Original Article

Structural Comparison for Identifying Protein Hotspots Using PhiDsc Method

Mohamad Hussein Hoballa [1] [ID], Changiz Eslahchi [1,2*] [ID]

[1] Department of Computer and Data Sciences, Shahid Beheshti University, Tehran, Iran
[2] School of Biological Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

Corresponding Author: Changiz Eslahchi, PhD, Professor, Department of Computer and Data Sciences, Shahid Beheshti University, Tehran, Iran. Tel: +98-2122431653, E-mail: ch-eslahchi@sbu.ac.ir

Abstract

Introduction: Somatic mutations in cancer are caused by a complex interaction of many starting and driving factors that work together to create a unique mutational landscape. During tumor growth, the controlled cellular environment restricts the alteration of only a few pathways. As a result, tumors that originate from various cell types frequently display similar genetic alterations. A noteworthy development in recent times is the increased detection of hotspot mutant residues located within particular genes. PhiDsc (Protein Functional Mutation Identification by 3D Structure Comparison), an innovative statistical technique developed for the purpose of detecting functional mutations in proteins that are prone to aberrations, is introduced in this study with a specific focus on the RAS and RHO protein families.

Materials and Methods: By combining 3D structural alignment and recurrence data, PhiDsc determines whether mutated residues within a protein family have the potential to be functionally significant. The protein relationships within families were determined using UniProtKB, and the structural alignment of similar proteins in three dimensions was executed using DALI. The RCSB Protein Data Bank was consulted for the protein structures. The extraction of mutational data for the pertinent proteins was performed using BioMuta. The 3D hotspot database was utilized to identify mutational hotspots within the protein families under investigation. PhiDsc is accessible for free at https://github.com/hobzy987/PhiDsc-DALI.

Results: The PhiDsc method successfully found both known and unknown mutational hotspots and changed residues in the RAS and RHO protein families. These changes are functionally important because they happen in or near active regions and domains that are important for protein-protein interactions.

Conclusions: PhiDsc, an innovative statistical method, effectively detects functional mutations in frequently aberrant genes through the selective targeting of altered residues located in protein families that are highly probable to have functional significance. The present study showcased the ability of PhiDsc to identify mutations that impact the development and advancement of cancer, with a specific focus on the RAS and RHO protein families.

Keywords: Cancer, Protein Structural Alignment, Functional Mutations

Citation: Hoballa MH, Eslahchi C. Structural Comparison for Identifying Protein Hotspots Using PhiDsc Method. J Appl Biotechnol Rep. 2023;10(4):1156-68. doi:10.30491/JABR.2023.383043.1601

Introduction

The initiation of cancer development occurs when genomic alterations and chromosomal abnormalities are acquired. These changes can arise from uncorrected errors that occur during DNA replication, repair, or exposure to mutagens.[1] Certain of these modifications not only expedite the buildup of somatic mutations[2] but also play a mechanistic role in the advancement of cancer. There is a hypothesis suggesting that a subset of cells carrying these "driver mutations" will benefit selectively from them, leading to the enhancement of characteristic features of cancer.[3] The mean quantity of driver mutations present in each tumor differs among distinct types of cancer.[4] On the contrary, it is normally assumed that the majority of somatic modifications, referred to as "passenger mutations," hold negligible to no functional importance.[5] Despite the progress made in the field of cancer genomics, the task of differentiating the scarce driver mutations present in a tumor from the vast background of passenger mutations continues to be formidable.

Mutational hotspots are genomic regions characterized by an elevated frequency of nucleotide substitutions, which are linked to both the genesis and advancement of tumors.[6] Prominent sequencing endeavors, including The Cancer Genome Atlas (TCGA),[7] the International Cancer Genome Consortium (ICGC),[8] and Project GENIE[9] have concluded with an abundance of potential hotspot mutations whose functional significance is uncertain. The data are efficiently represented and structured through the utilization of multiple platforms, such as BioMuta[10] and cBioPortal[11,12] that enable the downloading and examination of extensive cancer genomics datasets. The site of mutations most commonly observed is in exons, or the coding sequence of proteins. The functional impact of these entities is ascertained through two approaches:

direct analysis of their effect on the encoded protein or utilization of in silico bioinformatic methods for prediction.[13,14]

The functional significance of tumor mutations has traditionally been deduced from their statistical recurrence, with the underlying assumption that any aberrant changes identified in tumors are probably coincidental and lack functional implications.[15] Nevertheless, recent studies have revealed that the allocation of passenger mutations throughout the genomes of cancer cells is not coincidental.[16] Conversely, these entities tend to congregate in nucleotide sequence contexts that are impacted by the distinct mutational mechanisms exhibited by each tumor.[17,18] On the other hand, there is a theory that the location of functionally important residues along the protein sequence and the surrounding nucleotides affect the spread of driver mutations.[19,20] It might be hard to find functional mutations just by looking for recurrences because there may be underlying mutational mechanisms that target specific genomic contexts. This phenomenon may lead to the frequent alteration of residues, which may not inherently be involved in the progression of tumors.[21]

A multitude of methods are employed to identify hotspot and driver mutations by analyzing the frequency of mutations within a specific gene across a collection of tumor samples. MutSig[22] and MuSiC[23] are two such examples. In order to improve the detection of mutational hotspots in genes that undergo infrequent mutations, protein-level annotation methods are applied, including local-positional clustering,[24] phosphorylation site inclusion,[25] and data derived from paralogous protein domains.[26] These methods facilitate protein-level annotation,[27] aid in the detection of functional mutations in genes with uncommon mutation occurrences, and contribute to the analysis of three-dimensional protein structures.

Several methods are available for predicting functional alterations across diverse protein sequences and structures. Methods including 3DHotspots,[28] Hotspot3D,[29] Mutation3D,[30] and Signatures of Cancer Mutation Hotspots in Protein Kinases[31] improve our understanding of genetic abnormalities by utilizing the three-dimensional structure of proteins and 3D reconstruction of protein networks.[32] Instead, methods like PinSnps,[33] StructMAn,[34] Hot-MAPS,[35] and SpacePAC,[36] as well as SAAMBE-3D,[37] which use somatic cancer mutation-enriched protein-protein interactions,[38] are more useful. The objective of these methods is to gain an understanding of how mutations affect signal transduction, activation cascade proteins, and the function of the original protein. Methods based on individual protein structures or three-dimensional reconstruction of protein networks have improved the detection of mutational clusters in cancer.[39,40] These techniques have unveiled functional implications, including folding-free energy and protein monomer stability. MutaGene and other methods take into account the specific DNA sequence context in which protein-altering mutations occur.[41]

Some methods need structures as input, while others only need sequences. It can be hard to group methods based on what they need as input, but the results of these methods often show if a predicted mutation happens at a hotspot residue. Still, there are problems to solve, especially when only driver genes are being studied and it's hard to tell the difference between driver-specific mutations and passenger mutations in the same gene. So, it is important to look at protein sequences in the context of a bigger picture in order to figure out how mutations change the p53 protein's functional sites and shape.

To address these issues, we introduce PhiDsc, denoting Protein Functional Mutation Identification by 3D Structure Comparison. PhiDsc circumvents these challenges by identifying cancer-causing changes in a target protein through an assessment of three-dimensional structural similarity, protein folding data, and mutation frequency within the same gene family. By examining frequently altered regions of the protein family and comparing the three-dimensional structures of human wild-type proteins within the same family, PhiDsc identifies potential functional mutations.

PhiDsc integrates these methods by classifying protein families based on functional regions, hotspot mutations, and sequence similarity. Consequently, PhiDsc establishes connections between extensively studied proteins with functional mutations and lesser-explored proteins within the same family, achieved by comparing the three-dimensional structures of related domains.

