

$$P = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{a! b! c! d! n!}$$

# **PHYLOGENE** for time resolved cost-effective bioinformatics / biostatistics

## **CORAVVALID:**

### ***Complementary analysis of identification and relative quantification data***

#### **Input data**

Output analysis files from **genomics/transcriptomics/proteomics/metabolomics**  
Phylogene has an expertise in using databases and can issue recommendations for  
selecting the best sources for performing this work.

#### **1 Match research**

Research of matches between elements recognized by basic analysis as  
interesting (identification, overexpression, underexpression) and specific  
interest data (such as **metabolic pathways, biological or molecular functions,  
cellular components, interactors ; potential orthologs...**), which is contained in  
specialized databases depending on the request (organisms, diseases,  
interactions databases)

*Querying and retrieving data from databases, using interesting elements  
identifiers used as access keys. Used databases might be for instance Biocyc,  
Gene Ontology, Kegg or Reactome.*





## 2 Group processing and ORA(Over-representation analysis)

Matches processing in order to sort them based on relevant criteria, then to build synthetic data groups known as over-representation groups. Statistical score calculation, allowing to concur on group composition random or determinist nature.

*Statistical techniques used for these calculations are z-scores, Fisher's exact test and adjusted permutation p-values.*

## 3 Complementary validation

Statistical error correction procedure performed in order to enforce reliability, with ORA analytic score modification and final group validation, which results in data concentration to ascertain which of those are relevant. Multiple database results comparison is used to fine tune groups and broaden data range.

*Statistical error correction techniques performed are Bonferroni and Benjamini-Hochberg procedures.*

## Output data

Results are displayed as lists of elements of interest (such as **metabolic pathways, biological or molecular functions, cellular components, interactors; potential orthologs...**) which are linked to sub-groups of primary detection elements, statistical scores being associated with these to allow for their validation.

## Applications:

**Identification or quantification results from genomics/ transcriptomics/ proteomics / metabolomics**

**Drug and active ingredient effect objectivation / Immune activated cells characterization / Biomarker research and validation / Production process perfecting and follow-up / Cell recombination checking / Drug discovery / Target discovery**

## Quality:

**PHYLOGENE** is driven by a Quality Assurance policy validated by an ISO 17025 COFRAC accreditation.

We provide you a unique contact for your project driving, from specifications up to results