Regression Modeling and Meta-Analysis of Diagnostic Accuracy of SNP-Based Pathogenicity Detection Tools for UGT1A1 Gene Mutation

Fakher Rahim,1 Hamid Galehdari,2 Javad Mohammadi-asl,3 and Najmaldin Saki4

1 Golestan Blv. Toxicology Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
2 Genetic Department, Faculty of Science, Shahid Chamran University, Ahvaz, Iran
3 Department of Medical Genetics, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
4 Research Center of Thalassemia & Hemoglobinopathy, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

Correspondence should be addressed to Fakher Rahim; bioinfo2003@ajums.ac.ir

Received 11 May 2013; Revised 30 June 2013; Accepted 12 July 2013

1. Introduction

Crigler-Najjar syndrome (CNS) (MIM nos. 218800, 606785) type I and type II are inherited as autosomal recessive conditions that is resulted from mutations in the UGT1A1 gene (UGT1A1; MIM nos. 191740) [1–4]. Type I is characterized by almost complete absence of UGT1A1 enzyme activity, and these patients are refractory to phenobarbital treatment, while type II is a less severe form of deficiency [5, 6]. Patients with CNS are at permanent risk of developing severe neurologic complications such as hearing problems, mental retardation, and choreoathetosis due to severe unconjugated hyperbilirubinemia [7]. It is well known that UGT1A1 is expressed specifically in the liver and that it is difficult to perform an expression analysis directly on the patients by invasive liver biopsy but to state that the mutation causes inactivation of the enzyme you could perform an in vitro functional study by cloning the mutated CDNA of UGT1A1 in an expression vector. The constructs could be transfected in hepatic cell lines as HepG2 or HUH7. The expression analysis on these cells overexpressing the mutated forms of UGT1A1 will allow you to finally demonstrate the inactivation of the enzyme [8]. The UGT1A1 gene comprises five consecutive exons located on chromosome 2q37 by which complete or partial inactivation of any exon causes CNS. Single variations in deoxyribonucleic acid (DNA) base pairs responsible for protein, called coding which is single nucleotides polymorphism (SNP), leads to changes in amino acids that ultimately affect the protein structure and function. Different such types of SNPs include, missense mutations, nonsense, silent
mutations, and splice-site mutations. The majority of missense mutations leads to considerable variation in the protein structure and function, causing the disease symptoms. Data about nonsynonymous SNPs exists in public repositories such as SWISSPROT [9], dbSNP [10], and HGVBASE [11].

Genetic methods including the detection of genes linked to the disease phenotypes and the identification of aberrant functions of these genes have, in recent years, provided worthy understanding into the biological foundations of genetic mutation [12]. The present review summarized all available evidence on the accuracy of SNP-based pathogenicity detection tools and introducing regression model based on different scores including functional scores, mutation score, and genomic variation degree and compared the results to the published clinical result.

2. Materials and Methods

2.1. SNP Data Sources and Collection. An inclusive search was done to find all CNS-related mutations. The major data repositories, including HGMD, dbSNP, SNPdb, and Ensembl, were reviewed. All CNS-related mutations were extracted and double checked for duplicated queries and then tabulated (Table 1).

2.2. Inclusion Criteria. Only UGT1A1-gene-related missense mutations were included.

2.3. Exclusion Criteria. Other types of mutation such as synonymous or nonsense were excluded.

2.4. Data Extraction. The pathogenicity prediction was done using SNP-based detection tools including SIFT [13], PHD-SNP [14], PolyPhen2 [15], fathmm [16], Provean [17], and Mutpred [18]. Then a regression model was designed using functional scores, mutation score, and genomic variation degree. For each SNP-based pathogenicity detection tool and our regression model, we extracted a 2 × 2 table including positive prediction of the disease (True Positive, TP), negative prediction as neutral (true negative, TN), positive prediction in nondisease (false positive, FP), and negative prediction in disease (false negative, FN). In order to assess the phenotypic characterization and clinical features of the disease of interest, we searched databases, including SWISSPROT [9], dbSNP [10], Ensembl [19], OMIM [20], DECIPHER [21], and HGVBASE [11]. Furthermore, we compared the results of SNP-based pathogenicity detection tools with the results of phenotypic description tools. Then we calculated the diagnostic odds ratio (diagnostic OR), which is a single indicator of test performance and varies between 0 and infinity [22].

