Mining Transcriptomics Data on VEuPathDB: An Exercise
Co-Expression of genes in *Aspergillus niger*

VEuPathDB
Eukaryotic Pathogen, Vector & Host
Informatics Resources

The search question

A major obstacle in understanding biological systems lies in the vast number of genes in many genomes whose functions remain unknown. The fungal genus *Aspergillus* includes key human pathogens, model species, and industrially valuable microorganisms. *Aspergillus niger* is widely utilized for the production of organic acids, proteins, and holds potential for discovering novel bioactive secondary metabolites. But only a small minority of predicted open reading frames in the *A. niger* genome have been experimentally validated.

An04g07430 is an *Aspergillus niger* protein coding gene with little functional annotation. By **finding genes that are expressed at the same time as An04g07430**, we may find clues about its function and the biological processes that it participates in.
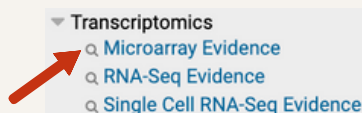

What platform is best suited for this question?

FungiDB
Fungal & Oomycete Informatics Resources

This knowledgebase has data and informatics resources for oomycetes and diverse fungal taxa including *Aspergillus.*

What search is best suited for this question?

▼ Transcriptomics
  🔍 Microarray Evidence
 🔍 RNA-Seq Evidence
  🔍 Single Cell RNA-Seq Evidence

The analysis of gene expression data from **microarray data** can identify groups of genes that exhibit similar expression levels across different conditions.

What is a possible search strategy that will answer this question?

Use the **Microarray Evidence** data and **Co-expression search** to find the co-expression network for the gene of interest. Further, explore the genes in this network using the column histogram tool and enrichment analyses to glean information about the function of this group of co-expressed genes.

1. Navigate to the appropriate search

- Navigate to FungiDB.org
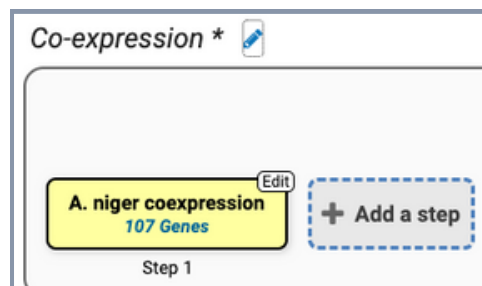- From the sidebar or header, search for "microarray" and click on Microarray Evidence
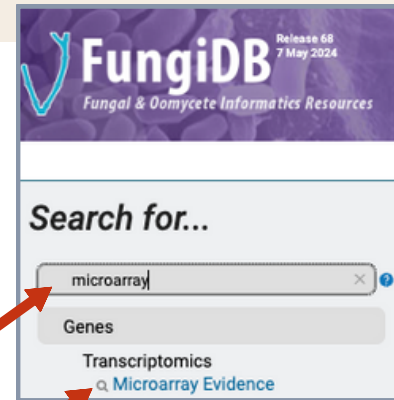
2. Choose the appropriate experiments/ samples & parameters within the search

