GENE EXPRESSION ANALYSIS IN HEREDITARY DISEASES USING THE TOOL ASACO

AUTOMATIC AND SERIAL ANALYSIS OF CO-EXPRESSION (ASACO)

Author: Silva Escalera, Elena Mª

Academic tutor: Antonio J. Pérez Pulido

Area de Genética, University Pablo de Olavide

INTRODUCTION

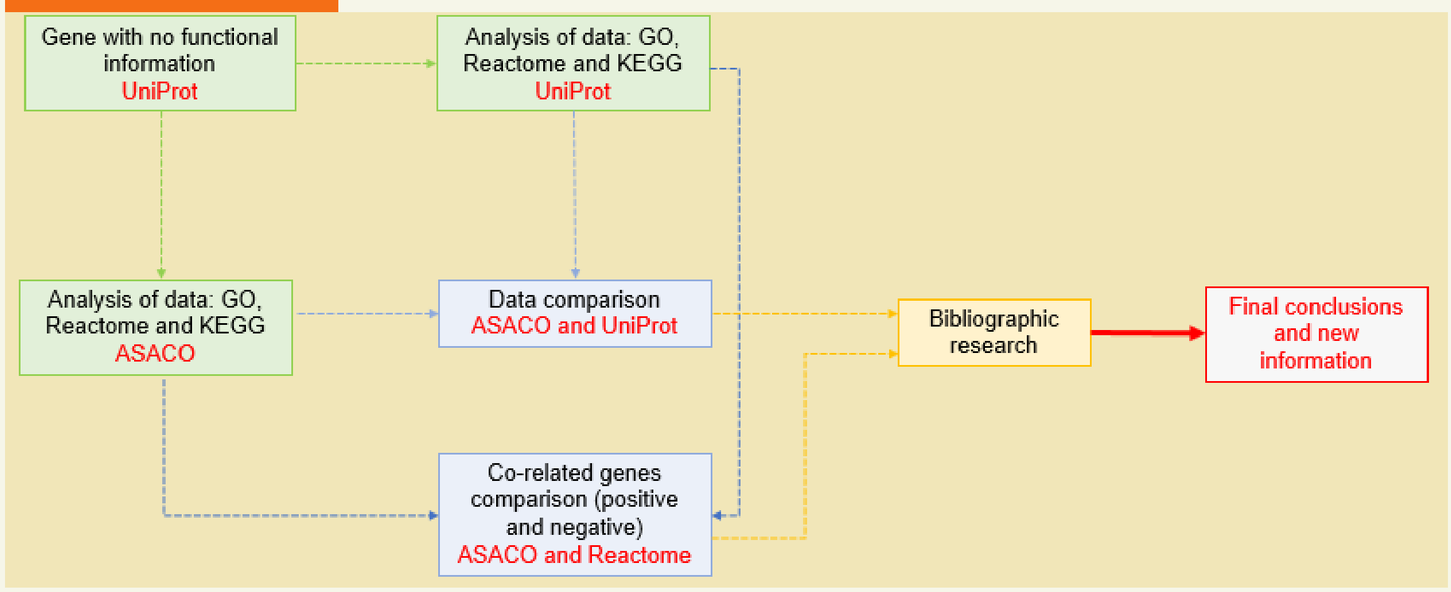
In recent years, Bioinformatics has positioned itself as a highly demanded discipline within the scientific field thanks to the recent advances which have allowed a significant growth in the information available about gene sequences and gene expression. Specifically, gene expression analysis has proven to be a very useful technique for creating knowledge about different complex hereditary diseases by obtaining new data that allows us to understand their actions, similarities with other pathologies, genetic changes and even regulatory drugs.

For the present project the tool Automatic and Serial Analysis of CO-expression (ASACO) developed by the UPOBioinfo Group, in the Bioinformatics Unit of the CABD (Centro Andaluz de Biología del Desarrollo) is used. This tool analyzes the expression of a gene giving putative both positive and negative correlators.

OBJECTIVE

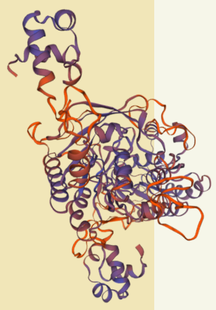
To analyze genes for which functional information is not available and to functionally annotate them with a new tool based on transcriptomic data to obtain relevant data about the diseases in which they are involved.

