The Journal of Molecular Diagnostics, Vol. ■, No. ■, ■ 2022

the Journal of Molecular Diagnostics
jmdjournal.org

PirePred

An Accurate Online Consensus Tool to Interpret Newborn Screening—Related Genetic Variants in Structural Context

Juan José Galano-Frutos,*† Helena García-Cebollada,*† Alfonso López,*† Mireia Rosell,‡ Xavier de la Cruz,§¶ Juan Fernández-Recio,†‡ and Javier Sancho*†‖

From the Department of Biochemistry and Molecular and Cell Biology, Faculty of Science, University of Zaragoza, Zaragoza; Biocomputation and Complex Systems Physics Institute (BIFI),† Joint Units BIFI-IQFR (CSIC) and GBs-CSIC, University of Zaragoza, Zaragoza; Research Unit in Clinical and Translational Bioinformatics,§ Vall d'Hebron Institute of Research (VHIR), Universitat Autònoma de Barcelona, Barcelona; Institut Catala per la Recerca i Estudis Avançats (ICREA),¶ Barcelona; Instituto de Ciencias de la Vid y del Vino (ICVV),‡ CSIC—Universidad de La Rioja—Gobierno de La Rioja, Logroño; and Aragon Health Research Institute (IIS Aragón),‖ Zaragoza, Spain*

Accepted for publication January 5, 2022.

Address correspondence to Javier Sancho, Department of Biochemistry, Molecular and Cell Biology, Faculty of Science, University of Zaragoza, Zaragoza 50009, Spain. E-mail: jsancho@unizar.es.

PirePred is a genetic interpretation tool used for a variety of medical conditions investigated in newborn screening programs. The PirePred server retrieves, analyzes, and displays in real time genetic and structural data on 58 genes/proteins associated with medical conditions frequently investigated in the newborn. PirePred analyzes the predictions generated by 15 pathogenicity predictors and applies an optimized majority vote algorithm to classify any possible nonsynonymous single-nucleotide variant as pathogenic, benign, or of uncertain significance. PirePred predictions for variants of clear clinical significance are better than those of any of the individual predictors considered (based on accuracy, sensitivity, and negative predictive value) or are among the best ones (for positive predictive value and Matthews correlation coefficient). PirePred predictions also outperform the comparable *in silico* predictions offered as supporting evidence, according to American College of Medical Genetics and Genomics guidelines, by VarSome and Franklin. In addition, PirePred has very high prediction coverage. To facilitate the molecular interpretation of the missense, nonsense, and frameshift variants in ClinVar, the changing amino acid residue is displayed in its structural context, which is analyzed to provide functional clues. PirePred is an accurate, robust, and easy-to-use tool for clinicians involved in neonatal screening programs and for researchers of related diseases. The server can be freely accessed and provides a user-friendly gateway into the structural/functional consequences of genetic variants at the protein level. *(J Mol Diagn 2022, ■: 1—20; https://doi.org/10.1016/j.jmoldx.2022.01.005)*

Newborn screening is widely used for the early detection of different conditions that can evolve to produce severe health problems, including death. Newborn screening programs began in developed countries in the mid-1960s,[1] and although they have since been extended to most countries, there are still vast regions with a poor or no implementation.[2] The number of tested conditions greatly varies from country to country, with the most comprehensive programs including >50 conditions. Screening is commonly based on a blood sample taken from the newborn, which is analyzed using mass spectrometry to identify metabolites with anomalous concentrations.[3] In some cases, the analyses include some targeted genetic screening to detect pathogenic variants in the newborn. When metabolic conditions are detected, the suspected genes may also be sequenced to identify variants that could help to explain the detected condition, to improve diagnosis, or to guide the treatment. Implementation of direct genetic screening programs are

Supported by 2014—2020 Interreg V-A Spain-France-Andorra (POCTEFA) grant EFA086/15, Mineco grant PID2019-107293GB-I00, and Gobierno de Aragón grants E45_17R and LMP30_18. H.G.-C. is the recipient of an FPU16/04232 doctoral contract from MCIN.

Disclosures: None reported.

https://doi.org/10.1016/j.jmoldx.2022.01.005

likely in the not too distant future, which might flood hospital genetics and pediatrics departments with new variants to interpret.[4]

The interpretation of genetic variants may pose different degrees of difficulty. When the detected variant has been described and its occurrence and clinical significance have been carefully annotated in a genetic variants database (eg, ClinVar[5]), the task may be easy. In many cases, however, the variant is poorly characterized or has been found for the first time, and its clinical significance is not firmly established. In the latter cases of variants of uncertain significance (VUSs), the interpretation of the variant involves predicting its phenotypic outcome using bioinformatics tools. For the simpler case of single-nucleotide variants (SNVs) related to monogenic disorders, a multiplicity of pathogenicity predictors[6] are available for researchers and clinicians to anticipate the probable phenotypic consequence of the variation. However, the global accuracy of such predictors seems to have reached a plateau at approximately 85% for binary (pathogenic/benign) predictions.[7] In addition and not infrequently, contradicting predictions are provided for the same variant by different predictors,[8] which has stimulated the recommendation of obtaining several outcomes; however, these outcomes should only be trusted in cases in which highly or totally coincidental predictions are gathered.[9] Sometimes a number of predictions on the same variant are used to generate a consensus prediction applying combination rules, such as a majority vote. For clinicians who are not necessarily expert bioinformaticians, dealing with the generation and analysis of several predictions on the same variant poses considerable difficulties.

Genetic variations in protein coding regions are responsible for the synthesis of proteins with changes in the amino acid sequence. In the simple and most frequent case of missense SNVs,[10] the protein product will bear one amino acid change whose location in the three-dimensional structure of the protein will determine[11–13] whether the protein will be unstable or aggregation prone and will display reduced affinity for substrates or partner proteins, reduced catalytic activity, or any other feature detrimental to its biological activity. Visualizing and analyzing the effect of the variation at the protein structural level is not a guarantee for a precise molecular interpretation, but it can be of help. Further development of more accurate predictors is envisioned when simulation methods providing detailed structural analysis can be incorporated to existing machine learning–based predictors.[14] However, incorporating even simple structural analysis methods to the practice of genetic interpretation is hampered, on the one hand, by some knowledge gaps that exist between the genetics and structural biology fields and, on the other, because detailed structural information on the concerned gene product, usually a protein, may be missing or incomplete. The experimentally determined structural coverage of the human proteome constitutes only 18% of the human protein residues, which can be increased to 50% by homology modeling (*https://swissmodel.expasy.org/repository/species/9606*, last accessed May 22, 2021). Recent massive modelling of the human proteome[15] may further increase this structural coverage.

Newborn screening focuses on several dozen conditions, which are often monogenic and typically associated with well-known genes that code for enzymes for which structural information exists or can be obtained in most cases through homology modeling. This article describes PirePred (*https://pirepred.bifi.es*, last accessed November 20, 2021), an accurate and comprehensive online tool for the easy interpretation of SNVs (missense, nonsense, and frameshift) in 58 genes associated with the principal conditions investigated in newborn screening programs.[16] PirePred aligns with American College of Medical Genetics and Genomics (ACMG) guidelines[9] by providing computational supporting evidence for the interpretation of variants. PirePred combines 15 individual predictors and readily provides, for any nonsynonymous SNV in those 58 genes, a consensus prediction (benign, VUS, or pathogenic) that outperforms the individual predictors in accuracy and in several other relevant quality prediction metrics. Moreover, PirePred provides updated structural models (X-ray or homology) for the corresponding proteins on which any variant described in ClinVar can be readily displayed and further analyzed in the structural context.

Materials and Methods

Server Implementation

PirePred uses Bootstrap version 4.0 (Bootstrap, San Francisco, CA; *https://getbootstrap.com*) for the presentation in the client side (front-end). AJAX technology allows the implementation of requests to the server (gene, protein, or disease) by the user. In the back-end, PHP version 7.3 (Coretechs, Kensington, MD; *https://www.php.net/downloads.php*) connects (in real time) with the ClinVar[5] API and obtains the data query in XML, which is formatted and displayed in a list. The protein structure related to the selected entity (gene, protein, or disease) is shown through the open-source JavaScript viewer JSmol version 14.30.0 (Jmol Development Team).[17] Displayed predictions for all the real and potential variants of a selected gene caused by an SNV are obtained from the dbNFSP version 4.1a repository.[6] The whole information is returned to the user through the interface. Figure 1 includes a general scheme of [F1] the PirePred server implementation. Moreover, Supplemental Table S1 shows PirePred compatibility with the most extended browsers used along with the most common operating systems. An updated browser is recommended, and JavaScript must not be disabled.

Protein Structures and Models

The protein structures depicted with the JSMol version 14.30.0 molecular viewer[17] after selecting a gene, protein,

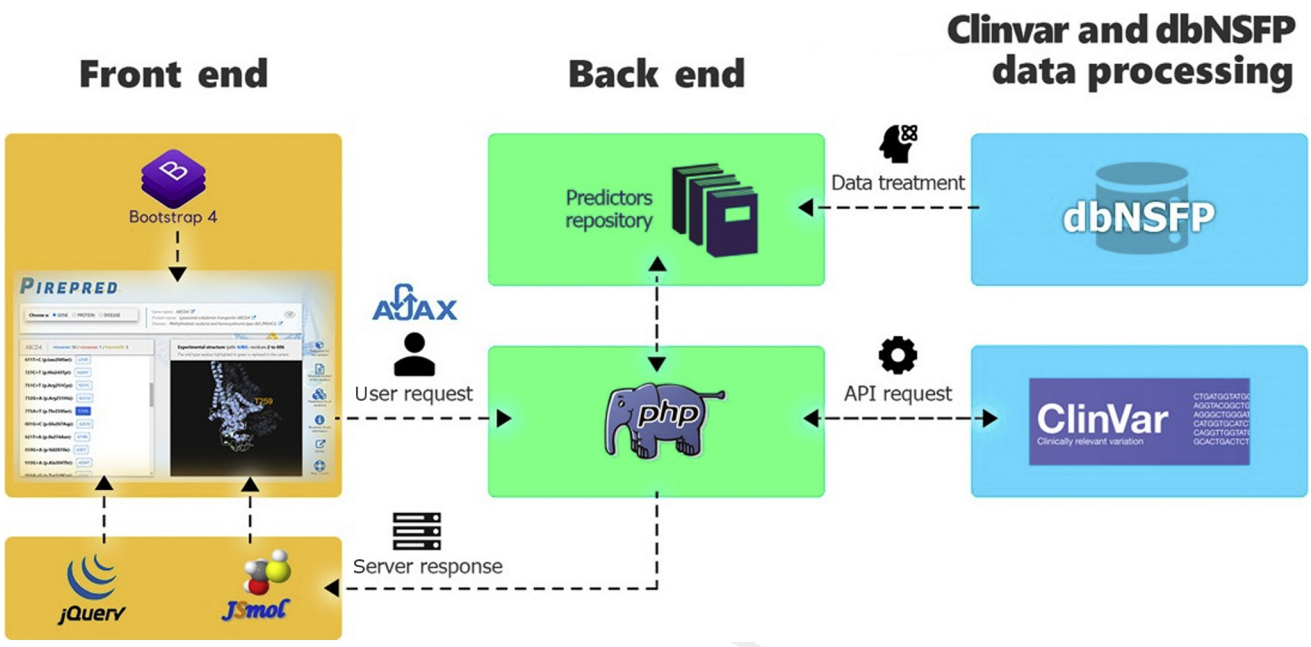


Figure 1 PirePred server's scheme. Overview of the PirePred web server implementation and the involved technologies.

or disease in the PirePred home screen are i) experimentally determined structures available in the Protein Data Bank (PDB) and having a reasonable coverage of the gene sequence, ii) single-template homology models obtained through the SwissModel server or from its repository,[18] iii) single-template threading models obtained from i-Tasser server,[19] or iv) combined structures built by joining an experimentally solved fragment of the protein with a modeled fragment of its unsolved region. For the latter case, the modeled fragments were obtained through the multi-template homology modeling server GPCR-SSFE version 2.0[20] or through the SwissModel[18] (Table 1). These combined models were manually built by following structure alignment (of common solved regions) and distance criteria while avoiding intermonomer and intramonomer clashes according to the solved part of the biological assembly (*SLC25A13* gene). For the special case of the *TSHR*'s combined structure (PDB 2XWT plus a GPCR-SSFE multitemplate homology modeling) (Table 1), because there is no overlapping region between the combined domains, a threading model (i-Tasser) of the whole thyroid-stimulating hormone receptor protein was first obtained, which was used as a geometrical guide for the interdomain distance to properly join the two domains.

