Review

Cancer Research

Evaluating Robustness and Sensitivity of the NanoString Technologies nCounter Platform to Enable Multiplexed Gene Expression Analysis of Clinical Samples

Margaret H. Veldman-Jones[1], Roz Brant[1], Claire Rooney[1], Catherine Geh[1], Hollie Emery[1], Chris G. Harbron[1], Mark Wappett[1], Alan Sharpe[1], Michael Dymond[1], J. Carl Barrett[2], Elizabeth A. Harrington[1], and Gayle Marshall[1]

Abstract

Analysis of clinical trial specimens such as formalin-fixed paraffin-embedded (FFPE) tissue for molecular mechanisms of disease progression or drug response is often challenging and limited to a few markers at a time. This has led to the increasing importance of highly multiplexed assays that enable profiling of many biomarkers within a single assay. Methods for gene expression analysis have undergone major advances in biomedical research, but obtaining a robust dataset from low-quality RNA samples, such as those isolated from FFPE tissue, remains a challenge. Here, we provide a detailed evaluation of the NanoString Technologies nCounter platform, which provides a direct digital readout of up to 800 mRNA targets simultaneously. We tested this system by examining a broad set of human clinical tissues for a range of technical variables, including sensitivity and limit of detection to varying RNA quantity and quality, reagent performance over time, variability between instruments, the impact of the number of fields of view sampled, and differences between probe sequence locations and overlapping genes across CodeSets. This study demonstrates that Nanostring offers several key advantages, including sensitivity, reproducibility, technical robustness, and utility for clinical application. *Cancer Res; 75(13); 2587–93. ©2015 AACR.*

Introduction

A key constraint of translational research within a clinical trial setting is that there is often a limited amount of tissue from which to carry out biomarker analyses. Further, this tissue is frequently archival and stored in formalin-fixed paraffin embedded (FFPE) blocks. Traditional methods of gene expression analysis have limitations for clinical application. For example, RT-PCR measures the expression of one gene at a time, whereas multiplex expression profiling techniques such as microarrays, covering many thousands of transcripts, are often expensive and lack flexibility and reproducibility when evaluating low-quality RNA samples such as those from FFPE (1–4). Platforms that enable multiplexed analysis of biomarkers from limited amounts of poor-quality material are therefore very attractive.

The Nanostring Technologies nCounter platform is a relatively new technology and has been used within various clinical and research applications (2–13). The automated nCounter platform hybridizes fluorescent barcodes directly to specific nucleic acid sequences, allowing for the nonamplified measurement of up to 800 targets within one sample (1, 14). A number of papers have shown that the Nanostring platform is comparable with other technologies (11, 12, 14); however, these studies have not assessed the limit of detection (LOD) and the robustness of the platform, which is essential when evaluating precious clinical trial samples.

We present a comprehensive evaluation of the nCounter platform analyzing all technical variables to establish the utility of this platform for multiplexed gene expression analysis of clinical tumor samples. We investigated platform robustness with a range of clinical test samples, assessed a number of protocol variables, and evaluated the sensitivity and specificity of the platform when using low input and/or quality RNA test samples. We analyzed a wide range of tumor tissue types, including diffuse large B-cell lymphoma (DLBCL), gastric, lung, breast, pancreas, melanoma, xenograft, and blood with similar observations. We find the nCounter platform favorable over other techniques based upon sensitivity, technical reproducibility, robustness, ease of use, hands-on analysis time, and utility for clinical application.

Materials and Methods

nCounter specifics

Probe sequences were custom designed and manufactured by NanoString, unless otherwise specified. Gene names were

[1]AstraZeneca Oncology and Discovery Sciences Innovative Medicines, Alderley Park, Macclesfield, United Kingdom. [2]AstraZeneca, Waltham, Massachusetts.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

E.A. Harrington and G. Marshall contributed equally to this article.

Corresponding Author: Margaret H. Veldman-Jones, AstraZeneca, Oncology Innovative Medicines, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, United Kingdom. Phone: 44-1625-233767; Fax: 44-1625-514289; E-mail: Margaret.Veldmanjones@astrazeneca.com

doi: 10.1158/0008-5472.CAN-15-0262

anonymized due to confidentiality restraints and lack of relevance to the purpose of this article. Each CodeSet includes a number of housekeeping genes to correct for RNA input amount and/or quality differences. Housekeeping genes were selected from publicly available databases based on stability and detectable expression levels across the tissue type of interest. Generated data went through an internal QC process and choice of housekeeping genes refined if necessary. While one CodeSet contained four housekeeping genes, all others contained a minimum of eight.

