Journal of Breast Cancer

J Breast Cancer 2018 September; 21(3): 277-287 https://doi.org/10.4048/jbc.2018.21.e39

Check for updates

ORIGINAL ARTICLE

Long Noncoding RNA Signature and Disease Outcome in Estrogen Receptor-Positive Breast Cancer Patients Treated with Tamoxifen

Gen Wang, Xiaosong Chen, Yue Liang, Wei Wang, Yan Fang, Kunwei Shen

Comprehensive Breast Health Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Purpose: Recent data have shown that the expression levels of long noncoding RNAs (lncRNAs) are associated with tamoxifen sensitivity in estrogen receptor (ER)-positive breast cancer. Herein, we constructed an lncRNA-based model to predict disease outcomes of ER-positive breast cancer patients treated with tamoxifen. **Methods:** LncRNA expression information was acquired from Gene Expression Omnibus by re-mapping pre-existing microarrays of patients with ER-positive breast cancer treated with tamoxifen. The distant metastasis-free survival (DMFS) predictive signature was subsequently built based on a Cox proportional hazard regression model in discover cohort patients, which was further evaluated in another independent validation dataset. **Results:** Six lncRNAs were found to be associated with DMFS in the discover cohort, which were used to construct a tamoxifen efficacy-related lncRNA signature (TLS). There were 133 and 362 patients with TLS high- and low-risk signatures in the discover cohort. Both univariate and multivariate analysis demonstrated that TLS was associated with DMFS. TLS high-risk patients had worse outcomes than low-risk patients, with a hazard ratio of 4.04 (95% confidence interval, 2.83–5.77; $p<0.001$). Both subgroup analysis and receiver operating characteristic analysis indicated that TLS performed better in lymph node-negative, luminal B, 21-gene recurrence score high-risk, and 70-gene prognosis signature high-risk patients. Moreover, in a comparison of the 21-gene recurrence score and 70-gene prognosis signature, TLS showed a similar area under receiver operating characteristic curve in all patients. Gene Set Enrichment Analysis indicated that TLS high-risk patients showed different gene expression patterns related to the cell cycle and nucleotide metabolism from those of low-risk patients. **Conclusion:** This six-lncRNA signature was associated with disease outcome in ER-positive breast cancer patients treated with tamoxifen, which is comparable to previous messenger RNA signatures and requires further clinical evaluation.

Key Words: *Breast neoplasms, Long noncoding RNA, Neoplasm metastasis, Prognosis, Tamoxifen*

INTRODUCTION

Breast cancer is one of the most common fatal malignant tumors [1]. Estrogen receptor (ER)-positive breast cancer accounts for more than 60% of all breast cancer cases [2], for which endocrine therapy is among the most effective treatments. Tamoxifen, a selective estrogen-receptor modulator, is widely used in the adjuvant treatment of patients with ER-positive breast cancer [3]. However, *de novo* or acquired resistance still occurs. Approximately 30% of ER-positive patients do not respond to adjuvant tamoxifen treatment [4]. To determine whether ER-positive patients require further adjuvant treatment beyond tamoxifen, clinical and pathological parameters must be identified to predict the disease outcome following tamoxifen therapy. For patients with a high risk of relapse, additional treatment, such as chemotherapy, may be needed to decrease disease recurrence. Breast cancer tumor are heterogeneous and routine clinical and pathological factors, including age, menopausal status, ER positivity, progesterone receptor (PR) positivity, human epidermal growth factor receptor 2 (HER2) status, and Ki-67 expression level, cannot accurately predict disease outcomes following tamoxifen treatment [5]. In recent decades, several multi-gene assays, such as

Correspondence to: **Kunwei Shen**
Comprehensive Breast Health Center, the 22nd Floor, 1st Building, Ruijin Hospital, 197 Ruijin Er Road, Shanghai 200025, China
Tel: +86-21-64370045-602208, Fax: +86-21-64370045-602208
E-mail: kwshen@medmail.com.cn

This work was supported by grants from National Natural Science Foundation of China (grant number: 81472462), Medical Guidance Foundation of Shanghai Municipal Science and Technology Commission (grant number: 15411966400), Technology Innovation Act Plan of Shanghai Municipal Science and Technology Commission (grant number: 15411952500, 15411952501).

Received: March 25, 2018 **Accepted:** August 6, 2018

http://ejbc.kr | pISSN 1738-6756
eISSN 2092-9900

21-gene recurrence score, 70-gene prognosis signature, and intrinsic subtype signature, were developed and approved by the U.S. Food and Drug Administration for predicting disease outcomes of ER-positive patients, leading to more individualized administration of chemotherapy and endocrine therapy for these patients [6].

Long noncoding RNAs (lncRNAs) are transcripts longer than 200 nucleotides without protein translational potential. More than 90% of the human genome is transcribed into nonprotein coding RNAs, indicating the potentially important roles of these sequences in cancer progression in addition to messenger RNAs (mRNAs) [7]. In recent years, numerous lncRNAs such as *UCA1*, *DSCAM-AS1*, and *HOTAIR* were found to be associated with tamoxifen sensitivity in ER-positive breast cancer, suggesting that lncRNAs can be applied as prognosis biomarkers in tamoxifen-treated patients with breast cancer [8-10]. A predictive model integrating multiple lncRNAs may more accurately predict outcomes than a single lncRNA. Additionally, an lncRNA-based predictive model may provide prognosis information based on pre-existing mRNA signatures, which may cooperate with mRNA prediction models and improve outcome prediction in tamoxifen therapy.

