

Mohamed Abuelanin

✉ mabuelanin@gmail.com  mr-eyes.com  [mabuelanin](https://www.linkedin.com/in/mabuelanin)  [mr-eyes](https://github.com/mr-eyes)

Professional Profile

Computational biologist and software engineer developing scalable methods for genome assembly, structural variant discovery, and analysis of large public sequencing datasets. My current work focuses on improving detection and interpretation of autism-associated rare variants in short-read data from approximately 2,600 quartet families, each including an autistic proband, an unaffected sibling, and both biological parents, especially within duplicated and repetitive regions better resolved by CHM13/T2T references. I also build open-source bioinformatics software and reusable infrastructure for sequence quality control, k-mer analysis, comparative genomics, and public-data reuse.

Research Focus

- Developing approaches to detect autism-relevant variants from short-read Simons cohort data in segmental duplications and other difficult genomic regions, with emphasis on CHM13-resolved sequence.
- Applying long-read chromatin-conformation methods, including CiFi, to assemble and interpret structurally complex loci such as *AVPR1A*, a gene linked to social behavior in voles and relevant to human neurodevelopmental biology.
- Building scalable software and workflows that make large public sequencing resources easier to search, quality control, compare, and reuse.

Education

University of California, Davis

Ph.D. in Computer Science | Thesis: Petabyte-scale sequence characterization

Davis, CA

Sep. 2020 – Jun. 2024

Nile University

M.Sc. in Communication and Information Technology | Thesis: kmer-based sequence partitioning

Giza, Egypt

Mar. 2017 – Aug. 2020

Higher Technological Institute

B.Sc. in Electrical/Biomedical Engineering | Thesis: ML-Based Taxonomic Classification

Cairo, Egypt

Sep. 2011 – Jan. 2017

Research and Professional Experience

University of California, Davis – Genome Center & MIND Institute

Davis, CA

Postdoctoral Researcher, Dept. of Biochemistry and Molecular Medicine

Jul. 2024 – Present

- Develops computational approaches for identifying autism-associated rare variants from short-read sequencing data in segmental duplications and other complex genomic regions.
- Develops methods for detecting and characterizing rare human contigs in underrepresented Khoe-Sān genomes.
- Uses CiFi-based comparative genomics to resolve complex vole loci such as *AVPR1A*, a social-behavior gene with human neurodevelopmental relevance.

University of California, Davis – Titus Brown's Lab, School of Veterinary Medicine

Davis, CA

Graduate Student Researcher – Scalable framework for massive-scale sequence QC

Sep. 2020 – Jun. 2024

- Developed scalable methods for massive-scale sequence characterization and quality control across public sequencing datasets.
- Led grant-supported research improving discoverability, reuse, and meta-analysis of Common Fund and canine sequencing data.
- Created and maintained open-source tools for lightweight sequence assessment, k-mer analysis, and data reuse workflows.

Additional Teaching and Engineering Experience

- **Marine Biological Laboratory (MBL)**, Woods Hole, MA – Teaching Assistant, Gene Regulatory Networks Workshop, Oct. 2023.
- **University of California, Davis**, Davis, CA – Instructor, Data Analysis Collaboratory Workshop, Jun. 2023.

Research Funding and Applicant-Led Projects

- NIH Research Grant #1R03OD034502-01 – Primary Researcher / Project Lead** Sep. 2022 – Sep. 2025
"Data-driven search of Common Fund datasets for better discoverability and novel meta-analysis" \$312,000
- Applicant-led research project. I developed, led, implemented, and authored the core computational work; the award was administered through the institutional PI.
- Morris Animal Foundation Grant D22CLP-505 – Primary Researcher / Project Lead** Jun. 2022 – Oct. 2025
"Maximizing the scientific utilization of Golden Retriever Lifetime Study's datasets" \$304,000
- Applicant-led research project. I developed, led, implemented, and authored the core computational work; the award was administered through the institutional PI.
- Ferrini Research Award – UC Davis School of Veterinary Medicine** 2023
Large-scale sequence-based quality control of canine SRA sequencing data

