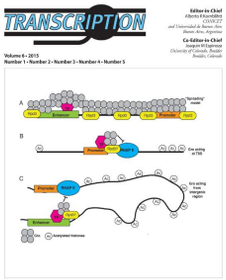

Transcription

ISSN: 2154-1264 (Print) 2154-1272 (Online) Journal homepage: http://www.tandfonline.com/loi/ktrn20

lnc00673 (ERRLR01) is a Prognostic Indicator of Overall Survival in Breast Cancer

Ubaidat Abdul-Rahman, Balázs Gyoörffy & Brian D. Adams

To cite this article: Ubaidat Abdul-Rahman, Balázs Gyoörffy & Brian D. Adams (2017): lnc00673 (ERRLR01) is a Prognostic Indicator of Overall Survival in Breast Cancer, Transcription, DOI: 10.1080/21541264.2017.1329684

To link to this article: http://dx.doi.org/10.1080/21541264.2017.1329684

View supplementary material

Accepted author version posted online: 10 Aug 2017.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=ktrn20

Download by: [Australian Catholic University] **Date:** 12 August 2017, At: 05:46

Check for updates

lnc00673 (*ERRLR01*) is a Prognostic Indicator of Overall Survival in Breast Cancer

Ubaidat Abdul-Rahman[1], Balázs Győrffy[2,3], and Brian D. Adams[1,4,5]

[1]The RNA Institute, University at Albany, State University of New York, 1400 Washington Ave, Albany, NY 12222, USA

[2]MTA TTK Lendület Cancer Biomarker Research Group, Hungarian Academy of Sciences, Magyar Tudósok körútja 2., 1117, Budapest, Hungary

[3]Semmelweis University 2nd Dept. of Pediatrics, Tűzoltó utca 7--9., 1094, Budapest, Hungary

[4]Investigative medicine Program, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520, USA

[5]Corresponding Author: Brian D. Adams, Ph.D. Principle Investigator / Research Faculty , The RNA Institute / State University of New York, 1400 Washington Ave, Albany, NY 12222 Teaching Faculty, Investigative Medicine Program, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520 , **Telephone:** 518-437-4447**; Fax:** (518) 437--4456 , **Email:** brian.adams@yale.edu or bdadams@albany.edu

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Abstract

LncRNAs are novel noncoding RNAs involved in the epigenetic regulation of gene expression by recruiting ribonucleoprotein complexes to specific genomic loci to initiate histone methylation and/or other chromatin modifications. LncRNAs themselves function as tumor suppressors or oncogenes, depending on the gene regulatory networks they govern. We identified lnc00673 (*ERRLR01*) as a marker of overall survival (OS) in breast cancer patients. Specifically, *ERRLR01* levels were elevated in triple-negative breast cancer (TNBC) as compared to Luminal- A, Luminal-B, and HER2 breast cancer subtypes. *ERRLR01* levels were also inversely correlated with breast cancer survival across all breast cancer patients. Upon stratification, OS in ERα^{-} tumors correlated with negative survival outcomes, while in ERα^{+} tumors, *ERRLR01* correlated with positive outcomes. This suggests *ERRLR01* is modulated by hormone signaling in breast cancer. Gene-network analysis revealed *ERRLR01* correlated with distinct pathways including “epithelial development” and “cellular differentiation". These data suggest *ERRLR01* could be an oncogene in TNBC, as well as a biomarker in breast cancer patients.

Keywords

lncRNA, Breast Cancer, Survival Outcomes, Estrogen Signaling, Hormone- Regulation, ERRLR01

Accepted Manuscript

Introduction

Long-noncoding RNAs function as decoys, regulators of translation, and/or molecular scaffolds that recruit chromatin modifying enzymes to distinct genomic loci[1-4]. LncRNA transcripts are transcribed in sense and/or anti-sense orientation[5,6,7], and lncRNA transcription depends upon a tightly controlled regulatory network that as of yet, is still not well understood. The ENCODE Consortium determined the abundance and location of many lncRNAs across several species, which in various cases are conserved through positional synteny[8,9]. In many cases, lncRNAs reside near protein coding regions and function in *cis* to maintain the surrounding chromatin in an open and epigenetically active state. In other cases, lncRNAs can be transcribed in an anti-sense orientation to a protein-coding gene, which in turn recruits histone chromatin-modifying complexes that support gene silencing[10,11]. The nomenclature for a particular lncRNA is derived, in part, from the closest neighboring protein coding gene. For instance, lincRNA-p21 neighbors *CDKN1A*, regulates CDKN1A (p21) expression, and when lincRNA-p21 is knocked down phenocopies the effects imparted by the loss of CDKN1A[12,13]. Additionally, genes such as ANRIL (antisense noncoding RNA in the INK4 locus), which is an antisense transcript produced within the CDKN2B locus, recruits PRC1/2 complexes and inhibits CDKN2B gene expression[14,15].

There has been a recent effort to elucidate lncRNA function by identifying particular cellular states and/or pathophysiologies associated with dysregulated lncRNA expression[16]. In this context, modulating lncRNA abundance, either through gain- or loss-of-function approaches serves as an opportunity to understand which gene-regulatory networks a specific lncRNA is associated with. Furthermore, alteration of lncRNA expression that in turn promotes changes in cellular phenotypes indicates a particular lncRNA could be a driver of a disease or pathophysiological state. Several lncRNAs have been identified as important mediators of disease states, such as *PTCSC3* in papillary thyroid carcinoma, and *ANRIL* in type-2 diabetes[17,18]. Additionally, *MALAT-1* is expressed in the heart, lung, and kidneys, but is elevated in lung tumor tissue and has elevated biological activity in metastatic lung adenocarcinoma as compared to a benign or pre-malignant state[19-22]. *HOTAIR*, initially described in fibroblasts, regulates the cluster of HOX genes through *cis* as well as *trans*-regulatory mechanisms[16], and is associated with oncogenesis. *HOTAIR* expression levels are high in several tumors types including lung, breast, and prostate. Furthermore, ectopic expression of *HOTAIR* results in aberrant PRC2 function and improper recruitment of PRC2-associted complexes to the correct genomic loci. This process facilitates promiscuous gene regulation allowing for a pro-tumorigenic state[23]. Several other lncRNAs including *PCAT-1* and *GAS5* also have tumorigenic function; however, there is a growing need to elucidate the function of the nearly 118,000 recently annotated human lncRNAs. Given lncRNAs serve as decoys, function as scaffolds that mediate protein-protein interactions, and promote chromatin remodeling through recruitment of DNA- modifying enzymes to distinct genomic loci, the capability to tease out specific mechanisms associated with a particular lncRNA remains a challenge. However, determining putative lncRNA functions can be initiated through investigation of associative miRNA and mRNA gene regulatory networks, which could then be utilized to determine if a lncRNA harbors putative

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

oncogenic potential when dysregulated.

In this study, we identified *ERRLR01*, as a lncRNA whose expression was differentially abundant across four breast cancer subtypes within Affymetrix and The Cancer Genome Atlas (TCGA) datasets. Furthermore, *ERRLR01* expression correlated with overall survival (OS) and relapse-free survival (RFS) outcomes within an Affymetrix dataset. Previous reports indicated *ERRLR01* was a prognostic marker in non-small cell lung cancer[24], and functioned as a pro- metastatic lncRNA in melanoma[25]. In our model, *ERRLR01* was aberrantly expressed across breast cancer tumor subtypes and cell lines. Therefore, we performed gene network analysis and identified a series of pathways highly correlated with *ERRLR01* expression, and further recognized putative miRNA-lncRNA interactions that could be responsible for modulating *ERRLR01* expression in specific cellular contexts. Overall, these studies indicate *ERRLR01* may function as an oncogene in breast cancer, and that anti-sense strategies developed to target this lncRNA may eventually become a new therapeutic approach to treat breast cancer patients.

