Check for updates

Interpreting the spectrum of gamma-secretase complex missense variation in the context of hidradenitis suppurativa—An *in-silico* study

Dillon Mintoff[1,2], Nikolai P. Pace[3,4]* and Isabella Borg[1,3,5]

[1]Department of Pathology, Faculty of Medicine and Surgery, University of Malta, Msida, Malta, [2]Centre for Molecular Medicine and Biobanking, University of Malta, Msida, Malta, [3]Centre for Molecular Biology and Biobanking, University of Malta, Msida, Malta, [4]Department of Anatomy, Faculty of Medicine and Surgery, University of Malta, Msida, Malta, [5]Department of Pathology, Mater Dei Hospital, Msida, Malta

Hidradenitis suppurativa (HS) is a disease of the pilosebaceous unit characterized by recurrent nodules, abscesses and draining tunnels with a predilection to intertriginous skin. The pathophysiology of HS is complex. However, it is known that inflammation and hyperkeratinization at the hair follicle play crucial roles in disease manifestation. Genetic and environmental factors are considered the main drivers of these two pathophysiological processes. Despite a considerable proportion of patients having a positive family history of disease, only a minority of patients suffering from HS have been found to harbor monogenic variants which segregate to affected kindreds. Most of these variants are in the ɣ secretase complex (GSC) protein-coding genes. In this manuscript, we set out to characterize the burden of missense pathogenic variants in healthy reference population using large scale genomic dataset thereby providing a standard for comparing genomic variation in GSC protein-coding genes in the HS patient cohort.

KEYWORDS

hidradenitis suppurativa, genetics, gamma secretase complex, pathophysiology, nicastrin

Introduction

Hidradenitis suppurativa (HS) is chronic inflammatory disease of the pilosebaceous unit (PSU) characterized by recurrent nodules, abscesses and draining tunnels. These lesions can manifest in any part of the body but have a predilection to intertriginous zones such as the axillae and groin. The primary pathogenic mechanisms driving HS are inflammation and hyperkeratinization (Zouboulis et al., 2020). Several environmental and lifestyle factors play a role in disease development and progression, with obesity and smoking having strong, well-established associations with the disease (König et al., 1999; Kromann et al., 2014). In addition, HS has an underlying genetic association, with

significant narrow-sense heritability estimates (77%) being reported in the literature and common forms of HS demonstrating strong familial segregation (Straalen et al., 2020). Despite several reports implicating ɣ secretase complex (GSC) variants in HS, evidence for direct causal mechanisms is lacking (Gao et al., 2006; Sellheyer and Krahl, 2008; Al-Ali et al., 2010; Miskinyte et al., 2012; Jiao et al., 2013; Nomura et al., 2014; Zhang and Sisodia, 2015; Pink et al., 2016; Ratnamala et al., 2016). For example, *in vitro* studies of variants presumed to be pathogenic, have been demonstrated to lack functional significance in relation to disease-associated pathways (Zhang and Sisodia, 2015). Conflicting results have also been obtained with regards to differential expression of genes coding for proteins of the GSC in serum and skin (Wang et al., 2010; Zouboulis et al., 2021). Therefore, despite a putative monogenic disease driver in a subset of HS patients harboring pathogenic variants, predominantly in genes coding for the protein subunits of the ɣ secretase complex (GSC) (Pace et al., 2022), the exact nature of the genetic risk in the disease remains elusive.

The GSC is a member of the intramembrane-cleavage protease (I-CliP) family that cleaves several protein substrates (Wolfe et al., 1999; Merilahti et al., 2017). The most established pathogenic role of the GSC is in familial Alzheimer's disease (fAD) through dysregulation of amyloid precursor protein (APP) cleavage and the generation of amyloid ß in neural tissues (Hur, 2022). Interestingly, studies have shown that the prevalence of AD is increased four-fold in patients having a family history of HS, although no mechanistic explanation for this observation is available (Esme et al., 2020). On the other hand, the role of GSC activity in non-neural tissues such as keratinocytes and cutaneous fibroblasts is less clear. A putative mechanism implicates dysregulation of NOTCH pathway signaling in the pathophysiology of HS (Melnik and Plewig, 2013). However conflicting results have been obtained in translational studies (Frew, 2019; Frew and Navrazhina, 2020). Importantly, the molecular etiology of HS and fAD do not directly overlap. Of the GSC protein-coding genes, monogenic HS is mostly related to pathogenic variation in *NCSTN* (OMIM: #142690, Acne inversa, Familial 1), unlike fAD which is predominantly associated with *PSEN* mutations (Xiao et al., 2021; Pace et al., 2022). Additional loci involved in HS include *PSTPIP1*, *MEFV*, *NLRP3*, *IL1RN*, *NOD2* and *POFUT*, although their relevance to disease has been disputed. Despite the high proportion of cases with a family history of HS, a low prevalence of pathogenic GSC variation has been identified in sizeable patient cohorts with sporadic or familial forms of HS (Duchatelet et al., 2020). In addition, the role of common variants and polygenic risk scores in HS has not yet been evaluated.

Although the pathophysiological relevance of the GSC in humans has been extensively described, especially in Alzheimer's disease (Hur, 2022), the variety and molecular impact of coding variants in a healthy reference population has not been evaluated. In this manuscript we use a large-scale population genomic dataset derived from the Genome Aggregation Database (GnomAD) to map the frequency and spectrum of missense variants in the four genes (*NCSTN, PSEN1, PSENEN, APH1A*) encoding protein subunits of the GSC. Evolutionary conservation metrics and *in-silico* analysis of pathogenicity are assessed, and the impact of missense variants from GnomAD on protein stability and structure and inter-subunit interactions evaluated. Thus, we provide a concise summary of the burden of GSC coding variation and its impact on protein structure in a large dataset that is routinely employed as a reference population in clinical genetic studies. Importantly, rare coding variants having deleterious *in silico* predictor scores are classified as variants of uncertain significance in the absence of definitive proof of pathogenicity, and thus they often confound molecular diagnosis. These data thereby address an important, but yet unmet, area of research which can be utilized to facilitate the interpretation of GSC missense variation in the context of associated conditions such as HS. This allows for a discussion on the implications of the findings in endotyping for precision medicine.

Materials and methods

Variant identification

Genomic variation in the four loci of the GSC were extracted from GnomAD v2.1.1, a large publicly-available collection of population variation from harmonized sequence data (https://gnomad.broadinstitute.org/) (Karczewski et al., 2020). The database contains a collection of putative loss of function variants (pLoF) from 125,748 exomes (containing 17.2 million variants) and 15,708 genomes (containing 229.9 million variants) from 141,456 individuals. Of these, 443,769 are high confidence pLoF variants, the majority of which involve canonical transcripts of 16,694 genes. Genomic positions are described with reference to the GRCh37/hg19 reference sequence. GnomAD contains data representative of major global European, Asian, African, and African/American ethnicities, along with additional 'other' populations that do not cluster with these major ethnicities. The GnomAD v2 cohort participants span a broad age range and the database does not include duplicate individuals, first or second-degree relatives to minimize inflation of rare variants. We excluded variants in GnomAD having ClinVar entries to obtain a better representation of unaffected reference populations and limit inclusion of variants reported in patient samples. pLoF variants, in-frame insertion/deletions, synonymous variants and intronic/untranslated region variants were excluded. We thus selected to focus on missense variants due to their direct clinical relevance in view of relative abundance and diversity in the human genome as well as their capacity to alter structure-function relationships at the protein level.

The observed/expected (o/e) ratio is defined as a continuous measure of how tolerant a gene is to a class of variation. A low o/e value indicates stronger evolutionary selection for that class of variation than a gene with a higher value. The o/e ratio for missense variants was compared across the four genes of the GSC to evaluate evolutionary gene constraint. The observed and expected counts depend on both gene and sample size. Thus, the 90% confidence interval for the o/e value is vital for interpretation of this metric and is presented as the 90% upper bound of the CI (termed loss-of-function observed/expected upper bound fraction, LOEUF) (Karczewski et al., 2020).

