Readers Insight
Publisher of open access books and journals
PJMLS
Print ISSN: 2707-4471. Online ISSN: 2707-448X

Research Article	Pak-Euro Journal of Medical and Life Sciences
DOI: 10.31580/pjmls.v6i3.3012	Copyright © All rights are reserved by Corresponding Author
Vol. 5 No. 4, 2022: pp. 386g-386u	
www.readersinsight.net/pjmls	**Revised:** August 31, 2023 **Accepted:** September 13, 2023
Submission: June 27, 2023	**Published Online:** September 30, 2023

A COMPREHENSIVE IN-SILICO EXPLORATION UNVEILING OF GENOMIC, STRUCTURAL AND FUNCTIONAL INSIGHTS

Fazal Shan[1]*, Ahmad Al Abad[2], Majid Ullah[3], Nighat Nawaz[4], Mihk Ikram[1]
Irshad Ahmad[1]*

[1]**Department of Molecular Biology and Genetics,** Khyber Medical University Peshawar, Pakistan

[2]**Department of Pathology,** Bannu Medical College, Bannu, Pakistan

[3]**Department of Medical Lab Technology,** Khyber Medical University Peshawar, Pakistan

[4]**Department of Chemistry,** Islamia College, Peshawar, Pakistan

***Corresponding Authors:** Fazalshan. E. mail: fazalshan.ibms@kmu.edu.pk
and Dr. Irshad Ahmad. E. mail: irshadahmad@kmu.edu.pk

Abstract

Our study explores the intricate realm of IKZF3, unraveling its genomic, structural, and functional nuances through a comprehensive In-silico analysis. IKZF3, a key transcription factor, plays a pivotal role in diverse biological processes, including immune response and cellular differentiation. Leveraging advanced computational tools, we scrutinized the genomic landscape of IKZF3, identifying key regulatory elements and potential binding sites. Our structural analysis unveiled novel insights into the three-dimensional architecture of IKZF3, shedding light on its conformational dynamics and potential interaction interfaces. Furthermore, we explored the functional implications of IKZF3 in cellular processes, emphasizing its role in modulating gene expression networks. By integrating multi-omics data, we uncovered potential downstream targets and pathways influenced by IKZF3 activity. The findings presented in this study provide a holistic understanding of IKZF3, bridging the gap between genomics and functional biology. This In-silico exploration not only contributes to the growing body of knowledge surrounding IKZF3 but also lays the foundation for future experimental investigations. The insights gained from this study pave the way for targeted experiments to validate the predicted genomic regulatory elements, decipher the functional consequences of IKZF3 interactions, and explore its therapeutic implications in various biological contexts. As we navigate the intricate landscape of transcriptional regulation, this work serves as a valuable roadmap for researchers seeking to unravel the complexities of IKZF3 and its multifaceted roles in cellular physiology.

Keywords: *Functional implications, Genomic Analysis, IKZF3, Structural Dynamics, Therapeutic Potential, Transcriptional Regulation*

INTRODUCTION

The gene *IKZF3* codes for a zinc-finger protein belonging to the Ikaros family, crucial for lymphocyte growth progress. This family includes Ikaros, Aiolos, and Helios, with *IKZF3* primarily influencing B lymphocytes. Ikaros and Aiolos are known to impact gene expression by remodeling chromatin and altering DNA accessibility and structure. Both homodimers and heterodimers of these proteins regulate gene expression in B lymphocytes. *IKZF3*'s role in immune cell differentiation and maturation is crucial, and its involvement in autoimmune diseases and hematological malignancies underscores its significance, making it a subject of in-depth molecular exploration (1, 2). In the realm of structural annotation, researchers utilize various bioinformatic techniques to unravel the genetic and genomic architecture of the *IKZF3* gene. This includes identifying exon-intron boundaries, promoter regions, and regulatory elements, as well as analyzing genetic variations, such as single nucleotide polymorphisms (SNPs), within the gene's sequence. These insights provide a foundation for understanding the gene's

regulatory mechanisms and its potential implications in various biological processes (3). Functional annotation, on the other hand, focuses on deciphering the protein product of the *IKZF3* gene. This entails examining the protein's domain structure, motifs, post-translational modifications, and the prediction of its three-dimensional conformation. In silico methods also allow for the exploration of protein-protein interactions, identification of binding partners, and prediction of downstream pathways influenced by the *IKZF3* protein. These analyses offer a comprehensive perspective on how *IKZF3* participates in cellular and molecular processes (4).

In recent years, in silico techniques have become indispensable tools in genomics and functional genomics research. With the availability of vast genomic databases and powerful computational algorithms, scientists can gain insights into the role of *IKZF3* in diverse biological contexts, such as hematopoiesis, immune response, and autoimmune diseases. This in silico approach not only enhances our understanding of the gene's fundamental biology but also has significant implications for therapeutic target discovery and the development of precision medicine approaches (5).

In this study, we will delve into the in silico structural and functional annotation of the *IKZF3* gene. By leveraging computational resources and techniques, we aim to uncover hidden facets of this gene's biology that may have far-reaching implications in the fields of immunology, oncology, and autoimmune diseases.

MATERIALS AND METHODS

SEQUENCE RETRIEVAL, HOMOLOGY, AND PHYLOGENETIC TREE ANALYSIS

In order to determine the genomic localization of the IKZF3 gene, the human genome draft sequence was meticulously examined by conducting a thorough search on the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/)(6). This allowed researchers to pinpoint the specific location of the IKZF3 gene within the human genome. Additionally, to gain valuable insights into the gene's evolutionary history and variation among different species, a phylogenetic analysis was carried out. Peptide sequences of the IKZF3 gene from various species, including humans, Rattus norvegicus, Mus musculus, Bos tauru and Balaenoptera davidsoni were obtained from the Uniprot database (http://www.uniprot.org/). The phylogenetic analysis is a fundamental tool that helps researchers understand the evolutionary trends of the IKZF3 gene across different species. To construct this phylogenetic tree, an online multiple sequence alignment tool, MUSCLE, was employed. This tool utilizes the neighbor-joining method, a widely used algorithm for building phylogenetic trees. Default parametric settings were applied during the analysis.

PHYSICO-CHEMICAL CHARACTERIZATION OF IKZF3

PROTPARAM SERVER

To analyze physical and chemical parameters, as well as structural and functional motifs for the human IKZF3 protein sequence, the Protparam Server was utilized. This online tool (http://web.expasy.org/protparam/) provides valuable information about various protein properties (7).

SIGNALP V4.1

SignalP v4.1 online tool (http://www.cbs.dtu.dk/services/SignalP/) use for Signal peptide prediction for IKZF3. This tool is widely used to identify signal peptides and cleavage sites in protein sequences, which are critical for protein trafficking and localization (8).

PROTSCALE SERVER

The Protscale Server (http://web.expasy.org/protscale/) was employed to analyze several physiochemical properties for the human IKZF3 protein. These properties included polarity, percentage of buried residues, average flexibility, bulkiness, hydrophilicity, and relative mutability. Protscale is a valuable resource for characterizing the physico-chemical properties of protein sequences (7).

m SERVER

To analyze overall physical and chemical parameters, as well as structural and functional motifs for the human IKZF3 peptide sequence, the Protparam Server was utilized. This online tool (http://web.expasy.org/protparam/) provides valuable information about various protein properties (7).

