Katarzyna Wręczycka, Ph.D.

Leonhardstrasse 1, 8001 Zurich, Switzerland
+48 503 777 010 / +41 178 4466628
katwre@gmail.com
katwre.github.io
github.com/katwre
linkedin.com/in/katwre
scholar.google.com/citations?user=katarzynawreczycka

Profile

Bioinformatics Data Scientist with a PhD in Computational Molecular Biology and over 10 years of experience. Expert in applying machine learning to large-scale genomic and multi-omics data for biomarker discovery and (epi)genomics data analysis. Proficient in Python and R programming, data visualization, and creative thinking. Thrives in dynamic, international, and highly collaborative environments.

Experience

04.2025 – 06.2025

Bioinformatician | Research Assistant II

ETH Zürich & Functional Genomics Center Zürich (FGCZ) of the University of Zurich (UZH)

- von Meyenn Lab (40% position):
  - **RNA-seq and CUT&RUN** analysis of adipose tissue to investigate epigenetic memory of obesity after weight loss.
  - Maintenance of lab's **NextFlow Tower** with the ETH Zürich cluster (**SLURM**), with **Singularity** containers and **Spack** for reproducible workflows.
- Bar-Nur Lab (40% position):
  - **Single-cell RNA-seq and WGS** analysis for generating functional xenogeneic germ cells in rodents via blastocyst complementation. Assembly of chromosome Y in *Mus minutoides*.
  - **RNA-seq** analysis aiming to assess pathways driving muscle aging to target sarcopenia with natural compounds.
- Functional Genomics Center Zurich University of Zurich (20% position):
  - **Deployed Shiny applications** via ShinyProxy, accessible at shiny-public.fgcz.uzh.ch for the von Meyenn's lab (snRNA-seq: MHUO, Mouse epiAT memory, Human AT memory)
  - Running and maintaining **reproducible workflows in Ruby on Rails, R and bash** on the FGCZ servers for the Bar-Nur's lab.
  - **Single-nucleus RNA-seq** analysis contributing to a cell-level kidney reference atlas comparing ATG7-dependent autophagy and proximal tubule dysfunction in disease versus healthy tissue.

02.2023 – 01.2025

Bioinformatician

Remote from Berlin, Germany

Engaged in B2B contracts with bioinformatics, data science and software consultancy company Ardigen (CRO department) and with Selvita S.A., as well as freelance consultancy. Collaborated on diverse projects with pharmaceutical companies and startups. Key highlights include:

- **Machine learning for drug development:** Applied **positive unlabeled learning with random forests and SVM** (with pulearn library) and **autoencoder** (in PyTorch) to identify transcription factors as drug targets, integrating clinical trial insights (Merck, Oncology Data Science Department).
- **Single-cell RNA-seq analysis:** Studied the cGAS–STING pathway, immune response, and DNA damage in cancer (in R/Seurat)(Merck, Oncology Bioinformatics Department).

- **Biomarker discovery:** Collaborated closely with data scientists and experimental scientists in R&D and Clinical Development teams to identify tumor-associated antigens and gene signatures for patients with limited treatment options. Leveraged RNA-seq, WGS, external **clinical trial** databases, and performed **survival analysis** to support therapeutic decision-making (Merck, Oncology Bioinformatics Department).
- **Integrative omics data analysis:** investigated glia-to-neuron reprogramming in Alzheimer's disease by integrating RNA-seq (differential gene expression), ATAC-seq (chromatin accessibility), ChIP-seq (H3K4me2/H3K27me3), and BS-seq (DNA methylation). Built reproducible **nextflow** pipelines and applied **R/Bioconductor** tools (DESeq2, methylKit, genomation) for data processing and analysis. Performed enhancer motif selection in enhancers (MonaLisa; Machlab et al., 2022), Gene Ontology and GSEA, and GWAS enrichment. Collaborated with a data scientist to apply **transformer-based model** (**Enformer**, Avsec et al., Nature Methods, 2021) to identify and characterize target genes and regulatory enhancers (Stardustries).
- **Software development for web applications:** Enhanced genomic data visualization and integrated in-house and publicly available sequencing (epi)genomic data in IGV app browser using **JavaScript and Python** (Novo Nordisk).
- Gene fusion detection pipeline: Consulted on setting up a **nextflow** pipeline to detect and visualize gene fusions in dog cell-free DNA (testblu diagnostics).
- **Cloud infrastructure and workflow automation:** Built **Docker** images, utilized Kubernetes, and deployed workflows on **AWS**.

