

Deciphering Zinc Finger Proteins' Role in Breast Cancer Progression: Homology-Based Analysis of Etiology and Prognosis.

Mihaela Tzvetkova

Received July 26, 2023

Accepted September 17, 2023

Electronic access October 31, 2023

Breast cancer (BC) is the most common cancer among women worldwide. Cancer is characterized by multi-faceted progression that mandates the continuous expression and activation of various transcription factors (TFs) responsible for cellular growth and sustenance. Previous research has revealed that numerous zinc finger proteins (ZFPs) are implicated in breast cancer. This study illuminates the intricate role of zinc finger proteins in BC progression and etiology, harnessing their critical function as transcription factors. The primary focus is on C2H2-type ZFPs, a significant subset of ZFPs, examining their unique characteristics and potential contribution to BC. To identify novel ZFPs implicated in BC, a dual approach is employed: applying homology-based prediction using the Basic Local Alignment Search Tool (BLAST), and another technique that scrutinizes both highly conserved and less-conserved regions. The proteins identified through homology are subjected to an exhaustive literature review to assess their role in BC and other cancers and are further evaluated for their prognostic significance using Kaplan-Meier survival plots derived from cBioPortal data. 27 out of 34 of the identified ZFPs were previously implicated in some cancer, and 7 portrayed a statistically significant Kaplan-Meier plot. The study culminates with an in-depth exploration of a protein (ZFP461) that demonstrated a statistically significant association with survival outcomes, utilizing resources from cBioPortal and the STRING platform. The findings not only reinforce the pivotal role of ZFPs in BC but also emphasize the value of homology-based prediction in uncovering BC's intricate etiology, thereby paving the way for future therapeutic advancements.

Introduction

Most instances of morbidity and mortality due to malignant tumors in women are caused by breast cancer (BC)¹. Despite substantial advances in diagnosis and treatment of BC, due to its complex and context-dependent nature, the molecular mechanisms underlying its development, progression, and resistance to therapies are still not fully understood². A vast array of genes are implicated in the initiation, progression, and metastasis of BC, many of which are controlled by specific regulatory proteins such as transcription factors³. Transcription factors possess domains through which they bind to DNA of promoter or enhancer regions of specific DNA sequences, thereby activating their transcription. They also have a region that communicates with RNA polymerase II or different transcription factors, which then influences the volume of messenger RNA (mRNA) generated by the gene. TFs can be categorized into various key groups, based on their structure and the manner in which they interact with DNA⁴. Zinc-finger proteins (ZFPs) constitute the most extensive group of transcription factors in the human genome, encoded by 2% of human genes⁵. ZFPs' key feature is a zinc finger domain, a compact structural motif wherein a zinc ion is coordinated by a combination of cysteine and histidine residues that stabilizes the structures of ZFPs, thereby enabling them to in-

teract with numerous molecules⁶. The variety in the structures and functions of zinc finger units allows these proteins to participate in a wide range of biological activities, such as growth, differentiation, metabolism, and autophagy. Over recent years, an increasing number of studies have highlighted the potential influence of ZFPs in the advancement of several cancers⁷. Distinctive molecular signatures, characterized by unique gene and microRNA expression profiles, are inherent to each type of cancer. Consequently, the regulatory roles of the same microRNA can diverge across different cancer types⁸. This extends to the regulation of ZFPs, whose expression profiles are under the control of these microRNAs, leading to significant variance in their expression among diverse cancer types. Furthermore, ZFPs are subject to post-translational modifications influenced by various environmental stimuli, which activate precise cellular signaling pathways⁹. Particularly in the case of BC, these findings underscore the probable paramount role of the tumor microenvironment and specific molecular subtypes in dictating ZFP functionality. Due to the context-dependent nature of ZFPs in influencing BC, more research should be pursued to further understand their roles and mechanisms of action. As of now, eight unique types of zinc finger domains have been documented, including the Cys2His2 (C2H2)-like, Gag knuckle, Treble clef, Zinc ribbon, Zn2/Cys6, TAZ2 domain-like, Zinc

binding loops, and Metallothionein. In the context of BC, the various subtypes of zinc finger domains have not been extensively explored, with the exception of the C2H2 and Metallothionein domains, which have been shown to act as prognostic biomarkers¹⁰. Notably, the C2H2-type ZFPs represent the most frequently investigated subtype in this context and therefore will be a primary focus of this study. C2H2-type ZFPs, the largest subset within the comprehensive ZFP family, exhibit defining features including tandem zinc finger motifs and additional functional domains such as the Krüppel-associated box (KRAB). These domains are pivotal in modulating subcellular localization, DNA-binding affinity, and gene expression by dictating selective interactions among transcription factors and other cellular constituents¹¹. Current research endeavors have highlighted the integral roles of these domains in BC pathogenesis. For instance, ZFPs containing a KRAB domain, mostly acting as transcription inhibitors, have been implicated as oncogenes and tumor suppressors within the context of BC. ZFPs containing a KRAB domain have also been indicated to foster the phenotype of cancer stem cells through interactions with the tripartite motif-containing 28 (TRIM28) protein (which is itself a ZFP) [11]. This interplay among C2H2-type ZNFs, their constituent domains, and their implications in the progression of BC underscores the potential utility of these proteins as novel foci for cancer research—not to mention that due to the vast nature of ZFPs, many have not been studied. This emphasizes the necessity for further exploration and identification of novel C2H2-type ZNFs in BC, potentially offering groundbreaking insights into disease etiology and therapeutic innovation. A way to identify novel ZFPs involved in BC is through homology interference, or homology-based prediction. This approach is grounded in the principle of sequence homology, the notion that similar DNA and protein sequences often yield similar functions¹². The credibility of this approach has been underscored by numerous empirical studies. For instance, Brachat et al. (2003) used sequence homology in the budding yeast *Saccharomyces cerevisiae* to predict the functions of proteins in the fungus *Ashbya gossypii*, and their findings were subsequently validated through experimental methods, demonstrating the high degree of accuracy attainable with homology-based prediction¹³. Further evidence of the viability of homology-based prediction was provided by Gough et al. (2001). In their research, they extended the approach to predict the functions of entire protein families within the genome of the bacterium *Escherichia coli*¹⁴. Remarkably, their predictions were accurate for 80% of the protein families under study. The highly conserved domains of ZFPs (ZFPs) is a direct consequence of their crucial function as transcription factors. Hence, similarities in the sequence of those areas might indicate similar functions in these areas across different ZFPs. This observation leads to the hypothesis that the application of homology

interference could provide increased precision and reliability in this context. Nevertheless, this proposition remains speculative and requires thorough exploration and substantiation. The pronounced conservation within ZFP domains presents a plausible premise for the potential utility of homology interference, yet definitive outcomes necessitate further empirical examination. These conservation properties suggest that a local alignment tool such as the Basic Local Alignment Search Tool (BLAST) could suffice in identifying homologous proteins, particularly in the context of their implications for BC. BLAST's functionality prioritizes the identification of homologous sequences based on the most conserved regions, making it well-suited to the high conservation characteristic of ZFPs' functional units¹⁵. Nevertheless, it should be acknowledged that less conserved regions, potentially overlooked by local alignment tools such as BLAST, may also bear significant relevance to BC. Hence, in addition to BLAST pairwise alignment, EMBOSS that scrutinizes both highly conserved and less conserved regions has also been utilized in this study. The comparison of data generated from these different alignment methodologies facilitates an enhanced comprehension of the applicability of various analytical techniques in the investigation of the role ZFPs play in BC. Intriguingly, despite the potential effectiveness of this dual analytical approach, its utilization remains limited within cancer research, specifically concerning ZFPs. By adopting this novel methodology, we aim to not only provide new insights into the function of ZFPs in BC but also enrich the methodological toolkit employed in the wider field. The resultant homologous proteins were subjected to a literature review, with a focus on previous articles that deliberated on the role of these proteins in BC or other malignancies. Further validation was achieved via Kaplan-Meier survival plots, which contrasted ZFP expression against survival outcomes for a substantial cohort of 8881 breast tumor samples, sourced from the cBioPortal database. One of the proteins that displayed a statistically significant correlation with patient survival was subjected to an in-depth analysis utilizing the resources available in the cBioPortal database and a robust pathway analysis conducted using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) platform. The primary objective of this study is to identify novel ZFPs involved in breast cancer through the utilization of homology interference. We seek to elucidate the mechanisms of action and potential therapeutic implications of these ZFPs, offering insights into disease etiology and therapeutic innovation. Although a significant amount of research has been conducted on ZFPs in various cancers, most commonly colorectal cancer, there is a noticeable scarcity of research papers and reviews specifically addressing the role of ZFPs in breast cancer. This gap underscores the urgent need to compile and synthesize all existing information on the role of ZFPs in breast cancer to enhance understanding and guide future research in

this area. The functional characterization of ZFPs in cancer biology is often a subject of debate, with disparate studies attributing conflicting roles to specific ZFPs—some identifying them as oncogenes that drive tumor progression, and others as tumor suppressors that inhibit oncogenesis. Such inconsistencies underscore the imperative for more research to delineate a more definitive role of ZFPs in tumorigenesis. Furthermore, there is a need to pinpoint the ZFPs that have the most substantial impact on BC metastasis and survival, as this could lead to the identification of novel and effective therapeutic strategies. Moreover, a limited number of studies have utilized biocomputational methods to specifically assess ZFPs in cancer. Most existing studies generally focus on oncoproteins as a whole, and include only a handful of ZFPs in the analysis. In contrast, this study is exclusively centered on ZFPs, providing a unique platform for comparison between each other and a compelling rationale for their significance. By exclusively focusing on ZFPs and employing a combination of homology-based function prediction and biocomputational methods including survival analysis, this study aims to overcome the listed limitations of past research. The results obtained from this approach lent further credence to the putative role of this protein in BC, thereby reinforcing the validity of employing homology-based function prediction for ZFPs in cancer research.

