Evaluation of reference genes and normalization strategy for quantitative real-time PCR in human pancreatic carcinoma

Beatrice Mohelnikova-Duchonova\textsuperscript{a,b}, Martin Oliverius\textsuperscript{c}, Eva Honsova\textsuperscript{d} and Pavel Soucek\textsuperscript{a,*}
\textsuperscript{a}Department of Toxicogenomics, National Institute of Public Health, Prague, Czech Republic
\textsuperscript{b}First Faculty of Medicine, Charles University in Prague, Czech Republic
\textsuperscript{c}Department of Transplantation Surgery, Institute of Clinical and Experimental Medicine, Prague, Czech Republic
\textsuperscript{d}Department of Clinical and Transplantation Pathology, Institute of Clinical and Experimental Medicine, Prague, Czech Republic

Abstract. Histologically verified pairs (n = 10) of pancreatic tumors and non-neoplastic tissues were used for quantitative real-time PCR and the stability of 24 reference genes was analyzed with geNorm and NormFinder software. Raw $C_q$ values correlated with the degree of RNA degradation. This correlation was abolished by normalization to $C_q$ of 18S endogenous control gene. Both geNorm and NormFinder programs suggested EIF2B1, ELF1, MRPL19, and POP4 as the same most stable genes. We have thus identified suitable reference genes for future expression studies in pancreatic carcinoma. Normalization method reducing the effects of RNA degradation on the quality of results was also developed.

Keywords: Pancreas, carcinoma, transcript, quantification, reference gene, normalization

1. Introduction

Pancreatic cancer is the fourth leading cause of cancer death in the United States [1] as well as in the Czech Republic [2]. It has a dismal five-year survival less than 5% [3], primarily related to the fact that disease-specific symptoms occur late in the course of the disease. Pancreatic cancer belongs to one of the most difficult conditions to treat [4] and its etiology and molecular pathogenesis is still poorly understood.

Gene expression profiles are commonly used for identification of groups of genes involved in specific functional aspects of tumor biology and for the development of candidate biomarkers [5]. A number of methods have been developed to study gene expression.

The relative quantification by real-time PCR method (qPCR) is adequate strategy for most purposes where investigation of physiological changes in gene expression levels is conducted [6–8]. There are several essential requirements for successful expression analysis. The first and crucial step is correct sample collection, handling and preparation [9]. By the time of diagnosis no more than 20% of patients with pancreatic carcinoma have surgically operable disease [10]. Thus, sampling in large specialized centers rather than in numerous smaller surgical departments is necessary for practical reasons. Typically, pancreatic tumors have low neoplastic cellularity and a predominance of non-neoplastic fibrous (or desmoplastic) stroma. This is rather unique to duct adenocarcinomas of the pancreas; in contrast to infiltrating carcinomas arising in other organs. To overcome this obstacle potentially leading to spurious results, two approaches are generally used. Firstly, microdissection or other methods of purification of the epithelial component are utilized. Second-
ly, differences between macrodissected pancreatic tumors and non-neoplastic control tissues and eventually specimen from patients with chronic pancreatitis as an intermediate group are analyzed with stable cell lines as correctors [11]. In pancreatic tissues specifically, high-quality RNA preparation as the prerequisite condition for successful analysis presents an extremely difficult challenge. Subsequent assessment of RNA quality and quantity is the next critical step in expression analysis [6]. Pancreatic tissues contain a large amount of endogenous RNases. RNAses are a major secretory product of normal pancreatic acinar cells and cause extensive degradation of mRNA in pancreatic tissues. Moreover, there is a frequent loss of acinar cells during development of infiltrating pancreatic cancers due to atrophy or destruction of the gland by the tumor growth [11]. The content of RNase secreting cells thus may vary among tumors. Thus for laboratories aiming at expression profiling of pancreatic cancer, it is essential to introduce specific protocol for the handling of a pancreatic specimen. RNA isolation procedure and rigorous control of RNA and cDNA quality and quantity. The appropriate choice of internal standards (reference genes) is critical for qPCR performance. Ideally, the internal standards should be constitutively expressed in all cell types under study. Internal standards should also be independent of experimental conditions and unaffected by human disease conditions. Formerly, a number of housekeeping genes, which are necessary for basic cell survival, like glucose-6-phosphate dehydrogenase (GAPDH), ribosomal RNA subunits (18S and 28S rRNA), β-actin or cyclophilins were used as reference genes. However, it has become clear that these highly popular reference genes show distinct differences in gene expression in certain tissue types and considerably vary between normal and malignant tissues [12–14].

