Expression and Function of C1orf132 Long-Noncoding RNA, in Breast Cancer Cell Lines and Tissues

Afsaneh Malekzadeh Shafaroudi
Department of Biology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad
Ali Sharifi-Zarchi
Department of Computer Engineering, Sharif University of Technology, Tehran
Saeid Rahmani
Department of Computer Engineering, Sharif University of Technology, Tehran
Nahid Nafisi
Surgical Department, School of Medicine, Iran University of Medical Sciences, Tehran
Seyed Javad Mowla
Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran
Andrea Lauria
Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy
Salvatore Oliviero
Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy
Maryam Moghadam Matin (matin@um.ac.ir)
Department of Biology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad

Research Article

Keywords: bioinformatics, C1orf132, RNA

Posted Date: January 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-144862/v1

Abstract

MIR29B2CHG/C1orf132 is the host gene for generating miR-29b2 and miR-29c. Here, we employed bioinformatics and experimental approaches to decipher expression of C1orf132 in breast cancer cells and tissues. Our data demonstrated a significant downregulation of C1orf132 in triple-negative breast cancer. We also predicted a putative promoter for the longer transcripts of C1orf132. The functionality of the distal promoter was confirmed by transfecting MCF7 cells with a C1orf132 promoter-GFP construct. Knocking-out the promoter by means of CRISPR/Cas9 approach revealed no expression alteration of neighboring genes, CD46 and CD34. However, the expression of miR-29c was reduced by half, suggesting an enhancer effect of the distal promoter on miR-29c generation. Furthermore, the promoter knock-out an elevation of migration ability in MCF12A edited cells. Moreover, the expressions of cell mobility genes e.g., CDH2, FGF2, FGFR1 and the stem cell and EMT-associated transcription factor ZEB1 were significantly elevated in edited cells. RNA sequencing data on the edited and unedited cells revealed many up- and down-regulated genes involved in various cellular pathways, including epithelial to mesenchymal transition and mammary gland development pathways. Altogether, we are reporting the existence of an additional/distal promoter with an enhancer effect on miR-29 generation and an inhibitory effect on cell migration.

Introduction

Breast cancer (BC) is the leading cause of cancer related death in women worldwide, with an estimated 2,100,000 new cases and 627,000 deaths in 2018[1,2]. In Iran, the age distribution of BC is nearly 10 years lower than their counterparts in many countries[3]. Despite the increased incidence rate, BC is highly curable, if diagnosed and appropriately treated at early stages. Clinically, classification of BC is based on histopathological findings of the patients' tissue samples obtained through surgery or biopsy. This includes the assessment of the receptor status of tumor cells namely estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2)[4,5]. One important subtype of BC is the triple negative BC (TNBC), in which neither of the aforementioned receptors are expressed. TNBC tumors make up 10–30% of all breast cancers and are reported to be the most aggressive form of BC with high histological grade, increased mitotic count, central necrosis and margins of invasion. It is also associated with younger age of incidence and poorer prognosis, in comparison to non-TNBC patterns[6], as well as an association with shorter duration of breastfeeding, higher age at first pregnancy and lower parity[7].

Ductal carcinoma *in situ* (DCIS) is a pre-invasive form of BC, which 20–50% of them may progress towards invasive ductal carcinoma (IDC) and spread to other tissues during progression of the disease. However, there is no way to determine which lesions would be stable without treatment and which ones will become invasive and require chemotherapy[8].

Long non-coding RNAs (lncRNAs) are important but poorly conserved transcribed RNA molecules with at least 200 nt in length which are involved in complex biological as well as pathological processes. Their

mechanism of action in gene expression regulation can be summarized in four different ways: I) serving as a signal molecule, II) acting as a molecular decoy and/or sponge, III) guiding some changes in *cis* or *trans* gene expression, and IV) acting as scaffolds[9,10]. Dysregulation of several lncRNAs has already been reported in breast cancer. They exert either oncogenic or tumor-suppressive functions with significant impacts on various biological and pathological processes, especially in the progression of malignant tumors[11].

C1orf132 is an intergenic lncRNA located on 1q32.2 between *CD34* and *CD46* protein coding genes. *C1orf132* (also known as *MIR29B2CHG*) is the host-gene for miR-29b2 and miR-29c with a molecular size more than 20 kb. Its promoter is more likely to be hypermethylated in basal-like breast tumors, comparing to the low grade ones[12]. Bhardwaj and colleagues used next generation RNA sequencing technique on a breast cancer progression cell line model and discovered that the tumor suppressor miR-29c is highly down-regulated during the course of breast cancer cells progression toward TNBC. Moreover, their data revealed that higher miR-29c expression, is associated with a longer survival rate and lower distant metastasis[13].

Here, we explored the expression pattern of *C1orf132* in different types of breast cancer cell lines and tissues. We also discovered a distal promoter area with an enhancing function on miR-29b2/c expression, and an inhibitory effect on cell cycle and cell migration.

Results

C1orf132 locus and its transcripts.

C1orf132, also known as the host gene for miR-29b2 and miR-29c, is located between two protein coding genes, *CD34* and *CD46* (Fig. 1). There are several transcripts with various lengths and number of exons of *MIR29B2CHG* located between the two aforementioned coding genes. No ortholog gene for *C1orf132* has been reported in mouse, mostly because lncRNAs are not conserved during evolution. For this reason, we searched for potential lncRNAs overlapping with miR-29b2 and miR-29c in mouse genome. Interestingly, we discovered a lncRNA with conserved chromosomal location and transcriptional direction, in reference to its human counterpart (Fig. 1).

The human MiR precursor has 5 or 6 exons with its promoter located (~20 kb) upstream of the miR29b2c (Fig. 2; indicated with blue color). In addition, we noticed several spliced variants of the gene, some started at apparently different start sites, located upstream of the known promoter for miR-29 host gene variant. The latter finding suggested that the lncRNA could bear some more roles, rather than just being served as a precursor to generate miR-29. Using bioinformatics tools such as Fantom5 supported the existence of at least two distal potential start sites (p1 and p2) for the transcription of long forms of *C1orf132*, with the putative p2 promoter containing a CpG island (Fig. 2A-C). ENCODE data exhibited the expression and chromatin modifications (H3K27ac and H3K4me3) of C1orf132 locus in MCF7 cells, while the active marks were mostly concentrated on two start sites. One for the promoter of miR-29

precursor and the other one for the potential promoter of the longer transcript (Fig. 2D-G). According to the ENCODE ChIP-seq data, the latter promoter could interact with more than 80 different transcription factors (Fig. 2H-J).

C1orf132 was significantly down-regulated in TNBC tissue samples.

Expression analysis of different breast cancer subtypes in TCGA revealed a downregulation of the longer transcripts of *C1orf132* in TNBC patients. According to this data, the expression level of C1orf132 transcripts originating from p2 promoter were significantly decreased in ER/PR negative and TNBC patients, compared to the ER/PR positive as well as non-TNBC ones (Fig. 3A). To confirm this experimentally, we performed qRT-PCR and measured the level of C1orf132 long isoform in tumor samples along with their adjacent non-tumor tissues. The real-time PCR results on 52 paired tissue samples confirmed the same pattern of significant (p<0.05) down-regulation of *C1orf132* in TNBC vs. non-TNBC samples (Fig. 3B). However, we found no significant correlation between this isoform and other subtypes of breast cancer (data not shown).

C1orf132 transcribed from the p2 promoter localizes in the nucleus.