Protocol was followed according to standard nCounter instructions, with the exception of the creation of a master-mix containing hybridization buffer, Reporter ProbeSet and Capture ProbeSet (volume:volume ratio of 1:1:0.5), of which 25 μL was added to 5 μL target RNA. A pre-prepared master-mix resulted in a more efficient workflow when setting up multiple cartridges. No detrimental effect was observed on the data using this amendment in the protocol instructions as assessed over a large number of CodeSets and sample numbers (data not shown).

GEN2 Prep Station incubation time was set at the higher sensitivity setting (3 hours) and 280 fields of view (FOV) were routinely captured, unless otherwise noted.

All assays described in this article are currently identified as research use only.

Cell lines and reference RNA

The DLBCL cell line SU-DHL-10 was purchased from Cambridge Enterprise (DSMZ) in 2010, and HBL-1 was obtained from Professor Masafumi Abe (Fukushima Medical University, Fukushima, Japan), under license from Tokyo Medical and Dental University, Tokyo, Japan. NCI-H520 cells were obtained from the ATCC. All cell lines were tested for authenticity by genotyping prior to use. Cells were maintained in 5% CO_2 at 37°C and cultured in RPMI supplemented with 10% to 15% FBS and 1% L-glutamine. Universal Human Reference RNA (UHR) was purchased from Stratagene (catalog #75000-41).

Clinical samples

Commercially available DLBCL tissue samples were purchased from OriGene. A number of matching samples were included and designated as Fresh (RNA extracted fresh by OriGene), frozen (tissue snap frozen in liquid nitrogen) and FFPE (tissue fixed in formalin).

Other clinical samples (prostate, gastric, and lung FFPE) were purchased from Avaden Biosciences, Asterand, acquired through our alliance with Manchester Cancer Research Centre, UK, or procured through our local UK BioBank.

All required consents for these exploratory analyses were obtained for all clinical material samples. Prior to processing, each sample was reviewed by an internal certified pathologist to confirm disease diagnoses and to verify tumor content. Macrodissection of FFPE tissues ensured that only tumor regions were analyzed in downstream applications.

RNA extraction

Cell lines and frozen tissues were extracted using the RNeasy kit (Qiagen), and tissues were homogenized using the Tissue Lyser II system (Qiagen). FFPE tissues were extracted using the RNeasy FFPE extraction kit (Qiagen). RNA quantity and quality were assessed by Nanodrop 2000 and RNA 6000 Nano Kit (Agilent), respectively. Protocols were followed according to the manufacturer's instructions. Choice of extraction kit was extensively evaluated separately (data not shown).

Analysis

nCounter data were normalized through an internally developed Pipeline Pilot Tool (NAPPA, publicly available on the Comprehensive R Archive Network, CRAN, Harbron & Wappett (2014) R package: NAPPA http://CRAN.R-project.org/package=NAPPA).

In brief, data were $\log_2$ transformed after being normalized in two steps: raw NanoString counts were first background adjusted with a Truncated Poisson correction using internal negative controls followed by a technical normalization using internal positive controls. Data were then corrected for input amount variation through a Sigmoid shrunken slope normalization step using the mean expression of housekeeping genes. A transcript was designated as not detected if the raw count was below the average of the 8 internal negative control raw counts plus 2 SDs reflecting approximately a 95% confidence interval. Transcripts below the LOD were not adjusted to background but kept their original values. We included data points below the LOD in most of our plots but omitted them from the analysis where appropriate, because we see merit in using these data points for certain data analyses (e.g., gene signature scores).

Results

Evaluation of platform-associated variables

The nCounter platform consists of two instruments, the Prep Station, a liquid handling robot that performs the purification of the hybridized complexes and their immobilization onto the surface of a cartridge, and the Digital Analyzer (DA), a scanner that identifies and counts the barcodes captured for each sample. The Prep Station processes 12 samples per run and loads them into a 12-lane cartridge. The DA can hold 6 cartridges and measures the number of FOV at four settings: 25, 100, 280, and 555. The total hands-on time for 12 reaction samples is approximately 15 minutes, and the remainder of the protocol is a fully automated process. In order to apply optimal settings in evaluating test sample variables, we started with testing variability inherent to the platform.