Variant deleteriousness and conservation scoring metrics

The deleteriousness of filtered missense variants was analyzed using three *in silico* tools - Combined Annotation Dependent Deletion (CADD), PrimateAI and MetaSVM. CADD (https://cadd.gs.washington.edu/) integrates multiple annotations into a single metric (C-Score) by contrasting variants that survived natural mutation with simulated variants (Kircher et al., 2014). CADD prioritizes functional, deleterious and pathogenic variants. The resultant C-score correlates with allelic diversity and variant pathogenicity. These scores are transformed into a PHRED-like rank score relative to the genome wide distribution of scores across all potential single nucleotide variants in the human genome (8.6×10^9 variants) (Rentzsch et al., 2019). A scaled CADD score of ≥ 20 indicates that a variant is among the top 1% most deleterious genomic variation. PrimateAI is a trained deep residual network, which interprets amino acid sequence flaking a variant of interest and the orthologous sequence alignment in other species to classify the likelihood of pathogenicity of human variants having uncertain significance (Sundaram et al., 2018). The PrimateAI score ranges from 0 to 1, with variants having a score closer to 1 being more likely to be pathogenic. MetaSVM(Liu et al., 2016) is an ensemble score using Support Vector Machine (SVM) that integrates nine prediction scores (SIFT, PolyPhen-2, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy and PhyloP) and allele frequencies in the 1,000 genome project (1KGP) database. The tool has been trained on 36,192 single nucleotide variants (SNVs) from UniProt.

Conservation scores (Rate4Site) were calculated using the ConSurf webserver tool (https://consurf.tau.ac.il/). This server enables estimation of the evolutionary conservation of amino/nucleic acid positions in a macromolecule based on the phylogenetic relations between homologous sequences (Ashkenazy et al., 2016). The degree of conservation at a specific residue correlates to its structural and functional relevance. ConSurf provides accurate computation of

evolutionary rate by using an empirical Bayesian method. The tool normalizes scores so that the average score for all residues is zero whilst the standard deviation is one. Negative values represent slowly evolving, highly conserved sites with the most conserved position having a score of −1.

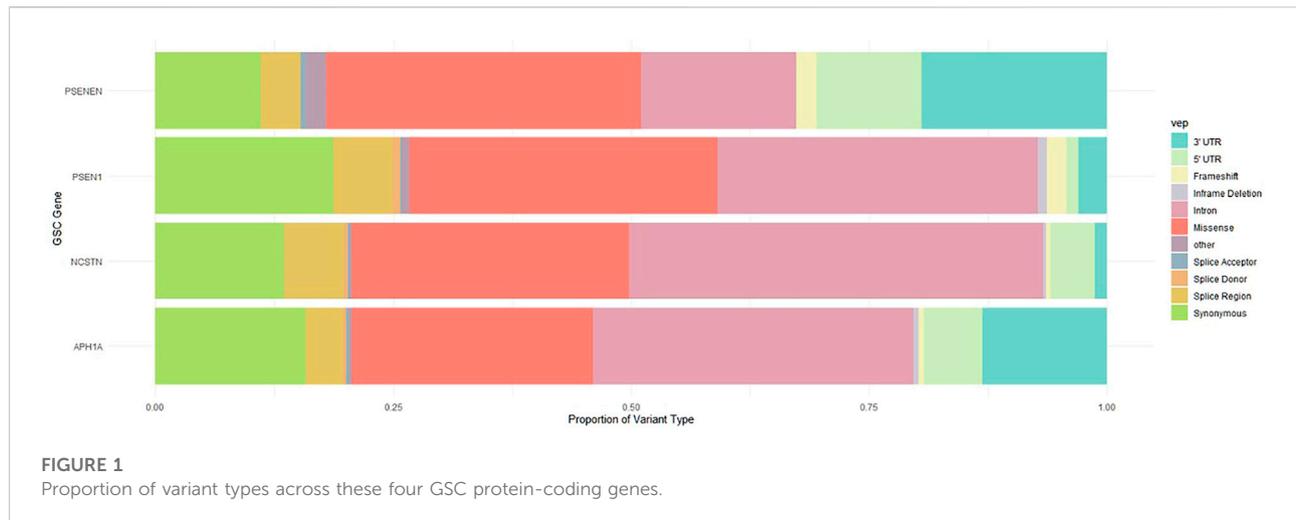## The effects of missense variants on protein stability and inter-subunit interaction

The filtered missense variants were cross-referenced to Missense3D-DB, a database of precomputed structural predictions for approximately 4 million human missense variants (Khanna et al., 2021). Missense3D enables evaluation of the impact of missense variants on protein structure. It assesses 17 different structural features that are essential for protein confirmation and stability, such as stearic clashes and disallowed phi/psi angles, to identify substitutions that are conformationally deleterious (Venselaar et al., 2010). VarSeq software (Golden Helix) was used for variant interpretation and annotation (Kopanos et al., 2019).

Protein structures as predicted by AlphaFold were utilized using the following identifiers; AF-Q96BI3-F1 (APH1A), AF-Q92542-F1 (Nicastrin), AF-P49768-F1 (PSEN1) and AF-Q9NZ42-F1 (PSENEN). Structural representations were generated using the PyMOL molecular visualization system (PyMOL, 2020). Protein domains were identified using InterPro Database14 and drawn using ProSite online software (https://prosite.expasy.org/). Protein domains were related to the distribution of evolutionary conservation and deleteriousness across primary structure as outlined above.

The estimated stability spectrum of all identified missense variants in GSC was evaluated using FoldX 5.1 (Schymkowitz et al., 2005). To determine how a single point substitution affects protein stability, $\Delta\Delta G$ (kcal/mol) was calculated. This metric quantifies the difference in the change of Gibbs-free energy and is a measure of the change in energy between the folded and unfolded states ($\Delta G_{folding}$) and the change in $\Delta G_{folding}$ when a missense variant is present. $\Delta\Delta G$ is thus a predictor of whether a missense variant will be energetically favorable in terms of protein stability. Using FoldX 5.1, the energies for the wild type ($\Delta G_{fold,WT}$) and the mutant ($\Delta G_{fold,MT}$) protein were computed to give the stability change ($\Delta\Delta G_{fold} = \Delta G_{fold,MT} - \Delta G_{fold,WT}$).

The change in Gibbs-free energy ($\Delta\Delta G_{folding}$ in kcal/mol) was categorized based on the following thresholds: highly stabilizing (≤ -1.84), stabilizing (-1.84 to -0.92), slightly stabilizing (-0.92 to -0.46), neutral (-0.46 to $+0.46$), slightly destabilizing ($+0.46$ to $+0.92$), destabilizing ($+0.92$ to $+1.84$), and highly destabilizing ($>+1.84$) (Studer et al., 2014).

A systematic *in-silico* evaluation of residues at inter-subunit protein-protein interactions (PPI) in the GSC was performed using BUDE Alanine Scan (https://pragmaticproteindesign.bio.

FIGURE 1
Proportion of variant types across these four GSC protein-coding genes.

