Predicting interaction-specific protein–protein interaction perturbations by missense variants with MutPred-PPI

Ross Stewart[1], Florent Laval[2,3,4,5,6,7,8], Georges Coppin[2,3,4,6], Kerstin Spirohn-Fitzgerald[2,3,4], Maxime Tixhon[2,3,4,9], Tong Hao[2,3,4], Michael A. Calderwood[2,3,4], Matthew Mort[10], David N. Cooper[10], Marc Vidal[2,3], Predrag Radivojac[1,*]

[1]Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA
[2]Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA, USA
[3]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA
[4]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA
[5]TERRA Teaching and Research Centre, University of Liège, Gembloux, Belgium
[6]Laboratory of Viral Interactomes, GIGA Institute, University of Liège, Liège, Belgium
[7]Laboratory of Molecular and Cellular Epigenetics, GIGA Institute, University of Liège, Liège, Belgium
[8]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada
[9]Computational Biology and Bioinformatics, Université Libre de Bruxelles, Brussels, Belgium
[10]Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, UK

*Corresponding author: Predrag Radivojac (predrag@northeastern.edu)

Abstract

Disruption of protein–protein interactions (PPIs) is a major mechanism of a variant's deleterious effect. Computational tools are needed to assess such variants at scale, yet existing predictors rarely consider loss of specific interactions, particularly when variants perturb binding interfaces without significantly affecting protein stability. To address this problem, we present MutPred-PPI, a graph attention network that predicts interaction-specific (edgetic) effects of missense variants by operating on AlphaFold 3-based protein complex contact graphs with protein language model embeddings imposed upon nodes. We systematically evaluated our model with stringent group cross-validation as well as benchmark data recently collected within the IGVF Consortium. MutPred-PPI outperformed all baseline methods across all evaluation criteria, achieving an AUC of 0.85 on seen proteins and 0.72 on previously unseen proteins in cross-validation, demonstrating strong generalizability despite scarce training data. To demonstrate biomedical relevance, we applied MutPred-PPI to variants from ClinVar, HGMD, COSMIC, gnomAD, and two *de novo* neurodevelopmental disorder-linked datasets. Disease-associated variants from ClinVar and HGMD showed strong enrichment for both quasi-null and edgetic effects, whereas population variants from gnomAD increasingly preserved interactions with higher allele frequencies. Notably, we observed a strong edgetic disruption signature in highly recurrent cancer variants from both the full COSMIC dataset and a subset of variants from oncogenes. Recurrent tumor suppressor gene variants and autism spectrum disorder-associated variants exhibited moderate quasi-null enrichment, whilst neurodevelopmental disorder-linked variants showed a weak edgetic disruption signature. These results indicate distinct PPI perturbation mechanisms across disease types and show that MutPred-PPI captures functionally relevant molecular effects of pathogenic variants.

1 Introduction

Protein–protein interactions (PPIs) govern cellular function, with the human proteome exhibiting up to two million interactions [1]–[3]. Genetic variants residing in distinct binding interfaces of a protein can have vastly different functional consequences depending upon which PPIs are disrupted [4]. Whereas some variants cause complete loss of function (LoF) and interactions, "edgetic" variants selectively disrupt PPIs while maintaining other interactions [5], [6]. Variants exhibiting these edgetic mechanisms have been implicated in various diseases [5]. For example, cancer-associated mutations in the coiled-coil domain of BRCA1 impair homologous recombination repair by disrupting its interaction with PALB2, while preserving other functions [7], [8]. This interaction-specific nature of variant effects necessitates computational methods capable of predicting interaction perturbations at the resolution of individual protein pairs.

Most variant effect predictors assess overall pathogenicity or other properties without considering specific interaction partners [9], [10]. Methods such as REVEL [11], MutPred2 [12], and AlphaMissense [13] accurately predict variant pathogenicity [14] and have been recommended for clinical use [15], but prediction of interaction-specific effects remains limited. Existing interaction-specific approaches such as SAAMBE-3D [16] and mCSM-PPI2 [17] predict quantitative changes in binding free energy ($\Delta\Delta G$) induced by mutations in experimental protein complexes, rather than outputting a probabilistic score that indicates PPI perturbation. SWING [18], a recent advance, uses a sliding window approach to model interactions as language, predicting probabilistic interaction-specific perturbations from amino acid sequence alone. However, stringent evaluation on held-out proteins remains necessary to support generalization.

The convergence of recent developments in machine learning and biotechnology platforms is creating new opportunities for accurate prediction of PPI perturbations. AlphaFold 3 [19] mitigates the requirement for experimental structures by providing accurate structural predictions for protein complexes of interest. Protein language models (pLMs), such as ProtT5 [20], generate per-residue embeddings that capture evolutionary and functional information without requiring costly multiple sequence alignments. Furthermore, recent variant interaction profiling through the Impact of Genomic Variation on Function (IGVF) Consortium [21] provides newly generated experimental data for the training and validation of machine learning models.

In this work, we build upon these developments and present MutPred-PPI, a graph attention network (GAT) [22] approach for predicting interaction-specific effects of missense variants. Our method improves upon previous work by: (1) using AlphaFold 3 to generate protein complex structures for interacting protein pairs; (2) exploiting pre-trained pLM embeddings that generalize across protein families; and (3) employing a two-stage training strategy that first learns general stability disruption patterns before fine-tuning on interaction-specific effects. We validate our approach using stringent cross-validation and benchmark schemes that test generalization to unseen proteins and demonstrate superior performance compared to existing methods. Application to disease-associated variants revealed distinct PPI perturbation patterns between disease types, supporting PPI perturbation as a key pathogenic mechanism. Overall, these results provide new insights into molecular mechanisms of genetic disease consequent to mutation.

2 Methods

2.1 Preprocessing

The end-to-end pipeline of MutPred-PPI is shown in Figure 1. Each data point, which we refer to as a variant-partner pair, is represented as a triplet (i, v, p) where i is the wild-type interactor protein, v specifies the missense variant within i, and p is a known interaction partner of i.

First, the amino acid sequences of i and p are entered into AlphaFold 3 (default parameters) to model the protein complex. Then, we compute the contact graph from the AlphaFold 3 structure, with nodes representing amino acid residues and edges representing contacting residues. We define residues as in contact if any heavy atom distance is ≤ 4.5 Å [23].

