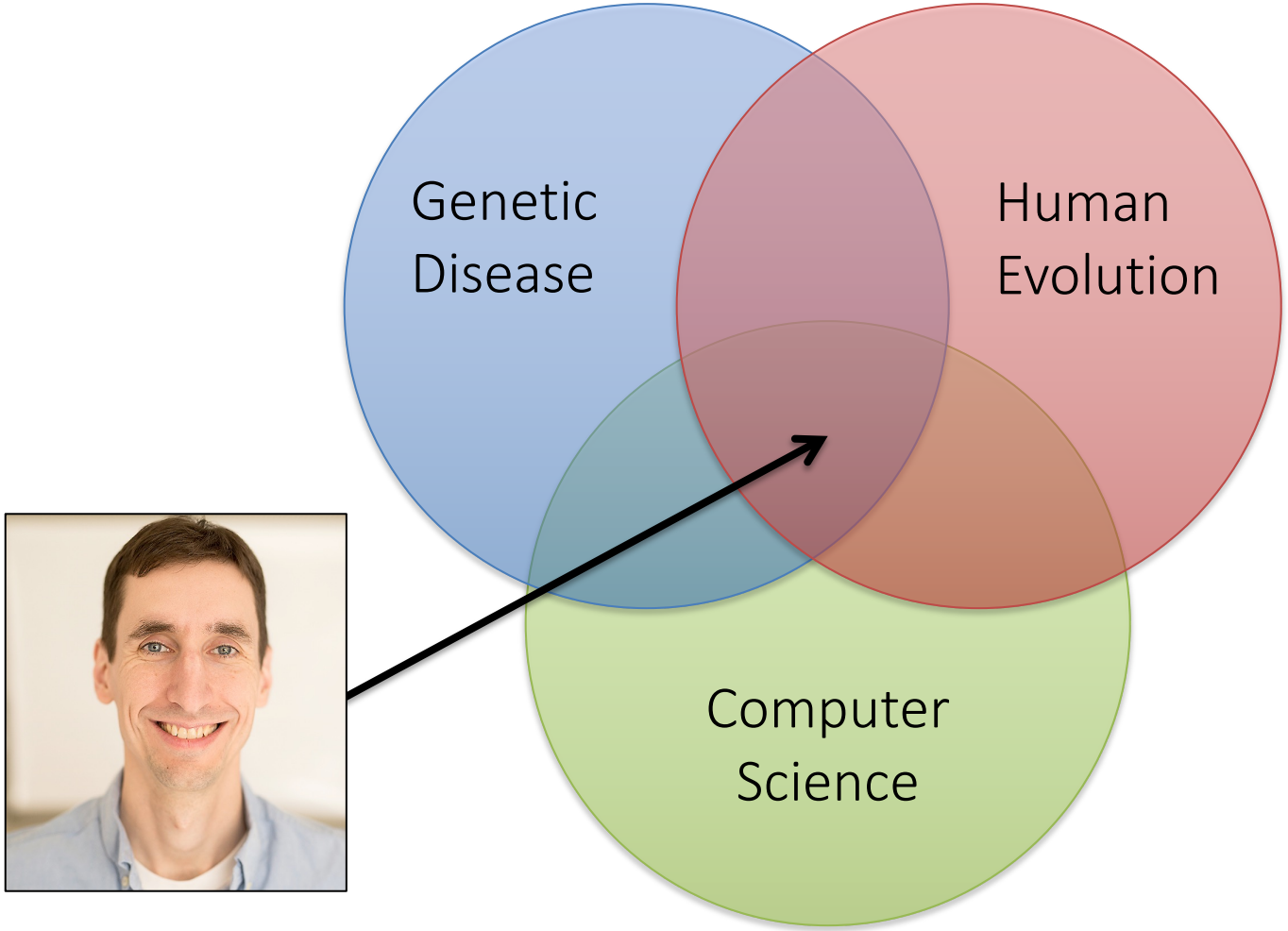
Personalized Structural Biology for Genome Interpretration

Tony Capra
Associate Professor – BCHSI, Epi&Biostats
http://www.capralab.org/

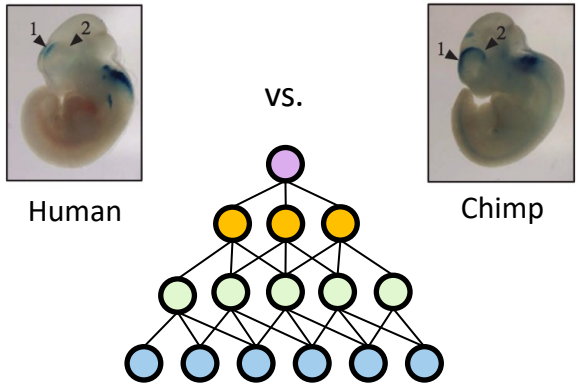
Pizza Talks 2022/3


Genetic
Disease
Human
Evolution
Computer
Science

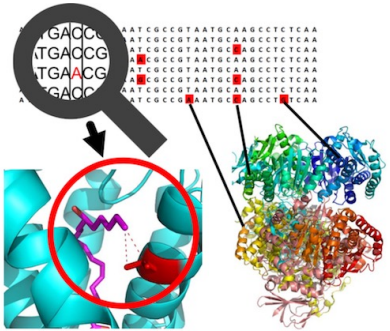
Research Foci

1. Recent Human Evolution

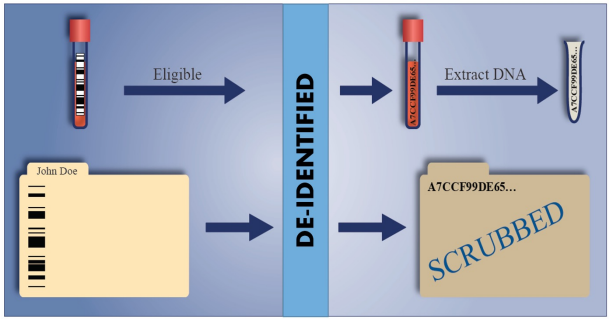
2. Machine Learning + Gene Regulation



3. Personalized Structural Biology



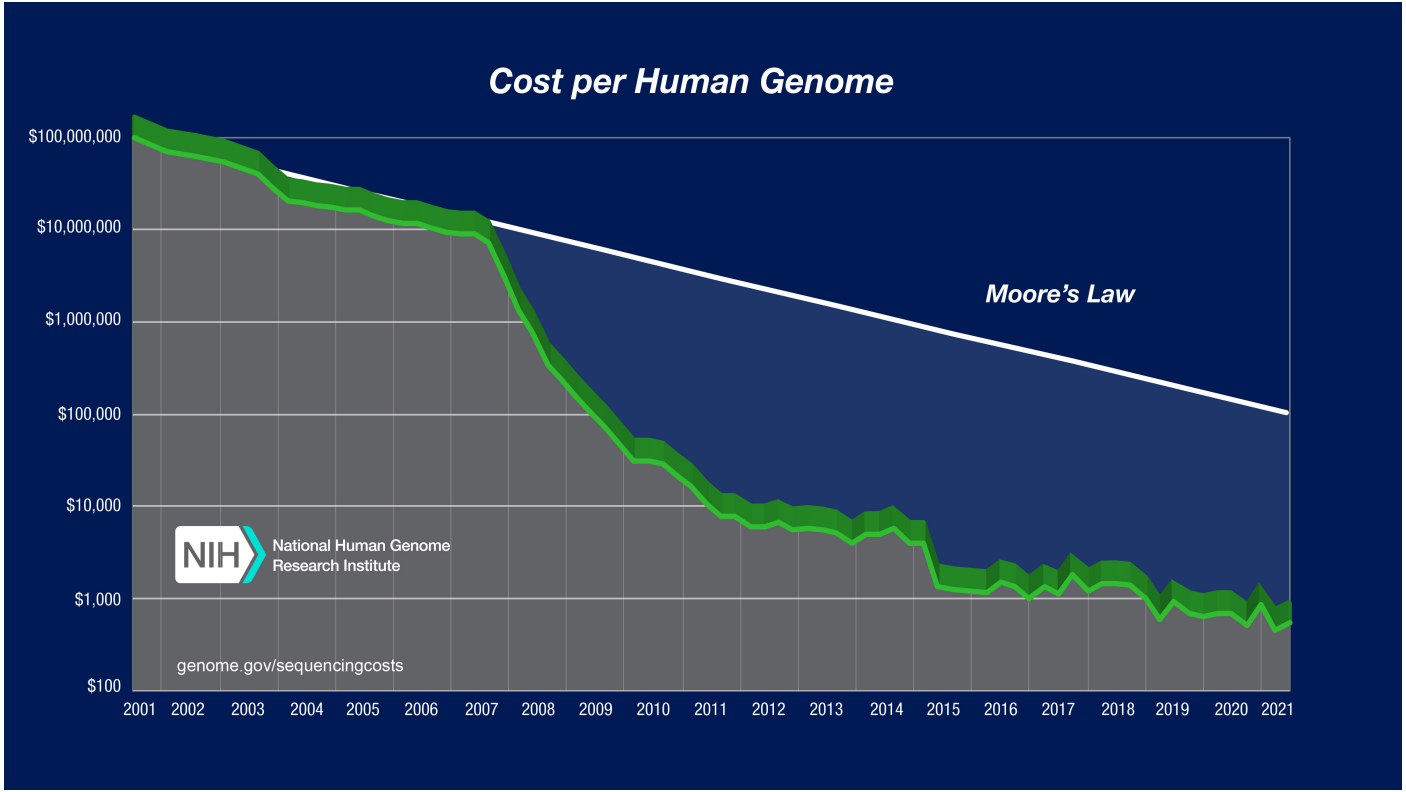
4. EHRs, Biobanks, and Genetics



How many protein sequence altering variants does your genome carry?

1000 Genomes Project; gnomAD

Genome sequencing is “cheap”



How many protein sequence-altering variants does your genome carry?

- 10,000–12,000 protein altering variants
- 50–100 protein truncating variants

1000 Genomes Project; gnomAD

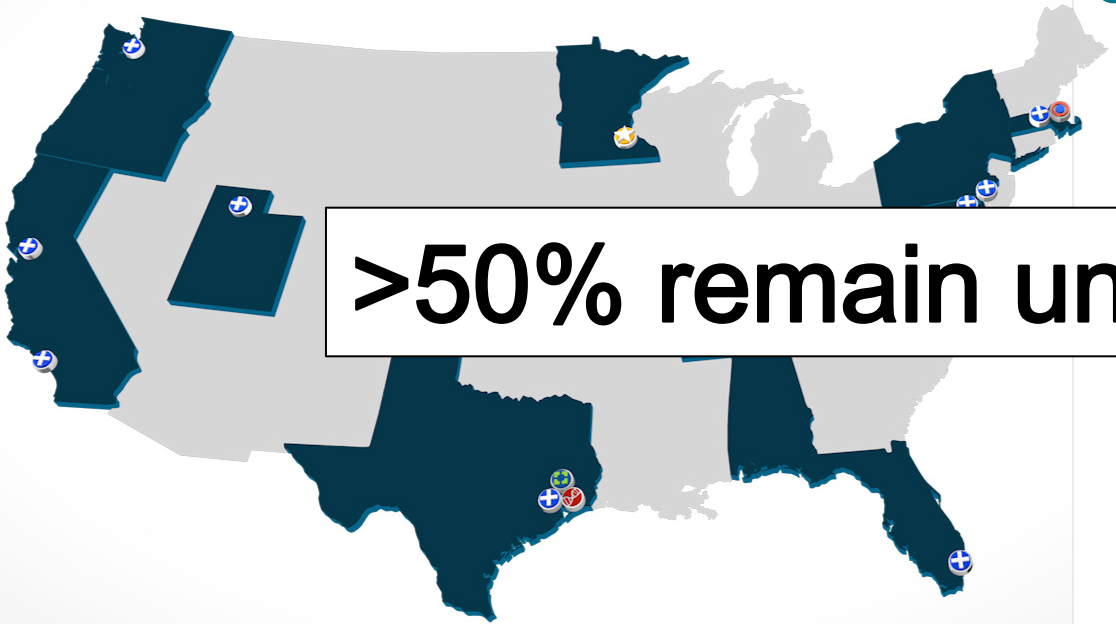
Each of us carries thousands of rare coding germline variants.