FUNCTIONAL DOMAIN ASSESSMENT OF IKZF3 PROTEIN

DOMAIN MAPPING OF DISEASE MUTATIONS (DMDM) DATABASE

The DMDM database (http://www.bioinfo.umbc.edu/dmdm/) was used to study the functional domains of the IKZF3 protein. DMDM employs a Hidden Markov Model-based sequence alignment tool (HMMer) to map genetic polymorphisms, such as mutations and variations, onto the protein's domain structure. Default parametric settings were applied for this analysis.

BIOMUTA V3.0 DATABASE

To check for disease associations with genetic polymorphisms, the BioMuta v3.0 database (https://hive.biochemistry.gwu.edu/biomuta) was utilized. This database is particularly curated for single-nucleotide variations (SNVs) and their associations with diseases, with a focus on oncogenesis.

POLYMORPHISM ANALYSIS OF IKZF3 PROTEIN

POLYPHEN PREDICTION TOOL

Polyphen prediction tools were employed to assess the impact of variations on the functionality of the IKZF3 protein. The PolyPhen tool was used to annotate coding nonsynonymous single-nucleotide polymorphisms (SNPs). It employs a high-quality multiple sequence alignment pipeline based on machine learning methods, optimized for high-throughput analysis of next-generation sequencing data.

SUB-CELLULAR LOCALIZATION ANALYSIS

CELLO SERVER

The sub-cellular localization of the IKZF3 protein was evaluated using the CELLO server (http://cello.life.nctu.edu.tw/). CELLO is a computational tool that predicts the sub-cellular localization of proteins based on various features derived from their primary amino acid sequences.

TMHMM AND HMMTOP SERVERS

The transmembrane structural domains within the IKZF3 protein were assessed using the TMHMM and HMMTOP servers. These servers utilize hidden Markov models to predict transmembrane helices in protein sequences. This information is essential for understanding the protein's membrane topology and its potential roles in cellular processes.

UNIPROT

UniProt (http://www.uniprot.org/) was also utilized as a resource for sub-cellular localization information. UniProt is a comprehensive database that provides detailed information about proteins, including their sub-cellular localization, functional annotations, and other relevant data.

METHYLATION SITE ANALYSIS

The analysis of methylation sites in the IKZF3 gene was conducted using the MethyCancer tool, which can be accessed at http://methycancer.psych.ac.cn/MethyCancer.do. This resource provides valuable information regarding DNA methylation patterns and their associations with cancer and other biological processes (9).

PREDICTION OF POST-TRANSLATIONAL MODIFICATIONS (PTMS)

For the prediction of post-translational modifications (PTMs) on the IKZF3 protein, researchers utilized various servers provided by the Center for Biological Sequence Analysis CBS (http://www.cbs.dtu.dk/services/).

PREDICTION OF PTMS

NetCGlyc 1.0

This tool was used to predict sites for glycosylation on the IKZF3 protein. Glycosylation is a common PTM that involves the attachment of sugar molecules to proteins, often influencing their structure and function (10).

NetNGlyc 1.0

NetNGlyc was utilized to predict N-linked glycosylation sites on the IKZF3 protein. N-linked glycosylation is a specific type of glycosylation with important regulatory roles. (11)

NetOGlyc 4.0

This tool was employed to predict sites for O-linked glycosylation on the IKZF3 protein. O-linked glycosylation involves the attachment of sugar molecules to specific amino acid residues. (12)

NetPhos 3.1

NetPhos was used to predict phosphorylation sites on the IKZF3 protein. Phosphorylation is a common PTM that plays a crucial role in cellular signaling and regulation.(13)

ProP 1.0

The was used for the prediction of an unspecified PTM. The specific nature of this modification is not mentioned in the provided information.(14)

FUNCTIONAL PROTEIN ASSOCIATION NETWORK ANALYSIS

STRING V10 DATABASE

The analysis was performed using STRING v10 (http://string-db.org/), a comprehensive database that specializes in protein-protein interaction data and functional associations. STRING provides information on direct and indirect interactions between proteins, helping researchers to understand how proteins cooperate in various biological processes (15).

MUTANT PROTEIN STRUCTURE BUILDING IDENTIFIED BY DMDM

TEMPLATE SEARCH VIA BLAST

To predict the protein structure via homology modeling, the normal IKZF3 crystal structure was obtained from the Protein Data Bank (PDB) at www.rcsb.org was initially searched using BLAST. To identify suitable templates with good sequence identity and coverage, typically requiring a query sequence identity greater than 40% and query coverage exceeding 50%. This crystal structure served as a template for generating 3D structures for each mutant sequence using the Swiss Model online protein prediction server (https://swissmodel.expasy.org/). The resulting structures were assessed using the Rampage server to create Ramachandran plots, allowing for the evaluation of predicted structure quality. Visualization of the predicted protein structures was performed using Biovia Discovery Studio version 4.5. Additionally, the effects of mutations on the protein's 3D structure were observed through protein structure alignment using the SuperPose server at https://wishart.biology.ualberta.ca/SuperPose/.

SWISS MODEL PROTEIN PREDICTION

The identified similar domains were used for the structure prediction via the Swiss Model online protein prediction server (http://swissmodel.expasy.org/).The structure was built domain by domain, considering the structural similarities.

QUALITY ASSESSMENT

The predicted protein structure was subjected to quality assessment using the Rampage server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). The Ramachandran plot was analyzed to evaluate the quality of the predicted structure. The Ramachandran plot helps assess the stereochemical quality of the model. Visualization:

The predicted protein structure was visualized using Biovia Discovery Studio v4.5 (https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/). This software allows for the visualization and analysis of 3D protein structures.

PREDICTION OF DRUG BINDING POCKETS PROFILE

To assess the drugability profile of selected proteins and predict the quality and quantity of drug binding sites, we utilized the DoGSite scorer server. This tool was employed to identify binding pockets within the proteins and evaluate their respective Drug scores. During this analysis, we collected information about pocket size and drug scores. The descriptors used for scoring included volume, hydrophobicity, and shape, all of which fell within a range of 0 to 1. A score of 1 indicated the highest potential for druggability. We considered drug binding pockets with a predicted drug score greater than 0.5. In cases where multiple pockets exceeded this cutoff score, we selected the top five pockets for further investigation.

RESULTS

SEQUENCES' RETRIEVAL AND HOMOLOGY TREE CONSTRUCTION

The gene *IKZF3* is located on the long arm (q arm) of human chromosome 17, specifically in the 17q12-q221.1 region. It spans from position 39864312-39757715, covering a total of 106597base pairs and is composed of 09 exons that encode a protein with 509 amino acids. ERBB2, L0C111089947, MIEN1, KRT8P4, ZPBP2, GSDMB, GRB7, RPL39P4 are the neighboring gene. It also known as AIO, IMD84, AIOLOS, ZNFN1A3. Figure 1a-c in the provided information illustrates the chromosomal location of *IKZF3*, its neighboring genes, and their respective positions and exon structures. A peptide sequence analysis of the *IKZF3* gene from humans and various other species was conducted using MUSCLE. The results of this analysis were used to create a homology tree (depicted in Figure 2). In this tree, human *IKZF3* was found to be closely related to great apes such as Rattus norvegicus (D3ZSW3_RAT) and, suggesting that they belong to the Mammalia family. Ma's nigh monkey (A0A2K5EG69_AOTNA) also grouped closely with human *IKZF3*.

