



Sherloc Prediction Challenge

**Predict the classification of novel variants encountered
across genetic tests conducted by Invitae**

05 | 2022

Rachel Hovde, Peter Combs, Alex Colavin



CAGI slide pack guidelines

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

Rachel Hovde, Peter Combs, Alex Colavin

- 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.



INVITAE: A HIGH VOLUME CLINICAL GENETIC TEST PROVIDER

ClinVar

ClinVar

Search ClinVar for gene symbols, HGVS expressions, conditions, and more

Search

[Advanced](#)

[Help](#)

[Home](#)

[About](#)

[Access](#)

[Help](#)

[Submit](#)

[Statistics](#)

[FTP](#)



COVID-19 Information



[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

Submitters and their submissions

This page lists all submitters and the summary of their contributions, ordered inversely by the number of submissions. We acknowledge their support.

Note that the number of submissions for any submitter may be different from the number of variants found in a search based on the name of that submitter. A submission is for the interpretation of a variant and a condition; if the submitter has interpreted the variant with respect to more than one condition, then the count of submissions will be higher than the search result.

Submitter	Maximum review status	Total submissions	Submissions with interpretations	Total Genes	Last updated
Invitae	Assertion criteria	606139	606136	14180	Nov 23, 2021
GeneDx	Assertion criteria	308152	308022	29439	Jan 20, 2022
Illumina Laboratory Services; Illumina	Assertion criteria	208884	208884	5749	Dec 20, 2021
Ambry Genetics	Assertion criteria	72615	72615	1722	Dec 09, 2021

SHERLOC: A SEMI-QUANTITATIVE VARIANT CLASSIFICATION SYSTEM

#1 Five types of evidence, considered in a hierarchical approach:

Population Data
Variant Type
Clinical Observations
Experimental Studies
Indirect and Computational

Categories

#2 Rule based scoring for each individual piece of evidence

B	P
	*
	*
	*

Criteria & Scores
(1–5 pts)

#3 Point score thresholds to determine final classification based on the ACMG suggested five-tier classification system



Classification Thresholds

Pathogenic: ≥ 5 pts

Benign: ≥ 4 pts

CAGI Sherlock Challenge

- **Predict variant pathogenicity**
 - 3 subchallenges: missense, small intronic indel, small exonic indel
 - No team indicated that they wanted to be graded specifically on a subchallenge
- **Test set:**
 - ~70k variants
 - Most ClinVar submissions held back until post-CAGI

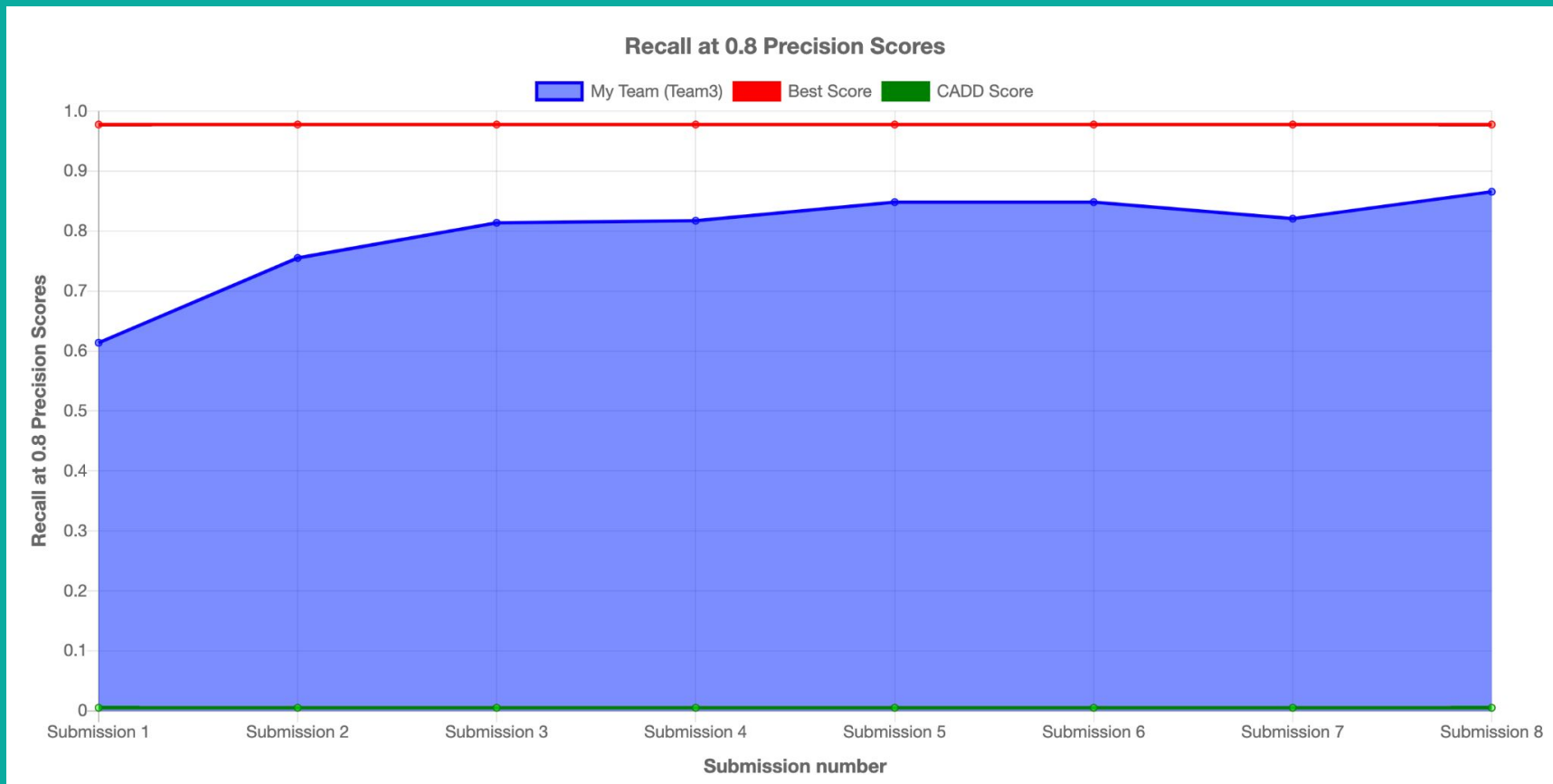
Subchallenge	total	benign	path
All	71427	61673	9754
Missense	5676	3295	2381
Small intronic indel	3299	3216	83
Small exonic in-frame indel	466	192	274
Other	61986	54970	7016

Challenge goals

- Encourage contestants to build robust, clinically-applicable models
- Highlight innovative techniques
- See whether differing approaches to variant interpretation give consistent results
- Invite scientists to revisit published variant interpretations using new techniques
- Pilot to open-source Invitae software:
 - Confidential code sharing
 - Ongoing evaluation of published models



Progress Tracker



ClinVar overlap

~8% of test set variants found in December 2021 ClinVar release . . .

But, turned out it didn't significantly impact the scores.