Materials and Methods:

Gene chip database construction

To develop survival analysis software, a gene chip database was firs established as described previously[26]. In brief, a GEO search was made to identify breast cancer gene expression datasets with available survival data, with each dataset containing at least 30 patient samples. The raw .CEL files were downloaded and MAS5 normalized in the R statistical environment (http://www.r-project.org) using the Affy Bioconductor library. Database quality control and removal of duplicate samples were performed as described[27]. This analysis methodology was then utilized on a curated Affymetrix dataset described in Gyorffy, B *et. al.,* 2013[39]. Patient samples from the Affymetrix datasets were primarily of Caucasian origin, and biopsies were obtained from the primary breast tumor. This curated dataset contains over 2000 breast cancer patient samples, all containing survival endpoints, and a majority of the samples are ERα+[39].

Specifically, Affymetrix dataset analysis was performed whereby expression of *ERRLR01* in each patient sample was determined using probe set 227452_. The original signal was MAS5 normalized, and listed in rank-order via spearman rank analysis[38,39]. For *ERRLR01*, the probe set 227452_at was used. This probe set has a specificity of one *ERRLR01* transcript and a coverage of 100% of the transcript variants of *ERRLR01*[28]. *ESR1* and *HER2* status was determined for each sample using the probe sets 205225_at and 216836_s_at as described earlier[29]. To determine receptor status, the raw expression cutoff values of 500 and 1,150 were used for *ESR1* and *HER2*, respectively.

RNA-Seq database construction

RNA-Seq measurement for breast cancer patients[30] were published by The Cancer Genome Atlas (TCGA) of the National Cancer Institute (https://cancergenome.nih.gov/) and we downloaded the pre-processed level 3 data generated using the Illumina HiSeq 2000 RNA Sequencing Version 2 platform. The primary tumor samples from this dataset were obtained from patients of Caucasian origin. For these

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

samples, expression levels were determined using a combination of MapSplice and RSEM. We have combined the individual patient files in R using the plyr package. TCGA datasets also underwent Kaplan-Meir survival analysis utilizing the statistical methodologies highlighted below.

Quantitative Real Time PCR

For qPCR, cells were lysed in TRIzol (Life Technologies), and total RNA was used as the input for subsequent RT-PCR reactions. For lncRNA analysis, cDNA synthesis and qRT-PCR were performed according to the miScript II RT Kit (Qiagen) and miScript SYBR® Green PCR Kit (Qiagen) protocols respectively[31]. MiRNA levels were normalized to RNU6B levels. For mRNA analysis, cDNA synthesis and qRT-PCR were performed per the RT^2 First Strand Kit and RT^2 qPCR Primer assay protocols from Qiagen.

Genetic-Network and Go Term Analysis

For gene pathways analysis, Spearman-Rank analysis was performed on the entire Affymetrix data. We identified the top 200 genes most highly correlated with *ERRLR01* across all patient specimens. The top 200 genes are equivalent to roughly the top 5% of genes in the Spearman- Rank analysis. We utilized this gene set to perform GO term analysis using the WebGestalt algorithm[55]. In these studies, we performed a series of data queries including GO term analysis, KEGG analysis, and miRNA target analysis. The algorithm compares the enrichment of genes uploaded in the program compared to reference gene names of the entire human genome. Utilizing hypergeometric analysis with Benjamini and Hochberg multiple test adjustment (i.e., FDR Analysis), we identified significant gene pathways using a cutoff of *p value* < 0.05. Only pathways containing 6 genes or more from the genes loaded into the algorithm were analyzed. Spearman-Rank analysis was also utilized to identify genes that were the most anti-correlated with *ERRLR01*. Similar statistical analysis was performed, and the resulting data is highlighted in the supplemental data.

Statistical analysis

Molecular subtypes were designated per the StGallen guidelines using expression of ESR1, HER2, and MKI67 (PMID: 23917950). This includes a TNBC cohort (ESR1 and HER2 negative patients), a HER2-enriched cohort (HER2 positive and ESR1 negative patients), a Luminal A cohort (ESR1 positive HER2 negative with low MKI67), and a Luminal B cohort (ESR1 positive and HER2 positive as well as ESR1 positive HER2 negative with high MKI67 expression).

Survival analysis was performed in the R statistical environment (http://www.r-project.org) using the survival Bioconductor library. For the expression of *ERRLR01*, each percentile (of expression) between the lower and upper quartiles was computed and the best performing threshold was used as the final cutoff in the Cox regression analysis. Kaplan–Meier survival plot, and the hazard ratio with 95% confidence intervals and log-rank P value were calculated and plotted in R. Kruskal-Wallis test was used to compare continuous expression among multiple cohorts and Mann-Whitney test was used to compare two cohorts. Statistical significance was set at $p<0.05$.

Results

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

ERRLR01 is Highly Expressed in Estrogen Receptor Negative Tumor Subtypes

We performed an *in-silico* screen of lncRNAs that were differentially expressed across breast cancer patients, or between normal and breast cancer tissue samples. We identified *ERRLR01* as a long noncoding RNA expressed at higher levels in breast cancer patient samples as compared to normal breast tissue. We utilized numerous databases including MiTranscriptome[32], which is a meta-assembly of 6,503 RNA-Seq libraries (***Figure S1***). *ERRLR01* is a recently described lncRNA that has sequence similarity to the steroid receptor RNA activator 1, and hence termed *ERRLR01* (SRA-like-non-coding RNA). *ERRLR01* is located on chromosome 17q24.3, a region associated with genomic breakpoints within a variety of cancer types including breast, lung, colorectal, and uterine leiomyomas[33-37]. Additionally, *ERRLR01* was identified to play a role in melanoma invasion[25], and was linked with worse overall survival in melanoma patients. *ERRLR01* expression was also previously reported to be elevated in melanoma cell lines post invasion, suggesting *ERRLR01* promotes invasion. Given the role of *ERRLR01* in melanoma, and the finding that *ERRLR01* was elevated in breast cancer specimens as compared to normal breast tissue, we decided to investigate whether *ERRLR01* could serve as a biomarker for breast cancer progression and/or disease onset.

ERRLR01 is a Predictor of Overall Survival in Breast Cancer

We analyzed Affymetrix U133 gene chip data, as well as the breast cancer TCGA dataset and determined that across four distinct breast cancer subtypes, as determined by PAM50 classification, that *ERRLR01* was elevated in TNBC patient samples (***Figure 1***). The analysis of the Affymetrix dataset was performed by Spearman rank analysis whereby expression of *ERRLR01* in each patient sample was determined by probe set 227452_ and whereby the signal

was MAS5 normalized, and listed in rank-order[38,39]. The average rank-order across all patients within the sample subset was determined and plotted as a box-whisker-plot (***Figure 1a***). ERRLR01 expression was elevated in both TNBC as well as $HER2^+$ patient subtypes when compared to Luminal-A and Luminal-B subtypes. This indicated *ERRLR01* was inversely correlated with ERα status in breast cancer patients. To validate this result, we assessed the expression of *ERRLR01* in the breast cancer TCGA dataset[30]. Here, *ERRLR01* expression was determined using the MapSplice Algorithm[40], and the resultant log(2) transformed normalized data was plotted as a box-whisker-plot (***Figure 1b***). In this dataset ERRLR01 expression was elevated in both TNBC and HER2 positive patient samples. Together, results from both patient datasets indicated that *ERRLR01* expression was elevated in $ERα^-$ tumor subtypes, as compared to $ERα^+$ tumor subtypes (*p value* $< 1 \times 10^{-16}$, as determined by Kruskal-Wallis analysis).