The purpose of our study was to assess performance of 24 reference genes by qPCR in a series of 10 macrodissected pairs of pancreatic carcinomas and adjacent non-neoplastic tissues. The robustness of qPCR analysis on samples with quality unavoidably compromised by RNA degradation was assessed as a major endpoint.

2. Material and methods

2.1. Samples

Tissue specimens were obtained from 10 surgically treated pancreatic carcinoma patients diagnosed over a period between August 2008 and May 2010. The resection specimens were immediately transferred from the Surgery to the Pathology Department, then macrodissected, snap-frozen in liquid nitrogen and stored at −80°C until RNA preparation. Histological diagnoses of pancreatic carcinomas including types, stages, and grades of the tumors were performed according to AJCC rules for classification [15]. Only histologically confirmed tumor and non-neoplastic tissue samples were used for RNA analysis. The following data on patients were retrieved from medical records: age, sex, date of diagnosis, pTNM stage, histological type and grade of the tumor (Table 1). All patients were asked to read and sign an informed consent in accordance with the requirements of the Ethical Commission of the Institute of Clinical and Experimental Medicine in Prague.

2.2. Isolation of total RNA and cDNA synthesis

Tissue samples were homogenized by mechanical disruption using Precellys instrument (Bertin Technologies, Montigny-le-Bretonneux, France) at a speed of 6,500 for 15 sec. Total RNA was isolated using Trizol Reagent according to manufacturer’s protocol
<table>
<thead>
<tr>
<th>Gene symbol (ID)</th>
<th>Assay ID</th>
<th>Location of probe</th>
<th>Amplicon length (bp)</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH (2597)</td>
<td>Hs02758991</td>
<td>3'-most, exon 7–8 boundary</td>
<td>93</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>HPRT1 (3251)</td>
<td>Hs99999909</td>
<td>3'-most, exon 6–7 boundary</td>
<td>100</td>
<td>hypoxanthine phosphoribosyltransferase 1</td>
</tr>
<tr>
<td>GUSB (2990)</td>
<td>Hs00999908</td>
<td>3'-most, exon 11–12 boundary</td>
<td>81</td>
<td>glucuronidase, beta</td>
</tr>
<tr>
<td>ACTB (80)</td>
<td>Hs00357333</td>
<td>exon 1–2 boundary</td>
<td>77</td>
<td>actin, beta</td>
</tr>
<tr>
<td>B2M (567)</td>
<td>Hs09999907</td>
<td>3'-most, exon 2–3 boundary</td>
<td>75</td>
<td>beta-2-microglobulin</td>
</tr>
<tr>
<td>HMBS (3145)</td>
<td>Hs00692979</td>
<td>3'-most, exon 1–2 boundary</td>
<td>64</td>
<td>hydroxyacyl-CoA dehydrogenase</td>
</tr>
<tr>
<td>IPO8 (10526)</td>
<td>Hs00183531</td>
<td>3'-most, exon 20–21 boundary</td>
<td>71</td>
<td>importin 8</td>
</tr>
<tr>
<td>PGK1 (5230)</td>
<td>Hs00943178</td>
<td>3'-most, exon 5–6 boundary</td>
<td>73</td>
<td>phosphoglucomutase 1</td>
</tr>
<tr>
<td>TRP (6908)</td>
<td>Hs00427621</td>
<td>exon 3–4 boundary</td>
<td>65</td>
<td>TATA box binding protein</td>
</tr>
<tr>
<td>TFRC (7037)</td>
<td>Hs00998911</td>
<td>exon 14</td>
<td>105</td>
<td>transferrin receptor</td>
</tr>
<tr>
<td>YWHAZ (7534)</td>
<td>Hs00237047</td>
<td>exon 1–2 boundary</td>
<td>70</td>
<td>tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, 1a, cytoplasmic</td>
</tr>
<tr>
<td>PPIA (5478)</td>
<td>Hs00156570</td>
<td>exon 3–4 boundary</td>
<td>94</td>
<td>peptidylglycine alpha-hydroxylating monooxygenase A</td>
</tr>
<tr>
<td>POLR2A (5430)</td>
<td>Hs00172187</td>
<td>3'-most, exon 1–2 boundary</td>
<td>61</td>
<td>polymerase (RNA) II (DNA-directed) poly(A) polymerase</td>
</tr>
<tr>
<td>GADD45A (1647)</td>
<td>Hs00196054</td>
<td>exon 3–4 boundary</td>
<td>123</td>
<td>growth arrest and DNA-damage-inducible, alpha</td>
</tr>
<tr>
<td>EIF2B1 (1967)</td>
<td>Hs00426752</td>
<td>3'-most, exon 4–5 boundary</td>
<td>75</td>
<td>eukaryotic translation initiation factor 2B, subunit 1</td>
</tr>
<tr>
<td>ABL1 (25)</td>
<td>Hs00245445</td>
<td>exon 3–4 boundary</td>
<td>91</td>
<td>c-abl oncogene 1, non-receptor tyrosine kinase</td>
</tr>
<tr>
<td>ELF1 (1997)</td>
<td>Hs00152844</td>
<td>3'-most, exon 2–3 boundary</td>
<td>76</td>
<td>E74-like factor 1</td>
</tr>
<tr>
<td>MRPL19 (9801)</td>
<td>Hs00608519</td>
<td>exon 2–3 boundary</td>
<td>72</td>
<td>mitochondrial ribosomal protein L19</td>
</tr>
<tr>
<td>PESI (23481)</td>
<td>Hs00362795</td>
<td>3'-most, exon 13–14 boundary</td>
<td>56</td>
<td>peptidylglycine alpha-hydroxylating monooxygenase A</td>
</tr>
<tr>
<td>RPL20 (6156)</td>
<td>Hs00256497</td>
<td>exon 4–5 boundary</td>
<td>149</td>
<td>ribosomal protein L30</td>
</tr>
<tr>
<td>RPS17 (6218)</td>
<td>Hs00236586</td>
<td>exon 2–3 boundary</td>
<td>88</td>
<td>ribosomal protein S17</td>
</tr>
<tr>
<td>POP1 (10775)</td>
<td>Hs00198357</td>
<td>exon 3–4 boundary</td>
<td>68</td>
<td>processing of precursor 4, ribonuclease P/MRP subunit</td>
</tr>
<tr>
<td>PSMC4 (5704)</td>
<td>Hs00197826</td>
<td>exon 6–7 boundary</td>
<td>83</td>
<td>proteasome 26S subunit, ATPase, 4</td>
</tr>
<tr>
<td>18S (10008588)</td>
<td>Hs00392899</td>
<td>not applicable</td>
<td>61</td>
<td>18S ribosomal RNA</td>
</tr>
</tbody>
</table>

