## mgr inż. Maciej Migdał

# Modeling of transcription factors influence on gene expression based on data obtained using next-generation sequencing methods

**Rozprawa na stopień doktora nauk medycznych i nauk o zdrowiu**

**w dyscyplinie nauki medyczne**

Promotor: dr hab. Cecilia Winata

Laboratorium Genomiki Rozwoju Danio Pręgowanego,
Międzynarodowy Instytut Biologii Molekularnej i Komórkowej
w Warszawie

Obrona rozprawy doktorskiej przed Radą Dyscypliny Nauk Medycznych
Warszawskiego Uniwersytetu Medycznego

Warszawa 2023 r.

# Abstract in English

The presented collection of articles providing the basis for this doctoral dissertation consists of three publications: Pawlak et al. „Dynamics of cardiomyocyte transcriptome and chromatin landscape demarcates key events of heart development”, *Genome Res.*, 2019; Migdał et al. „Multi-omics analyses of early liver injury reveals cell-type-specific transcriptional and epigenomic shift”, *BMC Genomics*, 2021; and Migdał et al. „xcore: an R package for inference of gene expression regulators”, *BMC Bioinformatics*, 2023. The selected publications aimed to elucidate the principles of gene regulation, emphasizing on the identification of transcription factors (TFs) and DNA regulatory elements (REs) which constitute gene regulatory networks underlying various cellular processes. To achieve this principal aim, we utilized next-generation sequencing (NGS) data collected from organisms at various levels of biological complexity, including *in vivo* data from zebrafish (*Danio rerio)* and *in vitro* human cell lines. These were analyzed using either established bioinformatic algorithms or those which I developed, including NGS data processing, motif enrichment analysis, and machine learning using penalized linear models. The key assumption of the work is the causal relationship between TF, RE, and the transcriptional outcome of their target genes. Based on this assumption, the analytical frameworks exemplified in this collection of articles approaches the problem of transcriptional regulation mechanism using information on gene expression and the activity of RE.

The “Introduction” chapter provides a brief introduction to the topic of gene regulation and the approaches I used in the bioinformatic analysis of the experimental data obtained from experiments employing NGS. Its three subsections give an overview of the included articles, emphasizing on the common gene regulation theme of these studies and the differences in the employed analytical methodologies. The following three chapters contain the copies of the included articles; the associated supplementary materials can be accessed in their on-line forms. Finally, the dissertation is concluded with a “Summary and conclusions” chapter. The co-authors contribution statements for each publication can be found attached following the last chapter.