

Research Article

DV200 Index for Assessing RNA Integrity in Next-Generation Sequencing

Takehiro Matsubara ¹, Junichi Soh ^{2,3}, Mizuki Morita,⁴ Takahiro Uwabo,⁴ Shuta Tomida,⁵ Toshiyoshi Fujiwara,⁶ Susumu Kanazawa,⁷ Shinichi Toyooka,³ and Akira Hirasawa⁸

¹Okayama University Hospital Biobank, Okayama University Hospital, Japan

²Department of Surgery, Division of Thoracic Surgery, Kindai University Faculty of Medicine, Japan

³Department of General Thoracic Surgery, Breast and Endocrinological Surgery, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Japan

⁴Department of Biomedical Informatics, Okayama University Graduate School of Interdisciplinary Science and Engineering in Health Systems, Japan

⁵Department of Biobank, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Japan

⁶Department of Gastroenterological Surgery, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Japan

⁷Department of Radiology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Japan

⁸Department of Clinical Genomic Medicine, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Japan

Correspondence should be addressed to Junichi Soh; drjsou7@gmail.com

Received 13 June 2019; Revised 29 December 2019; Accepted 6 February 2020; Published 27 February 2020

Academic Editor: Fengjie Sun

Copyright © 2020 Takehiro Matsubara et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Poor quality of biological samples will result in an inaccurate analysis of next-generation sequencing (NGS). Therefore, methods to accurately evaluate sample integrity are needed. Among methods for evaluating RNA quality, the RNA integrity number equivalent (RINe) is widely used, whereas the DV200, which evaluates the percentage of fragments of >200 nucleotides, is also used as a quality assessment standard. In this study, we compared the RINe and DV200 RNA quality indexes to determine the most suitable RNA index for the NGS analysis. Seventy-one RNA samples were extracted from formalin-fixed paraffin-embedded tissue samples ($n = 30$), fresh-frozen samples ($n = 25$), or cell lines ($n = 16$). After assessing RNA quality using the RINe and DV200, we prepared two kinds of stranded mRNA sequencing libraries. Finally, we calculated the correlation between each RNA quality index and the amount of library product (1st PCR product per input RNA). The DV200 measure showed stronger correlation with the amount of library product than the RINe ($R^2 = 0.8208$ for the DV200 versus 0.6927 for the RINe). Receiver operating characteristic curve analyses revealed that the DV200 was the better marker for predicting efficient library production than the RINe using a threshold of >10 ng/ng for the amount of the 1st PCR product per input RNA (cutoff value for the RINe and DV200, 2.3 and 66.1%; area under the curve, 0.99 and 0.91; sensitivity, 82% and 92%; and specificity, 93% and 100%, respectively). Our results indicate that NGS libraries prepared using RNA samples with the DV200 value > 66.1% exhibit greater sensitivity and specificity than those prepared with the RINe values > 2.3. These findings suggest that the DV200 is superior to the RINe, especially for low-quality RNA, because it is a more consistent assessment of the amount of the 1st NGS library product per input.

1. Introduction

Next-generation sequencing (NGS) has become an essential technology in molecular biology research and clinical assessment [1–3]. However, the quality of the input biological samples has a critical effect on NGS results. It is important to grasp the quality of NGS results before conducting NGS analyses in order to avoid wasting precious samples and to minimize cost and labor.

Several RNA quality indexes have been developed, including RNA integrity number equivalent (RINe) and DV200 metrics (percentage of RNA fragments > 200 nucleotides in size). RINe is generally and widely used for assessing RNA integrity, and it is based on a mathematical model that calculates an objective quantitative measurement of RNA degradation that represents the relative ratio of signal in the fast zone to the 18S peak signal.

The DV200 was developed by Agilent in 2014 as a tool to more accurately assess the quality of RNA samples (<http://urx.red/OB4Y>) and used as an RNA quality assessment standard even in the protocol published by Illumina. Values indicative of high quality can be obtained with the DV200 even for samples exhibiting weak 18S and 28S peaks if there is a sufficient volume of RNA fragments greater than 200 nt in length. However, the best practice for evaluating RNA quality remains uncertain. In this study, we compared the two RNA quality indexes in terms of the amount of the 1st PCR product as preparation for NGS analyses in order to determine a much more suitable RNA quality index.

