BIOINFORMATICS ORIGINAL PAPER Vol. 25 no. 19 2009, pages 2588–2594
doi:10.1093/bioinformatics/btp439

Gene expression

Testing the hypothesis of tissue selectivity: the intersection–union test and a Bayesian approach

K. Van Deun[1,*], H. Hoijtink[2], L. Thorrez[3], L. Van Lommel[3], F. Schuit[3] and I. Van Mechelen[1]

[1]Center for Computational Systems Biology SymBioSys, Katholieke Universiteit Leuven, 3000 Leuven, Belgium, [2]Department of Methodology and Statistics, University of Utrecht, 3508 TC Utrecht, The Netherlands and [3]Gene Expression Unit, Katholieke Universiteit Leuven, 3000 Leuven, Belgium

Received on April 7, 2009; revised on July 8, 2009; accepted on July 13, 2009

Advance Access publication August 11, 2009

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Finding genes that are preferentially expressed in a particular tissue or condition is a problem that cannot be solved by standard statistical testing procedures. A relatively unknown procedure that can be used is the intersection–union test (IUT). However, two disadvantages of the IUT are that it is conservative and it conveys only the information of the least differing target tissue–other tissue pair.

Results: We propose a Bayesian procedure that quantifies how much evidence there is in the overall expression profile for selective over-expression. In a small simulation study, it is shown that the proposed method outperforms the IUT when it comes to finding selectively expressed genes. An application to publicly available data consisting of 22 tissues shows that the Bayesian method indeed selects genes with functions that reflect the specific tissue functions. The proposed method can also be used to find genes that are underexpressed in a particular tissue.

Availability: Both MATLAB and R code that implement the IUT and the Bayesian procedure in an efficient way, can be downloaded at http://ppw.kuleuven.be/okp/software/BayesianIUT/.

Contact: katrijn.vandeun@psy.kuleuven.be

1 INTRODUCTION

Gene expression is often profiled for several tissues to get insights into gene function and regulation (Dezso *et al.*, 2008; Liu *et al.*, 2008; Su *et al.*, 2004). In this respect, finding genes that are selectively expressed in a particular tissue is of importance to unravel the biological processes taking place in the particular tissue and to identify candidate biomarkers (Klee, 2008; Liang *et al.*, 2006; Su *et al.*, 2004). It is important to make a distinction between three types of tissue-selectivity [see Klee (2008) for a review]: (i) the gene is only expressed in a particular tissue; usually this type of expression is called 'tissue-specific' (Skrabanek and Campagne, 2001); (ii) the gene is expressed at approximately the same level in all tissues except one [called categorical tissue specificity by Schug *et al.* (2005)]; or (iii) the gene is over- or underexpressed in a particular tissue compared with all the other tissues (Kadota *et al.*, 2003, 2006). The latter broadest type of preferential expression in a particular tissue is often called tissue-selective (Greller and Tobin, 1999; Liang *et al.*, 2006). Note that tissue-specific as defined by Skrabanek and Campagne (2001) is a special case of categorical tissue specificity and that categorical tissue specificity is a special case of tissue-selectivity. In accordance with these different definitions of tissue-selectivity, different methods to find genes with the particular expression profile have been proposed. The remainder of the article focuses on the broadest class of tissue-selectivity, namely relative over- or underexpression in a particular tissue compared with all the other tissues. All methods are explained and illustrated for selective over-expression, but the generalization to selective underexpression is straightforward.

To find tissue-selective genes, we found two approaches in the literature that rely on a firm statistical framework, hereby reducing arbitrary choices to a minimum. A first approach is unsupervised and relies on outlier detection to scan expression profiles for outlying values (Kadota *et al.*, 2003, 2006). The resulting tissue-selective profiles can be selectively expressed in more than one tissue and can be both up- or downregulated in these tissues. A drawback of the method is that it cannot account for biological and technical variation because either all replicate values are included in the analysis (with the likely outcome that only some replicate values will be detected as outlying), or a single representative measure (e.g. average over the replicates) has to be used in the analysis. The second approach is supervised and constructed for the case of replicate arrays for each tissue (Liang *et al.*, 2006). It relies on hypothesis testing procedures to test whether a gene is selectively overexpressed in a particular tissue. *t*-tests are used to measure how significant the difference in expression of each of the target tissue–other tissue pairs is and a gene is declared tissue-selective when each of the differences is significant. The problem of multiple testing is accounted for by using the Tukey–Kramer multiple comparison procedure. Although the obtained tissue-selective genes are claimed to have higher expression in the target tissue than in *each of* the other tissues, the statistical procedure used is tailored to find those genes that are significantly higher expressed in the target compared with *at least* one of the other tissues. To understand why this is the case, observe that Tukey–Kramer and other common multiple comparison procedures like Bonferroni and Dunnett (Dunnett, 1955) control the chance of wrongfully rejecting the null hypothesis of no significantly different target–other tissue pair against the alternative of *at least one* significantly different pair. What is needed, is a test that controls

*To whom correspondence should be addressed.

the chance of wrongfully rejecting the null hypothesis against the alternative of *all* significantly different target–other tissue pairs.

A test that would have the desired alternative hypothesis of significantly higher expression in the target tissue compared with each of the other tissues, was proposed by Berger (1982) and is known as the intersection–union test (IUT; Berger and Hsu, 1996). However, the IUT test has two disadvantages. First, it is conservative implying that many tissue-selective genes would be missed; and second, the obtained results are not very informative as the test only indicates whether a gene is tissue-selective or not without any distinction in the degree of tissue selectivity (see also Allison *et al.*, 2006). As an alternative, we would like to propose a Bayesian procedure.

