

Predicting the Effect of Variants of Unknown Significance in Molecular Tumor Boards with the VUS-Predict Pipeline

Vanessa SCHLOTZIG^{a,*}, Kevin KORNRUMPF^{a,*}, Alexander KÖNIG^b,
Tim TUCHOLSKI^a, Jonas HÜGEL^{c,d}, Tobias R. OVERBECK^e, Tim BEISSBARTH^{a,d},
Raphael KOCH^e and Jürgen DÖNITZ^{a,1}

^aDept. of Medical Bioinformatics, University Medical Center Göttingen, Germany

^bDept. of Gastroenterology, gastrointestinal Oncology and Endocrinology,
University Medical Center Göttingen, Germany

^cDept. of Medical Informatics, University Medical Center Göttingen, Germany

^dCampus Institute Data Science (CIDAS), Göttingen, Germany

^eDept. of Hematology and Medical Oncology,
University Medical Center Göttingen, Germany

Abstract. Precision oncology utilizing molecular biomarkers for targeted therapies is one of the hopes to treat cancer. The availability of patient specific molecular profiling through next-generation sequencing, though, increases the amount of available data per patient to an extent that computational support is required to identify potential driver alterations for targeted therapies and rational decision-making in molecular tumor boards (MTBs). For some genetic variants evidence-based drug recommendations are available in public databases, but for the majority, the variants of unknown significance (VUS), this clinical information is missing. Additionally, for most of these variants no information about the functional impact on the protein is accessible. To acquire maximal functional evidence for VUS, the VUS-Predict pipeline collects estimations about the effect of a VUS by integrating multiple pre-existing tools. Pre-existing tools implement different approaches for their predictions, which are summarized by our newly developed tool with a common score and classification in neutral or deleterious variants. The primary tools are chosen based on their sensitivity and specificity on well-known variants of the transcription factor TP53. Resulting negative and positive predictive values are used to calibrate the VUS-Predict pipeline. Further, the pipeline is evaluated using data from public cancer databases and cases of the MTB in Göttingen, both also in comparison with the ensemble method REVEL. The results show that VUS-Predict has clear advantages in a clinical setting due to clear and traceable predictions. In particular, VUS outperforms REVEL in the real-life setting of a MTB. Likewise, an evaluation on variants of public cancer databases confirms the good results of VUS-Predict and shows the need for a reliable gold standard and unambiguous results of the tools under test.

¹ Corresponding Author, Jürgen Dönitz, University Medical Center Göttingen (UMG), Goldschmidtstr. 1, 37077 Göttingen, Germany; E-mail: juergen.doenitz@bioinf.med.uni-goettingen.de. * equal contribution.

Keywords. variants of unknown significance, molecular tumor board, VUS-Predict pipeline

1. Introduction

Precision oncology depends on the determination and functional interpretation of patient specific biomarkers, e. g. gene variants, to identify specific therapies. While the molecular diagnosis is often established by sequencing pipelines for the whole genome or known oncogenes, the interpretation of the found gene variants is still challenging. Clinical studies elucidating the functional and clinical relevance are scarce and only exist for a small number of variants. To integrate genetic variants, cancer entity and drug recommendations, an evidence level can be assigned and this information is made available in public databases such as CIViC, GDKD, OncoKB [1-3] and others. For variants that are not covered by the public databases no evidence-based recommendation can be given. However, an estimation about the putative effect is the basis for further studies and might support establishing novel treatment approaches.

So far, several pre-existing tools utilize different approaches to predict the effect of a mutation, e.g., based on sequence or protein structure information. Each approach has its own characteristics resulting in specific sensitivity and specificity. E.g., Missense3D [4] depends on protein 3D structure and has a very high specificity but lower sensitivity due to limited availability of 3D structures. Collecting information for a *variant of unknown significance* (VUS) by integrating the results from different tools is useful and already done for example by Ensembl VEP (*Variant Effect Predictor*) [5] or REVEL [6]. VEP aims at a complete representation of the available tools and shows an overview page with all single results. A single summarizing result is not available and might also not be intended in this context because the tools are not selected based on their characteristics but completeness. REVEL, in contrast, implements a machine learning algorithm, based on Random Forests, and outputs a score for a queried variant [6]. However, while this approach might summarize and simplify the interpretation of genetic variants, it is not possible to trace back the decision and it is up to the user to decide if the score is already in the range of a pathogenic variant.