Results

Identification of known ZFPs

Eight ZFPs with a well-characterized role in BC were identified using the literature review criteria established above. For oncogenic ZFPs, which include ZFX, ZEB1, ZNF711, and ZNF367, the predominant roles are in promoting metastasis, enhancing cell growth, and conferring resistance mechanisms like chemotherapy resistance. Their interaction pathways often involve essential BC-associated genes such as NANOG, TGF- β , and VEGFA. Intriguingly, most of these ZFPs interact with pathways pivotal for cell growth, such as Wnt/B-catenin and JAK/STAT, emphasizing their critical roles in disease progression. On the other hand, tumor suppressor ZFPs, including ZNF385B, ZNF24, ZNF668, and ZFP82 (or ZNF545), primarily play roles in inhibiting tumor growth processes like angiogenesis, promoting DNA repair, and inducing cell cycle arrest. Their interactions tend to revolve around prominent tumor-suppressing agents such as p53 and VEGF. It's notable that the downregulation or suppression of these ZFPs correlates with poor prognosis and aggressive BC subtypes. Table 1 below summarizes the roles, principal interactions, and potential therapeutic implications of these ZFPs:

Identification of homologous ZFPs The bioinformatics

analysis identified a total of 34 homologs for four of the known ZFPs (ZFPs) - ZEB1 (9 homologs), ZNF24 (3 homologs), ZNF711 (10 homologs), and ZFP82 (12 homologs). ZFX, ZNF668, ZNF385B, and ZNF367 did not have homologs passing the predefined criteria. All of the parameters associated with the ZFPs are shown in Table 2 and discussed below. Examining the ZEB1 homologs, ZEB2 displays the highest query coverage of 92% with an E-value of 0.0 and a percent identity of 44.48% from BLAST. This is further corroborated by EMBOSS analysis where ZEB2 shows a sequence identity of 42.6%, a similarity of 56.6%, reflecting the highest alignment score of 2270.5. Despite the sequence identity being moderate, the significant overlap in protein sequences coupled with a low E-value and supporting EMBOSS data affirms a strong homology between ZEB1 and ZEB2, likely a result of shared roles in biological processes such as the epithelial-mesenchymal transition. Other ZEB1 homologs including ZNF552, ZNF232, ZNF461, ZKSCAN1, ZNF135, ZNF529, ZFP1, and ZNF275 show lower query coverage (17-28%). Nonetheless, the significance of the alignments, indicated by E-values ranging from 1.0E-17 to 7.0E-20, and the sequence identities (49.33% to 54.67%) imply probable homology and functionally or structurally important shared segments between ZEB1 and these proteins. This is supported by the EMBOSS analysis, which despite showing considerably lower sequence identity, ranging from 6.6% (ZNF275) to 15.5% (ZNF135), also suggests non-trivial similarity percentages, suggesting that they may conserve residues that have similar physicochemical properties. One noteworthy finding is ZNF461, which shows a sequence identity of 38.6% with a lower similarity of 31.8% in the EMBOSS analysis, indicating that a portion of identical residues might not perform similar functions or occupy similar structural roles. This highlights a potential complexity in the evolutionary relationship between ZEB1 and ZNF461. The homolog cluster of ZNF24, as revealed through BLAST analysis, primarily includes ZNF396, ZBED9, and ZSCAN12. ZNF396 is especially significant, demonstrating the highest query coverage of 96% and a remarkably low E-value of 8.0E-59. This suggests a substantial sequence overlap and alignment with ZNF24. Coupled with a percent identity of 56.77%, these findings indicate that ZNF396 shares more than half of the same amino acid residues at the aligned positions with ZNF24, suggesting a close evolutionary relationship. Complementing this, ZBED9 and ZSCAN12, although showing lower query coverages of 89% and 74% respectively, still present with low E-values, indicative of their significant homology with ZNF24. ZSCAN12, despite its lower query coverage, interestingly exhibits a higher percent identity of 58.04% compared to ZBED9's 53.18%, hinting at more pronounced sequence conservation with ZNF24. For ZBED9, the EMBOSS results reveal a contrasting narrative with a much

ZFPs	Role in BC	Promotes/Suppresses Cancer	Notable Interactions and Pathways	Potential as Therapeutic Target
ZFX	Enhances cell growth and metastasis; confers self-renewal capabilities and resistance to chemotherapy ^{16, 17, 18, 19, 20}	Promotes	Activates NANOG gene, thereby interacting with TGIF-B, Wnt/B-catenin JAK/STAT ¹⁶	Inhibition shown to suppress cancer progression.
ZEB1	Facilitates Epithelial-Mesenchymal Transition (EMT) thereby promoting metastasis and angiogenesis ^{21, 22, 23, 24, 25}	Promotes	Responds to signals like TGF- β , represses E-cadherin, activates VEGFA expression	Silencing restores sensitivity to EGFR-TKI
ZNF711	Regulates gene expression, associated with BC staging and prognosis ^{26, 27, 28, 29}	Promotes	Interacts with PHF8, KDM5C, correlated with ER and HER-2 expression	Biomarker for disease progression and targeted therapy
ZNF367	Promotes metastasis ^{30, 31, 32}	Promotes	Interacts with BRG1, activates CIT and TP53BP2, inhibits Hippo pathway, activates YAP1	Inhibition of YAP1 reduces metastasis ^{33, 34, 35, 36}
ZNF24	Suppresses VEGF expression, inhibits angiogenesis ^{37, 38, 39}	Suppresses	Binds to VEGF promoter region tumor growth	Therapies that enhance ZNF24's function could inhibit
ZNF668	Enhances p53's tumor-suppressing capabilities promotes DNA repair ^{40, 41, 42}	Suppresses	Interacts with MDM2, p53, Tip60	Therapies that enhance ZNF668's function could enhance tumor suppression and DNA repair
ZFP82 or ZNF545	Halts cell proliferation, induces cell cycle arrest, triggers apoptosis ^{43, 44, 45}	Suppresses	Upregulates c-Jun/AP1, BAX, p53, Caspase 3 proliferation and induce apoptosis	Restoring ZNF545 expression could halt cell in BC cells

Table 1: A table with custom font style.

lower sequence identity of 7.0%, despite the BLAST results suggesting higher sequence conservation. This discrepancy suggests that while ZBED9 may exhibit overall homology with ZNF24, the precise sequence conservation is potentially concentrated in specific areas. These regions might hint at functionally critical areas within the protein structure, such as binding sites, active sites, or other functional domains that have been preserved across evolutionary lines due to their significance in the protein's function. Such evolutionary conservation, therefore, may implicate ZBED9 in analogous cellular pathways or mechanisms to those ZNF24 operates in. As such, it invites speculation about a potential role for ZBED9 in the complex disease networks of BC. Similarly, the EMBOSS data further reinforce the homology between ZNF24 and ZSCAN12. Despite the lower BLAST query coverage, ZSCAN12 demonstrates a higher sequence identity of 14.8% and similarity of 20.7% in the EMBOSS analysis. This correlation of findings suggests that ZSCAN12, much like ZNF396, could share important structural or functional features with ZNF24. The homology results for ZNF711 display lower query coverage values compared to those of ZNF24, ranging from 45% to 53%. The E-values, however, are still low, indicating significant alignments. The percent identities range from 33.07% to 35.48%, which are lower than those for ZNF24, suggesting less sequence conservation. Nonetheless, the percent identity still indicates a measure of homology, suggesting that these proteins share a portion of the same amino acid residues with ZNF711 at the aligned positions. Building upon this, the EMBOSS analysis offers a nuanced understanding of the homologous relationships. Notably, ZNF429, ZNF43, ZNF431, ZNF66, ZNF708, ZNF737, ZNF85, ZNF93, ZNF98, and ZNF99 all demonstrate identity percentages between 17.4% and 20.3%, which, while lower than the BLAST percent identities, still show a degree of homology. Among these, ZNF43, with an identity of 20.3%, shows the highest sequence identity, suggesting a homologous relationship with ZNF711. Despite the lower sequence identities compared to the BLAST results, these proteins might still share functionally important residues or regions with ZNF711. Of particular interest is ZNF91, which, despite having a relatively low identity percentage of 10.6% and a high gap percentage of 68.3%, still achieves a comparable score of 582, indicating a specific, yet scattered, conservation of sequences. This hints at potential hotspots of functional importance within the protein structure. Furthermore, the similarity percentages, ranging from 27.2% to 30.9%, highlight the possibility of amino acid substitutions that still maintain similar properties and functionalities, which is another significant facet of understanding protein evolution. The BLAST analysis of ZFP82 homology unveils an exceptional degree of sequence overlap with other proteins, most with a query coverage of 99% and even attaining a full 100% in the case

of ZNF546. Virtually null E-values across the board point to highly significant alignments, signifying strong homologous relationships. ZNF383 is a slight outlier with an E-value of 2.0E-179, which, despite being higher than the rest, is still a highly significant alignment. The impressive percent identities evident in the BLAST results speak volumes about the shared evolutionary paths of these proteins. ZFP30 and ZFP14 command attention with their highest values at 75.42% and 69.92%, respectively, reflecting an extensive sharing of identical amino acid residues with ZFP82. The EMBOSS results bolster these insights. Both ZFP30 and ZFP14 maintain their top ranks with identities at 75.1% and 69.7%, respectively, further validating their close ties with ZFP82. The low percentage of gaps, particularly for ZFP30, indicates a high level of sequence structure conservation, providing more weight to their close evolutionary relationships. Other proteins, including ZNF420, ZNF780A, ZNF283, ZNF546, ZNF850, ZNF565, ZNF540, ZNF571, ZNF461, and ZNF383, display percent identities ranging from 51.29% to 56.79% in BLAST results, indicative of considerable homology. The EMBOSS results, while showing slightly lower identities, consistently reflect this trend of homology. Standouts like ZNF461 and ZNF565 showcase identities over 50% and low gap percentages in the EMBOSS results, suggesting they share significant sequence conservation with ZFP 82. Such high percent identity values suggest a close evolutionary relationship, and the aligned sequences share a majority of the same amino acid residues. To summarize, both ZFP82 and ZNF24, which act as suppressors, exhibit the highest level of sequence conservation with their identified homologs. In contrast, ZEB1 and ZNF711, which are promoters of cancer, display less sequence conservation. However, their homologs still indicate a substantial degree of homology, underscoring the intricate and complex nature of protein evolution and functional diversity. These findings provide a robust starting point for further experimental investigation into the shared and unique roles of these ZFPs.

