Evaluation of suitable control genes for quantitative polymerase chain reaction analysis of maternal plasma cell-free DNA

QIWEI YANG1,2, XIUYING LI2, HASSAN ABDPELLAH AHMED ALI1,2, SHAN YU1,2, YUCHENG ZHANG2, MEI WU2, SUJIE GAO1,2, GUANJIE ZHAO1,2, ZHENWU DU1,3 and GUIZHEN ZHANG1,3

1Central Laboratory, Second Hospital, Jilin University, Changchun, Jilin 130041;
2Scientific Research Center, China-Japan Union Hospital, Jilin University, Changchun, Jilin 130033;
3Orthopaedic Institute, Second Hospital, Jilin University, Changchun, Jilin 130041, P.R. China

Received November 16, 2014; Accepted August 4, 2015

DOI: 10.3892/mmr.2015.4334

Abstract. The content stability of commonly used control genes is considered to vary significantly in different independent experimental systems, either in the expression of RNA expression or in the level of DNA content. The present study aimed to examine a panel of six common control genes, including β-globin (HBB), telomerase (TERT), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), albumin (ALB), β-actin (ACTB) and T cell receptor γ (TRG), in order to evaluate and validate the most reliable control genes for quantitative polymerase chain reaction (qPCR) in investigations for the analysis of fetal-derived DNA and maternal-derived DNA in maternal plasma to enable non-invasive prenatal assessment.

Introduction
Following the confirmation by Lo et al (1) of the presence of cell-free fetal DNA (cffDNA) in maternal plasma and serum in 1997, investigations have focussed on the utilization of cffDNA in non-invasive prenatal testing (NIPT). To date, cffDNA analysis is widely used in numerous NIPT, including for fetal gender determination (2), Rhesus blood group D (RhD) antigen status determination (3), and for the assessment of monogenic diseases and chromosomal aneuploidies prenatally (4).

CffDNA is widely accepted to originate predominantly from the product of placenta trophoblast apoptosis (5), and exhibits a distinctive molecular characteristic. CffDNA molecules are generally <300 bps in length, while the maternally-derived cell-free plasma DNA are >300 bps in length (6,7). Based on these observations, it is possible to separate cffDNA molecules from the overwhelming quantity of maternal-derived DNA.

Several techniques are used for analyzing cffDNA, including methylated DNA immunoprecipitation, digital polymerase chain reaction (PCR) and massively parallel sequencing (8-10). Quantitative PCR (qPCR) is the most fundamental, cost efficient and common method used for cffDNA analysis, however, its accuracy is affected by a number of external and internal factors, including the quantity of the initial samples, the quality of templates and the PCR efficiency (11). Therefore, it is necessary to normalize the gene level. At present, the use of control genes as a standard normalizer (12) is the most common method. Control genes are commonly defined as genes, which ubiquitously exist at stable levels in different biological contexts and are used to confirm the presence and quality of DNA in each sample, as well as measure the quantity of total (maternal and fetal) DNA in each sample (13). However, no single universal and entirely constant control gene has been reported. Accumulating evidence has indicated that the content levels of widely used control genes vary significantly in different independent studies (14-16). Therefore, it is essential to compare and evaluate the content stability of each control gene prior to use for normalization in cffDNA analysis. To the best of our current knowledge, the commonly used control genes for cffDNA analysis are selected, almost without any preliminary evaluation of their content suitability.

Keywords: quantitative polymerase chain reaction, control gene, cell-free fetal DNA, maternal plasma DNA, normalization

Correspondence to: Professor Guizhen Zhang or Dr Zhenwu Du, Central Laboratory, Second Hospital, Jilin University, 265 Ziqiang Street, Changchun, Jilin 130041, P.R. China
E-mail: zhangguizhenjlu@163.com
E-mail: doom99106@163.com
The present study aimed to examine the content stability of six commonly used control genes, which exist as differently sized maternal plasma DNA molecules, including those >300 bps, considered maternally-derived DNA, and <300 bps, considered fetally-derived DNA. These control genes are β-globin (HBB), telomerase (TERT), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), albumin (ALB), β-actin (ACTB) and T cell receptor γ (TRG), and they were selected based on previous reports on cffDNA using the qPCR method. In the present study, three common statistical algorithm programs, geNorm (17), NormFinder (18) and BestKeeper (19), were used to evaluate the content stabilities of the six genes. The results of the present study aimed to reveal optimal control gene selections for further investigations on cffDNA.

