Gene 658 (2018) 28–35

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Analysis of the miRNA–mRNA–lncRNA network in human estrogen receptor-positive and estrogen receptor-negative breast cancer based on TCGA data

Bin Xiao[a,1], Weiyun Zhang[a,1], Lidan Chen[a], Jianfeng Hang[a], Lizhi Wang[a], Rong Zhang[a], Yang Liao[a], Jianyun Chen[a], Qiang Ma[b,c,d], Zhaohui Sun[a,*], Linhai Li[a,*]

[a] Department of Laboratory Medicine, Guangzhou General Hospital of PLA, Guangzhou, Guangdong 510010, China
[b] Institute of Antibody Engineering, School of Biotechnology, Southern Medical University, Guangzhou, China
[c] State Key Laboratory of Organ Failure, Institute of Antibody Engineering, School of Biotechnology, Southern Medical University, Guangzhou, China
[d] Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University, China

ARTICLE INFO

Keywords:
Breast cancer
ER
miRNA–lncRNA–mRNA network
Prognosis

ABSTRACT

Estrogen receptor-positive (ER^+) and ER-negative (ER^-) subtypes of breast cancer have distinct clinical outcomes because they respond differentially to endocrine therapies. We aimed to comprehensively analyze differentially expressed microRNA (miRNAs), long non-coding RNAs (lncRNAs) and mRNAs in different ER subtypes as well as to identify prognosis-related RNAs. The expression levels of miRNAs, lncRNAs, and mRNAs between breast cancer and normal samples were compared using data from The Cancer Genome Atlas database. Differentially expressed miRNAs, lncRNAs and mRNAs between ER^+ and ER^- samples were also screened. An ER subtype-related miRNA–lncRNA–mRNA network was constructed. lncRNAs and mRNAs in this network were further subjected to an analysis of their associations with patient prognosis. Sets of differentially expressed miRNAs, lncRNAs, and miRNAs between breast cancer and normal samples were identified among which 14 miRNAs, 78 lncRNAs, and 475 mRNAs were differentially expressed between ER subtypes. Relationships between these RNAs were analyzed. The resultant ER subtype-related miRNA–lncRNA–mRNA network consisted of 14 nodes, among which LINC0092 and chromosome 2 open reading frame 71 (C2orf71) were correlated with better prognosis of breast cancer. LINC0092 was co-expressed with SFRP1 and RGMA and regulated by hsa-miR-449a and hsa-miR-452–5p. C2orf71 was co-expressed with LINC00511 and regulated by hsa-miR-184. Cross-talk among differentially expressed miRNAs, lncRNAs, and miRNAs may be an important feature in ER^+ and ER^- subtypes of breast cancer. LINC0092 and C2orf71, two of these cross-talking RNAs, may serve as novel prognostic predictor of breast cancer because of their close associations with prognosis.

1. Introduction

Breast cancer is one of the most commonly diagnosed cancers and one of the leading causes of cancer death in women (Siegel et al., 2016). The estrogen receptor (ER) plays an important role in the development and progression of most breast cancers (Rochefort et al., 2003; Hayes and Lewis-Wambi, 2015). Upon stimulation with estrogen, the ER signaling pathway promotes cell proliferation by activating the transcription of related genes (Hall et al., 2001). Endocrine therapies targeting the ER signaling pathway are standard and effective treatments for ER-positive (ER^+) breast cancer (Rochefort et al., 2003; Hayes and Lewis-Wambi, 2015). However, ER-negative (ER^-) breast cancer has poor response to endocrine therapies, and it generally shows a worse prognosis than its ER^+ equivalent (Rochefort et al., 2003; Hayes and Lewis-Wambi, 2015). In addition to the ER status, variations in the expression of a wide range of other molecules, including protein-

Abbreviations: ER^+, estrogen receptor-positive; ER^-, ER-negative; miRNAs, microRNA; lncRNAs, long non-coding RNAs; C2orf71, chromosome 2 open reading frame 71; ER, estrogen receptor; ncRNA, non-coding RNA; BPs, biological processes; BCAR4, breast cancer anti-estrogen resistance protein 4; HOTAIR, HOX transcript antisense RNA; GAS5, growth arrest specific 5; TCGA, The Cancer Genome Atlas; RNAseqV2, RNA-Seq Version 2; RNA-Seq, RNA sequencing; P.Value, p value; adj.P.Value, adjusted p value; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; OS, overall survival; DFS, disease-free survival; AURKA, aurora kinase A; MYBL2, MYB proto-oncogene like 2; PI3, peptidase inhibitor 3; SFRP1, secreted frizzled-related protein 1

