



# OPEN Network pharmacology and AI in cancer research uncovering biomarkers and therapeutic targets for RALGDS mutations

S. Mohammed Zaidh<sup>1,2</sup>, Hariharan Thirumalai Vengateswaran<sup>1</sup>, Mohammad Habeeb<sup>1</sup>, Kiran Balasaheb Aher<sup>3</sup>, Girija Balasaheb Bhavar<sup>4</sup>, N. Irfan<sup>1✉</sup> & K NV Chenchu Lakshmi<sup>5</sup>

The lack of target therapies is accountable for the higher mortality of various types of cancer. To address this issue, we selected a target mutated Kirsten rat sarcoma virus oncogene homologue, which plays a significant role in various cancers. Our study aims to identify selective biomarkers and develop diagnostic and therapeutic strategies for KRAS-associated genes using artificial intelligence. Initially, Genomic data, cancer epidemiology, proteomics network interactions, and omics enrichment were analyzed. Structured E-pharmacophore model aided in capturing the binding cavity using eraser algorithms and fabricating a new selective lead compound for the KRSA. The selective molecule was abridged inside the binding cavity and stability was validated through 100 ns molecular dynamics simulations. Epidemiological-neural network studies indicated KRAS mutations leads 40 types of cancer, exclusively pancreatic and colorectal cancers, with diploid and missense mutations as primary factors. Pathway analysis highlighted the involvement of the MAPK and RAS signaling pathways in cancer development and proteomics analysis identified RALGDS as a key protein. Protein-based pharmacophore analysis mapped the biologically active features such as donor, acceptor and aromatic ring with the designed ligands. The results of interaction interpretation illustrate that the amino acid Tyr566 formed an H-bond interaction with the amine group of the octyl ring system and 20 amino acids crafted to properly orient the molecule to fit inside the polar cavity of KRAS protein. The MMGBSA score of  $-53.33$  kcal/mol conformed to the well-configured binding with KRSA and realistic model simulation exposed the  $\pi$ - $\pi$ ,  $\pi$ -cationic and hydrophobic interactions stabilised the molecule inside the KRSA protein throughout 100 ns simulation. The study demonstrates the vitality of AI and network pharmacology to identify potential-target biomarkers for KRAS-associated genes, paving the way for improved cancer diagnostics and therapeutics.

**Keywords** KRAS–RALGDS, Genomics statistics, Pathway analysis, Multi-omics, Molecular modelling

The most often mutated oncogene (Kirsten Rat Sarcoma Viral Oncogene Homologue (KRAS)) is responsible for a diversity of malignancies, such as NSCLC-non-small-cell lung cancer<sup>1</sup>, PDAC-pancreatic ductal adenocarcinoma, & CRC-colorectal cancer<sup>2–6</sup>. The discovery of certain inhibitors and the recognition of driver genes have completely changed the way that cancer is treated, greatly increasing progression-free survival and lowering toxicity in targeted therapy<sup>7–10</sup>. One of the extremely aggressive and severe diseases is pancreatic cancer which has a poor prognosis and limited available treatments. Mutations in the KRAS oncogene are a key factor among the many genetic changes linked to the onset and spread of pancreatic cancer. Roughly 90% of individuals with pancreatic cancer have KRAS mutations, which are among the most frequent genetic abnormalities discovered in these patients<sup>11–14</sup>. Targeted therapies for other mutations, like EGFR and ALK gene fusion, have advanced; however, KRAS has proven to be a difficult disease to treat, with few effective treatments available until sotorasib for the KRAS (G12C) subtype was recently approved<sup>15</sup>. Because of its inherent qualities, targeting KRAS has

<sup>1</sup>Crescent School of Pharmacy, BS Abdur Rahman Crescent Institute of Science and Technology, Chennai 600048, India. <sup>2</sup>D3 Drug Tech Lab Pvt Ltd, Chennai 600048, India. <sup>3</sup>Department of Pharmaceutical Quality Assurance, Shri Vile Parle Kelavani Mandal's Institute of Pharmacy, Dhule, Maharashtra 424001, India. <sup>4</sup>Department of Pharmaceutical Chemistry, Shri Vile Parle Kelavani Mandal's Institute of Pharmacy, Dhule, Maharashtra 424001, India. <sup>5</sup>Department of Pharmaceutical Chemistry, KL College of Pharmacy, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, A.P 522302, India. ✉email: nirfanahamed1985@live.com

proven to be particularly difficult in contrast to other oncogenes. There has been limited success or selectivity in several indirect ways, such as focusing on downstream signalling effectors, using epigenetic techniques, and employing synthetic lethality tactics. Furthermore, KRAS mutant patients frequently show poor response to conventional therapy, resulting in the need for more research in this field<sup>16–19</sup>. The development of inhibitors that selectively target the KRAS oncogene presents a viable path in the search for targeted treatments for pancreatic cancer. However, considering the complexity and dynamic nature of KRAS signalling, this has proven to be a difficult undertaking<sup>20</sup>. According to research by Anabela Ferreira et al., and team, the RAL guanine nucleotide dissociation stimulator (RALGDS) functions as a key downstream effector in KRAS signaling, acting as a GTP/GDP exchange factor to promote the GDP-GTP conversion for RAS-like (RAL) proteins. This activation of RAL GTPases by RALGDS contributes significantly to pro-survival mechanisms, supporting cellular proliferation and cell cycle progression. Recent studies highlight RALGDS's essential role in modulating KRAS-driven oncogenic pathways, linking RALGDS activity to enhanced tumor growth and metastasis in various cancers. RALGDS thus emerges as a critical mediator in KRAS signaling, with potential therapeutic targeting implications. Continued research focuses on the mechanistic pathways and regulatory interactions of RALGDS within the KRAS network to uncover novel intervention strategies. An essential modulator of the KRAS pathway, RALGDS (RAS-specific guanine nucleotide exchange factor) has drawn interest as a possible zone for the fabrication of novel molecules to treat pancreatic cancer. An important development in the ongoing therapy of pancreatic cancer is the discovery of RALGDS-targeted inhibitors. By selectively targeting RALGDS, these inhibitors aim to interfere with the KRAS signalling cascade, which may impede the growth and viability of cancer cells that carry KRAS mutations<sup>21</sup>. These inhibitors offer hope for more potent and focused therapy for pancreatic cancer patients, potentially enhancing their prognosis and quality of life, even if research and clinical application of these drugs are still in their early stages<sup>22</sup>. The discovery of oncogenes linked to rodent sarcoma virus genes, including Kirsten Rat Sarcoma Virus and Murine Sarcoma Virus, began the quest towards understanding KRAS. These oncogenes are homologous to human KRAS, which has been critical in starting a series of signalling molecules that affect cell proliferation, differentiation, chemotaxis, and apoptosis. We shall<sup>23</sup> explore the complex world of KRAS as we go along, including its protein, gene, carcinogenic function, and the exciting opportunities it offers for cancer treatment<sup>24,25</sup>.

To find potential therapeutic targets and the underlying molecular mechanisms influenced by the inhibitors, gene enrichment analysis for K-Ras-associated RALGDS targeted inhibitors involves gathering gene expression data from altered and unaltered gene tissues, selecting gene sets associated with K-Ras and RALGDS pathways, preprocessing the data for quality, statistically comparing gene expression changes like mutation, amplification etc.<sup>26</sup>, visualizing the results, interpreting the physiological importance of these changes, and validating the findings. STRING is a database of anticipated and known protein-protein interactions<sup>27</sup>. They involve direct (physical) & indirect (functional) correlations<sup>28</sup>. The process of docking K-Ras gene-associated RALGDS inhibitor pancreatic cancer-targeted drugs involves predicting the binding affinity, finding of allosteric site of the protein by the aid of an eraser algorithm and the type of interaction of these newly identified lead skeletons by computationally evaluating their interactions with the RALGDS protein<sup>29</sup>. This procedure helps in the discovery of possible lead compounds for pancreatic cancer treatment. Analysing these inhibitors' pharmacokinetic characteristics is crucial after docking experiments. To determine whether a drug is appropriate for a given therapeutic use, dosage, and possible adverse effects, pharmacokinetics evaluates the drug's absorption, distribution, metabolism, and excretion in the body<sup>22</sup>. Molecular docking and pharmacokinetic analysis are combined in this integrated strategy to evaluate the safety and efficacy of newly identified lead skeletons as possible therapeutics. The discovery of oncogenes linked to rodent sarcoma virus genes, including Kirsten Rat Sarcoma Virus and Murine Sarcoma Virus, began the quest towards understanding KRAS<sup>30</sup>. These oncogenes are homologous to human KRAS, which has been critical in starting a series of signalling molecules that affect cell proliferation, differentiation, chemotaxis, and apoptosis. We shall explore the complex world of KRAS as we go along, including its protein, gene, carcinogenic function, and the exciting opportunities it offers for cancer treatment<sup>31</sup>. A computational strategy that streamlines the drug discovery process by utilizing in silico methods to predict and refine interactions between drug candidates and biological targets. Methods such as molecular docking & molecular dynamics (MD) simulations enable researchers to examine how small molecules interact with proteins, evaluate the stability of these interactions, and investigate their behavior in a dynamic, physiological context. CADD enhances the identification of potential drug candidates, minimises reliance on extensive laboratory experiments, and supports the development of more targeted and cost-effective therapies.