FOV settings. First, we investigated which one of the four FOV settings generates the most reproducible data. Supplementary Fig. S1A shows the mean SD calculated across three, independent, fresh clinical DLBCL samples. Genes were grouped arbitrarily according to their expression level into three categories containing equal numbers of high, medium, or low expressing genes. The same cartridge was rescanned at 25, 100, 280, and 555 FOVs on the DA. As expected, variability is highest for the lowest expressed genes (<1 SD) and reduced systematically with an increasing number of FOVs captured. Only a minimal improvement was observed when reading samples at the highest setting of 555 FOV compared with 280 FOV. Therefore, 280 FOV was chosen as the optimal setting for subsequent experiments due to the added benefit of reduced scanning time compared with 555 FOV.

Cartridge lanes and digital analyzer slots. Fading of signal due to scanning at 280 FOV reduced raw data counts on average by 26.8% per gene after 10 scans. The internal top positive control

top also reflected this with a decreased raw count of 2.5% on average per scan (Supplementary Fig. S1B). Although the normalized $\log_2$ values were minimally affected, we wanted to avoid any impact of this small but significant variable on the outcome of the data. Therefore, cartridge lane-to-lane variability and DA slot-to-slot variability were determined at the lowest FOV setting of 25. Supplementary Fig. S1C shows the variability of a representative high and low expressing gene from a Universal Human Reference RNA (UHR) sample replicated across the 12 lanes of a cartridge and read in the 6 slots of the DA. The variability of a high expressed gene was extremely low, and although variability increased with a lower expressed gene, it remained low ($\log_2$ SD < 0.5) between cartridge lanes and DA slots. It was concluded that using different lanes on a cartridge and different slots on the DA does not introduce significant bias to the data.

Prep station. In order to meet clinical trial throughput demand, we investigated whether prepping cartridges across two different GEN2 Prep stations would add variability to the data. Based on independent replicates of a UHR sample across two GEN2 Prep stations, the coefficient of variation (CoV) was determined to be lower between Prep stations (4.67%) than within Prep stations (10.01%). There was also no evidence of a systematic trend when calculating the mean SD between all replicates (Supplementary Table S1). Taken together, these data suggest that no significant variability is introduced by using two different GEN2 Prep stations within the same laboratory, and daily throughput can therefore be increased by using both Prep stations simultaneously. However, random sample distribution across cartridges is the best practice to reduce any sources of systematic variability.

Technical reproducibility was also compared between a GEN1 (FOV setting 600) and GEN2 (FOV setting 280) Prep Station based in two different laboratories and showed good correlation of results (Supplementary Fig. S1D).

Evaluation of sample-associated variables

Having established the robustness of the nCounter and identified optimal conditions, we investigated data reproducibility and platform sensitivity in a range of tumor samples.

Technical and biologic replicates. Technical variability between identical samples was assessed on two main variables: hybridization timing (replicates processed on the same day on the same cartridge vs. different days) and sample quality (FFPE tissue, poor; fresh-frozen tissue, good). Matched DLBCL samples were randomized on cartridges across the two variables. Figure 1A illustrates that no difference in technical reproducibility could be detected between hybridization timing or samples of different quality as measured by the absolute difference between replicate samples (hybridization timings grouped together, no difference seen). On average, the R^2 score was greater than 0.98 for all variables tested (not shown).

Biologic sample replicates were assessed using a gastric tissue sample set. RNA was extracted from two sections that were 5 to 10 μm apart within the tumor block, and NanoString gene expression profiles were generated. Overall, good correlation was seen between sections from the same tumor with a mean intrasample CoV of 6.47%, which was lower than the mean intersample CoV of 12.09%. R^2 scores ranged from 0.33 to 0.70 for intersample correlations and 0.59 to 0.93 for intrasample correlations (Table 1), demonstrating the ability of NanoString to discern differences in gene expression between different tumor samples, beyond those arising from biologic variability within a sample.

As expected, these results show that, in general, biologic variability between samples is far greater than technical variability.

Sample quality. One of the key advantages of the NanoString technology is the absence of an amplification step and direct measurement of target molecules, thus avoiding any bias. This makes it an attractive platform to evaluate degraded mRNA such as that extracted from FFPE samples, which are often obtained from clinical trial investigations. Therefore, mRNA expression levels between matching fresh, frozen, and FFPE sample types were compared.