In the present study, we used Gene Expression Omnibus (GEO) data to select lncRNAs related to tamoxifen sensitivity and construct an lncRNA-based signature to predict disease outcomes of ER-positive breast cancer patients treated with tamoxifen, which was then evaluated in an independent validation cohort. Additionally, this lncRNA-based signature was compared with the 21-gene recurrence score and 70-gene prognosis signature in these patients to investigate the potential clinical implications of this approach.

METHODS

Tamoxifen-treated patients with breast cancer profiling database

Datasets in the GEO meeting the following criteria were included in our study and formed the discover and validation datasets. First, gene expression data of patients was acquired using the Affymetrix HG-U133 A (GPL96) or HG-U133 Plus 2.0 (GPL570) microarray platform (Affymetrix, Santa Clara, USA). Next, breast cancer patients were with a full record of their ER status and treated with tamoxifen adjuvant treatment. Finally, distant metastasis-free survival (DMFS) records were available for each patient in these datasets. Since all the patients' data were obtained from public available GEO, ethic approvals of study and informed consent were already handled when they were submitted to GEO.

Determination of lncRNA expression by re-mapping approach

The lncRNA expression data from gene expression chips was obtained and analyzed as previously reported [11-13]. Briefly, the Robust Multichip Average package and Combat function in the Surrogate Variable Analysis package were utilized to normalize raw data among different datasets [14, 15]. Moreover, the Guided Principal Components Analysis package was utilized to assess the batch effect before and after normalization [16]. The statistical variance and the first two principal components from each batch were compared before and after normalization for the discover and validation datasets, respectively. The *p*-value of batch variance was analyzed accordingly.

To determine lncRNA expression, probes from the arrays were aligned to the human genome (GRCh38/hg38) using SeqMap [17] such that probes matching the lncRNA chromosomal positions from GENCODE (http://www.gencodegenes.org; GRCh38, release 25) were identified [18].

Cox proportional hazards regression prediction model construction

To identify lncRNAs correlated with DMFS, univariate Cox proportional hazards regression analysis was firstly performed to evaluate the relationship between each lncRNA and DMFS in the discover cohort. Only lncRNAs related to DMFS with $p < 0.002$ were considered statistically significant. Each lncRNA was evaluated by stratifying patients in the discover cohort with Cutoff Finder [19]. Multivariate Cox proportional hazards regression analysis was conducted by inputting the significant lncRNAs. By integrating these prognosis-related lncRNAs, a Cox proportional hazards regression prediction model was established. The predictive risk score of each patient was then calculated from the linear combination of lncRNA expression with its regression coefficients generated in multivariate Cox proportional hazards regression analysis. Separated by the optimized cutoff score value, patients were classified into a low-risk group with relatively good survival and a high-risk group with poor survival. Survival curves were derived using the Kaplan-Meier method with log-rank tests to evaluate differences in DMFS between the low- and high-risk groups with R package "survminer". Receiver operating characteristic (ROC) analysis was performed to assess the predictive capability of the model with the R package "survivalROC" [20].

Inference of biological processes driven tamoxifen resistance in tamoxifen efficacy-related lncRNA signature high-risk patients

To evaluate the biological processes driving tamoxifen resistance in the tamoxifen efficacy-related lncRNA signature

https://doi.org/10.4048/jbc.2018.21.e39

(TLS) high-risk patients, Gene Set Enrichment Analysis (GSEA; http://www.broadinstitute.org/gsea) was performed to detect the biological differences between TLS high-risk and low-risk patients using MSigDB c2: curated gene sets: all canonical pathways [21,22]. Gene sets associated with TLS high-risk patients and identified with a false discovery rate (FDR) <0.01 and $p<0.005$ were considered as statistically significant. Furthermore, we attempted to annotate the potential function of each lncRNA in the TLS. Because GSEA failed to show significant results for individual lncRNAs in our analysis, a previously described method was adopted [11,13,23]. Briefly, mRNAs highly correlated with each lncRNA were identified in the discover dataset by Pearson correlation analysis (top 1.0%). Next, the positively or negatively correlated mRNAs were input into the Database for Annotation, Visualization, and Integrated Discovery (DAVID; version 6.8; https://david.ncifcrf.gov/) [24,25]. Finally, DAVID functional annotations with FDR <0.01 and $p<0.005$ were visualized with the Enrichment Map plugin in Cytoscape (version 3.2.0; http://www.cytoscape.org) for each lncRNA [26].

Statistical analysis

All data in this study was analyzed with R software (version 3.3.1; http://www.r-project.org/) and Bioconductor (http://www.bioconductor.org/). PAM50 intrinsic subtypes, 21-gene recurrence score (Oncotype DX®; Genomic Health, Redwood City, USA) and 70-gene prognosis signature (MammaPrint®; Agendia, Amsterdam, the Netherlands) were obtained by using the "genefu" package in R [27]. A *p*-value less than 0.05 was considered significant.