## Material and methodology

### Genomic analysis

KRAS-associated gene data were collected and screened by Database CBioportal Cancer Genomics (<https://www.cbioportal.org/>). The CBioportal database was utilized to comprehensively examine patients, analyzing both altered and unaltered genes, including mutation amplifications, deep deletions, multiple alterations, and various types of cancer. Additionally, the study predicted copy number alterations (MSK-IMPACT), splice variants of uncertain significance (VUS), and in-frame mutations<sup>32</sup>. Pathway Analysis of Genes Kras-associated genes were studied on the reactome pathway facilitated the identification of gene pathways, distinguished by its high-performance memory capabilities. To achieve this, the over-representation analysis method is segmented into four distinct steps, each employing specific data structures to enhance performance and minimize memory usage. Pathway analysis methods predominantly serve the analysis of omics data acquired from high-throughput technologies. The discussion focuses on the strengths and weaknesses associated with storing data directly in a relational database for in-database analyses. However, relational databases may encounter challenges in managing large-scale omics data due to limitations in scalability and performance when handling complex data structures and extensive datasets. In contrast, the new pathway analysis approach offers a more optimized

solution by leveraging tailored data structures for each analysis step<sup>33</sup>. This enhances processing speed and overall performance, leading to more comprehensive and accurate pathway analysis results.

### Proteomics AI and multi-omics analysis

Multi-omics is the integrated application of various high-throughput screening technologies, including genomics, transcriptomics, single-cell transcriptomics, proteomics, metabolomics, and spatial transcriptomics. These approaches significantly advance disease research<sup>34</sup>.

$$D_{\text{integrated}} = \sum_{i=1}^n w_i \times D_i$$

Where  $D_i$  represents the dataset from omics source  $i$ , and  $w_i$  is the assigned weight for each data type to optimize predictive accuracy. Grid-based cluster algorithms and AI-based (string) databases centred on protein network interaction. These grid techniques partition a structure and convert the data space into a finite number of cells<sup>35</sup>. Clusters are sites that have a higher density of related cells. Because the majority of grid-based clustering approaches rely on cell density computations, these algorithms are density-based. It was depicted we should identify clusters of protein-protein interaction by using grid-based algorithms with identified genes was analysed through the cystoscope software and formed a network and highly identified position<sup>36,37</sup>. The multi-omics analysis was calculated on the Oncogenes of KRAS-associate. to study the Protein-protein interaction and analyse the closeness to predict the therapeutic target. The heat map genes and overlap were identified in the Metascape package and were utilized for gene enrichment analysis, aiming to explore the biological processes and molecular activities associated with unique proteins, tissues, and gene ontology protein domains<sup>38</sup>. Proteins were ranked using various methods such as molecular function, biological process, SMART protein domains, and UniProt annotation by analysing the log ratio of proteins' Pearson's  $r$  value and  $p$ -value<sup>39,40</sup>. Notably, the RALGDS is a potential target for KRAS-associated genes.

### RALGDS lead design and fabrication

In the study of RALGDS interactions, the Schrodinger Maestro software package was used (Version 13.9.132). The protein RALGDS was initially prepared as the source entry, undergoing preprocessing steps that involved adjusting hydrogen atoms and assigning bond orders via the CCD database. A zero-order method was used for the generation of metal-disulfide bonds. With the pH set to 2.0, Epik was utilized to generate the heteroatom states, while the PROPKA algorithm optimized the structure by orienting sample water molecules to assign hydrogen bonds accurately. Subsequently, the protein was minimized using the OPLS4 force field with a heavy atom RMSD convergence set at 0.30 Å. To identify the ligand binding pocket, site map detection was performed using the eraser algorithm. The receptor grid was generated based on the D score and boundary box constructed surrounding the binding site, which identified the location as a potential binding pocket. The scaling VdW radius was set to 1.0, and a cutoff of 0.25 was used for partial charges. The enclosing box was centred on the peptide centroid within the workspace, with the docking peptide length specified at 14 Å<sup>41,42</sup>. With 4 Å crop site maps at the nearest site point, the Site Map methodology was cast off to spot the binding pocket in the protein standard grid. To dock the ligand an allosteric grid is built based on the site map locations. Ligand designing and fabrication were conducted by protein-based pharmacophore modelling, utilizing active site ligand properties to detect features such as hydrogen bond donors, acceptors, aromatic and aliphatic. A scaling factor of 0.80 and a partial charge cutoff of 0.15 for van der Waals radii were applied to establish the ligand docking technique. Following the import of the grid file, flexible ligand sampling and extra precision XP setup were executed. This process included sampling nitrogen inversion, considering rigid conformations, and utilizing all predefined functional clusters to bias torsion samples. Epik state penalties remained incorporated into the final dock score. Additionally, the energy window for ring sampling was static at 2.5 kcal/mol for conformer generation, and a minimization distance-dependent dielectric constant value of 2.0 was employed.

### MD simulation studies

Using this technique, the best-identified docked result and MMGBSA result studies are conducted for additional confirmation to investigate the physical conformational changes and set the solvated models. The complex was first solvated using the buffer box size scheming technique and the pre-defined SPC solvent model, with the limit condition set to orthorhombic box shape. Na<sup>+</sup> & salt-negative ions (Cl<sup>-</sup>) were added to the system to neutralise it and reduce the last capacity based on the complex protein's surface occupation. Molecular dynamic simulation was applied to the solvated type. The trajectory recording interval was set to 100 ps and the simulation time was set to 100 ns with a 1.2 energy gap<sup>43-45</sup>. An entire 1000 frames are produced at a temperature of the NPT ensemble class of 300 K with 1.01325 bar pressure, with the model relaxed earlier to simulation. The last comprehensive simulation, with trajectory quality ensured, underwent analysis with a block length averaging 10.0 ps<sup>46,47</sup>. This protocol for quality analysis precise the 100 ns simulation, presenting graphs of Total Energy (TE) (kcal/mol), Potential Energy (PE) (kcal/mol), Temperature (T) (K), Pressure (P) (bar), and Volume (V) (Å<sup>3</sup>) parameters. The Trajectory was analysed by using the GPA accelerated Desmond package. Throughout the 100ns period times, they observed the RMSD, and RMSF by using a mathematical model and also analysed the protein-ligand contacts, ligand-protein contacts, and heat map interaction<sup>48</sup>.

$$\text{RMSD} = \sqrt{\sum_{i=1}^N (x_i - x_1)^2 / N} \quad (1)$$

$$\text{RMSF} = \sqrt{\frac{1}{N} \sum_j^N (xi(j) - (xi))^2} \quad (2)$$

### ADMET prediction

ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) forecasts were shown using the Graph-based signature approach. Understanding a compound's pharmacokinetics is crucial for defining its drug-likeness. Values from medicines with comparable in vivo experimental data were utilized to assess query structures intended for drug discovery<sup>49</sup>. The tool receives the canonical form of the lead compound as input and incorporates various parameters such as lipophilicity, polarity, molecular weight, saturation, and insolubility. Furthermore, the tool provides bioavailability scores, drug-likeness assessments, adherence to the Lipinski rule of 5, & toxicity evaluations, amongst other parameters<sup>50,51</sup>. These data collectively contribute to the assessment of lead compound properties, aiding in the identification of promising candidates for drug development.

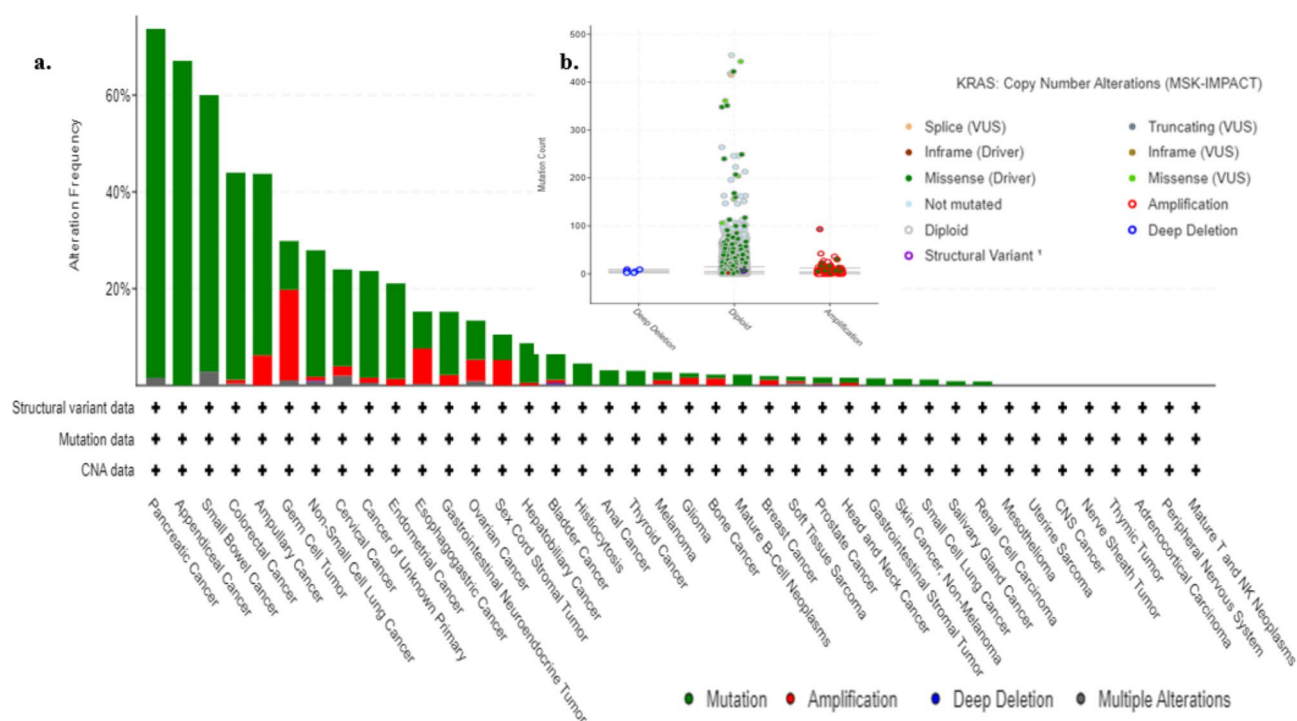
## Result and discussion

### Genomic analysis

Genomic analysis conducted using CBioPortal for cancer genomics focused on the KRAS gene. The most common alteration observed was KRAS mutations, followed by amplifications. These alterations were predominantly found in pancreatic cancer. The KRAS gene is implicated in approximately 40 different types of cancers. It is illustrated in Fig. 1A. The genomic analysis of KRAS copy number alterations using CBioPortal sheds light on the significance of these variations in cancer progression. The results show that diploid KRAS genes with missense driver mutations are the most common, with a high log ratio of 480.

