Reviewing and assessing existing meta-analysis models and tools

Adaptively Weighted (AW) Fisher's method [6]:

The adaptive weighted statistics can be derived following the step by step calculations below:

1. Suppose we have a total of G genes in a differentially expressed data across j studies. Define the observed weighted statistics as

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

and similarly

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

where w_g is the corresponding weight for gene g across J studies, that is, $w_g = (w_{g1}, \ldots, w_{gJ})$ and P_{gj} is the p-value of gene g in the jth study with $1 \leq g \leq G$, $1 \leq j \leq J$ and $1 \leq b \leq B$, where B is the total number of times permutation occurs in each study.

2. Estimate the p-value of the observed weighted statistics of a given gene g and weight w_g as,

$$pU\left(u_g(w_g)\right) = \frac{\sum_{b=1}^{B}\sum_{g'=1}^{G} I\{u_{g'}^{(b)}(w_g) \geq u_g(w_g)\}}{B \cdot G}$$

Likewise compute,

$$pU\left(u_g^{(b)}(w_g)\right) = \frac{\sum_{b'=1}^{B}\sum_{g'=1}^{G} I\{u_{g'}^{(b')}(w_g) \geq u_g^{(b)}(w_g)\}}{B \cdot G}$$

3. Using a) and b), calculate the optimal weight as

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

and likewise

$$w_g^{(b)*} = \arg\min_{w_g \in W} pU\left(u_g^{(b)}(w_g)\right)$$

where $W = \{w \mid w_g \in \{0,1\}\}$. This optimal weight gives indication of which studies contribute to the statistical significance or differentially expressed evidence of the meta-analysis. Hence, the adaptive weighted statistics

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

is the p-value of the minimum p-value among all possible weights. Similarly,

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

4. Next, evaluate the corresponding p-value of the AW-statistics S_g as,

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

5. Also, evaluate the corresponding q-value of the AW-statistics S_g as,

$$q\left(S_g\right) = \frac{\hat{\pi_0}\sum_{b=1}^{B}\sum_{g'=1}^{G} I\{S_{g'}^{(b)} \leq S_g\}}{B\sum_{g'=1}^{G} I\{S_{g'} \leq S_g\}}$$

where

$$\hat{\pi_0} = \frac{\sum_{g=1}^{G} I\{p(V_g) \in A\}}{l(A) \cdot G}$$

is the estimated proportion of null genes with $A = [0.5, 1]$ and $l(A) = 0.5$ [6, 8] where $I\{\cdot\}$ is an indicator function. Hence, the detected genes will be the list of all genes whose $q(V_g) \leq 0.05$.

For example, let's assume from four different studies, gene D is with p-values $= (1, 1, 1e^{-4}, 1)$ and gene E is with p-values $= (1e^{-2}, 1e^{-2}, 1e^{-2}, 1e^{-2})$, then the adaptively weighted Fisher's method will generate adaptive weights $w = (0, 0, 1, 0)$ for gene D indicating statistical significance only in the third study and $w = (1, 1, 1, 1)$ for gene E indicating statistical significance in all four studies.

Fixed effect model (FE) [1, 9]:

FE model assumes that all the studies under consideration shares a common/fixed true effect size. 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 inverse-variance-weighted effect-size estimator of the true effect size is given by

$$\overline{T} = \frac{\sum_{j=1}^{J} w_j T_j}{\sum_{j=1}^{J} w_j}.$$

The variance of $\overline{T}$ is calculated as,

$$v = \frac{1}{\sum_{j=1}^{J} w_j},$$

and its standard error,

$$SE(\overline{T}) = \sqrt{v}.$$

By default, $\overline{T}$ also follows a normal distribution and the test statistics can therefore be calculated as

$$S_{FE} = \frac{\overline{T}}{SE(\overline{T})} = \frac{\sum_{j=1}^{J} w_j T_j}{\sqrt{\sum_{j=1}^{J} w_j}}$$

which follows $\mathcal{N}(0,1)$ under the null hypothesis that there is no association. If we assumed 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. This model is mostly applied to GWAS dataset [1, 9].

Random effect model (RE) [1, 9]:

Unlike the FE model, the RE model assumes that the effect size for each study in the meta-analysis are different and the effect sizes are drawn from a normal distribution with mean μ and variance σ^2. The RE analysis approach is to decompose the observed variance into its two component parts, i.e., the within-study and between-study variance. Using the Cochran's Q test statistics to evaluate the between-study heterogeneity as explained earlier, that is,

$$\begin{aligned} Q &= \sum_{j=1}^{J} w_j (T_j - \overline{T})^2 \\ &= \sum_{j=1}^{J} w_j T_j^2 - \frac{\left[\sum_{j=1}^{J} w_j T_j\right]^2}{\sum_{j=1}^{J} w_j} \end{aligned}$$

The between-study variance

$$\tau^2 = \begin{cases} \frac{Q-df}{c}, & \text{if } Q > df \\ 0, & \text{if } Q \leq df \end{cases}$$

where

$$c = \sum w_j - \frac{\sum w_j^2}{\sum w_j}.$$

Similarly, in RE model the weight assigned to each study,

$$w_j^* = v_j^{*-1}$$

, where

$$v_j^* = v_j + \tau^2$$

with v_j and τ^2 the within-study and between-study variance respectively. Also, the inverse-variance-weighted effect-size estimator of the true effect size is given by

$$\overline{T}^* = \frac{\sum_{j=1}^{J} w_j^* T_j}{\sum_{j=1}^{J} w_j^*}.$$

The variance of $\overline{T}^*$ is calculated as,

$$v^* = \frac{1}{\sum_{j=1}^{J} w_j^*},$$

and its standard error,

$$SE(\overline{T}^*) = \sqrt{v^*}.$$

The test statistics is similarly calculated as

$$S_{RE} = \frac{\overline{T}^*}{SE(\overline{T}^*)} = \frac{\sum_{j=1}^{J} w_j^* T_j}{\sqrt{\sum_{j=1}^{J} w_j^*}}$$

and the p-value is

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

for one-tailed test and

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

for a two-tailed test, where Φ is the standard normal cumulative distribution function. This model is mostly applied to GWAS dataset [1, 9].

