3′-UTR shortening contributes to subtype-specific cancer growth by breaking stable ceRNA crosstalk of housekeeping genes

Zhenjiang Fan[1], Soyeon Kim[2,3], Yulong Bai[4], Brenda Diergaarde[4,5], Steffi Oesterreich[6,7,8], Hyun Jung Park[4†].

[1]Department of Computer Science, University of Pittsburgh, Pittsburgh, PA, USA

[2]Department of Pediatrics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA

[3]Division of Pulmonary Medicine, Children's Hospital of Pittsburgh UPMC, Pittsburgh, PA, USA

[4]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

[5]Hillman Cancer Center, University of Pittsburgh Medical Cancer, Pittsburgh, PA, USA

[6]Women's Cancer Research Center, UPMC Hillman Cancer Center, Pittsburgh, PA, USA

[7]Magee-Women's Research Institute, Magee-Women's Research Hospital of University of Pittsburgh Medical Center, Pittsburgh, PA, USA

[8]Department of Pharmacology and Chemical Biology, University of Pittsburgh, Pittsburgh, PA, USA

[†] Corresponding author: hyp15@pitt.edu (H.J.P.)

ABSTRACT

Shortening of 3′UTRs by alternative polyadenylation (APA) is a post-transcriptional mechanism that can regulate expression of hundreds to thousands of genes in human cancers. In breast cancer, different subtypes of tumor samples, such as estrogen receptor positive and negative (ER+ and ER-), are driven by distinct molecular mechanisms, suggesting possible differences in post-transcriptional regulation between the subtype tumors. However, the distinct 3′-UTR shortening (3′US) profiles between the subtype tumors do not clearly signify mechanistic differences. In this study, based on the profound tumorigenic role of 3′US through competing-endogenous RNA (ceRNA) network (3′US-ceRNA effect), we hypothesize that 3′US-ceRNA effect, not 3′US cis effect, drives subtype-specific tumor growth. However, we found that different sample size available in the subtypes bias the size of the ceRNA networks, disrupting the fair comparison of the 3′US-ceRNA effect. By addressing this bias and building tumor ceRNA networks comparable between the subtypes, we identified a novel subset of housekeeping (HK) genes as stable and strong miRNA sponges (sponge HK genes) that synchronize ceRNA networks of normal samples adjacent to the subtype tumor samples. We

further found that global and distinct 3′US events effectively break their stable sponge effect in a subtype-specific fashion, especially in association to aggressive and metastatic phenotypes for ER- tumors. Knockdown of NUDT21, a master 3′-UTR regulator in HeLa cells, confirmed that the causal role of 3′US-ceRNA effect dysregulating sponge HK genes for tumor growth. In spite of the high and stable expression across diverse physiological conditions, HK genes have been under-characterized especially in post-transcriptional regulation. In this study, we identified a novel subset of HK genes, sponge HK, that play important roles in maintaining normal ceRNA networks and driving subtype-specific tumor growth in their interaction with global 3′US events in tumor.

INTRODUCTION

Approximately, 70% of human genes contain multiple polyadenylation (polyA) sites in the 3′-untranslated region (3′-UTR)[1]. Through alternative polyadenylation (APA) during transcription, messenger RNAs (mRNA) from the same gene can have various 3′-UTR lengths. Since the 3′-UTR contains regulatory regions including microRNA (miRNA) target sites, mRNAs with shortened or lengthened 3′-UTRs may diversify regulation landscape, for example miRNA binding landscape, in pathological conditions such as cancer[2], [3]. In human cancer, 3′-UTR lengthening (3′UL) has been associated with cell senescence[4] with implications for tumor-associated processes, such as cell cycle inhibition, DNA damage/repair process, and tumor suppression [5]–[8]. Widespread 3′-UTR shortening (3′US) has been implicated for oncogene activation in cancer cell lines[1]. Further, 3′US events add prognostic power beyond common clinical and molecular covariates in cancer patients [9] and are associated with drug sensitivity in cancer cell lines [10]. These results suggest that APA events, both 3′-UTR shortening and lengthening, play important roles in cancer etiology and treatments.

The 3′-UTR is also implicated in competing-endogenous RNA crosstalk (ceRNA)[11]. CeRNAs co-regulate each other RNAs through competing for binding miRNAs. In diverse types of cancer, ceRNA regulation involves numerous established oncogenes and tumor suppressor genes[12] and makes a post-transcriptional level of molecular pathway interactions for tumorigenesis [13]. When 3′-UTR shortening genes lose miRNA target sites on their 3′-UTRs and do not sequester the miRNAs, the associated miRNAs bind to the 3′-UTR of the ceRNA partners, which would be competing for the binding of the miRNAs. As a result, 3′-UTR shortening disrupts ceRNA crosstalk (3′US-ceRNA) for growth of diverse cancer types, including breast cancer[14].

Breast cancer can be classified into two major subtypes based on estrogen receptor (ER) status, a central component of the pathological evaluation of breast cancer[15]. Estrogen receptor negative (ER-) breast tumors have unique growth mechanism compared to estrogen receptor positive (ER+) breast tumors. Thus, ER- breast cancer has a worse prognosis than ER+ [16] with a more aggressive phenotype[17], [18]. Also, while ER+ cancers can be treated with endocrine therapy, blocking ER activity or depleting estrogen levels, the unique growth mechanism of ER- breast cancers immobilize this therapeutic approach.

To develop targeted therapies that effectively treat ER- breast cancer, its molecular dynamics needs to be understood comprehensively. Based on the profound tumorigenic effect of 3′US-ceRNA[14], we hypothesize that 3′US-ceRNA effects specific to ER- breast tumors contribute to the unique growth mechanism. In this study, by addressing a quantitative challenge due to different sample sizes between ER+ and ER- breast tumor samples, we corroborated this hypothesis. Especially, in our analyses, we identified a novel subset of housekeeping (HK) genes (sponge HK) in synchronizing ceRNA networks in normal samples adjacent to the subtype tumor samples. Further, we showed that sponge HK genes are likely *trans* targets of 3′US-ceRNA effects playing important roles in subtype-specific tumor growth. Since subtyping human cancers based on molecular features enables more accurate treatment plans, our analyses provide a systematic understanding of post-transcriptional dynamics for ER- breast tumors and shed novel insights into the development of effective therapies for ER- breast cancer.

Global 3′-UTR shortening and 3′-UTR lengthening events differ between ER+ and ER-

To study the role of 3′-UTR shortening genes in subtype-specific breast tumors, we divided 97 breast tumors and the adjacent "normal" samples available in TCGA (see Methods) into 77 ER-positive (ER+) and 20 ER-negative (ER-) sample pairs. Using DaPars [9], we identified 3′UTR shortened (3′US, ΔPDUI < -0.2 and FDR < 0.05) and 3′UTR lengthened (3′UL, ΔPDUI > 0.2 and FDR < 0.05) genes. There are similar numbers of 3′US genes and 3′UL genes in the subtypes (in terms of union set (**Fig. 1A, B, C**) and intersection (**S. Fig. 1A, 1B**)), suggesting that they play equally important roles. However, we found that 3′US genes are more recurrent (occurring in > 20% of the tumor samples[9], **Fig. 1D, E, S. Fig. 1C, 1D**) both in the ER+ and ER- tumors (e.g. $P=5.0\times10^{-5}$ for ER+). The lower recurrence of 3′UL genes partially explains why previous studies, which were based on recurrent events[9], [10], observed fewer 3′UL genes than 3′US. With our interest in the common mechanism of APA events, we will focus on 3′UL genes. Further analyses showed that 3′US and 3′UL play distinct roles in cancer. First, the recurrent 3′US and 3′UL genes show little overlap in ER+ and ER- tumors (1 and 13 genes in common, $P=1.27e^{-6}$ and $P=3.97e^{-9}$, respectively, **Fig. 1D, 1E**). Second, the number of 3′UL events is not correlated with that of 3′US events across the tumor samples (P=0.35 for ER+ and P=0.61 for ER-, **Fig. 1B, 1C**). Third, IPA pathway analysis (**S. Table 1**, **S. Fig. 1E**) shows that the recurrent 3′US and 3′UL genes are enriched for distinct sets of molecular pathways in ER+ and ER- tumor samples. The IPA pathway analysis further suggests that 3′UL or 3′US genes themselves have limited roles for subtype-specific cancer progression in spite of the global occurrence, in that small numbers of pathways are significantly ($P<10^{-2}$) enriched for them and only a couple of them are "cancer" pathways (those with keyword "cancer"). Previously, we showed a profound role of 3′US genes on tumorigenesis not in cis, but in trans in their interaction with ceRNAs (3′US-ceRNA effect) [14]. Since ER- specific 3′US genes do not show strong cis effect in terms of molecular pathway, we further hypothesize that the widespread recurrent 3′US genes specific to ER- tumors make 3′US-ceRNA effect for ER- specific cancer growth.

Two-step Pairwise Normalization of ER+ and ER- ceRNA network

To study subtype-specific 3′US-ceRNA effect, we set out to compare the ceRNA networks from ER+ and ER- tumor and normal sample pairs. Computationally, ceRNA networks build edges between the genes if the genes share a significant number of microRNA (miRNA) target sites and their expression levels across samples are correlated (co-expression)[14], [19]. For gene pairs that share a significant number of miRNA target sites (FDR < 0.05 based on hypergeometric test e.g. [14], [19]), the co-expression cutoff needs to be determined to build comparable ceRNA networks for ER+ and ER-. Using the common co-expression cutoff (e.g., Pearson's $\rho > 0.6$) would inflate the number of edges for ER- compared to ER+ (160,687 in ER- normal vs. 88,275 in ER+ normal, **S. Fig. 2A**). Our simulation study (using a cutoff of Pearson's ρ 0.6) further confirmed that the network size difference is attributable to the sample size difference. When smaller numbers of ER+ normal samples are used to construct the ceRNA network, the network size increase (**S. Fig. 3A**), consistent with the trend for ER- and ER+ network size (**S. Fig. 2A**). In particular, the simulation shows that the actual size of the ER- normal network (with 20 samples) falls in the non-outlier range of ER+ normal network size when 20 samples are used.

