Chilean Journal of Statistics
Vol. 1, No. 1, April 2010, 91–102

Special Issue "In Memory of Pilar Loreto Iglesias Zuazola"
Research Paper

Significance test for comparing digital gene expression profiles: Partial likelihood application

Leonardo Varuzza and Carlos Alberto de Bragança Pereira

Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil

(Received: 24 September 2009 · Accepted in final form: 22 January 2010)

Abstract

Most of the statistical tests currently used to detect differentially expressed genes are based on asymptotic results and may not be appropriate for low expression tags. Another problem is the common use of a single canonical cutoff, the critical level, for the *p*-values of all the tags, without taking into consideration the type II error and the highly variable character of the total frequency of each tag. This work reports the development of an exact significance test for comparing digital expression profiles that, in contrast to the χ^2 test, does not use asymptotic considerations. The test allows the use of a tag-customized critical significance level which minimizes a linear combination of type I and type II errors. Hence, the critical significance level is a function of the total tag expression. This feature implies that the identification of differentially expressed tags can be reliably determined, in a manner that depends on both: the *p*-value and the critical level calculated for each tag. We implemented this test on `kemp`, a `C` language program available under the general public license (GNU) at `http://code.google.com/p/kempbasu`.

Keywords: Bioinformatics · Gene expression · SAGE · Significance test.

Mathematics Subject Classification: Primary 62F03 · Secondary 62P10.

1. Introduction

Crucial events in the biology of living organisms, such as cell differentiation and specialization, depend on minute variations of gene expression under different conditions and/or temporal events. A key approach to elucidate a gene function is to quantify and compare the expression level of a large set of genes in different tissues or developmental stages, or under different conditions/treatments. This task can be performed by using large-scale hybridization to microarrays or by counting gene tags or signatures using methods such as serial analysis of gene expression (SAGE) (Velculescu et al., 1995) and massively parallel signature sequencing (MPSS) (Brenner et al., 2000). By comparing transcript expression profiles among different samples, one can identify differentially expressed genes associated with a particular tissue and/or condition. Unlike microarrays analysis, SAGE and MPSS

*Corresponding author. Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, Cidade Universitária, 05508-090 São Paulo, SP, Brazil. Email: varuzza@gmail.com (L. Varuzza); cadebp@gmail.com (C.A.B. Pereira)

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

do not require any prior knowledge of the transcript sequences. These techniques provide a digital profiling and permit to estimate the relative abundance of mRNA molecules of a transcriptome based upon two main premises. First, these methods assume that each position-specific short sequence tag can unequivocally identify its corresponding transcript. Second, tag counts are considered representative of the abundances of the corresponding mRNAs of the transcriptome: i. e., every mRNA copy has the same chance of being counted as the corresponding tag of the library. The selection of a specific tag sequence, from the total pool of transcripts, can be well approximated as a sampling with replacement (Stollberg et al., 2000).

Several statistical tests have been devised to deal with the problem of comparing digital expression profiles and identifying differentially expressed genes (Audic and Claverie, 1997; Baggerly et al., 2004; Robinson and Smyth, 2007; Stekel et al., 2000; Thygesen and Zwinderman, 2006; Vêncio et al., 2004) and reviewed by Man et al. (2000), Romualdi et al. (2001) and Ruijter et al. (2002). Most of these tests, including the χ^2 test as the most popular representative, rely on asymptotic approximations and, as such, may perform inadequately when the sample size is small, as is the case of low expression tags. Another problematic aspect of detecting differentially expressed genes by significance tests is the use of a single critical level, such as the canonical values 0.1, 0.05 or 0.01, which, apart from the common use, do not present any particular advantage. The significance test introduced here does not use any asymptotic considerations for the test statistic distribution: hence, it should produce more appropriate results for low frequency tags than the asymptotic ones, as the case of the χ^2 significance test for homogeneity. Also, our test uses a tag-customized critical level which minimizes a linear combination of type I and type II errors. As seen in the sequel, for tags with high expression, the corresponding p-values of the two tests –the one introduced here and the classical χ^2– are almost equal, as illustrated by Figure 3. On the other hand, for low frequencies, it seems that the p-value obtained for the new test is closed to the average of the χ^2, see Figure 4. The exact test p-value is obtained using the multinomial cumulative distribution function, a discrete function with jumps in its possible observations. Contrarily, the χ^2 distribution function is absolutely continuous and has, evidently, no jumps.

2. Methods

An exact significance test is described in the sequel. Assuming $k(>1)$ libraries, let m be the number of distinct tags and n_j the total number of tags in jth library, for $j = 1, \ldots, k$. The frequency of the ith tag in the jth library is denoted by x_{ij} with $n_j = x_{1j} + \cdots + x_{mj}$. The basic statistical model can be stated as:

(i) For each j, the random vector $X_{.j} = (x_{1j}, \ldots, x_{mj})$ is distributed as a multinomial of order m with parameters n_j and $P_{.j} = (p_{1j}, p_{2j}, \ldots, p_{mj})$.

