



OPEN

## Reproducible and sensitive micro-tissue RNA sequencing from formalin-fixed paraffin-embedded tissues for spatial gene expression analysis

Hiroko Matsunaga<sup>1</sup>, Koji Arikawa<sup>1</sup>, Miki Yamazaki<sup>2,3</sup>, Ryota Wagatsuma<sup>2,3</sup>, Keigo Ide<sup>2,3</sup>, Ashok Zachariah Samuel<sup>1</sup>, Kazuya Takamochi<sup>4</sup>, Kenji Suzuki<sup>4</sup>, Takuo Hayashi<sup>5</sup>, Masahito Hosokawa<sup>1,2,3,6</sup>, Hideki Kambara<sup>1,7</sup> & Haruko Takeyama<sup>1,2,3,6</sup>✉

Spatial transcriptome analysis of formalin-fixed paraffin-embedded (FFPE) tissues using RNA-sequencing (RNA-seq) provides interactive information on morphology and gene expression, which is useful for clinical applications. However, despite the advantages of long-term storage at room temperature, FFPE tissues may be severely damaged by methylene crosslinking and provide less gene information than fresh-frozen tissues. In this study, we proposed a sensitive FFPE micro-tissue RNA-seq method that combines the punching of tissue sections (diameter: 100  $\mu\text{m}$ ) and the direct construction of RNA-seq libraries. We evaluated a method using mouse liver tissues at two years after fixation and embedding and detected approximately 7000 genes in micro-punched tissue-spots (thickness: 10  $\mu\text{m}$ ), similar to that detected with purified total RNA (2.5 ng) equivalent to the several dozen cells in the spot. We applied this method to clinical FFPE specimens of lung cancer that had been fixed and embedded 6 years prior, and found that it was possible to determine characteristic gene expression in the microenvironment containing tumor and non-tumor cells of different morphologies. This result indicates that spatial gene expression analysis of the tumor microenvironment is feasible using FFPE tissue sections stored for extensive periods in medical facilities.

Integrated analysis of cell spatial location and gene expression information using histological tissue sections is a powerful tool for determining tissue function, particularly pathological changes, under diseased conditions<sup>1–7</sup>. Initially, hybridization-based methods, such as FISH or MERSFISH, were developed for the spatial analysis of genes expressed in vivo using fluorescent probes prepared in advance<sup>1–3</sup>. However, issues such as limited adaptation to only known genes and limited simultaneous detection of multiple genes were reported. With the rapid progress of next-generation sequencing technology, spatial transcriptome technology has recently garnered attention for comprehensive sequence-based analysis<sup>4,6,7</sup>. All tissue functions involve complex cell–cell interactions<sup>8–10</sup>. Examining the gene expression profiles of morphologically specific tissue domains could help determine the details of cellular interactions in such microenvironments. Fresh-frozen (FF) tissue-based sampling is widely used for mainstream gene expression analyses<sup>1–4,6,7</sup>, owing to the relative ease of extraction of high-quality RNA. However, it is difficult to preserve FF tissue samples for long periods, and the method does not allow detailed morphological observation.

Morphological characteristic-based tissue diagnostic approaches are valuable in the clinical field<sup>11–13</sup>. Cellular structures and tissue morphologies are generally well-preserved in formalin-fixed paraffin-embedded (FFPE)

<sup>1</sup>Research Organization for Nano and Life Innovation, Waseda University, Tokyo, Japan. <sup>2</sup>Department of Life Science and Medical Bioscience, Waseda University, Tokyo, Japan. <sup>3</sup>Computational Bio Big-Data Open Innovation Laboratory, AIST-Waseda University, Tokyo, Japan. <sup>4</sup>Department of Thoracic Surgery, Juntendo University School of Medicine, Tokyo, Japan. <sup>5</sup>Department of Human Pathology, Graduate School of Medicine, Juntendo University, Tokyo, Japan. <sup>6</sup>Institute for Advanced Research of Biosystem Dynamics, Waseda Research Institute for Science and Engineering, Waseda University, Tokyo, Japan. <sup>7</sup>Frontier Biosystems, Inc., Tokyo, Japan. ✉email: haruko-takeyama@waseda.jp

tissues, and hence, they are preferred over FF tissues for sample preservation in pathological diagnosis. In addition, the potential for long-term storage at room temperature is an important advantage of FFPE samples. Archived FFPE tissue samples often serve as invaluable resources for diagnosing disease pathogenesis. However, the crosslinking of proteins and nucleic acids may occur during formalin fixation and cause RNA fragmentation<sup>14,15</sup>. This in turn affects the quality of FFPE-derived RNA; hence, whole-transcriptome analysis of FFPE tissues using standard protocols is challenging. Recent studies have examined the feasibility of transcriptome analysis using purified total RNA from FFPE tissues<sup>16–18</sup> and also compared the gene expression in FF and FFPE tissues<sup>19–23</sup>. However, obtaining gene expression information using methylene-crosslinked RNA derived from FFPE tissues was challenging, and each method required a pre-treatment process involving extraction and purification from one or more tissue sections as the starting material with the commercially available kits (e.g. Nucleo spin Total RNA FFPE XS kit (Takara, Shiga, Japan) requires 1–3 thin sections of 10  $\mu\text{m}$  thickness, and PureLink™ FFPE RNA Isolation Kit (Thermo Fisher Scientific, Waltham, MA, USA) requires 3–8 thin sections of 10  $\mu\text{m}$  thickness).

Spatial transcription profiling is a powerful tool for evaluating the functional characteristics of cells in a typical tissue architecture. It is required to focus on minute, specific regions for understanding the interaction between tissue morphology and corresponding gene expression. Obtaining gene expression information with minimum loss of RNA molecules from a very small tissue section is critical for the success of spatial transcriptomic analysis. In a previous study, we targeted very small (< 100  $\mu\text{m}$ ) micro-tissues containing several dozen cells for constructing RNA-sequencing (RNA-seq) libraries<sup>24,25</sup>. Typical kit-based RNA extraction protocols are not suitable for such small samples, as a certain quantity of RNA molecules are potentially lost during nucleic acid purification, which may involve the use of columns. It is desirable to develop a protocol that is compatible for use with small tissue sections, minimizes sample loss, and allows processes from RNA extraction to cDNA construction to be performed in an uncomplicated manner. Previously, we had reported the development of a high-speed apparatus for the punching and collection of micro-punched tissue spots (PuTi-spots) (diameter: 100  $\mu\text{m}$ ) from FF tissue slices (thickness: 10–20  $\mu\text{m}$ ) while observing the morphological characteristics under a microscope. Using this punching device, PuTi-spots could be collected from both FF and FFPE tissue sections.

