Banerjee *et al.*
Middle East Fertility Society Journal (2024) 29:43
https://doi.org/10.1186/s43043-024-00202-9

Middle East Fertility Society Journal

CORRESPONDENCE

Open Access

From uncertain to certain—how to proceed with variants of uncertain significance

Emili Banerjee[1], Suman Pal[2], Abhijit Biswas[2] and Koutilya Bhattacharjee[2*]

Abstract

With the increased next generation sequencing (NGS) based genetic diagnosis due to technological boon, the biomedical world is getting a substantial number of single nucleotide variations (SNVs) every day along with other genetic variations. The detected SNVs may or may not have clinical significance. Based on different levels of study, these SNVs are categorized either as disease associated or not disease associated. However, there exists another category called as "uncertain" where the scientific literature has scanty of data. These "uncertain" or "variants of uncertain significance (VUS)" has become the greatest challenge for the diagnostic fraternity since no specific decision can be taken by them for the persons carrying the VUS. Therefore, there exists a huge knowledge gap that needs to be addressed for better patient care. The present study aims to find out the possible ways of investigation that may help in reducing this knowledge gap so that decisive approaches can be made against VUS for better and accurate patient care.

Keywords Uncertain, VUS, Variants of uncertain significance, Genetic diagnosis, SNVs, NGS test

Background

Genetic diagnosis has evolved tremendously from the last decade due to technological inventions and advancements. Apart from the classical dideoxy DNA sequencing, next generation sequencing (NGS) based exome sequence analyses have remarkably improved the diagnosis of different disorders due to nucleotide variations. This, in particular, has become very effective in reducing the diagnostic odyssey in many cases of rare genetic disorders. The single nucleotide variations (SNVs) are differentiated into five major groups, pathogenic, likely pathogenic, uncertain, likely not pathogenic/little clinical significance and not pathogenic/low clinical significance as per the American College of Medical Genetics (ACMG) guidelines, 2015 [1]. Among the five groups, four are decisive for the biomedical fraternity and the clinicians. But the group named as uncertain, which are commonly called as variants of uncertain significance (VUS), has become the greatest challenge to the medical genetics nowadays.

As per the ACMG guidelines, VUS are carrying pathogenicity probability between 0.05 and 0.949 [1]. This is the widest probability range among all of them. Any nucleotide or genomic variations that are not yet shown by laboratory studies to cause any loss of function or gain of function of any type come under VUS. It may consist large genomic duplications, any frameshift variants, promoter region variants, regulator region variants, intronic variants, missense alterations, small in-frame insertions/deletions and/or any silent variants [1]. Thus, effectively any variant that is yet to be studied for its functional significance comes under this umbrella category called VUS. This knowledge gap is becoming a major concern for the clinicians, geneticists and genetic counsellors as no decisive diagnosis can be made when a VUS is reported. The

*Correspondence:
Koutilya Bhattacharjee
koutilya@rkmvccrahara.org; sendtokoutilya@gmail.com
[1] Neuberg Center for Genomic Medicine, Neuberg Diagnostics, Ahmedabad, India
[2] Pharmacology and Disease Biology Lab, UG and PG Department of Zoology, Ramakrishna Mission Vivekananda Centenary College, Khardaha, West Bengal, India

present authors aim to propose possible investigation lines that can lead to a decision in cases of VUS.

Line of investigation

There are many ways of investigating the functional importance of any VUS. A comprehensive study would always be ideal one to establish the functional importance of any unreported variation, since every line has its own limitations.

Laboratory studies

Comparative laboratory studies between healthy individuals and patients (case–control) are the most acceptable way to establish the functional importance of any genetic variation. Otherwise, studies among the patients are also acceptable. However, limiting issue is about the availability of patient sample. If the study can be done from blood, saliva, urine or faeces, the study is in general possible, otherwise collection of tissue biopsy samples is very cumbersome and requires multiple ethical as well as administrative procedures. Furthermore, it is very unusual that a specific variation with particular functional importance being reported in human genome will also be there in the same genetic region among any laboratory animal model.

