

ILYES BAALI

📍 New York, NY | 📞 (646) 236-4417 | ✉️ ilb4001@med.cornell.edu | 🌐 ilyesbaali | 📧 ilyes495 | U.S. Permanent Resident

SUMMARY

Computational biologist (Ph.D., Weill Cornell Medicine / Memorial Sloan Kettering Cancer Center) developing machine-learning methods for RNA biology and translational oncology.

- Analyzes large-scale multi-omic data: bulk and single-cell RNA-seq, CLIP-seq, Ribo-seq, and spatial transcriptomics.
- Builds deep learning models on biological sequences, including protein language models and genomic sequence-to-function models, with 10+ years of Python experience.
- Publishes and presents across fields: *Nature*, *Bioinformatics*, *Scientific Reports*, and *Blood*; oral presentations at ISMB/ECCB 2025 and ISMB 2021.
- Mentors undergraduate and graduate researchers, including 4 research interns supervised at MSKCC.

EDUCATION

Ph.D., Computational Biology

Jul. 2026

Tri-Institutional PhD Program in Computational Biology & Medicine, Weill Cornell Medicine / Memorial Sloan Kettering Cancer Center
New York, NY

Dissertation: *Cellular context in protein–RNA interactions*: computational methods to quantify RBP-mediated translational control (applied to MSI2 in AML) and to diagnose when *in vivo* (eCLIP) binding reflects true sequence specificity. Advisor: Prof. Quaid Morris (MSKCC).

M.Sc., Computer Engineering

Aug. 2020

Antalya Bilim University

Antalya, Turkey

Thesis: *DriveWays*, overlapping cancer driver modules from multi-omics data. GPA 4.0/4.0. Three publications (one first-author).

B.Sc., Computer Engineering

Jun. 2017

Antalya Bilim University

Antalya, Turkey

Graduated with high honors, ranked 3rd in class; best senior project award. One first-author publication.

B.Sc., Electrical & Electronics Engineering

Jun. 2017

Antalya Bilim University

Antalya, Turkey

Graduated with high honors, ranked 2nd in class. Three publications.

RESEARCH EXPERIENCE

Doctoral Researcher

Jul. 2021 – Present

Morris Lab, Memorial Sloan Kettering Cancer Center & Weill Cornell Medicine

New York, NY

- Built an RBP-agnostic CLIP model to test whether eCLIP binding reflects an RBP's intrinsic sequence specificity.
- Reanalyzed 240+ ENCODE eCLIP experiments with this model and found that only 32% reflect intrinsic sequence specificity.
- Showed that non-specific *in vivo* signal concentrates in purine/G-rich, exon-junction-localized RBPs and reflects condensate (stress-granule, nuclear-speckle) localization rather than sequence.
- Developed an intrinsic-specificity score that flags non-sequence-specific eCLIP binding without re-

quiring *in vitro* data.

- Benchmarked protein language model (ESM) and AlphaFold-derived representations for predicting RBP binding specificity, outperforming sequence- and structure-based baselines.
- Built and deployed Nextflow pipelines on SLURM HPC clusters for RNA-seq, iCLIP/eCLIP, HyperTRIBE, and Ribo-seq/RiboSTAMP data, processing multi-terabyte genomic datasets.
- Designed a 10x Visium spatial transcriptomics pipeline to characterize tumor microenvironment composition, enabling cell-type deconvolution.
- Demonstrated that the RBP MSI2 promotes translation of a leukemic stem-cell program in AML through a cooperative RBP module (co-first-author manuscript in review).
- Developed a method to quantify translation efficiency from RiboSTAMP editing data.
- Supervised 4 undergraduate interns and graduate rotation students in CLIP-seq analysis, ML model fine-tuning, and HPC usage (see Mentoring).
- Communicated results to mixed computational, wet-lab, and clinical audiences; organized a year-long lab-meeting series with automated feedback collection.

Visiting Researcher

Morris Lab, University of Toronto

Jun. 2019 – Sep. 2019

Toronto, Canada

- Designed and implemented a multitask deep learning model (PyTorch) to predict *in vivo* RBP binding sites from *in vitro* binding preferences using paired RNAcompete/eCLIP data; work later presented at Genome Informatics 2021 and ISMB 2022.

Research Assistant

Computational Biology Lab (Profs. H. Kazan & C. Erten), Antalya Bilim University

Oct. 2018 – Jul. 2020

Antalya, Turkey

- Developed DriveWays, a graph-based algorithm for identifying overlapping cancer driver modules from TCGA multi-omics data (first-author publication in *Scientific Reports*).
- Performed classification and survival analyses for MEXCOWalk, a random-walk method for cancer module discovery (*Bioinformatics*).
- Co-supervised a final-year undergraduate independent study project, guiding experimental design, data analysis, and report writing.

