

Article

# Identification and Validation of Novel Reference Genes in Acute Lymphoblastic Leukemia for Droplet Digital PCR

Vanessa Villegas-Ruiz <sup>1,\*</sup> , Karina Olmos-Valdez <sup>1</sup>, Kattia Alejandra Castro-López <sup>1</sup>, Victoria Estefanía Saucedo-Tepanecatí <sup>1</sup>, Josselen Carina Ramírez-Chiquito <sup>1</sup>, Eleazar Israel Pérez-López <sup>1</sup>, Isabel Medina-Vera <sup>2</sup>  and Sergio Juárez-Méndez <sup>1,\*</sup> 

<sup>1</sup> Experimental Oncology Laboratory, Research Department, National Institute of Pediatrics, Mexico City 04530, Mexico; kary\_230794@hotmail.com (K.O.-V.); kattia.castrolpz@gmail.com (K.A.C.-L.); saucedo9116@hotmail.com (V.E.S.-T.); carina-rc@outlook.com (J.C.R.-C.); eleazarisraelp@yahoo.com.mx (E.I.P.-L.)

<sup>2</sup> Research Methodology Department, National Institute of Pediatrics, Mexico City 04530, Mexico; isabelj.medinav@gmail.com

\* Correspondence: vanessavillegasruiz@yahoo.com.mx (V.V.-R.); sjuarezm@pediatria.gob.mx (S.J.-M.)

Received: 25 March 2019; Accepted: 7 May 2019; Published: 17 May 2019



**Abstract:** Droplet digital PCR is the most robust method for absolute nucleic acid quantification. However, RNA is a very versatile molecule and its abundance is tissue-dependent. RNA quantification is dependent on a reference control to estimate the abundance. Additionally, in cancer, many cellular processes are deregulated which consequently affects the gene expression profiles. In this work, we performed microarray data mining of different childhood cancers and healthy controls. We selected four genes that showed no gene expression variations (*PSMB6*, *PGGT1B*, *UBQLN2* and *UQCR2*) and four classical reference genes (*ACTB*, *GAPDH*, *RPL4* and *RPS18*). Gene expression was validated in 40 acute lymphoblastic leukemia samples by means of droplet digital PCR. We observed that *PSMB6*, *PGGT1B*, *UBQLN2* and *UQCR2* were expressed ~100 times less than *ACTB*, *GAPDH*, *RPL4* and *RPS18*. However, we observed excellent correlations among the new reference genes ( $p < 0.0001$ ). We propose that *PSMB6*, *PGGT1B*, *UBQLN2* and *UQCR2* are housekeeping genes with low expression in childhood cancer.

**Keywords:** references genes; droplet digital PCR; gene expression; leukemia

## 1. Introduction

Cancer is one of the most common health problems in the world and the National Cancer Institute (NCI) estimates 1,735,350 new cases in the United States in 2018; of these, 10,270 cases are childhood cancer (<https://www.cancer.gov>). Thus, the identification and implementation of new markers that help to diagnose, prognose and target cancer are urgent priorities. Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy worldwide and in Mexico [1]. The five-year survival is > 90% in developed countries [2–4]; however, in developing countries such as Mexico, the survival rates are very low [5], possibly due to late diagnosis and several other factors, including risk classification, cytogenetics and immunological and molecular alterations [6–9].

In the postgenomics era, several groups have focused on omics to study different alterations in cancer. Currently, several molecular markers have been identified in a wide variety of malignancies, including SNPs, DNA gain and loss, epigenetic modification, coding RNA, noncoding RNA and protein expression and their modifications. The cytogenetic alterations in several types of cancer have an important impact on the clinical prognosis of ALL, such as *BCR-ABL* [10], *FUS-ERG* [11],

*ETV6-RUNX1* [12], *E2A-PBX1* [13] and *KMT2A-AFF1*. However, less than 20% of leukemia patients have an alteration.

Gene expression (GE) is a spectacular cell process that is tissue-driven and the dynamics of GE are modulated by several factors that include: environmental, microenvironmental, intracellular or extracellular processes, among others. Nevertheless, RNA diversity is not completely clear; bioinformatic studies have shown that only ~2% of the transcriptome promotes protein diversity, and the alternative splicing (AS) of mRNA provides an exceptional capacity to generate a complex proteome. Additionally, RNA expression has been used as a molecular marker in diverse malignancies, such as mendelian disorders [14], tuberculosis [15] and some types of cancer [16]. Nevertheless, its measurement is not completely uniform because it is dependent on cellular conditions, such as stress, temperature, O<sub>2</sub> concentrations and overgrowth in vitro, which could influence modifications in GE. Thus, it is very important to establish methods for RNA quantification and reference genes (calibrator or housekeeping) that do not show variations between patients and healthy conditions. The expression level of specific transcripts in clinical diagnosis plays an important role in prognosis, treatment assignment and overall survival.

Furthermore, the GE level could be influenced by several technical factors, including contamination by alcohol, phenol, proteins; RNA integrity; and cDNA synthesis. In addition, RNA quantification could have some technical variations caused by RNA absolute quantification. Thus, the use of reference genes (housekeeping) is very important for normalizing GE and detecting changes in the expression of potential biomarkers. Third-generation PCR, known as droplet digital PCR (ddPCR), has proven to be a powerful tool for determining gene expression. It has higher accuracy, sensitivity and several advantages because it has minor inhibitory effects on PCR and internal normalization is unnecessary for detection, however it is recommendable for gene expression analysis with ddPCR [17]. It provides a great opportunity to absolutely measure the nucleic acid (DNA/RNA) and identify rare allelic variants, nonabundant RNA transcripts and copy number variations (CNVs), among others. This platform has proven to be a powerful tool with the ability to precisely count the absolute abundance of transcripts or DNA. The molecular abundance is critical in the diagnosis of several pathologies, and the clinical relevance in prognosis, such as detecting HER2 and EEF2 in breast cancer [18], improves the precision and accuracy of diagnosis.

Reports of potential molecular GE markers in ALL remain uncertain due to the lack of uniformity and reproducibility in the method criteria of real-time PCR, which focus mainly on normalization genes, and then the measured level expression of candidate genes in the study. The selection of a housekeeping gene is the first step in the normalization of an interest gene. However, several studies on RNA transcriptomics analysis in ALL have reported using a single and conventional constitutive gene; unfortunately, it is still common practice today [19,20]. This failure can lead to inaccuracy and miscalculation on GE and poor reproducibility. Therefore, it is necessary to improve critical variables in the quantitative transcripts to improve the precision and accuracy of potential GE in ALL [21].

On the other hand, ddPCR is a powerful method for absolute quantification of DNA/RNA. Moreover, GE is not constant in all transcriptomes; some transcripts are overexpressed and some others are poorly expressed. Thus, the aim of this work was to identify new candidate RNA transcripts with constitutive expression levels obtained by microarray data mining that did not have variations in the level expression using ddPCR in ALL. We optimized the ability to digitally count potential RNA transcripts and determined homogeneous expression versus level expression of housekeeping genes for ALL. These high quantitative and precise constitutive transcripts in ALL allow for the availability and validation of ALL samples using ddPCR. Therefore, it is of great importance to develop and identify new reference genes (NRGs) with low expression levels in childhood cancer, in special ALL and then have precise comparisons with classical reference genes (CRGs) of low or high expression using ddPCR.

