Innov Biosyst Bioeng, 2023, vol. 7, no. 4, 36–47
doi: 10.20535/ibb.2023.7.4.288017

Original article

IN SILICO ANALYSIS OF ANTI-CERVICAL CANCER DRUG OFF-TARGET EFFECTS ON DIVERSE PROTEIN ISOFORMS FOR ENHANCED THERAPEUTIC STRATEGIES

A. Iqbal[1], F. Ali[1], S. Choudhary[1], A. Qayyum[1], F. Arshad[1], S. Ashraf[1], M. Aziz[1], A.U. Shakil[1], S.A. Sehgal[2], M. Hussain[1], M. Sajid[1*]

[1]Faculty of Life Sciences, Department of Biotechnology, University of Okara, Okara, Punjab, Pakistan
[2]Department of Bioinformatics, Islamia University Bahawalpur, Bahawalpur, Punjab, Pakistan

*Corresponding author: infobiotec@uo.edu.pk

Received 25 September 2023; Accepted 20 November 2023

Background. Cervical cancer is a serious medical condition that affects hundreds of thousands of individuals worldwide annually. The selection and analysis of suitable gene targets in the early stages of drug design are crucial for combating this disease. However, overlooking the presence of various protein isoforms may result in unwanted therapeutic or harmful side effects.
Objective. This study aimed to provide a computational analysis of the interactions between cervical cancer drugs and their targets, influenced by alternative splicing.
Methods. Using open-access databases, we targeted 45 FDA-approved cervical cancer drugs that target various genes having more than two distinct protein-coding isoforms. To check the conservation of binding pocket in isoforms of the genes, multiple sequence analysis was performed. To better understand the associations between proteins and FDA-approved drugs at the isoform level, we conducted molecular docking analysis.
Results. The study reveals that many drugs lack potential targets at the isoform level. Further examination of various isoforms of the same gene revealed distinct ligand-binding pocket configurations, including differences in size, shape, electrostatic characteristics, and structure.
Conclusions. This study highlights the potential risks of focusing solely on the canonical isoform, and ignoring the impact of cervical cancer drugs on- and off-target effects at the isoform level. These findings emphasize the importance of considering interactions between drugs and their targets at the isoform level to promote effective treatment outcomes.

Keywords: cervical cancer; isoforms; molecular docking; interaction analysis; bioinformatics approaches.

Introduction

In developing countries, cervical cancer is the main cause of cancer-related deaths and loss of life loss [1]. Several years earlier than the median age at which breast, lung, and ovarian cancers are diagnosed, cervical cancer is commonly diagnosed in the fifth decade of life [2]. Ninety percent of the 270,000 cervical cancer fatalities in 2015 occurred in low- and middle-income countries (LMIC), where mortality was 18 times higher than in developed nations [3]. Nearly all cervical cancers are caused by high-risk subtypes of the human papillomavirus (HPV), and screening and vaccination programs are effective preventive measures against HPV [4]. The two most prevalent histological subtypes (squamous cell carcinoma, and adenocarcinoma) account for 70% and 25% of all cervical malignancies, respectively [5, 6]. The major decrease in cervical cancer mortality has been attributed to the development and implementation of screening programs [7]. Cervical cancer has a poor prognosis following metastasis or recurrence, with a 5-year overall survival (OS) rate of approximately 17% [8]. To improve the efficacy of cervical cancer treatment, it is crucial to identify novel therapeutic targets and survival-associated biomarkers.

Major innovations in large-scale multi-omics research provide a unique perspective for systems biology analyses of the emergence and spread of cancers. HPV contributes to the development of cervical cancer, which is considered to be a virus-driven malignancy. Early HPV infection may simply be a result of external causes, such as changes in the genome, eventually causing cervical epithelial cells to become malignant (e.g., gene fusion, non-coding RNAs, copy number variation, DNA methylation, and somatic DNA mutations) [9–13]. Transcriptomic and epigenetic modifications have been the focus of several prospective studies. However, Alternative splicing (AS) in cancer post-transcriptional protein isoforms has not yet been thoroughly studied.

© The Author(s) 2023. Published by Igor Sikorsky Kyiv Polytechnic Institute.
This is an Open Access article distributed under the terms of the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/), which permits re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

In eukaryotes, a remarkable biological process known as alternative splicing, which promotes proteomic diversity, allows a single gene to express several protein isomers. In humans, where more than 94% of genes are alternatively spliced, the occurrence and properties of alternative splicing are highly diverse [14–16]. This method enables cancer cells to generate abnormal proteins with altered functional domains, which promotes carcinogenesis [17–19]. In malignancies, these domain changes can lead to complicated remodeling and protein-protein interactions. Some essential oncogenic splicing variations can control tumor epithelial-to-mesenchymal transition and biological processes in cancer stem cells [20]. Gene expression is properly controlled in a context-specific manner, even if gene isoforms appear to have different and sometimes even opposing functions.

Aberrant protein isoforms that cause diseases have the potential to be effective drug targets in addition to serving as significant biomarkers [21, 22]. In this study, we examined the effectiveness of FDA-approved drugs against various cervical cancer-related gene isoforms. Using structural analysis and clinical data on the expression of these genes, we curated the drug interaction data for various isoforms of different genes implicated in cervical cancer, and evaluated their effectiveness against protein isoforms. In this study, we primarily focused on cervical cancer and examined whether the drugs were effective against target gene isoforms.

Methods

Collection of genes and their protein isoforms

We identified genes associated with cervical cancer using the COSMIC database [23], an online resource for somatically acquired mutations reported in human cancers. More than 30 genes may contribute to cervical cancer (Supplementary File 1). Based on the number of patient samples, the top five genes out of 30 were selected and used for further analysis. The Ensemble genome database [24] was used to curate the gene isoforms and protein sequences of these genes. Using the COSMIC Mutation ID, mutations were identified in the genes and matched with the variants of each gene isoform using the Ensemble database.

Curation of drugs-target interaction data

Using the Drug Gene Interaction Database (DGIdb) [25], we curated the FDA Approved drugs for our genes. Using this database, more than 40 drugs that received FDA approval were identified. These drugs were retrieved from the Drug Bank [26] and cheMBL [27].

Sequence analysis of gene isoforms

To check the conservation of binding pocket in isoforms of the genes, Binding Pockets of the canonical proteins were predicted through the COACH Server (https://zhanggroup.org/COACH/). We identified domains from the EMBL-EBI InterPro database [28] and aligned them with the sequences of the canonical protein and its protein isoforms. Using the Bioconductor program msa, which offers a selection of alignment techniques and produces alignment plots in LaTeX format, we created numerous sequence alignments. Using the Cluster Omega method in the msa package, we aligned the binding site sequence with all isoforms of the same gene.

