



OPEN RNA-seq analysis of wild-type and mutated *TBPL1* gene in breast cancer cells lines through CRISPR/Cas9 approach reveals novel molecular signatures

Rabia Mishal¹, Jorge Meléndez-Zajgla², Bertha Rueda-Zarazúa^{2,3},
María Luisa Labra-Barrios¹, Carlos Alberto Castañón-Sánchez^{4,9}, Salvador Uribe Carvajal⁵,
Laura Padierna-Mota⁶, José Manuel Hernández-Hernández¹, Gloria Leon-Avila⁷,
Armando Pérez Rangel¹, Edgar Hernández-Martínez¹, Erika Beatriz Angeles-Morales⁸,
Ibrahim Khalil Albalawi¹ & Juan Pedro Luna-Arias¹✉

Breast cancer is the leading cause of death among women globally. Several genes have been found to be transcriptionally dysregulated in cancer, according to recent studies. TATA-box binding protein (TBP) and its two paralogs, *TBPL1* and *TBPL2*, play roles in human transcription. The *TBPL1* gene is implicated in colorectal carcinomas by suppressing the expression of miR-18a. However, its function in breast cancer remains undisclosed. *TBPL1* is distantly related to TBP and possesses a 40% similarity with TBP's core domain. In the present study, we explored the potential role of the *TBPL1* gene in transcriptome regulation by knocking out the *TBPL1* gene through the CRISPR/Cas9 method. Following the knockout of the *TBPL1* gene, we examined the gene transcription patterns and compared them to wild-type cell lines. We observed disparate signatures of upregulated and downregulated genes in wild-type and mutated conditions. Healthy breast MCF-12F, and T47D, SKBR3, and MDA-MB-231 breast cancer cell lines were assessed, as these cancer cells exhibit overexpression of the *TBPL1* gene. Next-generation sequencing data revealed distinct marker genes regulated by the *TBPL1* gene and their potential involvement in cell migration, proliferation, anti-apoptosis, and metastasis. Additionally, we also discovered novel lncRNAs implicated in the transcriptome analysis of the *TBPL1* knocked-out gene. Our investigation indicated that this gene might affect varied stages of breast cancer cell lines' cellular properties, such as cell duplication, morphology, and growth. It might also contribute to tumor formation in more aggressive cell lines like MDA-MB-231 in vivo.

Keywords TBP family, Breast cancer cell lines, Gene expression profiles, CRISPR/Cas9, *TBPL1* knockout

¹Department of Cell Biology, Center for Research and Advanced Studies of the National Polytechnic Institute (Cinvestav), Av. IPN 2508, Col. San Pedro Zacatenco, Alcaldía Gustavo A. Madero, 07360 Mexico City, Mexico. ²Laboratory of Functional Genomics of Cancer, National Institute of Genomic Medicine (INMEGEN), Periferico Sur 4809, Arenal Tepepan, Tlalpan, 14610 Mexico City, Mexico. ³Biological Sciences Graduate Program, National Autonomous University of Mexico (UNAM), Mexico City, Mexico. ⁴Laboratory of Biomedical Research, Subdirección de Enseñanza e Investigación, Regional High Specialty Hospital of Oaxaca, C. Aldama s/n, Paraje "El Tule, 71294 San Bartolo Coyotepec, Oaxaca, Mexico. ⁵Department of Molecular Genetics, Institute of Cellular Physiology, National Autonomous University of Mexico (UNAM), Mexico City, Mexico. ⁶Laboratorios de Especialidades Inmunológicas (LEI), Av. Gran Canal s/n, Locales 3 y 4, Ampliación Casas Alemán, Alcaldía Gustavo A. Madero, 07580 Mexico City, Mexico. ⁷Departamento de Zoología, Escuela Nacional de Ciencias Biológicas del Instituto Politécnico Nacional (ENCB-IPN), Unidad Profesional Lázaro Cárdenas, Prolongación de Carpio y Plan de Ayala s/n, Col. Santo Tomás, Alcaldía Miguel Hidalgo Mexico City, 11340 Mexico City, Mexico. ⁸Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas del Instituto Politécnico Nacional (ENCB-IPN), Unidad Profesional Lázaro Cárdenas, Prolongación de Carpio y Plan de Ayala s/n, Col. Santo Tomás, Alcaldía Miguel Hidalgo Mexico City, 11340 Mexico City, Mexico. ⁹Deceased: Carlos Alberto Castañón-Sánchez. ✉email: jpluna@cinvestav.mx

Breast cancer is the leading cause of cancer-related deaths among women globally. Dysfunctional signaling pathways, genetic alterations in regulatory sequences or chromatin regulators, and changes in gene expression regulation at the transcription level are hallmarks of cancerous cells¹. In eukaryotes, RNA polymerase II (RNAP II) transcribes all protein-coding genes, as well as miRNAs and lncRNAs². The transcription initiation of protein-coding genes requires the basal transcription factors TFIID, TFIIA, TFIIB, TFIIE, and TFIIH. These factors are sequentially assembled on the core promoter to recruit the TFIIF-RNAP II complex, forming the transcription preinitiation complex (PIC)^{3,4}. The core promoter spans about 100 bp around the transcription start site (TSS) and includes motifs such as the TATA-box, BRE, Inr, and DPE⁵.

The TATA-box binding protein (TBP) plays a role in RNA polymerase I, II, and III transcription in eukaryotes⁶. TBP is the initial member of a compact eukaryotic family of TBP paralogs^{6,7}. TAFs (TBP-associated factors) along with TBP, assemble into the TFIID complex, which appears in various forms depending on the tissue and cell types⁸. TBP consists of two domains, the C-terminal (core) and the N-terminal domains. In particular, the C-terminal domain is highly conserved, features a saddle-shaped structure, and carries both a convex and concave surface⁹. The convex surface binds to the TATA-box through the minor groove utilizing an induced fit model⁹. Intriguingly, the N-terminal domain varies in sequence and size among species yet remains conserved among vertebrates¹⁰.

TBPL2 (TBP2 or TRF3) is the most recent and final member described in the TBP family^{11,12}. It has only been found in vertebrates and shows 95% similarity to TBP's core domain^{7,10}. It both stimulates basal transcription in TATA-box promoters and interacts with TFIIA and TFIIB in vitro^{7,10}. Some plants also have two *TBP* genes, very similar in sequence, including *Zea mays*¹² and *Arabidopsis thaliana*^{13,14}. *TBPL2* is expressed in oocytes, early embryos, and embryonic stem cells, with conserved patterns in frogs, zebrafish, and mice^{10,15}. The absence of TBP and the abundance of TBPL2 indicates that TBPL2 functions as a TBP replacement factor in several biological systems^{12,15–17}.

The TATA-box binding protein-like 1 (TBPL1, TLF, or TRF2) is a less conserved distant metazoan paralog^{10,18–22}. It showcases only 40% similarity with the TBP core domain²³. TBPL1 is unique in that it cannot bind directly to the TATA-box^{10,18–22}. However, it does interact with TFIIA and TFIIB^{10,18–23}. TBPL1 forms a complex with TFIIA in HeLa cells²⁴ but also interacts with the TFIIA-like factor (ALF), which is a germ cell-specific paralog of TFIIA²⁵. Notably, TBPL1 negatively modulates transcription by sequestering TFIIA^{26,27}. Therefore, the TBPL1 interaction with TFIIA activates the transcription of the TATA-less promoter of the *NF1* gene²⁶. TBPL1 inhibits transcription of TATA-box-containing promoters, while simultaneously activating transcription on TATA-less promoters in chickens²⁷. The role of TBPL1 in worms, frogs, and fishes has been examined using knockdown and dominant negative strategies^{28–32}. In these organisms, deletion of *TBPL1* leads to an early halt in embryonic development and further downregulation of certain gene subsets¹⁰. Numerous transcripts in frogs require TBPL1 but not TBP or TBPL2³⁰. TBPL1 does not affect normal embryogenesis in mice, but it is crucial for proper spermatogenesis^{33,34}. This might suggest a functional specialization in different species^{10,18–22,33,35–38}.

TBPL1 has recently been linked to the development and progression of cancer. The downregulation of miR-18a³⁹ and miR-133b⁴⁰ was demonstrated in colorectal cancer cells and tissues. These miRNAs regulate the expression of *TBPL1*, and their overexpression drives a significant decrease in the invasion and proliferation of cancer cells^{39,40}. In agreement with these findings, a study of salivary gland tumors revealed that the *TBPL1* gene was part of a fusion product with the *Nuclear Factor I B (NFIB)* gene, suggesting its participation in oncogenic chromosome rearrangements⁴¹. The *TBPL1* gene was significantly mutated in samples of pulmonary carcinoid tumors⁴². Another study on the pathways affected by genomic changes revealed that TBPL1 is associated with the NF- κ B signaling pathway, which is linked to cell proliferation, metastasis, and epithelial-mesenchymal transition⁴². TBPL1 has also been involved in cancer in other studies; for example, in a study conducted using Mendelian Randomization analysis, this transcriptional factor was identified as one of five genes whose expression showed a significant causal relationship with prostate cancer risk⁴³.

The existence of TBP paralogs raises the question of why more than one TBP-related factor is required in organisms from metazoans to vertebrates, where only one such protein would have soundly executed this function in transcription. Global gene expression analysis offers one plausible hypothesis requiring additional components of the basal transcription machinery to accommodate more elaborated gene regulatory frameworks in these organisms²¹. Switching of TBP family members during development and differentiation is another intriguing explanation for multiple TBP paralogs¹⁰.

The aim of this study was to determine the transcriptome of the *TBPL1* gene in breast cancer cell lines and the subsequent effect of the knocking out of this gene on breast cancer regulation. For this purpose, the *TBPL1* gene was mutated using CRISPR/Cas9 technology, and a comparative analysis of wild-type versus knocked-out cell lines was performed using RNA-seq analysis. We obtained the transcriptomes of the healthy MCF-12F breast cell line and the T47D, SK-BR-3, and MDA-MB-231 breast cancer cell lines, which revealed differential expression patterns of downstream genes, including the upregulation of various novel transcripts. These revealed differential expression patterns of downstream genes, with the upregulation of several novel transcripts. The current study suggests a potential role of the *TBPL1* gene in cancer aggressiveness, cell migration, invasion, metastasis, and the EMT pathways, as well as the involvement of different genes in regulating apoptosis in breast cancer cell lines. Consequently, this indicates the involvement of the *TBPL1* gene in cancer regulation.

Results

TBP and *TBPL1* genes show differential gene expression behavior among different breast cancer cell lines

Human cells possess three members of the TBP family: *TBP*, *TBPL1*, and *TBPL2*⁷. Initially, we determined the mRNA gene expression levels of these genes via end-point RT-PCR, using the HCC-1937, MCF-7, T47D,

SK-BR-3, and MDA-MB-231 breast cancer cell lines, with the MCF-12F and MCF-10A healthy breast cell lines selected as controls (Fig. 1). *TBP* and *TBPL1* are expressed in all cell lines at varied levels, but *TBPL2* is not (Fig. 1B–D). As controls, we also assessed the expression of the *B2M* gene, which encodes the Beta-2-Microglobulin (Fig. 1E), as well as the integrity of *RNA28S* and *RNA18S* ribosomal transcripts in total RNA samples (Fig. 1A).

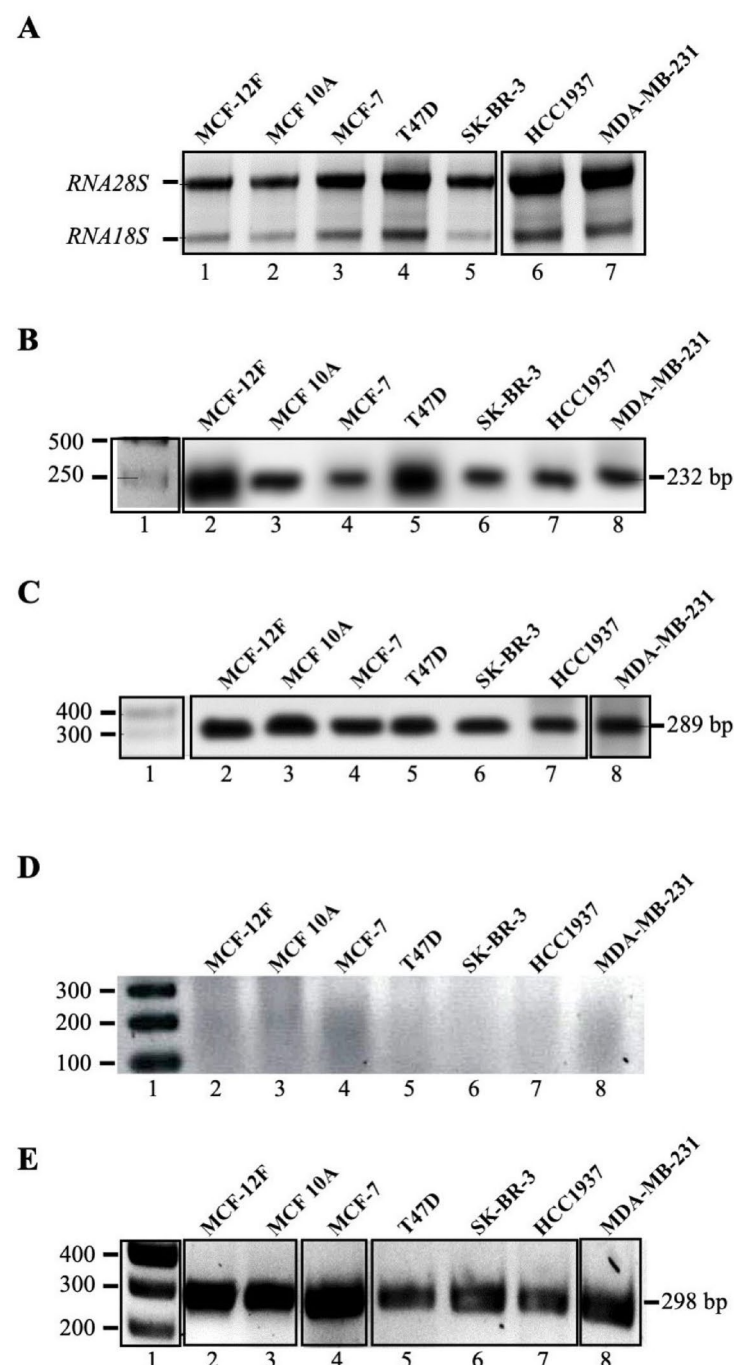


Fig. 1. *TBP* and *TBPL1* genes, but not *TBPL2*, are expressed in healthy and breast cancer cell lines. Total RNA was isolated with Trizol and end-point RT-PCR was performed with the oligonucleotides shown in Table 2. (A) Electrophoresis on 1% bleach agarose gel showing the integrity of the *RNA28S* and *RNA18S* ribosomal genes in MCF-12F and MCF-10A healthy cell lines (lanes 1 and 2, respectively), and MCF-7, T47D, SK-BR-3, and MDA-MB-231 breast cancer cell lines (lanes 3 to 7, respectively). (B) Amplification of a 232 bp DNA fragment corresponding to the *TBP* gene in all breast cell lines. (C) Amplification of a 289 bp DNA fragment corresponding to the *TBPL1* gene in all cell lines. (D) *TBPL2* gene was not expressed in any of the healthy cell lines or breast cancer cell lines. (E) As an internal control, a 298 bp DNA fragment corresponding to the *B2M* gene was amplified in all the breast cell lines studied.

