Prepared by Team Rabbit

Accelerating Pre-IND by Mapping Disease Mechanisms to Therapeutic Candidates

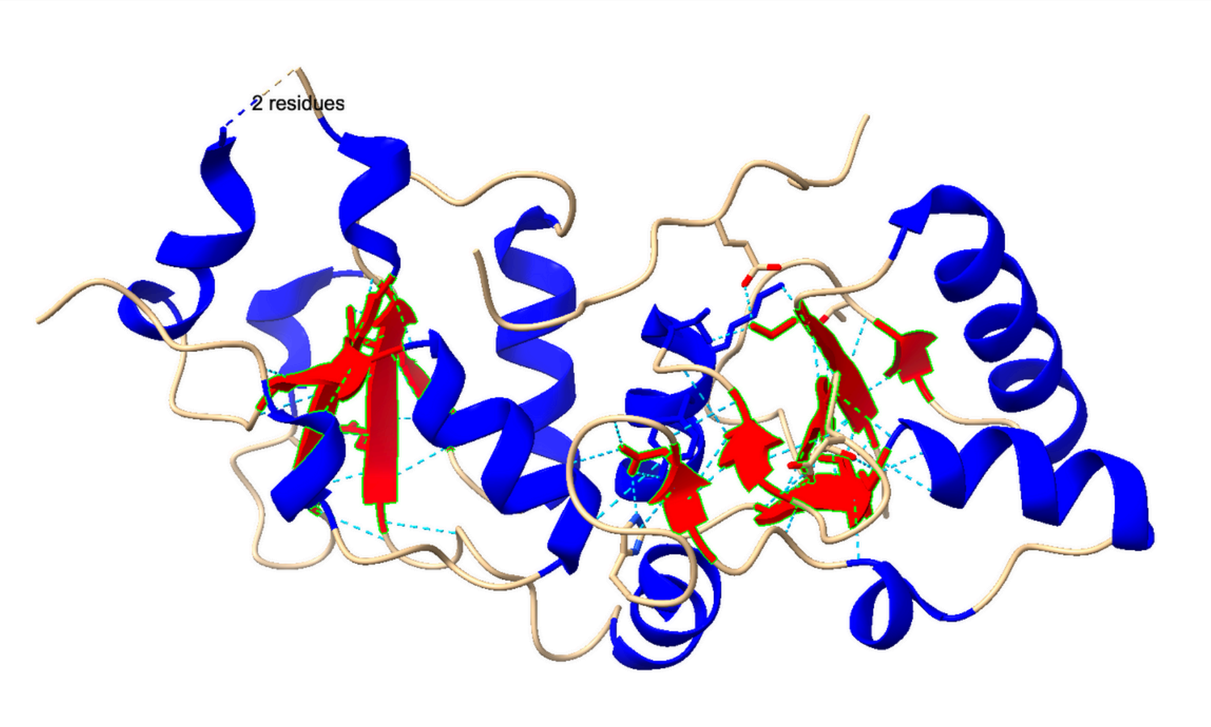


Photo of BRCA1 BRCT (PDB: 4OFB) from ChimeraX

A Framework for Disease Diagnosis and Drug Repurposing

#HACKRARE

Background & Motivation - Drug Discovery

- Recent work illustrates how **diverse mutations** (missense, repeat expansions, splice defects) **drive rare disease** through multiple molecular mechanisms.
- Multi-omics (Baxter et al.) and family-based genomic studies (Posey et al.) reveal how specific variants disrupt protein structure or cell signaling (as with huntingtin), underscoring the **need for a semi-quantitative framework to classify** loss of function, gain of function, or dominant negative.
- Yet, rare disease drug development is **costly**—pre-IND alone can reach $5.6 million (Pharmagellan)—and that excludes foundational research critical for understanding and validating targets. Because patient populations and biospecimens are limited, finding robust early-stage data can be slow, driving total costs and timelines even higher.

Year 1	4.500
Year 2	1.125
Total Pre-Clinical R&D (in millions)	**$5.63**

Why This Matters for Drug Discovery

- **Gain of Function**: Design inhibitors/antagonists to dampen overactivity
- **Loss of Function**: Pursue protein replacement, gene therapy, or approaches that enhance residual activity
- **Dominant Negative**: Block mutant interference with wild-type protein to avert dysfunction
- Key Benefit: Mechanism-informed therapies streamline treatment design by targeting disease root cause.

Core Objectives:

Mechanism Profiling:

- Classify gene variants (LoF, GoF, DN) with a confidence score to guide foundational research on disease progression.

