

Reviewing and assessing existing meta-analysis models and tools

Funmilayo L. Makinde, Milaine SS. Tchamga, James Jafali, Segun Fatumo, Emile R. Chimusa, Nicola Mulder and Gaston K. Mazandu 

Corresponding author: Gaston K. Mazandu, Department of Pathology, Division of Human genetics, University of Cape Town, Health Sciences Campus Anzio Rd, Observatory, 7925, South Africa. E-mail: gaston.mazandu@uct.ac.za

Abstract

Over the past few years, meta-analysis has become popular among biomedical researchers for detecting biomarkers across multiple cohort studies with increased predictive power. Combining datasets from different sources increases sample size, thus overcoming the issue related to limited sample size from each individual study and boosting the predictive power. This leads to an increased likelihood of more accurately predicting differentially expressed genes/proteins or significant biomarkers underlying the biological condition of interest. Currently, several meta-analysis methods and tools exist, each having its own strengths and limitations. In this paper, we survey existing meta-analysis methods, and assess the performance of different methods based on results from different datasets as well as assessment from prior knowledge of each method. This provides a reference summary of meta-analysis models and tools, which helps to guide end-users on the choice of appropriate models or tools for given types of datasets and enables developers to consider current advances when planning the development of new meta-analysis models and more practical integrative tools.

Key words: meta-analysis; predictive power; sample size; data integration; cohort study; experimental study

Introduction

Recent advances in high-throughput technologies and computational scanning approaches have led to the availability of enormous amounts of publicly available clinical and biological datasets [1–4]. These datasets, however, are often produced by different research projects or sites, and in most cases, often have some limitations when trying to combine them, including increased heterogeneity, as samples type collected are generally not standardized across sites; limited sample size in each site,

which may lead to insufficient power for inferences; unbalanced sample sizes between cases and controls, which may yield high biases in the outcome and often inconclusive results; and also batch effects and noise introduced into datasets from the technical origin that do not reflect biological variation [5, 6]. Combining these datasets is advantageous to researchers in increasing statistical power to detect biological phenomena from studies where logistical considerations restrict sample size or in studies that require the sequential hybridization of arrays [5].

Funmilayo L. Makinde, PhD candidate in the Computational Biology Division at University of Cape Town in collaboration with the African Institute for Mathematical Sciences (AIMS) South Africa.

Milaine SS. Tchamga, Postdoctoral fellow in the Division of Human Genetics at University of Cape in collaboration with the African Institute for Mathematical Sciences (AIMS) South Africa.

James Jafali, Researcher at Pathogen Biology Research Group, Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Malawi.

Segun Fatumo, Assistant Professor at London School of Hygiene and Tropical Medicine, University of London, UK.

Emile R. Chimusa, Associate Professor at the Division of Human Genetics, Department of Pathology, University of Cape Town, South Africa.

Nicola Mulder, Professor and Head of the Computational Biology Division at University of Cape Town, South Africa.

Gaston K. Mazandu, Senior Lecturer position at the Division of Human Genetics, Department of Pathology at University of Cape Town, and Associate Researcher at the African Institute for Mathematical Sciences (AIMS), South Africa.

Submitted: 11 April 2021; Received (in revised form): 7 July 2021

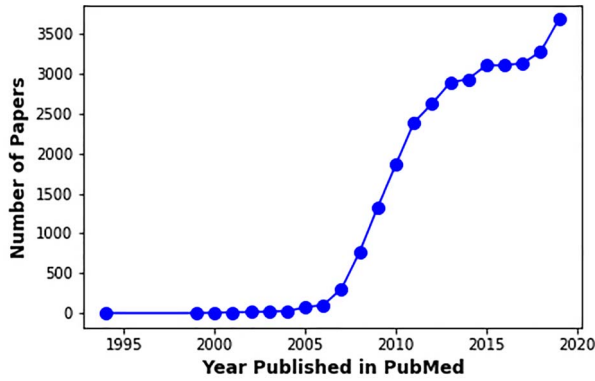


Figure 1. Number of publications on the application of meta-analysis to GWAS or gene expression datasets over the years. Search query used in PubMed search: [(meta-analysis[Title/Abstract]) AND (gene expression[Title/Abstract]) OR (genome wide association[Title/Abstract])].

Through funder or journal requirements, many of the large datasets which have accumulated from different studies are available in public data repositories and databases including Gene Expression Omnibus (GEO) @ <http://www.ncbi.nlm.nih.gov/geo/>, The Cancer Genome Atlas (TCGA) @ <http://cancergenome.nih.gov>, Array Express @ <http://www.ebi.ac.uk/arrayexpress/> and Sequence Read Archive (SRA) @ <http://www.ncbi.nlm.nih.gov/sra/>. There is potential value in combining related datasets for further analysis, but the challenges mentioned earlier limit this potential. To address this, effective statistical analyses have been widely suggested in biomedical research [7], one of which is meta-analysis. Meta-analysis is described as a statistical analysis of a large collection of analysis results from individuals to integrate the findings to increase statistical power and obtain more precise effect size estimates [8, 9]. In meta-analysis, two types of information integration are often considered: horizontal meta-analysis and vertical integrative analysis [10].

Horizontal meta-analysis, also known as multi-study data integration combines multiple genomic studies (e.g. multiple microarray, methylation or GWAS studies) to increase statistical power, accuracy and validation [10], while vertical integrative analysis combines multi-omics data (e.g. gene expression, CNV, genotyping, methylation, somatic mutation, miRNA and clinical variables) to investigate disease subtypes, disease associated or driving genes and related regulatory networks [11]. Over the years, there has been exponential growth in the application of meta-analysis to different types of datasets, most especially Genome Wide Association studies (GWAS) and microarray gene expression datasets (see Figure 1), which have generated promising results.

Several statistical methods have been proposed for performing meta-analysis [6, 7, 11–21]. This article comparatively reviews three different categories of these meta-analysis techniques: Combining P-values, Combining Effect Sizes and Combining Ranks. We applied these methods to datasets retrieved, assessed their pros and cons, and discussed each performance based on our result as well as assessment results from previous research. Finally, we briefly discuss available meta-analysis software such as METASOFT [22], MANTRA [11], GWAMA [23], METAL [24], rmeta [25], MetABEL [26], catmap [27], Metafor [28] and MetaOmics [29]. This provides information to help orient the implementation of new models and practical integrative tools, as well as enables researchers to choose appropriate models or tools for given types of datasets.

Meta-analysis methods

Meta-analysis techniques have been widely developed and applied in genomic applications, especially for combining multiple transcriptomic studies. The most commonly used meta-analysis methods in the analysis of clinical and transcriptomic datasets mainly aim to identify genes associated with an outcome in the case where the expected outcome is continuous, multi-class or survival censored, or to identify differentially expressed genes in the case where the expected outcome is binary [6, 29]. These meta-analysis methods used for combining such data are grouped into three categories, namely, combining P-values, combining effect sizes and combining ranks. Figure 2 shows a graphical representation of the most commonly used meta-analysis methods, indicating the time each method was developed, while Table 1 provides a summary of existing meta-analysis methods and their features. We discuss each category in detail below.

Combining P-values methods

Combining P-values is a popular approach used in meta-analysis. It provides simplicity and flexibility to different outcome variables including survival, multi-class and continuous outcomes. There are several methods of P-value combinations from independent statistical tests with each having different statistical properties [6, 29, 30]. Next, we provide a summary of some methods used in combining P-values across several studies.

