Reviewed Preprint
v1 • October 31, 2025
Not revised

Reviewed Preprint
v2 • May 8, 2026
Revised by authors

Mutational and Expression Profile of ZNF217, ZNF750, ZNF703 Zinc Finger Genes in Kenyan Women Diagnosed with Breast Cancer

Michael Kitoi[1,3] ✉, **John Gitau**[2], **Godfrey Wagutu**[3], **Kennedy Mwangi**[4], **Florence Ngonga**[1], **Francis Makokha**[3]

[1]Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya • [2]Département de Chimie, Université du Quebéc à Montréal, Montréal, Canada • [3]Directorate of Research and Innovation, Mount Kenya University, Thika, Kenya • [4]International Livestock Research Institute, Nairobi, Kenya

For correspondence: Michael.o.kitoi@gmail.com

Competing interests: No competing interests declared

Funding: See page 20

Reviewing editor: Yongliang Yang, Shanghai University of Medicine and Health Sciences, China

eLife Assessment

This study presents a **valuable** finding on the mutational landscape and expression profile of ZNF molecules in 23 Kenyan women with breast cancer. The evidence supporting the claims of the authors is **solid**, although inclusion of a larger number of patient samples, more statistical details and sufficient comparison with existing large-scale datasets would have strengthened the study. The work will be of interest to medical biologists working in the field of breast cancer.

https://doi.org/10.7554/eLife.108076.2.sa4

Abstract

Objectives To characterize the somatic mutational spectrum and transcriptomic expression of the zinc-finger genes ZNF217, ZNF703, and ZNF750 in Kenyan women with breast cancer, and to explore their associations with clinicopathologic features.

Methods Whole-exome sequencing and RNA-sequencing were performed on paired tumor and adjacent normal tissues from 23 consented patients treated at two Kenyan referral hospitals. Variants were called with Mutect2 using a study-specific panel of normal; functional consequences were annotated with VEP. After featureCounts quantification, differential expression was analyzed in DESeq2 (fold-change ≥ 1.5, $p < 0.05$). Two-sample *t*-tests (mutations) and ANOVA (expression) evaluated relationships with HER2 status and clinical stage.

Results A total of 358 somatic mutations were detected: 170 in ZNF217, 24 in ZNF70*3* and 164 in ZNF750. Single-nucleotide substitutions (319 SNPs) dominated, with C→T and A→G changes most common; 27 deletions and 2 insertions were also observed. Frameshift events in ZNF217 and ZNF703 introduced premature stop codons predicted to truncate protein function. All three genes were significantly up-regulated in tumors versus normal (ZNF217 p = 0.0004), with the greatest expression in HER2-positive tumors and in stages 2–3 disease. Mutation burden for each gene did not differ by HER2 status ($p > 0.56$) or by stage ($p > 0.32$).

Conclusions Kenyan breast tumors harbor frequent, functionally relevant mutations and marked over-expression of ZNF217, ZNF703, and ZNF750. These alterations, especially the pronounced up-regulation of ZNF217, highlight the trio's potential as diagnostic or prognostic biomarkers and warrant larger studies to validate their clinical utility and suitability as therapeutic targets in sub-Saharan African populations.

Introduction

The global incidence of cancer reported approximately 20 million new cases in 2022 including 11.6% female breast cancer case which ranked as the second most common cancer worldwide (Bray et al., 2024). Cancer research indicates that breast cancer represented 6.7 percent of total cancer-related fatalities among the 9.7 million deaths (Bray et al., 2024). The GLOBOCAN 2022 report shows breast cancer takes the lead position as the most common cancer affecting Kenyan females with a rate of 25.6% for new cancer cases (Global Cancer Observatory, 2022). Breast cancer stands as the main cause of fatalities in Keenya where the annual new cases reached 7,243 (16.2% of all cases) while mortality rates showed 3,398 deaths (Global Cancer Observatory, 2022).

Recent genomic studies have proven that Zinc Finger proteins (ZNF) play an essential role in the development of breast cancer (Stradella et al., 2022). Breast cancer pathogenesis shows ZNF217, ZNF750 and ZNF703 as potential contributors because several mutations and dysregulated expression patterns during development have been linked to disease progression. ZNF217 functions as a normal cellular repressor that controls both cell survival and growth dynamics (Krig et al. (2010). The overexpression of ZNF217 occurs when amplifications or mutations take place causing gene expression disturbances as well as aggressive breast cancer development that affects African ancestry populations (Ansari-Pour et al., 2021). Breast cancer in the African population undergoes more aggressive development due to ZNF217 non-coding mutations (Ansari-Pour et al., 2021). Scientific studies have demonstrated that ZNF217 frame-shift mutations damage protein structures which results in improper gene expression patterns that create cancer development (Krig et al., 2010a). Breast cancer cells express ZNF217 at 20q13 amplification regions that drive the transcription and protein synthesis of Erb-B2 Receptor Tyrosine Kinase 3 (ErbB3) (Krig et al., 2010).

ZNF750 gene plays a regulatory role in epidermal differentiation through its control of differentiation-specific genes located on chromosome 17q25.3 (Butera et al., 2020). ZNF750 functions as a tumor suppressor in normal tissue because it blocks cell multiplication and facilitates cell differentiation. Genomic abnormalities in ZNF750 disrupt crucial cell functions that then increase cancer development probabilities according to (Butera et al., 2020) across various aggressive breast cancer variants. Similarly, ZNF750 gene fusions promote cancer-related signal pathways for breast cancer progression. Butera et al. (2020) demonstrate that ZNF750 gene mutations are connected with two types of breast cancer: aggressive triple-negative breast cancer (TNBC) and lymph node metastasis. Mutations in ZNF750 lead to dysfunctional cellular processes in breast tissue because they generate both the condition of hyper-expression and hypo-expression of ZNF750 (Cassandri et al., 2020).

The transcription factor ZNF703 functions normally by controlling developmental processes as well as cellular differentiation pathways when present on chromosome 8p12 (Klæstad et al., 2021). The study of ZNF703 in breast cancer shows that it activates oncogenic signaling through excessive protein production, which advances tumor growth by increasing cell survival and growth potential (Zhang et al., 2022). The poor clinical outcome and aggressive features of breast cancer are strongly associated with high ZNF703 expression levels, especially in patients with luminal B breast cancer subtypes (Marzbany et al., 2019). The research on ZNF703 gene-copy associations with breast cancer proliferation and outcome and Luminal subtype revealed that tumor cells with high copy numbers showed increased proliferation and aggressive traits in 7 percent of cases, most frequently within the Luminal B subgroup (Klæstad et al., 2021).

In the current study, therefore, whole genomic exome and RNA-sequencing data of tumor and adjacent normal tissues of 23 Kenyan women with breast cancer were analyzed. We aimed (i) to define the somatic-mutation map of the three genes, ZNF217, ZNF703 and ZNF750, (ii) to measure their level of dysregulation in cancer and normal breast tissue, and (iii) to investigate links between alterations of these genes and clinicopathologic parameters. Verifying the molecular phenotypes of these ZNF genes in a Kenyan population will progress precision-oncology studies in sub-Saharan Africa and provide the foundations of subsequent biomarker validation and identification of new drug targets.

Materials and methods

Patients and Samples

This research collected its data from a pilot study examining breast cancer progression in Kenyan patients {Institutional Ethics Review Committee (IERC), Ref: 2018/REC-80 (v7); License No: NACOSTI/P/22/21374}. After patients granted their consent at AIC Kijabe Hospital and Aga Khan University Hospital, Nairobi, purposive sampling enrolled 23 patients with malignant breast cancer. Fresh tumor tissue together with adjacent normal tissue were extracted from 23 patients, while 46 patient samples were sent to the National Cancer Institute at Bethesda MD USA for sequencing procedures.

Whole-exome sequencing and RNA-sequencing

Tang et al. (2023) provided the methods for processing tumors together with normal samples. Whole-exome sequencing together with RNA-sequencing was done using a Novaseq system which produced paired end reads of 150bp at a total read depth of 30 million according to Tang et al. (2023). Tumor sample analysis was done at 250x depth while normal adjacent tissue analysis reached 150x depth (Tang et al., 2023).

Reads mapping and variant calling

The workflow proceeded with quality control evaluations of raw whole-exome data (WED) and RNA-seq data from twenty-three samples through implementation of FastQC (ref). MultiQC (Ewels et al., 2016) generated the summary of results. The reads underwent trimming for low quality and adapter elements through Trimmomatic before undergoing FASTQC and MultiQC quality screening. The hg38 human reference genome received mapped BWA MEM (ref) reads from those samples successful during the quality control phase. The analysis report demonstrated that mapping fully succeeded for greater than 95% of input reads. The Picard tool performed read duplication followed by the addition of read-groups into the deduped BAM files. The base quality recalibration process utilized GATK according to G. Zhang et al. (2015). The somatic variant calling process in paired tumor normal mode occurred through MuTect2 (ref) (Chen et al., 2020).

The analysis made use of the panel of normal (PON) derived from normal reads obtained through the study. Normal samples entered Mutect2 with tumor-only mode in order to produce the panel of normal according to Chen et al. (2020). The CreateSomaticPanelOfNormals tool (O'Donovan, 2019) performed variant call aggregation of individual results from separate processing of each normal sample. The somatic variant detection accuracy increased because this method helped identify and eliminate technical artifacts and germline variants that normally exist in normal samples. The PON accompanied variant detection to remove artifacts together with germline variants while ensuring proper identification of somatic mutations. The variant tool set normalization (VT) (Bayat et al., 2017) served as the variant normalization method, which involved GATK filtering, while the variant effect predictor (VEP) (Salz et al., 2023) delivered functional/consequence annotations. Each annotation score (whether, moderate, high or low) originated from the gene effect analysis. The variants linked to ZNF genes can be viewed using the Integrative Genomic Viewer (IGV) (ref) (Robinson et al., 2017).