Drug Identification:

- Match each mechanism type with a potential therapeutic strategy (agonists/enhancers for LoF, inhibitors for GoF, disruptors for DN) to streamline pre-IND decision-making.

Our Solution:

- We use an **ensemble of models** with insights from **four databases.**
- To automate the process of manually analyzing literature to classify disease mechanism using an **LLM**.
- This allows clients to determine the **mechanisms** that underly genetic mutations.

GoFCards

Ensembl

GOF LOF

G2P

Gene2Phenotype, GoFCards, Yeast2Human, LoGoFunc

```python
import requests

def fetch_clingen_gene_disease_interaction(gene_symbol):
    """Fetch gene-disease interaction summary from ClinGen API."""
    url = f"https://search.clinicalgenome.org/kb/gene-validity?search={gene_symbol}"

    try:

        response = requests.get(url, timeout=100)
        response.raise_for_status()
        print("hey")
        print("response.json() = " + response.json())
        data = response.json()

        print("hyello")

        # Debugging: Print the API response to verify structure
        print("API Response:", data)

        # Extract relevant summary fields
        interactions = []
        for record in data.get("records", []):  # Iterate through multiple records
            interactions.append({
                "Gene": gene_symbol,
                "Disease": record.get("diseaseLabel", "N/A"),
                "Classification": record.get("classification", "N/A"),
                "Mode of Inheritance": record.get("modeOfInheritance", "N/A"),
                "Evidence Level": record.get("evidenceLevel", "N/A"),
                "Last Evaluated": record.get("lastEvaluated", "N/A"),
                "Curator": record.get("curator", "N/A"),
            })

            print(interactions.gene_symbol)

        return interactions if interactions else [{"Gene": gene_symbol, "Disease": "No interactions found"}]

    except requests.exceptions.RequestException as e:
        print(f"Error fetching ClinGen data for {gene_symbol}: {e}")
        return [{"Gene": gene_symbol, "Disease": "Error retrieving data"}]
```

