

Performance of in silico tools for the evaluation of p16INK4a (CDKN2A) variants in CAGI

Marco Carraro¹, Giovanni Minervini¹, Manuel Giollo^{1,2}, Yana Bromberg^{3,4,5}, Emidio Capriotti⁶, Rita Casadio⁷, Roland Dunbrack⁸, Lisa Elefanti⁹, Pietro Fariselli¹⁰, Carlo Ferrari², Julian Gough¹¹, Roger Hoskins¹², Panagiotis Katsonis¹³, Emanuela Leonardi¹⁴, Olivier Lichtarge^{13,15,16,17}, Chiara Menin⁹, Pier Luigi Martelli⁶, Abhishek Niroula¹⁸, Lipika R. Pal¹⁹, Susanna Repo²⁰, Maria Chiara Scaini⁹, Mauno Vihinen¹⁸, Qiong Wei⁷, Qifang Xu⁷, Yuedong Yang²¹, Yizhou Yin^{19,22}, Jan Zaucha¹¹, Huiying Zhao²³, Yaoqi Zhou²¹, Steven E. Brenner¹², John Moulton^{19,24}, Silvio C.E. Tosatto^{1,25,*}.

1 Department of Biomedical Sciences University of Padova, Viale G. Colombo 3, 35121, Padova, Italy.

2 Department of Information Engineering, University of Padova, Via Gradenigo 6, 35121 Padova, Italy.

3 Department of Biochemistry and Microbiology, Rutgers University, 76 Lipman Dr, New Brunswick, NJ 08901, USA.

4 Department of Genetics, Rutgers University, 145 Bevier Rd, Piscataway, NJ 08854, USA.

5 Technical University of Munich Institute for Advanced Study, (TUM-IAS), Lichtenbergstr. 2a, 85748 Garching/Munich, Germany.

6 BioFolD Unit, Department of Biological, Geological, and Environmental Sciences (BiGeA), University of Bologna, Via F. Selmi 3, Bologna, 40126, Italy.

7 Biocomputing Group, Dept BiGeA, University of Bologna, via San Giacomo 9/2, 40126 Bologna, Italy.

8 Institute for Cancer Research, Fox Chase Cancer Center, 333 Cottman Ave, Philadelphia PA 19111, USA.

9 Immunology and Molecular Oncology Unit, Veneto Institute of Oncology, Padua I-35128, Italy.

10 Dept BCA, University of Padua, viale dell'Università 16, 35020 Legnaro (PD), Italy.

11 Department of Computer Science, University of Bristol, Bristol, UK.

12 Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720-3102.

13 Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, Texas.

14 Department of Woman and Child Health, University of Padova, I-35128 Padova, Italy.

15 Department of Biochemistry & Molecular Biology, Baylor College of Medicine, Houston, Texas

16 Department of Pharmacology, Baylor College of Medicine, Houston, Texas

17 Computational and Integrative Biomedical Research Center, Baylor College of Medicine, Houston, Texas

18 Protein Structure and Bioinformatics Group, Department of Experimental Medical Science, Lund University, BMC B13, 221 84 Lund, Sweden.

19 Institute for Bioscience and Biotechnology Research, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850.

20 EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.

21 Institute for Glycomics and School of Information and Communication Technology, Griffith University, Gold Coast, QLD 4222, Australia.

22 Computational Biology, Bioinformatics and Genomics, Biological Sciences Graduate Program, University of Maryland, College Park, MD 20742, USA.

23 Institute of Health and Biomedical Innovation, Queensland University of Technology, Queensland, 4622, Australia.

24 Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742.

25 CNR Institute of Neuroscience, Padova, Italy.

**Corresponding Author*

Supplementary Material

Performance Measures

All quantitative measures used for the determination of a ranking between the submissions are listed:

- **Pearson CC (PCC):** *R* function *cor* was used. Check the reference manual for details.
- **Pairwise Kendall CC (KCC):** the *R* function *cor* was used. Check the reference manual for details.
- **Root Mean Square Error (RMSE):** defined as

$$\sqrt{\frac{\sum(p_i - r_i)^2}{n}}$$

where n is the number of mutations, p is the vector of predictions and r is the vector of experimental values.