- Most of these variants have not been functionally characterized
  - aka Variants of Unknown Significance (VUS)
- Most of these variants are benign.

How do we distinguish these from causal rare disease variants?

Undiagnosed Diseases Network

Solving medical mysteries through team science

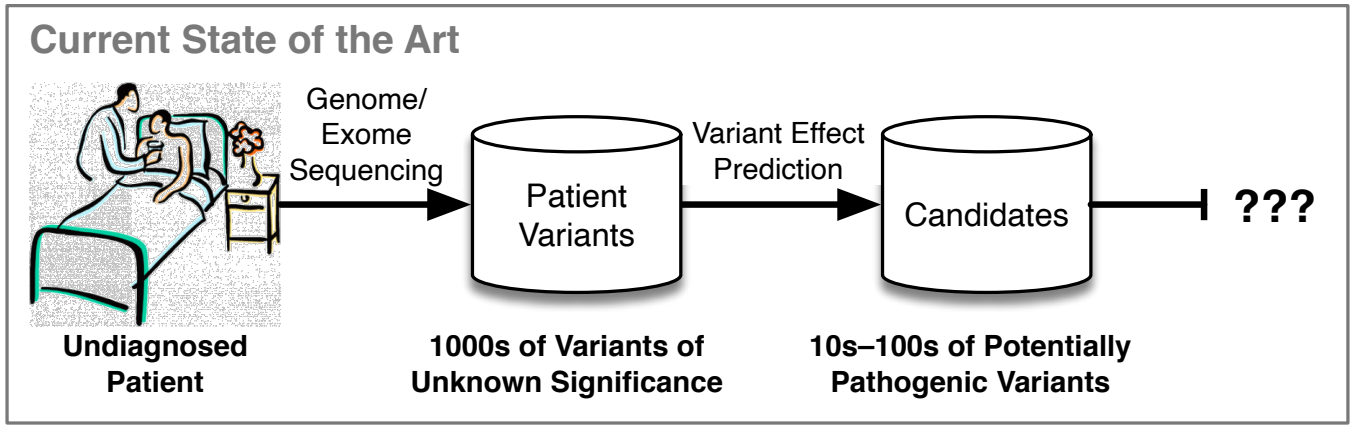


>50% remain unsolved

- **Baylor College of Medicine** — Houston, TX
- **Baylor College of Medicine and University of Oregon** — Houston, TX
- **Brigham and Women's Hospital, Boston Children's Hospital, Massachusetts General Hospital** — Boston, MA
- **Duke University** — Durham, NC
- **Harvard Medical School and University of Alabama at Birmingham** — Boston, MA
- **Mayo Clinic** — Rochester, MN
- **National Institutes of Health** — Bethesda, MD
- **Stanford Medicine** — Stanford, CA
- **UCLA** — Los Angeles, CA
- **University of Miami School of Medicine** — Miami, FL
- **University of Utah** — Salt Lake City, UT
- **University of Washington School of Medicine and Seattle Children's Hospital** — Seattle, WA
- **Vanderbilt University Medical Center** — Nashville, TN
- **Washington University in St. Louis** — St. Louis, MO

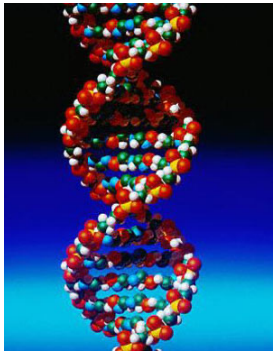
Clinical site · Coordinating center · DNA sequencing core · Central Biorepository · Model Organisms Screening Center · Metabolomics Core

Twelve clinical sites, a coordinating center, a sequencing core, a metabolomics core, two model organisms screening centers, and a central biorepository



How can **we** help distinguish the 1000s of rare benign variants from causal rare disease variants?

