WP.

SOPHiA DDM™ Community variant classification connects experts and drives collective intelligence

Accurate and consistent interpretation of genetic variants is a cornerstone of clinical genomics, yet it remains one of the most complex and time-consuming aspects of genetic testing. While public databases and standardized guidelines such as those from the American College of Medical Genetics and Genomics (ACMG) have brought greater structure to variant classification, many variants - particularly variants of uncertain significance (VUS) - still lack sufficient evidence for confident interpretation. SOPHiA GENETICS addresses this challenge through the SOPHiA DDM™ Platform, which not only automates variant detection, annotation, and classification, but also enables users to contribute to and benefit from one of the largest expert networks in the world. This whitepaper explores how the SOPHiA DDM™ variant flagging feature fosters community-driven classification, generating a collective intelligence that enhances interpretation accuracy, supports clinical decision-making, and strengthens the predictive performance of the proprietary AROMA AI agent.

Table of Contents

Addressing variant classification challenges

The accurate and timely interpretation of genomic variants identified during genetic testing is crucial for delivering the highest standards of patient care. Sequence variant interpretation requires integration of diverse sources of evidence, from population data to functional studies, to support the link between a specific variant and a disease. However, synthesizing multiple complex data sources is both challenging and time-consuming, making variant interpretation a persistent hurdle in genetic testing.

To standardize germline variant interpretation across laboratories and increase classification accuracy, the ACMG partnered with the Association for Molecular Pathology (AMP) to provide a systematic framework for collecting and assessing evidence such as population data, computational data, functional data, and segregation data to support the determination of variant classification as “benign”, “likely benign”, “uncertain significance”, “likely pathogenic”, or “pathogenic”.[1]

Multiple internationally recognized databases enable the global community of clinical laboratories to share their variant classifications amongst each other, with the aim of creating a collective intelligence. ClinVar,[2] ClinGen,[3] COSMIC,[4] DECIPHER,[5] and dbSNP[6] are some of the most well-known of these databases, with ClinVar remaining the only database where submission of variant pathogenicity is open to all, and not just to specific pre-curated expert groups. Due to this open access, ClinVar contains a staggering 3+ million unique variants.[2]

Despite the implementation of these global variant interpretation standards and access to global databases, 36% of ClinVar variants remain classified as VUS.[7] Indeed, the total number of VUS in ClinVar increased 5-fold between 2020 and 2023, partly as a result of the trend to use less targeted, larger gene panels such as exomes and genomes. VUS can lead to complexity in reporting by leading to inconclusive genetic testing reports. As a result, there is a pressing need to identify and apply innovative solutions that accelerate and facilitate the variant interpretation process, particularly in relation to VUS.

This whitepaper outlines how SOPHiA DDM™ users contribute to the SOPHiA GENETICS Community by sharing their pathogenicity assessment of identified variants. This variant flagging amongst the SOPHiA GENETICS Community of experts creates collective insights to increase confidence when interpreting variants, particularly VUS. We report details of the nearly one million variant flags in SOPHiA DDM™, their concordance with other databases, how they contribute to the advanced performance of the proprietary AROMA predictive algorithm, and the novel insights gained from this community - one of the largest networks of healthcare institutions in the world.

SOPHiA DDM™ for variant detection, annotation, and classification

SOPHiA DDM™ is a cloud-native AI SaaS platform for the analysis, standardization, and interpretation of multiomic healthcare data. In Genomics, it utilizes patented state-of-the-art machine learning algorithms to efficiently call, annotate, and classify variants from raw next-generation sequencing (NGS) data. Intuitive features streamline variant prioritization, simplify interpretation, and expedite reporting through a user-friendly, web-based interface (*Fig. 1*).

Robust data security is guaranteed, with full compliance to IVDR, HIPAA, GDPR, and ISO/IEC 27001, 27017 & 27018, to ensure the highest standards of data privacy and cloud security.