Expression of homologous ZFPs in BC:

Upon completing the review of the ZFP homologs and their implications in cancer, certain patterns have emerged. A substantial 79% of these ZFP homologs have been implicated in different types of cancer, which underscores their significant role in cancer biology and the potential impact of those whose roles are yet to be deciphered in BC. Guided by the principle of sequence homology, we hypothesized that if a certain ZFP plays a suppressive or promoting role in BC, its homologous counterparts would exhibit similar properties. Of the twelve homologs examined for the tumor suppressor ZFP82, fifty percent were implicated in various cancers, with two specif-

Zinc-Finger Protein	Name of protein	Query cover	Blast	Per. Identity %	length	identity	EMBOSS similarity	gaps	score
			E value	%					
ZEB1	ZEB2	92%	0	44.48%	1246	42.60%	56.60%	15.60%	2270.5
	ZNF552	21%	7.00E-20	50.67%	1209	8.20%	12.80%	74.70%	242
	ZNF232	18%	2.00E-18	53.33%	1175	10.10%	15.70%	68.70%	248.5
	ZNF461	27%	3.00E-18	54.67%	1296	38.60%	31.80%	25.80%	202.5
	ZKSCAN1	28%	1.00E-17 5	4.67%	1222	10.40%	17.80%	66.20%	227
	ZNF135	17%	1.00E-17	53.33%	1238	15.50%	22.90%	55.40%	301
	ZNF529	17%	2.00E-17	49.33%	1233	12.50%	19.50%	64.50%	197
ZNF24	ZNF396 or ZNF367	96%	8.00E-59	56.77%	380	12.90%	18.90%	57.10%	31
	ZBED9	89%	5.00E-48	53.18%	1337	7.00%	9.60%	86.40%	440
	ZSCAN12	74%	3.00E-46	58.04%	622	14.80%	20.70%	70.70%	429.5
ZNF711	ZNF91	53%	7.00E-49	34.57%	1493	10.60%	16.50%	68.30%	582
	ZNF98	46%	5.00E-48	34.76%	895	17.80%	27.20%	45.90%	569
	ZNF85	46%	5.00E-47	33.07%	902	19.20%	30.50%	41.20%	588
	ZNF99	46%	5.00E-47	33.07%	1046	19.30%	30.30%	40.20%	574.5
	ZNF43	46%	5.00E-47	33.07%	989	20.30%	30.90%	37.20%	602.5
	ZNF429	49%	3.00E-46	33.07%	966	17.40%	28.90%	46.70%	582.5
	ZNF737	46%	4.00E-46	35.48%	862	19.50%	30.40%	44.20%	578
	ZNF431	45%	4.00E-46	33.24%	869	19.40%	29.70%	40.70%	569.5
	ZNF66	46%	1.00E-45	34.93%	879	19.20%	29.80%	43.00%	580
ZFP82	ZNF93	46%	1.00E-45	34.32%	898	19.20%	30.50%	41.10%	583
	ZFP30	99%	0	75.42%	535	75.10%	84.50%	3.40%	2218
	ZFP14	99%	0	69.92%	535	69.70%	79.80%	30.60%	2042
	ZNF420	99%	0	56.79%	736	41.00%	49.70%	31.90%	1619
	ZNF780A	99%	0	75 1.55%	649	47.30%	59.20%	19.10%	1597.5
	ZNF283	99%	0	56.12%	643	39.30%	47.70%	33.10%	1376
	ZNF546	100%	0	55.25%	811	38.80%	46.60%	34.40%	1587.5
	ZNF850	98%	0	51.29%	1093	28.50%	34.90%	51.50%	1570.5
	ZNF565	99%	0	55.75%	642	52.60%	65.50%	6.70%	1533
	ZNF540	99%	0	52.47%	662	48.50%	59.20%	19.80%	1640
	ZNF571	98%	0	51.55%	642	45.80%	55.90%	22.10%	1533
	ZNF461	99%	0	54.24%	551	52.80%	64.40%	5.30%	1468.5
ZNF383	99%	2.00E-179	56.72%	534	52.40%	61.60%	11.20%	1417.5	

Table 2: Results of homology interference from BLASTp and EMBOSS, ranked by E value.

ically linked to BC. ZNF565 and ZNF540 emerged as having a suppressive role congruous to ZFP82 (Table 3). Meanwhile, ZFP14, ZFP30, ZNF283, ZNF546, ZNF461, ZNF780A, and ZNF383 showed divergent roles, including promoting metastasis, elevating cancer risk through missense variants, and involvement in chromosomal rearrangements and differential expression in ovarian and lung cancers (Table 3). For ZNF420, ZNF850, and ZNF571, their roles in cancer remain unclarified, underscoring the need for further research. Observations for the tumor promoter ZEB1 and its homologs are similar. Out of ten homologs, eight showed relevance in cancer-related research, with ZEB2, ZKSCAN1, ZNF232, ZNF461, ZNF529, ZFP1, and ZNF275 paralleling ZEB1's tumor-promoting roles. However, ZNF135, ZNF382, and ZNF552 displayed contrasting roles, manifesting as suppressive tendencies, varying cancer stage impacts, altered carcinoma expression, and drug resistance (Table 3). As with the previous group, the roles of ZFP1 and ZNF275 are not yet defined and warrant further research. ZNF24, another tumor suppressor, has three homologs—ZNF367, SCAND3/ZBED9, and ZSCAN12—that are all linked with varying cancer roles or diseases. Specifically, ZNF367 is associated with breast cancer progression and displays a multifaceted influence on cancer biology [64]; SCAND3/ZBED9 acts as a methylation biomarker in hepatocellular carcinoma, especially valuable for AFP-negative cases [65]; and ZSCAN12 is indicative of tumor-infiltrating lymphocyte densities in gastric carcinoma [66] (Table 3). Finally, ZNF711, a recognized tumor promoter, had eight of its eleven homologs implicated in various cancers, with ZNF91, ZNF98, ZNF99, and ZNF93 exhibiting cancer-promoting roles akin to ZNF711 (Table 3). Conversely, ZNF43, ZNF431, and ZNF66 were associated with earlier cancer stages, drug resistance, and reduced mRNA expression in head and neck cancers, respectively (Table 3). The roles of ZNF85, ZNF429, ZNF737, and ZNF708 remain uncertain, highlighting the need for further investigation in these areas.

Among the 34 homologous ZFPs analyzed, seven exhibited statistically significant survival plots, elucidating their potential roles in breast cancer (BC). Notably, six out of these seven proteins, specifically ZEB2, ZNF420, ZNF46, ZNF529, and ZNF850, were associated with unfavorable survival outcomes, as manifested by HRs greater than 1. An HR greater than 1 signifies an elevated risk of an adverse event, such as mortality. This relationship accentuates their correlation with adverse patient prognosis in other malignancies, as delineated in the literature review (Table 4), indicating that augmented protein expression is linked with deteriorated patient prognosis. Conversely, ZKSCAN1, with an HR value less than one, denoted a diminished risk of adverse outcomes, thereby exhibiting a protective effect. This finding is somewhat unexpected, given the established role of ZKSCAN1 as a facilitator of lung

cancer, insinuating a multifaceted function in diverse tumor milieus. ZNF420 and ZNF850 lacked pre-existing literature that could furnish insights into their potential repercussions in cancer, underlining the necessity for additional investigation. Among the other ZFPs, solely ZEB2 has been formerly associated with BC in scientific literature, whereas ZNF461 and ZNF529 have been examined in the context of non-small cell lung cancer and hepatocellular carcinoma, respectively. This disclosure underscores the imperative for more exhaustive research into the functions of other ZFPs in the BC milieu. Except for ZNF529, these proteins display some of the most conspicuous results in homology interference (as indicated in Table 2). It is crucial to recognize that the absence of substantial representation in Kaplan-Meier plots does not necessarily signify that the other identified homologous proteins lack relevance in BC. Contrarily, several experimentally validated homologous ZFPs known to play roles in BC were not depicted, reflecting the variability and complexity of their potential contributions. This phase of the research was primarily conducted as a means of highlighting the most significant findings of the study, not to disregard the contributions of other potentially relevant proteins.

Further analysis:

ZNF461 was further investigated for a role in BC due to the lack of prior research in its involvement in BC and its expression being significantly inversely proportional with survival rate in BC (Tables 3 and 4). Based on the protein-protein pathway generated by the STRING database, ZNF461 is predicted to interact with numerous proteins that have been shown to be involved in BC. In the STRING database every interaction is rated by a confidence score from 0.1-1, representing how strong the evidence that supports the interaction is (1 being the highest). The proteins with the highest confidence scores were TOX2, HDAC2, and PHGDH exhibiting confidence scores in the range of 0.73-0.93.

TOX2, also known as Thymocyte selection-associated HMG box gene 2, is a transcription factor that belongs to the TOX subfamily of high mobility group box (HMG-box) superfamily which also includes TOX, TOX3, and TOX4⁸³. Abnormal expression of TOX2 has been associated with various types of cancer. In BCs, it was found that TOX2 was hypermethylated in the promoter region, leading to its silencing. This hypermethylation was found in 23% of breast tumors studied. When TOX2 was silenced through methylation, expression of the two TOX2 transcripts was significantly reduced. Interestingly, the knockdown of TOX2 impacted multiple pathways, suggesting a potential role for TOX2 in modulating the tumor microenvironment⁸⁴. Additionally, a recent study on natural killer/T-cell lymphoma (NKTL) found that increased TOX2 levels led to the growth and spread of

Table 3 Literature review on the identified homologous ZFPs and their known links to cancers and other selected indications⁴⁶⁻⁸².