2. Materials and Methods

2.1. Data Collection. Seventy-one specimens were obtained at four sections of Okayama University Hospital (Center for Clinical Oncology; Department of Hematology, Oncology and Respiratory Medicine; Department of Respiratory Medicine; and Department of Thoracic, Breast and Endocrinological Surgery) during their own studies. All study protocols were approved by the Institutional Review Board/Ethical Committee of Okayama University, Okayama, Japan (reference numbers K1603-066, K1512-024, K1505-033, K1605-022, and K1808-009), and all participants signed written informed consent. Each section consigned an analysis by NGS to our biobank for its own research purpose and provided collected samples to Okayama University Hospital Biobank. This study uses only the data obtained in the steps of RNA extraction from the provided collected samples, preparation of the NGS library, and the NGS analysis, which were conducted at Okayama University Hospital Biobank. Detailed information regarding each of the samples used in the study is shown in Supplemental Table 1.

2.2. RNA Extraction and Quality Evaluation. RNA was extracted from frozen samples ($n = 25$) and cell lines ($n = 16$) using the RNeasy Mini kit (Qiagen, Hilden, Germany) or from formalin-fixed paraffin-embedded (FFPE) tissue samples ($n = 30$) using the RNeasy FFPE kit. RINe values were automatically determined on the basis of electropherograms generated using TapeStation HS RNA ScreenTape (Agilent Technologies, Santa Clara, CA, USA). We calculated the

DV200 values on the basis of the same electropherograms using TapeStation Analysis software.

2.3. NGS Library Construction. NGS libraries were prepared using TruSeq RNA Access (Illumina, San Diego, CA, USA) ($n = 63$) or TruSight RNA Pan-Cancer (Illumina, San Diego, CA, USA) ($n = 8$). The NGS library preparation kits utilized the same workflows: fragmentation, cDNA synthesis, 1st PCR, hybridization, 2nd PCR, and cleanup, although hybridization probes were different. The amount of the NGS library product was quantified using a Qubit 2.0 fluorometer (Thermo Fisher, Waltham, MA, USA).

2.4. Receiver Operating Characteristic Curve Analysis. We generated receiver operating characteristic (ROC) curves using JMP 9.0.2 software (SAS Institute Japan, Osaka, Japan). We determined >10 ng/ng for the 1st PCR product per input RNA as the threshold on the basis of the following factors: (1) 200 ng of 1st PCR product is needed to proceed to the 2nd PCR step for the NGS library preparation and (2) the minimum recommended input volume of RNA is 20 ng, as determined according to the following formula: $200 \text{ ng } 1^{\text{st}} \text{ PCR product} / 20 \text{ ng input volume} = 10 \text{ ng/ng}$.

3. Results

3.1. Distribution of RINe and DV200 Values. The median values (range) were 2.1 (1.0–9.6) for the RINe and 66.1% (24.6–97.3%) for the DV200. The DV200 values were relatively scattered, whereas the RINe values exhibited two peaks at approximately 1 to 4 and 8 to 10 (Figures 1(a) and 1(b); Supplementary Table 1).

3.2. Correlation between DV200 and RINe. As shown in Figure 1(c), the RINe and DV200 values were correlated ($R^2 = 0.6944$). It should be noted that 12 of 32 (37.5%) samples with a low RINe value (<5) exhibited a high DV200 value (>70%), suggesting that the DV200, compared with RINe, has the potential to increase the number of samples available for the following assays.

3.3. RINe and DV200 Values and NGS Library Preparation. The median of the 1st NGS library product per input was 41.0 ng/ μ l (0.01–129.5 ng/ μ l) (Supplemental Table 1). Both the RINe and DV200 values correlated positively with the amount of the 1st NGS library product, although the DV200 exhibited a better correlation than the RINe index ($R^2 = 0.8208$ versus 0.6927, respectively) (Figures 2(a) and 2(b)). The fresh and FFPE samples were extracted and analyzed separately from the other samples to investigate the effects of different sample types. In the fresh samples, a high RINe value, a high DV200 value, and a sufficient amount of the 1st NGS library product were obtained (more than 8.3, 89.32, and 73.15 ng/ng, respectively), even though the R^2 value of the RINe was higher than that of the DV200 (Figures 2(c) and 2(d)). Although the amount of the 1st NGS library product was low in all FFPE samples, the DV200 showed better R^2 value than the RINe (0.0294 versus 0.0006), indicating that the DV200

