Biotech Club Mini Missions

Instructor Application - Spring 2026

Instructor Name	[Your Full Name]
Email	[your.email@university.edu]
Year	[e.g., Junior]
Project Title	**DNA Trait Predictor: AI-Powered Eye Color, Hair Color & Ancestry**

Why I Want to Teach This Workshop

Everyone is curious about their genetics—why do I have green eyes? Where did my red hair come from? This workshop answers those questions by teaching students to build an AI system that predicts visible traits from DNA data. It's hands-on genetics without needing a lab.

What makes this project special is how beginner-friendly it is. Students work with simple CSV files containing SNP data (the genetic 'spelling differences' between people), learn basic machine learning, and see immediate, visual results. No complex bioinformatics required—just Python, pandas, and scikit-learn.

By the end, students will have built a GUI application where they can input DNA markers and watch the AI predict traits in real-time. It's the perfect bridge between biology and computer science, showing that genetics is just pattern recognition with fascinating data.

Project Overview

The Big Idea: Build an AI-powered trait predictor that analyzes DNA markers (SNPs) to predict eye color, hair color, and ancestry—complete with a user-friendly GUI.

What Students Will Build:

A complete machine learning pipeline with:

- CSV data parser that reads public SNP datasets (from OpenSNP or simulated data)
- Feature extraction module to identify the SNPs that matter for each trait
- Three machine learning models:

○ **Eye Color Classifier**: Predicts blue, green, hazel, or brown (using 6 key SNPs like rs12913832 in HERC2)

○ **Hair Color Classifier**: Predicts black, brown, blonde, or red (using SNPs in MC1R, TYR, TYRP1)

○ **Ancestry Predictor**: Estimates continental ancestry percentages (European, African, East Asian, South Asian, Native American)

• tkinter GUI where users input SNP values and see predictions instantly with confidence scores

Why This Works Without a Lab

Public Datasets: We use real genetic data from OpenSNP (open-source genetic database) where people voluntarily share their 23andMe results

Simulated Data: For training, we generate realistic SNP combinations based on known trait genetics

CSV Format: All data is simple text files with columns like: SNP_ID, Chromosome, Position, Genotype (AA/AG/GG)

Tech Stack (All Beginner-Friendly)

Core Python Libraries:

• pandas: Load and process CSV files

• scikit-learn: Train Random Forest classifiers (simple, interpretable ML)

• numpy: Basic array math

• tkinter: Built-in Python GUI library (no installation needed!)

Data Sources:

• OpenSNP.org: Public genetic data repository

• HIrisPlex: Published research data on eye/hair color genetics

Why This Project is Perfect for Beginners

No Complex Setup: Everything runs on standard Python—no cloud services, no special software

Data is Just CSVs: Students already know Excel/spreadsheets, so genetic data feels familiar

Instant Visual Feedback: The GUI shows trait predictions immediately—students can test their own 'DNA' and see results

Relatable Science: Everyone can relate to eye color and hair color—way more engaging than abstract algorithms

Portfolio-Ready: They'll have a working app with a GUI to show employers or grad schools

4-Week Workshop Curriculum

Each week includes a 90-minute teaching session followed by a 60-minute hands-on lab.

Week	Teaching Component	Hands-On Activity
Week 1	**DNA 101 & Data Basics** • What is DNA? The ATCG code • SNPs: single-letter differences in DNA • How traits are determined (genetics 101) • CSV format and pandas basics	**Load & Explore SNP Data** Students download OpenSNP CSV files, load them with pandas, and explore the data structure. They'll filter for specific SNPs (like rs12913832 for eye color) and visualize genotype distributions.
Week 2	**Machine Learning Basics** • What is classification? • Training data vs. test data • Random Forest intuition (decision trees) • scikit-learn workflow	**Train Eye Color Predictor** Students build a Random Forest classifier to predict eye color (blue/green/brown) from 6 SNPs. They'll split data into train/test sets, train the model, and evaluate accuracy. Deliverable: a .pkl model file.
Week 3	**Multi-Trait Models** • Hair color genetics (MC1R gene) • Ancestry & population genetics • Model evaluation (accuracy, confusion matrix) • Confidence scores & probabilities	**Build Hair & Ancestry Models** Students train two more classifiers: hair color (black/brown/blonde/red) and ancestry (5 continental groups). They'll create a unified prediction pipeline that loads all

		three models and runs predictions on new data.
Week 4	**GUI Development & Deployment** • tkinter basics (widgets, layouts, events) • Loading models in production • User input validation • Displaying results with confidence bars	**Create the Trait Predictor App** Students build a GUI with input fields for SNP genotypes (dropdown menus: AA/AG/GG), a 'Predict' button, and result displays showing trait predictions with confidence percentages. App includes a file upload option for full CSV analysis.

Week 5: Demo Day (Optional Extension)

Students present their apps to the club! They can test each other's 'DNA' (using simulated data), compare model accuracy, and discuss extensions like adding skin tone prediction or integrating with real 23andMe data.

Key Takeaways for Students

Technical Skills

- Python data processing with pandas (CSV parsing, filtering, feature extraction)
- Machine learning with scikit-learn (Random Forest classifiers, train/test splits, evaluation metrics)
- GUI development with tkinter (layout managers, event handlers, file dialogs)
- Model persistence (saving/loading .pkl files)
- Data visualization (matplotlib for confidence bars and distributions)

Biology Concepts

- SNPs and genetic variation
- Mendelian vs. complex traits
- Population genetics and ancestry
- Ethics of genetic testing (privacy, bias, accuracy)

Portfolio Deliverables

- GitHub repo with complete codebase (data pipeline + ML models + GUI)
- Demo video showing the app in action
- Technical writeup explaining the genetics, ML approach, and results

Commitment Acknowledgment

I understand and commit to the following:

- Preparing detailed lesson plans and code examples for each week
- Leading 90-minute teaching sessions and 60-minute hands-on labs weekly
- Providing office hours or Slack support for student questions
- Curating or creating training datasets (simulated SNP data + public datasets)
- Attending instructor meetings and collaborating with other Mini Missions leads

_____

Signature

_____

Date

Supporting Materials (Optional)

Resume/CV: [link or attach]

GitHub: [github.com/yourprofile]

LinkedIn: [linkedin.com/in/yourprofile]