Date: December 3rd, 2018

Department of Translational Molecular Pathology (TMP)
Immune Profiling Laboratory
Cancer Immune Monitoring and Analysis Center
Division of Pathology and Lab Medicine
The University of Texas MD Anderson Cancer Center

Directors:
Ignacio I. Wistuba, MD, Professor and Chair, TMP
Raja Luthra, PhD, Professor, Hematopathology and TMP
Dzifa Douse, PhD, Scientific Project Director, TMP

Nanostring Analytical Validation

Version 2.0

The nCounter Analysis System from Nanostring Technologies Inc. utilizes a novel digital barcode technology and single molecule imaging for the direct hybridization and detection of hundreds of unique transcripts in a single reaction with high levels of precision and sensitivity (< 1 copy per cell). The nCounter Pan Cancer Panel performs multiplex gene expression analysis with 770 genes from 24 different immune cell types, common checkpoint inhibitors, CT antigens, and genes covering both the adaptive and innate immune response.

Nanostring nCounter Platform	
(i) Accuracy	Accuracy was measured by the closeness of replicate analysis outcome to the original result for 29 Formalin Fixed Paraffin Embedded (FFPE) tissue samples representing Breast Tumor (6), Breast Normal (6), Tonsil (6), Lung Tumor (5) and Lung Normal (6). Intra assay (Runs 3, 4) and Inter assay (Runs 5, 6, 7, 8) assessments were also performed.
(ii) Analytical sensitivity	RNA input of 0 ng, 62.5 ng, 125 ng, 250 ng, 500 ng and 625 ng from two Breast FFPE tissue samples were analyzed, both intra and inter assay, for determining the lower limit of detection based on input quantification. (Runs 1 and 2). The limit of the assay was 125 ng with an optimal RNA input of 250ng.
(ii) Precision: Intra and Inter-assay reproducibility.	Coefficient of variation (% CV) was determined for intra- and inter-assay consistency in 6 FFPE breast tumor samples with the ideal RNA input (250 ng) tested in 2 runs with duplicates in each run. Intra-assay %CV observed was % CV < 13, and inter-run CV was % CV <14 (Run 3 and 4) for all 770 genes assessed (Genes with low expression levels contribute to large % CV).
(iii) Analytical reproducibility	Inter-run and inter-operator reproducibility was determined in 6 FFPE lung tumor samples and 6 tonsil samples in 2 runs from 2 different operators (Run 5 and 6). Inter-run reproducibility observed was %CV<15) for all 770 genes assessed (Genes with low expression levels contribute to large % CV).
(ix) Analytical specificity	Analytical sensitivity was determined by using an orthogonal platform (RNA-seq) as reference. Receiver Operator Curve (ROC) was determined between the two platforms. Normalized counts data from three (3) FFPE lung cancer patient samples (in triplicate runs) using RNA-seq were used. Normalized counts from 5 FFPE lung cancer samples (in duplicate) run on the NanoString platform (Run 5) was used to compare with that from RNA-seq. Area under the curve (AUC) was 0.876 (Specificity of 89.9%, sensitivity of 70.2).

(x) Any other performance characteristics required for assay performance	All of the required equipment, have annual service contracts with regular Preventive Maintenance performed to maintain optimal calibration and performance. All other small equipment such as multi-channel pipettes and laboratory material have calibration performed by certified vendors.

1. ***Analyte(s):*** RNA extracted from 30 Formalin Fixed Paraffin Embedded (FFPE) Tissue Samples.

2. ***Technical platform(s):***
This report describes the analytical validation parameters (RNA input, analytical sensitivity, reproducibility and precision) of nCounter platform from Nanostring technology at the department of Translational Molecular Pathology (TMP) at MD Anderson Cancer Center.

3. ***Reagents and Equipment***
nCounter Master Kit
nCounter Code Sets
Nuclease Free Water
nCounter Prep plates
nCounter Cartridge
Thermal Cycler
Ice bucket
Pipettes
Nanostring Prep Station
Nanostring nCounter cartridge reader
Nanostring nSolver analysis software

4. ***Data Summary***
Gene expression on the nCounter platform (prep Station and Digital Analyzer – Nanostring technology, Seattle, USA) was assessed with the nCounter PanCancer Immune profiling panel. The nCounter® PanCancer Immune Profiling Panel is a highly multiplexed gene expression panel designed to quantitate 770 genes that fall into four functional categories; such as genes which identify immune cells, genes that assist with assessing immunological function and response to immunotherapy, genes identifying tumor-specific antigens and Housekeeping genes that facilitate sample-to-sample normalization.

The platform estimates the quantity of each mRNA transcript using a multiplexed hybridization system and digital readouts of fluorescent barcoded probes that are hybridized to each transcript. Raw counts resulting from the analysis were normalized. Normalized data was analyzed using NanoString's nSolver version 4.0 software and log2 transformed for further statistical analyses. Statistical Analyses was performed by R version 3.3.0. Background was defined as the mean of the negative controls plus two standard deviations. Data distribution plots were generated with log2 transformed raw data, and the background level was labeled as vertical line. Clustering analysis was performed on normalized data from all genes with the Pearson correlation distance and ward.D2 linkage rule. The Pearson correlation coefficients were calculated among duplicates. The slope of trendline was obtained by fitting a linear regression of data from duplicates.

Sample distribution in each run were designed for achieving different analytical validation goals.

- <u>Analytical sensitivity:</u> RNA inputs of 0 ng, 62.5 ng, 125 ng, 250 ng, 500 ng and 625 ng from two Breast FFPE tissue samples were duplicated across Runs 1 and 2 to determine optimal input RNA.

- Intra- and Inter-assay reproducibility: 29 Formalin Fixed Paraffin Embedded (FFPE) tissue samples representing Breast Tumor (6), Breast Normal (6), Tonsil (6), Lung Tumor (5) and Lung Normal (6). Intra assay (Runs 3, 4) and Inter assay (Runs 5, 6, 7, 8) were tested in replicates.
- Inter- Operator reproducibility: 6 lung and 6 tonsil samples were tested by different operators in 2 runs - Runs 5 and 6.
- Tumor – Normal differentiation: 12 Normal (non – tumor)specimens were tested on Runs 7 and 8 and cluster analysis performed on the output data from Runs 3 to 8 to delineate between normal and tumor samples.
- Analytical Specificity: Normalized counts data from three (3) FFPE lung cancer patient samples (in triplicate runs) using RNA-seq were used and compared with normalized counts from 5 FFPE lung cancer samples (in duplicate) run on the NanoString platform (Run 5).

5. *Analytical Sensitivity: RNA input*

To determine the optimal input RNA for reproducible data and the effects of suboptimal input on the assay's outcome, two breast cancer samples with different concentration ranging from 0 ng to 625 ng (0ng, 62.5ng, 125ng, 250ng, 500ng and 625ng) were analyzed in two runs. Table 1 shows input characteristics of the samples tested. The limit of detection based on input quantification and optimal input RNA for reproducible data was determined.

Table 1. Sample information.

Tumor Type	Sample ID	Run 1 RNA input (ng)	Run 2 RNA input (ng)
Breast	18-150	0	0
Breast	18-150	62.5	62.5
Breast	18-150	125	125
Breast	18-150	250	250
Breast	18-150	500	500
Breast	18-150	625	625
Breast	18-148	0	0
Breast	18-148	62.5	62.5
Breast	18-148	125	125
Breast	18-148	250	250
Breast	18-148	500	500
Breast	18-148	625	625

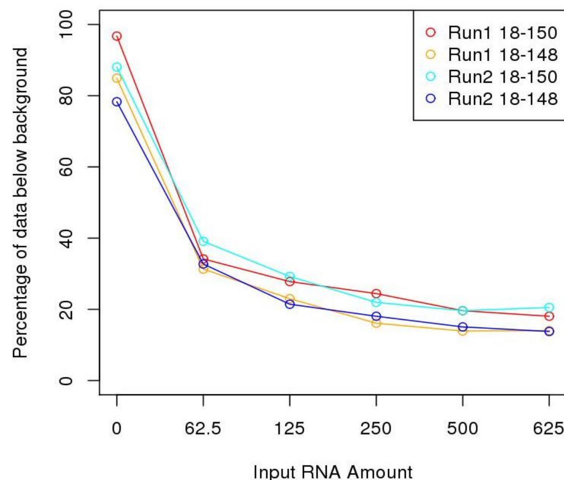
As per Nanostring user manual, background cut-off is determined by calculating the mean of the negative controls plus two standard deviations. The limit of detection based on input quantification was assessed by the percentage of data that was lower than background cut-off. (Table 2 and Figure 1). With increasing RNA input quantity the percentage of data below the background cut-off was observed to approach a plateau from RNA input of 125 ng onwards confirming the limit of detection as 125ng RNA input with the optimal RNA input to be 250ng.

