Received: 17 June 2019 | Accepted: 27 June 2019

DOI: 10.1002/humu.23858

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

SPECIAL ARTICLE

VIPdb, a genetic Variant Impact Predictor Database

Zhiqiang Hu[1*] | Changhua Yu[1,2*] | Mabel Furutsuki[1,3] | Gaia Andreoletti[1] | Melissa Ly[1,4] | Roger Hoskins[1] | Aashish N. Adhikari[1] | Steven E. Brenner[1]

[1]Department of Plant and Microbial Biology, University of California, Berkeley, California

[2]Department of Bioengineering, University of California, Berkeley, California

[3]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California

[4]Division of Data Sciences, University of California, Berkeley, California

Correspondence
Steven E. Brenner, Department of Plant and Microbial Biology, 111 Koshland Hall #3102, University of California, Berkeley, CA 94720-3102.
Email: brenner@compbio.berkeley.edu

Present address
Gaia Andreoletti, Bakar Computational Health Sciences Institute, University of California, San Francisco, California and Department of Pediatrics, University of California, San Francisco, California.

Funding information
NIH, Grant/Award Numbers: U41 HG007346, R13 HG006650; TATA Consultancy Services Ltd.

Abstract

Genome sequencing identifies vast number of genetic variants. Predicting these variants' molecular and clinical effects is one of the preeminent challenges in human genetics. Accurate prediction of the impact of genetic variants improves our understanding of how genetic information is conveyed to molecular and cellular functions, and is an essential step towards precision medicine. Over one hundred tools/resources have been developed specifically for this purpose. We summarize these tools as well as their characteristics, in the genetic Variant Impact Predictor Database (VIPdb). This database will help researchers and clinicians explore appropriate tools, and inform the development of improved methods. VIPdb can be browsed and downloaded at https://genomeinterpretation.org/vipdb.

KEYWORDS

genotype-phenotype relationship, SNV phenotype, SV impact, variant impact, variant impact prediction, VIPdb

Genomic data hold the promise of revolutionizing our understanding and treatment of human disease. There is a growing gap between the rapid increase of data generation and the functional interpretation of these data (Lappalainen, Scott, Brandt, & Hall, 2019; Muir et al., 2016). Understanding the effects of genetic variants on phenotype is crucial for several fields, including annotation of large-impact variants in resequencing projects (e.g., Pabinger et al., 2014), interpretation of genome-wide association results (e.g.,Visscher et al., 2017), and detection of disease-related variants in a clinical context (e.g., Richards et al., 2015). However, it is presently impractical to experimentally and clinically study effects of all possible genetic variants. Therefore, it is crucially important to develop computational tools for variant impact prediction. To date, over a hundred of tools and databases have been developed for specific contexts. Tools can be developed for specific types of variants. For example, most tools are developed for predicting impact of single nucleotide variations (SNVs), while a smaller number can address small insertions and deletions (indels), and only a handful of tools are able to predict impact of large structural variations (SVs). Moreover, the impact of a variant can be evaluated at different levels, such as molecular damage (e.g., gene expression (Zhou et al., 2018), protein functionality (Vaser, Adusumalli, Leng, Sikic, & Ng, 2016), RNA splicing [Park, Pan, Zhang, Lin, & Xing, 2018]) and pathogenicity (e.g., visible phenotypes [Landrum et al., 2016]). As these tools' purpose differ, the information or training data they draw upon is also diverse, including but not limited to, pathogenicity and clinical phenotypes, cross-species and within-species sequence conservation, protein sequence and protein structure, RNA splicing, RNA binding protein recognition, miRNA binding activity, transcriptional gene regulation, epigenetic signals and translation efficiency. In addition, some tools are optimized for a specific species, and some are developed for a specific disease or gene. The main

*Zhiqiang Hu and Changhua Yu are joint first authors.

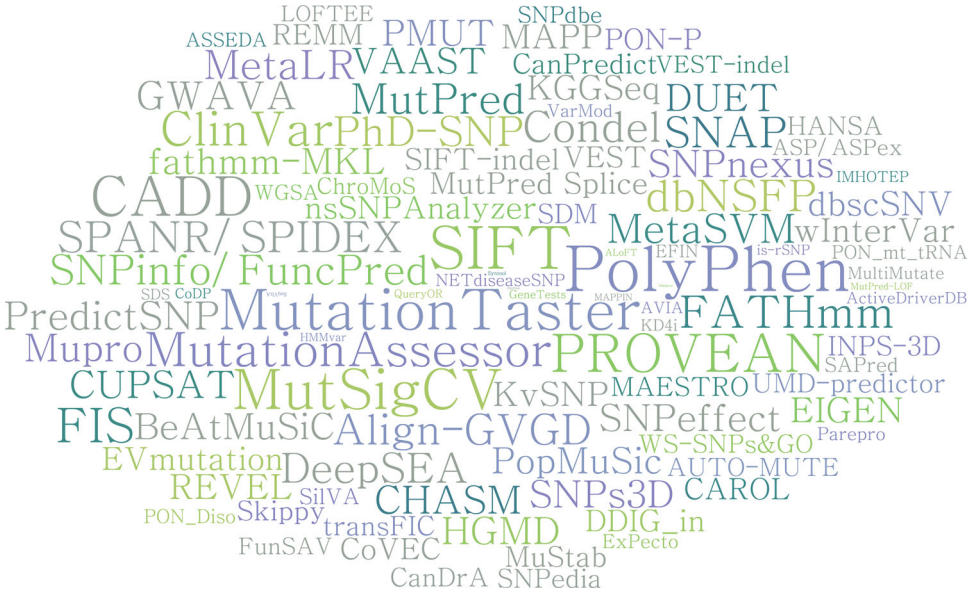

FIGURE 1 Wordle of variant impact predictors. Only tools primarily designed for variant impact prediction were plotted. Character sizes represent the logarithm of recent citations (from Jan 1, 2017 to Jun 1, 2019). For tools with multiple publications, we plotted the sum of all citations. Colors were randomly selected. The Wordle tool (http://www.wordle.net) was used to generate the plot

TABLE 1 Tools and databases used for genomic variant impact predictions

Tool	Primarily for variant impact prediction	Reference
ActiveDriverDB	● DB	Krassowski et al. (2017)
AGGRESCAN		Conchillo-Solé et al. (2007)
AGGRESCAN3D		Zambrano et al. (2015)
Align-GVGD	●	Tavtigian et al. (2006)
ALoFT	●	Balasubramanian et al. (2017)
ANNOVAR		Yang and Wang (2015)
ASP/ASPex	●	Marini, Thomas, and Rine (2010)
ASSEDA	●	Nalla and Rogan (2005)
AUTO-MUTE	●	Masso and Vaisman (2010)
AVIA	●	Vuong et al. (2015)
BeAtMuSiC	●	Dehouck, Kwasigroch, Rooman, and Gilis (2013)
CADD	●	Kircher et al. (2014)
CanDrA	●	Mao et al. (2013)
CanPredict	●	Kaminker, Zhang, Watanabe, and Zhang (2007)
CAROL	●	Lopes et al. (2012)
CHASM	●	Wong et al. (2011)
ChroMoS	●	Barenboim and Manke (2013)
ClinPred	●	Alirezaie, Kernohan, Hartley, Majewski, and Hocking (2018)
ClinVar	● DB	Landrum et al. (2016)
CoDP	●	Terui, Akagi, Kawame, and Yura (2013)
CoMEt		Leiserson, Wu, Vandin, and Raphael (2015)
Condel	●	González-Pérez and López-Bigas (2011)
COSMIC	DB	Tate et al. (2019)
CoVEC	●	Frousios, Iliopoulos, Schlitt, and Simpson (2013)

(Continues)

TABLE 1 (Continued)

