Nikolay S. Markov

nikolai.markov@icloud.com | mxposed.github.io | Google Scholar profile ⟳

updated September 2025

Computational biology researcher and leader. After pivoting my career from software engineering leadership to reunite with my passion for biological research in 2017, I came to the Pulmonary division at Northwestern. Here, I set up computational biology operations from sample tracking to interpretable multimodal analyses to support translational research in lung diseases, building the foundation for the division NIH funding of $5M+ and impactful publications. My interest is to use vast amount of published and private multimodal data via novel machine learning models to improve our understanding of common diseases and factors that determine patient individual outcomes.

Expertise

- **Research leadership**
  Cross-functional collaboration, project management, team management, mentorship, scientific communication

- **Machine learning**
  Deep learning frameworks (torch, jax), large language models, multimodal fusion, embedding representations, bayesian modeling

- **Healthcare and biological data**
  Electronic health records (EHR), clinical notes analysis, single-cell RNA-seq, TCR-seq, spatial transcriptomics, imaging data, statistics

- **Software & tools**
  Python, R, SQL, Linux, computational clusters, interactive visualizations, pipelines

Education

Ph.D. in Computational Biology, Northwestern University, Chicago, USA — 2022–2025 (expected in December)
Driskill Graduate Program in Life Sciences

M.S. in Bioinformatics, Newcastle University, Newcastle upon Tyne, UK — 2017–2018
With distinction

Undergraduate coursework in Biology, Genetics, Moscow State University, Moscow, Russia — 2003–2006

Work Experience

Ph.D. researcher, Division of Pulmonary and Critical Care Medicine, Feinberg School of Medicine, Northwestern University, Chicago, USA — 2022–present
- Led 4 large scientific collaborative projects with 500+ patients and multimodal data to publication
- Acquired external funding for my training (competitive AHA predoctoral fellowship, $67,000)
- Authored and co-authored 18, including 4 first/co-first, publications or preprints
- Led analysis of scRNAseq patient samples in the context of their clinical course with novel machine learning approach (clustering of patient-day representations and late fusion)
- Identified cell population as a biomarker and potential therapeutic target of ILD in SSc (*bioRxiv* 2025)
- Consulted 7 Northwestern grad students, postdocs and faculty on deep learning, data science and data visualization, including setting up and training in paw tracking on videos for mouse experiments

Research data analyst, bioinformatics, Division of Pulmonary and Critical Care Medicine, Feinberg School of Medicine, Northwestern University, Chicago, USA — 2019–2022
- Created data processing pipelines, data exploration and management infrastructure for the division
- Delivered analytical insights from scRNAseq and other data to principal investigators for 7 publications
- Formulated activated T cell–macrophage circuit in *Nature* 2021, which supported successful clinical trials of Auxora in COVID-19 (NCT04345614)
- Supported grant writing for U19, R01 and other NIH grants for the division, resulting in $5M+ funding
- Hired and trained incoming data analysts to grow the team and replace myself

Head of maintenance tools development group, Yandex, Moscow, Russia — 2014–2017
- Managed a team of 6 engineers: hiring, mentoring, resolving conflicts, improving performance
- Synthesized internal customers' needs into technical roadmaps for supporting web-services
- Owned various web-services to improve employees' workflows

Full-stack software engineer, Yandex, Moscow, Russia — 2007–2014
- Automated deploy workflows of system administrators for better consistency and transparency
- Deployed and maintained various web-services to improve employees' workflows

Key publications

(* denotes first/co-first contribution)

* Profibrotic monocyte-derived alveolar macrophages as a biomarker and therapeutic target in systemic sclerosis-associated interstitial lung disease. ***bioRxiv***. 10.1101/2025.08.07.669006 ⟳ — 2025

Defining and benchmarking open problems in single-cell analysis. ***Nature Biotechnlogy***. 10.1038/s41587-025-02694-w ⟳ — 2025

* A distinctive evolution of alveolar T cell responses is associated with clinical outcomes in unvaccinated patients with SARS-CoV-2 pneumonia. ***Nature Immunology***. 10.1038/s41590-024-01914-w ⟳ — 2024

* Machine learning links unresolving secondary pneumonia to mortality in patients with severe pneumonia, including COVID-19. ***The Journal of Clinical Investigation (JCI)***. 10.1172/JCI170682 ⟳ — 2023

An integrated cell atlas of the lung in health and disease. ***Nature Medicine***. 10.1038/s41591-023-02327-2 ⟳ — 2023

UCSC Cell Browser: visualize your single-cell data. ***Bioinformatics***. 10.1093/bioinformatics/btab503 ⟳ — 2021

* Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. ***Nature***. 10.1038/s41586-020-03148-w ⟳ — 2021