The SynMall resource for characterizing the functional impact of synonymous variation

Chen Ye, Xiaoyan Li, Na Cheng, et al.

Genome Res. 2026 36: 421-431 originally published online December 17, 2025
Access the most recent version at doi:10.1101/gr.281257.125

Resource

The SynMall resource for characterizing the functional impact of synonymous variation

Chen Ye,[1] Xiaoyan Li,[1] Na Cheng,[2] Yansen Su,[3] and Junfeng Xia[1]

[1]*Information Materials and Intelligent Sensing Laboratory of Anhui Province and School of Life Sciences and Medical Engineering, Anhui University, Hefei, Anhui 230601, China;* [2]*School of Biomedical Engineering, Anhui Medical University, Hefei, Anhui 230032, China;* [3]*School of Artificial Intelligence, Anhui University, Hefei, Anhui 245041, China*

Synonymous single-nucleotide variants (sSNVs) are increasingly recognized as contributors to disease, yet existing variant annotation databases offer limited functional insights for sSNVs. Here, we present SynMall, a comprehensive resource designed to decipher the functional impact of synonymous variation. SynMall catalogs 25 million potential human sSNVs and integrates evolutionary and population information of sSNVs from 45 non-human species. For each human sSNV, SynMall provides multilevel annotations that combine American College of Medical Genetics and Genomics (ACMG)–aligned variant interpretation information, such as allele frequencies and functional effects, with more than 100 descriptors at the DNA, RNA, and protein levels. These include both handcrafted features and embeddings from large language models to support advanced representation learning. To prioritize pathogenic sSNVs, we have developed SynScore, a machine learning framework that integrates ACMG guidelines and diverse biological characteristics. Benchmark comparisons show that SynScore achieves state-of-the-art performance, validating its effectiveness for genome-wide pathogenicity inference. Furthermore, SynMall enables mechanistic exploration by investigating in silico assessments and curated literature evidence to evaluate sSNV effects on miRNA–mRNA interactions, mRNA splicing, mRNA stability, and codon usage. By consolidating these features into a unified platform, we anticipate that SynMall will serve as a valuable resource for elucidating the functional role of synonymous mutations.

[Supplemental material is available for this article.]

Synonymous single-nucleotide variants (sSNVs), which do not alter the amino acid sequence encoded by the affected codon, have historically been regarded as functionally neutral (Chamary et al. 2006). However, with the advent of high-throughput sequencing technologies, accumulating evidence has revealed that sSNVs may account for 6%–8% of all driver mutations in cancer genetics owing to single-nucleotide substitutions (Supek et al. 2014) and are associated with various diseases (Sauna and Kimchi-Sarfaty 2011). Extensive research into the functional mechanisms of sSNVs has demonstrated their capacity to influence a wide range of biological processes and molecular functions, including transcriptional regulation, mRNA splicing, translation kinetics, and protein expression (Liu 2020; Bailey et al. 2021). For instance, a recent study identified an sSNV, *PIBF1* c.954G>A, that induced exon skipping, leading to a downstream frameshift and premature termination, as confirmed by RNA analysis (Byrne et al. 2023). Similarly, a novel pathogenic sSNV in the *WAS* gene was found to disrupt splicing, thereby affecting protein expression (Sun et al. 2024). However, the functional relevance of most sSNVs remains unknown. In ClinVar (Landrum et al. 2020), sSNVs with a clear "benign/likely benign" or "pathogenic/likely pathogenic" clinical significance account for only ~2% of all sSNVs in the human genome. Furthermore, the impact of sSNVs extends beyond humans, influencing other species as well (Wang et al. 2021; Zhao et al. 2021; Tan et al. 2023), and has applications in breeding optimization. Also, the neutral information implied by mutations tolerated under natural selection makes mutations in closely related species one of the key focuses (Gao et al. 2023). Nevertheless, no dedicated resource currently exists for sSNVs in non-human species.

Given the growing evidence linking sSNVs to disease, several platforms have been developed to investigate their roles. SynMICdb (Sharma et al. 2019) focuses on sSNVs in cancer genomes, providing integrated data on gene and cancer associations. Another resource, dbDSM (Wen et al. 2016), catalogs disease-related sSNVs manually curated from published literature. However, these databases are highly specialized, covering less than ~3% of sSNVs in the human genome, and thus cannot serve as a comprehensive resource. On the other hand, whole-genome variation resources such as VarCards2 (Wang et al. 2024), FAVOR (Zhou et al. 2023), and CADD v1.7 (Schubach et al. 2024) encompass broad coverage of variant types and interpretations but fall short in elucidating the specific mechanisms by which sSNVs contribute to pathogenicity. Moreover, the sequence and structural properties of variants are widely used in the development of predictive models for sSNVs and demonstrate discriminative power (Livingstone et al. 2017; Zhang et al. 2017). However, obtaining diverse representations of sSNVs often requires complex workflows and programming expertise.

To address these limitations and combine the strengths of existing resources, we develop SynMall, a comprehensive platform designed to store all potential sSNVs across the entire human genome, along with sSNVs from other species. SynMall provides detailed variant interpretations under the American College of

Corresponding authors: jfxia@ahu.edu.cn, lixiaoyan@ahu.edu.cn
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.281257.125.

Medical Genetics and Genomics (ACMG) guidelines (Richards 2015) and biological knowledge, incorporating data on allele frequencies (AFs), in silico predictions, disease associations, de novo variants, regulatory information, and more. To support further exploration, SynMall offers a suite of feature-oriented computational tools to characterize sSNVs and assess their impact on biological processes. By effectively utilizing the information from published literature, SynMall offers up-to-date, curated knowledge for researchers. Regarding non-human species, SynMall also incorporates 5,033,859 sSNVs from 29 vertebrate species, of which 495,726 are mapped to the human genome through multiple sequence alignment (MSA) and share human annotation information. In addition, SynMall includes 11,841 genotype-to-phenotype associations from 21 domesticated animals and cultivated plants. To enhance usability, we design a user-friendly web interface that supports a wide range of functionalities. We anticipate that SynMall will serve as a universal resource for advancing research into the role of sSNVs.

Results

Overview of SynMall

SynMall is a genome-wide platform for sSNVs that provides comprehensive characterization and supports both biological and computational studies. As shown in Figure 1A, first, SynMall identifies all potential sSNVs on protein-coding transcripts and integrates records from widely used genome-wide variant databases. Second, it collects extensive annotations, categorized into mutation level (describing the variant itself) and region level (describing the surrounding genomic context of the variant). Third, SynMall assembles three core modules: a customizable batch annotation module built on integrated data; a feature extraction module that captures contextual properties of sSNVs across DNA, RNA, and protein levels; and an analysis module that investigates potential biological mechanisms affected by sSNVs. Fourth, SynMall also continuously curates and updates structured knowledge from recent literature. Beyond human data, SynMall incorporates sSNVs from non-human species: (1) it collects vertebrate sSNVs from multiple sources and maps them onto the human genome via MSA; (2) it also integrates genotype–phenotype associations for sSNVs from economically important species; and (3) it compiles population AFs from diverse primates and maps the corresponding sites to the human reference genome. A cross-comparison between SynMall and other variant databases (Supplemental Table S1) shows that SynMall outperforms in functionality, resource, and usability. The following sections describe the main modules of SynMall in greater detail.

Universal annotation resource characterizes and prioritizes genome-wide sSNVs

In the annotation module, SynMall integrates diverse resources to comprehensively describe sSNVs (Fig. 1B), including 32 in silico prediction scores, 14 external databases, 13 population AF sources, 10 conservation-related properties, and 10 types of regulatory information.

