

CAGI slide pack guidelines

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

Melissa Cline

- 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 2013. 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, but bear in mind that we have limited means of enforcing them, and therefore the restrictions cannot be guaranteed.



Germline Cancer Challenge: ENIGMA Assessment

Melissa Cline

July 8, 2018

ENIGMA: ClinGen Expert Panel for BRCA1 and BRCA2

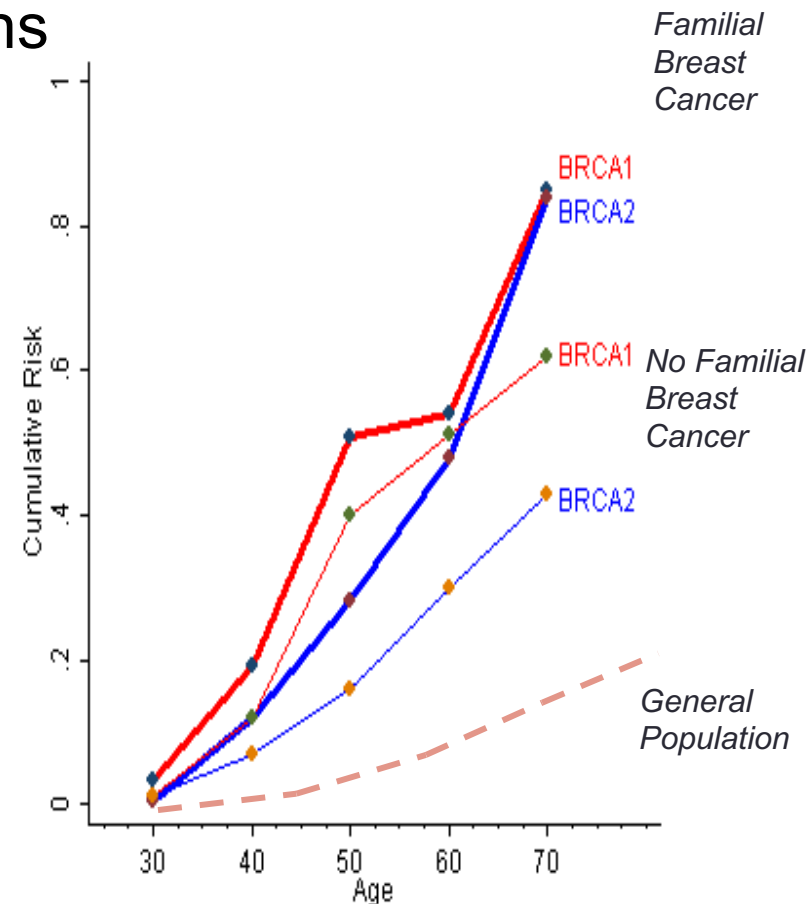
**Evidence-based Network for the
Interpretation of Germline Mutant Alleles**

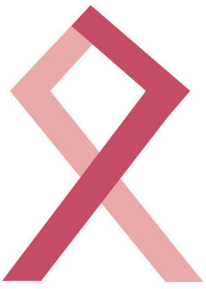
AGATGC**VARIANT**TGAGTTTGT**CLASSIFICATION**GT
GTGAACAGACACTGAAATATTTTCTAGGAATTGCGGGA
GGA**ENIGMA**GACATTC
ATG**ENIGMA**TTTGAA
ATT**ENIGMA**CTCACC
TGCAGAAGAATCTGAACATAAAAACAACAATTACGAAC
C**BREAST****CANCER**AACCTAT**GENES**TTAAAACTCCAC

Motivation for BRCA

BRCA variation is relatively common with well known medical implications

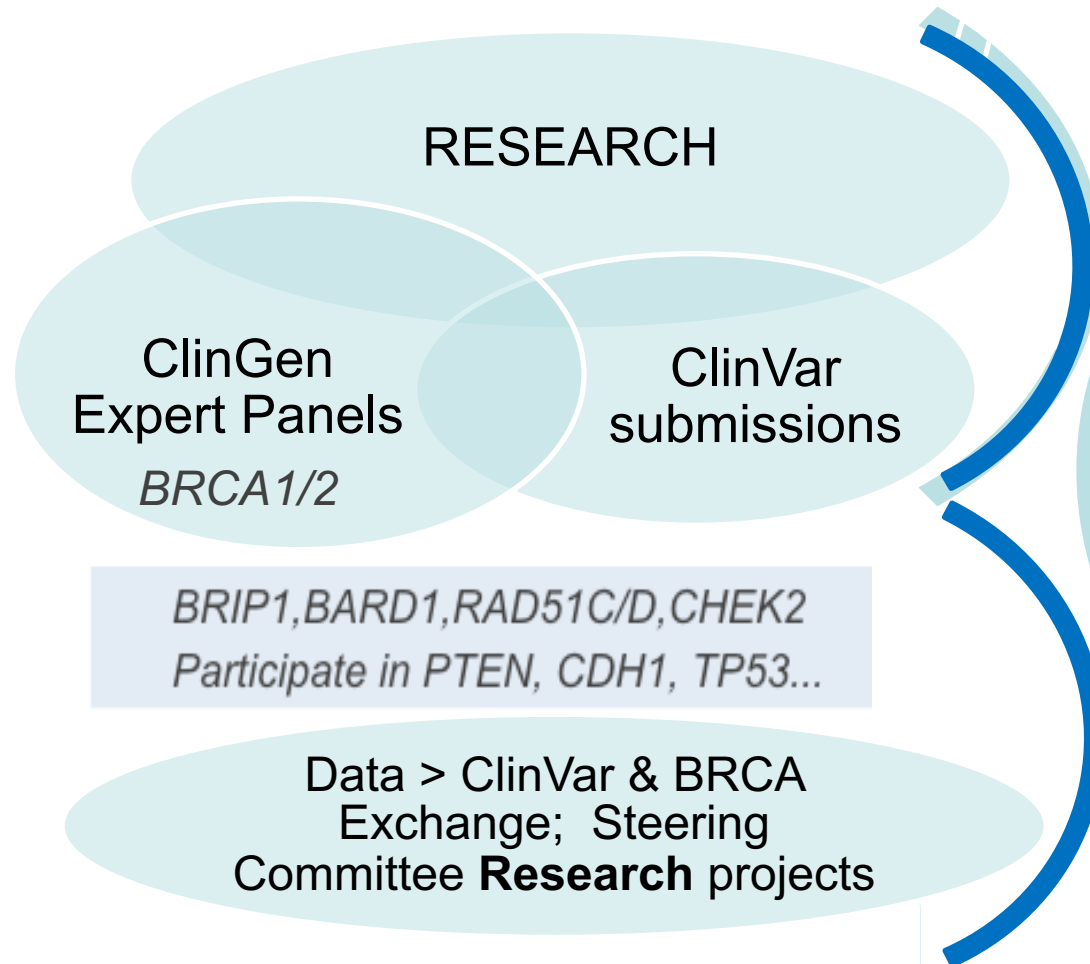
- Lifetime risk of developing breast or ovarian with pathogenic BRCA mutation
- Men with pathogenic BRCA mutations are also at risk for prostate cancer
- Drug treatment: PARP inhibitors show effectiveness for BRCA1/2 patients



