# nanoString
### TECHNOLOGIES

# nCounter®
## Expression Data Analysis Guide

**NanoString Technologies, Inc.**

530 Fairview Ave N
Suite 2000
Seattle, Washington 98109

**www.nanostring.com**

**Tel:**    206.378.6266
888.358.6266

**E-mail:** info@nanostring.com

## Molecules That Count®

## FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

### Intellectual Property Rights

This nCounter® Analysis System manual and its contents are the property of NanoString Technologies, Inc. ("NanoString"), and is intended solely for the use of NanoString customers, for the purpose of operating the nCounter Analysis System. The nCounter Analysis System (including both its software and hardware components) and this User Guide and any other documentation provided to you by NanoString in connection therewith, are subject to patents, copyright, trade secret rights and other intellectual property rights owned by, or licensed to, NanoString. No part of the software or hardware, may be reproduced, transmitted, transcribed, stored in a retrieval system, or translated into other languages without the prior written consent of NanoString.

### Limited License

Subject to the terms and conditions of the nCounter Analysis System contained in the product quotation, NanoString grants you a limited, non-exclusive, non-transferable, non-sublicensable, research use only license to use the proprietary nCounter Analysis System only in accordance with the manual and other written instructions provided by NanoString. Except as expressly set forth in the terms and conditions, no right or license, whether express, implied or statutory, is granted by NanoString under any intellectual property right owned by, or licensed to, NanoString by virtue of the supply of the proprietary nCounter Analysis System. Without limiting the foregoing, no right or license, whether express, implied or statutory, is granted by NanoString, to use the nCounter Analysis System with any third party product not supplied or licensed to you by NanoString, or recommended for use by NanoString in a manual or other written instruction provided by NanoString.

### Trademarks

NanoString Technologies, NanoString, nCounter and Molecules That Count are registered trademarks or trademarks of NanoString Technologies, Inc., in the United States and/or other countries. All other trademarks and/or service marks not owned by NanoString that appear in this manual are the property of their respective owners.

### Copyright

© 2011 NanoString Technologies, Inc. All rights reserved.

**nanoString**
TECHNOLOGIES

# CONTENTS

nanoString
T E C H N O L O G I E S

## PREFACE

### Purpose

This manual describes in detail the data output by the nCounter Digital Analyzer and provides guidance for interpreting and troubleshooting data generated with custom nCounter Gene Expression, miRGE CodeSets and miRNA Panel assays.

### Conventions Used

The following conventions are used throughout this manual and are described below for your reference:

### Fonts

Special font formatting is used in this manual. Such formatting conventions are used in specific instances as described below:

**TIP**   Information contained in a Tip may offer helpful suggestions, alternative procedures, methods and/or shortcuts.

**NOTE**   This note type emphasizes general information.

**CAUTION**   This note type presents essential content indicating that the potential exists for assay failure, diminished data quality, and/or a loss of data if the information presented is ignored.

**WARNING**   This note type indicates that a potential hazard to your personal safety, or the potential for equipment damage exists.

**BOLD**   When appearing in text or in a procedure, the bold text serves to highlight a specific button, key stroke, or menu option available.

• **Bold** text may appear elsewhere to highlight important text or terms.

• **Green** text is used to help the reader identify active hyperlinks.

**ITALICS**   Used to emphasize an important word or expression within the text.

• Formatting of a book title, journal, or other documentation.

• Used to indicate the special or unusual meaning of a word or phrase.

### Contact Information

**NanoString Technologies, Inc.**

530 Fairview Ave N
Suite 2000
Seattle, Washington 98109 USA

**Tel:**   206.378.6266

888.358.NANO (6266)

**Fax:**   206.378.6288

**E-mail:**  support@nanostring.com

# 1 Data Analysis

## Introduction

NanoString Technologies nCounter assays are designed to provide a single-tube, ultra-sensitive, reproducible, and highly multiplexed method for detecting nucleic-acid targets across all levels of biological expression. These assays provide a method for direct detection of targets with molecular barcodes without the use of reverse transcription or amplification. nCounter assays are processed on the fully automated Prep Station and data are collected and tabulated by the Digital Analyzer in sets of twelve. This manual describes the data output by the nCounter Digital Analyzer and provides guidance for interpreting and troubleshooting your data.

## Assay QC Metrics

All nCounter data contain imaging quality control (QC) metrics that should be assessed prior to analysis and interpretation of data. See the Troubleshooting Guide *(Chapter 2)* or contact NanoString Support for help in identifying the cause of deviations in your data.

