

RESEARCH ARTICLE

Open Access



Evaluation of in silico pathogenicity prediction tools for the classification of small in-frame indels

S. Cannon, M. Williams, A. C. Gunning and C. F. Wright*

Abstract

Background The use of in silico pathogenicity predictions as evidence when interpreting genetic variants is widely accepted as part of standard variant classification guidelines. Although numerous algorithms have been developed and evaluated for classifying missense variants, in-frame insertions/deletions (indels) have been much less well studied.

Methods We created a dataset of 3964 small (< 100 bp) indels predicted to result in in-frame amino acid insertions or deletions using data from gnomAD v3.1 (minor allele frequency of 1–5%), ClinVar and the Deciphering Developmental Disorders (DDD) study. We used this dataset to evaluate the performance of nine pathogenicity predictor tools: CADD, CAPIPE, FATHMM-indel, MutPred-Indel, MutationTaster2021, PROVEAN, SIFT-indel, VEST-indel and VVP.

Results Our dataset consisted of 2224 benign/likely benign and 1740 pathogenic/likely pathogenic variants from gnomAD (n = 809), ClinVar (n = 2882) and, DDD (n = 273). We were able to generate scores across all tools for 91% of the variants, with areas under the ROC curve (AUC) of 0.81–0.96 based on the published recommended thresholds. To avoid biases caused by inclusion of our dataset in the tools' training data, we also evaluated just DDD variants not present in either gnomAD or ClinVar (70 pathogenic and 81 benign). Using this subset, the AUC of all tools decreased substantially to 0.64–0.87. Several of the tools performed similarly however, VEST-indel had the highest AUCs of 0.93 (full dataset) and 0.87 (DDD subset).

Conclusions Algorithms designed for predicting the pathogenicity of in-frame indels perform well enough to aid clinical variant classification in a similar manner to missense prediction tools.

Keywords Pathogenicity, In-frame indels, Variant interpretation, Pathogenicity prediction

Background

Next generation DNA sequencing (NGS) is transforming healthcare by facilitating novel understanding of disease and uptake of precision medicine initiatives [1,

2]. Genetic variation is widespread, with every individual carrying > 200 very rare coding variants [3], so molecular diagnosis of monogenic disorders requires expert clinical and scientific interpretation of variants detected by NGS. Classifying the pathogenicity of candidate causal variants is essential for robust diagnosis and management of genetic disorders. To this end, numerous in silico pathogenicity prediction algorithms have been developed and are widely used as evidence when interpreting genetic variants. The use of pathogenicity predictors is supported by current guidelines

*Correspondence:

C. F. Wright

Caroline.Wright@exeter.ac.uk

Department of Clinical and Biomedical Sciences (Medical School), Faculty of Health and Life Sciences, University of Exeter, Research, Innovation, Learning and Development Building, Royal Devon and Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

from the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) [4] and, more recently, the UK Association for Clinical Genomic Science (ACGS) [5], through the PP3/BP4 criteria. Pathogenicity prediction algorithms incorporate various lines of evidence to predict the impact of variation on protein function, including evolutionary inter-species sequence conservation [6], physico-chemical distances between amino acids [7] as well as integrated tests for identifying regulatory features [8]. Some also incorporate human variation and disease data [9] by querying gene or variant level understanding from open source [10] or proprietary [11] databases. These aggregated data are used to generate statistical prediction models, such as supervised machine learning classifiers [12], which produce a score used to assign pathogenicity status to a given variant.

Most pathogenicity predictors have been developed to predict the effect of missense substitutions [13, 14], which are primarily caused by single nucleotide variants (SNVs) in the protein-coding regions of the genome. However, small insertions and deletions (indels) account for between 13 and 18% of all variation in the human genome [15, 16], both within and outside protein-coding regions, and have been linked to numerous rare heritable diseases [17] as well as cancerous somatic mutations [18]. Approximately 40% of coding indels are in-frame [19], defined as a nucleotide length (n), wholly divisible by three, which results in the removal or addition of $n/3$ amino acids. Unlike frame-shifting indels, which are generally assumed to cause loss-of-function, the insertion or deletion of a small number of amino acids is likely to have a similarly deleterious effect on a protein as substitution of one amino acid for another. Indeed, missense variants and in-frame indels are frequently grouped together as “protein altering variants” and overall assumed to have “moderate” impact [20].

Numerous small in-frame indels have been shown to cause monogenic disease, most famously (p.Phe508del) in *CFTR* [21]. However, in general, the classification of in-frame indels has been much less well studied than missense and loss-of-function variants. To this end, we created a novel dataset of previously classified in-frame indels, constructed from three databases, two open source (gnomAD [22] and, ClinVar [10]) and one managed access (Diagnosing Developmental Disorders study (DDD) [23]), and use this dataset to evaluate the performance of nine in silico prediction algorithms. We show that although the accuracy of pathogenicity classifiers varies across tools, overall the performance is comparable to those designed for missense variants.

