

# Meta analyses for Omics integration

---

Workshop - Omics Integration and Systems Biology

**Ashfaq Ali**

NBIS, SciLife lab, Lund University, Sweden

updated: 2021-04-20

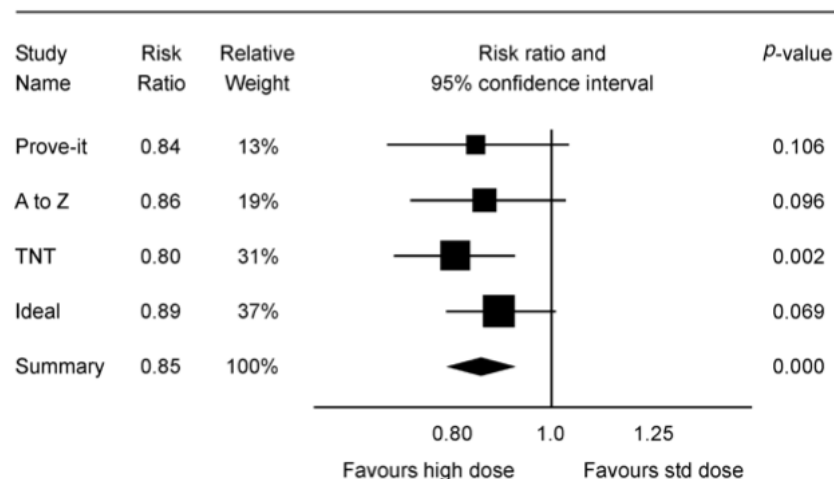
# Meta analyses and its components



An example comparing effect of increased dosage of statins on myocardial infarction.

In conclusive evidence.

Impact of Statin Dose  
On Death and Myocardial Infarction



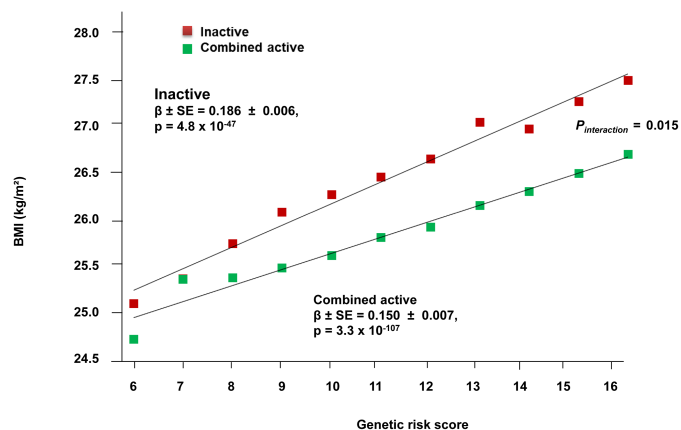
Cannon *et al* 2006

- Individual studies
  - Effect Sizes
  - Precision and Weights
  - P value and confidence intervals
- Summary
  - Effect Size
  - P Value
  - Precision
- Heterogeneity of effect size
  - A measure of consistency (later)

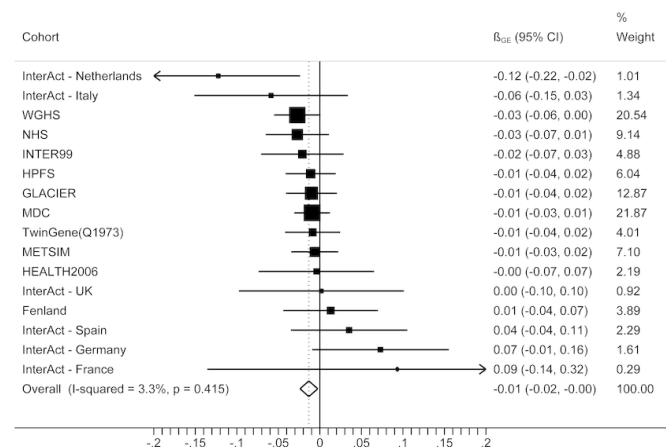
# Meta analyses relevant Example from Genetic Epidemiology

- Hypothesis: Genetic risk score for obesity has more effect on inactive people

$$GRS = \sum_{i=1}^n risk\ alleles$$



- Met Analyses gives non significant results



Ahmed et al/2013

# Why do meta analyses



The goal of meta analyses is to contextualize the results of any study in the context of all the other studies