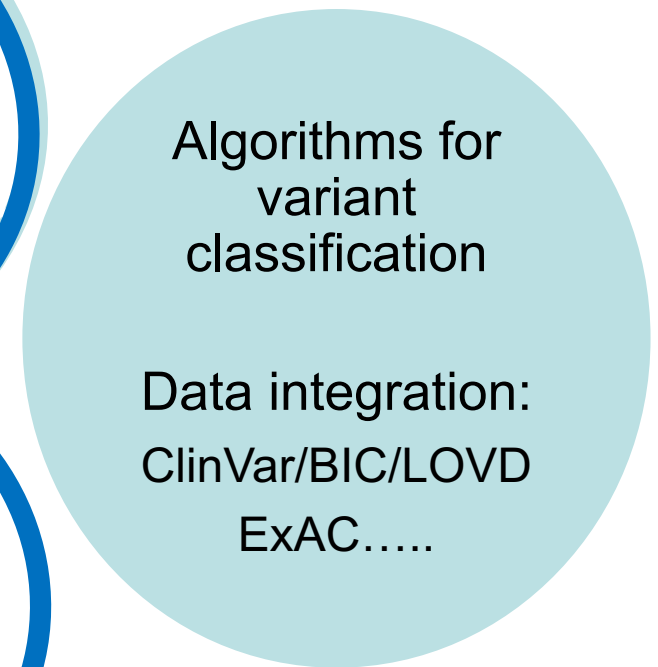


BRCA EXCHANGE: BRCA DATA AGGREGATION AND ANALYSIS

ENIGMA



BRCA Exchange



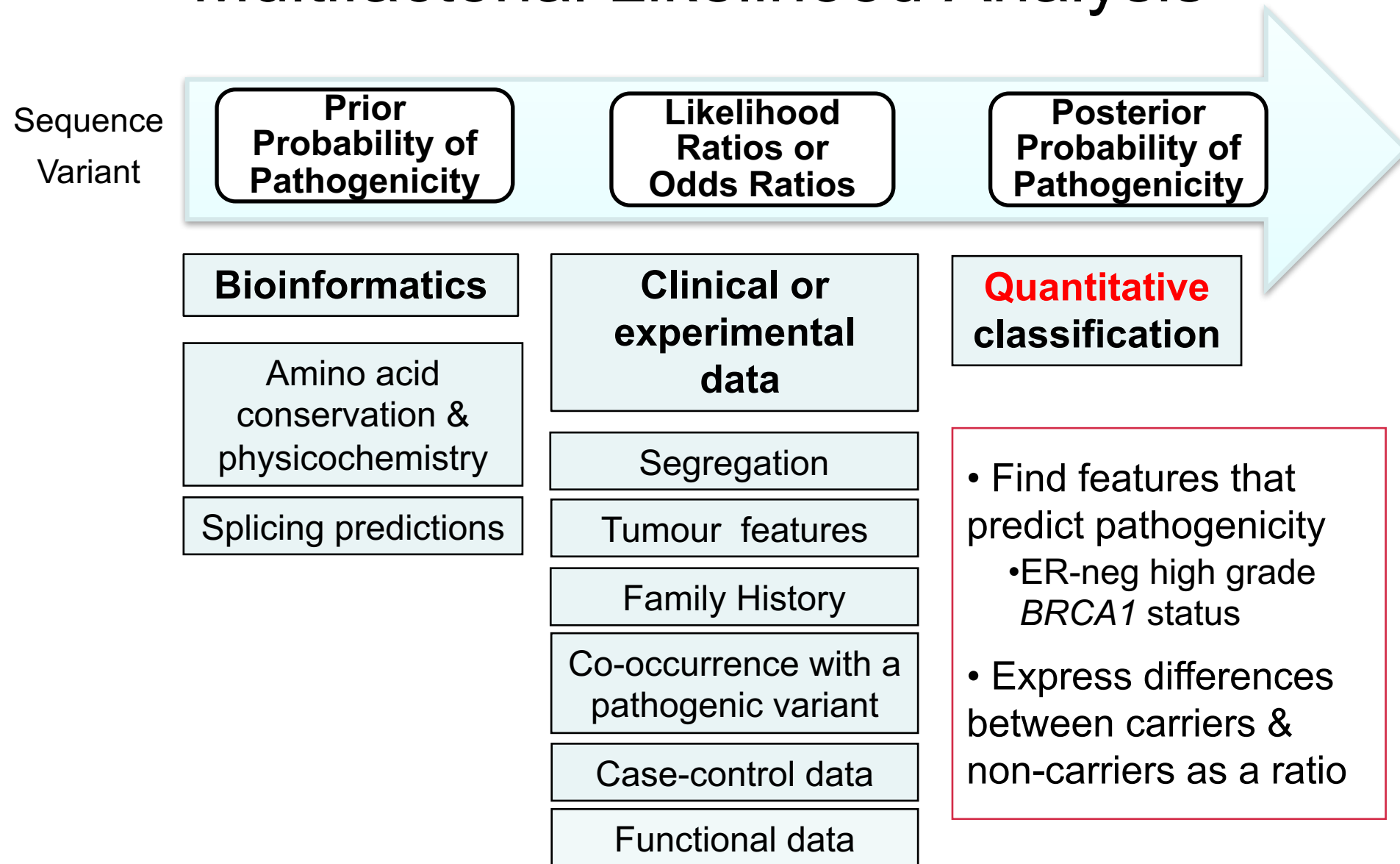
Classification categories

Five-tier systems most used for germline disease risk variants

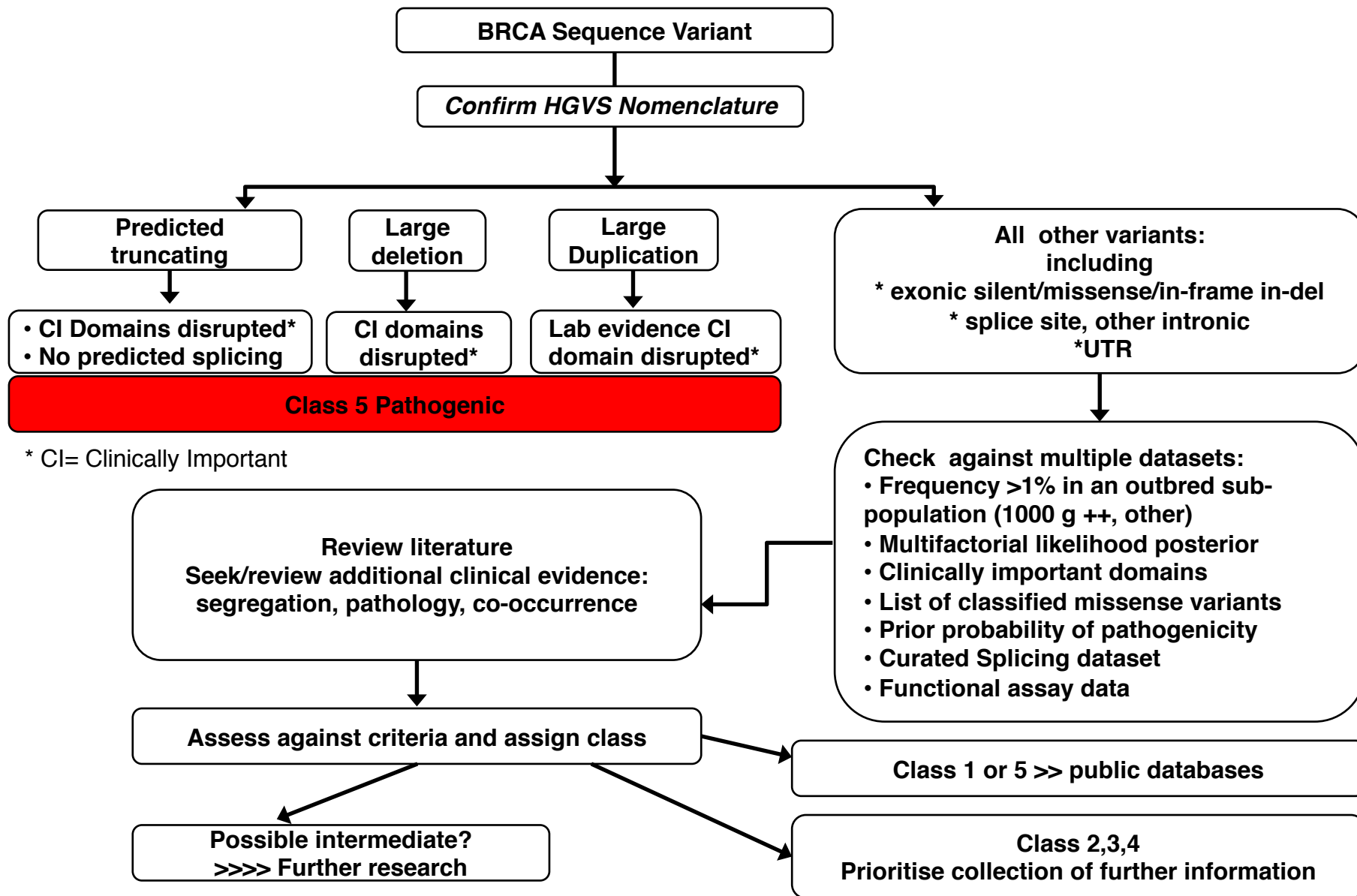
- ACMG – American College of Medical Genetics
 - Richards et al *Genet Med*, 2015, 17: 405
- IARC – International Agency for Research into Cancer
 - Plon et al *Hum Mutat* 2008, 29:1282
 - also promotes **research** to classify variants where necessary