Next, we generate pLM embeddings with ProtT5 (ProtT5-XL-UniRef50) [20] for the wild-type sequences of both i and p, as well as the mutant sequence of i (denoted i_v). The pLM embeddings of i_v and p are then mapped onto the contact graph of the protein complex as node features. We chose to incorporate the mutant embeddings of the interactor protein rather than wild-type embeddings to capture variant-specific context, as computing separate AlphaFold 3 complexes for each variant would have

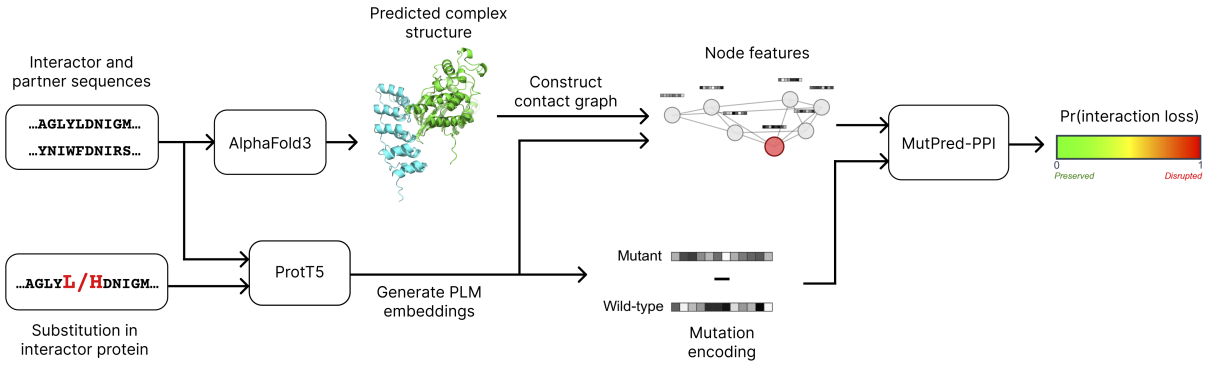

Figure 1: MutPred-PPI end-to-end pipeline. Wild-type sequences of i and p are input to AlphaFold 3 to generate the complex structure. Wild-type and mutant sequences are input to ProtT5 to generate embeddings, which are passed through MutPred-PPI to give Pr(interaction loss).

been computationally prohibitive, and AlphaFold models have been shown to be insensitive to structural changes induced by missense variants [24].

Finally, the mutation is encoded as the difference between the mutant and wild-type residue embeddings at the variant position, followed by z-score normalization across variants. We feed both the node features and the mutation encoding into MutPred-PPI to predict the probability of interaction loss.

2.2 Model architecture

The model architecture, as shown in Figure 2, consists of three main components: graph attention layers to learn local structural context, a mutation processor to encode mutation effects, and a prediction head for binary classification. Based on the principle that interface mutations are enriched for PPI-perturbing effects [25], our model architecture prioritizes the local structural context at the mutation site where protein–protein contacts are directly affected, rather than global protein features. A detailed mathematical representation of the model architecture is shown in Supplementary Section 1.

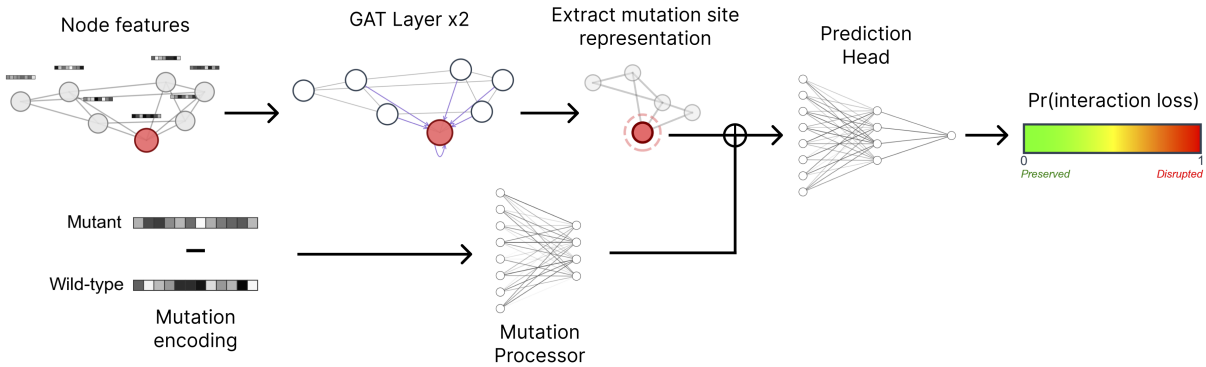

Figure 2: MutPred-PPI model architecture. The node features are propagated through both GAT layers to capture local structural context. The mutation encoding is processed separately, then combined with the node representation at the mutation site for final prediction. Symbol $\oplus$ denotes concatenation.

2.3 Training procedure

We employed a two-stage transfer learning strategy. MutPred-PPI was first pretrained to predict protein stability disruption using experimental structures of monomeric proteins, then fine-tuned for interaction-specific perturbations using AlphaFold 3 complexes. Variants were classified as stability-preserving if $|\Delta\Delta G| < 0.5\,\text{kcal/mol}$ and stability-disrupting if $|\Delta\Delta G| \geq 1.5\,\text{kcal/mol}$ [26]. The stability pretraining dataset (176,908 variants) was substantially larger than the available PPI perturbation data (5,894 variant-partner pairs), enabling the model to learn generalizable mutational representations before specializing in edgetic effects. Class weighting was applied to address class imbalance in both pretraining and fine-tuning.

2.4 Data

We combined three main datasets for model training and evaluation: PPI perturbation data of variants across Mendelian disorders from Sahni *et al.* [27], population variant PPI perturbation data from Fragoza *et al.* [28], and stability data for pretraining from Blaabjerg *et al.* [29]. For additional performance evaluation, we used in-house benchmark data recently generated through the IGVF Consortium, which we refer to as VarChAMP (Variant Characterization Across the Mendelian Proteome). The VarChAMP dataset was generated through yeast two-hybrid assays using histidine dropout media supplemented with 1 mM 3-Amino-1,2,4-triazole (3-AT), retaining only variants with successfully recovered sequences. Detailed dataset statistics are shown in Table 1. For all datasets, we removed duplicate data points, those with conflicting labels, and those with failed structure generation.

Table 1: Dataset statistics for model training and evaluation. Triplets refer to variant-partner combinations.

