Abstract:

Now we use long,non coding RNAs(lncRNAs)which are in the next generation sequencing technologies in breast cancer research . lncRNA are a type of RNA, defined as being transcripts with lengths exceeding 200 nucleotides that are not translated into protein .Their important roles in the regulation of cancer-related pathways in addition to deregulation of their expression in a number of cancers have suggested that they can be used as markers for cancer detection and prognosis, as well as targets for cancer treatment. They get a subset of 18 cell line from previously publish RNA-seq dataset of 675 cancer cell and analyses it by detailed comparison of differentially expressed lncRNAs in each breast cancer sub-type with normal-like breast epithelial cells. They use Gene Expression Profiling Interactive Analysis (GEPIA2) and data from The Cancer Genome Atlas (TCGA) and The Genotype-Tissue .Expression (GTEx) project to identify lnsRNAs with invasive breast cancer.

Introduction:

Breast cancer is the most common cancer diagnosed in women. Because of RNA sequencing (RNA-seq) ,We discovered that a lot of cells produce RNA that don't produce protien. Non coding RNAs have important roles like potential disease modifiers and could be exploited as biomarkers and/or therapeutic targets . There are some types of non coding RNAs. Abundant and functionally important types of non-coding RNAs include transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), as well as small RNAs such as microRNAs, siRNAs, piRNAs, snoRNAs, snRNAs, exRNAs, scaRNAs and the long ncRNAs such as Xist and HOTAIR . We will focus on lncRNA and how lncRNAs are altered contribute to cancer, along with discovering their normal physiological roles. To understand which lncRNAs are specifically be linked to breast cancer, it is important to examine their expression profiles in various cell lines .The classification of invasive breast lesions into molecular subtypes based on the presence or absence of receptors for hormones,oestrogen (ER) and progesterone (PR) along with human epidermal growth factor-2 (HER2/ERBB2).These difference have the basis of the molecular classification of breast cancer into four major groups:(1) luminal A (2) luminal B, HER2 enriched and basal-like (3) Luminal A involves cancer cells that are ER and/or PR positive, HER2-negative and low levels of the cell cycle-regulated protein Ki-67 (4) Luminal B cancers exhibit lower ER/PR expression, with variable HER2 levels and high levels of protein Ki-67

There are also preinvasive forms of breast cancer - ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS) – distinguished by their sites of origin within the ducts or the lobules of the breast

Breast cancer cell lines are useful in knowing the details of biological processes involved with cancer initiation and progression and also split to types as breast cancer tumours (1) basal A cluster (2)basal B cluster and they don't appear in in primary tumours. To determine lncRNAs, We use classification of breast cancer cell lines

By usingGene Expression Profiling Interactive Analysis (GEPIA2) and data from The Cancer Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTEx) project ,they determined .novel, uncharacterised lncRNAs, LOC101448202, LOC105372471 and LOC105372815

Related work:

We are used to using genomic alterations, hormone receptor status and changes in cancer-related proteins to provide new avenues for targeted therapies on breast cancer research . Now we use lncRNAs in our research.

Klijn et al (2015):We use RNA-seq data and Data was retrieved from the EMBL-European Genome-Phenome Archive (EGA) servers under EGAD00001000725. Breast cancer cell line RNA-seq data files were identified using the metadata file provided EGA. Klijn paper described RNA-seq and single nucleotide polymorphism (SNP) array analysis of 675 human cancer cell lines so we knew that we had at least three to four representative lines from each group, i.e. luminal A, luminal B, HER2 positive, basal A and basal B.

Nucleic Acids
Res.,47(2019): Using GEPIA2 To examine the clinical significance of identified lncRNAs, we used GEPIA2 to explore data from TCGA and GTEx databases.

Nat. Commun., 7(2016): Our results are in agreement with previous work describing the oncogenic role of this RNA

J. Breast Canc., 20(2017): ZNF667-AS1 is supposed to be a tumour suppressor role of this lncRNA in breast cancer.

Methodology &Results:

First step we used RNA-seq data which retrieved from the EMBL-European Genome-Phenome Archive (EGA) servers . Second step we Select of breast cancer cell lines .We identify breast cancer cell line RNA-seq data files by using the metadata file provided EGA.We select 18 lines luminal A (BT-483, CAMA-1, KPL-1, MCF-7), luminal B (MDA-MB-330, UACC-812, ZR-75-30), HER2 enriched (MDA-MB-453, SK-BR-3, UACC-893), basal-like type A (BT-20, MDA-MB-436, MFM-223), basal-like type B (CAL-120, MDA-MB-157, MDA-MB-231).Third step we identify lncRNAs in RNA-seq datasets. We choose the forward reads in a zipped Fastq format then decrypted and unzipped into Fastq format.We downloud and decryption of the RNA-seq data by using Java shell provided by the EGA.FastQC was used to recheck the quality of the RNA-seq data .The reference genome annotation file was used for GRCh38 with the command “-t *lnc_RNA” to select for lncRNA.We use HTSeq to perform read counts then by using Excel we compiled the counts into a data frame and imported into R Studio for statistical analyses.For statistical analysis of differential lncRNA expression between the breast cancer cell lines we use The package DESeq2. We also use other packages utilised were pheatmap and EnhancedVolcano.We chose to trim our results by eliminating non-significant results by setting an adjusted p-value of 0.01 then arranged from lowest to highest.The fourth step Expression of lncRNAs in breast tumour samples and patient survival analyses and we use Gene Expression Profiling Interactive Analysis 2 (GEPIA2).Fifth step is Cell culture.We use MCF10A, MCF7 and MDA-MB-231 cells SK-BR-3, ZR-75-30.Sixth step is RNA analysis by qRT-PCR.By using TRIzol (Thermo Fisher Scientific) Total RNA was extracted from cells. qRT-PCR was performed using the StepOnePlus™ Real-Time PCR System and StepOnePlus software (Applied Biosystems).The seventh step is Statistics and code availability.Most statistical analyses were performed in R (version 3.5.2).