The panel of normal (PON) used for analysis stemmed from normal reads, which were collected for this purpose. Mutect2 with tumor-only mode processing generated the panel of normal from test samples according to Chen et al. (2020). Each individual normal sample received independent processing for variant calling before the CreateSomaticPanelOfNormals tool combined the output (O'Donovan, 2019). The somatic variant detection accuracy increased because this method helped identify and eliminate technical artifacts and germline variants that normally exist in normal samples. The PON operated as a filtering tool during variant calling to delete technical artifacts together with germline variants which ensured precise somatic mutation identification. The variant tool set normalization (VT) (Bayat et al., 2017) served as the variant

normalization method which involved GATK filtering while variant effect predictor (VEP) (Salz et al., 2023) delivered functional/consequence annotations. The annotation process determined the impact levels (whether, moderate, high or low) of each gene. The variants linked to ZNF genes can be viewed using the Integrative Genomic Viewer (IGV) (ref) (Robinson et al., 2017).

Gene Expression Analysis

Quality assessment through FASTQC with MultiQC tools was conducted before performing the RNA-seq analysis for the 46 processed samples. Testing of sample quality yielded satisfactory results so researchers performed human reference genome mapping of the samples using the STAR aligner. The program calculated the reference genome comparisons with each read sequence by determining the number of mismatches (NM). The featureCounts software (Liao et al., 2014) determined the quantitative count of reads per gene. DESeq2 software executed normalization procedures on the gene expression data (Malick et al., 2022). Each sequencing read proceeded to produce raw gene count data after the quantification process. The raw counts measured the sequencings' numbers that aligned to each gene and served as the starting point for normalization procedures. DESeq2 applied size factor normalization which generated unique size factors for each sample to normalize the varying sequencing depths in the data (Malick et al., 2022). Size factors were determined by finding the median value of dividing gene counts by their corresponding geometric mean across all measured genes. The normalization procedure ended by producing a count matrix that contained normalized gene expression measurements for every sample.

Differential Gene Expression (DGE) Analysis and Principal Component Analysis (PCA)

The analysis of Differential Gene Expression (DGE) through DESeq2 occurred in R (Malick et al., 2022). The biological variance calculation for each experimental condition combines replicated results to establish expression distinctions between two specific conditions. The identified differentially expressed genes were identified through a combination of ≥ 1.5-fold change and statistical significance ≤ 0.05 (Gitau et al., 2024). Principal Component Analysis (PCA) using Python processed normalized gene expression data in order to detect expression profile patterns and reduce dataset dimensions of ZNF217, ZNF750, and ZNF703 genes (Greenacre et al., 2022). R was used for detailed visualization of ZNF217, ZNF750, and ZNF703 gene differential expression through ggplot2 boxplot creation.

Statistical Significance

Simple t-test established the expression level differences' significance between normal and tumor samples through a statistical significance analysis.

Results

Patients and sample characteristics

The research has detailed the demographic and clinical/pathological characteristics of the 23 studied patients in Table 1. Participation groups included cancer stages together with human epidermal growth factor receptor 2 (HER2) expression results. Table 1 reveals that 52.2% of patients showed HER2+ status while 8.6% were in cancer stage I and 34.7% had stage III and 47.8% were at stage II of their condition.

ZNF703 Gene Mutations

A total of 24 gene mutations were detected in the 23 patient samples including 20 SNPs combined with 3 deletions and a single insertion (Figure 1A). A total of twelve patients among the twenty-three samples (12) exhibited ZNF703 gene mutations while eleven patients did not show any mutations in the gene (Figure 1B). Patient 237 demonstrated the greatest mutating activity with

Sample numbers	AGE	Systolic	Diastolic	heart rate	Weight	BMI	Cancer Stage	HER2
201	65	142	94	94	59	22.5	2	Positive
203	30	101	64	75	70.66	27.8	3	Positive
205	51	137	92	79	75	23.68	3	Positive
207	72	172	85	71	50	23.14	2	Negative
209	66	147	82	100	70	30.30	2	Negative
211	45	133	96	82	84	32.81	1	Positive
213	43	150	87	64	84	31.42	2	Negative
215	44	129	83	96	75.5	25.52	3	Positive
217	41	116	60	86	74.4	29.06	2	Positive
219	36	114	72	75	67.3	26.26	3	Positive
221	53	123	86	81	78	31.96	2	Negative
223	58	113	67	98	73.8	27.95	2	Negative
225	53	147	83	96	70	28.8	1	Positive
227	60	150	92	75	79	30.9	1	Negative
229	54	119	81	80	101	37.8	2	Positive
231	72	134	91	74	71	27.7	2	Negative
233	44	144	94	88	63	25.6	2	Positive
235	49	120	84	90	98.5	34.9	3	Negative
237	66	124	55	78	63	24.9	3	Positive
239	53	112	67	78	61	21.1	1	Negative
241	58	133	88	71	90		2	Negative
243	51	137	83	119	70	27.7	3	Positive
245	66	153	87	118	66	24.8	3	Negative

Table 1.

four different mutations, yet 14 patients presented one mutation and 8 patients did not have mutagenic changes (Figure 1C). The C/T substitutions appeared as the dominant substitution type whereas G to C occurred least frequently (Figure 1C). Study results showed four missense types among the classified variants and patient 237 held the sole frameshift mutation in the ZNF703 gene (Table 2). The variants affecting this gene generated moderate effects across samples except for patient 237 who experienced a high impact due to a frameshift mutation (Table 2). Missense types resulted in moderate consequences for ZNF703 genes of patient 201 and 205 and were also observed in patients 223 and 233 but to a lower extent. A high-degree of influence resulted from the in-frame deletion which replaced Alanine 517. The in-frame deletion demonstrated homozygous removal affecting both alleles of this gene when visualized through IGV (Figure 1D).

ZNF750 Gene Mutations

A total of 164 mutations were identified among the 23 patients who exhibited 154 SNPs and 9 deletions, together with 1 insertion (Figure 2A). About three samples contained 15 to 12 and 11 SNP mutations while two other samples showed fewer than three SNPs in total (Figure 2B).

Vertical transitions based on A/G substitutions constituted 40 mutations in the gene pool, yet A/C substitutions and A/T substitutions, together with C/A and G/T substitutions, totaled fewer than 10 counts (Figure 2C). The highest occurring indel mutation is TTTTG/T which occurred five times, and all other mutations occurred only once during the analysis (Figure 2D).

The C to T DNA change at positions 82,831,023 and 82,831,023 in two specimens led to a coding sequence mutation transforming amino acid valine 478' to isoleucine, which shares comparable hydrophobic properties (Table 3). A genetic alteration at position 82,830,234 in sample 231 modified Asp to Asn according to the information contained in Table 3. At position 82,831,982 of sample 241, the C/T change transformed arginine (Arg) into glutamine (Gln at position 158' as listed in Table 3.

The majority of T/C SNPs resulted in no change of the amino acid sequence as Ala 575' remains constant in the protein chain (Figure 2E) (Table 4). Three specimens demonstrated T to C substitution at position 82,831,280 of the gene, leading to missense mutations which changed glutamine (Gln) at position 392 into arginine (Table 4 and Figure 2E). The T/C substitution at position 82,831,752 generated a missense mutation altering methionine (Met) at position 235' to valine (Val) based on the data in Table 4.

These gene alterations at position 354 produced no impact on the ZNF750 protein because they maintained amino acid threonine (Thr) while position 165' retained valine (Val), resulting in an unmodified subsequent protein structure (Table 5). At position 82,831,752 of the gene, twelve annotated A/G were found to be synonymous because proline (Pro) at specific positions maintained the same amino acid sequence without affecting protein functionality (Table 6).

ZNF217 Gene Mutations

Investigators uncovered 170 mutations along with 145 SNPS and 24 deletions and one insertion throughout the 23 samples (Figure 3A). A total of 105 C to T substitutions were the most frequent mutations among ZNF217 changes while G to C substitutions only happened less than two times (Figure 3B). GAGAC/C in-frame deletion represented the most often detected mutation in patients with eight mutations among all samples (Figure 3B). A total of 36 mutations among all identified changes have received annotations. Recorded mutations comprised of C/T, C/A and TA/T variants with 22 mutations classified as missense and 13 as synonymous mutations and one classified as a frameshift disturbance (Figure 3D).

Twelve mutations identified turned out to be missense variations, and two mutations were defined as synonymous. Ten C to T mutations found across the samples resulted in the creation of ZNF217 gene variants with isoleucine (Ile) at position 739 instead of valine (Val) due to a change in

Table 2.

Patient ID	Position	Substitution /Indel	Mutation Type	Impact Level	Amino Acid
201	37,697,439	C/T	Missense	MODERATE	Ser 177' Phe
	37,698,452	T/C	Synonymous	LOW	His 517'
205	37,698,165	C/G	Missense	MODERATE	Pro 422' Ala
215	37,698,266	G/T	Synonymous	LOW	Pro 455'
223	37,697,290	C/A	Missense	MODERATE	Ala 130' Asp
233	37,698,467	G/C	Synonymous	LOW	Pro 455'
	37,697,422	C/T	Missense	Moderate	Ser 174' Leu
237	37,698,415	GCGC CGCCGC/G	In-frame deletion	HIGH	Ala 514' del

Figure 1. Mutation profile of ZNF7O3 gene in 23 samples.

A) Type of variants that occurred in the gene. B) Distribution of substitutions per sample. C) Distribution of substation type. D) The mutation viewed under IGV showing deletion of GCGCCCGCCGC and insertion of G in sample 237.

