The American Journal of Human Genetics, Volume *106*

Supplemental Data

An Integrated Deep-Mutational-Scanning Approach Provides Clinical Insights on *PTEN* Genotype-Phenotype Relationships

Taylor L. Mighell, Stetson Thacker, Eric Fombonne, Charis Eng, and Brian J. O'Roak

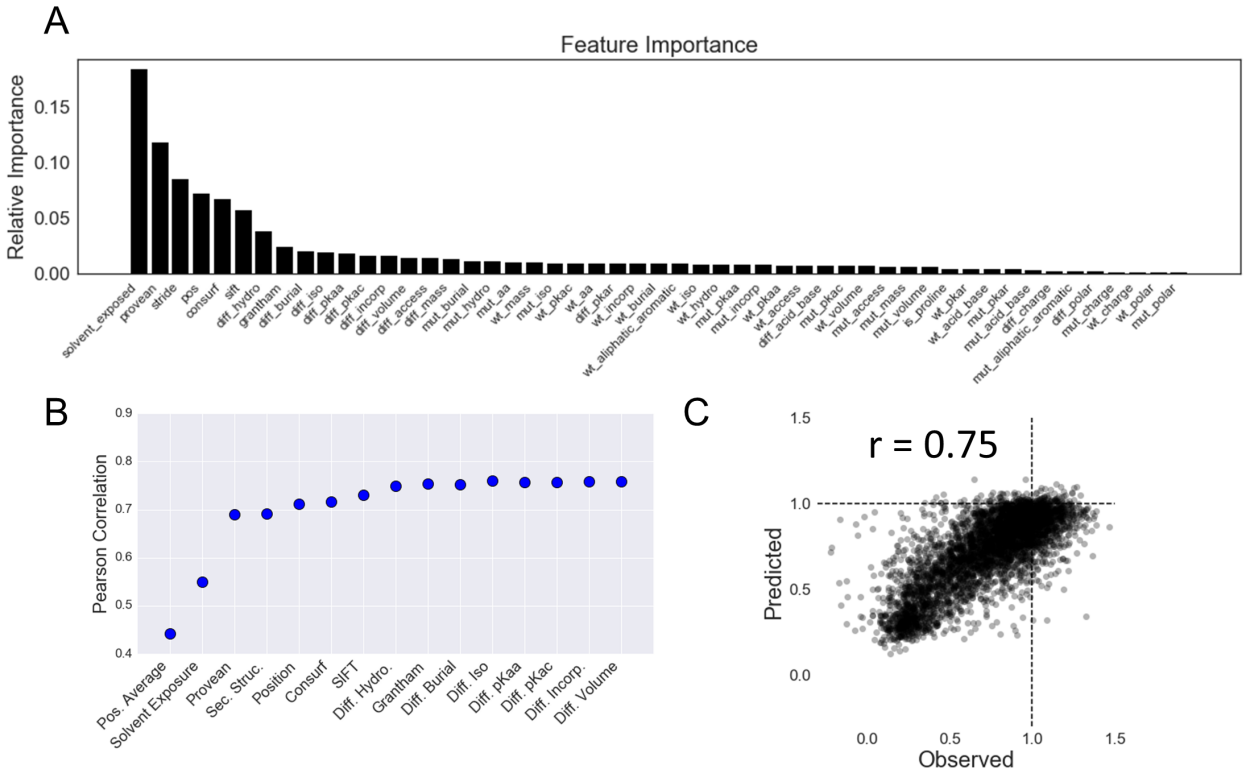

Figure S1. Imputation of missing abundance scores

We used 51 evolutionary, biophysical, biochemical, or *in silico* variant effect predictors to train a random forest machine learning model to predict the effect of unseen variants (See Mighell et al., 2018).

(A) Feature importance of all features (besides position average) from the full dataset, calculated as the relative increase in error upon random permutation of each feature.

(B) Pearson correlations between predicted and observed variant scores. The first model includes only position average, while each successive row includes that feature as well as all features to the left. Features in order include; “Pos. Average”: average abundance score of other single amino acid mutations at that position or neighboring positions (see Materials & Methods); “Solvent Exposure”: calculated with GETAREA web tool; “Provean”: mutation effect predictions; “Sec. Struc.”: secondary structure, enumerated with STRIDE; “Position” : position in primary sequence; “Consurf”: evolutionary conservation; “SIFT”: mutation effect predictor; “Diff. Hydro”: difference in hydropathy between wildtype and variant amino acid; “Grantham”: Grantham substitution score; “Diff. Burial”: difference in burial between wildtype and variant amino acid; “Diff. Iso”: difference in isoelectric point between wildtype and variant amino acid; “Diff pKaa”: difference in amino pKa between wildtype and variant amino acid; "Diff. PKac": difference in carboxyl pKa between wildtype and variant amino acid; "Diff. Incorp": difference in protein incorporation rate between wildtype and variant amino acid; "Diff. Volume": difference in volume between wildtype and variant amino acid. Note: To ensure that our feature selection approach was consistent across subsets of the data, we iteratively repeated this procedure using 90% subsets of the data. We found that the set of top 12 features were consistent across all folds and that the median ranking of features across the folds was the same as what was originally used for modeling.

