*NAR Genomics and Bioinformatics*, 2026, **8**, lqaf211
https://doi.org/10.1093/nargab/lqaf211
**Standard Article**

OXFORD

# SyMetrics: an integrated machine learning model for evaluating the pathogenicity of synonymous variants in the human genome

**Linnaeus Bundalian [1,2], Martina Schmidt Strnadová[3], Felix Garten[1], Susanne Horn [3], Udo Stenzel[3], Denny Popp[1], Johannes R. Lemke[1], Saskia Biskup[4], Björn Schulte[4], Patrick May [5], Frank Bösebeck[6], Antje Garten[7], Doreen Thor[3], Angela Schulz[3], Julia Hentschel[1], Janet Kelso [8], Torsten Schöneberg[3,9], Diana Le Duc [1,2,8,10,*]**

[1]Institute of Human Genetics, University of Leipzig Medical Center, Leipzig, Saxony 04103, Germany
[2]Institute for Clinical Genetics, Technische Universität Dresden, and National Cancer Center (NCT) Dresden, Dresden, Saxony 01307, Germany
[3]Rudolf Schönheimer Institute of Biochemistry, Medical Faculty, University of Leipzig, Leipzig, Saxony 04103, Germany
[4]CeGaT GmbH, Tübingen, Baden-Württemberg 72076, Germany
[5]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette 4365, Luxembourg
[6]Agaplesion Diakonie Clinic Rottenburg, Rottenburg, Lower Saxony 27356, Germany
[7]Hospital for Children and Adolescents and Center for Pediatric Research (CPL), University of Leipzig, Leipzig, Saxony 04103, Germany
[8]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Saxony 04103, Germany
[9]School of Medicine, University of Global Health Equity, Kigali 6955, Rwanda
[10]Center for Diagnostics GmbH, Department of Genetics, Chemnitz Clinics, Chemnitz, Saxony 09116, Germany

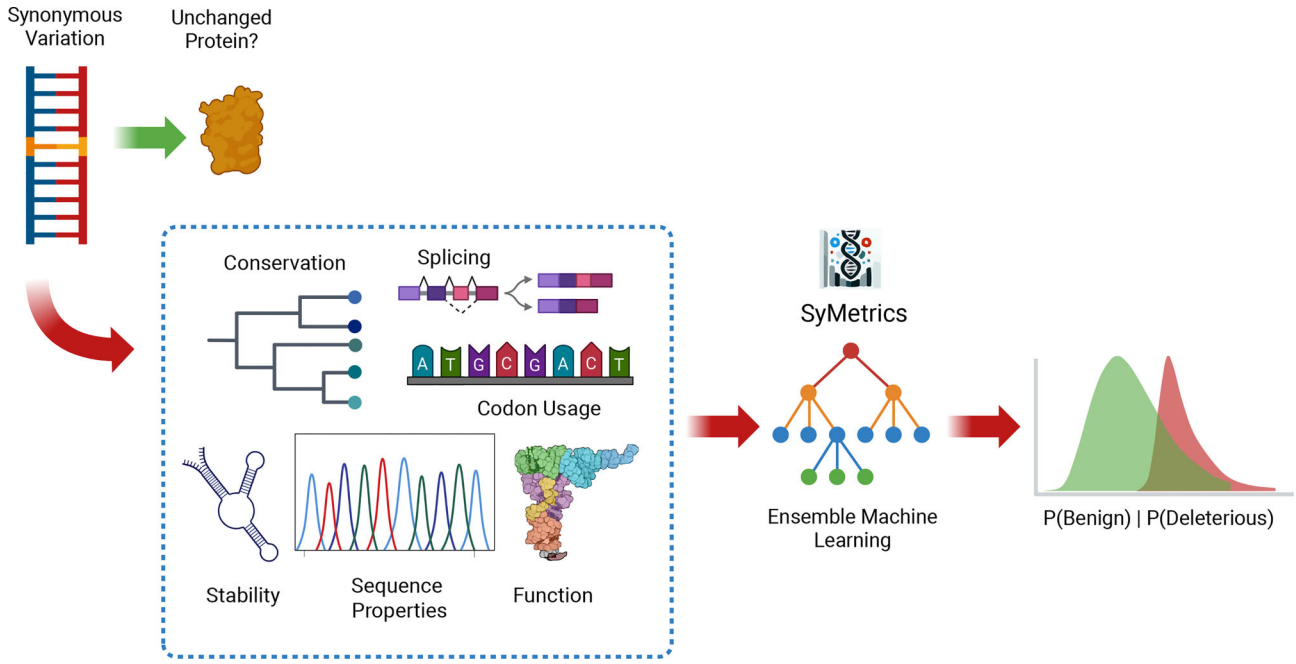
*To whom correspondence should be addressed. Email: diana.leduc@ukdd.de

## Abstract

Synonymous single nucleotide variants (sSNVs), traditionally seen as neutral, are now recognized for their biological impact. To assess their relevance, we developed SyMetrics, a framework that integrates predictors of splicing, RNA stability, evolutionary conservation, codon usage, synonymous variation effects, sequence properties, and allele frequency. We analyzed all possible sSNVs across the human genome, and our machine-learning model achieved 97% accuracy in distinguishing deleterious from benign variants, with a ROC–AUC of 0.89, outperforming individual predictors. Our estimates indicate that about 1.98 ± 0.17% of sSNVs absent from population databases are damaging (roughly 900 000 sSNVs), with an odds ratio of 3.87 for deleteriousness compared to common sSNVs ($P < 0.05$). To validate predictions, we performed functional assays on selected sSNVs in the *AVPR2* gene and additionally used available large scale mutagenesis screens of *RAD51C* and *BAP1* variants. In a clinical cohort, we identified 15 predicted deleterious sSNVs in genes linked to patient phenotypes; 9 were classified as (likely) pathogenic while 6 were variants of uncertain significance (VUS) per American College of Medical Genetics guidelines. For three VUS, segregation data supported their suspected inheritance patterns (*de novo*, X-linked). Our findings underscore the functional importance of sSNVs. To support further research and clinical applications, we provide a Python package and web application (https://symetrics.org/) for evaluating these variants comprehensively.

**Received:** April 26, 2025. **Revised:** November 3, 2025. **Accepted:** December 1, 2025

## Graphical abstract




## Introduction

Recent advancements in genomics and molecular diagnostics have significantly enhanced our understanding of genetic variation. Emerging sequencing technologies have provided invaluable resources for research, enabling deeper exploration of the human genome [1, 2] and greater insights into the mechanisms underlying various diseases and complex traits [3]. However, despite substantial progress in diagnostic yield [4], a significant diagnostic gap remains, particularly in rare diseases, with approximately 50% of cases still undiagnosed [5, 6].

Single nucleotide variations (SNVs) constitute a major component of human genomic variation and have drawn considerable attention from the medical and research communities. A coding SNV involves a single nucleotide change [7], which can be classified as either synonymous (sSNV) or nonsynonymous (nSNV). sSNVs do not alter the amino acid sequence during translation [1, 3, 8–13] and are often referred to as silent mutations [1, 10]. In contrast, nSNVs change the protein sequence and potentially its function, making them the primary focus of disease and trait-related studies [8, 14].

Traditionally, human genetics, mutagenesis screens, and evolutionary analyses have suggested that sSNVs are more likely to be neutral compared to nSNVs [15]. However, a growing body of research challenges this assumption, with studies indicating that sSNVs can have significant functional consequences [15, 16]. Recent findings suggest that purifying selection, a form of natural selection that removes harmful genetic variants, also acts on synonymous mutations [17], reinforcing the idea that they are not necessarily neutral. Even without altering amino acid sequences, sSNVs can impact gene expression and protein function through mechanisms such as modifying mRNA stability, altering splicing efficiency, and affecting translation kinetics [13].

Furthermore, the human genome contains alternative open reading frames (ORFs) embedded within canonical mRNAs [18, 19]. Although these alternative ORFs are less well-annotated and often exhibit lower expression levels, they add an additional layer of complexity to genetic interpretation. Notably, an sSNV in a canonical transcript may result in a nonsynonymous change when translated in an alternative reading frame, further highlighting the functional potential of sSNVs.

Given this complexity, annotating phenotype-affecting variants as "synonymous" may be misleading in clinical settings. To address this, it has been proposed that functionally consequential sSNVs be termed "unsense" mutations for more precise communication [20]. However, existing methods for quantifying variant effects are largely tailored to nSNVs [21, 22], with few tools specifically designed for assessing sSNVs. In response, several synonymous variant effect predictors have been developed, including Silent Variant Analyzer (SilVA) [23], Identification of Deleterious Synonymous Variants (IDSV) [10], and others [24]. These initial efforts have demonstrated the need for robust predictors and have driven further interest in developing and benchmarking bioinformatic tools for synonymous variant interpretation [24, 25].

In this study, we sought to understand how sSNVs disrupt biological processes or impair gene function. To develop a predictive framework, we integrated existing functional consequence scores (synVEP) [22], splicing effect predictors (MES, SpliceAI) [4, 26], RNA stability and folding metrics (SURF) [27], phylogenetic conservation scores (GERP++) [28], codon usage measures (RSCU and dRSCU), and sequence properties (e.g. CpG sites, CpG exons, and relative distance to end of mature mRNA and pre-mRNA) [29]. To classify sSNVs, we trained an ensemble machine learning model using a dataset of known benign and deleterious sSNVs [30], incorporating synonymous variant effect measures and allele frequency as features. Our model estimates that ~1.98 ± 0.17% (95% CI) of synonymous variants absent from population databases are deleterious. Exemplarily, functional validation demonstrated that predicted deleterious variants in the vasopressin receptor type 2 (*AVPR2*) gene impairs function. Finally, we re-

analyzed a large clinical cohort, identifying potentially deleterious sSNVs in genes associated with the probands' phenotypes. To facilitate research and clinical applications, we provide the results through the Python library (https://pypi.org/project/symetrics/) and web platform—SyMetrics (https://symetrics.org/), enabling researchers and clinicians to differentiate deleterious sSNVs from benign ones.

## Materials and methods

### OBSERVED and "NOT SEEN" synonymous variants

The sSNVs in the synVep or generated database were classified as follows:

- ***observed***: variants which can be found in the gnomAD v4.1.0 database with allele count >1 (AC > 1).
- ***not seen***: variants which cannot be found in the gnomAD v4.1.0 database.
- ***singleton***: variants which can be found in the gnomAD v4.1.0 database but with AC = 1.
- ***unobservable***: variants which are not likely to be observed in the future were discarded in the previous study as it was deemed to have little to no effect in the analysis.

We regrouped the variants by putting both ***singleton*** and ***observed*** together as **observed** and ***unobservable*** and ***not seen*** as **not seen** while using only the variants which passed the QC filter of latest gnomAD version (v4.1.0).

### sSNV Effect predictor, features, and scores

There are several existing tools and effect predictors designed to target sSNV [21, 22, 31]. We used SynVep, SpliceAI, SURF RNA stability, and SILVA to generate scores. For SILVA, we focused on the intermediate output of the model, as these are easier to explain and relate to biological implications and relevance. We used these tools to acquire scores reflecting the sSNV's metrics related to its consequences: splicing effects, RNA stability, phylogenetic and conservation scores, codon usage, and sequence properties, as described in Supplementary Table S1—Model metrics.

#### Functional effect—synVEP

One of the recently developed tools for assessing the functional impact of synonymous variants is the machine learning-based tool synVep. This tool assigns scores to sSNVs—synVEP scores—by using a sequential extreme gradient boosting model to differentiate pathogenic from benign variants. The score generated reflects the potential impact of the sSNVs. The implications of the scores are shown in the following notation:

$$s = \{0,\ x < 0.5\ 1,\ x \geq 0.5\} \quad (1)$$

Where 0 = no effect and 1 = effect [25].