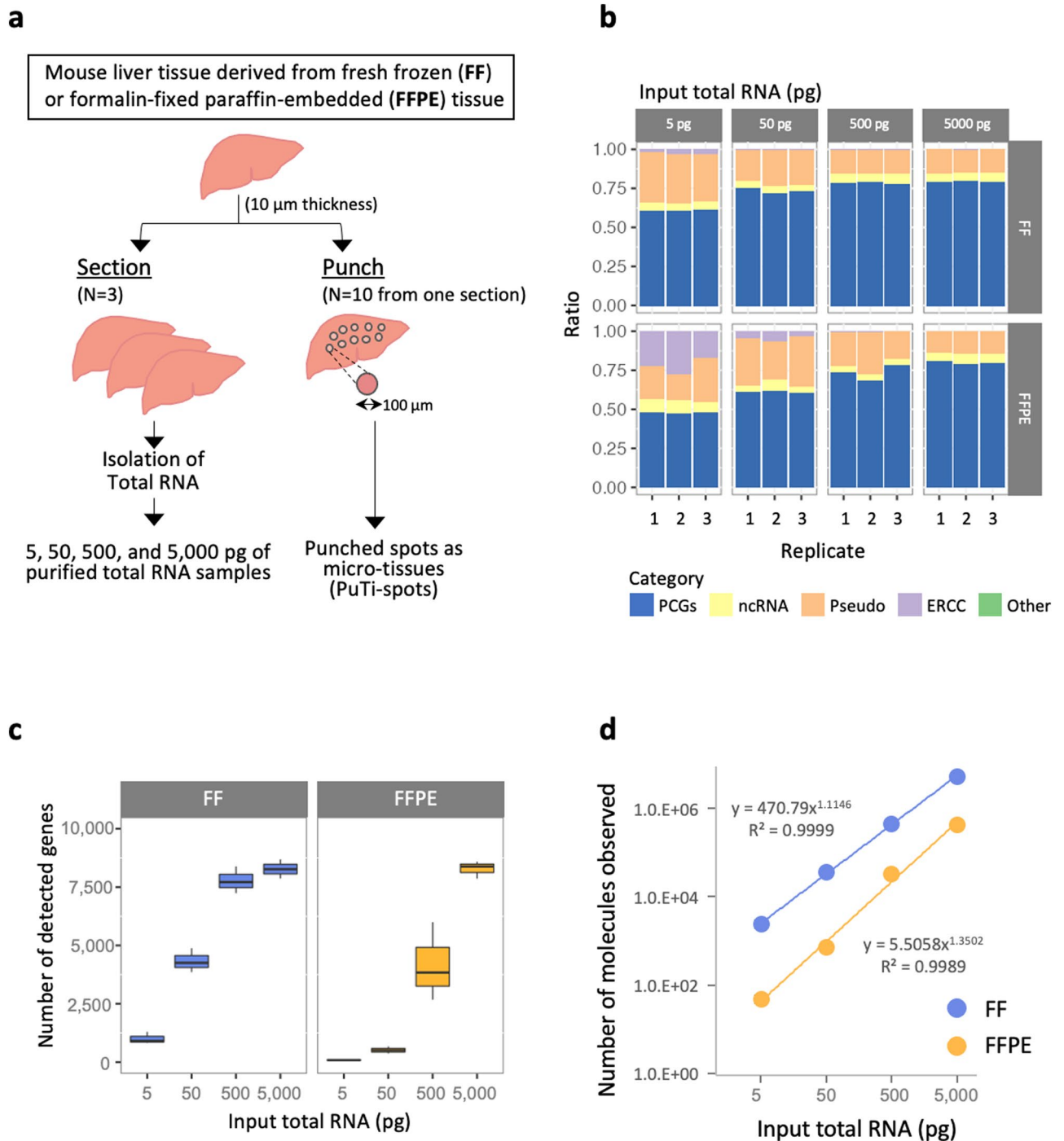
In this study, we report gene expression analysis using very small FFPE tissue sections, a process that has not been conducted with much success to date. We developed an upgraded RNA-seq library preparation protocol for FFPE PuTi-spots using a surfactant, proteinase K, and heating for membrane dissolution and decrosslinking, based on a previously reported method<sup>25</sup>. We critically compared the detection sensitivity and accuracy of RNA-seq achieved using this method with those achieved using serially diluted purified total RNA as the starting material. Further, we performed a thorough comparison of the RNA-seq results obtained using FFPE and FF tissues. Finally, RNA-seq analysis of the PuTi-spots was performed using FFPE clinical specimens to evaluate the detection of genes in low-quality tissue samples with degraded of nucleic acids after long-term archiving at room temperature for more than 6 years.

## Results

**Sample summary.** An overview of the samples used for the assessments in this study is provided in Fig. 1a. Frozen mouse liver tissue stored at  $-80\text{ }^{\circ}\text{C}$  for 2 years after embedding was used as the FF sample. Similarly, mouse liver tissue fixed in 4% formalin for 24 h, embedded in paraffin, and stored at  $4\text{ }^{\circ}\text{C}$  for 2 years was used as the FFPE sample. In both cases, purified total RNA was analyzed using thin tissue sections (thickness: 10  $\mu\text{m}$ ) with commercially available RNA extraction kits. Total RNA quality was assessed based on the RNA integrity number (RIN)<sup>26,27</sup> estimated by electropherograms using Tape Station (Agilent Technologies, Santa Clara, CA, USA) (Table 1). For the FF sample, 18S and 28S rRNA peaks could be clearly identified, and the average RIN value was estimated to be 8.4 (N = 3) (Supplementary Fig. 1). Conversely, the FFPE sample showed no observable rRNA peaks, and the average RIN value was 3.5 (N = 3), which was significantly lower than that of the FF sample. To test the possibility of RNA degradation, the DV200 values (percentage of RNA fragments with > 200 nucleotides) were calculated based on the results of electropherograms<sup>28</sup>. The DV200 value for the FF sample was 98.3%, indicating almost no degradation. The DV200 value for the FFPE sample was 86.3%, which was relatively lower.

Although there are only few changes in gene expression at different sites in the liver<sup>29</sup>, three adjacent tissue slices (thickness: 10  $\mu\text{m}$ ) were used as replicas for total RNA extraction to minimize the variations in gene expression between samples owing to spatial heterogeneity (N = 3). For the same reason, 10 adjacent PuTi-spots ( $\phi = 100\text{ }\mu\text{m}$ ) collected from the one identical section with the punching device<sup>24</sup> were used as replicas for PuTi-spots evaluation (N = 10).

**Evaluation of RNA-seq profiles using purified total RNA from FF and FFPE tissue specimens.** First, we performed RNA-seq of FFPE and FF tissue samples, beginning with different initial purified total RNA contents, to assess the potential impact of RNA degradation on the results of gene expression analysis. Four purified total RNA samples with 5, 50, 500, and 5000 pg (= 5 ng) of total input RNA per reaction were prepared by serial dilution. The smallest input total RNA content of 5 pg per reaction was selected because the total RNA content per cell was expected to range from 5 to 50 pg based on our experiences and as reported in a previous study<sup>30</sup>. PuTi spots are assumed to contain 10–50 cells based on the diameter and thickness of the spot and the nuclear staining pattern. Therefore, a series of total RNA reaction mixtures with RNA content of up to 5000 pg was used to match the RNA content in the PuTi-spot. Three replicates were used for each case (N = 3). The average number of sequence-reads per sample was 0.6 million, satisfying the number of reads required to detect expressed genes<sup>31</sup>. The proportion of reads assigned to protein-coding genes (PCGs) in FF samples plateaued at 80% of the total reads (Fig. 1b). In FFPE, the ratio of PCGs reached 80% at 5000 pg of input. Compared



