

CODING FOR BIOLOGIST PROJECT TOPICS

- 1 **Genomic Data Visualization Tool** - Develop an integrated bioinformatics tool for visualizing and exploring genomic data, including gene expression profiles with its mutations
- 2 **3D Protein Structure and Drug Visualizer and Analysis Tool** - Develop an interactive platform for 3D protein structure visualization, allowing detailed analysis of protein-ligand interactions
- 3 **Microbiome Diversity and Functional Analysis Tool** - Build a comprehensive tool for analyzing microbiome datasets, providing rich visualizations of microbial diversity, composition, and functional capabilities
- 4 **Cancer Mutation Signature Analysis App** - Develop a user-friendly software tool to analyze mutational signatures in cancer genomics, uncovering patterns that can guide therapeutic strategies
- 5 **Gene Expression Classification Web App** - Design a Streamlit app to upload gene expression data and classify disease vs. normal using ML models (SVM, Random Forest).
- 6 **Protein Stability Predictor Based on Mutations** - Use Biopython and ML to predict the effect of mutations on protein stability using sequence and structure info.
- 7 **Rare Disease Variant Prioritizer Using NLP + ML** - Use text mining and ML to prioritize candidate variants in rare disease cases based on literature co-occurrence.
- 8 **miRNA Target Prediction using ML** - Use sequence features to train a model that predicts miRNA-mRNA interactions.
- 9 **Single-cell RNA-seq Clustering and Cell Type Classification** - Perform clustering of scRNA-seq data to uncover cellular heterogeneity.
- 10 **Epigenetic Pattern Prediction Using ML** - Predict DNA methylation or histone modification sites using sequence and structural features.
- 11 **RNA-Seq Analysis Pipeline (Snakemake/Nextflow)** - Build an automated RNA-seq pipeline including trimming, alignment, and expression quantification.
- 12 **Variant Calling Pipeline for NGS Data** - Create a complete script-based pipeline from FASTQ to VCF using tools like STAR, GATK.
- 13 **Automatic Literature Curation from PubMed** - Build a PubMed scraper using Python and NCBI Entrez API to collect articles by keywords.
- 14 **In-Silico Mutation Analysis on Disease Genes** - Study the structural and functional impact of known SNPs in disease-related genes.
- 15 **Host-Pathogen Interaction Analysis** - Map interactions between host and viral proteins in datasets (e.g., SARS-CoV-2, HIV).
- 16 **RNA Editing Site Exploration in Neurodegeneration** - Identify RNA editing sites in ALS or Alzheimer's patient RNA-seq datasets.
- 17 **AI-Based Drug Side Effect Prediction from Molecular Data** - Use ML to predict potential adverse drug reactions based on chemical descriptors and targets.

- 18 **AI Model for Identifying Functional lncRNAs** - Use AI to distinguish functional vs. non-functional long non-coding RNAs using multi-feature inputs.
- 19 **Prediction of Cancer Progression Using Multi-Omics Data** - Combine transcriptomics, proteomics, and metabolomics data to predict cancer progression using ensemble machine learning models
- 20 **Analysis of Tumor Microenvironment Using Single-Cell Data** - Utilize unsupervised learning to identify cell populations and interactions in the tumor microenvironment
- 21 **Identifying De Novo SCN1A Mutations in Dravet Syndrome** - Combining AlphaMissense Deleteriousness Scores with dbSNP Absence and Parental Confirmation
- 22 **Pharmacogenomic Variant Identification in CYP2C19** - Utilizing AlphaMissense to Predict Functional Impact of Rare Variants Beyond dbSNP-Curated Alleles
- 23 **Population-Specific Pathogenic Variant Burden in PAH for Phenylketonuria** - Integrating AlphaMissense, dbSNP Population Data, and ClinVar
- 24 **Developing Allosteric Inhibitors of Plasmodium falciparum DHODH via Fragment Linking for Malaria**
- 25 **Leveraging Cryo-EM Structures of the MC4R-Gs Complex for Agonist Design in Obesity**
- 26 **Structure-Based Optimization of RSV Fusion (F) Protein Inhibitors for Respiratory Syncytial Virus Infection**
- 27 **CFTR Gene Variant Analysis for Cystic Fibrosis using ClinVar API and Python**
- 28 **Virtual Screening + ML Re-Ranking for MC4R Agonists**
- 29 **Deep Learning for Predicting Type 2 Diabetes from Gut Microbiota**
- 30 **Multi-Class Disease Prediction Using Microbiome Data**
- 31 **Time Series Analysis of Cell Growth or Infection Progression**
- 32 **Codon Optimization Tool for Synthetic Genes** - Create a codon usage bias calculator for optimizing gene sequences for different expression hosts (e.g., E. coli vs. human).