CAGI slide pack guidelines

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

Susanna Zucca

- 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



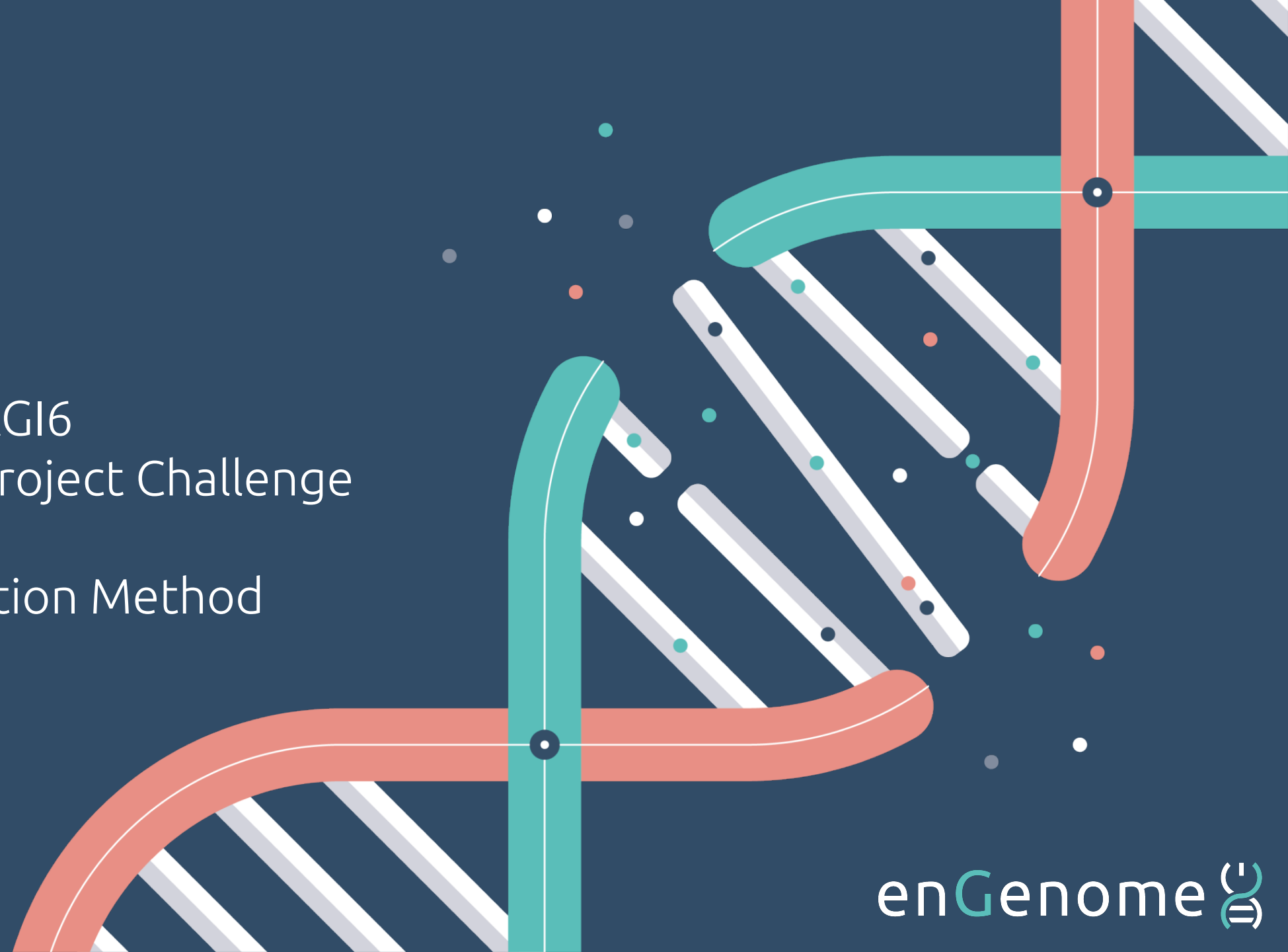
Permission request regarding CAGI 6 recordings and hybrid presentations

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 CAGI RGP: The E-team, 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.