Central Dogma of Molecular Biology



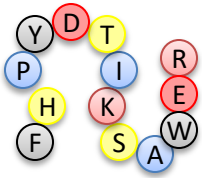
DNA

Function

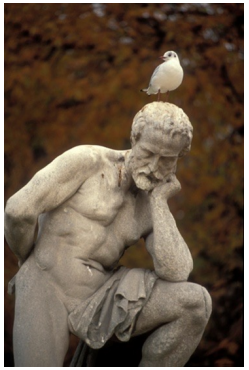
Transcription

RNA

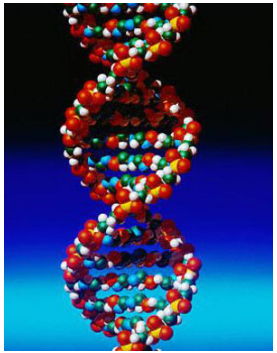
Translation



Protein



Central Dogma of Structural Biology



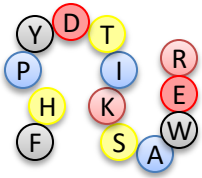
DNA

Function

Transcription

RNA

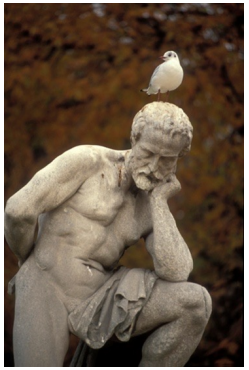
Translation

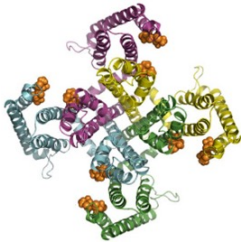


Protein

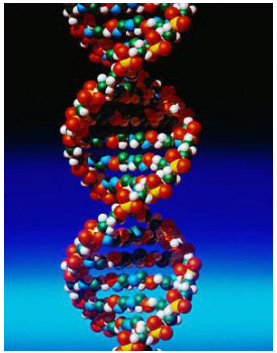
Folding

Structure



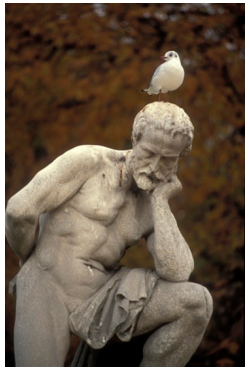


Central Dogma of Structural Biology



DNA

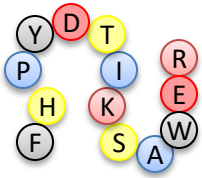
Function



Transcription

RNA

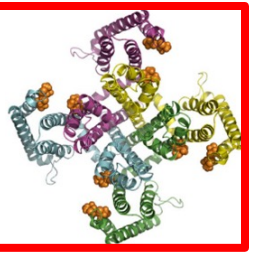
Translation



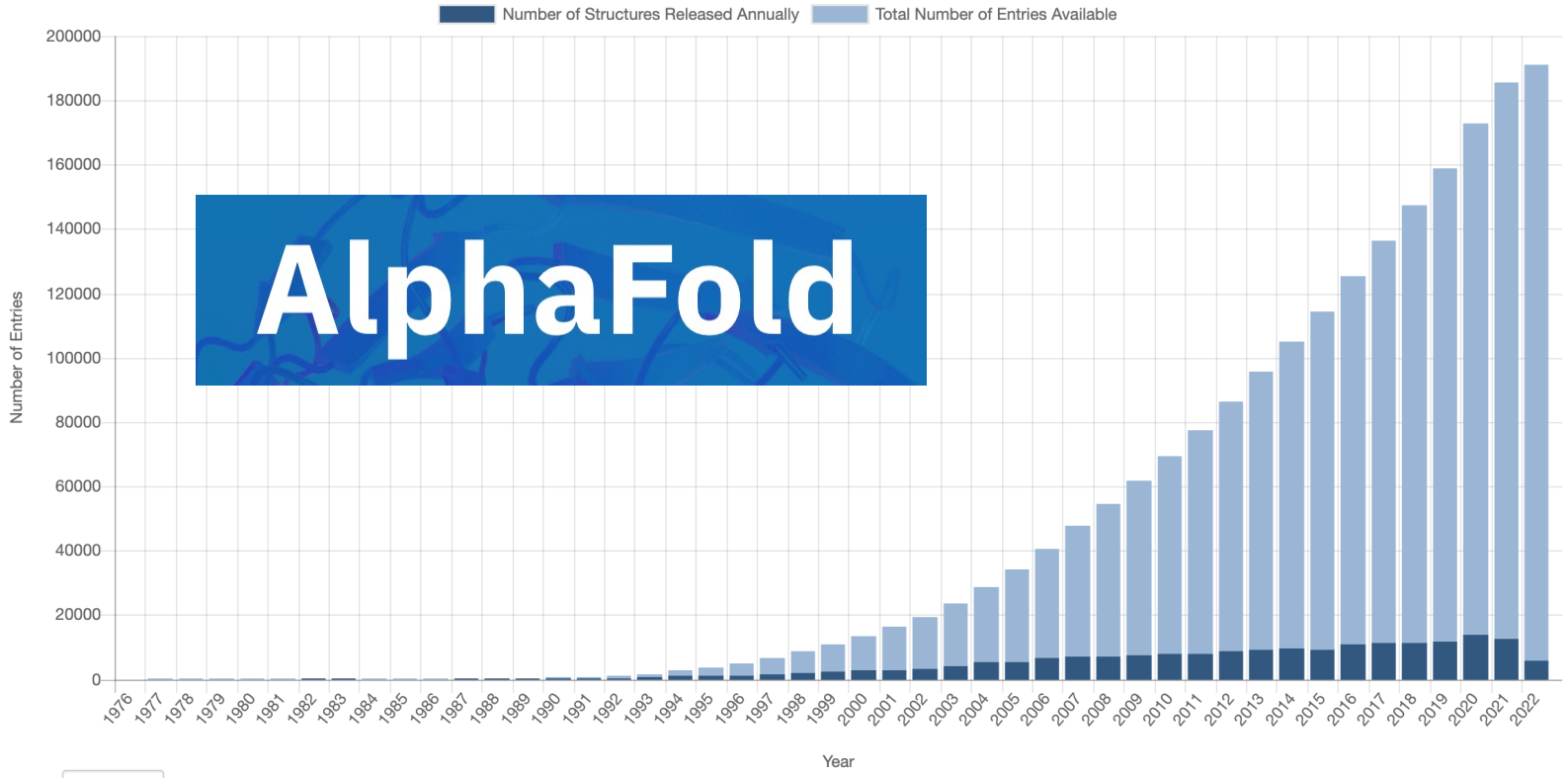
Protein

Folding

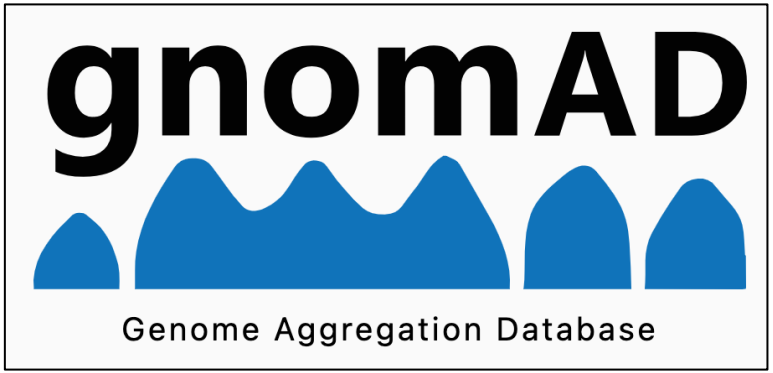
Structure



>200,000 3D structural models

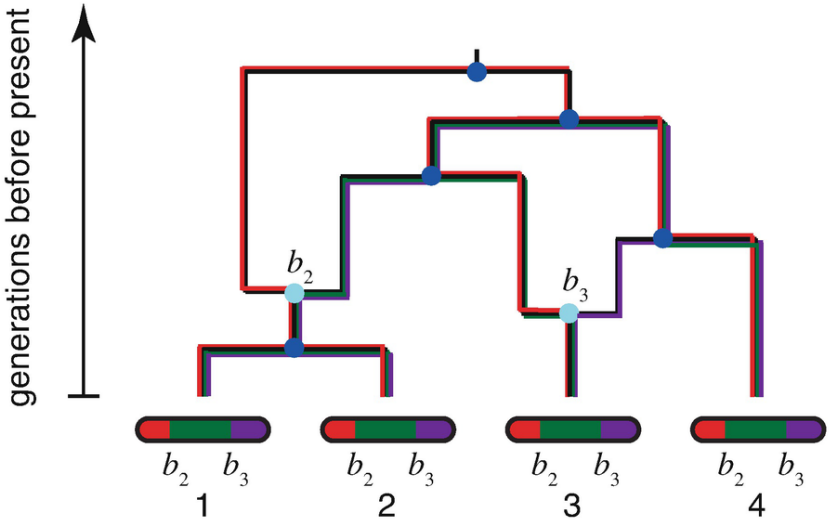


>200 million human protein variants



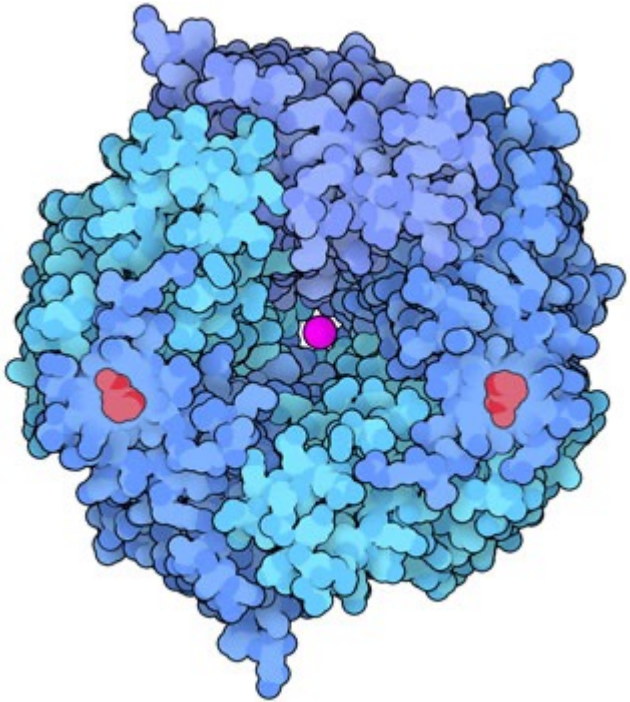
>140,000 exomes and genomes

Opportunity!



Population Genetics

+



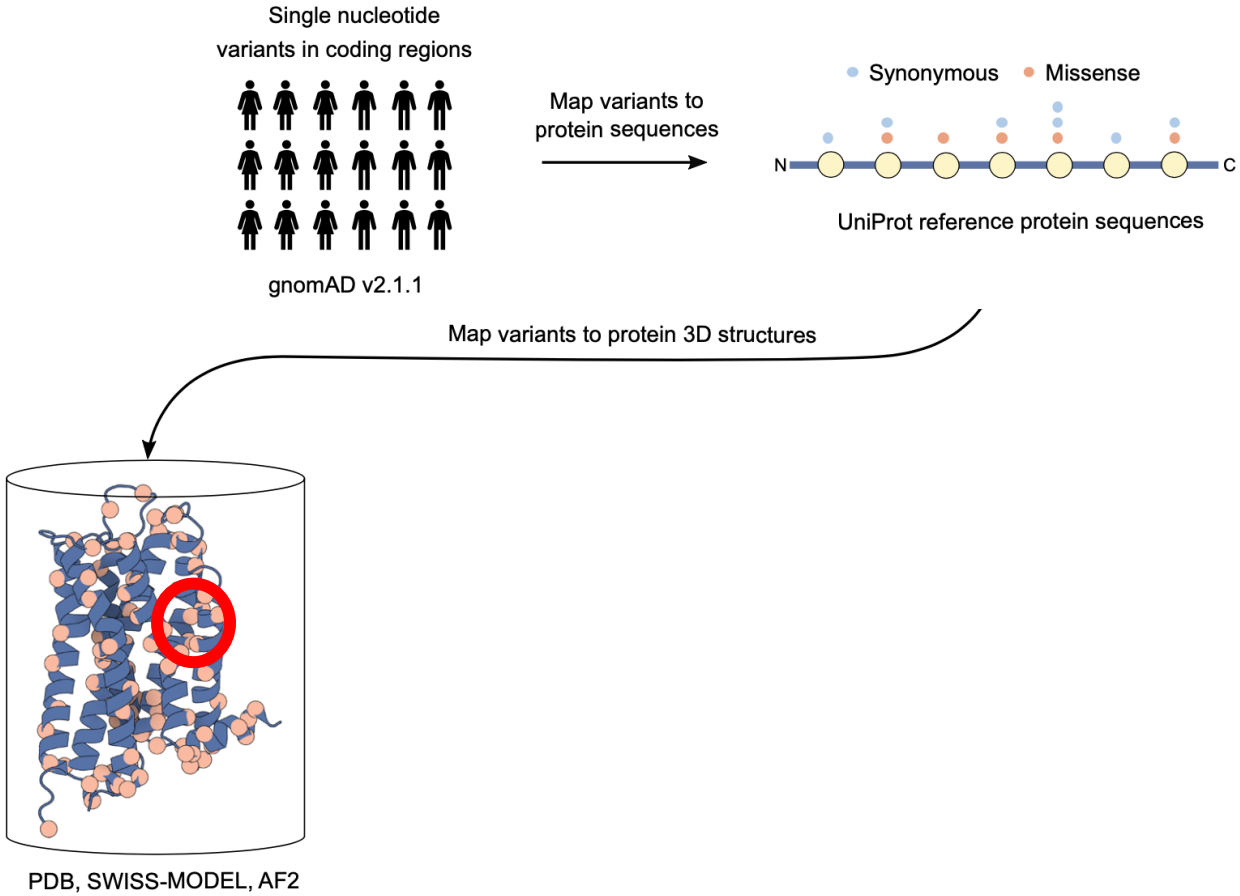
Structural Biology

Which spatial regions of human proteins tolerate genetic variation?

Intuition:

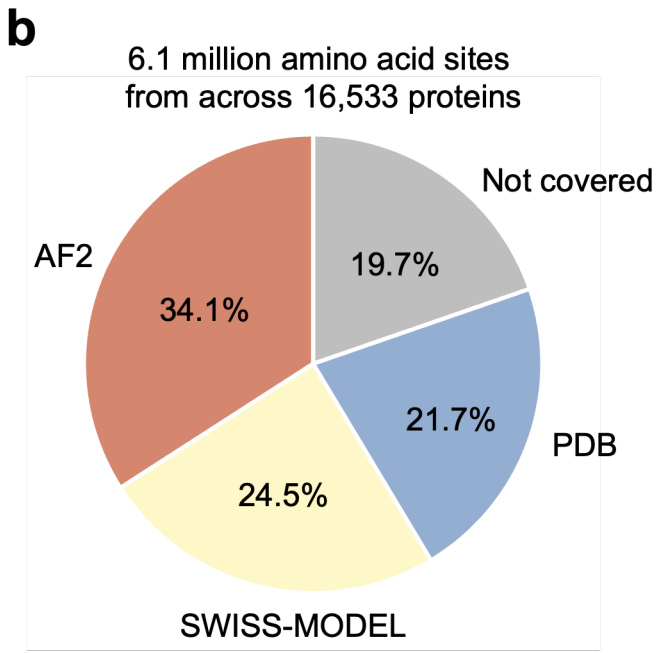
Spatial regions that do not vary in large healthy populations are functionally constrained.

Variants in these regions are candidates for disease.



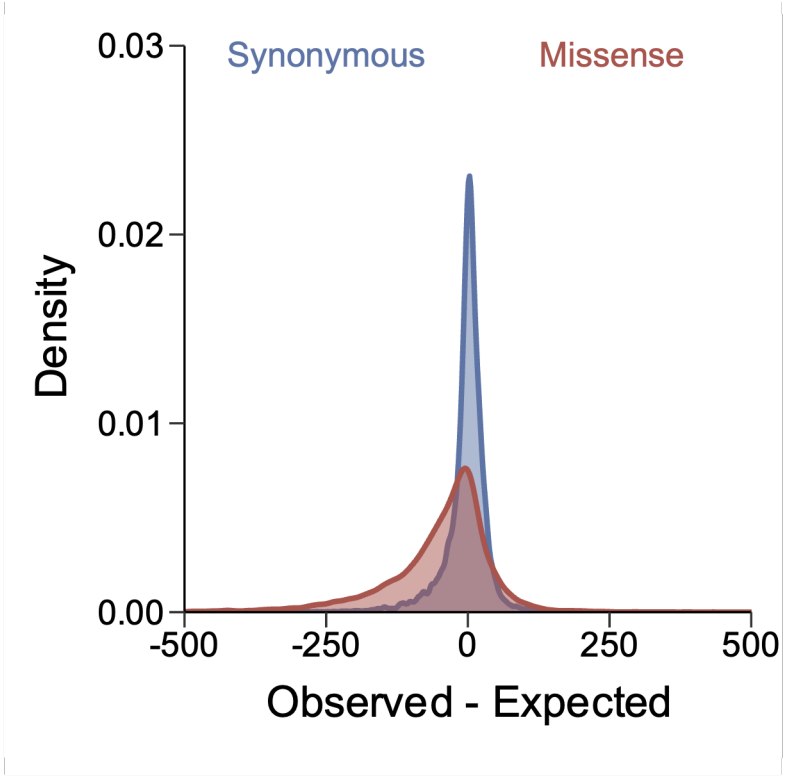
Li et al. 2022; https://github.com/CapraLab/cosmis

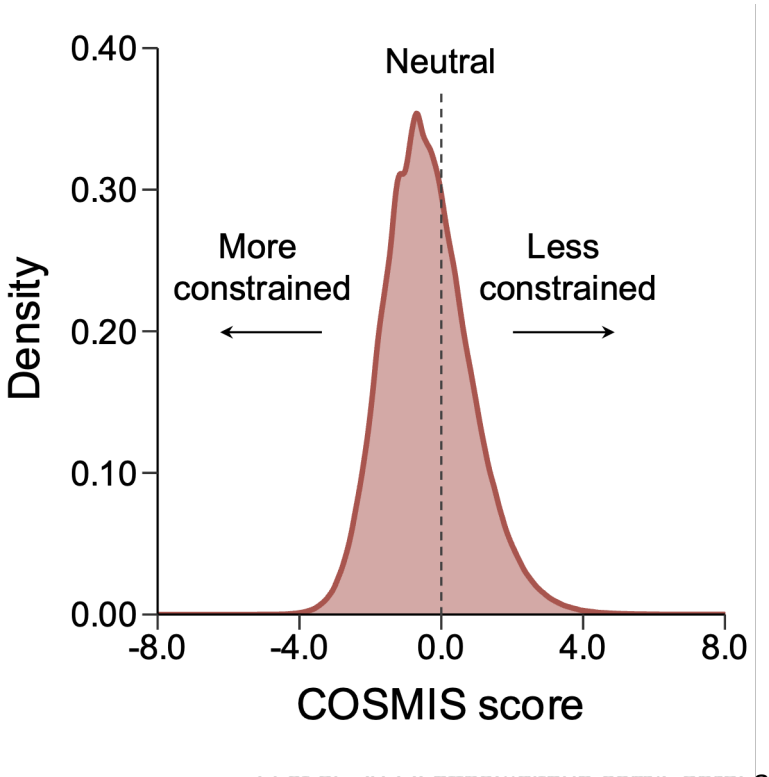
Alphfold substantially increased our coverage of protein space