10.2015 – 11.2021

PhD Student in Computational Molecular Biology

Bioinformatics and Omics Data Science Platform at the Berlin Institute of Medical Systems Biology, Max Delbrück Center in Berlin.

- **Epigenomics and ChIP-seq Analysis**: Investigated high-density transcription factor binding sites and identified spurious ChIP-seq peaks linked to RNA structures called DNA-RNA hybrids (R-loops), and hypo-methylated regions using **elastic net regression and PCA** [2]. Integrated data from publicly available resources, including ENCODE and Roadmap Epigenomics, as well as experimental datasets from ChIP-seq, ATAC-seq, Bisulfite-seq, and DRIP-seq/RDIP-seq studies.
- **R packages Development**: Maintained, developed and provided user support for the **Bioconductor R package** genomation (github.com/BIMSBbioinfo/genomation). Developed R package motifActivity for transcription factor network reconstruction from sequencing data using rank-based estimation linear models (github.com/katwre/motifActivity).
- **Co-authored a peer-reviewed review on statistical methods** for analyzing Bisulfite-seq data (accompanied by publicly available code at github.com/BIMSBbioinfo/Strategies_for_analyzing_BS-seq) [4].
- **Bisulfite-seq Pipeline**: Implemented a scalable **snakemake** pipeline for Bisulfite-seq data analysis using TrimGalore!, BWA-meth/Bismark, samblaster, MethylKit/methylDackel, and genomation R packages, published as part of the PiGX: Pipelines in Genomics project [3]. Tested and applied it on the MDC BIMSB cluster (**Grid Engine**) and BIH cluster (**SLURM**).
- **DNA Methylation Biomarkers for Cardiovascular Disease**: Identified DNA methylation biomarkers for acute coronary syndrome by analyzing Bisulfite-seq data from cell-free DNA of heart tissue and plasma, uncovering gene and transcription factor associations. My contribution included preprocessing raw Bisulfite-seq data, calling differentially methylated regions, associating them with genes, and transcription factors in cardiovascular diseases [1].
- **Omics data analysis**: Conducted omics analysis of high-risk pediatric neuroblastomas, identifying DNA methylation biomarkers and transcription factor networks specific to high-risk cases from solid tumors and cell-free DNA in urine. Applied **machine learning methods** (incl. **elastic net, ridge regression, XGBoost**) and integrated Bisulfite-seq, RNA-seq, and ChIP-seq datasets.

03 – 09.2015

Visiting Predoctoral Researcher

Bioinformatics and Omics Data Science Platform at the Berlin Institute of Medical Systems Biology, Max Delbrück Center in Berlin.

Awarded a €7,000 scholarship by the Institute of Computer Science, Polish Academy of Sciences (IPI PAN), co-financed by the European Union. Supervised by Dr. Altuna Akalin.

- Developed and maintained **Bioconductor R package genomation**, designed for the annotation, visualization, and summary of genomic data (github.com/BIMSBbioinfo/genomation).
- Investigated phantom peaks (false positives) in publicly available ChIP-seq datasets, contributing foundational work later expanded during Ph.D. research [2].
- Collaborated with research teams at IPI PAN, delivering monthly progress presentations and detailed reports.

Education

10.2015 – 05.2021

Ph.D. in Computational Molecular Biology

Humboldt University, Department of Biology, Berlin, Germany & Berlin Institute for Medical Systems Biology at the Max Delbrück Center for Molecular Medicine, Berlin, Germany

Dissertation: "Integrative analysis on the effect of DNA methylation on gene regulation in tumor and healthy tissues". Advisors: Dr. Altuna Akalin and Prof. Dr. Uwe Ohler

10.2012 – 06.2014

M.Sc. in Bioinformatics and Systems Biology

University of Warsaw, Department of Mathematics, Informatics and Mechanics, Warsaw, Poland