For a VUS prediction tool in a clinical setting, though, it is important that it collects data from multiple established tools with complementary and non-redundant features to provide a clear and transparent output.

With the VUS-Predict pipeline, we aim to create a tool as additional source of information to find clinical decisions for VUS. Our goal is to address the following areas: Based on their implemented approach, tools have been selected that cover different methods to determine the effect of mutations in a protein. The scores of the primary tools can be queried, together with an overall score and a label that predicts the effect of the variance to the function of the protein. The final score is calculated by taking the positive and negative predictive value of each tool into account which is calibrated on known variants of the well-studied transcription factor TP53. With this approach, the VUS-Predict pipeline is established and currently evaluated in a real-life setting of a molecular tumor board (MTB) at the *University Medical Center Göttingen* (UMG).

2. Results

2.1. *The VUS-Predict pipeline*

The VUS-Predict pipeline is designed as an integrative service of specialized tools to predict the effect of variants of unknown significance. Following a literature research, ten tools [4,7-15] have been chosen for an evaluation based on keyword search, references, expert knowledge and possibility of a local installation. Table 1 lists their implemented approach as well as their sensitivity and specificity regarding 100 well known variants of TP53, annotated by the International Agency for Research on Cancer (IARC). All considered variants were either non-functional or functional to receive unambiguous results. For this reason, partially functional and supertrans variants were not a part of this process due to a lack of information concerning the effect on the protein. In this evaluation, all non-functional TP53 variants represent deleterious mutations and functional variants neutral mutations. For the different approaches we selected one unique representative. The candidate in the three categories we selected based on good, but also balanced statistical performance. The selected methods and representative tools are: 1) Missense3D [4] as it conducts a structural analysis of the proteins' 3D structure to determine damages and with a nearly perfect precision and specificity. 2) FATHMM [10] which predicts the extent of sequence conservation by Hidden Markov Models to distinguish deleterious from neutral variants. It competes well against other conservations tools and provides a great sensitivity. 3) SIFT [15] since it takes into account sequence homology and the differences in physical characteristics through amino acid exchanges.

Table 1. Summary of the characteristics of the tools that have been evaluated as candidates for the VUS-Predict pipeline.

Tool	Approach	Accuracy	Precision	Sensitivity	Specificity
Align GVGD	conservation	0.70	0.94	0.62	0.89
BLOSUM62	similarity	0.53	0.74	0.54	0.50
CADD	multiple sources	0.80	0.81	0.94	0.43
FATHMM	conservation	0.85	0.84	0.99	0.50
HOPE	multiple sources	0.80	0.82	0.92	0.47
Missense3D	3D structure	0.44	1.00	0.26	1.00
MASS	conservation	0.80	0.83	0.90	0.54
MutationTaster2	conservation	0.78	0.81	0.90	0.46
PolyPhen-2	several features	0.80	0.86	0.86	0.64
SIFT	homology, amino acid characteristics	0.82	0.81	0.99	0.39

To investigate the impact on the overall score for each selected tool, again annotated TP53 variants from the IARC database were used to calibrate the pipeline. For this calibration process 324 missense variants were considered, 288 damaging and 36 neutral. With the aforementioned classification of functional and non-functional mutations, all 324 missense variants were predicted by the previously selected tools. The results of these predictions are collected in a confusion matrix for each tool. As solely the amount of true positive predictions among all positive predictions mattered, and vice versa for the negative class, the positive and negative predictive values were chosen as metrics, PPV respectively NPV.

