Kathryn (Katy) Monopoli, PhD

Machine Learning Scientist *| Drug Discovery* (508)-367-9077 | krmonopoli@gmail.com | LinkedIn | github

Machine learning scientist with biotech and pharma industry experience developing predictive models that advance therapeutic programs. Background in experimental biology provides practical intuition for data quality, assay design, and cross functional decision making, enabling stronger collaboration with R&D teams. Known for clear scientific communication across teams and for delivering computational solutions that create measurable impact in drug discovery.

KEY EXPERTISE

1. ML and Predictive Modeling for Therapeutic Design

- Build and deploy models that improve drug design decisions for RNA/oligonucleotide and protein/biologics-based modalities
- Develop sequence, structure, and assay-level representations that enable high accuracy in small or noisy data regimes
- Integrate ML outputs with experimental pipelines to guide prioritization, mechanism insight, and translational decisions

Key Skills: Python, PyTorch, TensorFlow, Scikit learn, XGBoost, deep and embedding models, GNNs, representation learning, feature engineering, small-data methods, model evaluation and interpretability, high throughput screening analysis, dose response modeling, ETL and QC pipelines, SQL, data visualization

2. Software Engineering and Pipeline Development for Drug Design

- Architect computational pipelines and production tools for large scale sequence design and ML-driven decision support
- Lead cross functional engineering efforts to define requirements, coordinate milestones, and deliver reproducible workflows
- Build scalable data systems and APIs that support model training, experimental annotation, and downstream integration

Key Skills: Python, Java, C++, Docker, AWS and HPC environments, Docker, Git, API and workflow design, SQL and database tools, containerized pipelines, reproducible research practices, project and team coordination, Jira, Confluence

3. Cross-Domain Scientific Communication and Collaboration

- Present complex computational results to diverse scientific, clinical, and executive audiences with clarity and precision
- Bridge ML, chemistry, and biology contexts to support decision making across therapeutic programs and stakeholders
- Lead collaborations spanning biotech startups, large pharma, and academic groups while mentoring junior scientists

Key Skills: Scientific communication, public speaking, data visualization, technical writing, cross functional collaboration (with chemists, biologists, clinicians, shareholders), project management, mentoring, model explainability for scientific partners

PROFESSIONAL EXPERIENCE

UMass Chan Medical School — RNA Therapeutics Institute | Worcester, MA 2019–2025

Graduate Researcher | Computational Biology and ML

AI Methods for Therapeutic Design and Biomedical Data Analysis

- Developed ML algorithms enabling design of 2,000+ fully modified siRNAs, resulting in 34 advancing to drug discovery
- Built predictive models (ensemble, regularized linear models, neural networks, semi-supervised learning) using rigorous cross validation and holdout testing along with integration with experimental data (MLOps)
- Applied representation learning and embedding based feature strategies that outperformed traditional engineered features.
- Designed data partitioning and training schema for small, imbalanced biological datasets to improve early generalization
- Validated models with experimental data to produce actionable insights

ML Engineering, Pipelines, and Tooling

- Built end-to-end ML pipelines for high throughput NGS data performing QC, feature extraction, training, tuning, and reporting.
- Deployed models within a production web platform integrating scoring, design tools, and visualization for collaborators.
- Created reproducible workflows on Linux and HPC systems using containers, schedulers, and Git based versioning.
- Designed database schemas and lightweight APIs to manage experimental data, annotations, and model outputs.

Collaboration, Communication, and Impact

- Collaborated with chemists, biologists, and clinicians to define features/labels and refine models based on results.
- Co-authored publications and patents on ML driven therapeutic design, including work in preparation for Nature Biotech.
- Presented methods and findings to interdisciplinary global conference audiences

siRNA.ai | AI-powered Therapeutic siRNA Design Portal | Worcester/Boston, MA 2023-2025

Software Development Architect | Predictive siRNA Drug Design

- Led the development of a web based ML powered therapeutic design platform integrating algorithms, data, and UX tools
- Coordinated with software engineers, data scientists, and experimentalists to define requirements, set milestones, and delivered on time and under budget
- Mentored junior teams on ML best practices, reproducible research, and communication with non-technical stakeholders.
- Resulted in over 5000 siRNA design requests globally.
- Led to partnerships with early biotech startups and big pharma companies including Novartis

Advirna - Oligonucleotide Therapeutics | Cambridge, MA 2016-2018
Associate Scientist | Bioinformatics & Database Engineer

- Improved predictive algorithms for oligo design, increasing design throughput while improving experimental validation accuracy.
- Trained interpretable ML models optimized for small, noisy datasets using regularized approaches, enabling stakeholders to understand key features driving predictions.
- Designed and deployed a cloud backed MySQL database for R&D data

Institute for Protein Design (David Baker Lab) | Seattle, WA 2016-2017
Researcher | Computational Biophysics

- Contributed C++ software components to the Rosetta protein folding and design suite
- Ran large scale protein modeling and design calculations on cluster and cloud infrastructure