* Corresponding authors.
E-mail addresses: zhaohui3@126.com (Z. Sun), mature303@126.com (L. Li).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.gene.2018.03.011
Received 22 June 2017; Received in revised form 9 November 2017; Accepted 3 March 2018
Available online 05 March 2018
0378-1119/

coding RNA and non-coding RNA (ncRNA), may be involved in the development and progression of breast cancers associated with different ER statuses. Among ncRNAs, microRNA (miRNA) and long non-coding RNA (lncRNA) have drawn particular attention as increasing evidence has shown their involvement in tumorigenesis (Garzon et al., 2009; Morris and Mattick, 2014; Hayes and Lewis-Wambi, 2015; Huarte, 2015).

miRNAs are small regulatory RNAs of approximately 22 nucleotides, which interact with target mRNAs to regulate the expression of the gene by inhibiting translation or by degrading target mRNAs (Garzon et al., 2009; Morris and Mattick, 2014). Approximately 30% of human genes are regulated by miRNAs (Garzon et al., 2009), and it is suggested that dysregulation of miRNAs promotes breast cancer metastasis (Hayes and Lewis-Wambi, 2015). For example, miRNAs of the well-known let-7 family are tumor suppressors targeting oncogenic mRNAs, including mRNAs of the Ras family, high mobility group A protein 2 (HMGA2), and c-myc; the expression of let-7 miRNAs are downregulated in breast cancer (Garzon et al., 2009; Hayes and Lewis-Wambi, 2015). The expression of another miRNA, miR-21, is upregulated in breast cancer and it functions as oncogene-targeting tumor suppressors, such as phosphatase and tensin homolog on chromosome 10 (PTEN), programmed cell death protein 4 (PDCD4), and tropomyosin 1 (TPM1) (Garzon et al., 2009; Hayes and Lewis-Wambi, 2015).

In contrast, lncRNAs are > 200 nucleotides in length and function differently from miRNAs (Morris and Mattick, 2014). In addition to mediating the inhibition of translation and degrading target mRNAs, they have also been shown to interact with miRNAs, thereby blocking the effects of miRNAs on mRNAs (Wang and Chang, 2011; Guttman and Rinn, 2012). lncRNAs may also interact with DNA and proteins, mediating transcriptional regulation and chromatin modification (Wang and Chang, 2011; Guttman and Rinn, 2012). lncRNAs are engaged in various biological processes (BPs), including cell cycle, apoptosis, and differentiation (Hu et al., 2012). Similar to miRNAs, an increasing number of lncRNAs play a role in breast cancer, including breast cancer anti-estrogen resistance protein 4 (BCAR4), HOX transcript antisense RNA (HOTAIR),and growth arrest specific 5 (GAS5) (Wang and Chang, 2011; Hayes and Lewis-Wambi, 2015; Huarte, 2015). Both BCAR4 and HOTAIR are oncogenes that promote breast cancer metastasis and could be used as predictors of the prognosis of breast cancer, whereas GAS5 is a tumor suppressor that induces cell cycle arrest and apoptosis (Wang and Chang, 2011; Hayes and Lewis-Wambi, 2015; Huarte, 2015). BCAR4 is also known for its relationship with anti-estrogen resistance (Hayes and Lewis-Wambi, 2015; Huarte, 2015).

Notably, a recent study revealed the differences in the expression levels of miRNAs, lncRNAs, and mRNAs between MCF-7 (ER^+) and MDA-MB-231 (ER^-) breast cancer cell lines (Wu et al., 2015). An lncRNA–mRNA co-expression network and miRNA–mRNA interaction network were also established and integrated to identify miRNAs, lncRNAs, and mRNAs associated with ER expression. Two mRNAs, CD74 and formin-like 2; an miRNA miR-19a; and an lncRNA, deleted in lymphocytic leukemia 1, correlated with ER expression (Wu et al., 2015). The important roles that miRNAs and lncRNAs play in breast cancer suggest the importance of examining ER subtype-related miRNAs, lncRNAs, and mRNAs as well as the regulatory networks of these RNAs.