Li et al. 2022; https://github.com/CapraLab/cosmis

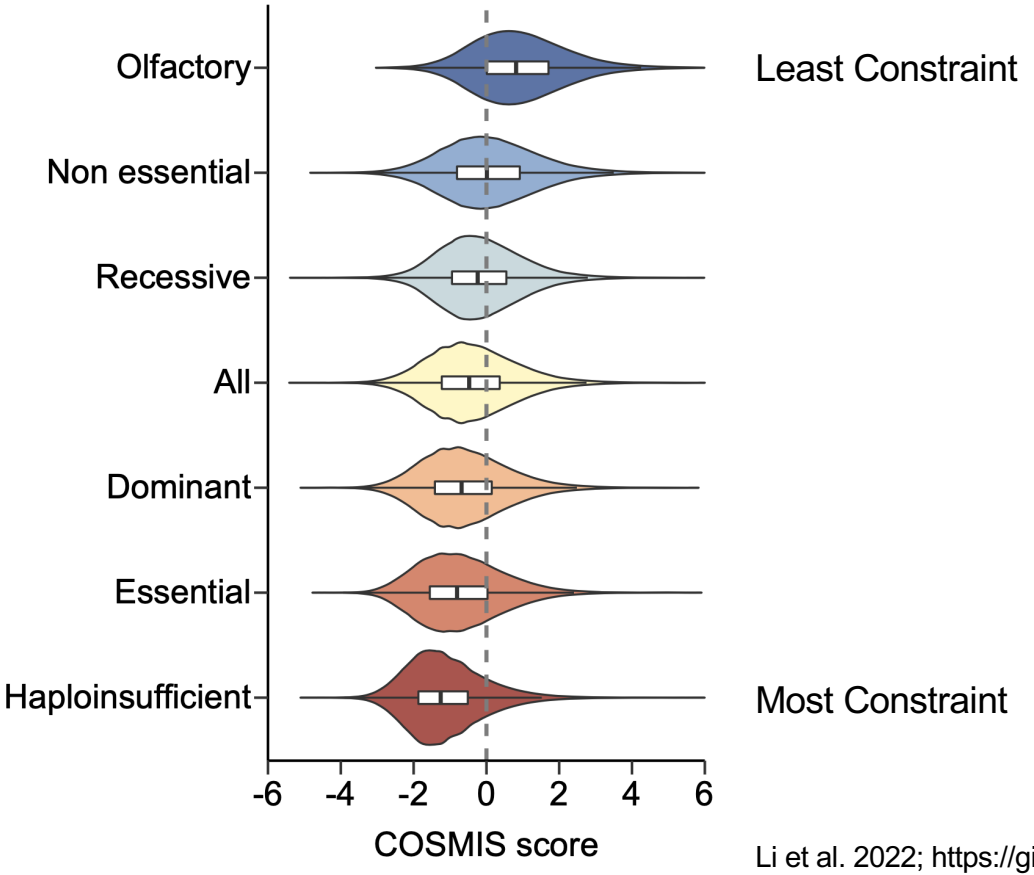
COSMIS quantifies depletion of missense variants in contact sets





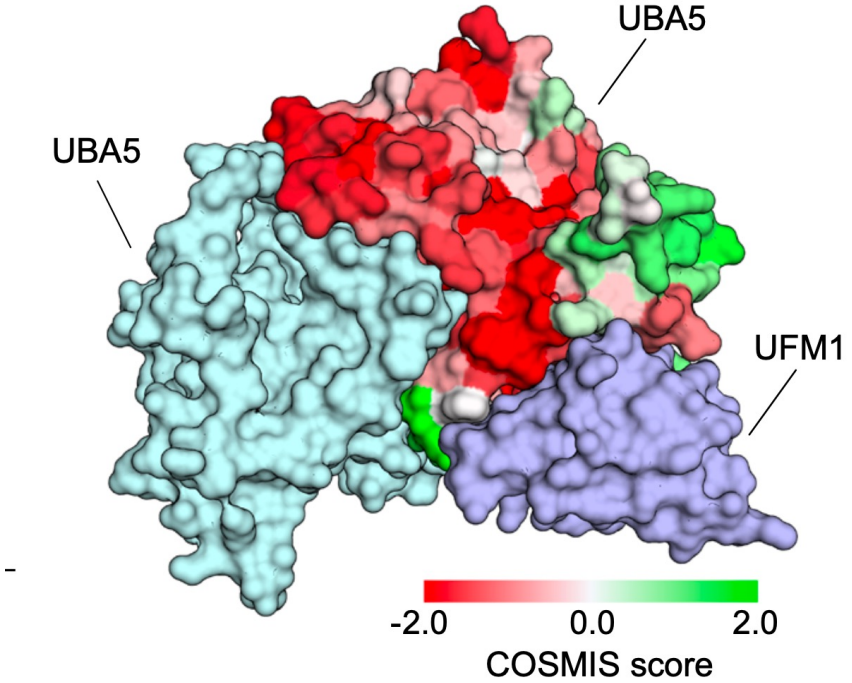
Li et al. 2022; https://github.com/CapraLab/cosmis

A range of mutational constraint…

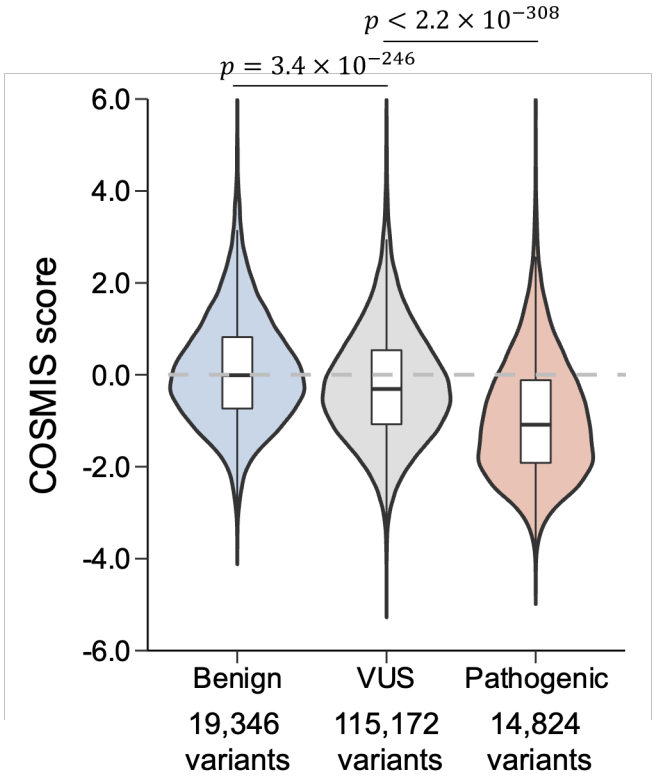


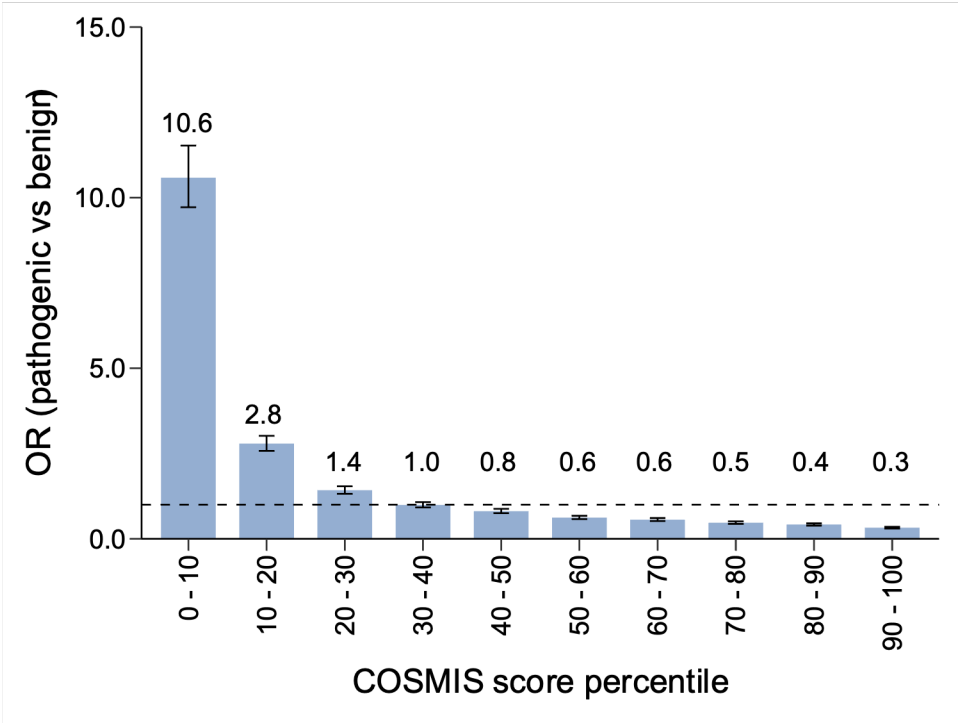
Li et al. 2022; https://github.com/CapraLab/cosmis

Which spatial regions are constrained?



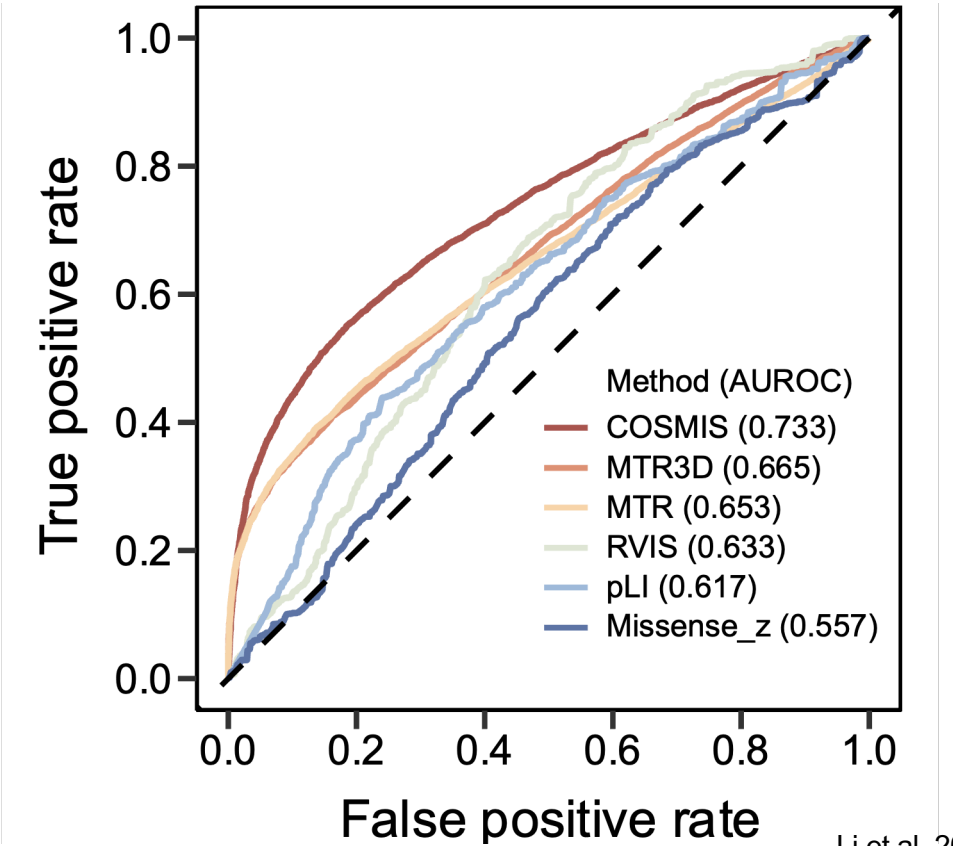
COSMIS strongly predicts pathogenicity





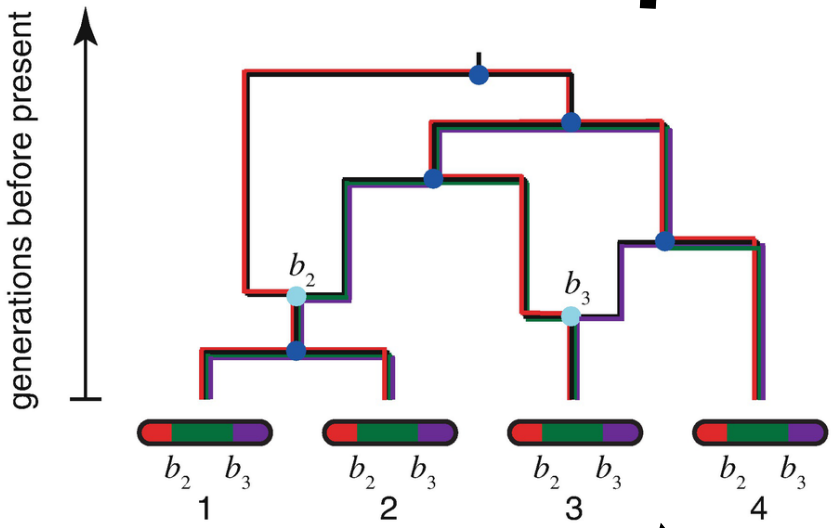
Li et al. 2022; https://github.com/CapraLab/cosmis

COSMIS outperforms other constraint metrics

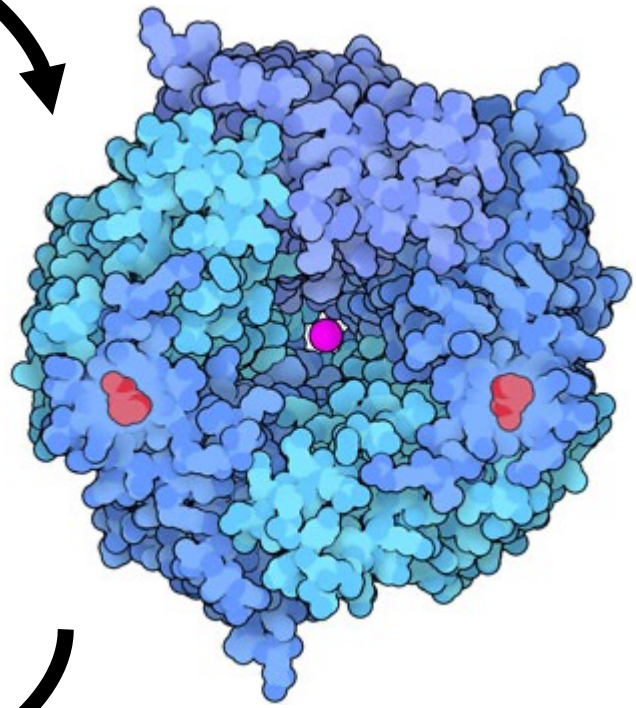


Li et al. 2022; https://github.com/CapraLab/cosmis

Opportunity!