Zinc-Finger Protein	Homologous ZFPs	Role in BC/Cancer
ZEB1	ZEB2	found to be upregulated in BC, and to contribute to BC progression by suppressing E-cadherin expression [46-48]
	ZNF552	elevated expression was associated with improved disease-free survival in BC [49]
	ZNF232	differentially expressed in ovarian cancer and BC [50]
	ZNF461	involved in immune response for promoting non-small cell lung cancer progression [51]
	ZKSCAN1	indirectly promotes non-small cell lung cancer and lung adenocarcinoma progression [52] [53]
	ZNF135	highly methylated in breast and colon cancer [54]
	ZNF529	ZNF529-AS1, a long non-coding RNA highly expressed in hepatocellular carcinoma (HCC), correlates with tumor stage, pathological grade, and poor prognosis, and its knockdown inhibits HCC cell invasion and migration [55]
	ZFP1	no information found
	ZNF275	no information found
ZNF24	ZNF367	promotes BC progression, cell proliferation, invasion, and adhesion; its downregulation, along with upregulation of miR-21-5p, inhibits BC migration and invasion, demonstrating its complex role in cancer biology [56]
	SCAND3/ZBED9	serves as a novel methylation biomarker in hepatocellular carcinoma, enhancing early detection, especially in alpha-fetoprotein (AFP)-negative cases, and acting as a predictive indicator for portal vein tumor thrombus (PVTT) [57]
	ZSCAN12	signifies densities of tumor-infiltrating lymphocytes in gastric carcinoma [58]
ZNF711	ZNF91	related to the occurrence and development of bladder cancer [59-60], colorectal cancer (CRC) [61], ovarian cancer [62-63], and non-small cell lung cancer [64]
	ZNF98	promotes BC and lung cancer [65]
	ZNF85	no information found
	ZNF99	kidney cancer patients with somatic mutation in ZNF99, resulting in gain of function, had lower survival rate [66]
	ZNF43	ZNF43-methylated normal tissue is associated with earlier colorectal cancer stages [67]
	ZNF429	the gene with the highest somatic mutation frequency in thymic carcinoma [68]
	ZNF737	no information found
	ZNF431	role in resistance to Imatinib Mesylate in gastrointestinal stromal tumor [69]
	ZNF66	low mRNA expression in head and neck cancers [70]
	ZNF93	when highly expressed, promotes proliferation and migration of ovarian cancer cells and relates to poor prognosis [71-72]
	ZNF708	no information found
ZFP82	ZFP30	leads to BC metastasis via KAP1-mediated activation [73]
	ZFP14	deletion is protective against prostate cancer; somatic copy gains are linked to poor prognosis in BC [74]
	ZNF420	no information found
	ZNF780A	contributed to the formation of a novel RUNX1::ZNF780A chimera, resulting from chromosomal rearrangements associated with leukemia [75]
	ZNF283	contributed largely to the excess of missense variants, thereby increasing low-penetrance BC risk [76]
	ZNF546	elevated expression in ovarian cancer [77-78]
	ZNF850	no information found
	ZNF565	suppression induced BC cell death upon PI3K inhibition [79]
	ZNF540	tumor suppressor in lung cancer, and high levels of expression led to better survival rates in cervical cancer patients [80-81]
	ZNF571	no information found
	ZNF461	involved in immune response for promoting non-small cell lung cancer progression [51, 82]
ZNF383	promotes lung cancer by upregulating the expression of ERCC1 and HDAC7 [83]	

NKTL and the overproduction of PRL-3, an oncogenic phosphatase known for its role in cancer survival and metastasis. The study also found that TOX2 overexpression was nega-

tively associated with patient survival in NKTL⁸⁵. HDAC2 is significantly upregulated in cancer specimens and BC cell lines compared to non-cancerous tissues and normal cell lines.

Homologous Protein	Hazard Ratio	p-value	q-value
ZEB2	1.99	0.050	0.320
ZFP14	1.55	0.021	0.049
ZSCAN1	0.58	0.042	0.200
ZNF420	1.45	0.054	0.119
ZNF461	1.67	0.006	0.008
ZNF529	1.63	0.008	0.020
ZNF850	1.56	0.021	0.050

Table 3: Significance of Kaplan-Meier survival plots for the homologous proteins yielding a p value less than or equal to 0.05

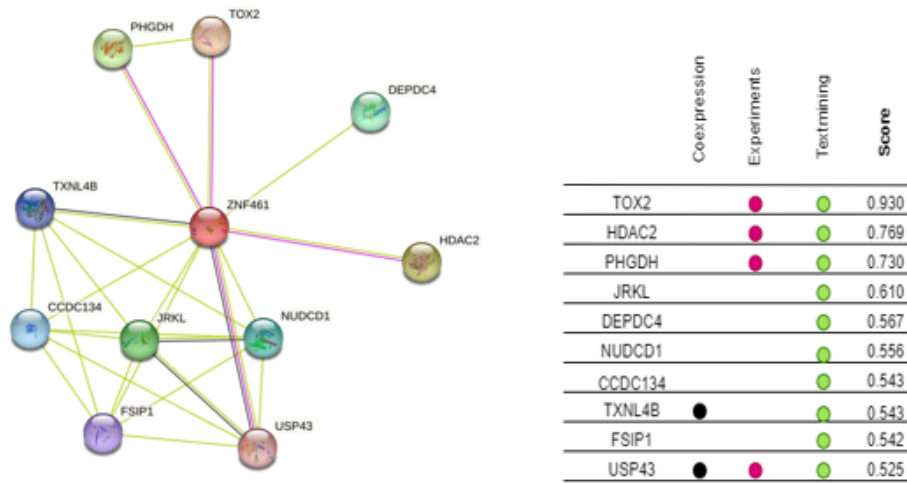


Fig. 1: Protein interactions of ZNF461 in Homo sapiens. (a) Predicted protein interactions using the STRING database (b) Information describing the specific type of protein interaction and the prediction score.

Conversely, the expression of miR-646, a microRNA that regulates gene expression, was decreased in clinical BC specimens and cells. The study concluded that miR-646 can potentially suppress the progression and proliferation of BC cells by inhibiting HDAC2⁸⁶. In another study, HDAC2 and another enzyme, HDAC3, were found to be significantly overexpressed in less differentiated breast tumors and associated with negative hormone receptor status. High expression of HDAC2 was significantly correlated with the overexpression of the HER2 protein, which can promote the growth of cancer cells, and the presence of nodal metastasis, which indicates the spread of cancer to lymph nodes. The study concluded that HDAC2 and HDAC3 are strongly expressed in more aggressive types of tumors⁸⁷. Phosphoglycerate dehydrogenase (PHGDH) diverts glycolytic intermediates into the serine biosynthesis pathway and is known to be overexpressed in a subset of BCs⁸⁸. Indeed, PHGDH exists in a region of chromosome 1p commonly amplified in BC. In total, 18% of patient-derived BC cell lines and 6% of primary tumors have

amplifications in PHGDH. study identified 39 metabolic enzymes essential for the survival of BC cells. Among these, the serine synthesis pathway was particularly significant, and the first enzyme in the pathway, PHGDH, was overexpressed in BC. The study demonstrated that inhibition of PHGDH impeded BC cell proliferation⁸⁹. The other proteins identified had a lower confidence score of 0.61- 0.52. Nevertheless, several have also been shown to have a role in BC. JRKL, alongside FA2H, corresponds to altered enhancer-promoter interactions (EPIs) and might be potential targets for acquired resistance to doxorubicin, a common chemotherapy drug used to treat BC^{90, 91}. NUDCD1, or NudC domain containing 1, has been found to have a significant role in triple-negative BC (TNBC), which is the most malignant subtype of BC, and to be successfully inhibited by HDAC7⁹². It has been found that BC cells and tissues consistently demonstrated elevated FSIP1 expressions, which correlated with poor overall survival^{93, 94}. USP43 also facilitates BC cell proliferation and metastasis⁹⁵.

The genomic profile of the ZNF461 gene was then eval-

uated through an Oncoprint analysis (Figure 4). Most studies reported an amplification of ZNF461, suggesting that it is a commonly occurring event in tumors. A smaller proportion of studies reported deletions of this gene, indicating that deletions, although less frequent, are still a significant event. Moreover, 2.6% of the studies reported mutations in the ZNF461 gene. These mutations included truncating mutations, missense mutations, and structural variant mutations, all of which have been reported across the studies analyzed. Truncating mutations lead to a shortened protein product, whereas missense mutations result in a single amino acid change in the protein product. Structural variant mutations involve larger changes in the DNA structure, such as deletions, duplications, inversions, or translocations. Gene amplification, as observed for ZNF461, signifies an augmentation in the copy number of the gene encoding this particular protein within neoplastic cells, compared to their non-malignant counterparts. However, it should be noted that this is not the sole determinant of gene activity. Gene expression levels and subsequent protein production could also be modulated by the interplay of various transcription factors and epigenetic modifications that regulate the accessibility and activation of the gene. Amplification events can potentially trigger a surge in protein expression levels. In the event that this protein functions as a growth-promoting factor, it could facilitate oncogenesis and further promote tumor progression. Such genetic behavior might elucidate the observed correlation between ZNF461 amplification and unfavorable prognostic outcomes in BC.

Due to the large and variable nature of the dataset encompassing 8881 samples, alteration frequency across the different BC subtypes was not directly available. Therefore, it became necessary to gather such data from the dataset that demonstrated the highest levels of ZNF461 alterations, specifically The Metastatic BC Project, consisting of 301 patients and 379 samples. Our analysis revealed that the expression of the ZNF461 protein was most highly amplified in distal and lobular carcinoma as well as invasive lobular carcinoma (Figure 5). Interestingly, ZNF461 expression did not show significant amplification when tumors presented a more invasive disease phenotype. Instead, mutations in the gene were noted, thereby hinting at its potential role in metastasis or cancer progression. Such a role would be reflective of the ZFP having a similar role as its homolog, ZEB1, which is known for its role in metastasis. Additionally, it would make sense when the proteins that ZNF461 was predicted to interact with are taken into consideration (Figure 3).

Further analysis of the mutations associated with ZNF461 suggested that they are most frequently encountered at the C2H2 domain (300-500), which is its most highly-conserved domain and which is associated with DNA binding (Figure 6). These mutations may disrupt its DNA binding capacity, which

could alter downstream gene regulation and potentially impact the protein's functional role. Given the role of ZNF461 as a transcription factor interacting with identified oncogenes, mutations within this domain could result in the dysregulation of gene expression. Such anomalies are consistent with the hallmarks of cancer, implying that the specific site of mutations within the ZNF461 protein may play a significant role in BC pathogenesis and progression.

The Metastatic BC Project dataset was subsequently searched to discern potential co-expression between ZNF461 and the predicted protein interactants which have demonstrated associations with BC. Significant co-expression was observed for proteins JRKL, DEPDC4, and NUDCD1 (Figure 7). However, it's worth noting that while the p-values indicate statistical significance, the correlation coefficients, particularly for NUDCD1, were relatively weak. This observation suggests an enhanced impetus for further investigation into the roles and potential interactions of these particular proteins. The evident co-expression between ZNF461 and these proteins might indicate the probability of these interactions occurring within the BC context, but the strength of these relationships needs more scrutiny. Given the known involvement of JRKL, DEPDC4, and NUDCD1 in BC, the detected co-expression, albeit weak for some, further bolsters the possibility of ZNF461 functioning as an integral component in BC pathology. The reason why other interacting proteins were not co-expressed could be that ZNF461 impedes their transcription rather than promotes it. Nevertheless, it should be noted that the sample size for which the analysis was done was small (n=167), making it harder to provide both statistical significance and strong correlation.