## 2. Materials and methods

### 2.1. Gene Expression Data Mining

The data were obtained from the ArrayExpress and we selected the Affymetrix GeneChip 1.0. We downloaded 768 microarray experiments that corresponded to childhood malignancy, healthy tissue and cancer cell lines. The data included in the study are listed in (Table S1). Then, we detected microarray quality control (MQC) variations according to a previous study [22–24]. We excluded all samples that showed changes in the intensity signal of the microarray controls, similar to the previous analysis [24]. Our QC analysis included 193 microarrays in the next analysis (Table S2). The microarray data analysis was archived using Partek Genomics version 6.6, Santa Clara, CA, USA. Bioinformatic analysis was performed using quantile normalization, probeset summarization by Median Polish, background correction by robust microarrays analysis (RMA) and, finally, the data underwent  $\log_2$  transformation. To identify genes that do not vary in expression levels by comparison between healthy tissues and tumors (including cancer cell lines), we used the exon intensity signals analyzed by ASANOVA with an unadjusted  $p$ -value  $> 0.95$ .

### 2.2. Tumor Samples

In this study, 40 samples of bone marrow diagnosed with ALL were obtained with previous signed informed consent. The protocol was approved by the Institutional Ethics Committee (INP protocol 060/2016) in accordance with the Declaration of Helsinki. The bone marrow samples were treated with lymphoprep density gradient medium (STEMCELL Technologies, USA) according to the protocol for cell pellets after RNA isolation. We used RNA from the SUP-B15 cell line to standardize the methodologies in this study. SUP-B15 was cultured in Iscove's modified Dulbecco's medium according to the manufacturer's instructions (ATCC). The culture medium was supplemented with 20% fetal bovine serum (Biowest, Riverside, Kansas, USA) and penicillin/streptomycin (100 U/mL, ATCC, Manassas, VA, USA). Cells were cultured at 37 °C, 5% CO<sub>2</sub> and a humidified atmosphere.

### 2.3. RNA Extraction

The cultured cells and patient cells were disrupted using a TissueLyser system (Qiagen, Valencia, CA, USA) for 60 s at 25 Hz. Total RNA was extracted using 1 mL of TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. RNA was resuspended in 45  $\mu$ L of DEPC water and quantified using a NanoDrop One UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

Then, RNA was quantified with the Qubit RNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) using a Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) according to the protocol. RNA samples were stored at  $-70$  °C until further use.

### 2.4. cDNA Synthesis

Before cDNA synthesis, RNA samples were subjected to DNase treatment (Thermo Fisher Scientific, Waltham, MA, USA), which included 1000 ng of total RNA, 1  $\mu$ L of 10 $\times$  Buffer and 1U of DNase. DEPC-treated water was added for a final reaction volume of 10  $\mu$ L. The reaction was incubated at 37 °C for 30 min, the cycle was stopped, 1  $\mu$ L of 50 mM EDTA was added and the reaction was incubated at 65 °C for 10 min. cDNA synthesis was performed using the RevertAid Transcriptase Kit (Thermo Fisher Scientific, Waltham, MA, USA). The master mix contained 1 $\times$  reaction buffer, 1 mM dNTP Mix, 100 pmol Random hexamers, 10 U RiboLock RNase Inhibitor and 200 U RevertAid Reverse Transcriptase and DEPC water up to 20  $\mu$ L. The reactions were incubated at 25 °C for 10 min, 42 °C for 60 min and 70 °C for 10 min. The final concentration of cDNA samples was 50 ng/ $\mu$ L.

### 2.5. RT-PCR Amplification

We evaluated the cDNA synthesized using 25 ng of cDNA for end point PCR. The reaction contained 1× KAPA2G Mix, 0.5 μM forward and reverse primers and DEPC water up to 15 μL. The primers are listed in Table 1. The amplification program consisted of predenaturation at 95 °C for three minutes, followed by 38 cycles of 95 °C for 15 s, extension for 15 s (T<sub>m</sub> for each constitutive transcript are shown in Table 1) and 72 °C for 15 s, with a final extension of 72 °C for five minutes in a Proflex PCR System thermal cycler (Applied Biosystem Inc., Foster City, CA, USA). The PCR products were separated by electrophoresis in 2.0% agarose gels (g/v) and stained with Sybr Gold (Thermo Fisher Scientific, Waltham, MA, USA) (1:10,000) at 90 volts for 35 min in 0.5% TAE solution. The amplification sequence was as follows: RPL4-a and RPS18-a first, followed by the rest of the genes listed with the letter q for quantitative RT-PCR.

**Table 1.** Primer list.

Name	Sequences 5' → 3'	Amplicon Size (bp)	T <sub>m</sub> (°C)
<i>ACTB</i>	Fw 5'-TCACAAATGTGGCCGAGGACTTT-3' Rv 5'-AGAAGTGGGGTGGCTTTTAGGATG-3'	115	60
<i>GAPDH</i>	Fw 5'-CTCAACGACCACTTTGTCAAGCTC-3' Rv 5'-CTCTCTCCTCTTGCTCTTGCT-3'	147	60
<i>RPL4-a</i>	Fw 5'-CGAATGAGAGCTGGCAAAGGCAAA-3' Rv 5'-ACGCCAAGTGCCGTACAATTCATC-3'	243	60
<i>RPL4</i>	Fw 5'-GTGGGACGTTTCTGCATTG-3' Rv 5'-TGTGCATGGGAAGATTGTAGT-3'	112	60
<i>RPS18-a</i>	Fw 5'-AATCCACGCCAGTACAAGATCCCA-3' Rv 5'-TTTCTTCTTGGACACACCCACGGT-3'	241	58
<i>RPS18</i>	Fw 5'-CAGCCAGGTCCTAGCCAATG-3' Rv 5'-CCATCTATGGGCCGAATCT-3'	82	60
<i>UBQLN2</i>	Fw 5'-CAGCCTGAAGGATCAGTGTAGT-3' Rv 5'-AGGGTCTCTTTATGGGAGAAGC-3'	84	60
<i>UQCR2</i>	Fw 5'-CCTGCGGGGTGATGTTGATA-3' Rv 5'-CAGCTACTTCCCAACGACGA-3'	83	60
<i>PGGT1B</i>	Fw 5'-CTGTGGTTTCCGAGGCTCTT-3' Rv 5'-GCATGAGAGGCCAGTGTAGG-3'	118	60
<i>PSMB6</i>	Fw 5'-TGACACCTATTCACGACCGC-3' Rv 5'-GGACCAGTGGAGGCTCATTC-3'	129	60

### 2.6. Quantitative RT-PCR

Thereafter, cDNA of the SUP-B15 cell line was analyzed by quantitative RT-PCR through amplification of *ACTB*, *GAPDH*, *RPL4*, *RPS18*, *PGGT1B*, *PSMBG*, *UBQLN2* and *UQCR2* (Table 1). We used the Kapa Sybr Fast qPCR Master Mix 2× Kit (Kapa Biosystems Inc., Wilmington, MA, USA). The reaction included master mix with the addition of DEPC water up to 10 μL, 2× Kapa Sybr Fast reagent qPCR Master Mix containing MgCl<sub>2</sub> at a final concentration of 2.5 mM, 10 μM forward and reverse primers and 10 ng of cDNA. Quantitative RT-PCR was performed on a Step One Real-Time PCR System (Applied Biosystems Inc., Foster City, CA, USA). The reactions were incubated at 95 °C for 10 min, then 40 cycles for amplification at 60 °C with a denaturation at 95 °C for 15 s and finally a melting curve at 53 °C for 1 min with a quantification every 0.3 °C for 15 s up to 95 °C.