Five matching DLBCL tissue samples were analyzed, and only a minimal difference was observed between matching fresh and frozen samples compared with matching fresh and FFPE samples based on the absolute difference of the replicates (Fig. 1B). Clustering analysis confirmed that fresh and frozen tissues correlated better than fresh and FFPE tissues for 4 of 5 matching samples. All matching tissues clustered together as expected (Supplementary Fig. S2). Encouragingly, the difference in matching fresh and frozen replicates was comparable with that seen with technical replicates. The quantitative correlation was also very robust (11), demonstrating that the nCounter platform can generate high-quality data on FFPE material to enable clinically relevant mRNA gene expression studies. Numerous other publications have independently confirmed this conclusion (1–4).

Sample quantity. Next, we compared the optimal input amount of fresh versus FFPE tissue in a titration experiment and showed that equivalent data could be generated from 100-ng good-quality mRNA (RIN >8.6) versus 400-ng poor quality RNA (RIN < 2.4) based on the detection level (Fig. 2A). Following on from this, we determined the minimum input amount of FFPE material required to get the most data from clinical samples. Based on the previous observation, RNA input was started at the more practical amount of 100 ng titrating down in 2-fold steps. As shown in Fig. 2A, this starting amount is only marginally compromised for genes at the lowest expression level compared with a fresh sample (on average 5% increase in genes below the LOD). Figure 2B (top) illustrates the number of genes that fall below the LOD relative to the 100-ng input sample. Although genes at the lowest expression level were most affected, it was surprising to see that even at an input RNA amount of 6.25 ng, the more highly expressed genes remained unaffected and correlations were still good (Fig. 2B, bottom). Depending on the expression level of the genes of interest present in the CodeSet, the acceptable amount of input RNA for a FFPE sample can be much lower than 100 ng based on these results.

Another way of improving results for samples with a compromised quantity and/or quality is to concentrate them (e.g., with vacuum, precipitation, or column technology) in order to increase the input amount. However, this process often results in a total yield reduction without improving the final concentration (data not shown), which is not acceptable when dealing with precious clinical trial samples. We investigated whether we could increase the fixed input volume of 5 μL up to 15 μL using