#### Splicing effect—SpliceAI and MES

One of the most important aspects that essentially affects gene activities as well as protein diversity is splicing—the process involving the removal of introns to join exons together hence forming the mature messenger RNA—and disrupting this process can lead to aberrant production of messenger RNA (mRNA) which can manifest as disease or unexpected phenotypes [32, 33]. SpliceAI is a deep learning tool developed to predict the probability of whether a given position of pre-mRNA sequence is a splice acceptor or donor site in terms of Delta Scores (Δ Score) for acceptor loss, donor loss, acceptor gain, and donor gain [34]. High scores, at least >0.22 [35], are indicative that the position has a higher probability of being a splice acceptor or donor site. However, the threshold may vary depending on the context. In this study, we used the maximum delta scores (Max Δ Score or MAX_DS) among the four delta scores for a given genetic variant to simply indicate that the variant indeed can cause splicing. Additionally, we used scores generated from MaxEntScan (MES) to estimate the splice site motif or junction strength [26, 36]. The values represent the maximum efficiency of splicing at the sites. We used MES ≥ 3 as a threshold to indicate a relevant and potential influence of the genetic variation to splicing. The established threshold was based on the empirical observations on overall and combined score distribution of both observed and not-seen and is best to represent at least the top 2% of the population.

#### RNA stability—SURF

Summarized RNA Folding (SURF) is another metric under consideration that focuses on how genetic variants, especially sSNVs, affect RNA stability and folding. It calculates a broad range of RNA folding metrics for each SNV and evaluates the changes induced by this genetic variation, including, but not limited to, free energy, edge distance, and centroid distance. From these metrics, a unified score is derived to represent the impact of the variants under observation. For our analysis, we used a SURF threshold of ≥5 to indicate that the variant may have a potential influence on RNA stability [27].

#### Conservation and phylogenetic relevance—GERP++

The evolutionary constrained elements were estimated using GERP++ which is a valuable tool capable of quantifying the level of evolutionary constraint which can be indicative of their functional influence and selection pressure. This tool calculates site-specific rejected substitution scores (RS) where higher positive scores imply more deleterious genetic variants and/or stronger selection [28] similar to its predecessor GERP [37]. The threshold is set as GERP++ ≥ 4, which is considered to have a large deleterious effect associated with a high selection coefficient [38, 39, 40].

#### Codon usage

Two essential metrics used for codon usage analysis are RSCU (Relative Synonymous Codon Usage) and dRSCU, which represent the standardized differences in RSCU values. RSCU assumes that all synonymous codons are used equally, and based on this assumption, it compares the observed and expected frequencies of codons. A value >1 implies overrepresentation, while a value <1 indicates under-representation [41, 42]. Therefore, for this score, we established two criteria to observe the outcomes: one for when the threshold is RSCU > 1 and another for RSCU < 1. On the other hand, dRSCU helps assess the deviation of specific codons from common usage patterns in a dataset [43]. The range of observed values for dRSCU is from 0 to 2, and we decided to set the threshold at dRSCU > 1, as we aim to focus on values that are significantly different from 0, where there are minimal or no observed standard differences.

#### Sequence features

Although nucleotide sequences may appear complex and unintuitive at first glance, closer examination reveals recurring

patterns that are associated with molecular signatures and biological processes [44, 45]. We used CpG and relative distance to pre mRNA (f_premrna) and mature mRNA (f_mrna). CpG sites are DNA regions where cytosine (C) and guanine (G) and the overall occurrence of CpG contents provides a baseline reference with which actual observations can be compared. Significant deviations from the expected CpG might suggest functional relevance. The features f_mrna and f_premrna capture the relative position of the genetic variant in a mature and pre mRNA transcript [29].

## Gene-wise scoring function and constraints group comparison

With the listed effect measures, we aimed to identify signals that could be induced by different features related to these measures. This approach provides insight into the relevance of the features as well as the effectiveness of the selected scoring scheme. Specifically, we determined which genes had more sSNVs with high scores, i.e. where the score for a particular effect measure exceeded the set threshold, in relation to the proportions of the observed and not seen groups. To assess the difference in proportions between these groups for each given effect measure/scoring scheme, we used a $Z$ proportion test.

$$z = \frac{(\widehat{p}_1 - \widehat{p}_2)}{\sqrt{\hat{p}\ (1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{2}$$

Where:

- $\hat{p}$ —the pooled proportion of variants meeting the given criterion and score threshold for both NOT SEEN and OBSERVED.
- $\hat{p}_1$—the proportion of NOT SEEN variants meeting a given criterion and score threshold.
- $\hat{p}_2$—the proportion of OBSERVED variants meeting a given criterion and score threshold.
- $n_1$—the number of NOT SEEN variants.
- $n_2$—the number of OBSERVED variants
- $z$—the resulting statistic from the proportion test as a function of the selected/given variant effect measure (*synonymous Z-score*).

Using the proportion also allows to account for the length of the genes since with this approach we normalized the aggregated value by the number of observed values. A synonymous $Z$-score means that there are more not seen variants having a high effect measure score compared to the observed variants and lower score means otherwise.

The variant effect and gene-wise scores can aid in the deterministic interpretation of the results. The combined implications of the scores can be represented as follows:

$$V(x) = \begin{cases} 0, & s < t \\ 1, & s \geq t \end{cases} \ where\ 0 = no\ effect,\ 1 = effect \tag{3}$$

Where:

- $t$ is the set threshold for the selected effect measure
- $s$ is the effect measure score
- $V(x)$ is whether that score implies that the variant has an effect or none.

And,

$$G(x) = \begin{cases} 0, & s < t \\ 1, & s \geq t \end{cases} \ where\ 0 = tolerant,\ 1 = intolerant \tag{4}$$

Where:

- $t$ – is the set threshold for the selected effect measure
- $s$ – is the gene-wise score
- $G(x)$ – whether the score implies that the gene is tolerant or not.

## Functional implications and enrichment analysis

To further interpret the scores, we related them to existing gene constraints given in gnomAD—specifically, *pLI* and *Missense Z scores*. After scoring the genes, we grouped them into categories based on whether they had HIGH or LOW scores in the selected gene constraints. This categorization was done to determine if genes with high gene constraints tend to have higher gene-wise synonymous scores as described in the previous step. The group definition is as follows:

$$group = \begin{cases} LOW, & score < thr \\ HIGH, & score \geq thr \end{cases} \tag{5}$$

Where:

$$score \in \{\ pLI,\ Missense\ Z\ Score\} \tag{6}$$

To test in which functional pathway the genes identified to have a high gene-wise score play a role, we performed enrichment analysis and overrepresentation analysis across different gene sets in Gene Ontology (GO), KEGG, and Reactome using the R package clusterProfiler [46, 47], which uses hypergeometric test [48] to assess the probability of observing at least $k$ genes from the list of the pathway database:

$$P(X \geq k\ |n; N; K) = \sum_{i=k}^{n} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}} \tag{7}$$

## Variant classification model

### Dataset for known deleterious and benign sSNV

The known disease-causing dataset, representing the positive class, was retrieved and benchmarked from an existing study on synonymous variants [30]. The list consisted of variants from the Database of Deleterious Synonymous Mutations (dbDSM v1.2) [49] and the ClinVar database, with variants labeled as "likely pathogenic" and "pathogenic" considered as true positives. The negative class was represented by variants labeled as "likely benign" and "benign" in ClinVar [30, 50]. The pooled dataset included 4696 benign sSNVs and 367 deleterious sSNVs, all mapped to the GRCh37 reference genome. For model training and validation, the dataset was split into 70% for training, 15% for testing, and 15% for validation, while Synthetic Minority Oversampling (SMOTE) [51] was applied exclusively to the training dataset to address the class imbalance at 7:100 ratio which is a moderate class imbalance (https://developers.google.com/machine-learning/crash-course/overfitting/imbalanced-datasets?hl=en, accessed 19 August 2025). The application of SMOTE has proven effective for minority class-prevalence below 15% or 20%

**Table 1.** Performance metrics are used to evaluate the models

| Statistical measure | Formula | Explanation |
|---|---|---|
| Accuracy | $\frac{TP+TN}{TP+FP+TN+FN}$ | Ratio of correctly classified sets and the total number of elements in the set which is being classified [53], this estimates the probability of the true value of the class and the model's overall effectiveness. |
| Recall | $\frac{TP}{TP+FN}$ | Probability of the label to be the positive class |
| Precision | $\frac{TP}{TP+FP}$ | Estimates the model's predictive value of a positive/negative class [54] |
| F1 Score | $\frac{2\ x\ Precision\ x\ Recall}{Precision+recall}$ | The harmonic mean of precision and recall [53] |
| AUC | $\frac{sensitivity+specificity}{2}$ | Also known as also known as balanced accuracy [54] |

demonstrating a significant improvement in F1-score and recall which is important to our use case [52]. This method augments the minority class by randomly selecting an instance and finding its *k*-nearest neighbors, generating a new instance that is a convex combination of the selected instance and its nearest neighbors. This helps the model build larger decision regions around the minority class points without compromising data integrity [51]. We verified the integrity of the resampled data by checking the distance between clusters formed from the resampled data and their original class, confirming that the resampled clusters were near the original ones (Supplementary Fig. S4).

### Model selection and evaluation

For the ease of model selection, we used the Python library—Lazy Predict (https://lazypredict.readthedocs.io/en/latest/, accessed 19 August 2025). This library automates the training and evaluation of multiple classification models from the most traditional up to the advanced methods. The output does not give the models themselves but rather yields the metrics of the baseline models that can aid in the selection. Upon selection, it is still necessary to train it in the main library supporting the selected model from the Lazy Predict module. The following metrics and/or statistical measures (Table 1) were used to evaluate the performance of the models.

### Symetrics web and stand-alone package

We created a python package (https://pypi.org/project/symetrics) and a web platform (https://symetrics.org) to allow developers, clinicians, and researchers to access the records of the synonymous variants along with the selected variant effect measures as well as the predictions.

## Functional assay and analysis of selected synonymous human genetic variant materials

If not stated otherwise, materials used for cell culture were obtained from Sarstedt or ThermoFisher. Standard chemicals were purchased from Merck or Carl Roth GmbH. Primers were synthesized by Microsynth.

### Cloning and introduction of point mutation

The genomic sequence of human *AVPR2* consisting of three exons and two introns was amplified using human genomic DNA. Thereby, an N-terminal hemagglutinin (HA)-tag and a C-terminal FLAG-tag were introduced after the start codon and in front of the stop codon, respectively. Both tags have been shown to not interfere with receptor expression and function [55, 56]. Furthermore, the amplification primers:

- forward primer: 5′-TGTACCCCTACGACGTCCCCG ACTACGCCCTCATGGCGTCCACCACTTCCG-3′
- reverse primer:

5′-ATCATGTCTGGATCCACTAGTCACTTAT CGTCA TCGTCCTTATAATCCGATGAAGTGTC CTTGGCC-3′ contained restriction sites for *Aat*II and *Spe*I to allow for cloning into the pcDps-derived plasmid pL [57]. To insert the point mutations, quick change mutagenesis was performed. Thus, two complement primers were designed carrying the desired point mutation in the middle of the primer sequence (Supplementary Table S1—Primers used to insert point mutations). PCR was performed using Phusion high fidelity polymerase (ThermoFisher) according to the manufacturer's protocol using *AVPR2* pL as template plasmid at an annealing temperature of 55°C. Prior to transformation into chemical competent cells, template plasmid was digested using DpnI (NEB). Sanger sequencing was performed by Microsynth to ensure the correct sequences of the mutants derived.

### Cell culture

HEK293T cells were cultured in DMEM media (ThermoFisher) supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin, and 100 µg/ml streptomycin at 37°C in a humidified atmosphere containing 5% $CO_2$. Cells were regularly split twice a week. For the different analysis, cells were counted using a Neubauer chamber and seeded into poly-L-lysine (Sigma–Aldrich)-coated well plates with the following cell numbers: 96-well plates—20 000 cells in 200 µl media, 48-well plates—40 000 cells per well in 500 µl media, six-well plates—300 000 cells in 2 ml media.

### Determination of total receptor expression

HEK293T cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) media as described above. Six-well plates were transfected with plasmid DNA (125 ng) using Lipofectamine2000 (ThermoFisher) according to the manufacturer's protocol. A sandwich ELISA taking advantage of the HA- and Flag-tags were used to estimate the total amounts of receptor proteins as previously described [58]. In brief, after lysis of transfected HEK293T cells (10 mM Tris–HCl, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, 1% sodium deoxycholate, and 0.2 mM NP-40) the total expression of receptors was analyzed with an anti-FLAG-M2-antibody (Sigma–Aldrich) coated to microtiter plates (10 µg/ml anti-Flag-antibody in 0.15 M sodium tetraborate/HCl, pH 8). After washing with PBS-T (0.05% Triton X-100 in PBS), unspecific binding was blocked with DMEM + 10% FBS for 1 h at 37°C followed by incubation of the solubilized cells for 1 h. Detection was achieved using an anti-HA peroxidase-conjugated antibody (Roche) diluted 1:1000 in blocking solution and *o*-phenylenediamine (10 mg in 0.1 M citric acid, 0.1 M $Na_2HPO_4$, and 0.2% $H_2O_2$). Using 1 M HCl the reaction was stopped, and OD readings were recorded at 492 and 620 nm (Tecan Sunrise).