Database Analysis

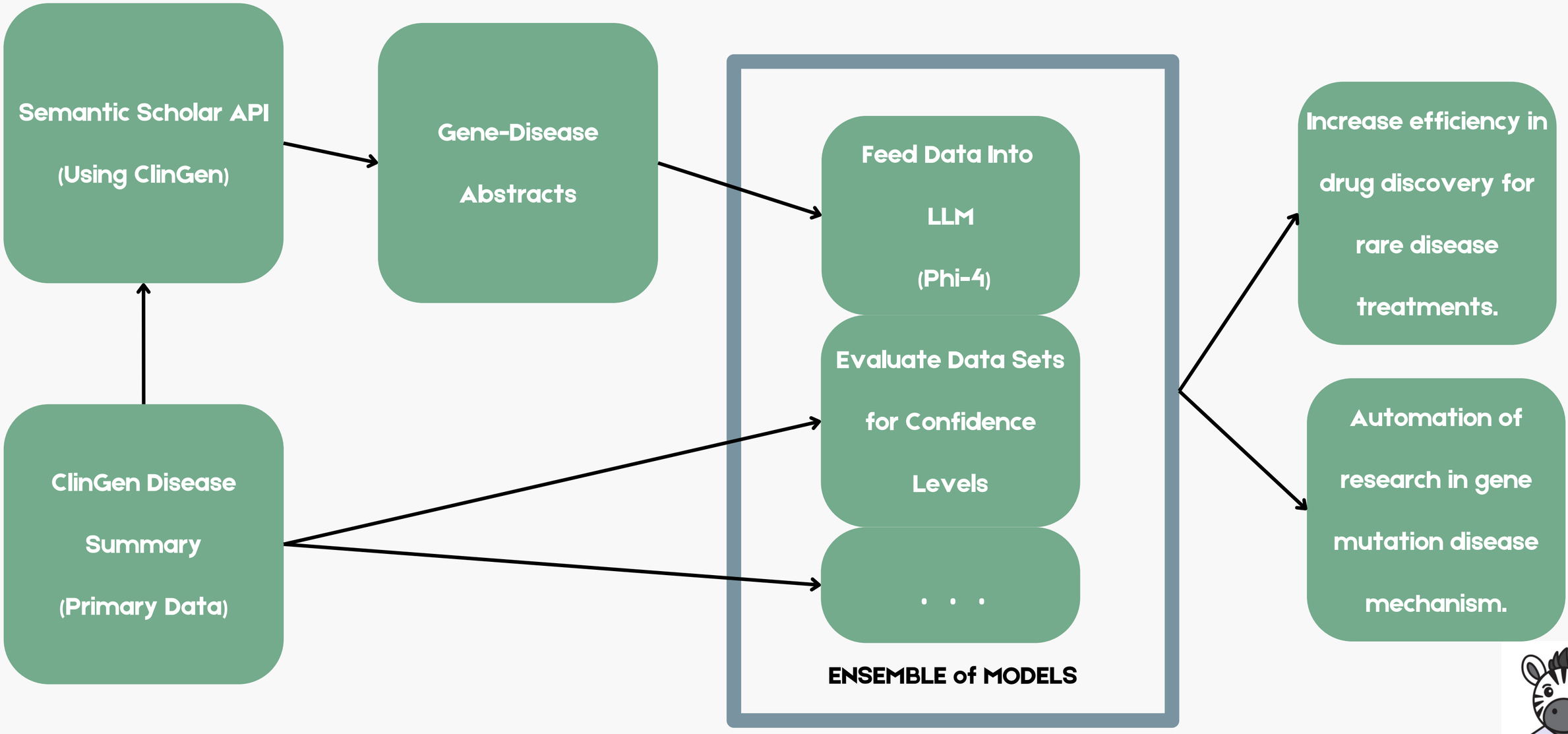
LLM scraped 2604 Papers

Collected Data

Prediction

#HACKRARE

Methods



#HACKRARE

Challenges We Faced:

- Learning **database file** types
  - Accessible vs. non-accessible.
- Databases not providing APIs or download sources.
- Multiple **aliases** for the same disease to sort through.
- Limited **time** to balance learning and gathering a multitude of datasets.
- Many biotech companies do not **publish detailed data** on how specific variants disrupt protein structure to protect their **intellectual property**, which in turn limits the availability of variant-specific mechanistic information for broader research and drug development efforts.

Database									
G2P_DD_202	**G2P00786**	ABCD1	loss of functic	ADRENOLEUKODYSTROPHY; ALD; ALDP; AMN	[1, 0, 0, 1]	ABCD1-related adrenoleukodystrophy			
G2P_Skin_20	**G2P00786**	ABCD1	loss of functic	ADRENOLEUKODYSTROPHY; ALD; ALDP; AMN	[1, 0, 0, 1]	ABCD1-related adrenoleukodystrophy			
G2P_Eye_202	**G2P00786**	ABCD1	loss of functic	ADRENOLEUKODYSTROPHY; ALD; ALDP; AMN	[1, 0, 0, 1]	ABCD1-related adrenoleukodystrophy			
LoGoFunc	**CM000636**	ABCD1	LOS	X	[1, 0, 0, 1]				

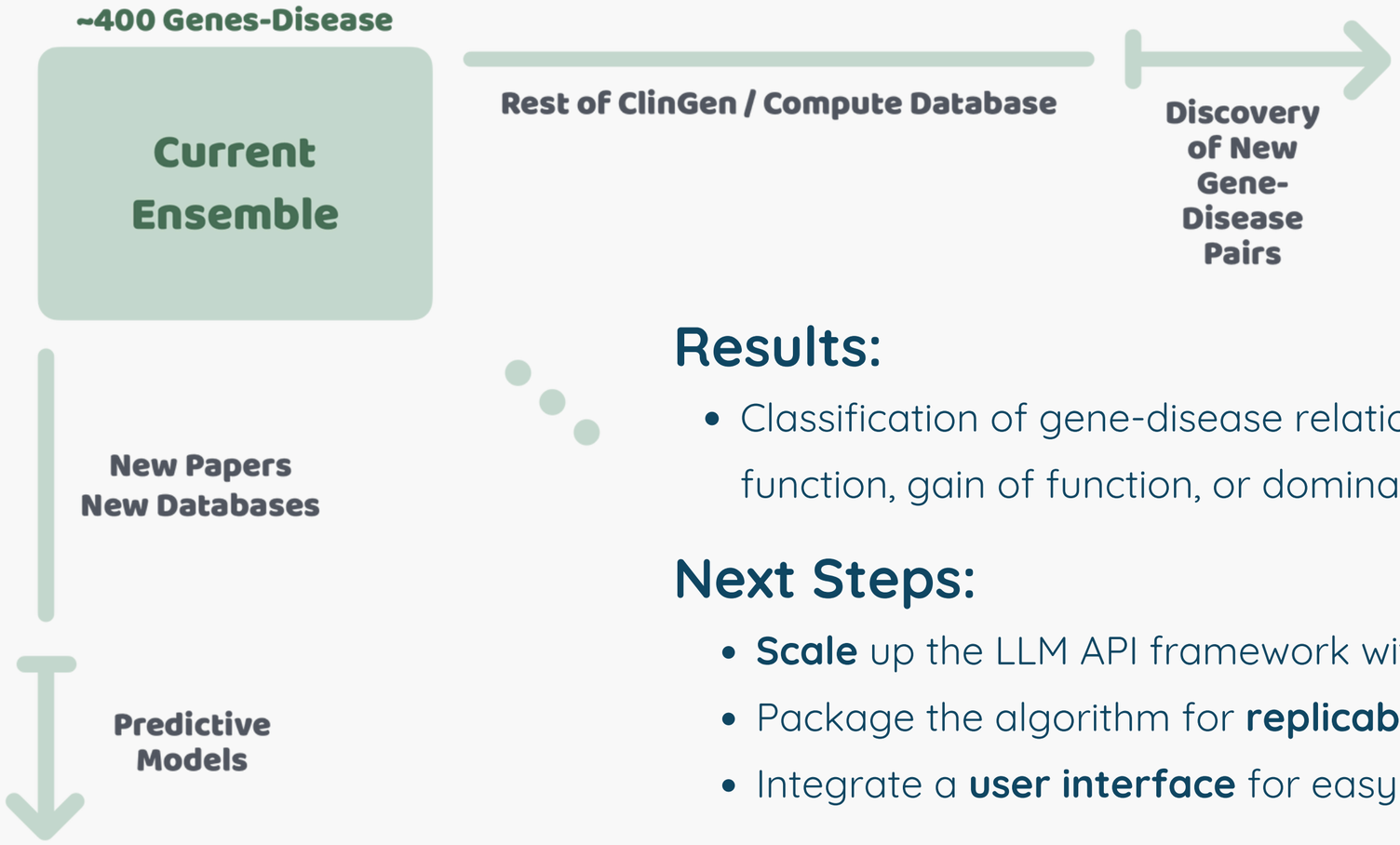
Fig. 1 Demonstration of Multiple Aliases for the Same Disease