Table 2. Percentage of Data below background.

Tumor Type	Sample ID	RNA input (ng)	Run 1 (%)	Run 2 (%)
Breast	18-150	0	96.75%	88.05%

Breast	18-150	62.5	34.16%	39.09%
Breast	18-150	125	27.79%	29.22%
Breast	18-150	250	24.42%	21.95%
Breast	18-150	500	19.61%	19.61%
Breast	18-150	625	18.05%	20.52%
Breast	18-148	0	84.94%	78.31%
Breast	18-148	62.5	31.30%	32.73%
Breast	18-148	125	22.99%	21.43%
Breast	18-148	250	16.10%	18.05%
Breast	18-148	500	13.90%	15.06%
Breast	18-148	625	14.03%	13.77%

Figure 1. Percentage of data below background and RNA input.



Clustering and correlation analyses. Clustering analysis on log2 normalized data with Pearson correlation distance and ward.D2 linkage rule is showed in Figure 2. Samples with no RNA input clustered together whereas independently of RNA input amount the 2 samples tested, cluster in 2 clusters. Within each sample, replicates with same RNA input amount clustered together demonstrating the good inter-run consistency of the assay. Pearson correlation coefficient between sample replicates with RNA input >0ng showed high correlation values (rho>9) independently of RNA input. A distribution of Pearson correlation coefficients is shown in Figure 3.

Figure 2. Clustering Analysis of log 2 normalized data from Run 1 and Run 2. (Samples labeled as Sample ID, Run ID and RNA input).

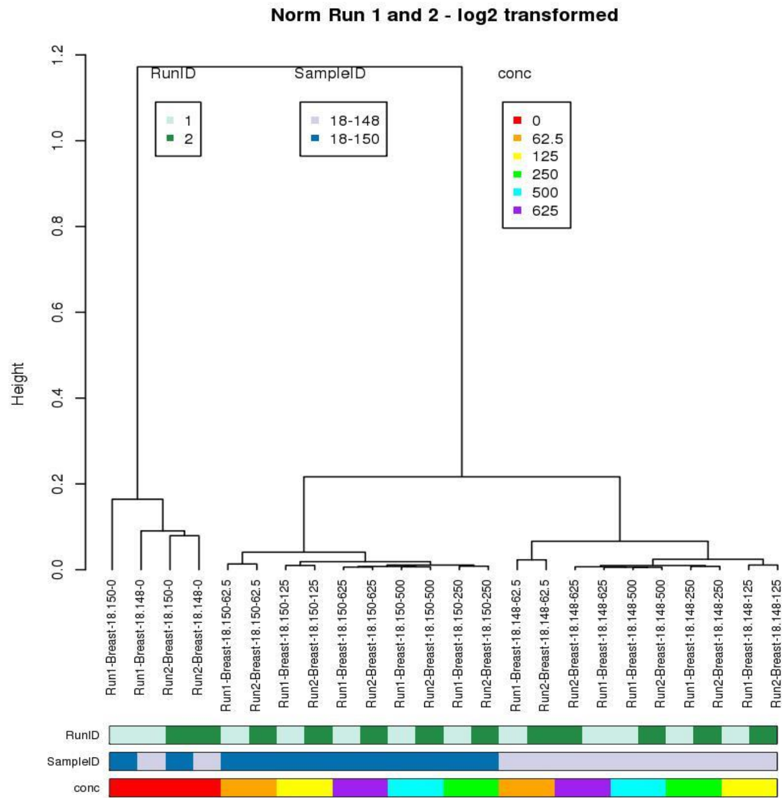
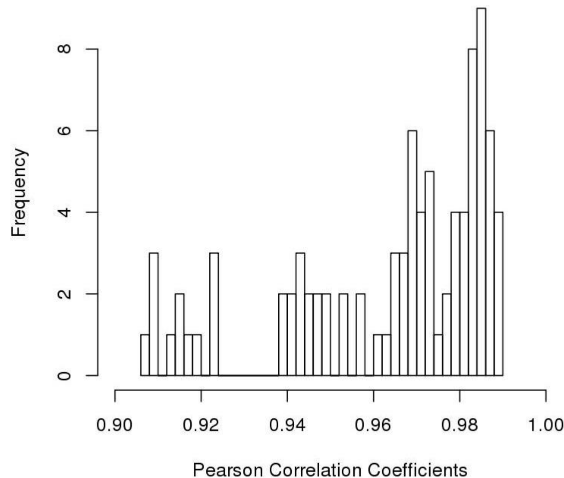


Figure 3. Histogram distribution of Pearson correlation coefficients of sample replicates in Run 1 and 2.



Conclusion

Higher input RNA quantity correlates to more data above the background cutoff, approaching a plateau from 125ng onwards (no significant difference beyond 500 ng of RNA input). Data obtained from 250 ng RNA input were similar to that of 500 ng indicating the optimal RNA input to be 250 ng with the lower limit of detection based on quantification as 125 ng.

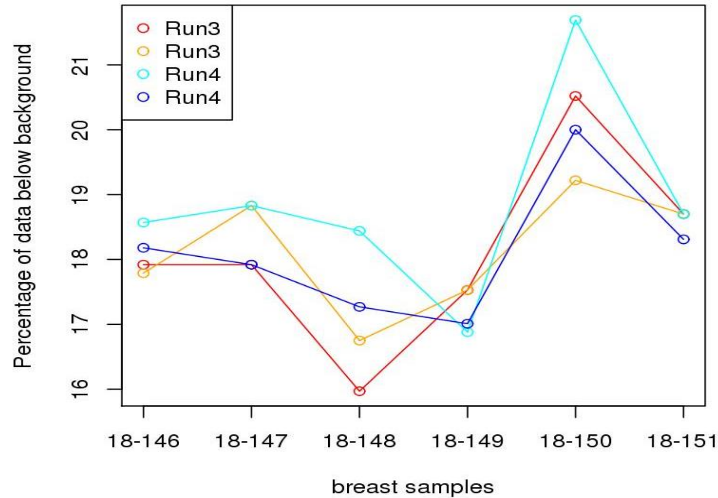
6. Precision: Intra and Inter Assay Variability

To determine intra-run and inter-run consistency of Nano string assay, 6 FFPE tumor breast samples (Sample ID: 18-146, 18-147, 18-148, 18-149, 18-150 and 18-151) were tested using optimal RNA input (250 ng) in duplicates in two runs (Run 3 and Run 4). Data percentage below background was assessed (Table 3, Figure 4).

Table 3. Percentage of data below background in Run 3 and Run 4.

Tumor Type	Sample ID	RNA input (ng)	Run 3 (%)	Run 4 (%)
Breast	18-146.1	250	17.92 %	18.57%
Breast	18-146.2	250	17.79 %	18.18%
Breast	18-147.1	250	17.92%	18.83%
Breast	18-147.2	250	18.83%	17.92%
Breast	18-148.1	250	15.97%	18.44%
Breast	18-148.2	250	16.75%	17.27%
Breast	18-149.1	250	17.53%	16.88%
Breast	18-149.2	250	17.53%	17.01%
Breast	18-150.1	250	20.52%	21.69%
Breast	18-150.2	250	19.22%	20%
Breast	18-151.1	250	18.70%	18.70%
Breast	18-151.2	250	18.7%	18.31%

Figure 4. Percentage of data below background in Run 3 and Run 4.