Materials and methods

Plasma sample collection and DNA extraction. The present study was approved by the Ethical Committee of Second Hospital, Jilin University (Jilin, China). For the investigation, 2 ml of peripheral blood was collected from the cubital vein of 20 pregnant females (gestational age, 18.67±0.58 weeks) and written informed consent was obtained from each individual prior to commencement of the investigation. The blood samples were anticoagulated using EDTA (1.5%). DNA was extracted from the plasma of each sample using a QIAamp DNA mini kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions, within 4 h of blood collection.

Separation of maternal- and fetal-derived DNA. The extracted DNA was subjected to 1% agarose gel electrophoresis (Invitrogen Life Technologies, Carlsbad, CA, USA) (7,20), and was visualized under ultraviolet light (G-2008, Peiqing Science & Technology, Shanghai, China). Each lane was cut at a position of 300 bps into two discrete sections, according to the DL500 DNA marker (Takara Bio, Inc., Otsu, Japan) and extracted from the agarose using a AxyPrep Gel Extraction kit (Axygen Biosciences, Union City, CA, USA), according to the manufacturer's instructions. DNA with a length <300 bps was defined as fetally-derived DNA (fetal group) and DNA with lengths >300 bps was defined as maternal-derived DNA (maternal group).

qPCR analysis. The subsequent qPCR analysis was performed using an ABI PRISM 7500 Sequence Detection system (Applied Biosystems Life Technologies, Foster City, CA, USA). The primers of the control genes were synthesized by Sangon Biotech Co., Ltd. (Shanghai, China) and the sequences are presented in Table I.

The qPCR reactions were performed in a 20 µl volume containing DNA (8 ng) using a SYBR Premix Ex Taq kit (Takara Bio, Inc.), including 10 µl SYBR Premix Ex Taq (2X), 0.4 µl ROX Reference Dye (50X) and 1 µl forward/reverse primer (10 µM each), made up to 20 µl with deionized water, according to the manufacturer's instructions. The amplification was performed using an ABI PRISM 7500 Sequence Detection system (Applied Biosystems Life Technologies) and subjected to the following cycling steps: Initial step of 95°C for 10 min, followed by 50 cycles of 95°C for 15 sec, 58°C for 15 sec and 72°C for 30 sec. Each assay was performed four times. The results of the qPCR results were subjected to 1% agarose gel electrophoresis. To estimate the efficiencies of amplification, a standard curve was generated by Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) for each primer pair based on four points of serial 2-fold dilutions of the DNA template and performed qPCR reactions as described above.

Statistical analysis. Microsoft Excel was used to calculate the mean and standard deviation (SD) values. The amplification efficiencies were calculated using the slope of the calibration curve with the equation, E = 2^(-1/slope), following which the correlation coefficients (R² values) were determined.

The content stabilities of the six candidate control genes were assessed using three commonly used programs: geNorm (http://medgen.ugent.be/~jvdesomp/genorm/), NormFinder (http://moma.dk/normfinder-software) and BestKeeper (http://www.gene-quantification.de/bestkeeper.html), according to the manufacturer's instructions. In geNorm and NormFinder, the threshold cycle (Ct) values were converted into relative quantities using the 2^(-ΔΔCt) formula (ACTB = Ct - lowest Ct). For BestKeeper, the raw Ct values were used directly. These three programs were all based on Microsoft Excel, using different algorithms to determine the stability of the control genes.

Results

Amplification performance of primers. The qPCR amplification product was detected using 1% agarose gel electrophoresis and was of the expected size with no primer dimers (Fig. 1A). A single peak was obtained in each amplification reaction during the analysis of the dissociation curves, which confirmed the specific amplification of the primers (Fig. 1B). The sequences, corresponding amplicon sizes, PCR efficiencies of the primers and R² values are listed in Table I.

Amplification profile of the candidate control genes. The amplification profiles of the candidate control genes were estimated according to the Ct values of the six biological samples. As shown in Fig. 2, the mean Ct values of each gene in the maternal group and fetal group were determined.

The control genes exhibited Ct values varying between 32.78, for ACTB, and 38.74, for TERT, in the total samples (Fig. 2A). Among these genes, ACTB exhibited the lowest Ct value (32.25-33.43) and TERT exhibited the highest Ct value (37.23-40.22), followed by GAPDH (36.13-39.48), as shown in Fig. 2A. Among these six candidate control genes, TRG was the most variable in terms of content, with a high SD value (2.63) in the maternal group (Fig. 2B). ACTB was the candidate control gene with the lowest SD values (0.34 and 0.49 in the maternal and fetal group, respectively; Fig. 2B and C). No significant difference in Ct values were observed between the maternal and fetal group for any of the genes.