#HACKRARE

Results & Future Expansion

Gene Symbol	HGNC ID	Disease Label	MONDO ID	Most Likely Mechanism	Score
BRCA1	HGNC:1100	Breast-ovarian Cancer	MONDO:0011450	LOF	7
SOS 1	HGNC:11187	Noonan Syndrome 4	MONDO:0012547	GOF	3
C1QTNF5	HGNC:14344	Inherited Retinal Dystrophy	MONDO:0019118	DN	6

Fig. 1 excerpt from data



Results:

- Classification of gene-disease relationships to lead to either a loss of function, gain of function, or dominant negative and an associated score

Next Steps:

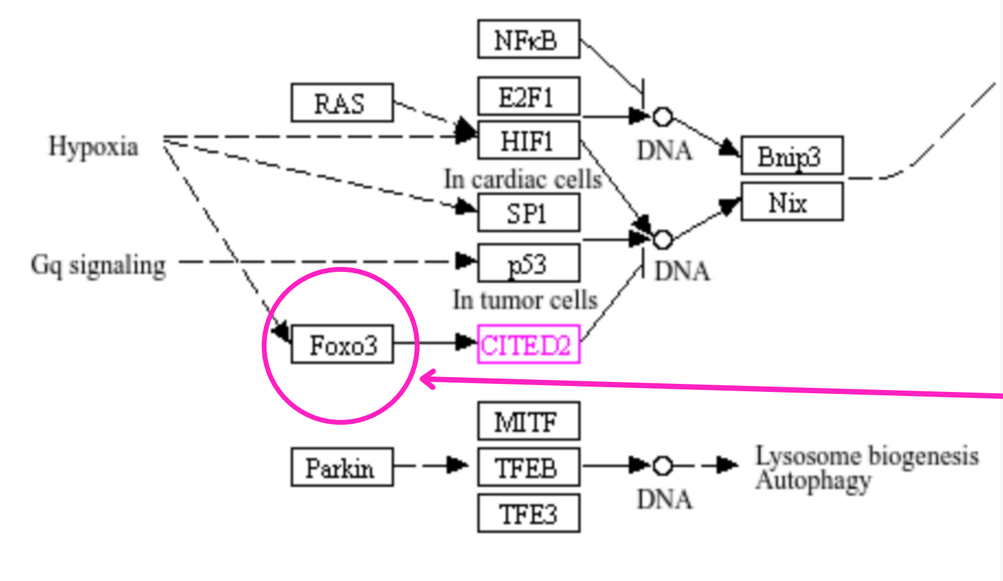
- **Scale** up the LLM API framework with more data.
- Package the algorithm for **replicability**.
- Integrate a **user interface** for easy provider access.