In this article, we first briefly describe the IUT and we introduce a Bayesian alternative. Subsequently, both procedures are compared in a simulation study and the Bayesian procedure is used to find the tissue-selective genes for a panel of 22 tissues.

2 METHODS

2.1 Intersection–union test

Assume some specific gene and a specific target tissue t. Let us further denote by H_{0j} the partial null hypothesis that the gene under study is expressed equally or higher in tissue j than in the target tissue. Furthermore, we denote by H_{1j} the partial alternative that that the gene is expressed higher in the tissue target t. Then the gene is selectively upregulated in target tissue t, if the compound null hypothesis that H_{0j} holds for at least one tissue ($j \neq t$), is rejected against the compound alternative hypothesis that for all tissues ($j \neq t$) H_{1j} holds. Using formal notation, the set of compound hypotheses is composed by $H_0 = \bigcup H_{0j}$ and $H_1 = \bigcap H_{1j}$. Note that this is different from common multiple comparison procedures where the underlying set of compound hypotheses is that all partial null hypotheses hold (i.e. $H_0 = \bigcap H_{0j}$) against the alternative that at least one partial null hypothesis can be rejected (i.e. $H_1 = \bigcup H_{1j}$). Berger (1982) introduced a procedure to test the composite null hypothesis $H_0 = \bigcap H_{0j}$ against the composite alternative $H_1 = \bigcup H_{1j}$ (see also Berger and Hsu, 1996). A result is declared significant by this test at level α, if it holds that each partial null hypothesis H_{0j} can be rejected at level α. As proven by Berger, the significance level of his test is less than or equal to the significance level used for each of its implied partial tests.

The IUT can be applied to the problem of finding tissue-selective upregulated genes as follows. For a particular gene, test each target tissue–other tissue pair at the desired significance level (e.g. 0.05) using a suitable test-statistic like the t-test. Only when all pairs yield a significant result, the gene is declared selectively overexpressed. Note that to account for the problem of testing multiple genes, the significance level used can be adapted using Bonferonni's or Sidak's correction. An adaptation to finding underexpressed genes is straightforward by testing the partial null hypothesis that the gene under study is expressed equally or *lower* in tissue j than in the target tissue against the partial alternative that the gene is expressed lower in the target tissue. An efficient implementation of this approach in MATLAB or R can be found online (http://ppw.kuleuven.be/okp/software/BayesianIUT/).

Often the significance level of the IUT is (much) less than α such that the procedure is conservative (a gene will not be easily declared to be selectively upregulated in the tissue). This yields a very low false discovery rate, however, at the cost of many false negatives. Deng *et al.* (2008) proposed an adjusted IUT for the special case of two independent tests that is less conservative. Note, however, that it is not suitable to find tissue-selective genes (the common target makes that the tests are dependent and usually interest is in comparing the target with more than two other tissues). Another disadvantage of the IUT is that it indicates whether a gene is selectively overexpressed or not, but not to which degree. No distinction is made, for example, at the 0.05 level of significance between a tissue-selective upregulated gene that has P-value of 0.03 for each of the target–other tissue pairs and a gene that has P-value of 0.0001. The IUT can be made somewhat more informative by reporting the largest P-value of the partial tests (Tuke *et al.*, 2009), thus focusing only on the least differing target tissue–other tissue pair.

2.2 Bayesian evaluation of the constrained hypothesis

In this section, we describe a Bayesian alternative for the IUT to evaluate the hypothesis of tissue-selective overexpression. Note that the procedure will be explained for one gene. The data consist of $i = 1, \ldots, N$ expression levels y_i for $j = 1, \ldots, J$ tissues. The total sample size $N = \sum_j N_j$ with N_j denoting the number of replications for tissue j which means that i is nested in j. So, for example, for $j = 1$, $i = 1, \ldots, N_1$, for $j = 2$, $i = N_1 + 1, \ldots, N_1 + N_2$. The model for the expression level is

$$y_i = \mu_1 d_{1i} + \cdots + \mu_J d_{Ji} + \epsilon_i, \tag{1}$$

where μ_j denotes the population mean of tissue j, d_{ji} is 1 if the expression was obtained for tissue j and 0 otherwise and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. The density of the data for this ANOVA model is

$$f(\boldsymbol{y} \mid \boldsymbol{d}_1, \ldots, \boldsymbol{d}_J, \mu_1, \ldots, \mu_J, \sigma^2) = \prod_j \prod_{i \in j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{1}{2} \frac{(y_i - \mu_1 d_{1i} - \cdots - \mu_J d_{Ji})^2}{\sigma^2}, \tag{2}$$

where $\boldsymbol{y} = [y_1, \ldots, y_N]$ and $\boldsymbol{d}_j = [d_{j1}, \ldots, d_{jN}]$.