### Field of View (FOV) Counted

The nCounter Digital Analyzer images each lane in discrete units, called fields of view (FOV). Optical issues, such as an inability to focus due to bubbles or insufficient oiling of the cartridge, can prevent successful imaging of a FOV. The Digital Analyzer reports the number of FOVs successfully imaged as FOV Counted. Significant discrepancy between the number of FOV for which imaging was attempted (FOV Count) and for which imaging was successful (FOV Counted) may be indicative of an issue with imaging performance.

### Binding Density

In order to count accurately, the Digital Analyzer only counts the codes that are unambiguously distinguishable. If codes overlap with each other in an image, the Digital Analyzer does not interpolate what those codes are; rather, it simply does not count them. This provides increased confidence that the molecular counts you receive are from truly recognizable codes. Under most conditions, discarding a few unrecognizable codes does not impact the data. However, if there are too many overlapping codes in the image, significant data loss could occur. To determine image saturation, the nCounter Digital Analyzer calculates a Binding Density for each lane as it processes the images. The Binding Density is a measure of the number of optical features per square micron. The Binding Density is useful for determining whether or not data collection has been compromised due to image saturation.

Typically, the range for binding density will be between 0.05 and 2.25. In this range, reporter distribution on the slide surface is such that very few, if any, reporters on the slide surface overlap, enabling the Digital Analyzer to accurately tabulate counts for each reporter species. A binding density greater than 2.25 is indicative of a large number of overlapping reporters on the slide surface. Thus, the counts observed in lanes with a Binding Density greater than 2.25 have had significant numbers of codes ignored, affecting quantification, and potentially, linearity of the assay. Note that because the Binding Density metric includes many system controls that are independent of your sample, it is not a good indicator of how well your sample performed and should only be used as an imaging QC.

# CodeSet Internal Controls

The External RNA Control Consortium (ERCC) is a group of industry representatives established to develop RNA control transcripts that can be used to assess technical performance in gene expression assays. NanoString has adopted sequences developed and tested by the ERCC for positive and negative hybridization controls. The ERCC control sequences are not homologous to any known organism, are applicable and transferable in all CodeSets, and generate consistent results in gene expression analyses. Reporter probes designed against ERCC transcript sequences are pre-mixed into every CodeSet, and are therefore available for use in data analyses.

> **ⓘ  CAUTION:** If you plan to use additional ERCC transcripts in your experiments, please contact NanoString to verify that they are compatible with assay controls before proceeding.

**TABLE 1:** Internal assay controls by assay type.

| Assay Type | Positive Controls (POS) | Negative Controls (NEG) | Other Controls |
|---|---|---|---|
| mRNA Gene Expression | 6 hybridization | 8 hybridization | |
| miRGE | 6 hybridization<br>2 ligation / hybridization | 8 hybridization | |
| miRNA Panels | 6 ligation / hybridization | 8 ligation / hybridization | Between 4-5 mRNA hybridization house-keeping genes<br><br>2 optional additional non-mammalian miRNA spike-in controls |

## mRNA Gene Expression CodeSets

Each mRNA Expression CodeSet contains probes designed against fourteen ERCC transcript sequences. Six of these sequences are used as positive hybridization controls and eight are designed as negative controls.

For each positive control, in vitro transcribed RNA targets are pre-mixed with the Reporter CodeSet during manufacturing. Six different transcripts are provided, each at one of the following concentrations in the 30 µL hybridization reaction: 128 fM, 32 fM, 8 fM, 2 fM, 0.5 fM, and 0.125 fM. This range of concentrations corresponds to the expression levels of most mRNAs of interest present in 100 ng of total RNA.

For ERCC negative control probes, target transcripts are absent. These negative controls can be used to estimate the non-specific background in your experiment, as described in the section titled *Assessing Background.*

## miRGE CodeSets

Each miRGE CodeSet contains probes designed against sixteen ERCC transcript sequences. Six of these sequences are used as positive hybridization controls, two are used as ligation controls and eight are designed as negative controls.

For each positive hybridization control, in-vitro transcribed RNA targets are pre-mixed with the Reporter CodeSet during manufacturing. Six different transcripts are provided, each at one of the following concentrations in the 35 µL hybridization reaction: 110 fM, 27.5 fM, 6.9 fM, 1.7 fM, 0.4 fM and 0.1 fM.

For each ligation control, synthetic small RNAs, similar to miRNAs, are provided for addition to the miRNA sample preparation reaction as described in nCounter miRGE Assay Manual and should be present in the hybridization at 110 fM and 27 fM respectively. These small RNA controls must be ligated and hybridized in the same manner as an endogenous miRNA; thus, they act as controls for the miRNA species in the experiment, from sample preparation through hybridization.

