**HBB DNA Mutation Effect Analysis Workflow**

**1. Title**

**HBB DNA Mutation Effect Analysis Workflow**  
*Predicting DNA mutation effects using Python.*

**2. Introduction / Overview**

This workflow describes the step-by-step process for predicting the effects of DNA mutations in the **HBB gene** using Python. The analysis includes:

* Cleaning mutation data from ClinVar.
* Classifying mutations as **Silent, Missense, Nonsense, or Unknown.**
* Generating **summary tables** and **visualizations.**
* Reference files are provided for reproducibility.

**3. Data Preparation**

* **ClinVar mutation file:** clinvar\_result.txt.
* Extract **HBB gene-related mutations** using Python.
* Save the cleaned data as hbb\_mutations.csv

**Note:** Detailed mutation list is in CSV file.

**4. Reference Sequence**

* FASTA file: HBB.fasta
* Length: 441 bp.
* Used as reference for mutation mapping.

**5. Python Analysis Workflow**

**Step 1: Required Libraries**

import pandas as pd

from Bio import SeqIO

import matplotlib.pyplot as plt

**Step 2: Load Data**

mutations = pd.read\_csv("hbb\_mutations.csv")

record = SeqIO.read("HBB.fasta", "fasta")

ref\_seq = record.seq

mutations.head()

**Step 3: DNA Mutation Effect Prediction**

def dna\_effect\_predictor(cDNA\_change):

if '\*' in str(cDNA\_change):

return "Nonsense"

if 'p.=' in str(cDNA\_change):

return "Silent"

if 'p.' in str(cDNA\_change):

return "Missense"

return "Unknown"

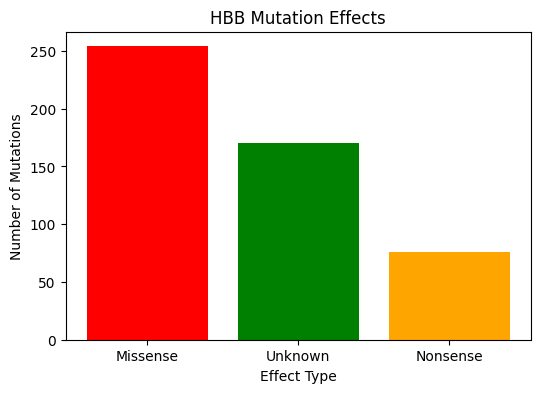
mutations['DNA\_Effect'] = mutations['Name'].apply(dna\_effect\_predictor)

mutations.head()

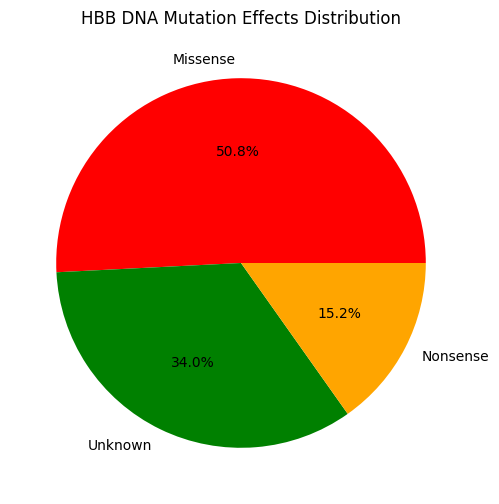
**Step 4: Summary Table**

| **Effect** | **Count** |
| --- | --- |
| Missense | 254 |
| Unknown | 170 |
| Nonsense | 76 |

**Step 5: Visualizations**

**Bar Graph:**  


**Pie Chart:**



**6. Conclusion**

* Total mutations analyzed: 500
* Missense mutations: 254
* Nonsense mutations: 76
* Unknown mutations: 170
* Graphs provide a clear visual overview.
* Workflow is reproducible using provided files.

**7. Files**

* <https://github.com/maryamhamda2002-wq/HBB_Mutation_Analysis/commit/a4219f661adfa1c9a2210ba4320f6cef02851d24>

Link → Python notebook containing the full workflow: loading CSV, predicting DNA mutation effects, summary tables, and visualizations.

* → Cleaned HBB mutation dataset (ClinVar data).
* → Reference DNA sequence of HBB gene.

**8. References**

* ClinVar database: <https://www.ncbi.nlm.nih.gov/clinvar>
* Biopython library: <https://biopython.org/>
* Python 3.x environment
* 
* 