

RESEARCH

Open Access



Genetic polymorphism in untranslated regions of PRKCZ influences mRNA structure, stability and binding sites

Aneela Mustafa¹, Maria Shabbir^{1*}, Yasmin Badshah¹, Khushbukhat Khan², Fizzah Abid¹, Janeen H. Trembley^{3,4,5}, Tayyaba Afsar⁶, Ali Almajwal⁶ and Suhail Razak^{6*}

Abstract

Background Variations in untranslated regions (UTR) alter regulatory pathways impacting phenotype, disease onset, and course of disease. Protein kinase C Zeta (PRKCZ), a serine-threonine kinase, is implicated in cardiovascular, neurological and oncological disorders. Due to limited research on PRKCZ, this study aimed to investigate the impact of UTR genetic variants' on binding sites for transcription factors and miRNA. RNA secondary structure, eQTLs, and variation tolerance analysis were also part of the study.

Methods The data related to PRKCZ gene variants was downloaded from the Ensembl genome browser, COSMIC and gnomAD. The RegulomeDB database was used to assess the functional impact of 5' UTR and 3' UTR variants. The analysis of the transcription binding sites (TFBS) was done through the Alibaba tool, and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) was employed to identify pathways associated with PRKCZ. To predict the effect of variants on microRNA binding sites, PolymiRTS was utilized for 3' UTR variants, and the SNPinfo tool was used for 5' UTR variants.

Results The results obtained indicated that a total of 24 variants present in the 3' UTR and 25 variants present in the 5' UTR were most detrimental. TFBS analysis revealed that 5' UTR variants added YY1, repressor, and Oct1, whereas 3' UTR variants added AP-2alpha, AhR, Da, GR, and USF binding sites. The study predicted TFs that influenced PRKCZ expression. RNA secondary structure analysis showed that eight 5' UTR and six 3' UTR altered the RNA structure by either removal or addition of the stem-loop. The microRNA binding site analysis highlighted that seven 3' UTR and one 5' UTR variant altered the conserved site and also created new binding sites. eQTLs analysis showed that one variant was associated with PRKCZ expression in the lung and thyroid. The variation tolerance analysis revealed that PRKCZ was an intolerant gene.

Conclusion This study laid the groundwork for future studies aimed at targeting PRKCZ as a therapeutic target.

Keywords UTR, PRKCZ, Non-coding SNPs, *In-silico* tools, miRNA, Oncological disorders

*Correspondence:

Maria Shabbir
mshabbir@asab.nust.edu.pk
Suhail Razak
smarazi@ksu.edu.sa

¹Department of Healthcare BiotechnologyAtta-Ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Sector H-12, Islamabad 44000, Pakistan

²Cancer Clinical Research Unit, Trials360, Lahore, Pakistan

³Minneapolis VA Health Care System Research Service, Minneapolis, MN, USA

⁴Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA

⁵Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA

⁶Department of Community Health Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

The human genome predominantly comprises of non-coding DNA [1]. Non-coding DNA, although not translated into proteins, plays a significant role in gene expression regulation during the transcription phase. The non-coding regions harbor untranslated regions (UTRs) that include 5' UTRs and 3' UTRs. 5' UTRs are situated upstream of the protein-coding region, while 3' UTRs are present downstream of protein-coding regions. UTRs, as we know, are not just passive bystanders in the process of gene expression. They are vital components for gene expression regulation and pre-mRNA processing, which affect protein outcomes [2, 3]. The 5' UTRs control translation initiation by serving as an entry point for ribosome and interaction with TFs [4]. However, 3'UTRs are involved in the mRNA subcellular location, stability, and polyadenylation [5, 6].

Various studies in the literature have highlighted SNPs in UTR regions cause disease development, including cancer [7, 8]. UTR functions are commonly disrupted in oncological disorders [9, 10]. SNPs in the 3' UTR region are reported to modify the cellular homeostasis balance toward cancer [11]. 5' UTR mutations in BCL2 CXCL14 and TAOK2 genes caused abnormal gene translation [12]. Furthermore, the 5' UTR mutation in the APC gene added a false start codon, which hindered the translation initiation process, leading to colon cancer development [13]. 5' UTR (g.-130G>A) variant in the MATE2-K gene associated with poor metformin response [14]. PRKCZ is a serine-threonine kinase that belongs to the atypical group of protein kinase C [15]. PRKCZ is present on chromosome 1(1p36.33–p36.2) on the forward strand [16]. PRKCZ enzyme is polyfunctional, regulating cellular homeostasis and behavioral phenotype [17, 18]. PRKCZ is a regulator of tumorigenesis due to the downstream target of the P13K, and it also activates ERK/MAPK and NF-kb, JAK/STAT pathways [19–21]. PRKCZ expression dysregulations lead to poor prognosis and a lower survival rate, as reported in different cancers such as prostate cancer, breast cancer, glioblastoma, colon cancer, and ovarian cancer [21–25]. A splice variant in the 3' region of PRKCZ was highly expressed in malignant prostate cancer cells and tissues [26]. rs436045 of PRKCZ was significantly associated with type 2 diabetes (T2DM) pathogenesis [27]. A previous study found that PRKCZ promoter methylation was involved in T2DM development [28]. PRKCZ hypermethylation induced EMT in Head and neck squamous cell carcinoma [29].

The present study was mainly focused on untranslated region variants, i.e., 5' UTR and 3' UTR, of PRKCZ. The current research aimed to explore the impact of PRKCZ 5'UTR and 3'UTR variants on transcription factor binding sites and the impact of TFs on gene regulation and miRNA-mediated interactions. The study also focused on

studying the impact of 5'UTR and 3'UTR variants on the stability of mRNA secondary structure. eQTLs analysis was targeted to investigate the effect of UTR variants on the expression of PRKCZ in different tissues. Variation tolerance analysis was targeted to analyze the probability of variation in the gene with disease onset. This study provides preliminary in-silico based analyses that can be investigated further to understand the functional and clinical implications of PRKCZ UTR variants.

Methods

Data collection

The data relating to PRKCZ gene variants was downloaded from the Ensembl genome browser [30], COSMIC [31] and gnomAD [32]. The data contained variant IDs, allele variations and genomic coordinates. Ensembl genome browser includes a wide range of SNPs, both common and rare variants from various sources [30, 33], gnomAD focuses on genomic variants from whole-genome sequencing data [32], and COSMIC primarily contains somatic mutations from cancer genomes. It has lower UTR variants because it mainly focuses on coding region mutations [34]. The variant data obtained from the databases comprised two categories: coding regions and non-coding regions. As per the research objective, non-coding region variants encompassing 5' UTR and 3' UTR were included in the study and coding region variants were excluded. In order to process high-quality data, all the UTRs were screened out from the databases and thoroughly investigated to remove redundant data. The UTR SNPs were mapped on the PRKCZ genome sequence (GRCh38/hg38 genome assembly) according to the information retrieved from the databases.

Regulatory function analysis

The RegulomeDB database [35] was accessed to determine the functional impact of 5' UTRs and 3'UTRs variants. This tool classifies the variants into six classes ranging from 1 to 7, and Classes are further subdivided into subcategories, namely 1a, 1b, 2a, 2b, 3a, and 3b and it also provides scores to UTR variants: with variants having a score of 1 or nearly equal to 1 have functional dysregulations and score less than 1 indicate lower functional dysregulation (Supplementary Material 1, Table S1).