Methods

Selection of zinc-finger proteins involved in BC:

Specific zinc-finger proteins associated with BC were systematically identified using PubMed, Web of Science, and Embase databases, employing "zinc-finger proteins" and "breast cancer" as key search terms. These databases were selected for their extensive coverage of biomedical literature, reputation for quality, and inclusivity of research across various disciplines relevant to zinc-finger proteins and breast cancer. Review articles were manually searched for references to original research articles on the topic. To identify specific research articles discussing the role of individual ZFPs in BC, advanced search was performed via "ZFP name" (e.g., "ZEB1"), boolean operator "and", and "BC". Research articles that failed to fulfill the inclusion criteria shown in Figure 1 were subsequently excluded. Zinc-finger proteins that demonstrated an association with BC in more than three unique research articles that fulfilled the inclusion criteria were recorded in Table 1. While this strategy may

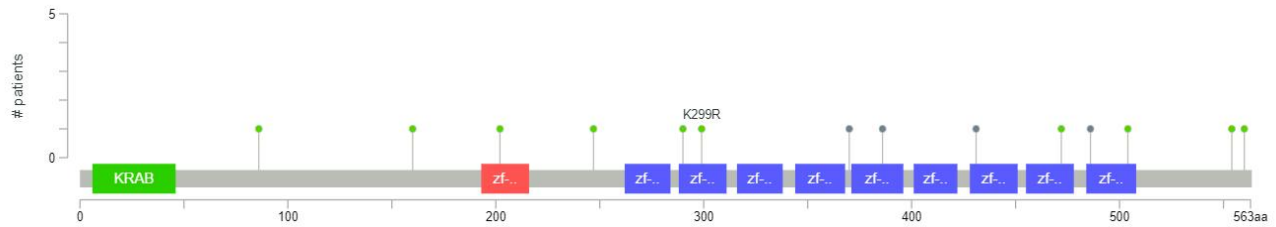


Fig. 4: Diagram showing distribution of disease-linked mutations across ZNF461. Green circles indicate missense mutations while gray circles indicate truncating mutations.

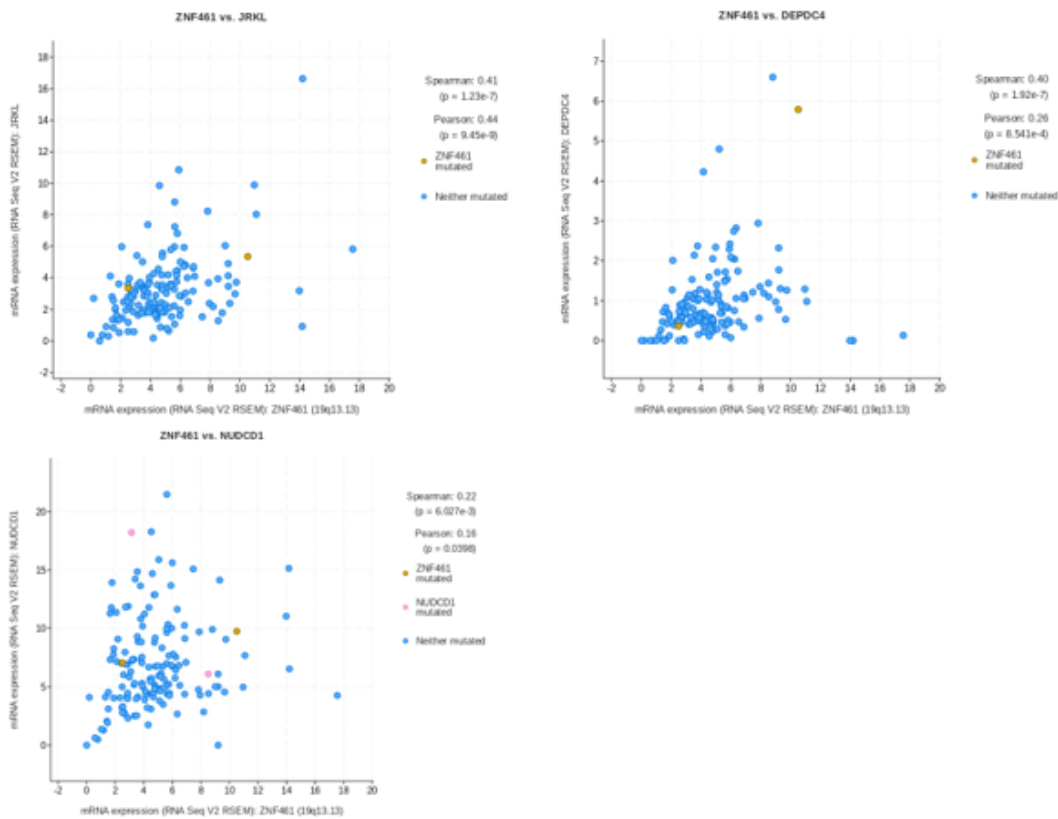


Fig. 5: Correlation of ZNF461 mRNA Expression with mRNA Levels of JRKL, NUDCD1, and DEPC40, and Mutation Status Across Samples

Identifying homologous ZFPs:

The amino acid sequence of the identified zinc-finger proteins was obtained from the National Center for Biotechnology Information (NCBI) in the FASTA format. A sequence similarity search was conducted using the Basic Local Alignment Search Tool (BLAST) for proteins (BLASTp). Protein:protein alignments were used for the search as amino acid

sequences are more conserved due to selective pressure, meaning changes that negatively affect the protein's function may be selected against during evolution. In contrast, DNA alignments can be more variable due to codon degeneracy, and their statistics are often less accurate due to the larger size of DNA databases⁹⁶. In BLASTp, the non-redundant database was selected for the search, which was confined to Homo sapiens to yield the most relevant results. The criteria for evaluating ho-

Inclusion criteria	Exclusion criteria
<ul style="list-style-type: none"> • Study Types: Peer-reviewed original research articles, including observational studies (e.g., cohort, case-control, cross-sectional studies), randomized controlled trials, and experimental studies. • Subjects: Studies on human subjects, in vivo or in vitro (cell lines), with breast cancer. • Intervention/Exposure: Studies that focus on the role of C2H2 Zinc Finger proteins in breast cancer development, progression, or treatment (expression, molecular interactions, or patient outcomes associated with C2H2 Zinc Finger proteins) • Publication Language: Studies published in English. • Publication Date: Studies published in the last 10 years. 	<ul style="list-style-type: none"> • Study Types: Review articles, commentaries, editorials, letters, case reports, and conference abstracts. • Intervention/Exposure: Studies that mention C2H2 Zinc Finger proteins but do not investigate their role in breast cancer. • Publication Language: Studies not available in English (unless translation resources are available). • Publication Date: Studies published more than 10 years ago. • Data Availability: Studies with insufficient data (e.g., missing key outcomes, sample size, etc.)."

Fig. 6: Inclusion and exclusion criteria for research articles on the role of ZFPs in BC

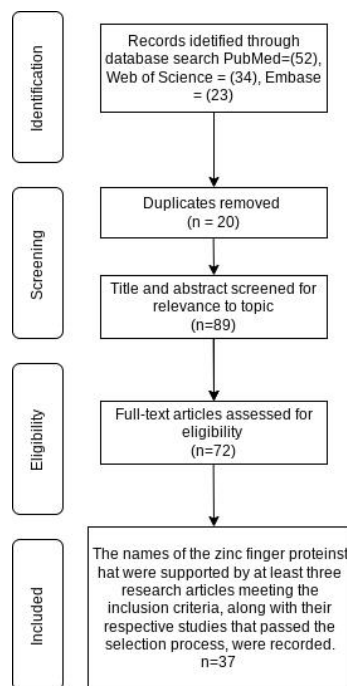


Fig. 7: Selection procedure for selected ZNFs based on research articles discussing their role in BC

mology were established based on the methodology employed by Boekhorst et al. (2007)⁹⁷, which includes the following parameters:

- An E-value threshold of 1e-5 or lower was chosen. The

E-value parameter estimates the expected number of chance hits in a database search of specific size. A lower E-value signifies a reduced potential for random matches, facilitating the identification of more statistically signifi-

cant alignments¹⁵.

- A minimum sequence identity of 30% was used as a cutoff. Sequence identity, the percentage of identical matches between two sequences, indicates evolutionary proximity. Higher percentages suggest a greater likelihood of homology¹⁵.
- A query coverage of at least 40% was required. Query coverage represents the proportion of the query sequence that well-aligns with the subject sequence, thereby ensuring that a substantial portion of the query protein is encompassed in the alignment¹⁵.

For the proteins that passed the criteria and are known to have multiple isomers, the accession ID of the full-length isomers were used, and duplicates were removed. The same method was employed if the protein appeared as unnamed. The results obtained are shown in Table 2. While BLASTp is particularly effective at identifying highly conserved domains, as is the case with ZFPs, it predominantly highlights the most similar regions and may overlook less conserved regions in multi-domain proteins, such as the C2H2 proteins. Such regions could potentially provide additional evidence for homology. Therefore, in order to conduct a more comprehensive analysis, the proteins passing the initial BLASTp screening were subsequently subjected to a pairwise sequence alignment using NEEDLE EMBOSS for further validation of homology¹². False positives might occur when non-homologous sequences share a high level of similarity due to convergent evolution or random chance, while false negatives might result from overly stringent thresholds that exclude genuine homologous sequences. Given the objectives of the present investigation, a scenario involving false positives is deemed more acceptable. Hence, through the usage of multiple databases and the usage of lenient boundaries, the methodology is designed to minimize the incidence of false negatives. The parameters used for evaluating the NEEDLE EMBOSS analysis include:

- **Alignment Score:** The score is a measure of the overall similarity between the two sequences, including both matches and gaps. A higher score generally indicates a better alignment⁹⁸.
- **Percentage Identity:** This refers to the percentage of residues that are identical in both sequences within the alignment. It is a direct measure of sequence conservation⁹⁸.
- **Percentage Similarity:** This parameter considers both identical and similar residues (for instance, residues that share similar physicochemical properties) within the alignment. It gives a broader view of sequence resemblance than percentage identity⁹⁸. The results obtained are shown in Table 2.

