|-------|----------------------------------------------------------------------------|------------------------------------------|------------------------------------------|---------------------------------------|
| HIROMI SAKAMOTO |                      | HIROMI SAKAMOTO | 0000-0002-0776-9918 | Department of Biobank and Tissue Resources, Fundamental Innovative Oncology Core, National Cancer Center Research Institute, Tokyo, Japan | Sample selection | Sample collection |                                        |
| TERUHIKO YOSHIDA |                      | TERUHIKO YOSHIDA | 0000-0002-0776-9918 | Department of Biobank and Tissue Resources, Fundamental Innovative Oncology Core, National Cancer Center Research Institute, Tokyo, Japan | Sample selection | Sample collection |                                        |
| TAKAHIRO OCHIYA |                      | TAKAHIRO OCHIYA | 0000-0002-0776-9918 | Laboratory of Integrative Oncology, National Cancer Center Research Institute, Tokyo, Japan | Supervisor |                                        |                                        |
| MITSUHITO SASAKI |                      | MITSUHITO SASAKI | 0000-0002-0776-9918 | Department of Molecular and Cellular Medicine, Institute of Medical Science, Tokyo Medical University, Tokyo, Japan | Supervisor |                                        | Project administration                |
| MASAUMI IKEDA |                      | MASAUMI IKEDA | 0000-0002-0450-2086 | Department of Hepatobiliary and Pancreatic Oncology, National Cancer Center Hospital East, Kashiwa, Japan | Sample selection | Sample collection |                                        |
| MASASHI KUDO |                      | MASASHI KUDO | 0000-0003-2161-1807 | Department of Hepatobiliary and Pancreatic Surgery, National Cancer Center Hospital East, Kashiwa, Japan | Sample selection |                                        |                                        |
| NAOTO GOTOHDA |                      | NAOTO GOTOHDA | 0000-0002-4468-5844 | Department of Hepatobiliary and Pancreatic Surgery, National Cancer Center Hospital East, Kashiwa, Japan | Sample selection |                                        |                                        |
| SHUICHI MITSUNAGA |                   | SHUICHI MITSUNAGA | 0000-0001-9496-3278 | Department of Hepatobiliary and Pancreatic Oncology, National Cancer Center Hospital East, Kashiwa, Japan | Sample selection | Supervisor |                                        |
| TAKESHI KUWATA |                      | TAKESHI KUWATA | 0000-0002-4468-5844 | Department of Genetic Medicine and Services, National Cancer Center Hospital East, Kashiwa, Japan | Sample selection |                                        |                                        |
| TAKASHI KOJIMA |                      | TAKASHI KOJIMA | 0000-0002-4468-5844 | Department of Gastrointestinal Oncology, National Cancer Center Hospital East, Kashiwa, Japan | Sample selection |                                        |                                        |
| TATSUO MURANO |                      | TATSUO MURANO | 0000-0003-0423-2151 | Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East, Kashiwa, Japan | Sample selection | Project administration |                                        |
| First author | Corresponding author | Author name | ORCID | Affiliation                                                                 | Contribution 1          | Contribution 2          | Contribution 3          |
|--------------|----------------------|-------------|-------|----------------------------------------------------------------------------|------------------------|------------------------|------------------------|
| Tomonori Yano |                      |             |       | Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East, Kashiwa, Japan | Sample selection         | Project administration |
| Taiki Yamaji |                      |             | 0000-0002-4799-8782 | Division of Epidemiology, National Cancer Center Institute for Cancer Control, National Cancer Center, Tokyo, Japan | Sample selection         | Project administration |
| Takahisa Matsuda |                |             | 0000-0002-9244-2820 | Cancer Screening Center, National Cancer Center Hospital, Tokyo, Japan | Sample selection         | Project administration |
