Long non-coding RNA (lncRNA) transcripational landscape in breast cancer identifies LINC01614 as non-favorable prognostic biomarker regulated by TGFβ and focal adhesion kinase (FAK) signaling

Radhakrishnan Vishnubalaji1, Hibah Shaath1, Eyad Elkord1 and Nehad M. Alajez1

Abstract
Long non-coding RNAs (lncRNAs) represent a class of epigenetic regulators implicated in a number of physiological and pathological conditions. Herein, we characterized the lncRNA expression portrait from 837 patients with invasive breast cancer and 105 normals from the cancer genome atlas (TCGA), which revealed eighteen upregulated and forty-six downregulated lncRNAs. Clustering analysis revealed distinct lncRNA profile for the triple negative breast cancer (TNBC) and normal breast tissue, while less separation was observed among the HER2+HR−, HER2−HR+, HER2−HR− molecular subtypes. LINC01614, and LINC01235 correlated with worse disease-free survival (DFS), while the expression of Inc-LRR1–1, Inc-ODF3B–2, AC015712.5, Inc-LAMB3–1, Inc-SPP2–3, and Inc-MAP9–2 correlated with better DFS. The expression of LINC01235 correlated with worse overall survival (OS), while the expression of MIR205HG, Inc-MAP2K6–5, FGF14-AS2, Inc-SPP2–3 correlated with better OS. Highest expression of LINC01614 was observed in progesterone receptor (PR)+, Estrogen receptor (PR)+, and HER2+ tumors, while lowest expression was in TNBC. Concordantly, LINC01614 was highly expressed in the luminalB/HER2+ subtype from the SRP062132 dataset. Elevated expression of LINC01614 was subsequently validated in primary breast cancer tissue and breast cancer cell lines. Bioinformatics and pathway analyses on LINC01614high vs. LINC01614low BC tissue revealed TGFBI and ECM as the most activated networks in LINC01614high tumors. Concordantly, strong correlation between the expression of LINC01614 and COL10A1 ($R^2 = 0.6929$), SPOCK1 ($R^2 = 0.5156$), ZEB1 ($R^2 = 0.3372$), TGFBI ($R^2 = 0.2978$), TGFB1 ($R^2 = 0.1985$), ACTA2 ($R^2 = 0.1833$), and TAGLN ($R^2 = 0.1909$) was observed. Mechanistically, exogenous TGFBI induced LINC01614 expression in the BT474 triple positive BC model, while small-molecule inhibition of transforming growth factor β (TGFβ, SB-431542) or focal adhesion kinase (FAK, PF-573228) abrogated LINC01614 expression. Our data revealed the lncRNA transcription landscape in breast cancer and its molecular subtypes. Our data provide novel insight implicating LINC01614 as unfavorable prognostic marker in BC, its association with the HR+/HER2+ BC molecular subtype and its regulation by TGFβ and FAK signaling.

