Supplementary Information for
MobiDetails: Online DNA variants interpretation

David Baux[1,*], Charles Van Goethem[1], Olivier Ardouin[2], Thomas Guignard[3], Anne Bergougnoux[1], Michel Koenig[1], Anne-Françoise Roux[1]

[1]Laboratoire de Génétique Moléculaire, CHU de Montpellier, Université de Montpellier, France
[2]Plateau de Médecine Moléculaire et Génomique, CHU de Montpellier, Université de Montpellier, France
[3]Unité de Génétique Chromosomique, CHU de Montpellier, Université de Montpellier, France
*Correspondence : david.baux@inserm.fr

Table of contents :

Supplementary Method 1:

MobiDetails is a python Flask (https://palletsprojects.com/p/flask/) application built on top of a PostgreSQL database (Supplementary Figure 1). The database handles gene coordinates, variants and user-specific data. Many external annotations (such as ClinVar[1]) are stored as compressed flat files (often in VCF format) and queried via tabix (https://www.htslib.org/) by the Flask engine. The engine either renders HTML/CSS/JS web pages using jinja2 templates (https://jinja.palletsprojects.com/en/2.11.x/) or returns a JSON object, depending on the caller (web navigation or API call). The API can be accessed via a swagger application for testing purpose (https://mobidetails.iurc.montp.inserm.fr/MDAPI/).

Supplementary Table 1: External resources included in MobiDetails

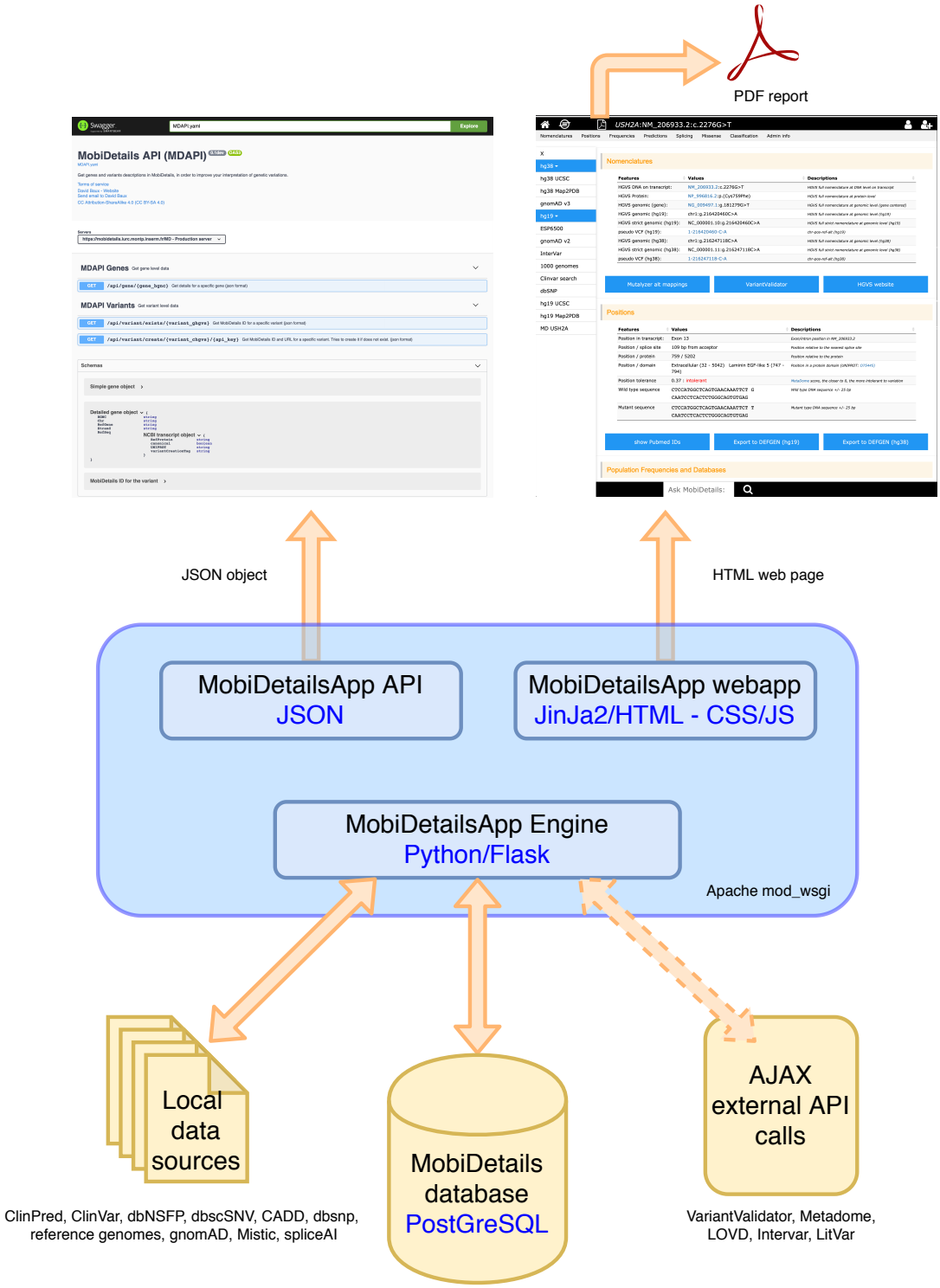
Resource	Version	Access	Reference
CADD	1.5	Local VCF	[2]
ClinPred	2018_hg19	Local file	[3]
ClinVar	20200830	Local VCF	[1]
dbMTS	1.0	Local file	[4]
dbNSFP	4.0a	Local file	[5,6]
dbscSNV	1.1	Local file	[7]
dbSNP	154	Local VCF	[8]
Eigen	1.1	From dbNSFP or dbMTS	[9]
FatHMM	2.3	From dbNSFP	[10]
gnomAD exome/genome	2.0.1	Local VCF	[11]
gnomAD genome	3	Local VCF	[11]
InterVar	NA	API	[12]
LitVar	1	API	[13]
LOVD	2-3	API	[14]
MaxEntScan	2004	Perl script	[15]
MetaDome	1.0.1	API – Local file	[16]
MetaSVM-LR	NA	From dbNSFP	[17]
Mistic	1	Local file	[18]
MPA	1	Internally computed	[19]
Mutalyzer	2.0.32	API	[20]
Polyphen-2	2.2.2	From dbNSFP	[21]
REVEL	NA	From dbNSFP	[22]
SIFT	Ensembl 66	From dbNSFP	[23]
spliceAI	1.3	Local VCF	[24]
TogoWS	NA	API	[25]
VariantValidator	1.0.4	API	[26]