### Determination of cell surface expression

Cell surface expression of *AVPR2* mutants was determined in HEK293T cells using a direct ELISA [59] (Supplementary Fig. S5). In this assay, the anti-HA-POD-conjugated antibody binds to the introduced N-terminal HA-tag of receptors

present at the cell surface. Thereto, 24 h after seeding into 48-well plates, cells were transfected with 25 ng of different plasmid DNA using Lipofectamine2000 (ThermoFisher) according to the manufacturer's protocol. Forty-eight hours post transfection, cells were fixed with formaldehyde for 20 min at room temperature before blocking (DMEM + 10% FBS) for 1 h at 37°C. Later, cells were incubated with an anti-HA peroxidase-conjugated antibody (Roche) diluted 1:1 000 in blocking solution. Detection of receptor expression on the cell surface was performed using *o*-phenylenediamine solved in substrate buffer (0.1 M citric acid and 0.1 M $Na_2HPO_4$) containing 0.2% $H_2O_2$. The reaction was stopped after 3 min by adding 1 M HCl. OD values at 492 nm were measured using the Sunrise microplate reader (Tecan) and normalized by subtracting the background OD at 620 nm. Expression is given after subtracting the OD values of mock-transfected cells in percentage compared to wt *AVPR2* pL.

#### Determination of receptor activity

The AVPR2 is coupled to the $G_s$ protein and its activation results in an increase in intracellular cAMP which can be detected using the AlphaScreen™ cAMP functional assay (PerkinElmer). Thus, 24 h after seeding into 96-well plates, cells were transfected with 12.5 ng of plasmid DNA per well. Forty-eight hours after the transfection, cells were washed with phenol red-free DMEM containing 1 mM IBMX (Sigma–Aldrich), an inhibitor of phosphodiesterases. Cells were then stimulated for 30 min 1 μM AVP (Sigma–Aldrich) at 37°C in media containing IBMX. To stop the stimulation, plates were placed onto ice, media was removed, and 20 μl of lysis buffer (20 mM Tween 20, 1 mM IBMX, 5 mM HEPES, and 0.1% BSA diluted in $ddH_2O$, pH = 7.4) was added to each well. Samples were stored at −20°C until measurement. Determination of cAMP was performed according to the manufacturer's specifications using EnVision 2105 Multimode Plate Reader (PerkinElmer).

#### Statistical analysis

Statistical and graphical analyses were performed using Python. All experiments were performed as technical triplicate in four separate batches. Each tested variant is presented as % of the corresponding wild type value in the batch. Statistical significance was determined using the Wilcoxon–Mann–Whitney test. *P*-values below 0.05 were considered to be significant.

### Functional and experimental validation of SyMetrics with large scale saturation mutagenesis screens of *RAD51C* and *BAP1* variants

To further validate SyMetrics with experimentally tested sSNVs, we leveraged recently published large scale saturation mutagenesis screens performed on human tumor suppressor genes i.e. *RAD51C* and *BAP1* via Saturation Genome Editing (SGE). SGE is a high-throughput multiplexed assay of variant effect (MAVE) which enables direct measurement of the impact of thousands of genetic variants on cellular fitness [60, 61].

#### Methodological approach for MAVE

A Cas8/sgRNA complex and a variant repair library were introduced to human near-haploid (HAP1) cells. Given that *RAD51C* and *BAP1* are genes of known functional relevance, the presence of deleterious alleles is expected to deplete the variant carrying cells over time. This depletion was captured on different time points for both studies—*RAD51C* (Days 4, 7, and 14) [61] and *BAP1* (Days 4, 7, 10, 14, and 21) [62]. The functional scores were calculated from the log2-fold changes (LFCs) of the variant abundance and were normalized with respect to their apparent growth rates per unit time. Scores are classified into deleterious (score < 0), enriched (score > 0), or unchanged. Those variants classified as deleterious (score < 0) are variants which reduce cell viability and proliferation [61, 62].

#### BAP1 and RAD51C high-resolution functional mapping and assessment via SGE

This experiment involved the use of 9188 unique *RAD51C* variants across its 9 coding sequence exons which include over 99.5% of all possible coding sequence single nucleotide alterations. The functional classification resulted in >99.9% accuracy/concordance with clinical data [61]. For *BAP1*, 18 108 unique variants were characterized of which 6196 were found to have abnormal functions. The SGE functional scores for *BAP1* variants have shown >99% sensitivity and >98% specificity in variant classification. It has demonstrated a concordance of 90% between independent SGE libraries indicating the robustness of the approach [62].

### RNA analysis of a sSNV in *ARHGEF9* identified in a patient cohort

RNA was isolated from cultivated fibroblasts and subsequently cDNA was synthesized. Parts of *ARHGEF9* were amplified by long-range PCR and sequenced using Oxford Nanopore Technologies (ONT) using the 96 native barcoding library prep kit (SQK-NBD114.96) according to the manufacturer's instructions. Pool of barcoded libraries was loaded onto a MinION flow cell type R10.4.1 and sequenced on a MinION device Mk1b (ONT). Raw data was base-called, and demultiplexed using dorado (ONT) and subsequently aligned against the Human reference genome build hg38 using minimap2 in splice-aware mode. Aligned reads were analyzed using the Integrative Genomics Viewer (IGV). Presence of the variant on DNA level was confirmed by extracting DNA from the same cultivated fibroblasts and subsequent Sanger sequencing of parts of the *ARGHEF9* gene.

## Results

### Genes with known functional relevance are more likely to exhibit intolerance to synonymous variation

We initially hypothesized that if synonymous variation affects gene function, genes intolerant to loss-of-function or missense variation would be more likely to show depletion of sSNVs compared to neutrally evolving genes.

To test this, we used a pooled Z proportion test (synonymous Z-score) to calculate the difference in the proportion of observed and not observed sSNVs (i.e, those absent in the general population) within each gene. A high proportion of not observed variants may indicate negative selection due to functional impairment.

For this analysis, we used all human protein-coding transcripts from Ensembl Biomart [63], filtering out those with unknown nucleotides, missing start/stop codons, or patched

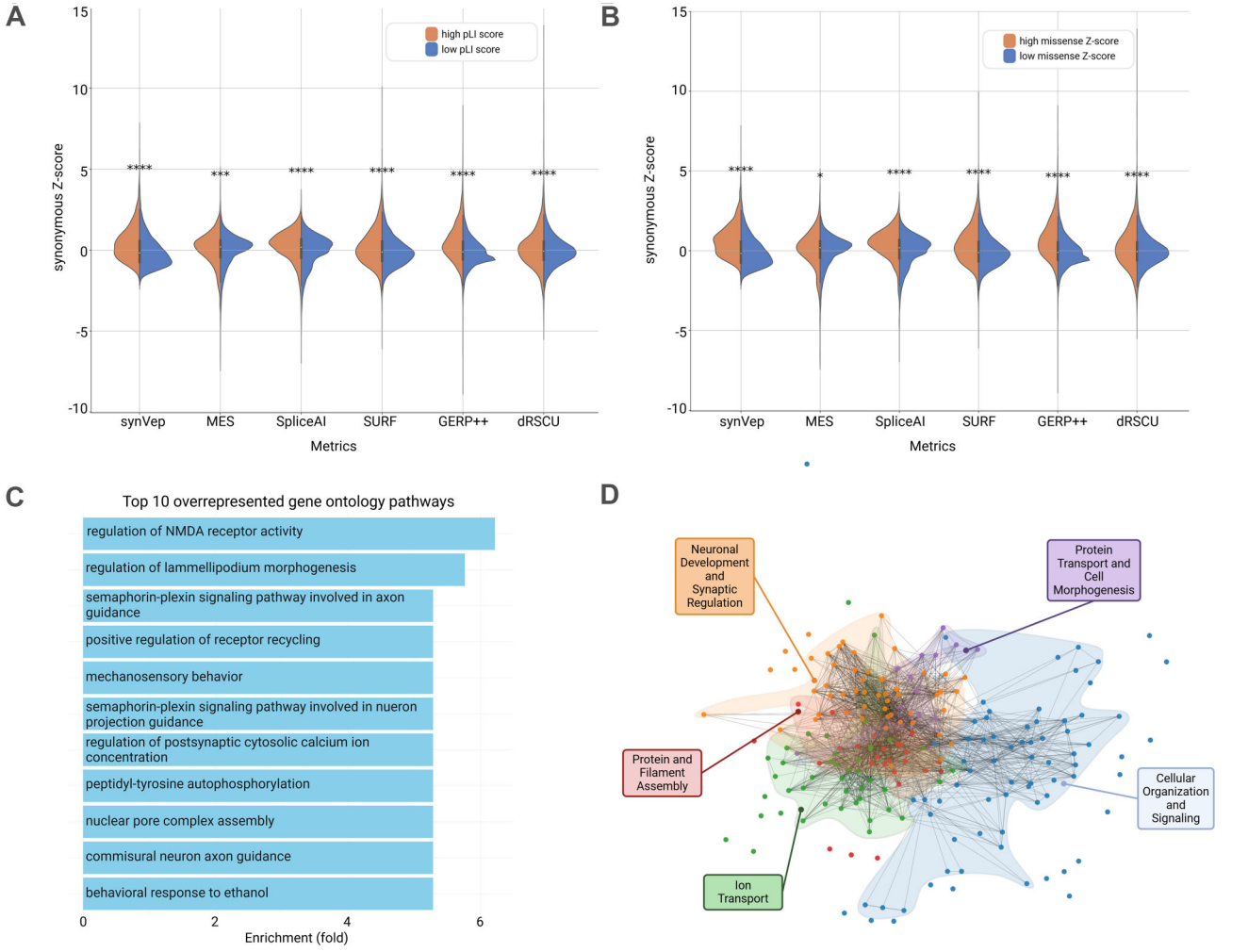



**Figure 1.** Comparison of *Z*-scores for synonymous variants across previously reported metrics in genes categorized by (**A**) high and low pLI (probability of loss-of-function intolerance) scores or by (**B**) high and low missense *Z*-scores (intolerance to missense variation). The significantly (Mann–Whitney test) higher synonymous *Z*-scores were observed among the metrics derived from synVep, MES, SpliceAI, SURF, GERP++, and dRSCU, suggesting higher scores for genes that are intolerant to functional variation. (**C**) Top 10 overrepresented GO categories for genes with high synonymous *Z*-scores ($\geq$ 1.96) in at least one metric. Genes with high synonymous *Z*-scores show enrichment among pathways related to brain and nervous system pathways and cellular activities. (**D**) Clustering of GO pathways based on gene community detection. The clustering of pathways is based on the co-occurrence of genes with high synonymous *Z*-scores across different GO categories. The analysis reveals distinct clusters of pathways that are enriched in genes (nodes in the network) frequently found together (edges that connect nodes), which identified five key biological themes. $*$ *P*-value $< 0.05$, $***$ *P*-value $\leq 0.001$, $****$ *P*-value $\leq 0.0001$.

chromosome IDs. We annotated sSNVs using synVEP v1 [22] with an updated reference from gnomAD v4.1 [64], classifying variants as either "observed" or "not seen."

We then calculated *Z*-scores (from proportion of "not seen" to "observed" variants for a gene) for various previously reported metrics related to sSNVs: synVEP (functional effect of synonymous variants), MES (splice site strength), SpliceAI (splice site prediction), GERP++ (evolutionary conservation), RSCU and dRSCU (codon usage), CpG and CpG exon content, and distances to pre-mRNA (f_premRNA) and mRNA (f_mRNA). Only sSNVs passing synVEP quality checks were included.

Genes depleted from a specific class of variation, based on gnomAD data, where high pLI ($\geq$ 0.8) indicates intolerance to loss-of-function and high missense *Z*-score ($\geq$ 3) suggests intolerance to missense variation [64], were considered functionally constrained. We compared synonymous *Z*-scores between genes with high and low pLI, as well as between genes with high and low missense *Z*-scores.