Dataset	**Proteins**	**Pairs**	**Variants**	**Triplets**	**Disruptive**	**Non-disruptive**
PPI Perturbation Data						
Sahni et al. (Mendelian)	537	559	489	1,487	532	955
Fragoza et al. (Population)	1,868	2,252	1,904	4,515	847	3,667
VarChAMP (IGVF)	429	436	633	2,935	2,067	868
Combined Training Sets						
Sahni, Fragoza	2,083	2,710	2,310	5,894	1,395	4,499
Sahni, Fragoza, VarChAMP	2,367	3,118	2,918	8,827	3,466	5,361
Stability Pretraining Data						
Blaabjerg et al.	45	-	176,908	-	112,028	64,880

2.5 Evaluation

To mitigate information leakage commonly arising in paired protein input schemes (due to sequence similarity and shared individual proteins) [30], we first grouped each variant-partner pair by clustering the concatenated sequence of both wild-type proteins. To cluster sequences, we used CD-HIT [31] with a 50% sequence identity threshold. We then evaluated the model through 30 iterations of 10-fold group cross-validation (grouped by sequence cluster) on the training dataset as well as blind-test predictions on the benchmark dataset. To aggregate predictions during cross-validation, we averaged the performance across each fold and iteration, weighted by the number of test data points. For benchmark predictions, we constructed an ensemble of 10 models (one per cross-validation fold), averaging predictions across models. Owing to the heterogeneity between the two PPI perturbation training datasets, we also evaluated the effects of fine-tuning on only the dataset of Mendelian variants versus the combined dataset.

Test predictions were binned into three classes, per Park and Marcotte [30], because protein- and variant-level information leakage can still occur from the pair input scheme. Class one represents predictions with both proteins shared in the training data. Class two represents predictions with one protein shared in the training data. Class three represents predictions with neither protein shared in the training data. Variant-partner pairs are never shared across training and test partitions; however, shared variants may appear in classes one and two with different partner combinations. Class three ensures both proteins and variants are unseen.

We compared MutPred-PPI to SWING [18] (trained on identical cross-validation partitions) and two pretrained proxy models: SAAMBE-3D [16] and MutPred2 [12]. In the official implementation for SWING, a language model is pretrained on training and test data (without labels), and then a discriminator is trained on embeddings generated from that model. To eliminate any information leakage from the test distribution, we evaluated SWING under two scenarios: pretraining on the training and test data (test pretrain), and pretraining on training data only (blind-test). This blind-test training scenario ensured that test data were truly unseen at inference time. To rank SAAMBE-3D predictions, we used the predicted change in binding free energy induced by a mutation (operating on the AlphaFold 3 structure). SAAMBE-3D predictions were binned into test classes based on overlap with its own training data (experimental PDB structures), which differed from our cross-validation splits. MutPred2, a pathogenicity predictor, determined baseline performance because variant pathogenicity can serve as a proxy for interaction loss. However, MutPred2 does not incorporate partner protein information and has constant predictions for a variant across distinct partners (and hence, test classes did not apply). For blind-test evaluation on the benchmark dataset, variants overlapping with the training data of each model were excluded; thus, potential information leakage in classes one and two arose only at the protein

level. MutPred2 predictions were not filtered based on variant overlap as MutPred2 was not trained on PPI perturbation data.

2.6 Predictions on variant repositories

After fine-tuning on a combined dataset containing the original training data of Mendelian and population variants as well as VarChAMP benchmark data, we applied the resulting ensemble (from cross-validation) to variants from multiple disease and population variant repositories. By aggregating predictions across partners, we predicted edgotypes (i.e., patterns of interaction perturbations induced by a mutation) of variants from ClinVar [32], COSMIC [33], HGMD [34], gnomAD [35], and two *de novo* neurodevelopmental disorder datasets [12], [36]; see Supplementary Section 2 for detailed sources.

First, we gathered variants from each repository alongside their respective partner proteins from BioGRID [37] (physical binding evidence only). Owing to the computational cost of AlphaFold 3, for large datasets, we selected a subset of interactor proteins and up to 10 partner proteins (for each interactor protein) for structure generation and predicted interaction perturbation for all variants in those protein pairs. For variants in the partner protein, we flipped the protein roles (treating the partner as the interactor) before model input. Additionally, we downloaded 126,118 high-confidence precomputed AlphaFold 3 predictions from ProtVar [38] and generated predictions for variants in either protein.

Variants in the partner protein of generated complexes, as well as either protein of precomputed complexes, often had far fewer than 10 partners tested, reducing confidence in edgotype classification. To mitigate this limitation, we only analyzed variants with at least three tested partner proteins when three or more partners existed in BioGRID; otherwise variants were processed normally. Protein pairs with either protein shared in the training dataset were excluded.

Variants were classified as "quasi-wild-type" if all partner interactions were predicted to be unperturbed, "quasi-null" if all were predicted to be perturbed, and "edgetic" if predictions were mixed across partners [27]. Given the class-weighted training objective, we applied a threshold of 0.5 to determine whether an interaction was considered perturbed. Because interaction loss probabilities across partners are not independent, and model outputs may be miscalibrated, a probabilistic formulation (e.g., $\Pr(\text{quasi-null} \mid i, v) = \prod_p \Pr(\text{interaction loss} \mid i, v, p)$) was inappropriate.

2.7 Computational efficiency

A single NVIDIA V100 GPU processes approximately 100 variant-partner pairs per minute with precomputed AlphaFold 3 structures and contact graphs. AlphaFold 3 structures (using MMseqs2 [39] for sequence alignment) and derived contact graphs can take several minutes to generate per protein complex. Structural features should be generated once per protein pair and cached for all variants. ProtT5 embedding generation is the primary computational cost after structure prediction.

3 Results

3.1 Performance Evaluation

Cross-validation results on the training dataset containing Mendelian and population variants are shown in Figure 3. MutPred-PPI outperformed all compared methods across test classes. SWING (test pretrain) performed well on class one but showed substantial performance degradation on classes two and three. By contrast, MutPred-PPI maintained more stable performance across classes. SAAMBE-3D showed limited performance across classes, and MutPred2 provided a partner-agnostic baseline.

The blind-test evaluation results on the VarChAMP benchmark dataset are shown in Figure 4. MutPred-PPI outperformed all compared methods across test classes, with an expected modest performance decline from class one to three. Surprisingly, SWING (test pretrain) often performed worse than SWING (blind-test). MutPred2 achieved comparable performance to MutPred-PPI on class three, and SAAMBE-3D showed minimal predictive signal.

To compare performance across datasets, we also fine-tuned MutPred-PPI on the Mendelian dataset alone. During cross-validation (Supplementary Figure S1), MutPred-PPI similarly outperformed all baseline methods across test classes. Interestingly, evaluating both MutPred-PPI training configurations on the VarChAMP benchmark dataset (Supplementary Figure S2) showed that including population variants in the fine-tuning dataset reduced class three performance.