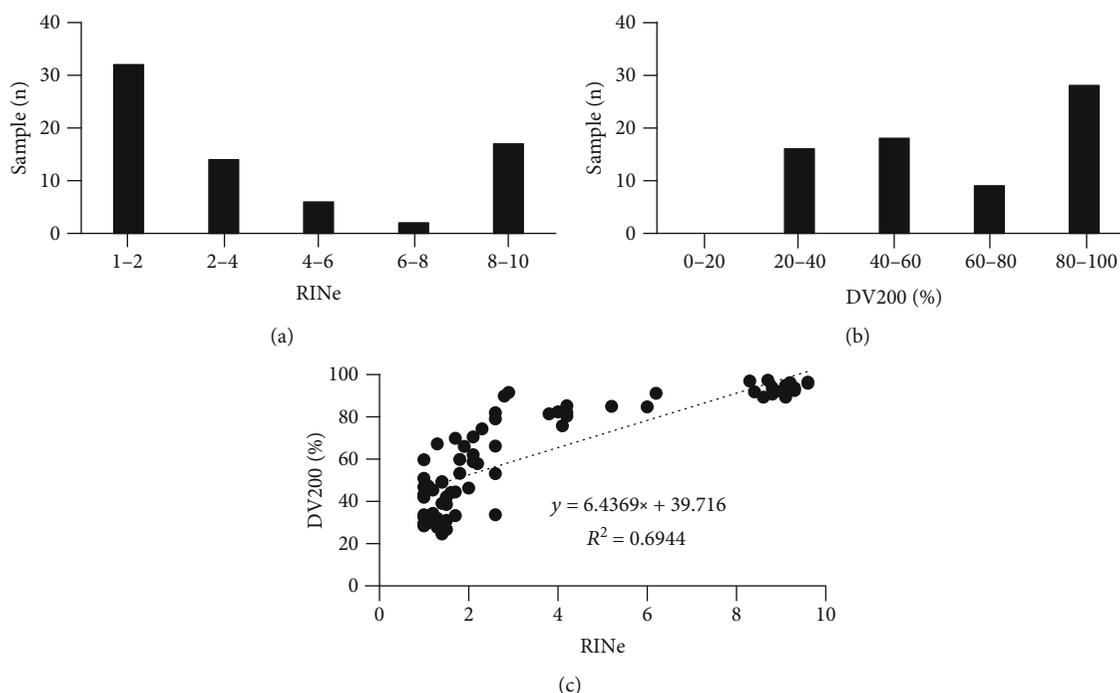


FIGURE 1: Relationship between RINe and DV200 values. (a, b) Distribution of RINe and DV200 values. Graphs show the distribution of RINe and DV200 values categorized in 2-point and 20-point increments, respectively. (c) Correlation between RINe and DV200 values. RINe and DV200 values were determined using TapeStation 2200.

is useful for evaluating RNA in low-quality samples such as FFPE.

3.4. Receiver Operating Characteristic Curve Analyses. The analysis of ROC curves indicated that the optimal RINe and DV200 cutoff values were 2.3 and 66.1%, respectively, when >10 ng/ng for the 1st PCR product per RNA input was considered a sufficient amount for NGS. The area under the curve (AUC) for the DV200 was 0.99, with sensitivity of 92% and specificity of 100%, whereas the AUC for the RINe was 0.91, with sensitivity of 82% and specificity of 93% (Figures 3(a) and 3(b)).

4. Discussion

Remarkable progress in development of NGS technologies has made it possible to analyze a variety of specimens, including highly degraded materials such as 10-year-old FFPE samples [4]. The RINe has been widely used as an indicator of RNA quality in NGS, microarray, and qPCR [5–7]. However, the DV200 is more suitable than the RINe for quantification of RNA because it can be applied to evaluate not only RNAs extracted from fresh or frozen samples but also samples with lower RINe values, such as RNAs extracted from FFPE samples [8, 9]. In our study, the DV200 showed better correlation with the amount of the 1st NGS library product compared with the RINe even for low-quality samples such as FFPE. Recently, paraffin-embedded RNA metrics (PERM) is also proposed as a novel indicator that is based on the intensity of fluorescence at specific time points using the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA)

[10]. Although we attempted to perform a PERM analysis, unfortunately, the TapeStation used in this study did not support the PERM analysis.

Furthermore, our study also revealed that the DV200 with a cutoff value of 66.1% provided greater AUC, sensitivity, and specificity than the RINe (cutoff value 2.3) on the basis of the analysis of ROC curves. These results indicate that the DV200 with a cutoff value of 66.1% is more useful than the RINe for predicting whether a sufficient amount of high-quality 1st NGS product can be obtained.