This indicates a strong association between diploid KRAS alterations and cancer development, positioning them as potential targets for cancer therapies. Amplifications, though observed less frequently (log ratio of 100), also contain missense driver mutations. While these alterations are not as prevalent as diploid forms, they may still contribute to cancer progression, suggesting that both diploid and amplified KRAS alterations play important roles in driving oncogenesis, albeit through distinct mechanisms<sup>52</sup>. The presence of deep deletions in KRAS, although less common, suggests the involvement of additional, more complex factors in cancer progression. Although deletions typically lead to loss of function, in the case of KRAS, their impact may be less significant compared to diploid or amplified states.

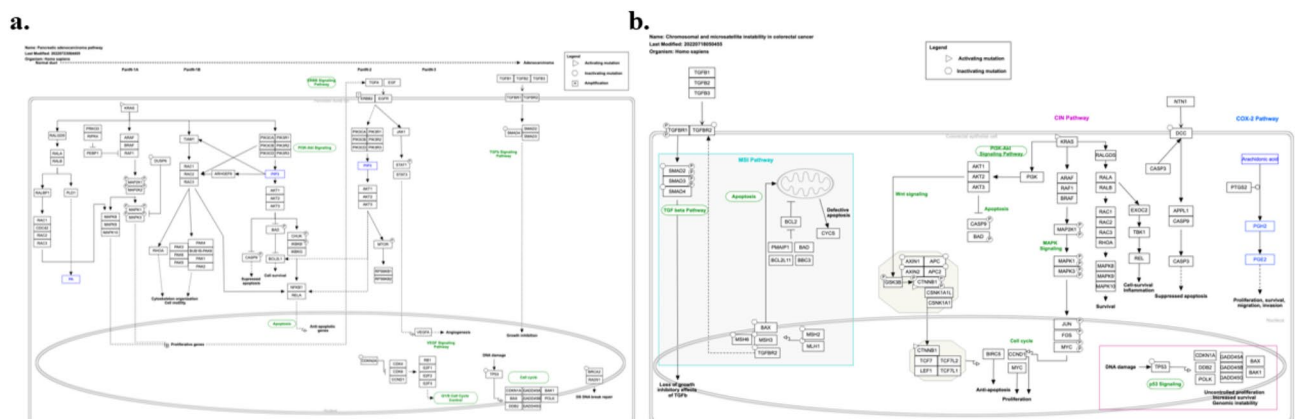
The pathways were presented majorly with Infiltrating ductal adenocarcinoma is the most predominant form of pancreatic malignancy, often referred to as pancreatic ductal adenocarcinoma (PDA) by researchers<sup>53</sup>. The development of PDA involves a stepwise development from normal ductal epithelium to aggressive cancer, with distinct precursor lesions known as Pancreatic Intraepithelial Neoplasia (PanINs) playing a crucial role in this process<sup>54</sup>. The genetic alterations associated with PDA occur at different stages of this progression. Early in the process, we observe the overexpression of the HER-2/neu (ERBB2) gene and then the occurrence of triggering



**Fig. 1.** (a) KRAS Frequency alteration and types of cancer analysis (b) Copy number alteration of KRAS gene.

point mutations in the K-RAS gene. At an intermediate stage, there is inactivation of the p16 gene (CDKN2A). As the disease advances, the inactivation of essential genes like p53, SMAD4, and BRCA2 takes place relatively late in the series of events. These genetic mutations have a deep impact on the development and progression of PDA as illustrated in Fig. 2a. K-RAS, a key player in this situation, triggers various downstream signalling pathways. While it's common to interpret EGF receptors as upstream activators of RAS proteins, it's essential to note that they can also function as RAS signal transducers through a mechanism where RAS induces the autocrine initiation of the EGFR group ligands. Furthermore, PDA exhibits wide genomic unpredictability and aneuploidy, which means that the cancer cells have an abnormal number of chromosomes. This genomic instability can be recognized by factors like telomere attrition in addition to mutations in genes like p53 and BRCA2, which probably contribute to these characteristics. Additionally, the deactivation of the SMAD4 tumour suppressor gene results in the forfeiture of the inhibitory effect of the transforming growth factor-beta signalling pathway, which is another serious aspect of PDA development. To summarize, the progression of pancreatic ductal adenocarcinoma and the associated genetic mutations follow a defined sequence, as illustrated in Fig. 1b, from left to right<sup>55</sup>. Understanding these molecular events is important for improving our knowledge of PDA and developing more effective diagnostic and treatment strategies CRC initiates since the epithelial cells lining the colon and rectum, and its development involves the buildup of specific genetic modifications in oncogenes in addition to tumour suppressor genes (TSGs). There are two key mechanisms of genomic variability related to the progression of sporadic CRC. The initial is chromosomal instability (CIN), a result of genetic variations that activate oncogenes like K-ras and deactivate TSGs like p53, DCC/Smad4, also APC. Another mechanism is microsatellite instability (MSI), which arises after the deactivation of DNA mismatch repair genes MLH1 or MSH2 due to hypermethylation of their promoters and minor mutations in genes by coding microsatellites, such as by way of transforming growth factor receptor II (TGF-RII) in addition BAX<sup>56,57</sup>.

In addition to sporadic cases, there are hereditary syndromes with germline changes in definite genes, such as mutations in the tumour suppressor gene APC on chromosome 5q in ancestral adenomatous polyposis (FAP) and mutated DNA mismatch repair genes in hereditary nonpolyposis colorectal cancer (HNPCC)<sup>60</sup>. The greatest shared mutation in colon cancer includes the deactivation of APC. In cases where APC doesn't have a deactivating mutation, there are frequently triggering mutations in  $\beta$ -catenin. To develop cancer, together alleles must change. When carriers have APC deactivating mutations, the danger of colorectal tumours at age 40 is nearly 100%<sup>61</sup>. The influence of KRAS alterations deeply depends on the order in which mutations occur. Main KRAS changes usually result in self-limiting hyperplastic or marginal lesions, nevertheless, if they happen afterward a previous APC mutation, they often evolve into cancer. KRAS transformation is also a strong forecaster of deprived response to panitumumab and cetuximab treatment in CRC<sup>62</sup>. It is depicted in Fig. 2b. Testing for “activating” mutations in the KRAS gene is currently the most reliable way to forecast a colorectal cancer patient's response to EGFR-inhibiting medications. Loss of heterozygosity (LOH) of DCC in area 18q21 is one of the greatest recurrent genetic irregularities in advanced CRC furthermore, TGF $\beta$  signalling is active in nearly CRC cells with MSI alterations in TGFBR2, occurring in 30–50% of CRC. Individuals whose tumours direct the altered version of the KRAS gene do not answer to cetuximab or panitumumab. DCC is a gene with a conditional role, acting as both a tumour suppressor and an oncogene. Once DCC exists and is not triggered by netrin, it promotes programmed cell death and represses tumour formation<sup>63</sup>. However, once DCC exists and netrin-activated, it supports cell survival and stands in as an oncoprotein. Abnormal overexpression of cyclooxygenase-2 (COX-2) is also understood to play a significant part in CRC growth. The tumorigenic properties of COX-2 are recognized in the generation of PGE2 (prostaglandin E2), and enhanced ranges of PGE2 have been observed in colorectal adenomas and carcinomas. COX-2 in addition to PGE2 regulates various processes in colorectal tumours, including proliferation, survival, migration, and invasion.” This elaboration provides a more comprehensive understanding of the genetic and molecular factors contributing to colorectal cancer development and progression.



