Human Molecular Genetics, 2015, Vol. 24, No. 24 6910–6920

doi: 10.1093/hmg/ddv391
Advance Access Publication Date: 22 September 2015
Original Article

ORIGINAL ARTICLE

3′UTR shortening and EGF signaling: implications for breast cancer

Hesna Begum Akman[1], Merve Oyken[1], Taner Tuncer[1], Tolga Can[2] and Ayse Elif Erson-Bensan[1,*]

[1]Department of Biological Sciences and [2]Department of Computer Engineering, M.E.T.U., Ankara 06800, Turkey

*To whom correspondence should be addressed. Tel: +90 (312)2105043; Fax: +90 (312)2107976; Email: erson@metu.edu.tr

Abstract

Alternative polyadenylation (APA) plays a role in gene expression regulation generally by shortening of 3′UTRs (untranslated regions) upon proliferative signals and relieving microRNA-mediated repression. Owing to high proliferative indices of triple negative breast cancers (TNBCs), we hypothesized APA to cause 3′UTR length changes in this aggressive subgroup of breast cancers. Our probe-based meta-analysis approach identified 3′UTR length alterations where the significant majority was shortening events (~70%, 113 of 165) of mostly proliferation-related transcripts in 520 TNBC patients compared with controls. Representative shortening events were further investigated for their microRNA binding potentials by computational predictions and dual-luciferase assay. *In silico*-predicted 3′UTR shortening events were experimentally confirmed in patient and cell line samples. To begin addressing the underlying mechanisms, we found CSTF2 (cleavage stimulation factor 2), a major regulator of 3′UTR shortening to be up-regulated in response to epidermal growth factor (EGF). EGF treatment also resulted with further shortening of the 3′UTRs. To investigate the contribution of CSTF2 and 3′UTR length alterations to the proliferative phenotype, we showed pharmacological inhibition of the EGF pathway to lead to a reduction in *CSTF2* levels. Accordingly, RNAi-induced silencing of *CSTF2* decreased the proliferative rate of cancer cells. Therefore, our computational and experimental approach revealed a pattern of 3′UTR length changes in TNBC patients and a potential link between APA and EGF signaling. Overall, detection of 3′UTR length alterations of various genes may help the discovery of new cancer-related genes, which may have been overlooked in conventional microarray gene expression analyses.

Introduction

mRNA UTRs have long been known to have roles in mRNA transport to cytosol, stability, localization and translational efficiency (1). Recently, we came to better appreciate the significance of 3′ UTRs due to the presence of microRNA and/or regulatory RNA binding sites that may eventually affect protein levels. Given the functional significance of 3′UTRs in post-transcriptional regulation through these binding sites, length of these regions becomes important. Major determinant of 3′UTR length (i.e. from the STOP site to the poly(A) tail) is the polyadenylation signal position and mRNA cleavage at a more distal site. Approximately, 70% of human mRNAs are predicted to contain alternative polyadenylation (APA) sites allowing an immense potential of transcript variability (2). Such isoforms may differ in their half-lives, stability and translational efficiencies. Possibly due to this diverse isoform generation potential, APA is seen in both embryonic development and later during growth/differentiation of adult cells across different tissues (3,4).

While APA of few isolated cases were reported earlier (e.g. *IgM* and *DHFR*) (5–8), advancements in high throughput expression and sequencing approaches showed APA to be surprisingly widespread. In highly proliferative cells, a transcriptome wide tendency to switch to proximal poly(A) site usage was initially shown during the activation of primary murine $CD4^+$ T lymphocytes (3). Other studies showed the association of 3′UTR shortening and proliferation (4). Interestingly, the opposite pattern, i.e. lengthening of 3′UTRs, was reported in embryonic development

Received: July 27, 2015. **Revised and Accepted:** September 15, 2015

in mouse (9), fruit fly (10) and zebrafish (11). On the other hand, induced pluripotent stem cells demonstrate a global 3′UTR shortening pattern (9). These findings point out the dynamic regulation of APA in proliferating and differentiating cells.

Given the widespread use and a potential to generate numerous isoforms, altered APA patterns are gaining attention in cancer, a disease characterized by uncontrolled cell proliferation. A potential contributor to altered APA is the changes in APA machinery protein levels. Initial evidence supporting the role of key proteins in APA comes from B cell differentiation where up-regulation of *CSTF2* (cleavage stimulation factor 2), a major regulator of 3′UTR shortening, leads to a switch from distal to proximal poly(A) site selection for the IgM heavy chain transcript (12). *CSTF2* was also shown to be up-regulated upon T-cell stimulation and lead to shorter mRNA isoform generation (13).

TNBC is a heterogeneous subgroup that constitutes 15–20% of all breast cancer cases. TNBCs are immunohistologically characterized by ER, progesterone receptor and HER2/neu negativity and are aggressive with high proliferative indices, earlier relapse times and higher mortality rates (14). Owing to lack of specific targeted therapies, TNBC patients are generally restricted to surgery and conventional systemic chemotherapies. As a result, compared with the considerable benefits of targeted therapies in ER +, HER2/neu+ tumors, TNBC is a challenge in terms of treatment in today's clinical practice. While, within the last decade, high throughput expression analysis provided valuable information on deregulated gene expression patterns in TNBCs, it is clear that we need to better understand the molecular biology behind the TNBC cases. Therefore, we aimed to investigate 3′UTR length changes in TNBC patients compared with normal breast tissue using a computational meta-analysis approach for APA-regulated gene discovery. This way we were able to generate a pipeline to reanalyze a large number of patient and control datasets for APA-based 3′UTR length changes. Here, we present data on the potential role of APA and a shift to shorter 3′UTR usage in TNBCs.

Results

APA in TNBCs

We tested whether TNBCs have altered 3′UTR isoforms due to APA in reference to their high proliferation rates. APA-based 3′ UTR length changes were detected using a probe-based analysis tool, APADetect (described in Materials and Methods). Using 520 TNBC and 82 normal control gene chip analysis results and genome-wide poly(A) site positions, transcript specific probes were grouped as proximal or distal sets. APA events were then detected as ratios of proximal/distal probe intensity means (hereafter referred to as Short/Long Ratio, SLR in short) in TNBCs compared with normal breast tissue (Workflow is summarized in Supplementary Material, Fig. S1).

3′UTR shortening events in TNBC

SLRs of 165 mRNAs were found to be significantly different in 520 TNBCs compared with controls (Fig. 1A). Out of 165 significant APA cases, 68.5% (113 of 165) were shortening and 31.5% (52 of 165) were lengthening events (Supplementary Material, Table S1). An ontology analysis showed enrichment in overall proliferation-related functions and metabolism-related processes (Fig. 1B). Because number of 3′UTR shortening was prominent to lengthening events, six mRNAs showing high SLRs were selected to be investigated in detail. *IQCK* (IQ Domain-Containing Protein K), *USP9X* (Ubiquitin Specific Peptidase 9 X-Linked), *RPL13* (Ribosomal Protein L13), *SNX3* (Sorting Nexin-3), *TOP2A* (Topoisomerase DNA II alpha), and *YME1L1* (YME1-like 1 ATPase) had significantly high SLRs in TNBC patient samples compared with controls (Fig. 1C). ROC analysis further confirmed the specificity and sensitivity of the individual SLR values of these six genes in their abilities to separate TNBCs from controls (Fig. 1D).