For ERCC negative control probes, target transcripts are absent. These negative controls can be used to estimate the non-specific background in your experiment as described in the section titled *Assessing Background.*

## miRNA Panel Assay Kit

Each miRNA Panel Assay Kit contains probes designed against fourteen ERCC transcript sequences. Six of these sequences are utilized as positive ligation controls and eight are utilized as negative controls. For each positive control, a separate ERCC RNA sequence of similar size to a miRNA is supplied. Six different small RNA controls are provided such that after dilution and addition to the miRNA sample preparation reaction, each will be at one of six different concentrations in the 30 μL hybridization: 128 fM, 32 fM, 8 fM, 2 fM, 0.5 fM, and 0.125 fM. These controls must be ligated and hybridized in the same manner as an endogenous miRNA; thus, they act as controls for the entire miRNA analysis process from sample preparation through hybridization.

The miRNA Panel Assay Kit also contains probes for a set of species-specific mRNA housekeeping genes that do not have to be ligated to be counted.

Additionally, the miRNA Panel Assay Kit contains probes designed against non-mammalian small RNAs that can be used to measure RNA purification efficiency if so desired.

For ERCC negative control probes, target transcripts are absent. These negative controls can be used to estimate the non-specific background in your experiment as described in the section titled **Assessing Background.**

# Normalization

The data produced by the nCounter Digital Analyzer are exported as a Reporter Code Count (RCC) file. RCC files are comma-separated text (.csv) files that contain the counts for each gene in a sample. The data for each sample hybridization are contained in a separate RCC file. Prior to comparing data between hybridizations, slight differences in hybridization, purification, binding efficiency and other experimental variables must be normalized. To accomplish this, NanoString recommends utilizing the internal positive controls that are present in each CodeSet. Since these targets are independent of the sample, normalizing using these controls will help to eliminate variability unrelated to the sample.

## Positive Control Normalization

Positive control normalization can be used to normalize all platform associated sources of variation (e.g. automated purification, hybridization conditions, etc). This type of normalization will not account for differences in sample input between technical/biological replicates. If the calculated positive control scaling factor described below is outside a range of 0.3 - 3, it may indicate significant under-performance of a lane or lanes. Care should be taken when interpreting results from such experiments.

### Method

The positive spike-in RNA hybridization controls for each lane may be used to estimate the overall efficiency of hybridization and recovery for each lane. To do so, a lane specific value representative of positive control counts should be calculated, e.g. sum of positive control counts, geometric mean of control counts (see explanation of geometric mean below), etc. The average of these calculated values across all lanes is used as the reference against which each lane is normalized. A scaling factor is then calculated for each of the lanes based on the calculated value for the positive controls in each lane relative to the average of this value for the positive controls across all lanes. This normalization factor may then be used to adjust the counts for each gene target and negative controls in the associated lane.

> **NOTE:** For miRGE CodeSets, NanoString recommends normalizing mRNA probes relative to the six hybridization-only controls and the miRNA probes relative to the ligation controls.

nanoString
TECHNOLOGIES

**WORKFLOW EXAMPLE**

1. For each lane (sample), calculate the sum, average, or geomean of the positive controls.
2. Calculate the average of the sums, averages, or geomeans across all lanes.
3. Divide this average by the sum, average, or geomean in each lane to calculate a lane-specific scaling factor.
4. Multiply all the counts in a lane (POS, NEG and all gene counts) by the lane-specific scaling factor.

Following this technical standardization, additional normalization may be required prior to analysis. The best normalization strategy for your application will ultimately depend on experimental setup. Several simple, effective normalization procedures are outlined below along with a brief discussion of their merits and limitations.

## Reference (Housekeeping) Gene Normalization

Reference gene normalization may be performed in order to adjust counts of all probes relative to a probe (or set of probes) that are not expected to vary between samples or replicates. Reference gene normalization assumes that some of the targets sequences recognized by the CodeSet are consistent in their expression levels. The choice of reference genes to include in the CodeSet is therefore a crucial part of experimental design. Nanostring recommends at least three reference genes, but the more that are included, the more accurate the normalization will be. For advice in choosing reference genes, please contact Nanostring support or e-mail **support@nanostring.com**.

CodeSet content normalization corrects for differences in sample input between assays and is typically performed after positive control normalization. While reference genes will also be affected by systemic variability, a single-step normalization using only the reference genes will not distinguish systemic variability from sample variability.