SynMall catalogs 92 million transcript-level sSNVs, corresponding to 25,735,732 sSNVs at the genomic level (Fig. 1C). Notably, a substantial majority (72.02%) are unobserved in existing data (Fig. 1D), highlighting the need to prioritize sSNVs and gather supporting evidence. Based on the assembled features, we therefore design a machine learning strategy, SynScore (see Methods), which integrates clinically important ACMG annotations, such as functional score and AF, along with biological characteristics, including regulatory elements and epigenetic information. In the independent test set, pairwise comparisons with other variant effect predictors (VEPs) show that SynScore achieves the highest AUC and AUPR (Fig. 1E; Supplemental Fig. S1A). When evaluated against other VEPs on a common subset without missing values (Supplemental Fig. S1B,C), SynScore also exhibits the highest AUC (0.922) and AUPR (0.951), demonstrating that the annotation framework constructed in SynMall can effectively distinguish pathogenic sSNVs. The strong performance of SynScore stems from its integration of functional scores and conservation information (Supplemental Fig. S1D), as well as additional features rarely included in previous tools, such as primate AFs and superenhancers, which also demonstrate notable importance. Although the CADD functional score exhibits the highest feature importance, removing it reduces performance by only 0.3% (Supplemental Fig. S1E), indicating that SynScore does not solely rely on CADD and that other features can provide complementary information.

We analyze SynScore for all sSNVs located in actionable genes (Miller et al. 2023). As shown in Figure 1F, we exclude *TTN* because it is considered actionable only for truncating variants. In contrast, genes such as *MYH7*, *PTEN*, *MAX*, *TPM1*, *MYH11*, and *CACNA1S*, which are related to cardiovascular and cancer diseases, exhibit higher significance or mean SynScore values, suggesting that sSNVs in these genes are more likely to be functional. Moreover, sSNVs with higher SynScore values (Fig. 1G) tend to be enriched in genes with a lower loss-of-function observed/expected upper-bound fraction (LoEUF) from gnomAD (Chen et al. 2024b), in which a lower LoEUF indicates stronger constraint. This underscores the potential of SynScore as a prioritization strategy for candidate pathogenic sSNVs. We further rank all sSNVs in COSMIC v100 (Tate et al. 2019) by SynScore and select the top 0.1%, corresponding to 318 genes. Using DAVID (Sherman et al. 2022) for KEGG pathway enrichment analysis (Fig. 1H), we find that these genes are enriched in pathways such as arrhythmogenic right ventricular cardiomyopathy and hypertrophic cardiomyopathy, suggesting their contribution to cardiac pathogenesis.

To further investigate the details, we visualize the distribution of sSNVs across the coding sequence (CDS) of these actionable genes (Fig. 1I; Supplemental Fig. S2). A Mann–Whitney *U* test reveals significant differences in SynScore distribution between the splice site consensus (SSC) regions and internal exons, with sSNVs located at SSC regions showing higher SynScores, indicating a higher pathogenic potential. Furthermore, spatial clustering analysis using InterPro (Blum et al. 2025) reveals that high-SynScore sSNVs tend to localize in disordered regions of the encoded proteins. Specifically, in *MYH11*, *TPM1*, and *MAX*, high-scoring sSNVs are enriched in coil and consensus disorder prediction regions (Fig. 1J; Supplemental Fig. S3), consistent with previous findings on regSNPs-splicing (Zhang et al. 2017). Although sSNVs do not alter amino acid sequences, codon changes may influence translation initiation or cotranslational folding, thereby leading to functional consequences (Diederichs et al. 2016). In the following, we further investigate the computational tools and resources in SynMall. The demonstrations include SpliceBERT, GPN-MSA, ERNIE-RNA, ISM, ESEFinder 3.0, SpliceAI, MMSplice, CADD-Splice, dbscSNV, DeltaSplice, RNAfold, sincFold, RNAsnp, TarBase, Mimosa, TargetNet, tRNA adaptation index (gtAI), codon adaptation index (CAI), and relative synonymous codon usage