Binary effect model (BE) [1]:

BE is a new type of random effect model of meta-analysis that captures studies with or without an effect together. This model is the weighted sum of z-scores method where the *m-values*, the posterior probability that effect exists in each study of a meta-analysis, are incorporated into the weights. More likely, it assigns more weight to studies predicted to have an effect and lesser weight to the studies predicted not to have an effect. So let $z_j = \frac{T_j}{\sqrt{v_j}}$ be the z-score of the *jth* study. Just as the test statistics of FE model can be written in the form of weighted sum of z-score, that is,

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

The binary effect model statistics is given as

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

where the weight $\sqrt{w_j} \approx \sqrt{Np(1-p)}$ with N the sample size and p the effect size (minor allele) frequency and $\sqrt{w_j} \approx \sqrt{N}$ when the effect size is the same between studies. m_j is the corresponding m-value of study j and it is given as

$$m_j = \frac{\pi N(T_j; \mu, v_j)}{(1-\pi)N(T_j; 0, v_j) + \pi N(T_j; \mu, v_j)}$$

where π is the prior probability that each study will have an effect, i.e.,

$$\pi = P(X_j = 1)$$

where X_j is a random variable such that

$$X_j = \begin{cases} 1, & \text{if study} j \text{ has an effect} \\ 0, & \text{if if study} j \text{ does not have an effect} \end{cases}$$

and an assumption of a beta prior is made on π, that is, $\pi \sim Beta(\alpha, \beta)$ where α and β can be chosen but most likely $\alpha = 1$ and $\beta = 1$. The other terms are explained thus, if there is no effect in study j,

$$P(T_j|noeffect) = N(T_j; 0, v_j),$$

and

$$P(T_j|effect) = N(T_j; \mu, v_j),$$

if there is effect in study j. $N(\cdot)$ is the probability density function of a normal distribution and μ is the unknown true effect size. This model is mostly applied to GWAS dataset [1].

Bayesian meta-analysis:

Bayesian approach meta-analysis assumes that a given population cluster with the same ethnic group will possibly shares the same effect size , but there is difference in effect sizes among different population clusters. 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. 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 Bayes' factor

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

where

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

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 as discussed below. The likelihood

$$f(\mathbf{b}, \mathbf{s}|\theta) = f(\mathbf{b}, \mathbf{s}|\beta) = \prod_{j=1}^{J} f(b_j, s_j|\theta_j),$$

and

$$f(b_j, s_j|\theta_j) \propto \frac{1}{s_j} \exp\left[-\frac{(b_j - \beta_j)^2}{2s_j^2}\right]$$

However, the population specific allelic effect β is determined by assigning the populations to ethnic cluster, and this is based on the assumption that effect sizes are likely to vary among different ethnic groups. Hence,

$$f(b_j, s_j|\theta_j) = f(b_j, s_j|K, \mathbf{C}, \psi) \propto \frac{1}{s_j} \exp\left[-\frac{(b_j - \sum_{k=1}^{K} T_{jk}\psi_k)^2}{2s_j^2}\right]$$

where $\mathbf{C} = \{C_1, C_2, \ldots, C_K\}$ is the cluster center, ψ is the corresponding allelic effect and the tessellation $T_{jk} = 1$ if population P_j is assigned to the cluster with center C_k and 0, otherwise.

The prior density function $f(\theta|M_0)$, of parameters under null model M_0 is given by,

$$f(\theta|M_0) = \begin{cases} 1, & \text{if } \beta = 0 \\ 0, & \text{otherwise} \end{cases}$$

while under the alternative model, M_1,

$$f(\theta|M) \propto f(K)(N-K)!\frac{\exp(-\sigma)}{\sigma}\prod_{k=1}^{K}\exp\left(-\frac{(\psi_k-\mu)^2}{2\sigma^2}\right)$$

where

$$f(K) = \begin{cases} \frac{1}{2}, & \text{if } K = 1 \\ \frac{2^{N-1}}{2^K(2^{N-1}-1)}, & \text{otherwise} \end{cases}$$

N is the total number of different populations and the cluster allelic effect have a prior $N(\mu, \sigma)$ distribution, independent of $\mathbf{C}$, where μ has a prior uniform distribution and σ has a prior exponential distribution with expectation 1. Since the marginal likelihood $f(\mathbf{b}, \mathbf{s}|M)$ can't be evaluated directly, the joint posterior density of

$$f(\theta|\mathbf{b}, \mathbf{s}, M) \propto f(\mathbf{b}, \mathbf{s}|\theta)f(\theta|M)$$

is considered instead and it is approximated using the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm.

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. This model is mostly applied to GWAS dataset [7].

RankProd (RP) & RankSum (RS) methods:

Suppose we have a total of G genes in a 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 (in a decreasing/increasing order if we are interested in up-regulated/down-regulated genes respectively). On one hand, considering a single replicate, it follows that under the null hypothesis (no differentially expressed genes present in the dataset), the rank of a gene in the list generated comes from a uniform distribution, that is,

$$P(r_i = g) = \frac{1}{G}$$

where $g \in \{1, \ldots, G\}$. On the other hand, while considering J replicates, it is expected that not all replicates will have same gene at the top of there list. Thus, the probability of a gene being ranked first in each replicate is $\frac{1}{G^J}$.

Hence, 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 whereby the datasets we are analysing is an unpaired dataset. For example having treatment (T) vs control (C) experiment. The RP/RS is performed using the following algorithm. Suppose we have n studies with (n_{iT}, n_{iC}) replicates where $i = 1, 2, \ldots, J$.