Although the network size difference results in a systematic bias in network comparison, it is not straightforward to address the bias. To remove the bias, one might want to sample ER+ normal samples to the same number of samples available for ER- (n=20). However, the resulting ceRNA networks constructed from the subsample lose topological consistency within them (**S. Fig. 2B**), making it difficult to represent the ceRNA dynamics in ER+ tumors. Also, if the ER- ceRNA network is constructed using the expression correlation cutoff with the same statistical significance to ER+ (based on a permutation test, see Methods), it will drastically deflate the number of edges (**S. Fig. 2C**), resulting in another systematic bias for comparison. To address this issue, we first constructed the reference network from normal samples of larger size (ER+) using the common correlation cutoff (Pearson's $\rho > 0.6$). Based on the assumption that normal samples would have similar molecular dynamics between ER+ and ER-, we sought to find an expression correlation cutoff for ER- normal network that yields the most topological similarity to the ER+ reference network. To estimate topological similarity, we employed a normalized Laplacian Matrix Eigenvalue Distribution that discovers ensembles of Erdős–Rényi graphs better than other metrics, such as Sequential Adjacency or Laplacian[20] (see Methods). While ER- normal network topology changes drastically using different correlation cutoff values (**S. Fig. 3B, 3C**), we found that a cutoff of 0.68 makes the ER- normal network most similar to the ER+ reference network (2,846 and 2,864 nodes for ER+ and ER-, respectively **S. Fig. 2D**). Using another measure for topological similarity, average clustering coefficient[21], the cutoff of 0.68 is supported again that normal ER- network with correlation cutoff 0.68 makes the closest average clustering coefficient to the reference network (0.39 for the ER- network with cutoff 0.68 and 0.40 for the reference ER+ network, **S. Fig. 2D**). We further applied the subtype-specific cutoffs (0.68 for ER- and 0.6 for ER+) to build tumor ceRNA networks (1,392 and 2,039 nodes for ER+ and ER-, respectively). Since this method normalizes ER+ and ER- normal ceRNA networks by identifying correlation cutoffs for topological similarity and builds ER+ and ER- tumor ceRNA networks separately using the subtype-specific correlation cutoffs, we call it two-step network normalization. Subsequent analyses show the importance of this normalization

approach in inspecting subtype-specific 3′US-ceRNA effect based on the number of genes in the network.

3′UTR shortening is associated with the aggressive metastatic phenotypes of ER- tumors in ceRNA.

To identify the subtype-specific effect of 3′US-ceRNA, we compared ER- and ER+ ceRNA networks after two-step normalization. Especially, due to the aggressive phenotypes, we will study the effect of 3′US-ceRNA on ER- tumors from now on. Among 1,783 ceRNA partners of 521 3′US genes (3′US ceRNA partners) in normal ER- ceRNA network, 498 (27.9%) are found only in ER- (ER- 3′US ceRNA partners), whereas the other 1,285 (72.1%) are also in ER+ as 3′US ceRNA partners (common 3′US ceRNA partners, **Fig. 2A**). We found that 118 IPA canonical pathways significantly ($P < 0.01$) enriched for the ER- 3′US ceRNA partners (**S. Table 2**) are linked with several aspects of ER- specific tumor phenotypes (**Fig. 2B**). The first set of the pathways are "cancer" pathways. For example, the Ingenuity Pathway Analysis knowledgebase "Molecular Mechanisms of Cancer" pathway ($P=10^{-5.25}$) includes a set of genes disruptions of which drive tumor progression. Specific to breast cancer, the enrichment of the "Breast Cancer Regulation by Stathmin1" ($P=10^{-3.92}$) pathway is interesting, since overexpression of Stathmin1 correlates with loss of the ER [22] and with aggressive breast tumor phenotypes[23]. The second category of pathways underlies the aggressive metastasis of ER- tumors more directly. For example, eight pathways were found associated with breast tumor metastasis [24], and we found that five of them are significantly enriched for ER- 3′US ceRNA partners with the exception of PI3K/AKT, the enriched p-value of which is just below our cutoff ($P=10^{-1.95}$). Further, previous studies have associated breast tumor malignancy and poor survival with abnormal control of Ephrin A (reviewed in [25]), which is strongly enriched for ER- 3′US ceRNA partners (P-val=$10^{-5.05}$). The functional enrichment analyses show that ER- specific 3′US ceRNA partners regulate pathways for cancer signaling and aggressive metastatic phenotypes in normal samples. However, in ER- tumors, 81.7% of 3′US ceRNA partners lost the ceRNA relationship (**S. Fig 4A)**, likely losing the normal control. Since most of them (1285 of 1,690 (76.0%), **Fig. 2A**) were shared with ER- normal samples, ER- and ER+ share less 3′US ceRNA partners in tumor (35 of 416 (8.41%) 3′US ceRNA partners in ER- shared with ER+ (**Fig. 2C**)) that were enriched in cancer signaling pathways in normal samples. Consequently, subtype-specific 3′US-ceRNA effect interrupts pathways responsible for cancer signaling and metastasis in a subtype-specific fashion.

Housekeeping genes keep ER+ and ER- normal ceRNA networks to similar topology

Further, we categorized housekeeping (HK), tumor-associated (tumor suppressors or oncogenes, TA), and transcription factor (TF) genes with possible sponge effect (>5 miRNA binding sites in the 3′UTR). Out of 3,804 HK[26], 932 TA[27], and 1,020 TF genes[28] curated in public databases (see Methods), the ceRNA networks consist of 3-fold more HK genes than TA or TF genes with the sponge effect (**Fig. 3A** for normal and **S. Fig. 5A** for tumor). Expectedly, HK genes form 5 ~ 7 fold more edges than the other gene classes in both normal and tumor ceRNA

networks for ER+ and ER- subtypes (**S. Fig. 5B**). Due to their active roles in cell maintenance[26], HK genes are expected to maintain constant expression levels under most physiological conditions[26]. Consistent with the observation, the 958 HK ceRNA genes (**S. Fig. 5C**) express as highly as but with a significantly ($P=1.72e^{-54}$) less variation across the normal samples than 1,906 non-HK ceRNA genes in the network (**Fig. 3B**). With our observation that the HK genes contain more miRNA binding sites than the other genes (P=0.05, **Fig. 3C**), they should function as stable sponges for miRNAs[29]. Thus, with a significantly large number of the HK ceRNA genes ($P=8.77e^{-771}$, **Fig. 3D**) between ER- and ER+ normal samples, we hypothesize that they keep ER- and ER+ normal ceRNA networks in similar topology. To test this hypothesis, from normal ER- and ER+ ceRNA networks, we first selected edges involving the HK ceRNA genes to form subnetworks (HK ceRNA networks). Further, by selecting the same number of edges not involving HK genes 100 times, we form 100 non-HK ceRNA networks. Since the HK ceRNA networks are significantly more similar between ER+ and ER- ($P < 0.01$) than 100 non-HK ceRNA networks through miRNA sponge effect, we identified a novel role of HK genes that make normal ceRNA crosstalk consistent between the subtypes.

3′US disrupts ceRNA crosstalk of housekeeping genes in tumor.

To examine the impact of 3′US on the HK ceRNA genes, we selected 727 genes (75.8%) that are connected to 3′US genes (3′US HK ceRNA partners) out of 958 HK genes on the ER- normal network, which is in the same scale as the other classes of genes that are known to be regulated by 3′US genes[12], [14] (61.8% from 317 TA genes and 90.2% from 271 TF genes). Additionally, 3′US HK ceRNA partners are connected to as many 3′US genes as the other classes of genes are (**Fig. 4A**). Compared to 231 HK genes on the ER- normal network that are not connected to 3′US genes, those that are ceRNA partners of 3′US genes (3′US HK ceRNA partners) are more highly connected in the network (**Fig. 4B**). The high connectivity of the HK genes suggests the important roles of 3′US-ceRNA effect for ER- normal ceRNA network. Previously, we showed that 3′US represses genes in tumor if they were the ceRNA partners in normal [14]. Hence, repression of 3′US HK ceRNA partners in tumor (**Fig. 4C)** signifies that they are indeed in ceRNA relationship with 3′US genes in normal. Further, simulation studies and cell line experiments have shown that ceRNA relationships propagate through the ceRNA network[19], [30]. In the same vein, ceRNA relationship changes, either loss or gain (tumor vs. normal), also could propagate[13]. Thus, when the ceRNA relationship of HK genes is disrupted in tumor due to 3′US, the disrupted ceRNAs should further disrupt the ceRNA relationship with their ceRNA partners (**S. Fig. 6A**). As 727 3′US HK ceRNA partners lost higher ratios of ceRNA partners in tumor (**Fig. 4D**), 3′US disrupts stable ceRNA network topology of these HK genes and their roles in the normal ceRNA network.

3′US represses housekeeping genes to promote tumor growth

To gain insights into the cause-and-effect relationship between 3′US-ceRNA effect and HK gene repression and between HK gene repression and tumor growth, we revisited a previous study[14], [31], in which NUDT21 knockdown (KD) induced broad 3'UTR-shortening, leading

to enhanced tumor growth in HeLa cells and glioblastoma *in vitro* and *in vivo* (available in GSE42420[14] and GSE78198[31]). First, we chose 11,431 genes that are expressed in the experiment data (avg. FPKM > 1). Among them, we further chose 4,430 genes that would work as miRNA sponges (>5 miRNA binding sites). To identify ceRNA relationship between the genes, we will solely use significance of miRNA binding site overlap (FDR < 0.05), since the other criteria for the ceRNA identification, co-expression, cannot be effectively estimated from two replicates of NUDT21 KD experiments. This ceRNA network includes 860 3′US genes that are connected with 2,449 genes in the network. Among these 3′US ceRNA partners, a significant portion of them (705, 28.8%) are HK genes, while 184 are TA and 163 are TF genes. Especially, it is interesting to note that HK genes in the network are only either 3'US genes (n=298) or 3'US ceRNA partners (n=705). On the other hands, almost half of TA and TF genes in the network are not connected with 3'US genes (149 of 333 (44.7%) and 147 of 310 (47.4%) for TA and TF, respectively). These results suggest that 3'US is designed to affect HK genes through ceRNA. Further, expectedly, among 2,449 3'US ceRNA partners in the network, 705 HK genes are more repressed than 184 TA and 163 TF genes or than 298 HK 3'US genes (**Fig. 5A,** P-value=0.01 and 0.05, 0.002, respectively) in the network. These results confirm that 3′US-ceRNA effect represses HK genes also in the NUDT21 KD experiment.

To assess the impact of this repression on tumor growth, we further conducted IPA analysis on 705 HK 3'US ceRNA partners in comparison to the other 2,410 HK genes not in the network. First, although there are much less HK 3'US ceRNA partners than the other HK genes, they are enriched for more IPA Diseases & Functions terms. While the IPA analysis gives N/A for the terms that are so lowly enriched that cannot be estimated, a significantly more number of the terms are estimated N/A for the other HK genes than for HK 3'US ceRNA partners (185 and 296, respectively, P-value<$2.2e^{-16}$). Further, for the terms that are non-N/As in either HK 3'US ceRNA partners or the other HK genes, we substitute N/As with the minimum value across the estimated p-values. Then, the difference of p-values in HK 3'US ceRNA partners vs. the other HK genes also shows that more terms are significantly enriched for HK 3'US ceRNA partners (254 terms with better p-values for HK 3'US ceRNA partners and 141 for the other HK genes, P-value<$2.2e^{-16}$). This trend is more pronounced for the terms that are important for cancer. For example, IPA terms with keywords “Cell”, “Cancer (or Tumor)”, “Apoptosis (, Death, or Necro)”, and “Growth (, Proliferation, or Progression)” are significantly (P-value<$2.2e^{-16}$) enriched in HK 3'US ceRNA partners, while certain terms for general biological processes such as “RNA” are enriched in the other HK genes (**Fig. 5B**). Altogether, the results show that 3'US-ceRNA effect in the NUDT21 KD experiment, which promotes tumor growth, represses a particular set of HK genes to disrupt pathways that are important for cancer.