This analysis indicated that *ERRLR01* may be inversely correlated with hormone signaling, as hormone-sensitive tumors expressed lower levels of *ERRLR01* than those with hormone- independent tumors. To test this hypothesis Kaplan-Meir survival analysis was performed, coupled with Cox regression analysis on the Affymetrix dataset described above. We determined that within all breast cancer patient samples, high *ERRLR01* expression inversely correlated with RFS (***Figure 2a***); HR = 1.5 with a *p value* $< 5 \times 10^{-6}$. When segregating patients by ERα status, we found that $ERα^+$ patients with high expression of *ERRLR01* had

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

better overall survival (OS) rates than those with low *ERRLR01* expression (***Figure 2c***); HR = 0.75 with a *p value* < 0.0046. We did not perform RFS analysis after patient stratification due to smaller sample size and reduced statistical power. Interestingly, in patients with ERα⁻ tumors, *ERRLR01* expression was inversely correlated with OS (***Figure 2b***); HR = 1.31 with a *p value* < 0.095. This was a unique finding and suggests *ERRLR01* is regulated by ERα or that ERα signaling controls *ERRLR01* expression.

To test this notion further, we assessed *ERRLR01*–related OS in breast cancer patients after PAM50 sub-stratification. Within Basal-like (i.e., TNBC) and HER2-positive breast cancer categories high *ERRLR01* expression inversely correlated with OS, while in Luminal-A and Luminal-B categories *ERRLR01* expression correlated with OS (***Figure S2***). We attempted to validate this survival relationship with *ERRLR01* using the TCGA datasets, however, we did not find any significant survival benefit based on differential *ERRLR01* expression. It should be noted that survival data in TCGA is harder to determine, given a majority of samples were acquired from primary tumor samples, with only 5--10yr follow-up.

Characterization of the ERRLR01 Locus and Interacting Transcription Factors

Given the significant *ERRLR01*-related OS between hormone-dependent and hormone- independent breast cancer patients, we wanted to elucidate a putative mechanism for such a relationship. To do this we first determined the genomic location of *ERRLR01* and inquired as to which putative regulatory moieties could serve as signals to recruit potential RNA and protein binding components to the *ERRLR01* transcript. *ERRLR01* is positionally conserved via synteny between the human and mouse genome (***Figure 3a***). A moderate level of transcription was observed from both RNA-Seq data on 9 cell lines from ENCODE, and from lncRNA RNA-Seq datasets where *ERRLR01* was expressed highest in testes and brain tissue, but lower in normal breast and other tissues. Furthermore, evidence of CpG promoter methylation was observed (Methylation Score = 117) using the UCSC genome browser dataset. Overall these data support the notion that *ERRLR01* is expressed at low to moderate levels in most normal tissue specimens or cell lines. These findings are also in line with the Affymetrix and TCGA datasets, where the raw FPKMs for *ERRLR01* were 10 or less.

Upon further characterization of the *ERRLR01* genomic loci we identified H3K4Me1 and H3K4Me3 marks between *ERRLR01* exons three and four, and in some cases overlapped with CTCF binding sites, as well as other transcription factors (***Figure 3b***). The convergence of CTCF binding on the *ERRLR01* genomic loci was of particular interest since CTCF is a well- documented 17b-estradiol regulated protein whose function is that of an insulator to prevent enhancer element binding to particular promoters[41-47]. As an example, CTCF is a factor known to be regulated by ERα in MCF-7 cells, and therefore could also affect *ERRLR01* levels in these estrogen-responsive cell lines. This notion is supported by GEO datasets, (see ***Figure S3***). Therefore, *ERRLR01* could operate as an oncogene as is suggested by some groups, yet in hormone-sensitive tumors a 17b estradiol-CTCF regulatory axis could be reducing the activity of *ERRLR01*. In support of this, at least three ERα binding sites were identified on the 3' end of the *ERRLR01* transcript. Whether these sites are functional or serve as co-activator versus co- repressor binding sites is unclear.

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

Further *in silico* analysis indicated that the family of GATA transcription factors (including GATA3) also bound to the exonic regions of *ERRLR01* (***Figure 3b***). Further work is required to determine the functionality of these sites.

Potential RNA- and miRNA- Binding Sites within the ERRLR01 Locus

To further understand the potential RNA-regulatory network that *ERRLR01* is a part of, we utilized several databases to interrogate whether certain RNA-binding proteins interacted with *ERRLR01*, or whether specific miRNAs could bind the *ERRLR01* transcript. The strongest RNA binding protein interaction we identified was FUS (***Figure 4***). FUS binding to *ERRLR01* was confirmed through the StarBase 2.0 algorithm[45] and is denoted as a HHF3_128133 binding site (***Figure 4a***). Other RNA binding protein sites were identified through HITS-CLIP data; however, a majority of these sites require confirmation through RNA-IP experiments before concluding these interactions are relevant. It was interesting to note that DGCR8 bound to intronic regions of *ERRLR01* indicating that a miRNA sequence may be derived from this host sequence.

Next we asked whether there were any regulatory interactions between miRNAs and the *ERRLR01* transcript. Using the miRdB algorithm[46] we observed several putative miRNA interactions across *ERRLR01*, and highlighted the top 5 miRNA candidates (***Figure 4b***). Many of these miRNAs have unknown functions, however miR-515 is a well-studied oncogene in breast cancer[47-50], and is involved in mediating or regulating EMT[51-53]. miR-515 is also an estrogen regulated gene[48], where ERα can bind the promoter of miR-515 and regulate gene expression in a 17b-estradiol-dependent manner.

Overall these bioinformatic analyses indicated that *ERRLR01* was a highly regulated lncRNA predictive of overall survival in breast cancer patients, and could be regulated by 17b-estradiol signaling in breast cancer. To test this notion, we screened for *ERRLR01* expression across several breast cancer cell lines. We confirmed that several TNBC cell lines harbored high expression of *ERRLR01* as compared to $ERα^+$ cell lines, and normal HMECs (***Figure 5***). The high expression of *ERRLR01* in TNBC cell lines indicates that *ERRLR01* may be repressed by 17b-estradiol signaling. However, this analysis was done under steady state growth conditions whereby 17b-estradiol as present in the growth media. However, a recent study by Lin *et al.*[54], indicated that in MCF-7 cells, *ERRLR01* levels are low under hormone-depleted conditions yet upon 17b-estradiol addition *ERRLR01* levels increase within 3hrs (***Figure S3***). The consequences of this rise in *ERRLR01* levels requires elucidation, as numerous transcription factors and RNA binding proteins may be modulated by this change in *ERRLR01* levels. It would be interesting to determine in follow-up studies whether direct modulation of *ERRLR01* induces functional and/or phenotypic consequences in breast cancer cells.