**Figure 1.** Evaluation of RNA-seq profiling of purified total RNA derived from FFPE and FF tissues. **(a)** Overview of sample preparation for evaluation using mouse liver tissue. Purified total RNA samples were prepared using the RNeasy Mini Kit (QIAGEN) for FF tissues or RNeasy FFPE Kit (QIAGEN) for FFPE tissues. Three adjacent tissue slices (thickness: 10  $\mu\text{m}$ ) were used as section-replicates (N=3) and 10 PuTi-spots ( $\phi = 100 \mu\text{m}$ ) collected from a single tissue section were used as punch-replicates (N=10). **(b)** Mapping category of reads. We evaluated whether reads were assigned to five categories: PCGs, ncRNAs, pseudo genes, ERCC of internal standard RNAs, or others. As the input increased, the ratio assigned to the PCGs increased, and above a certain quantity, the values were saturated. It is also apparent that the assignment of reads to ERCC increases when PCGs have fewer reads, starting at 5 or 50 pg. **(c)** Number of detected genes. The number of genes detected for respective input RNAs was plotted. Libraries were prepared using total RNA purified from FF or FFPE tissue sections (N=3) as the starting material. PCGs were identified from the respective libraries. **(d)** Number of RNA molecules from total RNA contributing to the reaction. We estimated the number of RNA molecules that contribute to the reaction with reference to ERCC, which was used as an internal standard, that is, the value of  $(\text{sum of TPMs of the RNA}) \times (\text{sum of the number of molecules of the input ERCC}) / (\text{sum of TPMs of the ERCC}) = (\text{sum of the number of input RNA molecules})$  was plotted. FFPE formalin-fixed paraffin-embedded, FF fresh frozen, RNA-seq RNA-sequencing, PCGs protein-coding genes, ncRNAs non-coding RNAs, TPM transcripts per kilobase million.

	Preparation	N	RNA quality	
			RIN	Dv200
Fresh frozen	2017_09	3	8.4	98.3%
FFPE	2017_07	3	3.5	86.3%

**Table 1.** Sample quality assessments. *RIN* RNA integrity number, *Dv200* percentage of RNA fragments with more than 200 nucleotides.

to that in FF samples, the value in FFPE samples increased to 80%, with an order of magnitude higher amount of inputs (5000 pg). The number of genes detected in the FF sample also plateaued at an input of 500 pg, whereas the genes detected in the FFPE sample reached a similar number at an input of 5000 pg, which was at a higher order of magnitude (Fig. 1c). This indicates that between FF- and FFPE-derived samples, owing to RNA degradation, the actual number of RNA molecules in the mixture participating in the reaction could be different from the total RNA input.

We then estimated the number of RNA molecules that contributed to cDNA production. The calculations were performed using ERCC RNA Spike-Ins (Thermo Fisher Scientific, Waltham, MA, USA), which is a commercially available external RNA control, and calculated values of transcripts per kilobase million (TPM). According to our estimate, we assumed that the ratio of the total number of ERCC molecules used in the reaction to the total TPM of ERCC (obtained from the RNA-seq results) is equal to the ratio of the total number of RNA molecules contributing to cDNA production to the total TPM of RNA. The number of molecules presumed to have contributed to the reaction increased by an order of magnitude relative to the abundance ratio of the dilution series of total RNA from 5 pg to 5000 pg used for the reaction (Fig. 1d). This was anticipated, implying that the number of molecules contributing to the reaction increases with the total RNA used in the reaction. Conversely, the estimated number of molecules contributing to the reaction in FF and FFPE tissues differed by almost an order of magnitude (Fig. 1d). This implies that the reads assigned to PCGs plateaued at a value less than that in the FF samples by an order of magnitude because the number of RNA molecules contributing to the reaction in the FF-derived purified total RNA was greater by one order of magnitude. This result suggests that the 3'-end polyA sequence of the mRNA was not sufficiently captured by the poly-T primer used in the reverse transcription reaction owing to progressive degradation or incomplete decrosslinking of the purified total RNA molecules from FFPE tissues. The dissociation of methylene crosslinking was challenging, even for tissues that had been stored under relatively mild conditions (4 °C). In the commercial kit used in this study, decrosslinking was performed by heat treatment. Although heat treatment is known to cause the effective dissociation of methylene crosslinking, it may also induce RNA degradation. Even when the total RNA derived from FFPE tissue was purified by the optimization protocol recommended in the kit, the reaction efficiency was lower than that achieved with the total RNA derived from FF tissue, corresponding to a reduction by an order of magnitude in the input volume.

**RNA-seq analysis of PuTi-spots from mouse liver specimens.** RNA-seq analyses of mouse liver PuTi-spots were performed for validating the sequence quality of the FFPE samples using our proposed method. The micro-punched circular sections were estimated to contain 10 to 50 cells per spot, as determined by the counting of hematoxylin–eosin (HE)-stained nuclei. An overview of the workflow from PuTi-spot collection to RNA-seq library preparation is shown in Fig. 2a. After deparaffinization of the FFPE tissue slices, the regions of interest in the tissue were selected based on morphology and micro-punched using our sampling device<sup>24,25</sup>. FF tissue was also used for evaluation after 2 years of storage at –80 °C embedded in super cryoembedding medium (SCEM) solution (Leica Microsystems, Tokyo, Japan). Before micro-punching, the FF tissue sections were lightly fixed in 99.5% ethanol immediately after sectioning to inhibit RNA degradation<sup>25</sup>. PuTi-spots were recovered directly into microtubes from both FFPE and FF samples and were then subjected to cell membrane disruption by proteinase K treatment and mRNA purification using polyT magnetic beads (PPP, purification by proteinase K and polyT magnetic beads)<sup>25</sup>. For FFPE tissues, heat treatment was performed prior to magnetic bead treatment to induce the dissociation of methylene crosslinking. As previously reported, we confirmed that the PPP process is highly effective at improving the detection of PCGs in the cDNA synthesis of FF PuTi-spots<sup>25</sup>. In FFPE PuTi-spots, the number of genes detected without PPP treatment was 2055 ± 269 (mean ± standard deviation, N = 4), whereas the number of genes detected with PPP involving permeabilization was 5335 ± 325 (mean ± standard deviation, N = 4) (Fig. 2b) (PuTi-spot N = 4 was used as a replicate for evaluating reaction conditions). The number of detected genes increased to 7066 ± 733 (mean ± standard deviation) with the addition of heat treatment to the PPP for the dissociation of methylene crosslinking. We used ten replicates (N = 10) for this condition, which is N = 4 in addition to six PuTi-spots, using the conditions determined by the above study. Since the purpose of this study is to evaluate the quality of RNA-seq in FFPE tissue compared to FF tissue, we increased the number of replicates in the latter.

The average number of genes detected in one PuTi-spot was 8688 for FF and 7066 for FFPE samples. This corresponded to an input of 500 to 5000 pg of total RNA in purified total RNA analysis (Fig. 2c). Since the total RNA content per cell is approximately 50 pg, 50 cells would be roughly equivalent to a starting material of 2500 pg of total RNA. The number of molecules contributing to the reaction was also calculated as previously performed for the purified total RNA sample. The estimated number of molecules contributing to the reaction was 10<sup>6</sup> for FF and 10<sup>5</sup> for FFPE samples, equivalent to a total RNA input of 2500 pg (Fig. 2d). In summary, we confirmed

that using the proposed method, RNA present in the PuTi-spot can be recovered for RNA-seq analysis with efficiency comparable to that achieved with purified total RNA extraction using a commercially available kit.