Variant Class	Predictive Testing of relatives	Surveillance for Relatives	Research Testing of Relatives
Pathogenic	Yes	Yes – high-risk	No
Likely pathogenic	Yes	Yes- high-risk	Yes
Uncertain	No	Based on family history & other risk factors	Yes
Likely not pathogenic or Likely Benign	No	Treat as “no pathogenic variant detected”	Yes
Not pathogenic or Benign	No	Treat as “no pathogenic variant detected”	No

Multifactorial Likelihood Analysis



ENIGMA BRCA Classification



Class 5 Pathogenic:

Posterior Probability
>0.99

Nonsense/frameshift
disrupting clinically
important domain(s)

Variant allele producing
only mRNA transcript(s)
that disrupt clinically
important domains

Large deletion of
exon(s) that disrupts
clinically important
domain(s) OR is
proven to lead to a
frameshift

Large duplication
confirmed to encode a
frameshift

Class 4 Likely pathogenic:

Posterior Probability
0.95-0.99

Variant alters initiation
methionine site

Canonical splice site,
untested for allele-
specific splicing
AND not predicted to
produce in-frame RNA
isoforms

Same amino acid
change as known
pathogenic missense
AND no splicing
aberration
AND not in controls

In-frame deletion of
codon for which a
pathogenic missense
has been described
AND low prediction to
alter splicing
AND not in controls

Class 3 Uncertain:

Posterior Probability
0.05-0.949

Canonical splice site
predicted to produce
in-frame RNA
isoforms,
OR
Allele leads to
transcripts that
encode (partially)
functional protein
AND no additional
evidence to classify

Insufficient evidence
to classify

Potential intermediate
risk
as indicated by
multiple apparently
conflicting points of
evidence regarding
variant pathogenicity

Class 2 Likely not pathogenic:

Posterior Probability
0.001-0.049

Same amino acid
change as class 1
missense
AND low prediction to
alter splicing

Synonymous
substitution
AND low prediction to
alter splicing

*More to come soon:
For deep intronic &
promoter/UTR*

Class 1 Not pathogenic:

Posterior Probability
<0.001

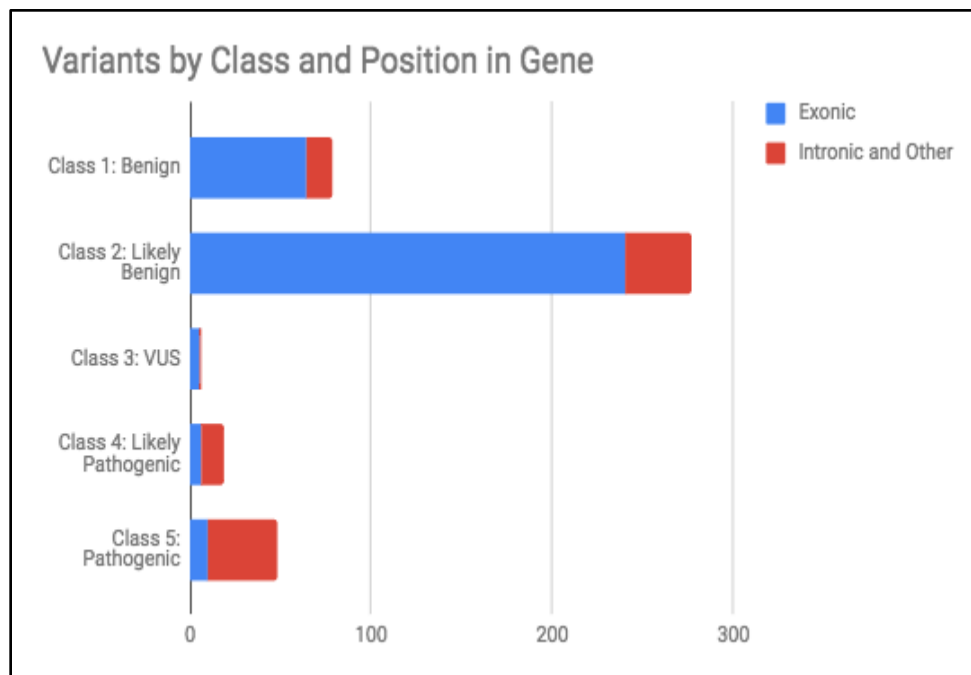
Allele frequency $\geq 1\%$
in reference controls

Same amino acid
change as class 1
missense
AND no splicing
effect proven

Exonic
substitution/ins-
del/synonymous with
prior probability ≤ 0.02
OR intronic
AND low prediction to
alter splicing OR no
splicing effect proven
AND co-occurrence
in trans with
pathogenic and no
Fanconi-like features
OR Allele frequency
 ≥ 0.001 and < 0.01 in
reference controls

Class 5 Pathogenic:	Class 4 Likely pathogenic:	Class 3 Uncertain:	Class 2 Likely not pathogenic:	Class 1 Not pathogenic:
Posterior Probability >0.99	Posterior Probability 0.95-0.99	Posterior Probability 0.05-0.949	Posterior Probability 0.001-0.049	Posterior Probability <0.001
Nonsense/frameshift d in	Variant alters initiation	Canonical splice site		Allele frequency ≥1% controls
Class 5 Pathogenic	Class 4 Likely Pathogenic	Class 3 Uncertain	Class 2 Likely Benign	Class 1 Benign
> 0.99	0.95 – 0.99	0.05 – 0.949	0.001 – 0.049	< 0.001
important domains		<u>AND</u> no additional evidence to classify	Synonymous substitution <u>AND</u> low prediction to alter splicing	Exonic substitution/ins- del/synonymous with prior probability ≤0.02 OR intronic <u>AND</u> low prediction to alter splicing OR no splicing effect proven <u>AND</u> co-occurrence <i>in trans</i> with pathogenic and no Fanconi-like features OR Allele frequency ≥0.001 and <0.01 in reference controls
Large deletion of exon(s) that disrupts clinically important domain(s) <u>OR</u> is proven to lead to a frameshift	Same amino acid change as known pathogenic missense <u>AND</u> no splicing aberration <u>AND</u> not in controls	Insufficient evidence to classify		
Large duplication confirmed to encode a frameshift	In-frame deletion of codon for which a pathogenic missense has been described <u>AND</u> low prediction to alter splicing <u>AND</u> not in controls	Potential intermediate risk as indicated by multiple apparently conflicting points of evidence regarding variant pathogenicity	<i>More to come soon: For deep intronic & promoter/UTR</i>	