SOPHiA DDM™ Community variant classification

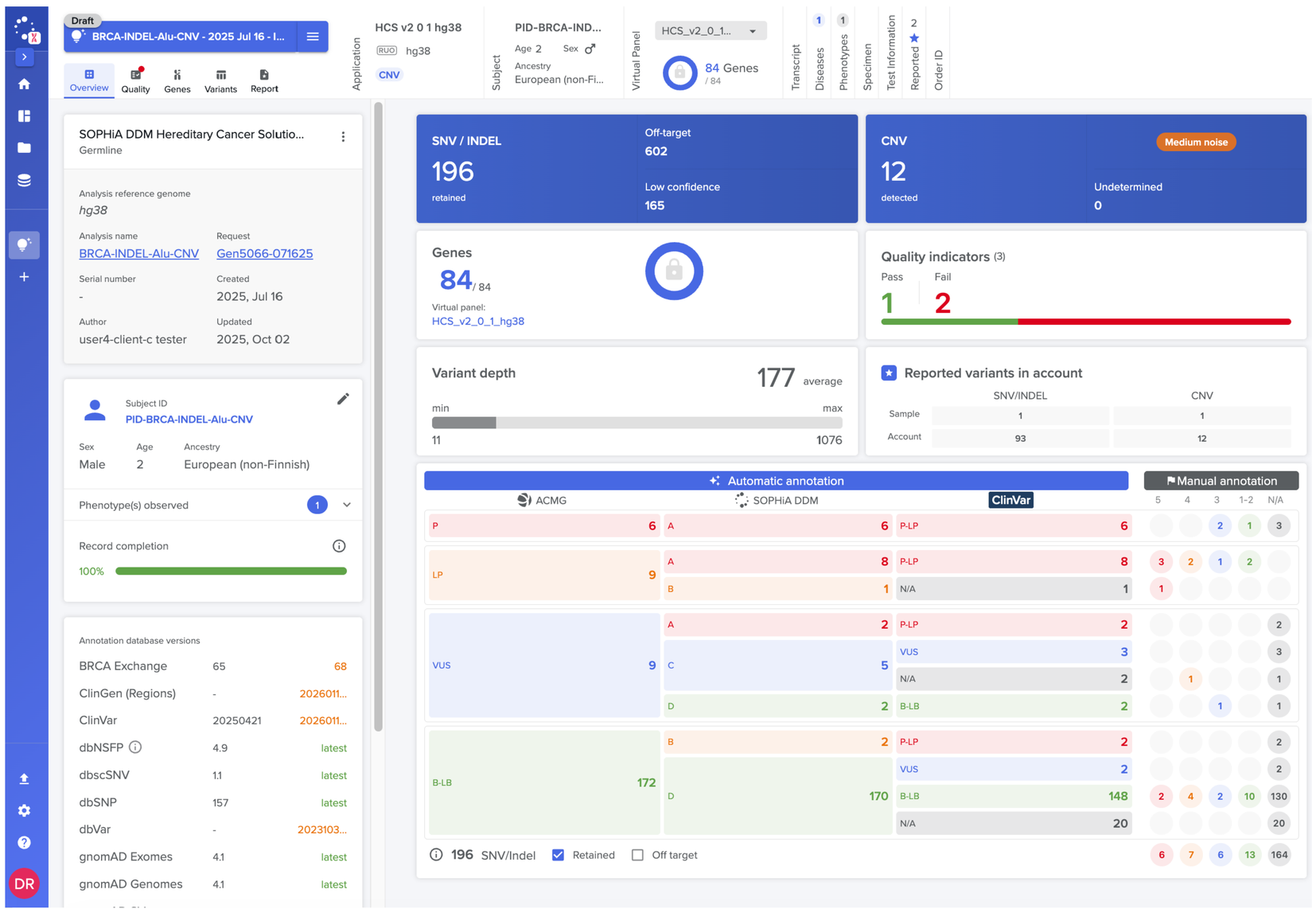

Figure 1 | The SOPHiA DDM™ variant interpretation view provides intuitive access to key interpretive information. Automatic annotation classifies variants according to ACMG and ClinVar criteria, as well as using the SOPHiA DDM™ proprietary AROMA variant classification algorithm (centre). Variants classified by the SOPHiA DDM™ community of experts are displayed in the user annotation section (right).

