Gen-ie revolutionizes early diagnosis by harnessing the power of genomic sequencing to detect rare diseases, dramatically reducing the often lengthy and challenging journey patients endure to obtain a diagnosis. Our advanced sequencing engine can analyze and sequence 500 rare disease genes, identifying specific mutations in nitrogenous bases within DNA that drive disease development. By processing the entire human genome—comprising 3 billion base pairs—Gen-ie delivers precise insights into the genetic foundations of rare diseases, enabling faster and more accurate diagnoses.

A compelling example of the transformative potential of whole-genome sequencing (WGS) comes from a national health system study that employed WGS to uncover new genetic causes and enhance the diagnosis of rare diseases. Among 13,037 participants, 1,138 out of 7,065 individuals with extensive phenotyping received a genetic diagnosis, including 9,802 patients with rare disorders. The study identified 95 Mendelian gene-disease associations, 11 of which were discovered after 2015, with 79 confirmed as causal. Additionally, WGS revealed four novel non-coding variants that contribute to disease by disrupting genes such as ARPC1B, GATA1, LRBA, and MPL. Analysis of data from the UK Biobank further demonstrated that rare alleles influence extreme red blood cell traits. This research highlights the effectiveness of WGS in both diagnostic and etiological discovery, advancing the study of rare diseases in clinical settings (Turro et al., 2020). WGS is undoubtedly the future of pharmacogenomics and rare disease diagnosis.

For data analysis, we utilized the ClinGen API, focusing on individualized research into proteins and mutated genes. Key parameters included the gene region responsible for the mutation (e.g., specific chromosome), haploinsufficiency, triplosensitivity, and hereditary inheritance patterns (e.g., X-linked, Autosomal Dominant). We developed a scoring matrix (1-10) with a color gradient, derived from data analyzed via the ClinGen API. Unique values were assigned to categories based on the relevance of research to the disease, using a metric system of standard operating procedures (SOPs). A higher SOP number indicates greater relevance. Relevance calculations were based on classifications such as definitive (1.0), strong (0.9), medium (0.5), weak (0.3), no known disease (0.05), disputed (0.05), and refuted (0.1). Additionally, gene mutations were categorized by their mechanisms, such as Loss of Function (LOF), Gain of Function (GOF), and Dominant Negative (DN), represented as columns in the matrix.

Through UniProt analysis, we found that positive haploinsufficiency and certain hereditary inheritance patterns, such as X-linked recessive and autosomal recessive, correlate with LOF. Conversely, GOF was associated with autosomal dominant inheritance, positive triplosensitivity, and mutations in chromosomal regions 2, 3, 6, 12, 13, 14, 17, and 19. However, some data were incomplete or lacked clinical research, leading us to classify the remaining gene mutations as DN using a process of elimination.

When a user inputs a genome sequence in FASTA format, the system maps it against our database to check for mutations in the 500 rare disease genes. It also identifies whether the gene is associated with GOF, LOF, or DN, along with a relevancy score based on the studies conducted on the gene. The system achieves an estimated 70% efficiency in sorting gene data functions, with the remaining 30% limited by insufficient data. The framework is designed for efficiency, featuring a sleek, user-friendly interface with multifunctional capabilities tailored for senior researchers. In the future, the efficiency of data processing can be improved by expanding research on understudied genes and implementing tools like PolyPhen-2 to predict the potential impact of amino acid substitutions on protein structure and function based on physical and comparative considerations.

This project exemplified exceptional teamwork and synergy, with all participants collaborating effectively and leveraging each other's strengths. We used Excel for gene mutation categorization, Python for coding, Reactome to identify pathways and their links, PubMed for data research, UniProt for individualized protein and gene analysis, Google Scholar for academic paper searches, and NIH resources for cross-referencing protein functions and locations.

In summary, Gen-ie represents a groundbreaking advancement in the diagnosis of rare diseases, combining cutting-edge genomic sequencing with a robust analytical framework to provide accurate, efficient, and user-friendly solutions for researchers and clinicians alike. By addressing the challenges of rare disease diagnosis, Gen-ie paves the way for faster, more precise medical interventions, ultimately improving patient outcomes and transforming the landscape of rare disease research and treatment.

Bibliography

Turro, E., Astle, W. J., Megy, K., Gräf, S., Greene, D., Shamardina, O., Lango Allen, H., Sanchis-Juan, A., Frontini, M., Thys, C., Stephens, J., Mapeta, R., Burren, O. S., Downes, K., Haimel, M., Tuna, S., Deevi, S. V. V., Aitman, T. J., Bennett, D. L., ... Ouwehand, W. H. (2020). Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*, 583(7814), 96–102. <https://doi.org/10.1038/s41586-020-2434-2>

**Introduction & Problem Statement**

1. What specific problem does your research address, and why is it important?

**RELEVANCE:**  
Does this project address the unique challenges of (a) rare disease(s) as outlined in the problem statement?

**IMPACT:**  
What potential for impact does this project have?

**NOVELTY:**  
Does this project approach the prompt with creativity and unique thinking?

**FEASIBILITY:**  
How tractable would it be to implement the project?

**SCALABILITY:**  
Does the project address the challenges of accessibility on a global scale?

**PRESENTATION:**  
Does this project clearly and persuasively communicate its impact?

For patients with *r*are disease, their journey does not stop at diagnosis. As technology has advanced remarkably in recent years, the search has expanded for treatment methods that better manage disease symptoms with fewer adverse effects. A key part of this effort is pinpointing how a genetic mutation leads to the development of disease. Due to the substantial variability of evidence in the literature, identifying a mechanism of disease for established gene-disease relationships is difficult, but can lead to the development of more targeted therapies, as well as more accurate diagnoses and prognoses.

1. How does this problem impact the field of study (e.g., healthcare, biochemistry, molecular biology)?
2. What gaps in existing research does your project aim to fill?

**Literature Review**

1. What are the key studies or findings relevant to your research topic?
2. What theories or models provide a foundation for your work?

**Data & Methods**

1. What type of data did you work with, and how was it collected?
2. What methodologies did you use to analyze the data?
3. What challenges did you face in data collection or analysis?

**Results**

1. What were the key findings of your research?
2. Were there any unexpected results?

**Discussion**

1. What do your results mean in the broader context of your field?
2. How do they contribute to existing knowledge or challenge current understandings?
3. What are the potential real-world implications of your findings?
4. How do your findings support or contradict previous research?
5. What alternative explanations could account for your results?

**Challenges & Future Directions**

1. What limitations did you encounter in your study?
2. How could future research build upon your findings?
3. What practical applications or advancements could result from your research?
4. What modifications would you make if you repeated the study?