Allele frequency

As per the general rule, any SNV having an allele frequency less than 1% of the general population will be called as mutation and any SNV having greater than 1% frequency will be called as polymorphisms [2]. If the diagnosed SNV(s) comes under the reported polymorphisms from different existing open databases like "NCBI" and "Ensemble", then it signifies that probably the SNVs has least functional significance in the gene function. However, if the diagnosed SNV(s) is/are de novo or novel, then any population data will be scanty in any databases. Therefore, to be ascertain of the clinical significance of any VUS from allele frequency is practically very hard to achieve, since the status VUS comes due to not availability of these data sets.

Reverse phenotyping

This is a new approach of genomic diagnosis, which has become very effective in correlating genomic variations with disease. First genomic variations are identified through NGS-based tests and/or other molecular diagnostic tools and based on the genotype probable damages are targeted and diagnostically confirmed. This approach is also called as "genotype first" [2] as the approach first identifies the variation and then based on the genotype, phenotype is targeted for identification. Reverse phenotyping has emerged as a very good diagnostic tool in case of rare genetic disease.

Genotype–phenotype correlation

Correlating the disease phenotype with the detected SNV genotype is done through a population data driven statistical test. A dedicated online database called "GPCards" [3] (http://www.genemed.tech/gpcards) is there to find the genotype–phenotype correlation status of all the SNVs reported in scientific literatures. However, the correlation status of any SNV having scanty of specific population data or of any de novo or novel SNV is hard to determine from this database. For such situations, population data is extremely essential to be studied.

Family history and segregation analyses

It is one of the most important factors in decision-making for VUS. If any diseased proband carries a VUS, with other members of the family having the disease, it becomes easier to study the clinical significance of the VUS. Confirmation of the reported VUS among diseased family members indicates correlation or association of the VUS with the disease. Therefore, taking proper family history while counselling the patient is of utmost importance.

In silico *predictions*

Splicing variants as well as frameshift variants that change a major portion of protein primary structure are clinically significant for any protein function and can be designated as pathogenic or likely pathogenic even before any laboratory study. Similar situation is there for nonsense variants that cause protein truncation. There are a number of databases, tabulated as Table 1, that harbour structural as well as functional significances of reported SNVs, point mutations and small indel variations.

The majority of SNVs cause missense variations whose clinical significance is needed to be determined or at least predicted. For such cases, there are a number of web-based tools that can predict the effect of the SNV(s) on protein function. Choudhury et al. [10] classified some of these mutation analyses tools into two primary groups. First group predicts the local effect of the amino acid substitution while the second group analyses the effect at structure of the protein. However, an exhaustive classification of the applications and limitations of all these web-based amino acid substitution effect prediction tools are scanty in the scientific literature. An attempt is being made in the present manuscript to fill up this knowledge gap. To predict the effect of any amino acid substitution in the protein function, following factors are needed to be considered:

Table 1 List of the public databases harbouring the details of single nucleotide variations (SNVs) along with their possible utilities that can help to understand the functional implications of any variation

Sl No	Name	URL	Remarks	Reference
1	dbSNP	http://www.ncbi.nlm.nih.gov/SNP/	dbSNP helps to connect variations which include mutations with clinical significance and polymorphisms to other sequences available on NCBI by utilizing BLAST and E-PCR analysis of the flanking sequences that immediately surrounds the concerned variation	[4]
2	Ensembl	https://www.ensembl.org/	Ensembl provides reference data for the interpretation of the genome for desired species. Annotation of genome assemblies from various sources, consisting genes, regulatory regions, variants and comparative data helps to build bridge between scientific research and genome interpretation. There are many integrated tools and the most important of these is Ensembl Variant Effect Predictor (VEP) and it provides the effect of variants under consideration on genes, transcripts and protein sequence, along with regulatory regions	[5]
3	Online Mendelian Inheritance in Man (OMIM)	http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim	OMIM is the fundamental reservoir of well organized, curated data on genes and the phenotypes associated with it and the inter-relatedness between them. The entries in OMIM are derived from the peer-reviewed biomedical literature and are effectively curated into structured entries	[6]
4	(PharmGKB)	http://www.pharmgkb.org/	The PharmGKB website gives a wide range of pharmacogenomics data, from annotations of the primary literature to guidelines for fine-tuning drug treatment on the basis of genetic information	[7]
5	ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/	ClinVar is a public repository of human genetic variants and interpretations of the significance of the variants to disease	[8]
6	Human Gene Mutation Database (HGMD)	http://www.hgmd.cf.ac.uk/ac/index.php	HGMD consists of an extensive collection of published mutations of the germline in nuclear genes that are thought to be potential candidates to cause, or have some close association with human inherited disease	[9]