Materials and Methods
Datasets and Software
UniProtKB: UniProtKB (Universal Protein Knowledgebase)[42] is a comprehensive database that provides users access to a vast amount of protein sequence and functional information. One of the many databases that UniProtKB provides is the Protein Families Database. The Protein Families Database is a collection of protein families and domains curated and classified based on their sequence, structure, and functional similarities. It includes over 20,000 protein families and more than 200,000 domains, covering many biological functions and processes.

Each protein family in the database is annotated with information on its function, subcellular localization, evolutionary history, and links to relevant literature and other resources. The database also includes tools and resources for analyzing and visualizing protein family data, such as sequence alignments, phylogenetic trees, and domain architectures.

The Protein Families Database is an important protein structure and function resource for researchers. It provides information on the relationships between different protein families and their evolutionary histories. It is used in various applications, from predicting protein function and structure to understanding the molecular mechanisms of disease.

RCSB: The Research Collaboratory for Structural Bioinformatics (RCSB)[43] is a non-profit entity committed to enabling the

public to utilize a wide range of data pertaining to the structures of biological macromolecules. At the core of its objectives lies the RCSB Protein Data Bank (PDB), an all-encompassing repository.

The RCSB PDB is the preeminent and most widely utilized repository of three-dimensional structures pertaining to complexes, nucleic acids, and proteins. The application provides users with a suite of tools to analyze and visualize the more than 180,000 experimentally determined structures that it contains.

In addition to the PDB, the RCSB furthers the pursuit of knowledge by offering supplementary resources. Providing an extensive collection of specialized educational materials and resources designed for students, educators, and researchers, the PDB-101 educational portal is an invaluable asset. The Ligand Explorer application facilitates the investigation of macromolecular structures by providing users with the ability to search for particular ligands that are of interest within the binding sites. In addition, knowledge is enhanced through the assistance of the RCSB Chemical Component Dictionary, which details the chemical components that make up biological macromolecules.

The RCSB is dedicated to the promotion of scientific knowledge and education through the distribution of superior structural data and resources. Its primary objective is to facilitate unrestricted entry into the vast collection of structural data at the disposal of the scientific community, thereby aiding in the investigation of the composition, operation, and interrelationships of biological macromolecules.

BioMuta: The BioMuta dataset[44] is a comprehensive knowledge base of somatic mutations identified in human cancers. It includes information on over 4.5 million mutations from over 36,000 tumors across 37 cancer types.

Each mutation in the BioMuta dataset is annotated with its genomic location, nucleotide change, amino acid change, and functional impact. In addition, the dataset provides information on the frequency of each mutation in different cancer types and its association with specific pathways and functional domains.

The BioMuta dataset also has annotations for mutations that might be clinically important, like those that change drug targets or are linked to drug resistance. This information can be used to inform the development of personalized cancer therapies. BioMuta functions as a curated repository of single-nucleotide variations associated with cancer, aggregating information from an array of sources. The variations have been systematically gathered from reputable databases such as COSMIC,[46] ClinVar,[47] CIVIC,[48] and UniProtKB. Utilizing BioMuta, all the protein family's residues members are annotated with mutational and hotspot data (version 4)[10] and 3Dhotspots.[39] By proactively procuring information from scholarly articles and utilising automated analyses of publicly accessible datasets, such as TCGA[7] and ICGC,[8] the database is enhanced. Three-dimensional hotspots manifest as a discernible characteristic of BioMuta. Statistically significant mutations are encapsulated within these hotspots, which are strategically arranged within three-dimensional protein structures. Notably, these mutations are identified in the context of cancer. By identifying particular mutational sites in the dataset, hotspot mutations furnish an invaluable asset in the quest to comprehend mutational trends that are linked to cancer. Therefore, the BioMuta dataset is a valuable resource for cancer genomics researchers, providing a comprehensive view of the somatic mutations that occur in human cancers and their potential clinical significance. It has been used in numerous studies to identify potential therapeutic targets and biomarkers for cancer diagnosis and prognosis.

DALI: The DALI (Distance-matrix ALIgnment)[45] algorithm is extensively employed in the comparison and alignment of protein structures. It operates on the principle of utilizing a distance matrix, which is a matrix comprising the pairwise distances between every pair of residues in a protein structure. The Dali algorithm identifies similar regions in two protein structures by utilizing their distance matrices and a heuristic approach. The initial step involves the computation of dihedral angles, a collection of structural descriptors that characterize the local geometry of protein structures. The alignment of these structures is then accomplished via dynamic programming, with the alignment score optimized by minimizing the distance between equivalent residues in the two structures.

Numerous studies have utilized the Dali algorithm to identify functionally significant regions in proteins, predict the structure and function of proteins, and aid in drug discovery by identifying potential small-molecule inhibitor targets. This algorithm is extensively used in structural biology due to its high precision and capacity to detect similarities among distantly related proteins.

PHIDSC Algorithm

The PHIDSC method revolves around a protein denoted as P, characterized by a recognized three-dimensional structure and m amino acid residues. This method encompasses six distinct stages. Initially, leveraging data from UniProtKB, the entire ensemble of proteins belonging to the same family as P is identified and consolidated into a list termed B(P). Subsequently, all human proteins in B(P) with three-dimensional structures (PDB) from the Protein Data Bank are selected to form the set A(P).

The resultant matrix, denoted as E(P), delineates the outcomes when aligning the three-dimensional structures of the target proteins in A(P) with the three-dimensional structure of P. Subsequent to this alignment, a probability is computed for each residue of P through the use of the 3Dhotspot database and BIOMUTA V4. This computation involves retrieving mutational information for each protein

in A(P). The derived probabilities and scores undergo thorough examination to pinpoint any potential functional mutations in protein P. The ensuing sections offer a detailed elucidation of each phase integral to the entire process.

Step 1: Define the Protein List A(P)
In the first step of the algorithm we do:

- The database utilization involves the application of the UniProtKB database in order to ascertain the proteins that are linked to a particular protein, represented as P, and its associated family.
- Utilizing the Protein Data Bank at RCSB (PDB) is how the three-dimensional structure of protein P is determined. The structures of both wild-type and modified proteins are contained in this repository.
- In order to optimize the alignment phase, the algorithm chooses the variant of protein P that has undergone the fewest modifications, thereby guaranteeing that it does not contain any mutations. The selected variant precisely matches the length of P.
- The final stage, known as list formation, culminates in the generation of the protein list A(P), which is represented as $A(P) = \{P_1, P_2, P_3 \ldots P_n\}$.

Step 2: 3D structure alignment
In this step of the algorithm the three-dimensional structures of the target proteins within A(P) are aligned with the three-dimensional structure of the reference protein, P. The alignment is conducted using Dali method.

Step 3: Define Matrix E(P)
In the third step of the algorithm we do:

- Matrix Construction: The matrix E(P), denoted as E(P) = $[a^j_{k_{ij}}]$, is formed with m rows representing the number of amino acids in protein P and n columns representing the number of proteins in A(P).
- Element Representation: Each element, $a^j_{k_{ij}}$, within the matrix denotes the type of amino acid present in the sequence of protein P_j that aligns with the ith amino acid in protein P.
- Index Explanation: Here, k_{ij} signifies the position number of the amino acid in the sequence of protein P_j that aligns with the ith amino acid in protein P.

The matrix E(P) serves as a comprehensive representation of the aligned amino acid residues across the protein family, aiding in subsequent analyses.

Step 4: Obtain Mutational Details for Each Protein in A(P)
The fourth step of the algorithm involves accessing the BioMuta database in order to obtain exhaustive information regarding mutations that impact each protein in the protein family (A(P)). As a result, every residue in the protein family is systematically annotated with mutational and hotspot data acquired from BioMuta.