(ii) Assuming that the libraries have been collected independently, the random vectors X_j, $j = 1, \ldots, k$ are mutually statistically independent.

(iii) Considering that, for digital expression profiles, parameters $P_{.j}$ and n_j usually assume, respectively, low and high values. Audic and Claverie (1997), Cai et al. (2004) and Zhu et al. (2008) have considered that the distribution of tag frequency x_{ij} can be approximated by a Poisson distribution with mean $n_j p_{ij}$.

(iv) Considering the ith tag, assuming that the libraries have been collected independently and denoting the vectors of observations and parameters as $X_{i.} = (x_{i1}, \ldots, x_{ik})$ and $P_i. = (p_{i1}, \ldots, p_{ik})$, the full model for this single ith tag is as

$$f(X_i.|P_i.) = \frac{(n_1 p_{i1})^{x_{i1}} \cdots (n_k p_{ik})^{x_{ik}}}{(x_{i1})! \cdots (x_{ik})!} \exp\{-(n_1 p_{i1} + \cdots + n_k p_{ik})\}. \qquad (1)$$

2.1 Partial likelihood

Using an alternative parameterization, we rewrite Equation (1) as

$$\begin{aligned} f(X_{i\cdot}|P_{i\cdot}) &= f(X_{i\cdot}|\Pi_i,\theta_i) = f(X_{i\cdot}|y_i,\Pi_i)g(y_i|\theta_i) \\ &= \left\{\frac{y_i!}{x_{i1}!\times\cdots\times x_{ik}!}\pi_{i1}^{x_{i1}}\times\cdots\times\pi_{ik}^{x_{ik}}\right\}\left\{\frac{\theta_i^{y_i}\exp(-\theta_i)}{y_i!}\right\}. \end{aligned} \tag{2}$$

Here, we have

$$y_i = x_{i1}+\cdots+x_{ik};\ \ \theta_i = n_1p_{i1}+\cdots+n_kp_{ik};\ \ \pi_{ij} = \frac{n_jp_{ij}}{\theta_i};\ \text{and}\ \Pi_i = (\pi_{i1},\ldots,\pi_{ik}).$$

Also, g is the marginal density of y_i, the total expression of the ith tag.

The consequence of this parameterization is that the original likelihood becomes the product of a multinomial probability function by a Poisson probability function. It is noteworthy that the parameters, Π_i and θ_i, are variation independent; i.e., the value of one does not bring any information about the value of the other – the parameter space of (Π_i,θ_i) is the Cartesian product of the parameter spaces of Π_i and of θ_i –. According to Base (1977) and Cox (1975), for inferences about Π_i (or θ_i) one may only consider the multinomial (or Poisson) probability factor of Equation (2) as the likelihood function.

Under the full likelihood model (1), the null hypothesis –ith tag has the same expression in all libraries– is, for any p in the interval (0, 1), given by H: $P_{i.} = (p_{i1},\ldots,p_{ik}) = (p,\ldots,p)$. Under the new parameterization and the use of the partial likelihood approach, for $n = n_1+\cdots+n_k$, the null hypothesis becomes a simple (single point) null hypothesis as H': $\Pi_i = \Pi_0 = (n_1/n,\ldots,n_k/n)$. The partial likelihood approach, discussed by Cox (1975) and Pereira and Lindley (1987), simplifies considerably the problem of comparing the expressions of the ith tag over the k libraries. Hence, under the null hypothesis, the likelihood is simply a multinomial probability function evaluated for Π_0. The two new statistical models –under the null and alternative hypotheses– to be compared are, respectively,

$$H'\colon\quad f(x_{i1},\ldots,x_{ik}|y_i,\Pi_0) = \frac{y_i!}{n^{y_i}}\prod_{j=1}^{k}\frac{n_j^{x_{ij}}}{x_{ij}!}\quad\text{and} \tag{3}$$

$$A'\colon\quad f(x_{i1},\ldots,x_{ik}|y_i,\Pi_i) = (y_i!)\prod_{j=1}^{k}\frac{\pi_{ij}^{x_{ij}}}{x_{ij}!}. \tag{4}$$

2.2 Significance level

According to Cox (1977) and Kempthorne (1976), a significance test is a method that measures the consistency of the data with the null hypothesis. The commonly used index to perform this task is the well known p-value. We refer to Kempthorne and Folks (1971) for important discussions on the evaluation of p-values. If R is a statistic, a real function of the sample observations, in which small values of R cast doubt about the null hypothesis H, the p-value associated to the observed r, is the probability under H of the event $\{R \leq r\}$, that is, $pv = \mathrm{P}\{R \leq r|H\}$. The consequence of this definition is that the random variable R must be a function that produces an order in the sample space. In this ordered sample space, the sample points with low order favor the alternative hypothesis, whereas those with a higher order support the null hypothesis.