ed.ac.uk/balas/) (Wood et al., 2020). This tool enables characterization of the thermodynamic stability of PPI hotspot residues using computational alanine scanning. *In-silico* alanine scanning mutagenesis identifies residues which contribute most to the change in the free energy of binding ΔΔG (referred to a "hot spots" (Clackson and Wells, 1995)) defined as ΔΔG = ΔG^{mut}–ΔG^{wt} where ΔG^{wt} refers to the binding free energy upon complexing of the wild-type residue and ΔG^{mut} to that with alanine-mutated proteins. Neutral residues are defined as residues which when mutated fail to meet the ΔΔG ≥2.0 kcal/mol hot spot residue criterion (Morrow and Zhang, 2012).
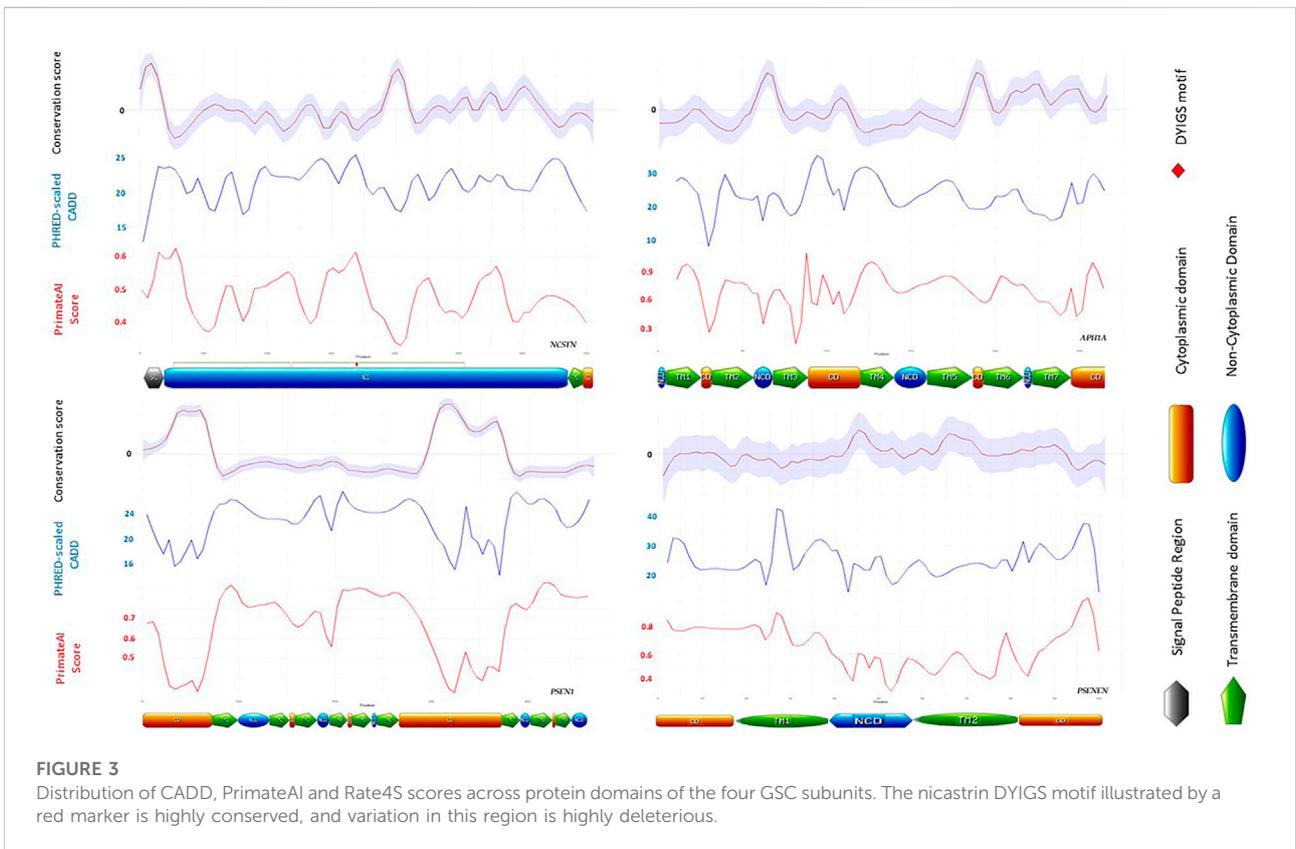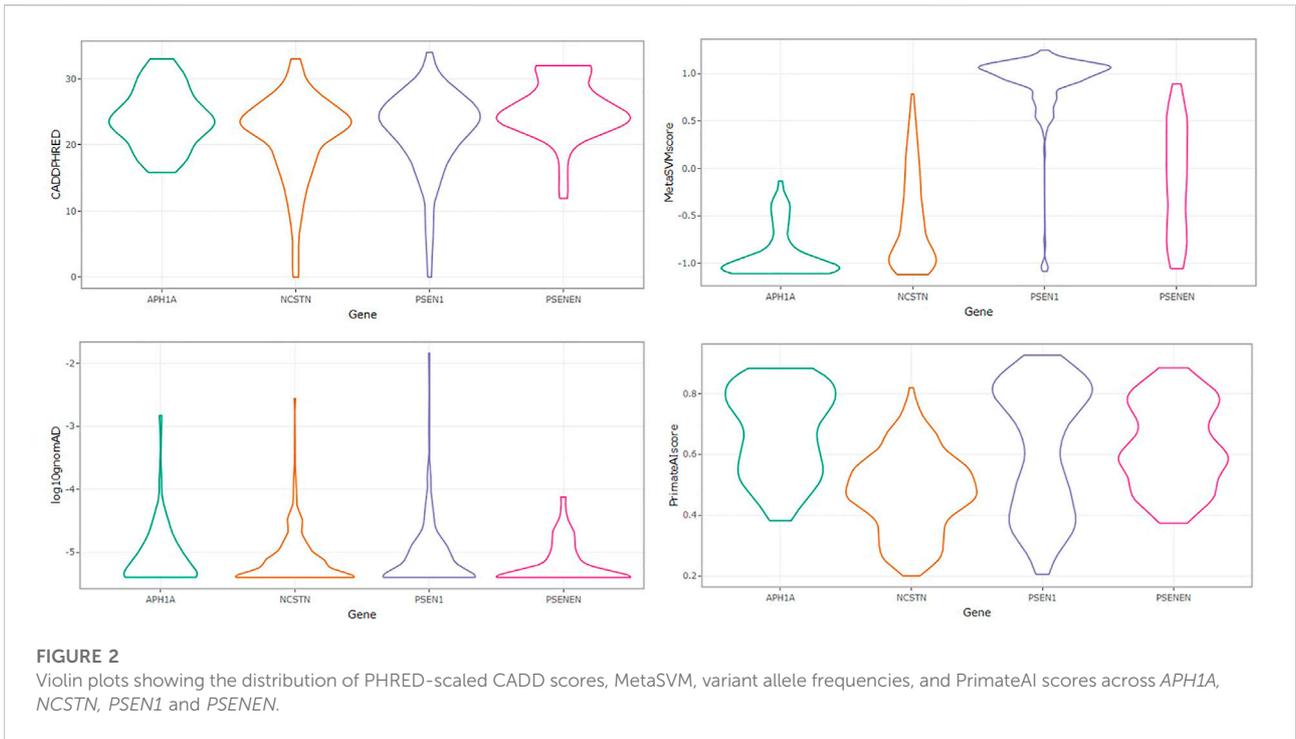
Results

The spectrum of missense variants in ɣ secretase complex protein-coding genes

Using aggregate, multi-ancestry whole exome and whole genome sequencing data from GnomAD, a broad spectrum of variants in *NCSTN*, *PSEN1*, *PSENEN* and *APH1A* were identified. The proportion of variant types across these four GSC protein-coding genes is illustrated in Figure 1. As expected, deleterious variants that disrupt translation through nonsense, frameshift or splice-altering effects are maintained at low frequencies in human populations, whereas missense and synonymous variants are more abundant. A total of 524 missense variants at these four loci were filtered from GnomAD, with most variants ($n = 300$) located in *NCSTN,* the largest of the four genes, followed by 131, 48, 45 missense variants in *PSEN1*, *APH1A* and *PSENEN* respectively. The global observed/expected (o/e) evolutionary constraint score for missense variants in the GSC genes varies, with an o/e ratio of 0.79 (0.72–0.87) for *NCSTN*, 0.62 (0.55–0.71) for *PSEN1*, 0.75 (0.59–0.98) for *PSENEN* and 0.54 (0.46–0.65) for *APH1A*. This indicates that *APH1A* is less

tolerant to missense variation compared to the other GSC loci. In comparison, the four loci have a high probability of being loss-of-function intolerant (pLI) with the loss-of-function o/e upper bound fraction as described by GnomAD being lower than 0.5 (Karczewski et al., 2020).

The distribution of variant allele frequencies across the four loci is shown in Figure 2 with a higher median variant allele frequency in *APH1A* and *NCSTN*. Importantly, the four loci are characterized by abundant rare variants and a relative paucity of missense variants exceeding 0.1% in minor allele frequency. Most of the identified missense variants are singletons carried by only a single individual in GnomAD. This corresponds to the negatively skewed distribution of $\log_{10}$ allele frequency around -5.5 in the four genes. We also compared the distribution of CADD and PrimateAI scores across the four GSC loci. Significantly higher CADD scores occur in *PSENEN*, *PSEN1* and *APH1A* compared to *NCSTN* (Kruskal–Wallis ANOVA, $p < 0.05$). PrimateAI scores were also significantly higher in *PSEN1*, *APH1A* and *PSENEN* relative to *NCSTN* (Kruskal–Wallis ANOVA, $p < 0.05$). The CADD and PrimateAI scores correlate well ($r_s = 0.627$, $p < 0.001$) and both functional prediction algorithms thus illustrate an identical pattern of deleteriousness across the four genes. The distribution of MetaSVM scores was significantly higher in *PSEN1* compared to the other three loci (Kruskal–Wallis ANOVA, $p < 0.05$) as shown in Figure 2.

Given these findings, we set out to map the conservation and pathogenicity scores of missense variants across the primary amino acid sequence of each of the four GSC genes. As expected, a clear interrelatedness of conservation and deleteriousness is demonstrated, with the most conserved residues being also the most deleterious based on Phred-CADD and PrimateAI scores as illustrated in Figure 3. We also evaluated the distribution of CADD, PrimateAI and Rate4S scores across protein domains. A higher distribution of CADD scores was observed in the peptidase domain of nicastrin (Residues 254–468) compared

FIGURE 2
Violin plots showing the distribution of PHRED-scaled CADD scores, MetaSVM, variant allele frequencies, and PrimateAI scores across *APH1A, NCSTN, PSEN1* and *PSENEN*.