Introduction
Breast cancer (BC) is the most common cancer type in females worldwide1. The molecular mechanisms involved in BC pathogenesis have been thoroughly studied, leading to BC classification into three major subtypes: Luminal which is positive for estrogen (ER+) and progesterone...
| Ensembl Gene ID   | LNCipedia gene ID | log fold Tum vs Nor | fold Tum vs Nor | Tum vs Nor p value (raw) | Tum vs Nor p value (adj) | Status |
|-------------------|-------------------|---------------------|----------------|--------------------------|--------------------------|--------|
| ENSG00000230838.1 | LINCO1614         | 2.341821475         | 5.069422738    | 5.65E-81                 | 3.76E-79                 | Up     |
| ENSG00000230499.6 | Inc-MAPK15–6      | 1.767786642         | 3.405315908    | 3.82E-82                 | 2.61E-80                 | Up     |
| ENSG00000223808.1 | AC044784.1, Inc-GATA3–7 | 1.610083258     | 3.052694584    | 1.25E-16                 | 1.13E-15                 | Up     |
| ENSG00000259187.1 | Inc-TRIM69–1      | 1.590879481         | 3.01239282     | 1.81E-44                 | 4.73E-43                 | Up     |
| ENSG00000261039.1 | LINC02544         | 1.398432474         | 2.636150018    | 7.47E-43                 | 1.85E-41                 | Up     |
| ENSG00000223638.1 | Inc-TAF3–1        | 1.21241703          | 2.317255348    | 9.35E-12                 | 1.84E-12                 | Up     |
| ENSG00000272993.1 | Inc-HIST2H2AA4–1  | 1.185802408         | 2.274988663    | 1.78E-16                 | 3.74E-16                 | Up     |
| ENSG00000243350.1 | GATA3–AS1         | 1.148653771         | 2.217069154    | 3.27E-13                 | 2.39E-12                 | Up     |
| ENSG00000232627.2 | C4A–AS1           | 1.105244735         | 2.1513537      | 7.23E-16                 | 6.20E-15                 | Up     |
| ENSG00000272766.1 | Inc-ODF3B–2       | 1.077677457         | 2.11065509     | 6.08E-19                 | 1.98E-18                 | Up     |
| ENSG00000272661.1 | Inc-KLHC7B–2      | 1.048411607         | 2.06825147     | 4.74E-19                 | 4.87E-18                 | Up     |
| ENSG00000225373.2 | Inc-TINCR–1       | 1.044336743         | 2.062417976    | 1.83E-17                 | 1.73E-16                 | Up     |
| ENSG00000262176.1 | Inc-HIST2H2AA3–3 | 1.026047258         | 2.036437106    | 3.92E-35                 | 7.61E-34                 | Up     |
| ENSG00000251141.1 | MRPS50–DT         | 1.10254162          | 2.02860429     | 1.58E-07                 | 7.36E-07                 | Up     |
| ENSG00000268919.3 | Inc-KCNK6–1       | 1.001259226         | 2.00796371     | 8.83E-23                 | 1.07E-21                 | Up     |
| ENSG00000228223.1 | HCG11             | −1.017371351        | −2.02427373    | 2.51E-53                 | 8.53E-52                 | Down   |
| ENSG00000235387.1 | na                | −1.020539284        | −2.028677144   | 1.96E-59                 | 7.71E-58                 | Down   |
| ENSG00000228162.1 | Inc-SPP2–3        | −1.043000128        | −2.060508089   | 9.21E-115                | 1.41E-112                | Down   |
| ENSG00000272639.1 | AC015712.5        | −1.05048961         | −2.07132647    | 1.95E-49                 | 5.91E-48                 | Down   |
| ENSG00000259616.1 | Inc-BIRC2–4       | −1.069704648        | −2.09900361    | 2.70E-26                 | 3.84E-25                 | Down   |
| ENSG00000264868.1 | Inc-STEAP4–1      | −1.080167322        | −2.114281279   | 8.30E-28                 | 1.25E-26                 | Down   |
| ENSG00000250538.1 | Inc-MAP9–2        | −1.096920898        | −2.138976893   | 1.05E-104                | 1.21E-102                | Down   |
| ENSG00000243836.1 | WDR86–AS1         | −1.151186013        | −2.220955695   | 1.05E-116                | 1.73E-114                | Down   |
| ENSG00000178947.8 | SMIM10L2A         | −1.161204888        | −2.23641294    | 2.98E-104                | 3.39E-102                | Down   |
| ENSG00000267194.1 | Inc-MAP2K6–5      | −1.16295703         | −2.239158576   | 5.17E-82                 | 3.52E-80                 | Down   |
| ENSG00000233429.5 | HOTAIRM1          | −1.179118648        | −2.264384021   | 1.06E-87                 | 8.10E-86                 | Down   |
| ENSG00000271738.1 | Inc-TSPAN14–1     | −1.191983318        | −2.28466075    | 1.60E-57                 | 6.01E-56                 | Down   |
| ENSG00000214548.10| Inc-DLK1–15       | −1.199354187        | −2.29638526    | 4.55E-110                | 6.30E-108                | Down   |
| ENSG00000268164.1 | na                | −1.20443904         | −2.30449867    | 2.30E-93                 | 2.01E-91                 | Down   |
| ENSG00000260025.1 | Inc-FEZ2–7        | −1.20586992         | −2.30673212    | 5.20E-63                 | 2.24E-61                 | Down   |
| ENSG00000245812.2 | LINC02202         | −1.227351365        | −2.341367447   | 7.20E-190                | 9.16E-187                | Down   |
| ENSG00000262179.2 | MYMX              | −1.233789324        | −2.351839037   | 8.67E-99                 | 8.49E-97                 | Down   |
| ENSG00000186594.8 | MIR22HG           | −1.237175737        | −2.357365953   | 4.69E-86                 | 3.43E-84                 | Down   |
| ENSG00000270547.1 | LINC01235         | −1.239824501        | −2.361698012   | 6.31E-42                 | 1.50E-40                 | Down   |
| ENSG00000255248.2 | MIR100HG          | −1.261015808        | −2.396644303   | 3.69E-100                | 3.82E-98                 | Down   |
receptors (PR+), human epidermal growth factor receptor 2 (HER2+), and basal-like tumors, which lacks hormone receptor and HER2 expression, called triple-negative breast cancer (TNBC)\textsuperscript{2–4}. While this classification has led to better stratification of BC, this disease is still associated with high mortality rate underscoring a need to develop novel molecular signature for better stratification and for prediction of disease outcome.

Although a number of gene-expression signatures have been reported as diagnostic and prognostic biomarkers in BC, the role of non-coding RNAs in this domain is just beginning to unfold. Genome wide transcriptome studies have revealed the existence of large number of long non-coding RNAs (≥200 nucleotides)\textsuperscript{5,6}. Current GENCODE database (version 30) revealed the existence of approximately 16,193 lncRNAs and 14,706 pseudogenes in the human genome.

LncRNAs are involved in regulating various biological processes, including tumor-suppressor and oncogenic pathways and may serve as prognostic markers in BC. A number of oncogenic (H19, SRA, LSINCT5, Zfas1, lncRNA-Smad7, LOC554202, HOTAIr, SOX2OT and FAL1) and tumor suppressor (GAS5 and XIST) lncRNAs have been identified in BC; however their regulation and the mechanisms of action for the majority of lncRNAs remains to be unraveled\textsuperscript{6,7}.

In this study, we characterized the lncRNA expression portrait from 837 patients with invasive BC and 105 normal breast tissues from the cancer genome atlas (TCGA) breast cancer dataset.