Gene isoforms expression in normal and tumors samples

We examined the clinical data offered by UCSC Xena [29] for patients with cervical cancer, which is an online resource for analyzing multi-omics, clinical, and phenotypic data. We used UCSC Xena to compare TCGA tumor samples with normal GTEx samples to evaluate whether protein-coding isoforms are upregulated or downregulated in cervical cancer. The expression of gene isoforms was examined in normal patient samples using GTEx and in tumor samples using TCGA, both of which were drawn from 307 Cervical Cancer Samples available in the UCSC Xena database. We also visualized the exon structure of the gene isoforms to better understand the pattern of alternative splicing in various isoforms of the genes.

Structure prediction of protein isoforms and ligand docking

To better understand the associations between proteins and their ligands (drugs), we predicted the 3D structures of protein isoforms using a number of tools for structural level study of the different isoforms of the proteins. Protein isoform structures were predicted using the structure prediction tools trRosetta [30], Robetta [31], Swiss-Model [32], and I-TASSER [33]. Furthermore, the ERRAT quality factor and the favored, allowed, and disabled regions in the Ramachandran plot were used to evaluate the predicted structures. After evaluation, we used SiteMap53 [34] to determine the drug target region in those protein isoforms'

3D structures. The predicted 3D structures of the protein isoforms were prepared for docking analysis using Chimera 1.15 rc. We used Pyrex software to perform ligand-protein docking analysis, and considered a number of drugs that have already been approved for such proteins to check the effectiveness of these drugs against various protein isoforms that are affected by disease. Poses of the protein–ligand complexes were captured to further analyze the pocket sizes, shapes, and electrostatic surfaces of the docked protein isoforms.

Interaction analysis

The Discovery Studio 2021 Client was used to examine the protein-ligand complexes. We examined how the drug, which has high binding affinity for canonical proteins, interacts with different protein isoforms. Furthermore, we examined the interactions between the hydrophobic and hydrogen sites in different docked protein isoforms.

Results

Drugs Target Genes have multiple isoforms

More than 30 genes linked to cervical cancer were identified to have missense mutations (Supplementary File 1). Five genes were selected for further analysis considering the number of patient samples. We identified FDA-approved drug interactions to analyze interactions between drugs and their target protein isoforms. We retrieved more than 145 entries belonging to five distinct Cervical Cancer genes.

A partial list of the summary tables is presented in Table 1. We found that the bulk of the candidate genes had two or more transcribed spliced variants and protein isoforms (Fig. 1).

Our findings demonstrate that the majority of cancer drug target genes undergo splicing and produce many gene isoforms that may be functionally distinct and react with drugs in different ways, highlighting the significance of obtaining protein isoforms and alternative splicing in drug development.

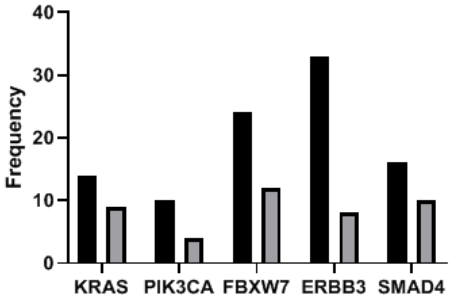

Figure 1: Number of transcript variants and protein-coding isoforms of canonical proteins: ■ – transcripts variants, ▭ – protein isoforms

Table 1: FDA Approved Drugs against target genes and protein-coding isoforms

Genes	Abbreviations	FDA approved drugs	Number of transcript	Number of protein isoforms	References
KRAS	Kirsten rat sarcoma virus	Cetuximab, AZD-4785, Selumetinib, CC-223, AZD-8835, PD-0325901, Trametinib, Ridaforolimus	14	9	[35–38]
SMAD4	Mothers against decapentaplegic homolog 4	Lysine, Sapanisertib, Fluorouracil, Alectinib, Crizotinib, Cetuximab, Gemcitabine, Irionotecan, Carboplatin, Placlitaxel	10	4	[39–42]
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	Trastuzumab, Temsirolimus, Serabelisib, Taselisib, CC-223, INK-1117, Alpelisib, Buparlisib, Capivasertib	16	10	[43, 44]
ERBB3	Human epidermal growth factor receptor 3	Pertuzumab, Trastuzumab, MM-121, AV-203, AMG-888, Patritumab, Duligotuzumab, Sapitinib, MM-111	33	12	[45–47]
FBXW7	F-Box and WD Repeat Domain Containing 7	Temsirolimus, Sirolimus, Regorafenib, Vorinostat Belinostat, Entinostat, Docetaxel, Vorinostat, AR-42	24	8	[35, 48, 49]

Differences in binding pockets among protein isoforms

Using several sequence alignments, we identified the precise interaction residues in the drug-binding region of each protein isoform. We performed multiple sequence alignments between the Pfam functional domains, canonical proteins, protein isoform sequences, and predicted protein-binding pockets. Here, we describe the sequence alignment plots of a few genes.

Cellular functions essential for cancer development, such as cell growth, proliferation, motility, survival, and metabolism, are regulated by Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha (PIK3CA) [50]. PIK3CA has four protein isoforms (PIK3CA-201, PIK3CA-203, PIK3CA-204, and PIK3CA-205). Protein isoforms PIK3CA-203 & 204 have 21 and 118 residues, respectively. Comparison with the predicted pocket binding of the canonical protein. It was found that these residues lack residues in the binding pocket, which indicates that these residues completely lack the predicted pocket binding (Fig. 2). The canonical proteins and protein isoforms, PIK3CA-201 and 205, were found to have identical sequences in the predicted binding pocket. However, we found variations in the C-terminal regions and domain PF00454 of the protein isoforms PIK3CA-201 and 205 (Fig. 2). We examined the C-terminal region of the canonical protein, PIK3CA-201 & 205, and the Pfam domain PF00454 to further explain this variation. In previous studies, we have found that the C-terminal region is necessary for catalysis. This has been suggested to be a crucial PI3Ks regulating component [51]. The Pfam domain is a domain of the p100α catalytic subunit of PIK3CA. However, in USP13-PIK3CA, the entire C-terminal region was replaced by USP13, which affected its catalytic activity. Since PIK3CA-201 and PIK3CA-205 have the same upstream regions, the fusion proteins produced by the two protein isoforms should ideally have the same structure. Additionally, we aligned two other USP13-PIK3CA protein sequences in the FusionGDB database to support this claim, and all sequences had overlapping interference residues with the predicted pocket binding (Supplementary File 2). These sequence-level data indicate that the drug may target all