Data distribution. Figures 5 and 6 show the distribution of log2 transformed raw data from samples in Run 3 (background cutoff 18.24) and Run 4 (background cutoff 19.06).Background cutoff is illustrated as the green vertical line. Similar distribution pattern was observed for same sample within the run as well as across runs.

Figure 5. Data distribution patterns for sample replicates in Run 3.

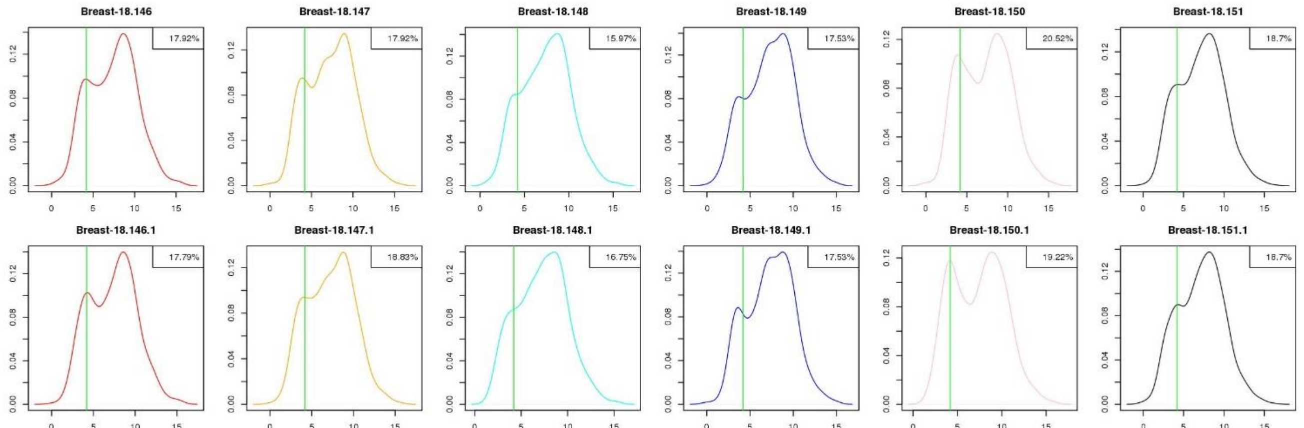
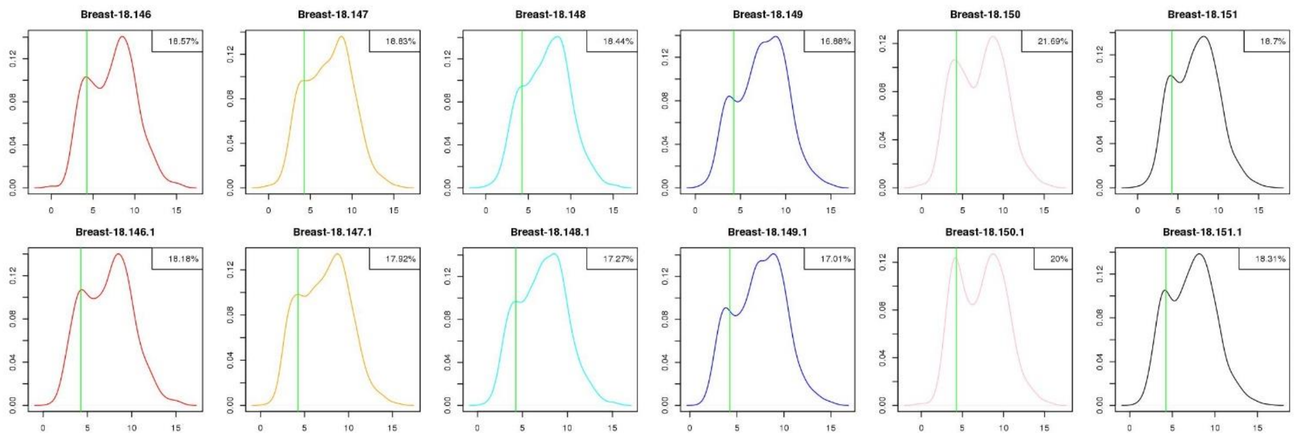


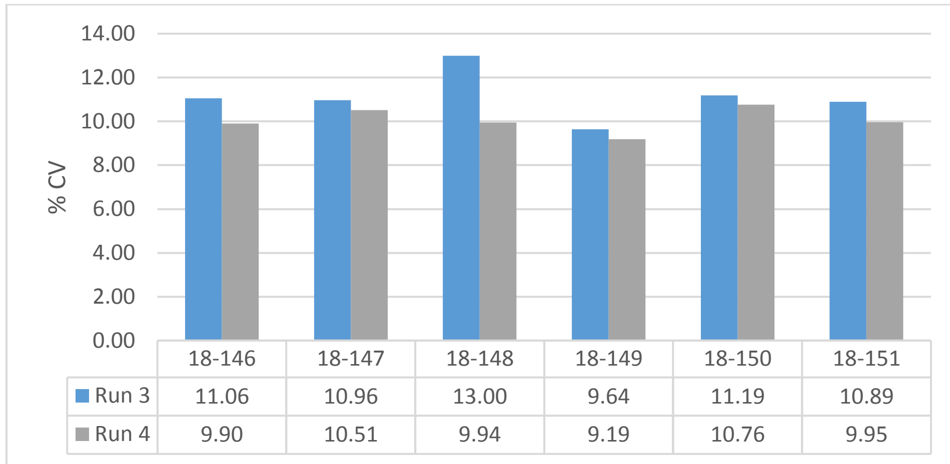
Figure 6. Data distribution patterns for sample replicates in Run 4.



Coefficient of Variation (CV). CV and %CV calculations were performed on the gene expression data as well as the percentage of data below the cut-off for replicate samples within the run and across the runs to assess the extent of variability in relation to the mean of the population.

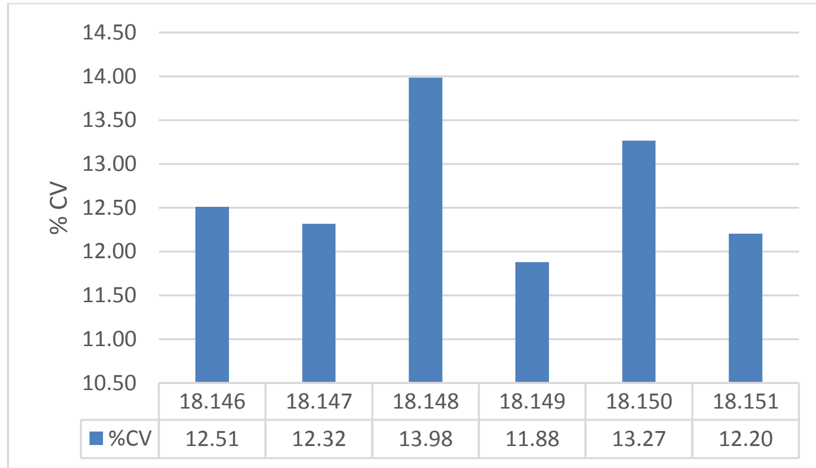
Intra-assay. Intra-run % CV for each gene of the Encounter platform was calculated. The average % CV of all genes for each sample in Run 3 and 4 are shown in Figure7. Intra-assay CV observed was % CV < 13 (Run 3 and 4) for all 770 genes assessed (Genes with low expression levels contribute to large % CV).

Figure 7. Sample Intra-run % CV in Run 3 and Run 4.



Inter-assay .Inter-run % CV for each gene of the Encounter platform was calculated. The average % CV for each sample is shown in Figure 8. Average %CV across all samples was 12.69 % (Run 3 and 4) for all 770 genes assessed (Genes with low expression levels contribute to large % CV).