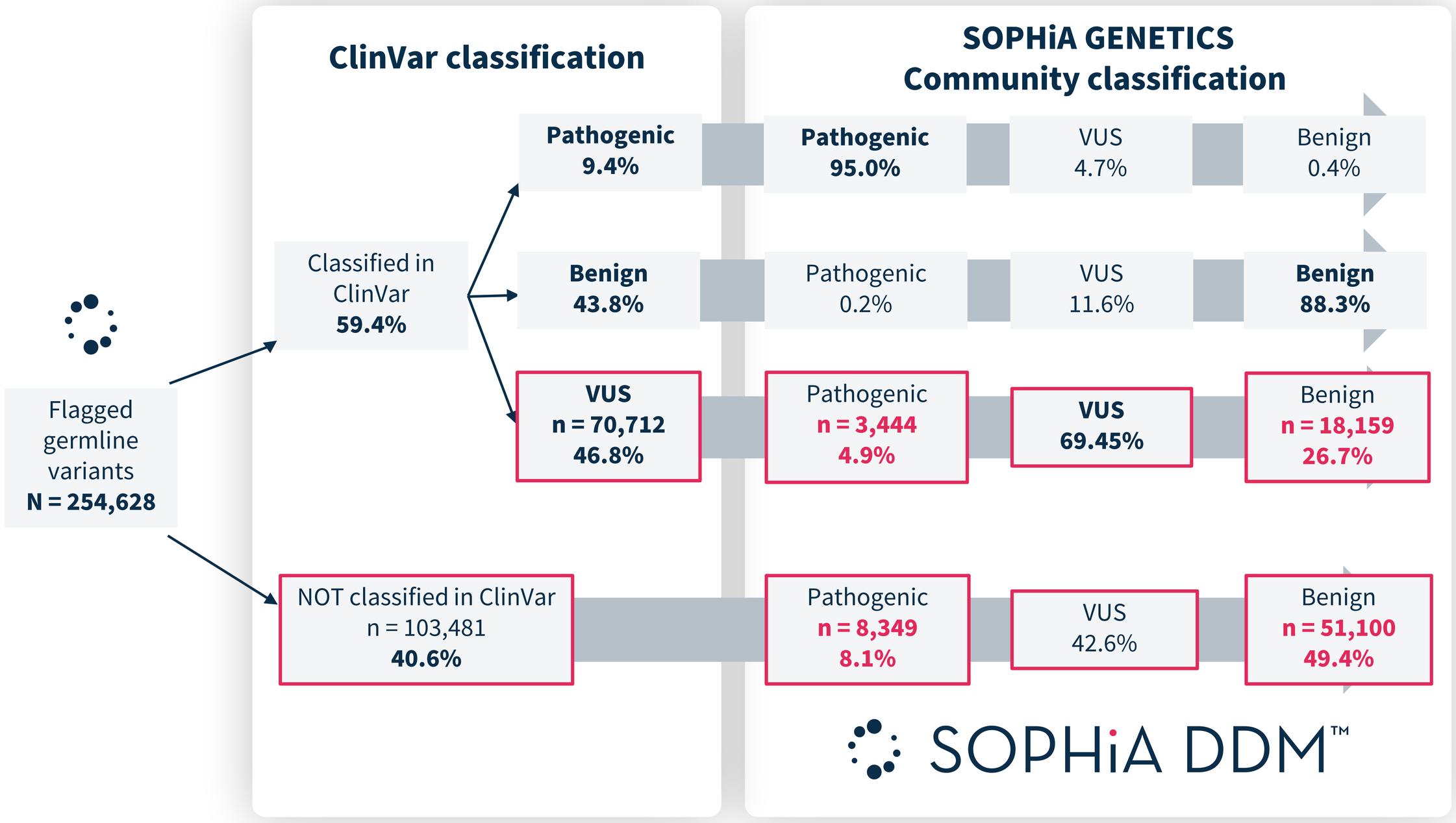

Figure 3 | Additional insights provided by the SOPHiA GENETICS Community over ClinVar for germline variant classification of ClinVar VUS and variants not classified in ClinVar. Percents in the SOPHiA GENETICS Community classification box (right) are the percent of the total number of variants with a given ClinVar classification (left) reported with a certain SOPHiA DDM™ classification e.g., 95% of variants classified as Pathogenic in ClinVar were also classified as Pathogenic by the SOPHiA GENETICS Community (calculated from data collected in 2024).

The SOPHiA DDM™ variant flagging feature provides unique collective intelligence

The SOPHiA GENETICS Expert Community is one of the largest global networks of healthcare institutions. It provides unique knowledge of variant effect, pathogenicity, and association with disease that extends beyond what is available in public databases (*Fig. 2*). Using the variant flagging feature, users can classify variants as benign, likely benign, VUS, likely pathogenic, or pathogenic based on their own assessment of the evidence found in databases, as well as the phenotype and experimental evidence available and generated in their lab. All users can easily access this information within the SOPHiA DDM™ variant interpretation view (*Fig. 1*).

A

Number of unique variants classified	Number of variant classifications
482,684	888,751