In addition to the 1st NGS library product per input, we examined the effect of RNA quantification on quality metrics of RNA sequencing (RNA-seq): duplicates, reads not mapped, and nonspecific matches. As shown in Supplemental Figure 1, the DV200 showed better R^2 values than the RINe. Consistent with our report, another study reported a positive correlation between the DV200 value and the number of uniquely mapped NGS reads, which are reads mapped to one region of the reference genome [11]. By contrast, sample selection based on the RINe values reportedly provides no advantage for determining the quality of NGS reads [12]. In order to analyze some functional relationships between the RNA quality and the result of RNA-seq, we analyzed the transcripts per million (TPM) of protein coding genes (Supplemental Figure 2), and we found that the total TPM of protein coding genes in all the fresh samples (RINe > 8.0 and DV200 $> 89\%$) exceeds 950,000 (meaning 95% of total RNA-seq reads). This result suggests that RNA-seq with high-quality input RNA using TruSeq RNA Access library preparation protocols could capture the whole picture of gene expression of protein coding genes with the least

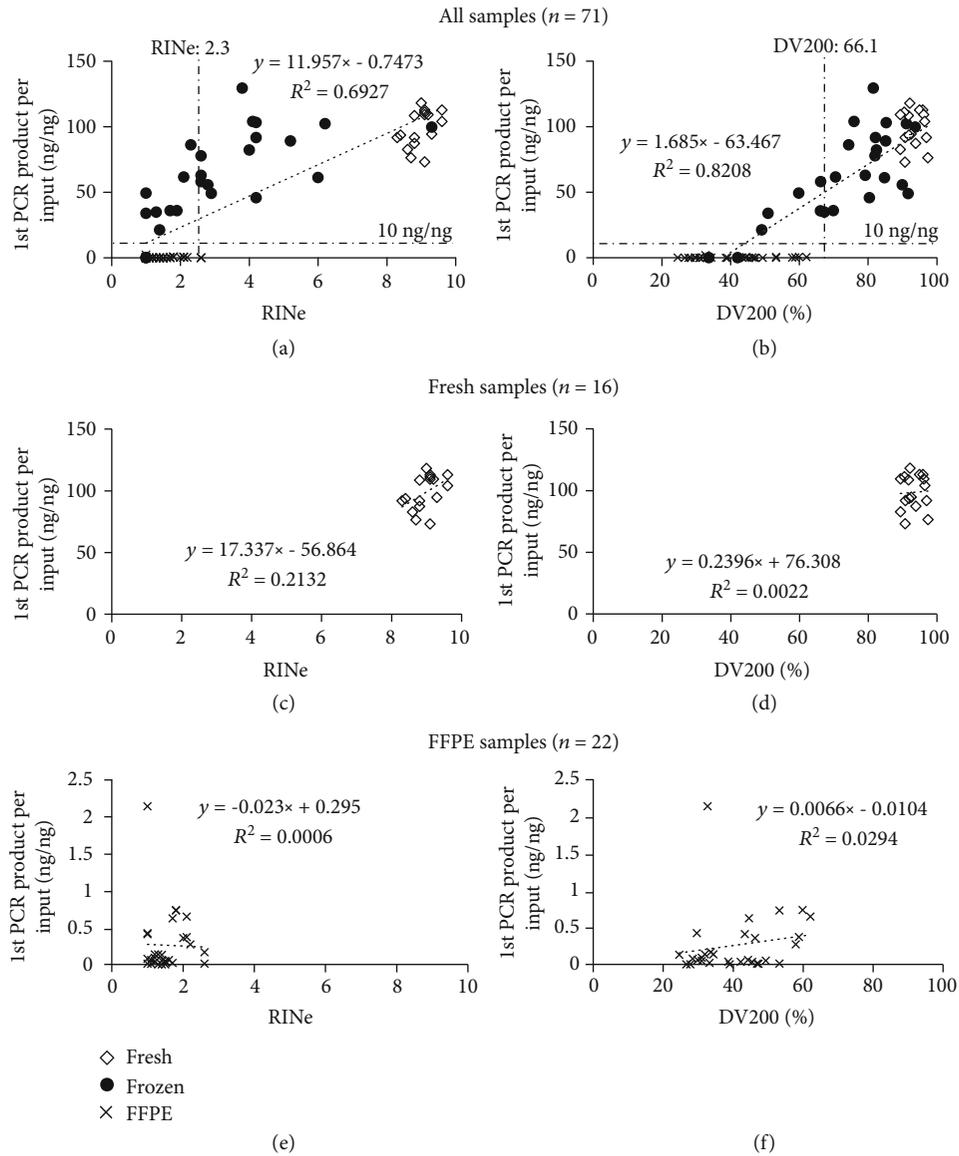


FIGURE 2: Correlation between RNA quality indexes and NGS library yields. (a, b) Correlation between RINe and DV200 of all samples in terms of amount of the 1st NGS library product per input (◇: fresh, ●: frozen, and ×: FFPE). NGS libraries were prepared using TruSeq RNA Access or TruSight RNA Pan-Cancer. The amount of the 1st NGS library product was measured using Qubit. Threshold lines are drawn for the amount of the 1st NGS library product (10 ng/ng) and the cutoff value (RINe: 2.3 and DV200: 66.1) on the basis of receiver operating characteristic (ROC) curve analysis. (c)–(f) are segregated from all samples depending on sample type: fresh sample (◇: c, d) or FFPE sample (×: e, f).

information loss. On the other hand, the total TPM of protein coding genes in all the FFPE samples (RINe < 3.0 and DV200 < 55%) ranged from 675,000 to 778,000 (meaning 67.5–77.8% of total RNA-seq reads) with one outlier (578,000). This result suggests that RNA-seq for low-quality input RNA may lead to the gene expression profiles with some information loss due to the potential RNA degradation/fragmentation. The total TPM of protein coding genes in frozen samples ranged widely from 150,000 to 970,000 on the basis of their RNA quality. On the other hand, in samples with RINe values of 2 or less, some samples had TPM values of more than 800,000, but others

had TPM values of 800,000 or less. This result suggests that careful interpretation is required when using RNA with an RINe value of 2 or less.