- **Amount of predictions within standard deviations (PWSD):** assuming a normal distribution for the experimental measures and the predicted values, *p-value* from the *student's t-test* for each mutation was used. Then, the number of *p-value* greater or equal than 0.05 was simply counted. Due to the difference in the standard deviations of the submissions, this index was even calculated using a fixed standard deviation of 10%.
- **Area Under the Curve (AUC):** the *performance* function of the *ROCR* *R* package was used. Considering that the experimental measures e are real values (i.e. between 50 and 100), e were converted to classes. AUC for threshold 65, 75 and 90 were computed, according to data provider suggestions.

Final Ranking

The final ranking of a submission i was calculated using KCC, RMSE, AUC (threshold = 75) and PWSD (fixed standard deviation of 10%) where R_i is the score based on which the final ranking was

defined :

$$R_i = RANK(KCC_i) + RANK(RMSE_i) + RANK(AUC_i) + RANK(PWSD_i)$$

The lower the number, the better the ranking of the submitter. These parameters were used because they provide different rankings. Statistical significance (with student t-test) of the ranking value was also computed in order to detect predictors with similar performances. The code used for assessment is available upon request to the corresponding author.

Predictor method descriptions

Participants used a wide range of approaches to generate predictions. A short description of each group's method follows.

Bromberg laboratory (submissions 2, 11, 17)

Predictions of change in proliferation rate were made using SNAP (Screening for Non-Acceptable Polymorphisms) (Bromberg and Rost, 2007; Hecht et al., 2015), a neural network-based method for evaluation of functional effects of single amino acid substitutions. For each substitution, SNAP produces a score, computed as the mean prediction of ten neural nets trained on different folds of SNAP cross-validation. SNAP scores ≤ 0 indicate no change in protein function (neutral), while SNAP scores > 0 indicate functional effect (non-neutral). Although SNAP scores do not map directly to percentage change in proliferation rate, we have previously observed (Bromberg et al., 2013) that increased scores correlate with the severity of protein function change. Here, we used the scores as reference and made predictions of change in proliferation rate in three different ways described below. Note that for submission, standard deviation values were interpreted in a non-statistical fashion.

- A) For neutral predictions we report no change, proliferation rate=0.5. For non-neutral predictions SNAP scores ≥ 60 indicated proliferation=1.0. SNAP scores in 0 to 60 range were normalized to represent the 0.5 to 1.0 range in proliferation rate. “Standard deviation” was pre-set at 0.05
- B) Predictions computed as in (A), but “standard deviation” was additionally computed from the scores of each of the ten SNAP-component networks
- C) Predictions estimated as: SNAP score $\geq 37 \rightarrow$ 1.0 proliferation rate (severe effect), score $\geq 27 \rightarrow$ 0.8 (moderate effect), and score $\geq 16 \rightarrow$ 0.6 (mild effect). “Standard deviation” was 0.01 for SNAP score ≥ 60 , 0.15 for score ≤ 0 , and the intermediates normalized within this range.

BioFold laboratory (submissions 5, 12, 18)

SNPs&GO is a machine-learning tool implementing a Support Vector Machine algorithm (Calabrese et al., 2009). This method predicts the impact of nonsynonymous Single Nucleotide Variants (nsSNVs) taking in input ~50 features derived protein sequence/profile, calculated using BLAST algorithm (Altschul et al., 1997), and functional information encoded by the Gene Ontology terms. SNPs&GO returns in output a score that represents the probability of each nsSNV to be pathogenic. The last version of SNPs&GO (Capriotti et al., 2013; Capriotti and Altman, 2011), maintained by the BioFold Unit (<http://snps.biofold.org/snps-and-go>), has been trained and tested on a set ~38,000 missense mutations extracted from SwissVar database (<http://swissvar.expasy.org/>). When tested in cross-validation on this dataset, SNPs&GO achieved an overall accuracy of ~81% and an AUC of 0.88. In previous independent benchmark tests SNPs&GO resulted among the best methods for predicting deleterious nsSNVs (Thusberg et al., 2011).