As a control, six random genes were selected from the group of genes that had unchanged SLRs in TNBCs compared with controls. Individual analysis of these genes showed no significant change in the means and variances of SLRs in the TNBC and controls (Supplementary Material Fig. S2A). Among these unchanged SLRs was *CDC6* which we have earlier reported to undergo 3′UTR shortening in response to estrogen stimulation in estrogen receptor positive (ER+) breast cancer cells (16). The unchanged SLR of *CDC6* in TNBC patients compared with normal controls was confirmed by RT-qPCR (Supplementary Material, Fig. S2B).

High SLRs and relapse free survival

We grouped patients with top 25% (high) and bottom 25% (low) SLR values for these six significant 3′UTR shortening cases and compared the two group's relapse free survival times up to 10 years. Interestingly, high versus low *SNX3, YME1L1* and *USP9X* SLR values correlated with poor and good prognosis of patients with hazard ratios of 4.75 ($P < 0.001$), 3.362 ($P < 0.0001$) and 1.99 ($P < 0.005$) (95% CI), respectively (Fig. 1E). No significant difference was detected for the high and low SLR groups of *IQCK, RPL13* or *TOP2A*. Given the correlation with prognosis, *SNX3, YME1L1* and *USP9X* transcripts were used as the test set for future analysis. High SLRs for *TOP2A* and *SNX3* SLRs in TNBCs also showed correlation with tumor grades (Supplementary Material, Table S2).

Short isoform expression

Next, short and long isoform expression levels and SLRs were quantified in TNBC cell lines to validate the *in silico* patient results for YME1L1, *SNX3* and *USP9X*. In all three cases, SLRs were indeed significantly higher in breast cancer cell lines compared with normal breast tissue (Fig. 2A). 3′RACE (rapid amplification of cDNA ends) further confirmed the presence of the short isoforms (Supplementary Material, Figs S3–S5). MCF10A is an immortalized, non-tumorigenic cell line, it is of triple negative nature (17), and interestingly, SLR values were also high in MCF10A cells compared with normal breast, possibly due to high proliferative rates (18).

MicroRNA target site abundance for long 3′UTRs

3′UTR shortening may be a means to evade negative regulation of microRNAs (3). Therefore, using three different microRNA binding prediction tools (PITA, FindTar3 and TargetScan), we searched for common microRNA binding patterns for the 3′ UTRs. Results showed few microRNA predictions for the short isoforms, whereas a higher number of microRNA binding predictions were made for the extended region of the longer isoforms (Fig. 2B). Because experimentally confirmed 3′UTR boundaries were not known, *IQCK* and *TOP2A* were not included in the analysis. Among experimentally known 3′UTR boundaries, *SNX3* was the only exception with high number of microRNA prediction sites on the short 3′UTR compared with the extended region of the long isoform. To test whether short 3′UTRs indeed resulted with increased protein levels and to further investigate the *SNX3* case, short and long (short + extended) 3′UTR isoforms of *SNX3*

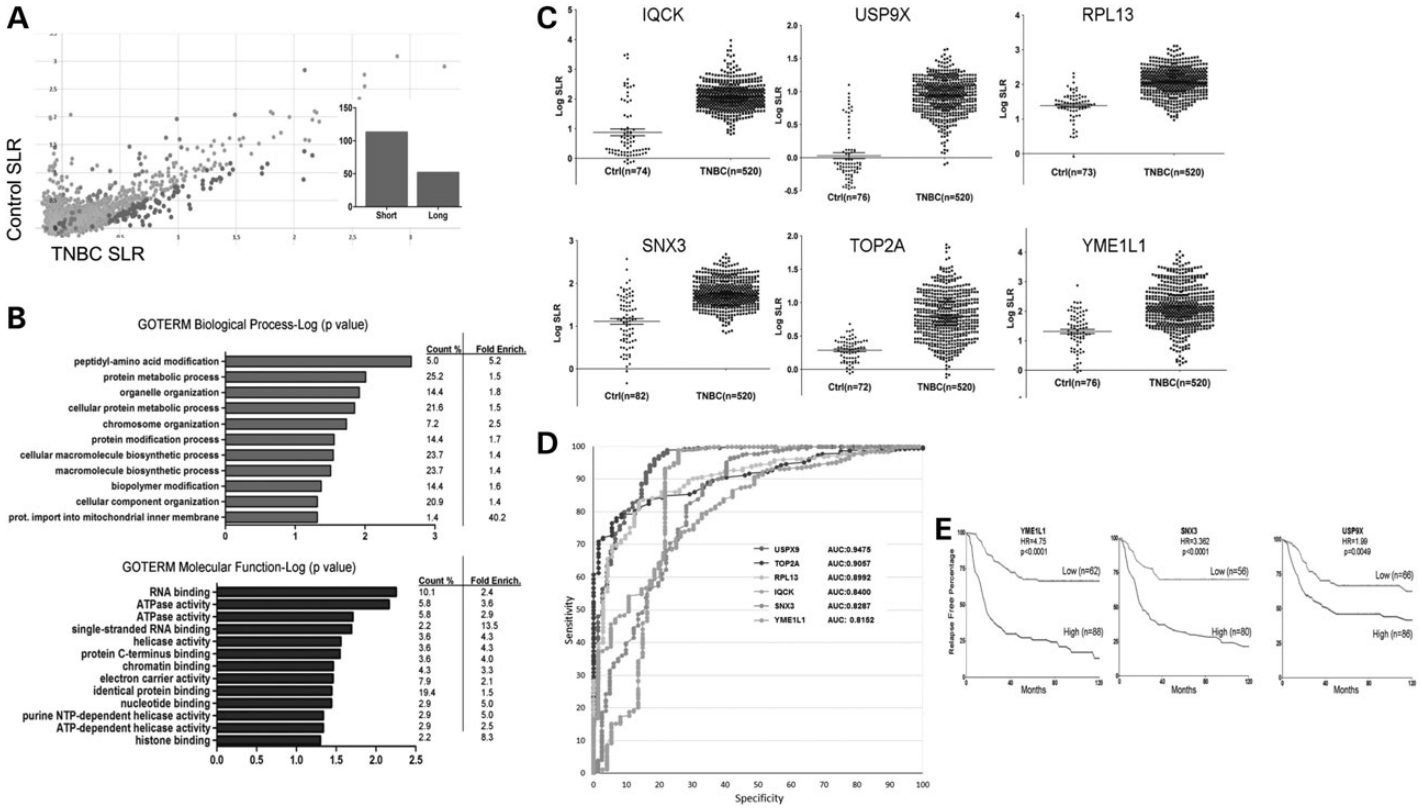

Figure 1. 3′UTR shortening events in TNBCs. (**A**) APA events in 520 TNBCs compared with 82 normal breast tissues. Two thousand and one hundred and eighty-two genes passed stringency filters of APADetect and were analyzed by SAM (15) to evaluate statistical significance. Out of 165 significant APA events, 68.5% (113 of 165) were 3′UTR shortening, and 31.5% (52 of 165) were 3′UTR lengthening events. (**B**) DAVID Ontology analysis of genes undergoing APA and 3′UTR length changes. (**C**) Individual SLR values of *IQCK*, *USP9X*, *RPL13*, *SNX3*, *TOP2A* and *YME1L1* as the most significantly shortened 3′UTRs in TNBCs compared with normal breast tissue (control group, Ctrl). Scatter plots were drawn using the GraphPad software (*P* values were <0.0001). The control/sample numbers vary due to outliers excluded by APADetect stringency filters. (**D**) Receiver operator characteristic (ROC) curve for the most significantly shortened 3′UTRs in 520 TNBCs compared with controls. Area under curve values for each gene's individual SLR is given in the key. (**E**) Relapse free survival rates of TNBC patients with high (highest 25%) or low (lowest 25%) SLRs for *YME1L1*, *SNX3* and *USP9X* for a period of 120 months (10 years). Hazard ratios (95% CI) were 4.75 ($P < 0.001$), 3.362 ($P < 0.0001$) and 1.99 ($P < 0.005$), respectively.