The likelihood ratio, used in most statistical methods, is an appropriate statistic for ordering the sample space relative to the null hypothesis, Kempthorne and Folks (1971) and Pereira and Wechsler (1993). Considering the ith tag, the corresponding likelihood ratio is the maximum of the likelihood function under H, Equation (3), divided by the overall maximum of the likelihood function; Equation (4) in which $\frac{x_{ij}}{y_i}$ replaces π_{ij}. Let $R_i(W)$ be the likelihood ratio for the ith tag evaluated at any possible observation $W = (w_1, \ldots, w_k)$ and r_i the actual observed value, $R_i(X_i) = r_i$. The p-value for the ith tag is $pv_i = \mathrm{P}\{R_i(W) \leq r_i | H\}$ and the likelihood ratio is

$$R_i(W) = \left(\frac{y_i}{n}\right)^{y_i} \prod_{j=1}^{k} \frac{n_j}{w_j}. \tag{5}$$

Considering the tail set $T = \{W | R(W) \leq r_i\}$, the new p-value has the expression

$$pv_i = \sum_{w \in T} \frac{y_i!}{n^{y_i}} \prod_{j=1}^{k} \frac{n_j^{w_j}}{w_j!}. \tag{6}$$

The use of likelihood ratios for computing p-values has been already addressed by Neyman and Pearson (1928), Pereira and Wechsler (1993) and Dempster (1997).

A serious limitation of this exact p-value calculation is that the number of points in the sample space grows exponentially with the dimension of the problem (the number of libraries in the present case). To overcome this problem, we use an algorithm based on the Monte Carlo method. In order to test and validate this method, we developed `kemp`, a `C` language program named after Prof. Oscar Kempthorne, an English statistician who has produced an extensive work on the topic of significance tests.

2.3 Critical significance level

In order to have inferential meaning, the evaluation of a significance level should help one to decide in favor or against a null hypothesis. Hence, a decision rule must be stated. For instance, the critical level is the cutoff between reject/accept actions. In any digital expression profile, the relative abundance can vary dramatically from tag to tag, implying that the use of a single critical significance level for all tags may be unfair for those tags with low frequencies. For this reason, we decided to calculate a critical level for each particular tag, according to the recommendations of DeGroot (1986). Thus, we used the optimum procedure of the decision theory, which minimizes the risk function $a\alpha + b\beta$, a linear combination of the two kinds of errors: α and β, corresponding to errors of type I (false positive) and type II (false negative), respectively.

The value of α is the probability of the critical region using the parameter value defined by the null hypothesis. Conversely, computation of β is more complex, since A, the alternative hypothesis, is composed rather than a single point hypothesis. To solve the problem of defining the appropriate β, we consider the average of the probability of the complement of the critical region over all possible single parameter points within the parameter set defined by the alternative hypothesis. Perform this computation is equivalent to use a uniform prior over the parameter space and consider the predictive distribution for this prior choice. Fortunately, the predictive is a uniform discrete distribution in the sample space. Hence, it is a constant equal to the inverse of the number of points of the sample space. The number of points of the sample space for k libraries is equal to the combination of $(y + k - 1)$ taken y to y, i.e., $\frac{(y+k-1)!}{y!(k-1)!}$. Therefore, the aforementioned average to obtain β is the number of points within the acceptance region, divided by this constant.

To choose the critical level, one should choose, among all possible critical regions, the one that produces the minimum of $a\alpha + b\beta$. The value of α obtained with this choice of critical region is the critical level that is used in our test. The equation

$$S = \frac{10(\alpha - pv)}{\alpha} \tag{7}$$

corresponds to a score, based on practical results, to order the differentially expressed tags according to the relative "distance" between their corresponding p-value, pv, and the critical level, α.

3. Results

3.1 The critical level as a function of y_i

In order to establish an automatic procedure to discriminate tags according to their differential expression status, we defined two sets of weights for type I and type II errors, respectively: $(a = 1, b = 1)$ and $(a = 4, b = 1)$. These values were arbitrarily defined based on several analyses of SAGE (serial analysis of gene expression) data from human libraries, followed by experimental validations with real-time PCR (data not shown). See Silveira et al. (2008) for additional discussion on this matter. The α critical levels were computed for a range of possible values of y_i (the total tag frequency) and for k values (number of libraries) ranging from 2 to 5. Since the calculation of α is a computer intensive task, we estimated a polynomial approximation of this critical level for each value of k. This result was incorporated on `kemp`, our implemented software for the exact significance test, and is available in Appendix (Supplementary Material). Figures 1 and 2 present the dilog graphics of the critical level values and the corresponding adjusted functions, assuming k values of 2 and 5, respectively.

As can be seen, the weights $a = 4$ and $b = 1$ generates critical level values that are consistently lower than those obtained using $a = b = 1$. This result can be ascribed to weight 4 used for the α error, which leads to a greater minimization of type I error. Also, for both sets of weights, when y presents high values, the critical level is much more stringent than the canonical values 0.1, 0.05 and 0.01.