### Table 1 continued

| Ensembl Gene ID | LNCipedia gene ID | log fold Tum vs Nor | fold Tum vs Nor | Tum vs Nor p value (raw) | Tum vs Nor p value (adj) | Status |
|-----------------|-------------------|---------------------|---------------|-------------------------|-------------------------|--------|
| ENSG00000272143.1 | FGF14-AS2         | -1.285565793        | -2.437776382  | 8.61E-100               | 8.62E-98                | Down   |
| ENSG00000236333.3 | TRHDE-AS1         | -1.286805825        | -2.439874827  | 8.55E-139               | 1.16E-136               | Down   |
| ENSG00000257877.1 | Inc-MAPKAPK5-1    | -1.291454567        | -2.447748728  | 4.55E-147               | 1.25E-144               | Down   |
| ENSG00000267532.2 | Inc-SLC16A11-7    | -1.32087847         | -2.517667645  | 7.47E-147               | 2.50E-144               | Down   |
| ENSG00000234456.3 | MAGI2-AS3         | -1.339265343        | -2.530224405  | 1.04E-159               | 4.58E-157               | Down   |
| ENSG00000272327.1 | Inc-NRG1-3        | -1.352076922        | -2.552793645  | 8.52E-68                | 4.13E-66                | Down   |
| ENSG00000258545.1 | RHOF1-AS1         | -1.35631857         | -2.560866164  | 6.26E-114               | 9.16E-112               | Down   |
| ENSG00000267653.1 | Inc-ABC5-6        | -1.372853753        | -2.589823452  | 1.55E-108               | 2.04E-106               | Down   |
| ENSG00000229108.1 | LINC0287          | -1.395635686        | -2.631044566  | 6.69E-164               | 3.41E-161               | Down   |
| ENSG00000230148.4 | HOXB-AS1          | -1.42667247         | -2.68825962   | 8.69E-41                | 6.89E-41                | Down   |
| ENSG00000228971.2 | Inc-RWDD3-5       | -1.450759321        | -2.73318843   | 2.47E-108               | 3.21E-106               | Down   |
| ENSG00000249669.3 | CARMN             | -1.474915257        | -2.779673159  | 1.36E-234               | 1.73E-230               | Down   |
| ENSG00000230937.5 | MIR205HG          | -1.500142384        | -2.828706285  | 1.10E-27                | 1.65E-26                | Down   |
| ENSG00000180139.10| ACTA2-AS1         | -1.56395501         | -2.955482333  | 1.37E-118               | 2.33E-116               | Down   |
| ENSG00000231367.1 | Inc-ATL2-1        | -1.570416598        | -2.969904619  | 4.73E-83                | 3.27E-81                | Down   |
| ENSG00000258663.1 | Inc-RTL1-1        | -1.573129653        | -2.975494916  | 3.68E-96                | 3.40E-94                | Down   |
| ENSG00000255471.1 | Inc-FZD4-1        | -1.665847283        | -3.172999476  | 1.39E-151               | 5.52E-149               | Down   |
| ENSG00000267519.2 | Inc-C19orf57      | -1.684869673        | -3.21513511   | 2.96E-81                | 1.98E-79                | Down   |
| ENSG00000267047.1 | Inc-SLC16A11-7    | -1.78333068         | -3.442205117  | 1.36E-109               | 1.85E-107               | Down   |
| ENSG00000238018.1 | Inc-RTN4-3        | -1.826091508        | -3.545751691  | 7.48E-143               | 2.21E-140               | Down   |
| ENSG00000272761.1 | Inc-CCDC80-5      | -1.868752854        | -3.652167299  | 2.80E-47                | 7.92E-46                | Down   |
| ENSG00000229645.4 | Inc-SYNE3-1       | -2.025900744        | -4.072714669  | 5.48E-134               | 1.32E-131               | Down   |
| ENSG00000228639.2 | Inc-SLC39A11-10   | -2.048436659        | -4.136574768  | 6.54E-44                | 1.68E-42                | Down   |
| ENSG00000254148.3 | Inc-SLC39A11-10   | -2.181617437        | -4.536618794  | 6.46E-45                | 1.71E-43                | Down   |
| ENSG00000227591.1 | Inc-LAMB3-1       | -2.192798277        | -4.571914035  | 1.67E-169               | 1.12E-166               | Down   |
| ENSG00000269936.2 | CARMN             | -2.778459057        | -6.86191136   | 5.33E-170               | 3.77E-167               | Down   |
Our comparative analysis identified eighteen upregulated and forty-six downregulated lncRNAs in breast cancer compared to normal tissue. Interestingly, we identified eleven lncRNAs: LINC01614, LINC01235, lnc-LRR1–1, lnc-ODF3B-2, AC015712.5, lnc-LAMB3–1, lnc-SPP2–3, lnc-MAP9–2, MIR205HG, lnc-MAP2K6–5 and FGF14-AS2 to be associated with BC patient outcome. In particular, our data provided novel insight implicating LINC01614 as unfavorable prognostic marker in BC, its association with the HR+/HER2+ BC molecular subtype and its regulation by TGFβ and FAK signaling.

Fig. 1 lncRNA transcriptional portrait in breast cancer compared to normal tissue. a Hierarchical clustering of breast cancer (n = 837) and normal (n = 105) breast tissue based on differentially expressed lncRNA transcripts. Each column represents one sample and each row represents a transcript. Expression level of each transcript (log2) in a single sample is depicted according to the color scale. b Principal component analysis (PCA) for the lncRNA transcriptome of breast cancer and normal breast tissue. c Venn diagram depicting the overlap between tumors samples from tumor cluster 2 (T-cluster 2) on the further right of the heatmap in panel (a) and the indicated breast cancer molecular subtype.
Materials and methods

Data analyses and bioinformatics

Long noncoding RNA (lncRNA) expression from 837 invasive breast carcinoma and 105 normal subjects were retrieved from The Atlas of Noncoding RNAs in Cancer (TANRIC; http://bioinformatics.mdanderson.org/main/TANRIC:Overview) database. Expression data were subsequently imported into Altanalyze v.2.1.0 software as described before. Hierarchical clustering was performed using cosine for columns and cosine for rows while principal component analysis was performed to assess the relatedness of samples. Gene expression for the same cohort was retrieved from the cBioPortal for Cancer Genomics (https://www.cbioportal.org/) database as we described before.

RNA-Seq data analysis

Raw RNA sequencing data were retrieved from sequence read archive (SRA) database under accession no. SRP062132. Data were retrieved using the SRA toolkit version 2.9.2 as previously described. Pair end reads were aligned to the hg19 human reference genome in CLC Genomics Workbench-12 (QIAGEN, Germany). The abundance of the expression of transcripts was measured as the score of TPM (Transcript Per Kilobase Million) mapped reads in CLC Genomics Workbench 12.

