

CAGI slide pack guidelines

I agree to my slides being used according to the following terms:

_____Gabriel Cia_____

- Authors: please include THIS slide at the beginning of your presentation to indicate that you agree with reuse of your slides in CAGI presentations, according to the terms below, and digitally sign by writing your name above.
- To disseminate the CAGI results, I encourage all CAGI participants to remix these CAGI slides for inclusion in their oral and poster presentations
- CAGI participants may use or publish any content of these slides only in compliance with the CAGI Data Use Agreement
- CAGI participants may not cite or publish any content of these slides except with the written permission of the originating author or in the primary CAGI publications
- CAGI participants must acknowledge the original authors of these slides and CAGI 6. Include the acknowledgement banner across the bottom
- CAGI participants must include the CAGI credits slide in their presentations
- Slides with a red slash (no remix) may not be included in a presentation unless all attendees are registered CAGI participants who signed the data use agreement
- Slides with a red X over the entire slide may not be used in any circumstances
- Slides with a blue slash (embargoed) may not be used unless the embargo is explicitly lifted
- Slides without slash or X may be tweeted unless flagged with the “No tweet” icon.
- CAGI participants using these slides must explain the “No tweet” icon, which means that the slide content should not be disseminated outside the presenting conference hall. If the presentation venue permits pervasive tweeting, you may not include these slides in your talk
- Note for slide authors: we expect all CAGI participants to abide by these restrictions, and will make best effort to ensure they are followed. However, bear in mind that we have limited means of enforcing them, and therefore the restrictions cannot be guaranteed.

Permission request regarding CAGI 6 recordings and hybrid presentations

I understand that my presentation as associated with CAGI 6 will be recorded and broadcasted via the internet through CAGI's designated technology provider.

As the copyright holder to this presentation entitled "Leveraging structural, evolutionary and contextual information to develop explainable models for deleterious variant prediction", I hereby agree that the copyright for the online broadcast and recording of my presentation be released to CAGI and my recording be uploaded on to the CAGI website.

Leveraging structural, evolutionary and contextual information to develop explainable models for deleterious variant prediction

by Gabriel Cia, Fabrizio Pucci, François Ancien, Daniele Raimondi, Wim Vranken, Marianne Rومان



COMPUTATIONAL BIOLOGY
AND BIOINFORMATICS LAB



Leveraging structural, evolutionary and contextual information to develop explainable models for deleterious variant prediction

by Gabriel Cia, Fabrizio Pucci, François Ancien, Daniele Raimondi, Wim Vranken, Marianne Rومان



COMPUTATIONAL BIOLOGY
AND BIOINFORMATICS LAB



Deleterious variant predictors



Deleterious variant predictors

Multiple Sequence
Alignment (**MSA**)

S	Y	K	V	V	T	V
A	Y	K	V	V	S	V
S	Y	K	V	I	T	V
A	Y	K	V	V	T	T
S	Y	K	V	I	T	V
T	Y	K	V	V	T	T



Deleterious variant predictors

Multiple Sequence
Alignment (**MSA**)