B

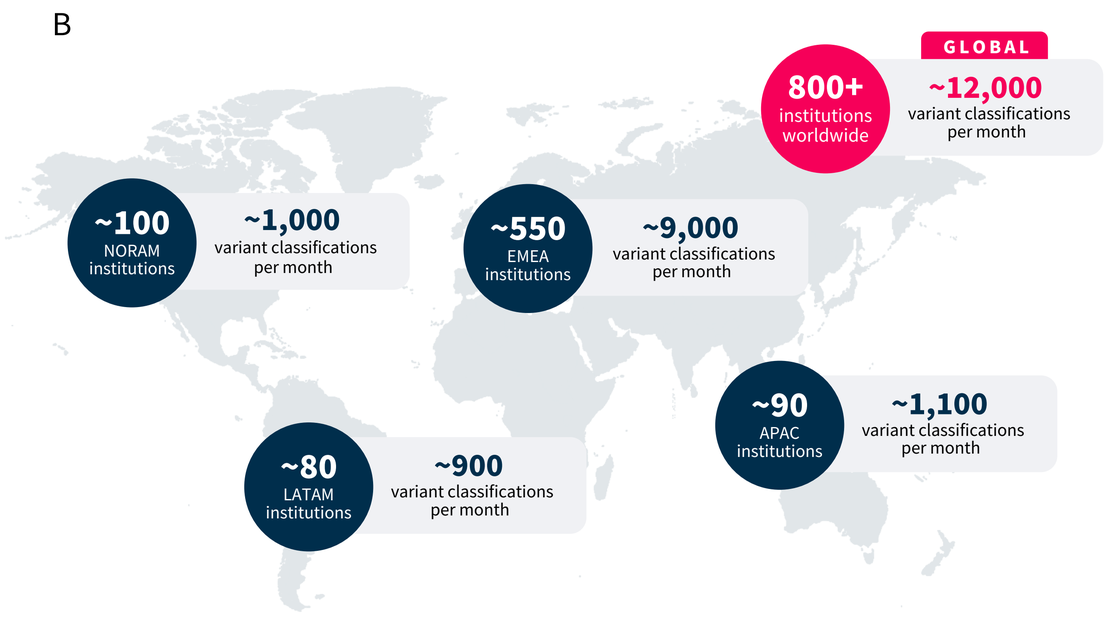

Figure 2 | Variant classifications entered by users of the SOPHiA DDM™ Platform using the variant flagging feature. *A.* Number of unique variants classified and total number of variant classifications since first use of the SOPHiA DDM™ Platform (based on data collected up to the end of 2025). *B.* Number of variant classifications per month, by region (based on data from 2024).

Over the course of analysis with the SOPHiA DDM™ Platform, more than a quarter million germline variants (N = 254,628) have been classified by the SOPHiA GENETICS Community of experts (*Fig. 3*). ~60% of these variants were also classified in ClinVar, while ~40% are unique to the SOPHiA GENETICS Community. SOPHiA GENETICS Community classifications are largely consistent with those in ClinVar, with 89% concordance for benign and pathogenic classifications. However, SOPHiA DDM™ provides substantial additional insights for ClinVar VUS – 4.9% of these variants (n = 3,444) were classified as pathogenic and 26.7% (n = 18,159) as benign through the variant flagging feature. Furthermore, for variants not classified in ClinVar, 8.1% (n = 8,349) were classified as pathogenic and 49.4% (n = 51,100) as benign by the SOPHiA GENETICS Community, demonstrating that variant flagging in SOPHiA DDM™ frequently extends insights beyond data available in public databases.

A global analysis of the variants classified by the SOPHiA GENETICS Community found that SOPHiA DDM™ users are in strong agreement with each other; most variant flags support the same classification (*Fig. 4*). This was especially clear when variants were flagged as either pathogenic or benign, with flags rarely conflicting. This demonstrates the strength of the network consensus within the SOPHiA GENETICS Community.

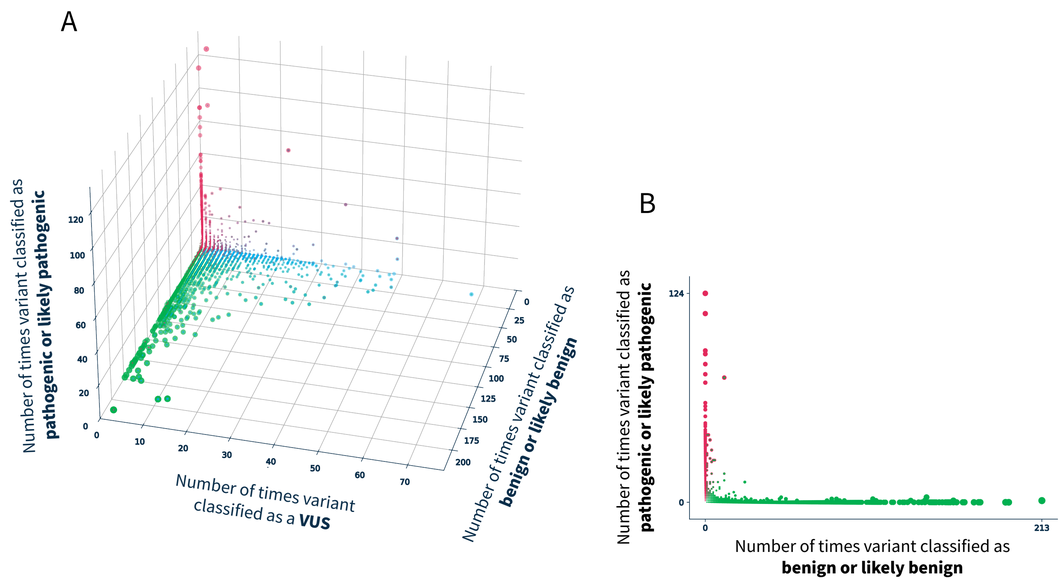

Figure 4 | Concordance of SOPHiA DDM™ user variant classification The axes represent the number of times a variant (each dot represents a variant) has been flagged with that pathogenicity classification (based on data from 2024). *A.* All flags; *B.* Pathogenic/likely pathogenic and benign/likely benign flags only.

Pink = pathogenic/likely pathogenic; Green = benign + likely benign; Blue = VUS.

The SOPHiA GENETICS Community continues to grow and strengthen

The SOPHiA GENETICS Community network is rapidly growing, with thousands of variants reported each year through the SOPHiA DDM™ variant flagging feature.

Since 2015, between 70-130,000 new variant classifications have been entered each year for 45-77,000 unique germline and somatic variants per year (*Fig. 5*), demonstrating that the SOPHiA GENETICS Community database is consistently growing in depth and breadth.