Fig. 2 Disease-free survival (DFS) of breast cancer patients based on lncRNA expression. Kaplan-Meier DFS analysis for Inc-LRR1–1 (a), LINC01614 (b), Inc-ODF3B-2 (c), LINC01235 (d), AC015712.5 (e), Inc-LAMB3–1 (f), Inc-SPP2–3 (g), and Inc-MAP9–2 (h) in the TCGA BC cohort. Significance was calculated using the log-rank test. p values are indicated on each plot.
Expression of LINC01614 in each molecular subtype was plotted using Graphpad Prism 6.0 software (Graphpad Software, San Diego, CA, USA).

Gene set enrichment and modeling of gene interactions networks
Upregulated genes in the LINC01614high BC group were imported into the Ingenuity Pathways Analysis (IPA) software (Ingenuity Systems; www.ingenuity.com/) and were subjected to functional annotations and regulatory network analysis using upstream regulator analysis (URA), downstream effects analysis (DEA), mechanistic networks (MN) and causal network analysis (CNA) prediction algorithms. IPA uses precise algorithm to predict functional regulatory networks from gene expression data and provides a significance score for each network according to the fit of the network to the set of focus genes in the database. The p value is the negative log of P and represents the possibility that focus genes in the network being found together by chance.

Statistical and survival analysis
Kaplan-Meir survival analysis and plotting were conducted using IBM SPSS statistics version 24 software. For survival analysis, patients were grouped into high or low based on LINC01614 log2 gene expression. The log-rank test was used to compare the outcome between expression groups. Statistical analyses to compare specific gene expression and graphing were performed using Graphpad Prism 6.0 software. Unpaired two-tailed t-test and p value of <0.05 was considered significant as we described before.

Cell culture, recombinant TGFβ treatment, and small molecule inhibition
Human breast cancer cell lines (BT474, T47D, MDAMB453, ZR751, MCF7, HCC70, HS578T, MDAMB468, BT549 and MDAMB231) were cultured in Dulbecco’s modified Eagle’s medium/RPMI 1640 supplemented with D-glucose 4500 mg/l, 4 mM L-glutamine and 110 mg/l sodium pyruvate, 10% fetal bovine serum and 1x penicillin–streptomycin (Pen-Strep) (all purchased from Gibco-Invitrogen, Waltham, MA, USA). The triple positive BC cell line (BT474) was treated with rhTGFβ (10 ng/ml, Peprotec, London, UK), TGFβ inhibitor (SB-431542; 10 μM, Selleckchem Inc., Houston, TX, USA), FAK inhibitor (PF-573228; 5 μM, Selleckchem Inc., Houston, TX, USA) and combination of rhTGFβ and TGFβ inhibitor. Pharmacological inhibition of TGF-β and FAK pathways were conducted as we previously described. Briefly, 0.2 × 10⁶ cells/well were cultured in 6 well plates (duplicate) and incubated for 48 hours and subsequently the
Fig. 4 (See legend on next page.)
expression of LINC01614 was measured using qRT-PCR. Assays were carried out with appropriate DMSO control.

LncRNA validation using qRT-PCR

Tumor tissue (TT) specimens from eight BC tissue and adjacent normal tissue (NT) were obtained from treatment-naive BC patients prior to surgery with a proper written informed consent. The study was approved by Qatar Biomedical Research Institute, Doha, Qatar (Protocol no. 2017–006). Total RNA was extracted from eight primary BC tissue, adjacent normal tissue, and from a panel of breast cancer cell lines using Norgen RNA/DNA/Protein Purification Plus Kit (Norgen Biotek Corp, Ontario, Canada) as per the manufacturer’s instructions. Expression level of LINC01614 was validated using SYBR Green-based quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR). The total RNA (500 ng) was reverse transcribed into complementary DNA (cDNA) using a High Capacity cDNA Reverse Transcript Kit (catalogue No. 4368814; ABI) according to the manufacturer’s protocol. Relative levels of lncRNA was determined using the cDNA as template in real-time PCR analysis using the Applied Biosystems QuantStudio 6/7 Flex Real-time PCR system. Primer sequences used in the current study were: LINC01614 F: 5′-AACCAAGAGCGAAGGCAAAGA-3′; LINC01614 R: 5′-GCTTGGA-CACGACCCCTAGC-3′; GAPDH F: 5′-CTGGTAAAGTTGATATTGTTGCCAT-3′; and GAPDH R: 5′-TGGAATCATATTGGAACATGTAAACC-3′. The relative expression level was calculated using −ΔΔCT, GAPDH was used as an endogenous control.

Results

Expression profiling of IncRNAs from the TCGA BC dataset compared to normal breast tissue

Expression data for 12727 lncRNAs from 837 patients with invasive BC and 105 normal breast tissue were retrieved from TANRIC database and were subjected to differential expression analysis, which identified 18 upregulated and 46 downregulated lncRNAs (±22, FDR p ≤ 0.05; Table 1). Hierarchical clustering revealed three major clusters, where breast cancer samples clustered at both sides while normals clustered in the middle (Fig. 1a). Principle component analysis (PCA) also revealed clear separation of normal from breast cancer based on lncRNA expression (Fig. 1b). Interestingly, we observed majority of patients in the further right cluster (76%; Fig. 1a) to be of the TNBC molecular subtype (Fig. 1c).

Identification of IncRNAs associated with overall survival (OS) and disease-free survival (DFS) in the TCGA BC dataset

The differentially expressed lncRNAs from the profiling study were subsequently subjected to Kaplan-Meir survival analysis. Our data revealed LINC01614 and LINC01235 to predict worse DFS of BC patients (Fig. 2b and d; log-rank test = 0.07 and 0.08, respectively). On the other hand, Inc-LRR1–1, Inc-ODF3B–2, AC015712.5, Inc-LAMB3–1, Inc-SPP2–3, and Inc-MAP9–2 were associated with better DFS (Fig. 2a, c, e–h). Interestingly, the expression of LINC01235 correlated with worse DFS and worse OS (Figs. 2d and 3b), while the expression of MIR205HG, Inc-MAP2K6–5, FGFI4–AS2, Inc-SPP2–3 correlated with better OS (Fig. 3a, c–e). Taken together, our data revealed LINC01614 as the only upregulated lncRNAs in BC and associated with worse DFS in the TCGA dataset.

