Scientific article

Multicopy reference assay (MRef) — a superior normalizer of sample input in DNA copy number analysis

Samuel Long, Quan Peng, John DiCarlo, Yexun Wang, and Eric Lader

QIAGEN Sciences, Inc. 6951 Executive Way, Frederick, Maryland 21703, USA

Abstract: Copy number variations (CNVs) and alterations (CNAs) are a source of genetic diversity in humans and are often pathogenic. Numerous CNVs and CNAs are being identified with various genome analysis platforms, including array comparative genomic hybridization (aCGH), single nucleotide polymorphism (SNP) genotyping platforms, and next-generation sequencing. Independent verification of copy number changes is a critical step. Quantitative real-time PCR (qPCR) is a classic method to verify microarray copy number findings. Traditional copy number assays that use qPCR typically rely on a putative single-copy gene reference assay (e.g., RNase P or TERT) to normalize the DNA input for downstream $\Delta\Delta C_T$-based copy number calculation for comparison to a reference genome. When applied to cancer samples, these single-copy reference assays may no longer be a reliable indicator of DNA input due to the presence of complex chromosome composition (both in chromosome number and structure). To meet the need for an accurate DNA input normalizer, especially for heterogeneous tumor samples, QIAGEN developed a multicopy reference (MRef) assay for real-time PCR copy number analysis. This assay, in conjunction with QIAGEN's greater than 10 million genomewide copy number assays and pathway- and disease-focused copy number PCR arrays (Figure 1), provides a successful solution for copy number analysis. This article will address the assay design considerations, development, and performance of this multicopy reference (MRef) assay.

Introduction

Relative copy number determination using qPCR usually relies on the use of a reference assay to normalize the DNA input (i.e., the amplifiable DNA amount in the amplicon size range of the reference assay). The C_T for the locus of interest is normalized against the C_T of the reference assay in the same sample, and the ΔC_T thus obtained will then be used in $\Delta\Delta C_T$ analysis to enable copy number calculation. This calculation is usually performed in reference to a two-copy or known copy reference genome(s)/sample(s). Therefore, accurate representation of the DNA input by the reference assay is one of the most important requirements for any copy number analysis using qPCR.