Methods

Benchmark dataset generation

Variants were retrieved from gnomAD (v3.1.1) [22], ClinVar [10], and the DDD study deposited in DECIPHER [23], all accessed 18 March 2021, before filtering for suitability for this study (Fig. 1). Briefly, variants in genome build GRCh38 were included if they were evenly divisible by 3 and < 100 base-pairs in length. Assumed benign variants with a minor allele frequency 1–5% were retained from the gnomAD population database, while variants classified as likely pathogenic (LP), pathogenic (P), benign (B) or likely benign (LB) were retained from the two clinical datasets. Identical variants in more than one database were retained from only one using the preferential order of DDD, ClinVar then gnomAD, and variants with conflicting annotations between databases were removed. The resulting variants were annotated by the Ensembl Variant Effect Predictor (VEP) [20]. Those annotated as “inframe_insertion” or inframe_deletion” with biotype “protein coding” and a single protein consequence per variant were selected ($n = 3964$; Table 1, Additional file 2: Table S1). A subset of potentially novel variants from the DDD study, which were not present in either ClinVar or gnomAD ($n = 151$), was used as an additional test set because these variants are unlikely to have been previously encountered by the tools.

Tool selection and benchmarking

For inclusion in this study, pathogenicity prediction tools were identified from the literature and had to be either (i) accessible through a webserver or (ii) downloadable for use on a local server. We evaluated the performance of nine pathogenicity prediction tools, using their default classification threshold criteria: CADD [24], CAPICE [25], PROVEAN [26], FATHMM-indel [27], MutationTaster2021 [28], MutPred-Indel [29], SIFT-indel [12], VEST-indel [30], and VVP [31] (Table 2). Standard performance metrics (sensitivity, specificity, positive and negative predictive values) and the Matthews Correlation Coefficient (MCC) [32] were calculated for all tools. Receiver-operator characteristics (ROC) and the area under the ROC curve (AUC) were determined for all tools apart from SIFT-indel and MutationTaster2021 which produced binary classifications. All above analyses were repeated using the DDD-only subset. We also considered the effect of protein length on the ability of software to classify variants by grouping variants into four bins of amino acid length (1, 2–4, 5–10 and 11+).

Results

Benchmark datasets contained a good balance of pathogenic and benign insertions and deletions

Our dataset consisted of 3964 small in-frame indels from 1820 genes, including 1246 insertions and 2718