The first step in results is Bioinformatic identification of lncRNA differentially expressed in malignant versus non-malignant breast cancer cell lines.Based on their molecular subtypes, we limited 675 human cancer cell lines down to 17 breast cancer cell lines, ensuring that each group contained at least three to four representative lines, such as luminal A, luminal B, HER2 positive, basal A, and basal B.This resulted in our working RNA-seq dataset from 18 cell lines.Using the multivariate data analysis approach principal component analysis, we investigated the variation of the selected cell lines (PCA). In terms of lncRNA expression, the luminal A, luminal B, and HER2 enriched cell lines clustered together. Other malignant subtypes exhibited more variation in basal B cell lines;While non-malignant and malignant cell lines showed degrees of variation, basal A did not. MCF10A, the normal-like line, and MCF10DCIS.com, the DCIS line, grouped together and exhibited limited variation.Second step is Bioinformatic identification of lncRNAs differentially expressed in breast cancer cell lines divided by hormone/receptor status.The malignant cell lines were then sorted into three groups according on their hormone/receptor sensitivity: ER/PR positive, HER2 sensitive, and ER/PR/HER2 negative.Notably, DSCAM1-AS1 was the most significant upregulated lncRNA; while LOC101927136 was the most significant downregulated lncRNA.Third step is Evaluation of breast cancer cell lines by molecular subtypes for lncRNA expression.The normal-like breast cell line, MCF10A, was chosen as the basis for comparison in a category in the design matrix. Using our DESeq2 data, we examined cell lines based on molecular groupings – DCIS, luminal A, luminal B, HER2 enriched, basal-like type A and basal-like type B.In a recent study, DSCAM-AS1 was shown to regulated the cell cycle at the G1/S transition, increasing cell proliferation.LINC00885 and MUC5B-AS1 are two lncRNAs that have previously been linked to cancer types other than breast cancer, although CELF2-AS1 has no known cancer relationship. Most intriguingly, we discovered a few uncharacterized overexpressed lncRNAs, including LOC101448202, LOC105372471, and LOC105372815.Forth step is Assessment of clinical relevance of lncRNAs in breast cancer using GEPIA2. GEPIA2 was used to investigate data from the TCGA and GTEx databases. CELF2-AS1, DSCAM-AS1, ELFN1-AS1, LINC00885, and ZNF667-AS1 were shown to have correlations with breast cancer in GEPIA2 using our curated list. Each lncRNA has a survival plot and a comparative expression plot (tumour vs. normal tissue). GEPIA2 was used to create breast cancer survival analysis plots (Kaplan-Meier) for lncRNAs (A) CELF2-AS1, (C) DSCAM-AS1, (E) ELFN1-AS1, (G) LINC00885, and (I) ZNF667-AS1. Comparative expression of the same lncRNAs (B) CELF2-AS1, (D) DSCAM-AS1, (F) ELFN1-AS1, (H) LINC00885, and (J) ZNF667-AS1 in breast cancer tumour tissues (red) against normal tissue samples (grey)generated using GEPIA2. Fifth step is Experimental validation of lncRNA expression by qRT-PCR.We used qRT-PCR to examine at lncRNA expression in breast cancer cell lines representative of each molecular subtype and the normal-like line, MCF10A, for six lncRNAs on our curated list (CCAT1, DSCAM-AS1, LINC00885, LOC105372815, MUC5B-AS1 and ZNF667-AS1).qRT-PCR was performed using cDNA synthesized from total RNA isolated from MCF10A (normal-like), MCF10DCIS.com (DCIS), MCF7 (luminal A), ZR-75-30 (luminal B), SK-BR-3 (HER2 positive), and MDA-MB-231 (basal B) cells. Relative expression of lncRNAs (A) CCAT1; (B) DSCAM-AS1; (C) LINC00885; (D) LOC105372815; (E) MUC5B-AS1 and (F) ZNF667-AS1, as compared to GAPDH, are shown, using one-way ANOVA. The qRT-PCR studies verified lncRNA expression in the tested cell lines in large part in accord with our bioinformatic analysis. CCAT1 lncRNA was shown to be extremely weakly expressed in most breast cancer cell lines examined, with the greatest expression in the normal-like line, MCF10A, similar to the results of DESeq analysis.The luminal A (MCF7), luminal B (ZR-75-30), and HER2 positive (SK-BR-3) lines had the greatest levels of DSCAM-AS1, whereas the basal-like line had essentially little expression. Although DSCAM-AS1 was one of the most important and highly expressed lncRNAs in the ER/PR/HER2 negative and basal A subtypes, this appears to be inconsistent.Interestingly, unlike the other breast cancer cell lines studied, qRT-PCR analysis of LINC00885 showed the lowest expression of this lncRNA in MCF7 cells   in the basal-like line (MDA-MB-231), which agrees with our bioinformatic study. Each lncRNA for LOC105372815 and MUC5B-AS1 was expressed at a higher level in some cell lines than in MCF10A, although there was consistent .low expression for each of them