Personal Genomics

STUDENTS / UNIVERSITIES

Qudsi Aljabiri
Mahmut Emre Boğa
Nurettin Nusret Curabaz

Betül Çelik
Ceren Karakuş
Gloriano Sherifi

Zeynep Şencan
Neval Yaprak

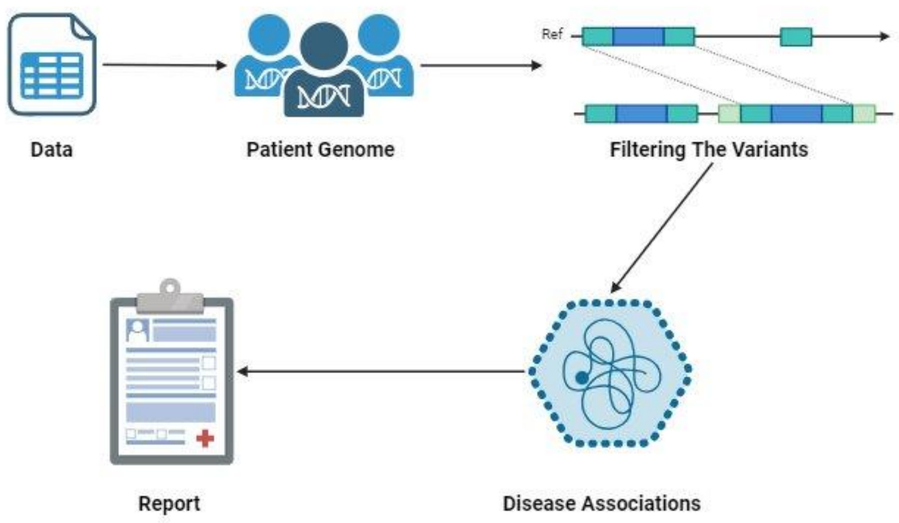
SUPERVISOR(S)