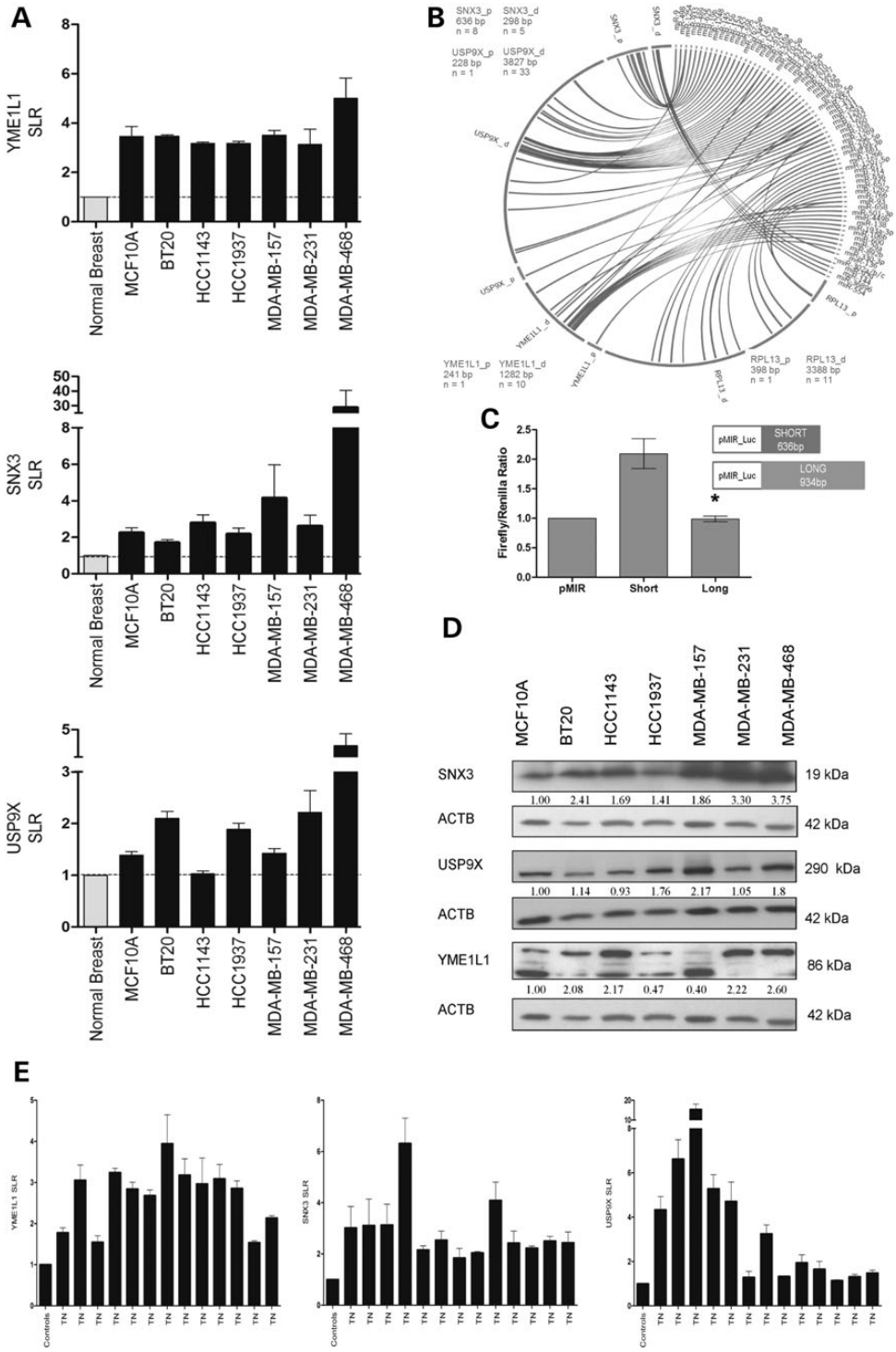

Figure 2. Experimental confirmation of 3′UTR shortening events. (**A**) RT-qPCR analysis of *YME1L1*, *SNX3* and *USP9X*. The fold change for the short and long isoforms was normalized against the reference gene, *SDHA*. Quantification was done using the reaction efficiency correction and ΔΔCq method. Experiment was repeated three times. (**B**) Circos plot shows short (shown as 'S') and long (extended) 3′UTRs (shown as 'L') and corresponding microRNAs predicted by bioinformatics tools described in Materials and Methods. Total 3′UTR sizes and microRNA binding prediction positions are drawn to Proximal and distal 3'UTR targeting microRNAs are shown. (**C**) *SNX3* short and long 3′UTRs were cloned into the 3′UTR of Firefly luciferase gene in pMIR for the dual-luciferase assay. pMIR–short, pMIR–long and pMIR–empty vectors were co-transfected with phRL–TK into MCF10A cells. 24 h after transfection, Firefly/Renilla luciferase read-outs from the constructs were normalized to that of empty pMIR, which was set to 1. Experiment was repeated three times.*$P < 0.05$. (**D**) Western blot analysis of SNX3, USP9X and YME1L1 in TNBC cell lines compared with MCF10A. ACTB antibody was hybridized to membranes for equal loading. Protein band intensities were divided to that of ACTB to assess equal loading. Densitometry was done using the ImageJ software and by MCF10A normalization. YME1L is an integral membrane protein and may have post-translational modifications recognized by the antibody. (**E**) RT-qPCR of SLRs in 4 normal (Controls) and 13 TNBC patient samples. Patient SLRs were normalized to that of control SLRs which were set to 1. Experiment was repeated three times. All values are mean ± SEM.

were cloned into 3′UTR of a luciferase reporter construct. Compared with empty vector constructs (i.e. no 3′UTR cloned downstream of luciferase), short 3′UTR of *SNX3* resulted with increased luciferase activity compared with the longer 3′UTR isoform (Fig. 2C), showing that the short 3′UTR was indeed resulting with higher protein levels.

Next, while mRNA levels may not always correlate with protein levels due to other post-transcriptional and post-translational regulatory mechanisms, we examined the protein levels of these high SLR mRNAs by western blotting (Fig. 2D). For SNX3, USP9X and YME1L1, 50% (3 of 6), 17% (1 of 6) and 67% (4 of 6) of cell lines had more than 2-fold expression compared with already highly proliferative MCF10A cells, respectively.

To evaluate how these findings reflected patient breast tumors, we performed RT-qPCR with 13 TNBC patient samples to detect SLRs for our test genes. We detected high SLRs in patient samples compared with four normal control breast tissue samples (Fig. 2E). Therefore, both the cell line and patient sample data confirmed the high SLRs.

Increased SLRs in TNBCs versus ER+ patients

About 70% of all human breast cancers are ER+ and depend on estrogen to proliferate (19). Therefore, we investigated how specific our findings were for TNBCs compared with ER+ patients ($n = 207$) (20). Interestingly, APA-dependent SLR increase was significantly higher in TNBCs compared with ER+ breast cancers especially for *SNX3*, *USP9X*, and *YME1L1* (Supplementary Material, Fig. S6). There was also a group of genes that had no significant SLR change between TNBCs and ER+ breast cancers (e.g. *RPL13*, *IQCK*).

