Physiol Genomics 30: 363–370, 2007.
First published April 24, 2007; doi:10.1152/physiolgenomics.00251.2006.

GeneChip, geNorm, and gastrointestinal tumors: novel reference genes for real-time PCR

Mark Kidd,[1] Boaz Nadler,[2] Shrikant Mane,[3] Geeta Eick,[1] Maximillian Malfertheiner,[1] Manish Champaneria,[1] Roswitha Pfragner,[4] and Irvin Modlin[1]

[1]*Gastrointestinal Research Group and* [3]*Keck Laboratory, Yale University School of Medicine, New Haven, Connecticut;* [2]*Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel and* [4]*Institute of Pathophysiology, Centre for Molecular Medicine, Medical University of Graz, Austria*

Submitted 15 November 2006; accepted in final form 23 April 2007

Kidd M, Nadler B, Mane S, Eick G, Malfertheiner M, Champaneria M, Pfragner R, Modlin I. GeneChip, geNorm, and gastrointestinal tumors: novel reference genes for real-time PCR. *Physiol Genomics* 30: 363–370, 2007. First published April 24, 2007; doi:10.1152/physiolgenomics.00251.2006.—Accurate quantitation of target genes depends on correct normalization. Use of genes with variable tissue transcription (*GAPDH*) is problematic, particularly in clinical samples, which are derived from different tissue sources. Using a large-scale gene database (Affymetrix U133A) data set of 36 gastrointestinal (GI) tumors and normal tissues, we identified 8 candidate reference genes and established expression levels by real-time RT-PCR in an independent data set ($n = 42$). A geometric averaging method (geNorm) identified *ALG9*, *TFCP2*, and *ZNF410* as the most robustly expressed control genes. Examination of raw C_T values demonstrated that these genes were tightly correlated between themselves ($R^2 > 0.86$, $P < 0.0001$), with low variability [coefficient of variation (CV) <12.7%] and high interassay reproducibility ($r = 0.93$, $P = 0.001$). In comparison, the alternative control gene, *GAPDH*, exhibited the highest variability (CV = 18.1%), was significantly differently expressed between tissue types ($P = 0.05$), was poorly correlated with the three reference genes ($R^2 < 0.4$), and was considered the least stable gene. To illustrate the importance of correct normalization, the target gene, *MTA1*, was significantly overexpressed ($P = 0.0006$) in primary GI neuroendocrine tumor (NET) samples (vs. normal GI samples) when normalized by $geNorm_{ATZ}$ but not when normalized using *GAPDH*. The $geNorm_{ATZ}$ approach was, in addition, applicable to adenocarcinomas; *MTA1* was overexpressed ($P < 0.04$) in malignant colon, pancreas, and breast tumors compared with normal tissues. We provide a robust basis for the establishment of a reference gene set using GeneChip data and provide evidence for the utility of normalizing a malignancy-associated gene (*MTA1*) using novel reference genes and the geNorm approach in GI NETs as well as in adenocarcinomas and breast tumors.

GAPDH; housekeeping; microarray

REAL-TIME RT-PCR IS A highly sensitive and reproducible technique with a wide dynamic range of amplification that allows quantitation of target genes. Typically, a reference gene (variously known as a "housekeeping," "maintenance," "endogenous control," or "reference" gene) is amplified together with the target gene of interest to calibrate for several sources of experimental variability including the amount of starting material, RT enzymatic efficiencies, and differences between tissues or cells in overall transcriptional activity (15). The amount of target gene in each specimen assayed is reported as the ratio of the target gene relative to the reference gene, with changes in sample gene expression compared with either an external standard or a reference sample (48).

Despite publications identifying numerous novel housekeeping genes (47), the majority of investigators use only a single gene from a small panel of housekeeping genes, comprising *β-actin* (*ACTB*), *β2-microglobulin*, *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*), *hydroxymethylbilane synthase*, *hypoxanthine phosphoribosyl-transferase 1* (*HPRT1*), *ribosomal protein L13a*, ribosomal 18S RNA, *succinate dehydrogenase complex* subunit A, *TATA box binding protein*, *ubiquitin C*, and *tyrosine 3-monooxygenase* to normalize real-time PCR results (1). Genes commonly used for normalization in neuroendocrine tumor (NET) tissue studies include *GAPDH* (23, 25, 37, 46), *ACTB* (37), *HPRT1* (34), and *glucose phosphate isomerase* (4). A growing body of evidence, however, indicates that transcript levels of these commonly used housekeeping genes may vary considerably in different tissue types or under different experimental conditions, and that a "universal control gene" does not exist (15, 44, 47, 48). Furthermore, a recent study (45) revealed that conventional normalization strategies based on a single housekeeping gene can lead to normalization errors of up to 3- and 6-fold in 25 and 10% of cases, respectively, with some cases showing error values >20-fold (45).

The determination of a panel of genes that have robust expression in the specific experimental system being studied is essential to ensure accurate normalization and interpretation of results. A software program written by Vandesompele et al. (45), geNorm, identifies the most stably expressed gene or set of genes from among a pool of genes and estimates the number of genes required to calculate a robust normalization factor based on the geometric mean of these genes. The gene stability measure "*M*" that geNorm determines is defined as the average pairwise variation of a particular gene with all other potential reference genes (45). This measure is based on the principle that the expression ratio of two ideal control genes should be identical in all samples; thus genes with the lowest *M* value are the most stably expressed. This normalization algorithm has been successfully applied and validated in several recent studies (11, 13, 30–33, 36), and the clinical applicability of this approach has been recently highlighted in prostate cancer tumors (31). Normalization of expression levels of the matrix metalloproteinase (MMP) *RECK* by geNorm resulted in significantly different results (25% decrease vs. matched normal tissue) compared with the standard normalization approach using *ACTB* (10% increase in *RECK* in tumors). Expression of

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).

Address for reprint requests and other correspondence: I. Modlin, Yale Univ. School of Medicine, TMP202, 333 Cedar St., New Haven, CT (e-mail: imodlin@optonline.net).

this gene is normally decreased in tumor samples. Choosing the "correct" normalization approach therefore is both essential and critical for obtaining reliable expression data, particularly when clinical markers are evaluated.