1. Form J pairwise comparison, i.e., $J_i = n_{iT} \times n_{iC}$ comparison and evaluate J list of pairwise ratios within each study of their fold changes, i.e., $\frac{T_{ip}}{C_{iq}}$ where $p = 1, 2, \ldots, n_{iT}$ and $q = 1, 2, \ldots, n_{iC}$.

2. Rank the ratios within each comparison with the largest $\implies$ *rank 1*, where r_{gi} is the rank of the g^{th} gene in the i^{th} comparison.

3. Determine the RP/RS for each gene as;

$$RP_g = \left(\prod_i r_{gi} \right)^{\frac{1}{J}},$$

and

$$RS_g = \frac{(\sum_i r_{gi})}{J},$$

respectively, with $J = J_1 + J_2 + \cdots + J_n$.

4. Independently permute expression value within each single array relative to gene ID, repeat step (a)-(c) and obtain the statistic $RP_g^{(b)}$

5. Repeat step (d) B times, form reference distribution with $RP_g^{(b)}$, $(b = 1, \ldots, B)$ and $(g = 1, \ldots, G)$, determine p-value and FDR associated with each gene using the formula given below:

$$p_g = \frac{\sum_b \sum_g I\left(|RP_g^{(b)}| \leq RP_g\right)}{G \cdot B}$$

$$FDR_g = \frac{\sum_b \sum_g I\left(|RP_g^{(b)}| \leq RP_g\right)}{B \sum_g I\left(|RP_g^{(b)}| \leq RP_g\right)}$$

where $I(\cdot)$ is an indicator function. This model is mostly applied to gene expression dataset [3, 4]

Heterogeneity

Heterogeneity is the difference in effect sizes between studies. This can be caused by so may factors including; *genetic factor* as a result of difference in the populations between studies, *environmental factor* due to using subjects from different regions, *design factor* which causes statistical heterogeneity when the true effect size is invariant, heterogeneity shown by collected markers as a result of difference in the linkage disequilibrium structures between studies and also heterogeneity caused by different imputation accuracies and different genotyping errors when studies use different genotyping platforms.

Once this factors has been identified in meta-analysis result, it will be easier to know what approach to take by correctly interpreting the cause of heterogeneity which may lead to a better understanding of the disease mechanism in the case of genetic or environmental factor or, designing a more effective replication study in the case of statistical heterogeneity caused by the design factor. The existence of true heterogeneity among studies can be determined using different test statistics such as, the Cochran's Q test statistics [2, 5] , that is,

$$Q = \sum_{j=1}^{J} w_j (T_j - \overline{T})^2$$

where J is the number of studies, w_j is the inverse of the within-study variance for the jth study, T_j is the effect size estimate of study j and $\overline{T} = \frac{\sum_{j=1}^{J} w_j T_j}{\sum_{j=1}^{J} w_j}$ is the inverse-variance-weighted effect-size estimator of the true effect size. Q test statistic follows a χ^2 distribution with degree of freedom $df = J - 1$ under the assumption of genetic homogeneity. Another alternative heterogeneity test statistics which were reported to be more robust than Q when considering small number of studies are; the I^2 index [2, 5] ,

$$I^2 = \frac{[Q - (J-1)]}{Q} \times 100\%$$

$$I^2 = \begin{cases} \frac{[Q-(J-1)]}{Q} \times 100\%, & \text{if } Q > df \\ 0, & \text{if } Q \leq df \end{cases}$$

the H statistics [2],

$$H = \sqrt{\frac{Q}{J-1}}$$

and the R statistics [2],

$$R = \sqrt{\frac{\sum w_j}{\sum w_j^*}} = \sqrt{\frac{\sum w_j}{\sum (w_j^{-1} + \tau^2)}}$$