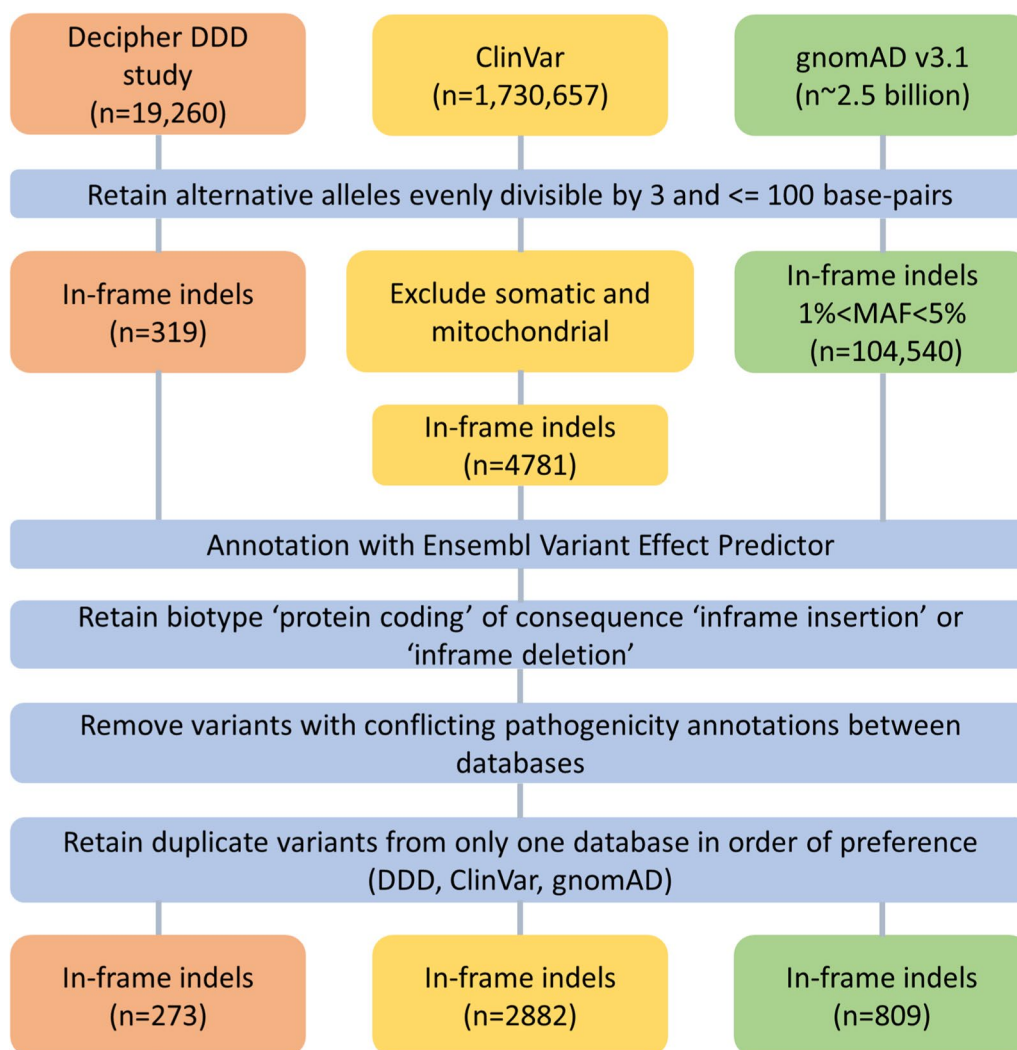


Fig. 1 Flowchart of dataset construction. We included in-frame indels from ClinVar, gnomAD and the DDD study (deposited in DECIPHER). *SNV* single nucleotide variant, *MAF* Minor allele frequency

Table 1 Number of variants from each database included in our benchmark dataset

Database	Benign	Likely benign	Likely pathogenic	Pathogenic	Total
ClinVar	424	803	796	859	2882
DDD	15	173	49	36	273
gnomAD	809	0	0	0	809
Total	1248	976	845	895	3964

deletions from gnomAD (n = 809), ClinVar (n = 2882) and DDD (n = 273) (Fig. 1). Of these, 2224 were B/LB and 1740 were P/LP ranging in size from 1–48 amino acids for insertions and 1–66 amino acids for deletions (Fig. 2). The longest pathogenic and benign

deletions were 32 and 66 residues, and the longest pathogenic and benign insertions were 28 and 48 residues, respectively. Variants were distributed across 1820 protein-coding genes (mean = 2.18, SD = 3.64, min = 1, max = 66). The proportion of benign/pathogenic variants varied across genes linked with monogenic disease. Some genes had almost exclusively benign variants in our dataset, e.g. *DSPP* [MIM:125485] (B/LB = 25, P/LP = 0) and *ARID1B* [MIM:614556] (B/LB = 62, P/LP = 1); some had almost exclusively pathogenic variants, e.g. *LDLR* [MIM:606945] (B/LB = 0, P/LP = 63), *FBNI* [MIM: 134797] (B/LB = 0, P/LP = 20) and *NFI* [MIM:613113] (B/LB = 0, P/LP = 17); and some had similar numbers of pathogenic and benign variants, e.g. *CREBBP* [MIM:600140] (B/LB = 10, P/

Table 2 Pathogenicity predictors and default or recommended classification thresholds used in this study

Software	Input data format	Command line interface	Genome Build (GRCh)	Benign threshold	Pathogenic threshold
CADD	VCF	+	37/38	< 20	> = 20
CAPICE	VEP annotated TSV	+	37/38	< 0.02	> = 0.02
FATHMM-indel	VCF	-	37	> = 0.5	< 0.5
MutPred-Indel	Peptide (FASTA)	+	37/38	< 0.672	> = 0.672
MutationTaster 2021	VCF/genomic position/transcript specific	-	37	Benign	Deleterious
PROVEAN	VCF subset (CSV)	+	37	> 2.5	< = 2.5
SIFT-indel	VCF subset (CSV)	-	37/38	Neutral	Damaging
VEST-indel	VCF	+	37/38	< 0.5	> = 0.5
VVP	VCF	+	37/38	< 57	> = 57

VCF Variant call format, VEP Variant effect predictor, TSV tab separated values, CSV comma separated values