This results in a comprehensive representation of the mutational landscape within the protein family, which can be utilized as a foundation for subsequent research.

Step 5: Rank the Residues
In the fifth step of the algorithm each amino acid within the members of A(P) is assigned a score based on the specific mutational data associated with that amino acid. The scoring is as follows:

Let a^t_k be the kth amino acids of protein P_t. Define:

$$m(a^t_k) = \begin{cases} 1 & \text{if } a^t_k \text{ is reported as mutation in biomuta} \\ 2 & \text{if } a^t_k \text{ is reported as hotspot in 3Dhotspots database} \\ 0 & \text{otherwise(either non} - \text{aligned or not mutated)} \end{cases}$$

Let the ith row of the matrix E(P) be $[a^1_{k_{i1}}, a^2_{k_{i2}}, \ldots, a^n_{k_{in}}]$, $1 \leq i \leq m$. The following score is assigned to i^{th} amino acids of P:

$$S(i) = \sum_{j=1}^{n} m\left(a^j_{k_{ij}}\right)$$

The probability associated with each acquired score S(i) at each position (row in the matrix E(P)) is determined to assess its statistical significance. Let protein P_t have m_t amino acids of which l_t are mutated in BioMuta. Define:

$$P(a^t_k) = \begin{cases} \dfrac{l_t}{m_t} & , m(a^t_k) > 0 \\ 1 - \dfrac{l_t}{m_t}, & m(a^t_k) = 0 \end{cases}$$

To differentiate between non-mutated and non-aligned residues (both with score $m(a^t_k) = 0$), and considering that functional changes are encoded in the alignments, an assumption is made. Then, if in $a^j_{k_{ij}}$ (j), j is a gap, we assume $P\left(a^j_{k_{ij}}\right) = 1$.

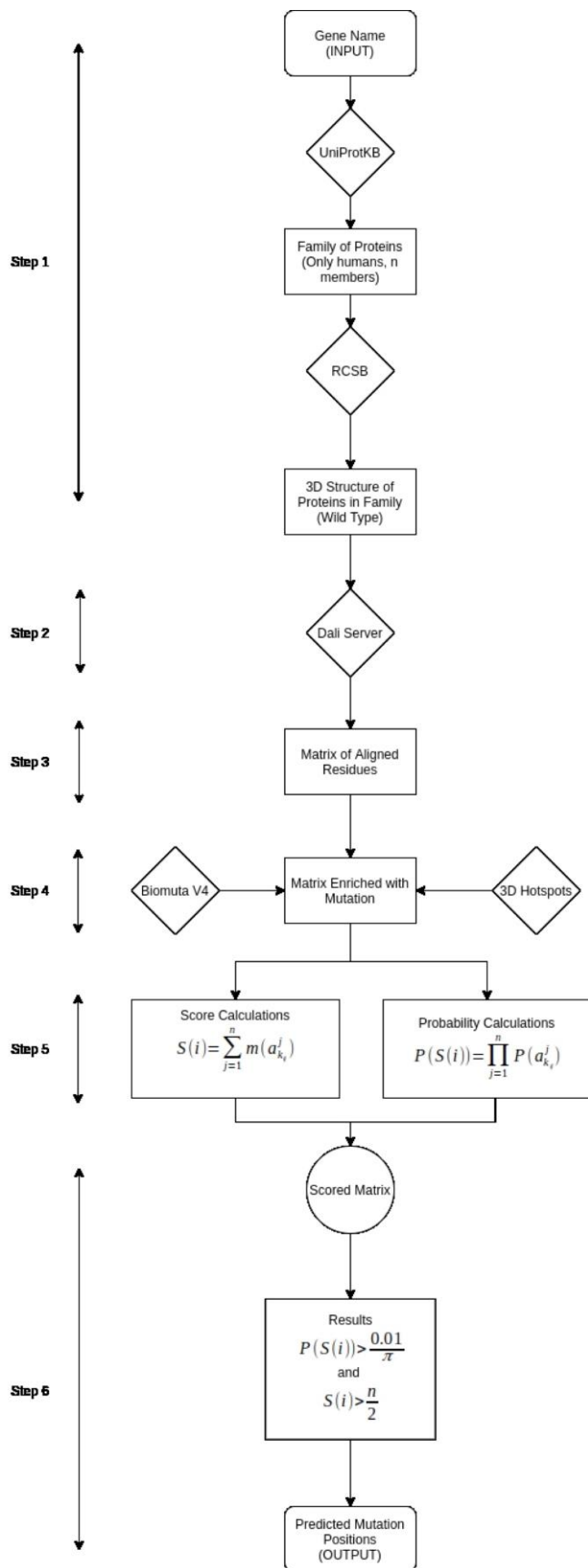Now the overall probability P(S(i)) is calculated as:

$$P(S(i)) = \Pi_{j=1}^n P(a^j_{k_{ij}})$$

This step establishes a ranking system based on the mutational data, providing insights into the significance of amino acids in the context of functional changes.

Step 6: Select candidates.
In the final step of the algorithm the ith amino acid of protein P

Figure 1. The PHIDSC Technique. The method starts by gathering family members; the algorithm then gets the 3D structures from RCSB, pairs members with the input protein, and then enriches mutations in the alignments before calculating scores and probability.

is chosen as a candidate functional mutation based on the following conditions:

- $P(S(i)) < \frac{0.01}{n}$, (following the Bonferroni correction)
- $S(i) > \frac{n}{2}$

This step ensures a stringent selection process, where candidates are chosen if their associated probabilities are below a specified threshold (adjusted for multiple testing) and if the cumulative score $S(i)$ exceeds half of the total number of proteins (n/2). This selection mechanism aims to identify potential functional mutations with a high degree of confidence.

The method is schematically described in Figure 1.

Leave-One-Out Cross-Validation (LOOCV)

In the LOOCV approach, one data point is systematically excluded from the training set in each iteration. The model is trained on n−1 samples, with the excluded sample serving as the validation set. The predictive accuracy of the model is evaluated for the omitted data point in each iteration, and this process is repeated for all possible combinations, resulting in n iterations. The overall efficacy of the model is determined by averaging the errors across all iterations, providing a comprehensive assessment of its generalization capability. For a specific protein P, $A_i(P) = \{P_1, P_2, P_3,…P_n\} - \{P_i\}$ is utilized to generate PHIDSC predictions. This involves using the protein family set $A_i(P)$ in each iteration to predict functional mutations for P. The robustness of these predictions is evaluated by collecting anticipated functional mutations for each iteration ($1 \le i \le n$). A predicted functional mutation is considered robust if it is consistently indicated in at least 80% of all rounds, ensuring reliability and consistency in the prediction outcomes.

Residue Interaction Network

In order to assess the physical consequences of mutations on the configuration and operation of proteins, the Residue Interaction Network (RIN) approach is applied. Chang et al.[20] established that mutations that transpire in close proximity to other hotspot mutations within the three-dimensional structure of a protein are more probable to be classified as hotspot mutations. User-defined RINs are produced by employing the RINalyzer tool,[49] which is an integral component of Cytoscape,[51] a molecular interaction network analysis platform. These RINs are derived from the three-dimensional protein structure acquired from the RCSB Protein Data Bank. The RINerator, an important part of the RINalyzer, checks the strength of different types of biological interactions,[50] such as hydrogen bonds, contacts/clashes, and hydrogen atoms. RINalyzer, a Java plugin for Cytoscape, facilitates the examination and visualization of

molecular interaction networks through an intuitive interface. The results from cBioPortal,[11,12] which is a curated mutation dataset made up of different cancer samples, are compared to the results of RIN's interacting residues. It is through this comparative analysis that we learn more about how mutations affect protein interactions, which is a major contribution to the field of cancer-associated mutations as a whole.

