



ELSEVIER

Contents lists available at ScienceDirect

MethodsX

journal homepage: [www.elsevier.com/locate/mex](http://www.elsevier.com/locate/mex)

## Method Article

## Methods for multiple outcome meta-analysis of gene-expression data

Konstantina E. Vennou<sup>a</sup>, Daniele Piovani<sup>b,c</sup>, Panagiota I. Kontou<sup>a</sup>,  
Stefanos Bonovas<sup>b,c</sup>, Pantelis G. Bagos<sup>a,\*</sup><sup>a</sup> Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia 35131, Greece<sup>b</sup> Department of Biomedical Sciences, Humanitas University, Milan, Italy<sup>c</sup> IBD Center, Humanitas Clinical and Research Center - IRCCS, Milan, Italy

## A B S T R A C T

Meta-analysis is a valuable tool for the synthesis of evidence across a wide range study types including high-throughput experiments such as genome-wide association studies (GWAS) and gene expression studies. There are situations though, in which we have multiple outcomes or multiple treatments, in which the multivariate meta-analysis framework which performs a joint modeling of the different quantities of interest may offer important advantages, such as increasing statistical power and allowing performing global tests. In this work we adapted the multivariate meta-analysis method and applied it in gene expression data. With this method we can test for pleiotropic effects, that is, for genes that influence both outcomes or discover genes that have a change in expression not detectable in the univariate method. We tested this method on data regarding inflammatory bowel disease (IBD), with its two main forms, Crohn's disease (CD) and Ulcerative colitis (UC), sharing many clinical manifestations, but differing in the location and extent of inflammation and in complications. The Stata code is given in the Appendix and it is available at: [www.compgen.org/tools/multivariate-microarrays](http://www.compgen.org/tools/multivariate-microarrays).

- Multivariate meta-analysis method for gene expression data.
- Discover genes with pleiotropic effects.
- Differentially Expressed Genes (DEGs) identification in complex traits.

© 2020 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

## A R T I C L E I N F O

*Method name:* Multivariate meta-analysis*Keywords:* Multiple outcome, Pleiotropic effects, Microarrays, Meta-analysis*Article history:* Received 2 November 2019; Accepted 16 February 2020; Available online 21 February 2020DOI of original article: [10.1016/j.ygeno.2019.09.019](https://doi.org/10.1016/j.ygeno.2019.09.019)

\* Corresponding author.

*E-mail addresses:* [vennou\\_k@yahoo.gr](mailto:vennou_k@yahoo.gr) (K.E. Vennou), [dpiovani@hotmail.com](mailto:dpiovani@hotmail.com) (D. Piovani), [pkontou@dib.uth.gr](mailto:pkontou@dib.uth.gr) (P.I. Kontou), [sbonovas@gmail.com](mailto:sbonovas@gmail.com) (S. Bonovas), [pbagos@compgen.org](mailto:pbagos@compgen.org) (P.G. Bagos).<https://doi.org/10.1016/j.mex.2020.100834>

2215-0161/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license.

(<http://creativecommons.org/licenses/by/4.0/>)

## Specification Table

Subject Area:	Biochemistry, Genetics and Molecular Biology
More specific subject area:	Bioinformatics
Method name:	Multivariate meta-analysis
Name and reference of original method:	This is a co-submission. The original method is: <a href="https://doi.org/10.1016/j.ygeno.2019.09.019">https://doi.org/10.1016/j.ygeno.2019.09.019</a>
Resource availability:	The Stata code is available at <a href="http://www.compgen.org/tools/multivariate-microarrays">www.compgen.org/tools/multivariate-microarrays</a> and in supplementary material

## Method details

In gene expression experiments, the mRNA levels of thousands of genes are measured simultaneously in tissue samples. After the initial pre-processing steps that may differ among methods (normalization etc.), the identification of differentially expressed genes (DEGs<sup>1</sup>; genes that are differentially expressed, that is, up- or down-regulated, under different conditions) is a procedure usually reduced to a statistical test for the equality of means between groups (e.g. *t*-test). Concerning meta-analysis of microarray and RNAseq studies there are three broad categories of statistical methods that make use of effect sizes, p-values or ranks. For several reasons, the effect size approach which is based on common practices in meta-analysis, is preferable (in particular for the problem at hand) and thus it was advocated early in the literature [1]. However, to handle the problem of small sample size and non-normal data, most authors suggest a type of correction for calculating the statistical significance.

