Original Paper

Interactive Web-Based Resource for Annotation of Genetic Variants Causing Hereditary Angioedema (HADA): Database Development, Implementation, and Validation

Alejandro Mendoza-Alvarez[1], MSc; Adrián Muñoz-Barrera[2], BSc; Luis Alberto Rubio-Rodríguez[2], BSc; Itahisa Marcelino-Rodriguez[1], PhD; Almudena Corrales[1,3], CLT; Antonio Iñigo-Campos[2], CLT; Ariel Callero[4], MD, PhD; Eva Perez-Rodriguez[4], MD; Jose Carlos Garcia-Robaina[4], MD; Rafaela González-Montelongo[2], PhD; Jose Miguel Lorenzo-Salazar[2], MSc; Carlos Flores[1,2,3,5], PhD

[1]Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain

[2]Genomics Division, Instituto Tecnológico y de Energías Renovables, Santa Cruz de Tenerife, Spain

[3]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain

[4]Allergy Unit, Hospital Universitario Nuestra Señora de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain

[5]Instituto de Tecnologías Biomédicas, Universidad de La Laguna, Santa Cruz de Tenerife, Spain

Corresponding Author:
Carlos Flores, PhD
Research Unit
Hospital Universitario Nuestra Señora de Candelaria
Universidad de La Laguna
Carretera del Rosario s/n
Santa Cruz de Tenerife, 38010
Spain
Phone: 34 922602938
Email: cflores@ull.edu.es

Abstract

Background: Hereditary angioedema is a rare genetic condition caused by C1 esterase inhibitor deficiency, dysfunction, or kinin cascade dysregulation, leading to an increased bradykinin plasma concentration. Hereditary angioedema is a poorly recognized clinical entity and is very often misdiagnosed as a histaminergic angioedema. Despite its genetic nature, first-line genetic screening is not integrated in routine diagnosis. Consequently, a delay in the diagnosis, and inaccurate or incomplete diagnosis and treatment of hereditary angioedema are common.

Objective: In agreement with recent recommendations from the International Consensus on the Use of Genetics in the Management of Hereditary Angioedema, to facilitate the clinical diagnosis and adapt it to the paradigm of precision medicine and next-generation sequencing–based genetic tests, we aimed to develop a genetic annotation tool, termed Hereditary Angioedema Database Annotation (HADA).

Methods: HADA is built on top of a database of known variants affecting function, including precomputed pathogenic assessment of each variant and a ranked classification according to the current guidelines from the American College of Medical Genetics and Genomics.

Results: HADA is provided as a freely accessible, user-friendly web-based interface with versatility for the entry of genetic information. The underlying database can also be incorporated into automated command-line stand-alone annotation tools.

Conclusions: HADA can achieve the rapid detection of variants affecting function for different hereditary angioedema types, and further integrates useful information to reduce the diagnosis odyssey and improve its delay.

(J Med Internet Res 2020;22(10):e19040) doi: 10.2196/19040

KEYWORDS

XSL•FO
RenderX

Introduction

Hereditary angioedema (HAE) is a rare genetic disease caused by an increase of vascular permeability, generating recurrent acute swelling episodes commonly localized on the face, trunk, and extremities. HAE can also be life-threatening when the upper airways or the tissues from the oral cavity are affected [1,2]. Because of its nonspecific clinical signs, HAE is poorly recognized by physicians. Therefore, delayed or ambiguous diagnoses are common, increasing the risk of patient morbidity and mortality [3]. With exceptions for particular subtypes, the recent international consensus guidelines recommend the use of genetic testing to reach a definitive HAE diagnosis in clinical practice [4]. However, with increasing recognition that the genetic cause of HAE is more complex than previously anticipated [5], the tasks involved in the identification of the genetic defect in each patient are increasingly demanding [6].

With the decreasing cost and increasing sequencing throughput of next-generation sequencing (NGS), the screens for gene defects can now be accomplished through a simultaneous evaluation of gene sets, the exome, or the whole genome as indicated [7]. However, access to either high-performance computational equipment and trained bioinformatics personnel or, alternatively, to paywalled cloud-based all-inclusive solutions is the most limiting factor for NGS to be efficiently used in clinical settings [5]. To facilitate the identification of pathogenic variants in HAE patients, Kalmár et al [8] developed HAEdb, an online locus-specific database to centralize the information of the genetic alterations, which allows researchers to retrieve mutation information and contribute new detected variants. However, HAEdb focuses only on the most frequently affected gene (SERPING1, encoding the C1 esterase inhibitor [C1-INH]) and uses a data retrieval scheme based on matching with the existing records, requiring the user to prioritize the most likely variant affecting function. Besides, HAEdb has not been updated to include the pathogenic classification according to the standard guidelines established by the American College of Medical Genetics and Genomics (ACMG) [9], which would allow for standardization between laboratories and studies, while guiding global efforts for improving NGS-based diagnosis to move toward the precision medicine paradigm. More recent efforts such as that of Ponard et al [10] have aimed to fill this gap and updated the mutational spectrum in two of the known HAE genes (SERPING1 and F12) based on the Leiden Open Variant Database (LOVD) v3.0, a platform-independent framework for the maintenance and curation of a web-based database of genetic variants [11]. Nevertheless, these resources are ill-adapted to current NGS technologies and to the evolving knowledge of the genetic causes of HAE.

To fill this gap, we here present the Hereditary Angioedema Database Annotation (HADA) tool, a freely accessible, user-friendly, and versatile web-based interface to facilitate the identification of the genetic variants causing HAE.

Methods

Gene and Variant Extraction to Build the Database

We retrieved the variants found among HAE cases and relatives from HAEdb [8] (accessed June 18, 2020) and LOVD v3.09 (Figure 1). These databases contain hundreds of records for variants in the SERPING1 or F12 genes detected in clinical studies of families with one or more members affected by HAE. Additionally, a search was performed on January 1, 2020 in PubMed for the terms "angioedema" and "mutation" with the aim of retrieving other HAE genes from the literature. From this search, studies focusing on acquired forms of angioedema were ruled out, and an in-depth analysis of each prioritized study was performed. The VarSome database [12] was also screened to retrieve variants, as well as to update the genetic nomenclature of variant descriptors. The information described in the corresponding articles where each variant was described was manually inspected to verify that the original descriptions were accurate.

