Data in Brief

Transcriptional profiling of differentially expressed long non-coding RNAs in breast cancer

Lizhen Wang a,1, Xiaokun Shen a,1, Bojian Xie b,1, Zhaosheng Ma b, Xiaobing Chen c,⁎, Feilin Cao b,⁎

a Department of Pediatrics, Taizhou Hospital, Wenzhou Medical University, Taizhou, Zhejiang, China
b Department of Surgical Oncology, Taizhou Hospital, Wenzhou Medical University, Taizhou, Zhejiang, China
c Department of Internal Oncology, Henan Cancer Hospital, The Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, Henan, China

A R T I C L E   I N F O

Article history:
Received 17 September 2015
Accepted 27 September 2015
Available online 30 September 2015

Abstract

Long non-coding RNAs (lncRNAs) are subclass of non-coding RNAs that have been recently shown to play critical roles in cancer biology. However, little is known about their mechanistic role in breast cancer pathogenesis, especially in triple-negative breast carcinomas (TNBC) that have particular poor outcomes.

This study was specifically designed to identify the signatures relevant lncRNAs in breast cancer and characterize lncRNAs that modulate the phenotype. Here we provide detailed methods and analysis of microarray data, which is deposited in the Gene Expression Omnibus (GEO) with the accession number GSE64790. The basic analysis as contained in the manuscript published in Oncotarget with the PMID 26078338. These data can be used to further elucidate the mechanisms of breast cancer.

© 2015 Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

2. Experimental design, materials and methods

2.1. Experimental design

Paired TNBC tissues and adjacent normal breast tissues from every subject were obtained and used for long non-coding RNAs profiling.

2.2. Materials and methods

2.2.1. Patient material

All patients are informed and have declared written informed consent that their samples can be used for research. All patients received tumor resection at Zhejiang Taizhou Hospital and were diagnosed with TNBC histopathologically after surgery. Immunohistochemical staining of estrogen receptor/progesterone receptor and ErbB2 receptor in 3 samples are shown in Fig. 1. There was no radiotherapy or chemotherapy prior to surgery. 3 paired samples were used for microarray analysis of lncRNAs. Demographic and clinical characterizations of the study population are summarized in Table 1.

Tissue sections and specimen were prepared by pathologists prior to snap freezing and subsequent storage in liquid nitrogen prior to sectioning. Biobanking and handling of the tissues followed the BRISQ guidelines. 20–30 μm sections/sample (depending on tissue size) were made in a cryotome and used for RNA extraction. All samples had at least 80% tumor cell content.

2.2.2. RNA isolation

Total RNA, including small RNAs, was isolated using the mirVanaTM RNA Isolation Kit (Ambion, Foster City, CA, United States) according to the manufacturer’s instructions. RNA concentration was measured using the Nano Drop ND-2000 (Thermo Scientific), and RNA quality was evaluated on Agilent Bioanalyzer 2100 (Agilent Technologies). Only samples with an RNA Integrity Number (RIN) greater than or equal to 7 were used for further downstream processing.
2.2.3. In vitro transcription, labeling and hybridization

The sample labeling, microarray hybridization and washing were performed based on the manufacturer’s standard protocols. Briefly, mRNA was purified from total RNA after removal of rRNA by using an mRNA-ONLY Eukaryotic mRNA Isolation Kit (Epicenter Biotechnologies, USA). Then, each sample was transcribed to double strand cDNA, then synthesized into cRNA and labeled with Cyanine-3-CTP. The labeled cRNAs were hybridized onto the Human lncRNA array V4.0 (4 × 180 K, Agilent), including the global profiling of 78,243 human lncRNAs and 30,215 coding transcripts. After washing, the arrays were scanned with the Agilent Scanner G2505C (Agilent Technologies).

