Mikhail Kolmogorov

Tenure-Track Standtman Investigator
Cancer Data Science Laboratory
National Cancer Institute

Email: *mikhail.kolmogorov@nih.gov*
Web: *https://ccr.cancer.gov/staff-directory/mikhail-kolmogorov*

Research Interests

The focus of my research is *computational biology* - algorithms, mathematical models and tools aimed to answer fundamental questions about living systems through the analysis of large-scale biological data. Specifically, I have extensive experience in computational genomics and proteomics, with a focus on the design and development of efficient algorithmic approaches that bridge theoretical models and real experimental data.

Research Positions

- **Tenure-Track Stadtman Investigator** *Jan 2022 - now*
  National Cancer Institute, Bethesda, MD, USA

- **Posdoctoral Fellow** *Jan 2021 - Dec 2021*
  University of California, Santa Cruz, USA
  Advisors: Benedict Paten

- **Posdoctoral Fellow** *Oct 2019 - Dec 2020*
  University of California, San Diego, USA
  Advisors: Rob Knight and Pavel Pevner

Education

- **PhD, Computer Science** *Sep 2014 - Sep 2019*
  Department of Computer Science and Engineering,
  University of California, San Diego, USA
  Thesis: Algorithms for long-read assembly
  Advisor: Pavel Pevzner

- **MSc, Bioinformatics** *Sep 2012 – Jul 2014*
  St. Petersburg Academic University, Russia
  Thesis: Genome assembly using multiple references
  Advisor: Son Pham

- **BSc, Applied Mathematics and Informatics** *Sep 2008 - Jul 2012*
  Department of Mathematics, ITMO University,
  St.Petersburg, Russia
  Thesis: Variational method for the minimum energy path calculation
  Advisor: Igor Popov

Publications

- **Mikhail Kolmogorov***, Kimberley J Billingsley*, Mira Mastoras, Melissa Meredith, Jean Monlong, Ryan Lorig-Roach, (17 authors), Karen H Miga, Miten Jain, Winston Timp, Adam M Phillippy, Mark Chaisson, Fritz J Sedlazeck, Cornelis Blauwendraat*, Benedict Paten*, “Scalable Nanopore sequencing of human genomes provides a comprehensive view of haplotype-resolved variation and methylation”, *Nature Methods* (2023) PMID: 37710018, *corresponding authors

- Isabel Rodriguez, Nicole M Rossi, Ayse Keskus, Yi Xie, Tanveer Ahmad, Asher Bryant, Hong Lou, Jesica Godinez Paredes, Rose Milano, Nina Rao, Sonam Tulsyan, Joseph F Boland, Wen Luo, Jia Liu, Tim O'Hanlon, Jazmyn Bess, Vera Mukhina, Daria Gaylalova, Yuko Yuki, Laksh Malik, Kimberley Billingsley, Cornelis Blauwendraat, Mary Carrington, Meredith Yeager, Lisa Mirabello, **Mikhail Kolmogorov**, Michael Dean. “Insights into the Mechanisms and Structure of Breakage-Fusion-Bridge Cycles in Cervical Cancer using Long-Read Sequencing”, AJHG (2024)

- Ekaterina Kazantseva, Ataberk Donmez, Mihai Pop, **Mikhail Kolmogorov. “**stRainy: assembly-based metagenomic strain phasing using long reads”. bioRxiv preprint (2023) https://doi.org/10.1101/2023.01.31.526521

- Derek M. Bickhart*, **Mikhail Kolmogorov***, Elizabeth Tseng, Daniel Portik, Anton Korobeynikov, Ivan Tolstoganov, Gherman Uritskiy, *(9 authors)* & Tim Smith "Generation of lineage-resolved complete metagenome-assembled genomes by precision phasing." *Nature Biotechnology, 40 (5), 711-719 (2022)*

- Dmitry Antipov, Mikhail Rayko, **Mikhail Kolmogorov** & Pavel Pevzner. “viralFlye: assembling viruses and identifying their hosts from long-read metagenomics data”. *Genome Biology, 23 (1), 1-21 (2022)*

