

Abdullah (Arby) Abood

571-274-4329 | arbyaboodphd@gmail.com | www.arby-abood.com | <https://github.com/aa9gj> | [LinkedIn](#) | US Citizen

OBJECTIVE

Senior data scientist/computational biologist building production-facing analytics and LLM-enabled systems in regulated environments. Proven track record delivering end-to-end pipelines (Nextflow/SLURM), predictive modeling, and knowledge graph + RAG solutions. Lead cross-functional collaborations across Regulatory, Clinical, Product Development, and R&D to reduce cycle time and standardize evidence-backed decision making.

WORK EXPERIENCE

COLGATE-PALMOLIVE | SENIOR SCIENTIST

Research Triangle Park, NC | Jun 2023 – Present

- Architected and built a retrieval-augmented LLM pipeline (Python) that generates 80% complete nutritional-claims substantiation by grounding generation in AAFCO regulations, Global Nutrient Standards, and 12K historical claims from Veeva, with computed Dry Matter Basis/digestibility metrics and audit-ready references. This project reduced review cycle time 50% and eliminated manual copy/paste across NA, EU, and global markets by standardizing claim templates and scaling the system cross-functionally.
- Built disease prediction (XGBoost, Random Forest) and biological age estimation models (Lasso-penalized Cox PH, Gompertz) on CBC/chemistry data (n~1,000 patients), achieving AUC up to 0.92 and enabling risk stratification for downstream studies.
- Designed and deployed a GraphRAG pipeline (Neo4j-backed knowledge graph) that ingests final study reports locally on HPC using open-source models for ingestion, embeddings, and NER including Nemotron-Parse, BGE, and openbioNER coupled with hybrid retrieval using closed-models via API to enable automated summarization for scientific query answering and decision support.
- Developed reproducible Nextflow pipelines for bulk RNA-seq, scRNA-seq, and WGBS (short- and long-read), improving QC standardization, portability, and run-to-run reproducibility on HPC.

UNIVERSITY OF VIRGINIA | COMPUTATIONAL BIOLOGIST

Charlottesville, VA | Jul 2018 – May 2023

- Identified causal isoforms in bone mineral density GWAS by integrating sQTL data and Bayesian analysis, refined with long-read RNAseq and proteomics, leading to the implication of genes in osteoporosis through *in-vitro* validation

Relevant work:

- **Abood A et al. Long-read proteogenomics to connect disease-associated sQTLs to the protein isoform effectors of disease.**

American Journal of Human Genetics, 2024

- Castaldi P, **Abood A et al. Bridging the splicing gap in human genetics with long-read RNA sequencing: finding the protein isoform drivers of disease.** Human Molecular Genetics, 2022

- Uncovered long non-coding RNAs impacting BMD GWAS loci through eQTL, Bayesian colocalization, TWAS, and allelic imbalance, revealing key non-coding elements of the bone transcriptome and their roles in osteoporosis.

Relevant work:

- **Abood A et al. Identification of known and novel long non-coding RNAs potentially responsible for the effects of BMD GWAS loci.** J Bone Miner Res. 2022.

- **Abood A, Farber CR. Using "-omics" Data to Inform Genome-wide Association Studies (GWASs) in the Osteoporosis Field.** Current Osteoporosis Reports. 2021.

EDUCATION

Ph.D. COMPUTATIONAL BIOLOGY

Charlottesville, VA | May 2023

UNIVERSITY OF VIRGINIA

Teaching Assistant (TA): Core Course in Integrative Biosciences (BIMS 6000), Recent Advances in Public Health Genomics (PHS 5705)

SPECIALIZED TRAINING

BIOMEDICAL DATA SCIENCE TRAINING GRANT

Awarded NIH T32 Biomedical Data Science Training Grant supporting advanced research in multi-omics integration, and high-dimensional clinical data analysis to advance understanding of disease mechanisms.