These values also form the basis for the calculation of the overall score for the VUS-Predict pipeline which ranges from -1 to +1. Thereby, a value smaller than 0 indicates a

deleterious, a value greater than 0 a neutral variant, and a score of 0 that no decision was possible. An indecisive score occurs if the requested variant cannot be found by any of the implemented tools. For variants, the overall score is calculated as a weighted average of the prediction tools’ results. Each weight reflects either the positive predictive value for a deleterious prediction or the negative predictive value for a neutral classification. In case of a deleterious prediction, the positive predictive value is multiplied by -1 and, in case of a neutral classification, the negative predictive value is multiplied by 1, see also (1).

$$f_i = \begin{cases} tool_{[i]} * PPV, & \text{if } tool_{[i]} < 0 \\ tool_{[i]} * NPV, & \text{if } tool_{[i]} \geq 0 \end{cases} \tag{1}$$

$$\frac{1}{n} \sum f_i$$

The comparison of the PPV and NPV between VUS-Predict and the included tools, FATHMM, SIFT and Missense3D, reveals that for the PPV it is in the last place but all tools have comparable and high numbers (Figure 1A). The case for the NPV differs, the tools differ a lot and the VUS-Predict pipeline has the best result, followed by FATHMM.

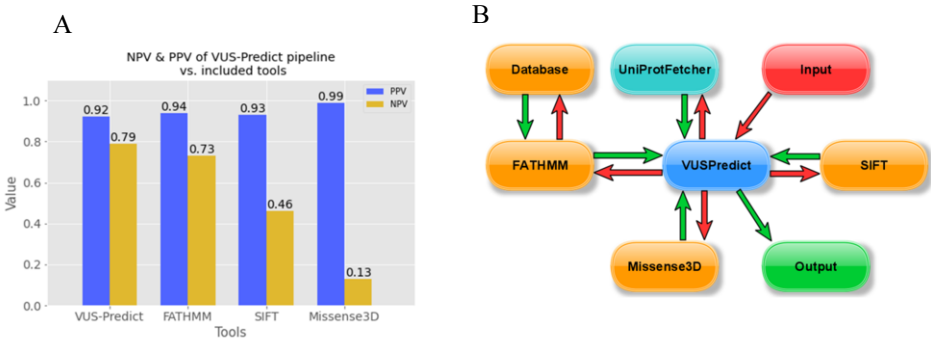


Figure 1. The left image is an overview of the PPV and NPV for each single tool used in the pipeline and the scores of VUS-Predict itself. The right image shows an overview of the VUS-Predict pipeline and how it is connected with other tools.

2.2. *Evaluation of the VUS-Predict pipeline on public available data sources of evidence based recombinations and comparison with REVEL*

To estimate the prediction performance of the VUS-Predict pipeline, we used known and evidence based variant drug combinations from public cancer databases. We extracted the single nucleotide variants (SNVs) from CIViC, GDC and OncoKB and used them as input for the VUS-Predict pipeline. In addition, we ran REVEL on the same input set and plotted the results in Figure 2. The color of the bars represents the outcome of the VUS-Predict pipeline, where green refers to variants predicted as neutral, and the x-axis shows the score of REVEL. The output from REVEL is a continuous score between 0 and 1, where higher scores are more likely to be pathogenic. Missing a distinct threshold, the performance of REVEL as a number of neutral or pathogenic predictions cannot be

given. Instead, the number of scores in bins of 0.1 are plotted. Overall, out of 470 SNVs 78 (17%) are predicted to be neutral (47% Table 2). A NPV cannot be determined as only variants that can be used as indicators for a specific treatment are included in the cancer databases.