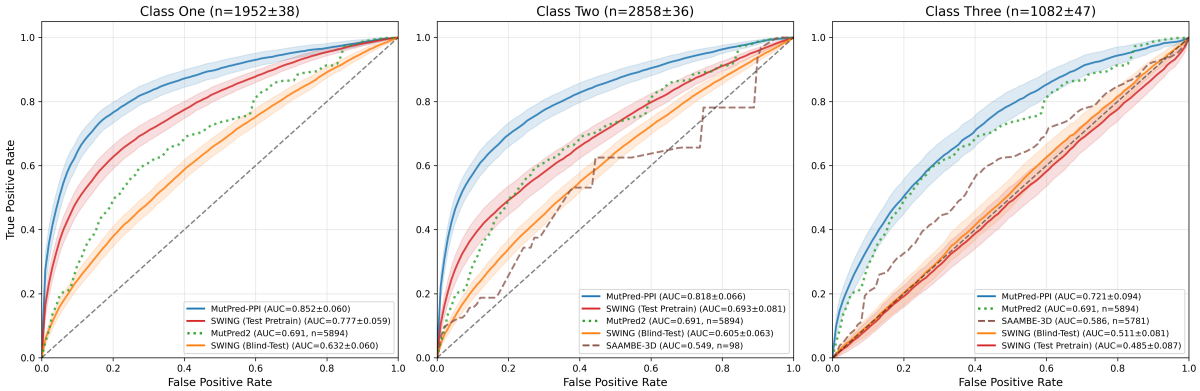

Figure 3: Cross-validation performance on combined Mendelian and population variant training data ($n = 5{,}894$). Solid lines show mean ROC curves across 30 iterations of 10-fold group cross-validation; shaded regions indicate ± 1 standard error of the mean computed across the 10 folds. Classes one, two, and three represent predictions with both proteins shared, one protein shared, and neither protein shared between training and test data, respectively. MutPred2 predictions are shown constant across test classes (partner-agnostic). Subplot titles show test set sizes. Proxy methods (dashed and dotted lines) display sample sizes in parentheses. Curves with $n < 30$ were excluded.

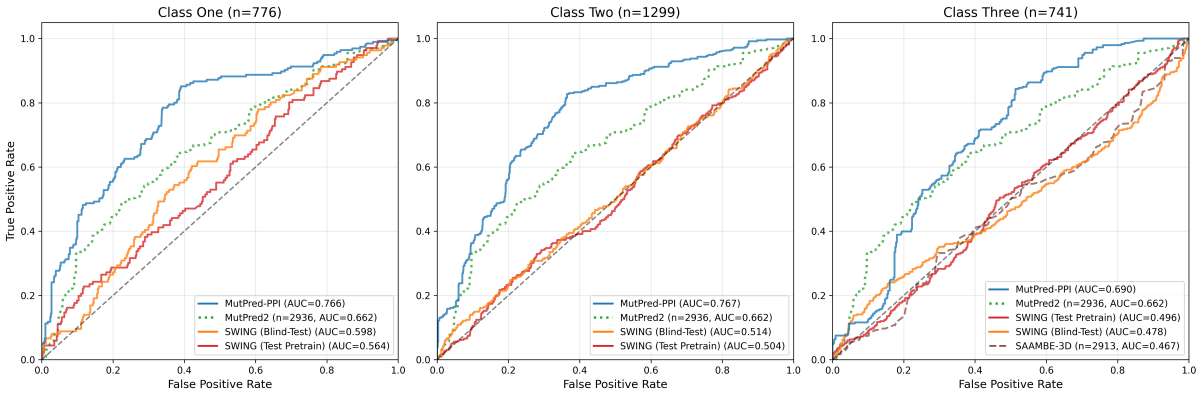

Figure 4: Blind-test ROC curves on the VarChAMP benchmark dataset ($n = 2{,}816$). Classes one, two, and three represent predictions with both proteins shared, one protein shared, and neither protein shared between training and test data, respectively. MutPred2 predictions are shown constant across test classes (partner-agnostic). Subplot titles show test class sizes. Proxy methods (dashed and dotted lines) display sample sizes in parentheses. Curves with $n < 30$ were excluded.

3.2 Application to disease and population variant repositories

After fine-tuning a final ensemble on the combined training dataset with VarChAMP data (cross-validation performance shown in Supplementary Figure S3), we applied MutPred-PPI to variants from ClinVar, COSMIC, HGMD, gnomAD, and two *de novo* neurodevelopmental disorder datasets. Variant groups were constructed by stratifying each repository: ClinVar by clinical significance; COSMIC by recurrence and gene classification (oncogene vs. tumor suppressor gene, TSG); HGMD as a single group; gnomAD by allele frequency (AF); and two independent *de novo* neurodevelopmental disorder datasets (NDD and ASD) by case/control status. NDD includes cases and controls across four disorders (autism spectrum disorder, intellectual disability, schizophrenia, and epileptic encephalopathy), while ASD contains autism spectrum disorder cases only.

For each variant group, we calculated the enrichment or depletion of quasi-null and edgetic variants relative to the gnomAD population baseline as shown in Figure 5. The enrichment trend E was calculated as

$$E = \frac{f_{obs} - f_{gnomAD}}{f_{obs} + f_{gnomAD}} \tag{1}$$

where $f_{obs}, f_{gnomAD} \in [0, 1]$ are the observed rates (relative frequencies) of quasi-null or edgetic variants in the variant group of interest and gnomAD, respectively. The trend is positive if $f_{obs} > f_{gnomAD}$ (enrichment) and negative if $f_{gnomAD} > f_{obs}$ (depletion). The theoretical range of E is $[-1, 1]$, where values approaching 1 indicate strong enrichment and values approaching -1 indicate strong depletion.

Statistical significance was assessed using bootstrap resampling with 100,000 iterations. Enrichments were significant if their sign remained consistent across $(1 - \alpha_{\text{Bonferroni}})$ of bootstrap samples, with Bonferroni correction ($\alpha = 0.05$, $n_{\text{tests}} = 64$) applied across all comparisons.

