**Supplementary Table S1. Summary of comparing computational methods.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Methods** | **Core model** | **Damaging** | **Tolerable** | **Training data** | **Testing data** | **Update information** |
| **Class one**: function prediction methods | | | | | | |
| FATHMM (1) | Hidden Markov models | ≤ 1.5 | > 1.5 | SNVs from HGMD (November 2011) and UniProt (November 2011) | SNVs from VariBench database (November 2011), SwissVar (February 2011) and four cancer-associated  genes (*BRCA1, MSH2, MLH1,* and *TP53*) (Hicks et al. 2011) | Current version: FATHMM v2.3 |
| fitCons (2) | INSIGHT (Inference of Natural Selection from Interspersed Genomically coHerent elemenTs) | > 0.7 | < 0.7 | Genomes of 54 unrelated human individuals | three types of functional elements | Current version: V1.01 28-Aug-2014 |
| LRT (3) | Likelihood ratio test of codon neutrality | < 0.001 | > 0.001 | Coding sequences of 32 vertebrate species | three human genomes (Levy et al. 2007; Wang et al. 2008; Wheeler et al. 2008) | Last updated: November 20, 2009 |
| Mutation Assessor (4) | Combinatorial entropy formalism | > 1.9 | ≤ 1.9 | SNVs from COSMIC database ( release 49) | SNVs from UniProt (HUMSAVAR, release 2010\_08), IARC TP53 database and COSMIC database ( release 49) | Dec 31, 2015 Release3 |
| MutationTaster (5) | Naive Bayes classifier | > 0.5 | ≤ 0.5 | SNVs from dbSNP, OMIM, HGMD and the literature | SNVs from dbSNP, OMIM, HGMD and the literature | Apr 2014 Mutation Taster2 published, training SNVs from 1000 G and HGMD |
| PolyPhen2-HDIV (6) | Naive Bayes classifier | > 0.453 | < 0.453 | SNVs from UniRef100 (release 15.12 of 15-Dec-2009) and UniProtKB/Swiss-Prot (release 57.12 of 15-Dec-2009) | SNVs from UniProtKB/Swiss-Prot (release 57.12 of 15-Dec-2009) | Last updated: Mar 08, 2012; current version: PolyPhen-2 v2.2.2 (r394) Feb 23, 2012 |
| PolyPhen2-HVAR (6) | Naive Bayes classifier | > 0.447 | < 0.447 | SNVs from UniRef100 (release 15.12 of 15-Dec-2009) and UniProtKB/Swiss-Prot (release 57.12 of 15-Dec-2009) | SNVs from UniProtKB/Swiss-Prot (release 57.12 of 15-Dec-2009) | Last updated: Mar 08, 2012; current version: PolyPhen-2 v2.2.2 (r394) Feb 23, 2012 |
| PROVEAN (7) | Delta alignment score | ≤ 2.5 | > -2.5 | SNVs from UniProt/HUMSAVAR (Release 2011\_09) | SNVs from UniProt (Release 2011\_09) and experimental datasets from mutagenesis experiments  previously carried out for the E.coli LacI protein (Markiewicz et al. 1994) and the human  tumor suppressor TP53 protein | Last updated: Jan 30, 2015 current version: v1.1.5 May 7, 2014 |
| SIFT (8) | Position-specific scoring matrix | ≤ 0.05 | > 0.05 | 1,750 deleterious and 2,254 tolerant nsSNVs of E. coli LacI gene | 4004 substitutions from LacI (Markiewiczet al. 1994; Suckow et al. 1996), 336 substitutions from HIV-1 protease (Loeb et al. 1989), and 2015 substitutions from bacteriophage T4 lysozyme (Rennell et al. 1991) | Last updated: Aug 2011; current version: SIFT v. 1.03 |
| VEST3 (9) | Random Forest | > 0.5 | < 0.5 | SNVs from HGMD (2012v2) and the exome sequencing project (ESP6500 accessed 07/2012) | SNVs from HGMD (2012v2) and the exome sequencing project (ESP6500 accessed 07/2012) | most recent version: VEST-4， Positive class expanded and updated to HGMD (2017.1)，Neutral class changed to ExAC Release 1 (2/2017) |
| **Class two**: conservation methods | | | | | | |