A question that remains unresolved is how to identify the appropriate housekeeping gene(s) for a particular experimental system. The choice of reference genes is often based on historical precedent, e.g., *GAPDH* or *β-actin*, both originally used in Northern blots and RNase protection assays (15). The measurement of transcript levels of thousands of genes simultaneously using gene chip technology has, however, provided large databases of transcript information from various experimental systems that can then be used as alternative resources to identify reference genes (47). In this manuscript, we propose the approach of using these data resources for the identification of novel endogenous control genes that can be utilized to provide a more robust basis for the establishment of a reference gene set in a particular tissue or experimental set-up. Our report documents the identification and evaluation of a panel of eight reference genes as well as the commonly used *GAPDH* for real-time PCR normalization in a series of gastroenteropancreatic NETs (GEP NETs; previously referred to as carcinoid tumors), normal gastrointestinal (GI) tissues and a novel small intestinal carcinoid cell line (KRJ-I) (35). These 8 novel genes were selected based on statistical algorithms (outlier detection, robust feature selection methodology) applied to GeneChip data from 36 Affymetrix U133A GeneChip samples (GEP NETs and normal tissue) to detect genes with low variability. Expression levels of these reference genes were then measured by real-time RT-PCR in an independent set of GI tissue samples ($n = 42$) and the small intestinal carcinoid cell line (KRJ-I). The geometric averaging method (45) was used to identify the most robustly expressed control genes in the tissue set we interrogated. To further assess the efficacy of this strategy, the expression level of *MTA1*, a gene previously identified by our group as a potential marker of small intestinal NET malignancy (22), was then examined in the same set of GI tissue samples to establish which approach (geNorm or *GAPDH*) was superior for measuring target gene expression in normal GI tissue and GI NETs. Thereafter, to evaluate the utility of our normalization approach in non-NETs, we prospectively examined the expression levels of *MTA1* in adenocarcinomas ($n = 44$) from the colon, pancreas, and breast and compared this with normal tissue ($n = 26$) from these GI and non-GI sites.

These approaches resulted in the identification of three reference genes, *ALG9*, *TFCP2* and *ZNF410*, that may be used for a robust normalization of target gene expression measured by real-time PCR in both GI and non-GI adenocarcinomas and NET samples.

MATERIALS AND METHODS

Tissue Samples

GI NETs and normal tissues. Seventy-eight tissue samples from normal GI tract and NETs of the GI tract were selected for this study, comprising 32 NETs (24 small bowel and 8 stomach), 5 liver and 3 lymph node metastases from NETs, 6 GI stromal tumors (GISTs) and 3 adenocarcinomas (2 gastric, 1 colon), and 29 normals including samples from the small bowel ($n = 11$), stomach ($n = 15$), liver ($n = 2$), and lymph nodes ($n = 1$) (Table 1) (2).

Table 1. *Samples used for GeneChip, reference gene identification, and real-time PCR analysis*

Type	GeneChip	RT-PCR
Stomach		
Normal ($n = 15$)	4	11
Carcinoid ($n = 8$)	5	3
GIST ($n = 5$)	5	0
Adenocarcinoma ($n = 2$)	1	1
Small bowel		
Normal ($n = 11$)	4	7
Carcinoid ($n = 24$)	11	13
GIST ($n = 1$)	1	0
Colon		
Adenocarcinoma ($n = 1$)	1	0
Lymph node		
Normal ($n = 1$)	0	1
Metastasis ($n = 3$)	0	3
Liver		
Normal ($n = 2$)	1	1
Metastasis ($n = 5$)	3	2

GIST, gastrointestinal stromal tumor.

Thirty-six samples were included in the GeneChip analyses while forty-four were used as an independent set to verify gene expression by real-time PCR (Table 1). GISTs were included in the GeneChip analysis, as these stromal tumors share some neuroendocrine features with carcinoids (6), whereas gastric adenocarcinomas and gastric neuroendocrine carcinomas (malignant NETs) are both reported to share an undifferentiated phenotype. Similarly, neither is responsive to the growth regulatory effects of hypergastrinemia (27). RNA was also extracted from the KRJ-I cell line, a small intestinal carcinoid cell line recently characterized by our group (35), for real-time PCR.

GI and non-GI adenocarcinomas and normal tissues. Seventy tissue samples comprising forty-four adenocarcinomas (21 colon, 10 pancreas, and 13 breast) and twenty-six normal tissues including samples from the colon ($n = 10$), pancreas ($n = 8$), and breast ($n = 8$) were studied. Because *MTA1* overexpression is associated with malignancy (evidence of lymph node metastases) in these tumors (10, 17, 18, 41), we categorized each tumor using previously reported criteria (10, 17, 18, 41) into either tumors of low malignant potential (no pathological evidence of lymph node metastasis) or tumors with a high malignant potential (pathological evidence of metastasis).

All samples were obtained from the Cooperative Human Tissue Network and Yale New Haven Hospital. Tissues were either frozen at −70°C or placed in liquid nitrogen for storage before RNA extraction.

GeneChip

RNA extraction. Total RNA was extracted from the 36 samples indicated in Table 1 using TRIzol (Invitrogen) followed by Qiagen RNeasy kit, and the RNA quality was assessed using Agilent Bioanalyzer (Agilent Technologies, Palo Alto, CA) to visually verify the absence of genomic DNA contamination, integrity, and ratio of 28S and 18S bands. Only samples with an absorbance ratio at 260 and 280 nm (A_{260}/A_{280}) ≥1.9 were used. Ten micrograms of very-high-quality total RNA were provided to the Keck Affymetrix facility where cRNA labeling, hybridization (U133A GeneChip), and data analysis were performed as described previously (22).

Hybridization. The Affymetrix U133A array consists of ~22,000 probe sets targeting 18,400 transcripts and variants, including 14,500 well-characterized human genes (http://www.affymetrix.com/products/arrays/specific/hgu133.affx). The hybridized arrays were scanned using a confocal laser fluorescence scanner (Agilent Microarray Scanner, Agilent

Technologies). Arrays were scaled to an average intensity of 500 and analyzed independently using Microarray Suite (MAS) 5.0 software (Affymetrix, Santa Clara, CA).

Statistical analyses of Affymetrix GeneChip data. The aim of the statistical analysis was to robustly identify candidate genes to be reference genes. To this end, raw expression data for each of the 36 samples and 22,838 probes/sample on the Affymetrix U133A chips were log transformed using Matlab (v.7; The Mathworks, Natick, MA). This is a standard approach that compresses the dynamic range of expression values, thus facilitating data interpretation (12). Log transformation of candidate reference genes previously identified in other tissues (45) in our sample data set identified a mean expression range of 6.2–10.1 with a standard deviation range of 0.23–0.49. Real-time PCR studies have previously determined that genes with large standard deviations could not be considered "stable" genes using the geNorm criteria (45). Since our aim was the identification of stable reference genes, we focused on target genes that showed a low standard deviation (<0.22) based on accepted array analysis criteria (12). To exclude genes that were called "absent" on arrays, we used a lower limit of mean log-transformed expression of 4.