The quantitative analysis of the mRNA levels of *TBP* and *TBPL1* genes via RT-qPCR showed distinct expression patterns for each studied cell line (Fig. 2). *TBP* showed consistent expression levels in healthy cell lines (MCF10A and MCF-12F). However, its expression differed in breast cancer cell lines. *TBP* was found to be underexpressed in HCC1937, MCF7, and T47D breast cancer cell lines (Fig. 2A). Conversely, it was overexpressed in the more aggressive SK-BR-3 and MDA-MB-231 breast cancer cell lines. Interestingly, there may be a direct correlation between the expression level of this gene and the molecular classification of the breast cancer cell lines. MCF-7 and T47D, which belong to the Luminal A subtype, are not highly aggressive and non-invasive breast cancer cell lines displaying similarly low levels of *TBP* mRNA gene expression.

Contrarily, SK-BR-3—classified as enriched in the HER2 receptor—showed a 400% overexpression of *TBP*. This level could be tied to the aggressiveness, invasiveness, and ability of SK-BR-3 to form undifferentiated tumors in immunocompromised mice⁴⁴. Notably, MDA-MB-231, a cell line that is triple-negative for estrogen (ER), progesterone (PR), and Her2 receptors, revealed nearly 750% *TBP* gene overexpression compared to healthy cell lines. This is in line with its characteristics—the ability to invade, proliferate, and form tumors in mice, as well as the poorest patient survival prognosis for triple-negative cancer. Conversely, HCC1937—despite being triple-negative—exhibited a lower *TBP* gene expression level than MDA-MB-231, possibly due to its stage IIB grade 3 designation per the TNM classification system.

RT-qPCR analysis of the *TBPL1* gene in healthy control cell lines revealed consistent expression levels. In HCC1937 and MCF7 cell lines, we observed a reduction in *TBPL1* expression by 60% and 50%, respectively, compared to healthy cell lines. Conversely, T47D, SK-BR-3, and MDA-MB-231 demonstrated increased *TBPL1* expressions of 30%, 90%, and 120%, respectively (Fig. 2B). This uptick in *TBPL1* expression in the latter two cell lines potentially correlates with their heightened cancer aggression. Notably, the expression patterns of the *TBPL1* gene mirror those found in *TBP* expression in HCC1937, characterized by relatively low *TBP* gene expression compared to MDA-MB-231. Therefore, the low *TBP* and *TBPL1* gene expression levels might be linked to the reassuring prognosis of the stage IIB breast cancer subtype, as opposed to the more severe prognosis related to the triple-negative MDA-MB-231 of stage IV, despite being triple-negative as well.

Silencing of the *TBPL1* gene in breast cancer cell lines

Given that *TBP* is known to be vital for cell survival in several organisms due to its crucial role in transcription conducted by the three RNA polymerases, we concentrated our study on the role of *TBPL1* in these cell lines. Initially, we assessed the quantity of *TBPL1* expressed in cells via immunofluorescence and confocal microscopy (Fig. 3). Our focus was particularly on T47D, SK-BR-3, and MDA-MB-231 as these cell lines exhibited overexpression of the *TBPL1* mRNA (Fig. 2). The *TBPL1* protein was observed to be overexpressed, primarily localizing in the cytoplasm in all three breast cancer cell lines, compared with the MCF-12F healthy cell line. This observation corroborated the overexpression determined by RT-qPCR.

To ascertain the potential function of *TBPL1* in breast cancer cell lines, we employed the CRISPR/Cas 9 method to knockout this gene, using the U6-gRNA Cas-92A + GFP-RM CRISPR plasmid. We confirmed the knockout through RT-PCR in transfected MCF-12F, T47D, SK-BR-3, and MDA-MB-231 cell lines (Fig. 4B,D,E,H, correspondingly), compared with untransfected cell lines (Fig. 4A,C,E,G). Notably, no *TBPL1* gene expression was observed in all transfected cell lines. We also documented the integrity of the *RNA28S* and *RNA18S* ribosomal transcripts in total RNA samples from all four cell lines (Fig. 4I,J). High-quality RNA samples were prepared in triplicates, culminating in 24 samples (Suppl. Table 1), which were used for next-generation sequencing (NGS) under both wild-type and mutated conditions.

Western blot assays were used to determine the *TBPL1* protein levels in both healthy and knocked-out cells. T47D, SK-BR-3, and MDA-MB-231 cells were noted to express the *TBPL1* protein (Fig. 5, lanes 4, 6, and 8), except MCF12F cells (Fig. 5, lane 2). Knocked-out cells exhibited a significant decrease in *TBPL1* in T47D and MDA-MB-231 cells (Fig. 5, lanes 5 and 9, respectively), while there was no detection of *TBPL1* in SK-BR-3 and MCF12F cells (Fig. 5, lanes 7 and 3, respectively). Tubulin protein showcased expression in both wild and mutated cells. Despite *TBPL1* mRNA levels not being detected by RT-PCR, the amounts of protein expression varied. This is potentially due to a longer half-life for the protein in T47D and MDA-MB-231 cells. Alternatively, it could be attributed to the cells not being incubated with G-418 immediately following transfection with the plasmid.

Transcriptomic analyses using next-generation sequencing technology

The RNA-Seq analyses are divided into two sections: the first part shows the differential gene expression found in healthy vs. breast cancer cell lines, and the second displays the comparison between the wild-type and the corresponding *TBPL1* mutated cell line.

Differential expression analysis of healthy vs breast cancer cell lines

The differential gene expression analysis was conducted between the MCF-12F healthy cell line and the SK-BR-3, T47D, and MDA-MB-231 breast cancer cell lines. We identified 3860 genes that exhibited significant changes in their expression levels based on false discovery rate (FDR) values for overexpressed and underexpressed genes. The volcano plots display the correlation between statistical significance ($\log_{10}$ FDR) and the $\log_2$ fold change in gene expression for T47D (Fig. 6A), SK-BR-3 (Fig. 5B), and MDA-MB-231 (Fig. 5C), each compared to MCF-12F.

The Principal Component Analysis (PCA) plot illustrates that the three experiments for each cell line are differentially grouped (Fig. 6D). The three breast cancer cell lines are distinctly separated from the healthy cell line. As anticipated, the T47D (Luminal A) appears more closely related to SK-BR-3 (HER2 overexpressing), yet both are distant from MDA-MB-231 (Triple Negative) due to their molecular classifications.

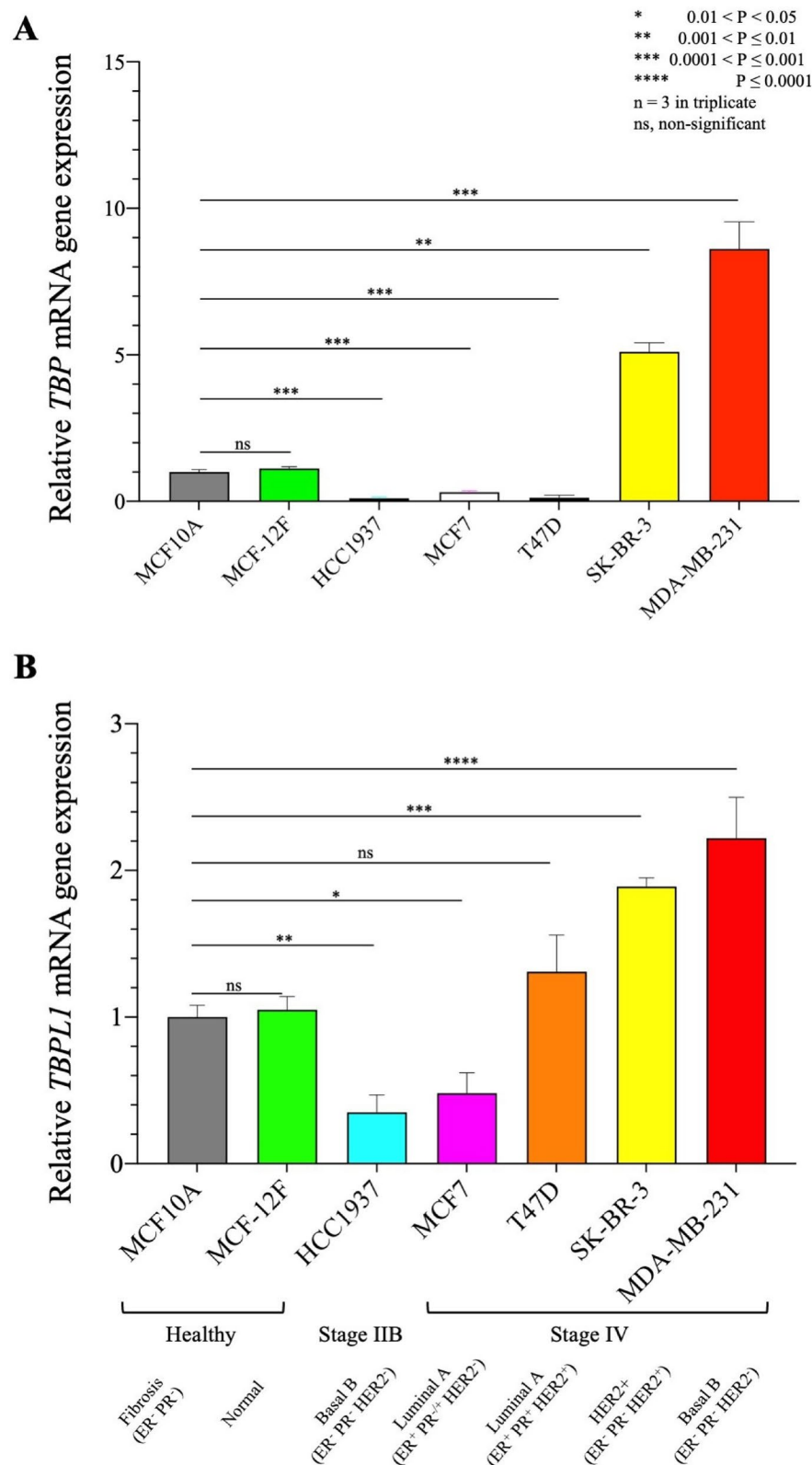


Fig. 2. *TBP* and *TBPL1* genes are differentially expressed in breast cancer cell lines according to the stage and molecular classification. **(A)** Relative *TBP* mRNA expression levels were evaluated by RT-qPCR using the *RNA18SN1* gene as a housekeeping gene and normalized to the values obtained for MCF10A. **(B)** RT-qPCR evaluated relative *TBPL1* mRNA expression levels. The results shown are from three independent experiments in triplicate, and the analysis of the results was performed with the $2^{(-\Delta\Delta Ct)}$ algorithm. Statistics analysis was carried out with unpaired Student's t-test with Welch's correction., and the asterisks show the p-value for significance.

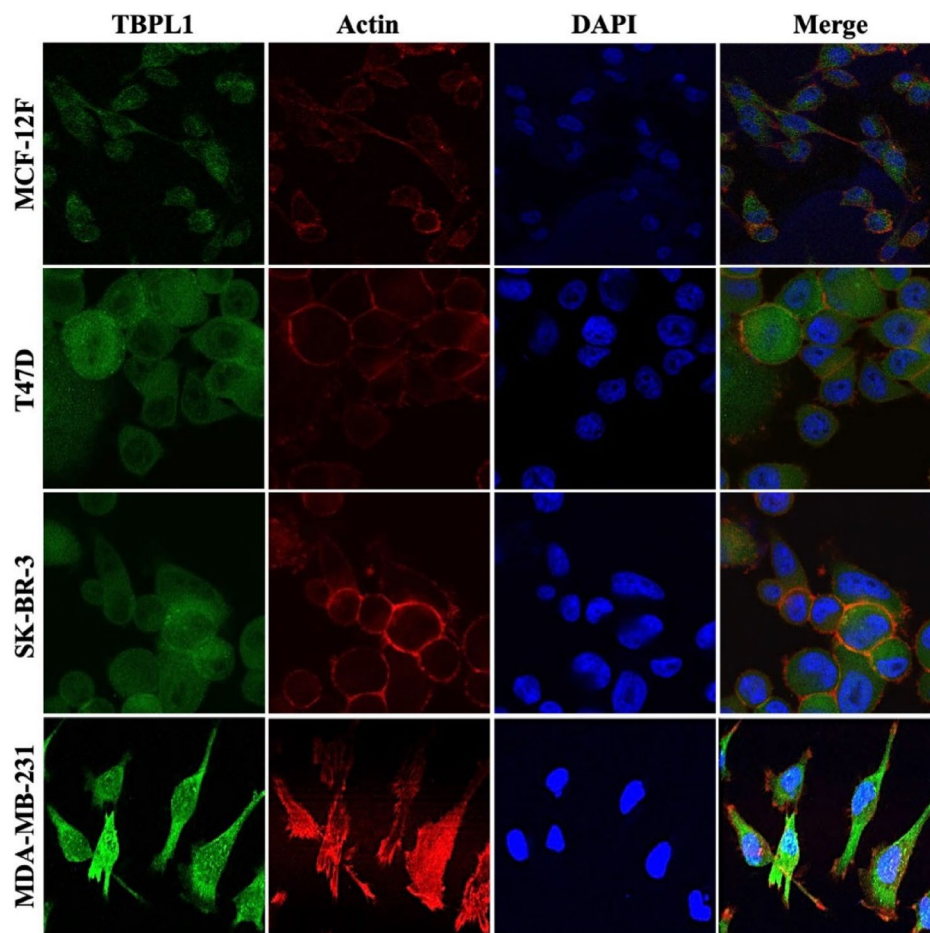


Fig. 3. Intracellular localization of TBPL1 protein by immunofluorescence and confocal microscopy in healthy and breast cancer cell lines. Cells were fixed and incubated with antibodies against TBPL1 (1:200 dilution) and stained with Rhodamine-Phalloidin (1:100 dilution) for actin cytoskeleton and with Hoechst 33,342 for the nuclei.

The genes that altered their expression level across all studied breast cell lines were categorized using the k-means algorithm. The heatmaps show that distinctive expression patterns are present in each cell line (Fig. 6E). For instance, MCF-12F presents an exclusive cluster of 1,684 upregulated genes, only seen in this healthy cell line (Fig. 6E, cluster 1). Notably, the same signature appears downregulated in breast cancer cell lines. SK-BR-3 possesses a unique cluster of 780 genes that are specifically overexpressed (Fig. 6E, cluster 2). Similarly, T47D exhibits 795 upregulated genes (Fig. 6E, cluster 3), while the more aggressive cell line, MDA-MB-231, also presents a distinct signature of 582 overexpressed genes (Fig. 6E, cluster 4). Intriguingly, the genes seen in clusters 2, 3, and 4 are downregulated in the healthy cell line, hinting at their potential role in breast cancer.

The main pathways affected in breast cancer cell lines compared to healthy cell lines are displayed in Fig. 7. The Luminal A T47D breast cell line shows alterations in the expression levels of genes related to ribosome function, including the mitochondrial ribosome (Fig. 7A). These genes, which are part of the large or small subunits, were primarily overexpressed. Some of these genes have also been implicated in coronavirus disease. Other significant pathways involve cancer-related proteoglycans, which are chiefly overexpressed. Many of these are connected to the cytoskeleton and focal adhesion, including Laminin (LAMC2, LAMB3), Actinin (ACTN1, ACTN4), Collagen (COL4A1, COL4A6), Tenascin C (TNC), and Integrin (ITGA7, ITGA1). Transcription factors like MYC, ESR1, and MDM2 were also overexpressed, but others, such as HIF1A, were downregulated. Additionally, factors related to cell cycle control, such as CDKN1A, and other signaling molecules, including RRAS2, WNT4, WNT5A, and ERBB4, were affected.