Against this background, in our study, we attempted to reveal ER subtype-related miRNAs, lncRNAs, and mRNAs in human breast cancer using data from The Cancer Genome Atlas (TCGA). A more thorough analysis of ER subtype-related RNA–RNA relationships, including lncRNA–mRNA co-expression, miRNA–lncRNA interactions, and miRNA–mRNA interactions, was performed. Based on RNA–RNA relationships, we constructed a more systematic miRNA–lncRNA–mRNA network and also identified prognosis-related RNAs in this network.

2. Methods

2.1. Data source

RNA-Seq Version 2 (RNAseqV2) exon sequencing data, miRNA sequencing (miRNA-Seq) data, and clinical data of breast cancer were downloaded from the TCGA database (https://tcga-data.nci.nih.gov/) in April 2017. A total of 929 samples (842 breast cancer samples and 87 normal samples) were included in miRNA-Seq data and 1205 samples (1093 breast cancer samples and 112 normal samples) were included in RNA sequencing (RNA-Seq) data. All data were acquired using the Illumina HiSeq 2000 RNA-Seq platform.

2.2. lncRNA re-annotation

Information about the chromosome site, row count, and reads per kilo base per million reads was included in RNAseqV2 exon data. Annotations of chromosome sites of lncRNA and protein-coding RNA (V25) in the Genecode database (https://www.gencodegenes.org/) (Harrow et al., 2012) were used to re-annotate RNAseqV2 exon data. If the start site of an exon was located in an lncRNA from the Genecode database and its sequence was also complementary to the lncRNA, the exon would be defined as lncRNA. Similarly, an exon would be defined as mRNA if its start site was located in an mRNA from the Genecode database, with its sequence also complementary to the mRNA.

2.3. Screening of differentially expressed miRNAs, lncRNAs, and mRNAs

A total of 807 ER^+ breast cancer samples, including 123 HER2-positive and 436 HER2-negative samples, and 238 ER^- breast cancer samples, including 40 HER2-positive and 126 HER2-negative samples were available on TCGA. Genes with a low expression level (expression value of zero, accounting for > 80% of the genes) were removed from TCGA sequence data. Raw counts of lncRNAs, mRNAs, and miRNAs were transformed to log-CPM values and used for linear modeling using the edge R package under R (version 3.4, http://www.bioconductor.org/packages/release/bioc/html/edgeR.html) (Robinson et al., 2010; McCarthy et al., 2012). The mean-variance relationship was adjusted using precision weights calculated by the voom function.

Differences in the expression levels of lncRNAs, mRNAs, and miRNAs in breast cancer samples compared with those in normal samples were analyzed using the Bayesian *t*-test in the limma package (version3.10.3, http://www.bioconductor.org/packages/2.9/bioc/html/limma.html) (Smyth, 2005). The resulting p value (P.Value) of each gene was adjusted by multiple testing using the Benjamini & Hochberg method (Benjamini and H.Y., 1995). mRNAs with adjusted p value (adj.P.Value) < 0.05 and $|\log_2 FC \text{ (fold change)}| > 2.0$ were considered as differentially expressed mRNAs. lncRNAs and miRNAs with adj.P.Value < 0.05 and $|\log_2 FC| > 1.5$ were considered as differentially expressed lncRNAs and miRNAs, respectively. The differentially expressed lncRNAs, mRNAs, and miRNAs were used for sample hierarchical cluster analysis using the hclust function under R.

Samples with information about the ER status were used for further analysis of differences in the expression between ER^+ and ER^- subtypes using the same procedure as described above. RNAs with adj.P.Value < 0.05 and $|\log_2 FC| > 1.0$ were considered to be ER subtype-related lncRNAs, mRNAs, or miRNAs.

2.4. ER subtype-related functional and pathway enrichment analyses

Gene Ontology (GO, http://www.geneontology.org/) analysis (Ashburner et al., 2000; Gene Ontology, 2015) and Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.kegg.jp/kegg/pathway.html) analysis (Kanehisa and Goto, 2000; Kanehisa et al., 2016) are tools that are widely used for the functional annotation of genes and proteins. GO and KEGG pathway analyses of subtype-related