Expression of homologous ZFPs in BC:

A comprehensive literature review was conducted to investigate if the identified homologous ZFPs have been previously examined in the context of BC or general oncology. To carry out this research, the names of the homologous ZFPs were used in conjunction with the Boolean operator 'AND' and the terms 'BC' or 'cancer'. These terms were employed in a systematic search across four major databases: PubMed, Google Scholar, Embase, and Web of Science. The outcomes of this literature search are arranged in Table 3. Further confirmation of identified proteins' potential role in BC was conducted through the cBioPortal database, where their expression was compared with survival outcomes of patients. The cBioPortal database serves as a valuable resource in cancer research, providing extensive collections of cancer genomic data [100]. Researchers can access and analyze large-scale datasets, including gene expression profiles and clinical information, to gain insights into various cancer types. To establish a representative cohort of BC patients, the analysis included all samples available (n=8808). Data downloaded from a publicly available cBioPortal database does not require ethical approval. All patients whose samples were used in this analysis signed informed consent⁹⁹. Kaplan-Meier survival plots are commonly used in survival analysis to visualize and compare survival curves among different groups¹⁰⁰. In this study, these plots were utilized to assess the association between the expression of the ZFP and survival outcomes in BC patients. The generation of Kaplan-Meier survival plots and subsequent statistical analyses were facilitated by cBioPortal's integrated tools. Within the platform, the samples were divided into two groups based on their expression status for the ZFPs: the "altered" group (with expression) and the "unaltered" group (without expression). For each group, the cBioPortal platform generated survival plots. A comparison of the survival curves of the 'altered' and 'unaltered' groups allowed for the evaluation of whether ZFP expression significantly correlated with survival outcomes in BC patients. The statistical significance of this correlation was determined using p-values and q-values provided by the cBioPortal platform, while the hazard ratios (HR) were calculated separately by obtaining a text file of the data used in the Kaplan-Meier plots, ordering the data into a table using the python programming language, and analyzing it with CoxPHFitter from the lifelines library in python¹⁰¹. Adjacent with an HR value, a p-value was also calculated and checked for similarity to the one calculated in the cBioPortal itself. For all ZFPs the p-values were the same, indicating correct calculation of the HR value. In the context of this study, a small p-value ($p < 0.05$) indicates that the differences in survival outcomes between the 'altered' and 'unaltered' groups are statistically significant and therefore unlikely to be due to random chance. On the other hand, the q-value is a measure to control

the false discovery rate (FDR), an essential factor to consider when conducting multiple comparisons, as is common in genomics research. A smaller q-value represents a lower chance of a false positive result, which enhances confidence in the validity of a significant result, especially when numerous tests are performed. The hazard ratio (HR) is a useful measure in survival analysis, indicating the likelihood of death occurring in the 'altered' group compared to the 'unaltered' group. An HR greater than 1 suggests a higher risk in the 'altered' group, whereas an HR less than 1 suggests a lower risk⁹⁹. The results of this analysis are presented in Table 4.

Further analysis:

Based on the literature review and Kaplan-Meier survival plots, ZNF461 was chosen for further analysis with the aim of further investigating its possible role in BC and hence usage of homology-based prediction as a valid methodology. A protein-protein interaction (PPI) network for ZNF461 was established using the STRING database, providing an understanding of the potential interacting proteins and their implications in BC¹⁰². Genomic data derived from cBioPortal also served to examine ZNF461 expression and genetic alterations in BC samples. This process included the generation of an OncoPrint, a graphical representation of the alterations of ZNF461 in the samples (e.g., amplifications, deep deletions, structural variants, and mutations)¹⁰³. To visualize the frequency of those alterations, they were evaluated across the different study cohorts and BC subtypes. Mutation locations within the ZNF461 gene were also analyzed, contributing to an enriched comprehension of mutation-induced functional impacts. Finally, a co-expression analysis was conducted to identify genes with correlated mRNA or protein expression levels. Such an analysis is particularly important due to the role of ZNF461 as a transcription factor, and it could support the presence of the gene networks identified through the STRING database in BC.

Discussion

This study not only broadens the molecular understanding of BC by identifying potential new ZFPs in BC, but it also lays the groundwork for future detailed investigations. An immediate avenue for follow-up is validating the roles of these identified ZFPs. Techniques like co-immunoprecipitation (Co-IP) can be employed to profile interacting proteins, revealing the immediate molecular neighborhood of our hits. Furthermore, chromatin immunoprecipitation sequencing (ChIPseq) could shed light on the genes directly regulated by these ZFPs in both healthy and BC cells, providing insight into their mechanistic role in the disease. Another critical aspect for vali-

ation is probing disease-associated mutations in these ZFPs from patient sequencing data. Recognizing prevalent mutations may unveil structural or functional vulnerabilities that can be therapeutically targeted. These newly identified ZFPs offer promise as potential therapeutic targets, with potential therapies either inhibiting or amplifying their activity based on their role in tumor progression. The use of Kaplan-Meier plots in this research proves invaluable in discerning the prognostic significance of these proteins. Identifying ZFPs that are associated with aggressive tumor phenotypes could aid in the early prognosis of high-risk patients, while understanding the functional role of specific ZFPs in tumorigenesis could facilitate the development of targeted therapies. Additionally, evaluating the diagnostic potential of ZFPs, either as biomarkers or as potential therapeutic targets, could enhance the accuracy of breast cancer detection and treatment. If specific ZFPs consistently correlate with reduced survival rates in subsequent studies, they could become vital diagnostic markers. This would assist clinicians in identifying patients at elevated risk and consequently influence therapeutic decisions. Considering that ZFPs are pivotal regulators of gene expression, unveiling new ZFPs associated with breast cancer facilitates the exploration of the downstream genes or pathways they govern. This opens the door to a richer understanding of the molecular drivers of the disease.

Moreover, correlations between the expression or mutation status of these ZFPs and therapeutic outcomes could steer the development of tailored treatment plans, pushing the boundaries of personalized medicine. A deeper exploration of the underlying biological mechanisms is necessary to elucidate how ZFPs contribute to breast cancer progression or regulation. This involves investigating the interactions between ZFPs and other cellular components, such as DNA, RNA, and proteins, and understanding their role in the signaling pathways associated with breast cancer. It is also essential to explore the potential of ZFPs in modulating the tumor microenvironment, which plays a critical role in cancer progression and response to therapy. A comprehensive understanding of the role of ZFPs in breast cancer will not only impact clinical practice by enabling the development of more precise diagnostic tools and therapeutic strategies but will also improve patient outcomes by facilitating early detection and personalized treatment plans. This will contribute to the ongoing efforts to push the boundaries of personalized medicine and improve the overall survival rates of breast cancer patients. If specific ZFPs are found to be associated with early stages of breast cancer or with particularly aggressive forms of the disease, they could be used as biomarkers for early detection and prognosis. For instance, a high expression level of a particular ZFP in a patient's tumor tissue or blood sample could indicate a higher risk of rapid disease progression or metastasis. This information would allow clinicians to classify patients into differ-

ent risk categories and tailor their surveillance and treatment strategies accordingly, ultimately leading to better patient outcomes. Understanding the functional roles of different ZFPs in breast cancer progression will enable the development of targeted therapies. For instance, if a specific ZFP is found to play a critical role in tumor growth or invasion, inhibitors targeting this ZFP could be developed. This targeted approach would potentially lead to more effective treatments with fewer side effects, as it would specifically target the cancer cells without affecting the healthy ones. The expression or mutation status of ZFPs could also be used to predict a patient's response to specific therapies. For example, if a particular ZFP mutation is associated with resistance to a specific chemotherapy agent, patients with this mutation could be directed towards alternative treatments, thereby avoiding the toxicity and inefficacy of a treatment that is unlikely to be beneficial. Regular monitoring of ZFP expression or mutation status during the course of treatment could provide valuable information about the disease progression and the effectiveness of the ongoing treatment. For example, a decrease in the expression level of a ZFP associated with tumor aggressiveness could indicate a positive response to treatment, while an increase could signal a need to modify the treatment plan.

By incorporating these strategies into clinical practice, clinicians will be better equipped to make informed decisions regarding the diagnosis, prognosis, and treatment of breast cancer patients, ultimately leading to improved patient outcomes. This will contribute to the ongoing efforts to push the boundaries of personalized medicine and improve the overall survival rates of breast cancer patients. By juxtaposing the ZFPs highlighted in this research with those previously linked to breast cancer, the study fills in the existing knowledge chasms, underscoring the pivotal role of transcriptional regulation in advancing disease understanding. The recently identified ZFPs represent a promising frontier in the development of therapeutic interventions, given their potential to be harnessed for either inhibition or amplification, contingent on their specific roles in tumor progression. Nevertheless, this endeavor is not without its challenges. One concern is the potential for off-target effects, whereby the modulation of ZFP activity inadvertently impacts other cellular processes, precipitating unforeseen side effects. This is a consequence of the intricate network of interactions in which these proteins are involved, and the potential for cross-reactivity with other cellular components. Moreover, the technical hurdles associated with the development of targeted therapies cannot be understated. Crafting molecules with the requisite specificity to selectively target the ZFPs of interest, without inadvertently interacting with other proteins, presents a significant challenge. This necessitates a rigorous optimization process to ensure not only the efficacy of the therapeutic agent but also its safety profile. The role of ZNF461 in BC presents a

multifaceted subject for discussion, necessitating a thorough exploration of its potential functions, interactions, and overall relevance to the biological underpinnings of BC. ZNF461, inversely proportional to survival rate in BC, emerged as a significant area of interest, especially given its potential prognostic implications. Its proposed interactions with key proteins already implicated in BC, namely TOX2, HDAC2, and PHGDH, underscore a mechanistic interconnectedness that could provide critical insights into the cellular processes modulated by ZNF461 such as modulating the tumor microenvironment, promoting cellular proliferation, and facilitating metastasis. The genomic profiling of ZNF461 revealed a predominant state of amplification, punctuated by sporadic instances of deep deletions. Such amplification events are potentially linked to a surge in protein expression levels. Assuming ZNF461 functions as a growth-promoting factor, this could invariably facilitate oncogenesis and further propel tumor progression. This hypothesis aligns with the observed correlation between ZNF461 amplification and unfavorable prognostic outcomes in BC. Furthermore, the distinct patterns of ZNF461 amplification and mutations across different BC subtypes and stages suggest a nuanced role for ZNF461 in cancer progression or metastasis. This hypothesis is further substantiated by the predicted interactions between ZNF461 and proteins integral to cancer survival, metastasis, and tumor aggressiveness. Critically, mutations were frequently observed within the C2H2 domain of ZNF461, a region associated with DNA binding. Such mutations could potentially disrupt DNA binding capacity, consequently altering downstream gene regulation and impacting ZNF461's functional role as a transcription factor interacting with established oncogenes. This hypothesis aligns with the broader cancer hallmark of dysregulated gene expression and implies that the specific loci of mutations within ZNF461 may be pivotal in BC pathogenesis and progression.