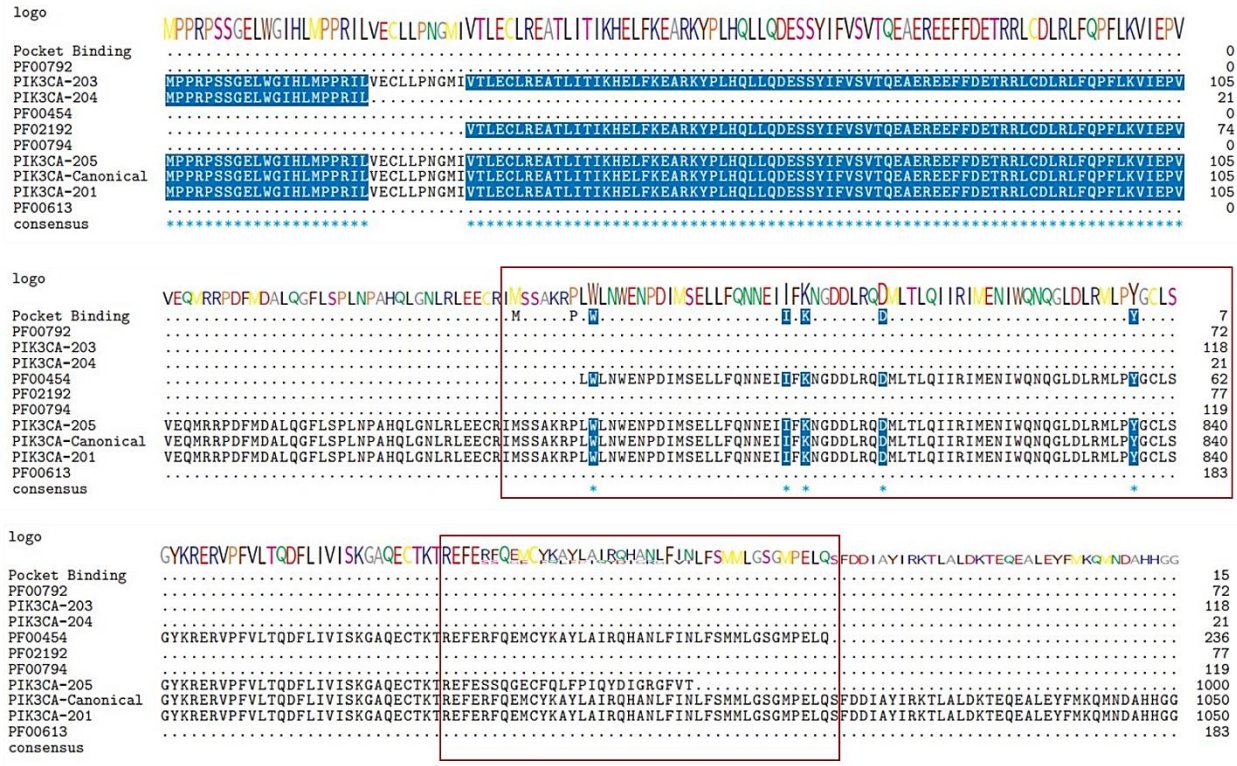

Figure 2: Sequence alignments of predicted pocket binding residues of PIK3CA isoforms. The predicted binding pocket residues; aligned Pfam domains; and PIK3CA-201, PIK3CA-203, PIK3CA-204, and PIK3CA-205 are shown from top to bottom. Each line included the sequence logo of the consensus sequences at the top. Residues in the sequence that coincided with anticipated binding residues are shown in blue. Purple color suggests that this residue is conserved in approximately 50% of all sequences. Similar amino acids are shown under the pink shading

USP13-PIK3CA fusion protein splice-variant protein isoforms. As a result, splice-variation within the PIK3CA gene does not influence binding to its targets in protein isoforms PIK3CA-201 and 205, whereas it may affect PIK3CA-203 and 204, which do not have a predicted binding pocket.

In 285 cervical cancers, PIK3CA 16 targetable oncogenic mutations were found to be the most common oncogenic mutation, with KRAS mutation coming in second. However, despite significant efforts, cancers with KRAS mutations remain challenging to treat because of the plasticity of tumor cells and acquisition of additional mutations. Multiple sequence alignments of KRAS protein isoforms (KRAS-201, 202, 205, 203, 204, 210, and 214) indicated that isoforms 201, 202, 205, and 203 have binding residues and are thus likely targets of drugs, while isoforms 204 and 207 lack binding pockets and are not predicted to be targets of KRAS-targeting drugs, as shown in Fig. 3. Further investigation revealed variations in KRAS isoforms 202, 205, 203, and 204 at the C terminal compared to KRAS isoform 201. These findings suggest that further efforts are required to specifically target KRAS protein isoforms.

High gene isoforms expression in tumor tissues

Using clinical information from the UCSC Xena, which is accessible through several projects (TCGA, GTEx, and TARGET), we were able to determine the expression of gene isoforms. We observed the expression of PIK3CA and KRAS gene isoforms in TCGA samples of cervical and breast cancers, as shown in Fig. 4A. The expression of the gene isoforms was nearly the same in both the cancer types. The gene isoform (PIK3CA-204/ENST00000477735.1) is not expressed in tumor and normal samples, and is thus ignored. The gene isoform (PIK3CA-203/ENST00000468036.1) was highly expressed in TCGA tumor samples, in contrast to normal GTEx samples. Although we previously found that gene isoform-203 does not have a predicted binding pocket, we observed that tumor cells expressed it. This should be included in future studies to examine the on- and off-target effects of these drugs.

Using transcriptome expression data from TCGA repository, it was possible to compare the expression of KRAS gene isoforms

KRAS-202/ENST00000311936.7,
KRAS-203/ENST00000556131.1,
and KRAS-204/ENST00000557334.5)

in cervical and breast samples (Fig. 4B). Compared to normal samples, tumor samples were shown to have higher levels of KRAS-203 expression. Sequence analysis of FBXW7, ERBB3, and SMAD4 is shown in Supplementary File 3. Future studies analyzing the on- and off-target effects of drugs should consider these gene isoforms, as they are expressed in tumors.

