

CAGI slide pack guidelines

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

Yana Bromberg

- 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.



CAGI 2018: PTEN and TPMT challenge evaluations

Yana Bromberg

Dept. of Biochemistry and Microbiology

Dept. of Genetics, Rutgers University

Institute for Advanced Study, TU Munich

July 5th, 2018



Experimental reference

- What are the scores?
 - A library of thousands of PTEN and TPMT mutations was assessed to measure the stability of the variant protein using a multiplexed variant **stability** profiling (VSP) assay, which detects the presence of EGFP fused to the mutated PTEN and TPMT protein respectively... The challenge is to predict the effect of each variant on TPMT and/or PTEN protein stability.

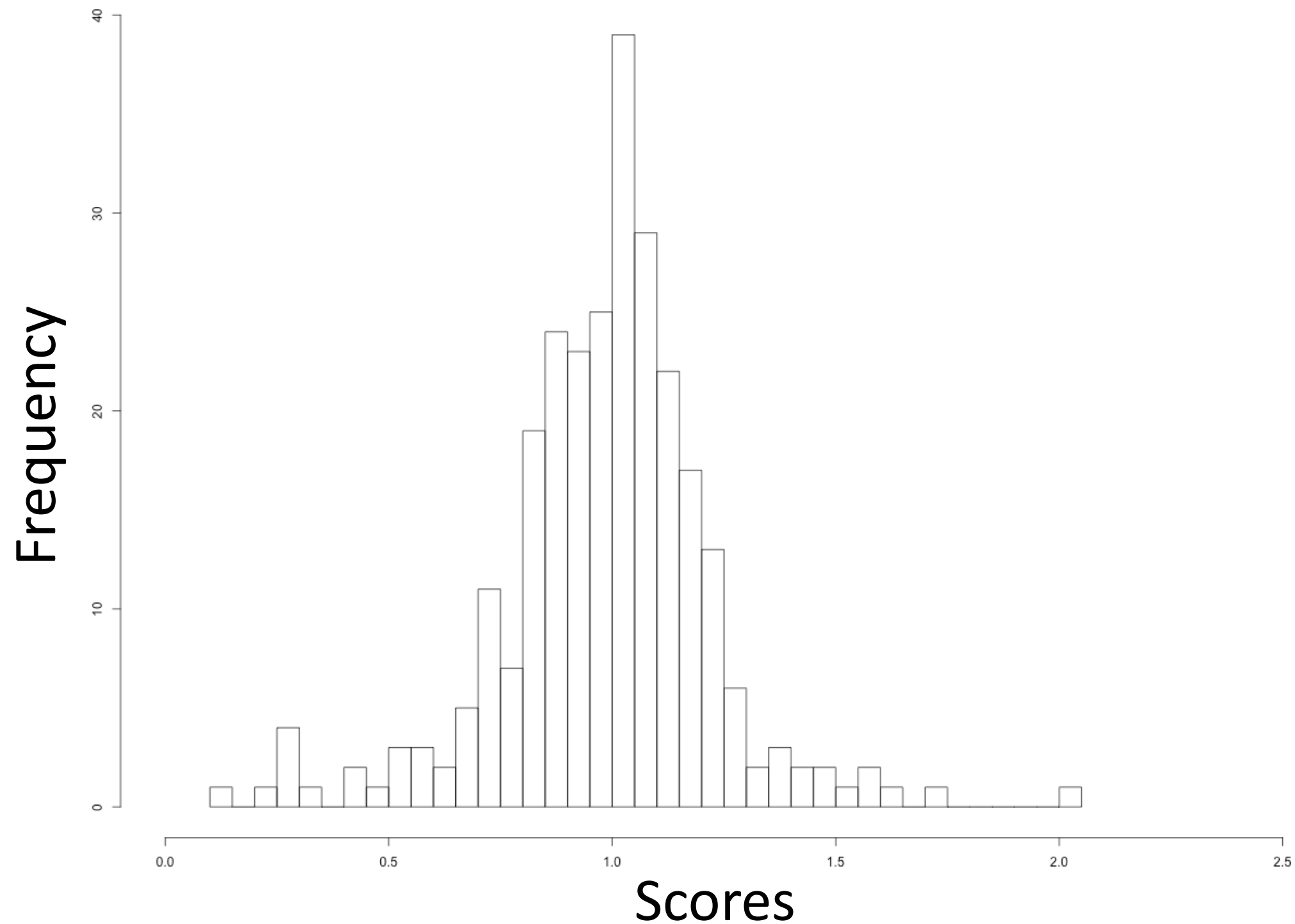
Experimental reference

- What are the scores?
 - Effect of each variant on TPMT or PTEN protein stability, with 0 meaning unstable, 1 meaning wild-type stability and >1 meaning more stable than wild type.
 - Scores and confidence intervals were computed from replicate data for 3,736 PTEN missense variants, and 2,924 TPMT missense variants (6,660 total)

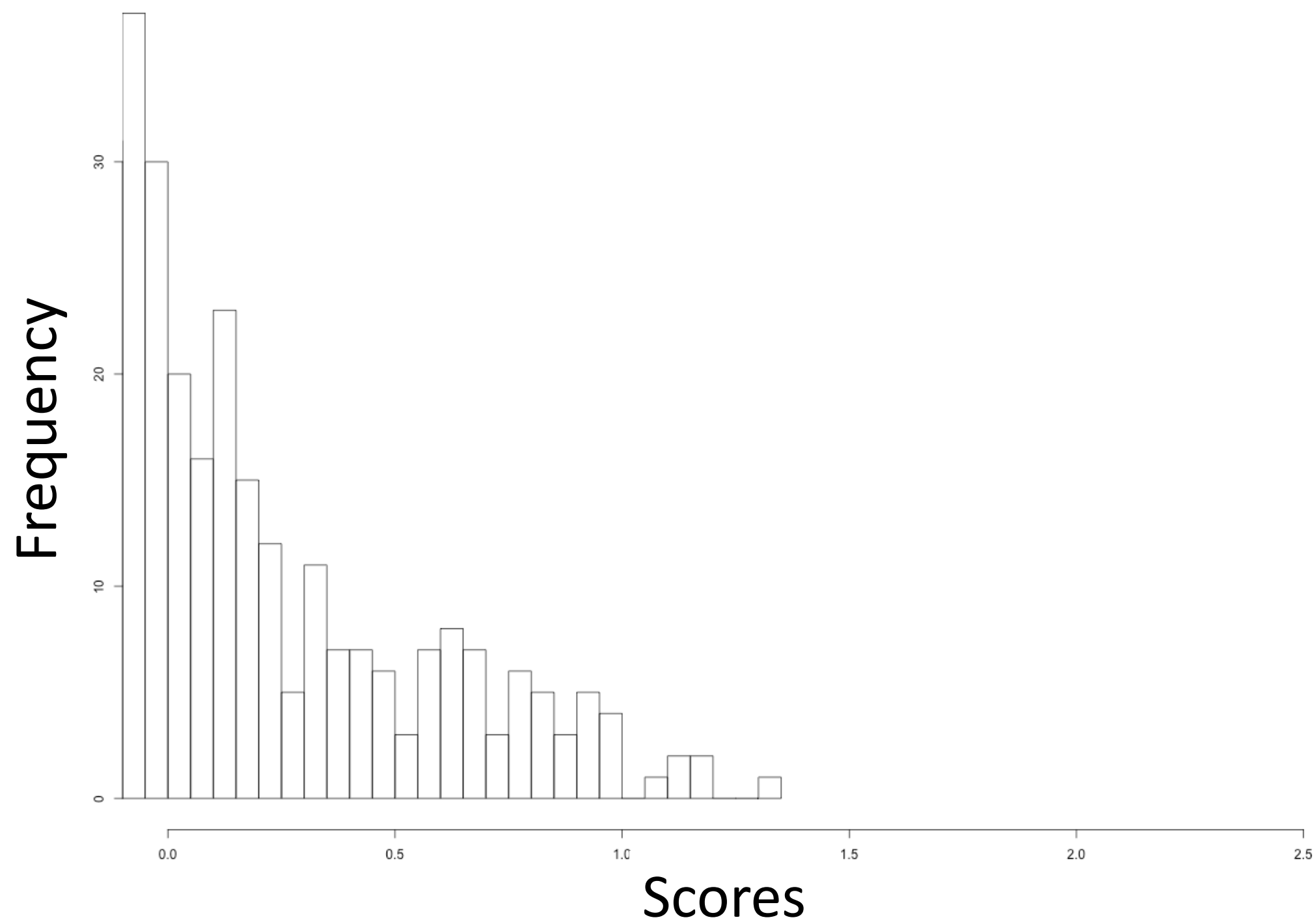
Experimental reference QUESTIONS

- What are the scores?
 - Effect of each variant on protein stability
 - Range -0.975 to 2.198. Why are there negative scores?
 - Effects computed from replicate data for 3,736 PTEN and 2,924 TPMT missense variants (6,660 total)
 - 7,953 entries provided (4,001 PTEN and 3,952 TPMT). Why are there more entries than expected?
 - 317 “toX” (nonsense?) variants.
 - 273 synonymous (?) variants. What DNA sequence is this?
 - 73 variants with no replicates

