

Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



www.rsc.org/molecularbiosystems

Q-GDEMAR: A GENERAL METHOD FOR THE IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES IN MICROARRAYS WITH UNBALANCED GROUPS

Daniel V. Guebel,^{a,b} Montserrat Perera-Alberto,^{a,c} Néstor V. Torres^{a*}

^aSystems Biology and Mathematical Modelling Group. Science Faculty. University of La Laguna. Tenerife. Canary Islands. Spain.

^bBiotechnology Counselling Services. Buenos Aires. Argentina.

^cDepartment of Anatomy, Pathology, Histology and Physiology. University of La Laguna. Tenerife. Canary Islands. Spain.

* (NVT) ntorres@ull.edu.es

Short title: Gaussian-deconvolution of microarrays

Abstract

Microarray analysis is a powerful tool to simultaneously determine the pattern of transcription of large amounts of genes. For data post-processing distinct computational methods are currently used that, however, lead to different results regarding the genes differentially expressed. A new methodology for microarray data analysis named Q-GDEMAR is herein presented. It combines the quantile characterization of the entire distribution together with the Gaussian deconvolution of the central region of the microarray data distribution. Three discriminant variable variants are proposed that allow summarizing data and comparing groups even when their size is strongly unbalanced. In addition, a simple procedure to compute the false discovery rate (FDR) is also presented. The performance of the method is compared with that observed when using LIMMA (Linear Models Microarray) software as reference. In 58 out of 68 cases, Q-GDEMAR showed a higher sensitivity than LIMMA to detect differentially expressed genes ($p=1 \times 10^{-10}$). The proposed method does not produce biased information, detecting genes with high sensitivity equally well in both tails of the distribution ($p=0.7428$). Moreover, all detected genes were associated to very low levels of FDR (median value=0.67%, interquartile range=0.87%). Q-GDEMAR can be used as general method for microarray analysis, but is particularly indicated when the conditions to be compared are unbalanced. The superior performance of Q-GDEMAR is the consequence of its higher discriminative power and, the fact that it yields a univocal correspondence between the p-values and the values of the discriminating variable. Q-GDEMAR was tested only with Affymetrix microarrays. However, given that it operates after the step of data standardization, it can be used with the same quality features on any of the available mono- or dual-channel microarray platforms.

1. Introduction

cDNA- and oligonucleotide-based microarrays are routinely applied in the study of complex diseases in several biomedical fields, ranging from neurobiology and cancer to connective tissue or infectious pathologies.¹⁻³ There are multiple protocols and software packages for the data analysis of microarrays.⁴⁻¹⁰ However, the meta-analysis of different microarray studies related to a same pathology usually yields very different results, no matter if the studies were carried out with the same or different analytical platforms,¹¹⁻¹⁵ thus limiting the translation of the microarray's findings to the clinical practice.^{16,17}

Although microarray discrepancies can be partially attributed to some unavoidable biological and analytical variability, the observed differences can also be the consequence of operational factors during the post-processing steps, once the data have been standardized as we discuss below.

Microarray data are routinely subjected to logarithmic transformation in order to "normalize" the distribution of such data. Normalization should increase the Gaussian character of the distribution (a feature required for t-Student, ANOVAs, and estimation of Bayesian probabilities). Hence, skewness and/or kurtosis should be minimized while the problem of variances heterocedasticity alleviated.¹⁸ However, the efficiency of the logarithmic transformation in achieving an improved Gaussian character is not always guaranteed.¹⁹ An alternative to avoid these problems is the use of non-parametric methods, but these methods are far less powerful than the parametric ones when the size of the samples is not big enough.²⁰

One of the main objectives pursued through the microarray studies is to identify the "differentially expressed genes", i.e., those genes whose level of expression changes along the different conditions tested (e.g., age, sex, disease stage, treatments). The differentially expressed genes are recognized on the basis of their associated "statistical effects", which usually requires the comparison of the "treatment groups" against a "control group" in terms of their mean values.

In order to discriminate the statistical effects of the tested conditions upon the genes, there are two forms to summarize the means' comparisons: by the "Fold-Change (FC)-Difference"²¹ or by the $\log_2(\text{FC})$.²² Importantly, one or other form lead to identify different sets of differentially expressed genes.²³ In addition, neither of these forms account for the differences in the sizes of the groups. Optimal experimental designs require of balanced groups. Otherwise, the means of the groups will have different reliability. Nevertheless, due to practical limitations the balance condition frequently cannot be fulfilled. In these cases, the detection of differentially expressed genes poses additional problems.²⁴

The statistical effects are assessed by ranking the values of the summarizing variable selected according to the significance of two magnitudes: the **p-value** ($p = \text{probability}(\text{error type } I \leq \alpha) = \text{probability}(\text{false positive})$),²⁵ and the false discovery rate (**FDR**). This latter is defined as the "expected" value of the ratio

between the number of false positives and the total number of significant values declared. From the Bayesian perspective, and according to the type of assumptions considered, there are several forms to estimate the expected value for the ratio that defines FDR.²⁶⁻²⁹ In any case, while the p-values give the significance of the statistical effect of a gene as if its expression were determined independently, the FDR protects against false inferences due to the problem of multiple comparisons involved when thousands of genes are being determined simultaneously.²⁹

Another important question driving our search for an alternative procedure to compute the FDR values, is our verification that current algorithms, such as Empirical-Bayes²⁶ and Benjamini-Hochberg²⁷, actually do not provide a univocal correspondence between the p-values (or FDR) computed and the discriminating variable as should be expected (Figure 1).

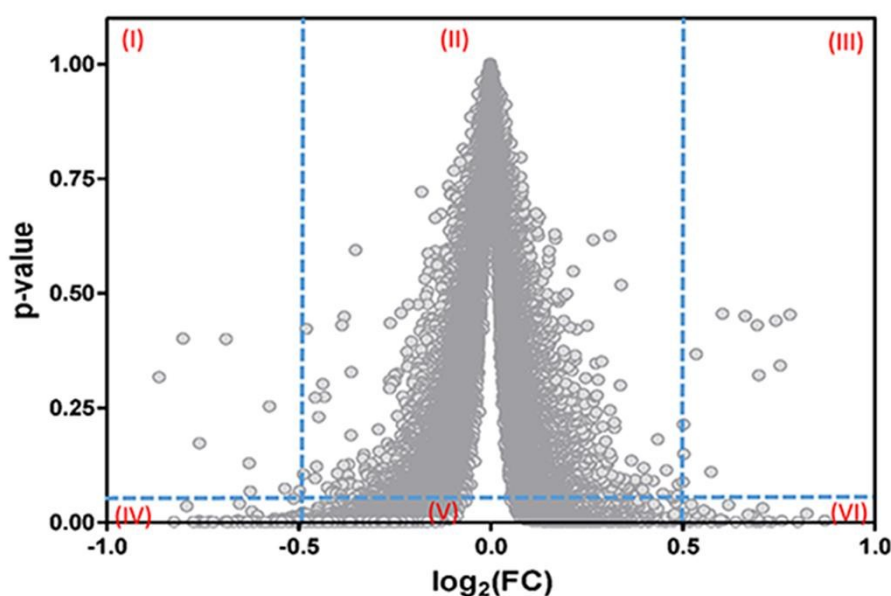


Figure 1. Failure of the current algorithms implemented in LIMMA to establish a univocal correspondence between the magnitudes of fold-change (FC) and its associated probability level of significance (p-value). This analysis was performed on data from GSE48060³⁰ but a similar result is obtained with data from other eleven microarrays. A similar profile of dispersion is also observed for the FDR values of this and other microarrays (results not shown). Roman numerals in the graph indicate the presence of six regions. The three upper regions differ in their significance with respect to the corresponding homologous at the bottom. Current algorithms as Empirical Bayes and Benjamini-Hochberg, lead to a situation in which the same value of FC admits multiple p-values (or FDR). Some of them can be interpreted as significant while others are interpreted as non-significant.

A practical consequence of the absence of a univocal relation between the p-values (or FDR values) and the fold-change (FC) in Figure 1, is that this situation will contribute to some degree of misclassification of the genes in the microarray data. In fact, by taking two vertical lines at two feasible cut-off values of the discriminant variable at the x-axis (e.g., $\log_2(\text{FC}) = \pm 0.5$), and one horizontal line

at a feasible limit of significance (e.g., p -value=0.05), six regions are generated. The lack of unicity can be seen by comparing any of the juxtaposed regions (I and IV, II and V, III and VI) because a same value of FC appears associated with very low and very high p -values in all these cases.