2.5. Statistical Analysis. All the analyses were done by SPSS 16.0. A regression model was designed using three categories, including functional score [23], structural score (GV, genomic variation score) [24], and conservation score [25]. Each SNP-based pathogenicity detection tool was compared by the reference values using logistic regression. The sensitivity (Sn), specificity (Sp), accuracy (ACC), diagnostic OR, and Matthew’s correlation coefficient (MCC) were calculated using the following formula:

\[
\text{Sensitivity (Sn)} = \frac{TP}{TP + FN},
\]

\[
\text{Specificity (Sp)} = \frac{TN}{TN + FP},
\]

\[
\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP},
\]

\[
\text{Diagnostic OR} = \frac{Sn/1 - Sn}{1 - Sp/Sp},
\]

\[
\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FN) (TP + TN) (FN + TN) (FN + FP)}}.
\]

The metadisk was used to calculate individual and pooled diagnostic OR, sensitivity, specificity, negative likelihood ratio, and positive likelihood ratio [26]. We also compared the AUC (area under curve), which is a popular index of the overall performance of a test, using the summary receiver operating characteristic (SROC) curve [27].

3. Results

Overall, 59 different SNPs related to missense mutations in the UGT1A1 gene were reviewed using the designed protocol (Figure 1). Our regression model was as \( y = 3.39 + (-0.24 \times \text{functional score}) + (-0.14 \times \text{GV score}) + (-2.44 \times \text{conservation score}) \). Comparing the diagnostic OR, our model showed high detection potential (diagnostic OR: 16.71, 95% CI: 3.38–82.69) (Figure 2). The highest MCC and ACC was belonged to our suggested model (46.8% and 73.3%, followed by SIFT (34.19% and 62.71%) (Table 2). The SROC curves reflected an acceptable and fairly good overall diagnostic performance for our suggested model compared to the SNP-based pathogenicity detection tools (Figure 3). The AUC analysis showed a significance overall performance of our suggested model compared to the selected SNP-based pathogenicity detection tool (Table 3).

4. Discussion

Since the late 1990s, the initiation of research using genetic testing or molecular medicine, development of diagnostics accuracy tests, and molecular assays that measure levels of genes or specific mutations are used to provide a specific therapy for an individual’s diseases. We suggested a regression model based on different scores including functional scores, conservation score, and genomic variation degree and compared the results to the published clinical result as reference. We observed the effect of a set of disease-causing missense mutations, determined from the general population. The susceptibility of Mendelian inherited disease is most frequently associated with SNPs; hence, the mechanisms by which this occurs are still poorly known. From a biological
Table 1: Prediction results of SNP-based pathogenicity detection tools compared with the published results.