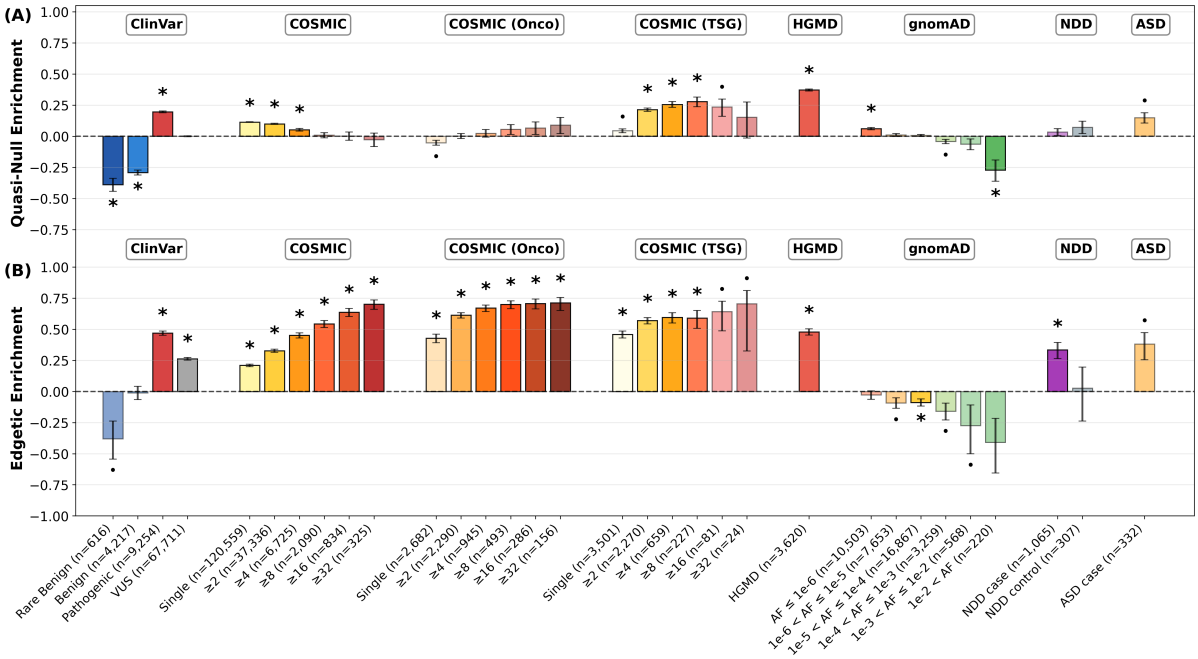

Figure 5: Edgotype enrichment of each variant group with respect to gnomAD across 100,000 bootstrap iterations using all available partners per variant. COSMIC, COSMIC (Onco), and COSMIC (TSG) are binned by recurrence; gnomAD is binned by allele frequency (AF). Rare benign variants are defined as AF $\leq 1\%$. Error bars indicate 68% confidence intervals. Asterisks indicate statistical significance with Bonferroni correction ($\alpha = 0.05$, $n_{\text{tests}} = 64$). Dots indicate significance at the uncorrected $\alpha = 0.05$ level. (A) Quasi-null enrichment. (B) Edgetic enrichment. VUS: variants of uncertain significance.

We observed that variant groups have differential partner coverage (Supplementary Table S1). To account for potential biases this might introduce, we also performed a partner-controlled enrichment analysis by sampling exactly three partners per variant in each bootstrap iteration (Supplementary Figure S4).

3.2.1 ClinVar and HGMD variants

In ClinVar, pathogenic variants show significant enrichment of quasi-null and edgetic effects, while benign variants are depleted for quasi-null effects. Rare benign variants (AF $\leq 1\%$) exhibit similar depletion. Variants of uncertain significance (VUS) show intermediate PPI perturbation effects between pathogenic and benign, while also being enriched for edgetic effects with respect to the population. HGMD disease-linked variants, similar to ClinVar pathogenic variants, show substantial enrichment for both quasi-null and edgetic effects. In the partner-controlled analysis, these enrichment patterns largely hold, with VUS exhibiting weak enrichment for both quasi-null and edgetic effects, and benign variants showing strong depletion for both effects.

3.2.2 Cancer-associated variants

We analyzed cancer-associated somatic mutations across three gene groups: the full dataset (COSMIC), strictly oncogenes (COSMIC (Onco)), and strictly TSGs (COSMIC (TSG)). In both the full dataset and oncogenes, we observe monotonic trends of increasing edgetic enrichment with recurrence. The full dataset also shows the opposite trend for quasi-null enrichment in statistically significant bins. By contrast, while also exhibiting similar edgetic enrichment trends, TSG variants display a monotonic trend of increasing quasi-null enrichment with recurrence across statistically significant bins, although this trend weakens in the highest recurrence bins ($\geq$16 and $\geq$32) where sample sizes are limited ($n = 81$ and $n = 24$, respectively). These patterns largely persist in the partner-controlled analysis, although edgetic enrichment trends in TSG variants are no longer observed.

3.2.3 Population variants

As allele frequency increases in gnomAD population variants, we observe trends towards depletion of both quasi-null and edgetic effects, exhibiting near-monotonic patterns. In the partner-controlled analysis, these patterns hold although are less extreme.

3.2.4 Neurodevelopmental disorder variants

In the NDD dataset (variants across the four neurodevelopmental disorders), we observe significant edgetic enrichment in case variants but not controls. Case variants from the ASD-specific dataset, although not statistically significant after Bonferroni correction ($n = 332$ variants), exhibit substantial enrichment for quasi-null and particularly edgetic effects. In the partner-controlled analysis, NDD and ASD-specific case variants lose statistically significant edgetic enrichment, and ASD-specific variants show stronger quasi-null enrichment.

4 Discussion

Transfer learning enables generalization to unseen proteins. MutPred-PPI outperformed all compared methods and demonstrated superior generalization to unseen proteins. In cross-validation, the performance of SWING dropped from AUC 0.78 (class one) to 0.49 (class three), while MutPred-PPI maintained a more stable performance (0.85 to 0.72). When performing an ablation study by removing the pretraining step and training on PPI perturbation data alone, we observed protein-level overfitting similar to SWING (data not shown), which justified the use of a two-stage training strategy.

Disease variants are enriched for PPI perturbation. Across HGMD and ClinVar, pathogenic variants showed strong enrichment for quasi-null and edgetic effects, while ClinVar benign variants showed strong depletion of quasi-null effects, with depletion patterns persisting in rare benign variants. Disease-associated variants are known to frequently disrupt protein stability [40] and be enriched at PPI interfaces [41]. Our results suggest that pathogenic variants perturb PPIs through LoF and edgetic mechanisms [27], which is consistent with these established principles. The trend of VUS demonstrating an intermediate functional disruption effect between pathogenic and benign variants is also consistent with experimental findings [42].

Highly recurrent cancer variants are enriched for edgetic mechanisms. The recurrence of variants shows a strong correlation with predicted edgetic effect in cancer variants. This aligns with established cancer biology, as variants with higher recurrence are more likely to be oncogenic drivers [43], and the alteration of PPIs plays a major role in oncogenic signaling [44]. Cancer driver genes are strongly enriched for mutations in PPI interfaces [41], [45], particularly at central network positions [46]. Although MutPred-PPI does not predict gain of interactions, the monotonic trend of increasing edgetic enrichment from single-occurrence to high recurrence observed in both the full dataset of COSMIC variants and the subset of oncogene variants indicates positive selection for network-rewiring mutations in oncogenes. The absence of quasi-null enrichment further suggests that edgetic disruption, rather than protein destabilization, can contribute to gain of function, as demonstrated by recurrent *PIK3CA* helical domain mutations that selectively disrupt inhibitory p85 binding, consequently causing gain of function [47].