Figure 1. Schematic representation of the steps involved in gene and variant extraction, annotation, and database curation. All genetic variants reported in the articles studying hereditary angioedema (HAE) families in PubMed, HAEdb, VarSome, and the Leiden Open Variation Database (LOVD) were collected (n=502). ANNOVAR was used for annotation of frequencies, pathogenic predictors, and pathogenic classifications, among other information. Variants with a minor allele frequency below 1%, declared as pathogenic in HAEdb or as unknown but with a pathogenic prediction were kept as the set of known variants affecting function (n=450).
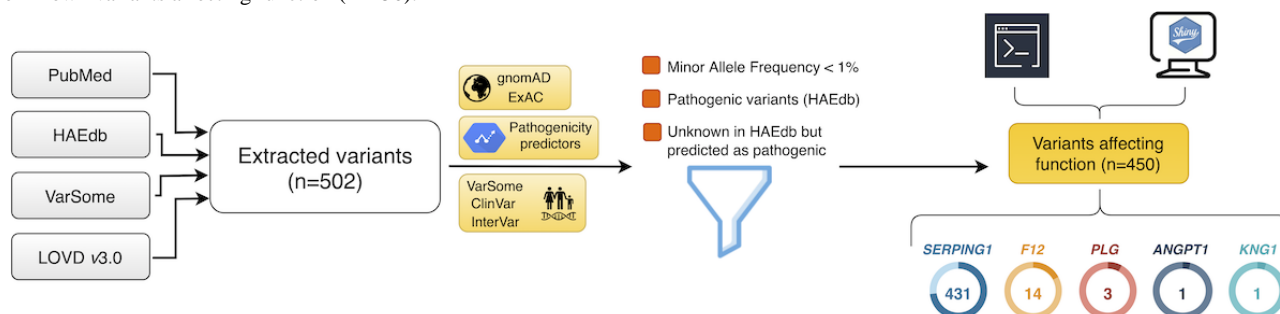
Given that a confident clinical interpretation of structural variants is challenging and that the standards for reporting those involving copy number variants have only recently been set [13], the criteria indicated above were only applied to single nucleotide variants (SNVs) and small insertion/deletions (indels). Therefore, the 45 gross mutations (ie, structural variants in the SERPING1 gene) from HAEdb that have been identified in HAE families were not integrated into HADA.

Database Variant Annotation

According to their chromosomal coordinates, the selected variants were first adapted to the GRCh37/hg19 reference

genome, exonic locations, Human Genome Variation Society (HGVS) nomenclature, coding effect, and PubMed citation. ANNOVAR v18.04.16 [14] was used to annotate according to RefGene, the allele frequency in gnomAD v2.1.1 [15] and ExAC [16] (November 29, 2015 release), dbSNP build 150 information, and precalculated pathogenicity predictors (SIFT, PolyPhen2, MutationTaster, CADD, DANN, MetaSVM, LRT, and phastCons mammalian). Pathogenic probabilities according to ClinPred [17] and the ACMG pathogenic classification as determined by ClinVar [18] (March 5, 2019 release), InterVar [19] (January 18, release), and VarSome (accessed June 13, 2020) were also annotated (Figure 1).

Database Curation

Many original descriptions of the genetic studies in HAE did not clearly declare the causality of the reported variants, thereby increasing the difficulty to identify variants with effects on the disease as opposed to variants without a disease effect [4]. To facilitate the interpretation of HADA results, we imposed the following filters for a variant to be designated as a variant affecting function (Figure 1): (1) variants described as pathogenic in HAEdb; (2) variants with unknown effects in HAEdb but predicted as pathogenic, likely pathogenic, or of uncertain significance (VUS) by VarSome, ClinVar, or InterVar (a flag will advise users in the case that contradictory classification information is reported for a variant); and (3) a reported minor allele frequency<1% in gnomAD, either on the overall sample or for non-Finish Europeans (because most of the genetic studies in HAE to date have focused on European families).

Finally, to facilitate the identification of potentially novel variants in the user's uploaded data, HADA also includes all SNVs and small indels that were not classified as a variant affecting function using the criteria indicated above from dbSNP build 150, ClinVar, and InterVar located within 50-bp flanking regions of all exons from known HAE genes.

Implementation of HADA

The database is built on MongoDB v4.2.1, a NoSQL query engine, to speed up database user queries and the variant calling file (VCF)-oriented analysis (Figure 2). HADA is built in Shiny v1.3.2, an R v3.6.1 package (R Foundation for Statistical Computing, Vienna, Austria) for building web apps. Specifically, we used ShinyJS v1.0 to run JavaScript code within the web app frontend and Plotly v4.9.0 to generate interactive plots. ANNOVAR v18.04.16 is used to provide annotations from the database in the uploads. To preserve potentially sensitive information included in the uploaded VCFs, HADA uses an encrypting code based on Cryfa [20] that automatically secures the access to the file and decrypts it once returned to the user. This provides a high level of security and transfers the data control to the user. No sensitive sample information is stored or maintained in the server.