Tool	Primarily for variant impact prediction	Reference
CUPSAT	●	Parthiban, Gromiha, Abhinandan, and Schomburg (2007)
DANN		Quang, Chen, and Xie (2015)
DBD-Hunter		Gao and Skolnick (2008)
dbNSFP	● DB	X. Liu et al. (2016)
dbscSNV	●	Jian, Boerwinkle, and Liu (2014)
dbSNP	DB	Sherry et al. (2001)
dbVar	DB	Lappalainen et al. (2013)
DDIG_in	●	Livingstone et al. (2017)
DeepSEA	●	Zhou and Troyanskaya (2015)
DFIRE/DDNA2		B. Xu, Yang, Liang, and Zhou (2009)
DUET	●	Pires, Ascher, and Blundell (2014)
Dmutant		Zhou and Zhou (2002)
EFIN	●	Zeng, Yang, Chung, Lau, and Yang (2014)
EGAD		Pokala and Handel (2005)
EIGEN	●	Ionita-Laza, McCallum, Xu, and Buxbaum (2016)
Eris		Yin, Ding, and Dokholyan (2007)
EVmutation	●	Hopf et al. (2017)
Exome Variant Server (EVS)	DB	Altshuler et al. (2010)
Exomiser		Smedley et al. (2015)
FATHmm	●	Shihab et al. (2014)
fathmm-MKL	●	Shihab et al. (2015)
FIS	●	Reva, Antipin, and Sander (2011)
fitCons		Gulko, Hubisz, Gronau, and Siepel (2015)
FOLD-RATE		Gromiha, Thangakani, and Selvaraj (2006)
FoldAmyloid		Garbuzynskiy, Lobanov, and Galzitskaya (2009)
FOLDEF(core of FOLDX)		Schymkowitz et al. (2005)
FunSAV	●	M. Wang et al. (2012)
GeneTests	DB	Pagon et al. (2002)
GenoCanyon		Lu et al. (2015)
GERP + +		Davydov et al. (2010)
GWAVA	●	Ritchie, Dunham, Zeggini, and Flicek (2014)
HANSA	●	Acharya and Nagarajaram (2012)
HGMD	● DB	Stenson et al. (2017)
HMMvar	●	M. Liu, Watson, and Zhang (2015)
HOPE		Dunlavy, O'Leary, Klimov, and Thirumalai (2005)
Human Splicing Finder		Desmet et al. (2009)
IMHOTEP	●	Knecht et al. (2016)
INPS-3D	●	Savojardo, Fariselli, Martelli, and Casadio (2016)
INSIGHT		Thompson et al. (2014)
is-rSNP	●	Macintyre, Bailey, Haviv, and Kowalczyk (2010)
K-FOLD		Capriotti and Casadio (2007)
KD4i	●	Bermejo-Das-Neves, Nguyen, Poch, and Thompson (2014)
KGGSeq	●	M. J. Li et al. (2017)
KvSNP	●	Stone and Sidow (2005)

(Continues)

TABLE 1 (Continued)

Tool	Primarily for variant impact prediction	Reference
LocTree		Goldberg et al. (2014)
LOFTEE	●	Karczewski et al. (2019)
LUMC LSDB	DB	Fokkema et al. (2011)
LS-SNP/PDB		Ryan, Diekhans, Lien, Liu, and Karchin (2009)
MAESTRO	●	Laimer, Hofer, Fritz, Wegenkittl, and Lackner (2015)
MAPP	●	Stone and Sidow (2005)
MAPPIN	●	Gosalia, Economides, Dewey, and Balasubramanian (2017)
MaxEnt		Yeo and Burge (2004)
MetaLR	●	Dong et al. (2015)
MetaSVM	●	Dong et al. (2015)
MMSplice	●	Jun Cheng et al. (2019)
MultiMutate	●	Deutsch and Krishnamoorthy (2007)
Mupro	●	Cheng, Randall, and Baldi (2006)
MuSiC		Dees et al. (2012)
MuStab	●	Teng, Srivastava, and Wang (2010)
MutationAssessor	●	Reva et al. (2011)
MutationTaster	●	Schwarz, Cooper, Schuelke, and Seelow (2014)
MutPred Splice	●	Mort et al. (2014)
MutPred	●	B. Li et al. (2009)
MutPred-LOF	●	Pagel et al. (2017)
MutSigCV	●	Lawrence et al. (2013)
MuX-48		Kang, Chen, and Xiao (2009)
MuX-S		Kang et al. (2009)
NeEMO		Giollo, Martin, Walsh, Ferrari, and Tosatto (2014)
NETdiseaseSNP	●	Johansen, Izarzugaza, Brunak, Petersen, and Gupta (2013)
nsSNPAnalyzer	●	Bao, Zhou, and Cui (2005)
OMIM	● DB	Amberger and Hamosh (2017)
OncodriveCLUST		Tamborero, Gonzalez-Perez, and Lopez-Bigas (2013)
PAGE		Tartaglia, Cavalli, Pellarin, and Caflisch (2005)
Panther		Mi et al. (2017)
PantherPSEP		Tang and Thomas (2016)
Parepro	●	Tian et al. (2007)
PASTA		Walsh, Seno, Tosatto, and Trovato (2014)
Personal Genome Project	DB	Shringarpure and Bustamante (2015)
PharmGKB	● DB	Thorn, Klein, and Altman (2013)
phastCons		Siepel et al. (2005)
PhD-SNP	●	Capriotti, Calabrese, and Casadio (2006)
Phen-Gen		Javed, Agrawal, and Ng (2014)
phyloP		Pollard, Hubisz, Rosenbloom, and Siepel (2010)
PMUT	●	López-Ferrando, Gazzo, De La Cruz, Orozco, and Gelpí (2017)
PolyPhen	●	Adzhubei et al. (2010)
PON_Diso	●	Ali, Urolagin, Gurarslan, and Vihinen (2014)
PON_mt_tRNA	●	Niroula and Vihinen (2016)
PON-P	●	Niroula, Urolagin, and Vihinen (2015)

(Continues)

TABLE 1 (Continued)

Tool	Primarily for variant impact prediction	Reference
PopMuSic	●	Dehouck, Kwasigroch, Gilis, and Rooman (2011)
PPT-DB		Wishart et al. (2008)
PredictSNP	●	Bendl et al. (2016)
ProA		Fang, Gao, Tai, Middaugh, and Fang (2013)
PROFcon		Punta and Rost (2005)
PROVEAN	●	Choi and Chan (2015)
QueryOR	●	Bertoldi et al. (2017)
REMM	●	Smedley et al. (2016)
REVEL	●	Ioannidis et al. (2016)
SAAPdap/ SAAPred		Hurst et al. (2009)
SAPred	●	Ye et al. (2007)
Scide		Dosztányi, Magyar, Tusnády, and Simon (2003)
Scpred		Kurgan, Cios, and Chen (2008)
SDM	●	Pandurangan, Ochoa-Montaño, Ascher, and Blundell (2017)
SDS	●	Preeprem and Gibson (2014)
SeqVItA	●	Dharanipragada, Seelam, and Parekh (2018)
SIFT	●	Sim et al. (2012)
SIFT-indel	●	Hu and Ng (2013)
SignalP		Petersen, Brunak, Von Heijne, and Nielsen (2011)
SilVA	●	Buske, Manickaraj, Mital, Ray, and Brudno (2013)
SInBaD		Lehmann and Chen (2013)
SiPhy		Garber et al. (2009)
Skippy	●	Woolfe, Mullikin, and Elnitski (2010)
SNAP	●	Hecht, Bromberg, and Rost (2015)
SNPdbe	●	Schaefer, Meier, Rost, and Bromberg (2012)
SNPedia	● DB	Cariaso and Lennon (2012)
SnpEff		Cingolani, Platts et al. (2012)
SNPeffect	●	De Baets et al. (2012)
SNPinfo/FuncPred	●	Z. Xu and Taylor (2009)
SNPnexus	●	Dayem Ullah et al. (2018)
SNPs3D	●	Tian et al. (2007)
SPANR/SPIDEX	●	Xiong et al. (2015)
SPF_Cancer		Capriotti and Altman (2011)
SuRFR		Ryan, Morris, Porteous, Taylor, and Evans (2014)
Syntool	●	Zhang et al. (2017)
TANGO		Rousseau, Schymkowitz, and Serrano (2006)
TransComp		Qin, Pang, and Zhou (2011)
transFIC	●	Gonzalez-Perez, Deu-Pons, and Lopez-Bigas (2012)
UMD-predictor	●	Frédéric et al. (2009)
VAAST	●	Hu et al. (2013)
Variant Tools		Peng (2015)
VariBench		Sasidharan Nair and Vihinen (2013)
VariSNP		Schaafsma and Vihinen (2015)
VarMod	●	Pappalardo and Wass (2014)

