Received: 2 March 2020 | Accepted: 3 March 2020
DOI: 10.1002/gcc.22843

RESEARCH ARTICLE

WILEY

Immuno-oncology gene expression profiling of formalin-fixed and paraffin-embedded clear cell renal cell carcinoma: Performance comparison of the NanoString nCounter technology with targeted RNA sequencing

Suranand B. Talla[1] | Eugen Rempel[1,2] | Volker Endris[1] | Maximilian Jenzer[4] | Michael Allgäuer[1] | Constantin Schwab[1] | Daniel Kazdal[1] | Fabian Stögbauer[3] | Anna-Lena Volckmar[1] | Ildiko Kocsmar[1] | Olaf Neumann[1] | Peter Schirmacher[1,2] | Stefanie Zschäbitz[4] | Stefan Duensing[5] | Jan Budczies[1,2] | Albrecht Stenzinger[1,2] | Martina Kirchner[1,2]

[1]Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany

[2]German Cancer Consortium (DKTK), Heidelberg Partner Site, Heidelberg, Germany

[3]Institute of Pathology, Technical University of Munich, Munich, Germany

[4]Department of Medical Oncology, National Center for Tumor Diseases (NCT) Heidelberg, Heidelberg, Germany

[5]Molecular Urooncology, Department of Urology, University Hospital Heidelberg, Heidelberg, Germany

Correspondence
Martina Kirchner, PhD, Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany.
Email: martina.kirchner@med.uni-heidelberg.de

Albrecht Stenzinger, MD, Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany.
Email: albrecht.stenzinger@med.uni-heidelberg.de

Abstract

Inflammatory gene signatures are currently being explored as predictive biomarkers for immune checkpoint blockade, and particularly for the treatment of renal cell cancers. From a diagnostic point of view, the nCounter analysis platform and targeted RNA sequencing are emerging alternatives to microarrays and comprehensive transcriptome sequencing in assessing formalin-fixed and paraffin-embedded (FFPE) cancer samples. So far, no systematic study has analyzed and compared the technical performance metrics of these two approaches. Filling this gap, we performed a head-to-head comparison of two commercially available immune gene expression assays, using clear cell renal cell cancer FFPE specimens. We compared the nCounter system that utilizes a direct hybridization technology without amplification with an NGS assay that is based on targeted RNA-sequencing with preamplification. We found that both platforms displayed high technical reproducibility and accuracy (Pearson coefficient: ≥0.96, concordance correlation coefficient [CCC]: ≥0.93). A density plot for normalized expression of shared genes on both platforms showed a comparable bi-modal distribution and dynamic range. RNA-Seq demonstrated relatively larger

Suranand B. Talla and Eugen Rempel share first authorship. Albrecht Stenzinger and Martina Kirchner share last authorship.

signaling intensity whereas the nCounter system displayed higher inter-sample variability. Estimated fold changes for all shared genes showed high correlation (Spearman coefficient: 0.73). This agreement is even better when only significantly differentially expressed genes were compared. Composite gene expression profiles, such as an interferon gamma (IFNg) signature, can be reliably inferred by both assays. In summary, our study demonstrates that focused transcript read-outs can reliably be achieved by both technologies and that both approaches achieve comparable results despite their intrinsic technical differences.

KEYWORDS

FFPE, immune checkpoint blockade (ICB), NanoString, renal cell cancer, RNA-Seq

1 | INTRODUCTION

Renal cell carcinoma (RCC) ranks globally as the sixth most frequent cancer in men and tenth in women, accounting for 5% and 3% of all cancer diagnoses, respectively.[1] Clear cell renal cell carcinomas (ccRCC) make up a significant proportion (approx. 70%) of RCCs and display prolific angiogenesis due HIF-induced vascular endothelial growth factor (VEGF) overexpression following a loss-of-function mutation of the *von Hippel-Lindau* gene.[2,3] This molecular feature can be therapeutically exploited by combinatorial therapy regimens using tyrosine kinase inhibitors (TKIs) and anti-VEGF antibodies.[4,5] In addition to TKIs, the results of the phase III Checkmate-214 trial led to the approval of the combination of ipilimumab and nivolumab for patients with intermediate and poor risk metastatic RCC (mRCC) as first line treatment in 2018.[6] Since anti-VEGF therapy is reported to enhance the antitumor activity by increasing T-cell infiltration, MHC class-I expression, and reversing myeloid immunosuppression,[7-12] combinatorial treatment regimens are investigated in several clinical trials. Due to prolonged progression free survival and overall response rate in favor for pembrolizumab plus axitinib compared to sunitinib, the first combination treatment for a VEGFR-inhibitor and anti-PD-1 antibody was approved in 2019 (Keynote-426 trial) followed shortly by the approval of avelumab plus axitinib (JAVELIN-Renal 101).[13] The positive results of these trials changed current guidelines making dual treatment—either tyrosine kinase inhibitor plus immune checkpoint blockade (ICB) or double ICB—the current first line standard therapy. Even though immune checkpoint inhibition and VEGF-targeted therapy have become a standard-of-care for metastatic RCC (mRCC), potential biomarkers for therapy response and innate and acquired resistance mechanisms are still being investigated and are not available in clinical settings yet. A recent study showed that RNA-based signatures interrogating angiogenesis and specific inflammatory conditions, such as upregulation of interferon-gamma signaling, were associated with prolonged progression-free survival under combinatorial therapy with sunitinib and atezolizumab.[14] Pointing in a similar direction, an exploratory analysis of the JAVELIN 101 trial suggested that focused gene expression profiling might identify patients who benefit the most from these novel combinatorial therapies.[15] Implementation of such biomarkers in a routine diagnostic setting requires the use of specific assays that are able to robustly interrogate formalin-fixed and paraffin-embedded (FFPE) clear cell renal cell carcinoma (ccRCC) samples but comparative studies analyzing technical aspects of different high throughput platforms are lacking. As this information is crucial for predictive algorithms in the future, we investigated two commercially available immune gene expression assays in a head-to-head comparison: the nCounter analysis system based Pan Cancer Immune Profiling Panel (NanoString Technologies, Seattle, WA) that utilizes a direct hybridization technology without amplification[16] and the Oncomine Immune Response Research Assay (Thermo Scientific, Waltham, MA), which is based on targeted RNA-sequencing with preamplification.[17]

2 | MATERIALS AND METHODS

2.1 | Tumor samples

Tumor specimens from patients with clear cell renal cell carcinoma (ccRCC) were obtained from the NCT Tissue Bank Heidelberg at the Institute of Pathology Heidelberg (IPH). A total of 34 tumor samples of 14 renal cell carcinoma patients were subjected to gene expression profiling on both platforms. We analyzed 34 samples on both platforms. Six of these failed RNA-Seq and two nCounter analysis by either not producing enough read counts or not passing quality control. This led to 27 evaluable tumor samples on both platforms (Table S1).

2.2 | RNA extraction procedure

FFPE samples of ccRCC patients were stained with hematoxylin and eosin to identify and mark tumor regions for microdissection and to estimate tumor cell content. Mean tumor cellularity was 65% (minimum: 30%, maximum: 95%) (Table S1). RNA from 5 μm thick FFPE tissue sections was extracted using the Maxwell extraction system (Promega, Madison, WI) following the manufacturer's instructions and quantified by Qubit (Thermo Fisher Scientific, Waltham, MA).