RESULTS

Establishment of discover and validation datasets

A total of 1,056 tamoxifen-treated patients with ER-positive breast cancer from GSE6532, GSE9195, GSE17705, GSE19615, GSE26971, and GSE45255 datasets were enrolled in this study. In detail, 197 ER-positive patients from GSE6532 (GPL96 platform part) and 298 ER-positive patients from GSE17705 (GPL96) treated only with 5 years of tamoxifen were combined as the discovery dataset for lncRNA-based model construction. Additionally, 88 patients from GSE6532 (GPL570 platform part), 77 patients from GSE9195 (GPL570), 62 patients from GSE19615 (GPL96), 258 patients from GSE26971 (GPL96), and 74 patients from GSE45255 (GPL96) who were also treated with 5 years of tamoxifen were combined as the validation dataset. Principal component analysis revealed no significant variance among batches for both the discover and validation datasets after normalization (Supplementary Figures 1 and 2, available online).

In the discover cohort, there were 315 and 164 patients with negative and positive lymph nodes. A total of 147 patients were classified as luminal A subtype, while 348 patients had other intrinsic subtypes of tumors. For the 21-gene recurrence score, 253 tumors were classified as recurrence score low and medium, while 242 patients had high recurrence score tumors. In terms of the 70-gene prognosis signature, there were 48 and 447 patients with low-risk and high-risk signatures

Table 1. Distribution of patients and parameters correlated with DMFS in discover set

Characteristic	No.	Univariate analysis			Multivariate analysis		
		HR	95% CI	*p*-value	HR	95% CI	*p*-value
Lymph node status				0.001			0.445
Negative	315	1			1		
Positive	164	1.99	0.62–6.37		1.58	0.49–5.09	
Unknown	16						
TLS				<0.001			<0.001
Low-risk	362	1			1		
High-risk	133	4.04	2.83–5.77		3.40	2.34–4.96	
Intrinsic subtype				0.016			0.768
Luminal A	147	1			1		
Others	348	1.68	1.10–2.58		0.93	0.59–1.49	
Gene21				<0.001			0.008
Low & medium-risk	253	1			1		
High-risk	242	2.27	1.56–3.29		1.72	1.15–2.58	
Gene70				0.028			0.460
Low-risk	48	1			1		
High-risk	447	2.63	1.08–6.44		1.42	0.56–3.60	

DMFS = distant metastasis-free survival; HR = hazard ratio; CI = confidence interval; TLS = tamoxifen efficacy-related long noncoding RNA signature; Gene21 = 21-gene recurrence score; Gene70 = 70-gene prognosis signature.

Table 2. Distribution of patients and parameters correlated with DMFS in validation set

Characteristic	No.	Univariate analysis			Multivariate analysis		
		HR	95% CI	*p*-value	HR	95% CI	*p*-value
Tumor size (cm)				0.055			-
≤2	239	1			-		
>2	308	1.45	0.99–2.12		-	-	
Unknown	11						
Lymph node status				0.293			-
Negative	285	1			-		
Positive	231	1.23	0.84–1.81		-	-	
Unknown	42						
TLS				0.006			0.023
Low-risk	380	1			1		
High-risk	178	1.66	1.15–2.40		1.54	1.06–2.24	
Intrinsic subtype				0.057			-
Luminal A	180	1			-		
Others	378	1.49	0.99–2.26		-	-	
Gene21				0.038			0.103
Low & medium-risk	244	1			1		
High-risk	314	1.49	1.02–2.17		1.38	0.94–2.02	
Gene70				0.032			0.144
Low-risk	94	1			1		
High-risk	464	1.90	1.05–3.45		1.58	0.86–2.92	

DMFS = distant metastasis-free survival; HR = hazard ratio; CI = confidence interval; TLS = tamoxifen efficacy-related long noncoding RNA signature; Gene21 = 21-gene recurrence score; Gene70 = 70-gene prognosis signature.

