biology

MDPI

Article

A TRIM Family-Based Strategy for TRIMCIV Target Prediction in a Pan-Cancer Context with Multi-Omics Data and Protein Docking Integration

Yisha Huang †, Jiajia Xuan †, Jiayan Liang, Xixi Liu, Yonglei Luo, Xuejuan Gao * and Wanting Liu *

MOE Key Laboratory of Tumor Molecular Biology and Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes, Institute of Life and Health Engineering, College of Life Science and Technology, Jinan University, Guangzhou 510632, China; yolololohuang@outlook.com (Y.H.); xuanjj@jnu.edu.cn (J.X.); roshell@stu2024.jnu.edu.cn (J.L.); xixiliu@stu2023.jnu.edu.cn (X.L.); yonglei1115@stu.jnu.edu.cn (Y.L.)

* Correspondence: tgaoxj@jnu.edu.cn (X.G.); wtliu@jnu.edu.cn (W.L.)

† These authors contributed equally to this work.

Simple Summary

The E3 TRIM family plays a key role in cancer, but identifying which proteins they interact with is costly and challenging. Current prediction methods struggle because they rely on incomplete data or overlook how these interactions change in diseases like cancer. To solve this issue, we studied the largest TRIM subfamily (CIV), collecting hundreds of their known interactions from past experiments. We observed that CIV proteins consistently correlate with DEGs in cancers, unlike other TRIM members. Using this pattern—along with structural features and cancer-specific data—we built a computational tool called TRIMCIVtargeter to predict new CIV targets. Notably, our approach avoids artificial assumptions and accounts for disease context. This tool provides researchers with a faster, more accurate way to uncover TRIM-related cancer mechanisms, potentially accelerating the discovery of new therapeutic targets. By focusing on real-world biological trends, TRIMCIVtargeter advances our understanding of how these proteins contribute to cancer and offers a framework for studying other protein families in disease.

Abstract

The TRIM CIV subfamily, distinguished by its C-terminal PRY-SPRY domains, constitutes nearly half of the human TRIM family and plays pivotal roles in cancer progression through ubiquitination. Identifying TRIM CIV substrates and interactors has emerged as a critical approach for elucidating tumorigenesis. Current protein–protein interaction (PPI) prediction models face challenges, including an inherent deficiency of negative datasets, biased feature integration, and the absence of a cancer-specific interaction context. To achieve the precise identification of TRIMCIV targets, we developed TRIMCIVtargeter with predictive models that systematically integrates multi-dimensional PPI features—expression differences and correlations in specific cancer, comparable protein-docking scores, and cancer-specific context. Learning from the functional and structural interaction features between 718 experimentally validated TRIM–target pairs, two types of SVM-based binary models were independently trained using proteomic and transcriptomic data. Our models achieved robust prediction performance in cancers utilizing a fair feature space and circumventing hypothetical non-interacting pairs. TRIMCIVtargeter not only provides a cancer-related resource for studying TRIMCIV-mediated regulatory mechanisms but also offers a new perspective for family-specific PPI prediction, holding significant implications

check for updates

Academic Editors: Min Zhao, Ruifeng Hu and Dario Di Silvestre

Received: 27 April 2025
Revised: 9 June 2025
Accepted: 20 June 2025
Published: 22 June 2025

Citation: Huang, Y.; Xuan, J.; Liang, J.; Liu, X.; Luo, Y.; Gao, X.; Liu, W. A TRIM Family-Based Strategy for TRIMCIV Target Prediction in a Pan-Cancer Context with Multi-Omics Data and Protein Docking Integration. *Biology* **2025**, *14*, 742. https://doi.org/10.3390/biology14070742

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

for biomarker discovery and therapeutic targeting in oncology. The online platform of TRIMCIVtargeter is now available.

Keywords: TRIM family; PPI prediction; cancer; multi-omics; machine learning; protein docking

1. Introduction

The tripartite motif (TRIM) family has more than 70 members in humans, and it is classified into 12 subfamilies (CI to UC) on the basis of their domain organization, with the CIV subfamily accounting for nearly half of all members [1]. Most TRIM proteins contain a RING-finger domain at the N-terminus, which confers E3 ubiquitin ligase activity, while the C-terminal domains play crucial roles in recognizing substrate and interactors, collectively referred to as "targets" in this study [2,3]. The TRIM family has been extensively implicated in various cellular processes, particularly in carcinogenesis [4]. Notably, the CIV subfamily, which comprises over half of the TRIM family and is characterized by C-terminal PRY-SPRY domains, plays a key role in regulating cell progression through complex pathways across multiple cancer types [5,6]. The PRY-SPRY domain of the TRIM protein has been reported to serve as a critical component for substrate recognition and interactor binding in various biological pathways. For instance, TRIM21 interacts with GSDMD via its PRY-SPRY domain, stabilizing GSDMD expression in quiescent cells [7]. TRIM15 interacts with TAK1 and inhibits its K63-linked ubiquitination depending on its PRY-SPRY domain. [8] Similarly, the binding activity of TRIM25 also depends on its PRY-SPRY domain [9]. Current research efforts are focused on identifying potential targets of TRIM members through experimental approaches, which are essential for understanding tumorigenesis and discovering biomarkers in specific cancer contexts [2,10]. However, experimentally identifying TRIM-specific targets from the vast pool of cellular proteins is both labor-intensive and costly. Consequently, predictive models for protein–protein interactions (PPIs) became desirable to facilitate the identification of potential TRIM targets.

With the growing availability of biological data, computational approaches have been developed to predict PPIs based on the characteristics of interacting proteins, including sequence homology [11–13], co-expression patterns [14,15], Gene Ontology (GO) similarity [16,17], network topology [16,18], geometric features [19], and the physicochemical properties of binding regions [20,21]. Feature extraction and model construction have become mainstream methods in PPI prediction [22]. However, since only positive PPIs are well-documented in the literature, the lack of verified negative or non-interacting protein pairs poses a significant challenge in model training [23]. It is impractical to experimentally validate a comprehensive set of negative PPIs, and negative interaction databases such as Negatome provide very limited pairs, showing that they are barely available for model training [24]. Thus, researchers often resort to generating negative datasets based on hypotheses. EnPPIpred selected random protein pairs which are not reported in a negative dataset [25]. Zhang et al. generated non-PPI datasets based on low sequence similarity [26]. N Khunlertgit et al. considered the topological characteristic of interaction networks [27]. DPPN-SVM assumed non-interaction based on differential subcellular localization indirectly measured via co-expression and interaction networks [28]. However, these strategies introduce biases and limit model generalizability. For instance, it is widely acknowledged that proteins can translocate between organelles, such as mitochondria and chloroplasts, to participate in dynamic biological pathways [29,30]. Defining non-interacting protein pairs based solely on localization disregards these findings and can introduce significant biases.

Furthermore, biased features in protein–protein interaction increase dataset sparsity, ultimately compromising prediction accuracy. Liang et al. enriched GO terms and network topology in predictive models, which successfully improve the identification of well-studied "hub" proteins but still struggle to predict interactors for less-characterized proteins [16]. Current PPI prediction tools such as protein2vec [31], TANGO [32], and TransformerGO [33], which rely on GO annotation information, all exhibit inherent bias toward hub genes. Consequently, proteins outside of research hotspots often yield inaccurate predictions or cannot be predicted at all, particularly for novel proteins. An additional critical consideration is that protein interactions function within specific biological contexts, including tissue origin and disease phenotypes. However, most existing computational tools for PPI prediction tools, including docking programs, frequently overlook disease-specific factors in protein interactions, limiting their predictive accuracy in disease-related research.

Apart from physical interaction, functional interaction also presented a signal for protein interaction, assuming PPIs that display similar expression pattern are functionally associated [34]. The STRING database summarized predicted functional interactions of various organisms [35]. TRIM proteins mainly regulate cancer development through substrate ubiquitination, and many experiments also validated the co-expression pattern between TRIM members and their targets at the expression level [2]. For example, TRIM25 activated Nrf2 via ubiquitination-mediated Keap1 degradation, positively associated with Nrf2 expression and negatively with Keap1 expression in hepatocellular carcinoma [36]. Similarly, TRIM27 mediates the ubiquitination of TBK1 through SHP2 recruitment, showing a direct co-expression relationship with SHP2 [37]. In glioblastoma, a tissue microarray analysis of glioma samples revealed an inverse relationship between TRIM21 and TIF1γ expression levels, while TRIM21 was positively correlated with β-catenin levels [38]. The TRIM family has been identified as a key player in various cancers through interactions with differentially expressed genes (DEGs), which are widely recognized in tumorigenesis research and have been incorporated into model training such as cancer prediction [39] and overall survival prediction [40]. For example, TRIM59 is upregulated in gastric tumors to promote p53 degradation via ubiquitination [41], while TRIM31 competitively interacts with p53 in breast cancer, leading to its stabilization and activation [42]. In antiviral immune responses, ARRDC4 recruits TRIM65 to promote K63-linked ubiquitination of MDA5 [43]. These findings highlight DEGs as crucial TRIM targets in disease regulation. With the development of high-throughput methods, extensively available multi-omics data has been widely employed to elucidate the functional interaction of PPI in cancers. For example, Chen et al. investigated co-expression patterns between E3 ligases and their interacting substrates using pan-cancer multi-omics data and subsequently incorporated these features into their predictive model [44].

This study introduces a TRIM family-based methodology for protein–protein interaction (PPI) prediction. By systematically compiling and manually curating TRIM–target pairs from the literature, we leveraged the distinct and balanced representation of TRIMCIV and non-CIV subfamilies to develop binary classifiers capable of distinguishing TRIMCIV-specific interactions from the broader TRIM target pool. To construct a robust predictive framework, we integrated multi-omics data—including cancer proteomics and transcriptomics—with computational docking pipelines. An MS-based model leveraging cancer proteomic data and an RNAseq-based model utilizing transcriptomic data was trained and subsequently deployed in TRIMCIVtargeter, an online platform designed to facilitate the discovery of TRIMCIV targets across diverse cancer types. By providing a systematic resource for studying TRIM-mediated regulatory mechanisms in cancer, this work also presents a new avenue for PPI prediction.