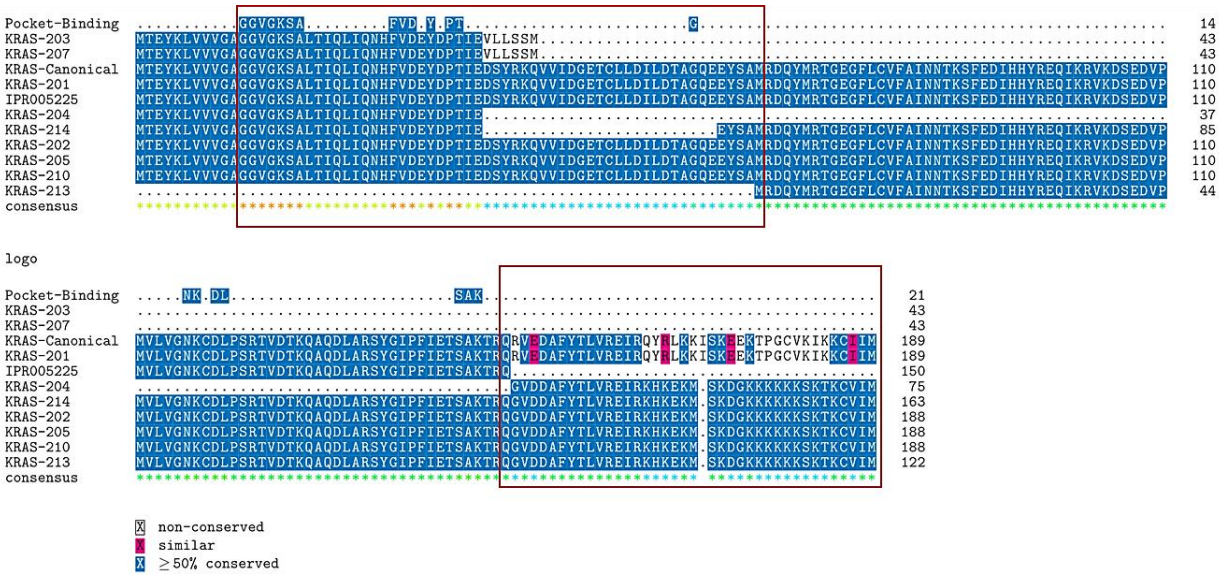

Figure 3: Sequence alignments of the predicted pocket-binding residues of various KRAS protein isoforms. Using Bioconductor software msa, Cluster Omega was used to align the binding residues with the protein isoform sequences. The predicted binding pocket residues, aligned Pfam domains, and KRAS protein isoforms are shown from top to bottom. Each line included the sequence logo of the consensus sequences at the top. Residues in the sequence that coincided with anticipated binding residues are shown in blue. Purple color suggests that this residue is conserved in approximately 50% of all sequences. Similar amino acids are shown under the pink shading

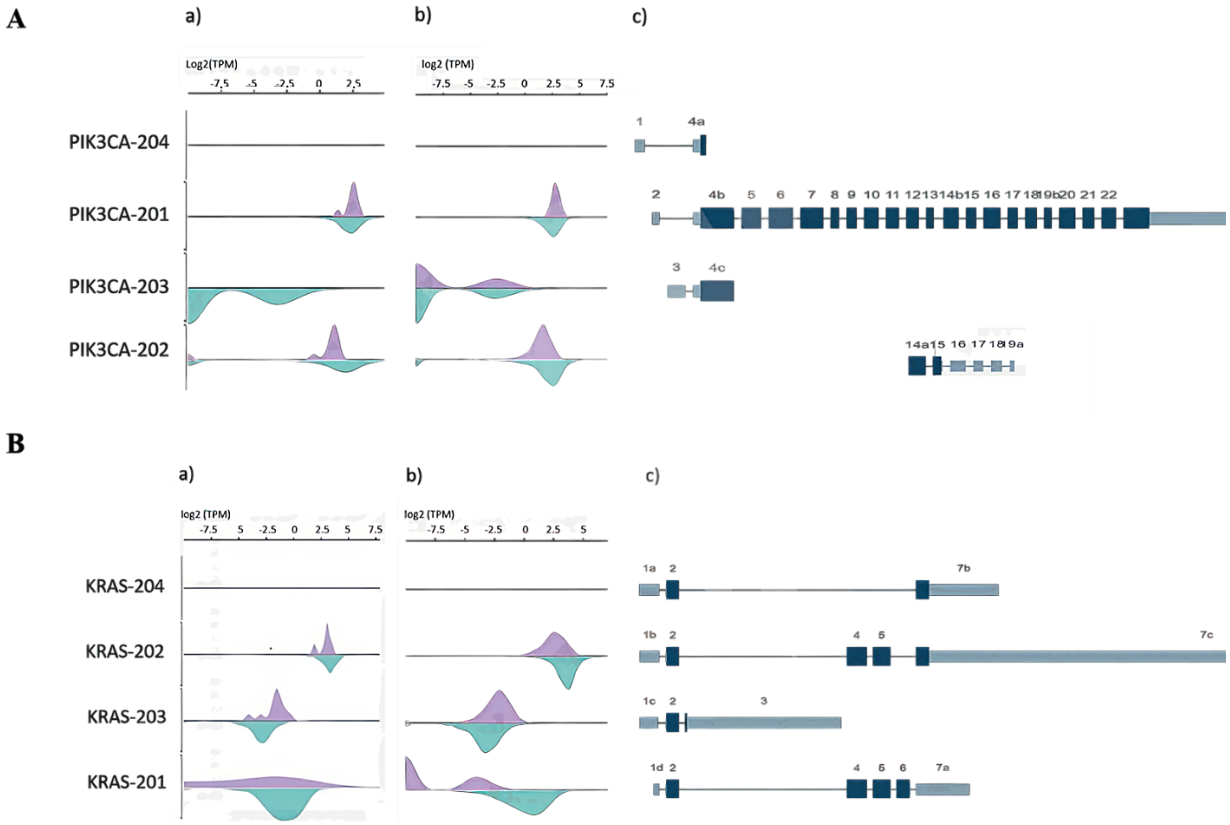

Figure 4: PIK3CA isoform expression and exon structure (A) and KRAS isoform expression and exon structure (B). Green density represents log2(TPM) from normal GTEx samples, whereas purple density represents those from (a) TCGA Cervical Cancer samples and (b) TCGA Breast Cancer samples. Density plots and c) exon-structure plots following the same sequence. Each plot was generated using the UCSC Xena browser [52]

Interactions of drugs on structural level

Although we have shown changes in binding pockets across gene isoforms at the sequence level, structural-level research is the only way to gain more solid proof that drugs bind to their target protein isoforms in distinct ways. We studied the KRAS gene, which has seven distinct protein isoforms, together with known drugs that target them, to understand how a certain drug molecule interacts with several isoforms of a protein.

The three-dimensional (3D) structure of each protein isoform was predicted using various databases. The best predicted structures were projected to have ERRAT scores greater than 94. The structures with poor ERRAT values were further improved.

Next, using Pyrex, we conducted docking analysis while considering the selection of drugs that have been identified to target this disease protein target. After analyzing the docked positions, we observed that, although some drugs bind similarly to protein isoforms, others bind extremely differently. For instance, protein isoformsKRAS-203, 204, and 207 showed low binding affinity with FDA Approved drugs (Table 2). This finding supports our previous finding that these protein isoforms have very small sequences and do not have a predicted binding pocket. All the other protein isoforms of KRAS (KRAS-201, 202, 205, 210, 213, and 214) have high binding affinities. AZD-4785 had good KRAS-201, 202, 205, and 214 scores, respectively. These six protein isoforms had strong binding affinities for Trametinib, although KRAS-202 had a low binding affinity. All protein isoforms showed good binding affinity with ridoforolimus. While the remaining drugs showed good binding affinities with these protein isoforms, certain protein isoforms displayed lower affinities.

In the case of PIK3CA, the protein isoforms PIK3CA-203 and 204 showed low binding affinity for approved FDA Drugs, as these protein isoforms have short sequences and do not have a predicted binding pocket (Table 3). PIK3CA-201 & 205 showed the best binding affinity for drugs. Temsirolimus showed good binding affinity with all protein isoforms.