To confirm the potential activity of distal promoters, two reporter constructs were generated, namely pEGFP-1-p1 and pEGFP-1-p2 (Fig. 4A). The constructs were then used for transient transfection of MCF7 cells, along with pEGFP-C1 vector as a positive control (Fig. 4B). The level of transfection efficiency was assessed by the percentage of fluorescent cells. The presence of the GFP signal revealed the activity of the putative p2 promoter, in contrast to the predicted p1 promoter, which failed to show any activity (Fig. 4C, 4D(. Motivated with this finding, we decided to do functional analysis on the p2 promoter by deletion of this region, and to investigate its potential effects on various cellular behaviors as well as transcriptome via RNA-seq technique.

To investigate the subcellular localization of C1orf132 in MCF12A cells, we amplified the RNA extracts fractioned by Ambion cell fractionation buffer. qRT-PCR data on nuclear and cytoplasmic fractions demonstrated that C1orf132 was prevalently present within the nucleus ($p<0.05$). U1 and B2M were used as controls for nuclear and cytoplasmic compartments, respectively (Fig. 4E).

Considering the activity of p2 region, we decided to target it with three pairs of guide RNAs (Fig. 5A) after transfecting gRNAs to MCF12A cells, the best sharp PCR product of 438 bps (caused by a genomic deletion of about 2000 bps) belonged to guides number 1&3 (Fig. 5B). Genomic real-time PCR on MCF12A and edited colonies confirmed that the colonies were knocked down and that the colony number 5 had the lowest expression of edited promoter (Fig. 5C).

Flow cytometry analysis revealed a G2/M arrest during the cell cycle progression of the edited compared to unedited MCF12A cells (p<0.05 Fig. 6A and 6B). Data obtained from three different biological repeats demonstrated that the putative p2 promoter knock-down could delay the cell cycle progression and significantly increased the number of cells in G2/M phase.

Promoter deletion enhanced the migration ability of MCF12A cells.

As seen in Fig. 6, the initial scratch width was monitored in wound healing experiments in edited (top) vs. unedited MCF12A cells (bottom). The results obtained at time points 0 and 24 h revealed that the migration ability is significantly improved in edited cells, compared to the unedited cells (p<0.05; Fig. 6C and 6D).

Gene expression profiling after *C1orf132* knock down

The gene expression profiles obtained by RNA-seq in two biological replicates of edited and wild type cells demonstrated 1565 genes with differential expression (adjusted p<0.001), with 551 upregulated and 1014 downregulated genes (Supplementary Fig. 1A). Moreover, two dimensional PCA graph of the data for the two replicates of *C1orf132* edited and unedited MCF12A cells showed a nicely separated replicates of edited cells from wild type cells (Supplementary Fig. 1B). Accordingly, the same data emerged from the heatmap of differentially expressed genes between edited and unedited MCF12A cells (Supplementary Fig. 1C).

As lncRNAs could have a general enhancer role on neighboring genes, we first looked at any potential altered expression of upstream and downstream genes of the edited region. According to the RNA-seq analysis, the two neighboring genes, *CD34* and *CD46*, were not affected in edited cells, however, the expression of mir29b2c precursor declined in edited cells compared to the wild type cells (Fig. 6E). According to GO results, the differential expression analysis of the RNA-seq data revealed several up- and down-regulated genes involved in various biological processes, including the positive regulation of vascular endothelial cell proliferation and regulation of endothelial cell chemotaxis to fibroblast growth factor biological processes. *FGF2* and *FGFR1* genes which are responsible for the above-mentioned processes, were upregulated more than 4- fold in edited cells (Fig. 6F). FGF2 is a tumor cell survival factor that helps the cells to escape apoptosis. Moreover, in edited cells the expression fold change for *CDH2* (N-Cadherin), as a mesenchymal marker, was upregulated more than twelve folds, compared to the unedited cells. In contrast, the expression of *CDH1* (E-Cadherin), as an epithelial marker, was downregulated in edited cells (near zero) compared to the wild type MCF12A cells. Here, *ZEB1* as an EMT transcription factor showed about 15-fold upregulation in edited cells (Fig. 6F). *ESR1*, the gene coding for estrogen recepror alpha (ERα), is under-expressed along with gap junction β-2 (GJB2/CX26) and *WNT4* in edited MCF12A cells to near zero.

Discussion

lncRNAs are the largest class of transcripts in human cells, and also the least understood ones. Among various mechanisms attributed to them, being the host of small RNAs, including miRNAs, is the best-known function of these regulatory molecules. However, it is not clear if this type of lncRNAs have additional functions other than being simply the host genes for small RNAs. *C1orf132* or *MIR29B2CHG* is the host gene of two miRNAs, miR-29b2 and miR-29c. Here, we investigated if this lncRNA could have other functions in human breast cancer cells.

There are few reports on *C1orf132* gene, and its cellular role. It has been reported as one of the five candidate genes which their CpG islands methylation was used to predict individual's age in forensic medicine[14,15]. Indeed, *C1orf132* gene provided the highest accuracy in predicting the age[16].

In contrast to coding genes, long non-coding genes have low conservation during evolution, which makes it difficult to find their orthologs in other species. Here, we used a simple logic to find *C1orf132* ortholog gene in mouse by looking at genomic location of mouse miR-29b2 and miR-29c. As we expected, there was a lncRNA gene in mouse hosting both miRNAs. Surprisingly, *C1orf132* and its mouse counterpart have an exact chromosomal position, locating between *CD34* and *CD46* genes, and are transcribed in the same direction, considering their neighboring genes. Logically, there is no need for positional genomic conservation of ortholog miRNA host genes. Indeed, this evolutionary conservation suggests that *C1orf132* might have other functions, in addition to being a host gene for miR-29b2c.

Our *in silico* studies revealed two potential promoters, in addition to the previously known miRNAs precursor promoter[12], as well as two groups of long and short transcripts for *C1orf132*. The reason for generating short and long transcripts of the gene is not known, however, we had a similar pattern of short and long transcripts for *PSORS1C3* in a previous study[17]. There is a possibility that this kind of gene expression regulation by alternative splicing in a tissue- and cell-specific manner is a more general feature of lncRNAs and their possible dysregulation could contribute to pathogenesis of many cancers[18]. To explore potential functions of the long transcripts of *C1orf132*, we suppressed their expression by excising the distal promoter by means of CRISPR/Cas9 method. It has been already claimed that lncRNAs could have an enhancer effect on their neighboring genes[19]. However, we observed no changes in the expression level of *CD34* and *CD46* in edited cells, both in our RNA-seq and real-time PCR data; ruling out an enhancer role of these transcripts on their neighboring genes. Interestingly, there existed a significant downregulation of miR-29c-3p in edited cells, suggesting an enhancer role of distal promoter on proximal promoter of the gene. The later finding makes it very difficult to interpret the RNA-seq results of the edited cells. Indeed, it is impossible to discriminate the direct effects of C1orf132 from an indirect effect mediated by altered expression of miR-29b2 and miR-29c.

In a bioinformatics analysis of 20 published studies, Yan and colleagues reported that the low expression of miR-29a/b/c is associated with poor prognosis of malignant neoplasms, and could be used as a key biomarker to predict cancers progression and recurrence[20]. Furthermore, these miRNAs exhibit both tumor suppressive and oncogenic roles in different cancers. This might be due to the differences in sample types, location, size, or the period of follow-up study[20]. For example, miR-29c was reported to be upregulated in non-small cell lung cancer (NSCLC) tissues, whereas it was downregulated in lung cancer patients subjected to radiotherapy[21,22].

We also observed an enhanced migration ability in edited cells, suggesting an invasion suppressor role for C1orf132. This is in agreement with Zhang et al. report in which a significant increase of invading cells emerged in miR-29 suppressed HepG2 cells[23]. Jiang et al. reported that miR-29c was remarkably

decreased in pancreatic cancer cells and that it had association with shorter overall survival and tumor recurrence in pancreatic cancer patients[24]. It has been well documented that miR-29c is lost as early as the preneoplastic stage of TNBC tumorigenesis, where its downregulation could cause a worse overall survival via regulating several target genes[13].

