

A decorative graphic on the left side of the slide. It features a series of circles in teal and pink colors, connected by thin lines that create a diagonal path from the top-left towards the bottom-right. The circles vary in size and opacity, with some being solid and others semi-transparent.

RareLink

**Deep Learning For Detecting
Genetic Interactions Within Rare
Diseases**

Shivam Gandhi, Eren Shin, David An, Xichen Zhang

INTRODUCTION

In the current era, the existence of rare diseases still brings great suffering to patients and families. As a result, there are still many questions to be explored around the genetic architecture of rare diseases.

First, we have to understand that one of the main challenges in the genetic structure of rare diseases comes from the pervasive genetic interactions, a phenomenon known as **epistasis**.



NORD[®]

National Organization
for Rare Disorders

Example: Patients with Epidermolysis Bullosa:

There may be 500,000+ people living with EB worldwide.

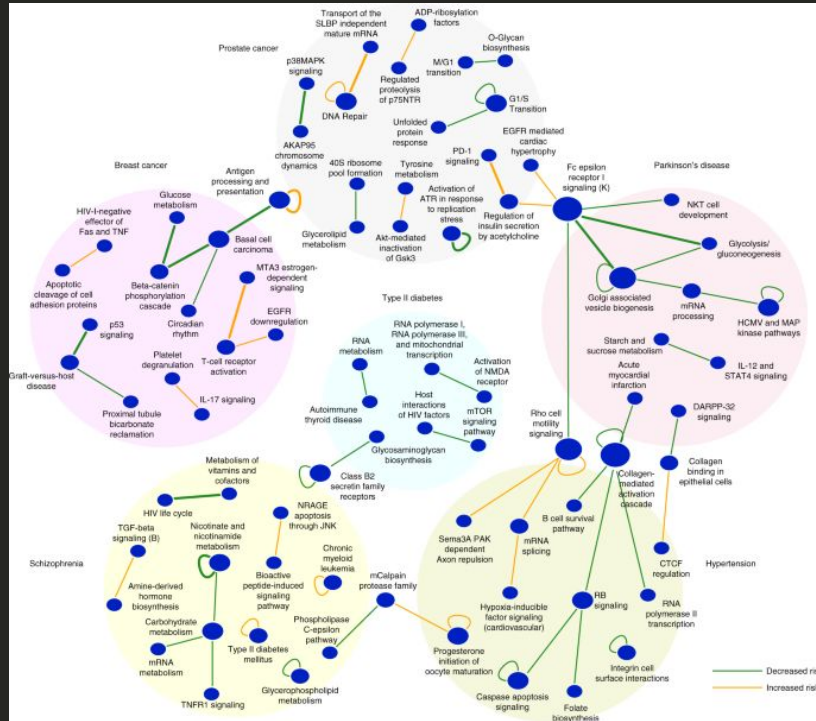
Healthcare cost:

Ranges from \$50,000 to over \$400,000 per year per patient in U.S.

Treatments:

Very Limited!

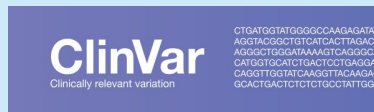
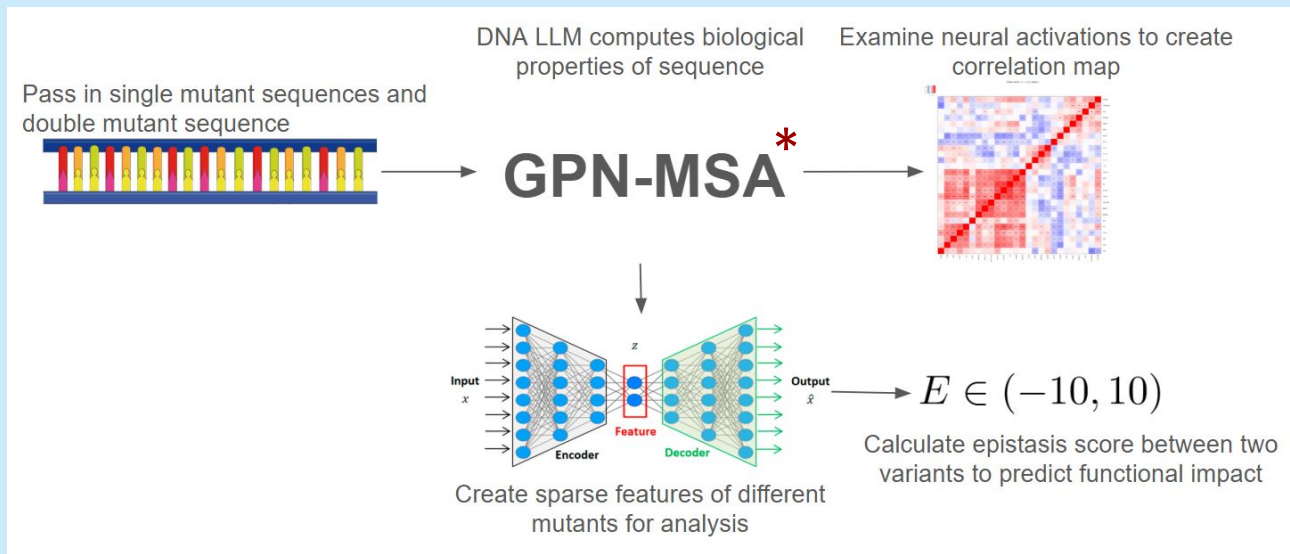
Genetic Interactions Complicate Our Understanding of Disease