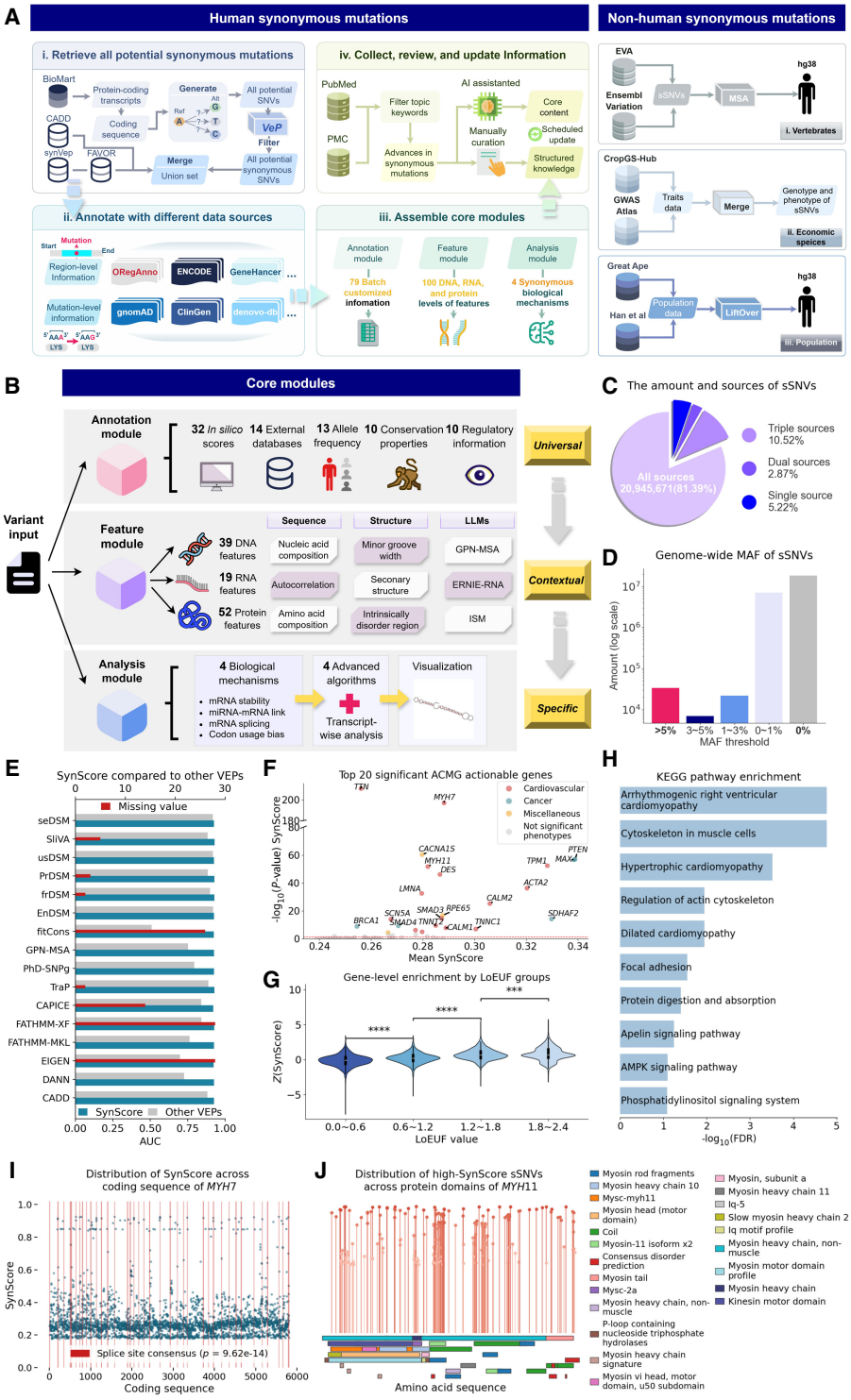

Figure 1. Architecture of the SynMall database and its annotation-based analyses. (*A*) An overview of SynMall, including integrated resources for both human and non-human sSNVs. (*B*) SynMall assembles three core modules: the annotation module offers 79 universal annotations; the feature module provides more than 100 contextual features for each variant; and the analysis module focuses on four sSNV-specific biological mechanisms with predictive algorithms and visualizations. (LLMs) Large language models. (*C*) Variants are aggregated from four resources: transcript-generated, CADD v1.7, synVep, and FAVOR. (All sources) sSNVs recorded in all four data sets, (single/dual/triple sources) variants present in only some of them. (*D*) Genome-wide minor allele frequency (MAF) landscape of sSNVs, integrated from multiple population sequencing projects. (*E*) Comparison of SynScore with other variant effect predictors (VEPs) on an independent test set. (*F*) SynScore patterns of sSNVs in ACMG-actionable genes. (*G*) *Z*(SynScore): gene-level enrichment of high-score sSNVs across loss-of-function observed/expected upper-bound fraction (LoEUF) groups. Significance (two-sided Mann–Whitney *U* test) is indicated by asterisks: (***) $P<0.001$, (****) $P<0.0001$. (*H*) Top 10 overrepresented KEGG pathways for genes with high SynScores in COSMIC. (*I*) Positional distribution of SynScore along the coding sequence of *MYH7*. (*J*) Spatial clustering of *MYH11* sSNVs with SynScore > 0.5 across protein domains.

(RSCU). References for all tools and resources mentioned are provided in the Supplemental Documents.

Scalable contextual features capture discriminative latent patterns

Compared to the annotation module, the feature module enables flexible exploration of sSNV-related features at varying lengths. It integrates 39 DNA-level, 19 RNA-level, and 52 protein-level features to describe the context surrounding sSNVs (Fig. 1B). To extract sequence- and structure-based properties, we assemble a suite of tools that generate high-throughput predictions of DNA shape features, RNA secondary structures with minimum free energy (MFE), diverse sequence-derived attributes, and residue-level protein characteristics such as intrinsic disorder, secondary structure, and solvent accessibility. Beyond these handcrafted features, SynMall also integrates 10 biological language models, considering their growing success in variant effect prediction tasks (Cheng et al. 2023). At the DNA level, we include models that distinguish at multiple-sequence alignment, long-range dependency modeling, reverse-complement consistency, and epigenetic feature prediction. At the RNA level, SynMall incorporates models with strong performance in splicing prediction and RNA structural stability assessment. At the protein level, SynMall adopts an advanced model capable of extracting rich structural information from standalone protein sequences.

To demonstrate the utility of the feature module, we analyze a curated data set (Supplemental Methods) using extracted DNA, RNA, and protein features. We submit the job after selecting all feature categories; uploading a VCF file; choosing context lengths of 16 bp, 70 bp, and 10 amino acids; and specifying hg38 (Fig. 2A). We then compare the feature distributions between benign and pathogenic sSNVs. At the DNA level (Fig. 2B), ACT-count and electrostatic potential from the DNA shape differ significantly between the two groups, consistent with the importance of DNA shape features reported in EPEL (Bi et al. 2025). At the RNA level, nucleotide chemical properties and ΔMFE show distributional differences (Fig. 2C). Although the ΔMFE effect is modest, aligning with synVep (Zeng et al. 2021), in which mRNA stability features show limited predictive power. At the protein level, pseudo-K-tuple-reduced amino acid composition (PseKRAAC) and absolute solvent accessibility differ between pathogenic and benign variants (Fig. 2D), consistent with findings from regSNPs-splicing (Zhang et al. 2017). We further use a 129 bp context length at the DNA and RNA levels and 43 amino acids at the protein level to visualize embeddings from biological language models. Using SpliceBERT (Fig. 2E), GPN-MSA (Fig. 2F), ERNIE-RNA (Fig. 2G), and ISM (Fig. 2H) as examples, SynMall generates embeddings for both reference and alternate sequences. UMAP (McInnes et al. 2018) shows that "reference" and "alternate" embeddings do not separate benign and pathogenic variants, whereas "difference" embeddings offer slightly better discrimination. Because of the close-by sampling strategy of the curated data set, positive and negative samples are close in genomic positions and therefore share highly similar sequence contexts. As a result, embeddings of sSNVs from language models still require further feature extractors to achieve better separability. The result provides a straightforward demonstration of the module's capabilities, highlighting several features that may help distinguish between benign and pathogenic sSNVs. It encourages users to further explore combinations of features and context lengths to identify those with the greatest predictive power for sSNV pathogenicity.

Synonymous-specific mechanisms interpret the functional impact of sSNVs

The analysis module focuses on four major biological mechanisms of sSNVs: mRNA secondary structure, miRNA–mRNA binding, mRNA splicing, and codon usage bias (Lin et al. 2023). It integrates classic and advanced algorithms to provide intuitive, visual interpretations of sSNV effects, including transcript-specific analyses (Fig. 1B).

To demonstrate its utility, we first analyze *COL4A3* c.765G>A, a pathogenic sSNV that disrupts splicing (Daga et al. 2025). ESEFinder 3.0 identifies motifs for SF2/ASF (IgM-BRCA1), SF2/ASF, and SC35, which serve as key splicing factors, in the reference sequence. After the mutation, the motif scores for SF2/ASF (IgM-BRCA1) and SF2/ASF change by 0.29 and 0.45, and the SC35 is lost, suggesting a potential disruption of splicing regulation (Fig. 3A). To further evaluate splicing effects, SynMall integrates multiple prediction tools. SpliceAI predicts a high donor loss score of 0.90, and MMSplice gives a donor score of −4.70, both indicating donor site disruption. CADD-Splice assigns a Phred score above 15, suggesting a potentially deleterious splicing variant. dbscSNV reports scores of 0.99 and 1.00 from random forest and AdaBoost models, indicating strong splicing disruption (Fig. 3B). Together, these results suggest that the mutation disrupts normal splicing.

Next, to illustrate the module's flexibility, we analyze *ACE* c.2350G>A, a sSNV reported to reduce mRNA stability (Liang et al. 2013). We assess its effect under various sequence contexts (50 bp, 100 bp, 200 bp, and entire transcript) in the canonical transcript ENST00000290866. In the 50-bp context, RNAfold and sincFold predict ΔMFE values of +1.60 and +4.60, respectively (Fig. 3C), suggesting increased instability. RNAsnp yields a significant *P*-value of 0.07, indicating substantial local structural impact. Across longer contexts, RNAfold's ΔMFE ranges from +0.30 to +2.00 (Supplemental Fig. S4). Then, we examine the same sSNV in another transcript, ENST00000582761, under a 100-bp context. sincFold and RNAfold predict ΔMFE values of +3.33 and +2.00, respectively, and RNAsnp produces a *P*-value of 0.03, all suggesting a more unstable structure compared with the canonical transcript (Supplemental Fig. S5). Overall, these results indicate that this mutation leads to mRNA structural instability under different transcript isoforms.

The analysis module also evaluates sSNV effects on miRNA–mRNA interactions. For example, we examine the sSNV *DFFA* c.795G>A in relation to hsa-miR-9-5p. SynMall identifies the mutation within a predicted miRNA target site (Fig. 3D) and retrieves interaction data from TarBase, which reports experimentally validated binding at Chr 1: 10,461,689–10,461,701. TargetScan predicts a high context score percentile of 79.00, suggesting the site as a favorable target. By evaluating miRNA binding to the mRNA before and after the mutation, Mimosa predicts a classification of one, and TargetNet predicts a probability of ∼0.84 for both cases, indicating the interaction is likely preserved, with higher values reflecting stronger confidence in miRNA–mRNA binding. Together, these results suggest that although the sSNV resides within a miRNA target site, it is unlikely to disrupt the interaction.

