TOMAS: A novel TOpology-aware Meta-Analysis approach applied to System biology

Tin Nguyen
Department of Computer Science
Wayne State University
Detroit, MI 48202, USA
tin@wayne.edu

Diana Diaz
Department of Computer Science
Wayne State University
Detroit, MI 48202, USA
dmd@wayne.edu

Sorin Draghici
Department of Computer Science
Wayne State University
Detroit, MI 48202, USA
sorin@wayne.edu

ABSTRACT

With the explosion of high-throughput data, an effective integrative analysis is needed to decipher the knowledge accumulated in multiple studies. However, batch effects, patient heterogeneity, and disease complexity all complicate the integration of data from different sources. Here we introduce TOMAS, a novel meta-analysis framework that transforms the challenging meta-analysis problem into a set of standard analysis problems that can be solved efficiently. This framework utilizes techniques based on both p-values and effect sizes to identify differentially expressed genes and their expression change on a genome-scale. The computed statistics allow for topology-aware pathway analysis of the given phenotypes, where topological information of genes is taken into consideration. We compare TOMAS with four meta-analysis approaches, as well as with three dedicated pathway analysis approaches that employ multiple datasets (MetaPath). The eight approaches have been tested on 609 samples from 9 Alzheimer's studies conducted in independent labs for different sets of patients and tissues. We demonstrate that the topology based meta-analysis framework overcomes noise and bias to identify pathways that are known to be implicated in Alzheimer's disease. While presented here in a genomic data analysis application, the proposed framework is sufficiently general to be applied in other research areas.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Statistical computing; J.3 [**Life and Medical Sciences**]: Biology and genetics

General Terms

Algorithms

Keywords

meta-analysis, effect size, pathway

BCB'16, October 2–5, 2016, Seattle, WA, USA.
 ISBN 978-1-4503-4225-4/16/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2975167.2975168.

1. INTRODUCTION

Advanced techniques in sequencing and microarray assays have transformed biological research by enabling comprehensive monitoring of biological systems. Vast amounts of data of all types have accumulated in many public repositories, such as Gene Expression Omnibus (GEO) [1], Array Express [2], and The Cancer Genome Atlas (TCGA) (`http://cancergenome.nih.gov`). Gene expression data, as measured by microarray and high-throughput sequencing, are particularly abundant in public repositories, such that many diseases are represented by a dozen studies or more.

A typical comparative analysis of molecular data generally yields a set of genes that are differentially expressed (DE) between the conditions. These sets of DE genes contain the genes that are likely to be involved in the biological processes responsible for the disease. However, such sets of genes are usually insufficient to reveal the underlying biological mechanisms. Therefore, researchers have developed a variety of knowledge bases that map genes to functional modules. These knowledge bases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) [3] or Reactome [4], contain graphs that describe how genes interact together to accomplish specific biological processes. Over-Representation Analysis (ORA) [5], Gene Set Enrichment Analysis (GSEA) [6], Gene Set Analysis (GSA) [7], and Impact Analysis [8], are examples of approaches designed to identify the pathways that are relevant in a given condition.

Remarkably, due to inherent bias present in individual studies, independent studies of the same disease often yield completely different lists of differential expressed (DE) genes, making interpretation extremely difficult [9, 10]. This problem is not resolved by simply analyzing the same individual datasets at the system level, as pathway analysis results are also often inconsistent as well [11]. Effective meta-analysis approaches, which are statistical methods for the quantitative analysis of independent but related studies, are needed to unify the biological knowledge spread over such similar studies with apparently incongruent results.

Meta-analysis of gene expression data has primarily been used for DE gene detection [12]. Early meta-analyses simply performed the intersection or union of DE gene lists obtained from individual studies [13, 14], resulting in a single list which is either too conservative or too inclusive, respectively. Rhodes et al. [15] were among the earliest to apply Fisher's method [16] for DE gene detection. Since then, other p-value based meta-analysis methods have been applied, such as Stouffer's method [17], minP [18], and maxP [19].

Recently, meta-analysis has also been used to combine multiple experiments for on the pathway level [11, 20, 21, 14]. Kaever et al. [20] used Fisher's and Stouffer's method to combine p-values of pathways from independent studies. Nguyen et al. [11] added another level of meta-analysis to make better use of large number of samples within individual studies. Shen et al. [21] developed a dedicated approach, named MetaPath, that performs meta-analysis at both the gene and pathway level separately, and then combines the results to give the final p-value and ranking of pathways.

The major drawback of these p-value-based meta-analysis approaches is that they neglect the actual expression changes, i.e. effect sizes. This results in a critical information loss. While p-value is partly a function of effect size, it is also partly a function of sample size [22]. For example, with large sample size, a statistical test will almost always demonstrate a significant difference, unless the effect size is exactly zero. In reality, any individual study will include some degree of batch effects, sampling/study bias, noise, and measurement errors. Simply combining individual p-values would not be able to correct such problem. On the contrary, meta-analysis of effect sizes across all studies would definitely compensate for and eliminate such random effects [22, 23].

Another limitation of p-value based meta-analyses is that they work under the assumption that the p-values provided by the individual statistical tests follow a uniform distribution under the null hypothesis. Previous reports describe non-uniform distributions of p-values under the null as due to specific factors such as improper normalization, cross-hybridization, poorly characterized variance, heteroskedasticity in microarray data analysis [24, 25].

Here we propose TOMAS (TOpology-aware Meta-Analysis applied to System biology), a new meta-analysis approach that utilizes techniques based on both p-values and effect sizes to combine independent studies. Our contribution is two-fold. First we use empirical distributions to calculate p-values for individual studies. This approach avoids making assumptions about null distributions of gene expression and thus compensates for potential bias. Second, the meta-analysis of effect sizes accurately estimates the central tendency of expression change for individual genes. The estimated genome-scale expression change allows for topology-aware analysis, in which gene interaction and signal propagation are taken into consideration.

We illustrate the new approach using 609 samples from 9 Alzheimer's studies conducted in independent labs for different sets of patients and tissues. We compare TOMAS with 7 other approaches: GSEA and GSA combined with Fisher's method [16] and addCLT [11], plus 3 MetaPath approaches [21]. TOMAS outperforms existing approaches to identify pathways relevant to the disease. Our results suggest that the combination of both p-value based and effect based meta-analysis techniques provides more power and robustness than each taken alone. While presented here in the context of pathway analysis, the framework can be modified to adapt to other domains or applications, such as biomarker detection, genome-wide association studies (GWAS), or enrichment analysis (Gene Ontology, gene set analysis).

2. METHODS

The pipeline consists of three main modules: i) identifying genes that are differentially expressed under the disease condition, ii) estimate the expression change (effect size) of the genes, and iii) perform pathway analysis using the calculated statistics. The first and second modules essentially represent two different meta-analysis approaches at the gene level – one is based on p-value while another is based on effect-size. In Figure 1, the purple arrows show the detailed steps of the first module while the blue arrows display the steps of the second module. The results obtained from the two modules then serve as the input for the topology-aware pathway analysis.

