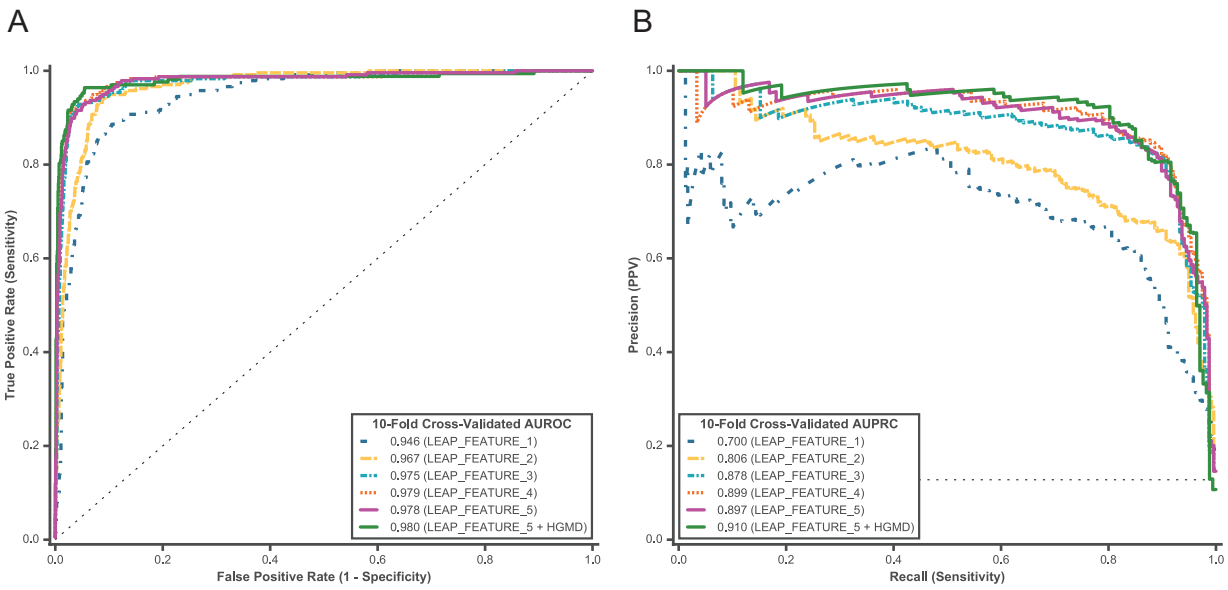
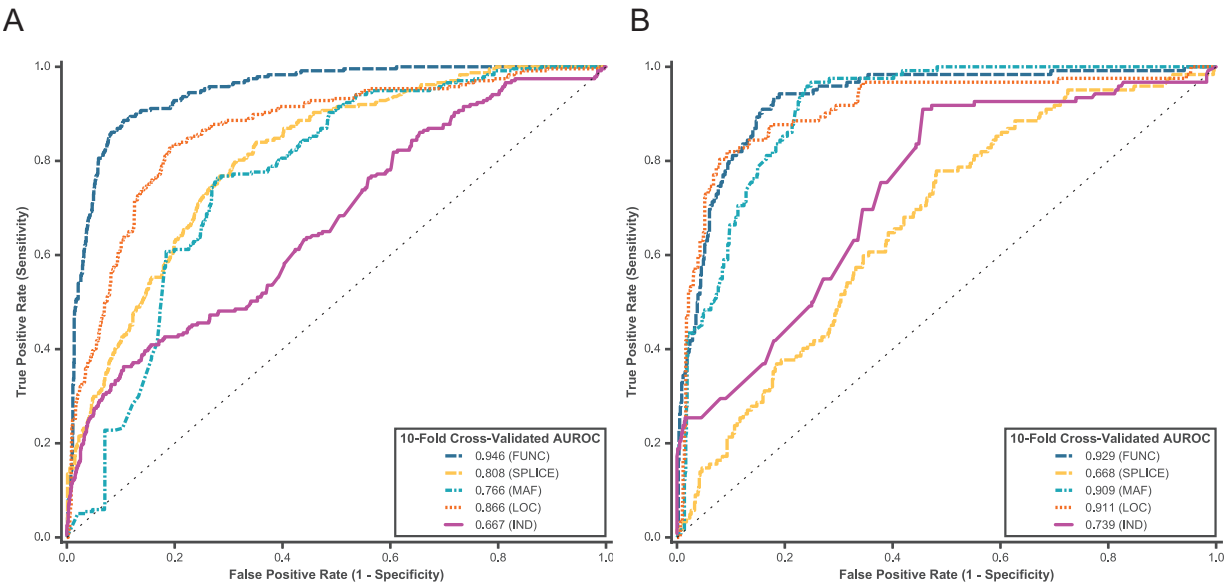
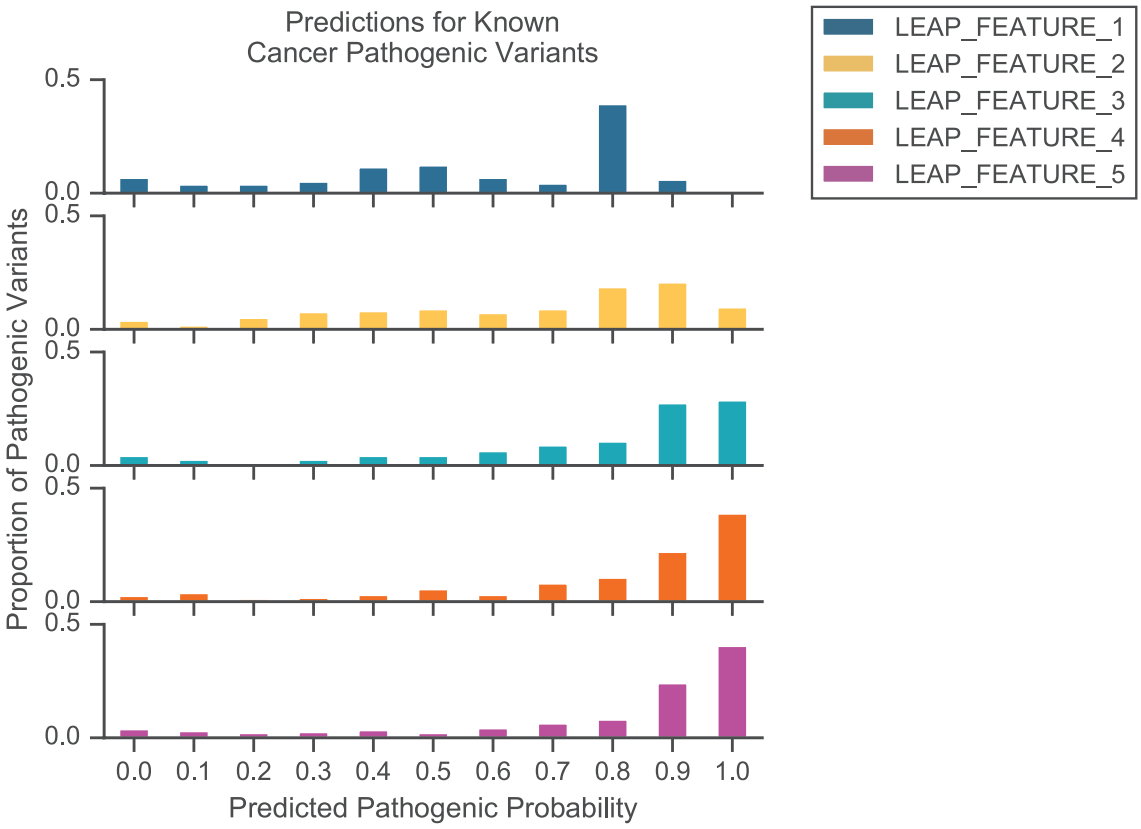
Supp. Figure S1. 10-fold cross-validated predictions were assessed from a binary L2-regularized logistic regression model for feature comparison, with HGMD features considered. Predictions for P/LP and B/LB hereditary cancer variants were assessed using (A) AUROC and (B) AUPRC. Feature comparison models are described in Table 2.