Table 2: Binding Affinity Values of the KRAS-Canonical protein and its protein isoforms

Drugs	KRAS-Canonical	KRAS-201	KRAS-202	KRAS-205	KRAS-214	KRAS-213	KRAS-210	KRAS-203	KRAS-204
AZD-4785	–6.9	–7.3	–7.2	–7	–7.2	–6.4	–7.9	–5.7	–4
AZD-8835	–8.3	–8.2	–8.2	–8.7	–7.9	–7.7	–8.5	–6.3	–4.9
CC-223	–7.2	–7.7	–7.5	–7.5	–7.7	–7.3	–7.7	–6.1	–4.5
PD-0325901	–7.7	–7.3	–6.9	–6.9	–7.4	–6.3	–6.8	–5.4	–4.3
Ridaforolimus	–10.1	–10.2	–9.7	–9.6	–10.6	–10.7	–9.6	–8.8	–5.3
Selumetinib	–7.2	–7.2	–7.6	–7.7	–7.1	–6.6	–7.4	–5.9	–4.8
Trametinib	–8.2	–9	–7.9	–9.6	–9	–8.6	–8.9	–6.9	–5.5

Table 3: Binding affinity values of the PIK3CA-Canonical, PIK3CA-201, 205, 203, and 204

Drugs	PIK3CA-Canonical	PIK3CA-201	PIK3CA-205	PIK3CA-203	PIK3CA-204
CC-223	–8.6	–8.3	–8	–6.4	–6.4
ALPELISIB	–8.9	–8.8	–9.5	–7.5	–7.7
BUPARLISIB	–8.6	–8.2	–8.1	–6.2	–6.4
CAPIVASERTIB	–9.5	–9.6	–8.9	–6.6	–6.6
INK-1117	–9	–9	–9	–6.8	–6.8
SERABELISIB	–8.9	–9.1	–9	–6.8	–6.8
TASELISIB	–9.7	–9.7	–8.2	–7.2	–6.9
TEMSIROLIMUS	–9.5	–11.6	–10.2	–9	–9
TRASTUZUMAB	–10.5	–9.6	–9.6	–6.8	–7.7

Table 4: Hydrogen and Hydrophobic interactions of the docked protein isoforms with drugs

Protein	Hydrogen Interactions	Hydrophobic Interactions
PIK3CA-Canonical	GLU, ASN, ASP, ASP, TYR	THR
PIK3CA-201	ASP, ASN, LYS, SER	LYS, ASP, ASN, PRO, GLN
PIK3CA-205	ARG, ASP, ASP, LYS, PHE	GLU, TYR, LYS
PIK3CA-203	SER,THR	ARG, GLU

To explain how different pocket sizes, shapes, and electrostatic potential surfaces may create an illusion similar to the binding mode even when the scores are the same in some instances. Here, we examined the temsirolimus binding mode in all four protein isoforms and discovered that while the binding scores were similar, the binding patterns varied greatly, as shown in (Fig. 5). The molecular docking results for FBXW7, ERBB3, and SMAD4 are shown in Supplementary File 4. These results led us to hypothesize that, despite the identity of the ligand-binding residues, the binding pocket structures change in size, form, and dynamic properties, resulting in different binding patterns for a single drug in several protein isoforms with various binding affinity values.

Interaction analysis of the target protein isoforms was performed to determine the type and number of interactions between the docked tesmilorous and PIK3CA protein isoforms. When a complex has a significant number of hydrogen bonds together with a small number of salt bridges, hydrophobic contacts, and $\pi-\pi$ interactions, it is said to be strong. To determine the number of interactions generated by each molecule, we tested each docked drug differently (Fig. 6). According to the interaction study, complexes with strong binding affinities produced the most hydrogen bonds (Table 4).

PIK3CA-Canonical and protein isoforms 201 and 205 were shown to have strong interactions, whereas the docked PIK3CA-203 complex was found to have weak interactions.

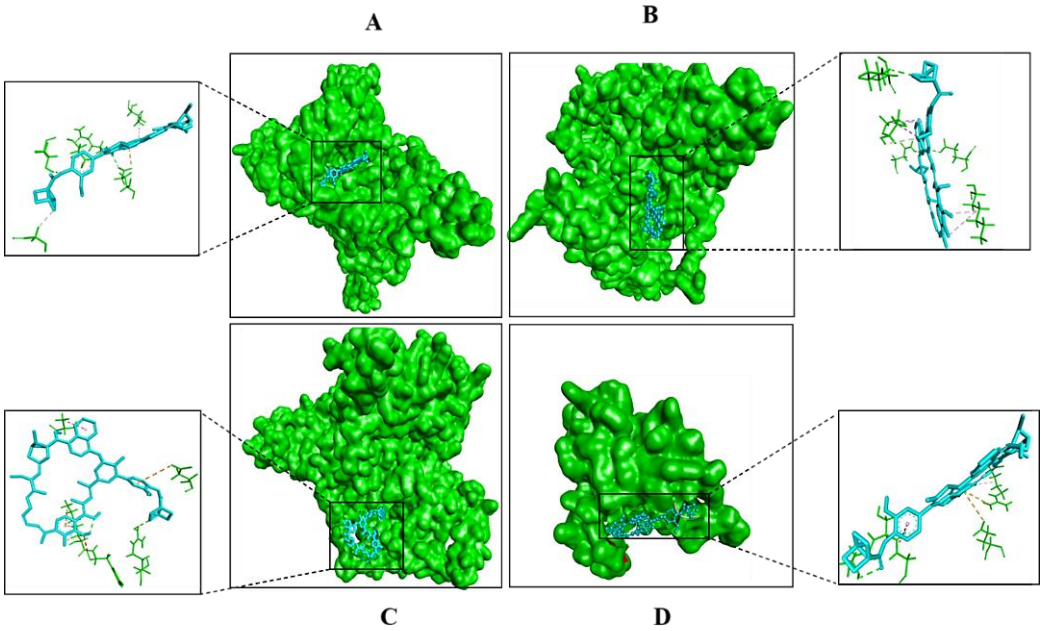

Figure 5: Ligand-binding pocket of PIK3CA protein isoforms: (A) Canonical Protein, (B) PIK3CA-201, (C) PIK3CA-205, and (D) PIK3CA-202 with the drug temsirolimus