Models retrieved from the SwissModel repository were chosen from the list available for each gene not only based on the criterion of having the highest possible coverage and Q_{mean} (integrated quality parameter) but also looking for a template that matched the right oligomerization state predicted or stated for the protein. The same criterion was applied for selecting the best templates for the models built *de novo* through this server. Except for the thyroid-stimulating hormone receptor protein, all the modeled structures were initially modeled with both Swiss-Model and i-Tasser, and the models' quality tested using Molprobity.[21] Table 1 specifies for each gene the origin of the associated protein structure visualized with JSmol. More detailed information about every structure/model is given in an *ad hoc* button included in PirePred (see *Results*).

Structural Context Analysis

Solvent accessible surface areas (SASAs) were calculated for the multimeric and monomeric forms of each protein using an *ad hoc* DSSP[22]-based script. The monomeric form of a multimeric model is obtained by isolating the chain of interest. Relative SASAs for each residue are expressed as the percentage of exposure in the folded protein relative to the mean exposure of that residue type according to a database of folded proteins[23] (the 5% increase in SASA values commonly issued by DSSP, as reported by Estrada et al,[23] was corrected for the calculation of the relative exposure). Moreover, to predict whether a residue is on the multimeric interaction surface, the SASA value was first obtained for the given residue in the multimeric assembly and then compared with that obtained for the residue in the monomeric form. Thus, if the quotient between the multimeric and monomeric SASA values is <0.9, the residue is considered to be buried in the multimeric form and, therefore, to be part of the interaction surface. In addition, the location of the variant in a catalytic or functionally relevant site is checked up by searching the affected residue in the SITE annotations of the original PDB file (if available) and

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:37 am ■ EO: JMDI-D-21-

Table 1 Genes, Related Entries in Relevant Databases, and Details on the Associated Protein Structures Depicted in PirePred

Gene	Ensembl gene set entry*	Associated protein UniProt entry†	Associated disease OMIM entry‡	Current ClinVar SNVs§ (missense/ nonsense/ frameshift), *n*	Protein structure displayed in JSmol (PDB or model template/coverage)¶
ABCD4	ENSG00000119688	O14678	614857	60/1/4	EM structure (PDB: 6JBJ/93%)
ACADM	ENSG00000117054	P11310	201450	151/16/35	X-ray structure (PDB: 4P13*/92%)
ACADS	ENSG00000122971	P16219	201470	108/8/18	X-ray structure (PDB: 2VIG/92%)
ACADSB	ENSG00000196177	P45954	610006	41/4/2	X-ray structure (PDB: 2JIF/88%)
ACADVL	ENSG00000072778	P49748	201475	358/27/87	X-ray structure (PDB: 2UXW/86%)
ACAT1	ENSG00000075239	P24752	203750	100/8/26	X-ray structure (PDB: 2IB8/92%)
ACSF3	ENSG00000176715	Q4G176	614265	96/27/19	SwissModel st-HM (template PDB: 4GXR.A/93%)
ARG1	ENSG00000118520	P05089	207800	62/8/21	X-ray structure (PDB: 3GMZ/98%)
ASL	ENSG00000126522	P04424	207900	105/18/14	X-ray structure (PDB: 1K62*/97%)
ASS1	ENSG00000130707	P00966	215700	118/13/19	X-ray structure (PDB: 2NZ2/98%)
BCKDHA	ENSG00000248098	P12694	248600	91/23/20	X-ray structure (PDB: 1OLX/88%)
BCKDHB	ENSG00000083123	P21953	248600	113/23/30	X-ray structure (PDB: 1OLX/85%)
BTD	ENSG00000169814	P43251	253260	210/28/52	SwissModel st-HM (template PDB: 4CYF.A/90%)
CBS	ENSG00000160200	P35520	236200	213/18/22	X-ray structure (PDB: 4L3V/90%)
CFTR	ENSG00000001626	Q20BH0	219700	1045/225/259	EM structure (PDB: MSM*/79%)
CPT1A	ENSG00000110090	P50416	255120	103/21/9	SwissModel st-HM (template PDB: 2H3W.A/78%)
CPT1B	ENSG00000205560	Q92523	601987	1/0/0	SwissModel st-HM (template PDB: 1T7N.A/78%)
CPT1C	ENSG00000169169	Q8TCG5	616282	37/0/2	SwissModel st-HM (template PDB: 1T7N.A/75%)
CPT2	ENSG00000157184	P23786	255110	196/28/54	SwissModel st-HM (template PDB: 4EP9.A/95%)
CYP11B1	ENSG00000160882	P15538	202010	92/17/17	X-ray structure (PDB: 6M7X/92%)
CYP17A1	ENSG00000148795	P05093	202110	42/9/8	X-ray structure (PDB: 6CIZ/92%)
CYP21A2	ENSG00000231852	P08686	201910	74/7/13	X-ray structure (PDB: 4Y8W*/89%)
DBT	ENSG00000137992	P11182	248600	61/20/15	SwissModel st-HM (template PDB: 2IHW.Q-X/49%)
ETFA	ENSG00000140374	P13804	231680	40/5/5	X-ray structure (PDB: 1EFV/94%)
ETFB	ENSG00000105379	P38117	231680	27/1/2	X-ray structure (PDB: 1EFV/99%)
ETFDH	ENSG00000171503	Q16134	231680	143/11/21	SwissModel st-HM (template PDB: 2GMH.A/94%)
FAH	ENSG00000103876	P16930	276700	91/20/20	SwissModel st-HM (template PDB: 2HZY.A/99%)
FCGR2A	ENSG00000143226	P12318	152700	3/1/0	X-ray structure (PDB: 1H9V*/54%)
GALT	ENSG00000213930	P07902	230400	262/37/40	X-ray structure (PDB: 6GQD*/91%)
GCDH	ENSG00000105607	Q92947	231670	178/27/19	X-ray structure (PDB: 2R0N/89%)
HADHA	ENSG00000084754	P40939	609015, 609016	86/30/30	X-ray structure (PDB: 6DV2/95%)
HADHB	ENSG00000138029	P55084	609015	56/3/4	X-ray structure (PDB: 6DV2/91%)
HBB	ENSG00000244734	P68871	140700, 613985, 603902, 603903	424/30/123	X-ray structure (PDB: 1DXT/100%)
HCFC1	ENSG00000172534	P51610	309541	103/0/0	X-ray structure (PDB: 4GO6/10%)
HMGCL	ENSG00000117305	P35914	246450	51/15/10	X-ray structure (PDB: 2CW6/91%)
HPD	ENSG00000158104	P32754	276710	39/4/3	X-ray structure (PDB: 3ISQ/96%)
HSD3B2	ENSG00000203859	P26439	201810	24/12/4	SwissModel st-HM (template PDB: 3WJ7.A/96%)
IVD	ENSG00000128928	P26440	243500	81/9/17	X-ray structure (PDB: 1IVH/92%)
LMBRD1	ENSG00000168216	Q9NUN5	277380	34/1/4	i-Tasser st-TM (template PDB: 3G5U.A/100%)
MCCC1	ENSG00000078070	Q96RQ3	210200	128/17/25	SwissModel st-HM (template PDB: 5CSL.A/92%)

(*table continues*)

Table 1 (*continued*)

Gene	Ensembl gene set entry*	Associated protein UniProt entry†	Associated disease OMIM entry‡	Current ClinVar SNVs§ (missense/nonsense/frameshift), *n*	Protein structure displayed in JSmol (PDB or model template/coverage)¶
MCCC2	ENSG00000131844	Q9HCC0	210210	107/15/14	SwissModel st-HM (template PDB: 3U9R.F/95%)
MLYCD	ENSG00000103150	O95822	248360	81/2/5	X-ray structure (PDB: 4F0X/92%)
MMAA	ENSG00000151611	Q8IVH4	251100	65/24/20	X-ray structure (PDB: 2WWW/83%)
MMAB	ENSG00000139428	Q96EY8	251110	72/7/12	X-ray structure (PDB: 2IDX/72%)
MMACHC	ENSG00000132763	Q9Y4U1	277400	79/26/37	X-ray structure (PDB: 3SC0/84%)
MMADHC	ENSG00000168288	Q9H3L0	277410	34/5/10	X-ray structure (PDB: 5CV0/57%)
MMUT	ENSG00000146085	P22033	251000	196/56/51	X-ray structure (PDB: 3BIC*/95%)
MTHFR	ENSG00000177000	P42898	236250	102/13/14	X-ray structure (PDB: 6FCX*/91%)
PAH	ENSG00000171759	P00439	261600	543/58/91	X-ray structure (PDB: 6N1K*/93%)
PAX8	ENSG00000125618	Q06710	218700	38/2/2	NMR structure (PDB: 2K27/35%)
PCCA	ENSG00000175198	P05165	606054	105/21/29	SwissModel st-HM (template PDB: 3N6R/91%)
PCCB	ENSG00000114054	P05166	606054	124/16/27	SwissModel st-HM (template PDB: 3N6R/96%)
SLC22A5	ENSG00000197375	O76082	212140	261/27/37	SwissModel st-HM (template PDB: 6N3I.A/69%)
SLC25A13	ENSG00000004864	Q9UJS0	605814	78/18/19	Combined model: X-ray (PDB: 4P5V) + SwissModel st-HM (template PDB: 1OKC.A/76%)
SLC25A20	ENSG00000178537	O43772	212138	31/5/5	SwissModel st-HM (template PDB: 6GCI.A/96%)
TAT	ENSG00000198650	P17735	276600	33/4/5	X-ray structure (PDB: 3DID/88%)
TGFB1	ENSG00000105329	P01137	131300	29/1/0	X-ray structure (PDB: 5VQP*/83%)
TSHR	ENSG00000165409	P16473	275200	69/2/3	Combined model: X-ray (PDB: 2XWT) + GPCR-SSFE 2.0 mt-HM (PDB: many/67%)

*Associated entries in Ensembl (human) database (*https://www.ensembl.org/Homo_sapiens/Info/Index*, last accessed June 1, 2021).
†Associated entries in UniProtKB database (*https://www.uniprot.org/uniprot*, last accessed June 1, 2021).
‡Associated entries in Online Mendelian Inheritance in Man (OMIM) database (*https://www.omim.org*).
§Number of entries per gene in ClinVar database by July 19, 2021 (*https://www.ncbi.nlm.nih.gov/clinvar*).
¶Type of structure (experimental or modeled) shown and available in PirePred. Experimental method and server (program) used to build models or fragments are indicated. Combined models conformed by an experimentally solved protein fragment plus a modeled one, as indicated in *Materials and Methods*. Indicated coverage percentages calculated from the fraction of residues in the depicted structure (in multimeric assemblies the chain with the highest coverage is taken) related to the residues number in the protein canonical sequence. Detailed information about every structure [eg, resolution, missing residues, mutations in the original structure (*), renumbering, template(s), coverage(s), and the server used for modeling when it is the case] can be accessed through the structure/model information button (Figure 3) associated with each protein in PirePred.

mt-HM, multitemplate homology model; PDB, Protein Data Bank; SNV, single-nucleotide variant; st-HM, single-template homology model; st-TM, single-template threading model.

in the Catalytic Site Atlas[24] (*https://www.ebi.ac.uk/thornton-srv/m-csa/*, last accessed May 1, 2021).