Identification of transcription factor binding sites

The Alibaba tool was utilized to analyze the impact of UTR variants on transcription binding sites (TFBS) [36]. The tool combines TRANSFAC database data and generates pairwise alignment to find the TFBS. Alibaba tool takes the FASTA format as an input and provides TFBS as an output. It facilitates the identification of the number and type of transcription factors that bind to both

the wild-type and mutant sequences. To elucidate the upstream and downstream proteins of PRKCZ that regulate diverse transcription factors in the nucleus, a thorough analysis was conducted through pathway mapping and gene annotations obtained from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database [37]. The pathway was visually represented using Microsoft PowerPoint.

microRNA annotation and structure prediction

To predict the effect of UTR variants on microRNA binding sites, PolymiRTS [38] was utilized for 3' UTR variants, and the SNPinfo tool [39] was used for 5' UTR variants. The PolymiRTS tool takes variant ID as an input and provides miRNA ID, functional class and context+score as an output. The database classified the effects of these variants into four functional categories: 'D', 'N', 'C', and 'O'. The 'D' class indicates that the derived allele affects a conserved miRNA site, the 'N' class disrupts a non-conserved miRNA site, the 'C' class creates a new miRNA site, and the 'O' class provides no information on the ancestral allele. Additionally, the PolymiRTS Database provided insight into the Context+score change, where higher negative scores suggest a greater likelihood of disease development due to miRNA binding site disruption. Similarly, the SNPinfo tool also takes a variant ID as input and provides a transcription binding site if present, its genomic location and allele alteration as an output.

RNA secondary structure prediction

The study investigated the effect of both 5' UTR, located upstream of the initiation site, and 3' UTR, located downstream of the termination site of mRNA transcript, on the pre-processing, stability, and translation of mRNA structures. The RNAFold [40] web server was utilized to predict the secondary structure of mRNA and the minimum free energy value as a measure of the stability of the mRNA structure. RNA secondary is an important component in transcription and translation [6, 41]. RNAFold utilizes dynamic programming and McCaskill PF algorithms to predict mRNA secondary structure. It takes DNA sequence as an input to assess RNA secondary structure.

Expression quantitative trait loci (eQTLs) prediction

The GTEx portal [42] was employed to discover the single tissue eQTLs. eQTLs are genetic variants linked to various gene expression variations [43]. The GTEx portal takes variant ID as input and provides information on gene expression variations in different tissues, exclusively human tissues. The tool is an invaluable source for understanding the impact of non-coding variation correlated with other diseases, such as cancer, diabetes, stroke, etc.

Variation tolerance analysis

The Human Gene Database [44] was utilized to determine the genetic variation tolerance in the PRKCZ gene. The data provides a residual variation tolerance score and gene damage index (GDI). The residual variation score shows the gene's susceptibility to genetic variation, while GDI characterizes the relationship between gene damage and the likelihood of disease.

Results

Data collection and processing

Data for both the 3' and 5' UTR variants was obtained from the Ensembl genome browser. The total number of 3' and 5' UTR variants obtained from the Ensembl genome browser were 576, including 139 5' UTR variants (Supplementary Material S2, Table S2) and 223 3' UTR variants (Supplementary Material S3, Table S3). Furthermore, 192 variants were downloaded from genomAD, consisting of 124 5' UTR variants and 68 3' UTR variants. The COSMIC database only contained 22 variants, all of which were 3' UTR variants (Fig. 1A). The present study exclusively incorporated UTR variants that exhibited nucleotide changes, while those variants that lacked information related to nucleotide changes were excluded from the analysis (Fig. 1B).

Regulatory functional analysis

The present study conducted a detailed regulatory analysis of both the 3' and 5' UTR variants through utilizing the RegulomeDB tool. The RegulomeDB tool provided information on 100 5' UTR variants (Fig. 2A) (Supplementary material 4, Table S4) and 154 3' UTR variants (Supplementary material S5, Table S5). Furthermore, variants with a probability score ranging from 0.60 to 1 were included for functional analysis, whereas those with a score lower than 0.60 were excluded. After applying the pre-determined inclusion and exclusion criteria, a total of 25 5' UTR variants and 24 3' UTR variants were identified to have a score above the 0.60 threshold. This information is depicted in Fig. 3A and B. Figure 3C provides a visual representation of the regulomeDB class rank to which the identified variants belong. Additionally, Fig. 3D displays the percentage distribution of the variants across the different rank classes.

Transcription factor binding site analysis

A comparative analysis of the 5' UTR wild type and mutant sequences was conducted, which revealed that of the 25 variants analyzed, 11 variants caused the deletion of transcription factor binding sites, 8 variants led to the creation of a new site, whereas 6 variants caused no effect on TFBS (Supplementary Material S6, Table S6) (Fig. 4A). However, 5' UTR variants (rs1389053287, rs1282623960, rs192386882, and, rs57743955) led to the

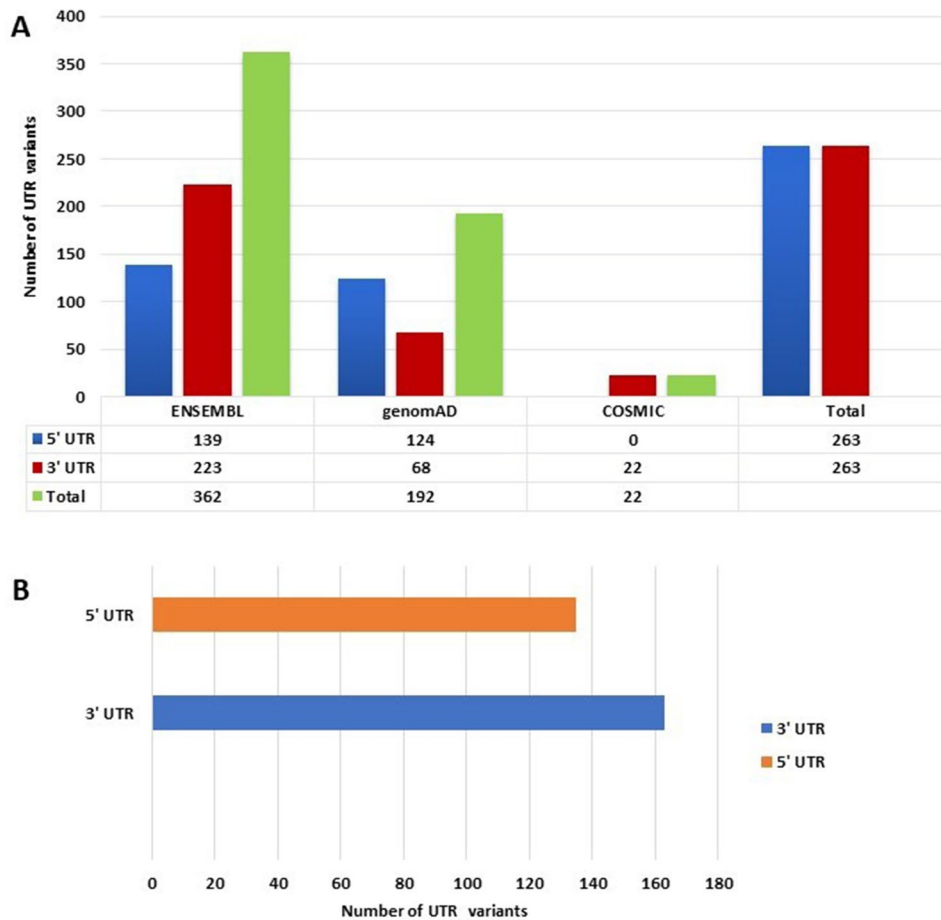


Fig. 1 PRKCZ corresponding UTR variants data retrieved from multiple sources, i.e., Ensembl genome browser, genomeAD, and Cosmic Panel (A) shows the number of 3' and 5' UTR variants obtained from databases, while (B) exhibits the UTR variants chosen for further analysis