Distribution of synonymous variant scores

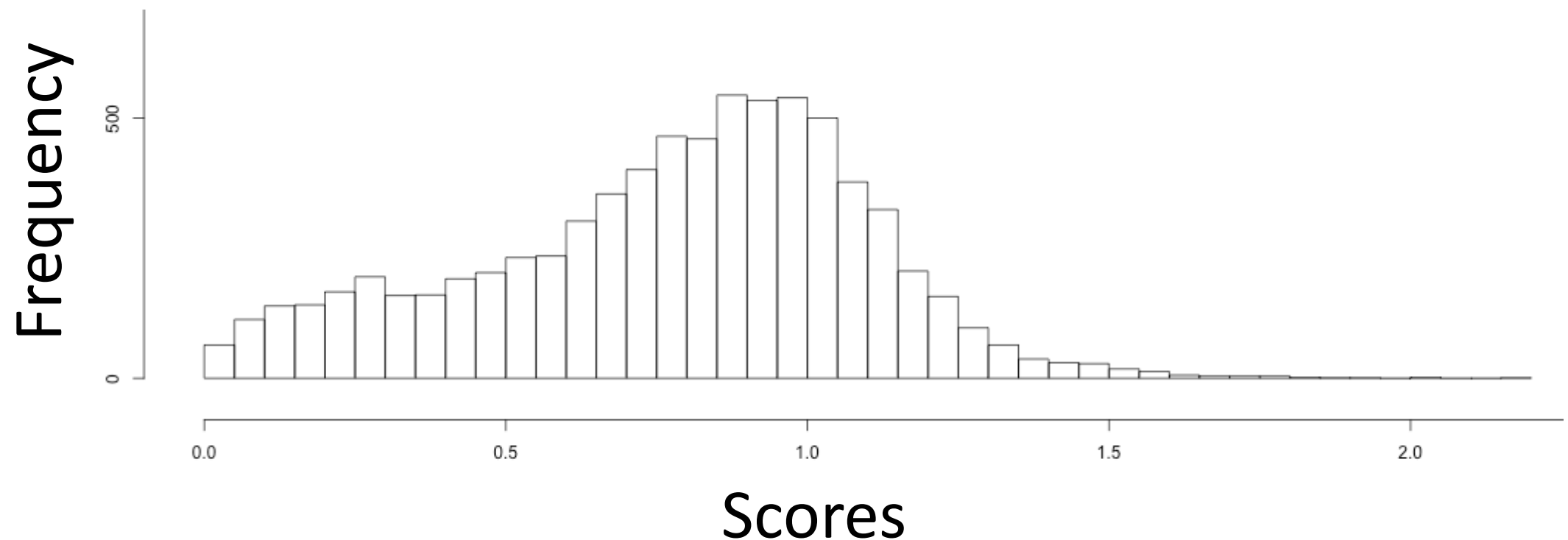


Distribution of nonsense variant scores



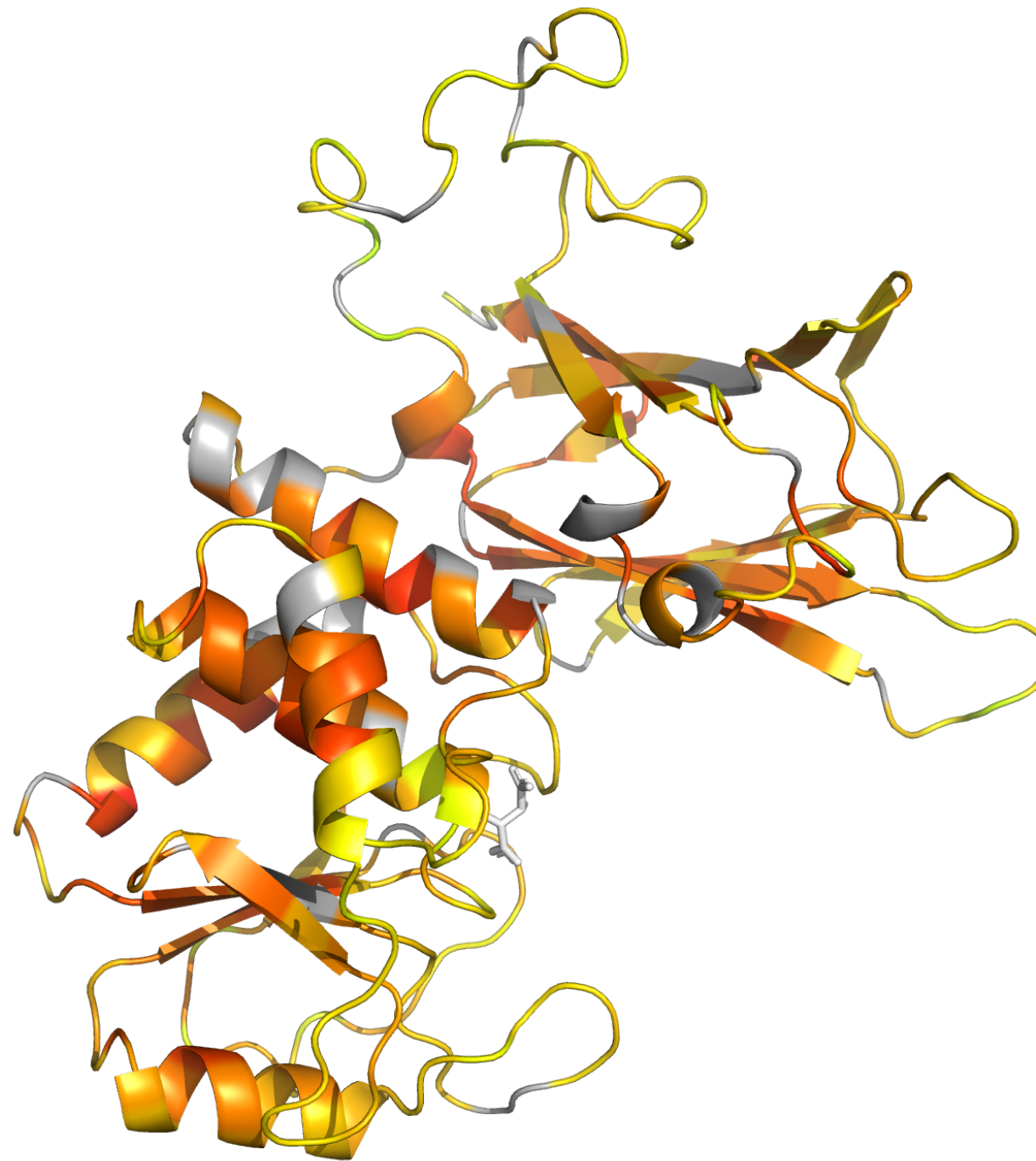
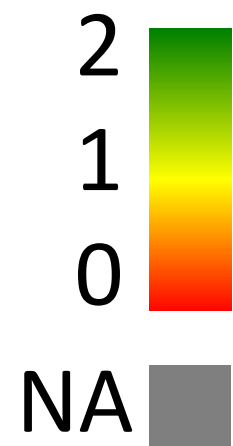
Reduced Reference Set

- 7,953 entries in the reference
 - 7,636 no “toX” (nonsense?)
 - 7,473 scores ≥ 0
- 3,860 PTEN and 3,613 TPMT variants considered



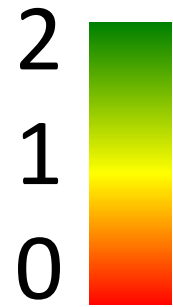
PTEN (1D5R)