### Method

Because reference genes are often expressed at different levels, the geometric mean of the reference genes for each lane should be utilized to calculate scaling factors. Like a simple average (arithmetic mean), geometric mean indicates the central tendency of a set of numbers. However, unlike a simple average, the geometric mean is less sensitive to variation in the magnitude of count levels between probes. A geometric mean is obtained by multiplying each value in a data set and then taking the nth root (where n is the count of numbers in the set) of the resulting product.

$$\text{Geometric Mean} = \sqrt[n]{x^1 x^2 \cdots x^n}$$

The average of these geometric means across all lanes is used as the reference against which each lane is normalized. A normalization factor is then calculated for each of the lanes based on the geometric mean of counts for the reference gene(s) in each lane relative to the average geometric mean of counts for the reference gene(s) across all lanes. This normalization factor is then used to adjust the counts for each gene target and controls in the associated lane.

**WORKFLOW EXAMPLE**

1. For each lane calculate the geometric mean of the reference genes.
2. Calculate the average of the geometric means across all lanes.
3. Divide this average by the geometric mean in each lane to get a lane-specific normalization factor.
4. Multiply all the endogenous counts in a lane by that lane's normalization factor. Do not multiply the internal assay controls by this factor.

## Global Normalization

With CodeSets containing a large number of probes (> 300), it may be assumed that while some targets may increase or decrease in any given sample, these are likely to represent a small portion of the total number of targets assayed and that the overall level of expression within a sample will be the same. In these cases, global normalization methods which utilize large numbers of reporters to generate normalization factors may be employed.

> **NOTE:** It is important to note that if a significant fraction of probes exhibit differential expression from sample to sample, this method may not be appropriate.

## Total Counts

The sum of all counts in each sample can serve as an approximation of the total nucleic acid expression within a sample. With this method, a normalization factor is generated using the sum of all counts in a lane relative to the mean sum of all counts in each lane. However, if a large number of targets vary, or if a few vary by orders of magnitude, this method may skew the data inappropriately and not produce representative results. This method works best to compare similar sample types (cells or tissues) with similar overall expression profiles.

### WORKFLOW EXAMPLE

1. For each lane, calculate the sum of all the endogenous gene probe counts. Do not include the internal positive and negative controls in this calculation.
2. Calculate the average of the sums across all lanes.
3. Divide this average by the sum in each lane to calculate a lane-specific normalization factor.
4. Multiply all the probe counts in a lane by the lane-specific normalization factor.

## Average Counts

Similar to the method described above for total counts, the average of all counts in each sample can serve as an approximation of the total nucleic acid expression within a sample. With this method, a normalization factor may be generated using the mean of all counts in a lane relative to the average mean of all counts in each lane. This normalization method has limitations similar to the Total Counts normalization method.

### WORKFLOW EXAMPLE

1. For each lane, calculate the sum of all the endogenous gene probe counts. Do not include the internal positive and negative controls in this calculation.
2. Calculate the average of the sums across all lanes.
3. Divide this average by the sum in each lane to calculate a lane-specific normalization factor.
4. Multiply all the probe counts in a lane by the lane-specific normalization factor.

## Geometric Mean

With this method, a normalization factor should be generated using the geometric mean of the approximately 50-100 targets with the highest counts. This minimizes the impact that non-expressed targets might have on the normalization factor, i.e. the majority of system noise is excluded from contributing to normalization. To normalize to the geometric mean, probe counts are reordered from most to least abundant and then a normalization factor is calculated using only robustly expressed genes. As with normalization to the sum of all counts described above, this method works best to compare similar sample types (cells or tissues) with overall similar miRNA expression profiles.

### WORKFLOW EXAMPLE

1. Reorder the genes from highest to lowest abundance. Make sure the order is identical in all samples. For this reordering, the average expression of each gene could be used as an index of reordering (preferred), or a single reference lane could be used as the index.
2. For each lane, calculate the geometric mean of the top 50-100 expressing genes.
3. Calculate the average of the geometric means across all lanes.
4. Divide this average by the geometric mean in each lane to calculate a lane-specific normalization factor.
5. Multiply all the counts in a lane by the lane-specific normalization factor.

## nanoString
### TECHNOLOGIES

# Assessing Background

Accurate estimation of probe background is essential for interpreting expression data. NanoString includes several probes in each CodeSet for which no target is expected to be present. These negative controls can be used to estimate background values in your experiment, but may not capture minor probe-specific differences in background for every probe in a CodeSet.

## Using the Negative Control Probes

Each CodeSet includes several reporters for which no transcript is supplied. These probes are designed against ERCC sequences using the same probe design criteria as all other reporters in a CodeSet. These negative controls can therefore be used to estimate systematic background counts within any single hybridization reaction. Background estimation using these probes may be performed in many ways, each with an associated affect on stringency and false discovery rate. Below is an example of several methods that may be used to estimate background.