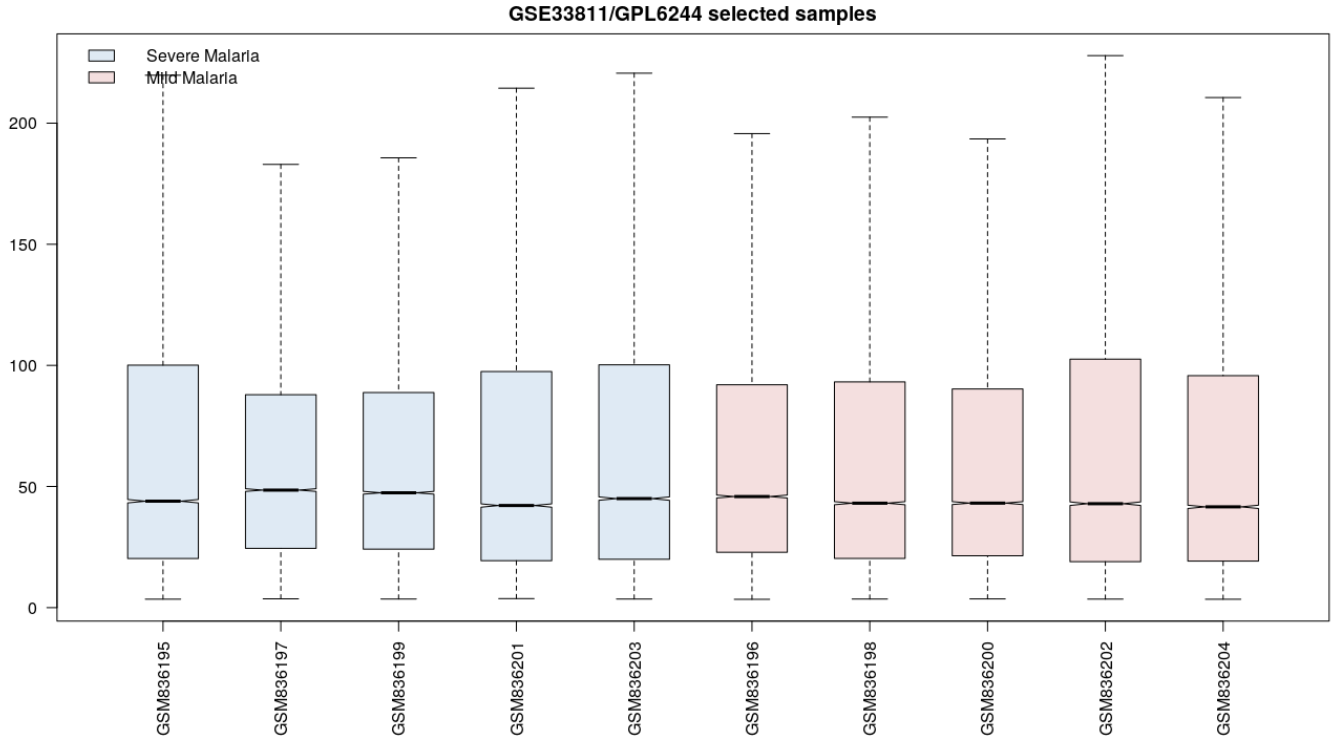
where the between-study variance

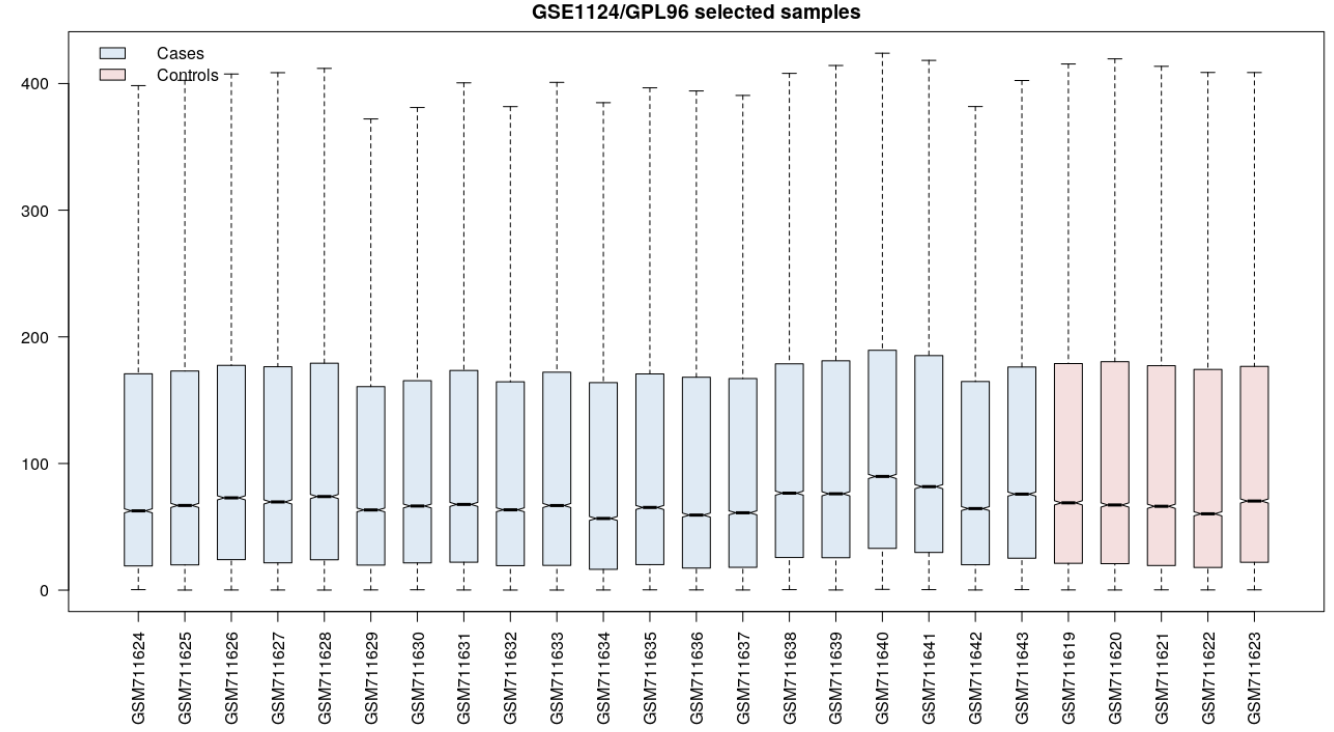
$$\tau^2 = \begin{cases} \frac{Q-df}{c}, & \text{if } Q > df \\ 0, & \text{if } Q \leq df \end{cases}$$

