

Received: 2016.01.18
Accepted: 2016.04.15
Published: 2016.04.28

A Simple Method for Optimization of Reference Gene Identification and Normalization in DNA Microarray Analysis

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDEF **Federico M. Casares**

Department of Bioinformatics, MitoGenetics Research Institute, Farmingdale, NY, U.S.A.

Corresponding Author: Federico M. Casares, e-mail: federico.casares@mitogenetics.com
Source of support: Departmental sources

Background: Comparative DNA microarray analyses typically yield very large gene expression data sets that reflect complex patterns of change. Despite the wealth of information that is obtained, the identification of stable reference genes is required for normalization of disease- or drug-induced changes across tested groups. This is a prerequisite in quantitative real-time reverse transcription-PCR (qRT-PCR) and relative RT-PCR but rare in gene microarray analysis. The goal of the present study was to outline a simple method for identification of reliable reference genes derived from DNA microarray data sets by comparative statistical analysis of software-generated and manually calculated candidate genes.


Material/Methods: DNA microarray data sets derived from whole-blood samples obtained from 14 Zucker diabetic fatty (ZDF) rats (7 lean and 7 diabetic obese) were used for the method development. This involved the use of software-generated filtering parameters to accomplish the desired signal-to-noise ratios, 75th percentile signal manual normalizations, and the selection of reference genes as endogenous controls for target gene expression normalization.

Results: The combination of software-generated and manual normalization methods yielded a group of 5 stably expressed, suitable endogenous control genes which can be used in further target gene expression determinations in whole blood of ZDF rats.

Conclusions: This method can be used to correct for potentially false results and aid in the selection of suitable endogenous control genes. It is especially useful when aimed to aid the software in cases of borderline results, where the expression and/or the fold change values are just beyond the pre-established set of acceptable parameters.

MeSH Keywords: **Data Interpretation, Statistical • Diabetes Mellitus, Type 2 • Microarray Analysis**

Full-text PDF: <http://www.basic.medscimonit.com/abstract/index/idArt/897644>

 2467

 6

 1

 19



Background

The generation of very large amounts of gene microarray data poses a challenge, not only for its processing but also for its interpretation, due to intrinsic false discovery rates. In addition, the problem of background noise along with differences in hybridization efficiencies is also an important factor generating variability within and among microarray chips, constituting major confounding elements in gene expression analysis. Some of the most advanced commercially available software can automatically account for most, but not all, of these challenges.

In general, the persistence of confounding elements generates the need for appropriate data normalization methods, such as using the specific n^{th} percentile signal intensity value of a particular array. Often, software-generated normalization methods have already incorporated the n^{th} -percentile approach (e.g., 75th percentile) along with some background-subtraction mechanism [1]. In this regard, dealing with the assay's inherent background noise becomes critical to account for signal stringency. Hence, the importance of using filtering parameters to accomplish the desired signal-to-noise ratios becomes obvious.

Moreover, it becomes necessary to use housekeeping or reference genes as endogenous controls for further gene signal normalization. This is not a common use in gene microarray analysis, where log transformation, background subtraction, and n^{th} percentile normalizations have been the norm [1]. In this regard, the use of endogenous control genes is prerequisite in qRT-PCR and relative RT-PCR [2–4]. The principle behind this methodology consists of simply using widely expressed genes that do not respond to most treatments as references to compare to genes of interest (target genes) that do change. This helps in the proper interpretation of gene expression patterns and in calculating relative gene expression fold changes between treatment groups, minimizing technique-derived experimental errors. Thus, the same rationale should apply in gene microarray analysis. However, selecting the right endogenous control genes for normalizing data can be difficult since these widely expressed reference genes are not truly universal. In this regard, there have been observed changes in reference gene expression with different treatments as well as tissue-specific differential reference gene expression patterns [2–10]. For this reason, a systematic method must be determined for its use in the selection of array-specific (i.e., tissue- and taxa-specific) endogenous control genes based on a pool or pools of pre-established and widely used housekeeping or reference genes.