2. Materials and Methods

2.1. TRIM–Target Interaction Database Construction

To construct a comprehensive TRIM–target interaction database, we retrieved publications focusing on Homo sapiens from the past decade using Google Scholar with the query keyword "TRIM* + ubiquitination." Targets (substrates and interactors) of TRIM family members were manually curated based on key patterns found in the literature, including "TRIM* mediates/promotes/regulates the ubiquitination/degradation of TARGET", "TRIM* targets TARGET for degradation", "TRIM* interacts with/facilitates TARGET", "TRIM*-TARGET ubiquitin signaling", "TARGET is ubiquitinated and degraded by TRIM*", and "K*-linked ubiquitination by TRIM* stimulates/regulates TARGET". Here, TRIM* denotes a specific TRIM member, and TARGET refers to the corresponding substrate or interactor. The gene names of identified targets were standardized using the uniprot database (https://www.uniprot.org/, accessed on 16 June 2024). As a result, 718 unique TRIM–target pairs with 474 unique targets were manually curated to establish a high-confidence reference dataset.

2.2. Structural Alignment

TM-align is a structural alignment algorithm used to compare 3D protein structures based on their topology. A TM-score > 0.5 is recognized as a reliable indicator of significant structural similarity between proteins, as established by benchmark studies on protein structure alignment [45]. In this study, TM-align was employed to evaluate the structural similarity of C-terminal domains among TRIM family members. A TM-score > 0.5 was considered indicative of structural similarity, with higher scores reflecting greater structural resemblance.

2.3. Expression Data Collection, Processing, and Analysis

Protein expression datasets in specific cancer types were collected from supplementary data in publications. A total of 9 mass spectrometry (MS)-quantified datasets covering 8 cancer types (Samples n = 905) were selected, with each containing over 6000 proteins (Table S2). Due to the limited availability of MS datasets, we supplemented our analysis with transcriptomic data from the UCSC Xena database (TCGA, TARGET, and GTEx projects; https://xenabrowser.net/datapages/, accessed on 21 July 2024), where the study-specific biases of transcriptomic data were mitigated, enabling a comparative analysis between TCGA and GTEx [46]. The RNA profiles for proteins was also recommended as the strategy in testing the reliability of provided interactions [23]. We followed the same strategy recommended in [46] to cope with the insufficiency of normal samples. We obtained a total of 26 RNAseq datasets covering 24 cancer types (Samples n = 11,815; Table S2).

To identify differentially expressed genes (DEGs) at both the protein and RNA levels, gene names were standardized using the UniProt database. For the proteomics dataset, protein expression values were $\log_2$-transformed, and genes with >50% missing values across samples were excluded. Missing values were then imputed using the k-Nearest Neighbors (KNN) algorithm. For the transcriptomic dataset, low-count genes were filtered out using edgeR (v4.0.16), and expression matrices were normalized using the trimmed mean of M-values (TMM) method [47]. Only primary isoforms were retained, and duplicate gene expressions were averaged. DEGs were identified using limma (v3.58.1) with an adjusted p-value threshold of < 0.01 (FDR-corrected). Differential expression cutoffs were set at $|\log FC| > 0.5$ for the MS-based model and $|\log FC| > 1$ for the RNAseq-based model.

2.4. Correlation Analysis

In each cancer dataset, correlations between TRIM members and DEGs were assessed by spearmanr, with a significance threshold of $p < 0.01$. Overall, 17,336 significantly

correlated DEGs and 71 TRIM members (including 36 from the CIV subfamily and 35 from other subfamilies) were selected for physical interaction analysis.

2.5. Docking Pipeline

Given that TRIM subfamilies are classified based on their C-terminal domains, we extracted C-terminal sequences of specific TRIM members using domain information from the PDB database. To evaluate TRIM–target binding affinity, we employed a ZDock-Rosetta-ZRank2 docking pipeline [48–50]. The performance of this docking pipeline has been assessed as a successful docking strategy with acceptable predictions by a community-wide experiment CAPRI project in multiple rounds, including rounds 6–11 [51], rounds 13–19 [52], and rounds 20–26 [53]. Target protein structures with Swiss-Prot-reviewed annotations were retrieved from the AlphaFold database (https://alphafold.ebi.ac.uk/, accessed on 1 August 2024). To prepare for docking, the full-length structures of the target and the C-terminal domains of TRIM members were marked. A small amount of PDB fails (~2%) in docking were attributed to overly large or technical errors. ZDock was run to generate 2000 conformations per TRIM-DEG pair. To include the hydrogen bond, Rosetta was used to add hydrogen atoms to the conformations, followed by ZRank2 scoring to rank out the best conformations. Each TRIM–target pair underwent the same evaluation workflow, receiving an interaction score (zrank_score) supported by the ZRank2 function, where lower scores indicate stronger binding affinity, and we assessed the interacting power including the Van der Waals energy, electrostatics energy, and desolvation. Given the computational demands of protein docking, we preprocessed TRIM-correlated DEGs and ultimately conducted docking for 364,068 TRIM-DEG pairs.

2.6. Predictor Building

To construct predictive models for TRIMCIV target identification, we first filtered reported TRIM–target pairs by selecting those with significant differential expression and correlation across different cancer datasets, as well as those successfully assessed for interaction by docking. The resulting dataset was divided into two groups: TRIMCIV–target pairs (labeled as TRUE) and TRIMelse–target pairs (TRIM member outside CIV subfamily and corresponding targets, labeled as FALSE). Due to differences in expression levels across datasets, we prepared two separate training datasets, respectively, for the MS-based model and the RNAseq-based model. These models aim to classify targets of TRIMCIV members.

To identify an optimal classification algorithm, we evaluated various machine learning models including Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Logistic Regression (LR), and Naïve Bayes (NB) and ultimately selected SVM due to its good performance. The SVM classifiers were fed with the logFC of target, correlation efficiency (R), disease type, and docking scores (zrank_score) between TRIM and target. Since expression data were derived from different disease backgrounds, fold change values for the same gene vary across different cancers. As SVM requires training data to be standardized, all numerical features were standardized before training.

Model training employed the scikit-learn python package (1.5.1), which utilizes the LIBSVM library as its backend [54]. The training datasets, consisting of binary labels (TRUE or FALSE), were mapped into a high-dimensional feature space via a kernel function. The optimal hyperplane was determined to maximize the margin between positive and negative instances. To enhance model performance, the optimal hyperparameter was found using grid search [55]. The final SVM models were implemented using a radial basis function (RBF) kernel, which showed robust sensitivity and specificity during validation. The decision function and kernel function were defined as follows:

$$f(x) = \sum_{i \in SV} y_i \alpha_i K(x_i, x) + b \quad (1)$$

$$K(x_i, x_j) = exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right) = exp(-\gamma|x_i - x_j|^2) \quad (2)$$

In the formula, y_i denotes the class label of the support vector (with TRUE and FALSE mapped to 1 and 0, respectively). The coefficients α_i are obtained from the optimization process, and b represents the bias term. The kernel function $K(x_i, x_j)$ is the RBF kernel, where x_i and x_j are feature vectors. $\gamma = -\frac{1}{2\sigma^2}$ is a hyperparameter that controls the width of the RBF kernel. A larger γ value results in a narrower kernel, emphasizing local data points, whereas a smaller γ value leads to a smoother decision boundary. In this study, the squared Euclidean distance between them is given by

$$|x_i - x_j|^2 = (disease_i - disease_j)^2 + (logFC_i - logFC_i)^2 + (R_i - R_j)^2 + (zrank_score_i - zrank_score_j)^2 \quad (3)$$

In SVM-based model training, the regulation factor, C, was used to control the trade-off between achieving a low training error and a low testing error, effectively regularizing the model to prevent overfitting. In this study, the MS-based model and RNAseq-based model were both trained with regulation factor C = 1 and γ = 1 after grid search for optimal hyperparameters.

Model training was generally conducted in a two-step process using a stratified shuffled fivefold cross-validation scheme [56]. First, a grid search was performed to optimize the hyperparameter of the SVM via fivefold cross-validation. The dataset was reshuffled, and hyperparameters were applied to model training and evaluation using fivefold cross-validation. Finally, the binary classifiers were retrained on the full dataset using the optimal hyperparameters before deployment.

2.7. Evaluation of Model Performance

To assess model performance, we used the Receiver Operating Characteristic (ROC) curve, which evaluates sensitivity (precision, PRE) and specificity (SPE) across different classification thresholds. The classification metric also included accuracy (ACC), recall (REC), F1 score (F1), and Matthews Correlation Coefficient (MCC), defined as follows:

$$\mathrm{PRE} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (4)$$

$$\mathrm{SPE} = 1 - \frac{\mathrm{FP}}{\mathrm{FP+TN}} \quad (5)$$

$$\mathrm{ACC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+FN+TN+FP}} \quad (6)$$

$$\mathrm{REC} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (7)$$

$$\mathrm{F1} = \frac{2\times\mathrm{PRE}\times\mathrm{REC}}{\mathrm{PRE+REC}} \quad (8)$$

$$\mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP+FP})(\mathrm{TP+FN})(\mathrm{TN+FP})(\mathrm{TN+FN})}} \quad (9)$$

where TP and FP represent the numbers of true and false positives, while TN and FN are the numbers of true and false and negatives.

To evaluate model generalization, fivefold cross-validation was performed. The dataset was stratified into five approximately equal subsets, with four subsets used for training and the remaining subset used for testing. This process was repeated five times, ensuring that each subset served as a test set once. The numbers of TPs and FPs were averaged across the five iterations to compute the overall TP/FP ratio, as well as the final sensitivity and specificity values used for ROC analysis.