3.2 Comparison between kemp p-value and χ^2 p-value

The χ^2 homogeneity significance test is widely used and described in the literature for comparison of digital expression profiles (Man et al., 2000; Romualdi et al., 2001). We decided to compare our significance level (p-value) to the χ^2 test. We used a data set composed of four SAGE libraries derived from human brain tissues and potentially containing genes involved in increased risk of Alzheimer Disease (GEO accession code GSE6677). The tags were arbitrarily separated into two groups according to their total frequency (y). Thus, we calculated the significance levels using the Kemp method and the χ^2 test for both, the high-expression tags ($y > 50$, Figure 3) and low expression tags ($y \leq 50$, Figure 4). Figure 3 illustrates that there is a good agreement between the significance levels (0.999 Pearson correlation); it becomes better as y increases. Conversely, Figure 4 shows that when the expression of the tags is relatively low, the two significant levels present a lower agreement (0.66 correlation). This result is in consonance with what we should expect, since the χ^2 test is an asymptotic approximation and is based in a continuous distribution function, whereas our proposed significance test is exact and based in a discrete frequency distribution that have jumps in its distribution function. We notice that the new exact p-value average is closed to the χ^2 p-value and vice-versa.

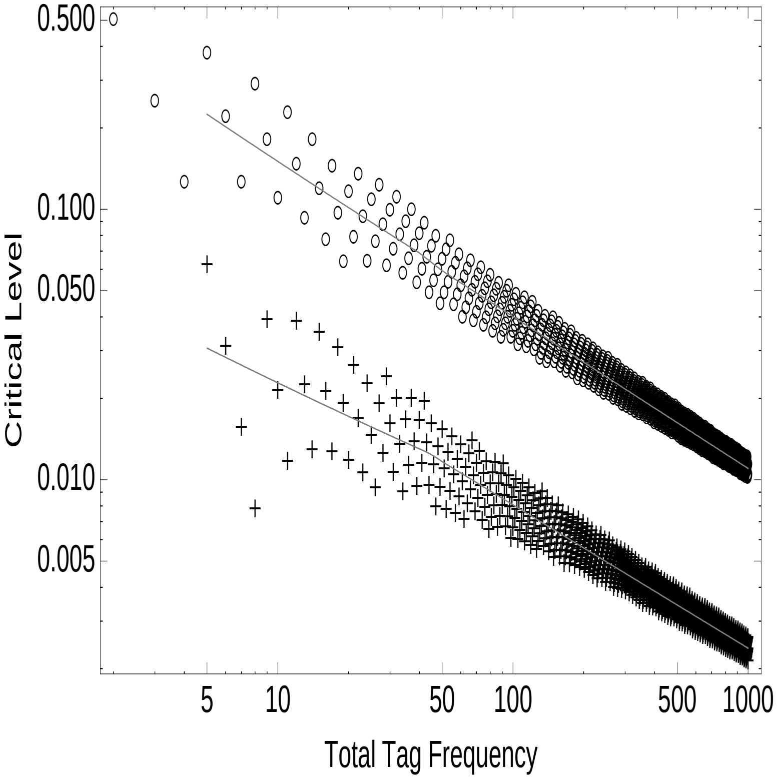

Figure 1. Dilog graphic with the simulated values of the critical level and the fitted function for $k = 2$. Critical levels with weights $a = b = 1$ are indicated by and with weights $(a = 4, b = 1)$ are indicated by $+$.

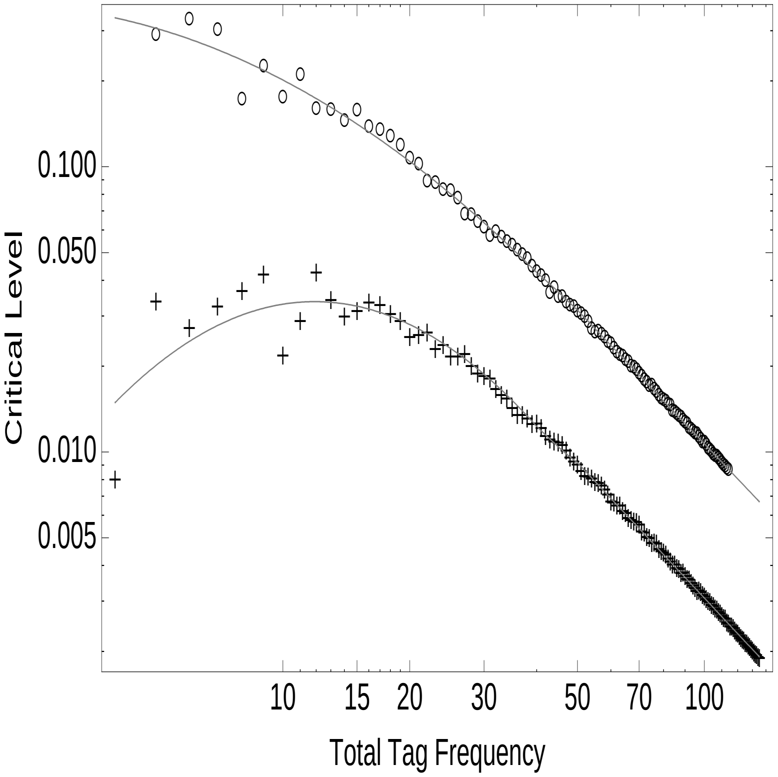

Figure 2. Dilog graphic for $k = 5$ with the simulated values of the critical level and the fitted function. Critical levels with weights $a = b = 1$ are indicated by $\circ$ and with weights $(a = 4, b = 1)$ are indicated by $+$.