In the present examination, dependent measure data sets were derived from paired DNA microarray gene expression analyses performed on whole-blood samples from homozygous ZDF rats exhibiting clinically-relevant type 2 symptomatology

in comparison to heterozygous healthy lean controls [11]. In this regard, the ZDF rat has been well established in the biomedical literature as a high-resolution translational model for elucidation of underlying pathophysiological mechanisms critically linked to advanced therapeutic development for major human disorders, including type 2 diabetes [12,13], cardiovascular disease [14], renal disease [15], atherosclerosis [16–18], and rheumatoid arthritis [11]. In addition, a list of potentially suitable endogenous control genes for the study of whole-blood ZDF rat samples is provided.

Material and Methods

The analytical software used in this examination was Agilent's GeneSpring GX, ver.13.1.1. Manual calculations were performed using the Microsoft Excel basic package. The gene microarray data (Agilent single-color expression) was obtained from a published study [11] using whole-blood samples collected from 7 twelve-week-old male homozygous (Fa/Fa) leptin receptor-deficient ZDF rats exhibiting a full-fledged type 2 diabetic phenotype highlighted by hyperglycemia, hyperlipidemia, liver hypertrophy, increased water consumption, and urine output, and from 7 twelve-week-old male heterozygous (Fa/fa) healthy lean controls. Briefly, the study animals were housed 2 per cage and maintained in an Innovive caging system (San Diego, CA). The rooms were lit for 12 hours from 7:00 AM to 7:00 PM, each day, using artificial light. Animals had free access to water and Purina 5008 rodent food (Waldschmidt's, Madison, WI) for the duration of the study (7 weeks) [11]. The study was approved by the Institutional Animal Care and Use Committee (IUCAC, Study Number SNY1301). Animal care and all technical procedures were performed by PhysioGenix, Inc. staff in accordance with the established protocols in the *National Institute of Health Guide for Care and Use of Laboratory Animals (Eighth Edition)*.

The initial data were processed using Agilent's feature extraction software, followed by analysis using the microarray platform software GeneSpring and by enhancement through a manual optimization method, as follows:

First, using the microarray software, the following filtering parameters were implemented: a filter by flags (e.g., "detected", "not detected", and "compromised") where irregular features (or signals) were discarded, and a signal-to-noise ratio of 2, which was chosen as the lower limit in at least 1 of the subject groups. In addition, a list of 34 annotated reference genes previously identified in the biochemical literature [2–8,19] was built and used to filter the experimental data.

Second, the signal intensities of the reference genes passing the above-mentioned filters were manually divided by the 75th

percentile value of the corresponding arrays and the resulting values were used to calculate fold changes in gene expression between the healthy lean and the diabetic obese groups by simple division (i.e., diabetic obese value/healthy lean value). It should be noted that a gene variation was deemed biologically irrelevant when its fold change value was defined as $-1.2 < x < 1.2$.

Statistical analyses

Software-generated data was compared using moderated *t* test method with Benjamini-Hochberg multiple testing correction. The *t* test was used for evaluating the manually normalized data.

Results

Application of filtering parameters

The first step before data analysis deals with signal quality control and the setting of filtering parameters in a particular data set. In this evaluation, the filters previously described were used to identify potential endogenous control gene candidates from the list of 34 widely used reference genes [2–8,19] (Table 1). After this initial filtering process, in which genes not meeting the signal quality criteria were filtered out (i.e., flagged as “compromised”; $S/N < 2$), a working list of 18 gene probes corresponding to 16 endogenous gene candidates was made (Table 2). It should be noted that there can be more than 1 probe per gene, each having a different sequence, thus hybridizing to a different region of the gene transcript.

Endogenous control gene determination

Software-generated gene selection

Suitable endogenous control genes should exhibit minimal-to-no expression variation between groups (e.g., control vs. treatment), in this particular study, between diabetic obese and healthy lean groups. In this evaluation, an absolute fold change value of 1.2 was set as the limit for the gene selection criterion. In this way, further filtering by fold change yielded 3 suitable endogenous control gene candidates, which can be seen in Figure 1. In addition, Table 3 shows these software-selected genes (Hsp90ab1, Pum1, and Srsf4) along with their fold change and *p*-values.

