

# MUALLA GELDEGUL

✉ mualla.unsall@gmail.com

☎ (+90) 541 652 13 41

📍 Pendik, Istanbul

🌐 mualla-unsall

📅 23/07/1997

🔄 unsalmualla

## About

---

I am a Master's student in Bioinformatics and Systems Biology. I aim to apply the scientific discipline I have developed during the analysis of complex data and my theoretical knowledge in data science/machine learning to solve real-world problems in the technology sector. I am eager to transform my analytical thinking skills into professional competence by participating in large-scale data projects and to continuously learn.

## Education

---

2024 – Present  
Istanbul, Türkiye

**MSc, Bioinformatics and System Biology, Istanbul University**  
Following a curriculum focused on Data Science and Machine Learning, currently researching the integration of machine learning techniques with high-dimensional scRNA-seq data analysis for the master's thesis.

2015 – 2019  
Zonguldak, Türkiye

**BSc, Molecular Biology and Genetics, Zonguldak Bülent Ecevit University**

## Projects

---

### Bayesian Modeling & Analysis Framework for High-Dimensional Data | Ongoing,

*Python (Scanpy, Pandas, Scikit-learn), R (Seurat), VB-GMM*

Developing a modular analysis pipeline to process large-scale, high-dimensional genomic matrices containing thousands of genes and cells. Implementing Variational Bayesian Gaussian Mixture Models (**VB-GMM**) using **Scikit-learn** to uncover latent cell subpopulations, moving beyond traditional deterministic clustering methods. Leveraging **dimensionality reduction** techniques (PCA, UMAP, t-SNE) and utilizing **Matplotlib/Seaborn** to visualize complex vector spaces and cluster distributions. Engineering a reproducible workflow that transforms raw sequencing counts into structured **AnnData** and **DataFrame** objects for advanced statistical inference.

### Comparative ML Analysis on High-Dimensional Genomic Data <sup>🔗</sup>, Python, Scikit-learn

Processed high-dimensional RNA-seq data (features >> samples) to distinguish primary and metastatic samples. Evaluated the performance of **Supervised Learning** algorithms (SVM, Random Forest, Naive Bayes) on complex numerical arrays.

Spearheaded the **Unsupervised Learning** module; implemented K-Means clustering and optimized cluster separation on **sparse data structures** to uncover hidden molecular patterns. Demonstrated how data science techniques effectively solve classification problems in high-dimensional biological vector spaces.

### Automated Data Integration Tool *🔗, Python (openpyxl)*

Developed a custom Python automation tool to extract and integrate **heterogeneous data structures** (DNA Methylation & mRNA matrices) from multi-sheet Excel workbooks.

Leveraged the **openpyxl** library to parse semi-structured spreadsheets and convert raw data into **clean, structured arrays**, significantly reducing manual effort. Engineered a reusable pipeline that merges disconnected datasets into a **unified feature matrix**, ensuring data integrity for downstream statistical analysis.

### End-to-End RNA-Seq Analysis Pipeline *🔗, Galaxy Platform*

Designed and executed a comprehensive **bioinformatics pipeline** on the Galaxy platform to analyze raw RNA-sequencing data. Orchestrated the full workflow from **Quality Control (QC)** and read mapping to **Differential Expression Analysis**, ensuring reproducibility of results. Transformed complex datasets into actionable insights using advanced **visualization techniques** (PCA, Volcano Plots, Heatmaps).

Interpreted statistical outputs to identify significant gene expression patterns.

### Microarray Data Analysis *🔗, Excel*

Conducted a comprehensive **Microarray analysis** to investigate the correlation between DNA methylation and gene expression levels across various cancer types. Created **Heatmaps** and advanced visualizations in Excel to identify and communicate complex data patterns. Performed **comparative statistical analysis** to translate high-dimensional biological data into actionable scientific insights.

## Professional Experience

---

06/2023 – 07/2025

Istanbul, Türkiye

### Molecular Biologist, *Gen Era Diagnostics*

Collaborated within the molecular biology team to execute **high-throughput** DNA Microarray and Methylation Profiling workflows. Managed the critical **pre-analytical phase** by isolating high-quality DNA from complex biological matrices (prenatal, FFPE, blood), ensuring data integrity for downstream analysis.

11/2021 – 05/2023

Istanbul, Türkiye

### Molecular Biologist, *Gelişim Tıp Laboratuvarları*

Conducted high-precision genomic data analysis using specialized software tools (Coffalyser, SeqScape, Chromas, IGV). Processed raw sequencing data generated from Sanger Sequencing and MLPA workflows to identify genetic variants. Responsible for the interpretation and **visualization of complex genomic signals**, ensuring diagnostic accuracy for diverse genetic panels.

## Skills

---

### Python

Pandas, NumPy,  
Scikit-learn,  
Matplotlib/Seaborn,  
Scanpy

### R

Bioconductor,  
ggplot2, Seurat

### Data Visualization

### Data Storytelling

### Statistical Modeling

### EDA

### SQL

### Excel (Advanced)

## Languages

---

**English** — Professional Working Proficiency