EMT is characterized by a loss of function of the E-cadherin adhesion protein (*CDH1*) in epithelial tissues; and breast cancer is served as a good example of this process. In addition to their role in normal cells, classical cadherins (E- and N-cadherins) have a distinguished role in transforming malignant cells as well as tumor progression, in a process termed 'cadherin switching'[25,26]. This process is regulated by EMT transcription factors, including ZEB family. ZEB1 and ZEB2 are EMT transcription factors, which synergistically increase tumor invasion and cell migration[27]. ZEB1 helps epigenetic silencing of *CDH1*, by bringing several enzymes to the *E-cadherin* promoter for epigenetic induction or by inhibiting the expression of stemness-repressing miRNAs to cause dynamic transition of non-cancerous stem cells into cancer stem cells (CSCs) and vise versa[28,29]. In this study, we demonstrated that a cadherin switching evidently took place through a downregulation of *CDH1* and an upregulation of *CDH2* by almost 12-fold, as well as by almost 14-fold overexpression of *ZEB1*.

Interaction of N-cadherin (CDH2) and FGF receptor (FGFR) leads to their stabilization on the cell surface. In this way, N-cadherin increases tumor cell interactions with endothelial and mesenchymal cells[25]. On the other hand, fibroblast growth factor 2 (FGF2) which can induce proliferation of neoplastic cell and tumor hormone-independent growth[30], is a tumor cell survival factor, which inhibits cell apoptosis through an autocrine secretory loop. In addition, FGF2 can be an activator of angiogenesis during tumor mass growth and metastasis[31]. In this study, we demonstrated that *FGF2* and *FGFR1* were differentially over expressed in edited MCF12 cells by almost four folds. By these results, the pathway analysis of our data by Enrichr was nicely validated for up-regulated genes.

On the contrary, we evaluated the expression of *ESR1*, *GJB2* and *WNT4* genes using qRT-PCR and found out that these genes are expressed at very low (near to zero) levels, in edited cells. GJB2 and ESR1 with a role in "mammary gland development pathway-pregnancy and lactation" according to WikiPathways 2019 Human were among the downregulated differentially expressed genes in edited vs. wild type MCF12A. Estrogen receptor α (encoded by *ESR1*), which is expressed in normal breast epithelium has an essential growth and differentiation role and also is associated with the growth and survival of breast epithelial cancer cells[32] as well as a part at progression initiation of estrogen-dependent breast cancers[33]. As mentioned earlier, a significant progressive loss of miR-29c, as a result of breast cancer cell line progression model to TNBC, was expected[13]. Therefore, a diminution of *ESR1* expression is in line with the miR-29c deduction which we have shown in our edited cells. We also observed a significant reduction of GJB2 in our edited cells. *GJB2* encodes a member of the gap junction protein family. GJB2 protein is often found in the epithelial cells of the mammary duct (luminal) and can form a gap connection channel in the form of homodimers or heterodimers[34]. This protein in breast cancer was shown to be a tumor-

suppressor gene and its very low expression in breast tumor tissues might be associated with hypermethylation[35].

Our results revealed a down regulation of *WNT4* in our *C1orf132* knocked-down MCF12A cells. Wnt-4 is a member of the WNT family, which encode important signaling proteins and play a vital role in some developmental and oncological processes. Several reports have shown the association of Wnt-4 to epithelial cells, like its expression in normal murine keratinocytes[36] and also in the epidermal cells throughout skin development[37] as well as its expression necessity for normal mammary gland development[38]. Seitoh and colleagues reported that in keratinocyte cell lines, a poor differentiation and more malignant phenotype is related to the loss of Wnt-4 gene expression. In another study, a similar association between Wnt4 and mesenchymal to epithelial transition during kidney development has been nicely shown[39]. In contrast, Vouyovitch and colleagues reported that depletion of WNT4 in MCF12 cells inhibited cellular proliferation, and upregulated WNT4 in non-malignant breast cells, stimulated growth, inhibited apoptosis and increased cell migration and EMT[40].

As the sole reference on C1orf132 lncRNA and cancer, Peng et al. detected two lncRNAs (C1orf132 and TMPO-AS) as a prognostic signature for lung adenocarcinoma patients, which was validated in two independent datasets, and also by the GO enrichment analysis. By pathway analysis, they suggested that down regulation of *C1orf132* was associated with a poor prognosis in lung cancer patients, probably by deregulating the “cell cycle and cell adhesion molecules” pathways in cancer cells[41]. Finally, our qRT-PCR data on *C1orf132* knocked-down cells were in line with the RNA-seq and bioinformatics data. Altogether, our data suggest a dysregulated cell cycle and cell adhesion in breast cell line, which might be the cause of aberrant proliferation and migration of edited MCF12 cells.

Methods

Ethical statement.

This research involved collecting human tissues from Khatam-ol-Anbia and Rasule-Akram hospitals, with no experimenting on human subjects or animals. *In vitro* experiments on commercial cell lines and pathological samples were approved as a Ph.D. thesis proposal, by Ferdowsi University of Mashhad (code number: IR.UM.REC.1399.104).

Clinical tissue samples.

In this study, a total number of 52 pathological tissue samples with a diagnosis of primary breast cancer were collected. None of the patients had been treated with preoperative radiotherapy, chemotherapy, or other relevant modalities. Breast cancer tissue samples along with their matched adjacent apparently normal tissues were collected, immediately preserved in liquid nitrogen, and then stored at -80ºC until analysis. All patients signed informed consents and agreed to the use of their surgical specimens for research.

Bioinformatics analysis.

We used UCSC genome browser to scan the genomic area around *C1orf132* (located at chr1: 207,978,592-208,052,441 (hg19)) for potential promoter activity. Histone modifications by ChIP-seq from ENCODE (H3K27ac and H3K4me3) and DNA-seq were used to detect the active area near transcription start site (TSS). Using ENCODE ChIP-seq data for different transcription factors (TFs), a list of TFs which bind to *C1orf132* potential promoters was extracted. Fantom5 data were used to determine which genomic regions the reads originated from.

To evaluate the correlation of C1orf132 transcripts originating from the putative p2/distal promoter with the tumor state of breast tissues, we performed data mining of The Cancer Genome Atlas (TCGA) breast cancer subtypes to study their association with the expression level of *C1orf132*.

Cloning the putative promoters of C1orf132.

The potential promoter regions (p1 and p2) for C1orf132, according to the bioinformatics analysis, were amplified from genomic DNA using Accu *Taq* polymerase kit (Invitrogen, USA) and specific primers (supplementary Table 1). The amplified products (1203 bp for p1 and 2143 bp for p2), were cloned into the promoterless green fluorescent protein (pEGFP-1) reporter vector, using flanking sequences on primers for *Age*I/*Sal*I restriction sites (Supplementary Table 1). The accuracy of the cloned constructs was confirmed by DNA sequencing (Europhins, Italy).

Cell culture and transfection.

MCF7 and MCF12A were obtained from IIGM cell bank. MCF7 cell line was cultured in high glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), and 1% penicillin/streptomycin. MCF12A cells were cultivated in a special medium containing DMEM/F12 (Thermo Fisher Scientific, USA), 5% heat inactivated horse serum (Gibco, USA), 20 ng/ml recombinant human EGF (AF100-15 Peprotech, USA), 0.5 µg/ml hydrocortisone, 100 ng/ml cholera toxin (Vibrio cholerae C8052, Sigma, USA), 10 µg/ml human recombinant insulin (Zinc solution 12585-014, Gibco, USA) and 1% penicillin/streptomycin, and incubated at 37ºC with 5% humified CO_2. To examine the promoter activity of the cloned regions, the recombinant vectors were used to transfect MCF7 cells. All experiments were done with complete medium in at least triplicates.