To identify differentially expressed genes, we first calculate p-values for each gene in each study, and then combine the p-values for each gene across independent studies. Study-specific and gene-specific empirical distributions of t-statistics are calculated by randomly assigning class labels to samples. Each distribution consists of $1,000$ random t-statistics. The left- and right-tailed p-values are calculated by comparing the actual t-statistic obtained from the real grouping against these empirical distributions. The p-values are then combined separately for the left and right tails. The final p-value for the gene is set to twice the minimum of the two combined p-values (see Section 2.1 for more details).

To estimate the central tendency of effect sizes, we first compute standardized mean difference (SMD) for each gene in each study. We next estimate the overall effect size using the random-effects model and the REstricted Maximum Likelihood (REML) algorithm [26]. This overall estimated effect size represents the change in expression of the gene under the effect of the disease (Section 2.2).

The combined p-values and the estimated effect sizes represent the evidence from which we will infer the differential expression between the two phenotypes. These statistics can be used as the input for analyses that are routinely done in pathway analysis (Section 2.3). The difference is that these statistics are gathered from multiple independent studies and thus are expected to accurately represent the real expression change of genes on a genome-scale. Technical details of each step are described in the following sections.

2.1 Computing cumulative p-values

2.1.1 Empirical hypothesis testing

In this framework, we use the t-statistic as the discriminating statistic between the conditions (disease vs. healthy). Formally, denoting x_1 and x_2 as the two groups of measurements to be compared, the two-sample t-statistic for unequal variances is defined as:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (1)$$

where $\bar{x}_1$ and $\bar{x}_2$ are sample means, n_1 and n_2 are sample sizes, s_1^2 and s_2^2 are sample variances.

Considering the dataset DS_i ($i \in [1..m]$), we randomly divide the biological samples of the dataset into two groups and then calculate the t-statistic for each gene. We repeat this step $1,000$ times to construct the empirical distributions $\xi_1^i, \xi_2^i, \ldots, \xi_n^i$ (where n is the number of genes). Since these distributions are constructed from the random sampling of the same (mixed) set of measurements, they can be considered null distributions. The t-statistics obtained from the real grouping are compared against these distributions to provide a test for differential expression.

For the gene j, we have m left-tailed p-values $p_{jl}^1, \ldots, p_{jl}^m$ where l denotes "left-tailed". We combine these m values

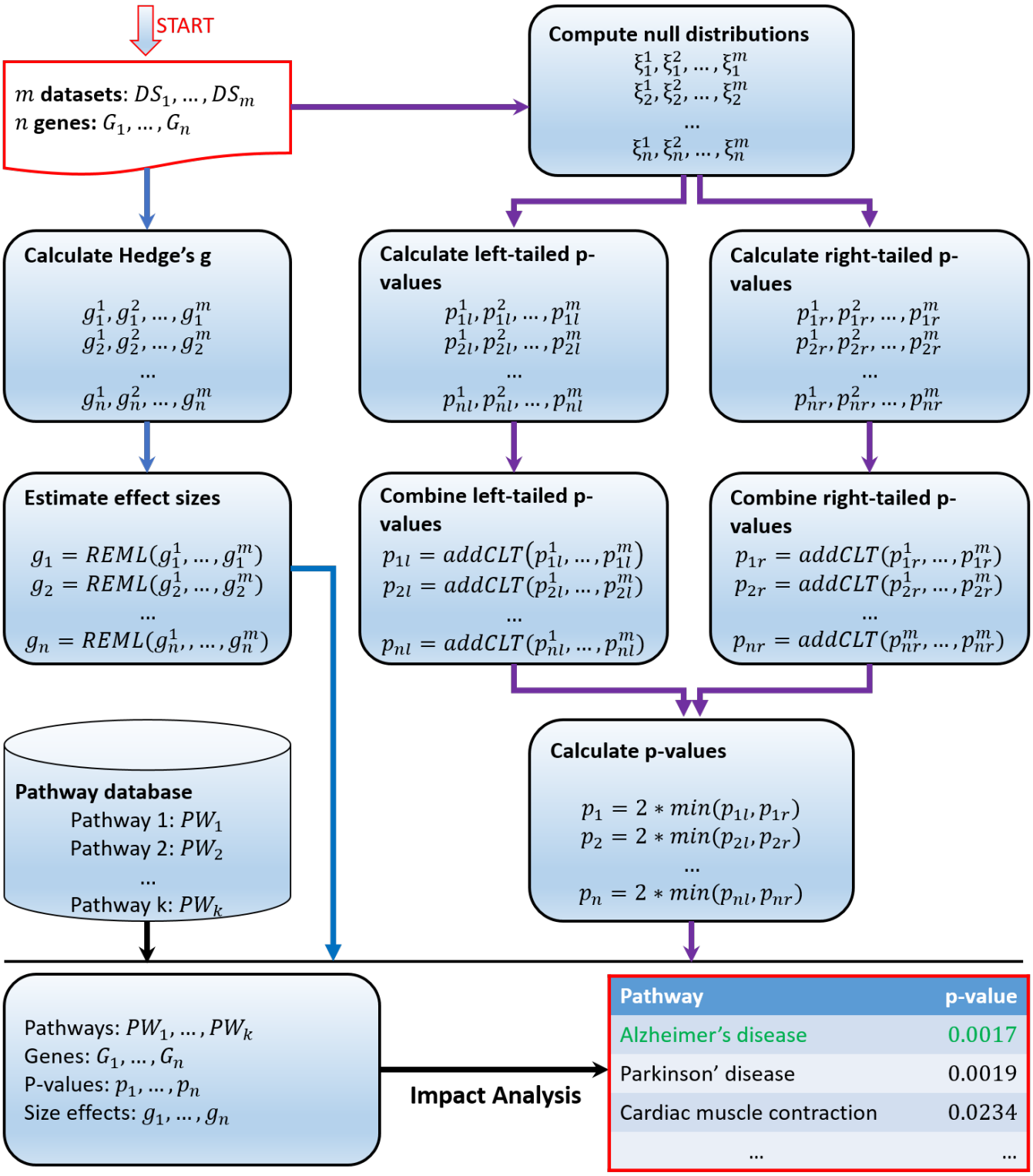

Figure 1: The overall pipeline of TOMAS. The input consists of m independent datasets, which have n genes in common. The blue arrows show the computation of effect sizes. For each gene, we first calculate Hedge'g in each study and then estimate the overall effect size using the REstricted Maximum Likelihood (REML) algorithm. The purple arrows show the computation of empirical p-values. Study specific, gene specific empirical distribution of t-statistics are calculated by randomly assigning class labels to samples. The left- and right-tailed p-values are calculated by comparing the actual t-statistic obtained from the real grouping against these empirical distributions. The p-values are then combined separately for the left and right tails. The final p-value of the gene is set to twice the minimum of the two combined p-values. The effect sizes and the combined p-values of the genes then serve as input of Impact Analysis.

using addCLT (see Section 2.1.2) to get a single value $p_{jl} = addCLT(p_{jl}^1, \ldots, p_{jl}^m)$. Similarly, we calculate the right-tailed p-value $p_{jr} = addCLT(p_{jr}^1, \ldots, p_{jr}^m)$, in which r denotes "right-tailed". The final p-value for the gene will be set to:

$$p_j = 2 * \min(p_{jl}, p_{jr}) \tag{2}$$

This combined p-value of the gene represents how unlikely the cumulative differential expression is observed by chance.