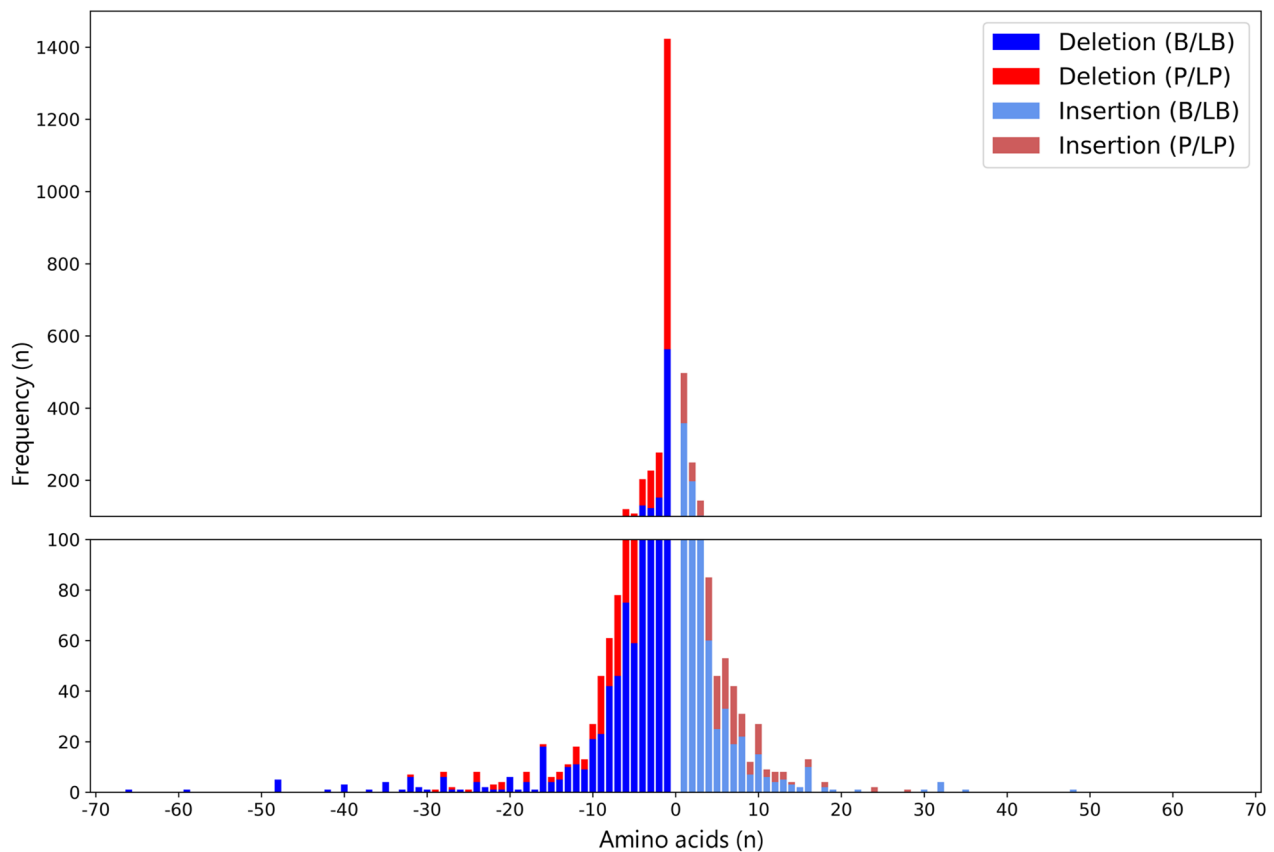


Fig. 2 Histogram showing length and pathogenicity classification of our benchmark dataset. In-frame indels ($n = 3964$, deletions = 2718, insertions = 1246) were taken from gnomAD, ClinVar and the DDD study. *B/LB* benign/likely benign (blue), *P/LP* pathogenic/likely pathogenic (red)

LP = 10) and *ARX* [MIM:300382] (B/LB = 10, P/LP = 9). The DDD-only dataset consisted of 151 novel in-frame indels from 119 genes, including 81 B/LB and 70 P/LP

variants ranging in size from 1–11 to 1–13 amino acids for insertions and deletions, respectively, all in genes

where rare deleterious variants are known to cause developmental disorders.

Performance was generally high across all tools using our full dataset, but some tools performed substantially worse using a smaller, novel variant dataset

For the full dataset, 3615–3963 (91–99%) of variants were classified by each tool and 3522 (89%) were classified by every tool. Of the latter, 556 (15.8%) were universally categorised correctly by all nine tools as pathogenic ($n = 179$, 5.1%) or benign ($n = 377$, 10.7%). Sensitivity and specificity ranged from 0.30–0.99 to 0.61–0.97, respectively (Table 3, Fig. 3A). For the smaller DDD-only novel dataset, 143–151 (95–100%) variants were classified by each tool and 141 (93%) were classified by every tool. Of these, 14 (9.9%) were universally categorised correctly by all nine tools as pathogenic ($n = 8$, 5.7%) or benign ($n = 6$, 4.2%). Sensitivity ranged from 0.24 to 0.97 and specificity range from 0.14 to 0.8 (Table 3, Fig. 3B).

Sensitivity decreased for most tools between the full dataset and the DDD-only subset, apart from FATHMM-indel that remained the same (0.94), as well as CADD and SIFT-indel which increased from 0.49 to 0.64 and 0.82 to 0.86, respectively; MutationTaster2021 showed the

largest decrease in sensitivity from 0.98 to 0.72. Specificity decreased for all tools between the two datasets with CADD and SIFT-indel decreasing the least from 0.92 to 0.80 and 0.61 to 0.51, respectively; VVP decreased the most from 0.67 to 0.14. These observations were recapitulated in the MCC metric, where VVP and MutationTaster2021 decreased the most by 0.48 and 0.44, whereas CADD and SIFT-Indel decreased the least by 0.02 and 0.04, respectively. PROVEAN, VEST-indel and FATHMM-indel showed similar performance in the DDD-only subset with MCC metrics of 0.51, 0.53 and 0.51, respectively.