The analysis module also examines codon usage changes caused by sSNVs. For example, *USP28* c.1344T>G changes the codon from GTT to GTG, increasing the gtAI and the CAI, suggesting enhanced translation efficiency and preference for highly expressed codons. Although the RSCU slightly decreases, both codons remain preferentially used (Fig. 3E). SynMall also

incorporates tissue-specific codon usage data, showing that GTG is generally more abundant than GTT across tissues (Fig. 3F).

Literature mining reveals recent findings and curated evidence

SynMall integrates the latest research on sSNVs to keep users up to date with ongoing discoveries. The "trend" page features a monthly keyword cloud highlighting emerging topics and includes distilled summaries of recent studies (Supplemental Fig. S6A,B). To address the limited availability of pathogenic evidence for sSNVs, SynMall continuously curates relevant literature. To date, we have manually reviewed 149 recent publications, adding 88 sSNVs annotated with ACMG evidence levels, which are compiled in the "curation" section. To further aid in understanding the

Figure 2. Case usage in the SynMall feature module. (*A*) Workflow for feature extraction from a given variant call format (VCF) file. (*1*) Select the molecular category: DNA, RNA, or protein. (*2*) Choose feature types to compute. (*3*) Upload the VCF file. (*4*) Specify the sequence context length and reference genome version. (*B*) DNA-level features distinguishing functional sSNVs. (*C*) RNA-level features distinguishing functional sSNVs. (*D*) Protein-level features distinguishing functional sSNVs. (PseKRAAC) Pseudo K-tuple-reduced amino acid composition. (*E*) UMAP visualization of SpliceBERT embeddings with a 129-bp window. (*F*) UMAP visualization of GPN-MSA embeddings with a 129-bp window. (*G*) UMAP visualization of ERNIE-RNA embeddings with a 129-bp window. (*H*) UMAP visualization of ISM embeddings with a 43-amino-acid window. (Reference) Embeddings of the reference sequence surrounding the mutation, (alternate) embeddings of the alternate sequence, (difference) alternate minus reference embeddings.

molecular mechanisms underlying sSNV pathogenicity, SynMall adopts the classification system defined in dbDSM. Among all annotated mechanisms, splicing regulation is the most prevalent, accounting for 56% of cases (Supplemental Table S2). In addition, SynMall provides comprehensive literature-based information, including associated diseases or disorders, key sentences, and

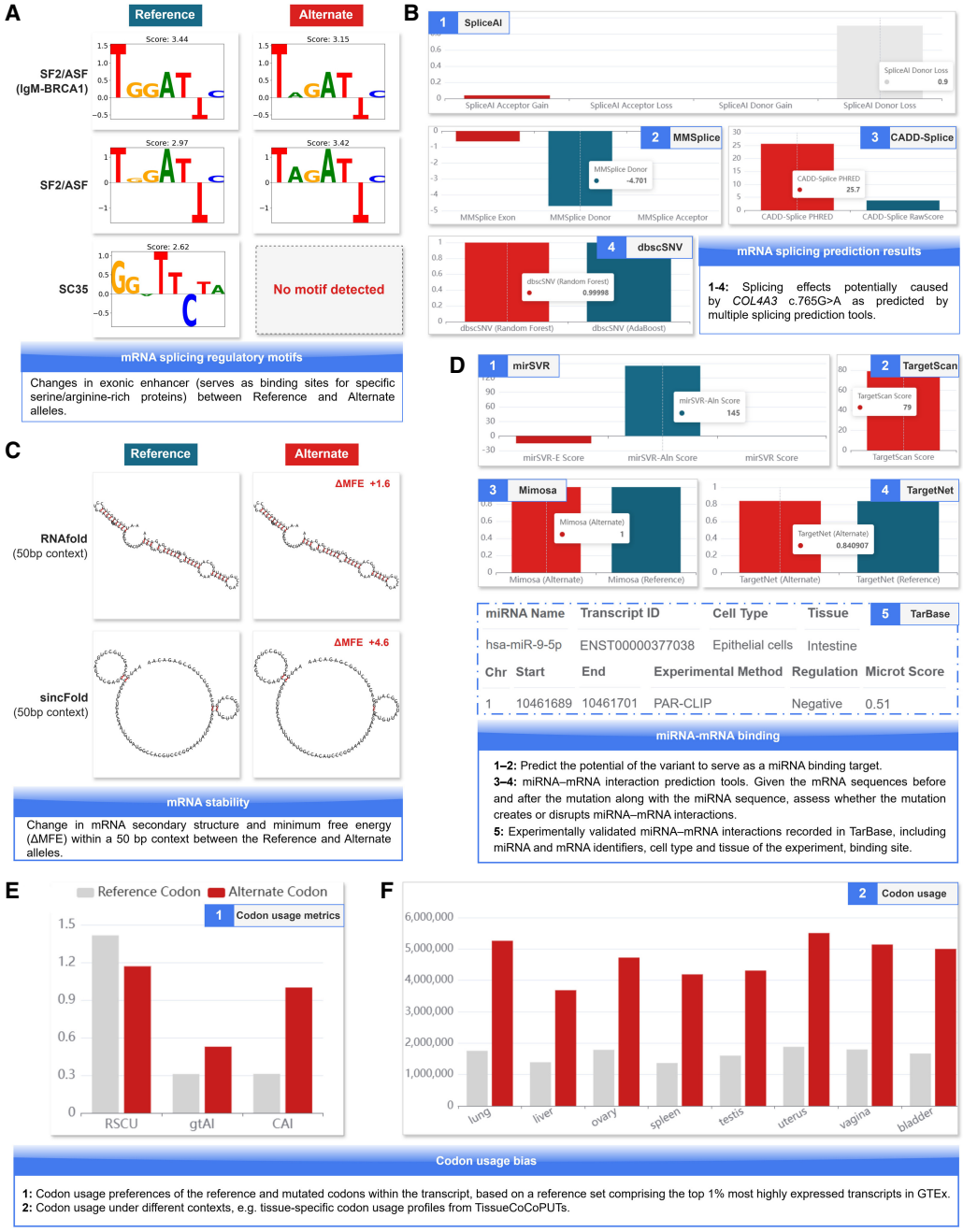

Figure 3. Case usages in the analysis module of SynMall. (*A*) SynMall uses ESEFinder 3.0 to compare the mRNA sequences before and after the *COL4A3* c.765G > A mutation, assessing whether the variant alters serine/arginine-rich (SR) protein binding sites and may affect alternative splicing regulation. (*B*) SynMall provides splice prediction scores for *COL4A3* c.765G > A: (*1*) SpliceAI predicts the gain or loss of donor/acceptor sites caused by the mutation; (*2*) MMSplice evaluates the likelihood of the site being a functional donor or acceptor; (*3*) CADD-Splice assesses the pathogenicity of splicing variant; and (*4*) dbscSNV uses the ensemble model to estimate the mutation's likelihood of altering splicing. (*C*) SynMall predicts the mRNA secondary structures of *ACE* c.2350G > A using RNAfold and sincFold and calculates changes in minimum free energy (MFE) to evaluate structural stability. (*D*) SynMall analyzes *DFFA* c.795G > A for miRNA–mRNA interaction changes: (*1*,*2*) mirSVR and TargetScan predict whether the variant is located in a potential miRNA binding site; (*3*,*4*) SynMall uses Mimosa and TargetNet to assess whether the mutation creates or disrupts interactions with specific miRNAs based on altered mRNA sequences; and (*5*) SynMall checks whether the variant overlaps experimentally validated miRNA–mRNA interaction sites, including binding positions, experimental context, and more. (*E*) SynMall presents codon-level metrics for *USP28* c.1344T > G, including relative synonymous codon usage (RSCU), codon adaptation index (CAI), and gene-specific tRNA adaptation index (gtAI). (*F*) The codon usage count of GTT and GTG across various tissue types.

mutation details, allowing users to trace and validate each curated evidence item (Supplemental Fig. S6C).