Fisher’s method: Fisher developed this method in 1925 [8]. It is the most straightforward statistical approach used in combining P-values across studies to estimate the accuracy of a unified P-value [7]. It combines log-transformed P-values from independent datasets/studies. The Fisher’s statistical formula is given by

$$S_{\text{Fisher}} = -2 \sum_{j=1}^J \log(P_j)$$

where P_j is the P-value for the j^{th} study, and S_{Fisher} follows a chi-squared distribution under the null hypothesis (assuming null P-values are uniformly distributed), with $2J$ degrees of freedom, where J is the total number of studies [8]. One major constraint this method has is that all studies are weighted uniformly, and this is likely inappropriate when combining GWAS studies with different sample sizes [6]. Also, it is impossible to estimate the average magnitude of differential expression when working with P-values [31]. An implementation of this method is found in *metap* an R package for meta-analysis of significance values [32].

Stouffer method (Z-transform test): Stouffer introduced this method in 1949 [18]. Unlike the Fisher method, Stouffer’s method adopted a different approach by using a simple transformation to transform the P-values P_j from an individual study into a new variable say:

$$z_j^* = \Phi^{-1}(1 - P_j)$$

where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution which follows a standard normal $N(0,1)$ under the null hypothesis [18]. Stouffer’s statistical formula is given by

$$S_{\text{Stouffer}} = \frac{1}{\sqrt{J}} \sum_{j=1}^J z_j^*$$

which also follows a normal distribution to obtain its combined P-value. Advantages of this method include: (i) It is more robust

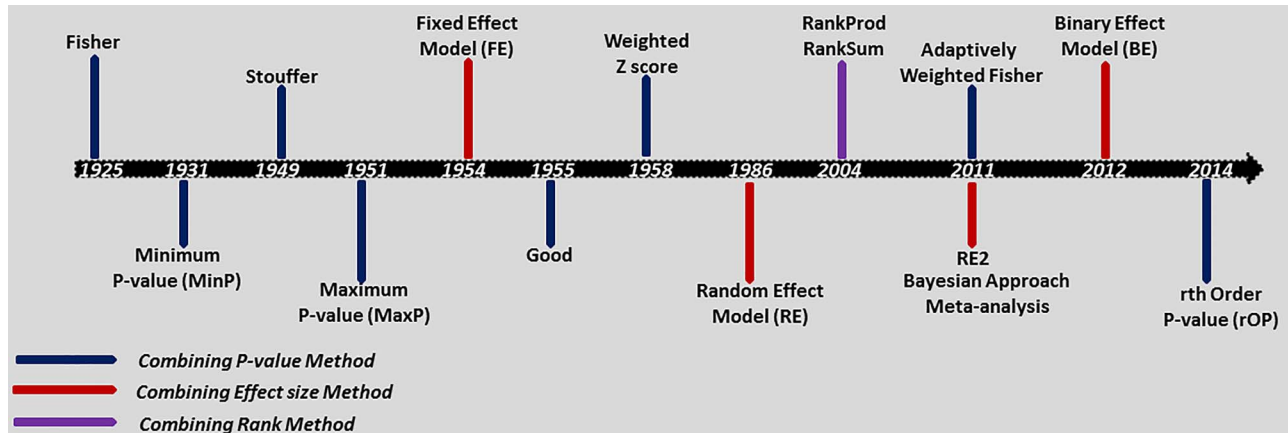


Figure 2. Timeline of existing meta-analysis methods showing the evolution of the meta-analysis model development. (Fisher [8], MinP [19] Stouffer [18], MaxP [21], FE [37], Good [13], weighted z-score [33], RE [10, 20], RankProd/Ranksum [15, 41], adaptive weighted Fisher [16], RE2 [22], BE [10] and roP [17]).

Table 1. Summary of existing meta-analysis methods—Yes means feature is present and No implies absence of feature in each model

Model	AWF	Baye	BE	FE	Fisher	MaxP	MinP	RankProd/ RankSum	RE	roP	SR/PR	Stouffer	WZscore
Effect size	no	yes	yes	yes	no	no	no	no	yes	no	yes	no	no
Fold change	no	no	no	no	no	no	no	yes	no	no	no	no	no
Heterogeneity estimate	no	no	no	no	no	no	no	no	yes	no	no	no	no
HS _A	no	no	no	no	no	yes	no	no	yes	no	yes	no	no
HS _B	yes	no	no	no	yes	no	yes	yes	no	yes	no	yes	no
HS _r	no	no	no	no	no	no	no	no	yes	yes	no	no	no
M-value	no	no	yes	no	no	no	no	no	no	no	no	no	no
MCMC	no	yes	no	no	no	no	no	no	no	no	no	no	no
P-value	yes	no	no	no	yes	yes	yes	no	no	yes	yes	yes	yes
Sample size	no	no	no	no	no	no	no	no	no	no	no	no	yes
Standard error	no	no	no	yes	no	no	no	no	yes	no	no	no	no
Weight	yes	no	yes	yes	no	no	no	no	yes	no	no	no	yes
Z-score	no	no	yes	yes	no	no	no	no	yes	no	no	no	yes

and has little loss of efficiency compared to combining effect sizes methods where estimated effect sizes and standard error are considered as input; (ii) It is useful in cases where effect size estimates cannot be retrieved [18].

Weighted Z-score method: This method is a major improvement over the Stouffer method and was introduced by Liptak in 1958 [33]. In this case, P-values are transformed to Z-scores in a one-to-one transformation and also incorporate weight. Unlike the Stouffer method, the weighted Z-score is more efficient and powerful by allowing different weights for different studies [7]. Its statistical formula is given by

$$S_{wz} = \frac{\sum_j z_j w_j}{\sqrt{\sum_j w_j^2}}$$

where

$$z_j = \Phi^{-1} \left(1 - \frac{P_j}{2} \right) * \text{sign}(\nabla_j)$$

with P_j the P-value for the j^{th} study, ∇_j , the direction of the effect for study j and weight $w_j = \sqrt{N_j}$ with N_j being the sample size of the j^{th} study [34]. In the case where a study has unequal numbers of cases and controls, Willer et al. [24] recommended that the

effective sample size should be calculated as

$$N_{\text{eff}} = \frac{4}{\left(\frac{1}{N_{\text{cases}}} + \frac{1}{N_{\text{ctrls}}} \right)}$$

Also S_{wz} follows a normal distribution to obtain its combined P-value. An implementation of this method is the METAL software which provides a fast and efficient meta-analysis of genome wide association scans.

Adaptively Weighted (AW) Fisher's method: This method is an extension of the Fisher method and was developed by Li and Tseng in 2011 [16]. The method includes adaptive weights for individual studies in order to characterize effective studies contributing to the meta-analysis for an improved biological interpretation and statistical power of the results. It searches for all possible weights that will be identified as the best adaptive weight of an individual study with the smallest derived P-value. The adaptive weighted statistic given by

$$S_g = pU \left(u_g \left(w_g^* \right) \right)$$

is the P -value of the minimum P -value among all possible weights, where the observed weighted statistics

$$u_g(w_g) = - \sum_{j=1}^J w_{gj} \log(P_{gj}),$$

and

$$w_g^* = \arg \min_{w_g \in W} U(u_g(w_g)),$$

is the optimal weight that gives an indication of which studies contribute to the statistical significance or differentially expressed evidence of the meta-analysis. The corresponding P -value of the AW-statistics S_g is calculated as

$$p(S_g) = \frac{\sum_{b=1}^B \sum_{g'=1}^G I\{S_{g'}^{(b)} \leq S_g\}}{B * G}$$

(See Supplementary File for details). One disadvantage of this method is its inadequacy for traditional meta-analysis in evidence-based medicine research or epidemiology, and it also introduces bias toward studies with consistent significant effects [16].