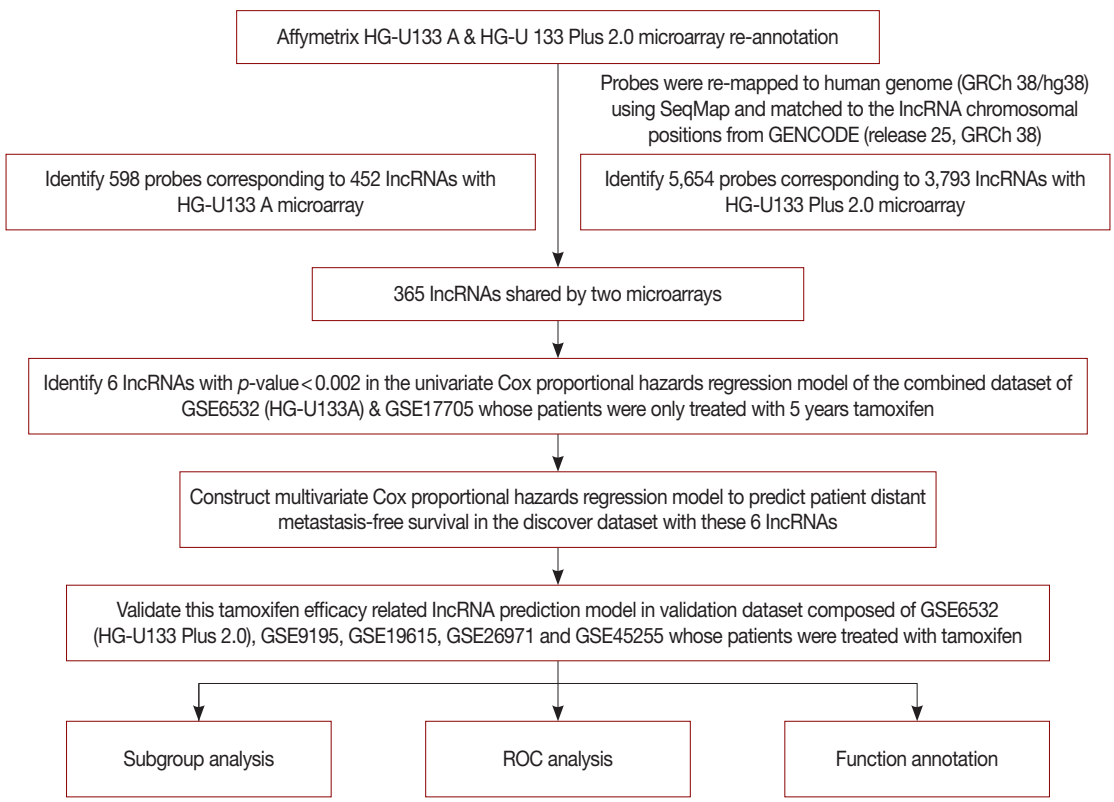


Figure 1. The diagram of the construction and validation of the tamoxifen efficacy-related long noncoding RNA (lncRNA) signature.
ROC = receiver operating characteristic.

https://doi.org/10.4048/jbc.2018.21.e39

(Table 1). The medium follow-up period was 7.4 (0.0–16.3) years and 126 patients (25.5%) experienced distant metastasis in the discover cohort.

A total of 558 patients were included in the validation cohort. Two hundred thirty-nine patients had tumors no larger than 2.0 cm, while 308 patients had tumors larger than 2.0 cm. A total of 285 patients were lymph node-negative and 231 patients were lymph node-positive. Additionally, 180 patients were classified as luminal A subtype and 378 as other subtypes. In terms of 21-gene recurrence classification, there were 244 and 314 patients classified as having low to medium and high recurrence scores, respectively. Meanwhile, 94 and 464 patients had low and high 70-gene prognosis scores, respectively (Table 2). After a medium follow-up of 6.3 (0.0–17.6) years, 117 patients (21.0%) had distant metastasis events.

Identification of lncRNAs associated with disease outcome in the discover dataset

After establishing the discover and validation datasets, we re-annotated the probes corresponding to lncRNAs for the HG-U133A and HG-U133 Plus 2.0 Affymetrix platform. A total of 598 probes corresponding to 452 lncRNAs were obtained for the HG-U133A microarray, while 5,654 probes were matching with 3,793 lncRNAs in the HG-U133 Plus 2.0 microarray. There were 365 lncRNAs overlapping between the two platforms. Figure 1 shows a diagram of the data analysis and model construction.

In the discover group of 495 ER-positive patients treated with tamoxifen, six of these 365 overlapping lncRNAs were significantly associated with DMFS in univariate Cox proportional hazard regression analysis (Table 3). Each of these six lncRNAs was capable of classifying patients into high- and low-risk groups, which could predict disease outcomes in this cohort of patients (Figure 2).

Establishment of TLS in the discover dataset

These six lncRNAs were then integrated in a multivariate Cox proportional hazard regression model to construct TLS, which assessed each tamoxifen-treated patient with an individual risk score. TLS scores were calculated as follows: TLS