These data suggest that the DV200 index is superior to RINe for assessing RNA integrity in order to obtain NGS results worthy of evaluation.

In general, the time required for tissue acquisition, fixation, and preservation is important for RNA quality [13, 14]. However, unfortunately, we could not obtain detailed information including ischemia time, interval from sample collection to formalin fixation, and formalin fixation duration. Currently, we are planning to obtain the duration of

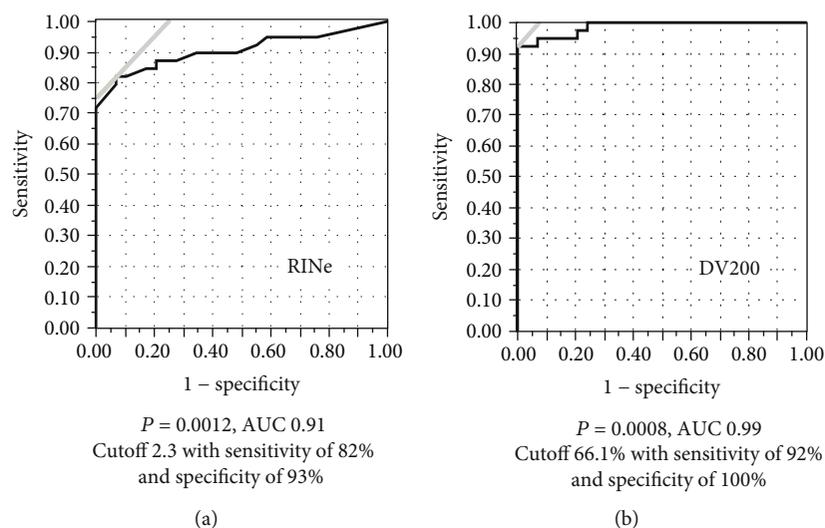


FIGURE 3: Receiver operating characteristic curves for RINe and DV200. ROC curves for the RINe and DV200 indexes indicating the most efficient amount (more than 10 ng/ng) of the 1st PCR library product per input. The area under the curve for DV200 was greater than that for the RINe index (0.99 with $P = 0.0008$ and 0.91 with $P = 0.0012$, respectively).

processing for sample preservation to investigate the effect of the duration of the preservation process on RNA quality as well as NGS libraries.

5. Conclusion

The DV200 index is a more consistent assessment of the amount of the 1st NGS library product per input than the RINe index, especially for low-quality RNA. Therefore, we conclude that the DV200 is a beneficial RNA quality index for NGS analyses using degraded RNA samples such as those extracted from FFPE samples.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

We thank Hisao Kubo (Center for Clinical Oncology, Okayama University Hospital), Go Makimoto (Department of Hematology, Oncology and Respiratory Medicine, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences), Kadoaki Oohashi, Katsuyuki Kiura (Department of Respiratory Medicine, Okayama University Hospital), Hidejiro Torigoe (Department of Thoracic, Breast and Endocrinological Surgery, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences), and Hisao Higo (Department of Hematology, Oncology and Respiratory Medicine, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical

Sciences) for providing data from their own studies, and we thank Hiroko Hanafusa, Natsumi Hibino, and Ruriko Ogawa (Okayama University Hospital Biobank, Okayama University Hospital) for technical support. This research was partially supported by Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (16K10681, 16K19454, and 15H04830), the Translational Research Program; Strategic PROMotion for practical application of INnovative Medical Technology (TR-SPRINT), and the program for an Integrated Database of Clinical and Genomic Information from the Japan Agency for Medical Research and Development (AMED).