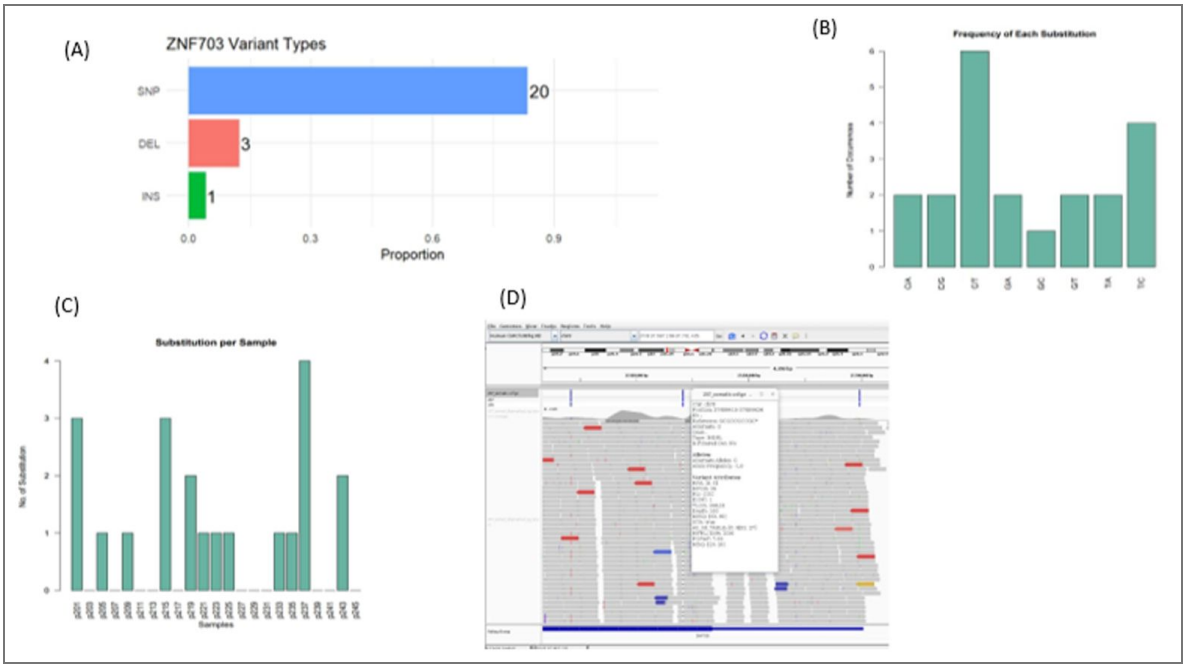


Figure 2. **Mutation profile of ZNF75O gene in 23 patients.**

A) Type of variants that occurred in the gene. B) Distribution of substitutions within the gene per sample, C) The frequency by which the substitutions occurred in the gene, D) The distribution of indels in the gene, E) The impact of mutation on ZNF75O gene.

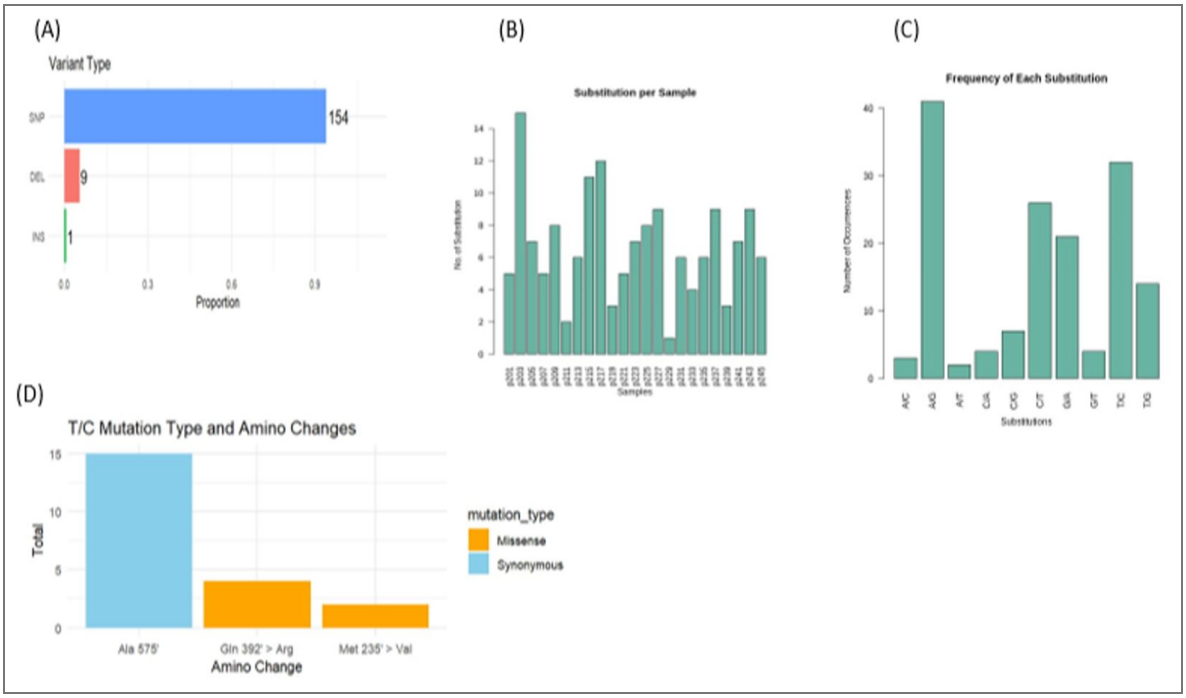


Table 3.

Sample ID	Position	Mutation	Type	Impact	Amino acid
217	82,831,023	C/T	Missense	MODERATE	Val 478'> Ile
231	82,830,234	C/T	Missense	MODERATE	Asp 694' > Asn
241	82,831,982	C/T	Missense	MODERATE	Arg 158' Gln
243	82,831,023	C/T	Missense	MODERATE	Val 478' Ile

Table 4.

Sample ID	Position	Mutation	Type	Impact	Amino acid
203	82,830,589	T/C	Synonymous	LOW	Ala 575'
	82,831,280	T/C	Missense	MODERATE	Gln 392' > Arg
205	82,830,589	T/C	Synonymous	LOW	Ala 575'
207	82,830,589	T/C	Synonymous	LOW	Ala 575'
215	82,830,589	T/C	Synonymous	LOW	Ala 575'
	82,831,752	T/C	Missense	MODERATE	Met 235' >Val
217	82,830,589	T/C	Synonymous	LOW	Ala 575'
	82,831,280	T/C	Missense	MODERATE	Gln 392' > Arg
221	82,830,589	T/C	Synonymous	LOW	Ala 575'
225	82,830,589	T/C	Synonymous	LOW	Ala 575'
227	82,830,589	T/C	Synonymous	LOW	Ala 575'
	82,830,589	T/C	Synonymous	LOW	Ala 575'
237	82,830,589	T/C	Synonymous	LOW	Ala 575'
	82,831,752	T/C	Missense	MODERATE	Met 235' > Val
	82,832,178	T/C	Missense	MODERATE	Asn 93' Asp
239	82,830,589	T/C	Synonymous	LOW	Ala 575'
241	82,830,589	T/C	Synonymous	LOW	Ala 575'
	82,831,280	T/C	Missense	MODERATE	Gln 392' > Arg
243	82,830,589	T/C	Synonymous	LOW	Ala 575'
	82,831,280	T/C	Missense	MODERATE	Gln 392' Arg
245	82,830,589	T/C	Synonymous	LOW	Ala 575'

Table 5.

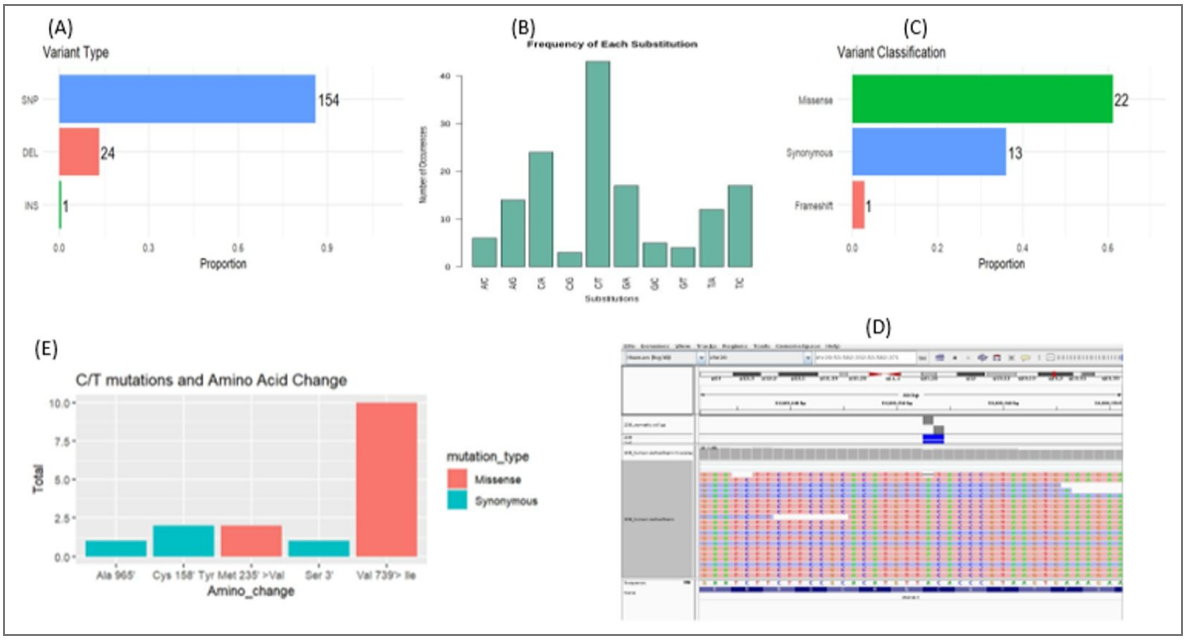
Sample ID	Position	Mutation	Type	Impact	Amino Acid
201	82,831,393	G/A	Synonymous	LOW	Thr 354'
205	82,832,218	G/A	Synonymous	LOW	Pro 566'
217	82,832,090	G/A	Synonymous	LOW	Val 165'
233	82,831,393	G/A	Synonymous	LOW	Thr 354'
	82,831,429	G/A	Synonymous	LOW	Thr 342'
243	82,831,393	G/A	Synonymous	LOW	Thr 354'

Table 6.

Sample ID	Position	Mutation	Type	Impact	Amine Acid
203	82,830,616	A/G	Synonymous	LOW	Pro 566'
205	82,830,616	A/G	Synonymous	LOW	Pro 566'
207	82,830,616	A/G	Synonymous	LOW	Pro 566'
215	82,830,616	A/G	Synonymous	LOW	Pro 566'
217	82,830,616	A/G	Synonymous	LOW	Pro 566'
221	82,830,616	A/G	Synonymous	LOW	Pro 566'
227	82,830,616	A/G	Synonymous	LOW	Pro 566'
237	82,830,616	A/G	Synonymous	LOW	Pro 566'
239	82,830,616	A/G	Synonymous	LOW	Pro 566'
241	82,830,616	A/G	Synonymous	LOW	Pro 566'
243	82,830,616	A/G	Synonymous	LOW	Pro 566'
245	82,830,616	A/G	Synonymous	LOW	Pro 566'

Figure 3. Mutation profile of ZNF217 gene in 23 patients.