| Shoichiro Tsugane |               |             | 0000-0003-4105-2774 | Division of Cohort Research, National Cancer Center Institute for Cancer Control, National Cancer Center, Tokyo, Japan | Supervisor               |                        |
| Kazuki Hashimoto |                  |             | 0000-0002-8324-2202 | Department of Breast and Endocrine Surgery, National Center for Global Health and Medicine, Tokyo, Japan | Sample selection         | Sample collection       | Project administration |
| Kazuhiko Yamada |                 |             |       | Department of Surgery, National Center for Global Health and Medicine, Tokyo, Japan | Sample collection         | Sample selection         |                        |
| Nobuyuki Takemura |               |             |       | Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, National Center for Global Health and Medicine, Tokyo, Japan | Sample collection         | Sample selection         |                        |
| Kyoji Ito |                      |             |       | Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, National Center for Global Health and Medicine, Tokyo, Japan | Sample collection         | Sample selection         |                        |
| Fuminori Mihara |                    |             |       | Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, National Center for Global Health and Medicine, Tokyo, Japan | Sample collection         | Sample selection         |                        |
| Akihiko Shimomura |                  |             | 0000-0002-2557-8170 | Department of Breast and Medical Oncology, National Center for Global Health and Medicine, Tokyo, Japan | Supervisor               |                        |                        |
| Kunitoshi Shigeyasu |                |             | 0000-0002-1266-2875 | Department of Gastroenterological Surgery, Dentistry and Pharmaceutical Sciences, Okayama University Graduate School of Medicine, Okayama, Japan | Sample selection         | Sample collection       |                        |
| Kazuhiro Noma |                      |             | 0000-0002-5584-3908 | Department of Gastroenterological Surgery, Dentistry and Pharmaceutical Sciences, Okayama University Graduate School of Medicine, Okayama, Japan | Sample selection         | Sample collection       |                        |
| Toshiyoshi Fujiwara |                   |             | 0000-0002-5377-6051 | Department of Gastroenterological Surgery, Dentistry and Pharmaceutical Sciences, Okayama University Graduate School of Medicine, Okayama, Japan | Sample selection         | Sample collection       |                        |
| Author | ORCID ID | Affiliation | Contribution 1 | Contribution 2 | Contribution 3 |
|--------|----------|-------------|----------------|----------------|----------------|
| Hideki Yamamoto | 0000-0003-4919-3483 | Department of Clinical Genomic Medicine, Dentistry and Pharmaceutical Sciences, Okayama University Graduate School of Medicine, Okayama, Japan | Sample selection | Project administration | Supervisor |
| Mizuki Morita | 0000-0001-8592-5499 | Okayama University Hospital Biobank, Okayama University Hospital, Okayama, Japan | Sample collection | Project administration | |
| Shinichi Toyooka | 0000-0002-7588-6745 | Department of General Thoracic Surgery and Breast and Endocrinological Surgery, Dentistry and Pharmaceutical Sciences, Okayama University Graduate School of Medicine, Okayama, Japan | Supervisor | | |
| Akihiro Tamori | 0000-0001-9119-4864 | Department of Hepatology, Graduate School of Medicine, Osaka City University, Osaka, Japan | Sample selection | Project administration | |
| Tasuku Nakabori | 0000-0002-9486-850X | Department of Hepatobiliary and Pancreatic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| Kenji Ikezawa | 0000-0001-9517-8760 | Department of Hepatobiliary and Pancreatic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| Kazuyoshi Ohkawa | 0000-0001-9700-2472 | Department of Thoracic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| Kazumi Nishino | 0000-0001-7517-8670 | Department of Thoracic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| Toru Kuma | 0000-0001-9700-2472 | Department of Thoracic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| Kei Kunimasa | 0000-0001-7517-8670 | Department of Thoracic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| Kazumi Nishino | 0000-0001-7517-8670 | Department of Thoracic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| Toshihiro Kudo | 0000-0001-7517-8670 | Department of Thoracic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| Naotoshi Sugimoto | 0000-0001-7517-8670 | Department of Thoracic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| Masayoshi Yasui | 0000-0001-7517-8670 | Department of Thoracic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| Takeshi Omori | 0000-0001-7517-8670 | Department of Thoracic Oncology, Osaka International Cancer Institute, Osaka, Japan | Sample selection | | |