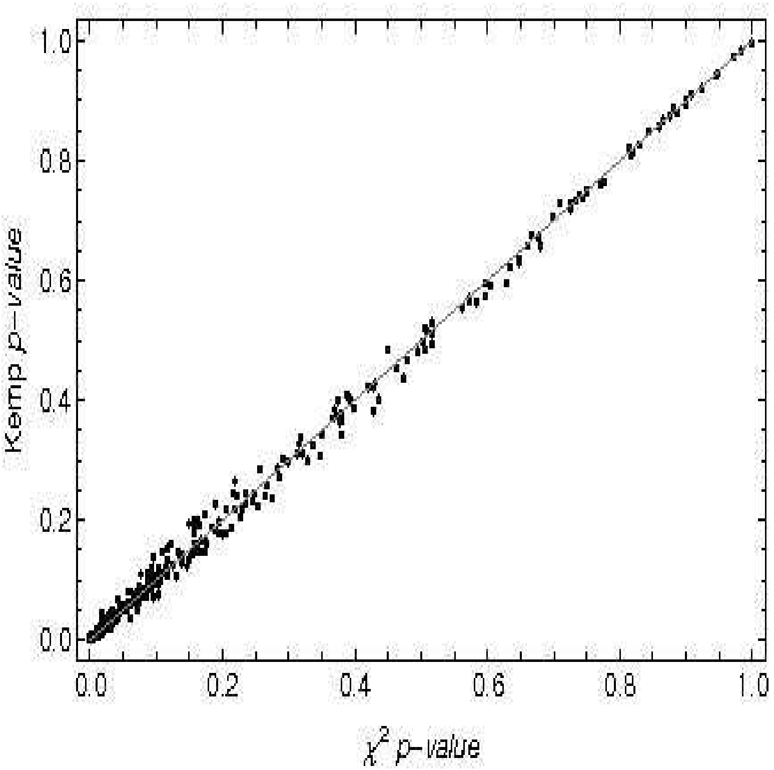

Figure 3. Relation between kemp and χ^2 significance levels for tags with $y > 50$.

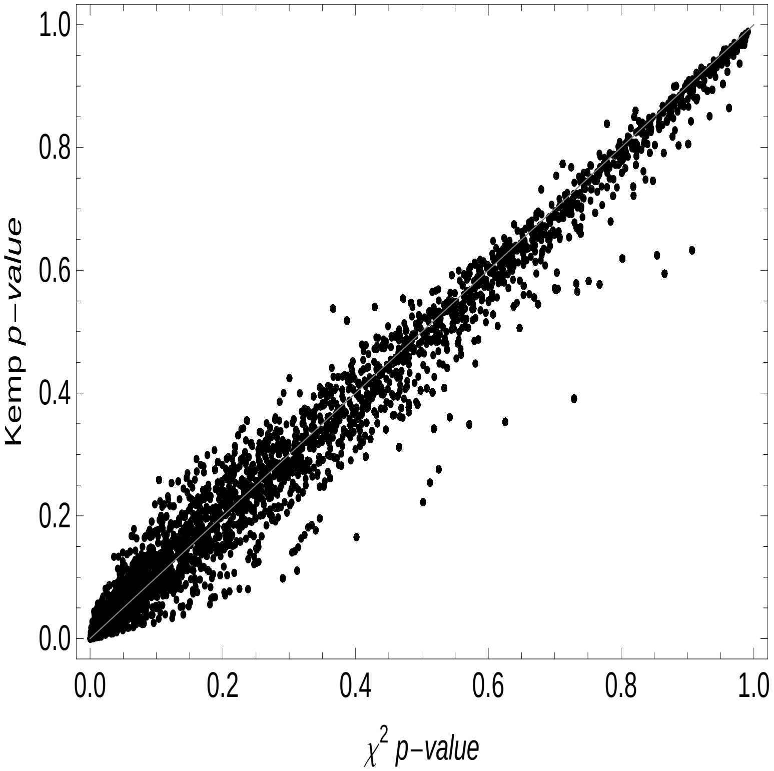

Figure 4. Relation between kemp and χ^2 significance levels for tags with $y \leq 50$.

4. Discussion

This paper introduces a simple method for the comparison of digital expression profiles. The method was implemented on the open source program `kemp`. Several statistical tests have been used to evaluate SAGE data and identify differentially expressed tags. Some of these tests have been compared by different groups (Man et al., 2000; Romualdi et al., 2001; Ruijter et al., 2002). The general conclusion was that the classical χ^2 test, originally introduced by Karl Pearson (Pearson, 1900), was equivalent to and even outperformed other available tests, including the test of Audic and Claverie (1997) and the R statistic of Stekel et al. (2000). The χ^2 test has the advantage of being simple and can be applied to a broad range of problems. However, given the asymptotic character of the χ^2 test, it may not be recommended for the analysis of low frequency tags. Due to this feature, we conceived our test using the original definition of more extreme sample points, without any asymptotic result. As a consequence, our test may be more realistic than the χ^2 test for low expression tags since uses the exact frequency distribution. Corroborating this fact, a comparison of the p-values calculated by `kemp` and the χ^2 p-value, showed a good agreement, Figure 3, for high expression tags where continuity approximation is computationally recommended. For low frequency tags, for which the χ^2 test could become inappropriate, Figure 4 shows disagreement between the two p-values. Thus, we believe that the test proposed here and implemented on `kemp`, represents an improvement for the analysis of digital expression profiles since, besides been exact, it allows the use of optimal critical values obtained by minimization of linear combinations of the two kinds of errors.