		Subcategory Missense	Subcategory Small intronic deletion or insertion	Subcategory Small exonic in-frame deletion or insertion	Uncategorized
interpretation	in_ClinVar				
benign	False	1002	945	30	4750
	True	356	112	16	1641
likely_benign	False	1545	2073	124	46688
	True	392	86	22	1891
likely_pathogenic	False	860	45	85	1121
	True	245	3	14	109
pathogenic	False	885	29	159	5155
	True	391	6	16	631



Teams

1. **3 billion:** rules-based approach on evidence from published models
2. **Bologna Biocomputing:** Missense-only; used published SNPs&GO model (SVM)
3. **Lichtarge:** Regression on scores from published models, chosen separately for each subchallenge
4. **njust-monash biohacker:** 2-layer recurrent neural net
5. **Strokach:** Gradient-boosted trees on features from deep neural nets
6. **SYSU-SAIL:** Ensemble of custom gradient-boosted trees + published predictors
7. **VARITY:** Gradient-boosted trees on quality-weighted training examples



Submission format

Variant ID

Prediction:

-1 = benign, 1=pathogenic, 0=VUS

Confidence Score:

Used for ranking the predictions from most to least certain, and determining cutoffs for precision/ negative predictive value



Quantitative scoring

Metrics:

- Recall for predictions above a confidence threshold that enforces:
 - 95% precision (12 pts)
 - 90% precision (8 pts)
- True negative rate for predictions above a confidence threshold that enforces:
 - 95% negative predictive value (12 pts)
 - 90% negative predictive value (8 pts)
- Number of variants which only that single team classified correctly (5 pts total)



Results

For simplicity, only showing the overall scores with ClinVar variants excluded

7 teams. Most teams provided multiple models. Except for the Unique Variants metric, each model treated separately.

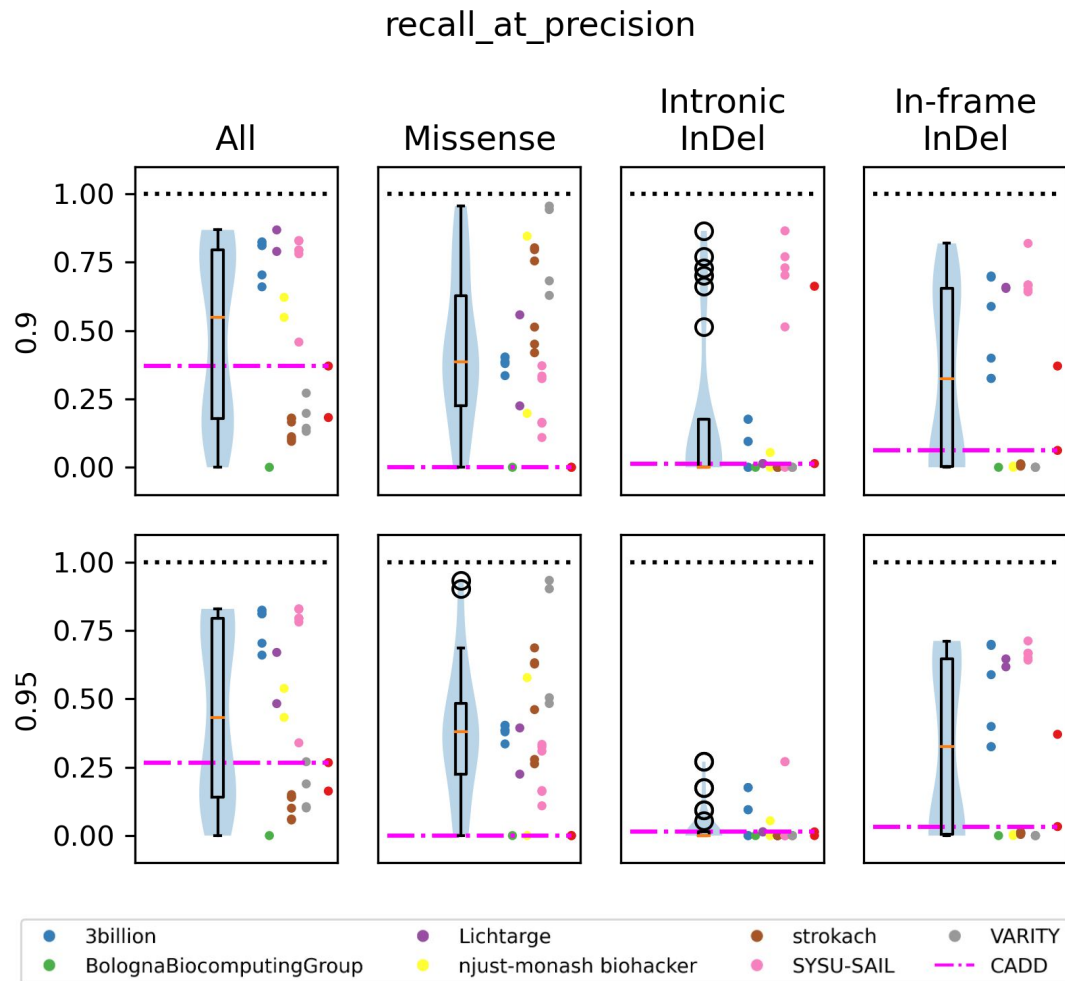


Recall at Precision (12/8 pts)

Ordering by the team-provided scores, what fraction of all the pathogenic variants were they able to call with X% precision

Specialization: VARITY very good at missense, SYSU-SAIL at indels.