In the case of SK-BR-3, some molecules involved in nucleotide metabolism were found to be either upregulated or downregulated (Fig. 7B). The PI3K-AKT signaling pathway was identified as altered. Other affected pathways in this cell line included those genes involved in Focal Adhesion, Human Papillomavirus infection, TNF signaling pathway, and efferocytosis, the latter being involved in the clearance of apoptotic cells. Finally, a few pathways were impacted in MDA-MB-231 (Fig. 7C), including the guidance of axons, coronavirus diseases, and assembly of ribosomes.

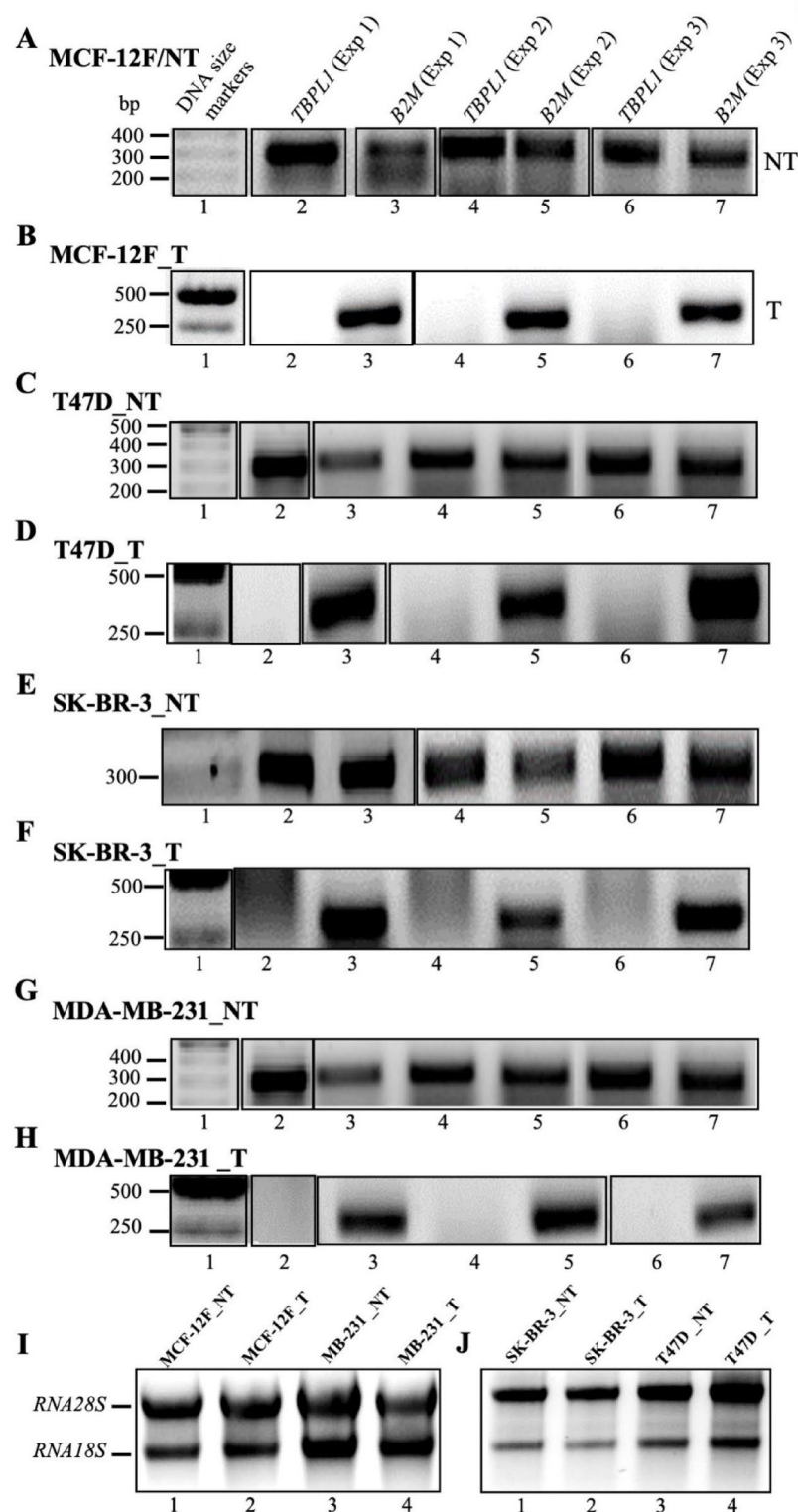


Fig. 4. Determination of the *TBPL1* gene expression levels in untransfected and transfected healthy and breast cancer cell lines with the U6-gRNA Cas-92A + GFP-RM CRISPR plasmid. The *TBPL1* gene mRNA levels were measured by end-point RT-PCR in all cell lines. As a control, the *B2M* mRNA expression level was also determined in all samples. RNA samples of the three independent experimental replicates for nontransfected and transfected cell lines were used as templates for RT-PCR assays before sending them for RNAseq analyses. (A, C, E, G) MCF-12, T47D, SK-BR-3, and MDA-MB-2321 nontransfected (NT) cell lines. (B, D, F, H) MCF-12, T47D, SK-BR-3, and MDA-MB-231 transfected (T) cell lines. (I) RNA gel electrophoresis shows the integrity of *RNA18S* and *RNA28S* ribosomal genes for all healthy and breast cancer cell lines.

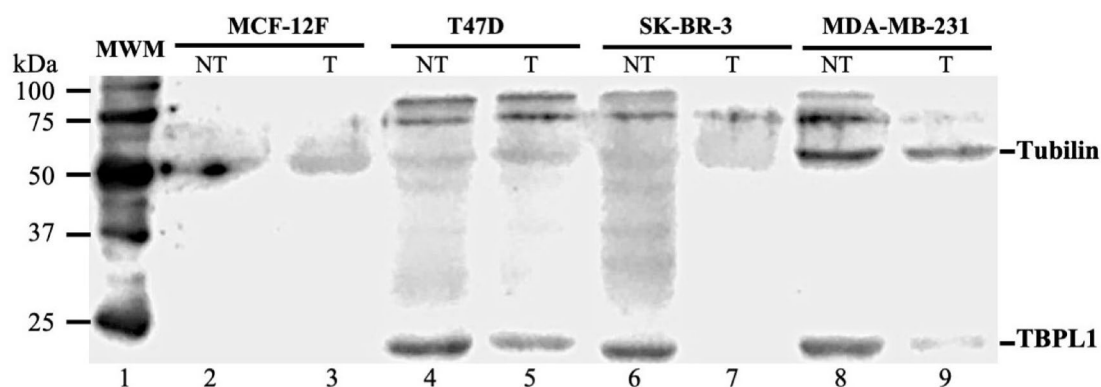


Fig. 5. Determination of the TBPL1 protein expression level by Western blot. Protein extracts from nontransfected and transfected cell lines with the U6-gRNA Cas-92A + GFP-RM CRISPR plasmid were analyzed by Western blot. Membranes were probed with anti-TBPL1 and anti-Tubulin antibodies and revealed for chemoluminescence as described in Methods.

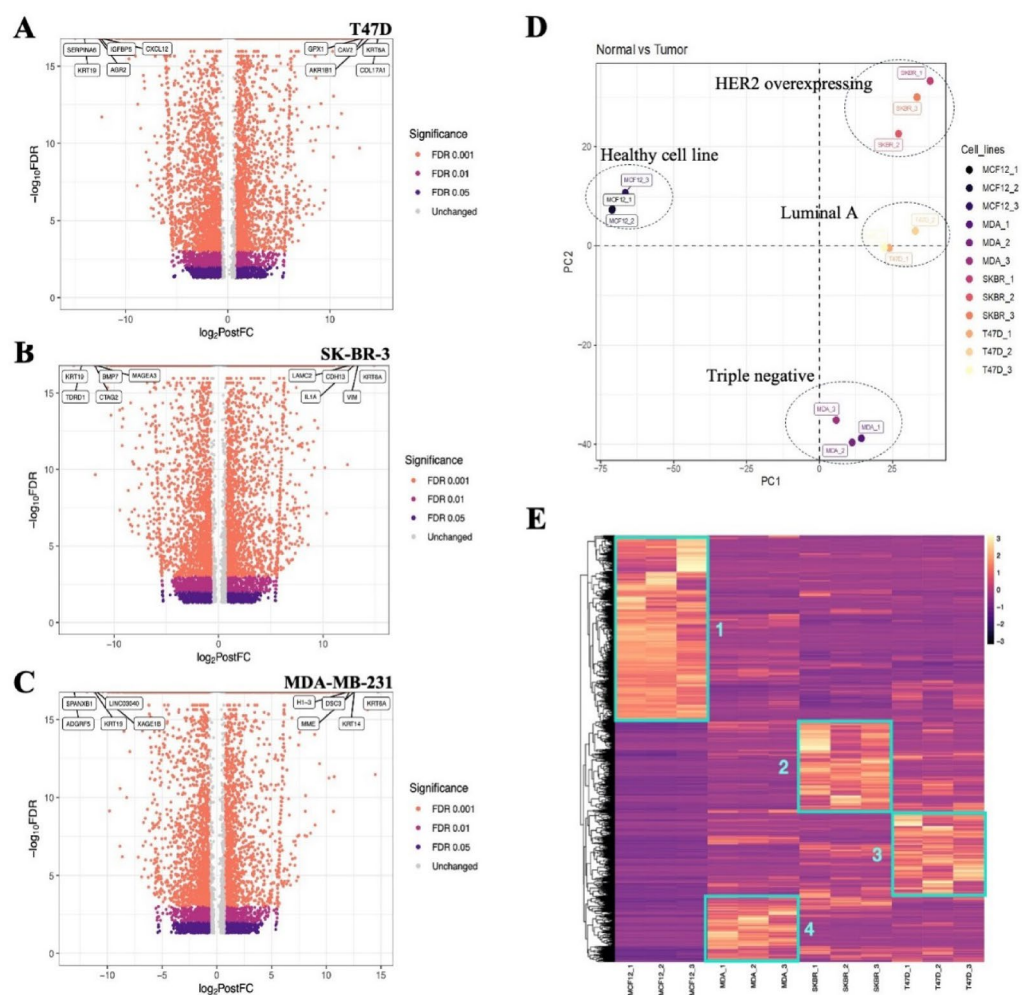


Fig. 6. Gene expression profiles of healthy and breast cancer cell lines. RNAseq data for each cell line were obtained as described in Methods. Comparisons were then carried out between each breast cancer cell line and the healthy breast cell line. Volcano plot of the genes identified in T47D (A), SK-BR-3 (B), and MDA-MB-231 (C). (D) Principal Component Analysis (PCA) of the data obtained for the three replicates for each cell line. (E) Hierarchical clustering of the genes that were found up-regulated and down-regulated in the healthy and breast cancer cell lines.

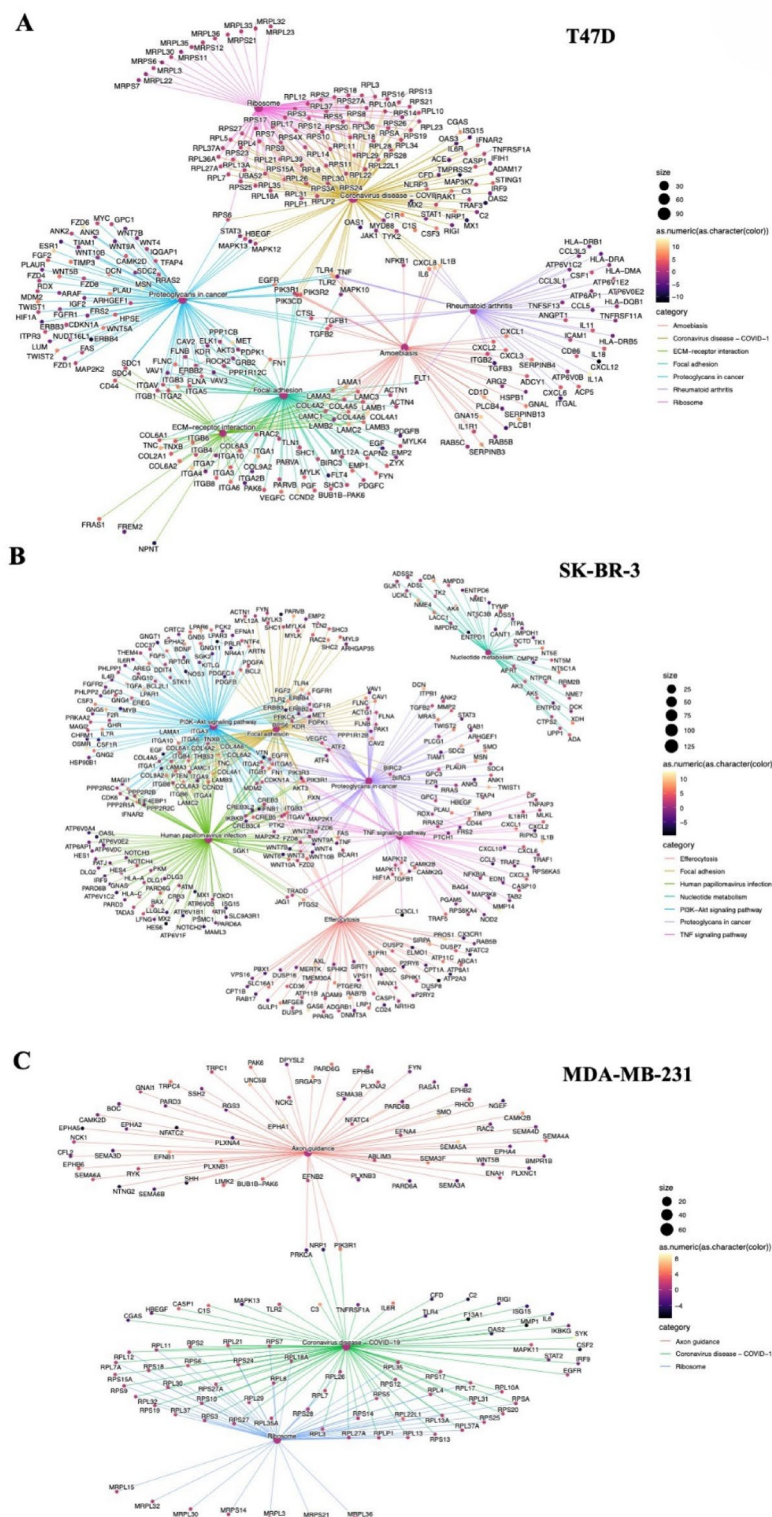


Fig. 7. Enriched pathways (KEGG) Interactomes of the genes affected in breast cancer cell lines compared to the healthy MCF-12F breast cell line. T The overexpressed or subexpressed genes affected in breast cancer cell lines compared to the healthy breast cells are shown and classified according to the main pathways involved. (A) T47D. (B) SK-BR-3. (C) MDA-MB-231.