To identify biologically relevant genes whose expression was neither dependent on the cell cycle nor transcriptionally linked, we *1*) determined whether expression correlated with cell proliferation, *2*) determined whether expression was highly correlated, and *3*) focused on genes with known biological function. Probes that showed a correlation of $r^2 > 0.6$ (regression analysis) with probes for the proliferation-associated genes, *Ki-67* and *PCNA,* or probes that were highly correlated with one another were excluded from further analyses. Finally, only those genes with average raw expression values of >300 in all 36 samples were retained, resulting in a panel of 8 potential housekeeping genes that could be evaluated because of the availability of Assays-on-Demand products for real-time PCR analysis.

Real-Time PCR

RNA extraction. Total RNA (2 μg) was extracted from 114 samples using TRIzol (Invitrogen) and then cleaned using the Qiagen RNeasy kit in conjunction with the DNeasy Tissue kit (Qiagen) to ensure that no contaminating genomic DNA was present (21). The clean RNA was then converted to cDNA using the High Capacity cDNA Archive kit (Applied Biosystems).

Real-time amplification. Real-time RT-PCR analysis was performed using the Assays-on-Demand products listed in Table 2 (21). All samples were adjusted to 20 ng/μl cDNA before experiments; 1 μl of template cDNA was used per reaction. In addition to the eight candidate reference genes, the expression of *MTA1*, previously shown by our group to be a potential marker of neoplasia in GI NETs (22, 24), was evaluated in all GI samples ($n = 42$) using the following Assays-on-Demand product: Hs00183042_m1 (22). The expression of *GAPDH* (Hs99999905_m1) was also assessed, since this is the most commonly used reference gene for PCR normalization (15, 44). In a second set of studies in 70 tissue samples, *ALG9*, *TFCP2*, *ZNF410*, *GAPDH*, and *MTA1* levels were measured by real-time PCR in colon, pancreas, and breast samples.

For analysis of KRJ-I RNA, in addition to analyzing *MTA1* expression, *Ki-67*, which has been proposed as a prognostic marker of GI NET neoplasia (29, 38), and two other genes, namely *MAGE-D2* and *NAP1L1*, shown by our group to be increased in expression in GI NETs (22, 24) were also evaluated (Assays-on-Demand: Hs00267195_m1, Hs00374760_m1, Hs00748775_m1).

Cycling and fluorescence detection were undertaken using the ABI7900 Sequence Detection System. Non-RT controls were included in triplicate in each real-time RT-PCR experiment to ensure the absence of genomic DNA contamination. Cycling was performed under standard conditions (TaqMan Universal PCR Master

Table 2. *Gene symbol and function of 8 candidate neuroendocrine reference genes*

Gene Symbol	Name	Chr No.	Function	Assays-on-Demand
ALG9	Asparagine-linked glycosylation 9 homolog	11	Catalyzes the transfer of mannose from Dol-P-Man to lipid-linked ligosaccharides	Hs00226894_m1
MCRS1	Microspherule protein (cell cycle-regulated factor p78)	12	Modulates the transcription repressor activity of DAXX by recruiting it to the nucleolus	Hs00197096_m1
NCOR1	Nuclear receptor co-repressor 1	17	Mediates the transcriptional repression activity of some nuclear receptors by promoting chromatin condensation, thus preventing access of the basal transcription	Hs00196920_m1
POLR2B	Polymerase (RNA) II (DNA directed) polypeptide B, 140 kDa	4	DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the 4 ribonucleoside triphosphates as substrates	Hs00265358_m1
TFCP2	Transcription factor CP2	12	α-Globin transcription factor, CP2, with homology to *Drosophila* transcription factor Elf-1/NTF-1, dimerizing with NFE2 to form the stage selector protein complex.	Hs00232185_m1
UBR2	Ubiquitin protein ligase E3 component n-recognin 2	6	Ubiquitin ligase protein, which is a component of the N-end rule pathway. Recognizes and binds to proteins bearing specific amino-terminal residues that are destabilizing according to the N-end rule, leading to their ubiquitination and subsequent degradation	Hs00389767_m1
ZNF410	Zinc finger protein 410	14	Transcription factor that activates transcription of matrix-remodeling genes such as MMP1 during fibroblast senescence.	Hs00221828_m1
ZW10	Centromere/kinetochore protein zw10 homolog.	11	Essential component of the mitotic checkpoint, which prevents cells from prematurely exiting mitosis. Required for the assembly of the dynein-dynactin complex onto kinetochores. Its association with a SNARE-like complex suggests a role in membrane traffic between the endoplasmic reticulum and the Golgi	Hs00190911_m1

Chr, chromosome.

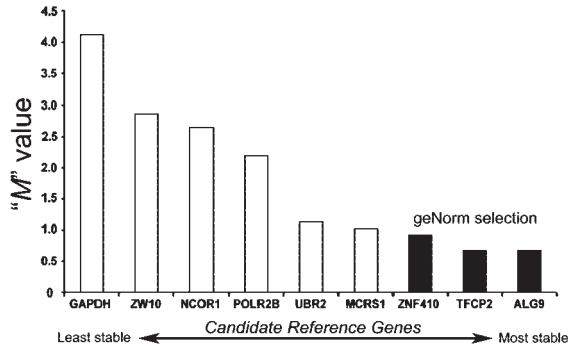

Fig. 1. Stability of *GAPDH* and the 8 candidate reference genes in gastrointestinal (GI) neuroendocrine tumors (NETs) and normal tissue calculated using geNorm. The 3 genes selected by geNorm as showing the least variation between samples (and hence the greatest stability or robust expression) are indicated on the graph. The "*M*" value is a measure of gene stability and defined as the average pairwise variation of a particular gene with all other potential reference genes (45).

Mix protocol), and the raw cycle threshold (C_T) values were exported.

Determination of stable control genes. The geNorm VBA applet for Microsoft Excel was used to determine the most stable genes from among the eight candidate reference genes (45). Raw C_T values were transformed to quantities using the comparative delta C_T method (26) where the highest relative quantity for each gene is set to 1 for input into geNorm. The gene expression stability (*M*) value for each gene was calculated by geNorm (Fig. 1). To estimate how many reference genes should be used, normalization factors based on the geometric mean of the expression levels of the "*n*" no. of best reference genes were calculated by stepwise inclusion of an extra less-stable reference gene (45).