Fractionation assay.

We examined the nuclear vs. cytoplasmic subcellular localization of C1orf132 in MCF12A cells by fractionation with cell fractionation buffer (Ambion, USA), according to the manufacturer's instructions. RNA was extracted to assess the relative proportion of C1orf132 in the nuclear and cytoplasmic fractions. The transcription levels of beta 2-microglobulin (*B2M*) as a cytoplasmic marker, U1 as a nuclear marker and C1orf132 were then assessed using quantitative reverse transcription- polymerase chain reaction (qRT-PCR).

Promoter activity reporter assay.

MCF7 cells were seeded in 48-well plates (SPL Life Sciences, South Korea). After complete adhesion, cells were transfected with 1 µg of vector (pEGFP-1-p1 or pEGFP-1-p2) using Lipofectamine LTX & PLUS reagent (Invitrogen, USA), according to the manufacturer's instructions. pEGFP-1 is a promoterless vector, containing eGFP reporter. 48 h after transfection and under a fluorescent microscope, the presence of the GFP signal was monitored in MCF7 cells transfected with pEGFP-C1, as the positive control, and pEGFP-1-p1 or pEGFP-1-p2 vectors.

Promoter deletion using CRISPR/Cas9 system.

In order to suppress *C1orf132* expression, we decided to delete its p2 promoter using CRISPR/Cas9 system. Three pairs of different guide RNAs (gRNAs) were designed (Supplementary Table 1) to target the putative sequences of the p2 transcription start site, using http://crispr.mit.edu. The three gRNAs where then cloned into TOPO-TA gRNA vector. Briefly, guide RNA plasmid backbone (TOPO-TA gRNA) was linearized using the *Bbs*I enzyme (New England Biolabs, USA) and digested at 37˚C for 1 h. Annealing took place using sense and anti-sense oligonucleotides in buffer 2 (NEB), before placing the reaction in a thermocycler with the ramp of 0.1˚C/sec from 95˚C to 25˚C. To ligate the gRNA within the linearized vector, T4 DNA ligase (Thermo Fisher Scientific, USA) was used and the reaction was incubated at room temperature for 60 min. Then, *E. coli* cells were transformed with gRNA vectors using heat-shock protocol followed by plating the cells on LB/ampicillin (100 µg/ml) plates overnight. Colony-PCR was performed to verify the sequence of the gRNA plasmids, using the reverse oligo of guides and forward primer of TOPO vector.

Different pairs of gRNA-containing vectors were co-transfected along with wild type-Cas9 vector (PX458) into MCF12A cells by lipofectamine LTX & PLUS reagent (Invitrogen, USA). Single cells expressing GFP were selected by a cell sorter machine (BD FACSCelesta), 24 h after transfection. Edited colonies were investigated for deletion, by DNA extraction using Puregene Core kit A (Qiagen, USA), according to the manufacturer's instructions, followed by performing PCR using flanking primers (Edited-Test) and DNA sequencing (Europhins, Italy).

Wound healing assay.

Edited and unedited MCF12A cells were seeded into 24-well plates (with 6 repeats) and grew to 90% confluency. The monolayers were scratched using a 200 µl pipette tip and then the floating cells were removed by several washes with phosphate buffered saline (PBS). Subsequently, the cells were incubated at 37ºC for 24 h, before being photographed. The migration area of wound healing was border-lined and analyzed, using the ImageJ software.

Cell cycle analysis by flow cytometry.

Triplicates of three different densities (10000, 20000, and 30000) of edited vs. unedited MCF12A cells were seeded into 12-well plates. The Vybrant™ DyeCycle™ Violet Stain kit (Invitrogen, USA), which is capable of entering living cells and staining DNA, was used to examine the cell cycle 24 h later. Briefly, cells were washed in cold PBS and then harvested, and resuspended in the complete medium. Next, flow cytometry tubes each containing 1 ml of cell suspension in complete medium at a concentration of 1×10^6 cells/ml were prepared. 1 µl of Vybrant™ DyeCycle™ Violet Stain was added to each tube (final concentration of 5 µM) and mixed well. After 1 h incubation at 37˚C with protection from light, the samples were analyzed without washing or fixing on a flow cytometer (BD FACSCelesta) using laser beam.

RNA extraction and qRT-PCR.

A small amount of frozen patient tissue samples was lysed for RNA extraction using RNSol (ROJE, Iran)/or QIAZOL (Qiagen, USA). Total RNA was extracted according to the manufacturer's instructions. 1 µg of RNA was first treated with DNase I (Thermo Fisher Scientific, USA) in order to eliminate any traces of DNA contamination, and then reverse transcribed using PrimeScript first strand cDNA synthesis kit (TaKaRa, Japan). The expression of target genes was evaluated using BioFACT™ 2X real-time PCR master mix (BioFact, South Korea) through ABI StepOne real-time PCR system. The sequence of primers used for quantifying each target gene can be found in Supplementary Table 1. Relative expression of target genes to *B2M* was calculated according to $2^{-\Delta\Delta Ct}$ method. For the cells, the expression of target genes was evaluated using SensiFAST SYBR No-ROX one-step kit (Bioline. USA Cat. No: BIO-72001), according to the manufacturer's instructions. 30 ng of RNA was used in one step real-time PCR using Rotor-Gene instrument (Qiagen, USA).

TaqMan miR assay for miR-29c-3p detection.

Hsa-miR-29c-3p (ID 000587), and U6 snRNA (ID 001973) TaqMan miR assays were ordered from Applied Biosystems (Foster City, CA). Total RNA extraction followed by qRT-PCR assay were performed to determine the expression level of miR-29c-3p in edited versus unedited MCF12A cells, according to the manufacturer's protocol. Briefly, total RNA extraction was performed by Qiazol (Qiagen, USA), followed by reverse transcription using SuperScript II Reverse Transcriptase (Thermo Fisher Scientific, USA), before doing qPCR using Taqman® Universal PCR Master Mix II (Thermo Fisher Scientific, USA). The samples were then incubated for 10 min at 95ºC for initial denaturation, and then subjected to 40 PCR cycles, each consisting of 95ºC for 15 sec and 60ºC for 60 sec. U6 was used as the internal control of miR-29c-3p. The $2^{-\Delta\Delta Ct}$ method was used to analyze microRNA levels.

Statistical analysis.

All statistical analyses were performed using GraphPad Prism 6 software. Student *t*-test was employed to investigate the significance of observed differences of gene expression alterations. All tests were done in 3 biological repeats and values are reported as mean± standard deviation (SD). *P* values less than 0.05 were considered as statistically significant.

Declarations

Acknowledgments:

We are grateful to Professor Daniela Taverna and Ms. Safoora Torkashvand for their sincere helps in sample collections and experimental procedures. This work was supported by research grants from Ferdowsi University of Mashhad (Grant number 47470), and NIMAD (Grant number 976872).

M.M. designed experiment, performed the neural differentiation part and edit the manuscript; S.J.M. supervised the research, designed the experiments, analyzed data and edit the final draft of the manuscript.

Author contributions:

A.M.S. designed and performed experiments, analyzed data and wrote the first draft of the manuscript

A.S. Z. designed the bioinformatics studies, interpret the data, edit the manuscript.

S.R. Performed bioinformatics studies, interpret the data

N. N. provides the clinical samples and clinicopathologic information of the patients, edit the manuscript

S. J. M. analyzed data and edit the final draft of the manuscript.

A. L. performed and analyzed the RNA seq part and edit the manuscript,

S. O. supervised the research, designed the experiments, analyzed data and edit the final draft of the manuscript.

M. M. M. supervised the research, designed the experiments, analyzed data and edit the final draft of the manuscript.

Competing Interests

The authors declare no competing interests.