addition of entirely new YY1, repressor, Oct1, and YY1 mutant TFBS respectively (Fig. 4B). Based on the comparative analysis of 3' UTR, it was found that 4 variants resulted in the deletion of TFBS, while 14 variants led to the addition of TFBS. Additionally, 6 variants did not affect the TFBS (Supplementary Material S77, Table S7) (Fig. 4C). 3' UTR variants (rs1017752535, rs755735985, rs375170466, rs1006027187, and rs371058786) led to the addition of AP-2alpha, AhR, Da, GR, and USF mutant TFBS respectively (Figs. 4D and 5). TFs identified through the Alibaba tool were also analyzed to investigate their impact on PRKCZ expression. Moreover, KEGG pathway analysis revealed that different signaling pathways were involved in PRKCZ gene expression. UTR genetic variants could lead to overexpression of certain TFs *c-myc*, *c-Jun* and *fos*, PDK and NF-KB complex, which ultimately could initiate positive feedback that promoted PRKCZ overexpression. These pathways encompass MAPK, WNT canonical and planner polarity pathway, NF-KB, and PI3K pathway. MAPK pathway was found to activate *c-myc*, *c-Jun* and *fos*, while NF-KB pathway promoted the transcription of NF-KB complex

transcription. Similarly, the WNT canonical and planner polarity pathway also activated *c-Jun*, *c-myc* and PDK. All these TFs mentioned triggered PRKCZ translation and expression. Moreover, PRKCZ underwent activation within the cytoplasm through the AKT and PDK pathway via PI3K activation. Additionally, PRKCZ exerted an influence on the activity of ERK and RAC1 proteins. Consequently, PRKCZ influenced ERK and RAC1 to enhance the activity of *C-myc*, *C-Jun*, and *Fos*.

Analyzing the impact of UTR variants on miRNA binding sites

The PolymiRTS database contained valuable data regarding variants that affected miRNA binding. Seven 3' UTR were observed to alter the miRNA binding sites, including its genomic location, ancestral allele, functional class, and Context+score (Table 1). Moreover, it was found that one SNP affected multiple miRNAs. SNPS rs80119748 altered the binding site of *hsa-miR-3186-3p* at conserved site, rs11553733 altered the binding site of various miRNAs; it created a new binding site of *has-miR-4746-3p* and *hsa-miR-876-3p*, and it altered the binding site of

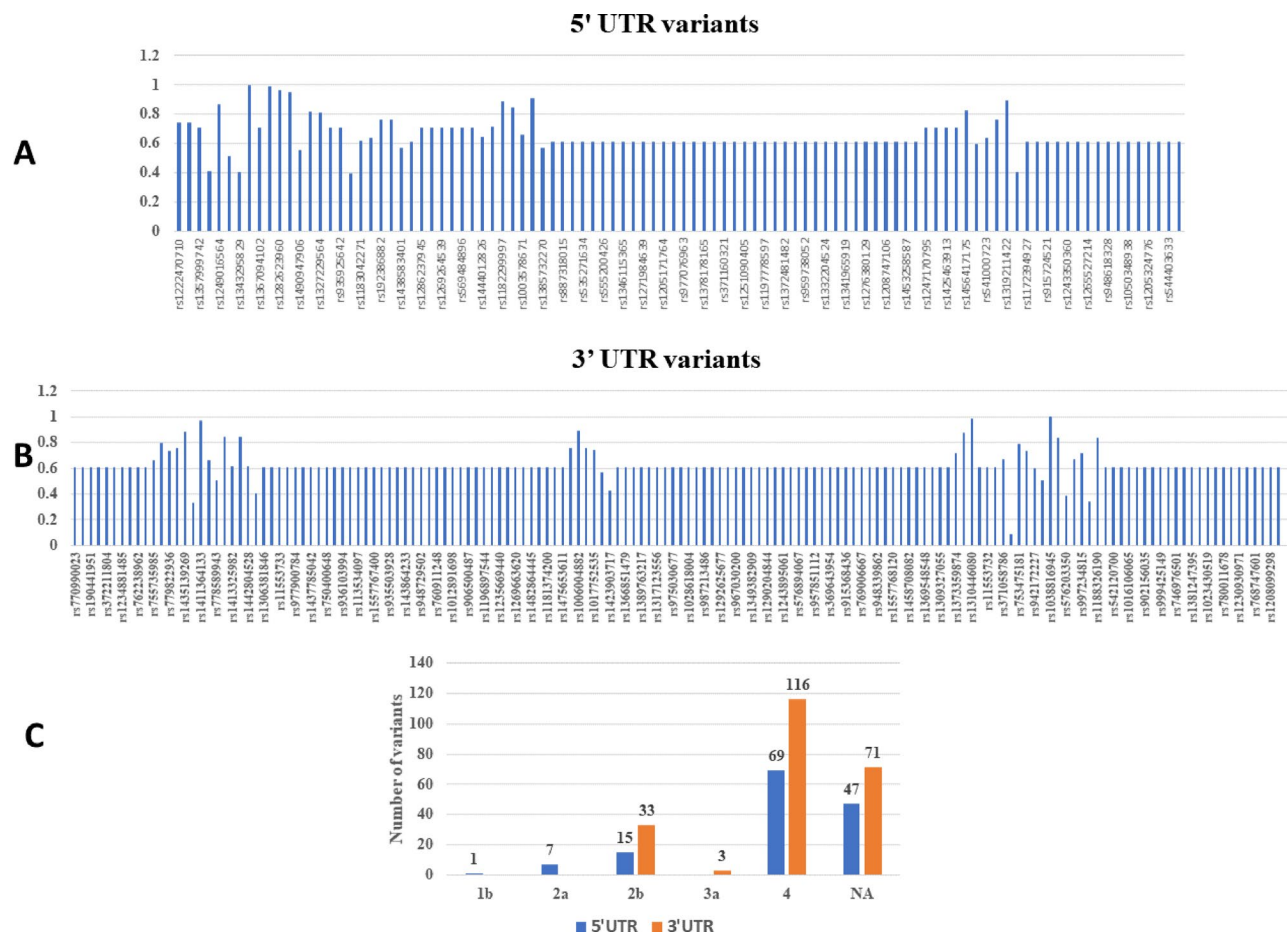


Fig. 2 Functional analysis of UTR variants by RegulomeDB. The data is segmented by score and rank for both 3' UTR and 5' UTR. **(A)** displays the score-wise category for 3' UTR variants, while **(B)** shows the score-wise classification for 5' UTR variants. A score of 1 or near 1 indicates a higher probability of affecting miRNA function. **(C)** demonstrates the rank-wise categorization for different classes for both 3' UTRs and 5' UTRs

hsa-miR-140-5p hsa-miR-6499-3p at the conserved site, rs113534097 impacted the binding site of two miRNA such as hsa-miR-188-5p, hsa-miR-6866-3p by creating new binding sites. rs143864233 affected the binding sites of hsa-miR-1307-3p, hsa-miR-1469, hsa-miR-4638-5p at conserved sites. rs202100130 altered the binding site of hsa-miR-3943 at the conserved site, whereas it created new binding sites for following miRNA such as hsa-miR-4313, hsa-miR-3127-3p, hsa-miR-6769a-3p, hsa-miR-6756, 3p, rs1050601, hsa-miR-132-5p, hsa-miR-3148 d, hsa-miR-6124 d, hsa-miR-4497, hsa-miR-4664-3p, hsa-miR-4668-5p, hsa-miR-6784-5p. The SNPinfo database indicated that rs57743955 5' UTR variant created TFBS. Among 5'UTR variants, it was found that solely one variant had a discernible impact on the TFBS (Table 2).