CeRNAs of housekeeping genes in ER- tumor are associated with ER- specific growth.

To understand the general subtype-specific impact of HK genes in ceRNA, we analyzed the ceRNA partners of HK genes (HK ceRNA partner) in ER+ and ER- tumors. With ceRNA network reduced in tumor due to 3′US [14] and miRNA expression decrease[32], HK genes lost the ceRNA relationship in both ER+ and ER- (972 of 1,635 (59.4%) lost in ER- and 1,330 of

1,688 (78.8%) in ER+, **S. Fig. 5B**). The loss breaks the high overlap of HK ceRNA partners between ER+ and ER- normal samples (**Fig. 5A**), resulting into 505 and 144 HK ceRNA partners unique in ER- and ER+ tumor (ER- and ER+ HK ceRNA partners), respectively (**Fig. 6B**). While it is known that cell growth and cell cycle regulations are different in the subtypes[33]–[35], our analysis shows that the subtype-specific HK ceRNAs are enriched for cell growth- and cell cycle-related pathways. First, we found that ER- HK ceRNA partners are enriched for pathways associated to growth factor (with keyword "GF") (**S. Table 4**). Especially, EGF (P-val=$10^{-2.99}$) activates cell cycle progression in ER- tumors[36], and expression of VEGF (P-val=$10^{-2.42}$) is associated to ER- tumors[37]. Also, both EGF and VEGF are suspected to proliferate ER- tumors when estrogen cannot sustain them[37]. Second, cell cycle pathways are enriched for ER+ specific HK ceRNA partners, suggesting that ER-regulated cell cycle[38], [39] differentiates ER+ and ER- cancer partially at the ceRNA level. Especially, since regulation of cell cycle, G1- and S-phase and their transition ratio, is crucial for ER+ tumor's proliferation (reviewed in [40]), it is interesting that cell cycle regulation pathways for various phases (G1/S or G2/M) of various mediators (Estrogen or Cyclins) are enriched with 144 ER+ HK ceRNA partners. Third, considering that the enrichment analysis was for the disjoint sets of genes (505 unique to ER- and 144 unique to ER+), it is interesting that unique HK ceRNA partners of both subtypes are significantly enriched for some "cancer" pathways e.g. "Molecular Mechanisms of Cancer", because it shows that HK ceRNAs are involved in cancer mechanisms equally significantly but in a subtype-specific fashion.

DISCUSSION

To investigate the role of 3′US-ceRNA effect [14] for estrogen receptor positive (ER+) vs. ER- breast tumors, we constructed the ceRNA networks for the subtype tumors comparable to each other by addressing the bias due to the different number of samples (72 for ER+ and 20 for ER- in TCGA). A fair comparison of the networks suggests that 3′US disrupts the ceRNA network for ER- tumors' aggressive phenotypes. Further, we revealed a role of 3′US-ceRNA effect on housekeeping (HK) genes. Although HK genes highly and stably express in diverse biological contexts[41], our understanding of their roles is limited, especially with regards to ceRNA crosstalk. For the first time, our results identified a general novel role of HK genes in maintaining ceRNA network on physiological conditions.

Further analyses show that 3′US disrupts ceRNA crosstalk of HK genes in a subtype-specific fashion. First, we defined a novel subset of HK genes that are *trans* target of 3′US-ceRNA effect (3′US ceRNA HK genes) enriched in important pathways in association to ER-'s aggressive phenotype. Since they are much less than the other HK genes in number (e.g. 705 3′US ceRNA HK genes vs. 2,401 HK genes in the NUDT21 KD experiment), our definition may shed novel insights into the efficient development of new therapeutic targets. Especially, 3′US ceRNA HK genes are repressed (**Fig. 4C**) and lose ceRNA relationships in ER- tumors (**Fig. 4D**) via certain miRNAs that are released from global 3′US events[13], [14]. Thus, treating ER- tumors with miRNAs involved in the effects is expected to mitigate ER-'s aggressive phenotype.

In network analysis, a network of interest is often compared to a reference network. However, if the networks are constructed from different numbers of samples, the comparison will be biased due to the sample size difference (**S. Fig. 2, 3**). To compare the ceRNA networks between ER- and ER+ tumors, we assumed that normal samples should have similar molecular dynamics and found the subtype-specific cutoff values for normal ceRNA networks. Then, we assumed that normal and tumor ceRNA network should share the cutoff within each subtype and applied the subtype-specific cutoff value to construct the tumor ceRNA network for each subtype (two-step pairwise normalization). As the resulting ceRNA networks facilitate novel discoveries on the subtype-specific 3′US-ceRNA effect, we expect that the two-step pairwise normalization method can further help normalize biological networks constructed with the different number of samples if the matched normal samples are available.

METHODS

TCGA breast tumor RNA-seq data and identification of breast cancer subtypes.

Quantified gene expression files (RNASeqV1) for primary breast tumors (TCGA sample code 01) and their matching solid normal samples (TCGA sample code 11) were downloaded from the TCGA Data Portal[42]. We used 97 breast tumor samples that have matched normal tissues, which were further categorized into 77 estrogen receptor positive (ER+) and 20 estrogen receptor negative (ER-). For ER+ and ER-, we collected both normal (ER+ normal and ER- normal) and tumor (ER+ tumor and ER- tumor) samples. A total of 10,868 expressed RefSeq genes (fragments per kilobase of transcript per million mapped reads (FPM) ≥ 1 in > 80% of all samples) were selected for downstream analyses.

Selection of miRNA target sites

Predicted miRNA-target sites were obtained from TargetScanHuman version 6.2[43]. Only those with a preferentially conserved targeting score (Pct) more than 0 were used[9]. Experimentally validated miRNA- target sites were obtained from TarBase version 5.0[44], miRecords version 4[45] and miRTarBase version 4.5[46]. The target sites found in indirect studies such as microarray experiments and high-throughput proteomics measurements were filtered out [47]. Another source is the microRNA target atlas composed of public AGO-CLIP data[48] with significant target sites (q-value < 0.05). The predicted and validated target site information was then combined to use in this study.

Statistical significance of Pearson correlation coefficient

The implementation of the Pearson r function is provided by a python package, SciPy, and available at https://scipy.org/, which returns the calculated correlation coefficient and a 2-tailed p-value for testing non-correlation. The Pearson correlation coefficient measures the linear relationship between two variables (e.g. gene X and gene Y) and when the two covariates follow binormal distribution, we can assume that their Pearson's correlation follows student t distribution. The p-value is calculated by three steps: 1) calculating the value of the Pearson's correlation t, 2) defining the degree of freedom df (N-2, where N is the sample size), 3) getting the probability of having t or more extreme than t from a Student's t-distribution with the degrees of freedom df. We used hypergeometric test in Scipy to estimate significant of miRNA binding site overlap between genes.

Housekeeping, transcription factor and tumor-associated genes

Housekeeping genes are required for the maintenance of basic cellular functions that are essential for the existence of a cell, regardless of its specific role in the tissue or organism. Generally, housekeeping (HK) genes are expected to be expressed at relatively constant rates in

most non-pathological situations[41]. We used 3,804 HK genes defined in RNA-Seq data for 16 normal human tissue types: adrenal, adipose, brain, breast, colon, heart, kidney, liver, lung, lymph, ovary, prostate, skeletal muscle, testes, thyroid, and white blood cells[26].

Transcription factors (TFs) play an important role in the gene regulatory network. We used 2,020 TF genes defined in TFcheckpoint database[28], in which TF information is collected from 9 different resources followed by manual inspections for sequence-specific DNA-binding RNA polymerase II TF.

The tumor-suppressor genes and oncogenes were defined by the TUSON algorithm from genome sequencing of > 8,200 tumor/normal pairs[27], in particular residue-specific activating mutations for oncogenes and discrete inactivating mutations for tumor-suppressor genes. TUSON computationally analyzes patterns of mutation in tumors and predicts the likelihood that any individual gene functions as a tumor- suppressor gene or oncogene. We used 466 oncogenes and 466 tumor suppressor genes at the top 500 in each prediction (after subtracting 34 genes in common).

Building subtype ceRNA networks

For each of the breast cancer data (ER+ normal, ER+ tumor, ER- normal, and ER- tumor) that we defined above, we constructed a ceRNA network based on microRNA (miRNA) target site share and expression correlation[14], [19]. The same miRNA target site information was determined regardless of the subtypes, resulting into a miRNA target site share network (based on FDR > 0.05 in hypergeometric test with miRNA target site information). And given the same miRNA target site share network, the expression correlation information for each subtype will select ceRNA network edges for each subtype.

We first constructed the ER+ normal reference ceRNA network by applying a traditional correlation cutoff (>=0.6) on the miRNA target site share network. Then, to identify ER- normal ceRNA network comparable to ER+ normal reference ceRNA network, we applied different correlation cutoff values (0 to 1 with a step size of 0.01) on the miRNA target site share network for ER- normal samples, and select the correlation cutoff values that makes ER- normal ceRNA network most similar to ER+ normal reference ceRNA network. To estimate topological similarity, we employed normalized Laplacian Matrix Eigenvalue Distribution that discovers ensembles of Erdős–Rényi graphs better than other metrics such as Sequential Adjacency or Laplacian[20]. After identifying the ER+ normal reference network and the corresponding ER- normal network, we used the same cutoffs (0.6 for ER+ subtypes and 0.68 for ER- subtypes) to construct the ER+ tumor network and the ER- tumor network, respectively.

Estimating topological similarity

To identify the structural equivalence between two networks, we employed spectral analysis not only to identify the structural similarities, but also to track down the underlying dynamic behavior changes between them. Spectral clustering on networks uses the eigenvalues of several

matrices, such as adjacency matrix, the Laplacian matrix, the normalized Laplacian matrix. In this research, we used the normalized Laplacian matrix since it involves both the degree matrix and adjacency matrix, where the degree matrix can identify the node related equivalence of networks and the adjacency matrix can capture the structural equivalence of networks. Another very important reason of using the normalized Laplacian eigenvalue matrix is that it is more sensitive to small changes because it considers more information[17].

For network G, the normalized Laplacian of G is the matrix:

$$N = D^{-1/2} - LD^{-1/2} \tag{1}$$

where L is the Laplacian matrix of G and D is the degree matrix. The Laplacian matrix L is defined as: $L = D - A$, where A is the adjacency matrix of G.