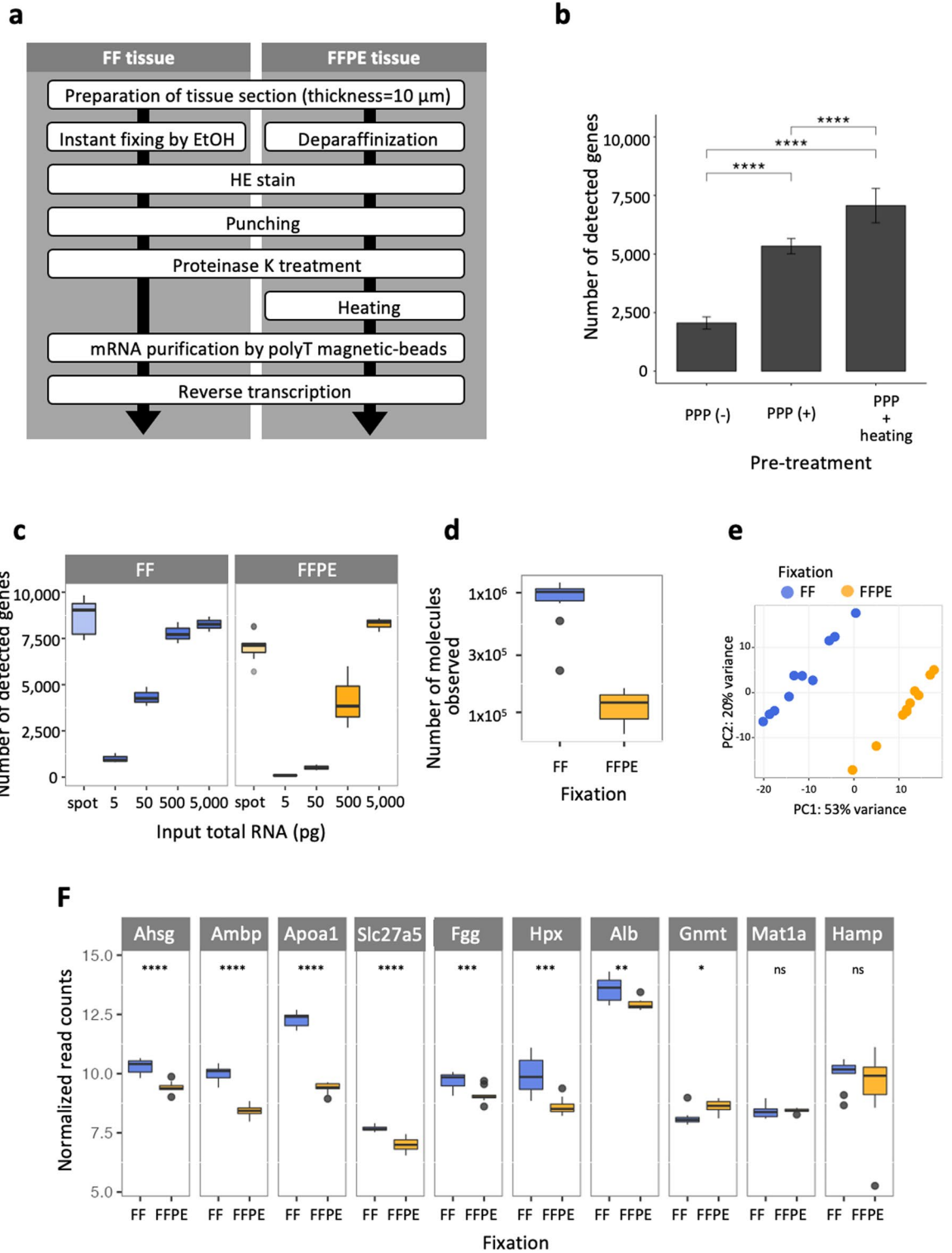
**Evaluation of gene detection sensitivity in PuTi-spots of mouse liver specimens.** The gene expression profile of the PuTi-spots was evaluated. Ten PuTi-spots were collected from a 10  $\mu\text{m}$ -thick tissue section of mouse liver and used for evaluation as ten replicates. The proportion of gene expression levels and detection frequency for both FF and FFPE samples were comparable in all ten replicates (Supplementary Fig. 2). In the principal component analysis (PCA) of gene expression, two different clusters corresponding to FF and FFPE samples were observed (Fig. 2e). This suggests that the differences in the detected gene pattern depend on the tissue preservation condition. We compared the expression levels of ten genes expressed specifically in the mouse liver<sup>32</sup> in FF or FFPE tissue spots. The normalized read counts showed that the *Gnmt*, *Mat1a*, and *Hamp* genes showed almost no difference in expression between FF and FFPE samples, whereas the seven other genes showed significant differences in expression based on the fixation conditions (Fig. 2f). The differences in the normalized read counts for each gene between PuTi-spots and purified total RNA were not substantial, suggesting that the differences indicated in Fig. 2f did not result from differences in the methods used for cDNA library preparation, including RNA purification. Therefore, we aligned the sequencing reads to each gene using Integrative Genomics Viewer<sup>33–35</sup>. The results showed that FFPE samples used for this evaluation showed a significant decrease in reads with alignment at approximately 500 bases from the 3'-end of each gene, regardless of whether the starting material was purified total RNA or a PuTi-spot (Supplementary Fig. 3). This suggests that cDNA obtained from FFPE samples is fragmented to approximately 500 bases or less owing to nucleic acid degradation by methylene crosslinking. Thus, the number of reads in FFPE-derived samples tends to decrease compared to that in the FF-derived sample because the reads are biased toward the 3'-end. Since the *Hamp* gene was only approximately 400 bases, the reads mapped to the entire gene region without bias, which resulted in no difference between the normalized read counts in FF and FFPE samples. In contrast, the *Mat1a* gene spans approximately 3.5 kb, which is greater than the average length of the genes assessed in this study (approximately 1.6 kb). In both FF and FFPE samples, there was a depression in read alignment in the middle of the last exon at the 3'-end (the last exon was approximately 1.9 kb in length), possibly because the complete sequence was not obtained, similar to that obtained for the cDNA template, even from FF samples. The results of RNA-seq using PuTi-spots suggested that the differences in gene expression profiling could be attributed to the method of tissue preservation rather than the method of library preparation, based on the starting materials. The bias in gene expression profiling between FF- and FFPE-derived samples could be eliminated by, for example, focusing only on the 3'-end.

**RNA-seq analysis of PuTi-spots from tumor and non-tumor areas in resected pathological specimens of lung cancer.** We applied the methods discussed above to a pathological specimen to evaluate their potential for clinical application. FFPE tissues derived from human lung cancer were used as the specimen, which had been embedded 6 years ago and were stored at room temperature. As in the previous analyses, total RNA was extracted from the thin sections (10  $\mu\text{m}$ -thick) using a commercially available kit and used for RNA quality evaluation. In addition, PuTi-spots were prepared from sections adjacent to those used for RNA assessment. Punching was performed in each area, distinguishing between tumor and non-tumor areas, based on the discretion of the pathologist (Fig. 3a). The RIN value was 1.8 and the DV200 value was 52%, which were indicative of progressive nucleic acid degradation by crosslinking.