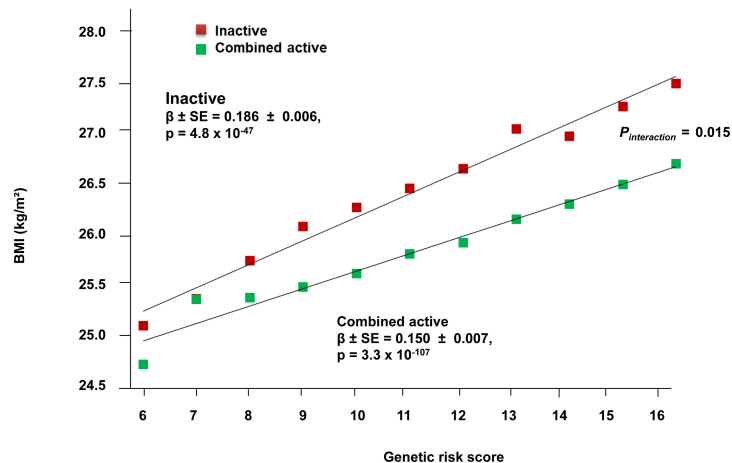
- **Statistical Significance**
  - Is there a significant association?
- **Clinical importance of the effect**
  - estimate the effect size as accurately as possible
  - quantify the extent of the variance and consider the implications
- **Consistency of effects**
  - whether or not the effect size is consistent across the body of data

# Implications of dimensionality in genome wide omics studies

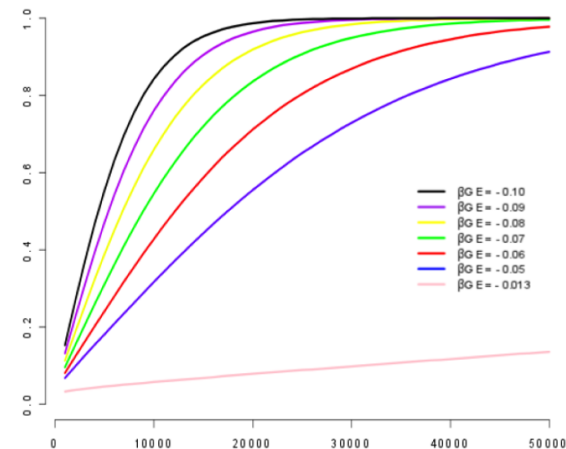


- **Winners Curse** "When statistical power is low, estimates of the odds ratio from a genome-wide association study, or any large-scale association study, will be upwardly biased"
  - The problem is widely known in **GWAS** and **eQTL** studies.

Consider [Ahmed et al.](#)



Effect size = 0.186



X- axis : Sample size

Y- axis :statistical power

# Meta analyses techniques

1. Combining P- values,
2. Combining effects.
3. Combining rank statistics

# p values



Fisher's method: Sum of minus log-transformed  $P$ -values where larger Fisher score reflects stronger aggregated differential expression evidence.

$$p = -2 \times \sum_{i=1}^k \ln (p_i)$$

- Pearson's method

$$p = -2 \times \sum_{i=1}^k \ln (1 - p_i)$$

- Stouffer's Method

$$Z_i = \Phi^{-1} (1 - p_i)$$

$\Phi$  is the standard normal cumulative distribution function.

$$\frac{\sum_{i=1}^k Z_i}{\sqrt{k}}$$

*it allows including weights for the studies. In this case, the statistic is*

- minP:  $\min (p_1, p_2, \dots, p_i, \dots, p_k)$
- maxP:  $\max (p_1, p_2, \dots, p_i, \dots, p_k)$

## 2. Effect size based analyses



### Fixed effect models

FEM combines the effect size across  $K$  studies by assuming a simple linear model with an underlying true effect size plus a random error in each study

$$\overline{T}_{\cdot} = \frac{\sum \omega_i T_i}{\sum \omega_i}$$

### Random effect model

REM extends FEM by allowing random effects for the inter-study heterogeneity in the model.

$$\overline{T}_{\cdot}^* = \frac{\sum_{i=1}^k \omega_i^* T_i}{\sum_{i=1}^k \omega_i^*}$$

$\omega$  are the different weights assigned to each study, that is, the inverse within-study variance  $V(T_i)$  and  $\omega_i = \frac{1}{V(T_i)}$



# Heterogeneity of effect sizes



The statistic that represents the total variance,  $Q$ , is defined as (Cochran's  $Q$ ) which is computed by summing the squared deviations of each study's estimate from the overall meta-analytic estimate

$$Q = \sum_{i=1}^k \omega_i (T_i - \bar{T})^2$$

where  $T_i$  is the observed effect,  $\omega_i$  is the calculated weights for the FEM and  $\bar{T}$  is the combined effect calculated for the FEM

A test for heterogeneity examines the null hypothesis that all studies are evaluating the same effect. The usual test statistic, weighting each study's contribution in the same manner as in the meta-analysis.