A) Type of variants that occurred in the gene, B) Distribution of substitutions within the gene per sample, C) The frequency by which the substitutions occurred in the gene, D) Variant classification of all the mutations, E) The type of mutations causes by C/T substitution, and F) The frameshift mutation of TA T substituted by a termination codon (Ter) as viewed under IGV.



base coding (Figure 3E). Two analyzed samples showed a C/T base substitution at position 53,582,354 that equated to a synonymous substitution of Cys to Tyr at position 158 (Figure 3E and Table 7).

The nucleotide mutation C/A at position 53,582,438 within sample 205 caused a phenotypic variation where cysteine (Cys) at position 130 was mutated into phenylalanine (Phe) (Table 8). Through its synonymous mutation at position 53,582,080, Sample 217 maintains an identical amino acid sequence for alanine located at position 249 according to Table 8. A C/A substitution found at position 53,582,400 in sample 229 produces a missense mutation that changes amino acid valine at position 143 into phenylalanine (Table 8).

At nucleotide position 53,576,873, the C/A substitution leads to a missense mutation that switches alanine to serine at position 631 (Table 8). Two C/A mutations occur in Sample 245, where the first substitution at position 53,577,058 leads to Val in position 569, and another mutation at position 53,582,169 converts Val to Phe found at position 220 (Table 8). The position 53,582,352 TA/T mutation produces a frameshift type mutation (Figure 4E). The frameshift mutation leads to a premature translation termination because it replaces position 158 cysteine (Cys) with a termination codon (Ter) (Table 9).

The research team evaluated the mutation rates between the three genes while comparing results with HER2 status. Researchers applied two-sample t-tests to evaluate significance in the data relationship. Mutation counts of the ZNF217 gene displayed similar median counts between HER2-positive cases and those that were HER2-negative during analysis. However, the distribution of mutation levels was more variable in HER2-positive samples. Results from the t-test yielded no remarkable statistical difference ($t = -0.022$, $df = 21$, $p = 0.9825$) as shown in Figure 4A. The relationship between ZNF703 mutation counts and HER2 status failed to show any statistical significance ($t = -0.587$, $df = 21$, $p = 0.5637$) at the same time that the data exhibited elevated variability in HER2-positive samples. Mutations of ZNF750 produced identical distribution patterns between HER2-positive and HER2-negative groups since their interquartile ranges cross, and statistical tests showed no significant differences ($t = -0.587$, $df = 21$, $p = 0.5637$) (Figure 4C). Data shows that HER2 status has no impact on the mutation burdens found for the three genes ZNF217, ZNF703, and ZNF750 within this patient cohort.

Analysis of the three gene mutations showed no significant relationship with cancer staging. ANOVA statistical analyses established the evaluation of significance levels between Stage 1, 2 and 3 data points. Stage 2 exhibited minimal variations in total mutations but statistical analysis showed no significant difference ($p > 0.32$ for all stage comparisons) compared to other stages (Figure 5A). ZNF703 mutation counts demonstrated an upward trend between stage 1 and 3 according to ANOVA analysis results but there was no statistical significance ($p > 0.35$ in all pairwise tests). ZNF750 mutation data displayed an upward trend (Figure 5C) despite significant range association yet demonstrated no statistically meaningful differences between categories ($p > 0.35$ for all analyses). The stage of cancer does not create a clear link between mutation counts for ZNF217, ZNF703 or ZNF750.

Expression Profile for 23 Patients

Analysis of the three ZNF gene samples for the 23 paired breast tissues was conducted in R by using Principal Component Analysis (PCA). Three ZNF gene cluster analyses for 23 paired breast specimen samples occurred within R through the utilization of Principal Component Analysis (PCA) methods (Figure 6). The normal tissue group aligned closely in the cluster indicating uniform expression levels but the tumor group spread out widely, consistent with differences between various cancer stages (Figure 6). The Stage 1 tumors exhibited blue clusters that stayed compressed together while Stage 2 and Stage 3 tumors expanded in dispersion revealing elevated progression-related heterogeneity (Figure 6). The two principal components account for 58.46% and 24.77% of the variance independently so specific changes in gene expression in tumors explain most of the profile variability (Figure 6).

Table 7.

Sample ID	Position	Mutation	Type	Impact	Amino acid
201	53,576,549	C/T	Missense	MODERATE	Val 739' > Ile
207	53,576,549	C/T	Missense	MODERATE	Val 739' > Ile
209	53,576,549	C/T	Missense	MODERATE	Val 739' > Ile
213	53,575,869	C/T	Synonymous	LOW	Ala 965'
	53,576,549	C/T	Missense	MODERATE	Val 739' > Ile
215	53,576,549	C/T	Missense	MODERATE	Val 739' > Ile
217	53,576,549	C/T	Missense	MODERATE	Val 739' > Ile
	53,582,818	C/T	Synonymous	LOW	Ser 3'
223	53,576,549	C/T	Missense	MODERATE	Val 739' > Ile
225	53,576,549	C/T	Missense	MODERATE	Val 739' > Ile
227	53,576,549	C/T	Missense	MODERATE	Val 739' > Ile
235	53,576,549	C/T	Missense	MODERATE	Val 739' > Ile
239	53,582,354	C/T	Missense	MODERATE	Cys 158' Tyr
245	53,576,549	C/T	Missense	MODERATE	Cys 158' Tyr

Table 8.

Sample ID	Position	Mutation	Type	Impact	Amino acid
205	53,582,438	C/A	Missense	MODERATE	Cys 130' > Phe
217	53,582,080	C/A	Synonymous	LOW	Ala 249'
229	53,582,400	C/A	Missense	MODERATE	Val 143' > Phe
239	53,576,873	C/A	Missense	MODERATE	Ala 631' >Ser
241	53,582,080	C/A	Synonymous	LOW	Ala 249'
245	53,577,058	C/A	Missense	MODERATE	Gly 569' >Val
	53,582,169	C/A	Missense	MODERATE	Val 220' >Phe

Table 9.

Sample ID	Position	Mutation	Type	Impact	Amino acid
239	53,582,352	TA/T	Frameshift	High	Cys 158' Ter

Figure 4. The association between ZNF mutations and HER2 status.

A) The distribution of total ZNF217 mutations stratified by HER2 status; no significance difference (t= −0.022, df = 21. P= 0.9825), B) The distribution of ZNF7O3 mutations stratified by HER2 status; no statistical difference (t = −2.45, df = 21, p = 0.0229), C) The distribution of ZNF750 mutations stratified by HER2 status, no statistical difference (t= −0.587, df = 21, p= 0.5637).

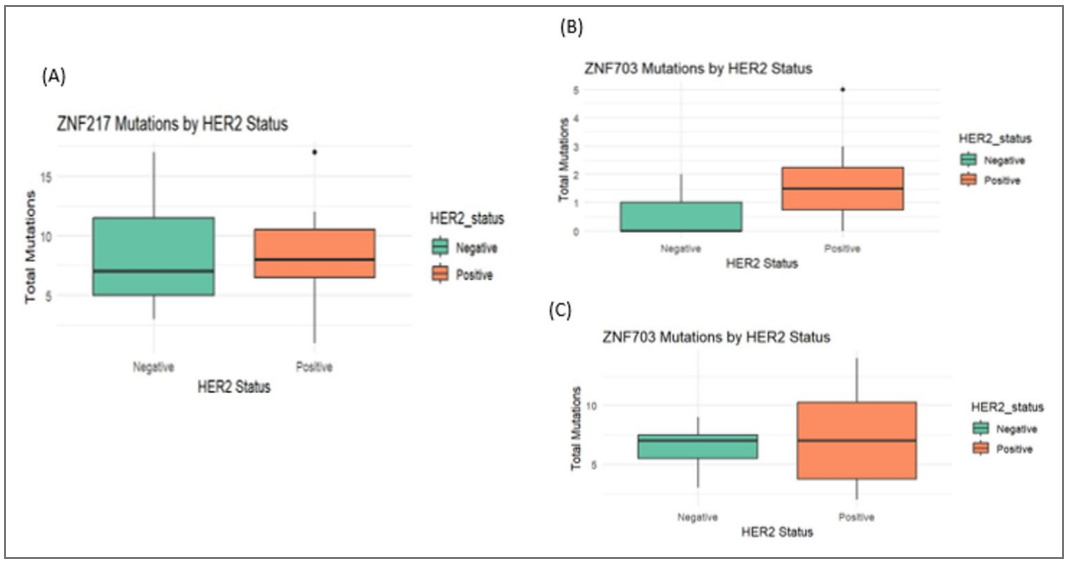
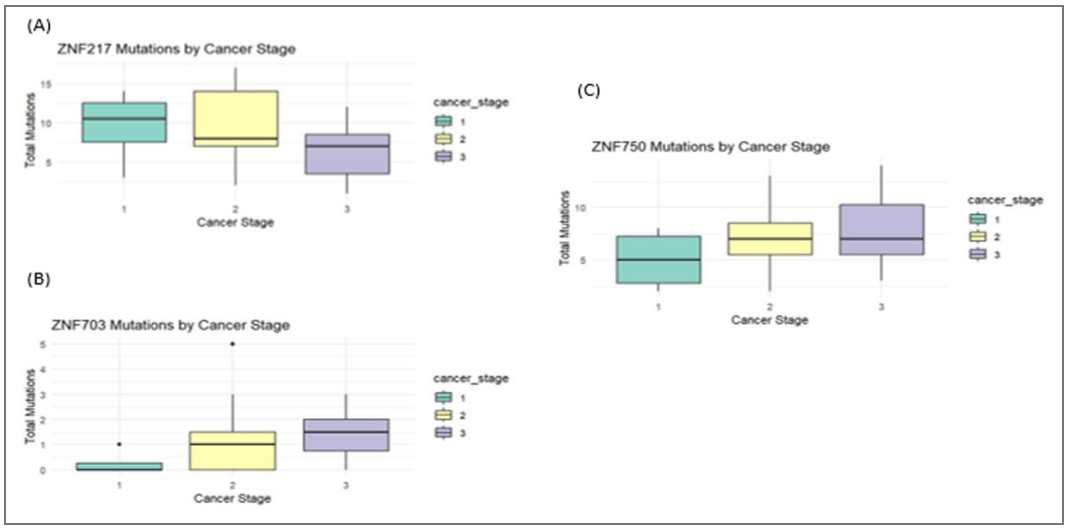


Figure 5. The association between ZNF mutations and Cancer Stage.