(C) 10-fold cross validation demonstrated high accuracy of the final model, yielding 0.75 Pearson correlation between predicted and observed variant scores.

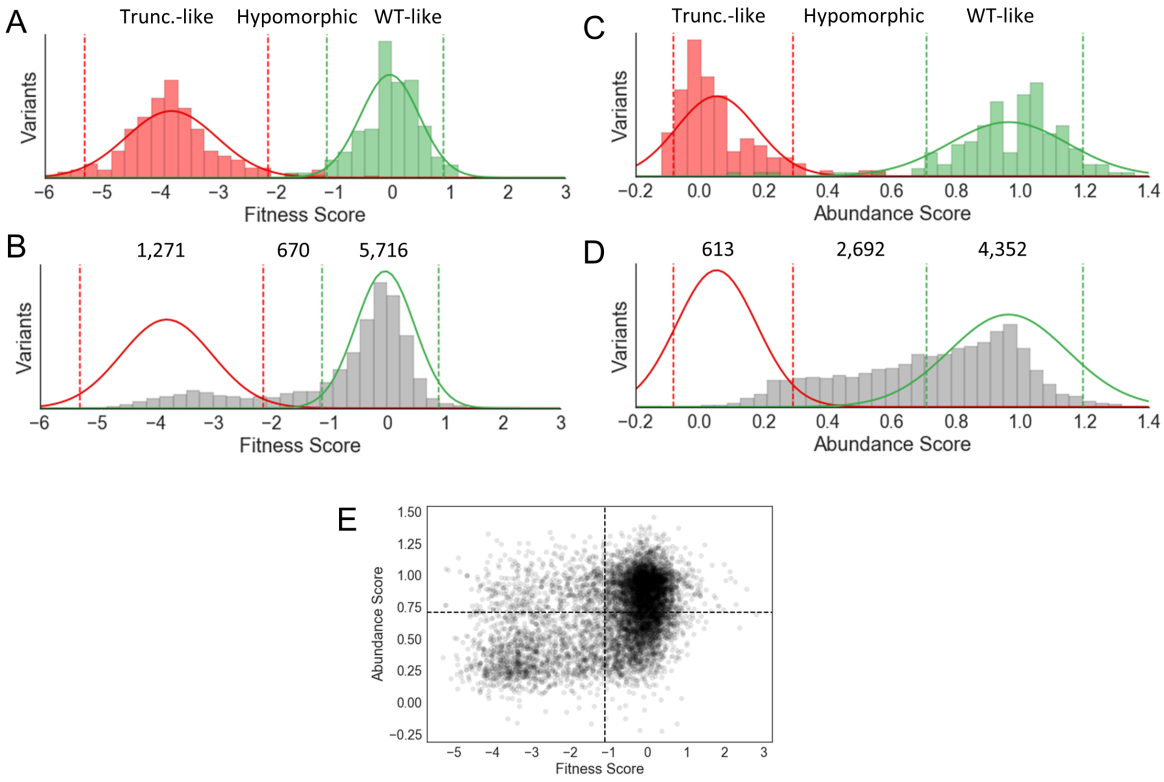

Figure S2. Scaling variant functional scores for integration

(A-B) Histograms showing distributions of fitness scores.

(A) We defined variant fitness scores in relation to the distribution of synonymous-wildtype variants (green) and early truncating nonsense variants (red). We used only truncating variants before the 350th position (i.e. excluding the unstructured tail) to ensure that this distribution represented true loss-of-function. Solid lines are Gaussian fits of the histograms. We drew cutoff lines (dashed lines) corresponding to the 2.5 and 97.5 percentile of these distributions.

(B) Fitness score distribution (gray) for all missense variants, including the high confidence measured and imputed scores, were compared to the synonymous/truncation cutoffs. We considered missense variants to be truncation-like if they fell within the 95 percentile bounds of the truncation distribution, and likewise for missense variants within the synonymous-wildtype distribution. Variants that fell between these two distributions were considered hypomorphic.

(C-D) Histograms showing distributions of abundance scores.

(C) As in (A), variant abundances scores were coded according to their relationship to truncating nonsense (red) and synonymous wild-type variants (green). Only truncating variants between the 30th and 300th position were used in order to exclude measurement artifacts at the termini. Solid lines are Gaussian fits of the histograms. Further, cutoff lines (dashed lines) were drawn at the 5th and 95th percentiles because the tails of the distributions were much longer for abundance scores relative to fitness scores.