RNA-seq was performed using the PuTi-spots collected from the tumor and non-tumor areas (each  $N = 9$ ) of the specimen (Fig. 3b). The number of detected genes was greater in the tumor area than in the non-tumor area, with an average of 4400 genes in the tumor area and 1800 genes in the non-tumor area (Fig. 3c). PCA showed that the gene expression pattern could be clearly categorized into two groups, one specific to the tumor area and the other to the non-tumor area (Fig. 3d). Conversely, the variation among PuTi-spots was higher in the tumor area than in the non-tumor area. In this specimen, cadherin binding was extracted by enriched gene ontology (GO) analysis from among the genes upregulated in the tumor area (Supplementary Fig. 4). Among the 44 genes with GO assigned to cadherin binding, *S100P* and *TAGLN2* were found to exhibit upregulation (Supplementary Table 1). *S100P* is a  $\text{Ca}^{2+}$ -binding protein that is overexpressed in various cancers and considered a tumor biomarker<sup>36</sup>. *TAGLN2* is a gene that has been suggested to be associated with cancer development, and suppressing this gene is considered to inhibit the growth, invasion, and metastasis of cancer cell<sup>37</sup>. Based on these results, it was indicated that even when PuTi-spots containing only a few dozen cells were collected from low-quality tissues with a DV200 value of 50%, a sufficient number of genes could be detected to understand the characteristics of cells with different morphologies, such as tumor and non-tumor cells, and to capture the different expression patterns functioning in the microenvironment.

## Discussion

We have demonstrated that direct RNA-seq from PuTi-spots (tissue spots with a diameter of 100  $\mu\text{m}$ ) derived from FFPE tissue stored for over 2 years can be used for gene expression analysis, with sensitivity comparable to that achieved with total RNA extracted using a commercially available kit. The key points of the featured experimental protocol are as follows. We demonstrated efficient membrane lysis and methylation crosslink dissociation for RNA extraction from the cells. For the FF tissue spots, membrane lysis was achieved using a surfactant and proteinase K, as reported previously<sup>25</sup>. For the FFPE tissue spots, we strengthened the process by using a surfactant, proteinase K, and heat treatment. Treatment with proteinase and heat was essential for the removal of proteins crosslinked with nucleic acids in the FFPE tissue. The library preparation protocol presented in this study can help minimize RNA loss by eliminating the need for RNA purification by column extraction. Consequently, we demonstrated that gene expression analysis can be performed with equal accuracy and sensitivity



◀**Figure 2.** Evaluation of RNA-seq profiling data derived from FFPE and FF PuTi-spots. **(a)** Sample preparation flow of cDNA library derived from FF or FFPE PuTi-spots as starting materials. The difference between the handling of FF and FFPE tissues is related to the appropriate pre-treatment process for each fixation method before the punching of a micro-region and the additional heat treatment step for decrosslinking in the FFPE tissue. **(b)** Pre-treatment assessment for dissociation of crosslinking in FFPE PuTi-spots. The treatment without PPP is indicated with PPP(-) (N = 4), with PPP is indicated with PPP(+) (N = 4), and the treatment with PPP and heating is indicated with PPP and heating (N = 10). **(c)** Comparison of detected gene number. The quantity of total RNA was assessed, which was equivalent to the number of genes detected in the micro-region samples. The number of genes detected from the micro-regions roughly corresponded to 500–5000 pg of total RNA. **(d)** As with the purified total RNA samples (Fig. 1d), we estimated the quantity of RNA contributing to the reaction based on the molecular weight of ERCC used as an internal standard. **(e)** PCA of gene expression data. PCA was performed using gene expression data from FF or FFPE-derived PuTi-spots. Two different clusters were identified by PCA. The variance in percentages accounted for by each principal component is shown on each axis. **(f)** Expression levels of liver-specific genes observed using FF and FFPE PuTi-spots. We focused on 10 genes specifically expressed in the mouse liver and compared the normalized read counts between FF and FFPE PuTi-spots used as starting materials (N = 10, respectively). Statistical analysis based on Welch's *t*-test. *P* value < 0.05 (\*), < 0.01 (\*\*), < 0.001 (\*\*\*), < 0.0001 (\*\*\*\*). *P* value > 0.05 is indicated by "ns". The *p* values were calculated and visualized using the ggplot2[40] (version 3.3.0) and ggpubr[39] (version 0.4.0) packages in R. *FFPE* formalin-fixed paraffin-embedded, *FF* fresh frozen, *PuTi-spot* micro-punched tissue spot, *RNA-seq* RNA-sequencing, *PCA* principal component analysis.

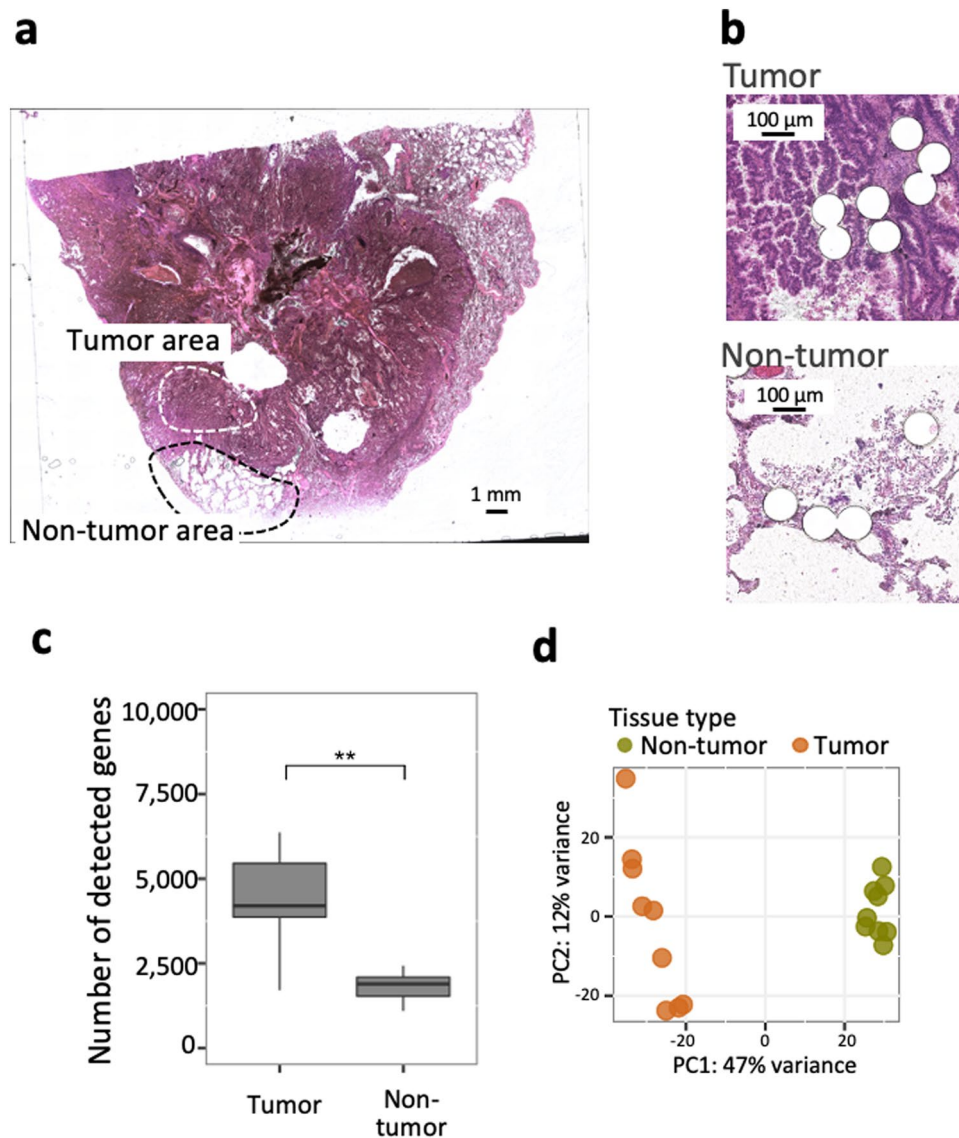
using mouse liver samples and purified total RNA, even in very small tissue sections with a diameter of 100 μm, which were collected from specific histomorphological locations.