The goal is to determine the support in the data for two hypotheses:

$$H_1 : \mu_1 > \{\mu_2, \ldots, \mu_J\}, \tag{3}$$

which states that μ_1 is larger than each of the means in the set $\{\mu_2, \ldots, \mu_J\}$ (the gene is tissue-selective), and

$$H_2 : \text{not } H_1. \tag{4}$$

Note that H_2 corresponds to H_0 of the IUT. Support in the data will be quantified using the Bayes factor (Kass and Raftery, 1995). Using Chib's approach (Chib, 1995) the Bayes factor of H_1 versus H_2 can be written as:

$$\mathrm{BF}_{12} = \frac{f(\boldsymbol{y} \mid \boldsymbol{\theta}, \sigma^2) h(\boldsymbol{\theta}, \sigma^2 \mid H_1) / g(\boldsymbol{\theta}, \sigma^2 \mid \boldsymbol{y}, H_1)}{f(\boldsymbol{y} \mid \boldsymbol{\theta}, \sigma^2) h(\boldsymbol{\theta}, \sigma^2 \mid H_2) / g(\boldsymbol{\theta}, \sigma^2 \mid \boldsymbol{y}, H_2)} \tag{5}$$

where $\boldsymbol{\theta} = [\mu_1, \ldots, \mu_J]$, $h(\cdot)$ denotes the prior distribution of the parameters for the hypothesis indicated and $g(\cdot)$ the posterior distribution. The dependence of $f(\cdot)$ and $g(\cdot)$ on $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_J$ is left implicit. Note that $\mathrm{BF}_{12} = 6$ implies that the support in the data for H_1 is six times as large as the support in the data for H_2.

It is convenient to write

$$\mathrm{BF}_{12} = \frac{\mathrm{BF}_{1u}}{\mathrm{BF}_{2u}}, \tag{6}$$

where u refers to $H_u : \mu_1, \ldots, \mu_J$, that is, an unconstrained model. Once a prior distribution has been specified for H_u the prior distribution of H_m for $m = 1, 2$ is obtained via

$$h(\boldsymbol{\theta}, \sigma^2 \mid H_m) = \frac{h(\boldsymbol{\theta}, \sigma^2 \mid H_u) I_{\boldsymbol{\theta} \in m}}{\int_{\boldsymbol{\theta}, \sigma^2} h(\boldsymbol{\theta}, \sigma^2 \mid H_u) I_{\boldsymbol{\theta} \in m} d\boldsymbol{\theta} d\sigma^2} = \frac{1}{c_m h(\boldsymbol{\theta}, \sigma^2 \mid H_u) I_{\boldsymbol{\theta} \in m}}, \tag{7}$$

where $I_{\boldsymbol{\theta} \in m} = 1$ if $\boldsymbol{\theta}$ is in agreement with the constraints of model m and 0 otherwise, and c_m is the proportion of H_u in agreement with H_m.

The prior distribution of H_u is

$$h(\boldsymbol{\theta}, \sigma^2 \mid H_u) = \prod_j \mathcal{N}(\mu_j \mid \mu_0, \tau_0) \text{Inv-}\chi^2(\sigma^2 \mid \nu, \sigma_0^2), \tag{8}$$

that is, the prior distribution of each μ_j is the same. With this specification, and independent of the choice of μ_0, τ_0, ν_0 and σ_0^2, $c_1 = 1/J$ and

$c_2=(J-1)/J$, which makes sense because there are J equivalent models in which one of the means is larger than the other means. Using (7) and (8), BF_{1u} can, for any value of $\boldsymbol{\theta}$ in agreement with H_1, be written as:

$$\begin{aligned}\mathrm{BF}_{1u}&=\frac{f(\boldsymbol{y}\,|\,\boldsymbol{\theta},\sigma^2)h(\boldsymbol{\theta},\sigma^2\,|\,H_1)/g(\boldsymbol{\theta},\sigma^2\,|\,\boldsymbol{y},H_1)}{f(\boldsymbol{y}\,|\,\boldsymbol{\theta},\sigma^2)h(\boldsymbol{\theta},\sigma^2\,|\,H_u)/g(\boldsymbol{\theta},\sigma^2\,|\,\boldsymbol{y},H_u)}\\&=\frac{Jh(\boldsymbol{\theta},\sigma^2\,|\,H_u)/((1/f_1)g(\boldsymbol{\theta},\sigma^2\,|\,\boldsymbol{y},H_u))}{h(\boldsymbol{\theta},\sigma^2\,|\,H_u)/g(\boldsymbol{\theta},\sigma^2\,|\,\boldsymbol{y},H_u)}\\&=Jf_1,\end{aligned}\tag{9}$$

where f_1 is the proportion of the unconstrained posterior distribution in agreement with the constraints of H_1 because:

$$\begin{aligned}g(\boldsymbol{\theta},\sigma^2\,|\,\boldsymbol{y},H_m)&=\frac{g(\boldsymbol{\theta},\sigma^2\,|\,\boldsymbol{y},H_u)I_{\boldsymbol{\theta}\in m}}{\int_{\boldsymbol{\theta}}g(\boldsymbol{\theta},\sigma^2\,|\,\boldsymbol{y},H_u)I_{\boldsymbol{\theta}\in m}d\boldsymbol{\theta}}\\&=\frac{1}{f_m g(\boldsymbol{\theta},\sigma^2\,|\,\boldsymbol{y},H_u)I_{\boldsymbol{\theta}\in m}}.\end{aligned}\tag{10}$$

Using a similar derivation for BF_{2u}, it is obtained that

$$\mathrm{BF}_{12}=\frac{\mathrm{BF}_{1u}}{\mathrm{BF}_{2u}}=\frac{Jf_1}{J/(J-1)f_2}.\tag{11}$$

See Klugkist and Hoijtink (2007) for a more elaborate discussion of the derivation of the Bayes factor for inequality constrained hypotheses in the context of ANOVA. As will be illustrated in Section 3.1, if the number of means smaller than the target mean increases, the Bayes factor increases in favor of the hypothesis of selective overexpression and if the number of means larger than the target mean increases, the Bayes factor increases in favor of the complement of the hypothesis of selective overexpression. Furthermore, note the following property of the Bayes factor. Consider the situation where all means have the same value. Then the expected value of $f_1=1/J$ and $f_2=(J-1)/J$, that is, $BF_{12}=1$. Stated otherwise, if all means are equal the Bayes factor is neutral with respect to the hypothesis of interest and its complement. According to the Bayes factor one or more means being equal to the mean of the target tissue is neither evidence in favor nor against the hypothesis of selective overexpression.