|   | SNP-ID     | Variant | SIFT | PhD-SNP | PolyPhen-2 | fathmm | Provean | MutPred | References |
|---|------------|---------|------|---------|------------|--------|---------|---------|------------|
| 1 | rs74720349 | V3G     | 1    | 0       | 0          | 0      | 0       | 1       | 0          |
| 2 | rs201984525| L11P    | 0    | 1       | 0          | 0      | 0       | 1       | 0          |
| 3 | rs11033541 | L15R    | 1    | 1       | 1          | 0      | 1       | 1       | 1          |
| 4 | rs72551339 | H39D    | 1    | 1       | 1          | 1      | 1       | 1       | 1          |
| 5 | rs140365717| E56A    | 1    | 0       | 1          | 1      | 1       | 1       | 1          |
| 6 | rs4148323  | G71R    | 0    | 1       | 1          | 0      | 0       | 0       | 1          |
| 7 | rs72551340 | F83I    | 0    | 1       | 0          | 1      | 1       | 0       | 1          |
| 8 | rs144217005| V109A   | 0    | 1       | 0          | 1      | 0       | 0       | 0          |
| 9 | rs140867457| I16K    | 0    | 0       | 0          | 1      | 1       | 0       | 0          |
| 10| rs200734586| K118N   | 0    | 0       | 0          | 1      | 0       | 1       | 0          |
| 11| rs72551341 | L175Q   | 1    | 0       | 1          | 1      | 1       | 1       | 1          |
| 12| rs72551342 | C177R   | 0    | 0       | 1          | 1      | 1       | 1       | 1          |
| 13| rs201093245| Y192C   | 1    | 0       | 1          | 1      | 1       | 1       | 0          |
| 14| rs72551343 | R209W   | 1    | 0       | 1          | 1      | 1       | 1       | 1          |
| 15| rs144398951| I215V   | 0    | 0       | 1          | 0      | 0       | 0       | 0          |
| 16| rs144721642| G225M   | 0    | 0       | 0          | 1      | 0       | 0       | 0          |
| 17| rs35003977 | V225G   | 0    | 0       | 0          | 1      | 1       | 1       | 1          |
| 18| rs35350966 | P229Q   | 0    | 1       | 1          | 1      | 0       | 1       | 1          |
| 19| rs47640261 | T232N   | 0    | 0       | 1          | 0      | 0       | 1       | 0          |
| 20| rs7307513  | S250P   | 0    | 1       | 0          | 1      | 0       | 1       | 0          |
| 21| rs41950052 | P267R   | 1    | 1       | 1          | 1      | 1       | 0       | 0          |
| 22| rs143072292| V273F   | 1    | 1       | 0          | 1      | 1       | 1       | 0          |
| 23| rs72551345 | G276R   | 1    | 1       | 1          | 1      | 1       | 1       | 1          |
| 24| rs72551347 | I294T   | 1    | 0       | 1          | 1      | 1       | 1       | 1          |
| 25| rs62625011 | G308E   | 1    | 1       | 1          | 1      | 1       | 1       | 1          |
| 26| rs11000345 | K317E   | 0    | 1       | 0          | 0      | 0       | 0       | 0          |
| 27| rs200903749| I322V   | 1    | 0       | 1          | 1      | 0       | 1       | 1          |
| 28| rs17851756 | I322T   | 1    | 1       | 1          | 1      | 1       | 1       | 1          |
| 29| rs20203422 | I329T   | 1    | 0       | 1          | 1      | 1       | 1       | 1          |
| 30| rs72551348 | Q331R   | 1    | 1       | 1          | 1      | 1       | 1       | 1          |
| 31| rs139607673| R336W   | 1    | 1       | 1          | 1      | 1       | 1       | 1          |
| 32| rs144978321| S343L   | 0    | 1       | 1          | 1      | 1       | 0       | 1          |
| 33| rs149750520| N344K   | 1    | 0       | 1          | 1      | 1       | 0       | 0          |
| 34| rs201372184| A346V   | 1    | 1       | 1          | 1      | 0       | 1       | 1          |
| 35| rs72551351 | Q357R   | 1    | 1       | 1          | 1      | 1       | 1       | 1          |
| 36| rs34946978 | P364L   | 0    | 0       | 1          | 1      | 1       | 1       | 0          |
| 37| rs5750087 | R367G   | 1    | 0       | 1          | 1      | 1       | 1       | 1          |
| 38| rs72551352 | A368T   | 1    | 0       | 1          | 1      | 1       | 1       | 1          |
| 39| rs72551353 | A368S   | 1    | 0       | 1          | 1      | 1       | 1       | 1          |
| 40| rs72551354 | S381R   | 1    | 1       | 1          | 1      | 0       | 1       | 1          |
| 41| rs143573365| V386I   | 0    | 0       | 1          | 1      | 0       | 0       | 0          |
| 42| rs28934877 | N400H   | 1    | 1       | 1          | 1      | 1       | 1       | 1          |
| 43| rs72551355 | A401P   | 0    | 1       | 1          | 1      | 1       | 1       | 1          |
| 44| rs140613392| R403H   | 0    | 0       | 1          | 1      | 1       | 1       | 1          |
| 45| rs36076514 | V411L   | 0    | 0       | 0          | 1      | 0       | 1       | 0          |
| 46| rs72551356 | K428E   | 0    | 0       | 1          | 1      | 1       | 1       | 1          |
| 47| rs20172337 | M441T   | 0    | 0       | 0          | 1      | 0       | 0       | 0          |
| 48| rs143033456| R442C   | 1    | 1       | 1          | 1      | 1       | 1       | 1          |
| 49| rs201427749| R450C   | 1    | 0       | 1          | 1      | 1       | 1       | 1          |
| 50| rs200370335| R450H   | 1    | 0       | 1          | 1      | 1       | 1       | 1          |
| 51| rs114982090| P451L   | 1    | 1       | 1          | 1      | 1       | 1       | 1          |
Table 1: Continued.