### 2.7. Droplet Digital PCR

ddPCR was performed using QX200 ddPCR EvaGreen SuperMix (Bio-Rad, Hercules, CA, USA) in a QX200 Droplet Digital PCR System (Bio-Rad). Briefly, a PCR mixture containing 10 μL of 2× QX200

ddPCR EvaGreen SuperMix, 0.5  $\mu\text{L}$  of each set of primers at 5  $\mu\text{M}$ , 4  $\mu\text{L}$  of cDNA and  $\text{H}_2\text{O}$  necessary for a total reaction of 20  $\mu\text{L}$ . The Automated Droplet Generator loaded 20  $\mu\text{L}$  of each reaction mix in the DG8 Cartridge onto QX200 Droplet Generation system for droplet generation. The droplets were transferred into a 96-well plate and subsequently amplified using the following cycling parameters: predenaturation step at 95  $^\circ\text{C}$ , 35 amplification cycles of 95  $^\circ\text{C}$  for 15 s, 60  $^\circ\text{C}$  for 30 s, one cycle of 4  $^\circ\text{C}$  for 30 s, 90  $^\circ\text{C}$  for 5 min and finally a hold at 4  $^\circ\text{C}$ . Finally, the dropletized PCR in a 96-well plate was read by the Q $\times$ 200 Droplet Reader. The Poisson-corrected determination of template concentration was calculated using QuantaSoft™ Analysis Pro Software (v1.0, Bio-Rad). For quantification, a minimum of 10,000 acceptable droplets per 20  $\mu\text{L}$  reaction was used, followed by manual selection of positive and negative droplet populations. All experiments included a no-cDNA template control (NTC) and 0.2 ng/ $\mu\text{L}$  of cDNA for CRGs, and 5 ng/ $\mu\text{L}$  of cDNA was used to quantify the number of copies/20  $\mu\text{L}$  by well for NRGs.

### 2.8. Statistical Analysis

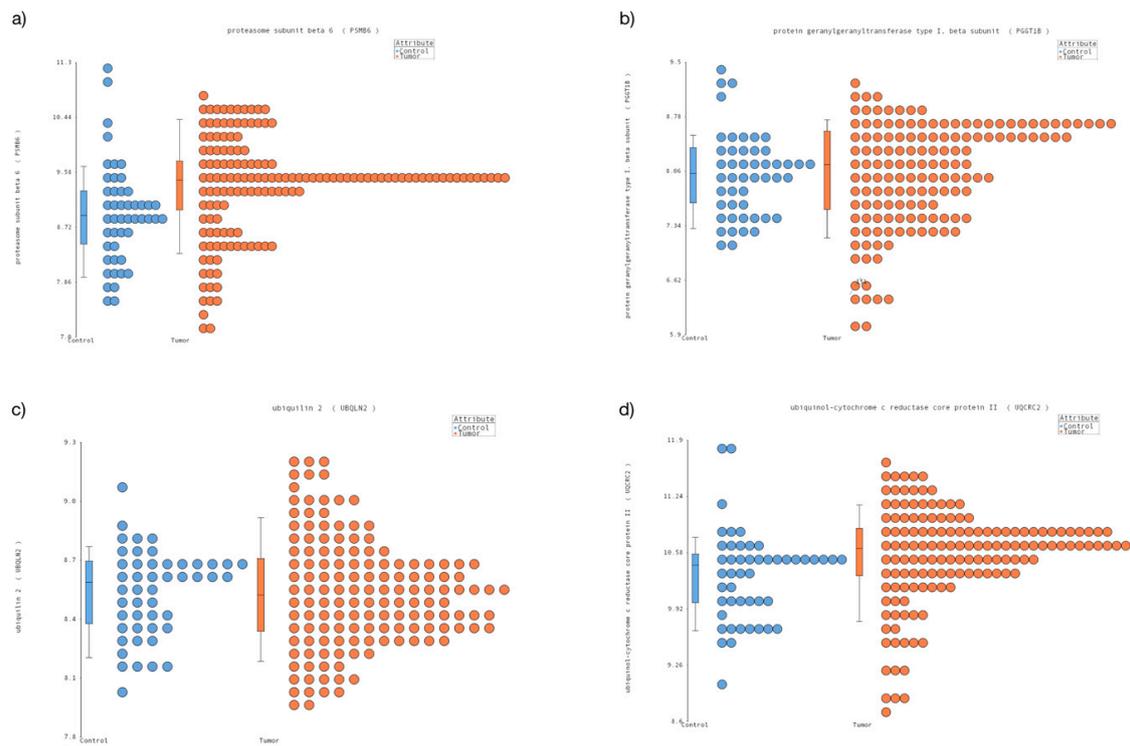
Variables were assessed using the Kolmogorov–Smirnov Z test to examine sample distribution. We correlated the absolute quantifications of the GE among *RPL4*, *RPS18*, *ACTB*, *GAPDH*, *UQCR2*, *PSMB6*, *UBQLN2* and *PGGT1B*. The correlation was achieved using the Pearson correlation coefficient. We consider statistical significance  $p < 0.05$ . Data analysis was performed using SPSS v.25 software for Macintosh (IBM Corp., Armonk, NY, USA). Curve fitting was performed in GraphPad Prism by least squares optimization.

## 3. Results

### 3.1. New Candidates Reference Genes in Pediatric Cancer

First, we performed data mining of different pediatric tumors and control tissues using Affymetrix GeneChip 1.0. Our data mining identified and downloaded 768 microarrays, including osteosarcoma, Hodgkin lymphoma, rhabdomyosarcoma, retinoblastoma, neuroblastoma, medulloblastoma, leukemia, bone marrow, retina, fibroblasts, skeletal muscle, bone, liver and different cancer cell lines of these tumors (Table S1). First, we identified the relationship by sample among microarray internal controls according to the manual of GeneChip 1.0 as follows: bioB, bioC, bioD and Cre at final concentrations of 1.5, 5, 25 and 100 pM, respectively. PolyA RNA controls were evaluated using Dap, Thr, Phe and Lys to final concentrations of 1:7500, 1:25,000, 1:50,000 and 1:100,000, respectively, as previously reported [24]. Our results show that ~75% of the available data showed differences in relation concentrations of almost one hybridization and/or polyA RNA control. Thus, we selected 193 microarray files that showed no differences among the controls (Table S2).