Mean Score

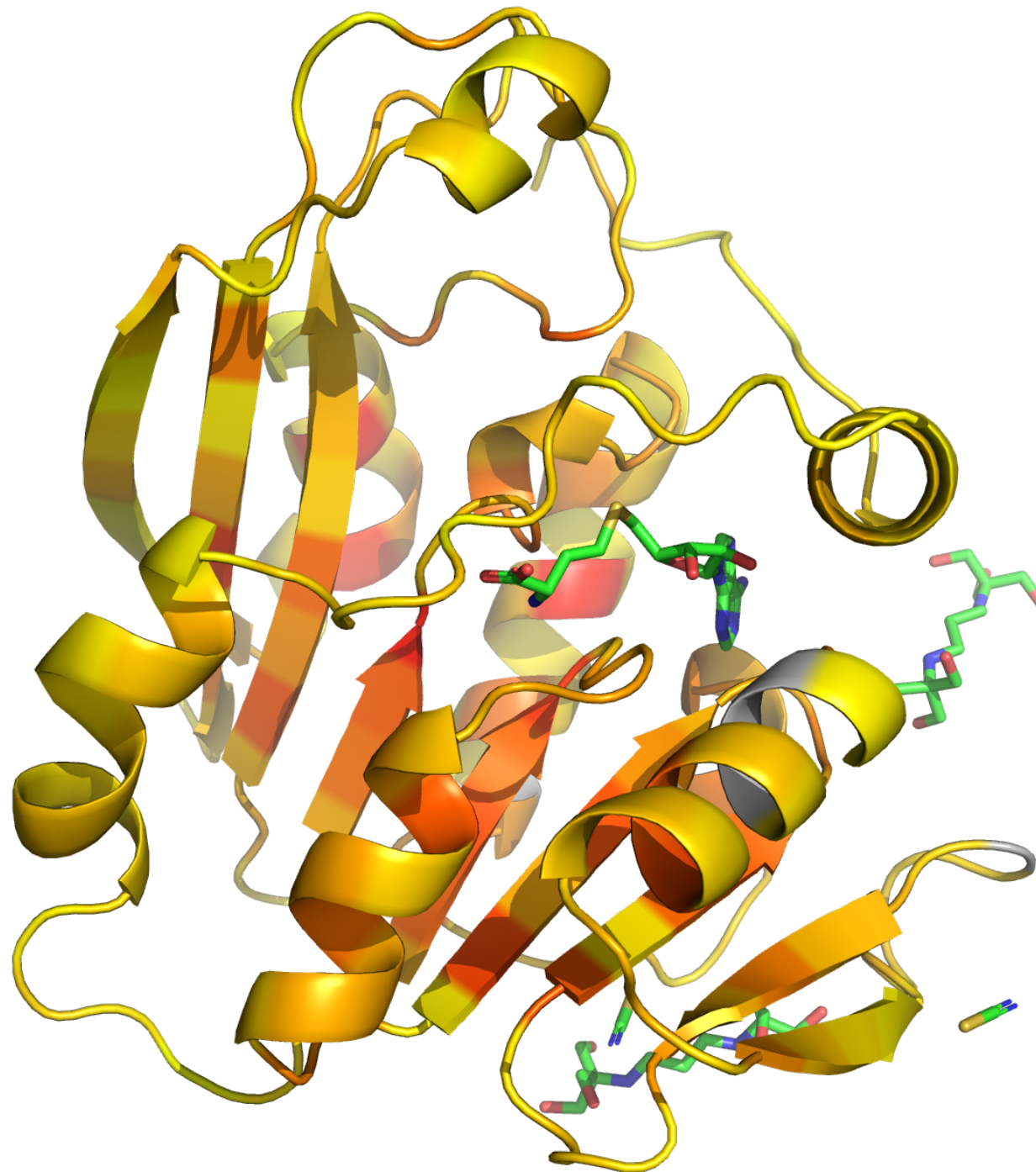


TPMT (2H11A)

Mean Score



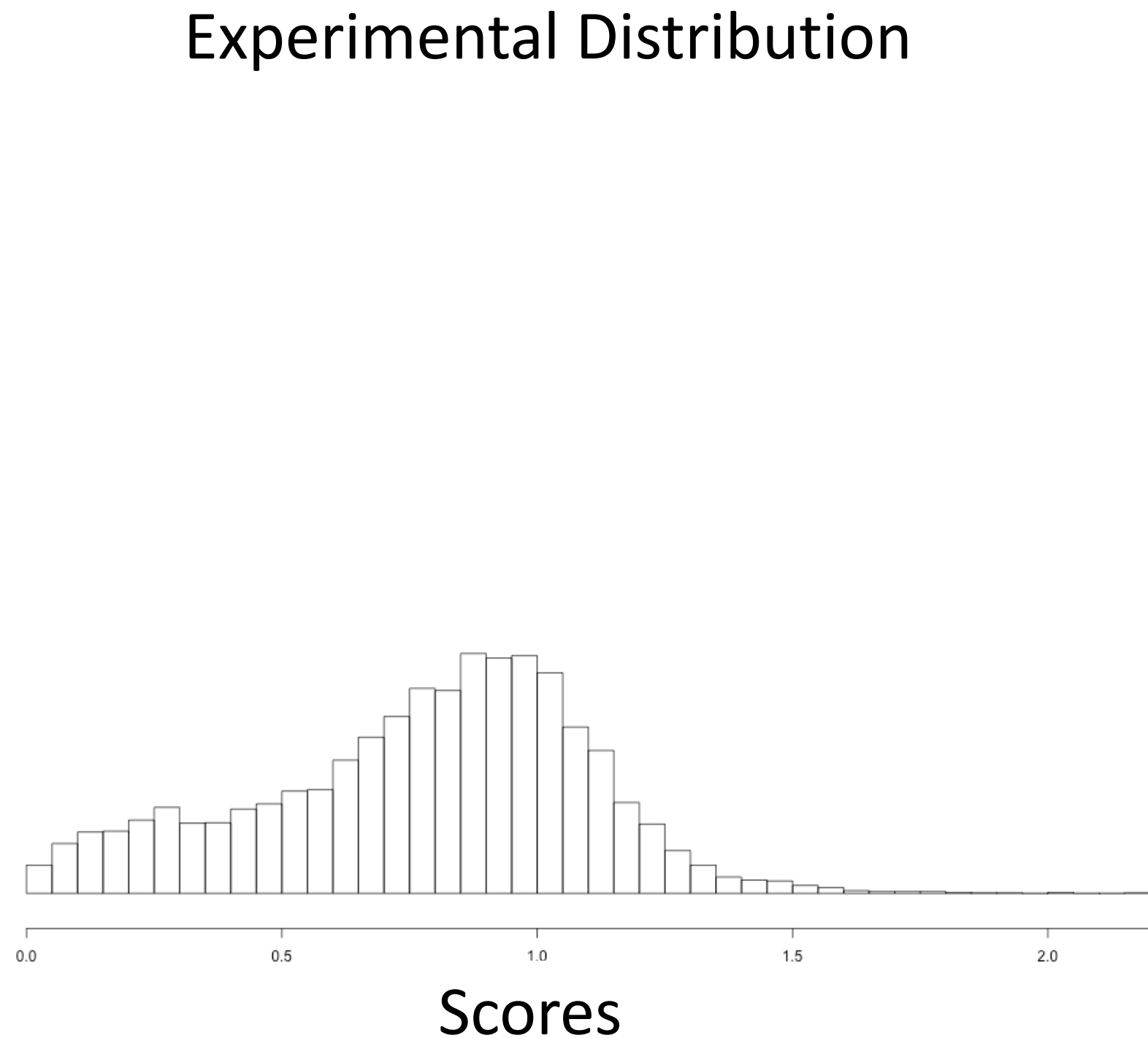
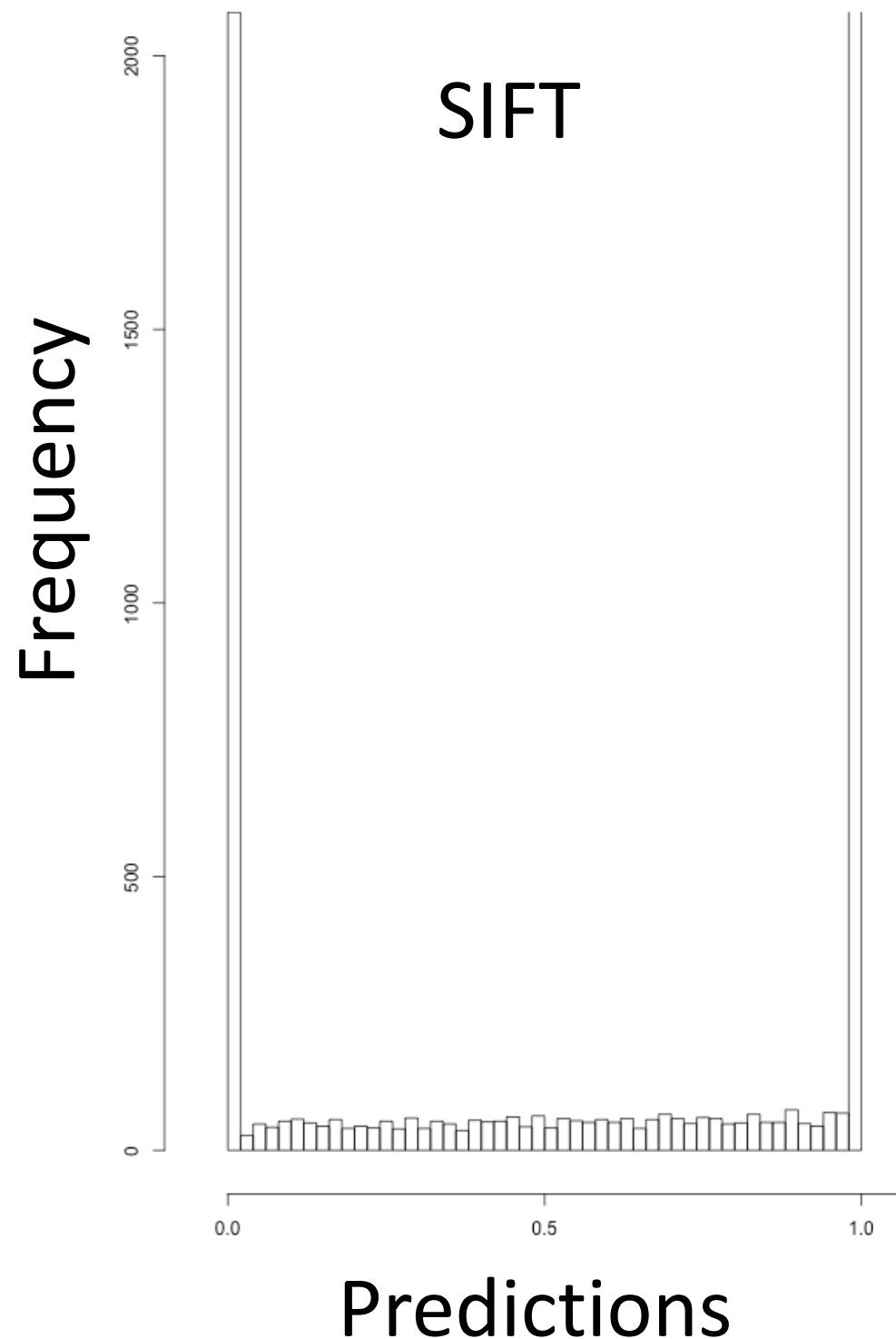
NA