Figure 8. Average sample % CV inter-run (Run 3 and Run 4).



For each sample the % CV of the percentage of data below the background cut-off, was calculated to assess the variability of the data output both intra-run and inter-run. The % CV for each sample is shown in Table 4 and Average %CV across all samples was 3.16 % (Run 3 and 4) demonstrating good concordance.

Table 4. Data Percentage below background in Run 3 and 4

Tumor Type	Sample ID	RNA input (ng)	Run 3 (% below background cut-off)		Run 4 (% below background cut-off)		%CV	Avgas %CV
			Sample-1	Sample-2	Sample-1	Sample-2		
Breast	18-146	250	17.92%	17.79%	18.57%	18.18%	1.89	3.16
Breast	18-147	250	17.92%	18.83%	18.83%	17.92%	2.85	
Breast	18-148	250	15.97%	16.75%	18.44%	17.27%	6.06	
Breast	18-149	250	17.53%	17.53%	16.88%	17.01%	1.98	
Breast	18-150	250	20.52%	19.22%	21.69%	20%	5.09	
Breast	18-151	250	18.70%	18.70%	18.70%	18.31%	1.05	

Clustering and Correlation Analysis. Clustering analysis on log 2 normalized data with Pearson correlation distance and ward.D2 linkage rule showed all replicate samples were clustered together (Figure 9). Pearson correlation coefficient between sample replicates showed high correlation values (rho>9) independently of the assay run as seen in Figure 10.

Figure 9. Clustering Analysis of log 2 normalized data of sample replicates in Run 3 and 4. (Samples labeled as Sample ID, Run ID and RNA input).

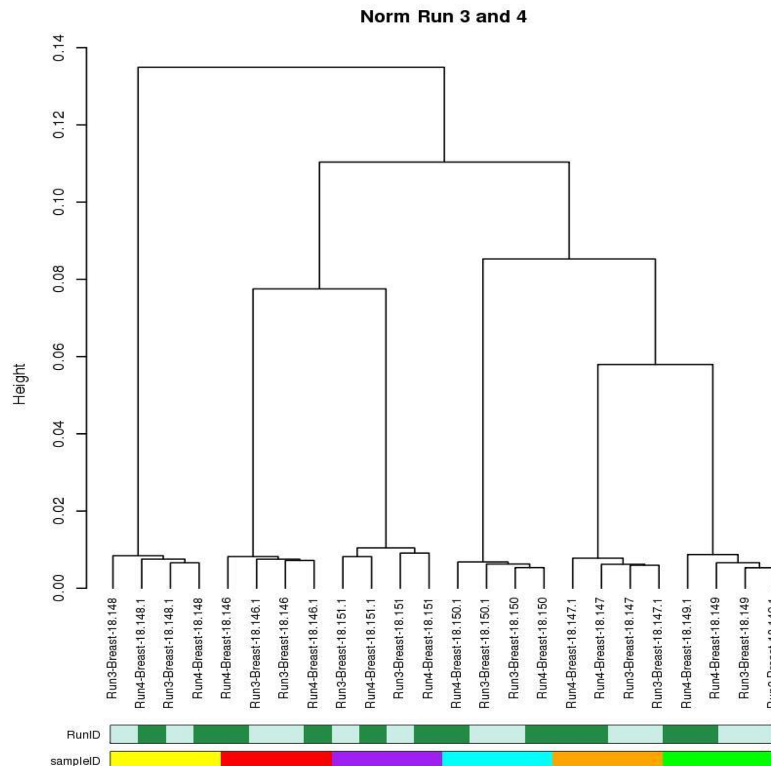
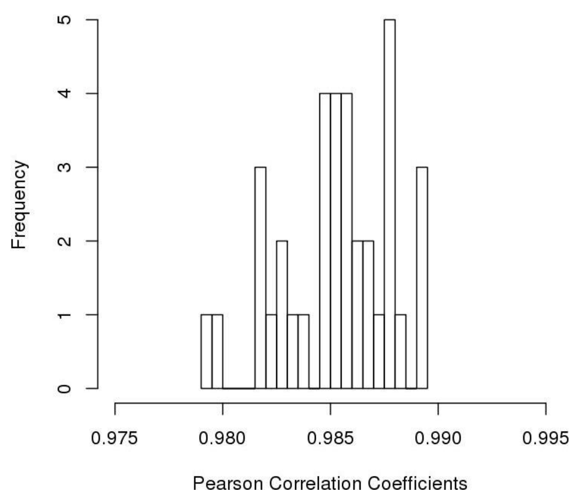


Figure 10. Histogram distribution of Pearson correlation coefficients of sample replicates in Run 3 and 4.



Slope Values. Slope values were determined by fitting linear model of duplicates. Tables 5 and 6 show slope values for intra- and inter-assay reproducibility respectively. Slope trendline was calculated with a value close to 1 in both intra- and inter-assay as shown in Figure 11.

Table 5. Slope for Intra-Assay reproducibility (Run 3 and Run 4)

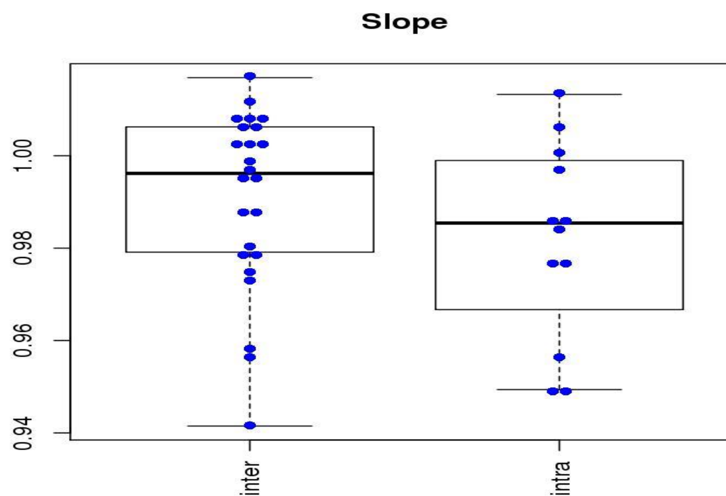
Run ID, Sample ID (Duplicate)	Slope	Type
Run3-Breast-18.146	0.949	intra
Run3-Breast-18.147	0.956	intra
Run3-Breast-18.148	0.997	intra
Run3-Breast-18.149	0.984	intra
Run3-Breast-18.150	0.976	intra
Run3-Breast-18.151	0.977	intra
Run4-Breast-18.146	1.000	intra
Run4-Breast-18.147	1.006	intra
Run4-Breast-18.148	1.013	intra
Run4-Breast-18.149	0.986	intra
Run4-Breast-18.150	0.986	intra
Run4-Breast-18.151	0.949	intra

Table 6. Slope for Inter-Assay reproducibility (Run 3 and Run 4)

Run ID, Sample ID	Slope	Type
Run3-vs-Run4-18.146.1	0.941	inter
Run3-vs-Run4-18.146.2	0.957	inter
Run3-vs-Run4-18.146.3	0.979	inter
Run3-vs-Run4-18.146.4	1.002	inter
Run3-vs-Run4-18.147.1	0.997	inter
Run3-vs-Run4-18.147.2	1.011	inter
Run3-vs-Run4-18.147.3	0.959	inter
Run3-vs-Run4-18.147.4	0.975	inter

Run3-vs-Run4-18.148.1	1.007	inter
Run3-vs-Run4-18.148.2	1.002	inter
Run3-vs-Run4-18.148.3	0.994	inter
Run3-vs-Run4-18.148.4	0.979	inter
Run3-vs-Run4-18.149.1	0.978	inter
Run3-vs-Run4-18.149.2	0.988	inter
Run3-vs-Run4-18.149.3	0.972	inter
Run3-vs-Run4-18.149.4	1.007	inter
Run3-vs-Run4-18.150.1	1.007	inter
Run3-vs-Run4-18.150.2	1.016	inter
Run3-vs-Run4-18.150.3	0.994	inter
Run3-vs-Run4-18.150.4	1.007	inter
Run3-vs-Run4-18.151.1	0.997	inter
Run3-vs-Run4-18.151.2	1.003	inter
Run3-vs-Run4-18.151.3	1.005	inter
Run3-vs-Run4-18.151.4	0.987	inter

Figure 11. Slope trend-line Boxplots for Run 3 and 4, Intra – and inter-assay.