Fig. 1. Multiple sequence alignment of IKZF3 gene

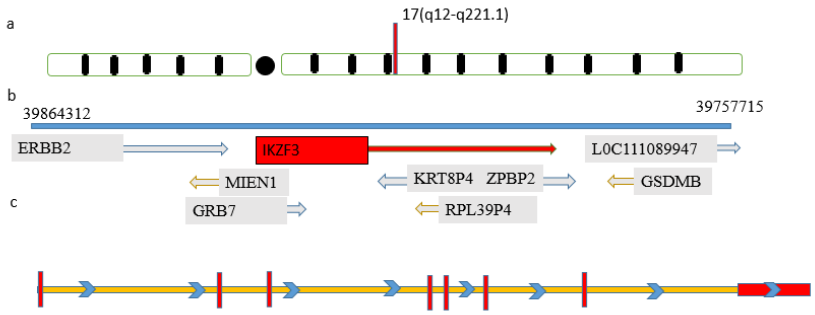


Fig. 1. IKZF3 gene location, structure and domains. a Chromosomal localization of D IKZF3 gene on q arm of human chromosome 17 highlighted in red , contains 09 exons (red bars)

PREDICTION OF SIGNAL PEPTIDE AND STRUCTURAL DOMAINS OF *IKZF3*

The SignalP Server analysis for *IKZF3* indicated that no signal peptide site is predicted, with a score of 0.002, as depicted in Fig. 3(a). The server also calculated C-scores, S-scores, and Y-scores for *IKZF3*. Notably, all these scores are below the standard value of 0.5. This suggests that no signal peptide to be present in the *IKZF3* protein. In terms of transmembrane domains, information obtained from the THMM database, which is curated and reviewed, indicated that *IKZF3* total length is 509 amino acids, and according to predictions, it does not contain any transmembrane helices (TMHs). The expected number of amino acids within transmembrane helices is very low, at 0.02079, suggesting an absence of such structures. Similarly, no amino acids within the first 60 residues are expected to be part of transmembrane helices, with an expected count of 0. The total probability of an N-terminal signal peptide is quite low, at 0.00273. According to the TMHMM2.0 output, the protein is classified as being outside of the membrane from amino acid 1 to 509, indicating that it is not integrated within the lipid bilayer shown in Fig. 3(b).

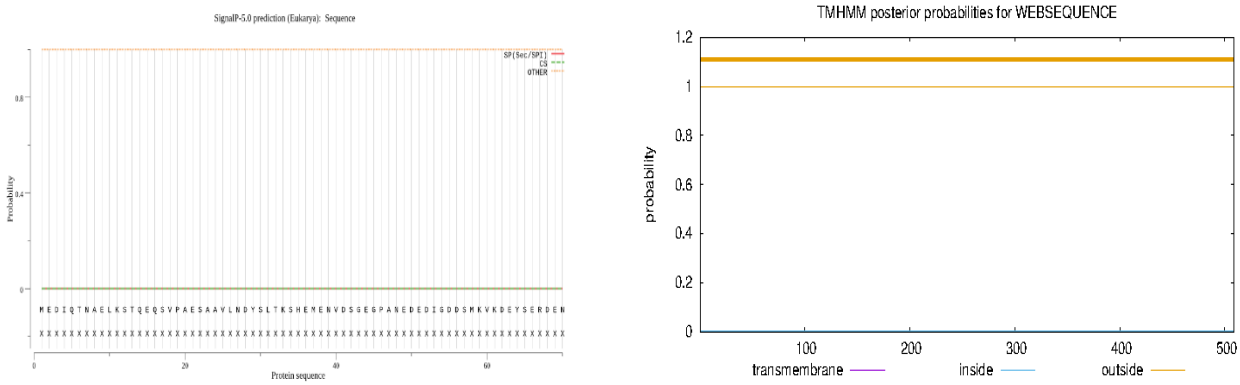


Fig. 3(a). Signal peptide not detected for IKZF3 protein using peptide sequence . **2 (b)** *transmembrane domains, information obtained from the THMM database*

FUNCTIONAL DOMAIN ANALYSIS

DMDM serves as a centralized portal for the comprehensive study of functional domains within *IKZF3*, offering valuable insights as illustrated in Figure 3 and detailed in Table I. Some major domains analyzed in the protein. The C2H2 zinc finger (zf_C2H2), characterized by conserved cysteines and histidines coordinating a zinc ion in a specific pattern, and forms a domain with crucial residues in stabilizing its fold. This domain, often involved in DNA binding, notably recognized in transcription factor Sp1, displays specificity beyond the typical consensus sequence, as observed in high-affinity binding to sequences such as the GAG (=CTC) repeat in the wt1 promoter. The ZnF_C2H2 domain is a common structural motif found in human proteins, occurring 723 times across 1165 different proteins. However, no specific pairwise interactions have been identified for this domain, and a total of 249 mutations have been observed within it, with 28 of these mutations associated with disease. This suggests its potential role in various cellular functions and its relevance in disease mechanisms.

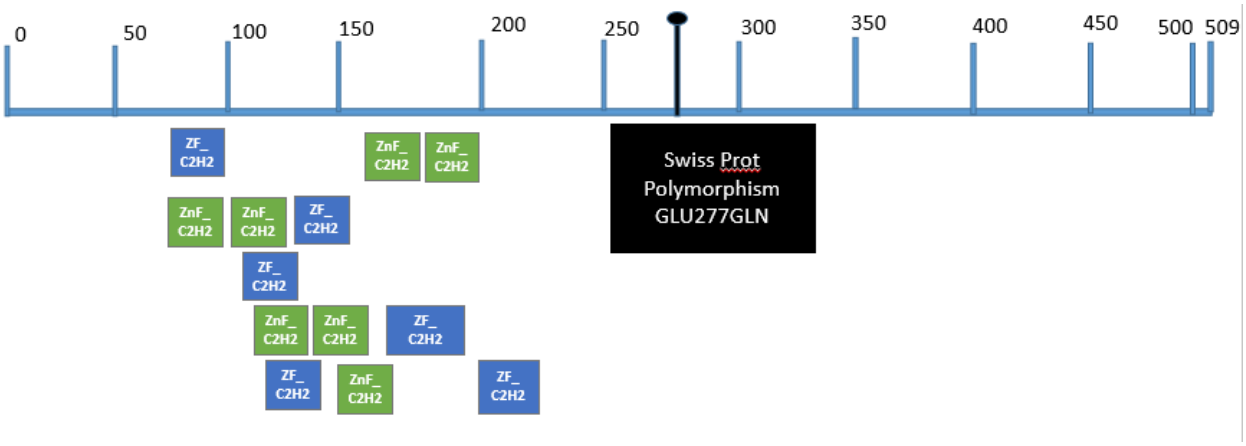


Fig. 4. IKZF3 contains various domains ZnF_C2H2 - smart00355, zf-C2H2 - pfam00096

