JellyByte

Rare Diseases Pathway Exploration Framework

Ilan Mitnikov

MIT Grad Student

Molecular Machines Group

David Magrefty

MIT Grad Student

Andrew Lo Lab

Liyam Chitayat

MIT Grad Student

Ed Boyden Lab

Itamar Chinn

MIT Grad Student

Barzilay Lab

Predicting the effects of rare disease mutations on a system biology scale “virtual cell”

We have 3 goals

01 Compare system effects of rare disease mutations to create bigger cohorts of patients

02 Identify the 5,000 unknown patients who have CMT4J using biomarker analysis

03 Identify common pathways between existing drug targets and rare disease impact to repurpose drugs

We combined 9 data sources and applied 3 different machine learning models over hundreds of proteins to build a “virtual cell”

Data Sources

- Uniprot - Protein sequences and metadata
- BioGrid - Protein-protein interactions
- RHEA - Metabolic pathways
- BRENDA - Metabolic pathways
- ProteinAtlas - targets of FDA approved drugs
- Chebi - Metabolite metadata
- PubMed - Literature
- Protein Data Bank (PDB) - Protein structures
- AlphaFold Database - Predicted protein structures

Machine Learning Models

- AlphaMissense (predictions)
  - Identifying pathogenic mutations
- PeSTo
  - Predicted protein-protein and protein-ligand interactions sites
- AlphaFold-Multimer
  - Predicted protein complex structures

BRENDA ChEBI RCSB PDB PROTEIN DATA BANK AlphaFold Protein Structure Database

UniProt Rhea THE HUMAN PROTEIN ATLAS PubMed® BioGRID

PeSTo

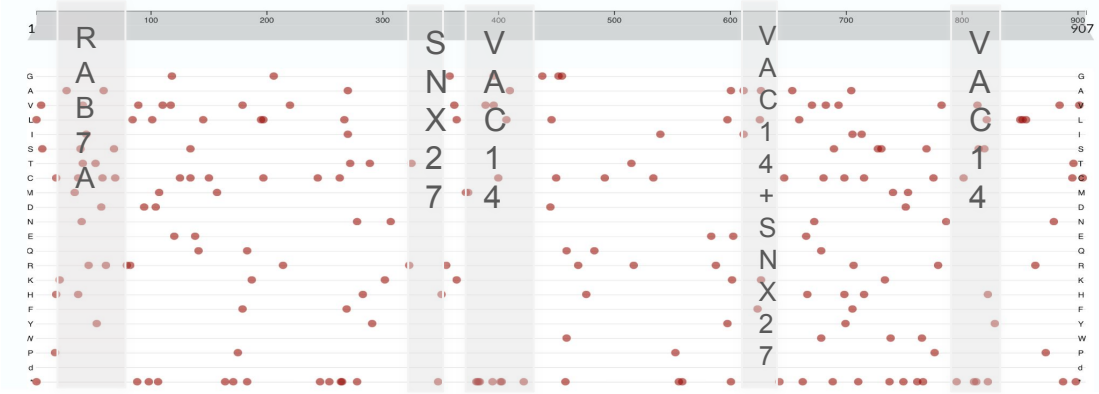
Google DeepMind

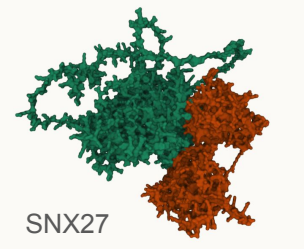
AlphaFold-Multimer

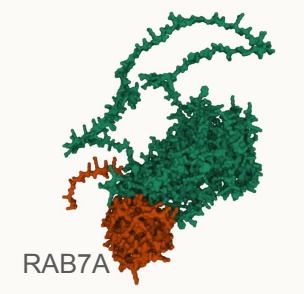
AlphaMissense

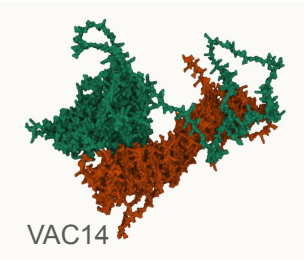
Using interaction site models we predict “pathway impact score” from pathogenic mutations which inform changes on “virtual cell”