By contrast, TSG variants show a monotonic trend of increasing quasi-null enrichment with recurrence in statistically significant bins, while also exhibiting weak edgetic enrichment (edgetic trends are lost in the partner-controlled analysis). This difference reflects distinct selective pressures, where recurrent mutations in tumor suppressors favorably exhibit complete LoF mechanisms, as demonstrated experimentally for *RUNX1* [48]. Additionally, the contrast between enrichment trends of cancer and inherited disease mutations aligns with prior observations that these mutation types exhibit distinct molecular disruption patterns [49].

Population variants show depletion of PPI perturbation with increasing allele frequency. Among population variants, increasing allele frequencies correlate with benign status due to purifying selection against deleterious mutations [50], [51], and hence PPI-perturbing variants. This depletion trend is observed in both gnomAD predictions and experimentally profiled population variants [28], and contrasts with the enrichment observed in disease-associated variants. Nevertheless, while depleted, common population variants are not necessarily devoid of PPI perturbation mechanisms.

Neurodevelopmental disorder variants show variable PPI perturbation patterns. NDD case variants showed significant edgetic enrichment. However, this differs between analyses, suggesting weak

overall enrichment. This weaker signature may reflect that *de novo* missense variants in neurodevelopmental disorders often show more subtle functional effects [52], [53], and that not all *de novo* case variants are causative, unlike ClinVar and HGMD variants with stringent curation criteria. ASD-specific variants exhibit moderate enrichment for both quasi-null and edgetic effects, although the partner-controlled analysis suggests primarily quasi-null enrichment. Larger datasets are needed to more definitively characterize PPI perturbation mechanisms in neurodevelopmental disorders.

Implications for clinical variant interpretation. The ability of MutPred-PPI to predict interaction-specific effects of mutations has direct implications for clinical variant interpretation. Identifying specific PPIs that are perturbed offers mechanistic insights into disease etiology, revealing disrupted pathways and explaining the molecular basis of pathogenicity. For VUS, the observed intermediate PPI perturbation enrichment between pathogenic and benign variants suggests that MutPred-PPI could aid in variant reclassification, particularly within ACMG/AMP frameworks [54], where interaction-specific predictions provide mechanistic evidence beyond traditional pathogenicity scores, similar to splicing disruption predictions [55]. Our predictions could also prove useful for precision medicine approaches by indicating whether mutating a protein of interest would perturb specific PPIs while leaving others intact.

Limitations. Our per-variant predictions may underestimate edgetic effects due to limited partner sampling and differential coverage across variant groups. This differential coverage reflects potential BioGRID ascertainment bias [56], and that disease variants are enriched in hub proteins with central network positions [46], [57]. However, our partner-controlled enrichment analysis (Supplementary Figure S4) showed that disease enrichment trends largely hold when controlling for this bias, thereby supporting the biological validity of our findings.

Future work. The field of predicting interaction-specific effects of mutations remains underdeveloped owing to problem difficulty and the scarcity of high-quality, context-specific experimental data. Continued production of such data remains paramount for advancing computational predictors towards accurate mechanistic models. Further computational development necessitates advances in predictor accuracy, prediction of interaction gain, extensions to other macromolecules and ligands, and coverage of additional variant types. The consequences of alternative splicing on PPI perturbations also remain a challenging prediction task [58] and could be addressed through a modified variant effect prediction framework. As these methods mature, they will become increasingly valuable for understanding how genetic variation causes disease through interactome-level effects.

Acknowledgements

We thank the VarChAMP group from the IGVF Consortium for providing access to the benchmark dataset. We acknowledge the use of AlphaFold 3 (Google DeepMind) and the Catalogue of Somatic Mutations in Cancer (COSMIC, Wellcome Sanger Institute) under academic licenses. This work was supported by the NIH awards U01HG012022 (P.R.), R01GM145937 (P.R.), and UM1HG011989 (M.V.).

Competing Interests: The authors declare the following competing interests. D.N.C. and M.M. acknowledge Qiagen Inc. for their financial support through a License Agreement with Cardiff University. The remaining authors declare no competing interests.

Code and Data Availability

Source code and trained models are available at `https://github.com/rosstewart/MutPred-PPI`. The processed datasets used for training, evaluation, and predictions are available at `https://doi.org/10.5281/zenodo.17645488`. Provided AlphaFold 3 predictions are subject to AlphaFold Server Output Terms of Use [19]. VarChAMP data were unpublished at the time of submission and hence were excluded. We were unable to deposit COSMIC and HGMD variant-partner interactions due to licensing restrictions; these should be obtained directly from their respective sources.

References

[1] K. Venkatesan, J.-F. Rual, A. Vazquez, *et al.*, "An empirical framework for binary interactome mapping," *Nature Methods*, vol. 6, pp. 83–90, 2009. DOI: `10.1038/nmeth.1280`.

[2] M. P. H. Stumpf, T. Thorne, E. de Silva, *et al.*, "Estimating the size of the human interactome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 19, pp. 6959–6964, 2008. DOI: 10.1073/pnas.0708078105.

[3] J. Lugo-Martinez, D. Zeiberg, T. Gaudelet, N. Malod-Dognin, N. Przulj, and P. Radivojac, "Classification in biological networks with hypergraphlet kernels," *Bioinformatics*, vol. 37, no. 7, pp. 1000–1007, 2021. DOI: 10.1093/bioinformatics/btaa768.

[4] X. Wang, X. Wei, B. Thijssen, J. Das, S. M. Lipkin, and H. Yu, "Three-dimensional reconstruction of protein networks provides insight into human genetic disease," *Nature Biotechnology*, vol. 30, no. 2, pp. 159–164, 2012. DOI: 10.1038/nbt.2106.

[5] Q. Zhong, N. Simonis, Q.-R. Li, *et al.*, "Edgetic perturbation models of human inherited disorders," *Molecular Systems Biology*, vol. 5, p. 321, 2009. DOI: 10.1038/msb.2009.80.

[6] N. Sahni, S. Yi, Q. Zhong, *et al.*, "Edgotype: A fundamental link between genotype and phenotype," *Current Opinion in Genetics & Development*, vol. 23, no. 6, pp. 649–657, 2013. DOI: 10.1016/j.gde.2013.11.002.

[7] S. M.-H. Sy, M. S. Y. Huen, and J. Chen, "PALB2 is an integral component of the BRCA complex required for homologous recombination repair," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 17, pp. 7155–7160, 2009. DOI: 10.1073/pnas.0811159106.

[8] T. K. Foo and B. Xia, "BRCA1-dependent and independent recruitment of PALB2-BRCA2-RAD51 in the DNA damage response and cancer," *Cancer Research*, vol. 82, no. 18, pp. 3191–3197, 2022. DOI: 10.1158/0008-5472.CAN-22-1535.