Ogün Adebali

Sabancı Üniversitesi

pure PROGRAM FOR UNDERGRADUATE RESEARCH

ABSTRACT



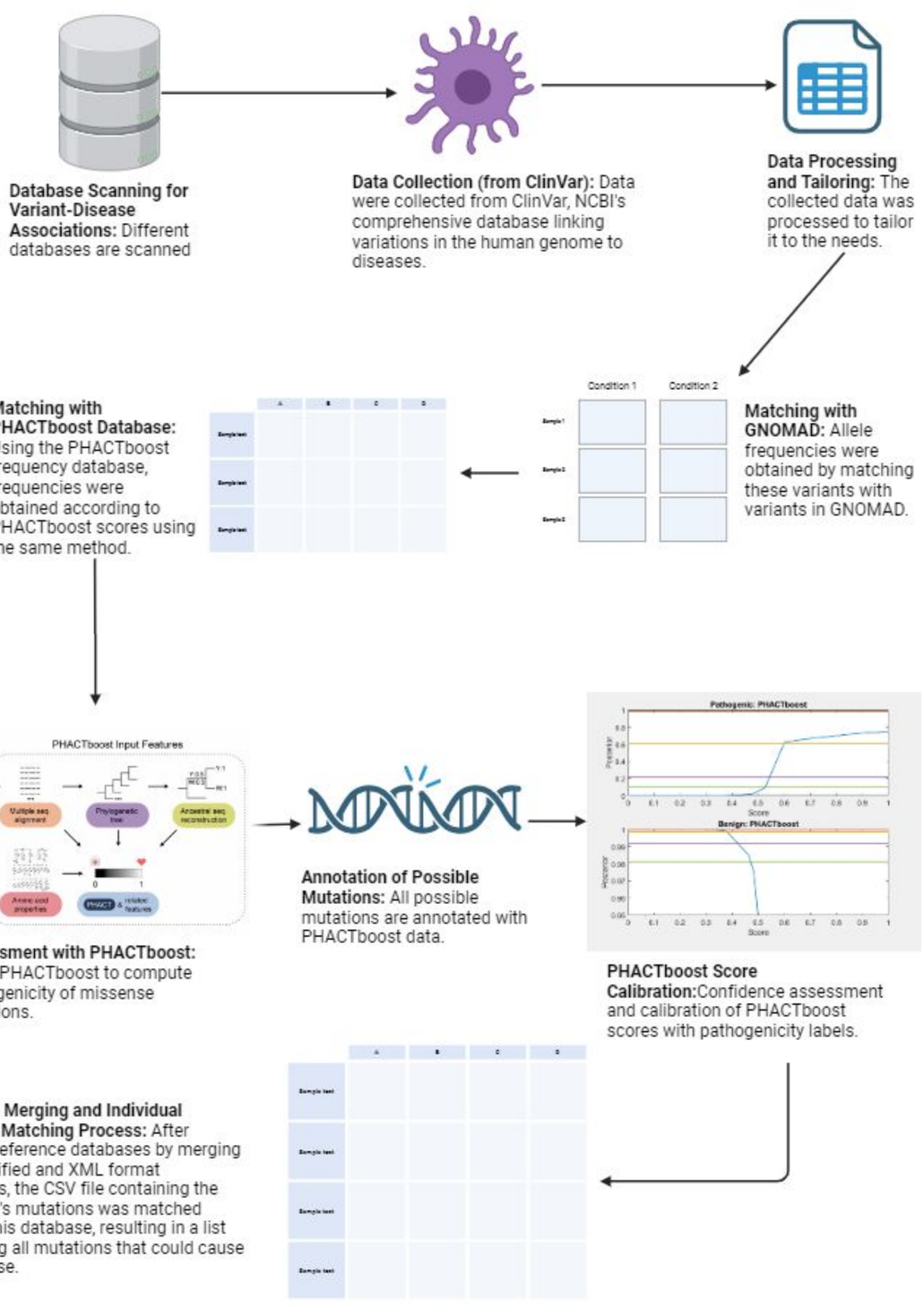
This project aims to improve algorithms for predicting disease risk through genetic analysis. It began by manually constructing disease-specific databases and understanding both databases and biology. Data collection and processing were automated using gnomAD to increase efficiency. A modeling script was developed to analyze genetic information, generating sample reports for experts and patients. Key tools include PHACTboost for estimating pathogenicity of missense mutations, ClinGen Calibrator