Table I. Functional domains of IKZF3

Family	Domain name	Cd accession	Start	End
znf-C2h2	Zinc-finger double domain	Smart00355	59	81
znf-C2h2	Zinc-finger double domain	Smart00355	87	109
znf-C2h2	Zinc-finger double domain	Smart00355	112	134
znf-C2h2	Zinc-finger double domain	Smart00355	140	162
znf-C2h2	Zinc-finger double domain	Smart00355	146	168
znf-C2h2	Zinc-finger double domain	Smart00355	174	196
zf-C2h2	Zinc-finger double domain	Pfam00096	59	81
zf-C2h2	Zinc-finger double domain	Pfam00096	112	134
zf-C2h2	Zinc-finger double domain	Pfam00096	115	137
zf-C2h2	Zinc-finger double domain	Pfam00096	146	168
zf-C2h2	Zinc-finger double domain	Pfam00096	168	190
zf-C2h2	Zinc-finger double domain	Pfam00096	202	224

PREDICATION OF POLARITY, % BURIED RESIDUES, AVERAGE FLEXIBILITY, BULKINESS, HYDROPHILICITY, AND RELATIVE MUTABILITY

IKZF3 was subjected physio-chemical analyses (polarity, % buried residues, average flexibility, bulkiness, hydrophilicity, and relative mutability) using Protscale Server available at expasy platform. Higher the generated score greater is the probability for each parameter. The values for polarity are calculated with highly 222 aa polar residue to be with 39.771and least polar one 126 aa having value 0.782. The bulkiness are between 0.403 (45aa, 46aa) and 0.857 (463 aa), while accessibility is calculated to be 0.221 (470 aa) and 0.745 (265 aa). The least buried residue is 222 with calculated values of 0.120and most buried one is 22aa valuing 0.709. The relative mutability calculations showed a patch of highly mutable region comprising of residue 309 with value 0.783, while least mutable residue was 204 aa with values 0.267. The average flexibility values are in the range of 0.0.296 (470 aa) to 0.856 (53 aa). While Hydrophobicity is estimated using Hopp and Woods method, and the values ranged from 0.319 (463 aa) to 0.861 (65 aa). The maximum and minimum scores were calculated for *IKZF3* protein show in Table II.

Table II. Physio-chemical analysis of IKZF3 protein by Protscale server ExPasy

Polarity / Zimmerman				**% Bulkiness**				**% Accessible residues**			
Highest		**Lowest**		**Highest**		**Lowest**		**Highest**		**Lowest**	
Position	Score	Position	Score	Position	Score	Position	Score	Position	Score	Position	Score
222	39.771	126	0.782	463	0.857	45, 46	0.403	265	0.745	470	0.221
% Buried residues				**Relative mutability**				**Average Flexibility**			
Highest		**Lowest**		**Highest**		**Lowest**		**Highest**		**Lowest**	
Position	Score	Position	Score	Position	Score	Position	Score	Position	Score	Position	Score
22	0.709	222	0.120	309	0.783	204	0.267	53	0.856	470	0.296
Hphob./Hopp and Woods											
Highest		**Lowest**		**Highest**		**Lowest**		**Highest**		**Lowest**	
Position	Score	Position	Score	Position	Score	Position	Score	Position	Score	Position	Score
65	0.861									463	0.319

GENETIC POLYMORPHISMS AND CANCER ASSOCIATION FOR *IKZF3*

In this study, genetic polymorphisms in the *IKZF3* gene were analyzed using two different databases, DMDM and BioMuta. While DMDM identified one specific polymorphisms (p.GLU277GLN) in *IKZF3* shown in figure 5, it did not associate these variations with any disease condition. On the other hand, BioMuta provided a comprehensive overview of *IKZF3* genetic polymorphisms associated with various cancer types and their frequencies. The observed single nucleotide polymorphisms (SNPs) for the *IKZF3* gene in different cancer types were as follows 65 in uterine melanoma, 30 in uterine and lungs cancer , 29 in malignant glioma, 106 in skin cancer, 15 in kidney cancer, 14 in breast cancer, 27 in stomach cancer and 26 in liver cancer (Fig. 6a). Among these, the top three most frequent SNPs were located at amino acid position 322 with nine variants, position 458 with eight variants, position 47 with 7 variants. The functional impact of somatic cancer single-nucleotide variants (SNVs) in *IKZF3* was predicted using PolyPhen2. Furthermore, a domain-wise analysis of mutations frequency distribution revealed a higher number of mutations in the zinc finger znf_c2h2 (249), with 28 classified as pathogenic mutations, and others as lacking functional information. It is crucial as it determines the subcellular location of the protein, which is the plasma membrane. In summary, the study provides insights into genetic polymorphisms in the *IKZF3* gene and their potential associations with various cancer types, offering valuable information on the functional impact of somatic cancer SNVs within different protein domains.

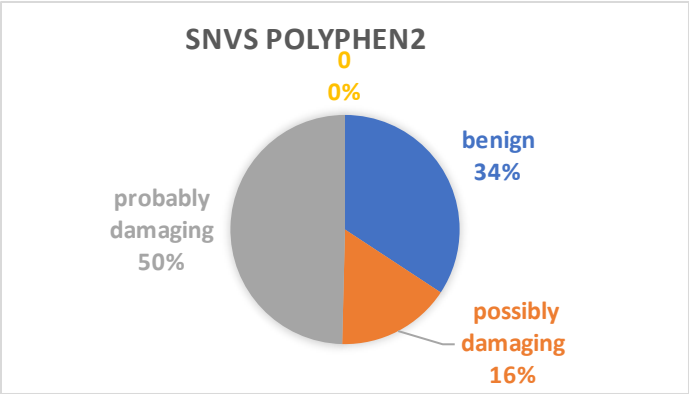


Fig. 5. PolyPhen2 tool classified SNVs three categories benign, probably damaging and possibly damaging