Manual normalization

The manual method involving 75th percentile normalization of background-subtracted signals along with fold change calculations yielded 5 potentially suitable endogenous control

candidates. Three were the same as the software-generated genes, plus 2 additional genes – Dimt1 and Gusb (Table 4). These genes exhibited fold change values < 1.2 and > -1.2 , with *p*-values considered not significant ($p > 0.05$). In this regard, after manual normalization, there was an additional gene, Decr1, that exhibited an acceptable fold change value of 1.19 but had a *p*-value < 0.05 and hence was not selected (Table 2). Table 5 shows an arbitrary gene grouping based on signal intensity values (low, medium, high). This helps in the selection of suitable endogenous control genes because, as mentioned earlier, these genes should ideally be chosen so that they encompass a large signal intensity spectrum in such a way that it compensates for the potentially diverse copy numbers of target genes (translated as signal intensities) [3]. Finally, Table 6 shows the Agilent probe sequences of each of the endogenous control genes selected by this method.

Discussion

Filtering parameters to accomplish the desired signal-to-noise ratios

Gene microarray data need to be adequately filtered. The first step before data analysis involved a quality control step in which irregular signals or “compromised” features are removed (i.e., filter by flags: detected; not detected; compromised). Often, in order to accomplish an acceptable microarray signal intensity level, a signal should be at least twice as strong as that of the background (i.e., signal-to-noise ratio ≥ 2) and, depending on the desired stringency level, this filter cut-off can be set to a signal-to-noise ratio of 3 or higher. Normally, gene microarray technologies produce a consistent background signal whose mean level information can be easily obtained from the raw image data (e.g., using feature extraction software). A microarray platform’s software automatically subtracts calculated background signal from raw signal values, effectively yielding processed raw signal values. To filter out genes whose processed raw signal values are less than twice the background ($S/N < 2$), a filter (i.e., processed raw signal cut-off) should be set to a lower limit equivalent to the array’s mean background signal value. In this way, if the processed (i.e., background-subtracted) raw signal is added to the technology’s mean background signal value, it will be equivalent to $S/N=2$. Processed data falling below the $S/N=2$ level were eliminated from the final data set. Moreover, this filtering by expression level was applied so it would accept a gene when in at least 1 of the subject groups studied (e.g., control; treatment “A”; and treatment “B”) is detectable since, for example, a given treatment/s could cause downregulation of a gene below a level corresponding to $S/N=2$. The same applies when a gene is only detectable after a treatment. In this way, when a particular gene or group of genes was present at $S/N \geq 2$ in at least 1 of the subject groups, then those genes passed the filter.

Table 1. List of commonly used reference genes.

Gene symbol	Gene name	Chromosome
A4galt	alpha 1,4-galactosyltransferase	chr7
Actb	Actin, beta	chr12
B2m	beta-2 microglobulin	chr3
Cck	Cholecystokinin	chr8
Cry2	Cryptochrome 2 (photolyase-like)	chr3
Csnk1g2	Casein kinase 1, gamma 2	chr7
Decr1	2,4-dienoyl CoA reductase 1, mitochondrial	chr5
Dimt1	DIM1 dimethyladenosine transferase 1 homolog (<i>S. cerevisiae</i>)	chr2
Eef1a1	Eukaryotic translation elongation factor 1 alpha 1	chr8
Farp1	FERM, RhoGEF (Arhgef) and pleckstrin domain protein 1 (chondrocyte-derived)	chr15
Fpgs	Folylpolyglutamate synthase	chr3
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase	chr4
Gins2	GINS complex subunit 2 (Psf2 homolog)	chr19
Gusb	Glucuronidase, beta	chr12
Hmbs	Hydroxymethylbilane synthase	chr8
Hprt1	Hypoxanthine phosphoribosyltransferase 1	chrX
Hsp90ab1	Heat shock protein 90 alpha (cytosolic), class B member 1	chr9
Mapre2	Microtubule-associated protein, RP/EB family, member 2	chr18
Pex16	Peroxisomal biogenesis factor 16	chr3
Pgk1	Phosphoglycerate kinase 1	chrX
Polr2a	Polymerase (RNA) II (DNA directed) polypeptide A	chr10
Ppia	Peptidylprolyl isomerase A (cyclophilin A)	chr14
Ppib	Peptidylprolyl isomerase B	chr8
Pum1	Pumilio RNA-binding family member 1	chr5
Rpl4	Ribosomal protein L4	chr8
Rplp2	Ribosomal protein, large P2	chr1
Sdha	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	chr1
Srsf4	Serine/arginine-rich splicing factor 4	chr5
Tbp	TATA box binding protein	chr1
Tfrc	Transferrin receptor	chr11
Trap1	TNF receptor-associated protein 1	chr10
Ubc	Ubiquitin C	chr12
Ywhag	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma	chr12
Ywhaz	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta	chr7

List of 34 widely used housekeeping/reference genes screened for endogenous control gene selection.