Variants Data Set

The data set used to train and test the PirePred consensus classifying algorithm derives from ClinVar.[5] To conform the data set, all missense and nonsense SNVs in the genes of interest (10,052 variants) were considered. Among them, those with at least one star of review status annotation in ClinVar (7895 variants) were initially retrieved (Table 2) and, according to their annotated effects, were assigned to a ternary classification. Thus, variants annotated as benign or likely benign in ClinVar (270 variants) were assigned to the benign group, whereas those annotated as pathogenic or likely pathogenic (2154 variants) were assigned to the Pathogenic group. The 2424 variants in either of these two groups jointly conform the data set used for training and testing as will be detailed. On the other hand, variants annotated in ClinVar as conflicting interpretations of pathogenicity or uncertain significance were not used further for training or testing.

Training and Testing a Consensus Classifier for Missense and Nonsense SNVs

For the PirePred consensus classifier, the ClinVar annotations for the variants in the training data set were compared

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:37 am ■ EO: JMDI-D-21-

Table 2 Composition of the Data Set Retrieved from the ClinVar Database and of the Filtered Set Used to Train the PirePred Consensus Classifier

Data set	Entries, *n*	Relative to the total, %
All ClinVar*	10,052	100.00
Review status with 1+ star	7895	78.54
1+ Star and not uncertain or conflicting†	2424	24.11
1+ Star and pathogenic or likely pathogenic‡	2154	21.43
1+ Star and benign or likely benign§	270	2.69

Data retrieved by June 1, 2021 (*https://www.ncbi.nlm.nih.gov/clinvar*).
*Related to missense, nonsense, and frameshift variants.
†Training data set includes 1687 missense and 737 nonsense variants.
‡Group of ClinVar variants named in this work as pathogenic.
§Group of ClinVar variants named in this work as benign.

with the corresponding binary predictions issued by 15 pathogenicity predictors[8,25–38] (Supplemental Table S2). Initially, 18 pathogenicity predictors were considered based on their performance. Those predictors are implemented in the dbNFSP version 4.1a repository[6] from where their predictions were obtained. Four of these 18 predictors (MVP,[34] CADD,[38] REVEL,[32] and MutPred[33]) do not issue binary predictions but rather scores. Their scores were converted into binary predictions by selecting thresholds that maximized the Matthews correlation coefficient (MCC) value of the binary predictions obtained on the training data set. The external validation group was omitted (Figure 2), whereas accuracy, sensitivity, and positive predictive value (PPV) were kept at >90% of their maximum value (Supplemental Figure S1). Threshold values of 1000 and 1400 were tried for each predictor (Supplemental Table S3) to select the final ones (0.5, 0.6, 0.88, and 3.0, respectively) (Supplemental Table S4). For the remaining 14 categorical predictors, only benign and pathogenic predictions were taken into account (VUS or unavailable predictions discarded). To further refine the initial selection of 18 predictors, their individual and relative performances were compared. For this purpose, a correlation matrix (Supplemental Figure S2) was obtained based on the prediction scores issued for all possible variants of the 58 genes analyzed in PirePred. FATHMM[39] was discarded because of the low correlation shown with most other predictors (Supplemental Figure S2). MetaLR[8] and BayesDel[40] were also discarded to avoid redundancy because they were highly correlated with MetaSVM[8] ($r = 0.92$) and REVEL[32] ($r = 0.92$), respectively. After this refinement, 15 predictors (Supplemental Table S2) remain. They issue the binary predictions that are taken by the PirePred consensus classifier.

PirePred provides, using a modified version of the majority vote algorithm, a ternary consensus classification of variants (benign, VUS, or pathogenic). The classification is based on the fraction of benign predictions issued by the 15 predictors for a given variant, which is used to classify the variant by means of two classifying thresholds. A lower threshold, pathogenic to VUS, is used to separate the variants that will be classified as pathogenic from the rest, whereas a VUS-to-benign higher threshold is used to separate the variants that will be classified as benign from the rest. Variants getting a fraction of benign predictions between the two thresholds or equaling one of them will be classified as VUS. The selection of the best values for the pathogenic-to-VUS and VUS-to-benign thresholds used to classify missense variants has been performed through heatmap analysis of the following quality metrics: PPV, negative predictive value (NPV), MCC, accuracy, sensitivity, specificity, and coverage. For nonsense variants, the selection of the pathogenic-to-VUS and VUS-to-benign thresholds did not require heatmap analysis (further details in *Results*).

For the purpose of training the classifier and to try to detect potential overfitting, the missense variants of the training data set (Table 2) were split into five groups of approximately equal size (Figure 2), each having the same ratio of benign and pathogenic variants. Variants that affect a same residue were always kept in the same group. Groups 1 to 4 were used for training and evaluation tries [learning curve (LC) and leave one out (LOO)], whereas group 5 was omitted and only used to perform an external validation (EV) of the final classifier. First, to ensure that the size of the training data set is enough to represent the variability of the whole data set, the dependence of the thresholds and of the quality metrics on the size of the training data set was analyzed. For this purpose, the model was trained using one (LC 1), two (LC 2), or three groups (LC 3) and the quality metrics described above determined on the remaining three,

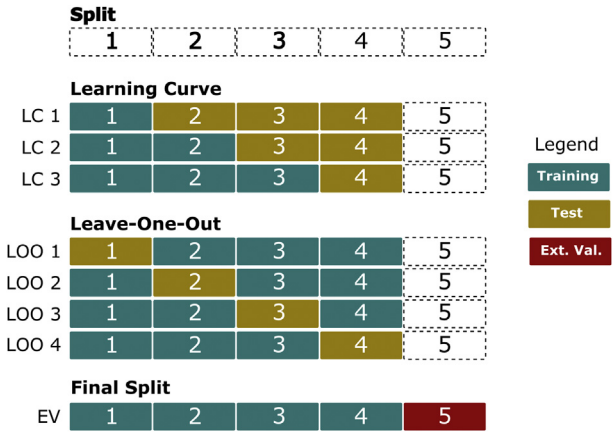
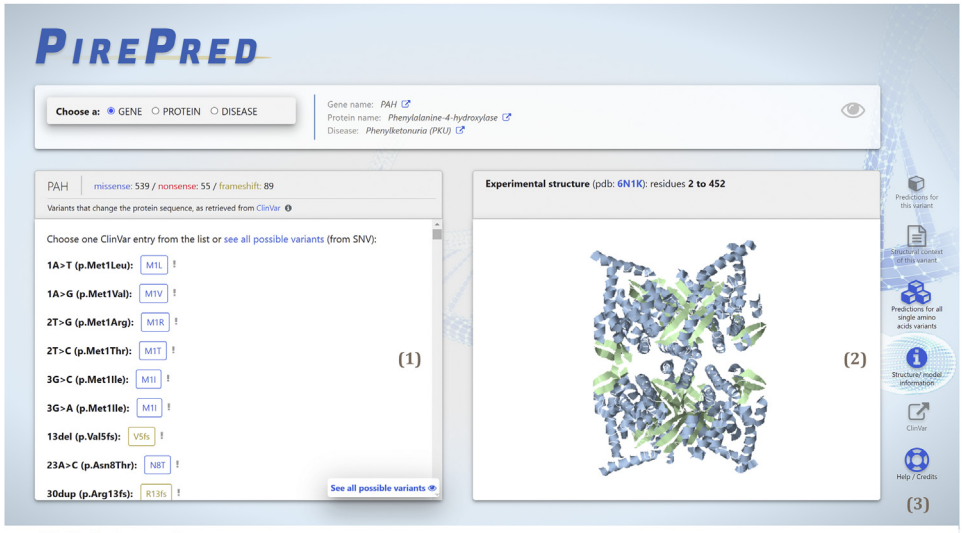


Figure 2 Splitting workflow followed for the missense variants data set to train the classifier and to assess the potential presence of overfitting. The missense variants' data set (1687 variants) is split into five groups in which group 5 is left out and only used for the final external validation (EV, red background) of the classifier. The different splits defined are named after the initials of the training or validation method they have been used for (learning curve, leave one out, and EV). The groups in the different learning curve and leave-one-out tries may be used for training (green background) or testing (brown background), depending on the try.

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:37 am ■ EO: JMDI-D-21-



(1) Variants panel
(2) JSmol panel
(3) Toolbar

Figure 3 Panels that contain or access the relevant information in PirePred. The variants panel (1) displays the list of variants retrieved (in real time) from ClinVar (missense, nonsense, and frameshift) associated with each gene, protein, or disease selected. Direct access to the prediction output for all possible single-nucleotide variants (SNVs) is also given through the prediction for all single amino acid variants button. The JSmol panel (2) allows the depiction of real, modeled, or combined protein structures gathered or built for the server. The toolbar (3) includes the access to the main predictive output by PirePred (prediction for this variant and predictions for all single amino acid variants buttons), to extra information on structural/functional features when the structural fragment that contains the SNV is available (structural context of this variant button), to some detailed information on the biological assembly structure (structure/model information button), to the associated ClinVar entry (ClinVar button), and to the user's Help of PirePred (help/credits button).

two, or one group, respectively (Supplemental Table S5). Second, to assess the goodness of the model and the variability of the quality metrics with the composition of the training and testing data sets, an LOO analysis was performed (LOO 1 to LOO 4) (Figure 2). In this way, quality metrics (Supplemental Table S5) were obtained for each of the four testing groups, after having used the corresponding other three other for training. Finally, after having selected the pathogenic-to-VUS and VUS-to-benign thresholds, the same metrics were evaluated on the external validation group (fifth group) to ensure that they are consistent with the values and variability previously determined in the LOO analysis (Supplemental Table S5). Selection of the pathogenic-to-VUS and VUS-to-benign thresholds for nonsense variants is described in *Results*. As for frameshift variants, most of the predictors used in the PirePred classifier (Supplemental Table S2) do not predict them. Therefore, PirePred cannot classify this type of variants. Nevertheless, frameshift variants are listed and structurally represented in the PirePred server, where the user is advised that they are usually pathogenic.

Feasibility of Incorporating Additional Genes or Novel Predictors to PirePred

Novel genes associated with newborn screening or those already known for which high-quality structural information becomes available at the protein level can be easily incorporated into the PirePred server. The classification algorithm has been trained with and relies on published predictions from reported predictors, summarized in the dbNFSP version 4.1a repository.[6] Therefore, the quality of the predictions will remain stable at the present level without a need for recalibration when new predictors are released because they should not necessarily be incorporated into PirePred. However, new predictors or new versions of the existing ones can be easily incorporated into PirePred after recalibration of the pathogenic-to-VUS and VUS-to-benign thresholds (see above), which may provide opportunities for further improvement of the overall quality metrics.