In N, each of its entry elements is given by:

$$N_{i,j} = \begin{cases} 1, & if\ i = j\ and\ \mathrm{degree}(v_i) \neq 0 \\ -\dfrac{1}{\sqrt{\mathrm{degree}(v_i)\,\mathrm{degree}(v_j)}}, & if\ i \neq j\ and\ v_i\ is\ adjacent\ to\ v_j \\ 0, & otherwise \end{cases} \tag{2}$$

where degree (vertex v) is the function that return the degree of the vertex v.

To assess how close two eigenvalue distributions of network G_1 and G_2 are, we used the Kolmogorov–Smirnov test (KS test), which is defined as:

$$K_{n,m} = \sup_x |dist_{1,n}(x) - dist_{2,m}(x)| \tag{4}$$

where $dist_{1,n} and\ dist_{2,m}$ are the empirical distribution functions of the first and the second eigenvalue distribution respectively, and $\sup_x$ is the supremum of the set of distances.

By using the normalized Laplacian Matrix and KS test, ER+ normal reference network G_{ref}^{ER+} is compared with a ER- normal subnetwork with a particular correlation cutoff i G_i^{ER-} in the following three steps:

1) Compute the normalized Laplacian metrics N_{ref}^{ER+} and N_i^{ER-} from G_{ref}^{ER+} and G_i^{ER-} respectively.
2) Compute the eigenvalues E_{ref}^{ER+} and E_i^{ER-} from N_{ref}^{ER+} and N_i^{ER-} respectively.
3) Compute the KS statistic between E_{ref}^{ER+} and E_i^{ER-}.

The third step test the null hypothesis that eigenvalues E_{ref}^{ER+} and E_i^{ER-} are drawn from the same continuous distribution. If the two-tailed p-value returned by the KS test is high, then we cannot reject the hypothesis that G_{ref}^{ER+} and G_i^{ER-} are the same network. In another word, the higher the p-value is, the more similar G_{ref}^{ER+} and G_i^{ER-}.

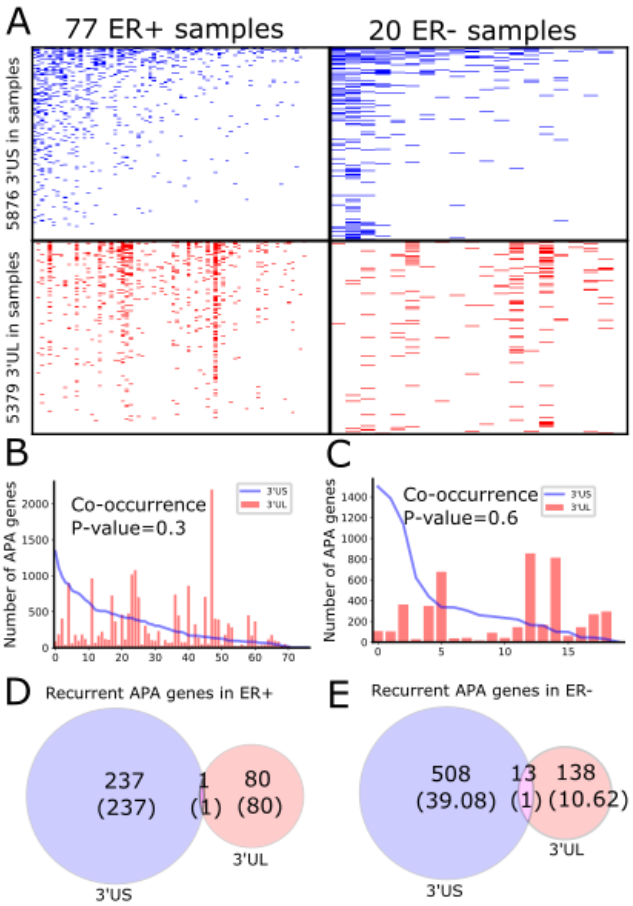


Figure 1. Global APA events distinct for ER+ and ER-. (A). Heatmaps showing the genes with 3′US (top panel) or 3′UL (bottom panel) in ER+ samples (left column) or ER- samples (right column). In ER+ (B) and ER- (C), the number of APA genes (3′US in line and 3′UL in red bar) in the tumor-normal sample pairs ordered as in Fig. A. Overlap of the recurring (>20% in tumor samples) 3′US and 3′UL genes in ER+ (D) and ER- (E).

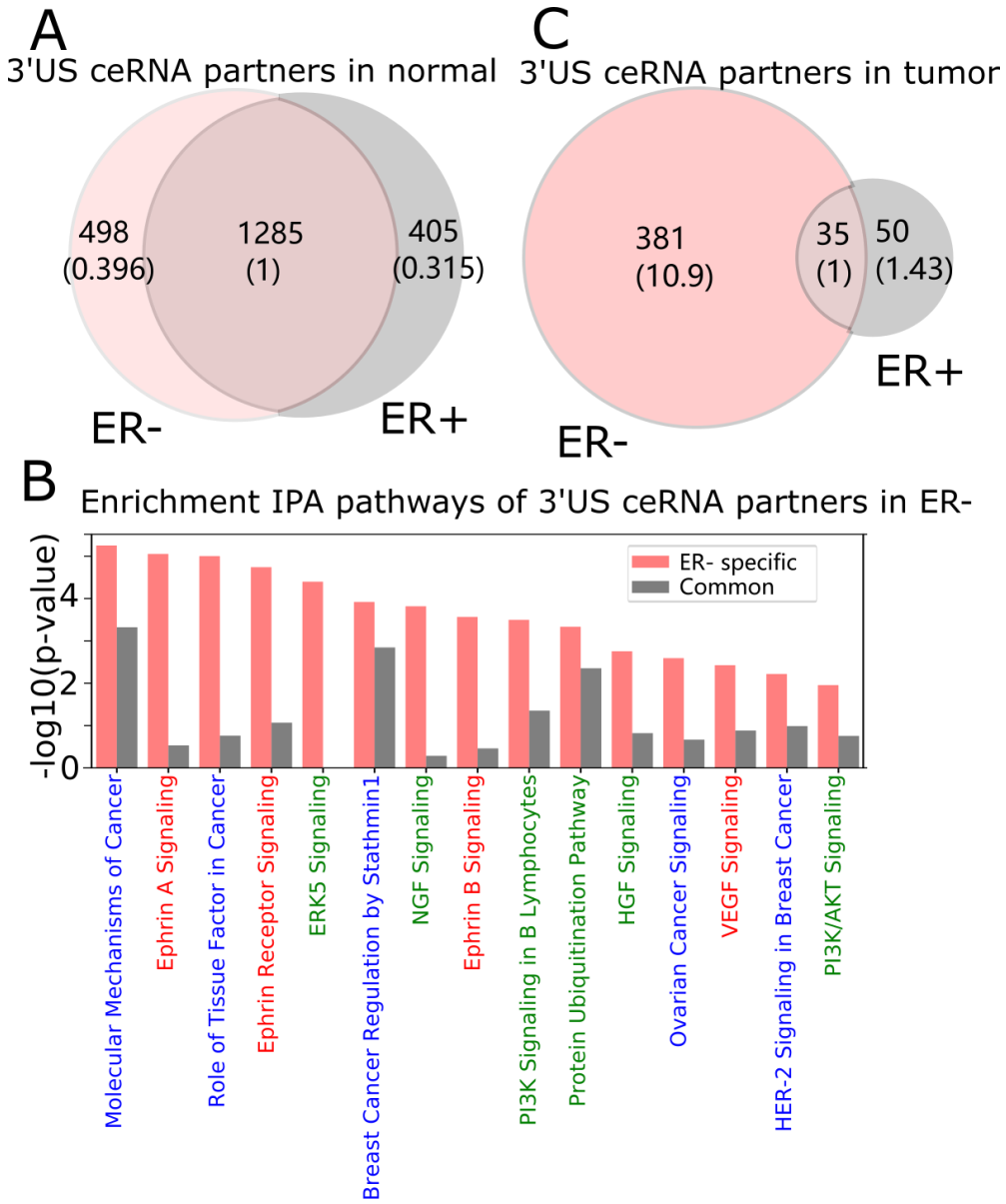


Figure 2. 3′UTR shortening is associated to ER-'s aggressive phenotypes in ceRNA. (A) Intersection of 3′US ceRNA partners between ER- and ER+ normal ceRNA networks. (B) IPA canonical pathways significantly ($P < 0.01$) enriched for the ER- 3′US ceRNAs. Pathways are colorcoded by keyword, "Cancer" in blue, "Signaling" in red and those associated with aggressive phenotypes[24] in green. (C) Intersection of 3′US ceRNA partners between ER- and ER+ tumor ceRNA networks.

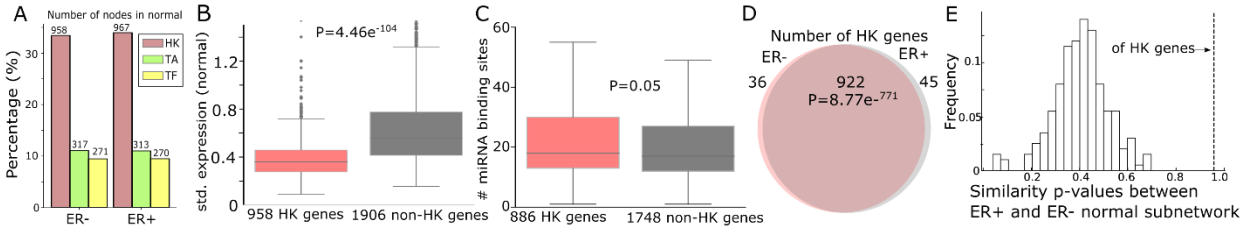


Figure 3. Housekeeping genes make consistent ceRNA networks between ER- and ER+ normal samples. (A) The number (and the percentage to the total number of nodes in the networks) of housekeeping (HK), tumor-associated (TA), or transcription factor (TF) genes in ER+ and ER- normal ceRNA networks. (B) Standard deviation of gene expressions of 958 HK genes and 1,906 non-HK genes in ER- normal ceRNA network. (C) Number of miRNA binding sites on the 3′UTR of 886 HK and 1,748 non-HK genes in the network (those that have miRNA binding site information). (D) Number of HK genes shared by ER- and ER+ normal ceRNA networks (with those in common). (E) Distribution of the similarity p-values between subnetworks sampled with 922 HK genes from ER+ and ER- normal networks (horizontal dotted line) and 200 subnetworks sampled with 1,990 non-HK genes to the same number of HK genes (bar). The higher the p-value is, the more similar the networks are[20].