Hence, while regions I and III are associated to potentially false negative cases, the region V is associated to potentially false positive results. Note that due to the problem of multiple simultaneous comparisons, the horizontal line should actually be taken at values quite lower than $p=0.05$. This last situation will improve identifying the significant points located in the new, diminished regions IV-V-VI, but at the cost of worsening the reliability of the significance of the points located in the new, expanded regions I and III.

From Figure 1 it can be concluded that of the six regions generated, only the regions named as IV and VI are suitable for the detection of the differentially expressed genes. This explains why the current methods need to apply a double filtering criterion (FC and p -value) instead of a unique one, filtering either for FC or for p .^{31,32}

Importantly, the absence of a univocal correspondence between the p -values and fold-change values can explain the dissociation between the biological and the statistical criteria observed with the interpretation of some microarrays.^{12,33} It is worth to note that if the $\log(1/p)$ were considered instead of the p -values at the y-axis, Figure 1 will turn into the well-known “volcano” plot.^{31,32}

Likewise, when current algorithmic methods was tested for the reliability of the FDR values generated, we observed that sometimes none of the thousands of analysed genes resulted acceptable according to their FDR values, even when many of the same genes were associated to highly significant p -values. Moreover, in other instances the current algorithmic methods yield very significant p -values ($p=10^{-8}$ to $p=10^{-12}$) even when the empirical distribution of the experimental data cannot produce values beyond the order of $p=10^{-5}$ to $p=10^{-6}$.

Again, the excessive dispersion of the p -values for a given value of fold-change in Figure 1 indicate that these values don't act as asymptotic limits to the p -values obtained empirically from the microarray samples as can be expected. This is a shortcoming of the method because the computation of the FDR by current algorithms depends strongly on the accuracy of the individual p -values previously obtained.^{33,34}

In view of the methodological limitations previously commented, in the present paper we introduce a novel alternative for the post-processing analysis of the microarray data. At the same time, we introduce an alternative method to compute FDR. Both methods are oriented to optimize the identification of the differentially regulated genes. Finally, we will assess the performance of the proposed methodology by comparing it with LIMMA, a program included in the Bioconductor Project (<https://www.bioconductor.org>).

LIMMA is a particularly reliable reference for this comparison. First, it is a classic, almost standard method that is routinely updated.^{35,36} Also, because to make the comparison we have to make combined use of its underlying algorithms (Empirical Bayes and Benjamini-Hochberg), which with slight variants are common to other programs.⁸ Finally, the core of LIMMA services can be additionally accessed by friendly interfaces, such the cases of the portal

Babelomics^{9,37} and GEO2R (NIH, USA),³⁸ both facilitating the operation of this program to any user.

In the Material and Method section the rationale of the Q-GDEMAR is presented as well as the logic of the flux of operations implied in our algorithms and the definition of the discriminant variables used. In the Results and Discussion, we examine the efficiency of the current procedure of microarray normalization; we assess the overall performance of Q-GDEMAR when compared with LIMMA and characterize the relative efficiency of the discriminant variables used. Finally, we discuss the differences of Q-GDEMAR with respect to Cauchy Distribution. Note that detailed protocols for computing the p-values and FDR by our alternative approaches, as well as three tables containing data of interest for the analysis performed are provided as Electronic Supplementary Materials (ESI).

Materials and Methods

2.1 Compared microarrays. The following twelve sets of microarray data were downloaded from the NCBI Gene Expression Omnibus database (GEO): GSE48060,³⁰ GSE54992,³⁹ GSE5713,⁴⁰ GSE5281,⁴¹ GSE1919,⁴² GSE36297,⁴³ GSE48754,⁴⁴ GSE28619,⁴⁵ GSE34308,⁴⁶ GSE1297,⁴⁷ GSE11882,⁴⁶ GSE46922,⁴⁸ where GSE indicates the code of accession to the GEO Data-Sets (<http://www.ncbi.nlm.nih.gov/gds>). As proof of concept, the data-sets were processed in parallel following the proposed method and the LIMMA,³⁵ being this latter processed through the web interface GEO2R provided by NIH (USA, <http://www.ncbi.nlm.nih.gov/geo/info/geo2r.html>). It is worth mentioning that our study was designed to address a heterogeneous group of biomedical conditions. In addition, the data-sets cover an ample range of unbalanced situations regarding the number of control and treatment samples. These two experimental groups were being subjected to three discriminant variables in Q-GDEMAR (see Table S1, ESI), as well as tested with the standard method LIMMA. With the aim of gaining statistical power in the assessment of the methods, we decided to restrict the comparison of their performance to unique contrasts between the two experimental groups in each microarray, even when more complex analysis could be also possible in some cases.

2.2 Q-GDEMAR: Quantile characterization and Gaussian DEconvolution of the central region of the MicroARray data distribution

Microarrays can determine the expression of up to fifty five thousand genes simultaneously. Each gene is monitored by specifically designed “probe-sets”, comprising several synthetic probes targeted to different segments of the gene sequence (Microarray Probing Mapping, available in: http://www.ensembl.org/info/genome/microarray_probe_set_mapping.html).

The experimental evidence supports that most of the monitored genes usually do not vary their expression along the tested conditions, or if they do, only moderate stochastic fluctuations around a well-identifiable mean value are observed. In addition, two minor groups of genes, those that are up- and down-regulated can also be recognized. Hence, the problem of microarray analysis seems to lie on

the deconvolution of three different populations of genes. However, given that the entire distribution of the data can be known, it is only necessary to identify precisely the genes whose expression does not vary in order to spot the occurrence of deregulated genes at the tails of the curve.

2.2.1 Parametric and Non-Parametric Combined Approach

A key feature of our method is that the central region of the microarray data distribution usually shows a Gaussian behaviour even when the entire distribution does not. So, it is always possible to determine a linear central region of the distribution by plotting the cumulated probability (following a “normal” probability scale) against the values of the discriminating variable chosen to represent the experimental data (see Figure 2.A). There are also two regions that depart from the central straight-line; the deviation from the Gaussian model is caused by the occurrence of deregulated genes. We will take advantage of these two opposite features (see Figure 2.B).

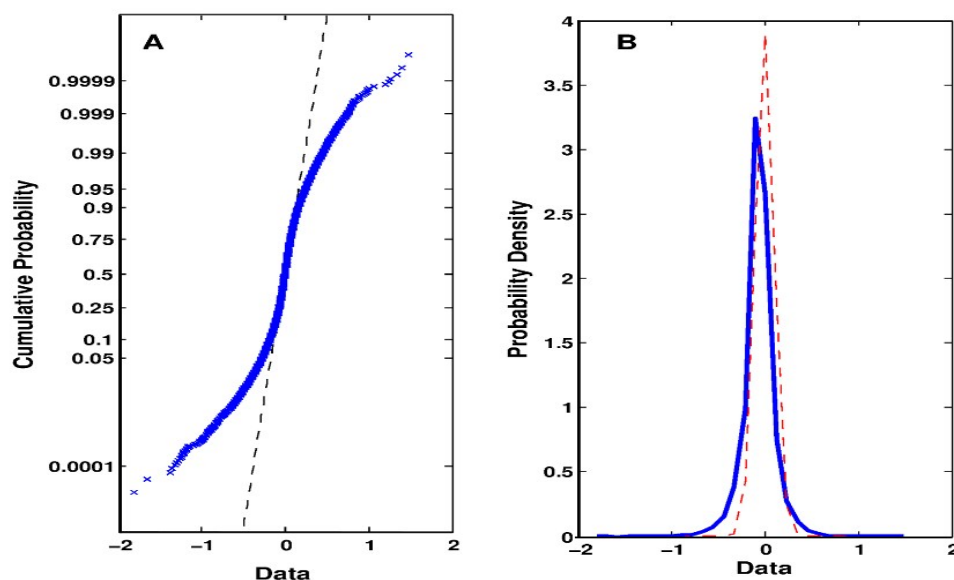


Figure 2. Evidences and principles that support the Q-GDEMAR method. **A.** The intersection between the experimental data (*crossed, blue markers*) and the theoretical straight-line expected for a Gaussian distribution (*dashed, black line*) shows that the Gaussian model fits well in the central region of the data (i.e., the region of genes whose expression does not change significantly). **B.** Consequently, it is possible to establish the variation in the probability profile of the entire set of experimental data (*solid, blue curve*) with respect to the Gaussian model fitted to the central region (*dashed, red curve*). Herein, “Data” refers to the expression of each gene as summarized by the difference $\Delta = \text{median}(\log_2(\text{Treatment Group})) - \text{median}(\log_2(\text{Control Group}))$. Original data were taken from GSE48060³⁰ but the profiles depicted are also representative of the rest of the microarrays analysed.