## Mean of Negative Control Probes

The average of all the negative controls in a lane could be defined as background for that lane. If this value is used, false discovery will be expected to be high, since by definition half of the known negative probes will exhibit counts above the average background value.

**WORKFLOW EXAMPLE**

For each lane, calculate the average of the NEG control counts. This is the background threshold.

## Mean Plus Standard Deviation of Negative Control Probes

The average plus a multiple of the standard deviation of all negative control probe counts in a lane could be defined as background for that lane. In this case, the false discovery would be expected to be lower to much lower than a simple average, depending on the multiple of the standard deviation used. For example, average plus three standard deviations will be more stringent than average plus two standard deviations, reducing the false positive rate (but potentially increasing the false negative rate).

**WORKFLOW EXAMPLE**

1. For each lane, calculate the average of all the NEG control counts.
2. For each lane, calculate the standard deviation of all the NEG control counts.
3. Add a multiple of the standard deviation to the average. For example, multiply the standard deviation by two before adding to the average.
4. This is the background threshold.

## Highest Negative Control Probe Value

Counts for the negative control probe exhibiting the greatest value within a lane is chosen as background. This method may significantly reduce the rate of false discovery relative to using the mean.

**WORKFLOW EXAMPLE**

For each lane, determine the maximum of the NEG counts. This is the background threshold.

## Using Candidate Targets as Negative Controls

If your CodeSet contains probes for targets which are known to have no expression (such as in a gene knock-out study) in your samples of interest, it is possible to use these probes to determine background in much the same manner as described above for the NanoString supplied negative control probes. These candidate target negatives may be used alone or in conjunction with the NanoString supplied negative controls. As more negative control probes are used to estimate background, confidence that the determined background threshold accurately represents system background increases.

## Using an RNA-free Control

To use this method, you will need to perform nCounter assays without sample input (blanks), ideally in triplicate so that statistical tests may be applied *(see the section below for an example of how an RNA-free input assay may be used to assess detection)*. This will generate a measurement of how each individual reporter interacts with components of the system during an experiment. This is the most robust means of determining background and a means to assess background in a probe-specific manner.

## Background Subtraction

Once a background threshold has been determined (either in a probe-specific or more global manner), it is possible to subtract out background counts in order to determine true counts. In most cases this is unnecessary, especially when overall background is low. In cases where probe-specific background is high, background subtraction may improve downstream data analysis, such as fold-change estimation. In general, background subtraction will amplify estimations of fold-change.

**NOTE:** For probes exhibiting counts near background levels, subtraction may result in negative or zero values being obtained. If fold change determination for such a target is desired, the value for this target should be set equal to the background level in that lane (or some other arbitrary value, such as 1). Fold change determinations of genes at or near background levels should be considered an estimate.

## Differences Between mRNA and miRNA Probes

nCounter miRNA probes rely on end ligation of specific DNA oligonucleotides to each miRNA, a process described in the *nCounter miRNA Assay Manual* and *nCounter miRGE Assay Manual*. Due to this ligation event, average background for miRNA probes can differ from mRNA probes.

nanoString
T E C H N O L O G I E S

## Determining the Presence or Absence of Target Transcripts

Once an estimation of background has been obtained, a presence/absence call for any given reporter probe target may be generated. As with the determination of background, there are many different methods for assessment of target presence/absence and the best method is likely to depend on experimental design and nature of the question addressed by a particular experiment. Below are some methods of making presence/absence calls.

### Using the Background Threshold

The simplest method of determining presence/absence of a target is to make a comparison of gene counts generated relative to background counts for that lane. This method of making presence/absence calls may be utilized regardless of the number of replicates per sample, but may have limited statistical power depending on the method used to determine background *(see the section above for different methods of estimating the background threshold)*. When using this method, it is important that all comparisons are between normalized candidate target reporter probes and normalized background counts for that lane.

#### WORKFLOW EXAMPLE

For each lane, determine whether the gene counts are greater than the background threshold count value for that lane. If they are, then expression for that gene is detected in that sample.

### Statistical Tests

To determine with statistical significance whether a target transcript is present in a sample, technical or biological replicate assays are necessary (minimum n=3). After normalizing replicate assays, the probability that the target is above background can be found by a variety of means. Below is an illustration of how a Welch's *t*-test may be used.

