

# Curriculum Vitae

Johannes Köster

July 13, 2024

## Personal information

**Homepage:** <https://koesterlab.github.io>

**Birth:** May 5, 1985 in Gelsenkirchen, Germany

**Nationality:** German

**Family status:** Married, two children

## Positions

### Current positions

**since 2023: Full professor (W3, tenure track)** for Bioinformatics and Computational Oncology (Lehrstuhl Bioinformatische Algorithmen in der Onkologie) at the University of Duisburg-Essen.

**since 2020: DKTK Investigator** at the German Consortium of Translational Cancer Research (DKTK) of WTZ and DKFZ.

**since 2023: Scientific coordinator** for the core “computational biomedicine” in the BIOME graduate school of the University of Duisburg-Essen.

**since 2016: Affiliated consulting scientist** in Myles Brown's group (Division of Molecular and Cellular Oncology, Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School), USA.

### Previous positions

**2017 – 2023: Group leader** (permanent position), Algorithms for reproducible bioinformatics (<https://koesterlab.github.io>), Genome Informatics, Institute of Human Genetics, Faculty of Medicine, University of Duisburg-Essen. Currently eight group members (graduate students).

**2016 – 2017: Researcher** (NWO Veni funded, independent) at Life Sciences, Centrum Wiskunde & Informatica, Amsterdam, Netherlands.

**2016:** **Postdoctoral research fellow** under the supervision of Alexander Schönhuth, Life Sciences, Centrum Wiskunde & Informatica, Amsterdam, Netherlands.

**2016:** **Consultant** (data analysis and workflow management) for Juno therapeutics, 307 Westlake Avenue, Seattle, USA.

**2015 – 2016: Postdoctoral research fellow** in Xiaole Shirley Liu's group (Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health) and Myles Brown's group (Division of Molecular and Cellular Oncology, Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School), USA.

**2011 – 2015: Research fellow** at both the Chair of Genome Informatics of Sven Rahmann, University of Duisburg-Essen and in the “European Network for Cancer in Children and Adolescents” project under Alexander Schramm, University Hospital Essen, Germany.

## Education

**2011-2015: Ph.D. in Computer Science (grade “ausgezeichnet”/summa cum laude)** at TU Dortmund, Germany. Thesis title: “Parallelization, Scalability, and Reproducibility in Next-Generation Sequencing Analysis”. Supervisor: Sven Rahmann, Chair of Genome Informatics, University of Duisburg-Essen. Second assessor: Axel Mosig, Bioinformatics Group, Ruhr University Bochum.

**2005 – 2010: Diploma (equiv. to M.Sc.) in Computer Science (grade 1.0/A with honors)** at TU Dortmund, Germany. Thesis title: “Propagating Interaction Logic toward Predictive Protein Hypernetworks”. Supervisors: Sven Rahmann, Bioinformatics for High-Throughput-Technologies, Computer Science XI, TU Dortmund and Eli Zamir, Systems Biology of Cell-Matrix Adhesion, Department of Systemic Cell Biology, Max Planck Institute of Molecular Physiology Dortmund.

## Awards

- 2019: **Friedmund-Neumann-Prize** for outstanding fundamental research on reproducibility in biomedicine. Schering Foundation. Amount: 10,000 €
- 2011: **Hans-Uhde-Prize** by the ThyssenKrupp Uhde GmbH for extraordinary diploma thesis with practical application in Computer Science. Amount: 500 €
- 2021: **Young scientist best paper award** of the Faculty of Medicine, University of Duisburg-Essen for the paper "Varlociraptor: enhancing sensitivity and controlling false discovery rate in somatic indel discovery".
- 2019: **Young scientist best paper award** of the Faculty of Medicine, University of Duisburg-Essen for the paper "Bioconda: sustainable and comprehensive software distribution for the life sciences".
- 2014: **NVIDIA Hardware Grant** received for publication "Massively parallel read mapping on GPUs with the q-group index and PEANUT". Amount: 2000 € (Tesla GPU)

## Organized Events

- Nov 2021: **PC member:** Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB) 2021
- Feb 2019: **Co-Organizer:** Data structures in Bioinformatics (DSB) 2019, Dortmund, Germany
- Jun 2018: **Co-Organizer:** Lorentz workshop, Leiden, The Netherlands: Making sense of millions of single cells: approaching the era of Single Cell Data Science, experimental advances and computational challenges
- Sep 2017: **PC member:** CIBB 2017, Cagliari, Italy
- Mar 2017: **Organizer:** Tutorial "Reproducible data analysis with Snakemake" at CWI Amsterdam
- Feb 2017: **Co-Organizer and session chair:** workshop "data structures in bioinformatics" (DSB) 2017 (<https://dsb2017.github.io>)