A) The distribution of total ZNF217 mutations across cancer stages 1, 2, and 3; ANOVAtest, no significant differences between the groups (stage 2 vs. stage 1: p = 0.999; stage 3 vs. stage 1: p = 0.539; stage 3 vs. stage 2: p = 0.321), B) The distribution of ZNF7O3 mutations across cancer stages 1, 2, and 3; ANOVAtest, no statistically significant differences between any pair of cancer stages (stage 2 vs. stage 1: p = 0.456; stage 3 vs. stage 1: p = 0.358; stage 3 vs. stage 2: p = 0.946), C) The distribution of ZNF75O mutations across cancer stages 1, 2, and 3, ANOVA test, (stage 2 vs. stage 1: p = 0.507; stage 3 vs. stage 1: p = 0.351; stage 3 vs. stage 2: p = 0.896).



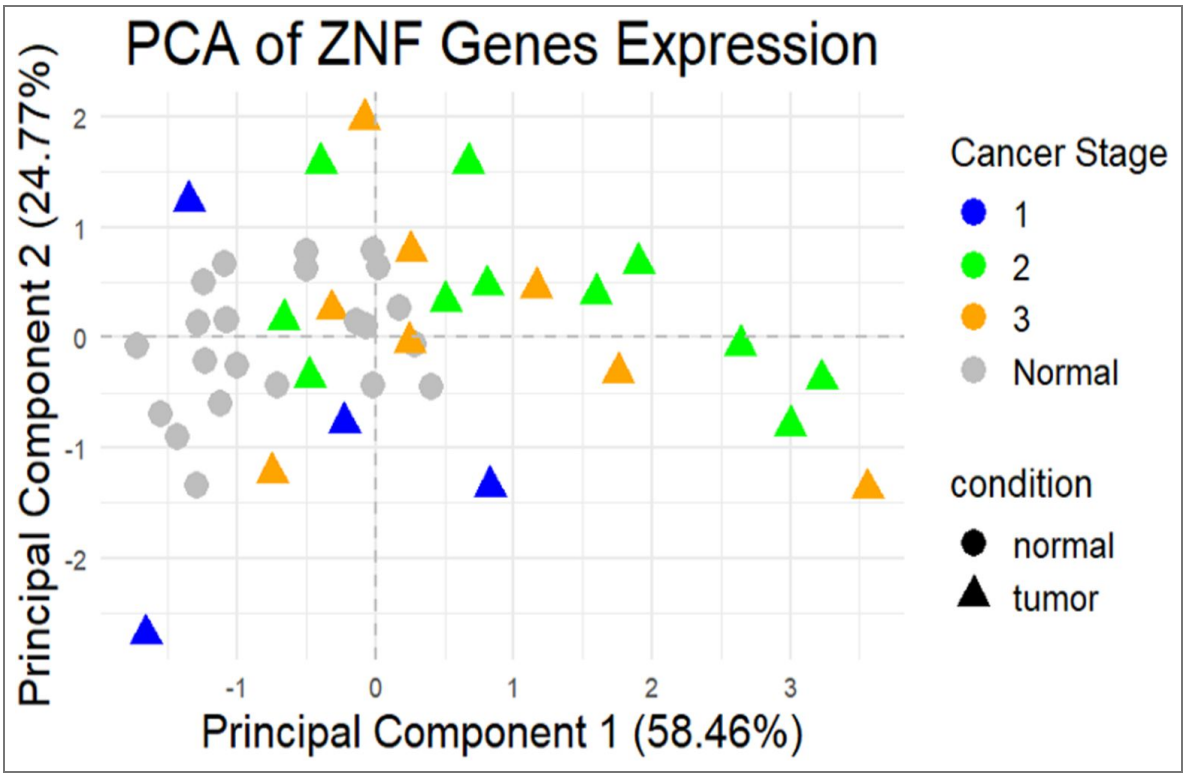


Figure 6. Principal Component Analysis (PCA) of ZNF217, ZNF7O3, and ZNF75O expression in 23 paired breast tissues.

Normal samples (gray) cluster tightly, while tumor samples show increasing dispersion across stages (Stage 1: blue, Stage 2: green, Stage 3: orange), reflecting greater variability with progression. The first two components explain 58.46% and 24.77% of the variance, driven by tumor-specific changes in expression.

The evaluation of differentially-expressed genes (DGE) occurred with normalized data. The research analysis indicated that all three target genes displayed important expression intensities ($P < 0.05$) in examined tumor tissues. The experiment results confirmed ZNF217 genes as significantly active ($P = 0.0004$) while ZNF750 genes showed significant expression ($p = 0.001$) and ZNF703 genes also exhibited significant expression (0.002) in tumor samples (Figure 7).

The research evaluated ZNF750 gene expression based on patient cancer progression levels. The expression levels of this gene were enhanced in tumor samples obtained from patients in stages 2 and 3. Analysis of stage 1 tissue samples yielded five results which two of them indicated elevated ZNF750 gene activity in healthy tissues. Some patients diagnosed with stage 1 breast cancer show elevated ZNF750 gene expression in their healthy breast tissue which suggests the gene might play a regulatory part during the first stages of cancer advancement (Figure 8A). The results from the Kruskal-Wallis test showed no significant differences between cancer stages based on $\chi^2 = 5.6918$ ($df = 2$) at $p = 0.05808$. This study examined how the gene expression levels changed when considering both HER2 protein levels and cancer development stage. The analysis showed HER2-positive stage 2 patients had the highest ZNF750 expression levels yet stage 3 showed higher expression levels in HER2-negative patients compared to HER2-positive patients (Figure 8B). The stage 1 results demonstrated that HER2-positive breast tumors demonstrated greater gene expression than their HER2-negative counterparts (Figure 8B). The results from paired t-test revealed no significant statistical difference between the two groups ($t = 0.18044$, $df = 20.998$, $p = 0.8585$) when comparing HER2-positive and HER2-negative expression. Stage 2 cancer cells with elevated HER2 expression likely support tumor progression during intermediate stages but stage 3 cancer cells with HER2-negative status may compensate their more advanced condition through increased gene expression.

Researchers evaluated ZNF703 gene expression data to match it against the cancer stage evaluations of the specimens under study. The gene exhibited elevated expression levels throughout tumor tissue at stages 2 and 3 based on Figure 9A. Within stage 1 and the five analyzed samples two demonstrated elevated ZNF703 gene expression at the normal tissue level (Figure 9A). The expression level of ZNF703 showed no statistical difference between different cancer stages based on the Kruskal-Wallis test ($\chi^2 = 1.2261$, $df = 2$, $p = 0.5417$). Analysis of ZNF703 gene expression took place based on HER2 status and cancer stage distribution. In stages 1 and 3 ZNF703 showed its highest expression in HER2-negative samples than in HER2-positive samples (Figure 9B). According to a paired t-test analysis the expression levels between HER2-positive and HER2-negative groups matched up with no significant statistical difference at $t = 0.36584$, $df = 18.554$, and $p = 0.7186$. The ZNF703 gene shows higher importance for HER2-negative breast cancer development in stages one and three along with advanced cases because it promotes tumor progression beyond HER2 signaling mechanisms.

Most research samples documented enhanced ZNF217 gene expression levels in normal tissues with five specimens showing increased gene activity specifically in normal tissue (Figure 10A). Stage 1 saw three out of five cases where the gene had elevated expression levels in normal breast tissue according to Figure 10A. The Kruskal-Wallis analysis demonstrated that cancer stages did not affect ZNF217 expression levels because the results indicated $\chi^2 = 2.9136$, $df = 2$, $p = 0.233$. Analysis of ZNF217 gene expression involved evaluation through normal and HER2 status comparison across different cancer stages. The gene expression levels in stage 1 and 2 samples with HER2 positivity outpaced HER2-negative results (Figure 10B). The analysis of Figure 10B revealed greater ZNF217 gene expression in cancer samples that were HER2-negative during stage 3. A paired t-test analysis revealed no significant difference in expression patterns between HER2-positive and HER2-negative groups ($t = -0.62653$, $df = 20.973$, $p = 0.5377$). The ZNF217 gene functions differently according to HER2 status when it accelerates early-stage HER2-positive cancer progression before taking precedence in late-stage HER2-negative cancer development.

Gene regulation patterns follow distinct routes based on the effects of genetic mutations on ZNF217 together with ZNF750 and ZNF703. Structural modification from frameshift mutations linked to ZNF217 expression reductions yet these defects have probable functional failures (Figure 11A). Conversely, missense mutations had positive correlations with ZNF217 expression because

Figure 7. DGE for ZNF75O, ZNF217, and ZNF7O3 genes.

The significant expression ($p < 0.05$) is normalized for each gene in both conditions, with normal samples represented in blue and tumor samples in red.