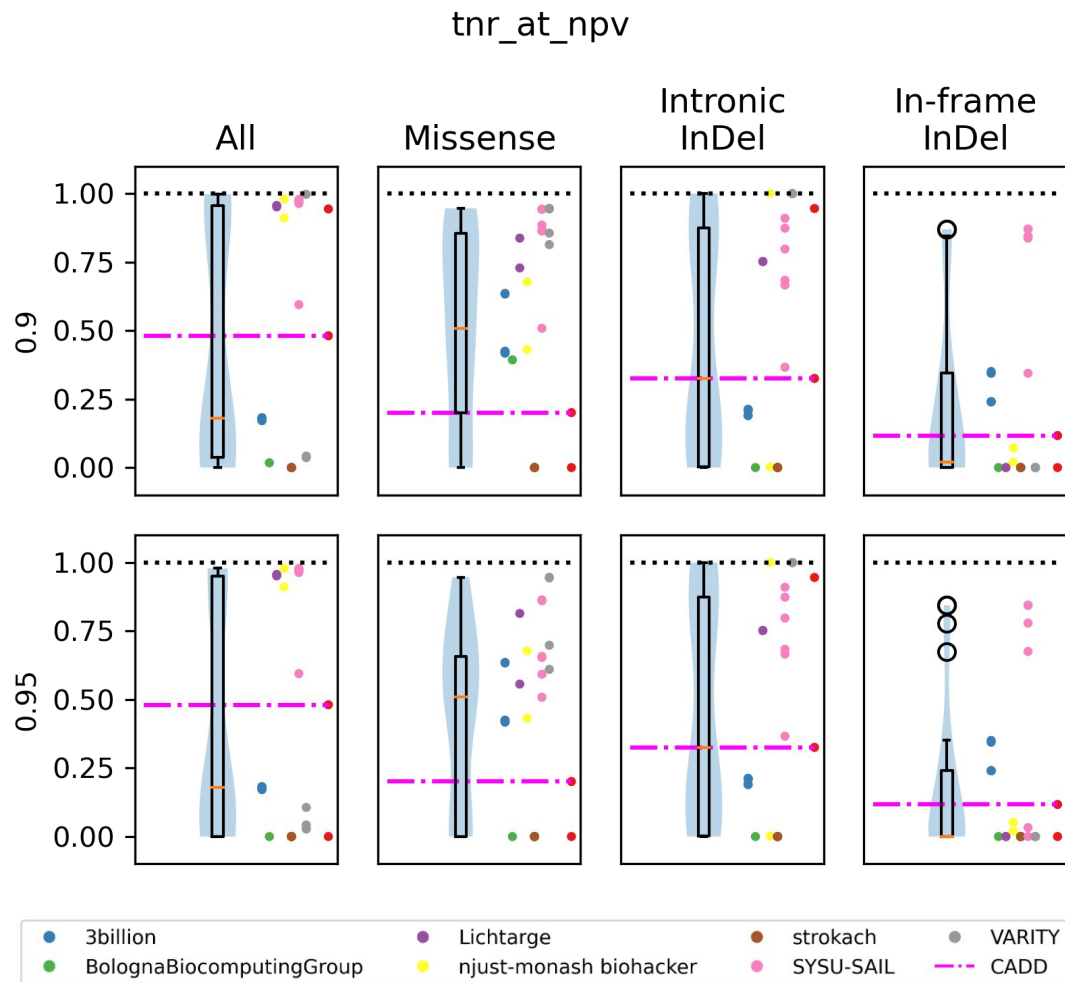
Most of 3billion's performance comes from variants not in one of these categories



True Negative Rate at NPV (12/8 pts)

Once again, SYSU-SAIL
specialized at indels (particularly
in frame)

VARITY Model 4 had a huge drop
between NPV=.90 (99.7%) and
NPV=.95 (10.6%)

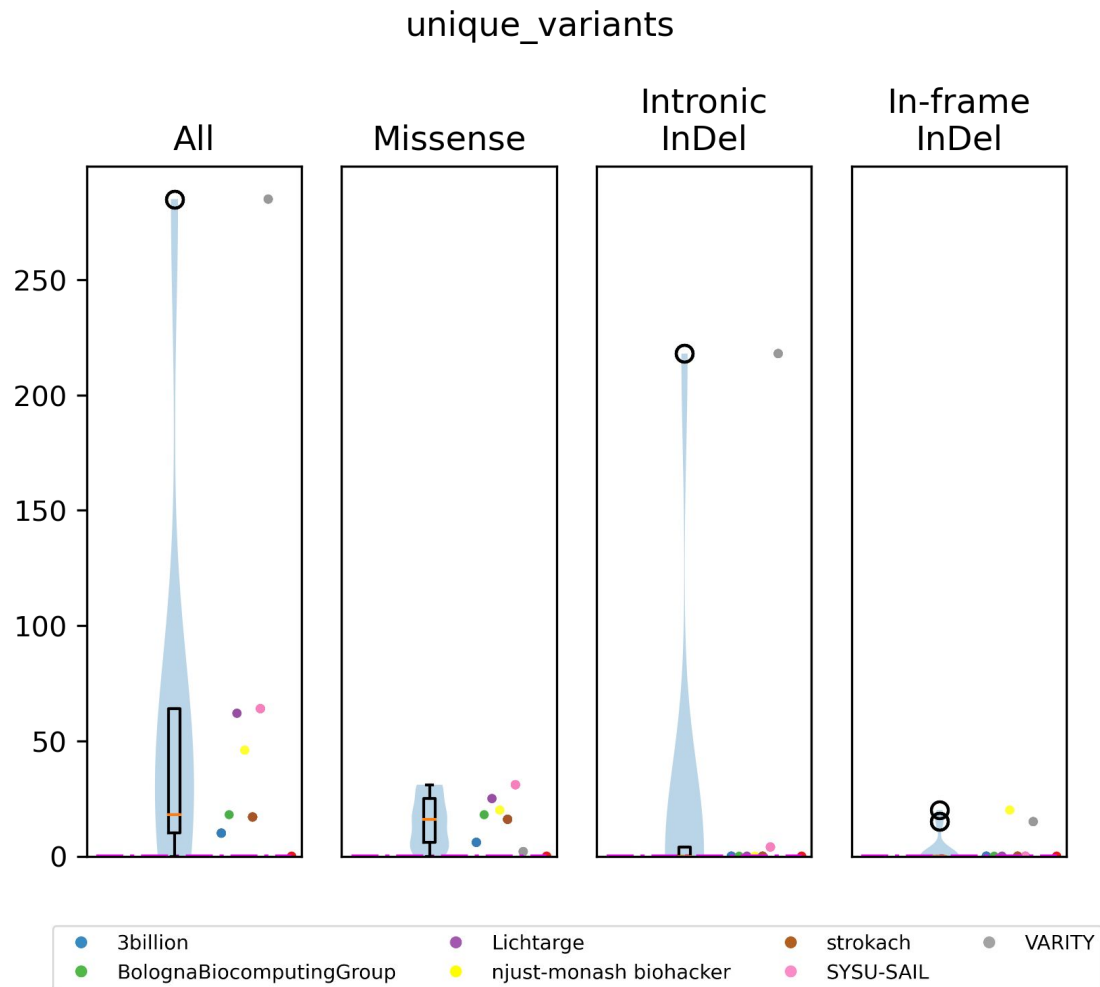


Unique Variants (5 pts)

Number of variants where the consensus across a team's submissions are correctly called, concordant, *and* not called by any other team

Most teams got 10's of variants

VARITY had 294 variants (218 of which were intronic InDels)