- We select the nearest neighbors of our target gene - in this case the FIG4 gene
- We predict the protein complex structure using AlphaFold-Multimer to predict whether known or predicted pathogenic variants lie on the protein-protein interaction interface and will impact protein-protein interactions (bioRxiv 2024.02.19.580970)
- We predict the protein-ligand interfaces using PeSTo to identify whether known or predicted pathogenic variants lie on the protein-protein interaction interface and will impact enzymatic activity
- We give an “impact score” depending on how close to an interface the variant is and apply these scores to the “virtual cell” interactions









Pathway and Biomarker Identification

– Excluded water and other very common molecules
(glutathione, 10.59)
(1-phosphatidyl-1D-myo-inositol 3,4-bisphosp(5-), 9.29)
(1-phosphatidyl-1D-myo-inositol 3,5-bisphosp(5-), 7.96)
(1-phosphatidyl-1D-myo-inositol 3,4,5-trisphosp(7-), 6.79)
(oxalic acid, 5.37)
(D-glucopyranose, 5.02)
('HC02053[c]', 4.59)
(2-oxoglutaric acid, 4.37)
(succinic acid, 4.34)
(1-phosphatidyl-1D-myo-inositol 4-phosphate(3-), 3.77)
(1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate, 3.48)
(phosphatidate(2-), 3.28)
(1-phosphatidyl-1D-myo-inositol 5-phosphate(3-), 3.07)
('HC02052[c]', 2.87)
('HC02050[c]', 2.87)
('HC02086[c]', 2.87)
('HC02054[c]', 2.87)
('HC02051[c]', 2.87)
('HC02093[m]', 2.87)
(sulfate, 2.56)
(D-glucopyranose 6-phosphate, 2.31)
('pail5p_hs[g]', 1.44)
(D-glucopyranose 1-phosphate, 1.27)
('alpa_hs[c]', 1.26)
(thiosulfate(2-), 1.12)
(9H-xanthine, 0.87)
((S)-malic acid, 0.84)
(adenine, 0.76)
(uracil, 0.72)
(hypoxanthine, 0.7)
(guanine, 0.65)
(D-ribosylnicotinic acid, 0.62)
(2-deoxy-D-ribofuranose 1-phosphate, 0.58)
(alpha-D-ribose 1-phosphate, 0.52)
(nicotinamide, 0.45)
(sn-glycerol 3-phosphate, 0.42)
('ggn[c]', 0.4)
(thymine, 0.26)

Drug Repurposing

We calculate the pathway similarity between FIG4 and all other 854 cured diseases with the hope to find a combination of available drugs which could alleviate the symptoms of rare diseases