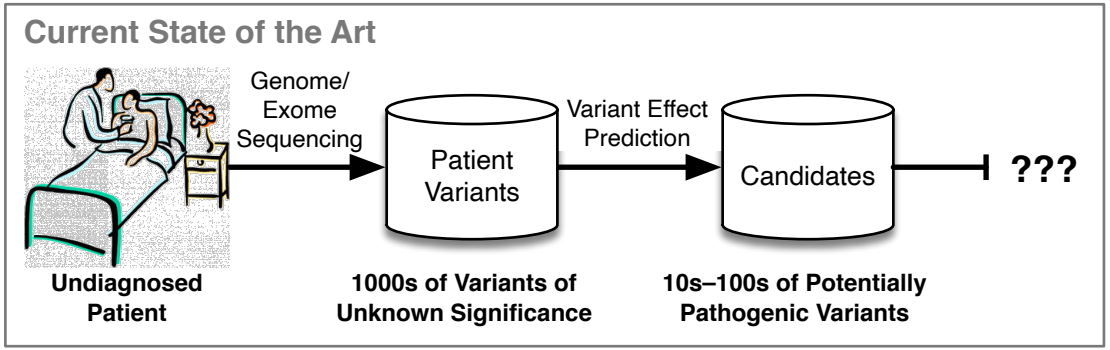
+



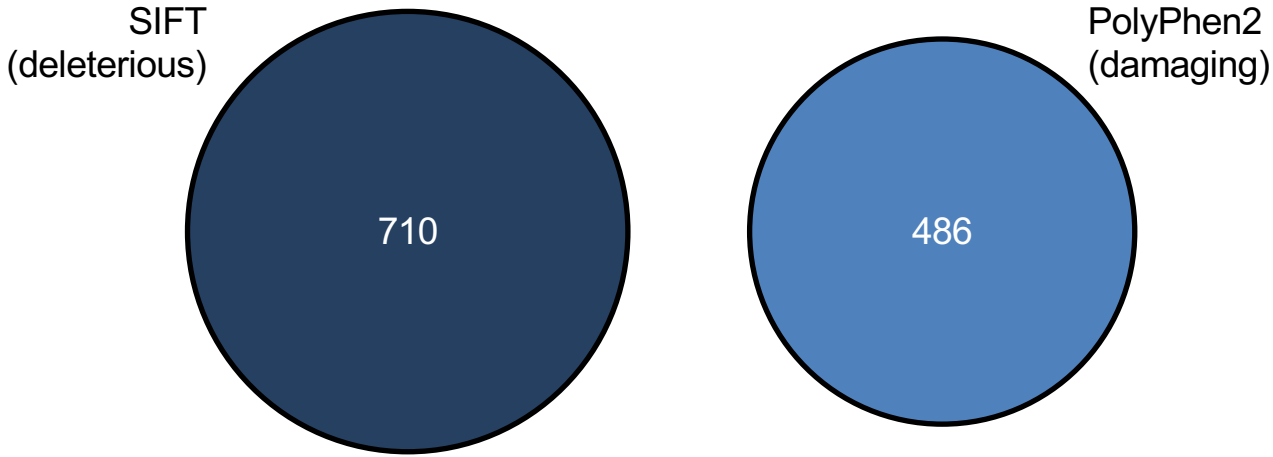
Population Genetics

Structural Biology

What can we do to help?

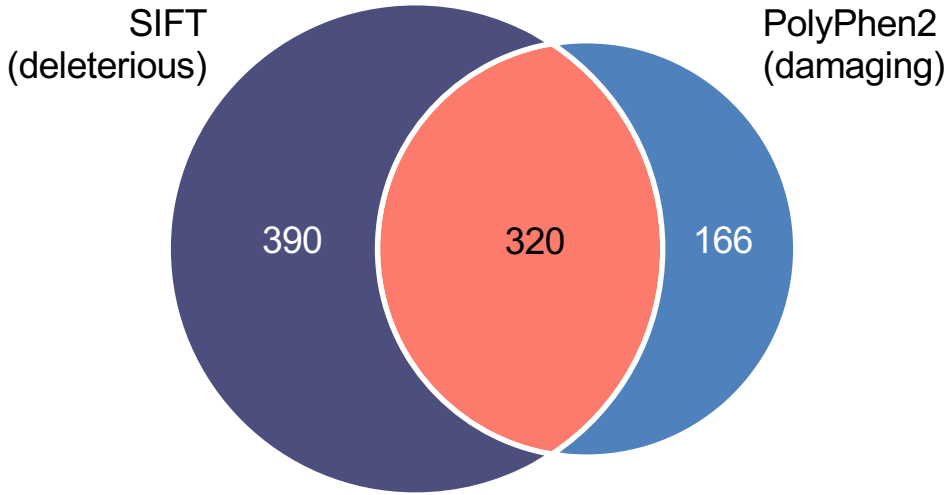


Variant Effect Predictions Disagree



Analyzed 1400 VUS

Variant Effect Predictions Disagree



Only 44% agree!

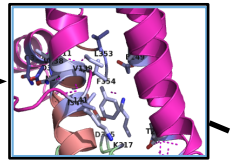
Variant Effect Prediction Challenges

- Methods frequently disagree
- May not be applicable across human populations
- Do not provide mechanistic justification for predictions

GOAL: Evaluate functional effects in a less biased / more interpretable way.

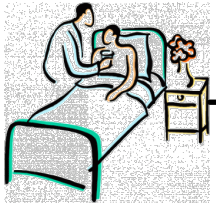
Hypothesis I:
3D Structure Disruption

Protein Structures



Actionable Hypotheses about Variant Effects

Current State of the Art



Genome/ Exome Sequencing

Patient Variants

Variant Effect Prediction

Candidates

???

Undiagnosed Patient

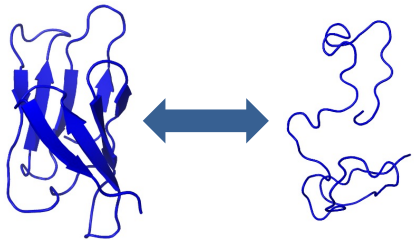
1000s of Variants of Unknown Significance

10s–100s of Potentially Pathogenic Variants

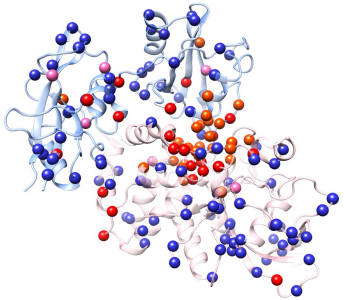
Treatment Decisions

Undiagnosed Diseases Network

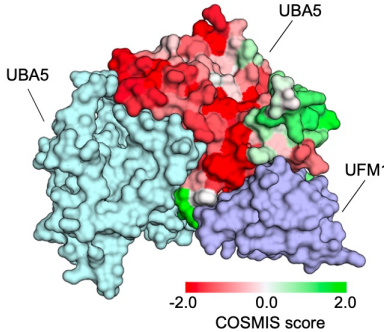
Tools for computational interpretation of structural effects of patient variants



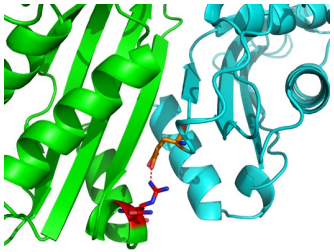
Stability



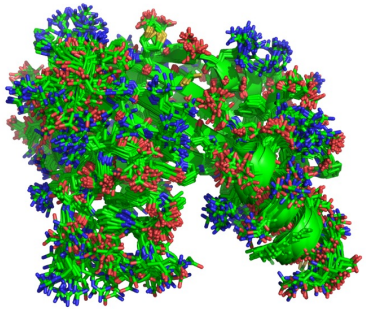
Variant Clustering



Evolutionary Constraint

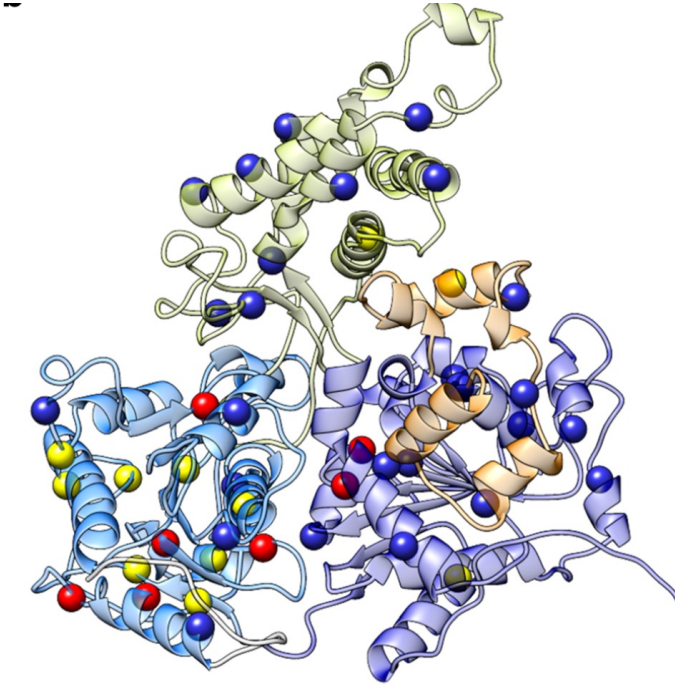


Interactions



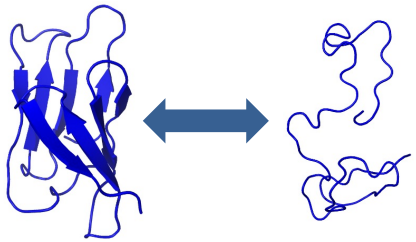
Dynamics

Pathogenic Proximity (PathProx)

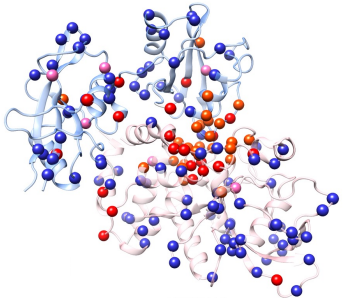


Sivley et al. AJHG 2018; Sivley et al. BMC Bioinformatics 2018

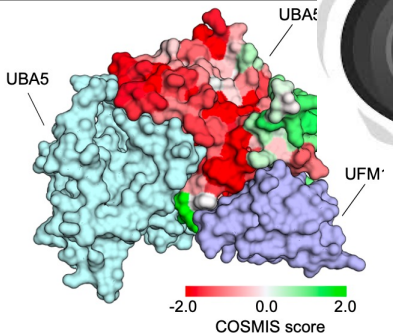
Tools for computational interpretation of structural effects of patient variants



