

Katarína Grešová

contact@katarinagresova.com · katarinagresova.com · Berlin, Germany · github.com/katarinagresova

Machine learning for RNA regulation — sequence models, the interpretability that makes them arguable, and the benchmarks that decide whether any of it is real. Seven years of production software engineering before the doctorate. Available for teaching, consulting and advisory work.

RESEARCH EXPERIENCE

2025 – present	Postdoctoral researcher Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center, Berlin Interpretable machine learning for RNA regulation: what in an mRNA's own sequence sets how much protein a cell makes from it, and what that explains about health and disease.
2023 – 2025	Researcher Intelligent Systems for Complex Data group, Masaryk University, Brno Protein-structure similarity search at proteome scale — adapting pretrained networks to produce structure embeddings, and the pipelines to run them on HPC.
2024 – 2025	Pre-doctoral fellow Laboratory of Genetics and Genomics, National Institute on Aging, NIH, Baltimore Senescent-cell classification from single-cell long-read sequencing in the ageing mouse brain, and shared senescence markers across cell types from transcriptomic and proteomic data.
2023 – 2024	Visiting researcher Alexiou Lab, University of Malta, Msida Led the miRBench project. Taught deep learning and interpretability, and supervised four junior researchers.
2020 – 2023	Research specialist Alexiou group, CEITEC, Masaryk University, Brno Genomic Benchmarks, attribution-based interpretation of sequence models, and deep learning for RNA. Ran the group's workshops and hackathons.

EDUCATION

2020 – 2025	Ph.D. in Genomics and Proteomics Masaryk University, Brno Dissertation: Modeling small RNA binding rules using machine learning, supervised by Panagiotis Alexiou. Awarded summa cum laude.
2018 – 2020	M.Sc. in Bioinformatics and Biocomputing Brno University of Technology Thesis: Bacteria classification into taxonomic categories based on properties of 16S rRNA. Graduated with honours.
2015 – 2018	B.Sc. in Information Technology Brno University of Technology Thesis: Searching semantically annotated texts. Graduated with honours.

RECOGNITION

Ph.D. awarded summa cum laude, Masaryk University, 2025 · Pre-doctoral fellowship, National Institute on Aging, NIH, 2024 – 2025 · Oxford Machine Learning Summer School, health and finance tracks, 2022 · Top-five poster, ENBIK 2022 · M.Sc. and B.Sc. both graduated with honours, Brno University of Technology

SELECTED PUBLICATIONS

16 in total, 5 as first or co-first author

2026	SenCat: cataloging human cell senescence through multi-omic profiling of multiple senescent primary cell types <i>Molecular Cell</i> — Shared second author
2025	miRBench: novel benchmark datasets for microRNA binding site prediction that mitigate against prevalent microRNA frequency class bias <i>Bioinformatics</i> — Co-first author
2024	RBP-Tar: a searchable database for experimental RBP binding sites <i>F1000Research</i> — first author
2023	Genomic benchmarks: a collection of datasets for genomic sequence classification <i>BMC Genomic Data</i> — first author
2023	Using attribution sequence alignment to interpret deep learning models for miRNA binding site prediction <i>Biology</i> — first author

2025 - present	Genomic Benchmarks QC Scores a genomic dataset for the shortcuts a classifier could exploit — bias, duplicates, train/test leakage — before you train on it. genomic-benchmarks.github.io/genomic-benchmarks-qc/
2024 - 2025	miRBench Benchmark datasets and a Python package for miRNA target-site prediction, built to remove the frequency bias that inflates published scores. github.com/katarinagresova/miRBench
2021 - 2023	Genomic Benchmarks Eight curated datasets for genomic sequence classification, each with an honest split and a baseline worth beating. github.com/ML-Bioinfo-CEITEC/genomic_benchmarks
2022 - 2023	Attribution sequence alignment Aligns attribution profiles across many sequences, making the patterns a network has learned visible as motifs a biologist can argue with. doi.org/10.3390/biology12030369
2023 - 2026	AlphaFind and AlphaFind2 Structure-similarity search across all 200 million-plus predicted structures in AlphaFold DB. Public web service. alphafind.ics.muni.cz/
2023 - 2024	RBP-Tar Searchable database of experimentally determined RNA-binding protein sites, pulled out of published CLIP experiments. rbp-tar.biodata.ceitec.cz

SOFTWARE ENGINEERING

2023	Machine learning engineer Melown Technologies, Brno
2021 - 2022	Machine learning specialist Proficio / Databy, Brno
2020 - 2021	Software developer Oracle NetSuite, Brno
2018 - 2020	Software developer Tieto, Brno
2016 - 2018	Research scholar Knowledge Technology group, Brno University of Technology

TEACHING

2024	Deep Learning for Genomics Three-day course, University of Malta
2023	Explainability for Deep Neural Networks MALTAomics Summer School, University of Malta
2021	Inside the black box: explainable deep learning for biological data Tutorial, ECCB 2021
2020 - 2021	Introduction to Bioinformatics Teaching assistant, Faculty of Medicine, Masaryk University
ongoing	The Missing Skills open course notes
2021 - 2022	Biology Crash Course open course notes
2023 - present	Supervision 3 master's students and bachelor interns; hackathons and workshops organised
2021 - 2026	Talks and posters ECCB, the Microsymposium on RNA Biology, ENBIK (top-five poster) and departmental seminars — 9 in total

SKILLS

Python — PyTorch, TensorFlow, scikit-learn, Hugging Face, pandas, NumPy, Jupyter

Also Java, C++, R, SQL and rather a lot of Bash

Snakemake pipelines, Git, CI/CD, HPC

Deep learning, explainable AI, model benchmarking, NLP, computer vision

Genomics: sequence models, RNA-seq (bulk and single-cell), long-read sequencing, CLIP and chimeric-read protocols, protein structure

Full publication list, code and course material at katarinagresova.com