(Continues)

TABLE 1 (Continued)

Tool	Primarily for variant impact prediction	Reference
VarWalker		Jia and Zhao (2014)
VEP		McLaren et al. (2016)
VEST	●	Carter, Douville, Stenson, Cooper, and Karchin (2013)
VEST-indel	●	Douville et al. (2016)
Waltz		Maurer-Stroh et al. (2010)
WGSA	●	X. Liu et al. (2016)
wInterVar	●	Q. Li and Wang (2017)
WoLF-PSORT		Horton et al. (2007)
WS-SNPs&GO	●	Capriotti et al. (2013)
Zyggregator		Tartaglia and Vendruscolo (2008)

Note: Tools that are primarily for predicting variant impact are indicated. All items shown are tools, except databases are also flagged with "DB".

challenges in using these tools include their large number and high diversity, lack of consistent output formats, unclear performance characteristics and limited use guidelines. The critical Assessment of Genome Interpretation (CAGI) conducts community experiments for a particular purpose to provide objective assessment of variant impact predictions of phenotype (Hoskins et al., 2017, Andreoletti et al. (this issue)).

Here, we describe the Variant Impact Predictor Database (VIPdb). In this database, we have attempted to collect a comprehensive list of genetic variant impact prediction tools, as well as tools that while not primarily designed for this goal, nonetheless contribute to this purpose, such as tools estimating conservation scores and databases holding population allele frequencies (Figure 1 and Table 1). We also include some particularly relevant databases. This database will help researchers choose appropriate tools, and inform the development of improved methods.

VIPdb is a publication-based database. It originated as a table of resources (Brenner, 2007) and a pilot impact tool list, which was first generated in 2010, manually updated in an ad hoc fashion and available via CAGI (https://genomeinterpretation.org/impact) since

TABLE 2 Description of fields in VIPdb (Table S1), that we collected for each tool or reference

Fields		Description
Ref ID		VIPdb literature reference ID
Tool ID		VIPdb tool or database ID
Name		Tool or database name
Database?		A binary value indicating if it is a database
Primarily for variant impact prediction		A binary value indicating if the tool is primarily designed for predicting variant impact information, or the database provides such information
Variant type	SNVs	A binary value indicating if the tool or database addresses single nucleotide variations
	Indel	A binary value indicating if the tool or database addresses small insertions and deletions
	SVs	A binary value indicating if the tool or database addresses large structural variations
	Nonsynonymous/nonsense	A binary value indicating if the tool addresses nonsynonymous and/or nonsense variants
	Synonymous	A binary value indicating if the tool addresses synonymous variants
	Splicing	A binary value indicating if the tool addresses splicing variants
	Regulatory regions	A binary value indicating if the tool addresses other regulatory variants, e.g., variants in enhancer region
Gene-specific		List of gene(s), if the tool is designed for specific gene(s).

(Continues)

TABLE 2 (Continued)

Fields			Description
In dbNSFP Academic			A binary value indicating if the tool is within dbNSFP academic database
In dbNSFP Commercial			A binary value indicating if the tool is within dbNSFP commercial database
License (Note: Users must evaluate the suitability of each tool and its licensing for their application. The license characterizations are for convenience, and we make no claims regarding their accuracy)		Free for academic use?	A binary value indicating if the tool or database is free for academic use
		Free for commercial use?	A binary value indicating if the tool or database is free for commercial use
		Description	Extracted description of the license
Downloadable precalculated annotation?			A binary value indicating if precalculated annotations/scores are available
Standalone?			A binary value indicating if a stand-alone version is available
Web server?			A binary value indicating if a web server is available
Website accessible as of Jun 1, 2019			A binary value indicating the homepage accessibility
Homepage			Homepage of the tool or database
Source code accessible			Source code link, if available
Reference and Citations	Per paper info	Latest publication?	A binary value indicating if this paper is the latest publication of the tool or database
		Title	Title of the publication
		Doi	DOI of the publication
		PubMed ID	PubMed ID of the publication
		Year	Published year of the publication
		Y<2014	Number of citations before 2014
		Y2014	Number of citations during 2014
		Y2015	Number of citations during 2015
		Y2016	Number of citations during 2016
		Y2017	Number of citations during 2017
		Y2018	Number of citations during 2018
		2019–2019.6	Number of citations between Jan 1, 2019 and Jun 1, 2019
		Y>2015	Number of citations after Jan 1, 2015
		Y>2017	Number of citations after Jan 1, 2017
		Total	Total citation number of the publication
	All tool citation		Total citation number of the tool (citations of multiple reference summed together)
Update date			Last update date for this reference

then. To collect a near-comprehensive list of recent tools, we manually reviewed all publications that cited at least one of the pioneering variant impact prediction tools, including SIFT (Kumar, Henikoff, & Ng, 2009; Sim et al., 2012; Vaser et al., 2016), or PolyPhen (Adzhubei et al., 2010; Adzhubei, Jordan, & Sunyaev, 2013; Ramensky, Bork, & Sunyaev, 2002); as well as annotation tools, such as ANNOVAR (K. Wang, Li, & Hakonarson, 2010; Yang & Wang, 2015) and SnpEff (Cingolani, Patel et al., 2012; Cingolani, Platts et al., 2012). Tools in dbNSFP (X. Liu, Wu, Li, & Boerwinkle, 2016) were also included. In addition, a small number of tools and databases were further collected by direct searches via PubMed database with a various key words list. We excluded LSDBs (Cotton et al., 2008), many of which can be found via LUMC LSDB (https://grenada.lumc.nl/LSDB_list/lsdbs). Next, we collected information about these included tools (Table 2), including their primary purposes (nonsynonymous/nonsense, splicing or other regulatory variants), accepted variant types (SNPs, indels, or SVs), platforms (standalone or online tool), use license (free for academic or for commercial use), whether in dbNSFP database (X. Liu et al., 2016), current website/source code accessibility and citations for publications (from Scopus database).

VIPdb is browsable at the CAGI website (https://genomeinterpretation.org/vipdb) and can be fully downloaded

as a spreadsheet (Table S1; latest version available at above website). We accept submissions of new resources.

ACKNOWLEDGMENT

The CAGI experiment coordination is supported by NIH U41 HG007346 and the CAGI conference by NIH R13 HG006650. This work was supported by TATA Consultancy Services Ltd.

DATA AVAILABILITY STATEMENT

VIPdb is browsable at https://genomeinterpretation.org/vipdb; and the full database can be also downloaded.