* assay recommended by the manufacturer (for criteria see www.appliedbiosystems.com);
* Based on assay re-evaluation in August 2010 with NCBI Entrez Gene annotations, this assay may detect transcript(s) from off target gene(s) (see www.appliedbiosystems.com).

Table 2: Information about reference genes and the respective Assays

B. Mohelnikova-Duchonova et al. / Evaluation of reference genes for qPCR in human pancreatic cancer

qPCR was performed in 7500 Real-Time PCR System by help of TaqMan® Custom Plates (Applied Biosystems, Foster City, CA). For analysis of reference genes 4x24 well format was used comprising of TaqMan® Gene Expression Assays with optimized primer and probe sets (Table 2). The following criteria were applied to the selection of TaqMan® Gene Expression Assays: i/ exon-exon boundaries span where possible in order to minimize signals from traces of
contaminating DNA, ii/ as short as possible amplicons and iii/ location near to 3'-end of the transcript to reduce influence of RNA quality on qPCR. Priority was given to assays from the validated endogenous control set. The list of reference genes and respective assays is given in Table 2. Each reaction contained in total 20 μl of premixed cDNA and 2x TaqMan® Gene Expression Master Mix (Applied Biosystems) in 1:1 ratio. The amount corresponding to 25 ng of total RNA used for cDNA synthesis was loaded per each sample well. Cycling parameters were: 2 min. at 50°C and initial denaturation at 95°C for 10 min, followed by 50 cycles consisting of denaturation at 95°C for 15 sec and annealing at 60°C for 60 sec.