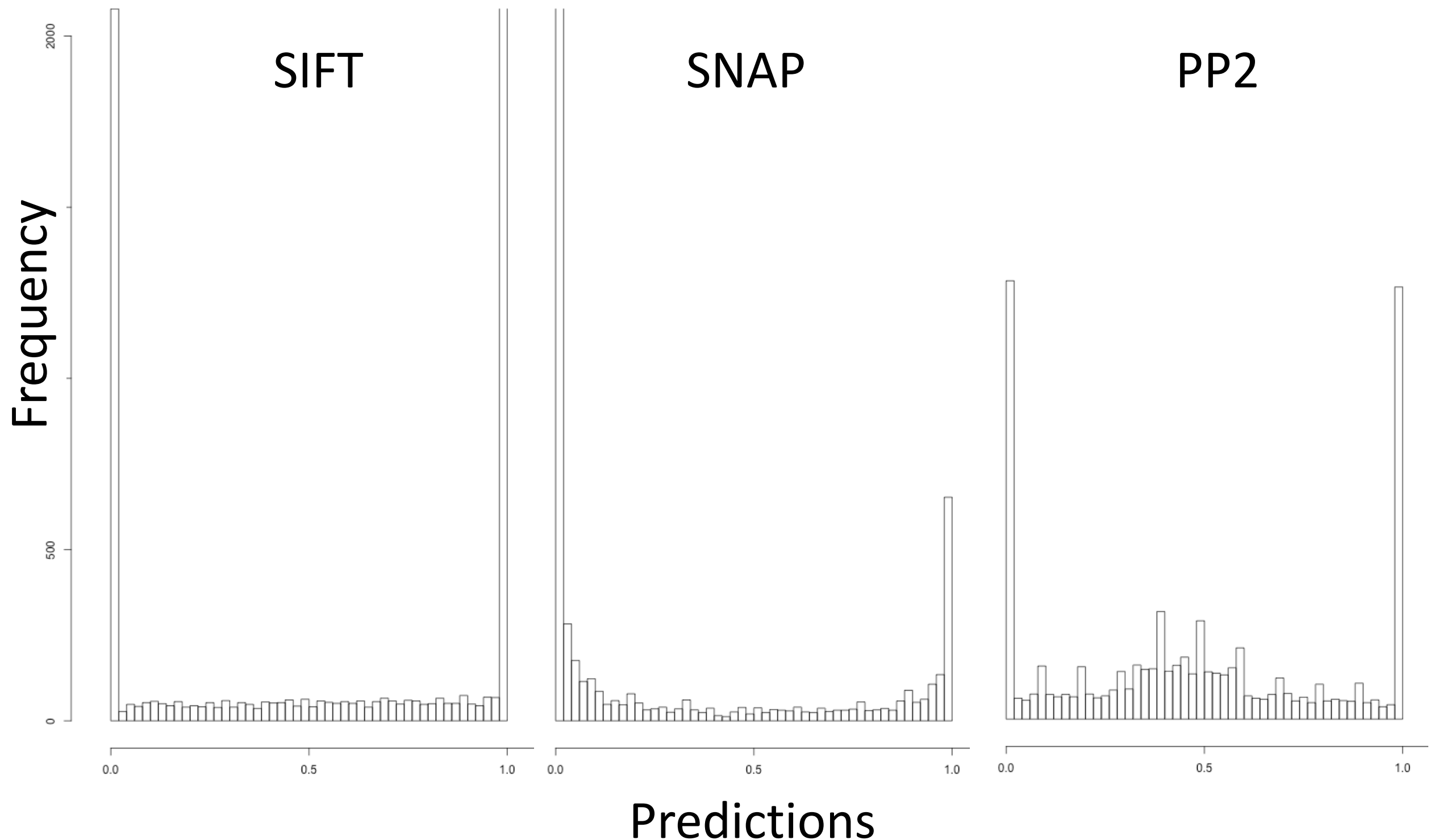
Predictions

- Eight groups submitting
 - Multiple submissions each
- Plus baseline methods **that are NOT meant to predict effects on stability:**
 - SNAP, PolyPhen (PP2), SIFT