Supp. Figure S2. 10-fold cross-validated predictions were assessed from binary L2-regularized logistic regression models, each trained on features from a single feature category from Table 1 to assess feature category contribution to model performance for (A) hereditary cancer and (B) cardiovascular disorder variants.

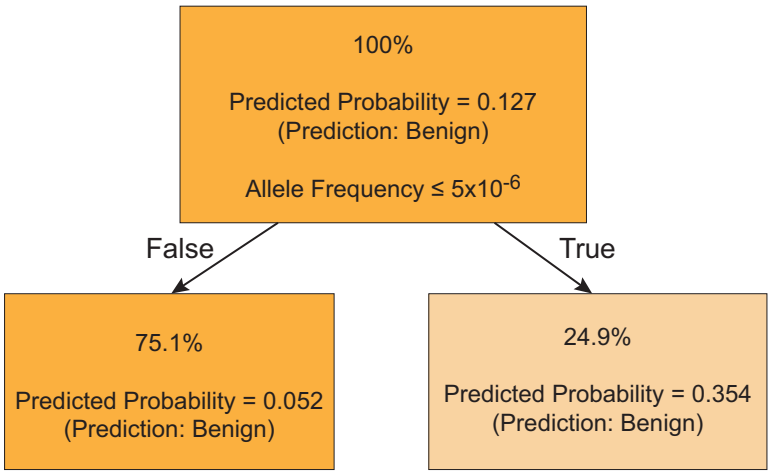


Supp. Figure S3. Distributions of 10-fold cross-validated predictions from binary L2-regularized logistic regression models with increasing levels of feature category inclusion are shown, for P/LP variants only. Feature comparison models are described in Table 2, and AUROC and AUPRC are listed in Figure 1.