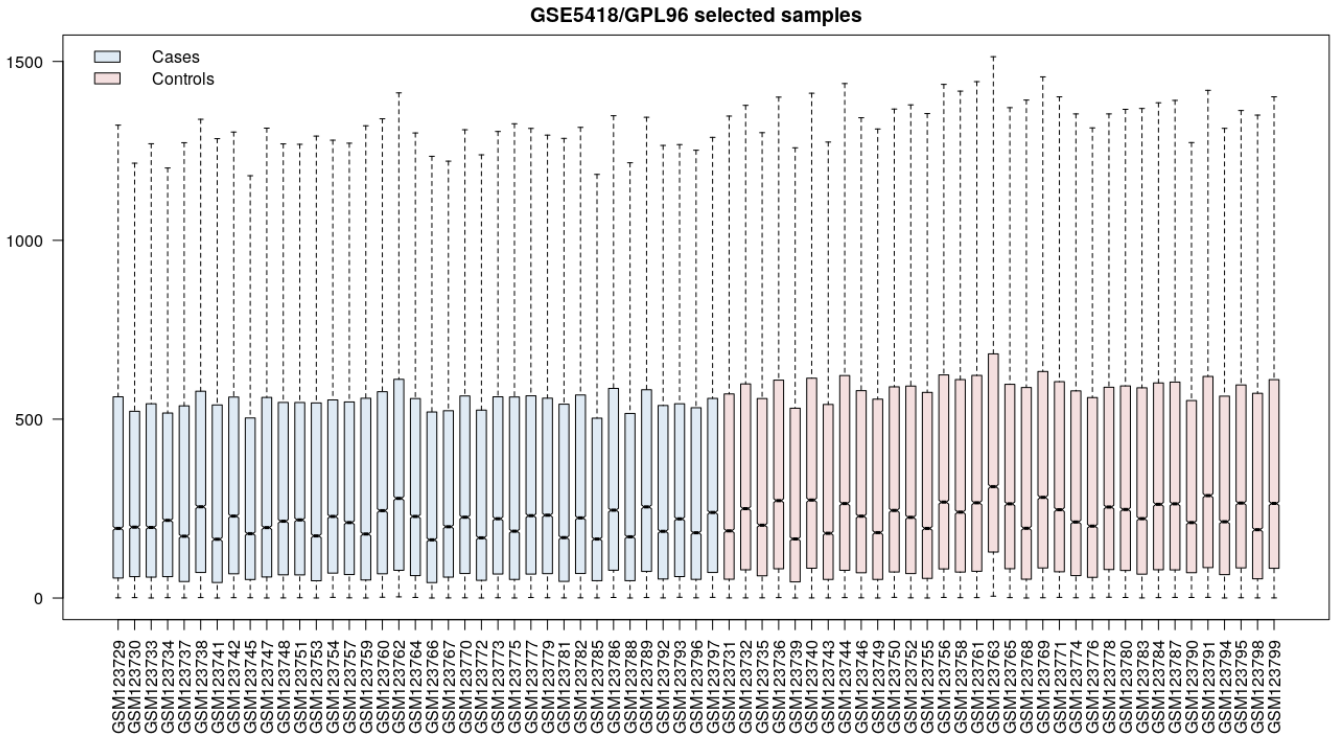
where

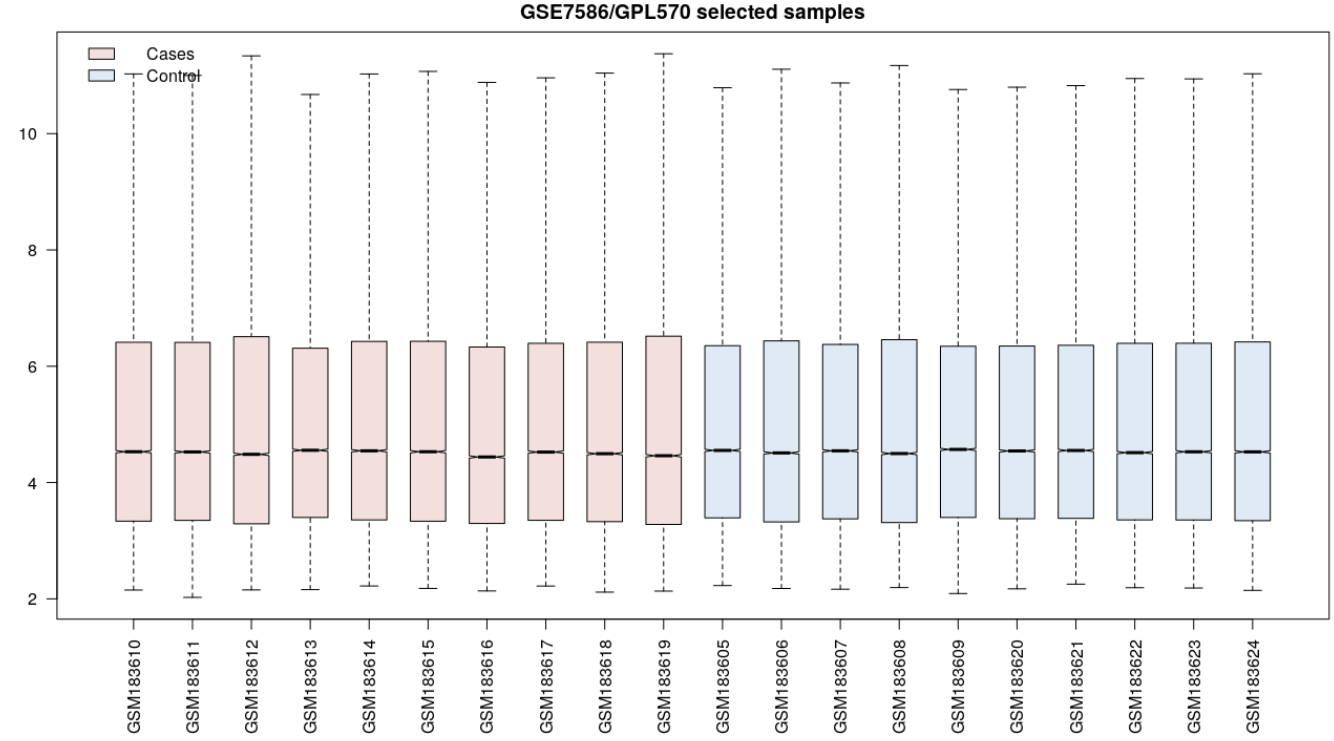
$$c = \sum w_j - \frac{\sum w_j^2}{\sum w_j}.$$