Figure 2. Schematic representation of the HADA architecture and user interface. HADA is hosted on a server at TeideHPC, through a UI-based on Shiny. When the user sends a query to the server, it is encrypted and interrogates the HADA database, and returns matched variants affecting function associated with hereditary angioedema. The app also searches the data for other variants located within known hereditary angioedema genes to facilitate the identification of potentially novel variants. Once finished, an email with the link to download the annotated variant calling file (VCF) is sent to the user. At the end of the process, uploaded and annotated files are deleted from the server.
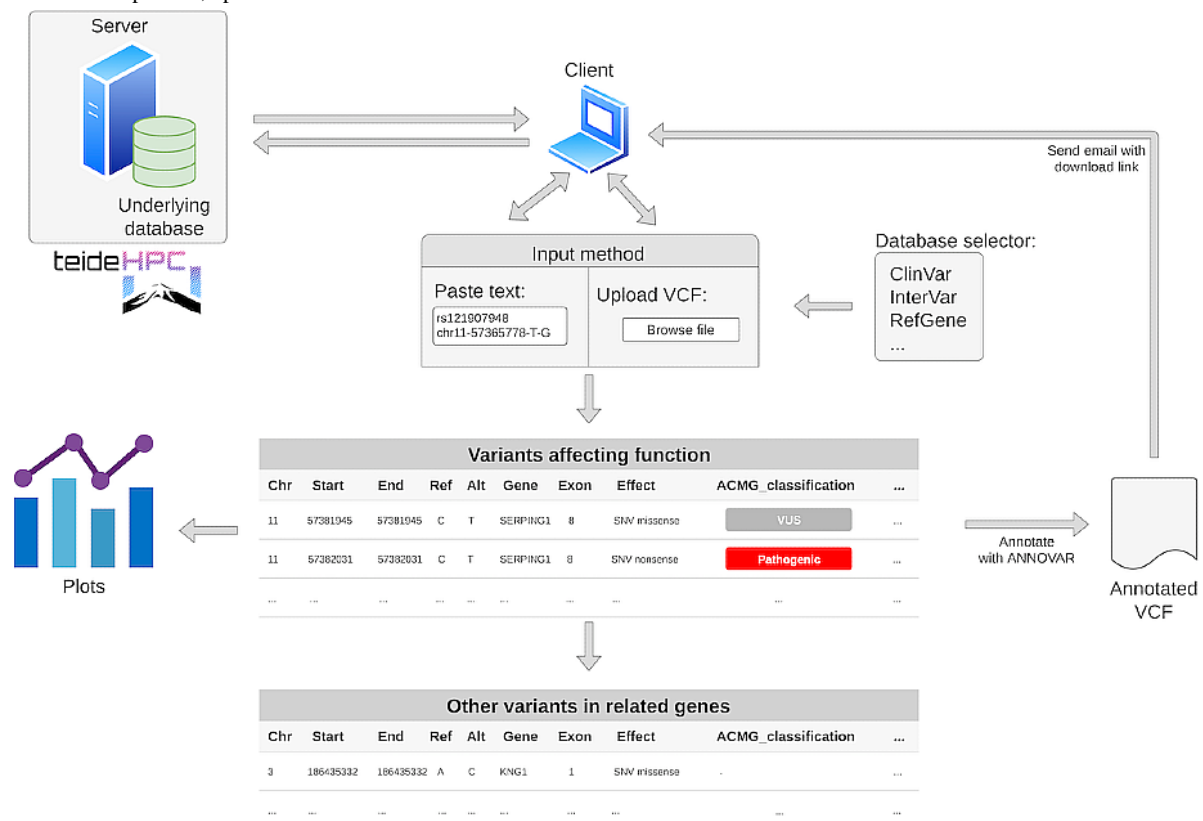
The curated database used by HADA is also available as a separate download at github [21] and it has been configured to annotate VCFs using ANNOVAR from the command line interface tool so that it can be easily incorporated into routine standalone NGS bioinformatics workflows. The HADA web interface is publicly accessible [22] hosted on TeideHPC premises [23].

Validation in an HAE Family

To validate HADA with comprehensive NGS data, we generated whole-exome sequencing (WES) data from a patient with HAE and his two family members recruited by the Allergy Service from the Hospital Universitario Nuestra Señora de Candelaria (HUNSC), Santa Cruz de Tenerife, Spain. The study was approved by the HUNSC Ethics Committee and written informed consent was obtained from the patients. The index patient was a 31-year-old male who visited the emergency department more than 5 times due to acute angioedema attacks with facial and cutaneous symptoms, manifesting as episodes affecting the upper airways from the age of 23 years. His biochemical blood test showed normal levels of C1-INH (47 mg/dL) based on the N Antisera to Human Coagulation Factors and C1 Inhibitor kit (Siemens Healthcare Diagnostics, Marburg) and reduced C1-INH activity (10%) (Berichrom C1 inhibitor, Siemens Healthcare Diagnostics, Marburg, Germany), suggesting HAE type II. His mother (60 years old) and a sister (23 years old) reported no symptoms of HAE, and also consented to participate in the study.

Briefly, sequencing libraries were prepared from DNA extracted using a commercial column-based kit (GFX kit, GE Healthcare, Little Chalfont, UK) using Nextera DNA Exome Kit (Illumina Inc, San Francisco, CA). The TapeStation 4200 system (Agilent Technologies, Santa Clara, CA) was used for library size estimation and the concentration was determined by the Qubit dsDNA HS Assay (Thermo Fisher Scientific, Waltham, MA). Libraries were sequenced on a HiSeq 4000 Sequencing System (Illumina Inc) with paired-end 75-base reads. Libraries were sequenced along with 1% of PhiX control V3 to an average depth of 50X after removal of duplicate reads. Sequencing reads were preprocessed with bcl2fastq v2.18 and mapped to hg19/GRCh37 with the Burrows-Wheeler Aligner 0.7.15-r1140 [24]. Resulting BAM files were processed with Qualimap v2.2.1 [25], SAMtools v1.3 [26], and Picard v2.10.10 [27]. Variant calling was performed using HaplotypeCaller following the Best Practices workflow recommendations for germline variant calling in GATK (v3.8) [28], obtaining one VCF for each sequenced individual. A high confidence set of variants was obtained after applying the following filters: variants with PASS, read depth$\geq$20, genotype quality$\geq$100, mapping quality$\geq$50. The transition-to-transversion nucleotide substitution ratio was also obtained as a quality control for the refined set of variants. Sequencing data and variant identification were obtained by the Instituto Tecnológico y de Energías Renovables (Santa Cruz de Tenerife, Spain).

Results

Overall Characteristics of the Database

A total of 502 variants from 5 genes causally linked to HAE were retrieved from PubMed, HAEdb, VarSome, and LOVD. A subset of 450 variants was designated as variants affecting function, including variants reported in the following 5 genes: *SERPING1* (n=431), *F12* (n=14), *PLG* (n=3), *ANGPT1* (n=1), and *KNG1* (n=1) (Figure 1). Three of these genes were discovered in the last 2 years based on WES approaches in HAE patients without C1-INH defects but not carrying *F12* variants affecting function [29-31]. ClinVar offered very limited information on this set of variants affecting function, as only 34 of them (7.6%) had corresponding ACMG class assignment: 2 are reported as benign or likely benign, 6 are indicated as VUS, and 26 are classified as pathogenic and likely pathogenic. InterVar included information for half of the set (226/450, 50.2%). However, 171 (75.7%) of these were classified as VUS. VarSome was the only resource that allowed assigning ACMG classes to all retrieved variants affecting function. According to VarSome, most of the HAE variants affecting function are classified as pathogenic (183/450, 40.6%) or likely pathogenic (171/450, 38.0%) (Figure 3). Although VarSome did not classify any of the HAE variants affecting function as benign or likely benign, 96 of them (21.3%) were still reported as VUS. Besides, precalculated pathogenicity predictors were available for a mean of 243 of the variants affecting function in the database. Taken together, these results highlight the existing gap in current interpretations of variant pathogenicity [32,33].