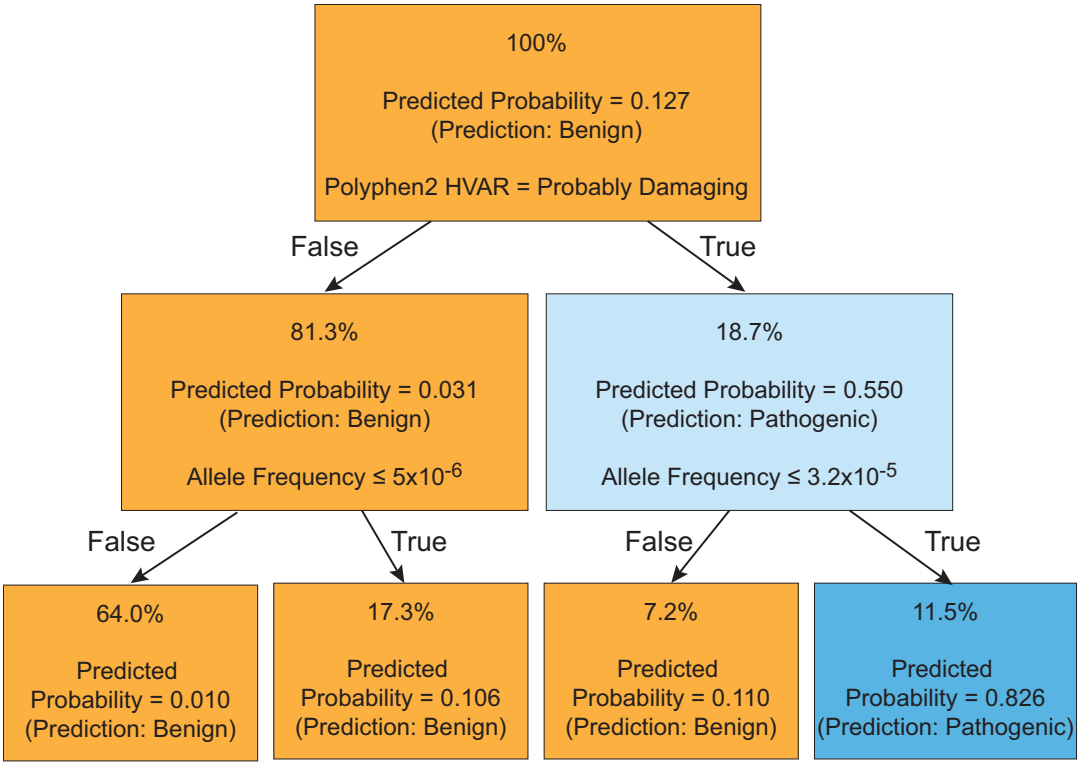


Supp. Figure S4. An illustrated example of how non-linearity between features may be automatically captured by decision trees. The percent of variants, pathogenicity probability (percent of variants at each node that are P/LP, without VUS), predicted classification, feature name, and decision-tree-generated cutoffs are shown at each node. (A) Decision tree visualization with only gnomAD overall population frequency as a feature. (B) Decision tree visualization with gnomAD overall population frequency and Polyphen2 HVAR functional prediction as features.