Note that in the case presented above, the “entire” data-set does not match a true Gaussian model (Figure 2.A) despite its bell-shaped distribution (Figure 2.B). However, the central region of the distribution complies with a Gaussian model in both plots (Figure 2.A and 2.B). Importantly, the narrow dispersion of the

Gaussian central region observed in Figure 2.B allows us to establish well-defined limits to the normal stochastic fluctuation for genes whose expression does not change significantly. This deconvolution process results in a gain of sensitivity in the determination of the differentially expressed genes (significant genes are given by the difference of areas between the blue and the red curves at the tails of the entire distribution in Figure 2.B). At the same time, the risk of accepting a high number of false positive results is greatly diminished (note the reduced area under the tails of the red curve of Gaussian central region in Figure 2.B). Hence, the value of FDR will result drastically reduced.

2.2.2 Q-GDEMAR Operations

The logic behind the Q-GDEMAR method is shown in Figure 3. As Q-GDEMAR operates down-stream of the standardized data, a pre-processing step has to be applied when raw data of the microarray are used. This can be accomplished by the usual routines using free software available (e.g., RMA-Express, GC-RMA or MAS5, among others).^{18, 50-52} Note that data downloaded from the GEO database is already standardized.

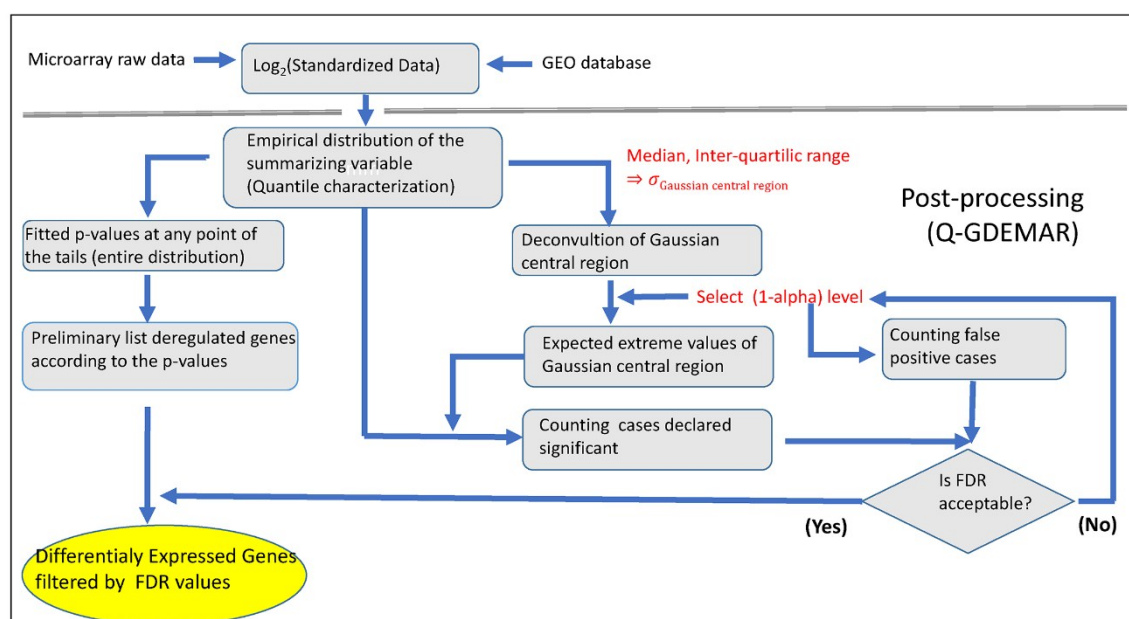


Figure 3. Flux diagram of the Q-GDEMAR method. See Appendices 1-2 (ESI) for technical details.

In microarray analysis, the variations between treatment conditions are determined by computing the “statistical effects”, i.e., the differences (or ratios) between the means of the groups compared. Means summarize the distribution of the experimental data in a unique, “representative” value but it is not the only possible parameter of centrality. In Q-GDEMAR the comparisons are

implemented by using three types of discriminating variables: one based on the difference between the means but corrected by their weighted average, and other two which are direct variants of median-based comparisons.

Note that if non-balanced groups are compared in terms of means, each mean has a different degree of reliability and hence, some form of weighted average should be applied. In this regard, current methods have limitations.²² Instead, as used in the present method and shown in Figure 2, the control group, with 6 replicates, and the treatment group, with 9 replicates, were compared on the basis of the difference of their $\log_2(\text{medians})$ along $n=43742$ genes without considering the actual number of replicates. Further details about the three types of discriminating variable used to deal with unbalanced groups are presented in Section 2.3.

In order to deal with the determination of the “significance” of the computed differences, in our method the deregulated genes are assessed by two complementary techniques. The first technique relies on the empirical modelling of the entire data distribution; the second is based on the identification of genes showing expression values beyond the extreme limits expected for the central Gaussian population of data.

The first approach is suitable to compute the significance of individual effects for each gene analysed. In this case, p-values will be computed based on the determination of some critical quantiles from the entire data distribution (see Appendix 1, ESI). In doing so, the individual p-values are obtained without the need of assuming the Gaussian character of the overall data distribution. The second approach is suitable to compute the false discovery rate (FDR) (see Appendix 2, ESI).

In contrast with current algorithms, such as that used by LIMMA, our method does not compute FDR for individual genes but for the set of differentially expressed genes on each tail of the overall distribution. Note that working together, the features of both combined approaches lead to a situation in which the identified groups of up- and down-regulated genes acquire different values of FDR and comprise different number of genes.

2.3 Discriminating Variables: In order to account directly or indirectly for the possible unbalance in size among the compared groups, we assayed three variants of discriminating variables. These variants are the following:

2.3.1 Differences of means considering the size of the groups

To alleviate the detrimental effect of unbalanced groups, we define a discriminating variable named as $\Delta 1$ -Difference. It allows to summarize the contrasts as is shown in equation 1, where the sub-indices X and C denote the “treatment condition” and the “control condition” respectively.

$$\Delta 1 \text{ Difference} = \frac{1}{WA} (MEAN_X - MEAN_C) \quad (1)$$

Note that $MEAN_X$ and $MEAN_C$ refer to means computed on the previously \log_2 -transformed data (equation 2 and equation 3), while **WA** is a correction factor computed as weighted average between the means of transformed data, being n_i the number of samples included in each of the compared groups i (eq.4).

$$MEAN_X = \text{mean}(\log_2(\text{standardized values}_X))$$

(2)

$$MEAN_C = \text{mean}(\log_2(\text{standardized values}_C))$$

(3)

$$WA = \frac{\sum_i n_i MEAN_i}{\sum_i n_i}$$

(4)

2.3.2 Differences of medians without considering the size of the groups

In contrast with the mean value, median is a “robust” statistical parameter. It can deal with the problem of the excessive within-group heterogeneity, being less sensitive to outliers. Moreover, the median provides reliable estimation of the centrality value of any distribution, particularly when significant skewness and /or kurtosis are present. On these basis, we define the discriminating variable named $\Delta 2$ -Difference as follows:

$$\Delta 2 \text{ Difference} = \text{median}(\log_2(\text{standardized values}_X)) \\ - \text{median}(\log_2(\text{standardized values}_C))$$

(5)

2.3.3 Ratio of medians without considering the size of the groups

Under the same considerations of robustness and variability of the medians that applied in the case of $\Delta 2$ -Difference, another discriminating variable **Ratio 2** can be defined as:

$$\text{Median Ratio} = \frac{\text{median}(\log_2(\text{standardized values}_X))}{\text{median}(\log_2(\text{standardized values}_C))} \quad (6)$$

In spite that the proposed median-based discriminatory variables only indirectly consider the fact that median behaves itself as a random variable, this does not result detrimental (see Section 3.5). In this median-based approach, the effects of the main comparisons (i.e., treatment vs. control) are “confounded” with the possible effects associated to the variation of the median with the size of the samples. We have not knowledge about a law governing this type of variation, such neither about some equivalent to Central Limit theorem, accounting for the distribution of replicated means, but for the case of the medians since it could compensate the variation of the median value with the size of the sample.

Since median is a more robust parameter than the mean, it is expected that the median values performance will be better that the obtained by the mean values. Particularly in the case of microarrays, the impact on median variation with the size of the sample might be lower that the impact due to other sources of

variation present, such as the biological and technical variability. In fact, classic t-student fails to deal with the low-expression genes, forcing current methods of microarray analysis to use modified forms of the t-statistic (regularized, penalized, or stabilized t-tests).^{8,28,32} From this “adapted” t-statistics, the p-values and the FDR are derived but neither of these forms are free of some strong assumptions.^{8,28} Hence, results obtained from the current algorithms are valid up to a given extent^{8,55} (see Section 4).