(D) As in (B), abundance score distribution (gray) for all missense variants, including the high confidence measured and imputed scores, were compared to the synonymous/truncation cutoffs. For abundances scores, we considered missense variants to be truncation-like if they fell within the 90 percentile bounds of truncation distribution, and likewise for missense variants within the synonymous-wildtype distribution. Variants that fell between these two distributions were considered hypomorphic.

(E) Scatterplot of all missense variants plotted as a function of fitness and abundance scores. Pearson's r = 0.43.

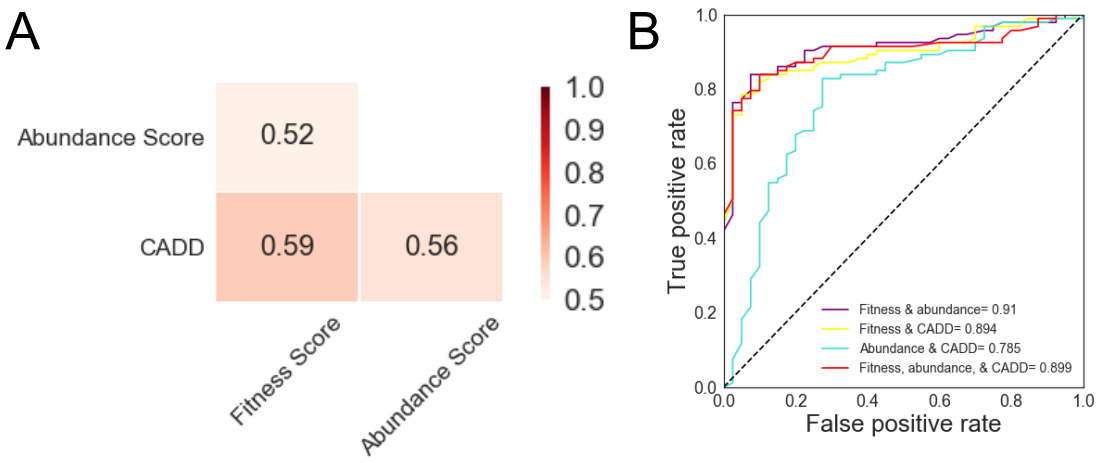

Figure S3. Logistic regression optimization for predicting pathogenic vs. benign PTEN variation

(A) Spearman correlation of features used in the modeling.

(B) Receiver operator characteristic curves with corresponding area under the curve for the various models tested. Model weights reported in Table S7.

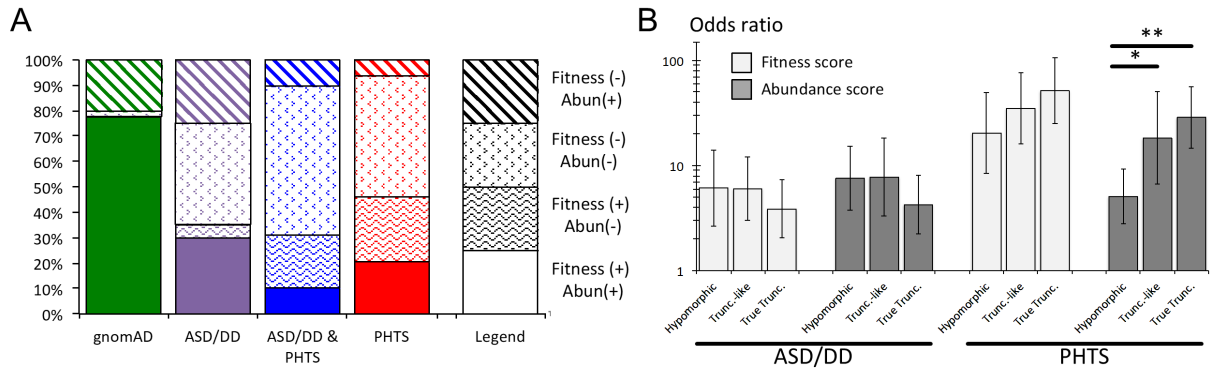

Figure S4. Risk scores for fitness and abundance exposures

(A) Variants in gnomAD or different clinical categories occupy different fitness/abundance quadrants (as shown in Figure 5C).

(B) We calculated logistic regression-based odds ratios for individuals to develop ASD/DD or PHTS features as a function of exposure to hypomorphic scoring, truncation-like scoring, or true truncation variants. Odds ratios are calculated as a comparison between the variant class of interest and the wildtype-like scoring variants. Odds ratios and 95% confidence intervals reported in Table S8. *p<0.05, **p<0.01.