

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

- Developed predictive genetic models scoring 36M+ target-indication pairs to de-risk early-stage drug development.
- Automated MR and Coloc pipelines to validate mechanism chains and predict clinical trial success.
- Architected a unified GWAS catalog using out-of-core DuckDB/Polars ETL pipeline to harmonize FinnGen, PAN-UKB, MVP.
- Designed scalable precomputation pipelines (GWAS, PGS, burden tests) with fast query interfaces for rapid association lookups.
- Built an automated post-ETL QC and validation framework with strict statistical checks to ensure decision-grade data quality.
- 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

- Prioritized Alzheimer's therapeutic targets by integrating multi-omics evidence data into knowledge graphs and ML models.
- Expanded the AD Atlas (R Shiny, Neo4j), an interactive network platform enabling rapid disease pathway exploration
- Engineered graph abstraction layers to translate heterogeneous data into interpretable network topologies.
- Deployed reproducible harmonization pipelines processing high-dimensional omics to extract robust molecular signatures.

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.
- 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).
- Presented consortia findings and authored two manuscripts currently under revision at Nature Communications.

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).
- Computed reverse-planning polygons for where a water point must lie to cover a selected building/site.
- Developed a full-stack AWS web app for interactive mapping, real-time routing, and resource planning in the field.
- Delivered system to the 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

- Instructed hands-on computational genomics and mentored students in complex trait analysis.
- Developed and refined interactive JupyterLab course materials to optimize practical workshop flow.
- Taught UNIX, GWAS 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 multi-center GWAS data from joint replacement cohorts to identify novel genetic risk factors.
- Harmonized Illumina exome-array datasets, engineering QC workflows to resolve cross-center batch effects.
- Built QC workflows to detect batch effects, technical bias, and cross-center inconsistencies.
- Executed large-scale genotype imputation via the Michigan Imputation Server using the HRC reference panel.

Education

Technical University of Munich (TUM) - Experimental Medicine PH.D. / DR.RER.NAT. IN SCHOOL OF MEDICINE (MGC)	<i>Munich, DE</i> <i>Dec 2024</i>
Thessaly University - Computer Science and Biomedical Informatics MSC. IN INFORMATICS AND COMPUTATIONAL BIOMEDICINE	<i>Lamia, GR</i> <i>April 2020</i>
Piraeus University - Department of Statistics and Insurance Science PTYCHION (4-YEAR BSC.) IN STATISTICS AND ACTUARIAL SCIENCE	<i>Athens, GR</i> <i>February 2018</i>

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.
Logic	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 DNAxexus , 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.