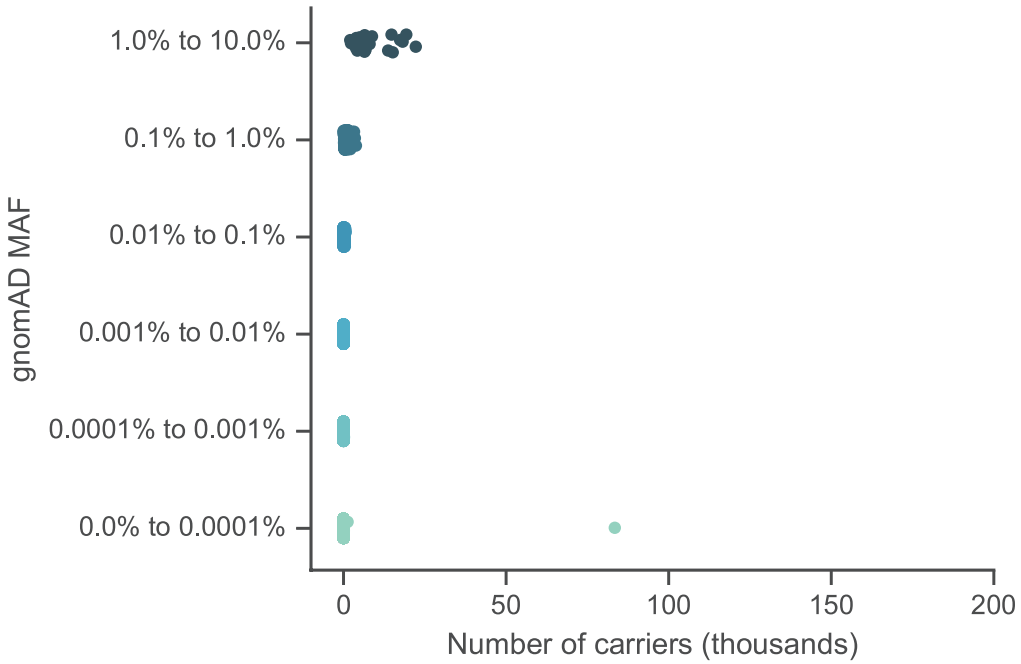
A



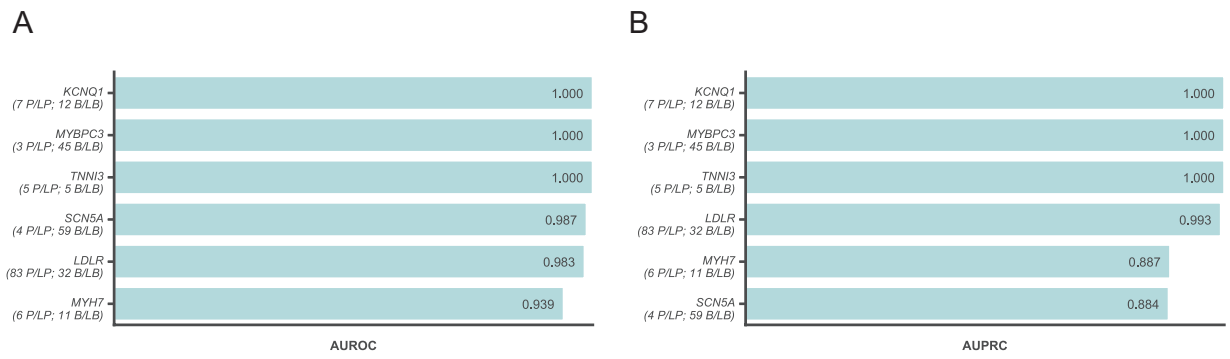
B



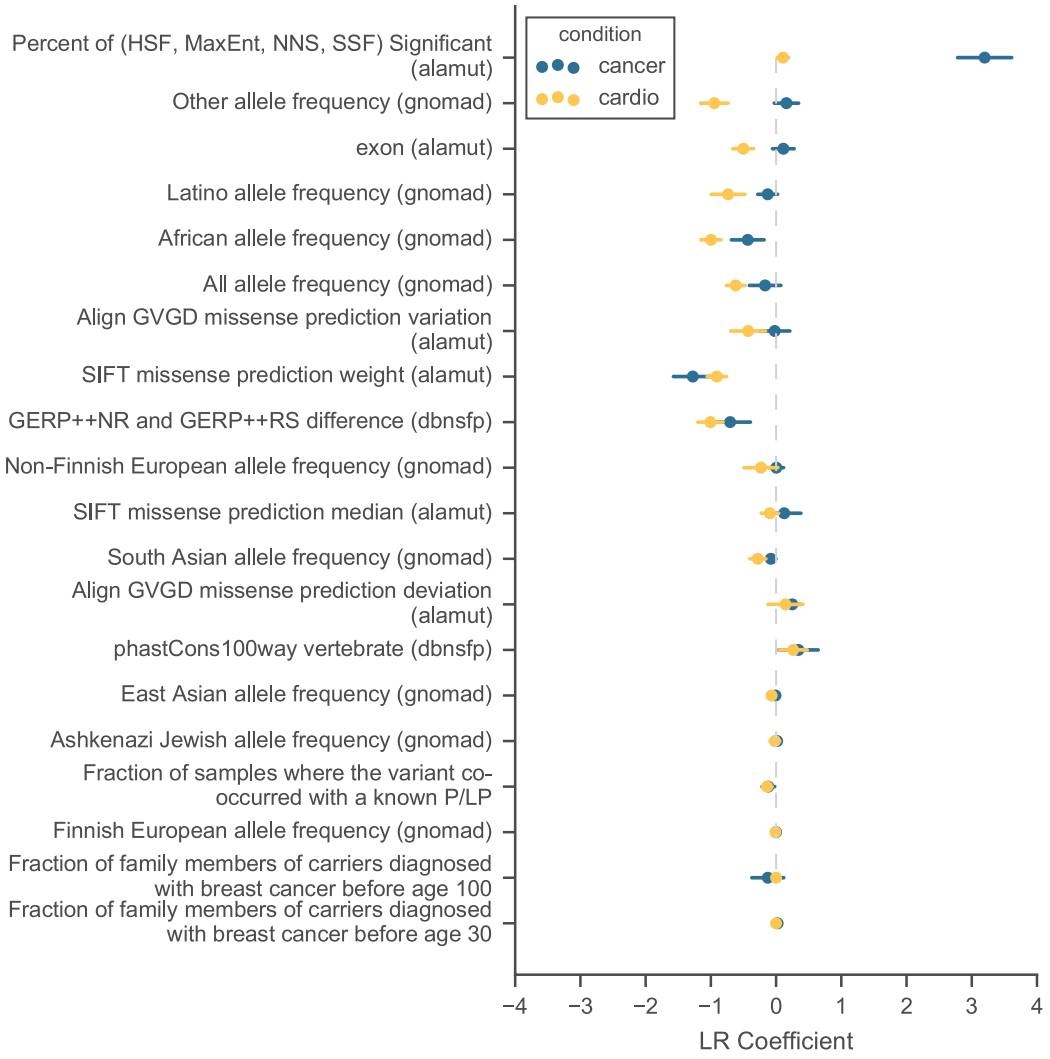
Supp. Figure S5. The number of carriers detected in the Color variant database was compared with overall population frequency from gnomAD. Each point represents a unique hereditary cancer missense variant. The majority (73%) of missense variants detected were very rare (<0.001%) or absent from gnomAD. Variants with gnomAD population frequency greater than 10% make up 0.2% of missense variants detected, and were not shown in this plot.