### *t*-Test

The Welch's *t*-test can be used to assess the likelihood that a set of gene counts is significantly different than the set of background counts. This type of *t*-test does not make an assumption of equal variance between test populations, but does assume that variance within each test population is normally distributed. If the *p*-value of a one-tailed, heteroscedastic (normalized replicate gene counts vs. all negative control counts) is less than 0.05, and the counts for the target are greater than the background threshold, then there is at least a 95% probability that the counts for that gene are significantly above the background threshold. The stringency of the presence/absence test is affected by the choice of *p*-value confidence level and method of determining a background threshold. For example, a choice of $p<0.05$ and background defined as the average of the negative controls will be less stringent than a choice of $p<0.01$ and background equal to the highest negative control probe counts.

#### WORKFLOW EXAMPLE

1. Using a one-tailed, heteroscedastic *t*-test, calculate a *p*-value comparing the replicate gene counts to the background counts.

2. Calculate the average of the replicate gene counts.

3. If the average replicate gene counts are greater than the background threshold, and the *p*-value is less than your acceptable threshold confidence level, then the gene is detectable in the samples.

## Using RNA-free Input Assays

If you have run an RNA-free (blank) assay, the presence or absence of transcripts may also be determined by comparing the counts for each gene in the presence of RNA to the counts for that gene in the absence of RNA. The difference between counts given for each code in the presence and absence of RNA can then be tested for statistical significance.

### WORKFLOW EXAMPLE

1. To use this method, you will need to perform nCounter analyses in triplicate using your CodeSet, but no sample input. This will generate a measurement of how each individual code interacts with components of our system during an experiment.

2. Normalize the counts obtained in the RNA-free assays to the positive spikes as described above in the Positive Control Normalization section.

3. Perform a one-tailed heteroscedastic $t$-test comparing the replicate counts obtained in the presence of RNA to the replicate counts obtained in the absence of RNA. This will give you a $p$-value for this test.

4. For each gene, subtract the average count in the RNA-free assays from the average count in the samples containing RNA. This subtracts out the unique background for each code from the sample measurements using that code.

5. As described above, if the $p$-value from the $t$-test is less than 0.05, and the background subtracted measurement is greater than zero, then the gene of interest is detected in that sample with a 95% confidence level.

## Calculating the Percent Coefficient of Variation

In replicate (technical or biological) measurements (minimum, n=3), it may be desirable to analyze the variance of measurement in order to assess the overall precision of the counts for any probe. This measure of probe precision can be expressed as the coefficient of variation percentage (%CV) of the replicates. Determination of %CV can be helpful in identifying experimental outliers and for any given reporter probe is expected to decrease as counts increase.

### WORKFLOW EXAMPLE

1. For each gene calculate the average counts across a set of samples.

2. For the same genes, calculate the standard deviation across the same set of samples.

3. For each gene, divide the standard deviation by the average and then multiply by 100 to obtain %CV.

nanoString
TECHNOLOGIES

# 2    Troubleshooting Guide

This table presents potential solutions to problems that might arise with nCounter data. For further questions or explanations, please contact your Field Applications Scientist (FAS) or **support@nanostring.com**. Please have the log files downloaded and ready to send for any questions about nCounter data or system performance. Issues are marked in color to indicate applicable assay in the following manner:

🟠 **Gene Expression Custom CodeSets and Panels**

🔵 **miRNA Panels**

🟢 **miRGE CodeSets**

**TABLE 2:** Troubleshooting guide.

| | Issue | Diagnosis | Possible Cause | Possible Solution |
|---|---|---|---|---|
| 🟠🔵🟢 | No counts for POS, NEG and endogenous probes | FOVCounted = 0 Binding density = 0 | Inability to focus | Clean oil off objective and cartridge. Apply fresh oil to each, re-scan. |
| | | FOVCounted is high, binding density is at least 0.02 | Part of CodeSet left out | Repeat hybridization. |
| | | FOVCount and binding density cannot be obtained | Scanner malfunction | Download log files and contact support@nanostring.com |
| 🟠🔵🟢 | Low counts for POS, NEG and all endogenous probes | FOVCounted is low, binding density at least 0.02 | Inefficient focusing | Clean oil off objective and cartridge. Apply fresh oil to each, re-scan. |
| | | FOVCounted is high, binding density is at least 0.02 | RNAse contamination; Organic solvent contamination | Repeat hybridization after cleaning up sample. |
| | | Cracked coverslip | Cartridge dropped; cartridge subjected to extreme temperature variations | Examine bottom of cartridge for defects.<br><br>Download log files of scan and contact support@nanostring.com |
| | | CodeSet concentration too low, or CodeSet was degraded | Spinning CodeSet too hard, or vortexing CodeSet when setting up hybridization | Do not spin tubes containing CodeSet in anything but picofuges.<br><br>Mix gently by flicking tubes. |