Non-human sSNV resource

In addition to human sSNVs, SynMall incorporates data from non-human species. For primates, it includes 594,937 sSNVs across five species and maps the sSNVs to the human genome. Notably, 509,873 human sSNVs exhibit a minor AF > 5% in at least one primate, suggesting potential benign effects owing to evolutionary tolerance. For vertebrates, SynMall collects sSNVs and maps them to the human genome through MSA, thus sharing functional annotation. As a result, 495,726 sSNVs are successfully aligned to human coordinates. Additionally, for economically important species, SynMall integrates 11,841 genotype-to-phenotype associations of sSNVs spanning 732 traits, facilitating research on the potential roles of sSNVs in breeding and trait improvement.

Case applications using SynMall

To demonstrate the practical application of SynMall, we conduct two case studies using the keywords "cardiovascular" and "breast," focusing on interpreting variants of uncertain significance (VUSs), as our analysis indicates that high-SynScore sSNVs are enriched in actionable genes associated with cardiovascular and cancer diseases.

In the first case, searching "cardiovascular" (Supplemental Fig. S7A) retrieves multiple external resources (Supplemental Fig. S7B). We focus on ClinVar variants labeled "uncertain significance" and rank them by SynScore. The top variant, 3:g.38557231C > T, links to detailed annotations (Supplemental Fig. S7C) covering conservation, in silico scores, and AF. SynMall assigns a PP3 (multiple lines of computational evidence support a deleterious effect on the gene or gene product) level under ACMG guidelines, as multiple tools indicate a functional effect. The analysis module (Supplemental Fig. S7D) shows that dbscSNV and DeltaSplice predict altered splicing with potential acceptor loss; ESEFinder 3.0 identifies involvement in two serine/arginine-rich (SR) protein motifs, which involve accurate splicing; RNAfold predicts reduced mRNA stability; and the sSNV changes a preferred codon to an unpreferred one, potentially lowering translation efficiency. Through a literature search, SynMall identifies two recently reported sSNVs linked to cardiovascular disease.

In the second case, we use "breast" as the keyword (Supplemental Fig. S8A). Following the same procedure as in the first case, we focus on the ClinVar VUS 13:g.32357903A > T with the highest SynScore (Supplemental Fig. S8B). By clicking the variant, detailed annotations show SynMall assigns PP3 and PM2 (absent from large sequencing project) levels under ACMG guidelines, as multiple computational tools predict a functional effect, and the variant is absent from large sequencing projects (Supplemental Fig. S8C). The analysis module shows that DeltaSplice and SpliceAI predict donor gain, suggesting a novel splice site; sincFold indicates altered mRNA secondary structure; and the sSNV caused reduced RSCU (Supplemental Fig. S8D). A literature search identifies eight additional records, illustrating that sSNVs can affect splicing and transcriptional regulation, contributing to breast cancer and carcinogenesis.

These two cases demonstrate that SynMall not only provides diverse resources for diseases of interest but also assigns predictive SynScores and ACMG evidence levels for VUS, aiding target prioritization. Additionally, it highlights potential molecular mechanisms of sSNVs through computational analyses and offers sSNVs with literature-supported evidence, serving as a reference for the same keyword context.

Web interface of SynMall

Construction and main content

We provide a concise overview of the web interface and its functionalities. The homepage serves as the central navigation hub, offering access to four primary entries: query, annotation, feature, and analysis (Fig. 4A). A quick search entrance is embedded on the homepage, supporting diverse search patterns and reference genomes (Fig. 4B). To accommodate batch processing and customized annotation, the platform accepts up to 50,000 variants in a VCF file per job submission (Fig. 4C). SynMall enables users to retrieve DNA, RNA, and protein properties of sSNVs by only uploading a VCF file (Fig. 4D). The analysis module is specifically designed to investigate well-researched mechanisms of sSNVs (Fig. 4E). The literature module provides manually extracted knowledge, detailing how sSNVs contribute to disease by influencing specific biological processes (Fig. 4F). Additionally, a species browse page offers access to data from diverse non-human species hosted on SynMall (Fig. 4G).

Diverse query strategy

SynMall features three primary search modes. First is single query (Supplemental Fig. S9A): SynMall supports four common query options by default, including Ensembl transcript ID, genomic region, genomic coordinate, and dbSNP ID. Additionally, it supports five more search options for human, including HGVSc, HGVSg, gene name, Ensembl gene ID, and RefSeq ID. Second is keyword query (Supplemental Fig. S9B): Users can search for relevant resources using natural language terms or phrases. Third is batch query (Supplemental Fig. S9C): A batch query function allows users to submit up to 1000 entries. It takes ~7.35 sec to query 1000 variants through the "genomic coordinate" format and 2.99 sec to query 1000 variants through the "dbSNP ID" format (Supplemental Table S3).

Bulk self-service annotation

We offer annotations across six major categories (Supplemental Fig. S10A). Users need to upload a VCF file and specify the reference genome version of the input (Supplemental Fig. S10B). SynMall supports annotating more than 50,000 variants in a single run.

Convenient species browser

For vertebrate species, users can click the "evidence status" and "hg38" switch to refine their results with records possessing evidence support and mutations that can be mapped to the human genome (Supplemental Fig. S11A). Regarding economic species, users can filter records based on traits of interest and *P*-value, enabling the extraction of significant genotype–phenotype associations (Supplemental Fig. S11B). As for primate species, users can narrow down entries by AF (orange box). Additionally, selecting a specific primate species (blue box) can add a nonnull condition for this species (Supplemental Fig. S11C).

Database update

To maintain the timeliness and accuracy of SynMall, we perform regular updates across four major resource types: variants,