Gene Regulatory Network Analysis Associated with ERRLR01 Expression

Given the evidence for *ERRLR01* to be an estrogen-regulated gene in breast cancer, we wanted to identify signaling pathways that may be part of an *ERRLR01*-mediated RNA network. To do this we employed GO term analysis using the WebGestalt algorithm[55]. We extracted the

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

gene expression profiles from all patients within the Affymetrix dataset and performed Spearman-Rank analysis to determine which genes were most correlated with *ERRLR01* expression (***Figure 6***). We used as a cut-off the 200 most positively correlated *ERRLR01* genes, which is approximately the top 5% genes derived from the analysis. We then performed a series of GO term and KEGG analysis using WebGestalt against the entire human reference genome, and found significantly enriched *ERRLR01* pathways (***Figure 6a***). We identified pathways such as Wnt Receptor Signaling ($p<7.70 \times 10^{-2}$), Epithelial Differentiation ($p<7.7 \times 10^{-2}$), and Epithelial Development ($p<7.70 \times 10^{-2}$). Importantly, pathways such as Epithelial Development and Hormone Signaling were recurring pathways associated with *ERRLR01*-correlated genes. Therefore, we conclude that *ERRLR01* is a lncRNA tightly regulated and serves to control proper developmental timing of mammary gland differentiation and development.

Similar analysis was performed on the 200 most inversely correlated *ERRLR01*-associated genes (***Figure 6b***). Pathways such as Small Molecule Metabolic Processes ($p<8.80 \times 10^{-3}$), Cellular Lipid Metabolic Processes ($p<8.80 \times 10^{-3}$), and Reproductive System Development ($p<3.12 \times 10^{-2}$) were identified. Similar analysis was performed on the TCGA dataset (see ***Figures S4-S5***).

Discussion

ERRLR01 is located on chr17:70,396,217-70,590,488, and was first identified by Schmidt *et al*., where *ERRLR01* transcriptionally upregulated MMP9 in in melanoma cell lines[25]. In this study, *ERRLR01* was compared across 150 clinical samples from TCGA and found to be elevated in tumors with greater thickness. Furthermore, patient samples that were greater than 14.1 RPKM had a worse OS outcome than patients with less than 14.1 RPKM. This indicated that *ERRLR01* operated as an oncogene in melanoma. *ERRLR01* was described as an oncogene in other cancer models as well, including lung and pancreatic cancer[24,56]. In lung cancer *ERRLR01* expression associated with higher TMN stages as well as lymph node metastasis. *ERRLR01* also functioned as a pro-metastatic factor in melanoma, since siRNA knockdown studies indicate that loss of *ERRLR01* resulted in reduced invasion as measured by Boyden-Chamber Matrigel assays[25]. In pancreatic cancer, the direct evidence for *ERRLR01* as an oncogene is less well determined, as a particular pancreatic cancer risk variant was identified within the *ERRLR01* transcript that generated a miR-1231 binding site, which presumably supports oncogenesis. More specifically, *ERRLR01* modulated PTPN11 degradation by promoting E3 ligase induced ubiquitination of PTPN11 and diminished ERK oncogenic signaling. Therefore, in pancreatic cancer the presumption is that *ERRLR01* operates as a tumor suppressor by dampening the pro-proliferative MAPK/ERK signaling pathway.

ERRLR01 could also be a potential therapeutic target given knockdown of *ERRLR01* in lung cancer cell lines and in mouse xenograft models resulted in reduced cell viability, and reduced cellular growth *in vivo*. *ERRLR01* does this by binding directly to LSD1 (KDM1A) and functions as a chaperone protein to recruit this histone demethylase to NCALD[24]. LSD1 is known to regulate cellular differentiation and cell cycle progression, while NCALD is a visinin-like protein 1 sub family of EF-hand calcium-binding proteins that is

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

downregulated in patients with poor prognosis tumors and those harboring poorly differentiated ovarian tumors[57]. Therefore, the mechanism of action is such that *ERRLR01* recruits LSD1 to epigenetically silence NCALD to promote tumorigenesis. The mechanism of action in melanoma is slightly different given *ERRLR01* serves as a scaffold to bring two protein components in proximity to each other, Brn3a and AR. The interaction with *ERRLR01* allows these protein heterodimers to bind to specific chromatin regions, within a site located in the promoter of MMP9[25]. The transcriptional upregulation of MMP9 imparted by *ERRLR01* promotes a metastatic phenotype in melanoma cell lines. Overall, these studies highlight the notion that *ERRLR01* mediates oncogenic or tumor suppressor functions dependent upon the cellular context, which is true for many non-coding RNAs.

While *ERRLR01* is described as a chaperone protein, our bioinformatics analysis identified *ERRLR01* as part of a gene-regulatory network involving miRNA sponging and interactions with specific RNA-binding proteins. This was an important analysis given the biological significance of both small and long noncoding RNAs are becoming increasingly appreciated. We further identified several SNPs across the *ERRLR01* transcript, however it was unclear as to whether any of these polymorphisms were associated with disease risk and/or onset. We intend on pursuing this line of analysis given genome-wide association studies (GWAS) have identified cancer risk loci outside protein-coding regions, such as *PTCSC3* in papillary thyroid carcinoma, and *ANRIL* in type-2 diabetes[17,58,59].

A major finding of this study was that 17b-estradiol regulated *ERRLR01* expression. A few studies have identified specific lncRNAs to be regulated by 17b-estradiol. For instance, *linc00160* expression can be modulated by 17b-estradiol, however the mechanism of action is unclear[60]. Similarly, well-studied lncRNAs have been identified to be regulated by 17b-estradiol, including H19, HOTAIR, and MALAT-1[61-63]. In these studies, it is unclear how these lncRNAs relate to survival in breast cancer patients, and furthermore, if these lncRNAs are associated with ERα-stratified overall survival. In this study, we provide strong evidence that *ERRLR01* is prognostic in breast cancer, that *ERRLR01*–related OS is dependent upon ERα status, and that *ERRLR01* itself is regulated by ERα signaling and17b-estradiol-associated coregulatory proteins in breast cancer. Overall, the continued efforts to understand this epigenetic regulation mediated by lncRNAs in the context of cancer development will aid in the generation of more effective therapeutic strategies to treat the disease[64].

Disclosure of Conflicts of Interest

B.D.A holds patent interests with, and consults with AUM LifeTech. B.D.A is also the President of The Brain Institute of America, LLC. The other authors have no conflicts of interest to disclose.

Acknowledgements:

We would like to thank our many colleagues for the review of this manuscript. We thank the State of New York and The Research Foundation for startup funds to B.D.A for making this manuscript possible. B.G. was supported by the OTKA 108655 grant. We also thank Kim DeWeerd at the Molecular Core and Tissue Core for providing the equipment necessary to perform these experiments.