2.4. Analyses of stability of reference genes

Firstly, ΔCq was calculated for each sample where \( \Delta C_q = C_q(\text{sample}) - C_q(\text{calibrator}) \), i.e. non-neoplastic sample with the lowest average \( C_q \) from all tested samples, i.e. ZPI5. Then relative quantity \( Q = e^{\Delta C_q} \), where efficiency \( E \) was considered as 100%, i.e. \( 2^{\Delta C_q} \), value was calculated and used for analysis of stability of reference genes. The study design adhered to the MIQE Guidelines (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) published in 2009 [9]. geNorm version 3.5 of March 2007 [17] and NormFinder version 19 of June 2009 [18] software programs were used for evaluation of stability of reference genes according to manuals available at the website for their download. Differences between groups were analyzed by non-parametric Wilcoxon test and correlations were analyzed by non-parametric Spearman’s test using SPSS version 15.0 software.

3. Results

3.1. RNA quantity and quality assessment

The total RNA was isolated from surgically removed tumor and non-neoplastic tissues of 10 patients with pancreatic carcinoma diagnosis. The median RIN of tumor and non-neoplastic tissue samples was 6.5 ± 1.0 (range 4.2–7.4) and of 3.3 ± 1.6 (2.1–6.5), respectively. RIN values differed significantly between tumors and non-neoplastic tissues (p = 0.005, Wilcoxon test). cDNA was synthesized using 2 μg of total RNA. Samples free of genomic DNA contamination as assessed by PCR for ubiquitin C fragment were used for subsequent analyses.
3.2. Quantitative real-time PCR

We analyzed transcript levels of 24 candidates for reference genes for qPCR studies of pancreatic tumor and non-neoplastic tissue samples. The mean raw $C_q$ values for all genes except 18S in tumors and non-neoplastic tissues were $27.06 \pm 3.47$ and $27.27 \pm 2.88$, respectively. The median raw $C_q$ values in tumors and non-neoplastic tissues were 27.02 and 27.24, respectively. The raw $C_q$ values were highly dependent on the degree of RNA degradation represented by RIN. Significant correlation of RIN with the mean of raw $C_q$ for all studied genes except 18S was observed in control tissues ($p = 0.010$, Fig. 3A). Non-significant trend was found in tumors ($p = 0.074$, Fig. 3A) and the effect became highly significant when all tissues were analyzed together ($R^2 = 0.57$, $p < 0.001$; data not shown). After normalization of data by calculation of $C_q$(reference gene)/$C_q$(18S endogenous control gene) none of the correlations with RIN remained significant ($p = 0.253$ in controls, $p = 0.501$ in tumors, and $p = 0.841$ in all tissues; Fig. 3B).

3.3. Identification of reference genes

The relative quantity $Q$ defined in Materials and Methods was used for analysis of stability of reference genes by geNorm and NormFinder. Sample ZPI5 with average $C_q = 25.27 \pm 2.44$ and RIN = 5.3 was used as calibrator. Both programs showed similar results, i.e. identified the same genes among the five most stable genes ($EIF2B1$, $ELF1$, $MRPL19$ and $POP4$) and suggested 2–3 reference gene combinations as sufficient for qPCR in pancreatic carcinoma. qPCR efficiency was then evaluated for the four best reference genes, i.e. $EIF2B1$, $ELF1$, $POP4$ and $MRPL19$ by assessment of serial dilutions of calibrator sample (ZPI5). Analysis of the efficiency, $R^2$, and slope of calibration curves and $C_q$ in non-template controls (NTCs) proved that all genes were suitable for further use in qPCR as reference genes (Table 3).

4. Discussion

Research of pancreatic cancer is hindered by difficulties with the recruitment of pancreatic cases with brief survival and poor performance status. Additionally, in inoperable patients (the majority in pancreatic cancer) the verification of diagnosis is complicated by the lack of pathological specimen. Moreover, pancreatic specimens present a hurdle in isolation of intact RNA due to the presence of high amounts of RNases. Therefore, it is essential to validate the protocol of sample handling and RNA isolation from pancreatic tissues for qPCR purposes and minimize the effect of poor RNA quality.
The aim of our study was to identify the most stable expressing endogenous control genes in tissue specimens from pancreatic cancer patients as potential reference genes for qPCR. Special attention was paid to the evaluation of the effect of RNA stability and assessment whether or not proper normalization can influence this effect.