LINC01614 expression correlates with HER2−HR− invasive breast cancer molecular subtype

LINC01614 was the most highly expressed lncRNA (5.0 FC, p (adj) = 3.7 × 10−79) in breast cancer compared to normal tissue. We subsequently validated the expression of LINC01614 in a cohort of breast cancer patients, which revealed elevated expression of LINC01614 in BC compared to adjacent normal tissue (5.9 FC, p = 0.0007, Fig. 4a). Similarly, LINC01614 expression was detected in a panel of BC cell lines, where highest expression was observed in the BT474 triple positive BC cell line (Fig. 4b). We subsequently sought to determine if IncRNA expression can discriminate breast cancer with various molecular subtypes. To that end, the 837 BC samples were divided into HER2−HR−, HER2+HR−, HER2+HR+, and TNBC and were subjected to the marker finder algorithm in Altanalyze v.2.1.0 compared to 115 normal breast...
| IncRNAs enriched according to breast cancer molecular subtype | HER2-HR+ | HER2 > HR+ | HER2 > HR^- | Normal |
|-------------------------------------------------------------|----------|------------|-------------|--------|
| TNBC                                                       | LNCipedia gene ID | LNCipedia gene ID | LNCipedia gene ID | LNCipedia gene ID | LNCipedia gene ID | LNCipedia gene ID | LNCipedia gene ID |
| ENSG00000263680.2 Inc-STAT9-3                               | ENSG000002480082 | ENSG000002666491.1 | ENSG000002701451.1 | ENSG000002320444.1 | ENSG00000264016.1 | Inc-TMEM22-1   |
| ENSG00000227036.2 Inc-SLAC9-11                               | ENSG000002490421 | ENSG000002701451.1 | ENSG00000256940.1 | ENSG00000234566.3 | MAGI2-AS3        |
| ENSG00000225177.1 Inc-CCDC2-BA-1                             | ENSG00000266999.1 | ENSG00000261235.1 | ENSG00000241684.1 | ENSG00000263400.2 | ENSG0000026105.1 |
| ENSG00000229124.2 VIM-AS1                                    | ENSG000002673482 | ENSG000002575086.1 | Inc-RPMI14B1            | ENSG00000226917.1 |
| ENSG00000258910.2 LINCD19-5                                   | ENSG000002632141 | ENSG00000263466.1 | Inc-FBXO47-2             | ENSG00000267047.1 |
| ENSG00000246334.2 PRR7-AS1                                   | ENSG00000248360.3 | ENSG00000230883.1 | LNC01614                | ENSG00000235904.1 |
| ENSG00000227524.1 Inc-KLF11-3                                 | ENSG000002351064 | Inc-BRD3OS-1         | ENSG00000237325.1 | ENSG00000235904.1 |
| ENSG00000258938.1 Inc-RALGA1-1                               | ENSG000002630111.1 | Inc-ZNF205-S1        | ENSG00000235491.1 | ENSG00000235904.1 |
| ENSG00000224167.1 Inc-FAM19A-3-2                              | ENSG000002253613.1 | ENSG00000261005.1 | Inc-MLLT6-1              | LNC01257        |
| ENSG00000227392.1 HPN-AS1                                    | ENSG000002059132.7 | ENSG00000228613.1 | Inc-PKDIN-1              | ENSG0000026996.2 |
| ENSG000002250439.6 LINC01356                                  | ENSG0000022314393 | ENSG00000205917.1 | Inc-AHIC1                | ENSG00000267097.1 |
| ENSG00000237571.4 Inc-DGCR9-6-6                               | ENSG000002618011.1 | ENSG000002620281.1 | Inc-SL3CA8B-2             | ENSG00000267047.1 |
| ENSG00000254615.2 Inc-OXR1-1                                 | ENSG000002268913.1 | ENSG00000263653.2 | Inc-PKDIN-1              | ENSG00000267047.1 |
| ENSG000002272192.1 Inc-MFTIR1-1                              | ENSG00000261664.1 | ENSG00000263975.1 | Inc-PKDIN-1              | ENSG00000267047.1 |
| ENSG00000229891.1 LINC01315                                   | ENSG000002628041.1 | ENSG00000267702.3 | Inc-TTC31-4              | ENSG00000267047.1 |
| ENSG00000261175.1 LINC01288                                   | ENSG000002608101.1 | ENSG00000268532.1 | Inc-SRAN3-1              | ENSG00000267047.1 |
| ENSG00000234899.5 Inc-SL3CA11-10                              | ENSG000002303142.1 | ENSG00000213793.3 | Inc-ZNF320-1             | ENSG00000267047.1 |
| ENSG00000251161.2 Inc-VPS18-1                                | ENSG00000267715.1 | ENSG00000234703.1 | Inc-CLIC6-3              | ENSG0000026237.1 |
| ENSG00000272430.1 Inc-SY15-3                                  | ENSG000002501011.1 | ENSG000002236791.1 | LIN01705                | ENSG00000223124.1 |
| ENSG00000257718.1 CPEN8-AS1                                  | ENSG000002620191.1 | ENSG00000259802.1 | Inc-CML1                 | ENSG00000223124.1 |
| ENSG000002151904.1 Inc-RAB23-3-5                              | ENSG000002367031.1 | ENSG00000266040.1 | Inc-EPOP-2               | ENSG00000226459.1 |
| ENSG000002272620.1 Inc-SORCS2-1                              | ENSG000002601361.1 | ENSG00000250081.1 | Inc-ADAMTS6-1            | ENSG00000226459.1 |
| ENSG00000226419.2 SLC16A1-AS1                                | ENSG00000264589.1 | ENSG00000232940.1 | ENSG00000254966.1 | AC10397.1        |
| TNBC | HER2-HER+ | HER2+ | HER2-HER+ | Normal |
|------|-----------|-------|-----------|--------|
| ENSG00000267287.1 | Inc-NFATC1| Inc-FAM153C| Inc-ZC3H12D-2 | ENSG00000259702.1 | Inc-PGPEP1| ENSG00000245812.2 | LINC02202 |
| ENSG00000253348.1 | Inc-KCNMB1 | Inc-SLC7A13 | Inc-MARCH10 | ENSG00000240497.2 | Inc-TMEM212 | ENSG00000238120.1 | LINC01589 |
| ENSG00000270170.1 | NCBP2 | Inc-AC012313.2 | LINC065 | ENSG00000273328.1 | Inc-KRBOX1-1 | ENSG00000230587.1 | LINC02580 |
| ENSG00000233654.1 | Inc-MFSD6 | Inc-SLCP1 | Inc-GATA2-AS1 | ENSG00000273237.1 | Inc-HNRNPA2B1 | ENSG00000272316.1 | Inc-RAB23-53 |
| ENSG00000252721.1 | KIRREL3 | Inc-PTX4-2 | LINC02681 | ENSG00000252543.1 | Inc-MYOC1 | ENSG00000272143.1 | FGFR4-AS2 |
| ENSG00000237976.1 | Inc-PSMB4 | Inc-DPH1-1 | Inc-PXDN-2 | ENSG00000226733.1 | Inc-C6orf141 | ENSG00000232079.2 | LINC01697 |
| ENSG00000231641.1 | Inc-ATP6V1B2-3 | Inc-NKRH2-3 | Inc-RH04 | ENSG00000234072.1 | Inc-SNIXR1 | ENSG00000261625.1 | Inc-MIRPRF1 |
| ENSG00000228109.1 | MELT | Inc-ATPB1-10 | LINC01863 | ENSG00000238128.1 | Inc-CYB561 | ENSG00000253864.1 | CARM1 |
| ENSG00000263893.1 | Inc-SOX9-5 | LINC01863 | ENSG00000265428.1 | Inc-CYB561 | ENSG00000253864.1 | CARM1 |
| ENSG00000254148.3 | Inc-SLC39A11-10 | LINC01863 | ENSG00000225383.1 | Inc-8MK | ENSG00000229645.4 | Inc-SYNE3-1 |
| ENSG00000228639.2 | Inc-SLC39A11-10 | LINC01863 | ENSG00000225383.1 | Inc-8MK | ENSG00000229645.4 | Inc-SYNE3-1 |
| ENSG00000245614.2 | DDX11A-1 | LINC01863 | ENSG00000225383.1 | Inc-8MK | ENSG00000229645.4 | Inc-SYNE3-1 |
| ENSG00000260420.1 | LIN2D18 | LINC01863 | ENSG00000245080.4 | MIR135G | ENSG00000231493.3 | PGM5-AS1 |
| ENSG00000222451.3 | LNCBPRESS1 | LINC01863 | ENSG00000207406.1 | Inc-TP53INP1 | ENSG00000226813.1 | Inc-TRMT61B-1 |
| ENSG00000232785.2 | Inc-FAM207A | LINC01863 | ENSG00000267116.1 | Inc-SPAG4-1 | ENSG00000255860.1 | RYTK1-AS1 |
| ENSG00000244717.1 | Inc-CDK18 | LINC01863 | ENSG000002253217.1 | Inc-SPAG1-1 | ENSG00000255860.1 | RYTK1-AS1 |
| ENSG00000251113.6 | Inc-NNB-1 | LINC01863 | ENSG00000262016.2 | Inc-SLCP18 | ENSG00000267116.1 | Inc-SPAG4-1 |
Table 2 continued