ORCID

Zhiqiang Hu http://orcid.org/0000-0001-8854-3410
Gaia Andreoletti http://orcid.org/0000-0002-0452-0009
Aashish N. Adhikari http://orcid.org/0000-0003-4305-9494
Steven E. Brenner http://orcid.org/0000-0001-7559-6185

REFERENCES

Acharya, V., & Nagarajaram, H. A. (2012). Hansa: An automated method for discriminating disease and neutral human nsSNPs. *Human Mutation*, *33*(2), 332–337. https://doi.org/10.1002/humu.21642

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics*, *76*(Suppl. 76), 7.20.1–7.20.41. https://doi.org/10.1002/0471142905.hg0720s76

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. https://doi.org/10.1038/nmeth0410-248

Ali, H., Urolagin, S., Gurarslan, O., & Vihinen, M. (2014). Performance of protein disorder prediction programs on amino acid substitutions. *Human Mutation*, *35*(7), 794–804. https://doi.org/10.1002/humu.22564

Alirezaie, N., Kernohan, K. D., Hartley, T., Majewski, J., & Hocking, T. D. (2018). ClinPred: Prediction tool to identify disease-relevant nonsynonymous single-nucleotide variants. *The American Journal of Human Genetics*, *103*(4), 474–483.

Altshuler, D., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., & Consortium, G. P. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061–1073. https://doi.org/10.1038/nature09534

Amberger, J. S., & Hamosh, A. (2017). Searching Online Mendelian Inheritance in Man (OMIM): A knowledgebase of human genes and genetic phenotypes. *Current Protocols in Bioinformatics*, *58*, 1.2.1–1.2.12. https://doi.org/10.1002/cpbi.27

De Baets, G., Van Durme, J., Reumers, J., Maurer-Stroh, S., Vanhee, P., Dopazo, J., & Rousseau, F. (2012). SNPeffect 4.0: On-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Research*, *40*(D1), D935–D939. https://doi.org/10.1093/nar/gkr996

Balasubramanian, S., Fu, Y., Pawashe, M., McGillivray, P., Jin, M., Liu, J., & Gerstein, M. (2017). Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. *Nature Communications*, *8*(1), 382. https://doi.org/10.1038/s41467-017-00443-5

Bao, L., Zhou, M., & Cui, Y. (2005). nsSNPAnalyzer: Identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Research*, *33*(Suppl. 2), W480–W482. https://doi.org/10.1093/nar/gki372

Barenboim, M., & Manke, T. (2013). ChroMoS: An integrated web tool for SNP classification, prioritization and functional interpretation. *Bioinformatics*, *29*(17), 2197–2198. https://doi.org/10.1093/bioinformatics/btt356

Bendl, J., Musil, M., Štourač, J., Zendulka, J., Damborský, J., & Brezovský, J. (2016). PredictSNP2: A unified platform for accurately evaluating SNP effects by exploiting the different characteristics of variants in distinct genomic regions. *PLOS Computational Biology*, *12*(5), e1004962.

Bermejo-Das-Neves, C., Nguyen, H. N., Poch, O., & Thompson, J. D. (2014). A comprehensive study of small non-frameshift insertions/deletions in proteins and prediction of their phenotypic effects by a machine learning method (KD4i). *BMC Bioinformatics*, *15*(1), 111. https://doi.org/10.1186/1471-2105-15-111

Bertoldi, L., Forcato, C., Vitulo, N., Birolo, G., De Pascale, F., Feltrin, E., & Valle, G. (2017). QueryOR: A comprehensive web platform for genetic variant analysis and prioritization. *BMC Bioinformatics*, *18*(1), 225. https://doi.org/10.1186/s12859-017-1654-4

Brenner, S. E. (2007). Common sense for our genomes. *Nature*, *449*(7164), 783–784. https://doi.org/10.1038/449783a

Buske, O. J., Manickaraj, A., Mital, S., Ray, P. N., & Brudno, M. (2013). Identification of deleterious synonymous variants in human genomes. *Bioinformatics*, *29*(15), 1843–1850. https://doi.org/10.1093/bioinformatics/btt308

Capriotti, E., & Altman, R. B. (2011). A new disease-specific machine learning approach for the prediction of cancer-causing missense variants. *Genomics*, *98*(4), 310–317. https://doi.org/10.1016/j.ygeno.2011.06.010

Capriotti, E., Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, *22*(22), 2729–2734. https://doi.org/10.1093/bioinformatics/btl423

Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P. L., Altman, R. B., & Casadio, R. (2013). WS-SNPs&GO: A web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics*, *14*(3), S6.

Capriotti, E., & Casadio, R. (2007). K-Fold: A tool for the prediction of the protein folding kinetic order and rate. *Bioinformatics*, *23*(3), 385–386. https://doi.org/10.1093/bioinformatics/btl610

Cariaso, M., & Lennon, G. (2012). SNPedia: A wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Research*, *40*(Database issue), D1308–D1312. https://doi.org/10.1093/nar/gkr798

Carter, H., Douville, C., Stenson, P. D., Cooper, D. N., & Karchin, R. (2013). Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics*, *14*(3), S3.

Cheng, J., Nguyen, T. Y. D., Cygan, K. J., Çelik, M. H., Fairbrother, W. G., & Gagneur, J. (2019). MMSplice: Modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biology*, *20*(1), 48.

Cheng, J., Randall, A., & Baldi, P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function, and Genetics*, *62*(4), 1125–1132. https://doi.org/10.1002/prot.20810

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, *31*(16), 2745–2747. https://doi.org/10.1093/bioinformatics/btv195

Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., & Lu, X. (2012). Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in Genetics*, *3*, 35. https://doi.org/10.3389/fgene.2012.00035

Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., & Ruden, D. M. (2012). A program for annotating and predicting the effects of

single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Conchillo-Solé, O., de Groot, N. S., Avilés, F. X., Vendrell, J., Daura, X., & Ventura, S. (2007). AGGRESCAN: A server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics*, *8*, 65. https://doi.org/10.1186/1471-2105-8-65

Cotton, R., Auerbach, A., Beckmann, J., Blumenfeld, O., Brookes, A., Brown, A., & Greenblatt, M. (2008). Recommendations for locus-specific databases and their curation. *Human Mutation*, *29*(1), 2–5.

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP+ +. *PLOS Computational Biology*, *6*(12), e1001025. https://doi.org/10.1371/journal.pcbi.1001025

Dayem Ullah, A. Z., Oscanoa, J., Wang, J., Nagano, A., Lemoine, N. R., & Chelala, C. (2018). SNPnexus: Assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Research*, *46*, 109.

Dees, N. D., Zhang, Q., Kandoth, C., Wendl, M. C., Schierding, W., Koboldt, D. C., & Ding, L. (2012). MuSiC: Identifying mutational significance in cancer genomes. *Genome Research*, *22*(8), 1589–1598. https://doi.org/10.1101/gr.134635.111

Dehouck, Y., Kwasigroch, J. M., Gilis, D., & Rooman, M. (2011). PoPMuSiC 2.1: A web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*, *12*, 151. https://doi.org/10.1186/1471-2105-12-151

Dehouck, Y., Kwasigroch, J. M., Rooman, M., & Gilis, D. (2013). BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Research*, *41*(Web Server issue), W333–W339. https://doi.org/10.1093/nar/gkt450

Desmet, F. O., Hamroun, D., Lalande, M., Collod-Bëroud, G., Claustres, M., & Béroud, C. (2009). Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*, *37*(9), e67–e67. https://doi.org/10.1093/nar/gkp215

Deutsch, C., & Krishnamoorthy, B. (2007). Four-body scoring function for mutagenesis. *Bioinformatics*, *23*(22), 3009–3015. https://doi.org/10.1093/bioinformatics/btm481

Dharanipragada, P., Seelam, S. R., & Parekh, N. (2018). SeqVItA: Sequence variant identification and annotation platform for next generation sequencing data. *Frontiers in Genetics*, *9*, 537.