| GERP++ (10) | Maximum likelihood evolutionary rate estimation | > 2 | < 2 | Genomes of 34 mammals | Genomes of 33 other mammalian | not updated |
| phastCons (11) | Two-state phylogenetic hidden Markov Model | > 0.999 | ≤ 0.999 | Genomes of seven vertebrates  Genomes of 20 mammals | five vertebrate genomes, four insect genomes, two Caenorhabditis genomes, and seven Saccharomyces genomes. | Current version: PHAST 1.4 October, 2016, Genomes of 100 vertebrates |
| PhyloP (12) | Distributions of the number of substitutions based on a phylogenetic hidden Markov model | > 2 | < 2 | Genomes of seven vertebrates,  Genomes of 20 mammals | 100,000 fourfold degenerate sites extracted from alignments of up to 19 species for the 44 ENCODE regions (Margulies et al. 2007) | Current version: PHAST 1.4 October, 2016, Genomes of 100 vertebrates |
| SiPhy (13) | Inferring nucleotide substitution pattern  per site | > 12 | < 12 | Genomes of 29 mammals | ENCODE regions (Birney et al., 2007) | Current version: 0.5 May 1, 2009 |
| **Class three**: ensemble methods | | | | | | |
| CADD (14) | Linear kernel support vector machine | > 20 | ≤ 20 | 16,627,775 high-frequency human-derived alleles and 49,407,057 “simulated” variants | SNVs from *MLL2* gene (Makrythanasis et al), ESP, *HBB* gene and ClinVar database (release date June 16 2012), somatic mutations from p53, variants from two enhancers and one promoter | Current version: Webserver v1.3  CADD v1.1: Dec 2, 2014, a slightly extended and updated annotation set |
| DANN (15) | Deep neural network | > 0.99 | < 0.99 | 16,627,775 high-frequency human-derived alleles and 49,407,057 “simulated” variants | 3,325,555 “observed” variants and “simulated” variants | not updated |
| Eigen (16) | Hierarchical model | > 0 | < 0 | Variants from dbNSFP v2.7 | variants from ClinVar database, *MLL2, CFTR, BRCA1* and *BRCA2* genes for Mendelian diseases, de novo mutations associated with ASD, EPI, ID and SCZ, GWAS and eQTL SNPs, noncoding cancer mutations from the COSMIC database | Current version: Eigen and Eigen-PC v1.1 |
| FATHMM-MKL (17) | Multiple kernel learning | > 0.5 | ≤ 0.5 | SNVs from HGMD (release 2013.4) and 1000G | SNVs from HGMD (release 2013.4), 1000G and ClinVar | Current version: FATHMM v2.3 |
| GenoCanyon (18) | Statistical Model | > 0.999 | ≤ 0.999 | ENCODE project | variants from ClinVar in June 2014 | Not updated |
| M-CAP (19) | Gradient boosting trees | > 0.025 | ≤ 0.025 | SNVs from HGMD Pro version 2015.2 and ExAC version 0.3 (Jan 13, 2015 release) | rare SNVs from HGMD Pro version 2015.2, 1000G, patient exomes and Mendelian mutations associated with *BRCA1, BRCA2, CFTR, MLL2* | Not updated |
| MetaLR (20) | logistic regression | > 0.5 | ≤ 0.5 | SNVs from UniProt database | SNVs from 57 publications (after 1 January 2011) from the journal Nature Genetics, CHARGE sequencing project and VariBench dataset II | Not updated |
| MetaSVM (20) | support vector machine | > 0 | ≤ 0 | SNVs from Uniprot database | SNVs from 57 publications (after 1 January 2011) for the journal Nature Genetics, CHARGE sequencing project and VariBench dataset II | Not updated |
| REVEL (21) | Random Forest | > 0.4 | < 0.4 | SNVs from HGMD version 2015.2,  Missense exome sequencing variants from ESP、ARIC、KGP | SNVs from SwissVar disease (release 2015\_10), ClinVar database and variants from  ESP、ARIC、KGP | Not updated |