2.8. Web Interface Building

For the rapid screening of TRIM–target interactions, we developed an interactive web platform using Vite, Vue3, and ExpressJs. The platform integrates preprocessed

datasets of TRIM–differentially expressed gene (DEG) pairs across 26 cancer types, with comprehensive metadata available on the data page. Our MS-based model comprises 79,720 high-confidence TRIM–target pairs, while the RNAseq-based model expands coverage to 1,356,496 putative interactions.

To assist users in selecting the candidates from multiple predicted interactions, we implemented ZRscore, a projection-based ranking metric that integrates both structural and functional evidence. This composite score combines normalized protein docking affinity scores (zrank_score) with co-expression correlation coefficients (R) through a geometric transformation. Each data point $(Zrankscore_{norm_i}, R)$ undergoes normalization where zrank_score values are scaled to a [0, 1] range, denoted as $Zrankscore_{norm_i}$. These normalized docking scores are then paired with their corresponding co-expression values and projected onto a reference line defined by the equation $R = 2Zrankscore_{norm_i}$ within a two-dimensional coordinate space, defined as

$$ZRscore_i = \frac{2Zrankscore_{norm_i} + |R_i|}{\sqrt{5}} \quad (10)$$

A higher ZRscore indicates a more confident candidate. This projection approach, analogous to Principal Component Analysis (PCA), emphasizes the combined contribution of both features while maintaining their relative weights.

3. Results and Discussions

3.1. TRIM Family Overview and Reported TRIM–Target Pair Database Construction

We first constructed the human TRIM family landscape based on domain composition and observed that TRIMCIV constitutes approximately 50% (36/76) of the TRIM family members (Figure 1A). The TRIMCIV group is characterized by a PRY-SPRY domain at the C-terminus, a critical component mediating protein–protein interaction, particularly in immune signaling pathways [1,6,57]. To further confirm the structural similarity of the C-terminus within the TRIM family, we conducted pairwise structural alignments using TM-align in bidirectional comparisons, ensuring a comprehensive and accurate assessment of structural similarity (Table S1). The results demonstrate that TRIMCIV members share highly similar C-terminal domains (overall TM-score > 0.5), distinguishing them from other TRIM groups (Figure 1B). While some TRIM family members (TRIM14, TRIM16, CMYA5, and TRIML2) in the uncategorized group (UC group) also contain the PRY-SPRY domain in their C-terminal region and their functional domain in the UC group has yet to be well characterized, we excluded these members to minimize interference in the identification of TRIMCIV targets in subsequent analyses.

As the framework of this study (Figure 2), we first curated reported substrates and interactors and constructed a TRIM–target interaction database, containing 281 pairs within the TRIMCIV group and 437 pairs in other TRIM groups, with a total of 474 unique targets (Table S1). The TRIM–target pair database is also publicly accessible at https://bioinformaticsscience.cn/trimcivpred/#/trim-ref (accessed on 16 June 2024). These TRIM–target pairs were subsequently analyzed using proteomics and transcriptomic datasets to investigate their differential expression levels and correlations across various cancer types, as well as their physical interactions. These features constituted the dimensions of TRIM–target pairs and were subsequently utilized to train the two models based on the source of the expression data. The trained models were then integrated into the TRIMCIVtargeter public platform for application.

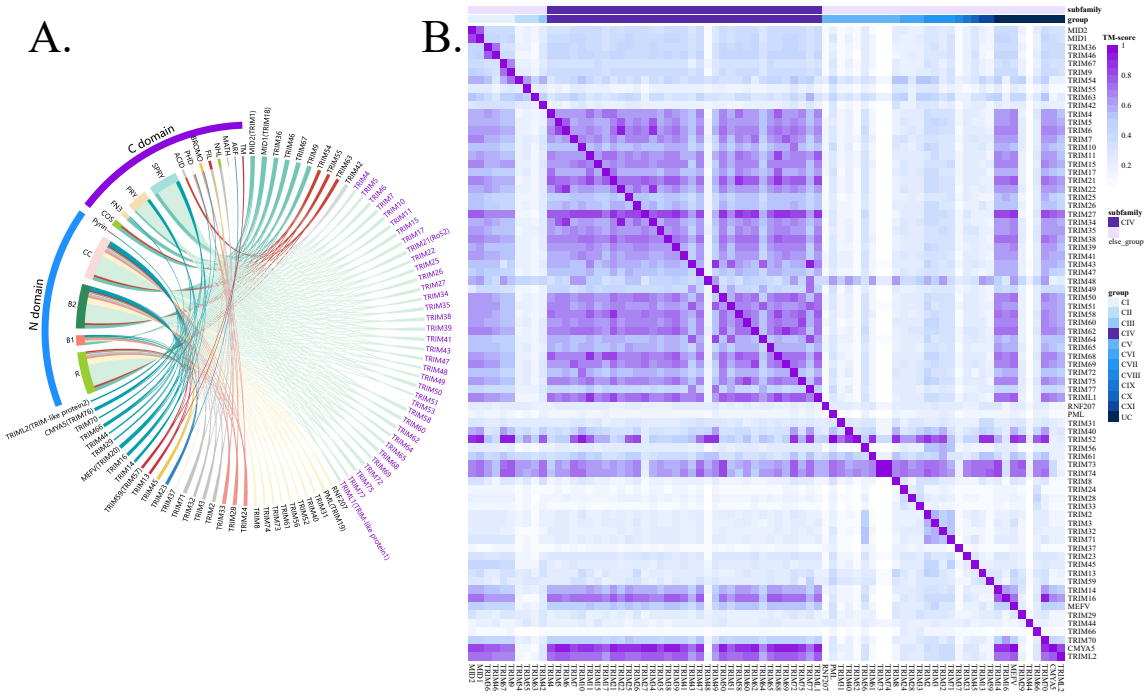

Figure 1. Overview of TRIM family and TRIMCIVtargeter. (**A**) Classification of TRIM family members based on C-terminal domain composition; TRIM CIV members are shown in purple. (**B**) Structural alignment of TRIM C-termini assessed by TM-align, with TM-score > 0.5 indicating significant similarity.

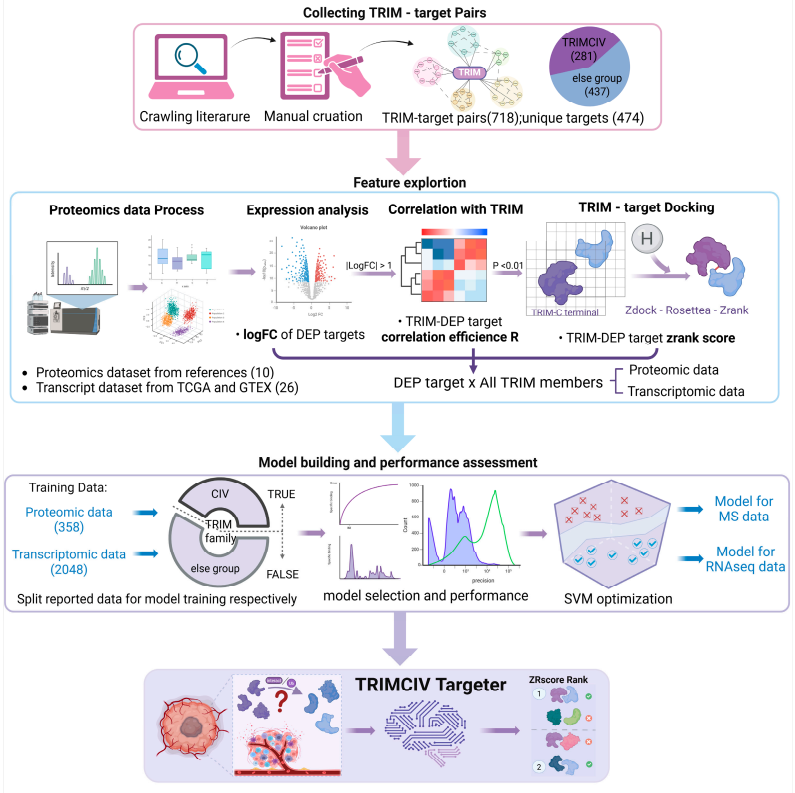

Figure 2. The framework of TRIMCIVtargeter. Reported TRIM targets were manually curated from the literature, and unique TRIM–target pairs were analyzed based on expression correlation and physical interactions via docking. Features including fold change, correlation coefficient (R) in specific cancer types, and interaction assessment (zrank_score) were integrated into the feature space for SVM-based model training. Two independent models were trained using gene expression in proteomics and transcriptomics, ultimately forming the TRIMCIVtargeter platform.

3.2. Distinct Correlation of TRIMCIV with DEG

We analyzed multi-omics expression data across 26 cancer types, including 8 MS datasets from supplementary files of publication and 26 RNAseq datasets from the UCSC database. Differentially expressed genes (DEGs) were identified from each dataset (with an adjusted *p*-value threshold of <0.01 and $|\log FC| > 1$) (Figure 3A; Table S2). Significantly expressed targets were filtered to examine the expression correlation with TRIM proteins across different cancer types (Figure 3B; Table S3). Our analysis of TRIM protein expression correlations across various cancer types revealed distinct patterns, particularly within the TRIMCIV family. As individual TRIM proteins typically interact with different targets, there are rare cases where multiple TRIMs (both CIV and non-CIV members) bind the same protein. Nevertheless, our co-expression analysis demonstrated that TRIMCIV members exhibit consistent correlation trends with DEGs across multiple cancer types. For instance, in the correlation landscape of breast invasive carcinoma and glioblastoma multiforme, TRIMCIV proteins exhibited a positive correlation with upregulated genes and a weaker correlation with downregulated genes (Figure 3A). More details of the co-expression relationship are included in Table S3. While these trends were not equally pronounced across all 35 cancer types analyzed, the recurrent correlations observed in multiple datasets suggest biological relevance among TRIMCIV members, supporting their utility for model construction.