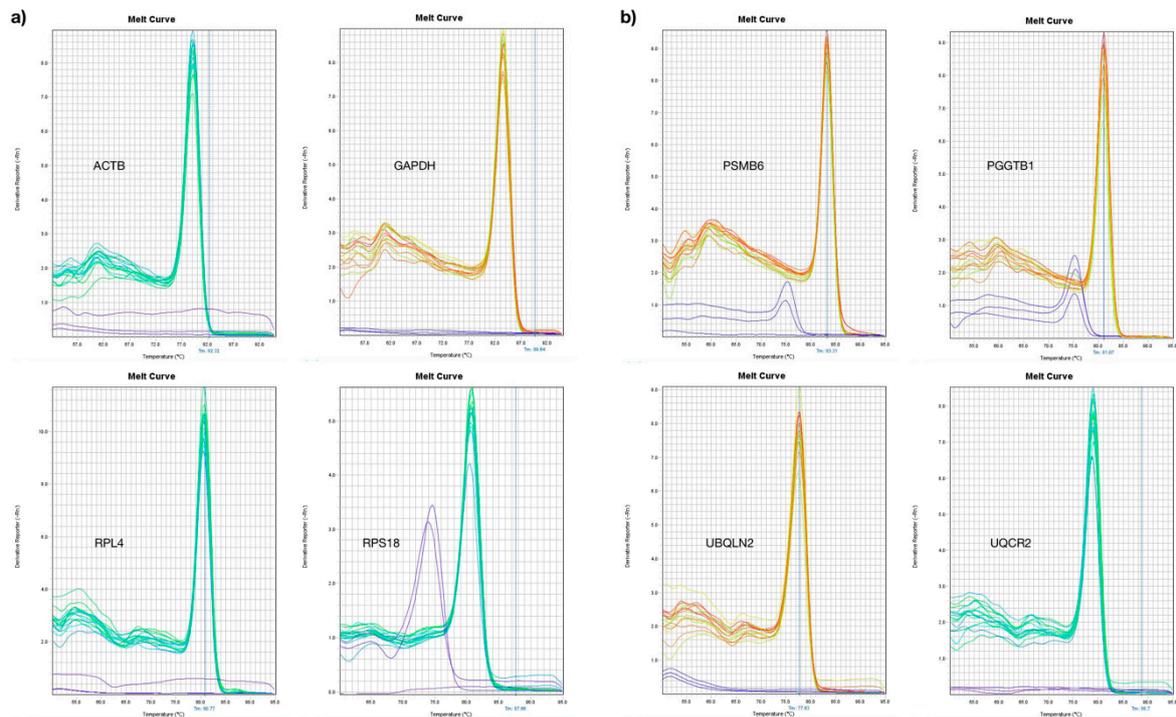
Then, we classified the microarray files in healthy tissues (control,  $n = 45$ ) and cancer tissues (tumor,  $n = 148$ ) and we performed a comparative analysis using a control as a reference; however, we selected the transcripts without varying GE. Our analysis showed that 189 transcripts from healthy and tumor samples did not show significant changes in GE. After visual inspection, we selected 10 transcripts (Figure 1, Figure S1), however only four transcripts (*PSMB6*, *PGGT1B*, *UBQLN2* and *UQCRC2*, Figure 1) were validated by quantitative RT-PCR and droplet digital PCR.



**Figure 1.** Nonvarying genes were expressed in childhood cancer and healthy tissue. The dot plot shows the level of expression on the  $x$  axis. (a) *PSMB6* (b) *PGGT1B*, (c) *UBQLN2*, (d) *UQCR2*. In the four transcripts, we observed homogenous expression between healthy tissues and cancer samples. The blue dots show the healthy tissue and the orange dots show the childhood cancer.

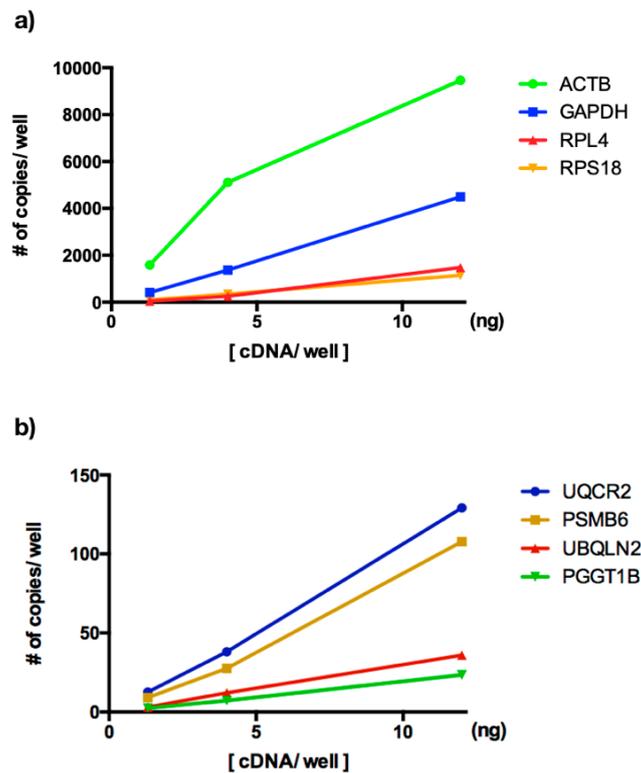
### 3.2. ddPCR Assay Optimization for Analysis of New Reference Genes

To contrast our results, we selected four reported CRGs (*GAPDH*, *ACTB*, *RPS18* and *RPL4*). Thus, we compare the GE by end point PCR. Although all samples were quantified by spectrometry and fluorometry, we observed some variations in GE, most likely due to sample conditions and RNA quality, among others. Thus, we evaluated all transcripts using qRT-PCR using five serial dilutions with an initial concentration of 50 ng/ $\mu$ L, with a factor dilution of 1:5. The results showed good kinetic and linearity of amplification for all genes that were evaluated. We observed no unspecific amplicons in the melting curves for the positive samples and we observed apparent dimer primers in *RPS18*, *PSMB6P* and *PGGT1B* (Figure 2).



**Figure 2.** Melting curves of the expressed nonvarying genes. (a) The classical reference genes *ACTB*, *GAPDH*, *RPL4* and *RPS18* are shown in the figure. We used five serial dilutions of cDNA SUP-B15 and we observed one curve as expected. Only *RPS18* showed up as a different curve in the no-template control, probably due to primer dimers. (b) The plot shows the new reference genes *PSMB6*, *PGGTB1*, *UBQLN2* and *UQCR2*. We observed one curve in the positive dilution. In *PSMB6* and *PGGTB1*, we observed other curves in the no-template control, probably caused by primer dimers.