2.3 | NanoString analysis

Isolated RNA was used for mRNA expression analysis on the nCounter-based NanoString instrument. RNA of 100 ng of each sample was hybridized using the Human Pan Cancer Immune Profiling Panel on the nCounter system (both NanoString Technologies, Seattle, WA), following the manufacturer's recommendations. Absolute read counts were quantified by the nCounter digital analyzer at the Institute of Pathology, Heidelberg. The Human Pan Cancer Immune Profiling Panel is designed to quantitate 730 target genes, 40 housekeeping genes, and additional positive and negative controls. Target genes mainly fall into three categories: Immune cell identification, estimation of immune cell function, and identifying tumor-specific antigens. For downstream analysis, absolute read counts of all panel genes were extracted from the nSolver software (NanoString Technologies). Target genes were normalized to 40 reference genes and fold changes and associated statistics were performed with the R-based NanoStringDiff package.[18] Fold change of each gene was calculated as the ratio of the average gene expression in the male group to that of the female group. Of note, for studying the effect of RNA input concentration on nCounter performance, absolute read counts detected at all RNA inputs were normalized individually by positive controls alone (technical normalization) and by positive control plus housekeeping genes.

2.4 | RNA sequencing

Complementary DNA (cDNA) was synthesized using RNA extracted from tumor samples of renal cell carcinoma patients. RNA of 10 ng from each sample was reverse transcribed using SuperScript VILO cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham, MA) at recommended PCR conditions. Oncomine Immune Response Research Assay (Thermo Fisher Scientific, Waltham, MA) libraries were prepared using the Ion AmpliSeq targeted sequencing technology (Thermo Fisher Scientific, Waltham, MA), as per manufacturer's instructions. The Oncomine panel consists of 395 genes focused on various immunological mechanisms including T-cell receptors, tumor infiltration of immune cells, immune checkpoints, and other prominent immune functions. Complementary DNA was amplified with the Oncomine primer pool targeting 395 genes. Amplicons were partially digested using the FuPa Reagent (Thermo Fisher Scientific, Waltham, MA). Afterwards, adapters were ligated (Ion Xpress Barcode Adapters, Thermo Fisher Scientific, Waltham, MA). Libraries were finally purified using AMPure XP magnetic beads (Beckman Coulter, Krefeld, Germany), and quantified using qPCR (Ion Library Quantitation Kit, Thermo Fisher Scientific, Waltham, MA) on a StepOne Plus qPCR machine (Thermo Fisher Scientific, Waltham, MA). All libraries were diluted to a concentration of 100 pM. Eight sample libraries were pooled together and amplified on Ion Spheres using the Ion Chef system (Thermo Fisher Scientific). Barcoded libraries were loaded onto an Ion 520 chip and sequenced using the Ion S5 XL System. The Torrent Suite software (v5.0.2) and the immuneResponseRNA plugin (both from Thermo Fisher Scientific) were used to extract absolute read count data, reads per million (RPM) and associated quality parameters. DESeq2[19] was employed for the normalization of gene read counts. Fold change of each gene was measured as the ratio of mean gene expression in the male group to that of female group. The description of the two workflows is shown in Figure 1.

2.5 | cDNA synthesis and qRT-PCR

Complementary DNA (cDNA) was synthesized by using SuperScriptIV VILO Master Mix (Thermo Fisher Scientific, Waltham, MA) as per manufacturer's instructions. Quantitative polymerase chain reactions were performed to find out relative gene expression using iTaq Universal SYBR Green Supermix (Bio-Rad, Dreieich, Germany) on QuantStudio5 (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA). The primer sequences of the primers used in qRT-PCR are listed in Table S2. All genes were normalized by *PPIA* and *RPS13* housekeeping genes. The average Ct of two housekeeping genes was subtracted from Ct values of target genes. After subtraction of housekeeping genes Ct values of target genes were compared with RNA-Seq and nCounter normalized read counts for linear correlation analysis.

2.6 | Statistical analyses

The R software (v.3.5.1; R Core Team, 2018) and GraphPad Prism version 7 (GraphPad Software, La Jolla, CA) were utilized for generating all graphs and correlation analyses. Pearson correlations were calculated for assessment of the intra-platform reproducibility, effect of RNA input on nCounter performance, and gene level correlations between the two methods. Concordance correlation coefficients (CCCs) were used to assess the precision and accuracy of intra-platform reproducibility and the effect of RNA concentration on nCounter.[20-22] Signature genes (*CD8A*, *EOMES*, *PRF1*, *IFNG*, and *CD274*) were combined into a single score by first *z*-score normalizing genes across all samples, and then averaging *z*-scores across the genes in the signature to create a single signature score for each sample. Besides correlation statistics, inter-platform concordance was also studied by Bland-Altman plot, using all measurable shared gene fold changes, that demonstrates the agreement between paired quantitative measurements and is widely regarded as the gold standard.[23] This is based on the statistical concept that displaying the differences between the pairs of quantitative measurements might offer better insight into the pattern and the "true" extent of agreement.

The significance of differential gene expression was assessed using likelihood ratio test for NanoString data (*glm.LRT* function implemented in the NanoStringDiff package) and Wald's test for RNA-Seq data. Subsequently, *P* values were corrected for multiple testing using the Benjamini-Hochberg method.[24] Lists of differentially expressed genes were generated controlling the false discovery rate (FDR) at 5%. Differentially expressed genes detected by one of the

platforms were considered as confirmed by the other platform when the raw *P* value was smaller than .05 (Tables S3 and S4).

3 | RESULTS

3.1 | Reproducibility of nCounter and RNA-Seq analysis

To examine the technical reproducibility, we assessed a limited set of samples (n = 3) on both platforms with each three technical replicates. All possible pairwise correlations were performed. Figures 2 and S1 illustrate scatterplots generated by correlating $\log_2$ RPM (reads per million) values from RNA-Seq and $\log_2$ transformed absolute read count values from nCounter. Both platforms demonstrated very high intra platform reproducibility (Pearson correlation (*r*) ≥0.96). Estimated concordance correlation coefficients for technical replicates (CCCs), that assert precision and accuracy, reached high levels too, providing further evidence to support this finding.

3.2 | Effect of RNA input concentration on nCounter performance

To study the effect of RNA input concentration on the nCounter performance, we measured gene expression at different amounts of RNA. Altogether, six concentrations were used: 100, 50, 25, 12.5, 6.25, and 3.125 ng (manufacturer's recommendation is to use at least 50 ng). The results of two independent runs are presented as pair plots in Figures 3A and S2A. We found that a high correlation is preserved even at 50, 25 and 12.5 ng when compared to 100 ng, as indicated by Pearson coefficients of at least 0.93; however, CCC plummeted sharply at 12.5 ng in two instances. Moreover, for all decreasing concentrations (50, 25, 12.5, 6.25, 3.125 ng), we also noticed strong correlation with the 100 ng sample for all genes having expression levels of 6 or above (log scale) while lowly expressed genes were barely detected in all diluted samples. Additionally, histograms and associated density plots showed gradual compression of bi-modal gene distribution in response to decreasing RNA input. Together, this illustrates that the nCounter system may provide reliable results down to RNA concentration of as low as 25 ng, an important finding since clinical samples are often far from ideal.