| Normal | HER2− | HER2+ | HER2−HR | HER2+HR |
|--------|--------|--------|---------|---------|
| ENSG00000226561.1 | BNSG000002227543.3 | ENSG00000225096.1 | ENSG0000022302541.1 | BNSG000002227543.3 |
| PCAP3-AS1 | LINC01717 | Inc-PTR1-1 | Inp-CHD1-1 | Inc-TRIM48-1 |
| In-1 | In-1 | In-1 | In-1 | In-1 |
| ENSG00000226561.1 | BNSG000002227543.3 | ENSG00000225096.1 | ENSG0000022302541.1 | BNSG000002227543.3 |
| PCAP3-AS1 | LINC01717 | Inc-PTR1-1 | Inp-CHD1-1 | Inc-TRIM48-1 |
| In-1 | In-1 | In-1 | In-1 | In-1 |
| ENSG00000226561.1 | BNSG000002227543.3 | ENSG00000225096.1 | ENSG0000022302541.1 | BNSG000002227543.3 |
| PCAP3-AS1 | LINC01717 | Inc-PTR1-1 | Inp-CHD1-1 | Inc-TRIM48-1 |
| In-1 | In-1 | In-1 | In-1 | In-1 |

Mechanistic network analysis predicts activation of TGFβ1 and ECM pathways in LINC01614high BC tissue