NA: not available

Local resources are either automatically updated (ClinVar), i.e. a script periodically checks whether a more recent version of the resource has been released, or must be updated manually, when updates are quite rare (gnomAD, dbNSFP…). The Metadome API returns for each transcript a JSON file containing all the scores for each position of the encoded protein: MobiDetails calls it the first time a gene is visited and keeps a local copy of the JSON file for each gene. Then when a variant page is called, the data is retrieved from the local copy.

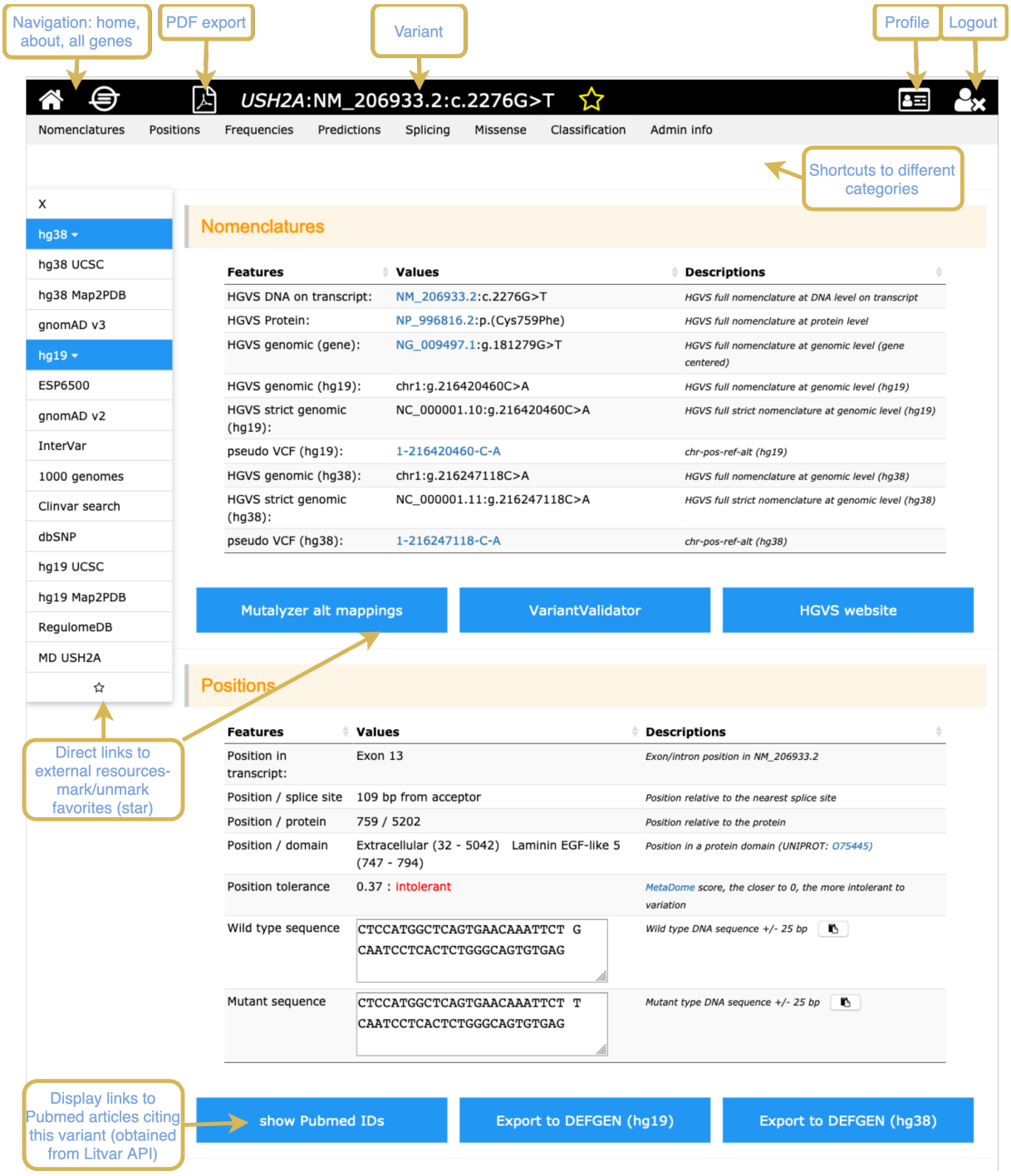
VariantValidator API is called when a variant is annotated for the first time. Mutalyzer API is called when a variant is annotated for the first time from its dbSNP identifier (rs id).

Supplementary Figure 1: Global architecture of MobiDetails



The main engine manages the PostGreSQL relational database which stores the genes and variants data. On query, it uses tabix (https://www.htslib.org/) to retrieve data from compressed files, and sends AJAX http requests to selected external APIs. Data sources are listed in Suppl. Table 1. SQL querys and data extracted from the files are treated to be rendered as JSON objects (REST API calls) or web pages.

Supplementary Figure 2: Full view of the variant web page



.

.

.

.

.

.

.

Population Frequencies and Databases