References

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA. Cancer J. Clin.* **68**, 7–30 (2018).
3. Sharifian, A. *et al.* Burden of Breast Cancer in Iranian Women is Increasing. *Asian Pac J Cancer Prev.* **16**, 5049–5052 (2015).
4. Curtis, C. *et al. The Cancer Genome Atlas Network. Nature.* **490**, 61–70 (2012).

5. Vuong, D., Simpson, P. T., Green, B., Cummings, M. C. & Lakhani, S. R. Molecular classification of breast cancer. *Virchows Arch.* **465**, 1–14 (2014).
6. Hon, J. D. C. *et al.* Breast cancer molecular subtypes: From TNBC to QNBC. *Am. J. Cancer Res.* **6**, 1864–1872 (2016).
7. Nafissi, N. *et al.* Relationships between Reproductive Risk Factors for Breast Cancer and Tumor Molecular Subtypes. *Multidiscip Cancer Investig.* **19**, 1767–1770 (2018).
8. Lesurf, R. *et al.* Molecular Features of Subtype-Specific Progression from Ductal Carcinoma In Situ to Invasive Breast Cancer. *Cell Rep.* **16**, 1166–1179 (2016).
9. Wang, K. C. & Chang, H. Y. *NIH Public Access.* **43**, 904–914 (2012).
10. Ling, H. *et al.* Junk DNA and the long non-coding RNA twist in cancer genetics. *Oncogene.* **34**, 5003–5011 (2015).
11. Wu, Y. *et al.* The role of lncRNAs in the distant metastasis of breast cancer. *Front. Oncol.* **9**, 407–420 (2019).
12. Poli, E. *et al.* Molecular subtype-specific expression of MicroRNA-29c in breast cancer is associated with CpG dinucleotide methylation of the promoter. *PLoS One.* **10**, 1–14 (2015).
13. Bhardwaj, A. *et al.* Regulation of miRNA-29c and its downstream pathways in preneoplastic progression of triple-negative breast cancer. *Oncotarget.* **8**, 19645–19660 (2017).
14. Freire-Aradas, A. *et al.* Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system. *Forensic Sci. Int. Genet.* **24**, 65–74 (2016).
15. Zbieć-Piekarska, R. *et al.* Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Sci. Int. Genet.* **17**, 173–179 (2015).
16. Jung, S. E. *et al.* DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples. *Forensic Sci. Int. Genet.* **38**, 1–8 (2019).
17. Mirzadeh Azad, F., Malakootian, M., Mowla, S. J. & lncRNA PSORS1C3 is regulated by glucocorticoids and fine-tunes OCT4 expression in non-pluripotent cells. *Sci. Rep.* **9**, 1–9 (2019).
18. Ma, X. *et al.* Overexpressed long noncoding RNA CRNDE with distinct alternatively spliced isoforms in multiple cancers. *Front. Med.* **13**, 330–343 (2019).
19. Dennis, M. K. *et al.* Long non-coding RNAs with enhancer-like finction in humans. *Cell.* **127**, 358–366 (2012).
20. Palareti, G. *et al.* Comparison between different D-Dimer cutoff values to assess the individual risk of recurrent venous thromboembolism: Analysis of results obtained in the DULCIS study. *Int. J. Lab. Hematol.* **38**, 42–49 (2016).
21. Zhu, W. *et al.* Expression of miR-29c, miR-93, and miR-429 as potential biomarkers for detection of early stage non-small lung cancer. *PLoS One.* **9**, 1–7 (2014).

22. Arechaga-Ocampo, E. *et al.* Tumor suppressor miR-29c regulates radioresistance in lung cancer cells. *Tumor Biol.* **39**, (2017).
23. Zhang, N. S., Dai, G. L. & Liu, S. J. MicroRNA-29 family functions as a tumor suppressor by targeting RPS15A and regulating cell cycle in hepatocellular carcinoma. *Int. J. Clin. Exp. Pathol.* **10**, 8031–8042 (2017).
24. Jiang, J. *et al.* Reduction of miR-29c enhances pancreatic cancer cell migration and stem cell-like phenotype. *Oncotarget.* **6**, 2767–2778 (2015).
25. Wheelock, M. J., Shintani, Y., Maeda, M., Fukumoto, Y. & Johnson, K. R. Cadherin switching. *J. Cell Sci.* **121**, 727–735 (2008).
26. Afzal Ashaie, M., Hoque Chowdhury, E. & Cadherins The Superfamily Critically Involved in Breast Cancer. *Curr. Pharm. Des.* **22**, 616–638 (2016).
27. Zhang, Y., Xu, L., Li, A. & Han, X. The roles of ZEB1 in tumorigenic progression and epigenetic modifications. *Biomed. Pharmacother.* **110**, 400–408 (2019).
28. Wellner, U. *et al.* The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat. Cell Biol.* **11**, 1487–1495 (2009).
29. Chaffer, C. *et al.* Poised chromatin at the ZEB1 promoter enables cell plasticity and enhances tumorigenicity. *Cell.* **154**, 61–74 (2013).
30. Giulianelli, S. *et al.* FGF2 induces breast cancer growth through ligand-independent activation and recruitment of ERα and PRBΔ4 isoform to MYC regulatory sequences. *Int. J. Cancer.* **145**, 1874–1888 (2019).
31. Yu, P. J., Ferrari, G., Galloway, A. C., Mignatti, P. & Pintucci, G. Basic fibroblast growth factor (FGF-2): The high molecular weight forms come of age. *J. Cell. Biochem.* **100**, 1100–1108 (2007).
32. Carausu, M. *et al.* ESR1 mutations: a new biomarker in breast cancer. *Expert Rev. Mol. Diagn.* **19**, 599–611 (2019).
33. Tomita, S. *et al.* Estrogen receptor a gene ESR1 amplification may predict endocrine therapy responsiveness in breast. *Cancer science.* **100** (6), 1012–1017 (2009).
34. Teleki, I. *et al.* Correlations of differentially expressed gap junction connexins cx26, cx30, cx32, cx43 and cx46 with breast cancer progression and prognosis. *PLoS One.* **9**, 1–12 (2014).
35. Shettar, A., Damineni, S., Mukherjee, G. & Kondaiah, P. Gap junction β-2 expression is negatively associated with the estrogen receptor status in breast cancer tissues and is a regulator of breast tumorigenesis. *Oncol. Rep.* **40**, 3645–3653 (2018).
36. Saitoh, A., Hansen, L. A., Vogel, J. C. & Udey, M. C. Characterization of Wnt gene expression in murine skin: Possible involvement of epidermis-derived Wnt-4 in cutaneous epithelial-mesenchymal interactions. *Exp. Cell Res.* **243**, 150–160 (1998).
37. Reddy, S. *et al.* Characterization of Wnt gene expression in developing and postnatal hair follicles and identification of Wnt5a as a target of Sonic hedgehog in hair follicle morphogenesis. *Mech. Dev.* **107**, 69–82 (2001).

38. Heikkil, M., Peltoketo, H. & Vainio, S. Wnts and the female reproductive system. *J. Exp. Zool.* **290**, 616–623 (2001).
39. Stark, K., Vainio, S., Vassileva, G. & McMahon, A. P. Epithelial transformation of metanephric mesenchyme in the developing kidney regulated by Wnt-4. *Nature.* **372**, 679–683 (1994).
40. Vouyovitch, C. M., Perry, J. K., Liu, D. X., Bezin, L. & Vilain, E. WNT4 mediates the autocrine effects of growth hormone in mammary carcinoma cells. *Endocr. Relat. Cancer.* **23** (7), 571–585 (2016).
41. Peng, F. *et al.* Differential expression analysis at the individual level reveals a lncRNA prognostic signature for lung adenocarcinoma. *Mol. Cancer.* **16**, 1–12 (2017).