Table 2. Filtered reference gene candidates for endogenous control gene selection.

Probe name	Gene symbol	Gene name	Fold change	p-value	Chromosome
A_64_P050964	Cry2	Cryptochrome 2 (photolyase-like)	-1.46	7.40×10 ⁻³	chr3
A_42_P526030	Decr1	2,4-dienoyl CoA reductase 1, mitochondrial	-1.32	1.91×10 ⁻²	chr5
A_44_P524471	Dimt1	DIM1 dimethyladenosine transferase 1 homolog (<i>S. cerevisiae</i>)	-1.26	1.60×10 ⁻¹	chr2
A_64_P232432	Gapdh	Glyceraldehyde-3-phosphate dehydrogenase	-2.83	2.20×10 ⁻³	chr4
A_64_P052510	Gapdh	Glyceraldehyde-3-phosphate dehydrogenase	-2.88	3.40×10 ⁻³	chr4
A_64_P073003	Gusb	Glucuronidase, beta	-1.34	1.90×10 ⁻³	chr12
A_44_P421363	Hmbs	Hydroxymethylbilane synthase	-2.38	3.00×10 ⁻⁴	chr8
A_43_P11257	Hprt1	Hypoxanthine phosphoribosyltransferase 1	-2.30	7.00×10 ⁻⁴	chrX
A_64_P045716	Hsp90ab1	Heat shock protein 90 alpha (cytosolic), class B member 1	1.08	1.99×10 ⁻²	chr9
A_64_P140020	Hsp90ab1	Heat shock protein 90 alpha (cytosolic), class B member 1	-1.60	8.30×10 ⁻³	chr9
A_64_P047724	Mapre2	Microtubule-associated protein, RP/EB family, member 2	-1.85	3.50×10 ⁻²	chr18
A_42_P492082	Pex16	Peroxisomal biogenesis factor 16	2.31	7.40×10 ⁻³	chr3
A_64_P058353	Ppia	Peptidylprolyl isomerase A (cyclophilin A)	-1.80	5.30×10 ⁻³	chr14
A_43_P13976	Ppib	Peptidylprolyl isomerase B	-1.71	3.00×10 ⁻⁴	chr8
A_42_P767897	Pum1	Pumilio RNA-binding family member 1	-1.11	1.28×10 ⁻²	chr5
A_64_P080678	Sdha	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	-1.67	3.60×10 ⁻³	chr1
A_42_P816010	Srsf4	Serine/arginine-rich splicing factor 4	-1.08	3.96×10 ⁻¹	chr5
A_44_P416641	Ywhaz	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta	-2.71	9.00×10 ⁻⁴	chr7

List of 18 gene probes resulting from software-generated filtering of 34 widely used housekeeping/reference genes. The fold change and *p*-values represent the variation between diabetic obese and healthy lean subjects. Statistics: Moderated *t*-test method with Benjamini-Hochberg Multiple Testing Correction.

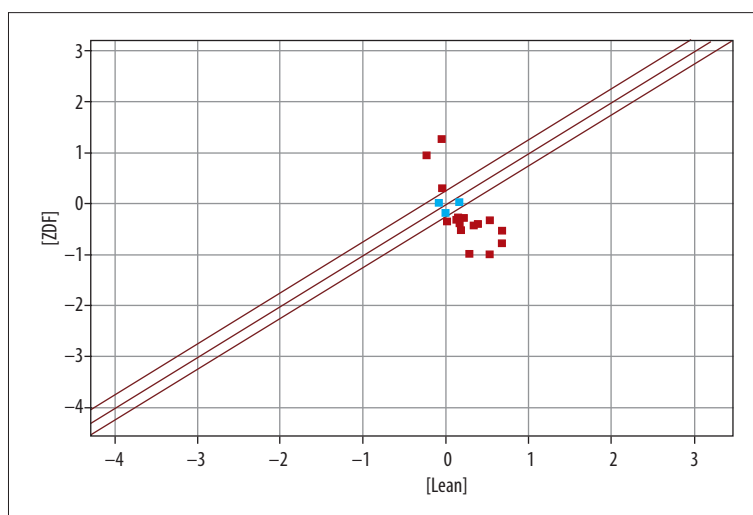


Figure 1. Scatter Plot Visualization of Software-generated Endogenous Control Gene Candidates. Scatter plot comparison between healthy (Lean) and diabetic obese (ZDF) rats. Outer lines delimit fold change values ≤ 1.2 and ≥ -1.2 . Highlighted in blue are those genes with the least modulation differences between the two groups.