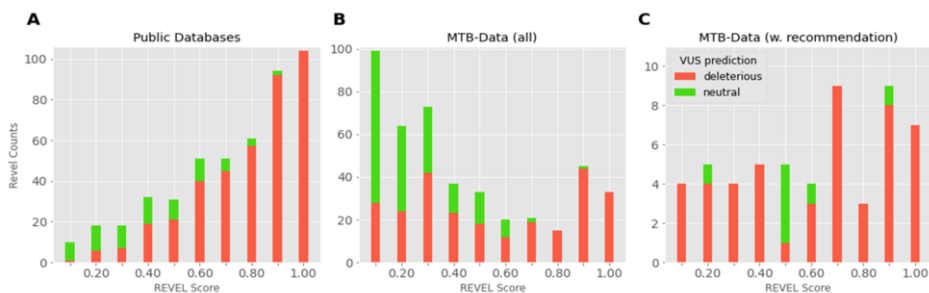


Figure 2. Evaluation of VUS-Predict and REVEL on the public databases CIViC, GDKG and OncoKB. The height of the bars represents the frequency of REVEL scores, the color indicates the prediction of VUS-Predict.

2.3. Evaluation in a molecular tumor board

Of particular interest is the performance of the VUS-Predict pipeline in a real MTB and the question how the interpretation of VUS can be used in this setting. For that 82 patient data sets from recent 14 tumor board sessions have been assessed with VUS-Predict in retrospective and compared with the recommendations of the MTB. In these 82 patients, a total of 862 variants were identified, of which we obtained 440 variants where VUS-Predict and REVEL returned a prediction (Table 2). For 175 of these variants there is an agreement between VUS-Predict and the MTB in the sense that the prediction is neutral and no recommendation was based on this variant. 210 variants on 96 genes were marked as “deleterious” but had no subsequent recommendation. Notably, this number also contains those cases where a variant is recognized as pathogenic but cannot be targeted or the therapy would not be suitable for the patient. Indeed, the affected gene in 52 of these cases was either TP53 or KRAS. Both are common oncogenes but can only be targeted in rare cases. Automatic rating of variants should provide valuable information as well as point to possible targets. VUS-Predict is not a substitute for decisions made by clinical experts with the knowledge of the complete patient’s history. This is also obvious in 7 (out of 55) cases where VUS-Predict output was “neutral” but was assigned to a recommendation in the MTB. In two cases the statement of the pipeline can be questioned. With a very low score of 0.07, SIFT and FATHMM counted for neutral and overruled Missense3D. In a case of a variant in LRB1B, a high tumor mutational burden value (TMB) was present and this combination was the reason for the recommended therapy. The rationale for the remaining cases are a well-known driver gene (ERBB2), expert knowledge and patient history.

3. Material and Methods

The VUS-Predict pipeline is implemented as a RESTful service to serve as a stand-alone information resource or to accompany tools for a MTB [16]. The pipeline itself also has a modular architecture. Each included tool is integrated by an adapter that encapsulates the characteristics of the tool and provides a common API to the VUS-

Predict service. The loose coupling allows easy extension with more tools and to choose the optimal framework and programming language for each tool. E. g. for the FATHMM adapter Python and the Flask framework was selected, because the origin project uses these settings and could be extended this way. For the other parts Java and Quarkus are used. The SIFT adapter is an example that queries itself a REST interface, in this case VEP. The other adapters use local data and methods. An overview of the architecture is shown in Figure 1B.

To start the pipeline, a HUGO gene symbol, amino acid exchange, and genomic location need to be provided. In the cases where the genomic location was missing, it was retrieved by a self-implemented coordinate converter service. Afterwards, the pipeline requests the UniProt API for the corresponding gene identifier since Missense3D and FATHMM require a UniProtID for a successful request. The UniProtIDs are filtered by HUGO gene symbol, organism, and status to prefer reviewed entries.

Next, Missense3D, FATHMM, and SIFT are queried. Both, Missense3D and FATHMM require a UniProtID and the protein description of the variant, while SIFT can be queried by providing the genomic description of the mutation. From these returned prediction results, the overall score is calculated as described above and returned to the user.

Accuracy, precision, sensitivity and specificity were calculated with the standard formulas for a binary classifier. True positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) were used as input. The accuracy can be evaluated with $(TP + TN) / (TP + FP + TN + FN)$, the precision with $TP / (TP + FP)$, the sensitivity with $TP / (TP + FN)$ and the specificity with $TN / (TN + FP)$.