Tool performance was generally independent of indel length

We investigated the tools' performance for insertions and deletions separately, and whether their performance was influenced by indel length (grouped into bins of 1, 2–4, 5–10 and 11 + amino acids inserted/deleted). We observed very little difference in performance between groups of variants (Additional file 1: Figs. S2 and S3) shows this in more detail), despite an increase in the proportion of benign variants with increasing indel length.

Table 3 Performance metrics for all indel pathogenicity prediction tools tested

Tool	TP	FP	TN	FN	Total (%)	Sens	Spec	LR+	LR–	PPV	NPV	AUC	MCC
<i>All variants (1740 pathogenic; 2224 benign)</i>													
CADD	852	176	2047	886	3961 (99.9)	0.49	0.92	6.19 ^S	0.55	0.83	0.70	0.86	0.47
CAPICE	1611	715	1500	129	3955 (99.7)	0.93	0.68	2.87 ⁺	0.11 ⁺	0.69	0.92	0.91	0.61
FATHMM-Indel	1626	572	1622	111	3931 (99.2)	0.94	0.74	3.59 ⁺	0.09 ⁺	0.74	0.94	0.91	0.68
MutPred-Indel	516	72	2116	1211	3915 (98.8)	0.30	0.97	9.08 ^S	0.73	0.88	0.64	0.81	0.37
MutationTaster2021	1675	99	1976	36	3786 (95.5)	0.98	0.95	20.52 ^S	0.02 ^S	0.94	0.98	–	0.93
PROVEAN	1482	581	1470	82	3615 (91.2)	0.95	0.72	3.35 ⁺	0.07 ^S	0.72	0.95	0.93	0.66
SIFT-Indel	1406	827	1287	313	3833 (96.7)	0.82	0.61	2.09	0.30 ⁺	0.63	0.80	–	0.43
VEST-indel	1560	385	1758	179	3882 (98.0)	0.90	0.82	4.99 ⁺	0.13 ⁺	0.80	0.91	0.93	0.71
VVP	1720	728	1481	20	3949 (99.6)	0.99	0.67	3.00 ⁺	0.02 ^S	0.70	0.99	0.87	0.67
<i>DDD subset (70 pathogenic; 81 benign)</i>													
CADD	45	16	65	25	151 (100)	0.64	0.80	3.25 ⁺	0.45	0.74	0.72	0.78	0.45
CAPICE	64	44	37	6	151 (100)	0.91	0.46	1.68	0.19 ⁺	0.59	0.86	0.82	0.41
FATHMM-indel	66	38	43	4	151 (100)	0.94	0.53	2.01	0.11 ⁺	0.63	0.91	0.74	0.51
MutPred-Indel	17	16	64	53	150 (99.3)	0.24	0.80	1.21	0.95	0.52	0.55	0.64	0.05
MutationTaster2021	50	19	61	19	149 (98.7)	0.72	0.76	3.05 ⁺	0.36 ⁺	0.72	0.76	–	0.49
PROVEAN	60	32	45	6	143 (94.7)	0.91	0.58	2.19	0.16 ⁺	0.65	0.88	0.86	0.51
SIFT-indel	60	39	41	10	150 (99.3)	0.86	0.51	1.76	0.28 ⁺	0.61	0.80	–	0.39
VEST-indel	62	29	51	8	150 (99.3)	0.89	0.64	2.44 ⁺	0.18 ⁺	0.68	0.86	0.87	0.53
VVP	68	70	11	2	151 (100)	0.97	0.14	1.12	0.21 ⁺	0.49	0.85	0.64	0.19

Findings from the entire dataset are included in the top table, and just the novel (DDD-only) subset in the bottom table. Relative strength of likelihood ratios for application of 'strong' or 'moderate' evidence under the ACMG/ACGS variant classification criteria [33] are denoted as: ^SHigh relative strength, ⁺Medium relative strength. Additional file 1: Fig. S1 shows ROC-AUC curves

TP True positive, FP False positive, TN True negative, FN False negative, LR+ positive likelihood ratio, LR– negative likelihood ratio, PPV Positive predictive value, NPV negative predictive value, AUC Area under the curve, MCC Matthews correlation coefficient