We have found that under the same standard conditions of sample collection (tumor and non-neoplastic tissues macrodissected and snap frozen at the same time), storage and RNA isolation, the quality of RNA (by RIN comparison) from non-neoplastic pancreatic specimens is considerably lower than that from the respective tumor samples ($p = 0.005$). This was true for all of the inspected ($n = 10$) sample pairs and therefore we can speculate that the RNAse content may generally be higher in non-neoplastic tissues. It is assumed that the ischemia time from ligation of the duodenal/pancreatic vessels to the withdrawal of tissue for analysis is too long and causes massive RNA degradation. Since this ischemia time is an unalterable component of the operation, RNA analyses of pancreatic tissues present very difficult procedure [19]. For isolation of total RNA, Trizol Reagent-based procedure was used according to the recent literature [20,21]. Thus, the method of RNA isolation most probably did not cause the observed differences in RNA stability between samples. Our observation confirmed the previously published results that RNA isolated from solid tissues usually has the RIN between 6 and 8, but that from
The gastrointestinal tract has great RIN variation with RIN around 4 under experimental conditions with bovine organs [22]. The relatively low RIN values also precluded laser capture microdissection, which lowers the RIN by 1–2 points in pancreatic specimens [21]. Fleige et al. [6] found a significant impact of RIN on results of expression analysis, mainly the quantification cycle (Cq, formerly Ct) and minor effect on PCR efficiency. Antonov et al. [23] demonstrated in a biologically homogenous system of intact and partially hydrolyzed RNA derived from a cell line that degradation-related shifts of Cq values can be compensated by calculating delta Cq values between test genes and the mean Cq values of several reference genes. These delta Cq values were less sensitive to fragmentation of the RNA and were unaffected by varying amounts of input RNA [23]. In our experimental setting RIN significantly correlated with mean raw Cq values of 24 tested genes. Therefore, we used normalization of data by calculation of Cq (reference gene)/Cq (18S endogenous control gene). This normalization abolished the observed correlation with RIN. Thus, we demonstrated that the degradation-related Cq shifts caused by RNA fragmentation can be compensated by normalizing of the Cq of the reference gene and potentially the gene of interest to Cq of the endogenous 18S control gene.

We have screened 24 candidates for reference genes selected preferably from commercially available validated endogenous controls. There are several approaches for selection of most stable reference genes for qPCR in the literature, including the use of various programs for evaluation of results. In order to provide an overview of their performance, we have used two examples of the most frequently used software (NormFinder and geNorm) for evaluation of results. Both of these programs gave consistent results, although when rigorously evaluated, the most stable genes were not picked in the same order. However, EIF2B1, ELF1, MRPL19 and POP4 were among the first five most stable genes in both NormFinder and geNorm output. The amplification efficiency of target genes and reference genes should be tested before their use in relative quantification by qPCR. Slope and linearity of calibration curves should be analyzed as well [24]. We have shown that all selected genes perform well in qPCR (efficiency in the range 90–110%, R2 ≥ 0.998) and thus we recommend the use of these genes as reference genes for qPCR in pancreatic cancer research. Rubie et al. [25] tested 21 reference genes by qPCR in five gastrointestinal cancers. Tumor and normal tissues from 10 patients with pancreatic carcinoma were followed, amongst others. 18S-RNA, glutaminyl-tRNA synthetase (QARS), phosphomannomutase (PMMA1), RNA polymerase II polypeptide L (POLR2L) and β-glucuronidase (GUSB) showed the most stable expression. Analysis of agarose gel electrophoretic profiles was used as RNA quality control instead of RIN. In this study the comparative ∆Cq method was used. Results were then transformed to linear form by calculation of 2-∆Cq (mean Cq pathological tissue-mean Cq calibrator) [25]. Our study of the same size (n = 10) considered a bit different selection of genes (7 of 21 previously used genes were included in our study). However, GUSB was not confirmed among the five most stable genes by any of the programs used in our study. HPRT, GAPDH, PGK1, ACTB, and B2M did not perform well in both our and the previous study [25]. The published microarray data contain more than 2,500 potential reference genes [26]. It seems intriguing to compare the results of this study with appropriate selection of the most stable reference genes from microarray data. However, the recently discussed incom-
plete information on tissue composition and/or the use of well-defined tissues severely hampers any meaningful microarray data analysis and conclusive data interpretation [27]. In context with the fact that repeatability of microarray data is considered limited [28] such comparison may be an interesting goal for future investigations after major shortcomings are resolved.

In conclusion, we have identified four potential candidates as reference genes for future qPCR studies in pancreatic cancer research. We also suggest the normalization method which may reduce the effect of poor RNA quality on the quality of results.

Acknowledgements

This study was supported by grants of Grant Agency of the Ministry of Health of the Czech Republic, grant no. 97994 and 10230-3.

References

Submit your manuscripts at http://www.hindawi.com