The only question remaining is the estimation of f_1 because $f_2=1-f_1$. Using a very large (but finite) number for τ_0, any number for μ_0, $\nu=-2$ and $\sigma_0^2=0$, that is very uninformative priors, the following algorithm renders a sample from the unconstrained posterior distribution of $\mu_1,\ldots,\mu_J,\sigma^2$ and an estimate of f_1:

- Step 1: assign initial values: $\mu_j=\bar{y}_j$ for $j=1,\ldots,J$ and $\sigma^2=1/N\sum_i(y_i-\bar{y}_1d_{1i}-\cdots-\bar{y}_Jd_{Ji})^2)$, where $\bar{y}_j$ denotes the sample average for tissue j.
- Step 2: for $j=1,\ldots,J$ sample μ_j from $g(\mu_j\,|\,\sigma^2,\boldsymbol{y})$ which is a normal distribution with mean $\bar{y}_j$ and variance σ^2/N_j.
- Step 3: verify whether or not the current values of $\mu_1,\ldots,\mu_J$ are in agreement with the constraints of H_1.
- Step 4: sample σ^2 from $g(\sigma^2\,|\,\mu_1,\ldots,\mu_J,\boldsymbol{y})$ which is a scaled inverse chi-square distribution with degrees of freedom $N-2$ and scale parameter $1/(N-2)\sum_i(y_i-\mu_1d_{1i}-\cdots-\mu_Jd_{Ji})^2$.
- Iterate Steps 2 through 4. The proportion of vectors $\mu_1,\ldots,\mu_J$ sampled in Step 3 in agreement with the constraints of H_1 is an estimate of f_1.

Note that this algorithm has two favorable properties. First, our approach is objective in the following sense: (i) due to the use of vague prior distributions, the posterior is proportional to the likelihood, that is, f_1 and f_2 are completely determined by the data, and (ii) since the prior is the same for each mean, c_1 and c_2 do not depend on the prior. Second, no burn-in is needed because convergence is almost instantaneously as we sample from an inverse normal chi-square distribution and because we initialize the algorithm with the sample average and sample variance. In applications, we used 5000 iterations in the Gibbs sampler. Testing for selective underexpression can be done by adapting the constraints in H_1; on the algorithmic level this only influences Step 3. An efficient implementation of this algorithm in MATLAB or R can be found online (http://ppw.kuleuven.be/okp/software/BayesianIUT/).

3 RESULTS

We compared the performance of the IUT and Bayesian procedure in finding tissue-selective overexpressed genes using simulated and real data. Real data were obtained after robust multichip analysis (RMA) preprocessing and a $\log_2$ transformation. Because the simulation can be used as a reference for the empirical data, several parameters in the simulation were chosen the same as for these data. These are the number of tissues (22), the number of replications per tissue (three to five), the total number of expression values for a gene (70), the SD per tissue calculated over the replications ($s=0.04$, corresponding to the median SD in the data) and the overall mean expression level ($\mu=7.3$).

3.1 Simulation

Here, we compare the performance of the two testing procedures in a controlled setting using simulated data. Two influencing factors are of interest, namely the amount of support in the data and the degree to which the expression differs between the target and other tissues (the so-called effect size; Cohen, 1969). First, for the amount of support four levels will be considered: (i) complete support by all 21 tissues to which the target is compared; (ii) one tissue that is neutral, that is, has the same mean as the target tissue, and 20 supporting; (iii) one tissue not supporting and 20 supporting; and (iv) none of the 21 tissues supporting. Second, the effect size δ is manipulated at two levels: one with a considerable overlap of the sampling distributions, $\delta=0.5\,s=0.02$, and one with a small overlap $\delta=2\,s=0.08$. The data are generated from a normal distribution with SD $s=0.04$ and mean equal to 7.3 for the target, to $7.3-\delta$ for a tissue supporting the hypothesis and to $7.3+\delta$ for a tissue not supporting the hypothesis. Note that the effect sizes used may seem too small for what can be expected from expression data. However, using effect sizes larger than approximately three leads to an almost complete separation of the sampling distributions such that performance of both the IUT and Bayesian approach would be almost perfect.

The results for the simulation are summarized in Table 1 that reports the proportion of genes for which the data support that they are selectively overexpressed in the target tissue. The panels correspond to different decision rules for tissue-selective overexpression: in the first panel, this is a significant result for the IUT; in the second panel, this is a Bayes factor larger than one (more support for H_1 than for H_2); and in the third panel this is for a Bayes factor larger than 32 (support for H_1 is 32 times larger than the support for H_2). The rows correspond to the four different amounts of support, while the two columns of each panel correspond to the two levels of effect size. Histograms of the log transformed Bayes factor[1] are shown in Figure 1. These were used to determine the cutoff of 32 (corresponding to a $\log_2$ value equal to 5).