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

Contributions

Manuscript Writing – Ubaidat Abdul-Rahman, Balázs Győrffy, and Brian D. Adams

TGCA and Affymetrix Bioinformatic Analysis – Balázs Győrffy

KEGG and Pathway Analysis – Brian D. Adams

Concept and Design - Brian D. Adams

Experimental Design - Ubaidat Abdul-Rahman and Brian D. Adams

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

References

1. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81,** 145–166 (2012).
2. Rinn, J. L. LncRNAs: Linking RNA to chromatin. *Cold Spring Harb. Perspect. Biol.* **6,** (2014).
3. Huarte, M. The emerging role of lncRNAs in cancer. *Nat. Med.* **21,** 1253–61 (2015).
4. Goff, L. A. & Rinn, J. L. Linking RNA biology to lncRNAs. *Genome Research* **25,** 1456– 1465 (2015).
5. Morris, K. V. Long antisense non-coding RNAs function to direct epigenetic complexes that regulate transcription in human cells. *Epigenetics: official journal of the DNA Methylation Society* **4,** 296–301 (2009).
6. Villegas, V. E. & Zaphiropoulos, P. G. Neighboring gene regulation by antisense long Non- Coding RNAs. *International Journal of Molecular Sciences* **16,** 3251–3266 (2015).
7. Morris, K. V. & Vogt, P. K. Long antisense non-coding RNAs and their role in transcription and oncogenesis. *Cell Cycle* **9,** 2544–2547 (2010).
8. Elnitski, L. L. *et al.* The ENCODEdb portal: Simplified access to ENCODE Consortium data. *Genome Res.* **17,** 954–959 (2007).
9. ENCODE Project Consortium, T. E. P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (80-.).* **306,** 636–40 (2004).
10. Werner, A. Biological functions of natural antisense transcripts. *BMC Biol.* **11,** 31 (2013).
11. Werner, A. Natural Antisense Transcripts. *RNA Biol.* **2,** 53–62 (2005).
12. Tang, S.-S., Zheng, B.-Y. & Xiong, X.-D. LincRNA-p21: Implications in Human Diseases. *Int. J. Mol. Sci.* **16,** 18732–18740 (2015).
13. Dimitrova, N. *et al.* LincRNA-p21 Activates p21 In cis to Promote Polycomb Target Gene Expression and to Enforce the G1/S Checkpoint. *Mol. Cell* **54,** 777–790 (2014).
14. Pasmant, E., Sabbagh, A., Vidaud, M. & Bièche, I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J.* **25,** 444–448 (2011).
15. Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **17,** 47–62 (2016).
16. Prensner, J. R. & Chinnaiyan, A. M. The emergence of lncRNAs in cancer biology. *Cancer Discovery* **1,** 391–407 (2011).
17. Cheetham, S. W., Gruhl, F., Mattick, J. S. & Dinger, M. E. Long noncoding RNAs and the genetics of cancer. *Br. J. Cancer* **108,** 2419–25 (2013).
18. Jendrzejewski, J. *et al.* The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc Natl Acad Sci U S A* **109,** 8646–8651 (2012).
19. Gutschner, T. *et al.* The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* **73,** 1180–1189 (2013).
20. Zhu, L., Liu, J., MA, S. & Zhang, S. Long Noncoding RNA MALAT-1 Can Predict Metastasis and a

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

Poor Prognosis: a Meta-Analysis. *Pathol. Oncol. Res.* **21,** 1259–1264 (2015).

21. Zhou, Y. *et al.* The long noncoding RNA MALAT-1 is highly expressed in ovarian cancer and induces cell growth and migration. *PLoS One* **11,** (2016).
22. Arun, G. *et al.* Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev.* **30,** 34–51 (2016).
23. Hajjari, M. & Salavaty, A. HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer Biol. Med.* **12,** 1–9 (2015).
24. Shi, X. *et al.* Upregulation of long intergenic noncoding RNA 00673 promotes tumor proliferation via LSD1 interaction and repression of NCALD in non-small-cell lung cancer. *Oncotarget* **7,** (2016).
25. Schmidt, K. *et al.* The lncRNA ERRLR011 Mediates Melanoma Invasion through a Conserved SRA1-like Region. *Cell Rep.* **15,** 2025–2037 (2016).
26. Mihály, Z. *et al.* A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. *Breast Cancer Research and Treatment* **140,** 219–232 (2013).
27. Győrffy, B. *et al.* RecurrenceOnline: an online analysis tool to determine breast cancer recurrence and hormone receptor status using microarray data. *Breast Cancer Res. Treat.* **132,** 1025–1034 (2012).
28. Li, Q., Birkbak, N. J., Gyorffy, B., Szallasi, Z. & Eklund, A. C. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics* **12,** 474 (2011).
29. Gyorffy, B. *et al.* RecurrenceOnline: An online analysis tool to determine breast cancer recurrence and hormone receptor status using microarray data. *Breast Cancer Res. Treat.* **132,** 1025–1034 (2012).
30. Cancer, T. & Atlas, G. Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).
31. Adams, B. D. *et al.* MiR-34a silences c-SRC to attenuate tumor growth in triple-negative breast cancer. *Cancer Res.* **76,** 927–939 (2016).
32. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47,** 199–208 (2015).
33. Cotterill, S. Chromosome 14. *Cancer Genet.* http://www.cancerindex.org/geneweb/clinkc14.htm- (2015).
34. Zhang, X., Cowper-Sal-lari, R., Bailey, S. D., Moore, J. H. & Lupien, M. Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res.* **22,** 1437–1446 (2012).
35. Pezzolo, A. *et al.* Loss of 10q26.1-q26.3 in association with 7q34-q36.3 gain or 17q24.3- q25.3 gain predict poor outcome in pediatric medulloblastoma. *Cancer Lett.* **308,** 215–224 (2011).
36. Sun, J. *et al.* Association between sequence variants at 17q12 and 17q24.3 and prostate cancer risk in European and African Americans. *Prostate* **68,** 691–697 (2008).
37. Solomon, E., Borrow, J. & Goddard, a D. Chromosome aberrations and cancer. *Science* 14 **254,** 1153–60 (1991).
38. Szász, A. M. *et al.* Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget* **7,** 49322–49333 (2016).

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

39. Gyorffy, B., Surowiak, P., Budczies, J. & Lánczky, A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* **8,** (2013).
40. Wang, K. *et al.* MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* **38,** (2010).
41. Phillips, J. E. & Corces, V. G. CTCF: Master Weaver of the Genome. *Cell* **137,** 1194–1211 (2009).
42. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cell* **60,** 676– 684 (2015).
43. Ong, C. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Publ. Gr.* **15,** 234–246 (2014).
44. Ohlsson, R., Bartkuhn, M. & Renkawitz, R. CTCF shapes chromatin by multiple mechanisms: The impact of 20 years of CTCF research on understanding the workings of chromatin. *Chromosoma* **119,** 351–360 (2010).
45. Yang, J. H. *et al.* StarBase: A database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.* **39,** (2011).
46. Wong, N. & Wang, X. miRDB: An online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* **43,** D146–D152 (2015).
47. Mattiske, S., Suetani, R. J., Neilsen, P. M. & Callen, D. F. The oncogenic role of miR-155 in breast cancer. *Cancer Epidemiology Biomarkers and Prevention* **21,** 1236–1243 (2012).
48. Pinho, F. G. *et al.* Downregulation of microRNA-515-5p by the estrogen receptor modulates Sphingosine kinase 1 and breast cancer cell proliferation. *Cancer Res.* **73,** 5936–5948 (2013).
49. Pardo, O. E. *et al.* miR-515-5p controls cancer cell migration through MARK4 regulation. *EMBO Rep.* **17,** 570–84 (2016).
50. Gilam, A. *et al.* Involvement of IGF-1R regulation by miR-515-5p modifies breast cancer risk among BRCA1 carriers. *Breast Cancer Res. Treat.* **138,** 753–760 (2013).
51. Craene, B. De & Berx, G. Regulatory networks defining EMT during cancer initiation and progression. *Nat. Rev. Cancer* **13,** 97–110 (2013).
52. Abdelmohsen, K., Srikantan, S., Kuwano, Y. & Gorospe, M. miR-519 reduces cell proliferation by lowering RNA-binding protein HuR levels. *Proc Natl Acad Sci U S A* **105,** 20297–20302 (2008).
53. Abdelmohsen, K. *et al.* miR-519 suppresses tumor growth by reducing HuR levels. *Cell Cycle* **9,** 1354–1359 (2010).
54. Lin, Z., Reierstad, S., Huang, C. C. & Bulun, S. E. Novel estrogen receptor-?? binding sites and estradiol target genes identified by chromatin immunoprecipitation cloning in breast cancer. *Cancer Res.* **67,** 5017–5024 (2007).
55. Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: An integrated system for exploring gene sets in various

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

biological contexts. *Nucleic Acids Res.* **33,** (2005).