Additionally, we sought to study the effect of input RNA amount on gene read count intensity. To this end, absolute read counts were normalized separately by positive controls and housekeeping (HK) genes (Figures 3B and 2B). Further, distribution intensities for highly (top row; ≥100 read counts) and lowly (bottom row; <100 read counts) expressed genes were examined separately following normalization. Genes found to be highly and lowly expressed at an initial concentration of 100 ng were further monitored for their distribution at decreasing concentrations. HK-normalized, highly expressed genes

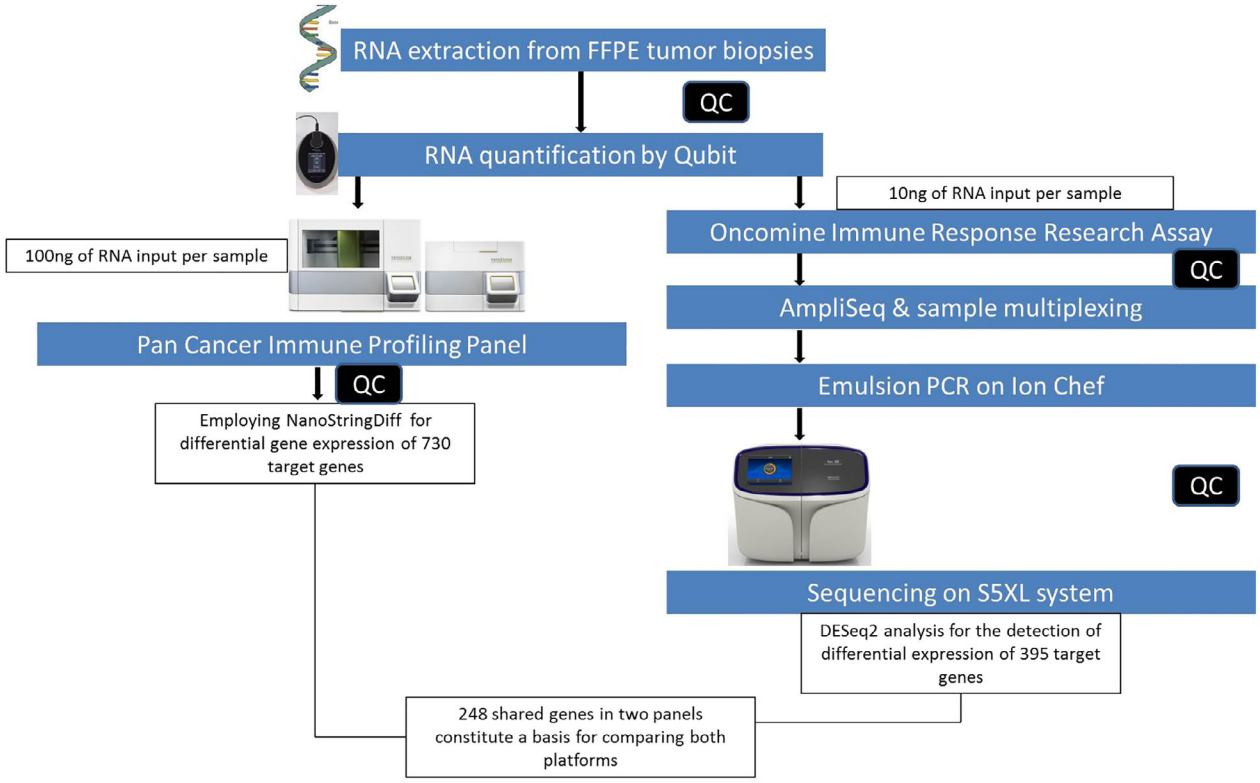

FIGURE 1 Platform comparison of workflows for differential gene expression analysis. The Pan Cancer Immune Profiling panel based on the nCounter technology was compared with the Oncomine Immune Response Research Assay sequenced by Ion Torrent semiconductor sequencing technology [Color figure can be viewed at wileyonlinelibrary.com]

showed consistent gene distribution and median values at RNA amounts of as low as 12.5 ng, arguing that nCounter might still be able to detect highly expressed genes even at very low RNA levels. HK-normalized, lowly expressed genes showed a comparable median only at 50 ng of RNA. Below this threshold, median values declined sharply, emphasizing that nCounter might require at least 50 ng RNA to detect lowly expressed genes. However, HK-normalization evens out the effect of input RNA differences on read counts that we wanted also to determine. Therefore we performed technical normalization with positive controls only. Following positive control normalization only, highly expressed genes showed a gradually decreasing median expression, whereas lowly expressed genes showed a sharp decline of median gene expression levels even at 50 ng of RNA; beyond that it is plateaued. Taken together, this suggests that nCounter is highly sensitive to RNA concentrations. However, highly expressed genes may still be robustly be detected even at as low as 12.5 ng but lowly expressed genes might require the recommended minimum of 50 ng. Table S5 shows the percentage of genes with at least 100 read counts after positive control (PC) and housekeeping gene (HK) normalization. This is also shown in Figures 3 and S2, respectively.

3.3 | Correlating performance metrics of nCounter and Ion Torrent RNA-Seq techniques

For a technical comparison of the expression patterns obtained by both platforms, pre-processed, normalized and $\log_2$ transformed read counts of 248 shared genes (ie, genes included in both panels) were juxtaposed (listed in Table S6). The density plot in Figure S3A shows the averaged distributions of signal intensities for 27 common samples with 0.95 confidence intervals. Both panels exhibited a bi-modal distribution. Inter-sample variability was visibly larger for the nCounter platform. The dynamic ranges of signal intensities were comparable, whereas the mode of the distribution (most common signal intensity) was stronger for RNA-Seq. In keeping with this, the number of genes with at least 100 normalized read counts in the 27 analyzed samples were found to be higher for RNA-Seq (Figure S3B).