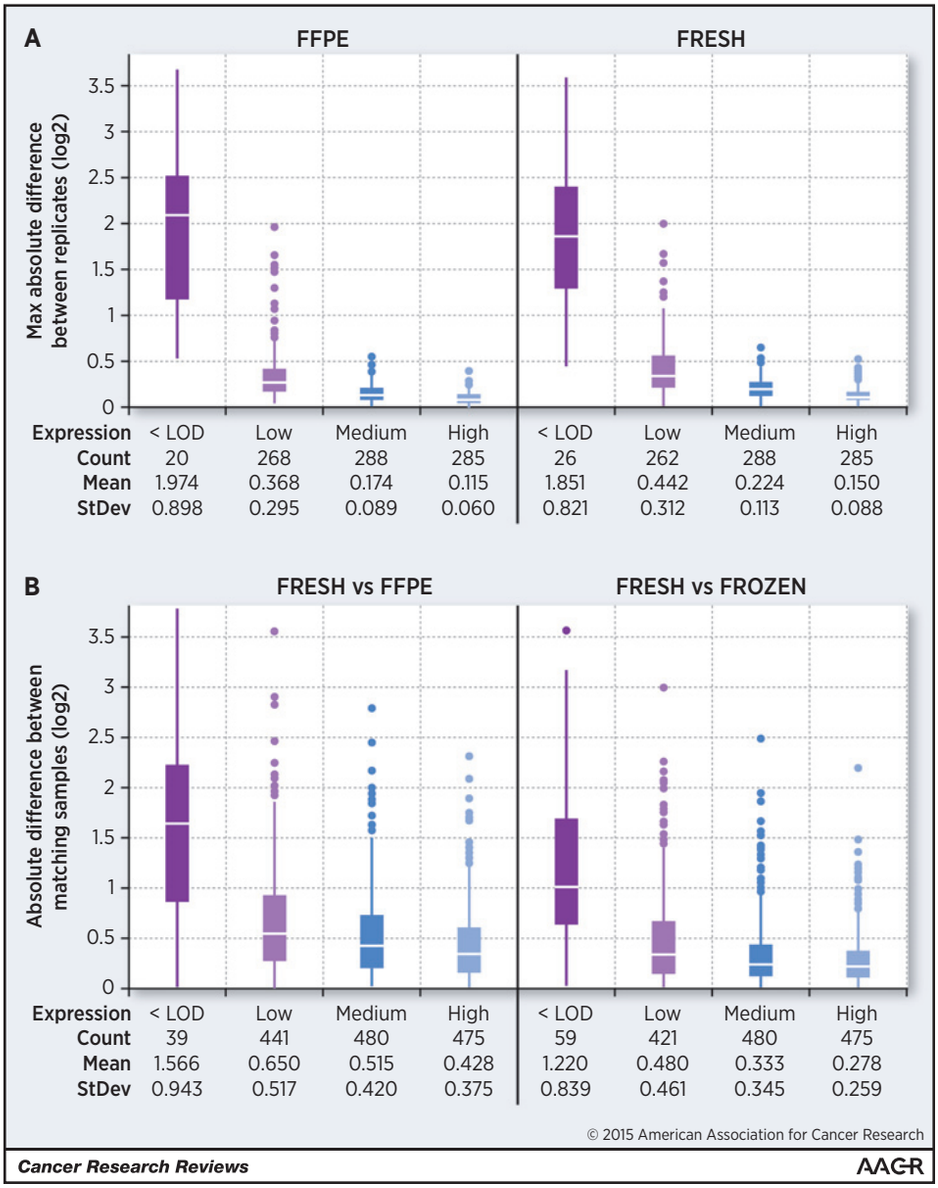

Figure 1.
A, box plots showing technical replicate difference for samples of low quality (FFPE) and high quality (FRESH). A total of 287 genes were ranked on expression level and divided in equal groups of low, medium, or high expression for both sample categories. Input RNA FFPE, 400 ng; fresh, 100 ng clinical DLBCL matching samples. Maximum absolute difference (of normalized $\log_2$ count) between four technical replicates calculated; summary of three samples shown for each expression level category. Gene designated below LOD if 3 or 4 samples out of four replicates were not detected. Data were normalized against 33 housekeeping genes. Expression, expression level of genes; count, number of genes; mean, mean absolute $\log_2$ difference. B, box plots showing comparison of matching clinical DLBCL fresh vs. frozen and fresh vs FFPE tissues. A total of 287 genes were ranked on expression level and divided in equal groups of low, medium, or high expression for both categories. Input RNA FFPE, 400 ng; fresh and frozen, 100 ng. The absolute difference (of normalized $\log_2$ count) between matching samples was calculated; summary of 5 samples is shown for each expression level category. Gene designated below LOD if one or both replicates were not detected. Data were normalized against 33 housekeeping genes. Expression, expression level of genes; count, number of genes; mean, mean absolute $\log_2$ difference.

1 μL increments (30 μL up to 45 μL final hybridization volume) in order to increase the input amount and showed no detrimental effect on correlation (Supplementary Fig. S3). Having repeated this result with another CodeSet (consisting of 770 genes), we feel confident the input volume can be increased from 5 μL up to 10 μL, thus allowing a 2-fold increase in the input RNA amount potentially increasing the number of genes above the LOD.

Other parameters evaluated

Cartridge stability. To determine the length of time a prepped cartridge can be stored at 4°C, we rescanned a cartridge four times over a 9-month period at 280 FOV. The largest effect on raw endogenous gene counts (down 67%) and captured FOVs (down 38%) occurred in the first 4 months of storage (Supplementary Fig. S4A), caused most likely by damaged reporters and slight warping of the cartridge. The effect stabilized over the next 5 months tested. The NAPPA normalization tool corrects for the number of FOV captured, and although the effect seemed significant on the raw data counts, normalization all but erased this effect with an R^2 score of 0.965 between data obtained on the first read (month 0) versus the last read (month 9; Supplementary Fig. S4B). Although good-quality data can be obtained from a cartridge stored for at least 9 months at 4°C, optimal data with minimal loss of low abundance genes are obtained in the first 3 months after cartridge preparation.

CodeSet stability. CodeSet stability beyond its expiry date was investigated next. Supplementary Fig. S5A shows the correlation plot of 6 independent fresh DLBCL clinical tissue samples tested with a 30 gene CodeSet 3.5 years after its manufacture. Some variability was noted with genes under or close to the LOD (addressed in more detail in "Technical and biologic replicates").