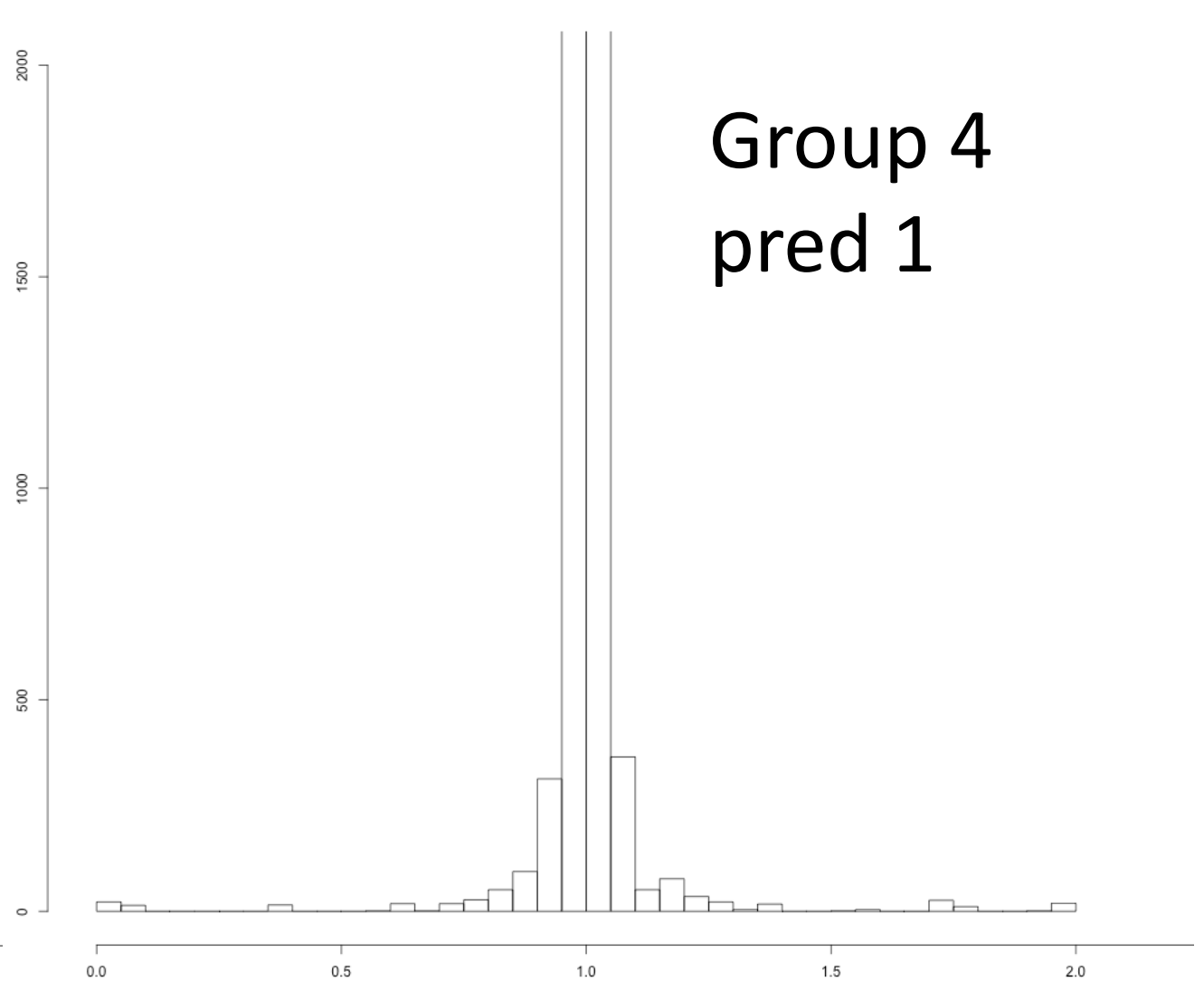
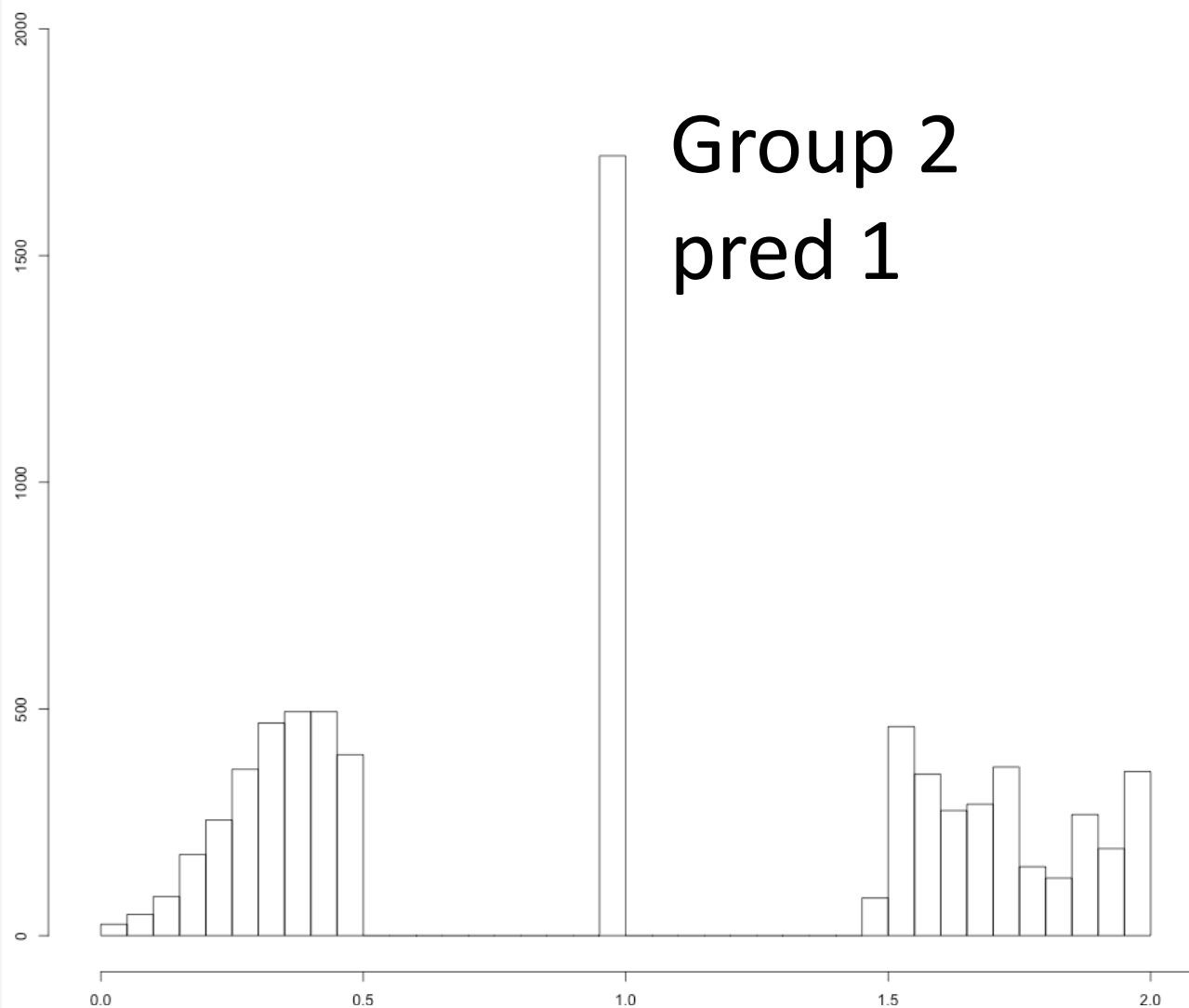
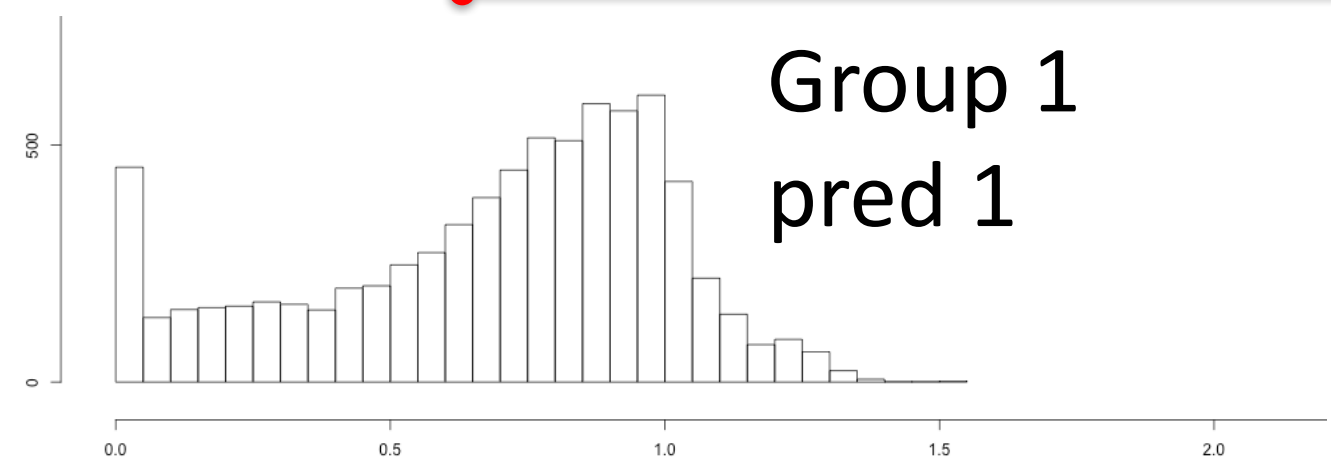
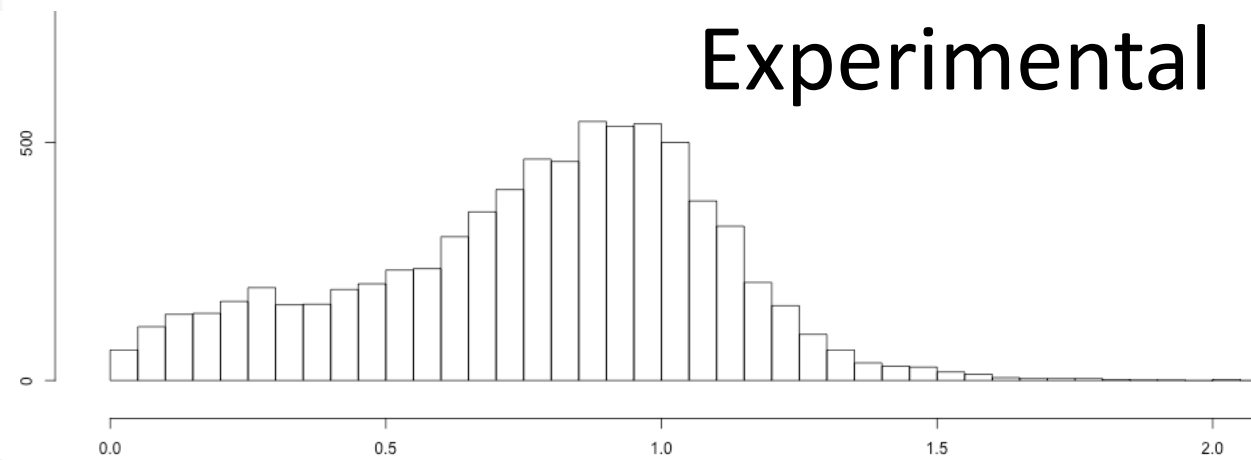
Prediction distributions of baseline predictors do not reflect experimental design



Prediction distributions of baseline predictors do not reflect experimental design



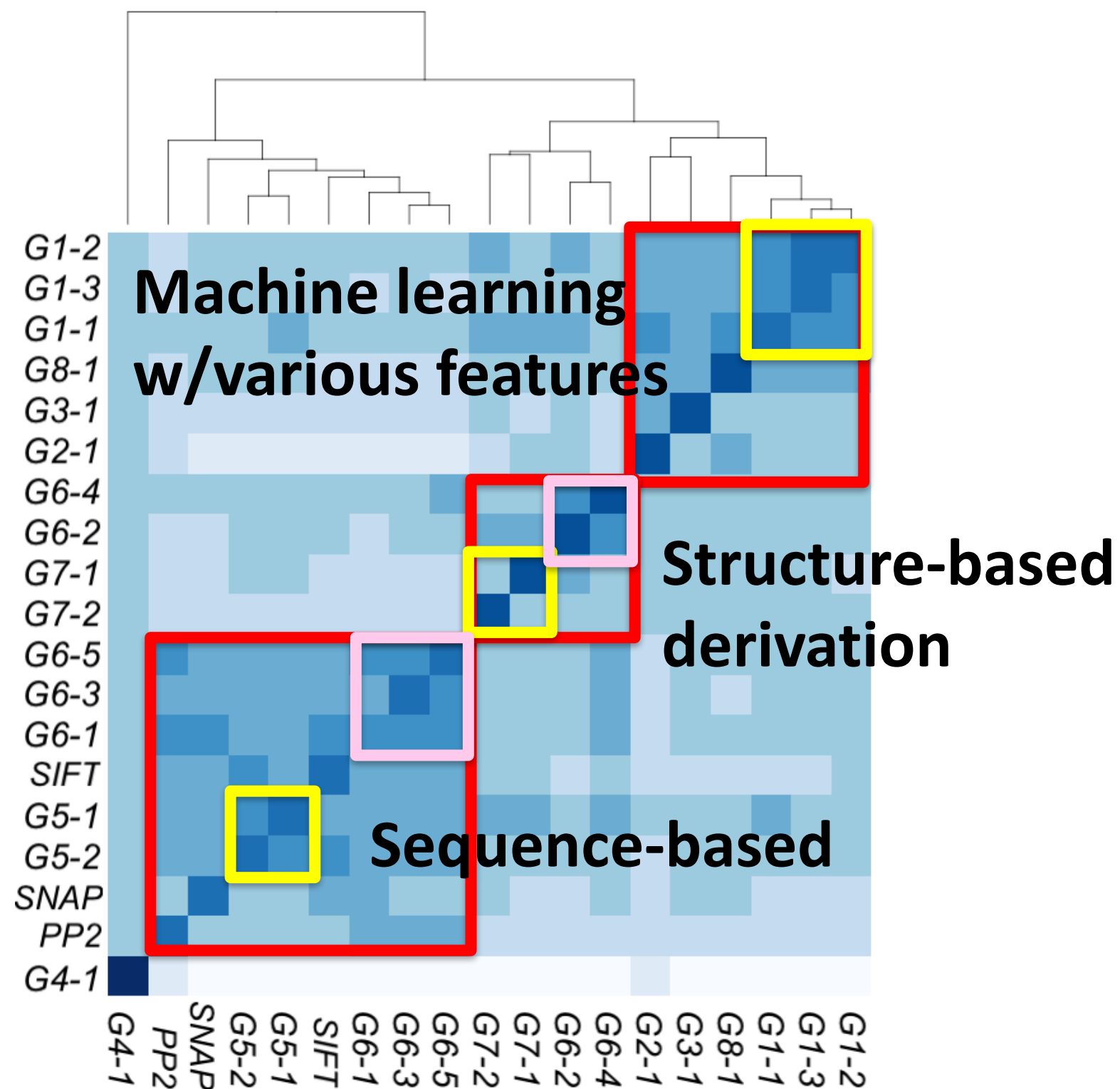
Prediction distributions also differ drastically across methods