There are situations though, in which we have a comparison of more than two groups. In general, we may encounter multiple outcomes, multiple risk factors or multiple treatments. For instance, in a meta-analysis of microarray studies, there are cases in which different medications (treatments) are administered or different exposures (risk factors) are recorded. In another potential example several (related) diseases are investigated against a common healthy group (multiple outcomes). Such examples include diseases of similar etiology like autoimmune diseases (Multiple Sclerosis, Rheumatoid Arthritis and Type I Diabetes Mellitus), or diseases with similar manifestations, like Crohn's disease (CD<sup>2</sup>) and Ulcerative colitis (UC<sup>3</sup>). In all cases, we are usually interested in finding differentially expressed genes common in all conditions, or genes that differ among the conditions. However, the estimates obtained are not independent since a common comparison group is included, and the usual methods for univariate meta-analysis are not adequate, especially when it comes to comparing the estimates. The above-mentioned problems are dealt with in the framework of multivariate meta-analysis, in which we are interested in performing a joint modeling of the different quantities of interest [2–4]. The multivariate meta-analysis enables the researcher to perform global tests in order to avoid the inflation of type I error rate due to multiple comparisons. Additionally, one can estimate the covariance matrix for the parameters of interest, which could be of importance when one wants to perform a comparison for the equality of the two estimates and construct a confidence interval for the difference or some other function of the estimates [5]. Finally, a major advantage of multivariate meta-analysis is that it can accommodate, under the missing at random (MAR<sup>4</sup>) assumption, studies reporting only one of the parameters of interest, resulting in the so-called borrowing strength from external studies [6]. In all cases, applying a multivariate methodology and pooling data from multiple sources may prove important in increasing the power of the analysis and allowing, under a unified analysis, the identification of genes that play a central role in these conditions.

<sup>1</sup> DEGs - Differentially Expressed Genes.

<sup>2</sup> CD - Crohn's Disease.

<sup>3</sup> UC - Ulcerative Colitis.

<sup>4</sup> MAR - Missing at random.

We used Stata<sup>5</sup> [7], a general-purpose statistical software package widely used in biostatistics, epidemiology and econometrics. Stata allows user-written commands and there is an active community of scientists supporting the development of new methods. We implemented the multivariate meta-analysis method for gene expression data and we applied it to data regarding inflammatory bowel disease (IBD<sup>6</sup>), with its two main forms, CD and UC, sharing many clinical manifestations, but differing in the location and extent of inflammation and in complications. The application of the method and the results of the meta-analysis are discussed in detail in [8].

The model we use is based on the standard model of multivariate meta-analysis [2–4]. We are interested in comparing the mean difference of a continuous outcome ( $X$ ), measured in study  $i = 1, 2, \dots, k$ . The effect sizes for the comparison of the two different disorders, are the standardized mean differences, which is referred to as Cohen's  $d$  [9]. It measures the difference in gene expression in CD and UC patients against controls (healthy subjects), respectively, and it will be estimated by:

$$\hat{d}_{1i} = \frac{\bar{X}_{1i} - \bar{X}_{0i}}{S_{pooled,i}} \text{ and } \hat{d}_{2i} = \frac{\bar{X}_{2i} - \bar{X}_{0i}}{S_{pooled,i}} \tag{1}$$

Where  $X_{1i}$ ,  $X_{2i}$  and  $X_{0i}$  are the means of the expression of the particular gene in CD, UC and control groups respectively.  $S_{pooled,i}$  is the pooled standard deviation, in study  $i$ , given by:

$$S_{pooled,i} = \sqrt{\frac{(n_{0i} - 1)S_{0i}^2 + (n_{1i} - 1)S_{1i}^2 + (n_{2i} - 1)S_{2i}^2}{n_{0i} + n_{1i} + n_{2i} - 3}} \tag{2}$$

with  $n_{1i}$ ,  $n_{2i}$  and  $n_{0i}$  being the sample size of each group and  $S_{1i}$ ,  $S_{2i}$  and  $S_{0i}$  the gene expression standard deviation for each group.