Conclusion. %CV, slope and slope trendline calculations performed on the gene expression data and on the percentage of data below the cut-off, for replicate samples within the run and across the runs confirmed good concordance within replicates in the same run and across runs.

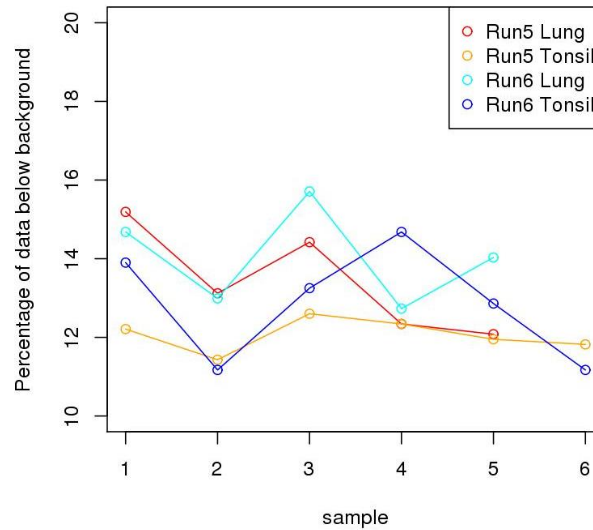
7. *Inter-operator Reproducibility*

To asses inter-operator reproducibility, 5 FFPE lung tumor samples (Lung 18-138, 18-140, 19-141, 18-142, 18-124) and 6 FFPE tonsil samples (Tonsil 18-132, 18-133, 18-134, 18-135, 18-136, 18-137) were analyzed in two runs performed by two different operators (Run 5 and Run 6). Amount off RNA input was the optimal 250 ng. Data percentage below background was assessed (Table 7 and Figure 12).

Table 7. Percentage of data below background in Run 5 and Run 6.

Tissue Type	Sample ID	RNA input (ng)	Run 5 (%)	Run 6 (%)	% CV	Average %CV
Lung	18-138	250	15.19%	14.68%	2.41	4.76
Lung	18-140	250	13.12%	12.99%	0.70	
Lung	18-141	250	13.12%	12.99%	0.70	
Lung	18-142	250	12.34%	12.73%	2.20	
Lung	18-124	250	12.08%	14.03%	10.56	
Tonsil	18-132	250	12.21%	13.90%	9.15	
Tonsil	18-133	250	11.43%	11.17%	1.63	
Tonsil	18-134	250	12.60%	13.25%	3.56	
Tonsil	18-135	250	12.34%	14.68%	12.25	
Tonsil	18-136	250	11.95%	12.86%	5.19	
Tonsil	18-137	250	11.82%	11.17%	3.99	

Figure 12. Percentage of data below background in Run 5 and Run 6



Clustering and Correlation Analysis. Clustering analysis on log 2 normalized data with Pearson correlation distance and ward.D2 linkage rule showed all replicate samples were clustered together (Figure 12). Pearson correlation coefficient between sample replicates showed high correlation values (rho>9) as observed in Figure 13.

Figure 13. Clustering Analysis of log 2 normalized data of sample replicates in Run 5 and Run 6. (Samples labeled as Sample ID, Run ID and RNA input).

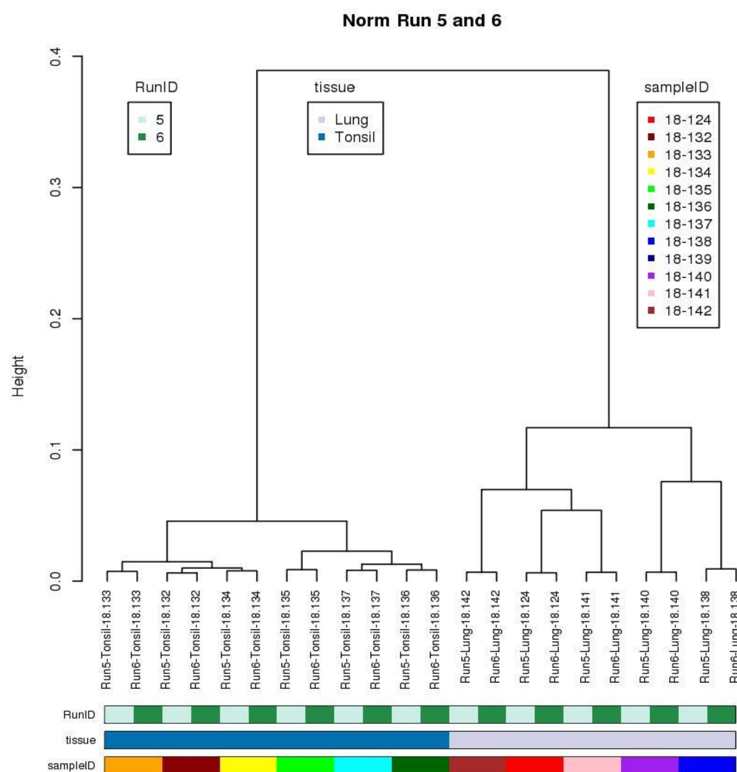
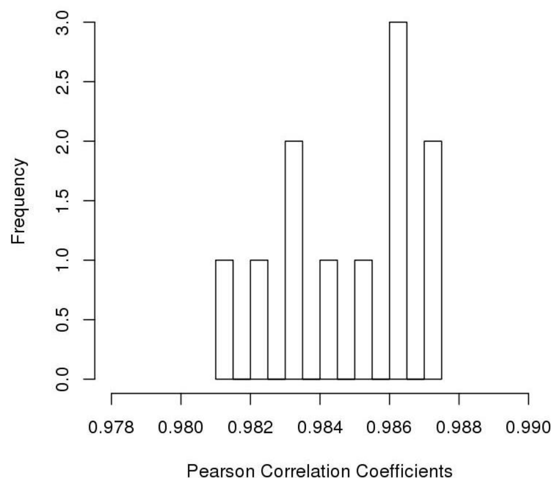


Figure 14. Histogram distribution of Pearson correlation coefficients of sample replicates in Run 5 and Run 6.

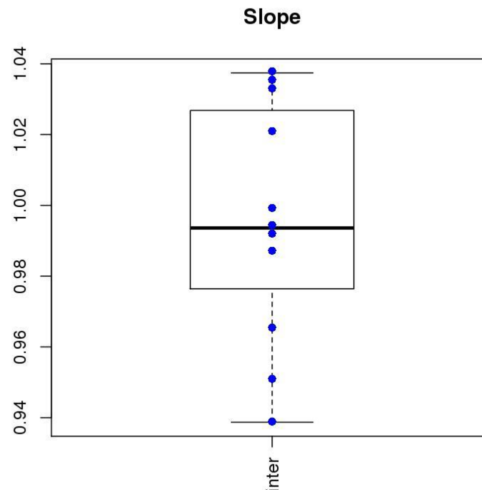


Slope Values. Slope values were determined by fitting linear model of Run 5 and 6 per individual operator. s. Table 8 and Figure 15 show slope values for inter-operator reproducibility. Slope values observed were close to 1 showing reproducibility of the data.