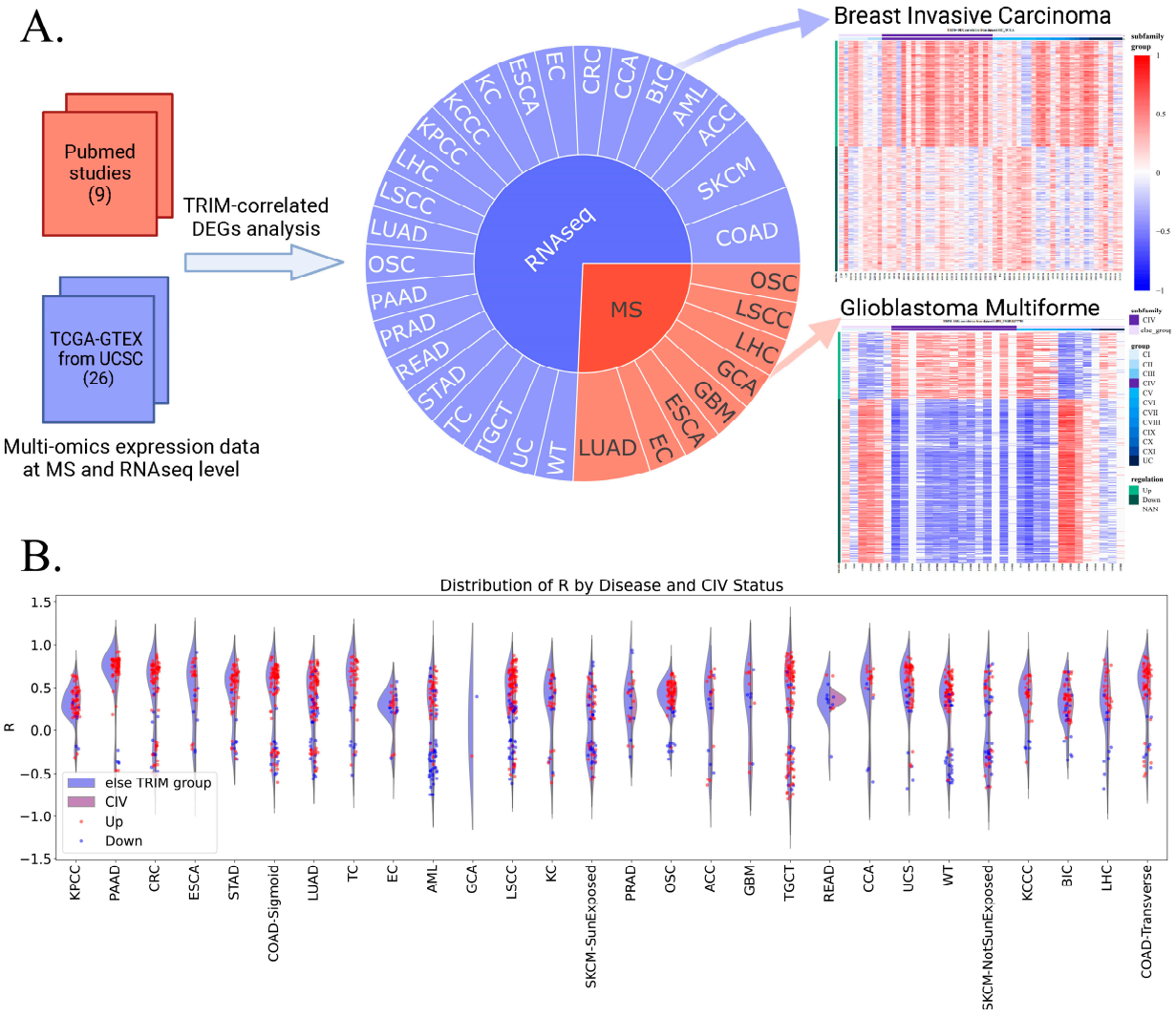

Figure 3. The correlation between TRIM proteins and DEGs across cancers. (**A**) A total of 9 MS datasets and 26 RNAseq datasets covering 26 cancer types were utilized to analyze correlations between TRIM

proteins and differentially expressed targets, revealing distinct clustering patterns within the TRIM-CIV group. (**B**) The R distribution between TRIM and targets across datasets. The cancer types include Adrenocortical Cancer (ACC), Acute Myeloid Leukemia (AML), Breast Invasive Carcinoma (BIC), Cholangiocarcinoma (CCA), Colon Adenocarcinoma (COAD), Colon and Rectal Cancer (CRC), Uterine Corpus Endometrioid Carcinoma (EC), Esophageal Carcinoma (ESCA), Kidney Chromophobe (KC), Kidney Clear Cell Carcinoma (KCCC), Kidney Papillary Cell Carcinoma (KPCC), Liver Hepatocellular Carcinoma (LHC), Lung Squamous Cell Carcinoma (LSCC), Lung Adenocarcinoma (LUAD), Ovarian Serous Cystadenocarcinoma (OSC), Pancreatic Adenocarcinoma (PAAD), Prostate Adenocarcinoma (PRAD), Rectum Adenocarcinoma (READ), Skin Cutaneous Melanoma (SKCM), Stomach Adenocarcinoma (STAD), Thyroid Carcinoma (TC), Testicular Germ Cell Tumor (TGCT), Uterine Carcinosarcoma (UC), Wilms Tumor (WT), Endometrial Carcinoma (EC), Glioblastoma multiforme (GBM), Gastric cancer (GCA), and High-Grade Serous Ovarian Carcinoma (OSC).

3.3. Interaction Scoring via Protein Docking

While the co-expression pattern of TRIM interactions does not necessarily indicate the physical interaction, we carried out protein docking to provide a more reliable measurement of their physical interactions [58]. To evaluate TRIM–target interactions, we filtered 244 reported targets based on significantly differential expression and significant correlation with TRIM proteins in cancers. The C-terminal domains of TRIM members were isolated and docked against their reported targets 2000 times (Figure 4A,B; Table S1). Protein docking was performed using a recommended pipeline integrating ZDOCK for initial docking, Rosetta for hydrogen addition, and ZRank2 for interaction scoring [48]. Although TRIMCIV proteins share a similar PRY-SPRY domain, the limited reported interactions and variability in the C-terminal domain pose challenges in definitively distinguishing TRIMCIV from other TRIM groups (Figure 4C; Figure S1). This docking protocol has been widely applied and assessed as an acceptable strategy by CAPRI [51–53], and its reliability was further examined, with sampled complexes consistently localizing near the PRY-SPRY domain of TRIMCIV proteins, thereby confirming biologically plausible binding poses (Figure S2). Further analysis revealed distinct docking score distributions between TRIM-CIV and non-CIV TRIMs. The CIV group predominantly exhibited interaction scores in the range of -20 to -80, whereas non-CIV TRIMs displayed a more uniform score distribution (Figure 5; Wilcoxon test, ** $p < 0.01$). This suggests that the PRY-SPRY domain of TRIMCIV proteins tends to engage targets with moderate to strong binding power (around -40), though structural variations in interacting partners may also influence binding strength. To ensure comparability, each TRIM–target pair was processed through the same pipeline, enabling the evaluation of different TRIM proteins with the same target or different targets with the same TRIM protein. Here, we utilized docking scores as a quantitative feature for TRIM–target interactions to facilitate further model exploration.

3.4. Evaluation of Prediction Model

Based on our investigation and computational analysis, the reported TRIM–target pairs were characterized using multi-omics cancer data, including target protein logFC values, correlation coefficients (R), disease types, and physical interaction scores (zrank_score). Pairs were filtered using the following thresholds: logFC > 0.5 in proteomics or logFC > 1 in transcriptomics (adjusted $p < 0.01$, FDR), along with significant expression correlation (R, $p < 0.01$). The protein pairs were then divided into TRUE and FALSE datasets for model training and evaluation, employing shuffle cross-validation (Figure 4A). Ultimately, we compiled proteomic data (8 cancer types, $n = 440$) and transcriptomic data points (24 cancer types, $n = 4438$) for predictive modeling (Figure 5B,C). Due to both research bias (with TRIM21, TRIM25, and TRIM28 having more reported substrates and interactors) and our selection criteria (logFC, R, and p-value cutoffs), the TRIM–target pairs were unevenly distributed across TRIM proteins.

However, since TRIM labels and their targets were excluded from the feature space during model training, this imbalance had a minimal impact on model performance.

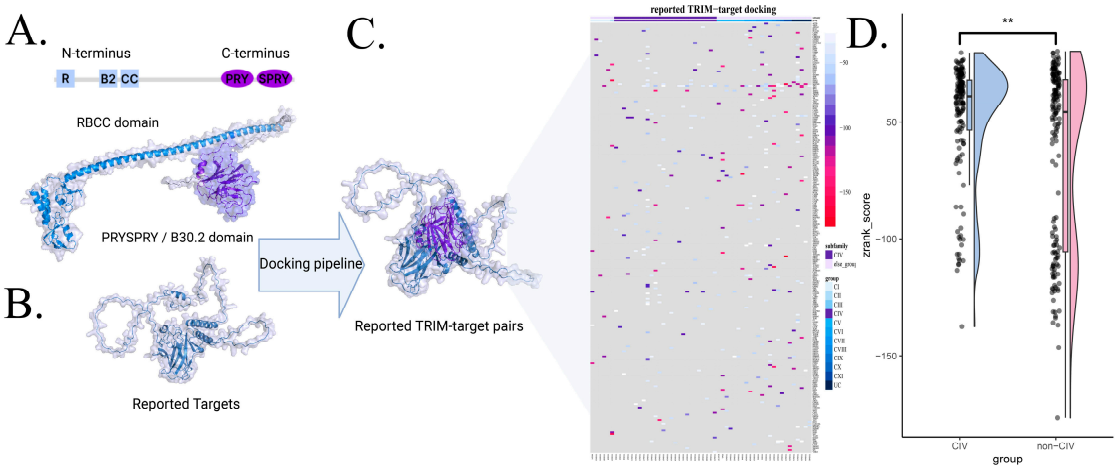

Figure 4. The schematic of docking reported TRIM–target pairs with the following pipelines: zdock was used to find the conformation, and Rosetta was used to add the hydrogen of conformation for ZRank2 to assess the best formation with the interaction score. (**A**) TRIMCIV structural model (e.g., TRIM21, P19474). TRIMCIV proteins exhibit a conserved domain architecture, with an N-terminal RING-finger domain, one or two zinc-finger B-boxes (B1 and B2), and a coiled-coil region, while the C-terminal PRY-SPRY/B20.3 domain is implicated in substrate recognition. (**B**) An example of a target structure (e.g., P53 and P04637). (**C**) Physical interactions of reported TRIM-target pairs assessed using the docking pipeline. Targets were filtered based on significantly differential expression and significant correlation with TRIM members. (**D**) The interacting score distribution of reported TRIM–target pairs. ** $p < 0.01$.