In this work, we use the t-statistic is as the discriminating statistic between the two phenotypes. However, we acknowledge that the t-statistic can be substituted by any other statistics to suit the purpose of the analysis.

2.1.2 Combining p-values

Existing methods of combining independent p-values include Fisher's method [16], Stouffer's method [17], minP [18], maxP [19], and addCLT [27, 11, 28]. Fisher's method uses the negative log product of the p-values as the summary statistic. Under the null hypothesis, this statistic follows a chi-square distribution with $2m$ degrees of freedom. Similarly, the test statistic of Stouffer's method is the sum of p-values transformed into standard normal variables, divided by the square root of m. This summary statistic follows a standard normal distribution if the null hypothesis is true. A major drawback of Fisher's and Stouffer's method is that they are sensitive to outliers. For example, if one of the individual p-value approaches zero, the combined p-value will be zero regardless of other individual p-values. The same is true for the minP's and maxP's statistic, where outliers can greatly influence the combined p-value.

On the contrary, the addCLT method uses the average of p-values as the test statistic and therefore it is more robust against extreme p-values. Denoting the individual p-values to be combined as $P_1, P_2, \ldots, P_m$, the summary statistic is defined as $X = \frac{\sum_{i=1}^{m} P_i}{m}$ ($X \in [0, 1]$). The probability density function (pdf) is derived from a linear transformation of the Irwin-Hall distribution [29, 30] as follows:

$$f(x) = \frac{m}{(m-1)!} \sum_{i=0}^{\lfloor m \cdot x \rfloor} (-1)^i \binom{m}{i} (m \cdot x - i)^{m-1} \tag{3}$$

When m is large, the computation of Equation (3) can lead to underflow/overflow problems. Therefore, we use the Central Limit Theorem [31] to estimate this distribution in this case. The variable X is the mean of m independent and identically distributed (i.i.d.) random variables that follow a uniform distribution with a mean of $\frac{1}{2}$ and a variance of $\frac{1}{12}$. From the Central Limit Theorem [31], the average of such m i.i.d. variables follows a normal distribution with mean $\mu = \frac{1}{2}$ and variance $\sigma^2 = \frac{1}{12m}$, i.e. $X \sim \mathcal{N}(\frac{1}{2}, \frac{1}{12m})$ for sufficiently large values of m.

2.2 Estimating overall effect sizes

2.2.1 Standardized mean difference

Since gene expression is scaled differently in each study, it is more reasonable to use the standardized mean difference (SMD) to measure expression changes, instead of the raw mean difference. Here we briefly describe two popular SMD metrics: *Cohen's d* [32] and *Hedge's g* [33].

Consider a study composed of two independent groups. Let $\bar{x}_1$ and $\bar{x}_2$ represent the sample means for that gene in the two groups, n_1 and n_2 the number of samples in each group. The pooled standard deviation of the two groups and Cohen's d are calculated as follows:

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{4}$$

$$d = \frac{\bar{x}_1 - \bar{x}_2}{S_{pooled}} \tag{5}$$

We note that Cohen's d, which is based on sample averages, tends to overestimate the population effect size for small samples. Let n be the degrees of freedom used to estimate S_{pooled}, i.e. $n = n_1 + n_2 - 2$. The corrected effect size, or Hedges' g [33], can be computed as follows:

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

$$g = J \cdot d \tag{7}$$

where Γ is the gamma function. In this work, we use Hedge' g as the standardized mean difference (SMD) between disease and control groups for each gene.

2.2.2 Restricted maximum likelihood estimation

Consider a collection of m studies where the effect size estimates, $y_1, \ldots, y_m$ have been derived as shown in Equation (7). A fixed-effects model would assume that there is one true effect size which underlies all of the studies in the analysis, such that all differences in observed effects are due to sampling error. However, this assumption is implausible since it cannot account for heterogeneity between studies [34, 35, 36] (see Appendix A for details).

In contrast, the random-effects model allows for variability of the true effect. For example, the effect size might be higher (or lower) in studies where the participants are older, or have a healthier lifestyle compared to others. The random-effects model assumes that each effect size estimate can be decomposed into two variance components by a two stage hierarchical process [35, 37, 38]. The first variance represents the variability of the effect size across studies, and the second variance represents the sampling error within each study. We can write the random-effects model as:

$$y_i = \mu + N(0, \sigma^2) + N(0, \sigma_{\epsilon_i}^2) \tag{8}$$

where μ is the central tendency of the effect size, $N(0, \sigma^2)$ represents the error term by which the effect size in the i^{th} study differs from the central tendency μ, and $N(0, \sigma_{\epsilon_i}^2)$ represents the sampling error.

The derivation and formulation of the REstricted Maximum Likelihood (REML) algorithm has been described in the literature [35, 39, 40] (see Appendix B for details). The *REML estimator* of $\hat{\mu}$ and the *combined p-value* of individual genes serve as the input of Impact Analysis (Section 2.3).

2.3 Topology-aware pathway analysis

Impact Analysis [41, 8] combines two types of evidence: (i) the over-representation of DE genes in a given pathway, and (ii) the perturbation of that pathway, caused by disease, as measured by propagating expression changes through the pathway topology. These two aspects are captured by the probability values, P_{NDE} and P_{PERT}, respectively.

The first p-value, P_{NDE}, is obtained using the hypergeometric model, which is the probability of obtaining at least

the observed number of differentially expressed genes. The input of the hypergeometric test consists of: i) the set of all measured genes, and ii) the set of DE genes. The former is the set of genes that are common in all datasets while the latter is the set of genes that have a combined $p \leq 0.01$.