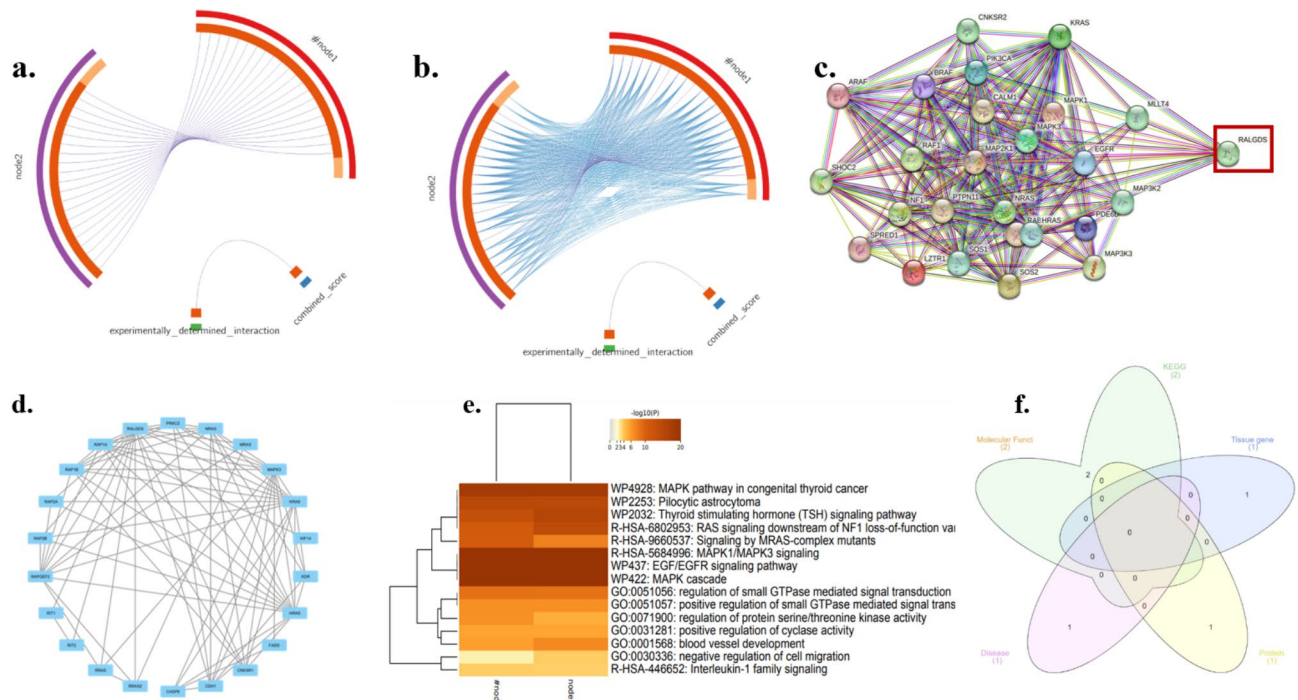
**Fig. 2.** (a) Pathway of pancreatic cancer (b) Pathway of colorectal cancer. Figure is reproduced from Ref<sup>58,59</sup>.

### Proteomics AI and multi-omics analysis

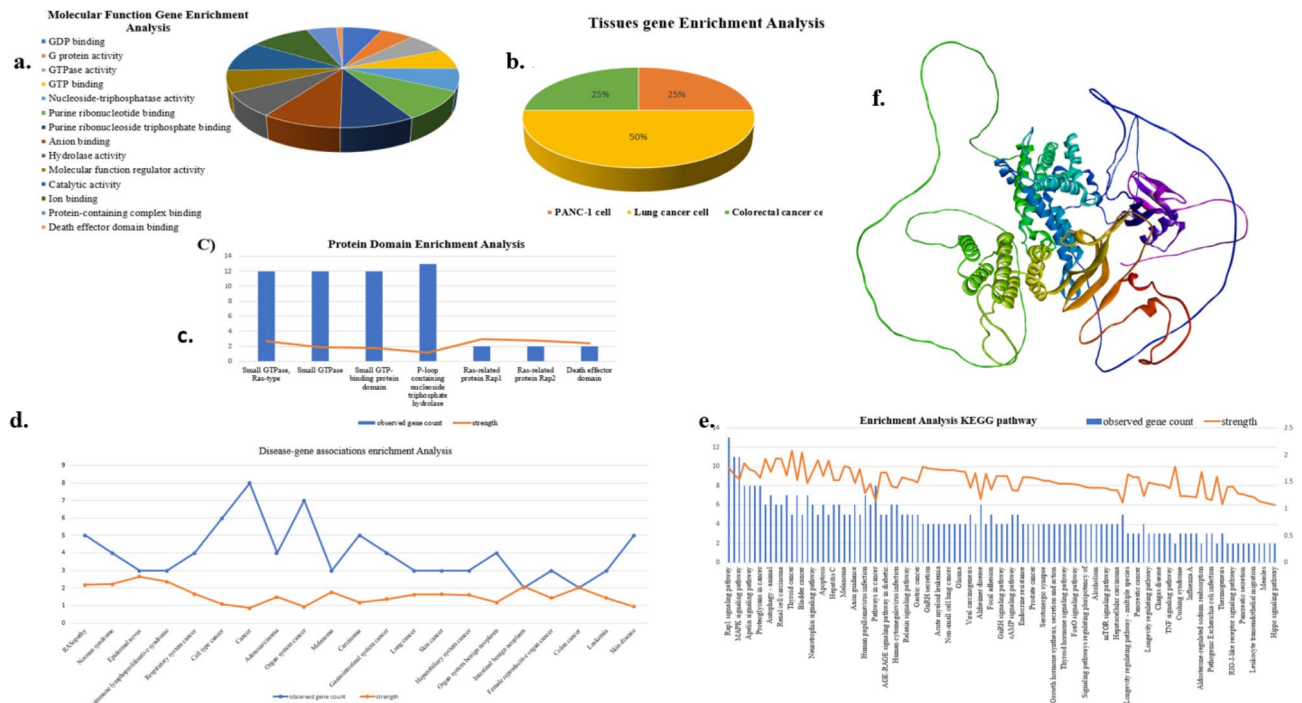
Incorporating AI into network pharmacology provides two primary advantages through a deep interpretable network relationship inference framework. First, the hierarchical structure and biologically relevant nodes allow for interpretable feature learning, giving biological context to the model's predictions. Second, this approach expands the feature space by learning both the network's features and their associations, effectively addressing the challenges of low generalization and limited interpretability commonly faced by AI-based models in biological networks. Additionally, advancements in language model technologies, particularly large language models (LLMs), have helped mitigate previous constraints related to computational resources and data volume. These developments facilitate the conception of robust deep interpretable models, streamlining network pharmacology methodologies and enhancing the utility of AI in this domain.

Figure 3a illustrates the gene level, identical genes are linked by purple curves. Figure 3b at the common term level, blue curves connect genes fitting to the similar enriched ontology term, revealing functional similarities. The internal circle signifies gene lists, with dark orange indicating genes present in numerous lists in addition to light orange denoting genes exclusive to a specific list. This visual analysis provides insights into both genetic overlap and shared biological functions, providing an understanding of the relationships between the gene sets<sup>64</sup>. The AI proteomics analysis was done on the Strings db database to identify the protein-protein interaction and identify the nodes as illustrated in Fig. 3c. As a result, the KRAS genes have high nodes and closeness was found in the RALGDS gene marked in Fig. 3d. Moreover, we examined and projected gene overlaps, heat maps, and protein enrichment analyses such as molecular function, tissue gene, protein domain, KEGG pathway, and disease association to further validate<sup>65</sup>. The protein network cluster has 23 proteins - protein has interacted strongly and a large number of nodes in the RALGDS protein, which was predicted on the cystoscope. It illustrates the comparison of two gene lists, showing overlap at two levels<sup>66</sup>. The heatmap of gene enrichment analysis was done on the Metascape software to predict the pathway of the genes based on the log ratio. From 0 to 20, they are mostly the highest signalling RAL guanine nucleotide dissociation stimulator (RALGDS) functions as a key downstream effector in KRAS signaling, acting as a GTP/GDP exchange factor to promote the GDP-GTP conversion for RAS-like (RAL) proteins. This activation of RAL GTPases by RALGDS contributes significantly to pro-survival mechanisms, supporting cellular proliferation and cell cycle progression. Recent studies highlight RALGDS's essential role in modulating KRAS-driven oncogenic pathways, linking RALGDS activity to enhanced tumour growth and metastasis in various cancers. RALGDS thus emerges as a critical mediator in KRAS signalling, with potential therapeutic targeting implications<sup>21,67</sup>. Ongoing research is centered on exploring the mechanistic pathways and regulatory interactions of RALGDS within the KRAS network, aiming to identify novel intervention strategies targeting the MAPK cascade and EGFR signaling pathways. Understanding these interactions is crucial, as they play a significant role in cancer progression and therapeutic resistance. In addition, a predictive analysis was conducted to identify genes associated with thyroid cancer, with the results visualized in Fig. 3e.

The omics analysis was performed on the major five components of the matching proteins which are represented in Fig. 4 as a molecular function, tissue gene, protein domain, and disease-associated KEGG



**Fig. 3.** (a) and (b) KRAS associated overlapped genes similarity (c) and (d) Network pharmacology of KRAS-associated genes (e) Heat map of gene enrichment KRAS (f) components of gene enrichment.