Results

PirePred Interrogation and Display

The PirePred home screen offers the user the possibility to choose, through a main selection panel, one gene, protein, or disease among 58 commonly investigated in neonatal screening programs. The 58 entries in each of these three selection modes are listed alphabetically. By selecting a gene, protein, or disease, all the associated missense, nonsense, and frameshift variants reported to date in the ClinVar database[5] are retrieved and listed in the variants panel (Figure 3). At the same time, a structure of the concerned protein (experimental, homology model, or combined biological assembly) is depicted in the JSmol panel through the JSmol viewer. A toolbar at the right side includes several functionalities. The prediction for this variant [F3]

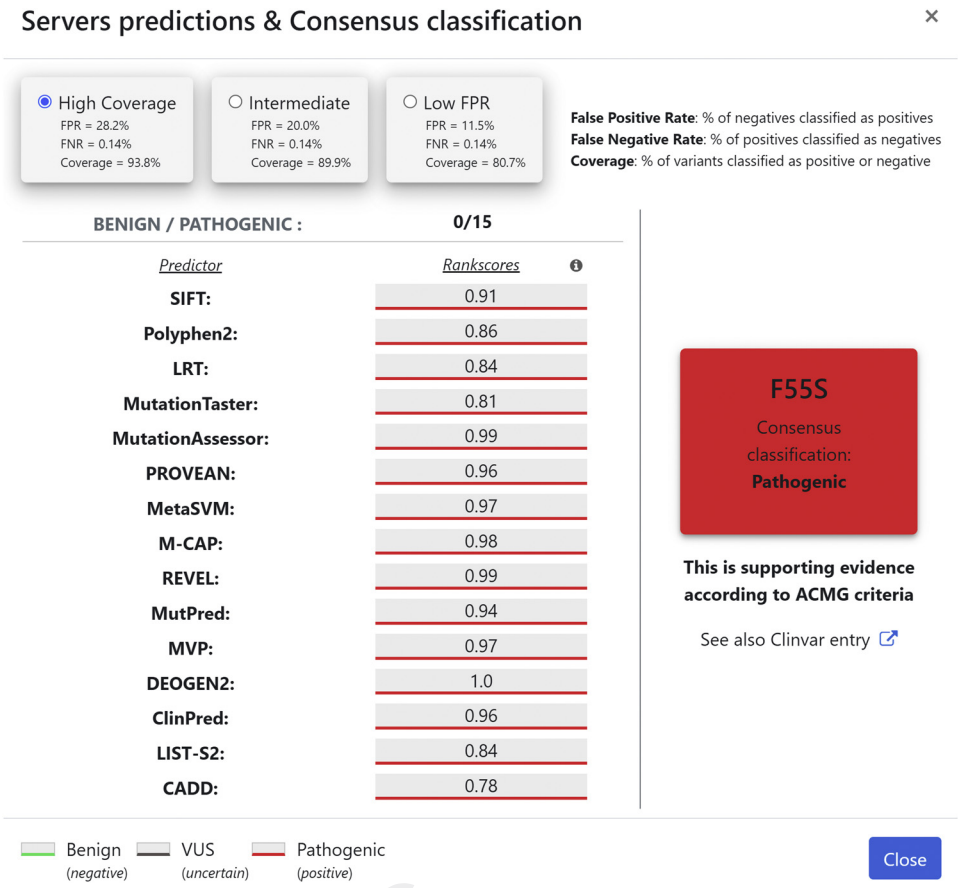


Figure 4 Prediction for this variant button output [example for a single-nucleotide variant (SNV)]. Panel displayed after clicking the prediction for this variant button. A ternary consensus classification for the SNV selected is given by PirePred for each of the three different quality parameters setups [high coverage as default, intermediate, and low false-positive rate (FPR) buttons] to the user (a detailed explanation about these alternative options is given in *Discussion* and is also indicated when hovering over the corresponding buttons). The relying prediction outputs obtained by the 15 prediction tools are also listed. ACMG, American College of Medical Genetics and Genomics; FNR, false-negative rate; FSSS, XXX; VUS, variant of uncertain significance. Q6

print & web 4C/FPO

button gives access to the PirePred ternary consensus classification (benign, VUS, or pathogenic) (Figure 4) for the [F4] variant (only for SNVs, both missense and nonsense) selected in the variants panel. This ternary classification relies on binary predictions given by 15 well-established predictors (Supplemental Table S2), which were selected as explained in *Materials and Methods*. Importantly, users interested in getting the ternary consensus prediction for SNVs (missense or nonsense) not yet reported in ClinVar can do so without having to upload any data. The predictions for all single amino acids variants button gives access to a table that contains the consensus ternary classification for all possible SNVs related to the selected gene, [F5] protein, or disease (Figure 5). The table also includes the predictions issued by the individual predictors, as taken from dbNFSP version 4.1a.[6] PirePred, therefore, enables the user not only to access predictions for currently described SNVs but also to anticipate the effect of any other variant of this type not yet described in patients.

PirePred relates the consensus predictions given for the SNVs to the structural/functional features of the protein concerned. Once a specific variant is selected, the structure depicted in the JSmol viewer zooms in and centers on the changing amino acid residue (Figure 6, A). In the case of [F6] missense SNVs, stability issues associated with the variation's susceptibility to disrupt the protein local conformation or alter neighboring intermonomeric interfaces can be visually evaluated by the user. For nonsense and frameshift variants, suggestive representations of the affected protein segment and a self-explanatory text indicating the putative disruptive structural consequences at the protein level will appear in the top of the JSmol panel (Figure 6, B and C). Moreover, SNVs linked to residues located at missing fragments of the protein structure (nonmodeled parts of the protein, see below on structural coverage) appear labeled with an exclamation mark at their right in the variants panel and will not be represented in the JSmol panel.

For users searching for a reasoned explanation to the consensus classification given for the variant, the structural context of this variant button provides access to relevant structural/functional information on the original residue changed and on its protein context (Figure 7, A). Thus, the [F7] monomeric and multimeric solvent exposure of the changing residue relative to its mean exposure in the unfolded

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:37 am ■ EO: JMDI-D-21-

Mutation	Benign/ Pathogenic	SIFT	Polyphen2	LRT	Mutation Taster	Mutation Assessor	PROVEAN	MetaSVM	M-CAP	REVEL	MutPred	MVP	DEOGEN2	ClinPred	LIST-S2	CADD
M1I	4/10	0.91	0.47	0.1	0.81	NA	0.2	0.98	1.0	0.84	1.0	0.97	0.72	0.74	0.65	0.51
M1K	4/10	0.91	0.54	0.1	0.81	NA	0.13	0.98	1.0	0.86	1.0	0.98	0.71	0.76	0.65	0.6
M1L	6/8	0.91	0.43	0.1	0.81	NA	0.17	0.96	1.0	0.85	1.0	0.94	0.71	0.34	0.65	0.23
M1R	4/10	0.91	0.54	0.1	0.81	NA	0.14	0.98	1.0	0.85	1.0	0.98	0.72	0.75	0.65	0.61
M1T	4/9	0.91	0.54	0.1	0.81	NA	0.18	0.98	1.0	0.87	NA	0.98	0.74	0.73	0.65	0.56
M1V	5/9	0.91	0.43	0.1	0.81	NA	0.19	1.0	1.0	0.86	1.0	0.95	0.72	0.62	0.65	0.4
S2A	10/5	0.09	0.01	0.13	0.81	0.07	0.17	0.86	0.92	0.83	0.03	0.76	0.69	0.09	0.04	0.19
S2C	9/6	0.57	0.01	0.13	0.59	0.11	0.3	0.94	0.93	0.8	0.26	0.92	0.74	0.23	0.15	0.45
S2F	9/6	0.65	0.29	0.13	0.59	0.11	0.27	0.96	0.96	0.81	0.22	0.92	0.73	0.32	0.21	0.48
S2P	10/5	0.48	0.24	0.13	0.81	0.11	0.12	0.89	0.94	0.84	0.03	0.87	0.71	0.2	0.11	0.23
S2T	10/5	0.47	0.01	0.13	0.81	0.11	0.16	0.85	0.86	0.83	0.07	0.78	0.68	0.14	0.08	0.11
S2Y	9/6	0.68	0.32	0.13	0.59	0.11	0.21	0.96	0.95	0.82	0.27	0.93	0.71	0.32	0.19	0.45
T3A	10/5	0.01	0.01	0.32	0.39	0.02	0.06	0.92	0.8	0.81	0.02	0.85	0.65	0.1	0.11	0.11
T3I	9/6	0.53	0.15	0.32	0.81	0.07	0.17	0.93	0.89	0.79	0.16	0.93	0.73	0.28	0.22	0.26
T3N	8/7	0.54	0.19	0.32	0.81	0.07	0.08	0.93	0.9	0.78	0.01	0.92	0.68	0.34	0.21	0.23
T3P	9/6	0.48	0.19	0.32	0.39	0.07	0.11	0.93	0.89	0.82	0.01	0.89	0.73	0.16	0.12	0.19

Figure 5 Predictions for all single amino acid variants button output (example for a gene). Table displayed after clicking the predictions for all single amino acid variants button. The table includes both the PirePred ternary classification and the 15 individual prediction outputs for all possible single-nucleotide variants. In the table, the PirePred predictions are given for each of the three selectable quality parameters setups [high coverage as default, intermediate, and low false-positive rate (FPR)]. ACMG, American College of Medical Genetics and Genomics; FNR, false-negative rate; VUS, variant of uncertain significance.

ensemble, its presence in the interaction surface between monomers if any, or in a spot annotated as a SITE in the PDB file (cofactor binding, catalytic, or assembling) or spot registered as part of an active site in the Catalytic Site Atlas, is provided. Moreover, the structure/model information button leads to detailed information about the shown structure [ie, resolution, missing residues, mutations in the original structure, renumbering, template(s), coverage(s), and the server used for modeling if appropriate] (Figure 7, B). In addition, the atomic coordinates of the structure depicted can be downloaded, which will allow the user to perform a more detailed analysis with other viewers and modeling programs. Finally, the ClinVar button provides a direct link to the ClinVar[5] entry of the selected SNV, and the Help/Credits button enables direct access to a tutorial-like Help for PirePred users.

Structural Coverage

As many as 41 of the 58 proteins encoded by the genes analyzed in PirePred have experimentally determined structures available in the PDB. In most cases, the structural coverage (percentage of residues with determined atomic coordinates) is reasonable. For 2 of these 41 proteins, the coverage was lower, and an experimentally solved fragment was combined with a modeled segment of the missing part to increase the number of residues covered, as described in *Materials and Methods*. Of the remaining 17 structures, 16 were obtained by single-template homology modeling through the SwissModel server[18] ($n = 6$) or directly retrieved from its repository ($n = 10$). For the remaining protein (the product of the *LMBRD1* gene) the best structure model, according to Molprobity[21] (not shown), was obtained from the i-Tasser server[19] (Table 1). The means ± SD structural coverage of the structures shown in PirePred is 85% ± 16%. Thirty-four of these structures cover >90% of their protein sequences, whereas only three cover <50% (Table 1). PirePred thus provides an enhanced experimental coverage for some proteins, as in the cases of the combined structures (*SLC25A13* and *TSHR* genes) and reliable homology models for some of the proteins analyzed.

Training of the Consensus Classifier

The data set with 2424 ClinVar entries (Table 2) was used to select the pathogenic-to-VUS and VUS-to-benign thresholds, which define the intervals that serve to classify the variants as pathogenic, VUS, or benign. Thresholds were

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:37 am ■ EO: JMDI-D-21-

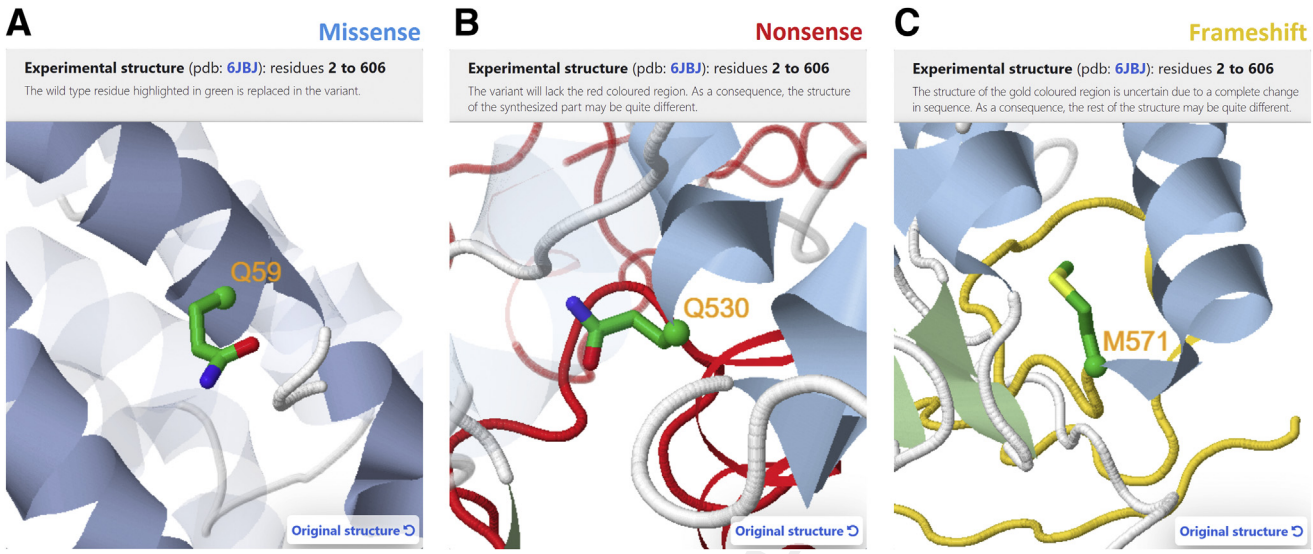


Figure 6 Representation modes in JSmol for the three different types of variants considered in PirePred. **A**–**C:** Missense (**A**), nonsense (**B**), and frameshift (**C**). Self-explanatory texts on the top of each panel describe what it is shown. For nonsense and frameshift variants, the putative consequences triggered by the amino acid change are indicated.