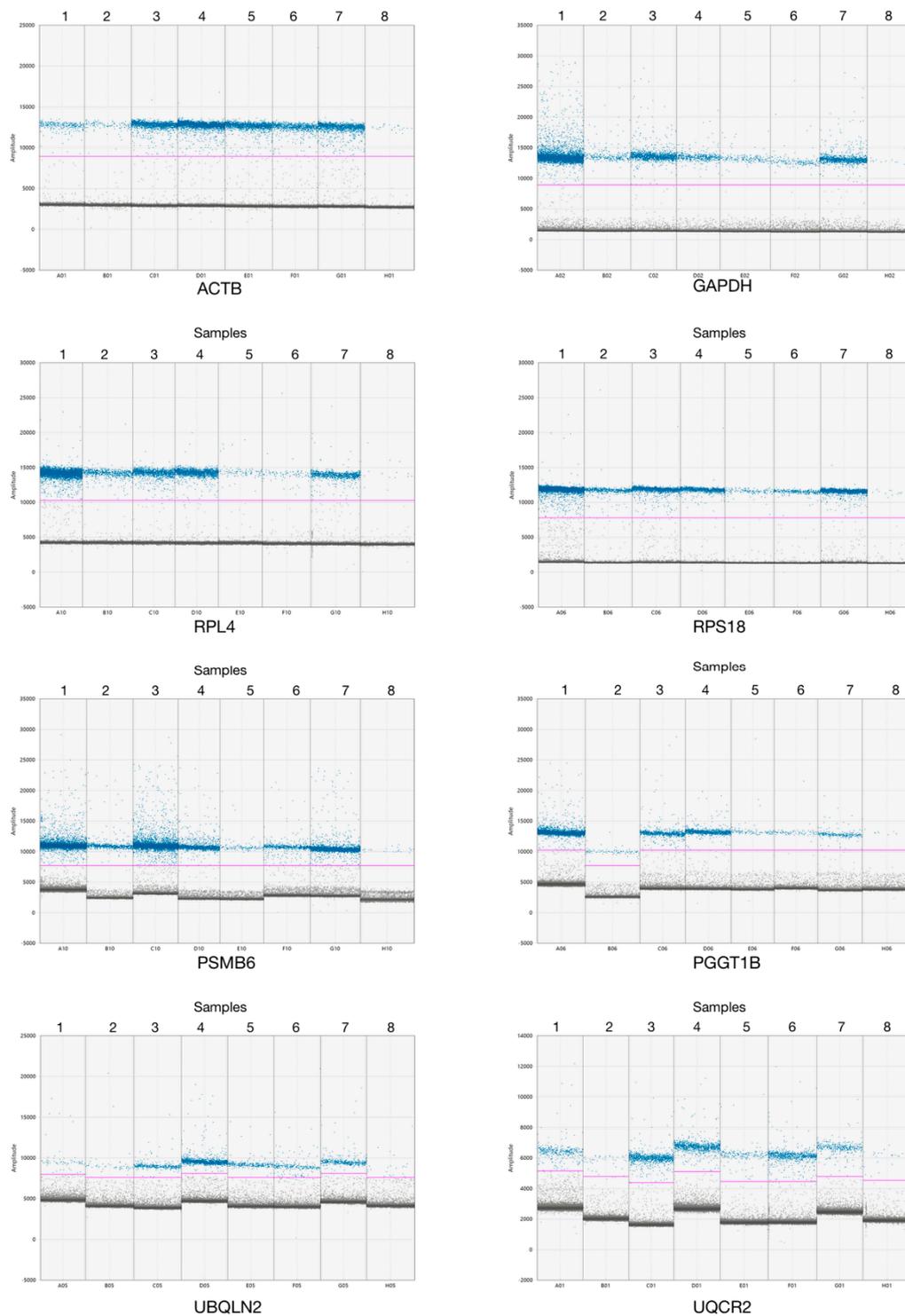
Next, we established ddPCR conditions for CRGs and NRGs as follows. First, we optimized the assay conditions to minimize droplet rain in terms of cDNA concentration, annealing temperature and primer concentration. We constructed an amplification curve using the SUP-B15 cell line for the eight transcripts using three samples tested with a serial dilution of 1:3. We obtained good linearity for eight genes with a regression coefficient of 0.99. Second, we plotted ng/reaction versus number of copies/well for eight reference genes and, interestingly, we observed that CRGs are expressed ~100 times more than NRGs (Figure 3). Additionally, we observed that the reaction was saturated to 12 ng/well for CRGs (Figure S2A), while the NRGs showed low expression at the same template concentration (Figure S2B), avoiding droplet rain in ddPCR for a highly concentrated sample.



**Figure 3.** Amplification curves by absolute quantification. The  $y$ -axis shows # copies/well and the axis  $x$ -axis shows the cDNA concentration used in ddPCR. (a) Copy number transcript by well of *ACTB*, *GAPDH*, *RPL4* and *RPS18*. *ACTB* was the most highly expressed housekeeping gene. (b) Copy number transcript by well of *PSMB6*, *PGGT1B*, *UBQLN2* and *UQCR2*. *UQCR2* was the most highly expressed following *PSMB6*.

### 3.3. Reference Genes Validation in Acute Lymphoblastic Leukemia

We quantified the absolute RNA expression of all evaluated transcripts. Our results showed that 0.8 ng/well is enough to determine the expression of these transcripts (CRGs), while 20 ng/well is enough to evaluate the transcripts with low expression (NRGs). After that, we evaluated the eight transcripts in 40 ALL samples. We observed a different copy number of reference transcripts, although all samples were quantified by spectrometry and fluorometry, of which eight representative samples of all reference genes were plotted (Figure 4). We observed differences in GE between CRGs and NRGs by sample. These results suggest that the ddPCR assay allowed us to discriminate genes with low and high expression.

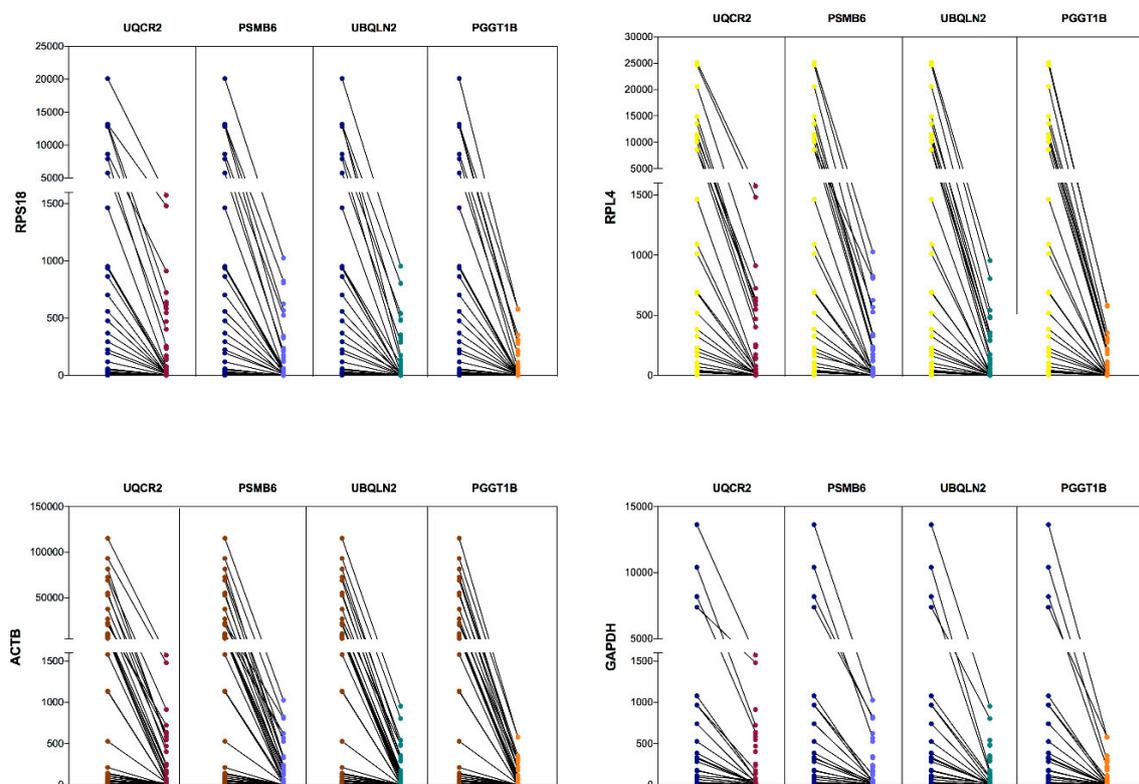


**Figure 4.** Droplet digital PCR amplification of the eight patients with acute lymphoblastic leukemia (ALL). Each plot represents the amplification of the eight reference genes evaluated in eight samples. The y-axis shows the amplitude signal and the x-axis shows the samples. Blue dots indicate positive amplification droplets and gray dots indicate negative amplification droplets. The pink line represents the cut-off of the positive and negative droplets.