Box plot of Malaria Study

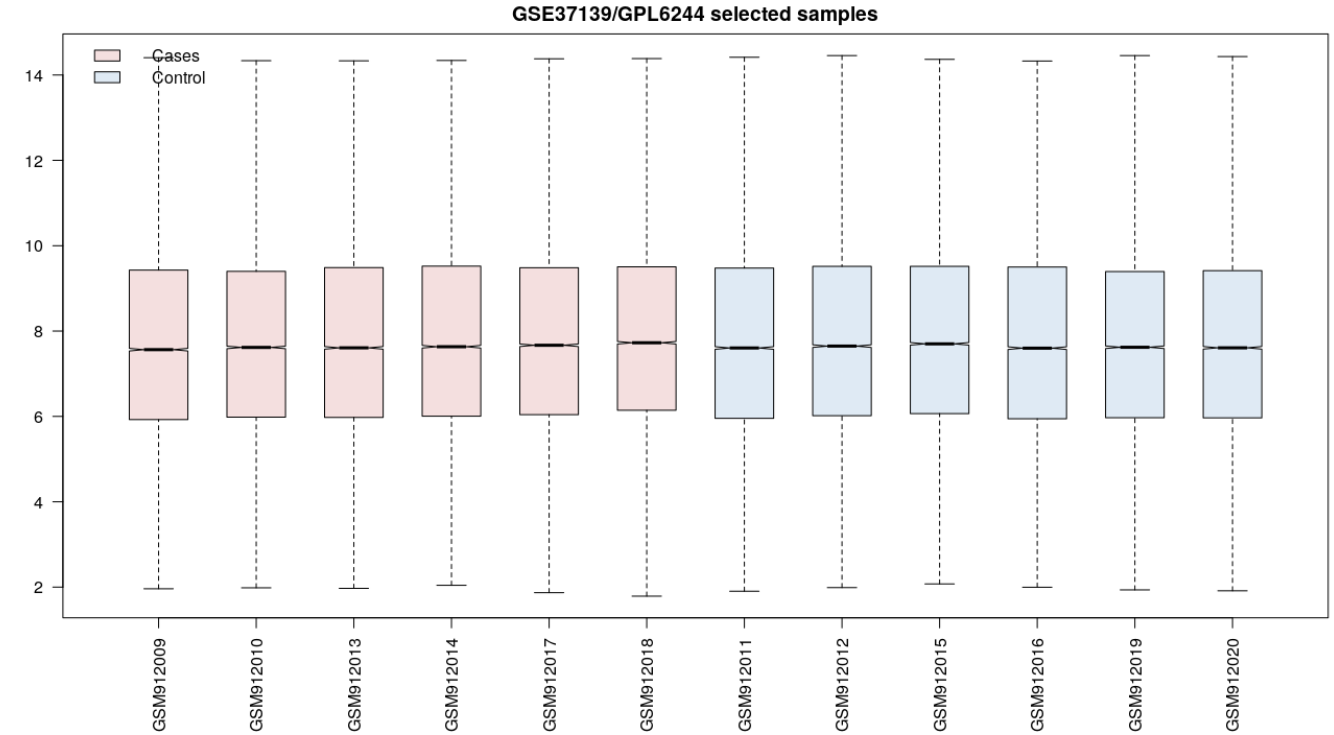

GSE33811/GPL6244 selected samples
Severe Malaria
Mild Malaria
200
150
100
50
0
GSM836195
GSM836197
GSM836199
GSM836201
GSM836203
GSM836196
GSM836198
GSM836200
GSM836202
GSM836204

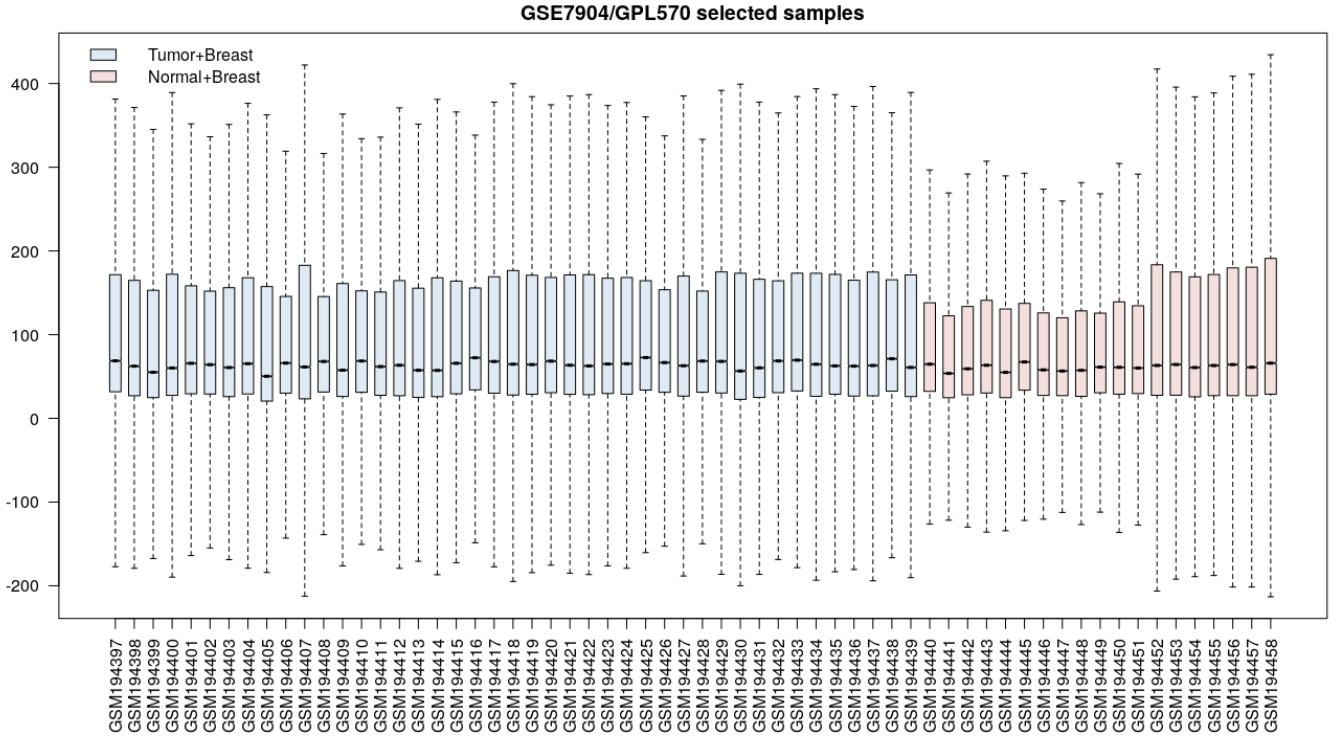

GSE1124/GPL96 selected samples
Cases
Controls
400
300
200
100
0
GSM711624
GSM711625
GSM711626
GSM711627
GSM711628
GSM711629
GSM711630
GSM711631
GSM711632
GSM711633
GSM711634
GSM711635
GSM711636
GSM711637
GSM711638
GSM711639
GSM711640
GSM711641
GSM711642
GSM711643
GSM711619
GSM711620
GSM711621
GSM711622
GSM711623