Table 3. Software-generated endogenous control gene candidates.

Probe name	Gene symbol	Gene name	Fold change	p-value	Chromosome
A_64_P045716	Hsp90ab1	Heat shock protein 90 alpha (cytosolic), class B member 1	1.08	1.99×10^{-2}	chr9
A_42_P816010	Srsf4	Serine/arginine-rich splicing factor 4	-1.08	3.96×10^{-1}	chr5
A_42_P767897	Pum1	Pumilio RNA-binding family member 1	-1.11	1.28×10^{-2}	chr5

The use of Agilent's GeneSpring GX 13.1.1 software yielded 3 housekeeping/reference genes as suitable candidates for endogenous control genes. The fold change and p-values represent the variation between diabetic obese and healthy lean subjects. Statistics: Moderated *t*-test method with Benjamini-Hochberg multiple testing correction.

Table 4. Endogenous control genes selected after manual normalization.

Probe name	Gene symbol	Gene name	Software generated		Manually normalized		Chromosome
			Fold change	p-value	Fold change	p-value	
A_44_P524471	Dimt1	DIM1 dimethyladenosine transferase 1 homolog (<i>S. cerevisiae</i>)	-1.26	1.60×10^{-1}	-1.13	3.46×10^{-1}	chr2
A_64_P073003	Gusb	Glucuronidase, beta	-1.34	1.90×10^{-3}	-1.19	9.82×10^{-2}	chr12
A_64_P045716	Hsp90ab1	Heat shock protein 90 alpha (cytosolic), class B member 1	1.08	1.99×10^{-2}	1.16	2.19×10^{-1}	chr9
A_42_P767897	Pum1	Pumilio RNA-binding family member 1	-1.11	1.28×10^{-2}	-1.01	9.58×10^{-1}	chr5
A_42_P816010	Srsf4	Serine/arginine-rich splicing factor 4	-1.08	3.96×10^{-1}	1.02	9.28×10^{-1}	chr5

Manual normalization yielded 2 additional genes, Dimt1 and Gusb, to the original list of 3 software-generated endogenous control candidates. Statistics: Student *t*-test was used for evaluating the manually normalized data.

Endogenous control gene selection and normalization

As mentioned earlier, selecting the right endogenous control genes for data normalization can be difficult due to gene expression changing with treatments or to tissue-specific differential gene expression patterns [4,9,10]. For this purpose, based on several important publicly available studies [2–8,19], a list of 34 widely used reference genes was built to be evaluated with the experimental data.

Ideally, it is preferable to select more than 1 endogenous control gene and to average their values. In this regard, it is recommended that, when possible, the endogenous control genes be chosen so that they will exhibit different signal intensity levels (e.g., low, medium, and high) [3]. This would account for the differences in copy number (i.e., signal intensities) among the target genes. Hence, the signal intensity levels of the potential endogenous control gene candidates were compared to be sure they spanned a relatively wide range. Moreover, suitable endogenous control genes should be selected such

that each is involved in a different cellular function and/or is found in different chromosomes [5,6]. Although this may not always be possible, it is recommended that at least 2 of these criteria be satisfied (Table 5).

Finally, in order to overcome or minimize inter-array differences, scaling to the n^{th} percentile is recommended (in this case, to the 75th percentile) [1]. If using a linear scale (i.e., not log-normalized), as in this case, the processed (background-subtracted) signal intensity values are divided by the 75th percentile value corresponding to the particular array. This scaling is applied to both potentially suitable endogenous control genes and target genes.

Normalization of target genes

After finding and normalizing suitable endogenous control genes, the next step is to use them to normalize genes of interest to calculate their expression pattern through fold change values. In this regard, the problem with software-generated

Table 5. Gene expression levels by signal intensity value.