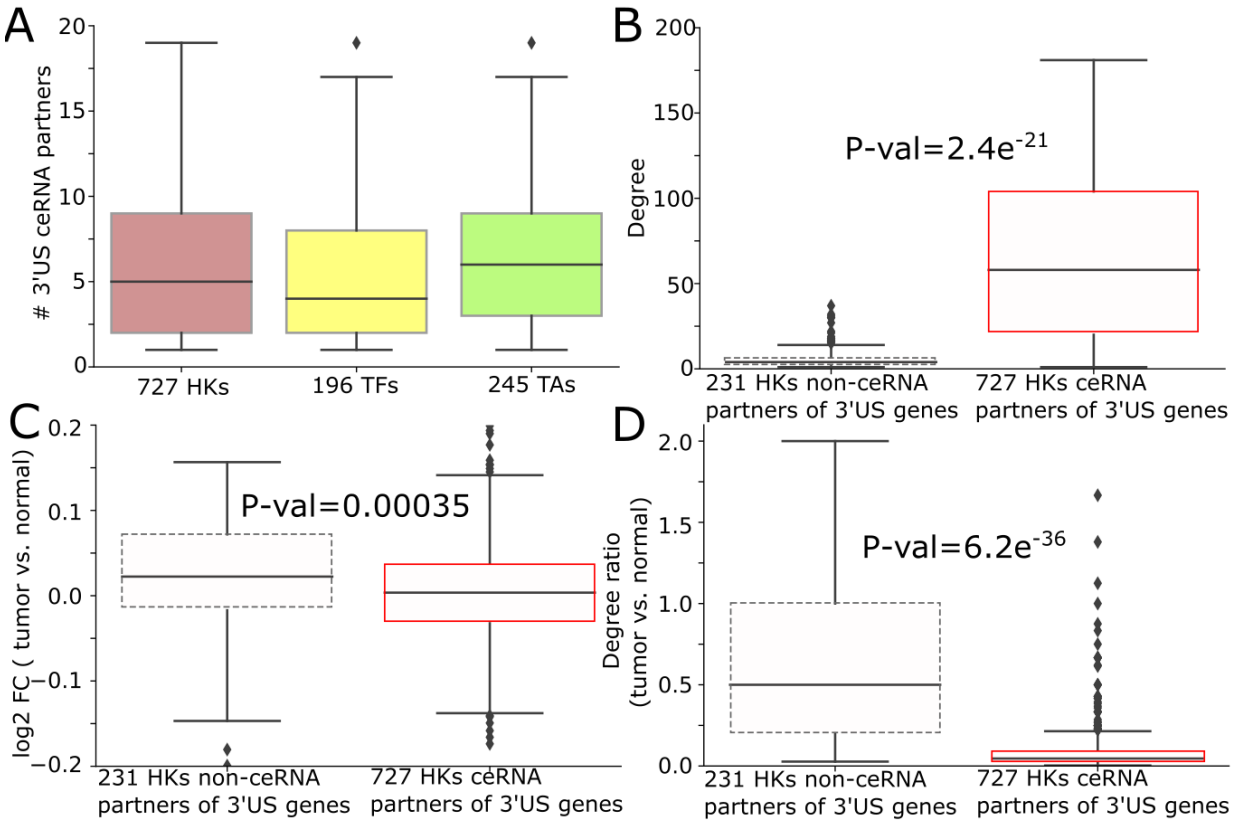


Figure 4. 3′US disrupts ceRNA relationship of HK genes. (A) The number of 3′UTR shortening genes connected to housekeeping (HK), transcription factor (TF), and tumor-associated (TA) genes. Degree (# neighbors in ERN normal ceRNA network) (B), log2 fold chance (tumor vs. normal) (C), degree ratio (tumor vs. normal) (D) of 727 and 231 HK genes that are ceRNA partners of 3′US genes or not, respectively. Since degree ratio in (D) represents the ratio of the number of neighbors retained in tumor, low degree values of 727 3′US HK ceRNA partners represents their higher loss of ceRNA neighbors in tumor.

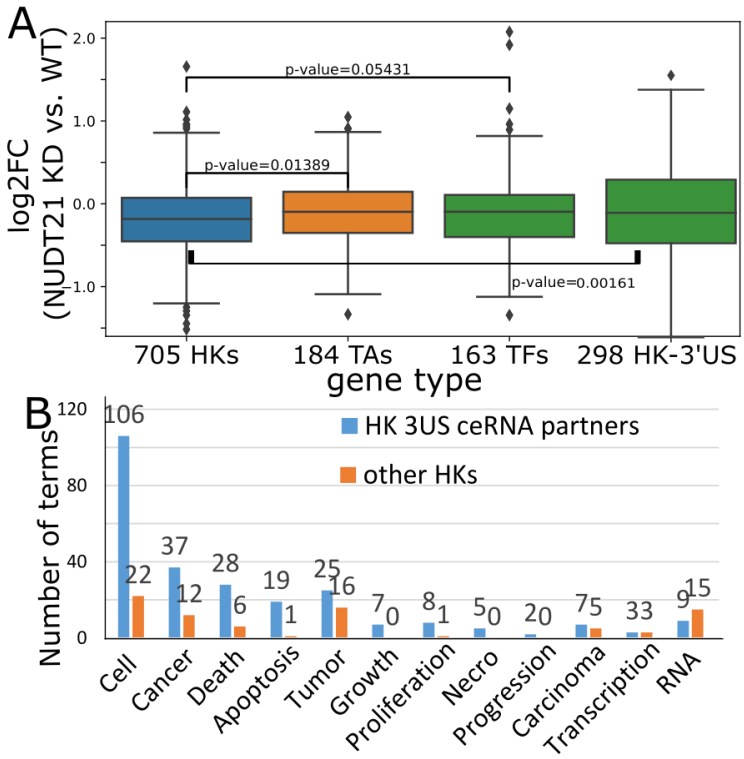


Figure 5 (A). log2FC (NUDT21 KD vs. WT) of 705 HK, 184 TA, and 163 TF genes that are (potential) ceRNA partners of 3'US genes. log2FC of 298 HK genes that are 3'US genes on the rightmost box. (B). Number of terms with the keyword indicated on the x-axis. Numbers on bar represent the actual number of terms.