Differential expression analysis of the wild-type and *TBPL1* knocked-out gene in the healthy breast and breast cancer cell lines

The effect of knocking out the *TBPL1* gene using CRISPR/Cas9 technology on the expression profile of both healthy and breast cancer cell lines is depicted (Fig. 8). Volcano plots reveal minimal changes in the expression profiles of the wild-type compared to each mutant breast cell line. The MCF-12F healthy breast cell line exhibited the most dramatic changes (Fig. 8A), followed by the breast cancer cell lines SK-BR-3 (Fig. 8C), MDA-MB-231 (Fig. 8D), and T47D (Fig. 8B). Wild-type and mutant MCF-12F cell replicates were differentially grouped (Fig. 8E). Parallel behavior was observed for T47D cells and SK-BR-3 cells. Conversely, the wild-type and mutant cells corresponding to MDA-MB-231 appeared closer, suggesting minimal effects of *TBPL1* gene knockout on the gene expression profile. Hierarchical clustering revealed a set of genes (Fig. 8F, 1) that showed decreased or unchanged expression in all *TBPL1* knockout cell lines relative to their wild-type counterparts. Cluster 4 in wild-type MCF-12F cells was found to be upregulated; however, the silencing of the *TBPL1* gene led to downregulation. For cluster 3 in wild-type SK-BR-3 cells, there was an initial upregulation which turned into downregulation after the *TBPL1* gene was silenced. Cluster 1 remained unaffected by *TBPL1* knockout in all cells.

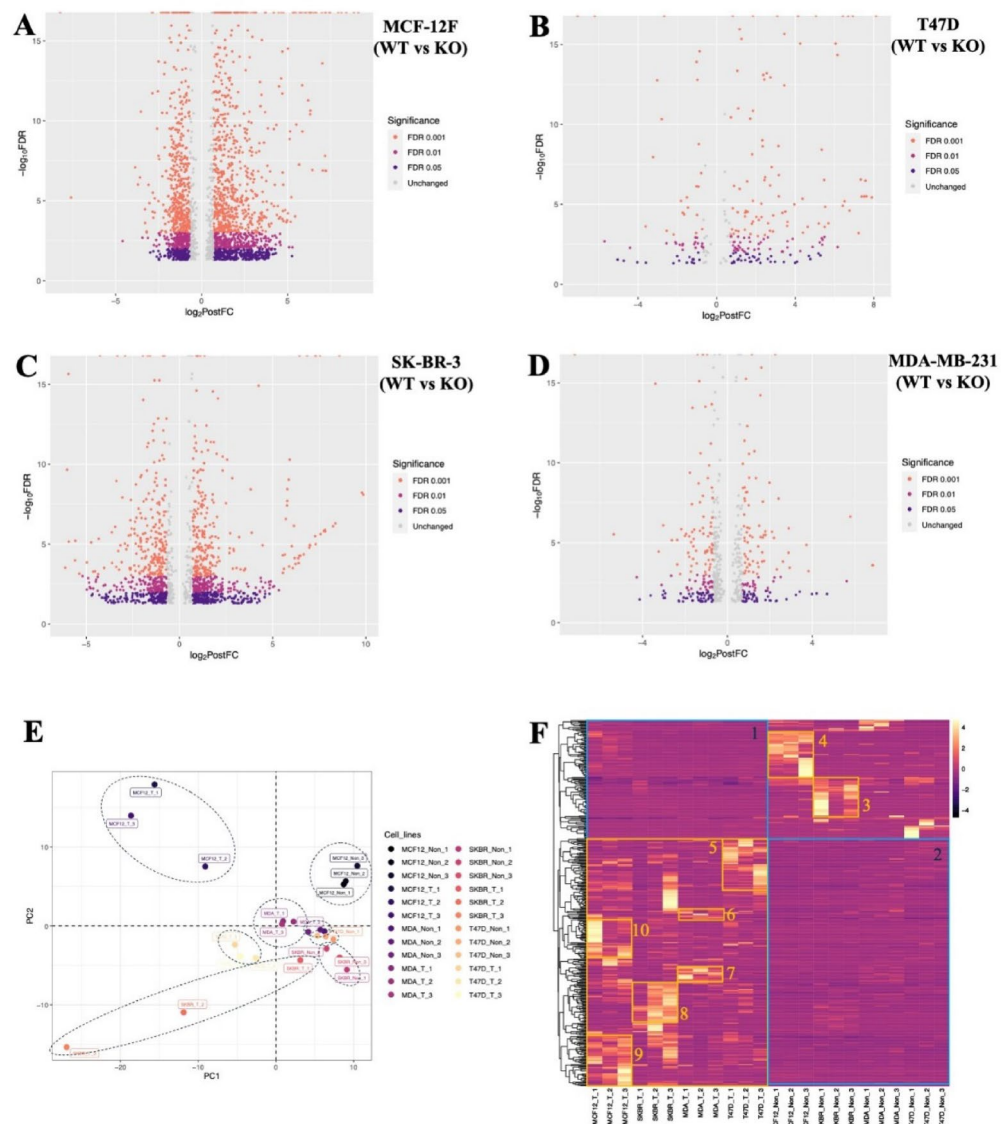


Fig. 8. Gene expression profiles of the wild type and *TBPL1* knocked out gene determined in healthy and breast cancer cell lines. Knock-out cell lines were obtained by transfection of the U6-gRNA Cas-92A + GFP-RM CRISPR plasmid with Lipofectamine as described in Materials and Methods. Comparisons of RNAseq data were then carried out between non-transfected and transfected cell lines. Volcano plot of the genes identified in MCF-12F (A), T47D (B), SK-BR-3 (C), and MDA-MB-231 (D). (E) Principal Component Analysis (PCA) of the data obtained for the three replicates for each cell line. (F) Hierarchical clustering of the genes found overexpressed and subexpressed in the healthy and breast cancer cell lines.

In cluster 2, which was downregulated in all wild-type cell lines, specific gene groups altered significantly following *TBPL1* gene knockout. For instance, cluster 5 was upregulated in T47D knockout cells. Similarly, the knockout of the *TBPL1* gene in MDA-MB-231 cells led to the overexpression of clusters 6 and 7, whereas in SK-BR-3, cluster 8 was upregulated. For the healthy cell line, clusters 9 and 10 were also observed to be upregulated after the mutation of the *TBPL1* gene.

In the case of MCF-12F, the most impacted pathways include those involved in infections such as Influenza A, Epstein-Barr virus, Measles, and SARS-CoV-2, as well as ribosome function, the Tricarboxylic acid cycle, Glyoxylate and dicarboxylate metabolism, Antigen processing and presentation, the mRNA surveillance pathway, and Mitophagy. Several affected genes are also connected to diseases including Parkinson's, Huntington's, and diabetic cardiomyopathy, as well as multiple disease neurodegeneration pathways (Fig. 9A,B). The *TBPL1* gene knockout in T47D impacted pathways associated with various viral infections, including human papillomavirus, SARS-CoV-2, Measles, Hepatitis C, Influenza A, and Epstein-Barr virus (Fig. 9C,D). The influenced pathways in SK-BR-3 featured cellular senescence, Lysosome function, and nucleocytoplasmic transport (Fig. 9E,F). Lastly, the affected pathways in MDA-MB-231 involved primary immunodeficiency, antigen processing and presentation, ferroptosis, glutathione metabolism, central carbon metabolism, glycolysis/gluconeogenesis, and fructose and mannose metabolism (Fig. 9G,H).

Thus, the effect of knocking out the *TBPL1* gene varies in each cell line. Even among breast cancer cells, the effects significantly differ, reflecting the heterogeneity of gene expression patterns and likely the differences in their genomes.

Discussion

TBPL1 is a member of the TBP family. Prior studies have demonstrated that *TBPL1* binds to active promoters in haploid testis cells in mice, alongside TBP and TAF³²—however, unique regulatory mechanisms might exist in the germline. Species-specific differences in *TBPL1* functions could potentially be explained by variations in the *TBPL1* core domain. While the core domain of TBP is over 80% conserved across species, *TBPL1*'s core domain only maintains 40–45% amino acid conservation among metazoans¹⁰. *TBPL1* controls specific subsets of genes that are distinct from TBP. Additionally, our findings indicate that the *TBPL1* gene manages different subsets of genes in healthy versus breast cancer cell lines, and in wild-type versus mutated conditions. It is essential to noteworthy that although the contributions of *TBPL1* described thus far in healthy cells are primarily related to transcriptional regulation, genetic alterations have also been identified in cancer cells. For example, in a study analyzing salivary gland tumors, a novel gene fusion between the *NFIB* and *TBPL1* genes was identified⁴¹, suggesting functions beyond the control of gene expression for this transcription factor. Likewise, in a study using pulmonary carcinoid tumors, the *TBPL1* gene was identified as one of the significantly mutated genes⁴². This research underscores the significance of investigating this transcription factor as a key target in cancer biology.

TBPL1 and *TBP* are the only genes transcribed in the examined cell lines, while *TBPL2* remained unexpressed (Fig. 1). The *TBP* gene was detected at deficient levels in HCC1937, MCF7, and T47D but was overexpressed in the more aggressive and poor-prognosis SK-BR-3 and MDA-MB-231 cell lines. The *TBPL1* gene was also overexpressed in these cell lines; in the first set of cell lines, it was underexpressed. This pattern of overexpression has also been observed in colorectal cancer, where increased *TBPL1* expression has been linked to an increased capacity for cancer cell invasion and proliferation^{39,40}. Although the mechanisms by which *TBPL1* expression is controlled in breast cancer have not been investigated, similar regulatory mechanisms are likely to exist, especially in more aggressive cell lines such as SK-BR-3 and MDA-MB-231. T47D exhibited a pattern in which these two genes were overexpressed (Fig. 2). This pattern strongly insinuates a distinct gene expression profile for each cell line. The expression of the *TBPL1* protein was validated by immunofluorescence and confocal microscopy (Fig. 3), RT-PCR (Fig. 4), and Western blot (Fig. 5). MCF-12F showed lower expression than the other cell lines.

In the first segment of this work, the transcriptomes of T47D, SK-BR-3, and MDA-MB-231 breast cancer cell lines were contrasted with the healthy MCF-12F breast cell line (Fig. 6). The T47D cell line, a Luminal A and PR+ that displays chemo-resistance. T47D showcases differentially expressed genes across various pathways like amyotrophic lateral sclerosis, chemical carcinogenesis reactive oxygen species, Coronavirus disease 2019 (COVID-19), diabetic cardiomyopathy, protein processing in the endoplasmic reticulum, ribosome, and thermogenesis (Fig. 7A). The downregulated genes *SERPINA6*, *IGFBP5*, *CXCL12*, *KRT19*, and *ARG2*, along with the upregulated *GPX1*, *CAV2*, *KRT6A*, *AKR1B1*, and *COL17A1* genes, were selected for this cell line. *SERPINA6*, found in both liver and breast cancer cell lines, encodes the alpha-globulin protein with a corticosteroid-binding capacity domain. *IGFBP5* inhibits both the growth and metastatic abilities of cancer cells⁴⁵; hence, it was seen as downregulated in T47D. The chemokine *CXCL12*, prevalent in typical sites of tumor metastasis such as lungs, lymph nodes, bone marrow, liver, and animal models⁴⁶, works by attracting T-lymphocytes and monocytes. Its subexpression potentially implies a negative regulation of the immune response.

SK-BR-3 is a poorly differentiated HER2+ overexpressing cell line, isolated from the pleural effusion of a patient with metastasis. It has shown a downregulation of *KRT19*, *BMP7*, *MAGEA3*, *TDRD1*, and *CTAG2* genes. *KRT19* has been acknowledged to regulate breast cancer properties by activating the AKT signaling pathway⁴⁷. However, recent studies have illuminated that modulation of *KRT19* expression can lead to contrasting effects on cell proliferation, survival, invasion, migration, and apoptosis, and these effects vary across different cancer cell types⁴⁸.

BMPs, or Bone Morphogenetic Proteins, are cytokines; extracellular signaling molecules that bind to the TGF-beta protein family. These proteins regulate gene transcription by signaling via specific serine-threonine receptors and intracellular SMAD proteins^{49,50}. This signaling cascade is regulated extensively on various levels⁵¹. Moreover, BMPs are involved in the activation and interaction with other signaling pathways, like the MAPK

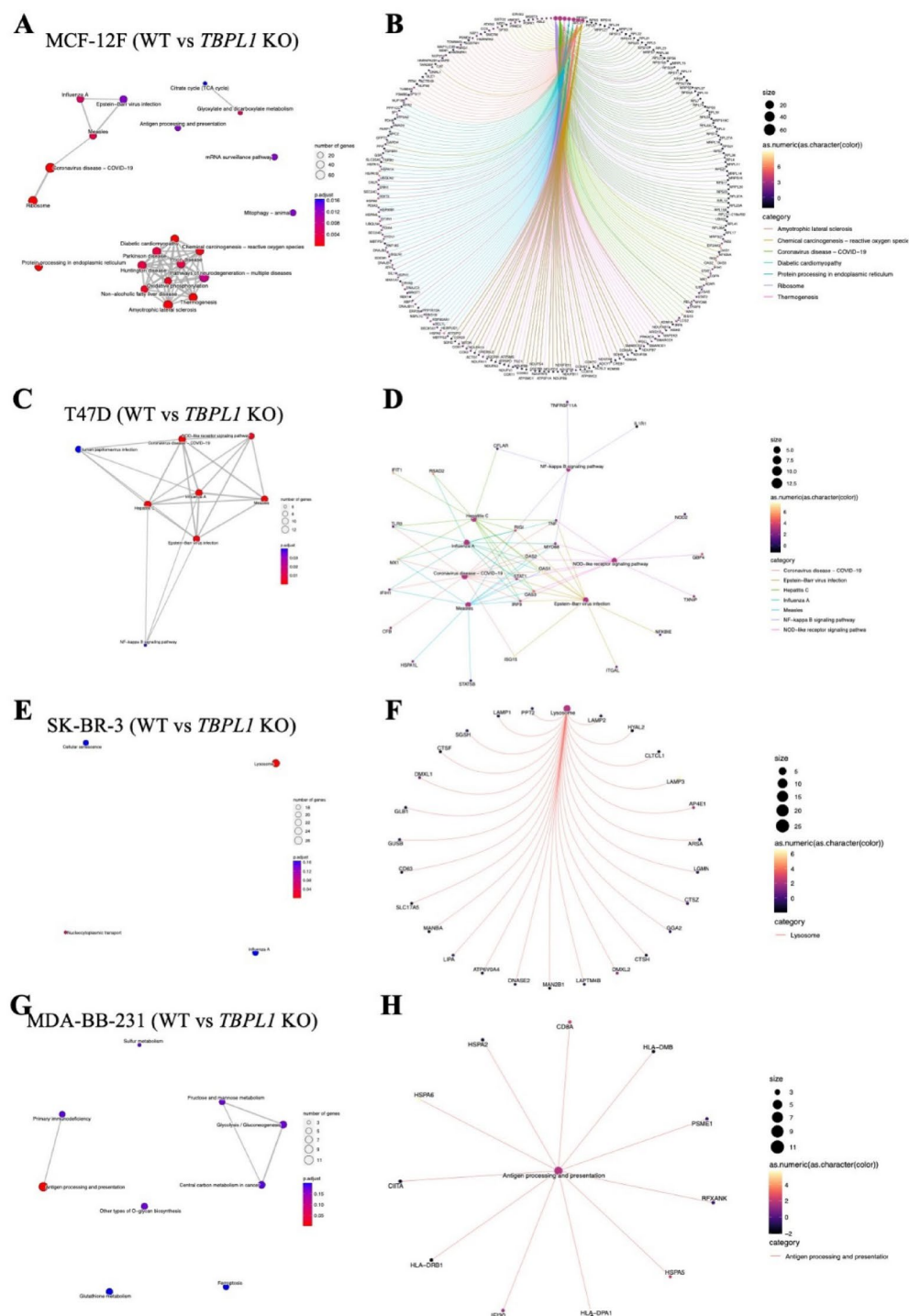


Fig. 9. Enriched pathways (KEGG) and gene expression levels affected by the knocking out of the *TBPL1* gene. Pathways and genes affected by the *TBPL1* gene knocking out in relation to the wild-type cells: MCF-12F (A, B), T47D (C, D), SK-BR-3 (E, F), MDA-MB-231 (G, H).

pathway⁵². BMPs influence a range of cellular processes including proliferation, differentiation, and apoptosis in assorted cell types⁵³.