Figures

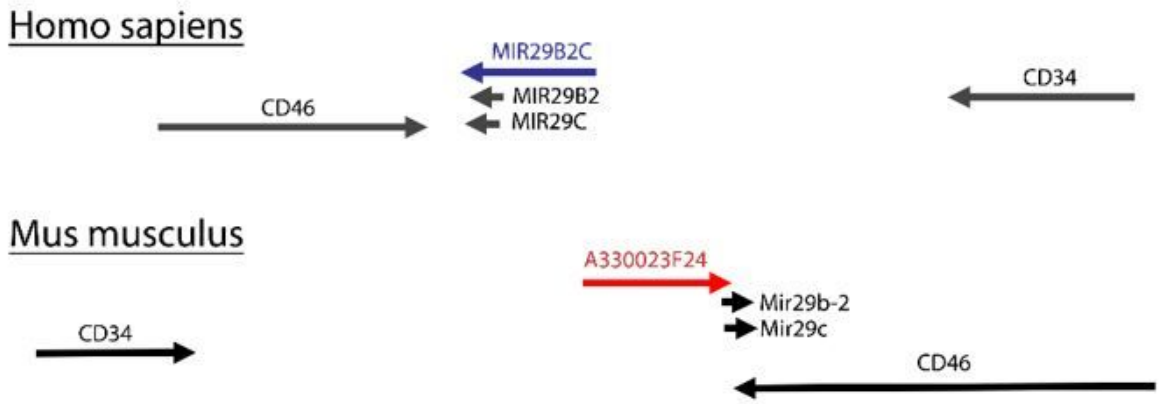

Figure 1

Genomic location of C1orf132 long-noncoding RNA in human and mouse. Top: The genomic location of miR-29b2, miR-29c and their precursor, MIR29B2CHG, in human. As shown, the region encoding the lncRNA is located between CD34 and CD46 genes. Bottom: By searching the genomic location of miR-29 in mouse, we found out that the lncRNA A330023F24 is the host gene for miR-29b-2 and miR-29c. Interestingly the chromosomal location and transcription direction of the lncRNA is conserved between human and mouse.

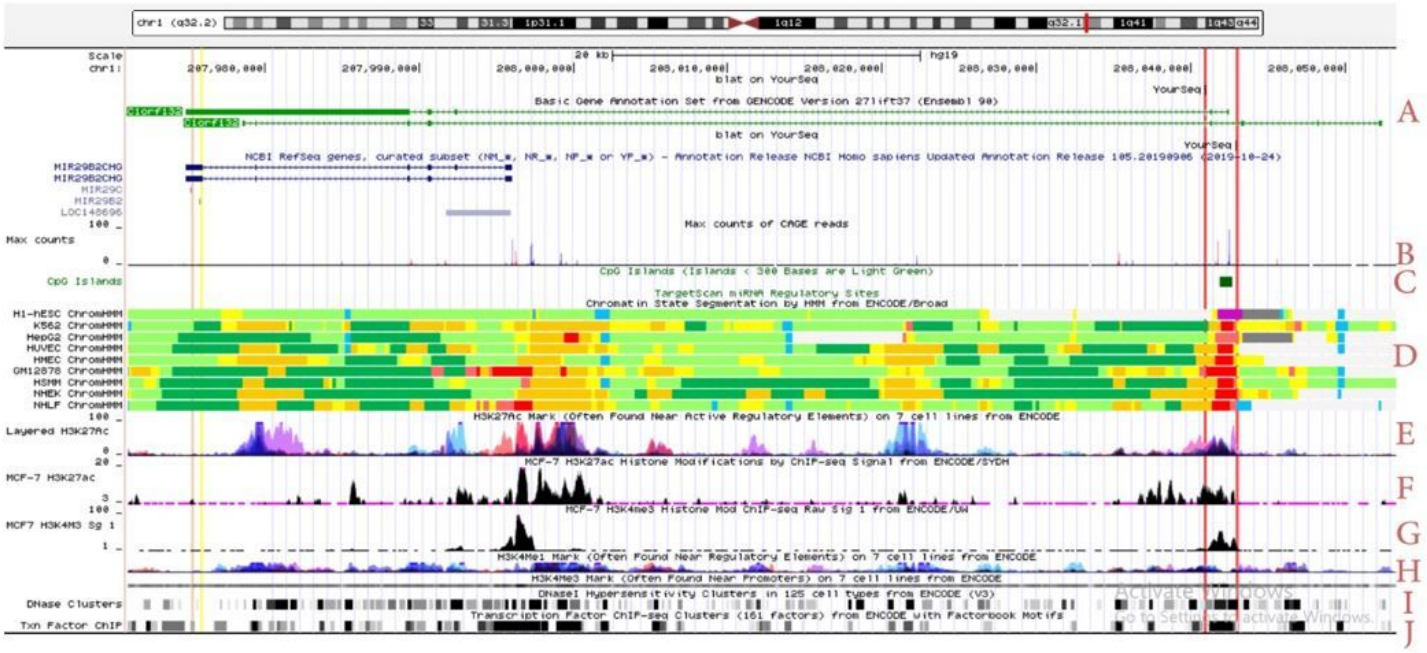

Figure 2

The genomic features of miR29b2c precursor. To make it easier to follow, the location of miR-29b2 and miR-29c are highlighted with yellow and purple colors, respectively, and the two potential distal transcription start sites of MIR29B2CHG with red color. As it is evident in the figure, there are several spliced variants for miR29b2c precursor. In addition to the miR29b2c precursor (short variants), there are long variants of C1orf132 (A). Other bioinformatics features (UCSC Genome Browser hg19) support the existence of other start sites for C1orf132 transcription, including: Max counts of CAGE reads belonging to FANTOM5 summary (B), CpG island (C), active promoter predicted element according to Broad ChromHMM (D), active or potentially active regulatory element of layered and MCF7 cells H3K27ac mark (E & F), active or potentially active promoter and enhancer regions marks by H3K4me3 and H3K4me1, respectively (G & H), DNase cluster (I), and transcription factor ChIP (J).