Analyzing the impact of UTR variants on RNA secondary structure

Out of 25 5' UTR, eight variants were found to alter the RNA secondary structure. rs1450239046 removed one stem-loop. Conversely, rs12822623960 removed

two stem loops, rs1471967977 removed two stem loops and increased the positional entropy, rs1319211422 removed one stem-loop and created a larger stem-loop, and rs1292130789 removed three stem-loops. Variant rs192386882 also caused the removal of three stem loops and caused an increase in the positional entropy, rs145641717175 rearranged nucleotides in the stem-loop, enlarged it, and increased the positional, and rs1444012826 added one stem loop and decreased the positional entropy (Fig. 6A). Out of 24 3' UTR, 6 variants either removed or added stem loops. The rs1038816956 removed a stem-loop leading to a decrease in positional entropy; the rs754595816 added one stem-loop, causing a larger stem-loop; the rs753475181 added two stem-loops, leading to high positional energy.the rs1006004882 and rs779822936 added one stem-loop, resulting in a decrease in positional entropy and the rs755735985 added a stem-loop, causingan increase in positional entropy (Fig. 6B).

The non-coding variants exerted varied effects on the mRNA structural stability by either enhancing or reducing it or causingno effect. The results showed that among

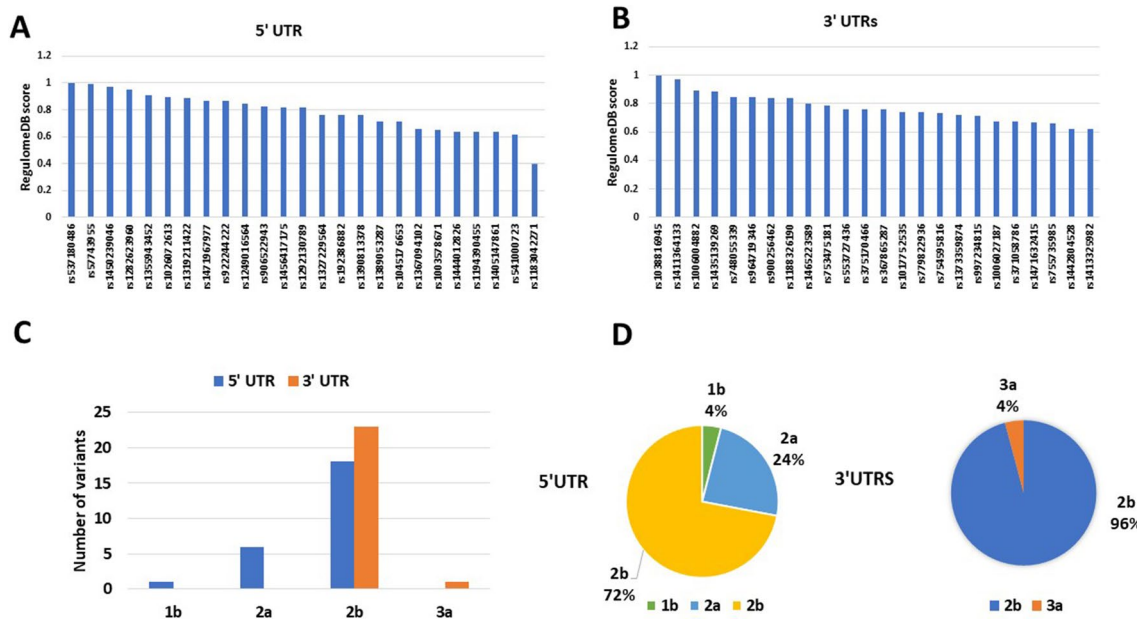


Fig. 3 UTR variants included for further analysis based on score and rank. **A**) selected 5' UTRs along with their score probability, **B**) selected 3' UTRs along with their score probability, **C**) the identified variant's class ranking, and **D**) the percentage distribution of both 5' and 3' UTR variants across the rank classes

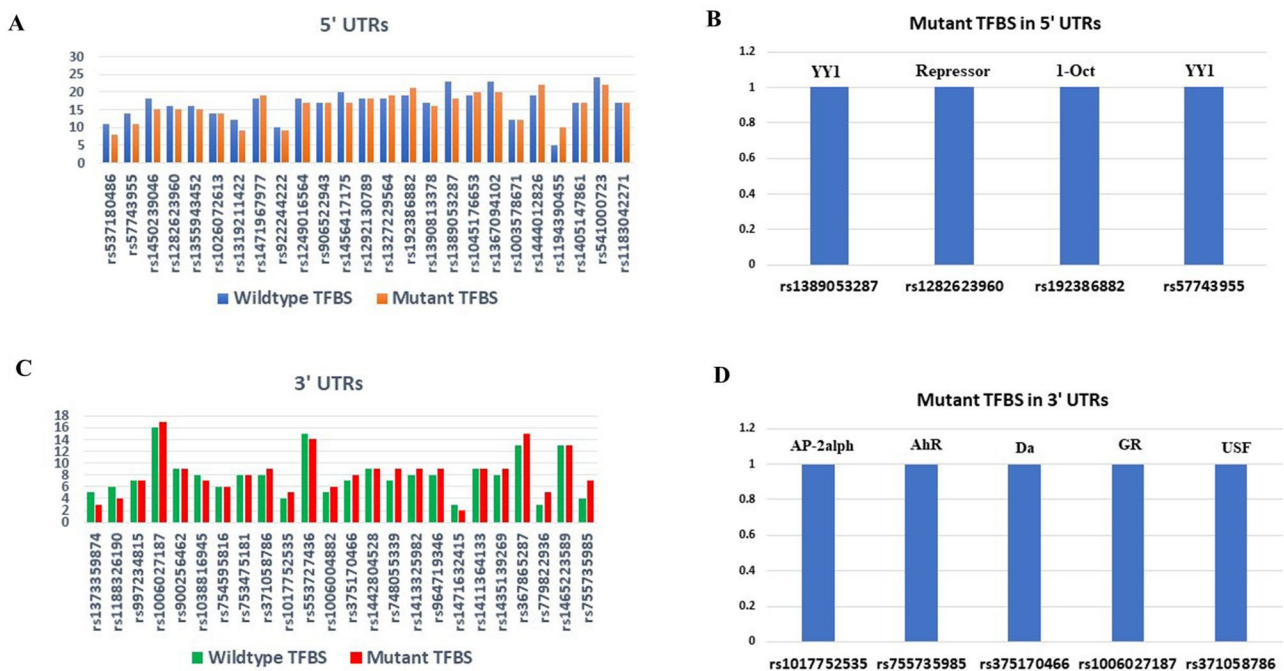


Fig. 4 Comparative analysis of TFBS by Alibaba web server for wild type and variant. **A**) illustrates the comparative analysis of 5' UTR for wildtype and variant, **B**) showcases the mutant TFBS induced by 5' UTR variants (rs1389053287, rs1282623960, rs192386882, and, rs57743955) **C**) displays the comparative study of the 3' UTRs for wildtype and variant and **D**) illustrates the mutant TFBS induced by 3' UTR variants (rs1017752535, rs755735985, rs375170466, rs100602187, and rs371058786)

5' UTR, ten variants caused an increase, six caused a decrease, and nine variants had no effect on the mRNA stability (Fig. 7A) (Supplementary Material S8, Table S8). Moreover, 3' UTR analysis showed that among all 3' UTR variants, seven variants led to an increase, six variants

caused a decrease, and ten variants had no effect on the mRNA structure stability (Fig. 7B) (Supplementary Material S9, Table S9). Furthermore, a collective analysis of 5' UTR and 3' UTR variants effect on mRNA stability is depicted in Fig. 7C.