Our analysis revealed that genes under stronger selection (high pLI or missense *Z*-score) exhibited significantly higher synonymous *Z*-scores (Fig. 1A and B) for several metrics: synVEP, MES, SpliceAI, SURF (RNA stability), GERP++, and dRSCU. However, RSCU, CpG, CpG exon, and distances to pre-mRNA and mRNA did not show significantly higher synonymous *Z*-scores in functionally constrained genes (Supplementary Fig. S1).

This suggests that the high proportion of not seen sSNVs in these genes may reflect their deleteriousness, leading to negative selection. Among 3295 genes with high pLI, 674 had a high synonymous *Z*-score ($\geq$ 1.96) across all significant metrics. Similarly, 394 of the 1589 genes with a high missense *Z*-score showed a high synonymous *Z*-score across all significant metrics. In total, we identified 1865 genes with a high

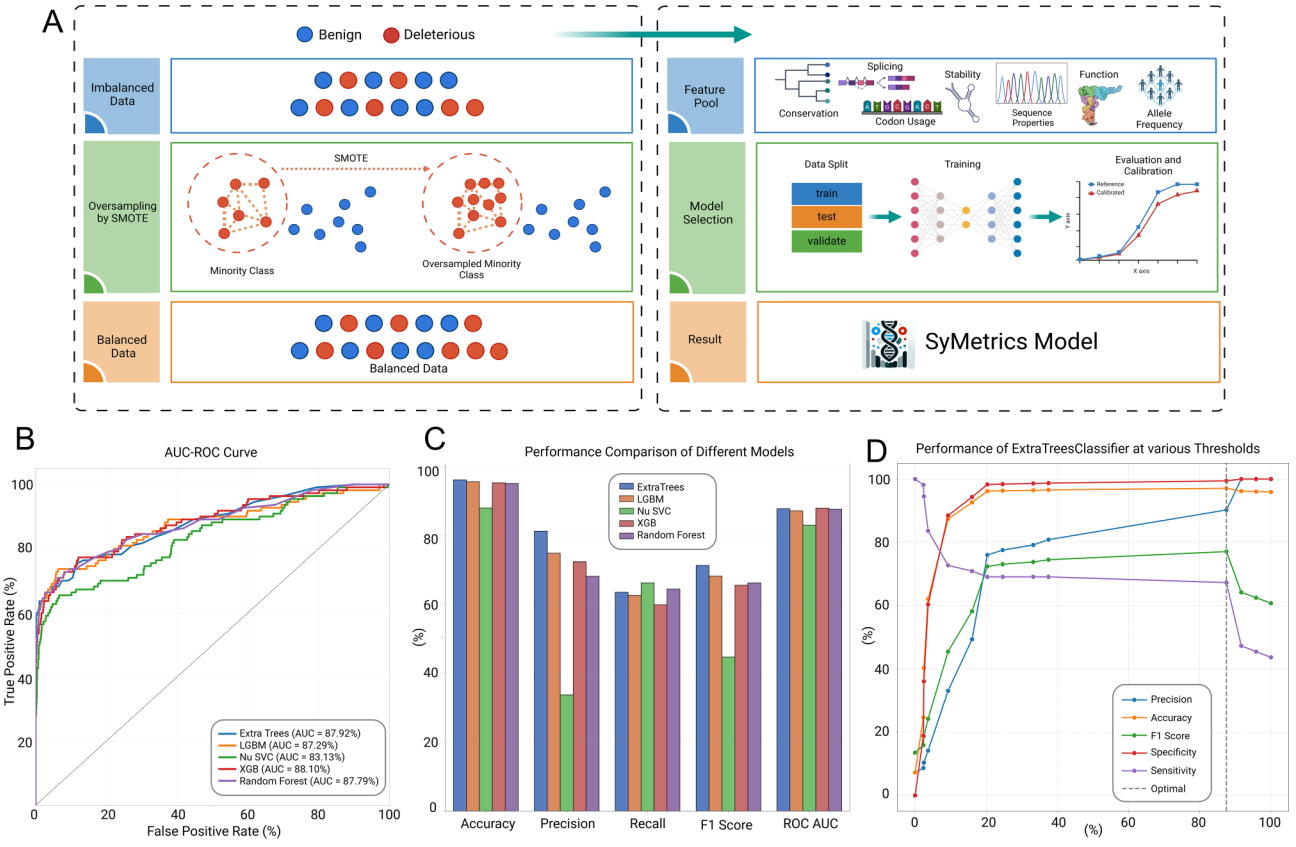



**Figure 2.** (**A**) Overview of model construction. This flowchart outlines the process of model construction, encompassing data preprocessing, training, evaluation, and calibration. During data preprocessing, an oversampling strategy was applied using SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalances. The balanced dataset was then annotated with scores corresponding to selected features and subsequently divided into training, testing, and validation subsets. Before finalizing the model, we calibrated it to optimize and balance the performance metrics. (**B**) Area Under the Curve (AUC) for Receiver Operating Characteristic (ROC) to evaluate model performance evaluation of top 5 models identified by Lazy Predict. All models show good discrimination between classes (AUC > 0.8). (**C**) Comparison of top 5 models identified by Lazy Predict. The dataset consisted of 457 synonymous variants (benign: 234, deleterious: 223). The different models were used to classify the variants. The model based on *ExtraTreesClassifier* (blue) outperforms other models in most metrics (metrics definition available in Table 1). (**D**) Threshold adjustment affects the performance of *ExtraTreesClassifier*. Using the calibrated model, we determined the best threshold that can yield a good balance of all the metrics. We used the prediction probabilities and adjusted the threshold from 0 to 1 and recalculated the metrics based on each threshold. The best threshold (marked with gray vertical line) shows the point where all the metrics have high values and hence represents the optimum. (Created in BioRender. Garten, A. (2025) https://BioRender.com/ ove998a)

synonymous *Z*-score (≥1.96) in at least one metric, regardless of their intolerance to loss-of-function or missense variation (Supplementary Table S1).

To explore whether genes intolerant to synonymous variation are enriched in specific biological processes, we conducted a gene ontology (GO) enrichment analysis. The top 10 over-represented pathways were predominantly related to synaptic and neuronal development (Fig. 1C).

For further insight, we applied the Leiden Algorithm [65], a complex network analysis method for community detection. This approach groups genes based on their GO network connections, revealing functional relationships. We configured the analysis with a maximum of 10 iterations and a fixed random seed for reproducibility, resulting in five clusters. Most genes in these clusters were associated with nervous system function and cellular processes (Fig. 1D).

Taken together, the depletion of synonymous variation in functionally constrained genes suggests that sSNVs may have functional effects and are subject to negative selection, as recently proposed by Gudkov *et al.* [17].

## ExtraTreesClassifier best discriminates between "benign" and "deleterious" classes in a public dataset

While deterministic approaches provide explainability, they are insufficient for reliably determining the deleteriousness of an sSNV. To achieve a more generalized and conclusive classification of sSNVs as either "benign" or "deleterious," we developed a supervised ensemble machine learning model. This model integrates 11 effect measures related to splicing, RNA stability, distance to premature and mature mRNA, codon usage, and CpG content, along with allele frequency in the general population, to predict the probability of deleteriousness (workflow in Fig. 2A). Our approach leverages existing public databases of known deleterious and benign genetic variants [30], including the Database of Deleterious Synonymous Mutations (dbDSM v1.2) [49] and ClinVar database [30, 50].

We incorporated 12 features—the 11 effect measures plus allele frequency—and used Lazy Predict (https://lazypredict.readthedocs.io/en/latest/, accessed 19 August 2025) in Python to estimate the performance of different models before building them. The best-performing models in terms of accu-

**Table 2.** Model Comparison of Lazy Predict Top 5 Models

| Model | Accuracy | Precision | Recall | F1 Score | ROC AUC |
|---|---|---|---|---|---|
| Extra Trees | 0.963 | 0.814 | 0.636 | 0.714 | 0.879 |
| LGBM | 0.958 | 0.750 | 0.627 | 0.683 | 0.873 |
| Nu SVC | 0.882 | 0.338 | 0.664 | 0.448 | 0.831 |
| XGB | 0.955 | 0.725 | 0.600 | 0.657 | 0.881 |
| Random Forest | 0.953 | 0.683 | 0.645 | 0.664 | 0.878 |

The selection of the best model is based on their performance on accuracy, precision, recall, F1 Score, and ROC–AUC. Note that emphasis on the balance of precision and recall should be considered along with accuracy. (Note: ExtraTrees—Extremely Randomized Trees, LGBM—Light Gradient Boosting Machine, Nu SVC—Nu Support Vector Classification, XGB—Extreme Gradient Boosting).

racy, balanced accuracy, F1-score, and ROC–AUC were predominantly tree-based ensemble machine learning models (Supplementary Table S1—Ranked models in Lazy Predict; Supplementary Fig. S2).

We then trained and evaluated the top five models predicted to perform best. Consistent with the Lazy Predict evaluation, the Extremely Randomized Trees (*ExtraTreesClassifier*) emerged as the top-performing model (Fig. 2C and Table 2). ExtraTrees is an ensemble learning technique that trains multiple decision trees and aggregates their outputs to classify variants. Unlike standard decision trees, ExtraTrees introduces additional randomness by selecting features for splitting at random, reducing bias while maintaining diversity among trees [66].

To enhance model reliability, we calibrated the trained ExtraTrees model using hyperparameters identified via Lazy Predict. We assessed model accuracy using stratified 10-fold cross-validation, yielding a mean accuracy of 96.3% and a ROC–AUC of 0.89, meaning the model correctly distinguishes between deleterious and benign variants 89% of the time.

The final calibrated model, trained on the entire dataset, achieved 97% accuracy, a ROC–AUC of 0.87, and higher precision compared to the uncalibrated version (Supplementary Fig. S3). While a default classification threshold of 0.5 is commonly used, we evaluated performance across multiple thresholds to identify an optimal balance between accuracy, precision, and recall. The most balanced performance was observed at $t = 0.875$, with: accuracy = 97.10%, precision = 90.23%, recall/sensitivity = 67.27%, specificity = 99.43%, and F1 Score = 0.77 (Fig. 2D). The full reference table for threshold adjustments is available in Supplementary Table S1—Threshold Adjustment. The model output score represents the probability of a variant being deleterious, termed SyMetrics Probability.

Our analysis demonstrates that *ExtraTreesClassifier* is the most effective model for distinguishing between benign and deleterious sSNVs, achieving high accuracy and robustness through ensemble learning and calibrated threshold optimization.

## SyMetrics demonstrates a strong performance in predicting the deleteriousness of sSNVs compared to individual variant effect measures

We further evaluated the performance of SyMetrics by comparing it against individual variant effect measures—synVEP, SpliceAI, SURF, MES, and GERP++—to assess its predictive power relative to the individual effect measures incorporated in the model. These metrics were selected because they directly relate to deleteriousness and selection. Additionally, we compared SyMetrics to CADD [67], an established and reliable score for detecting deleterious variants.

For this analysis, only variants found in dbDSM v1.2 [30, 49] were retained to represent the positive class. dbDSM v1.2 is a manually curated database of deleterious sSNVs compiled from published literature and resources such as ClinVar, GRASP, GWAS Catalog, GWASdb, PolymiRTS, PubMed, and Web of Knowledge. The negative class (neutral variants) was extracted from VariSNP [21, 68], which contains benign variants.

The variants were annotated based on their determined classes per variant effect measures using a simple deterministic scheme presented as:

Where:

- $t$ is the set threshold for the selected effect measure.
- $s$ is the effect measure score.
- $V(x)$ is whether that score implies that the variant is deleterious or benign.

The thresholds selected per variant effect measures are set as follows: synVEP ($t = 0.5$), SpliceAI ($t = 0.5$), SURF ($t = 4$), MES ($t = 3$), GERP++ ($t = 4$), and CADD ($t = 20$).

The varying sensitivity and specificity among individual effect measures indicate that no single mechanism (e.g. splicing via SpliceAI or RNA stability via SURF) can fully capture the deleteriousness of sSNVs. However, combining multiple effect scores enhances predictive accuracy, as reflected in the SyMetrics model performance metrics: accuracy (92.69%), F1-Score (0.92), and ROC–AUC (0.93). Among the individual effect measures, SyMetrics outperformed all others, correctly identifying all 234 negative variants and 188 of the 223 positive variants (Fig. 3).