| First author | Corresponding author | Author name | ORCID | Affiliation                                                                 | Contribution 1            | Contribution 2            | Contribution 3                        |
|--------------|----------------------|-------------|-------|----------------------------------------------------------------------------|---------------------------|---------------------------|----------------------------------------|
| Hiroshi Miyata |                     | Department of Gastroenterological Surgery, Osaka International Cancer Institute, Osaka, Japan | Sample selection         | Sample collection            |
| Toru Kimura  |                     | Department of General Thoracic Surgery, Osaka International Cancer Institute, Osaka, Japan   | Sample selection         | Sample collection            |
| Tomohiro Maniwa |                   | Department of General Thoracic Surgery, Osaka International Cancer Institute, Osaka, Japan   | Sample selection         | Sample collection            |
| Jiro Okami   |                     | Department of General Thoracic Surgery, Osaka International Cancer Institute, Osaka, Japan   | Sample selection         | Sample collection            |
| Hiroki Kusama |                     | Dept. of Breast and Endocrine Surgery, Osaka International Cancer Institute, Osaka, Japan   | Sample selection         | Sample collection            |
| Nobuyoshi Kittaka |               | 0000-0002-0906-1155 | Dept. of Breast and Endocrine Surgery, Osaka International Cancer Institute, Osaka, Japan | Sample selection         | Sample collection            |
| Takahiro Nakayama |               | 0000-0002-3713-1624 | Dept. of Breast and Endocrine Surgery, Osaka International Cancer Institute, Osaka, Japan | Sample selection         | Sample collection            |
| Masashi Nakayama |                | Department of Urology, Osaka International Cancer Institute, Osaka, Japan                | Sample selection         | Sample collection            |
| Yasutomo Nakai |                     | Department of Urology, Osaka International Cancer Institute, Osaka, Japan                | Sample selection         | Sample collection            |
| Kazuo Nishimura |                     | Department of Urology, Osaka International Cancer Institute, Osaka, Japan                | Sample selection         | Sample collection            |
| Shoji Yotsui  |                     | Clinical Laboratory, Osaka International Cancer Institute, Osaka, Japan                | Sample collection         |                             |
| Takashi Yamamoto |                  | Clinical Laboratory, Osaka International Cancer Institute, Osaka, Japan                | Sample collection         |                             |
| Tomoyuki Yamasaki |                | Clinical Laboratory, Osaka International Cancer Institute, Osaka, Japan                | Sample collection         |                             |
| Emi Yamashita |                     | Department of Medical Informatics, Osaka International Cancer Institute, Osaka, Japan | Sample selection         | Project administration       |
| Kazune Saito  |                     | Next-generation Precision Medicine Research Center, Osaka International Cancer Institute, Osaka, Japan | Sample selection         | Sample collection            |
| Keiichi Yoshida |                    | Next-generation Precision Medicine Research Center, Osaka International Cancer Institute, Osaka, Japan | Sample selection         | Sample collection          | Project administration |
| First author | Corresponding author | Author name | ORCID | Affiliation | Contribution 1 | Contribution 2 | Contribution 3 |
|-------------|----------------------|-------------|-------|-------------|----------------|----------------|----------------|