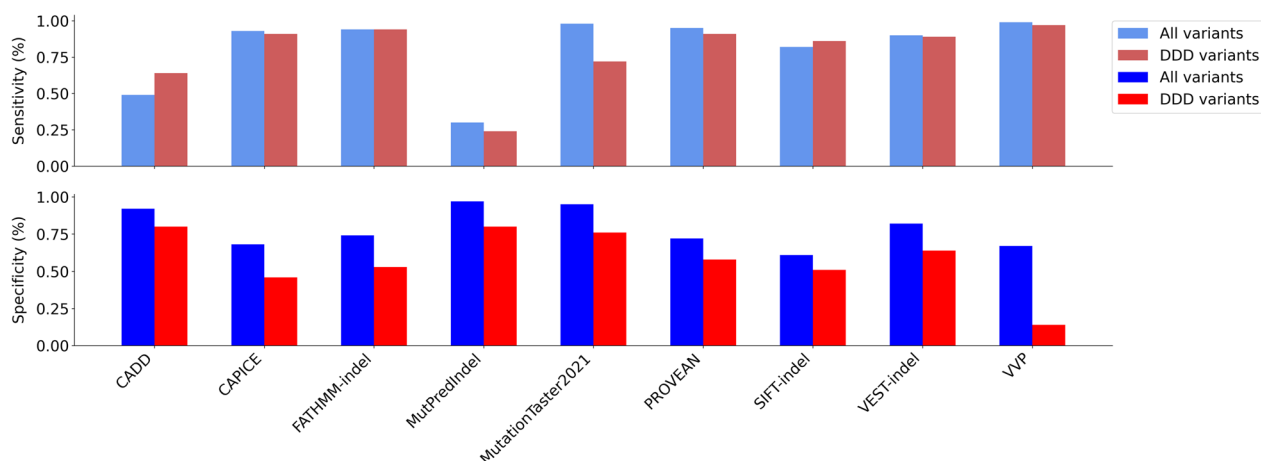


Fig. 3 Performance of pathogenicity prediction tools for in-frame indels. Sensitivity (top) and specificity (bottom) of nine pathogenicity prediction tools based on classification of 3964 in-frame-indels from ClinVar, DDD and gnomAD databases (blue), as well as a DDD-only subset of 151 variants (red)

Discussion

We tested the performance of nine pathogenicity prediction tools on a dataset of 3964 in-frame indels and a smaller subset of 151 novel, clinically classified indels that are not readily accessible from public databases. We show that the performance of these tools is generally good across a range of indel lengths, with AUCs of 0.81–0.93. As expected, most tools performed less well in the smaller novel subset, with AUCs of 0.64–0.87, which likely reflects the use of publicly accessible datasets in the tools’ classification method or training data.

Of the nine tools tested, MutationTaster2021 had the highest sensitivity and specificity when tested on all variants but also showed the greatest decrease in sensitivity when tested on the DDD only dataset. Since gnomAD variants were used as benign training cases and ClinVar and HGMD [11] as pathogenic training cases [28], this may reflect some overfitting [34] and potentially suggests a lower performance for previously unobserved variants. FATHMM-indel, CAPICE, VEST-indel and PROVEAN performed comparably well, although PROVEAN and VEST-indel classified fewer variants than CAPICE and FATHMM-indel. It should be noted that some tools (e.g. CADD, CAPICE, MutationTaster2021, PROVEAN, VVP) were not designed specifically for use with in-frame indels and were trained primarily on SNVs, whilst other tools (e.g. VEST-indel, FATHMM-indel, MutPred-Indel, SIFT-indel) were optimised particularly for the classification of indels. We have previously demonstrated that standard pathogenicity predictors such as SIFT and Polyphen-2 classified missense variants with AUCs of between 0.85–0.87 for a publicly accessible “open”

dataset, and 0.70–0.72 for a restricted access “clinical” dataset [34], which is a comparable performance to the indel pathogenicity predictors tested here. However, the newer meta-predictors Revel [35] and ClinPred [36] produced AUCs of 0.97–0.99 and 0.82–0.81 for open and clinical datasets of missense variants [34], respectively, outperforming all the indel pathogenicity prediction tools tested here. Nonetheless, similar to many missense pathogenicity predictors, the likelihood ratios calculated for in-frame indel predictors using our dataset (Table 3) support their use at either ‘supporting’ or ‘moderate’ towards the PP3 and BP4 criteria of the ACMG/ACGS recommendations, although none of the tools reach the moderate threshold in the DDD subset [33, 37].

We found that the pathogenicity predictor tools varied substantially in input requirements and their ease of use. For example, seven of the tools tested require variants to be uploaded in VCF format as input and five of these also offer a downloadable command line interface (Table 2). Tools with these two features are typically well suited for integration into analysis pipelines; however, ease of installation, additional required software and metadata dependencies varied. For example, MutPred-Indel contained all the necessary metadata to make pathogenicity predictions, but required variants to be input in FASTA format, which is not routinely used in a clinical genetic testing setting, as well as installation of a specific version of MATLAB. Similarly, PROVEAN offered a command line option but also required local installations of NCBI-Blast, and the NCBI nr protein database. The requirement for advanced bioinformatics skills to operate a tool will

adversely affect its utility, particularly for routine diagnostics use. In contrast, CAPICE used an Ensembl-VEP annotated TSV file as input, and we found it the easiest to install and quickest to use.

