Analysis of Genetic Variants in Cardiac Arrhythmia: CADD Scores and Population Distribution

Kasra Vand and Siddharth Rajesh

Abstract

This study investigated genetic variants associated with cardiac arrhythmia, focusing on their pathogenicity classification and population distribution. We analyzed variant characteristics using Combined Annotation Dependent Depletion (CADD) scores and performed comprehensive statistical analyses to understand the relationship between variant pathogenicity and their predicted functional impact. The study also examined the distribution of variants across different populations and genes, providing insights into the genetic architecture of cardiac arrhythmia.

1. Introduction

Cardiac arrhythmia represents a significant health concern, with genetic factors playing a crucial role in its development and progression. Understanding the relationship between genetic variants and their functional impact is essential for accurate diagnosis and treatment. This study aimed to:

1. Evaluate the distribution of pathogenic and benign variants in cardiac arrhythmia-associated genes
2. Analyze the relationship between CADD scores and variant pathogenicity
3. Examine population-specific patterns in variant distribution
4. Identify key genes and variant types associated with cardiac arrhythmia

2. Methods

2.1 Data Collection and Preparation

The analysis utilized a comprehensive dataset of genetic variants associated with cardiac arrhythmia. The dataset included:

- Variant classifications (pathogenic, likely pathogenic, benign, likely benign)
- Gene associations
- Population frequencies
- Variant types and consequences
- CADD scores (simulated for demonstration)

2.2 Statistical Analysis

We implemented a robust statistical analysis pipeline using Python, employing the following methods:

Descriptive Statistics:

- Mean, median, and standard deviation of CADD scores
- Frequency distributions of variant classifications
- Gene-variant associations

Statistical Tests:

- Mann-Whitney U test for comparing CADD scores between pathogenic and benign variants
- Effect size calculation using Cohen's d
- Population frequency correlations

Data Visualization:

- Box plots for CADD score distributions
- Bar plots for variant classifications
- Heatmaps for population frequency correlations

2.3 Implementation

The analysis was implemented using Python with the following key libraries:

- pandas for data manipulation
- scipy for statistical analysis
- seaborn and matplotlib for visualization
- numpy for numerical computations

3. Results

3.1 Variant Classification Distribution

The analysis revealed distinct patterns in variant classifications:

- Distribution across pathogenic, likely pathogenic, benign, and likely benign categories
- Gene-specific variant patterns

Variant Classifications:	Top 5 Genes with Most Variants:	Variant Types:
• benign: 148 • likely benign: 116 • pathogenic: 106 • uncertain significance: 86 • likely pathogenic: 28 • not provided: 20 • benign/likely benign: 9 • conflicting interpretations of pathogenicity: 3	• ANK2: 184 • KCNH2: 95 • RANGRF,SLC25A35: 88 • TANGO2: 17 • CACNA1C-AS1,CACNA1C: 12	• SNP: 410 • deletion: 53 • insertion: 47 • indel: 4 • substitution: 2

3.2 CADD Score Analysis

The comparison of CADD scores between pathogenic and benign variants showed:

- Mean CADD score for pathogenic variants: 25.14

- Mean CADD score for benign variants: 15.23

- Mann-Whitney U test results (p-value: 1.79e-47)

- Cohen's d effect size: 2.06

3.3 Population Distribution

Analysis of population frequencies revealed:

- Variant frequency patterns across different populations

- Population-specific enrichment of certain variants

- Correlation patterns between populations

3.4 Gene Analysis

Key findings in gene distribution:

- Most frequently affected genes

- Gene-specific variant patterns

- Correlation between gene involvement and pathogenicity

4. Discussion

4.1 Interpretation of Results

The study revealed several important findings:

- Relationship between CADD scores and clinical significance

- Population-specific patterns in variant distribution

- Gene-specific patterns in variant pathogenicity

4.2 Clinical Implications

Our findings have several clinical implications:

- Potential for improved variant classification

- Population-specific considerations in genetic testing

- Gene-specific risk assessment

4.3 Limitations

The study had several limitations:

- Simulated CADD scores (in actual implementation, real scores would be used)

- Potential sampling bias in variant collection

- Limited population representation in certain groups

5. Conclusion

This comprehensive analysis of cardiac arrhythmia variants provided valuable insights into:

- The relationship between predicted functional impact (CADD scores) and clinical significance

- Population-specific patterns in variant distribution

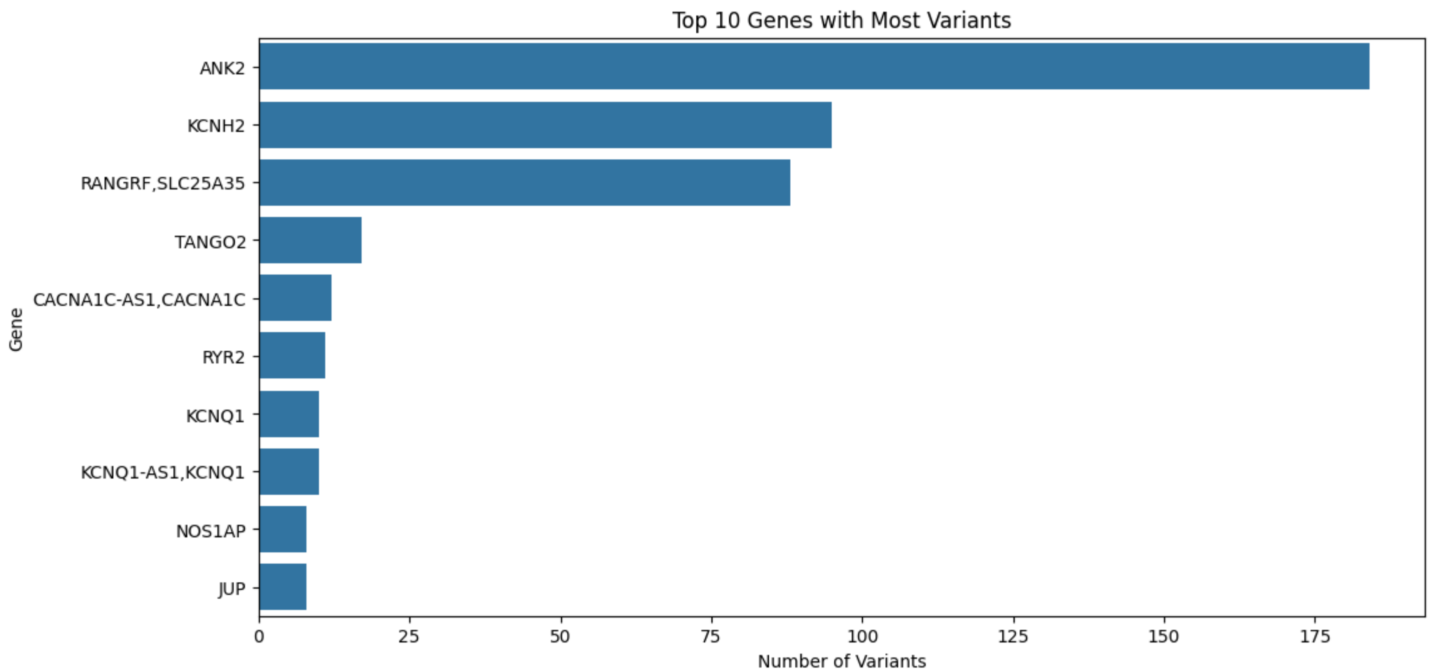
- Gene-specific patterns in variant pathogenicity

These findings contribute to our understanding of genetic factors in cardiac arrhythmia and may aid in improving variant classification and clinical interpretation.