Contents

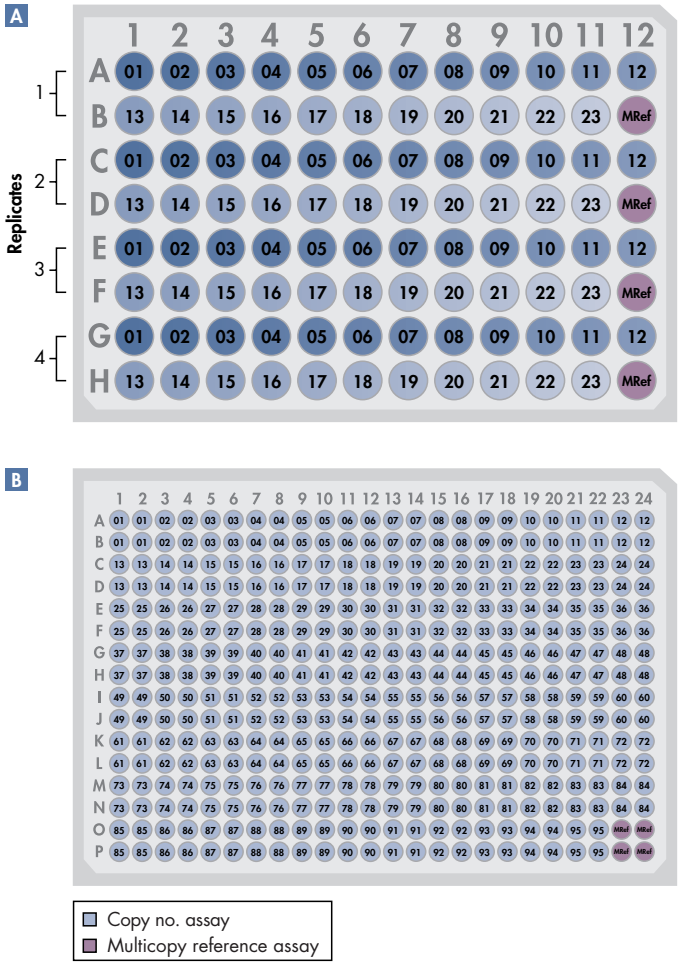

Figure 1. qBiomarker Copy Number PCR Array layout. A The 96-well plate layout for a 23-gene panel allows 4 technical replicates for one sample for 23 genes, (each gene assay is represented by they numbered blue wells) and a multicopy reference assay (MRef, in purple). B The 384-well plate layout for a 95-gene panel has 4 replicates of each gene assay and a multicopy reference assay (MRef). In addition, alternative plate layouts such as the 23-gene panel on a 384-well plate, a 23-gene panel on a 100-well Rotor-Disc, and a custom array layout with variable configurations are available. The $\Delta\Delta C_t$ method coupled with a t-test is used to calculate copy number value. The same data analysis principle applies to individual assays as well.

Current commonly used reference assays for copy number analysis are typically for single-copy genes such as RNase P (RPPH1) (1) and TERT (2). These assays have two significant drawbacks when used for DNA input normalization. Firstly, these assays rely on the single-copy gene locus being unaffected in all of the experimental samples. For example, an amplification or deletion of RNaseP or TERT would significantly affect the calculated copy number for a gene of interest (Table 1). Indeed, amplifications of both of these commonly used genes have been reported in the literature (3–4). Secondly, in some cancer cells, complex chromosomal composition can occur due to a combination of chromosome number alteration (in some cells as high as 84 chromosomes per cell) and chromosome structure change (such as marker chromosomes that contain three or more individual chromosome segments). In such cells, the single-copy assays are usually no longer accurate indicators of DNA input amount. For example, in a DNA sample where the average modal chromosome number is 80, using RNase P as a reference assay can inflate an amplified cancer gene's calculated copy number to about twice the actual copy number.

Table 1. Genomic changes affect single gene references

CNV reference	Copy number per diploid human genome	C_T change (reference assay)	GOI copy number (real)	COI copy number (calculated)	CNV calling for GOI
MRef Reference Assay	36	0	2	2.00	NC: Correct
	36 + 1	–0.04	2	1.94	NC: Correct
	36 – 1	+0.04	2	2.06	NC: Correct
Single-copy gene	2	0	2	2.00	NC: Correct
	2 + 1	–0.58	2	1.33	Loss: False
	2 – 1	+ 1	2	4.00	Gain: False

Multicopy reference assay design considerations

An ideal reference assay for DNA input normalization should fulfill three major criteria. With regard to copy number of the reference assay in the genome, ideally the reference assay should be present >20 times in a normal diploid genome. This will minimize the effect of local genomic amplification or deletion events on the copy number of the reference assay itself per genome equivalent amount of DNA. Secondly, distribution of the reference assay in the genome is crucial. Ideally, the many copies of the reference assay should be randomly distributed in the genome, such that the copies are located in different chromosomes and show no concentration on one or a few chromosomes or a chromosome arm. Finally, the reference assay must fulfill an amplicon size requirement. As the copy number assays and reference assay are also intended to be used on FFPE samples, ideally the reference assay amplicons should be 100 base pairs or below, and all amplicons should be within ~10 base pairs of the same length for all recognized sites in the genome.

Development of a multicopy reference assay (MRef)

Through bioinformatics analysis, QIAGEN identified a multicopy reference assay (MRef) that fulfills the above criteria. This primer assay recognizes 41 identical primer pair binding sites in the genome, potentially recognizes 66 near-match binding sites (a near-match binding site is defined as a pair of potential primer binding sequences that have a mismatch to one of the primers, or one mismatch to each of the primers in the primer pair), and is predicted to generate amplicons that are all ~80 base pairs in length. Melt curve analysis indicates that the primer assay produces more than one distinct amplicon in the genome, as expected (Figure 2).

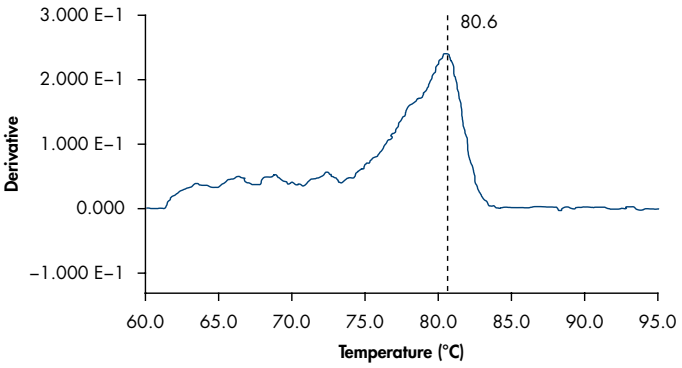

Figure 2. Dissociation curve of the multi-copy reference assay (MRef) on Promega human genomic DNA.