Further, we sought to investigate the gene-wise correlation between both platforms. To this end, Pearson correlations for 248 shared genes were calculated (listed in Table S7). We plotted the histogram for gene level correlation coefficients in Figure 4A. Out of

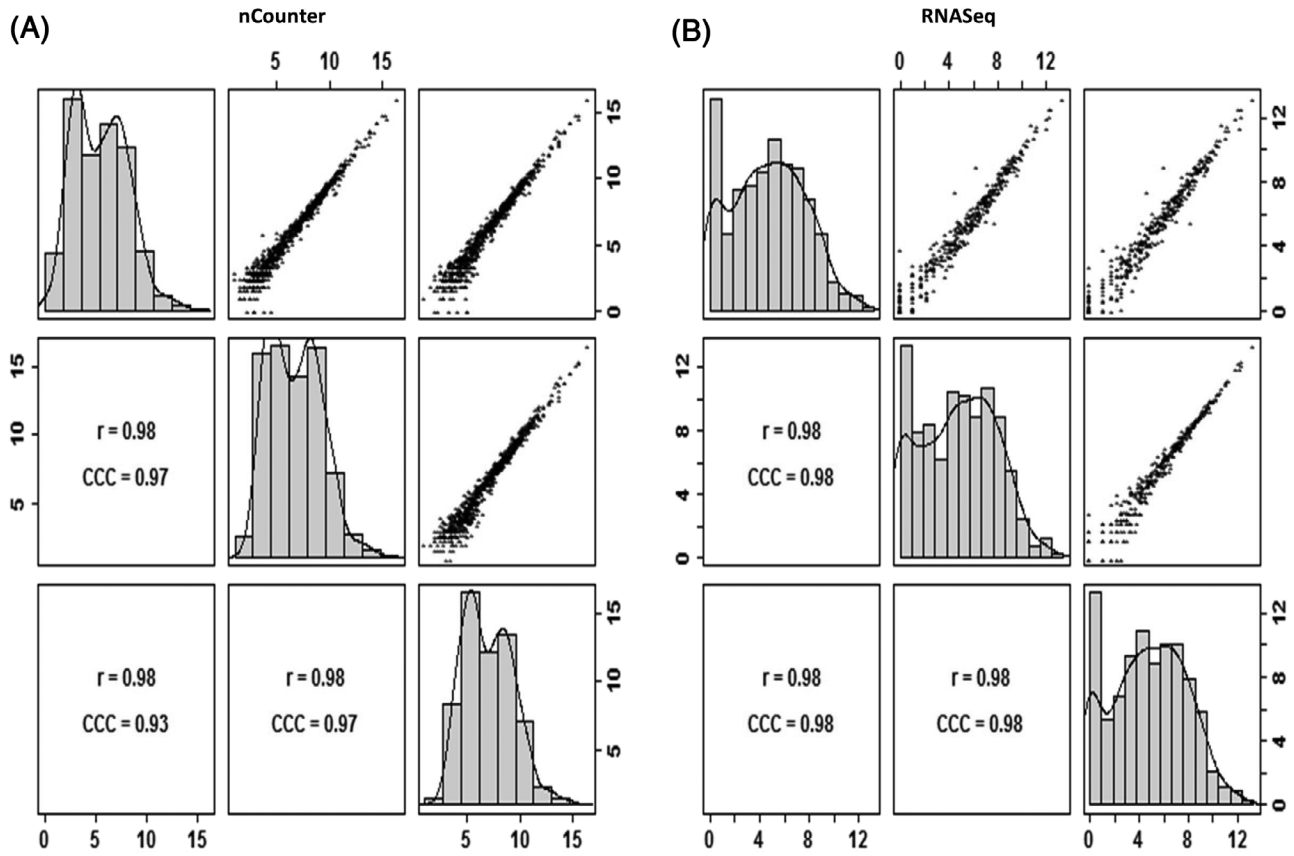

FIGURE 2 Demonstration of RNA-Seq and nCounter reproducibility. A, Scatter plots are generated using $\log_2$ absolute read counts from three technical replicates of a representative sample analyzed on nCounter. All 730 genes of the nCounter gene panel are included. B, Pairwise correlations among three technical replicates of a representative sample are estimated by $\log_2$ RPM (reads per million) values generated by RNA-Seq. All 395 genes of the Oncomine panel are included. For both technical platforms associated pairwise Pearson correlations (r) and concordance correlation coefficients CCCs for technical replicates are given. Read count distributions for each replicate are illustrated by respective histograms and density plots