selected separately for missense and for nonsense (only SNVs, not those nonsense variants derived from base duplications or indels) variants. For the missense variants (n = 1687), heatmaps for different quality prediction metrics (PPV, NPV, MCC, accuracy, and coverage) were calculated for all the LC, LOO, and EV tries described in Figure 2. The heatmaps obtained for the final EV try (training with groups 1 to 4 and evaluation on external group 5) are depicted in [F8] Figure 8, which are similar to those obtained for the different LC and LOO tries. For all the tries described in Figure 2, pathogenic-to-VUS and VUS-to-benign threshold values were chosen by applying the following rules: i) the pathogenic-to-VUS threshold is selected to maximize the PPV; ii) the VUS-to-benign threshold is selected to maximize the NPV; and iii) if several threshold values yield similar PPV or NPV (≥0.95), the selection is done to maximize MCC, coverage, and accuracy. in that order of priority. Supplemental Table S5 summarizes the thresholds selected for the training sets considered together with the PPV, NPV, MCC, accuracy, and coverage for those training sets and for their corresponding test sets. In all cases, pathogenic-to-VUS and VUS-to-benign thresholds of 0.4 and 0.7, respectively, were selected.

The LC approach showed similar PPV, NPV, MCC, accuracy, and coverage values independently of training and test set size (Supplemental Table S5). On the other hand, the PPV, NPV, MCC, accuracy, and coverage values obtained for the LOO approach (Supplemental Table S5) and for the final external validation (Supplemental Table S5) indicated that the predictions done for variants that are outside the

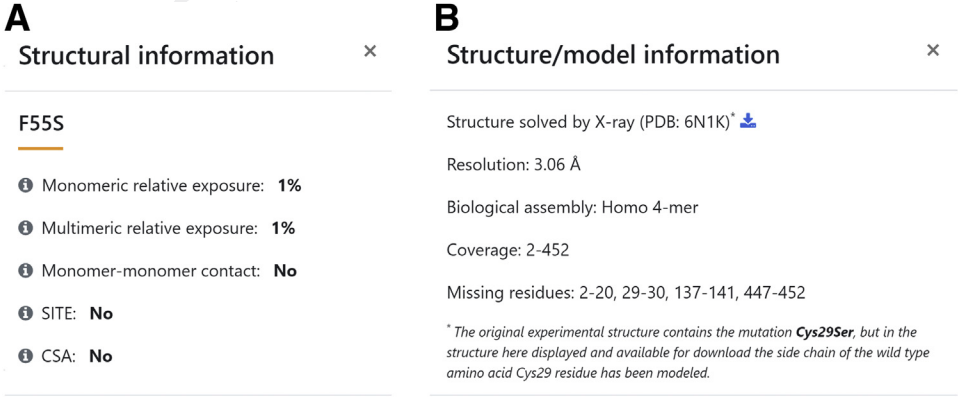


Figure 7 Structural context of this variant and structure/model information buttons outputs [example for a single-nucleotide variant (SNV)]. **A:** Panel displayed after clicking the structural context of this variant button. It contains information on structural functional features in the context of an SNV. They may help to provide, at the protein level, an explanation of the observed phenotype. **B:** Panel obtained after clicking the structure/model information button. It details information on the biological assembly structure depicted in the JSmol viewer. CSA, Catalytic Site Atlas.

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:37 am ■ EO: JMDI-D-21-

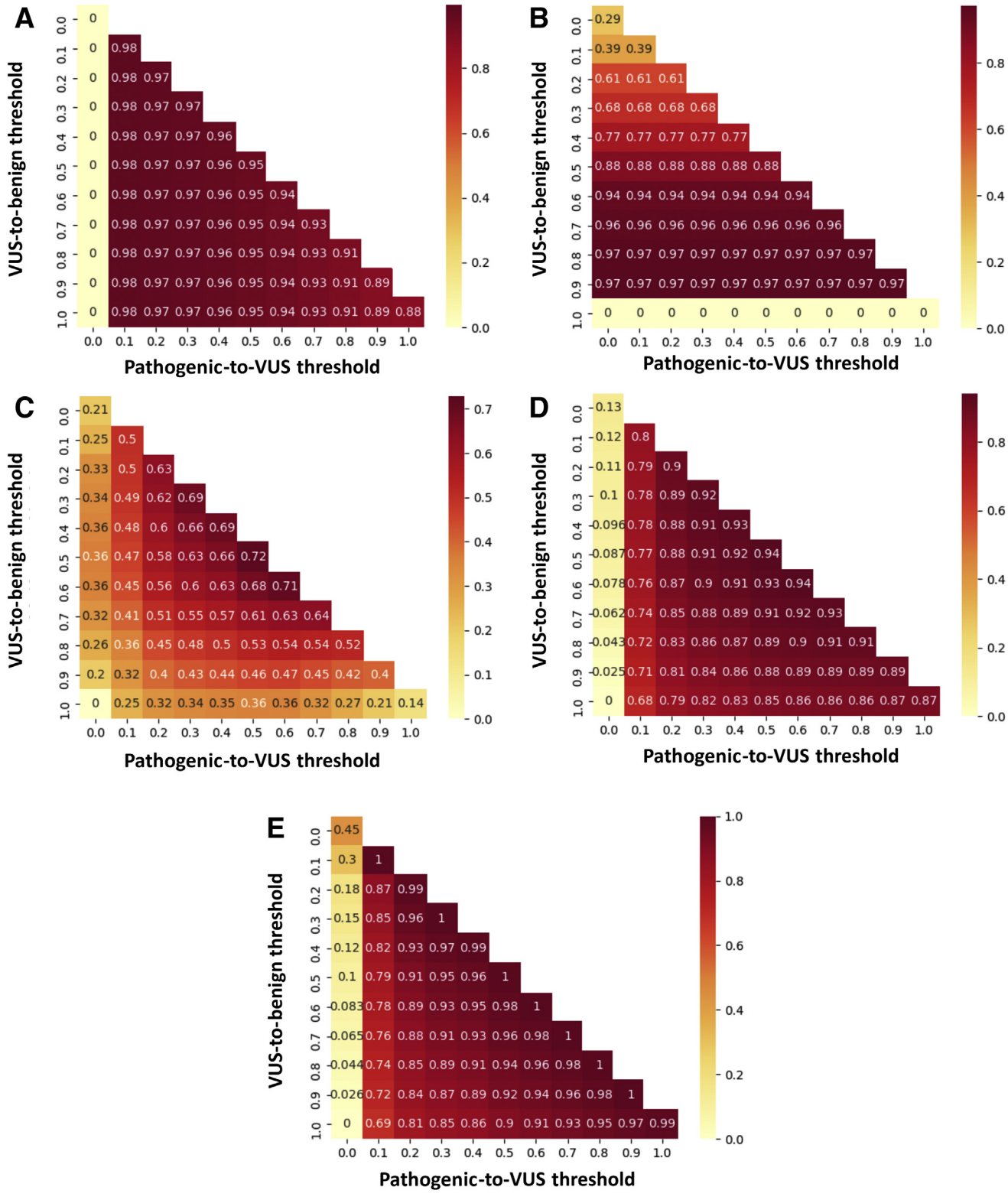


Figure 8 Heatmaps of quality prediction metrics for missense variants used to select the pathogenic—to—variant of uncertain significance (VUS) and VUS-to-benign thresholds for the PirePred ternary classifier. Heatmaps were obtained from the final external validation (EV) (Figure 2). **A**—**E:** For the indicated combinations of pathogenic-to-VUS (*x* axis) and VUS-to-benign (*y* axis) classifying thresholds, the corresponding values of positive predictive value (**A**), negative predictive value (**B**), Matthews correlation coefficient (**C**), accuracy (**D**), and coverage (**E**) are displayed on a background colored as indicated in the gradient bar at the right of each plot. Values for pathogenic-to-VUS threshold higher than VUS-to-benign thresholds have no physical sense and are not shown. The first column on the left in the heatmap for positive predictive value and the **bottom row** in the heatmap for negative predictive value are artifacts caused by the mathematical impossibility of diving 0 by 0.

Table 3 Predictive Statistics and Quality Metrics Obtained for the PirePred Consensus Classifier and the 15 Individual Predictors Selected from dbNFSP Version 4.1a and Used in the Majority Vote Algorithm

PirePred or predictor	Variants with a binary prediction*	Nonpredictor pathogenic variants†	Nonpredictor benign variants†	TP	FP	TN	FN	Coverage‡
PirePred‖‖								
High coverage (standard)	2274	43	107	2108	76	87	3	0.938
Intermediate	2118	115	129	2036	54	87	3	0.899
Low FPR	1955	317	152	1834	31	87	3	0.807
Mutation Taster	2419	3	2	2117	123	145	34	0.998
CADD	2424	0	0	1963	68	202	191	**1.000*****
LRT	2214	183	27	1730	84	159	241	0.913
ClinPred	1724	699	1	1372	22	247	83	0.711
REVEL	1687	736	1	1365	70	199	53	0.696
MetaSVM	1687	736	1	1270	66	203	148	0.696
LIST-S2	1639	772	13	1302	114	143	80	0.676
MVP	1536	826	62	1249	90	118	79	0.634
M-CAP	1563	740	121	1344	62	87	70	0.645
PROVEAN	1534	870	20	1133	79	171	151	0.633
SIFT	1534	870	20	1158	86	164	126	0.633
DEOGEN2	1523	867	34	1151	74	162	136	0.628
PolyPhen-2	1553	843	28	1159	68	174	152	0.641
Mutation Assessor	1450	938	36	1135	106	128	81	0.598
MutPred	1178	1046	200	1039	19	51	69	0.486

(*table continues*)

*Number of variants obtained from the ClinVar group 1+ star and not uncertain or conflicting in Table 2 (2424) minus those in the third and fourth columns from the left.

†Number of variants classified (PirePred) or predicted (individual predictors) as VUSs or with an unavailable prediction.

‡Coverage is the fraction of variants in the data set (2424 total variants) that are classified as benign or pathogenic (ie, those not classified as VUS or unpredicted).

§Accuracy is calculated as (TP + TN)/(TP + FP + TN + FN), which represents the percentage of correct predictions of all binary predictions.

¶Matthews correlation coefficient: $MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}}$.

‖PPV = TP/(TP + FP), which is the fraction of pathogenic predictions corresponding to pathogenic variants; NPV = TN/(TN + FN), which is the fraction of benign predictions corresponding to benign variants.

**Sensitivity = TP/(TP + FN).

††Specificity = TN/(TN + FP).

‡‡FPR = FP/Negatives = FP/(TN + FP + Nonpredictor Benign Variants).

§§FNR = FN/Positives = FN/(TP + FN + Nonpredictor Pathogenic Variants).

¶¶Fraction of variants not being predicted as positive or negative. Equals 1 − coverage.

‖‖The PirePred server classifies variants using thresholds (pathogenic to VUS and VUS to benign) based on the fraction on benign predictions recovered for the variant. The pathogenic-to-VUS threshold for missense variants can be modified to increase specificity at the expense of coverage. PirePred achieves maximal coverage in standard mode using the standard pathogenic-to-VUS threshold of 0.4, whereas in intermediate and low FPR modes, predictions are issued using pathogenic-to-VUS thresholds of 0.24 and 0.08, respectively.