for variant classification, and resources like ClinVar and OrphaPacket. The goal is to enhance understanding of genetic risk factors and offer solutions for personalized health management.

OBJECTIVE

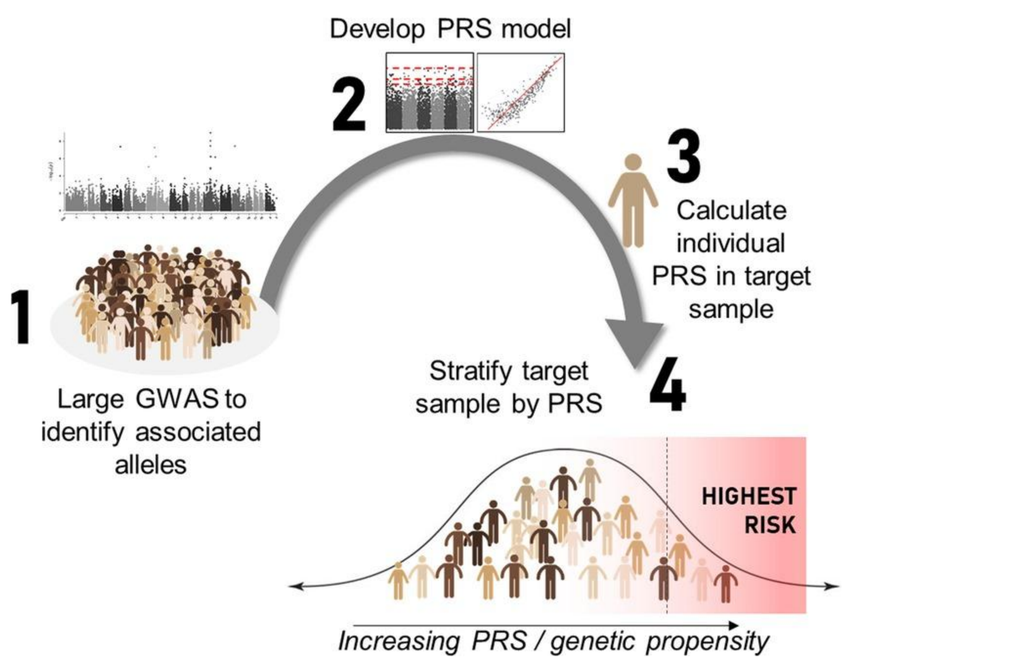
This project focuses on developing algorithms to predict disease risk through genetic analysis. It aims to create comprehensive mutation databases and models linking these mutations to genomic profiles. By generating detailed reports for patients and healthcare professionals, it seeks to provide actionable genetic risk information and advance personalized health management.