56. Zheng, J. *et al.* Pancreatic cancer risk variant in LINC00673 creates a miR-1231 binding site and interferes with PTPN11 degradation. *Nat. Genet.* **48,** 747–757 (2016).
57. Nymark, P. *et al.* Integrative analysis of microRNA, mRNA and aCGH data reveals asbestos- and histology-related changes in lung cancer. *Genes Chromosom. Cancer* **50,** 585–597 (2011).
58. Pasmant, E., Sabbagh, A., Vidaud, M. & Bièche, I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J.* **25,** 444–448 (2011).
59. Jendrzejewski, J. *et al.* The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc. Natl. Acad. Sci.* **109,** 8646–8651 (2012).
60. Jonsson, P. *et al.* Single-Molecule Sequencing Reveals Estrogen-Regulated Clinically Relevant lncRNAs in Breast Cancer. *Mol. Endocrinol.* **29,** 1634–45 (2015).
61. Sun, H. *et al.* H19 lncRNA mediates 17β-estradiol-induced cell proliferation in MCF-7 breast cancer cells. *Oncol. Rep.* **33,** 3045–3052 (2015).
62. Zhao, Z., Chen, C., Liu, Y. & Wu, C. 17β-Estradiol treatment inhibits breast cell proliferation, migration and invasion by decreasing MALAT-1 RNA level. *Biochem. Biophys. Res. Commun.* **445,** 388–393 (2014).
63. Bhan, A. & Mandal, S. S. Estradiol-Induced Transcriptional Regulation of Long Non-Coding RNA, HOTAIR. *Methods Mol. Biol.* **1366,** 395–412 (2016).
64. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521,** 232–236 (2015).

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

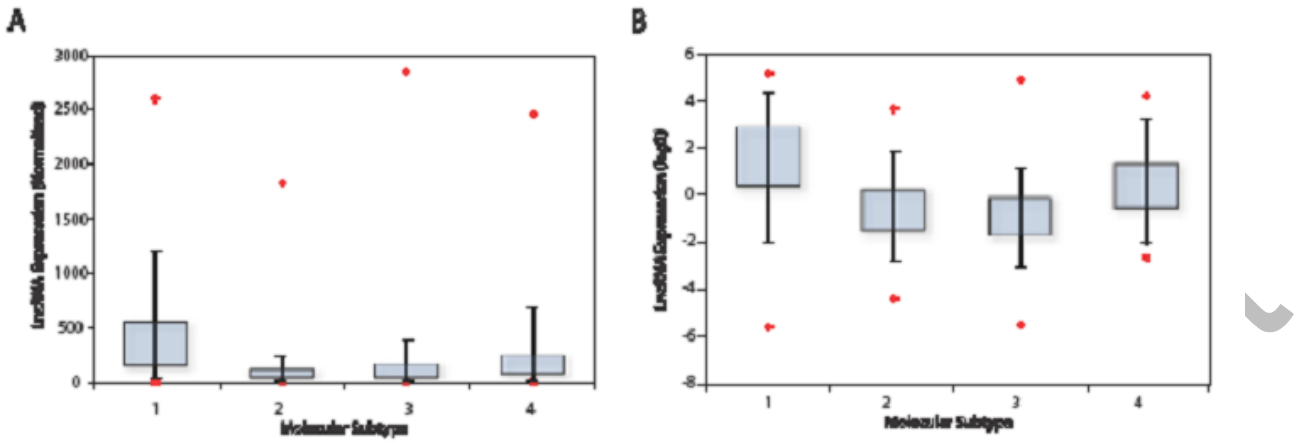

Figure 1: LncRNA expression in breast cancer patient populations which was acquired using an Affymetrix U133A array dataset, and is depicted by box-whisker-plots. In this analysis, *ERRLR01* expression was stratified into 4 subpopulations, and the mean rank expression was reported **(panel A)**. 1 = TNBC, n = 577; 2 = Luminal A, n = 1432; 3 = Luminal B, n = 632; 4 = HER2+, n = 301. *ERRLR01* expression in the TCGA dataset **(panel B)**, and data is presented in a log(2) transformed format. 1 = TNBC, n = 154; 2 = Luminal A, n = 91; 3 = Luminal B, n = 538; 4 = HER2+, n = 53. *denotes significance at $p<1 \times 10^{-16}$ as determined by Kruskal-Wallis one-way variance analysis testing.

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

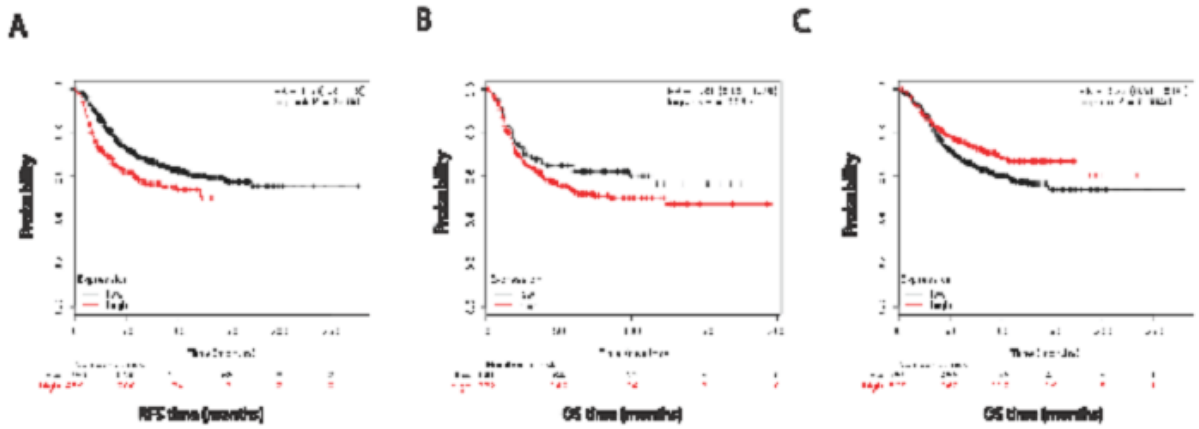

Figure 2: Kaplan-Meir and cox regression analysis of *ERRLR01* levels and OS in breast cancer patients. Data was obtained from Affymetrix datasets, and analyzed on KM-Plotter software. Panel **(A)** represents RFS in all breast cancer patients, **(B)** depicts OS in ER-negative breast cancer patients, and panel **(C)** depicts OS in ER-positive breast cancer patients. Hazard ratios and log- rank *p values* are reported for each analysis. *ERRLR01* correlates with poor survival outcomes in “all patients” as well as “ER-negative patients”, while positively correlates with OS in “ER- positive patients”.