1. Position of the amino acid concerned—whether the amino acid is present in functional site of the protein or not. Functional sites include active site of enzyme, binding site of other molecule for primary function, allosteric sites or transmembrane domain binding site. Presence in any particular domain or in the loop region.
2. Effect of the change in local ionization—whether the local charge distribution remains same or alters in tolerable range or alters to intolerable situation.
3. Effect of the substitution on global structure of the protein—whether the substitution has any significant effect on the tertiary or quaternary structure of the concerned protein.
4. Whether the unchanged amino acid is conserved—to check the whether the wild-type amino acid is conserved among all the reported species in different databases.
5. Whether the SNV(s) can be functionally significant intronic variant that play role in protein expression or alteration of protein primary structure.

To evaluate all the aforesaid five parameters, different web-based prediction tools can be used. An exhaustive list of the tools and their specific utilities are summarized in Supplementary Table 1.

Uncertainty to certainty—how to proceed

Therefore, to start with a VUS for estimating its clinical significance, the primary mode of investigation will be assessment of associated clinical parameters as far as practicable along with simultaneous in silico predictions. This combination type of analysis would be the reverse phenotyping. Based on the VUS type, i.e. whether the variant is in the non-coding region or in coding regions, specific tools have to be selected to predict specific aspect as was mentioned earlier. The clinical correlation with genotype comes under the genotype–phenotype correlation study. This will lead the investigator to understand whether the variant has any possible functional implications or not. If there is no correlation, the VUS may get designation shift from class 3 to class 2 (not likely pathogenic). But, on the contrary, if reverse

phenotyping predicts the VUS to be clinically correlated, further investigations are required using targeted Sanger sequencing-based confirmation of the presence/absence of the variant among the parents of the index person as part of segregation analysis. If the VUS is inherited, the inheritance pattern could be either dominant or