METHODOLOGY



RAD51C

For the initial experiment, the RAD51C gene has been used. This gene codes for the RAD51C protein, which is one of the 5 paralogs of RAD51, and it is involved in DNA double-strand break repair by homologous recombination. The variations affecting this gene are the cause of two different diseases. On the one hand, one of the complementation groups of Fanconi Anemia, specifically O, which is a hereditary disease that affects all components of the bone marrow and leads to problems such as anemia, leukopenia, thrombopenia, malformations, development of tumors, etc. On the other hand, it is also the cause of familial ovarian and breast cancer, which is characterized by a genetic predisposition to develop ovarian and breast cancer.

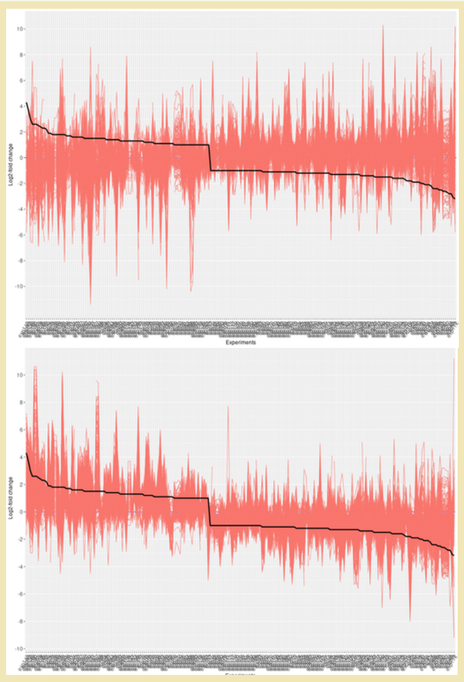


ANALYSIS

Before using ASACO we used Expression Atlas, which is a database that offers information on gene expression patterns between different species and biological conditions from transcriptomics studies.

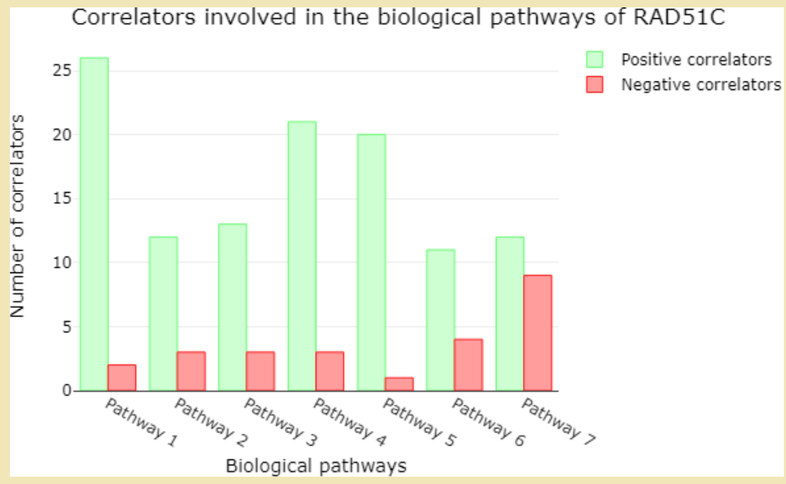
Next, we analyze our gene with ASACO and obtain the negative and positive correlators of our gene, in addition to its gene expression profile. As we can see in the graphs, the positive correlation shows a behavior similar to that of our gene, while the negative correlation shows an inverse behavior.