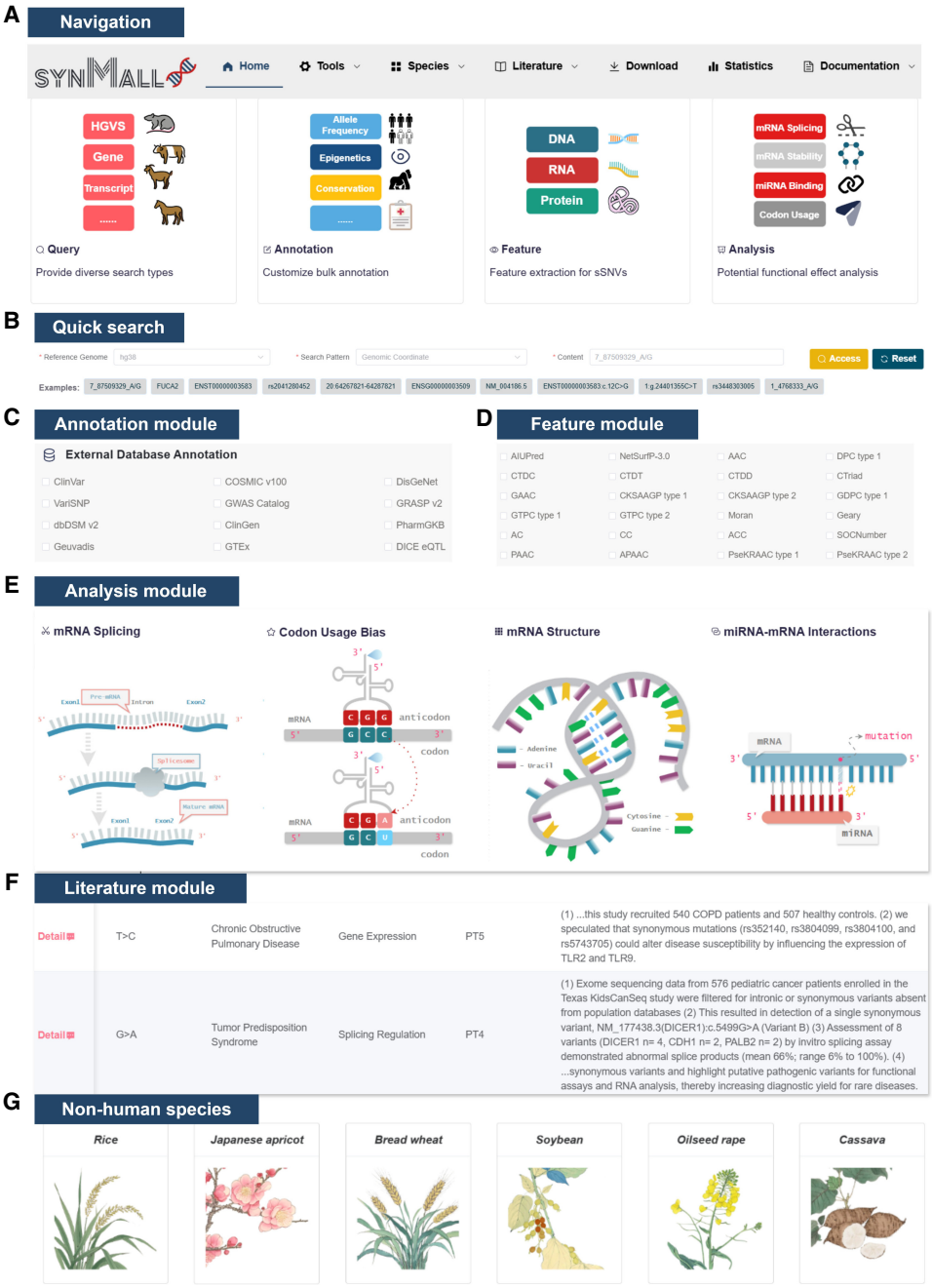

Figure 4. The overview interface of SynMall. (*A*) Navigation bar. (*B*) Quick search entrance on the "home" page. (*C*) Annotation module on the "tools" page. (*D*) Feature module on the "tools" page. (*E*) Analysis module on the "tools" page. (*F*) Curation knowledge on the "literature" page. (*G*) Entrance to different non-human species on the "species" page.

literature, external database annotations, and variant-centered computational tools. Given SynMall already provides all human sSNVs, for non-human species, we aggregate data annually from Ensembl Variation (Hunt et al. 2018), EVA (Cezard et al. 2022), GWAS Atlas (Liu et al. 2023), and CropGS-Hub (Chen et al. 2024a) and will add sSNVs from the Genome Variation Map (Li et al. 2021) within the next year, which recently provided downloadable variant consequence annotations. SynMall retrieves literature daily from PMC and PubMed and uses AI models to summarize content. Each quarter, we manually curate full-text articles to extract structured knowledge. External resources with periodic updates, such as ClinVar and COSMIC, are synchronized quarterly. New data types, including VEPs and feature extraction tools, are integrated semiannually with an emphasis on tools relevant to sSNVs.

Discussion

Recent research has increasingly demonstrated the functional impact of sSNVs, revealing that sSNVs can indeed pose potential risks

to human health and contribute to pathogenicity. Despite these advances, the functional significance of a vast number of sSNVs remains unknown. To facilitate the interpretation of sSNVs, we present SynMall, a comprehensive platform that systematically catalogs 25 million potential human sSNVs and further incorporates sSNVs from 30 vertebrate and primate species, as well as genotype–phenotype associations from 21 domesticated animals and crops. To achieve multilevel sSNV characterization, SynMall combines 79 annotation resources covering AF, evolutionary conservation, and functional prediction and further derives more than 100 descriptors at the DNA, RNA, and protein levels. These features incorporate both handcrafted properties and embeddings generated from 10 cutting-edge biological language models, enabling advanced representation learning and predictive analyses. Following ACMG guidelines and biological characteristics, SynMall develops a genome-wide prioritization model, SynScore, which demonstrates superior performance compared with existing VEPs. Furthermore, SynMall supports mechanistic interpretation by providing in silico evaluations of sSNV impacts on miRNA–mRNA interactions, mRNA splicing and stability, and codon usage. It also curates literature-centered evidence, including ACMG levels, disease information, and pathogenic mechanisms, thereby offering a valuable resource for understanding the pathogenicity of sSNVs. Despite its extensive data and functionalities, SynMall has certain limitations that should be acknowledged. First, SynMall currently lacks tools and data for tissue-, cell-, and disease-specific analyses, making it unable to assess sSNVs across different contexts. In the future, we plan to incorporate cell-specific computational tools and resources, such as DeepFun (Pei et al. 2021) and WebCSEA (Dai et al. 2022). Second, we employ widely used and advanced algorithms to calculate the effects of sSNVs on mRNA splicing, mRNA structure, and more. However, these results should be considered as indicative rather than definitive, as they serve primarily as references for further investigation. Additionally, for non-human species, annotations are currently limited to sSNVs that can be mapped to the human reference genome based on MSA, as most existing annotation resources are specifically designed for human variants. In the literature section, we currently provide curated information from recently published studies; nevertheless, we plan to implement updates covering all literature related to sSNVs. Going forward, we are committed to continuously updating and maintaining SynMall to enhance its utility as a robust platform for sSNV research.

Methods

Data source

SynMall first retrieves human protein-coding transcripts from BioMart (Kinsella et al. 2011) and generates all possible single-nucleotide variants (SNVs) in coding regions. We use the VEP (McLaren et al. 2016) tool to identify SNVs with synonymous consequences. These variants are further integrated and deduplicated from widely used variation databases, including FAVOR, CADD v1.7, and synVep, all of which catalog potential sSNVs across the human genome. For sSNVs in non-human species, we obtain and integrate genome-wide association study (GWAS) data for sSNVs from CropGS-Hub and the GWAS Atlas. Additionally, we collect vertebrate sSNVs from the EVA and Ensembl Variation. For primates, sSNVs and AF data are retrieved from the great ape (Prado-Martinez et al. 2013) and Han et al. (2019) data sets. Detailed processing steps are provided in Supplemental Table S4 and Supplemental Methods.

SynMall compiles a comprehensive set of annotations following ACMG guidelines. (1) disease-associated information—disease or phenotype associated with sSNVs, indicating deleterious effects; (2) AF—population AFs from multiple cohorts; (3) in silico predictions—functional impact and conservation scores from computational tools; and (4) de novo variants—observed in probands but absent in parents, indicating high pathogenic potential. We further integrate diverse information to comprehensively characterize sSNV biological functions. For splicing dysregulation, SynMall examines whether sSNVs overlap with splice site consensus regions and 2033 exonic splicing regulatory motifs (Supplemental Table S5; Supplemental Methods). SynMall also integrates ESEFinder 3.0 to identify exonic splicing enhancers specific to SR-rich proteins. For miRNA–mRNA interactions, we collect 2,425,678 experimentally validated miRNA–mRNA pairs from TarBase-v9.0 and assess whether sSNVs are located within these interaction regions. Regarding codon alteration, SynMall incorporates codon usage data under different contexts. SynMall annotates sSNVs located within known m^6A sites, as they can promote tumorigenesis by disrupting m^6A-dependent mRNA metabolism. For DNA-level regulatory elements, it collects enhancer- and promoter-associated data as well as curated regulatory annotations. SynMall also incorporates epigenetic features, including histone modifications, chromatin states, and transcription factor binding sites. We provide full documentation of all resources and tools used in SynMall, including their descriptions and references, in the Supplemental Documents and on the website at https://bioinfo.ahu.edu.cn/synMall/#/guide.