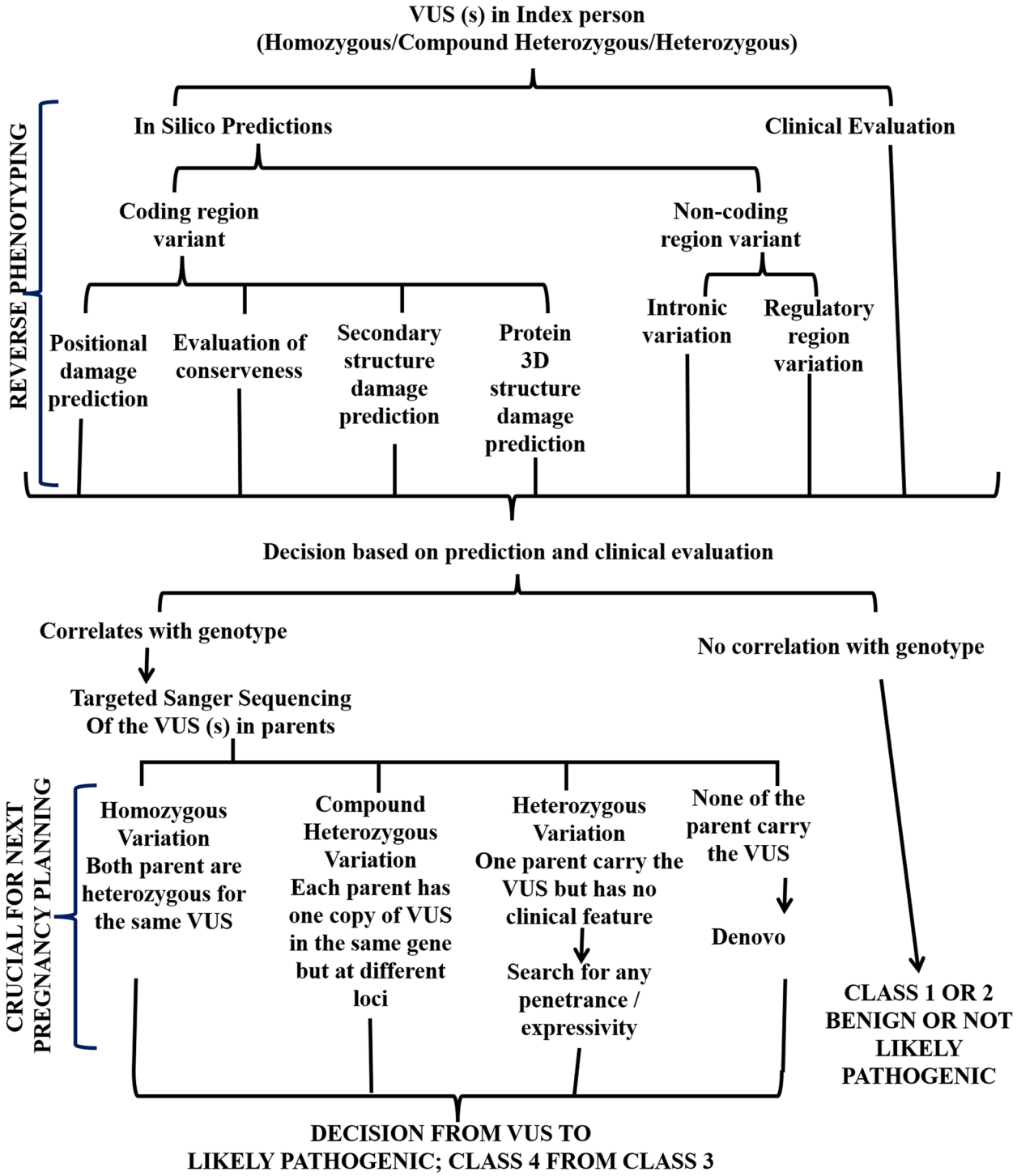

Fig. 1 Flow diagram of approaches to understand the possible functional implications of a VUS. The primary approach will always be through reverse phenotyping for finding the clinical correlation of the VUS. Followed by this, parental segregation analysis should be done for further confirmation of the functional implications of the VUS. Through these all studies, a VUS can be proposed to be redesignation to class 4 or class 5 or class 2

recessive. If dominant, it is expected that the parent carrying the VUS will also be affected with similar clinical conditions. But when the VUS is inherited in recessive mode, only homozygous recessive parent will be having similar phenotype. Targeted Sanger sequencing may also find both parents as heterozygous for the VUS(s) or each have different VUS(s) within the same gene leading to a compound heterozygosity in the index person. In both cases, parents will never be affected or having any clinical correlation. There may be another problematic situation for the clinician/genetic counsellor where parents may carry the VUS(s) (homozygous recessive or heterozygous dominant mode) but without any clinical correlation similar to the index person. In other words, they have the VUS(s) with proper genetic dosage but without any disease effect. This may definitely occur if the concerned gene has differential genetic penetrance or expressivity which needs further study to confirm. Lastly, the VUS(s) may occur in the index person in a de novo way. In such cases, he/she will be the only one person for clinical correlation. After completing the entire analysis, the VUS may be proposed to be reclassified as class 4 (likely pathogenic) or 5 (pathogenic). The following flowchart summarizes the aspects and ways of analysing VUS (Fig. 1).

Clinical insights

Recent studies reported presence of significant percentage of VUS among prenatal cases [11, 12], which is arising the question of taking decision about the VUS(s) present in the foetus. Predictive analysis of any VUS may help to understand the potential effect(s) that the variant can cause. This is extremely important for the clinicians as well genetic counsellors when the analysis is being done for a foetus and the would-be parents seek information about the complexities that may arise. To worsen the scenario, it may also happen that the VUS present in the foetus is a de novo variation or variation with differential genetic penetrance or expressivity. In all cases, further diagnosis is solely dependent on the probable functional implications of the VUS. Therefore, predictive VUS effect analysis may give some critical input about its functional implications for the genetic counsellor and/or the clinician for helping the would-be parents to take decision of continuing the pregnancy. This critical input is very important for precisive patient care. Apart from this, VUS analysis may also help in reducing the diagnostic odyssey in many rare as well as common genetic diseases.