Casadio laboratory (submission 3)

The effect of the P16 variants has been predicted with SNPs&GO (Calabrese et al., 2009). All the variants in the training set (18 variations) and in the testing set (10 variations) are classified with SNPs&GO as correlated to disease. Each prediction of SNPs&GO is supplemented with a reliability index (RI) ranging from 0 to 10. The rate of proliferation for each variant has been set to be proportional to the RI (proliferation = $0.1 \cdot \text{RI}$). On the 18 variation of the training set, the prediction resulted in a 0.76 Pearson's correlation index ($p = 0.00023$).

Dunbrack laboratory (submission 7)

For predicting the phenotypes of missense mutations, we have extensively explored the utility of information derived from biological assemblies that are present in protein crystal structures. For more than 50% of structures, the biological assembly is actually different from the asymmetric unit that crystallographers use to model the experimental crystallographic data. For structures with more than one protein in the biological assembly (whether the same sequence or different), we find the accessible surface area from biological assemblies provides a statistically significant improvement in prediction over the accessible surface area of monomers from protein crystal structures. We developed support vector machine models, trained on sequence-based features such as the difference between wildtype and mutant position-specific profile scores, conservation scores, and disorder predictions, and structural features such as the solvent-accessible surface area of the wildtype residue in all available biological assemblies of the target protein. We applied these support vector machines to several CAGI targets using both homo- and heterooligomeric structures and our program BAM (BioAssemblyModeler), which models the three-dimensional structures of proteins and protein complexes based on biological assemblies deposited in the PDB. In a small number of cases, the oligomeric structure changed a prediction from neutral to deleterious, because the surface area of the residue in the complex was different from the monomer of the same protein. In the case of p16, the experimental structure of the

protein in a hetero-tetramer with CDK6 (PDB entry 1BI7) was used as a source of structural information.

Gough laboratory (submissions 8, 14, 19)

The Gough Group approach relies on Functional Annotation Through Hidden Markov Models (FATHMM) developed by Hashem Shihab (Shihab et al., 2013).

The JackHMMER program from the HMMER3 package was ran to search for p16 homologous sequences within the UniRef database (Suzek et al., 2007) and generate a profile HMM from the multiple sequence alignment (Eddy, 1998). The probabilities of the wild-type and mutant amino-acids were extracted from the Dirichlet mixtures (Sjölander et al., 1996) describing the profile HMMs and used to determine the score for each missense variant:

$$score = \ln \frac{P_m/(1 - P_m)}{P_w/(1 - P_w)}$$

where P_w and P_m are the respective wild-type and mutant probabilities in the match state.

Negative scores indicate a deleterious effect, while positive scores indicate a favourable substitution. In order to match the submission format, scores were mapped such that a FATHMM score of zero corresponds to a negative control baseline (50% proliferation rate), while the remaining scores were mapped to the 0-100% range by adding or subtracting the proportionate percentage with regard to the maximum absolute FATHMM score obtained.

Lichtarge laboratory (submissions 4)

The Evolutionary Action (EA) measures the fitness effect of coding mutations, analytically. It models evolution as a mapping of genotypes (γ) to phenotypes (φ) in the fitness landscape via an evolutionary function (f). Assuming genotypes are evolvable, f should be differentiable, such that we can follow

