Katarzyna Wręczycka, Ph.D.

✉ Zürich, Switzerland | katwre@gmail.com | +48 503 777 010 | +41 78 446 66 28

🌐 katwre.github.io | linkedin.com/in/katwre | github.com/katwre

👤 scholar.google.com/citations?user=katarzynawreczycka

Profile

Bioinformatics Data Scientist with 9+ years of applied research and healthcare industry in precision medicine and clinical genomics. Skilled in building and evaluating ML models - regression, classification, generative AI (VAEs), and transformers - and in developing workflows for biomarker discovery and translational research. Thrive in international, cross-functional, and creative environments.

Experience

04.2025 – 07.2025

Bioinformatician | Research Assistant II

ETH Zürich & Functional Genomics Center Zürich (FGCZ) of the University of Zurich (UZH)
Employed by the Prof. von Meyenn Lab (40% position), Prof. Bar-Nur Lab (40% position) and FGCZ & UZH (20% position)

- **Gene regulation analysis (ML)**, and data visualization to uncover insights driving:
  - muscle aging (RNA-seq), chimera organogenesis (snRNA-seq),
  - assembly of Y chromosome of *Mus minutoides* (WGS),
  - yo-yo effect (RNA-seq, CUT&TAG).
- **Set-up ETL Nextflow Tower for NGS (MLOps):**
  - Set-up reproducible end-to-end ETL Nextflow pipelines (RNA-seq, CUT&TAG) with Nextflow Tower on SLURM (ETH Zürich cluster) with containerization (Docker/Singularity) and Spack.
  - Maintained in-house ETL pipelines in Ruby on Rails, R, and Bash on FGCZ servers.
- **Interactive web-based visualizations:**
  - Interactive R Shiny apps via ShinyProxy for scRNA-seq data.
  - shiny-public.fgcz.uzh.ch (snRNA-seq: MHUO, Mouse epiAT memory, Human AT memory)
- **Cross-disciplinary collaboration** with wet-lab scientists to interpret models and experiments.

02.2023 – 01.2025

Bioinformatics Data Scientist

Remote from Berlin, Germany

Engaged in B2B contracts with bioinformatics, data science and software consultancy company Ardigen (CRO department) and with Selvita S.A.. Key highlights include:

- **Healthcare AI and oncology (ML/AI):**
  - Applied ML/AI for tumor antigen (TAAs) nomination and patient stratification for clinical trials.
  - Performed survival analysis, classification (PU learning) and variational autoencoders (VAE).
  - Integrated cancer gene expression, mutation and drug response data: scRNA-seq, TCGA, TEMPUS, GenomeOncology.
  - Stakeholder: Oncology Data Science Department at Merck.
- **Biomarker discovery in Alzheimer's disease (ML/AI):**
  - Target gene and enhancer discovery in glia-to-neuron reprogramming.
  - Applied ML/AI analysis, incl. a transformer for enhancer detection called Enformer by Avsec et al., and LASSO for DNA motifs detection by Machlab et al..
  - Integrated multi-omics datasets: RNA-seq, ATAC-seq, Bisulfite-seq, and H3K27ac ChIP-seq.
  - Stakeholder: start-up Stardustries.
- **Interactive web-based visualization:**
  - Interactive visualisations of gene isoforms, and integration of in-house and publicly available sequencing (epi)genomic data.
  - Implemented in the IGV app browser and CLI in Javascript and Python.
  - Stakeholder: Novo Nordisk.
- **Cross-disciplinary collaboration:**
  - With software engineers, data scientists, UX/UI designers, and wet-lab scientists to translate user needs into robust, scalable technical solutions.
  - Set priorities and made context-aware decisions.
- **Stakeholder communication & presentation:**
  - Scrum project management
  - Project leading, planning and mentoring, presentations of results

- **MLOps and quality:**
  - Containerized workflows (Docker) on Kubernetes and AWS
  - Version control (git, gitlab/github)
  - Automated tests and documentation (Diátaxis)