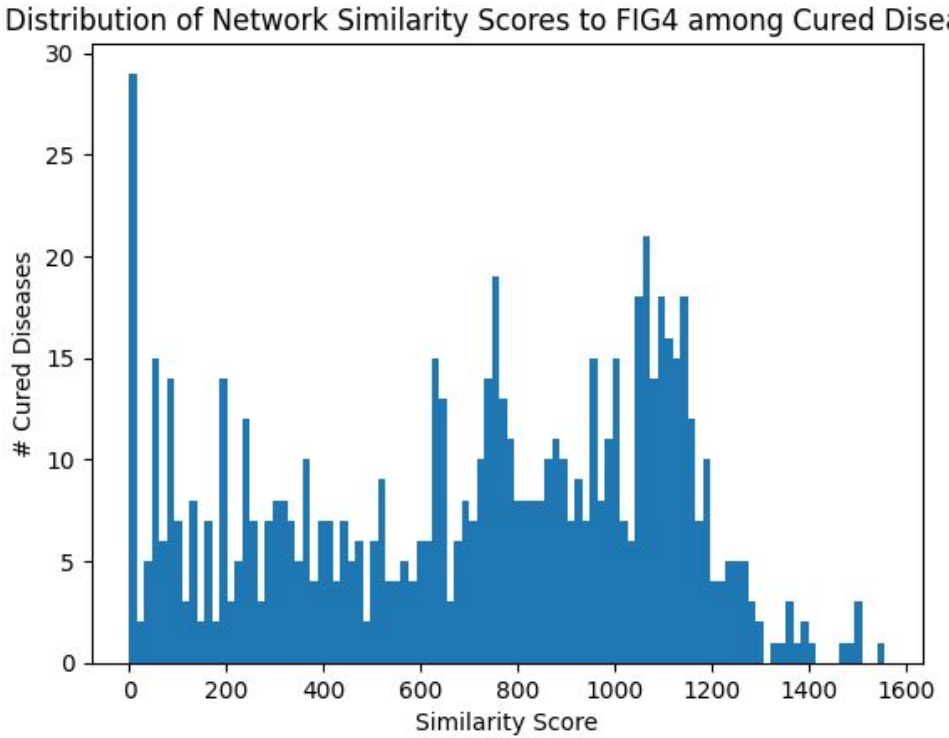
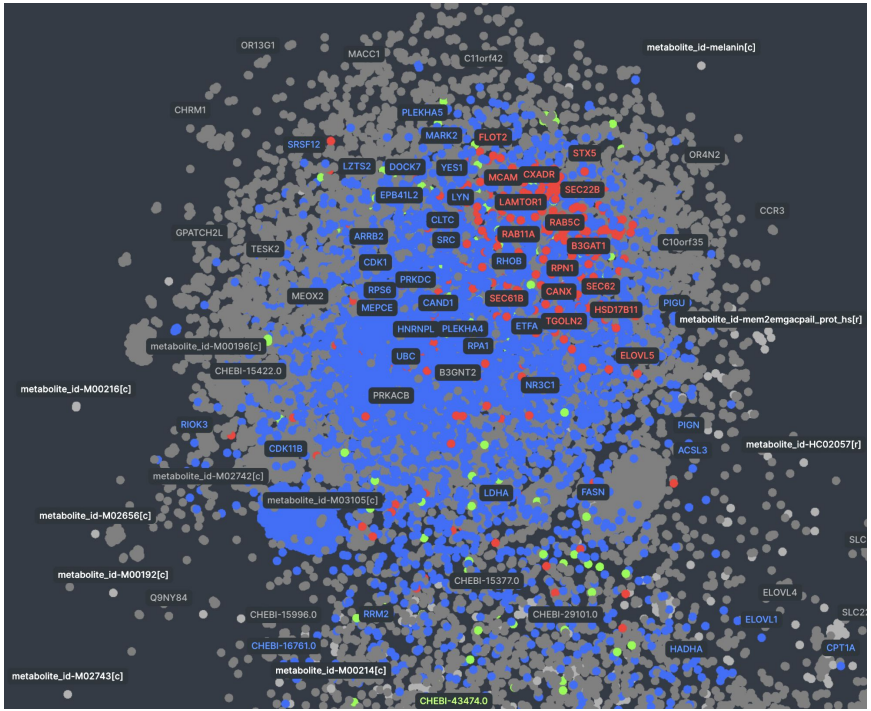


FIG 4 and 1A1COL



The future of creating a “virtual cell”