Upstream regulator analysis (URA) on the upregulated genes in LINC01614high revealed significant enrichment for a number of networks including TGFβ1, lipopolysaccharide, TGFβ, SP1, bleomycin, SMAD3, WNT3A, EDN1, dihydrotestosterone, and AGT (Fig. 6a). Highest enrichment was for the TGFβ1 network (Z score = 5.6; Fig. 6a). Mechanistic network analysis predicted TGFβ1 to directly activate the SMAD2, NFKB1A and SP1 through TGFβ (direct activation) and TNF (inconsistent sate), and to inhibit MYC through FGF2 (direct activation) and inhibit SMAD7 through TGFβ with higher confidence level (Fig. 6b). Concordantly, LINC01614 expression demonstrated significant positive correlation with various members of the TGFβ signaling pathways (COL10A1 ($R^2 = 0.7$), SPOCK1 ($R^2 = 0.5156$), ZEB1 ($R^2 = 0.3372$), TGFBI ($R^2 = 0.2978$), TGBF1 ($R^2 = 0.1985$), ACTA2 ($R^2 = 0.1833$), and TALSN ($R^2 = 0.1909$)) in the TCGA BC cohort (Fig. 6c). Moreover, we observed several collagen
family members to be upregulated in LINC01614\textsuperscript{high} BC, suggesting enhanced extracellular matrix (ECM) formation. Illustration of the ECM network in LINC01614\textsuperscript{high} BC mapped by IPA is shown in Fig. 6d. The color shade intensity of the node correlates with the expression level of the indicated genes. Therefore, our molecular and network analyses revealed strong correlation between LINC01614 expression, TGF\textbeta\textsuperscript{1} and ECM signaling.

**Mechanistically, recombinant TGF\textbeta\textsuperscript{1} induced LINC01614 expression, while pharmacological inhibition of TGF\textbeta\textsuperscript{1} signaling (using SB-431542) and FAK (using PF-573228) inhibited LINC01614 expression in BC cells (Fig. 6e), thus implication TGF\textbeta\textsuperscript{1} and FAK signaling in regulating LINC01614 expression in BC cells.**

**Discussion**

In recent years, IncRNAs have emerged as key players in regulating cellular functions, differentiation and disease progression, including cancer, through epigenetics, chromatin remodeling, transcriptional and post-transcriptional regulation\textsuperscript{5,14}. While the number of annotated IncRNAs in the human genome has increase dramatically, functional characterization of IncRNAs and their utilization as disease biomarkers is begging to unfold. In current study, we analyzed the IncRNA transcriptome from the TCGA breast cancer dataset and performed thorough survival and bioinformatics analyses which revealed eighteen upregulated and forty-six downregulated lncRNAs in BC compared to normal breast tissue. Additionally, our data identified different IncRNA signatures associated with various BC molecular subtypes (HER2\textsuperscript{+}HR\textsuperscript{+}, HER2\textsuperscript{+}HR\textsuperscript{-}, HER2\textsuperscript{-}HR\textsuperscript{+}, and TNBC) as well as those specific to normal breast tissue. Interestingly, our data revealed a distinct IncRNA cluster for the TNBC tumors, while such segregation was less evident among the other molecular subtypes (HER2\textsuperscript{+}HR\textsuperscript{-},
Fig. 6 (See legend on next page.)
HER2+ HR−, HER2− HR+). This could be attributed to the tissue of origin for various BC molecular subtypes, where TNBC originates from ductal cells, while the HR+ and HER2− originate from luminal cells.

Our analyses identified eleven lncRNAs (LINC01614, LINC01235, Inc-LRR1-1, Inc-ODF3B-2, AC015712.5, Inc-LAMB3-1, Inc-SPP2-3, Inc-MAP9-2, MIR205HG, Inc-MAP2K6-5 and FGF14-AS2) whose expression correlated with patient outcome. Among the identified lncRNAs, LINC01614 and LINC01235 correlated with worse DFS, while LINC01235 correlated with worse OS. Interestingly, LINC01235 was downregulated in BC compared to normal tissue, while at the same time it predicted worse DFS and OS. It is plausible that due to the large heterogeneity of BC cases included the TCGA BC cohort, the expression pattern for LINC01235 did not correlate with survival data. Additionally, a previous study reported LINC01235 (also called FLJ41200; ENSG00000270547.1) as cancer-related genes that mapped telomeric and centromeric to CD274 (PDL-1) at 9p23 in small-cell lung carcinoma, suggesting possible link between LINC01235 and immune regulation in cancer.

Interestingly, our data revealed over-expression of LINC01614 in BC compared to normal tissue and its elevated expression correlated with worse DFS. More in-depth analysis revealed LINC01614 to be highly expressed in ER+ (log2 exp = 2.1), in PR+ (log2 exp = 2.2) and HER2+ (log2 exp = 2.623), while TNBC exhibited lowest expression (log2 exp = 1.1). Those data were further validated in a second cohort where highest expression was observed in the luminal B/HER2+ molecular subtype while lowest expression was observed in the HER2-TNBC molecular subtype. The expression of LINC01614 did not correlate with BC disease stage, suggesting alteration in LINC01614 expression as an early feature during BC development and progression. Concordant with our data, LINC01614 expression has been linked to lung adenocarcinoma and the LINC01614-containing signature predicted OS and DFS in patients with esophageal squamous cell carcinoma. Recently, LINC01614 has also been reported as one of the lncRNA associated survival of ER+ BC patients.

To gain more insight into plausible molecular mechanisms of LINC01614 expression and function, we dicotomized the TCGA BC cohort into LINC01614high and LINC01614low and subsequently retrieved and identified mRNA transcripts upregulated in the LINC01614high group, which revealed 187 upregulated transcripts. Interestingly, IPA analysis on the upregulated gene list suggested strong correlation between LINC01614 expression and enriched functional categories associated with tumor cell movement and invasion. Nonetheless, LINC01614high expression was most significantly associated with TGFβ signaling, suggesting possible induction of LINC01614 by TGFβ signaling. Additionally, LINC01614high tumors exhibited high expression of several collagens, suggesting possible association between LINC01614 expression and enhanced ECM formation. It is noteworthy that ECM itself could be regulated by TGFβ signaling. Mechanistic investigation validated induction of LINC01614 by TGFβ, while its expression was inhibited by small molecule inhibitor of TGFβ and FAK, suggesting its regulation by TGFβ and FAK signaling.