Figure 3. Precalculated pathogenic scores by the American College of Medical Genetics and Genomics (ACMG) pathogenic class for variants affecting function recorded in HADA. Panel A: VarSome proportions of ACMG classes among the hereditary angioedema variants affecting function. Panels B–F: pathogenicity prediction scores by ACMG pathogenic classes as provided by VarSome.
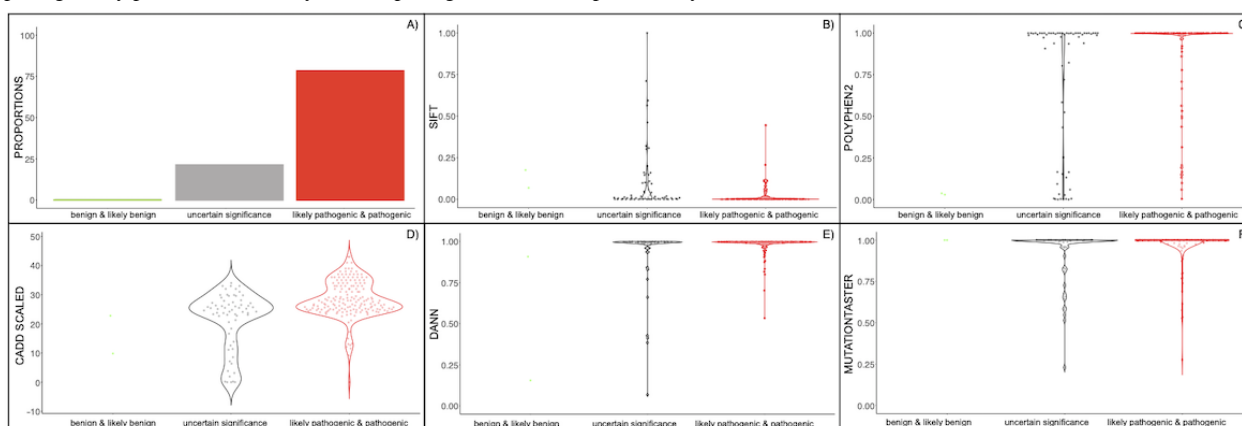
HADA Interface and Usage

HADA has been developed as a user-friendly graphical web-interface tool. It is designed with the objective of enabling the analysis of genetic variants regardless of the detection approach used for the screening of the genetic causes of HAE (Figure 4). As such, HADA is compatible with individual VCFs that have been obtained at any scale (gene panel, exome, or whole genome) and by any NGS technology. Alternatively, the user could opt to provide hg19 chromosome coordinates of the variant(s) of interest as the rsID number, HGVS coordinate, or amino acid change, among others. The latter is still a common standard in laboratories relying on Sanger sequencing results. All of the annotation information used to construct the underlying database can be selected by the user to extract the available information. To facilitate a prospective analysis of each query variant, references to the corresponding articles reporting the variant are also offered (in chronological order). Direct links to external tools (ie, VarSome), which allows accessing graphical information of gene transcripts, complement the possibilities of HADA. Despite the fact that sensitive patient information is not necessary for HADA to provide results, HADA integrates an automatic data encryption algorithm for the queries based on VCF data, ensuring a password-secured encryption of the input data while uploaded and of the decryption when results are downloaded. Furthermore, the information provided by the user is not stored permanently by the server.

The data processing workflow is shown in Multimedia Appendix 1. Once the data are uploaded, the variants contained in the VCF or that were provided by the user in chromosomal coordinates are matched against HADA. Automated detection and information extraction occur on the fly. Matched variants are subsequently shown in tabular format, accompanied by information previously selected by the user. HADA provides the ACMG classifications for each variant according to a color key to facilitate the identification of the HAE variants affecting function in the queries. HADA also suggests the subtype of HAE for which the variant affecting function is involved.

We anticipate that a user query might involve variants that have not yet been described among HAE genes (ie, novel HAE-related variants) or that have not been described in the scientific literature as an affecting functional variant to date. HADA integrates a genomic coordinates file that allows for the retrieval of predictions for any variant with any reference/alternative allele combination in any exonic region (plus a 50-bp flanking region on both ends of each locus) of the known HAE genes. Query results provided by HADA can be downloaded by the user in plain-text format in a comma-separated file, or as an annotated VCF in cases in which a VCF was uploaded by the user.

Finally, publication-ready plot generation and downloading are also incorporated in HADA. These functions plot graphs that summarize the information obtained from input data (ie, a VCF), as well as charts summarizing the proportions of the ACMG pathogenic classes predicted by VarSome, ClinVar, and InterVar separately, and the subtype of HAE that is associated with a variant affecting function that is identified in the query.

Figure 4. Selected snapshots of the HADA graphical interface with step-by-step instructions. Users can query variants of interest or alternatively upload an individual variant calling file (VCF) in the home page (A). The detected variants tab (B) shows the existence of variants affecting function in the query, as well as other variants in hereditary angioedema (HAE) genes. Plots are also generated and available for users to download (C).