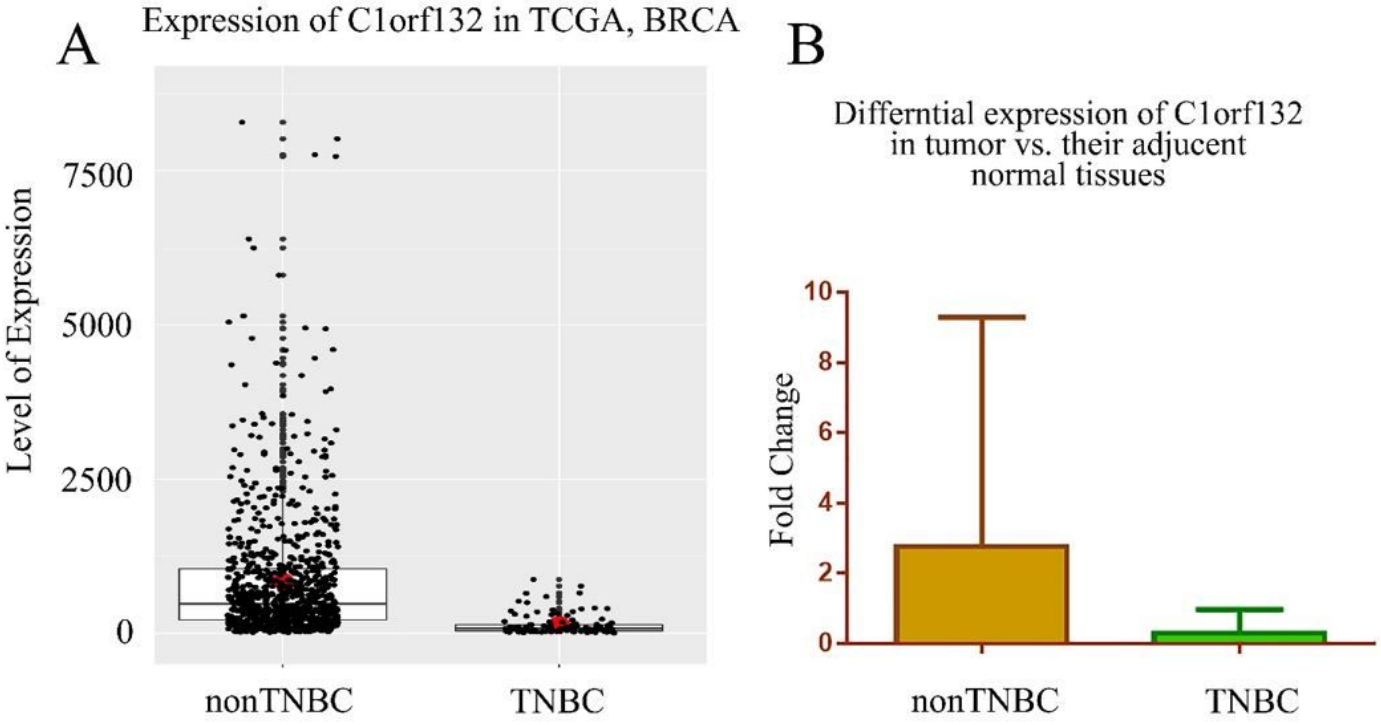

Figure 3

Downregulation of C1orf132 in TNBC tissue samples. A) The data mining of TCGA was performed for breast cancer samples for the expression of long isoform of C1orf132 transcribed from the potential distal promoter (ENST00000608023.5_1) on the (–) strand. According to this data, the expression of C1orf132 was significantly decreased in Triple negative breast cancer (TNBC) samples of BRCA compared to the non-TNBC samples. B) The qRT-PCR results on 52 fresh tumor and their adjacent apparently normal breast tissue samples demonstrated a significant reduction of C1orf132 expression in TNBC, compared to the non-TNBC tumor tissues (P value < 0.05).

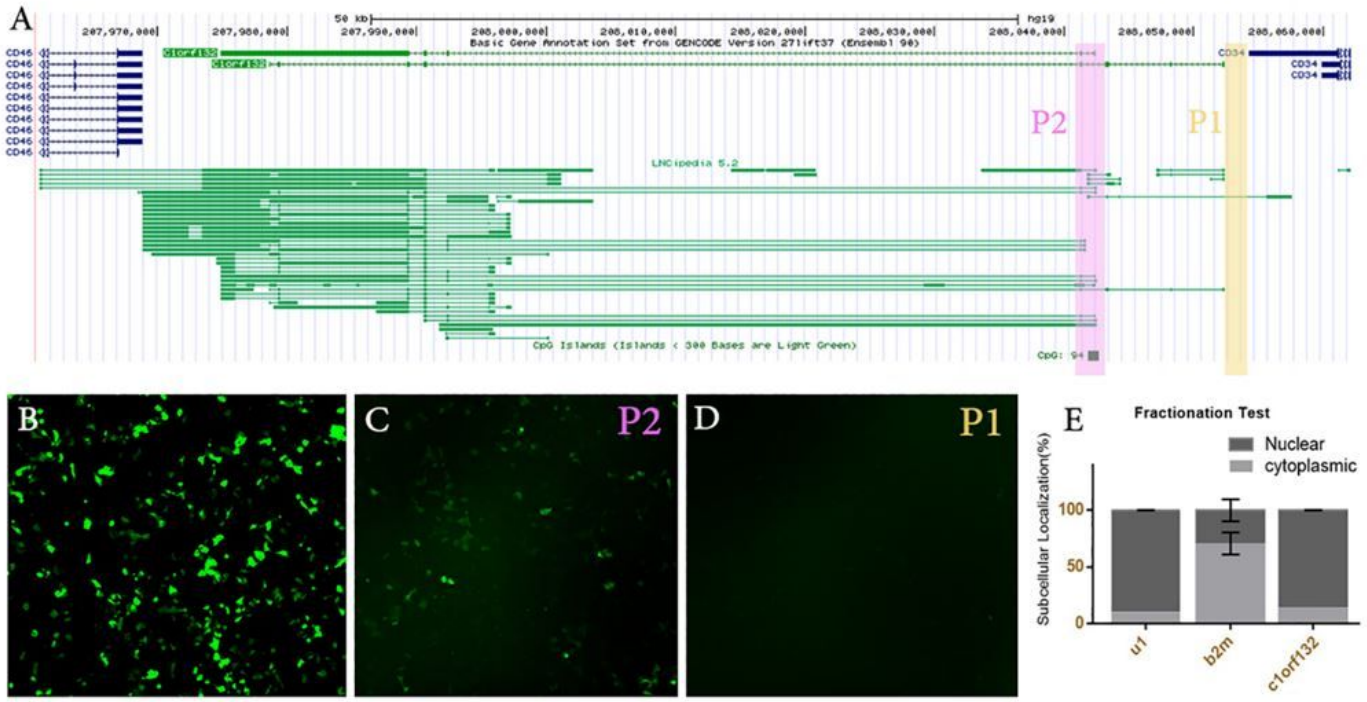

Figure 4

C1orf132 has at least two active promoters. A) The bioinformatics data suggest the existence of two additional potential promoters (p1 and p2), in addition to its conventional promoter. The predicted sequences correlated to the putative promoters were cloned into a promoterless pEGFP-1 vector. MCF7 cells transfected with: B) pEGFP-C1 vector as a positive control, C) pEGFP-1-p2 for the putative p2 promoter region, D) pEGFP-1-p1 for the putative p1 promoter region. Note that only p2 region exhibited promoter activity. E) The subcellular localization of C1orf132 following nuclear and cytoplasmic subcellular fractionation. C1orf132 is primarily localized within the nuclear compartment. U1 and B2M were used as markers for nuclear and cytoplasmic compartments, respectively.