P values are obtained by comparing the statistic with a  $\chi^2$  distribution with  $k-1$  degrees of freedom (where  $k$  is the number of studies).

$$I^2 = 100 \times (Q - df) / Q$$

, where  $Q$  is Cochran's heterogeneity statistic and  $df$  the degrees of freedom

# Ranks based meta analyses



- rankProducts

$$RP_g = \prod_i^k r_{ig}$$

- Rank Sums

$$RS_g = \sum_i^K r_{ig}$$

# Meta analyses of Expression based Omics (Motivation)



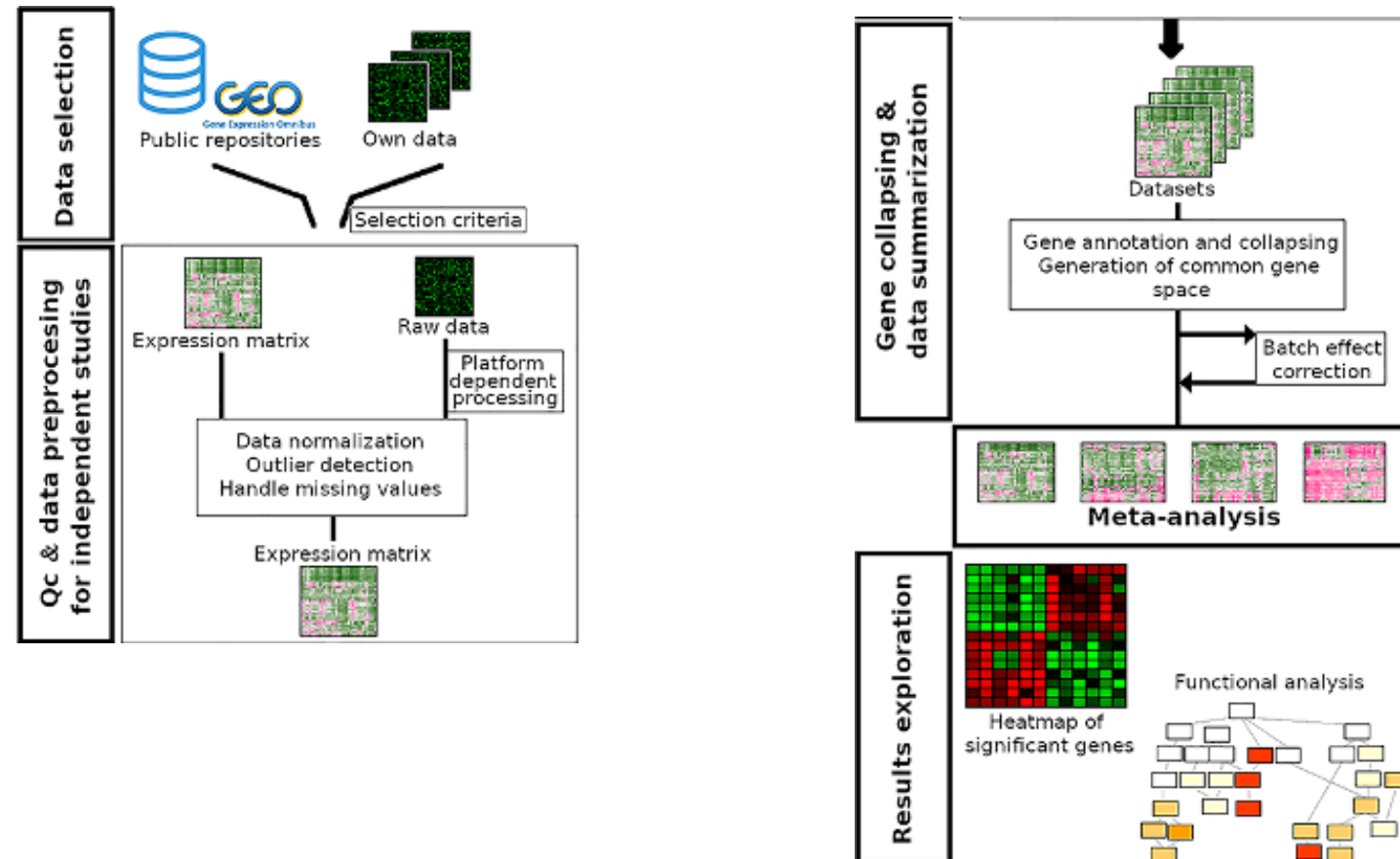
- Vertical Integration
  - Perform analysis of omics data either across experiments on the same samples
  - Very rare in clinical setting
  - Require data from same individuals
- Horizontal integration techniques cross studies on the same variables
  - Meta analyses is one of the techniques
  - Sample sizes small
  - Study specific biases
  - Batch effects