| SNP-ID         | Variant | SIFT | PhD-SNP | PolyPhen-2 | fathmm | Provean | MutPred | References |
|----------------|---------|------|---------|------------|--------|---------|---------|------------|
| rs115410088    | F460L   | 1    | 0       | 1          | 1      | 1       | 1       | 0          |
| rs72551358     | E463A   | 1    | 1       | 1          | 1      | 1       | 1       | 0          |
| rs115944950    | E463D   | 0    | 1       | 1          | 1      | 0       | 1       | 0          |
| rs72551359     | L474M   | 1    | 1       | 1          | 1      | 0       | 0       | 0          |
| rs150687296    | R475H   | 1    | 0       | 1          | 1      | 1       | 1       | 0          |
| rs34993780     | S488C   | 0    | 1       | 1          | 1      | 1       | 1       | 0          |
| rs72551360     | V499M   | 1    | 0       | 1          | 1      | 0       | 0       | 1          |
| rs199723856    | A511P   | 0    | 1       | 1          | 1      | 0       | 1       | 0          |

Disease: 1; neutral: 0; references: OMIM, PMID, SNPdbe, HGMD, and Swissvar results.

Table 2: Calculated Matthew’s correlation coefficient (MCC) and accuracy (ACC) of the selected SNP-based pathogenicity detection tools and suggested model.

| Detection tools | TP  | FP  | FN  | TN  | MCC  | ACC  |
|-----------------|-----|-----|-----|-----|------|------|
| SIFT            | 19  | 7   | 15  | 18  | 34.19% | 62.71% |
| PHD-SNP         | 1   | 27  | 1   | 34  | 3.39%  | 55.56% |
| PolyPhen2       | 23  | 3   | 20  | 13  | 29.89% | 61.02% |
| fathmm          | 1   | 27  | 1   | 34  | 3.39%  | 55.56% |
| Provean         | 20  | 6   | 18  | 15  | 29.99% | 59.32% |
| MutPred         | 23  | 3   | 20  | 13  | 29.89% | 61.02% |
| Model           | 26  | 2   | 14  | 18  | 46.80% | 73.33% |

TP: true positive; TN: true negative; FP: false positive; FN: false negative; MCC: Matthew’s correlation coefficient; ACC: accuracy.

Figure 1: Flowchart of searching for SNPs.
Sensitivity (95% Cl) 0.56 (0.38–0.73) 0.50 (0.01–0.99) 0.53 (0.38–0.69) 0.50 (0.01–0.99) 0.53 (0.36–0.69)
Pooled sensitivity = 0.55 (0.48 to 0.62)

\( \chi^2 = 0.51; df = 6 (P = 0.9977) \)

Inconsistency (I²) = 0%

Specificity (95% Cl) 0.72 (0.51–0.88) 0.56 (0.42–0.68) 0.81 (0.54–0.96) 0.56 (0.42–0.68) 0.71 (0.48–0.99)
Pooled specificity = 0.66 (0.59 to 0.72)

\( \chi^2 = 16.14; df = 6 (P = 0.013) \)

Inconsistency (I²) = 62.8%

Figure 2: The individual and pooled diagnostic OR, sensitivity, specificity, negative likelihood ratio, positive likelihood ratio.

Table 3: Area under curve for all the selected SNP-based pathogenicity detection tools.

| Tools     | Area | Std. error | P value | 95% Confidence Interval |
|-----------|------|------------|---------|-------------------------|
| Model     | .639 | .071       | .046    | .499 (.778)              |
| SIFT      | .527 | .075       | .716    | .380 (.675)              |
| PolyPhen2 | .516 | .076       | .829    | .367 (.666)              |
| PHD-SNP   | .571 | .074       | .345    | .426 (.716)              |
| Provean   | .587 | .074       | .249    | .442 (.732)              |
| Fathmm    | .560 | .075       | .427    | .413 (.707)              |
| Mutpred   | .580 | .075       | .288    | .433 (.727)              |

*Significant, P < .05.

point of view, the mutated residues are important for the proper functioning of a suitable protein structure [28].