XSL•FO
RenderX
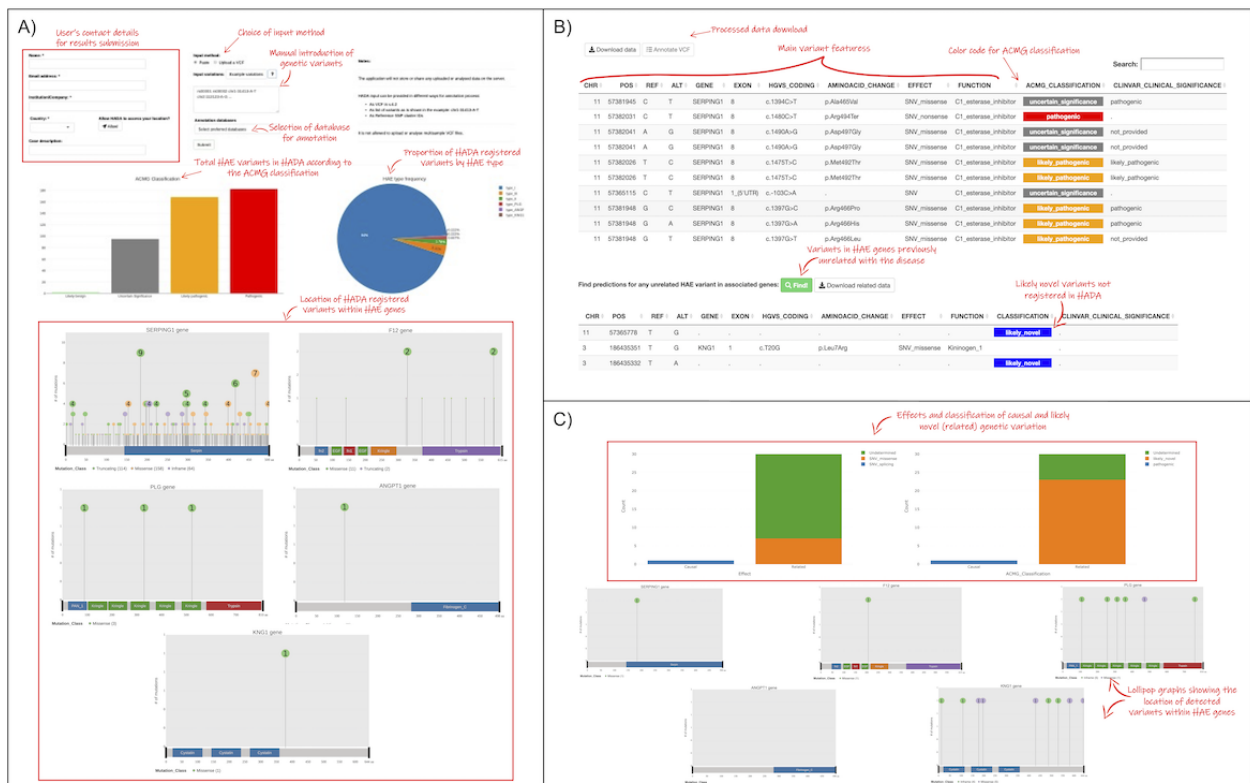
Validation of HADA in an HAE-Affected Family

To demonstrate the utility of HADA for the NGS-based identification of the variant affecting function involved in an HAE type II family, WES data (45 Mb) from the index patient, and the unaffected mother and sister were obtained. This detected a mean of 16,650 high-confidence variants per individual. Sequencing metrics and variant calling results are shown in Table 1. Resulting individual VCFs were processed with HADA, which reported a variant c.1396C>A (rs28940870) within *SERPING1* both in the index case and the sister that was heterozygous in both cases (compatible with an autosomal dominant inheritance pattern). The depth of coverage at this locus was 72X and 69X, respectively.

Table 1. Summary of the sequencing results in the validation study.

Feature	Index	Sister	Mother
Median insert size (bp[a])	297	268	268
Total reads (millions)	130.4	100.5	148.7
Aligned reads (%)	99.5	100	100
Mean coverage (%)	49.3	44.3	54.2
Targeted fraction ≥ 30X (%)	66	55	76
Ti/Tv[b]	3.16	3.19	3.22
Number of variants[c]	16,394	16,231	17,325
SNVs[d]	16,087	15,917	16,984
Indels[e]	307	314	341

[a]bp: base pairs.

[b]Ti/Tv: ratio of transitions to transversions.

[c]High confidence variants (FILTER=PASS; total depth≥20; genotype quality≥100; and mapping quality ≥ 50).

[d]SNVs: single nucleotide variants.

[e]Indels: insertion/deletion.

Most pathogenic prediction scores provided by HADA supported the variant as deleterious/damaging, and ACMG class predictors classified it either as a pathogenic/likely pathogenic variant or VUS (Table 2). This variant predicts an amino acid change (p.Arg466Ser) affecting the catalytic center of the protein, which has been previously described in independent HAE type II cases and was shown to be responsible for reduced C1-INH activity [34-36]. These observations are fully compatible with the clinical and biochemical findings supporting a diagnosis of HAE type II in the index case. Surprisingly, the sister was asymptomatic and had never experienced angioedema attacks, which further exemplifies the incomplete penetrance of HAE or that disease onset has not yet occurred. In any case, the provided genetic information could be used to anticipate potential clinical symptoms.

Table 2. American College of Medical Genetics and Genomics (ACMG) class predictions and pathogenic scores provided by HADA for the c.1396C>A variant found in a hereditary angioedema–affected family.

Predictor	Estimation
ACMG class prediction	
VarSome	VUS[a]
ClinVar	Pathogenic
InterVar	Likely pathogenic
Pathogenic predictors	
SIFT[b]	Damaging
Polyphen2[c]	Benign
MutationTaster	Disease-causing
CADD[d]	31
DANN[e] score	0.997
DANN rankscore	0.798
LRT[f]	Deleterious
MetaSVM	Damaging

[a]VUS: variant of uncertain significance.

[b]SIFT: Sorting Intolerant From Tolerant.

[c]PolyPhen2: Polymorphism Phenotyping v2.

[d]CADD: Combined Annotation Dependent Depletion.

[e]DANN: Deleterious Annotation of genetic variants using Neural Networks.

[f]LRT: likelihood ratio test.

[g]MetaSVM: meta-analytic support vector machine.

Discussion

Principal Findings

Here, we present HADA, a web-based analysis tool to facilitate the identification of the genetic variants causing HAE. With HADA, we aimed to provide a user-friendly tool to assist in the diagnosis of HAE that is adapted to NGS technologies and to the evolving knowledge of the causes of HAE. HADA has potential to reduce the time to interpret the detected variants in the NGS era by aggregating data from multiple sources, a process that commonly takes several hours to complete if it is handled manually [37,38]. At the moment, HADA integrates information from 450 SNVs and indels from 5 genes (*SERPING1*, *F12*, *PLG*, *ANGPT1*, and *KNG1*) that we classified as variants likely affecting function. While curating this information, we found that variant descriptions in HAE cases did not follow a standard [4], and many of the studies did not clearly declare the causality of the reported variants [35,39]. This situation helps to explain why up to 21.3% (n=96) of the simple bona fide variants affecting function continue to be reported as VUS in the best case. Finally, we demonstrated the utility of HADA for explaining HAE type II in a family that was assessed by WES in three family members, and offered conclusive information about the existence of the previously described variant c.1396C>A (p.Arg466Ser) in *SERPING1*.