Results
PHIDSC is applied to HRAS from the RAS[59] subfamily and RhoA from the RHO[60] subfamily of proteins.

HRAS
A(HRAS) = DIRAS1, DIRAS2, GEM, KRAS, NRAS, RAP1A, RAP1B, RAP2A, RASL12, REM1, REM2, RERG, RRAD, RRAS, and RRAS2 was the family group of HRAS.

Dali successfully matched 98 percent of the HRAS residues to the residues of each family member (Table 1), demonstrating the target protein's strong structural resemblance to the families of related proteins. (Additional files HRAS alignment.) As a result, PHIDSC scored 168 of 189 HRAS's residues (89%) and predicted 13 residues as functional mutations (Table 2). All these predicted functional mutations passed cross-validation tests (Figure 2) and were consistently predicted by six different algorithms to be efficient and protein-modifying.
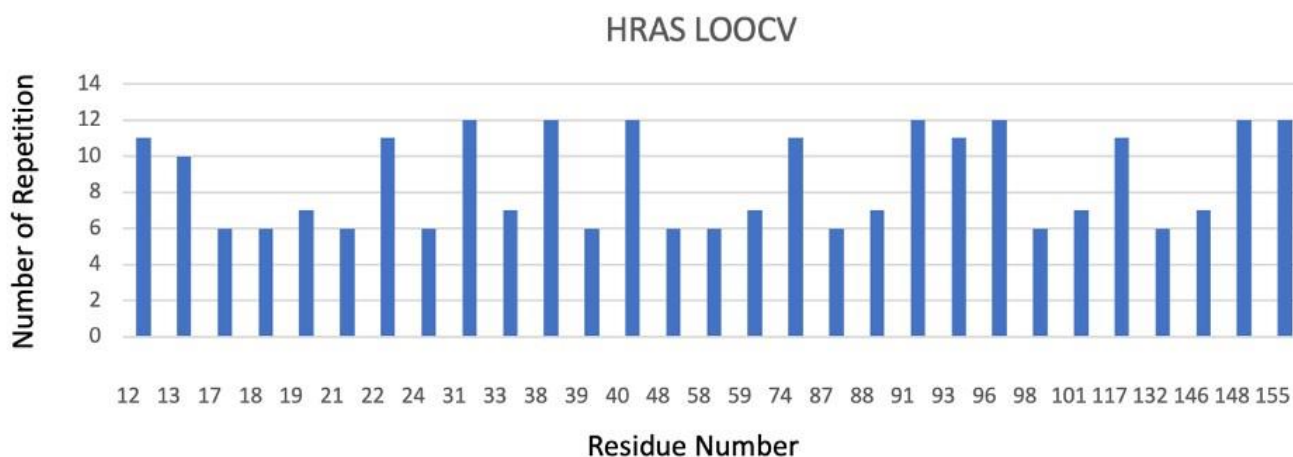
Utilizing the HRAS structure, RIN is produced (RCSB database ID 4Q21, with 168 residues). In the functional domains of the protein (G boxes, Switches I and II, GDI, and GEF interaction sites, and GTP/MG^{2+} binding domain), thirteen putative functional alterations shared 58 nearby residues. Additionally, the cBioPortal reports that 25 of these 58 residues were altered in human cancers.

Table 1. The Data Shows the Degree to Which Each Protein (HRAS-P01112) is Structurally Aligned with the Other Members of its Protein Family

Protein	RALA	RALB	RAP1A	RAP1B	RAP2A	KRAS	RASL12	NRAS	RERG	RIT1	RRAS2	RRAS	Median
PDB ID	2BOV	2KWI	1C1Y	4DXA	1KAO	3GFT	3C5C	3CON	2ATV	4KLZ	2ERY	2FN4	
Alignment	100	100	97.619	98.214	98.214	98.809	95.238	92.857	98.809	92.261	97.619	100	98.214

Table 2. PhiDsc's List of Potential HRAS Functional Mutations. The PHIDSC score p-value and projected interacting residues from the RIN analysis are used to order the locations of the residues. When available, the dbSNP polymorphism ID or COSMIC mutation reference is also included

HRAS										
Residue Nu	p-value	Interacting Residue								Mut Ref Nu
12	2.66E-07	11	16							COSM483
74	1.72E-06	5	70	71	73	75				COSM5991570
13	2.57E-06	117								COSM486
93	6.18E-06	81	82	90	91	113	137			COSM9497546
91	8.72E-06	87	88	90	93	95				COSM6476473
22	1.38E-05	18	19	20	32	26	28	146	149	152	COSM6923245
96	1.54E-05	9	10	11	92	93	97	98	99	100	RS889495169
117	1.85E-05	13	14	83	84	116	119	120	144		CSOM304967
31	3.84E-05	30	33							COSM6915342
40	4.20E-05	20	24	32	38	39	54	55	57		RS763920334
155	5.08E-05	79	144	151	152	153	159				COSM9515051
148	5.23E-05	119	145	150							COSM6903495
38	5.93E-05	39	40	57							RS750680771

Figure 2. The Data Displays the LOOCV for the Protein HRAS; the number of repetitions for each residue in the system's iterations is displayed (>80%), indicating that the system's results are reliable since they are achieved in all LOOCV iterations.

Residues 12, 13, 74, and 93, known to be crucial functionals and often mutated in different cancer types, made up the top four PHIDSC predictions in HRAS.[61] Residues 12 and 13 form a domain interacting with GTP/Mg^{2+} and guanine nucleotide dissociation inhibitors (GDI).[62] Typically, it is seen in malignancies such as bladder cancer,[63] thyroid cancer,[64] and other diseases such as Costello syndrome[61] and Schimmelpenning-Feuerstein-Mims syndrome.[63,65]

Endometrioid cancer and sebaceous carcinoma have mutations in residue 74, but few samples of prostate cancer exhibit mutations in residue 93.[66] The Ensemble Learning Approach for Stability Prediction of Interface and Core Mutations (ELSPIC) claims that,[67] residue 93, is found in the protein's core; it is likely that it has a direct impact on the structure and functionality of the protein.

Despite not being detected in any protein domains, three of the 13 putative functional mutations in HRAS were discovered close to the junction of exons 3 and 4 at position 97. Finally, phosphorylation site residue 96 has been found; the other residues, as indicated in (Figure 3), were found in functional protein domains.
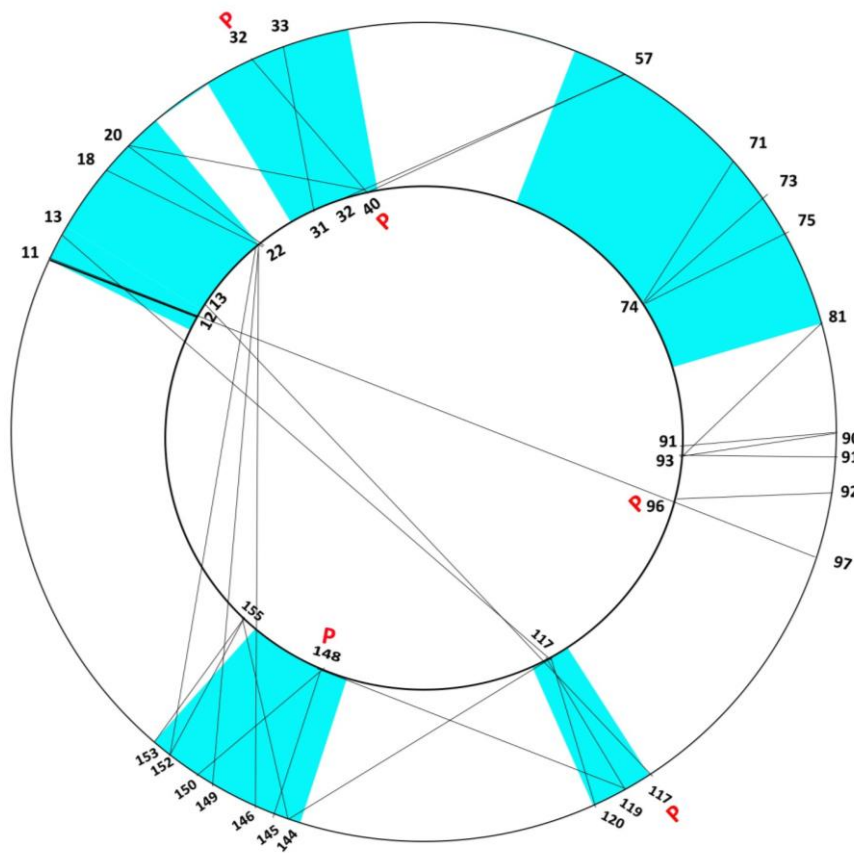
RhoA