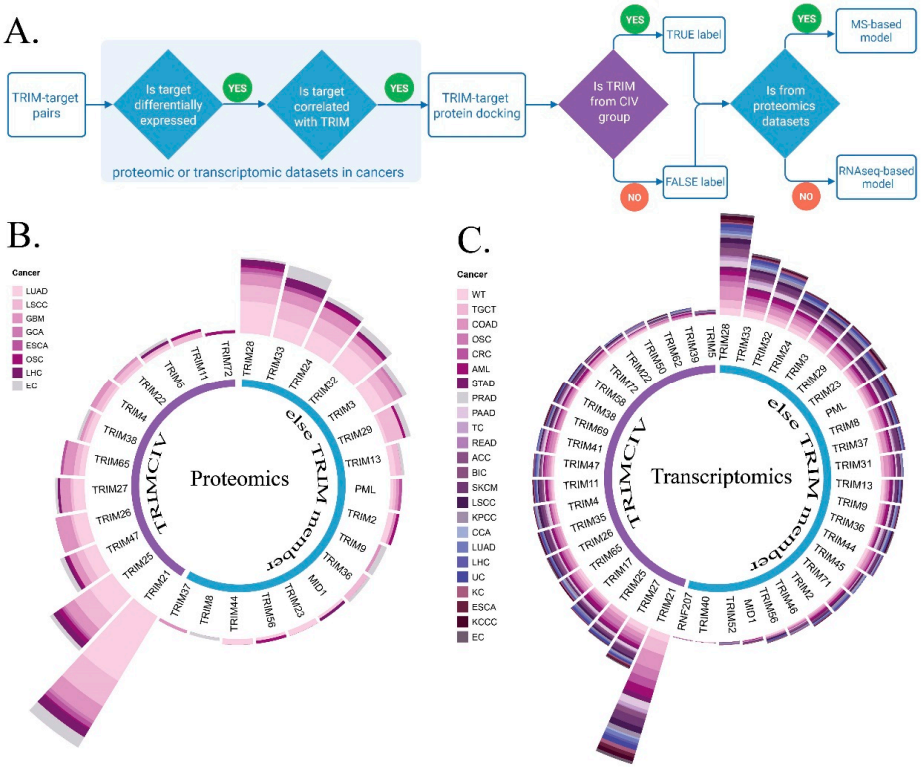

Figure 5. Dataset generation from proteomic and transcriptomic analyses across cancer types. (**A**) Workflow for generating and processing training datasets used to construct proteomics-based (MS) and transcriptomics-based (RNAseq) predictive models. (**B**,**C**) Composition of TRIM–target pairs identified in (**B**) proteomic and (**C**) transcriptomic datasets across different cancer types.

Five candidate machine learning algorithms were selected to develop models predicting TRIMCIV targets in high-dimensional feature spaces constructed from log2-fold change (logFC) values, TRIM-DEG correlation coefficients (R), interaction scores (ZRank_scores), and cancer types by hot coding. In both MS data and RNAseq data, SVM outperformed KNN, LDA, LR, and NB in various metrics evaluated using fivefold cross-validation (Figure 6A,B). Therefore, the SVM was optimized using grid search to construct an MS-based model and an RNAseq-based model. The Receiver Operating Characteristic (ROC) curves of both models achieved robust performance, with an area under the ROC curve (AUC) of 0.77 in the MS-based model and 0.74 in the RNA model (Figure 6C; Table 1). In a comparative analysis, the performance of the MS-based model and the RNAseq-based model appears to be similar. However, the MS-based model exhibits a slight advantage, indicating the better predictive capacity of the MS-based model for TRIMCIV targets when leveraging features derived from proteomics data (Figure 6D). As proteomic data directly reflects protein-level expression and physical interactions, the MS-based model should offer higher confidence in cases where the target proteins are detectable by mass spectrometry.

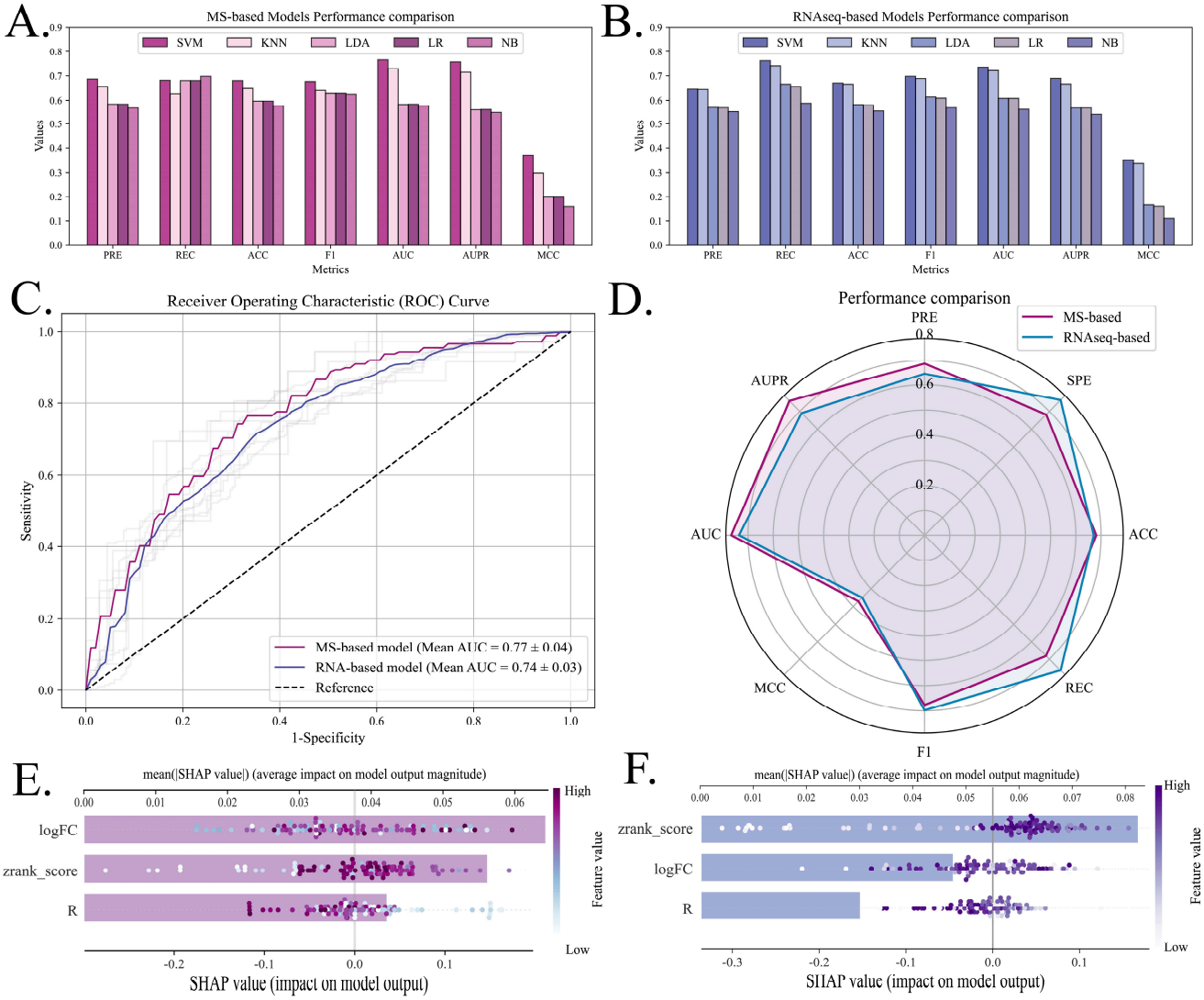

Figure 6. Performance evaluation of MS-based and RNAseq-based predictive models. (**A**,**B**) Performance comparison of five candidate machine learning algorithms (Support Vector Machine, k-Nearest Neighbors, Linear Discriminant Analysis, Logistic Regression, and Naïve Bayes) for (**A**) MS-based and (**B**) RNAseq-based models. (**C**,**D**) Receiver Operating Characteristic (ROC) curves averaged across

five iterations of fivefold cross-validation, with overall performance metrics comparing (**C**) MS-based and (**D**) RNAseq-based models. (**E**,**F**) Combination of feature density scatter plot and important bar plot with SHAP to interpret MS-based model and RNAseq-based model, respectively. Lower abscissa represents SHAP value (lighter dots represent higher eigenvalues and vice versa), while upper *x*-axis represents feature importance scores in development of model.

Table 1. Final models' performance after being trained with optimized hyperparameters and balanced datasets.