Conclusion

One of the greatest challenges for the biomedical genetics fraternity, primarily the genetic counsellors (GC), is to understand the clinical significance of the VUS and to help the clinician in taking decisions about the index person. The situation becomes complex when a couple with a diseased first child comes to the GC for planning a healthy pregnancy and that first index child is diagnosed to carry VUS(s). To make it worse if similar couple with a diseased child come with a running pregnancy. The decision of letting the pregnancy go or terminate depends upon the diagnosis. In such cases, determination of clinical significance of the VUS(s) remains very crucial not only from a treatment point of view but also an ethical issue of terminating a pregnancy is also associated.

The next step after prediction analyses is to report for the sake of science, knowledge and precision treatment in future. The simple fact is, a VUS is called VUS only because no data or report about it is there in the scientific literature.

However, there are important issues regarding the reporting. First, bioinformatic analyses are time consuming and require expertise which may not always be possible for the clinician, or GC or the genome analyst. It is obvious that, this trio of clinician, genome analyst and GC may come from different units and do not have any liaison among them. Secondly, since checking the parental segregation is very crucial for predictive analyses, a couple without any plan for future pregnancy will never be interested in proceeding for further diagnosis. So, question mark lies for availability of the samples. Thirdly, if the parents somehow agree to give blood sample, it is unethical for the clinician and GC or diagnostic company to charge then for paying the test cost, since they do not have any future plan that be obtained from the test results. Situation is worse among the low-income groups who are yet to afford the necessary tests even.

Thus, a regular object-oriented expert work force with competent infrastructure and ample funding is required for completion of all the analyses of VUS. The present author foresights formation of a global VUS consortium where the found VUS details may be registered by corresponding biomedical team for further analyses of all kind to take decision about its clinical significance and upload the details in a global publicly accessible database. Only then, the VUS(s) will reach to any significant decision.

Abbreviations

ACMG	American College of Medical Genetics
DNA	Deoxyribonucleic acid
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing
SNV	Single nucleotide variation
VUS	Variants of uncertain significance

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43043-024-00202-9.

Additional file 1: Supplementary Table 1 [13–43].

Acknowledgements
None.

Authors' contributions
Emili Banerjee—manuscript writing and conceptualization. Suman Pal—table preparation, bioinformatic studies, manuscript writing. Abhijit Biswas—table preparation, bioinformatic tool evaluation, manuscript writing. Koutilya Bhattacharjee—conceptualize the manuscript, primary writing, critical inputs.

Funding
The authors declare that there is no funding to be acknowledged.
Availability of data materials.
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that there is no conflict of interest.