The discrimination between high and low expression tags must be performed in such a way as to consider both statistical and biological relevance. Since housekeeping genes are expressed in high levels, the absolute number of counts may present a considerable variation across distinct libraries. These differences, nevertheless, are meaningless from a biological standpoint. Conversely, some functionally important genes present a relatively low expression and exert their activity by altering tiny amounts of their expression among the different tissues and/or conditions (Wang, 2006). Therefore, a tag presenting a differential expression with low counts would not be considered as significant by methods that use fixed critical levels, thus leading to a misinterpretation of the data and loss of potentially valuable information. This fact motivated us to calculate the critical level of each particular tag taking into account its total frequency. If population parameters are not exactly in the null sharp hypothesis set (a set with a smaller dimension than the alternative hypothesis set), highly expressed tags have smaller p-values than low expression tags. If one fixes the critical level, the minimum type I error, the type II error decreases drastically when the total expression increases. Hence, it becomes difficult to accept a null hypothesis for large sample sizes and to reject it for small-sized samples. For example, when comparing the expression of two 10,000-tag libraries, a tag presenting counts of 7 and 21, respectively, would show a p-value of 0.013. Conversely, a tag with counts 10 and 30 would result in a p-value of 0.002. Considering a cutoff of 0.01, the last tag would be considered as equally expressed within the two libraries, whereas the former tag would be interpreted as being differentially expressed. This fact motivated us to calculate the critical level of each particular tag as a function of the tag total frequency. The critical region in our method is the one that minimizes a linear combination of type I (the critical level α) and type II (β) errors. With this tag-customized approach, both tags of the example above would be classified as differentially expressed, since their corresponding critical levels would be 0.015 and 0.013, respectively. The method is still coherent, since tags with very low frequencies, even presenting differential counts, lead to high significance levels. For instance, in the aforementioned example, tag counts of 1 and 3 would result in a p-value of 0.63, a much higher value than the calculated cutoff of 0.03. Concluding, our method judges the tags in a fairer manner, since the cutoff value is customized to any particular tag, according to its expression level. This fact suggests that even for high expressed tags our method should be used to obtain the adequate critical levels.

Some other methods (Baggerly et al., 2004; Robinson and Smyth, 2007; Thygesen and Zwinderman, 2006; Vêncio et al., 2004) have been proposed for the comparative analysis of SAGE data. However, since these methods are designed for comparing groups of libraries, their use is severely restricted in experiments where a single library is represented in each category or condition.

Concluding, the `kemp` significance test reported in the present work, and implemented in standalone open-source programs, extend the set of currently available statistical tests for digital expression profiles. Also, we believe that they offer some advantages over other reported tests, including a more adequate treatment of low expression tags and the automatic calculation of a customized critical level. We also developed Bayesian version of the digital expression test, which uses the FBST test (Pereira et al., 2008) and is named Basu in honor of Prof. Debrabata Basu. The reader can find this program in the same package as `kemp`. If a Bayesian cutoff, well defined as the one presented here, can be developed, we will present another paper comparing Basu and Kemp tests.

System Requirements

The source code of `KempBasu` package and a executable binary for MS Windows are publicly available at the address `http://code.google.com/p/kempbasu` and are distributed under the general public license (GNU). The code depends on `glib`, `GSL` and `Judy` libraries, and if the `Pthreads API` is available, `KempBasu` can be run using multiple processors. Tested platforms include Linux, MacOSX and MS Windows.

Appendix
Supplementary Material about the Adjust of Kemp Cutoff Functions

To establish the function of y that gives the approximate α, we consider the pairs $\{L = \log(y); \log(\alpha)\}$ and use the least squares method piecewisely in two difference regions of y values: [1;50] and [51;10,000]. The former region adjusts a second degree polynomial $\alpha_k^{(1)}(y) = a_k L^2 + b_k L + c_k$ and the latter region adjusts a line $\alpha_k^{(2)}(y) = u_k + v_k L$. Tables 1 and 2 present the coefficient values for those functions for weights $(a = 4, b = 1)$ and $(a = 1, b = 1)$, respectively . The linear and the quadratic functions are combined, for each value of k, into a single continuous function by the Equation (8). Figures 1 to 2 show the function adjustment to the calculated pairs.