Stable performance of the MRef assay in 129 research samples from 9 ethnic populations

In order to have a universal reference assay for human research samples, the normalization assay must detect a similar number of primer binding sites across different ethnic populations. QIAGEN tested the copy number stability of the MRef assay in a total of 129 normal human genomic DNA samples from 9 ethnic populations (Figure 3). In these samples, the RB1 gene is assumed to be present in 2 copies. The MRef assay and a qBiomarker Copy Number PCR Assay for RB1 were tested against each DNA sample in quadruple replicate reactions, and the ΔC_T between the average C_Ts of the MRef assay and the RB1 assay was calculated for each individual DNA. The ΔC_Ts cluster around 6.54 (1 S.D. = 0.14). This result demonstrates the relatively stable copy number of the MRef assay in all the individual samples tested.

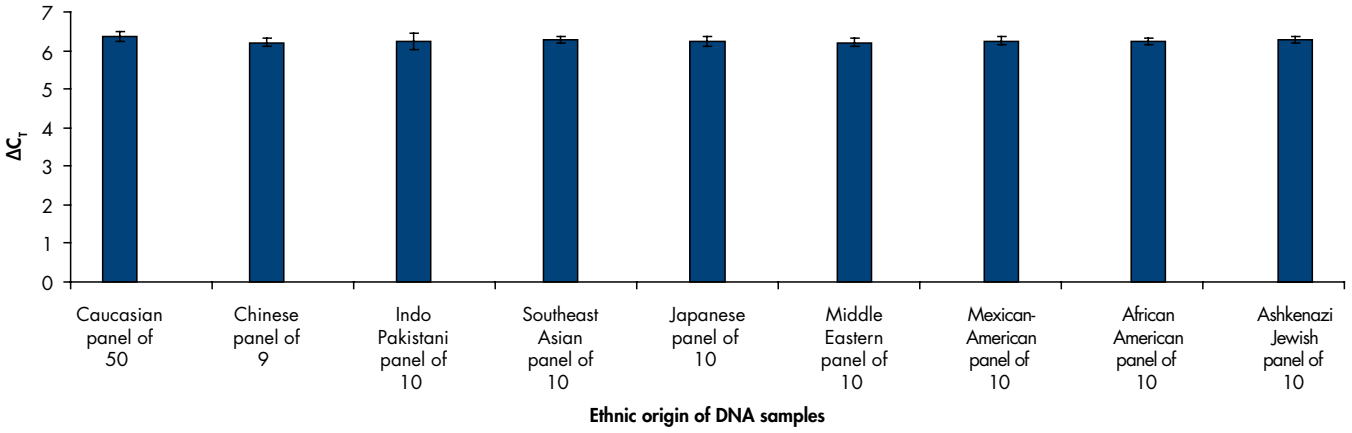

Figure 3. Stable performance of the MRef assay in 129 DNA research samples from 9 major human populations. The MRef (multicopy reference) assay and a qBiomarker Copy Number PCR Assay for RB1 were tested against normal human genomic DNA obtained from 129 individuals from 9 major ethnic populations (Coriell). Each assay and DNA sample combination was run in quadruple reactions. ΔC_T between the average C_Ts of the MRef assay and the RB1 assay was calculated for each individual DNA. Data from each DNA sample was grouped by ethnic origin and averaged, and the average ΔC_T for each population was plotted. The ΔC_Ts cluster around 6.54 (1 S.D. = 0.13). The RB1 gene is assumed to be present at 2 copies in all normal human genomic DNA samples.

