Currently, millions of transcriptomics datasets (the global measurement of gene expression) remain deposited in publicly available databases. These sets, created to answer specific questions about the gene expression level, contain information that has not yet been used in scientific research. Skillful processing of these datasets will allow the discovery of remaining, unused information resources, which can be further used to formulate new research hypotheses, without the need to conduct further costly experiments.

This project proposes the use of ICA analysis (independent component analysis) and the subsequent delimitation of iModulons (independently modulated sets of genes) to extract unused information from the deposited transcriptomic data. This strategy implements mathematical modeling and machine learning techniques, which will increase the efficiency and ‘discovery potential’ of this project. In this way, it will be possible to analyze and synthesize hundreds of data sets, each containing information on thousands of genes, in a comprehensive, systematic, and coordinated way.

It is planned to perform ICA-iModulon analysis on transcriptomic datasets for the yeast species *Yarrowia lipolytica*. This yeast is a valuable scientific research object and an organism used in industrial practice (including, the production of feed, sweetening agents, and recombinant proteins). In addition to the data deposited in publicly available databases, it is planned to develop additional transcriptomic data sets to strengthen the input sets and the ability to select iModulons induced under specific conditions. This is an important element of this project because, in the next step, the results of the ICA-iModulon analysis will be used to design a synthetic promoter (an element regulating gene expression) to be implemented in the controlled production of recombinant proteins. Recombinant proteins are highly valuable biotechnological products used in the production of industrial enzyme preparations, medicines (diagnostic tests or vaccines), cleaning products, and cosmetics. Inducible regulatory elements (the aforementioned synthetic promoters) enabling controlled recombinant protein production are of high value in research and industrial practice.

Therefore, the implementation of this project will lead to the acquisition of new knowledge (characteristics of yeast species used in research and industry), but will also allow for the practical use of this knowledge.