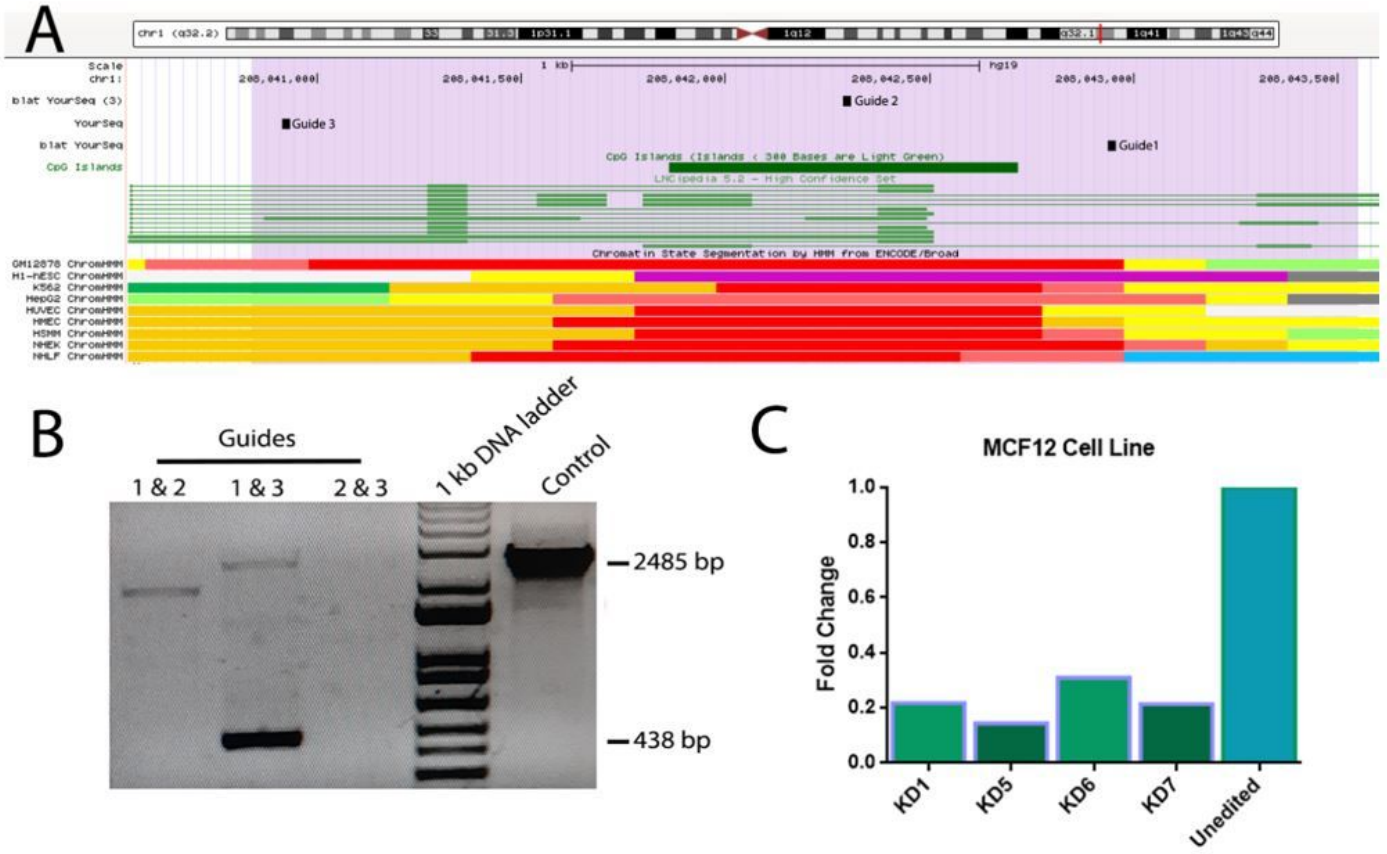

Figure 5

Removing C1orf132 p2 genomic region using three different guide RNAs. A) The genomic locations targeted with guide RNAs are depicted as guides 1-3 on putative p2 region (highlighted by purple color), B) Following transfection of different pairs of guide RNAs into MCF12A cells, use of gRNAs 1&3 led to generation of a sharp PCR product of 438 bps, caused by a genomic deletion of about 2000 bps. C) Genomic DNA real-time PCR for C1orf132 on 4 different MCF12A edited cell colonies, in comparison to unedited cells.

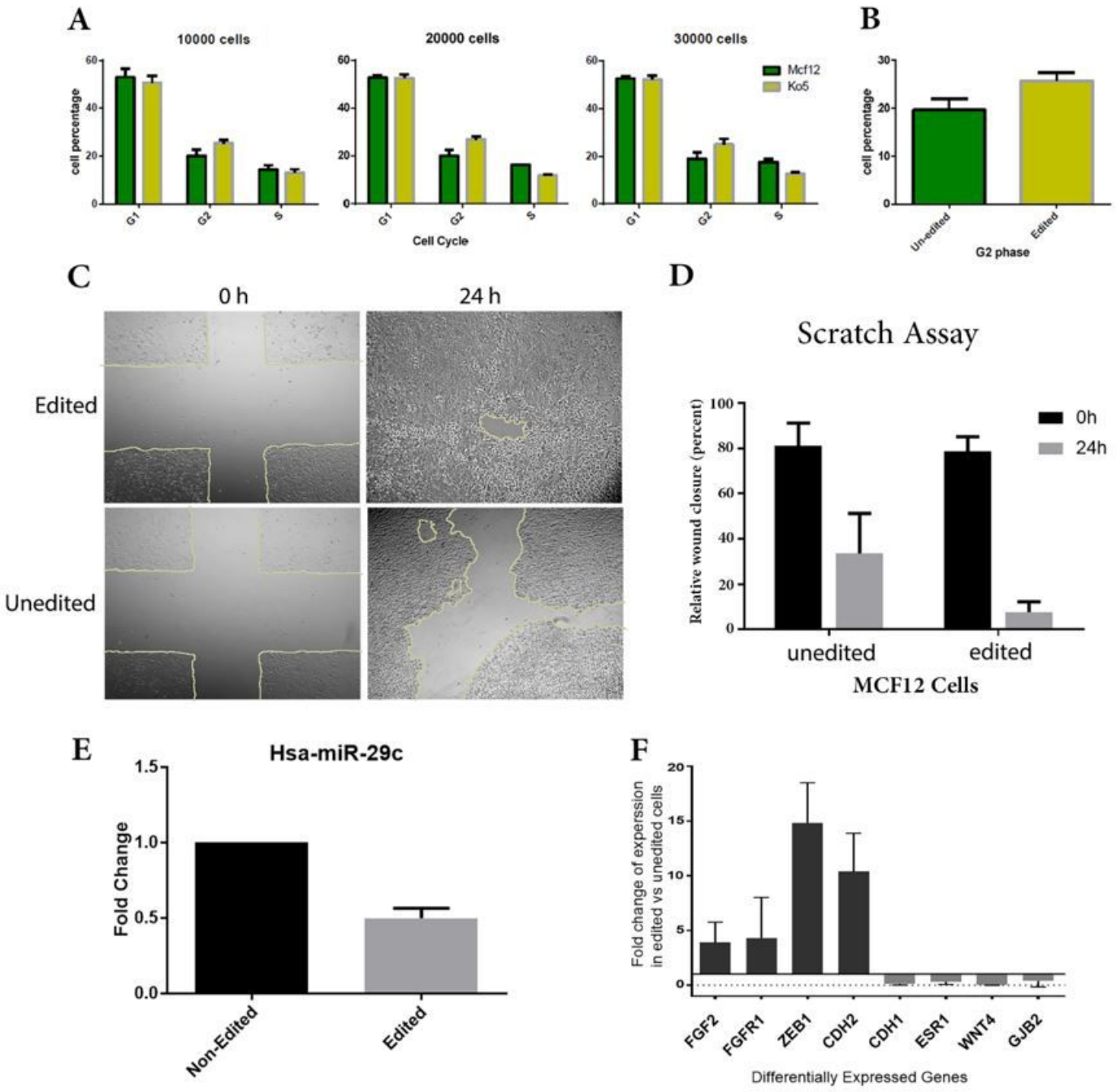

Figure 6

Functional analysis of C1orf132. A) Flow cytometry was used to explore cell cycle distribution for edited (C1orf132-promoter eliminated) vs. unedited MCF12A cells. B) Data obtained from three different biological repeats demonstrated a G2/M arrest in edited cells ($p<0.05$). C and D) Scratch (wound healing) assay was employed as a tool to monitor cell migration in edited (C1orf132-promoter eliminated) vs. unedited MCF12A cells. As it is clearly demonstrated in microscopic pictures, after 24 h the uncovered areas of the plates were significantly decreased ($p<0.05$) in cells edited for C1orf132. E) The expression level of miR-29c-3p in C1orf132-promoter eleminated cells has been significantly reduced by half, compared to wild type MCF12A. F) Several genes were differentially expressed in edited vs. unedited

MCF12A cells, FGF2 and its receptor FGFR1, ZEB1 and CDH2 were overexpressed almost 4 to 15 times in edited cells. Moreover, the expression level of some genes which are mostly epithelial makers, CDH1, ESR1, WNT4 and GJB2 were reduced in edited cells. As expected, the expression level of C1orf132 had been decreased as well. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryFig1andtable1.pdf