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

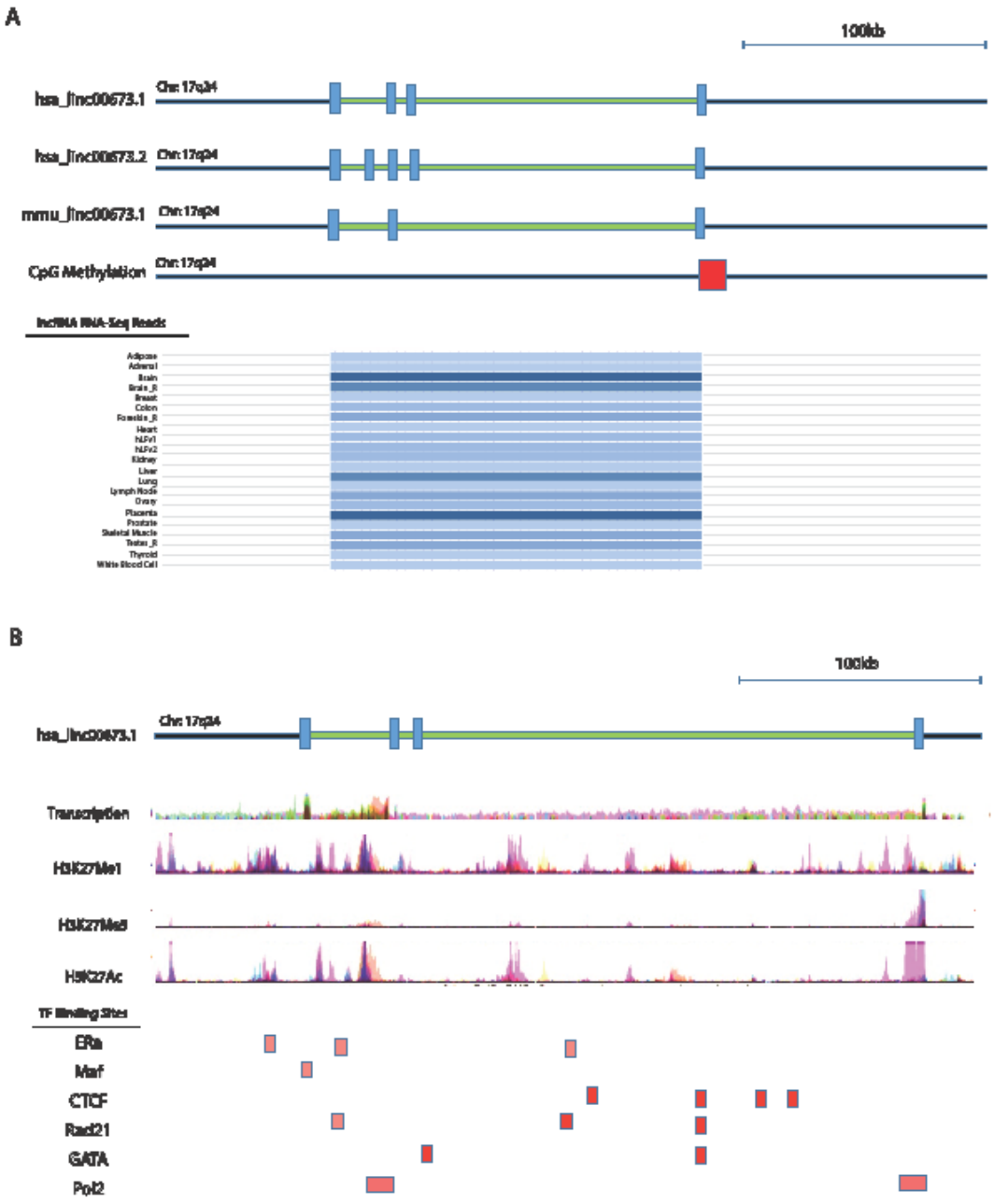

Figure 3: Genomic analysis of the *ERRLR01* gene region and the regulatory factors that interact with *ERRLR01* using the UCSC Genome browser. Transcription orientation is depicted from right (5' end) to left (3' end). **(A)** *ERRLR01.1* is a 4-exon gene conserved between mouse and human, and is expressed at high levels in testis, placenta, and brain tissue. **(B,** *Top Panel***)** Detectable transcription is present across a panel of cell lines with evidence of H3K4Me1 modification. **(B,** *Bottom Panel***)** Potential estrogen-regulated transcription factor binding locations are denoted, such as the GATA family of transcription factors, CTCF, and ERα itself.

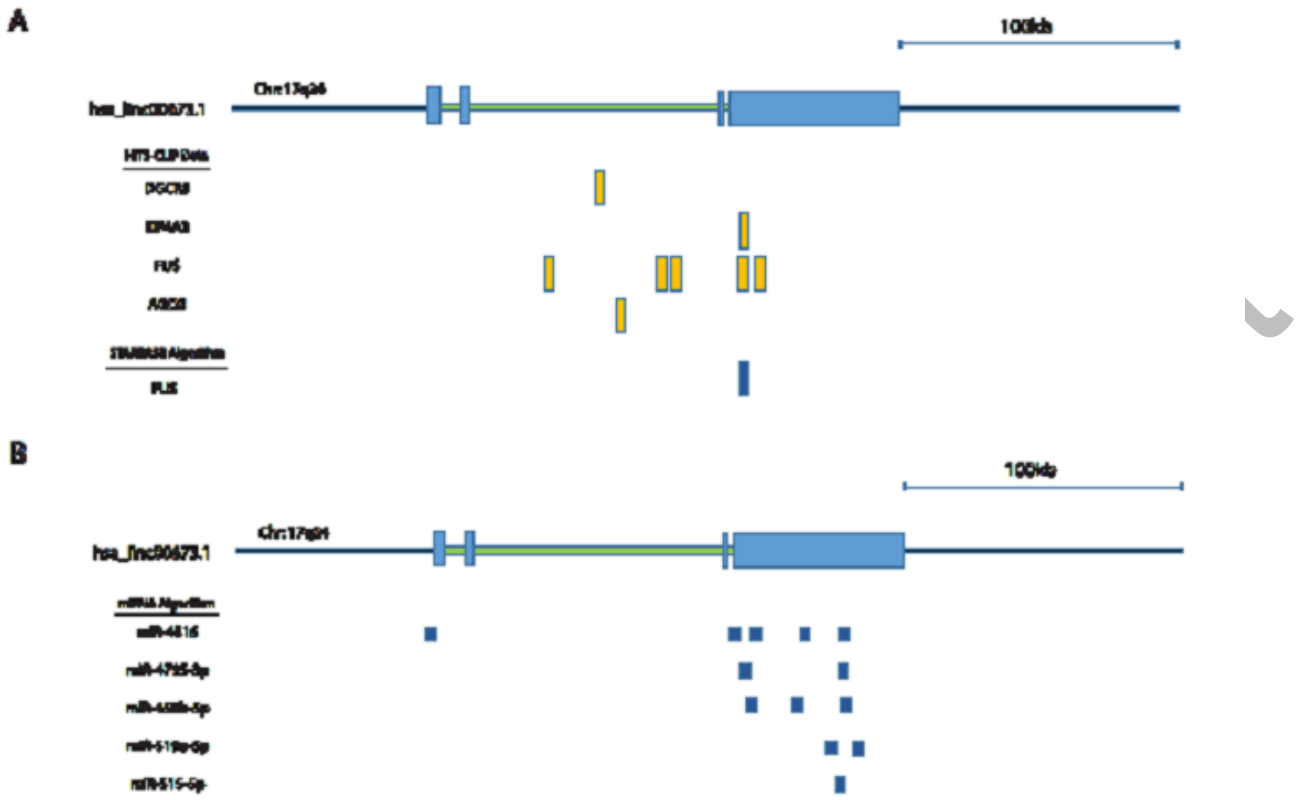

Figure 4: (A) *In silico* assessment of RNA binding proteins that interact with *ERRLR01* as determined by HITS-CLIP experiments. StarBase 2.0 was utilized to confirm FUS binding to the ERRLR01 locus. The *ERRLR01* transcript is depicted in the 3' to 5' orientation, as it is transcribed in the anti-sense orientation (minus strand of the DNA).