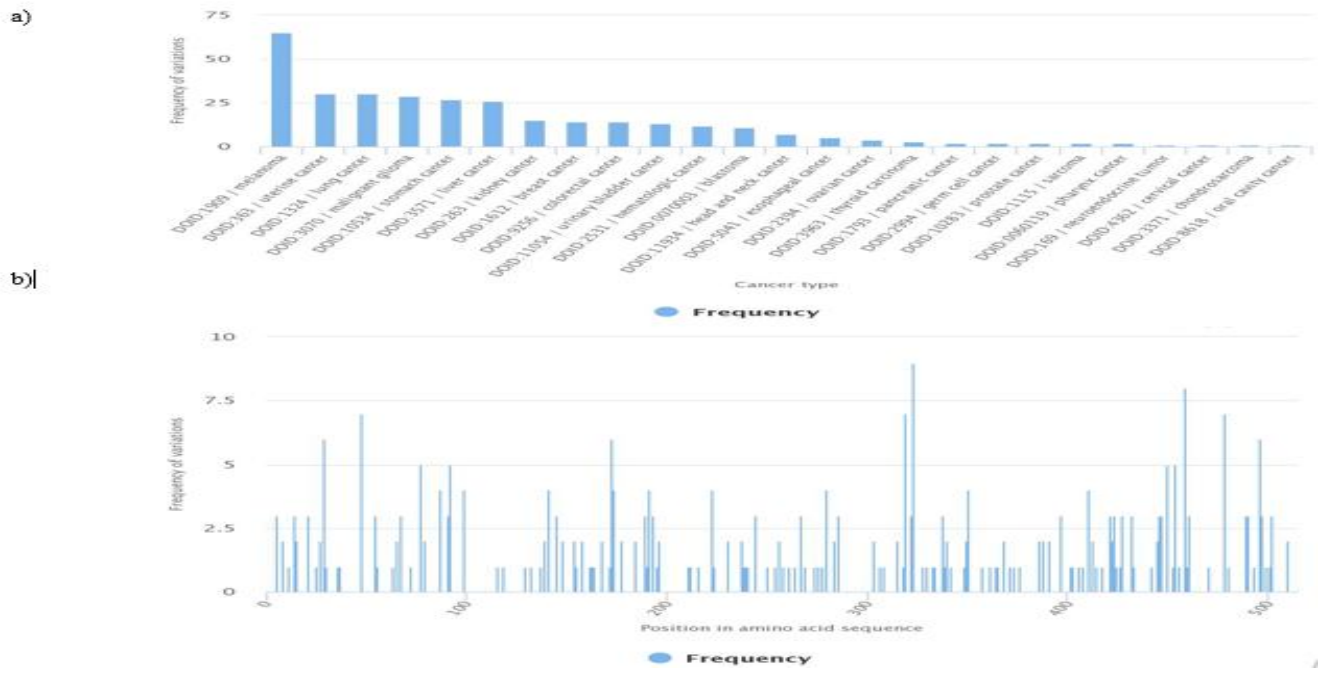


Fig. 6. Examining the SNPs in the IKZF3 gene among cancer patients. (a) Observed specific genetic variations in the IKZF3 gene associated with different types of cancer. **(b)** The frequency of SNPs in the IKZF3 gene across its entire length in cancer patients, *determining the positions where these genetic variations occur*

POST-TRANSLATIONAL MODIFICATIONS OF *IKZF3*

The CBS prediction server was utilized to identify several potential post-translational modifications for the *IKZF3* protein. The chosen cutoff value for each parameter was set at 0.5, with higher values indicating a greater likelihood of the modification occurring. Specifically, the ProP 1.0 prediction server predicated no propeptide signal cleavage site is situated between Arg(R)/Lys (K) Fig. 7(a). Regarding glycosylation, the NetCGlyc 1.0 server did not detect any C-mannosylation sites on the protein. However, the NetAcet 1.0 server identified a total of N-terminal acetylation sites. Additionally, the NetNGlyc 1.0 and NetOGlyc 4.0 servers calculated 1 N-linked glycosylation sites and not predicted any O-linked glycosylation sites, respectively. Phosphorylation sites were assessed using the NetPhos 3.1 server, which predicted a total of 103 potential phosphorylation sites on the *IKZF3* protein. Notably, serine residues were expected to host the highest number of phosphorylation sites (68), followed by threonine with 23 sites, and tyrosine with only 12 sites.

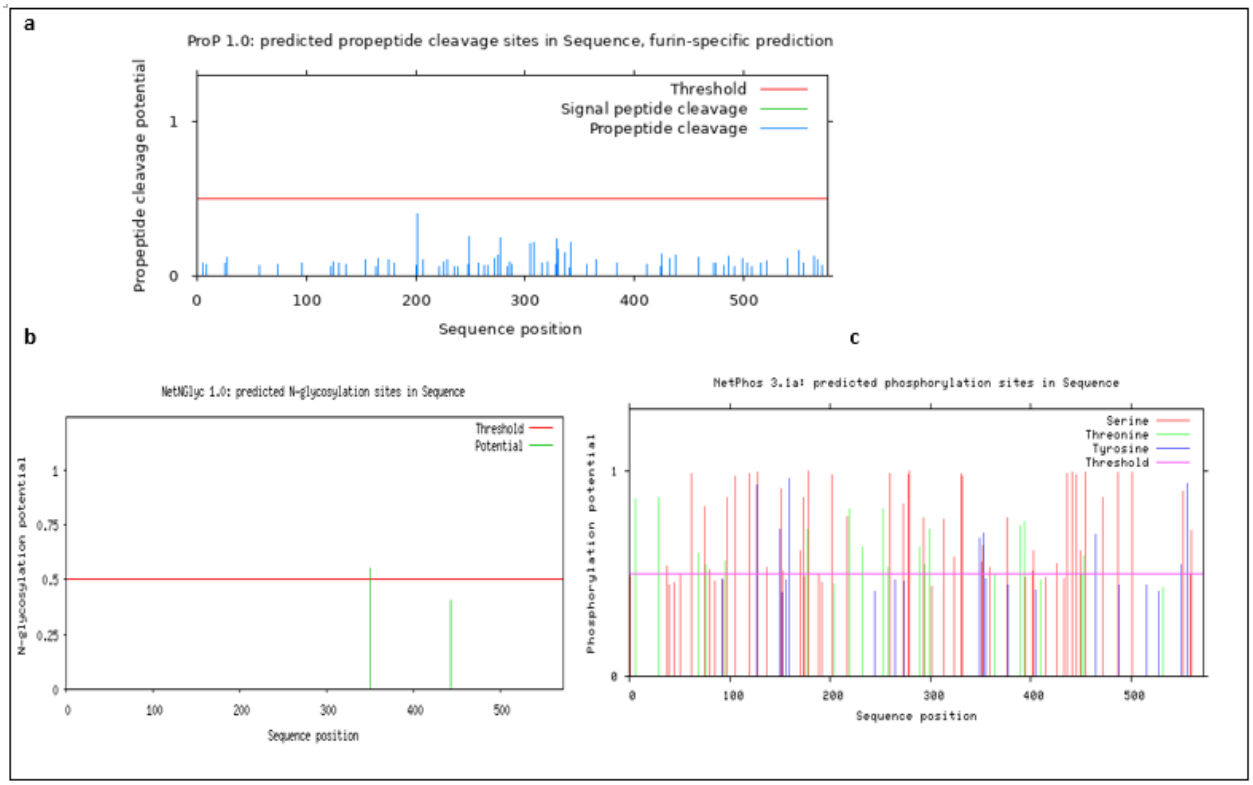


Fig. 7. Predicted post-translational modifications of the IKZF3 protein based. **(a)** Propeptide signal cleavage site between Arg(R)/Lys graph (K). **(b)** N-glycosyltion sites graph. **(c)** Phosphorylation sites graph

PHYSIOLOGICAL CHARACTERIZATION OF *IKZF3* PROTEIN USING PROTPARAM ANALYSIS

The ProtParam server was used to predict various physiological characteristics of *IKZF3*. It was found that *IKZF3* has a molecular weight of approximately 58023.15 Da and a molecular formula speculated to be $C^{2486}H^{3914}N^{734}O^{795}S^{38}$. The total number of atoms in the protein is 8130. The theoretical isoelectric point (pI) of *IKZF3* is estimated to be 8.12. In terms of stability, the half-life of *IKZF3* was estimated to be 30 hours in mammalian reticulocytes when tested in vitro, and it was greater than 20 hours when tested in yeast in vivo. The protein was found to be slightly unstable, with an instability index of 53.71. Additionally, the aliphatic index of *IKZF3* was determined to be 61.49. The estimated grand average of hydropathicity (GRAVY) for *IKZF3*1 is -0.793 , indicating its hydrophilic nature.