BMP7 also plays a significant function in cell invasion, proliferation, and migration as demonstrated in MDA-MB-231 cells⁵⁴. It has also been proven to inhibit cellular proliferation in several other cell lines⁵⁵. In an in vivo xenograft mouse model using MDA-MB-231 cells, BMP7 has displayed contradictory effects such as reducing tumor growth⁵⁶. Furthermore, BMPs have been implicated to play a role in tumor angiogenesis and metastasis⁵⁷.

The *MAGEA12* gene is found to be downregulated in the SK-BR-3 cell line compared to the healthy MCF-12F cell line. It is significantly associated with aggressiveness in breast cancer, regardless of hormone receptor subtype status. This suggests that the *TBPL1* gene may be involved in the invasive properties of the SK-BR-3 cell line, which could contribute to its heightened aggressiveness. Furthermore, genes with a downregulated *MAGEA12* signature are implicated in the migration and invasion of breast cancer cells. The regulation of *MAGEA12* expression could, therefore, play an integral role in determining the morphology of aggressive breast cancer cells⁵⁸.

The *CTAG2* gene (Cancer/Testis Antigen 2 gene) has been linked with diseases such as melanoma and other forms of cancer, inclusive of bladder and prostate cancers, as well as epileptic encephalopathy. In the case of breast cancer, members of the CTA family have been implicated in promoting invasion⁵⁹. The Cancer-Testis antigens *SPANX-A/C/D* and *CTAG2* further contribute to the invasiveness of breast cancer⁶⁰.

The *LAMC2*, *CDH13*, *KRT6A*, *IL1A*, and *VIM* genes were upregulated in the SK-BR-3 cell line. Aberrant *LAMC2* (Laminin) expression strongly correlates with metastasis and poor prognosis in lung cancer⁶¹. Existing research reports a positive correlation between *LAMC2* and *EGFR* expressions in bladder cancer, oral squamous carcinoma, cholangiocarcinoma, and multiple human cancer cell lines, including those of breast, neoplastic, and thyroid cancers⁶². *LAMC2* has also been seen to bind to *EGFR* in breast carcinoma and stimulate *MAPK* activation⁶³. Methylation of the *CDH13* gene in breast cancer has emerged as a potent diagnostic biomarker⁶⁴. Keratin 6A (*KRT6A*), a type II keratin involved in the epidermalization of squamous epithelium^{65,66}, plays a crucial role in cell migration, specifically keratinocyte migration⁶⁷. It further promotes the growth and metastasis of lung adenocarcinoma via the induction of the epithelial-mesenchymal transition⁶⁷. *IL1A* (interleukin-1 alpha) is implicated in cell proliferation and invasion⁶⁸. *VIM* (Vimentin), a cytoskeletal protein forming type III intermediate filaments, participates in cell migration, motility, and adhesion. In solid cancers, overexpression of vimentin is linked with EMT and metastasis⁶⁹. In conclusion, the overexpression of these genes in the SK-BR-3 breast cancer cell line is consistent with their known metastatic and cancerous properties at the molecular level.

MDA-MD-231 is the most aggressive subtype with a poor prognosis, characterized by the triple-negative status for ER, PR, and HER2- receptors. One of the significantly impacted pathways is related to ribosomes—this includes the protein-coding genes of the two subunits (Fig. 7C). These genes are overexpressed and their expression levels correlate with the invasiveness and tumor formation capacity of this cell line. Another pathway influenced is the so-called “Axon Guidance”, an embryonic pathway typically activated only during development. Key molecules in this process include the *NFATC2* gene, part of the nuclear factors of the activated T cells family. This transcription factor, localized in the cytoplasm, stimulates the transcription of immune-related genes post-activation by the T-cell receptor. Additionally, *NFATC2* is stimulated through the TGF- β 1 pathway and can activate the *MMP13* gene in MDA-MB-231 through interaction with ATF3. This activation is crucial for invasiveness⁷⁰.

Finally, there are some shared pathways between COVID-19 and cancer. For instance, both conditions involve a pro-inflammatory environment and the NF- κ B pathway. IL-6 is associated with the JAK-STAT pathway. Moreover, there is significant engagement of the same type of immune response through IFN1A. The common pathways between these two conditions include cytokines, IFN1A, Androgen Receptors, and immune checkpoint pathways⁷¹.

In MDA-MB-231 cells, downregulated genes include *SPANXB1*, *LINC03040*, *ADGRF5*, *KRT19*, and *Xage1B*. Conversely, the genes *H1-3*, *DSC3*, *KRT6A*, *MME*, and *KRT14* were upregulated. The human sperm protein associated with the nucleus on the X chromosome (*SPANX*) genes family, comprises five known members (*SPANX-A1*, *-A2*, *-B*, *-C*, and *-D*). These genes encode cancer/testis-specific antigens, which are potential cancer immunotherapy targets. *SPANXB1* promotes triple negative subtype breast cancer metastasis⁷². *LINC00460* expression is a potential marker for aggressive phenotypes in distinct tumors, including locally advanced lung cancer and basal-like BRCA⁷³. Notably, overexpression of *ADGRF5* has been seen in breast cancer. Breast cancer patients with high *ADGRF5* expression have a significantly shorter recurrence-free survival and distant metastasis-free survival compared to those with low expression of *ADGRF5*⁷⁴. *KRT19*, recent studies have shown, can regulate cancer properties by modulating signaling pathways like the *EGR1/PTEN/AKT*, *Wnt/ β -catenin/Notch*, and others. Knocking down *KRT19* promotes breast cancer cell proliferation, migration, and sphere formation through *NUMB*-dependent crosstalk in the *Wnt/Notch* signaling pathway⁵⁰. The activation of the human cancer/testis antigen gene, *XAGE-1*, in tumor cells correlates with CpG island hypomethylation, and it sees high expression in lung adenocarcinoma⁷⁵. The *DSC3* gene, a desmosome adhesion molecule that is responsible for normal physiological integrity and cell adhesion, is upregulated in the MDA-MB-231 cell line. Its expression changes during carcinogenesis, and *DSC3* is a p53-responsive gene. Like other adhesion molecules, it reportedly provides a barrier to the immune infiltration of tumors⁷⁶. Lastly, the expression of membrane metalloendopeptidase (*MME*) has been found in various cancer types. *MME* is overexpressed in pancreatic endocrine tumors and colorectal carcinoma but reduced in lung and ovarian cancer, suggesting cell type-specific effects of *MME*⁷⁷. In MDA-MB-231, this marker gene is upregulated.

The second part involved determining the effect of the *TBPL1* gene knockout. We found different signatures when the cells have the intact or mutated *TBPL1* gene (Fig. 8), with the healthy cell line being the most affected. For the *TBPL1* knocked out MCF-12F cell line, a set of genes including *NPIP3*, *ESNG00000281383*, *ZNF252P*, *TAF4*, *VP53*, *INFL3*, *RSAD2*, *SPATASL1*, and *DLX2* were upregulated. However, these were turned off in the wild-type MCF-12F cell line. Conversely, other gene sets, including *RAP1*, *TNFRSF10A*, *ZNF432*, *MYDGF*, *VSP39*, *CMTR*, *ZNF815P*, and *ENSG00000270066*, were found to be downregulated in the *TBPL1* mutated MCF-12F cell line.

The nuclear pore complex interacting protein, *NPIP3*, exhibits high expression, which contributes to better survival in normal cells and is regulated by estrogen receptors. *RSAD2*, a protein featuring the S-adenosyl-L-methionine (SAM) domain, impedes the replication of DNA or RNA viruses. Although it plays a detrimental

role in endometrial cancer, it also functions as a protein marker indicating a poor prognosis in cancer cells. Moreover, MCF-12F, a non-cancerous cell line, typically displays minimal expression of RSAD2 (Viperin) under wild-type conditions. ENSG00000281383, a newly identified gene product, was discovered during the knockout of the *TBPL1* gene in various cell lines, yet its function remains undefined. Although INFL3 is expressed in BRCA, its presence is somewhat affected in normal cells. This gene influences cytokine, antitumor, and antiviral regulation via the JAK-STAT pathway.

Contrarily, some of the downregulated genes in *TBPL1* muted cells include *RAP1A*. This gene is a part of the Ras oncogene family. Diseases associated with *RAP1A* consist of Kabuki syndrome⁷⁸ and Tuberous Sclerosis⁷⁹. Rap1 plays a varied role in tumor initiation and progression. For instance, the activation of Rap1 induces tumor initiation and epithelial-mesenchymal transition (EMT) through Notch signaling. Activated Rap1 may stimulate EGFR and Src/FAK, leading to integrin-mediated cell adhesion in cancer when overexpressed⁸⁰. However, low levels of the RAP1 gene in normal cells can cause cell senescence and growth inhibition⁸¹.

TNFRSF10A encodes a member of the TNF-receptor superfamily. This receptor is activated by the tumor necrosis factor-related apoptosis-inducing ligand (TNFSF10/TRAIL) which transduces a cell death signal, inducing apoptosis⁸². This gene is downregulated under *TBPL1* mutated conditions, highlighting the importance of the *TBPL1* gene in the survival of normal cells and its potential contributions to cell signaling pathways. ZNF432 is a newly identified HR repressor involved in the cellular response to DNA damage and acts as a PARP inhibitor⁸³.

The MDA-MB-231 cell line demonstrated an upregulation of *ENSG00000285238*, *HKDC1*, *RASA4*, *HSPH1*, *THAP10*, *TEX30*, *AVP11*, and *RASGRP3* genes within *TBPL1* mutated cells. However, the genes *MUCSB*, *CYP3A5*, *KCNE3*, *LINCO1934*, *JAK3*, *GPC2*, *LCDR*, and *ADGRA2* appeared downregulated in the *TBPL1* KO cell lines, despite their overexpression in wild-type cells. *ENSG00000285238* is a novel transcript of unknown function. that signifies the principal component of the nucleosome remodeling and deacetylase complex, playing a crucial role in epigenetic transcriptional repression. *HKDC1* is encoded as a member of the hexokinase protein family. This encoded protein is intrinsically involved in glucose metabolism, and a reduced expression may correlate with gestational diabetes mellitus. Notably, elevated gene expression might be linked with an unfavorable prognosis in hepatocarcinoma⁸⁴. *RASA4* encodes a member of the GAP1 family of GTPase-activating proteins, suppressing the Ras/mitogen-activated protein kinase pathway in response to Ca^{2+} . Strikingly, it activates Ras GTPase functionality.

Consequently, Ras shifts from the active GTP-bound state to the inactive GDP-bound state, no longer initiating downstream pathways that regulate gene expression, cell growth, and differentiation⁸⁵. The *HSPH1* gene encodes a member of the heat shock protein 70 family. The encoded protein functions as a nucleotide exchange factor for the molecular chaperone heat shock cognate 71 kDa protein (Hsc70). Furthermore, this protein has a separate though related role as a holdase, preventing the aggregation of misfolded proteins, including the cystic fibrosis transmembrane conductance regulator (CFTR) protein. Elevated expression of this protein has been observed in various human cancers⁸⁶.

The *CYP3A5* gene is overexpressed in prostate cancer⁸⁶. However, it is upregulated in the MDA-MB-231 cell line, while the *TBPL1* KO cell line showed significant downregulation. *LINCO1934* signifies the long intergenic non-protein-coding RNA1934, a non-annotated gene found upregulated in the wild-type MDA-MB-231 cell line, with significant downregulation in the KO cell line. Many novel transcripts have been found in the NGS data; most are long noncoding RNAs, so their functions remain unknown. Significantly, *JAK3* (Janus kinase 3) contributes to the pathogenesis of natural killer/T-cell lymphoma (NKTCL)⁸⁷, suggesting that it could be involved in immunogenic pathways in this cell line.

The SK-BR-3 cell line exhibited different signatures. In the *TBPL1* OK cell line, *IFI27*, *CMPK2*, *RSAD2*, *HSPA7*, *MRC1*, *IL6*, *IFNL3*, *LINCO1087*, and *RSC1A1*, were upregulated. Conversely, *GCNT3*, *LINCO1484*, *LINCO1293*, *CELF3*, *ART1*, *FOXJ1*, *TEMEM92*, *RENBP*, *HBE1*, and *LEMD1* genes were found to be downregulated in the same cells. However, these genes exhibited the opposite effect in the wild-type *TBPL1* SK-BR-3 cells.

The T47D *TBPL1* KO cell line exhibited upregulation of the following genes: *APOL1*, *LINC02574*, *SLC2A3P1*, *IGHG4*, *MIR210HG*, and *TRIM38*. Conversely, *FIBIN*, *ENSG00000264668*, *ENSG00000270066*, *ENSG00000144785*, *DUSP13*, *ASNSP1*, and *RPGP* genes were downregulated in the same cells.

The heatmap comparison of *TBPL1* mutant cells to wild-type cell lines yields remarkable results (Fig. 8F). These could be attributable to the disparate origins and genomic backgrounds of the cells under scrutiny. A comprehensive examination of our data may elucidate the crucial roles the *TBPL1* gene potentially plays a key role in the biology of both healthy and breast cancer cells. We also searched for *TBP*, *TBPL1*, and *TBPL2* expression levels in breast cancer samples compared to normal breast tissue in the GEPIA2 database. The only genes being expressed correspond to *TBP* and *TBPL1*, which confirms our results. We also looked up the three genes' survival curves (Kaplan-Meier plots) (Supplementary Fig. 1). When *TBP* has low expression, survival is 40% at 7,500 days; in comparison, when *TBP* has a high expression level, survival is 0% at 7,500 days. For the *TBPL1* gene, at low expression, the survival rate is between 25 and 50% at 7500 days, compared to approximately 12.5% when the *TBPL1* gene is overexpressed. *TBPL2* has no statistically significant difference. Thus, this analysis reveals an important role of *TBP* and *TBPL1* genes in breast cancer. Further, the roles of the lncRNAs and the miRNAs not yet evaluated in this study should also be closely examined.

Moreover, the knockout of the *TBPL1* gene reduces cell viability by 75% in the healthy breast cell line (MCF-12F) and 37.5% in the T47D Luminal A breast cancer cell line, while in SK-BR-3 (Her2 overexpressing) and MDA-MB-231 (triple negative) breast cancer cell lines, the viability was reduced to 50%, revealing that it is an essential gene (Suppl. Figure 2). Notably, our data show that *TBPL1* depletion affects both healthy and cancerous cells, suggesting that its role in cancer is context-dependent. For instance, in a recent randomized Mendelian study, it was observed that a high level of *TBPL1* expression is associated with a lower risk of developing prostate

Cell line	Subtype	Molecular markers	Incidence	Description
MCF-12F	Normal basal-epithelial	Does not express receptors like ER	Used as a healthy model in breast cancer studies	Non-tumorigenic* epithelial cell line
T47D & MCF-7	luminal A	ER + and/or PR +, HER2 – and low Ki67	The most common subtype (50–60% of all cases)	Exhibit robust cell-to-cell junctions ⁽³⁶⁾
				The relapse rate is the lowest in comparison to other types
				The recurrence is mainly in the boneGood prognosis for the patient
SK-BR-3	HER2-enriched	HER2/Neu receptor overexpression ER +, PR – and any Ki-67	20–30% of all cases of breast cancer	Aggressive biological and clinical behavior
HCC-1937	Basal-like B	ER –, PR – and HER2 –	TNBC cell line with stage IIB grade 3	Derived from primary ductal carcinoma
MDA-MB-231	Basal-like	ER –, PR –, HER2 –, CK5/6 +, and/or EGFR +	80% of basal-like subtype is triple-negative	Good prognosis for the patient
				15–20% of all cases of breast cancer
				Aggressive clinical behavior
				High rate of metastasis to the brain and lung
				Poor patient prognosis ⁽¹⁰⁾

Table 1. Subtype and characteristics of healthy and breast cancer cell lines used. *Information obtained from the American Type Culture Collection (ATCC) web site.