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., & Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human Molecular Genetics*, *24*(8), 2125–2137. https://doi.org/10.1093/hmg/ddu733

Dosztányi, Z., Magyar, C., Tusnády, G. E., & Simon, I. (2003). SCide: Identification of stabilization centers in proteins. *Bioinformatics*, *19*(7), 899–900. https://doi.org/10.1093/bioinformatics/btg110

Douville, C., Masica, D. L., Stenson, P. D., Cooper, D. N., Gygax, D. M., Kim, R., & Karchin, R. (2016). Assessing the pathogenicity of insertion and deletion variants with the variant effect scoring tool (VEST-Indel). *Human Mutation*, *37*(1), 28–35.

Dunlavy, D. M., O'Leary, D. P., Klimov, D., & Thirumalai, D. (2005). HOPE: A homotopy optimization method for protein structure prediction. *Journal of Computational Biology*, *12*(10), 1275–1288. https://doi.org/10.1089/cmb.2005.12.1275

Fang, Y., Gao, S., Tai, D., Middaugh, C. R., & Fang, J. (2013). Identification of properties important to protein aggregation using feature selection. *BMC Bioinformatics*, *14*(1), 314. https://doi.org/10.1186/1471-2105-14-314

Fokkema, I. F., Taschner, P. E., Schaafsma, G. C., Celli, J., Laros, J. F., & den Dunnen, J. T. (2011). LOVD v.2.0: The next generation in gene variant databases. *Human Mutation*, *32*(5), 557–563. https://doi.org/10.1002/humu.21438

Frousios, K., Iliopoulos, C. S., Schlitt, T., & Simpson, M. A. (2013). Predicting the functional consequences of non-synonymous DNA sequence variants—evaluation of bioinformatics tools and development of a consensus strategy. *Genomics*, *102*(4), 223–228.

Frédéric, M. Y., Lalande, M., Boileau, C., Hamroun, D., Claustres, M., Béroud, C., & Collod-Béroud, G. (2009). UMD-predictor, a new prediction tool for nucleotide substitution pathogenicity—Application to four genes: FBN1, FBN2, TGFBR1, and TGFBR2. *Human Mutation*, *30*(6), 952–959. https://doi.org/10.1002/humu.20970

Gao, M., & Skolnick, J. (2008). DBD-Hunter: A knowledge-based method for the prediction of DNA-protein interactions. *Nucleic Acids Research*, *36*(12), 3978–3992. https://doi.org/10.1093/nar/gkn332

Garber, M., Guttman, M., Clamp, M., Zody, M. C., Friedman, N., & Xie, X. (2009). Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics*, *25*(12), i54–i62. https://doi.org/10.1093/bioinformatics/btp190

Garbuzynskiy, S. O., Lobanov, M. Y., & Galzitskaya, O. V. (2009). FoldAmyloid: A method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics*, *26*(3), 326–332. https://doi.org/10.1093/bioinformatics/btp691

Giollo, M., Martin, A. J. M., Walsh, I., Ferrari, C., & Tosatto, S. C. E. (2014). NeEMO: A method using residue interaction networks to improve prediction of protein stability upon mutation. *BMC Genomics*, *15*, S7. https://doi.org/10.1186/1471-2164-15-S4-S7

Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., & Rost, B. (2014). LocTree3 prediction of localization. *Nucleic Acids Research*, *42*(W1), W350–W355. https://doi.org/10.1093/nar/gku396

Gonzalez-Perez, A., Deu-Pons, J., & Lopez-Bigas, N. (2012). Improving the prediction of the functional impact of cancer mutations by baseline tolerance transformation. *Genome Medicine*, *4*(11), 89. https://doi.org/10.1186/gm390

González-Pérez, A., & López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American Journal of Human Genetics*, *88*(4), 440–449. https://doi.org/10.1016/j.ajhg.2011.03.004

Gosalia, N., Economides, A. N., Dewey, F. E., & Balasubramanian, S. (2017). MAPPIN: A method for annotating, predicting pathogenicity and mode of inheritance for nonsynonymous variants. *Nucleic Acids Research*, *45*(18), 10393–10402.

Gromiha, M. M., Thangakani, A. M., & Selvaraj, S. (2006). FOLD-RATE: Prediction of protein folding rates from amino acid sequence. *Nucleic Acids Research*, *34*(Web Server issue), W70–W74. https://doi.org/10.1093/nar/gkl043

Gulko, B., Hubisz, M. J., Gronau, I., & Siepel, A. (2015). A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nature Genetics*, *47*(3), 276–283. https://doi.org/10.1038/ng.3196

Hecht, M., Bromberg, Y., & Rost, B. (2015). Better prediction of functional effects for sequence variants. *BMC Genomics*, *16*(Suppl 8), S1. https://doi.org/10.1186/1471-2164-16-S8-S1

Hopf, T. A., Ingraham, J. B., Poelwijk, F. J., Schärfe, C. P. I., Springer, M., Sander, C., & Marks, D. S. (2017). Mutation effects predicted from sequence co-variation. *Nature Biotechnology*, *35*(2), 128–135. https://doi.org/10.1038/nbt.3769

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: Protein localization predictor. *Nucleic Acids Research*, *35*(Suppl. 2), W585–W587. https://doi.org/10.1093/nar/gkm259

Hoskins, R. A., Repo, S., Barsky, D., Andreoletti, G., Moult, J., & Brenner, S. E. (2017). Reports from CAGI: The critical assessment of genome interpretation. *Human Mutation*, *38*(9), 1039–1041.

Hu, H., Huff, C. D., Moore, B., Flygare, S., Reese, M. G., & Yandell, M. (2013). VAAST 2.0: Improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution

matrix. *Genetic Epidemiology*, *37*(6), 622–634. https://doi.org/10.1002/gepi.21743

Hu, J., & Ng, P. C. (2013). SIFT Indel: Predictions for the functional effects of amino acid insertions/deletions in proteins. *PLOS One*, *8*(10), e77940. https://doi.org/10.1371/journal.pone.0077940

Hurst, J. M., McMillan, L. E. M., Porter, C. T., Allen, J., Fakorede, A., & Martin, A. C. R. (2009). The SAAPdb web resource: A large-scale structural analysis of mutant proteins. *Human Mutation*, *30*(4), 616–624. https://doi.org/10.1002/humu.20898

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., & Sieh, W. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American Journal of Human Genetics*, *99*(4), 877–885. https://doi.org/10.1016/j.ajhg.2016.08.016

Ionita-Laza, I., McCallum, K., Xu, B., & Buxbaum, J. D. (2016). A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nature Genetics*, *48*(2), 214–220. https://doi.org/10.1038/ng.3477

Javed, A., Agrawal, S., & Ng, P. C. (2014). Phen-gen: Combining phenotype and genotype to analyze rare disorders. *Nature Methods*, *11*(9), 935–937. https://doi.org/10.1038/nmeth.3046

Jia, P., & Zhao, Z. (2014). VarWalker: Personalized Mutation network analysis of putative cancer genes from next-generation sequencing data. *PLOS Computational Biology*, *10*(2), e1003460. https://doi.org/10.1371/journal.pcbi.1003460

Jian, X., Boerwinkle, E., & Liu, X. (2014). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Research*, *42*(22), 13534–13544. https://doi.org/10.1093/nar/gku1206

Johansen, M. B., Izarzugaza, J. M. G., Brunak, S., Petersen, T. N., & Gupta, R. (2013). Prediction of disease causing non-synonymous SNPs by the artificial neural network predictor netdiseaseSNP. *PLOS One*, *8*(7), e68370. https://doi.org/10.1371/journal.pone.0068370

Kaminker, J. S., Zhang, Y., Watanabe, C., & Zhang, Z. (2007). CanPredict: A computational tool for predicting cancer-associated missense mutations. *Nucleic Acids Research*, *35*(Suppl. 2), W595–W598. https://doi.org/10.1093/nar/gkm405

Kang, S., Chen, G., & Xiao, G. (2009). Robust prediction of mutation-induced protein stability change by property encoding of amino acids. *Protein Engineering, Design and Selection*, *22*(2), 75–83. https://doi.org/10.1093/protein/gzn063

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., & Birnbaum, D. P. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210. https://doi.org/10.1101/531210

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315. https://doi.org/10.1038/ng.2892

Knecht, C., Mort, M., Junge, O., Cooper, D. N., Krawczak, M., & Caliebe, A. (2016). IMHOTEP—A composite score integrating popular tools for predicting the functional consequences of non-synonymous sequence variants. *Nucleic Acids Research*, *45*(3), e13–e13.