The patient data used for the evaluation has been extracted from the clinical systems and covers the cases from the MTB of the University Medical Center Göttingen in the time period March 2020 to March 2021. The data were filtered, analyzed and evaluated in Python v3.9 using the NumPy v1.20.2 and pandas v1.2.3 packages. The graphics were created with the matplotlib v3.4.1 library.

Table 2. The data from the MTB and public databases are preprocessed and used as input for VUS-Predict and REVEL. The table shows the number of data used as input and the output of the two classification tools.

	Evaluation of VUS-Predict based on the MTB	Evaluation of VUS-Predict based on public DB
Number of Patients	82	-
Single Nucleotide Variants (SNVs)	862	1,945
SNVs with VUS-Predict & REVEL returning a prediction	440	470
VUS-Predict prediction: deleterious	258	392
VUS-Predict prediction: deleterious with MTB-recommendation	48	-
VUS-Predict prediction: deleterious, but not used for a recommendation	210	-
VUS-Predict prediction: neutral	182	78
VUS-Predict prediction: neutral without recommendation	175	-
VUS-Predict prediction: neutral with MTB-recommendation	7	-

4. Discussion

Molecular tumor boards are an emerging option to identify targeted treatment options for patients where established therapies failed. For these MTB, though, new processes have to be established and growing amounts of sequence data have to be filtered and investigated. To this end, computational support is urgently needed to filter and classify data as well as to annotate the data with additional information, e.g., references to clinical studies. The pathogenicity of a gene variant is a basic and important information, that are estimated by a diversity of tools and some meta tools combining those. VUS-Predict implements a new approach that combines the traceability of e.g. VEP with an overall score known from Machine Learning based tools like REVEL. While VUS-Predict includes only a few tools, the method to take the NPV and PPV into account leads to a convincing prediction performance.

The VUS-Predict pipeline proved to be a useful tool in our MTB to provide information about variants of unknown significance. The architecture allows easy use and interpretation of the results. More tools can be integrated to cover more approaches to determine the effect of VUS or to improve the performance of VUS-Predict. This gets obvious in the two cases in the evaluation in the MTB where Missense3D was overruled. The design of VUS-Predict taking the NPV and PPV into account already prefers the precision of Missense3D but fails short with a score of 0.07, where it gets outvoted by the predictions of SIFT and FATHMM. More examples for calibration or an additional boost value could help here. Additional primary tools might help filling gaps; however, it is important not to bias the score by over-representing one category of tools. A two-step method could be here interesting to first combine similar tools and then combine the intermediate scores of the categories into a final score. The reliable prediction of the effect of VUS is important if clinical studies report results rather on genes or pathways instead of on the individual genetic variants. In such cases the MTB has to decide, if the observed mutation might have an effect on the function of the protein and can be targeted by a drug. To support this decision-making in a rational and transparent manner, we here established the VUS-Predict pipeline. This pipeline can be used as a tool for molecular tumor boards, as exemplified for the MTB in Göttingen, mainly to filter out gene variants that are neutral for the protein and therefore not a reasonable target for therapy.

Notably, applying our VUS-Predict pipeline to real world use cases, the pipeline performed even better than on the data of the public cancer databases (Figure 2A). This example might show how integration of genetic and clinical data and systematic filtering of genetic variants might inform robust interpretation of genetic alterations.

Furthermore, our newly developed VUS-Predict pipeline seems to outcompete alternative approaches like REVEL. However, a direct comparison of our VUS-Predict pipeline and REVEL is difficult, as REVEL uses a continuous score and no clear threshold. The comparison of the results of the two evaluations still shows that VUS-Predict can at least keep up with REVEL and has advantages in the use of MTBs compared with REVEL and VEP.

Declarations

Ethical vote: Ethic commission of UMG, J. Brockmüller, 27/8/20, 3. 9. 2020

Conflict of Interest: The authors declare that there is no conflict of interest.