Implications of APA in TNBCs

Because hundreds of transcripts going through APA-dependent isoform generation may lead to variations in protein levels, these proteins alone or in combination may contribute to the tumorigenesis processes. Therefore, to look into the possible mechanism behind increased SLRs in TNBCs, we turned to epidermal growth factor (EGF) signaling as one of the prominent proliferation pathways. First, we detected expression of CSTF2 (cleavage stimulation factor, 3′ pre-RNA, subunit 2, 64 kDa), a major determinant of proximal polyadenylation site selection (13), (21) in EGFR positive TNBC cell lines (Fig 3A). Next, to investigate a potential link between *CSTF2*, APA and EGF signaling in TNBCs, we treated MCF10A and MDA-MB-468 cells with 50 ng/ml EGF for 12 and 24 h. Compared with untreated cells, *CSTF2* mRNA and protein levels increased in EGF-treated MCF10A and MDA-MB-468 cells (Fig. 3B). Accordingly, EGF treatment resulted with increased SLRs of *SNX3, USP9X* and *YME1L1* (Fig. 3C). Selective pharmacological inhibition of the EGFR pathway by AG1478 decreased pEGFR levels and reversed the increase of *CSTF2* mRNA and protein levels in response to EGF in both cells. Of note, AG1478 treatment resulted with a further decrease of CSTF2 levels compared with unstimulated cells (Fig. 4A). Because we have detected a pattern of 3′UTR length changes in proliferation-related genes (Fig. 1B) and an EGF responsive increase of *CSTF2* mRNA and protein in TNBCs (Fig. 3B), we investigated how important APA may be in the proliferation of TNBC cells. However, considering the fact that inhibition of EGF signaling can have various complex downstream effects on cells, we chose to silence *CSTF2* alone by RNAi (Fig. 4B). Compared with empty vector or control oligo transfected cells, *CSTF2* silenced MCF10A and MDA-MB-468 cells both had a significantly decreased (30–25%) proliferation rate detected by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide assay (MTT) (Fig. 4C), suggesting CSTF2 as a member of APA machinery to have an overall effect on transcripts that in turn change the growth characteristics of cells.

Discussion

In this study, we provide initial evidence on 3′UTR length changes in TNBC patients compared with healthy individuals. Our computational approach confirmed by experimentation showed a pattern of 3′UTR shortening in TNBC patient samples and cell lines. Considering that at least 70% of human mRNAs are predicted to contain APA sites allowing an immense potential of transcript variability (2), it is likely that numerous APA events and consequent changes in 3′UTR lengths may alter the protein expression profiles of cancer cells. Because probe-based analysis is limited to original chip design decisions and reported poly(A) sites, it is possible that APA-based 3′UTR length alterations we report here under-represent the actual complexity. However, one of the major advantages of our approach is the ability to easily analyze hundreds of existing patient datasets as a starting point for a general look on APA-based changes and novel disease-related gene discovery. Indeed, transcripts with significant 3′UTR length changes we chose as a test group were functionally highly relevant to cancer. For example, overexpression of *USP9X* is reported in a range of different cancers including breast. USP9X has been linked to beta-catenin stability, TGF-beta signaling through deubiquitination of SMAD4 as well as regulation of the AMP-activated protein kinase family (22–26). In general, deregulated ubiquitin specific proteases have been implicated in breast cancers by various groups including ours (27,28). *SNX3* is known to control Wingless/Wnt secretion through regulating retromer-dependent recycling (29). Wnt signaling has recently been associated with metastasis in triple negative breast cancers (30). *YME1L1* is required for apoptotic resistance, cristae morphogenesis, and cell proliferation (31) and has been implicated in cancer development and progression as a MYC-responsive gene (32,33).

Given that CSTF2 levels increased in response to EGF and increased expression of APA machinery proteins can affect poly(A) site selection decisions, EGF responsive APA is likely to alter 3′ UTR lengths of hundreds of transcripts at the same time. Overall these changes may have a cumulative effect on cell proliferation as suggested by the effect of on RNAi silenced *CSTF2* on cellular proliferation. These findings were in agreement with earlier observations that oncogenic transformation tends to favor 3′UTR shortening events, leading to higher levels of proteins (34). Therefore, we suggest EGF signaling and its pro-survival effect to be at least partially due to increased APA events. Hence overactive signaling pathways, not only in breast cancers but also other malignancies, may be exerting their proliferative effects partly by APA and 3′UTR length alterations of a set of genes. However, individual shortening and lengthening events should be considered case by case given the complexity of UTR-dependent positive or negative regulatory mechanisms.

For future work, other cancers with overactive EGFR signaling as well as other pro-survival signals will be of interest to investigate for APA-based 3′UTR length changes. Such alterations are likely to take roles in proliferation and cancer-related pathways that support tumor initiation and/or maintenance (34). Because 3′UTR length changes may have been overlooked in earlier conventional gene expression analyses, functional understanding of these 3′UTR changes then can be exploited to develop more effective treatment options for patients.