# Data integration using meta analyses for Omics



1. Detect/validate differentially expressed genes
  - Drug targets
  - Disease Biomarkers
2. Detect/validate differentially regulated pathways
3. Detect/ co-expression network

# Workflow for Meta analyses for Differential Expression



- MetaDE: meta analyses based on differential expression
- Meta analyses based on pathways
  - Effect size(Fixed effect, Random effect)
  - P Value
  - Vote counting
- MetaPath: Meta analyses based on pathways
  - At gene level
  - Pathway level
  - Hybrid
- MetaDCN: Meta analyses based on networks
  - Systematically identify co-expression modules and build based on across study consensus.

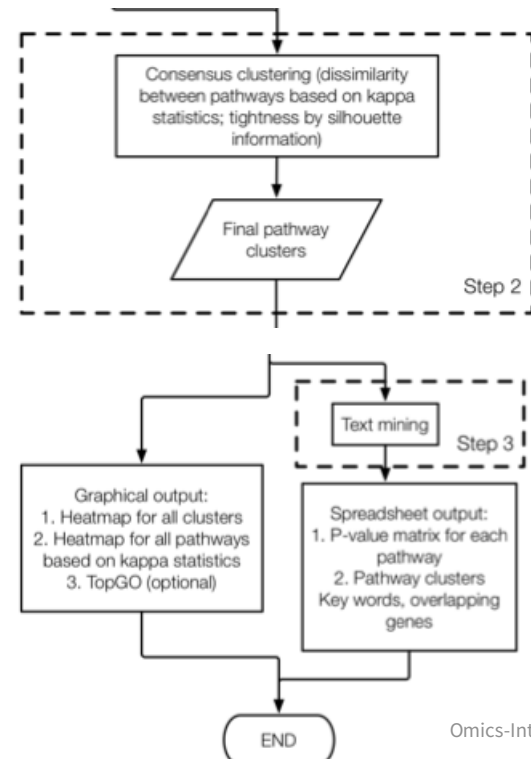
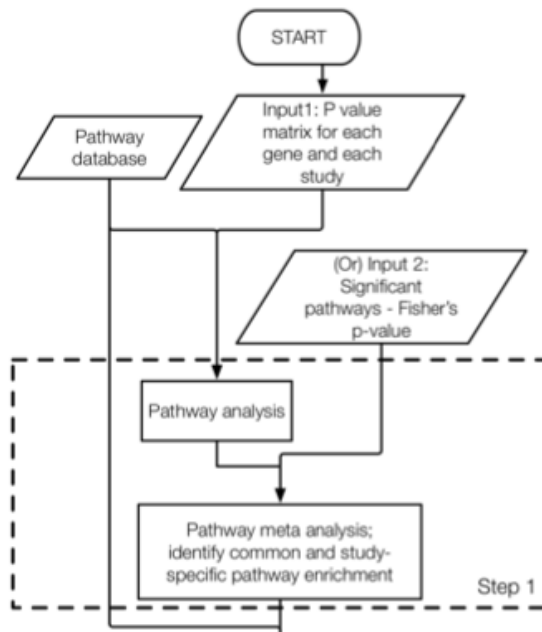
# Meta-Analysis for Pathway Enrichment (MAPE)



- Usual: compare the number of differentially expressed (DE) genes in and outside a pathway using [Fisher's exact test](#)
- Or [Gene Set Enrichment Analysis \(GSEA\)](#)

[MetaPath](#) and [CPI Workflow](#).