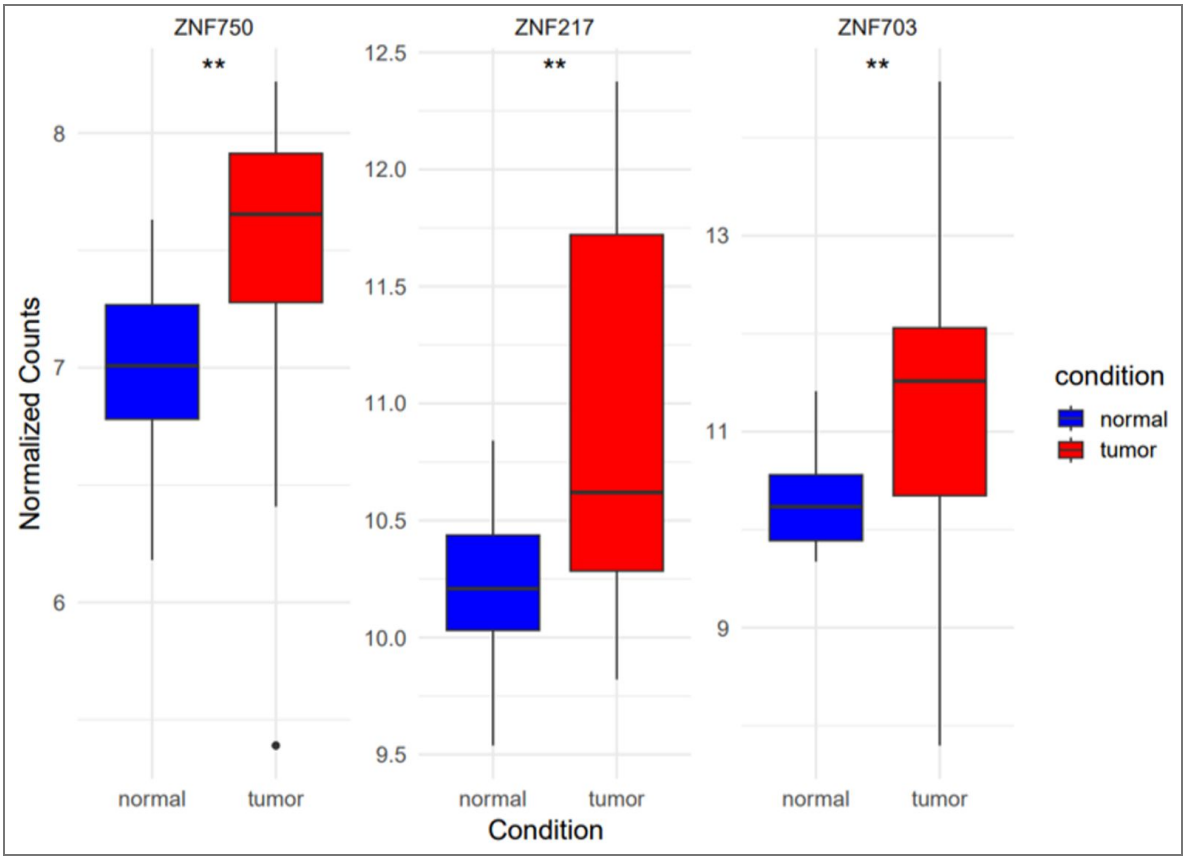


Figure 8.

A) The distribution of RNA-Seq expression data for ZNF75O genes based on cancer stage. Each dot represents the difference in expression between normal and tumor sample, those above the gradient line represents ZNF75O genes highly expressed in tumor samples for the 23 patients; those below the line represents the gene highly expressed in normal sample. B) The box plot showing Log2 fold change (>2 or <l/2) of expression counts of ZNF75O gene per HER2 status and cancer stage of the 23 patients.

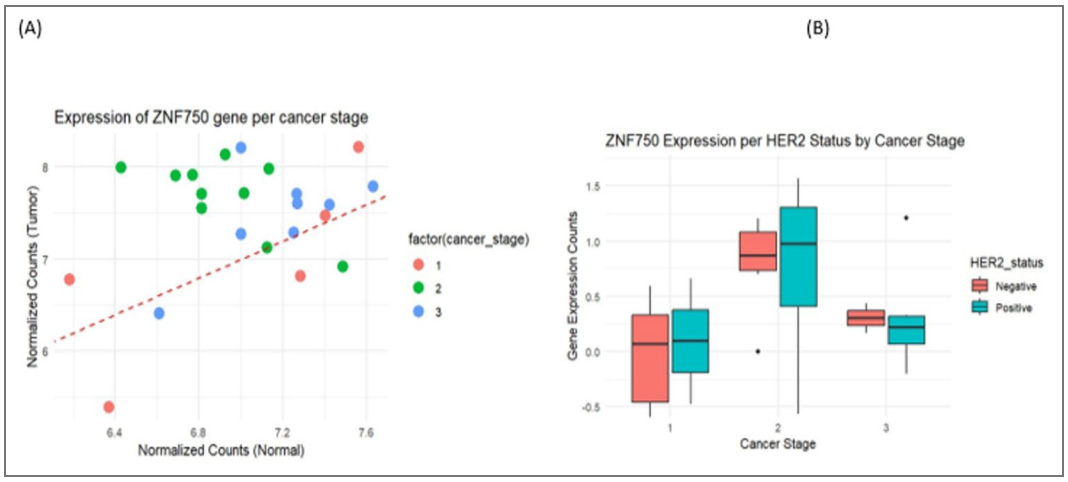


Figure 9.

A) The distribution of RNA-Seq expression data for ZNF7O3 genes based on cancer stage. Each dot represents the difference in expression between normal and tumor sample, those above the gradient line represents ZNF75O genes highly expressed in tumor samples for the 23 patients. B) Log2 fold change (>2 or <1/2) of expression counts of ZNF7O3 gene per HER2 status and cancer stage of the 23 patients.

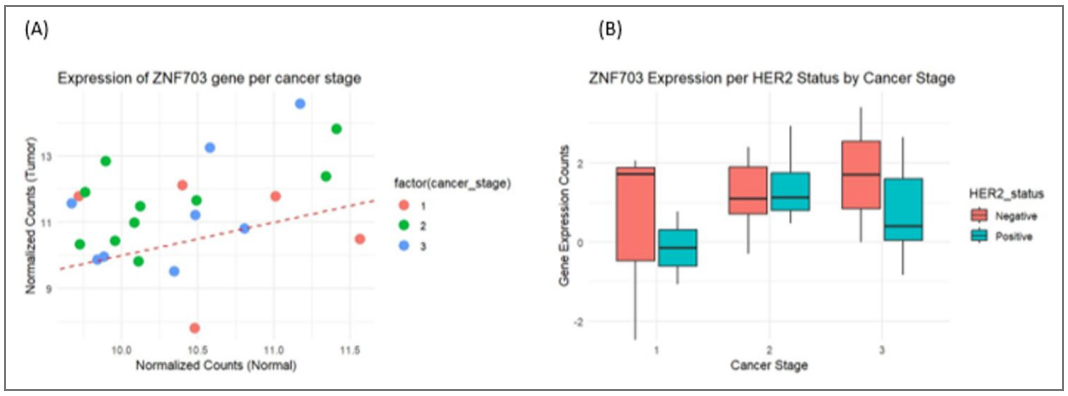
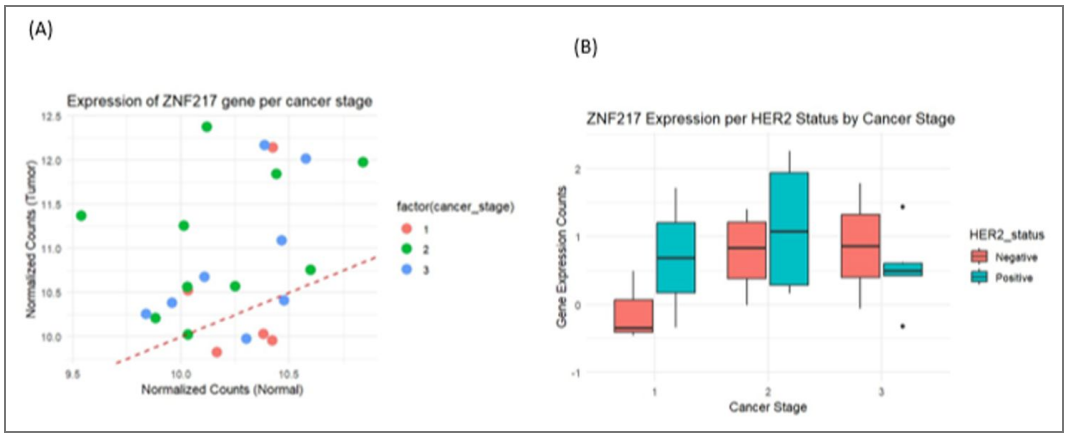


Figure 10.

A) The distribution of RNA-Seq expression data for ZNF217 genes based on cancer stage. Each dot represents the difference in expression between normal and tumor sample, those above the gradient line represents ZNF75O genes highly expressed in tumor samples for the 23 patients. B) Log2 fold change (>2 or <l/2) of expression counts of ZNF217 gene per HER2 status and cancer stage of the 23 patients..



they facilitated this elevation. Total SNPs in ZNF703 demonstrated robust positive association with expression alterations and missense mutations showed moderate contribution to expression elevation through known oncogenic functions (Figure 11B). Total SNP burden in ZNF750 showed a direct correlation with gene expression variations because elevated mutational load powerfully manipulates regulatory patterns (Figure 11C). The analysis showed synonymous mutations presented minimal effect on gene expression across all sequences in the three genes studied. Breast cancer gene expression differences primarily result from functional mutations that lead to missense alterations yet frameshift mutations can reduce expression levels when they affect protein stability.

Discussion

The research analyzed the ZNF217, ZNF750, and ZNF703 genomic alterations and mRNA levels within 23 Kenyan breast cancer patients. The examination revealed 359 genetic alterations that primarily exhibited C>T nucleotide changes in each assessed gene. The transitions identified through genomic analysis become essential in cancer research because Alexandrov et al. (2020) explains they commonly appear as DNA damage markers from oxidative stress in cancerous cells. Tissues from tumor areas demonstrated substantial elevation of ZNF217 together with ZNF703 and ZNF750 transcriptional levels compared to normal tissues ($P<0.05$), yet ZNF217 exhibited the greatest detected difference. The Principal Component Analysis (PCA) data effectively distinguished tumor sample data from normal sample data. The first two principal components in the PCA plot explain 58.46% and 24.77% of the variance which evidences that tumor-specific alterations in the three gene expressions drive most variations in measurement profiles (Gitau et al., 2024). The expression patterns differed according to HER2 status together with cancer stage progression indicating contextual roles for each gene in tumor developments. Studies showed that ZNF750 expression levels rose progressively from stage 2 through stage 3 tumor samples yet certain stage 1 specimens displayed elevated expression in healthy tissues thus indicating that ZNF750 works as an early regulatory factor (Gitau et al., 2024).

ZNF217 mutations included an in-frame GAGAC/C deletion which presented a strong potential to disrupt protein structure. An in-frame GAGAC/C deletion in the gene sequence is relevant because it may disrupt protein functioning which can advance breast cancer through modified EMT and TGF-β signaling mechanisms (Vendrell et al., 2012). A frameshift mutation of TA/T during analysis for sample 239 transformed the amino acid cysteine at position 158 into a termination codon (Ter) leading to early translation termination and functional damage to its cell cycle and apoptosis regulatory actions. Evidences indicate truncating ZNF217 mutations can cause deregulated cell proliferation through destruction of its gene-controlling abilities on apoptosis regulation (Krig et al., 2010b). Our research supports data showing how the frameshift mutation might push cells toward malignancy because it abets uncontrolled cell growth while demonstrating the importance of this mechanism due to ZNF217 serving as an ErbB3 pathway activator which commonly plays a role in breast cancer development (Beroukhim et al., 2007). The ZNF217 gene showed multiple missense mutations which contained exchanges at cysteine (C) 158 position that substituted the amino acid with tyrosine (Y). The conversion of a polar amino acid into a nonpolar type leads to substantial structural modifications of ZNF217 protein that could modify its DNA-binding features essential for regulatory activity (Briest et al., 2023). The aggressive cancer phenotype may result from this breast cancer-specific mutation of ZNF217 because ZNF217 frequently shows amplification in breast cancer tissues (Vendrell et al., 2012). Research results indicated that ZNF217 expression levels were elevated in most examined normal tissue samples where five samples displayed quantitatively higher ZNF217 protein levels. The research indicates ZNF217 regulates early cancer development stages before oncogenesis completely manifests (Tang et al., 2023).