As is the case for many other rare diseases, genetic testing in HAE has become an important step to reduce the diagnostic odyssey [40], increase the diagnostic yield, and tailor treatments [4,5]. The widespread adoption of NGS technology in clinical settings has led to the emergence of a wide variety of bioinformatics tools to assist and accelerate the detection and interpretation of associated genetic variants and their impact on disease risks. In recent years, an array of variant prioritizers and interpreters have been developed to obtain optimal rankings for the variants causing the disease. However, NGS-based solutions have not been adopted in the HAE community until recently [29-31]. Part of the explanation may reside in the fact that only two causal genes were known until 2018, and therefore most clinical diagnoses could only be made based on clinical symptoms and the biochemical measurements of C4 and C1-INH levels in plasma, along with C1-INH activity [5]. Importantly, public resource updates to assist in variant interpretation lag behind the pace of genetic discoveries. This is the case of Simply-ClinVar [41] or ClinVar, which are still outdated with respect to the genetic causes of HAE, reporting information for only two of the associated genes (*SERPING1* and *F12*). Simply-ClinVar also suggests *SLC34A1* as another HAE gene, although this evidence is unsupported by the current literature. This situation is changing currently, and the benefits of using NGS technologies to assess multiple genes simultaneously for HAE diagnosis are now clearer [5]. In fact, in patients with normal levels and activity of C1-INH, genetic testing is recommended for the routine diagnosis of many HAE subtypes. Furthermore, according to Germenis et al [4] and the international consensus on genetic aspects of HAE, it is widely

recognized that C4 and C1-INH plasma measurements can generate inconclusive results even in patients with genetic defects in C1-INH. Based on this, HADA will surely help practitioners to adapt to the NGS-based genetic screens of HAE cases.

HADA has some strengths and limitations for the interpretation of genetic variants involved in the causes of HAE. Among the limitations, the tool does not currently assess VCFs from families or trios and does not allow inferring whether a variant affecting function has a de novo origin from the patient's data. Similarly, HADA uses GRCh37/hg19 coordinates, which is still considered to be the gold standard in clinical settings [42]. Despite the fact that HADA can report the existence of variants residing in introns near the key elements for splicing, as has been recently found for a type I HAE case [43,44], novel variants affecting function residing deep within intron positions will remain undetected. The main reason is the limited capacity of current algorithms to predict the pathogenic potential of deep intronic variants [45]. Similarly, a variable proportion of information remains unavailable for many of the variants affecting function collected in the underlying database. For example, allelic frequencies are currently available from gnomAD and ExAC for as few as 11 to 20 (2%-4%) of the variants affecting function in HAE. As an example, the variant detected in the family analyzed in this study is a bona-fide causal variant of HAE. However, there are no records of this variant in the gnomAD or ExAC databases. Similarly, not all variants affecting function in HAE are described in current ClinVar or InterVar versions. Collectively, this situation adds to the challenge of better predicting disease effects of many of the variants that are declared as VUS.

Structural variants have an important role in HAE due to the presence of Alu element–related reorganizations affecting function in *SERPING1*, potentially causing HAE type I [46,47]. However, structural variants will not be interpreted by HADA at the moment as they have not been collected in the database. One of the main reasons for this is that, unlike the case for SNVs and small indels, current locus-oriented databases such as HAEdb do not provide the chromosomal coordinates of recorded structural variants from HAE families, making it difficult to integrate in HADA.

Among the strengths of HADA, we highlight the versatility of the tool to allow performing simple queries or to directly upload VCFs. This offers the possibility of analyzing a variant detected at any throughput scale, using either NGS technologies or Sanger sequencing. This feature is a great advantage compared to HAEdb or LOVD, which are not adapted for the automated analysis of NGS data (eg, from a VCF), serving only for single-variant queries, and neither of these previous tools has been subjected to a curation process or the inclusion of novel HAE genes. In addition to the imposed filters to construct the curated database of HADA, allele frequencies are also provided for all populations that are currently collected in the gnomAD and ExAC population databases. To cover the possibility of new discoveries, HADA also includes information from all other described variants in the coding regions, the 5′ and 3′ untranslated regions, and the 50-bp flanking regions (where most splicing variants affecting function are located) to easily flag novel variants in the uploaded data. Finally, despite the fact that its use does not require sensitive information for the analysis, HADA automatically conducts a password-secured encryption of the data and does not permanently store any data on the server.

We anticipate that HADA will be constantly updated to register new variants affecting function and associated information, changes in the pathogenic classification, and the possibility to assess structural variants, and GRCh38/hg38 coordinate and reference sequence number conversions. This will be based on third-party requests and internal updates according to novel discoveries and standards.

Conclusions

To adapt the genetic diagnosis of HAE to the era of NGS-based genomic medicine, we have developed HADA as a free and publicly available tool for simplifying the identification of simple variants affecting function in HAE. HADA will allow users to focus on biologically relevant questions instead of having to learn to install software dependencies, variant annotation tools, and become familiar with the UNIX command line. The main advantages of HADA are that it is focused on a disease, its ease of use, the ability to display specific and curated information of HAE either from individual or VCF queries, and that it is freely available. By combining these features into a single graphical and interactive tool, we expect that variant prioritization in HAE will become easier, faster, and standardized.

Authors' Contributions

AA, AB, and CF wrote the manuscript and designed the figures. AA, LR, IR, and CF built and curated the database. ArC and JR contributed information for HAE clinical aspects and sample collection of the tested family. AlC, RM, AC, JS, and AA managed

and sequenced the DNA samples of the family tested and performed the variant calling. AB, JS, and AA developed the front end. CF designed the project and obtained funding. All the authors revised and approved the final version of the manuscript.

Conflicts of Interest

Takeda Pharmaceutical Company funded a travel grant to AA. The rest of the authors declare no conflicts of interest.