Table 1. Biologic replicate R^2 scores for two adjacent samples for each of 6 gastric FFPE samples with a CodeSet consisting of 498 genes

R^2	S1 (A)	S2 (A)	S3 (A)	S4 (A)	S5 (A)	S6 (A)
S1 (B)	0.93	0.42	0.44	0.67	0.60	0.61
S2 (B)	0.35	0.93	0.35	0.46	0.36	0.44
S3 (B)	0.55	0.54	0.59	0.61	0.46	0.53
S4 (B)	0.69	0.41	0.46	0.84	0.70	0.47
S5 (B)	0.55	0.33	0.34	0.67	0.70	0.36
S6 (B)	0.60	0.52	0.45	0.51	0.41	0.88

NOTE: Input RNA amount was 100 ng; genes below LOD were omitted from analysis. Values in black box indicate intrapatient correlation; other values indicate interpatient correlation.
Abbreviations: S1 (A), sample 1 block A; S1 (B), sample 1 block B, etc.

This limited dataset suggests that a CodeSet is stable for at least 1.5 years after its stated expiry date.

Freeze–thawing of CodeSet. The number of freeze/thaw cycles a CodeSet could undergo before data were adversely affected was also assessed. An identical UHR sample was used in triplicate (from aliquots) for each freeze/thaw cycle. A CodeSet consisting of 463 genes was allowed to thaw to room temperature for 20 minutes, used, and refrozen at −80°C for 1 hour before commencing the next freeze/thaw cycle. Defrosting the original CodeSet aliquot was designated the first freeze/thaw cycle. The total difference in hybridization time between the first and fourth freeze/thaw cycles was 5 hours (range, 22–27 hours). Although the effect of a difference in total hybridization time has not been formally assessed, it is not likely to have a significant impact on the data as shown earlier when assessing variability between identical technical replicate samples. The results showed that, on average, a reduction of just under 10% in raw data gene counts was seen after 4 freeze/thaw cycles (increasing from 1.2% to 2.9% to 9.3% after each freeze/thaw cycle). Low abundance genes are most affected and as a consequence some fall below the LOD after four freeze/thaw cycles, but no impact on overall correlation was seen after normalization with an R^2 score of 0.99 (Supplementary Fig. S5B). After repeating the experiment using a different codeset, we concluded that at least one freeze/thaw cycle of a CodeSet is acceptable.