Note: we performed a comparative study of 23 pathogenicity-computation methods, including 10 function prediction methods, 4 conservation methods and 9 ensemble methods. The predicted pathogenicity scores and the cutoff values for distinguishing deleterious missense variants were sourced from dbNSFP database (https://sites.google.com/site/jpopgen/dbNSFP).

**Reference**:

1. Shihab,H.A., Gough,J., Cooper,D.N., Stenson,P.D., Barker,G.L.A., Edwards,K.J., Day,I.N.M. and Gaunt,T.R. (2013) Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Hum. Mutat.*, **34**, 57–65.

2. Gulko,B., Hubisz,M.J., Gronau,I. and Siepel,A. (2015) A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nat. Genet.*, **47**, 276–83.

3. Chun,S. and Fay,J.C. (2009) Identification of deleterious mutations within three human genomes. *Genome Res.*, **19**, 1553–1561.

4. Reva,B., Antipin,Y. and Sander,C. (2011) Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res.*, **39**.

5. Schwarz,J.M., Rödelsperger,C., Schuelke,M. and Seelow,D. (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods*, **7**, 575–576.

6. IA Adzhubei, S Schmidt L Peshkin,V.R. and A.G. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.

7. Choi,Y., Sims,G.E., Murphy,S., Miller,J.R. and Chan,A.P. (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One*, **7**.

8. Kumar,P., Henikoff,S. and Ng,P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, **4**, 1073–1081.

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

10. Davydov,E. V., Goode,D.L., Sirota,M., Cooper,G.M., Sidow,A. and Batzoglou,S. (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.*, **6**.

11. Spieth,J., Hillier,L.W. and Wilson,R.K. (2005) Evolutionarily conserved elements in vertebrate , insect , worm , and yeast genomes. 10.1101/gr.3715005.

12. Siepel,A., Pollard,K.S. and Haussler,D. (2006) New methods for detecting lineage-specific selection. *Lect. Notes Comput. Sci.*, **3909**, 190–205.

13. Garber,M., Guttman,M., Clamp,M., Zody,M.C., Friedman,N. and Xie,X. (2009) Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics*, **25**, 54–62.

14. Kircher,M. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. g*, **46**, 310–315.

15. Quang,D., Chen,Y. and Xie,X. (2015) DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*, **31**, 761–763.

16. Ionita-Laza,I., McCallum,K., Xu,B. and Buxbaum,J.D. (2016) A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.*, 10.1038/ng.3477.

17. Shihab,H. a, Rogers,M.F., Gough,J., Mort,M., Cooper,D.N., Day,I.N.M., Gaunt,T.R. and Campbell,C. (2015) An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*, **31**, 1536–1543.

18. Lu,Q., Hu,Y., Sun,J., Cheng,Y., Cheung,K.-H. and Zhao,H. (2015) A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data. *Sci. Rep.*, **5**, 10576.

19. Jagadeesh,K.A., Wenger,A.M., Berger,M.J., Guturu,H., Stenson,P.D., Cooper,D.N., Bernstein,J.A. and Bejerano,G. (2016) M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.*, **48**, 1581–1586.

20. Dong,C., Wei,P., Jian,X., Gibbs,R., Boerwinkle,E., Wang,K. and Liu,X. (2015) Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.*, **24**, 2125–2137.

21. Ioannidis,N.M., Rothstein,J.H., Pejaver,V., Middha,S., McDonnell,S.K., Baheti,S., Musolf,A., Li,Q., Holzinger,E., Karyadi,D., *et al.* (2016) REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.*, **99**, 877–885.