**Fig. 4.** Enrichment analysis (a) Molecular function (b) tissue gene (c) protein domain (d) Disease associate enrichment analysis (e) KEEG enrichment analysis (f) RALGDS protein.

pathway<sup>68</sup>. We also identified the matching proteins KRAS set of RALGDS protein in the analysis and also identified the genes associated with function strength and count. Disease, molecular function, proteins, KEGG, and tissues are the five primary components of enrichment analysis. These are represented in Fig. 4. The Molecular Function component shows 14 available components, illustrated in Fig. 4a. The major components include Anion binding, with matching proteins found to be RAP2A, RRAS, RAP1B, KRAS, RRAS2, MAPK3, KDR, RAPGEF2, MRAS, RAP2B, RIT2, KIF14, RIT1, NRAS, RALGDS, RAP1A, PRKCZ, and HRAS. The purine ribonucleoside triphosphate binding component matches proteins such as RAP2A, RRAS, RAP1B, KRAS, RRAS2, MAPK3, KDR, MRAS, RAP2B, RIT2, KIF14, RIT1, NRAS, RAP1A, PRKCZ, and HRAS. Similarly, the purine nucleotide binding component corresponds to proteins like RAP2A, RRAS, RAP1B, KRAS, RRAS2, MAPK3, KDR, RAPGEF2, MRAS, RAP2B, RIT2, KIF14, RIT1, NRAS, RAP1A, PRKCZ, and HRAS. These three components are majorly present in the molecular function enrichment analysis. The tissue enrichment analysis unveiled three distinct cancer types, notably showcasing lung cancer's dominance, encompassing about 50% of the tissues examined. Among the discerned proteins, KRAS, CDH1, NRAS, and HRAS emerged as key players within this context. Additionally, pancreatic and colorectal cancers emerged within the same tissue milieu, featuring the prominent presence of KRAS and CDH1 proteins. This insightful data is eloquently depicted in Fig. 4b. and In Fig. 4c, the protein domain enrichment analysis reveals a notable emphasis on the P-loop containing nucleoside triphosphate hydrolase. This domain garnered particular prominence, showcasing a rich abundance of proteins including RAP2A, RRAS, RAP1B, KRAS, RRAS2, MRAS, RAP2B, RIT2, KIF14, RIT1, NRAS, RAP1A, RALGDS and HRAS. This insightful data underscores the intricate molecular landscape under scrutiny, examined proteins. The analysis of the KEGG pathway revealed a comprehensive array of 104 components, prominently featuring pathways such as Rap1 signalling, MAPK1 signalling, and several cancer signalling pathways. Within these pathways, a cohort of matching proteins was identified, including RRAS, RAP1B, KRAS, CDH1, MAPK3, KDR, RAPGEF2, MRAS, NRAS, RAP1A, RALGDS, PRKCZ, and HRAS. This extensive repertoire of proteins underscores their crucial roles in modulating various cellular processes and highlights their significance in the context of KEGG pathway analysis. The disease enrichment analysis revealed a significant gene count associated with cancer, indicating a substantial presence of genes implicated in various types of cancer. This comprehensive analysis, depicted in Fig. 4d, illuminated the diverse spectrum of cancer types involved. Notably, matching proteins such as KRAS, RRAS2, CDH1, KDR, CASP8, RALGDS, RIT1, NRAS, and HRAS were identified within this enriched Analysis as shown in Fig. 4e. proteins KRAS, NRAS, and HRAS are crucial components of the RAS signaling pathway, which regulates cell proliferation, differentiation, and survival. Mutations in these proteins often result in their continuous activation, leading to uncontrolled cell growth and cancer progression. The role of CDH1, which is involved in cell-cell adhesion, is particularly important in preventing cancer metastasis; loss of its function can contribute to tumor invasion and metastasis. CASP8, on the other hand, plays a key role in apoptosis, and its dysregulation can lead to evasion of programmed cell death, thus promoting cancer cell survival. KDR (VEGFR-2) is a vital mediator in angiogenesis, the process by which new blood vessels form to supply nutrients to growing tumors. By targeting KDR, it may be possible to inhibit tumor vascularization and slow cancer growth. Additionally, proteins such as RALGDS and

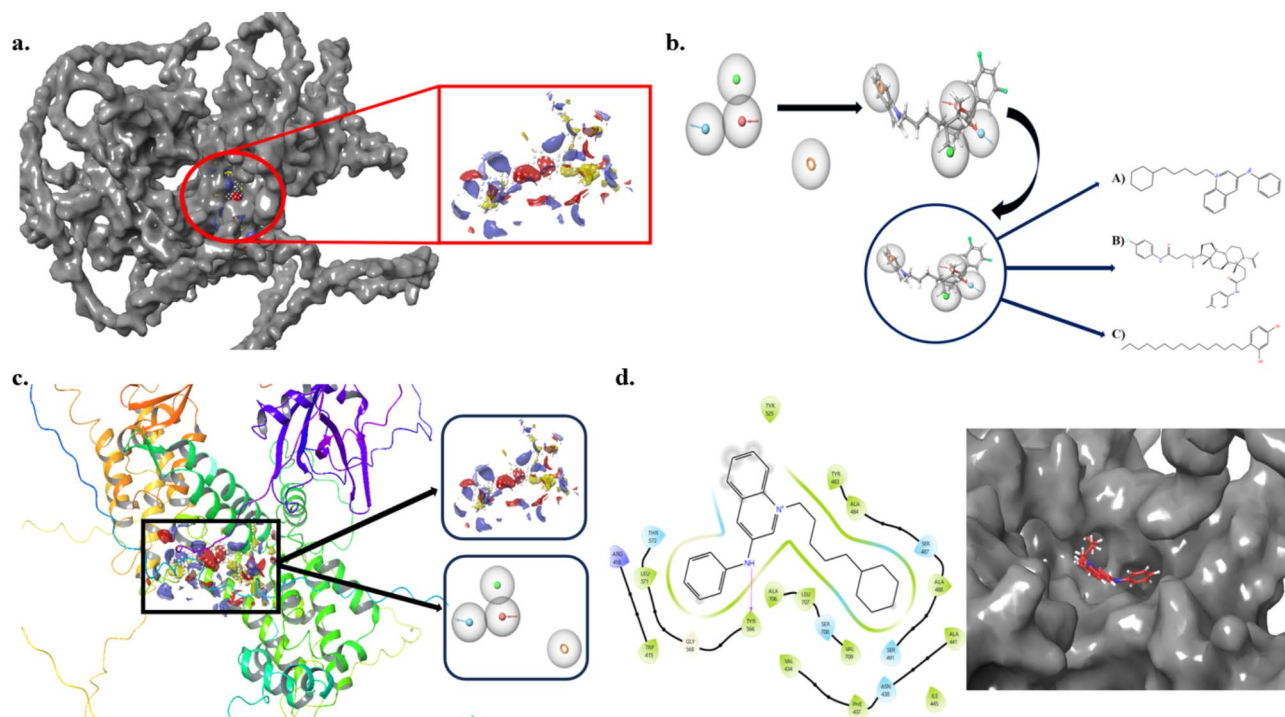
RRAS2 are involved in regulating cell signaling related to proliferation and differentiation, which can influence tumor development and progression. The abundance of these proteins underscores their potential roles in the pathogenesis and progression of cancer. RALDGS protein is proposed as a versatile guanine nucleotide exchange factor (GEF), triggering RalA and RalB GTPases to orchestrate intracellular transport<sup>69</sup>. Acting as an effector for R-Ras, H-Ras, K-Ras, and Rap, RALDGS plays a significant role in transducing signals downstream of these GTPases, influencing cellular processes and *In vivo* studies. Gonzalez et al. (2005) experimentally proved the formation of cancer in mouse model in Fig. 4f. This multifaceted protein's interactions and regulatory functions contribute to the coordination of intracellular transport and the maintenance of cellular homeostasis, describing its significance in fundamental cellular processes.

### RALDGS lead design and fabrication

The Alpha fold protein AF-Q12967 does not contain a catalytic pocket ligand, hence, we performed the binding pocket analysis in the maestro Schrodinger package using the binding pocket prediction module. First, the protein was prepared, pre-processed, removed water, and minimized and given the command binding site detection to identify the binding site it was illustrated in Fig. 5a<sup>68,70</sup>. Binding site selection was identified by the D score which was greater than 1, so it confirmed the organic compound binding protein. In Fig. 5c Ligand ASL properties were predicted for the pharmacophore developing (chain. name A (res.num 398 | res.num 401 | res.num 415 | res.num 416 | res.num 418 | res.num 419 | res.num 420 | res.num 433 | res.num 434 | res.num 436 | res.num 437 | res.num 438 | res.num 480 | res.num 483 | res.num 519 | res.num 524 | res.num 525 | res.num 529 | res.num 565 | res.num 566 | res.num 567 | res.num 568 | res.num 569 | res.num 571 | res.num 572 | res.num 575 | res.num 709 | res.num 711 | res.num 712 | res.num 713 | res.num 714)) by the that the forming of the pharmacophore developed the molecule with proof H-Bond donor & acceptor and aromatic system.

The nature of binding pocket analysis showed the Hydrophobic residues (Val, Leu, Phe, Met) stabilize the protein structure and fold to form binding pockets, and the Polar, charged residues (Ser, Thr, His, Asp, Arg, Lys) aid to substrate binding, and act as a potentially catalytic amino acid. The aromatic residues (Tyr, Trp) contribute to structural stability and ligand recognition. Two additional flexible residues (Gly, Pro) facilitate structural adjustments to form proper configurational ligand binding.

We designed an extensive set of molecules and analyzed their pharmacophoric features to identify three lead compounds, as illustrated in Fig. 5b. These molecules were developed based on the E-pharmacophore model, which included four key features: aromatic rings, hydrogen bond donors, and acceptors. During molecular docking, it was observed that the octyl aromatic ring system of one of the selected molecules fit well within the protein's binding cavity. The docking score, along with the MMGBSA binding energy calculation, yielded a value of -53.53 kcal/mol. The 2D interaction map is depicted in Fig. 5d, with additional details in Supplementary Fig. 1. The binding interactions revealed one hydrogen bond with residue TYR 566. Hydrophobic interactions were noted with several amino acids, including TYR 524, TYR 483, ALA 484, ALA 488, ALA 441, ILE 445, PHE 437, VAL 434, VAL 709, ALA 706, LEU 707, and TRP 415. Additionally, polar interactions were observed with



**Fig. 5.** (a) Binding pocket of RALDGS (b) RALDGS based on pharmacophore to lead generation (c) Binding pocket of RALDGS based on pharmacophore (d) Lead Interaction.

Lead molecule	Glide score (Kcal/mol)	MMGBSA (Kcal/mol)
A	− 6.026	− 53.53
B	− 5.928	− 51.06
C	− 5.260	− 36.66

**Table 1.** Docking score & MMGBSA score of the lead molecule.

residues ASN 438, SER 491, SER 708, SER 487, and THR 572. A positively charged interaction was identified with ARG 418. These results, including the MMGBSA it was depicted in (Table 1) The polar solvation energy, calculated using the Generalized Born model, was balanced effectively by the non-polar solvation contributions, which were optimized by the hydrophobic interactions observed with residues are ALA 484, TYR 483, and PHE 437.