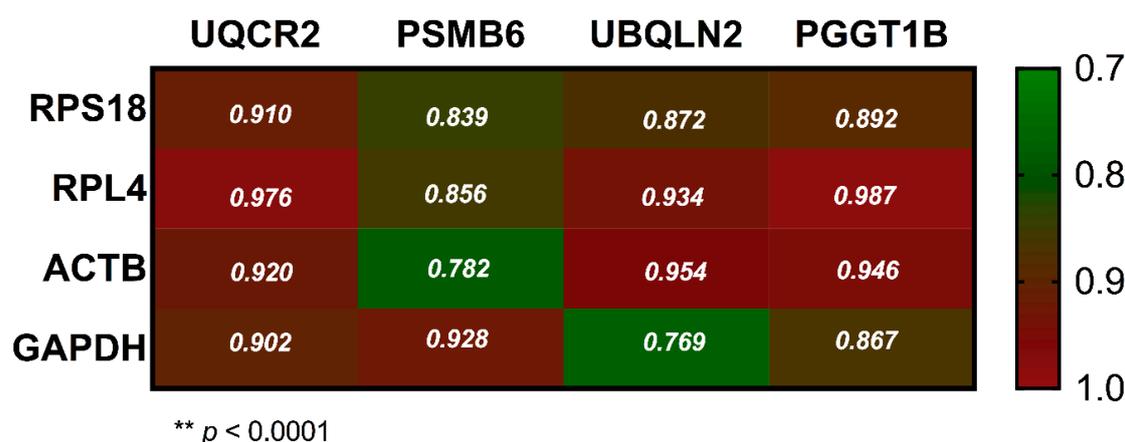
### 3.4. Comparative Expression Between CRGs and NRGs in Leukemia

To confirm whether the expression of each type of gene was the same, we plotted a bar graph to show paired data for the same sample. The scatterplots shown in the y-axis represent the number

of copies/ng normalized for CRGs and NRGs. First, each CRGs (*RPL4*, *RPS18*, *GAPDH*, *ACTB* and *GAPDH*) were compared with the four NRGs (*PSMB6*, *PGGT1B*, *UBQLN2* and *UQCR2*) in the copies/ng normalized in all samples that were evaluated. We systematically reviewed all comparisons and each scatterplot revealed the same patterns of change and clearly showed that the level of expression of CRGs was higher than that of NRGs (Figure 5). However, the level of expression for each CRGs in 40 ALL samples was proportional with respect to NRGs. This outcome revealed that patterns of expression level were similar between CRGs and NRGs. For this, it was necessary to evaluate the correlation of the absolute quantification of the expression between CRGs and NRGs (Figure 6). As expected, we obtained a high correlation between the gene studies, with the minimum value of 0.782 for *PSMB6* versus *ACTB*, while the highest correlation value was 0.987 for *RPL4* versus *PGGT1B*, both with  $p$ -values  $< 0.0001$ . The Spearman correlation helped to identify that all NRGs have a high value correlation with all CRGs; interestingly, *UQCR2* and *PGGT1B* showed a major correlation with all CRGs.



**Figure 5.** The gene expression comparison between the classical reference versus new reference genes in ALL. The top left plot shows the expression level of the *RPS18* gene versus *UQCR2*, *PSMB6*, *UBQLN2* and *PGGT1B*. The top right shows the level of *RPL4* gene expression versus *UQCR2*, *PSMB6*, *UBQLN2* and *PGGT1B*. The lower left shows the expression levels of *ACTB* versus *UQCR2*, *PSMB6*, *UBQLN2* and *PGGT1B*. The lower right shows the expression levels of the *GAPDH* gene versus *UQCR2*, *PSMB6*, *UBQLN2* and *PGGT1B*.



**Figure 6.** Pearson's correlations of the classical references versus new reference genes. We observed significant correlations  $p < 0.0001$  for all transcripts evaluated. The correlation is shown at the intersection. We observed that *UQCR2* and *PGGT1B* showed the best correlations with *RPS18*, *RPL4* and *GAPDH*. The lowest correlation was observed between *UBQLN2* versus *GAPDH* and *PSMB6* versus *ACTB* with 0.769 and 0.782, respectively.

#### 4. Discussion

Cancer is an important public health problem worldwide. In Mexico, leukemia is the most common childhood malignancy. There has been an increase in the number of new cases diagnosed, suggesting that the number of patients will be higher in the coming years. The challenge in cancer is the detection and sensitivity of the biomarkers. Many studies have focused on the identification of molecular markers in several types of cancer using microarray and next-generation sequencing.

In the last year, public data has been useful to identify molecular patterns in several types of cancer, mutations [25], AS [26] and molecular classification [24]. We performed microarray data mining of several types of childhood cancers and we inspected the quality control microarrays and observed that 70% of the data were rejected as we previously reported [22].

On the other hand, qRT-PCR quantification was performed based on a standard curve and the GE results are expressed as a fold change [27–29]. ddPCR is a robust methodology that has revolutionized nucleic acid quantification in providing sensitivity detection [30–33]. In ddPCR, each reaction is analyzed independently, the reaction is fractionated and the account positives may then use a Poisson correction to determine the absolute counts by target present in each sample [34] and facilitate measurement of individual target molecules [32]. Thus, ddPCR is a powerful method that achieves reproducibility, accuracy and sensitivity for the detection of molecular markers, disease monitoring and viral load, among others [32]. Therefore, ddPCR has been employed in the identification of low abundance molecules, such as fusion genes, minimal residual disease, SNPs and GE, among others [32,35,36].

RNA is a dynamic molecule of which its expression is tissue-dependent and the RNA quantification is dependent on cDNA synthesis efficiencies, sample degradation and RNA isolation [37]. Thus, it is very important to use the reference genes to normalize GE. Additionally, ddPCR provides high sensitivity for the detection and expression of relevant transcripts which is very important for normalization versus genes of equal or similar expression. Therefore, the comparison between genes of low and high expression would be incompletely comparable data because in ddPCR analysis, the determination of positive and negative droplets is a critical step; thus, the cDNA will be adjusted for unsaturation of the positive drop.

Several reports have focused on the comparison of the reference genes between qPCR versus ddPCR [17,38]. However, there are no reports about the use of the reference genes as normalizers of GE using ddPCR. In this study, we established the amount of cDNA/well in eight transcripts to be absolute quantification which allowed us to identify that the CRGs are expressed ~100 times more than NRGs (Figure 3). Previously, the transcript accounts were evaluated in the leukemia cell line

SUP-B15 and subsequently normalized to 20 ng/well (Figure 3). Our results showed that *ACTB* was the most highly expressed gene, with the following relationships: 1:2.1, 1:6.4, 1:8.3, 1:73, 1:88, 1:263, 1:404, *ACTB:GAPDH*, *ACTB:RPL4*, *ACTB:RPS18*, *ACTB:UQCR2*, *ACTB:PSMB6*, *ACTB:UBQLN2* and *ACTB:PGGT1B*.

To date, there are a few reference genes to use in ddPCR for GE. However, the validation of GE in ALL showed that the expression level among samples was different despite the fact that the samples were quantified by fluorometry and spectrometry and the samples were adjusted to 1 µg for cDNA synthesis (Figure 4). These results suggest that some technical factors affect the efficiency of the RT reaction or are characteristic *per se* of the sample [37]. However, when we analyzed the expression of the eight transcripts by sample, we observed that the proportion of the expression is conserved (Figure 5). Some reports have shown that the reference genes change the GE between study conditions when a relative quantification is performed by qPCR [39].

In our study, we analyzed the expression of eight reference transcripts using qRT-PCR and ddPCR. After that, we validated the expression in 40 patients with ALL for ddPCR and the CRGs were expressed ~100 times more than the NRGs. Additionally, we observed a high correlation among the eight transcripts analyzed (Figure 6). Our results suggest that the new transcripts could be used as reference genes in ALL.

In this work, we propose four NRGs that could be used to normalize genes of which have expression levels that are inconspicuous and provide a better comparison between samples and experimental conditions. The RNA molecule is very dynamic and its quantification is dependent on cDNA synthesis, RNA integrity and characteristics of the sample. Therefore, the absolute RNA quantification of the relevant genes in cancer is necessary to normalize a reference gene with a similar level of expression to obtain accurate results.