CAGI6
Rare Genome Project Challenge

eVai Prediction Method



enGenome

UNIVERSITÀ
DI PAVIA



enGenome

Enhance healthcare
with advanced genomics

Outline



Background



eVai

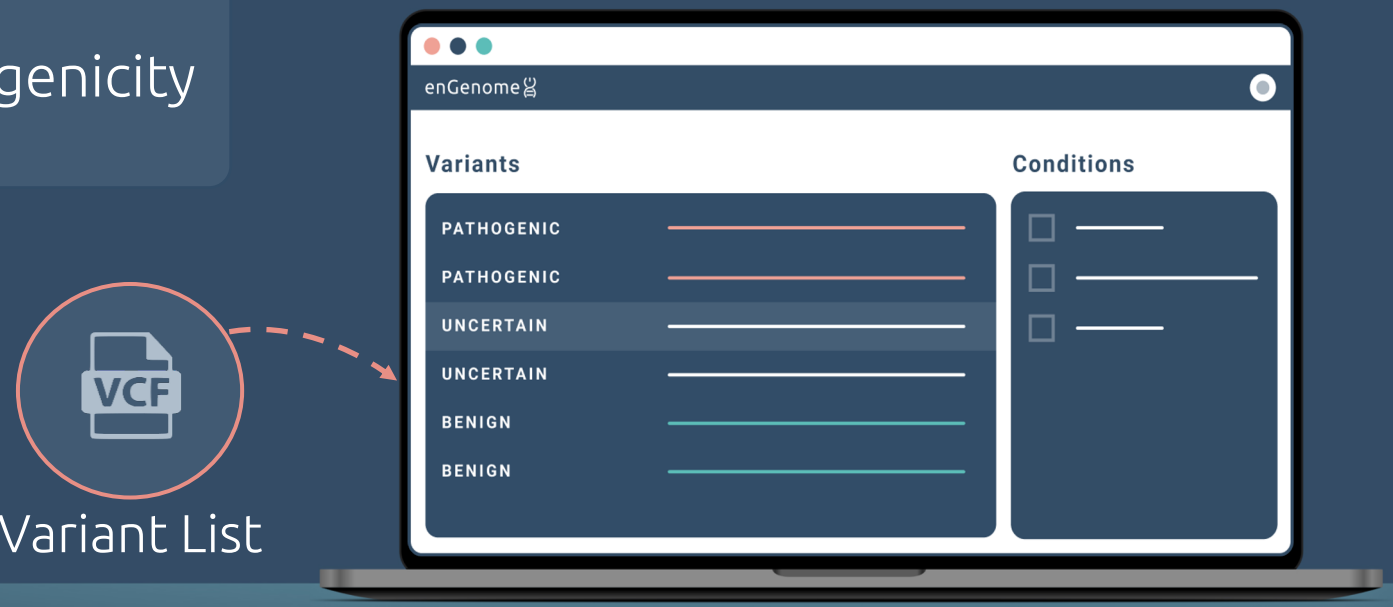


Suggested diagnosis

eVai

Software as a Service for Clinical Interpretation and Reporting of Genomic Variants

Artificial Intelligence and Bioinformatics to
CLASSIFY and **PRIORITIZE**
genomic variants according to pathogenicity



Variant List

Variant Report

PROBAND | Add Phenotype | Search by genomic variant or hgvs | Search

Filters

SNV / INDEL (29263) | CNV (253)

Export | # Variants: 29263 | Group by gene

Pathogenic (11)	P	N	SETBP1	missense_variant	p.Asp868Asn, c.2602G>A
Likely pathogenic (5)	P		SEC23B	missense_variant	p.Glu109Lys, c.325G>A
Uncertain significance (4)		Rc	SPTBN2	missense_variant	p.Ser1463Leu, c.4388C>T
Uncertain significance (4)		Rc	VARS2	splice_acceptor_variant...	c.2557-2A>G
Uncertain significance (4)			NOS3	missense_variant	p.Val914Ala, c.2741T>C
Uncertain significance (3.5)			TECTA	missense_variant	p.Asp414Val, c.1241A>T
Uncertain significance (3.5)			BCL2	inframe_deletion	p.Ser51del, c.151_153delTCC
Uncertain significance (3)		N	ANKRD26	frameshift_variant	p.Pro596fs, c.1788delC
Uncertain significance (3)		N	ANK3	intron_variant	c.2841+112T>C
Uncertain significance (3)			SBF2	5_prime_UTR_variant	c.-73C>T
Uncertain significance (3)		N	KIAA1033	splice_donor_variant...	c.1893+1G>A
Uncertain significance (3)			SLC7A7	missense_variant	p.Pro149Leu, c.446C>T
Uncertain significance (3)			GATM	stop_gained	p.Arg194*, c.580C>T
Uncertain significance (3)			CRYAA	missense_variant	p.Arg103Cys, c.307C>T
Uncertain significance (3)		N	FN1	frameshift_variant	p.Ser1709fs, c.5126delG
Uncertain significance (3)			DAZL	missense_variant	p.Tyr223Asn, c.667T>A

Conditions

Associations: 3

Schinzel-Giedion syndrome

Condition | Phenotypes (67 Total)