Our data also revealed elevated expression of LINC01614 in HER2+ BC tumors. Interestingly, we observed significant correlation between LINC01614 expression and HER2 mutation status in BC (supplementary figure 1). HER2+ (erbB2) represent 25 to 30% of breast cancer patients and is elevated expression has been associated with more aggressive BC phenotype and shorter DFS and OS. Additionally, activation of HER2 has been linked to Epithelial-to-mesenchymal transition (EMT), hence endowing cancer cells with a more aggressive and invasive phenotype. Interestingly, HER2 and TGFβ signaling cooperated in the induction of cellular processes associated with tumorigenic development in immortalized mammary epithelial cell line. Additionally, overexpression of HER2 activated the TGFβ/SMAD signaling pathway and induced SNAIL, SLUG and ZEB-1 expression and subsequent acquisition of mesenchymal phenotype. These published reports are consistent with our current data linking LINC01614 to TGFβ signaling and HER2+ molecular subtype.

Conclusions
Our data revealed the lncRNA transcriptional landscape in breast cancer and identified the lncRNA signatures...
associated with each molecular subtype. Specifically, our data provide novel insight implicating LINC01614 as unfavorable prognostic marker in BC, and its association with the HR+/HER2+ BC molecular subtype and its regulation by TGFβ and FAK signaling.

Acknowledgements
This work was supported by a start-up grant [QB13] for Dr Nehad Alajez from Qatar Biomedical Research Institute, Qatar Foundation. The publication of this article was funded by the Qatar National Library.

Conflict of interest
The authors declare that they have no conflict of interest.

Publisher’s note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

The online version of this article (https://doi.org/10.1038/s41420-019-0190-6) contains supplementary material, which is available to authorized users.

Received: 22 May 2019 Accepted: 5 June 2019
Published online: 24 June 2019

References
1. Polyak, K. Heterogeneity in breast cancer. J. Clin. Investig. 121, 3786–3788 (2011).
2. Sorlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl Acad. Sci. USA 98, 10869–10874 (2001).
3. Perou, C. M. et al. Molecular portraits of human breast tumours. Nature 406, 747–752 (2000).
4. Onitilo, A. A., Engel, J. M., Greenlee, R. T. & Mukesh, B. N. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clin. Med. Res. 7, 4–13 (2009).
5. Beckedorff F. C., Amaral M. S., Deocesano-Pereira C., Verjovski-Almeida S. Long non-coding RNAs and their implications in cancer epigenetics. Biosci. Rep. 2013 33, 667–675.
6. Liu, H. et al. Long non-coding RNAs as prognostic markers in human breast cancer. Oncotarget 7, 20584–20596 (2016).
7. Liu, Y., Sharma, S. & Watabe, K. Roles of IncRNA in breast cancer. Front. Biosci. 7, 94–108 (2015).
8. Olsson, A. et al. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. Nature 537, 698–702 (2016).
9. Vishnubalaji, R. et al. Molecular profiling of ALDH1 (+) colorectal cancer stem cells reveals preferential activation of MAPK, FAK, and oxidative stress pro-survival signalling pathways. Oncotarget 9, 13551–13564 (2018).
10. Leinonen, R., Sugawara, H. & Shumway, M. International Nucleotide Sequence Database C. The sequence read archive. Nucleic acids Res. 39(Database issue), D19–D21 (2011).
11. Bredell, M. et al. A network model of a cooperative genetic landscape in brain tumors. JAMA 302, 261–275 (2009).
12. Alajez, N. M. Large-scale analysis of gene expression data reveals a novel gene expression signature associated with colorectal cancer distant recurrence. PLoS ONE 11, e0167455 (2016).
13. Al-toub, M. et al. Pleiotropic effects of cancer cells’ secreted factors on human stromal (mesenchymal) stem cells. Stem Cell Res. Ther. 4, 114 (2013).
14. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. Nat. Rev. Genet. 10, 153–159 (2009).
15. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70 (2012).
16. George, J. et al. Genomic amplification of CD274 (PD-L1) in small-cell lung cancer. Clin. Cancer Res. 23, 1220–1226 (2017).
17. Qiu, M. et al. A novel IncRNA, LUADT1, promotes lung adenocarcinoma proliferation via the epigenetic suppression of p27. Cell Death Dis. 6, e1858 (2015).
18. Huang, G. W. et al. A three-IncRNA signature predicts overall survival and disease-free survival in patients with esophageal squamous cell carcinoma. BMC Cancer 18, 147 (2018).
19. Zhong, L. et al. A six-long non-coding RNAs signature as a potential prognostic marker for survival prediction of ER-positive breast cancer patients. Oncotarget 8, 67861–67870 (2017).
20. Verrecchia, F. & Mauvel, A. Transforming growth factor-beta signaling through the Smad pathway: role in extracellular matrix gene expression and regulation. J. Invest. Dermatol. 118, 211–215 (2002).
21. Hynes, N. E. & Stern, D. F. The biology of erbB-2/neu/HER-2 and its role in cancer. Biochim. et. Biophys. Acta 1198, 165–184 (1994).
22. Slamon, D. J. et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235, 178–182 (1987).
23. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. Cell 144, 646–674 (2011).
24. Gupta, P. & Srivastava, S. K. HER2 mediated de novo production of TGFbeta leads to SNAI-driven epithelial-to-mesenchymal transition and metastasis of breast cancer. Mol. Oncol. 8, 1532–1547 (2014).
25. Ueda, Y. et al. Overexpression of HER2 (erbB2) in human breast epithelial cells unmasks transforming growth factor beta-induced cell motility. J. Biol. Chem. 279, 24505–24513 (2004).