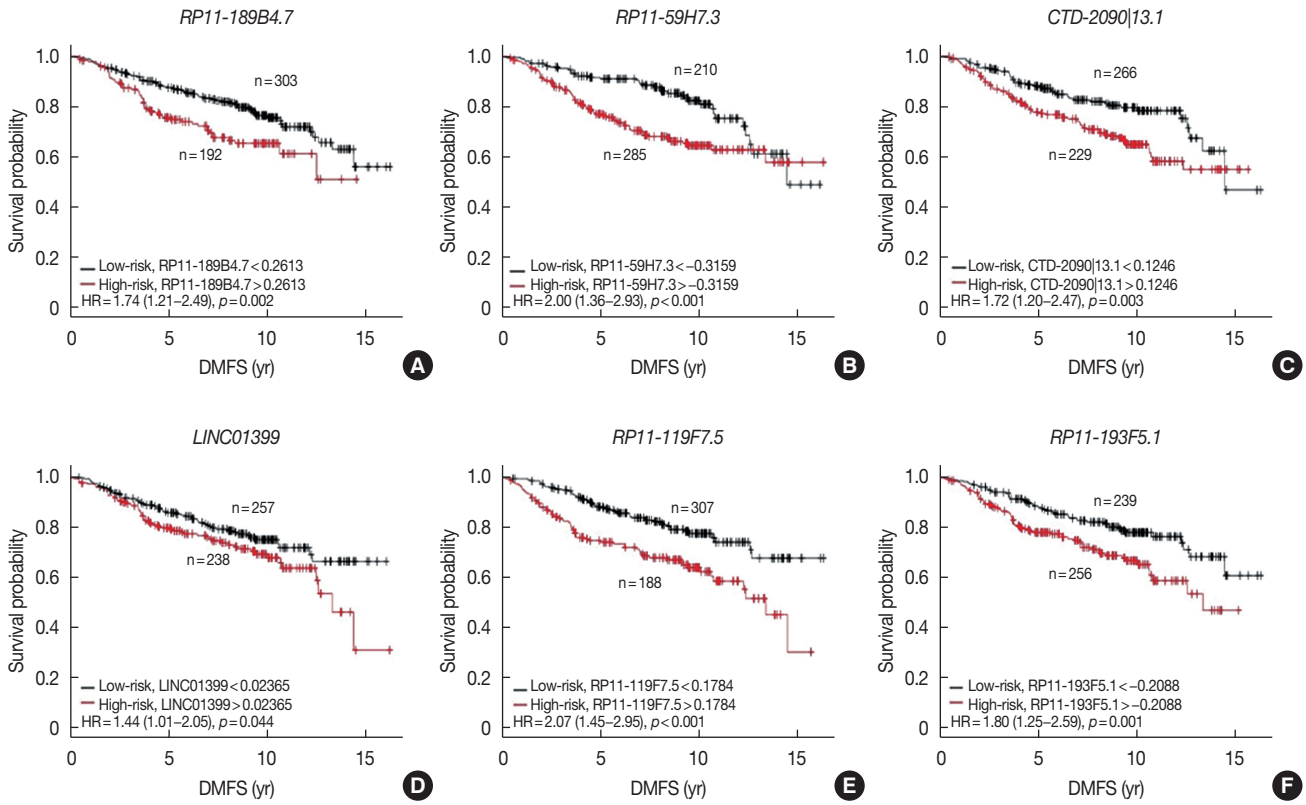


Figure 2. The six lncRNAs selected for the construction of the tamoxifen efficacy-related long noncoding RNA (lncRNA) signature. Six lncRNAs were identified with *p*-value less than 0.002 in the univariate Cox proportional hazards regression analysis of distant metastasis-free survival (DMFS) for the discover dataset. Each of them successfully divided patients in the discover dataset into high-risk and low-risk groups. (A) *RP11-189B4.7*, (B) *RP11-59H7.3*, (C) *CTD-2090|13.1*, (D) *LINC01399*, (E) *RP11-119F7.5*, and (F) *RP11-193F5.1*.
HR=hazard ratio.

Table 3. LncRNAs identified to be associated with DMFS in discover data set

Gene ID	Gene symbol	Chromosome position (GRCh38)	*p*-value	HR
ENSG00000277228.1	*RP11-189B4.7*	Chr13: 46,474,246-46,493,268 (+)	<0.001	1.47
ENSG00000259732.1	*RP11-59H7.3*	Chr15:59,121,034-59,133,250 (+)	<0.001	1.41
ENSG00000234277.2	*CTD-2090|13.1*	Chr1:227,393,591-227,431,035 (+)	<0.001	1.37
ENSG00000233080.2	*LINC01399*	Chr22:35,119,824-35,231,056 (–)	0.001	1.37
ENSG00000260400.1	*RP11-119F7.5*	Chr10:68,698,500-68,700,794 (+)	<0.001	1.35
ENSG00000258892.1	*RP11-193F5.1*	Chr14:60,879,714-60,982,585 (+)	0.001	1.34

lncRNA = long noncoding RNA; DMFS = distant metastasis-free survival; HR = hazard ratio.