Genetic variation in phenotype of the diseases is often difficult to detect because of the complex genetic nature of these species. Using functional characteristics of the genetic mutation will provide a powerful tool to uncovering genetic traits in more complex species and provide novel insights into the molecular mechanisms of the diseases [29]. More importantly, the associations between genetic variations of SNPs of candidate genes that are selected to represent the phenotype are variable and an important feature from the disease study point of view [30].

Sensitivity was not reduced, while higher sensitivity was observed in our suggested model followed by PolyPhen2, Mutpred, and SIFT. We compared our suggested model to several well-established SNP-based pathogenicity detection tools, by which the satisfactory performance of our model and SIFT indicates the importance of a mutation position in the context of the entire protein. It is therefore reasonable to believe that analyzing the results of some SNP-based pathogenicity detection tools such as, our proposed model, SIFT and PolyPhen2 is both feasible and promising but not very excellent.

Saunders and Baker [31] and Bao and Cui [32] claimed that in case of unavailability of the conservation score, structural characteristics are valuable predictors. In this study we support using the sequence conservation score which is
a good predictor and showed that an acceptable level of accuracy is achieved using the conservation score. Dobson et al., used machine learning methods to measure the sequence conservation score and showed that it is the most powerful single predictor and reported a high level of accuracy using the conservation score alone [33]. They also reported higher accuracy in structural characteristics in combination with the conservation score. We also showed that structural characteristics in combination with the conservation score improves prediction accuracy and can reduce the error rate of the conservation score alone.

Ng and Henikoff used sequence and/or structure to predict the effect of a missense mutation on protein function in a mathematical model and claimed that their suggested model is a good SNP-based pathogenicity detection tools [13]. Capriotti et al. [14] developed a mathematical method that started from the protein sequence information, which can predict whether a new phenotype derived from a nsSNP can be related to a genetic disease in humans. They reported more than 74% accuracy in predicting whether a single point mutation can be disease related or not. Stitzel et al. [15] introduced a tool based on the hidden Markov models (HMM) for analyzing sequence homology of SNPs and reported 68% accuracy in predicting whether a single point mutation can be disease related or not. Shihab et al. [16], described a functional analysis Through Hidden markov models software and server and reported 71% accuracy in the predicton, which was less than SIFT (74%) but equal to PolyPhen2 (71%). Choi et al. [17] developed a new algorithm, which provides a generalized approach to predict the functional effects of protein sequence variations including single or multiple amino acid substitutions and in-frame insertions and deletions. They reported 84.8% accuracy compared to SIFT (84.5%) and PolyPhen2 (84.7%) in whether predicting that mutation can be disease related or not. In the present study we observed the highest accuracy with our suggested model as 73.33% compared with SIFT (62.71%) followed by PolyPhen2 and Mutpred (61.02%, in both).

5. Conclusions

Our suggested model is comparable to the well-established SNP-based pathogenicity detection tools and can appropriately reflect the role of a disease-associated SNP in both local and global structures. A major drawback of the weighted SNP-based pathogenicity detection tools is the inherited restriction that falls within conserved protein domains. Hence, unlike other sequence-based prediction tools, which are too slow for practical use in large-scale sequencing projects, the weighted tools are computationally inexpensive and fast. Although the accuracy of our suggested model is not relatively high, the functional impact of the pathogenic mutations at the protein level is highlighted, which improves the understanding of the molecular basis of mutation pathogenesis.

References

[1] B. Lodoso Torrecilla, E. Palomo Atance, C. Camarena Grande et al., "Crigler-Najjar syndrome: diagnosis and treatment," Anales de Pediatría, vol. 65, no. 1, pp. 73–78, 2006.
[2] K. M. Nair, P. Lohse, and S. Nampoothiri, “Crigler-Najjar syndrome type 2: novel UGT1A1 mutation,” *Indian Journal of Human Genetics*, vol. 18, no. 2, pp. 233–234, 2012.