In the median-based approach, the implementation of the Gaussian-deconvolution operation imposes practical boundaries to the confounding effect. The analysis carried out on twelve real microarray data-sets, showing that the sensitivity of gene detection and FDR of our median-based approach outperformed the one by LIMMA algorithms (see section 2.3) let us to conclude that variation of the median with the size of the sample is not significant relative to other major sources of variation present in the system.

2. Results and Discussion

3.1 How normalizing are the current normalization procedures?

As was already commented, it is usually assumed that log-transformation is a universal recipe to get “normalized” data. However, an important residual trend between the variance with respect to the mean was verified when analysing the microarray data obtained from GSE54992, being both in \log_2 -transformed units (see Figure 4.A and Figure 4.B).

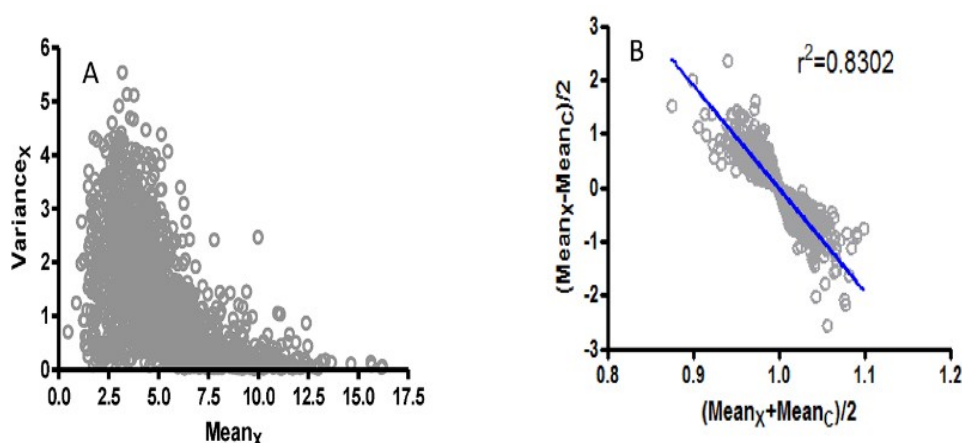


Figure 4. Pattern of dispersion in microarray data taken from GSE54992.³⁹ **A.** The variance is represented as function of the mean value for the group with pathology. **B.** The semi-distance between pathology and control groups is presented as a function of the average of expression in the same groups.

Different degrees of residual trend also persist with the regression of the variance over the mean in the other microarrays analysed herein. For example, the GSE48060 (potential function, $R^2 = 0.362$), GSE46922 (polynomial 5th degree,

$R^2=0.6359$), and GSE11882 (linear trend, $R^2=0.8020$), all of them showed significant patterns of residual trend of data after the log-transformation (data not shown).

In order to avoid this residual trend, we assayed other types of data transformation. By doing so, we observed that sometimes the type of transformation needed depends on the combination of factors analysed (e.g., “old-female” could require a transformation different from that for “old-male”, and different from “young-female” or “young-male”). Therefore, it appears not to be recommendable to apply different data transformation when analysing different factors such as sex and gender simultaneously.

Note that equality of variances is a desirable requisite to be fulfilled when applying Empirical Bayes, which is the main method implemented when using LIMMA and many other programs. However, the presence of residual trends does not constitute a problem in the Q-GDEMAR method because it operates on the empirical distribution of the data, without any further assumption.

3.2 Overall method performance assessment

The biological interpretation of microarrays results is influenced by the type of discriminant variable chosen as well as for the type of computational algorithm applied for the data post-processing. To assess both effects, we tested the three discriminant variants of Q-GDEMAR (Median Ratio, $\Delta 1$ -Difference, $\Delta 2$ -Difference) (Section 2.3), and their performance was compared against LIMMA program.

To that purpose, twelve microarray data-sets (Section 2.1) were evaluated by the number of up- and down regulated genes detected (Section 2.2). We also computed the FDR associated to each microarray analysed by Q-GDEMAR, but using our own approach (Appendix 2, ESI). In the case of LIMMA, the number of genes detected depends on the combined use that this program makes of the Empirical Bayes (for the estimation of the p-value),²⁶ and the FDR (computed by Benjamini-Hochberg approach on the basis of the p-values).²⁷

Actually, there is a unique definition for FDR, independently of the way in which its value is estimated.²⁶⁻²⁹ Thus, the number of genes considered as differentially expressed in LIMMA was taken from the set of genes obtained by this program, but fulfilling the condition that their FDR values were lower or equal than the FDR value determined by Q-GDEMAR. In this way, the comparisons between both methods were done at the same level of FDR (see Figures 5A, 5B, and 5C).

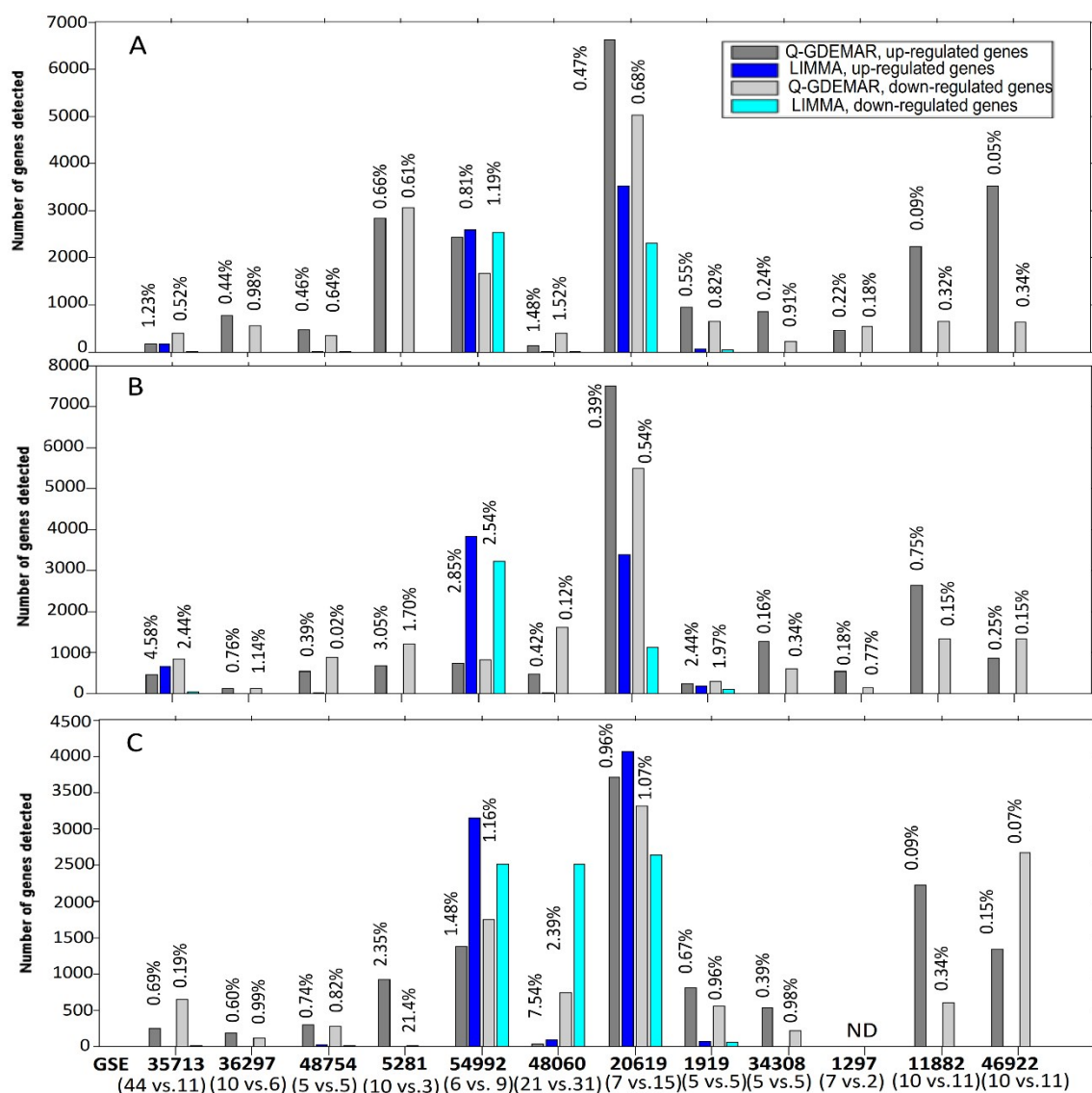


Figure 5: Comparative performance of Q-GDEMAR and LIMMA methods along twelve microarray data sets according to the number of genes detected and the FDR achieved. Q-GDEMAR was evaluated through three discriminating variants: Median Ratio (*top figure, A*), Δ -2 Difference (*middle figure, B*), Δ -1 Difference (*lower figure, C*). The values above the bars indicate the percentage of FDR achieved in each case, at which comparison is done. The values between parentheses indicate the number of control and treatment samples which are being compared in each microarray. For details on the data, see Table S1 (ESI).