Multimedia Appendix 1

Schematic representation of the variant calling file (VCF) variant matching against the curated database and data downloading in HADA.
[PNG File , 1417 KB-Multimedia Appendix 1]

References

1. Nzeako UC, Frigas E, Tremaine WJ. Hereditary angioedema: a broad review for clinicians. Arch Intern Med 2001 Nov 12;161(20):2417-2429. [doi: 10.1001/archinte.161.20.2417] [Medline: 11700154]

2. Agostoni A, Aygören-Pürsün E, Binkley KE, Blanch A, Bork K, Bouillet L, et al. Hereditary and acquired angioedema: problems and progress: proceedings of the third C1 esterase inhibitor deficiency workshop and beyond. J Allergy Clin Immunol 2004 Sep;114(3 Suppl):S51-131 [FREE Full text] [doi: 10.1016/j.jaci.2004.06.047] [Medline: 15356535]

3. Caballero T, Baeza ML, Cabañas R, Campos A, Cimbollek S, Gómez-Traseira C, Spanish Study Group on Bradykinin-Induced Angioedema (SGBA). Consensus statement on the diagnosis, management, and treatment of angioedema mediated by bradykinin. Part II. Treatment, follow-up, and special situations. J Investig Allergol Clin Immunol 2011;21(6):422-441; quiz 442 [FREE Full text] [Medline: 21995176]

4. Germenis AE, Margaglione M, Pesquero JB, Farkas H, Cichon S, Csuka D, Hereditary Angioedema International Working Group. International Consensus on the Use of Genetics in the Management of Hereditary Angioedema. J Allergy Clin Immunol Pract 2020 Mar;8(3):901-911. [doi: 10.1016/j.jaip.2019.10.004] [Medline: 31669336]

5. Marcelino-Rodriguez I, Callero A, Mendoza-Alvarez A, Perez-Rodriguez E, Barrios-Recio J, Garcia-Robaina JC, et al. Bradykinin-Mediated Angioedema: An Update of the Genetic Causes and the Impact of Genomics. Front Genet 2019;10:900. [doi: 10.3389/fgene.2019.00900] [Medline: 31611908]

6. Loules G, Zamanakou M, Parsopoulou F, Vatsiou S, Psarros F, Csuka D, et al. Targeted next-generation sequencing for the molecular diagnosis of hereditary angioedema due to C1-inhibitor deficiency. Gene 2018 Aug 15;667:76-82. [doi: 10.1016/j.gene.2018.05.029] [Medline: 29753808]

7. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet 2014 Sep;30(9):418-426. [doi: 10.1016/j.tig.2014.07.001] [Medline: 25108476]

8. Kalmár L, Hegedüs T, Farkas H, Nagy M, Tordai A. HAEdb: a novel interactive, locus-specific mutation database for the C1 inhibitor gene. Hum Mutat 2005 Jan;25(1):1-5. [doi: 10.1002/humu.20112] [Medline: 15580551]

9. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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):405-424. [doi: 10.1038/gim.2015.30] [Medline: 25741868]

10. Ponard D, Gaboriaud C, Charignon D, Ghannam A, Wagenaar-Bos IGA, Roem D, et al. SERPING1 mutation update: Mutation spectrum and C1 Inhibitor phenotypes. Hum Mutat 2020 Jan;41(1):38-57. [doi: 10.1002/humu.23917] [Medline: 31517426]

11. Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat 2011 May;32(5):557-563. [doi: 10.1002/humu.21438] [Medline: 21520333]

12. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. Bioinformatics 2019 Jun 01;35(11):1978-1980 [FREE Full text] [doi: 10.1093/bioinformatics/bty897] [Medline: 30376034]

13. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med 2020 Feb;22(2):245-257. [doi: 10.1038/s41436-019-0686-8] [Medline: 31690835]

14. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010 Sep;38(16):e164 [FREE Full text] [doi: 10.1093/nar/gkq603] [Medline: 20601685]

15. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Genome Aggregation Database Consortium, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020 May;581(7809):434-443 [FREE Full text] [doi: 10.1038/s41586-020-2308-7] [Medline: 32461654]

16. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016 Aug 18;536(7616):285-291 [FREE Full text] [doi: 10.1038/nature19057] [Medline: 27535533]

17. Alirezaie N, Kernohan KD, Hartley T, Majewski J, Hocking TD. ClinPred: Prediction Tool to Identify Disease-Relevant Nonsynonymous Single-Nucleotide Variants. Am J Hum Genet 2018 Oct 04;103(4):474-483 [FREE Full text] [doi: 10.1016/j.ajhg.2018.08.005] [Medline: 30220433]

18. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 2014 Jan;42(Database issue):D980-D985 [FREE Full text] [doi: 10.1093/nar/gkt1113] [Medline: 24234437]

19. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet 2017 Feb 02;100(2):267-280 [FREE Full text] [doi: 10.1016/j.ajhg.2017.01.004] [Medline: 28132688]

20. Hosseini M, Pratas D, Pinho AJ. Cryfa: a secure encryption tool for genomic data. Bioinformatics 2019 Jan 01;35(1):146-148 [FREE Full text] [doi: 10.1093/bioinformatics/bty645] [Medline: 30020420]

21. HADA download. github. 2020. URL: https://github.com/genomicsITER/HADA [accessed 2020-09-22]

22. HADA interface. URL: http://hada.hpc.iter.es/ [accessed 2020-09-22]

23. Teide HPC. URL: https://teidehpc.iter.es/ [accessed 2020-06-15]

24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009 Jul 15;25(14):1754-1760 [FREE Full text] [doi: 10.1093/bioinformatics/btp324] [Medline: 19451168]

25. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, et al. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 2012 Oct 15;28(20):2678-2679. [doi: 10.1093/bioinformatics/bts503] [Medline: 22914218]

26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009 Aug 15;25(16):2078-2079 [FREE Full text] [doi: 10.1093/bioinformatics/btp352] [Medline: 19505943]

27. Picard. URL: http://broadinstitute.github.io/picard [accessed 2020-09-22]

28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010 Sep;20(9):1297-1303. [doi: 10.1101/gr.107524.110] [Medline: 20644199]