ANALYSIS OF METHYLATION SITES AND SUBCELLULAR LOCALIZATION OF *IKZF3* GENE AND PROTEIN

The methylation sites of the *IKZF3* gene were predicted using the Methycancer tool. The analysis revealed DNA methylation occurring on Chromosome No. 17, specifically between the start site at

35,171,724 and the end site at 35,276,967. Furthermore, the subcellular localization of *IKZF3* protein was assessed using Cello v2.5, which indicated a nuclear localization score of 4.769. This information provides valuable insights into the regulation and localization of *IKZF3*, shedding light on its potential role in cellular processes.

TRANSMEMBRANE STRUCTURE ANALYSIS OF *IKZF3* PROTEIN USING TMHMM SERVERS

The TMHMM servers were employed to evaluate the transmembrane structural aspects within the *IKZF3* protein. These servers use hidden Markov models to predict transmembrane helices present in protein sequences shown in figure 9. Upon analysis of the *IKZF3* protein, which has a length of 509 amino acids, the TMHMM prediction indicated that there are no predicted transmembrane helices within the sequence. The expected number of amino acids in transmembrane helices was very low at 0.02079, and there were no transmembrane helices predicted within the first 60 amino acids.

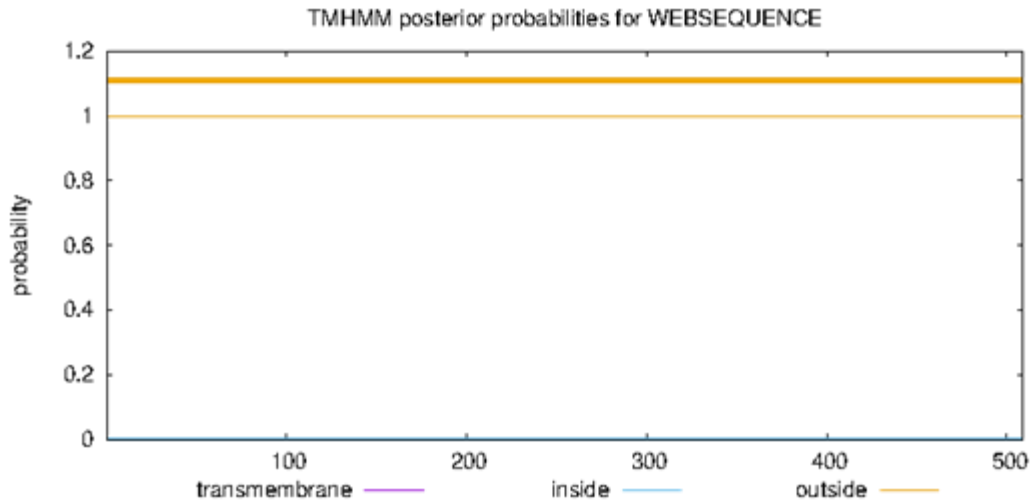


Fig. 8. IKZF3 Transmembrane helices graph

HOMOLOGY MODELING

IKZF3 WILD AND MUTANT PROTEIN STRUCTURE PREDICTION VIA HOMOLOGY MODELING

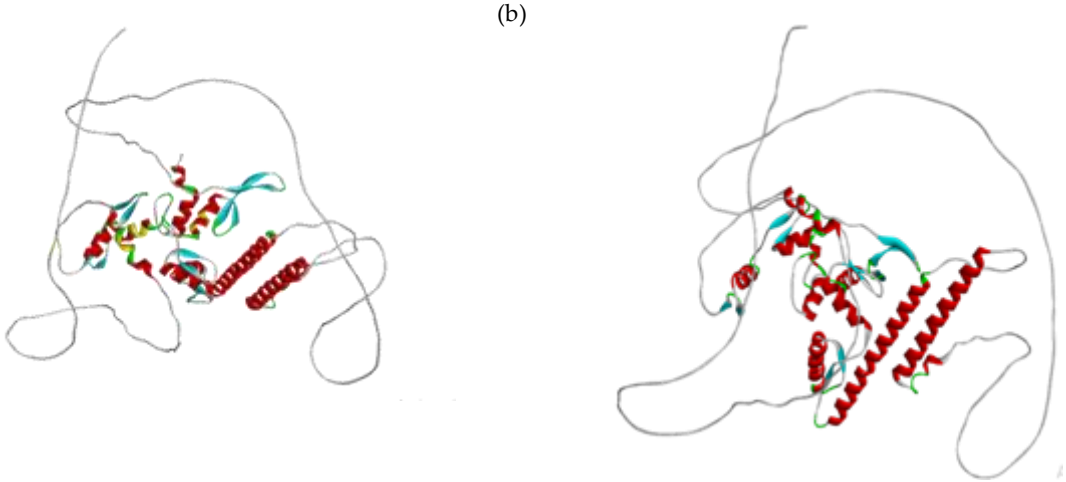
The crystal structures of the wild-type *IKZF3* protein were unavailable on the PDB database. We developed both wild-type and mutant alpha fold model. The mutant alpha fold model by introducing amino acid modifications into the protein sequence. Subsequently, model structures for the *IKZF3* wild-type and mutant proteins were constructed using the Swiss Model. The selection of the best structures was based on high Global Model Quality Estimation (GMQE) scores, ranging from 0 to 1, indicating their reliability (refer to Figure 10a-b). For validation, homologous structures of the mutant proteins were chosen, and their stereochemical quality was assessed through Ramachandran plot analysis (see Figure 10a-d). Overall, the variant structures exhibited no significant changes, but a closer examination of the Ramachandran plots revealed no effects on the placement and strain of amino acids within the structures when compared in a comparative manner.