Received: 2 June 2024 Accepted: 11 August 2024
Published online: 16 August 2024

References

1. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. 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. Genet Med. 2015 May;17(5):405–24. https://doi.org/10.1038/gim.2015.30. Epub 2015 Mar 5.
2. Trent RJ (2012) Molecular medicine. Academic Press, Sydney, Ronald J Trent
3. Li B, Wang Z, Chen Q, Li K, Wang X, Wang Y, Zeng Q, Han Y, Lu B, Zhao Y, Zhang R, Jiang L, Pan H, Luo T, Zhang Y, Fang Z, Xiao X, Zhou X, Wang R, Zhou L, Wang Y, Yuan Z, Xia L, Guo J, Tang B, Xia K, Zhao G, Li J (2021) GPCards: an integrated database of genotype-phenotype correlations in human genetic diseases. Comput Struct Biotechnol J 22(19):1603–1611. https://doi.org/10.1016/j.csbj.2021.03.011
4. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29(1):308–311. https://doi.org/10.1093/nar/29.1.308
5. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, Berry A, Bhai J, Bignell A, Billis K, Boddu S, Brooks L, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, Donaldson S, El Houdaigui B, El Naboulsi T, Fatima R, Giron CG, Genez T, Martinez JG, Guijarro-Clarke C, Gymer A, Hardy M, Hollis Z, Hourlier T, Hunt T, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Marugán JC, Mohanan S, Mushtaq A, Naven M, Ogeh DN, Parker A, Parton A, Perry M, Piližota I, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Pérez-Silva JG, Stark W, Steed E, Sutinen K, Sukumaran R, Sumathipala D, Suner MM, Szpak M, Thormann A, Tricomi FF, Urbina-Gómez D, Veidenberg A, Walsh TA, Walts B, Willhoft N, Winterbottom A, Wass E, Chakiachvili M, Flint B, Frankish A, Giorgetti S, Haggerty L, Hunt SE, Ilsley GR, Loveland JE, Martin FJ, Moore B, Mudge JM, Muffato M, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Dyer S, Harrison PW, Howe KL, Yates AD, Zerbino DR, Flicek P. Ensembl 2022. Nucleic Acids Res. 2022;50(D1):D988-D995. https://doi.org/10.1093/nar/gkab1049.
6. Amberger JS, Hamosh A (2017) Searching Online Mendelian Inheritance in Man (OMIM): a knowledgebase of human genes and genetic phenotypes. Curr Protoc Bioinformatics. 58:1.2.1-1.2.12. https://doi.org/10.1002/cpbi.27
7. Barbarino JM, Whirl-Carrillo M, Altman RB, Klein TE (2018Jul) PharmGKB: a worldwide resource for pharmacogenomic information. Wiley Interdiscip Rev Syst Biol Med. 10(4):e1417. https://doi.org/10.1002/wsbm.1417. (Epub 2018 Feb 23)
8. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR (2018) ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 46(D1):D1062–D1067. https://doi.org/10.1093/nar/gkx1153
9. Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, Hayden M, Heywood S, Millar DS, Phillips AD, Cooper DN (2020) The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. Hum Genet 139(10):1197–1207. https://doi.org/10.1007/s00439-020-02199-3. (Epub 2020 Jun 28)
10. Choudhury A, Mohammad T, Anjum F, Shafie A, Singh IK, Abdullaev B, Pasupuleti VR, Adnan M, Yadav DK, Hassan MI (2022) Comparative analysis of web-based programs for single amino acid substitutions in proteins. PLoS ONE 17(5):e0267084. https://doi.org/10.1371/journal.pone.0267084
11. Mardy AH, Wiita AP, Wayman BV, Drexler K, Sparks TN, Norton ME (2021) Variants of uncertain significance in prenatal microarrays: a retrospective cohort study. BJOG 128(2):431–438. https://doi.org/10.1111/1471-0528.16427
12. Cornthwaite M, Turner K, Armstrong L, Boerkoel CF, Chang C, Lehman A, Nikkel SM, Patel MS, Allen MV, Langlois S (2022) Impact of variation in practice in the prenatal reporting of variants of uncertain significance by commercial laboratories: need for greater adherence to published guidelines. Prenat Diagn 42(12):1512–1524. https://doi.org/10.1002/pd.6232
13. Ng PC, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31(13):3812–3814. https://doi.org/10.1093/nar/gkg509
14. Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet Chapter 7(Unit7):20. https://doi.org/10.1002/0471142905.hg0720s76
15. Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C (2018) FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. Bioinformatics 34(3):511–513. https://doi.org/10.1093/bioinformatics/btx536
16. Capriotti E, Fariselli P (2017) PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. Nucleic Acids Res 45(W1):W247–W252. https://doi.org/10.1093/nar/gkx369
17. Choi Y, Chan AP (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. (Epub 2015 Apr 6)
18. Reva B, Antipin Y, Sander C (2007) Determinants of protein function revealed by combinatorial entropy optimization. Genome Biol 8(11):R232. https://doi.org/10.1186/gb-2007-8-11-r232
19. Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11(4):361–362. https://doi.org/10.1038/nmeth.2890
20. Niroula A, Urolagin S, Vihinen M (2015) PON-P2: prediction method for fast and reliable identification of harmful variants. PLoS ONE 10(2):e0117380. https://doi.org/10.1371/journal.pone.0117380
21. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, Mort M, Cooper DN, Sebat J, Iakoucheva LM, Mooney SD, Radivojac P (2020) Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat Commun 11(1):5918. https://doi.org/10.1038/s41467-020-19669-x
22. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R (2013) WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genomics. Suppl 3(Suppl3):S6. https://doi.org/10.1186/1471-2164-14-S3-S6
23. López-Ferrando V, Gazzo A, de la Cruz X, Orozco M, Gelpí JL (2017) PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. Nucleic Acids Res 45(W1):W222–W228. https://doi.org/10.1093/nar/gkx313

