Reference Genes for Normalization of Expression Data

Overview

The accuracy and reliability of gene expression results are dependent upon the proper normalization of the data against internal reference genes. Usually one or more "housekeeping" genes are chosen as reference, since they often display uniform expression during various phases of development, across different tissue types, and under different environmental and experimental conditions. Artifactual changes in gene expression measurements can be avoided by assuring that equal numbers of cells are used in each experiment. However, controlling the number of cells used can be difficult when assaying tissue samples such as biopsies. Controlling the amount of total RNA used in each experiment can also reduce errors, but mRNA content per unit mass of total RNA may vary between samples, and quantification of total RNA in very small samples can be problematic. Differences in sample preparation, such as variable degradation of RNA, can also introduce error. Normalization against internal reference genes can help correct for these differences in numbers of cells, absolute mRNA content and sample preparation.

nCounter™ Analysis System

The NanoString™ nCounter Analysis System delivers direct, multiplexed measurement of gene expression, providing digital readouts of the relative abundance of hundreds of mRNA species simultaneously. The nCounter Analysis System is based on gene-specific probe pairs that are hybridized to the sample in solution. The protocol obviates any enzymatic reactions that might introduce bias in the results (Figure 1, step 1). The Reporter Probe carries the fluorescent signal; the Capture Probe allows the complex to be immobilized for data collection. Up to 550 pairs of probes specific for a particular set of genes are combined with a series of internal controls to form a CodeSet. After hybridization of the CodeSet with target mRNA, samples are transferred to the nCounter Prep Station (Figure 1, step 2) where excess probes are removed and probe/target complexes are aligned and immobilized in the nCounter Cartridge. Cartridges are then placed in the nCounter Digital Analyzer for data collection (Figure 1, step 3). Each target molecule of interest is identified by the "color code" generated by the ordered fluorescent segments present on the reporter probe. The reporter probes on the surface of the cartridge are then counted and tabulated for each target molecule

Reference Genes Available for Use with the nCounter™ Analysis System

Nanostring has characterized a set of human reference genes across a wide variety of tissue types on the nCounter analysis system. Genes with low, medium and high expression levels were selected to provide a diverse set that will give optimal performance with the nCounter system. We have chosen genes that are commonly used in gene expression profiling experiments. The table in the appendix provides the NCBI RefSeq database accession number, gene ID, the official gene symbol and full name for each reference gene in the CodeSet.

To illustrate the expression levels determined using the reference gene CodeSet, we hybridized the CodeSet with 100 ng of total RNA from each of 12 human tissues (purchased from Stratagene and BioChain Institute, Inc.) using the standard nCounter protocol. The counts from each hybridization were normalized to the internal positive controls to account for slight differences in assay efficiencies (see nCounter Data Analysis Guidelines). The expression levels for the reference genes across the 12 tissues were divided into 3 groups: low, medium and high expressers (Figure 2).