Clearly, the IUT is conservative. Even in the most favorable case (complete support, large effect size), only 26% of the genes are declared to be tissue-selective,[2] corresponding to a false negative rate of 74%. Compare this with the corresponding results for the Bayesian procedure (middle panel of Table 1): the majority of the Bayes factors indicate that H_1 is supported, even when the effect size

[1]Bayes factors below 0.0001 were set equal to 0.0001, while Bayes factors over 10 000 were set equal to 10 000.

[2]This number drops to 0 when accounting for the multiple testing associated to the testing of 5000 genes (results not shown).

Table 1. Proportion of tissue-selective overexpressed genes

	IUT		BF > 1		BF > 32	
	$\delta = 0.5s$	$\delta = 2s$	$\delta = 0.5s$	$\delta = 2s$	$\delta = 0.5s$	$\delta = 2s$
Complete support of H_1	0.002	0.257	0.588	0.996	0.024	0.689
One tissue neutral w.r.t. H_1	0.002	0.043	0.568	0.941	0.020	0.301
One tissue not supporting H_1	0	0	0.5	0.238	0.014	0.002
No support at all of H_1	0	0	0.09	0	0	0

Different panels correspond to different decision rules: a gene is declared tissue-selectively over-expressed (hypothesis H_1) in the first panel, if the IUT rejects H_0 against H_1; in the second panel, if the Bayes factor >1; and in the third panel, if the Bayes factor >32. The different rows correspond to different amounts of support of H_1, while the different columns of each panel correspond to different effect sizes.

is small. As can be seen in the third row of Table 1, this more liberal character of the Bayesian approach leads to many false positive results, but can be solved by requiring that the Bayes factor should be >32. In this case, as shown in the right panel of Table 1, the false positive rate is close to zero yet the false negative rate is much smaller than for the IUT. The second row of Table 1 presents the situation where one tissue has the same expression level as the target, and the other expression levels are smaller than the expression level of the target tissue. As elaborated in the previous section for the Bayesian approach, expression levels equal to the expression level of the target tissue are neither evidence in favor nor against the hypothesis of interest. This is reflected by an increased rate of genes detected as selectively overexpressed for both the IUT and Bayes factor. Taking 32 as a demarcation value for the Bayes factor yields 30% of the genes detected as selectively overexpressed. In case that the pair of tissues cannot be considered as a functionally equivalent group, these are false positives and avoiding them can be solved by taking a much higher Bayes factor (e.g. $\log_2 = 10$, see Fig. 1B) as a demarcation value. Then, also the Bayesian procedure becomes conservative. Note that adding more tissues with an expression level equal to the target tissue, will quickly reduce the proportion supported to almost zero (e.g. with two tissues having the same expression as the target, the 0.30 drops to 0.16). As illustrated in Table 1, this holds also for a lower effect size ($\delta = 0.5s$) for the tissues that support the hypothesis H_1. The fourth row shows that the performance of both the IUT and the Bayes factor is rather good if there is no support at all for the hypothesis of interest.

To illustrate that the Bayesian procedure is more informative than reporting the largest P-value of the partial tests involved in the IUT, we simulated data for which one tissue has the same mean as the target tissue (7.3). The means of the remaining tissues steadily varied from 7.3 to $7.3 - 3s = 7.3 - 0.12$ (corresponding to an effect size that varies from 0 to 3), this is increasingly supporting the hypothesis of selective overexpression in the target tissue. For each effect size, 50 replicates (genes) were generated. A scatterplot for the IUT is depicted in Figure 2A, where the largest P-value of the partial tests is plotted against the effect size for the IUT; a scatterplot for the Bayesian procedure is depicted in Figure 2B where the log of the Bayes factor is plotted against this same effect size. Clearly, the Bayesian procedure detects that the support of the 20 tissues increases, while this information is not captured by the IUT.

Advantages of the IUT over the Bayesian approach are that it is less computationally intensive and has a clear cutoff. First, running the simulation discussed above on an Intel Core Duo took less than a second for the IUT and less than 10 min for the Bayesian approach (with 5000 iterations): this is for all eight conditions and with 5000 replications per condition (in practice, this corresponds to eight analyses of 5000 genes). Note that although the Bayesian procedure is much slower, the time required for the analysis of a large dataset using an ordinary desktop is still very reasonable. Second, the conventional use of 0.01 or 0.05 as a cutoff for significance