### MD simulation studies

The quality of the molecular dynamic simulation results was thoroughly assessed, and the limit values found were illustrated in Fig. 6a; Table 2 to be within desirable ranges, indicating a high level of reliability and accuracy. Specifically, the multifaceted simulation exhibited stable volume (V), pressure (P), temperature (T), potential energy (PE), and overall energy profiles throughout the simulation.

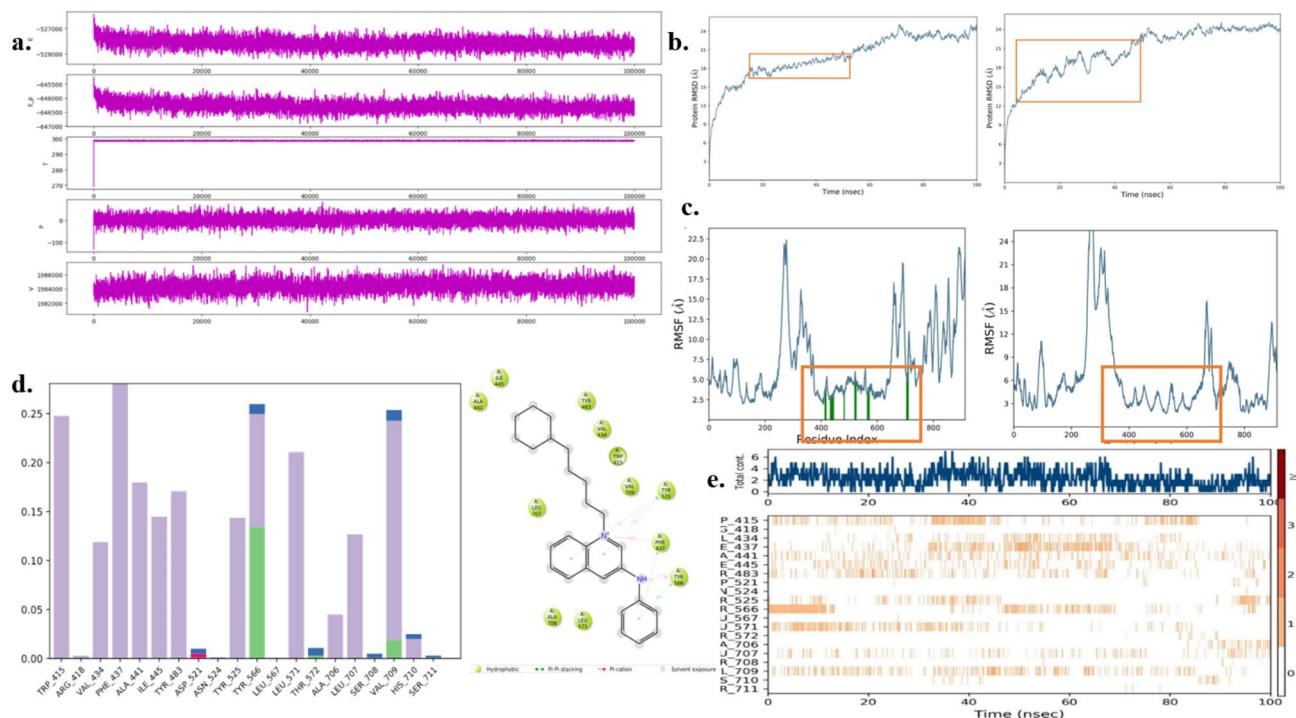
The RMSD graph was obtained from the 100ns molecular dynamics (MD) simulation, as illustrated in Fig. 6b. showcased the behavior of the protein-ligand complex. It was observed that the protein-ligand complex exhibited fluctuations within the range of 1–2 angstroms, indicating relatively stable behavior compared to the protein alone. Conversely, the protein alone demonstrated greater fluctuations, surpassing 3 angstroms, suggesting less stability in isolation. In conclusion, the analysis reveals that the protein-ligand complex, particularly involving the RALGDS protein, maintains stability throughout the MD simulation.

The protein Root Mean Square Fluctuation (RMSF) analysis revealed that the protein-ligand complex exhibited reduced fluctuation, measuring approximately 22.5 angstroms, compared to the plain protein which displayed higher fluctuation at around 25 angstroms. Moreover, the interaction analysis revealed that amino acid interactions between the ligand molecule and the protein occurred consistently from frames 400 to 700 of the simulation. This time frame indicates the formation and stability of interactions between the ligand and the protein residues. These findings are visually represented in Fig. 6c. These interactions primarily involved  $\pi$ – $\pi$  stacked interactions and H-bond interactions with active amino acids, notably TYR 525 and PHE 437 as illustrated in Fig. 6d. The heat map analysis performed throughout the simulation period revealed significant interactions between the ligand and the protein, with particularly strong involvement of residue TYR 566, as depicted in the accompanying. The analysis indicated that more than 20 amino acid interactions were maintained consistently throughout the simulation, as shown in Fig. 6e. These interactions highlight the stability and robustness of the binding. By the end of the simulation, it was evident that the newly designed octyl aromatic ring system demonstrated stable binding within the pocket of the RALGDS protein.

### ADMET prediction

The ADMET profile was determined using the PKCSM database, allowing us to pinpoint three leads for comparison. We verified parameters such as water solubility, Caco2 permeability, intestinal absorption, and skin permeability. Through this analysis, lead 2 emerged as having excellent absorption capabilities<sup>71</sup>. When it comes to P-glycoprotein substrate, these leads demonstrated no inhibitory effects. Specifically, for P-glycoprotein I and II inhibitors, lead 3 displayed no inhibitory effect, lead 1 is a substrate of CYP2D6 and CYP3A4 enzymes and functions as an inhibitor of CYP1A2, CYP2D6, and CYP3A4. Lead 2, on the other hand, is only a substrate for CYP3A4 and does not inhibit any of the listed cytochrome p450 enzymes.

Lead 3 acts as a substrate for CYP3A4 and inhibits CYP1A2 but does not inhibit CYP2D6 Comparing these properties, lead 1 stands out for its dual substrate activity with CYP2D6 and CYP3A4 and broad inhibition profile, while lead 2 primarily serves as a substrate for CYP3A4 without inhibitory effects In Fig. 7c. Lead 3 shares substrate activity with cyp3a4 but differs in its selective inhibition of CYP1A2 only in [supplementary 1](#). The ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties of three lead compounds (lead 1, lead 2, and lead 3) reveal distinct profiles. All leads exhibit similar absorption characteristics in terms of water solubility, caco-2 permeability, intestinal absorption in humans, and skin permeability. Lead 3 is the most likely to be a substrate for p-glycoprotein and also the most potent inhibitor of p-glycoprotein (i) while lead 1 shows the highest inhibitory activity for p-glycoprotein (ii) Regarding total clearance, lead 1 is eliminated the fastest, followed by lead 3 and lead 2. In terms of renal oct2 substrate likelihood, lead 3 is the most probable, followed by lead 2 and lead 1. This comprehensive analysis indicates varying degrees of absorption and interaction as shown in Figure A with p-glycoprotein, clearance rates, and renal transport among the three leads, providing valuable insights for further development. The provided image features two 3D graphs comparing three leads (Lead 1, Lead 2, and Lead 3) on distribution and toxicity profiles as depicted in Figures d, e & f. Graph A assesses the volume of distribution (Vdss), fraction unbound in humans, blood-brain barrier (BBB) permeability, and central nervous system (CNS) permeability. Graph B evaluates the maximum tolerated dose



**Fig. 6.** (a) Quality Simulation Analysis 100 ns (b) RMSD fluctuation lead and protein (c) RMS fluctuation protein-ligand complex and protein (d) Interaction of protein-ligand contacts 100 ns time and Interaction map of interaction (e) Heat map 100ns protein complex interaction.