Probe name	Gene symbol	Fold change	p-value	Manually normalized signal	Chromosome
A_64_P232432	Gapdh	-2.42	3.35×10 ⁻⁴	14.85	chr4
A_64_P052510	Gapdh	-2.46	9.48×10 ⁻⁴	14.48	chr4
A_64_P047724	Mapre2	-1.61	1.34×10 ⁻²	11.41	chr18
A_64_P058353	Ppia	-1.65	6.42×10 ⁻³	6.53	chr14
A_44_P416641	Ywhaz	-2.38	4.88×10 ⁻⁵	6.29	chr7
A_43_P13976	Ppib	-1.53	5.73×10 ⁻⁴	6.11	chr8
A_64_P045716	Hsp90ab1	1.16	2.19×10⁻¹	6.07	chr9
A_44_P421363	Hmbs	-1.99	7.29×10 ⁻³	4.18	chr8
A_43_P11257	Hprt1	-2.08	9.15×10 ⁻⁵	3.44	chrX
A_42_P816010	Srsf4	1.02	9.28×10⁻¹	2.65	chr5
A_64_P140020	Hsp90ab1	-1.44	5.42×10 ⁻⁴	2.58	chr9
A_64_P080678	Sdha	-1.49	4.16×10 ⁻⁴	2.42	chr1
A_42_P526030	Decr1	-1.19	2.77×10 ⁻²	1.26	chr5
A_42_P492082	Pex16	3.49	1.24×10 ⁻¹	0.4	chr3
A_64_P073003	Gusb	-1.19	9.82×10⁻²	0.33	chr12
A_64_P050964	Cry2	-1.27	6.54×10 ⁻²	0.28	chr3
A_44_P524471	Dimt1	-1.13	3.46×10⁻¹	0.27	chr2
A_42_P767897	Pum1	-1.01	9.58×10⁻¹	0.22	chr5

Selected endogenous control candidate genes shown in bold font. Signal Intensity: 10–15 = high □; 1–9.99 = med ■; 0–0.99 = low ▣. Arbitrary signal intensity-based separation of mean background-subtracted signals normalized by their corresponding microarray's 75th-percentile value. Note that the genes selected (in bold) show fold changes between 1.2 and -1.2, with p-values greater than 0.05, indicating that changes in expression were not statistically significant. Also note that although gene Decr1 exhibits a fold change below 1.2, its p-value is lower than 0.05 and therefore this gene was not selected. Statistics: Student t-test was used for evaluating the manually normalized data.

Table 6. Gene probe sequences.

Probe name	Gene symbol	Sequence
A_44_P524471	Dimt1	CAGAAGATTCAGTATAGCCGATAAAAATACAGCAGATCCTAACCAACACAGGTTTTAGTG
A_64_P073003	Gusb	AGAGGTTACGGTTCAGTGCCGAGGACCCAGTGTATGGGAAGCAGACCGTTTACATTCTAA
A_64_P045716	Hsp90ab1	TCTCATGAAGGAGACACAGAAGTCCATCTACTATATCACTGGTGAGAGCAAAGAGCAGGT
A_42_P767897	Pum1	AAGTACACCTATGGCAAGCACATCCTGGCCAAGCTTGAGAAGTACTACATGAAGAATGGT
A_42_P816010	Srsf4	CTTGTAATAGCACAGTCAAGAGAAATGGATACCTGCATAGCCCATAGGAAGTAACACTG

This table shows the gene probe sequences for *Rattus norvegicus*, corresponding to Agilent's microarray technology.

microarray gene signal intensities becomes more evident at the time of their fold change determination. This challenge is not only observed with the calculated fold changes, but also with the corresponding p-values, which may not be statistically

significant (e.g., >0.05) and hence, relevant genes may be filtered out. However, with the normalization method described above, along with the utilization of reliable endogenous control genes, this can be corrected.