calculus to solve for the phenotype change due to a mutation: $d\phi = f'(\gamma) \cdot d\gamma$, where $d\phi$ is the action of the mutation $d\gamma$ on fitness and $f'(\gamma)$ is the sensitivity of the mutated site to genotype changes. To compute the action of any mutation, $f'(\gamma)$ is approximated with Evolutionary Trace (ET) ranks of importance (Lichtarge et al., 1996) and $d\gamma$ with inverse amino acid substitution log-odds. ET measures the importance of protein residues by accounting for the phylogenetic distances between homologous sequences ($d\phi$) that vary at a residue position ($d\gamma$). The substitution log-odds are the tendencies of amino acids to substitute one another in proteins and reflect the differences in various physicochemical properties of amino acids. We calculated these log-odds using data from numerous homologous sequence pairs, accounting for the different functional importance of the amino acids. The computed fitness change $d\phi$, or Evolutionary Action score, has been shown to correlate with experimental loss of function, clinical association, morbidity, and mortality (Katsonis and Lichtarge, 2014; Neskey et al., 2015). EA is available for non-profit use at <http://mammoth.bcm.tmc.edu/EvolutionaryAction>.

Moult laboratory (submissions 9, 15, 20)

We adopted an ensemble-like approach to predict relative cell growth rate for p16 mutations. For training, 30 cell proliferation assay (Ruas et al., 1999) data points for p16 missense mutations at one of 25 positions were assembled from the literature and from the information provided with the challenge. We used six methods to analyze the function impact of the missense mutations: SNPs3D stability method (Yue and Moult, 2006), SNPs3D sequence profile method (Yue and Moult, 2006), Polyphen-2 (Adzhubei et al., 2010), SIFT (Kumar et al., 2009), CHASM (Carter et al., 2009), and Condel (González-Pérez and López-Bigas, 2011). Two multivariate linear models were used to fit the training data. The input are scores of the six methods, together with the total numbers of deleterious predictions and of neutral predictions. The output is the relative cell growth rate. The first linear model was fitted to the raw data, and had an intercept. The second model was adjusted so that 50% growth rate

corresponded to the most neutral value for each method. Both models achieved an encouraging R-squared (0.61 and 0.91 respectively), as well as a significant P-value. Regression residual errors at a similar level were achieved by leave-one-out cross validation. We then predicted the relative cell growth rate for variants in the challenge set using the two models. Because the cell lines used in the training set and the challenge set are different, and we observed relatively high cell growth rates in our predictions on the challenge set, we added a third submission in which predicted values are reduced by a factor of 1.47. We used the residual standard errors of the corresponding models as the confidence of the predictions.

Yang & Zhou laboratory (submissions 10, 16, 21, 22)

Mutation free energy: The mutation free energy (ddG) was calculated using ROSETTA3 (Rohl et al., 2004) and DMUTANT (Zhou and Zhou, 2002) with default option, respectively. The option for ROSETTA3 is "-ddg::iterations 20 -ddg::dump_pdbs true -ddg::local_opt_only true -ddg::min_cst false -ddg::mean true -ddg::min false -ddg::sc_min_only true -in:file:fullatom".

Evolution term (dPSSM): PSI-BLAST (Altschul et al., 1997) was employed to obtain the PSSM for a given wild-type sequence. The evolution preference value was defined as the difference of the PSSM values between the wild-type residue and the mutated residue.

SVM: The server was trained based on the above three features with 19 supplied experimental value by using libsvm (Chang and Lin, 2011).