2.2.4. Microarray data quality control

Since three technical replicate hybridizations were performed and later averaged, care was taken to ensure high repeatability between technical replicates. First, raw and normalized log2 data for each sample

---

Table 1

| Patient number | Patient code | Sex | Age (years) | Diagnosis     | Surgery date | ER   | PR  | HER2 |
|----------------|--------------|-----|-------------|----------------|--------------|------|-----|------|
| 1              | WMY          | F   | 50          | Breast cancer  | 2012.07.12   | –    | –   | 1-2+ |
| 2              | ZHX          | F   | 49          | Breast cancer  | 2012.07.11   | –    | –   | 2+   |
| 3              | CYF          | F   | 46          | Breast cancer  | 2012.07.11   | –    | –   | 2+   |

---

Fig. 1. Immunochemical staining of estrogen receptor/progesterone receptor and ErbB2 receptor of 3 TNBC samples for microarray analysis.

Fig. 2. Representative box plot of raw data from three technical replicate hybridizations of a single sample. For all samples, the box plots revealed median-centered raw data distributions, which were further refined during normalization. Overall, this points to high repeatability of technical replicate hybridizations.
were plotted using the R function boxplot. Control and flagged probes were not included. A representative box plot is shown in Fig. 2. While this analysis is designed to identify hybridizations that have intensity distributions different from those of their technical replicates, we did not find any instances of this. This analysis also ensures that the normalization has correctly centered the distributions of each replicate microarray.

Next, we compared scatter plots of raw and normalized log2 data for each sample using the R function pairs. Only data with a P-value detected < 0.01 were included. A representative scatter plot is shown in Fig. 3. Scatter plots were viewed in conjunction with Pearson correlation tables. Correlation values were calculated from both raw and normalized log2 intensities for each technical repeat. Only probes with P-value detected < 0.01 were included in the calculation. Scatter plots confirmed high repeatability among technical replicates.

2.2.5. Data normalization

Feature Extraction software (version 10.7.1.1, Agilent Technologies) was used to analyze array images and extract the raw data. Genespring (Version 12.5, Agilent Technologies) was employed to finish the basic analysis of the raw data. To begin with, the raw data were normalized with the quantile algorithm. The probes that had at least 1 condition out of 2 conditions flagged as “P” were chosen for further data analysis. Differentially expressed IncRNAs and mRNAs were then identified through fold-change as well as P values calculated with t-test. The threshold set for up- and down-regulated genes was fold change ≥2.0 and P value ≤0.05. Afterwards, hierarchical clustering was performed to display the distinguishable IncRNAs and mRNAs expression patterns among the samples.

2.2.6. Statistical analysis

The Statistical Program for Social Sciences (SPSS) 18.0 software (SPSS, Chicago, IL, United States) was employed to perform all the statistical analyses. All data were expressed as the mean ± SD or proportions where appropriate. For comparisons, paired t-tests and unpaired t-tests were performed where appropriate. P values of 0.05 (two-tailed) were considered statistically significant.

3. Discussion

TNBC accounts for approximately 10–25% of all breast cancers and is of particular clinical interest due to its tendency to affect younger women and refractory to currently available targeted therapy. The molecular mechanisms for aggressive clinical behavior of TNBC are not fully understood [1–3]. We have described here a very unique data set of patients with triple-negative breast carcinomas. This data set is further supported by additional clinical data that could be exploited in deciphering clinical associations; however, the small sample size limits this applicability. This data set has been used as a source of hypothesis generation to discern and investigate IncRNAs that impact breast carcinomas. The outcome of such analyses has led to the generation of our recent manuscript.

Conflict of interest

The authors declare no financial conflicts of interest.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (81202103, 31370810, 81172240, 81472714, and 81171914) and Henan outstanding youth fund projects (2014011016).

References

[1] S.K. Pal, B.H. Childs, M. Pegram, Triple negative breast cancer: unmet medical needs. Breast Cancer Res. Treat. 125 (3) (2011) 627–636, http://dx.doi.org/10.1007/s10549-010-1293-1.
[2] T. Hung, H.Y. Chang, Long noncoding RNA in genome regulation: prospects and mechanisms. RNA Biol. 7 (5) (2010) 582–585.
[3] F. Di Gesualdo, S. Capaccioli, M. Lulli, A pathophysiological view of the long non-coding RNA world. Oncotarget 5 (22) (2014) 10976–10996.