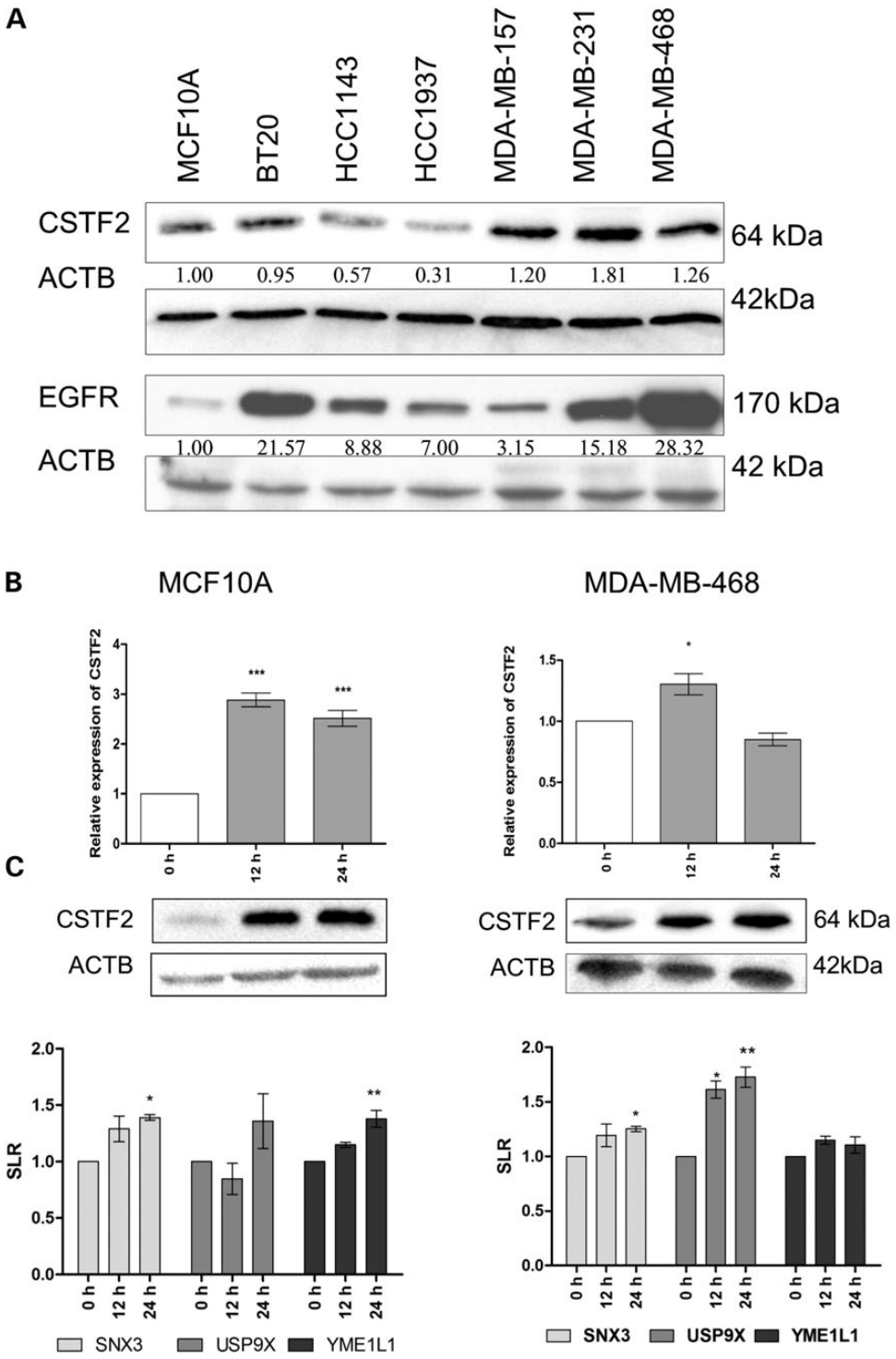

Figure 3. APA and EGF signaling. (**A**) CSTF2 and EGFR expression in TNBC cell lines detected by western blotting and normalized to MCF10A. (**B**) MCF10A and MDA-MB-468 cells were treated with EGF (50 ng/ml) for 12 and 24 h. *CSTF2* mRNA levels were detected by RT-qPCR and protein levels by western blotting. ACTB antibody was hybridized to the same membranes to assess equal loading. Experiment was repeated three times. (**C**) SLR increase in *SNX3*, *USP9X* and *YME1L1* in EGF-treated MCF10A and MDA-MB-468 cells. All values are mean ± SEM. $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.0001$.

Materials and Methods

Data

CEL files microarray results were downloaded from NCBI Gene Expression Omnibus (GEO) repository (35). Five hundred and twenty comparable datasets for TNBC samples came from GSE31519 (36) and 82 histologically normal epithelium and cancer-free prophylactic mastectomy patients were used (42 from GSE20437 (37), 15 from GSE9574 (38), 6 from GSE6883 (39), 6 from GSE26910 (40), 5 from GSE21422 (41), 7 from GSE3744 (42) and 1 from GSE2361 (43)). ER+ patient ($n = 209$) data were from GSE2034 (20).

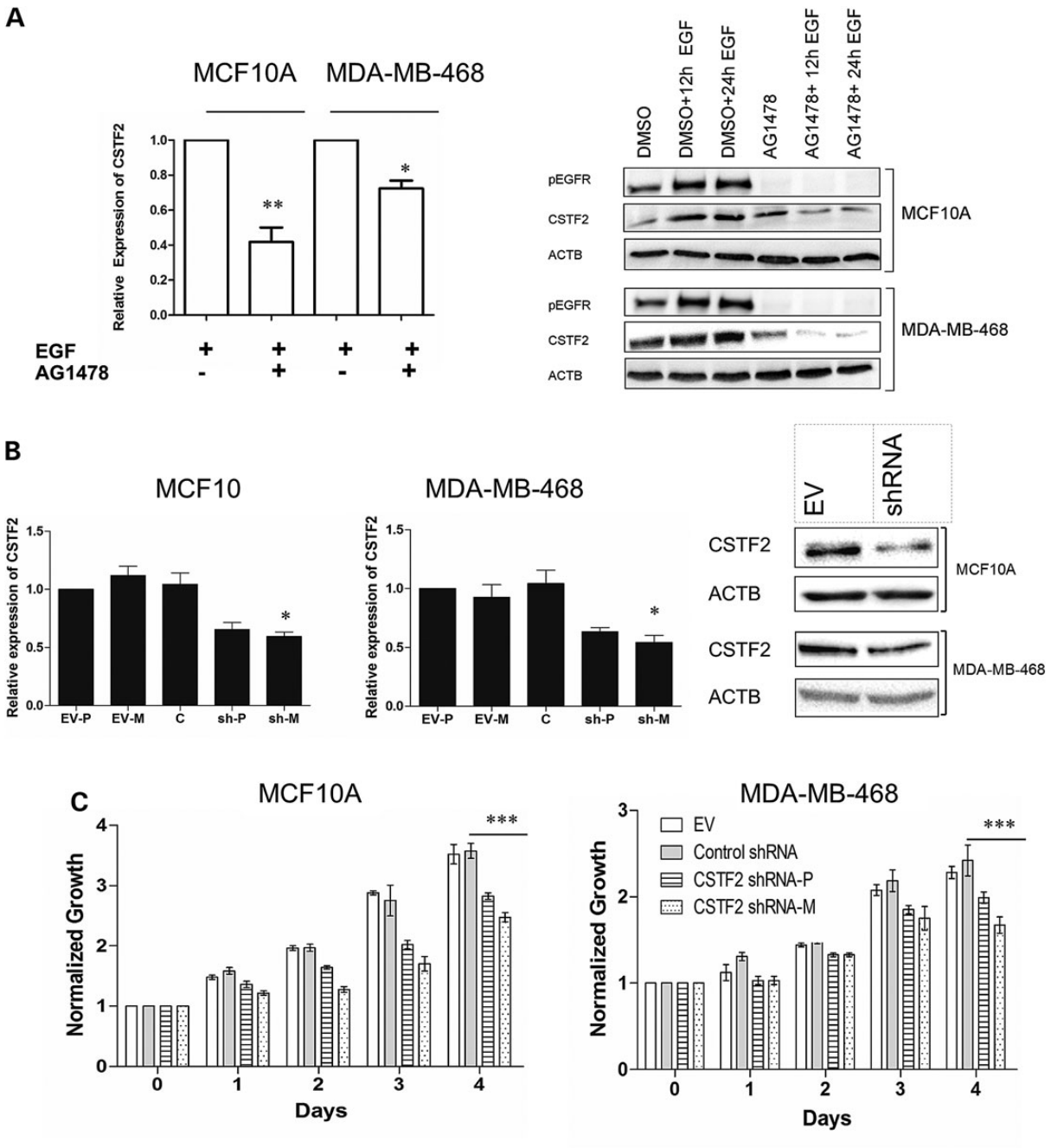

Figure 4. Effect of pharmacological and RNAi inhibition of *CSTF2* (**A**) Cells were pretreated with the selective EGFR inhibitor AG1478, then with EGF for 12 h. Relative *CSTF2* levels were detected by RT-qPCR. pEGFR and CSTF2 protein levels in EGF and AG1478 treated cells were detected by western blotting. (**B**) shRNA oligo harboring vectors were transfected into MCF10A and MDA-MB-468 cells. EV-P (polyclonal empty vector), EV-M (monoclone empty vector), C (control oligo harboring vector), sh-P (CSTF2 shRNA, polyclonal), sh-M (CSTF2 shRNA, monoclone). Relative *CSTF2* levels were detected by RT-qPCR. CSTF2 protein levels were confirmed to be down-regulated in MCF10A by ~60% and in MDA-MB-468 cells by ~30% detected by densitometric analysis (**C**) *CSTF2* silenced clones (P and M) demonstrated decreased proliferation quantified by MTT assay compared with controls (empty vector and control shRNA transfected cells). All values are mean ± SEM. $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.001$.