Krassowski, M., Paczkowska, M., Cullion, K., Huang, T., Dzneladze, I., Ouellette, B. F. F., & Reimand, J. (2017). ActiveDriverDB: Human disease mutations and genome variation in post-translational modification sites of proteins. *Nucleic Acids Research*, *46*(D1), D901–D910.

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), 1073–1082. https://doi.org/10.1038/nprot.2009.86

Kurgan, L., Cios, K., & Chen, K. (2008). SCPRED: Accurate prediction of protein structural class for sequences of twilight-zone similarity with predicting sequences. *BMC Bioinformatics*, *9*, 226. https://doi.org/10.1186/1471-2105-9-226

Laimer, J., Hofer, H., Fritz, M., Wegenkittl, S., & Lackner, P. (2015). MAESTRO-multi agent stability prediction upon point mutations. *BMC Bioinformatics*, *16*(1), 116.

Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., & Maglott, D. R. (2016). ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*, *44*(D1), D862–D868. https://doi.org/10.1093/nar/gkv1222

Lappalainen, I., Lopez, J., Skipper, L., Hefferon, T., Spalding, J. D., Garner, J., & Church, D. M. (2013). DbVar and DGVa: Public archives for genomic structural variation. *Nucleic Acids Research*, *41*(Database issue), D936–D941. https://doi.org/10.1093/nar/gks1213

Lappalainen, T., Scott, A. J., Brandt, M., & Hall, I. M. (2019). Genomic analysis in the age of human genome sequencing. *Cell*, *177*(1), 70–84.

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., & Getz, G. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, *499*(7457), 214–218. https://doi.org/10.1038/nature12213

Lehmann, K. V., & Chen, T. (2013). Exploring functional variant discovery in non-coding regions with SInBaD. *Nucleic Acids Research*, *41*(1), e7–e7. https://doi.org/10.1093/nar/gks800

Leiserson, M. D. M., Wu, H. T., Vandin, F., & Raphael, B. J. (2015). CoMEt: A statistical approach to identify combinations of mutually exclusive alterations in cancer. *Genome Biology*, *16*(1), 160. https://doi.org/10.1186/s13059-015-0700-7

Li, B., Krishnan, V. G., Mort, M. E., Xin, F., Kamati, K. K., Cooper, D. N., & Radivojac, P. (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*, *25*(21), 2744–2750. https://doi.org/10.1093/bioinformatics/btp528

Li, M. J., Li, M., Liu, Z., Yan, B., Pan, Z., Huang, D., & Wang, J. (2017). Cepip: Context-dependent epigenomic weighting for prioritization of regulatory variants and disease-associated genes. *Genome Biology*, *18*(1), 52. https://doi.org/10.1186/s13059-017-1177-3

Li, Q., & Wang, K. (2017). InterVar: Clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *American Journal of Human Genetics*, *100*(2), 267–280. https://doi.org/10.1016/j.ajhg.2017.01.004

Liu, M., Watson, L. T., & Zhang, L. (2015). Predicting the combined effect of multiple genetic variants. *Human Genomics*, *9*, 18. https://doi.org/10.1186/s40246-015-0040-4

Liu, X., White, S., Peng, B., Johnson, A. D., Brody, J. A., Li, A. H., & Gibbs, R. (2016). WGSA: An annotation pipeline for human genome sequencing studies. *Journal of Medical Genetics*, *53*(2), 111–112.

Liu, X., Wu, C., Li, C., & Boerwinkle, E. (2016). dbNSFP v3.0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Human Mutation*, *37*(3), 235–241. https://doi.org/10.1002/humu.22932

Livingstone, M., Folkman, L., Yang, Y., Zhang, P., Mort, M., Cooper, D. N., & Zhou, Y. (2017). Investigating DNA-, RNA-, and protein-based features as a means to discriminate pathogenic synonymous variants. *Human Mutation*, *38*(10), 1336–1347.

Lopes, M. C., Joyce, C., Ritchie, G. R., John, S. L., Cunningham, F., Asimit, J., & Zeggini, E. (2012). A combined functional annotation score for non-synonymous variants. *Human Heredity*, *73*(1), 47–51.

Lu, Q., Hu, Y., Sun, J., Cheng, Y., Cheung, K. H., & Zhao, H. (2015). A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data. *Scientific Reports*, *5*, 10576. https://doi.org/10.1038/srep10576.

López-Ferrando, V., Gazzo, A., De La Cruz, X., Orozco, M., & Gelpí, J. L. (2017). PMut: A web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Research*, *45*(W1), W222–W228. https://doi.org/10.1093/nar/gkx313

Macintyre, G., Bailey, J., Haviv, I., & Kowalczyk, A. (2010). is-rSNP: A novel technique for in silico regulatory SNP detection. *Bioinformatics*, *26*(18), i524–i530. https://doi.org/10.1093/bioinformatics/btq378

Mao, Y., Chen, H., Liang, H., Meric-Bernstam, F., Mills, G. B., & Chen, K. (2013). CanDrA: Cancer-specific driver missense mutation annotation

with optimized features. *PLOS One*, *8*(10), e77945. https://doi.org/10.1371/journal.pone.0077945

Marini, N. J., Thomas, P. D., & Rine, J. (2010). The use of orthologous sequences to predict the impact of amino acid substitutions on protein function. *PLOS Genetics*, *6*(5), 3. https://doi.org/10.1371/journal.pgen.1000968

Masso, M., & Vaisman, I. I. (2010). AUTO-MUTE: Web-based tools for predicting stability changes in proteins due to single amino acid replacements. *Protein Engineering, Design and Selection*, *23*(8), 683–687. https://doi.org/10.1093/protein/gzq042

Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., De La Paz, M. L., Martins, I. C., Reumers, J., & Rousseau, F. (2010). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nature Methods*, *7*(3), 237–242. https://doi.org/10.1038/nmeth.1432

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., & Cunningham, F. (2016). The Ensembl variant effect predictor. *Genome Biology*, *17*(1), 122. https://doi.org/10.1186/s13059-016-0974-4

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., & Thomas, P. D. (2017). PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research*, *45*(D1), D183–D189. https://doi.org/10.1093/nar/gkw1138

Mort, M., Sterne-Weiler, T., Li, B., Ball, E. V., Cooper, D. N., Radivojac, P., & Mooney, S. D. (2014). MutPred splice: Machine learning-based prediction of exonic variants that disrupt splicing. *Genome Biology*, *15*(1), R19.

Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D. J., Salichos, L., & Rozowsky, J. (2016). The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biology*, *17*(1), 53.

Nalla, V. K., & Rogan, P. K. (2005). Automated splicing mutation analysis by information theory. *Human Mutation*, *25*(4), 334–342. https://doi.org/10.1002/humu.20151

Niroula, A., Urolagin, S., & Vihinen, M. (2015). PON-P2: Prediction method for fast and reliable identification of harmful variants. *PLOS One*, *10*(2), e0117380.