- Meyer, F., Fritz, A., Deng, Z. L., Koslicki, D., Gurevich, A., Robertson, G., (28 authors), **Kolmogorov M**., (64 authors), McHardy, A. C. “Critical Assessment of Metagenome Interpretation-the second round of challenges”. *Nature methods 19 (4), 429-440 (2022)*

- S Nurk, S Koren, A Rhie, M Rautiainen, AV Bzikadze, A Mikheenko, (50 authors), M **Kolmogorov**, (65 authors), Karen H Miga, Adam M Phillippy, "The complete sequence of a human genome." *Science* 376 (6588), 44-53 (2022)

- A Bankevich, A Bzikadze, **M Kolmogorov**, D. Antipov & PA Pevzner. “Multiplex de Bruijn graphs enable genome assembly from long, high-fidelity reads”. *Nature Biotechnology 1-7 (2022)*

- ED Jarvis, G Formenti, A Rhie, A Guarracino*, (35 authors),* **M Kolmogorov***, (45 authors)* Kerstin Howe, Karen H Miga, Human Pangenome Reference Consortium “Automated assembly of high-quality diploid human reference genomes*”, bioRxiv (2022) https://doi.org/10.1101/2022.03.06.483034*

- Shafin, K., Pesout, T., Chang, P.C., Nattestad, M., Kolesnikov, A., Goel, S., Baid, G., **Kolmogorov, M**., Eizenga, J.M., Miga, K.H. and Carnevali, P.. “Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads”. *Nature Methods 18 1322–1332 (2021)*

- **Mikhail Kolmogorov,** Derek M. Bickhart, Bahar Behsaz, Alexey Gurevich, Mikhail

Rayko, Sung Bong Shin, Kristen Kuhn, Jeffrey Yuan, Evgeny Polevikov, Timothy P. L. Smith and Pavel A. Pevzner. “metaFlye: scalable and accurate long-read metagenome assembler”. *Nature Methods* (2020) *doi: 10.1038/s41592-020-00971-x*

- **Mikhail Kolmogorov,** Jeffrey Yuan, Yu Lin and Pavel Pevzner. “Assembly of long error-prone reads using repeat graphs”. *Nature Biotechnology* 37 540-546 (2019)

- Alla Mikheenko and **Mikhail Kolmogorov. “**Assembly Graph Browser: interactive visualization of assembly graphs”. *Bioinformatics* 35 3476-3478 (2019)

- Evgeny Polevikov and **Mikhail Kolmogorov. “**Synteny paths for assembly graphs comparison”. *Workshop on Algorithms in Bioinformatics 2019, Niagara Falls, NY*

- **Mikhail Kolmogorov**, Joel Armstrong, Brian J. Raney, Ian Streeter, Matthew Dunn, Fengtang Yang, Duncan Odom, Paul Flicek, Thomas Keane, David Thybert, Benedict Paten and Son Pham. “Chromosome assembly of large and complex genomes using multiple references”. *Genome Research* 28 (11), 1720-1732 (2018)

- Alex Bishara, Eli L Moss, **Mikhail Kolmogorov**, Alma Parada, Ziming Weng, Arend Sidow, Anne E Dekas, Serafim Batzoglou, Ami S Bhatt. “High-quality genome sequences of uncultured microbes by assembly of read clouds”. *Nature Biotechnology* 36, 1067–1075 (2018)

- Jingtao Lilue, Anthony G Doran, Ian T Fiddes, Monica Abrudan, Joel Armstrong, (22 authors), **Mikhail Kolmogorov,** (29 authors), Benedict Paten, Thomas M Keane. “Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci”. *Nature Genetics 50,* 1574–158 (2018)

- David Thybert, Maša Roller, Fábio CP Navarro, Ian Fiddes, Ian Streeter, Christine Feig, David Martin-Galvez, **Mikhail Kolmogorov**, (33 authors), Benedict Paten, Son Pham, Thomas M Keane, Duncan T Odom, Paul Flicek. “Repeat associated mechanisms of genome evolution and function revealed by the Mus caroli and Mus pahari genomes”. *Genome Research 28 (4),* 448-459 (2018)