The second p-value, P_{PERT}, depends on the identity of the specific genes that are differentially expressed as well as on the known interactions between the genes. It is calculated based on the perturbation factor in each pathway. The perturbation factor of a gene, $PF(g)$, is defined as:

$$PF(g) = \Delta E(g) + \sum_{u \in US_g} \beta_{ug} \cdot \frac{PF(u)}{N_{ds}(u)} \quad (9)$$

The first term represents the signed normalized expression change of the gene g, i.e. effect size. The second term is the sum of perturbation factors of upstream genes, normalized by the number of downstream genes of each such upstream gene. The value of β_{ug} quantifies the strength of interaction between u and g. By default, $\beta_{ug} = 1$ for *activation* and $\beta_{ug} = -1$ for *repression*. The net perturbation accumulation at the level of each gene, $Acc(g)$, is calculated by subtracting the observed expression change from the perturbation factor:

$$Acc(g) = PF(g) - \Delta E(g) \quad (10)$$

The total accumulated perturbation in the pathway is then computed as:

$$Acc(P_i) = \sum_{g \in P_i} Acc(g) \quad (11)$$

The null distribution of $Acc(P_i)$ is built by permuting both sample and gene labels of expression changes. The p-value, P_{PERT}, is then calculated by the probability of having values more extreme than the actually observed $Acc(P_i)$.

To compute P_{PERT}, the following input is required: the graphical representation of the pathway, the set of DE genes, and their estimated effect sizes. The graphical representation of pathways is obtained from KEGG; the list of DE genes is obtained from the combined p-values (Section 2.1.2); the estimated effect sizes are obtained from the random-effects model and the REML algorithm (Section 2.2.2).

The two p-values, P_{NDE} and P_{PERT}, are then combined using Fisher's method to get a p-value that represents how likely the pathway is impacted under the effect of the disease.

3. RESULTS

We compare the performance of TOMAS with 7 existing approaches: GSEA and GSA, each combined with Fisher's method and addCLT, plus 3 MetaPath approaches: gene level, pathway level and combined. We analyze 9 Alzheimer's gene expression datasets that were generated in independent labs for different sets of patients and tissues. The 9 datasets are GSE1297 (hippocampus), GSE4757 (entorhinal cortex), GSE5281 (entorhinal cortex, medial temporal gyrus, posterior cingulate, superior frontal gyrus, hippocampus, and primary visual cortex), GSE16759 (parietal lobe), GSE18309 (peripheral blood mononuclear cell), GSE28146 (hippocampus), GSE36980 (frontal cortex, temporal cortex, and hippocampus), GSE39420 (brain tissues), and GSE48350 (entorhinal cortex, post-central gyrus, hippocampus, and superior frontal gyrus). Table 1 shows the details of each dataset, such as the number of control and disease samples, tissues, and platforms. Pre-processing was performed on individual datasets using the *threestep* function from the package *affyPLM version 1.38.0* [42]. The parameters used for the *threestep* function are: robust multi-array analysis (RMA) background adjustment, quantile normalization, and median polish summarization. We use the KEGG database version 76, which includes 182 signaling pathways.

There is a dedicated pathway in KEGG, *Alzheimer's disease*, that was created precisely in order to describe the known mechanisms involved in this disease. However, it is well known that *Alzheimer's disease*, *Parkinson's disease*, and *Huntington's disease* share many signaling mechanisms and affect the same tissue (brain). The common elements include abnormal protein folding, endoplasmic reticulum stress, and ubiquitin mediated breakdown of proteins, leading to programmed cell death [43, 44, 45, 46]. In addition, previous studies [47] have shown the presence of a strong cross-talk that makes these three neurological diseases pathways appear as significant simultaneously, due to their dominant mitochondrial module. Therefore, we expect a good analysis method to find all three of these pathways as significant in this meta-analysis of Alzheimer's data.

For each of the nine datasets, GSEA produces a list of 182 p-values for 182 signaling pathways. In other words, nine p-values will be calculated for each pathway – one per study. These p-values are independent and thus can be combined using either Fisher's method or addCLT. Therefore, each meta-analysis method produces a list of pathways ranked according to the combined p-values. Similarly, we also combine GSA p-values using Fisher's and the addCLT method. The p-values of each list are then adjusted for multiple comparison using False Discovery Rate (FDR).

Table 2 displays the results obtained from the four approaches: i) GSA + Fisher's method, ii) GSA + addCLT, iii) GSEA + Fisher's method, and iv) GSEA + addCLT. The table shows the 24 top ranked pathways and the adjusted p-values for each of the 4 approaches. The horizontal lines across each list marks the cutoff of 1%. The three neurological disorder pathways, *Alzheimer's disease*, *Huntington's disease*, and *Parkinson's disease* are highlighted in green.

There are 23 significant pathways using GSA with Fisher's method, many of which are likely to be false positives. Each of the top 17 pathways has a combined p-value equal to zero because the p-value was zero for at least one of the datasets. The pathways *Alzheimer's disease*, *Huntington's disease*, and *Parkinson's disease* are not reported as significant and are ranked at the positions 26^{th}, 25^{th}, and 33^{rd}, respectively. GSA with addCLT produces a single false positive, the *Retrograde endocannabinoid signaling*, but it is also unable to identify any of the true positive pathways. The three neurological disorder pathways are also ranked at higher positions (4^{th}, 6^{th}, and 15^{th}).

Similarly, GSEA combined with Fisher's method and addCLT also fail to identify any of the three neurological disorder pathways as significant. GSEA with Fisher's method rank the pathways *Alzheimer's disease*, *Huntington's disease*, and *Parkinson's disease* at positions 8^{th}, 10^{th}, and 17^{th}, respectively; GSEA with the addCLT rank them at positions 49^{th}, 10^{th}, and 47^{th}, respectively. In summary, all four approaches fail to provide the needed power to identify the three neurological disorder pathways as significant at the significance threshold of 1%.

Table 1: Description of the 9 Alzheimer's gene expression datasets used in the experimental studies.

Accession ID	Control	Disease	Tissue	Platform
GSE1297	9	22	Hippocampus	Affymetrix Human Genome U133A
GSE4757	10	10	Entorhinal cortex	Affymetrix Human Genome U133 Plus 2.0
GSE5281	74	87	Entorhinal cortex, medial temporal gyrus, posterior cingulate, superior frontal gyrus, hippocampus, and primary visual cortex	Affymetrix Human Genome U133 Plus 2.0
GSE16759	4	4	Parietal lobe	Affymetrix Human Genome U133 Plus 2.0
GSE18309	3	3	Peripheral blood mononuclear cell	Affymetrix Human Genome U133 Plus 2.0
GSE28146	8	22	Hippocampus	Affymetrix Human Genome U133 Plus 2.0
GSE36980	47	32	Frontal cortex, temporal cortex, and hippocampus	Affymetrix Human Gene 1.0 ST
GSE39420	7	14	Brain tissues	Affymetrix Human Gene 1.1 ST
GSE48350	173	80	Entorhinal cortex, post-central gyrus, hippocampus, and superior frontal gyrus	Affymetrix Human Genome U133 Plus 2.0

Table 2: The 24 top ranked pathways and their adjusted p-values obtained by combining the GSA and GSEA p-values using Fisher's method and addCLT for Alzheimer's data. The horizontal lines show the 1% significance threshold. The neurological disorder pathways, *Alzheimer's disease*, *Parkinson's disease* and *Huntington's disease*, are highlighted in green. All the four meta-analysis approaches fail to identify the target pathway *Alzheimer's disease* as significant, and rank it at the positions 26^{th}, 4^{th}, 8^{th}, and 49^{th}, respectively.