Minimum P -value method (minP): Tippet in 1931 developed the minP method to detect differentially expressed genes whenever a small P -value exists in any one of the J studies [19]. Its statistical formula is written as

$$S_{\min P} = \min_{1 \leq j \leq J} P_j.$$

It uses the minimum P -value among the J studies which follows a Beta(α, β) distribution with degrees of freedom $\alpha = 1$ and $\beta = J$ as the test statistic under the null hypothesis [19].

Maximum P -value method (maxP): The maxP method was developed by Wilkinson [21] in 1951 to target differentially expressed genes that have small P -values in all studies. Its statistical formula is written as

$$S_{\max P} = \max_{1 \leq j \leq J} P_j.$$

It takes the maximum P -value which follows a Beta(α, β) distribution with degrees of freedom $\alpha = J$ and $\beta = 1$ as the test statistic under the null hypothesis [21].

r th ordered P -value method (rOP): Song and Tseng [17] developed the rOP method in 2014. This method uses the r -th ordered P -values from J combined studies. This method is considered a strong form of maxP (where $J/2 \leq r \leq J$) to identify candidate markers differentially expressed in most studies. The minP and maxP methods are special cases of rOP [17]. It also follows a Beta(α, β) distribution with degrees of freedom $\alpha = r$ and $\beta = J - r + 1$ as the test statistic under the null hypothesis [17].

The implementation of all these methods mentioned above can be found in MetaDE [35] an R package for microarray meta-analysis for differentially expressed gene detection. A significant advantage of using the combining P -value approaches is the allowance for standardization of the associations from genomic studies to a common scale. Compared to combining effect sizes, its simplicity and extensibility to different kinds of outcome variables is also a major advantage. Also, where the effect size may not be well defined, the association P -values can

still be calculated when the outcome variable is not binary [6]. Disadvantages of combining P -value include failure to provide an overall estimate of effect sizes, inaccessibility of between-study heterogeneity and it may be erroneous when there is inconsistency in the direction of effects of the combined studies [34]. The combining P -value methods are primarily applied to gene expression datasets [16, 31].

Combining effect sizes methods

The effect size combination method uses a two-group comparison to derive standardized effect sizes and its associated standard error across combined studies. Choi et al. [36] was among the first to apply the two most popular approaches namely fixed effects model (FE) and random effects model (RE) to microarray meta-analysis. In any given study of measuring differentially expressed genes, effect sizes are calculated using a standardized mean difference, that is,

$$d_j = \frac{T_j^* - C_j^*}{S_p}$$

where T_j^* and C_j^* are the treatment and control group of the j^{th} study, respectively, and S_p is the estimated standard deviation [31, 36]. Other effect size measures can be of log odds ratios or regression coefficients of the J independent studies [22]. We briefly illustrate the models used in combining effect sizes from multiple studies (See details in Supplementary File).

Fixed effect model (FE): Cochran in 1954, in his paper the combination of estimates from different experiments developed the fixed effect model (FE) [37], which assumes that all the studies under consideration share a common/fixed true effect size [10, 20]. Let T_j be the effect size estimate of each study which follows a normal distribution with mean μ and variance σ^2 . Also, let the weight assigned to each study $w_j = v_j^{-1}$ be the inverse of the variance, where v_j is the within-study variance for study j in a meta-analysis of J studies. The test statistic for the FE model is given by

$$S_{FE} = \frac{\sum_{j=1}^J w_j T_j}{\sqrt{\sum_{j=1}^J w_j}}$$

which follows $N(0,1)$ under the null hypothesis that there is no association. If we assume a one-tailed test, the P -value of the association is given by

$$p = 1 - \Phi(S_{FE})$$

while for a two-tailed test,

$$p = 2[1 - \Phi(|S_{FE}|)]$$

where Φ is the standard normal cumulative distribution function [10, 20]. According to Evangelou et al. [34], FE is the best approach for prioritizing and identifying phenotype associated SNPs when combining GWAS data [34]. A major advantage FE has over the random effect model (RE) is its ability to maximize discovery power [34, 38]. This model is mostly applied to GWAS datasets [10, 20].

Random effect model (RE): Unlike the FE model, the RE model assumes that the effect sizes for each study in the meta-analysis are different and are drawn from a normal distribution with

mean μ and variance σ^2 [10, 20]. The RE analysis approach is to decompose the observed variance into its two component parts, i.e. the within-study variance v_j and between-study variance τ^2 . DerSimonian and Laird [11] in 1986 proposed a method of moments estimator for τ^2 using Cochran's Q test statistics (See Supplementary File for details). The test statistics for RE model is given by

$$S_{\text{RE}} = \frac{\sum_{j=1}^J w_j^* T_j}{\sqrt{\sum_{j=1}^J w_j^*}}$$

where $w_j^* = v_j^{*-1}$, $v_j^* = v_j + \tau^2$ with v_j and τ^2 the within-study and between-study variance values, respectively. Similarly, the P-value is

$$p = 1 - \Phi(S_{\text{RE}})$$

for one-tailed test, and

$$p = 2[1 - \Phi(|S_{\text{RE}}|)]$$

for a two-tailed test, where Φ is the standard normal cumulative distribution function. In comparison to FE, RE has limited power in discovery effort. However, RE is better off when considering the generalized observed association, uncertainty across different studies and estimating average effect size of the associated variant [34, 38, 39]. This model is mostly applied to GWAS datasets [10, 20]. Furthermore, Han and Eskin [22] made improvements on the RE (see ref [22] for details). They developed RE2 by assuming no heterogeneity under the null hypothesis, which contradicts the former RE. This was done in order to improve discovery power in the presence of between-study heterogeneity in the effect sizes [22, 34]. This model is mostly applied to GWAS datasets [10, 20].

Binary effect model (BE): BE is a new type of random effect model of meta-analysis developed by Han and Eskin in 2012 [10]. This model captures studies with or without an effect together. This model is the weighted sum of z-scores method where the m-values, that is, the posterior probability that an effect exists in each study of a meta-analysis, are incorporated into the weights. It assigns more weight to studies predicted to have an effect and lesser weight to the studies predicted not to have an effect. Let $z_j = \frac{T_j}{\sqrt{v_j}}$ be the z-score of the jth study. The binary effect model statistic is given as

$$S_{\text{BE}} = \frac{\sum_{j=1}^J m_j \sqrt{w_j} z_j}{\sqrt{\sum_{j=1}^J m_j^2 w_j}},$$

where the weight $\sqrt{w_j} \approx \sqrt{Np(1-p)}$ with N the sample size and p the effect size (minor allele) frequency while $\sqrt{w_j} \approx N$ when the effect size is the same between studies. m_j is the corresponding m-value of study j (see Appendix for details). This model is mostly applied to GWAS datasets [10].

