Presentation Script

Slide 1: Title Slide

Welcome to my presentation on the 'Analysis of Protein Structure Deviations in SCN2A Mutations.' My name is Hannah Wimpy, and today, I will walk you through the work I've done to investigate the structural impacts of mutations in the SCN2A gene, particularly regarding their role in neurodevelopmental disorders like epilepsy and autism. The key hypothesis of this project is that mutations causing larger deviations in protein structure are more likely to be classified as pathogenic and associated with these disorders.

Slide 2: Introduction

SCN2A mutations have been linked to a range of neurodevelopmental disorders, including epilepsy and autism. The SCN2A gene encodes the sodium channel protein NaV1.2, which plays a critical role in neural signaling. Mutations in this gene can cause structural changes in the protein, leading to altered function and potentially severe clinical outcomes. This project focuses on using structural metrics such as Root Mean Square Deviation (RMSD) and Solvent Accessible Surface Area (SASA) to predict the pathogenicity of these mutations.

Definitions:

- RMSD (Root Mean Square Deviation): A measure of the average distance between atoms in protein structures, used to assess structural changes.
- SASA (Solvent Accessible Surface Area): The area of a protein that is accessible to a solvent like water, which helps assess changes in the protein's surface properties.

Slide 3: Project Motivation

Why is this important? SCN2A is crucial for neural development, and mutations in this gene can have profound impacts on brain function. By understanding how these mutations alter the protein's structure, we can better predict clinical outcomes and develop targeted treatments. This research aims to connect structural changes with the likelihood of diseases like epilepsy and autism.

Slide 4: Research Question and Hypothesis

The core research question is: Can structural deviations measured by RMSD and SASA predict the

pathogenicity of SCN2A mutations and their associated clinical outcomes? My hypothesis is that mutations causing larger deviations in protein structure, as measured by these metrics, are more likely to be pathogenic and associated with epilepsy or autism compared to other conditions.

Slide 5: Data and Methodology

This project uses variant data from the UniProt database, focusing on SCN2A mutations. For each variant, mutant protein structures were generated, and RMSD and SASA values were calculated. These values were then used to classify mutations based on their clinical outcomes. Tools like Modeller, UniProt, and machine learning models were used to conduct the analysis.

Slide 6: Data Visualization

The visualizations played a key role in understanding the relationships between RMSD, SASA, and clinical outcomes. By analyzing the distribution of these values, patterns were observed that supported the hypothesis: mutations with higher RMSD and SASA values tend to be pathogenic.

Slide 7: Data Processing Pipeline

The data processing pipeline involves several steps: fetching variant data from UniProt, generating mutant protein structures, calculating RMSD and SASA, classifying mutations, and visualizing the results. This pipeline ensures that the data is processed systematically and that the results are reliable.

Slide 8: Key Functions and Code Highlights

Some critical functions include:

- fetch_uniprot_data(): Retrieves data from UniProt.
- generate_mutant_sequences(): Generates mutant protein sequences.
- structural_analysis(): Analyzes RMSD and SASA to assess the structural impact of mutations.

Slide 9: Results Overview

The analysis revealed that RMSD and SASA are effective in distinguishing pathogenic from benign mutations. Mutations with higher structural deviations were more likely to be classified as pathogenic, supporting the hypothesis. These findings suggest that RMSD and SASA can be used as reliable predictors of mutation impact.

Slide 10: Regression Analysis: True vs Predicted

This slide presents the regression analysis comparing true vs predicted values. The key metrics are:

- Mean Squared Error (MSE): {mse:.4f}
- R^2 Score: {r2:.4f}

The strong correlation supports the hypothesis that structural deviations are linked to pathogenicity.

Slide 11: Residual Analysis

This slide shows the residual analysis, highlighting the accuracy of the model's predictions. The residuals cluster near zero, indicating that the model captures the relationship between RMSD/SASA and pathogenicity effectively.

Slide 12: Classification Analysis Results

This slide shows the classification analysis results with a confusion matrix. Key metrics are:

- Accuracy: {accuracy_class:.4f}
- F1 Score: {f1_class:.4f}

These results show that the model effectively classifies pathogenic vs benign mutations.

Slide 13: Association Analysis Results

This slide shows the association analysis results, demonstrating how well the model predicts different clinical outcomes like epilepsy and autism. The key metrics are:

- Accuracy: {accuracy_assoc:.4f}
- F1 Score: {f1_assoc:.4f}

These results confirm that RMSD and SASA are relevant metrics for predicting clinical outcomes.

Slide 14: Interpretation of Results

In summary, the analysis shows that mutations with higher RMSD and SASA values are more likely to be pathogenic. This highlights the importance of structural changes in determining the clinical impact of SCN2A mutations and suggests that these metrics can be valuable tools in clinical prediction.

Slide 15: Conclusion and Implications

The findings of this project indicate that RMSD and SASA can be used to predict the pathogenicity of SCN2A mutations. This has significant implications for future research and potential clinical applications, including the development of targeted therapies for neurodevelopmental disorders.

Slide 16: Future Work

Future directions include expanding the dataset to include more variants, exploring additional structural metrics, and translating these structural insights into clinical applications. The goal is to improve the predictive power of models and apply these findings to personalized treatment strategies.

Slide 17: Challenges Faced and Overcoming Them

One of the biggest challenges was ensuring the accuracy of structural predictions. Predicting how mutations affect structure requires both a strong understanding of protein biology and proficiency with bioinformatics tools. I overcame these challenges by consulting literature, seeking expert advice, and rigorously testing the models.

Slide 18: Questions and Discussion

Thank you for your attention. I'm happy to answer any questions you may have about this project, the methods used, or the implications of the findings.