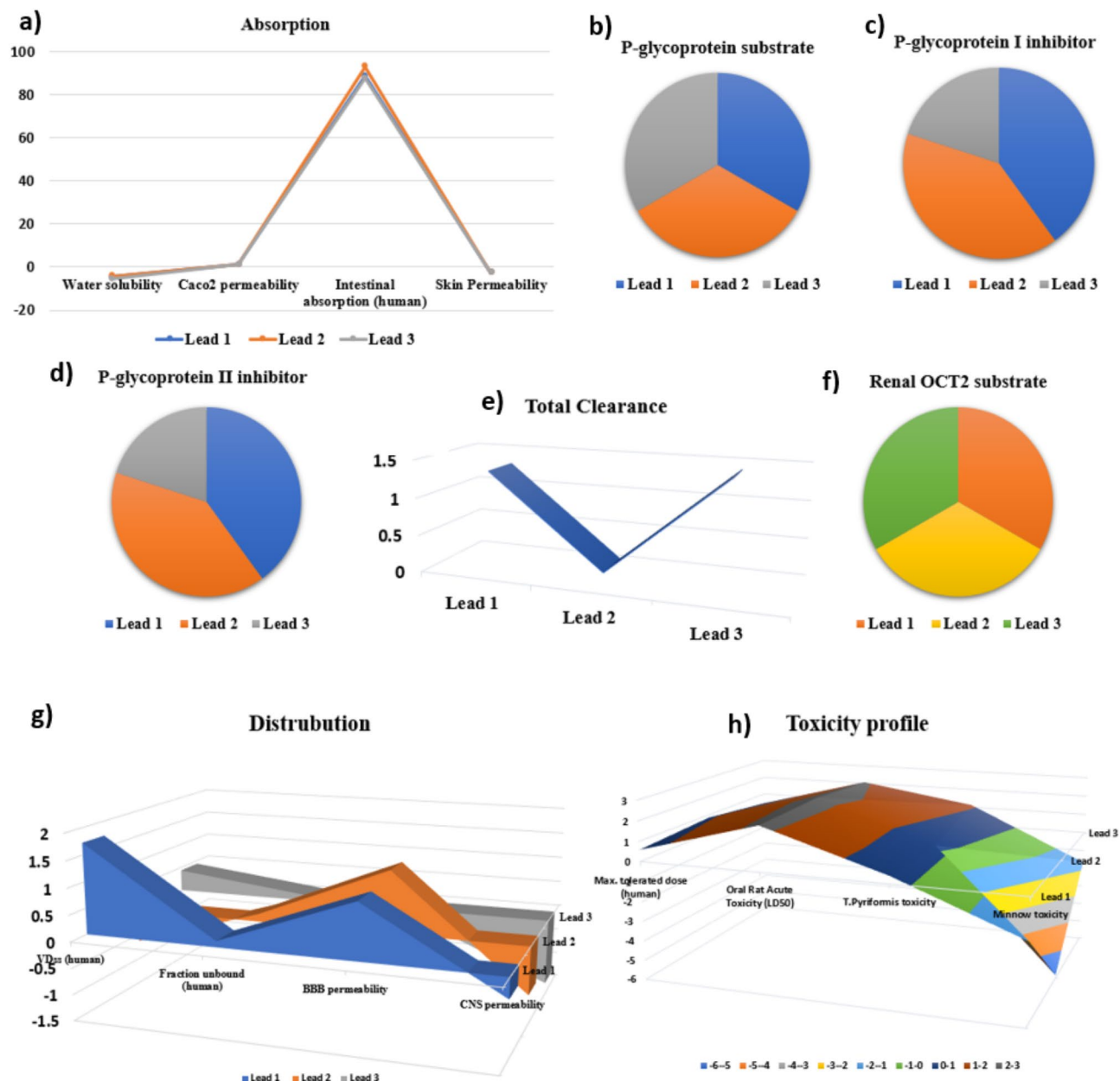
Properties	Average	Std. Dev	Slope (ps <sup>-1</sup> )
Quality values of NRPB1 protein complex simulation (100ns)			
Overall energy (kcal/mol)	- 527580.359	210.757	- 0.002
PE (kcal/mol)	- 646278.828	184.545	- 0.002
T (K)	298.707	0.378	0.000
P (bar)	1.183	21.049	0.000
V (Å <sup>3</sup> )	1984368.644	863.914	0.005

**Table 2.** Parameters of MD simulation quality analysis.

in humans, oral rat acute toxicity (LD50), and toxicity in *Tetrahymena pyriformis* and minnows. Lead 1 shows good CNS penetration but high environmental toxicity. In Figure h Lead 2 presents a balanced profile with moderate distribution and better safety. Lead 3 exhibits strong BBB and CNS permeability but poor overall distribution and lower safety. Based on these attributes.

## Conclusion

This research study employed an Artificial Intelligence-driven network biology approach to investigate the involvement of KRAS Oncogenes in diverse cancer types. By joining genomics statistics and accessing comprehensive genetic datasets from CBioportal cancer genomics, we developed an innovative drug discovery methodology. Through meticulous gene analysis encompassing Pathways, omics, and enrichment analyses, to study the biological function and molecular function of the protein we found that the RALGDS protein is a promising biomarker linked to KRAS. Following the modeling of the RALGDS protein, we utilized the eraser algorithm to identify the optimal binding cavity. Subsequently, we developed the single protein E-pharmacophore properties to design a molecule based on the ASL properties of the binding cavity. Upon docking the designed molecule, we obtained an MMGBSA score of -53.53 Kcal/mol and the octyl aromatic ring lead 1 has the best ADMET profile providing further validation. To enhance confirmation, we conducted Molecular Dynamics simulations to analyse the stability of the protein lead complex. Our findings indicate the formation of a stable complex, affirming the efficacy of the proposed molecule in RALGDS.



**Fig. 7.** (a) Absorption (b) Glycoprotein substrate (c) Glycoprotein I Inhibitors (d) Glycoprotein II Inhibitors (e) Total clearance (f) RENAL OCT2 (g) Distribution (h) Lead molecule toxicity profile.

## Data availability

Data is provided within the manuscript or supplementary information files.

Received: 13 September 2024; Accepted: 21 February 2025

Published online: 29 March 2025

## References

- Attieh, F. & Kourie, H. R. Breaking barriers: The latest insights into KRAS G12C inhibitors. *Future Oncol.* <https://doi.org/10.1080/14796694.2024.2394009> (2024).
- Hou, H. A. & Tien, H. F. Genomic landscape in acute myeloid leukemia and its implications in risk classification and targeted therapies. *J. Biomed. Sci.* **27**, 81 (2020).
- Mullard, A. Cracking KRAS. *Nat. Rev. Drug Discov.* **18**, 887–891 (2019).
- Vendramini, E. et al. KRAS, NRAS, and BRAF mutations are highly enriched in trisomy 12 chronic lymphocytic leukemia and are associated with shorter treatment-free survival. *Leukemia* **33**, 2111–2115 (2019).
- Oya, Y., Imaizumi, K. & Mitsudomi, T. The next-generation KRAS inhibitors... What comes after Sotorasib and adagrasib? *Lung Cancer* **194**, 107886 (2024).