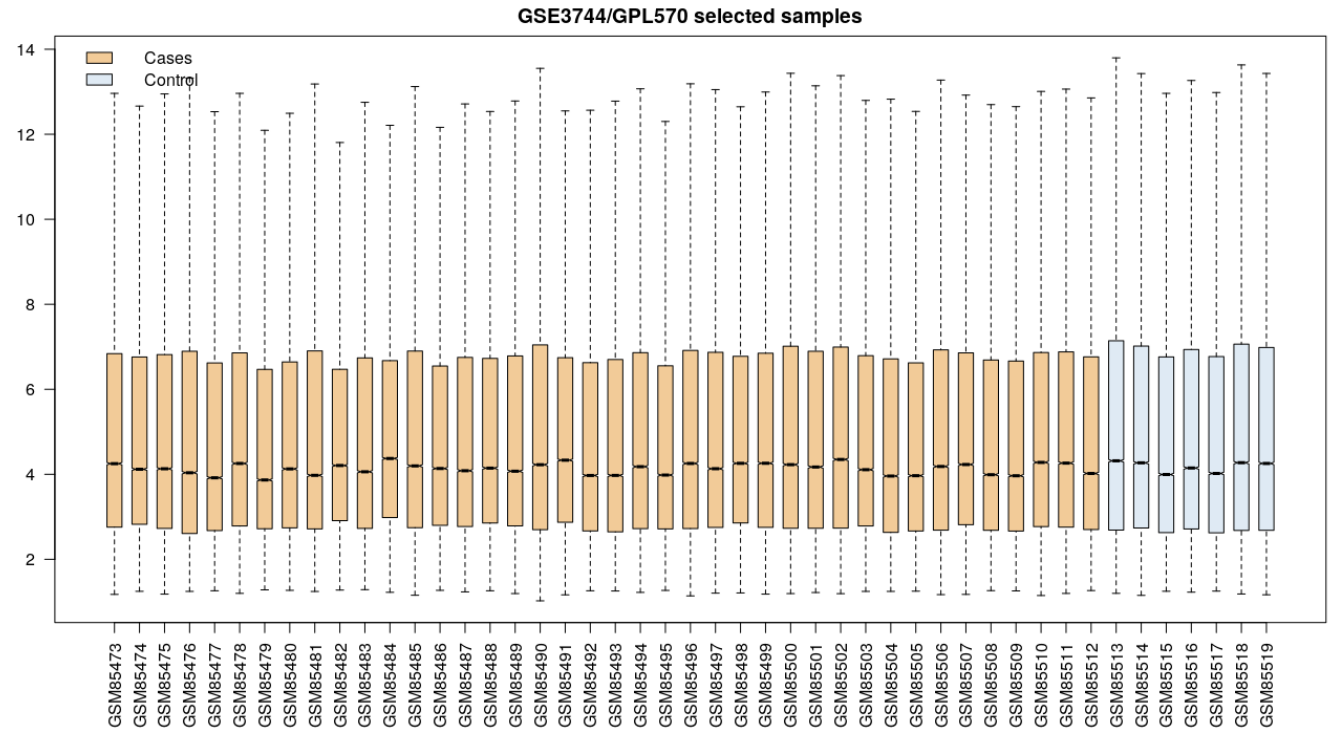


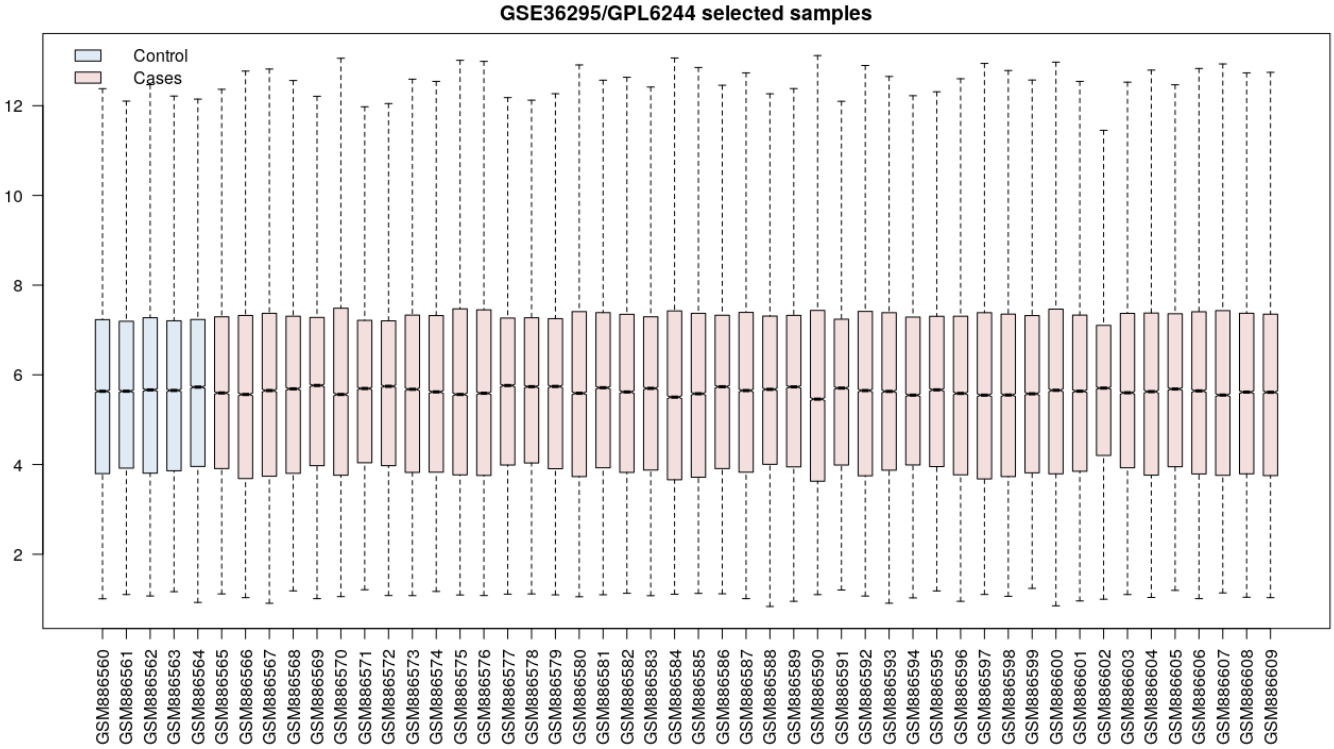


Box plot of Breast Cancer Study


GSE37139/GPL6244 selected samples
Cases
Control
14
12
10
8
6
4
2
GSM912009
GSM912010
GSM912013
GSM912014
GSM912017
GSM912018
GSM912011
GSM912012
GSM912015
GSM912016
GSM912019
GSM912020


GSE7904/GPL570 selected samples
Tumor+Breast
Normal+Breast
400
300
200
100
0
-100
-200
GSM194397
GSM194398
GSM194399
GSM194400
GSM194401
GSM194402
GSM194403
GSM194404
GSM194405
GSM194406
GSM194407
GSM194408
GSM194409
GSM194410
GSM194411
GSM194412
GSM194413
GSM194414
GSM194415
GSM194416
GSM194417
GSM194418
GSM194419
GSM194420
GSM194421
GSM194422
GSM194423
GSM194424
GSM194425
GSM194426
GSM194427
GSM194428
GSM194429
GSM194430
GSM194431
GSM194432
GSM194433
GSM194434
GSM194435
GSM194436
GSM194437
GSM194438
GSM194439
GSM194440
GSM194441
GSM194442
GSM194443
GSM194444
GSM194445
GSM194446
GSM194447
GSM194448
GSM194449
GSM194450
GSM194451
GSM194452
GSM194453
GSM194454
GSM194455
GSM194456
GSM194457
GSM194458


GSE3744/GPL570 selected samples
Cases
Control
14
12
10
8
6
4
2
GSM85473
GSM85474
GSM85475
GSM85476
GSM85477
GSM85478
GSM85479
GSM85480
GSM85481
GSM85482
GSM85483
GSM85484
GSM85485
GSM85486
GSM85487
GSM85488
GSM85489
GSM85490
GSM85491
GSM85492
GSM85493
GSM85494
GSM85495
GSM85496
GSM85497
GSM85498
GSM85499
GSM85500
GSM85501
GSM85502
GSM85503
GSM85504
GSM85505
GSM85506
GSM85507
GSM85508
GSM85509
GSM85510
GSM85511
GSM85512
GSM85513
GSM85514
GSM85515
GSM85516
GSM85517
GSM85518
GSM85519


GSE36295/GPL6244 selected samples
Control
Cases
12
10
8
6
4
2
GSM886560
GSM886561
GSM886562
GSM886563
GSM886564
GSM886565
GSM886566
GSM886567
GSM886568
GSM886569
GSM886570
GSM886571
GSM886572
GSM886573
GSM886574
GSM886575
GSM886576
GSM886577
GSM886578
GSM886579
