Meta-analysis in Bioinformatics

Roberto Pagliarini

What is Meta-analysis?

Statistical integration of independent studies

Applied in medicine, epidemiology, bioinformatics

Example: combine gene expression results from several microarray studies

Why Meta-analysis in Omics?

OMICS STUDIES = HIGH-DIMENSIONAL, SMALL SAMPLE SIZES

LIMITED REPRODUCIBILITY IN SINGLE STUDIES

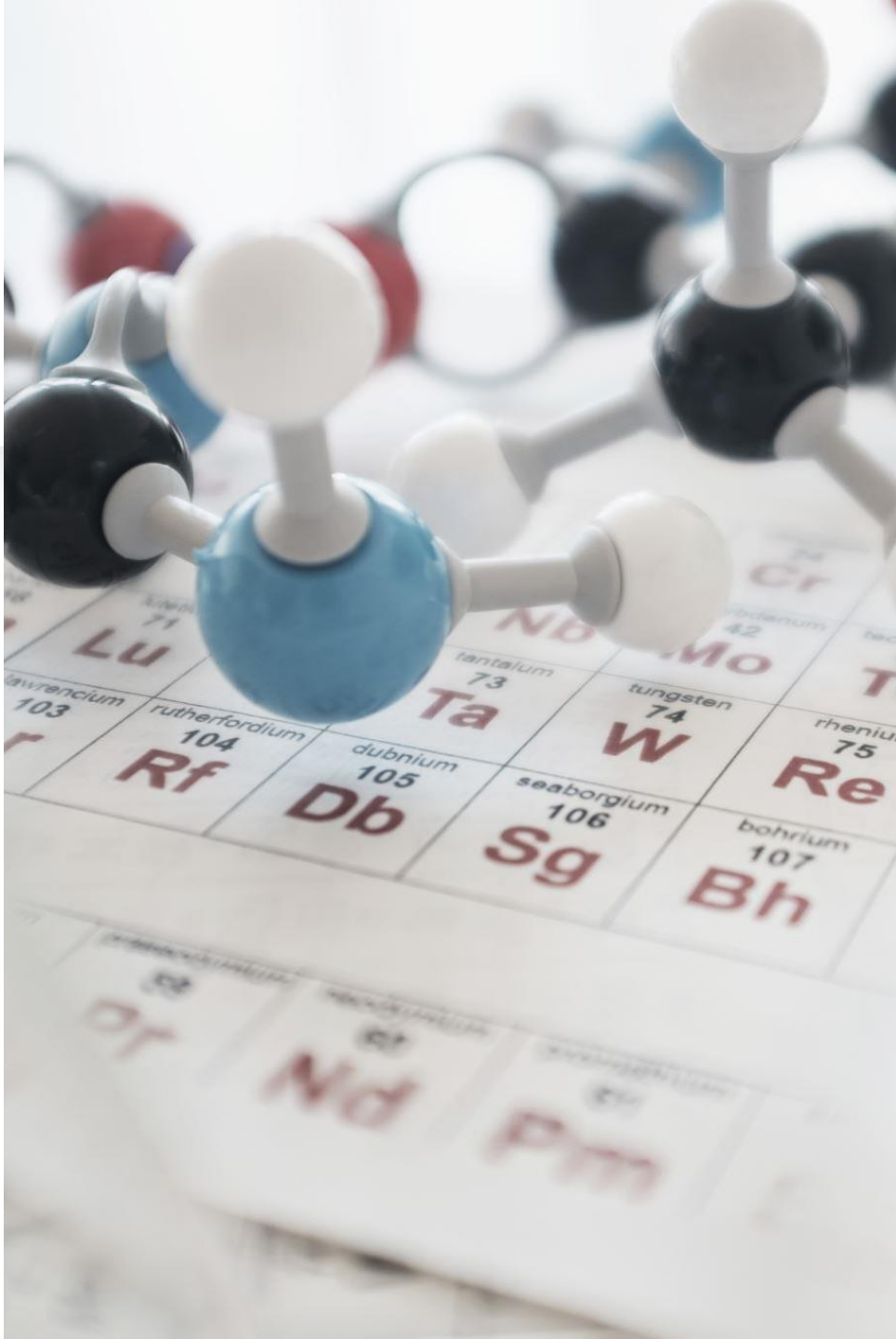
POOLING → ROBUST CONSENSUS FINDINGS

Bioinformatics Context

- Gene expression profiling
- GWAS, proteomics, metabolomics
- Multi-study integration for systems-level insights

Public Databases

- GEO (Gene Expression Omnibus)
- ArrayExpress (expression data)
- TCGA (cancer)
- EGA (genomics + clinical)
- Need preprocessing & harmonization



Challenges in Omics Meta-analysis

- Batch effects across platforms
- Heterogeneous annotations
- Missing phenotypic data
- Requires robust statistical corrections

Key R/Bioconductor Packages

- GEOquery → retrieve datasets
- limma → differential expression
- MetaDE, RankProd, MetaOmics → meta-analysis
- metafor → general meta-analysis framework

Workflow Diagram

1) Download datasets

2) Preprocess & normalize

3) Differential expression per dataset

4) Meta-analysis across datasets

5) Interpret consensus signatures

Case-Study GEO Breast Cancer

- Example datasets
  - GSE2034
  - GSE2990
- Case vs control groups
- Goal: consensus DE gene list

Step 1: Download & Inspect

- library(GEOquery)
- gse1 <- getGEO('GSE2034', GSEMatrix=TRUE)[[1]]
- expr1 <- exprs(gse1)
- pheno1 <- pData(gse1)

Step 2: Differential Expression Analysis

- library(limma)
- design <- model.matrix(~ factor(pheno1$cancer_status))
- fit <- lmFit(expr1, design)
- fit <- eBayes(fit)
- res <- topTable(fit, coef=2)

Step 3: Meta-analysis

- library(MetaDE)
- data.list <- list(dataset1=list(x=expr1, y=group1),
- dataset2=list(x=expr2, y=group2))
- meta.res <- MetaDE.rawdata(data.list, meta.method='Fisher’)
- summary(meta.res)

Step 4: Interpretation

- Consensus DE genes identified
- Use enrichment analysis (clusterProfiler)
- Biological interpretation: pathways, biomarkers