Superior performance of MRef over single-copy reference assay in copy number analysis

The performance of the MRef assay was compared to an established single-copy reference assay (RNase P) for DNA input normalization and the effect of including either reference in copy number call outcome. When MRef and RNase P were compared for their ability to normalize GRB7 gene copy number in mixed wild-type and tumor cell DNA (mimicking scenarios with heterogeneous tumor cells), the observed GRB7 copy numbers in general agreed well with the expected values when using the multicopy reference assay as the reference, while significant differences were observed between the observed GRB7 gene copy numbers and the expected values when RNase P was used as the reference assay (Figure 4A). RNase P normalization led to overestimation of the true copy number of GRB7. The average copy numbers of RNase P per normal genome copy amount of DNA were determined in 2 breast cancer cell lines, using commercially available human genomic DNA as a reference genome, and RNase P was found to be an unsuitable indicator of DNA input in these cancer cell lines (Figure 4B). The SKBR3 cancer cell line is known to have a complex chromosome composition due to a combination of chromosome number gain and structural changes. Collectively, the MRef assay was found to be less influenced by local genomic alterations, and is consequently a better indicator of DNA input than a single-copy reference assay.

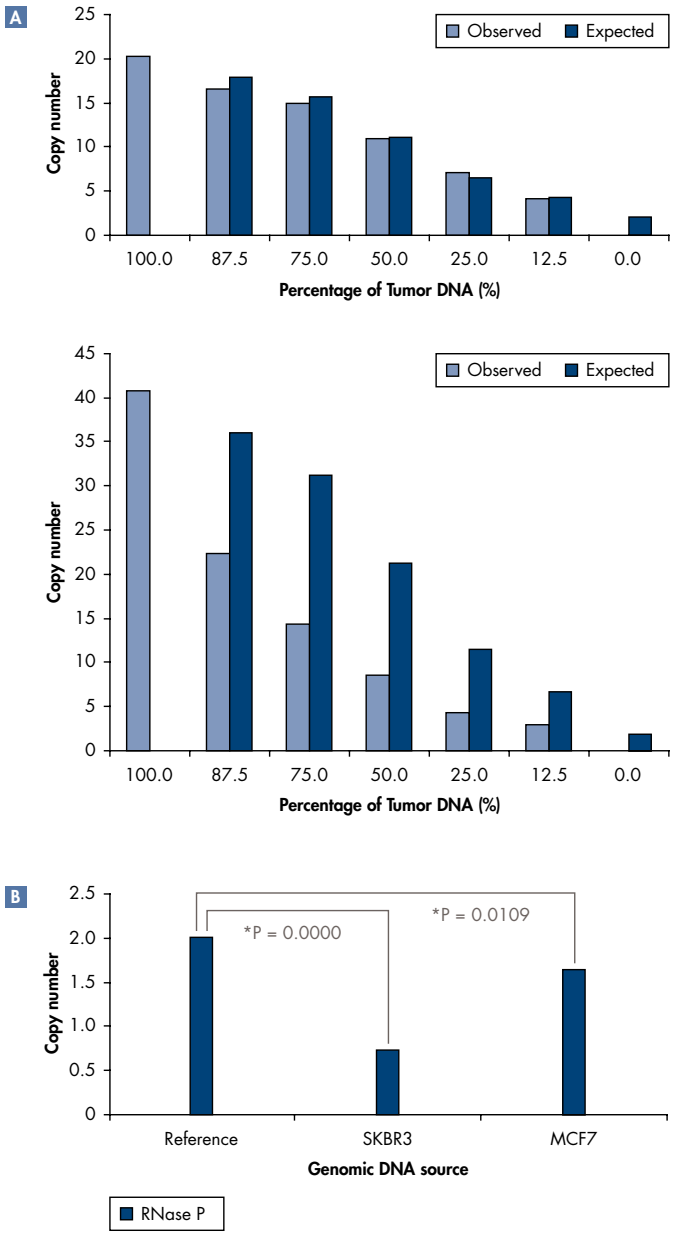

Figure 4. The multicopy reference assay (MRef) is superior to RNase P as a reference assay for copy number determination. A Tumor cell line DNA (SKBR3) and reference genomic DNA (Promega human genomic DNA) were mixed in different ratios (100%, 87.5%, 75%, 50%, 25%, 12.5%, and 0% SKBR3 cells, respectively), and the DNA mixes were tested for the gene copy number of GRB7 (a gene that is significantly amplified in SKBR3), using either the MRef assay (top) or the RNase P assay (bottom) as the reference. The GRB7 copy number in reference DNA is assumed to be 2. The "expected" GRB7 gene copy numbers in 87.5%, 75%, 50%, 25%, and 12.5% mixing ratio samples are calculated based on the GRB7 gene copy number in 100% SKBR3 gDNA sample, and the mixing ratio between the SKBR3 gDNA and reference genomic DNA. The observed GRB7 gene copy numbers in general agree well with the expected values when using the multicopy reference assay as the reference, while significant differences exist between the observed