S	Y	K	V	V	T	V
A	Y	K	V	V	S	V
S	Y	K	V	I	T	V
A	Y	K	V	V	T	T
S	Y	K	V	I	T	V
T	Y	K	V	V	T	T



Prediction



Molecular or biophysical
explanation



Insights for drug development

Solution: explainable models

Solution: explainable models

SNPMuSiC

Solution: explainable models

SNPMuSiC

Features

- Protein stability

Solution: explainable models

SNPMuSiC

Features

- Protein stability



Solution: explainable models

SNPMuSiC

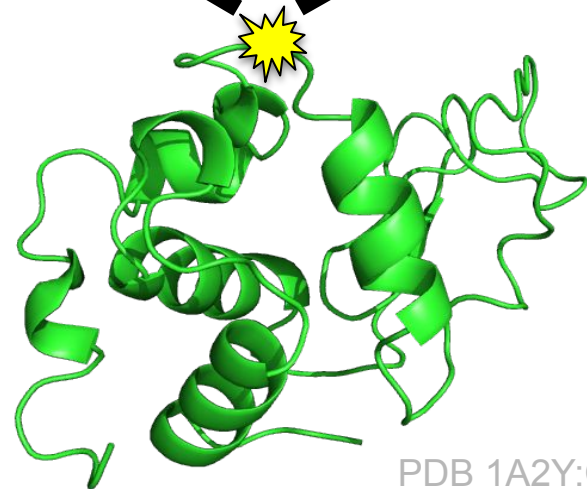
Features

- Protein stability



Deleterious

Neutral



Solution: explainable models

SNPMuSiC

Features

- Protein stability
 - Statistical potentials

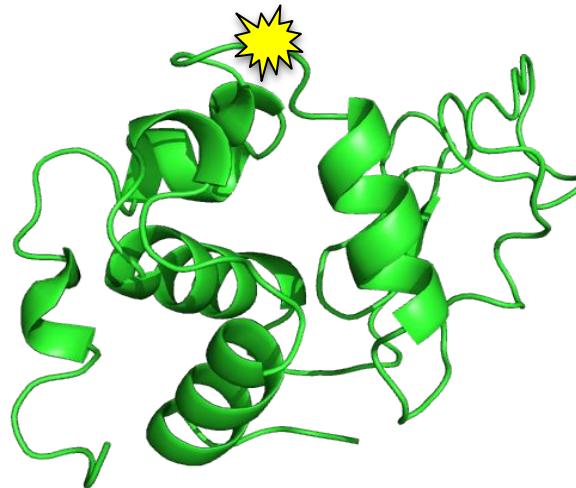
Solution: explainable models

SNPMuSiC

Features

- Protein stability
 - Statistical potentials

$$\text{Change in } \Delta G$$
$$\Delta G_{\text{WT}} - \Delta G_{\text{MUT}}$$



Solution: explainable models

SNPMuSiC

Features

- Protein stability
 - Statistical potentials
 - Change in residue volume

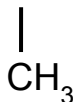
Solution: explainable models

SNPMuSiC

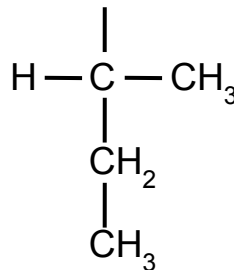
Features

- Protein stability
 - Statistical potentials
 - Change in residue volume

ALA



ILE



Solution: explainable models

SNPMuSiC

Features

- Protein stability
 - Statistical potentials
 - Change in residue volume
 - Residue solvent accessibility

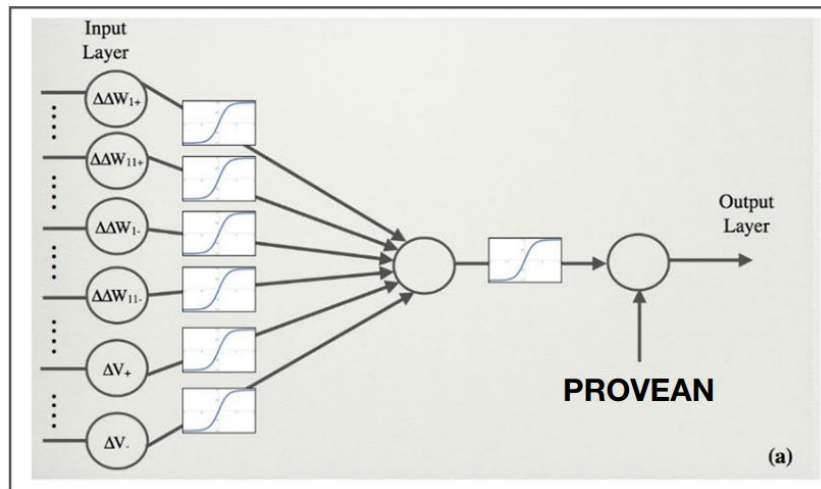
Solution: explainable models

SNPMuSiC

Features

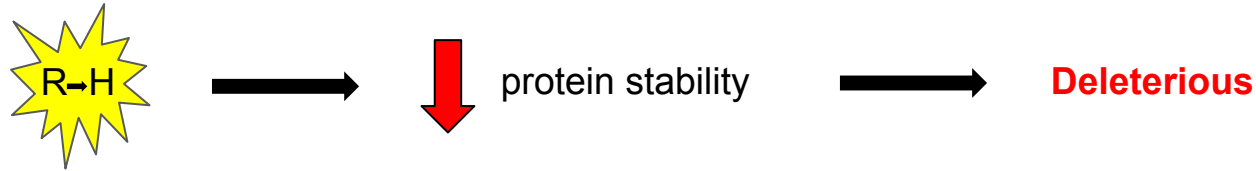
- Protein stability
 - Statistical potentials
 - Change in residue volume
 - Residue solvent accessibility
 - PROVEAN