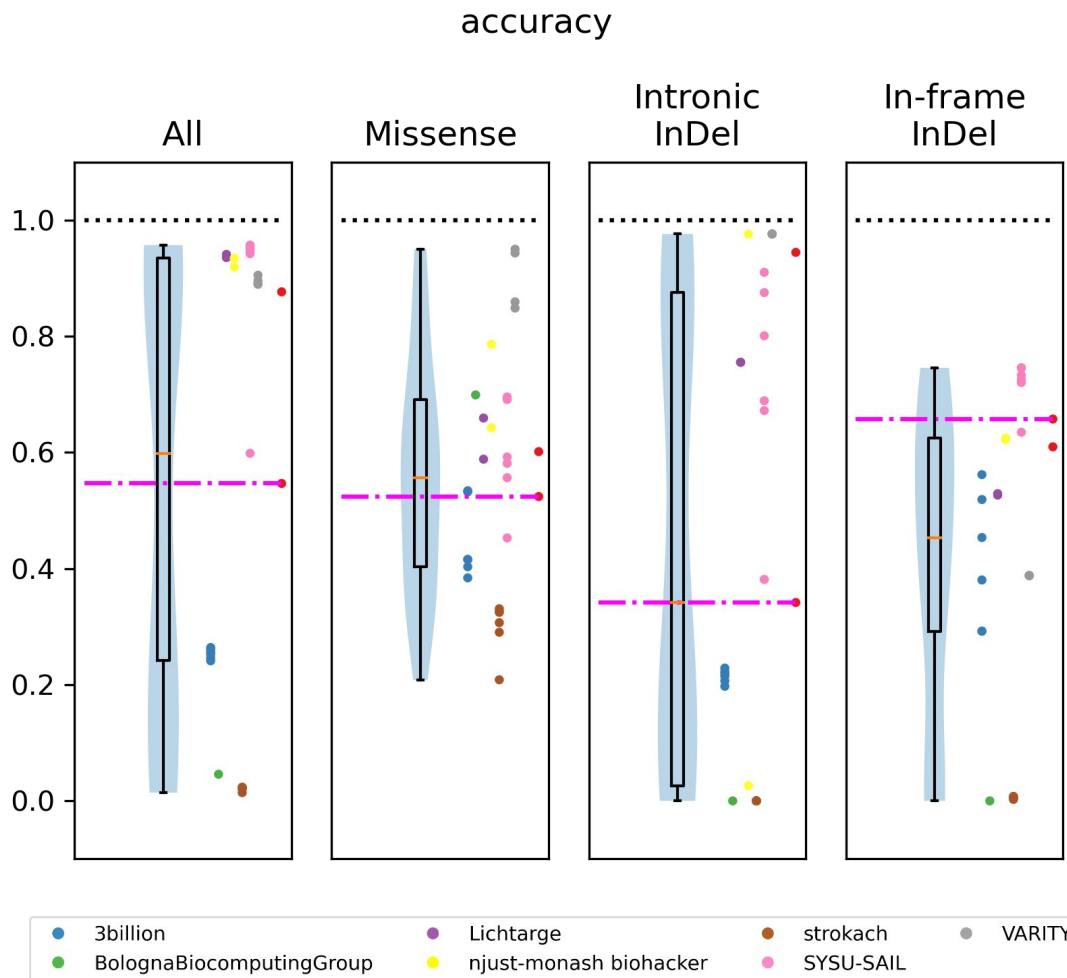


Accuracy (0 pts)

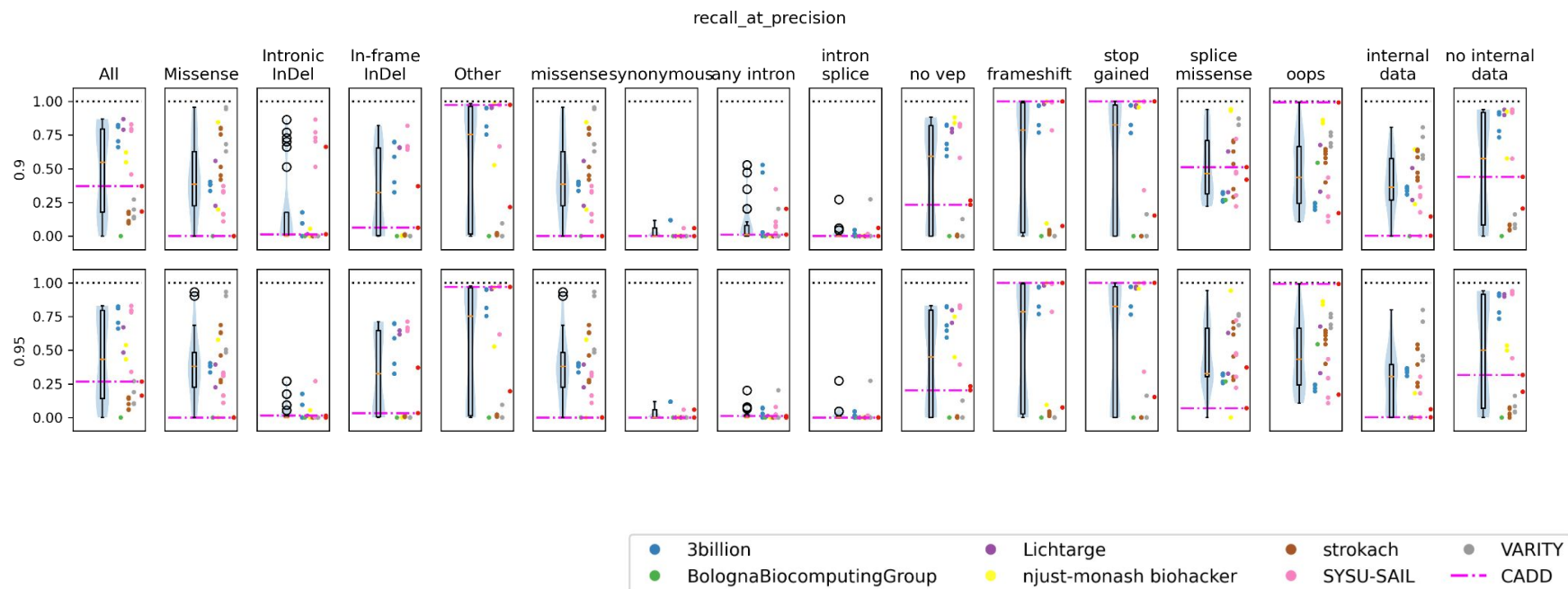
Not as informative due to the high number of benign variants

Doesn't fit our need to convey only high confidence calls to our patients

Not a factor in our scores, but teams Lichtarge, njust-monash, SYSU-SAIL, and VARITY have high accuracy overall (>0.9)



Can zoom into even finer sets to try to find where we might gain insight from particular models



Results

	Subchallenge	Final_Rank	Winner_CV_removed	Score_CV_removed	Winner CV_included	Score_CV_included
	All	1	SYSU-SAIL_m6	40.546346	SYSU-SAIL_m6	40.569282
	All	2	SYSU-SAIL_m4	40.404761	SYSU-SAIL_m4	40.433284
	All	3	SYSU-SAIL_m2	39.777288	SYSU-SAIL_m2	39.608461
	Subcategory relies on internal evidence	1	VARITY_m4	42.340426	VARITY_m4	41.984127
	Subcategory relies on internal evidence	2	VARITY_m1	37.920585	VARITY_m1	36.061121
	Subcategory relies on internal evidence	3	VARITY_m2	24.285346	3billion_m6	24.046472
	Subcategory Missense	1	VARITY_m4	40.322581	VARITY_m4	40.540541
	Subcategory Missense	2	VARITY_m1	39.769419	VARITY_m1	39.830134
	Subcategory Missense	3	njust-monash biohacker_m2	32.053457	njust-monash biohacker_m2	32.598676
	Subcategory Small exonic in-frame deletion or ...	1	SYSU-SAIL_m3	37.072490	SYSU-SAIL_m4	36.111802
	Subcategory Small exonic in-frame deletion or ...	2	SYSU-SAIL_m4	36.853360	SYSU-SAIL_m3	36.083517
	Subcategory Small exonic in-frame deletion or ...	3	SYSU-SAIL_m6	36.826480	3billion_m5	28.200207
	Subcategory Small intronic deletion or insertion	1	SYSU-SAIL_m2	38.289225	SYSU-SAIL_m2	38.294061
	Subcategory Small intronic deletion or insertion	2	SYSU-SAIL_m4	32.513049	SYSU-SAIL_m4	29.092044
	Subcategory Small intronic deletion or insertion	3	teamR_m2	25.038188	SYSU-SAIL_m3	27.639635
	Subcategory Other	1	SYSU-SAIL_m4	42.276964	Lichtarge_m1	42.803315
	Subcategory Other	2	SYSU-SAIL_m3	42.239535	SYSU-SAIL_m4	42.536914
	Subcategory Other	3	SYSU-SAIL_m2	42.227476	Lichtarge_m2	42.517016