In sum, SyMetrics outperforms individual variant effect measures in predicting the deleteriousness of sSNVs, achieving higher accuracy and reliability by integrating multiple predictive features.

## Predicted deleterious sSNVs are depleted in general population and enriched in genes under strong selection

To determine whether predicted deleterious sSNVs are depleted in the general population, we scored the generated sSNVs and categorized them by maximum allele frequency (AF) in all gnomAD v4.1 populations into five groups: very common, common, low frequency, rare, and not seen. We then performed a pairwise comparison of the number of sSNVs with a SyMetrics score exceeding the threshold ($t = 0.875$) across the following groups: common ($0.05 < AF < 0.1$) versus very common ($0.1 < AF < 1$), common versus low-frequency variants ($0.01 < AF < 0.05$), common versus rare ($0 < AF < 0.01$), and common versus not seen ($AF = 0$).

We observed significant differences in score distribution for all pairwise comparisons except for the very common group. This suggests that variants with higher SyMetrics scores are depleted in the general population due to selection.

To further investigate these findings, we calculated the odds ratio of a variant being deleterious in "very common, low-frequency, rare, and not seen" groups relative to the "common" group using Fisher's Exact Test. The predicted deleterious sSNVs were found to be significantly enriched among low-frequency, rare, and not seen groups (Fig. 4A).

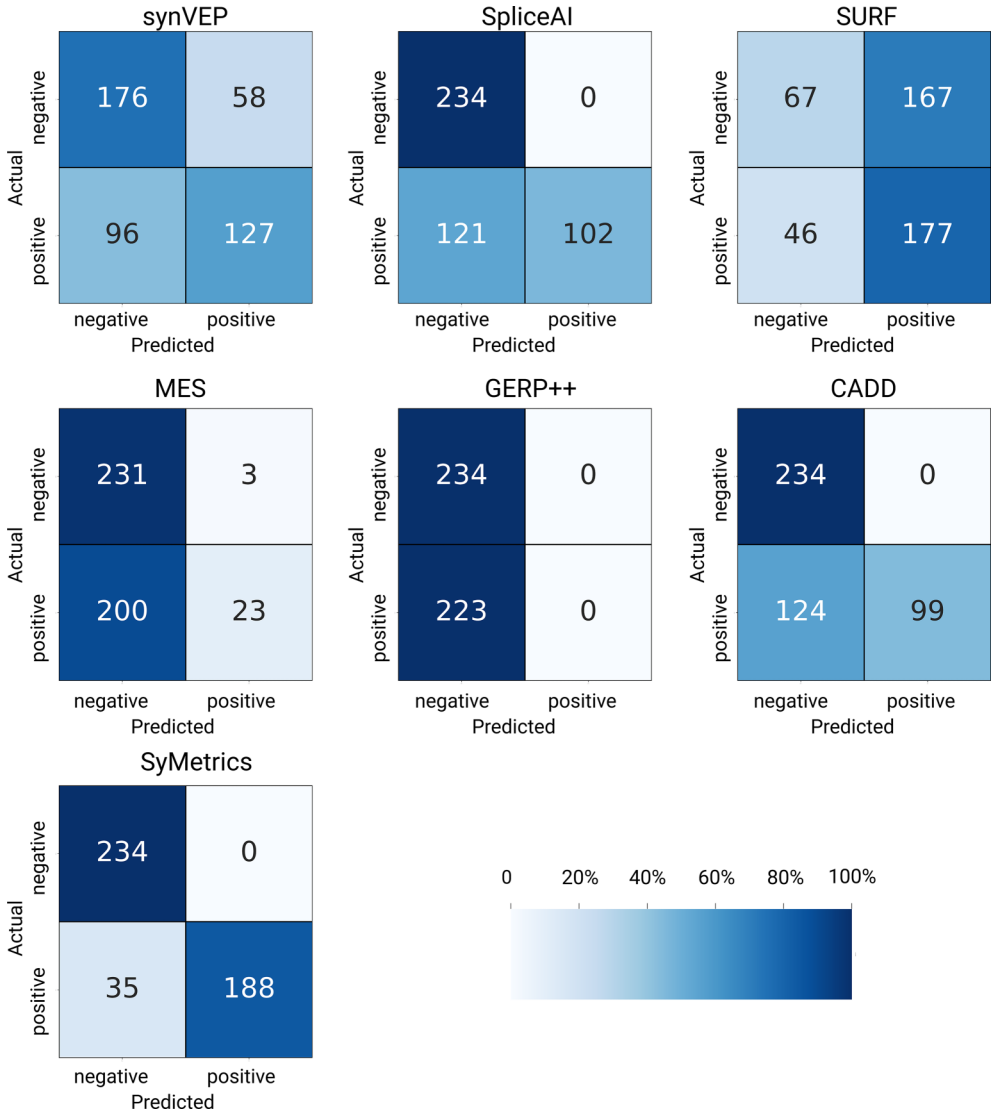



**Figure 3.** Heatmap of confusion matrices for variant effect predictions based on different metrics. The test dataset consisted of 457 synonymous variants (benign: 234, deleterious: 223). We calculated the number of true positives, false positives, true negatives, and false negatives, depicted across actual and predicted, for each metric based on the recommended thresholds. Compared to all other metrics, SyMetrics performed best in classifying the actual/true class among others.

Using stratified bootstrapping, we estimated the proportion of potentially deleterious synonymous variants: $1.98 \pm 0.15\%$ in the not seen category, $1.60 \pm 0.15\%$ in singletons, $1.07 \pm 0.13\%$ in rare variants, $0.53 \pm 0.1\%$ in common variants. This corresponds to $n = 931\,841$ (95% CI: 861 247–1 002 435) deleterious variants in the not seen group.

Next, we compared synonymous *Z*-scores using SyMetrics probability under the same grouping scheme, further stratifying genes by their pLI (probability of loss-of-function intolerance) and missense *Z*-score. As expected, genes under stronger selection (high pLI $\geq 0.8$ or missense *Z*-score $\geq 3$) exhibited significantly higher synonymous *Z*-scores (Fig. 4B), reinforcing the idea that sSNVs in these genes are absent in the general population due to their deleteriousness.

Altogether, we identified 309 genes with a high synonymous SyMetrics *Z*-score (>1.96), independent of the grouping scheme based on pLI or missense *Z*-score (Supplementary Table S1—High SyMetrics Z). In sum, predicted deleterious sSNVs are significantly depleted in the general population and enriched in genes under strong selection, highlighting their potential functional impact.

## Experimental validation of sSNVs affecting *AVPR2* function

To determine the extent to which our predictions can be validated, we conducted functional testing of eight sSNVs in the *AVPR2* gene. *AVPR2* encodes the arginine vasopressin receptor type 2, and pathogenic variants in this gene cause X-linked nephrogenic diabetes insipidus type 1. To mitigate potential bias toward functionally critical genes, we selected an X chromosome-located gene for which no constraint metrics are currently available in gnomAD v4.1. Additionally, we chose *AVPR2* because it does not exhibit an overall synonymous *Z*-score indicative of intolerance to synonymous variation (*Z*-score: 1.006).

None of the predicted deleterious variants was observed in the general population (gnomAD v4.1). The variants are dis-

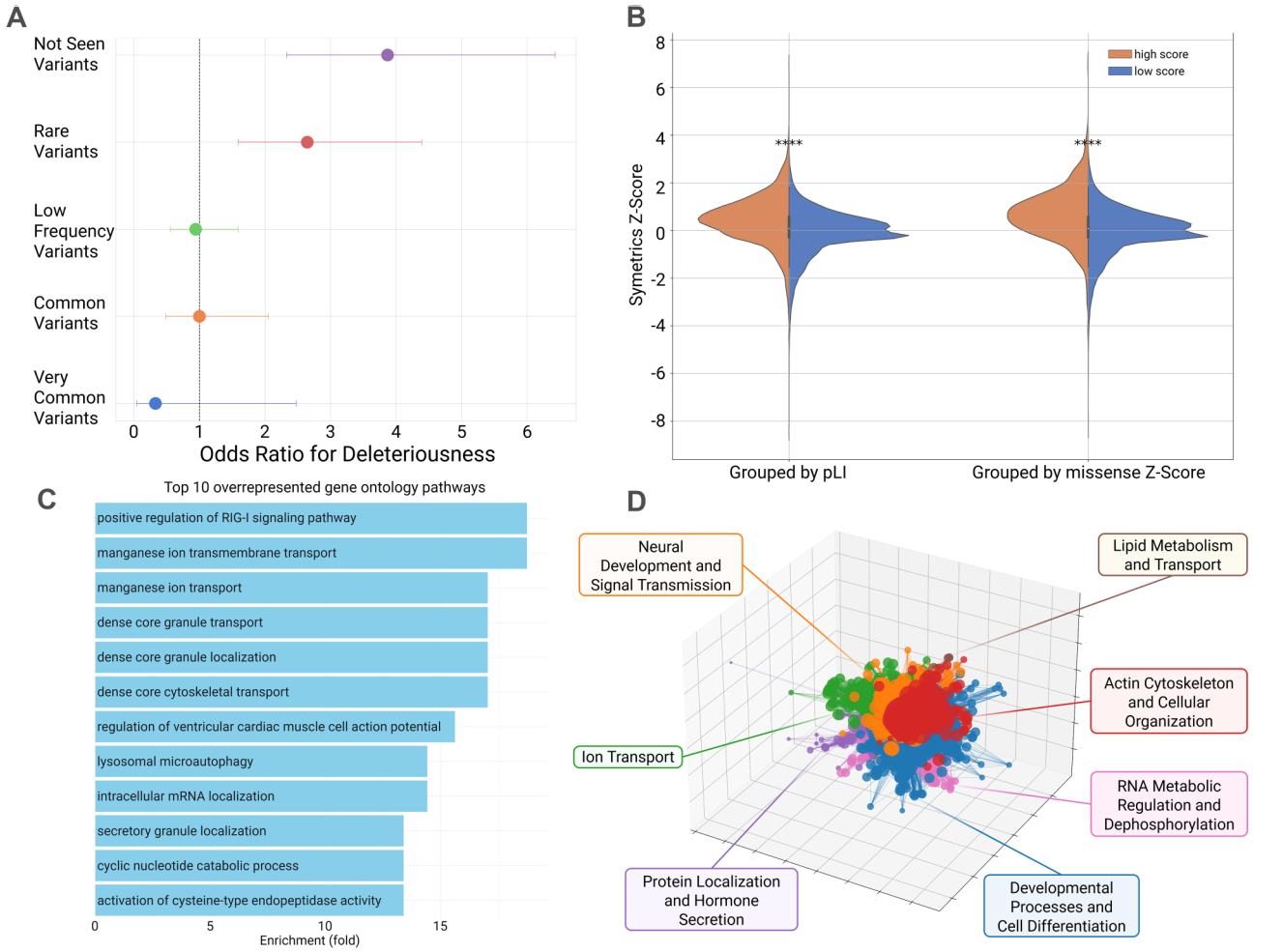



**Figure 4.** (**A**) Odds ratios for deleteriousness based on SyMetrics prediction across synonymous variants with different allele frequencies in the general population. Not seen variants = allele frequency is AF = 0 in gnomAD v4.1, rare variants = allele frequency is 0 < AF < 0.01 in gnomAD v4.1, low frequency variants = allele frequency is 0.01 < AF < 0.05 in gnomAD v4.1, common variants = allele frequency is AF > 0.05 in gnomAD v4.1, very common variants = allele frequency is 0.1 < AF < 1 in gnomAD v4.1. (**B**) Comparison of synonymous SyMetrics *Z*-scores for genes with high versus low pLI and missense *Z*-scores. Genes with high pLI and missense *Z*-score have significantly higher synonymous SyMetrics *Z*-scores **** *P*-value ≤ 0.0001, Mann–Whitney test. (**C**) Top 10 overrepresented gene ontology categories enriched in genes with high SyMetricss *Z*-score (≥ 1.96). Using the Mann–Whitney test, we tested whether the significant synonymous *Z*-scores of genes with high missense z score is higher than the one with low missense *Z*-scores. The significantly higher synonymous SyMetrics *Z*-scores were derived from the SyMetrics probability. (**D**) Clustering of GO pathways based on gene community detection. The clustering of pathways is based on the co-occurrence of genes with high synonymous SyMetrics *Z*-scores across different GO categories. The analysis reveals distinct clusters of pathways that are enriched in genes (nodes in the network) frequently found together (edges that connect nodes), which identified seven key biological themes, among which developmental processes and cell differentiation is the best represented category.

tributed across exons 2 and 3. Additionally, we included two predicted benign variants in our analysis.