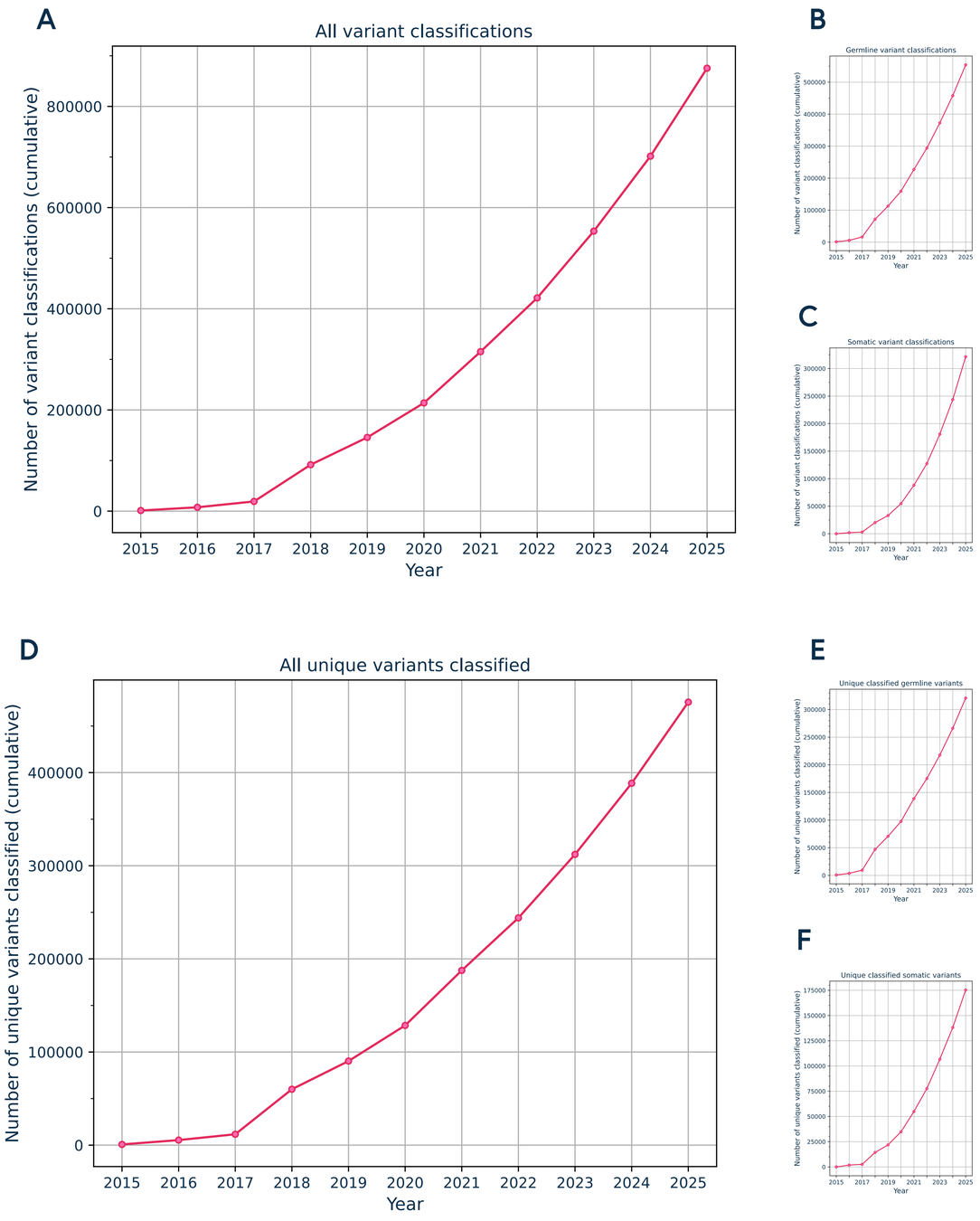

Figure 5 | Number of variant classifications and unique variants classified by the SOPHiA GENETICS Community over time using the SOPHiA DDM™ variant flagging feature. *A.* All variant classifications; *B.* Germline variant classifications; *C.* Somatic variant classifications; *D.* All unique classified variants; *E.* Unique classified germline variants; *F.* Unique classified somatic variants.

When evaluating the specific genes classified by SOPHiA DDM™ users, it becomes apparent that variant flagging is not specific to a single gene or small group, but that a diverse range of genes are covered. The gene with the most flagged germline variants is *TTN*, with 3.1% of germline variant classifications (n = 1,511) accounted for by this one gene (*Fig. 6*).