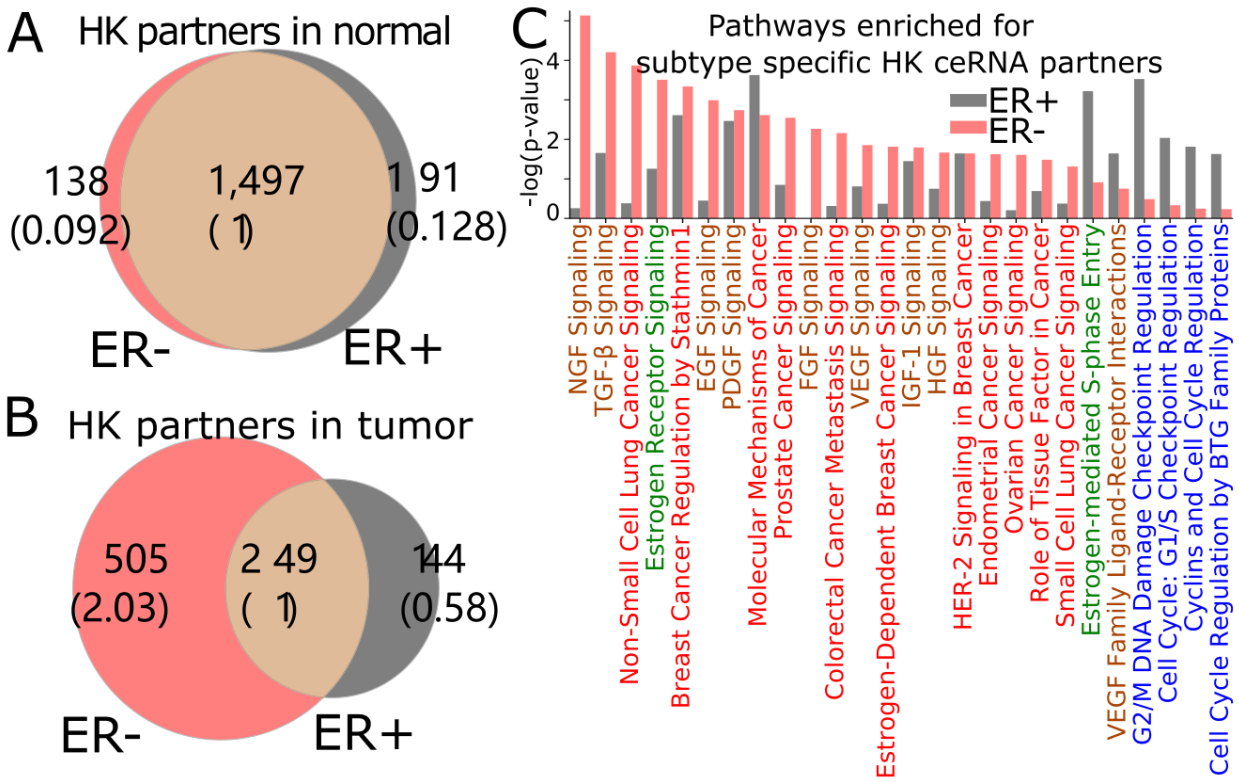
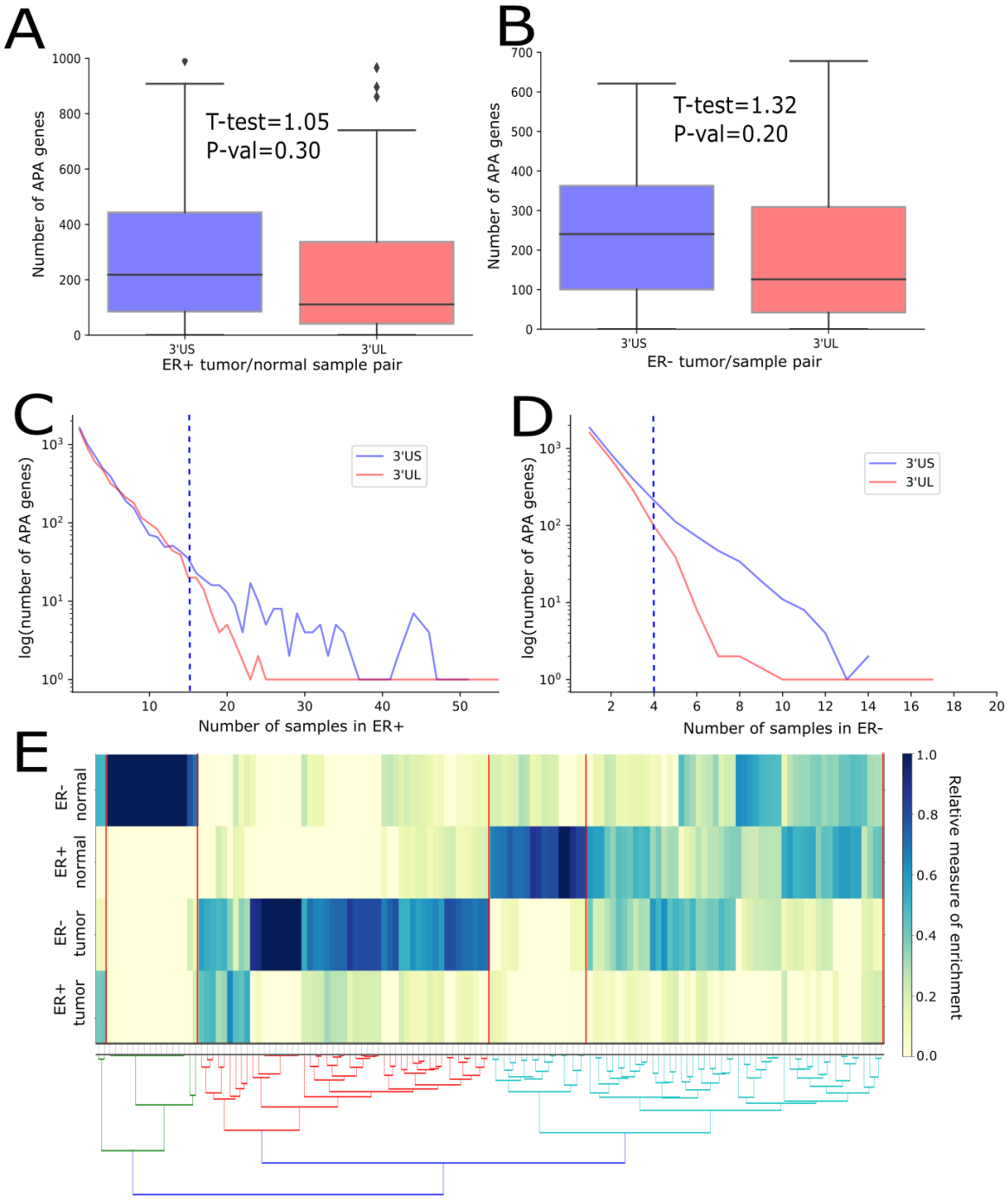
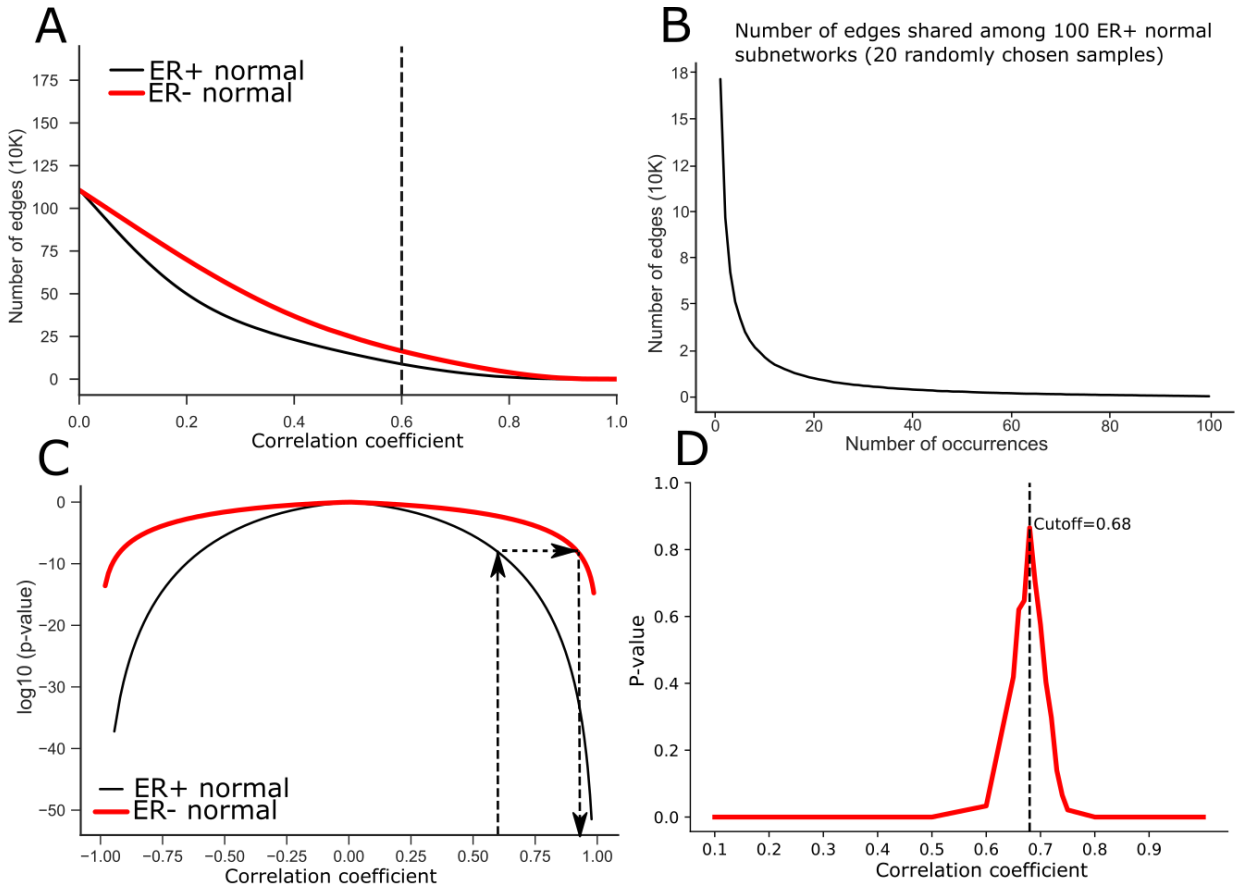


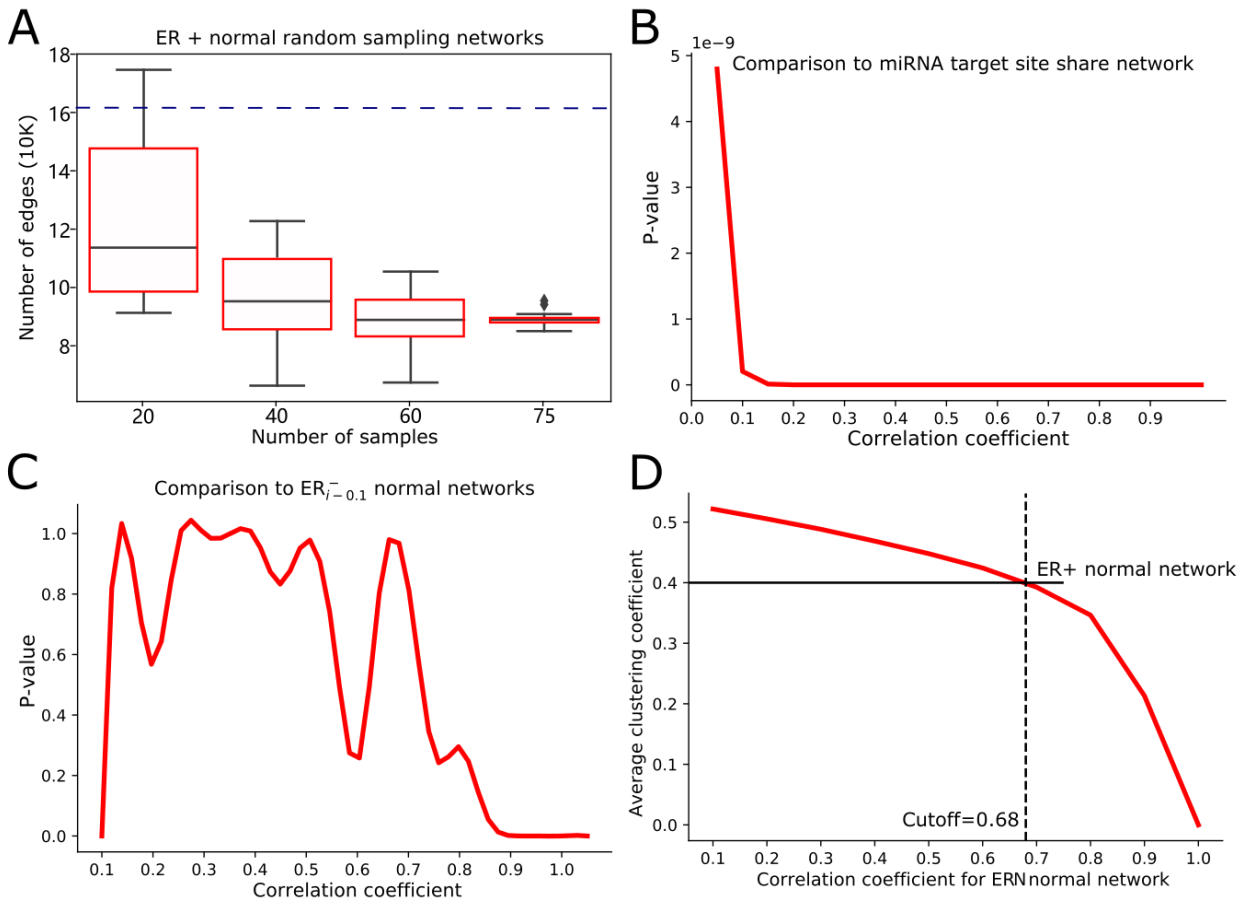
Figure 6. 3′US disrupts ceRNA relationship of HK genes for ER- specific growth. (A) Number of HK ceRNA partners unique and common to ER- and ER+ normal (left) and tumor (right) ceRNA networks. The numbers in parentheses are normalized to the number of genes shared between tumor and normal. (B) Degree in ER- tumor ceRNA network of 727 and 231 HK genes that are ceRNA partners of 3′US genes or not, respectively (C) IPA canonical pathways significantly ($P < 0.01$) enriched for ER+ and ER- specific HK ceRNA partners. Pathways are color-coded by keyword, “Cancer” in red, “GF” in brown, “Estrogen” in green, and “Cell Cycle” in blue.