We evaluated the impact of these variants on receptor function by measuring total cellular and cell surface expression (indicators of protein production, folding, and/or trafficking defects) and cAMP formation (agonist-induced receptor activation) in comparison with the wild-type receptor. We used minigene constructs containing all three exons, interrupted by their natural introns, cloned into a eukaryotic expression vector.

Among the investigated variants, three had a SpliceAI score > 0.5 (Supplementary Table S1--- *AVPR2* Variants). Interestingly, the variant NM_000054.7:c.27T > A, p.(Ala9=), with a SpliceAI score of 0.99 and SyMetrics 0.875, showed no functional effect (Fig. 5A–C). However, variant NM_000054.7:c.276A > G, p.(Gln92=) exhibited the lowest expression and the strongest functional effect among all tested variants (Fig. 5). While the SpliceAI score of this variant is 0.5, the SURF score is 27.54, suggesting high mRNA instability.

To assess whether a similar level of protein reduction could be observed for another splicing variant, we introduced a variant in the canonical splice site NM_000054.7:c.910 + 2T > C, p.?. This mutation is expected to result in the loss of an acceptor site, intron retention of 20 base pairs, and a premature stop codon. While the reduction in protein expression was significant, it was less pronounced than that observed for variant NM_000054.7:c.276A > G, p.(Gln92=).

For NM_000054.7: c.909A > G, p.(Gln303=), we also identified reduced cell surface expression, although to a lesser extent than in NM_000054.7:c.276A > G, p.(Gln92=). Splicing prediction indicates a donor site loss (SpliceAI = 0.62),

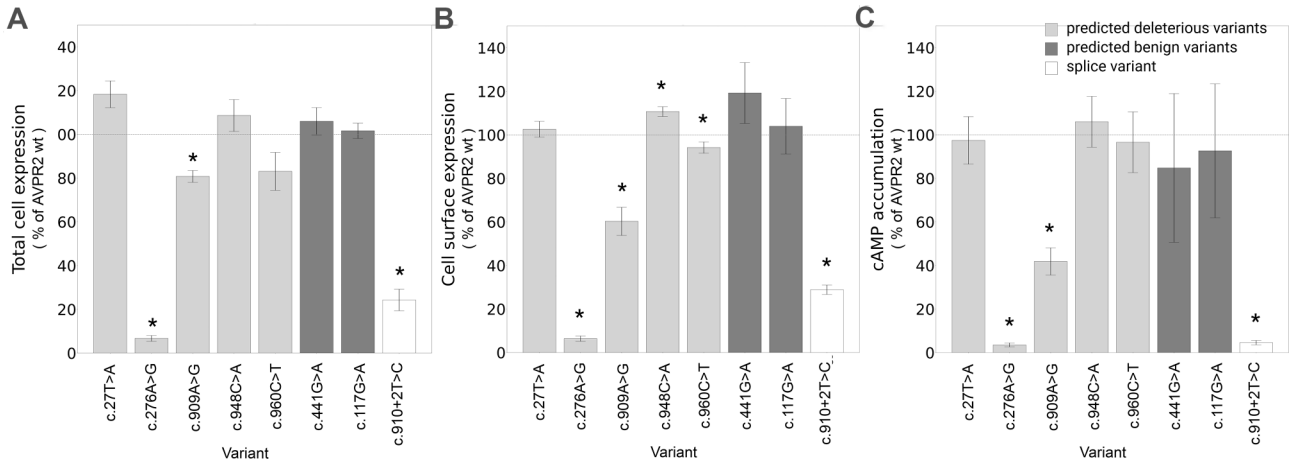



**Figure 5.** Synonymous variants affect the functionality of the human vasopressin 2 receptor (encoded by *AVPR2* gene). We predicted sSNVs that do not alter the amino acid residue but may affect receptor function—either predicted to be deleterious (light gray) or benign (dark gray). These variants were inserted into the genomic sequence of the human *AVPR2* and functionally tested. (**A**) First, total receptor protein expression was determined in an ELISA taking advantage of the N-terminal HA-tag and the C-terminal Flag-tag. (**B**) Then, cell surface expression of the receptor protein was determined. (**C**) Finally, cAMP accumulation was detected after receptor stimulation. For all variants, experiments show a consistent trend, albeit only for variants 2, 3, and 8, all experiments showed a significant result. ∗ *P*-value (Wilcoxon Mann–Whitney test) < 0.05. All experiments are presented as a percentage of the corresponding matched wild-type receptor. The dashed horizontal line represents the wild-type level (100%)

leading to the same transcript as the canonical splice site variant. However, NM_000054.7: c.909A > G, p.(Gln303=) exhibited significantly higher expression than the canonical splice site variant.

Lastly, we examined two predicted deleterious variants: NM_000054.7:c.948C > A, p.(Leu316=) and NM_000054.7:c.960C > T, p.(Thr320=), with SpliceAI scores of 0.23 and 0.42, respectively, suggesting a low probability of having a splicing effect. However, both had SURF scores exceeding the significance threshold of 5, suggesting potential effects on RNA stability. Interestingly, variant NM_000054.7:c.948C > A, p.(Leu316=) exhibited significantly higher cell surface expression (Fig. 5B), while variant NM_000054.7:c.960C > T, p.(Thr320=) showed significantly lower cell surface expression. However, total cell expression for these two variants did not reach significance, although the trend followed the same direction (Fig. 5). A similar pattern was observed in cAMP accumulation, but due to high variability, it did not reach statistical significance. We also tested two variants predicted to be benign - variants NM_000054.7:c.441G > A, p.(Ala147=) and NM_000054.7:c.117G > A, p.(Ala39=), both of which showed no significant difference from the wild-type.

In conclusion, among the five predicted deleterious variants, we observed significant alterations in cell surface expression in four cases, including one with higher expression than the wild-type. Considering RNA stability may improve predictions of a variant's deleteriousness, particularly in cases of potential gain-of-function effects, such as variant NM_000054.7:c.948C > A, p.(Leu316=).

## SyMetrics recapitulates experimentally determined effects of sSNVs in saturation genome editing of *BAP1* and *RAD51C*

To further validate the model's capability to predict the deleteriousness of sSNVs at scale, we utilized the variants filtered from the large-scale mutagenesis screening of *RAD51C* and *BAP1*. The sSNVs were filtered to only include the deleterious ones as indicated by their functional score in SGE (score < 0). For *BAP1*, we found 30 unique depleted sSNVs which are concordant to independent SGE libraries. SyMetrics predicted 25 of them to be deleterious. On the other hand, we found 15 *RAD51C* variants that were classified as depleted by SGE. Notably, 14 of these variants were predicted to be deleterious by SyMetrics demonstrating a high concordance and highlighting the model's predictive capability for functionally relevant sSNVs.

Altogether, the results of the validation against the *BAP1* and *RAD51C* SGE studies demonstrate effectiveness of the SyMetrics in predicting the functional relevance of sSNVs. The high concordance of the results highlights the feasibility of using SyMetrics to support and advance sSNV interpretation in clinical and research settings.

## SyMetrics predicts deleterious potential of sSNVs identified in a patient cohort

At the Human Genetics Institute at Leipzig University Clinics, we reanalyzed data from 11 903 patients who had undergone genetic evaluation either through exome/trio-exome sequencing or using a morbid gene panel (TruSight One, Illumina). We identified 767 774 unique variants, which were scored using SyMetrics. Of these, 8088 sSNVs were predicted to be deleterious.

To refine our analysis, we applied additional filters, requiring that the variant be absent from the general population and that the Human Phenotype Ontology (HPO) terms associated with the affected gene match those assigned to the patient, according to the HPOsim score [69].

We identified 15 variants in genes that matched the patient's phenotype and were predicted to be deleterious. Among these, nine variants were classified and reported as either likely pathogenic ($n = 4$) or pathogenic ($n = 5$). Additionally, six variants were reported as variants of uncertain significance (VUS). Parental segregation analysis was available for only three of these VUS cases.

In one case, the variant c.3015G > A, p.(Val1005=) in *IQSEC2*—a gene encoding a guanine nucleotide exchange factor for the ARF family of GTP-binding proteins—occurred *de novo*. Of all the effect measures considered, only synVEP was above the deleteriousness threshold. Notably, the variant had a SpliceAI score of only 0.47, meaning it could have been missed in routine clinical screening. Unfortunately, RNA analysis was not available for this proband.

In another case, a variant of uncertain significance c.1077G > A, p.(Lys359=) in *ARHGEF9*—a gene encoding collybistin, the guanine nucleotide exchange factor 9—was hemizygous in the affected proband and inherited from a carrier mother. Since developmental and epileptic encephalopathy 8 follows X-linked inheritance, segregation analysis alone could not clarify its pathogenicity. However, we recommended further familial segregation analysis, as a maternal uncle exhibited a similar phenotype. In this case, the SpliceAI score was only 0.18, meaning the variant would likely have been missed by routine clinical screening. However, both synVEP and SURF were above the threshold, suggesting potential RNA instability. RNA analysis based on cultivated fibroblasts from the heterozygous mother showed only transcripts from the wildtype allele confirming the suggested RNA instability. Hence, the variant should be denoted as c.1077G > A, r.0, p.(0) and was classified as likely pathogenic.

Similarly, we identified a patient who was a homozygous carrier of the variant c.729G > A, p.(Thr243=) in *SLC39A4*, a gene associated with acrodermatitis enteropathica. Both parents were carriers, making segregation analysis uninformative. However, the variant had a SpliceAI score of 0.76, strongly suggesting a splicing effect, for which we recommended RNA analysis.

In conclusion, applying SyMetrics in a clinical setting can be particularly beneficial for detecting deleterious variants that may not affect splicing and where segregation analysis can provide additional insights.

## Discussion

Recent developments in sequencing technologies have greatly improved our understanding of how genetic variation is associated with phenotypes or diseases. The role of sSNVs in the context of human disease has garnered increasing attention in recent years, challenging the traditional view that these variants are functionally neutral [15]. sSNVs, defined as single-nucleotide changes that do not alter the amino acid sequence of proteins, can nonetheless have significant biological implications. Evidence suggests that these variants can affect gene expression by interfering with mRNA splicing, stability, and translation efficiency, potentially contributing to disease susceptibility and phenotypic variation.

The scarcity of reliable genetic datasets on synonymous variants is evident in the comparatively lower level of research and investigation they receive, especially when compared to missense or nonsense variants. However, the association of synonymous variants with various rare diseases, as well as with cancer, where 6%–8% of pathogenic single nucleotide substitutions are synonymous variants, underscores the importance of studying these variants more comprehensively. These variants frequently act as driver mutations in human cancers.

The evaluation of sSNVs has been significantly advanced by the development of *in silico* scores that assess their potential impact on splicing, RNA stability, sequence conservation, translation efficiency, and other functional mechanisms. To explore whether genes known to be intolerant to loss-of-function or missense variation, which are often associated with disease, are also sensitive to synonymous changes, we compared the distribution of various functional scores between genes under strong selective constraints and those without such constraints. Specifically, we analyzed metrics assessing overall functional impact (synVEP), splicing effects (MES, spliceAI), RNA stability and folding (SURF), evolutionary conservation (GERP++), codon usage (RSCU, dRSCU), and sequence properties like CpG, CpG exons, and proximity to mRNA and pre-mRNA boundaries. Our analysis revealed a significantly higher distribution of five of these scores, related to splicing and RNA stability, for intolerant genes (Fig. 1). This is in line with accumulating evidence indicating that synonymous changes can modulate the stability of mRNA and its translation kinetics, ultimately affecting protein synthesis. Similarly, Gaither and colleagues showed that sSNVs that may disrupt mRNA structure have significantly lower rates in human populations [27]. Strong evidence that sSNVs can alter protein function and expression is provided by functional studies, which, although limited to particular genes, provide proof-of-concept. To this end, synonymous variants have been shown to influence blood group expression [70] or even cause hemophilia (c.459G > A (Val=) in the *F9* gene) by slowing factor IX translation and affecting its conformation, which results in decreased extracellular protein levels [71].