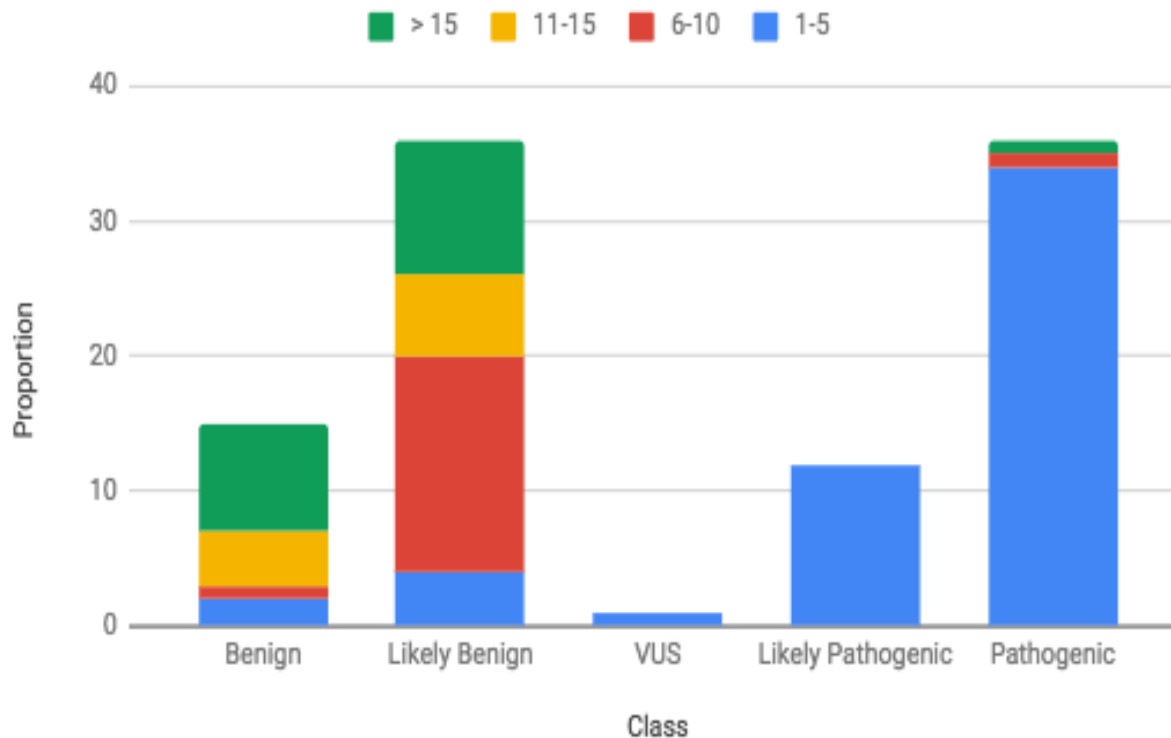
CAGI Prediction Targets



- Mostly missense substitutions
- A few in-frame deletions
- Some insertions, not predicted on
- Skewed for Class 2

Many of the intronic targets were pathogenic splice region variants

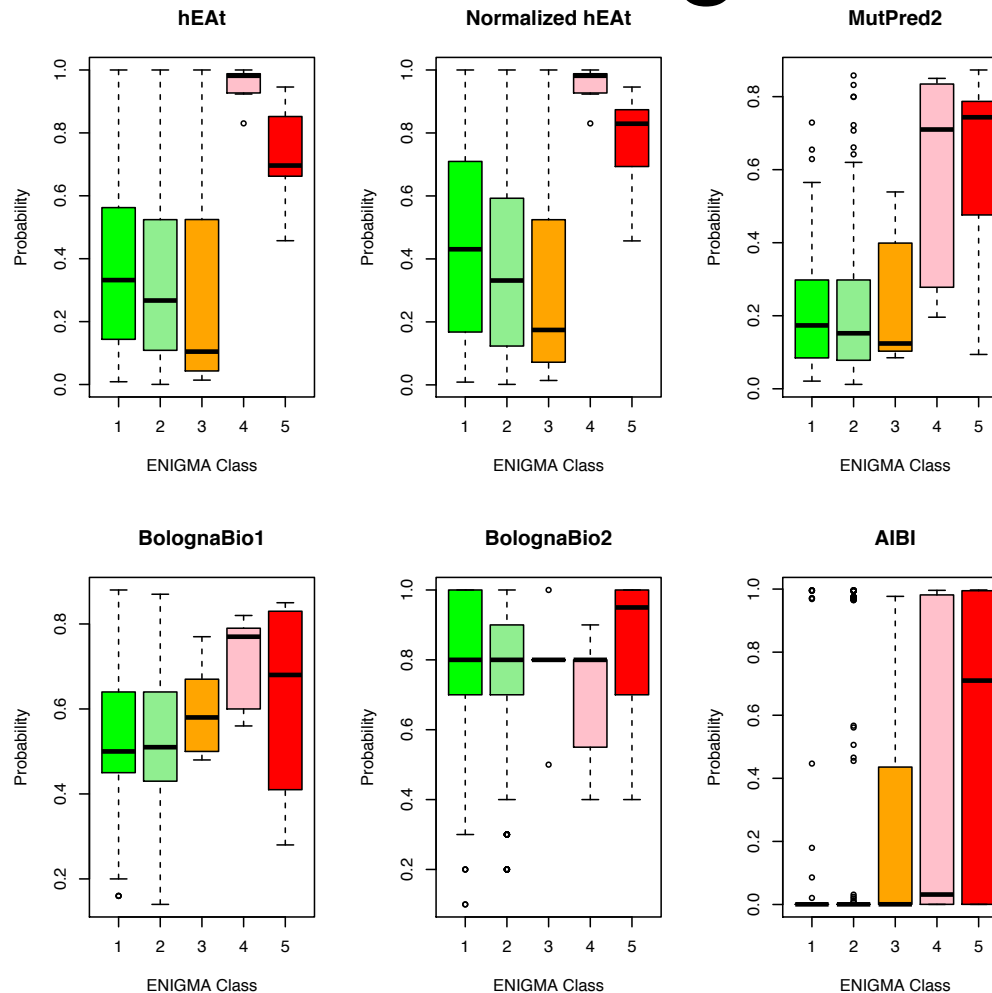
Intronic Variants by Class and Bases from the Splice Site