The variance of the effect size estimates  $d_{1i}$  and  $d_{2i}$  will be estimated by:

$$\widehat{var}(\hat{d}_{1i}) = s_{1i}^2 = \frac{1}{n_{1i}} + \frac{1}{n_{0i}} + \frac{\hat{d}_{1i}^2}{2n_i}, \quad \widehat{var}(\hat{d}_{2i}) = s_{2i}^2 = \frac{1}{n_{2i}} + \frac{1}{n_{0i}} + \frac{\hat{d}_{2i}^2}{2n_i} \tag{3}$$

and the covariance between the estimates of  $d_{1i}$  and  $d_{2i}$  will be

$$\widehat{cov}(\hat{d}_{1i}, \hat{d}_{2i}) = \frac{1}{n_{0i}} + \frac{\hat{d}_{1i}\hat{d}_{2i}}{2n_i} \tag{4}$$

with  $n_i = n_{0i} + n_{1i} + n_{2i}$  being the total sample size of the study [10]. The calculation of the within-studies covariance is very important since it plays a crucial role in the multivariate method.

Several statistical methods have been proposed to handle the issues of small sample sizes efficiently and calculate accurate  $p$ -values and confidence intervals. These methods include resampling methods such as bootstrap and permutation, Bayesian methods and penalized or regularized  $t$ -tests. The bootstrap and permutation are easily applied but would require enormous computational time. Regularized  $t$ -tests on the other hand, require additional modifications for the multivariate setting. Here, we resort to an old but reliable alternative that has been shown to perform well, at least for moderate sample sizes (i.e.  $\sim 5$ – $10$  per condition). The idea is based on the fact that the sample estimates of the standardized mean difference  $d$  has the tendency to overestimate the absolute value in small samples. This bias introduced by  $d$  can be corrected using the so-called Hedges'  $g$ , which generates an unbiased estimate. A correction factor, called  $J$ , is employed to convert from  $d$  to Hedges'  $g$ .

$$\hat{g}_{1i} = J(v_i)\hat{d}_{1i}, \quad \hat{g}_{2i} = J(v_i)\hat{d}_{2i} \tag{5}$$

Although approximate formulae exist, the exact formula for  $J$  is given by [11]:

$$J(v_i) = \frac{\Gamma\left(\frac{v_i}{2}\right)}{\sqrt{\frac{v_i}{2}}\Gamma\left(\frac{v_i-1}{2}\right)} \tag{6}$$

<sup>5</sup> Stata - Statistical software package created by StataCorp.

<sup>6</sup> IBD - Inflammatory Bowel Disease.

where  $v_i = n_{0i} + n_{1i} + n_{2i} - 3$ . When  $g$  is used, the variances will also change to:

$$\hat{v}ar(\hat{g}_{1i}) = (J(v_i))^2 \hat{v}ar(\hat{d}_{1i}), \quad \hat{v}ar(\hat{g}_{2i}) = (J(v_i))^2 \hat{v}ar(\hat{d}_{2i}) \quad (7)$$

Similarly,

$$\hat{c}ov(\hat{g}_{1i}, \hat{g}_{2i}) = (J(v_i))^2 \hat{c}ov(\hat{d}_{1i}, \hat{d}_{2i}) \quad (8)$$

In the multivariate random-effects meta-analysis, we assume that  $\mathbf{g}_i = (g_{1i}, g_{2i})$  is distributed following a multivariate normal distribution around the true means, according to the marginal model:

$$\mathbf{g}_i \sim MVN(\mathbf{g}, \mathbf{C} + \Sigma_i) \quad (9)$$

By  $\Sigma_i$  we denote the within-studies covariance matrix:

$$\Sigma_i = \begin{bmatrix} s_{1i}^2 & \rho_W s_{2i} s_{1i} \\ \rho_W s_{2i} s_{1i} & s_{2i}^2 \end{bmatrix} \quad (10)$$

The diagonal elements of  $\Sigma_i$  are the study-specific estimates of the variance (Eq. (3)), whereas the off-diagonal elements correspond to the pairwise within-studies covariances (Eq. (4)), for instance  $cov(g_{1i}, g_{2i}) = \rho_W s_{1i} s_{2i}$ . It is important to note that the elements of  $\Sigma_i$  are considered known quantities. In contrast, by  $\mathbf{C}$  we denote the between-studies covariance matrix, which is to be estimated during the fitting process:

$$\mathbf{C} = \begin{bmatrix} \tau_1^2 & \rho_B \tau_1 \tau_2 \\ \rho_B \tau_1 \tau_2 & \tau_2^2 \end{bmatrix} \quad (11)$$