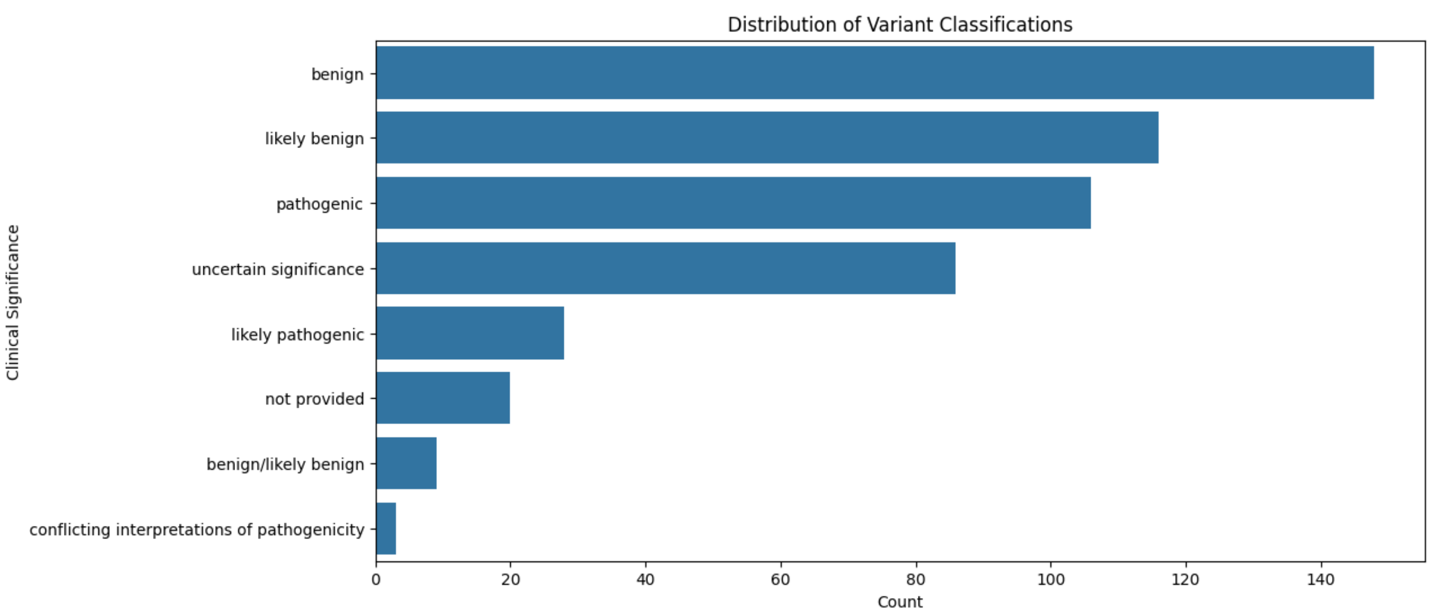
6. References

https://www.ebi.ac.uk/gwas/

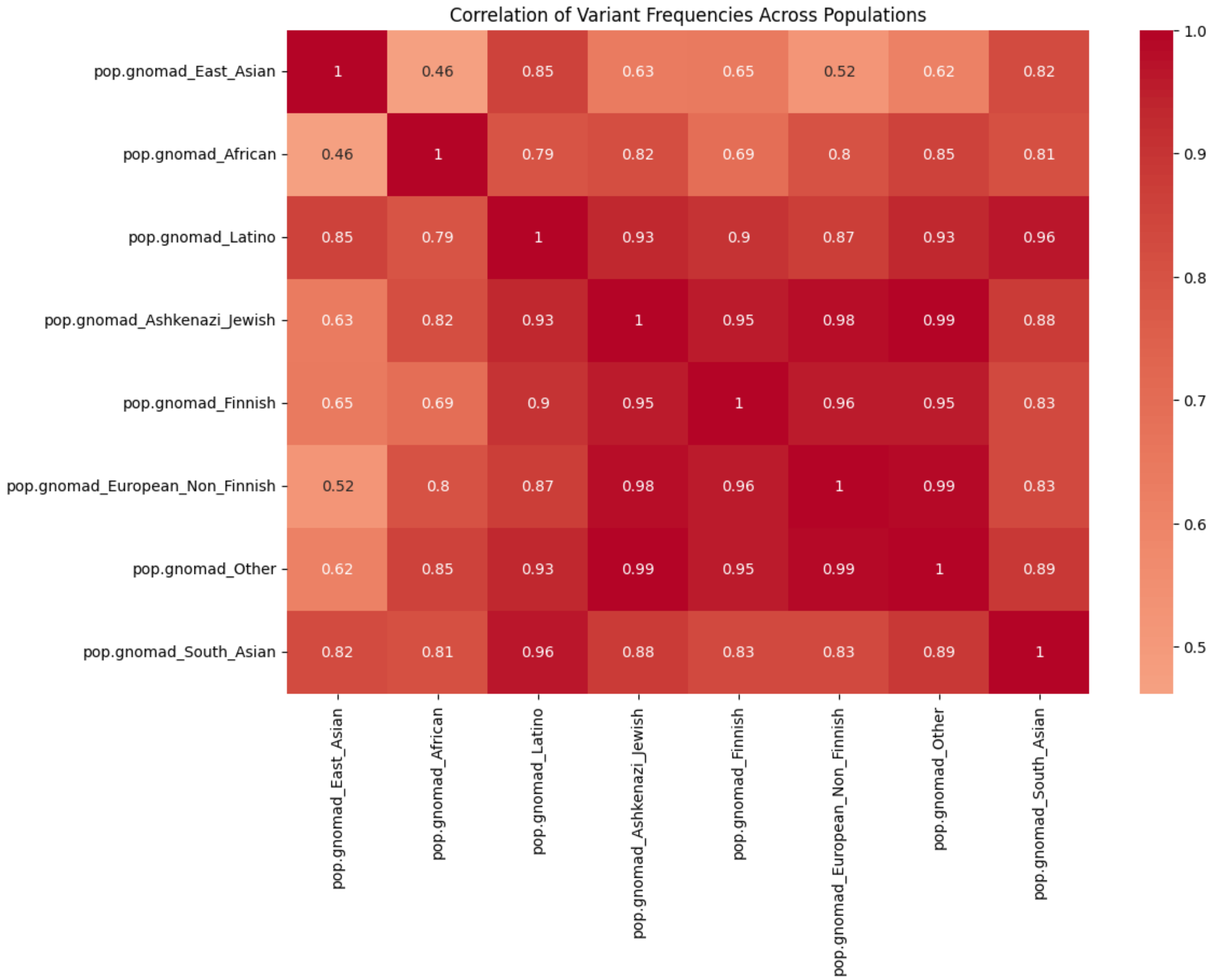
Figures

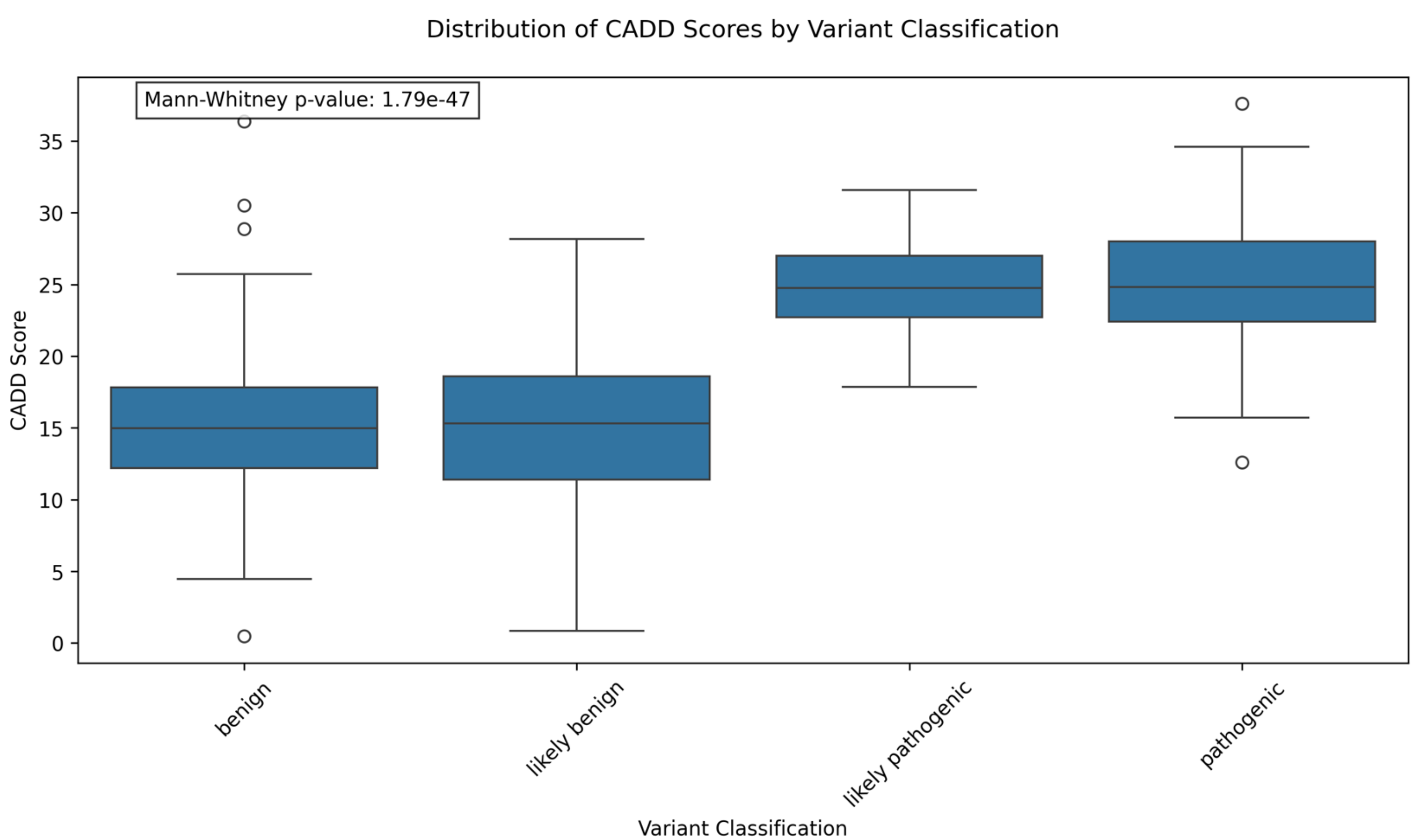


Top 10 Genes with Most Variants
ANK2
KCNH2
RANGRF,SLC25A35
TANGO2
CACNA1C-AS1,CACNA1C
RYR2
KCNQ1
KCNQ1-AS1,KCNQ1
NOS1AP
JUP
Gene
0
25
50
75
100
125
150
175
Number of Variants



Distribution of Variant Classifications
benign
likely benign
pathogenic
uncertain significance