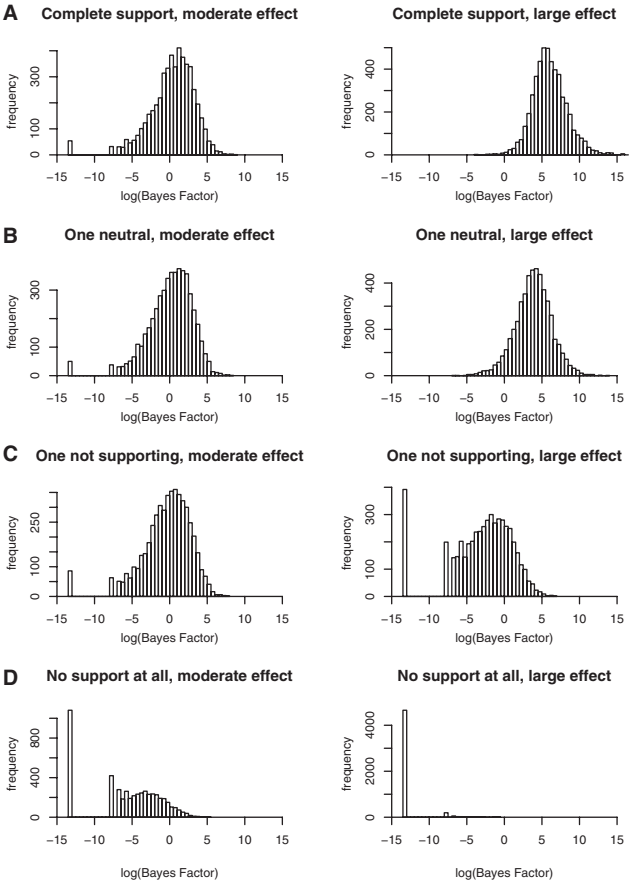

Fig. 1. Histograms of the $\log_2$ transformed Bayes factors. (**A**) Complete support of $H1$. (**B**) One tissue neutral with respect to H_1. (**C**) One tissue not supporting H_1. (**D**) None of the tissues supporting H_1. Panels at the left are obtained with a considerable overlap of the sampling distributions, at the right with a small overlap.

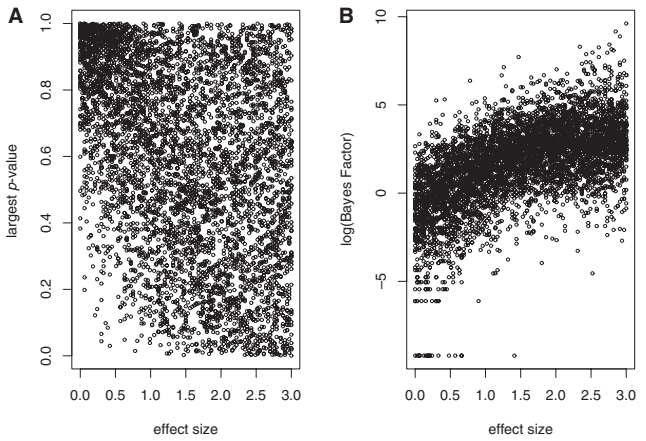

Fig. 2. Scatter plots for the largest *P*-value of the IUT (**A**) and of the $\log_2$ Bayes factors (**B**) in function of the effect size.

gives a clear rule. For the Bayes factor, there is not such a clear rule: Choosing a cutoff that yields a good balance between the number of true and false positives (this is not too conservative nor too liberal) depends on the number of tissues, their number of replicates and their effect sizes. Therefore, a simulation study should be performed prior to the analysis of the intended data; as an aid for potential users, we provided the script file of our simulation.

3.2 Tissue-selective genes

We used a publicly available microarray dataset (http://www.ncbi.nlm.nih.gov/geo/, accession number GSE9954; see also Thorrez *et al.*, 2008) that we generated via Affymetrix mRNA expression analysis using 430 2.0 arrays. This database consists of 22 different murine tissues, with 3–5 replicates for each tissue. Note that in the Bayesian procedure, the replications are supposed to follow a normal distribution; therefore, we took the $\log_2$ of the RMA preprocessed expression data.

For each tissue, the Bayesian procedure was used to find genes preferentially expressed in a particular tissue. We decided that a gene is tissue-selective upregulated when the Bayes factor was >32 on all probesets matching the gene. Genes that are tissue-selective are expected to be associated with the cellular processes which are the characteristic for the tissue. To assess whether this was the case for the gene sets identified by the Bayesian procedure, we tested functional overrepresentation of these genes using Ingenuity Pathway Analysis 7.1. The five most significant functions and diseases per tissue are listed in Table 2. It is apparent that most of these clearly reflect tissue-specific functions. This indicates that the underlying gene sets truly capture the tissue selectivity. For seminal vesicle only two functions reached significance and for salivary gland no significant function was found. Probably, this is due to the fact that these two tissues are studied by few researchers, leading to few publications on which Ingenuity Pathway Analysis can base its results.

It is important to realize that the obtained results depend strongly on the panel of tissues considered. Including more tissues will lead to a smaller number of genes denoted to be tissue-selective by the Bayesian procedure, but the biological specificity for these genes will be higher. For example, the panel used here contains three contractile tissues: gastrocnemius, diaphragm and heart; especially gastrocnemius and diaphragm have rather similar functions. Therefore, the number of tissue-selective genes is small (e.g. for gastrocnemius only 132 genes were found; see Table 2) and their functions are very specific (e.g. quantity of skeletal muscle associated to gastrocnemius and rib formation to diaphragm; see Table 2).

4 DISCUSSION

Finding tissue-selective genes is a recurring biological theme. As shown, the use of procedures that correct for multiple testing associated to the comparison of all target tissue–other tissue pairs is erroneous. A correct statistical procedure is the IUT. However, it is conservative and limited in information. As an alternative, we proposed a Bayesian procedure. In a simulation study, it was shown that this procedure is in most situations less conservative, while still keeping the number of false positives acceptable. Also, it is more informative than the IUT because it expresses how strongly the complete expression profile supports the hypothesis of tissue selectivity. Note that although we discussed the case of selective overexpression, both methods can also be used to find selective underexpressed genes.