To assess which genes appeared to be most intolerant to synonymous variation, we performed a gene ontology analysis, which revealed clustering of these genes in neuronal development and structural cellular processes (Fig. 1C and D). This further strengthens the idea that sSNVs in these genes could have significant functional consequences, and as these are fundamental processes under strong evolutionary constraint, any disruption could have detrimental effects on the organism.

While the above-mentioned individual scores are valuable for identifying variants that disrupt specific processes, we developed a comprehensive tool that predicts the overall deleteriousness of an sSNV by integrating multiple metrics. The allele frequency in the general population of the variants appears to be an important predictor of deleteriousness, as suggested by Gaither *et al.* and Zeng *et al.* [22, 27]. This is supported by our analysis, which shows a depletion of sSNVs in genes under constraint. Thus, in our model, we also integrated allele frequency in the general population (gnomAD v4.1) [64] as one of the parameters. We identified *ExtraTreesClassifier* to be the best-performing machine learning model, which, after calibration and setting the threshold to $t = 0.875$, reached an accuracy of 97.10%, a precision of 90.23%, and specificity of 99.43% (Fig. 2 and Table 1). Our model, which integrates multiple aggregated metrics, demonstrated improved accuracy in identifying both benign sSNVs and those with potential pathogenic consequences (Fig. 3), outperforming individual metrics as well as the widely established CADD score [67]. Interestingly, evolutionary conservation represented by GERP++ discerned all the benign variants but missed all known pathogenic sSNVs, suggesting that evolutionary conservation is a poor metric for synonymous variants. Conversely, SURF, which predicts RNA stability, managed to detect many of the true positives (177/223) but showed low specificity in discriminating against the true negatives (67/234) (Fig. 3). SpliceAI was able to properly de-

tect all true negatives but showed lower performance for the true positives (102/223). This may suggest that many of the sSNVs may have a functional consequence other than splicing. Overall, SyMetrics performed best, properly identifying all true negatives and 188/223 (ca. 84%) of the true positives (Fig. 3).

We were then prompted to determine which genes are depleted of synonymous variation based on the SyMetrics predictions. We initially showed that variants not present in the general population have an odds ratio of approximately 4 to be deleterious (Fig. 4A). Furthermore, we reproduced the synonymous constraint, as determined by the SyMetrics *Z*-score for the functionally relevant genes (Fig. 4B), and finally, we identified 309 genes with a high SyMetrics *Z*-score (Supplementary Table S1—High SyMetrics Z). Gene ontology clustering revealed seven key biological themes, among which developmental processes and cell differentiation were the best represented, followed by neuronal development and cytoskeletal and cellular organization. The depletion of such pathways from synonymous variation further underscores the functional relevance of sSNVs.

To confirm the functional importance of sSNVs, we tested predicted deleterious and benign variants in the *AVPR2* gene. The variant with the strongest effect was variant NM_000054.7: c.276A > G, p.(Gln92=) (Fig. 5). This variant had been previously reported in patients with diabetes insipidus. It was shown that it introduces an alternative acceptor site leading to a 5′ truncated exon 2 lacking its first 251 bases [72]. The truncation causes a frameshift followed by a premature stop codon, which may trigger nonsense-mediated decay. However, compared to a variant located in the canonical splice site, which also leads to a premature stop codon, the reduction in protein expression for variant NM_000054.7: c.276A > G, p.(Gln92=) was significantly lower. This may suggest that for variant NM_000054.7: c.276A > G, p.(Gln92=), RNA instability may additionally contribute to the almost absent expression of the protein, given the SURF score of 27.54. Moreover, for variant NM_000054.7: c.909A > G, p.(Gln303=), the splicing prediction is of a donor loss (SpliceAI = 0.62), which would lead to the same transcript as in the case of a variant in the canonical splice site. However, in the case of variant NM_000054.7: c.909A > G, p.(Gln303=), we observe much higher expression than in the case of the canonical splice site variant. While in our functional setup we cannot test whether splicing is affected, the higher expression level of this variant makes it very likely that splicing is not affected in this case. In turn, the reduced expression could be a result of altered RNA stability. Two of the tested variants did not have high SpliceAI scores. For these variants, we showed significantly altered cell surface expression (Fig. 5B). Variant NM_000054.7:c.948C > A, p.(Leu316=) showed higher expression, and the SURF score in this case was 5.93. Increased mRNA stability after synonymous variation had been demonstrated for the synonymous variant in the prothrombin gene, *F2* (NM_000506.4: c.1824C > T; p.Arg608=, SURF = 15.88). The variant causes increased prothrombin mRNA and plasma protein levels, such that carriers of the variant are at increased risk of thromboembolism [73]. Thus, considering RNA stability may improve the prediction for deleteriousness, especially in cases where the mechanism may be gain-of-function. Although the effect of variants NM_000054.7:c.948C > A, p.(Leu316=) and NM_000054.7:c.960C > T, p.(Thr320=) seems more limited, the consistent trend across all tests—total expression, cell surface expression, and cAMP accumulation (Fig. 5)—strongly suggests a functional impact.

Lastly, to check whether SyMetrics could be useful in a clinical setup, we revisited 11 903 cases presenting the suspicion of a genetic disease. We identified 15 sSNVs that were predicted to be deleterious, and the affected gene matched the proband's phenotype. Nine of these variants were classified according to American College of Medical Genetics (ACMG) criteria [74] as pathogenic or likely pathogenic. For the rest of 6 VUS, we had parental information only in three cases. One of the variants in c.3015G > A, p.(Val1005=) in *IQSEC2* occurred *de novo*, making it very likely to be causative for the reported symptoms. We recommended RNA analysis, which could shed more light on the functional mechanism, however the analysis was currently not possible, as no patient sample was available.

Furthermore, we identified a hemizygous variant c.1077G > A, p.(Lys359=) in *ARHGEF9* in a proband with epileptic encephalopathy and ataxia. The variant is inherited from a carrier mother, and there is a similarly affected maternal uncle, for whom no genetic testing is available. The phenotype is consistent with OMIM #300 607 (ARHGEF9-related disorder), and the pattern of inheritance in the family supports pathogenicity. Although the variant was initially classified as a variant of uncertain significance (VUS) according to ACMG criteria, our tool predicted pathogenicity through RNA instability. Subsequent RNA analysis indeed demonstrated RNA instability consistent with a loss-of-function effect. Based on these findings, the variant could be reclassified as likely pathogenic, confirming its role as the cause of disease in this family.

In conclusion, this study underscores the significant impact of sSNVs on essential biological processes, highlighting their potential role in disease mechanisms. The observed link between genetic variants and functional outcomes, especially in early-onset conditions like epileptic encephalopathy and related disorders, emphasizes the need to consider these variants in clinical research and diagnostics. Overlooking the functional relevance of sSNVs may lead to missed disease etiologies, ultimately limiting our understanding of the genetic foundations of complex disorders. To aid in this effort, we present a tool that predicts functionally relevant synonymous variations. This tool is freely available online and provides researchers with insights into the underlying mechanisms of a variant, displaying all relevant metrics and allele frequencies in the general population. As the availability of genetic data increases, we will likely better understand the "silence" of synonymous variants.

## Acknowledgements

We thank the German Research Foundation and University of Leipzig for the opportunity for entrusting the grants to support the projects in Le Duc Lab. The graphical abstract was created in BioRender. Garten, A. (2025) https://BioRender.com/6e69nny

The open source publication is supported by the Open Access Publishing Fund of Leipzig University.

*Author contributions*: Linnaeus Tapia Bundalian (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Investigation [equal], Methodology [equal], Software [lead], Validation [equal], Visualization [lead], Writing—original

draft [lead]), Martina Strnadová (Methodology [equal], Resources [equal], Validation [equal], Writing—review & editing [equal]), Felix Garten (Software [supporting], Writing—review & editing [equal]), Susanne Horn (Investigation [equal], Software [equal], Validation [equal], Writing—review & editing [equal]), Udo Stenzel (Resources [equal], Software [supporting]), Johannes R. Lemke (Writing—review & editing [equal]), Saskia Biskup (Writing—review & editing [equal]), Björn Schulte (Writing—review & editing [equal]), Patrick May (Writing—review & editing [equal]), Frank Bösebeck (Investigation [equal], Resources [equal], Validation [equal], Writing—review & editing [equal]), Antje Garten (Writing—review & editing [equal]), Doreen Thor (Investigation [equal], Validation [equal], Writing—review & editing [equal]), Angela Schulz (Investigation [equal], Validation [equal], Writing—review & editing [equal]), Julia Hentschel (Writing—review & editing [equal]), Janet Kelso (Writing—review & editing [equal]), Torsten Schöneberg (Conceptualization [equal], Investigation [equal], Resources [equal], Validation [equal], Writing—review & editing [equal]), Diana Le Duc (Conceptualization [equal], Formal analysis [equal], Funding acquisition [lead], Investigation [equal], Methodology [equal], Writing—original draft [equal], Writing—review & editing [equal]), and Denny Popp (Methodology [equal], Validation [equal], Writing—review & editing [equal]).

## Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

## Conflict of interest

The authors declare no competing interests.

## Funding

This study is funded by the Else Kroner-Fresenius-Stiftung 2020_EKEA.42 to D.L.D., the German Research Foundation SFB 1052 project B10 to D.L.D. and A.G, and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC 1423/2 (project number 421152132) to T.S. and D.T..

## Data availability

The processed files and databases used in this study can be found in Zenodo repository https://zenodo.org/records/16375535. The Python package can be found in PyPi repository, https://pypi.org/project/symetrics/. The web application is accessible through https://symetrics.org/.

## References

1. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* 2011;12:683–91. https://doi.org/10.1038/nrg3051
2. Winterer G, Goldman D. Genetics of human prefrontal function. *Brain Res Rev* 2003;43:134–63. https://doi.org/10.1016/S0165-0173(03)00205-4
3. Hunt RC, Simhadri VL, Iandoli M *et al.* Exposing synonymous mutations. *Trends Genet* 2014;30:308–21. https://doi.org/10.1016/j.tig.2014.04.006
4. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF *et al.* Predicting splicing from primary sequence with Deep learning. *Cell* 2019;176:535–48. https://doi.org/10.1016/j.cell.2018.12.015
5. Sciascia S, Roccatello D, Salvatore M *et al.* Unmet needs in countries participating in the undiagnosed diseases network international: an international survey considering national health care and economic indicators. *Front Public Health* 2023;11:1248260. https://doi.org/10.3389/fpubh.2023.1248260
6. Graessner H, Zurek B, Hoischen A *et al.* Solving the unsolved rare diseases in Europe. *Eur J Hum Genet* 2021;29:1319–20. https://doi.org/10.1038/s41431-021-00924-8
7. Katsonis P, Koire A, Wilson SJ *et al.* Single nucleotide variations: biological impact and theoretical interpretation. *Protein Sci* 2014;23:1650–66. https://doi.org/10.1002/pro.2552
8. Walsh IM, Bowman MA, Soto Santarriaga IF *et al.* Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc Natl Acad Sci USA* 2020;117:3528–34. https://doi.org/10.1073/pnas.1907126117
9. Dhindsa RS, Wang Q, Vitsios D *et al.* A minimal role for synonymous variation in human disease. *Am Hum Genet* 2022;109:2105–9. https://doi.org/10.1016/j.ajhg.2022.10.016
10. Shi F, Yao Y, Bin Y *et al.* Computational identification of deleterious synonymous variants in human genomes using a feature-based approach. *BMC Med Genomics* 2019;12:12. https://doi.org/10.1186/s12920-018-0455-6
11. Giacoletto CJ, Rotter JI, Grody WW *et al.* Synonymous variants of uncertain silence. *Int J Mol Sci* 2023;24:10556, https://doi.org/10.3390/ijms241310556 .
12. Lin BC, Katneni U, Jankowska KI *et al. In silico* methods for predicting functional synonymous variants. *Genome Biol* 2023;24:126. https://doi.org/10.1186/s13059-023-02966-1
13. Jankowska KI, Meyer D, Holcomb DD *et al.* Synonymous ADAMTS13 variants impact molecular characteristics and contribute to variability in active protein abundance. *Blood Adv* 2022;6:5364–78.
14. Ranganathan Ganakammal S, Alexov E. An ensemble approach to predict the pathogenicity of synonymous variants. *Genes* 2020;11:1102, https://doi.org/10.3390/genes11091102 .
15. Dhindsa RS, Wang Q, Vitsios D *et al.* A minimal role for synonymous variation in human disease. *Am Hum Genet* 2022;109:2105–9. https://doi.org/10.1016/j.ajhg.2022.10.016
16. Mello AC, Leao D, Dias L *et al.* Broken silence: 22, 841 predicted deleterious synonymous variants identified in the human exome through computational analysis. *Genet Mol Biol* 2024;46:e20230125. https://doi.org/10.1590/1678-4685-gmb-2023-0125
17. Gudkov M, Thibaut L, Giannoulatou E. Quantifying negative selection on synonymous variants. *Human Genetics and Genomics Advances* 2024;5:100262. https://doi.org/10.1016/j.xhgg.2024.100262
18. Kovacs E, Tompa P, Liliom K *et al.* Dual coding in alternative reading frames correlates with intrinsic protein disorder. *Proc Natl Acad Sci USA* 2010;107:5429–34. https://doi.org/10.1073/pnas.0907841107
19. Vasu K, Khan D, Ramachandiran I *et al.* Analysis of nested alternate open reading frames and their encoded proteins. *NAR Genom Bioinform* 2022;4:lqac076.
20. Vihinen M. When a synonymous variant is nonsynonymous. *Genes* 2022;13:1485. https://doi.org/10.3390/genes13081485
21. Ranganathan Ganakammal S, Alexov E. An ensemble approach to predict the pathogenicity of synonymous variants. *Genes* 2020;11:1102, https://doi.org/10.3390/genes11091102 .
22. Zeng Z, Aptekmann AA, Bromberg Y. Decoding the effects of synonymous variants. *Nucleic Acids Res* 2021;49:12673–91. https://doi.org/10.1093/nar/gkab1159