ACMG Classification: Pathogenic Confirm

Pathogenicity Score: 11

ACMG Criteria: Edit

PS1 | PM1 | PM2 | PM5 | PP3 | PP5

Gene: SETBP1

Inheritance: Autosomal dominant

MedGen ID: C0265227

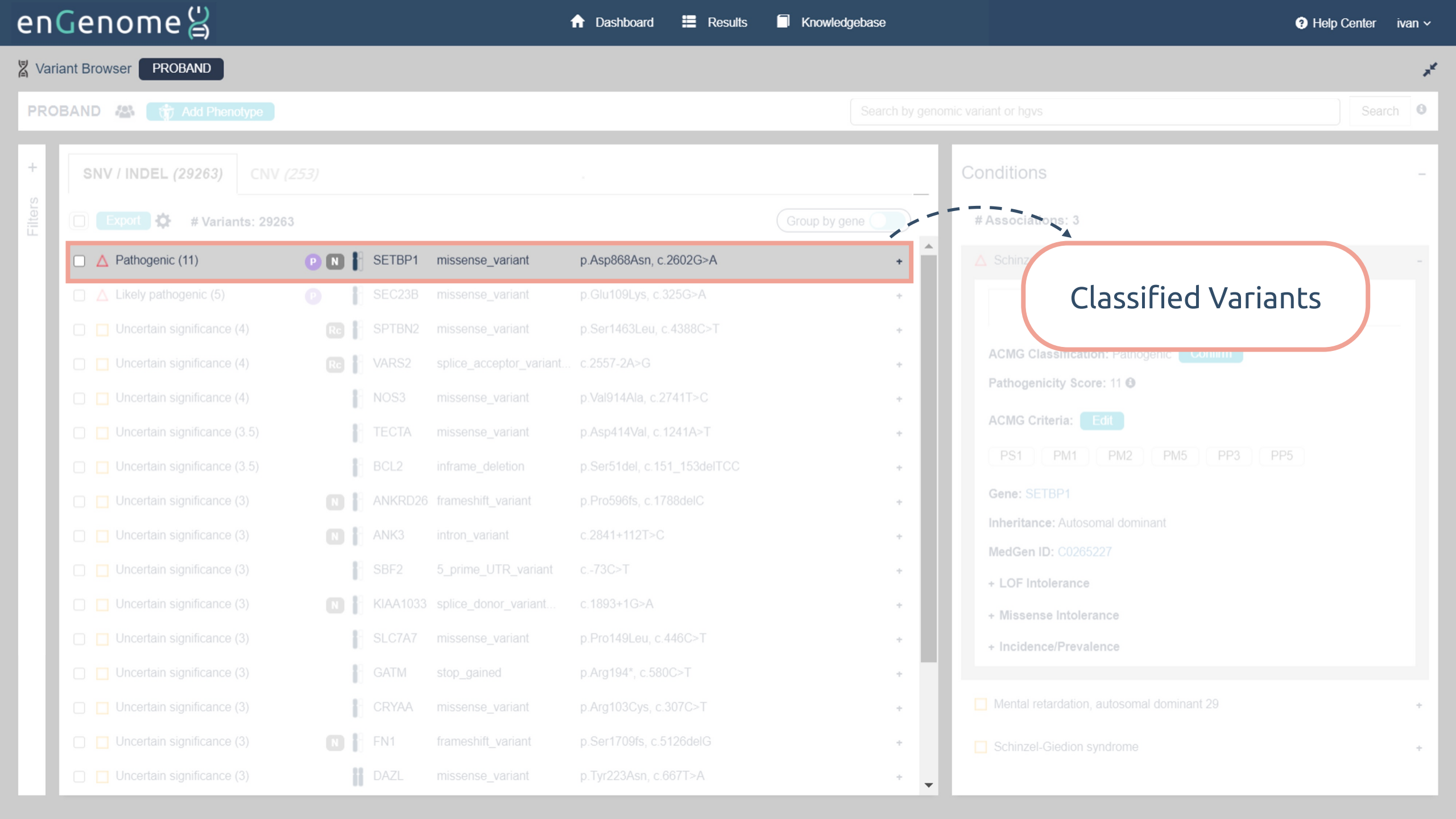
+ **LOF Intolerance**

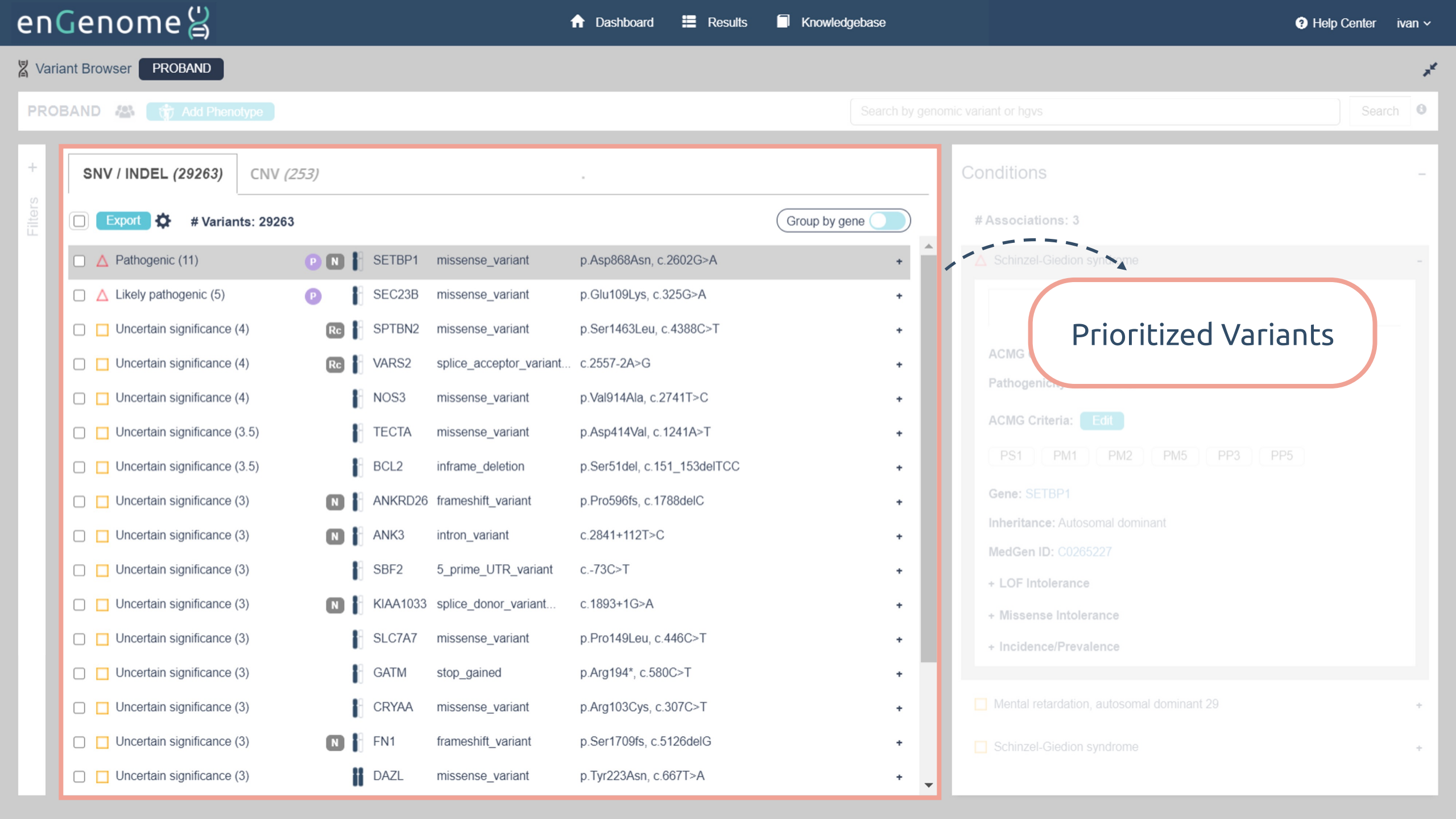
+ **Missense Intolerance**

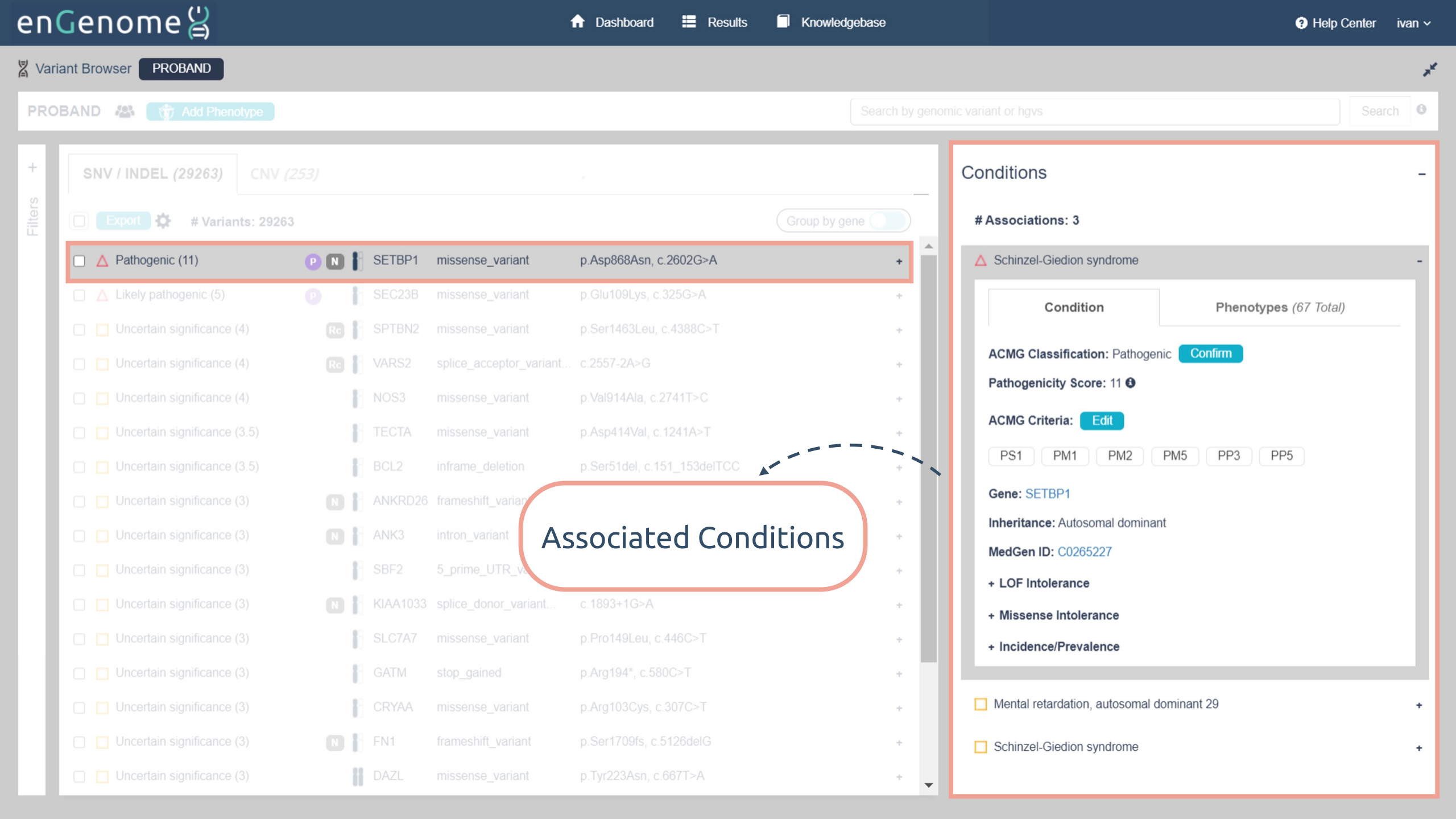
+ **Incidence/Prevalence**

Mental retardation, autosomal dominant 29

Schinzel-Giedion syndrome


enGenome
Dashboard
Results
Knowledgebase
Help Center
ivan
Variant Browser
PROBAND
PROBAND
Add Phenotype
Search by genomic variant or hgvs
Search
Filters
SNV / INDEL (29263)
CNV (253)
Export
Variants: 29263
Group by gene
Pathogenic (11) P N SETBP1 missense_variant p.Asp868Asn, c.2602G>A
Likely pathogenic (5) P SEC23B missense_variant p.Glu109Lys, c.325G>A
Uncertain significance (4) Rc SPTBN2 missense_variant p.Ser1463Leu, c.4388C>T
Uncertain significance (4) Rc VARS2 splice_acceptor_variant... c.2557-2A>G
Uncertain significance (4) NOS3 missense_variant p.Val914Ala, c.2741T>C
Uncertain significance (3.5) TECTA missense_variant p.Asp414Val, c.1241A>T
Uncertain significance (3.5) BCL2 inframe_deletion p.Ser51del, c.151_153delTCC
Uncertain significance (3) N ANKRD26 frameshift_variant p.Pro596fs, c.1788delC