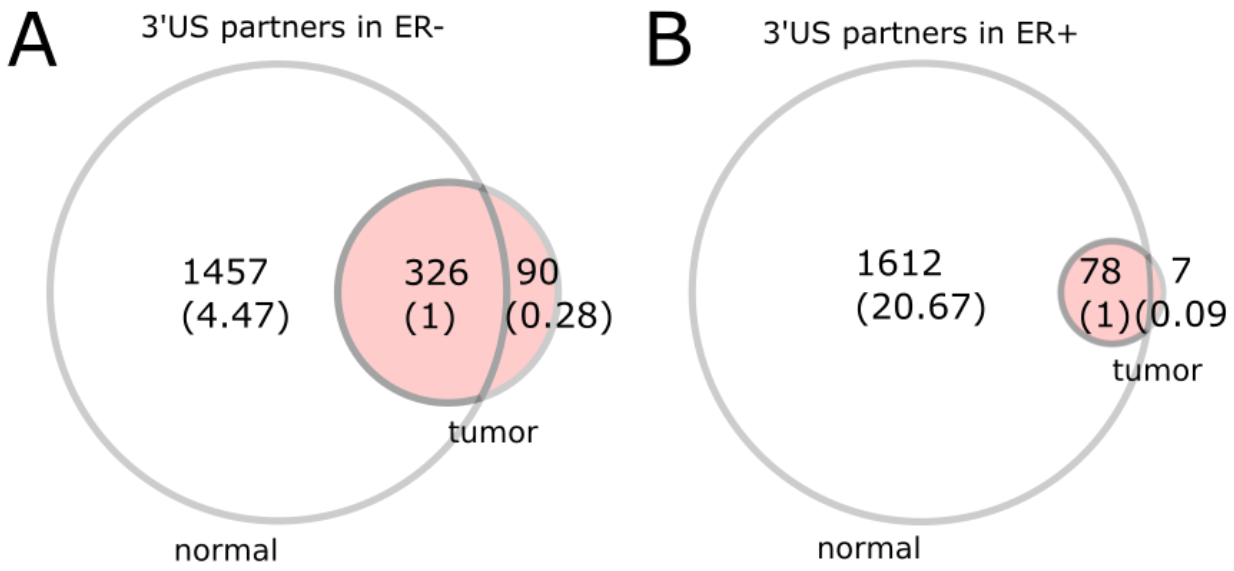
S. Figure 1. Boxplot showing the number of 3′US and 3′UL genes in each sample of ER+ (A) and ER- (B). T-test is t-test statistic value. Number of APA genes (y-axis blue in log10 for 3′US and red for 3′UL) common to ER+(C) and ER- (D) samples (x-axis) for ER+. The vertical dotted blue line marks the 20% threshold for recurrent events (E). IPA pathways enriched for 3′UL and 3′US genes in ER- and ER+. Colors represent enrichment of each pathway (column) for each class of genes (The higher the enrichment is, the higher the associated term is enriched). The red lines cut the pathways into 5 clusters, where each cluster is enriched in a set of genes.



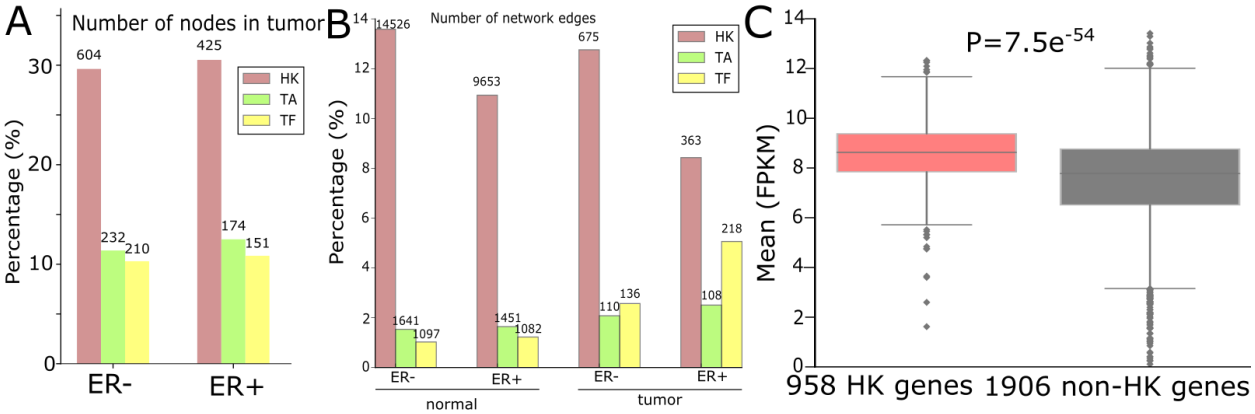
S. Figure 2. Two-step Pairwise Normalization of ER+ and ER- ceRNA network. (A) Number of edges in the ceRNA networks by the correlation coefficient cutoff (black and red line for ER+ and ER- networks, respectively). (B) The number of edges shared among 100 ER+ subnetworks from normal samples, where each of them was built by using 20 randomly chosen samples. (C) Statistical significance (p-value) achievable by using different correlation coefficient cutoff values for ER+ (black) and ER- (red) samples. Statistical significance for a correlation coefficient cutoff value is described in Methods. To achieve the same statistical significance of the traditional cutoff value (0.6) from ER+ to ER-, the cutoff value would inflate to 0.89, resulting in drastically a deflated number of edges. (D). Topological similarity (y-axis) between ER+ and ER- normal ceRNA networks by the cutoff value for ER- (x-axis). The bigger the p-value is, the more similar the two networks are (see Methods)[20]. The ER- normal network with the cutoff of 0.68 looks most similar to the ER+ normal reference network.



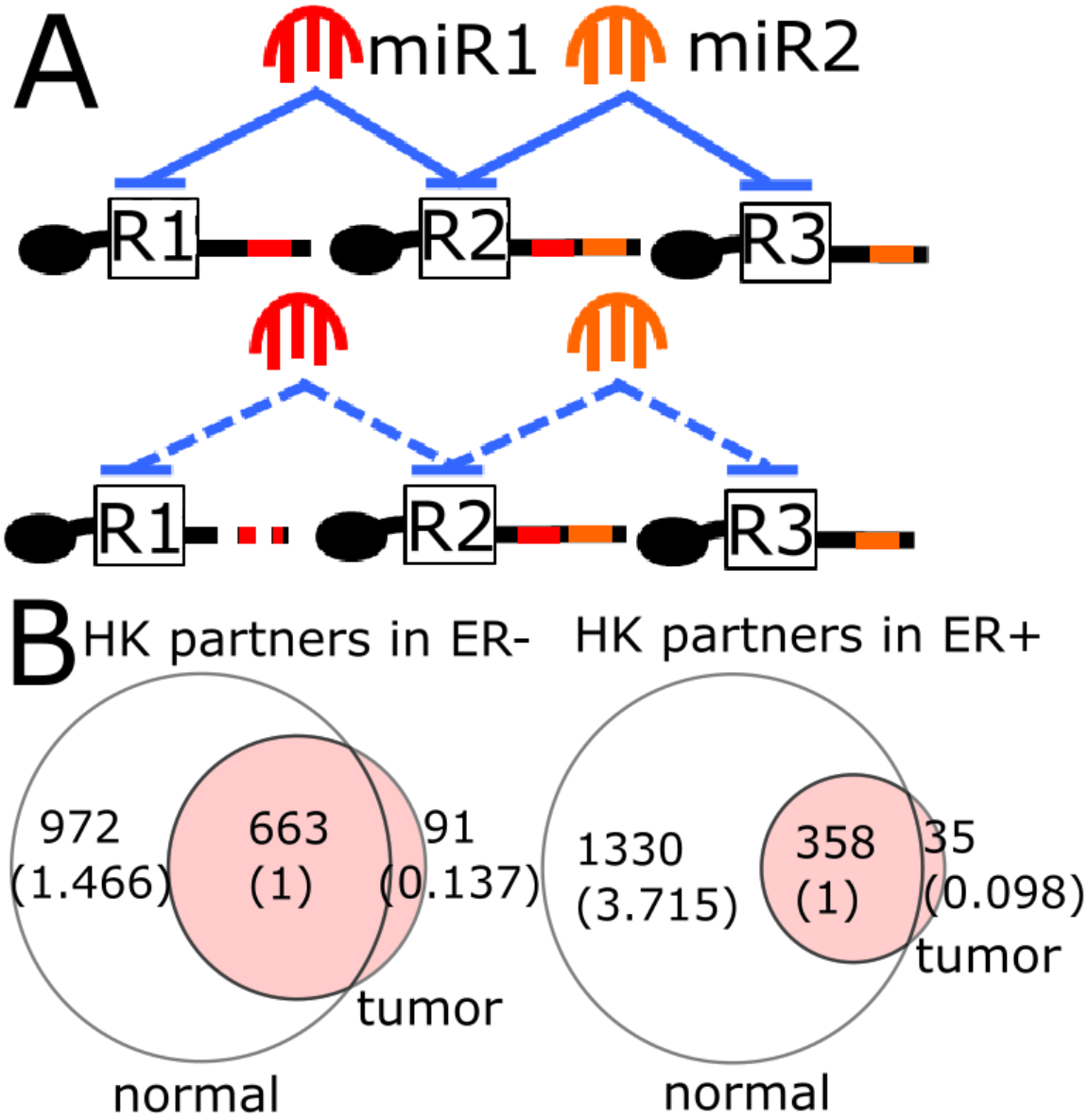
S. Figure 3. (A) Number of edges of networks from a subset of ER+ normal samples sampled in different size. Blue dotted line represents the number of edges of ER- normal network whose sample size is 20 (160,687) (B) Comparison of the miRNA target site share network to ER- normal networks with different correlation cutoff values (see Methods). As illustrated, ER- normal network topology changes drastically by different cutoff values in reference to the miRNA target site share network. (C) Comparison of ER- normal networks with previous correlation cutoff values in the stepwise increase (see Methods). (D) Comparing ER- normal ceRNA networks of different correlation cutoff values with the ER+ normal reference network in the average clustering coefficient. The average clustering coefficient for the ER+ normal reference network is 0.40 (indicated by the horizontal black line), which is quite close to the average clustering coefficient for the ER- normal network with a cutoff of 0.68 (indicated by the vertical dashed line).



S. Figure 4. Number of 3′US ceRNA partners in normal and tumor ceRNA networks found in (A) ER- and (B) ER+. The numbers in parentheses are normalized to the number of genes shared between tumor and normal.



S. Figure 5. (A) Number (and percentage to the total number of nodes in tumor networks) of HK genes and other important classes of genes in ER+ and ER- normal ceRNA networks. (B) The number (and the percentage to the total number of edges) of housekeeping (HK), transcription factor (TF), and 245 tumor-associated genes (TF) edges in ER- and ER+ for normal and tumor. (C) Average gene expression values of 1,003 HK genes and 1,990 non-HK genes in ER+ and ER- normal networks.



S. Figure 6. (A) Without 3'US (illustrated on top), R1 and R2 would compete for miR1 (red), forming ceRNA crosstalk and R2 and R3 would compete for miR2 (orange), forming ceRNA crosstalk. With 3'US on R1 (illustrated below), R1 would lose its ceRNA crosstalk with R2. Through indirect ceRNA effect that propagate the relationship loss, R2 would lose its ceRNA crosstalk with R3. (B) Numbers of HK ceRNA partners in ER- normal and in ER- tumor with overlap in common (The numbers in parentheses are normalized to the number of genes shared between tumor and normal). Numbers of HK ceRNA partners in ER+ normal and in ER- tumor (The numbers in parentheses are normalized to the number of genes shared between tumor and normal).

Acknowledgements

We thank Steffi Oesterreich, Ph.D., Professor and Vice Chair in Precision and Translational Pharmacology in the University of Pittsburgh and Susanne M. Gollin, Ph.D., Professor of the Department of Human Genetics in the University of Pittsburgh for their valuable suggestions. Also, this research was supported in part by the University of Pittsburgh Center for Research Computing through the resources provided.