Niroula, A., & Vihinen, M. (2016). PON-mt-tRNA: A multifactorial probability-based method for classification of mitochondrial tRNA variations. *Nucleic Acids Research*, *44*(5), 2020–2027. https://doi.org/10.1093/nar/gkw046

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., & Trajanoski, Z. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in Bioinformatics*, *15*(2), 256–278.

Pagel, K. A., Pejaver, V., Lin, G. N., Nam, H. -J., Mort, M., Cooper, D. N., & Radivojac, P. (2017). When loss-of-function is loss of function: Assessing mutational signatures and impact of loss-of-function genetic variants. *Bioinformatics*, *33*(14), i389–i398.

Pagon, R. A., Tarczy-Hornoch, P., Baskin, P. K., Edwards, J. E., Covington, M. L., Espeseth, M., & Palepu, R. D. (2002). GeneTests-GeneClinics: Genetic testing information for a growing audience. *Human Mutation*, *19*(5), 501–509. https://doi.org/10.1002/humu.10069

Pandurangan, A. P., Ochoa-Montaño, B., Ascher, D. B., & Blundell, T. L. (2017). SDM: A server for predicting effects of mutations on protein stability. *Nucleic Acids Research*, *45*(W1), W229–W235. https://doi.org/10.1093/nar/gkx439

Pappalardo, M., & Wass, M. N. (2014). VarMod: Modelling the functional effects of non-synonymous variants. *Nucleic Acids Research*, *42*(W1), W331–W336. https://doi.org/10.1093/nar/gku483

Park, E., Pan, Z., Zhang, Z., Lin, L., & Xing, Y. (2018). The expanding landscape of alternative splicing variation in human populations. *The American Journal of Human Genetics*, *102*(1), 11–26.

Parthiban, V., Gromiha, M. M., Abhinandan, M., & Schomburg, D. (2007). Computational modeling of protein mutant stability: Analysis and optimization of statistical potentials and structural features reveal insights into prediction model development. *BMC Structural Biology*, *7*, 54. https://doi.org/10.1186/1472-6807-7-54

Peng, B. (2015). Reproducible simulations of realistic samples for next-generation sequencing studies using variant simulation tools. *Genetic Epidemiology*, *39*(1), 45–52. https://doi.org/10.1002/gepi.21867

Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Pires, D. E. V., Ascher, D. B., & Blundell, T. L. (2014). DUET: A server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Research*, *42*(W1), W314–W319. https://doi.org/10.1093/nar/gku411

Pokala, N., & Handel, T. M. (2005). Energy functions for protein design: Adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *Journal of Molecular Biology*, *347*(1), 203–227. https://doi.org/10.1016/j.jmb.2004.12.019

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*, *20*(1), 110–121. https://doi.org/10.1101/gr.097857.109

Preeprem, T., & Gibson, G. (2014). SDS, a structural disruption score for assessment of missense variant deleteriousness. *Frontiers in Genetics*, *5*, 82.

Punta, M., & Rost, B. (2005). PROFcon: Novel prediction of long-range contacts. *Bioinformatics*, *21*(13), 2960–2968. https://doi.org/10.1093/bioinformatics/bti454

Qin, S., Pang, X., & Zhou, H. X. (2011). Automated prediction of protein association rate constants. *Structure*, *19*(12), 1744–1751. https://doi.org/10.1016/j.str.2011.10.015

Quang, D., Chen, Y., & Xie, X. (2015). DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*, *31*(5), 761–763. https://doi.org/10.1093/bioinformatics/btu703

Ramensky, V., Bork, P., & Sunyaev, S. (2002). Human non-synonymous SNPs: Server and survey. *Nucleic Acids Research*, *30*(17), 3894–3900.

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Research*, *39*(17), e118–e118. https://doi.org/10.1093/nar/gkr407

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., & Committee, A. L. Q. A. (2015). 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. *Genetics in Medicine*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Ritchie, G. R. S., Dunham, I., Zeggini, E., & Flicek, P. (2014). Functional annotation of noncoding sequence variants. *Nature Methods*, *11*(3), 294–296. https://doi.org/10.1038/nmeth.2832

Rousseau, F., Schymkowitz, J., & Serrano, L. (2006). Protein aggregation and amyloidosis: Confusion of the kinds? *Current Opinion in Structural Biology*, *16*(1), 118–126. https://doi.org/10.1016/j.sbi.2006.01.011

Ryan, M., Diekhans, M., Lien, S., Liu, Y., & Karchin, R. (2009). LS-SNP/PDB: Annotated non-synonymous SNPs mapped to Protein Data Bank structures. *Bioinformatics*, *25*(11), 1431–1432. https://doi.org/10.1093/bioinformatics/btp242

Ryan, N. M., Morris, S. W., Porteous, D. J., Taylor, M. S., & Evans, K. L. (2014). SuRFing the genomics wave: An R package for prioritising SNPs by functionality. *Genome Medicine*, *6*(10), 79. https://doi.org/10.1186/s13073-014-0079-1

Sasidharan Nair, P., & Vihinen, M. (2013). VariBench: A benchmark database for variations. *Human Mutation*, *34*(1), 42–49. https://doi.org/10.1002/humu.22204

Savojardo, C., Fariselli, P., Martelli, P. L., & Casadio, R. (2016). INPS-MD: A web server to predict stability of protein variants from sequence and structure. *Bioinformatics*, *32*(16), 2542–2544.

Schaafsma, G. C. P., & Vihinen, M. (2015). VariSNP, A benchmark database for variations from dbSNP. *Human Mutation*, *36*(2), 161–166. https://doi.org/10.1002/humu.22727

Schaefer, C., Meier, A., Rost, B., & Bromberg, Y. (2012). Snpdbe: Constructing an nsSnp functional impacts database. *Bioinformatics*, *28*(4), 601–602. https://doi.org/10.1093/bioinformatics/btr705

Schwarz, J. M., Cooper, D. N., Schuelke, M., & Seelow, D. (2014). Mutationtaster2: Mutation prediction for the deep-sequencing age. *Nature Methods*, *11*(4), 361–362. https://doi.org/10.1038/nmeth.2890

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: An online force field. *Nucleic Acids Research*, *33*(Web Server issue), W382–W388. https://doi.org/10.1093/nar/gki387

Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, *29*(1), 308–311.

Shihab, H. A., Gough, J., Mort, M., Cooper, D. N., Day, I. N. M., & Gaunt, T. R. (2014). Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Human Genomics*, *8*(1), 11. https://doi.org/10.1186/1479-7364-8-11

Shihab, H. A., Rogers, M. F., Gough, J., Mort, M., Cooper, D. N., Day, I. N. M., & Campbell, C. (2015). An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*, *31*(10), 1536–1543. https://doi.org/10.1093/bioinformatics/btv009

Shringarpure, S. S., & Bustamante, C. D. (2015). Privacy risks from genomic data-sharing beacons. *American Journal of Human Genetics*, *97*(5), 631–646. https://doi.org/10.1016/j.ajhg.2015.09.010

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., & Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, *15*(8), 1034–1050. https://doi.org/10.1101/gr.3715005

Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, *40*(W1), W452–W457. https://doi.org/10.1093/nar/gks539

Smedley, D., Jacobsen, J. O. B., Jäger, M., Köhler, S., Holtgrewe, M., Schubach, M., & Robinson, P. N. (2015). Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nature Protocols*, *10*(12), 2004–2015. https://doi.org/10.1038/nprot.2015.124

Smedley, D., Schubach, M., Jacobsen, J. O., Köhler, S., Zemojtel, T., Spielmann, M., & McMurry, J. A. (2016). A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *The American Journal of Human Genetics*, *99*(3), 595–606.

Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., & Cooper, D. N. (2017). The Human Gene Mutation Database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human Genetics*, *136*(6), 665–677. https://doi.org/10.1007/s00439-017-1779-6

Stone, E. A., & Sidow, A. (2005). Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Research*, *15*(7), 978–986. https://doi.org/10.1101/gr.3804205

Tamborero, D., Gonzalez-Perez, A., & Lopez-Bigas, N. (2013). Oncodrive-CLUST: Exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics*, *29*(18), 2238–2244. https://doi.org/10.1093/bioinformatics/btt395

Tang, H., & Thomas, P. D. (2016). PANTHER-PSEP: Predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics*, *32*(14), 2230–2232. https://doi.org/10.1093/bioinformatics/btw222

Tartaglia, G. G., Cavalli, A., Pellarin, R., & Caflisch, A. (2005). Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. *Protein Science*, *14*(10), 2723–2734. https://doi.org/10.1110/ps.051471205

Tartaglia, G. G., & Vendruscolo, M. (2008). The Zyggregator method for predicting protein aggregation propensities. *Chemical Society Reviews*, *37*(7), 1395–1401. https://doi.org/10.1039/b706784b

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., & Forbes, S. A. (2019). COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, *47*(D1), D941–D947. https://doi.org/10.1093/nar/gky1015

Tavtigian, S. V., Deffenbaugh, A. M., Yin, L., Judkins, T., Scholl, T., Samollow, P. B., & Thomas, A. (2006). Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *Journal of Medical Genetics*, *43*(4), 295–305. https://doi.org/10.1136/jmg.2005.033878

Teng, S., Srivastava, A. K., & Wang, L. (2010). Sequence feature-based prediction of protein stability changes upon amino acid substitutions. *BMC Genomics*, *11*(2), 1.

Terui, H., Akagi, K., Kawame, H., & Yura, K. (2013). CoDP: Predicting the impact of unclassified genetic variants in MSH6 by the combination of different properties of the protein. *Journal of Biomedical Science*, *20*(1), 25. https://doi.org/10.1186/1423-0127-20-25

Thompson, B. A., Spurdle, A. B., Plazzer, J. P., Greenblatt, M. S., Akagi, K., Al-Mulla, F., & Barbera, V. M. (2014). Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nature Genetics*, *46*(2), 107–115. https://doi.org/10.1038/ng.2854

Thorn, C. F., Klein, T. E., & Altman, R. B. (2013). PharmGKB: The Pharmacogenomics Knowledge Base. *Methods in Molecular Biology*, *1015*, 311–320. https://doi.org/10.1007/978-1-62703-435-7_20

Tian, J., Wu, N., Guo, X., Guo, J., Zhang, J., & Fan, Y. (2007). Predicting the phenotypic effects of non-synonymous single nucleotide polymorphisms based on support vector machines. *BMC Bioinformatics*, *8*, 450. https://doi.org/10.1186/1471-2105-8-450

Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., & Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols*, *11*(1), 1–9.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *The American Journal of Human Genetics*, *101*(1), 5–22.

Vuong, H., Che, A., Ravichandran, S., Luke, B. T., Collins, J. R., & Mudunuri, U. S. (2015). AVIA v2.0: Annotation, visualization and impact analysis of genomic variants and genes. *Bioinformatics*, *31*(16), 2748–2750. https://doi.org/10.1093/bioinformatics/btv200

Walsh, I., Seno, F., Tosatto, S. C. E., & Trovato, A. (2014). PASTA 2.0: An improved server for protein aggregation prediction. *Nucleic Acids Research*, *42*(W1), W301–W307. https://doi.org/10.1093/nar/gku399

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), e164–e164. https://doi.org/10.1093/nar/gkq603

Wang, M., Zhao, X. M., Takemoto, K., Xu, H., Li, Y., Akutsu, T., & Song, J. (2012). FunSAV: Predicting the functional effect of single amino acid variants using a two-stage random forest model. *PLOS One*, *7*(8), e43847. https://doi.org/10.1371/journal.pone.0043847

Wishart, D. S., Arndt, D., Berjanskii, M., Guo, A. C., Shi, Y., Shrivastava, S., & Lin, G. (2008). PPT-DB: The protein property prediction and testing database. *Nucleic Acids Research*, *36*(Suppl. 1), D222–D229. https://doi.org/10.1093/nar/gkm800

Wong, W. C., Kim, D., Carter, H., Diekhans, M., Ryan, M. C., & Karchin, R. (2011). CHASM and SNVBox: Toolkit for detecting biologically important single nucleotide mutations in cancer. *Bioinformatics*, *27*(15), 2147–2148. https://doi.org/10.1093/bioinformatics/btr357

Woolfe, A., Mullikin, J. C., & Elnitski, L. (2010). Genomic features defining exonic variants that modulate splicing. *Genome Biology*, *11*(2), R20.

Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., & Hughes, T. R. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *Science*, *347*(6218), 1254806–1254806.

Xu, B., Yang, Y., Liang, H., & Zhou, Y. (2009). An all-atom knowledge-based energy function for protein-DNA threading, docking decoy discrimination, and prediction of transcription-factor binding profiles. *Proteins: Structure, Function, and Bioinformatics*, *76*(3), 718–730. https://doi.org/10.1002/prot.22384

Xu, Z., & Taylor, J. A. (2009). SNPinfo: Integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Research*, *37*(Suppl. 2), W600–W605. https://doi.org/10.1093/nar/gkp290

Yang, H., & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature Protocols*, *10*(10), 1556–1566. https://doi.org/10.1038/nprot.2015.105

Ye, Z. Q., Zhao, S. Q., Gao, G., Liu, X. Q., Langlois, R. E., Lu, H., & Wei, L. (2007). Finding new structural and sequence attributes to predict possible disease association of single amino acid polymorphism (SAP). *Bioinformatics*, *23*(12), 1444–1450. https://doi.org/10.1093/bioinformatics/btm119

Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology*, *11*(2-3), 377–394. https://doi.org/10.1089/1066527041410418

Yin, S., Ding, F., & Dokholyan, N. V. (2007). Eris: An automated estimator of protein stability [2]. *Nature Methods*, *4*(6), 466–467. https://doi.org/10.1038/nmeth0607-466

Zambrano, R., Jamroz, M., Szczasiuk, A., Pujols, J., Kmiecik, S., & Ventura, S. (2015). AGGRESCAN3D (A3D): Server for prediction of aggregation properties of protein structures. *Nucleic Acids Research*, *43*(W1), W306–W313. https://doi.org/10.1093/nar/gkv359

Zeng, S., Yang, J., Chung, B. H.-Y., Lau, Y. L., & Yang, W. (2014). EFIN: Predicting the functional impact of nonsynonymous single nucleotide polymorphisms in human genome. *BMC Genomics*, *15*(1), 455.

Zhang, T., Wu, Y., Lan, Z., Shi, Q., Yang, Y., & Guo, J. (2017). Syntool: A novel region-based intolerance score to single nucleotide substitution for synonymous mutations predictions based on 123,136 individuals. *BioMed Research International*, *2017*, 5096208.

Zhou, H., & Zhou, Y. (2002). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Science*, *11*(11), 2714–2726. https://doi.org/10.1110/ps.0217002

Zhou, J., Theesfeld, C. L., Yao, K., Chen, K. M., Wong, A. K., & Troyanskaya, O. G. (2018). Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nature Genetics*, *50*(8), 1171–1179.

Zhou, J., & Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. *Nature Methods*, *12*(10), 931–934. https://doi.org/10.1038/nmeth.3547

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Hu Z, Yu C, Furutsuki M, et al. VIPdb, a genetic Variant Impact Predictor Database. *Human Mutation*. 2019;40:1202–1214. https://doi.org/10.1002/humu.23858