6. Mok, T. S. K. et al. KRYSTAL-12: Phase 3 study of adagrasib versus docetaxel in patients with previously treated advanced/metastatic non-small cell lung cancer (NSCLC) harboring a KRAS<sup>G12C</sup> mutation. *J. Clin. Oncol.* **42**, LBA8509–LBA8509 (2024).
7. Martínez-Jiménez, F. et al. A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* **20**, 555–572 (2020).
8. Zhou, M. et al. Recent progress in retinoblastoma: Pathogenesis, presentation, diagnosis and management. *Asia-Pacific J. Ophthalmol.* **13**, 100058 (2024).
9. Shi, Y. et al. Comprehensive pan-cancer analysis of 33 human cancers reveals immunotherapeutic value of focal adhesion tyrosine kinase. *Medicine* **103**, e37362 (2024).
10. Yu, X. et al. Cancer epigenetics: From laboratory studies and clinical trials to precision medicine. *Cell. Death Discov.* **10**, 28 (2024).
11. Hu, J. X. et al. Pancreatic cancer: A review of epidemiology, trend, and risk factors. *World J. Gastroenterol.* **27**, 4298–4321 (2021).
12. Zhou, J. D. et al. KRAS overexpression independent of RAS mutations confers an adverse prognosis in cytogenetically normal acute myeloid leukemia. *Oncotarget* **8**, 66087–66097 (2017).
13. Jerchel, I. S. et al. RAS pathway mutations as a predictive biomarker for treatment adaptation in pediatric B-cell precursor acute lymphoblastic leukemia. *Leukemia* **32**, 931–940 (2018).
14. del Re, M. et al. Implications of KRAS mutations in acquired resistance to treatment in NSCLC. *Oncotarget* **9**, 6630–6643 (2018).
15. Reita, D. et al. Direct targeting KRAS mutation in non-small cell lung cancer: Focus on resistance. *Cancers (Basel)* **14** (2022).
16. Dietlein, F. Identification of cancer driver genes based on nucleotide context. *Nat. Genet.* **52**, 208–218 (2020).
17. A compendium of mutational cancer driver genes. *Nat. Rev. Cancer*. <https://www.nature.com/articles/s41568-020-0290-x>
18. Hussain, A. & Brooks, C. L. III Guiding discovery of protein sequence-structure-function modeling. *Bioinformatics* **40** (2024).
19. Kumar, R. et al. Somatic mutational landscape across Indian breast cancer cases by whole exome sequencing. *Sci. Rep.* **14**, 18679 (2024).
20. Mao, Z. et al. KRAS(G12D) can be targeted by potent inhibitors via formation of salt bridge. *Cell. Discov.* **8**, 5 (2022).
21. Huang, L., Guo, Z., Wang, F. & Fu, L. KRAS mutation: From undruggable to druggable in cancer. *Signal. Transduct. Target. Ther.* **6**, 386 (2021).
22. Hu, Z. I. & O'Reilly, E. M. Therapeutic developments in pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* **21**, 7–24 (2024).
23. Cornilescu, G. et al. Natural product Graveoline modulates Kirsten rat sarcoma viral oncogene homologue (KRAS) membrane association: Insights from advanced spectroscopic studies. *ACS Pharmacol. Transl. Sci.* **7**, 1983–1995 (2024).
24. Simanshu, D. K., Nissley, D. V. & McCormick, F. RAS proteins and their regulators in human disease. *Cell* **170**, 17–33 (2017).
25. Nourbakhsh, M., Degen, K., Saksager, A., Tiberti, M. & Papaleo, E. Prediction of cancer driver genes and mutations: The potential of integrative computational frameworks. *Brief. Bioinform.* **25** (2024).
26. You, Y. et al. Artificial intelligence in cancer target identification and drug discovery. *Signal. Transduct. Target. Ther.* **7**, 156 (2022).
27. Akbarzadeh, S., Coşkun, Ö. & Günċer, B. Studying protein–protein interactions: Latest and most popular approaches. *J. Struct. Biol.* **216**, 108118 (2024).
28. Pepe, G. et al. Artificial intelligence methods enhance the discovery of RNA interactions. *Front. Mol. Biosci.* **9** (2022).
29. Han, C. W., Jeong, M. S. & Jang, S. B. Structure, signaling and the drug discovery of the Ras oncogene protein. *BMB Rep.* **50**, 355–360 (2017).
30. Weller, J. A. & Rohs, R. Structure-based drug design with a deep hierarchical generative model. *J. Chem. Inf. Model.* **64**, 6450–6463 (2024).
31. García-España, A. & Philips, M. R. Origin and evolution of RAS membrane targeting. *Oncogene* **42**, 1741–1750 (2023).
32. Lee, J. K. et al. Comprehensive pan-cancer genomic landscape of KRAS altered cancers and real-world outcomes in solid tumors. *NPJ Precis. Oncol.* **6**, 91 (2022).
33. Fabregat, A. et al. Reactome pathway analysis: A high-performance in-memory approach. *BMC Bioinform.* **18**, 142 (2017).
34. Chen, C. et al. Applications of multi-omics analysis in human diseases. *MedComm (Beijing)* **4** (2023).
35. Skzlarczyk, D. et al. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
36. Das, S. & Chakrabarti, S. Classification and prediction of protein–protein interaction interface using machine learning algorithm. *Sci. Rep.* **11**, 1761 (2021).
37. Daou, L. & Hanna, E. M. Predicting protein complexes in protein interaction networks using mapper and graph convolution networks. *Comput. Struct. Biotechnol. J.* **23**, 3595–3609 (2024).
38. Chen, J. et al. Multi-omics dissection of stage-specific artemisinin tolerance mechanisms in Kelch13-mutant plasmodium falciparum. *Drug Resist. Updates* **70**, 100978 (2023).
39. Zou, H. T., Ji, B. Y. & Xie, X. L. A multi-source molecular network representation model for protein–protein interactions prediction. *Sci. Rep.* **14**, 6184 (2024).
40. Mohammed Zaidh, S. et al. Genes adaptability and NOL6 protein Inhibition studies of fabricated flavan-3-ols lead skeleton intended to treat breast carcinoma. *Int. J. Biol. Macromol.* **127661** <https://doi.org/10.1016/j.jbiomac.2023.127661> (2023).
41. Irfan, N. et al. Active and allosteric site binding MM-QM studies of methylidene tetracyclo derivative in PCSK9 protein intended to make a safe antilipidemic agent. *J. Biomol. Struct. Dyn.* **1–10** <https://doi.org/10.1080/07391102.2023.2239928> (2023).
42. Navabshah, I. et al. Computational lock and key and dynamic trajectory analysis of natural biophors against COVID-19 Spike protein to identify effective lead molecules. *Mol. Biotechnol.* **63**, 898–908 (2021).
43. Ravi, G. N., Navabshah, I., Unnikrishnan, S. & Ramalingam, K. In silico protein–protein interaction of pterois volitans venom with cancer inducers of *Helicobacter pylori*. *Appl. Biochem. Biotechnol.* **194**, 354–367 (2022).
44. Irfan, N., Balasubramanian, S., Ali, D. M. & Puratchikody, A. Bioisosteric replacements of tyrosine kinases inhibitors to make potent and safe chemotherapy against malignant cells. *J. Biomol. Struct. Dyn.* **41**, 9437–9447 (2023).
45. Habeeb, M. et al. Development characterization and molecular simulation studies of metoclopramide HCl and tramadol HCl bilayer tablets. *Res. J. Pharm. Technol.* **529–534**. <https://doi.org/10.52711/0974-360X.2022.00085> (2022).
46. Haghighkhah, H., Ghalami Choobar, B. & Amjad-Iranagh, S. Effect of salt concentration on properties of mixed carbonate-based electrolyte for Li-ion batteries: A molecular dynamics simulation study. *J. Mol. Model.* **26**, 220 (2020).
47. Gregory, K. P., Elliott, G. R., Wanless, E. J., Webber, G. B. & Page, A. J. A quantum chemical molecular dynamics repository of solvated ions. *Sci. Data* **9**, 430 (2022).
48. RMSD/RMSF Analysis | BioChemCoRe (2018). <https://ctlee.github.io/BioChemCoRe-2018/rmsd-rmsf/>
49. Wu, F. et al. Computational approaches in preclinical studies on drug discovery and development. *Front. Chem.* **8**, 726 (2020).
50. Kumar, C. S., Narasu, M. L. & Singh, C. R. Pharmacokinetic analysis and structural optimization of inophyllamine-I to forecast as a possible drug candidate. *Phytomed. Plus* **3**, 100422 (2023).
51. Myung, Y., de Sá, A. G. C. & Ascher, D. B. Deep-PK: Deep learning for small molecule pharmacokinetic and toxicity prediction. *Nucleic Acids Res.* **52**, W469–W475 (2024).
52. Saito, Y. et al. Landscape and function of multiple mutations within individual oncogenes. *Nature* **582**, 95–99 (2020).
53. Orth, M. et al. Pancreatic ductal adenocarcinoma: Biological hallmarks, current status, and future perspectives of combined modality treatment approaches. *Radiat. Oncol.* **14**, 141 (2019).
54. Sarantis, P., Koustas, E., Papadimitropoulou, A., Papavassiliou, A. G. & Karamouzis, M. V. Pancreatic ductal adenocarcinoma: Treatment hurdles, tumor microenvironment and immunotherapy. *World J. Gastrointest. Oncol.* **12**, 173 (2020).
55. Zhao, M., Mishra, L. & Deng, C. X. The role of TGF- $\beta$ /SMAD4 signaling in cancer. *Int. J. Biol. Sci.* **14**, 111 (2018).
56. Schöniger, S. & Rüschoff, J. Mismatch repair deficiency and microsatellite instability. *Encyclopedia* **2**, 1559–1576 (2022).

57. González-García, A. et al. RalGDS is required for tumor formation in a model of skin carcinogenesis. *Cancer Cell*. **7**, 219–226 (2005).
58. Hanspers, K., Willighagen, E., Tan, A., Ehrhart, F. & Weitz, E. Chromosomal and microsatellite instability in colorectal cancer. Preprint at (2024). <https://www.wikipathways.org/instance/WP4216>
59. Xie, Y. Z. et al. Pancreatic adenocarcinoma pathway. *Front. Immunol.* **12** (2024).
60. Chan, D. K. H. & Buczacki, S. J. A. Tumour heterogeneity and evolutionary dynamics in colorectal cancer. *Oncogenesis* **10**, 1–9 (2021).
61. Aitchison, A. et al. APC mutations are not confined to hotspot regions in early-onset colorectal cancer. *Cancers (Basel)* **12**, 3829 (2020).
62. Saoudi González, N. et al. Cetuximab as a key partner in personalized targeted therapy for metastatic colorectal cancer. *Cancers (Basel)*. **16**, 412 (2024).
63. Jin, K., Qian, C., Lin, J. & Liu, B. Cyclooxygenase-2-prostaglandin E2 pathway: A key player in tumor-associated immune cells. *Front. Oncol.* **13** (2023).
64. Sánchez-Pla, A., Salicrú, M. & Ocaña, J. An equivalence approach to the integrative analysis of feature lists. *BMC Bioinform.* **20**, 1–13 (2019).
65. Zhang, P. et al. Network pharmacology: Towards the artificial intelligence-based precision traditional Chinese medicine. *Brief. Bioinform.* **25** (2023).
66. Shen, L. GeneOverlap: An R package to test and visualize gene overlaps (2023).
67. Ferreira, A. et al. Crucial role of oncogenic KRAS mutations in apoptosis and autophagy regulation: Therapeutic implications. *Cells* **11**, 2183 (2022).
68. Tolios, A. et al. Computational approaches in cancer multidrug resistance research: Identification of potential biomarkers, drug targets and drug-target interactions. *Drug Resist. Updates* **48**, 100662 (2020).
69. Zhou, W. et al. FordNet: recommending traditional Chinese medicine formula via deep neural network integrating phenotype and molecule. *Pharmacol. Res.* **173**, 105752 (2021).
70. Wang, F. et al. A novel SIK2 inhibitor SIC-19 exhibits synthetic lethality with PARP inhibitors in ovarian cancer. *Drug Resist. Updates* **101077** <https://doi.org/10.1016/j.drug.2024.101077> (2024).
71. El Rhabori, S. et al. Design of new molecules against cervical cancer using DFT, theoretical spectroscopy, 2D/3D-QSAR, molecular docking, pharmacophore and ADMET investigations. *Heliyon* **10**, e24551 (2024).

## Acknowledgements

We thank D3 Drug Tech Lab Pvt Ltd and Crescent Innovation and Incubation Council-Chennai for providing software packages. The authors, Dr. Kiran Balasaheb Aher and Dr. Girija Balasaheb Bhavar, express their sincere gratitude to the Management and Principal of Shri Vile Parle Kelavani Mandal's Institute of Pharmacy, Dhule, Maharashtra (India), for their invaluable support, which greatly contributed to the successful completion of this work.

## Author contributions

S. Mohammed Zaidh – Conceptualization, methodology, and Result Interpretation, Hariharan Thirumalai Vengateswaran – Writing and editing the Original Manuscript, Mohammad Habeeb – Formal Analysis, Kiran Balasaheb Aher – Data curation, Girija Balasaheb Bhavar – Validation, Irfan N\* – Investigation, Resource, Methodology and supervision and K.N.V. Chenchu Lakshmi – Administration.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-91568-x>.

**Correspondence** and requests for materials should be addressed to N.I.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025