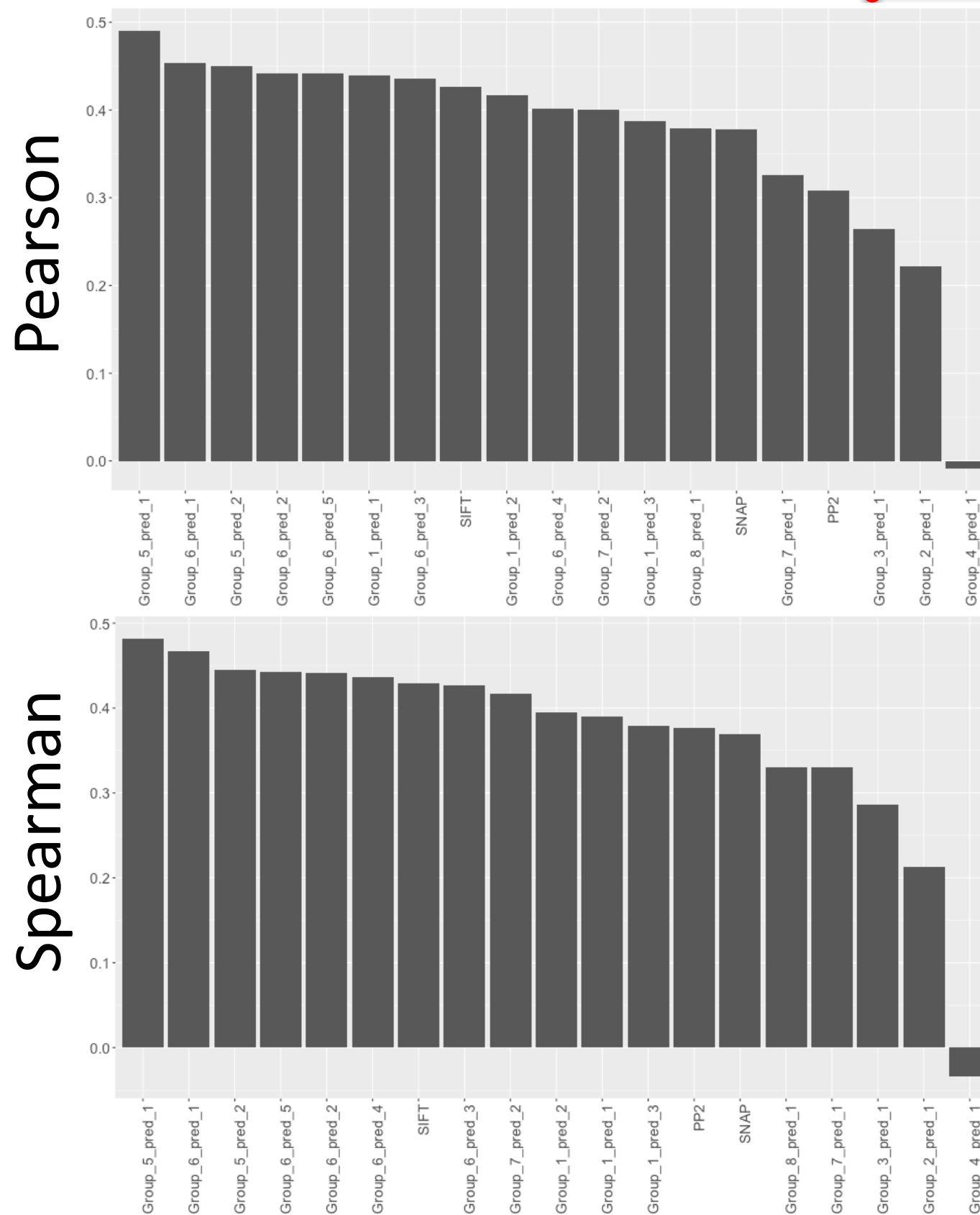
Evaluations

- Simple correlation
 - Pearson/Spearman
- Class based performance
 - Assign variants an experimental effect class
 - Tri-class:
 - Lower CI limit < 0.468 , most de-stabilizing (lowest 25%)
 - Upper CI limit > 1.2 , most stabilizing (high 25%)
 - WT: score in range $[0.9, 1.1]$
 - Bi-class:
 - Group stabilizing and de-stabilizing

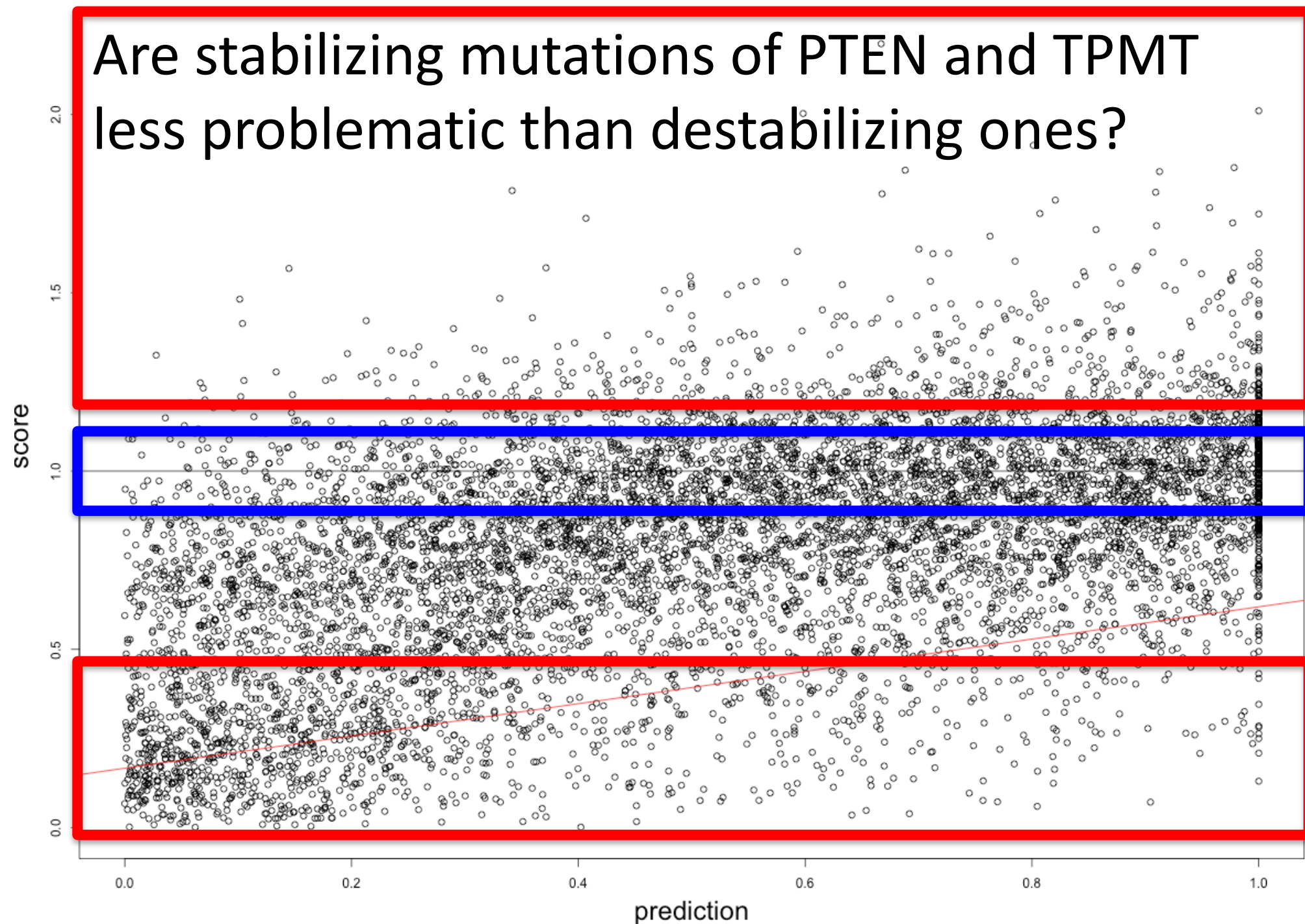
There is some correlation across predictors