PROJECT DETAILS

The backbone of our data was curated from ClinVar, which is the NCBI's comprehensive database for variation-to-disease correlation found in the human genome. After processing that information to suit our needs, we matched those variants with those found in gnomAD to obtain allele frequencies. The allele frequencies allowed us to further separate the benign variants in our dataset. Then, using the same linking method as before, we matched them with the Sabancı University Adebali Lab's PHACTboost: A Phylogeny-aware Boosting Algorithm to Compute the Pathogenicity of Missense Mutations to obtain corresponding PHACTboost scores of variants. This tool aids to further assess the pathogenicity of a novel variant we might come across in a patients genome. we relied on the tool . All of the possible mutations in the reported genes were annotated with the information generated from PHACTboost. Then, in order to assess the confidence of PHACTboost scores and their correspondence with pathogenicity labels, further data analysis and calibration were done. The data preparation process was applied to both a simplified variant summary database from ClinVar and their more comprehensive XML data file. These unified databases were then used as references, against which an individual's CSV file containing their mutations would be compared. The final output, after running the individual's data against the unified database, would be a list of all of the individual's mutations that could lead to disease.

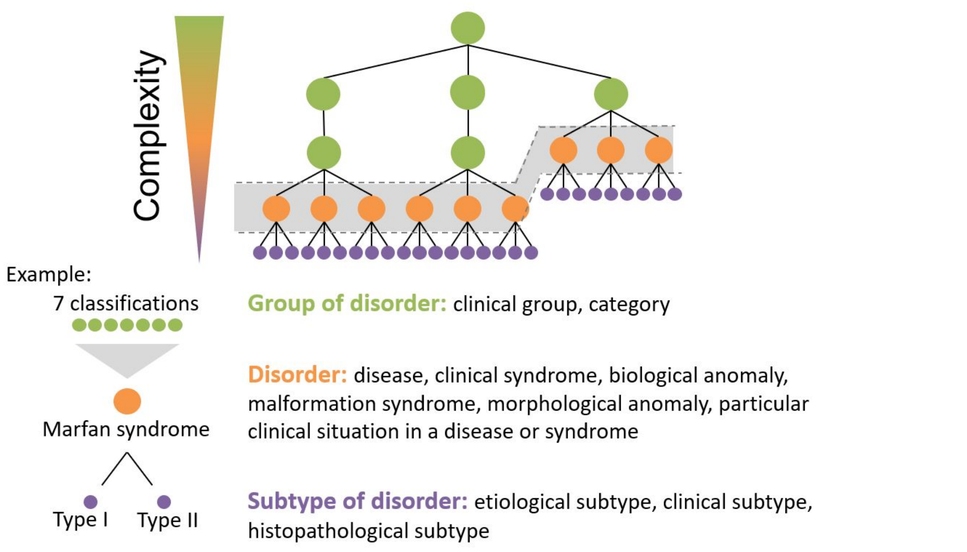


PROJECT DETAILS



For complex (polygenic) diseases, the underlying mechanisms are more complicated than simple (monogenic) diseases. While a specific combination of certain variants can lead to a disease phenotype, a smaller subset of those variants might not affect the patients health at all. In order to assess an individuals risk of having a complex disease, we gather variant effect weights curated by clinical studies for each disease phenotype from the GWAS database. Then, we obtain a sample population data from 1000 Genomes Project, quality check both the GWAS and the population data and then calculate the PRS scores for individuals in that population. This results in a distribution which we can compare our patients polygenic risk score calculation against in order to classify them as having high-risk or low-risk for the specific disease.

Reports have been prepared for the disease using Orphapacket and OMIM. The reports include the disease description, genetic information, clinical symptoms, prevalence and genetic mode of inheritance. The disease definition section describes the general definition of the disease, its etiology, clinical symptoms, course, synonyms and the broad groups of diseases to which it relates.

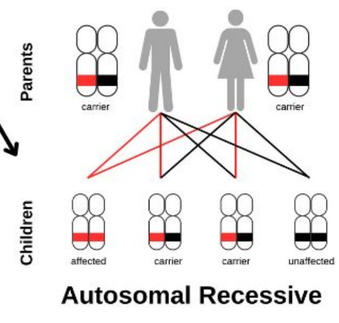


The Phenotypes section lists the common clinical manifestations of the disease and how often they occur. The Genetic Information section details disease-causing genes, mutations in these genes and genetic predisposing factors. The Prevalence section describes the incidence of the disease and prevalence rates in specific geographical areas. The Inheritance Type section describes how the disease is inherited, for example inheritance patterns such as autosomal dominant, autosomal recessive or X-linked recessive.

REPORT SECTION	CONTENT
Disease Definition	Genetic origin, Synonyms
Genetic Information	Mutations, Affected genes
Clinical Signs	Symptoms and their frequency
Prevalence Information	Regional prevalence rates
Genetic Inheritance	Mode of inheritance

GENE SYMBOL	GENE NAME	DISORDER- GENE ASSOCIATION TYPE
GENE 1	GENE NAME 1	Disease-causing germline mutation(s) in

PHENOTYPES	FREQUENCY
SYMPTOM 1	Occasional (29-5%)
SYMPTOM 2	Very frequent (99-80%)
SYMPTOM 3	Frequent (79-30%)
SYMPTOM 4	Very rare (<4-1%)



CONCLUSION

In conclusion, our method integrates data from ClinVar, gnomAD, and PHACTboost to accurately assess genetic variants and their potential to cause disease. By filtering out benign variants and utilizing PHACTboost scores, we ensure a precise evaluation of pathogenicity. For complex diseases, we calculate polygenic risk scores using variant effect weights from the GWAS database and compare these with population data from the 1000 Genomes Project to classify individuals' risk levels. The final reports, prepared using Orphapacket and OMIM, provide a detailed overview of each disease, covering definitions, genetic information, clinical manifestations, prevalence, and inheritance patterns, offering comprehensive insights to guide patient care and clinical decisions.

REFERENCES

[1] Dereli, O., Kuru, N., Akkoyun, E., Bircan, A., Tastan, O., & Adebali, O. (2024). PHACTboost: A Phylogeny-Aware Pathogenicity Predictor for Missense Mutations via Boosting. *Molecular Biology and Evolution*, *41*(7), msae136.

[2] Pejaver, V., Byrne, A. B., Feng, B. J., Pagel, K. A., Mooney, S. D., Karchin, R., ... & Topper, S. (2022). Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *The American Journal of Human Genetics*, *109*(12), 2163-2177.