A frameshift mutation found in ZNF703 displayed an alanine deletion at position 514 which was an in-frame variant. The scientific literature shows that ZNF703 frameshift mutations produce increased cell proliferation in breast cancer by modifying MAPK and PI3K/AKT signal pathways (X. Zhang et al., 2013). ZNF703 overexpression links to higher proliferation rates in HER2-negative

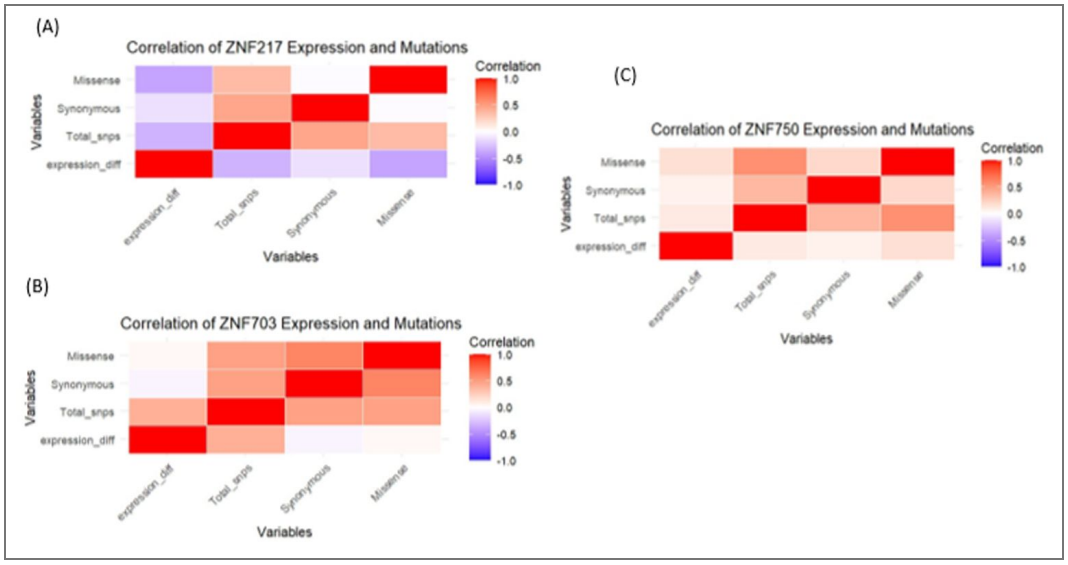


Figure 11. The correlation analysis between ZNF mutations and expression differences.

A) The correlation analysis between ZNF217 mutations and expression differences, B) The correlation analysis for the ZNF7O3 gene demonstrated notable associations, C) A correlation analysis between ZNF75O expression differences and mutation.

specimens because such mutations prove significant for both moderate-impact and HER2-negative cases (X. Zhang et al., 2022). The altered ZNF703 gene demonstrates how it promotes aggressive cancer growth in HER2-negative tumors which makes it a possible therapeutic target for those breast cancers. Twelve breast cancer cases provided evidence of increased ZNF703 expression in HER2-negative tumors across stage 1 and stage 3 samples showing its role in tumor expansion that bypasses HER2 pathways (Stradella et al., 2022).

Scientists believe the ZNF750 gene mutation profile acknowledges missense mutations affecting tumor-suppressive functions through DNA-binding domain amino acid substitutions at position 478 from valine (V) to isoleucine (I) (Butera et al., 2020). The mutations found in ZNF750 are highly important for triple-negative breast cancers since the loss of ZNF750 function remains linked to progressive disease features (Cassandri et al., 2020). The study demonstrates that ZNF750 mutations stand out as a biomarker that indicates aggressive breast cancer types within African populations along with their distinct mutational patterns compared to other research populations (Brewster et al., 2014). The different types of mutations found within ZNF217 ZNF703 and ZNF750 gene sequences within the Kenyan cohort seem to result from the combination of genetic and environmental elements that affect how breast cancer evolves in Africans. ZNF750 demonstrates increased expression specifically in HER2-positive stage 2 tumors according to expression analysis findings (Klæstad et al., 2021). Its oncogenic role seems to play a greater part in HER2-negative tumors that appear later in progression thus suggesting a compensatory mechanism.

Data availability

RNA sequence data were deposited in GEO (accession number GSE225846). All data generated or analyzed during this study are included in the manuscript and supporting files; source data files have been provided for all figures.

Additional information

Funding

Funder	Grant reference number	Author
Mount Kenya University (MKU)		Florence Ngonga Francis Makokha Michael Kitoi John Gitau Godfrey Kinyori Kennedy Mwangi

Author ORCID iDs

Michael Kitoi: https://orcid.org/0009-0007-8098-8547

References

Aganezov S., Yan S. M., Soto D. C., Kirsche M., Zarate S., Avdeyev P., Taylor D. J., Shafin K., Shumate A., Xiao C., *et al.* (2022) A complete reference genome improves analysis of human genetic variation. *Science* **376**:eabl3533 https://doi.org/10.1126/science.abl3533 | PubMed

Alessio B., Matteo C., Rugolo F., Massimiliano A., Gerry M (2020) The ZNF750–RAC1 axis as potential prognostic factor for breast cancer. *Cell Death Discovery* **6** https://doi.org/10.1038/s41420-020-00371-2 | PubMed

Alexandrov L. B., Kim J., Haradhvala N. J., Huang M. N., Tian Ng A. W., Wu Y., Boot A., Covington K. R., Gordenin D. A., Bergstrom E. N., *et al.* (2020) The repertoire of mutational signatures in human cancer. *Nature* **578**:94-101 https://doi.org/10.1038/s41586-020-1943-3 | PubMed

Ansari-Pour N., Zheng Y., Yoshimatsu T. F., Sanni A., Ajani M., Reynier J.-B., Tapinos A., Pitt J. J., Dentro S., Woodard A., *et al.* (2021) Whole-genome analysis of Nigerian patients with breast cancer reveals ethnic-driven somatic evolution and distinct genomic subtypes. *Nature Communications* **12**:6946 https://doi.org/10.1038/s41467-021-27079-w | PubMed

Bayat A., Gaëta B., Ignjatovic A., Parameswaran S (2017) Improved VCF normalization for accurate VCF comparison. *Bioinformatics* **33**:964-970 https://doi.org/10.1093/bioinformatics/btw748 | PubMed

Bray F., Laversanne M., Sung H., Ferlay J., Siegel R. L., Soerjomataram I., Jemal A. (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **74**:229-263 https://doi.org/10.3322/caac.21834 | PubMed

Brewster A. M., Chavez-MacGregor M., Brown P (2014) Epidemiology, biology, and treatment of triple-negative breast cancer in women of African ancestry. *The Lancet Oncology* **15**:e625-e634 https://doi.org/10.1016/S1470-2045(14)70364-X | PubMed

Butera A., Cassandri M., Rugolo F., Agostini M., Melino G (2020) The ZNF750–RAC1 axis as potential prognostic factor for breast cancer. *Cell Death Discovery* **6**:1-9 https://doi.org/10.1038/s41420-020-00371-2 | PubMed

Cassandri M., Butera A., Amelio I., Lena A. M., Montanaro M., Mauriello A., Anemona L., Candi E., Knight R. A., Agostini M., *et al.* (2020) ZNF750 represses breast cancer invasion via epigenetic control of prometastatic genes. *Oncogene* **39**:4331-4343 https://doi.org/10.1038/s41388-020-1277-5 | PubMed

Chen Z., Yuan Y., Chen X., Chen J., Lin S., Li X., Du H (2020) Systematic comparison of somatic variant calling performance among different sequencing depth and mutation frequency. *Scientific Reports* **10**:3501 https://doi.org/10.1038/s41598-020-60559-5 | PubMed

Ewels P., Magnusson M., Lundin S., Käller M (2016) MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**:3047-3048 https://doi.org/10.1093/bioinformatics/btw354 | PubMed

Gitau J., Kinyori G., Sayed S., Saleem M., Makokha F. W., Kirabo A (2024) The Association between the JAK-STAT Pathway and Hypertension among Kenyan Women Diagnosed with Breast Cancer. *bioRxiv* https://doi.org/10.1101/2024.06.07.597892 | PubMed

GLOBOCAN (2022) Global cancer statistics 2022-GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* https://doi.org/10.3322/caac.21834 | PubMed

Greenacre M., Groenen P. J. F., Hastie T., D'Enza A. I., Markos A., Tuzhilina E (2022) Principal component analysis. *Nature Reviews Methods Primers* **2**:1-21 https://doi.org/10.1038/s43586-022-00184-w

Klæstad E., Sawicka J. E., Engstrøm M. J., Ytterhus B., Valla M., Bofin A. M (2021) ZNF703 gene copy number and protein expression in breast cancer; associations with proliferation, prognosis and luminal subtypes. *Breast Cancer Research and Treatment* **186**:65-77 https://doi.org/10.1007/s10549-020-06035-0 | PubMed

Krig S. R., Miller J. K., Frietze S., Beckett L. A., Neve R. M., Farnham P. J., Yaswen P. I., Sweeney C. A (2010a) ZNF217, a candidate breast cancer oncogene amplified at 20q13, regulates expression of the ErbB3 receptor tyrosine kinase in breast cancer cells. *Oncogene* **29**:5500-5510 https://doi.org/10.1038/onc.2010.289 | PubMed

Krig S. R., Miller J. K., Frietze S., Beckett L. A., Neve R. M., Farnham P. J., Yaswen P. I., Sweeney C. A (2010b) ZNF217, a candidate breast cancer oncogene amplified at 20q13, regulates expression of the ErbB3 receptor tyrosine kinase in breast cancer cells. *Oncogene* **29**:5500-5510 https://doi.org/10.1038/onc.2010.289 | PubMed

Liao Y., Smyth G. K., Shi W (2014) featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**:923-930 https://doi.org/10.1093/bioinformatics/btt656 | PubMed

Malick S. B. A., Conteh F., Sawo M. (2022) Bioinformatics Analysis of Differentially Expressed Gene's in Breast Cancer Using DESeq2. Islamic University of Technology.