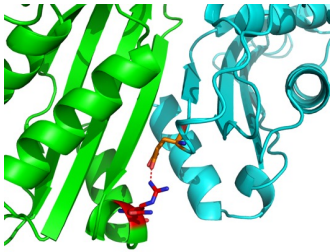
Stability



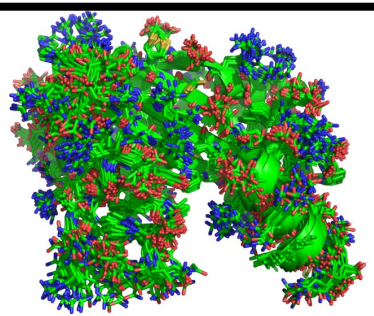
Variant Clustering



Evolutionary Constraint



Interactions



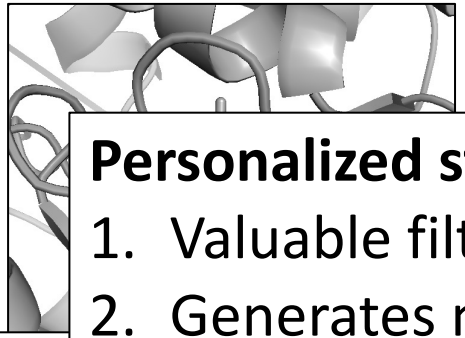
Dynamics

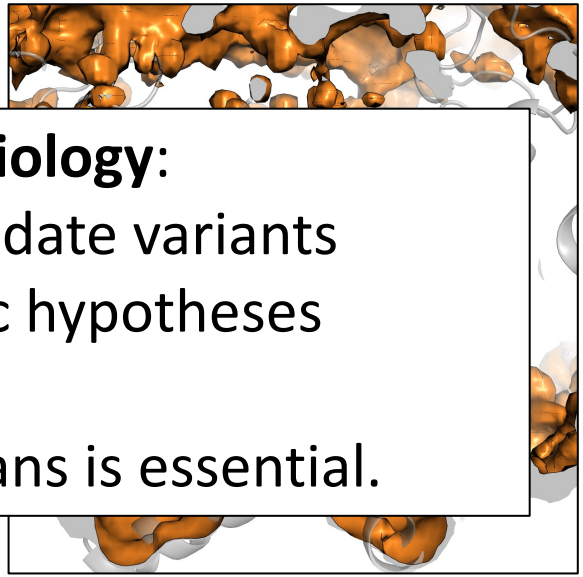
Incorporating 3D structure improves rare var interpretation

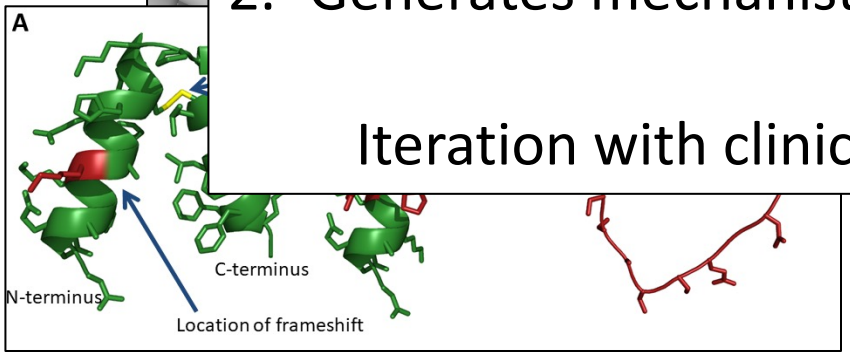
Jens Meiler PhD

Jonathan Sheehan PhD

Chris Moth PhD







Personalized structural biology:

1. Valuable filter of candidate variants
2. Generates mechanistic hypotheses

Iteration with clinicians is essential.

>100 patients analyzed — **>2100** variants considered

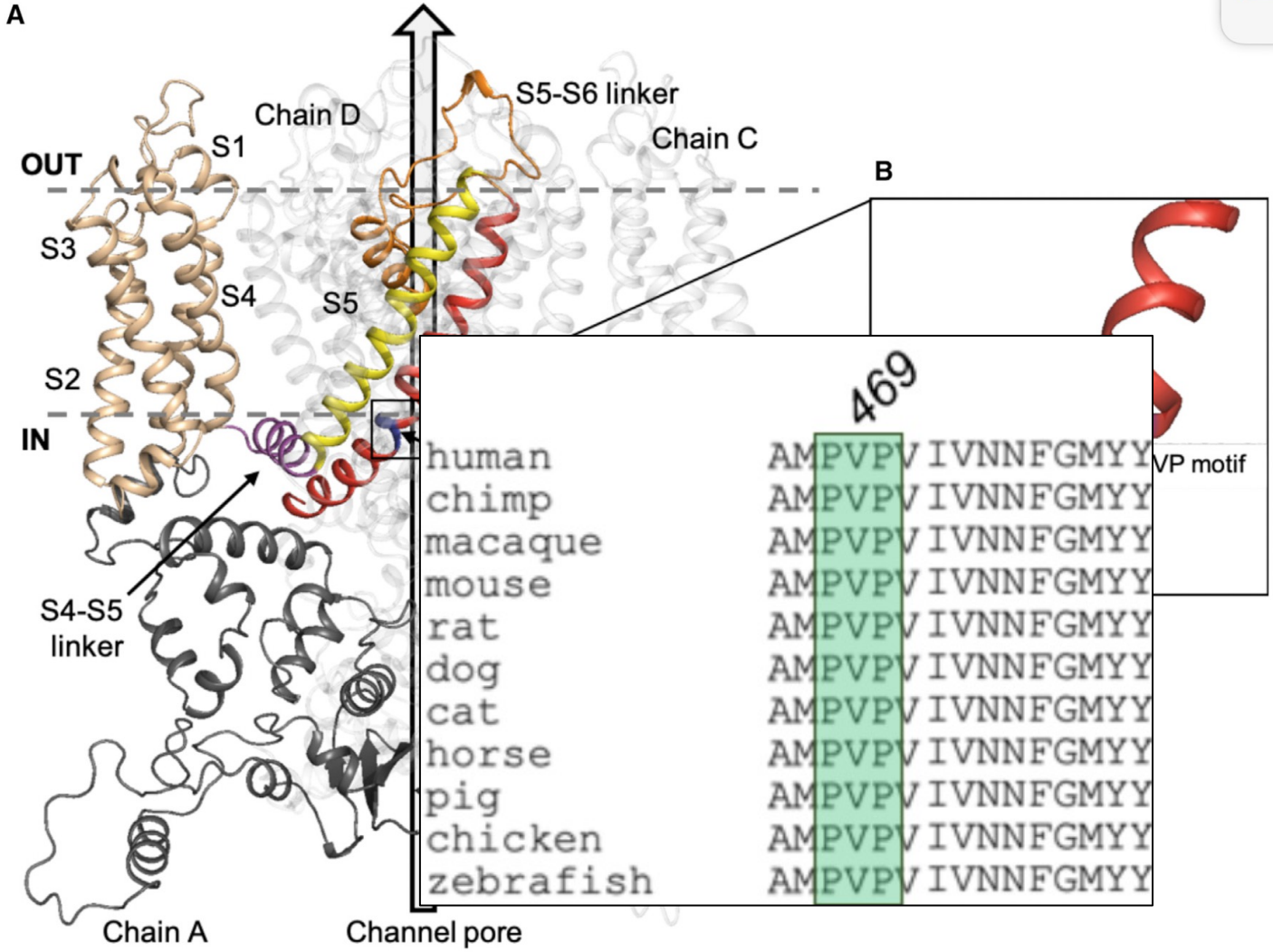
UDN Patient with DEE-like Symptoms

A 4-year-old boy at the Vanderbilt University UDN site presented with DEE-like phenotypes, including multiple types of refractory seizure and global developmental delay. Around 18 months of age, he developed generalized tonic clonic seizures, and was **diagnosed with Lennox-Gastaut syndrome**, a severe form of DEE. ***However, he continued to have frequent myoclonic absence seizures and occasional generalized tonic clonic seizure.***

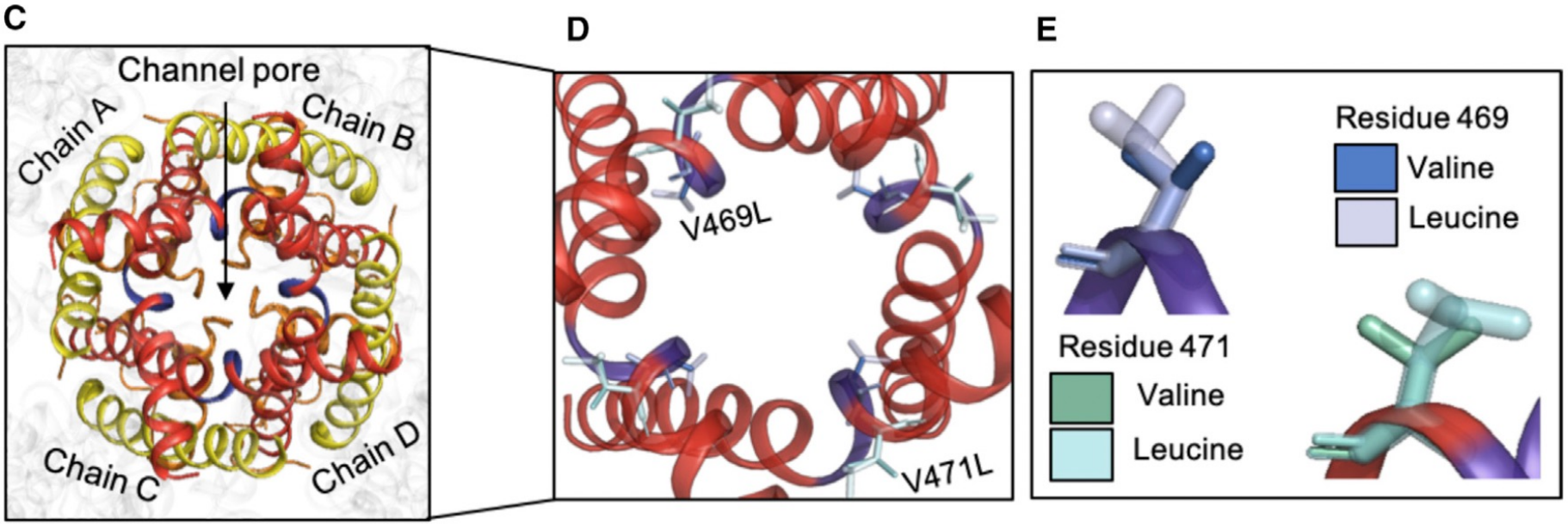
Negative on Athena epilepsy gene panel

WGS reveals KCNC2 p.V469

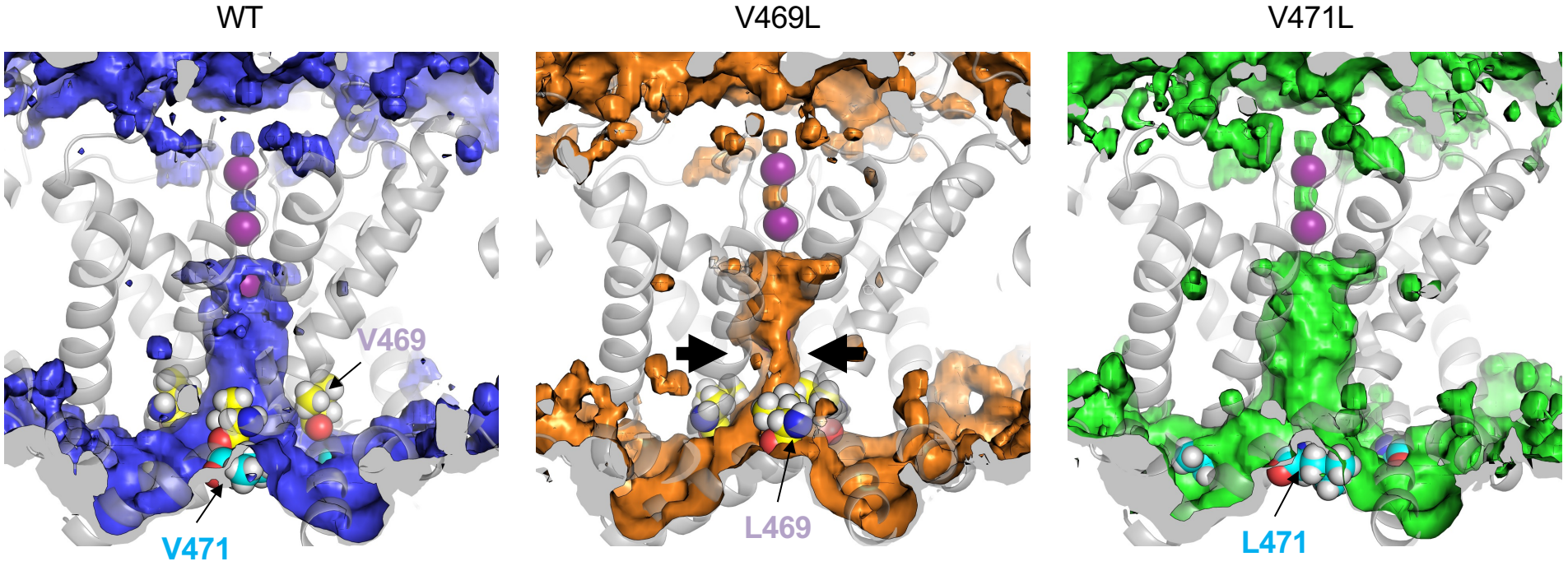
- homo-tetrameric voltage-gated potassium channel Kv3.2
- highly expressed in GABAergic interneurons in the CNS