- Choose the Coexpression search for the data set titled ***Aspergillus niger* gene co-expression network (Vera Meyer).** This data is the result of a meta-analysis (Schape et al Nucleic Acids Research 2019) of 155 publicly available transcriptomics analyses for *A. niger*, which were used to generate a genome-level co-expression network and sub-networks for >9,500 genes.
- Configure the **Coexpression search** to find the co-expression network for the gene **An04g07430**.
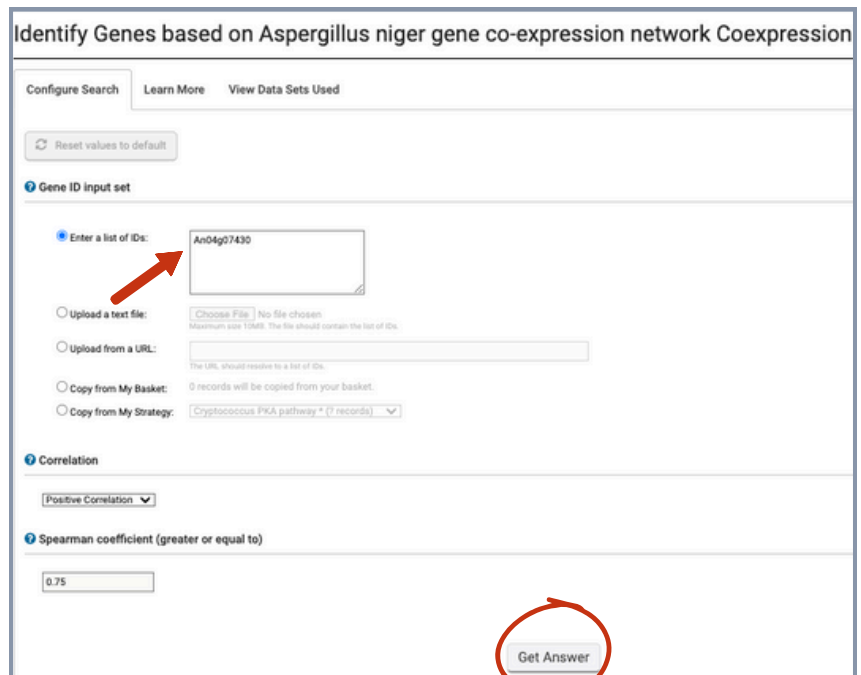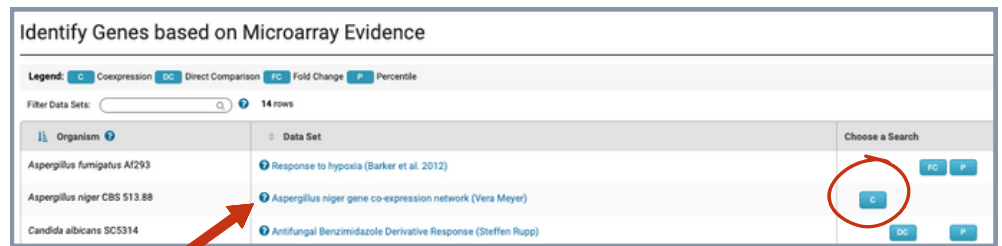
How many genes share the co-expression profile of An04g07430?

- What genes share the co-expression profile of An04g07430? Several genes have a correlation coefficient of 0.85 (see **red circle** above). What are these genes? Visit their gene pages by clicking on the Gene ID link to learn more.
- Scan the product description column (see **red rectangle** above) for genes with known functions.
- Use the Column Histogram tool (see **red arrow** above) to view a word cloud of the product descriptions in the column. Set the rank range to 25–50. **What words occur most often in the product descriptions of An04g07430 co-expressed genes?**
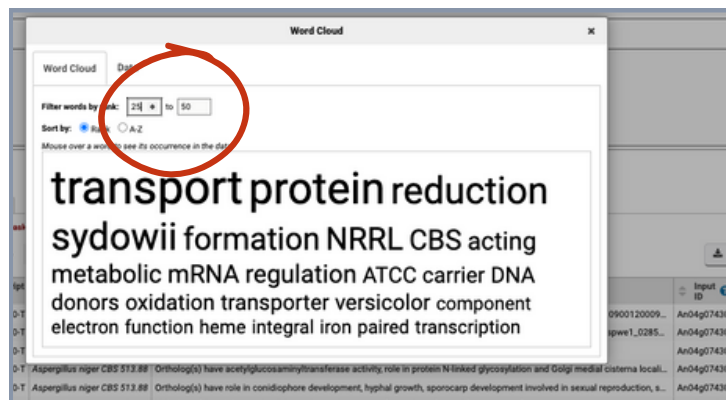
Run the **Gene Ontology Enrichment** analyses for Molecular Function, Cellular Component and Biological Processes. Do these provide information about what this group of co-expressed genes might be doing?

Questions? Comments? Write to help@veupathdb.org