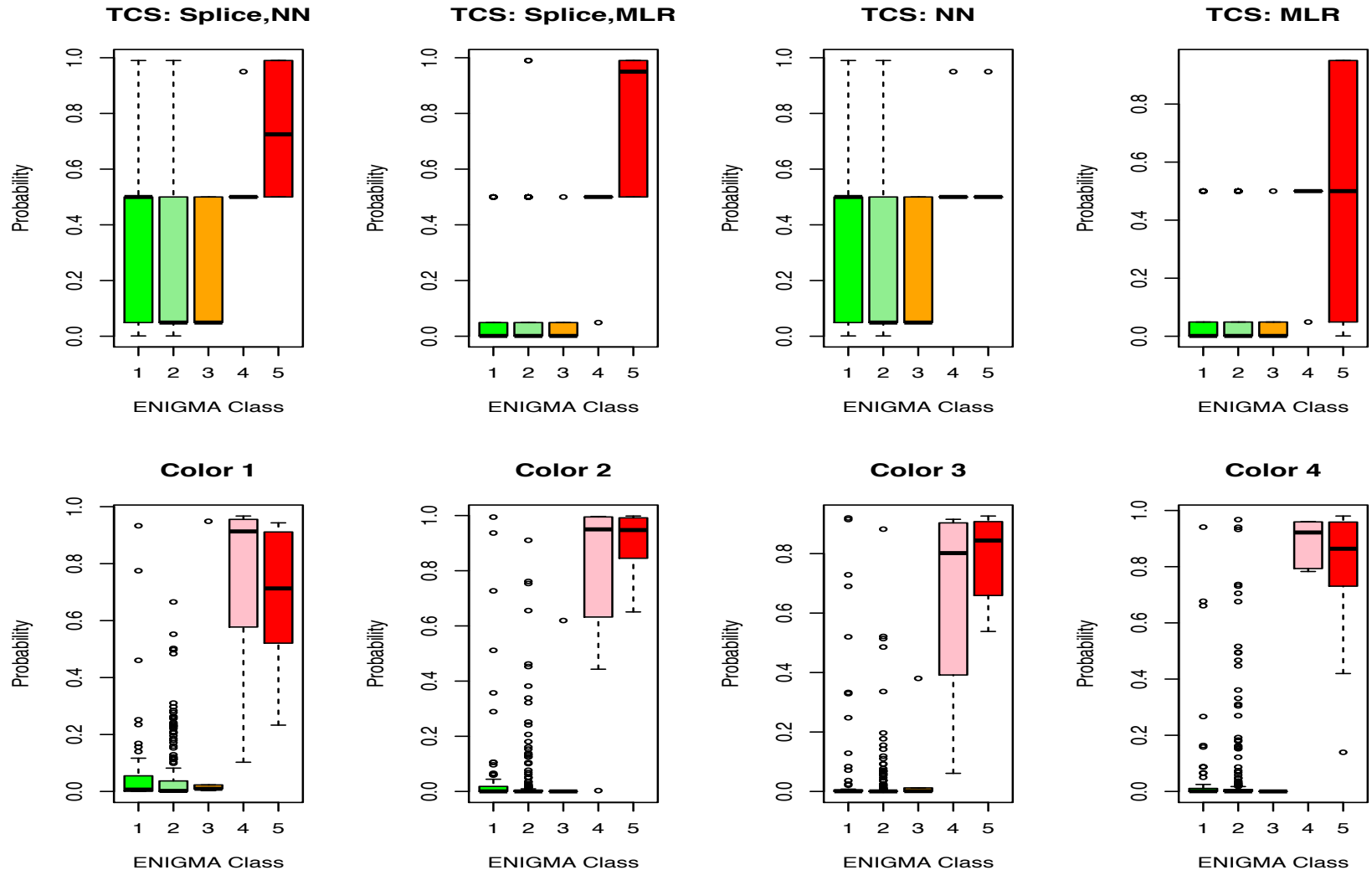
Prediction methods

Team	Methods
Lichtarge	hEAt: Evolutionary Action method, with and without normalization for fraction of affected transcripts
Mooney Radivojac	MutPred2
TCS_IN_02	1. Predicted splicing impact + neural network protein impact prediction 2. Predicted splicing impact + functional assay result prediction 3. Neural network protein impact prediction 4. MLR prediction of functional assay scores
Bologna Biocomputing	1. SNP&Go 2. Disease Index Matrix
AIBI	MutPred2 features, linear regression loss function trained on clinical significance classes.
Color Genomics	Logistic regression and random forest estimation, with and without family history observations

Qualitative results for Lichtarge, MutPred2, BiolognaBio, AIBI



Qualitative results for TCS_IN_02 and Color Genomics



Quantitative Assessment Process

Statistics:

- Multi-class and two-class AUC (Benign, vs. Pathogenic),
- RMSD, Pearson and Kendall Correlation
- Sensitivity, Specificity, Accuracy, F1 at empirically-derived thresholds with best F1

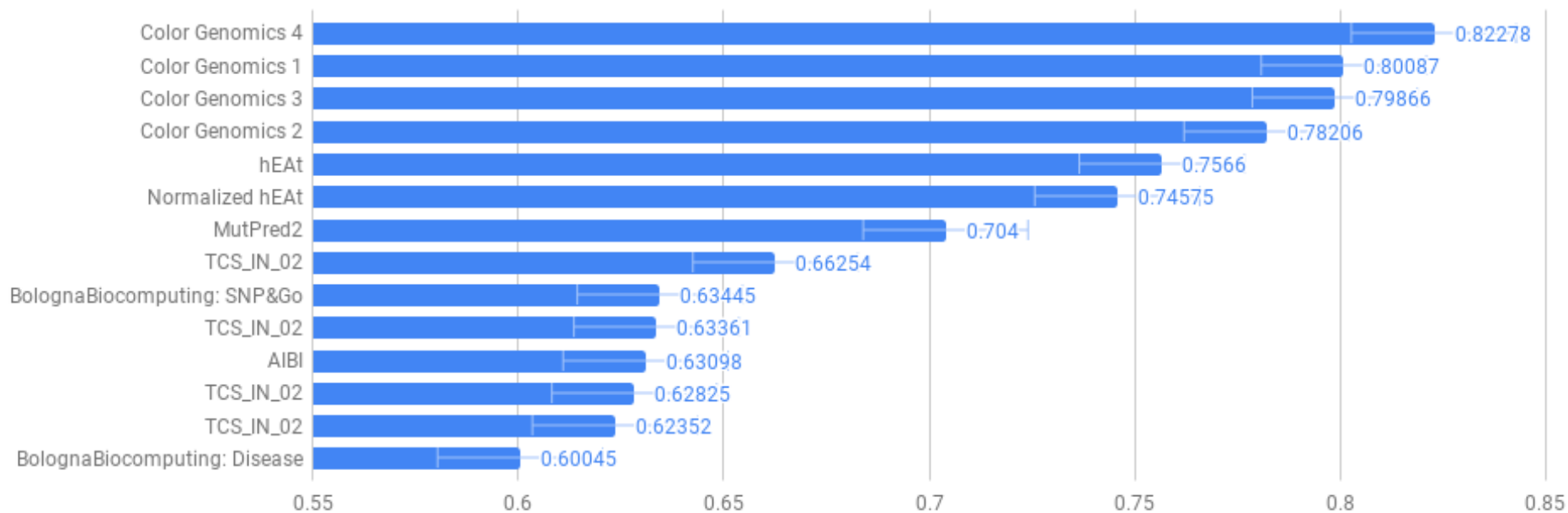
Confidence intervals:

- For N random bootstrap permutations, add Gaussian noise $N(0, SD^2)$
- Re-sample all statistics
- Estimate confidence intervals from the standard deviation of the statistics

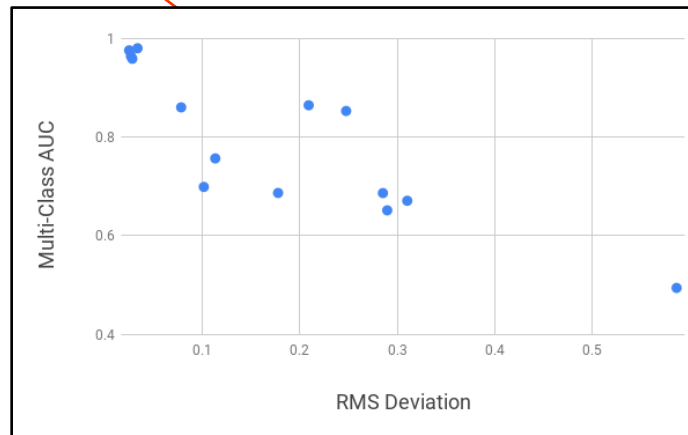
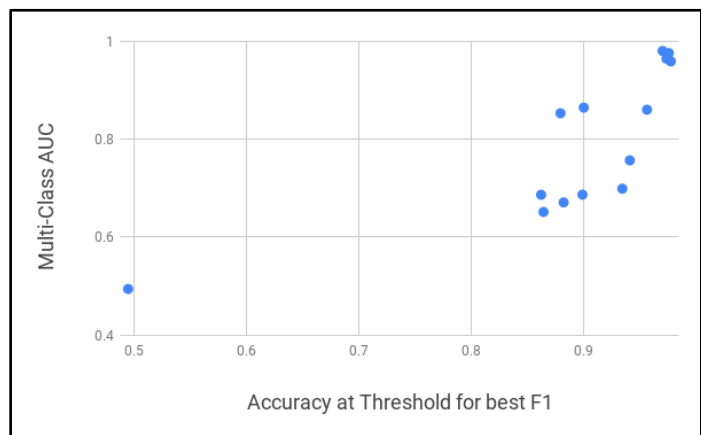
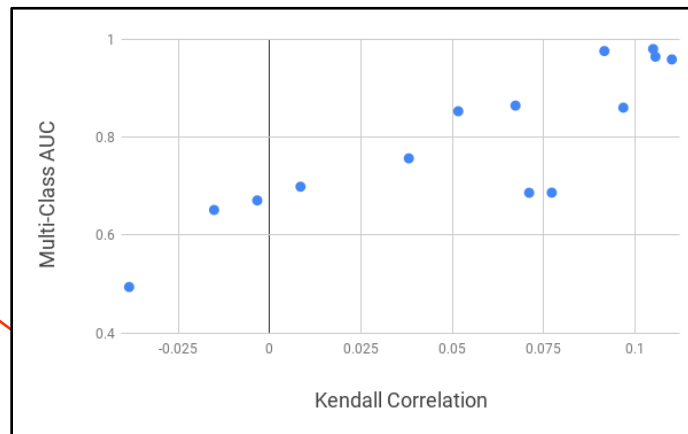
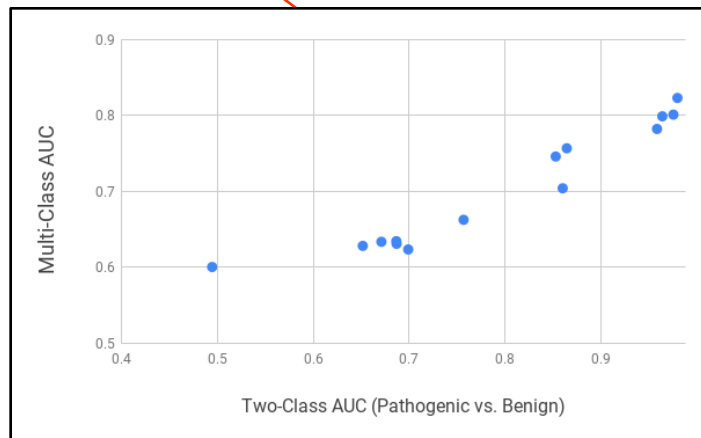
Ranking: by multi-class AUC