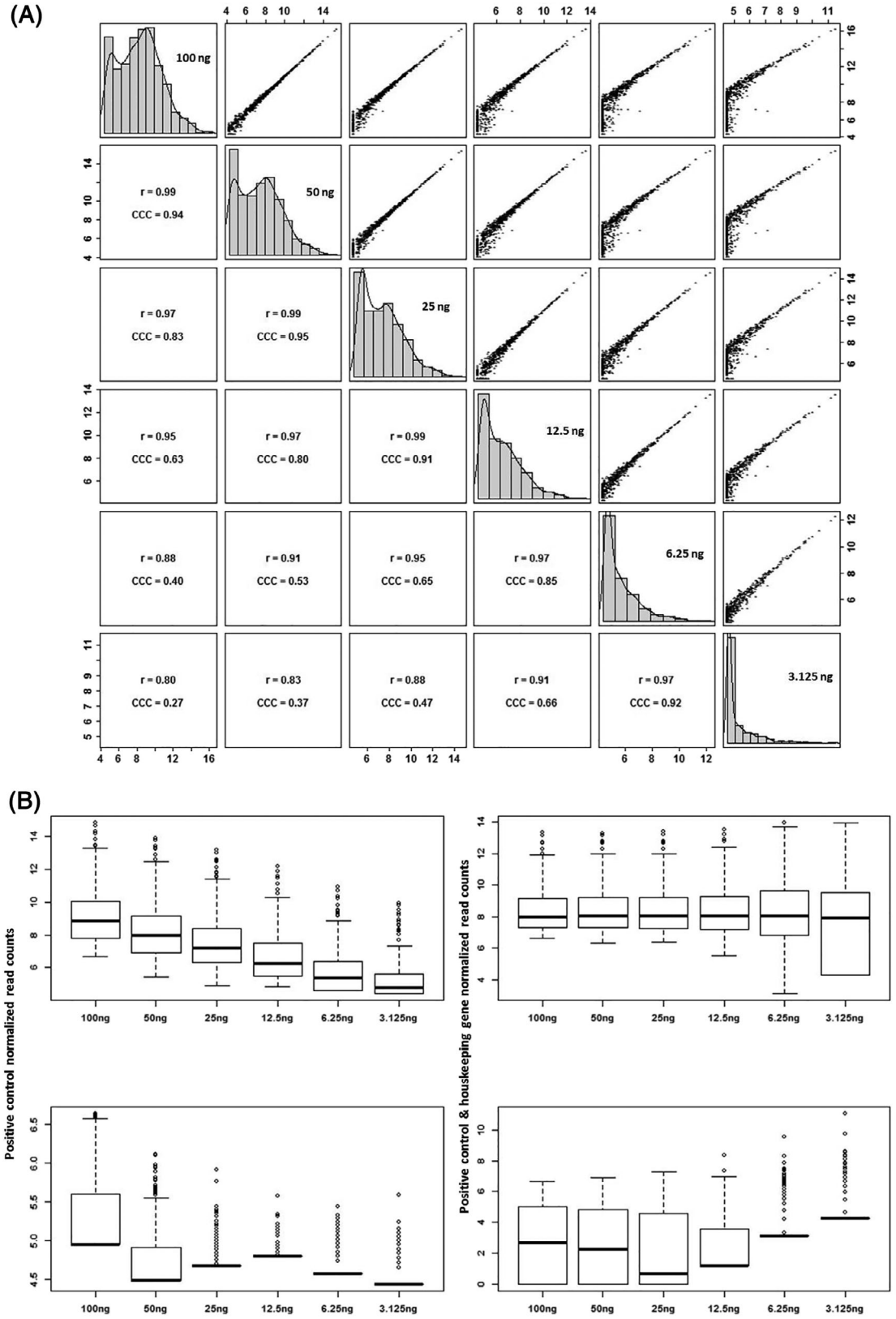

FIGURE 3 Effect of RNA input volume on nCounter performance. A, Scatter plots are generated using $\log_2$ absolute reads of 730 genes detected at different concentrations of RNA in nCounter experiments. Pearson correlations (*r*) and CCCs indicate correlations for all pairwise comparisons. B, Effect of RNA input variations on gene read counts and distribution. (First column) Box-plots illustrate distribution of gene read counts following positive control normalization. Highly (top; ≥100 read counts) and lowly (bottom; <100 read counts) expressed genes are studied separately. (Second column) Box-plots show read counts distribution of highly (top; ≥100 read counts) and lowly (bottom; <100 read counts) expressed genes after housekeeping gene normalization. Genes that are found to be highly or lowly expressed at 100 ng of RNA input are further measured at decreasing concentrations. Each box represents the interquartile range; line inside the box indicates median value. Whiskers represent minimum and maximum values. Outliers are mentioned above the whiskers

248 shared genes 226 demonstrated positive correlations, 16 had negative correlations and the associations of the remaining six genes were not measurable. The mean correlation for all genes was 0.45 with maximum and minimum correlations of 0.98 and −0.25.2, respectively. Furthermore, we analysed the association of the inter-platform correlation with the biological variance of each of the genes. Scatter

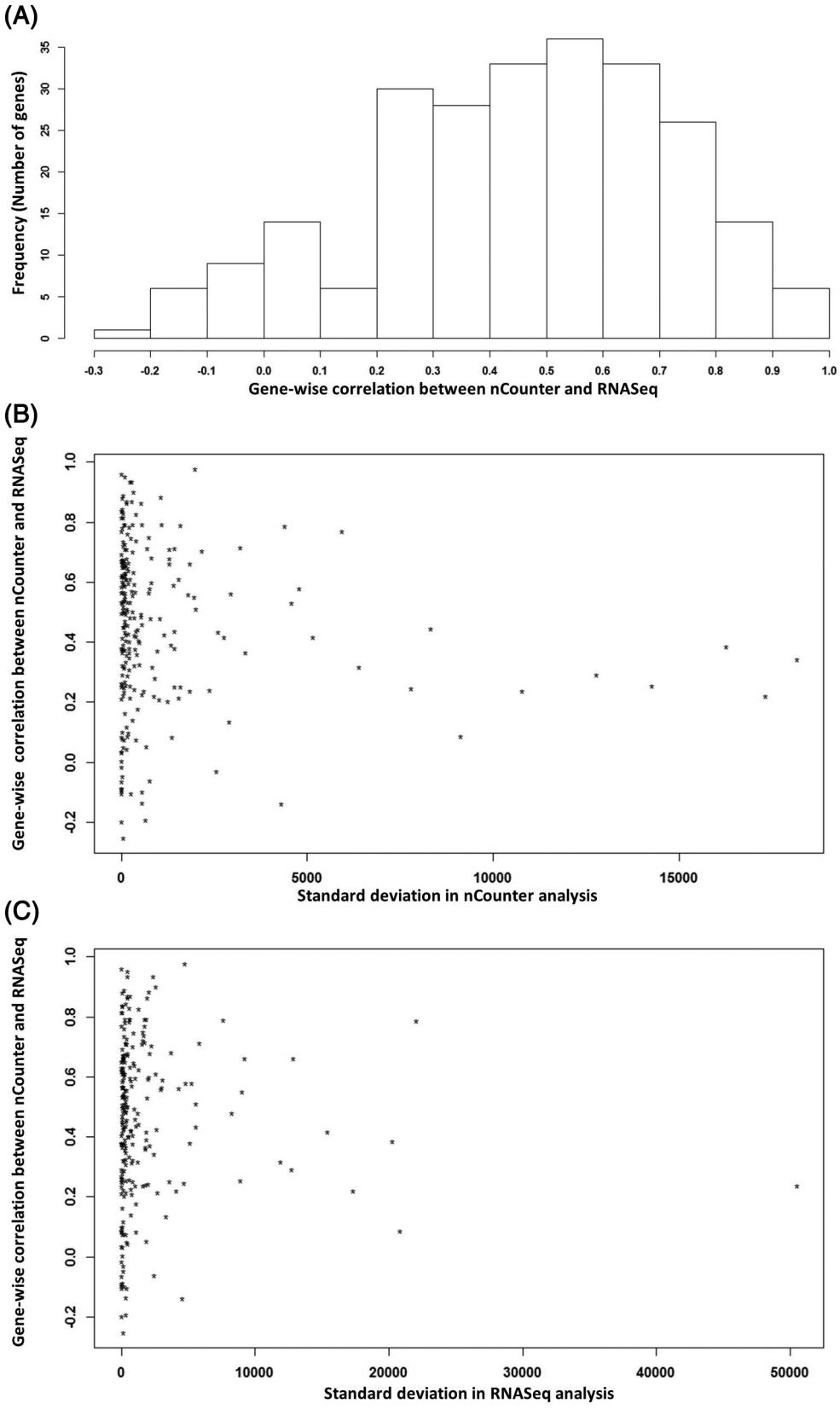

FIGURE 4 Gene-wise correlation analyses between both platforms. A, Histogram displays gene-wise correlations between nCounter and RNA-Seq. X-axis shows Pearson correlations for 248 shared genes and y-axis represents the frequency of genes for respective correlation intervals. Normalized read counts from both platforms are used for this correlation. B,C, Scatterplots display the association of gene-wise correlation between both platforms to SD of gene read counts

plots were generated by correlating the SD to respective gene wise correlation coefficients (Figure 4B, C). Scatter plots showed moderate association between biological variance and gene correlation. Taken together, gene-wise correlation analyses showed only moderate levels of correlation between the two platforms.

Next, to confirm the expression of genes estimated by nCounter and RNA-Seq, we performed experiments using qRT-PCR (Figure S4). Six genes (*CCL20*, *FAS*, *LILRB2*, *S100A8*, *CD63*, and *TGFB1*) were randomly selected and pair plots were generated by comparing normalized reads from three platforms. *CCL20* and *S100A8* showed a high correlation (Pearson correlation coefficients ≥0.92) among three platforms. However, this was not the case for all genes tested since *FAS* and *LILRB2* qRT-PCRs show high correlation only with nCounter (Pearson correlation ~0.7), but not with RNA-Seq. *CD63* and *TGFB1* had poor correlations among the three methods. Collectively, these data suggest that secondary confirmation of expression levels of individual genes, detected by nCounter and RNA-Seq, may be required.

Next, we sought to determine the concordance between nCounter and RNA-Seq. To this end, we correlated fold changes of shared genes (listed in Tables S3 and S4) generated by comparing differential gene expression between tumor samples from mal and female patients. A Spearman coefficient of 0.73 indicated strong correlation between the two platforms (Figure 5). Complementing this analysis, we also generated a Bland-Altman difference plot that compared the fold changes of the shared genes in both panels (Figure S5). The mean of differences (bias) is 0.33, indicating that the fold changes measured by RNA-Seq are slightly higher on average. Ideally, in the context of high precision of both platforms each dot displayed on the graph should rest on the *X*-axis. Still more than 95% of differences lie within the limits of agreement specified by the SD (bias ±1.96*SD).