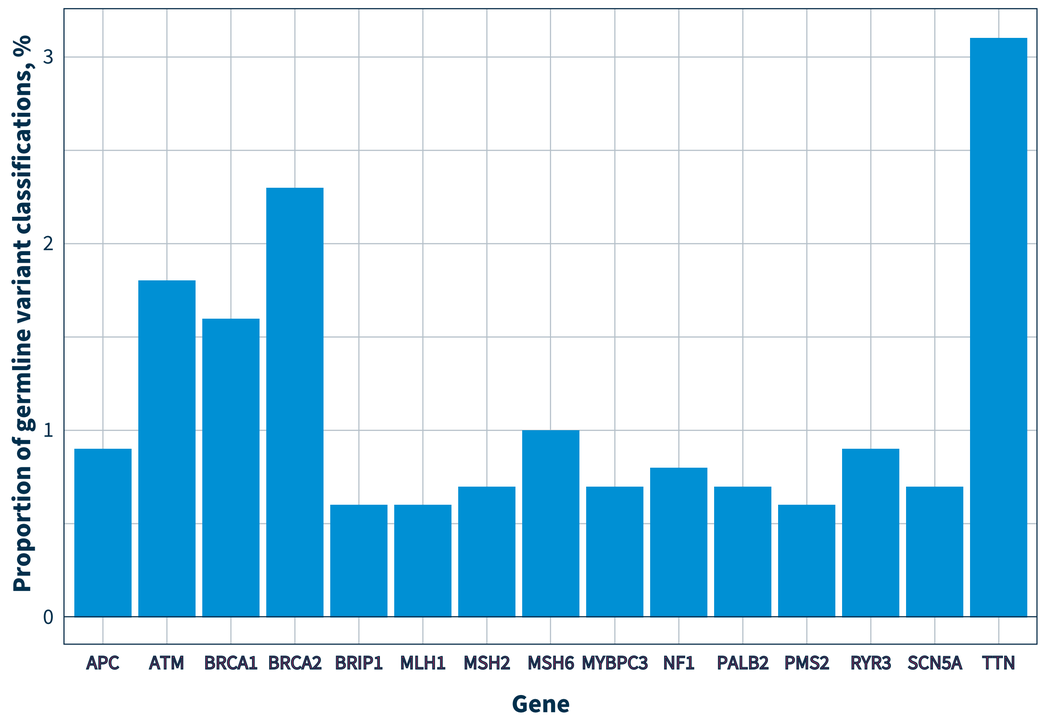

Figure 6 | Proportion of germline variant classifications associated with specific genes. The 15 genes with the highest frequency of germline variant flags are shown (based on data from 2024).

SOPHiA DDM™ users can easily leverage the SOPHiA GENETICS Community

Knowledge shared within the SOPHiA GENETICS Community can be easily accessed through the SOPHiA DDM™ Platform, which eases and speeds up variant interpretation by providing valuable information from a range of trusted sources in a single location (*Fig. 7*). For example, users can view global community variant frequencies alongside internal account frequencies in a user-friendly chart. In addition, user classifications of variants are displayed alongside the pathogenicity prediction by ACMG, SOPHiA DDM™'s proprietary AROMA algorithm, and ClinVar classification.

It is simple for SOPHiA DDM™ users to contribute to the SOPHiA GENETICS Community by using the variant flagging feature to manually classify the pathogenicity of a variant. In several mouse clicks it is possible to "Edit Variant Pathogenicity", with no complex form to fill in. Submitting a variant to ClinVar is slightly more involved, requiring the manual entry of details such as HGVS notation, genomic coordinates, pathogenicity classification, supporting evidence, and associated clinical significance.

Contributing to this collective intelligence though SOPHiA DDM™ strengthens the network consensus while also saving the classification to individual user accounts to be reported any time the same variant is detected in a sample.