Content stability of the candidate control genes. According to the geNorm database, HBB and GAPDH were the most stable genes, with the lowest M values, which were followed by ACTB, in the total sample and fetal group, whereas ALB and TERT were ranked as the most stable genes in the maternal group (Fig. 3A-C). TRG was considered to be an unstable gene in all three groups. Notably, almost none of the pairwise
variation values were below the cutoff value (V=0.15), with the exception of V5/6 in the fetal group (Fig. 3D). This result indicated that combining five control genes together in the fetal group increased the stability for normalization. No optimal combination number of control genes were identified for normalization in the other groups.

Table I. Primer sequences, product sizes and PCR efficiency.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Primer sequence</th>
<th>Product size (bp)</th>
<th>PCR efficiency</th>
<th>R²-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBB</td>
<td>F-GTGCACTGACTCTTGAGGAGA R-CCTTGTATCAACCTGCCCAG</td>
<td>101</td>
<td>2.58</td>
<td>0.97</td>
</tr>
<tr>
<td>TERT</td>
<td>F-GGTGAACTCTCGTGTTATGCAAR-GGCACACTGGGTGCTGCC</td>
<td>97</td>
<td>2.00</td>
<td>0.97</td>
</tr>
<tr>
<td>GAPDH</td>
<td>F-GGACTGAGGCTCACCACCTTTT R-GCATGGACTGTTGCTGCAAA</td>
<td>157</td>
<td>1.72</td>
<td>0.99</td>
</tr>
<tr>
<td>ALB</td>
<td>F-TGAAACATAGTTACCCCAAGAGTTT R-CCTCTCTCTGCAAAAGGTTGCTCATAT</td>
<td>80</td>
<td>1.79</td>
<td>0.99</td>
</tr>
<tr>
<td>ACTB</td>
<td>F-CCTGTACGCCAACAAGTGC R-ATACTCCTGCTTGATCC</td>
<td>211</td>
<td>2.08</td>
<td>0.98</td>
</tr>
<tr>
<td>TRG</td>
<td>F-AGGGTTGTTGGAATCAAGG R-CGTGCAACAAAGTTGTCCTC</td>
<td>160</td>
<td>1.82</td>
<td>0.97</td>
</tr>
</tbody>
</table>

PCR, polymerase chain reaction; F, forward; R, reverse; HBB, β-globin; TERT, telomerase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; ALB, albumin; ACTB, β-actin; TRG, T cell receptor γ.

Figure 1. Specificity of primers and amplicon length. (A) 1% agarose gel electrophoresis of amplified fragments. (B) Dissociation curve of six control genes. HBB, β-globin; TERT, telomerase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; ALB, albumin; ACTB, β-actin; TRG, T cell receptor γ; M, marker.

Figure 2. Mean Ct values of the candidate control genes in (A) total samples, (B) maternal groups and (C) fetal groups (n=20). The dot represents the arithmetic mean; the bar indicates the minimal to maximal Ct value (standard deviation). Ct, threshold cycle; HBB, β-globin; TERT, telomerase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; ALB, albumin; ACTB, β-actin; TRG, T cell receptor γ.
The results of the NormFinder analysis indicated that HBB was the most content stable gene, with a stability value of 0.325, in the total samples (Fig. 4A). The optimal combination of two genes was determined to be HBB and GAPDH. The stability value of the HBB/GAPDH combination (0.249) was lower than that of HBB alone (0.325), which suggested that the combination of these two genes provided higher stability, compared with HBB alone (Fig. 4A). HBB and ACTB were identified as the most content stable genes in the maternal and fetal group, respectively (Fig. 4B and C).

The results of the BestKeeper analysis revealed that HBB demonstrated the highest stability in the total samples and in the maternal group, whereas GAPDH was determined as the optimal performer in the fetal group (Fig. 5A-C). On examination of the variance (Fig. 5D-F), the SD value of TRG was >1.00 in the total samples and maternal group, therefore, it was considered unacceptable and eliminated from the stability analysis.