[9] P. Katsonis, K. Wilhelm, A. Williams, and O. Lichtarge, "Genome interpretation using in silico predictors of variant impact," *Human Genetics*, vol. 141, no. 10, pp. 1549–1577, 2022. DOI: 10.1007/s00439-022-02457-6.

[10] Y.-J. Lin, A. S. Menon, Z. Hu, and S. E. Brenner, "Variant Impact Predictor database (VIPdb), version 2: trends from three decades of genetic variant impact predictors," *Human Genomics*, vol. 18, no. 1, p. 90, 2024. DOI: 10.1186/s40246-024-00663-z.

[11] N. M. Ioannidis, J. H. Rothstein, V. Pejaver, *et al.*, "REVEL: An ensemble method for predicting the pathogenicity of rare missense variants," *American Journal of Human Genetics*, vol. 99, no. 4, pp. 877–885, 2016. DOI: 10.1016/j.ajhg.2016.08.016.

[12] V. Pejaver, J. Urresti, J. Lugo-Martinez, *et al.*, "Inferring the molecular and phenotypic impact of amino acid variants with MutPred2," *Nature Communications*, vol. 11, no. 1, p. 5918, 2020. DOI: 10.1038/s41467-020-19669-x.

[13] J. Cheng, G. Novati, J. Pan, *et al.*, "Accurate proteome-wide missense variant effect prediction with AlphaMissense," *Science*, vol. 381, no. 6664, eadg7492, 2023. DOI: 10.1126/science.adg7492.

[14] R. Rastogi, R. Chung, S. Li, *et al.*, "Critical assessment of missense variant effect predictors on disease-relevant variant data," *Human Genetics*, vol. 144, pp. 281–293, 2025. DOI: 10.1007/s00439-025-02732-2.

[15] T. Bergquist, S. L. Stenton, E. A. Nadeau, *et al.*, "Calibration of additional computational tools expands ClinGen recommendation options for variant classification with PP3/BP4 criteria," *Genetics in Medicine*, vol. 27, no. 6, p. 101 402, 2025, ISSN: 1098-3600. DOI: 10.1016/j.gim.2025.101402.

[16] S. Pahari, G. Li, A. K. Murthy, *et al.*, "SAAMBE-3D: Predicting effect of mutations on protein–protein interactions," *International Journal of Molecular Sciences*, vol. 21, no. 7, p. 2563, 2020. DOI: 10.3390/ijms21072563.

[17] C. H. M. Rodrigues, Y. Myung, D. E. V. Pires, and D. B. Ascher, "mCSM-PPI2: Predicting the effects of mutations on protein–protein interactions," *Nucleic Acids Research*, vol. 47, no. W1, W338–W344, 2019. DOI: 10.1093/nar/gkz383.

[18] J. C. Siwek, A. A. Omelchenko, P. Chhibbar, *et al.*, "Sliding Window Interaction Grammar (SWING): A generalized interaction language model for peptide and protein interactions," *Nature Methods*, vol. 22, pp. 1707–1719, 2025. DOI: 10.1038/s41592-025-02723-1.

[19] J. Abramson, J. Adler, J. Dunger, *et al.*, "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature*, vol. 630, pp. 493–500, 2024. DOI: 10.1038/s41586-024-07487-w.

[20] A. Elnaggar, M. Heinzinger, C. Dallago, *et al.*, "ProtTrans: Toward understanding the language of life through self-supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 7112–7127, 2022. DOI: 10.1109/TPAMI.2021.3095381.

[21] IGVF Consortium, "Deciphering the impact of genomic variation on function," *Nature*, vol. 633, no. 8028, pp. 47–57, 2024. DOI: 10.1038/s41586-024-07510-0.

[22] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=rJXMpikCZ.

[23] C. Yuan, H. Chen, and D. Kihara, "Effective inter-residue contact definitions for accurate protein fold recognition," *BMC Bioinformatics*, vol. 13, p. 292, 2012. DOI: 10.1186/1471-2105-13-292.

[24] G. R. Buel and K. J. Walters, "Can AlphaFold2 predict the impact of missense mutations on structure?" *Nature Structural & Molecular Biology*, vol. 29, no. 1, pp. 1–2, 2022. DOI: 10.1038/s41594-021-00714-2.

[25] A. David, R. Razali, M. N. Wass, and M. J. Sternberg, "Protein–protein interaction sites are hot spots for disease-associated nonsynonymous SNPs," *Human Mutation*, vol. 33, no. 2, pp. 359–363, 2012. DOI: 10.1002/humu.21656.

[26] Y. Bromberg and B. Rost, "Correlating protein function and stability through the analysis of single amino acid substitutions," *BMC Bioinformatics*, vol. 10, no. Suppl 8, S8, 2009. DOI: 10.1186/1471-2105-10-S8-S8.

[27] N. Sahni, S. Yi, M. Taipale, J. I. Fuxman Bass, *et al.*, "Widespread macromolecular interaction perturbations in human genetic disorders," *Cell*, vol. 161, no. 3, pp. 647–660, 2015. DOI: 10.1016/j.cell.2015.04.013.

[28] R. Fragoza, J. Das, S. D. Wierbowski, *et al.*, "Extensive disruption of protein interactions by genetic variants across the allele frequency spectrum in human populations," *Nature Communications*, vol. 10, no. 1, p. 4141, 2019. DOI: 10.1038/s41467-019-11959-3.

[29] L. M. Blaabjerg, M. M. Kassem, L. L. Good, *et al.*, "Rapid protein stability prediction using deep learning representations," *eLife*, vol. 12, e82593, 2023. DOI: 10.7554/eLife.82593.

[30] Y.-R. Park and E. M. Marcotte, "Flaws in evaluation schemes for pair-input computational predictions," *Nature Methods*, vol. 9, pp. 1134–1136, 2012. DOI: 10.1038/nmeth.2259.

[31] L. Fu, B. Niu, Z. Zhu, S. Wu, and W. Li, "CD-HIT: Accelerated for clustering the next-generation sequencing data," *Bioinformatics*, vol. 28, no. 23, pp. 3150–3152, 2012. DOI: 10.1093/bioinformatics/bts565.

[32] M. J. Landrum, J. M. Lee, M. Benson, *et al.*, "ClinVar: Improving access to variant interpretations and supporting evidence," *Nucleic Acids Research*, vol. 46, no. D1, pp. D1062–D1067, 2018. DOI: 10.1093/nar/gkx1153.

[33] J. G. Tate, S. Bamford, H. C. Jubb, *et al.*, "COSMIC: The catalogue of somatic mutations in cancer," *Nucleic Acids Research*, vol. 47, no. D1, pp. D941–D947, 2019. DOI: 10.1093/nar/gky1015.