Additionally, the observed co-expression between ZNF461 and JRKL, DEPDC4, and NUDCD1, albeit weak in some instances, suggests a potential interaction within the BC context. Given the established involvement of JRKL, DEPDC4, and NUDCD1 in BC, the detected co-expression, despite its relative weakness, strengthens the hypothesis that ZNF461 as a transcription factor functions as a central component in BC pathology. The absence of co-expression with other interacting proteins could potentially indicate that ZNF461 inhibits, rather than promotes, their transcription. However, the limited sample size for the analysis necessitates further investigation to substantiate both the statistical significance and the strength of the correlation. The genomic profile of ZNF461, its predicted interactions with key proteins, and the observed co-expression with specific proteins collectively suggest a nuanced role for ZNF461 as a transcription factor in BC pathogenesis and progression. Nonetheless, these findings neces-

sitate further validation and a more detailed mechanistic understanding to fully elucidate the role of ZNF461 in BC. The selection of online tools in this study enriched the analysis, yet the inclusion of additional tools might have enhanced its depth and breadth. A functional predictive analysis tool could have complemented the Kaplan-Meier analysis, particularly for assessing the potential of ZFPs as oncogenes. PolyPhen-2 (Polymorphism Phenotyping v2) could have been employed to predict the impact of amino acid substitutions on human protein structure and function, especially for ZFPs that displayed significant Kaplan Mayer plots but lacked literature review data. It would have also been interesting to compare the pathways of the homologous ZFPs to the identified ones to see if they have common interacting proteins, strengthening the legitimacy of the approach. Additionally, for a comprehensive pathway analysis of ZNF461 using STRING, tools like KEGG (Kyoto Encyclopedia of Genes and Genomes) could have contextualized the interactions within broader metabolic and signaling pathways. Such an approach would offer a panoramic view of potential impacts, deepening the understanding of these proteins' roles within cellular networks. Moreover, the inclusion of a domain-level homology analysis or a comparison of DNA binding motifs between different ZNFs could have been valuable. Mutations in these motifs can significantly influence the function of the ZNF. However, a major challenge in implementing this is that such data does not exist for many of the less-studied TFs.

This research approach demonstrates several strengths. A notable advantage is the comprehensive use of multiple tools and databases, such as BLASTp, NCBI, NEEDLE EMBOSS, cBioPortal, and STRING. Such diversity provides a holistic analysis, covering multiple aspects from protein function to clinical relevance. Moreover, the detailed validation process using both BLASTp and NEEDLE EMBOSS ensures a robust validation of homology. The approach also excels in its expression analysis, with Kaplan-Meier plots and hazard ratio evaluations providing a tangible clinical context for the identified ZFPs. The in-depth analysis of the selected protein, ZNF461, from its protein-protein interaction network to its co-expression analysis, offers a detailed perspective on its potential role in BC. Since a homology interference methodology hasn't been tested previously and has been used in the context of identifying novel oncogenes/tumor suppressor and in analyzing ZFPs, we have limited outlets for comparison. Nevertheless, ZNF711's role in BC was initially speculated in a paper on bioinformatics. When we were searching for homologous proteins to ZFX, ZNF711 was one of them. This in a way offers more confirmation for our findings. The methodology employed in this study has not been utilized specifically for the identification of new oncogenes and tumor suppressors, as well as for the analysis of zinc ZFPs. This limits the availability of comparable studies for reference. However, the

involvement of ZNF711 in BC was initially hypothesized in a bioinformatics paper¹⁰⁴. In this paper, we initially identified ZNF711 as a homologous protein to ZFX (oncogene). This prior indication of a connection between ZNF711 and BC through other bioinformatics approaches lends additional support to our findings and underscores the potential significance of this research.

However, the methodology is not without its limitations. While the databases used are reputable and commonly used in bioinformatics research, they each come with their own set of challenges and potential inaccuracies. Data entry errors, incomplete datasets, or even the algorithms they employ can introduce inconsistencies, making any findings based on them potentially error-prone. A second concern arises from the dependency on previously published research for expression analysis, which can restrict insights if the existing literature has its own biases or limitations. This would primarily be the case for the initial identification of ZFPs involved in BC and then for the literature review of the ZFP homologs' role in cancer, which are both valuable components influencing the legitimacy of the study. Another weakness is the assumption that sequence similarity directly correlates with functional relevance in BC. This may not always be accurate, and deviations from this assumption can influence the validity of the study's predictions. Nevertheless, the study aims to mitigate this error by ensuring further confirmation with diverse tools, especially with the further analysis on the role of ZNF461 in breast cancer. Moreover, as the research is fundamentally a bioinformatics analysis, all conclusions drawn are theoretical and lack experimental validation. Given the complexity of cancer, a bioinformatics-based approach can only provide preliminary insights. Concrete conclusions and therapeutic recommendations would necessitate further experimental investigations to validate the predictions made by this study. Overall, the findings of this research underscore the promise of leveraging homology interference as a potent tool for predicting cancer-associated biomarkers, specifically within the realm of ZFPs—a technique that remains markedly underutilized. This study provides an insightful and comprehensive synthesis of the function and impact of the investigated ZFPs and their homologs in the landscape of BC and wider oncology research. Additionally, it illuminates the potential avenues for future investigations, underscoring its significant implications for breakthroughs in cancer diagnostics and treatment strategies.

Acknowledgements

I would like to acknowledge my mentor, Ryan Prestil, for all of the guidance and support throughout the writing of this study.

References

- 1 B. Smolarz, A. Nowak and H. Romanowicz, *Cancers*, 2022, **14**, 2569.
- 2 M. Tufail, J. Cui and C. Wu, *American Journal of Cancer Research*, 2022, **12**, 2920.
- 3 N. Laham-Karam, G. Pinto, A. Poso and P. Kokkonen, *Frontiers in Chemistry*, 2020, **8**, year.
- 4 ScienceDirect Topics, *Transcription Factors - an overview*, 2012, <https://www.sciencedirect.com/topics/neuroscience/transcription-factors>.
- 5 T. Ravasi, *Genome Research*, 2003, **13**, 1430–1442.
- 6 S. Krishna, *Nucleic Acids Research*, 2003, **31**, 532–550.
- 7 J. Jen and Y. Wang, *Journal of Biomedical Science*, 2016, **23**, year.
- 8 J. Lu, G. Getz, E. Miska and et al., *Nature*, 2005, **435**, 834–838.
- 9 M. Cassandri, A. Smirnov, F. Novelli and et al., *Cell Death Discovery*, 2017, **3**, 1–12.
- 10 B.-H. Bay, R. Jin, J. Huang and P.-H. Tan, *Experimental Biology and Medicine*, 2006, **231**, 1516–1521.
- 11 A. Fedotova, A. Bonchuk, V. Mogila and P. Georgiev, *Acta Naturae*, 2017, **9**, 47–58.
- 12 S. Sinha, B. Eisenhaber and A. Lynn, *Bioinformatics: Sequences, Structures, Phylogeny*, 2018, pp. 289–305.
- 13 S. Brachat, F. Dietrich, S. Voegeli and et al., *Genome Biology*, 2003, **4**, R45.
- 14 J. Díaz-Mejía, M. Babu and A. Emili, *FEMS Microbiology Reviews*, 2009, **33**, 66–97.
- 15 J. Díaz-Mejía, M. Babu and A. Emili, *FEMS Microbiology Reviews*, 2009, **33**, 66–97.
- 16 D. Wheeler and M. Bhagwat, *BLAST QuickStart*, 2016, <https://www.ncbi.nlm.nih.gov/books/NBK1734/>.
- 17 K. Rajput, M. Ansari, S. Jha and et al., *bioRxiv*, 2022.
- 18 Y. Zhou, G. Ma, S. Peng and et al., *Thoracic Cancer*, 2021, **12**, 2427–2438.
- 19 S. Rhie, L. Yao, Z. Luo and et al., *Genome Research*, 2018, **28**, 310–320.
- 20 G. Carney, D. Bush, S. Sellers and et al., *ZFX expression is connected with improved survival in patients with breast cancer*, 2021, <https://doi.org/10.21203/rs.3.rs-592491/v1>.
- 21 O. Fedorova, A. Daks, S. Parfenyev and et al., *Biochemical and Biophysical Research Communications*, 2022, **589**, 29–34.
- 22 N. Mohammadi Ghahhari, M. Sznurkowska, N. Hulo, L. Bernasconi, N. Aceto and D. Picard, *Nature Communications*, 2022, **13**, year.
- 23 X. Zhang, Z. Zhang, Q. Zhang and et al., *Cell Death Disease*, 2018, **9**, year.
- 24 H. Wu, H. Zhong, G. Li and et al., *Journal of Translational Medicine*, 2020, **18**, year.
- 25 M. Perez-Oquendo and D. Gibbons, *Cancers*, 2022, **14**, 1864.
- 26 X. Li, L. Tian, L. Zhang, B. Xu, Y. Zhang and Q. Li, *OncoTargets and Therapy*, 2020, **13**, 6593–6601.
- 27 G. Wu, H. Peng, M. Tang and et al., *eBioMedicine*, 2021, **71**, 103558.
- 28 A. Köferle, A. Schlattl, A. Hörmann and et al., *Cell Reports*, 2022, **39**, 110636.
- 29 W. Ni, A. Perez, S. Schreiner, C. Nicolet and P. Farnham, *Nucleic Acids Research*, 2020, **48**, 5986–6000.
- 30 X. Wu, X. Zhang, L. Yu and et al., *Oncogene*, 2020, **39**, 2568–2582.
- 31 H. Zeng, T. Li, D. Zhai and et al., *International Journal of Biological Sciences*, 2020, **16**, 2084–2093.
- 32 G. Al-Awsi, F. Alsaikhan, R. Margiana and et al., *Stem Cell Research Therapy*, 2023, **14**, year.
- 33 C. Wei and X. Li, *BMC Cancer*, 2020, **20**, year.
- 34 N. Yan, C. Liu, F. Tian and et al., *International Journal of Genomics*, 2021, **2021**, 4301802.
- 35 E. Saberi Ansar, C. Eslahchii, M. Rahimi and et al., *BMC Medical Genomics*, 2019, **12**, year.
- 36 P. R. de Alcantara Filho, F. R. Mangone, A. C. Pavanelli, S. A. de Bessa Garcia, S. Nonogaki, C. A. de Toledo Osório, V. P. de Andrade and M. A. Nagai, *Breast Cancer Management*, 2018, **7**, BMT09.
- 37 Z. Harper, L. Yan, R. Loureiro and et al., *Cancer Research*, 2007, **67**, 8736–8741.
- 38 D. Jia, S. Hasso, J. Chan and et al., *Blood*, 2013, **121**, 707–715.
- 39 D. Jia, L. Huang, J. Bischoff and M. Moses, *The FASEB Journal*, 2014, **29**, 1371–1382.
- 40 R. Hu, G. Peng, H. Dai and et al., *Cancer Research*, 2011, **71**, 6524–6534.
- 41 R. Hu, *Function of ZNF668 in Cancer Development*, 2011, https://digitalcommons.library.tmc.edu/utgsbs_dissertations/186/?utm_source=digitalcommons.library.tmc.edu%2Futgsbs_dissertations%2F186&utm_medium=PDF&utm_campaign=PDFCoverPages, Accessed on July 15, 2023.
- 42 R. Hu, E. Wang, G. Peng, H. Dai and S. Y. Lin, *Cell Cycle*, 2013, **12**, 2033–2041.
- 43 Y. Xiao, T. Xiang, X. Luo and et al., *PLOS ONE*, 2014, **9**, e110990.
- 44 Y. Cheng, P. Liang, H. Geng and et al., *Molecular Cancer Research: MCR*, 2012, **10**, 925–936.
- 45 D. Yang, Y. Wu, Q. Xiang, D. Zhou, Z. Qiu, X. Yin, B. Zeng, J. Tang, X. Luo, W. Peng et al., 2021.
- 46 W. Si, W. Huang, Y. Zheng and et al., *Cancer Cell*, 2015, **27**, 822–836.
- 47 Y. He, J. Northey, A. Pelletier and et al., *Oncogene*, 2017, **36**, 3490–3503.
- 48 R. Xie, J. Tang, X. Zhu and H. Jiang, *Bioscience Reports*, 2019, **39**, year.
- 49 C. Fan, J. Du and N. Liu, *Computational and Mathematical Methods in Medicine*, 2021, **2021**, 1–15.
- 50 M. Bakhtiari, J. Park, Y. Ding and et al., *Nature Communications*, 2021, **12**, 2075.
- 51 J. Li, L. Bi, Z. Shi and et al., *Pathology - Research and Practice*, 2016, **212**, 549–554.
- 52 Y. Wang, R. Xu, D. Zhang, T. Lu, W. Yu, Y. Wo, A. Liu, T. Sui, J. Cui, Y. Qin et al., *Translational Lung Cancer Research*, 2019, **8**, 862.
- 53 N. Yu, H. Gong, W. Chen and W. Peng, *Translational Lung Cancer Research*, 2023, **14**, 1467–1476.
- 54 T. Xi and G. Zhang, *Pathology - Research and Practice*, 2017, **213**, 83–88.
- 55 Y. Zhang, T. Zhu, F. He, A. Chen, H. Yang and X. Zhu, *CARTILAGE*, 2021, **13**, 1457S–1464S.
- 56 L. Du, X. Tao and X. Shen, *Breast Cancer*, 2021.
- 57 F. Xu, L. Zhang, Y. Xu and et al., *Cancers*, 2020, **12**, 2332.
- 58 X. Wen, H. Jin, M. Li and et al., *Scientific Reports*, 2022, **12**, 862.
- 59 M. Unoki, J. Okutsu and Y. Nakamura, *International Journal of Oncology*, 2003, **22**, 1217–1223.
- 60 Y. Han, Q. Zheng, Y. Tian, Z. Ji and H. Ye, *Journal of Surgical Oncology*, 2019, **119**, 1145–1154.
- 61 Q. Zhao, F. Wang, Y. Chen and et al., *Nature Communications*, 2022, **13**, 2342.
- 62 J. Dufresne, P. Bowden, T. Thavarajah, A. Florentinus-Mefailoski, Z. Z. Chen, M. Tucholska, T. Norzin, M. T. Ho, M. Phan, N. Mohamed et al., *Clinical proteomics*, 2018, **15**, 1–19.
- 63 W. Fu, J. Zhao, W. Hu and et al., *Cancer Management and Research*, 2021, **13**, 5671–5681.
- 64 K. Agelopoulos, G. Richter, E. Schmidt and et al., *Clinical Cancer Research*, 2015, **21**, 4935–4946.
- 65 T.-C. Chang, *PhD thesis*, The George Washington University, 2017.
- 66 C. Chen, J. Wu, Y. Lin and et al., *International Journal of Molecular Sciences*, 2022, **23**, 10225.
- 67 W. Yang, S. J. Chen, X. Cheng and et al., *Annals of Translational*