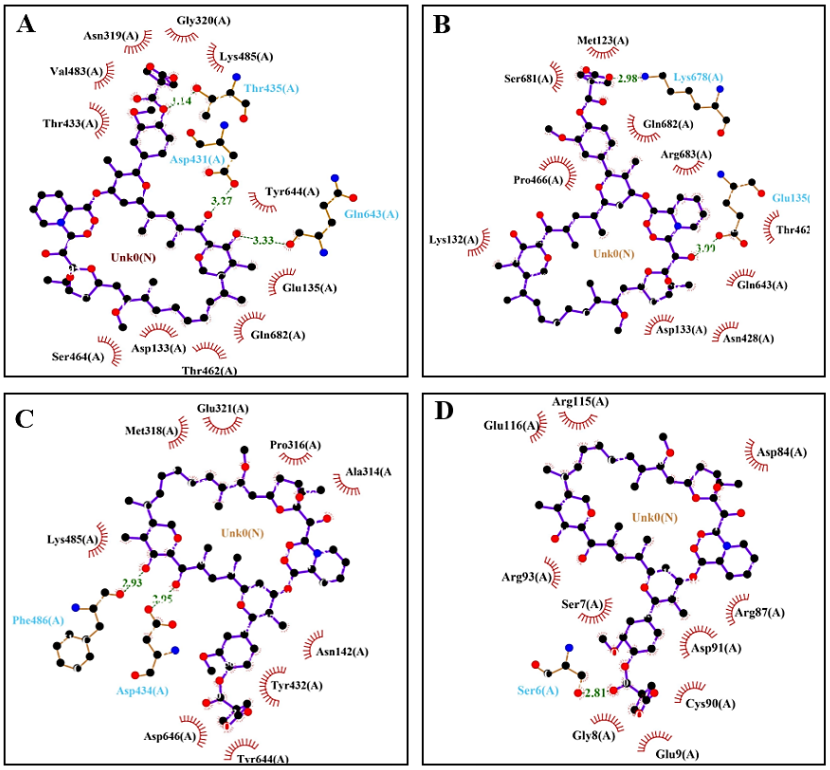

Figure 6: Ligplot analysis of the interactions between PIK3CA protein isoforms and temsirolimus. Hydrophobic interactions between amino acid residues are indicated by red arcs, whereas hydrogen bonds are represented by dashed green lines with specified bond lengths

Discussion

Although current target prediction methods have shown the accuracy of genomic, chemical, and pharmacological data in drug target interaction prediction, these methods frequently concentrate only on the canonical protein while disregarding the on- or even off-target isoform-level interactions that are linked to the action of the chemical [53]. Previous studies have related cancer-specific aberrant splicing to drug resistance. However, little is known about the therapeutic effects of this drug on specific tissues and its side effects on other tissues. Gene isoforms produced by alternative splicing can be expressed at different levels and exhibit various, perhaps conflicting, activities in various tissues and/or organs [54, 55]. We postulated that various protein isoforms formed by alternative splicing might develop into candidates for off-target or non-target drug interactions due to the presence or lack of target binding sequences in different alternative splicing of genes specifically involved in cervical cancer. Our findings show that most small-molecule therapeutic targets have a variety of protein isoforms. Therefore, it is feasible that most pharmacologically targeting gene isoforms have functional differences and show isoform-level changes in their interactions with the drug.

We found that KRAS-203 is highly expressed in tumor samples. Sequence alignment and data analysis of the gene expression patterns in the TCGA and GTEx datasets uncovered significant data, such as medicines that skip alternative gene isoforms that are also expressed in cancer but perhaps are not targeted, while the drugs that might possibly target alternative protein isoforms that are variously expressed across many normal tissues and are involved in the process of cancer development. Furthermore, the ability of the same medication to bind to several structurally related protein isoforms with various affinities was verified using a drug docking study and structural analysis of KRAS and PIK3CA proteins. These findings are basically two processes in which both could possibly lead to far-off impacts that could result in drug resistance.

In comparison to the canonical isoform, we observed low KRAS isoform expression in TCGA samples. We observed, via structural docking, that various medicines can interact with all protein isoforms in various ways. It remains unknown whether the secondary protein isoforms behave similarly to or differently from the downregulated primary isoform, carcinogenic, or overexpressed. In contrast, the different protein isoforms, with the exception of KRAS-204, which was not expressed in normal or tumor samples, showed variable and greater expression in healthy tissues than in tumor tissues. These protein isoforms can act as tumor suppressors or regulators, counteracting the functions of carcinogenic protein isoforms. The immediate inhibition of these protein isoforms may be undesirable under these conditions. Although the precise roles of these protein isoforms are still unknown, separating sites from non-targets at the splice level is a crucial step in the early stages of drug discovery.

Owing to restrictions on the availability of data, we were challenged by several limitations in the current study. The first is the lack of mapping of gene isoforms between public online databases and older studies. For example, differences in exon numbers are frequently reported between these two sources. Public databases such as Ensemble do not contain many gene isoforms previously described in the literature. This makes it extremely challenging to annotate these gene and protein isoforms both structurally and functionally. Therefore, the major aspects of our study are that the overexpression of protein isoforms that are more advantageous for the development of cancer should be suppressed, and the main aims for suppression should be the gene isoforms that are upregulated in cancer. This is obviously a restriction because these two hypotheses might be incorrect; however, we do not currently have any better methods for evaluating the roles of these unidentified protein isoforms. Furthermore, the inclusion of the actual gene-level expression of these gene isoforms will strengthen this claim. To our knowledge, there is currently no comprehensive database that includes the expression of all the protein isoforms on a complete proteome scale. In our opinion, the importance of identifying pharmacological targets at the protein isoform level should be emphasized. However, our results add to those of a recent study that identified mean mRNA expression across tissues and variance in expression across tissues as the two key characteristics that separate effective medications from ineffective ones [56].

Conclusions

This study highlights the potential risks of focusing solely on the canonical isoform, and ignoring the impact of cervical cancer drugs on- and off-target effects at the isoform level. Identifying additional cancer biomarkers at the isoform level and connecting them to treatment sensitivity using

computational methods is crucial. Our findings indicate that the protein isoforms have distinct binding pocket confirmations, which indicate the potential variations in drug binding and efficacy. Some isoforms completely lack the binding pocket, which highlights the importance of considering the drugs effectiveness across the isoforms. Some isoforms were found to be upregulated in tumor samples, suggesting them as potential therapeutic targets. Molecular docking analysis revealed that protein isoforms have varying binding affinity with FDA approved drugs, which is essential to predict the drug response and effectiveness. We expect that our findings will encourage further investigation into the possibility of designing protein isoform-level medication. Sufficient structural and functional knowledge of these isoforms is necessary to achieve this goal.

Interests disclisure

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Funding

None.

References

[1] Yang BH, Bray FI, Parkin DM, Sellors JW, Zhang ZF. Cervical cancer as a priority for prevention in different world regions: an evaluation using years of life lost. Int j Cancer. 2004;109(3):418-24. DOI: 10.1002/ijc.11719

[2] Stuver S, Adami H-O. Cervical cancer. New York: Oxford University Press; 2002.