**TABLE 2 (continued):** Troubleshooting guide.

| | Issue | Diagnosis | Possible Cause | Possible Solution |
|---|---|---|---|---|
| ● ● ● | Low counts for POS, NEG and all endogenous probes | | Scanner malfunction | Download log files and contact support@nanostring.com |
| ● | Low counts for POS and all endogenous probes | miRNA counts low, Housekeeping control counts high | Sample contamination causing ligation problem | Check purity of starting RNA samples; A260/A230 should be greater than 1.8.<br><br>Purify sample and repeat ligation and hybridization. |
| | | miRNA counts low, Housekeeping control counts high | Ligase not added | Repeat ligation and hybridization. |
| | | miRNA counts low, Housekeeping control counts high | Thermocycler problem | Check programs used and calibration of thermocycler.<br><br>Repeat ligation and hybridization on another machine. |
| ● ● ● | Low counts for endogenous probes, but POS counts high | System controls look normal, endogenous probe counts near background | Sample not added | Repeat hybridization. |
| | | System controls look normal, endogenous probe counts highly variable | RNA in sample extensively degraded | Repeat hybridization using more sample if available |
| ● | Low counts for ligation controls, but POS counts high | mRNA counts high, miRNA counts low | Sample contamination causing ligation problem | Check purity of starting RNA samples; A260/A230 should be greater than 1.8.<br><br>Purify sample and repeat ligation and hybridization. |
| | | mRNA counts high, miRNA counts low | Ligase not added | Repeat ligation and hybridization |
| | | mRNA counts high, miRNA counts low | Thermocycler problem | Check programs used and calibration of thermocycler.<br><br>Repeat ligation and hybridization on another machine. |

**nanoString**
TECHNOLOGIES

**TABLE 2 (continued):** Troubleshooting guide.

| | Issue | Diagnosis | Possible Cause | Possible Solution |
|---|---|---|---|---|
| 🟠🔵🟢 | All NEG counts are high | POS counts high, NEG counts high | Non-specific probe aggregation | Repeat experiment, setting up hybridizations at room temperature. |
| | | POS counts high, NEG counts high | Capture probe added too soon | Repeat experiment, making sure Capture Probes are added last and just before placing at 65°C. |
| | | POS counts low, NEG counts high | Degraded CodeSet | Repeat experiment with CodeSet stored at -80°C until just before use |
| 🔵🟢 | | POS counts high, NEG counts high | Reduced ligation specificity | Check programs used and calibration of thermocycler. Repeat ligation and hybridization on another machine. |
| 🟠🔵🟢 | One or two NEG counts are high, rest are low | POS counts high, one or two NEG counts high, endogenous counts look normal | Potential cross-hybridization with another sequence in sample | Continue with analysis, ignoring the sequences potentially cross-hybridizing. |
| 🟠🔵🟢 | POS, NEG, Endogenous counts don't make sense | Counts for all probes do not fit with expected control behavior or expected biology | Wrong RLF used | Re-scan cartridge, using correct RLF file (including version number). |
| | | Counts for all probes do not fit with expected control behavior or expected biology | Wrong CodeSet used | Repeat hybridization using correct CodeSet. |
| 🔵🟢 | One or two endogenous probe counts don't make sense | Counts for a few probes do not fit expectations, others do | Adjustment for probe specific or ligation-mediated background not performed | Repeat analysis using nSolver (miRNA) or making sure sample-free background counts are subtracted before normalization. |
| 🟠🟢 | | Counts for a few probes do not fit expectations, others do | Probe was designed against incorrect sequence | Contact support@nanostring.com |

## References

- Tong W. et al. Evaluation of external RNA controls for the assessment of microarray performance. Nature Biotechnology v.24 no.9 (1132-1139), 2006.
- Mestdagh P. et al. A novel and universal method for microRNA RT-qPCR data. Normalization. Genome Biology, 10:R64, 2009.
- Vandesompele et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:RESEARCH0034, 2002.

# 3 NanoString Glossary

This chapter contains definitions of terminology associated with the nCounter Analysis System and RCC file data fields.