	GSA + Fisher's method		GSA + addCLT	
	Pathway	p.fdr	Pathway	p.fdr
1	Retrograde endocannabinoid signaling	0	Retrograde endocannabinoid signaling	0.0072
2	Toxoplasmosis	0	Toxoplasmosis	0.0177
3	Long-term depression	0	Long-term depression	0.0177
4	Gap junction	0	Alzheimer's disease	0.0177
5	Amphetamine addiction	0	Morphine addiction	0.0177
6	Vasopressin-regulated water reabsorption	0	Huntington's disease	0.0177
7	Staphylococcus aureus infection	0	GABAergic synapse	0.0177
8	Small cell lung cancer	0	Epithelial cell signaling in Helicobacter pylori infection	0.0177
9	cAMP signaling pathway	0	Glutamatergic synapse	0.0218
10	Pathogenic Escherichia coli infection	0	Gap junction	0.0246
11	Adipocytokine signaling pathway	0	Amyotrophic lateral sclerosis (ALS)	0.0255
12	Platelet activation	0	Cardiac muscle contraction	0.0257
13	Phospholipase D signaling pathway	0	Oxytocin signaling pathway	0.0279
14	Ovarian steroidogenesis	0	Endocrine and other factor-regulated calcium reabsorption	0.0279
15	mRNA surveillance pathway	0	Parkinson's disease	0.0279
16	Maturity onset diabetes of the young	0	Non-alcoholic fatty liver disease (NAFLD)	0.0279
17	Chemokine signaling pathway	0	Synaptic vesicle cycle	0.0279
18	Synaptic vesicle cycle	0.0005	Transcriptional misregulation in cancer	0.0279
19	Glutamatergic synapse	0.0005	Circadian entrainment	0.0279
20	Dopaminergic synapse	0.0012	Serotonergic synapse	0.0279
21	Serotonergic synapse	0.0021	Dopaminergic synapse	0.0279
22	Endocrine and other factor-regulated calcium reabsorption	0.0077	Vibrio cholerae infection	0.0279
23	Morphine addiction	0.0092	Inflammatory bowel disease (IBD)	0.0279
24	Epithelial cell signaling in Helicobacter pylori infection	0.0105	HTLV-I infection	0.0279

	GSEA + Fisher's method		GSEA + addCLT	
	Pathway	p.fdr	Pathway	p.fdr
1	Amyotrophic lateral sclerosis (ALS)	0	Vascular smooth muscle contraction	0.2508
2	Serotonergic synapse	0	Vibrio cholerae infection	0.2508
3	Insulin secretion	0	Melanogenesis	0.5060
4	Cardiac muscle contraction	0	Renin-angiotensin system	0.5060
5	Endocrine and other factor-regulated calcium reabsorption	0.1155	Ras signaling pathway	0.8492
6	Synaptic vesicle cycle	0.1155	Dorso-ventral axis formation	0.8492
7	Vascular smooth muscle contraction	0.1666	Amyotrophic lateral sclerosis (ALS)	0.8492
8	Alzheimer's disease	0.1666	Adipocytokine signaling pathway	0.8492
9	Pancreatic secretion	0.1666	Inflammatory mediator regulation of TRP channels	0.8492
10	Huntington's disease	0.1666	Huntington's disease	0.8492
11	Glutamatergic synapse	0.2896	Prostate cancer	0.8492
12	Adrenergic signaling in cardiomyocytes	0.3053	GnRH signaling pathway	0.8492
13	Vibrio cholerae infection	0.3694	Thyroid hormone synthesis	0.8492
14	Renin-angiotensin system	0.3708	Rap1 signaling pathway	0.8492
15	Melanogenesis	0.3708	Synaptic vesicle cycle	0.8492
16	Amphetamine addiction	0.4514	Regulation of autophagy	0.8492
17	Parkinson's disease	0.4514	Sulfur relay system	0.8492
18	Ribosome biogenesis in eukaryotes	0.4514	Toxoplasmosis	0.8492
19	Thyroid hormone synthesis	0.4673	Pancreatic secretion	0.8492
20	Non-alcoholic fatty liver disease (NAFLD)	0.5247	Endometrial cancer	0.8492
21	Vasopressin-regulated water reabsorption	0.5391	Cell cycle	0.8492
22	Prostate cancer	0.5697	Vasopressin-regulated water reabsorption	0.8492
23	Dopaminergic synapse	0.5712	AMPK signaling pathway	0.8492
24	Wnt signaling pathway	0.5712	Alcoholism	0.8492

Table 3: The 10 top ranked pathways and FDR-corrected p-values obtained by combining Alzheimer's data using 4 approaches: MetaPath_G, MetaPath_P, MetaPath_I, and TOMAS. MetaPath_G, MetaPath_P, and MetaPath_I fail to identify the target pathway *Alzheimer's disease* as significant, and rank it at the positions 72^{th}, 90^{th}, and 28^{th}, respectively. TOMAS identifies the target pathway as significant and ranks it on top.

	MetaPath_G (gene-level)		MetaPath_P (pathway-level)	
	Pathway	p.fdr	Pathway	p.fdr
1	Type II diabetes mellitus	0.1850	Long-term depression	0.2162
2	Renin-angiotensin system	0.6165	Dorso-ventral axis formation	0.2412
3	Circadian rhythm	0.6483	Allograft rejection	0.2422
4	Shigellosis	0.7750	Circadian rhythm	0.2830
5	Proteoglycans in cancer	0.7820	Endocrine and other factor-regulated calcium reabsorption	0.2855
6	Salivary secretion	0.7822	Shigellosis	0.3350
7	Small cell lung cancer	0.7862	Renal cell carcinoma	0.3602
8	Taste transduction	0.7864	Gap junction	0.3728
9	Malaria	0.7872	VEGF signaling pathway	0.3875
10	Regulation of actin cytoskeleton	0.7913	African trypanosomiasis	0.4273

	MetaPath_I (both levels)		**TOMAS**	
	Pathway	p.fdr	Pathway	p.fdr
1	Shigellosis	0.3320	Alzheimer's disease	$<10^{-4}$
2	Renin-angiotensin system	0.3451	Parkinson's disease	$<10^{-4}$
3	Long-term depression	0.3533	Huntington's disease	$<10^{-4}$
4	Allograft rejection	0.3545	Synaptic vesicle cycle	$<10^{-4}$
5	Type II diabetes mellitus	0.3670	Non-alcoholic fatty liver disease (NAFLD)	0.0003
6	Endocrine and other factor-regulated calcium reabsorption	0.3750	Cardiac muscle contraction	0.0007
7	Dorso-ventral axis formation	0.3798	Epithelial cell signaling in Helicobacter pylori infection	0.0027
8	Circadian rhythm	0.4188	Vibrio cholerae infection	0.0170
9	Renal cell carcinoma	0.5656	Fc gamma R-mediated phagocytosis	0.0302
10	Gap junction	0.5934	Regulation of actin cytoskeleton	0.0427