Results

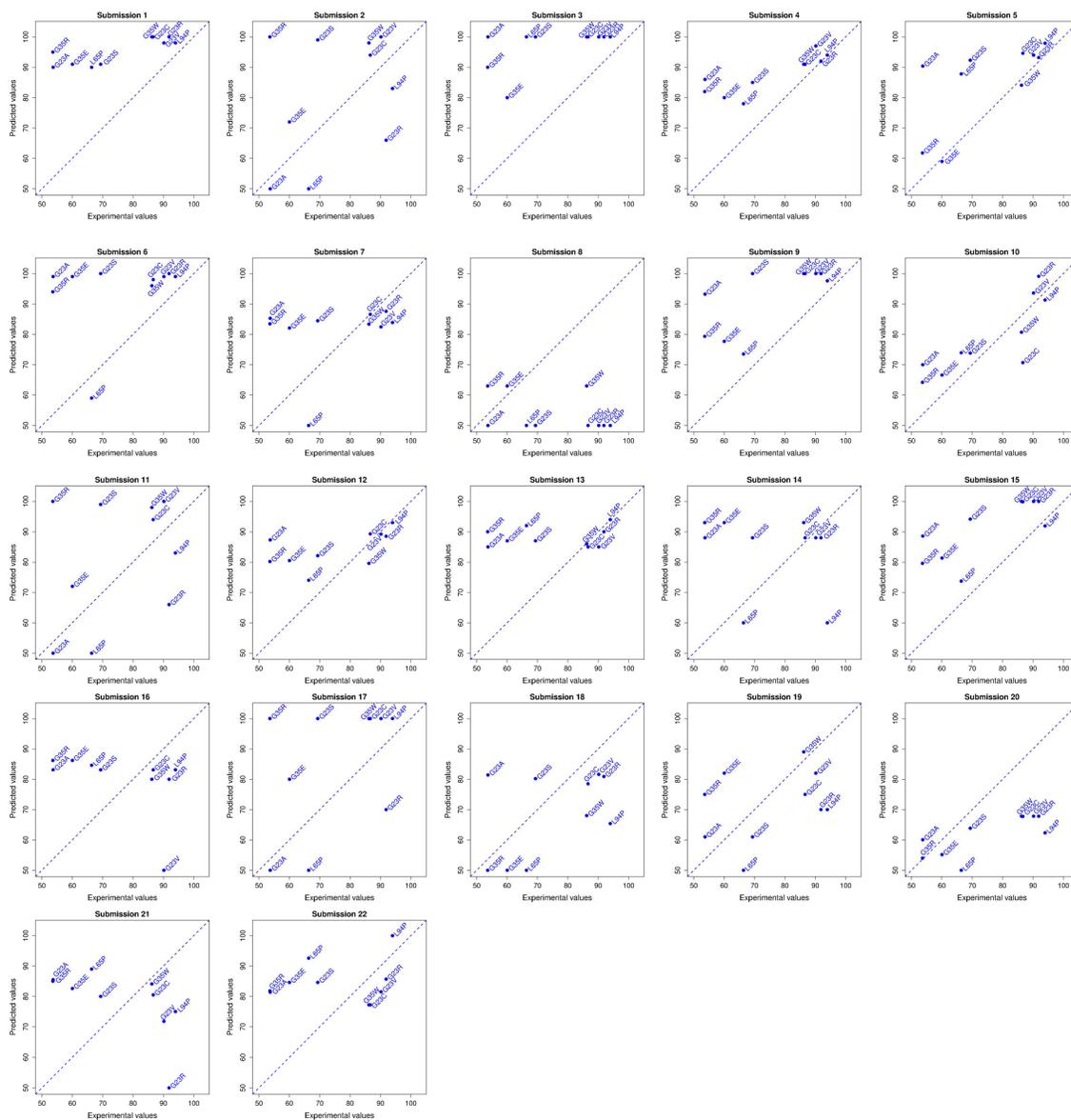
The submitted results include four groups of scaled data according to SVM output, dPSSM, ddG from ROSETTA3 and DMUTANTS. The Pearson correlation coefficients on the 10 test values are 0.83, -0.45, -0.61 and 0.15, respectively. The negative signs in the middle were caused by a sign error. Further analysis shows that the correlation coefficient for SVM is 0.937 after removing one outlier point.

Vihinen laboratory (submissions 6, 13)

The group used two methods for the predictions, PON-P (Olatubosun et al., 2012) and PON-P2 (Niroula et al., 2015). PON-P is a random forest-based metapredictor that utilizes results from four tolerance predictors (PhD-SNP, SIFT, PolyPhen-2, and SNAP) and a stability effect predictor (I-Mutant). PON-P was trained and tested on VariBench datasets (Sasidharan Nair and Vihinen, 2013) according to the established guidelines (Vihinen 2012, 2013). By bootstrapping, totally 200 random forest (RF) predictors were trained. Reliability score was computed based on the results of the RF predictors and the variants with a high reliability score were classified as pathogenic or neutral and those with a lower reliability as unknown.