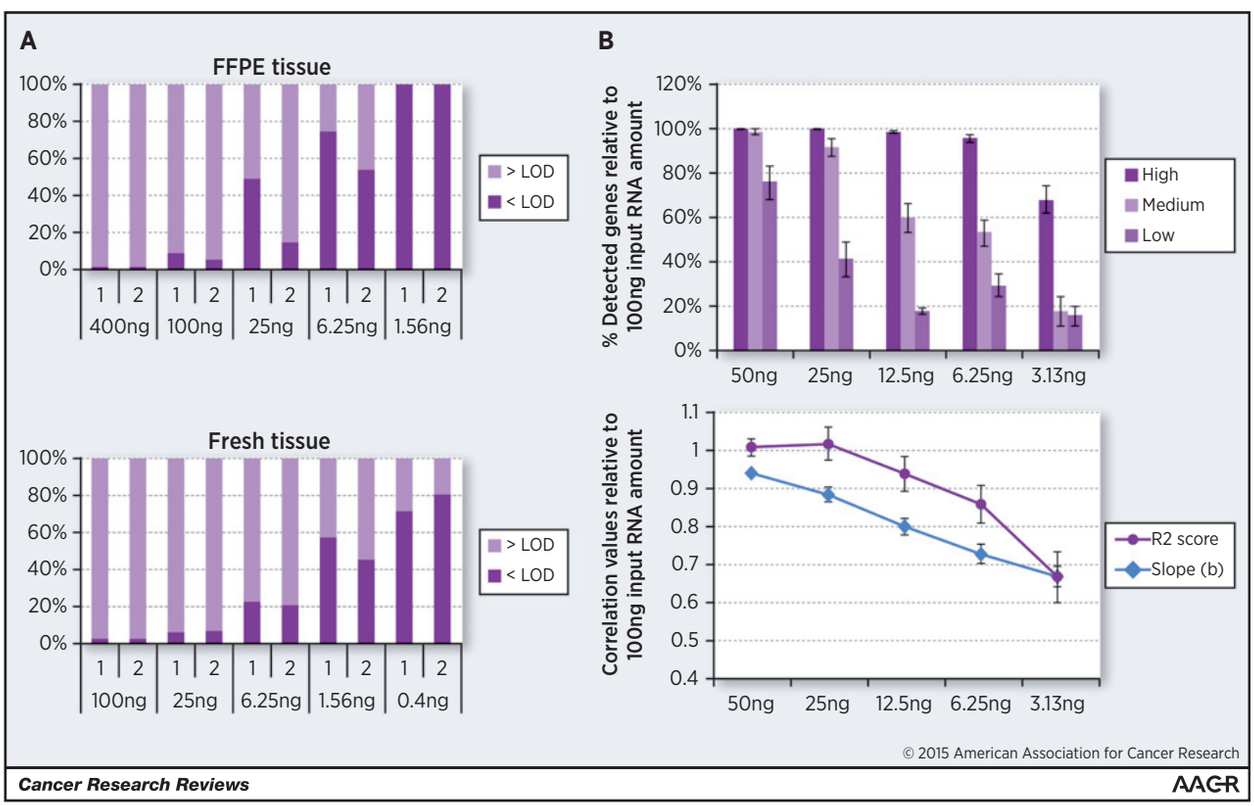

Figure 2.
A, titration of RNA input amount using two matching pairs of FFPE (top) and fresh (bottom) DLBCL tissues. The percentage of genes above the LOD versus below the LOD is illustrated from a CodeSet consisting of 320 genes, including 33 housekeeping genes. B, titration of RNA input amount down in 2-fold steps from 100 ng using four FFPE lung tissue samples (top). The percentage of detected genes relative to the highest input RNA amount of 100 ng is illustrated for genes grouped into equal numbers of high, medium, and low expression levels with a CodeSet consisting of 468 genes, including 8 housekeeping genes. Gene expression level ranges ($\log_2$): high, 8.6–13.7; medium, 7.1–8.6; low, <4.5–7.0. Correlation to 100-ng input amount assessed with mean R^2 scores and slope factors (bottom). Values based on a mean of four replicate lung FFPE samples.

Overlapping genes in different CodeSets. When data were compared from identical samples across different CodeSets (regardless of whether these were identical or not) containing overlapping target genes, we noticed a systematic shift of data in the same direction and some loss in correlation for the majority of genes. Not surprisingly, variability was greatest for genes below the LOD and for genes with nonidentical probe sequences (Supplementary Fig. S6). Unusually one gene with identical probe sequences was also identified as an outlier in the majority of samples investigated. This latter observation may warrant designing multiple probes to genes of high interest, although the shift in data and loss in correlation can be minimized by calibrating using one or more reference samples replicated against both CodeSets and directly comparing fold changes rather than absolute expression values.

Discussion

A number of studies performed to test nCounter platform– and test sample–associated variables have shown excellent robustness and reproducibility. The sensitivity of target detection remained impressive even at very low input RNA amounts and without the need for an amplification step. In instances where technical variables affected raw counts, these were negated by the NAPPA normalization tool. The main consequence of compromised sample quality, quantity, or processes was that lower expressed genes fell below the LOD, but even this was very minor in most cases.

We found the practical aspect of using the nCounter platform favorable over any other technique used (11) with minimal time needed and reduced margin for user error due to few handling steps during the preparation of reactions. The processing of cartridges with a maximum of 36 samples processed per Prep Station in one (10 hours) working day, improved to 72 samples a day when using two Prep Stations concurrently, is still a bottleneck. A method for increasing throughput utilizing a new sample plexing CodeSet configuration currently in development at NanoString is expected to improve this.

Careful CodeSet calibration consideration is necessary when test samples demand the use of more than one batch of CodeSet. Although we have shown that CodeSet batch effects are minimal, when directly comparing absolute gene expression values, the choice of your calibration sample(s) and calculation method for determining gene correction factors can influence the outcome of the data.

Designing probe sequences demands careful consideration in order to obtain the most accurate data. NanoString's default probes are designed to regions that favor hybridization efficiency rates, but it is important to confirm that the target region is expressed in the tissue, disease (preclinical and clinical), and isoform of interest, avoiding highly variable untranslated regions (often not expressed in cancer) and pseudogenes.

Other applications of the nCounter technology not covered here that warrant further evaluation are gene fusions (3, 12), miRNA (15), CNV (16), and protein analysis (17). Overall, this study demonstrates that nCounter technology offers several key advantages, including sensitivity, technical reproducibility, and robustness for analysis of FFPE samples and strong evidence of utility of the nCounter for mRNA gene expression analysis in clinical tissue.