[3] Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet. 2019;393(10167):169-82. DOI: 10.1016/S0140-6736(18)32470-X

[4] Crosbie EJ, Einstein MH, Franceschi S, Kitchener HC. Human papillomavirus and cervical cancer. Lancet. 2013;382(9895):889-99. DOI: 10.1016/s0140-6736(13)60022-7

[5] Small Jr W, Bacon MA, Bajaj A, Chuang LT, Fisher BJ, Harkenrider MM, et al. Cervical cancer: a global health crisis. Cancer. 2017;123(13):2404-12. DOI: 10.1002/cncr.30667

[6] Howlader N, Ries LA, Stinchcomb DG, Edwards BK. The impact of underreported Veterans Affairs data on national cancer statistics: analysis using population-based SEER registries. J Ntnl Cancer Inst. 2009;101(7):533-6. DOI: 10.1093/jnci/djn517

[7] Moore DH. Cervical cancer. Obstet Gynecol. 2006;107(5):1152-61. DOI: 10.1097/01.AOG.0000215986.48590.79

[8] Kim S, Choi H, Byun J. Overall 5-year survival rate and prognostic factors in patients with stage IB and IIA cervical cancer treated by radical hysterectomy and pelvic lymph node dissection. Int J Gynecolog Cancer. 2000;10(4):305-12. DOI: 10.1046/j.1525-1438.2000.010004305.x

[9] Network CGAR. Integrated genomic and molecular characterization of cervical cancer. Nature. 2017;543(7645):378. DOI: 10.1038/nature21386

[10] Oyervides-Muñoz MA, Pérez-Maya AA, Rodríguez-Gutiérrez HF, Gómez-Macias GS, Fajardo-Ramírez OR, Treviño V, et al. Understanding the HPV integration and its progression to cervical cancer. Infect Genet Evol. 2018;61:134-44. DOI: 10.1016/j.meegid.2018.03.003

[11] Wilting SM, Steenbergen RD. Molecular events leading to HPV-induced high grade neoplasia. Papillomavirus Res. 2016;2:85-8. DOI: 10.1016/j.pvr.2016.04.003

[12] Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, et al. A comprehensive pan-cancer molecular study of gynecologic and breast cancers. Cancer Cell. 2018;33(4):690-705. e9. DOI: 10.1016/j.ccell.2018.03.014

[13] Liu S, Zheng B, Sheng Y, Kong Q, Jiang Y, Yang Y, et al. Identification of cancer dysfunctional subpathways by integrating DNA methylation, copy number variation, and gene-expression data. Front Genet. 2019;10:441. DOI: 10.3389/fgene.2019.00441

[14] Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40(12):1413-5. DOI: 10.1038/ng.259

[15] Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456(7221):470-6. DOI: 10.1038/nature07509

[16] Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet. 2010;11(5):345-55. DOI: 10.1038/ng.259

[17] Zhang J, Manley JL. Misregulation of pre-mRNA alternative splicing in cancer. Cancer Discov. 2013;3(11):1228-37. DOI: 10.1158/2159-8290.CD-13-0253

[18] Lee SC-W, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. Nat Med. 2016;22(9):976-86. DOI: 10.1038/nm.4165

[19] Climente-González H, Porta-Pardo E, Godzik A, Eyras E. The functional impact of alternative splicing in cancer. Cell Rep. 2017;20(9):2215-26. DOI: 10.1016/j.celrep.2017.08.012

[20] Pradella D, Naro C, Sette C, Ghigna C. EMT and stemness: flexible processes tuned by alternative splicing in development and cancer progression. Mol Cancer. 2017;16:1-19. DOI: 10.1186/s12943-016-0579-2
[21] Safikhani Z, Smirnov P, Thu KL, Silvester J, El-Hachem N, Quevedo R, et al. Gene isoforms as expression-based biomarkers predictive of drug response in vitro. Nat Commun. 2017;8(1):1126. DOI: 10.1038/s41467-017-01153-8
[22] Ma J, Wang J, Ghoraie LS, Men X, Chen R, Dai P. Comprehensive expression-based isoform biomarkers predictive of drug responses based on isoform co-expression networks and clinical data. Genomics. 2020;112(1):647-58. DOI: 10.1016/j.ygeno.2019.04.017
[23] Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res. 2019;47(D1):D941-7. DOI: 10.1093/nar/gky1015
[24] Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, et al. Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods. 2021;18(10):1196-203. DOI: 10.1038/s41592-021-01252-x
[25] Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, et al. Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res. 2021;49(D1):D1144-51. DOI: 10.1093/nar/gkaa1084
[26] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074-82. DOI: 10.1093/nar/gkx1037
[27] Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1):D945-54. DOI: 10.1093/nar/gkw1074
[28] Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021;49(D1):D344-54. DOI: 10.1093/nar/gkaa977
[29] Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38(6):675-8. DOI: 10.1038/s41587-020-0546-8
[30] Du Z, Su H, Wang W, Ye L, Wei H, Peng Z, et al. The trRosetta server for fast and accurate protein structure prediction. Nat Protoc. 2021;16(12):5634-51. DOI: 10.1038/s41596-021-00628-9
[31] Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004;32(suppl_2):W526-31. DOI: 10.1093/nar/gkh468
[32] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-303. DOI: 10.1093/nar/gky427
[33] Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9:1-8. DOI: 10.1186/1471-2105-9-40
[34] Halgren TA. Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model. 2009 Feb;49(2):377-89. DOI: 10.1021/ci800324m
[35] Fan J, Bellon M, Ju M, Zhao L, Wei M, Fu L, et al. Clinical significance of FBXW7 loss of function in human cancers. Mol Cancer. 2022;21(1):87. DOI: 10.1186/s12943-022-01548-2
[36] Ross SJ, Revenko AS, Hanson LL, Ellston R, Staniszewska A, Whalley N, et al. Targeting KRAS-dependent tumors with AZD4785, a high-affinity therapeutic antisense oligonucleotide inhibitor of KRAS. Sci Transl Med. 2017 Jun 14;9(394):eaal5253. DOI: 10.1126/scitranslmed.aal5253
[37] De Roock W, Jonker DJ, Di Nicolantonio F, Sartore-Bianchi A, Tu D, Siena S, et al. Association of KRAS p.G13D mutation with outcome in patients with chemotherapy-refractory metastatic colorectal cancer treated with cetuximab. JAMA. 2010;304(16):1812-20. DOI: 10.1001/jama.2010.1535
[38] Maitra R, Seetharam R, Tesfa L, Augustine TA, Klampfer L, Coffey MC, et al. Oncolytic reovirus preferentially induces apoptosis in KRAS mutant colorectal cancer cells, and synergizes with irinotecan. Oncotarget. 2014;5(9):2807-19. DOI: 10.18632/oncotarget.1921
[39] Kloth JN, Kenter GG, Spijker HS, Uljee S, Corver WE, Jordanova ES, et al. Expression of Smad2 and Smad4 in cervical cancer: absent nuclear Smad4 expression correlates with poor survival. Modern Pathol. 2008;21(7):866-75. DOI: 10.1038/modpathol.2008.62
[40] Singhi AD, Ali SM, Lacy J, Hendifar A, Nguyen K, Koo J, et al. Identification of targetable ALK rearrangements in pancreatic ductal adenocarcinoma. J Natl Compr Canc Netw. 2017 May;15(5):555-62. DOI: 10.6004/jnccn.2017.0058
[41] Ganesh K, Shah RH, Vakiani E, Nash GM, Skottowe HP, Yaeger R, et al. Clinical and genetic determinants of ovarian metastases from colorectal cancer. Cancer. 2017;123(7):1134-43. DOI: 10.1002/cncr.30424
[42] Hassan B, Akcakanat A, Sangai T, Evans KW, Adkins F, Eterovic AK, et al. Catalytic mTOR inhibitors can overcome intrinsic and acquired resistance to allosteric mTOR inhibitors. Oncotarget. 2014;5(18):8544-57. DOI: 10.18632/oncotarget.2337
[43] Voutsadakis IA. PI3KCA mutations in uterine cervix carcinoma. J Clin Med. 2021 Jan 10;10(2):220. DOI: 10.3390/jcm10020220
[44] Juric D, de Bono JS, LoRusso PM, Nemunaitis J, Heath EI, Kwak EL, et al. A first-in-human, Phase I, dose-escalation study of TAK-117, a selective PI3Kα isoform inhibitor, in patients with advanced solid malignancies. Clin Cancer Res. 2017 Sep 1;23(17):5015-23. DOI: 10.1158/1078-0432.CCR-16-2888