Easy variants

Benign variants mostly replace with AAs of similar properties,
Pathogenic mostly with different properties

Benign

alt.pro	R	H	K	D	E	S	T	N	Q	A	V	I	L	M	F	Y	W	C	G	P
ref.pro																				
R	0	28	8	0	0	3	0	0	46	0	0	1	2	0	0	0	5	6	9	1
H	20	0	0	0	0	0	0	1	9	0	0	0	1	0	0	9	0	0	0	0
K	13	0	0	0	9	0	1	10	1	0	0	1	0	0	0	0	0	0	0	0
D	0	0	0	0	10	0	0	13	0	1	4	0	0	0	0	1	0	0	5	0
E	0	0	8	15	0	0	0	0	6	3	1	0	0	0	0	0	0	4	0	0
S	6	0	0	0	0	0	6	7	0	4	0	1	14	0	1	1	0	5	4	0
T	2	0	1	0	0	4	0	0	0	25	0	11	0	17	0	0	0	0	2	0
N	0	4	4	7	0	28	0	0	0	0	0	1	0	0	0	1	0	0	0	0
Q	16	6	3	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
A	0	0	0	3	3	7	35	0	0	0	33	0	0	0	0	0	0	0	2	6
V	0	0	0	0	0	0	0	0	0	7	0	28	11	22	1	0	0	0	2	0
I	0	0	0	0	0	0	6	0	0	0	22	0	6	3	1	0	0	0	0	0
L	2	2	0	0	0	2	0	0	0	0	10	0	0	2	5	0	0	0	5	0
M	1	0	0	0	0	0	11	0	0	0	22	8	7	0	0	0	0	0	0	0
F	0	0	0	0	0	1	0	0	0	0	0	0	4	0	0	1	0	0	0	0
Y	0	3	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	2	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
C	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	0
G	14	0	0	3	5	23	0	0	0	7	1	0	0	0	0	0	0	0	0	0
P	5	0	0	0	0	13	8	0	0	8	0	0	18	0	0	0	0	0	0	0

+ - P NP

Pathogenic

alt.pro	R	H	K	D	E	S	T	N	Q	A	V	I	L	M	F	Y	W	C	G	P
ref.pro																				
R	0	6	0	0	0	1	0	0	3	0	0	0	6	0	0	0	3	5	6	6
H	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
D	0	4	0	0	1	0	0	0	0	4	0	0	0	0	0	0	0	2	0	0
E	0	0	5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0
S	3	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	0	0
T	2	0	1	0	0	0	0	1	0	0	2	0	1	0	0	0	0	0	1	0
N	0	1	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
A	0	0	0	1	3	2	2	0	0	0	1	0	0	0	0	0	0	0	0	1
V	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
L	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
M	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	2	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Y	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0
C	5	0	0	0	0	1	0	0	0	0	0	0	0	0	4	9	1	0	4	0
G	12	0	0	10	5	7	0	0	0	3	6	0	0	0	0	0	0	6	0	0
P	1	1	0	0	0	1	0	0	0	0	0	0	4	0	0	0	0	0	0	0

+ - P NP

Hardest variants

“Hard” = Everybody disagreed with us

90/101 variants that nobody got right relied on internal Invitae data

Of the other 11: 6 missense, 2 frameshift, 2 intronic, and 1 synonymous variants.

NM_001142800.2(EYS):c.3809T>G

AddOn:Individual: Variant in trans with LP/P variant in affected individual

P 1

Individual: 3 unrelated case reports

P 2

NM_000816.3(GABRG2):c.1000G>A 1

Individual:Dominant: Heterozygous in unaffected adult, INTERMEDIATE penetrance or LATE onset

B 2

RQ [REDACTED] unaffected parent

Individual:Dominant: Heterozygous in unaffected adult, INTERMEDIATE penetrance or LATE onset

B 2

RQ [REDACTED] variant in unaffected mother (reviewed clinic note in proband's chart; [REDACTED])

SLC25A21, Exon 7, NM_030631.3:c.511G>A (p.Gly171Arg)

4 pathogenic calls from
participants

Protein predictors all point
towards pathogenic...

But roughly 1.2% AF in Ashkenazi
Jewish gnomAD data, including 2
homozygotes

Evidence Type	Scores
Not a CVID	B 0
Not in HGMD	B 0
Population:MAF:RECESSIVE: Very high in gnomAD	B 5
Not previously observed in affected individuals/families	B 0
InSilico: General protein predictors: all deleterious	P 0.5

Qualitative Scoring

- A topic for group discussion!
- What are our goals?
 - Fair, objective scoring: current CAGI
 - But also:
 - Scientific contribution
 - Clinical outcomes for patients
- We had sent out a proposed rubric relatively late in the game: not used in scoring this year
- Honorable mentions to team Strokach, who had a very well documented Github and showed their testing, and to team njust-monash biohacker, who described the logic behind their model and features in detail.



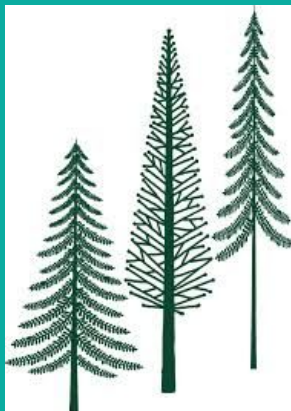
Qualitative scoring rubric

- a. **Sources:** Their model describes all data sources used in a clear format, ideally with links. (10pts)
- b. **Model features and Engineering:** Do they choose rational, biologically plausible features? Do they reason clearly about why they chose a particular model type/architecture? (10pts)
- c. **Scope:** Do they define their model scope appropriately? Can they articulate where their model should and should not be used? (10pts)
- d. **Clinical utility:** Do they discuss clinical use of their model, and does their proposed use take into account their expected True and False positive rates appropriately? (10pts)
- e. **Contribution:** Do they discuss what lessons other scientists can learn from their modeling work, or what directions it suggests for future research? (20pts)



Goals for next CAGI

- Explore private data-sharing via Invitae's homomorphic encryption tools
- Pilot open-source model API:
 - Allow participants to share their model without revealing the code
 - Enable queries for non-pre-computed variants
 - Allow retraining and re-evaluation to keep model “evergreen”



Questions?

