Xenofon **Giannoulis**

(+49) 15772182145 | xgiannoulis@proton.me | xenophong.github.io/web/

Work Experience

Statistical Geneticist — *Pheiron Gmbh,DE*

Platform Development & Scientific Discovery — *July 2025 – now*

- Designed precomputation pipeline supporting three analysis modes (single-variant GWAS, PGS-GWAS, burden tests).
- Engineered streaming ETL for genetics association results at hundreds-of-millions to billions of rows.
- Implemented out-of-core processing with low-memory execution and hour-scale end-to-end runtimes.
- Built post-ETL QC/validation framework with modular checks, resumable runs, and production monitoring.
- Added biological validation against external references to ensure decision-grade result quality.
- Developed external GWAS catalogue: harmonized public biobank summary statistics into a unified resource.
- Designed a query interface for fast gene-, variant-, and phenotype-level association lookups.
- Authored end-to-end documentation (architecture, workflows, benchmarks, troubleshooting) to enable team communication.

Postdoctoral Researcher — *Helmholtz Munich,DE*

Institute of Computational Neurobiology — *January 2025 – June 2025*

- Used knowledge-graph and network medicine methods to prioritize potential therapeutic targets in Alzheimer's disease.
- Integrated multi-omics data with graph-based and ML models to characterize disease pathways and candidate mechanisms.
- Built interactive exploration tools (R Shiny + Neo4j) for network-based hypothesis generation; AD Atlas.
- Developed reproducible pipelines to harmonize and analyze high-dimensional omics datasets from NIH-funded consortia.

Scientific Researcher — *Helmholtz Pioneer Campus,DE*

Institute of Translational Genetics — *June 2020 – December 2024*

- Built a multi-tissue atlas of mitonuclear gene regulation in GTEx v8 (684 donors, 48 tissues) using WGS + RNA-seq.
- Mapped 111 independent mtDNA cis-eQTL associations across mtDNA genes, including 86 novel signals.
- Implemented LMM eQTL workflows with rigorous QC, mtDNA haplogroup inference, and tissue-specific covariate control.
- Extended discovery to trans effects and causal follow-ups via colocalization, mediation testing, and Mendelian randomization.
- Identified genome-wide trans mito-nuclear regulatory links, including CNS-enriched cell-type interaction effects.
- Integrated eQTL results into network-based gene prioritization using Speos (GNN/MLP; TAG graph convolution).
- Built an RNA-seq pipeline to quantify mtDNA heteroplasmy and mtRNA modifications (12,577 samples, 49 tissues).
- Built coverage-aware association tests (LM + beta-binomial), finding 109 age and 784 expression associations.
- xCell interaction + mediation analyses found 9 age–heteroplasmy–mtDNA expression links.
- Presented results in interdisciplinary settings and contributed to manuscripts, figures, and reproducible analysis code.

Research Fellow — *LMU, Munich,DE*

Data Science for Social Good | Munich Center for Machine Learning — *August 2023 – September 2023*

- Built a geospatial planning tool for fire brigades to model hose reachability under real-world obstacles (stairs, rails, buildings).
- Generated reachability polygons from a selected water extraction point for a given hose length.
- Computed reverse-planning polygons for where a water point must lie to cover a selected building/site.
- Developed a full-stack web app for interactive mapping, real-time routing, and resource planning in the field.
- Deployed on AWS using a containerized architecture for reliable, scalable use by partners.
- Delivered to Federal Office for Disaster Control & Munich Fire Brigade; presented to the German federal parliament.

Instructor, Human Genetics of Complex Traits — *TUM Munich,DE*

Faculty of Medicine (ME1660) — *October 2022 – March 2023*

- Tutored and supported students in computational genomics through hands-on labs and office hours.
- Improved course materials and workshop flow for practical analysis in JupyterLab.
- Taught UNIX, QC, association testing, meta-analysis, and polygenic risk score methods.

Genomic Data Science Intern — *Helmholtz Munich,DE*

Institute of Translational Genomics — *October 2019 – April 2020*

- Analyzed GWAS data from multi-center joint replacement cohorts to identify genetic risk factors.
- Harmonized Illumina exome-array datasets across multiple centers and processing pipelines.
- Built QC workflows to detect batch effects, technical bias, and cross-center inconsistencies.
- Ran genotype imputation via the Michigan Imputation Server using the HRC reference panel.

Education

Technical University of Munich (TUM) - Experimental Medicine *Munich, DE*
Ph.D. / Dr.rer.nat. in School of Medicine (MGC) *Dec 2024*

Thessaly University - Computer Science and Biomedical Informatics *Lamia, GR*
MSc. in Informatics and Computational Biomedicine *April 2020*

Piraeus University - Department of Statistics and Insurance Science *Athens, GR*
Ptychion (4-year BSc.) in Statistics and Actuarial Science *February 2018*

Technical Experience

Platform/API
Queryable data catalogs and analytical storage: Apache Iceberg, Parquet, partitioning, schema evolution.
Analytical SQL and in-process analytics: SQL, DuckDB.
Interactive apps with RShiny + Neo4j for graph exploration and hypothesis generation.
REST query APIs for gene/variant/phenotype lookups and catalog search/querying.
REST services with Flask for geospatial routing and mapping workflows.
Data validation and monitoring for production pipelines (QC modules, checks, runbooks).

Engineering
Python, R, Bash, Linux for large-scale data processing and workflow automation.
Reproducible workflows, containerization, and cloud execution.
Knowledge graph / semantic web tooling: Protégé, RDF, SPARQL.
Certified Oracle MySQL Developer (1z0-882).
Version control with GitHub; documentation with LaTeX and RMarkdown.
AI-assisted development with Cursor, Copilot, and LLM-based tooling.

Genetics
GWAS, PheWAS, eQTL/ct-eQTL mapping, WGS, RNA-seq, variant calling, Illumina array processing.
Tools: Regenie, PLINK, LDAK, Bioconductor, MendelianRandomization, Coloc.
PRS construction, burden testing, fine-mapping, conditional analysis, annotation integration.
Multi-omics integration (genomic, transcriptomic, proteomic, regulatory).
Biobanks: GTEx, PsychEncode, UK Biobank, All of Us, FinnGen, MVP.

Statistics & ML
Linear mixed models (GRM/variance components), GLMs; heteroskedasticity-aware modeling.
Empirical inference: genotype permutations, simulations, likelihood-ratio tests; model evaluation.
Multiple testing control: BH/q-value FDR and Bonferroni at tissue/study-wide levels.
Causal follow-up: Bayesian colocalization, conditional/mediation analysis, Mendelian randomization.
Feature engineering, PCA, ensemble learning for gene prioritization.

Cloud
Cloud-based genomic workflows with DNAnexus, AWS/S3, and Docker.
Parallel execution, resumable runs, and cost-aware scaling for biobank-scale analyses.

Scientific Publications

Interplay between mitochondrial and nuclear DNA in gene expression regulation. Biorxiv
Xenofon Giannoulis, Simon Wengert, Florin Ratajczak, Matthias Heinig, Na Cai.
Under review at Nature Communications.

Tissue-Specific mtDNA Heteroplasmy Linked to Aging and Gene Expression. Biorxiv
Simon Wengert, **Xenofon Giannoulis**, Peter Kreitmaier, Holger Prokisch, Paolo Casale, Matthias Heinig, Na Cai.
Under review at Nature Communications.