Gene name (Gene ID)	Reference Sequences	Oligonucleotide sequence	Amplicon length (bp)
<i>TBP</i> (6908)	NM_003194.5	FW: 5'-ATGACTCCCGGAATCCCTA	232
	NM_001172085.2	RV: 5'-CCACTGCCTGTTGCTGCT	
<i>TBPL1</i> (9519)	NM_001253676.2	FW: 5'-CTTTGGAAGGAGCAAATGTAAT	289
	NM_004865.4	RV: 5'-GCATGAGGTCTATTGTTCTTTG	
<i>TBPL2</i> (387,332)	NM_199047.3	FW: 5'-AGGACAACAGCCCTTATATTTAG	251
		RV: 5'-ATAAGACCAGGAAACAGTTCAG	
<i>RNA18SN1</i>	NR_145820.1	FW: 5'-CCAGTAAGTGCGGGTCATAAG	95
		RV: 5'-GGCCTCACTAAACCATCCAA	
<i>B2M</i>	NM_004048.4	FW: 5'-ATGTCTCGTCCGTGGCCTTAGCT	294
		RV: 5'-ATACTCATCTTTTCAGTGGGGGT	

Table 2. Oligonucleotides used in RT-PCR and RT-qPCR assays. *TBP* and *TBPL1* genes have two messenger RNA sequences in the RefSeq database at NCBI. Therefore, primers were designed to amplify both isoforms for these genes if they are being expressed. FW, forward; RV, reverse.

cancer⁴³, which would position this transcription factor as a tumor suppressor in contrast to studies that have shown that an increase in the expression of this protein promotes cancer progression^{39,40}.

Methods

Cell cultures

Breast cancer cell lines (HCC-1937, MCF7, T47D, SK-BR-3, MDA-MB-231) and control cell lines MCF-12F and MCF-10A (Table 1) were obtained from the American Type Culture Collection (ATCC) and cultured as recommended. The MCF-12F and MCF-10A cell lines were cultivated in Dulbecco's Modified Eagle Medium (DMEM)/ F12 medium (Gibco BRL), supplemented with 10% fetal bovine serum (FBS), 10 ng/μL epidermal growth factor (EGF), 8.36 μg/mL insulin, 0.5 μg/mL hydrocortisone, 1 × L-glutamine, and 1 × penicillin–streptomycin. MDA-MB-231 and MCF-7 cell lines were cultured in DMEM, enriched with 10% FBS, 1 × sodium pyruvate, 1 × L-glutamine, 1 × non-essential amino acids (NEA), and 1 × penicillin–streptomycin solution. An additional 4.18 μg/mL of insulin was added to MCF-7 cells. T47D and HCC1937 cell lines were cultivated in RPMI 1640 medium enriched with 10% FBS, 1 × sodium pyruvate, 1 × L-glutamine, and 1 × penicillin–streptomycin. Lastly, the SK-BR-3 cell line was cultured in McCoy's medium with 10% FBS, 1 × L-glutamine, and 1 × penicillin–streptomycin. All cell lines were incubated at 37 °C in a Shell Lab incubator within a humidified atmosphere containing 5% CO₂ (v/v) until they achieved 80–90% confluence.

mRNA gene expression level of target genes by end-point RT-PCR and RT-qPCR

Total RNA was isolated from cells using the Trizol reagent (Invitrogen, Carlsbad, CA, USA), following the recommended protocol. Its quantity was determined using a Nanodrop 2000c system (Thermo Fisher Scientific). Oligonucleotides, as shown in Table 2, were designed using the PrimerQuest and OligoAnalyzer Tools from Integrated DNA Technologies (<https://www.idtdna.com/pages>) and the sequences were taken from the Reference Sequences (RefSeq) database at the National Center for Biotechnology Information (NCBI), National Library of Medicine of the National Institutes of Health (<https://www.ncbi.nlm.nih.gov/>). The Blast program at NCBI was

employed to verify the oligonucleotides' specificity for the target genes, using the *RNA18SN1* gene as a reference gene⁸⁸. The integrity of RNA was checked by conducting electrophoresis in a 1% agarose gel in 1× Tris Acetate-EDTA (TAE) buffer as described⁸⁹. The gel was stained using 0.5 µg/mL ethidium bromide and band visibility was achieved using a Gel Doc Gel Imager transilluminator (Bio-Rad).

Five micrograms of total RNA from each cell line were digested using 5 units of DNase I RNase-free (Roche, Mexico) at 37 °C for 1 h, after which the enzyme was inactivated at 70 °C for 15 min. The samples were then incubated in a solution containing 250 ng oligo dT₁₈ (SIGMA-ALDRICH), 1× SuperScript II RT buffer (New England BioLabs, Ipswich, Massachusetts), 0.5 mM dNTP mix (Roche), 10 mM DTT (New England BioLabs), and 40 U RNasin Ribonuclease Inhibitor (Promega) at 65 °C for 10 min. Following this, 200 units of SuperScript II Reverse Transcriptase (Invitrogen, Thermo Fisher Scientific) were added, and the mixture was incubated at 50 °C for 1 h. RNA removal was achieved by incubating the solution with 1 µL 0.2 M NaOH at room temperature for 10 min and then adding 2 µL TE at pH 8.0 to neutralize the pH.

The gene expression of target genes was verified using RT-qPCR with 2 µL cDNA solution to perform PCR amplification. This used 2 units per 100 µL of the total reaction volume of One *Taq* DNA polymerase (New England Biolabs), 200 nM of each oligonucleotide (Table 2), 1× *Taq* DNA polymerase buffer, and 2 mM MgCl₂ in a 10 µL final reaction volume. The PCR conditions were set at 94 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 1 min, annealing for 1 min at 58 °C, and extension at 72 °C for 1 min, implemented using an Eppendorf MasterCycler Personal PCR system. Amplicons were then analyzed by electrophoresis in a 1% agarose gel, stained with 0.5 µg/mL ethidium bromide, and visualized on a Bio-Rad Transilluminator (Table 2).

Relative expression of mRNA by RT-qPCR

For RT-qPCR assays, samples (100 ng) of total RNA were used with the KAPA SYBR FAST One-Step kit (Kapa Biosystems, Mexico), following the recommended protocol, using an Eco-Real Time PCR (Illumina) system. The PCR conditions were 50 °C for 30 min for reverse transcription, followed by 95 °C for 10 min, and 40 cycles, each consisting of 95 °C for 10 s, 62 °C for 30 s, and 72 °C for 30 s. The expression level of target genes was normalized with the *RNA18SN1* housekeeping gene. The relative gene expression was measured using the equation $2^{(-\Delta\Delta C_t)}$ with the v4.02 EcoStudy program (Illumina). Appropriate controls (no reverse transcriptase and no template reactions) were included in each 48-well plate. The GraphPad Prism 6.0 version program was used for all statistical analyses, utilizing a non-paired student T-test with Welch's corrections. Significant difference values were displayed as mean ± standard deviation for three biological repeats, each in triplicate.

Subcellular localization of TBPL1 by immunofluorescence and confocal microscopy

Half a million cells for each cell line studied were cultured in 6-well cell culture plates containing sterile glass coverslips (Corning) overnight at 37 °C in an incubation oven. Cells were washed twice with phosphate buffer saline (PBS) 1× for 5 min, then fixed with 2 mL of 4% paraformaldehyde (PFA) for 30 min at room temperature. They were washed three times with PBS 1× for 5 min with gentle shaking on a Rocker 35 EZ shaker. Following this, cells were permeabilized with 0.05% Triton X-100 for 30 min at room temperature, washed three times with PBS 1×, and incubated in 2 mL blocking solution (3% bovine serum albumin in PBS 1×) for 1 h at room temperature.

After washing once with PBS 1×, cells were incubated with a rabbit polyclonal anti-TBPL1 antibody (1:200 dilution) overnight at 4 °C. This antibody was a gift from Dr. László Tora from IGBMC, Strasbourg, France. Cells were then washed five times with PBS 1×, each for 10 min, and incubated with the Fluorescein-conjugated goat anti-rabbit IgG Antibody (1:200 dilution, Vector). Staining was done with Rhodamine-Phalloidin (1:100 dilution, Invitrogen) for 1 h at room temperature.

After incubation, the cells were again washed five times with PBS 1×, for 5 min each, and incubated with Hoechst 33,342 reagent (1.16 µg/mL, Thermo Fisher Scientific) for 3 min at room temperature. After two washes with PBS 1×, each for 5 min, cells were mounted on glass slides with Gelvatol 20/30 mounting medium (Sigma-Aldrich). Finally, cells were examined using a Leica TCS SP8 AOBS multiphoton confocal microscope (Leica Microsystems) at our department.

TBPL1 gene knockout

The CRISPR/Cas9 system was employed to knockout the *TBPL1* gene using a CRISPR all-in-one plasmid, specifically designed for this project by Sigma-Aldrich (Merck) and named the U6-grNA Cas-9-2A + GFP-RM CRISPR plasmid. This plasmid encodes the target guide RNA under the control of the U6 gene promoter and the Cas9 gene, which is fused at its 3'-end with an SV40 nuclear localization signal. The GFP gene and the bGH polyadenylation signal are also embedded, controlled by a human cytomegalovirus (CMV) promoter fused to a CMV enhancer, with the NeoR/KanR gene serving as an antibiotic selection marker. The short guide RNA sequence was chosen via a proprietary Sigma-Aldrich algorithm based on the following criteria: the single-guide RNA (sgRNA) sequence should have a high on-target score⁹⁰, contain a high specificity score (indicating low off-target effects), be complementary to a sequence located on a target-specific exon, and be located close to the N-terminus of the target gene⁸⁶. The target sequence of the ID/Clone name HS00003-0,009,519, corresponding to the *TBPL1* gene, was 5'-GCAGACAGTGATGTTGCAT, and the PAM sequence was TGG, exhibiting a calculated efficiency of 56.23 and specificity of 86.21, according to the program utilized. We confirmed its specificity for the *TBPL1* gene at the RefSeq database using the BLASTN 2.14.1 + program with the non-redundant nucleotide human genome (taxid:9605) database at the BLAST-NCBI web page (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The Non-Homologous End Joining (NHEJ) repair mechanism was chosen due to the error-prone nature of NHEJ, which enables the highly efficient creation of knockout-causing mutations in the target gene. These mutations appear at the site of the targeted double-stranded break more often than at

off-target sites. Generally, off-target cutting constitutes a meager 1% of total activity, while on-target cutting comprises the remaining 99%^{91,92}.

The plasmid was transformed into competent *Escherichia coli* TOP10 cells and was purified using the QIAGEN Plasmid Mega purification kit, as recommended. Fifty thousand cells from each cell line, set in a 24-well plate with sterile glass coverslips (Corning), were transfected with 2.5 µg of the plasmid using Lipofectamine 3000 (Thermo Fisher Scientific), in accordance with the manufacturer's protocol.

All cell lines were incubated at 37 °C in a humidified air atmosphere with 5% (v/v) CO₂ in a Sheldon Shel Lab 3517 Water Jacket CO₂ incubator. After transfecting cells with Lipofectamine 3000, they were incubated with the antibiotic G-418 for 48 h to augment the population of transfected cells in each cell line. Previously, the dosage of the antibiotic was determined by independently performing a kill curve for each cell line. The MCF-12F, MDA-MB-231, T47-D, and SK-BR-3 cell lines were incubated with 800 µg/mL, 1.2 mg/mL, 500 µg/mL, and 450 µg/mL of G-418 antibiotic, respectively. The plasmid transfection and expression of the Cas9-GFP fusion protein in cells were confirmed by the fluorescence emitted from the GFP 48 h post-transfection; this was observed through a Leica TCS SP8 AOBS Multiphoton Confocal Microscope (Leica Microsystems). Total RNA was extracted using the Trizol method as previously described. This was followed by multiple PBS-T washes each lasting 10 min and done under consistent agitation. mRNA expression level was determined by end-point RT-PCR in transfected and untransfected cells, with the expression of the B2M gene coding the β-2-Microglobulin protein used as a control. Amplicons were separated by electrophoresis on 1% agarose gels and were subsequently observed as mentioned earlier.

Detection of TBPL1 by western blot

Fifty thousand cells per well were cultured in 1 mL of supplemented culture medium and transfected with 2.5 µg/well for 48 h, as previously described, within a 24-well plate. Cells were treated with 1 mL of 0.25% trypsin and 0.02% EDTA in 1 × PBS for 1 min at room temperature until detachment from the culture plate. These cells were gathered through centrifugation in 1.5 mL Eppendorf tubes (Axygen) using a 5410 Eppendorf centrifuge for 1 min at room temperature, with the resulting pellet resuspended in 30 µL of 1 × Laemmli buffer. Protein samples (20 µg per lane) were boiled for 2 min before undergoing electrophoresis in a 10% sodium dodecyl sulfate–polyacrylamide gel (SDS-PAGE) using a Mini Protean Tetra Cell system (Bio-Rad). Proteins were consequently transferred to a nitrocellulose membrane (0.22 µm pore size, Amersham) with the help of a Mini Trans-Blot Cell (Bio-Rad) at 4 °C overnight. The membrane was subsequently blocked with a 5% skim milk solution in PBS containing 0.05% Tween 20 in PBS (PBS-T) overnight at 4 °C under gentle shaking, followed by overnight incubation at 4 °C with a 1:1,000 dilution of each anti-human TBPL1 rabbit monoclonal antibody (Boster Bio-Tech) and anti-tubulin mouse monoclonal primary antibody (Invitrogen) diluted in PBS-T with 5% skim milk under gentle shaking. This was followed by multiple PBS-T washes each lasting 10 min and done under consistent agitation. After ten washes, the membrane was incubated for 1 h at room temperature in a 5% skim milk in PBS-T solution containing a 1/20,000 dilution of the peroxidase-linked anti-rabbit IgG antibody (Vector), and with streptactin-peroxidase (Bio-Rad) to elucidate protein markers. This was followed by 1 h-long incubation with the goat HRP-linked anti-mouse IgG antibody (Thermo Fisher Scientific) at a 1:20,000 dilution and at room temperature. After a final series of ten washes in PBS-T, each lasting 10 min under gentle agitation, immunoreactivity was detected through chemiluminescence using the Pierce ECL Western Blotting Substrate Kit (Thermo Fisher Scientific) at a 1:8 substrate-to-buffer ratio. Images were captured with an Odyssey FC Imaging System Bioanalyzer (Licor).

RNA sequencing experimental methods and workflow

Whole transcriptome sequencing was conducted to achieve the gene expression profiles of the studied cell lines. All 24 paired-end samples were successfully analyzed for total RNA, and a single-end analysis was performed on 24 miRNA samples. It is noteworthy that three independent biological experiments were performed, each in triplicate, for RNA seq, RT-qPCR, and RT-PCR. These samples, frozen and shipped in dry ice, were sent to the Macrogen company located in Seoul, South Korea. RNA-Seq analysis was carried out using their standard protocol. Apparatuses such as the TruSeq Stranded Total RNA Library Prep Gold Kit and the Illumina NovaSeq6000 sequencer platform were utilized. The analysis was successfully completed on all 24 paired-end samples with a read length of 151 nucleotides. The data received via USB drive were relevant to all 24 tested samples for future downstream assays.