- Medicine*, 2021, **9**, 1659.
- 68 L. Rink, Y. Skorobogatko, A. Kossenkov and et al., *Molecular Cancer Therapeutics*, 2009, **8**, 2172–2182.
 - 69 Y. Tian, J. Wang, C. Qin and et al., *Frontiers in Genetics*, 2020, **11**, year.
 - 70 Z. Cui, Q. Fan and C. Jia, *Signal Processing*, 2020, **174**, 107614.
 - 71 A. Naghsh-Nilchi, L. Ebrahimi Ghahnavieh and F. Dehghanian, *Journal of Cellular and Molecular Medicine*, 2022, **26**, 4530–4547.
 - 72 J. Addison, C. Koontz, J. Fugett and et al., *Cancer Research*, 2015, **75**, 344–355.
 - 73 M. Kumaran, C. Cass, K. Graham and et al., *Scientific Reports*, 2017, **7**, year.
 - 74 I. Panagopoulos, K. Andersen, I. M. Johannsdottir, F. Micci and S. Heim, *Cancer Genomics & Proteomics*, 2022, **20**, 51–63.
 - 75 N. Li, S. Rowley, E. Thompson, S. McInerney, L. Devereux, D. Goode, J.-S. Lee, N. Li, G. Dite, R. Milne et al., *Breast Cancer Research*, 2018, **20**, year.
 - 76 S. Mamoor, *A gene expression signature of epithelial ovarian cancer in the blood*, OSFPREPRINTS, Year of access, <https://osf.io/4xmn6>.
 - 77 O. Morin, *A blood gene expression signature for epithelial ovarian cancer*, OSFPREPRINTS, https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwja96Pf35GAAXW-S_EDHfGdA68QFnoECCoQAQ&url=https%3A%2F%2Fosf.io%2Fpreprints%2Fj958v%2F&usq=AOvVaw2txJYFQMEB5tYqlrD4245t&cshid=1689459848019890&opi=89978449, Accessed on July 16, 2023.
 - 78 Y. Zwang, O. Jonas, C. Chen and et al., *eLife*, 2017, **6**, e24523.
 - 79 J. Sobocińska, J. Nowakowska, S. Molenda and et al., *Current Oncology*, 2022, **29**, 9896–9915.
 - 80 J. Zhang, Y. Wang, B. Yuan, H. Qin, Y. Wang, H. Yu, X. Teng, Y. Yang, J. Zou, M. Zhang et al., *Frontiers in Oncology*, 2023, **13**, 1112020.
 - 81 J. Li, L. Bi, Z. Shi and et al., *Pathology - Research and Practice*, 2016, **212**, 549–554.
 - 82 M. H. Wang, X. Du, Z. Su and X. Yu, *Laboratory Medicine*, 2020, **51**, 377–384.
 - 83 C. Liang, S. Huang, Y. Zhao, S. Chen and Y. Li, *Biomarker Research*, 2021, **9**, year.
 - 84 M. Tessema, C. Yingling, M. Grimes and et al., *PLoS ONE*, 2012, **7**, year.
 - 85 J. Zhou, S. H.-M. Toh, T. K. Tan and et al., 2023, **22**, year.
 - 86 N. Darvishi, K. Rahimi, K. Mansouri and et al., *Molecular and Cellular Probes*, 2020, **53**, year.
 - 87 N. Darvishi, K. Rahimi, K. Mansouri and et al., *Molecular and Cellular Probes*, 2020, **53**, year.
 - 88 J. W. Locasale, A. R. Grassian, T. Melman and et al., *Nature Genetics*, 2011, **43**, 869–874.
 - 89 X. Zhao, J. Fu, J. Du and W. Xu, *International Journal of Biological Sciences*, 2020, **16**, 1495–1506.
 - 90 L. Pangon, I. Ng, M. Giry-Laterriere and et al., *Oncogene*, 2016, **35**, 2834–2841.
 - 91 M. Hirao-Suzuki, T. Koga, G. Sakai and et al., *Biochemical and Biophysical Research Communications*, 2020, **531**, 215–222.
 - 92 M. Zhu, N. Liu, J. Lin, J. Wang, H. Lai and Y. Liu, *Oncology Letters*, 2023, **25**, 33.
 - 93 C. Liu, L. Sun, J. Yang and et al., *Proceedings of the National Academy of Sciences*, 2018, **115**, 13075–13080.
 - 94 M. Yan, J. Wang, Y. Ren and et al., *Cell Death & Disease*, 2019, **10**, 1–13.
 - 95 F. Lin, Z. Xie, L. Chang and et al., *International journal of clinical and experimental pathology*, 2017, **10**, 11014–11021.
 - 96 P. B. Hi and D. A. Liberles, *Protein Science*, 2016, **25**, 1168–1178.
 - 97 J. Boekhorst and B. Snel, *BMC Bioinformatics*, 2007, **8**, 356.
 - 98 J. Blazewicz, W. Frohberg, M. Kierzyńska, E. Pesch and P. Wojciechowski, *BMC Bioinformatics*, 2011, **12**, year.
 - 99 C. J. Creighton, *Current Protocols in Molecular Biology*, 2018, **121**, 19.14.1–19.14.13.
 - 100 J. T. Rich, J. G. Neely, R. C. Paniello, C. C. J. Voelker, B. Nussenbaum and E. W. Wang, *Otolaryngology–Head and Neck Surgery*, 2010, **143**, 331–336.
 - 101 C. Davidson-Pilon, *Journal of Open Source Software*, 2019, **4**, 1317.
 - 102 D. Szklarczyk, A. L. Gable, K. C. Nastou and et al., *Nucleic Acids Research*, 2020, **49**, D605–D612.
 - 103 M. Radovich, P. J. Kiel, S. M. Nance and et al., *Oncotarget*, 2016, **7**, year.
 - 104 L. Tian, L. Zhang, B. Xu, Y. Zhang and Q. Li, *OncoTargets and Therapy*, 2020, **13**, 6593–6601.