

Deep Learning of Transcriptional Regulation in *Issatchenkia orientalis*

Payam Dibaeinia¹ (dibaein2@illinois.edu), Shayan Tabe Bordbar¹, Shounak Bhogale², Veronika Dubinkina³, **Sergei Maslov³**, and **Saurabh Sinha^{1,3}**

¹Department of Computer Science, University of Illinois Urbana-Champaign, Urbana;
²Center for Biophysics and Quantitative Biology, University of Illinois Urbana-Champaign, Urbana; ³Department of Bioengineering and Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana

<https://cabbi.bio/research/conversion-theme/>

Project Goals: The project goal was to build a computational modeling tool to better understand transcriptional regulation by analyzing sequence and expression data. This tool can be used to identify key regulators of transcriptomic differences found in yeast strains that are tolerant to low pH, thus providing mechanistic insights that can be used to engineer better strains.

Understanding the “grammar” of transcriptional regulation can help us interpret the different patterns of gene expression in various experimental conditions and predict gene expression in untested conditions. Several computational models of gene expression have been developed in the past that aimed at modeling the expression driven by a promoter as a function of the sequence (*cis* elements) and the concentrations of transcription factors (TF) that bind to the promoter regions (*trans* elements). Here, we developed a ConvolutioNeural Network-based Sequence-to-Expression Prediction Tool (CoNSEPT). CoNSEPT extends previous transcriptional regulation models by capturing a wide variety of non-linear plausible mechanisms that might underlie the gene regulation. A trained CoNSEPT model can be used to identify the TFs’ roles and distance-dependent interactions between TFs, and to guide the future experimental design by predicting optimal *trans* conditions for a particular desired transcriptional profile. Here, we used CoNSEPT to decipher the transcriptional regulation in *Issatchenkia orientalis* (Io129 strain) using the mRNA expression profile of about 5,000 genes in 23 conditions with 22 hypothetical TFs. We showed that CoNSEPT yields a better fit to these data than a baseline linear model. Our future studies will include extending the model by employing a larger set of TFs and exploiting the trained model for suggesting new experimental conditions.

Funding statement: *This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.*