Design of SynScore

SynScore builds upon key annotations for ACMG variant interpretation, including functional scores and AF, together with biological features such as regulatory elements and epigenetic information. To avoid potential data leakage, functional scores derived from models trained on annotated coding SNVs are excluded. The model integrates both general and sex-specific AFs from gnomAD exomes and genomes, along with AFs from five primate species. A detailed list of feature descriptors is provided in Supplemental Table S6. During preprocessing, raw features are imputed and scaled using an iterative imputer and standard scaler implemented in scikit-learn v1.6.1 (Pedregosa et al. 2011). We curate a data set that contains 2362 sSNVs for training and 238 for testing (Supplemental Methods). LightGBM v4.6.0 (Ke et al. 2017) serves as the classification model. After parameter optimization, we randomly split the training data into five parts and perform fivefold cross-validation. After training, we select the models with the lowest validation loss from each fold and ensemble them by averaging their predictions. Finally, we calculate SynScores for all sSNVs in the human genome, representing their pathogenicity potential. Model performance is benchmarked against other VEPs using AUC and AUPR metrics (Supplemental Methods).

To assess whether higher SynScores are more enriched in genes under strong selection, we define high-score sSNVs for each gene using a threshold of 0.5 and compute the Z-score as

$$Z(SynScore) = \frac{N_o - N \cdot p_{high}}{\sqrt{N \cdot p_{high} \cdot (1 - p_{high})}},$$

where N_o is the observed number of high-score sSNVs, N is the total number of sSNVs, and p_{high} is the overall proportion of high-score sSNVs. We use a LoEUF score < 0.6 as a threshold to define genes under strong negative selection.

Implementation of feature extraction

For each sSNV, DNA sequences are retrieved based on the mutation site and context length through BEDTools (Quinlan and Hall 2010), whereas RNA sequences are extracted from the longest transcript associated with the mutation. Protein sequence outputs are derived from the amino acids encoded by the reference CDS. If the sSNV is near the start or end of the sequence, unavailable to retrieve full-length context, missing sequences are filled with the neighbor sequence. Subsequently, we compute the corresponding feature representation based on the extracted context sequences.

Computation of the analysis module

In the analysis module, SynMall primarily evaluates the effects of sSNVs based on the context sequences surrounding the mutation, such as whether an sSNV alters exonic splicing regulatory motifs depending on its position within the transcript. To assess mRNA structural changes, SynMall predicts the MFE of both the wild-type and mutant sequences and uses the MFE change to evaluate the impact of the sSNV on structural stability. For codon usage, SynMall calculates RSCU, CAI, and the gene-specific gtAI for each transcript, using the top 1% highest-expressing transcripts in GTEx (GTEx Consortium 2013) as the reference set. Regarding miRNA–mRNA interactions, SynMall accepts miRNA accessions from miRBase or user-defined miRNA sequences. It predicts potential interactions by comparing wild-type and mutant mRNA sequences with miRNA targets and assesses whether sSNVs disrupt interactions.

Update of literature resources

SynMall retrieves literature from PMC and PubMed daily using Biopython (Cock et al. 2009). We apply a large language model (Kimi Team: Bai et al. 2025) alongside manual inspection to present the essential results. To further extract key information, we review open-access publications, screening for pathogenic mechanisms and mutation information. To systematically assign evidence strength, we align the criteria with ACMG guidelines. Details on the query method and field descriptions are provided in the Supplemental Methods.

Data access

All data presented in SynMall are available at the link https://bioinfo.ahu.edu.cn/synMall. A downloadable version of SynMall is available from Figshare (https://doi.org/10.6084/m9.figshare.29281664). A manual containing all tools, resources, and interpretations of the database is provided in the Supplemental Documents. The code and data associated with this paper are available at GitHub (https://github.com/xialab-ahu/SynMall) and in the Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (U22A2038, 62402006, 62532017, and 62072003). We acknowledge the high-performance computing platform of Anhui University for providing computing resources and thank the members of our laboratory for their valuable contributions.

Author contributions: J.X. conceived the study and refined the manuscript. C.Y. collected and analyzed the data, developed the website, and drafted the manuscript. N.C. assisted in the literature review of the database. Y.S. contributed to the collection and application of computational tools. X.L. refined the manuscript and assisted in website design.

References

Bailey SF, Alonso Morales LA, Kassen R. 2021. Effects of synonymous mutations beyond codon bias: the evidence for adaptive synonymous substitutions from microbial evolution experiments. *Genome Biol Evol* **13:** evab141. doi:10.1093/gbe/evab141

Bi C, Shi Y, Xia J, Liang Z, Wu Z, Xu K, Cheng N. 2025. Ensemble learning-based predictor for driver synonymous mutation with sequence representation. *PLoS Comput Biol* **21:** e1012744. doi:10.1371/journal.pcbi.1012744

Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E, Pinto BL, Orr A, Paysan-Lafosse T, Ponamareva I, et al. 2025. InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res* **53:** D444–D456. doi:10.1093/nar/gkae1082

Byrne AB, Arts P, Ha TT, Kassahn KS, Pais LS, O'Donnell-Luria A, Broad Institute Center for Mendelian Genomics, Aguet F, Arachchi HM, Austin-Tse CA, et al. 2023. Genomic autopsy to identify underlying causes of pregnancy loss and perinatal death. *Nat Med* **29:** 180–189. doi:10.1038/s41591-022-02142-1

Cezard T, Cunningham F, Hunt SE, Koylass B, Kumar N, Saunders G, Shen A, Silva AF, Tsukanov K, Venkataraman S, et al. 2022. The European Variation Archive: a FAIR resource of genomic variation for all species. *Nucleic Acids Res* **50:** D1216–D1220. doi:10.1093/nar/gkab960

Chamary J-V, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet* **7:** 98–108. doi:10.1038/nrg1770

Chen J, Tan C, Zhu M, Zhang C, Wang Z, Ni X, Liu Y, Wei T, Wei X, Fang X, et al. 2024a. CropGS-Hub: a comprehensive database of genotype and phenotype resources for genomic prediction in major crops. *Nucleic Acids Res* **52:** D1519–D1529. doi:10.1093/nar/gkad1062

Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, Alföldi J, Watts NA, Vittal C, Gauthier LD, et al. 2024b. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625:** 92–100. doi:10.1038/s41586-023-06045-0

Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, Pritzel A, Wong LH, Zielinski M, Sargeant T, et al. 2023. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381:** eadg7492. doi:10.1126/science.adg7492

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25:** 1422–1423. doi:10.1093/bioinformatics/btp163

Daga S, Loberti L, Rollo G, Adamo L, Colavecchio OL, Brunelli G, Zguro K, Tripodi SA, Guarnieri A, Garosi G, et al. 2025. Slowly progressive autosomal dominant Alport syndrome due to COL4A3 splicing variant. *Eur J Hum Genet* **33:** 461–467. doi:10.1038/s41431-024-01706-8

Dai Y, Hu R, Liu A, Cho KS, Manuel AM, Li X, Dong X, Jia P, Zhao Z. 2022. WebCSEA: web-based cell-type-specific enrichment analysis of genes. *Nucleic Acids Res* **50:** W782–W790. doi:10.1093/nar/gkac392

Diederichs S, Bartsch L, Berkmann JC, Fröse K, Heitmann J, Hoppe C, Iggena D, Jazmati D, Karschnia P, Linsenmeier M, et al. 2016. The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. *EMBO Mol Med* **8:** 442–457. doi:10.15252/emmm.201506055

Gao H, Hamp T, Ede J, Schraiber JG, McRae J, Singer-Berk M, Yang Y, Dietrich ASD, Fiziev PP, Kuderna LFK, et al. 2023. The landscape of tolerated genetic variation in humans and primates. *Science* **380:** eabn8153. doi:10.1126/science.abn8197

GTEx Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45:** 580–585. doi:10.1038/ng.2653