SOPHiA DDM™ Community variant classification

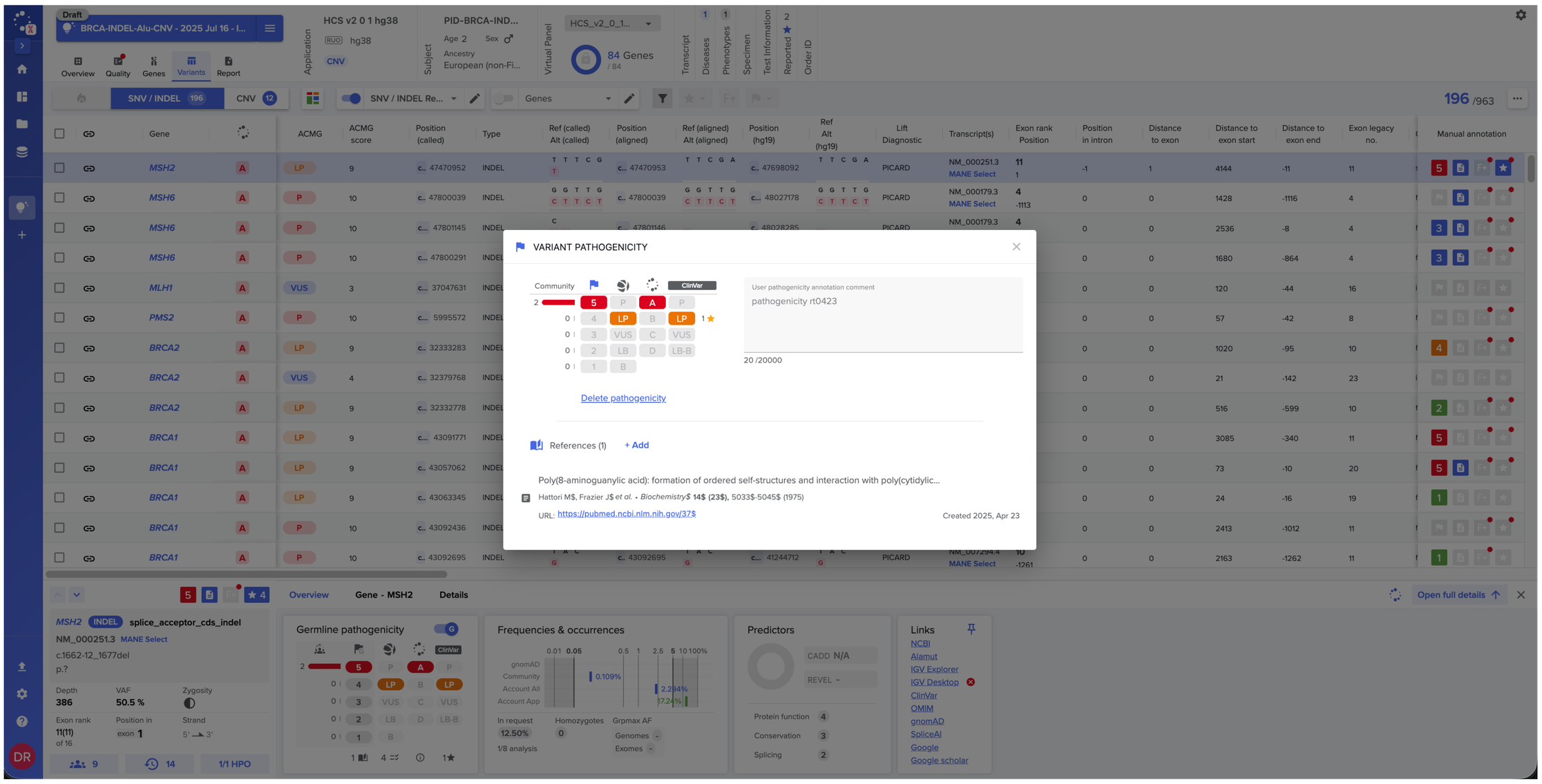

Figure 7 | SOPHiA DDM™ pathogenicity assessment window displaying comprehensive information for an Indel in *MSH2*. The main part of the screen shows the variant flagging feature that enables the manual classification/reclassification of a variant. The bottom panel provides variant classification information including user flags, ACMG, AROMA, and ClinVar, alongside global and account variant frequencies, predictors data, and links to more interpretation tools.

Community knowledge strengthens the predictive performance of the AROMA variant classification algorithm

Four proprietary algorithms optimize variant annotation and interpretation in SOPHiA DDM™. MOKA underpins variant annotation, integrating AROMA for pathogenicity prediction and ESPRESSO for phenotype-driven prioritization, which together form the EXCELSA combined variant prioritization score. The unique insights collected amongst the SOPHiA GENETICS Community strengthen the SOPHiA DDM™ AROMA predictive algorithm. The AROMA algorithm uses machine learning to classify variants in SOPHiA DDM™ as A - Pathogenic, B - Potentially Pathogenic, C - Variant of Uncertain Significance (VUS), or D - Benign. With SOPHiA GENETICS Community data at the core, machine learning techniques consider a broad range of base annotation features guided by ACMG guidelines, and knowledge generated by expert communities worldwide to predict the pathogenicity of variants. SOPHiA GENETICS Community data are curated by algorithms that extract key insights to maximize the accuracy of the AROMA algorithm. In this way, the SOPHiA DDM™ Platform leverages on the expert network of users to accelerate variant interpretation.

The updated AROMA algorithm (released in 2025) has a set of classification criteria which ingest data from the SOPHiA DDM™ variant flagging feature, Clinvar, ClinGen, and more, to predict the pathogenicity of each variant (*Fig. 8*). SOPHiA DDM™ automated ACMG classification is combined with SOPHiA DDM™ Community flags and data from curated datasets using machine learning to produce a final pathogenicity prediction.