Table 8. Slope for Inter-operator reproducibility (Run 5 and 6)

Run ID Sample ID	Slope	Type
Run5-vs-Run6-Lung-18.138	0.950	Inter-operator
Run5-vs-Run6-Lung-18.140	1.034	Inter-operator
Run5-vs-Run6-Lung-18.141	1.032	Inter-operator
Run5-vs-Run6-Lung-18.142	0.999	Inter-operator
Run5-vs-Run6-Lung-18.124	0.986	Inter-operator
Run5-vs-Run6-Tonsil-18.132	0.993	Inter-operator
Run5-vs-Run6-Tonsil-18.133	0.966	Inter-operator
Run5-vs-Run6-Tonsil-18.134	1.020	Inter-operator
Run5-vs-Run6-Tonsil-18.135	0.991	Inter-operator
Run5-vs-Run6-Tonsil-18.136	0.938	Inter-operator
Run5-vs-Run6-Tonsil-18.137	1.037	Inter-operator

Figure 15. Slope trend-line Boxplots for Run 5 and 6, Inter-operator Reproducibility.



Conclusion. %CV, Slope and slope trendline calculations performed on the percentage of data below the cut-off, for replicate samples across the runs confirmed good inter-operator concordance with an average % CV of 4.76%.

8. *Analytical Specificity: Normal vs Tumor FFPE samples*

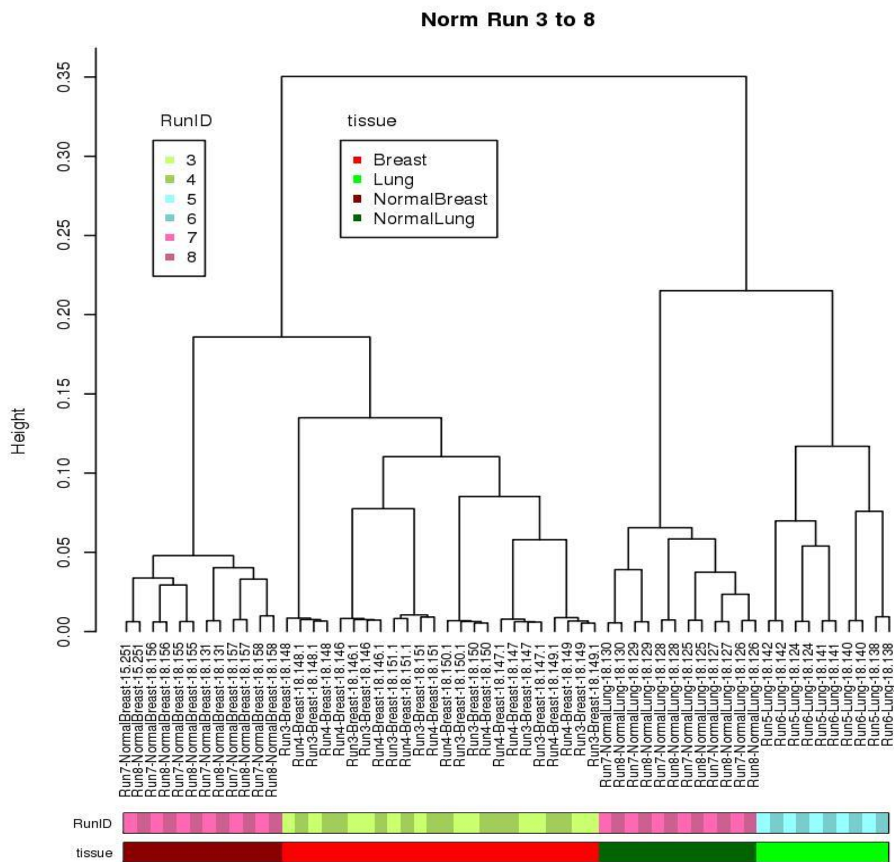
To assess the ability of Encounter Nano string platform to delineate between normal and tumor samples, data from the tumor sets (6 FFPE breast tumor and 5 FFPE lung samples) used to assess assay reproducibility was compared to 6 normal breast samples and 6 normal lung samples were analyzed in two runs (Run 7 and Run 8). Input RNA was 250 ng. Percentage of Data below the background cut-off for normal breast and lung normal samples was checked for concordance (Table 9).

Table 9. Percentage of data below background cut-off in Run 7 and Run 8.

Tissue Type	Sample ID	RNA input (ng)	Run 5 (%)	Run 6 (%)
Normal Lung	18-125	250	14.55 %	14.29%
Normal Lung	18-126	250	14.03%	14.55 %
Normal Lung	18-127	250	15.06%	14.42%
Normal Lung	18-128	250	16.1%	16.88%
Normal Lung	18-129	250	14.29%	15.45%
Normal Lung	18-130	250	13.12%	13.25%
Normal Breast	18-155	250	21.82%	22.34%
Normal Breast	18-156	250	21.43%	21.04%
Normal Breast	18-131	250	19.61%	17.53%
Normal Breast	18-157	250	21.69%	21.17%
Normal Breast	18-158	250	22.73%	22.73%
Normal Breast	18-251	250	20.26%	19.48%

Clustering and Correlation Analysis. Clustering analysis on log 2 normalized data with Pearson correlation distance and ward.D2 linkage rule showed all replicate samples were clustered together (Figure 16). Breast and lung samples were separated in different clusters. Normal samples were separated from tumor samples within the same tissue type.

Figure 16. Clustering Analysis of log 2 normalized data of sample replicates of normal breast, normal lung, and tumor lung and tumor breast samples. (Samples labeled as Sample ID, Run ID and RNA input).



Conclusion. Percentage of data below the cut-off, for replicate samples across the runs indicated good concordance between runs. Clustering analysis on log 2 normalized data demonstrated Breast and lung samples was separated in different clusters and Normal samples were distinct from tumor samples within the same tissue type.

9. *Analytical Specificity: NanoString vs RNA-seq for Lung Cancer samples*

To assess the specificity of the NanoString platform, an orthogonal platform (RNA-seq) was used as a reference and a Receiver Operator Curve (ROC) determined between the two platforms.

Sample Source: Normalized counts data from three (3) FFPE lung cancer patient samples (in triplicate) run using RNA-seq was obtained from a different lab in the institution. Normalized counts from 5 FFPE lung cancer samples (in duplicate) run on the NanoString platform was used to compare with that from RNA-seq. Data from both sources underwent a log2 transformation before comparing them.

Method and results: There were 733 genes in common between the RNA-seq and NanoString data, these genes were used for comparison between the two platforms. To assess the correlation between RNA-Seq data and NanoString data, we calculated mean expression of lung cancer samples in both RNA-Seq and NanoString data. Figure 17 shows a good correlation between the two platforms with a spearman's coefficient of 0.76. To identify highly expressed genes, we calculated CV (Coefficient of Variation) and mean expression values for each gene. We chose highly expressed genes by the following two criteria: Mean expression in upper quantile and CV < 0.5. Figure 18 shows CV vs mean expression for RNA-Seq and NanoString data, with the cutoffs labeled in red line. The genes on the right-bottom corner are defined as highly expressed genes. Using RNA-Seq data as the reference, we calculated sensitivity and specificity with different cutoffs on NanoString data and generated an ROC curve for the data. Figure 19 shows the ROC curve with an area under the curve (AUC) of 0.876, which is considered good. For example, at cutoffs labeled in red line (Figure 18), we have specificity of 89.9% and. sensitivity of 70.2%.