Bayesian meta-analysis: A Bayesian approach to meta-analysis was developed in 2011 by Morris [40] to specifically perform trans-ethnic meta-analysis [11]. It assumes that a given population cluster with the same ethnic group may share the same effect size, but there is a difference in effect sizes among different population clusters. It assume the observed effect size of the jth study $b_j \sim N(\beta_j, s_j)$ where s_j is the corresponding standard error and β_j is the population-specific effect for the jth population cluster. Then let M_0 be the null hypothesis of no association and M_1 the alternative hypothesis in a Bayesian

framework. The evidence of association can be assessed by means of Baye's factor

$$\Lambda = \frac{f(b, s|M_1)}{f(b, s|M_0)}$$

where

$$f(b, s|M) = \int_{\theta} f(b, s|\theta) f(\theta|M) d\theta$$

is the marginal likelihood of the observed effect size under the model M with θ denoting the unknown model parameter which includes the population-specific effect β and hyper-parameters relating to prior distribution (see Supplementary File for details). This approach performs better compared to fixed-effects and random effect meta-analysis, especially in terms of power to detect association and localization of the causal variant over a range of models of heterogeneity between ethnic groups and also has increased power and mapping resolution when the similarity in allelic effects between populations is well captured by their relatedness [36]. A minor limitation might be that its implementation can be computationally intensive [34]. This model is also mostly applied to GWAS datasets [11].

Even though combining effect size methods are well known, they are powerful and also have a strong Gaussian assumption on effect size. They often fail in transcriptomic data due to their restriction to only two group comparisons [29].

Combining ranks methods

Combining rank statistics methods are used to avoid results which can be influenced by outliers, which methods such as combining effect sizes and P-values may overlook. What the rank-based method does is to calculate the ranks of differentially expressed evidence for each gene in each study instead of calculating their effect size or P-value. Methods used in combining ranks are discussed below.

RankProd (RP) & RankSum (RS) methods: Breitling et al. [41] introduced the rank product (RP) method, which aims to detect differentially expressed genes [15]. Suppose we have a total of G genes in differential expression data across J replicated experiments. Let $r_{i,j}$ be the position of the i th gene in the j th replicate experiment in a list ordered according to fold changes. The rank product (RP) statistics for the i th gene is defined as the geometric mean of all the rank of genes obtained in each replicate. That is,

$$RP_i = \left(\prod_{j=1}^J r_{i,j} \right)^{\frac{1}{J}}$$

Also, the rank sum (RS) statistics is defined as the arithmetic mean of all the ranks. That is,

$$RS_i = \frac{1}{J} \sum_{j=1}^J r_{i,j}$$

In the case where the datasets to be analyzed are unpaired datasets, the RP/RS is performed using the algorithm explained in Supplementary File [15, 41]. This method has advantages over linear models, including fewer assumptions and better robustness, and a biologically intuitive fold change (FC), thereby increasing power in low sample size and/or large noise settings [41].

Sum of ranks (SR): This method applies a naive sum of the differentially expressed evidence ranks across studies. The test statistic of SR is obtained as follows:

$$SR_g = \sum_{j=1}^J R_{gj}$$

where R_{gj} is the rank of P -value of gene g among all genes in the j th study [42, 43].

Product of ranks (PR): Similarly, this method applies a naive product of the differentially expressed evidence ranks across studies. The test statistic of PR is obtained as follows:

$$PR_g = \prod_{j=1}^J R_{gj}$$

where R_{gj} is the rank of P -value of gene g among all genes in the j th study [42, 43].

The P -values of these test statistics can be calculated analytically or obtained from a permutation analysis as shown in Supplementary File. Note that genes with the smallest RP or RS values are the most likely to be upregulated or downregulated based on your choice of order when ranking the fold changes. One major disadvantage is that combining ranks methods only consider gene ranks rather than absolute expression values, which often leads to its robustness against heterogeneity across different studies. The implementation of this method can be found in a Bioconductor package (RankProd) [15] and it is mostly applied to differentially expressed datasets.

Hypothesis setting

Considering a meta-analysis of J combined transcriptomic studies, with each study j containing G genes, it is essential in meta-analysis to choose an appropriate method suitable for any type of research based on the objectives of the studies. This is due to the influence that the correctly identified method has on the final result of the differential expression analysis. Thus, depending on the aim of identifying different types of targeted markers based on the biological question or objectives under consideration, researchers have developed three main hypothesis settings as a guide in identifying targeted differentially expressed markers using the right meta-analysis method [16, 17, 44]. Given the null hypothesis for each gene g :

$$H_0 : \bigcap_j \{\theta_{gj} = 0\},$$

where θ_{gj} is the effect size of gene g and study j with $1 \leq g \leq G$, $1 \leq j \leq J$ and following Birnbaum [45], Li & Tseng [16] and Song & Tseng's [17] convention, the complementary hypotheses settings are:

(1) HS_A for detecting targeted biomarkers that are differentially expressed in all studies or cohorts

$$HS_A : \left\{ H_0 \text{ vs } H_a^{(A)} : \bigcap_j \{\theta_{gj} \neq 0\} \right\}$$

Under the alternative hypothesis $H_a^{(A)}$, gene g is identified only when it is differentially expressed in all studies. It also implies that for gene g , the effect sizes of all J combined studies are nonzero. The suitable meta-analysis methods for HS_A are maxP

[16, 17, 44], sum of ranks [29], product of ranks [29] and RE model [16, 17].

(2) HS_B for detecting targeted biomarkers that are differentially expressed in one or more studies

$$HS_B : \left\{ H_0 \text{ vs } H_a^{(B)} : \bigcup_j \{\theta_{gj} \neq 0\} \right\}$$

Under the alternative hypothesis $H_a^{(B)}$, gene g is identified only if it is differentially expressed in one or more studies. It also implies that for gene g , the effect sizes of at least one of the J combined studies are nonzero. This is with the hope that an extremely small P -value in one study is usually enough to influence the meta-analysis and bring about statistical significance. Suitable meta-analysis methods for HS_B are Fisher [16, 17, 29, 44], Stouffer [16, 17, 29, 44], AW-Fisher [16, 17, 29, 44], minP [16, 17, 29, 44], rankprod and the FE model [29].

(3) HS_r for detecting targeted biomarkers that are differentially expressed in most of the studies

$$HS_r : \left\{ H_0 \text{ vs } H_a^{(r)} : \sum I_{\{\theta_{gj} \neq 0\}} \geq r \right\}$$

where r is pre-specified with $\frac{1}{2} \leq r \leq J$ and $I_{\{\cdot\}}$ is an indicator function. Under the alternative hypothesis $H_a^{(r)}$, gene g is identified only if it is differentially expressed in the majority of studies. It also implies that for gene g , the effect sizes of at least r with $\frac{1}{2} \leq r \leq J$ of the J combined studies are nonzero. The suitable meta-analysis methods for HS_r are rOP [17, 29, 44] and the RE model [29].

In meta-analysis applications, it is more appealing to detect differentially expressed markers in all studies, that is HS_A , however, in the case where we have large J combined studies, it is difficult to detect those markers in all the studies. This is because there is always additional noise/bias present in experimental data. Thus, most researchers prefer the hypothesis HS_r which detects differentially expressed markers in the majority of studies, i.e. $>70\%$ of the studies. HS_B is useful when considering statistically significant biomarkers in at least one study and when heterogeneity is expected [17].

Meta-analysis software

There are several existing tools that are used to carry out meta-analysis. Table 2 provides some existing meta-analysis software and their respective attributes with a brief explanation as follows.