Statistics and Data Analysis

Statistical analyses were performed using Matlab and Prism 4 (GraphPad Software, San Diego, CA). These included linear regression analysis and two-tailed Mann-Whitney and Wilcoxon rank sum tests for nonparametric data. A P value <0.05 was considered significant.

RESULTS

GeneChip Analyses

Analysis of the 36-GeneChip data set resulted in a list of 56 probes representing 42 well-characterized genes. Three probes targeting the KIAA0676 protein demonstrated a correlation >0.6 and were excluded. After a filtering, using the remaining criteria, eight potential housekeeping genes remained. These 8 genes that showed high expression levels but low variability among the 36 samples evaluated are indicated in Table 2. None of the conventional NET housekeeping genes [*GAPDH* (25, 37, 46), *ACTB* (37), *HPRT1* (34), and *glucose phosphate isomerase* (4)] was present on this list. Proteins encoded by the eight novel candidate housekeeping genes have a diversity of functions, ranging from conversion of DNA to RNA (*POLR2B*) to transcriptional regulation (*TFCP2*, *ZNF410*, *NCOR1*) to glycosylation activity (*ALG9*), and are distributed over six chromosomes (chromosomes 4, 6, 11, 12, 14, and 17) (Table 2).

Normalization

The ranking of the eight potential reference genes and *GAPDH* examined according to stability (least variability) was as follows: *ALG9* > *TFCP2* > *ZNF410* > *MCRS1* > *UBR2* > *POLR2B* > *NCOR1* > *ZW10* > *GAPDH*, with *GAPDH* the least stably (most variably) expressed gene and *ALG9* the most stably (least variably) expressed gene (19). The three reference genes estimated by geNorm to provide the most reliable normalization factor were *ALG9*, *TFCP2*, and *ZNF410* ($geNorm_{ATZ}$).

An examination of the raw C_T values for each of these three genes and for *GAPDH* demonstrated that these four genes were differently distributed ($P < 0.001$; Kruskal-Wallis test). The three reference genes were both less variably distributed than *GAPDH* (coefficients of variation: 9.7–10.58 vs. 18.11%) (Fig. 2*A*), tended to be highly correlated (R^2 0.86–0.96, $P <$ 0.0001) (Fig. 2*B*), and exhibited a lower correlation with *GAPDH* (R^2 0.253–0.28, $P < 0.001$) (Fig. 2*C*).

Within the different tissues (stomach, small intestine, liver, lymph nodes), *GAPDH* demonstrated the greatest variability and was significantly differently distributed among these tissues ($P = 0.052$ vs. $P = 0.313$–0.467 for *ALG9*, *TFCP2*, and *ZNF410*; Kruskal-Wallis test).

Importance of Robust Gene Normalization in GI NET Malignancy

To establish the utility of the geNorm approach in GI NETs, we first evaluated the reproducibility of the normalization factor

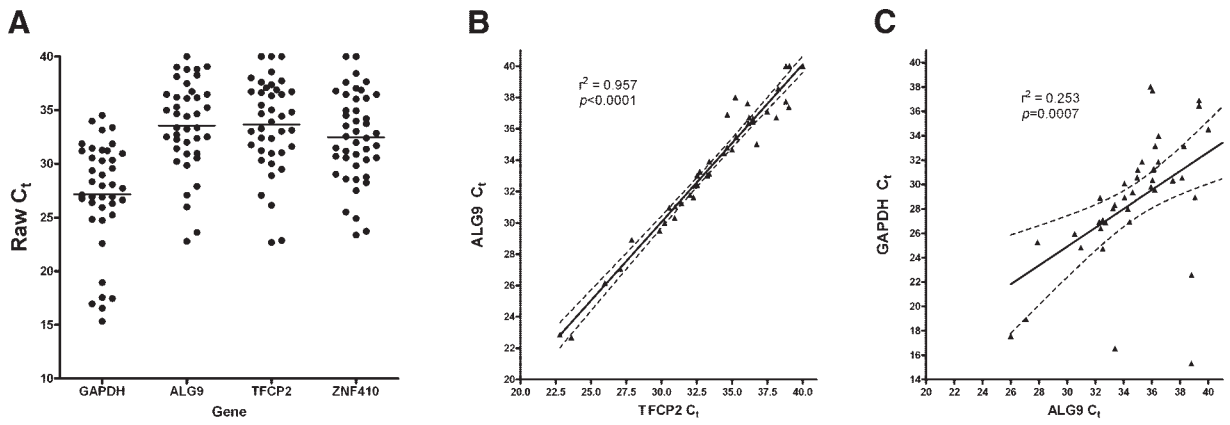

Fig. 2. Raw cycle threshold (C_T) values for each of the reference genes and GAPDH (*A*) and linear regressions between *ALG9* and *TFCP2* (*B*) *or GAPDH* and *ALG9* (*C*). Raw C_T values for *GAPDH* were lower than for the 3 reference genes and more variably distributed (*A*). Each reference gene was tightly correlated (*B*), but the relationship with GAPDH was less evident, although this remained significant (*C*).

calculated using geNorm from the three reference genes and compared this to *GAPDH*. An evaluation of the normalization factor demonstrated that this had a higher interassay reproducibility (2 separate real-time PCR studies) (Spearman = 0.93, P = 0.0011) than *GAPDH* (Spearman = 0.667, P = 0.0415).

We next compared the relative expression levels of *MTA1* normalized by $geNorm_{ATZ}$ with *MTA1* normalized by *GAPDH* in normal tissues to identify whether normalizing gene expression may have an organ-specific bias. An examination of the distribution of either *MTA1*/geNorm or *MTA1*/*GAPDH* within normal tissues demonstrated that there were no organ-specific differences in expression.

Having demonstrated no organ-specific differences, we examined the effect of each of the two normalization approaches on expression levels of *MTA1*. Normalization by $geNorm_{ATZ}$ resulted in significantly lower expression values for *MTA1* than when this gene was normalized with *GAPDH* ($P < 0.028$, 2-tailed Wilcoxon signed rank test) (19). Using regression analysis, we could identify no relationship between genes normalized by $geNorm_{ATZ}$ and *GAPDH* ($R^2 < 0.3$).

To evaluate the utility of each of these normalization approaches ($geNorm_{ATZ}$ and *GAPDH*), we next assessed the application of this stratification analysis strategy in normal tissue and compared this with primary NETs and metastases (Fig. 3). *MTA1* was significantly elevated (P = 0.0006, 2-tailed Mann-Whitney test) in primary NETs compared with normal mucosa (Fig. 3*A*). Levels of *MTA1* were increased, but not significantly, in metastases compared with normal tissue, but primary NET levels of *MTA1* were significantly elevated compared with metastases (P = 0.0314). When *GAPDH* was used to normalize the data, these trends were evident, but differences in expression were not significant between any of the samples (Fig. 3*B*).