FIGURE 3
Distribution of CADD, PrimateAI and Rate4S scores across protein domains of the four GSC subunits. The nicastrin DYIGS motif illustrated by a red marker is highly conserved, and variation in this region is highly deleterious.
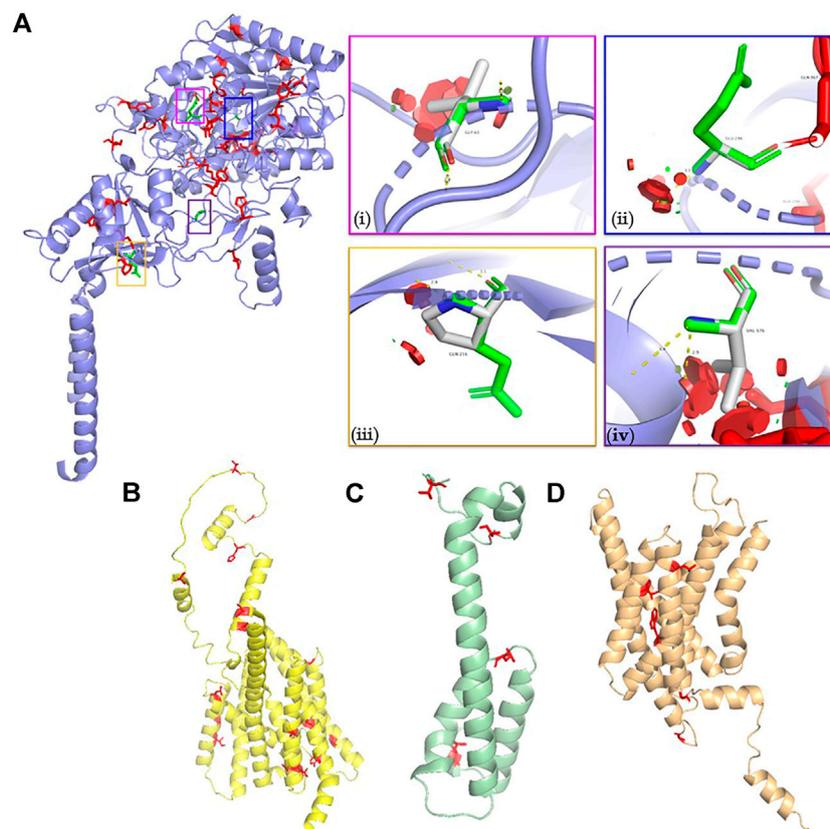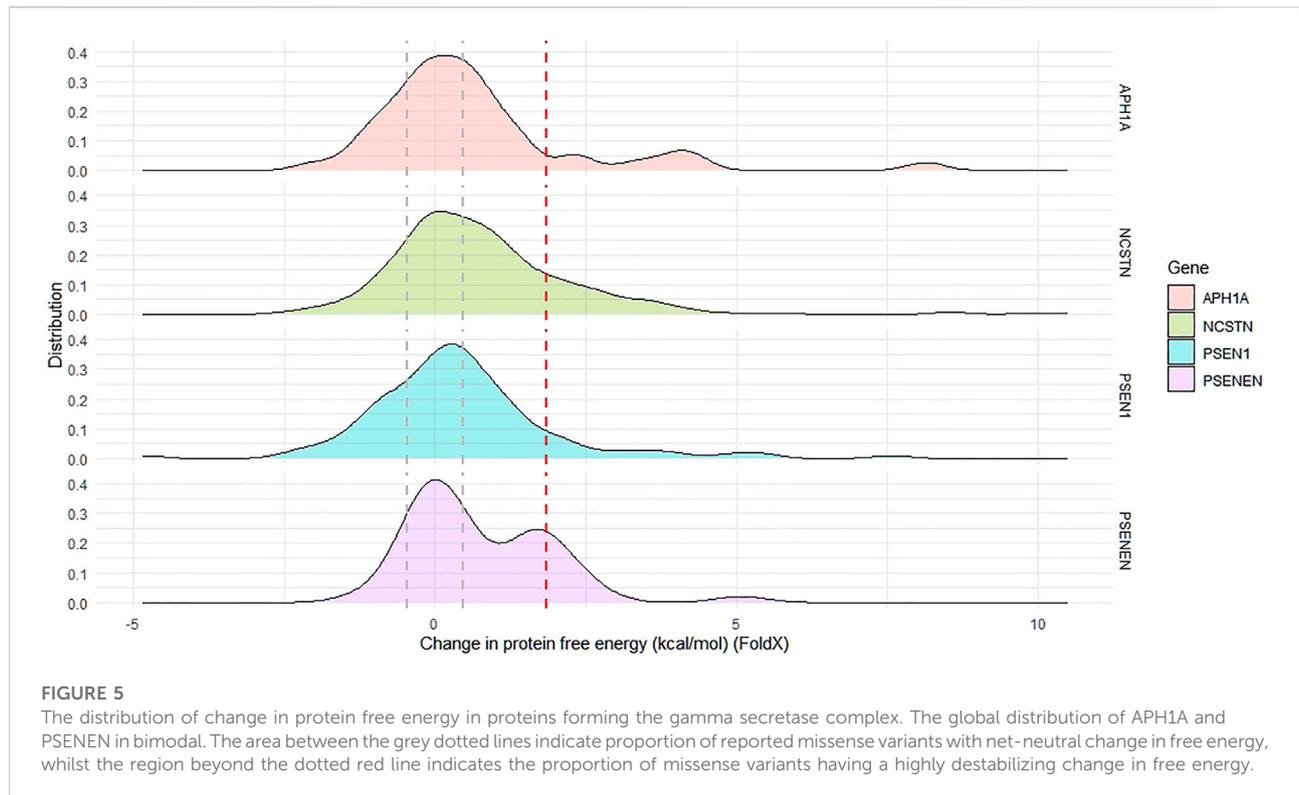
FIGURE 4
Molecular models of **(A)** NCSTN **(B)** PSEN **(C)** PSENEN and **(D)** APH1A proteins. Structurally deleterious residues are highlighted as red ball and stick residues. The structurally damaging *NCSTN* missense variants described in HS are illustrated in vignettes (i) p. Gly61Val (ii) p. Glu296Gly (iii) p. Gln216Pro, and (iv) p. Gly576Val. The p. Gly61Val triggers a disallowed phi/psi angle and replaces a buried glycine residue. The p. Gly296Gly variant also triggers a disallowed phi/psi angle. Finally, the p. Gly296Gly variant replaces a burred charged residue with an uncharged residue. It also leads to expansion of the cavity volume by 139.9 Å3.

to the other regions, although this did not exceed statistical significance thresholds. Notably, a high pathogenicity and conservation score was observed in the DYIGS motif located in a proposed nicastrin substrate-binding pocket (Xie et al., 2014). The four loci are characterized by a high burden of putatively deleterious variation, with most positions having CADD scores >20.

Impact on ɣ secretase complex protein structure

The burden of structurally deleterious missense variants in the GSC was evaluated using the Missense3D server as outlined in the methods. Approximately 13% of all the reported GnomAD coding variants in *NCSTN*, *APH1A* and *PSEN1* were predicted to be structurally deleterious. Conversely, a lower proportion (8%, 4/45) of variants in *PSENEN* were structurally deleterious. GSC variants with a structurally deleterious impact on protein

structure are highlighted red in ball and stick mode in Figure 4. In comparison, four (namely p. Gly61Val (Duchatelet et al., 2020), p. Gln216Pro (Zhang et al., 2013), p. Glu296Gly (Xu et al., 2016) and p. Gly576Val (Duchatelet et al., 2020) [residues in ball and sick mode highlighted in green]) of the ten *NCSTN* missense variants reported in HS are deleterious (40%, $\chi^2 = 6.01$, $p < 0.05$) (Pace et al., 2022). The criteria for deleteriousness is illustrated in Supplementary Table S1. This indicates significant enrichment for damaging missense *NCSTN* substitutions in HS. The normalized conservation score according to ConSurf for the wild type residues at these four positions are −0.885, −0.888, −0.540 and −0.889 respectively, indicating that these variants are slowly evolving, highly conserved sites. Conversely, none of the reported *PSEN* and *PSENEN* missense variants in HS are structurally deleterious. No disease associated *APH1A* missense variant has been reported to date, although a single missense variant in its homolog *APH1B* has been published in association with the HS phenotype (Theut Riis et al., 2020). This *APH1B* p. His170Arg substitution is

FIGURE 5
The distribution of change in protein free energy in proteins forming the gamma secretase complex. The global distribution of APH1A and PSENEN in bimodal. The area between the grey dotted lines indicate proportion of reported missense variants with net-neutral change in free energy, whilst the region beyond the dotted red line indicates the proportion of missense variants having a highly destabilizing change in free energy.

structurally damaging; however, the authors had queried the relevance of this gene in HS, since functional models demonstrate a higher cutaneous expression of *APH1A* rather than *APH1B* (Theut Riis et al., 2020).