METAL was initially released in 2008 and then later published in 2010 (<http://www.sph.umich.edu/csg/abecasis/metal/>) [24]. It is a fast and efficient computational tool for meta-analysis of large datasets such as genome wide association scans, and is the most widely used meta-analysis software package [7] which was designed to improve complex traits gene mapping studies. METAL was written in C++ and applies two main approaches in combining evidence for association from individual studies using appropriate weight. These approaches are (1) the weighted Z score approach as discussed above which is based on the sample size, P -value and direction of effect in each study, and (2) the effect-size based method weighted by the study-specific standard error. It requires effect size estimates and their standard errors to be in consistent units across studies.

Genome-Wide Association Meta-Analysis (GWAMA) (2010) [23] is an open-source software (<http://www.well.ox.ac.uk/GWA>

Table 2. Existing meta-analysis software and their attributes

Software	Graphical tool	Heterogeneity metric	Meta-analysis models	Preprocess software	Reference
catmap	NA	Q, I^2	FE, RE	NA	[27]
GWAMA	Manhattan, QQ plots	Q, I^2	FE, RE	SNPTEST, PLINK	[23]
MANTRA	NA	NA	Bayesian analysis	NA	[11]
MetABEL	Forest plot	Q, I^2	FE	ABEL	[26]
Metafor	Forest plot Funnel plot	Q, I^2	FE, RE, Mixed effect model	NA	[28]
METAL	NA	Q, I^2	Weighted Z-score	NA	[24]
MetaOmics	HeatMap	Q, I^2	MetaDE (Fisher, Stouffer, AWF, MinP, MaxP, roP, FE, RE, RankProd, RankSum, SR, PR)	MetaQC	[29]
METASOFT	Forest-PMPlot	Q, I^2	FE, RE, RE2, BE	NA	[22]
rmeta	Funnel plot	Q, I^2	FE, RE	NA	[25]

MA) designed for performing meta-analysis of summary statistics generated from genome-wide association studies of dichotomous phenotypes or quantitative traits. The software incorporates existing tools such as SNPTEST and PLINK to preprocess genome-wide association analysis files, aligns studies to the same reference strand irrespective of the genome-wide association genotyping product, and performs fixed effect meta-analysis, and in the presence of heterogeneity can perform random-effects meta-analysis. It calculates two heterogeneity measures (Cochran's Q statistic, I^2) of allelic effects between studies and provides graphical visualization of meta-analysis results using separate scripts for Manhattan and quantile-quantile (QQ) plots, and also accounts for automated genomic control for population structure.

METASOFT (<http://genetics.cs.ucla.edu/meta/>) is also an open-source software designed for performing a range of basic and advanced meta-analytic methods in an efficient manner. The methods incorporated in METASOFT include Fixed Effects model (FE) which is based on inverse-variance-weighted effect size as previously discussed, Random Effects model (RE) also based on inverse-variance-weighted effect size, Han and Eskin's Random Effects model (RE2) optimized to detect associations under heterogeneity and Binary Effects model (BE) optimized to detect associations when some studies do or do not have an effect. Similar to GWAMA, the heterogeneity measures used are Cochran's Q statistic and I^2 . It uses ForestPMPlot, a visualization tool to display results such as the P-value, study name, log odds ratio and its standard error and summary statistics for each study. It produces a computed m-value, that is, the posterior probability that the effect exists in each study and the PM-Plot visualizes the m-value of each study along with its P-value [22]. The MANTRA (Meta-ANalysis of Transethnic Association studies) software has been developed to implement two independent runs of the MCMC algorithm used in the Bayesian approach meta-analysis and also estimate the Baye's factor for each variant resulting in a summary output of the MCMC algorithm. This software is only available on request from the author [11].

MetaOmics is an analysis pipeline and browser-based software suite for transcriptomic meta-analysis. It is freely available at <https://github.com/metaOmics/metaOmics>. There are several R software packages and modules incorporated in MetaOmics. They include the *MetaPreprocess* module which allows users to input and preprocess multiple transcriptomic datasets for storage and later meta-analysis implementation of other analytical

modules, such as *MetaQC*, a microarray meta-analysis in quality control that provides a quantitative and objective tool to assist with study inclusion/exclusion criteria for meta-analysis, and *MetaDE*, a microarray meta-analysis which implements 12 major meta-analysis methods for differential expression (DE) analysis to identify candidate markers associated with disease outcome. Output results include differentially expressed gene lists with corresponding raw P-values, q-values and various visualization tools. Another module incorporated in this software is *MetaPath*, a microarray meta-analysis in pathway enrichment detection tool which is used to identify the pathways associated with disease outcome. Others are *MetaNetwork* used for detecting differential co-expression networks (DCN), *MetaPredict* for prediction analysis, *MetaClust* for disease subtype discovery and *MetaPCA* for dimension reduction and exploratory visualization [29, 35].

Some R software packages for meta-analysis include *rmeta* software which incorporates methods such as fixed and random effects meta-analysis models for two-sample comparisons and cumulative meta-analyses. It produces outputs which include standard summary plots, funnel plots and computes summaries and tests for association and heterogeneity [25]. *MetABEL* is an R software for meta-analysis of genome-wide association scans between quantitative or binary traits and SNPs [26] and *catmap* (*Case-Control and TDT Meta-Analysis Package*) is an R software package which performs meta-analyses on genetic case-control data by combining case-control and family-based (TDT) studies. It conducts fixed-effects (with inverse variance weighting) and random-effects meta-analyses on combined genetic data by specifically implementing a fixed-effects model and a random-effects model for combined studies [27]. Likewise, *Metafor* (*meta-analysis package for R*) an R package is made up of comprehensive functions used in meta-analysis. It includes functions to estimate effect sizes or outcome measures and fit fixed, random and mixed effect models to data, as well as performing meta-regression analyses. It outputs meta-analytic plots including forest and funnel plots [28].

Results

We used the GEO2R tool [46] to download four gene expression datasets each for malaria and breast cancer case studies from the Gene Expression Omnibus (GEO) repository. The GEO2R tool

Table 3. Results from assessing combining P-value meta-analysis methods using malaria study

Analysis	#Significant Genes	ACCURACY	F1SCORE	PRECISION	TPR
Individual-analysis					
GSE33811	562	0.4762	0.0571	0.3333	0.0313
GSE1124	3465	0.5333	0.3636	0.5263	0.2778
GSE7586	4739	0.5263	0.3077	0.5714	0.2105
GSE5418	6741	0.6000	0.4828	0.5833	0.4118
Meta-analysis					
AWFisher	4290	0.5500	0.4906	0.5652	0.4333
Fisher	2319	0.6000*	0.4783	0.6875	0.3667
MaxP	939	0.5167	0.1714	0.6000	0.1000
MinP	2319	0.6000*	0.4286	0.7500*	0.3000
rOP	1268	0.5333	0.2632	0.6250	0.1667
Stouffer	1903	0.5167	0.3256	0.5385	0.2333
Weighted Z-score	6648*	0.5833	0.6154*	0.5714	0.6667*

Table 4. Results from assessing combining P-value meta-analysis methods using breast cancer study

Analysis	# Significant Genes	ACCURACY	F1SCORE	TPR
Individual-analysis				
GSE37139	5142	0.5294	0.4783	0.4231
GSE7904	16 253	0.4648	0.4722	0.4474
GSE36295	6448	0.5800	0.5116	0.4231
GSE3744	10 240	0.5323	0.4314	0.3438
Meta-analysis				
AWFisher	10 087	0.4783	0.5862	0.7391
Fisher	7182	0.4783	0.5556	0.6522
MaxP	2995	0.5870	0.5128	0.4348
MinP	6926	0.4348	0.5000	0.5652
rOP	4288	0.6087*	0.5714	0.5217
Stouffer	6101	0.5000	0.5660	0.6522
Weighted Z-score	12 815*	0.5000	0.6462*	0.9130*

using the GEOquery and limma R packages from the Bioconductor project was also employed to analyze each of the dataset to produce P-values for individual differentially expressed genes. For the two case studies, results from individual analysis of the four datasets were combined using different combining P-value meta-analysis methods. Furthermore, the performance of each individual and meta-analysis results were assessed using evaluation metrics from a confusion matrix, that is, true positive rate (TPR), precision, accuracy and F1-score as shown in Table 3 and Table 4 for malaria and breast cancer case studies, respectively. Table 3 and Table 4 also include the total number of differentially expressed genes identified in both individual and meta-analysis by setting a threshold P-value of 0.05.