Author contributions: VS, RK, TB, JD: conception and design of the work; KK, VS, JD: data analysis; VS, TT: implementation of coordinate converter; JH: data acquisition; AK, TRO, RK: medical evaluation; JD, VS, KK, RK, TB, JH: writing the manuscript; All authors approved the manuscript in the submitted version and take responsibility for the scientific integrity of the work.

Acknowledgement: This work was supported by the Volkswagen Foundation within research project MTB-Report (ZN3424) and by the German Federal Ministry of Education and Research (BMBF) within the MyPathSem project (031L0024A). Furthermore, RK is supported by Deutsche Krebshilfe (70113602)

References

- [1] Griffith M, Spies NC, Krysiak K et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nature Genetics*. **49** (2017) 170–174.
- [2] Dienstmann R, Jang IS, Bot B et al. Database of Genomic Biomarkers for Cancer Drugs and Clinical Targetability in Solid Tumors, *Cancer Discov*. **5** (2015) 118–123. doi:10.1158/2159-8290.CD-14-1118
- [3] Chakravarty D, Gao J, Phillips S et al. OncoKB: A Precision Oncology Knowledge Base, *JCO Precision Oncology*. (2017) 1–16. doi:10.1200/PO.17.00011
- [4] Ittisoponpisan S, Islam SA, Khanna T et al. Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated?, *Journal of Molecular Biology*. **431** (2019) 2197–2212.
- [5] McLaren W, Gil L, Hunt SE et al. The Ensembl Variant Effect Predictor, *Genome Biology*. **17** (2016) 122. doi:10.1186/s13059-016-0974-4
- [6] Ioannidis NM, Rothstein JH, Pejaver V et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants, *Am J Hum Genet*. **99** (2016) 877–885.
- [7] Tavtigian SV, Byrnes GB, Goldgar DE et al. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications, *Hum. Mutat*. **29** (2008) 1342–1354.
- [8] Glusman G, Rose PW, Prlić A et al. Mapping genetic variations to three-dimensional protein structures to enhance variant interpretation: a proposed framework, *Genome Medicine*. **9** (2017) 113. doi:10.1186/s13073-017-0509-y
- [9] Rentzsch P, Witten D, Cooper GM et al. CADD: predicting the deleteriousness of variants throughout the human genome, *Nucleic Acids Research*. **47** (2019) D886–D894. doi:10.1093/nar/gky1016
- [10] Shihab HA, Gough J, Cooper DN et al. Predicting the functional consequences of cancer-associated amino acid substitutions, *Bioinformatics*. **29** (2013) 1504–1510. doi:10.1093/bioinformatics/btt182
- [11] Venselaar H, te Beek TA, Kuipers RK et al. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces, *BMC Bioinformatics*. **11** (2010) 548. doi:10.1186/1471-2105-11-548
- [12] Reva B, Antipin Y, and Sander C. Predicting the functional impact of protein mutations: application to cancer genomics, *Nucleic Acids Res*. **39** (2011) e118. doi:10.1093/nar/gkr407
- [13] Schwarz JM, Cooper DN, Schuelke M et al. MutationTaster2: mutation prediction for the deep-sequencing age, *Nat Methods*. **11** (2014) 361–362. doi:10.1038/nmeth.2890
- [14] Adzhubei LA, Schmidt S, Peshkin L et al. A method and server for predicting damaging missense mutations, *Nat Methods*. **7** (2010) 248–249. doi:10.1038/nmeth0410-248
- [15] Vaser R, Adusumalli S, Leng SN et al. SIFT missense predictions for genomes, *Nature Protocols*. **11** (2016) 1–9. doi:10.1038/nprot.2015.123
- [16] Perera-Bel J, Hutter B, Heining et al. From somatic variants towards precision oncology: Evidence-driven reporting of treatment options in molecular tumor boards. *Genome Med*. 2018 Mar 15;10(1):18.