Like other comparable studies, we were limited by several factors. Firstly, the veracity of the variant classifications taken from ClinVar, the DDD study and gnomAD is uncertain, and our benchmark dataset may include some erroneous variant classifications. We tried to minimise this issue by incorporating data from three different databases and by using minor allele frequency thresholds for benign variants. However, the low number of variants in the DDD subset ($n=151$) limits the comparison of tool performance metrics and a larger dataset would provide a more accurate assessment. Secondly, unlike missense variants caused by SNVs, in-frame indels are comparatively rare and are harder to detect robustly using NGS, and thus our dataset is relatively small. Evaluation of the performance of the tools versus indel length was further limited by the inverse correlation between frequency and variant length in the dataset, which limits the interpretability of tool performance for larger indels. Although large (>100 base-pair) in-frame indels exist, and may be either benign or pathogenic, these are difficult to detect using short-read NGS technologies, so were largely absent from the databases used here and excluded from our dataset. Finally, not all variants in our dataset were in genes linked with monogenic disease, particularly those from gnomAD, which potentially introduces a bias for tools that use gene-level data for classification. However, around 75% of genes present in our dataset contained variants from at least two of the databases, and a sensitivity analysis using only these variants produced similar results (data not shown).

Conclusions

We have shown that numerous *in silico* pathogenicity prediction tools perform well for in-frame indels using a benchmark dataset. We therefore suggest that genomic diagnostic laboratories should consider incorporating these tools—in the same manner as missense prediction tools—to aid variant classification. Our findings are consistent with previous studies [25, 27, 30] and, to the best of our knowledge, represent the largest independent assessment to date of pathogenicity predictors for in-frame indels.

Abbreviations

ACGS	UK Association for Clinical Genomic Science
ACMG	American College of Medical Genetics and Genomics
AMP	Association for Molecular Pathology
AUC	Area under the curve
B	Benign
CSV	Comma separated values

DDD	Deciphering developmental disorders
FP	False positive
FN	False negative
LB	Likely benign
LP	Likely pathogenic
LR+	Positive likelihood ratio
LR−	Negative likelihood ratio
MCC	Matthews correlation coefficient
MIM	Mendelian inheritance in man
NGS	Next generation sequencing
NPV	Negative predictive value
P	Pathogenic
PPV	Positive predictive value
ROC	Receiver operator curve
SD	Standard deviation
SNV	Single nucleotide variant
TP	True positive
TN	True negative
TSV	Tab separated values
VCF	Variant call format
VEP	Variant effect predictor

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-023-01454-6>.

Additional file 1: Fig. S1. ROC curves and AUCs using (A) 3964 in-frame indels from gnomAD, ClinVar and the DDD study (B) 151 in-frame indels only observed in the DDD study. Only tools producing continuous data as output are plotted ($n=7$). **Fig. S2.** Performance of nine pathogenicity prediction tools versus increasing variant length in a dataset of 3964 in-frame indels aggregated from gnomAD, ClinVar and the DDD study. **Fig. S3.** Sensitivity and specificity of 9 pathogenicity prediction algorithms with increasing variant length in a dataset of in-frame (A) deletions ($n=2718$) and (B) insertions ($n=1246$).

Additional file 2: Table S1. Inframe-indels and pathogenicity predictions from tools used in this study.

Acknowledgements

We wish to thank all the patients and family members whose data were used in the study.

Author contributions

CFW conceived and designed the study and provided the datasets. MW performed an initial analysis and write-up using a subset of tools, with assistance from ACG. SC performed the full data analysis and wrote the first draft of the manuscript. SC and CFW finalised the manuscript, and all authors approved it.

Funding

This work was supported by the Medical Research Council [MR/T00200X/1]. The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [HICF-1009-003], a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute [WT098051]. The views expressed in this publication are those of the author(s) and not necessarily those of the Wellcome Trust or the Department of Health.

Availability of data and materials

All variants in the final dataset are included in Additional file 2. The publicly available variant datasets are available from gnomAD [22] or ClinVar [10]. Genomic datasets from the DDD Study are available under managed access for research into developmental disorders via the European Genome-phenome Archive (EGAS00001000775). Individual pathogenic/likely pathogenic variants are openly accessible with phenotypes via DECIPHER [38].