APADetect

Probe-based expression analysis tool (APADetect) was developed to detect 3′UTR isoforms. APADetect is accessible online at http://www.ceng.metu.edu.tr/~tcan/APADetect/ and is provided as a command line Java program with source code. APADetect analyzes Affymetrix Human Genome U133A (HGU133A, GPL96) and U133 Plus 2.0 Arrays (HGU133Plus2, GPL570) using the consensus/exemplar sequences of a transcript and probe recognition target sites on the transcripts. There are 2411 and 3683 unique probe sets for a total of 2066 and 3067 genes on the HGU133A and HGU133Plus2 platforms that can be split into proximal and distal probe sets based on the positions of poly(A) sites taken from PolyA_DB (44). For individual genes, mean signal intensities of proximal and distal probe sets are calculated and used as indicators of 'Short' and 'Long' 3′UTR isoform abundance. Patient SLR (short/long ratio) values are then compared with the control group

to detect APA-based isoform abundance changes. APADetect has stringency filters to calculate SLRs. Size filter eliminates transcripts with only one distal or proximal probe for accuracy concerns. Outlier probe filter eliminates outlier probes using Iglewicz and Hoaglin's median method. For each proximal and distal subset, the median absolute deviation is computed as the median of absolute differences of individual probe intensities from their respective medians. The modified z-score of a probe p is then computed as:

$$z = \frac{0.6745\,(\text{intensity of } P - \text{median})}{\text{MAD}}.$$

Probes with absolute values of z-score greater than 3.5 are identified as outliers and are removed from subsequent analyses. After removal of the outlier probes, samples with only one proximal/distal probe are filtered out by re-application of the 'size filter' at the sample level. Outlier sample filter screens samples that deviate significantly from group medians. Distal filter eliminates probes that are significantly higher than proximal ones as proximal probes should recognize both the short and the long isoform. APADetect results are then tested with significance analysis of microarrays (SAM) for statistical significance after log normalization of SLR values (15). Sample numbers that passed the filters are indicated in the figures and legends. A work flow chart is also presented (Supplementary Material, Fig. S1).

Ontology and significance analysis

Significant APA events were analyzed using DAVID (45) and by scatter plots using the GraphPad software followed by the t-test for statistical significance. receiver operator curve (ROC) analysis was performed for sensitivity and specificity (46,47). For correlation analysis of SLRs and patient characteristics, an unpaired t-test with Welch's correction and F-test were used for statistical significance of group means and variances. For relapse free survival analysis, SLR values of patients were grouped as 'High' (top 25%) and 'Low' (bottom 25%). Statistical significance and P values were determined by log-rank test using GraphPad.

microRNA target sites on mRNAs

TargetScan (http://www.targetscan.org/), PITA (http://genie.weizmann.ac.il/) and FindTar3 (http://bio.sz.tsinghua.edu.cn/) were used to predict possible microRNA–mRNA interactions for the short and long 3′UTR isoforms and common predictions were shown on a Circos plot (48).

Samples and treatments

Cells were grown in following conditions; MCF10A in DMEM/F12 medium with 5% horse serum, 1% penicillin/streptomycin (P/S), 100 mg/ml EGF, 1 mg/ml hydrocortisone, 1 mg/ml cholera toxin, and 10 mg/ml insulin, BT20, HCC1143, HCC1937, MDA-MB-157, MDA-MB-231 and MDA-MB-468 in DMEM with Earle's salts and 10% FBS. All media contained 1% P/S. Cell lines were grown as monolayers and were incubated at 37°C with 95% humidified air and 5% CO_2. Before EGF treatments, MCF10A cells were grown in DMEM/F12 medium with 1% (P/S) alone, MDA-MB-468 cells were grown in phenol red-free medium supplemented with 10% charcoal stripped FBS and 1% P/S for 48 h. EGF-deprived cells were then treated with 50 ng/ml EGF for 12 and 24 h. Selective inhibitor of EGF receptor kinase, AG1478 (25 μM, Tocris), was given 12 h prior to EGF where appropriate. Treatments were repeated at least two independent times. Four normal and 13 TNBC patient cDNAs were used from Origene Breast Cancer cDNA array IV (BCRT504) (with three technical replicates).

Expression analysis

RNA isolation, quantification, cDNA synthesis and expression analysis were conducted according to MIQE Guidelines (49) and checklist. Total RNA was isolated using High Pure RNA Isolation Kit (Roche Applied Science). For DNase treatment, total RNA was incubated with DNase I (Roche Applied Science) at 37°C for 1 h followed by phenol-chloroform extraction. Lack of DNA contamination was confirmed by PCR. RNA concentration and purity were determined by NanoDrop ND1000 (Thermo Scientific). cDNAs (20 μl) were synthesized using the RevertAid First Strand cDNA Synthesis Kit (Fermentas) using 2 μg total RNA and oligo dT primers. For RT-qPCR, SYBR® Green Mastermix (Roche Applied Science) was used with the Rotor Gene 6000 (Corbett, Qiagen) cycler. Ten microliters reactions were performed with 300 nM of specific primer pairs. USP9X (NM_001039591), SNX3 (NM_003795), and YME1L1 (NM_014263) short and long 3′UTR isoforms were amplified using the following primer sets; USP9X 3′ UTR- short (product size: 135 bp) USP9X_SF:5′-TCCAGGTCCTC CAATCAAATAC-3′, USP9X_SR: 5′-GGATTGGGAAGAGGAGGATT AC T-3′, USP9X 3′UTR-long (product size: 190 bp) USP9X_LF:5′-GAA AGAAACAGCCCCCAGAA-3′, USP9X_LR: 5′-TCAGTTGGTG GAGGGCTTAG-3′, SNX3 3′-UTR-short (product size: 165 bp) SNX3_SF: 5′-GCCTGAAATTTGGCAAGAAG-3′, SNX3_SR: 5′-TCTT GTCAACTGCCAAAACAA-3′, SNX3 3′-UTR-long (product size: 150 bp) SNX3_LF: 5′-TCATTCCTGTAACTCCATTCCCT-3′, SNX3_LR: 5′-GCAGTTTTCAAATACACAAAGTGCT-3′, YME1L1 3′-UTR-short (product size: 187 bp) YME1L1_SF: 5′-TGATATGGATGCTTGCTGGT -3′, YME1L1_SR: 5′-GGGATGCTAA TTTGCAATAGG-3′, YME1L1 3′-UTR-long (product size: 150 bp) YME1L1_LF: 5′-TGCTTGGTTT CAAAAGCAAAAT-3′, YME1L1_LR: 5′-ATTCATCTGTAGATCCTAT AACAGCA-3′. Fold changes were normalized against the reference gene; SDHA (NM_004168) (50) amplified with the following primer sets: SDHA (product size: 86 bp) SDHA_F: 5′-TGGGAACAAGAGGG CATCTG-3′, and SDHA_R: 5′-CCACCAC TGCATCAATTCATG-3′. Reference gene, *SDHA*, expression is consistent across breast cancer cell lines (50). For all reactions, following conditions were used; incubation at 94°C for 10 min, 40 cycles of 94°C for 15 s, 56°C for 30 s, and 72°C for 30 s. For relative quantification, the reaction efficiency incorporated ΔΔCq formula was used (51). One-way ANOVA with Tukey's multiple comparison post-test was performed using GraphPad Prism (CA, USA). Experiments were performed three independent times.