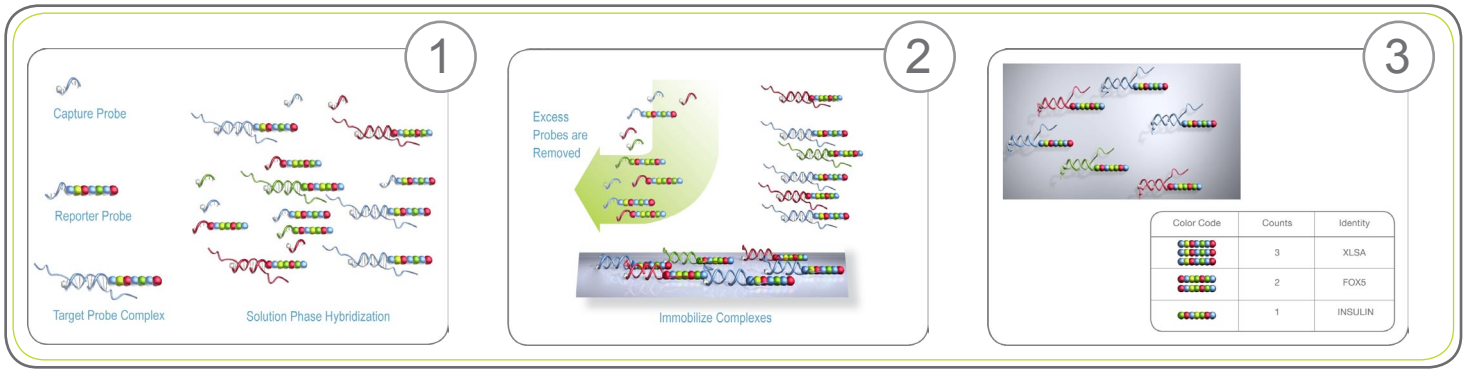

Figure 1. The Nanostring™ technology.

Figure 2. Expression levels of the 20 reference genes, in counts, across 12 human tissues, and divided into three groups: high (A), medium (B) and low (C) expressers.

These reference genes are available as a single CodeSet, or they can be selected individually as part of a user-defined custom CodeSet. This is not meant to be an all-inclusive set. If you would like to use reference genes not included in this list, we can work with you to select the best probes for a CodeSet including the reference genes of your choice.

Using Reference Genes in Data Analysis

Reference genes can be particularly useful for accurate assessment of expression levels in samples containing degraded mRNA. There are many valid approaches for normalization. Here, we illustrate how a single reference gene can be used in a simple linear normalization to correct for RNA degradation in an nCounter experiment. We used PGK1 as a reference gene since its expression does not vary significantly between the tissues being studied.

For this example, a CodeSet was manufactured containing 48 genes extensively characterized by the Microarray Quality Control (MAQC) consortium and the 18 Nanostring reference genes. The expression levels for these genes were measured in triplicate for total RNA from brain tissue and heart tissue. One sample of heart tissue RNA was degraded by heating at 95°C for 50 minutes, while the other was not treated. When the counts were normalized with the positive controls and averaged for the samples of each mRNA type (Figure 3), the RNA degradation affected the expression measurement. It was found that degradation reduced the apparent expression level of every gene by an average of about 34% (as indicated by the reduced slope of the line), compared to the intact heart RNA.

To correct for this apparent decrease in transcript abundance, we next calculated a PGK1 normalization factor. The positive control-normalized gene counts for PGK1 obtained in heart tissue for one of the assays were used as a reference. The corrected counts for PGK1 obtained for each assay were divided by those obtained from the first assay of the heart tissue sample to obtain a normalization factor. The positive control normalized gene counts for each gene in each assay were then divided by the appropriate normalization factor and averaged for the samples of each mRNA type to generate counts normalized to the PGK1 reference gene.

Once the counts are normalized to the reference gene PGK1 (Figure 3), the observed expression levels of the genes were equivalent in both the intact and undegraded mRNA samples (slope ~1.0). The counts in the two different samples could then be compared directly.