Disclosure of Potential Conflicts of Interest

C. Harbron is principal statistician at Roche. No potential conflicts of interest were disclosed by the other authors.

Acknowledgments

The authors thank Tom Liptrot, Paul Fisher, Shaun Grosskurth, Dan Stetson, Brian Dougherty, Emma Jones, Caroline Seiler, Philippa Webster (NanoString), and Christina Bailey (NanoString) for input in technical discussions.

nCounter and NanoString Technologies are registered trademarks of NanoString Technologies, Inc.

Grant Support

This study was entirely funded by AstraZeneca.

Received January 27, 2015; revised March 19, 2015; accepted March 19, 2015; published OnlineFirst June 11, 2015.

References

1. Reis PP, Waldron L, Goswami RS, Xu W, Xuan Y, Perez-Ordonez B, et al. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. BMC Biotechnol 2011;11:46.
2. Saba NF, Wilson M, Doho G, DaSilva J, Benjamin Isett R, Newman S, et al. Mutation and transcriptional profiling of formalin-fixed paraffin embedded specimens as companion methods to immunohistochemistry for determining therapeutic targets in oropharyngeal squamous cell carcinoma (OPSCC): a pilot of proof of principle. Head Neck Pathol 2014 Sep 19. [Epub ahead of print].
3. Scott DW, Chan FC, Hong F, Rogic S, Tan KL, Meissner B, et al. Gene expression-based model using formalin-fixed paraffin-embedded biopsies predicts overall survival in advanced-stage classical Hodgkin lymphoma. J Clin Oncol 2013;31:692–700.
4. Norton N, Sun Z, Asmann YW, Serie DJ, Necela BM, Bhagwate A, et al. Gene expression, single nucleotide variant and fusion transcript discovery in archival material from breast tumors. PLoS One 2013;8:e81925.
5. Nielsen T, Wallden B, Schaper C, Ferree S, Liu S, Gao D, et al. Analytical validation of the PAM50-based Prosigna breast cancer prognostic gene signature assay and nCounter analysis system using formalin-fixed paraffin-embedded breast tumor specimens. BMC Cancer 2014;14:177.
6. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW, et al. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. J Clin Oncol 2013;31:2783–90.
7. Martin JW, Chilton-MacNeill S, Koti M, van Wijnen AJ, Squire JA, Zielenska M. Digital expression profiling identifies RUNX2, CDC5L, MDM2, RECQL4, and CDK4 as potential predictive biomarkers for neo-adjuvant chemotherapy response in paediatric osteosarcoma. PLoS One 2014;9:e95843.
8. Lee J, Sohn I, Do I, Kim K, Park S, Park J, et al. Nanostring-based multigene assay to predict recurrence for gastric cancer patients after surgery. PLoS One 2014;9:e90133.
9. Sivendran S, Chang R, Pham L, Phelps R, Harcharik S, Hall L, et al. Dissection of immune gene networks in primary melanoma tumors critical for anti-tumor surveillance of patients with stage II-III resectable disease. J Invest Dermatol 2014;134:2202–11.
10. Stricker TP, Morales La Madrid A, Chlenski A, Guerrero L, Salwen HR, Gosiengfiao Y, et al. Validation of a prognostic multi-gene signature in high-risk neuroblastoma using the high throughput digital NanoString nCounter system. Mol Oncol 2014;8:669–78.
11. Veldman-Jones M, Lai Z, Wappett M, Harbron C, Barrett JC, Harrington EA, et al. Reproducible, quantitative and flexible molecular sub-typing of clinical DLBCL samples using the NanoString nCounter system. Clin Cancer Res 2014 Oct 9 [Epub ahead of print].

12. Northcott PA, Shih DJ, Remke M, Cho YJ, Kool M, Hawkins C, et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta Neuropathol 2011;123: 615–26.
13. Lira ME, Choi Y, Lim SM, Deng S, Huang D, Ozeck M, et al. A single-tube multiplexed assay for detecting *ALK*, *ROS1*, and *RET* fusions in lung cancer. J Mol Diagn 2014;16:229–43.
14. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 2008;26:317–25.
15. Mestdagh P, Hartmann N, Baeriswyl L, Andreasen D, Bernard N, Chen C, et al. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. Nat Methods 2014;11: 809–15.
16. Northcott PA, Shih DJH, Peacock J, Garzia L, Morrissy AS, Zichner T, et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. Nature 2012;488:49–56.
17. Ullal AV, Peterson V, Agasti SS, Tuang S, Juric D, Castro CM, et al. Cancer cell profiling by barcoding allows multiplexed protein analysis in fine-needle aspirates. Sci Transl Med 2014;6:219ra9.