The results of both the IUT and Bayesian procedure are highly dependent on the panel of tissues considered. These methods denote a gene as preferentially expressed in a particular tissue when expression in the tissue is higher than in each of the other tissues considered in the panel. Small panels can be expected to lead to large lists of tissue-selective genes containing many genes that are even more selectively overexpressed in a tissue not included in the panel. In an application of the Bayesian procedure to a panel of 22 tissues, we illustrated this by including two tissues with rather similar functionality, namely gastrocnemius and diaphragm. To avoid biologically flawed results, the proposed Bayesian method as well as the IUT should be applied with careful consideration of the tissues to include in the panel.

The use of the Bayesian procedure and the IUT is not limited to finding tissue-selective genes in normal tissues. Any problem involving the comparison of a reference group to each of the other groups (more than two) can be envisaged. Interesting applications are the comparison of normal tissue (persons) to several types of diseased tissue (persons). For example, Nishimura *et al.* (2007) aimed at genes susceptible of autism by comparing normal persons to each of two groups of autistic persons.

An advantage of the Bayesian approach is that it is flexible in the kind of hypotheses that can be tested with it. Not only can it be easily adapted to test the hypothesis H_1 of selective expression in a few (and not a single) tissue against the complement H_2: not H_1, but also against more specific alternative hypotheses H_2. For example, it may be of interest to know whether the data support that a particular gene is upregulated in a very specific tissue belonging to a group of related tissues (e.g. three neuronal tissues) rather than that it is upregulated overall in this group of related tissues . The procedure that was proposed here is general and can thus be easily adapted to test such hypotheses. On the other hand, the IUT always tests against H_2: not H_1. A more challenging adaptation of the Bayesian procedure would be the possibility to include equalities in the hypothesis (e.g. to find categorical tissue-specific genes) in a way that the procedure is efficient enough to deal with thousands of genes

Table 2. For each tissue, the genes identified by the Bayesian procedure to be tissue-selective were analyzed for functional overrepresentation

Tissue	Nr BF	Nr IPA	Top-5 functions enriched in gene set	Significance	Tissue	Nr BF	Nr IPA	Top-5 functions enriched in gene set	Significance
Gastrocnemius	132	77	Contraction of muscle	1.86E-04	Thymus	401	272	Developmental process of blood cells	2.05E-24
			Quantity of skeletal muscle	5.03E-03				Proliferation of lymphocytes	2.05E-24
			Disease of muscle	7.56E-03				Developmental process of leukocytes	1.19E-22
			Assembly of thin filaments	9.99E-03				Proliferation of T lymphocytes	7.89E-22
			Slow-channel congenital myasthenic syndrome	9.99E-03				Quantity of leukocytes	1.12E-19
Spleen	356	216	Immune response	2.28E-36	Small intestine	643	436	Cell death of colorectal cancer cell lines	1.34E-03
			Proliferation of lymphocytes	3.16E-30				Metabolism of nucleic acid component or derivative	1.34E-03
			Proliferation of leukocytes	3.16E-30				Cleavage of protein	1.34E-03
			Activation of leukocytes	3.60E-27				Transport of lipid	1.34E-03
			Activation of lymphocytes	3.71E-25				Infection of mammalia	1.51E-03
Liver	513	386	Metabolism of amino acids	7.21E-22	Eye	422	263	Vision of organism	8.93E-57
			Metabolic disorder	7.15E-18				Opthalmic disorder	1.94E-39
			Cholestasis	3.30E-15				Retinal degeneration	2.63E-33
			Metabolism of lipid	2.41E-14				Spinocerebellar ataxia, type 7	2.63E-33
			Hepatic system disorder	2.41E-14				Retinitis pigmentosa	2.43E-17
Brain	807	548	Neurological disorder	1.91E-25	ES cells	1212	862	Mitosis	1.79E-12
			Neurotransmission	1.91E-25				Processing of rRNA	4.90E-10
			Schizophrenia	3.17E-21				Modification of DNA	1.24E-09
			Huntington's disease	1.81E-17				Repair of DNA	8.13E-08
			Cognition	5.57E-13				Splicing of RNA	8.16E-08
Lung	239	175	Migration of cells	2.50E-09	Placenta	798	508	Cell movement	1.01E-07
			Development of blood vessel	4.34E-07				Development of blood vessel	1.57E-05
			Development of lung	1.79E-05				Morphogenesis of cells	2.55E-05
			Survival of rodents	3.99E-05				Growth of cells	6.09E-05
			Respiratory disorder	7.87E-05				Adhesion of eukaryotic cells	8.31E-05
Kidney	343	220	Renal and urological disorder	2.07E-07	Ovary	530	218	Reproductive system disorder	5.42E-10
			Bartter's syndrome	9.97E-06				Quantity of ovarian follicle	1.02E-07
			Transport of anion	2.53E-05				Ovulation	3.52E-06
			Metabolism of acyl-coenzyme A	4.61E-05				Development of ovary	5.14E-06
			Transport of phosphoric acid	1.23E-04				Ovarian failure	5.95E-06
Adrenal gland	160	112	Biosynthesis of steroid	7.38E-04	Fetus	383	252	Dupuytren contracture	3.45E-08
			Congenital adrenal hyperplasia	3.08E-03				Burn	4.83E-08
			Synthesis of hormone	3.57E-03				Development of connective tissue	1.75E-05
			Mitochondrial DNA depletion syndrome	3.57E-03				Development of skeleton	3.54E-05
			Blood pressure of organism	3.57E-03				Condensation of cartilage tissue	1.91E-04
Bone marrow	277	184	Immune response	1.08E-15	Testis	2398	969	Spermatogenesis	6.56E-27
			Degranulation of eukaryotic cells	3.76E-14				Gametogenesis	5.00E-26
			Arthritis	1.94E-13				Development of germ cells	1.18E-12
			Severe acute respiratory syndrome	1.94E-13				Development of spermatids	1.23E-12
			Inflammatory response	5.14E-12				Fertilization	2.00E-10
Adipose	273	202	Quantity of fatty acid	1.14E-11	Pituitary gland	352	195	Formation of cyclic AMP	2.91E-04
			Quantity of lipid	1.52E-10				Development of pituitary gland	4.84E-04
			Synthesis of triacylglycerol	1.88E-08				Processing of hormone	6.04E-04
			Uptake of carbohydrate	1.80E-06				Cyclopia of mice	6.23E-04
			Accumulation of triacylglycerol	2.54E-06				Release of cyclic AMP	7.69E-04
Diaphragm	248	166	Contraction of muscle	7.18E-12	Heart	844	344	Development of cardiac muscle	2.09E-07
			Development of muscle	1.57E-04				Cardiac contractility of heart	5.54E-05
			Myopathy	1.96E-03				Contraction of muscle	1.22E-04
			Glycolysis	1.14E-02				Metabolism of ATP	1.28E-04
			Formation of rib	1.36E-02				Cardiomyopathy	1.52E-04
Seminal vesicle	532	264	Glycosylation of amino acids	3.46E-02	Salivary gland	638	259	No significant associations	
			Development of baculum	3.46E-02					