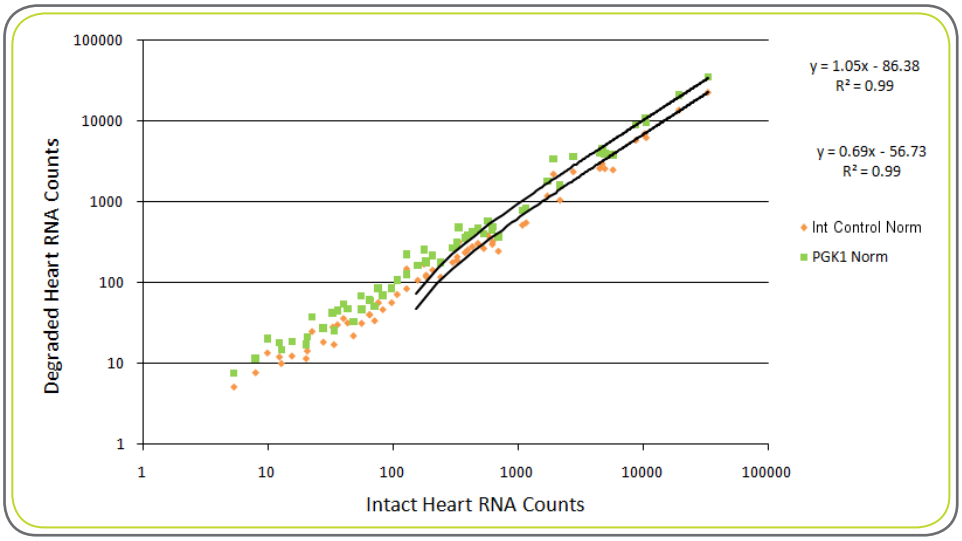

Figure 3. Concordance of expression levels using intact and degraded RNA

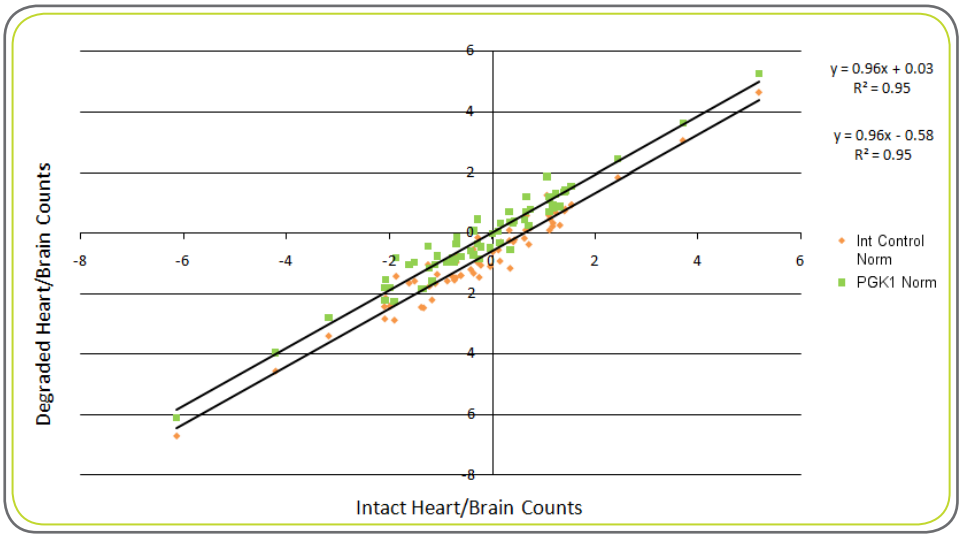

Figure 4. Concordance of calculated fold-changes in expression levels using intact and degraded RNA

Accurate Assessment of Changes in Gene Expression

The ultimate goal of gene expression experiments is to reliably determine changes in transcript abundance during various phases of development, across different tissue types, and under different environmental and experimental conditions. Normalization to reference genes can help assure accurate assessment of these changes. To illustrate how fold-changes are affected by housekeeping gene normalization, we analyzed counts in total brain RNA compared to either intact heart RNA or degraded heart RNA using the CodeSet described in the previous section. After corrected counts were averaged for the triplicates of each sample type, the ratio of heart/brain counts was determined using both intact and degraded heart mRNA samples. A log2 plot of the ratios was then used to determine the concordance of the fold changes in expression using intact and degraded heart mRNA (Figure 4).

In one case the fold-change ratios were determined using counts normalized only to the internal positive controls. The plot illustrates that the trend line does not pass through zero, indicating that the use of degraded mRNA can skew observed changes in expression. To correct for this anomaly, we compared the fold-changes calculated after normalizing to PGK1 When fold-change ratios were determined using counts that have been normalized to PGK1, the trend line passed through zero. This example demonstrates that reliable fold-change measurements can be made with degraded RNA by normalizing to a suitable reference gene.

Getting the Most From Your Reference Genes

The tissue comparisons shown in Figure 2 are meant as a guide for use with a particular tissue type in your system. Expression levels may vary from those illustrated, depending on the source and history of the tissue. You may wish to verify the relative expression levels of these genes in your samples before using them as a reference.