Dissertation: "Genome-wide analysis of noncoding copy-number variations in human genome". Advisor: Prof. Dr hab. Anna Gambin

10.2010 – 06.2012

B.Sc. in Bioinformatics and Systems Biology

University of Warsaw, Department of Mathematics, Informatics and Mechanics, Warsaw, Poland

Dissertation: "Potential regulatory region discovery method implementation". Advisor: Dr hab. Bartosz Wilczynski. I developed an algorithm based on Bernoulli distribution to identify conserved noncoding regulatory regions (github.com/katwre/bioinf/bio_motif_ensembl)

Technical skills

- **Programming & Tools:** In-depth knowledge of statistical programming, data analysis, and visualisations in R (CRAN/Bioconductor), Python (Biopython, NumPy, pandas, scikit-learn, seaborn, PyTorch); experienced in LaTeX, bash, and Javascript; unit tests (R/testthat; Python/unittest/pytest)
- **Data Analysis:** Statistical analysis of (epi)genetic data, including statistical tests (incl. t-tests, Wilcoxon tests), regression models (e.g., multivariate logistic regression, elastic net, survival analysis/Cox proportional hazards model), and classification algorithms (incl. random forest, SVM, KNN); unsupervised learning techniques, such as clustering (hierarchical clustering, K-means, EM algorithm) and dimensionality reduction (incl. PCA, t-SNE); and applied deep learning approaches (autoencoder).
- **Bioinformatics Pipelines:** Experienced in building and development of complex pipelines using workflow languages (snakemake, nextflow), and workflow engines (Galaxy)
- **High-Performance Computing:** Experienced in using traditional high-performance computing clusters and job schedulers (Grid Engine and SLURM) and Kubernetes, Database Engines (MySQL, PostgreSQL); created Docker/Singularity images; cloud computing - AWS and DigitalOcean
- **Version Control and Software Management:** proficient in Linux/Unix systems, git version control system, conda and GNU Guix package management systems
- **NGS Tools & Databases:** incl. samtools, BEDtools, GATK, IGV, Bowtie2, BWA, Bismark, BLAST and omics databases: incl. TCGA, RoadmapEpigenomics, TEMPUS, GTEx, Ensembl, NCBI, GenomOncology
- **Web development basics:** Django, CSS, Javascript, HTML, JQuery, and PHP

Language skills

- English - full professional proficiency, Polish - mother tongue, German - elementary proficiency

References

Prof. Dr. Ferdinand von Meyenn
Associate Professor
Department of Health Sciences and Technology at ETH Zurich
ferdinand.vonmeyenn@hest.ethz.ch

Dr. Altuna Akalin
Head of Bioinformatics & Omics Data Science Platform
The Max Delbrück Center for Molecular Medicine (MDC), the Berlin Institute for Medical Systems Biology (BIMSB)
altuna.akalin@mdc-berlin.de, +49 30 9406-4271

Dr. Olga Bogarytova
Data scientist, Translational Biomarker Lead, RU Oncology
Merck KGaA
Frankfurter Str. 250, 64293 Darmstadt, Germany
olga.bogatyrova@merckgroup.com

Publications

[1] R. R. C. Cuadrat, A. Kratzer, H. G. Arnal, R. A. C., K. Wreczycka, A. Blume, V. Ebenal, T. Mauno, B. Osberg, M. Moobed, J. Hartung, C. Seppelt, D. Meteva, A. Haghikia, D. Leistner, U. Landmesser and A. Akalin, "Cardiovascular disease biomarkers derived from circulating cell-free dna methylation", NAR Genomics and Bioinformatics, 2023. DOI: 10.1093/nargab/lqad061.

[2] K. Wreczycka, V. Franke, B. Uyar, R. Wurmus, S. Bulut, B. Tursun and A. Akalin, "HOT or not: examining the basis of high-occupancy target regions", Nucleic Acids Research, vol. 47, no. 11, pp. 5735–5745, May 2019. DOI: 10.1093/nar/gkz460.

[3] R. Wurmus, B. Uyar, B. Osberg, V. Franke, A. Gosdschan, K. Wreczycka, J. Ronen and A. Akalin, "PiGx: reproducible genomics analysis pipelines with GNU Guix", GigaScience, vol. 7, no. 12, Oct. 2018, giy123. DOI: 10.1093/gigascience/giy123.