RhoA, a member of the RHO[60] subfamily of proteins with

A(RhoA) = {RHOB, RHOC, RHOD, RHOQ, RHOU, RND1, RND3, RAC1, RAC2, RAC3, CDC42}.

The RCSB database retrieves 3D structure files for each member (if found in PDB) of A(RhoA). Table 3 displays the complete list of PDB structures. The input protein and each member of its family are then compared structurally in pairs using the Dali server. In the constructed alignments, 97 percent of the RhoA residues lined up with the residues of each family member. The target proteins support these findings and their respective protein families' striking structural similarities, as an outcome, 179 out of 193 residues were scored for RhoA.

In the last phase, the P-value of PHIDSC statistics is obtained for each target protein residue. For RhoA, eight potential functional mutations were found. The RhoA protein likely functional mutations generated by the PHIDSC method are shown in Table 4. All eight candidates successfully passed the cross-validation process (Figure 4). Remarkably, none of the cancer mutation databases identified any evidence of a mutation in RhoA residue 29. Nevertheless, mutation prediction tools suggest a potential alteration in the functional activity of RhoA, shedding light on the significance of these candidates in the context of cancer-associated mutations.

Figure 3. It Shows the Interacting Residues in the Outer Circle and the Potential Functional Alterations in the Inner Circle. As the canSAR BLACK system claims,[60] HRAS functional regions are shown by the blue areas. In addition, the connections between the outer ring of interacting residues and the inner circle of potential functional mutation reflect residue interactions. Only the HRAS residues altered in cBioPortal are shown in this image.

Table 3. The Data Demonstrates How Each Protein (RhoA) is Structurally Aligned with the Relevant Protein Family Member

Protein	RAC1	RAC2	RAC3	RHOB	RHOC	RHOD	RHOQ	RHOU	RND1	RND3	Median
PDB ID	1E96	1DS6	2C2H	2FV8	2GCN	2J1L	2ATX	2Q3H	3Q3J	2V55	
Alignment	99.441	99.441	93.854	95.53	98.324	87.709	99.441	94.413	97.206	100	97.765

Table 4. The Data Covers all Potential RhoA Functional Mutations Suggested by the PHIDSC Method. The table lists each possible functional mutation's interacting residues in the third column, their position number (P) in the first column, and their P-value in the second. The mutations' "COSM" letters denote that they were marked in the cosmic database as tumor-related mutations. At the same time, the mutations' "rs" letters imply that the Dpsnp database has annotations for them.

		RHOA								
Residue Number	p-value	Interacting Residue								Mut Ref Nu
111	3.07904E-05	78	79	80	109	110	177			COSM2849881
34	4.86819E-05	35								COSM2849895
139	9.956E-05	84	86	89	92	122	139	140	143	COSM2849897
168	0.000147858	170	171	172						COSM7114068
110	0.000209526	77	78	79	80	107	108	11		RS368767616
29	0.000224094	23	27	28	29	31				NO
172	0.000300752	46	48	168	169	172	174	175	176	COSM1309264
127	0.000484266	87	121	124	125	127	129	130	131	MU85445108

By utilizing the RCSB database's structure 1OW3, we constructed the Residue Interaction Network (RIN) for RhoA. In this network, the eight identified potential functional alterations are surrounded by 42 neighboring residues, with 18 of them already recognized as actual mutations according to the cBioPortal database (refer to Table 3 for interacting residues). Notably, RINalyzer data indicates that these neighbors of putative functional mutations are associated with Protein-Protein Interaction (PPI) functionals. These interacting residues, including position 127, are also situated in RhoA protein domains related to interactions with GTPase-activating proteins (GAP), Guanine Nucleotide Exchange Factors (GEF), and Guanine Nucleotide Dissociation Inhibitors (GDI), as well as phosphorylation sites. This underscores the crucial role of residue 127 in influencing the functional activity of RhoA (Figure 5).
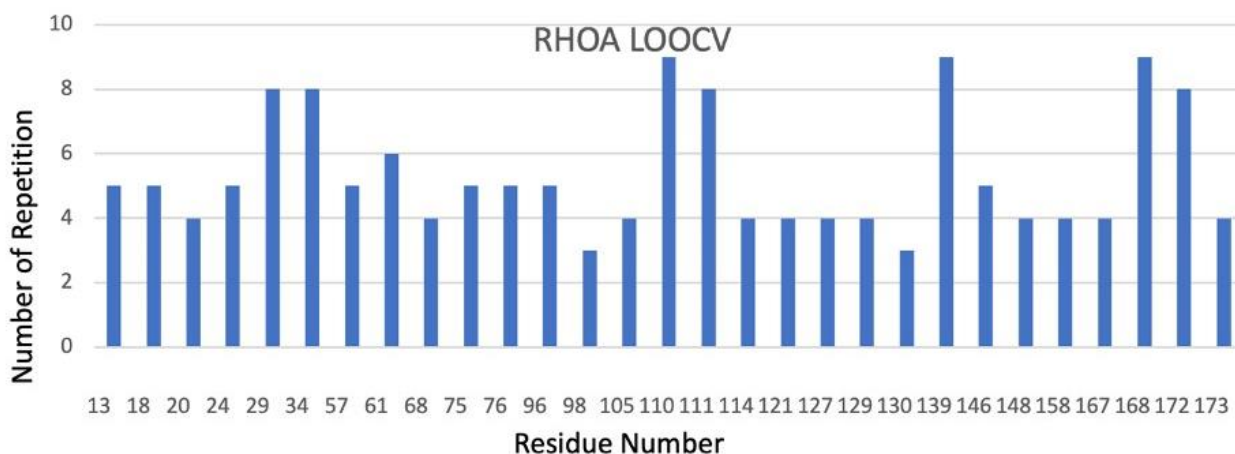
Among cancer samples, four RhoA residues (34, 139, 111, and 168) exhibited high scores (Table 4). The 3D structure analysis of RhoA reveals that residue 34 is in proximity to both the core region and the GTPase-Activating Protein (GAP) interaction site. Mutations at this site, as per data from ELASPIC and COSMIC, enhance the affinity for ARHGAP1, a crucial GAP protein for RhoA activation.
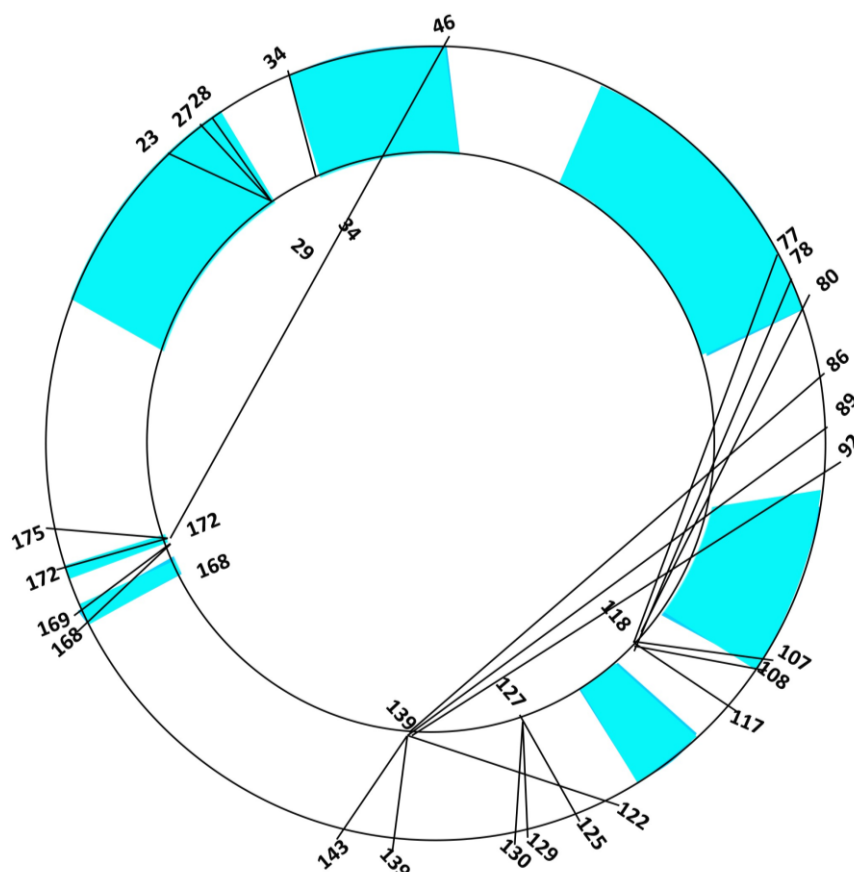
As for the RhoA mutant at residue 139, it was identified in one non-small cell lung cancer sample and two samples each of cervix and stomach cancer, based on COSMIC data. However, in the last two samples, it was not classified as a functional mutation.