Features	Values	Descriptions
gnomAD exome:	0.0009	*v2.0.1 Exomes global MAF*
gnomAD genome:	0.0009	*v2.0.1 Genomes global MAF*
gnomAD v3:	1.34030e-03	*v3 Genomes global MAF*
dbSNP rsid:	rs80338902	*Identifier for NCBI dbSNP*
Clinvar:	Pathogenic	*Clinvar interpretation*
Intervar:	Pathogenic	*Semi-automated ACMG classification - click on the intervar link to adjust*
LOVD Matches:	• Whole genome datasets • Global Variome shared LOVD • Retinal and hearing impairment genetic mutation database • MSeqDR-LSDB Mitochondrial Disease Locus Specific Database	*LOVD match in public instances*
LOVD Effect Reported:	• unknown: 1 • functionProbablyAffected: 16 • functionAffected: 178 • functionNotAffected: 2	*Effects reported by LOVD submitters*
LOVD Effect Concluded:	• functionAffected: 175 • notClassified: 20 • functionNotAffected: 2	*Effects concluded by LOVD curators*

Direct links to various LOVD instances containing the variant

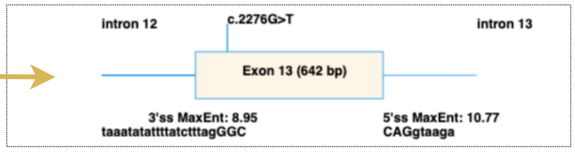
Overall predictions

Features	Values	Descriptions
CADD raw:	3.924570	*[-6.41;35.5] The higher the less likely to be observed*
CADD phred:	25.4	*Phred-like scaling of raw score*
Eigen raw:	1.04852691810073	*[-3.33;6.84] The higher the less likely to be observed*
Eigen phred:	15.52217	*Phred-like scaling of raw score*
MPA score:	10	*Raw score [0;10], 10: high impact*
MPA impact:	Clinvar pathogenic	*Impact type*