Han S, Andrés AM, Marques-Bonet T, Kuhlwilm M. 2019. Genetic variation in pan species is shaped by demographic history and harbors lineage-specific functions. *Genome Biol Evol* **11:** 1178–1191. doi:10.1093/gbe/evz047

Hunt SE, McLaren W, Gil L, Thormann A, Schuilenburg H, Sheppard D, Parton A, Armean IM, Trevanion SJ, Flicek P, et al. 2018. Ensembl variation resources. *Database (Oxford)* **2018:** bay119. doi:10.1093/database/bay119

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. 2017. LightGBM: a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, Long Beach, CA, pp. 3149–3157. Curran Associates, Red Hook, NY. doi:10.5555/3294996.3295074

Kimi Team: Bai Y, Bao Y, Chen G, Chen J, Chen N, Chen R, Chen Y, Chen Y, Chen Y, Chen Z, et al. 2025. Kimi k2: open agentic intelligence. arXiv:2507.20534 [cs.LG]. doi:10.48550/arXiv.2507.20534

Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011:** bar030. doi:10.1093/database/bar030

Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, Hoffman D, Jang W, Kaur K, Liu C, et al. 2020. ClinVar: improvements to accessing data. *Nucleic Acids Res* **48:** D835–D844. doi:10.1093/nar/gkz972

Li C, Tian D, Tang B, Liu X, Teng X, Zhao W, Zhang Z, Song S. 2021. Genome Variation Map: a worldwide collection of genome variations across multiple species. *Nucleic Acids Res* **49:** D1186–D1191. doi:10.1093/nar/gkaa1005

Liang S, Pan M, Hu N, Wu Y-Y, Chen H, Zhu J-H, Guan H-J, Sang A-M. 2013. Association of angiotensin-converting enzyme gene 2350 G/A polymorphism with diabetic retinopathy in Chinese Han population. *Mol Biol Rep* **40:** 463–468. doi:10.1007/s11033-012-2081-2

Lin BC, Katneni U, Jankowska KI, Meyer D, Kimchi-Sarfaty C. 2023. In silico methods for predicting functional synonymous variants. *Genome Biol* **24:** 126. doi:10.1186/s13059-023-02966-1

Liu Y. 2020. A code within the genetic code: codon usage regulates co-translational protein folding. *Cell Commun Signal* **18:** 145. doi:10.1186/s12964-020-00642-6

Liu X, Tian D, Li C, Tang B, Wang Z, Zhang R, Pan Y, Wang Y, Zou D, Zhang Z, et al. 2023. GWAS atlas: an updated knowledgebase integrating more curated associations in plants and animals. *Nucleic Acids Res* **51:** D969–D976. doi:10.1093/nar/gkac924

Livingstone M, Folkman L, Yang Y, Zhang P, Mort M, Cooper DN, Liu Y, Stantic B, Zhou Y. 2017. Investigating DNA-, RNA-, and protein-based features as a means to discriminate pathogenic synonymous variants. *Hum Mutat* **38:** 1336–1347. doi:10.1002/humu.23283

McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: Uniform Manifold Approximation and Projection. *J Open Source Softw* **3:** 861. doi:10.21105/joss.00861

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl variant effect predictor. *Genome Biol* **17:** 122. doi:10.1186/s13059-016-0974-4

Miller DT, Lee K, Abul-Husn NS, Amendola LM, Brothers K, Chung WK, Gollob MH, Gordon AS, Harrison SM, Hershberger RE, et al. 2023. ACMG SF v3.2 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* **25:** 100866. doi:10.1016/j.gim.2023.100866

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. 2011. Scikit-learn: machine learning in Python. *J Mach Learn Res* **12:** 2825–2830. doi:10.5555/1953048.2078195

Pei G, Hu R, Jia P, Zhao Z. 2021. DeepFun: a deep learning sequence-based model to decipher non-coding variant effect in a tissue- and cell type-specific manner. *Nucleic Acids Res* **49:** W131–W139. doi:10.1093/nar/gkab429

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* **499:** 471–475. doi:10.1038/nature12228

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842. doi:10.1093/bioinformatics/btq033

Richards S. 2015. 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* **17:** 405–424. doi:10.1038/gim.2015.30

Sauna ZE, Kimchi-Sarfaty C. 2011. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* **12:** 683–691. doi:10.1038/nrg3051

Schubach M, Maass T, Nazaretyan L, Röner S, Kircher M. 2024. CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. *Nucleic Acids Res* **52:** D1143–D1154. doi:10.1093/nar/gkad989

Sharma Y, Miladi M, Dukare S, Boulay K, Caudron-Herger M, Groß M, Backofen R, Diederichs S. 2019. A pan-cancer analysis of synonymous mutations. *Nat Commun* **10:** 2569. doi:10.1038/s41467-019-10489-2

Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. 2022. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* **50:** W216–W221. doi:10.1093/nar/gkac194

Sun Y, Song X, Pan H, Li X, Sun L, Song L, Ma F, Hao J. 2024. Wiskott-Aldrich syndrome: a new synonym mutation in the WAS gene. *Intractable Rare Dis Res* **13:** 69–72. doi:10.5582/irdr.2023.01102

Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. 2014. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156:** 1324–1335. doi:10.1016/j.cell.2014.01.051

Tan Y, Ma C-Y, Li X-X, Han G-D, Dong Y-W. 2023. Genome-wide sequencing identifies a thermal-tolerance related synonymous mutation in the mussel, *Mytilisepta virgata*. *Commun Biol* **6:** 5. doi:10.1038/s42003-022-04407-4

Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, et al. 2019. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* **47:** D941–D947. doi:10.1093/nar/gky1015

Wang SY, Cheng YY, Liu SC, Xu YX, Gao Y, Wang CL, Wang ZG, Feng TQ, Lu GH, Song J, et al. 2021. A synonymous mutation in *IGF-1* impacts the transcription and translation process of gene expression. *Mol Ther Nucleic Acids* **26:** 1446–1465. doi:10.1016/j.omtn.2021.08.007

Wang Z, Zhao G, Zhu Z, Wang Y, Xiang X, Zhang S, Luo T, Zhou Q, Qiu J, Tang B, et al. 2024. VarCards2: an integrated genetic and clinical database for ACMG-AMP variant-interpretation guidelines in the human whole genome. *Nucleic Acids Res* **52:** D1478–D1489. doi:10.1093/nar/gkad1061

Wen P, Xiao P, Xia J. 2016. dbDSM: a manually curated database for deleterious synonymous mutations. *Bioinformatics* **32:** 1914–1916. doi:10.1093/bioinformatics/btw086

Zeng Z, Aptekmann AA, Bromberg Y. 2021. Decoding the effects of synonymous variants. *Nucleic Acids Res* **49:** 12673–12691. doi:10.1093/nar/gkab1159

Zhang X, Li M, Lin H, Rao X, Feng W, Yang Y, Mort M, Cooper DN, Wang Y, Wang Y, et al. 2017. regSNPs-splicing: a tool for prioritizing synonymous single-nucleotide substitution. *Hum Genet* **136:** 1279–1289. doi:10.1007/s00439-017-1783-x

Zhao Y, Xie J, Wang S, Xu W, Chen S, Song X, Lu M, El-Kassaby YA, Zhang D. 2021. Synonymous mutation in *Growth Regulating Factor 15* of miR396a target sites enhances photosynthetic efficiency and heat tolerance in poplar. *J Exp Bot* **72:** 4502–4519. doi:10.1093/jxb/erab120

Zhou H, Arapoglou T, Li X, Li Z, Zheng X, Moore J, Asok A, Kumar S, Blue EE, Buyske S, et al. 2023. FAVOR: functional annotation of variants online resource and annotator for variation across the human genome. *Nucleic Acids Res* **51:** D1300–D1311. doi:10.1093/nar/gkac966

Received July 29, 2025; accepted in revised form December 15, 2025.