Residue 111 mutations were observed in a sample of individuals with stomach cancer, leading to an increased affinity for the CTRO protein, known for controlling cytokinesis by forming a contractile ring.[7]

A mutation in residue 168 was found to disrupt the interaction between RhoA and PKN1/PKN2, two proteins associated with prostate cancer that play essential roles in cell migration and proliferation.[69,70] This mutation also increased affinity with the KAPCA gene, linked to ovarian and breast cancer.[68] These findings shed light on the complex relationships and functional implications of RhoA mutations in various cancer types (Figure 5).

Figure 4. The Data Demonstrates the LOOCV for the Protein RhoA; the number of repetitions for each residue is shown in all iterations of the system, suggesting that the results of the system are robust since the original findings are attained in all LOOCV rounds.

Figure 5. It Demonstrates the Interacting Residues in the Outer Circle and the Potential Functional Alterations in the Inner Circle. RhoA functional zones are shown by the blue areas by the canSAR BLACK system.[60] In addition, the connections between the outer ring of interacting residues and the inner circle of potential functional mutation reflect residue interactions. Only the HRAS residues altered in cBioPortal are shown in this image.

Discussion

Our research investigates protein families classified by UniProtKB that consist of related entities with comparable sequences, structures, and functions. Our hypothesis posits that mutations that are shared among these protein families and are associated with the same cancer phenotype are prevalent in the same domain. This suggests that domains and modifications are common throughout the family. The proposed methodology employs scores as a means of statistically assessing whether these mutations indicate functional alterations in family-shared regions.

In order to authenticate our method, PhiDsc, an innovative approach to detect functional protein mutations, we performed experiments utilizing domains from two widely recognized protein families—HRAS and RhoA—both of which are implicated in the pathogenesis of cancer. We incorporated several factors into our methodology: the location of a mutation within the three-dimensional structure of the protein,[71] the frequency of its recurrence in human tumors,[72] and its correlation with established functional hotspot mutations in paralogous proteins belonging to the same family or possessing comparable domains.[73] In order to optimise statistical performance by reducing false positives, we implemented the Bonferroni restriction. PhiDsc presents

a thorough and validated methodology for linking mutant residues to particular functional regions of biological proteins, thereby providing significant contributions to the understanding of protein alterations associated with cancer.

The performance of PHIDSC was evaluated by means of an analysis of the HRAS and RhoA proteins. HRAS, as classified by the KEGG Pathway, is a GTPase protein that is a member of the RAS subfamily. It is involved in the regulation of a multitude of biological processes across 84 pathways. The residues in HRAS that are most significantly impacted, namely positions 12, 13, and 61, demonstrate correlations with a multitude of cancer subgroups.[73] Sustained HRAS activity is likely responsible for its ability to promote tumor development.

Likewise, according to the KEGG Pathway, RhoA is classified as a signaling G protein belonging to the RHO subfamily, which regulates an extensive array of biological processes spanning 43 pathways. The RhoA residues that undergo frequent alterations, particularly at positions 17 and 42, have been identified in numerous types of cancer.[74] The ongoing activation of RhoA within the protein is responsible for its oncogenic potential.

With the exception of one potential residue in RhoA, all residues identified by PHIDSC exhibited alterations in

cancer samples and various disorders, including Costello syndrome, which is associated with germline HRAS mutations.[75] Despite the rare occurrence of a specific putative functional mutation in cancer mutation datasets and its lack of prior recognition as a hotspot mutation, CanSar black[60] analysis revealed its presence in active functional domains of proteins or its interaction with functional residues within a more extensive network. Our study expanded its investigation beyond the Biomuta database by utilizing tumor samples from additional datasets, including COSMIC,[46] cBioPortal, and Dbsnp. This expansion unveiled potential functional hotspots that had not been detected in Biomuta. Notably, these were discovered to be infrequently modified residues, with the exception of RhoA residue 29. This implies that PHIDSC improves the capability of identifying functional mutations that occur infrequently and were not previously recognized as hotspots. Although the COSMIC and Dbsnp[76] databases did not contain any mutational records for RhoA residue 29, it was determined to be a highly variable location by the mutation analysis software MutaGene.[41] It is noteworthy to mention that COSMIC and Dbsnp utilize distinct curation processes to classify each mutation as an SNP.

As described in the section on findings, these results demonstrate that PhiDsc's predictions have the capacity to influence the structure and function of proteins by inferring the stabilising effect of point mutations using a variety of concepts. Nevertheless, additional experimental verification is essential in order to clarify the precise consequences of undetected mutations.

PHIDSC provided outcomes pertaining to particular hotspots of HRAS and RhoA that have been previously identified in cancer. These hotspots comprised residues 12, 13, and 117 out of the 12 predictions for HRAS, and residue 34 out of the 11 predictions for RhoA. The utility of PHIDSC goes beyond oncogenes; it can also detect functional mutations in tumor suppressor genes or any other protein family, as long as there are a significant number of members, even if the validation proteins selected are oncogenic. The utilized mutation profile data is deemed adequate and dependable.

This method has, nevertheless, two constraints: protein families consisting of a small number of members and the lack of a three-dimensional protein structure. In order to overcome these constraints, it is intended that a forthcoming tool update will prioritize functional protein domains over the entire protein. Additionally, the alignment comparison will incorporate the projected three-dimensional structure of the protein.

Conclusion

The demonstrated method stands out for its ability to enhance our understanding of proteins within the same family, particularly shedding light on the mechanism by which hotspot mutations propagate—a fact that has received relatively limited attention in prior research. Application of this approach to two distinct protein subfamilies (HRas and RhoA) successfully extracted several well-established hotspot mutations.