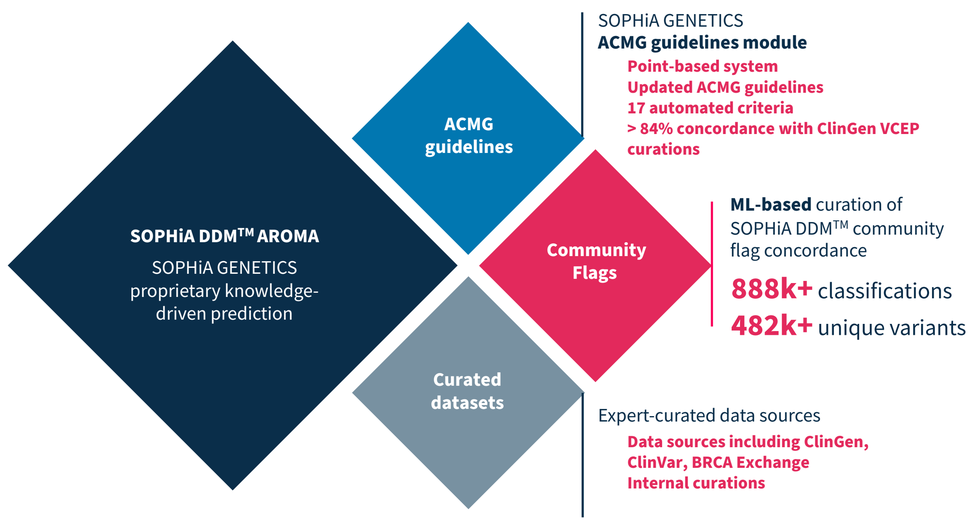

Figure 8 | Schematic of the AROMA predictive algorithm for germline variant classification in SOPHiA DDM™. AROMA is rooted in ACMG guidelines for variant interpretation and extended with machine-learning curated SOPHiA GENETICS Community data as well as other expert-curated datasets.

The AROMA algorithm delivers high analytical performance (*Fig. 9*). By integrating variant classification information from the SOPHiA GENETICS Community with internal SOPHiA DDM™ insights derived from the analysis of over 2 million genomic profiles, with data from global databases such as ClinVar and ClinGen, AROMA surpasses the performance of automated ACMG classification alone, achieving 99% sensitivity and precision.

Figure 9 | Analytical performance of the AROMA algorithm for variant classification. Sensitivity and precision of the AROMA algorithm (pink) versus automated ACMG classification (blue), benchmarked on an externally validated dataset.

Conclusions and future perspectives

The SOPHiA GENETICS Community of experts enhances variant interpretation by providing unique pathogenicity insights beyond public databases. With the intuitive SOPHiA DDM™ variant flagging feature, users can submit classifications in seconds, contributing to a global community that has amassed over 800,000 variant classifications, fostering a collaborative network of expertise.

SOPHiA GENETICS Community classifications of germline variants align closely with ClinVar but offer additional insights, particularly for VUS. More than 20,000 ClinVar-classified VUS have received refined interpretations, and nearly 60,000 variants absent from ClinVar have been classified as pathogenic or benign within SOPHiA DDM™. The strong consensus on pathogenicity assessment among users underscores the reliability of the network's collective expertise.

The SOPHiA DDM™ Platform facilitates seamless access to community-driven knowledge, accelerating and improving variant interpretation globally. This shared intelligence also strengthens the performance of the SOPHiA DDM™ AROMA predictive algorithm, which classifies variants with high analytical accuracy, streamlining decision making.

Recent advancements to the variant interpretation capabilities in SOPHiA DDM™ include the implementation of the oncogenicity algorithm, which is based on joint guidelines from ClinGen, CGC, and VICC,[8] and integrates SOPHiA GENETICS Community insights to improve the accuracy of somatic variant classification. Future developments will further enhance the SOPHiA DDM™ collective intelligence. The variant flagging feature will be upgraded to support expanded classifications, including pharmacogenomic associations and more granular VUS categories such as VUS-high, VUS-mid and VUS-low[9]. Further upgrades will incorporate functionality into the variant flagging feature that encourages users to submit their variant classifications to ClinVar. Additionally, upgrades to the Peer Networks within University Hospital Systems or across multiple institutions in the same country or region will enable more detailed variant flagging and facilitate expert discussions.

By leveraging the collective intelligence of a global expert network, SOPHiA GENETICS continues to drive advancements in precision medicine, empowering clinicians with faster, more accurate variant interpretations.

References

1. Richards S, Aziz N, Bale S, et al. Genet Med Off J Am Coll Med Genet. 2015;17(5):405-424.
2. Landrum MJ, Chitipiralla S, Brown GR, et al. Nucleic Acids Res. 2020;48(D1):D835-D844.
3. Rehm HL, Berg JS, Brooks LD, et al. N Engl J Med. 2015;372(23):2235-2242.
4. Kuchinski K, King N, Driggers J, et al. J Pharmacol Exp Ther. 2024;391(3):441-449.
5. Firth HV, Richards SM, Bevan AP, et al. Am J Hum Genet. 2009;84(4):524-533.
6. Phan L, Zhang H, Wang Q, et al. Nucleic Acids Res. 2025;53(D1):D925-D931.
7. Fowler DM, Rehm HL. Am J Hum Genet. 2024;111(1):5-10.
8. Horak P, Griffith M, Danos AM, et al. Genet Med Off J Am Coll Med Genet. 2022;24(5):986-998.
9. Bennett G, Karbassi I, Chen W, et al. MedRxiv Prepr Serv Health Sci. Published online November 13, 2024:2024.11.13.24317242.

About us

SOPHiA GENETICS (Nasdaq: SOPH) is a cloud-native healthcare tech company advancing data-driven medicine worldwide. Its SOPHiA DDM™ Platform delivers real-time insights from complex genomic and multimodal data, serving a global network of hospitals, labs, and biopharma. Learn more at **SOPHiAGENETICS.com** or on **LinkedIn**.

Where others see data, we see answers.

Want to know more?

Contact us at: info@sophiagenetics.com