29. Bafunno V, Firinu D, D'Apolito M, Cordisco G, Loffredo S, Leccese A, et al. Mutation of the angiopoietin-1 gene (ANGPT1) associates with a new type of hereditary angioedema. J Allergy Clin Immunol 2018 Mar;141(3):1009-1017. [doi: 10.1016/j.jaci.2017.05.020] [Medline: 28601681]

30. Bork K, Wulff K, Steinmüller-Magin L, Braenne I, Staubach-Renz P, Witzke G, et al. Hereditary angioedema with a mutation in the plasminogen gene. Allergy 2018 Feb;73(2):442-450. [doi: 10.1111/all.13270] [Medline: 28795768]

31. Bork K, Wulff K, Rossmann H, Steinmüller-Magin L, Braenne I, Witzke G, et al. Hereditary angioedema cosegregating with a novel kininogen 1 gene mutation changing the N-terminal cleavage site of bradykinin. Allergy 2019 Dec;74(12):2479-2481. [doi: 10.1111/all.13869] [Medline: 31087670]

32. Bean LJH, Hegde MR. Clinical implications and considerations for evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. Genome Med 2017 Dec 18;9(1):111 [FREE Full text] [doi: 10.1186/s13073-017-0508-z] [Medline: 29254502]

33. Niroula A, Vihinen M. How good are pathogenicity predictors in detecting benign variants? PLoS Comput Biol 2019 Feb;15(2):e1006481 [FREE Full text] [doi: 10.1371/journal.pcbi.1006481] [Medline: 30742610]

34. Aulak KS, Cicardi M, Harrison RA. Identification of a new P1 residue mutation (444Arg—>Ser) in a dysfunctional C1 inhibitor protein contained in a type II hereditary angioedema plasma. FEBS Lett 1990 Jun 18;266(1-2):13-16. [doi: 10.1016/0014-5793(90)81494-9] [Medline: 2365061]

35. Xu Y, Zhi Y, Yin J, Wang L, Wen L, Gu J, et al. Mutational spectrum and geno-phenotype correlation in Chinese families with hereditary angioedema. Allergy 2012 Nov;67(11):1430-1436. [doi: 10.1111/all.12024] [Medline: 22994404]

36. Gösswein T, Kocot A, Emmert G, Kreuz W, Martinez-Saguer I, Aygören-Pürsün E, et al. Mutational spectrum of the C1INH (SERPING1) gene in patients with hereditary angioedema. Cytogenet Genome Res 2008;121(3-4):181-188. [doi: 10.1159/000138883] [Medline: 18758157]

37. Harrison SM, Dolinsky JS, Knight Johnson AE, Pesaran T, Azzariti DR, Bale S, et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. Genet Med 2017 Oct;19(10):1096-1104. [doi: 10.1038/gim.2017.14] [Medline: 28301460]

38. Christensen KD, Vassy JL, Phillips KA, Blout CL, Azzariti DR, Lu CY, MedSeq Project. Short-term costs of integrating whole-genome sequencing into primary care and cardiology settings: a pilot randomized trial. Genet Med 2018 Dec;20(12):1544-1553. [doi: 10.1038/gim.2018.35] [Medline: 29565423]

39. Martinho A, Mendes J, Simões O, Nunes R, Gomes J, Dias Castro E, et al. Mutations analysis of C1 inhibitor coding sequence gene among Portuguese patients with hereditary angioedema. Mol Immunol 2013 Apr;53(4):431-434. [doi: 10.1016/j.molimm.2012.09.003] [Medline: 23123409]

40. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, et al. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am J Hum Genet 2017 May 04;100(5):695-705 [FREE Full text] [doi: 10.1016/j.ajhg.2017.04.003] [Medline: 28475856]

41. Pérez-Palma E, Gramm M, Nürnberg P, May P, Lal D. Simple ClinVar: an interactive web server to explore and retrieve gene and disease variants aggregated in ClinVar database. Nucleic Acids Res 2019 Jul 02;47(W1):W99-W105 [FREE Full text] [doi: 10.1093/nar/gkz411] [Medline: 31114901]

42. Guo Y, Dai Y, Yu H, Zhao S, Samuels DC, Shyr Y. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. Genomics 2017 Mar;109(2):83-90 [FREE Full text] [doi: 10.1016/j.ygeno.2017.01.005] [Medline: 28131802]

43. Vatsiou S, Zamanakou M, Loules G, Psarros F, Parsopoulou F, Csuka D, et al. A novel deep intronic SERPING1 variant as a cause of hereditary angioedema due to C1-inhibitor deficiency. Allergol Int 2020 Jan 17:online ahead of print [FREE Full text] [doi: 10.1016/j.alit.2019.12.009] [Medline: 31959500]

44. Hujová P, Souček P, Grodecká L, Grombiříková H, Ravčuková B, Kuklínek P, et al. Deep Intronic Mutation in SERPING1 Caused Hereditary Angioedema Through Pseudoexon Activation. J Clin Immunol 2020 Apr;40(3):435-446. [doi: 10.1007/s10875-020-00753-2] [Medline: 31982983]

45. Lord J, Gallone G, Short PJ, McRae JF, Ironfield H, Wynn EH, Deciphering Developmental Disorders study. Pathogenicity and selective constraint on variation near splice sites. Genome Res 2019 Feb;29(2):159-170. [doi: 10.1101/gr.238444.118] [Medline: 30587507]

46. López-Lera A, Garrido S, Roche O, López-Trascasa M. SERPING1 mutations in 59 families with hereditary angioedema. Mol Immunol 2011 Oct;49(1-2):18-27. [doi: 10.1016/j.molimm.2011.07.010] [Medline: 21864911]

47. Germenis AE, Speletas M. Genetics of Hereditary Angioedema Revisited. Clin Rev Allergy Immunol 2016 Oct;51(2):170-182. [doi: 10.1007/s12016-016-8543-x] [Medline: 27116602]

Abbreviations

ACMG: American College of Medical Genetics and Genomics
C1-INH: C1 esterase inhibitor
HADA: Hereditary Angioedema Database Annotation
HAE: hereditary angioedema
HGVS: Human Genome Variation Society
HUNSC: Hospital Universitario Nuestra Señora de Candelaria
indel: insertion/deletion
LOVD: Leiden Open Variant Database
NGS: next-generation sequencing
SNV: single nucleotide variant
VCF: variant calling file
VUS: variant of uncertain significance
WES: whole-exome sequencing