Uncertain significance (3) N ANK3 intron_variant c.2841+112T>C
Uncertain significance (3) SBF2 5_prime_UTR_variant c.-73C>T
Uncertain significance (3) N KIAA1033 splice_donor_variant... c.1893+1G>A
Uncertain significance (3) SLC7A7 missense_variant p.Pro149Leu, c.446C>T
Uncertain significance (3) GATM stop_gained p.Arg194*, c.580C>T
Uncertain significance (3) CRYAA missense_variant p.Arg103Cys, c.307C>T
Uncertain significance (3) N FN1 frameshift_variant p.Ser1709fs, c.5126delG
Uncertain significance (3) DAZL missense_variant p.Tyr223Asn, c.667T>A
Conditions
Associations: 3
Classified Variants
ACMG Classification: Pathogenic
Confirm
Pathogenicity Score: 11
ACMG Criteria:
Edit
PS1
PM1
PM2
PM5
PP3
PP5
Gene: SETBP1
Inheritance: Autosomal dominant
MedGen ID: C0265227
+ LOF Intolerance
+ Missense Intolerance
+ Incidence/Prevalence
Mental retardation, autosomal dominant 29
Schinzel-Giedion syndrome


enGenome
Dashboard
Results
Knowledgebase
Help Center
ivan
Variant Browser
PROBAND
PROBAND
Add Phenotype
Search by genomic variant or hgvs
Search
Filters
SNV / INDEL (29263)
CNV (253)
Export
Variants: 29263
Group by gene
Pathogenic (11) P N SETBP1 missense_variant p.Asp868Asn, c.2602G>A
Likely pathogenic (5) P SEC23B missense_variant p.Glu109Lys, c.325G>A