GRB7 gene copy numbers and the expected values when RNase P is used as the reference assay. B RNase P and other single-copy genes are not suitable normalizers for certain sample input. Genomic DNA was isolated from 2 breast cancer cell lines (SKBR3 and MCF7) using QIAGEN QIAamp DNA Mini Kit. The absolute average copy numbers of RNase P per normal genome copy amount of DNA were determined in these breast cancer cell line genomic DNAs with the $\Delta\Delta C_T$ method, using the multi-copy reference assay (MRef) as the normalization control of DNA input. Promega® human genomic DNA was used as the reference DNA and the absolute copy number of RNase P per normal genome in this DNA is assumed to be 2. The copy number of RNase P in both SKBR3 and MCF7 is significantly altered, demonstrating that this reference gene would be an unreliable normalizer for these samples.

MRef assay performance in cell line chromosomal aneuploidy detection

QIAGEN verified the performance of the MRef assay in combination with qBiomarker Copy Number PCR Assays in cell lines with known copy number alteration of the X chromosome, which had previously been identified using cytogenetic methods. Copy number assays for the X-chromosome genes androgen receptor (AR) and methyl CpG binding protein 2 (MECP2) revealed 1 copy of each gene in the XY sample, 3 copies in the XXX sample, and 4 copies in the XXXX sample, as expected, using normal female DNA as a 2 copy XX control (Figure 5). These results show the ability of qBiomarker Copy Number PCR Assays to accurately identify single-copy changes when using the MRef assay as the reference assay.

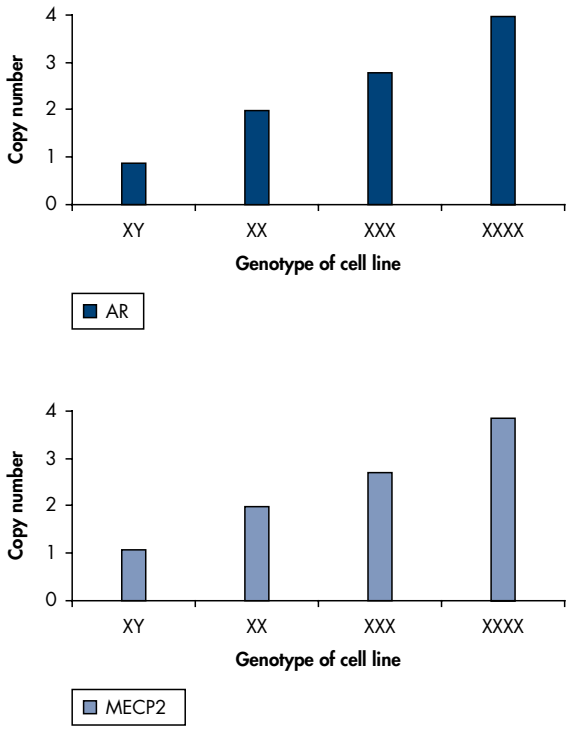

Figure 5. qBiomarker Copy Number PCR Assays accurately identify aneuploidy. Genomic DNA was isolated from 4 cell lines that contain 1 copy (XY, Coriell NA13619), 2 copies (XX, Coriell NA01921), 3 copies (XXX, Coriell NA03623) and 4 copies (XXXX, Coriell NA11226) of X-chromosome, respectively, using QIAGEN QIAamp DNA Mini Kit. Chromosomal aberrations had been previously identified by cytogenetic methods. qPCR copy number assays designed to target AR and MECP2 (both located on the X-chromosome) were tested against the 4 DNAs. The multicopy reference assay (MRef) was used to normalize the amount of DNA input. The $\Delta\Delta C_T$ method was used to calculate the gene copy number, using XX (Coriell NA01921) as a 2-copy reference. Each assay was tested against each sample in quadruple replicate reactions, and a t-test was performed.