Figure 5 shows some relevant features. First, LIMMA shows many gaps in gene-number count. This indicates that LIMMA was able to detect just a few genes (or none at all) at the low level of FDR. This is in strong contrast with the performance showed by Q-GDEMAR. Second, in 58 of the 68 comparisons done at similar level of FDR, Q-GDEMAR detected more genes than LIMMA did (see Table S1, ESI). Finally, different non-parametric statistical tests (Wilcoxon's Rank Test, Sign Test, and Fisher's Exact Test) confirmed the highly significant differences when Q-GDEMAR variants are overall compared against LIMMA method (probability p ranges from 1×10^{-9} to 9.85×10^{-10} depending on the statistical method).

Also important is the finding that algorithms implemented through LIMMA are affected by a bias. In fact, analysis of results from Table S1 shows that LIMMA detect systematically a higher number of up-regulated genes with respect to the number of down-regulated genes (Sign Test, $p=2.74 \times 10^{-4}$). Instead, this bias was not present in Q-GDEMAR even when different number of up- and down-regulated genes were always detected (Sign Test, $p=0.7428$).

Very often, researchers are forced to accept FDR values as high as 10-30% due to the difficulties found with their microarray data. In fact, in three of the twelve microarrays tested (GSE36297, GSE1297, and GSE46922), LIMMA showed a median value of FDR ranging between 17% and 46%, even considering only those genes with extreme expression (absolute t-Student ≥ 4). Instead, with the combined use of Q-GDEMAR and our method to compute FDR we can circumvent this limitation. In fact, we have obtained very low levels of FDR (median value=0.67%, interquartile range=0.87%; see Figures 5A-5C). It is worth noting that our FDR evaluations included the detection of genes in three microarrays (the GSE 46922, GSE5281, and GSE1297) for which the LIMMA algorithms did not detect any significant gene at all (FDR varied between 88.5% and 98.9%, see Table S1).

We thus conclude that Q-GDEMAR shows a superior performance with respect to LIMMA in the sensitivity to detect differentially expressed genes. Moreover, it does at very low levels of false discovery rate (FDR).

3.3 Discriminating variants' performance assessment

Although Q-GDEMAR outperformed LIMMA in 85.3% of the analysed cases, a ranking of efficiencies for the individual variants can be established within each data-set (see Table S1, ESI). The Median Ratio was the best discriminating variable in 50% of the cases analysed, $\Delta 2$ -Difference followed (33.33% cases), and lastly $\Delta 1$ -Difference (16.66% cases). This differential distribution is statistically significant ($\chi^2=12.04$, $p < 5 \times 10^{-3}$), thus confirming that discriminating variants differ in the number of genes detected in a given instance.

We have determined that the distribution curves of the three variants show quite similar profiles, but they do not keep a "strict" relation along their domains. Although each curve evolves monotonically, their mutual relations change according to the number of genes detected (see Figure 6).

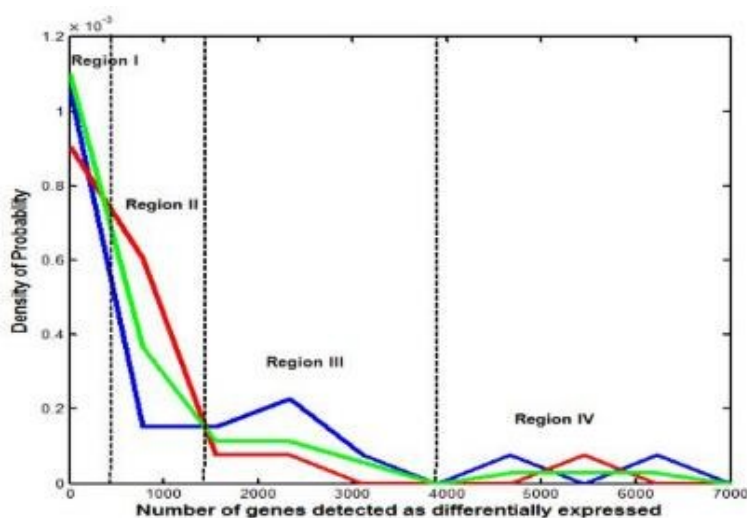


Figure 6. Comparative hierarchical order of the distributions curves corresponding to Q-GDEMAR variants as a function of the number of significant genes detected along the twelve microarrays analysed in Table S1. The discriminant variants compared are the Median Ratio (*blue line*), the $\Delta 2$ -Difference (*red line*) and, the $\Delta 1$ -difference (*green line*). Regions (I to IV) are defined by the zones of the domain where the change in the hierarchical order takes place.

The trends displayed in Figure 6 can explain the data observed in Table S1 (ESI), in the sense that Median Ratio outperforms the $\Delta 2$ - or $\Delta 1$ -Difference in some microarrays, but only within a given range of detected genes. The fact that hierarchical order among the curves in Figure 6 does not change point to point, but along a few, discrete intervals, suggests that the observed variations reflect true differences in the behaviour of the variants rather than mere noisy data .

Differences due to the unbalance in size among the compared groups can also influence the variants performance. This was observed in GSE5281 microarray, where a poor correlation between the average and the median values ($R^2=0.625$) is found in three samples of the treatment group. Nevertheless, the control group of the same microarray, which comprised ten samples, shows an excellent correlation ($R^2=0.999$; data not shown). Hence, the median ratio variant out-performs the $\Delta 1$ -Difference in the case of the GSE5281, because the better variant is based on median values, responding more accurately to the nature of the data (see Table 1S, ESI).

Moreover, “unbalanced sampling” could exert an “interaction effect” rather than a “fixed effect”. That is, the effect produced by size differences between the groups could vary depending on the high or low number of replications involved in a particular comparison. This is the case of the microarrays GSE35713 (44 control samples against 11 treatment samples) and GSE54992 (6 control samples against 9 treatment samples). In both cases, the $\Delta 1$ -Difference out-performed the median-based discriminating variants. This seems justified since in these cases, a high correlation between means and medians within each condition is guaranteed due to the high number of samples considered. In addition, the weight factor used as correction in $\Delta 1$ -Difference seems able to compensate the strongly unbalanced experimental design used. Besides, the number of genes detected in GSE35713 fall into the Region II of Figure 6, providing an additional reason for $\Delta 1$ -Difference variant out- performing the other discriminating variants.

In brief, given that the relative performance of each discriminant variant depends on the number of genes detected in each particular microarray, but the number of genes that will be detected is unknown beforehand, a practical consequence of Figure 6 is that there is no *a priori* preferred form for calculation of a suitable discriminating variable. The three variants have to be tested by Q-GDEMAR, and best selected for a given micro-array.

3.4 What determines the efficiency of the discriminant variants?

In order to quantify the efficiency of the discriminating variables we introduce the index Phi (Φ), defined as $\Phi = [1 - (\sigma_{central} / \sigma_{global})]$, where $\sigma_{central}$ is the standard deviation for the Gaussian central region of the data distribution (computed through the Gaussian deconvolution operation), and σ_{global} is the apparent standard deviation (computed upon the entire data distribution).

The index Φ characterizes how good is a given discriminant variant in separating the largest population of genes with invariant expression from those small groups of genes which are deregulated in a given condition. Note that if Φ index is zero, there is no gained discrimination from the central region (i.e., $\sigma_{Central} = \sigma_{Global}$). On the contrary, an ideal value of 1 would imply that the genes without significant variation would have a unique, constant value of expression (i.e., $\sigma_{Central} = 0$). In this case, the discrimination between the genes belonging to the Gaussian central region and those de-regulated genes localized at the tails of the “entire” data distribution will be maximum. Otherwise, as in general could be expected, the Φ index will take some value comprised between zero and one (i.e., $\sigma_{Central} < \sigma_{Global}$).

Importantly, the “index Φ ” shown to be significantly correlated with the “number of genes detected” by Q-GDEMAR variants (R-Pearson=0.806, $p < 1 \times 10^{-4}$), and this correlation persists even after being corrected by the kurtosis (partial correlation $R_{genes_Phi(kurtosis)} = 0.8418$) or the skewness (partial correlation $R_{genes_Phi(skewness)} = 0.8194$). Because other co-linear relations also occurred, the correlation matrix of system’s variables (see Table S3, ESI) was subjected to principal component analysis (see Figure 7).