Uncertain significance (4) Rc SPTBN2 missense_variant p.Ser1463Leu, c.4388C>T
Uncertain significance (4) Rc VARS2 splice_acceptor_variant... c.2557-2A>G
Uncertain significance (4) NOS3 missense_variant p.Val914Ala, c.2741T>C
Uncertain significance (3.5) TECTA missense_variant p.Asp414Val, c.1241A>T
Uncertain significance (3.5) BCL2 inframe_deletion p.Ser51del, c.151_153delTCC
Uncertain significance (3) N ANKRD26 frameshift_variant p.Pro596fs, c.1788delC
Uncertain significance (3) N ANK3 intron_variant c.2841+112T>C
Uncertain significance (3) SBF2 5_prime_UTR_variant c.-73C>T
Uncertain significance (3) N KIAA1033 splice_donor_variant... c.1893+1G>A
Uncertain significance (3) SLC7A7 missense_variant p.Pro149Leu, c.446C>T
Uncertain significance (3) GATM stop_gained p.Arg194*, c.580C>T
Uncertain significance (3) CRYAA missense_variant p.Arg103Cys, c.307C>T
Uncertain significance (3) N FN1 frameshift_variant p.Ser1709fs, c.5126delG
Uncertain significance (3) DAZL missense_variant p.Tyr223Asn, c.667T>A
Conditions
Associations: 3
Schinzel-Giedion syndrome
Prioritized Variants
ACMG Criteria: Edit
PS1 PM1 PM2 PM5 PP3 PP5
Gene: SETBP1
Inheritance: Autosomal dominant
MedGen ID: C0265227
+ LOF Intolerance
+ Missense Intolerance
+ Incidence/Prevalence
Mental retardation, autosomal dominant 29
Schinzel-Giedion syndrome