## General

| Term | Definition |
| --- | --- |
| Assay | Each nCounter assay is capable of simultaneous, single-tube, multiplexed detection of up to 800 targets. |
| Capture Probe | Biotinylated oligonucleotide complementary to the 3' portion of a target sequence. |
| Reporter Probe | Fluorescently barcoded oligonucleotide complementary to the 5' portion of a target sequence. |
| CodeSet | A collection of Capture Probes and Reporter Probes designed against specific target sequences. |
| miRNA Tag Reagent | A collection of oligonucleotides that facilitate specific ligation of individual miRNAs to a miRNA-specific tag in order to generate CodeSet targets. |
| Prep Station | An automated liquid-handling robot that utilizes magnetic bead-based purification to remove unbound CodeSet, non-target cellular transcripts and other cell debris after sample hybridization. In addition to hybridization purification of target/probe complexes, the Prep Station also automates immobilization of the sample onto an imaging surface for subsequent data collection. |
| Cartridge | A micro-fluidic device containing 12 lanes that allows distribution and immobilization of nCounter probes on a solid surface for subsequent imaging and counting by the nCounter Digital Analyzer. |
| Lane | Area of the cartridge surface over which nCounter probes are immobilized. Each lane is 2600 $\mu$m × 8000 $\mu$m. |
| Field of View (FOV) | Area of the cartridge surface discretely imaged by the Digital Analyzer. A single image the Digital Analyzer uses to count molecules. |
| Digital Analyzer | The nCounter Digital Analyzer collects data by taking images of the immobilized fluorescent reporters in the sample cartridge with a CCD camera through a microscope objective lens. The Digital Analyzer may collect data at one of four resolutions: max, high, medium and low, yielding data for hundreds of thousands of target molecules. The number of images (FOV) taken correlates linearly with the number of reporters counted and this, in part, determines the dynamic range and level of sensitivity in the system. Images are processed internally by the Digital Analyzer and the results are exported as a file that can be downloaded via memory stick. |
| Cartridge Definition File (CDF) | A file used by the Digital Analyzer to associate user-input sample information and a RLF with a cartridge. |

| Term | Definition |
|---|---|
| Reporter Library File (RLF) | A file used by the Digital Analyzer to associate target name and annotation with a unique fluorescent barcode after imaging and counting. |
| Reporter Code Counts (RCC) File | A data file (.csv) output by the Digital Analyzer that contains sample information, probe information and probe counts. |
| CodeSet Content | Probes within a CodeSet that hybridize to sequences of interest. For custom Gene Expression and miRGE CodeSets, this refers to the customer defined endogenous and housekeeping genes/probes. For miRNA panels, this refers to all miRNA and mRNA housekeeping probes. ERCC positive and negative control probes are not considered part of the CodeSet Content. |

## nCounter Digital Analyzer Data Output Terms

| Term | Definition |
|---|---|
| File Attributes | Data fields imported from CDF. |
|    File Name | RCC file name. Unique for each lane. |
|    Sample ID | Sample identifier specified by the user in the CDF. |
|    Sample Date | Date information specified by the user in the CDF. |
|    File Version | RCC Collector version number. |
|    GeneRLF | RLF used to associate gene names with counts specified by the user in the CDF. |
|    Comments | Text annotation field specified by the user in the CDF. |
| Lane Attributes | Data fields generated by the Digital Analyzer during a scan. |
|    Lane ID | Numerical identifier of imaged lane (always a value from 1 to 12). |
|    FOV Count | Total number of FOVs imaged per lane. |
|    Scanner ID | Numerical identifier of the Digital Analyzer. |
|    Stage Position | Numerical identifier stage position / cartridge scanned (always a value from 1 to 6). |
|    Binding Density | Number of fluorescent spots per $\mu m^2$ of the lane surface. |
|    Messages | Text output field for Digital Analyzer QCs. A binding density greater than 2.25 will generate a warning indication. |
| Reporter Counts | Data fields containing nCounter Reporter Probe information. |
| Code Class | Text field defining nCounter Reporter Probe pairs. |
|    Positive | A probe pair that hybridizes to synthetic RNA targets present in the hybridization reaction at known concentrations. |
|    Negative | Reporter probe pair for which no known target should exist within the endogenous sample. |
|    Housekeeping | Reporter probe pair for an endogenous target expected to vary minimally in expression level from sample to sample. |
|    Endogenous | Reporter probe pair for a target queried within a given sample. |
|    Message | Text field detailing important information to the end-user. |
| Name | Text field defining the gene name associated with an nCounter® Reporter Probe pair. |
| Accession | Text field defining the gene accession information associated with an nCounter Reporter Probe pair. |

**NanoString Technologies, Inc.**

530 Fairview Ave N
Suite 2000
Seattle, Washington 98109

**CONTACT US**

info@nanostring.com
**Tel:** (888) 358-6266
**Fax:** (206) 378-6288
www.nanostring.com

**SALES CONTACTS**

| | |
|---|---|
| United States: | **us.sales@nanostring.com** |
| Europe: | **europe.sales@nanostring.com** |
| Japan: | **japan.sales@nanostring.com** |
| Other Regions: | **info@nanostring.com** |