***Maximum (or minimum in the case of FPR, FNR, and uncertain rate) values highlighted in bold for each of the quality metrics obtained for PirePred (high coverage) and the listed predictors (per column).

†††Values highlighted in bold when the best performance is obtained by PirePred with intermediate or with low FPR prediction modes.

FN, false negative (false benign); FP, false positive (false pathogenic); FPR, false-positive rate; TN, true negative (true benign); TP, true positive (true pathogenic); VUS, variant of uncertain significance.

training set are as good as those for variants that are in, which is consistent with a lack of overfitting in the classifying algorithm.

For nonsense variants, no heatmaps were necessary to select the pathogenic-to-VUS and VUS-to-benign thresholds. On the one hand, only 1 of the 737 variants of this type present in the data set (Table 2) was annotated as benign, which is consistent with the idea that producing a protein lacking a fragment should normally trigger pathogenic consequences. On the other hand, because the fraction of benign predictions obtained for this single variant from the selected predictors was 0.33 (1 of 3 predictors) and there were only two pathogenic variants with a higher fraction of benign predictions, it was decided that nonsense variants with a fraction of $\geq$0.34 are considered VUSs. Thus, pathogenic-to-VUS and VUS-to-benign thresholds of 0.34 and 1 were set up, respectively, so that variants predicted as benign by a fraction of predictions $<$0.34 are classified as

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:37 am ■ EO: JMDI-D-21-

Table 3 (*continued*)

Accuracy[§]	MCC[¶]	PPV[‖]	NPV[‖]	Sensitivity**	Specificity[††]	FPR,[‡‡] %	FNR,[§§] %	Uncertain rate,[¶¶]%
0.965***	0.704	0.965	**0.967*****	**0.999*****	0.534	28.1	**0.1*****	6.2
0.974†††	0.761	0.974	**0.967**†††	**0.999**†††	0.617	20.0	**0.1**†††	10.1
0.983†††	**0.836**†††	0.983	**0.967**†††	**0.998**†††	0.737	11.5	**0.1**†††	19.3
0.935	0.630	0.945	0.810	0.984	0.541	45.6	1.6	0.2
0.893	0.563	0.967	0.514	0.911	0.748	25.2	8.9	**0.0*****
0.853	0.432	0.954	0.398	0.878	0.654	31.1	11.2	8.7
0.939	**0.794*****	**0.984*****	0.748	0.943	**0.918*****	8.1	3.9	28.9
0.927	0.721	0.951	0.790	0.963	0.740	25.9	2.5	30.4
0.873	0.586	0.951	0.578	0.896	0.755	24.4	6.9	30.4
0.882	0.529	0.919	0.641	0.942	0.556	42.2	3.7	32.4
0.890	0.520	0.933	0.599	0.941	0.567	33.3	3.7	36.6
0.916	0.522	0.956	0.554	0.950	0.584	23.0	3.3	35.5
0.850	0.514	0.935	0.531	0.882	0.684	29.3	7.0	36.7
0.862	0.526	0.931	0.566	0.902	0.656	31.9	5.9	36.7
0.862	0.530	0.940	0.544	0.894	0.686	27.4	6.3	37.2
0.858	0.537	0.945	0.534	0.884	0.719	25.2	7.1	35.9
0.871	0.503	0.915	0.612	0.933	0.547	39.3	3.8	40.2
0.925	0.521	0.982	0.425	0.938	0.729	**7.0*****	3.2	51.4

pathogenic, whereas those predicted by a fraction $\geq$0.34 are classified as VUSs.

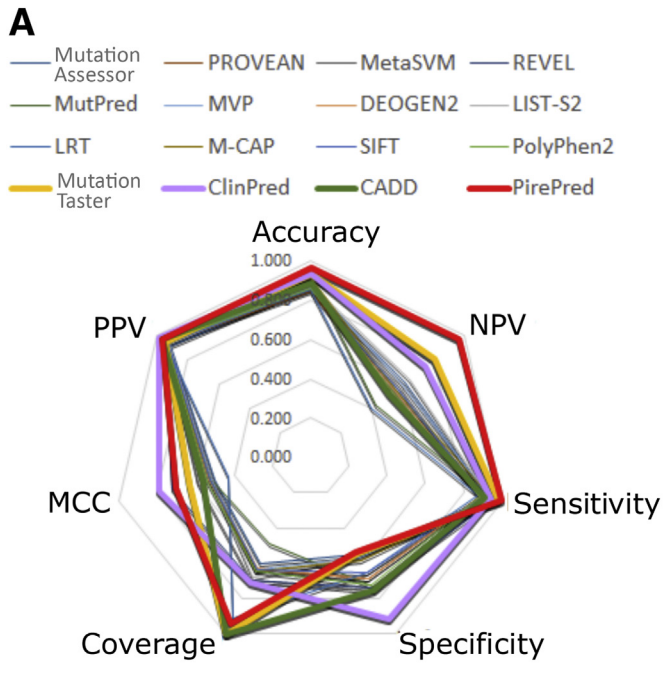
The indicated classification thresholds (0.4/0.7 for missense and 0.34/1.0 for nonsense variants) will be referred to as PirePred's standard thresholds and represent a fine compromise among several quality performance metrics, as described in the following section.

Discussion

Performance of the Consensus Classifier

The predictive statistics obtained by PirePred (with standard thresholds) and by each of the 15 individual binary predictors for variants described in ClinVar as 1+ star and not

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:38 am ■ EO: JMDI-D-21-



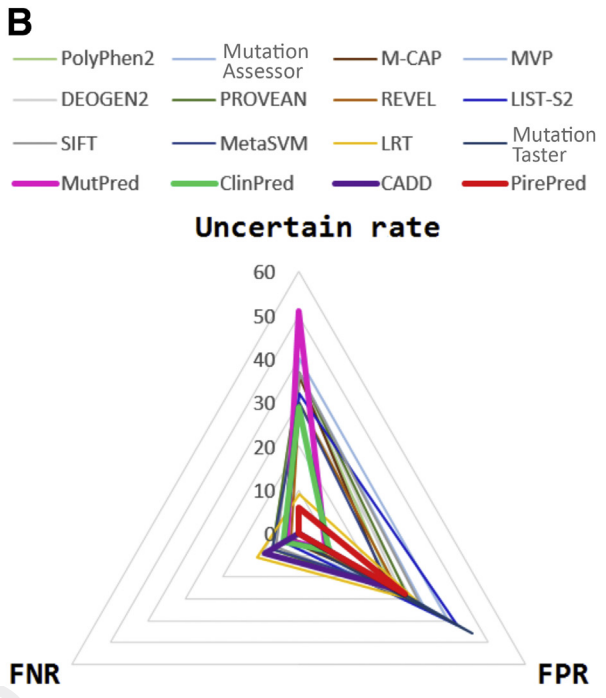


Figure 9 Comparison of quality metrics for PirePred and the selected 15 predictors used with the majority vote algorithm. **A:** Radar plot to compare the performance of PirePred (standard thresholds: high coverage mode) and the 15 individual predictors on the ClinVar data set. The seven quality metrics in Table 3 that in a perfect predictor would take the maximum value of 1 are shown with thick lines: PirePred and three other predictors (ClinPred, MutationTaster, and CADD) showing the best overall performances. **B:** Fault triangles of PirePred (standard thresholds: high coverage mode) and the 15 individual predictors. The three quality metrics in Table 3 [false positive rate (FPR), false negative rate (FNR), and uncertain rate] than would take the minimum value of zero in a perfect predictor are shown. Their values (in percentages) are joined by lines that define triangles. The smaller the edges of the triangle, the better the metrics represented. The fault triangle of an ideal predictor would consist of a single point at the center of the graph. MCC, Matthews correlation coefficient; NPV, negative predictive value; PPV, positive predictive value.

[T3] uncertain or conflicting (Table 2) are given in Table 3. This table includes the number of variants predicted by each predictor in the data set (also those not predicted or predicted as VUSs), the breakdown of true and false pathogenic [true positive (TP) and false positive (FP)] and true and false benign [true negative (TN) and false negative (FN)] predictions, and the quality prediction metrics obtained on a binary base (benign/pathogenic) for each predictor and for the PirePred consensus classifier. Variants classified by PirePred as benign can be aligned with ACMG guidelines as fulfilling the BP4 supporting benign criterion, whereas those classified as pathogenic fulfill the PP3 supporting pathogenic criterion. The superiority of PirePred in predictive power compared with the individual predictors can be noted for many of the metrics. Coverage reports the percentage of variants for which the classifying algorithm issues a prediction (benign or pathogenic). With the standard thresholds (high coverage button in Figures 4 and 5), the PirePred coverage (93.8%) is the third highest, only after those of CADD[38] (100%) and MutationTaster[28] (99.8%), slightly above that of LRT (91.3%), and significantly larger than those offered by the rest of predictors (≤71.1%). Accuracy reports the fraction of correct predictions of all predictions given. PirePred accuracy (96.5%) outperforms all others, followed by ClinPred[36] (93.9%) and MutationTaster[28] (93.5%). The lowest accuracy (85.0%) is provided by PROVEAN.[30] The MCC, which takes into account the four elements of the confusion matrix (TP, FP, TN and FN) is considered to provide a more balanced measure of the quality of a binary classification than accuracy alone. PirePred MCC (0.70) is only outperformed by ClinPred[36] (0.79) and REVEL[32] (0.72), whose respective coverages, 71.1% and 69.6%, are much lower than that of PirePred. On the other hand, PirePred accuracy and MCC are higher than those of the only two predictors (CADD[38] and MutationTaster[28]) that offer a larger coverage. The lowest MCC is provided by LRT[27] (0.43).

PirePred excels at identifying pathogenic variants. PPV and sensitivity in Table 3 indicate, respectively, the percentage of pathogenic predictions that correspond to pathogenic variants (according to ClinVar) and the percentage of pathogenic variants that are predicted as such. PirePred PPV (96.5%) is below those of ClinPred[36] (98.4%) and MutPred[33] (98.2%) and similar to that of CADD (96.7%), whereas PirePred sensitivity (99.9%) is the highest followed by MutationTaster[28] (98.4%). A good PPV and coverage performance of MutationTaster is guaranteed by the fact that this predictor automatically predicts as disease-causing any variant that is marked as pathogenic in ClinVar.

Concerning the identification of benign variants, PirePred greatly outperforms all the individual predictors in NPV (96.7%), followed by MutationTaster[28] (81%) and REVEL[32] (79%). This finding means that PirePred rarely misclassifies a pathogenic variant as benign. Thus, with standard thresholds, PirePred achieves excellent coverage and the highest accuracy, NPV, and sensitivity of all 15 individual predictors [F9] (Figure 9, A), and, as discussed, it does not seem to be overfitted to its training data set. Nevertheless, PirePred pays a price for those generally superior predictive performance metrics in its specificity value (53.4%), which is the lowest among those of the 15 individual predictors. However, the low specificity of a ternary classifier such as PirePred may arise for two reasons with very different implications for clinical use: it might reflect that many benign variants are misclassified as pathogenic, increasing the number of FP results or that they are declared as VUSs, not having an impact on the number of FP results. PirePred (with standard thresholds) declares approximately 40% of benign variants (107 of 270) as VUSs, which lowers the number of TN results but does not increase the number of FP results and lowers its specificity, defined as TN/(TN + FP). Importantly, this does not increase its FP rate (FPR) [defined as FP/(TN + FP + UB)] (Table 3), which reports the fraction of benign variants subjected to evaluation that are wrongly classified as pathogenic. The FPR of PirePred (28.1%) is close to the mean of the 15 individual predictors (27.9%).