Utility of Normalization Approach in Colon, Pancreas, and Breast Adenocarcinomas

To expand the scope of the utility of our normalization approach, we examined *MTA1* expression in colon, breast, and pancreas tumors, which are all reported to overexpress *MTA1*, particularly when categorized as malignant (10, 17, 18, 41).

In colon samples, tumors that were classified as malignant (i.e., associated with lymph node metastases) had significantly elevated levels of *MTA1* compared with either normal mucosa (P = 0.015, 2-tailed Mann-Whitney test) or nonmalignant tumors (P = 0.006, 2-tailed Mann-Whitney test) (Fig. 4*A*). When *GAPDH* was used to normalize the data, differences in expression levels were not significant between any of the samples (Fig. 4*B*).

In pancreas samples, malignant tumors had significantly elevated levels of *MTA1* compared with normal mucosa (P = 0.04) (Fig. 4*C*). Levels vs. nonmalignant tumors were different but did not achieve significance (P = 0.067). Similarly, *MTA1* normalized by *GAPDH* was also increased in malignant tumors compared with either the normal pancreas (P = 0.03) or nonmalignant tumors (P = 0.02) (Fig. 4*D*).

In breast tissue samples, malignant tumors had significantly elevated levels of *MTA1* compared with either normal breast tissue (P = 0.0007) or nonmalignant tumors (P = 0.0012) (Fig. 4*E*). When *GAPDH* was used to normalize the data, differences in expression levels were not significant between any of the samples (Fig. 4*F*).

KRJ-I Cell Line

Finally, we evaluated reference and target gene expression in the small intestinal carcinoid cell line KRJ-I. An analysis of *ALG9*, *TFCP2*, and *ZNF410* transcript expression in this cell line demonstrated that raw C_T values ranged from 24.9 to 27.4. GAPDH was expressed at higher levels in this cell line than the three reference genes (C_T = 21.1–21.7). Normalization of potential NET marker gene levels, *Ki-67*, *MAGE-D2*, *MTA1* and *NAP1L1*, by either $geNorm_{ATZ}$ or *GAPDH* confirmed lower values (~10-fold less) using the former approach. In general, both approaches were reproducible, although interassay variability using $geNorm_{ATZ}$ was less than when using *GAPDH,* and gene expression showed less of a difference when normalized using $geNorm_{ATZ}$ compared with *GAPDH* (Spearman r = 0.77, P = 0.051, vs. r = 0.54, P = 0.149).

DISCUSSION

The ideal normalization gene(s) should *1*) exhibit low variability from sample to sample, *2*) be critical to the activity of the cell, and *3*) be consistently expressed in the primary tissue used in the experimental system. Because samples (biopsy or surgical specimens) obtained at surgery or during examination are heterogenous in terms of their cellular composition, it is particularly important to define a panel of genes that will reliably allow normalization of transcript levels of target genes in these samples. Furthermore, the genetic signature of the metastatic tumor discovered in lymph or liver samples may be of interest in the understanding of the molecular determinants of metastases, and hence the ideal housekeeping gene(s) in a

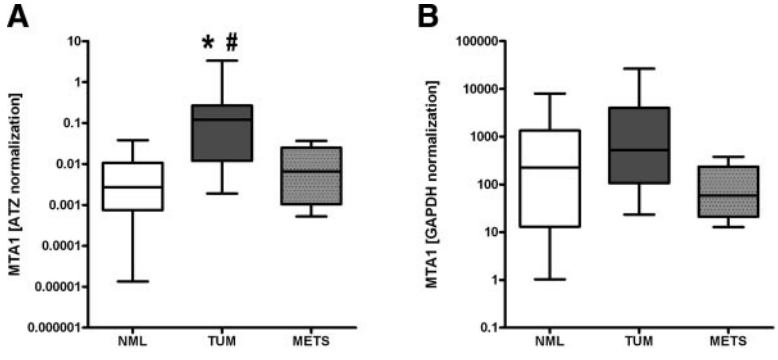

Fig. 3. *MTA1* gene expression normalized to either $geNorm_{ATZ}$ (ATZ; *A*) or *GAPDH* (*B*) in normal GI tissue (NML), primary tumors (TUM), and metastases (METS) using box and whisker plot analysis. The lower and upper quartiles and medians for each normalization approach are indicated. *MTA1* was significantly overexpressed in primary tumor when normalized by a combination of the 3 housekeeping genes, *ALG9*, *TFCP2*, and *ZNF410*. In contrast, normalization by *GAPDH* showed no significant overexpression of this candidate marker gene. *$P < 0.001$ vs. normal tissues. #$P < 0.05$ vs. metastases (2-tailed Mann Whitney test).