A previous report on intrinsic gene expression profiling bias unique to FFPE tissues had evaluated thousands of FFPE-derived tissue samples for comparison with FF tissues<sup>17</sup>. The report addressed that the gene expression quantification of FFPE-derived RNA was robust, consistent, and reproducible but had FFPE-specific biases. This suggests that there is a potential for a similar intrinsic bias in the gene expression of FFPE-derived PuTi-spots in our study. We speculate that this intrinsic bias leads to the differences in gene expression profiling due to the method of tissue preservation, FFPE and FF, also observed in our study. The differences in gene expression profiling between FFPE and FF tissues have been examined previously in several studies<sup>18–23</sup>, all of which similarly stated that there were differences specific to the tissue preservation method. Meanwhile, in previous analyses, the starting material was a thin section, with an area of approximately 5 mm<sup>2</sup> or greater, cut from an FFPE tissue block, from which total RNA (in the order of nanogram to micrograms) was extracted and prepared. To date, it has been almost impossible to perform gene expression analysis using very small FFPE tissue sections (diameter: 100 μm) as starting material. In this study, we have reported gene expression analysis using very small FFPE tissue sections (diameter: 100 μm) as starting material, which has not been conducted with significant success to date.

The appropriate thickness and diameter of PuTi-spots are expected to vary depending on the purpose of the study. Thickness of the sections was set to 10 μm because the cells did not overlap each other in the thickness direction and appeared clean in the morphological observation using HE staining. The decrease in mRNA levels is expected to lower the number of detected genes, but this may be because the cells were fragmented during sectioning. And this suggests that the cells were broken up during thinning and did not remain in the PuTi-spot due to the thinner section thickness in the first place. It is expected that the number of detected genes obtained from FFPE tissue sections will decrease if the sections are stored for a longer period of time than that reported in this report and if the storage is under room temperature environment. Some reports have described a significant impact of the preservation status of tissue blocks on the number of detected genes when evaluated using FFPE tissues preserved for more than 20 years<sup>16,18</sup>. In these reports, evaluation was performed on tissues with DV200 of 1–50%, indicating that the degradation of nucleic acids was greatly advanced. Based on the results of our report, it can be concluded that the decrease in the number of detected genes is not due to the fact that the tissue pieces are smaller, from a single thin slice to a PuTi-spot but is largely due to the quality of the tissue itself, including the procedure and environment for sectioning, fixing and storage. We will continue to establish the robustness of the method by applying the analysis of PuTi-spots to a variety of tissues with longer storage periods.

Lastly, the application of this method to clinical specimens showed that it is possible to detect genes even in tissue samples stored at room temperature for 6 years after embedding. Although it is expected that the gene groups identified differ based on the tissue sections used and the microenvironmental conditions, we believe the reliability of the detection results in this report to be considerably high, because the accuracy and sensitivity were evaluated carefully in a model case using mice and applied to clinical specimens. Cancer tissue does not only contain sections that can be clearly distinguished into tumor and non-tumor cells. It is also assumed that gene expression may differ between non-tumor areas adjacent to and distant from the tumor site.

By applying this method to the vast number of pathological specimens stored long-term at medical institutions, gene expression can be comprehensively analyzed from specimens with a clear pathological condition and prognostic course. This will lead to the accumulation of new knowledge for determining drug susceptibility and predicting prognosis. Furthermore, the ability to obtain a large amount of gene expression information from a very small amount of tissue fragments means that it will be possible to obtain an unprecedentedly large amount of genetic information without changing the amount of tissue collected from patients, which is expected to contribute to the clarification of etiology. We hope that our proposed method will find applications in pathology, drug discovery evaluation, and associated fields of medicine, in which transcriptome analysis of tissue preserved for long time periods is often necessary.



**Figure 3.** Evaluation of RNA-seq profiling of microenvironments in lung tumor specimens using PuTi-spots. **(a)** Morphological image of tumor and non-tumor areas in lung tumor specimens. Thin sliced sections (thickness: 10 μm) were obtained from a surgical resected sample that had been archived at room temperature for 6 years. After deparaffinization and HE staining, PuTi-spots were punched and collected under microscopic observation. **(b)** Image after PuTi-spot collection. Each white circle (diameter: 100 μm) corresponds to the location at which the PuTi-spot was punched out. **(c)** Comparison of the detected gene numbers. The quantity of total RNA samples was assessed, which was equivalent to the number of genes detected in the micro-region samples. The number of genes detected from the micro-regions roughly corresponded to 500–5000 pg of total RNA. Statistical analysis based on Welch's *t*-test. The *p* values were calculated and visualized using the ggplot2<sup>40</sup> (version 3.3.0) and ggpubr<sup>39</sup> (version 0.4.0) packages in R. \*\**P* value < 0.01. **(d)** PCA of gene expression data. PCA was performed using gene expression data from tumor or non-tumor-derived PuTi-spots (*N* = 10, respectively). Two different clusters were identified, and the distribution of the plots showed different trends between the groups. The amount of variance in percentages accounted for by each principal component is shown on each axis. *RNA-seq* RNA-sequencing, *PuTi-spot* micro-punched tissue spot, *HE* hematoxylin–eosin, *PCA* principal component analysis.

## Methods

**Ethics declarations.** All mice experiments were completed in compliance with the ARRIVE guidelines. All mice experiments were treated according to protocols approved by the Waseda University Animal Experimentation Committee (Approval no. 2017-A056, 2018-A067).

Human tissue specimens were collected and analyzed under a protocol approved by the institutional review boards of Juntendo University (No.2018090) and Waseda University (No.2017-G001). Informed consent was



obtained from the participants or their legal representatives. The study conformed to the Helsinki declaration (1964) and its amendments or comparable ethical standards.

**Tissue sources.** All mice used in this study (ICR, male, age > 2 months, Tokyo Laboratory Animals Science Co. Ltd., Tokyo, Japan) were treated according to protocols approved by the Waseda University Animal Experimentation Committee (Approval no. 2017-A056, 2018-A067). Mice were euthanized by inhalation of 2% isoflurane, and then dissected to collect tissue samples.

Clinical sample of lung cancer was acquired from surgically resected tumors at the Department of Human Pathology, Juntendo University School of Medicine. Immediately after acquisition, the tissues were fixed in 10% neutral-buffered formalin for 24 h at room temperature, embedded in paraffin after routine processing. The type of the lung cancer considered was papillary adenocarcinoma. Pathological diagnosis was based on the 2021 World Health Organization classification<sup>38</sup>. Tumor tissue specimens were collected and analyzed under a protocol approved by the institutional review boards of Juntendo University (No. 2018090) and Waseda University (No. 2017-G001). Informed consent was obtained from the participants or their legal representatives.

**Preparation of FF tissue sections.** A liver tissue resected from the mouse was immediately washed with cooled phosphate-buffered saline (PBS, pH 7.4, Thermo Fisher Scientific), immersed in SCEM solution (Leica Microsystems), and rapidly frozen in liquid nitrogen. The embedded frozen tissues were stored at  $-80^{\circ}\text{C}$  until use in experiments. For total RNA extraction (from bulk samples), thin sections (thickness: 20  $\mu\text{m}$ ) were used. For micro-region sampling, 20  $\mu\text{m}$ -thick sections were transferred to cryofilm (SECTION-LAB, Hiroshima, Japan) and immersed and fixed in 99.5% ethanol for 10 s.