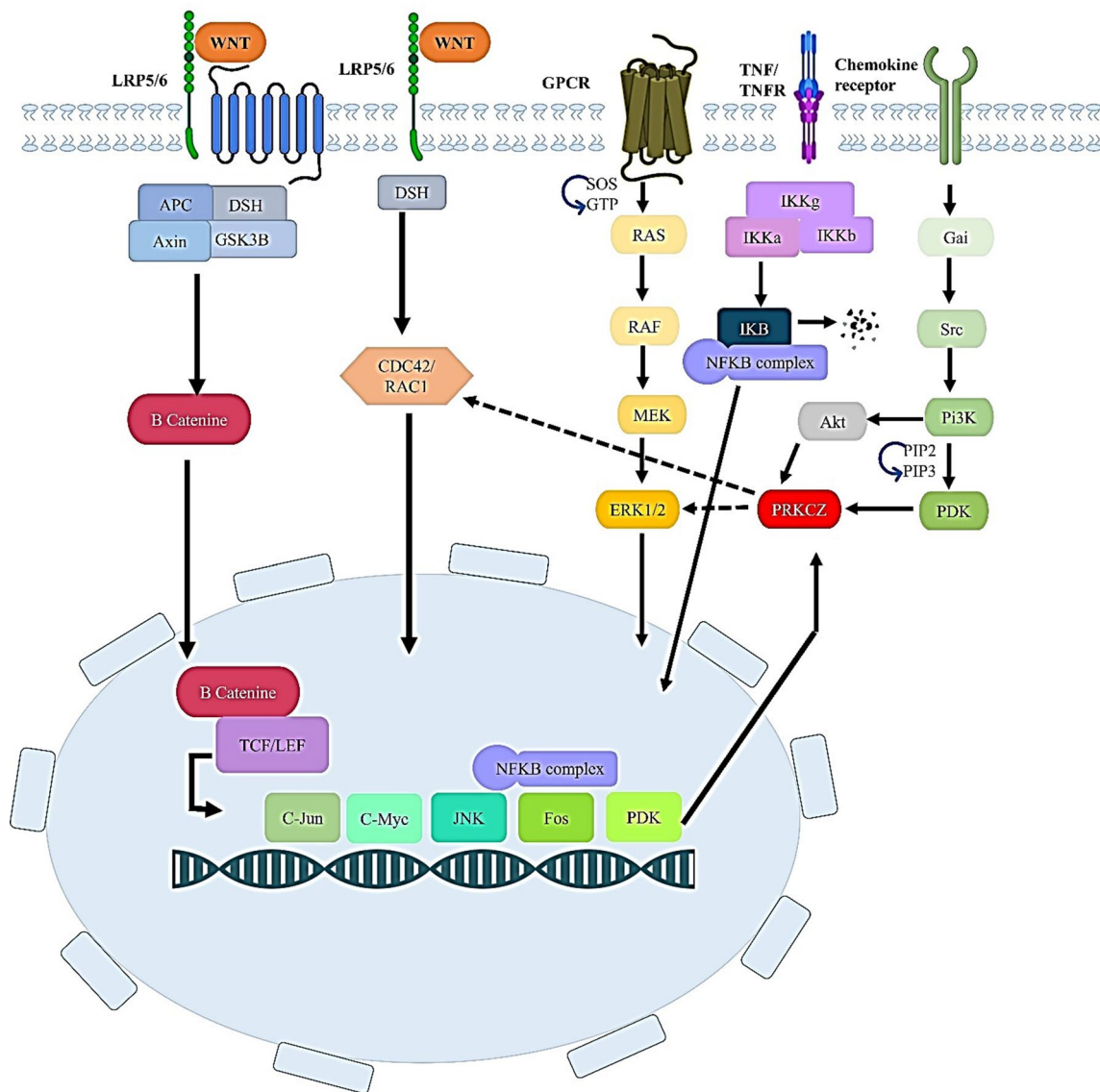


Fig. 5 PRKCZ signaling schematic representation constructed using the KEGG database. P13K, MAPK, NFkB WNT canonical and planner polarity pathway regulate PRKCZ transcription. Furthermore, PRKCZ interacts with ERK and CDC42/RAC1 to maintain positive feedback

Analyzing the impact of UTR variants on expression by eQTLs

The GTEx portal revealed that only one variant, rs17777295, was found to be expressed in the lungs and thyroid with a p-value of 1.9×10^{-116} and 1.5×10^{-19} , respectively (Table 3).

Analyzing variation tolerance

The Human Gene Database revealed that PRKCZ had a high probability of intolerance, with scores of 8.25% and 1.63%, respectively (Table 4). In practical terms, this means that genetic variants could impact PRKCZ functionally that could increase the likelihood of disease onset. Specifically, the database indicated that as gene

damage increases, the probability of the gene to cause disease decreases.

Discussion

Gene variations in UTR regions have been linked to various diseases. Literature has highlighted the involvement of UTR variants in several neurological and oncological disorders [45–48]. These regions are of extreme significance due to their essential function in regulating gene expression through the addition, deletion, or modification of transcription factors and transcription binding sites. However, the literature does not contain data concerning the role of PRKCZ UTR variants in disease onset. Experimental procedures can be inefficient in terms of both time and cost. Therefore, the present study utilized

Table 1 List of PRKCZ 3'UTR variants impacting 3'UTR region interaction with miRNA binding sites along with functional class and context+ score analyzed through PolymiRTS tool

Variant rsID	Location	Ancestral allele	Derived allele	miRNA ID	Functional class	context+ score change
rs80119748	1:2116455	G	G	hsa-miR-3186-3p	D	0.002
rs11553733	1:2116510	C	C	has-miR-4746-3p	D	-0.268
				hsa-miR-876-3p	D	-0.218
				hsa-miR-140-5p	C	-0.182
rs113534097	1:2116556	C	T	hsa-miR-6499-3p	C	-0.144
				hsa-miR-188-5p	C	-0.074
				hsa-miR-6866-3p	C	-0.074
rs143864233	1:2116570	G	G	hsa-miR-1307-3p	D	-0.461
				hsa-miR-1469	D	-0.366
				hsa-miR-4638-5p	D	-0.162
rs202100130	1:2116709	G	G	hsa-miR-6722-5p	C	-0.208
				hsa-miR-3943	D	-0.241
				hsa-miR-4313	D	-0.193
rs1050601	1:2116746	A	A	hsa-miR-3127-3p	C	-0.144
				hsa-miR-6769a-3p	C	-0.116
				hsa-miR-6756-3p	C	-0.201
rs1050601	1:2116746	A	C	hsa-miR-132-5p	D	-0.247
				hsa-miR-3148	D	0.027
				hsa-miR-6124	D	0.018
				hsa-miR-1180-3p	C	-0.227
				hsa-miR-4497	C	-0.28
				hsa-miR-4664-3p	C	-0.223
				hsa-miR-4668-5p	C	-0.063
hsa-miR-6784-5p	C	-0.293				

Table 2 PRKCZ 5'UTR variants modulating the miRNA TFBS analyzed through SNPinfo

Variant rsID	Location	Allele	TFBS
rs57743955	1: 1,971,873	C/T	Y

in-silico approaches to determine the effect of UTR variations in PRKCZ. The present study has examined

the impact of PRKCZ UTR variants on interactions with transcription factors and miRNA. RNA secondary structure, eQTLs, gene network, and variation tolerance analysis were also done.