One important step taken before the analysis of gene expression is to restrict the search to a specific gene list or lists pertaining to a more focused field of interest (e.g., a particular disease-related list of genes). This helps in the manageability of the data set by restricting it to a much lower number of genes. The next step will be to filter the data according to the filtering parameters depicted in the previous section. That is, filtering by flags, leaving out those having compromised signals, and then filtering by expression, leaving out genes whose processed (background-subtracted) raw signal values are less than twice the background ($S/N < 2$). Again, this is achieved by setting the processed signal's lower limit to the equivalent of the technology's mean background signal value. Once this selected group of genes is filtered, the next step is to manually scale the gene's processed raw signals to the 75th percentile, as described above. The resulting target gene values are then normalized by simple division using the combined value (i.e., mean) of the endogenous control or reference genes selected earlier, as follows: target gene value/endogenous control mean value, for each target gene. In this way, the values obtained can be used to compare control and treatment groups through fold change calculations (e.g., treated vs. control group) and the calculated p -values used to evaluate their statistical significance.

Conclusions

The ZDF rat is a proven model for the study of different comorbidities associated with type 2 diabetes. The results obtained

References:

- Reimers M: Making informed choices about microarray data analysis. *PLoS Comput Biol*, 2010; 6: e1000786
- Andersen CL, Jensen JL, Orntoft TF: Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res*, 2004; 64: 5245–50
- Lee S, Jo M, Lee J et al: Identification of novel universal housekeeping genes by statistical analysis of microarray data. *J Biochem Mol Biol*, 2007; 40: 226–31
- Vandesompele J, De Preter K, Pattyn F et al: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*, 2002; 3: RESEARCH0034
- Mane VP, Heuer MA, Hillyer P et al: Systematic method for determining an ideal housekeeping gene for real-time PCR analysis. *J Biomol Tech*, 2008; 19: 342–47
- Stamova BS, Apperson M, Walker WL et al: Identification and validation of suitable endogenous reference genes for gene expression studies in human peripheral blood. *BMC Med Genomics*, 2009; 2: 49
- Dheda K, Huggett JF, Bustin SA et al: Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques*, 2004; 37: 112–14, 116, 118–19
- Bar M, Bar D, Lehmann B: Selection and validation of candidate housekeeping genes for studies of human keratinocytes – review and recommendations. *J Invest Dermatol*, 2009; 129: 535–37
- Suzuki T, Higgins PJ, Crawford DR: Control selection for RNA quantitation. *Biotechniques*, 2000; 29: 332–37
- Thellin O, Zorzi W, Lakaye B et al: Housekeeping genes as internal standards: use and limits. *J Biotechnol*, 1999; 75: 291–95
- Kream RM, Mantione KJ, Casares FM, Stefano GB: Concerted dysregulation of 5 major classes of blood leukocyte genes in diabetic ZDF rats: A working translational profile of comorbid rheumatoid arthritis progression. *International Journal of Prevention and Treatment*, 2014; 3: 17–25
- Kakimoto T, Kimata H, Iwasaki S et al: Automated recognition and quantification of pancreatic islets in Zucker diabetic fatty rats treated with exendin-4. *J Endocrinol*, 2013; 216: 13–20
- Wang F, Guo X, Shen X et al: Vascular dysfunction associated with type 2 diabetes and Alzheimer's disease: A potential etiological linkage. *Med Sci Monit Basic Res*, 2014; 20: 118–29
- Carley AN, Severson DL: Fatty acid metabolism is enhanced in type 2 diabetic hearts. *Biochim Biophys Acta*, 2005; 1734: 112–26
- Zanchi C, Locatelli M, Benigni A et al: Renal expression of FGF23 in progressive renal disease of diabetes and the effect of ACE inhibitor. *PLoS One*, 2013; 8: e70775
- Mierzecki A, Kloda K, Bukowska H et al: Association between low-dose folic acid supplementation and blood lipids concentrations in male and female subjects with atherosclerosis risk factors. *Med Sci Monit*, 2013; 19: 733–39
- Stohr R, Federici M: Insulin resistance and atherosclerosis: Convergence between metabolic pathways and inflammatory nodes. *Biochem J*, 2013; 454: 1–11
- Kream RM, Mantione KJ, Casares FM, Stefano GB: Impaired expression of ATP-binding cassette transporter genes in diabetic ZDF rat blood. *International Journal of Diabetes Research*, 2014; 3: 49–55
- Wang T, Liang ZA, Sandford AJ et al: Selection of suitable housekeeping genes for real-time quantitative PCR in CD4(+) lymphocytes from asthmatics with or without depression. *PLoS One*, 2012; 7: e48367