MRef assay performance with a qBiomarker Copy Number PCR Array in FFPE sample copy number alteration screening

QIAGEN evaluated the effectiveness of qBiomarker Copy Number PCR Arrays, which contain the MRef assay as the DNA input normalizer for each sample, as a primary screening tool. A lung cancer copy number PCR array was used to detect copy number alteration in non-small cell lung cancer FFPE samples (Figure 6A). A panel of qBiomarker Copy Number PCR Assays was chosen based on previous reports of amplifications or deletions in lung cancer. In Figure 6A, the resulting array was used to compare a lung cancer FFPE sample (#256, purchased from Cybrdi, Inc.) and a normal lung tissue FFPE sample (N23, purchased from Cybrdi, Inc.). The fibroblast growth factor 3 gene, FGF3, showed statistically significant amplification in the lung cancer FFPE sample via the copy number PCR array, and an adjacent genomic qPCR assay confirmed amplification of this gene in the cancer sample compared to the normal sample (Figure 6B).

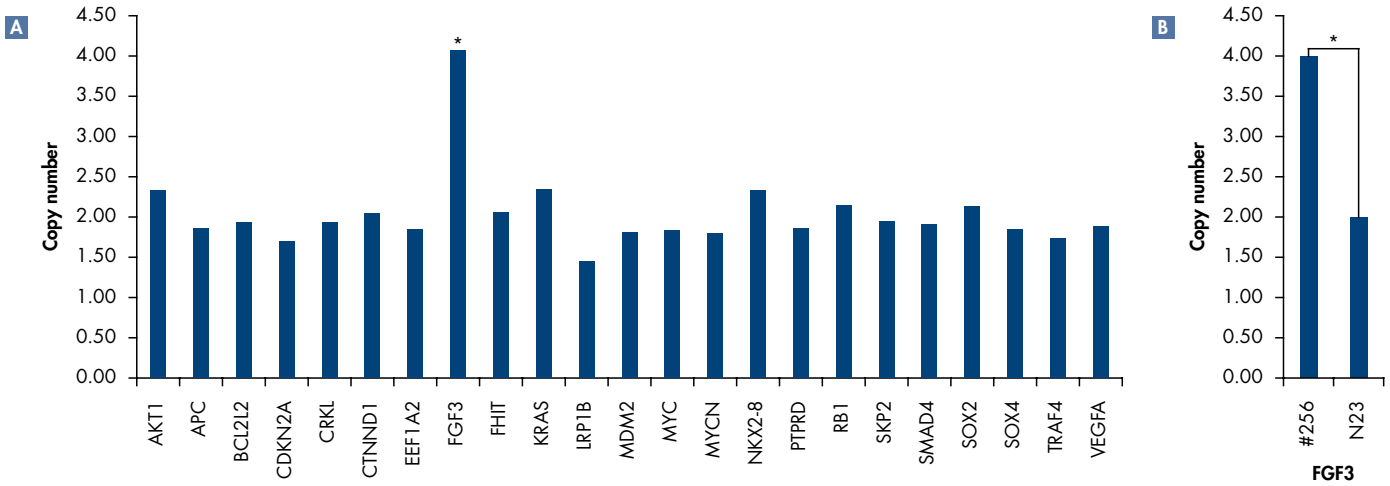

Figure 6. qBiomarker Copy Number PCR Array identifies a copy number alteration in a lung cancer FFPE sample that was subsequently verified using an adjacent genomic assay. A Genomic DNA was isolated from a non-small cell lung cancer FFPE sample (#256) and a healthy lung FFPE DNA (N23) (reference sample) using QIAGEN DNeasy Blood & Tissue Kit. DNA was profiled on a qBiomarker Copy Number PCR Array (cat. no. VAHS-00567) which contains a panel of lung cancer-specific genes previously reported to undergo amplifications or deletions in lung cancer. The samples were tested against each assay in quadruple replicate reactions. The MRef assay was used to normalize the amount of DNA input in each sample. The $\Delta\Delta C_T$ method was used to calculate the gene copy number for each locus in sample #256. A t-test was performed to determine if the observed deviation from 2 copies was statistically significant (* indicates $p<0.01$). The copy number PCR array identifies FGF3 as statistically significantly amplified. B Individual qPCR verification of FGF3 locus amplification in sample #256 compared to sample N23. An adjacent genomic assay (~200 bp away from the FGF3 assay (cat. no. VPH 111-0348167A) on the PCR array in A) was used for independent verification of the copy number PCR array result. The MRef assay was used to normalize the amount of DNA input in each sample, and the $\Delta\Delta C_T$ method was used to calculate the gene copy number for the FGF3 locus in sample #256.

