

Ilyes Baali

New York, NY

(646)-236-4417 | ilb4001@med.cornell.edu | LinkedIn: [ilyesbaali](#)

Education

Weill Cornell Medicine and Memorial Sloan Kettering Cancer Center, New York, NY

May 2025

- Ph.D. Candidate In Computational Biology at the Tri-I Computational Biology And Medicine Program

Antalya Bilim University, Antalya, Turkey

August. 2020

- M.S.C. IN COMPUTER ENGINEERING (3 publication, 1 first author)
 - Thesis title: DriveWays: A Method for Identification of Overlapping Cancer Driver Modules
- B.SC. IN COMPUTER ENGINEERING (1 publication, first author)
- B.SC. IN ELECTRICAL AND ELECTRONICS ENGINEERING (3 publications)

Jun. 2017

Jun. 2017

Experience

Morris Lab (Prof. Quaid Morris) MSKCC and WCM, New York, NY

Jul. 2021 – Present

DOCTORAL RESEARCHER

- Develop a machine learning-based recalibration model that can accurately predict the cellular binding sites of RNA-Binding Proteins (RBPs) given only their intrinsic binding preferences.
- Reanalyze all available CLIP data to identify biases and experimental artifacts in the dataset. We found evidence that as much as a third of the eCLIP data associated with ENCODE RBPs does not measure direct binding.
- Investigate various machine learning techniques such as AlphFold, for representing proteins to predict the binding preferences of RNA-binding Proteins (RBPs) from their amino-acid sequences.
- Analyze large genomic data to understand the role of MSI2-RBP in AML, this include running various bioinformatics pipelines to detect binding sites from HyperTRIBE assay and measure translation efficiency from RiboSTAMP assay.
- Collaborated with wet-lab teams to address their bioinformatics needs, prepared presentations to communicate the findings and results.
- Organize regular lab-meeting for a whole year, including scheduling presentations, developing automated scripts to collect and distribute presentation feedback.

Computational Biology Lab (Assoc. Prof. Hilal Kazan & Prof. Cesim Erten), Antalya, Turkey

Oct. 2018 - Jul.2020

RESEARCH ASSISTANT

- Developed a novel algorithm for identifying overlapping cancer driver modules using multi-omics data from The Cancer Genome Atlas (TCGA).
- Performed the classification and survival analysis for a novel approach named MEXCOWalk for identifying cancer driver modules.
- Co-supervised a final year undergraduate student's independent study project, providing support and guidance on various aspects including experimental design, data analysis, and report writing.

Morris Lab at University of Toronto (Prof. Quaid Morris), Toronto, Canada

Jun. 2019 – Sep.2019

VISITING RESEARCHER

- Designed and implemented a multitask deep learning model for predicting in vivo binding sites for RNA-binding proteins (RBPs) using in vitro binding preferences learned from paired data. This model utilizes advanced machine learning techniques to accurately predict binding site locations within the genome, which can provide valuable insights into the regulatory mechanisms of RBPs and their role in various biological processes.

Machine Learning Lab (Assoc. Prof. Hilal Kazan), Antalya, Turkey

Oct. 2016 - Aug. 2017

SENIOR PROJECT

- Participated in the Neuroblastoma data Integration challenge organized by the CAMDA consortium, with the goal of improving the prediction of clinical outcome, survival time, or disease mechanisms through the integration of multiple data types, as compared to the original expression only study conducted by the FDA SEQC.
- Integrated various types of data such as gene expression and aCGH data to improve the prediction of survival time.

Digital Media and Data Reconstruction Lab (Prof. Lizhuang Ma), Shanghai, China

Jun. 2016 - Sep. 2016

SUMMER INTERNSHIP

- Learned the principles of machine and deep learning methods, with a particular focus on convolutional neural networks.
- Built a face recognition system using the GoogleNet neural network, which achieved an accuracy of 94% on the LFW (Labelled Faces in the Wild) dataset.

Skills

Computational Biology

Perform various differential analyses, Analyze CLIP and RNA-editing based assays, Apply machine learning techniques to functional genomic data.

Deep Learning Frameworks

Pytorch, Keras, Tensorflow

Programming Languages

Python(+10 years), R, C/C++, JAVA, Matlab

Spoken Languages

French (good), Arabic (bilingual), Turkish (beginner)