A sometimes useful simplification can be made in Eq. (11) if we assume that the between-studies variances are equal [6]. In such case by letting  $\tau = \tau_1 = \tau_2$ ,  $\mathbf{C}$  reduces to:

$$\mathbf{C} = \begin{bmatrix} \tau^2 & \rho \tau^2 \\ \rho \tau^2 & \tau^2 \end{bmatrix} \quad (12)$$

This is particularly useful in comparison of several treatments and in cases where the direct comparisons are scarce, since with this parameterization the total number of estimated parameters is reduced. In any case, the final marginal model on which we base the inference is:

$$\begin{bmatrix} g_{1i} \\ g_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}, \begin{bmatrix} s_{1i}^2 + \tau_1^2 & \rho_W s_{1i} s_{2i} + \rho_B \tau_1 \tau_2 \\ \rho_W s_{1i} s_{2i} + \rho_B \tau_1 \tau_2 & s_{2i}^2 + \tau_2^2 \end{bmatrix} \right\} \quad (13)$$

This model can be fitted in any statistical package capable of fitting random-effects weighted regression models with an arbitrary covariance matrix, such as SAS<sup>7</sup>, R<sup>8</sup>, or Stata. Stata code for fitting the model is presented in the Appendix using `mvmeta` which performs inferences based on either Maximum Likelihood (ML<sup>9</sup>), the Restricted Maximum Likelihood (REML<sup>10</sup>) or the multivariate method of moments (MM<sup>11</sup>), which is the method of choice due to its non-iterative nature [12].

As we already mentioned, a major advantage of this model is that it can accommodate, under the MAR assumption, studies reporting only one of the parameters of interest, resulting in borrowing strength from external studies [6]. Another advantage is that, with the estimated variance-covariance matrix, it can be used to perform global tests for the effect sizes. For instance, we can test the null hypothesis that both estimates of  $g_1$  and  $g_2$  are equal to zero ( $H_0: g_1 = 0, g_2 = 0$ ). The test statistic,

<sup>7</sup> SAS - Statistical Analysis System, a statistical software developed by SAS Institute.

<sup>8</sup> R - Programming language and free software environment supported by the R Foundation for Statistical Computing.

<sup>9</sup> ML - Maximum Likelihood.

<sup>10</sup> REML - Restricted Maximum Likelihood.

<sup>11</sup> MM - Method of Moments.

$W$ , follows a chi-square ( $\chi^2$ ) distribution on 2 degrees of freedom and it can be obtained using the delta method:

$$W = \begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \end{bmatrix}^T \begin{bmatrix} \hat{v}ar(\hat{g}_1) & \hat{c}ov(\hat{g}_1, \hat{g}_2) \\ \hat{c}ov(\hat{g}_1, \hat{g}_2) & \hat{v}ar(\hat{g}_2) \end{bmatrix}^{-1} \begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \end{bmatrix} \sim \chi_2^2 \quad (14)$$

With this statistic one can test for pleiotropic effects, that is, for genes that influence both outcomes. More importantly, the test may discover genes that have a change in expression that is not detectable in the univariate setting. A similar test has been proposed for joint estimation of gene main effects and gene-environment interactions in GWAS<sup>12</sup> [13] and by members of our team for identifying pleiotropic effects in bivariate meta-analysis of GWAS [14].

Another useful test, would be that of the equality of the estimates. In this case the null hypothesis ( $H_0$ ) would be:  $D = g_1 - g_2 = 0$  and it can be tested using the Wald test (or an equivalent chi-square test):

$$\frac{\hat{D}}{\sqrt{\hat{v}ar(\hat{D})}} = \frac{\hat{g}_1 - \hat{g}_2}{\sqrt{\hat{v}ar(\hat{g}_1 - \hat{g}_2)}} \sim N(0, 1) \quad (15)$$

The variance of  $D$ , would be obtained easily by the delta method:

$$\hat{v}ar(\hat{g}_1 - \hat{g}_2) = \hat{v}ar(\hat{g}_1) + \hat{v}ar(\hat{g}_2) - 2\hat{c}ov(\hat{g}_1, \hat{g}_2) \quad (16)$$

This test is similar to the test for the genetic model of inheritance in genetic association studies [15,16] and it can also, in principle, be used to perform a comparison of the two estimates and construct a confidence interval for the difference or some other function of the estimates.