Figure 17. Correlation between RNA-seq and NanoString for FFPE lung cancer samples

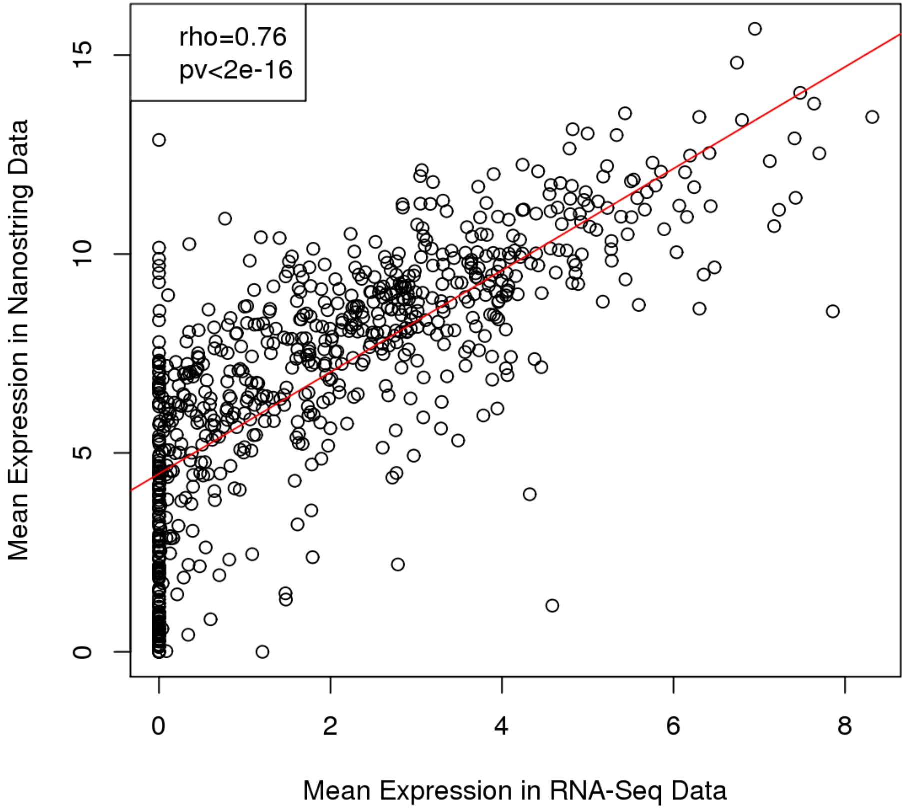


Figure 18. CV vs mean expression for RNA-Seq and Nanostring data

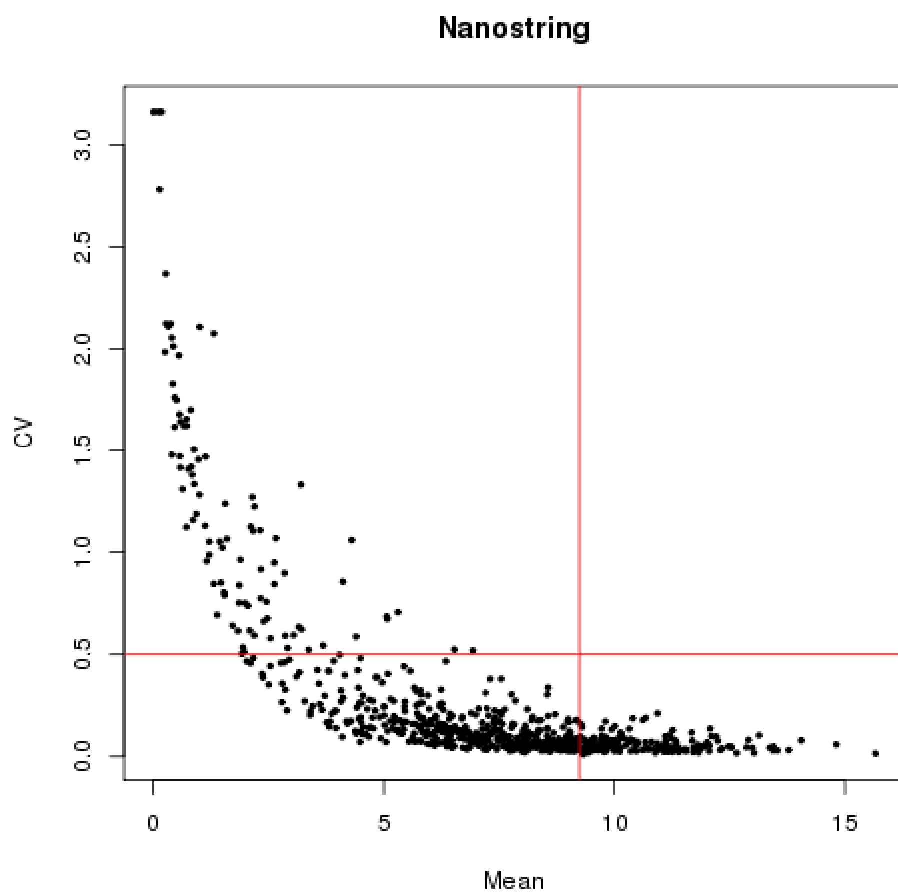
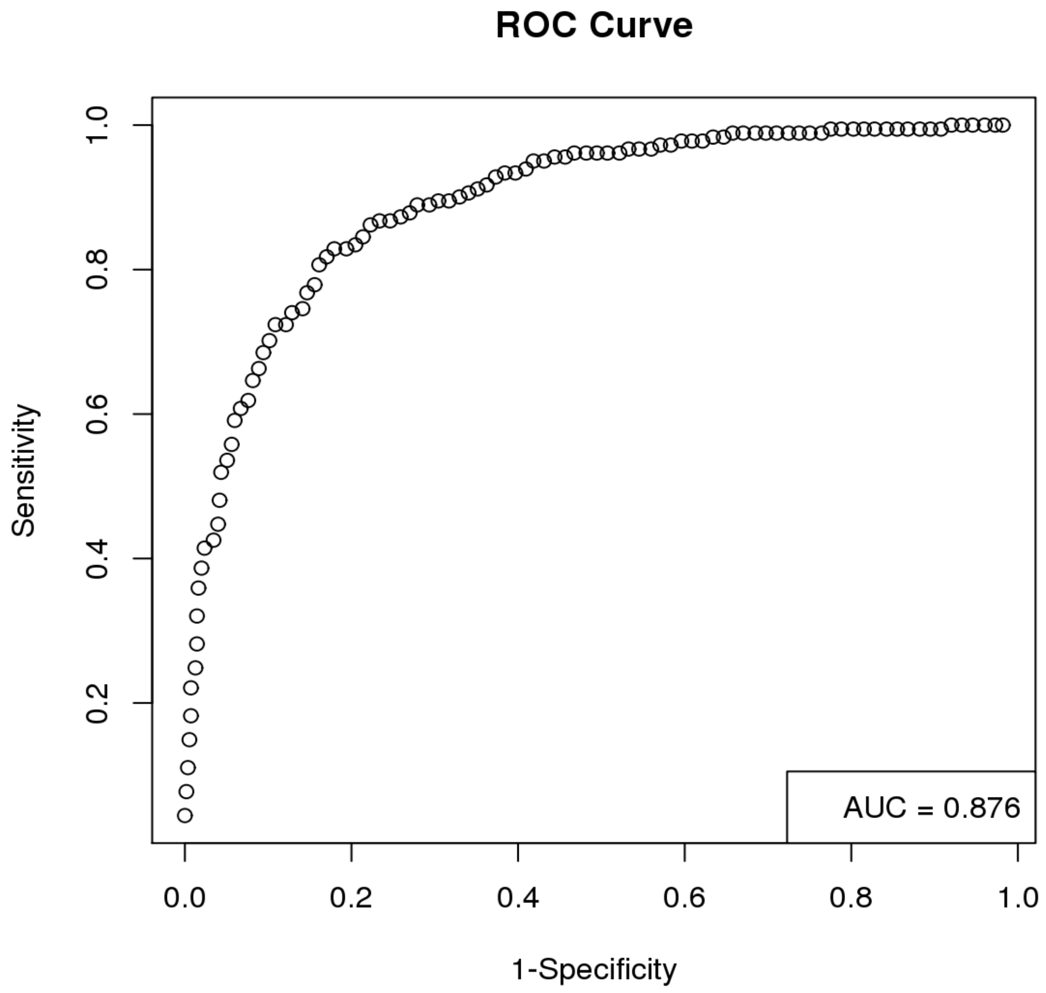