24. Pires DE, Ascher DB, Blundell TL (2014) mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics 30(3):335–342. https://doi.org/10.1093/bioinformatics/btt691
25. Worth CL, Preissner R, Blundell TL (2011) SDM—a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res 39(Web Server issue):W215–22. https://doi.org/10.1093/nar/gkr363.
26. Laimer J, Hiebl-Flach J, Lengauer D, Lackner P (2016) MAESTROweb: a web server for structure-based protein stability prediction. Bioinformatics 32(9):1414–1416. https://doi.org/10.1093/bioinformatics/btv769
27. Parthiban V, Gromiha MM, Schomburg D (2006) CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Res 34(Web Server issue):W239–42. https://doi.org/10.1093/nar/gkl190.
28. Rodrigues CHM, Pires DEV, Ascher DB (2021) DynaMut2: assessing changes in stability and flexibility upon single and multiple point missense mutations. Protein Sci 30(1):60–69. https://doi.org/10.1002/pro.3942
29. Schubach M, Maass T, Nazaretyan L, Röner S, Kircher M (2024) CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. Nucleic Acids Res 52(D1):D1143-D1154. https://doi.org/10.1093/nar/gkad989.
30. Rogers MF, Shihab HA, Gaunt TR, Campbell C (2017) CScape: a tool for predicting oncogenic single-point mutations in the cancer genome. Sci Rep 7(1):11597. https://doi.org/10.1038/s41598-017-11746-4
31. 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
32. Chen L, Jin P, Qin ZS (2016) DIVAN: accurate identification of non-coding disease-specific risk variants using multi-omics profiles. Genome Biol 17(1):252. https://doi.org/10.1186/s13059-016-1112-z
33. Yang H, Chen R, Wang Q, Wei Q, Ji Y, Zheng G, Zhong X, Cox NJ, Li B (2019) De novo pattern discovery enables robust assessment of functional consequences of non-coding variants. Bioinformatics 35(9):1453–1460. https://doi.org/10.1093/bioinformatics/bty826
34. Ionita-Laza I, McCallum K, Xu B, Buxbaum JD (2016) A spectral approach integrating functional genomic annotations for coding and noncoding variants. Nat Genet 48(2):214–220. https://doi.org/10.1038/ng.3477
35. Ioannidis NM, Davis JR, DeGorter MK, Larson NB, McDonnell SK, French AJ, Battle AJ, Hastie TJ, Thibodeau SN, Montgomery SB, Bustamante CD, Sieh W, Whittemore AS (2017) FIRE: functional inference of genetic variants that regulate gene expression. Bioinformatics 33(24):3895–3901. https://doi.org/10.1093/bioinformatics/btx534
36. Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, Khurana E, Gerstein M (2014) FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. Genome Biol 15(10):480. https://doi.org/10.1186/s13059-014-0480-5
37. Lu Q, Hu Y, Sun J, Cheng Y, Cheung KH, 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. https://doi.org/10.1038/srep10576
38. Huang YF, Gulko B, Siepel A (2017) Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. Nat Genet 49(4):618–624. https://doi.org/10.1038/ng.3810
39. Gussow AB, Copeland BR, Dhindsa RS, Wang Q, Petrovski S, Majoros WH, Allen AS, Goldstein DB (2017) Orion: detecting regions of the human non-coding genome that are intolerant to variation using population genetics. PLoS ONE 12(8):e0181604. https://doi.org/10.1371/journal.pone.0181604
40. Zhou L, Zhao F (2018) Prioritization and functional assessment of non-coding variants associated with complex diseases. Genome Med 10(1):53. https://doi.org/10.1186/s13073-018-0565-y
41. Zhang S, He Y, Liu H, Zhai H, Huang D, Yi X, Dong X, Wang Z, Zhao K, Zhou Y, Wang J, Yao H, Xu H, Yang Z, Sham PC, Chen K, Li MJ (2019) regBase: whole genome base-wise aggregation and functional prediction for human non-coding regulatory variants. Nucleic Acids Res 47(21):e134. https://doi.org/10.1093/nar/gkz774
42. Smedley D, Schubach M, Jacobsen JOB, Köhler S, Zemojtel T, Spielmann M, Jäger M, Hochheiser H, Washington NL, McMurry JA, Haendel MA, Mungall CJ, Lewis SE, Groza T, Valentini G, Robinson PN (2016) A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. Am J Hum Genet 99(3):595–606. https://doi.org/10.1016/j.ajhg.2016.07.005
43. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W (2016) REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet 99(4):877–885. https://doi.org/10.1016/j.ajhg.2016.08.016

Publisher's Note