As in every high-throughput method involving multiple testing, additional correction methods need to be applied afterwards to account for multiple-comparisons (multiple genes in this case) and various correction methods have been proposed for this task. These methods are grouped into two categories, the ones that control the family-wise error rate (FWER<sup>13</sup>) and the ones that control the False Discovery Rate (FDR<sup>14</sup>). The most common approach is to control the family-wise error rate using the Bonferroni [17] correction, which is easily applied and intuitive, but very conservative. Other popular methods used for multiple testing corrections are the methods proposed by Holm [18], Sidak [19] and Holland [20]. The most commonly used method in high-throughput experiments, however, is the method proposed by Benjamini and Hochberg, which instead of controlling FWER, controls the FDR [21]. FDR-controlling procedures are designed to control the expected proportion of rejected null hypotheses that were incorrect rejections. FDR-controlling procedures have greater power (i.e. they are able to detect more differences as statistically significant), at the cost of an increased Type I error rate. Stata implements easily most of the corrections mentioned above.

The method as presented here was applied to data regarding IBD, with its two main forms, CD and UC [8]. The datasets that were used in the multivariate meta-analysis contained data for 25,409 unique genes/loci from 251 CD patients, 175 UC patients and 159 healthy individuals, originating from 4 studies. The multivariate meta-analysis identified 249 statistically significant DEGs in patients with CD and 38 in patients with UC, at an FDR of 1%. The DEGs identified to be common to both diseases were 20. With the use of the global test ( $W$ ) we identified 260 statistically significant DEGs that rejected the null hypothesis that both  $g_1$ , and  $g_2$  are equal to zero. From the 260 statistically significantly DEGs, 53 were found to be statistically significant only by the global test and not in any of the diseases. These genes could not have been identified using univariate methods. The global test ( $D$ ) was used as well, but only two genes (TRIP6 and IGHG1) were found to have equal change in the expression levels. Afterwards, we used these DEGs for further investigation with the help of bioinformatics overrepresentation analysis (ORA<sup>15</sup>) methods, that identified significant pathways [8].

<sup>12</sup> GWAS - Genome-Wide Association Studies.

<sup>13</sup> FWER - Family-Wise Error Rate.

<sup>14</sup> FDR - False Discovery Rate.

<sup>15</sup> ORA - Overrepresentation analysis.

The method was applied to microarray data; however, it can easily be applied also to RNAseq data. As we already mentioned, a major difference of Microarrays and RNAseq lies in the pre-processing steps, but for the purposes of a meta-analysis the analyst will not have to deal with that; one will assume that the data are processed and normalized. In general, data from microarrays and RNAseq are analyzed using the same meta-analysis methods [22,23], and in some cases, they are even combined together [24]. Taking these into account, we can safely assume that the method will work easily with RNAseq data, except for the case of a very small number of replicates ( $n = 2$  or  $3$  per condition, which of course will be a problem also for microarrays). In this case, even though the correction method we used is reported to perform well, other more advanced options may be preferable (for instance to use a bootstrap correction).

To conclude, we have presented an application of the multivariate meta-analysis to gene expression data. The Stata code is given in the Appendix and it is available at [www.compgen.org/tools/multivariate-microarrays](http://www.compgen.org/tools/multivariate-microarrays). Contrary to common microarrays data formats, the code requires that samples are listed one per line, whereas the expression for the various genes (or probes) are given in columns. Moreover, numerical variables indicating the type of the sample (in the particular example: Healthy, UC, CD) and the actual study, are needed. It is the user's responsibility to ensure that the variables are coded appropriately. We need to mention that the method does not need all outcomes to be measured in every study. As we noted earlier, under the MAR assumption the method can incorporate studies that consider only  $g_1$ , only  $g_2$ , or both and in this case the increase in power is larger. However, for the proper estimation of between studies covariance, it is advised that a reasonable number of studies should report both outcomes. The code also works if a gene is present only in one study, but in this case the results are based only on the data provided by the particular study. Nevertheless, we also coded Eqs. (14) and (15) to report such results in order to prevent information loss. The code is provided as a .do file and in its current form reads a fixed-name .dta file and outputs the results in two separate files, one for genes for which meta-analysis was conducted and one for those that appear only in a single study. The user may change the file names accordingly. Moreover, the range of genes which will be analyzed are hardcoded, but this is easy to change. In its current form the code can handle two correlated outcomes but changing the inner loop in order to incorporate more outcomes is trivial. Extending the method to allow for more difficult settings, like network meta-analysis for mixed-treatment comparisons [25], will need additional modifications, especially when it comes to model specification.

## Funding

This work is part of Konstantina's Vennou thesis, which is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Strengthening Human Resources Research Potential via Doctorate Research" (MIS-5000432), implemented by the State Scholarships Foundation (IKY).