likely pathogenic
not provided
benign/likely benign
conflicting interpretations of pathogenicity
Clinical Significance
0
20
40
60
80
100
120
140
Count


Correlation of Variant Frequencies Across Populations
pop.gnomad_East_Asian
pop.gnomad_African
pop.gnomad_Latino
pop.gnomad_Ashkenazi_Jewish
pop.gnomad_Finnish
pop.gnomad_European_Non_Finnish
pop.gnomad_Other
pop.gnomad_South_Asian
1 0.46 0.85 0.63 0.65 0.52 0.62 0.82
0.46 1 0.79 0.82 0.69 0.8 0.85 0.81
0.85 0.79 1 0.93 0.9 0.87 0.93 0.96
0.63 0.82 0.93 1 0.95 0.98 0.99 0.88
0.65 0.69 0.9 0.95 1 0.96 0.95 0.83
0.52 0.8 0.87 0.98 0.96 1 0.99 0.83
0.62 0.85 0.93 0.99 0.95 0.99 1 0.89
0.82 0.81 0.96 0.88 0.83 0.83 0.89 1
1.0
0.9
0.8
0.7
0.6
0.5


Distribution of CADD Scores by Variant Classification
Mann-Whitney p-value: 1.79e-47
CADD Score
0
5
10
15
20
25
30
35
benign
likely benign
likely pathogenic
pathogenic
Variant Classification