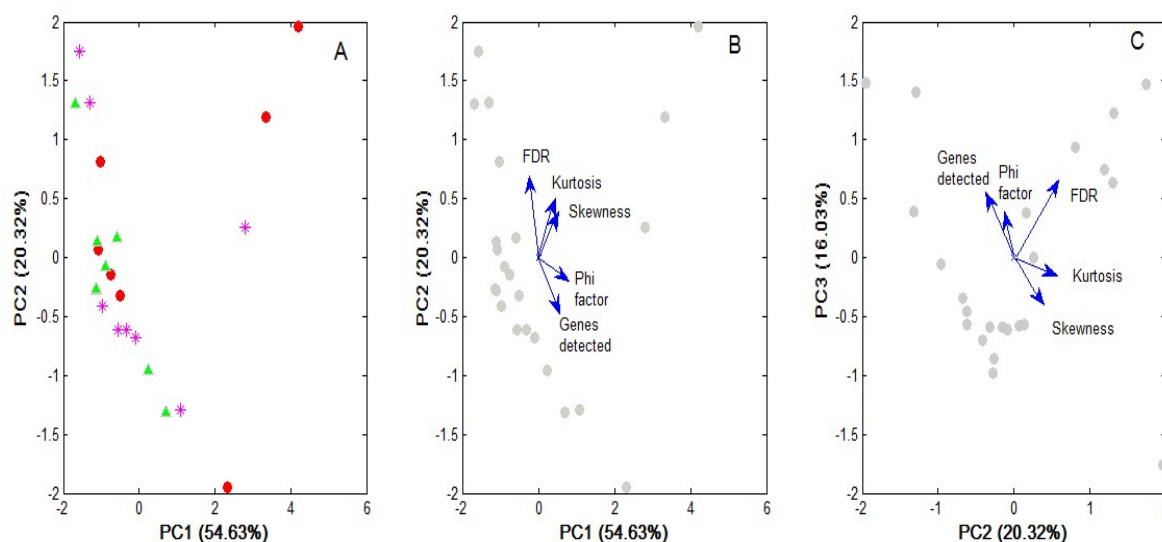


Figure 7: Eigen-decomposition of the correlation matrix of Table S3. The correlation matrix (apparent dimension=5) was projected on two bi-dimensional sub-spaces (PC1-PC2, PC2-PC3). **A.** The scores of the projected points are disaggregated according to the type of discriminant variables used for processing: Median Ratio (●), $\Delta 2$ -Difference (*), $\Delta 1$ -Difference (▲). **B and C.** Biplot representations of the points projected, but without disaggregation of the discriminating variants. The axis are the principal components (PC), and the values between parentheses give the percentage of variance explained by them. The relative position of the analysed variables in each sub-space is given by the eigenvectors of the correlation matrix (blue arrows). The closer the vectors, higher the cosine of the angle comprised, and hence, higher the correlation between the variables that they represent.

Figure 7A shown that no particular pattern of the data is detected among the three discriminating variants used. This is in line with our previous finding concerning the “shape” similarity observed with the distribution profiles associated to each variant (Figure 6). Importantly, data in Figure 7A support the hypothesis that differences among the discriminant variants are not produced by intrinsic differential features, but for the fact that discriminant variants could share a common, still unidentified factor that is present at different levels in a given microarray condition.

Moreover the eigen-analysis performed confirms the trend of “ Φ index” to be positively correlated with the “number of genes detected” (see Figure 7C and Figure 7B). That is, “ Φ index” can be the unifying factor, shared by the three discriminant variants, differing only in the level of efficiency achieved in each particular case.

Because PC1 together with PC2 explain 75% of the data variability, a main trend of “FDR” to negatively correlate with the “number of genes detected” can be established (see Figure 7B). However, a minor trend in which these variables appear in positive correlation also is observed (see Figure 7C). Logically, the index Φ correlated negatively with FDR in the first case (see Figure 7B), but positively in the second case (Figure 7C). Anyway, from a practical point of view, it should be remembered that although a high value of the Φ index could

increase the value of FDR, in our method FDR was usually very low (median value: 0.67%, interquartile range: 0.87%).

Concerning “kurtosis” and “skewness”, there is with both a main trend to worsen the FDR by increasing its value (see Figure 7B), but some minor cases exist in which neither of these variables affected the FDR (see Figure 7C). Moreover, skewness and kurtosis appeared as nearly orthogonal to “ Φ index” (see Figure 7B), which is opposite to the shown in the matrix of correlations where these variables have significant correlations (see Table S.2, ESI).

The apparent discrepancies are explained because correlations in Table S.2 were computed on the raw values of the variables, whereas in Figure 7 the variables are Z-standardized for the principal component analysis given that they come from a correlation matrix. Hence, while in the first case correlations are dominated by those variables with higher variance, in the second case this influence is neutralized.

A near orthogonality was observed between the pair skewness-kurtosis with respect to Φ index. This could mean either a non-linear relation between the “shape” of the initial distribution of the data and the level of gain of discriminating efficiency achieved through the Gaussian deconvolution or that there is no relation at all between these variables. A method based on specific algorithms to deal in complex way with the skewness and the kurtosis has been proposed to address the problem of the (non-linear) interactions between the variables.¹⁰ However, in our approach, these structural features of the entire distribution are not treated separately, but subsumed in the operation of Gaussian deconvolution.

The occurrence of several main trends in the principal component analysis coexisting with other minor cases in which previous relationships appear inverted or relaxed, denote segmentation of the data. By cluster analysis, we verified this segmentation (results not shown). This open the possibility of occurrence of “local” correlations rather than overall correlations. In spite of the data segmentation, we have verified by using the R Pearson correlation that “ Φ index” is strongly related to the “number of genes detected”. Analyses by partial correlation and principal component also confirm this relation. Although correlations do not necessarily imply causality, these findings strongly suggest that the efficiency of Gaussian deconvolution (i.e., the Φ index), effectively conditions the number of genes detected in the microarrays.

3.5 Is Cauchy distribution an alternative to Median Ratio in Q-GDEMAR?

It has been proposed that genes differentially expressed can be determined by models based on the Cauchy distribution,⁵³ a subtype of “alpha-stable” distributions.⁵⁴ Cauchy distribution is generated by the quotient of two normal, independent variables. Its main advantage is that median ratios of the samples converge to the true median ratio of the population while the tails of the distribution are “heavier” than the expected for a normal distribution, being thus closer to the microarrays reality.

Our Median Ratio discriminating variant is also computed as the quotient of two independent variables. However, neither the control nor the treatment conditions followed a normal distribution in the microarray data-sets that we have

analysed. Hence, it is not to be expected that Median Ratio follow a Cauchy distribution.

Moreover, Cauchy distributions are symmetric. Instead, in our Median Ratio distributions only a sub-set of microarrays are free of significant skewness. Interestingly, some particular subtypes of alpha-stable distributions allow some degree of a restricted asymmetry.⁵⁴ However, neither Cauchy nor alpha-stable distributions account for kurtosis, another prominent characteristic frequently observed in the Median Ratio distributions (see Table S2, ESI). Therefore, it should not be expected that Cauchy or alpha-stable distributions can be used as general method for microarray analysis.

As practical example of the consequences generated when the requisites of Cauchy distributions are not fulfilled, we analysed the microarray GSE 36297 with this method. The results are depicted in Figure 8.

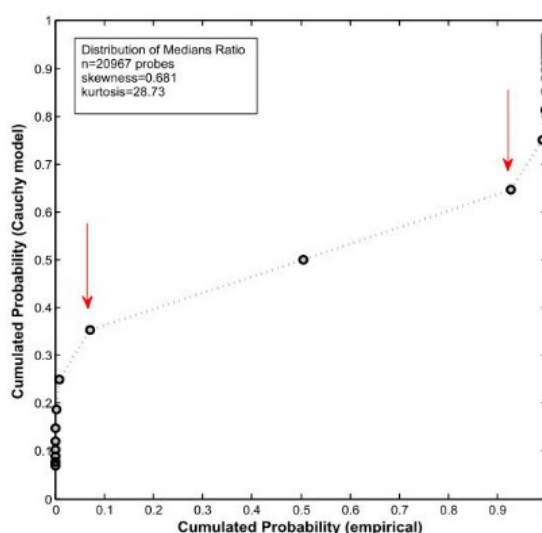


Figure 8: Q-Q plot between the “expected” cumulated probability distribution according to a Cauchy-based model and the “empirical” cumulated probability distribution. Computations were done upon data from microarray GSE36297,⁴³ which result in a Cauchy model with two parameters (location $t=0.9798$ and scale $s=0.0389$).