RNA-seq analysis involved several steps. Firstly, raw paired-end reads were cleaned using Trimmomatic⁹³. Cleaned reads were then mapped to the GRCh38 version of the human genome using the STAR aligner⁹⁴. Counting, normalization, and gene expression quantification were conducted using RSEM⁹⁵. To identify genes with differential expression, we considered those with a FDR less than 0.05 and a fold change of either less than 0.5 or greater than 2.

Downstream analysis for generating plots and illustrating each comparison was conducted using R packages (R 4.2.1). PCA, Venn diagrams, and volcano plots were created using ggplot2 (v 3.4.3), while heatmaps were produced with pheatmap (v.1.0.12) utilizing the K-means algorithm. Functional annotation and analysis were conducted using ClusterProfiler (v.4.6.2) with the Kyoto Encyclopedia of Genes and Genomes (KEGG) serving as the pathway reference (RefSeq (Assembly: GCF_000001405.40), with a significance threshold established at a p-value of <0.05. Protein–protein interaction networks were built with STRING, using stringApp (v.2.0.1) within Cytoscape (v.3.10.0).

The *RNA18SN1* gene functioned as a housekeeping gene to standardize the gene expression of *TBP* and *TBPL1* genes in RT-qPCR experiments. Each experiment was performed in triplicate, under identical conditions and incubation time. The total RNA from these samples was gathered at the same time. All PCR assays were conducted concurrently, also under consistent conditions. The *B2M* gene served as an internal control in PCR

experiments for the confirmation of knockout. Immunofluorescence and confocal microscopy experiments were performed concurrently and under identical conditions. All samples were in triplicate, and both mutated and wild-type cells were incubated for 48 h. For blot (WB) experiments, each lane utilized 20 µg of protein on average (protein content in 50,000 cells) determined with the Bradford Method⁹⁶, and the experiments were performed in triplicate using alpha-actin as an internal control. High-quality RNA was isolated for NGS experiments in triplicate for each cell line, thereby analyzing 24 samples in total. RNA quality was verified by the MacroGen company in Korea, where the dry ice-packed samples were sent for RNA-seq analysis.

It is noteworthy that three independent biological experiments were performed, each in triplicate, for RNA seq, RT-qPCR, and RT-PCR. All images in Figs. 1 and 4 were ready for publication using the PowerPoint software (Microsoft), by cutting and pasting in the final figure and only fitting the size of the images. No other software was used to manipulate images.

MTT cell viability assay

Cell viability was determined through the 3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide (MTT) assay⁹⁷. Fifteen thousand cells in 100 µL of the corresponding culture medium containing 10% FBS per well were inoculated in a 96-well plate (Corning) and incubated for 24 h at 37 °C and under an atmosphere of air containing 5% CO₂. Then, the culture medium was withdrawn from the well, and cells were transfected with 500 ng of the U6-gRNA Cas-9-2A + GFP-RM CRISPR plasmid following the Lipofectamine 3000's protocol (Invitrogen). Proper controls of untransfected cells were also prepared. After transfection, the culture medium was withdrawn from transfected and non-transfected cells and replaced with 100 µL MTT dilution [5 mg/mL in culture medium without FBS, sterilized by filtration through sterile 0.22 µm (pore size), 33 mm in diameter polyethersulfone filtration device (Millipore)], pipette mixed to properly distribute the solvent and incubated for four hours at 37 °C. Then, the MTT solution was withdrawn and washed once with 100 µL PBS 1 ×. Formazan crystals formed by living cells were solubilized with 200 µL isopropanol containing 40 mM HCl for 10 min and shaking at 30 rpm at room temperature. Finally, absorbance at 570 nm was immediately measured using a BioTek Epoch 2 Microplate Spectrophotometer (Agilent). As a reference, absorbance at 630 nm was also determined. As a negative control, the same amount of culture medium without serum was added to a well.

Data availability

The data from this study were submitted to the BioSample database at the National Center for Biotechnology Information. Our submission ID is PRJNA1100513, which will be publicly available (release date) by 2027–03–04. There is no temporarily accessible link for the BioSample database. Sequence Read Archive (SRA) data submission SUB14324503, "Transcriptome analysis of the wild-type and TBPL1 gene KO in healthy and breast cancer cell lines, Mar 19, 2024." Our SRA records will be accessible with the following link after the indicated release date: <https://www.ncbi.nlm.nih.gov/sra/PRJNA1100513>.

Received: 7 March 2024; Accepted: 11 July 2025

Published online: 28 October 2025

References

- Casamassimi, A. & Ciccodicola, A. Transcriptional regulation: Molecules, involved mechanisms, and misregulation. *Int. J. Mol. Sci.* **20**(6), 1281. <https://doi.org/10.3390/ijms20061281> (2019).
- Greber, B. J. & Nogales, E. The structures of eukaryotic transcription pre-initiation complexes and their functional implications. *Subcell Biochem.* **93**, 143–192. https://doi.org/10.1007/978-3-030-28151-9_5 (2019).
- Thomas, M. C. & Chiang, C. M. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41**(3), 105–178. <https://doi.org/10.1080/10409230600648736> (2006).
- Schier, A. C. & Taatjes, D. J. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* **34**(7–8), 465–488. <https://doi.org/10.1101/gad.335679.119> (2020).
- Juven-Gershon, T. & Kadonaga, J. T. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.* **339**(2), 225–229. <https://doi.org/10.1016/j.ydbio.2009.08.009> (2010).
- Müller, F., Zaucker, A. & Tora, L. Developmental regulation of transcription initiation: More than just changing the actors. *Curr. Opin. Genet. Dev.* **20**, 533–540. <https://doi.org/10.1016/j.gde.2010.06.004> (2010).
- Mishal, R. & Luna-Arias, J. P. Role of the TATA-box binding protein (TBP) and associated family members in transcription regulation. *Gene* **833**, 146581. <https://doi.org/10.1016/j.gene.2022.146581> (2022).
- Müller, F. & Tora, L. The multicoloured world of promoter recognition complexes. *EMBO J.* **23**, 2–8. <https://doi.org/10.1038/sj.emboj.7600027> (2004).
- Burley, S. K. X-ray crystallographic studies of eukaryotic transcription initiation factors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **351**(1339), 483–489. <https://doi.org/10.1098/rstb.1996.0046> (1996).
- Akhtar, W. & Veenstra, G. J. C. TBP-related factors: A paradigm of diversity in transcription initiation. *Cell Biosci.* **1**(1), 23. <https://doi.org/10.1186/2045-3701-1-23> (2011).
- Jallow, Z., Jacobi, U. G., Weeks, D. L., Dawid, I. B. & Veenstra, G. J. Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in *Xenopus*. *Proc. Natl. Acad. Sci.* **101**(37), 13525–13530. <https://doi.org/10.1073/pnas.0405536101> (2004).
- Bárfai, R. et al. TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Curr. Biol.* **14**(7), 593–598. <https://doi.org/10.1016/j.cub.2004.03.034> (2004).
- Vogel, J. M., Roth, B., Cigan, M. & Freeling, M. Expression of the two maize TATA binding protein genes and function of the encoded TBP proteins by complementation in yeast. *Plant Cell* **5**(11), 1627–1638. <https://doi.org/10.1105/tpc.5.11.1627> (1993).
- Heard, D. J., Kiss, T. & Filipowicz, W. Both Arabidopsis TATA binding protein (TBP) isoforms are functionally identical in RNA polymerase II and III transcription in plant cells: Evidence for gene-specific changes in DNA binding specificity of TBP. *EMBO J.* **12**(9), 3519–3528. <https://doi.org/10.1002/j.1460-2075.1993.tb06026.x> (1993).
- Savinkova, L. K., Sharypova, E. B. & Kolchanov, N. A. On the Role of TATA Boxes and TATA-Binding Protein in *Arabidopsis thaliana*. *Plants* **12**(5), 1000. <https://doi.org/10.3390/plants12051000> (2023).
- Gazdag, E. et al. TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. *Genes Dev.* **23**(18), 2210–2223. <https://doi.org/10.1101/gad.535209> (2009).

17. Yu, C. et al. TBPL2/TFIIA complex establishes the maternal transcriptome through oocyte-specific promoter usage. *Nat. Commun.* **11**(1), 6439. <https://doi.org/10.1038/s41467-020-20239-4> (2020).
18. Verma, N., Hung, K. H., Kang, J. J., Barakat, N. H. & Stumph, W. E. Differential utilization of TATA box-binding protein (TBP) and TBP-related factor 1 (TRF1) at different classes of RNA polymerase III promoters. *J. Biol. Chem.* **288**(38), 27564–27570. <https://doi.org/10.1074/jbc.C113.5030946> (2013).
19. Holmes, M. C. & Tjian, R. Promoter-selective properties of the TBP-related factor TRF1. *Science* **288**(5467), 867–870. <https://doi.org/10.1126/science.288.5467.867> (2000).
20. Jacobi, U. G. et al. TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation. *EMBO J.* **26**(17), 3900–3909. <https://doi.org/10.1038/sj.emboj.7601822> (2007).
21. Zehavi, Y., Kedmi, A., Ideses, D. & Juven-Gershon, T. TRF2: TRAnsForming the view of general transcription factors. *Transcription* **6**(1), 1–6. <https://doi.org/10.1080/21541264.2015.1004980> (2015).
22. Duttke, S. H., Doolittle, R. F., Wang, Y. L. & Kadonaga, J. T. TRF2 and the evolution of the bilateria. *Genes Dev.* **28**(19), 2071–2076. <https://doi.org/10.1101/gad.250563.114> (2014).
23. Dantonel, J. C., Wurtz, J. M., Poch, O., Moras, D. & Tora, L. The TBP-like factor: An alternative transcription factor in metazoa?. *Trends Biochem. Sci.* **24**(9), 335–339. [https://doi.org/10.1016/s0968-0004\(99\)01436-x](https://doi.org/10.1016/s0968-0004(99)01436-x) (1999).
24. Teichmann, M. et al. Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. *Proc. Natl. Acad. Sci. U. S. A.* **96**(24), 13720–13725. <https://doi.org/10.1073/pnas.96.24.13720> (1999).
25. Catena, R. et al. Proteolytic cleavage of ALF into alpha- and beta-subunits that form homologous and heterologous complexes with somatic TFIIA and TRF2 in male germ cells. *FEBS Lett.* **579**(16), 3401–3410. <https://doi.org/10.1016/j.febslet.2005.04.083> (2005).
26. Chong, J. A. et al. TATA-binding protein (TBP)-like factor (TLF) is a functional regulator of transcription: Reciprocal regulation of the neurofibromatosis type 1 and c-fos genes by TLF/TRF2 and TBP. *Mol. Cell. Biol.* **25**(7), 2632–2643. <https://doi.org/10.1128/MCB.25.7.2632-2643.2005> (2005).
27. Ohbayashi, T. et al. Vertebrate TBP-like protein (TLP/TRF2/TLF) stimulates TATA-less terminal deoxynucleotidyl transferase promoters in a transient reporter assay, and TFIIA-binding capacity of TLP is required for this function. *Nucleic Acids Res.* **31**(8), 2127–2133. <https://doi.org/10.1093/nar/gkg315> (2003).
28. Dantonel, J. C., Quintin, S., Lakatos, L., Labouesse, M. & Tora, L. TBP-like factor is required for embryonic RNA polymerase II transcription in *C. elegans*. *Mol. Cell.* **6**(3), 715–722. [https://doi.org/10.1016/s1097-2765\(00\)00069-1](https://doi.org/10.1016/s1097-2765(00)00069-1) (2000).
29. Kaltenbach, L., Horner, M. A., Rothman, J. H. & Mango, S. E. The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. *Mol. Cell.* **6**(3), 705–713. [https://doi.org/10.1016/s1097-2765\(00\)00068-x](https://doi.org/10.1016/s1097-2765(00)00068-x) (2000).
30. Veenstra, G. J., Weeks, D. L. & Wolffe, A. P. Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in *Xenopus*. *Science* **290**(5500), 2312–2315. <https://doi.org/10.1126/science.290.5500.2312> (2000).
31. Müller, F., Lakatos, L., Dantonel, J., Strahle, U. & Tora, L. TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. *Curr. Biol.* **11**(4), 282–287. [https://doi.org/10.1016/s0960-9822\(01\)00076-8](https://doi.org/10.1016/s0960-9822(01)00076-8) (2001).
32. Martianov, I., Velt, A., Davidson, G., Choukallah, M.-A. & Davidson, I. TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression. *Sci. Rep.* **6**, 32069. <https://doi.org/10.1038/srep32069> (2016).
33. Zhang, D., Penttilä, T. L., Morris, P. L., Teichmann, M. & Roeder, R. G. Spermiogenesis deficiency in mice lacking the Trf2 gene. *Science* **292**(5519), 1153–1155. <https://doi.org/10.1126/science.1059188> (2001).
34. Martianov, I. et al. Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Mol. Cell.* **7**(3), 509–515. [https://doi.org/10.1016/s1097-2765\(01\)00198-8](https://doi.org/10.1016/s1097-2765(01)00198-8) (2001).
35. Martianov, I. et al. Distinct functions of TBP and TLF/TRF2 during spermatogenesis: Requirement of TLF for heterochromatic chromocenter formation in haploid round spermatids. *Development* **129**(4), 945–955. <https://doi.org/10.1242/dev.129.4.945> (2002).
36. Hochheimer, A., Zhou, S., Zheng, S., Holmes, M. C. & Tjian, R. TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* **420**(6914), 439–445. <https://doi.org/10.1038/nature01167> (2002).
37. Kopytova, D. V. et al. Two isoforms of *Drosophila* TRF2 are involved in embryonic development, premeiotic chromatin condensation, and proper differentiation of germ cells of both sexes. *Mol. Cell. Biol.* **26**(20), 7492–7505. <https://doi.org/10.1128/MCB.00349-06> (2006).
38. Isogai, Y., Keles, S., Prestel, M., Hochheimer, A. & Tjian, R. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* **21**(22), 2936–2949. <https://doi.org/10.1101/gad.1608807> (2007).
39. Liu, G. et al. Tumor suppressor microRNA-18a regulates tumor proliferation and invasion by targeting TBPL1 in colorectal cancer cells. *Mol. Med. Rep.* **12**(5), 7643–7648. <https://doi.org/10.3892/mmr.2015.4335> (2015).
40. Xiang, K.-M. & Li, X.-R. MiR-133b acts as a tumor suppressor and negatively regulates TBPL1 in colorectal cancer cells. *Asian Pac. J. Cancer Prev.* **15**(8), 3767–3772. <https://doi.org/10.7314/apjcp.2014.15.8.3767> (2014).
41. Bubola, J. et al. Targeted RNA sequencing in the routine clinical detection of fusion genes in salivary gland tumors. *Genes Chromosomes Cancer* **60**(10), 695–708. <https://doi.org/10.1002/gcc.22979> (2021).
42. Asiedu, M.-K. et al. Pathways impacted by genomic alterations in pulmonary carcinoid tumors. *Clin. Cancer Res.* **24**(7), 1691–1704. <https://doi.org/10.1158/1078-0432.CCR-17-0252> (2018).
43. Qiu, X. et al. Mendelian randomization reveals potential causal relationships between cellular senescence-related genes and multiple cancer risks. *Commun. Biol.* **7**(1), 1069. <https://doi.org/10.1038/s42003-024-06755-9> (2024).
44. Fogh, J., Fogh, J. M. & Orfeo, T. One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice. *J. Natl. Cancer Inst.* **59**(1), 221–226. <https://doi.org/10.1093/jnci/59.1.221> (1977).
45. Zhang, B. et al. Secreted insulin-like growth factor binding protein 5 functions as a tumor suppressor and chemosensitizer through inhibiting insulin-like growth factor 1 receptor/protein kinase B pathway in acute myeloid leukemia. *Neoplasia* **47**, 100952. <https://doi.org/10.1016/j.neo.2023.100952> (2024).
46. Müller, A. et al. Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**(6824), 50–56. <https://doi.org/10.1038/35065016> (2005).
47. Saha, S. K. et al. KRT19 directly interacts with β -catenin/RAC1 complex to regulate NUMB-dependent NOTCH signaling pathway and breast cancer properties. *Oncogene* **36**(3), 332–349 (2016).
48. Saha, S. K., Yin, Y., Chae, H. S. & Cho, S.-G. Opposing regulation of cancer properties via KRT19-mediated differential modulation of wnt/ β -catenin/notch signaling in breast and colon cancers. *Cancers (Basel)* **11**(1), 99. <https://doi.org/10.3390/cancers11010099> (2019).
49. Itoh, S., Itoh, F., Goumans, M., Ten, J. & Dijke, P. Signaling of transforming growth factor-beta family members through Smad proteins. *Eur. J. Biochem.* **267**(24), 6954–6967. <https://doi.org/10.1046/j.1432-1327.2000.01828.x> (2000).
50. Miyazono, K., Dijke, P. T. & Heldin, C. H. TGF- β signaling by Smad proteins. *Adv. Immunol.* **75**, 115–157. [https://doi.org/10.1016/s0065-2776\(00\)75003-6](https://doi.org/10.1016/s0065-2776(00)75003-6) (2000).
51. Gupta, S. C., Kim, J. H., Prasad, S. & Aggarwal, B. B. Regulation of survival, proliferation, invasion, angiogenesis, and metastasis of tumor cells through modulation of inflammatory pathways by nutraceuticals. *Cancer Metastasis Rev.* **29**(3), 405–434. <https://doi.org/10.1007/s10555-010-9235-2> (2010).
52. Derynck, R. & Zhang, Y. E. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* **425**(6958), 577–584. <https://doi.org/10.1038/nature02006> (2003).