[3] H. Sagili, N. Pramy, D. Jayalaksmi, and R. Rani, “Crigler-Najjar syndrome II and pregnancy outcome,” *Journal of Obstetrics and Gynaecology*, vol. 32, no. 2, pp. 188–189, 2012.

[4] H. Aloulou, A. B. Thabet, S. Khanfir et al., “Type I Crigler Najjar syndrome in Tunisia: a study of 30 cases,” *Tunisie Medicales*, vol. 88, no. 10, pp. 707–709, 2010.

[5] A. Iolascon, A. Melon, B. Coppola, and M. C. Rosatelli, “Crigler-Najjar syndrome type II resulting from three different mutations in the bilirubin uridine 5’-diphosphate-glucuronosyltransferase (UGT1A1) gene,” *Journal of Medical Genetics*, vol. 37, no. 9, pp. 712–713, 2000.

[6] M. I. Shevell, B. Bernard, and J. W. Adelson, “Crigler-Najjar syndrome type I: treatment by home phototherapy followed by orthotopic hepatic transplantation,” *Journal of Pediatrics*, vol. 110, no. 3, pp. 429–431, 1987.

[7] M. I. Shevell, A. Majnemer, and D. Schiff, “Neurologic perspectives of Crigler-Najjar syndrome type I,” *Journal of Child Neurology*, vol. 13, no. 6, pp. 265–269, 1998.

[8] M. Kanou, T. Usui, H. Ueyama, H. Sato, I. Ohkubo, and T. Mizutani, “Stimulation of transcriptional expression of human UDP-glucuronosyltransferase 1A1 by dexamethasone,” *Molecular Biology Reports*, vol. 31, no. 3, pp. 151–158, 2004.

[9] B. Boeckxmann, A. Bairoch, R. Apweiler et al., “The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003,” *Nucleic Acids Research*, vol. 31, no. 1, pp. 365–370, 2003.

[10] S. T. Sherry, M.-H. Ward, M. Khodorov et al., “dbSNP: the NCBI database of genetic variation,” *Nucleic Acids Research*, vol. 29, no. 1, pp. 308–311, 2001.

[11] D. Fredman, G. Munns, D. Riös et al., “HGVbase: a curated resource describing human DNA variation and phenotype relationships,” *Nucleic Acids Research*, vol. 32, pp. DS16–DS19, 2004.

[12] A. Szalontai and K. Csizsar, “Genetic insights into the functional elements of language,” *Human Genetics*, 2013.

[13] P. C. Ng and S. Henikoff, “Predicting the effects of amino acid substitutions on protein function,” *Annual Review of Genomics and Human Genetics*, vol. 7, pp. 61–80, 2006.

[14] E. Capriotti, R. Calabrese, and R. Casadio, “Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models,” *Human Mutation*, vol. 34, no. 1, pp. 57–65, 2013.

[15] Y. Choi, G. E. Sims, S. Murphy, J. R. Miller, and A. P. Chan, “Predicting the functional effect of amino acid substitutions and indels,” *PLoS ONE*, vol. 7, no. 10, Article ID e46688, 2012.

[16] B. Li, V. G. Krishnan, M. E. Mort et al., “Automated inference of molecular mechanisms of disease from amino acid substitutions,” *Bioinformatics*, vol. 25, no. 21, pp. 2744–2750, 2009.

[17] A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini, and V. A. McKusick, “Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders,” *Nucleic Acids Research*, vol. 33, pp. D514–D517, 2005.

[18] H. V. Firth, S. M. Richards, A. P. Bevan et al., “DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources,” *The American Journal of Human Genetics*, vol. 84, no. 4, pp. 524–533, 2009.

[19] A. S. Glas, J. G. Lijmer, M. H. Prins, G. J. Bonsel, and P. M. M. Bossuyt, “The diagnostic odds ratio: a single indicator of test performance,” *Journal of Clinical Epidemiology*, vol. 56, no. 11, pp. 1129–1135, 2003.

[20] B. Reva, Y. Antipin, and C. Sander, “Predicting the functional impact of protein mutations: application to cancer genomics,” *Nucleic Acids Research*, vol. 39, no. 17, p. e118, 2011.