enGenome
Dashboard
Results
Knowledgebase
Help Center
ivan
Variant Browser
PROBAND
PROBAND
Add Phenotype
Search by genomic variant or hgvs
Search
Filters
SNV / INDEL (29263)
CNV (253)
Export
Variants: 29263
Group by gene
Pathogenic (11) SETBP1 missense_variant p.Asp868Asn, c.2602G>A
Likely pathogenic (5) SEC23B missense_variant p.Glu109Lys, c.325G>A
Uncertain significance (4) SPTBN2 missense_variant p.Ser1463Leu, c.4388C>T
Uncertain significance (4) VARS2 splice_acceptor_variant... c.2557-2A>G
Uncertain significance (4) NOS3 missense_variant p.Val914Ala, c.2741T>C
Uncertain significance (3.5) TECTA missense_variant p.Asp414Val, c.1241A>T
Uncertain significance (3.5) BCL2 inframe_deletion p.Ser51del, c.151_153delTCC
Uncertain significance (3) ANKRD26 frameshift_varian
Uncertain significance (3) ANK3 intron_variant
Uncertain significance (3) SBF2 5_prime_UTR_v
Uncertain significance (3) KIAA1033 splice_donor_variant... c.1893+1G>A
Uncertain significance (3) SLC7A7 missense_variant p.Pro149Leu, c.446C>T
Uncertain significance (3) GATM stop_gained p.Arg194*, c.580C>T
Uncertain significance (3) CRYAA missense_variant p.Arg103Cys, c.307C>T
Uncertain significance (3) FN1 frameshift_variant p.Ser1709fs, c.5126delG
Uncertain significance (3) DAZL missense_variant p.Tyr223Asn, c.667T>A
Associated Conditions
Conditions
Associations: 3
Schinzel-Giedion syndrome
Condition
Phenotypes (67 Total)
ACMG Classification: Pathogenic
Confirm
Pathogenicity Score: 11
ACMG Criteria:
Edit
PS1
PM1
PM2
PM5
PP3
PP5
Gene: SETBP1
Inheritance: Autosomal dominant
MedGen ID: C0265227
+ LOF Intolerance
+ Missense Intolerance
+ Incidence/Prevalence
Mental retardation, autosomal dominant 29
Schinzel-Giedion syndrome

How eVai works - Omics resources integration

Raw Variant List

1. INTERNAL KNOWLEDGEBASE

Each variant is
matched to the INTERNAL Knowledgebase
integrating more than 30 -omic resources



MedGen	ClinVar	HPO
RefSeq	Control Population	Functional Studies
In-silico Prediction	Splicing Prediction	Functional Domains

How eVai works - Omics resources integration

Raw Variant List

1. INTERNAL KNOWLEDGEBASE

2. USER KNOWLEDGEBASE

Each variant is
matched to the USER Knowledgebase
built by the the user over time