- **Mikhail Kolmogorov**, Eamonn Kennedy, Zhuxin Dong, Gregory Timp and Pavel Pevzner. “Single-Molecule Protein Identification by Sub-Nanopore Sensors“. *PloS Computational Biology* 13 (5), e1005356 (2017)

- Yu Lin*, Jeffrey Yuan*, **Mikhail Kolmogorov***, Max Shen and Pavel Pevzner. “Assembly of long error-prone reads using de Bruijn graphs”. *PNAS 113 (52),* E8396-E8405 (2016)*,* *equal contribution

- **Mikhail Kolmogorov**, Xiaowen Liu and Pavel Pevzner. “SpectroGene: a tool for proteogenomic annotations using top-down spectra”. *Journal of Proteome Research* 15 (1), 144-151 (2015)

- **Mikhail Kolmogorov**, Brian Raney, Benedict Paten, Son Pham. “Ragout - a reference-assisted assembly tool for bacterial genomes”. *ISMB 2014, Boston, MA*

- Ilya Minkin, Anand Patel, **Mikhail Kolmogorov**, Nikolay Vyahhi, Son Pham**. “**Sibelia: a scalable and comprehensive synteny block generation tool for closely related microbial genomes”. *Workshop on Algorithms in Bioinformatics 2013, Antibes, France*

Invited Conference Talks

- “Severus: accurate detection and characterization of somatic structural variation in tumor genomes using long reads”, AACR 2024, sponsored by Oxford Nanopore, San Diego 2024
- “Microbiome strain profiling using long reads”. Microbiome Data Congress, Boston, MA 2023
- “Nanopore sequencing reveal structural heterogeneity in canine osteosarcoma”, *Nanopore Community Meeting*, Houston, TX 2023
- “Towards population-scale nanopore sequencing: characterizing variation and methylation in 1000s brain genomes at NIH CARD”, *Biology of Genomes, Cold Spring Harbor, NY 2023*
- “Profiling Structural Variants and Complex Rearrangements in Cancer Genomes Using Long-Reads”, *TriOmics Summit, Boston, MA, 2022*
- “Long-read metagenomic assembly with metaFlye”, *Long Read Sequencing Workshop, The Jackson Laboratory, Farmington, CT, 2022*
- “Long-read assembly: are metagenomes easier than diploid genomes?”. *PAG 2020, San Diego, CA*
- “Long read genome and metagenome assembly using repeat graphs”. *Oxford Nanopore Community Meeting (virtual)*, 2020
- “Long read genome and metagenome assembly using repeat graphs”. *Annual Conference of the Russian-American Science Association (virtual)*, 2020
- “Assembly of long error-prone reads using repeat graphs”. *PAG 2018, San Diego, CA*

Invited Seminar Talks

- “De novo assembly of metagenomic strains using long reads”, DCEG Microbiome journal club, 2024
- “Napu and Severus: accurate and complete germline and somatic structural variation calling”, Oxford Nanopore Technologies webinar, 2024
- “Profiling Structural Variants and Complex Rearrangements in Cancer Genomes Using Long-Reads”, Children’s Mercy, Univ. of Missouri, 2023
- Structural Variation Analysis Using Long Reads, NCI Spring School on Algorithmic Cancer Biology, 2023
- “Profiling Structural Variation in “Healthy” and Cancer Genomes Using Long Reads”, Genetics Branch, NCI, 2023
- “Analyzing complex somatic rearrangements in cancer genomes using nanopore sequencing”, Oxford Nanopore webinar series, 2022

- “Profiling Structural Variants and Complex Rearrangements in Cancer Genomes Using Long-Reads”, Laboratory of Cell Biology seminar, National Cancer Institute 2022

- “Long-read, assembly-based methods for structural variant discovery”, Long Read Sequencing Seminar, Baylor College of Medicine (virtual) 2022

- “Algorithms for genome and metagenome assembly using long reads”, National Library of Medicine, NIH seminar series, 2022

- “Challenges in genome and metagenome assembly with long reads”, National Human Genome Research Institute, NIH seminar series, 2022

- “Algorithms for genome and metagenome assembly using long reads”, Computer Science Seminar, University of Maryland, 2022