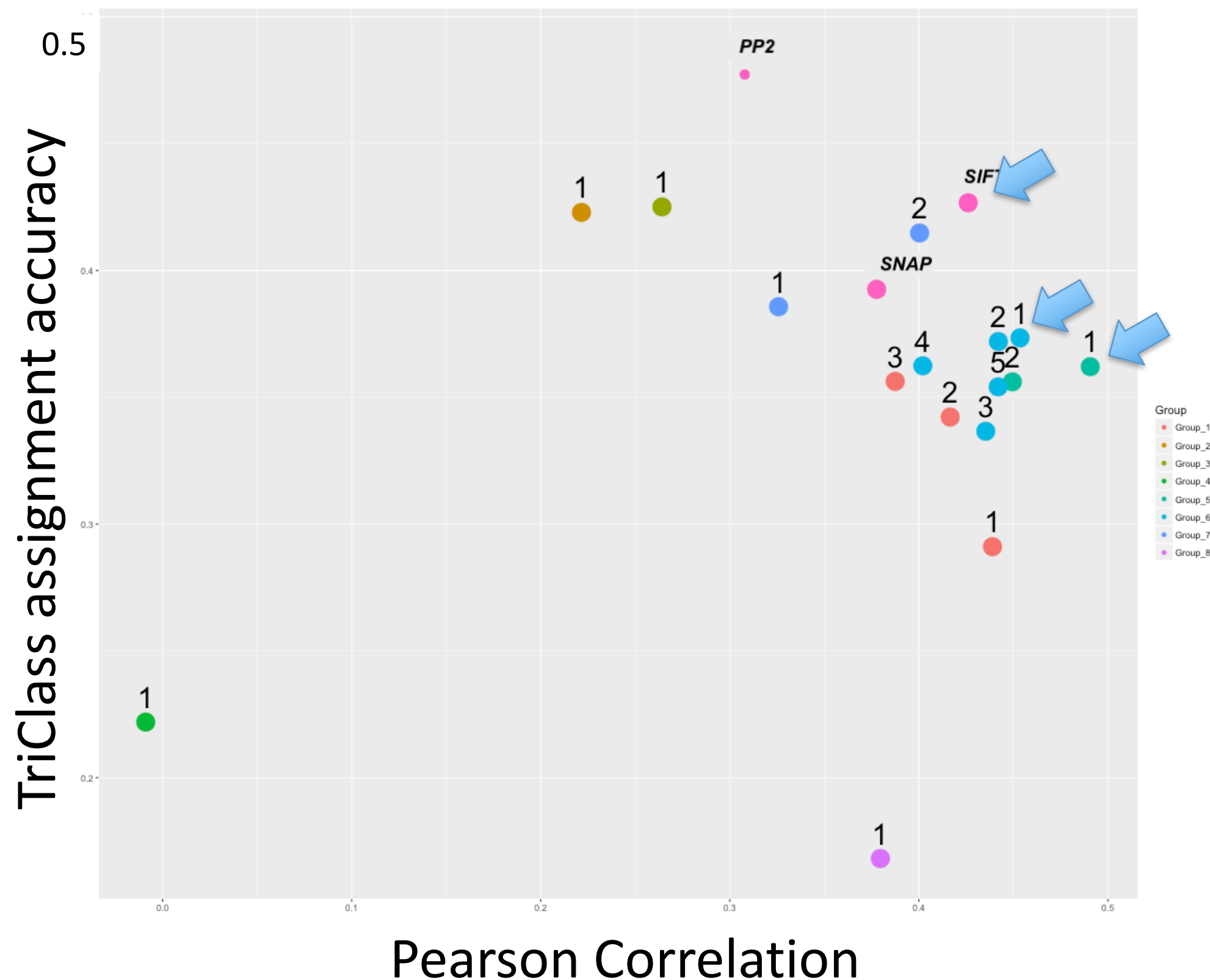
Correlation of predictions with experimental measurements is never over 0.5



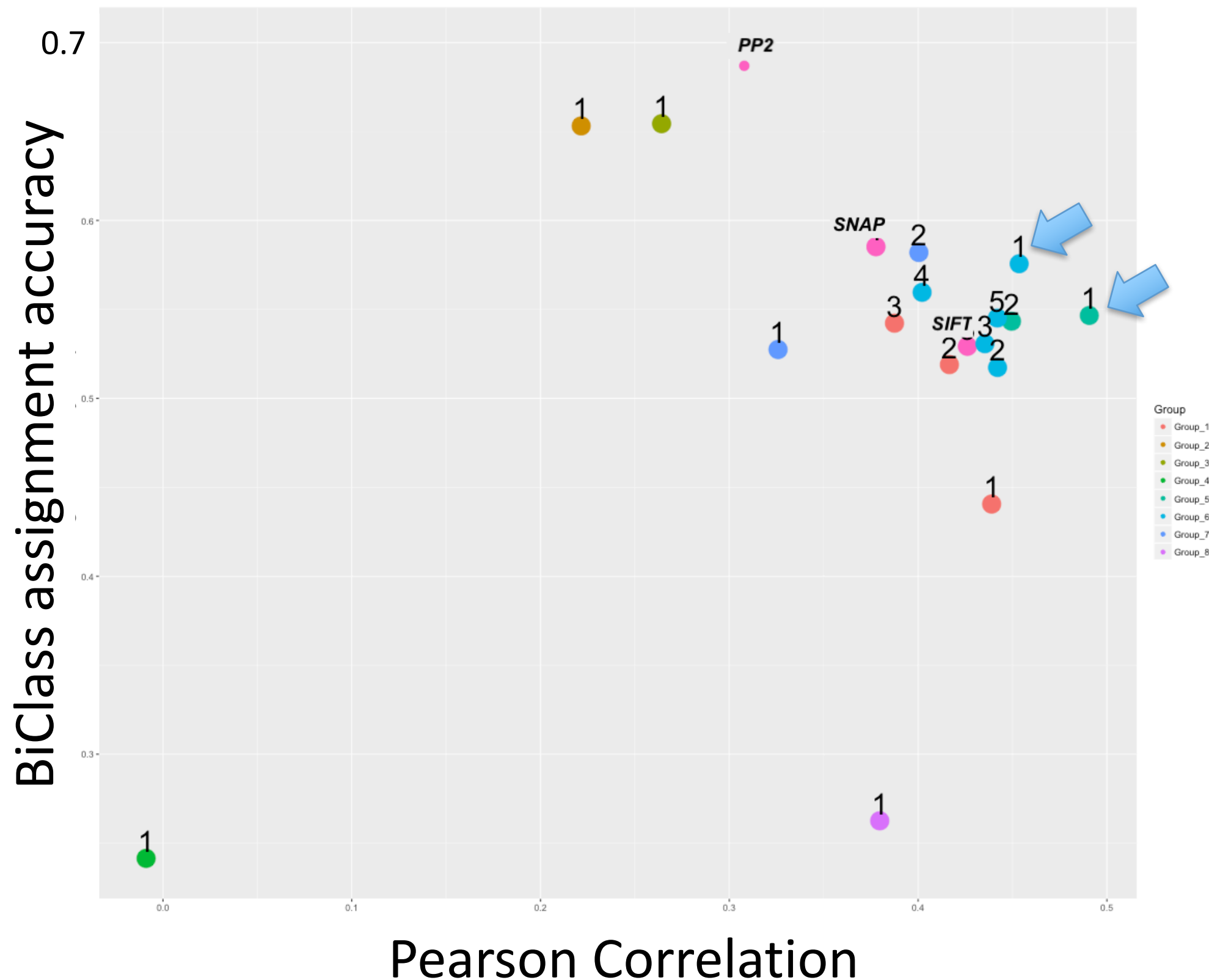
Group 1 Prediction 1 (highest correlation)