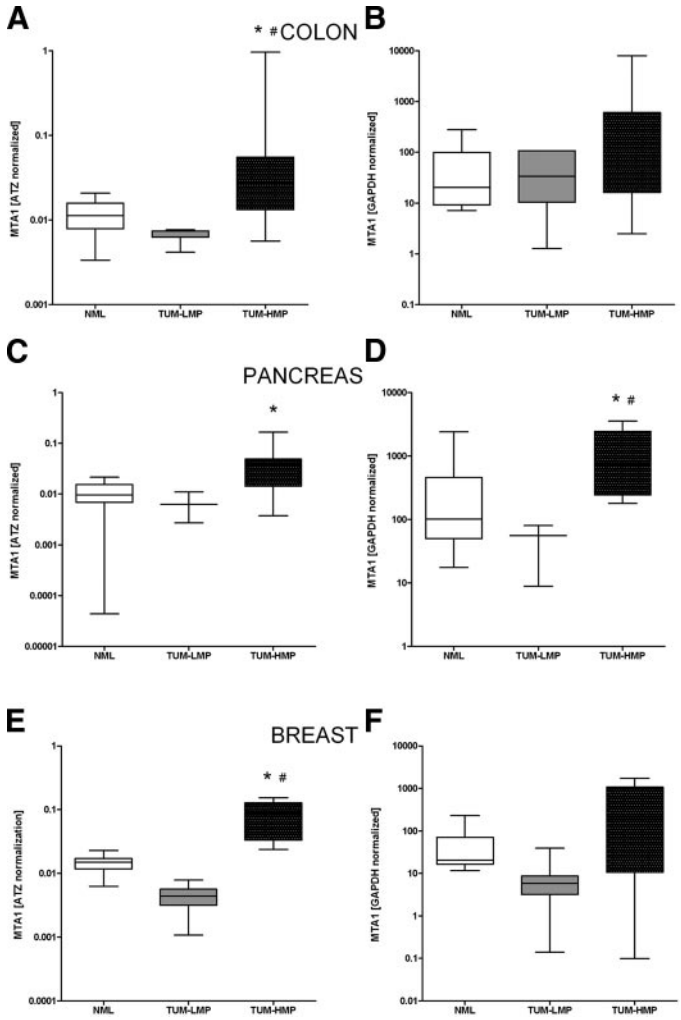

Fig. 4. Box and whisker plots of *MTA1* gene expression normalized to either geNorm$_{ATZ}$ (*A*, *C*, *E*) or *GAPDH* (*B*, *D*, *F*) in normal tissue (NML), primary adenocarcinomas with a low malignant potential (LMP; e.g., tumors without lymph node metastases: TUM-LMP), and primary adenocarcinomas with a high malignant potential (HMP; e.g., tumors with evidence of lymph node metastases: TUM-HMP) from the colon, pancreas, and breast. The lower and upper quartiles and medians for each normalization approach are indicated. *MTA1* was significantly overexpressed in HMP colon, pancreas, and breast tumors when normalized by a combination of the 3 housekeeping genes, *ALG9*, *TFCP2*, and *ZNF410*. In addition, HMP tumors of the colon and breast demonstrated significantly higher *MTA1* expression than LMP tumors from these organs. In contrast, normalization by *GAPDH* failed to show significant overexpression of *MTA1* in colon and breast tumors, irrespective of malignancy. Normalization of *MTA1* by *GAPDH* in pancreatic HMP tumors, however, resulted in significant overexpression compared with normal tissues and TUM-LMP. *$P < 0.04$ vs. normal tissues. #$P < 0.02$ vs. TUM-LMP (2-tailed Mann Whitney test).

panel should not vary from normal and malignant tissue to lymph node or liver tissue.

The selection of candidate reference genes that showed little variation but high expression on GeneChip arrays, followed by real-time PCR and geNorm$_{ATZ}$ analysis, resulted in the identification of three reference genes that can be used for normalization of PCR data from GI NETs and their metastases.

The three genes, *ALG9*, *TFCP2* and *ZNF410*, identified by geNorm to be the most robust reference genes exhibit the following features. ALG9 (asparagine-linked glycosylation 9 homolog) is encoded on chromosome 11 and catalyzes the transfer of mannose from Dol-P-Man to lipid-linked ligosaccharides (8); TFCP2 (transcription factor CP2) is encoded on chromosome 12 and is also recognized as α-globin transcription factor, CP2, with homology to *Drosophila* transcription factor Elf-1/NTF-1 (7); and ZNF410 (zinc finger protein 410), encoded on chromosome 14, is a transcription factor that activates transcription of matrix-remodeling genes such as MMP1 during fibroblast senescence (3). Expression of ZNF410 increases in senescent fibroblasts, but this only occurs at a protein level (3); mRNA levels appear to be constant throughout the fibroblast cell life span (3). Overall, relatively little is known about the transcriptional regulation of *ALG9*, *TFCP2* and *ZNF410*, although they are all considered to be transcriptionally ubiquitously expressed (3, 39). The different functions of these genes, a mannosyltransferase and transcription factors, involved in unrelated cellular processes further emphasizes their utility as reference genes. In our sample sets, these three genes had a significantly lower coefficient of variation than *GAPDH* and were not differently expressed in GI tissue and metastatic targets (lymph nodes and liver), and expression levels were tightly correlated.

Overexpression of *MTA1* mRNA and protein correlates with tumor invasion and metastasis in a variety of tumors including breast, hepatocellular, esophageal, gastric, pancreatic, and colorectal carcinomas (28, 40–43). Overexpression of this gene has been identified in small intestinal NETs (22), and expression levels are elevated in tumors with neuroendocrine features (14). An examination of the expression of this gene in normal GI tissue demonstrated that, following normalization, expression levels were similarly expressed in the stomach, small bowel, lymph nodes, and liver. Expression levels were significantly increased in primary NETs when appropriately normalized (geNorm$_{ATZ}$), but this difference was less evident following normalization with *GAPDH*. The latter was similar to the results of normalization of *MTA1* by the three less robustly expressed genes, *NCOR1*, *POL2RB* and *ZW10* (Fig. 1), which demonstrated an increase in expression in tumor samples but not significantly so ($P = 0.092$). These results demonstrating increased *MTA1* in primary tumors are supported by two earlier studies (14, 22) and suggest that a geNorm$_{ATZ}$ approach may be more sensitive than normalization approaches using *GAPDH* or other currently available reference genes.

Use of these three particular reference genes is not limited to normalization of marker genes in GI NETs. Approximately 40% of colon, pancreatic, and breast cancers overexpress *MTA1*, particularly when they exhibit lymph node metastases (10, 17, 18, 41). In our sample sets, *MTA1* expression following geNorm$_{ATZ}$ was significantly elevated in malignant tumors from each of these organs: colon, pancreas, and breast. Of note was the observation that, in colon and breast tumors, this approach segregated the malignant tumor group from the nonmalignant tumors, which were not different from normal tissue. Tumors from the pancreas were also segregated using *GAPDH* as a normalization strategy, indicating that, while our normalization approach has a broad utility, for more precise assessment of specific neoplasia, it may be necessary to consider identifying organ-specific reference genes. The approach of identifying candidate reference genes in microarray data-

bases provides an opportunity to identify suitable organ-specific reference genes for other organs and tumor types.

Identification of appropriate reference genes provides one mechanism to control for variation in real-time PCR studies. However, large interassay variability can be a feature of transcript measurements, particularly in clinical samples, with reports ranging from 2.7 to 25% (5, 9, 16). To minimize potential interassay variability, we examined gene expression in the small intestinal KRJ-I cell line. This was undertaken with the future objective of designing a PCR-based test useful in the clinical setting for quantitating transcript expression of candidate markers of malignancy.

KRJ-I is a continuous small intestinal carcinoid cell line with a rapid doubling time (1.7–1.9 days) that displays classic morphological and immunocytochemical features of an enterochromaffin cell carcinoid (20). It was established in 1992 from a multifocal ileal NET with an insular histological appearance (type I) (35). An examination of expression of four genes in this cell line demonstrated that geNorm$_{ATZ}$ for normalization was highly reproducible and provided the best interassay variability compared with a single gene-based normalization approach. This identification and characterization of marker gene expression in the KRJ-I cell line provides a reference point that can be included in future analyses of GI carcinoids and NETs. This will facilitate interassay comparisons, a requirement for any long-term laboratory analytical tool and of particular importance for tumors that are relatively rare (48).