**Preparation of FFPE tissue sections.** A liver tissue resected from the mouse was immediately washed with cooled PBS, immersed in 4% paraformaldehyde with PBS (Nacalai Tesque, Inc., Kyoto, Japan), and fixed at  $25^{\circ}\text{C}$  for 24 h. They were then immersed in 0.1 mol/L-PBS (pH 7.4, Nacalai Tesque, Inc.). After fixation, the tissues were dehydrated by soaking in 70%, 80%, 90%, and 95% ethanol solution for at least 30 min, followed by soaking in 99.5% ethanol solution for at least 30 min; the step was repeated four times, and the solution was changed each time. Hemo-Clear solution (FALMA, Tokyo, Japan) was used as an alternative to xylene for paraffin embedding. Paraplast (Merck, Darmstadt, Germany), a low-melting-point paraffin, was used for embedding.

**Collection of micro-tissue using the semi-automated micro-tissue punching system.** The micro-tissue collection device was fabricated by Frontier Biosystems (Tokyo, Japan). The sampling unit is designed such that it can be easily attached to a commercially available microscope. The device consists of a collection unit containing a tissue collection needle, a pumping unit for ejecting the punched tissue, and an injector unit that controls fluid pumping. The punched micro-tissues are ejected into an eight-micro tube strip. The site to be sampled is selected by observing the tissue under a microscope. Multiple locations can be automatically sampled by specifying the sampling site and the tube to be dispensed in the original system installed in the device.

For micro-tissue sampling, a hollow needle made of stainless steel, purchased from CASTEC (Kanagawa, Japan), was used. The tip of the sampling needle was knife-edged for smooth cutting of tissue sections. This sampling needle was connected to the injector via a polytetrafluoroethylene tube (Nichias, Tokyo, Japan). The tube was filled with a 99.5% ethanol solution. The internal parts of the collection needles were pre-washed with 70% ethanol, RNaseZap (Thermo Fisher Scientific), and 99.5% ethanol. A plastic Petri dish with a diameter of 35 mm was placed on the sample table. A 0.05 mm-thick silicone sheet (Kenneth, Osaka, Japan) was placed in the center of the Petri dish, and a cryofilm on which thin tissue sections were previously transferred was placed, with the surface of the tissue facing upward. In micro-tissue sampling, the tissue, the cryofilm on which the tissue was transferred, and the silicon sheet underlying it are captured together. The dispensed solution contained 99.5% ethanol. This was to inhibit RNA degradation. Tissue sections for micro-tissue sampling were stained with HE and then used. FF thin sections were stained after rapid fixation with 99.5% ethanol, whereas FFPE thin sections were stained after deparaffinization.

**Total RNA extraction from a tissue section.** Extraction of total RNA from FF or FFPE thin sections was performed using the RNeasy Mini Kit (QIAGEN, Hilden, Germany) or the RNeasy FFPE Kit (QIAGEN), respectively, according to the manufacturer's instructions. RINs of the extracted RNA were determined using TapeStation 4200 (Agilent, Tokyo, Japan), and concentration was determined using Qubit (Thermo Fisher). The samples were stored at  $-80^{\circ}\text{C}$  until use.

**cDNA library preparation.** Total RNA or a micro-tissue was used as a template for the cDNA library. Micro-tissue sections were dispensed with 99.5% ethanol during sampling and stored intact at  $-80^{\circ}\text{C}$ . A vacuum evaporator was used for cDNA library preparation to completely remove ethanol, followed by the subsequent reaction.

For cDNA library preparation, as the first step, the cell membranes were lysed with proteinase K, and the poly(A) RNA was purified using oligo(dT) magnetic beads (referred to as PPP), as previously described. Conducting the PPP process can reduce the carriers of non-coding RNA and increase the ratio of PCGs against the sequence-reads. For FFPE tissues, conducting the process at  $85^{\circ}\text{C}$  with 15 min of incubation followed by proteinase K treatment ensured the dissociation of the methylation crosslink. When purified total RNA was used as the starting material, 1  $\mu\text{L}$  of total RNA diluted to an appropriate concentration was used. When micro-tissue

sections were used as the starting material, ethanol was completely removed, as described above, and used for subsequent reactions. For each sample, a proteinase K reaction solution containing 5  $\mu\text{L}$  of PKD buffer (QIAGEN) and 0.31  $\mu\text{L}$  of ProK (QIAGEN) was added and incubated for 1 h at 56  $^{\circ}\text{C}$ . The solution was then mixed thoroughly with vortexing, spun down, and incubated at 85  $^{\circ}\text{C}$  for 15 min. Oligo(dT) magnetic beads (Dynabeads Oligo (dT)25, 61002, Thermo Fisher) used for mRNA purification were pre-washed three times with equal volumes of 1 $\times$  hybridization buffer (2 $\times$  SSPE, 0.05% Tween20, and 0.0025% RNase inhibitor) and then suspended in half the volume of 2 $\times$  hybridization buffer, in accordance with the instructions provided in the manual. The bead mixture solution was heated using a thermal cycler at 56  $^{\circ}\text{C}$  for 1 min, and then slowly cooled to 25  $^{\circ}\text{C}$  and allowed to stand for 10 min. The reaction mixture was then placed on a magnetic stand, and the supernatant was removed. The magnetic beads were washed twice with 100  $\mu\text{L}$  of ice-cold 1 $\times$  hybridization buffer and then washed with 100  $\mu\text{L}$  of ice-cold 1 $\times$  PBS supplemented with 0.0025% RNase inhibitor (Thermo Fisher), following which the wash solution was removed using a magnetic stand. The magnetic beads in which RNA was trapped were resuspended in 2.8  $\mu\text{L}$  of RNase-free water. The RNA bound to the complementary strands of poly(T) on magnetic beads was then denatured by heating at 80  $^{\circ}\text{C}$  for 2 min on a heat block. Finally, 2.5  $\mu\text{L}$  of the supernatant containing purified mRNA was recovered using a magnetic stand and used as material for cDNA library preparation. The total lysed and purified samples were directly processed according to the Smart-seq2 flow. The cDNA libraries were produced by adding ERCC at the same concentration (1:120,000 diluted ERCC spike-in mix) to all samples as an internal standard.

**Statistical analysis.** Statistical analyses were performed using R studio (ver 1.3.959) under R version 4.1.2 environment. The means of two independent groups were compared using Welch's t-test. The p-values were calculated and visualized using the R package "ggpubr"<sup>39</sup> based on "ggplot2"<sup>40</sup>.

**Sequencing and data analysis.** Amplified cDNA (0.25 ng) was used for preparing the sequencing library with the Nextera XT DNA Library Prep Kit (Illumina, San Francisco, CA, USA). Paired-end sequencing was performed on the MiSeq platform, with 75 bases each for read 1 (R1) and read 2 (R2). We trimmed the adapter sequences in all the sequence reads using Flexbar (ver. 3.5.0). The trimmed sequence reads were aligned to the Ensembl mouse reference genome (GRCm38 ver.92) for mouse liver tissue samples, including the ERCC sequences, using Hisat2 (ver. 2.1.0), with the default parameters. The gene expression levels, expressed as TPM, were calculated using Stringtie (ver. 2.1.7), with a transcriptome reference obtained from Ensembl. For comparison of the expression levels, the counts of mapped reads were normalized by the median of ratios in DESeq2<sup>41</sup>.

### Data availability

RNA-seq data were deposited in the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under the accession number PRJNA815867.