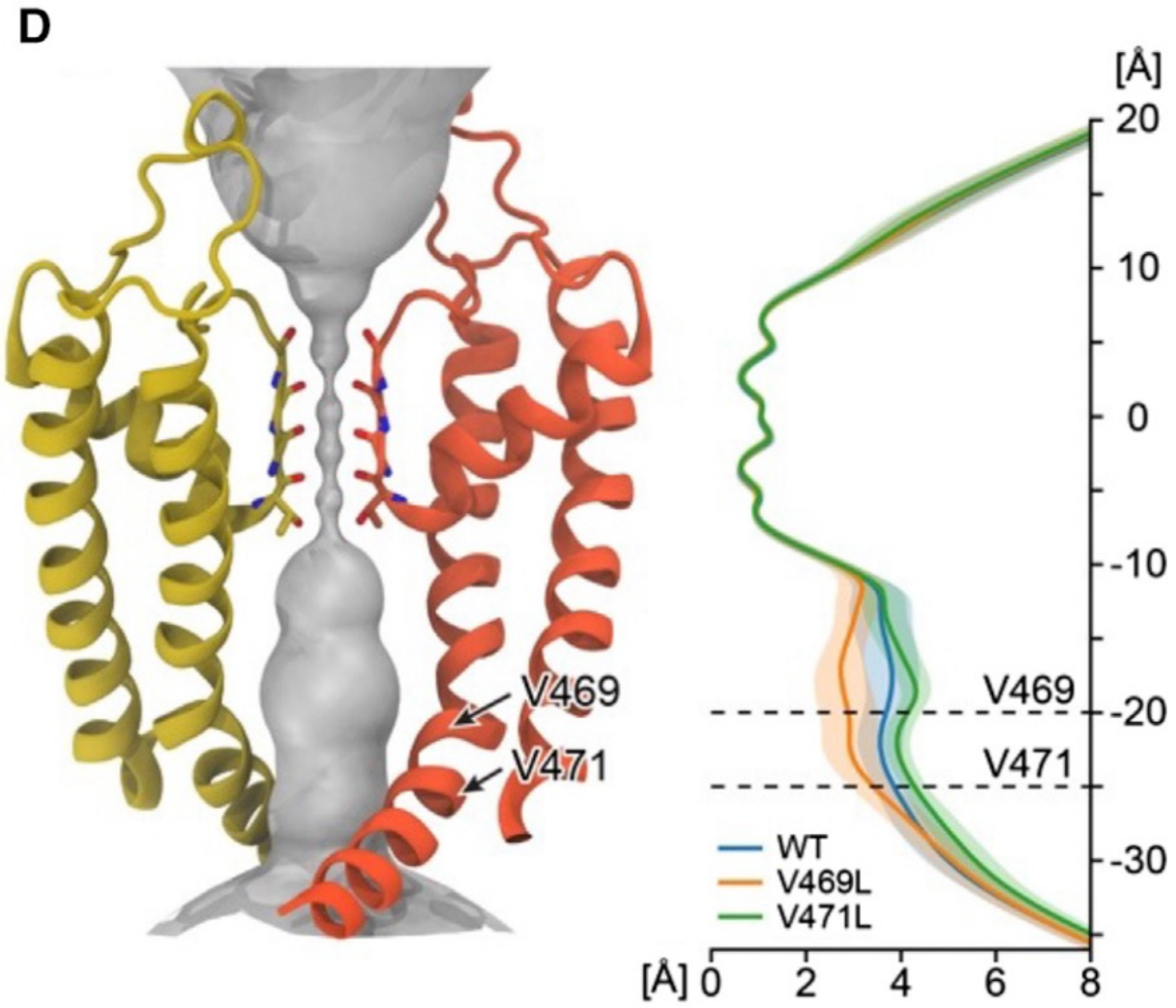
Leucine at position 469 is bulkier than native residue Valine and could block channel pore

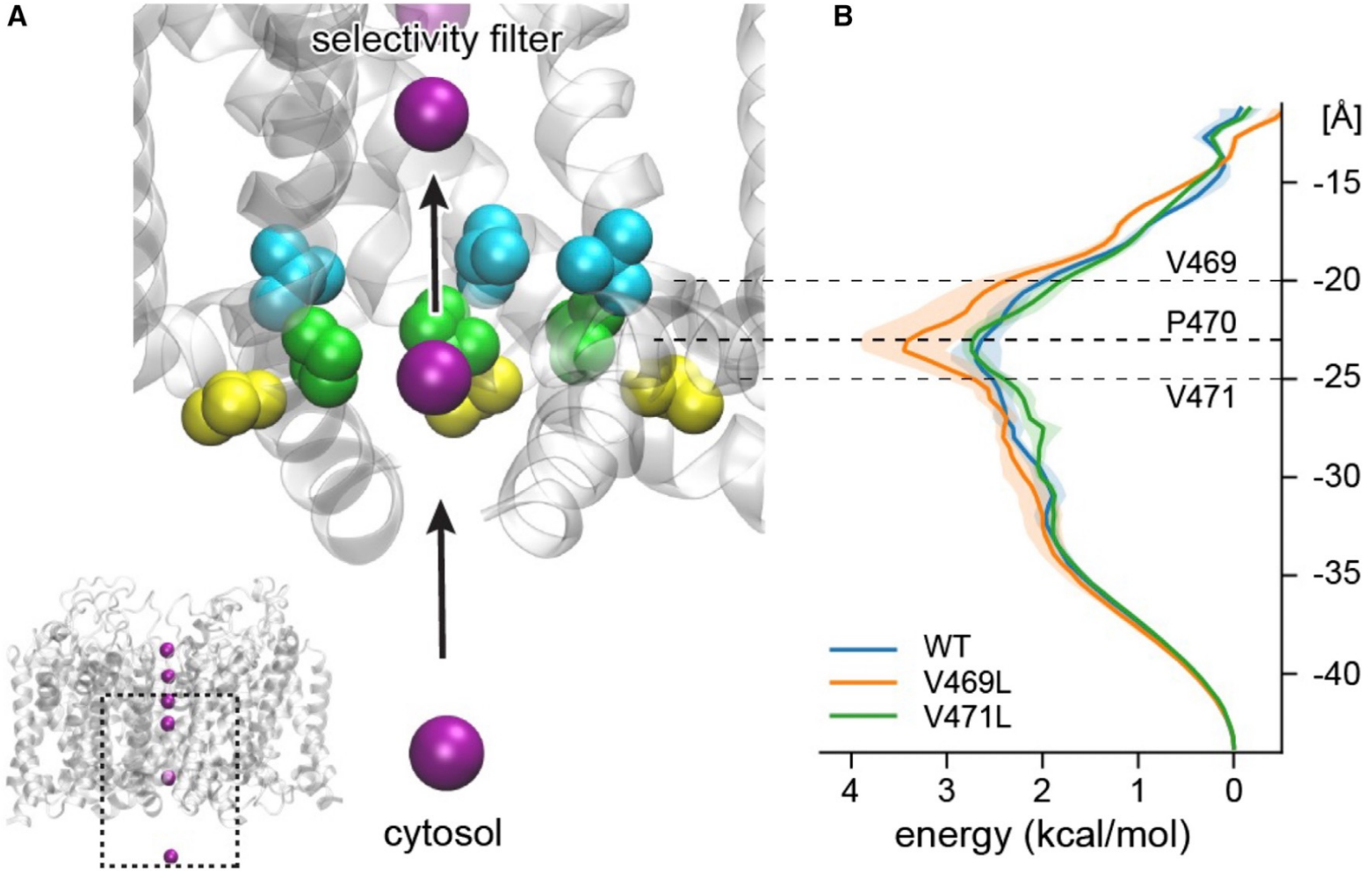


Conventional MD simulations reveal decrease in pore radius for V469L



- Isosurface representation of the average spatial density of water in MD simulations of WT (blue), V469L (orange), and V471L (green).
- Constriction and dewetting of the inner cavity observed in MD simulations of V469L.
- L471 appears to stabilize S6-S6 inter-subunit interactions which help to keep the inner channel gate open.


D
V469
V471
[Å]
20
10
0
-10
-20
-30
V469
V471
WT
V469L
V471L
[Å] 0
2
4
6
8


A
selectivity filter
cytosol
B
[Å]
-15
-20
-25
-30
-35
-40
V469
P470
V471
WT
V469L
V471L
4
3
2
1
0
energy (kcal/mol)