When you are choosing your own reference genes for a custom CodeSet, it is best to pick at least three reference genes which should be expressed at a relatively stable levels across your chosen experimental conditions. We recommend choosing one from the high expression category and at least one each from medium and low expressing categories. In order to improve the accuracy and sensitivity of your results, we do not recommend using multiple reference genes expressed at very high levels. You may want to check the literature to determine which genes may have been used for reference in experimental conditions that most closely match your system.

Extremely high-expressing genes such as ribosomal RNA transcripts or globin (in blood samples) should not be used as reference genes without modifying your CodeSet and protocol. If you need to use a very high expressing gene as a reference, please contact Nanostring to discuss possible approaches.

Points for Consideration

Note that the data shown in Figure 1 using the Nanostring reference gene set was derived from healthy, untreated tissue. These data could change for the same tissues that have been treated, or are diseased. It is always wise to check the literature or determine empirically if the reference genes being used are known to change their expression levels upon a particular treatment, or during a disease state.

If you wish to compare expression data from the nCounter Analysis System with data generated on other systems, it is imperative to use the same reference genes for normalization in both systems. This should eliminate any erroneous variability in relative expression levels and allow direct comparison of the data.

Human Reference Gene List

Accession	Gene ID	Official Symbol	Full Name
NM_001090.2	23	ABCF1	Homo sapiens ATP-binding cassette, sub-family F (GCN20), member 1 (ABCF1), transcript variant 2, mRNA
NM_001101.2	60	ACTB	Homo sapiens actin, beta (ACTB), mRNA
NM_000688.4	211	ALAS1	Homo sapiens aminolevulinate, delta-, synthase 1 (ALAS1), transcript variant 1, mRNA
NM_004048.2	567	B2M	Homo sapiens beta-2-microglobulin (B2M), mRNA
NM_004859.2	1213	CLTC	Homo sapiens clathrin, heavy polypeptide (Hc) (CLTC), mRNA
NM_000402.2	2539	G6PD	Homo sapiens glucose-6-phosphate dehydrogenase (G6PD), nuclear gene encoding mitochondrial protein, mRNA
NM_002046.3	2597	GAPDH	Homo sapiens glyceraldehyde-3-phosphate dehydrogenase (GAPDH), mRNA
NM_000181.1	2990	GUSB	Homo sapiens glucuronidase, beta (GUSB), mRNA
NM_000194.1	3251	HPRT1	Homo sapiens hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome) (HPRT1), mRNA
NM_005566.1	3939	LDHA	Homo sapiens lactate dehydrogenase A (LDHA), mRNA
NM_000291.2	5230	PGK1	Homo sapiens phosphoglycerate kinase 1 (PGK1), mRNA
NM_019014.3	84172	POLR1B	Homo sapiens polymerase (RNA) I polypeptide B, 128kDa (POLR1B), mRNA
NM_000937.2	5430	POLR2A	Homo sapiens polymerase (RNA) II (DNA directed) polypeptide A, 220kDa (POLR2A), mRNA
NM_000981.3	6143	RPL19	Homo sapiens ribosomal protein L19 (RPL19), mRNA
NM_001002.3	6175	RPLP0	Homo sapiens ribosomal protein, large, P0 (RPLP0), transcript variant 1, mRNA
NM_004168.1	6389	SDHA	Homo sapiens succinate dehydrogenase complex, subunit A, flavoprotein (Fp) (SDHA), nuclear gene encoding mitochondrial protein, mRNA
NM_003194.3	6908	TBP	Homo sapiens TATA box binding protein (TBP), mRNA
NM_178014.2	203068	TUBB	Homo sapiens tubulin, beta (TUBB), mRNA

Purchasing Information

Product Description	Catalog No.
nCounter™ Analysis System Includes the Prep Station and the Digital Analyzer	NCT-SYST
nCounter™ GX CodeSet Gene Expression Custom CodeSet	GXA-P1CS
nCounter™ Master Kit All reagents, sample cartridges, and consumables necessary for processing 48 Nucleic Acid Assays.	NAA-AKIT

NanoString Technologies, Inc.
530 Fairview Ave N, Suite 2000
Seattle, Washington 98109
(206) 378-NANO • (206) 378-6266
www.nanostring.com

v.20090608