Discussion

The identification of cfDNA in maternal plasma has become a primary target for NIPT (1). In healthy gravidae, cfDNA can be detected in maternal plasma as early as the seventh week following conception (21), which then increases as pregnancy progresses (22) and reaches a plateau in the ensuing three months, being cleared from the maternal plasma to become absent within 2 h of delivery (23). Furthermore, cfDNA molecules are generally shorter than 300 bps in length, whereas maternal-derived molecules are longer than 300 bps in length (6,7), which enables cfDNA molecules to be readily
separated from the original maternal DNA using electrophoresis. These properties have rendered cffDNA as an optimal material for NIPT. At present, qPCR is the most fundamental, cost efficient and commonly used method in investigations of cffDNA. Due to its low cost and ease of operation, qPCR is constantly being applied to attempt to diagnose numerous types of hereditary disease. To date, gender determination (24,25) and several diseases, including β-thalassemia (26,27), RhD fetal blood group genotyping (28-30), trisomy 21 (31) and X chromosome aneuploidies (32) have been successfully diagnosed using qPCR. In the process of quantitative investigations, control genes are important. A suitable control genes is required to be stably expressed in both maternal- and fetal-derived DNA. An ideal control gene in maternal plasma is that which is not affected or regulated by pregnancy conditions, stress response, stimulation or any other physiological or pathological state throughout the pregnancy process (33). However, there is accumulating evidence suggesting that the content levels of widely used control genes vary significantly in different independent investigations, for example the single-copy DNA control gene, HBB, which is used to represent the cell number has been suggested to be not the most reliable control gene (13). Our previous study also revealed that the content stability of widely used control genes for DNA demonstrated significant variation in the plasma DNA of pregnant and non-pregnant individuals (14). It is essential to normalize the control gene content levels and determine reliable control genes prior to any qPCR analysis. To the best of our knowledge, the present study is the first to evaluate the content stability of control genes commonly used in maternal- and fetal-derived DNA, respectively. The present study collected blood samples in the second trimester of gestation, at which stage the content of cffDNA is stable. Subsequently, six candidate control genes were assessed, including HBB, TERT, GAPDH, ALB, ACTB and TRG, which were estimated using the geNorm, NormFinder and BestKeeper statistical algorithms.

Onn analysis of the raw Ct values, ACTB exhibited the lowest mean Ct values, followed by HBB and TRG. By contrast, ACTB exhibited the lowest variation in content levels, indicated by the SD values, whereas TERT exhibited the highest mean Ct values. TRG exhibited the highest SD values, which indicated that its content varied markedly.

The optimal number of control genes for normalization was suggested by genes with a V-value below the cutoff value of 0.15 in geNorm (17). No optimal combination of the selected control genes had a V-value below the cutoff value, with the exception of the use of five genes in the fetal group. Thalita (35) reported that the combination of genes cannot increase the accuracy definitely and it is suggested, if conditions permit, that three of the most stable control genes are used, rather than a single gene (36). The number of control genes also depends on the experimental conditions.

Of note, the concentration of cffDNA in plasma is low (22) and the majority originates from the apoptosis of placental trophoblasts resulting in fragments shorter than 300 bps in length. These characteristics affect the PCR amplification of cffDNA, as the length of the cffDNA template is limited at 300 bps, whereas a longer template of the target gene increases the number of opportunities to be digested in the process of apoptosis (2). Therefore, amplicon sizes are required to be sufficiently short to ensure adequate effective templates for PCR amplification.
Increasingly, studies are focusing on the clinical application of cfDNA, which is relevant to NIPT. However, to the best of our knowledge, the control genes used in analysis of cfDNA are selected without confirmation of the content stability of these control genes in mature plasma DNA. The present study indicated that the most content stable control genes in maternal and fetal-derived DNA at the second trimester of gestational age, which can be used as a criterion in subsequent investigations.

In conclusion, the present study indicated that the content stability of control genes used for analyzing plasma DNA exhibited significant variation between maternal- and fetal-derived DNA, therefore, all qPCR performed to analyze cfDNA requires the initial selection of an appropriate control gene individually. The results of the present study also indicated that HBB in maternal- and fetal-derived DNA, and in maternal-derived DNA alone, and GAPDH in fetal-derived DNA enable efficient normalization for qPCR investigations in maternal plasma DNA. These results also present an appropriate strategy for the evaluation of candidate control genes for genomic DNA qPCR analysis.

Acknowledgements

This study was supported by the Project supported by the Key Foundation of Jilin Provincial Science & Technology Department, China (grant. no. 20130727038YY), the Jilin Provincial Science & Technology Department, China (grant. nos. 20120942 and 20110740) and the Jilin Provincial Development and Reform Commission, China (grant. no. 20101928).

References