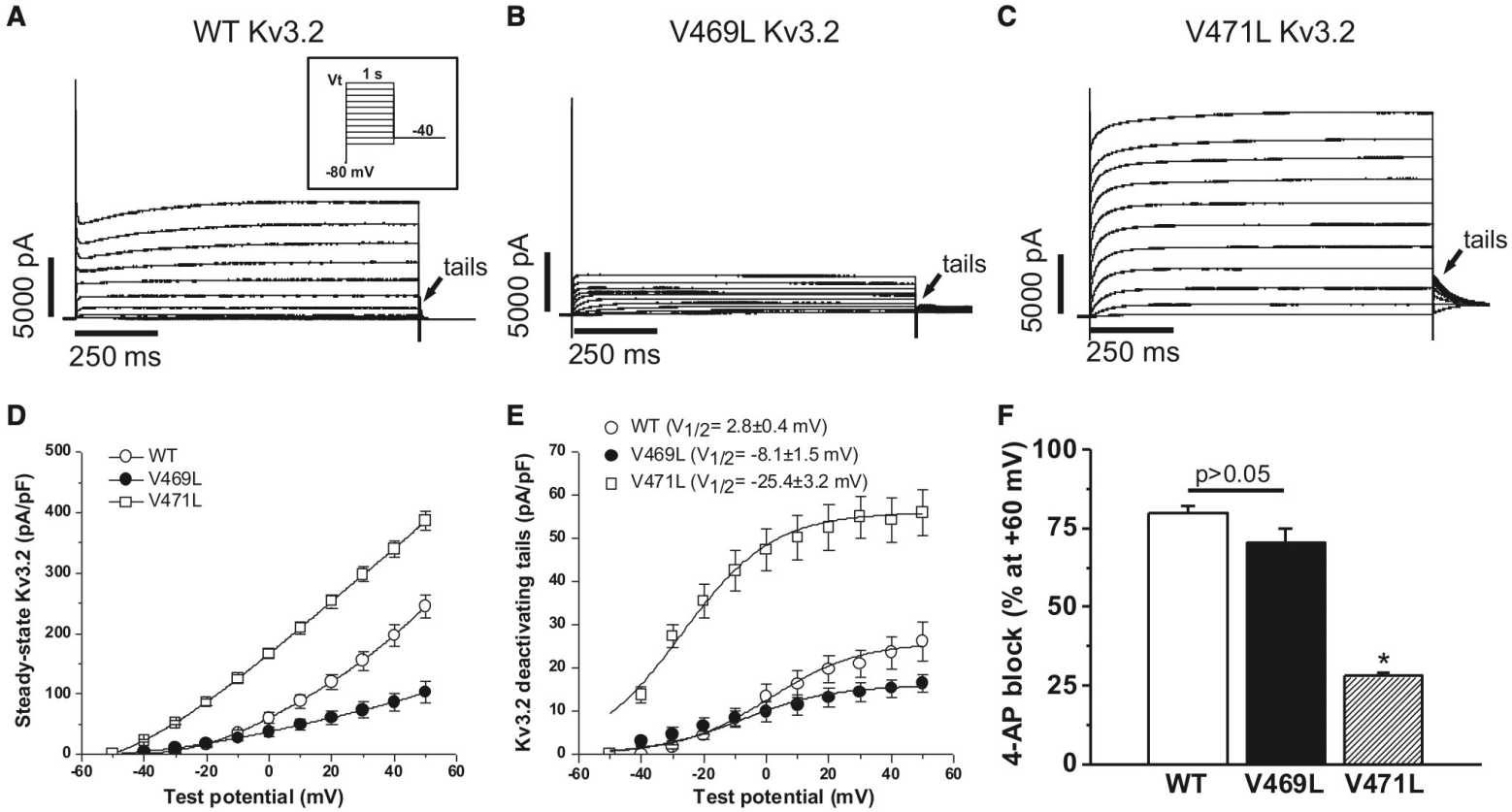
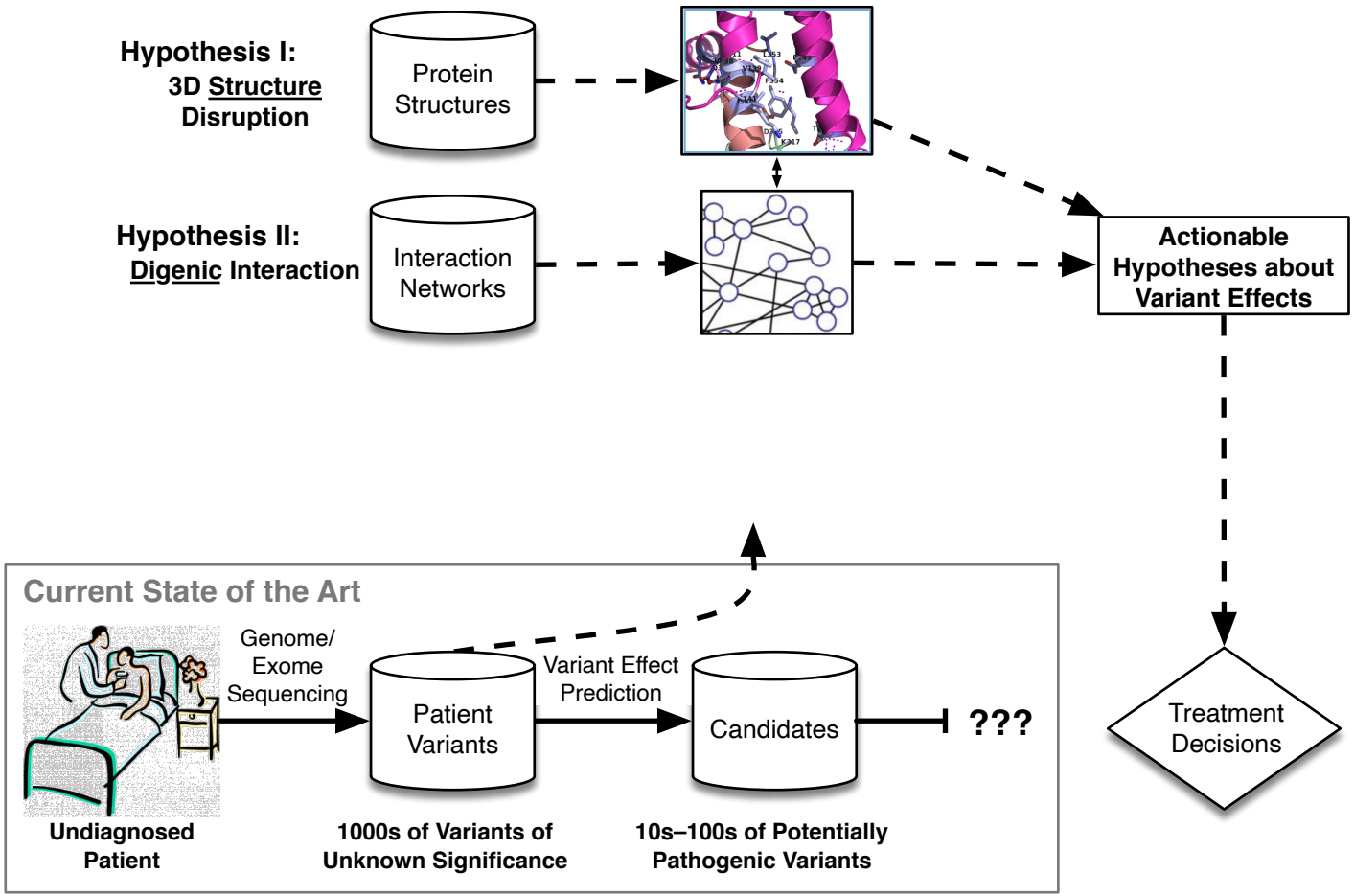
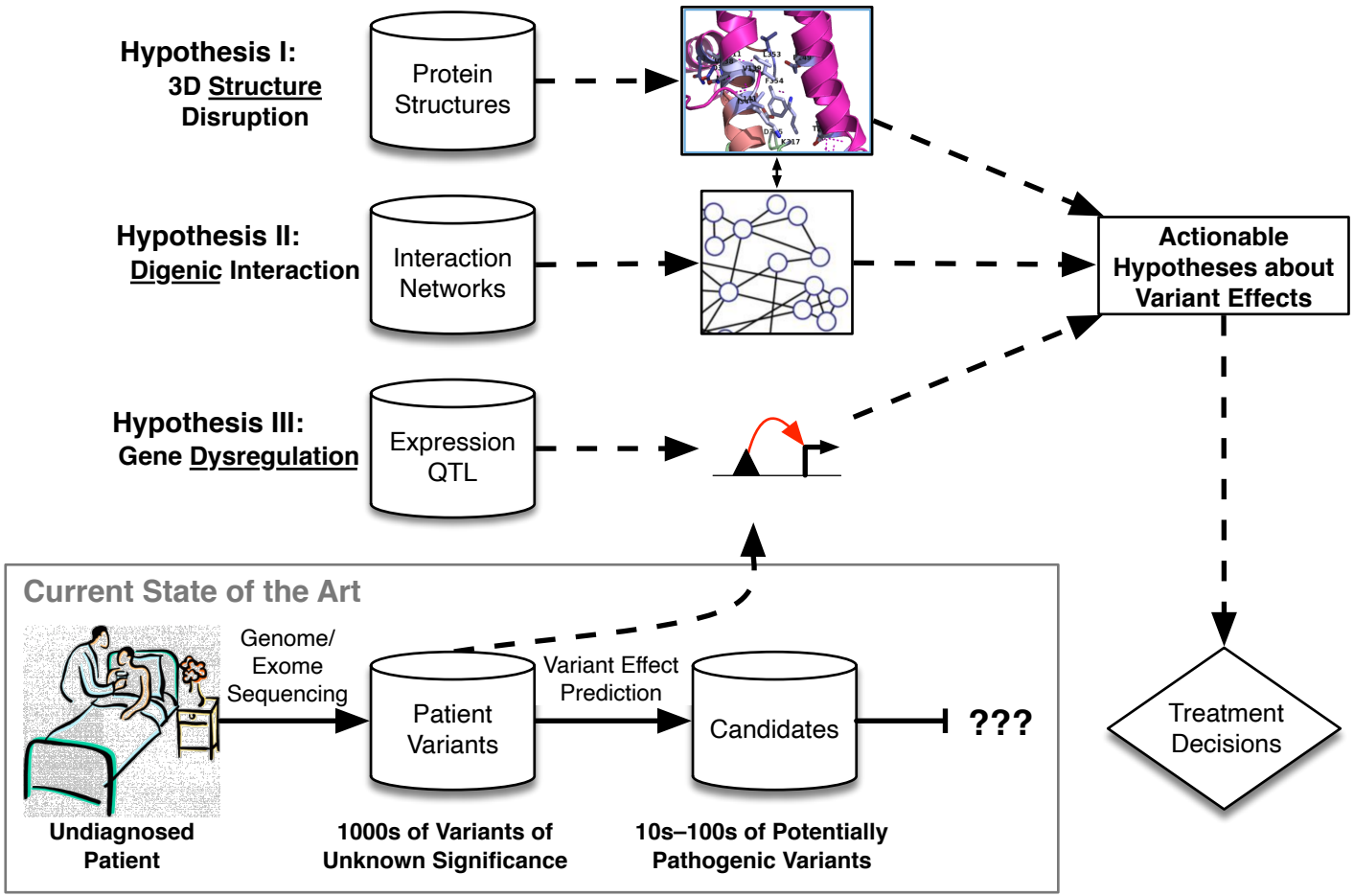


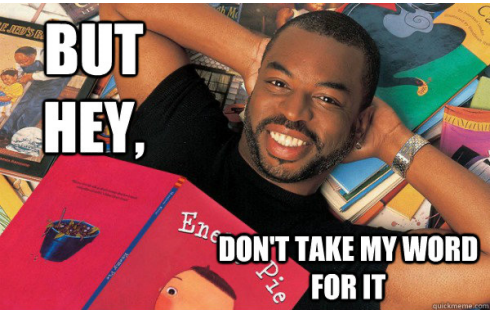
Figure 2. Candidate Kv3.2 variants cause loss and gain of channel function

KCNC2 Summary

- Protein structural modeling and MD simulations rationalize the mechanistic basis for the phenotypic heterogeneity of candidate variants
- Demonstrate heterogeneous loss-of-function and gain-of-function effects, despite both affecting the essential hinge region of Kv3.2
- Validate links between KCNC2 and heterogeneous DEE phenotypes
- Blueprint for integrating genetics, protein structural modeling, and experimental validation to develop mechanistic understanding of the molecular effects of *de novo* variants in rare disease.


Hypothesis I:
3D Structure Disruption
Protein Structures
Hypothesis II:
Digenic Interaction
Interaction Networks
Actionable Hypotheses about Variant Effects
Treatment Decisions
Current State of the Art
Genome/ Exome Sequencing
Patient Variants
Variant Effect Prediction
Candidates
???
Undiagnosed Patient
1000s of Variants of Unknown Significance
10s–100s of Potentially Pathogenic Variants


Hypothesis I:
3D Structure Disruption
Protein Structures
Hypothesis II:
Digenic Interaction
Interaction Networks
Hypothesis III:
Gene Dysregulation
Expression QTL
Actionable Hypotheses about Variant Effects
Treatment Decisions
Current State of the Art
Genome/ Exome Sequencing
Patient Variants
Variant Effect Prediction
Candidates
???
Undiagnosed Patient
1000s of Variants of Unknown Significance
10s–100s of Potentially Pathogenic Variants



https://doi.org/10.1038/s41467-022-30936-x OPEN

The 3D mutational constraint on amino acid sites in the human proteome

Bian Li[1,2✉], Dan M. Roden[2,3] & John A. Capra[1,4✉]

Do available protein 3D structures reflect human genetic and functional diversity?

Gregory Sliwoski, Neel Patel, R. Michael Sivley, Charles R. Sanders, Jens Meiler, William S. Bush, John A. Capra

Personalized structural biology reveals the molecular mechanisms underlying heterogeneous epileptic phenotypes caused by *de novo* KCNC2 variants

Souhrid Mukherjee,[1,6] Thomas A. Cassini,[11,17] Ningning Hu,[8,9,17] Tao Yang,[5,17] Bian Li,[1,5,6,17] Wangzhen Shen,[8] Christopher W. Moth,[6] David C. Rinker,[4,6] Jonathan H. Sheehan,[6,10] Joy D. Cogan,[2] Undiagnosed Diseases Network, John H. Newman,[3] Rizwan Hamid,[2] Robert L. Macdonald,[5,8] Dan M. Roden,[5,7,15] Jens Meiler,[4,5,6,7,12,13,14] Georg Kuenze,[4,6,12,*] John A. Phillips,[2,*] and John A. Capra[1,6,7,15,16,*]

Thank you!

Funding:

National Institutes of Health

march of dimes

Burroughs Wellcome Fund

http://www.capralab.org/

We are hiring!