## Invited talks

- Feb 2022 **Cambridge Genetics Seminar:** Reproducibility of NGS data analysis
- Dec 2021 **Helmholtz Center Dresden:** Sustainable data analysis
- Jul 2021 **Intelligent Systems for Molecular Biology Conference (ISMB), BIOINFO-CORE track:** Keynote on Snakemake
- Aug 2020 **University Medicine Mainz:** Reproducibility and transparency in bioinformatics
- May 2019 **NIH:** Reproducible data analysis with Snakemake
- May 2019 **Harvard University:** Towards a unified theory of variant calling
- Feb 2019 **Sanger Institute:** Reproducible data analysis with Snakemake
- Dec 2018 **Martin Luther University Halle-Wittenberg:** Reproducible data analysis with Snakemake
- Sep 2018 **Bernstein Conference on computational neuroscience, Berlin:** Reproducible data analysis with Snakemake
- Nov 2017 **Robert-Koch-Institute, Berlin:** Benchmarking RNA-seq data analysis
- Feb 2017 **Gutenberg University Mainz:** Snakemake and Bioconda
- Feb 2017 **Forschungszentrum Jülich:** Snakemake and Bioconda
- Nov 2016 **Helmholtz Open Science Workshop, Dresden, Germany:** Keynote about sustainable software development in science.
- Sep 2016 **Netherlands Cancer Institute (NKI), Amsterdam:** Snakemake and Bioconda.
- Jul 2016 **Rust bay area meeting, Mozilla Foundation Los Angeles:** Bioinformatics with Rust.
- Jun 2016 **Aquatic Ecosystem Research, University of Duisburg-Essen:** Snakemake and Bioconda.
- Jun 2016 **Leiden University Medical Center, Netherlands:** Algebraic variant calling, Snakemake and Bioconda.
- Jul 2015 **Broad Institute, Boston, USA:** Workflow Management with Snakemake
- Apr 2014 **Workshop on Fast Data Processing on GPUs, CUDA Center of Excellence, Dresden:** Massively parallel read mapping on graphics cards.
- Apr 2014 **DTL Focus Meeting, Utrecht, Netherlands:** Reproducible data analysis with Snakemake

## Teaching

- SS 2019** **Fachprojekt**<sup>1</sup> “Algorithm Engineering: Entwicklung einer Rust-Bibliothek am Beispiel von Wavelet Trees”, Faculty of Computer Science, TU Dortmund (4 SWS)
- WS 2018** **Fachprojekt** “Bioinformatik: Reproduzierbare Datenanalyse mit Snakemake am Beispiel der Bioinformatik”, Faculty of Computer Science, TU Dortmund (4 SWS)
- SS 2018** **Fachprojekt** “Algorithm Engineering: Entwicklung einer Rust-Bibliothek am Beispiel von Succinct Trees”, Faculty of Computer Science, TU Dortmund (4 SWS)
- since 2016** **18 invited tutorials** on reproducible data analysis with Snakemake (duration: 1 or 2 days; Sanger Institute, Harvard University, NIH, ETH Zürich, Berlin Institute of Health, Humboldt University Berlin, Norwegian University of Science and Technology, CWI Amsterdam, Universiteit Utrecht, INRA Toulouse, University of Duisburg-Essen, Wageningen University, AstraZeneca, RTG WisPerMed, Physalia Courses, Vlaams Instituut voor Biotechnologie).
- SS 2013** **One guest lecture** about variant calling with next-generation sequencing in the course “Statistic in Genetics”, Faculty of Statistics, TU Dortmund, Germany.
- WS 2012** **Two guest lectures** about protein networks, complex prediction and interaction dependencies in the course “Computational Omics”, Faculty of Computer Science, TU Dortmund, Germany.
- WS 2011** **Teaching assistant** for “Datastructures, Algorithms and Programming 1”, at the Chair of Computer Science X, TU Dortmund, Germany. (2 SWS)
- 2007 – 2009** **Student teaching assistant** for “Datastructures, Algorithms and Programming 1”, at the Chair of Computer Science X, TU Dortmund, Germany. (2 SWS)

## Reviewing

- since 2022** Grant application reviewer for Biotechnology and Biological Sciences Research Council (BBSRC), UK
- since 2021** Grant application reviewer for National Science Centre (NCN), Poland
- since 2024** Grant application reviewer for German Ministry of Research (BMBF)
- since 2012** Reviewer for various journals (Nature, Nature Biotechnology, Nature Communications, Nature Methods, Bioinformatics, BMC Bioinformatics, Genome Biology, Cell Systems, PLOS Computational Biology, GigaScience, PeerJ)

## Ten most important publications

1. Tüns, A.I., Hartmann, T., Magin, S., González, R.C., Henssen, A.G., Rahmann, S., Schramm, A., **Köster, J. (shared last author)** Detection and Validation of Circular DNA Fragments Using Nanopore Sequencing. *Frontiers in Genetics* 13 (2022).
2. Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., **Köster, J. (corresponding author)**. Sustainable data analysis with Snakemake. *F1000Res* 10, 33 (2021).
3. **Köster, J. (shared first and corresponding author)**, Dijkstra, L. J., Marschall, T., Schönhuth, A. Varlociraptor: enhancing sensitivity and controlling false discovery rate in somatic indel discovery. *Genome Biology* 21, 98 (2020).
4. Lähnemann, D., **Köster, J. (shared first and last author)**, Szczurek, E., McCarthy, D. J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., et al. Eleven grand challenges in single-cell data science. *Genome Biology* 21, 31 (2020).
5. **Köster, J. (corresponding author)**, Brown, M., Liu, X.S. A Bayesian model for single cell transcript expression analysis on MERFISH data. *Bioinformatics* 35, 995–1001 (2019).
6. Grüning, B., Dale, R., Sjödin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valieris, R., the Bioconda Team, **Köster, J. (corresponding author)**. Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat Methods* 15, 475–476 (2018).
7. **Köster, J.** Rust-Bio: a fast and safe bioinformatics library. *Bioinformatics* 32, 444–446 (2016).
8. Li, W., **Köster, J. (co-first author)**, Xu, H., Chen, C.-H., Xiao, T., Liu, J. S., . . . Liu, X. S. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biology*, 16(1), 281 (2015).
9. **Köster, J. (corresponding author)**, Rahmann, S. Massively parallel read mapping on GPUs with the q-group index and PEANUT. *PeerJ* 2, e606 (2014).
10. **Köster, J. (corresponding author)**, Rahmann, S. Snakemake – a scalable bioinformatics workflow engine. *Bioinformatics* 28, 2520–2522 (2012).

---

<sup>1</sup>A “Fachprojekt” is a hybrid 1-semester course in which students solve a computer science problem in small groups. It thereby consists of lecture, exercise, teamwork, and seminar components.