Dual-luciferase analysis

Short (636 bp) and long (934 bp) 3′UTR fragments of *SNX3* were PCR amplified with cloning primers using a high fidelity polymerase (Phusion polymerase) into pMIR-Report (pMIR) plasmid (Ambion). SacI and HindIII recognition site harboring cloning primers were as follows (restriction sites are in bold letters and recognition aiding sites are underlined):

SNX3-pMIR-F: CGAGCTCGAATTTGGCAAGAAGGGGCAA
SNX3-pMIR-R1: CCCAAGCTTGGGTGCATATCAGAAATGCATTT
SNX3-pMIR-R2: CCCAAGCTTGGGAATGGTATTTTATCTTTGAA

After sequence confirmation, MCF10A cells were co-transfected with pMIR (Firefly luciferase) (450 ng) and phRL–TK (Renilla Luciferase) (50 ng) in 24-well plates using 0.75 μl Fugene-HD (Roche). Twenty-four hours after transfection, cells were collected and dual-luciferase activities were measured using the

Modulus Microplate Luminometer (Turner Biosystems). The luminescence intensity of Firefly luciferase (pMIR) was normalized to that of Renilla luciferase (phRL-TK). Experiments were repeated three independent times with four replicates per experiment. One-way ANOVA followed by Tukey's multiple comparison test was used for statistical analysis.

Protein isolation and western blotting

Cells were washed with PBS and total protein isolations were done with M-PER Protein Extraction Reagent (Pierce). Protein extracts (50 µg) were denatured in 6X Laemmli buffer (12% SDS, 30% 2-mercaptoethanol, 60% Glycerol, 0.012% bromophenol blue, 0.375 M Tris) at 100°C for 5 min and were separated on a 8% polyacrylamide gel and transferred onto a PVDF membrane (Roche). The membranes were blocked in 5% BSA in TBS-T (Tris Buffer Saline-Tween, 20 mM Tris, 137 mM NaCl, pH: 7.6, 0.1% Tween 20) and were incubated overnight with the appropriate polyclonal primary antibodies: anti-USP9X (1:1000 dilution; Abcam ab99343), anti-SNX3 (1:200 dilution; Abcam ab174123), anti-YME1L1 (1:200 dilution; Abcam ab170123), anti-EGFR (1:200 dilution, Santa Cruz Biotechnology), anti-pEGFR (1:500 dilution, Abcam) and anti-CSTF2 (1:1000 dilution, Abnova) antibodies; followed by a 1 h incubation with the HRP-conjugated secondary antibodies. Proteins were visualized using an enhanced chemiluminescence kit (Clarity Western ECL Substrate; Bio-Rad). ACTB (sc-47778, Santa Cruz Biotechnology) was used for protein loading control. For blocking 5% BSA in TBS-T was used. Primary antibody was 1:2000 diluted and secondary anti-mouse antibody was 1:2000 diluted.

RACE

3′rapid amplification of cDNA ends was performed using the 3′ RACE Kit as suggested (Roche). Gene specific forward primers were designed for each gene and used SNX_F: GCCTGAAATTT GGCAAGAAG

USP9X_F1:CTGGCAATCCTCAGTACACTTACAAC

USP9X_F2:GAGATCACATAGTGCTAGGATGACACTT

YME1L1_F: TGATATGGATGCTTGCTGGT. Bands with the correct sizes were gel-extracted using Zymoclean™ Gel DNA Recovery Kit (Zymo Research). The gel-extracted PCR products were cloned to pGEM®-T Easy Vector (Promega). Cloning was confirmed with double digest using ApaI and SacI enzymes (Fermentas) and sequencing.

CSTF2 silencing and proliferation assay

shRNAs were designed for *CSTF2* according to pSUPER.retro.neo + GFP vector guidelines (OligoEngine). F: GATCCCCGGCTTTAGT CCCGGGCAGATTCAAGAGATCTGCCCG GGACTAAAGCCTTTTTA, R:AGCTTAAAAAGGCTTTAGTCCCGGGCAGATCTCTTG AATCTG CCCGGGACTAAA GCCGGG). Oligos were annealed and ligated to the vector as described earlier (28). After sequence confirmation, cells were transfected with shCSTF2, empty vector and control oligo constructs. Positive mono- and poly-clones were selected by 500 µg/ml and 750 µg/ml G418 (Roche) treatment for MCF10A and MDA-MB-468 cells, respectively. Cell proliferation was measured by MTT as described previously (28). Experiment was repeated three independent times with five replicates.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgments

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) 112S478. HBA was funded by the same grant. The authors would like to thank Dr M. Muyan for helpful discussions and critical reading of the manuscript.

Conflict of Interest statement. The authors declare no conflict of interest.

References

1. Moore, M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514–1518.
2. Derti, A., Garrett-Engele, P., Macisaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M. and Babak, T. (2012) A quantitative atlas of polyadenylation in five mammals. *Genome Res.*, **22**, 1173–1183.
3. Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A. and Burge, C. B. (2008) Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science*, **320**, 1643–1647.
4. Elkon, R., Ugalde, A.P. and Agami, R. (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.*, **14**, 496–506.
5. Alt, F.W., Bothwell, A.L., Knapp, M., Siden, E., Mather, E., Koshland, M. and Baltimore, D. (1980) Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3′ ends. *Cell*, **20**, 293–301.
6. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell*, **20**, 313–319.
7. Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L. and Wall, R. (1980) Two mRNAs with different 3′ ends encode membrane-bound and secreted forms of immunoglobulin mu chain. *Cell*, **20**, 303–312.
8. Setzer, D.R., McGrogan, M., Nunberg, J.H. and Schimke, R.T. (1980) Size heterogeneity in the 3′ end of dihydrofolate reductase messenger RNAs in mouse cells. *Cell*, **22**, 361–370.
9. Ji, Z., Lee, J.Y., Pan, Z., Jiang, B. and Tian, B. (2009) Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl Acad. Sci. U S A*, **106**, 7028–7033.
10. Hilgers, V., Perry, M.W., Hendrix, D., Stark, A., Levine, M. and Haley, B. (2011) Neural-specific elongation of 3′UTRs during Drosophila development. *Proc. Natl Acad. Sci. USA*, **108**, 15864–15869.
11. Li, Y., Sun, Y., Fu, Y., Li, M., Huang, G., Zhang, C., Liang, J., Huang, S., Shen, G., Yuan, S. *et al.* (2012) Dynamic landscape of tandem 3′UTRs during zebrafish development. *Genome Res.*, **22**, 1899–1906.
12. Takagaki, Y. and Manley, J.L. (1998) Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Mol. Cell.*, 2, 761–771.
13. Chuvpilo, S., Zimmer, M., Kerstan, A., Glöckner, J., Avots, A., Escher, C., Fischer, C., Inashkina, I., Jankevics, E., Berberich-Siebelt, F. *et al.* (1999) Alternative polyadenylation events contribute to the induction of NF-ATc in effector T cells. *Immunity*, **10**, 261–269.
14. Griffiths, C.L. and Olin, J.L. (2012) Triple negative breast cancer: a brief review of its characteristics and treatment options. *J. Pharm. Pract.*, **25**, 319–323.

15. Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. U S A*, **98**, 5116–5121.
16. Akman, B.H., Can, T. and Erson-Bensan, A.E. (2012) Estrogen-induced upregulation and 3′-UTR shortening of CDC6. *Nucleic Acids Res.*, **40**, 10679–10688.
17. Chavez, K.J., Garimella, S.V. and Lipkowitz, S. (2010) Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer. *Breast Dis.*, **32**, 35–48.
18. Soule, H.D., Maloney, T.M., Wolman, S.R., Peterson, W.D., Brenz, R., McGrath, C.M., Russo, J., Pauley, R.J., Jones, R.F. and Brooks, S.C. (1990) Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res.*, **50**, 6075–6086.
19. Harvey, J.M., Clark, G.M., Osborne, C.K. and Allred, D.C. (1999) Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J. Clin. Oncol.*, **17**, 1474–1481.
20. Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J. *et al.* (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*, **365**, 671–679.
21. Shell, S.A., Hesse, C., Morris, S.M. and Milcarek, C. (2005) Elevated levels of the 64-kDa cleavage stimulatory factor (CstF-64) in lipopolysaccharide-stimulated macrophages influence gene expression and induce alternative poly(A) site selection. *J. Biol. Chem.*, **280**, 39950–39961.
22. Deng, S., Zhou, H., Xiong, R., Lu, Y., Yan, D., Xing, T., Dong, L., Tang, E. and Yang, H. (2007) Over-expression of genes and proteins of ubiquitin specific peptidases (USPs) and proteasome subunits (PSs) in breast cancer tissue observed by the methods of RFDD-PCR and proteomics. *Breast Cancer Res. Treat.*, **104**, 21–30.
23. Taya, S., Yamamoto, T., Kanai-Azuma, M., Wood, S.A. and Kaibuchi, K. (1999) The deubiquitinating enzyme Fam interacts with and stabilizes beta-catenin. *Genes Cells*, **4**, 757–767.
24. Murray, R.Z., Jolly, L.A. and Wood, S.A. (2004) The FAM deubiquitylating enzyme localizes to multiple points of protein trafficking in epithelia, where it associates with E-cadherin and beta-catenin. *Mol. Biol. Cell*, **15**, 1591–1599.
25. Dupont, S., Mamidi, A., Cordenonsi, M., Montagner, M., Zacchigna, L., Adorno, M., Martello, G., Stinchfield, M.J., Soligo, S., Morsut, L. *et al.* (2009) FAM/USP9x, a deubiquitinating enzyme essential for TGFbeta signaling, controls Smad4 monoubiquitination. *Cell*, **136**, 123–135.
26. Al-Hakim, A.K., Zagorska, A., Chapman, L., Deak, M., Peggie, M. and Alessi, D.R. (2008) Control of AMPK-related kinases by USP9X and atypical Lys(29)/Lys(33)-linked polyubiquitin chains. *Biochem J.*, **411**, 249–260.
27. Sacco, J.J., Coulson, J.M., Clague, M.J. and Urbé, S. (2010) Emerging roles of deubiquitinases in cancer-associated pathways. *IUBMB Life*, **62**, 140–157.
28. Akhavantabasi, S., Akman, H.B., Sapmaz, A., Keller, J., Petty, E. M. and Erson, A.E. (2010) USP32 is an active, membrane-bound ubiquitin protease overexpressed in breast cancers. *Mamm. Genome*, **21**, 388–397.
29. Zhang, P., Wu, Y., Belenkaya, T.Y. and Lin, X. (2011) SNX3 controls Wingless/Wnt secretion through regulating retromer-dependent recycling of Wntless. *Cell Res*, **21**, 1677–1690.
30. Dey, N., Barwick, B.G., Moreno, C.S., Ordanic-Kodani, M., Chen, Z., Oprea-Ilies, G., Tang, W., Catzavelos, C., Kerstann, K.F., Sledge, G.W. *et al.* (2013) Wnt signaling in triple negative breast cancer is associated with metastasis. *BMC Cancer*, **13**, 537.
31. Stiburek, L., Cesnekova, J., Kostkova, O., Fornuskova, D., Vinsova, K., Wenchich, L., Houstek, J. and Zeman, J. (2012) YME1L controls the accumulation of respiratory chain subunits and is required for apoptotic resistance, cristae morphogenesis, and cell proliferation. *Mol. Biol. Cell*, **23**, 1010–1023.
32. Wan, D., Gong, Y., Qin, W., Zhang, P., Li, J., Wei, L., Zhou, X., Li, H., Qiu, X., Zhong, F. *et al.* (2004) Large-scale cDNA transfection screening for genes related to cancer development and progression. *Proc. Natl Acad. Sci. USA*, **101**, 15724–15729.
33. Guo, Q.M., Malek, R.L., Kim, S., Chiao, C., He, M., Ruffy, M., Sanka, K., Lee, N.H., Dang, C.V. and Liu, E.T. (2000) Identification of c-myc responsive genes using rat cDNA microarray. *Cancer Res.*, **60**, 5922–5928.
34. Mayr, C. and Bartel, D.P. (2009) Widespread shortening of 3′ UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, **138**, 673–684.
35. Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M. *et al.* (2011) NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res.*, **39**, D1005–D1010.
36. Rody, A., Karn, T., Liedtke, C., Pusztai, L., Ruckhaeberle, E., Hanker, L., Gaetje, R., Solbach, C., Ahr, A., Metzler, D. *et al.* (2011) A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast Cancer Res.*, **13**, R97.
37. Graham, K., de las Morenas, A., Tripathi, A., King, C., Kavanah, M., Mendez, J., Stone, M., Slama, J., Miller, M., Antoine, G. *et al.* (2010) Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. *Br. J. Cancer*, **102**, 1284–1293.
38. Tripathi, A., King, C., de la Morenas, A., Perry, V.K., Burke, B., Antoine, G.A., Hirsch, E.F., Kavanah, M., Mendez, J., Stone, M. *et al.* (2008) Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients. *Int. J. Cancer*, **122**, 1557–1566.
39. Liu, R., Wang, X., Chen, G.Y., Dalerba, P., Gurney, A., Hoey, T., Sherlock, G., Lewicki, J., Shedden, K. and Clarke, M.F. (2007) The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N. Engl. J. Med.*, **356**, 217–226.
40. Planche, A., Bacac, M., Provero, P., Fusco, C., Delorenzi, M., Stehle, J.C. and Stamenkovic, I. (2011) Identification of prognostic molecular features in the reactive stroma of human breast and prostate cancer. *PLoS One*, **6**, e18640.
41. Kretschmer, C., Sterner-Kock, A., Siedentopf, F., Schoenegg, W., Schlag, P.M. and Kemmner, W. (2011) Identification of early molecular markers for breast cancer. *Mol. Cancer*, **10**, 15.
42. Richardson, A.L., Wang, Z.C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J.D., Livingston, D.M. and Ganesan, S. (2006) X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell*, **9**, 121–132.
43. Ge, X., Yamamoto, S., Tsutsumi, S., Midorikawa, Y., Ihara, S., Wang, S.M. and Aburatani, H. (2005) Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics*, **86**, 127–141.
44. Zhang, H., Hu, J., Recce, M. and Tian, B. (2005) PolyA_DB: a database for mammalian mRNA polyadenylation. *Nucleic Acids Res.*, **33**, D116–D120.
45. Huang, d.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
46. Metz, C.E. (1978) Basic principles of ROC analysis. *Semin. Nucl. Med.*, **8**, 283–298.

47. Wray, N.R., Yang, J., Goddard, M.E. and Visscher, P.M. (2010) The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.*, **6**, e1000864.
48. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.
49. Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L. *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.*, **55**, 611–622.
50. Gur-Dedeoglu, B., Konu, O., Bozkurt, B., Ergul, G., Seckin, S. and Yulug, I.G. (2009) Identification of endogenous reference genes for qRT-PCR analysis in normal matched breast tumor tissues. *Oncol. Res.*, **17**, 353–365.
51. Fleige, S., Walf, V., Huch, S., Prgomet, C., Sehm, J. and Pfaffl, M. W. (2006) Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR. *Biotechnol Lett.*, **28**, 1601–1613.