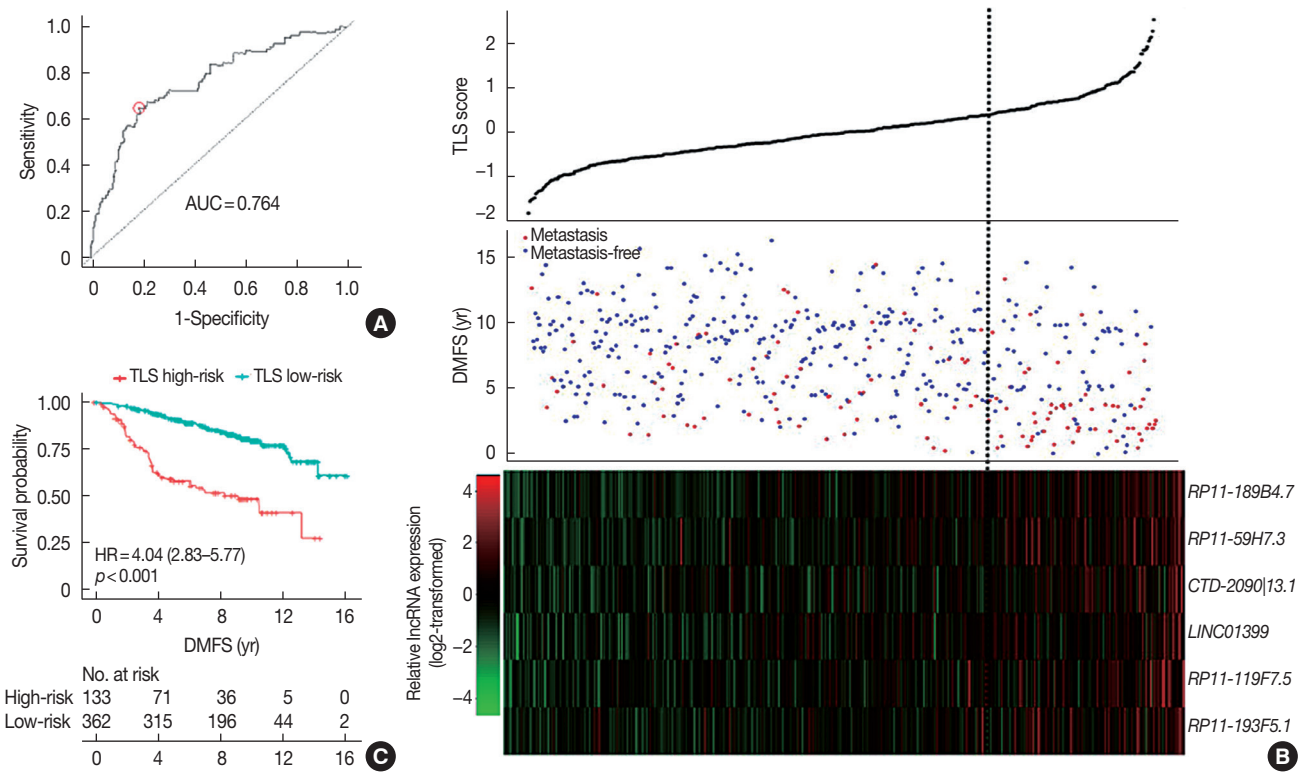


Figure 3. The prediction performance of the tamoxifen efficacy-related long noncoding RNA (lncRNA) signature (TLS) in the discover dataset after optimization. The TLS was optimized with best cutoff value. After that, the performance of TLS and expression profile of lncRNAs in TLS was analyzed in the discover dataset. (A) In the discover dataset, receiver operating characteristic (ROC) curve for the performance of TLS in distant metastasis-free survival (DMFS) was plotted with the corresponding area under the ROC curve (AUC) and the best cutoff for TLS score was determined. (B) The expression profile of the six lncRNAs in TLS, the risk score of TLS and patients' DMFS were integrated and then evaluated in the discover cohort. (C) Patients classified by TLS with optimized cutoff value were evaluated in Kaplan-Meier analysis in the discover dataset.
HR = hazard ratio.

score = 0.3753 × expression value of *RP11-189B4.7*+0.1399 × expression value of *RP11-59H7.3*+0.3604 × expression value of *CTD-2090|13.1*+0.1562 × expression value of *LINC01399*+0.4007 × expression value of *RP11-119F7.5*+0.0561 × expression value of *RP11-193F5.1*. The ROC curve of TLS was subsequently obtained with an area under the ROC curve (AUC) of 0.764 for 5 years of DMFS in the discover cohort (Figure 3A). In the ROC curve, by selecting the point nearest to the perfect prediction point, the cutoff value was set to 0.362 for TLS. Patients in the discover dataset were then separated into 133 high-risk cases and 362 low-risk cases (Table 1). The heat map revealed distinct lncRNA expression and DMFS in the TLS high- and low-risk groups (Figure 3B). Moreover, patients in the high-risk group had worse DMFS than those in the low-risk group by Kaplan-Meier analysis (Figure 3C). Univariate Cox proportional hazard regression demonstrated that lymph node status (hazard ratio [HR], 1.99; 95% confidence interval [CI], 0.62–6.37; $p = 0.001$), TLS (HR, 4.04; 95% CI, 2.83–5.77; $p < 0.001$), intrinsic subtype (HR, 1.68; 95% CI, 1.10–2.58; $p = 0.016$), 21-gene recurrence score (HR, 2.27; 95% CI, 1.56–