PON-P2 (Niroula et al., 2015) was developed to avoid the bottlenecks caused by the third party software that drastically reduced the speed of PON-P. PON-P2 does not utilize data from any other predictors, instead has features describing evolutionary conservation, biochemical properties of amino acids, Gene Ontology (GO) annotations and functional and structural annotations of variant sites. The method was trained and tested with datasets from VariBench. 200 RF predictors were trained using bootstrap training data and the method classifies the variants into pathogenic, neutral and unknown, similar to PON-P. A probabilistic approach was used to integrate information for functional and structural annotations of the variant site together with the RF predictions to obtain the final prediction.

Supplementary Figures



Supplementary Figure 1. Predicted vs. experimental values for all 22 submissions. The predicted value (y-axis) is plotted against the experimental value (x-axis) for all variants in each of the 22 submissions.

Supplementary Tables

Nucleotide variant	Protein variant	Proliferation rate
c.68G>A	p.Gly23Asp	1.00
c.71G>A	p.Arg24Gln	0.50
c.104G>C	p.Gly35Ala	0.80
c.104G>T	p.Gly35Val	0.90
c.170C>T	p.Ala57Val	0.50
c.179C>T	p.Ala60Val	0.90
c.178_179delinsCG	p.Ala60Arg	1.00
c.192_194dup	p.Leu65dup	1.00
c.199G>C	p.Gly67Arg	0.60
c.206A>G	p.Glu69Gly	0.70
c.220G>T	p.Asp74Tyr	1.00
c.229A>C	p.Thr77Pro	1.00
c.239G>C	p.Arg80Pro	0.90
c.241C>A	p.Pro81Thr	1.00
c.259C>T	p.Arg87Trp	0.70
c.290T>G	p.Leu97Arg	1.00
c.296G>C	p.Arg99Pro	1.00
c.301G>T	p.Gly101Trp	1.00
c.340C>T	p.Pro114Ser	0.90

Supplementary Table 1. p16INK4a proliferation rate training set. Variants are shown by their nucleotide and protein identifier, followed by the relative proliferation level. Proliferation levels were rescaled between 0.5 (wildtype-like) and 1 (tumor-like). The standard deviation in this training set is in the range 0.05 - 0.15 and not shown.

	G23S	G23R	G23C	G23A	G23V	G35R	G35W	G35E	L65P	L94P
S1	0	1	0	0	1	0	0	0	0	1
S2	0	0	1	1	0	0	0	0	0	0
S3	0	1	0	0	0	0	0	0	0	1
S4	0	1	1	0	1	0	1	0	0	1
S5	0	1	1	0	1	0	1	1	0	1
S6	0	1	0	0	0	0	0	0	1	1
S7	0	1	1	0	1	0	1	0	0	0
S8	0	0	0	0	0	0	0	1	0	0
S9	0	1	0	0	0	0	0	0	1	1
S10	1	1	0	0	1	0	1	1	1	1
S11	0	0	1	1	0	0	0	0	0	0
S12	0	1	1	0	1	0	1	0	1	1
S13	0	1	1	0	1	0	1	0	0	1
S14	0	1	1	0	1	0	1	0	1	0
S15	0	1	0	0	0	0	0	0	1	1
S16	0	0	1	0	0	0	1	0	0	0
S17	0	0	0	1	0	0	0	0	0	1
S18	0	0	1	0	0	1	0	0	0	0
S19	0	0	0	1	1	0	1	0	0	0
S20	1	0	0	1	0	1	0	1	0	0
S21	0	0	1	0	0	0	1	0	0	0
S22	0	1	1	0	0	0	0	0	0	1
Total	2	13	12	5	9	2	10	4	6	12

Supplementary Table 2. Correct predictions per variant. Submissions are shown as rows, followed by the total count of correct predictions. Columns list each variant of the p16INK4a challenge and whether the corresponding submission correctly predicted (1, grey background) the effect according to PWSD10. Notice how certain substitutions at the same position were more difficult to predict.

