



PROFESSIONAL SUMMARY

PhD candidate in Bioinformatics with hands-on expertise in multi-omics integration, transcriptomics, and methylomics for biomarker discovery in complex disease. Skilled in differential expression analysis, co-expression network analysis (WGCNA), cross-platform batch correction, and statistical genomics methods including Mendelian randomization and eQTL/mQTL integration. Proficient in R and Python for reproducible, end-to-end analysis pipelines, with additional experience in machine learning, web application development (R Shiny), and structured scientific communication. Motivated to apply integrative computational biology approaches to genomics and precision medicine research.

TECHNICAL SKILLS

Programming: R, Python, SQL, Linux/command-line, R Shiny

Bioinformatics: RNA-seq analysis, ChIP-seq analysis, microarray analysis, molecular docking (AutoDock Vina), CRISPR-Cas9 guide design, sequence alignment, probe-to-gene annotation

Transcriptomics & Methylomics: DESeq2, edgeR, Limma, functional enrichment and pathway analysis, cross-platform normalization

Multi-Omics Integration: DIABLO, mixOmics, WGCNA, horizontal/vertical/gene-level integration, sva/ComBat batch correction

Statistical Genomics: Mendelian randomization, eQTL integration, mQTL analysis, PCA and quality control, goodness-of-fit testing, contingency table analysis, t-tests, ANOVA, Pearson/Spearman correlation, linear regression diagnostics

Machine Learning / AI: Logistic regression, SVM, random forest, XGBoost, gradient boosting, KNN, deep neural networks, CNNs, GANs, Boruta/LASSO feature selection, hyperparameter tuning, cross-validation, SMOTE

Data Visualization: ggplot2, pheatmap, heatmaps, volcano plots, PCA plots, boxplots, violin plots, Matplotlib, Seaborn, Plotly, Tableau, Power BI

Databases & Data Acquisition: SQL/relational databases, NoSQL, NCBI GEO, PubMed, web scraping (BeautifulSoup, Pandas)

Workflow & Reproducibility: Jupyter notebooks, Git/GitHub, reproducible R/Python pipelines

EDUCATION

PhD in Bioinformatics | University of Lethbridge, Lethbridge, AB, Canada

Jan 2024 - Present

Research focus: multi-omics integration for biomarker discovery in rheumatoid arthritis.

Master of Science, Applied Genetics and Bioinformatics | Bangalore University, India

Completed Apr 2022

Major: Bioinformatics, Biostatistics, Applied Genetics. Minor: Biotechnology. GPA: 8.57/10, Grade A+.

Thesis: "Investigation of Anti-Arthritis Compounds from *Cinnamomum zeylanicum* through In-Silico Methods," Grade 9.5/10.

Honor: Gold Medal and 1st Rank Award, Vice Chancellor of Bangalore University (2022).

PUBLICATIONS

- Ghasemi Gojani, E., Rai, S., Norouzkhani, F., Shujat, S., Wang, B., Li, D., Kovalchuk, O., Kovalchuk, I. (2024). Targeting β -cell plasticity: A promising approach for diabetes treatment. *Current Issues in Molecular Biology*, 46(7), 7621-7667. DOI: 10.3390/cimb46070453; PubMed ID: 39057094.
- Rai, S. Feature selection and dimensionality reduction in multi-omics. Manuscript submitted for review (2026).
- Rai, S. (2022). Investigation of anti-arthritis compounds from *Cinnamomum zeylanicum* through in-silico methods. Master's Thesis, Department of Applied Genetics, Bangalore University. Awarded Gold Medal and university first rank; thesis grade 9.5/10.
- Multi-omics biomarker discovery in rheumatoid arthritis. Manuscript in preparation (2026).
- A web application for multi-omics rheumatoid arthritis data visualization. Manuscript in preparation (2026).

RESEARCH EXPERIENCE

PhD Researcher | University of Lethbridge, Lethbridge, AB

Jan 2024 - Present

- Conducted multi-omics data analysis for rheumatoid arthritis biomarker discovery, applying horizontal, vertical, and gene-level integration strategies using DIABLO and mixOmics frameworks.
- Performed differential gene expression analysis with DESeq2 and Limma across RNA-seq and microarray datasets.
- Implemented WGCNA to identify co-expression modules associated with phenotypic traits in transcriptomic data.
- Applied sva/ComBat batch correction to harmonize cross-platform and cross-study transcriptomic and methylomic datasets prior to integration.
- Investigated eQTL and mQTL relationships to link genetic and epigenetic variation to gene expression changes, and applied Mendelian randomization to evaluate causal relationships between molecular exposures and disease outcomes.
- Designed and applied hypothesis testing, correlation (Pearson/Spearman), regression, and classification analyses, including t-tests, ANOVA, chi-square/Kolmogorov-Smirnov tests, contingency table analysis, and linear regression with assumption validation, to characterize complex biological datasets.
- Implemented supervised and unsupervised machine learning models (logistic regression, SVM, random forest, XGBoost, deep neural networks) with Boruta and LASSO feature selection, and generated synthetic biomedical images using GANs to augment training datasets.
- Developed an R Shiny web application for interactive exploration of multi-omics rheumatoid arthritis data, with datasets acquired via NCBI GEO and PubMed scraping (BeautifulSoup, Pandas).