As can be seen in Figure 8, the fitting by a Cauchy distribution model is very poor. In fact, there is no quantile correspondence at the tails among of the compared distributions. The reason explaining the lack of fitting is the moderate level of kurtosis present in the data. This result is in contrast with the excellent performance obtained with the same microarray by Q-GDEMAR (see Tables S1 and S2, ESI).

It should be noted that neither Q-GDEMAR nor Cauchy-distribution assume that the overall data distribution is Gaussian. Hence, Q-GDEMAR also recognizes the problem of the “heavier tails”. But, Q-GDEMAR deals with “heavier tails” in a less restrictive and more accurate way than Cauchy-based models do.

3. Conclusions

The rigorous analysis of the microarray data is a hard, complex task. Its complexity arises from the convergence of multiple non-linear interacting factors. Some of them are the analytical variability, the biological variability inherent to the sampled subjects, the type of discriminating variable adopted, the absolute and the relative degree of unbalance between the compared groups. In order to fully rationalize these factors and to obtain an exhaustive explanation, a mathematical model where the multiple interrelations are represented is required.

Because such a model is rarely available, other alternatives should be envisaged. Our proposal intends to fill this need. Being aware of its shortcomings and limitations, the proposed method has nonetheless some significant advantages. First, in 58 over 68 cases compared Q-GDEMAR showed a higher sensitivity to detect genes differentially expressed ($p=1 \times 10^{-10}$). Second, it allows computing in a simple straightforward way two-tail FDR values, which are associated to the sub-populations of up- and down regulated genes rather than to each individual gene. Third, in addition to its higher sensitivity, the procedure that we have developed allowed us to achieve very low levels of FDR (median=0.67%, interquartile range=0.87%). Fourth, the method is statistically sound because univocal, well-defined p-values are assigned to the different values of the discriminant variable. Fifth, the outputs of the method, such as p-values and FDR, are independent from any particular theoretical assumption and thus, the risk to be under- or over-estimating the number of genes of interest is minimized.

Importantly, our method achieved high sensitivity for gene detection while maintaining very low levels of FDR, in strong contrast with algorithms commonly applied. For example, simulation studies of Yang and co-workers' using five popular methods for microarray post-processing analysis showed that their power of detection ranges between 10 to 60% when the size of replication varies between 3 to 12 samples; a maximum power of 80% is achieved only if more than 25 samples are used, an experimental condition rarely fulfilled.⁸ On the other hand, simulation studies of Mc Carthy and Smyth practiced on six current methods for post-processing analysis showed that these methods can maintain low FDR values only if the number of differentially expressed genes considered is limited to less than 200 genes. If the list were increased to 400 genes, a FDR value as high as 33% is generated.⁵⁵

There are no rules about what should be considered an "acceptable" value of FDR. A cost-benefit criterion based on practical considerations has to be used. The cut-off value of FDR should be decided taking into account the aim of the analysis, the inherent level of biological variability, the analytical variability of the microarrays, and the cost of the polymerase chain reaction (PCR) validating the analysis. In any case, the FDR values condition the number of genes that can

potentially be accepted as differentially expressed (the lower the FDR achieved, the higher is the number of identifiable deregulated genes). Many published microarray studies report a range from 10 to 500 top dysregulated genes. However, this criterion implicitly links the supposed biological importance exerted by one gene with the magnitude of the fold-change observed with it. However, it is a fact that many genes showing slight changes (~15-20%) can also exert important effects. In addition, the consistency of the microarray findings can also be tested through network analysis and ontology analysis, besides being validated by PCR and/or proteomics. These two complementary procedures, being context-dependent techniques, would require lists of candidate genes as long as possible. Again, the importance of obtaining high sensitivity of detection together with low FDR values, both features of our method, makes our approach very valuable (Section 3.2).

Although Q-GDEMAR was primarily devised to deal with the detection of differentially expressed genes in unbalanced groups, actually it is a general method for the post-processing analysis of c-DNA and oligonucleotide-based microarrays. As such, its application can be extended to other experimental designs not included in the present study. Hence, Q-GDEMAR can deal with paired samples, (1-way) ANOVA, and factorial designs, time-regressions, dosage-regression, or survival (COX) analysis. Importantly, the post-processing analysis of the dual-channel microarrays² can also be addressed by our method.

The common reason that justifies the ample scope of Q-GDEMAR is that in all the referred cases, the relation between treatments and control conditions finally converges in a unique, continuous, unimodal distribution to be analysed.

Q-GDEMAR and our procedure to compute FDR rely on the fact that most of the genes within a microarray, do not vary significantly their expression and follow a Gaussian distribution. This is not a supposition guided by mathematical convenience, but a fact supported by our analysis of the twelve microarray data-sets dealt with in this work (Figure 2.A, Section 2.1). What can change between different microarrays is not the presence of Gaussian central region, but the extent of its dispersion. According to the features of each microarray distribution, the operation of deconvolution provides an estimation of its interquartile range ($IQR=Q_{0.75}-Q_{0.25}$), which is related with the apparent $SD_{\text{Gaussian_central_region}}$ through equation 7 (Appendix 2, ESI). Different $SD_{\text{Gaussian_central_region}}$ values, through the interplay among equations 7-14, lead to different levels of FDR, and hence, to different number of genes detected as differentially expressed (Appendix 2, ESI).

So, either Q-GDEMAR or our procedure to compute FDR can function through the Gaussian deconvolution process, whichever the profile of the overall data distribution of the microarray.

Electronic Supplementary Materials (ESI)

Annex 1: Computation of p-values

Annex 2: Computation of false discovery rate (FDR)

Table S1: Comparison of the performance of the Q-GDEMAR method against

the results obtained by the reference LIMMA

Table S2: Efficiency of the different discriminant variants and shape characterization of the data distributions.

Table S3: Matrix of correlation between some intrinsic characteristics of the data distributions (Skewness and Kurtosis) and several output measurements (Φ index, Total genes detected, \overline{FDR}) computed upon the Q-GDEMAR variants shown in Table S2.

Acknowledgments. This work was funded by research grants from MINECO, Project BIO2014-54411-C2-2-R and the IMBRAIN project, REF. FP7-REGPOT-2012-CT2012-31637-IMBRAIN. DVG acknowledges the support given by the Campus Atlántico Tri-Continental (CEI-10/00018, Canary Islands, Spain). The authors thank Dr. Catalina Feledi for her valuable collaboration.

Authors Contribution. DVG was responsible for the design and conception of the study, MPA was responsible for the data acquisition and computing. All the authors contributed to the drafting of the manuscript, while NVT supervised the project and gave the final approval. All the authors agree on all the parts of the work.

References

1. N. Mah, A. Thelin, T. Lu, S. Nikolaus, T. Kühbacher, Y. Gurbuz, H. Eickhoff, G. Klöppel, H. Lehrach, B. Mellgård, C.M. Costello, S. Schreiber. A comparison of oligonucleotide and cDNA-based microarray systems. *Physiological Genomics*, 2004; 16: 361-370
2. Y. Guo, Q. Sheng, J. Li, F. Ye, D.C. Samuels, Y. Shyr. Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data. *PLoS One*, 2013; 8: e71462
3. K.J. Mantione, R.M. Kream, H. Kuzelova, R. Ptacek, J. Raboch, J. M. Samuel, G.B. Stefano. Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Med Sci Monit Basic Res*. 2014; 20: 138-42 (review)
4. R.C. Gentleman, V.J. Carey, D. M. Bate, B. Bolstadt, M. Detting et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, 2004; 5: R80.
5. G. R. Grant, E. Manduchi, C.J. Stoeckert. Analysis and management of microarray gene expression data. *Curr Protoc Mol Biol*, 2007; doi: 10.1002/0471142727.mb1906s77 (review).
6. S. K. Mohapatra, A. Krishnan. Microarray data analysis. *Methods Mol Biol*, 2011; 678: 27-43 (review).
7. K. Owzar, W.T. Barry, S.H. Jung. Statistical considerations for analysis of microarray experiments. *Clin Transl Sci*, 2011; 4(6): 466-77 (review).
8. D. Yang, R.S. Parrish, G.N. Brock. Empirical evaluation of consistency and accuracy of methods to detect differentially expressed genes based on microarray data. *Comput Biol Med*, 2014; 46:1-10.
9. R. Alonso, F. Salavert, F. Garcia-Garcia, J. Carbonell-Caballero, M. Bleda, L. Garcia-Alonso, et al. Babelomics 5.0: functional interpretation for new generations of genomic data. *Nucleic Acids Res*, 2015; 43: W117-W121.
10. J. Wang, M. Jia, L. Zhu, Z. Yuan, P. Li, C. Chang, J. Luo, M. Liu, T. Shi. Systematical Detection of Significant Genes in Microarray Data by Incorporating Gene Interaction Relationship in Biological Systems. *PLoS One*, 2010; 5(10): e13721.
11. J. P. Ioannidis, D. B. Allison, C. A. Ball, I. Coulibaly, X. Cui, et al. Repeatability of published microarray gene expression analyses. *Nat Genet*, 2009; 41: 149-55.
12. A. Campaign, Y. H. Yang. Comparison study of microarray meta-analysis methods. *Bioinformatics*, 2010; 11: 408-.
13. Y. Tan, Y. Liu. Comparison of methods for identifying differentially expressed genes across multiple conditions from microarray data. *Bioinformatics*, 2011; 7: 400-404.
14. F. Hong, R. Breitling. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics*, 2008; 24:374-82