SUPERIMPOSITION OF BOTH MUTANT AND WILD PROTEIN

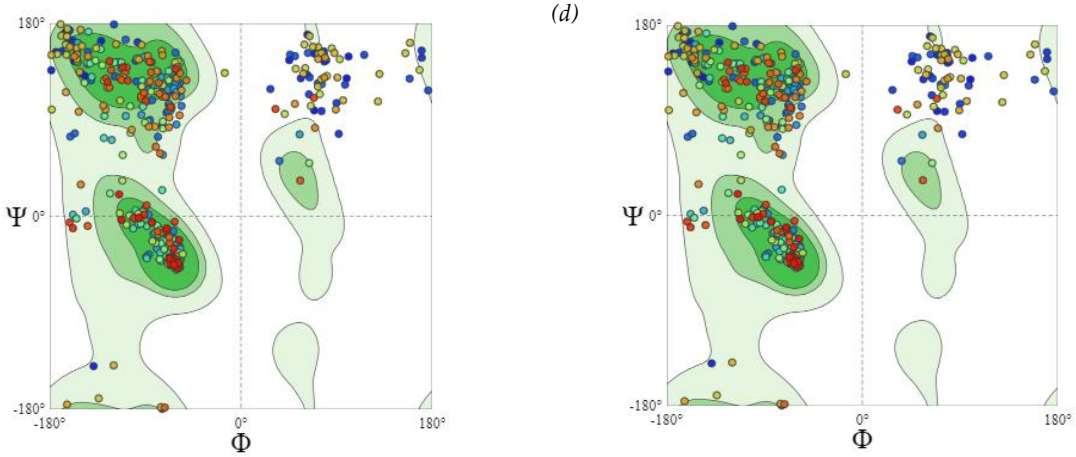
To investigate the impact of point mutations on the protein structure, we employed a consensus approach by integrating two complementary methods: mCSM (Mutation Cutoff Scanning Matrix) and SDM (Site Directed Mutator). The integration was achieved through the utilization of support vector machines (SVMs). To assess the structural changes, the "Protein Superposition Server" bioinformatics tool was employed. In this process, the wild-type and mutant protein structure in PDB formst as the input. Additionally, amino acid information, represented in one-letter codes, for both the wild-type and mutant

forms were incorporated (Fig. 9e). This comprehensive approach showed no positive effect of mutation was observed.

(a) (b)

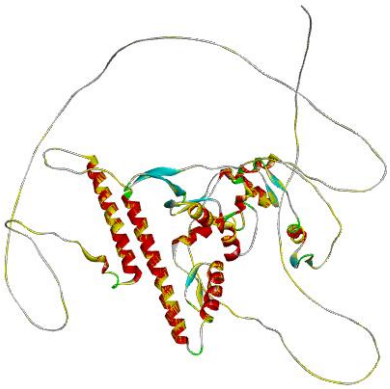


(c) (d)



Wild IKZF3 structure and Ramachandran plot *Mutant IKZF3 Structure and Ramachandran plot*

(e)



Model three-dimensional structure of IKZF3 variant and Superimpose onto the normal structure. variant in red and blue, while wild is represented in yellow

Fig. 9 (a-e). IKZF3 3D structure of variant, Wild along with Ramachandran plot and Superimposition onto the normal structure

COMPREHENSIVE ANALYSIS OF *IKZF3* PROTEIN: PROTEIN-PROTEIN INTERACTIONS AND FUNCTIONAL SIGNIFICANCE IN CELLULAR PROCESSES

Protein-protein interaction analysis was conducted using the STRING Database, resulting in a comprehensive network map (Fig. 9) that revealed multiple interactions. The network consisted of 21 nodes and 134 edges, with an average node degree of 12.8. The clustering coefficient was 0.799, indicating a significant level of interconnectedness among the proteins. The PPI enrichment value was notably low, at <1.0e-16 suggesting a significant enrichment of protein-protein interactions. (Fig. 9a). *IKZF3* protein involved in multiple process like Biological process GO: 0038172 interleukin-33-mediated signaling pathway GO: 0018076 N-terminal peptidyl-lysine acetylation GO: 0048293 regulation of isotype switching to IgE

isotypes GO: 0002826 negative regulation of T-helper 1 type immune response GO: 0033598 mammary gland epithelial cell proliferation. Molecular process GO: 0043425 bHLH transcription factor binding GO: 0001102 RNA polymerase II activating transcription factor binding GO: 0005160 transforming growth factor beta receptor binding GO: 0070888 E-box binding GO: 0001221 transcription cofactor binding. Cellular component GO: 0042025 host cell nucleus GO: 0005667 transcription regulator complex GO: 0000790 nuclear chromatin GO: 0005694 chromosome GO: 0005654 nucleoplasm. KEGG pathways hsa05321 inflammatory bowel disease, hsa05310 Asthma.

Fig. 10. *IKF3* Protein interaction with other multiple proteins, *PIK3CA, JAK1, SOS2, GRB2, LCK, HRAS*

ASSESSMENT OF *IKZF3* PROTEIN'S DRUGABILITY PROFILE THROUGH STRUCTURE-BASED ANALYSIS AND POCKET IDENTIFICATION

The assessment of drugability profiles based on protein structures is crucial for understanding the function of proteins. Utilizing advanced tools like protein pocket discovery, the DoGSite scorer (protein-plus) was employed to identify specific pockets within the protein. Among these pockets, those with drug scores exceeding 0.50 were initially chosen, and the best pockets were subsequently selected, focusing on those with a generated drug score of 0.80 or higher. In the case of *IKZF3*, the table 2 and Figure 10 display the best pockets. The pocket labeled "P-" received the highest drug score, followed by the second and third high-scoring pockets, denoted as "P-" and "P-", with drug scores of approximately 0 and 0, respectively.

Table III. IKFZ3 protein drug binding pockets interaction score and simple score

Pocket	Volume (A^3)	Surface (A^2)	Drug score	Simple score
P_1	1622.9	2323.98	0.82	0.74
P_0	7635.63	9039.78	0.8	0.7
P_2	1541.19	2316.33	0.8	0.71
P_3	491.72	957.43	0.75	0.42

Fig. 11. IKZF3 protein with drug binding pockets

DISCUSSION

IKZF3, a gene encoding a transcription factor, has been implicated in various diseases, including certain types of leukemia, autoimmune diseases, and inflammatory disorders. Its role in regulating immune system function and gene expression makes it a subject of ongoing research for potential therapeutic interventions in these conditions (16-18). The conducted study has provided a comprehensive analysis of the *IKZF3* gene and its protein product. The methodology employed a range of bioinformatics tools and databases to investigate the genomic localization, evolutionary history, physicochemical properties, functional domains, genetic polymorphisms, post-translational modifications, drug binding potential, and protein-protein interactions of *IKZF3*. The research thoroughly analyzed the IKZF3 sequence using UniProt, a comprehensive protein sequence database (19). The study systematically investigated proteins' structural features and annotation data, examining physico-chemical properties such as amino acid composition, coefficient, instability, GRAVY, aliphatic index, theoretical isoelectric point (pI), atomic composition, and molecular weight. The assessment aimed to gain insights into protein thermal stability, where a high aliphatic index indicates stability, reflecting a notable presence of hydrophobic amino acids. The protparam expassy tool was utilized for predicting these physico-chemical properties of the analyzed proteins (7). The protein was found to be slightly unstable, with an instability index of 53.71. phylogenetic tree prediction provides a framework for understanding the evolutionary context of genes, aids in functional annotation, and contributes to various aspects of biological research, including drug discovery and the exploration of the genetic basis of diseases (20). Our study analysis unveiled the evolutionary relationships of *IKZF3* across different species, emphasizing its close association with great apes and Ma's nigh monkey. The conservation of this gene among species underscores its evolutionary importance, particularly in the context of the Mammalia family. Such insights into evolutionary trends provide a backdrop for further functional investigations and comparative genomics. Structural annotation involved a meticulous analysis of the *IKZF3* protein, a zinc-finger protein belonging to the Ikaros family crucial for lymphocyte development. The identification of functional domains, such as the C2H2 zinc finger, through tools like the DMDM and BioMuta databases, shed light on potential genetic polymorphisms associated with various cancer types. The analysis of disease mutations within functional domains highlighted the significance of the zinc finger domain, emphasizing its role in disease mechanisms and suggesting its potential as a therapeutic target.