## 5. Conclusions

We validated the expression of the eight reference genes in 40 ALL patients and found that the classical reference gene had a high level of expression in contrast to the four NRGs *PSMB6*, *PGGT1B*, *UBQLN2* and *UQCR2*, which showed a high correlation with CGRs. Finally, we propose these new genes as housekeeping genes in childhood cancer. Therefore, the use of these transcripts as a normalizer will be dependent on the target expression level.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/10/5/376/s1>, Figure S1: Six candidate genes that did not vary in gene expression in childhood cancer and healthy tissues. Figure S2: Amplification linearity of reference genes by ddPCR assay. Table S1: Database of childhood cancer microarrays for data mining of ArrayExpress. Table S2: Database of childhood cancer microarrays that passed quality control and were then used for the gene expression analysis.

**Author Contributions:** Conceptualization, V.V.-R. and S.J.-M.; microarray data mining and PCR end point K.O.-V., K.A.C.-L., V.E.S.-T. and J.C.R.-C.; sample acquisition, E.I.P.-L.; ddPCR and qRT-PCR, V.V.-R.; statistical analysis, I.M.-V.; writing, K.O.-V., K.A.C.-L., V.E.S.-T., J.C.R.-C., V.V.-R. and S.J.-M.; writing—review and editing, V.V.-R. and S.J.-M.; supervision, V.V.-R.; funding acquisition, S.J.-M.

**Funding:** This work was supported by a Basic Science grant from SEP-CONACyT México (243233), a FOSISS grant (272633) and the Instituto Nacional de Pediatría, SSA (027/2015 and 060/2016).

**Conflicts of Interest:** The authors declare that they have no competing interests.

## References

1. Perez-Saldivar, M.L.; Fajardo-Gutierrez, A.; Bernaldez-Rios, R.; Martinez-Avalos, A.; Medina-Sanson, A.; Espinosa-Hernandez, L.; Flores-Chapa Jde, D.; Amador-Sanchez, R.; Penaloza-Gonzalez, J.G.; Alvarez-Rodriguez, F.J.; et al. Childhood acute leukemias are frequent in Mexico city: Descriptive epidemiology. *BMC Cancer* **2011**, *11*, 355. [CrossRef]
2. Pritchard-Jones, K.; Pieters, R.; Reaman, G.H.; Hjorth, L.; Downie, P.; Calaminus, G.; Naafs-Wilstra, M.C.; Steliarova-Foucher, E. Sustaining innovation and improvement in the treatment of childhood cancer: Lessons from high-income countries. *Lancet Oncol.* **2013**, *14*, e95–e103. [CrossRef]

3. Hunger, S.P. Expanding clinical trial networks in pediatric acute lymphoblastic leukemia. *J. Clin. Oncol.* **2014**, *32*, 169–170. [[CrossRef](#)]
4. Tsuchida, M.; Ohara, A.; Manabe, A.; Kumagai, M.; Shimada, H.; Kikuchi, A.; Mori, T.; Saito, M.; Akiyama, M.; Fukushima, T.; et al. Long-term results of tokyo children's cancer study group trials for childhood acute lymphoblastic leukemia, 1984–1999. *Leukemia* **2010**, *24*, 383–396. [[CrossRef](#)]
5. Magrath, I.; Steliarova-Foucher, E.; Epelman, S.; Ribeiro, R.C.; Harif, M.; Li, C.K.; Kebudi, R.; Macfarlane, S.D.; Howard, S.C. Paediatric cancer in low-income and middle-income countries. *Lancet Oncol.* **2013**, *14*, e104–e116. [[CrossRef](#)]
6. Moricke, A.; Reiter, A.; Zimmermann, M.; Gadner, H.; Stanulla, M.; Dordelmann, M.; Loning, L.; Beier, R.; Ludwig, W.D.; Ratei, R.; et al. Risk-adjusted therapy of acute lymphoblastic leukemia can decrease treatment burden and improve survival: Treatment results of 2169 unselected pediatric and adolescent patients enrolled in the trial all-bfm 95. *Blood* **2008**, *111*, 4477–4489. [[CrossRef](#)]
7. Flohr, T.; Schrauder, A.; Cazzaniga, G.; Panzer-Grumayer, R.; van der Velden, V.; Fischer, S.; Stanulla, M.; Basso, G.; Niggli, F.K.; Schafer, B.W.; et al. Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. *Leukemia* **2008**, *22*, 771–782. [[CrossRef](#)] [[PubMed](#)]
8. Schrappe, M.; Valsecchi, M.G.; Bartram, C.R.; Schrauder, A.; Panzer-Grumayer, R.; Moricke, A.; Parasole, R.; Zimmermann, M.; Dworzak, M.; Buldini, B.; et al. Late mrd response determines relapse risk overall and in subsets of childhood T-cell all: Results of the AIEOP-BFM-ALL 2000 study. *Blood* **2011**, *118*, 2077–2084. [[CrossRef](#)] [[PubMed](#)]
9. Moorman, A.V.; Ensor, H.M.; Richards, S.M.; Chilton, L.; Schwab, C.; Kinsey, S.E.; Vora, A.; Mitchell, C.D.; Harrison, C.J. Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: Results from the UK medical research council ALL97/99 randomised trial. *Lancet Oncol.* **2010**, *11*, 429–438. [[CrossRef](#)]
10. Gupta, S.K.; Bakhshi, S.; Chopra, A.; Kamal, V.K. Molecular genetic profile in BCR-ABL1 negative pediatric B-cell acute lymphoblastic leukemia can further refine outcome prediction in addition to that by end-induction minimal residual disease detection. *Leuk. Lymphoma* **2018**, *59*, 1899–1904. [[CrossRef](#)] [[PubMed](#)]
11. Zerkalenskova, E.; Panfyorova, A.; Kazakova, A.; Baryshev, P.; Shelihova, L.; Kalinina, I.; Novichkova, G.; Maschan, M.; Maschan, A.; Olshanskaya, Y. Molecular characteristic of acute leukemias with t(16;21)/FUS-ERG. *Ann. Hematol.* **2018**, *97*, 977–988. [[CrossRef](#)] [[PubMed](#)]
12. Vijayakrishnan, J.; Studd, J.; Broderick, P.; Kinnersley, B.; Holroyd, A.; Law, P.J.; Kumar, R.; Allan, J.M.; Harrison, C.J.; Moorman, A.V.; et al. Genome-wide association study identifies susceptibility loci for B-cell childhood acute lymphoblastic leukemia. *Nat. Commun.* **2018**, *9*, 1340. [[CrossRef](#)] [[PubMed](#)]
13. Hong, Y.; Zhao, X.; Qin, Y.; Zhou, S.; Chang, Y.; Wang, Y.; Zhang, X.; Xu, L.; Huang, X. The prognostic role of E2A-PBX1 expression detected by real-time quantitative reverse transcriptase polymerase chain reaction (RQ-PCR) in b cell acute lymphoblastic leukemia after allogeneic hematopoietic stem cell transplantation. *Ann. Hematol.* **2018**, *97*, 1547–1554. [[CrossRef](#)] [[PubMed](#)]
14. Kremer, L.S.; Bader, D.M.; Mertes, C.; Kopajtich, R.; Pichler, G.; Iuso, A.; Haack, T.B.; Graf, E.; Schwarzmayr, T.; Terrile, C.; et al. Genetic diagnosis of mendelian disorders via RNA sequencing. *Nat. Commun.* **2017**, *8*, 15824. [[CrossRef](#)]
15. Qian, Z.; Liu, H.; Li, M.; Shi, J.; Li, N.; Zhang, Y.; Zhang, X.; Lv, J.; Xie, X.; Bai, Y.; et al. Potential diagnostic power of blood circular RNA expression in active pulmonary tuberculosis. *EBioMedicine* **2018**, *27*, 18–26. [[CrossRef](#)]
16. Chen, L.; Lu, D.; Sun, K.; Xu, Y.; Hu, P.; Li, X.; Xu, F. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene* **2019**, *692*, 119–125. [[CrossRef](#)]
17. Taylor, S.C.; Laperriere, G.; Germain, H. Droplet digital PCR versus qPCR for gene expression analysis with low abundant targets: From variable nonsense to publication quality data. *Sci. Rep.* **2017**, *7*, 2409. [[CrossRef](#)] [[PubMed](#)]
18. Hindson, B.J.; Ness, K.D.; Masquelier, D.A.; Belgrader, P.; Heredia, N.J.; Makarewicz, A.J.; Bright, I.J.; Lucero, M.Y.; Hiddessen, A.L.; Legler, T.C.; et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **2011**, *83*, 8604–8610. [[CrossRef](#)] [[PubMed](#)]