https://doi.org/10.4048/jbc.2018.21.e39

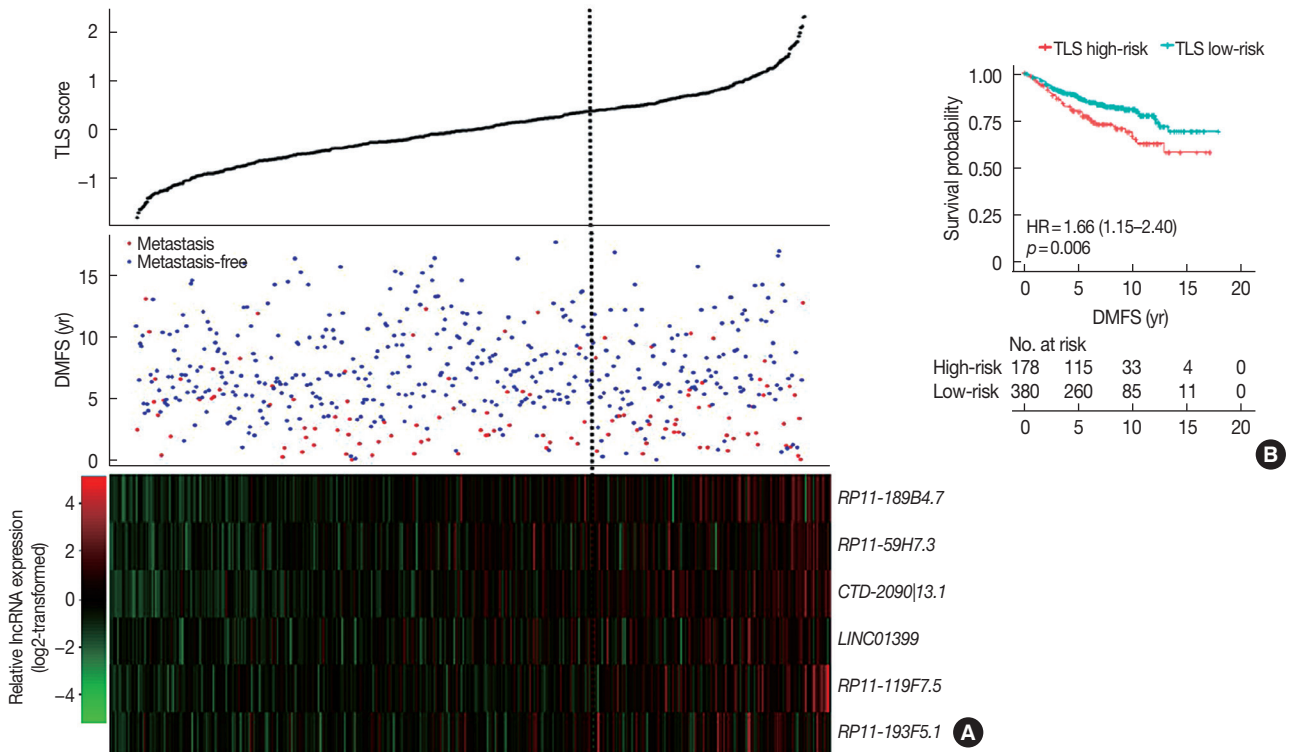


Figure 4. The evaluation of prediction capability of the tamoxifen efficacy-related long noncoding RNA (lncRNA) signature (TLS) in the validation dataset. (A) The expression profile of lncRNAs in TLS, the risk score of TLS and patients' distant metastasis-free survival (DMFS) were integrated and then evaluated in the validation cohort. (B) Patients classified by TLS with optimized cutoff value were evaluated in Kaplan-Meier analysis in the validation dataset.
HR=hazard ratio.

3.29; $p<0.001$) and 70-gene prognosis signature (HR, 2.63; 95% CI, 1.08–6.44; $p=0.028$) were related to DMFS in the discover cohort. Multivariate Cox proportional hazard regression analysis demonstrated that only TLS (HR, 3.40; 95% CI, 2.34–4.96; $p<0.001$) and 21-gene recurrence score (HR, 1.72; 95% CI, 1.15–2.58; $p=0.008$) were independent factors associated with DMFS in the discover cohort.

TLS validation

In the validation cohort, TLS separated patients into 178 high-risk cases and 380 low-risk cases (Table 2). This six lncRNA-based signature was also found to be associated with DMFS (Figure 4). In univariate Cox proportional hazard regression analysis, we found that the TLS (HR, 1.66; 95% CI, 1.15–2.40; $p=0.006$), 21-gene recurrence score (HR, 1.49; 95% CI, 1.02–2.17; $p=0.038$), and 70-gene prognosis signature (HR, 1.90; 95% CI, 1.05–3.45; $p=0.032$) were related to DMFS. Multivariate analysis showed that only TLS (HR, 1.54; 95% CI, 1.06–2.24; $p=0.023$) was independently correlated with DMFS.

Subgroup analysis of TLS and disease outcome

Overall, a total of 1,053 patients were in the discover and validation cohorts. There were 600, 395, and 58 patients with lymph node-negative, -positive, and unknown disease. Patients with known lymph node status were included in the following analysis. According to the PAM50 classification criteria, 327 cases were identified as luminal A subtype, 496 cases as luminal B subtype, 61 cases as HER2-enriched subtype, 44 cases as basal-like subtype and 125 cases as normal-like subtype. Patients with luminal A and B disease were included in further subgroup analysis. In terms of the 21-gene recurrence score, there were 232, 265, and 556 patients classified as having low-, medium-, and high-risk recurrence scores, respectively. For the 70-gene prognosis signature, 142 and 911 patients were classified as having low-risk and high-risk signatures. Subgroups analysis showed that TLS performed better in the lymph node-negative subgroup (HR, 3.28; 95% CI, 2.26–4.75; $p<0.001$) than in the positive subgroup (HR, 2.02; 95% CI, 1.39–2.93; $p<0.001$), better in the luminal B subgroup (HR, 2.41; 95% CI, 1.73–3.37; $p<0.001$) than in the luminal A subgroup (HR, 1.96; 95% CI, 1.07–3.58; $p=0.026$),