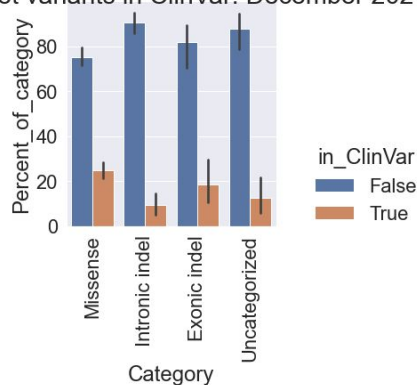


Appendix

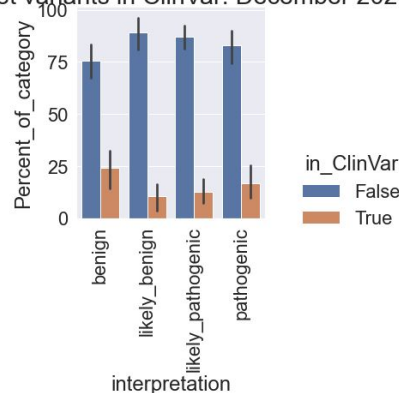
ClinVar Overlap

- Across subcategories, missense variants overrepresented in ClinVar
- Across interpretations, benign variants overrepresented, LB/LP had low overlap.
- Removing variants present in ClinVar did not greatly change maximum scores

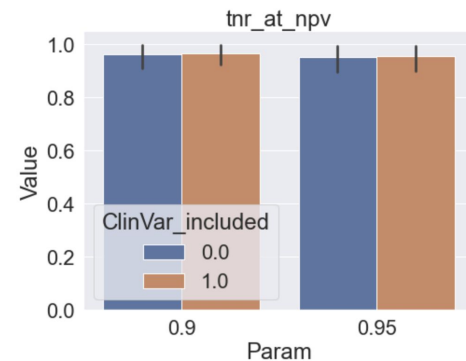
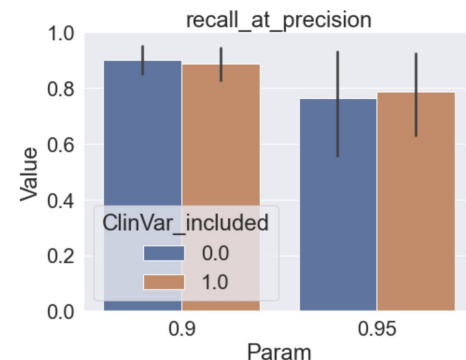
Test set variants in ClinVar: December 2021



Test set variants in ClinVar: December 2021



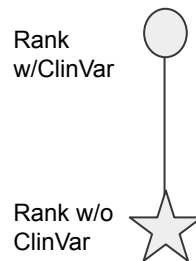
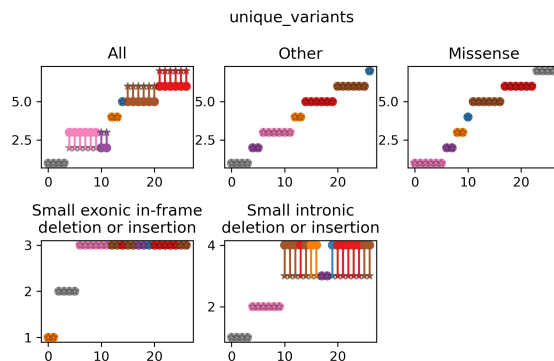
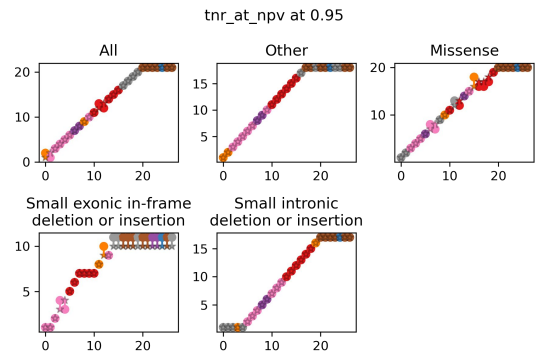
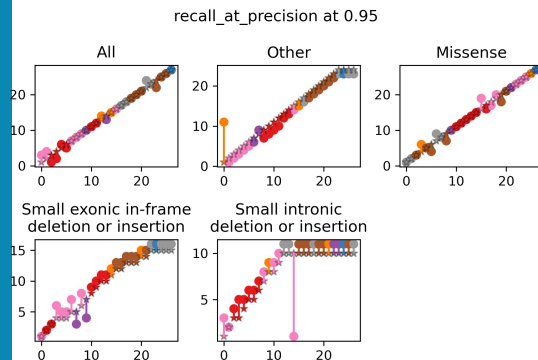
Maximum scores, with and without ClinVar variants



ClinVar Overlap

In most cases, scoring all variants vs excluding those in clinvar does not change overall ranks by much

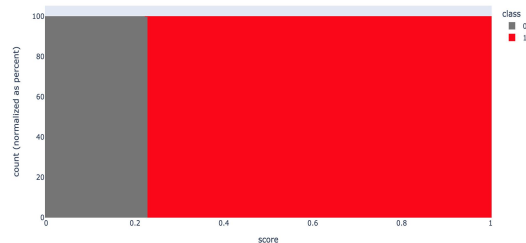
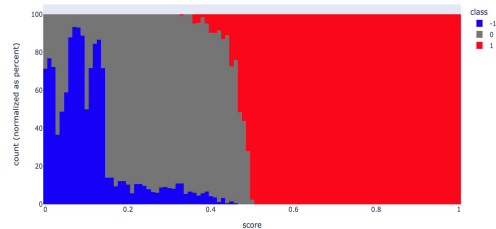
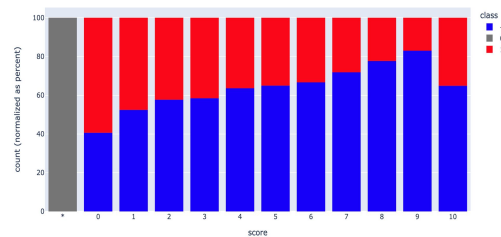
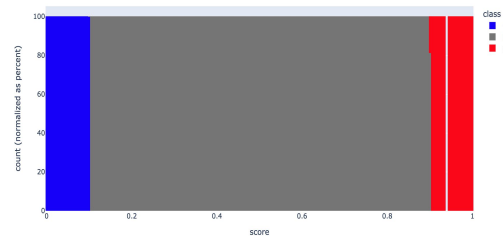
A few cases where in a subcategory, recall at 95% precision ranks change drastically



Confidence scores

We did not request a specific format for the scores, but did note it “may be used as a confidence rank for variants during assessment”