Splicing predictions

The exonic splicing context of the variant including natural splice sites scores is summarized in the graph below:

Representation of the exon, the variant position within the exon and natural 3' and 5' splice sites sequences and MaxEntScan values

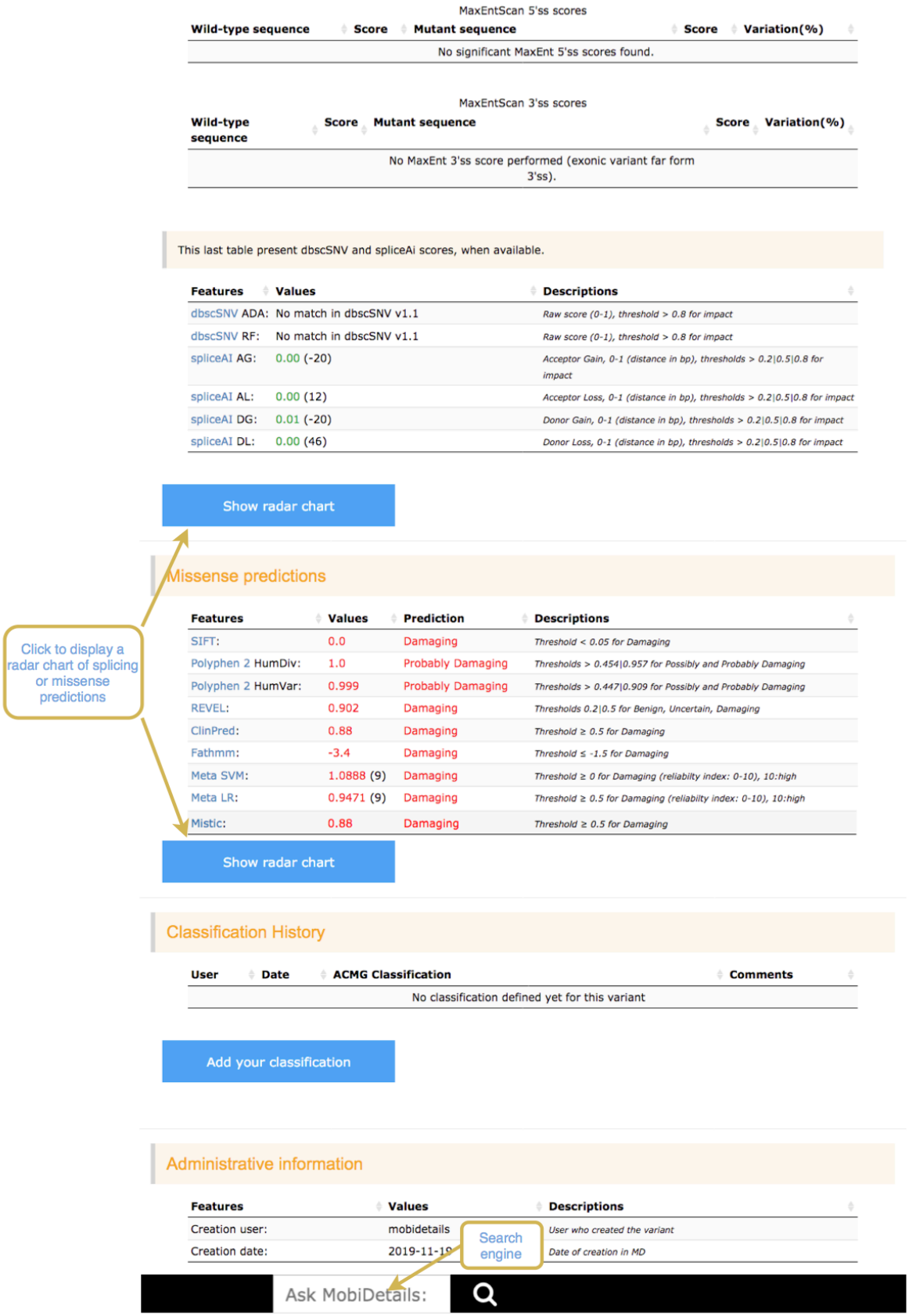


MaxEntScan scores are presented in the two following tables. Selected scores have:

- a |variation| > 15% and
- a raw score for mutant or wild-type of at least 3

The upper sequence shows the variation site in red, and the lower sequence the putative splice site considered by MaxEntScan (putative introns are shown as lower case and exons as upper case).