Quantitative Results

Multi-Class ROC AUC and Confidence Intervals



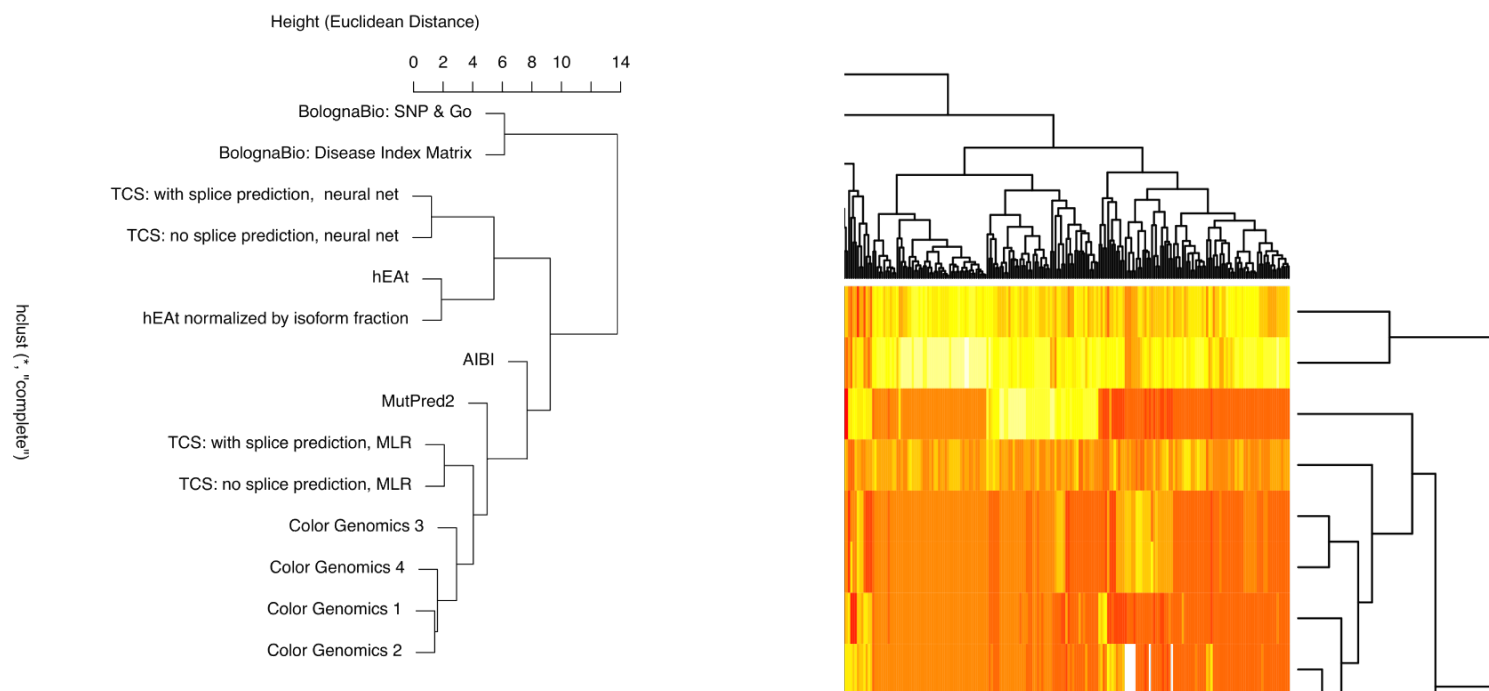
Observed no radical difference between different performance measures



Overall Observations

No single variant confounded all predictors

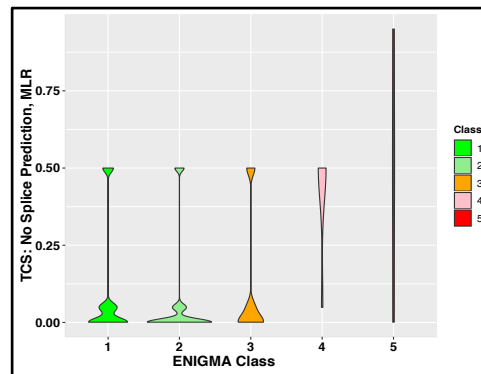
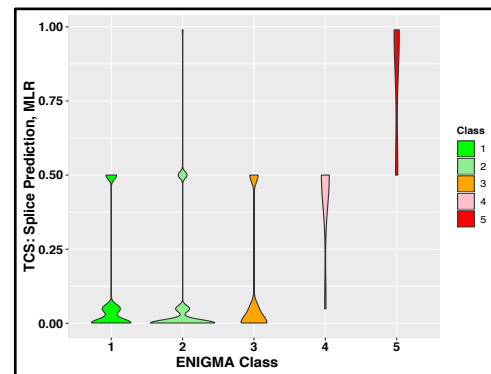
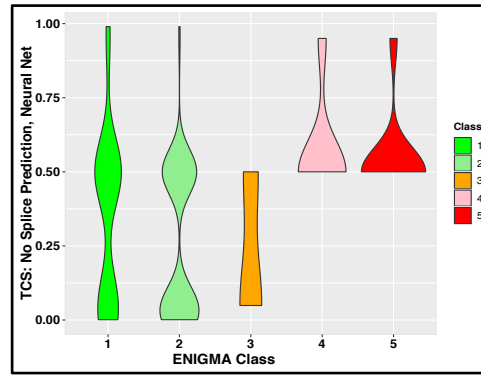
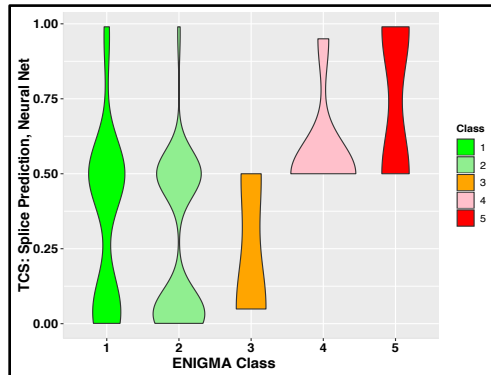
- There is opportunity for meta-prediction
- We can learn by comparing and contrasting methods