On the basis of the material we examined, it appears that the routine use of *GAPDH* to normalize data from either GI NETs or colon and breast adenocarcinomas should be avoided, as this gene showed the greatest fluctuation among the samples examined. In this study, we have provided the rationale for examination of large transcript databases to identify reference genes in experimental set-ups that can be used as viable alternatives to the traditionally accepted housekeeping genes used thus far. This approach may be widely applicable to other neoplasia or biological situations where quantification of transcript expression is a critical element of the study design.

GRANTS

This work was supported in part by National Cancer Institute Grant R01-CA-115825, the Bruggeman Medical Foundation and the William Z. and Ida Bess Novick Fund.

REFERENCES

1. **Abrahamsen HN, Steiniche T, Nexo E, Hamilton-Dutoit SJ, Sorensen BS.** Towards quantitative mRNA analysis in paraffin-embedded tissues using real-time reverse transcriptase-polymerase chain reaction: a methodological study on lymph nodes from melanoma patients. *J Mol Diagn* 5: 34–41, 2003.
2. **Ahlund L, Nilsson O, Kling-Petersen T, Wigander A, Theodorsson E, Dahlstrom A, Ahlman H.** Serotonin-producing carcinoid tumour cells in long-term culture. Studies on serotonin release and morphological features. *Acta Oncol* 28: 341–346, 1989.
3. **Benanti JA, Williams DK, Robinson KL, Ozer HL, Galloway DA.** Induction of extracellular matrix-remodeling genes by the senescence-associated protein APA-1. *Mol Cell Biol* 22: 7385–7397, 2002.
4. **Bloomston M, Durkin A, Yang I, Rojiani M, Rosemurgy AS, Enkmann S, Yeatman TJ, Zervos EE.** Identification of molecular markers specific for pancreatic neuroendocrine tumors by genetic profiling of core biopsies. *Ann Surg Oncol* 11: 413–419, 2004.
5. **Boyle D, Rosengren S, Bugbee W, Kavanaugh A, Firestein G.** Quantitative biomarker analysis of synovial gene expression by real-time PCR. *Athritis Res Ther* 5: R352–R360, 2003.
6. **Buragas M, Kidd M, Modlin IM, Cha C.** Multiple gastrointestinal stromal tumors and synchronous ileal carcinoids. *Nat Clin Pract Oncol* 2: 166–170; quiz 161 p. following 170, 2005.
7. **Chae JH, Kim CG.** CP2 binding to the promoter is essential for the enhanced transcription of globin genes in erythroid cells. *Mol Cell* 15: 40–47, 2003.
8. **Frank CG, Aebi M.** ALG9 mannosyltransferase is involved in two different steps of lipid-linked oligosaccharide biosynthesis. *Glycobiology* 15: 1156–1163, 2005.
9. **Garson J, Grant P, Ayliffe U, RFerns R, Tedder R.** Real-time PCR quantitation of hepatitis B virus DNA using automated sample preparation and murine cytomegalovirus internal control. *J Virol Methods* 126: 207–213, 2005.
10. **Giannini R, Cavallini A.** Expression analysis of a subset of coregulators and three nuclear receptors in human colorectal carcinoma. *Anticancer Res* 25: 4287–4292, 2005.
11. **Goossens K, Van Poucke M, Van Soom A, Vandesompele J, Van Zeveren A, Peelman LJ.** Selection of reference genes for quantitative real-time PCR in bovine preimplantation embryos. *BMC Dev Biol* 5: 27, 2005.
12. **Grant G, Liu J, Stoeckert C.** A practical false discovery rate approach to identifying patterns of differential expression in microarray data. *Bioinformatics* 21: 2684–2690, 2005.
13. **Hoerndli FJ, Toigo M, Schild A, Gotz J, Day PJ.** Reference genes identified in SH-SY5Y cells using custom-made gene arrays with validation by quantitative polymerase chain reaction. *Anal Biochem* 335: 30–41, 2004.
14. **Hofer M, Tapia C, Browne T, Mirlacher M, Sauter G, Rubin M.** Comprehensive analysis of the expression of the metastasis-associated gene 1 in human neoplastic tissue. *Arch Pathol Lab Med* 130: 989–996, 2006.
15. **Huggett J, Dheda K, Bustin S, Zumla A.** Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun* 6: 279–284, 2005.
16. **Hummel K, Lowe L, Bellini W, Rota P.** Development of quantitative gene-specific real-time RT-PCR assaays for the detection of measles virus in clinical specimens. *J Virol Methods* 132: 166–173, 2006.
17. **Iguchi H, Imura G, Toh Y, Ogata Y.** Expression of MTA1, a metastasis-associated gene with histone deacetylase activity in pancreatic cancer. *Int J Oncol* 16: 1211–1214, 2000.
18. **Jang KS, Paik SS, Chung H, Oh YH, Kong G.** MTA1 overexpression correlates significantly with tumor grade and angiogenesis in human breast cancers. *Cancer Sci* 97: 374–379, 2006.
19. **Johnsson F, Joelsson B, Gudmundsson K, Greiff L.** Symptoms and endoscopic findings in the diagnosis of gastroesophageal reflux disease. *Scand J Gastroenterol* 22: 714–718, 1987.
20. **Kidd M, Eick G, Modlin I, Pfragner R, Champaneria M, Murren J.** Further delineation of the continuous human neoplastic enterochromaffin cell line, KRJ-I, and the inhibitory effects of lanreotide and rapamycin. *J Mol Endocrinol* 38: 181–192, 2007.
21. **Kidd M, Eick G, Shapiro MD, Camp RL, Mane SM, Modlin IM.** Microsatellite instability and gene mutations in transforming growth factor-beta type II receptor are absent in small bowel carcinoid tumors. *Cancer* 103: 229–236, 2005.
22. **Kidd M, Modlin I, Mane S, Camp R, Eick G, Latich I.** The role of genetic markers, NAP1L1, MAGE-D2 and MTA1, in defining small intestinal carcinoid neoplasia. *Ann Surg Oncol* 13: 253–262, 2006.
23. **Kidd M, Modlin I, Mane S, Camp R, Shapiro M.** Q RT-PCR detection of Chromogranin A: a new standard in the identification of neuroendocrine tumor disease. *Ann Surg* 243: 273–280, 2006.
24. **Kidd M, Modlin IM, Mane SM, Camp RL, Eick GN, Latich I, Zikusoka MN.** Utility of molecular genetic signatures in the delineation of gastric neoplasia. *Cancer* 106: 1480–1488, 2006.
25. **Korbonits M, Jacobs RA, Aylwin SJ, Burrin JM, Dahia PL, Monson JP, Honegger J, Fahlbush R, Trainer PJ, Chew SL, Besser GM, Grossman AB.** Expression of the growth hormone secretagogue receptor in pituitary adenomas and other neuroendocrine tumor. *J Clin Endocrinol Metab* 83: 3624–3630, 1998.
26. **Livak KJ, Schmittgen TD.** Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method. *Methods* 25: 402–408, 2001.
27. **Modlin IM, Sachs G.** *Acid Related Diseases: Biology and Treatment* (2nd ed.). Philadelphia, PA: Lippincott, Williams & Wilkins, 2004.