[45] Kiavue N, Cabel L, Melaabi S, Bataillon G, Callens C, Lerebours F, et al. ERBB3 mutations in cancer: biological aspects, prevalence and therapeutics. Oncogene. 2020;39(3):487-502. DOI: 10.1038/s41388-019-1001-5

[46] Jaiswal BS, Kljavin NM, Stawiski EW, Chan E, Parikh C, Durinck S, et al. Oncogenic ERBB3 mutations in human cancers. Cancer Cell. 2013;23(5):603-17. DOI: 10.1016/j.ccr.2013.04.012

[47] Wada K, Lee JY, Hung HY, Shi Q, Lin L, Zhao Y, et al. Novel curcumin analogs to overcome EGFR-TKI lung adenocarcinoma drug resistance and reduce EGFR-TKI-induced GI adverse effects. Bioorg Med Chem. 2015 Apr 1;23(7):1507-14. DOI: 10.1016/j.bmc.2015.02.003

[48] He L, Torres-Lockhart K, Forster N, Ramakrishnan S, Greninger P, Garnett MJ, et al. Mcl-1 and FBW7 control a dominant survival pathway underlying HDAC and Bcl-2 inhibitor synergy in squamous cell carcinoma. Cancer Discov. 2013;3(3):324-37. DOI: 10.1158/2159-8290.CD-12-0417

[49] Villaruz LC, Socinski MA. Temsirolimus therapy in a patient with lung adenocarcinoma harboring an FBXW7 mutation. Lung Cancer. 2014 Feb;83(2):300-1. DOI: 10.1016/j.lungcan.2013.11.018

[50] Mendoza MC, Er EE, Blenis J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. Trends Biochem Sci. 2011;36(6):320-8. DOI: 10.1016/j.tibs.2011.03.006

[51] Miller S, Tavshanjian B, Oleksy A, Perisic O, Houseman BT, Shokat KM, et al. Shaping development of autophagy inhibitors with the structure of the lipid kinase Vps34. Science. 2010;327(5973):1638-42. DOI: 10.1126/science.1184429

[52] Goldman M, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. biorxiv. 2019 Mar 5:326470. DOI: 10.1101/326470

[53] Zhou L, Li Z, Yang J, Tian G, Liu F, Wen H, et al. Revealing drug-target interactions with computational models and algorithms. Molecules. 2019;24(9):1714. DOI: 10.3390/molecules24091714

[54] Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH-M. The functional consequences of alternative promoter use in mammalian genomes. Trends Genet. 2008 Apr;24(4):167-77. DOI: 10.1016/j.tig.2008.01.008

[55] Kalsotra A, Cooper TA. Functional consequences of developmentally regulated alternative splicing. Nat Rev Genetics. 2011;12(10):715-29. DOI: 10.1038/nrg3052

[56] Rouillard AD, Hurle MR, Agarwal P. Systematic interrogation of diverse Omic data reveals interpretable, robust, and generalizable transcriptomic features of clinically successful therapeutic targets. PLoS Comput Biol. 2018;14(5):e1006142. DOI: 10.1371/journal.pcbi.1006142

А. Ікбал[1], Ф. Алі[1], Ш. Чоудхарі[1], А. Каюм[1], Ф. Аршад[1], С. Ашраф[1], М. Азіз[1], А.У. Шакіл[1], Ш.А. Сегал[2], М. Хусейн[1], М. Саджид[1]

[1]Факультет наук про життя, кафедра біотехнології, Університет Окари, Окара, Пенджаб, Пакистан
[2]Кафедра біоінформатики, Ісламський університет Бахавалпур, Бахавалпур, Пенджаб, Пакистан

АНАЛІЗ *IN SILICO* НЕЦІЛЬОВОЇ ДІЇ ЛІКІВ ПРОТИ РАКУ ШИЙКИ МАТКИ НА РІЗНОМАНІТНІ ІЗОФОРМИ БІЛКА ДЛЯ РОЗШИРЕНИХ ТЕРАПЕВТИЧНИХ СТРАТЕГІЙ

Проблематика. Рак шийки матки є серйозним захворюванням, яке щорічно вражає сотні тисяч людей у всьому світі. Вибір і аналіз відповідних генів-мішеней на ранніх стадіях розробки ліків є вкрай важливим для боротьби з цією хворобою. Однак якщо залишити наявність різних ізоформ білка поза увагою, то можуть виникнути небажані терапевтичні або шкідливі побічні ефекти.
Мета. Провести обчислювальний аналіз взаємодії між ліками проти раку шийки матки та їхніми мішенями, на які впливає альтернативний сплайсинг.
Методика реалізації. Використовуючи бази даних відкритого доступу, ми націлилися на 45 схвалених FDA ліків від раку шийки матки, які спрямовані на різні гени, що мають більше двох різних ізоформ, які кодують білок. Щоб перевірити збереження кишені зв'язування в ізоформах генів, ми виконали множинний аналіз послідовностей. Щоб краще зрозуміти зв'язок між білками та схваленими FDA препаратами на рівні ізоформ, ми провели молекулярний докінг-аналіз.
Результати. Дослідження показує, що багатьом препаратам бракує потенційних мішеней на рівні ізоформ. Подальше дослідження різних ізоформ того самого гена виявило різні конфігурації кишень, що зв'язують ліганд, включаючи відмінності в розмірі, формі, електростатичних характеристиках і структурі.
Висновки. Це дослідження підкреслює потенційні ризики зосередження виключно на канонічній ізоформі й ігнорування цільового та нецільового впливу ліків проти раку шийки матки на рівні ізоформ. Ці результаті підкреслюють важливість розгляду взаємодії між ліками та їхніми мішенями на рівні ізоформ для сприяння ефективному лікуванню.

Ключові слова: рак шийки матки; ізоформи; молекулярний докінг; аналіз взаємодії; біоінформаційні підходи.