Miller J. K., Farnham P. J., Yaswen P., Sweeney C (2010) ZNF217, a candidate breast cancer oncogene amplified at 20q13, regulates expression of the ErbB3 receptor tyrosine kinase in breast cancer cells. *Oncogene* **29**:5500-5510 https://doi.org/10.1038/onc.2010.289 | PubMed

O'Donovan P. (2019) Identification of driver mutations and tumour evolution in HER2 positive breast cancer. Royal College of Surgeons in Ireland.

Robinson J. T., Thorvaldsdóttir H., Wenger A. M., Zehir A., Mesirov J. P (2017) Variant Review with the Integrative Genomics Viewer. *Cancer Research* **77**:e31-e34 https://doi.org/10.1158/0008-5472.CAN-17-0337 | PubMed

Salz R., Saraiva-Agostinho N., Vorsteveld E., van der Made C. I., Kersten S., Stemerdink M., Allen J., Volders P.-J., Hunt S. E., Hoischen A., *et al.* (2023) SUsPECT: A pipeline for variant effect prediction based on custom long-read transcriptomes for improved clinical variant annotation. *BMC Genomics* **24**:305 https://doi.org/10.1186/s12864-023-09391-5 | PubMed

Stradella A., Gargallo P., Cejuela M., Petit A., Bosch-Schips J., Carbonell P., Recalde S., Vethencourt A., Fernandez-Ortega A., Falo C., *et al.* (2022) Genomic characterization and tumor evolution in paired samples of metaplastic breast carcinoma. *Modern Pathology* **35**:1066-1074 https://doi.org/10.1038/s41379-022-01017-7 | PubMed

Tang W., Zhang F., Byun J. S., Dorsey T. H., Yfantis H. G., Ajao A., Liu H., Pichardo M. S., Pichardo C. M., Harris A. R., *et al.* (2023) Population-specific Mutation Patterns in Breast Tumors from African American, European American, and Kenyan Patients. *Cancer Research Communications* **3**:2244-2255 https://doi.org/10.1158/2767-9764.CRC-23-0165 | PubMed

Vendrell J. A., Thollet A., Nguyen N. T., Ghayad S. E., Vinot S., Bièche I., Grisard E., Josserand V., Coll J.-L., Roux P., *et al.* (2012) ZNF217 Is a Marker of Poor Prognosis in Breast Cancer That Drives Epithelial–Mesenchymal Transition and Invasion. *Cancer Research* **72**:3593-3606 https://doi.org/10.1158/0008-5472.CAN-11-3095 | PubMed

Voutsadakis I. A (2020) 8p11.23 Amplification in Breast Cancer: Molecular Characteristics, Prognosis and Targeted Therapy. *Journal of Clinical Medicine* **9** https://doi.org/10.3390/jcm9103079 | PubMed

Wang S., Wang C., Liu O., Hu Y., Li X., Lin B (2022) miRNA-651-3p regulates EMT in ovarian cancer cells by targeting ZNF703 and via the MEK/ERK pathway. *Biochemical and Biophysical Research Communications* **619**:76-83 https://doi.org/10.1016/j.bbrc.2022.06.005 | PubMed

Wang Y., Ma C., Yang X., Gao J., Sun Z. (2024) ZNF217: An Oncogenic Transcription Factor and Potential Therapeutic Target for Multiple Human Cancers. *Cancer Management and Research* https://doi.org/10.2147/cmar.s431135 | PubMed

Zhang G., Wang J., Yang J., Li W., Deng Y., Li J., Huang J., Hu S., Zhang B (2015) Comparison and evaluation of two exome capture kits and sequencing platforms for variant calling. *BMC Genomics* **16**:581 https://doi.org/10.1186/s12864-015-1796-6 | PubMed

Zhang X., Mu X., Huang O., Wang Z., Chen J., Chen D., Wang G (2022) ZNF703 promotes triple-negative breast cancer cells through cell-cycle signaling and associated with poor prognosis. *BMC Cancer* **22**:226 https://doi.org/10.1186/s12885-022-09286-w | PubMed

Zhang X., Mu X., Huang O., Xie Z., Jiang M., Geng M., Shen K (2013) Luminal Breast Cancer Cell Lines Overexpressing ZNF703 Are Resistant to Tamoxifen through Activation of Akt/mTOR Signaling. *PLOS One* **8**:e72053 https://doi.org/10.1371/journal.pone.0072053 | PubMed

Tang W., Zhang F., Byun J. S., Dorsey T. H., Yfantis H. G., Ajao A., Liu H, Pichardo M. S., Pichardo C. M., Harris A. R., *et al.* (2023) High neighborhood deprivation impacts DNA methylation and gene expression in cancer-related genes (RNA-Seq). NCBI Gene Expression Omnibus. ID GSE225846 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE225846

Peer reviews

Reviewer #1 (Public review):

Summary:

This manuscript investigates mutations and expression patterns of zinc finger proteins in Kenyan breast cancer patients. Whole-exome sequencing and RNA-seq were performed on 23 breast cancer samples alongside matched normal tissues.

Strengths:

Whole-exome sequencing and RNA-seq were performed on 23 breast cancer samples alongside matched normal tissues in Kenyan breast cancer patients. The authors identified mutations in ZNF217, ZNF703, and ZNF750.

Weaknesses:

(1) Research scope:

The results primarily focus on mutations in ZNF217, ZNF703, and ZNF750, with limited correlation analyses between mutations and gene expression. The rationale for focusing only on these genes is unclear. Given the availability of large breast cancer cohorts such as TCGA and METABRIC, the authors should compare their mutation profiles with these datasets. Beyond European and U.S. cohorts, sequencing data from multiple countries, including a recent Nigerian breast cancer study (doi: 10.1038/s41467-021-27079-w), should also be considered. Since whole-exome sequencing was performed, it is unclear why only four genes were highlighted, and why comparisons to previous literature were not included.

(2) Language and Style Issues

There are many typos and clear errors in the main text (e.g. (ref)).

Additionally, several statements read unnaturally. For example:

"Investigators uncovered 170 mutations ..." should instead be phrased as "We identified 170 mutations ...."

"The research team ..." should be rephrased as "Our team ...."

(3) Methods and Data Analysis Details

The methods section is vague, with general descriptions rather than specific details of data processing and analysis. The authors should provide:

(a) Parameters used for trimming, mapping, and variant calling (rather than referencing another paper such as Tang et al. 2023).

(b) Statistical methods for somatic mutation/SNP detection.

(c) Details of RNA purification and RNA-seq library preparation.

Without these details, the reproducibility of the study is limited.

(4) Data Reporting

This study has the potential to provide a valuable resource for the field. However, data-sharing plans are unclear. The authors should:

a) Deposit sequencing data in a public repository.

b) Provide supplementary tables listing all detected mutations and all differentially expressed genes (DEGs).

c) Clarify whether raw or adjusted p-values were used for DEG analysis.

d) Perform DEG analyses stratified by breast cancer subtypes, since differential expression was observed by HER2 status, and some zinc finger proteins are known to be enriched in luminal subtypes.

(5) Mutation Analysis

Visualizations of mutation distribution across protein domains would greatly strengthen interpretation. Comparing mutation distribution and frequency with published datasets would also contextualize the findings.

Comments on revisions:

The revised manuscript hasn't addressed any of these concerns. Careful proofreading is recommended, even if the authors do not intend to make further modifications to the manuscript.

https://doi.org/10.7554/eLife.108076.2.sa3

Reviewer #2 (Public review):

Summary:

This work integrated the mutational landscape and expression profile of ZNF molecules in 23 Kenyan women with breast cancer.

Strengths:

The mutation landscape of ZNF217, ZNF703, and ZNF750 were comprehensively studied and correlate with tumor stage and HER2 status to highlight the clinical significance.

Weaknesses:

The current cohort size is relatively small to reach significant findings, and targeted exploration on ZNF family without emphasizing the reason or clinical significance hinders the overall significance of the entire work.

https://doi.org/10.7554/eLife.108076.2.sa2

Reviewer #3 (Public review):

Summary:

This revised study analyzes the somatic mutational profiles and transcriptomic expression of three zinc-finger genes (ZNF217, ZNF703, ZNF750) in 23 Kenyan women with breast cancer, using whole-exome sequencing and RNA-sequencing of paired tumor-normal tissues. A total of 358 somatic mutations were detected, and all three genes were significantly upregulated in tumors compared to normal tissues (ZNF217 showing the most prominent difference). Higher expression was observed in HER2-positive tumors, though mutation burden for each gene did not correlate significantly with HER2 status or cancer stage. The findings provide preliminary evidence for the idenfication of diagnostic/prognostic biomarkers or therapeutic targets in sub-Saharan African populations.

Strengths:

The study's key strengths lie in its focus on an underrepresented Kenyan cohort, addressing a critical gap in sub-Saharan African breast cancer genomic research. It integrates DNA-level mutation analysis with RNA-level expression data, leveraging standardized bioinformatics pipelines (e.g., Mutect2 for variant calling, DESeq2 for differential expression) and rigorous quality control to deliver detailed insights into mutation types, functional impacts, and amino acid changes. Additionally, it explores gene expression patterns across different cancer stages and HER2 status subgroups, generating targeted hypotheses for future validation and enhancing the reliability of its findings.

Weaknesses:

The author has enhanced the descriptive depth of the study by adding details on mutations, expression subgroup analyses, and functional annotations but has not addressed the core weaknesses of small cohort size and lack of functional validation. While the revised version is more comprehensive in cataloging molecular alterations, it remains confined to descriptive analysis, with no substantial improvement in the reliability or generalizability of its conclusions.

https://doi.org/10.7554/eLife.108076.2.sa1

Author response:

I have updated the manuscript by correcting errors.

https://doi.org/10.7554/eLife.108076.2.sa0