23. Buske OJ, Manickaraj A, Mital S *et al.* Identification of deleterious synonymous variants in human genomes. *Bioinformatics* 2015;31:799. https://doi.org/10.1093/bioinformatics/btu765
24. Zeng Z, Bromberg Y. Predicting functional effects of synonymous variants: a systematic review and perspectives. *Front Genet* 2019;10:914. https://doi.org/10.3389/fgene.2019.00914
25. Zeng Z, Aptekmann AA, Bromberg Y. Decoding the effects of synonymous variants. *Nucleic Acids Res* 2021;49:12673–91. https://doi.org/10.1093/nar/gkab1159
26. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. In: *Proceedings of the seventh annual international conference on Research in computational molecular biology*. New York, NY, USA: ACM, 2003, 322–31. https://doi.org/10.1145/640075.640118
27. Gaither JBS, Lammi GE, Li JL *et al.* Synonymous variants that disrupt messenger RNA structure are significantly constrained in the human population. *Gigascience* 2021;10:giab023. https://doi.org/10.1093/gigascience/giab023
28. Davydov EV, Goode DL, Sirota M *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 2010;6:e1001025. https://doi.org/10.1371/journal.pcbi.1001025
29. Buske OJ, Manickaraj A, Mital S *et al.* Identification of deleterious synonymous variants in human genomes. *Bioinformatics* 2015;31:799. https://doi.org/10.1093/bioinformatics/btu765
30. Shi F, Yao Y, Bin Y *et al.* Computational identification of deleterious synonymous variants in human genomes using a feature-based approach. *BMC Med Genomics* 2019;12:12. https://doi.org/10.1186/s12920-018-0455-6
31. Zeng Z, Bromberg Y. Predicting functional effects of synonymous variants: a systematic review and perspectives. *Front Genet* 2019;10:914. https://doi.org/10.3389/fgene.2019.00914
32. Mount SM. A catalogue of splice junction sequences. *Nucleic Acids Res* 1982;10:459–72. https://doi.org/10.1093/nar/10.2.459
33. Will CL, Luhrmann R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 2011;3:a003707. https://doi.org/10.1101/cshperspect.a003707
34. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF *et al.* Predicting splicing from primary sequence with Deep learning. *Cell* 2019;176:535–48. https://doi.org/10.1016/j.cell.2018.12.015
35. Ha C, Kim J-W, Jang J-H. Performance evaluation of SpliceAI for the prediction of splicing of NF1 variants. *Genes* 2021;12:1308. https://doi.org/10.3390/genes12091308
36. Eng L, Coutinho G, Nahas S *et al.* Nonclassical splicing mutations in the coding and noncoding regions of the ATM gene: maximum entropy estimates of splice junction strengths. *Hum Mutat* 2004;23:67–76. https://doi.org/10.1002/humu.10295
37. Cooper GM, Stone EA, Asimenos G *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* 2005;15:901–13. https://doi.org/10.1101/gr.3577405
38. Huber CD, Kim BY, Lohmueller KE. Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. *PLoS Genet* 2020;16:e1008827. https://doi.org/10.1371/journal.pgen.1008827
39. Henn BM, Botigué LR, Peischl S *et al.* Distance from sub-saharan Africa predicts mutational load in diverse human genomes. *Proc Natl Acad Sci USA* 2016;113:E440–9. https://doi.org/10.1073/pnas.1510805112
40. Marsden CD, Vecchyo O-D, D. O'B *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc Natl Acad Sci USA* 2016;113:152–7. https://doi.org/10.1073/pnas.1512501113
41. Paulet D, David A, Rivals E. Ribo-seq enlightens codon usage bias. *DNA Res* 2017;24:303–210. https://doi.org/10.1093/dnares/dsw062
42. Xu X, Liu Q, Fan L *et al.* Analysis of synonymous codon usage and evolution of begomoviruses. *J Zhejiang Univ Sci B* 2008;9:667–74. https://doi.org/10.1631/jzus.B0820005
43. Cutter AD, Wasmuth JD, Blaxter ML. The evolution of biased codon and amino acid usage in nematode genomes. *Mol Biol Evol* 2006;23:2303–15. https://doi.org/10.1093/molbev/msl097
44. Liao X, Zhu W, Zhou J *et al.* Repetitive DNA sequence detection and its role in the human genome. *Commun Biol* 2023;6:954. https://doi.org/10.1038/s42003-023-05322-y
45. Zhang Z, Gerstein M. Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. *Nucleic Acids Res* 2003;31:5338–48. https://doi.org/10.1093/nar/gkg745
46. Yu G, Wang L-G, Han Y *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 2012;16:284–7. https://doi.org/10.1089/omi.2011.0118
47. Wu T, Hu E, Xu S *et al.* clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation* 2021;2:100141. https://doi.org/10.1016/j.xinn.2021.100141
48. Draghici S, Khatri P, Tarca AL *et al.* A systems biology approach for pathway level analysis. *Genome Res* 2007;17:1537–45. https://doi.org/10.1101/gr.6202607
49. Wen P, Xiao P, Xia J. dbDSM: a manually curated database for deleterious synonymous mutations. *Bioinformatics* 2016;32:1914–6. https://doi.org/10.1093/bioinformatics/btw086
50. Landrum MJ, Lee JM, Benson M *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016;44:D862–8. https://doi.org/10.1093/nar/gkv1222
51. Chawla NV, Bowyer KW, Hall LO *et al.* SMOTE: synthetic minority over-sampling technique. *Jair* 2002;16:321–57. https://doi.org/10.1613/jair.953
52. Abdelhamid M, Desai A. Balancing the scales: a comprehensive study on tackling class imbalance in binary classification. arXiv, https://arxiv.org/abs/2409.19751, 29 September 2024, preprint: not peer reviewed.
53. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *Bmc Genomics [Electronic Resource]* 2020;21:6. https://doi.org/10.1186/s12864-019-6413-7
54. Sokolova M, Japkowicz N, Szpakowicz S. Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In *AAAI Workshop* 2006;- Technical Report.Vol. WS-06-06, pp.24–9.
55. Sangkuhl K, Römpler H, Busch W *et al.* Nephrogenic diabetes insipidus caused by mutation of Tyr205: a key residue of V2 vasopressin receptor function. *Hum Mutat* 2005;25:505. https://doi.org/10.1002/humu.9337
56. Schoneberg T, Yun J, Wenkert D *et al.* Functional rescue of mutant V2 vasopressin receptors causing nephrogenic diabetes insipidus by a co-expressed receptor polypeptide. *EMBO J* 1996;15:1283–91. https://doi.org/10.1002/j.1460-2075.1996.tb00470.x
57. Bonner TI, Young AC, Brann MR *et al.* Cloning and expression of the human and rat m5 muscarinic acetylcholine receptor genes. *Neuron* 1988;1:403–10. https://doi.org/10.1016/0896-6273(88)90190-0
58. Römpler H, Yu H-T, Arnold A *et al.* Functional consequences of naturally occurring DRY motif variants in the mammalian chemoattractant receptor GPR33. *Genomics* 2006;87:724–32.
59. Liebing A-D, Krumbholz P, Stäubert C. Protocol to characterize gi/o and gs protein-coupled receptors in transiently transfected cells using ELISA and cAMP measurements. *STAR Protocols* 2023;4:102120. https://doi.org/10.1016/j.xpro.2023.102120
60. Gasperini M, Starita L, Shendure J. The power of multiplexed functional analysis of genetic variants. *Nat Protoc* 2016;11:1782–7. https://doi.org/10.1038/nprot.2016.135
61. Olvera-León R, Zhang F, Offord V *et al.* High-resolution functional mapping of RAD51C by saturation genome editing. *Cell* 2024;187:5719–5734.e19. https://doi.org/10.1016/j.cell.2024.08.039
62. Waters AJ, Brendler-Spaeth T, Smith D *et al.* Saturation genome editing of BAP1 functionally classifies somatic and germline

variants. *Nat Genet* 2024;56:1434–45. https://doi.org/10.1038/s41588-024-01799-3
63. Kinsella RJ, Kähäri A, Haider S *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* 2011;2011:bar030.
64. Karczewski KJ, Francioli LC, Tiao G *et al.* The mutational constraint spectrum quantified from variation in 141, 456 humans. *Nature* 2020;581:434–43. https://doi.org/10.1038/s41586-020-2308-7
65. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 2019;9:5233. https://doi.org/10.1038/s41598-019-41695-z
66. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn* 2006;63:3–42. https://doi.org/10.1007/s10994-006-6226-1
67. Rentzsch P, Witten D, Cooper GM *et al.* CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 2019;47:D886–94. https://doi.org/10.1093/nar/gky1016
68. Schaafsma GCP, Vihinen M. VariSNP, a benchmark database for variations from dbSNP. *Hum Mutat* 2015;36:161–6. https://doi.org/10.1002/humu.22727
69. Deng Y, Gao L, Wang B *et al.* HPOSim: an R package for phenotypic similarity measure and enrichment analysis based on the human phenotype ontology. *PLoS One* 2015;10:e0115692. https://doi.org/10.1371/journal.pone.0115692
70. de Coulgeans CD, Silvy M, Halverson G *et al.* Synonymous nucleotide polymorphisms influence <scp>D</scp>ombrock blood group protein expression in <scp>K</scp>562 cells. *Br J Haematol* 2014;164:131–41.
71. Simhadri VL, Hamasaki-Katagiri N, Lin BC *et al.* Single synonymous mutation in factor IX alters protein properties and underlies haemophilia B. *J Med Genet* 2017;54:338–45. https://doi.org/10.1136/jmedgenet-2016-104072
72. Schernthaner-Reiter MH, Adams D, Trivellin G *et al.* A novel AVPR2 splice site mutation leads to partial X-linked nephrogenic diabetes insipidus in two brothers. *Eur J Pediatr* 2016;175:727–33. https://doi.org/10.1007/s00431-015-2684-4
73. Pruner I, Farm M, Tomic B *et al.* The silence speaks, but we do not listen: synonymous c.1824C>T gene variant in the last exon of the prothrombin gene as a new prothrombotic risk factor. *Clin Chem* 2020;66:379–89. https://doi.org/10.1093/clinchem/hvz015
74. Durkie M, Cassidy E-J, Berry I *et al.* ACGS Best practice guidelines for variant classification in rare disease 2024. 2024. *14 July 2025*.



**Received:** April 26, 2025. **Revised:** November 3, 2025. **Accepted:** December 1, 2025