19. Zhao, M.Y.; Yu, Y.; Xie, M.; Yang, M.H.; Zhu, S.; Yang, L.C.; Kang, R.; Tang, D.L.; Zhao, L.L.; Cao, L.Z. Digital gene expression profiling analysis of childhood acute lymphoblastic leukemia. *Mol. Med. Rep.* **2016**, *13*, 4321–4328. [[CrossRef](#)] [[PubMed](#)]
20. Sakhinia, E.; Estiar, M.A.; Andalib, S.; Rezamand, A. Expression profiling of microarray gene signatures in acute and chronic myeloid leukaemia in human bone marrow. *Iran. J. Pediatr. Hematol. Oncol.* **2015**, *5*, 27–42.
21. Zhao, G.; Jiang, T.; Liu, Y.; Huai, G.; Lan, C.; Li, G.; Jia, G.; Wang, K.; Yang, M. Droplet digital PCR-based circulating microRNA detection serve as a promising diagnostic method for gastric cancer. *BMC Cancer* **2018**, *18*, 676. [[CrossRef](#)] [[PubMed](#)]
22. Villegas-Ruiz, V.; Moreno, J.; Jacome-Lopez, K.; Zentella-Dehesa, A.; Juarez-Mendez, S. Quality control usage in high-density microarrays reveals differential gene expression profiles in ovarian cancer. *Asian Pac. J. Cancer Prev.* **2016**, *17*, 2519–2525. [[PubMed](#)]
23. Villegas-Ruiz, V.; Juarez-Mendez, S. Data mining for identification of molecular targets in ovarian cancer. *Asian Pac. J. Cancer Prev.* **2016**, *17*, 1691–1699. [[CrossRef](#)] [[PubMed](#)]
24. Castillo-Rodriguez, R.A.; Davila-Borja, V.M.; Juarez-Mendez, S. Data mining of pediatric medulloblastoma microarray expression reveals a novel potential subdivision of the group 4 molecular subgroup. *Oncol. Lett.* **2018**, *15*, 6241–6250.
25. Wee, Y.; Liu, Y.; Bhyan, S.B.; Lu, J.; Zhao, M. The pan-cancer analysis of gain-of-functional mutations to identify the common oncogenic signatures in multiple cancers. *Gene* **2019**, *697*, 57–66. [[CrossRef](#)] [[PubMed](#)]
26. Kahles, A.; Lehmann, K.V.; Toussaint, N.C.; Huser, M.; Stark, S.G.; Sachsenberg, T.; Stegle, O.; Kohlbacher, O.; Sander, C.; The Cancer Genome Atlas Research Network; et al. Comprehensive analysis of alternative splicing across tumors from 8705 patients. *Cancer Cell* **2018**, *34*, 211–224. [[CrossRef](#)]
27. Bustin, S.A.; Beaulieu, J.F.; Huggett, J.; Jaggi, R.; Kibenge, F.S.; Olsvik, P.A.; Penning, L.C.; Toegel, S. Miqe precis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Mol. Biol.* **2010**, *11*, 74. [[CrossRef](#)]
28. Bustin, S.A. Quantification of mrna using real-time reverse transcription PCR (RT-PCR): Trends and problems. *J. Mol. Endocrinol.* **2002**, *29*, 23–39. [[CrossRef](#)]
29. Bustin, S.A.; Nolan, T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* **2004**, *15*, 155–166.
30. Huggett, J.F.; Foy, C.A.; Benes, V.; Emslie, K.; Garson, J.A.; Haynes, R.; Hellemans, J.; Kubista, M.; Mueller, R.D.; Nolan, T.; et al. The digital MIQE guidelines: Minimum information for publication of quantitative digital PCR experiments. *Clin. Chem.* **2013**, *59*, 892–902. [[CrossRef](#)]
31. Sanders, R.; Huggett, J.F.; Bushell, C.A.; Cowen, S.; Scott, D.J.; Foy, C.A. Evaluation of digital PCR for absolute DNA quantification. *Anal. Chem.* **2011**, *83*, 6474–6484. [[CrossRef](#)]
32. Sanders, R.; Mason, D.J.; Foy, C.A.; Huggett, J.F. Evaluation of digital PCR for absolute RNA quantification. *PLoS ONE* **2013**, *8*, e75296. [[CrossRef](#)]
33. Hindson, C.M.; Chevillet, J.R.; Briggs, H.A.; Gallichotte, E.N.; Ruf, I.K.; Hindson, B.J.; Vessella, R.L.; Tewari, M. Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods* **2013**, *10*, 1003–1005. [[CrossRef](#)]
34. Dube, S.; Qin, J.; Ramakrishnan, R. Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. *PLoS ONE* **2008**, *3*, e2876. [[CrossRef](#)] [[PubMed](#)]
35. Oxnard, G.R.; Paweletz, C.P.; Kuang, Y.; Mach, S.L.; O’Connell, A.; Messineo, M.M.; Luke, J.J.; Butaney, M.; Kirschmeier, P.; Jackman, D.M.; et al. Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. *Clin. Cancer Res.* **2014**, *20*, 1698–1705. [[CrossRef](#)]
36. Mencia-Trinchant, N.; Hu, Y.; Alas, M.A.; Ali, F.; Wouters, B.J.; Lee, S.; Ritchie, E.K.; Desai, P.; Guzman, M.L.; Roboz, G.J.; et al. Minimal residual disease monitoring of acute myeloid leukemia by massively multiplex digital PCR in patients with *npm1* mutations. *J. Mol. Diagn.* **2017**, *19*, 537–548. [[CrossRef](#)]
37. Stahlberg, A.; Kubista, M.; Pfaffl, M. Comparison of reverse transcriptases in gene expression analysis. *Clin. Chem.* **2004**, *50*, 1678–1680. [[CrossRef](#)] [[PubMed](#)]

38. Demeke, T.; Eng, M. Effect of endogenous reference genes on digital PCR assessment of genetically engineered canola events. *Biomol. Detect. Quantif.* **2018**, *15*, 24–29. [[CrossRef](#)] [[PubMed](#)]
39. Valente, V.; Teixeira, S.A.; Neder, L.; Okamoto, O.K.; Oba-Shinjo, S.M.; Marie, S.K.; Scrideli, C.A.; Paco-Larson, M.L.; Carlotti, C.G., Jr. Selection of suitable housekeeping genes for expression analysis in glioblastoma using quantitative RT-PCR. *Ann. Neurosci.* **2014**, *21*, 62–63. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).