MRef assay and qBiomarker Copy Number PCR Assays enable detection of 16% change in average gene copy number

The average copy number status of a set of genes that undergo frequent amplification or deletion in breast cancer was determined in the SKBR3 cell line with corresponding gene-specific qBiomarker Copy Number Assays and the MRef (Figure 7). Significant amplifications were observed in this cell line for ERBB2, MYC, ZNF217, and AURKA. In addition, a 0.51 copy change (25% gain) in average copy number and a 0.32 copy change (16% loss) in average copy number in CDK4 and FHIT were observed, respectively.

Gene	Gene ID	Copy number in SKBR3	p-value
AURKA	6790	7.66	0.0000
CDK4	1019	2.51	0.0062
ERBB2	2064	23.98	0.0000
FHIT	2272	1.68	0.0063
MYC	4609	21.79	0.0000
ZNF217	7764	11.67	0.0000

Figure 7. MRef assay combined with qBiomarker Copy Number PCR Assays enables detection of a 16% change in average copy number from a group of cells. SKBR3 cell line DNA was tested against qBiomarker Copy Number PCR Assays for 6 genes that have been frequently reported to undergo amplification or deletion in breast cancer samples. The multicopy reference assay (MRef) was used to normalize the amount of DNA input. The $\Delta\Delta C_T$ method was used to calculate the gene copy number, using Promega human genomic DNA as a 2-copy reference. Each assay was tested against each sample in quadruple replicate reactions, and a t-test was performed.

Conclusion

A suitable reference assay that accurately reflects the amount of DNA input is a critical factor in real-time PCR copy number profiling or measurement methods. Single-copy genes, on which conventional reference assays are based, are not an accurate indicator of DNA input in some sample types. The novel multicopy reference assay (MRef) described here recognizes a total of more than 100 loci that are randomly distributed in the genome with ~80 base pair amplicons. The copy number of the loci recognized by this assay remains constant in 129 healthy individuals from 9 ethnic populations. In titration experiments, the novel MRef demonstrates more accurate copy number measurement (i.e., agrees well with expected results) than conventional reference assays based on single-copy genes. In addition, qBiomarker Copy Number PCR Assays and Arrays using the novel MRef as the reference assay can detect chromosomal aneuploidy, be used as primary screening tools in fresh frozen and FFPE samples, and resolve as low as 16% alterations in average copy number.

In summary, this study has provided evidence of the limitations of single-copy reference genes, the superiority of MRef as a normalizer for the amount of DNA input, MRef's stable performance across human populations, and the usefulness of this assay in producing accurate copy number measurements.

References

1. Pau, C.P., Wells, S.K., Rudolph, D.L., Owen, S.M., and Granade, T.C. (2010) A rapid real-time PCR assay for the detection of HIV-1 proviral DNA using double-stranded primer. J. Virol. Methods **164**, 55.
2. Anedchenko, N. et al. (2008) Activation of the hTERT expression in squamous cell cervical carcinoma is not associated with gene amplification. Oncol. Rep. **20**, 469.
3. Conrad, D.F. et al. (2010) Origins and functional impact of copy number variation in the human genome. Nature **464**, 704.
4. Pinto, D. et al. (2011) Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nat. Biotechnol. **29**, 512.

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at www.qiagen.com or can be requested from QIAGEN Technical Services or your local distributor.

Australia ■ 1-800-243-800
Austria ■ 0800-281011
Belgium ■ 0800-79612
Brazil ■ 0800-557779
Canada ■ 800-572-9613
China ■ 800-988-0325
Denmark ■ 80-885945
Finland ■ 0800-914416
France ■ 01-60-920-930
Germany ■ 02103-29-12000
Hong Kong ■ 800 933 965
India ■ 1-800-102-4114
Ireland ■ 1800 555 049
Italy ■ 800-787980
Japan ■ 03-6890-7300
Korea (South) ■ 080-000-7145
Luxembourg ■ 8002 2076
Mexico ■ 01-800-7742-436
The Netherlands ■ 0800-0229592
Norway ■ 800-18859
Singapore ■ 1800-742-4368
Spain ■ 91-630-7050
Sweden ■ 020-790282
Switzerland ■ 055-254-22-11
Taiwan ■ 0080-665-1947
UK ■ 01293-422-911
USA ■ 800-426-8157

www.SABiosciences.com
www.qiagen.com