The energy stability landscape of γ secretase coding variants

The effect of all GnomAD-reported missense variants in the GSC on protein stability was investigated using FoldX as outlined in the methods. This analysis identifies positions that are mutation intolerant. Across the GSC, 46% of variants are destabilizing according to ΔΔG calculations. The distribution curves and ΔΔG cut-offs for each locus are shown in Figure 5. The global ΔΔG distribution for *APH1A* is also bimodal, with the top destabilizing variants in this gene being p. Ser213Arg, p. Trp174Arg and p. Tyr90Cys. Similarly, the ΔΔG distribution for *PSENEN* is bimodal, with the most destabilizing variants in this gene being p. Leu26Pro and p. Ser94Phe. On the other hand the global distribution of ΔΔG for all *NCSTN* and *PSEN1* variants in GnomAD is unimodal and skewed towards positive values. For the *NCSTN* polymorphisms, the topmost thermodynamically destabilizing variants with extreme ΔΔG values (>6.0 kcal/mol) are p. Val96Phe, p. Leu282Pro, p. Ser286Phe and p. Asp381His.

For *PSEN1*, the topmost thermodynamically destabilizing variants are p. Gly266Val, p. Leu364Pro, p. Ser367Arg and p. Ser313Pro. Across the four GSC loci, a significantly higher ΔΔG value was present in substitutions that had deleterious structural predictions by Missense3D.

In-silico alanine scanning identifies γ secretase hotspot residues at protein-protein interfaces

Protein-protein interactions are dependent on biochemical and/or electrostatic forces and are an essential element of the function of multi subunit complexes. Given their significant role in biological function, disruptions at interface residues can lead to disease. Five of the filtered GSC missense variants which are predicted to be structurally damaging by Missense3D are located at the protein's interface as per data from the RCSB's protein data bank (rcsb.org) (Berman et al., 2000). These include: *PSEN1*:p. Leu219Phe, *APH1A*:p.Gly233Arg and *PSENEN*:p.Leu3Gln, p. Leu26Pro and p. Gly99Asp. *In silico* alanine scanning using BudeAlaScan (Wood et al., 2020) identified hotspots which are presented in Supplementary Table S2. Of the filtered missense variants across the four GSC genes predicted to be structurally deleterious, only the *PSENEN*:p.Leu26Pro variant disrupts a hot spot residue.

On the other hand, of the pathogenic/likely pathogenic variants described in HS patients with multigenerational disease, only the *NCSTN*:p.Val224_Thr227del variant spans a hotspot residue. The deleted residues in this variant are located within an α-helix of NCSTN. Hydrophobic interactions between nicastrin residues Val224 and Ile225 and PSENEN's Pro101 are disrupted by this variant. Therefore, it is possible that the in-frame deletion p. Val224_Thr227 in *NCSTN* influences both the structure and the stability of the *NCSTN* ectodomain and sterically hinders GSC-substrate interactions (Mintoff et al., 2021).

Discussion

In this study we describe an *in-silico* evaluation of coding variation in the GSC, derived from the GnomAD reference population dataset. Several attributes of missense variation in *NCSTN*, *PSEN1*, *PSENEN* and *APH1A* are assessed, including evolutionary conservation and deleteriousness metrics and their relation to protein domains. The impact of amino acid substitutions on protein structure, thermodynamic stability and inter-subunit interaction hotspots is also evaluated. Genes of the GSC have an established role in the pathogenesis of both monogenic Alzheimer disease (primarily *PSEN1*) (Sims et al., 2020) and familial HS (primarily *NCSTN*), although genetic studies of HS are still in their infancy and characterized by relatively small numbers. In addition, variation within the GSC has been associated with complex non-Mendelian traits, ranging from educational attainment to forced expiratory volume in chronic obstructive pulmonary disease to apolipoprotein B levels (Lutz et al., 2015; Lee et al., 2018; Richardson et al., 2020). To the best of our knowledge, this study provides the first comprehensive characterization of GSC variation in a large population genomic dataset, and its interpretation in the setting of known HS pathogenic variation to date.

The analysis demonstrates a high burden of deleterious variation in the four GSC loci as determined by CADD and MetaSVM metrics, and a clear correlation between putative deleteriousness and evolutionary selection pressure along the primary structure of each protein. We also demonstrate significant enrichment for structurally damaging *NCSTN* missense variants in HS relative to GnomAD, with the affected residues being under strong selection pressure. *PSEN* and *PSENEN* missense variants reported in HS phenotypes are not structurally deleterious, suggesting that these can contribute to disease via different mechanisms. Energy calculations reveal that several substitutions are likely to cause a thermodynamically unfavorable effect, and analysis of protein-protein interfaces identifies hotspot residues where substitution is less energetically favorable and likely to cause destabilization of the complex.

The results of this analysis suggest that residues having low Rate4Site scores, high CADD and PrimateAI scores, high ΔΔG energies and structurally deleterious effects based on Missense3D should be prioritized for possible functional analysis to determine their impact on GSC function. While evolutionary conservation metrics provide valuable insight, it is important to consider that their substitution does not always result in disease. It has been postulated that the observed conservation is the result of reduced fitness that the nowadays-benign variants could elicit in affected hosts when competing for survival in the wild (Zaucha et al., 2020). Thus, conservation metrics should not be interpreted in isolation.

Although HS is a heritable trait (Bruinsma et al., 2021), only a minority of patients with apparent familial disease have underlying GSC pathogenic variation. Data from twin registries suggests that the primary mechanism underlying HS involves gene-gene and gene-environment interactions rather than monogenic variation (Straalen et al., 2020; Kjaersgaard Andersen et al., 2021). The extent and mechanisms by which variation in the GSC is implicated in the development and progression of HS is as yet to be elucidated. Indirect evidence in support of GSC involvement in the pathophysiology of HS includes the induction of disease in 75% of patients administered Nirogacestat, a selective pharmacological ɣ-secretase inhibitor (O'Sullivan Coyne et al., 2018). On the other hand, functional studies reveal that variation in GSC subunits alone (specifically *PSENEN*) may not suffice for the development of HS (Zhou et al., 2021).

Whole exome or whole genome sequencing remain the primary methodological approaches of choice for the genetic investigation of patients with suspected familial or syndromic HS (Rehm et al., 2013). Both techniques enable implementation of a biologically-agnostic, hypothesis-free approach. Exome sequencing has been shown to enable diagnoses of 20% of otherwise undiagnosed rare conditions (Gahl et al., 2012). One of the major drawbacks of expansive high-throughput sequencing is the detection of novel or rare missense variations of uncertain significance (VUSs). Such VUS are characterized by the absence of robust evidence to support relevance to a given clinical phenotype yet simultaneously possessing properties reminiscent of disease-causing variation. VUSs are frequent and thus pose a major clinical challenge (Cooper, 2015; Richards et al., 2015; Federici and Soddu, 2020). Rare or novel missense variants often remain relegated to the VUS category thus restricting their interpretation, clinical actionability and dissemination. As a consequence, VUSs place a burden on decision-making and communication. Importantly, population allele frequencies from GnomAD exomes and genomes are routinely employed by variant pathogenicity classification algorithms (Kopanos et al., 2019). Against this background, the present study gains significance as it characterizes GnomAD coding variants which create structural or functional alterations in the GSC.

Phenotypic heterogeneity adds to the complexity of HS, since the conditions presents with extensive variation in affected corporeal sites, underlying risk factors, clinical presentation, disease course and response to therapy. However, this can be considered an avenue for characterizing disease pathophysiology, as it also implies heterogenous pathogenic pathways in HS (Frew et al., 2015). This is reflected in patient endotype clustering. A proposed HS patient cluster having underlying GSC variation, was described as presenting with early-onset disease and a lean constitution, being more likely to have lesions in posterior sites, nodular disease, and suffer from co-morbid pilonidal sinus disease (González-Manso et al., 2021). On the other hand, only weak genotype-phenotype associations have been made. Patients harboring *NCSTN* variants have been described as being likelier to exhibit a phenotype characterized by axillary and mammary involvement, hypertrophic scars, more likely to be females and less likely to be smokers (Axillary-Mammary, LC1 phenotype (Canoui-Poitrine et al., 2013)) (Frew et al., 2019). Interestingly, this patient group is less likely to report a family history of disease. Such association is likely to be confounded by the fact that most HS patients are females having disease involving the axilla and/or breast (Dufour et al., 2014). Appropriate classification has been described as "essential" for providing personalized medicine for HS patients (Ingram and Piguet, 2013). It may also help to reduce the diagnostic delay which invariably leads to a worse prognosis (Kokolakis et al., 2020). Further inferred evidence for phenotypic and pathophysiological heterogeneity may be gained from investigating the genetic underpinnings of HS. The GSC, particularly nicastrin, may contribute more towards inflammation at PSU (Frew and Navrazhina, 2019; Hermasch et al., 2022), possibly through NOTCH pathways,. Other as yet unidentified genes may contribute more toward hyperkeratinization of the PSU's infundibular segment.