Biogen | Cambridge, MA 2013-2013
Intern | Antibody Discovery

- Evaluated antibodies and Fabs for binding, stability, and functionality using biophysical assays including BLI and ELISA
- Built Python tools for antibody sequence analysis to accelerate epitope mapping and CDR analysis

EDUCATION

UMass Chan Medical School and Worcester Polytechnic Institute *(dual degree)* | Worcester, MA 2019-2025
PhD, Computational Biosciences and Bioengineering
Advisors: Anastasia Khvorova (RNA therapeutics), Dmitry Korkin (Bioinformatics and CS)
Dissertation: AI and ML methods for predicting therapeutic siRNA efficacy and enabling data driven oligonucleotide design

University of Massachusetts Amherst | Amherst, MA 2015
MS, Molecular and Cellular Biology, Phi Beta Kappa, GPA 4.0
BS, Biochemistry and Molecular Biology**,** Honors College, *summa cum laude*, GPA 4.0 2014

TECHNICAL SKILLS

Languages: Python, R, Java, C++, SQL, Bash

Machine Learning and Modeling: TensorFlow, PyTorch, Scikit learn, XGBoost, LightGBM, Keras, representation learning, embedding models, supervised and semi supervised methods, ensemble approaches, feature importance analysis, cross validation strategies, model diagnostics for experimental readouts

Data Science and Experimental Analytics: Pandas, NumPy, SciPy, Statsmodels, Jupyter, QC and normalization pipelines for high throughput assays, statistical testing, dose response and curve modeling, dimensionality reduction, visualization of complex biological datasets, RDKit, Rosetta, AlphaFold, MLOps (MLFlow, Databricks), Django, Benchling integration

Bioinformatics and Sequencing Workflows: Biopython, pysam, HTSeq, DESeq2, Scanpy, processing and interpretation of NGS and transcriptomic data, alignment and quality assessment, sequence level feature generation, FASTQ/BAM handling and annotation

Infrastructure and R&D Tooling: Linux/Unix systems, Docker, Git and GitHub, AWS (EC2, S3), HPC environments with Slurm, relational databases (MySQL, SQLite), workflow automation, API development for model or assay data integration, data ingestion

SELECTED PUBLICATIONS - *Full Publication List on Google Scholar*

- **Monopoli KR**, Sostek B, Gross K, Alterman J, Korkin D, Khvorova A. (2025) Applying advanced AI methods to define features of functional therapeutic siRNAs. *Submitted to Nature Biotechnology.*
- Rivera Flores IV, **Monopoli KR**, Jackson S, Echeverria D, O'Reilly D, Brown RH, Khvorova A. (2024) Near Sequence Homology Does Not Guarantee siRNA Cross-Species Efficacy. *Nucleic Acid Therapeutics (NAT)*.
- Davis SM, Hildebrand S, **Monopoli KR**, et al. (2025) Systematic analysis of siRNA and mRNA features impacting fully chemically modified siRNA efficacy. *Nucleic Acids Research* (*NAR)*.
- **Monopoli KR**, Korkin D, Khvorova A. (2023) Asymmetric trichotomous data partitioning enables predictive ML models using limited siRNA efficacy datasets. *Molecular Therapy Nucleic Acids (MTNA)*.
- Shmushkovich T*, **Monopoli KR*** et al. (2018) Functional features defining the efficacy of cholesterol-conjugated self-deliverable chemically modified siRNAs. *NAR*. ***co-first authors***

PATENTS

1) Oligonucleotides for MAPT modulation (US 17/204,480)*****
2) Oligonucleotides for SNCA modulation (US 17/204,483)*****
3) Oligonucleotides for MSH3 modulation (US 63/012,603)*****
4) Oligonucleotides for SARS-CoV-2 modulation (US 17/333,839)

****= Licensed to Atalanta Therapeutics***

SELECTED TALKS

- ***(Invited Speaker)*** American Society of Gene and Cell Therapy (ASGCT) Meeting; 2026 May; Boston, MA; Designing Potent Therapeutic Fully Modified Therapeutic siRNAs using Advanced AI/ML Techniques. *(Oral Presentation, confirmed)*
- Oligonucleotide Therapeutics Society (OTS) Meeting; 2025 Oct; Budapest, Hungary; Accurate AI-Driven Prediction of Potent Therapeutic Fully Chemically Modified siRNAs. *(Oral Presentation, confirmed)*
- Cold Spring Harbor Laboratory Nucleic Acid Therapies Meeting; 2025 Mar; Cold Spring Harbor, NY; Advanced machine learning approach for accurate prediction of functional fully modified siRNAs. *(Oral Presentation)*
- Conference on Intelligent Systems for Molecular Biology (ISMB); 2022 Jul; Madison, WI; Trichotomous classification on small, limited datasets enables predictive model development for therapeutic small interfering RNA.*(Oral Presentation)*
- ***(Invited Speaker)*** Methods to apply and evaluate machine learning models on limited biological datasets through the lens of siRNA design. Talk presented at Oligonucleotide Therapeutics Society Webinar; 2021 Oct 29.