Comparing and Contrasting TCS team methods

Splicing Prediction

No Splicing Prediction



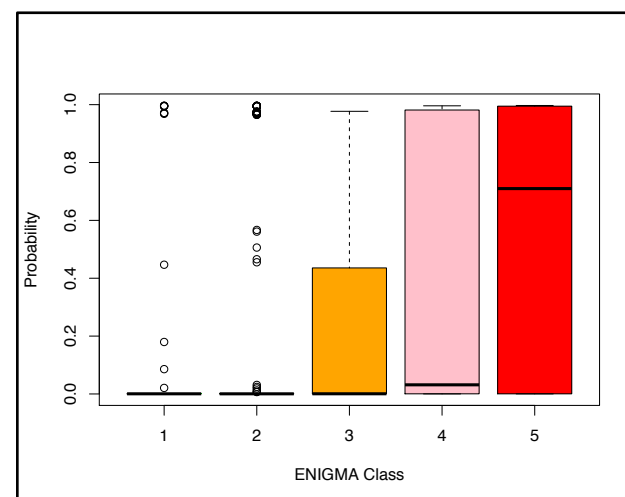
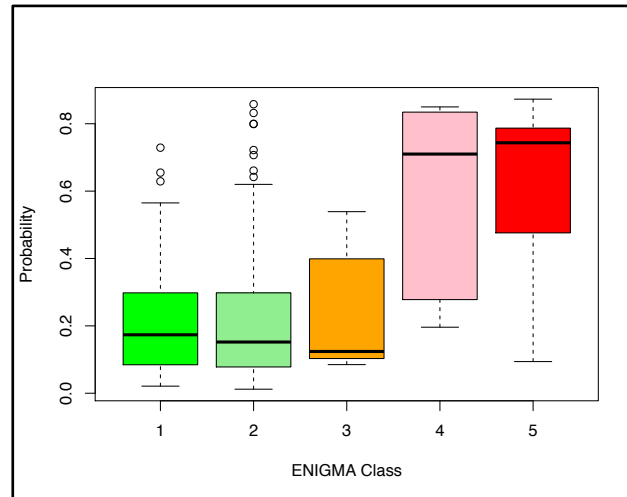
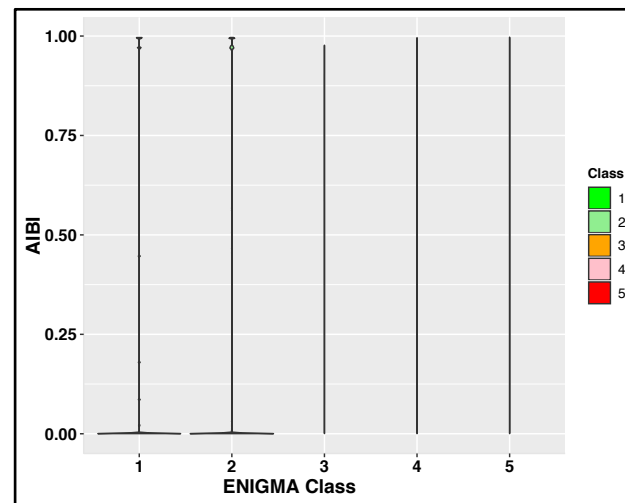
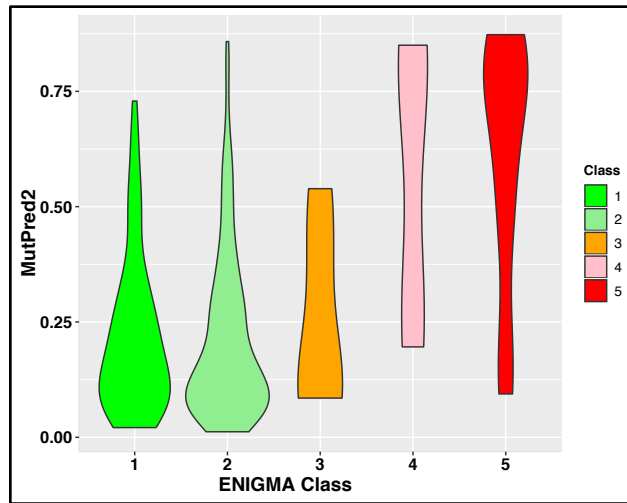
Neural Network

- Trained against variant interpretations
- Additional inputs on functional role of position, amino acid properties

Multiple Linear Regression

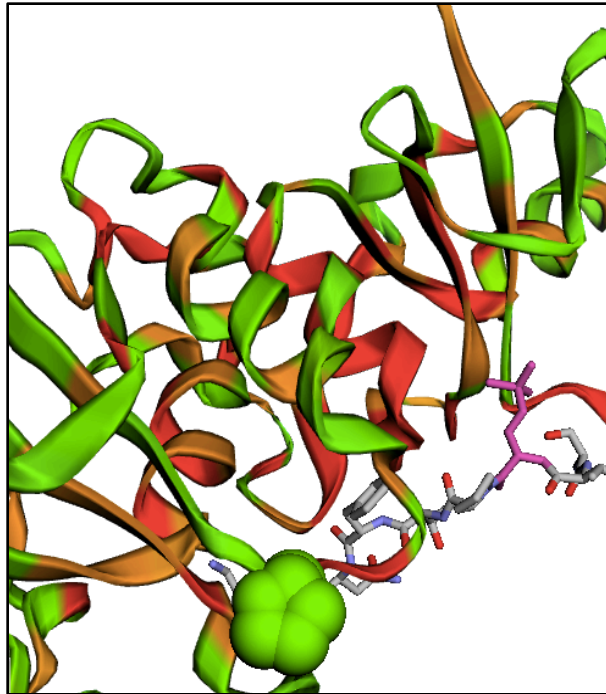
- Trained against functional assay results

Compare and Contrast: MutPred2 vs. AIBI (MutPred2)