A variety of interpretations from the teams about cutoffs for B/VUS/P

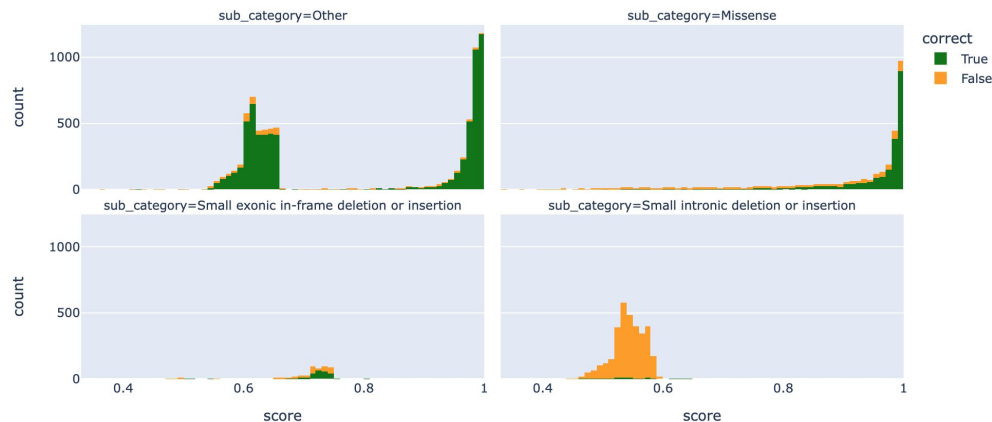
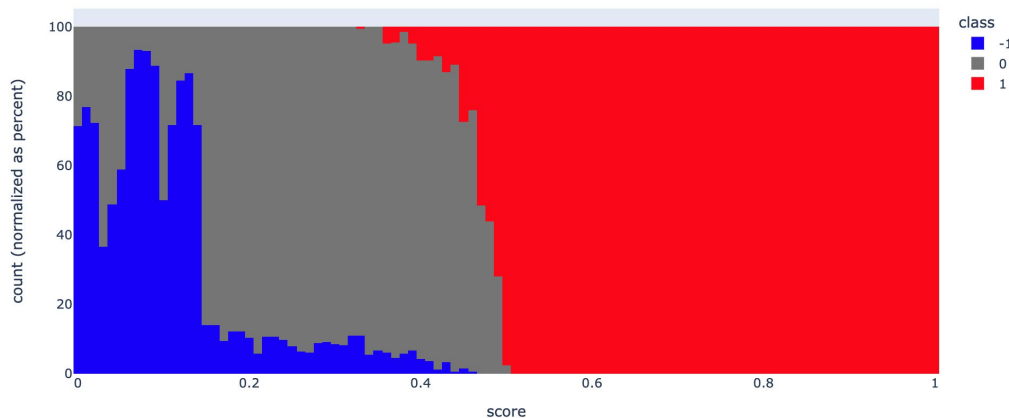


Good accuracy + inconsistent calibration = poor performance

Missense performance in the 80-90% range for pathogenic variants

“Other” category performed very well, recovering 97% of variants at 95% precision.

Overall Recall at 95% Precision was only about 47% due to missense and indels with inflated confidence



PNKP, Exon 17, NM_007254.3:c.1510del (p.Arg504Glyfs*)

1 team predicted benign, others
were VUS or equivocal

Leal et al (2018;
doi:[10.1007/s10048-018-0555-7](https://doi.org/10.1007/s10048-018-0555-7))
found that a termination codon
downstream of this one is
pathogenic

Evidence Type	Scores	@@	Notes
Not previously observed in affected individuals/families	B 0		pre-placed by CROP
Miscellaneous pathogenic evidence	P 0		RQ - epilepsy and microcephaly (not quite CRC), in trans with
Indirect: Fully encompasses a P/LP variant, LoF-established	P 5	Y	[UPGRADE VUS>P 4/1/21] p.Gln517* PMID: 30039206 (3 unrelated case reports + strong segregation) SOURCE: 30039206
Effect: Frameshift leading to an extension [LoF-est] OR [DN-est]	P 2.5		Added new Effect EV (EV0263) here and removed EV0019 "Effect: Truncating variant, may escape NMD [LoF-est] OR [DN-est]" previously used for Extensions.

NM_000384.2(APOB) :c.9176G>A

6 teams predicted B, 1 made no prediction

This variant is in a hot spot of residues that are recurrently targeted by disease-associated variants

Indirect: Disruption of essential amino acid residue

P 1.5

Individual: 2 unrelated case reports

P 1

Miscellaneous pathogenic evidence

P 0

Effect: Missense

P 0

Population:RECESSIVE: Allele count within pathogenic range (gnomAD)

P 1

Population:DOMINANT: Allele count within pathogenic range (gnomAD)

P 4

InSilico: Protein predictions - conflicting or insufficient data

P 0

Database: Variant present in ClinVar

B 0

InSilico FMP: Highly predictive pathogenic score

P 2

GBE1, Exon 11, NM_000158.3:c.1414 A>G (p.Lys472Glu)

3 teams predicted pathogenic

Within the FMP, functional, sequence, and physicochemical constraints are evaluated within representative phylogenetic families to determine patterns of constraint that are customized for each gene. Such customizations are inherently inaccessible to genome-wide interpretation engines or sequence-based modeling approaches and avoid generalizations of mechanisms of pathogenicity across biomolecules.



Evidence Type	Scores
Not a CVID	B 0
Not in HGMD	B 0
Population:Hom:MODERATE: 1 homozygote or hemizygote in gnomAD	B 1
InSilico FMP: Highly predictive benign score	B 2
Population:RECESSIVE: Allele count within pathogenic range (gnomAD)	P 1
Not previously observed in affected individuals/families	B 0
InSilico: Protein predictions - conflicting or insufficient data	P 0
Population:MAF:RECESSIVE: Somewhat high allele count: >8 in ExAC	B 1

PDE6B, Exon 10, NM_000283.3:c.1384G>A (p.Glu462Lys)

3 teams predicted pathogenic

0.2% allele frequency in gnomAD
South Asian population

Evidence Type	Scores
Not in HGMD	B 0
Not a CVID	B 0
InSilico: Protein predictions - conflicting or insufficient data	P 0
Population:Hom:MODERATE: 1 homozygote or hemizygote in gnomAD	B 1
Population:RECESSIVE: Allele count within pathogenic range (gnomAD)	P 1
Population:MAF:RECESSIVE: High in gnomAD	B 3
Population: Present in gnomAD but does not meet any threshold	P 0

EXT2, Exon 11, NM_207122.1:c.1685 G>A (p.Arg562Gln)

2 teams predicted pathogenic

0.23% AF in 1K Genomes AFR,
0.10% AF in gnomad AFR

Evidence Type

Scores

Not previously observed in affected individuals/families

B 0

InSilico: Predicted gain of a de novo splice site
(AlamutSplice)

P 0.5

Database: Variant present in ClinVar

B 0

Population:MAF:DOMINANT: High in gnomAD

B 3

InSilico: Protein predictions - conflicting or insufficient data

P 0

Not a CVID

B 0

Not in HGMD

B 0

Miscellaneous benign evidence

B 0

BRCA1, Intron 21, NM_007294.3:c.540 7-25T>A (Intronic)

4 teams predicted benign

Observed in Høberg-Vetti et al
(2016; 10.1038/ejhg.2015.196) to
affect splicing

Other evidence agrees, but is
overridden



INVITAE

© 2021 Invitae Corporation. All Rights Reserved.