Senior Project Student

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

Oct. 2016 – Aug. 2017

Antalya, Turkey

- Integrated gene expression and aCGH data to improve prediction of survival and clinical outcome in neuroblastoma for the CAMDA 2017 challenge; first-author publication in *Biology Direct* and oral presentation at CAMDA 2017.

Summer Research Intern

Digital Media & Data Reconstruction Lab (Prof. L. Ma), Shanghai Jiao Tong University

Jun. 2016 – Sep. 2016

Shanghai, China

- Built a face recognition system based on GoogLeNet, achieving 94% accuracy on the LFW benchmark.

TECHNICAL SKILLS

Multi-omics

RNA-seq (bulk & single-cell), spatial transcriptomics (10x Visium), CLIP-seq (iCLIP/eCLIP), Ribo-seq/RiboSTAMP, HyperTRIBE, differential expression, survival analysis, causal inference

Machine learning

PyTorch, Keras, TensorFlow; protein language models (ESM), AlphaFold integration; Bayesian inference (NumPyro); LLM tooling and prompt engineering for scientific workflows

Pipelines & HPC	Nextflow, Snakemake; SLURM cluster computing; Docker; git; reproducible research practices
Programming	Python (expert, 10+ years), R (proficient), C/C++, Java, MATLAB, SQL (basic)
Data & visualization	pandas, NumPy, scikit-learn, matplotlib/seaborn, ggplot2; publication-quality figures
Languages	Arabic (bilingual proficiency), French (professional), Turkish (conversational)

PUBLICATIONS

Peer-reviewed journal articles

- [1] Jiang Q., Raghavan M.S., Rodriguez A.M., Lallo M., ..., **Baali I.**, Kharas M.G., de Stanchina E., Urganci N., Shia J., Pe'er D., Sanchez-Vega F., Koche R., Morris Q., Chan J.M., Ganesh K. "ZFP36L2 orchestrates stress-adaptive plasticity in regeneration and cancer." *Nature* (2026).
- [2] **Baali I.**, Erten C., Kazan H. "DriveWays: a method for identifying possibly overlapping driver pathways in cancer." *Scientific Reports* 10:21971 (2020).
- [3] Ahmed R., **Baali I.**, Erten C., Hoxha E., Kazan H. "MEXCOWalk: mutual exclusion and coverage based random walk to identify cancer modules." *Bioinformatics* 36(3):872–879 (2020).
- [4] Al-Turjman F., **Baali I.** "Machine learning for wearable IoT-based applications: a survey." *Transactions on Emerging Telecommunications Technologies* (2019).
- [5] **Baali I.**, Acar A.E., Aderinwale T., HafezQorani S., Kazan H. "Predicting clinical outcomes in neuroblastoma with genomic data integration." *Biology Direct* 13:20 (2018).
- [6] Demirbas U., **Baali I.** "Power and efficiency scaling of diode pumped Cr:LiSAF lasers: 770–1110 nm tuning range and frequency doubling to 387–463 nm." *Optics Letters* 40:4615–4618 (2015).
- [7] Demirbas U., **Baali I.**, Acar A.E., Leitenstorfer A. "Diode-pumped continuous-wave and femtosecond Cr:LiCAF lasers with high average power in the near infrared, visible and near ultraviolet." *Optics Express* 23:8901–8909 (2015).
- [8] Beyatli E., **Baali I.**, Sennaroglu A., Sumpf B., Erbert G., Leitenstorfer A., Demirbas U. "Tapered diode-pumped continuous-wave alexandrite laser." *Journal of the Optical Society of America B* 30:3184–3192 (2013).

Workshop papers and published abstracts

- [9] Shi R., Dalal T., Fradkin P., Koyyalagunta D., ..., **Baali I.**, Wang B., Morris Q. "mRNABench: a curated benchmark for mature mRNA property and function prediction." Gen² Workshop on Generative AI in Genomics, *ICLR* (2026). openreview.net/forum?id=qBAtBfTvah.
- [10] Xie X.*, Herrejon Chavez F.*, **Baali I.***, ..., Morris Q.D., Vu L., Nguyen D.T.T., Kharas M.G. "Uncovering the HOXA9 translational regulatory complex that promotes AML leukemic stem cells." *Blood* 144 (Suppl. 1):4106 (2024). ASH Annual Meeting abstract.