53. Takenobu Katagiri, T. & Watabe, T. Bone morphogenetic proteins. *Cold Spring Harb. Perspect. Biol.* **8**(6), a021899. <https://doi.org/10.1101/cshperspect.a021899> (2016).
54. Emma-Leena Alarmo, E.-L. et al. BMP7 influences proliferation, migration, and invasion of breast cancer cells. *Cancer Let.* **275**(1), 35–43. <https://doi.org/10.1016/j.canlet.2008.09.028> (2009).
55. Rodriguez-Martinez, A. et al. Analysis of BMP4 and BMP7 signaling in breast cancer cells unveils time-dependent transcription patterns and highlights a common synexpression group of genes. *BMC Med. Genom.* **4**, 80. <https://doi.org/10.1186/1755-8794-4-80> (2011).
56. Johnstone, C. N. et al. Functional and genomic characterisation of a xenograft model system for the study of metastasis in triple-negative breast cancer. *Dis. Model. Mech.* **11**(5), 032250. <https://doi.org/10.1242/dmm.032250> (2018).
57. Ehata, S. & Miyazono, K. Bone morphogenetic protein signaling in cancer; some topics in the recent 10 years. *Front. Cell Dev. Biol.* **10**, 883523. <https://doi.org/10.3389/fcell.2022.883523> (2022).
58. Oh, C. et al. Epigenetic upregulation of MAGE-A isoforms promotes breast cancer cell aggressiveness. *Cancers* **13**(13), 3176. <https://doi.org/10.3390/cancers13133176> (2021).
59. Maine, E. A. et al. The cancer-testis antigens SPANX-A/C/D and CTAG2 promote breast cancer invasion. *Oncotarget* **7**(12), 14708–147026. <https://doi.org/10.18632/oncotarget.7408> (2016).
60. Naik, A., Lattab, B., Qasem, H. & Decock, J. Cancer testis antigens: Emerging therapeutic targets leveraging genomic instability in cancer. *Mol. Ther. Oncol.* **32**(1), 200768. <https://doi.org/10.1016/j.omton.2024.200768> (2024).
61. Liu, M. et al. LAMC2 promotes the proliferation of cancer cells and induce infiltration of macrophages in non-small cell lung cancer. *Ann. Transl. Med.* **9**(17), 1392. <https://doi.org/10.21037/atm-21-4507> (2021).
62. Garg, M., Braunstein, G. & Koeffler, H. P. LAMC2 as a therapeutic target for cancers. *Expert Opin. Ther. Targets.* **18**(9), 979–982. <https://doi.org/10.1517/14728222.2014.934814> (2018).
63. Schenk, S. et al. Binding to EGF receptor of a laminin-5 EGF-like fragment liberated during MMP-dependent mammary gland involution. *J. Cell Biol.* **161**(1), 197–209. <https://doi.org/10.1083/jcb.200208145> (2003).
64. Yang, J., Niu, H., Huang, Y. & Yang, K. A systematic analysis of the relationship of CDH13 promoter methylation and breast cancer risk and prognosis. *PLoS ONE* **11**(5), e0149185. <https://doi.org/10.1371/journal.pone.0149185> (2016).
65. Che, D. et al. KRT6A promotes lung cancer cell growth and invasion through MYC-regulated pentose phosphate pathway. *Front. Cell Dev. Biol.* **9**, 694071. <https://doi.org/10.3389/fcell.2021.694071> (2021).
66. Chang, H.-H., Dreyfuss, J. M. & Ramoni, M. F. A transcriptional network signature characterizes lung cancer subtypes. *Cancer* **117**(2), 353–360. <https://doi.org/10.1002/cncr.25592> (2011).
67. Yang, B. et al. KRT6A Promotes EMT and cancer stem cell transformation in lung adenocarcinoma. *Technol. Cancer Res. Treat.* **19**, 1533033820921248. <https://doi.org/10.1177/1533033820921248> (2020).
68. Yao, S. J., Ma, H.-S., Liu, G.-M., Gao, Y. & Wang, W. Increased IL-1 α expression is correlated with bladder cancer malignant progression. *Arch. Med. Sci.* **19**(1), 10. <https://doi.org/10.5114/aoms.2020.100677> (2020).
69. Wu, S., Du, Y., Beckford, J., Beckford, J. & Alachkar, H. Upregulation of the EMT marker vimentin is associated with poor clinical outcome in acute myeloid leukemia. *J. Transl. Med.* **16**(1), 170. <https://doi.org/10.1186/s12967-018-1539-y> (2018).
70. Rohini, M., Vairamani, M. & Selvamurugan, N. TGF- β 1-stimulation of NFATC2 and ATF3 proteins and their interaction for matrix metalloproteinase 13 expression in human breast cancer cells. *Int. J. Biol. Macromol.* **192**, 1325–1330. <https://doi.org/10.1016/j.jbiomac.2021.10.099> (2021).
71. Guan, X. et al. Androgen receptor activity in T cells limits checkpoint blockade efficacy. *Nature* **606**(7915), 791–796. <https://doi.org/10.1038/s41598-019-48064-w> (2022).
72. Kannan, A. et al. Cancer testis antigen promotes triple negative breast cancer metastasis and is traceable in the circulating extracellular vesicles. *Sci. Rep.* **9**(1), 11632. <https://doi.org/10.1038/s41598-019-48064-w> (2019).
73. Cisneros-Villanueva, M. et al. LINC00460 Is a dual biomarker that acts as a predictor for increased prognosis in basal-like breast cancer and potentially regulates immunogenic and differentiation-related genes. *Front. Oncol.* **11**, 628027. <https://doi.org/10.3389/fonc.2021.628027> (2021).
74. Jacenik, D., Hikisz, P., Beswick, E. J. & Fichna, J. The clinical relevance of the adhesion G protein-coupled receptor F5 for human diseases and cancers. *Biochim. Biophys. Acta Mol. Basis Dis.* **1869**(5), 166683. <https://doi.org/10.1016/j.bbadis.2023.166683> (2023).
75. Lim, J. H. et al. Activation of human cancer/testis antigen gene, XAGE-1, in tumor cells is correlated with CpG island hypomethylation. *Int. J. Cancer.* **116**(2), 200–206. <https://doi.org/10.1002/ijc.21007> (2005).
76. Dheeraj, A., Khamar, B. M., Shukla, C. P., Jain, N. K. & Modi, R. I. Desmocollin-3 and tumor immune microenvironment: Bioinformatics evaluation. *Cancer Res.* **83**(7_Supplement), 881–881. <https://doi.org/10.1158/2326-6074.TUMIMM24-B029> (2023).
77. Li, M. et al. Membrane metalloendopeptidase (MME) suppresses metastasis of esophageal squamous cell carcinoma (ESCC) by inhibiting FAK-RhoA signaling axis. *Am. J. Pathol.* **189**(7), 1462–1472. <https://doi.org/10.1016/j.ajpath.2019.04.007> (2019).
78. Bögershausen, N. et al. RAP1-mediated MEK/ERK pathway defects in Kabuki syndrome. *J. Clin. Invest.* **125**(9), 3585–3599. <https://doi.org/10.1172/JCI80102> (2015).
79. Brat, D. J. et al. Transgenic expression of dominant negative tuberin through a strong constitutive promoter results in a tissue-specific tuberous sclerosis phenotype in the skin and brain. *J. Biol. Chem.* **280**(7), 5870–5874. <https://doi.org/10.1074/jbc.M411768200> (2005).
80. Looi, C.-K., Hii, L.-W., Ngai, S. C., Leong, C.-O. & Mai, C.-W. The role of ras-associated protein 1 (Rap1) in cancer: Bad actor or good player?. *Biomedicine* **8**(9), 334. <https://doi.org/10.3390/biomedicine8090334> (2020).
81. Song, S. et al. Rap1-mediated nucleosome displacement can regulate gene expression in senescent cells without impacting the pace of senescence. *Aging Cell* **19**(1), e13061. <https://doi.org/10.1111/acer.13061> (2019).
82. Ramamurthy, V. et al. The structure of the death receptor 4-TNF-related apoptosis-inducing ligand (DR4-TRAIL) complex. *Acta Crystallogr. F Struct. Biol. Commun.* **71**(Pt 10), 1273–1281. <https://doi.org/10.1107/S2053230X15016416> (2015).
83. O'Sullivan, J. et al. ZNF432 stimulates PARylation and inhibits DNA resection to balance PARPi sensitivity and resistance. *Nucleic Acids. Res.* **51**(20), 11056–11079. <https://doi.org/10.1093/nar/gkad791> (2023).
84. Zhang, Z. et al. High expression of hexokinase domain containing 1 is associated with poor prognosis and aggressive phenotype in hepatocarcinoma. *Biochem Biophys. Res. Commun.* **474**(4), 673–679. <https://doi.org/10.1016/j.bbrc.2016.05.007> (2016).
85. Vigil, D., Cherfils, J., Rossman, K. L. & Der, C. J. Ras superfamily GEFs and GAPs: Validated and tractable targets for cancer therapy?. *Nat. Rev. Cancer* **10**(12), 842–857. <https://doi.org/10.1038/nrc2960> (2010).
86. Kimura, A. et al. Nuclear heat shock protein 110 expression is associated with poor prognosis and hyperthermo-chemotherapy resistance in gastric cancer patients with peritoneal metastasis. *World J. Gastroenterol.* **23**(42), 7541–7550. <https://doi.org/10.3748/wjg.v23.i42.7541> (2017).
87. Bellah, S. F., Salam, M. A., Billah, S. M. S. & Karim, M. R. Genetic association in CYP3A4 and CYP3A5 genes elevate the risk of prostate cancer. *Ann. Hum. Biol.* **50**, 63–74. <https://doi.org/10.1080/03014460.2023.2171122> (2023).
88. Koo, G. C. et al. Janus kinase 3-activating mutations identified in natural killer/T-cell lymphoma. *Cancer Discov.* **2**(7), 591–597. <https://doi.org/10.1158/2159-8290.CD-12-0028> (2012).
89. Fu, J. J. et al. Systemic inflammation is associated with differential gene expression and airway neutrophilia in asthma. *OMICS* **17**, 187–199. <https://doi.org/10.1089/omi.2012.0104> (2013).
90. Aranda, P. S., LaJoie, D. M. & Jorcyk, C. L. Bleach gel: A simple agarose gel for analyzing RNA quality. *Electrophoresis* **33**(2), 366–369. <https://doi.org/10.1002/elps.201100335> (2012).

91. Gandhi, S., Haeussler, M., Razy-Krajka, F., Christiaen, L. & Stolfi, A. Evaluation and rational design of guide RNAs for efficient CRISPR/Cas9-mediated mutagenesis in *Ciona*. *Dev. Biol.* **425**(1), 8–20. <https://doi.org/10.1016/j.ydbio.2017.03.003> (2017).
92. Naeem, M., Majeed, S., Hoque, M. Z. & Ahmad, I. Latest developed strategies to minimize the off-target effects in CRISPR-cas-mediated genome editing. *Cells* **9**(7), 1608. <https://doi.org/10.3390/cells9071608> (2020).
93. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> (2014).
94. Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinf.* **51**, 11.14.1–11.14.19. <https://doi.org/10.1002/0471250953.bi1114s1> (2015).
95. Lim, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* **12**, 323. <https://doi.org/10.1186/1471-2105-12-323> (2011).
96. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254. <https://doi.org/10.1006/abio.1976.9999> (1976).
97. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods*. **65**(1–2), 55–63. [https://doi.org/10.1016/0022-1759\(83\)90303-4](https://doi.org/10.1016/0022-1759(83)90303-4) (1983).

Acknowledgements

The authors are deeply grateful to The Consejo Nacional de Humanidades, Ciencia y Tecnología (Conahcyt, now renamed as Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti)) from Mexico for the financial support of this work with the grant Frontier Science 2019_116337 awarded to Dr. Juan Pedro Luna-Arias and Dr. Erika Beatriz Angeles-Morales. We are also grateful to the Department of Cell Biology at Cinvestav for the support of this project. We also thank Conahcyt for the Ph.D. fellowships 465228 and 752614 granted to Rabia Mishal and Edgar Hernández-Martínez, respectively. We are also grateful for the donation of the Shell Lab incubator and some culture media by the Laboratorio de Especialidades Inmunológicas, S.A. de C. (LEI).

Author contributions

RM, EBAM, and JPLA, designed the experimental approach, analyzed the results, and wrote the manuscript. JMZ, BRZ, RM, and JPLA, performed the bioinformatics analysis of RNAseq data. MLLB, CACS, LPM, cultured cells, performed immunofluorescence and confocal microscopy, transfected the cells, WBs, RT-qPCRs. APR, JMHH, GLA, performed analysis of the cytoskeleton, discussed data, and reviewed the manuscript. SUC, designed experiments and discussed results. RM, IKA, EHM, and JPLA, reviewed the manuscript and prepared figures.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-11634-2>.

Correspondence and requests for materials should be addressed to J.P.L.-A.

Reprints and permissions information is available at www.nature.com/reprints.

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

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

© The Author(s) 2025