Gene level, Pathway level or a hybrid approach.



# Meta-analyses Co-expression networks using MetaDCN

Differential co-expression (DC) refers to the change in gene–gene correlations between two conditions (e.g. cases and controls).

The differential correlation relationship could arise from meaningful biological sources as well as uncorrected technical biases

unwanted batch effect, or mixture of tissues could potentially contribute to co-expression relationships

differential co-expression may be confirmed across multiple datasets via meta-analyses to increase the detection power and stability. DC networks that are significant in one dataset may become more convincing if the DC patterns are preserved across multiple datasets.

## Disambiguation

*Here we are not discussing Network meta-analysis (NMA) which extends principles of meta-analysis to the evaluation of multiple treatments in a single analysis.*



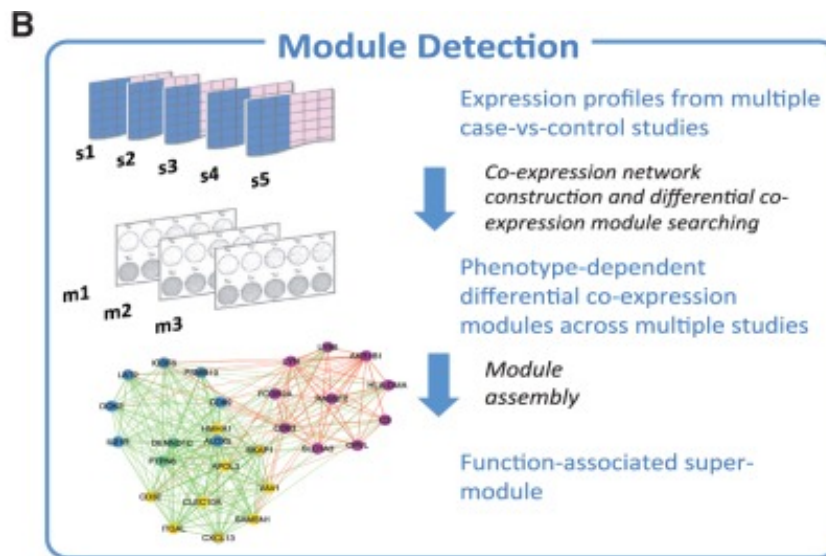
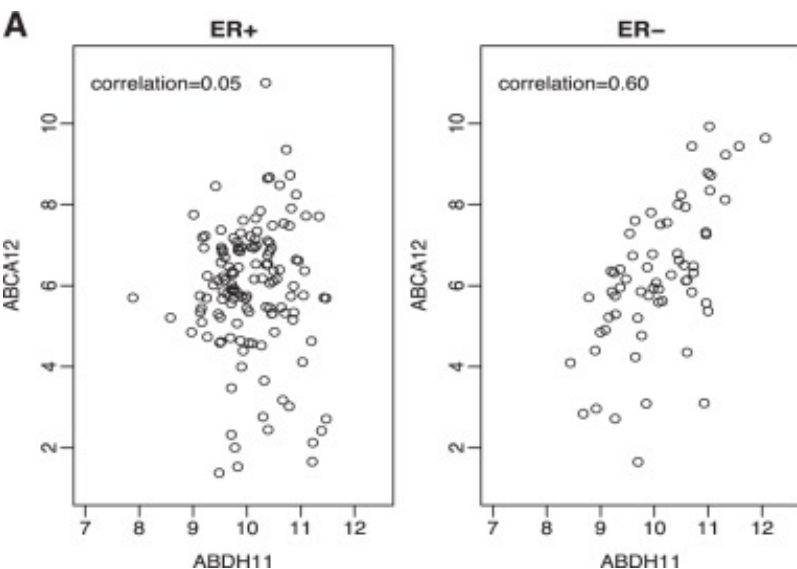
# Steps

## 1. \*Basic DC module detection

- Search for initial DC modules by calculating pair-wise gene-gene Spearman's correlations for robust comparisons
- Optimization by simulated annealing combined basic DC modules that share common pathway annotation into more interportable DC super-modules
- Control of false discovery rate

## 2. Differential Coexpression based on phenotype

## 3. DC supermodule assembly



# End of lecture