- “Generation of lineage-resolved complete metagenome-assembled genomes in complex microbial communities”. *PacBio SMRT Science Journal Club Seminar (virtual),* 2021

- “Completing human genome and microbiome: algorithms for long-read assembly”, *NYU Center for Human Genetics and Genomics symposium*, 2020

- “Metagenomic assembly using long reads”, Bioinformatics seminar series, *ITMO University, St. Petersburg, Russia*, 2020

- “Metagenomic assembly using long reads”, Metagenomic seminar series, *Universidad Miguel Hernandez*, 2020

- “Completing human genome and microbiome: algorithms for long-read assembly”, *Cornell Computational Biology symposium*, 2020

- “Assembly of long error-prone reads using repeat graphs”. *Center for Microbiome Innovation at UCSD*, 2019

- “Assembly of long error-prone reads using repeat graphs”. *Australian National University, Long-read, long-range workshop*, 2019

Contributed Conference Talks

- “metaFlye: Scalable long-read metagenome assembly using repeat graphs”. *Genome Informatics 2019, Cold Spring Harbor, NY* (selected abstract)

- **“**Synteny paths for assembly graphs comparison”. *Workshop on Algorithms in Bioinformatics 2019, Niagara Falls, NY* (proceedings author)

- “Assembly of long error-prone reads using repeat graphs”. *AGBT 2019, Marco Island, FL* (selected abstract)

- “Single-Molecule Protein Identification by Sub-Nanopore Sensors”. *HUPO 2018, Orlando, FL* (selected abstract)

- “Assembly of long error-prone reads using repeat graphs”. *RECOMB* 2018, *Paris, France* (proceedings author)

- “Single-Molecule Protein Identification by Sub-Nanopore Sensors“. *RECOMB-SEQ 2016, Santa Monica, CA* (proceedings author)

- “SpectroGene: a tool for proteogenomic annotations using top-down spectra”. *ASMS 2015, St. Louis, MO* (selected abstract)

- “Ragout - a reference-assisted assembly tool for bacterial genomes”. *ISMB 2014, Boston, MA* (proceedings author)

Mentoring and Advising

- Ekaterina Kazantseva, “Combining Genome Graphs and Phasing Information to Generate Complete Bacterial Haplotypes”. Master thesis, ITMO University, St. Petersburg, Russia, 2021-2022

- Mentor in the Science Mentors program for Russian-speaking students (*https://sites.google.com/view/thesciencementors/en*), 2020-2021

- Evgeny Polevikov. “Synteny paths for assembly graphs comparison”
  Master thesis, ITMO University, St. Petersburg, Russia, 2018-2019

- Alla Mikheenko. “Assembly Graph Browser: interactive visualization of assembly graphs”
  Research project, St. Petersburg Center for Algorithmic Bioinformatics, Russia, 2018-2019

Teaching

- Teaching assistant *Winter 2016*
  Algorithms in Molecular Biology. UC San Diego

- Teaching assistant *Spring 2013*
  Bioinformatics Seminar. St. Petersburg Academic University

- Lecturer, Bioinformatics Summer School 2016
  Moscow, Russia

- Lecturer, Bioinformatics Summer School 2014
  St. Petersburg, Russia

Fellowships

- CSE Doctoral Fellow Award (UCSD) Sep *2014 - Aug 2016*

Open Source Projects

- **Flye** – De novo assembler for single molecule sequencing reads (*main developer)*
  https://github.com/fenderglass/Flye

- **HapDup** - Generating phased diploid assembly using long reads only (*main developer*)
  https://github.com/fenderglass/HapDup

- **Ragout** – Chromosome-level scaffolding using multiple references (*main developer)* *https://github.com/fenderglass/Ragout*

- **Nano-Align** – Protein identification using a nanopore *(main developer)*
  https://github.com/fenderglass/Nano-Align

- **SpectroGene** – Gene prediction using top-down mass-spectrometry *(main developer)*
  https://github.com/fenderglass/SpectroGene

Industry Experience

- Research Intern *Summer 2017*
  Illumina, San Diego, CA, USA
  Mentors: Dorna Kashef and Serafim Batzoglou