From left to right, the different columns contain the tissue label (Tissue); the number of genes found to be tissue-selective (Nr BF); the number mapped to the pathway (Nr IPA); their five most significant functions and diseases (Top-5 functions enriched in gene set); and the significance of the functions and diseases obtained with the Benjamini–Hochberg corrected Fisher's exact test (Significance).

[see Klugkist and Hoijtink (2007) for a discussion of Bayes factors for equality and inequality constrained hypotheses]. Inclusion of such equality constraints for the IUT are discussed by Tuke *et al.* (2009) (an equivalence testing approach is used).

ACKNOWLEDGEMENTS

We wish to thank two anonymous reviewers for helpful comments.

Funding: the Research Fund of Katholieke Universiteit Leuven (SymBioSys: CoE EF/05/007 and GOA/2005/04).

Conflict of Interest: none declared.

REFERENCES

Allison,D.B. *et al.* (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat. Rev. Genet.*, **7**, 55–65.

Berger,R.L. and Hsu,J.C. (1996) Bioequivalence trials, intersecion-union tests, and equivalence confidence sets. *Stat. Sci.*, **11**, 283–319.

Berger,R.L. (1982) Multiparameter hypothesis testing and acceptance sampling. *Technometrics*, **24**, 295–300.

Chib,S. (1995) Marginal likelihood from Gibbs output. *J. Am. Stat. Assoc.*, **90**, 1313–1321.

Cohen,J. (1969) *Statistical Power Analysis for the Behavioral Sciences*. Academic Press, New York.

Deng,X. *et al.* (2008) Improving the power for detecting overlapping genes from multiple DNA microarray-derived gene lists. *BMC Bioinformatics*, **9** (Suppl. 6), S14.

Dezso,Z. *et al.* (2008) A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biol.*, **6**, 49.

Dunnett,C.W. (1955) A multiple comparison procedure for comparing several treatments with a control. *J. Am. Stat. Assoc.*, **50**, 1096–1121.

Greller,L.D. and Tobin,F.L. (1999) Detecting selective expression of genes and proteins. *Genome Res.*, **9**, 282–296.

Kadota,K. *et al.* (2003) Detection of genes with tissue-specific expression patterns using Akaike's information criterion procedure. *Physiol. Genomics*, **12**, 251–259.

Kadota,K. *et al.* (2006) ROKU: a novel method for identification of tissue-specific genes. *BMC Bioinformatics*, **7**, 294.

Kass,R. and Raftery,A. (1995) Bayes factors. *J. Am. Stat. Assoc.*, **90**, 773–795.

Klee,E.W. (2008) Data mining for biomarker development: a review of tissue specificity analysis. *Clin. Lab. Med.*, **28**, 127–143.

Klugkist,I. and Hoijtink,H. (2007) The Bayes factor for inequality and about equality constrained models. *Comput. Stat. Data Anal.*, **51**, 6367–6379.

Liang,S. *et al.* (2006) Detecting and profiling tissue-selective genes. *Physiol. Genomics*, **26**, 158–162.

Liu,X. *et al.* (2008) TIGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics*, **9**, 271.

Nishimura,Y. *et al.* (2007) Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. *Hum. Mol. Genet.*, **16**, 1682–1698.

Schug,J. *et al.* (2005) Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.*, **6**, R33.

Skrabanek,L. and Campagne,F. (2001) Tissueinfo: high-throughput identification of tissue expression profiles and specificity. *Nucleic Acids Res.*, **29**, E102.

Su,A.I. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.

Thorrez,L. *et al.* (2008) Using ribosomal protein genes as reference: a tale of caution. *PLoS ONE*, **3**, e1854.

Tuke,J. *et al.* (2009) Gene profiling for determining pluripotent genes in a time course microarray experiment. *Biostatistics*, **10**, 80–93.