Genomic studies in HS and the use of GnomAD in variant classification are characterized by several caveats that we will outline next. A comprehensive understanding of the impact of GSC missense variants on protein function is lacking. Attributing pathogenicity is difficult in the absence of *in-vitro* or *in-vivo* functional evidence. Few studies investigate changes in substrate binding affinity, NOTCH reporter assays or catalytic activity of the GSC, and thus follow-up of rare variants identified from sequencing studies is required. The spectrum of locus heterogeneity driving HS is unknown. Several independent reports have implicated *NCSTN* variation in lean subjects with atypical or familial HS. Yet, it is important to consider that selecting for such probands or kindreds can lead to an ascertainment bias that overestimates the contribution of the GSC in HS. It is thus reasonable to postulate that variants having reduced penetrance which predispose to milder or late-onset

disease and that do not drive familial segregation are not detected in cohort studies. Such hypomorphic or low-risk variants necessitate large discovery and replication cohorts. Ranola *et al.* show that comparing clinical cases to population controls derived from public databases ascertained using different criteria, can introduce false positives in genetic testing (Ranola et al., 2019). Important lessons can be derived from studies on transethnic comparison of polygenic risk scores. These highlight the need for an understanding of population-specific genetic risk and reinforce the rationale for large-scale genome-wide association studies in diverse populations (Duncan et al., 2019; Matsunaga et al., 2020).

In comparison to other monogenic diseases, the use of GnomAD-derived allele frequencies in HS carries some distinctions. Deriving allele frequencies from aggregate datasets unselected for disease, such as GnomAD, is a powerful approach that is core in ACMG-AMP classification criteria. However, rare variants causing late-onset disease are present in GnomAD. These can confound variant classification, and the causality of previously identified disease-associated variants has been disputed (Manrai et al., 2016; Kosmicki et al., 2017). For low-frequency missense variants that are not structurally deleterious and present in GnomAD, attributing causality requires functional evaluation. Conversely, it is impossible to ascertain whether diseases such as HS are represented in GnomAD, since phenotype and other individual-level data is not included. GnomAD excluded participants with severe pediatric diseases that impact on reproductive fitness. As HS is characterized by a significant diagnostic delay and extensive clinical heterogeneity in adult patients, it is likely that GnomAD participants bear HS-causal GSC variants. This reinforces the need for large biobank-driven prospective genomics HS research endeavors with carefully phenotyped cohorts.

In addition to the restrictions imposed by the absence of individual-level phenotype data in GnomAD, additional limitations must be acknowledged. GnomAD contains an overrepresentation of European participants, with poor representation from Asian, Oceanic and Middle Eastern populations. No knowledge on the penetrance of GSC alleles in HS exists. We focused on missense variants only. However, variants in introns, promoter regions or synonymous substitutions can impact on transcriptional efficiency or alter splicing. We did not stratify the analysis based on population substructure or gender data within GnomAD. Additionally, the alternate effect of substitutions on splicing isoforms that might offset the deleterious effects of variants was not considered (Kukurba et al., 2014; Miosge et al., 2015). The use of protein 3D structural data to computationally model the effects of missense variants can provide additional evidence for deleteriousness. Methods that assess the thermodynamic properties related to mutations, such as changes in protein stability consider only one possible mechanism that may affect

the phenotype (Glusman et al., 2017). It is likely that missense variants produce functional changes through alternate molecular mechanisms. *In-silico* pathogenicity classifiers were based on output from MetaSVM and CADD. Recent research suggests that gene-specific thresholds provide better performance for missense variants in a subset of genes (Sallah et al., 2022). MetaSVM has been shown to yield high performance in benchmark studies (Anderson and Lassmann, 2018) although, some investigations showed that it can exaggerate pathogenicity of variants in specific proteins (Zaucha et al., 2020). The overreliance on *in-silico* predictors in the absence of their validation in the setting of HS-causal GSC variation is another limitation. This study only made use of the GnomAD dataset, and did not consider additional sources of reference genomic data such as the United Kingdom Biobank and TopMed repositories (Sudlow et al., 2015; Taliun et al., 2021).

Conclusion

This study highlights the need for a greater effort to be made in the understanding of the sequence-function relationship of the GSC in HS. Importantly, a combined clinical and molecular characterization approach is required to capture subtle pathophysiological changes along the disease continuum involving both hyperkeratinization and inflammation. How alleles in GSC act to modify susceptibility in the context of a multifactorial trait, including interaction with complex polygenic risk phenotypes such as obesity or the metabolic syndrome, is also unknown. We envisage that the future of identifying and validating novel disease-gene candidates, may occur through interdisciplinary collaboration of clinical and research facilities (Rehm et al., 2013). Various hurdles need to be overcome for HS genomics research to achieve clinical significance, including deep phenotyping, screening of at-risk family members to delineate segregation, and detailed variant curation (Jabbour et al., 2021). Rare-variant burden enrichment analysis provides additional scope for HS and has recently been applied to GSC-associated syndromic forms of HS (Brandão et al., 2022). This study also highlights the necessity to harness insights from structural biology into clinical variant interpretation algorithms.

References

Al-Ali, F. M. S., Ratnamala, U., Mehta, T. Y., Naveed, M., Al-Ali, M. T., Al-Khaja, N., et al. (2010). Hidradenitis suppurativa (or Acne inversa) with autosomal dominant inheritance is not linked to chromosome 1p21.1-1q25.3 region. *Exp. Dermatol* 19, 851–853. doi:10.1111/j.1600-0625.2010.01088.x

Anderson, D., and Lassmann, T. (2018). A phenotype centric benchmark of variant prioritisation tools. *npj Genom. Med.* 3, 1–9. doi:10.1038/s41525-018-0044-9

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.962449/full#supplementary-material

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., et al. (2016). ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344–W350. doi:10.1093/nar/gkw408

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235

Brandão, L. A. C., Moura, R. R., Marzano, C., Moltrasio, P. M., Tricarico, S., and Crovella, S. (2022). Variant enrichment analysis to explore pathways functionality in complex autoinflammatory skin disorders through whole exome sequencing analysis. *Ijms* 23, 2278. doi:10.3390/ijms23042278

Bruinsma, R. L., Fajgenbaum, K., Yang, Y., Mohlke, K. L., Li, Y., and Sayed, C. (2021). Assessment of familial risk in patients with hidradenitis suppurativa. *Br. J. Dermatol* 184, 753–754. doi:10.1111/bjd.19664

Canoui-Poitrine, F., Le Thuaut, A., Revuz, J. E., Viallette, C., Gabison, G., Poli, F., et al. (2013). Identification of three hidradenitis suppurativa phenotypes: Latent class analysis of a cross-sectional study. *J. Investigative Dermatology* 133, 1506–1511. doi:10.1038/jid.2012.472

Clackson, T., and Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science* 267, 383–386. doi:10.1126/science.7529940

Cooper, G. M. (2015). Parlez-vous VUS? *Genome Res.* 25, 1423–1426. doi:10.1101/gr.190116.115

Duchatelet, S., Miskinyte, S., Delage, M., Ungeheuer, M.-N., Lam, T., Benhadou, F., et al. (2020). Low prevalence of GSC gene mutations in a large cohort of predominantly caucasian patients with hidradenitis suppurativa. *J. Investigative Dermatology* 140, 2085–2088. e14. doi:10.1016/j.jid.2019.10.025

Dufour, D. N., Emtestam, L., and Jemec, G. B. (2014). Hidradenitis suppurativa: A common and burdensome, yet under-recognised, inflammatory skin disease. *Postgrad. Med. J.* 90, 216–221. doi:10.1136/postgradmedj-2013-131994

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., et al. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* 10, 3328. doi:10.1038/s41467-019-11112-0

Esme, P., Esme, M., and Caliskan, E. (2020). Increased prevalence of family history of Alzheimer's disease in hidradenitis suppurativa: Cross-sectional analysis of 192 HS patients. *Dermatol. Ther.* 33. doi:10.1111/dth.14219

Federici, G., and Soddu, S. (2020). Variants of uncertain significance in the era of high-throughput genome sequencing: A lesson from breast and ovary cancers. *J. Exp. Clin. Cancer Res.* 39, 46. doi:10.1186/s13046-020-01554-6

Frew, J. W., Hawkes, J. E., Sullivan-Whalen, M., Gilleaudeau, P., and Krueger, J. G. (2019). Inter-rater reliability of phenotypes and exploratory genotype-phenotype analysis in inherited hidradenitis suppurativa. *Br. J. Dermatol* 181, 566–571. doi:10.1111/bjd.17695