The study presented here filtered out PRKCZ UTR variants with a high probability of functional dysregulations. The results are consistent with the past studies that

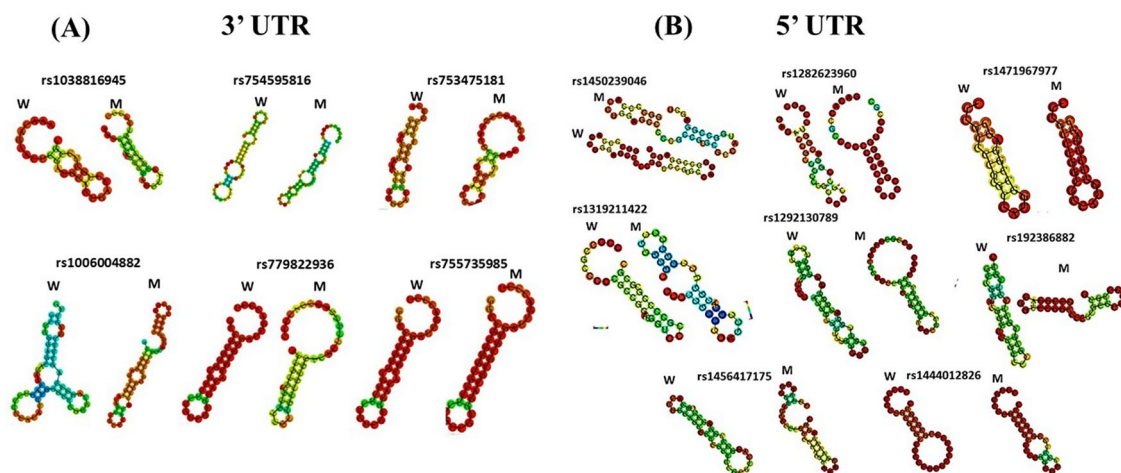


Fig. 6 PRKCZ RNA secondary prediction and positional entropy by RNAFold. (A) Shows RNA secondary structure for both wild type (W) and mutant (M) for 3'UTR variants, while (B) Shows RNA secondary structure for both wild type (W) and mutant (M) for 5'UTR variants. Positional entropy is denoted with different colors; red depicts the highest, and blue represents the lowest positional entropy

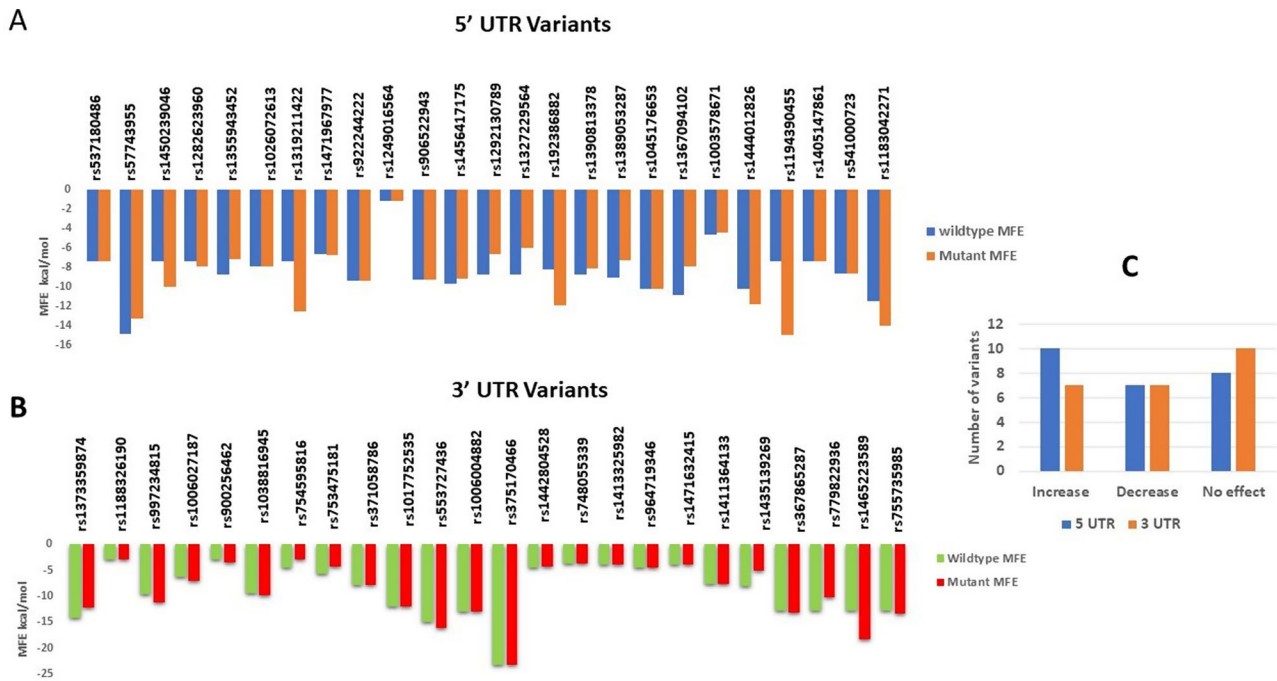


Fig. 7 Minimum Free Energy calculation by RNAFold due to UTR variants (A) variation in MFE in 5' UTR mutations, (B) variation in MFE in 3' UTR, and (C) shows a collective analysis of the decrease, increase, and no change in mRNA stability upon 5' UTR and 3' UTR variations

Table 3 PRK CZ UTR variant single tissue eQTL analysis by GTEx portal with the p-value

Variant ID	Organ	p-value
rs17777295	Lungs	1.9e-116
	Thyroid	1.5e-19

identified UTR variants in PRKCE, PKCI, and XRCC1 with functional implications [49–51]. Another study identified 97 non-coding SNPs involved in coronary artery disease development [52]. TFBS analysis revealed that 5' UTR variants added YY1, repressor, and Oct1, whereas 3' UTR variants added AP-2alpha, AhR, Da, GR, and USF binding sites. YY1 is a zinc finger transcription factor that regulates diverse functions such as proliferation, differentiation, and embryogenesis [53]. A previous study found that YY1 is closer to the translation start and had activator and repressor roles in gene regulation [54]. Repressor on the other hand, is a transcription factor that represses transcription by blocking the binding of different transcription factors to the promoter. The Oct1 transcription factor is pro-oncogenic and is involved in immune modulation, cytotoxic, and oxidative resistance, as well as metabolic reprogramming [55]. The role of AP-2alpha is significant in maintaining equilibrium between growth and differentiation. Its association with epidermal

growth factor is direct, and studies showed that it causes hyperproliferation in epidermal cancers [56]. AhR transcription factors play a role both in immunity toxicity and T cell development [57]. Da has an attachment site downstream of the transcription site where polymerase binds [58]. Da binding to the site where polymerase normally attaches affects gene expression regulation [59]. GR TF binding to glucocorticoid response elements (GREs) controls many physiological processes, such as developmental, metabolic, and immune [60, 61]. The USF transcription factor family is involved in various cellular processes, for example, iron homeostasis, embryonic development, and carcinogenesis [62]. Our findings support the conclusions of earlier studies that explored the functional implication of varying TFBS due to non-coding variants [49, 50, 63].