Model	PRE	SPE	ACC	REC	F1	MCC	AUC	AUPR
MS-based	0.687 ± 0.051	0.682 ± 0.130	0.681 ± 0.046	0.682 ± 0.130	0.677 ± 0.068	0.371 ± 0.082	0.767 ± 0.046	0.758 ± 0.029
RNAseq-based	0.645 ± 0.020	0.763 ± 0.058	0.671 ± 0.022	0.763 ± 0.058	0.698 ± 0.026	0.350 ± 0.049	0.736 ± 0.032	0.690 ± 0.051

However, considering the limited availability of proteomics data for target features and the potential utility of RNA-level expression data, both models were retained to ensure comprehensive and confident predictions. To further interpret the models, we utilized SHAP to illustrate the important features in the MS-based model and RNAseq-based model, respectively. The results show that logFC, R, and zrank_score were the top three dimensions contributing to the predictive model, which aligns with the biological knowledge underlying the feature selection process, validating the logical coherence of the feature set in capturing TRIM–target interactions (Figure 6E,F).

3.5. Web Server and Utility

To enable efficient access to TRIMCIV target predictions, we developed TRIMCIV-targeter, an interactive online platform (http://bioinformaticsscience.cn/trimcivpred/, accessed on 16 June 2024). Given the computational demands of protein docking, we pre-docked widely expressed TRIMCIV members against 13,091 differentially expressed human genes (at both the protein and RNA levels). The platform integrates preprocessed TRIM-DEG interaction data spanning multiple cancer types, comprising 79,720 pairs from the MS-based model and 1,356,496 pairs from the RNAseq model. Users can query potential interactions by gene name (with UniProt ID auto-completion for Swiss-Prot reviewed structures), select specific TRIMCIV members and cancer types, and retrieve candidate pairs predicted by either the MS-based model, RNAseq-based model, or both (Figure 7A). We also provided the TRIM family overview, landscape of pan-cancer datasets and manually curated reported TRIM-target information (Figure 7B). The results include detailed features for each TRIM–target pair, along with a ZRscore for candidate prioritization based on integrated R and zrank_score metrics (Figure 7C). A higher ZRscore indicates a more confident result. Other columns are also sortable for users to rank the candidate by their preferred features. While Gene Ontology (GO) terms were excluded from model training to avoid bias, the platform provides GO term overlap analysis between TRIM proteins and candidate targets as supplementary biological context (Figure 7C).

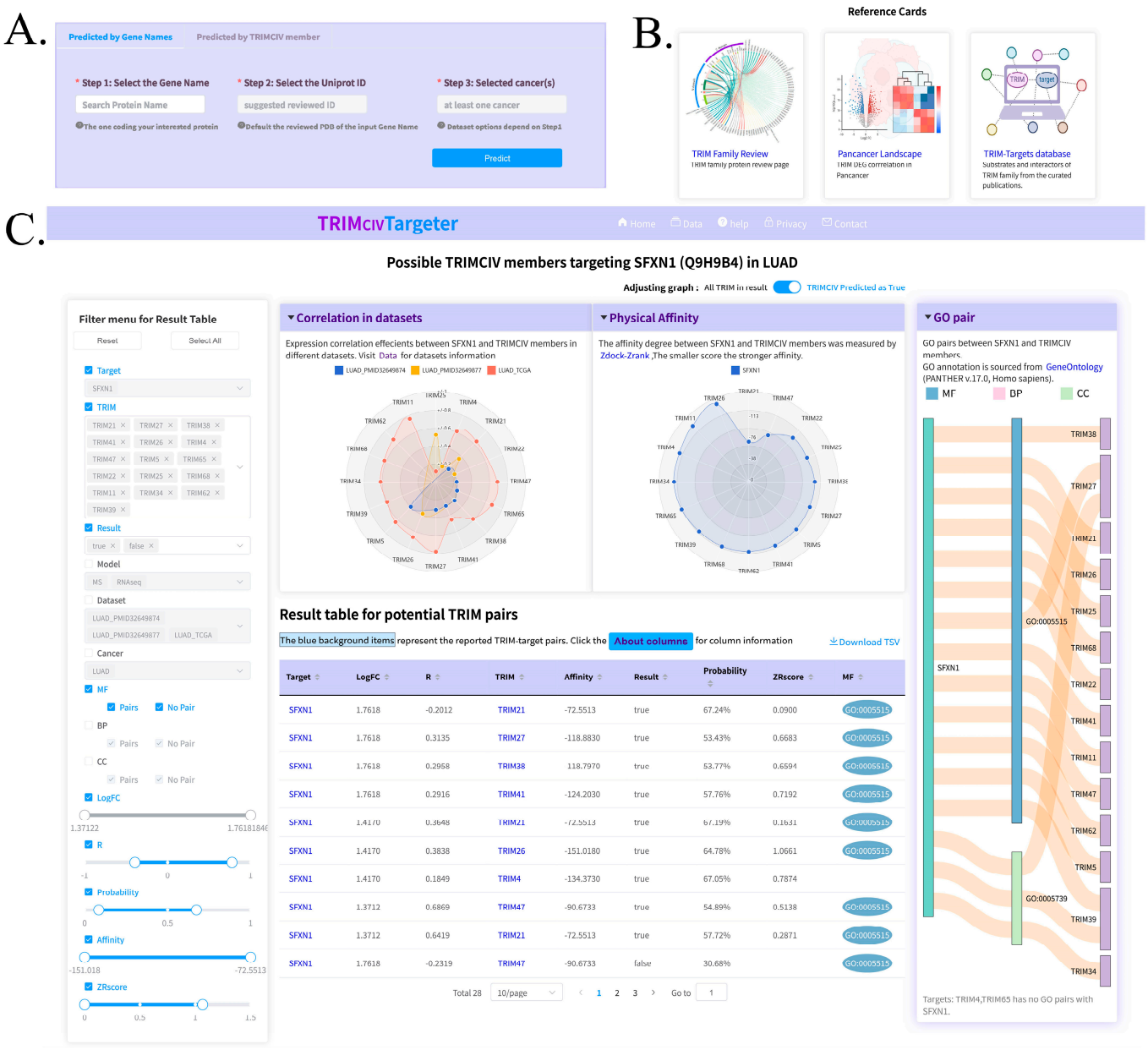

Figure 7. Overview of TRIMCIVtargeter platform. (**A**,**B**) Query interface and reference information of TRIMCIVtargeter designed to continuously integrate and learn from experimental validation data. * represents required input. (**C**) Result page displaying predicted TRIMCIV–target pairs and GO term associations between TRIM proteins and candidate targets.

4. Conclusions

This study systematically identified TRIM family targets by comprehensively analyzing reported interactions. Current PPI prediction models face inherent limitations, including the inherent scarcity of non-interaction data, feature selection biases, and a lack of disease-specific contextualization. To overcome these challenges, we developed a targeted computational approach focusing on specific subfamilies within the broader TRIM family framework. Our comprehensive characterization revealed that the CIV subfamily represents both a biologically and computationally optimal choice for several reasons: First, it constitutes the largest TRIM subgroup (36/76 members in humans) and contains a distinctive, conserved PRY-SPRY domain at the C-terminus—a well-documented protein interaction module that mediates target recognition. Second, our systematic literature curation showed that CIV members account for over one-third of all experimentally val-

idated TRIM–target interactions (281/718; Figure 2), providing substantial training data while maintaining balanced class distributions critical for machine learning applications. The biological relevance of the CIV subfamily-focused approach was further supported by multi-omics analyses. The integration of cancer transcriptomics and proteomics data revealed generally consistent TRIM-DEG correlation patterns among CIV members across multiple cancer types, indicating the relevance of biological function among the CIV subfamily. These observations align with the established structure–function relationships in protein interaction networks. As functional correlation does not necessarily indicate direct interaction, docking programs were implemented to further measure the physical interaction of TRIM–target pairs.

To construct a model for identifying the target of the CIV subfamily, we integrated multiple layers of biological evidence into a unified feature space, including cancer-specific fold changes in targets, expression correlations, and structural interaction measured by docking pipelines. According to different cancer data sources, we developed an MS-based model and an RNAseq-based model with particular attention to incorporating carefully validated true negative TRIMCIV–target pairs. Model evaluation confirmed confident predictive accuracy in identifying bona fide TRIMCIV targets. To bridge computational predictions with experimental research, we implemented these models in TRIMCIVtargeter, an interactive online platform that prioritizes high-confidence targets for experimental validation. This resource empowers researchers to efficiently explore TRIMCIV interactors and substrates implicated in cancer pathways. As multi-omics datasets and validated TRIMCIV–target interactions continue to expand, we anticipate further refinement of prediction accuracy through iterative model optimization.

Notably, rather than relying on hypothetical negative PPIs, TRIMCIVtargeter utilizes carefully curated positive and negative datasets derived from experimentally validated TRIM–target interactions in cancer. This approach ensures a balanced feature space while overcoming the common challenges of data sparsity and feature bias typically associated with hub proteins. However, this study also faces limitations. The model's performance may be constrained by the currently limited TRIM family PPI data and proteomics coverage. While our transcriptomic analysis included 24 cancer types, MS-based proteomic modeling was restricted to 8 types due to data availability. The next version should include improvements by expanding omics datasets, developing more robust feature engineering approaches, and implementing advanced model architectures to better capture subfamily interaction differences and enhance predictive performance. Furthermore, as the first dedicated PPI predictor for a specific protein family, TRIMCIVtargeter demonstrates robust performance while occupying a unique methodological niche, which also presents challenges for this work to compare with existing models.