Assessing the methods using the malaria case study, we deduced that the well-known Fisher and MinP methods perform better in the combining P-value category with an accuracy of 60% as shown in Table 3. However, the story was not the same while assessing the methods using the breast cancer case study. As we can see in Table 4, the rOP method has the highest prediction accuracy of 60.87%, classifying it as the best method. However, we noticed that the weighted z-score method happens to perform well by having the highest total number of significance genes in both case studies. Not only that, but it also had the highest F1-score of 61.54 and 64.62% for both malaria and breast cancer studies, respectively. F1 score seems to be a good criterion to choose the best method when aiming for a positive interpretation of biological result. Therefore, we recommend the use of the weighted z-score method, followed

by Fisher and so on when considering meta-analysis of gene expression datasets in combining P-value categories. However, it is evident that there might be changes in this recommendation with dataset complexity [8], for example, the maxP model performance varies across different datasets. This is in agreement with ‘no free lunch’ theorems for optimization, which highlight the danger of comparing models by their performance on a single dataset [47], suggesting that no single model can produce the best performance across all datasets.

Assessment of combining effect size methods is based on prior knowledge from the literature. Han and Eskin [22], in their paper, did a comparison of fixed effect model (FE), random effect model (RE) and Han and Eskin’s Random Effect model (RE2). Several analyses were done to justify the performance of these methods (see ref [22] for details) and they concluded that the FE model performs best by constantly achieving the highest statistical significance at low heterogeneity, while the RE2 model performs best when high heterogeneity exists between studies. Also, it was clearly stated that the RE model does not produce a more significant P-value than FE as expected when there is heterogeneity between studies thereby forfeiting its main goal [22]. Based on this, we highly recommend the use of the FE model for meta-analysis of GWAS when there is little or no heterogeneity between studies and RE2 when high heterogeneity between studies is observed. Furthermore, Han and Eskin (2012) [14] also recommend the binary effect model (BE), a type of RE approach that considers whether an effect is present or not between studies, to produce increased numbers of identified associations

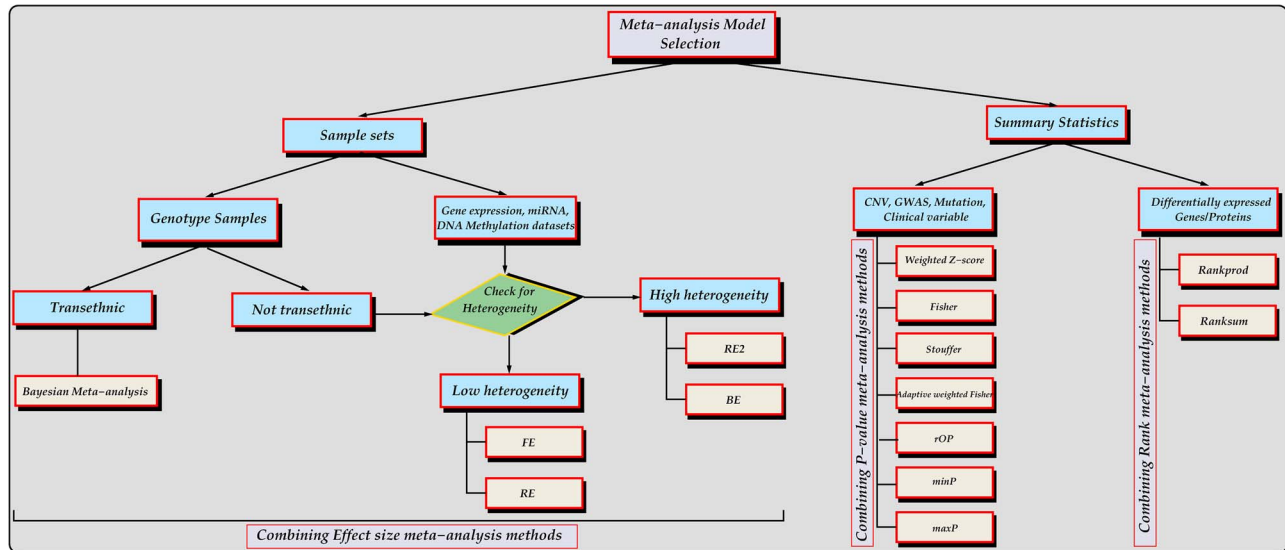


Figure 3. Decision tree of meta-analysis methods with respect to specific datasets. Method abbreviations: FE - Fixed effect, RE - Random effect, RE2 - Han and Eskin's random effect, BE - Binary effect, rOP - rth ordered P-value, minP - Minimum P-value, maxP - Maximum P-value.

showing heterogeneity [10]. However, Morris (2011) [40] have strongly recommended the use of Bayesian meta-analysis when considering meta-analysis of transethnic GWAS [11]. Combining rank methods are best used on differentially expressed datasets [15, 31, 41]. Hong et al. [31] suggest that preference be given to the RankProd/RankSum method when dealing with small sample size or large between-study variation. Based on these recommendations, Figure 3 provides a decision tree that guides users on the choice of methods with respect to specific dataset.

Discussion

We evaluated and compared the combining P-value meta-analysis methods, that is, Fisher, Stouffer, weighted z-score, adaptive weighted Fisher, rOP, minP and maxP methods, using two case studies. These case studies consist of eight publicly available gene expression datasets, including four Malaria and four breast cancer datasets. The reason why we chose two different case studies was to show how flexible these meta-analysis methods are, and how well their performance can be controlled under different levels of difficulty when combining multiple datasets from different platforms. The performance of these methods was assessed using evaluation metrics from a confusion matrix.

The four publicly available datasets on malaria used to evaluate meta-analysis methods include the Krupka et al. [48] dataset (GSE33811), which uses the *Affymetrix Human Gene 1.0 ST Array* platform with 10 samples of five Malawian patients with severe and mild malaria; samples from the *Affymetrix Human Genome U133A Array* platform of the Boldt et al. [49] dataset (GSE1124), with 25 samples of African children of which there are five asymptomatic *Plasmodium falciparum* infection (A); five uncomplicated malaria (U); five severe malarial anemia (A); five cerebral malaria (Ce) and five that are healthy. The Ockenhouse et al. [50] dataset (GSE5418), which also utilizes the *Affymetrix Human Genome U133A Array* platform with 71 samples from patients with naturally acquired malaria infection compared to those from volunteers in a challenge model vaccine trial; and the Muehlenbachs et al. [51] dataset (GSE7586), which uses the *Affymetrix Human Genome U133 Plus 2.0 Array* platform with

20 samples of women with 10 active and 10 negative placental malaria (PM) status.