Submission	Sens. 65	Spec. 65	B. Acc. 65	Sens. 75	Spec. 75	B. Acc. 75	Sens. 90	Spec. 90	B. Acc. 90
S1	1	0	0.5	1	0	0.5	1	0.29	0.64
S2	0.86	0.33	0.6	0.8	0.6	0.7	0.33	0.43	0.38
S3	1	0	0.5	1	0	0.5	1	0.29	0.64
S4	1	0	0.5	1	0	0.5	1	0.71	0.86
S5	1	0.67	0.83	1	0.4	0.7	1	0.57	0.79
S6	0.86	0	0.43	1	0.2	0.6	1	0.14	0.57
S7	0.86	0	0.43	1	0.2	0.6	0	1	0.5
S8	0	1	0.5	0	1	0.5	0	1	0.5
S9	1	0	0.5	1	0.2	0.6	1	0.43	0.71
S10	1	0.33	0.67	0.8	1	0.9	1	1	1
S11	0.86	0.33	0.6	0.8	0.6	0.7	0.33	0.43	0.38
S12	1	0	0.5	1	0.2	0.6	0.33	1	0.67
S13	1	0	0.5	1	0	0.5	0.33	0.86	0.6
S14	0.71	0	0.36	0.8	0.2	0.5	0	0.57	0.29
S15	1	0	0.5	1	0.2	0.6	1	0.57	0.79
S16	0.86	0	0.43	0.8	0	0.4	0	1	0.5
S17	0.86	0.33	0.6	0.8	0.4	0.6	0.67	0.43	0.55
S18	0.86	0.67	0.76	0.6	0.6	0.6	0	1	0.5
S19	0.71	0.33	0.52	0.4	0.8	0.6	0	1	0.5
S20	0.57	1	0.79	0	1	0.5	0	1	0.5
S21	0.86	0	0.43	0.4	0	0.2	0	1	0.5
S22	1	0	0.5	1	0	0.5	0.33	0.86	0.6

Supplementary Table 3. Complementary performance indices. Complementary performance measures are reported for three different levels of binary classification (Thresholds for proliferation levels: 65, 75, and 90). For each threshold, three performance indices are reported: Sensitivity, Specificity and Balanced Accuracy.

References

- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi:10.1038/nmeth0410-248
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Bromberg, Y., Kahn, P.C., Rost, B., 2013. Neutral and weakly nonneutral sequence variants may define individuality. *Proc. Natl. Acad. Sci. U. S. A.* 110, 14255–14260. doi:10.1073/pnas.1216613110
- Bromberg, Y., Rost, B., 2007. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* 35, 3823–3835. doi:10.1093/nar/gkm238
- Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P.L., Casadio, R., 2009. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.* 30,