Nevertheless, it is imperative to acknowledge certain limitations. The method heavily relies on the mutation profile of the family's proteins, demanding a comprehensive dataset for more precise results. Additionally, refinement is needed to ensure the accuracy of the alignment method. When comparing alternative alignment methods, divergent results were observed, with DALI consistently aligning with hotspots previously experimentally validated.

Despite these challenges, the versatility of PHIDSC is noteworthy. Its capabilities extend beyond oncogenes, offering the potential to identify functional mutations in protein types with substantial family sizes, including tumor suppressor genes. The robustness of the approach is fortified by the accurate and exhaustive utilization of mutation profile data.

Continual research efforts are dedicated to method improvement. This includes integrating anticipated structural models into alignment methodologies, with the aim of conducting a comparative analysis to determine the most efficient approach. Furthermore, the evaluation of protein domains, as opposed to entire proteins, is suggested to enhance alignment precision and generate more dependable outcomes. These proposed enhancements underscore a commitment to ongoing progress and innovation in the pursuit of precise identification of functional mutations.

References
1. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz Jr LA, Kinzler KW. Cancer genome landscapes.

science. 2013;339(6127):1546-58. doi:10.1126/science.1235122

2. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976;194(4260):23-8. doi:10.1126/science.959840

3. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. cell. 2011;144(5):646-74. doi:10.1016/j.cell.2011.02.013

4. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. Cell. 2017;171(5):1029-41. doi:10.1016/j.cell.2017.09.042

5. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med. 2018;10:25. doi:10.1186/s13073-018-0531-8

6. Baeissa H, Benstead-Hume G, Richardson CJ, Pearl FM. Identification and analysis of mutational hotspots in oncogenes and tumour suppressors. Oncotarget. 2017; 8(13):21290-304. doi:10.18632/oncotarget.15514

7. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol. 2015;2015(1):68-77. doi:10.5114/wo.2014.47136

8. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, et al. The international cancer genome consortium data portal. Nat Biotechnol. 2019;37(4):367-9. doi:10.1038/s41587-019-0055-9

9. The AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov. 2017;7(8):818-31. doi:10.1158/2159-8290.CD-17-0151

10. Dingerdissen HM, Torcivia-Rodriguez J, Hu Y, Chang TC, Mazumder R, Kahsay R. BioMuta and BioXpress: mutation and expression knowledgebases for cancer biomarker discovery. Nucleic Acids Res. 2018;46(D1):D1128-36. doi:10.1093/nar/gkx907

11. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401-4. doi:10.1158/2159-8290.CD-12-0095

12. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1. doi:10.1126/scisignal.2004088

13. Martelotto LG, Ng CK, De Filippo MR, Zhang Y, Piscuoglio S, Lim RS, et al. Benchmarking mutation effect prediction algorithms using functionally validated cancer-related missense mutations. Genome Biol. 2014;15(10):484. doi:10.1186/s13059-014-0484-1

14. Chen C, Hou J, Tanner JJ, Cheng J. Bioinformatics methods for mass spectrometry-based proteomics data analysis. Int J Mol Sci. 2020;21(8):2873. doi:10.3390/ijms21082873

15. Gonçalves E, Fragoulis A, Garcia-Alonso L, Cramer T, Saez-Rodriguez J, Beltrao P. Widespread post-transcriptional attenuation of genomic copy-number variation in cancer. Cell Syst. 2017;5(4):386-98. doi:10.1016/j.cels.2017.08.013

16. Dietlein F, Weghorn D, Taylor-Weiner A, Richters A, Reardon B, Liu D, et al. Identification of cancer driver genes based on nucleotide context. Nat Genet. 2020; 52(2):208-18. doi:10.1038/s41588-019-0572-y

17. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21. doi:10.1038/nature12477

18. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016; 534(7605):47-54. doi:10.1038/nature17676

19. Chakravorty D, Jana T, Das Mandal S, Seth A, Bhattacharya A, Saha S. MYCbase: a database of functional sites and biochemical properties of Myc in both normal and cancer cells. BMC Bioinform. 2017; 18:224. doi:10.1186/s12859-017-1652-6

20. Chang MT, Bhattarai TS, Schram AM, Bielski CM, Donoghue MT, Jonsson P, et al. Accelerating discovery of functional mutant alleles in cancer. Cancer Discov. 2018;8(2):174-83. doi:10.1158/2159-8290.CD-17-0321

21. Makova KD, Hardison RC. The effects of chromatin organization on variation in mutation rates in the genome. Nat Rev Genet. 2015;16(4):213-23. doi:10.1038/nrg3890

22. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499(7457):214-8. doi:10.1038/nature12213

23. Phan DL, Kim Y, Kim M. Music: Mutation analysis tool with high configurability and extensibility. 2018 IEEE international conference on software testing, verification and validation workshops (ICSTW), IEEE; 2018. pp. 40-46. doi:10.1109/ICSTW.2018.00026

24. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. Bioinformatics. 2013;29(18):2238-44. doi:10.1093/bioinformatics/btt395

25. Reimand J, Bader GD. Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers. Mol Syst Biol. 2013;9(1):637. doi:10.1038/msb.2012.68

26. Miller ML, Reznik E, Gauthier NP, Aksoy BA, Korkut A, Gao J, et al. Pan-cancer analysis of mutation hotspots in protein domains. Cell Syst. 2015;1(3):197-209. doi:10.1016/j.cels.2015.08.014

27. Cantor AJ, Shah NH, Kuriyan J. Deep mutational analysis reveals functional trade-offs in the sequences of EGFR autophosphorylation sites. Proc Natl Acad Sci USA. 2018;115(31):E7303-12. doi:10.1073/pnas.1803598115

28. Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. Genome Med. 2017;9:1-3. doi:10.1186/s13073-016-0393-x

29. Chen S, He X, Li R, Duan X, Niu B. HotSpot3D web server: an integrated resource for mutation analysis in protein 3D structures. Bioinformatics. 2020;36(12):3944-6. doi:10.1093/bioinformatics/btaa258

30. Meyer MJ, Lapcevic R, Romero AE, Yoon M, Das J, Beltrán JF, et al. mutation3D: cancer gene prediction through atomic clustering of coding variants in the structural proteome. Hum Mutat. 2016;37(5):447-56. doi:10.1002/humu.22963

31. Chen W, Li Y, Wang Z. Evolution of oncogenic signatures of mutation hotspots in tyrosine kinases supports the atavistic hypothesis of cancer. Sci Rep. 2018;8(1):8256. doi:10.1038/s41598-018-26653-5

32. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nat Biotechnol. 2012;30(2):159-64. doi:10.1038/nbt.2106

33. Lu HC, Herrera Braga J, Fraternali F. PinSnps: structural and functional analysis of SNPs in the context of protein interaction networks. Bioinformatics. 2016;32(16):2534-6. doi:10.1093/bioinformatics/btw153

34. Gress A, Ramensky V, Büch J, Keller A, Kalinina OV. StructMAn: annotation of single-nucleotide polymorphisms in the structural context. Nucleic Acids Res. 2016; 44(W1):W463-8. doi:10.1093/nar/gkw364

35. Tokheim C, Bhattacharya R, Niknafs N, Gygax DM, Kim R, Ryan M, et al. Exome-scale discovery of hotspot mutation regions in human cancer using 3D protein structure. Cancer Res. 2016;76(13):3719-31. doi:10.1158/0008-5472.CAN-15-3190

36. Ryslik G, Cheng Y, Zhao H. SpacePAC: Identifying mutational clusters in 3D protein space using simulation; 2013.

37. Pahari S, Li G, Murthy AK, Liang S, Fragoza R, Yu H, et al. SAAMBE-3D: predicting effect of mutations on protein–protein interactions. Int J Mol Sci. 2020;21(7): 2563. doi:10.3390/ijms21072563