Likewise, four publicly available datasets were used for breast cancer meta-analysis study. These datasets include the Ingles et al. [52] dataset (GSE37139), which uses the *Affymetrix Human Gene 1.0 ST Array* platform with 12 samples of MCF7 breast cancer cells; the Richardson et al. [53] datasets comprising 62 samples designed to compare expression of tumor to normal breast tissue on the *Affymetrix Human Genome U133 Plus 2.0 Array*, with 43 tumor and 19 normal breast samples (GSE7904) and another 47 samples of human breast tumor cases (GSE3744); the Merdad et al. [54] dataset (GSE36295), which utilizes the *Affymetrix Human Gene 1.0 ST Array* platform with 50 samples corresponding to 45 samples of breast cancer tissues and five samples of healthy breast tissues. Visualization of different datasets is represented using box plots (Figures in Supplementary File). Although the use cases stated above focus on human, mainly due to data availability, these meta-analysis methods can be applied to any organism, including plants [55].

Meta-analysis techniques have been widely used in several applications, especially in clinical and biomedical contexts, such as in experimental study, e.g. randomized controlled clinical trials (RCTs) [56, 57], as well as in observational studies for group comparisons (e.g. cohort and case-control studies) [58, 59]. As indicated previously, meta-analysis provides a systematic integrative and quantitative framework combining independent study outputs for the same biological question to draw more effective and accurate inferences, thus playing an essential role in evidence-based medicine [57]. In the context of RCTs, meta-analysis technique produces more precise estimate of medical treatment effect and effective health intervention strategies [56]. The two common meta-analysis biomedical use cases are the post-GWAS [60] and expression level pattern [61] analyses of multiple cohorts with the same or different phenotypes to identify contextual biological markers related to conditions under consideration with an increased statistical power [60, 61].

Though existing meta-analysis methods [10, 15, 22, 31, 41, 62–64] have been deployed in different applications, as highlighted above, there is still a limited ability to assess and compare the

performance of these different methods. Currently, a trial-and-error approach is applied to produce better, more significant results by fine-tuning features of the existing meta-analysis methods. So, there is a need to implement a meta-analysis integrative simulation framework, which assesses all existing methods to enable end users to select effective models for their applications. Furthermore, looking at the performance of different methods (see [Tables 3 and 4](#)), no meta-analysis method achieves a performance score higher than 75%, except for the weighted z-score method achieving approximately 91% true positive rate for the breast cancer dataset, likely at the cost of high false positives, considering the high number of significant genes predicted. Thus, there is a need for improvement on existing meta-analysis methods and tools by developing new models which account for the limitations of the existing models.

Conclusion

This article comparatively reviews several existing meta-analysis methods under three categories: Combining *P*-values, Combining Effect Sizes and Combining Ranks methods. We provide recommendations based on our results and findings while assessing the performance of different methods in our analysis and also from prior knowledge obtained from literature. We therefore recommend Combining *P*-values methods to researchers interested in carrying out meta-analysis of gene expression datasets. However, when performing GWAS meta-analysis, methods that fall under combining effect size category are best to use, while methods for combining rank category best suite meta-analysis of differentially expressed datasets.

Key Points

- Comprehensive summary and consistent classification of existing meta-analysis models and tools.
- There are three categories meta-analysis methods: Combining *P*-values, Combining Effect sizes and Combining Ranks.
- Assessing existing meta-analysis models and discussing the performance of different models.
- Revealing that effect size combination best fit sample sets, *P*-value combination best fit for sample data summary statistics and combining rank best suit differentially expressed genes.
- Guiding meta-analysis model and tool end-users based on data type driven performance evaluation.

Supplementary Data

Supplementary data are available online at <https://academic.oup.com/bib>.

Data availability statement

Datasets used in different illustrations in this review are freely available in public repositories.

Acknowledgements

The authors thank those who have contributed toward advancing meta-analysis techniques, as well as researchers who are making their datasets available in public data repositories and databases. We also thank everyone involved

with free software, from the core developers to those who contributed to the documentation, and those providing computational facilities, specifically the Centre for high performance computing (CHPC), South Africa (<https://www.chpc.ac.za>), as well as those who have helped in the preparation of this manuscript.

Funding

Some of the authors are supported partially by The German Academic Exchange Service (DAAD). This study is supported by the National Institutes of Health (NIH), USA, Common Fund under H3ABioNet (U24HG006941) and SADAaCC (U24HL135600). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

References

1. Mazandu GK, Hooper C, Opat K, et al. IHP-PING—generating integrated human protein–protein interaction networks on-the-fly. *Brief Bioinform* 2021; **22**: bbaa277. [10.1093/bib/bbaa277](https://doi.org/10.1093/bib/bbaa277).
2. Hulsen T, Jamuar SS, Moody AR, et al. From big data to precision medicine. *Front Med* 2019; **6**: 34.
3. Mazandu GK, Kyomugisha I, Geza E, et al. Designing data-driven learning algorithms: A necessity to ensure effective post-genomic medicine and biomedical research. In: *Artificial Intelligence - Applications in Medicine and Biology*. 5 Princes Gate Court. London, UK: IntechOpen Publisher, 2019, 3–18.
4. Mazandu GK, Mulder NJ. Generation and analysis of large-scale data-driven mycobacterium tuberculosis functional networks for drug target identification. *Advances in Bioinformatics* 2011; **2011**: 801478.
5. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; **8**(1):118–27.
6. Tseng GC, Ghosh D, Feingold E. Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res* 2012; **40**(9):3785–99.
7. Begum F, Ghosh D, Tseng GC, et al. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res* 2012; **40**(9):3777–84.
8. Campaign A, Yang YH. Comparison study of microarray meta-analysis methods. *BMC bioinformatics* 2010; **11**(1):408.
9. Normand ST. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med* 1999; **18**(3):321–59.
10. Metaomics. *Bioinformatics and Statistical Learning Group*, Department of Biostatistics, University of Pittsburgh. <http://www.pitt.edu/~tsengweb/MetaOmicsHome.htm> (Accessed March, 2021). 4200 Fifth Ave, Pittsburgh, PA 15260, United States.
11. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**(3):177–88.
12. Fisher RA. *Statistical methods for research workers*. Edinburgh: Oliver & Boyd, 1950.
13. Good IJ. On the weighted combination of significance tests. *J R Stat Soc B Methodol* 1955; **17**(2):264–5.
14. Han B, Eskin E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet* 2012; **8**(3):e1002555.
15. Hong F, Breitling R, McEntee CW, et al. Rankprod: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 2006; **22**(22):2825–7.