Declarations

Ethics approval and consent to participate

The DDD Study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC); informed consent to participate in the DDD study was obtained from all of the participants.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 5 September 2022 Accepted: 9 February 2023

Published online: 28 February 2023

References

- Liu Z, Zhu L, Roberts R, Tong W. Toward clinical implementation of next-generation sequencing-based genetic testing in rare diseases: Where are we? *Trends Genet.* 2019;35(11):852–67.
- Kanzi AM, San JE, Chimumkangara B, Wilkinson E, Fish M, Ramsuran V, et al. Next generation sequencing and bioinformatics analysis of family genetic inheritance. *Front Genet.* 2020;11:544162.
- Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, et al. Variant interpretation using population databases: Lessons from gnomAD. *Hum Mutat.* 2021;n/a(n/a).
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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(5):405–23.
- Ellard S, Baple EL, Callaway A, Berry I, Forrester N, Turnbull C, et al. ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020: ACGS; 2020 [Available from: <https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf>].
- Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310–5.
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248–9.
- Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day INM, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics (Oxford, England).* 2015;31(10):1536–43.
- Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc.* 2015;10(12):2004–15.
- Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42(Database issue):D980–5.
- Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, et al. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat.* 2003;21(6):577–81.
- Hu J, Ng PC. SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins. *PLoS ONE.* 2013;8(10):e77940-e.
- Ghosh R, Oak N, Plon SE. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol.* 2017;18(1):225.
- Cubuk C, Garrett A, Choi S, King L, Loveday C, Torr B, et al. Clinical likelihood ratios and balanced accuracy for 44 in silico tools against multiple large-scale functional assays of cancer susceptibility genes. *Genet Med.* 2021;23(11):2096–104.
- Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68–74.
- Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELS) in human genomes. *Hum Mol Genet.* 2010;19(R2):R131–6.
- Weedon MN, Ellard S, Prindle MJ, Caswell R, Lango Allen H, Oram R, et al. An in-frame deletion at the polymerase active site of POLD1 causes a multisystem disorder with lipodystrophy. *Nat Genet.* 2013;45(8):947–50.
- Baeissa HM, Pearl FMG. Identifying the impact of inframe insertions and deletions on protein function in cancer. *J Comput Biol.* 2020;27(5):786–95.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285–91.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol.* 2016;17(1):122.
- Liechti-Gallati S, Bonsall I Fau - Malik N, Malik N Fau - Schneider V, Schneider V Fau - Kraemer LG, Kraemer Lg Fau - Ruedeberg A, Ruedeberg A Fau - Moser H, et al. Genotype/phenotype association in cystic fibrosis: analyses of the delta F508, R553X, and 3905insT mutations. 1992(0031–3998 (Print)).
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434–43.
- Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *The Lancet.* 2015;385(9975):1305–14.
- Rentzsch P, Schubach M, Shendure J, Kircher MA-O. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med.* 2021;13(1):31.
- Li S, van der Velde KJ, de Ridder D, van Dijk ADJ, Soudis D, Zwerwer LR, et al. CAPICE: a computational method for consequence-agnostic pathogenicity interpretation of clinical exome variations. *Genome Med.* 2020;12(1):75.
- Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE.* 2012;7(10):e46688.
- Ferlaino M, Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, et al. An integrative approach to predicting the functional effects of small indels in non-coding regions of the human genome. *BMC Bioinform.* 2017;18(1):442.
- Steinhaus R, Proft S, Schuelke M, Cooper DN, Schwarz Jana M, Seelow D. MutationTaster2021. *Nucleic Acids Res.* 2021;49(W1):W446–51.
- Pagel KA, Antaki D, Lian A, Mort M, Cooper DN, Sebat J, et al. Pathogenicity and functional impact of non-frameshifting insertion/deletion variation in the human genome. *PLoS Comput Biol.* 2019;15(6):e1007112-e.
- Douville C, Masica DL, Stenson PD, Cooper DN, Gyax DM, Kim R, et al. Assessing the pathogenicity of insertion and deletion variants with the variant effect scoring tool (VEST-Indel). *Hum Mutat.* 2016;37(1):28–35.
- Flygare S, Hernandez EJ, Phan L, Moore B, Li M, Fejes A, et al. The VAAST Variant Prioritizer (VVP): ultrafast, easy to use whole genome variant prioritization tool. *BMC Bioinform.* 2018;19(1):57.
- Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics.* 2020;21(1):6.
- Pejaver V, Byrne AB, Feng B-J, Pagel KA, Mooney SD, Karchin R, et al. Evidence-based calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for clinical use of PP3/BP4 criteria. *bioRxiv.* 2022:2022.03.17.484479.
- Gunning AC, Fryer V, Fasham J, Crosby AH, Ellard S, Baple EL, et al. Assessing performance of pathogenicity predictors using clinically relevant variant datasets. *J Med Genet.* 2021;58(8):547–55.
- Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet.* 2016;99(4):877–85.
- Alirezai 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(4):474–83.
- Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, et al. Modeling the ACMG/AMP variant classification

guidelines as a Bayesian classification framework. *Genet Med.* 2018;20(9):1054–60.

38. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am J Hum Genet.* 2009;84(4):524–33.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