To further investigate the inter-platform concordance, we generated lists of differentially expressed genes from the data of one of the platforms controlling the FDR at 5% and checked if the differential gene expression could be confirmed by the other platform (Tables S3 & S4). In the nCounter analysis, we detected 24 significantly differential expressed genes, while we detected 9 significantly expressed genes in the RNA-Seq analysis. Five (21%) of the genes detected by nCounter could be confirmed by RNA-Seq, five (56%) of the genes detected by RNA-Seq could be confirmed by nCounter. Thus, while nCounter was more sensitive in the detection of differential expression in the study cohort, both platforms were suitable to identify differential gene expression in cohorts of FFPE ccRCC samples.

Additionally, as visualized in Figure 6, both platforms and assays were able to reliably infer the IFNg-signature[14] from FFPE ccRCC samples. Moreover, we also observed a high concordance (*P* value less

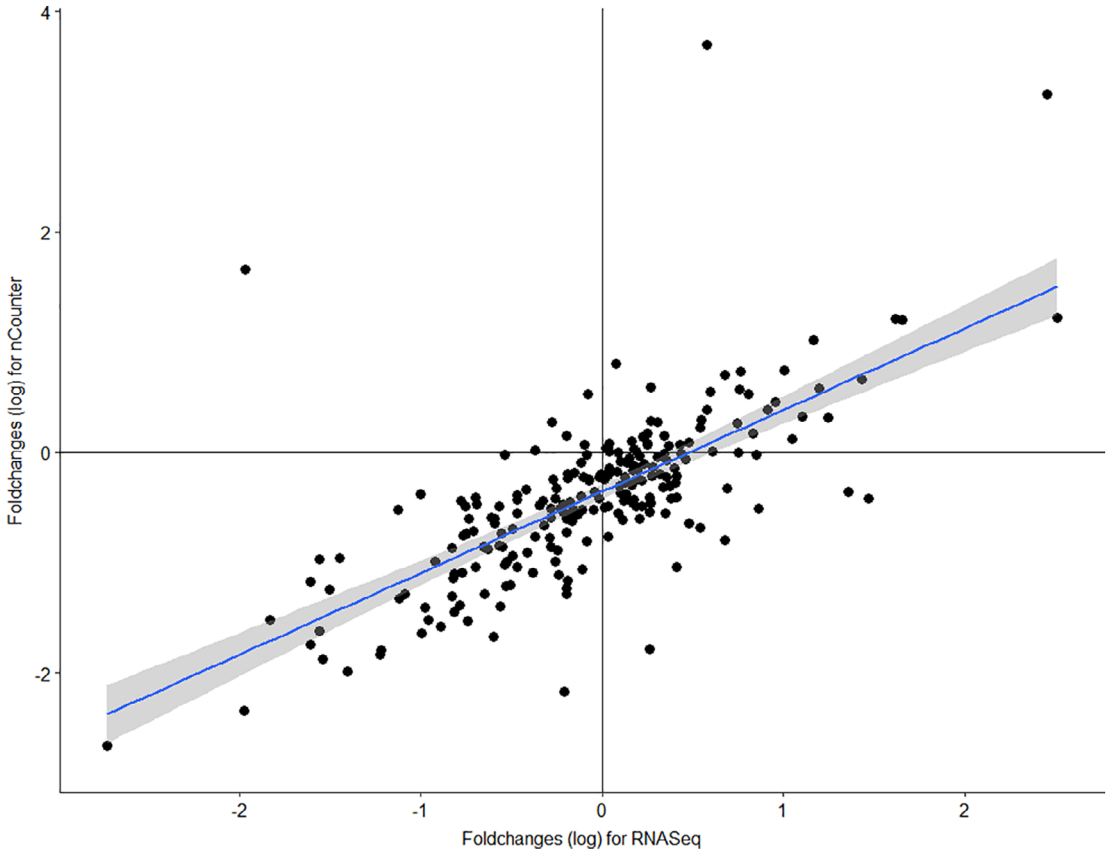

FIGURE 5 Demonstration of inter-platform concordance between nCounter and RNA-Seq. A total of 231 shared genes and their fold changes are taken into account. Scatter plot shows correlation between $\log_2$ fold changes estimated by nCounter and RNA-Seq. Correlation is quantified by Spearman coefficient (r = 0.73) [Color figure can be viewed at wileyonlinelibrary.com]