We also employ MetaPath to combine the 9 Alzheimer's studies. MetaPath performs meta-analysis at both gene and pathway levels separately, and then combines the results to give the final p-value and ranking of pathways. At the gene level, MetaPath calculates a t-statistic for each gene in each study, then combines them using the maxP method [19]. A pathway enrichment score is calculated using these genes, for each pathway, using a Kolmogorov-Smirnov test, and assessed for significance with a sample-wise permutation test. At the pathway level, MetaPath calculates pathway enrichment for each individual study, then combines the p-values, again using the maxP method [19]. Finally, p-values from the gene and pathway level are integrated using minP [18] to give the final p-value and ranking of pathways.

Table 3 shows the top 10 ranked pathways and adjusted p-values of the three MetaPath approaches: MetaPath_G (gene-level), MetaPath_P (pathway-level), and MetaPath_I (both levels). MetaPath_G identifies no significant pathway. The three pathways *Alzheimer's disease*, *Huntington's disease*, and *Parkinson's disease* are ranked at positions 72^{nd}, 18^{th}, and 42^{nd}, respectively. Similarly, MetaPath_P produces no significant pathway and ranks the three pathways at positions 90^{th}, 43^{rd}, and 161^{st}, respectively. In consequences, the combination of these two methods, MetaPath_I, also fails to identify the three neurological disorder pathways as significant (adjusted p-values 0.87, 0.97, and 0.91 with rankings 28^{th}, 89^{th}, and 63^{rd}, respectively).

Finally, we apply TOMAS to combine the 9 studies. Statistical tests were one-sided and performed independently for the two null hypothesis that no genes are over-expressed and no genes are under-expressed. The one-sided p-values are calculated and then combined using the addCLT method [11]. The final p-value for a gene is set to twice the minimum of the two one-sided p-values. The p-values for the whole set of genes are then adjusted for multiple comparisons using FDR. We use the standard 1% as the significance cutoff to identify the DE genes. For effect sizes, we first calculate Hedge's g [33] in each study, and then estimate the overall effect size for each gene using the REML algorithm [26]. The DE genes and their expression change (effect size) then serve as input for Impact Analysis [8].

The results of TOMAS are shown in Table 3. TOMAS identifies 7 pathways as significant using the 1% threshold. All three neurological disorder pathways are significant and are ranked at the very top. The target pathway *Alzheimer's disease* is the most significant with adjusted $p = 10^{-6}$.

There are totally $3,971$ KEGG genes that are measured in the 9 datasets. The framework identified 898 genes (23%) that are differentially expressed using the cutoff FDR=1%. The pathway *Alzheimer's disease* has 62 DE genes (out of 142); *Parkinson's disease* has 59 DE genes (out of 111); *Huntington's disease* has 77 DE genes (out of 155). Most of these DE genes belong to the mitochondrial module that is included in all the three neurological disease pathways.

Given that the pathway *Alzheimer's disease* is influenced by the mitochondrial compartment, which is strongly implicated in the disease [43, 44, 45, 46], it is not surprising that other pathways with strong mitochondrial components also garner high rankings. Previous studies [47] have shown the presence of a cross-talk that makes the neurological disease pathways, *Alzheimer's disease*, *Parkinson's disease* and *Huntington's disease*, along with *Cardiac muscle contraction* and *Non-alcoholic fatty liver disease (NAFLD)*, appear as significant simultaneously, due to their dominant mitochondrial module. *Cardiac muscle contraction* has a strong mitochondrial component and is highly dependent on calcium signaling, which is also prevalent in *Synaptic vesicle cycle*, *Alzheimer's disease*, and *Huntington's disease*. Ca2+ regulates mitochondrial metabolism, but calcium overload to

mitochondria can result in cell damage from reactive oxygen [48].

4. CONCLUSIONS

In this article, we present a novel topology-aware meta-analysis able to combine multiple studies and identify the signaling pathways that are significantly impacted in a given phenotype. This approach first calculates the empirical p-values for each gene in each study and then combines them using an approach based on the Central Limit Theorem. The combined p-value of a gene represents how unlikely the differential expression of the gene is observed by chance. At the same time, the framework also estimates the overall effect sizes using the REstricted Maximum Likelihood (REML) algorithm. The two statistics then serve as input for Impact Analysis which identifies the pathways that are impacted by the given disease.

To evaluate this framework, we examined 609 samples from 9 Alzheimer's gene expression datasets. TOMAS was compared against 7 different approaches, GSA and GSEA combined with Fisher's method and addCLT, plus three approaches implemented in the MetaPath package. We demonstrated that TOMAS outperforms existing approaches to correctly identify pathways relevant to the disease.

The main innovation of TOMAS is that it addresses the challenging meta-analysis problem by transforming it into a set of standard analysis problems that can be solved efficiently.

5. ACKNOWLEDGMENTS

This research was supported in part by the following grants: NIH R01 DK089167, R42 GM087013 and NSF DBI-0965741, and by the Robert J. Sokol Endowment in Systems Biology. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of the funding agencies.

6. REFERENCES

[1] T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L. Robertson, N. Serova, S. Davis, and A. Soboleva. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Research*, 41(D1):D991–D995, 2013.

[2] G. Rustici, N. Kolesnikov, M. Brandizi, T. Burdett, M. Dylag, I. Emam, A. Farne, E. Hastings, J. Ison, M. Keays, N. Kurbatova, J. Malone, R. Mani, A. Mupo, R. P. Pereira, E. Pilicheva, J. Rung, A. Sharma, Y. A. Tang, T. Ternent, A. Tikhonov, D. Welter, E. Williams, A. Brazma, H. Parkinson, and U. Sarkans. ArrayExpress update–trends in database growth and links to data analysis tools. *Nucleic Acids Research*, 41(D1):D987–D990, 2013.

[3] M. Kanehisa and S. Goto. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

[4] D. Croft, A. F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M. R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Hermjakob, L. Stein, and P. D'Eustachio. The Reactome pathway knowledgebase. *Nucleic Acids Research*, 42(D1):D472–D477, 2014.

[5] S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church. Systematic determination of genetic network architecture. *Nature Genetics*, 22:281–285, 1999.

[6] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceeding of The National Academy of Sciences of the Unites States of America*, 102(43):15545–15550, 2005.

[7] B. Efron and R. Tibshirani. On testing the significance of sets of genes. *The Annals of Applied Statistics*, 1(1):107–129, 2007.