28. **Moon WS, Chang K, Tarnawski AS.** Overexpression of metastatic tumor antigen 1 in hepatocellular carcinoma: relationship to vascular invasion and estrogen receptor-alpha. *Hum Pathol* 35: 424–429, 2004.
29. **Moyana TN, Xiang J, Senthilselvan A, Kulaga A.** The spectrum of neuroendocrine differentiation among gastrointestinal carcinoids: importance of histologic grading, MIB-1, p. 53, and bcl-2 immunoreactivity. *Arch Pathol Lab Med* 124: 570–576, 2000.
30. **Nicot N, Hausman JF, Hoffmann L, Evers D.** Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot* 56: 2907–2914, 2005.
31. **Ohl F, Jung M, Xu C, Stephan C, Rabien A, Burkhardt M, Nitsche A, Kristiansen G, Loening SA, Radonic A, Jung K.** Gene expression studies in prostate cancer tissue: which reference gene should be selected for normalization? *J Mol Med* 83: 1014–1024, 2005.
32. **Olsvik PA, Lie KK, Jordal AE, Nilsen TO, Hordvik I.** Evaluation of potential reference genes in real time RT-PCR studies of Atlantic salmon. *BMC Mol Biol* 6: 21, 2005.
33. **Ontsouka EC, Reist M, Graber H, Blum JW, Steiner A, Hirsbrunner G.** Expression of messenger RNA coding for 5-HT receptor, alpha and beta adrenoreceptor (subtypes) during oestrus and dioestrus in the bovine uterus. *J Vet Med A Physiol Pathol Clin Med* 51: 385–393, 2004.
34. **Papouchado B, Erickson LA, Rohlinger AL, Hobday TJ, Erlichman C, Ames MM, Lloyd RV.** Epidermal growth factor receptor and activated epidermal growth factor receptor expression in gastrointestinal carcinoids and pancreatic endocrine carcinomas. *Mod Pathol* 18: 1329–1335, 2005.
35. **Pfragner R, Wirnsberger G, Niederle B, Behmel A, Rinner I, Mandl A, Wawrina F, Luo JS, Adamiker D, Hoger H, Ingolic E, Schauenstein K.** Establishment of a continuous cell line from a human carcinoid of the small intestine (KRJ-I): characterization and effects of 5-azacytidine on proliferation. *Int J Oncol* 8: 513–520, 1996.
36. **Radonic A, Thulke S, Bae HG, Muller MA, Siegert W, Nitsche A.** Reference gene selection for quantitative real-time PCR analysis in virus infected cells: SARS corona virus, Yellow fever virus, Human Herpesvirus-6, Camelpox virus and Cytomegalovirus infections. *Virol J* 2: 7, 2005.
37. **Shida T, Furuya M, Nikaido T, Kishimoto T, Koda K, Oda K, Nakatani Y, Miyazaki M, Ishikura H.** Aberrant expression of human achaete-scute homologue gene 1 in the gastrointestinal neuroendocrine carcinomas. *Clin Cancer Res* 11: 450–458, 2005.
38. **Sokmensuer C, Gedikoglu G, Uzunalimoglu B.** Importance of proliferation markers in gastrointestinal carcinoid tumors: a clinicopathologic study. *Hepatogastroenterology* 48: 720–723, 2001.
39. **Swendeman SL, Spielholz C, Jenkins NA, Gilbert DJ, Copeland NG, Sheffery M.** Characterization of the genomic structure, chromosomal location, promoter, and development expression of the alpha-globin transcription factor CP2. *J Biol Chem* 269: 11663–11671, 1994.
40. **Toh Y, Ohga T, Endo K, Adachi E, Kusumoto H, Haraguchi M, Okamura T, Nicolson GL.** Expression of the metastasis-associated MTA1 protein and its relationship to deacetylation of the histone H4 in esophageal squamous cell carcinomas. *Int J Cancer* 110: 362–367, 2004.
41. **Toh Y, Oki E, Oda S, Tokunaga E, Ohno S, Maehara Y, Nicolson GL, Sugimachi K.** Overexpression of the MTA1 gene in gastrointestinal carcinomas: correlation with invasion and metastasis. *Int J Cancer* 74: 459–463, 1997.
42. **Toh Y, Pencil SD, Nicolson GL.** Analysis of the complete sequence of the novel metastasis-associated candidate gene, mta1, differentially expressed in mammary adenocarcinoma and breast cancer cell lines. *Gene* 159: 97–104, 1995.
43. **Toh Y, Pencil SD, Nicolson GL.** A novel candidate metastasis-associated gene, mta1, differentially expressed in highly metastatic mammary adenocarcinoma cell lines. cDNA cloning, expression, and protein analyses. *J Biol Chem* 269: 22958–22963, 1994.
44. **Tricarico C, Pinzani P, Bianchi S, Paglierani M, Distante V, Pazzagli M, Bustin SA, Orlando C.** Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. *Anal Biochem* 309: 293–300, 2002.
45. **Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F.** Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3: RESEARCH0034, 2002.
46. **Vikman S, Essand M, Cunningham JL, de la Torre M, Oberg K, Totterman TH, Giandomenico V.** Gene expression in midgut carcinoid tumors: potential targets for immunotherapy. *Acta Oncol* 44: 32–40, 2005.
47. **Warrington J, Nair A, Mahadevappa M, Tsyganskaya M.** Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics* 2: 143–147, 2000.
48. **Wong ML, Medrano JF.** Real-time PCR for mRNA quantitation. *Biotechniques* 39: 75–85, 2005.