The in-depth investigation into post-translational modifications, revealed crucial insights into the regulation of *IKZF3*'s stability and function. PTM is a crucial process occurring after a protein's synthesis and folding into its three-dimensional structure. It involves various chemical modifications like phosphorylation, acetylation, glycosylation, ubiquitination, and methylation. These modifications can impact a protein's activity, stability, localization, and interaction partners, ultimately influencing its biological function. PTMs play a vital role in cellular processes such as signal transduction, gene expression, and protein degradation. A comprehensive understanding of PTMs is essential for studying protein function and devising targeted therapies for a range of diseases (21). Physiological attributes such as polarity, % buried residues, average flexibility, bulkiness, hydrophilicity, and relative mutability were characterized, providing a holistic view of the protein's properties. The Protscale Server analysis further delineated the physico-chemical characteristics, offering a nuanced understanding of *IKZF3* at the molecular level. Prediction of methylation sites providing a systematic and computational approach to understanding the epigenetic regulation of genes. It contributes to our knowledge of gene function, disease mechanisms, and potential therapeutic targets (22). It highlighted the epigenetic regulation of IKZF3 on Chromosome 17, providing valuable information about its involvement in biological processes and potential associations with cancer. Subcellular localization predictions using the CELLO server indicated a nuclear localization score of 4.769, reinforcing the gene's role in transcriptional regulation and cellular processes. subcellular localization is a fundamental aspect of cell biology with implications for understanding cellular function, disease mechanisms, drug development, and biotechnological applications. It provides a spatial context that is essential for unraveling the complexity of cellular processes (23). The exploration of protein-protein interactions through the STRING Database revealed a complex network involving IKZF3 and multiple

proteins. The analysis of the protein's involvement in various biological processes, molecular functions, and cellular components, as indicated by Gene Ontology terms and KEGG pathways, underlined its diverse functional significance. Particularly, IKZF3's participation in pathways related to inflammatory bowel disease and asthma hints at its potential role in immune response regulation. Homology modeling provided three-dimensional structures for both wild-type and mutant IKZF3 proteins. Despite the introduction of amino acid modifications in the mutant, the structural analysis did not reveal significant alterations in the protein's conformation. This indicates a robust structural stability of IKZF3, essential for its functional integrity. The assessment of IKZF3's drugability profile, through structure-based analysis and pocket identification using the DoGSite scorer, identified potential binding pockets with high drug scores. The implications of these findings extend to therapeutic targeting, offering avenues for the development of precision medicine approaches in diseases associated with IKZF3 dysregulation. this study delves into the intricate landscape of the IKZF3 gene and its associated protein, providing a comprehensive resource for researchers and clinicians. The amalgamation of genomic, structural, functional, and regulatory insights equips the scientific community with a nuanced understanding of IKZF3's significance in cellular processes, its role in disease, and its therapeutic potential. The detailed exploration presented in this study not only contributes to the expanding field of *In-Silico* genomics but also lays the foundation for further experimental validations and clinical investigations. IKZF3 emerges as a pivotal player in the intricate orchestra of cellular regulation, and this study serves as a catalyst for future research endeavors aimed at unraveling its full biological repertoire.

CONCLUSION

The comprehensive genomic analysis of IKZF3 reveals its intricate structural dynamics and essential role in transcriptional regulation. The identified key functional implications open avenues for therapeutic exploration. This study enhances our understanding of IKZF3, shedding light on its potential as a therapeutic target for various diseases.

Acknowledgements:

We would like to express our sincere gratitude to KMU (Khyber Medical University) for their support and resources throughout the duration of this study. We also acknowledge the collaborative spirit and scholarly environment fostered by KMU, which has enriched our research experience and contributed to the advancement of scientific knowledge.

Conflict of Interest:

Authors have no conflict of interest.

References:

1. Hosokawa Y, Maeda Y, Takahashi E-i, Suzuki M, Seto M. Human aiolos, an ikaros-related zinc finger DNA binding protein: cDNA cloning, tissue expression pattern, and chromosomal mapping. Genomics. 1999;61(3):326-9.
2. Hong Y, Wang Q, Song Y-L, Liu G-M, Yang X, Yan M-X. Relationship between IKZF3 Gene Single Nucleotide Polymorphisms and Childhood Acute Lymphoblastic Leukemia. Zhongguo shi yan xue ye xue za zhi. 2021;29(3):690-5.
3. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. Journal of molecular biology. 1997;268(1):78-94.
4. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C. The Pfam protein families database. Nucleic acids research. 2012;40(D1):D290-D301.
5. El Yacoubi B, de Crécy-Lagard V. Integrative data-mining tools to link gene and function. Gene Function Analysis. 2014:43-66.
6. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N. Analysis tool web services from the EMBL-EBI. Nucleic acids research. 2013;41(W1):W597-W600.
7. Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PW, Kayastha AM. MFPPI–multi FASTA ProtParam interface. Bioinformation. 2016;12(2):74.

8. Emanuelsson O, Brunak S, Von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nature protocols. 2007;2(4):953.
9. He X, Chang S, Zhang J, Zhao Q, Xiang H, Kusonmano K, et al. MethyCancer: the database of human DNA methylation and cancer. Nucleic acids research. 2007;36(suppl_1):D836-D41.
10. Julenius K. NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. Glycobiology. 2007;17(8):868-76.
11. Gupta R, Jung E, Brunak S. Prediction of N-glycosylation sites in human proteins. 2004.
12. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KTB. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. The EMBO journal. 2013;32(10):1478-88.
13. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics. 2004;4(6):1633-49.
14. Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. Protein Engineering Design and Selection. 2004;17(1):107-12.
15. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic acids research. 2010;39(suppl_1):D561-D8.
16. Olsson L, Johansson B. Ikaros and leukaemia. British journal of haematology. 2015;169(4):479-91.
17. Li L, Ding X, Wang X, Yao Q, Shao X, An X, et al. Polymorphisms of IKZF3 gene and autoimmune thyroid diseases: associated with Graves' disease but not with Hashimoto's thyroiditis. Cellular Physiology and Biochemistry. 2018;45(5):1787-96.
18. Aggarwal K, Bansal V, Mahmood R, Kanagala SG, Jain R. Asthma and Cardiovascular Diseases: Uncovering Common Ground in Risk Factors and Pathogenesis. Cardiology in Review. 2023:10.1097.
19. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic acids research. 2000;28(1):45-8.
20. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ. A new view of the tree of life. Nature microbiology. 2016;1(5):1-6.
21. Perez-Vilar J, Randell SH, Boucher RC. C-Mannosylation of MUC5AC and MUC5B Cys subdomains. Glycobiology. 2004;14(4):325-37.
22. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C. NCBI GEO: archive for high-throughput functional genomic data. Nucleic acids research. 2009;37(suppl_1):D885-D90.
23. Jadot M, Boonen M, Thirion J, Wang N, Xing J, Zhao C. Accounting for protein subcellular localization: a compartmental map of the rat liver proteome. Molecular & cellular proteomics. 2017;16(2):194-212.