[8] S. Drăghici, P. Khatri, A. L. Tarca, K. Amin, A. Done, C. Voichiţa, C. Georgescu, and R. Romero. A systems biology approach for pathway level analysis. *Genome Research*, 17(10):1537–1545, 2007.

[9] P. K. Tan, T. J. Downey, E. L. Spitznagel Jr, P. Xu, D. Fu, D. S. Dimitrov, R. A. Lempicki, B. M. Raaka, and M. C. Cam. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Research*, 31(19):5676–5684, 2003.

[10] L. Ein-Dor, O. Zuk, and E. Domany. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *In Proceedings of the National Academy of Sciences*, 103(15):5923–5928, 2006.

[11] T. Nguyen, R. Tagett, M. Donato, C. Mitrea, and S. Drăghici. A novel bi-level meta-analysis approach-applied to biological pathway analysis. *Bioinformatics*, 32(3):409–416, 2016.

[12] G. C. Tseng, D. Ghosh, and E. Feingold. Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Research*, 40(9):3785–3799, 2012.

[13] F. Borovecki, L. Lovrecic, J. Zhou, H. Jeong, F. Then, H. Rosas, S. Hersch, P. Hogarth, B. Bouzou, R. Jensen, and D. Krainc. Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31):11023–11028, 2005.

[14] T. Manoli, N. Gretz, H.-J. Gröne, M. Kenzelmann, R. Eils, and B. Brors. Group testing for pathway analysis improves comparability of different microarray datasets. *Bioinformatics*, 22(20):2500–2506, 2006.

[15] D. R. Rhodes, T. R. Barrette, M. A. Rubin, D. Ghosh, and A. M. Chinnaiyan. Meta-analysis of microarrays interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Research*, 62(15):4427–4433, 2002.

[16] R. A. Fisher. *Statistical methods for research workers.* Oliver & Boyd, Edinburgh, 1925.

[17] S. Stouffer, E. Suchman, L. DeVinney, S. Star, and J. Williams, RM. *The American Soldier: Adjustment during army life*, volume 1. Princeton University Press, Princeton, 1949.

[18] L. H. C. Tippett. *The methods of statistics*. Williams & Norgate, London, 1931.

[19] B. Wilkinson. A statistical consideration in psychological research. *Psychological Bulletin*, 48(2):156, 1951.

[20] A. Kaever, M. Landesfeind, K. Feussner, B. Morgenstern, I. Feussner, and P. Meinicke. Meta-analysis of pathway enrichment: combining independent and dependent omics data sets. *PloS One*, 9(2):e89297, 2014.

[21] K. Shen and G. C. Tseng. Meta-analysis for pathway enrichment analysis when combining multiple genomic studies. *Bioinformatics*, 26(10):1316–1323, 2010.

[22] G. M. Sullivan and R. Feinn. Using effect size-or why the p value is not enough. *Journal of Graduate Medical Education*, 4(3):279–282, 2012.

[23] J. Cohen. Things i have learned (so far). *American Psychologist*, 45(12):1304, 1990.

[24] S. J. Barton, S. R. Crozier, K. A. Lillycrop, K. M. Godfrey, and H. M. Inskip. Correction of unexpected distributions of P values from analysis of whole genome arrays by rectifying violation of statistical assumptions. *BMC Genomics*, 14(1):161, 2013.

[25] A. A. Fodor, T. L. Tickle, and C. Richardson. Towards the uniform distribution of null P values on Affymetrix microarrays. *Genome Biology*, 8(5):R69, 2007.

[26] W. Viechtbauer et al. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3):1–48, 2010.

[27] T. Nguyen, D. Diaz, R. Tagett, and S. Draghici. Overcoming the matched-sample bottleneck: an orthogonal approach to integrate omic data. *Nature Scientific Reports*, 6:29251, 2016.

[28] T. Nguyen, C. Mitrea, R. Tagett, and S. Drăghici. DANUBE: Data-driven meta-ANalysis using UnBiased Empirical distributions - applied to biological pathway analysis. *Proceedings of the IEEE*, PP(99):1–20, March 2016.

[29] P. Hall. The distribution of means for samples of size n drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. *Biometrika*, 19(3-4):240–244, 1927.

[30] J. O. Irwin. On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's Type II. *Biometrika*, 19(3-4):225–239, 1927.

[31] O. Kallenberg. *Foundations of modern probability*. Springer-Verlag, New York, 2002.

[32] J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic Press, 2013.

[33] L. V. Hedges and I. Olkin. *Statistical method for meta-analysis*. Academic Press, 2014.

[34] G. A. Milliken and D. E. Johnson. *Analysis of messy data volume 1: designed experiments*, volume 1. Chapman & Hall/CRC, London, 2009.

[35] W. Viechtbauer. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3):261–293, 2005.

[36] J. E. Hunter and F. L. Schmidt. Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8(4):275–292, 2000.

[37] H. Goldstein. *Multilevel statistical models*, volume 922. John Wiley & Sons, New York, 2011.

[38] A. S. Bryk and S. W. Raudenbush. *Hierarchical linear models: applications and data analysis methods*. Sage Publications, Inc, 1992.

[39] D. A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358):320–338, 1977.

[40] R. R. Corbeil and S. R. Searle. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics*, 18(1):31–38, 1976.

[41] A. L. Tarca, S. Drăghici, P. Khatri, S. S. Hassan, P. Mittal, J.-s. Kim, C. J. Kim, J. P. Kusanovic, and R. Romero. A novel signaling pathway impact analysis. *Bioinformatics*, 25(1):75–82, 2009.

[42] B. M. Bolstad. *Low-level analysis of high-density oligonucleotide array data: background, normalization and summarization*. PhD thesis, University of California, 2004.

[43] R. H. Swerdlow. Brain aging, Alzheimer's disease, and mitochondria. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1812(12):1630–1639, 2011.

[44] A. Maruszak and C. Żekanowski. Mitochondrial dysfunction and Alzheimer's disease. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 35(2):320–330, 2011.

[45] X. Zhu, G. Perry, M. A. Smith, and X. Wang. Abnormal mitochondrial dynamics in the pathogenesis of Alzheimer's disease. *Journal of Alzheimer's Disease*, 33:S253–S262, 2013.

[46] H. W. Querfurth and F. M. LaFerla. Mechanisms of disease. *New England Journal of Medicine*, 362(4):329–344, 2010.

[47] M. Donato, Z. Xu, A. Tomoiaga, J. G. Granneman, R. G. MacKenzie, R. Bao, N. G. Than, P. H. Westfall, R. Romero, and S. Drăghici. Analysis and correction of crosstalk effects in pathway analysis. *Genome Research*, 23(11):1885–1893, 2013.