In this study, the reported targets of TRIM family proteins contain two types: canonical substrates undergoing ubiquitination (leading to either degradative or non-degradative outcomes) and binding partners involved in regulatory mechanisms. For example, TRIM21 interacts with PRMT5 (protein arginine methyltransferase 5), modulating TXNIP/p21 expression without inducing PRMT5 degradation [59]. The N protein of SARS-CoV-2 promoting TRIM25 interacts with G3BP2 (GTPase-activating protein SH3 domain–binding protein 2), inhibiting type I interferon production in the process of infection without involving the ubiquitination of TRIM25 [60]. As the first type of TRIM interaction only accounts for nearly half of the reported cases, the comprehensive prediction of TRIM-mediated PTM effects is particularly challenging. Furthermore, most existing studies focus on establishing protein–protein interaction networks in disease contexts rather than identifying precise modification sites—crucial information for understanding PTM hotspots. For instance, TRIM25-mediated ubiquitination accelerates RBPJ degradation via proteasome in

bladder cancer validated by immunoprecipitation without further studying the binding residues [61]. In vitro and in vivo assays demonstrated that whereas TRIM23 ubiquitinated ANO1, leading to its stabilization, TRIM21 ubiquitinated ANO1 and induced its degradation [62], yet there is no detailed PTM information for reference. As research progresses, we anticipate the accumulation of more comprehensive PTM data for TRIM E3 ligases to enable more accurate predictions. Although our current model predicts interactions at the binding level, it provides researchers with preliminary interaction information for the TRIMCIV subfamily to guide subsequent experimental investigations.

In conclusion, this study introduced a new methodology for PPI prediction in the TRIM family and developed the TRIMCIVtargeter platform for predicting potential TRIMCIV–target interactions. This work provides valuable insights for investigating cancer-specific protein interactions and demonstrates the potential of family-specific modeling to overcome limitations in conventional PPI prediction methods. This platform provides an entry for the practical use of models and valuable information to assist users in experimental research.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biology14070742/s1. Tables can be download at https://zenodo.org/records/14967126 (accessed on 16 June 2024). Figure S1: The TRIM–target pairs from reported interactions; Figure S2: Examples of TRIMCIV-target pair at different scores; Table S1: C-domain split information of TRIM, TMalign result of TRIM C domain, and reported TRIM–target pairs; Table S2: Source of multi-omics data and differential expressed genes across cancer types; Table S3: DEG-TRIM family-wide correlation across cancer types.

Author Contributions: Conceptualization, X.G. and W.L.; Methodology, Y.H.; Software, Y.H.; Validation, Y.H.; Formal Analysis, Y.H.; Investigation, Y.H., J.L., X.L., and Y.L.; Resources, Y.H., J.L., X.L., and X.G.; Data Curation, Y.H. and J.L.; Writing—Original Draft, Y.H.; Writing—Review and Editing, J.X. and W.L.; Visualization, Y.H.; Supervision, J.X. and W.L.; Project Administration, Y.H.; Funding Acquisition, X.G. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by "the Guangdong Basic and Applied Basic Research Foundation (2023A1515011221)", "the Fundamental Research Funds for the Central Universities" (21623335) and "Science and Technology Projects in Guangzhou" (2024A04J4125).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: TRIMCIVtargter is publicly available on the data page and reference cards of the following website https://bioinformaticsscience.cn/trimcivpred/ (accessed on 16 June 2024). The framework of the platform and model training process are available at https://github.com/yolololo-huang/TRIMCIVtargeter.git (accessed on 16 June 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Esposito, D.; Koliopoulos, M.G.; Rittinger, K. Structural determinants of TRIM protein function. *Biochem. Soc. Trans.* **2017**, *45*, 183–191. [CrossRef] [PubMed]
2. Hatakeyama, S. TRIM family proteins: Roles in autophagy, immunity, and carcinogenesis. *Trends Biochem. Sci.* **2017**, *42*, 297–311. [CrossRef] [PubMed]
3. Wang, H.-T.; Hur, S. Substrate recognition by TRIM and TRIM-like proteins in innate immunity. *Semin. Cell Dev. Biol.* **2021**, *111*, 76–85. [CrossRef]
4. Zhao, G.; Liu, C.; Wen, X.; Luan, G.; Xie, L.; Guo, X. The translational values of TRIM family in pan-cancers: From functions and mechanisms to clinics. *Pharmacol. Ther.* **2021**, *227*, 107881. [CrossRef]
5. Perfetto, L.; Gherardini, P.F.; Davey, N.E.; Diella, F.; Helmer-Citterich, M.; Cesareni, G. Exploring the diversity of SPRY/B30.2-mediated interactions. *Trends Biochem. Sci.* **2013**, *38*, 38–46. [CrossRef]

6. James, L.C.; Keeble, A.H.; Khan, Z.; Rhodes, D.A.; Trowsdale, J. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6200–6205. [CrossRef]
7. Gao, W.; Li, Y.; Liu, X.; Wang, S.; Mei, P.; Chen, Z.; Liu, K.; Li, S.; Xu, X.-W.; Gan, J.; et al. TRIM21 regulates pyroptotic cell death by promoting Gasdermin D oligomerization. *Cell Death Differ.* **2022**, *29*, 439–450. [CrossRef] [PubMed]
8. Roy, M.; Singh, K.; Shinde, A.; Singh, J.; Mane, M.; Bedekar, S.; Tailor, Y.; Gohel, D.; Vasiyani, H.; Currim, F.; et al. TNF-α-induced E3 ligase, TRIM15 inhibits TNF-α-regulated NF-κB pathway by promoting turnover of K63 linked ubiquitination of TAK1. *Cell. Signal.* **2022**, *91*, 110210. [CrossRef]
9. Choudhury, N.R.; Heikel, G.; Trubitsyna, M.; Kubik, P.; Nowak, J.S.; Webb, S.; Granneman, S.; Spanos, C.; Rappsilber, J.; Castello, A.; et al. RNA-binding activity of TRIM25 is mediated by its PRY/SPRY domain and is required for ubiquitination. *BMC Biol.* **2017**, *15*, 105. [CrossRef]
10. Huang, N.; Sun, X.; Li, P.; Liu, X.; Zhang, X.; Chen, Q.; Xin, H. TRIM family contribute to tumorigenesis, cancer development, and drug resistance. *Exp. Hematol. Oncol.* **2022**, *11*, 75. [CrossRef]
11. Guo, Y.; Yu, L.; Wen, Z.; Li, M. Using support vector machine combined with auto covariance to predict protein–protein interactions from protein sequences. *Nucleic Acids Res.* **2008**, *36*, 3025–3030. [CrossRef] [PubMed]
12. Zhang, C.; Freddolino, P.L.; Zhang, Y. COFACTOR: Improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic Acids Res.* **2017**, *45*, W291–W299. [CrossRef]
13. Hashemifar, S.; Neyshabur, B.; Khan, A.A.; Xu, J. Predicting protein–protein interactions through sequence-based deep learning. *Bioinformatics* **2018**, *34*, i802–i810. [CrossRef] [PubMed]
14. Zhang, Q.C.; Petrey, D.; Deng, L.; Qiang, L.; Shi, Y.; Thu, C.A.; Bisikirska, B.; Lefebvre, C.; Accili, D.; Hunter, T.; et al. Structure-based prediction of protein–protein interactions on a genome-wide scale. *Nature* **2012**, *490*, 556–560. [CrossRef] [PubMed]
15. De Bodt, S.; Proost, S.; Vandepoele, K.; Rouzé, P.; Van de Peer, Y. Predicting protein-protein interactions in Arabidopsis thaliana through integration of orthology, gene ontology and co-expression. *BMC Genom.* **2009**, *10*, 288. [CrossRef]
16. Li, Y.; Xie, P.; Lu, L.; Wang, J.; Diao, L.; Liu, Z.; Guo, F.; He, Y.; Liu, Y.; Huang, Q.; et al. An integrated bioinformatics platform for investigating the human E3 ubiquitin ligase-substrate interaction network. *Nat. Commun.* **2017**, *8*, 347. [CrossRef]
17. Zhang, Y.-H.; Huang, F.; Li, J.; Shen, W.; Chen, L.; Feng, K.; Huang, T.; Cai, Y.-D. Identification of Protein–Protein Interaction Associated Functions Based on Gene Ontology. *Protein J.* **2024**, *43*, 477–486. [CrossRef]
18. Lei, C.; Ruan, J. A novel link prediction algorithm for reconstructing protein–protein interaction networks by topological similarity. *Bioinformatics* **2013**, *29*, 355–364. [CrossRef]
19. Meyer, M.J.; Beltrán, J.F.; Liang, S.; Fragoza, R.; Rumack, A.; Liang, J.; Wei, X.; Yu, H. Interactome INSIDER: A structural interactome browser for genomic studies. *Nat. Methods* **2018**, *15*, 107–114. [CrossRef]
20. Jiménez, J.; Doerr, S.; Martínez-Rosell, G.; Rose, A.S.; De Fabritiis, G. DeepSite: Protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics* **2017**, *33*, 3036–3042. [CrossRef]
21. Wass, M.N.; Fuentes, G.; Pons, C.; Pazos, F.; Valencia, A. Towards the prediction of protein interaction partners using physical docking. *Mol. Syst. Biol.* **2011**, *7*, 469. [CrossRef] [PubMed]
22. Durham, J.; Zhang, J.; Humphreys, I.R.; Pei, J.; Cong, Q. Recent advances in predicting and modeling protein–protein interactions. *Trends Biochem. Sci.* **2023**, *48*, 527–538. [CrossRef] [PubMed]
23. Tang, T.; Zhang, X.; Liu, Y.; Peng, H.; Zheng, B.; Yin, Y.; Zeng, X. Machine learning on protein–protein interaction prediction: Models, challenges and trends. *Brief. Bioinform.* **2023**, *24*, bbad076. [CrossRef]
24. Blohm, P.; Frishman, G.; Smialowski, P.; Goebels, F.; Wachinger, B.; Ruepp, A.; Frishman, D. Negatome 2.0: A database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Res.* **2014**, *42*, D396–D400. [CrossRef]
25. Barman, R.K.; Jana, T.; Das, S.; Saha, S. Prediction of intra-species protein-protein interactions in enteropathogens facilitating systems biology study. *PLoS ONE* **2015**, *10*, e0145648. [CrossRef]
26. Zhang, L.; Yu, G.; Guo, M.; Wang, J. Predicting protein-protein interactions using high-quality non-interacting pairs. *BMC Bioinform.* **2018**, *19*, 105–124. [CrossRef]
27. Khunlertgit, N.; Yoon, B.-J. Incorporating topological information for predicting robust cancer subnetwork markers in human protein-protein interaction network. *BMC Bioinform.* **2016**, *17*, 143–152. [CrossRef] [PubMed]
28. Li, G.-P.; Du, P.-F.; Shen, Z.-A.; Liu, H.-Y.; Luo, T. DPPN-SVM: Computational identification of mis-localized proteins in cancers by integrating differential gene expressions with dynamic protein-protein interaction networks. *Front. Genet.* **2020**, *11*, 600454. [CrossRef]
29. Busch, J.D.; Fielden, L.F.; Pfanner, N.; Wiedemann, N. Mitochondrial protein transport: Versatility of translocases and mechanisms. *Mol. Cell* **2023**, *83*, 890–910. [CrossRef]
30. Li, H.-m.; Chiu, C.-C. Protein transport into chloroplasts. *Annu. Rev. Plant Biol.* **2010**, *61*, 157–180. [CrossRef]
31. Zhang, J.; Zhu, M.; Qian, Y. protein2vec: Predicting protein-protein interactions based on LSTM. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *19*, 1257–1266. [CrossRef]