.



Full view of a MobiDetails variant page for *USH2A* NM_206933.2:c.2276G>T variant (https://mobidetails.iurc.montp.inserm.fr/MD/variant/7). Up to 8 categories can be displayed

on the center panel depending on the nature of the variant (i.e. not all insertions/deletions have splicing predictions, non-missense variants do not present missense predictions, 3'UTR variants present miRNA target sites predictions…). The left panel is dedicated to direct external links and to the button star used to mark/unmark a variant as favorite when logged in. The header is used to present links for navigation (home, about…), for pdf export and administrative tasks (login/logout, register, profile page). Finally, the footer is dedicated to the search engine.

References

1. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: Public archive of interpretations of clinically relevant variants. Nucleic Acids Research. 2016;44(D1):D862-8.

2. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Research. 8 janv 2019;47(D1):D886-94.

3. Alirezaie N, Kernohan KD, Hartley T, Majewski J, Hocking TD. ClinPred: Prediction Tool to Identify Disease-Relevant Nonsynonymous Single-Nucleotide Variants. The American Journal of Human Genetics. oct 2018;103(4):474-83.

4. Li C, Mou C, Swartz MD, Yu B, Bai Y, Tu Y, et al. dbMTS: A comprehensive database of putative human microRNA target site SNVs and their functional predictions. Human Mutation. 2020;41(6):1123-30.

5. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Human mutation. août 2011;32(8):894-9.

6. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Human Mutation. mars 2016;37(3):235-41.

7. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Research. 16 déc 2014;42(22):13534-44.

8. Sherry ST. dbSNP: the NCBI database of genetic variation. Nucleic Acids Research. 2001;29(1):308-11.

9. Ionita-Laza I, McCallum K, Xu B, Buxbaum JD. A spectral approach integrating functional genomic annotations for coding and noncoding variants. Nat Genet. févr 2016;48(2):214-20.

10. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, et al. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. Human Mutation. janv 2013;34(1):57-65.

11. Genome Aggregation Database Consortium, Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. mai 2020;581(7809):434-43.

12. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. American journal of human genetics. 2 févr 2017;100(2):267-80.

13. Allot A, Peng Y, Wei C-H, Lee K, Phan L, Lu Z. LitVar: a semantic search engine for linking genomic variant data in PubMed and PMC. Nucleic Acids Research. 2 juill 2018;46(W1):W530-6.

14. Fokkema IF a. C, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. Human Mutation. 22 mai 2011;32(5):557-63.

15. Yeo G, Burge CB. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. Journal of Computational Biology. 2004;11(2-3):377-94.

16. Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. Human Mutation. 2019;40(8):1030-8.

17. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Human Molecular Genetics. 2015;24(8):2125-37.

18. Chennen K, Weber T, Lornage X, Kress A, Böhm J, Thompson J, et al. MISTIC: A prediction tool to reveal disease-relevant deleterious missense variants. Andrade-Navarro MA, éditeur. PLoS ONE. 31 juill 2020;15(7):e0236962.

19. Yauy K, Baux D, Pegeot H, Van Goethem C, Mathieu C, Guignard T, et al. MoBiDiC Prioritization Algorithm, a Free, Accessible, and Efficient Pipeline for Single-Nucleotide Variant Annotation and Prioritization for Next-Generation Sequencing Routine Molecular Diagnosis. The Journal of Molecular Diagnostics. juill 2018;20(4):465-73.

20. Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PEM. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. Human mutation. 2008;29(1):6-13.

21. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature methods. avr 2010;7(4):248-9.

22. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. The American Journal of Human Genetics. oct 2016;99(4):877-85.

23. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. janv 2009;4(7):1073-81.

24. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell. janv 2019;176(3):535-548.e24.

25. Katayama T, Nakao M, Takagi T. TogoWS: integrated SOAP and REST APIs for interoperable bioinformatics Web services. Nucleic Acids Research. 1 juill 2010;38(Web Server):W706-11.

26. Freeman PJ, Hart RK, Gretton LJ, Brookes AJ, Dalgleish R. VariantValidator: Accurate validation, mapping, and formatting of sequence variation descriptions. Human Mutation. janv 2018;39(1):61-8.