Indications for a Prevalence-Dependent Use of PirePred in a Clinical Setting

The practical usefulness of a predictive test, such as a variants classifier, is related to the prevalence of the condition investigated in the population analyzed.[41] At constant sensitivity and specificity (ie, having chosen a given test or predictor), an increase in the prevalence increases the PPV and decreases the NPV of the test, whereas a decrease in the prevalence decreases the PPV and increases the NPV. Therefore, when the test is used on a high prevalence population, a main concern may be the FN results (missed cases), whereas testing in a low prevalence population the concern may be the FP results (false alarms), which may translate in their further testing or undue treatment. In the context of genetic interpretation associated with newborn screening, if the variants subjected to analysis come from individuals who have been derived to DNA sequencing after a metabolic screening, the prevalence is expected to be high. However, if genetic screening programs for newborns become more common, the expected prevalence of any individual condition will be low, bearing in mind that the prevalence of the population is important to select the predictor/classifier and, given their quality metrics, to understand the implications in terms of false alarms and missed cases.

In this respect, the PirePred standard thresholds can be modified to adapt performance (which excels at not missing positive cases) to populations of lower prevalence. Specifically, the pathogenic-to-VUS threshold used for missense variants (0.4) can be changed to trade between specificity and coverage while leaving the excellent sensitivity virtually unchanged. Thus, in addition to the classification obtained using the standard thresholds, the PirePred server provides predictions using an intermediate pathogenic-to-VUS threshold of 0.24 for missense variants (intermediate button) or a lower threshold of 0.08 (lowFPR button) (Figures 4 and 5). The quality metrics obtained with those alternative thresholds are shown in Table 3. As anticipated, the standard pathogenic-to-VUS threshold (0.4) offers the higher coverage (94%) and a lower specificity (0.53), the intermediate pathogenic-to-VUS threshold (0.24) reduces the coverage to 90% but increases the specificity to 0.62, whereas the lower pathogenic-to-VUS threshold (0.08) further reduces the coverage to 81% and further increases the specificity to 0.74. The predictions obtained using the standard pathogenic-to-VUS threshold (high coverage button) are provided as the default by the server, but the user can click buttons to change from standard- to intermediate-(intermediate button) or to low-threshold—based (lowFPR button) predictions to adapt to the expected prevalence or to suit particular needs (Figures 4 and 5). In short, going from high coverage to lowFPR mode lowers the coverage but improves or leaves virtually unmodified all the other quality metrics for which a binary classification is issued (Table 3).

The performance of PirePred can be compared with the performance of the 15 related predictors by focusing on three metrics whose values are independent of the prevalence and may be of particular interest for its use in a clinical context: the FPR (percentage of benign variants predicted as pathogenic), the false-negative rate (FNR) (percentage of pathogenic variant predicted as benign), and the uncertain rate (percentage of variants not classified as positive or benign, which equals 1 − coverage). Figure 9, B presents a fault triangles plot built from the values of those three quality metrics for PirePred and the different predictors. Because the three metrics should get null values for a perfect predictor, the smaller the edges (closer vertices to the plot center) of the fault triangle the better. The PirePred fault triangle is one of the smaller ones because the FNR (0.1%) is the smallest one, the uncertain rate (6.2%) is the third smallest one, and the FPR (28.1%) is in the mean (Table 3).

The size and shape of the PirePred fault triangle change, depending on the value of the missense pathogenic-to-VUS threshold (Supplemental Figure S3). As the threshold is lowered, the FPR decreases and the uncertain rate increases. When PirePred is used with the lowFPR thresholds, the very low FNR of 0.1% of the standard thresholds is retained (few cases are missed), and the FPR is reduced to 11.5% (much less false alarms) at the expense of increasing the uncertain rate to 19.3%. In the lowFPR mode, PirePred displays the highest accuracy, MCC, PPV (together with ClinPred), NPV, sensitivity, and FNR. Its specificity is the fifth best one (but note that its FPR is second best after ClinPred) and the coverage is still the fourth best one.

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:38 am ■ EO: JMDI-D-21-

Table 4 Predictive Statistics and Quality Metrics Obtained by the PirePred Consensus Classifier, the 15 Individual Predictors Selected from dbNFSP Version 4.1a, and the *in Silico* Predictive Tools from VarSome and Franklin Online Servers

PirePred or predictor	Variants with a binary prediction*	Nonprediction pathogenic variants†	Nonprediction benign variants†	TP	FP	TN	FN	Coverage‡
PirePred***								
High coverage (standard)	200	9	15	177	10	13	0	0.893
Intermediate	187	17	20	169	5	13	0	0.835
Low FPR	164	37	23	149	2	13	0	0.732
Mutation Taster	223	0	1	186	23	14	0	0.996
CADD	224	0	0	149	3	35	37	**1.000**†††
LRT	209	12	3	158	16	19	16	0.933
ClinPred	155	69	0	115	3	35	2	0.692
REVEL	150	74	0	112	23	15	0	0.670
MetaSVM	150	74	0	108	20	18	4	0.670
LIST-S2	149	74	1	106	22	15	6	0.665
MVP	145	76	3	108	26	9	2	0.647
M-CAP	147	74	3	112	30	5	0	0.656
PROVEAN	146	78	0	95	9	29	13	0.652
SIFT	146	78	0	92	13	25	16	0.652
DEOGEN2	140	81	3	97	12	23	8	0.625
PolyPhen-2	140	81	3	88	8	27	17	0.625
Mutation Assessor	135	86	3	89	12	23	11	0.603
MutPred	121	84	19	99	7	12	3	0.540
Varsome (*in silico*)§§§	210	4	10	182	14	14	0	0.938
Franklin (*in silico*)¶¶¶	114	86	24	100	10	4	0	0.509

(*table continues*)

The assessed data set consisted of 224 newly reported variants (ClinVar) released after the full training and evaluation of PirePred reported in Table 3. Variants newly released in ClinVar [classified as likely benign, benign, likely pathogenic or pathogenic, with review status of one or more stars (*https://www.ncbi.nlm.nih.gov/clinvar/docs/review_status*)] released after the full training and evaluation of PirePred. A total of 43 of the 58 genes analyzed in PirePred are represented in this external set. Pathogenicity verdicts accessed by July 19, 2021 (*https://www.ncbi.nlm.nih.gov/clinvar*).

*Number of variants predicted as pathogenic or benign by PirePred or the individual predictors used in the majority vote algorithm.

†Number of variants classified (PirePred) or predicted (individual predictors) as VUSs or with an unavailable prediction.

‡Coverage is the fraction of variants in the data set (224 total variants) that are classified as benign or pathogenic (ie, those not classified as VUSs or unpredicted).

§Accuracy is calculated as (TP + TN)/(TP + FP + TN + FN), which represents the percentage of correct predictions of all binary predictions.

¶Matthews correlation coefficient: $MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}}$

‖PPV = TP/(TP + FP), which is the fraction of pathogenic predictions corresponding to pathogenic variants.

**NPV = TN/(TN + FN), which is the fraction of benign predictions corresponding to benign variants.

††Sensitivity = TP/(TP + FN).

‡‡Specificity = TN/(TN + FP).

§§FPR = FP/Negatives = FP/(TN + FP + Nonpredictor Benign Variants).

¶¶FNR = FN/Positives = FN/(TP + FN + Nonpredictor Pathogenic Variants).

‖‖Fraction of variants not being predicted as positive or negative. Equals 1 − coverage.

***The PirePred server classifies variants using thresholds (pathogenic to VUS and VUS to benign) based on the fraction of benign predictions recovered for the variant. The pathogenic-to-VUS threshold for missense variants can be modified to increase specificity at the expense of coverage. PirePred achieves maximal coverage in standard mode using the standard pathogenic-to-VUS threshold of 0.4, whereas in intermediate and low FPR modes, predictions are issued using pathogenic-to-VUS thresholds of 0.24 and 0.08, respectively.

†††Maximum (or minimum in the case of FPR, FNR, and uncertain rate) values highlighted in bold for each of the quality metrics obtained for PirePred (high coverage) and the listed predictors (per column).

‡‡‡Values highlighted in bold when the best performance is obtained by PirePred with intermediate or with low FPR thresholds.

§§§VarSome server (*https://varsome.com*, last accessed November 20, 2021).

¶¶¶Franklin server (*https://franklin.genoox.com*, last accessed November 20, 2021).

FN, false negative; FP, false positive; FPR, false-positive rate; TN, true negative; TP, true positive; VUS, variant of uncertain significance.

Table 4 (*continued*)

Accuracy[§]	MCC[¶]	PPV[‖]	NPV**	Sensitivity[††]	Specificity[‡‡]	FPR,[§§] %	FNR,[¶¶] %	Uncertain rate,[‖‖] %
0.950	0.706	0.947	**1.000**[†††]	**1.000**[†††]	0.565	26.3.	**0.0**[†††]	10.7
0.973[‡‡‡]	0.807	0.971	**1.000**[‡‡‡]	**1.000**[‡‡‡]	0.722	13.2	**0.0**[‡‡‡]	16.5
0.988[‡‡‡]	0.887	**0.987**[‡‡‡]	**1.000**[‡‡‡]	**1.000**[‡‡‡]	0.867	**5.3**[‡‡‡]	**0.0**[‡‡‡]	26.8
0.897	0.580	0.890	**1.000**[†††]	**1.000**[†††]	0.378	60.5	**0.0****	0.4
0.821	0.580	0.980	0.486	0.801	**0.921**[†††]	**7.9****	19.9	**0.0****
0.847	0.451	0.908	0.543	0.908	0.543	42.1	8.6	6.7
0.968**	**0.912****	**0.975****	0.946	0.983	**0.921**[†††]	**7.9**[†††]	1.1	30.8
0.847	0.572	0.830	**1.000**[†††]	**1.000**[†††]	0.395	60.5	**0.0**[†††]	33.0
0.840	0.538	0.844	0.818	0.964	0.474	52.6	2.2	33.0
0.812	0.437	0.828	0.714	0.946	0.405	57.9	3.2	33.5
0.807	0.386	0.806	0.818	0.982	0.257	68.4	1.1	35.3
0.796	0.336	0.789	**1.000**[†††]	**1.000**[†††]	0.143	78.9	**0.0**[†††]	34.4
0.849	0.623	0.913	0.690	0.880	0.763	23.7	7.0	34.8
0.801	0.498	0.876	0.610	0.852	0.658	34.2	8.6	34.8
0.857	0.606	0.890	0.742	0.924	0.657	31.6	4.3	37.5
0.821	0.569	0.917	0.614	0.838	0.771	21.1	9.1	37.5
0.830	0.552	0.881	0.676	0.890	0.657	31.6	5.9	39.7
0.917	0.665	0.934	0.800	0.971	0.632	18.4	1.6	46.0
0.933	0.657	0.929	**1.000**[†††]	**1.000**[†††]	0.500	36.8	**0.0**[†††]	6.2
0.912	0.500	0.909	**1.000**[†††]	**1.000**[†††]	0.286	26.3	**0.0**[†††]	49.1

Confirmation of PirePred Performance on Newly Reported Variants

One of the major problems that variants classifiers have to face is that of overfitting. When a classifier is overfitted to the training set, its performance on other variants may not be similarly good. Being aware of the problem, great care has been taken during the training of PirePred to avoid or at least to reduce overfitting. As described in the section reporting PirePred training, the metrics given in Table 3 are

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:38 am ■ EO: JMDI-D-21-

essentially independent of the size of the training and test sets and the precise composition of the training set, and, more importantly, they hold for the group of variants (Figure 2) that were not used in any of the training steps. In principle, all the above should suffice to indicate that PirePred is not overfitted to its training data set. However, surprises are not infrequent in related literature, and sometimes the performance of classifiers on newly reported variants are poorer than expected from their published metrics.[42]

In this respect, the most stringent test a classifier should pass is that of its performance on variants whose pathogenic or benign character was not known at the time the classifier was trained. The training/test set described in Table 2 and used to achieve the statistics described in Table 3 was confirmed by June 1, 2021. Since then, by July 19, 2021, 224 new variants with the 1+ star and not uncertain or conflicting status have been reported in ClinVar. They comprise 150 missense and 74 nonsense variants, which have been used to assess a real performance of PirePred and compared to that of the 15 selected predictors on this data set. The corresponding quality metrics are given in Table 4. These metrics are indeed similar to those given in Table 3. With standard thresholds (high coverage mode), PirePred is the best classifier in NPV, sensitivity and FNR, second in accuracy, third in PPV, and fourth in coverage. Although the specificity is still low for the reasons described in the performance section, it is nevertheless the fifth best at FPR (here at 26%; clearly better than the mean of 36%). With lowFPR thresholds, PirePred is the best in accuracy, PPV, NPV, sensitivity, FPR, and FNR, second in MCC, and third in specificity, while still keeping a coverage of 73% (the fourth higher). On the basis of these results, PirePred is proposed as the accurate, adaptable consensus classifier of choice for SNVs (missense and nonsense) occurring in 58 genes related to most of the conditions investigated in neonatal screening programs.