Evidence Type	Scores
InSilico: Predicted gain of a de novo splice site (AlamutSplice)	P 0.5
Miscellaneous pathogenic evidence	P 0
Miscellaneous pathogenic evidence	P 0
Individual: Observed in affected individuals not meeting case report criteria	P 0
ObservedImpact: Altered splicing (in vitro or in vivo), out-of-frame, may escape NMD [LoF]	P 3
Indirect: Loss of a functionally important domain	P 1.5
VariantDetails: Variant described with alternate nomenclature	P 0
Effect: Intronic variant outside the consensus splice site	B 2
Database: Variant is absent from clinically useful locus-specific databases	B 0
Population:DOMINANT: Allele count within pathogenic range (gnomAD)	P 4
InSilico: SpliceAI: Predicted splice site loss	P 4

ARSA, Exon 7, NM_000487.5:c.1197 C>G (p.His399Gln)

No team made a consistent prediction

biophysical analyses that examine the relationship between pathogenicity and the change in the Gibbs free energy of folding ($-\Delta\Delta G$) caused by missense mutations on molecular structures, as estimated from simulations employing efficient, purpose-specific, empirical force fields for statistical energy calculations

Evidence Type	Scores
Not a CVID	B 0
Not in HGMD	B 0
Population:RECESSIVE: Allele count within pathogenic range (gnomAD)	P 1
Indirect: Disruption of essential amino acid residue	P 1.5
InSilico: Protein predictions - conflicting or insufficient data	P 0
Population: Allele count within pathogenic range (ExAC)	P 0.5
InSilico FMP: Highly predictive pathogenic score	P 2

LZTR1, Exon 1, NM_006767.3:c.27dup (p.Gln10Alafs*24)

2 teams called benign

Reported in trans with
c.1943-256C>T in two brothers
with diagnosis of Noonan
Syndrome, exomes were done.
Deep intronic variant is P with
multiple publications/observations
in ClinVar (Broad, GeneDx).

Evidence Type	Scores	@@	Notes
Effect: Truncating variant, presumed loss of mRNA and protein [LoF-est]	P 5		Replaced EV0209 with EV0016, based on LOF established.
Miscellaneous pathogenic evidence	P 0		***Upgrade, VUS>P, based on change in gene curation (now LOF established) and PMID 32623905. 4/11/21***
Segregation: Weak segregation with disease	P 1		PMID 32623905. AR Noonan syndrome
Individual: 1 case report	P 0	Y	PMID 32623905 SOURCE: 32623905

P3H1, Intron 14, NM_022356.3:c.2055+18G>A (Intronic)

4 teams predicted benign

Case reports and experimental
splicing data point towards
pathogenic

Evidence Type	Scores	@@	Notes	Action
Not a CVID	B 0		pre-placed by CROP	✎
Effect: Silent and intronic changes outside of the consensus splice sites	B 2		pre-placed by CROP	✎
Population: Allele count within pathogenic range (ExAC)	P 0.6		pre-placed by CROP, Allele count: 1	✎
Database: Variant present in ClinVar	B 0		pre-placed by CROP	✎
Individual: 2 unrelated case reports	P 1	Y	PMID: 19088120 >> one of these is in trans with a P allele, but it's 1 point either way SOURCE: PMID: 27864101, 22615817, 3545499	✎
InSilico: Predicted strengthening of a cryptic acceptor or donor splice site: Strong	P 0.5		Originally "no splicing change" was placed, but I think this counts as "strong" change. MES = +20.9%, NNSPLICE = +41.7%, and HSF = +5.3%, which would be an average of 22.6% (>10% required for "strong" activation)	✎
ObservedImpact: Altered splicing (in vitro or in vivo), out-of-frame, may escape NMD [LoF]	P 3	Y	SOURCE: 19088120	✎
Miscellaneous pathogenic evidence	P 0		[upgrade] LP>P due to EV0333 (obs RNA altered splicing OOF, NMD-). 05/20/21	✎
Indirect: Truncation/Extension, 3 unrelated case reports with downstream C-terminus loss	P 1.5	Y	Essential domain (aa722-763): There is a motif at the very end of the protein (aa733-736) that is only present in the clinically relevant transcript (P3H1a) and not the two alternate transcripts (b/c). There are several variants that show the importance of this region, which I'm combining using the "overlapping CNV" concept. c.2164C>T (p.Q722*) = absent in ExAC, found but with c.10E 1704del (p.D320fs*16) in an affected individual, no protein expression SOURCE: 27864101, 22615817	✎

PMS2, Exon 8, NM_000535.5:c.825 A>G (Silent)

4 teams called benign

Van der Klift et al (2015; doi:
10.1002/mgg3.145) observed in
vivo and in vitro splicing changes

Likely pathogenic in ClinVar

Evidence Type	Scores	@@	Notes
InSilico: Predicted gain of a de novo splice site (AlamutSplice)	P 0.5		pre-placed by CROP. localSpliceEffect: New Acceptor Site
AddOn:Individual: Variant in trans with LP/P variant in affected individual	P 1	Y	PMID: 21261604 SOURCE: PMID: 28514183, 21261604, 26110232, 25856668
Experimental: In vitro splicing: Evidence for resulting in NMD+ variant	P 1		PMID 26247049 - minigene assays using two different vector system demonstrated exonic exclusion of the first 22 nt (figure S3). Comparable amount of mutant transcript expression with wild-type controls, so would be more than 90%. SOURCE: 26247049
ObservedImpact: Altered splicing (in vivo), NMD expected, observed in 1 individual [LoF]	P 3		PMID 26247049, delta 8p-22nts
Miscellaneous pathogenic evidence	P 0		02/2021 upgrade LP>P based on application of new RNA criteria, both EV330 and EV331 were applied
Effect: Silent variant not at end or beginning of exon	B 2		
Population:RECESSIVE: Allele count within pathogenic range (gnomAD)	P 4		pre-placed by CROP
Population:DOMINANT: Allele count within pathogenic range (gnomAD)	P 4		pre-placed by CROP
InSilico: SpliceAI: Predicted gain of a de novo splice site	P 4		pre-placed by CROP



ABCA4, Intron 13, NM_000350.2:c.1937+435C>G (Intronic)

4 teams predicted benign

Unrelated case reports and in vitro evidence in Sangermano et al (2019, doi: 10.1038/s41436-018-0414-9)

Evidence Type	Scores	@@	Notes
Not a CVID	B 0		pre-placed by CROP
Not in HGMD	B 0		pre-placed by CROP
Population: Insufficient coverage in ExAC	P 0		pre-placed by CROP. Allele count: 0; coverage 20x in -1.0%
Population: Absent in gnomAD	P 1		pre-placed by CROP
InSilico: General splicing predictors: raw data below thresholds for loss, gain and cryptic activation	B 1		pre-placed by CROP
Effect: Silent and intronic changes outside of the consensus splice sites	B 2		pre-placed by CROP
Experimental: In vitro splicing: Evidence for resulting in NMD+ variant	P 1		PMID: 30643219 SOURCE: 30643219
Individual: 3 unrelated case reports	P 2	Y	PMID: 30643219-2 CASES;RQ has dx of Stargardt, SOURCE: PMID: 30643219, Invitae
AddOn:Individual: Variant in trans with LP/P variant in affected individual	P 1		PMID: 30643219