David Novak

davidnovak9000@gmail.com david-novak-04b65989 davnovak.github.io

Bioinformatician with demonstrated proficiency in statistical data analysis, machine learning and software engineering

Employment History

2020–(Q4 2024) **Bioinformatics researcher.** FWO Strategic Basic research fellow at Saeys Lab, Center for Inflammation Research, VIB-UGent. Ghent, Belgium.
Research into applied ML and deep learning in cytometry and single-cell transcriptomics analysis pipelines: batch effect correction, large differential expression analyses. Algorithm development: interpretable dimensionality reduction.

2021–(Q4 2024) **Assistant lecturer (machine learning).** Department of Applied Mathematics, Computer Science and Statistics, Ghent University. Ghent, Belgium.
Guiding practical sessions and student projects. Consulting role for experimental design and HPC use.

2018–2020 **Programmer and bioinformatics researcher.** Childhood Leukaemia Investigation Prague, 2nd Faculty of Medicine, Charles University and Faculty Hospital Motol. Prague, Czech Republic.
Development of automated pipelines for processing of flow and mass cytometry clinical and research data. R and C++ development of a novel semi-automated trajectory inference solution.

2012–2020 **Translator and assistant.** MIDA Consulting. Prague, Czech Republic.
Czech-English & English-Czech translations and research into European Union subsidy programmes and maintenance of projects supported by them.

2018 **Lecturer (ESL).** Channel Crossings. Prague, Czech Republic.
Teaching individual English prep course for CAE qualification. English conversation course.

2017–2018 **Lecturer (algorithmisation).** Logiscool. Prague, Czech Republic.
Teaching principles of programming and algorithmisation to elementary-school students.

2015 **Research intern.** Cellular neurophysiology, Institute of Physiology, Czech Academy of Sciences. Prague, Czech Republic.
Primer design, bacterial transformation, plasmid preparation, IHC staining and fluorescent microscopy.

Education

2020 – (Q4 2024) **Ph.D. Bioinformatics** (ongoing) at Ghent University, Ghent, Belgium. Supervisor: Prof Yvan Saeys.
Topic: *Novel applications of structure learning and deep learning for single-cell data.*

2018 – 2020 **M.Sc. Bioinformatics** at Charles University, Prague, Czech Republic.
Thesis title: *Studying lymphocyte development using mass cytometry.*

2015 – 2018 **B.Sc. Biological Sciences** at Charles University, Prague, Czech Republic.
Thesis title: *Human lymphopoiesis and its examination via single-cell analysis.*

Published Research

Journal Articles

1 M. Bakardjieva, O. Pelák, M. Wentink, *et al.*, "Tviblindi algorithm identifies branching developmental trajectories of human b-cell development and describes abnormalities in rag-1 and was patients," *European Journal of Immunology*, vol. n/a, no. n/a, p. 2 451 004, DOI: https://doi.org/10.1002/eji.202451004. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/eji.202451004.

2. T. Liechti, S. Van Gassen, M. Beddall, *et al.*, "A robust pipeline for high-content, high-throughput immunophenotyping reveals age- and genetics-dependent changes in blood leukocytes," *Cell Reports Methods*, vol. 3, no. 10, p. 100 619, 2023, ISSN: 2667-2375. DOI: https://doi.org/10.1016/j.crmeth.2023.100619.

3. A. Couckuyt, R. Seurinck, A. Emmaneel, *et al.*, "Challenges in translational machine learning," *Human Genetics*, vol. 141, pp. 1–16, Sep. 2022. DOI: 10.1007/s00439-022-02439-8.

4. T. A. Liechti, M. Beddall, S. Van Gassen, *et al.*, "Leveraging high-dimensional flow cytometry to reveal the human immune system at a population-wide scale," *The Journal of Immunology*, vol. 206, no. 1*Supplement*, pp. 26.11–26.11, May 2021, ISSN: 0022-1767. DOI: 10.4049/jimmunol.206.Supp.26.11.

Conference Papers

1. D. Novak, S. Van Gassen, and Y. Saeys, "GroupEnc: Encoder with group loss for global structure preservation," in *BNAIC/BeNeLearn 2023*, Delft, Netherlands, 2023. URL: https://bnaic2023.tudelft.nl/static/media/BNAICBENELEARN_2023_paper_46.9317ce00beb72bf31803.pdf.

Pre-prints

1. D. Novak, C. de Bodt, P. Lambert, J. A. Lee, S. Van Gassen, and Y. Saeys, "Interpretable models for scRNA-seq data embedding with multi-scale structure preservation," 2024. DOI: 10.1101/2023.11.23.568428. eprint: https://www.biorxiv.org/content/early/2024/10/03/2023.11.23.568428.full.pdf.

2. J. Stuchly, D. Novak, N. Brdickova, *et al.*, "Deconstructing Complexity: A Computational Topology Approach to Trajectory Inference in the Human Thymus with tviblindi," May 2024. DOI: https://doi.org/10.7554/eLife.95861.1.

Skills and competencies

Languages	Strong reading, writing and speaking competencies for English, Czech, Slovak. Conversational German.
IT	Python, TensorFlow, PyTorch, sklearn, numba, matplotlib, R, ggplot2, Shiny, C++, OpenMP, .NET, xUnit, Java, JavaScript, React, HTML, CSS, SQL, Git, GitLab CI/CD, Slurm, Unix, Docker, AWS, NextFlow.
Misc.	Statistical data analysis, machine learning, NGS data analysis, computational cytometry and single-cell 'omics analyses, molecular biology theory, scientific writing, teaching.

In May 2024 I co-organised the Computational Cytometry Summer School at VIB-UGent, teaching a tutorial on statistical analyses and interpretation of experimental cytometry data.

In 2019 I co-taught R data analysis and statistics at the 2nd Faculty of Medicine, Charles University.

In 2014-2015 I volunteered at the Thomayer hospital in Prague as part of the Lékořice foundation, holding weekly visits with elderly hospitalised patients.

References

References from employers, colleagues and collaborators will be made available upon request.