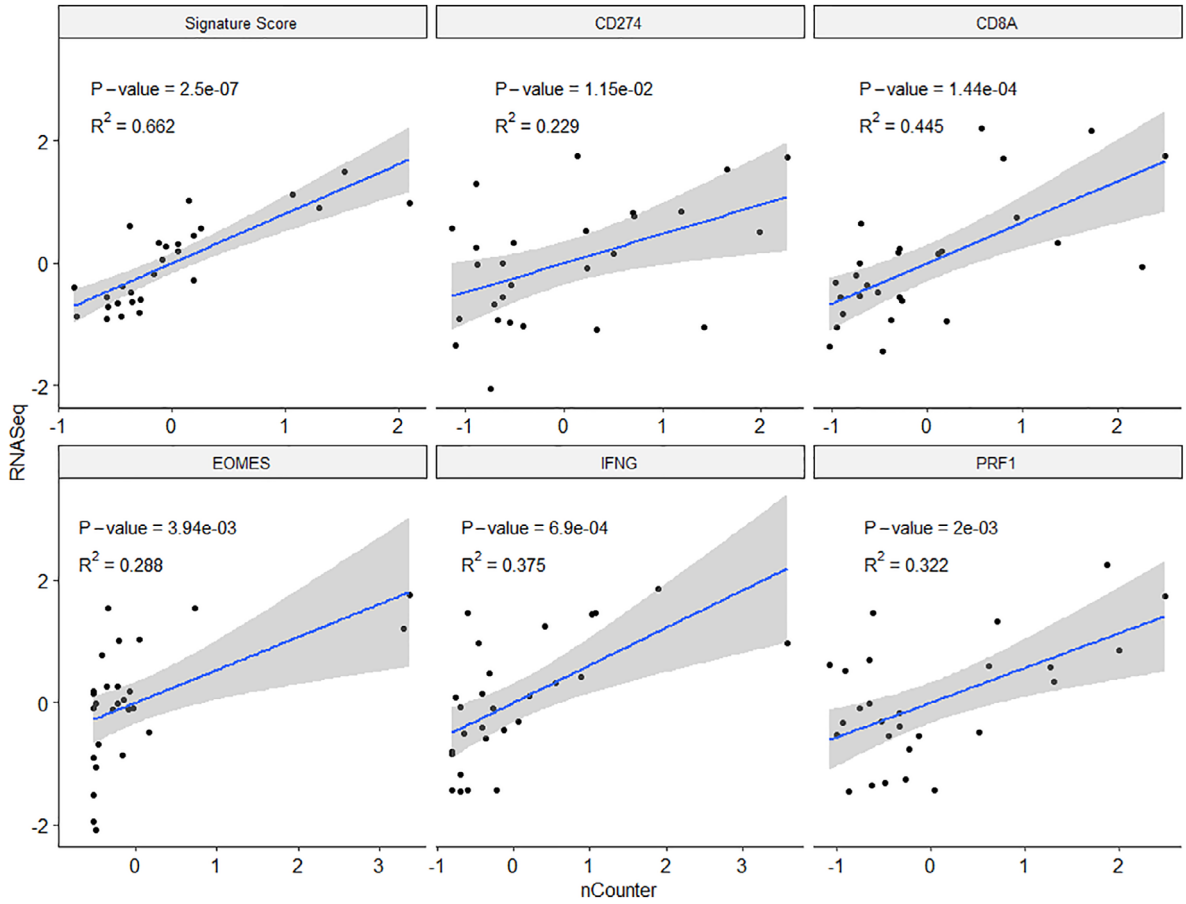

FIGURE 6 Scatterplots representing IFNg signature scores (first panel) and the expression of the individual corresponding genes. The *x*-axis displays the signature score or *z*-normalized expression of nCounter, the *y*-axis depicts corresponding values of RNA-Seq. The linear regression with corresponding smoothed confidence interval, *P* values and R^2 are added [Color figure can be viewed at wileyonlinelibrary.com]

than .015) when comparatively analyzing the expression levels of the individual genes that contribute to this signature.

4 | DISCUSSION

Until now, there are only a few studies that employed either the nCounter-based Pan Cancer Immune Profiling panel (NanoString Technologies)[25-27] or the Ion Torrent semiconductor RNA-Seq based Oncomine Immune Response Research Assay (Thermo Fisher Scientific)[28,29] in a few cancer entities including melanoma and head and neck cancer. To the best of our knowledge, our study is the first to compare these two technical platforms and to validate their performance using real-world FFPE clear cell RCC cases.

The major findings from the current study can be condensed into four aspects:

First, when read counts from three technical replicates were correlated, high intra-platform reproducibility (Pearson correlation of over 97%) was demonstrated, which is in line with previous reports.[16,30,31] High intra-platform reproducibility is considered crucial since it drastically reduces the number of technical control replicates needed, thereby enhances sample throughput and reduces costs. Additionally, concordance correlation coefficients (CCCs) revealed high accuracy and precision of both platforms.

Second, characteristics of overall expression and gene-wise correlations on both platforms were inspected by density plot and gene level correlation coefficients, respectively. Density plots disclosed a comparable bi-modal distribution that separated expressed from non-expressed genes. Besides possibly biological lower abundance of transcripts, degradation of RNA in FFPE samples might have also contributed to this phenomenon. Additionally, we observed a comparable dynamic range of both platforms, which might be an indicator of comparable sensitivity. However, we found the mode of RNA-Seq to be greater, which might be predominantly due to PCR-based target amplification prior to sequencing.

Third, we correlated gene expression levels obtained by qRT-PCR, nCounter, and RNA-Seq at the individual sample level. Comparative analyses showed heterogeneous results, which is in line with observations by others.[16,32-34] We tried to mitigate qRT-PCR variance by comparing only normalized values. Out of six genes tested, only *CCL20* and *S100A8* showed a high correlation. This might be due to apparently higher biological variance found for these two genes. Of note, whereas according to our experience nCounter based RNA profiling requires at least 50 ng of input RNA, RNA-Seq can work with

lower amounts of RNA (10 ng) and is therefore ideally suited for FFPE tissues with very limited RNA quantities.

Finally, we evaluated inter-platform concordance by correlating fold changes of all shared genes. Similar to previous studies, we first examined the cross-platform concordance between nCounter and RNA-Seq by correlation statistics.[35,36] However, correlation studies assess only linear associations between variables but not the differences, which in the present study were indispensable since we compared gene fold changes but not absolute read counts. In addition, correlation coefficients were computed assuming that expression values are not influenced by any measurement errors or sample quality, thus leading to a possible bias. To overcome this limitation, in the current study, we employed Bland-Altman difference plot,[37,38] which revealed high agreement between nCounter and RNA-Seq. In the analysis, nCounter (24 detected genes) was more sensitive than RNA-Seq (nine detected genes) in the detection of differential expression. This is possibly a consequence of the additional amplification steps inherent to the latter approach leading to larger variance of the measured expression levels and lower significance of the detected gene expression changes. Further studies in larger cohorts are warranted to confirm this observation. Besides, in order to really assess the concordance of the two technologies and protocols, samples were analyzed on both platforms in several runs on different days. Consequently, the agreement that is reflected in gene fold changes also includes possible technical and sample preparation variations. Of note, we compared the normalized data; thus, the level of agreement observed here might also partially be attributed to the normalization per se, besides inherent differences in the detection ability of both platforms. This being said, comparing un-normalized raw read counts could also lead to poor concordance because of potential batch effects in experiments and possible input variations, which could undervalue the "true" concordance. Additionally, we analyzed how well and reliably the IFNg signature reported by McDermott *et al*[14] can be read out from FFPE material using these two assays. The signature score displayed high concordance between the two the platforms suggesting that both assays could be used in routine diagnostics and clinical trials for this purpose.

In summary, we show that both platform and assays can be used for RNA profiling of FFPE clear cell renal cell cancer samples including the detection of an IFNg signature. The study also indicates that each technology has inherent advantages and disadvantages, which need to be considered prior to clinical implementation and testing of predictive biomarkers.

ACKNOWLEDGEMENTS

A. S., M. K., and V. E. conceived the idea and designed the study. M. K. and S. B. T. planned and conducted the experiments. Data acquisition, analysis, results interpretation and manuscript preparation were accomplished by M. K., S. B. T., M. A., J. B., E. R., and A. S. Statistical analyses were performed by S. B. T., E. R., and J. B., A. L. V., O. N., D. K. and I. K. participated in the planning of the experimental design and results interpretation. F. S. and C.S. selected regions of tumor in tissue sections for RNA extraction. P. S., S. D., M.J. and S. Z. kindly provided tumor biopsies and clinical data. All authors critically reviewed the manuscript and approved the final draft for the submission.

RESEARCH ETHICS

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the local Ethics Committee of the University of Heidelberg (S-864/2019).

CONFLICT OF INTEREST

AS is a consultant/advisory board member of AstraZeneca, Bristol-Myers Squibb, Bayer, Takeda, Seattle Genetics, Janssen, Pfizer, Novartis, and Thermo Fisher Scientific and received speaker's honoraria from BMS, MSD, Roche, Illumina, AstraZeneca, Novartis, and Thermo Fisher as well as research funding from Chugai and BMS. PS received advisory board honoraria from Pfizer, Roche, Novartis, and AstraZeneca as well as speaker's honoraria and research funding from Roche, AstraZeneca, and Novartis. All other authors declared no conflict of interest.

DATA AVAILABILITY STATEMENT

The generated data of this study are available from the corresponding author upon reasonable request.