[4] K. Wreczycka, A. Gosdschan, D. Yusuf, B. Grüning, Y. Assenov and A. Akalin, "Strategies for analyzing bisulfite sequencing data", Journal of Biotechnology, vol. 261, pp. 105–115, 2017, Bioinformatics Solutions for Big Data Analysis in Life Sciences presented by the German Network for Bioinformatics Infrastructure. DOI: https://doi.org/10.1016/j.jbiotec.2017.08.007.

Conferences & workshops

- attendance, PyData Berlin conference and workshops, 1-3 Sep. 2025, Berlin, Germany
- attendance, Federated Learning in Bioinformatics workshop by Swiss Institute of Bioinformatics, 29 Aug. 2025, Lugano, Switzerland
- speaker, "Assessing prognostic and diagnostic features of neuroblastoma methylome", DKTK - Young Academics Meeting, 6 Sep. 2019, Berlin Germany
- attendance, EMBL Conference: Cancer Genomics, 4–6 Nov. 2019, Heidelberg, Germany
- speaker, "Assesing prognostic and diagnostic features of high-dimentational (epi)genomics data for neuroblastoma", RECOMB, 5-8 May 2018, Washington, D.C., USA
- Europython conference and workshops, 10-16 July 2017, Rimini, Italy
- speaker, "HOT or not: Examining the basis of high-occupancy target regions", CRG-BIMSB MDC PhD Retreat, 15-18 Mai 2017, Girona, Spain
- EMBL-EBI Cancer Genomics Course, 20-23 June 2017, Hinxton, United Kingdom
- poster presentation "HOT or not: redefining the origin of high-occupancy target regions", EMBL conference From Functional Genomics to Systems Biology, 12-15 Nov. 2016, Heidelberg, Germany
- speaker, "Genomation: an R package to summarize, annotate and visualize genomic intervals", Bioinformatics Social Meetings Berlin, 16 Sep. 2016, Berlin Germany
- poster presentation "Genomation: an R package to summarize, annotate and visualize genomic intervals", the Max-Planck-Gesellschaft Otto Warburg Summer School, 6-11 Sep. 2015, Berlin, Germany
- Europython conference and workshops, 21-27 July 2014, Berlin, Germany

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Department of Health Sciences and Technology
Institute of Food, Nutrition and Health

ETH Zürich
Prof. Dr. Ferdinand von Meyenn
SLA C 93
Schorenstrasse 16
8603 Schwerzenbach, Switzerland

Telefon +41 44 655 74 05
Telefax +41 44 655 72 03
ferdinand.vonmeyenn@hest.ethz.ch
www.epigenetics.ethz.ch

Zurich, 2 September 2025

RE: Reference letter for Katarzyna Wreczycka

To whom it may concern

I am pleased to write this reference letter for Katarzyna Wreczycka, who worked in my group for Nutrition and Metabolic Epigenetics at ETH Zurich (epigenetics.ethz.ch) from April to June 2025, with a 40% appointment.
During her tenure, Kasia contributed to some of our current projects. She helped in the analysis of both RNA-seq and CUT&RUN datasets derived from adipose tissue, supporting our investigation into the epigenetic memory of obesity after weight loss. Kasia was also tasked to maintain and update the group's bioinformatics workflows on the ETH Euler cluster, including Nextflow Tower, SLURM, Singularity, and Spack.
In addition to her analytical work, Kasia provided technical support for our computational infrastructure and managed a successful migration of a Shiny application to ShinyProxy, hosted on the Functional Genomics Center Zürich servers (https://shiny-public.fgcz.uzh.ch/).
Her ability to quickly learn new technologies, work autonomously, was valued by her colleagues. Kasia consistently demonstrated professionalism, structurer working, and dedication, attributes which I am confident will allow her to thrive in any future role.
I strongly recommend Katarzyna Wreczycka for any opportunity that values computational and bioinformatics skills, and independent problem-solving. Please do not hesitate to contact me for further information.
Yours sincerely,

Prof. Dr. Ferdinand von Meyenn