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 (MDC-BIMSB).
- Led and contributed to end-to-end research projects applying advanced statistical and ML techniques, resulting in 4 **publications in top-tier journals (350+ citations)**.
- First co-author of a peer-reviewed **review on statistical methods** for Bisulfite-seq data analysis.
- **Genomic artifact detection (ML):**
  - Led ChIP-seq bias reduction project, linking spurious ChIP-seq signals to RNA:DNA hybrids, and open-chromatin and hypomethylated regions.
  - Performed ML analysis incl. elastic net, PCA.
  - Integrated large-scale omics data: ENCODE, RoadmapEpigenomics, publicly available ChIP-seq, RNA-seq, ATAC-seq, Bisulfite-seq, DRIP-seq.
  - Collaborated with Prof. Dr. Tursun's wet-lab to improve ChIP-seq protocol.
- **Snakemake ETL pipeline for Bisulfite-seq (MLOps):**
  - ETL pipeline implemented in snakemake, Python and R/Bioc.
  - Peer-reviewed publication as part of the PiGX: Pipelines in Genomics project.
  - Applied it on SLURM (Berlin Institute of Health at Charité), Grid Engine (MDC BIMSB).
- **cfDNA biomarker discovery for acute coronary syndrome (ML):**
  - Applied ML incl. logistic regression with overdisperssion correction.
  - Analyzed liquid biopsies (blood cell-free DNA) from patient-derived Bisulfite-seq data.
  - In close collaboration with clinician Prof. Dr. med. Ulf Landmesser at the Charité hospital.

03 – 09.2015 | Visiting Predoctoral Researcher

Bioinformatics and Omics Data Science Platform at Max Delbrück Center (MDC-BIMSB).
- Awarded a €7,000 scholarship by the Institute of Computer Science, Polish Academy of Sciences and co-financed by the European Union. Supervised by Dr. Altuna Akalin at MDC-BIMSB.
- Developed Bioconductor R package genomation for visualization and annotation of genomic data.

Education

Ph.D. Computational Molecular Biology, Humboldt University of Berlin, Berlin, Germany (2015–2021)
M.Sc. Computer Science and Bioinformatics, University of Warsaw, Poland (2012–2014)
B.Sc. Computer Science and Bioinformatics, University of Warsaw, Poland (2010–2012)

Skills

Technical skills:
- **Programming & Tools:** Python - NumPy, pandas, scikit-learn, seaborn, PyTorch, Matplotlib; R - CRAN, Bioconductor (incl. caret, tidyverse); unit tests: testthat/unittest/pytest; Javascript; bash
- **ML & AI:** statistical tests - incl. t-tests, Wilcoxon tests; regression - linear, logistic, Cox/survival analysis, elastic net/lasso/ridge; classification - random forest, XGBoost, SVM, LDA; clustering - K-means, EM, probabilistic models - Hidden Markov Models (HMMs), linear Gaussian state-space models; dimensionality reduction & factorization - PCA, t-SNE, MOFA, NMF; sampling/optimization - replica exchange monte carlo; deep learning - variational autoencoders (VAEs), CNNs, transformers, retrieval-augmented generation (RAG), federated learning
- **MLOps:** workflow languages - Nextflow, Snakemake; SLURM, Grid Engine, Docker/Singularity, Kubernetes, AWS, DigitalOcean, workflow engines - Galaxy
- **Databases:** MySQL, SQLite, PostgreSQL
- **Version Control and Software Management:** Linux/Unix, git, svn, conda, GNU Guix
- **Bioinformatic Tools:** samtools, BEDtools, GATK, IGV, Bowtie2, BWA, Bismark, BLAST
- **Omics Databases:** TCGA, Roadmap Epigenomics, TEMPUS, GTEx, Ensembl, NCBI, GenomOncology

Soft Skills:
- Cross-functional collaboration and independent project leadership across cross-disciplinary teams, clear communication of complex concepts to technical and non-technical audiences, adaptability and agility in evolving scientific and technical environments, analytical thinking and problem solving, project planning and mentoring, scientific writing and international presentation