Supplementary Materials

Supplementary 1. Supplemental Table 1. list of samples.

Supplementary 2. Supplemental Figure 1: correlation between RNA quality indexes and RNA-seq quality data. Supplemental Figure 2: functional relationship between the RNA quality and the result of RNA-seq.

References

- [1] Y. C. Zhang, Q. Zhou, and Y. L. Wu, "The emerging roles of NGS-based liquid biopsy in non-small cell lung cancer," *Journal of Hematology & Oncology*, vol. 10, no. 1, p. 167, 2017.
- [2] A. Fernandez-Marmiesse, S. Gouveia, and M. L. Couce, "NGS technologies as a turning point in rare disease research, diagnosis and treatment," *Current Medicinal Chemistry*, vol. 25, no. 3, pp. 404–432, 2018.
- [3] S. M. Aly and D. M. Sabri, "Next generation sequencing (NGS): a golden tool in forensic toolkit," *Archiwum Medycyny Sądowej i Kryminologii*, vol. 4, pp. 260–271, 2015.
- [4] W. Huang, M. Goldfischer, S. Babayeva et al., "Identification of a novel PARP14-TFE3 gene fusion from 10-year-old FFPE tissue by RNA-seq," *Genes, Chromosomes & Cancer*, vol. 54, no. 8, pp. 500–505, 2015.

- [5] D. L. Gibbons, W. Lin, C. J. Creighton et al., "Contextual extracellular cues promote tumor cell EMT and metastasis by regulating miR-200 family expression," *Genes & Development*, vol. 23, no. 18, pp. 2140–2151, 2009.
- [6] E. Sanz, L. Yang, T. Su, D. R. Morris, G. S. McKnight, and P. S. Amieux, "Cell-type-specific isolation of ribosome-associated mRNA from complex tissues," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 33, pp. 13939–13944, 2009.
- [7] C. Endrullat, J. Glökler, P. Franke, and M. Frohme, "Standardization and quality management in next-generation sequencing," *Applied & Translational Genomics*, vol. 10, pp. 2–9, 2016.
- [8] I. Sanchez, F. Betsou, and W. Mathieson, "Does vacuum centrifugal concentration reduce yield or quality of nucleic acids extracted from FFPE biospecimens?," *Analytical Biochemistry*, vol. 566, pp. 16–19, 2019.
- [9] L. Landolt, H. P. Marti, C. Beisland, A. Flatberg, and O. S. Eikrem, "RNA extraction for RNA sequencing of archival renal tissues," *Scandinavian Journal of Clinical and Laboratory Investigation*, vol. 76, no. 5, pp. 426–434, 2016.
- [10] J. Y. Chung, H. Cho, and S. M. Hewitt, "The paraffin-embedded RNA metric (PERM) for RNA isolated from formalin-fixed, paraffin-embedded tissue," *BioTechniques*, vol. 60, no. 5, pp. 239–244, 2016.
- [11] S. Santoro, I. D. Lopez, R. Lombardi et al., "Laser capture microdissection for transcriptomic profiles in human skin biopsies," *BMC Molecular Biology*, vol. 19, no. 1, p. 7, 2018.
- [12] I. Gallego Romero, A. A. Pai, J. Tung, and Y. Gilad, "RNA-seq: impact of RNA degradation on transcript quantification," *BMC Biology*, vol. 12, no. 1, p. 42, 2014.
- [13] H. Sun, R. Sun, M. Hao et al., "Effect of duration of ex vivo ischemia time and storage period on RNA quality in bio-banked human renal cell carcinoma tissue," *Annals of Surgical Oncology*, vol. 23, no. 1, pp. 297–304, 2016.
- [14] M. Srinivasan, D. Sedmak, and S. Jewell, "Effect of fixatives and tissue processing on the content and integrity of nucleic acids," *The American Journal of Pathology*, vol. 161, no. 6, pp. 1961–1971, 2002.