[34] P. D. Stenson, M. Mort, E. V. Ball, *et al.*, "The Human Gene Mutation Database (HGMD®): Optimizing its use in a clinical diagnostic or research setting," *Human Genetics*, vol. 139, no. 10, pp. 1197–1207, 2020. DOI: 10.1007/s00439-020-02199-3.

[35] S. Chen, L. C. Francioli, J. K. Goodrich, *et al.*, "A genomic mutational constraint map using variation in 76,156 human genomes," *Nature*, vol. 625, no. 7993, pp. 92–100, 2024. DOI: 10.1038/s41586-023-06045-0.

[36] J. M. Fu, F. K. Satterstrom, M. Peng, *et al.*, "Rare coding variation provides insight into the genetic architecture and phenotypic context of autism," *Nature Genetics*, vol. 54, no. 9, pp. 1320–1331, 2022. DOI: 10.1038/s41588-022-01104-0.

[37] R. Oughtred, J. Rust, C. Chang, *et al.*, "The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions," *Protein Science*, vol. 30, no. 1, pp. 187–200, 2021. DOI: 10.1002/pro.3978.

[38] J. D. Stephenson, P. Totoo, D. F. Burke, J. Jänes, P. Beltrao, and M. J. Martin, "ProtVar: Mapping and contextualizing human missense variation," *Nucleic Acids Research*, 2024. DOI: 10.1093/nar/gkae413.

[39] M. Steinegger and J. Söding, "MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets," *Nature Biotechnology*, vol. 35, no. 11, pp. 1026–1028, 2017. DOI: 10.1038/nbt.3988.

[40] Z. Wang and J. Moult, "SNPs, protein structure, and disease," *Human Mutation*, vol. 17, no. 4, pp. 263–270, 2001. DOI: 10.1002/humu.22.

[41] F. Cheng, J. Zhao, Y. Wang, and et al., "Comprehensive characterization of protein–protein interactions perturbed by disease mutations," *Nature Genetics*, vol. 53, no. 3, pp. 342–353, 2021. DOI: 10.1038/s41588-020-00774-y.

[42] L. M. Starita, N. Ahituv, M. J. Dunham, *et al.*, "Variant interpretation: Functional assays to the rescue," *The American Journal of Human Genetics*, vol. 101, no. 3, pp. 315–325, 2017. DOI: 10.1016/j.ajhg.2017.07.014.

[43] M. T. Chang, S. Asthana, S. P. Gao, *et al.*, "Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity," *Nature Biotechnology*, vol. 34, no. 2, pp. 155–163, 2016. DOI: 10.1038/nbt.3391.

[44] M. Sharifi Tabar, H. Francis, D. Yeo, C. G. Bailey, and J. E. Rasko, "Mapping oncogenic protein interactions for precision medicine," *International Journal of Cancer*, vol. 151, no. 1, pp. 7–19, 2022. DOI: 10.1002/ijc.33954.

[45] E. Porta-Pardo, L. Garcia-Alonso, T. Hrabe, J. Dopazo, and A. Godzik, "A pan-cancer catalogue of cancer driver protein interaction interfaces," *PLoS Computational Biology*, vol. 11, no. 10, e1004518, 2015. DOI: 10.1371/journal.pcbi.1004518.

[46] H. Nishi, M. Tyagi, S. Teng, *et al.*, "Cancer missense mutations alter binding properties of proteins and their interaction networks," *PLoS One*, vol. 8, no. 6, e66273, 2013. DOI: 10.1371/journal.pone.0066273.

[47] Y. Hao, B. He, L. Wu, *et al.*, "Nuclear translocation of p85β promotes tumorigenesis of PIK3CA helical domain mutant cancer," *Nature Communications*, vol. 13, no. 1, p. 2058, 2022. DOI: 10.1038/s41467-022-29585-x.

[48] K. Ozturk, R. Panwala, J. Sheen, *et al.*, "Interface-guided phenotyping of coding variants in the transcription factor RUNX1," *Cell Reports*, vol. 43, no. 7, p. 114 436, 2024. DOI: 10.1016/j.celrep.2024.114436.

[49] M. Mort, U. S. Evani, V. G. Krishnan, *et al.*, "In silico functional profiling of human disease-associated and polymorphic amino acid substitutions," *Human Mutation*, vol. 31, no. 3, pp. 335–346, 2010. DOI: 10.1002/humu.21192.

[50] K. J. Karczewski, L. C. Francioli, G. Tiao, *et al.*, "The mutational constraint spectrum quantified from variation in 141,456 humans," *Nature*, vol. 581, no. 7809, pp. 434–443, 2020. DOI: 10.1038/s41586-020-2308-7.

[51] J. A. Tennessen, A. W. Bigham, T. D. O'Connor, *et al.*, "Evolution and functional impact of rare coding variation from deep sequencing of human exomes," *Science*, vol. 337, no. 6090, pp. 64–69, 2012. DOI: 10.1126/science.1219240.

[52] Y. Jiang, J. Urresti, K. A. Pagel, A. B. Pramod, L. M. Iakoucheva, and P. Radivojac, "Prioritizing de novo autism risk variants with calibrated gene- and variant-scoring models," *Human Genetics*, vol. 141, no. 9, pp. 1595–1613, 2022. DOI: 10.1007/s00439-021-02356-2.

[53] F. K. Satterstrom, J. A. Kosmicki, J. Wang, *et al.*, "Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism," *Cell*, vol. 180, no. 3, pp. 568–584, 2020. DOI: 10.1016/j.cell.2019.12.036.

[54] S. Richards, N. Aziz, S. Bale, *et al.*, "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," *Genetics in Medicine*, vol. 17, no. 5, pp. 405–424, 2015. DOI: 10.1038/gim.2015.30.

[55] L. C. Walker, M. de la Hoya, G. A. Wiggins, *et al.*, "Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: Recommendations from the ClinGen SVI Splicing Subgroup," *The American Journal of Human Genetics*, vol. 110, no. 7, pp. 1046–1067, 2023. DOI: 10.1016/j.ajhg.2023.06.002.

[56] J. Gillis and P. Pavlidis, "Bias tradeoffs in the creation and analysis of protein-protein interaction networks," *Journal of Proteomics*, vol. 100, pp. 44–54, 2014. DOI: 10.1016/j.jprot.2014.01.020.

[57] H. Yu, P. M. Kim, E. Sprecher, V. Trifonov, and M. Gerstein, "Broker genes in human disease," *Genome Biology and Evolution*, vol. 2, pp. 815–825, 2010. DOI: 10.1093/gbe/evq064.

[58] M. A. Ghadie, L. Lambourne, M. Vidal, and Y. Xia, "Domain-based prediction of the human isoform interactome provides insights into the functional impact of alternative splicing," *PLoS Computational Biology*, vol. 13, no. 8, e1005717, 2017. DOI: 10.1371/journal.pcbi.1005717.