Received: 2 June 2022; Accepted: 3 November 2022

Published online: 14 November 2022

### References

- Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
- Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
- Stahl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
- Lein, E., Borm, L. E. & Linnarsson, S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358**, 64–69 (2017).
- Maniatis, S. *et al.* Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* **364**, 89–93 (2019).
- Rodrigues, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
- Wels, J., Kaplan, R. N., Rafii, S. & Lyden, D. Migratory neighbors and distant invaders: Tumor-associated niche cells. *Genes Dev.* **22**, 559–574 (2008).
- Sgro, A. E. *et al.* From intracellular signaling to population oscillations: Bridging size- and time-scales in collective behavior. *Mol. Syst. Biol.* **11**, 779 (2015).
- Dang, Y., Grundel, D. A. J. & Youk, H. Cellular dialogues: Cell-cell communication through diffusible molecules yields dynamic spatial patterns. *Cell Syst.* **10**, 82–98.e7 (2020).
- Asslaber, M. & Zatloukal, K. Biobanks: Transnational, European and global networks. *Brief. Funct. Genomic. Proteomic.* **6**, 193–201 (2007).
- Chetcuti, A. *et al.* Can archival tissue reveal answers to modern research questions?: Computer-aided histological assessment of neuroblastoma tumours collected over 60 years. *Microarrays (Basel)* **3**, 72–88 (2014).
- Hester, S. D. *et al.* Editor's Highlight: Dose-response analysis of RNA-Seq profiles in archival formalin-fixed paraffin-embedded samples. *Toxicol. Sci.* **154**, 202–213 (2016).
- Fraenkel-Conrat, H. & Olcott, H. S. The reaction of formaldehyde with proteins; cross-linking between amino and primary amide or guanidyl groups. *J. Am. Chem. Soc.* **70**, 2673–2684 (1948).
- Fraenkel-Conrat, H. & Olcott, S. H. Reaction of formaldehyde with proteins VI cross-linking of amino groups with phenol, imidazole, or indole groups. *J. Biol. Chem.* **174**, 827–843 (1948).
- Pennock, N. D. *et al.* RNA-seq from archival FFPE breast cancer samples: Molecular pathway fidelity and novel discovery. *BMC Med. Genomics* **12**, 195 (2019).
- Newton, Y. *et al.* Large scale, robust, and accurate whole transcriptome profiling from clinical formalin-fixed paraffin-embedded samples. *Sci. Rep.* **10**, 17597 (2020).

18. Zhao, Y. *et al.* Robustness of RNA sequencing on older formalin-fixed paraffin-embedded tissue from high-grade ovarian serous adenocarcinomas. *PLoS ONE* **14**, e0216050 (2019).
19. Li, J., Fu, C., Speed, T. P., Wang, W. & Symmans, W. F. Accurate RNA sequencing from formalin-fixed cancer tissue to represent high-quality transcriptome from frozen tissue. *JCO Precis. Oncol.* **2**, 1–9 (2018).
20. Lin, X. *et al.* A comparative analysis of RNA sequencing methods with ribosome RNA depletion for degraded and low-input total RNA from formalin-fixed and paraffin-embedded samples. *BMC Genomics* **20**, 831 (2019).
21. Marczyk, M. *et al.* The impact of RNA extraction method on accurate RNA sequencing from formalin-fixed paraffin-embedded tissues. *BMC Cancer* **19**, 1189 (2019).
22. Esteve-Codina, A. *et al.* A comparison of RNA-Seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. *PLoS ONE* **12**, e0170632 (2017).
23. Wimmer, I. *et al.* Systematic evaluation of RNA quality, microarray data reliability and pathway analysis in fresh, fresh frozen and formalin-fixed paraffin-embedded tissue samples. *Sci. Rep.* **8**, 6351 (2018).
24. Yoda, T. *et al.* Site-specific gene expression analysis using an automated tissue micro-dissection punching system. *Sci. Rep.* **7**, 4325 (2017).
25. Yamazaki, M. *et al.* Effective microtissue RNA extraction coupled with Smart-seq2 for reproducible and robust spatial transcriptome analysis. *Sci. Rep.* **10**, 7083 (2020).
26. Mueller, O. & Schroeder, A. RNA integrity number (RIN)—Standardization of RNA quality control. *Agil. Appl. Note.* **2011**, 1–8 (2004).
27. Schroeder, A. *et al.* The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **7**, 3 (2006).
28. Illumina. *Evaluating RNA Quality from FFPE Samples. Illumina Tech Note.* <https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/evaluating-rna-quality-from-ffpe-samples-technical-note-470-2014-001.pdf>:1–4 (2016).
29. Ding, C. *et al.* A cell-type-resolved liver proteome. *Mol. Cell. Proteomics* **15**, 3190–3202 (2016).
30. Qiagen. *How Much RNA Does a Typical Mammalian Cell Contain?: FAQ.*
31. Wu, A. R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11**, 41–46 (2014).
32. Song, Y., Ahn, J., Suh, Y., Davis, M. E. & Lee, K. Identification of novel tissue-specific genes by analysis of microarray databases: A human and mouse model. *PLoS ONE* **8**, e64483 (2013).
33. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
34. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant review with the integrative genomics viewer. *Cancer Res.* **77**, e31–e34 (2017).
35. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
36. Parkkila, S. *et al.* The calcium-binding protein S100P in normal and malignant human tissues. *BMC Clin. Pathol.* **8**, 2 (2008).
37. Kim, H. R., Park, J. S., Karabulut, H., Yasmin, F. & Jun, C. D. Transgelin-2: A double-edged sword in immunity and cancer metastasis. *Front. Cell Dev. Biol.* **9**, 606149 (2021).
38. Nicholson, A. G., Scagliotti, G., Tsao, M. S., Yatabe, Y. & Travis, W. D. 2021 WHO classification of lung cancer: A globally applicable and molecular biomarker-relevant classification. *J. Thorac. Oncol.* **17**, e80–e83 (2022).
39. *Ggpubr.* <https://rpkgs.datanovia.com/ggpubr/>.
40. Valero-Mora, P. M. ggplot2: Elegant graphics for data analysis. *J. Stat. Soft.* **35**(1), 1–3 (2010).
41. *Introduction to DGE-ARCHIVED.* [https://hbctraining.github.io/DGE\\_workshop/lessons/02\\_DGE\\_count\\_normalization.html\(DGE\)\\_workshop\\_02\\_DGE\\_count\\_normalization:20](https://hbctraining.github.io/DGE_workshop/lessons/02_DGE_count_normalization.html(DGE)_workshop_02_DGE_count_normalization:20).

## Acknowledgements

We thank Naoko Suzuki and Kiyofumi Takahashi for providing technical support. This research was supported by Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP21am0101104 and Research Support Project for Life Science and Drug Discovery (BINDS) under Grant Number 22ama121055. The super-computing resource was provided by the Human Genome Center (University of Tokyo).

## Author contributions

H.M., K.A. and H.T. conceived and designed the experiments. H.M., K.A., M.Y., M.H. and H.K. conducted the experiments and collected the data. H.M., K.A., M.Y., R.W. and K.I. performed computing analysis of the results. K.T., K.S., and T.H. collected and analyzed pathological data. H.M., S.A.Z., M.H. and H.T. wrote the manuscript.

## Competing interests

HK is a founder and shareholder of Frontier Biosystems, Inc., which provides the semi-automated micro-tissue punching system. The other authors have no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-23651-6>.

**Correspondence** and requests for materials should be addressed to H.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022