Figure 19. ROC curve



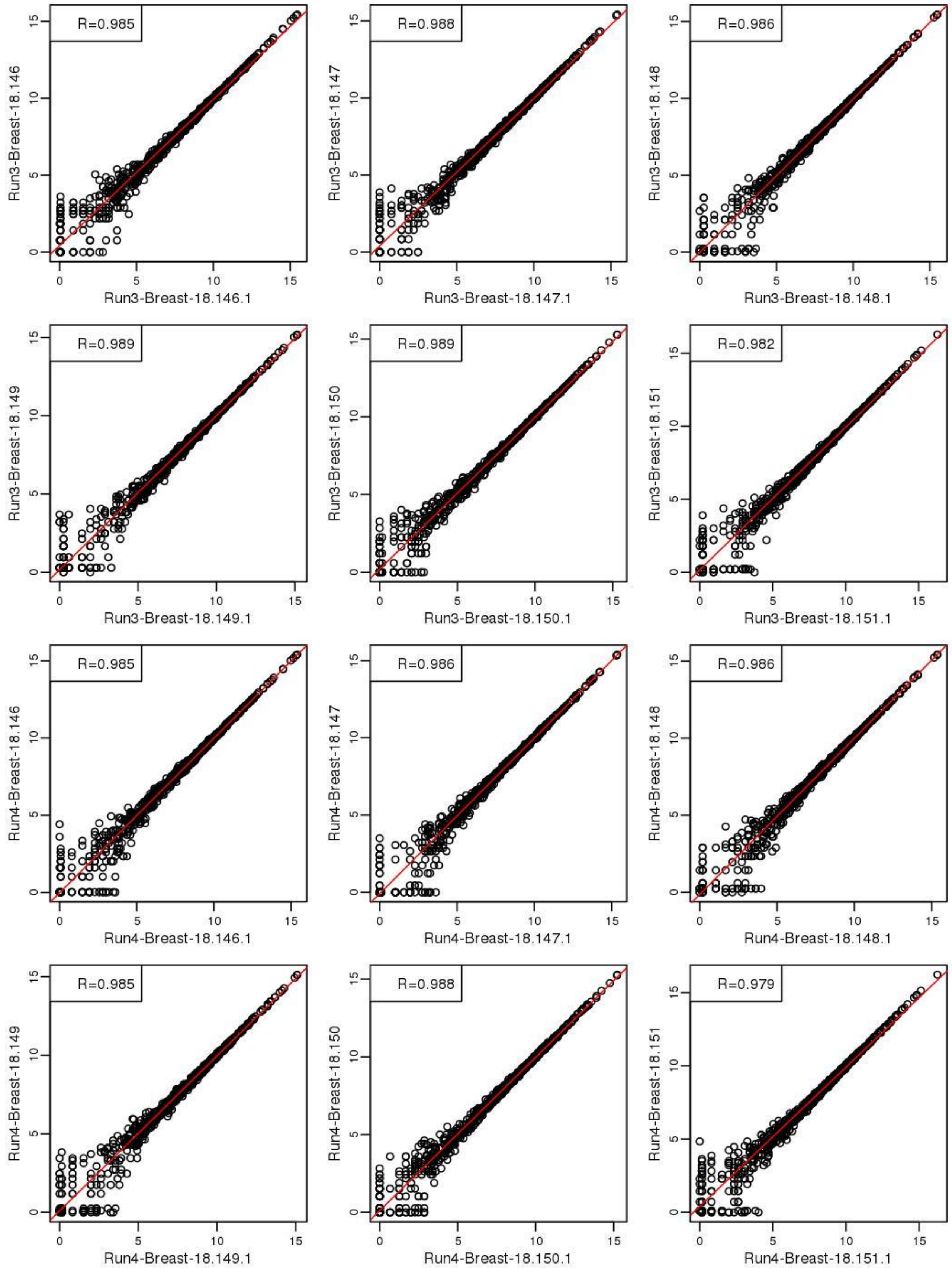
Director: Ignacio I Wistuba, MD, Professor and Chair; Dzifa Duose, PhD Scientific project Manager, Translational Molecular Pathology

[Signatures] and [dates]

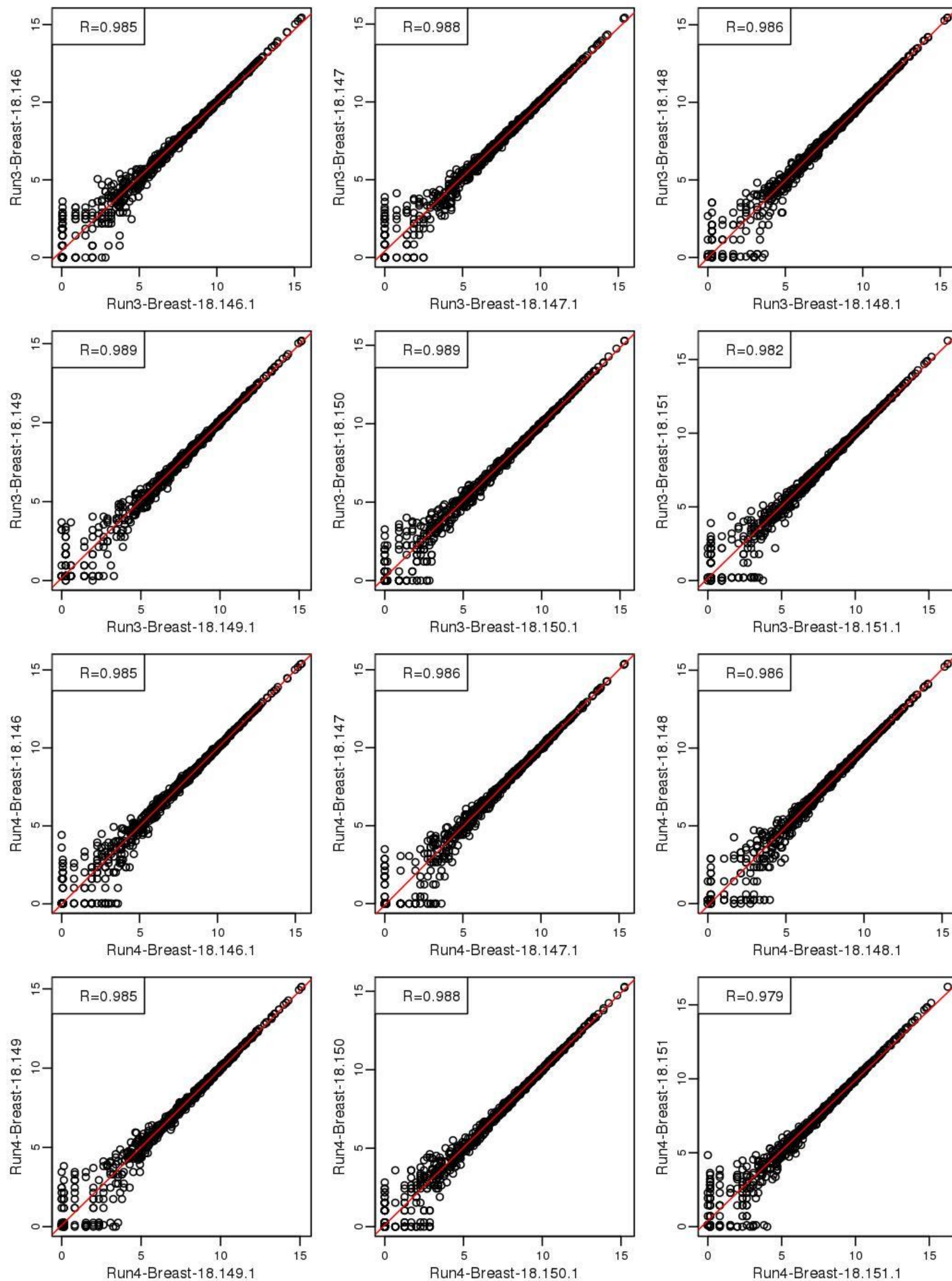
Signature/date: ______________12/03/2018______
Name: *Dzifa Duose, PhD*
Title: *Scientific Project Manager*

APPENDIX

Supplementary Figure 1. Correlation plots Intra-Run reproducibility (Run 3 and 4)



Supplementary Figure 2. Correlation plots Inter-Run reproducibility (Run 3 and 4)



Supplementary Figure 3. Correlation plots Inter-operator reproducibility (Run 5 and 6)