Over the hackathon we achieved

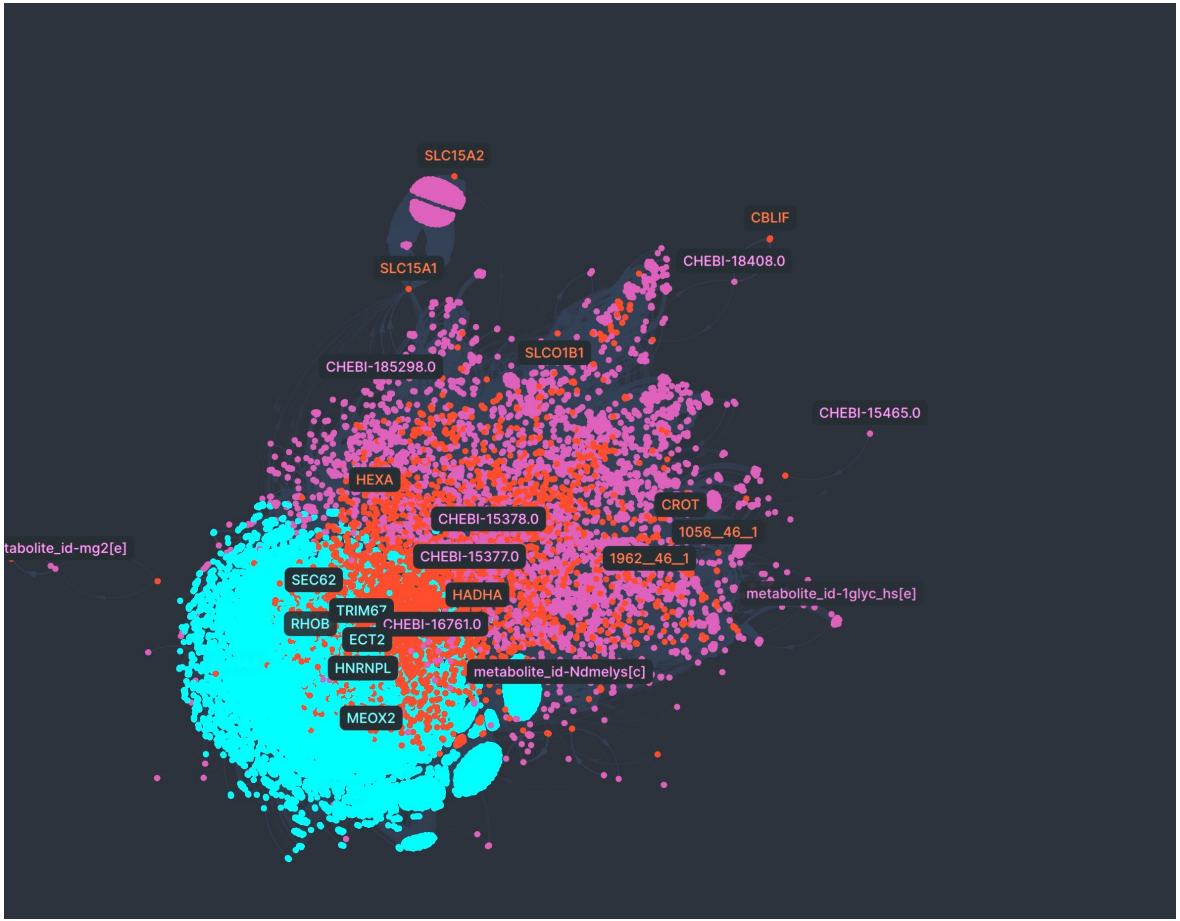
- Our project attempts to solve a key challenge in rare disease research in which we have low data and wish to more easily diagnose patients who are not immediately sent to full-genome sequencing (via biomarkers) and then efficiently propose potential therapeutic targets.
- We successfully created a first-of-its-kind “virtual cell” in which we can quickly use complex state-of-the-art machine learning methods to analyse the impacts of genetic mutations.
- While we present a first shot on goal and extract new insights in just one weekend, we expect this project to expand and evolve given additional time.

Importance of future work in this space and associated challenges

- In rare diseases we have extremely limited data about each disease and limited cohorts for both research and clinical trials
- Building a system-level “virtual cell” can take advantage of vast amount of literature and experimental data to alleviate issues caused by lower data
- The key challenges involve how to apply experimental data to understand the confidence we have in all of these interactions
- Additionally we would like to be able to develop a method to compare between different diseases efficiently in-silico to build bigger cohorts of patients across diseases

We plan to host the tool on a live website soon. Please reach out on Slack for immediate access!

David Magrefty, Itamar Chinn, Ilan Mitnikov, Liyam Chitayat


SLC15A2
CBLIF
SLC15A1
CHEBI-18408.0
SLCO1B1
CHEBI-185298.0
CHEBI-15465.0
HEXA
CROT
CHEBI-15378.0
1056__46__1
tabolite_id-mg2[e]
CHEBI-15377.0
1962__46__1
SEC62
HADHA
metabolite_id-1glyc_hs[e]
TRIM67
RHOB
CHEBI-16761.0
ECT2
HNRNPL
metabolite_id-Ndmelys[c]
MEOX2