Comparison with Computational Supporting Evidence from VarSome and Franklin Platforms

In recent years, several bioinformatics suites have been implemented that combine computational predictions with clinical support, segregation, or functional studies to assist variant calling. Two such platforms, Franklin (*https://franklin.genoox.com*, last accessed November 20, 2021) and VarSome,[43] use sets of rules that follow ACMG criteria.[9] The usefulness of those platforms to classify well-characterized variants is expected to be superior to that of purely computational tools. However, it is interesting to compare how their computational modules perform because, for novel variants, the computational prediction is the main supporting evidence for a provisional classification of the variant. Franklin and VarSome were compared with PirePred on the basis of their computational predictions for the 224 new variants with the 1+ star and not uncertain or conflicting status that were reported in ClinVar after the PirePred classifier was finalized (Table 4). Those Franklin and VarSome *in silico* predictions are mainly based on data present in a dbNSFP release previous to the reporting of those variants in ClinVar and are, therefore, as blind as those of PirePred. For each variant, the Franklin and VarSome predictions are compared in Supplemental Table S6 with those of PirePred and with the ClinVar classification. The pairwise prediction correlation between the pathogenicity verdicts issued by the several classifiers is given in Supplemental Table S7. PirePred (high coverage) and PirePred (intermediate) show the best correlations with ClinVar (0.727 and 0.745, respectively), followed by the *in silico* predictor of VarSome (0.697), PirePred (low FPR; 0.689), and the *in silico* predictor of Franklin (0.267). The poor correlation of Franklin (*in silico*) seems related to the fact that it does not issue predictions for nonsense SNVs, which appear to be introduced in its full evaluation scheme at the level of the ACMG criterion PVS1.[9] However, when VarSome or Franklin provide their verdicts using additional information (their global predictions use updated ClinVar data), their correlations with ClinVar are better (0.933 for Franklin and 0.911 for VarSome). The overall quality metrics of the Franklin and VarSome *in silico* predictions for the 224 variants are given in Table 4, compared with PirePred and the 15 individual predictors. PirePred outperforms Franklin *in silico* and VarSome *in silico* in accuracy, specificity, MCC, and PPV and equals them in sensitivity, NPV, and FNR. Its maximal coverage (0.893) is below that of VarSome (0.938) and above that of Franklin (0.509). It seems clear than using an integrated approach to classify variants is the best choice when sufficient variant characterization is available and that, for new variants, PirePred outperforms the *in silico* predictions in those two suites. The fine PirePred quality metrics observed for the 224 new variants with a one star ClinVar review status (Table 4) are essentially retained (Supplemental Table S8) when they are determined on the 26-variant subset classified with two or three stars.

Conclusions

PirePred is a unique and user-friendly genetic interpretation tool for protein sequence variants (missense, nonsense, and frameshift) associated with 58 genes relevant to newborn screening. Combining predictions from 15 predictors, PirePred provides accurate consensus classification (benign, VUS, or pathogenic) for any possible variant of these types that may arise in those 58 genes. The PirePred classification is computational supporting evidence (as defined by the ACMG guidelines), and it appears to be more accurate than equivalent supporting evidence from some comprehensive bioinformatics interpretation platforms. In addition, PirePred sets a focus on SNVs reported in ClinVar for additional structural evaluation. The affected protein residue is

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:38 am ■ EO: JMDI-D-21-

displayed in its structural protein context, which is analyzed to provide hints for a molecular interpretation of the predicted phenotype. The PirePred server can help both researchers and clinicians get a quick and reliable interpretation of SNVs in genes associated with most conditions currently investigated in neonatal screening.

Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.jmoldx.2022.01.005*.

References

1. Loeber JG, Platis D, Zetterström RH, Almashanu S, Boemer F, Bonham JR, et al: Neonatal screening in Europe revisited: an ISNS perspective on the current state and developments since 2010. Int J Neonatal Screen 2021, 7:15
2. Therrell BL, Padilla CD, Loeber JG, Kneisser I, Saadallah A, Borrajo GJC, Adams J: Current status of newborn screening worldwide: 2015. Semin Perinatol 2015, 39:171–187
3. Berry SA: Newborn screening. Clin Perinatol 2015, 42:441–453
4. Wojcik MH, Zhang T, Ceyhan-Birsoy O, Genetti CA, Lebo MS, Yu TW, Parad RB, Holm IA, Rehm HL, Beggs AH, Green RC, Agrawal PB, BabySeq Project Team: Discordant results between conventional newborn screening and genomic sequencing in the BabySeq Project. Genet Med 2021, 23:1372–1375
5. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR: ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018, 46:D1062–D1067
6. Liu X, Jian X, Boerwinkle E: dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Hum Mutat 2011, 32:894–899
7. Álvarez de la Campa E, Padilla N, de la Cruz X: Development of pathogenicity predictors specific for variants that do not comply with clinical guidelines for the use of computational evidence. BMC Genomics 2017, 18:569
8. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X: Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet 2015, 24:2125–2137
9. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL: Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015, 17:405–424
10. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM: Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 2012, 337:64–69
11. Navío D, Rosell M, Aguirre J, de la Cruz X, Fernández-Recio J: Structural and computational characterization of disease-related mutations involved in protein-protein interfaces. Int J Mol Sci 2019, 20:1583
12. Aguzzi A, O'Connor T: Protein aggregation diseases: pathogenicity and therapeutic perspectives. Nat Rev Drug Discov 2010, 9:237–248
13. Gao M, Zhou H, Skolnick J: Insights into disease-associated mutations in the human proteome through protein structural analysis. Structure 2015, 23:1362–1369
14. Galano-Frutos JJ, García-Cebollada H, Sancho J: Molecular dynamics simulations for genetic interpretation in protein coding regions: where we are, where to go and when. Brief Bioinform 2021, 22:3–19
15. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D: Highly accurate protein structure prediction for the human proteome. Nature 2021, 596:590–596
16. Watson MS, Mann MY, Lloyd-Puryear MA, Rinaldo P, Howell RR, Cordero J, American College of Medical Genetics Newborn Screening Expert Group: Newborn screening: toward a uniform screening panel and system - executive summary. Pediatrics 2006, 117:S296–S307
17. Herráez A, Hanson RM, Glasser L: Jmol: an open-source java viewer for chemical structures in 3D. Angel Biochem Mol Biol Educ 2009, 86:8
18. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, De Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T: SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018, 46:W296–W303
19. Yang J, Zhang Y: I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res 2015, 43:W174–W181
20. Worth CL, Kreuchwig F, Tiemann JKS, Kreuchwig A, Ritschel M, Kleinau G, Hildebrand PW, Krause G: GPCR-SSFE 2.0 - a fragment-based molecular modeling web tool for Class A G-protein coupled receptors. Nucleic Acids Res 2017, 45:W408–W415
21. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC: MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci 2018, 27:293–315
22. Kabsch W, Sander C: Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983, 22:2577–2637
23. Estrada J, Bernadó P, Blackledge M, Sancho J: ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. BMC Bioinformatics 2009, 10:104
24. Furnham N, Holliday GL, De Beer TAP, Jacobsen JOB, Pearson WR, Thornton JM: The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. Nucleic Acids Res 2014, 42:D485–D489
25. Ng P, Henikoff S: Predicting deleterious amino acid substitutions. Genome Res 2001, 11:863–874
26. Adzhubei I, Schmidt S, Peshkin L, Ramensky V, Gerasimova A, Bork P: A method and server for predicting damaging missense mutations. Nat Methods 2010, 7:248–249
27. Chun S, Fay J: Identification of deleterious mutations within three human genomes. Genome Res 2009, 19:1553–1561
28. Schwarz JM, Cooper DN, Schuelke M, Seelow D: MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 2014, 11:361–362
29. Reva B, Antipin Y, Sander C: Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 2011, 39:e118
30. Choi Y, Sims G, Murphy S, Miller J, Chan A: Predicting the functional effect of amino acid substitutions and indels. PLoS One 2012, 7:e46688
31. Jagadeesh K, Wenger A, Berger M, Guturu H, Stenson P, Cooper D: M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet 2016, 48:1581–1586

FLA 5.6.0 DTD ■ JMDI1187_proof ■ 12 February 2022 ■ 11:38 am ■ EO: JMDI-D-21-

32. Ioannidis N, Rothstein J, Pejaver V, Middha S, McDonnell S, Baheti S: REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet 2016, 99:877−885
33. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P: Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 2009, 25:2744−2750
34. Qi H, Chen C, Zhang H, Long JJ, Chung WK, Guan Y, Shen Y: MVP: predicting pathogenicity of missense variants by deep learning. BioRxiv 2018. doi:10.1101/259390
35. Raimondi T, Tanyalcin I, Ferté J, Gazzo A, Orlando G, Lenaerts T: DEOGEN2: prediction and interactive visualization of single amino acid variant deleteriousness in human proteins. Nucleic Acids Res 2017, 45:W201−W206
36. Alirezaie N, Kernohan KD, Hartley T, Majewski J, Hocking TD: ClinPred: prediction tool to identify disease-relevant nonsynonymous single-nucleotide variants. Am J Hum Genet 2018, 103:474−483
37. Malhis N, Jacobson M, Jones SJM, Gsponer J: LIST-S2: taxonomy based sorting of deleterious missense mutations across species. Nucleic Acids Res 2020, 48:W154−W161
38. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M: CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019, 47:D886−D894
39. Shihab H, Gough J, Cooper D, Stenson P, Barker G, Edwards K: Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat 2013, 34:57−65
40. Feng B-J: PERCH: a unified framework for disease gene prioritization. Hum Mutat 2017, 38:243−251
41. Bentley TG, Catanzaro A, Ganiats TG: Implications of the impact of prevalence on test thresholds and outcomes: lessons from tuberculosis. BMC Res Notes 2012, 5:563
42. Özkan S, Padilla N, Moles-Fernández A, Diez O, Gutiérrez-Enríquez S, de la Cruz X: The computational approach to variant interpretation: principles, results, and applicability. Edited by Lázaro C, Lerner-Ellis J, Spurdle A Clin. DNA Var. Interpret. Theory Pract. ed 1. Academic Press, 2021
43. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, Massouras A: VarSome: the human genomic variant search engine. Bioinformatics 2019, 35:1978−1980

Supplemental Figure S1 Metrics for the selection of threshold values to convert score-based predictors into binary ones. Plots for (**A**) REVEL predictor, (**B**) MutPred, (**C**) MVP and (**D**) CADD. In each panel, plots display the change of the indicated statistical metrics (see **inset** legend) for all the different thresholds tested for the predictor in question on the training set excluding the external validation split. Rules for selecting the optimal threshold are described in the main text.

Supplemental Figure S2 Correlation matrix for all predictors. Each cell in the matrix shows the Pearson correlation coefficient between the scores of the predictors of its row and column, colored according to the gradient bar at the **right**. Predictions from the selected predictors for all possible SNVs (as implemented in dbNFSP v4.1a[6]) of the genes covered by PirePred were taken to build the matrix.

Supplemental Figure S3 Fault triangle of the PirePred classifier as a function of the P-to-VUS threshold selected for missense variants. The threshold 0.4 (red triangle) is the value finally chosen for the standard (*High Coverage*) classifying setup among the three implemented in the server. For the sake of comparison, the thresholds used for the *Intermediate* and the *LowFPR* modes are 0.24 and 0.08, respectively, as described in the main text.