The current study identified 5' and 3' UTR variants affecting miRNA binding. 5' and 3' UTR variants disturb miRNA binding sites, affecting the translation process and leading to various pathologies [64]. A previous study found that 3' UTR SNP rs2229295 introduced binding sites for miR214-5p and miR550-5p, causing abnormal glucose metabolism [65]. MicroRNAs are essential components that play a crucial role in regulating the stability of mRNA [6]. The present study found that 3' UTR

Table 4 Genetic variation tolerance of PRK CZ along with residual variation intolerance score, Gene Damage Index Score and its impacts analyzed by the human Gene Database

Name	Residual Variation Intolerance Score	Impact	Gene Damage Index Score	Impact
PRK CZ	8.25%	Likely to be disease-causing	1.63%	Likely to be disease-causing

caused variation in miRNA binding, which could affect the translation process. 3' UTR SNP impact miRNA. In the past study, FOXC2 3' UTR SNPs (rs201118690, rs6413505, and rs201914560) caused binding site variation in miRNA (impact on three miRNA, hsa-miR-6886-5p, hsa-miR-6886-5p, and hsa-miR-6720-3p) [66]. 3' UTR (rs12516, rs3092995 and rs8176318) in BRCA1 impacted interaction with miR-103 seed sequence [65]. Another study showed that UTR variants impacted the binding of miRNA with PRKCE mRNA, causing PRKCE overexpression [49]. These findings align with prior studies that analyzed the impact of the 3' UTR variant on miRNA binding site and gene expression [45–48]. A previous study identified 110 3' UTR SNPs in 67 genes that impacted miRNA binding sites [67]. Both 3' and 5' UTR variants impacted mRNA secondary structure and stability. It had been emphasized in the literature that mRNA stability is a crucial factor for mRNA regulation and processing [68]. An in-silico study reported that the 3' UTR variant mutation was responsible for mRNA destabilization in the LIF gene, correlated to schizophrenia development [69]. The present study found that one variant was associated with tissue eQTLs. GTEx analysis had largely been employed in different studies to find the link between tissue eQTLs with coronary arteries, diabetes, and schizophrenia [70, 71]. According to the variation tolerance analysis conducted by the human gene database, it had been determined that the PRK CZ gene was intolerant, which put it at a higher risk of genetic variation. The GDI and residual intolerance score metrics are invaluable sources for determining the risk of genetic variation of a gene in the disease pathogenesis [72].

Conclusion

Our scientific investigation has employed in-silico approaches to determine the functional impact of UTR variants in PRK CZ. The present study has examined the involvement of TFs in regulating the expression of the PRK CZ gene. The PRK CZ UTR variants affected miRNA binding sites and mRNA stability. eQTLs analysis indicated that only one UTR variant showed expression in specific human tissues. Gene tolerance analysis revealed that PRK CZ is an intolerant gene that is prone to a high degree of genetic variation.

Limitations and Future Perspectives

It is imperative to validate the findings of the current study at the population level through genotyping following high-throughput sequencing techniques. Additionally, these discoveries can also be investigated in in-vitro and in-vivo models to determine the impact of variations in complex biological systems. Future studies may explore the potential correlation between identified TFs and PRK CZ by analyzing their co-expression.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-024-12900-8>.

Supplementary Material 1

Acknowledgements

The authors extend their appreciation to the Researchers Supporting project number (RSP2024R502), King Saud University, Riyadh, Saudi Arabia, for funding this project

Author contributions

Conceptualization, AM, MS, and YS; methodology, MS; experimentation validation AM, MS, TA, SR, YB, KK FA, formal analysis, AM, FA; investigation, KK, AA, FA; resources, MS, JHT, TA; data curation, ; writing—original draft preparation AM; writing—review and editing, FA; visualization, MS, supervision, MS; project administration, MS; All authors have read and agreed to the published version of the manuscript.

Funding

The authors extend their appreciation to the Researchers Supporting project number (RSP2024R502), King Saud University, Riyadh, Saudi Arabia, for funding this project. The funding body has no role in study design.

Data availability

Data are provided as supplementary material along with the manuscript. Raw data will be available from the corresponding author on request.

Declarations

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 20 May 2024 / Accepted: 4 September 2024

Published online: 13 September 2024

References

- Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.
- Pagni S, et al. Non-coding regulatory elements: potential roles in disease and the case of epilepsy. *Neuropathol Appl Neurobiol*. 2022;48(3):e12775.
- Wei W et al. Comprehensive characterization of posttranscriptional impairment-related 3'-UTR mutations in 2413 whole genomes of cancer patients. 2022. 7(1): p. 34.
- Leppik K, Das R, Barna M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol*. 2018;19(3):158–74.
- Chan JJ, Tabatabaieian H, Tay Y. 3' UTR heterogeneity and cancer progression. *Trends Cell Biol*. 2023;33(7):568–82.
- Wilkie GS, Dickson KS, Gray NK. Regulation of mRNA translation by 5'-and 3'-UTR-binding factors. *Trends Biochem Sci*. 2003;28(4):182–8.
- Mohammadi-Dehcheshmeh M et al. A transcription regulatory sequence in the 5' untranslated region of SARS-CoV-2 is vital for virus replication with an altered evolutionary pattern against human inhibitory microRNAs. 2021. 10(2): p. 319.
- Schuster SL et al. Multi-level functional genomics reveals molecular and cellular oncogenicity of patient-based 3' untranslated region mutations. 2023. 42(8).
- Mularoni L, et al. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol*. 2016;17:1–13.