32. Wang, H.; Zheng, H.; Chen, D.Z. TANGO: A GO-term embedding based method for protein semantic similarity prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *20*, 694–706. [CrossRef] [PubMed]
33. Ieremie, I.; Ewing, R.M.; Niranjan, M. TransformerGO: Predicting protein–protein interactions by modelling the attention between sets of gene ontology terms. *Bioinformatics* **2022**, *38*, 2269–2277. [CrossRef] [PubMed]
34. van Noort, V.; Snel, B.; Huynen, M.A. Predicting gene function by conserved co-expression. *TRENDS Genet.* **2003**, *19*, 238–242. [CrossRef] [PubMed]
35. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S.; et al. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646. [CrossRef]
36. Liu, Y.; Tao, S.; Liao, L.; Li, Y.; Li, H.; Li, Z.; Lin, L.; Wan, X.; Yang, X.; Chen, L. TRIM25 promotes the cell survival and growth of hepatocellular carcinoma through targeting Keap1-Nrf2 pathway. *Nat. Commun.* **2020**, *11*, 348. [CrossRef]
37. Zheng, Q.; Hou, J.; Zhou, Y.; Yang, Y.; Xie, B.; Cao, X. Siglec1 suppresses antiviral innate immune response by inducing TBK1 degradation via the ubiquitin ligase TRIM27. *Cell Res.* **2015**, *25*, 1121–1136. [CrossRef]
38. Li, Y.; Bao, L.; Zheng, H.; Geng, M.; Chen, T.; Dai, X.; Xiao, H.; Yang, L.; Mao, C.; Qiu, Y.; et al. E3 ubiquitin ligase TRIM21 targets TIF1γ to regulate β-catenin signaling in glioblastoma. *Theranostics* **2023**, *13*, 4919. [CrossRef]
39. Xiao, Y.; Wu, J.; Lin, Z.; Zhao, X. A deep learning-based multi-model ensemble method for cancer prediction. *Comput. Methods Programs Biomed.* **2018**, *153*, 1–9. [CrossRef]
40. Liang, J.-y.; Wang, D.-s.; Lin, H.-c.; Chen, X.-x.; Yang, H.; Zheng, Y.; Li, Y.-h. A novel ferroptosis-related gene signature for overall survival prediction in patients with hepatocellular carcinoma. *Int. J. Biol. Sci.* **2020**, *16*, 2430. [CrossRef]
41. Zhou, Z.; Ji, Z.; Wang, Y.; Li, J.; Cao, H.; Zhu, H.H.; Gao, W.-Q. TRIM59 is up-regulated in gastric tumors, promoting ubiquitination and degradation of p53. *Gastroenterology* **2014**, *147*, 1043–1054. [CrossRef]
42. Guo, Y.; Li, Q.; Zhao, G.; Zhang, J.; Yuan, H.; Feng, T.; Ou, D.; Gu, R.; Li, S.; Li, K. Loss of TRIM31 promotes breast cancer progression through regulating K48-and K63-linked ubiquitination of p53. *Cell Death Dis.* **2021**, *12*, 945. [CrossRef]
43. Meng, J.; Yao, Z.; He, Y.; Zhang, R.; Zhang, Y.; Yao, X.; Yang, H.; Chen, L.; Zhang, Z.; Zhang, H.; et al. ARRDC4 regulates enterovirus 71-induced innate immune response by promoting K63 polyubiquitination of MDA5 through TRIM65. *Cell Death Dis.* **2017**, *8*, e2866. [CrossRef]
44. Chen, D.; Liu, X.; Xia, T.; Tekcham, D.S.; Wang, W.; Chen, H.; Li, T.; Lu, C.; Ning, Z.; Liu, X.; et al. A multidimensional characterization of E3 ubiquitin ligase and substrate interaction network. *IScience* **2019**, *16*, 177–191. [CrossRef] [PubMed]
45. Zhang, Y.; Skolnick, J. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **2005**, *33*, 2302–2309. [CrossRef] [PubMed]
46. Wang, Q.; Armenia, J.; Zhang, C.; Penson, A.V.; Reznik, E.; Zhang, L.; Minet, T.; Ochoa, A.; Gross, B.E.; Iacobuzio-Donahue, C.A.; et al. Unifying cancer and normal RNA sequencing data from different sources. *Sci. Data* **2018**, *5*, 180061. [CrossRef] [PubMed]
47. Robinson, M.D.; Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **2010**, *11*, R25. [CrossRef]
48. Pierce, B.G.; Wiehe, K.; Hwang, H.; Kim, B.-H.; Vreven, T.; Weng, Z. ZDOCK server: Interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics* **2014**, *30*, 1771–1773. [CrossRef]
49. O'Meara, M.J.; Leaver-Fay, A.; Tyka, M.D.; Stein, A.; Houlihan, K.; DiMaio, F.; Bradley, P.; Kortemme, T.; Baker, D.; Snoeyink, J.; et al. Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J. Chem. Theory Comput.* **2015**, *11*, 609–622. [CrossRef]
50. Pierce, B.; Weng, Z. ZRANK: Reranking protein docking predictions with an optimized energy function. *Proteins Struct. Funct. Bioinform.* **2007**, *67*, 1078–1086. [CrossRef]
51. Wiehe, K.; Pierce, B.; Tong, W.W.; Hwang, H.; Mintseris, J.; Weng, Z. The performance of ZDOCK and ZRANK in rounds 6–11 of CAPRI. *Proteins Struct. Funct. Bioinform.* **2007**, *69*, 719–725. [CrossRef] [PubMed]
52. Hwang, H.; Vreven, T.; Pierce, B.G.; Hung, J.H.; Weng, Z. Performance of ZDOCK and ZRANK in CAPRI rounds 13–19. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 3104–3110. [CrossRef] [PubMed]
53. Vreven, T.; Pierce, B.G.; Hwang, H.; Weng, Z. Performance of ZDOCK in CAPRI rounds 20–26. *Proteins Struct. Funct. Bioinform.* **2013**, *81*, 2175–2182. [CrossRef] [PubMed]
54. Chang, C.-C.; Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 1–27. [CrossRef]
55. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [CrossRef]
56. Bradshaw, T.J.; Huemann, Z.; Hu, J.; Rahmim, A. A guide to cross-validation for artificial intelligence in medical imaging. *Radiol. Artif. Intell.* **2023**, *5*, e220232. [CrossRef]
57. Huang, Y.; Gao, X.; He, Q.Y.; Liu, W. A interacting model: How TRIM21 orchestrates with proteins in intracellular immunity. *Small Methods* **2024**, *8*, 2301142. [CrossRef]

58. Gromiha, M.M.; Yugandhar, K.; Jemimah, S. Protein–protein interactions: Scoring schemes and binding affinity. *Curr. Opin. Struct. Biol.* **2017**, *44*, 31–38. [CrossRef]
59. Li, Y.-H.; Tong, K.-L.; Lu, J.-L.; Lin, J.-B.; Li, Z.-Y.; Sang, Y.; Ghodbane, A.; Gao, X.-J.; Tam, M.-S.; Hu, C.-D.; et al. PRMT5-TRIM21 interaction regulates the senescence of osteosarcoma cells by targeting the TXNIP/p21 axis. *Aging* **2020**, *12*, 2507. [CrossRef]
60. Yang, Z.; Li, J.; Li, J.; Zheng, H.; Li, H.; Lai, Q.; Chen, Y.; Qin, L.; Zuo, Y.; Guo, L.; et al. Engagement of the G3BP2-TRIM25 Interaction by Nucleocapsid Protein Suppresses the Type I Interferon Response in SARS-CoV-2-Infected Cells. *Vaccines* **2022**, *10*, 2042. [CrossRef]
61. Tang, H.; Li, X.; Jiang, L.; Liu, Z.; Chen, L.; Chen, J.; Deng, M.; Zhou, F.; Zheng, X.; Liu, Z. RITA1 drives the growth of bladder cancer cells by recruiting TRIM25 to facilitate the proteasomal degradation of RBPJ. *Cancer Sci.* **2022**, *113*, 3071–3084. [CrossRef] [PubMed]
62. Cao, X.; Zhou, Z.; Tian, Y.; Liu, Z.; Cheng, K.O.; Chen, X.; Hu, W.; Wong, Y.M.; Li, X.; Zhang, H.; et al. Opposing roles of E3 ligases TRIM23 and TRIM21 in regulation of ion channel ANO1 protein levels. *J. Biol. Chem.* **2021**, *296*, 100738. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.