15. L.C. Chang, H. M. Lin, E. Sibille, G.C. Tseng. Meta-analysis methods for combining multiple expression profiles: comparisons, statistical characterization and an application guideline. *BMC Bioinformatics*, 2013; 4: 368.
16. P. Roepman. The future of diagnostic gene-expression microarrays: bridging the gap between bench and bedside. *Bioanalysis*, 2010; 2: 249-62.
17. M. Guarnaccia, G. Gentile, E. Alessi, C. Schneider, S. Petralia, S. Cavallaro. Is this the real time for genomics? *Genomics*, 2014; 103: 177-82.
18. B. M. Bolstad, R. A. Irizarry, M. Astrand, T. P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 2003; 19: 185-193.
19. M. Raimers. Statistical analysis of microarray data. *Addiction Biology*, 2005; 10: 23-35.
20. O. G. Troyanskaya, M. E. Garber, P. O. Brown, D. Botstein, R. B. Altman. Nonparametric methods for identifying differentially expressed genes in microarray data. *Bioinformatics*, 2002; 18: 1454-61.
21. L. Guo, E. K. Lobenhofer, C. Wang, R. Shippy, S. C. Harris, L. Zhang, et al. Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nature Biotechnology*, 2006; 24: 1162 – 1169.
22. V. G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci (USA)*, 2001; 98: 5116 –5121.
23. D. M. Witten, R. A. Tibshirani. A comparison of fold-change and the t-statistic for microarray data analysis. Stanford University. 2007; Available at <http://statweb.stanford.edu/~tibs/ftp/daniela-fold.pdf>
24. K. F. Kerr. Comments on the analysis of unbalanced microarray data. *Bioinformatics*, 2009; 25: 2035-2041.
25. D. B. Allison, X. Cui, G. P. Page, M. Sabripou. Microarray data analysis: from disarray to consolidation and consensus. *Nature Reviews Genetics*, 2006, 7; 55–65.
26. B. Efron, R. Tibshirani, J. D. Storey, V. Tusher. Empirical Bayes Analysis of a Microarray Experiment. *J American Statistical Association*, 2001, 96; 1151-1160.
27. Y. Benjamini, Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Royal Statistical Society, Series B (Methodological)*, 1995; 57: 289-300.
28. E. Hansen, K. F. Kerr. A comparison of two classes of methods for estimating false discovery rates in microarray studies. *Scientifica*, 2012; article ID: 519394. Available at: <http://www.hindawi.com/journals/scientifica/2012/519394/abs/>
29. K. Shedden, W. Chen, R. Kuick, D. Ghosh, J. Macdonald et al. Comparison of seven methods for producing Affymetrix expression scores

- based on False Discovery Rates in disease profiling data. *BMC Bioinformatics*, 2005; 6: 26
30. R. Suresh, X. Li, A. Chiriac, K. Goel, A. Terzic, C. Perez-Terzic, T. J. Nelson. Transcriptome from circulating cells suggest dysregulated pathways associated with long-term recurrent events following first-time myocardial infarction. *J. Molecular and Cellular Cardiology*, 2014; 74: 13–21.
 31. X. Cui, G. A. Churchill. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol*, 2003; 4(4):210.
 32. W. Li. Volcano plots in analyzing differential expressions with mRNA microarrays. *J Bioinform Comput Biol*, 2012; 10:1231003.
 33. M. R. Dalman, A. Deeter, G. Nimishakavi, Z-H. Duan. Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinformatics*, 2012; 13:S11.
 34. E. Hansen, K.F. Kerr. A comparison of two classes of methods for estimating false discovery rates in microarray studies. *Scientifica*, 2012; article ID 519394. Available in the web at: <http://www.hindawi.com/journals/scientifica/2012/519394/>
 35. G. K. Smyth. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 2004; 3(1): Article 3.
 36. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G. K. Smyth. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, 2015; 43: e47.
 37. I. Medina, J. Carbonell, L. Pulido, S.C. Madeira, S. Goetz et al. Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res*, 2010; 38: W210-3.
 38. T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim et al. NCBI GEO: archive for functional genomics data sets update. *Nucleic Acids Res*, 2013; 41: D991–D995.
 39. Y. Cai, Q. Yang, Y. Tang, M. Zhang, H. Liu et al. Increased Complement C1q Level Marks Active Disease in Human Tuberculosis. *PLoS ONE*, 2014; 9: e92340.
 40. H. Levy, X. Wang, M. Kaldunski, S. Jia, J. Kramer et al. Transcriptional signatures as a disease-specific and predictive inflammatory biomarker for type 1 diabetes. *Genes and Immunity*, 2012; 13, 593–604.
 41. W. S. Liang, T. Dunckley, T.G. Beach et al. Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain. *Physiological Genomics*, 2007, 28; 311-322.
 42. U. Ungethuen, T. Haeupl, H. Witt, D. Koczan, V. Krenn et al. Molecular signatures and new candidates to target the pathogenesis of rheumatoid arthritis. *Physiological Genomics*, 2010; 42A: 267-282.

43. J. M. Kristensen, V. Skov, S. J. Petersson, N. Ørtenblad et al. A PGC-1 α - and muscle fibre type-related decrease in markers of mitochondrial oxidative metabolism in skeletal muscle of humans with inherited insulin resistance. *Diabetologia*, 2014; 57:1006-1015.
44. D. R. Crooks, T.G. Natarajan, S. Y. Jeong, C. Chen et al. Elevated FGF21 secretion, PGC-1 α and ketogenic enzyme expression are hallmarks of iron-sulfur cluster depletion in human skeletal muscle. *Hum Mol Genet*, 2014; 23: 24-39.
45. S. Affò, M. Dominguez, J. J. Lozano, P. Sancho-Bru et al. Transcriptome analysis identifies TNF superfamily receptors as potential therapeutic targets in alcoholic hepatitis. *Gut*, 2013; 62: 452-60.
46. X. M. Wang, W. Y. Yik, P. Zhang, W. Lu et al. The gene expression profiles of induced pluripotent stem cells from individuals with childhood cerebral adrenoleukodystrophy are consistent with proposed mechanisms of pathogenesis. *Stem Cell Res Ther*, 2012; 3: 39.
47. E.M. Blalock, J.W. Geddes, K.C. Chen, N.M. Porter et al. Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc Natl Acad Sci (USA)*, 2004; 101: 2173-2178.
48. N.C. Berchtold, D. H. Cribbs, P.D. Coleman, J. Rogers et al. Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc Natl Acad Sci (USA)*, 2008; 105:15605-15610.
49. M. Jernås, Y. Hou, F. Strömberg Céлинд, L. Shao L et al. Differences in gene expression and cytokine levels between newly diagnosed and chronic pediatric ITP. *Blood*, 2013; 122:1789-1792.
50. T. M. Therneau, K.V. Ballman. What Does PLIER Really Do? *Cancer Informatics*, 2008; 6:423-431.
51. Z. Wu, R. A. Irizarry. Preprocessing of oligonucleotide array data. *Nature Biotechnology*, 2004; 22: 656 – 658.
52. Z. Wu, R. A. Irizarry. Description of GCRMA package. Oct.13 (2014) available from: <http://www.bioconductor.org>.
53. I.V. Yang, E. Chen, J. P. Hasseman, W. Liang, B.C. Frank et al. Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol*, 2002; 3(11):research0062.
54. D. Salas-Gonzalez, E.E. Kuruoglu, D. P. Ruiz. A heavy-tailed empirical Bayes method for replicated microarray data. *Computational Statistics and Data Analysis*, 2009; 53:1535–1546.
55. D.J. McCarthy, G.K. Smyth. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*, 2009; 25:765-7