10. Puente XS, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526(7574):519–24.
11. Zhang L et al. Functional SNP in the microRNA-367 binding site in the 3' UTR of the calcium channel ryanodine receptor gene 3 (RYR3) affects breast cancer risk and calcification. 2011. 108(33): pp. 13653–13658.
12. Zeraati M, et al. Cancer-associated noncoding mutations affect RNA G-quadruplex-mediated regulation of gene expression. *Sci Rep*. 2017;7(1):708.
13. Young B et al. P364: exploring the effects of a point mutation in the 5'UTR of APC found in a family with colon cancer. *Genet Med Open*, 2023. 1(1).
14. Choi J, et al. A common 5'-UTR variant in MATE2-K is associated with poor response to metformin. *Clin Pharmacol Ther*. 2011;90(5):674–84.
15. Newton AC. Protein kinase C: perfectly balanced. *Crit Rev Biochem Mol Biol*. 2018;53(2):208–30.
16. Wang H-F, et al. Hypermethylation of PRKCZ regulated by E6 inhibits invasion and EMT via Cdc42 in HPV-related head and neck squamous cell carcinoma. *Cancers*. 2022;14(17):4151.
17. Ohno S, Nishizuka Y. Protein kinase C isotypes and their specific functions: prologue. *J Biochem*. 2002;132(4):509–11.
18. Dempsey EC, et al. Protein kinase C isozymes and the regulation of diverse cell responses. *Am J Physiology-Lung Cell Mol Physiol*. 2000;279(3):L429–38.
19. Martin PM, Hussaini IM. PKC η as a therapeutic target in glioblastoma multiforme. *Expert Opin Ther Targets*. 2005;9(2):299–313.
20. Yang J-Q, et al. Loss of PKC α 1 impairs Th2 establishment and allergic airway inflammation in vivo. *Proc Natl Acad Sci*. 2009;106(4):1099–104.
21. Reina-Campos M, Diaz-Meco MT, Moscat J. The dual roles of the atypical protein kinase cs in cancer. *Cancer Cell*. 2019;36(3):218–35.
22. Carmo Ad, et al. PKC signaling in glioblastoma. *Cancer Biology Therapy*. 2013;14(4):287–94.
23. Yin J, et al. Association of PKC ζ expression with clinicopathological characteristics of breast cancer. *PLoS ONE*. 2014;9(3):e90811.
24. Islam SA, Patel R, Acevedo-Duncan M, BeBA-MCR. Protein kinase C- ζ stimulates colorectal cancer cell carcinogenesis via PKC- ζ /Rac1/Pak1/ β -Catenin signaling cascade. *Biochim et Biophys Acta -Molecular Cell Res*. 2018;1865(4):650–64.
25. Shimada E, et al. Methylation-mediated silencing of protein kinase C zeta induces apoptosis avoidance through ATM/CHK2 inactivation in dedifferentiated chondrosarcoma. *Br J Cancer*. 2022;126(9):1289–300.
26. Yao S, et al. Splice variant PRKC- ζ -PrC is a novel biomarker of human prostate cancer. *Br J Cancer*. 2012;107(2):388–99.
27. Li Y-F et al. *Protein kinase C ζ (PRKCZ) Gene is associated with type 2 diabetes in Han population of North China and analysis of its haplotypes*. *World Journal of Gastroenterology*, 2003. 9(9): p. 2078.
28. Zou L, et al. Hypermethylation of the PRKCZ gene in type 2 diabetes mellitus. *J Diabetes Res*. 2013;2013(1):721493.
29. Zhai Y, et al. High-efficiency brain-targeted intranasal delivery of BDNF mediated by engineered exosomes to promote remyelination. *Biomaterials Sci*. 2022;10(19):5707–18.
30. Howe KL, et al. Ensembl 2021. *Nucleic Acids Res*. 2021;49(D1):D884–91.
31. Forbes S, et al. COSMIC 2005. *Br J Cancer*. 2006;94(2):318–22.
32. Gudmundsson S, et al. Variant interpretation using population databases: lessons from gnomAD. *Hum Mutat*. 2022;43(8):1012–30.
33. Yates AD et al. *Ensembl 2020*. *Nucleic acids research*, 2020. 48(D1): pp. D682–D688.
34. Tate JG, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 2019;47(D1):D941–7.
35. Dong S, et al. Annotating and prioritizing human non-coding variants with RegulomeDB v. 2. *Nat Genet*. 2023;55(5):724–6.
36. Wu W, et al. Nuclear factor-kappab regulates the transcription of NADPH oxidase 1 in human alveolar epithelial cells. *BMC Pulm Med*. 2021;21:1–12.
37. Du J, et al. KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Mol Biosyst*. 2014;10(9):2441–7.
38. Özkan Oktay E, et al. In silico prediction and molecular docking of SNPs in nrp1 gene associated with SARS-CoV-2. *Biochem Genet*. 2024;62(1):156–75.
39. Kaur R, Singh J, Kaur M. Structural and functional impact of SNPs in P-selectin gene: a comprehensive in silico analysis. *Open life Sci*. 2017;12(1):19–33.
40. Langdon WB, Petke J, Lorenz R. *Evolving better RNAfold structure prediction. in Genetic Programming: 21st European Conference, EuroGP 2018*, Parma, Italy, April 4–6, 2018, Proceedings 21. 2018. Springer.
41. Fricke M, et al. Global importance of RNA secondary structures in protein-coding sequences. *Bioinformatics*. 2019;35(4):579–83.
42. Stanfill AG, Cao X. Enhancing research through the use of the genotype-tissue expression (GTEx) database. *Biol Res Nurs*. 2021;23(3):533–40.
43. Westra H-J, Franke L. *From genome to function by studying eQTLs*. *Biochimica et Biophysica Acta (BBA)-molecular basis of Disease*, 2014. 1842(10): pp. 1896–902.
44. Stenson PD, et al. The human gene mutation database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum Genet*. 2020;139:1197–207.
45. Dutton-Regester K, et al. A highly recurrent RPS27 5'UTR mutation in melanoma. *Oncotarget*. 2014;5(10):2912.
46. Ellingford JM, et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med*. 2022;14(1):73.
47. Hindorf LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci*. 2009;106(23):9362–7.
48. Schuster SL, Hsieh AC. The untranslated regions of mRNAs in cancer. *Trends cancer*. 2019;5(4):245–62.
49. Khan K, et al. PRKCE non-coding variants influence on transcription as well as translation of its gene. *RNA Biol*. 2022;19(1):115–29.
50. Shah H, et al. Investigation of UTR variants by computational approaches reveal their functional significance in PRKCI gene regulation. *Genes*. 2023;14(2):247.
51. Yoo SS, et al. Putative functional variants of XRCC1 identified by RegulomeDB were not associated with lung cancer risk in a Korean population. *Cancer Genet*. 2015;208(1–2):19–24.
52. Cheema AN, Rosenthal SL, Ilyas M, Kamboh. Proficiency of data interpretation: identification of signaling SNPs/specific loci for coronary artery disease. *Database*. 2017;2017:pbx078.
53. Hosea R, et al. Targeting transcription factor YY1 for Cancer Treatment: current strategies and future directions. *Cancers*. 2023;15(13):3506.
54. Whitfield TW, et al. Functional analysis of transcription factor binding sites in human promoters. *Genome Biol*. 2012;13:1–16.
55. Kim K-P, et al. Biological importance of OCT transcription factors in reprogramming and development. *Exp Mol Med*. 2021;53(6):1018–28.
56. Wang X, et al. AP-2 α : a regulator of EGF receptor signaling and proliferation in skin epidermis. *J Cell Biol*. 2006;172(3):409–21.
57. Attafi IM, Bakheet SA, Korashy HM. The role of NF- κ B and AhR transcription factors in lead-induced lung toxicity in human lung cancer A549 cells. *Toxicol Mech Methods*. 2020;30(3):197–207.
58. Burbach JP, Smits S, Smidt MP. Transcription factors in the development of midbrain dopamine neurons. *Ann N Y Acad Sci*. 2003;991:61–8.
59. Dennis DJ, Han S, Schuurmans C. bHLH transcription factors in neural development, disease, and reprogramming. *Brain Res*. 2019;1705:48–65.
60. Hoffman JA, et al. Single-cell RNA sequencing reveals a heterogeneous response to glucocorticoids in breast cancer cells. *Commun Biology*. 2020;3(1):126.
61. Wu W, et al. Glucocorticoid receptor activation signals through fork-head transcription factor 3a in breast cancer cells. *Mol Endocrinol*. 2006;20(10):2304–14.
62. Liu X, et al. Systematic analysis reveals distinct roles of USF family proteins in various cancer types. *Int J Biol Mark*. 2023;38(3–4):243–52.
63. Briones-Orta MA et al. *Osteopontin splice variants and polymorphisms in cancer progression and prognosis*. *Biochimica et Biophysica Acta -Reviews on Cancer*, 2017. 1868(1): pp. 93–108. A.
64. Baek D, et al. The impact of microRNAs on protein output. *Nature*. 2008;455(7209):64–71.
65. Chhichholiya Y, et al. SNPs in miRNAs and target sequences: role in cancer and diabetes. *Front Genet*. 2021;12:793523.
66. Nimir M et al. *In silico analysis of single nucleotide polymorphisms (SNPs) in human FOXC2 gene*. *F1000Research*, 2017. 6.
67. Karimzadeh MR, et al. MicroRNA binding site polymorphism in inflammatory genes associated with colorectal cancer: literature review and bioinformatics analysis. *Cancer Gene Ther*. 2020;27(10):739–53.
68. Dodson RE, Shapiro DJ. *Regulation of pathways of mRNA destabilization and stabilization*. 2002.
69. Moudi M, et al. Polymorphism in the 3'-UTR of LIF but not in the ATF6B gene associates with schizophrenia susceptibility: a case-control study and in silico analyses. *J Mol Neurosci*. 2020;70:2093–101.
70. de Goede OM, et al. Population-scale tissue transcriptomics maps long non-coding RNAs to complex disease. *Cell*. 2021;184(10):2633–48. e19.

71. Rao S, et al. Involvement of the long intergenic non-coding RNA LINC00461 in schizophrenia. *BMC Psychiatry*. 2022;22(1):59.
72. Itan Y, et al. The human gene damage index as a gene-level approach to prioritizing exome variants. *Proc Natl Acad Sci*. 2015;112(44):13615–20.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.