## Acknowledgments

The authors would like to thank the associate editor and the anonymous referees whose comments helped in improving the quality of the manuscript.

## Declaration of Competing Interest

The Authors confirm that there are no conflicts of interest.

## References

- [1] F. Hong, R. Breitling, A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments, *Bioinformatics* 24 (3) (2008) 374–382.
- [2] H.C. van Houwelingen, L.R. Arends, T. Stijnen, Advanced methods in meta-analysis: multivariate approach and meta-regression, *Stat. Med.* 21 (4) (2002) 589–624.

- [3] D. Mavridis, G. Salanti, A practical introduction to multivariate meta-analysis, *Stat. Methods Med. Res.* 22 (2) (2013) 133–158.
- [4] D. Jackson, R. Riley, I.R. White, Multivariate meta-analysis: potential and promise, *Stat. Med.* (2011).
- [5] D.G. Altman, J.M. Bland, Interaction revisited: the difference between two estimates, *BMJ* 326 (7382) (2003) 219.
- [6] J.P. Higgins, A. Whitehead, Borrowing strength from external trials in a meta-analysis, *Stat. Med.* 15 (24) (1996) 2733–2749.
- [7] StataCorp, *Stata 13 Base Reference Manual*, Stata Press, College Station, TX, 2013.
- [8] K.E. Vennou, et al., Multiple outcome meta-analysis of gene-expression data in inflammatory bowel disease, *Genomics* 112 (2) (2020) 1761–1767.
- [9] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn., Hillsdale, New Jersey, 1988 L.Erlbaum.
- [10] L.J. Gleser, I. Olkin, Stochastically Dependent Effect Sizes. *The Handbook of Research Synthesis and Meta-Analysis*, Russell Sage Foundation, 2009, pp. 357–376.
- [11] L.V. Hedges, Distribution theory for glass's estimator of effect size and related estimators, *J. Educ. Stat.* 6 (2) (1981) 107–128.
- [12] I.R. White, Multivariate random-effects meta-analysis, *Stata J.* 9 (2009) 40–56.
- [13] A.K. Manning, et al., Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP x environment regression coefficients, *Genet. Epidemiol.* 35 (1) (2011) 11–18.
- [14] C. Medina-Gomez, et al., Bivariate genome-wide association meta-analysis of pediatric musculoskeletal traits reveals pleiotropic effects at the SREBF1/TOM1L2 locus, *Nat. Commun.* 8 (1) (2017) 121.
- [15] P.G. Bagos, A unification of multivariate methods for meta-analysis of genetic association studies, *Stat. Appl. Genet. Mol. Biol.* 7 (1) (2008) Article 31.
- [16] P.G. Bagos, Genetic model selection in genome-wide association studies: robust methods and the use of meta-analysis, *Stat. Appl. Genet. Mol. Biol.* 12 (3) (2013) 285–308.
- [17] S. Dudoit, Y.H. Yang, M.J. Callow, T.P. Speed, Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments, *Statistica. Sinica.* 12 (1) (2000) 111–139.
- [18] S. Holm, A simple sequentially rejective multiple test procedure, *Scand. J. Stat.* 6 (1979) 65–70.
- [19] Z. Šidák, Rectangular confidence regions for the means of multivariate normal distributions, *J. Am. Stat. Assoc.* 62 (318) (1967) 626–633.
- [20] B. Holland, M. Copenhaver, An improved sequentially rejective bonferroni test procedure, *Biometrics* 43 (2) (1987) 417.
- [21] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B (Methodol.)* 57 (1) (1995) 289–300.
- [22] K. Tsuyuzaki, I. Nikaido, *metaSeq: Meta-Analysis of RNA-seq Count Data*, Tokyo University of Science, Tokyo, 2013.
- [23] A. Rau, G. Marot, F. Jaffrézic, Differential meta-analysis of RNA-seq data from multiple studies, *BMC Bioinform.* 15 (1) (2014) 91.
- [24] J. Jung, et al., Meta-analysis of microarray and RNA-Seq gene expression datasets for carcinogenic risk: an assessment of bisphenol a, *Mol. Cell. Toxicol.* 13 (2) (2017) 239–249.
- [25] G. Lu, A. Ades, Combination of direct and indirect evidence in mixed treatment comparisons, *Stat. Med.* 23 (20) (2004) 3105–3124.