1237–1244. doi:10.1002/humu.21047

- Capriotti, E., Altman, R.B., 2011. Improving the prediction of disease-related variants using protein three-dimensional structure. *BMC Bioinformatics* 12, 1–11. doi:10.1186/1471-2105-12-S4-S3
- 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, 1–7. doi:10.1186/1471-2164-14-S3-S6
- Carter, H., Chen, S., Isik, L., Tyekucheva, S., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Karchin, R., 2009. Cancer-Specific High-Throughput Annotation of Somatic Mutations: Computational Prediction of Driver Missense Mutations. *Cancer Res.* 69, 6660–6667. doi:10.1158/0008-5472.CAN-09-1133
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A Library for Support Vector Machines. *ACM Trans Intell Syst Technol* 2, 27:1–27:27. doi:10.1145/1961189.1961199
- Eddy, S.R., 1998. Profile hidden Markov models. *Bioinformatics* 14, 755–763. doi:10.1093/bioinformatics/14.9.755
- González-Pérez, A., López-Bigas, N., 2011. Improving the Assessment of the Outcome of Nonsynonymous SNVs with a Consensus Deleteriousness Score, *Condel. Am. J. Hum. Genet.* 88, 440–449. doi:10.1016/j.ajhg.2011.03.004
- Hecht, M., Bromberg, Y., Rost, B., 2015. Better prediction of functional effects for sequence variants. *BMC Genomics* 16, 1–12. doi:10.1186/1471-2164-16-S8-S1
- Katsonis, P., Lichtarge, O., 2014. A formal perturbation equation between genotype and phenotype determines the evolutionary action of protein coding variations on fitness. *Genome Res.* gr.176214.114. doi:10.1101/gr.176214.114
- Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1081. doi:10.1038/nprot.2009.86
- Lichtarge, O., Bourne, H.R., Cohen, F.E., 1996. An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* 257, 342–358. doi:10.1006/jmbi.1996.0167
- Neskey, D.M., Osman, A.A., Ow, T.J., Katsonis, P., McDonald, T., Hicks, S.C., Hsu, T.-K., Pickering, C.R., Ward, A., Patel, A., Yordy, J.S., Skinner, H.D., Giri, U., Sano, D., Story, M.D., Beadle, B.M., El-Naggar, A.K., Kies, M.S., William, W.N., Caulin, C., Frederick, M., Kimmel, M., Myers, J.N., Lichtarge, O., 2015. Evolutionary Action Score of TP53 Identifies High-Risk Mutations Associated with Decreased Survival and Increased Distant Metastases in Head and Neck Cancer. *Cancer Res.* 75, 1527–1536. doi:10.1158/0008-5472.CAN-14-2735
- Niroula, A., Urolagin, S., Vihinen, M., 2015. PON-P2: Prediction Method for Fast and Reliable Identification of Harmful Variants. *PLOS ONE* 10, e0117380. doi:10.1371/journal.pone.0117380
- Olatubosun, A., Väliäho, J., Härkönen, J., Thusberg, J., Vihinen, M., 2012. PON-P: integrated predictor for pathogenicity of missense variants. *Hum. Mutat.* 33, 1166–1174. doi:10.1002/humu.22102
- Rohl, C.A., Strauss, C.E.M., Misura, K.M.S., Baker, D., 2004. Protein structure prediction using Rosetta. *Methods Enzymol.* 383, 66–93. doi:10.1016/S0076-6879(04)83004-0
- Ruas, M., Brookes, S., McDonald, N.Q., Peters, G., 1999. Functional evaluation of tumour-specific variants of p16INK4a/CDKN2A: correlation with protein structure information. *Oncogene* 18, 5423–5434. doi:10.1038/sj.onc.1202918
- Sasidharan Nair, P., Vihinen, M., 2013. VariBench: a benchmark database for variations. *Hum. Mutat.* 34, 42–49. doi:10.1002/humu.22204
- Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L.A., Edwards, K.J., Day, I.N.M., Gaunt, T.R., 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* 34, 57–65. doi:10.1002/humu.22225

- Sjölander, K., Karplus, K., Brown, M., Hughey, R., Krogh, A., Mian, I.S., Haussler, D., 1996. Dirichlet mixtures: a method for improved detection of weak but significant protein sequence homology. *Comput. Appl. Biosci. CABIOS* 12, 327–345. doi:10.1093/bioinformatics/12.4.327
- Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., Wu, C.H., 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinforma. Oxf. Engl.* 23, 1282–1288. doi:10.1093/bioinformatics/btm098
- Thusberg, J., Olatubosun, A., Vihinen, M., 2011. Performance of mutation pathogenicity prediction methods on missense variants. *Hum. Mutat.* 32, 358–368. doi:10.1002/humu.21445
- Yue, P., Moulton, J., 2006. Identification and analysis of deleterious human SNPs. *J. Mol. Biol.* 356, 1263–1274. doi:10.1016/j.jmb.2005.12.025
- 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 Sci. Publ. Protein Soc.* 11, 2714–2726. doi:10.1110/ps.0217002