[48] P. S. Brookes, Y. Yoon, J. L. Robotham, M. Anders, and S.-S. Sheu. Calcium, ATP, and ROS: a mitochondrial love-hate triangle. *American Journal of Physiology-Cell Physiology*, 287(4):C817–C833, 2004.

APPENDIX

A. FIXED-EFFECT MODEL

The general form for the linear model is

$$y = X\beta + e \tag{12}$$

where $y \in R^k$ is a vector of k effect sizes, X is an design matrix, β is a vector of p unknown fixed effects, and e is a vector of random error, $e_i \sim N(0, \sigma^2)$, so the error are i.i.d. normal variables.

The likelihood function for (12) is given by

$$L(\beta, \sigma^2; y) = \prod_{i=1}^{k} f(y_i; X_i\beta, \sigma^2)$$
$$= \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(y_i - X_i\beta)^2)$$

where X_i is the i^{th} row of X, $f(y; \mu, \sigma^2)$ is the probability density function for a normal random variable y with mean μ and variance σ^2.

After removing the additive constant $-\frac{k}{2}\ln(2\pi)$, the log-likelihood function is then given by

$$l(\beta, \sigma^2; y) = -\frac{k}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^T(y = X\beta) \quad (13)$$

and the maximum likelihood estimates of μ and σ^2 are those values that maximize L or l.

Note that the ordinary least squares estimator of β is the value that minimize the sum of squares of the residuals, i.e. the difference between the data (y_i) and its estimated value ($X_i\mu$):

$$\sum_{i=1}^{n}(y_i - X_i\beta)^2 \quad (14)$$

This clearly yields the same estimate as the maximum likelihood method. Taking differentiation with respect to β and setting the derivative to zero will yield

$$\hat{\beta} = (X^TX)^{-1}X^Ty \quad (15)$$

and for σ^2

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T(y - X\hat{\beta})}{k} \quad (16)$$

When X is not full column rank, the generalized inverse of A is used, such that $AA^{-1}A = A$.

B. RANDOM-EFFECTS MODEL

More generally, the random-effects model given by Equation (8) can be written in the general linear mixed-effects model as

$$y = X\beta + Z\gamma + e \quad (17)$$

where $y \in R^k$ are the observed size effects, $X \in R^{k\times p}$ is a design matrix for $\beta \in R^p$ (vector of fixed effects parameters), Z is the design matrix for the $\gamma \in R^q$ (a vector of random effects parameters), and $e \in R^k$ is a vector of random error terms. We note that Equation (17) and Equation (8) are identical when y consists of the k effect size estimates, X is a vector composed entirely of 1's, β includes only the mean μ_θ, Z is the identity matrix, γ is comprised of the τ_i values at the population level.

As described above, we have $E(\gamma) = E(e) = 0$. In addition, from the independency between the true effect and sampling error, we also have $cov(\gamma, e) = 0$. Denoting $D = var(\gamma)$ and $V = var(y)$, we have

$$V = var(Z\gamma + e) = Zvar(\gamma)Z^T + var(e)$$
$$= ZDZ^T + var(e)$$

After assuming normality of the random terms in the model, we have $y \sim N(X\beta, V)$. Denoting the variance components in V by the vector σ^2, as described in [35, 39], we can write the log-likelihood function of β and σ^2 as:

$$l(\beta, \sigma^2; y) = -\frac{1}{2}\ln|V| - \frac{1}{2}(y - X\beta)^TV^{-1}(y = X\beta) \quad (18)$$

after removing the additive constant $-\frac{k}{2}\ln(2\pi)$. The log-likelihood function of μ_θ and σ^2_θ for Equation (8) can be written as follows:

$$l(\mu_\theta, \sigma^2; y) = -\frac{1}{2}\sum_{i=1}^{k}\ln(\sigma^2_\theta + \sigma^2_{\epsilon_i}) - \frac{1}{2}\sum_{k=1}^{k}\frac{(y_i - \mu_\theta)^2}{\sigma^2_\theta + \sigma^2_{\epsilon_i}} \quad (19)$$

Setting partial derivatives with respect to μ_θ and σ^2_θ equal to zero and solving the likelihood equations for the two parameter to be estimated, we obtain the mean and variance, i.e. $\hat{\mu}_\theta$ and $\hat{\sigma}^2_\theta$, of the size effects. Solution for the maximum likelihood (ML) method can be obtained by iterating between $\hat{\mu}_\theta$ and $\hat{\sigma}^2_\theta$.

However, the maximum-likelihood estimator of σ^2_θ tend to underestimate the population heterogeneity in finite samples by failing to account for the loss in degree of freedom that results from estimating β [35, 39, 40]. The restricted maximum likelihood method is based on the likelihood of a vector whose components are independent linear combinations of the observations. The basic idea is to end up with a random vector that contains all the information on the variance components but no longer contains information of the fixed effect parameters.

Denoting $r = rank(X)$, $r = 1$ for meta-analysis random-effects model. Denote $K \in R^{(N-r)\times n}$ as a matrix of full ranks, and $E(Ky) = 0$. Since $E(Ky) = KX\beta$ and $y \sim N(X\beta, V)$, we also have $KX = 0$ and $KY \sim N(0, KVK^T)$. For instance, $M = I - X(X^TX)^{-1}X^T$ would be a good example. The log-likelihood function for Ky is given by

$$l(\sigma^2; y) = -\frac{1}{2}\ln|V| - \frac{1}{2}\ln|X^TV^{-1}X| - \frac{1}{2\sigma^2}(y - X\hat{\beta})^TV^{-1}(y = X\hat{\beta}) \quad (20)$$

This simplifies to

$$l(\sigma^2_\theta; y) = -\frac{1}{2}\sum_{i=1}^{k}\ln(\sigma^2_\theta + \sigma^2_{\epsilon_i}) - \frac{1}{2}\ln\sum_{i=1}^{k}\frac{1}{\sigma^2_\theta + \sigma^2_{\epsilon_i}} - \frac{1}{2}\sum_{i=1}^{k}\frac{(y_i - \hat{\mu}^{(ML)}_\theta)^2}{\sigma^2_\theta + \sigma^2_{\epsilon_i}} \quad (21)$$

The REML estimator of σ^2_θ is then given by

$$\hat{\sigma}^2_\theta = \frac{\sum_{i=1}^{k} w_i^2[(y_i - \mu_\theta)^2 - \sigma^2_{\epsilon_i}]}{\sum_{i=1}^{k} w_i^2} + \frac{1}{\sum_{i=1}^{k} w_i} \quad (22)$$

and

$$w_i = \frac{1}{\hat{\sigma}^2_\theta + \sigma^2_{\epsilon_i}}$$
$$\hat{\mu}_\theta = \frac{\sum_{i=1}^{k} w_i y_i}{\sum_{i=1}^{k} w_i} \quad (23)$$

Again, the variance and mean are obtained in the same iterative manner as described for the regular maximum likelihood estimator.