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

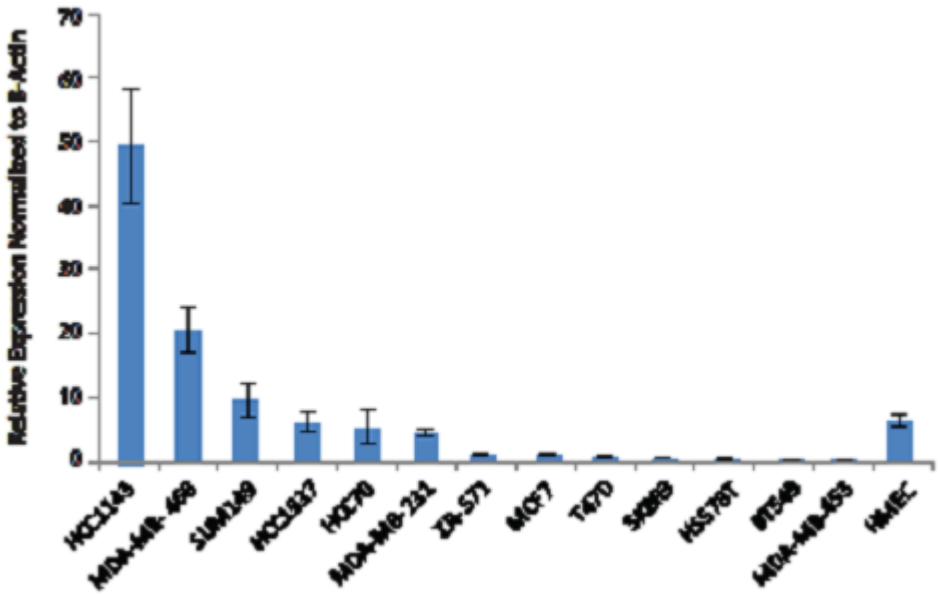

Figure 5: Quantitative PCR analysis of *ERRLR01* expression across a panel of breast cancer cell lines. Analysis indicates a few TNBC cell lines express high levels of *ERRLR01*, as compared to normal HMEC lines, as well as ERα+ cell lines.

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted Manuscript

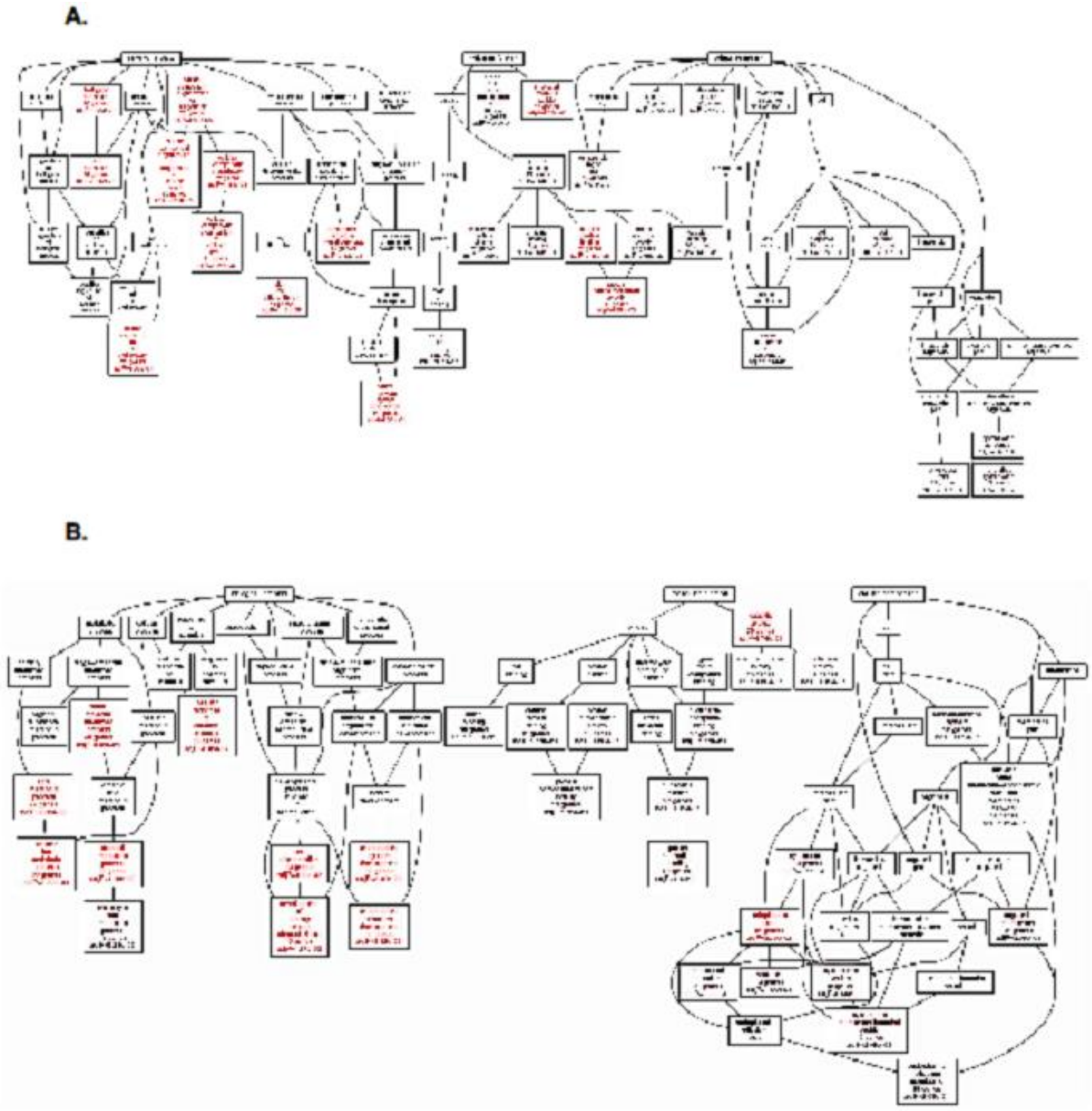

Figure 6: Go Term analysis of the top 200 positively correlated *ERRLR01* genes via spearman rank analysis of the Affymetrix dataset (HGU133). (**Panel A**) Analysis was performed in Basal-like breast tumors(n = 577), given *ERRLR01* expression is high / detectable within those samples. Analysis indicated *ERRLR01* levels correlated with “Central Nervous System Development”, “Stem Cell Differentiation”, “Protein Dimerization Activity”, and “Positive Regulation of Cell Proliferation”. (**Panel B**) Go Term Analysis of the top 200 negatively correlated *ERRLR01* genes via spearman rank analysis. Red Boxes highlight significant pathways and Black Boxes are highlighted as non-significant pathways.

Downloaded by [Australian Catholic University] at 05:46 12 August 2017

Accepted