Variants
Diseases
Genes
Family co-segregation
Functional Studies
Domains
Papers

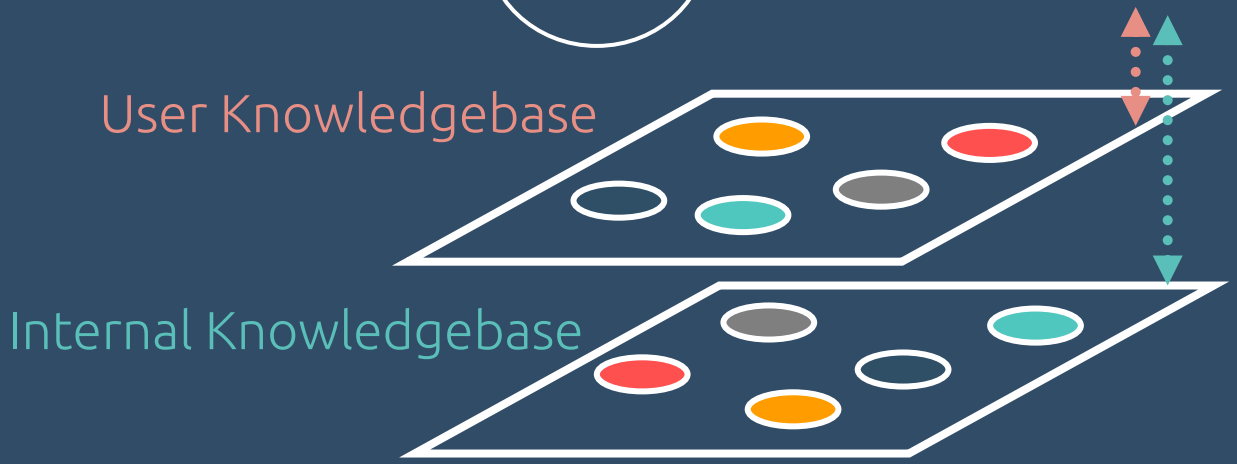
How eVai works - Expert system

VCF Raw Variant List

1. INTERNAL KNOWLEDGEBASE
2. USER KNOWLEDGEBASE
3. REASONING AND SCORING

On the whole variant set
eVAI verifies ACMG criteria
by reasoning on meta-knowledgebase

eVai inference engine



How eVai works - Expert system



Raw Variant List

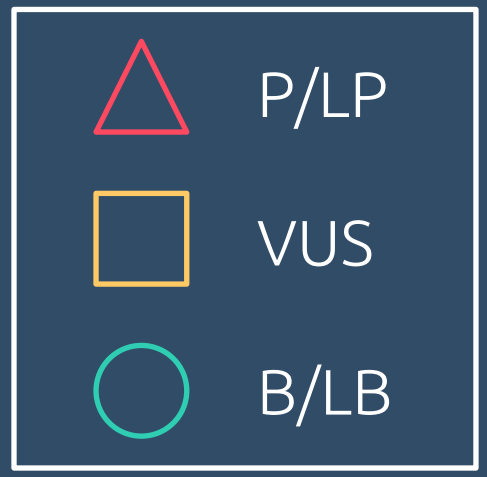
1. INTERNAL KNOWLEDGEBASE

2. USER KNOWLEDGEBASE

3. REASONING AND SCORING

4. ACMG-BASED REPORT

Each variant is
ACMG-Classified
Disease-based classification



Pathogenic

+4 Very strong

+3 Strong

+2 Moderate

+1 Supporting

Benign

-0.5 Supporting

-1 Strong

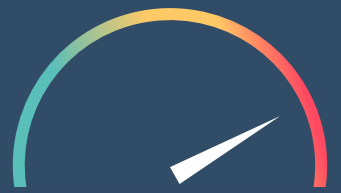
-1.5 Stand-Alone

How eVai works - Machine Learning model

VCF Raw Variant List

1. INTERNAL KNOWLEDGEBASE
2. USER KNOWLEDGEBASE
3. REASONING AND SCORING
4. ACMG-BASED REPORT
5. SUGGESTED DIAGNOSIS

Each variant is scored according to its clinical significance:
Suggested diagnosis