Competing interests

The authors declares no competing financial interests.

Funding

This work was supported partly by Biomedical Modeling Awards and the Joan Gollin Gaines Cancer Research Fund at the University of Pittsburgh to H.J.P..

Author Contributions H.J.P conceived the project and designed the experiments. F.Z. performed the data analysis. S.K. conducted statistical tests. H.J.P. wrote the manuscript with input from B.D., S.K..

REFERENCES

[1] C. Mayr and D. P. Bartel, "Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells.," *Cell*, vol. 138, no. 4, pp. 673–84, Aug. 2009.

[2] C. P. Masamha and E. J. Wagner, "The contribution of alternative polyadenylation to the cancer phenotype," *Carcinogenesis*, vol. 39, no. 1, pp. 2–10, 2018.

[3] S. Kim, Y. Bai, B. Diergaarde, G. C. Tseng, and H. J. Park, "Alternative Polyadenylation Modifies Target Sites of MicroRNAs with Clinical Potential for Breast Cancer Progression," *bioRxiv*, p. 601518, Jan. 2019.

[4] M. Chen *et al.*, "3 ′ UTR lengthening as a novel mechanism in regulating cellular

senescence," pp. 285–294, 2018.

[5] G. P. Dimri *et al.*, "A biomarker that identifies senescent human cells in culture and in aging skin in vivo.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 92, no. 20, pp. 9363–7, 1995.

[6] R. A. Busuttil, M. Rubio, M. E. T. Dollé, J. Campisi, and J. Vijg, "Oxygen accelerates the accumulation of mutations during the senescence and immortalization of murine cells in culture.," *Aging Cell*, vol. 2, no. 6, pp. 287–294, 2003.

[7] C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, and G. Kroemer, "The hallmarks of aging," *Cell*, vol. 153, no. 6, 2013.

[8] D. Muñoz-Espín and M. Serrano, "Cellular senescence: From physiology to pathology," *Nat. Rev. Mol. Cell Biol.*, vol. 15, no. 7, pp. 482–496, 2014.

[9] Z. Xia *et al.*, "Dynamic Analyses of Alternative Polyadenylation from RNA- Seq Reveal Landscape of 3 ' UTR Usage Across 7 Tumor Types," *Nat. Commun.*, pp. 1–38, 2014.

[10] Y. Xiang *et al.*, "Comprehensive Characterization of Alternative Polyadenylation in Human Cancer," vol. 110, no. November 2017, pp. 1–11, 2018.

[11] L. Salmena, L. Poliseno, Y. Tay, L. Kats, and P. P. Pandolfi, "A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?," *Cell*, vol. 146, no. 3, pp. 353–8, Aug. 2011.

[12] P. Sumazin *et al.*, "An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma.," *Cell*, vol. 147, no. 2, pp. 370–81, Oct. 2011.

[13] H. J. Park, S. Kim, and W. Li, "Model-based analysis of competing- endogenous pathways (MACPath) in human cancers," *PLoS Comput. Biol.*, vol. 22, no. 14, 2018.

[14] H. J. Park *et al.*, "3′ UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk," *Nat. Genet.*, vol. 50, pp. 783–789, 2018.

[15] M. E. H. Hammond *et al.*, "American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version).," *Arch. Pathol. Lab. Med.*, vol. 134, no. 7, pp. e48-72, Jul. 2010.

[16] S. Tsutsui, S. Ohno, S. Murakami, Y. Hachitanda, and S. Oda, "Prognostic value of epidermal growth factor receptor (EGFR) and its relationship to the estrogen receptor status in 1029 patients with breast cancer," *Breast Cancer Res. Treat.*, vol. 71, no. 1, pp. 67–75, 2002.

[17] M. Sheikh, G. M, P. P, F. JA, and R. H, "Why are estrogen-receptor-negative breast cancers more aggressive t," *Invasion Metastasis*, vol. 14, no. 1–6, pp. 329–36, 1994.

[18] A. Pergamenschikov *et al.*, "Molecular portraits of human breast tumours," *Nature*, vol. 406, no. 6797, pp. 747–752, 2002.

[19] U. Ala *et al.*, "Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 18, pp. 7154–9, Apr. 2013.

[20] R. Gera *et al.*, “Identifying network structure similarity using spectral graph theory,” *Appl. Netw. Sci.*, vol. 3, no. 1, p. 2, 2018.

[21] C. C. Friedel and R. Zimmer, “Inferring topology from clustering coefficients in protein-protein interaction networks,” *BMC Bioinformatics*, vol. 15, pp. 1–15, 2006.

[22] P. Curmi *et al.*, “Overexpression of stathmin in breast carcinomas points out to highly proliferative tumours,” *Br. J. Cancer*, vol. 82, no. 1, pp. 142–150, 2000.

[23] S. Obayashi *et al.*, “Stathmin1 expression is associated with aggressive phenotypes and cancer stem cell marker expression in breast cancer patients,” *Int. J. Oncol.*, vol. 51, no. 3, pp. 781–790, 2017.

[24] K. Krishnan *et al.*, “miR-139-5p is a regulator of metastatic pathways in breast cancer.,” *RNA*, vol. 19, no. 12, pp. 1767–80, Dec. 2013.

[25] D. Vaught, D. M. Brantley-Sieders, and J. Chen, “Eph receptors in breast cancer: Roles in tumor promotion and tumor suppression,” *Breast Cancer Research*. 2008.

[26] E. Eisenberg and E. Y. Levanon, “Human housekeeping genes, revisited.,” *Trends Genet.*, vol. 29, no. 10, pp. 569–74, Oct. 2013.

[27] T. Davoli *et al.*, “Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome.,” *Cell*, vol. 155, no. 4, pp. 948–962, Oct. 2013.

[28] K. Chawla, S. Tripathi, L. Thommesen, A. Lægreid, and M. Kuiper, “TFcheckpoint: A curated compendium of specific DNA-binding RNA polymerase II transcription factors,” *Bioinformatics*, vol. 29, no. 19, pp. 2519–2520, 2013.

[29] Y. Tay, J. Rinn, and P. P. Pandolfi, “The multilayered complexity of ceRNA crosstalk and competition,” *Nature*, vol. 505, no. 7483, pp. 344–352, Jan. 2014.

[30] M. Nitzan, A. Steiman-Shimony, Y. Altuvia, O. Biham, and H. Margalit, “Interactions between distant ceRNAs in regulatory networks,” *Biophys. J.*, vol. 106, no. 10, pp. 2254–2266, 2014.

[31] C. P. Masamha *et al.*, “CFIm25 links alternative polyadenylation to glioblastoma tumour suppression.,” *Nature*, vol. 510, no. 7505, pp. 412–416, May 2014.

[32] J. Lu *et al.*, “MicroRNA expression profiles classify human cancers.,” *Nature*, vol. 435, no. 7043, pp. 834–8, Jun. 2005.

[33] G. Hong *et al.*, “Genes Dysregulated to Different Extent or Oppositely in Estrogen Receptor-Positive and Estrogen Receptor-Negative Breast Cancers,” *PLoS One*, vol. 8, no. 7, p. e70017, 2013.

[34] M. C. Alles *et al.*, “Meta-Analysis and Gene Set Enrichment Relative to ER Status Reveal Elevated Activity of MYC and E2F in the ‘Basal’ Breast Cancer Subgroup,” *PLoS One*, vol. 4, no. 3, p. e4710, 2009.

[35] M. C. Abba *et al.*, “Gene expression signature of estrogen receptor α status in breast cancer,” *BMC Genomics*, vol. 6, pp. 1–13, 2005.

[36] D. K. Biswas, a. P. Cruz, E. Gansberger, and a. B. Pardee, “Epidermal growth factor-induced nuclear factor kappa B activation: A major pathway of cell-cycle progression in estrogen-receptor negative breast cancer cells,” *Proc. Natl. Acad. Sci.*, vol. 97, no. 15, pp. 8542–8547, Jul. 2000.

[37] D. Fuckar *et al.*, “VEGF expression is associated with negative estrogen receptor status in patients with breast cancer,” *Int. J. Surg. Pathol.*, vol. 14, no. 1, pp. 49–55, 2006.

[38] S. Javanmoghadam, Z. Weihua, K. K. Hunt, and K. Keyomarsi, “Estrogen receptor alpha is cell cycle-regulated and regulates the cell cycle in a ligand-dependent fashion,” *Cell Cycle*, vol. 15, no. 12, pp. 1579–1590, 2016.

[39] S. Paruthiyil, H. Parmar, V. Kerekatte, G. R. Cunha, G. L. Firestone, and D. C. Leitman, “Estrogen Receptor Inhibits Human Breast Cancer Cell Proliferation and Tumor Formation by Causing a G 2 Cell Cycle Arrest,” pp. 423–428, 2004.

[40] D. C. Henley, J. S. Foster, J. Wimalasena, P. Seth, and A. Bukovsky, “Multifaceted Regulation of Cell Cycle Progression by Estrogen: Regulation of Cdk Inhibitors and Cdc25A Independent of Cyclin D1-Cdk4 Function,” *Mol. Cell. Biol.*, vol. 21, no. 3, pp. 794–810, 2002.

[41] E. Eisenberg and E. Levanon, “Human housekeeping genes are compact,” *TRENDS Genet.*, vol. 19, no. 7, pp. 362–365, 2003.

[42] M. Goldman *et al.*, “The UCSC Cancer Genomics Browser: update 2013.,” *Nucleic Acids Res.*, vol. 41, no. Database issue, pp. D949-54, Jan. 2013.

[43] B. P. Lewis, C. B. Burge, and D. P. Bartel, “Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.,” *Cell*, vol. 120, no. 1, pp. 15–20, Jan. 2005.

[44] G. L. Papadopoulos, M. Reczko, V. a Simossis, P. Sethupathy, and A. G. Hatzigeorgiou, “The database of experimentally supported targets: a functional update of TarBase.,” *Nucleic Acids Res.*, vol. 37, no. Database issue, pp. D155-8, Jan. 2009.

[45] F. Xiao, Z. Zuo, G. Cai, S. Kang, X. Gao, and T. Li, “miRecords: An integrated resource for microRNA-target interactions,” *Nucleic Acids Res.*, vol. 37, no. November 2008, pp. 105–110, 2009.

[46] S.-D. Hsu *et al.*, “miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions.,” *Nucleic Acids Res.*, vol. 42, no. Database issue, pp. D78-85, Jan. 2014.

[47] H. Dvinge *et al.*, “The shaping and functional consequences of the microRNA landscape in breast cancer.,” *Nature*, vol. 497, no. 7449, pp. 378–82, May 2013.

[48] M. P. Hamilton *et al.*, “Identification of a pan-cancer oncogenic microRNA superfamily anchored by a central core seed motif.,” *Nat. Commun.*, vol. 4, p. 2730, Jan. 2013.