16. Li J, Tseng GC. An adaptively weighted statistic for detecting differential gene expression when combining multiple transcriptomic studies. *The Annals of Applied Statistics* 2011; 5(2A):994–1019.
17. Song C, Tseng GC. Hypothesis setting and order statistic for robust genomic meta-analysis. *The annals of applied statistics* 2014; 8(2):777.
18. Stouffer SA. A study of attitudes. *Sci Am* 1949; 180(5): 11–5.
19. Tippett LHC. *The methods of statistics: An introduction mainly for workers in the biological sciences*. Williams & Norgate: Verlag London, 1931.
20. Wang X, Chua H, Chen P, et al. Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum Mol Genet* 2013; 22(11):2303–11.
21. Wilkinson B. A statistical consideration in psychological research. *Psychol Bull* 1951; 48(2):156.
22. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* 2011; 88(5):586–98.
23. Mägi R, Morris AP. Gwama: software for genome-wide association meta-analysis. *BMC bioinformatics* 2010; 11(1):288.
24. Willer CJ, Li Y, Abecasis GR. Metal: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; 26(17):2190–1.
25. Lumley T. Rmeta: Meta-analysis Package for R. <https://CRAN.R-project.org/package=rmeta> (Accessed August, 2021). R Foundation for Statistical Computing, Vienna, Austria.
26. Struchalin M, Aulchenko Y. MetABEL: Meta-analysis of genome-wide SNP association results. <https://CRAN.R-project.org/package=MetABEL> (Accessed August, 2021). R Foundation for Statistical Computing, Vienna, Austria.
27. Nicodemus KK. Catmap: Case-Control and TDT Meta-Analysis Package. <https://CRAN.R-project.org/package=catmap> (Accessed August, 2021). R Foundation for Statistical Computing, Vienna, Austria)
28. Viechtbauer W: Metafor: Meta-analysis package for r. <https://cran.r-project.org/web/packages/metafor/> (Accessed August, 2021). R Foundation for Statistical Computing, Vienna, Austria.
29. Ma T, Huo Z, Kuo A, et al. MetaOmics: analysis pipeline and browser-based software suite for transcriptomic meta-analysis. *Bioinformatics* 2019; 35(9):1597–9.
30. Heard NA, Rubin-Delanchy P. Choosing between methods of combining-values. *Biometrika* 2018; 105(1):239–46.
31. Hong F, Breitling R. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics* 2008; 24(3):374–82.
32. Dewey M. Metap: meta-analysis of significance values. R Package Version 11. Available online: <https://cran.r-project.org/package=metap> (accessed August, 2021). R Foundation for Statistical Computing, Vienna, Austria.
33. Lipták T. On the combination of independent tests. *Magyar Tud Akad Mat Kutat Int Kozl* 1958; 3: 171–97.
34. Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet* 2013; 14(6):379–89.
35. Wang X, Kang DD, Shen K, et al. An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. *Bioinformatics* 2012; 28(19):2534–6.
36. Choi JK, Yu U, Kim S, et al. Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics* 2003; 19(suppl 1):i84–90.
37. Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954; 10(1):101–29.
38. Pereira TV, Patsopoulos NA, Salanti G, et al. Discovery properties of genome-wide association signals from cumulatively combined data sets. *Am J Epidemiol* 2009; 170(10): 1197–206.
39. Ioannidis JPA, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS ONE* 2007; 2(9):e841.
40. Morris AP. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* 2011; 35(8):809–22.
41. Breitling R, Armengaud P, Amtmann A, et al. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett* 2004; 573(1–3):83–92.
42. Chang LU, Lin HM, Sibille E, et al. Meta-analysis methods for combining multiple expression profiles: comparisons, statistical characterization and an application guideline. *BMC bioinformatics* 2013; 14(1):368.
43. Dreyfuss JM, Johnson MD, et al. Meta-analysis of glioblastoma multiforme versus anaplastic astrocytoma identifies robust gene markers. *Mol Cancer* 2009; 8(1):71.
44. Huo Z, Tang S, Park Y, et al. P-value evaluation, variability index and biomarker categorization for adaptively weighted fisher's meta-analysis method in omics applications. *arXiv preprint arXiv* 2017; 1708: 05084.
45. Birnbaum A. Combining independent tests of significance. *J Am Stat Assoc* 1954; 49(267):559–74.
46. Clough E, Barrett T. The gene expression omnibus database. In: *Statistical genomics*. Springer, 2016, 93–110.
47. Wolpert DH, Macready WG. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation* 1997; 1(1):67–82.
48. Krupka M, Seydel K, Feintuch CM, et al. Mild plasmodium falciparum malaria following an episode of severe malaria is associated with induction of the interferon pathway in malawian children. *Infect Immun* 2012; 80(3):1150–5.
49. Boldt ABW, Tong HV, Grobusch MP, et al. The blood transcriptome of childhood malaria. *EBioMedicine* 2019; 40: 614–25.
50. Ockenhouse CF, Hu W, Kester KE, et al. Common and divergent immune response signaling pathways discovered in peripheral blood mononuclear cell gene expression patterns in presymptomatic and clinically apparent malaria. *Infect Immun* 2006; 74(10):5561–73.
51. Muehlenbachs A, Fried M, Lachowitz J, et al. Genome-wide expression analysis of placental malaria reveals features of lymphoid neogenesis during chronic infection. *The Journal of Immunology* 2007; 179(1):557–65.
52. Ingles-Esteve J, Morales M, Dalmases A, et al. Inhibition of specific nf- κ b activity contributes to the tumor suppressor function of 14-3-3 σ in breast cancer. *PLoS ONE* 2012; 7(5):e38347.
53. Richardson AL, Wang ZC, Nicolo AD, et al. X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell* 2006; 9(2):121–32.
54. Merdad A, Karim S, Schulten H, et al. Expression of matrix metalloproteinases (mmps) in primary human breast cancer: Mmp-9 as a potential biomarker for cancer invasion and metastasis. *Anticancer Res* 2014; 34(3):1355–66.
55. Ngugi HK, Esker PD, Scherm H. Meta-analysis to determine the effects of plant disease management measures: review and case studies on soybean and apple. *Phytopathology* 2011; 101(1):31–41.

56. Su X, McDonough DJ, Chu H, et al. Application of network meta-analysis in the field of physical activity and health promotion. *J Sport Health Sci* 2020; **9**: 511, 20.
57. Berchiolla P, Chiffi D, Valente G, et al. The power of meta-analysis: a challenge for evidence-based medicine. *Euro Jnl Phil Sci* 2021; **11**: 7.
58. Mueller M, D'Addario M, Egger M, et al. Methods to systematically review and meta-analyse observational studies: a systematic scoping review of recommendations. *BMC Med Res Methodol* 2018; **18**(1):44.
59. Ngo-Bitoungui VJ, Belinga S, Mnika K, et al. Investigations of kidney dysfunction-related gene variants in sickle cell disease patients in Cameroon (Sub-Saharan Africa). *Front Genet* 2021; **12**: 595702.
60. Chimusa ER, Dalvie S, Dandara C, et al. Post genome-wide association analysis: dissecting computational pathway/network-based approaches. *Brief Bioinform* 2019; **20**(2): 690–700.
61. Toro-Domínguez D, Villatoro-García JA, Martorell-Marugán J, et al. A survey of gene expression meta-analysis: methods and applications. *Brief Bioinform* 2021; **22**(2): 1694–705.
62. Lee SY, Park YK, Yoon CH, Kim K, Kim KC. Meta-analysis of gene expression profiles in long-term non-progressors infected with HIV-1. *BMC Med Genomics* 2019 Dec; **12**(1):1–0.
63. Waldron L, Riestler M. Meta-analysis in gene expression studies. In: *Statistical Genomics*. New York, NY: Humana Press, 2016, 161–76.
64. Deelen J, Evans DS, Arking DE, et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat Commun* 2019; **10**(1):1–4.