$$\alpha_k(y) = \begin{cases} \alpha_k^{(1)}(y), & y < 40; \\ (1-\lambda)\alpha_k^{(1)}(y) + \lambda\alpha_K^{(2)}(y), & 40 \leq y < 50; \\ \alpha_k^{(2)}(y), & y \geq 50. \end{cases} \tag{8}$$

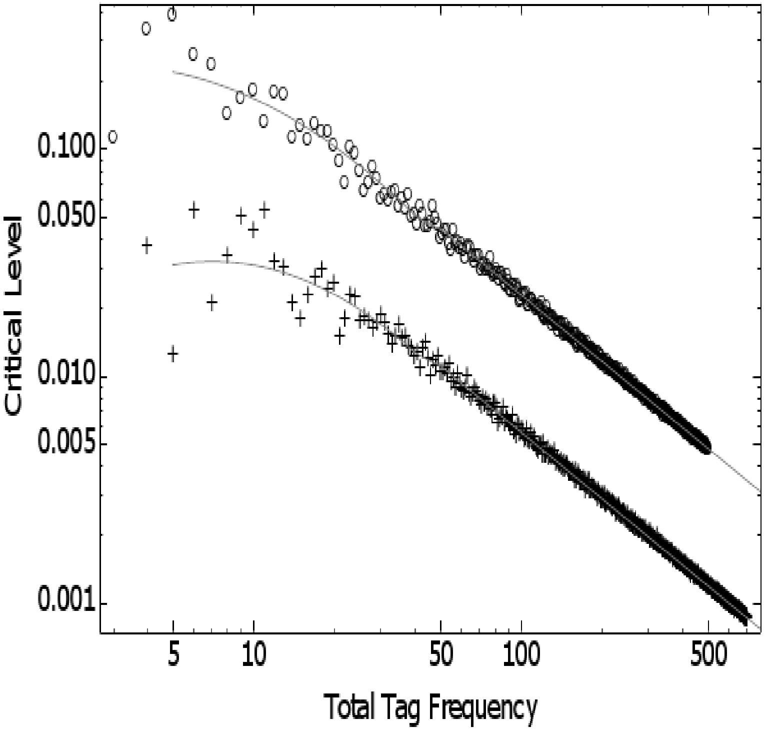

Figure 5. Dilog graph with the simulated values of the critical level and the fitted function for $k = 3$. Critical levels calculated with weights $(a = 1, b = 1)$ are indicated by $\circ$ and with weights $(a = 4, b = 1)$ are indicated by $+$.

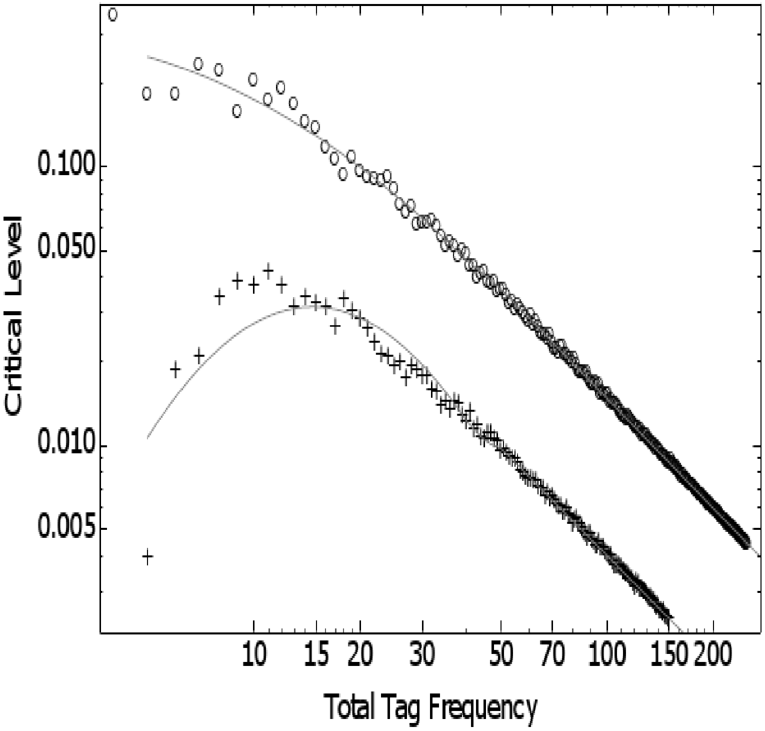

Figure 6. Dilog graph with the simulated values of the critical level and the fitted function for $k = 4$. Critical levels calculated with weights $(a = 1, b = 1)$ are indicated by $\circ$ and with weights $(a = 4, b = 1)$ are indicated by $+$.