Manuscripts in review or in preparation

- [11] Xie X.*, Herrejon Chavez F.*, **Baali I.***, ..., Morris Q.D., Vu L.P., Nguyen D.T.T., Kharas M.G. "A dynamic RNA-binding protein module coordinates translation of leukemic stemness programs." In review (2026).
- [12] **Baali I.**, Morris Q. "An intrinsic-specificity score and RBP-agnostic CLIP model to distinguish sequence-specific *in vivo* (eCLIP) binding from shared, condensate-driven signal." In preparation.

* Co-first authors.

TALKS & POSTER PRESENTATIONS

2026	Poster. "Bridging the gap: recalibrating <i>in vitro</i> models for accurate <i>in vivo</i> RBP binding predictions." <i>Cold Spring Harbor Laboratory 90th Symposium: AI in Biology</i> , Cold Spring Harbor, NY.
2025	Oral presentation. "Bridging the gap: recalibrating <i>in vitro</i> models for accurate <i>in vivo</i> RBP binding predictions." <i>ISMB/ECCB 2025</i> , Liverpool, UK.
2021 – 2022	Posters. "Predicting <i>in vivo</i> binding preferences of RNA-binding proteins using transfer learning." <i>ISMB 2022</i> , Madison, WI, and <i>Genome Informatics 2021</i> , Cold Spring Harbor Laboratory (virtual).
2021	Oral presentation. "DriveWays: a method for identifying possibly overlapping driver pathways in cancer." <i>ISMB 2021</i> (virtual).
2017	Oral presentation. "Predicting clinical outcomes in neuroblastoma with genomic data integration." <i>16th Critical Assessment of Massive Data Analysis (CAMDA)</i> , Prague, Czech Republic.
2013 – 2015	Co-authored conference contributions (laser physics, with U. Demirbas et al.): CLEO Europe 2015 (Munich; oral), CLEO 2014 (San Jose, CA; oral), Advanced Solid State Lasers 2013 (Paris; poster).

HONORS & AWARDS

- PhD Fellowship, Tri-Institutional PhD Program in Computational Biology & Medicine (2020–2026).
- Best Senior Project, Department of Computer Engineering, Antalya Bilim University.
- Full merit-based scholarship and High Honors in both engineering departments, Antalya Bilim University.

MENTORING & SUPERVISION

Summer 2026	Kate Zhang , research intern, Morris Lab: mapping shared CLIP binding hotspots, transcript regions bound by many RBPs independently of their sequence specificity.
Summer 2024	Defne Ceyhan and Jessica Lin , research interns, Morris Lab: predicting RBP cellular localization from CLIP data.
Summer 2023	Ravi Tapanmitra , research intern, Morris Lab: protein language models for predicting TCR–epitope recognition.
2023	Mentor , Tri-Institutional Mentorship Initiative: guided a prospective student through the graduate-school application process.
2019 – 2020	Co-supervised a final-year undergraduate independent study project, Computational Biology Lab, Antalya Bilim University.

PEER REVIEW, SERVICE & LEADERSHIP

2025	Reviewer, <i>Scientific Reports</i> .
2023 – 2025	Reviewer, <i>Machine Learning in Computational Biology (MLCB)</i> conference.
2024	Reviewer, <i>Great Lakes Bioinformatics Conference (GLBIO)</i> .
2023	Reviewer, <i>ICML 2023 Workshop on Computational Biology (WCB)</i> .
2024, 2025	Application reviewer, <i>Computational Biology Summer Program (CBSP)</i> , MSKCC: a 10-week mentored research internship for about 15 undergraduates in computer science and applied mathematics (MSKCC, Weill Cornell, Rockefeller).

Ongoing	Core organizer, <i>Algerian American Scientists Association (AASA)</i> : co-led numerous initiatives aimed at strengthening the Algerian scientific community in the United States.
2022 – 2023	Organizer, Morris Lab meeting series: year-long schedule with automated presentation-feedback collection.

HACKATHONS

Mar. 2026	Y Combinator Bio × AI Hackathon , San Francisco, CA. Built an automated protein-engineering pipeline that designs an SSTR2 GPCR biosensor, co-evolving cpGFP insertion sites and linker sequences with a genetic algorithm, scoring candidates by ESMFold and AlphaFold2-multimer structure prediction, and using a Claude agent to shortlist designs by folding quality and ligand coupling.
Oct. 2025	NYC AI Agents Hackathon , New York, NY. Built a real-time volunteer-matching agent that streams Bluesky posts, detects volunteering intent with an LLM, and matches posts to NYC opportunities using local vector embeddings, with a web interface and Datadog monitoring.
Jun. 2019	Hackbyte: Loblaw Digital × Microsoft Machine Vision Hackathon , Toronto, Canada. Built a computer-vision pipeline to monitor produce quality and freshness in grocery stores.