Journal Articles

- [1] **M. Abuelanin**, G. Kaya, J.A. Lake, C. Lambert, M.V. Wu, K. Krashenninnikova, J. Wood, N.G. Solomon, Z. Donaldson, K.L. Bales, K. Howe, J. Korlach, D. Manoli, J. Tollkuhn, M.Y. Dennis. "Single-library chromosome-scale diploid assemblies of vole genomes resolve a species-specific duplication implicated in pair bonding." *Cell Genomics*, accepted, 2026.
- [2] S.P. McGinty, G. Kaya, S.B. Sim, A. Makunin, R.L. Corpuz, M.A. Quail, **M. Abuelanin**, et al. "CiFi: accurate long-read chromosome conformation capture with low-input requirements." *Nature Communications*, 2025.
- [3] M. Shokrof, **M. Abuelanin**, C.T. Brown, T.A. Mansour. "The Great Genotyper: a graph-based method for population genotyping of small and structural variants." *GigaScience* 14, 2025.
- [4] L. Irber, N.T. Pierce-Ward, **M. Abuelanin**, H. Alexander, A. Anant, K. Barve, et al. "sourmash v4: A multitool to quickly search, compare, and analyze genomic and metagenomic data sets." *Journal of Open Source Software* 9(98), 6830, 2024.

Dissertation and Conference Papers

- [1] **M. Abuelanin**. "Scalable Computational Frameworks for Next-Generation Sequencing Analysis and Gene Set Integration." *Ph.D. Dissertation*, University of California, Davis, 2024.
- [2] **M. Abuelanin**, M. Fares, M. El-Hadidi. "Insilico Codon Bias Correction for Transgenic Biological Protein Sequences for Vaccine Production." *Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, 2020.

Selected Presentations

- **T2T F2F Meeting**, Santa Cruz, CA – Invited talk, Sept. 3–4, 2026.
- **Biology of Genomes 2026 (CSHL)**, CSHL, NY – Talk: "Chromosome-scale vole assemblies resolve a prairie vole-specific AVPR1A duplication," 2026.
- **Halloween Symposium, UC Davis Genome Center**, Davis, CA – Poster: "CiFi: A Low-Input Long-Read Method to Unmask Hidden Chromatin," Oct. 2025.
- **Plant and Animal Genome Conference (PAG 2025)**, San Diego, CA – Talk: "Revolutionizing Canine SRA Dataset Analysis for Enhanced Comparative and Reusability Insights," Jan. 2025.
- **Biology of Genomes 2025 (CSHL)**, CSHL, NY – Poster: "Structural Variant Discovery and Characterization from *de novo* Assembly of Khoe-San Genomes," 2025.
- **Plant and Animal Genome Conference (PAG 2022)**, San Diego, CA – Poster: "Fast Clustering of Hundreds of Microbiome Datasets," Jan. 2022.

Selected Open-Source Software and Public Infrastructure

- **snipe** – SRA-scale lightweight sequence QC and assessment.
- **cifiasm** – CiFi-based workflow for chromosome-scale diploid genome assembly.
- **Sequencing ROI** – Library-complexity prediction using alpha-diversity and rarefaction curves.
- **kProcessor** – Efficient sequence k-mer indexing and analysis.
- **DBRetina** – Gene overlap super-network analysis.
- **kSpider** – K-mer-based ultrafast pairwise DNA clustering.
- **sourmash** – Metagenome analysis and genome comparison using k-mers.
- **denovo-rnaseq** – De novo assembly, differential expression, and annotation Snakemake workflow.

- **guidedPartitioner** – Reference-guided RNA-seq read partitioning for accurate assembly.
- **kDecontaminer** – RNA-seq decontamination for cleaner de novo transcriptome assembly.
- **Codon Optimization Tool** – cDNA optimization using codon usage bias for DNA vaccine design.

Peer Review, Teaching, and Service

- **Nature Biotechnology** – Peer-reviewed journal 1 manuscript
- **Review Commons** – Journal-independent peer review platform 1 manuscript
- **RECOMB** – International Conference on Research in Computational Molecular Biology (USA) 1 paper
- **NILES** – Novel Intelligent and Leading Emerging Sciences Conference (Egypt) 2 papers
- Teaching and workshop instruction in gene regulatory networks, data analysis, machine learning, big data, advanced next-generation sequencing, programming, and embedded systems.

Technical Skills

Software Engineering: C++, Rust, Python. Probabilistic data structures, graph algorithms, scalable sequence analysis, interactive data visualization.

Genomics & Bioinformatics: Chromosome-scale de novo assembly, structural variant discovery in complex repetitive regions, comparative and population genomics, transcriptomics, epigenomics.

Workflows & Infrastructure: Snakemake, Nextflow, WDL. HPC, cloud, containerization.