Master's Research | Bangalore University, India

Completed Apr 2022

- Applied molecular docking (AutoDock Vina) and bioinformatics pipelines to screen 3,000 phytochemicals from *Cinnamomum zeylanicum*, identifying candidates with stronger predicted binding affinity than standard rheumatoid arthritis drugs.
- Investigated anti-arthritis compound activity through in-silico virtual screening.

Bioinformatics Intern | ArrayGene, Bangalore, India

Jan 2022 - Aug 2022

- Applied differential gene expression analysis using DESeq2 and edgeR to identify key genes across experimental conditions.
- Performed preprocessing and quality control of high-throughput sequencing data.
- Conducted functional enrichment and pathway analysis to interpret biological significance of identified genes.
- Developed heatmaps, volcano plots, and PCA plots to summarize results.
- Applied CRISPR-Cas9 guide design and evaluation, assessing on-target/off-target scores using CRISPR design tools.
- Performed analyses on Linux-based systems using R.

PROFESSIONAL EXPERIENCE

Bioinformatics Expert (Contract) | Mercor, Remote

Mar 2026 - Jun 2026

- Designed and reviewed bioinformatics and genomics benchmark tasks for AI model evaluation, defining rubrics for transcriptomics, genomics, and epigenomics workflows.
- Developed reproducible Jupyter notebook workflows to support benchmark task creation.
- Assessed model-generated analyses against rubrics, providing structured written feedback on analytical rigor and best practices.

Mentor | University of Lethbridge, Dept. of Biological Sciences

Jan 2025 - Present

- Guided and monitored independent students in bioinformatics, ML, and AI research problem design under a MITACS-supported program.
- Documented mentee progress and provided structured feedback to support project completion.

Data Scientist | SirpiData Science, Bangalore, India

Jun 2023 - Nov 2023

- Applied machine learning and spatial analytics to detect urban growth and environmental risk using satellite imagery processed via Google Earth Engine, reducing data preparation time by 30%.
- Developed geospatial data pipelines integrating multiple datasets into environmental and urban analysis reports; built interactive dashboards for stakeholders.
- Contributed to the IUDX (India Urban Data Exchange) project, delivering water-quality and urban-infrastructure insights.

Data Science Intern | DataMites, Bangalore, India

Aug 2022 - Jan 2023

- Directed predictive classification and regression models (logistic regression, random forest, XGBoost, gradient boosting) on Kaggle datasets, improving diagnostic accuracy by 12% and reducing prediction error by 15%; addressed class imbalance using SMOTE.
- Designed a CNN-based image classification model achieving 88-90% accuracy, evaluated via cross-validation and standard classification/regression metrics (F1, ROC-AUC, RMSE, R²).
- Led a team through three end-to-end projects spanning classification, regression, and image processing, from data collection through model deployment.

Medical Reviewer/Coder | Omega Healthcare Management Services, Bangalore, India

Apr 2022 - Sep 2022

- Reviewed and analyzed NSCLC patient records for completeness and accuracy, processing ~50 records weekly, and assigned ICD-10 codes per medical coding standards for research and clinical reporting.
- Maintained strict HIPAA compliance while collaborating with case managers to resolve ambiguous clinical documentation.

PROJECTS

Multi-Omics Web Application for Rheumatoid Arthritis Biomarker Discovery: Interactive tool for exploring integrated transcriptomic and methylomic RA data. Methods: R Shiny, DIABLO, mixOmics, WGCNA, sva/ComBat. Outcome: in-development web application for multi-omics data exploration; associated manuscript in preparation.

GEO/PubMed Data Acquisition Pipeline: Automated curation of transcriptomic datasets and literature references via web scraping (NCBI GEO, PubMed) using BeautifulSoup and Pandas, streamlining dataset acquisition for downstream analyses.

In-Silico Phytochemical Screening for Rheumatoid Arthritis (Master's Thesis): Molecular docking (AutoDock Vina) and virtual screening of 3,000 phytochemicals from *Cinnamomum zeylanicum*, identifying candidates with binding affinity exceeding standard RA drugs.

Synthetic Biomedical Image Generation: Generated synthetic biomedical images using Generative Adversarial Networks (GANs) on Kaggle datasets to augment model training data.

Lung Cancer Diagnostic Classification Model: Predicted lung cancer diagnosis from tabular data using logistic regression, random forest, XGBoost, KNN, and SVM with SMOTE for class imbalance, achieving 90-92% test accuracy.

RESEARCH INTERESTS

Multi-omics integration, transcriptomics, epigenomics/methylomics, biomarker discovery, systems biology, AI/ML for biomedical research, precision medicine.

ACCOMPLISHMENTS

- Gold Medal & 1st Rank Award, Vice Chancellor of Bangalore University, Master's degree (2022)
- University Third Rank Holder, Bachelor's degree (2019)
- Certified Data Scientist (2023)
- MITACS Scholarship Recipient (2024 - Present)