Frew, J. W., and Navrazhina, K. (2019). *In silico* analysis of gamma-secretase-complex mutations in hidradenitis suppurativa demonstrates disease-specific substrate recognition and cleavage alterations. *Front. Med.* 6, 206. doi:10.3389/fmed.2019.00206

Frew, J. W., and Navrazhina, K. (2020). No evidence that impaired Notch signalling differentiates hidradenitis suppurativa from other inflammatory skin diseases. *Br. J. Dermatol* 182, 1042–1043. doi:10.1111/bjd.18593

Frew, J. W., Vekic, D. A., Woods, J., and Cains, G. D. (2015). Phenotypic heterogeneity implies heterogeneous pathogenic pathways in hidradenitis suppurativa. *Exp. Dermatol* 24, 338–339. doi:10.1111/exd.12648

Frew, J. W. (2019). We need to talk about notch: Notch dysregulation as an epiphenomenon in inflammatory skin disease. *Br. J. Dermatol* 180, 431–432. doi:10.1111/bjd.17414

Gahl, W. A., Markello, T. C., Toro, C., Fajardo, K. F., Sincan, M., Gill, F., et al. (2012). The national institutes of health undiagnosed diseases program: Insights into rare diseases. *Genet. Med.* 14, 51–59. doi:10.1038/gim.0b013e318232a005

Gao, M., Wang, P.-G., Cui, Y., Yang, S., Zhang, Y.-H., Lin, D., et al. (2006). Inversa acne (hidradenitis suppurativa): A case report and identification of the locus at chromosome 1p21.1-1q25.3. *J. Investigative Dermatology* 126, 1302–1306. doi:10.1038/sj.jid.5700272

Glusman, G., Rose, P. W., Prlić, A., Dougherty, J., Duarte, J. M., Hoffman, A. S., et al. (2017). Mapping genetic variations to three-dimensional protein structures to enhance variant interpretation: A proposed framework. *Genome Med.* 9, 113. doi:10.1186/s13073-017-0509-y

González-Manso, A., Agut-Busquet, E., Romaní, J., Vilarrasa, E., Bittencourt, F., Mensa, A., et al. (2021). Hidradenitis suppurativa: Proposal of classification in two endotypes with two-step cluster Analysis. *Dermatology* 237, 365–371. doi:10.1159/000511045

Hermasch, M. A., Janning, H., Perera, R. P., Schnabel, V., Rostam, N., Ramos-Gomes, F., et al. (2022). Evolutionary distinct roles of γ-secretase subunit nicastrin in zebrafish and humans. *J. Dermatological Sci.* 105, 80–87. doi:10.1016/j.jdermsci.2022.01.001

Hur, J.-Y. (2022). γ-Secretase in Alzheimer's disease. *Exp. Mol. Med.* 54, 433–446. doi:10.1038/s12276-022-00754-8

Ingram, J. R., and Piguet, V. (2013). Phenotypic heterogeneity in hidradenitis suppurativa (acne inversa): Classification is an essential step toward personalized therapy. *J. Investigative Dermatology* 133, 1453–1456. doi:10.1038/jid.2012.476

Jabbour, A. J., Straalen, K. R., Colvin, A., Prens, E. P., and Petukhova, L. (2021). Clinical translation of hidradenitis suppurativa genetic studies requires global collaboration. *Br. J. Dermatol* 186, 183–185. doi:10.1111/bjd.20749

Jiao, T., Dong, H., Jin, L., Wang, S., and Wang, J. (2013). A novel nicastrin mutation in a large Chinese family with hidradenitis suppurativa. *Br. J. Dermatol* 168, 1141–1143. doi:10.1111/bjd.12135

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443. doi:10.1038/s41586-020-2308-7

Khanna, T., Hanna, G., Sternberg, M. J. E., and David, A. (2021). Missense3D-DB web catalogue: An atom-based analysis and repository of 4M human protein-coding genetic variants. *Hum. Genet.* 140, 805–812. doi:10.1007/s00439-020-02246-z

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. doi:10.1038/ng.2892

Kjærsgaard Andersen, R., Clemmensen, S. B., Larsen, L. A., Hjelmborg, J. V. B., Ødum, N., Jemec, G. B. E., et al. (2021). Evidence of gene-gene interaction in hidradenitis suppurativa: A nationwide registry study of Danish twins. *Br. J. Dermatol* 186, 78–85. doi:10.1111/bjd.20654

Kokolakis, G., Wolk, K., Schneider-Burrus, S., Kalus, S., Barbus, S., Gomis-Kleindienst, S., et al. (2020). Delayed diagnosis of hidradenitis suppurativa and its effect on patients and healthcare system. *Dermatology* 236, 421–430. doi:10.1159/000508787

König, A., Lehmann, C., Rompel, R., and Happle, R. (1999). Cigarette smoking as a triggering factor of hidradenitis suppurativa. *Dermatology* 198, 261–264. doi:10.1159/000018126

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., et al. (2019). VarSome: The human genomic variant search engine. *Bioinformatics* 35, 1978–1980. doi:10.1093/bioinformatics/bty897

Kosmicki, J. A., Samocha, K. E., Howrigan, D. P., Sanders, S. J., Slowikowski, K., Lek, M., et al. (2017). Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet.* 49, 504–510. doi:10.1038/ng.3789

Kromann, C. B., Ibler, K. S., Kristiansen, V. B., and Jemec, G. B. E. (2014). The influence of body weight on the prevalence and severity of hidradenitis suppurativa. *Acta Derm. Venerol.* 94, 553–557. doi:10.2340/00015555-1800

Kukurba, K. R., Zhang, R., Li, X., Smith, K. S., Knowles, D. A., How Tan, M. H., et al. (2014). Allelic expression of deleterious protein-coding variants across human tissues. *PLoS Genet.* 10, e1004304. doi:10.1371/journal.pgen.1004304

Lee, J. J., Wedow, R., Wedow, A., Okbay, E., Kong, O., Maghzian, M., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* 50, 1112–1121. doi:10.1038/s41588-018-0147-3

Liu, X., Wu, C., Li, C., and Boerwinkle, E. (2016). dbNSFP v3.0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum. Mutat.* 37, 235–241. doi:10.1002/humu.22932

Lutz, S. M., Cho, M. H., Cho, K., Young, C. P., Hersh, P. J., Castaldi, M.-L., et al. (2015). A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet.* 16, 138. doi:10.1186/s12863-015-0299-4

Manrai, A. K., Funke, B. H., Rehm, H. L., Olesen, M. S., Maron, B. A., Szolovits, P., et al. (2016). Genetic misdiagnoses and the potential for health disparities. *N. Engl. J. Med.* 375, 655–665. doi:10.1056/NEJMsa1507092

Matsunaga, H., Ito, K., Akiyama, M., Takahashi, A., Koyama, S., Nomura, S., et al. (2020). Transethnic meta-analysis of genome-wide association studies identifies three New loci and characterizes population-specific differences for coronary artery disease. *Circ Genomic Precis. Med.* 13, e002670. doi:10.1161/CIRCGEN.119.002670

Melnik, B. c., and Plewig, G. (2013). Impaired notch signalling: The unifying mechanism explaining the pathogenesis of hidradenitis suppurativa (acne inversa). *Br. J. Dermatology* 168, 876–878. doi:10.1111/bjd.12068

Merilahti, J. A. M., Ojala, V. K., Knittle, A. M., Pulliainen, A. T., and Elenius, K. (2017). Genome-wide screen of gamma-secretase-mediated intramembrane cleavage of receptor tyrosine kinases. *MBoC* 28, 3123–3131. doi:10.1091/mbc.E17-04-0261

Mintoff, D., Pace, N. p., Bauer, P., and Borg, I. (2021). A novel c.671_682del NCSTN variant in a family with hidradenitis suppurativa: A pilot study. *Clin. Exp. Dermatol* 46, 1306–1308. doi:10.1111/ced.14677

Miosge, L. A., Field, M. A., Sontani, Y., Cho, V., Johnson, S., Palkova, A., et al. (2015). Comparison of predicted and actual consequences of missense mutations. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5189–E5198. doi:10.1073/pnas.1511585112

Miskinyte, S., Nassif, A., Merabtene, F., Ungeheuer, M.-N., Join-Lambert, O., Jais, J.-P., et al. (2012). Nicastrin mutations in French families with hidradenitis suppurativa. *J. Investigative Dermatology* 132, 1728–1730. doi:10.1038/jid.2012.23

Morrow, J. K., and Zhang, S. (2012). Computational prediction of protein hot spot residues. *Curr. Pharm. Des.* 18, 1255–1265. doi:10.2174/138161212799436412

Nomura, Y., Nomura, T., Suzuki, S., Takeda, M., Mizuno, O., Ohguchi, Y., et al. (2014). A novel NCSTN mutation alone may be insufficient for the development of familial hidradenitis suppurativa. *J. Dermatological Sci.* 74, 180–182. doi:10.1016/j.jdermsci.2014.01.013

O'Sullivan Coyne, G., Woodring, T. S., Lee, C.-C. R., Chen, A. P., and Kong, H. H. (2018). Hidradenitis suppurativa-like lesions associated with pharmacologic inhibition of gamma-secretase. *J. Investigative Dermatology* 138, 979–981. doi:10.1016/j.jid.2017.09.051

Pace, N. P., Mintoff, D., and Borg, I. (2022). The genomic architecture of hidradenitis suppurativa-A systematic review. *Front. Genet.* 13, 861241. doi:10.3389/fgene.2022.861241

Pink, A. E., Dafou, D., Desai, N., Holmes, O., Hobbs, C., Smith, C. H., et al. (2016). Hidradenitis suppurativa: Haploinsufficiency of gamma-secretase components does not affect gamma-secretase enzyme activity*in vitro*. *Br. J. Dermatol* 175, 632–635. doi:10.1111/bjd.14621

PyMOL (2020). pymol.org. AvaliableAt: https://pymol.org/2/(Accessed 24, 11, 2020).