| Masayuki Ohue | | | | Next-generation Precision Medicine Research Center, Osaka International Cancer Institute, Osaka, Japan | Sample selection | Sample collection | |
| Masakazu Kohda | | | | Healthcare Business Department, PFDeNA, Inc., Tokyo, Japan | Data analysis | Project administration | Sample selection |
| Tatsuya Yamaguchi | | | | Healthcare Business Department, PFDeNA, Inc., Tokyo, Japan | Sample collection | Project administration | Data analysis |
| Masami Tanaka | | | | Healthcare Business Department, PFDeNA, Inc., Tokyo, Japan | Sample selection | Data analysis | Project administration |
| Takashi Nishizawa | | | | Healthcare Business Department, PFDeNA, Inc., Tokyo, Japan | Sample selection | Supervisor | Data analysis |
| Tetsuhiko Taira | | | | Sagara Hospital, Department of Medical Oncology, Sagara Hospital, Kagoshima, Japan | Sample selection | Sample collection | |
| Junko Kawano | | | | 0000-0002-4289-1492 | Sagara Hospital, Department of Breast and Thyroid Surgical Oncology, Sagara Hospital, Kagoshima, Japan | Sample selection | Sample collection | |
| Yasuaki Sagara | | | | 0000-0001-5705-1739 | Sagara Hospital, Department of Breast and Thyroid Surgical Oncology, Sagara Hospital, Kagoshima, Japan | Supervisor | | |
| Yosuke Horita | | | | Department of Medical Oncology, International Medical Center, Saitama Medical University, Hidaka, Japan | Sample selection | Sample collection | |
| Yoshiaki Mihara | | | | 0000-0002-3752-0494 | Department of Medical Oncology, International Medical Center, Saitama Medical University, Hidaka, Japan | Sample selection | Sample collection | |
| Tetsuya Hamaguchi | | | | 0000-0002-8353-5371 | Department of Medical Oncology, International Medical Center, Saitama Medical University, Hidaka, Japan | Project administration | | |
| Okihide Suzuki | | | | Department of Digestive Tract and General Surgery, Saitama Medical Center, Saitama Medical University, Kawagoe, Japan | Sample selection | Sample collection | Project administration |
| Yoichi Kumagai | | | | Department of Digestive Tract and General Surgery, Saitama Medical Center, Saitama Medical University, Kawagoe, Japan | Sample selection | Sample collection | |
| Hideyuki Ishida | | | | Department of Digestive Tract and General Surgery, Saitama Medical Center, Saitama Medical University, Kawagoe, Japan | Supervisor | | |
| Motoki Yamagishi | | | | Department of Urology, Showa University Fujigaoka Hospital, Yokohama, Japan | Sample selection | Sample collection | Project administration |
| First author | Corresponding author | Author name |
|--------------|----------------------|-------------|
| Hideaki Shimoyama | | Department of Urology, Showa University Fujigaoka Hospital, Yokohama, Japan |
| Haruaki Sasaki | | Department of Urology, Showa University Fujigaoka Hospital, Yokohama, Japan |
| Takehiko Nakasato | | Department of Urology, Kikuna Memorial Hospital, Yokohama, Japan |
| Takeshi Shichijo | | Department of Urology, Showa university, Tokyo, Japan |
| Takashi Fukagai | | Department of Urology, Showa university, Tokyo, Japan |
| Kota Nishimura | | Department of Urology, Showa University Koto Toyosu Hospital, Tokyo, Japan |
| Kidai Hirayama | | Department of Urology, Showa University Koto Toyosu Hospital, Tokyo, Japan |
| Masashi Morita | | Department of Urology, Showa University Koto Toyosu Hospital, Tokyo, Japan |
| Yuji Kudo | | Department of Thoracic and Thyroid Surgery, Tokyo Medical University, Tokyo, Japan |
| Susumu Takeuchi | | Department of Thoracic and Thyroid Surgery, Tokyo Medical University, Tokyo, Japan |
| Norihiko Ikeda | | Department of Thoracic and Thyroid Surgery, Tokyo Medical University, Tokyo, Japan |
| Naohiro Kamoda | | Department of Urology, Tokyo Medical University, Tokyo, Japan |
| Kazunori Namiki | | Department of Urology, Tokyo Medical University, Tokyo, Japan |
| Yoshio Ohno | | Department of Urology, Tokyo Medical University, Tokyo, Japan |
| Tomohiro Umez | | Department of Molecular Pathology, Tokyo Medical University, Tokyo, Japan |
| Yoshiki Murakami | | Department of Molecular Pathology, Tokyo Medical University, Tokyo, Japan |
| Masahiko Kuroda | | Department of Molecular Pathology, Tokyo Medical University, Tokyo, Japan |
| Co-corresponding | Motoki Abe | Preferred Networks, Inc., Tokyo, Japan |
| Co-corresponding | Yusuke Yamamoto | Laboratory of Integrative Oncology, National Cancer Center Research Institute, Tokyo, Japan |