Table 1. Coefficient values of fitted critical level functions for the minimization of $4\alpha + \beta$.

k	a_k	b_k	c_k	u_k	v_k
2	0.009580	-0.46312	-2.76474	-2.37781	-0.53012
3	-0.304365	1.18976	-4.60784	-0.71361	-0.96851
4	-0.931159	5.00318	-10.1863	0.38512	-1.28105
5	-0.685327	3.39467	-7.59502	1.47602	-1.57657
6	-0.914225	4.84175	-9.81444	1.93518	-1.70783

Table 2. Coefficient values of fitted critical level functions for the minimization of $\alpha + \beta$.

k	a_k	b_k	c_k	u_k	v_k
2	0.007480	-0.607463	-0.53588	-0.62914	-0.56174
3	-0.226299	0.503742	-1.75040	0.67763	-0.96817
4	-0.215143	0.334093	-1.38061	1.79399	-1.30545
5	-0.248689	0.369967	-1.13529	2.62984	-1.55664

Acknowledgements

Leonardo Varuzza received a PhD fellowship from CAPES. Carlos A.B. Pereira is research fellow of CNPq. The authors thank very much Dr A. Gruber and Dr H. Brentani for helpful comments in early versions of the paper.

References

Audic, S., Claverie, J.M., 1997. The significance of digital gene expression profiles. Genome Research, 7, 986-995.

Baggerly, K., Deng, L., Morris, J., Aldaz, C., 2004. Overdispersed logistic regression for SAGE: modeling multiple groups and covariates. BMC Bioinformatics, 5, 144.

Base, D., 1977. On the elimination of nuisance parameters. Journal of the American Statistical Association, 72, 355-366.

Brenner, S.E., Johnson, M., Bridgham, J., Golda, G., Lloyd, D., Jonhson, D., Luo, S., McCurdy, S., Foy, M., 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nature Biotechnology, 18, 630-634.

Cai, L., Huang, H., Blackshaw, S., Liu, J.S., Cepko, C., Wong, W.H., 2004. Clustering analysis of SAGE data using a Poisson approach. Genome Biology, 5, R51.

Cox, D.R., 1975. Partial likelihood. Biometrika, 62, 269-276.

Cox, D.R., 1977. The role of significant test(with discussion). Scandinavian Journal of Statistics, 4, 49-70.

DeGroot, M.H., 1986. Probability and Statistics. Addison-Wesley, Boston.

Dempster, A.P., 1997. The direct use of likelihood for significance testing. Statistics and Computing, 7, 247-252.

Kempthorne, O., 1976. Of what use are tests of significance and tests of hypothesis. Communications in Statistics — Theory and Methods, 5, 763-777.

Kempthorne, O., Folks, L., 1971. Probability, Statistics and Data Analysis. Iowa State University Press, Iowa.

Man, M.Z., Wang, X., Wang, Y., 2000. POWER SAGE: comparing statistical tests for SAGE experiments. Bioinformatics, 16, 953-959.

Neyman, J., Pearson, E.S., 1928. On the use and interpretation of certain test criteria for purposes of statistical inference: part I. Biometrika, 20, 175-240.

Pearson, K., 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling. Philosophical Magazine Series, 50, 157-175.

Pereira, C.A., Lindley, D.V., 1987. Examples questioning the use of partial likelihood. The Statistician, 36, 15-20.

Pereira, C.A.B., Wechsler, S., 1993. On the concept of p-value. Brazilian Journal of Probability and Statistics, 7, 159-177.

Pereira, C.A.B., Stern, J.M., Wechsler, S., 2008. Can a significance test be genuinely Bayesian? Bayesian Analysis, 3, 79-100.

Robinson, M., Smyth, G., 2007. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics, 9, 321-332.

Romualdi, C., Bortoluzzi, S., Danieli, G., 2001. Detecting differentially expressed genes in multiple tag sampling experiments: comparative evaluation of statistical tests. Human Molecular Genetics, 10, 2133-2141.

Ruijter, J., Van Kampen, A., Bass, F., 2002. Statistical evaluation of SAGE libraries: consequences for experimental design. Physiological Genomics, 11, 37-44.

Silveira, N.J., Varuzza, L., Machado-Lima, A., Lauretto, M.S., Pinheiro, D.G., Rodrigues, R.V., Severino, P., Nobrega, F.G., Silva, W.A., Pereira, C.A.B., Tajara, E.H., 2008. Searching for molecular markers in head and neck squamous cell carcinomas (HNSCC) by statistical and bioinformatic analysis of larynx-derived SAGE libraries. BMC Medical Genomics, 1, 56.

Stekel, D.J., Git, Y., Falciani, F., 2000. The comparison of gene expression from multiple cDNA libraries. Genome Research, 10, 2055-2061.

Stollberg, J., Urschitz, J., Urban, Z., Boyd, C.D., 2000. A quantitative evaluation of SAGE. Genome Research, 10, 1241-1248.

Thygesen, H.H., Zwinderman, A.H., 2006. Modeling SAGE data with a truncated gamma-Poisson model. BMC Bioinformatics, 7, 157.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. Science, 270, 484-487.

Vêncio, R.Z.N., Brentani, H., Patro, D.F.C., Pereira, C.A.B., 2004. Bayesian model accounting for within-class biological variability in serial analysis of gene expression (SAGE). BMC Bioinformatics, 5, 119.

Wang, S., 2006. Understanding SAGE data. Trends in Genetics, 23, 42-50.

Zhu, J., He, F., Wang, J., Yu, J., 2008. Modeling transcriptome based on transcript-sampling data. PLoS One, 3, e1659.