ORCID

Eugen Rempel https://orcid.org/0000-0002-7408-1785
Michael Allgäuer https://orcid.org/0000-0003-4518-7887
Albrecht Stenzinger https://orcid.org/0000-0003-1001-103X

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7-30.
2. George DJ, Kaelin WG Jr. The von Hippel-Lindau protein, vascular endothelial growth factor, and kidney cancer. *N Engl J Med*. 2003;349(5):419-421.
3. Kaelin WG Jr. The von Hippel-Lindau gene, kidney cancer, and oxygen sensing. *J Am Soc Nephrol*. 2003;14(11):2703-2711.
4. Clark JI, Wong MKK, Kaufman HL, et al. Impact of sequencing targeted therapies with high-dose Interleukin-2 immunotherapy: an analysis of outcome and survival of patients with metastatic renal cell carcinoma from an on-going observational IL-2 clinical trial: PROCLAIM(SM). *Clin Genitourin Cancer*. 2017;15(1):31-41. e34.
5. Motzer RJ, Nosov D, Eisen T, et al. Tivozanib versus sorafenib as initial targeted therapy for patients with metastatic renal cell carcinoma: results from a phase III trial. *J Clin Oncol*. 2013;31(30):3791-3799.
6. Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus Ipilimumab versus Sunitinib in advanced renal-cell carcinoma. *N Engl J Med*. 2018;378(14):1277-1290.
7. Elamin YY, Rafee S, Toomey S, Hennessy BT. Immune effects of bevacizumab: killing two birds with one stone. *Cancer Microenviron*. 2015;8(1):15-21.
8. Kusmartsev S, Eruslanov E, Kubler H, et al. Oxidative stress regulates expression of VEGFR1 in myeloid cells: link to tumor-induced immune suppression in renal cell carcinoma. *J Immunol*. 2008;181(1):346-353.
9. Hodi FS, Lawrence D, Lezcano C, et al. Bevacizumab plus ipilimumab in patients with metastatic melanoma. *Cancer Immunol Res*. 2014;2(7):632-642.
10. Roland CL, Lynn KD, Toombs JE, Dineen SP, Udugamasooriya DG, Brekken RA. Cytokine levels correlate with immune cell infiltration

after anti-VEGF therapy in preclinical mouse models of breast cancer. *PLoS One*. 2009;4(11):e7669.
11. Wallin JJ, Bendell JC, Funke R, et al. Atezolizumab in combination with bevacizumab enhances antigen-specific T-cell migration in metastatic renal cell carcinoma. *Nat Commun*. 2016;7:12624.
12. Osada T, Chong G, Tansik R, et al. The effect of anti-VEGF therapy on immature myeloid cell and dendritic cells in cancer patients. *Cancer Immunol Immunother*. 2008;57(8):1115-1124.
13. Rini BI, Plimack ER, Stus V, et al. Pembrolizumab plus Axitinib versus Sunitinib for advanced renal-cell carcinoma. *N Engl J Med*. 2019;380 (12):1116-1127.
14. McDermott DF, Huseni MA, Atkins MB, et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat Med*. 2018;24(6):749-757.
15. Choueiri TK, Albiges L, Haanen JBAG, et al. Biomarker analyses from JAVELIN Renal 101: avelumab + axitinib (A+Ax) versus sunitinib (S) in advanced renal cell carcinoma (aRCC). *Journal of Clinical Oncology*. 2019;37(15_suppl):101.
16. Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. 2008;26(3):317-325.
17. Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011;475 (7356):348-352.
18. Wang H, Horbinski C, Wu H, et al. NanoStringDiff: a novel statistical method for differential expression analysis based on NanoString nCounter data. *Nucleic Acids Res*. 2016;44(20):e151.
19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15 (12):550.
20. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255-268.
21. Miron M, Woody OZ, Marcil A, Murie C, Sladek R, Nadon R. A methodology for global validation of microarray experiments. *BMC Bioinform*. 2006;7:333.
22. Bassani NP, Ambrogi F, Biganzoli EM. Assessing agreement between miRNA microarray platforms. *Microarrays (Basel)*. 2014;3(4):302-321.
23. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135-160.
24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*. 1995;57(1):289-300.
25. Bresler SC, Min L, Rodig SJ, et al. Gene expression profiling of anti-CTLA4-treated metastatic melanoma in patients with treatment-induced autoimmunity. *Lab Invest*. 2017;97(2):207-216.
26. Sato-Kaneko F, Yao S, Ahmadi A, et al. Combination immunotherapy with TLR agonists and checkpoint inhibitors suppresses head and neck cancer. *JCI Insight*. 2017;2(18):1-18.
27. Xu C, Zhang Y, Rolfe PA, et al. Combination therapy with NHS-muIL12 and Avelumab (anti-PD-L1) enhances antitumor efficacy in preclinical cancer models. *Clin Cancer Res*. 2017;23(19):5869-5880.
28. Conroy JM, Pabla S, Glenn ST, et al. Analytical validation of a next-generation sequencing assay to monitor immune responses in solid tumors. *J Mol Diagn*. 2018;20(1):95-109.
29. Paluch BE, Glenn ST, Conroy JM, et al. Robust detection of immune transcripts in FFPE samples using targeted RNA sequencing. *Oncotarget*. 2017;8(2):3197-3205.
30. Veldman-Jones MH, Brant R, Rooney C, et al. Evaluating robustness and sensitivity of the NanoString technologies nCounter platform to enable multiplexed gene expression analysis of clinical samples. *Cancer Res*. 2015;75(13):2587-2593.
31. Veldman-Jones MH, Lai Z, Wappett M, et al. Reproducible, quantitative, and flexible molecular subtyping of clinical DLBCL samples using the NanoString nCounter system. *Clin Cancer Res*. 2015;21(10):2367-2378.
32. Fang Z, Cui X. Design and validation issues in RNA-seq experiments. *Brief Bioinform*. 2011;12(3):280-287.
33. Shi Y, He M. Differential gene expression identified by RNA-Seq and qPCR in two sizes of pearl oyster (Pinctada fucata). *Gene*. 2014;538 (2):313-322.
34. Wu AR, Neff NF, Kalisky T, et al. Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods*. 2014;11(1):41-46.
35. Forreryd A, Johansson H, Albrekt AS, Lindstedt M. Evaluation of high throughput gene expression platforms using a genomic biomarker signature for prediction of skin sensitization. *BMC Genomics*. 2014;15:379.
36. Kelly AD, Hill KE, Correll M, et al. Next-generation sequencing and microarray-based interrogation of microRNAs from formalin-fixed, paraffin-embedded tissue: preliminary assessment of cross-platform concordance. *Genomics*. 2013;102(1):8-14.
37. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1 (8476):307-310.
38. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 1995;346(8982):1085-1087.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Talla SB, Rempel E, Endris V, et al. Immuno-oncology gene expression profiling of formalin-fixed and paraffin-embedded clear cell renal cell carcinoma: Performance comparison of the NanoString nCounter technology with targeted RNA sequencing. *Genes Chromosomes Cancer*. 2020;59:406–416. https://doi.org/10.1002/gcc.22843