CLINICAL SCORE

enGenome

Genomic variants classification and prioritization

Validated on more than

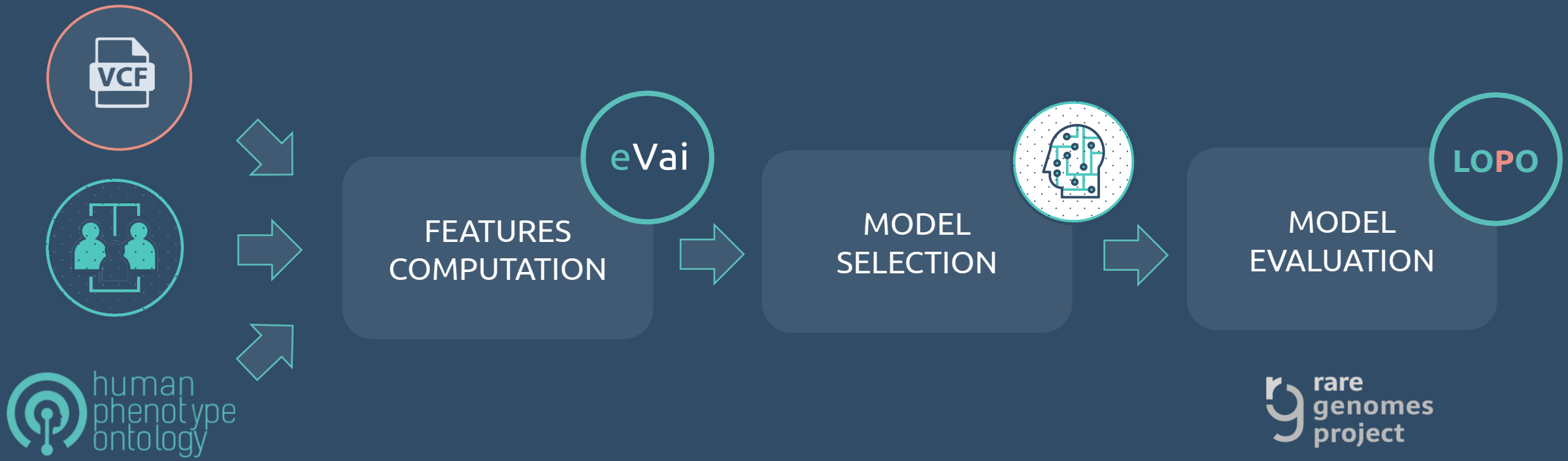
17.000 clinically-observed variants

Nicora, Limongelli et al, ***Human Mutation****, October 2018*
Nicora, Limongelli, Zucca et al, ***Scientific Reports****, January 2022*

Patent Application 102020000019180

Method for determining the pathogenicity/benignity of a genomic variant in relation to a given disease

The eVai Machine Learning approach



Features computation

What's the estimated pathogenicity of the variant?



Variant pathogenicity

Does the family segregation match the inheritance of the condition?



Family inheritance

How much overlap exists between patient's phenotypes and gene-associated phenotypes?



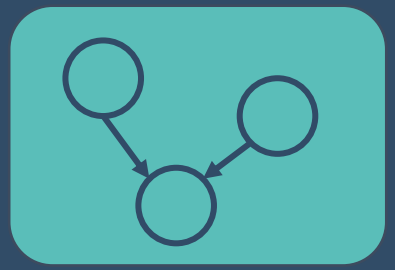
Phenotypic Similarity

Is the variant reliable?

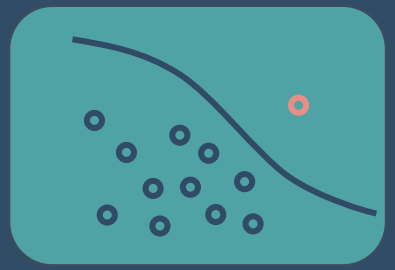


Variant quality

Model selection



MODEL 1



MODEL 2



MODEL 3



MODEL 4

Models were developed in Python through the scikit learn implementation.

Best models were selected based on their performances.

Model evaluation

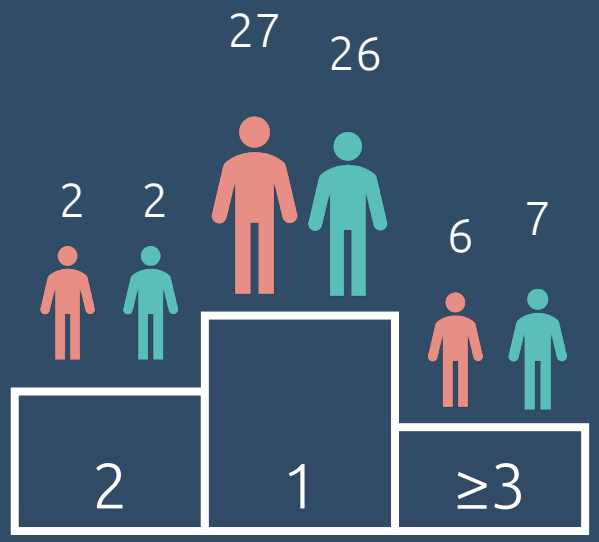
Leave-One-Proband-Out (LOPO) Cross Validation

Best performance on causative variant prioritization



Repeat for each proband

Prioritization performances on training set



Model 1:

- 77% of causative variants predicted pathogenic and ranked in 1° position (27/35)
- 97% of causative variants predicted pathogenic and ranked in the top 10 (34/35)

Model 2:

- 74% of causative variants predicted pathogenic and ranked in 1° position (26/35)
- 100% of causative variants predicted pathogenic and ranked in the top 10 (35/35)

Cagi6-RGP Test prediction

MODEL TRAINING → TEST PREDICTION

CAGI TRAINING COHORT
N=35



TRAINING

CAGI TEST COHORT
N=30



TEST

Thanks to

Development team

Ivan Limongelli
Federica De Paoli
Giovanna Nicora
Maria Giulia Carta

Advisory Board

Ettore Rizzo
Riccardo Bellazzi
Paolo Magni
Anne Bowcock

Variants Combinations: Classification & Prioritization

INPUT

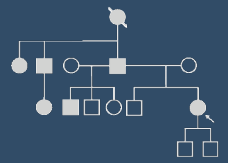
Patient's Variants



Patient's Clinical History

human phenotype ontology

Patient's Family History



METHOD

Machine Learning algorithm for digenic variant interpretation

DIVAs

Patent Pending

Training Dataset

DIDA: Digenic Diseases Database
1000 genome project
Internal enGenome Dataset
Manually curated Scientific Literature

OUTPUT



Classification and prioritization of combinations of variants on two genes

97% sensitivity
98% specificity

enGenome

Enhance healthcare with
advanced genomics



www.engenome.com
info@engenome.com