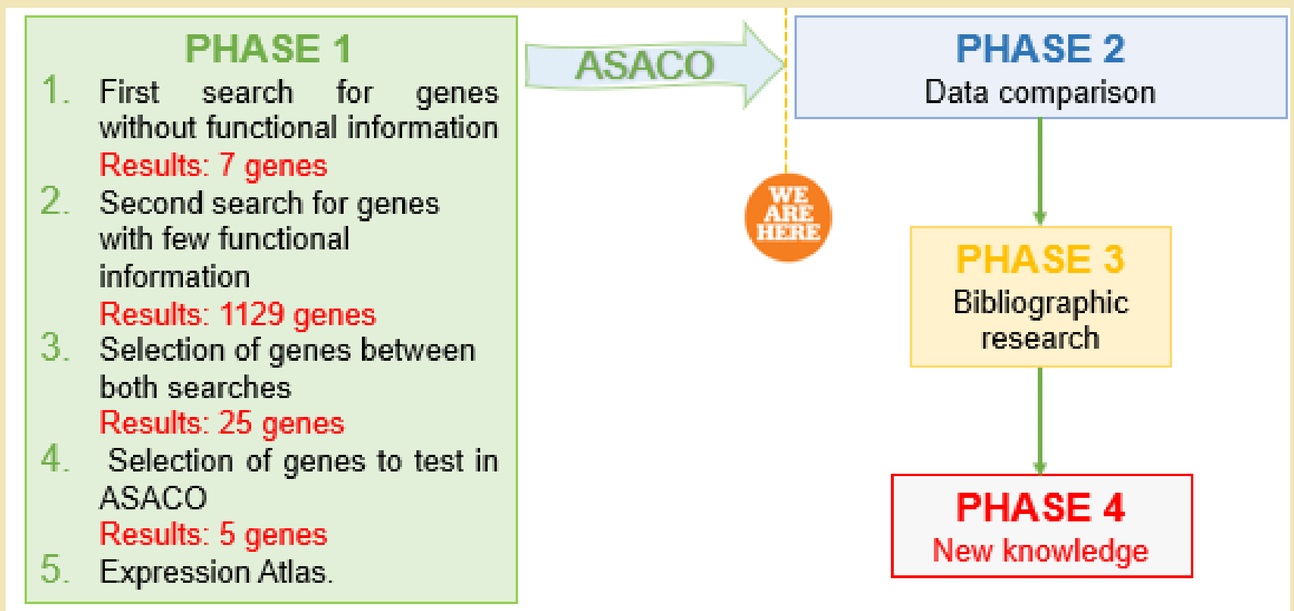
Log_2-fold change	Species	Gene name	Comparison	Experimental variables	Experiment name
4.3		RAD51C	'activated B cell' vs 'memory B cell'	cell type	Transcription profiling by array of human activated B cells, plasmablasts and plasma cells compared to memory B cells
3.7		RAD51C	'cells activated for 72 hr by adding PHA-L' vs 'control'	growth condition	Expression data from unactivated vs. activated PBMCs
-3.2		ENSG00000108384	'estrogen receptor alpha shRNA' vs 'scrambled shRNA'	RNA interference	RNA-seq of the human breast cancer ERÎ±-suppressed MCF-7(MCF-7/SP10+) cells and of their internal control MCF-7 (MCF-7/C) cells
-3.1		RAD51C	'differentiated; hAKPC-P' vs 're-differentiated; hIPod'	cell type, phenotype	Differentiation of human amniotic fluid kidney progenitor cells into podocytes and comparison with human conditionally immortalized podocytes
3		RAD51C	'plasmablast' vs 'memory B cell'	cell type	Transcription profiling by array of human activated B cells, plasmablasts and plasma cells compared to memory B cells



In addition, we obtained functional information of the gene and we compared it with the available information. From the number of matching annotations, we could obtain new information and checked in the available bibliography if the non-matching characteristics created new knowledge.

To get an overview of the results, we graphically compare the number of positively and negatively co-rrelated genes involved in the same biological pathways as RAD51C. From this graph we got new information that helped to make new questions; for example, the negative correlators acting as regulators or the similar number of positive and negative correlators in the last pathway (7). Again, these results needed to be compared with the available bibliography. Based on these ideas, the preliminary results with a well-known gene showed that its positive correlators had related functions with this gene.





Next, this procedure will be repeated starting with a list of 4 or 5 genes for which functional information is not available. To carry it out, we have begun searching for genes that do not have any type of functional annotation using UniProt database; however, since few results were obtained, a new search was made for genes whose cellular location was the only functional information available. The aim of this second experiment is to test whether the protocol followed with a well-known gene is useful to obtain new data while using genes without functional information.

CONCLUSION

- According to the results obtained in the first experiment carried out with RAD51C, the protocol followed and the tool used have proven to be useful in the study of the gene.
- Therefore, we expect to contribute to the creation of knowledge in the study of several hereditary diseases whose genes do not present functional information and to demonstrate the usefulness and value provided by the used tool.