#HACKRARE

Discussions - Drug Repurposing Case Study

CITED2 Gain-of-Function in Congenital Heart Defects

CITED2 is a transcriptional coactivator that regulates key developmental pathways which are crucial for heart formation. Our AI framework predicts a gain-of-function mutation with a overall confidence of 0.40 in CITED2. Suppressing CITED2-driven transcription may offer a therapeutic strategy to mitigate its pathological effects.



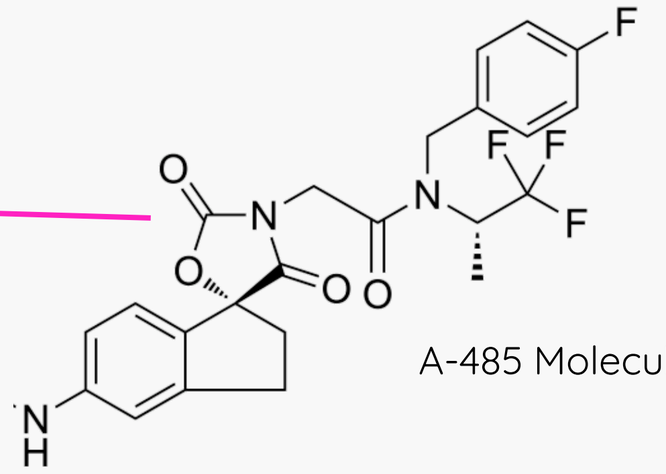
CITED2 Pathway via KEGG Pathway

A-485

Selective p300/CBP histone acetyltransferase inhibitor being studied in oncology.

Repurposing the Antagonist

Beyond its applications in cancers, A-485 has been shown to reduce lipogenesis in adipose tissue and the liver, while also decreasing hepatic glucose production through inhibition of FOXO1 acetylation. These secondary effects suggest broader metabolic modulation potential, highlighting its versatility for disease-specific repurposing in rare conditions.



A-485 Molecule

#HACKRARE

References

- Baxter RM, Sparks EA, Peak A, et al. “Understanding Rare Disease Pathogenesis: Tools, Technologies, and Translational Insights.” Frontiers in Molecular Biosciences. 2021;8:627267.
- Posey JE, O’Donnell-Luria AH, Chong JX, et al. “Insights into genetics, human biology and disease gleaned from family based genomic studies.” Genetics in Medicine. 2019;21(4):798–812.
- Ornitz DM, Legeai-Mallet L. “FGFR3 signaling in bone development and pathology.” Current Topics in Developmental Biology. 2017;123:113–142.
- Huntington’s Disease Collaborative Research Group. “A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington’s Disease Chromosomes.” Cell. 1993;72(6):971–983.
- Byers PH, Pyott SM. “Recessively inherited forms of osteogenesis imperfecta.” Annual Review of Genetics. 2012;46:475–497.
- Dietz HC, Cutting GR, Pyeritz RE, et al. “Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene.” Nature. 1991;352(6333):337–339.
- Stein, D., Kars, M.E., Wu, Y. et al. Genome-wide prediction of pathogenic gain- and loss-of-function variants from ensemble learning of a diverse feature set. Genome Med 15, 103 (2023). https://doi.org/10.1186/s13073-023-01261-9
- Wenjing Zhao, Youfu Tao, Jiayi Xiong, Lei Liu, Zhongqing Wang, Chuhan Shao, Ling Shang, Yue Hu, Yishu Xu, Yingluo Su, Jiahui Yu, Tianyi Feng, Junyi Xie, Huijuan Xu, Zijun Zhang, Jiayi Peng, Jianbin Wu, Yuchang Zhang, Shaobo Zhu, Kun Xia, Beisha Tang, Guihu Zhao, Jinchen Li, Bin Li, GoFCards: an integrated database and analytic platform for gain of function variants in humans, Nucleic Acids Research, Volume 53, Issue D1, 6 January 2025, Pages D976–D988,
- Dorrity, M. W., Queitsch, C., & Fields, S. (2019). High-throughput identification of dominant negative polypeptides in yeast. Nature methods, 16(5), 413–416. https://doi.org/10.1038/s41592-019-0368-0
- Gerasimavicius, Lukas; Livesey, Benjamin; Marsh, Joseph. (2021). Pathogenic loss-of-function, gain-of-function and dominant-neg mutations have profoundly different effects on protein structure: implications for variant effect prediction, [dataset]. University o Edinburgh. College of Medicine & Veterinary Medicine. Institute of Genetics and Cancer. https://doi.org/10.7488/ds/3153.