GSM886580
GSM886581
GSM886582
GSM886583
GSM886584
GSM886585
GSM886586
GSM886587
GSM886588
GSM886589
GSM886590
GSM886591
GSM886592
GSM886593
GSM886594
GSM886595
GSM886596
GSM886597
GSM886598
GSM886599
GSM886600
GSM886601
GSM886602
GSM886603
GSM886604
GSM886605
GSM886606
GSM886607
GSM886608
GSM886609

List of references

[1] Buhm Han and Eleazar Eskin. Interpreting meta-analyses of genome-wide association studies. *PLoS genetics*, 8(3):e1002555, 2012.

[2] Julian PT Higgins and Simon G Thompson. Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11):1539–1558, 2002.

[3] Fangxin Hong and Rainer Breitling. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics*, 24(3):374–382, 2008.

[4] Fangxin Hong, Rainer Breitling, Connor W McEntee, Ben S Wittner, Jennifer L Nemhauser, and Joanne Chory. Rankprod: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics*, 22(22):2825–2827, 2006.

[5] Tania B Huedo-Medina, Julio Sánchez-Meca, Fulgencio Marín-Martínez, and Juan Botella. Assessing heterogeneity in meta-analysis: Q statistic or i^2 index? *Psychological methods*, 11(2):193, 2006.

[6] Jia Li, George C Tseng, et al. An adaptively weighted statistic for detecting differential gene expression when combining multiple transcriptomic studies. *The Annals of Applied Statistics*, 5(2A):994–1019, 2011.

[7] Andrew P Morris. Transethnic meta-analysis of genomewide association studies. *Genetic epidemiology*, 35(8):809–822, 2011.

[8] John D Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.

[9] Xu Wang, Hui-Xiang Chua, Peng Chen, Rick Twee-Hee Ong, Xueling Sim, Weihua Zhang, Fumihiko Takeuchi, Xuanyao Liu, Chiea-Chuen Khor, Wan-Ting Tay, et al. Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Human molecular genetics*, 22(11):2303–2311, 2013.