Model architecture



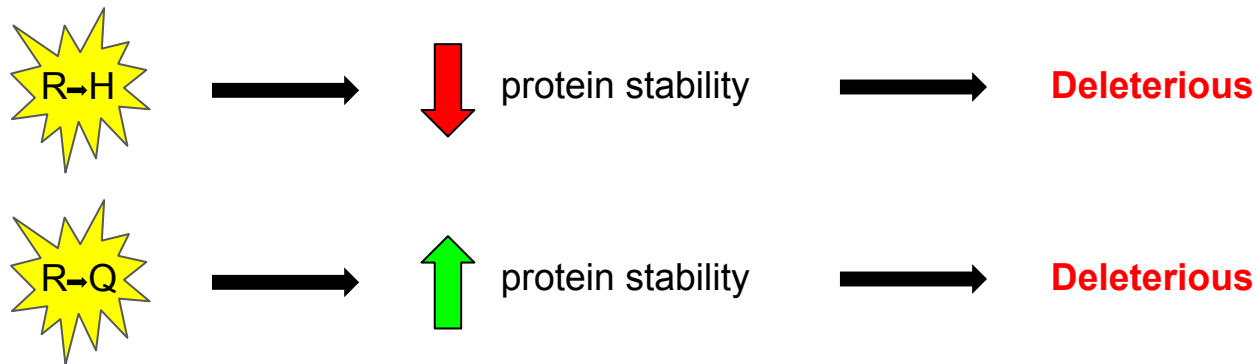
Solution: explainable models

SNPMuSiC



Solution: explainable models

SNPMuSiC



Solution: explainable models

SNPMuSiC

Performance

Solution: explainable models

SNPMuSiC

Performance

BACC	0.75
------	------

Solution: explainable models

SNPMuSiC

Performance

BACC	0.75
PPV	0.89

DEOGEN 2



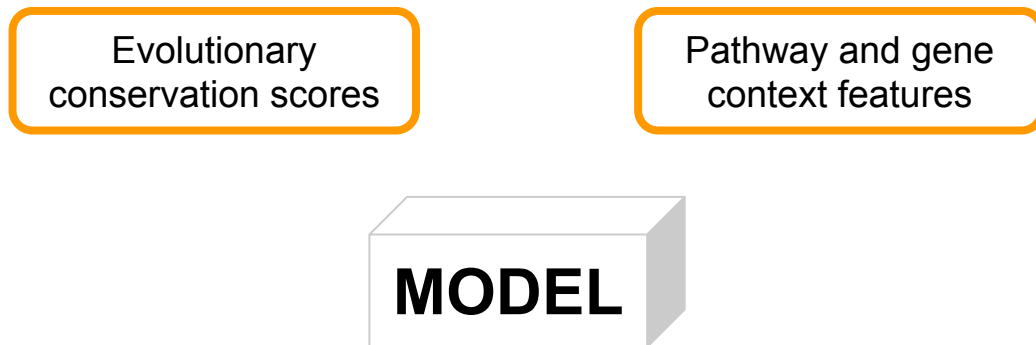
DEOGEN 2



Evolutionary
conservation scores



DEOGEN 2



DEOGEN 2



Evolutionary
conservation scores

Pathway and gene
context features

Tendency of PFAM
domains to be deleterious

MODEL

DEOGEN 2



Evolutionary
conservation scores

Pathway and gene
context features

Tendency of PFAM
domains to be deleterious

MODEL

Impact of mutation
on early folding

DEOGEN 2



Evolutionary
conservation scores

Pathway and gene
context features

Tendency of PFAM
domains to be deleterious

MODEL

Impact of mutation
on early folding

Presence of a PPI
patch

DEOGEN 2



**Diverse set
of features**

DEOGEN 2



**Diverse set
of features**



**Random
forest**

DEOGEN 2

Diverse set
of features



Determine the **weight** of
each group of features

 Explainability

DEOGEN 2

Diverse set
of features



Determine the **weight** of
each group of features

 Explainability

BACC = + 0.85

Annotate all missense challenge

Challenge: predict the functional effect (ie: neutral or deleterious) of ~82M human non-synonymous SNPs from the dbNSFP database.