Ranola, J. M. O., Tsai, G. J., and Shirts, B. H. (2019). Exploring the effect of ascertainment bias on genetic studies that use clinical pedigrees. *Eur. J. Hum. Genet.* 27, 1800–1807. doi:10.1038/s41431-019-0467-5

Ratnamala, U., Jhala, D., Jain, N. K., Saiyed, N. M., Raveendrababu, M., Rao, M. V., et al. (2016). Expanding the spectrum ofγ-secretase gene mutation-associated phenotypes: Two novel mutations segregating with familial hidradenitis suppurativa (acne inversa) and acne conglobata. *Exp. Dermatol* 25, 314–316. doi:10.1111/exd.12911

Rehm, H. L., Bale, S. J., Bayrak-Toydemir, P., Berg, J. S., Brown, K. K., Deignan, J. L., et al. (2013). ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med.* 15, 733–747. doi:10.1038/gim.2013.92

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., and Kircher, M. (2019). Cadd: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 47, D886–D894. doi:10.1093/nar/gky1016

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). ACMG laboratory quality assurance CommitteeStandards 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.* 17, 405–424. doi:10.1038/gim.2015.30

Richardson, T. G., Sanderson, E., Palmer, T. M., Ala-Korpela, M., Ference, B. A., Davey Smith, G., et al. (2020). Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable mendelian randomisation analysis. *PLoS Med.* 17, e1003062. doi:10.1371/journal.pmed.1003062

Sallah, S. R., Ellingford, J. M., Sergouniotis, P. I., Ramsden, S. C., Lench, N., Lovell, S. C., et al. (2022). Improving the clinical interpretation of missense variants in X linked genes using structural analysis. *J. Med. Genet.* 59, 385–392. doi:10.1136/jmedgenet-2020-107404

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: An online force field. *Nucleic Acids Res.* 33, W382–W388. doi:10.1093/nar/gki387

Sellheyer, K., and Krahl, D. (2008). What causes acne inversa (or hidradenitis suppurativa)? the debate continues. *J. Cutan. Pathol.* 35, 701–703. doi:10.1111/j.1600-0560.2008.01073.x

Sims, R., Hill, M., and Williams, J. (2020). The multiplex model of the genetics of Alzheimer's disease. *Nat. Neurosci.* 23, 311–322. doi:10.1038/s41593-020-0599-5

Studer, R. A., Christin, P.-A., Williams, M. A., and Orengo, C. A. (2014). Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2223–2228. doi:10.1073/pnas.1310811111

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. (2015). UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12, e1001779. doi:10.1371/journal.pmed.1001779

Sundaram, L., Gao, H., Padigepati, S. R., McRae, J. F., Li, Y., Kosmicki, J. A., et al. (2018). Predicting the clinical impact of human mutation with deep neural networks. *Nat. Genet.* 50, 1161–1170. doi:10.1038/s41588-018-0167-z

Taliun, D., Harris, D. N., Harris, M. D., Kessler, J., Carlson, Z. A., Szpiech, R., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299. doi:10.1038/s41586-021-03205-y

Theut Riis, P., Loft, I. C., Yazdanyar, S., Kjærsgaard Andersen, R., Pedersen, O. B., Ring, H. C., et al. (2021). Full exome sequencing of 11 families with Hidradenitis suppurativa. *J. Eur. Acad. Dermatol Venereol.* 35, 1203–1211. doi:10.1111/jdv.17095

van Straalen, K. R., Prens, E. P., Willemsen, G., Boomsma, D. I., and van der Zee, H. H. (2020). Contribution of genetics to the susceptibility to hidradenitis suppurativa in a large, cross-sectional Dutch twin cohort. *JAMA Dermatol* 156, 1359. doi:10.1001/jamadermatol.2020.3630

Venselaar, H., te Beek, T. A., Kuipers, R. K., Hekkelman, M. L., and Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinforma.* 11, 548. doi:10.1186/1471-2105-11-548

Wang, B., Yang, W., Wen, W., Sun, J., Su, B., Liu, B., et al. (2010). γ-Secretase gene mutations in familial acne inversa. *Science* 330, 1065. doi:10.1126/science.1196284

Wolfe, M. S., De Los Angeles, J., Miller, D. D., Xia, W., and Selkoe, D. J. (1999). Are presenilins intramembrane-cleaving proteases? Implications for the molecular mechanism of alzheimer's disease. *Biochemistry* 38, 11223–11230. doi:10.1021/bi991080q

Wood, C. W., Ibarra, A. A., Bartlett, G. J., Wilson, A. J., Woolfson, D. N., and Sessions, R. B. (2020). BAlaS: Fast, interactive and accessible computational alanine-scanning using BudeAlaScan. *Bioinformatics* 36, 2917–2919. doi:10.1093/bioinformatics/btaa026

Xiao, X., Liu, H., Liu, X., Zhang, W., Zhang, S., and Jiao, B. (2021). APP, PSEN1, and PSEN2 variants in Alzheimer's disease: Systematic Re-evaluation according to ACMG guidelines. *Front. Aging Neurosci.* 13, 695808. doi:10.3389/fnagi.2021.695808

Xie, T., Yan, C., Zhou, R., Zhao, Y., Sun, L., Yang, G., et al. (2014). Crystal structure of the γ-secretase component nicastrin. *Proc. Natl. Acad. Sci. U.S.A.* 111, 13349–13354. doi:10.1073/pnas.1414837111

Xu, H., Xiao, X., Hui, Y., Zhang, X., He, Y., Li, C., et al. (2016). Phenotype of 53 Chinese individuals with nicastrin gene mutations in association with familial hidradenitis suppurativa (acne inversa). *Br. J. Dermatol* 174, 927–929. doi:10.1111/bjd.14268

Zaucha, J., Heinzinger, M., Tarnovskaya, S., Rost, B., and Frishman, D. (2020). Family-specific analysis of variant pathogenicity prediction tools. *NAR Genomics Bioinforma.* 2, lqaa014. doi:10.1093/nargab/lqaa014

Zhang, C., Wang, L., Chen, L., Ren, W., Mei, A., Chen, X., et al. (2013). Two novel mutations of the NCSTN gene in Chinese familial acne inverse. *J. Eur. Acad. Dermatol Venereol.* 27, 1571–1574. doi:10.1111/j.1468-3083.2012.04627.x

Zhang, X., and Sisodia, S. S. (2015). Acne inversa caused by missense mutations in NCSTN is not fully compatible with impairments in Notch signaling. *J. Investigative Dermatology* 135, 618–620. doi:10.1038/jid.2014.399

Zhou, P., Liu, J., Xu, T., Guo, Y., Han, Y., He, Y., et al. (2021). Mutations in γ-secretase subunit-encoding PSENEN gene alone may not be sufficient for the development of acne inversa. *J. Dermatological Sci.* 103 (21), 73–81. doi:10.1016/j.jdermsci.2021.06.007

Zouboulis, C. C., Nogueira da Costa, A., Makrantonaki, E., Hou, X. X., Almansouri, D., Dudley, J. T., et al. (2020). Alterations in innate immunity and epithelial cell differentiation are the molecular pillars of hidradenitis suppurativa. *J. Eur. Acad. Dermatol Venereol.* 34, 846–861. doi:10.1111/jdv.16147

Zouboulis, V. A., Zouboulis, K. C., and Zouboulis, C. C. (2021). Hidradenitis suppurativa and comorbid disorder biomarkers, druggable genes, New drugs and drug repurposing-A molecular meta-analysis. *Pharmaceutics* 14, 44. doi:10.3390/pharmaceutics14010044