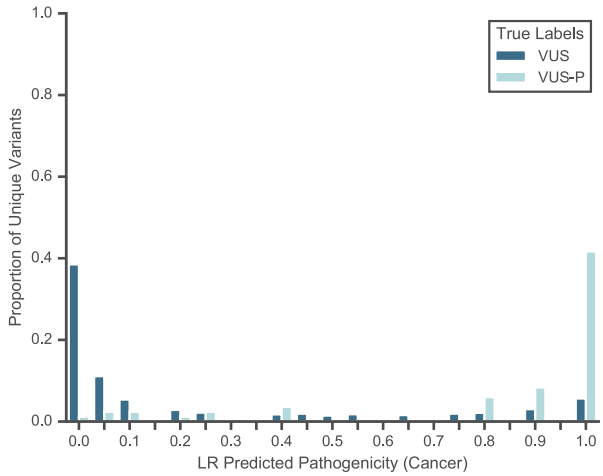
Supp. Figure S6. Gene holdout predictions from a binary L2-regularized logistic regression model using all feature categories (LEAP_MODEL_1 or LEAP_FEATURE_5) were assessed for robustness across different cardiovascular disorder genes. Performance was assessed with (A) AUROC and (B) AUPRC on predictions for variants in each gene withheld from model training. The number of actual P/LP and B/LB variants detected in each gene are listed below the gene name. Genes in which at least 3 P/LP variants were detected were included in this figure.



Supp. Figure S7. A comparison of coefficients between hereditary cancer and cardiovascular disorder L2-regularized binary logistic regression models is shown. Standard deviations of model coefficients were bootstrapped over 1000 models trained on data randomly sampled with replacement. Only common features between cancer and cardiovascular disorders were considered. The top 20 most significant features are shown, and significance was determined by trees feature importances which were scaled and averaged over the bootstrapped models. Features are sorted from largest difference in coefficients between cancer and cardio to smallest.



Supp. Figure S8. Distributions of VUS holdout predictions obtained from a binary logistic regression model trained on P/LP and B/LB variants are shown for hereditary cancer. Model performance on differentiating VUS-P and VUS was assessed with 0.863 AUROC and 0.048 AUPRC (0.029 precision and 0.762 recall using a 0.5 threshold).



Supp. Table S1. Variant database counts by classification groups. Binary classification models were trained on P/LP and B/LB variants only, while multiclass classification models were trained on P/LP, B/LB, and VUS. VUS-P indicates variants internally tagged as one additional piece of evidence away from an LP classification based on ACMG guidelines.

	Hereditary Cancer	Cardiovascular Disorders
P/LP	237	122
B/LB	1618	796
VUS	12371	4480
(VUS-P)	(84)	(19)
Binary total	1855	918
Multiclass total	14226	5398

Additional Files (.xlsx)
Supp. Table S2 Cancer full features list
Supp. Table S3 Cardiovascular full features list