Correlation measures alone don't do justice to extreme class predictions

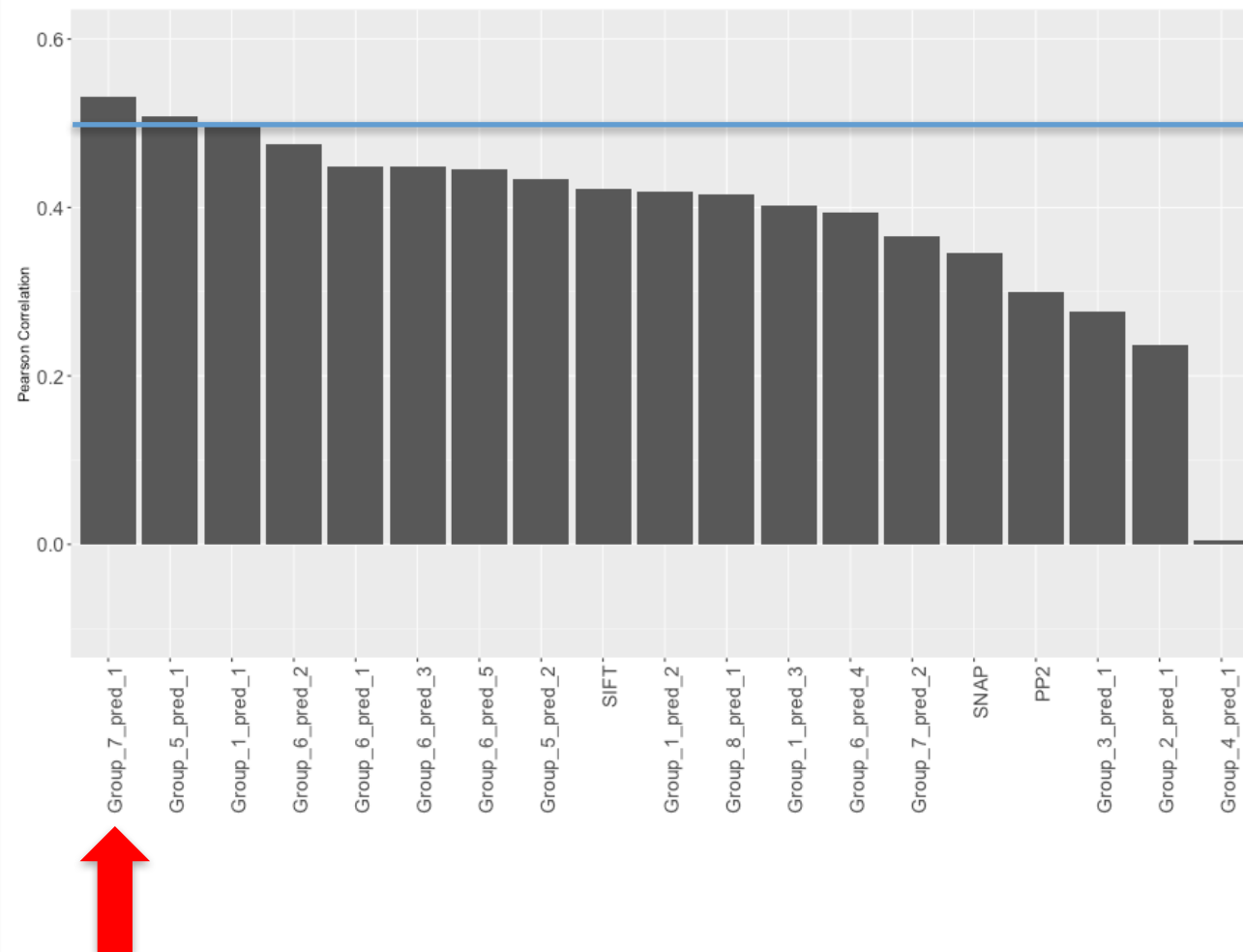


For most class-based predictors, effect/no-effect classifications are easier

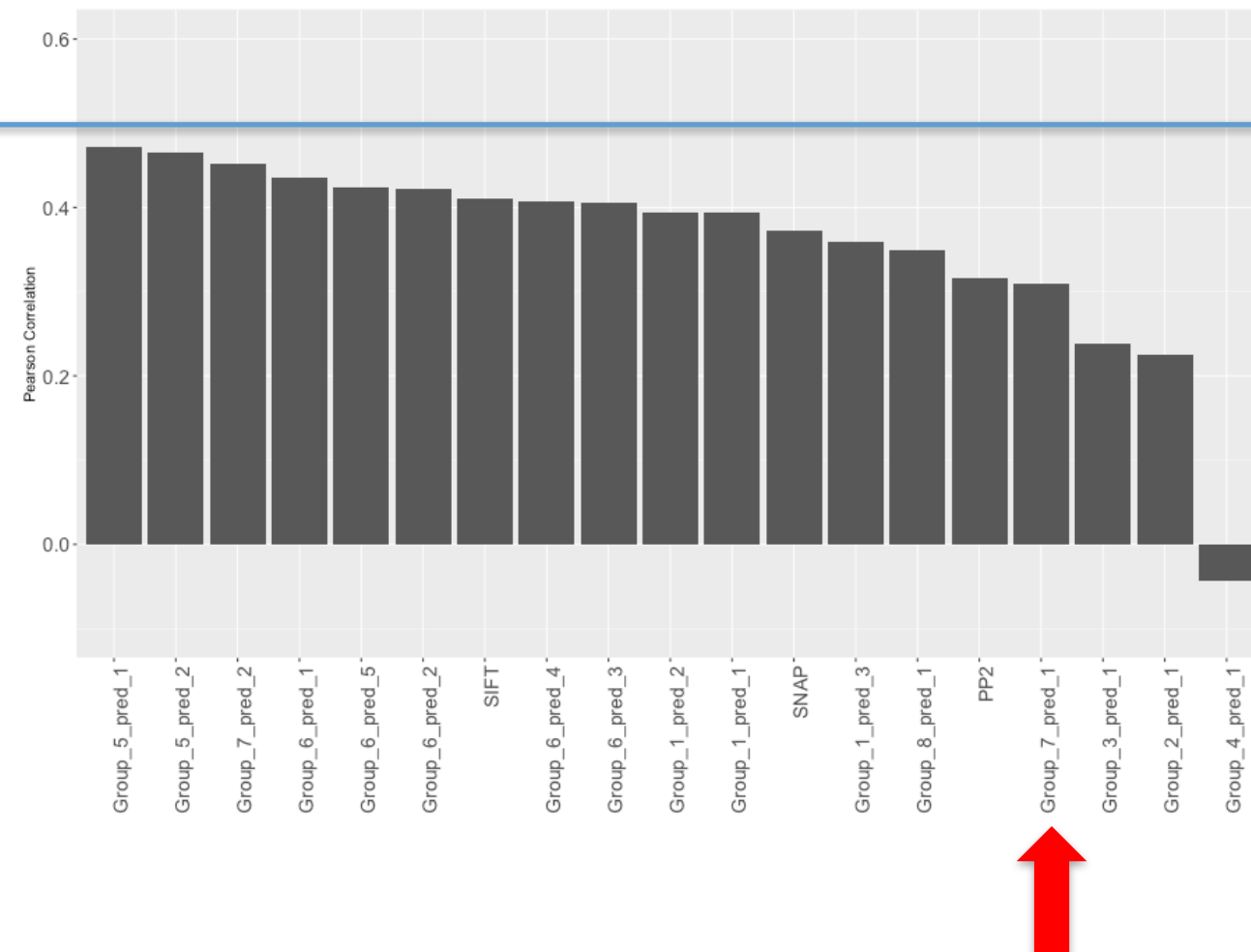


Predictors perform differently for different proteins

Pearson PTEN

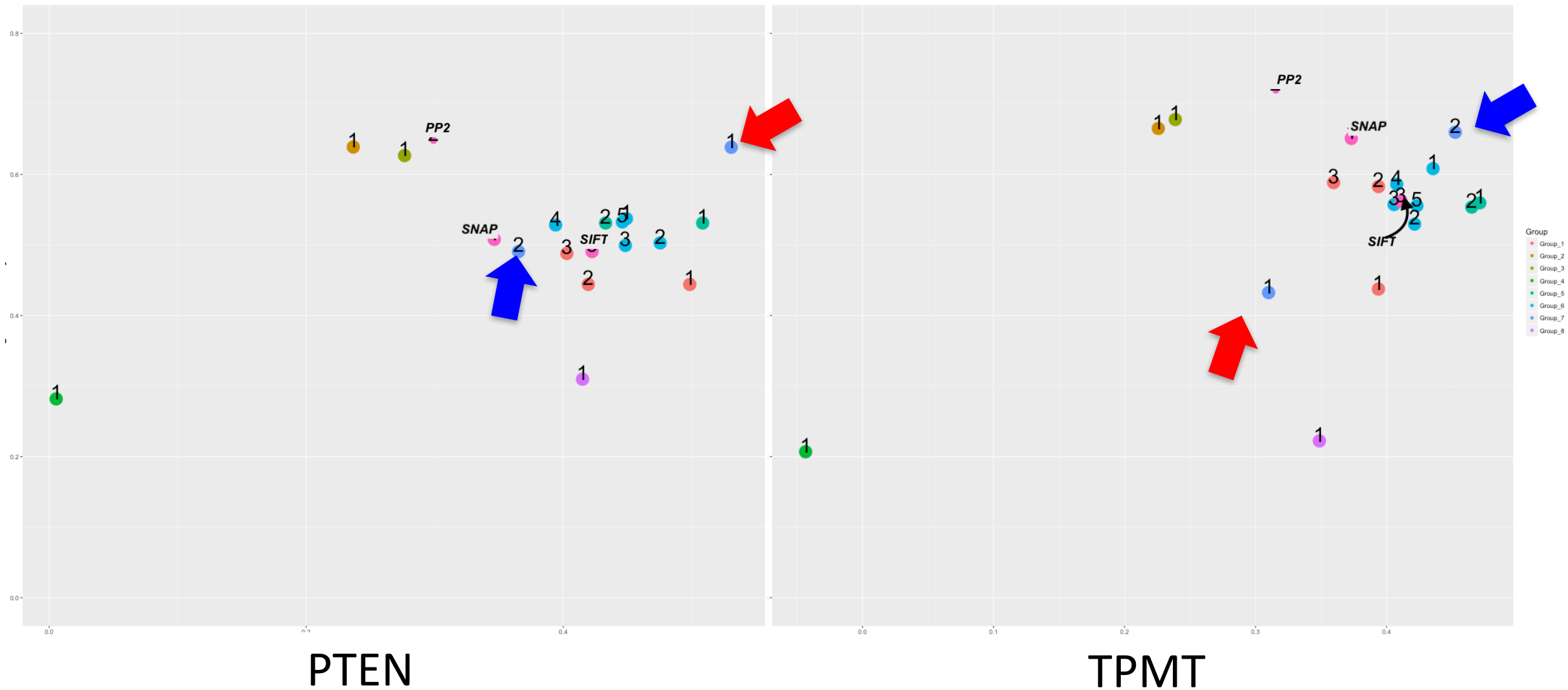


Pearson TPMT



Predictors perform differently for different proteins

BiClass assignment accuracy



Pearson Correlation

Drum Roll Please

- Best correlation across both proteins
 - Lichtarge lab (Group 5)
- **Best tri/bi class prediction across both proteins**
 - Soochow Lab (Groups 2 and 3)
- Best performance (biclass/correlation ratio) per protein
 - Dunbrack lab (Group 7)
- Best performance (biclass/correlation ratio) for both proteins
 - Tie Moulton lab (Group 6) and Lichtarge lab (Group 5)

Limitations of this evaluation

- Performance on just two proteins differs significantly enough that no clear leader/approach can be named for generic use
 - With the exception of outliers, most predictors perform about the same
 - How can we take SD of predictors into account?
- How can predictors be evaluated for predicting synonymous mutation effects, if the (DNA) mutation itself is not provided?
- Standard predictors (SNAP, SIFT, PP2) do OK but they shouldn't be used for these types of predictions!

Thank You!