- In cystic fibrosis, variations in *TGFB1* influence disease severity
- In retinitis pigmentosa, mutations in two genes together cause disease
- In sickle cell disease, levels of hemoglobin (BCL11A, HBS1L, etc.) determine the severity of sickling
- Dominance relationships make diagnosing autosomal recessive diseases difficult

RareLink: Detecting Genetic Interactions With Deep Learning

Recent advances in ML for biology makes DNA LLMs a prime candidate for detecting genetic interactions

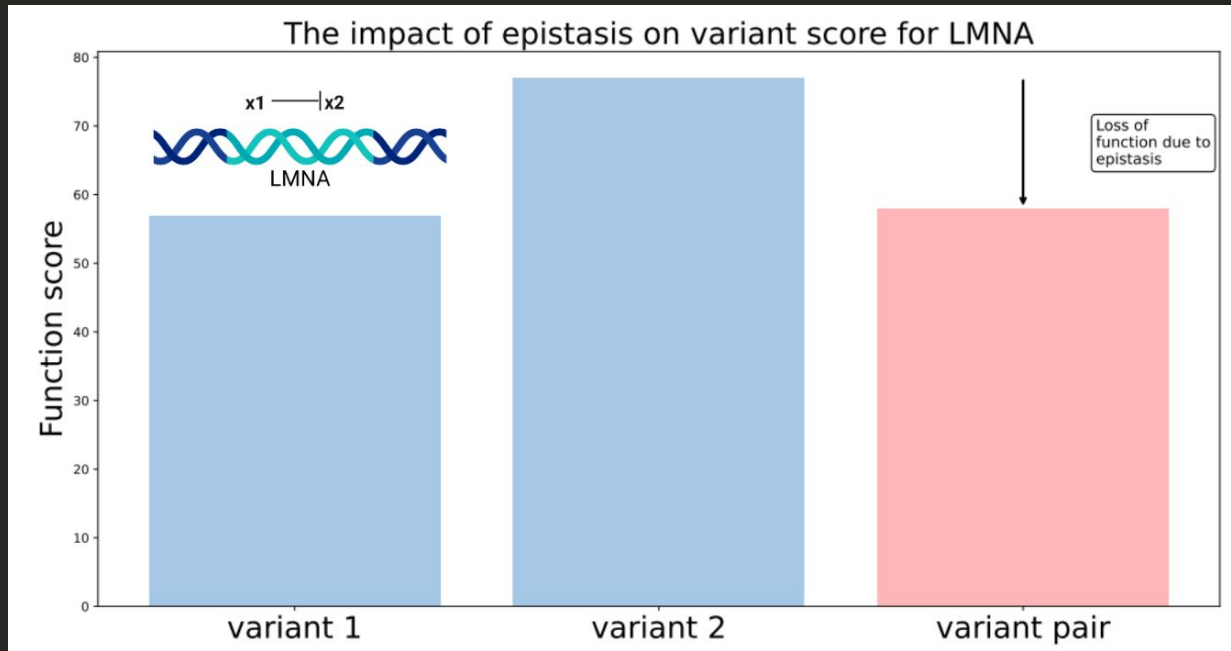


Data sources: ClinVar, OMIM, COSMIC

ML Models: GPN-MSA, Evo

**genomic pretrained network with multiple-sequence alignment*

RareLink predicts LoF genetic interaction between disease implicated variants in *LMNA*



From our function score, we calculate a value of epistatic strength E :

- $E \ll 0$: LoF interaction
- $E = 0$: no interaction
- $E \gg 0$: GoF interaction

RareLink Has a Web App to Aid Clinicians

Select a Gene

-- Select a Gene --

✓ BRCA1
BRCA2
TP53
EGFR
PTEN
KRAS
BRAF
MYC
ALK
NF1

Next

Variant Selection

-- Select Variant 1 --

-- Select Variant 2(Optional) --

Submit

Gene-Variant Disease Report

Home

Variants & Genes

Disease Report

Download

Search diseases...

Selected Gene & Variant

Gene: TP53

Variant(s): c.817C>T

Variant(s): c.215C>G

Edit Selection

Diseases Associated with Selected Gene & Variant

Disease Name	Strength of Evidence	Mechanism	Symptoms	References
Breast Cancer	0.00231	Dominant Negative	Breast mass, Lumps	Study 1 , Study 2
Ovarian Cancer	-6.375	Dominant Negative		

Pick one specific gene -> choose one/two variants -> Generate the Disease report

Future Directions

Scalability

- Use our dataset of all localized rare disease variant pairs in the human genome to flesh out the range of variants accessible on Rarelink's platform.

Biological Interpretability

- Sparse autoencoders can detect biologically relevant features of DNA sequences. Can we detect new disease mechanisms and potentially therapeutic targets?

Refining Our Metrics

- How valid are our epistasis metrics? Are the interactions we're detecting real? We aim to compare our metrics to data from CRISPR screens to get as accurate as possible

Fleshing Out a User Friendly Platform

- How will clinicians use RareLink to get patients a better diagnosis? How can we improve the workflow in rare disease care? We aim to do user interviews across providers and researchers.

Who are we?



Xichen Zhang

Undergraduate at SUNY Buffalo studying Applied Math with computing concentration.



Eren Shin

Alumnus of MIT 2023, Associate Computational Biologist at MGH/Broad Institute



David An

Undergraduate at Harvard studying Human Developmental and Regenerative Biology



Shivam Gandhi

PhD student at Harvard studying mathematical and machine learning methods for genetics