38. Wong ET, So V, Guron M, Kuechler ER, Malhis N, Bui JM, et al. Protein–protein interactions mediated by intrinsically disordered protein regions are enriched in missense mutations. Biomolecules. 2020;10(8):1097. doi:10.3390/biom10081097

39. Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. Genome Med. 2017;9:4. doi:10.1186/s13073-016-0393-x

40. Banerjee A, Mitra P. Estimating the effect of single-point mutations on protein thermodynamic stability and analyzing the mutation landscape of the p53 protein. J Chem Inf Model. 2020;60(6):3315-23. doi:10.1021/acs.jcim.0c00256

41. Goncearenco A, Rager SL, Li M, Sang QX, Rogozin IB, Panchenko AR. Exploring background mutational processes to decipher cancer genetic heterogeneity. Nucleic Acids Res. 2017;45(W1):W514-22. doi:10.1093/nar/gkx367

42. Breuza L, Poux S, Estreicher A, Famiglietti ML, Magrane M, Tognolli M, et al. The UniProtKB guide to the human proteome. Database. 2016;2016:bav120. doi:10.1093/database/bav120

43. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow GV, et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic Acids Res. 2021;49(D1):D437-51. doi:10.1093/nar/gkaa1038

44. Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol. 2016;34(2):155-63. doi:10.1038/nbt.3391

45. Holm L, Laakso LM. Dali server update. Nucleic Acids Res. 2016;44(W1):W351-5. doi:10.1093/nar/gkw357

46. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45(D1):D777-83. doi:10.1093/nar/gkw1121

47. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-7. doi:10.1093/nar/gkx1153

48. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet. 2017; 49(2):170-4. doi:10.1038/ng.3774

49. Doncheva NT, Klein K, Domingues FS, Albrecht M. Analyzing and visualizing residue networks of protein structures. Trends Biochem Sci. 2011;36(4):179-82. doi:10.1016/j.tibs.2011.01.002

50. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021;30(1):70-82. doi:10.1002/pro.3943

51. Holmås S, Riudavets Puig R, Acencio ML, Mironov V, Kuiper M. The Cytoscape BioGateway app: explorative network building from an RDF store. Bioinformatics. 2020;36(6):1966-7. doi:10.1093/bioinformatics/btz835

52. Stefl S, Nishi H, Petukh M, Panchenko AR, Alexov E. Molecular mechanisms of disease-causing missense mutations. J Mol Biol. 2013;425(21):3919-36. doi:10.1016/j.jmb.2013.07.014

53. Capriotti E, Fariselli P, Casadio R. I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005;33(suppl_2):W306-10. doi:10.1093/nar/gki375

54. Dehouck Y, Kwasigroch JM, Gilis D, Rooman M. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. BMC Bioinform. 2011;12(1):151. doi:10.1186/1471-2105-12-151

55. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40 (W1):W452-7. doi:10.1093/nar/gks539

56. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31(16):2745-7. doi:10.1093/bioinformatics/btv195

57. Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Res. 2006;34(suppl_2):W239-42. doi:10.1093/nar/gkl190

58. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat Commun. 2020;11(1):5918. doi:10.1038/s41467-020-19669-x

59. Kodaz H, Kostek O, Hacioglu MB, Erdogan B, Kodaz CE, Hacibekiroglu I, et al. Frequency of RAS mutations (KRAS, NRAS, HRAS) in human solid cancer. Breast Cancer. 2017;7(12):1-7.

60. Svensmark JH, Brakebusch C. Rho GTPases in cancer: friend or foe?. Oncogene. 2019;38(50):7447-56. doi:10.1038/s41388-019-0963-7

61. Bertola D, Buscarilli M, Stabley DL, Baker L, Doyle D, Bartholomew DW, et al. Phenotypic spectrum of Costello syndrome individuals harboring the rare HRAS mutation p. Gly13Asp. American Journal of Medical Genetics Part A. 2017;173(5):1309-18. doi:10.1002/ajmg.a.38178

62. Gripp KW, Baker L, Robbins KM, Stabley DL, Bellus GA, Kolbe V, et al. The novel duplication HRAS c. 186_206dup p.(Glu62_Arg68dup): clinical and functional aspects. Eur J Hum Genet. 2020;28(11):1548-54. doi:10.1038/s41431-020-0662-4

63. Homami A, Kachoei ZA, Asgarie M, Ghazi F. Analysis of FGFR3 and HRAS genes in patients with bladder cancer. Med J Islam Repub Iran. 2020;34:108. doi:10.34171/mjiri.34.108

64. Pozdeyev N, Rose MM, Bowles DW, Schweppe RE. Molecular therapeutics for anaplastic thyroid cancer. Seminars in cancer biology 2020;61:23-9. doi:10.1016/j.semcancer.2020.01.005

65. Gamayunov BN, Korotkiy NG, Baranova EE. Phacomatosis pigmentokeratotica or the Schimmelpenning-Feuerstein-Mims syndrome?. Clin Case Rep. 2016;4(6):564-7. doi:10.1002/ccr3.570

66. Kaur HB, Salles DC, Paulk A, Epstein JI, Eshleman JR, Lotan TL. PIN-like ductal carcinoma of the prostate has frequent activating ras/raf mutations. Histopathology. 2021;78(2):327-33. doi:10.1111/his.14224

67. Witvliet DK, Strokach A, Giraldo-Forero AF, Teyra J, Colak R, Kim PM. ELASPIC web-server: proteome-wide structure-based prediction of mutation effects on protein stability and binding affinity. Bioinformatics. 2016; 32(10):1589-91. doi:10.1093/bioinformatics/btw031

68. Papanikolaou N, Mantsou A, Kalosidis N. From driver mutations to driver cancer networks: Why we need a new paradigm. Cancer Stud. 2018;2(1):1.

69. Scott F, Fala AM, Pennicott LE, Reuillon TD, Massirer KB, Elkins JM, et al. Development of 2-(4-pyridyl)-benzimidazoles as PKN2 chemical tools to probe cancer. Bioorg Med Chem Lett. 2020;30(8):127040. doi:10.1016/j.bmcl.2020.127040

70. Yang CS, Melhuish TA, Spencer A, Ni L, Hao Y, Jividen K, et al. The protein kinase C super-family member PKN is regulated by mTOR and influences differentiation during prostate cancer progression. Prostate. 2017;77 (15):1452-67. doi:10.1002/pros.23400

71. Mucoz-Maldonado C, Zimmer Y, Medová M. A comparative analysis of individual RAS mutations in cancer biology. Front Oncol. 2019;9:1088. doi:10.3 389/fonc.2019.01088

72. Schaefer A, Reinhard NR, Hordijk PL. Toward understanding RhoGTPase specificity: structure, function and local activation. Small GTPases. 2014;5(2):e968004. doi:10.4161/21541248.2014.968004

73. Parker JA, Mattos C. The K-Ras, N-Ras, and H-Ras isoforms: unique conformational preferences and implications for targeting oncogenic mutants. Cold Spring Harb Perspect Med. 2018;8(8):a031427. doi:10.11 01/cshperspect.a031427

74. Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, et al. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. Nat Genet. 2014;46(6):583-7. doi:10.1038/ng.2984

75. Hartung AM, Swensen J, Uriz IE, Lapin M, Kristjansdottir K, Petersen US, et al. The splicing efficiency of activating hras mutations can determine costello syndrome phenotype and frequency in cancer. PLoS Genet. 2016; 12(5):e1006039. doi:10.1371/journal.pgen.1006039

76. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-11. doi:10.1093/nar/29.1.308