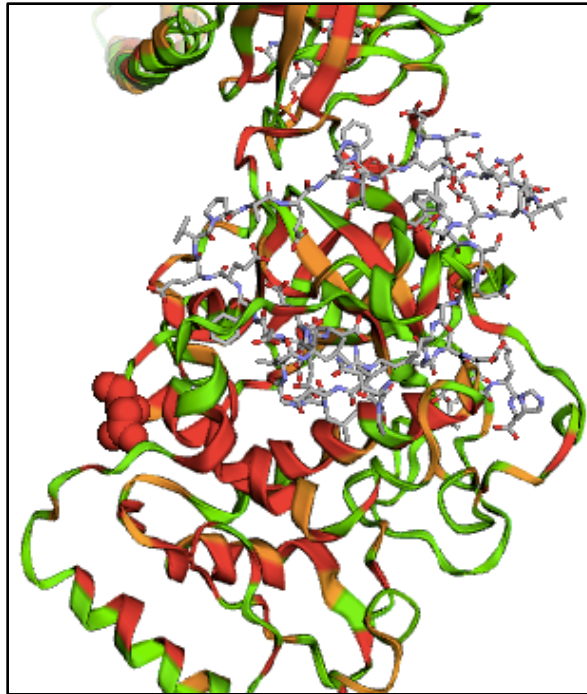


Hard Benign Variants: Conserved positions near surface of 3D structure

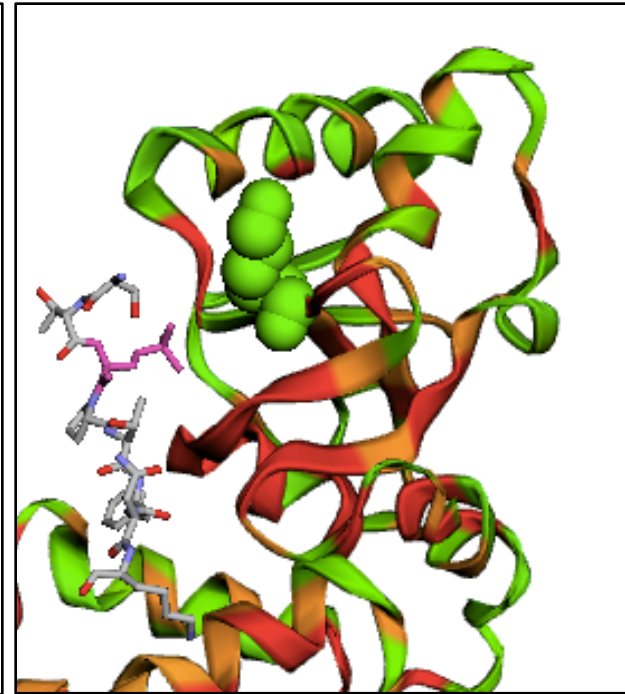
BRCA1 c.5312C>G
Pro -> Arg



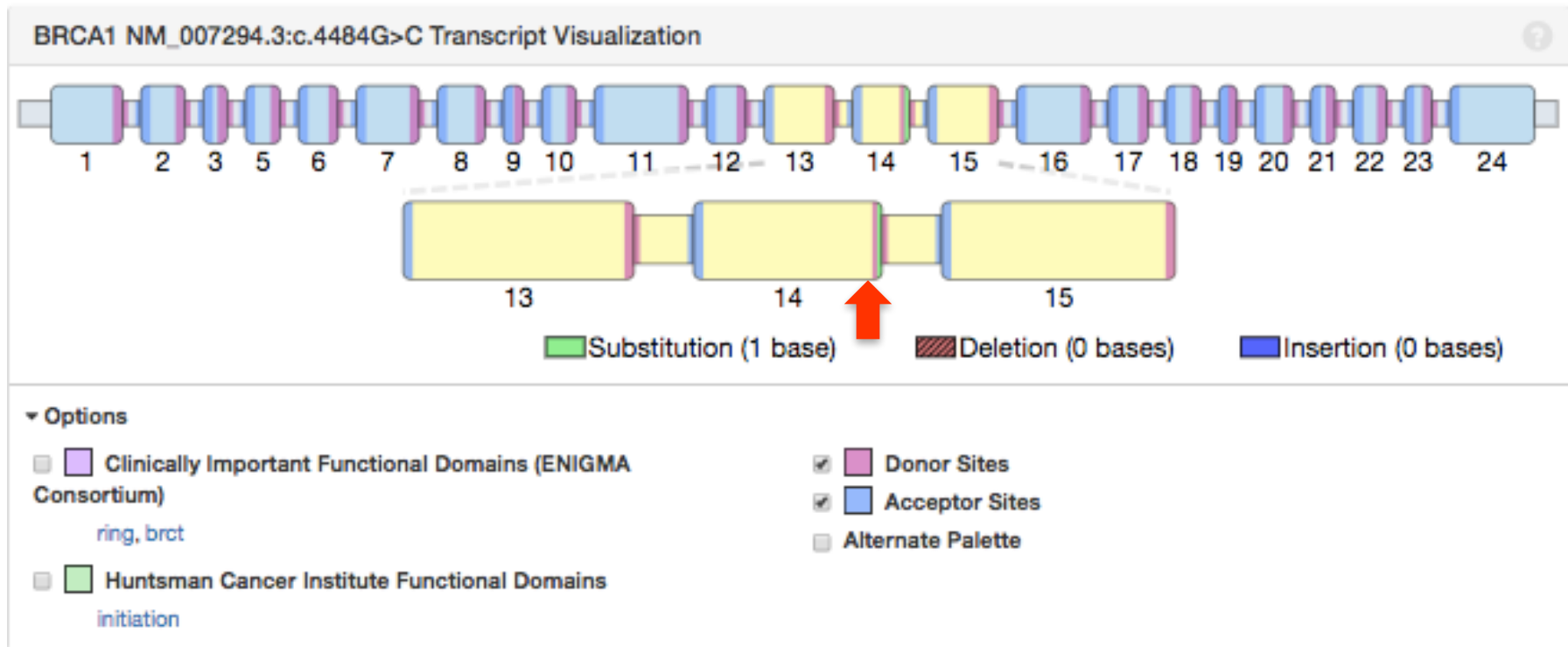
BRCA2 c.7994A>G
Gly -> Asp



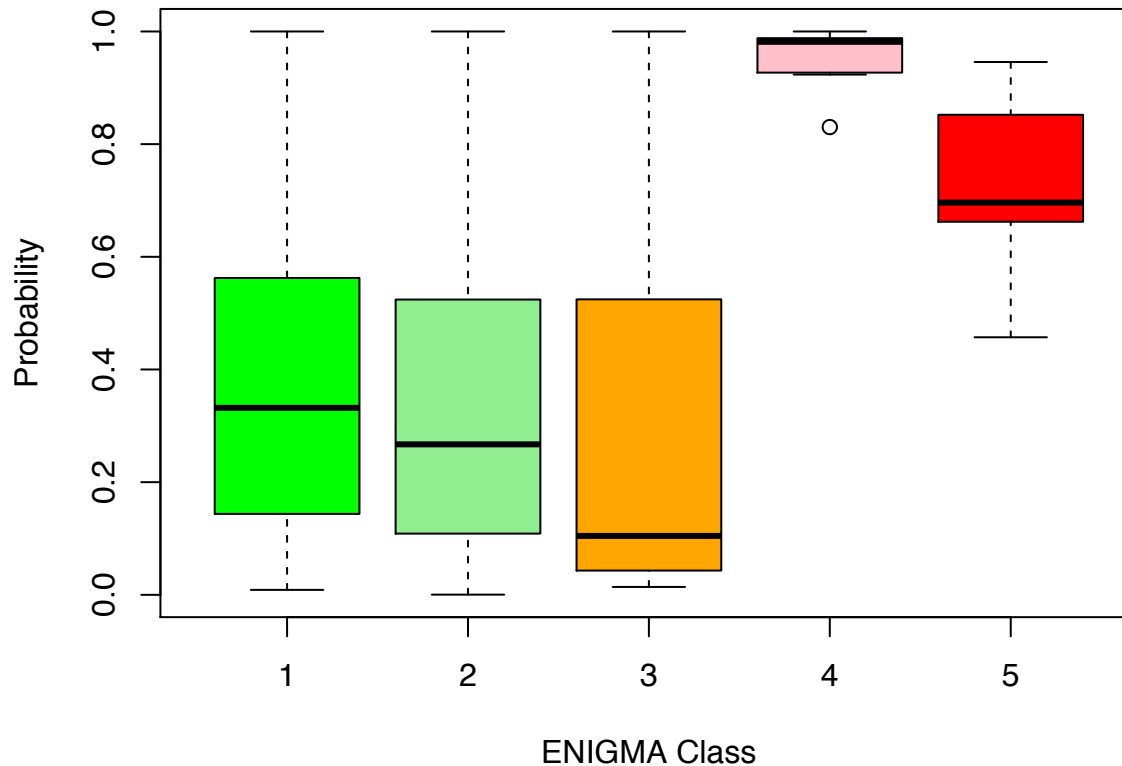
BRCA1:c.5068A>C
Lys -> Gln



Hard pathogenic variants: splice regions



Observation: methods have trouble distinguishing benign from likely benign, pathogenic from likely pathogenic



Rules for clinical significance are different from deleteriousness

Class 5 Pathogenic:

Posterior Probability
>0.99

Nonsense/frameshift
disrupting clinically
important domain(s)

Variant allele producing
only mRNA transcript(s)
that disrupt clinically
important domains

Large deletion of
exon(s) that disrupts
clinically important
domain(s) OR is
proven to lead to a
frameshift

Large duplication
confirmed to encode a
frameshift

Class 4 Likely pathogenic:

Posterior Probability
0.95-0.99

Variant alters initiation
methionine site

Canonical splice site,
untested for allele-
specific splicing
AND not predicted to
produce in-frame RNA
isoforms

Same amino acid
change as known
pathogenic missense
AND no splicing
aberration
AND not in controls

In-frame deletion of
codon for which a
pathogenic missense
has been described
AND low prediction to
alter splicing
AND not in controls

Nonsense/frameshift
disrupting clinically
important domain(s)

Variant allele producing
only mRNA transcript(s)
that disrupt clinically
important domains

Variant alters
initiation
methionine site

Same amino acid
change as known
pathogenic
missense
AND no splicing
aberration
AND not in controls

Rules for clinical significance are different from deleteriousness

Allele frequency $\geq 1\%$ in reference controls

**Same amino acid change as class 1 missense
AND no splicing effect proven**

**Same amino acid change as class 1 missense
AND low prediction to alter splicing**

**Synonymous substitution
AND low prediction to alter splicing**

Class 2 Likely not pathogenic:

Posterior Probability 0.001-0.049

**Same amino acid change as class 1 missense
AND low prediction to alter splicing**

**Synonymous substitution
AND low prediction to alter splicing**

*More to come soon:
For deep intronic & promoter/UTR*

Class 1 Not pathogenic:

Posterior Probability < 0.001

Allele frequency $\geq 1\%$ in reference controls

**Same amino acid change as class 1 missense
AND no splicing effect proven**

**Exonic substitution/ins-del/synonymous with prior probability ≤ 0.02
OR intronic
AND low prediction to alter splicing OR no splicing effect proven
AND co-occurrence *in trans* with pathogenic and no Fanconi-like features
OR Allele frequency ≥ 0.001 and < 0.01 in reference controls**

What value do we get from clinical significance?

- Clinical significance is not deleteriousness
- Deleteriousness is not penetrance
- ***For methods development***, three-class (Benign, VUS, Pathogenic) clinical significance is valuable, but further granularity might not be useful.

Acknowledgements

ENIGMA Consortium

Amanda Spurdle

Michael Parsons

CAGI 2018 Organizers and predictors

BRCA Exchange at UCSC and ETH Zurich

Rachel Karchin and CRAVAT team at JHU