Functional annotation of proteins: Application on MYBL2, a gene involved in the onset and progression of cancer

Author: Katarina Stanković, Mathematical Grammar School, *australiankat@gmail.com*
Mentor: Dr Branislava Gemović, Research Associate of Vinča Institute of Nuclear Sciences, Belgrade

1. Introduction

All known functions of proteins are housed in the Gene Ontology database (GO). They are represented as terms (e.g. GO:0006260 – DNA replication), which are hierarchically organized into three sub-ontologies: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC). Yet, for a great number of proteins, their functions are still unknown, including more than 6000 humane. Therefore, it is important to predict new functions of proteins with computational tools, where functions to be experimentally tested are filtrated, saving time and money. In this study, the functions of protein MYBL2, a gene with deregulated expression in cancer initiation and progression, were analysed. MYBL2 could then serve as a therapeutic target, where knowing all of its biological functions would be of significant benefit. **The aim of this study** is writing a computer program which extracts the results previously predicted by the function annotation algorithm, applying it on all human proteins (approx. 20000 proteins), as well as literature search for confirmation of predicted MYBL2 functions.

2. Methods and Materials

The program for extracting new functions of all human proteins, predicted by the function annotation algorithm previously developed in INS Vinča, is written in C programming language. Next, only predicted MYBL2 functions were isolated and analysed. Using the search engine PubMed, relevant scientific papers describing experimental evidence which confirm predicted MYBL2 functions, were identified. QuickGO was used for function analysis and visual representation (Figure 1).

3. Results and Discussion

An efficient program for extracting new functions of all human proteins was written and applied to protein MYBL2.

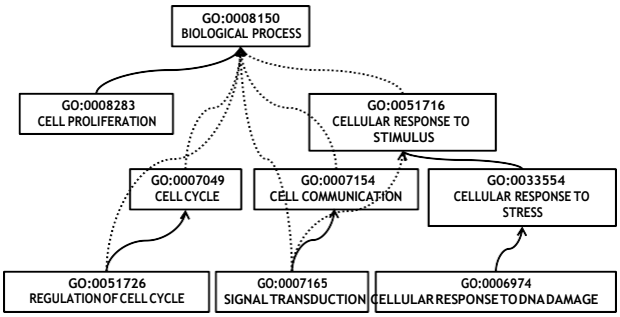

Figure 1 – Schematic diagram of a part of GO tree for MYBL2

The program extracted 16 predicted MYBL2 functions (Table 1). One prediction showed to be redundant; 3/15 functions were related to transcription, a previously known MYBL2 function; for 11/15 functions, references were found in which they are experimentally confirmed; only for 1/15 functions there wasn't evidence in original papers.

Table 1 – Predicted MYBL2 functions, grouped by BP sub-ontology of GO

GO term	Function	Confirm.
GO:0006259	DNA metabolic process	✓
GO:0006260	DNA replication	
GO:0006091	Generation of precursor metabolites and energy	
GO:0006974	Cellular response to DNA damage stimulus	✓
GO:0033554	Cellular response to stress	
GO:0051716	Cellular response to stimulus	
GO:0050896	Response to stimulus	
GO:0006950	Response to stress	
GO:0008283	Cell proliferation	✓
GO:0007165	Signal transduction	✓
GO:0023052	Signalling	
GO:0007154	Cell communication	
GO:0044700	Redundant term	
GO:0006383	Transcription from RNA polymerase III promoter	✓
GO:0045945	Positive regulation of transcription from RNA polymerase III promoter	
GO:0006359	Regulation of transcription from RNA polymerase III promoter	

4. Conclusion

Cancer is one of the most common and deadliest human diseases. Different processes lead to its occurrence and for the development of therapy, it is crucial to identify these processes and involved genes. In this study, computational tools were used with this aim and their role was shown on the example of MYBL2, which was found to be involved in cell proliferation, an important process for the pathogenesis of cancer. The written program for extracting new functions enables us to examine if there is experimental evidence for predicted protein functions. Probing literature, evidence was found for over 93% of predicted MYBL2 functions.

5. References

[1] V. R. Perović, doktorska studija "Razvoj multifunkcionalne bioinformatičke platforme zasnovane na potencijalu electron-jon interakcije bioloških molekula", (2013)

[2] Musa J, Aynaud MM, Mirabeau O, Delattre O, Grünewald TG. MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis. Cell Death Dis. 2017;8(6):e2895.

[3] Jiang Y, Oron TR, Clark WT, Bankapur AR, D'Andrea D, Lepore R, et al. An expanded evaluation of protein function prediction methods shows an improvement in accuracy. Genome Biol. 2016;17(1):184.