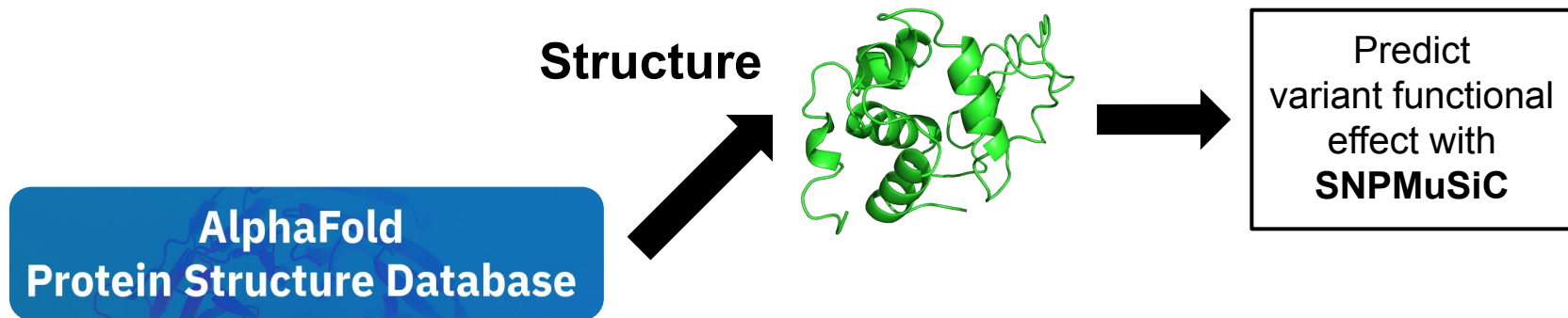
Annotate all missense challenge

Challenge: predict the functional effect (ie: neutral or deleterious) of ~82M human non-synonymous SNPs from the dbNSFP database.

AlphaFold
Protein Structure Database

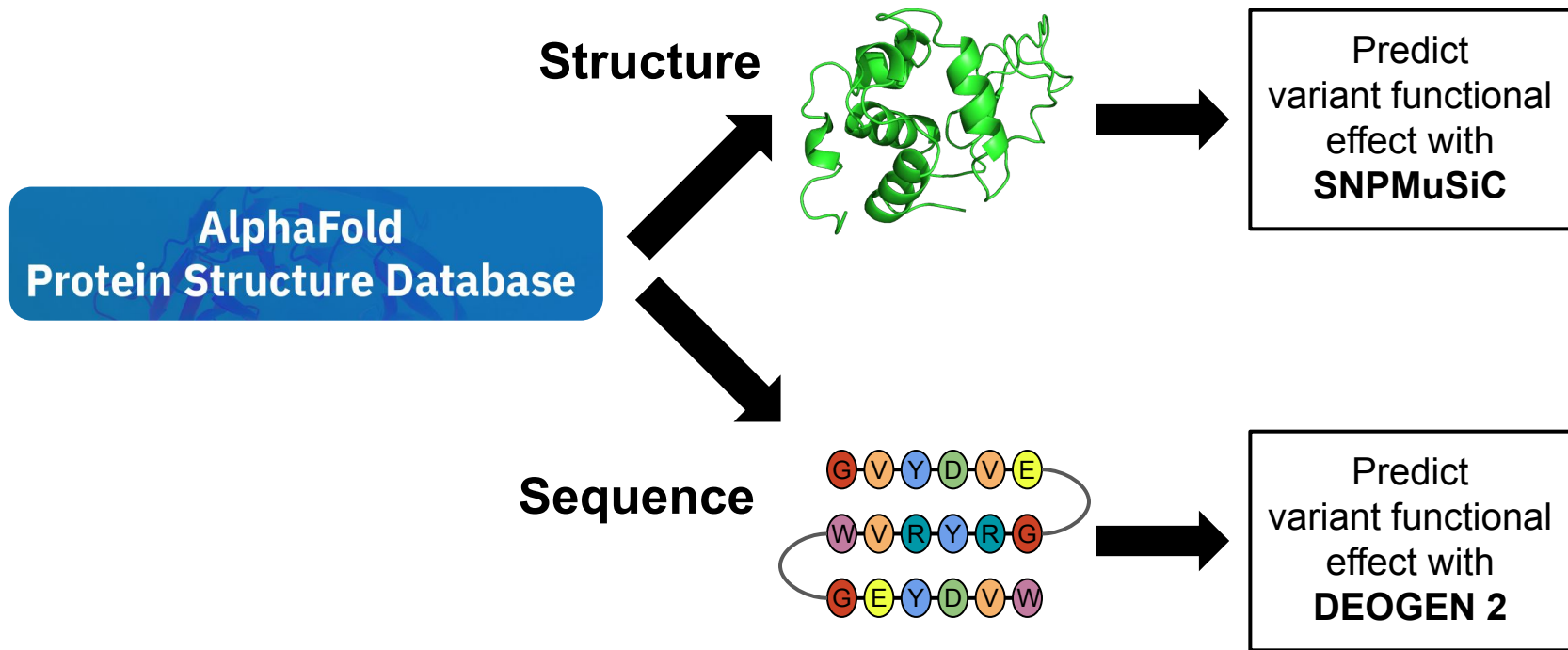
Annotate all missense challenge

Challenge: predict the functional effect (ie: neutral or deleterious) of ~82M human non-synonymous SNPs from the dbNSFP database.



Annotate all missense challenge

Challenge: predict the functional effect (ie: neutral or deleterious) of ~82M human non-synonymous SNPs from the dbNSFP database.



Annotate all missense challenge

Challenge: predict the functional effect (ie: neutral or deleterious) of ~82M human non-synonymous SNPs from the dbNSFP database.