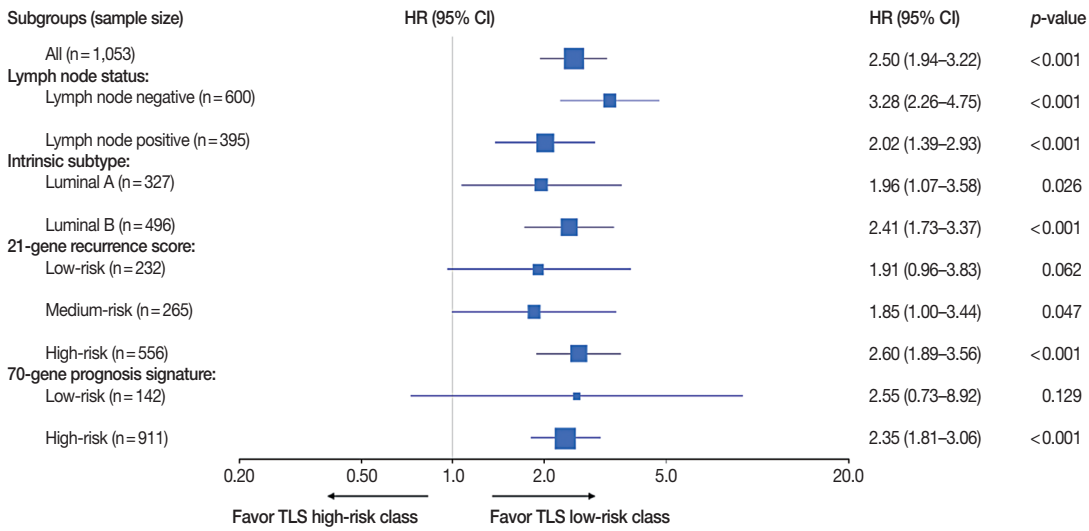


Figure 5. The evaluation of prediction power of the tamoxifen efficacy-related long noncoding RNA signature (TLS) in different subgroups of all tamoxifen treated breast cancer patients. Survival analysis of distant metastasis-free survival (DMFS) was performed to assess the prediction power of TLS in all tamoxifen treated patients, lymph node-negative subgroup, lymph node-positive subgroup, luminal A subgroup, luminal B subgroup, 21-gene recurrence score low-risk subgroup, 21-gene recurrence score medium-risk subgroup, 21-gene recurrence score high-risk subgroup, 70-gene prognosis signature low-risk subgroup and 70-gene prognosis signature high-risk subgroup.
HR = hazard ratio; CI = confidence interval.

better in the 21-gene recurrence score high-risk subgroup (HR, 2.60; 95% CI, 1.89–3.56; $p < 0.001$) than in the low-risk subgroup (HR, 1.91; 95% CI, 0.96–3.83; $p = 0.062$) and medium-risk subgroup (HR, 1.85; 95% CI, 1.00–3.44; $p = 0.047$) and better in the 70-gene prognosis signature high-risk subgroup (HR, 2.35; 95% CI, 1.81–3.06; $p < 0.001$) than in the low-risk subgroup (HR, 2.55; 95% CI, 0.73–8.92; $p = 0.129$) (Figure 5).

Comparison and integration of TLS with other gene models

The predictive accuracy of TLS was then compared with the 21-gene recurrence score and 70-gene prognosis signature in all patients with AUC values of 0.656, 0.635, and 0.631, respectively (Figure 6). In all lymph node-negative patients, the AUC values were 0.693, 0.676, and 0.623 for the TLS, 21-gene recurrence score and 70-gene prognosis signature in terms of DMFS outcome prediction. Similar results were observed in the lymph node-positive subgroup (AUC values of 0.627, 0.573, and 0.611, respectively). For different intrinsic subtypes of all patients, TLS was not superior to the 21-gene recurrence score and 70-gene prognosis signature in the luminal A subtype (AUC values of 0.548, 0.688, and 0.534, respectively). However, TLS performed much better than the 21-gene recurrence score and 70-gene prognosis signature in the luminal B subtype with AUC values of 0.659, 0.580, and 0.580, respectively. Furthermore, integration of TLS into the current 21-gene recurrence score and 70-gene prognosis signature improved the disease outcome prediction power (Figure 6).

Identification of biological processes driving tamoxifen resistance in TLS high-risk patients

GSEA was performed to determine the biological difference between the TLS high-risk and low-risk group patients and biological processes driving tamoxifen resistance in TLS high-risk patients. In GSEA, gene sets significantly associated with TLS high-risk patients were identified in the discover cohort (FDR < 0.01, $p < 0.005$) (Supplementary Table 1, available online) . Our result indicated that most gene sets associated with TLS high-risk patients were mainly correlated with cell cycle (most gene sets were correlated with G0/G1 phase, some were related to S phase and G2/M phase) and nucleotide metabolism (such as transcription, DNA and RNA synthesis). We also attempted to annotate the function of each lncRNA in the TLS by DAVID annotation analysis whose results are shown in Supplementary Figure 3 (available online).

DISCUSSION

In this study, a 6-lncRNA signature TLS was generated to predict tamoxifen sensitivity and survival outcomes of ER-positive breast cancer patients treated with tamoxifen. Multivariate Cox proportional hazard regression analysis demonstrated that TLS was independently correlated with the DMFS of ER-positive patients in both the discover and validation co-

https://doi.org/10.4048/jbc.2018.21.e39

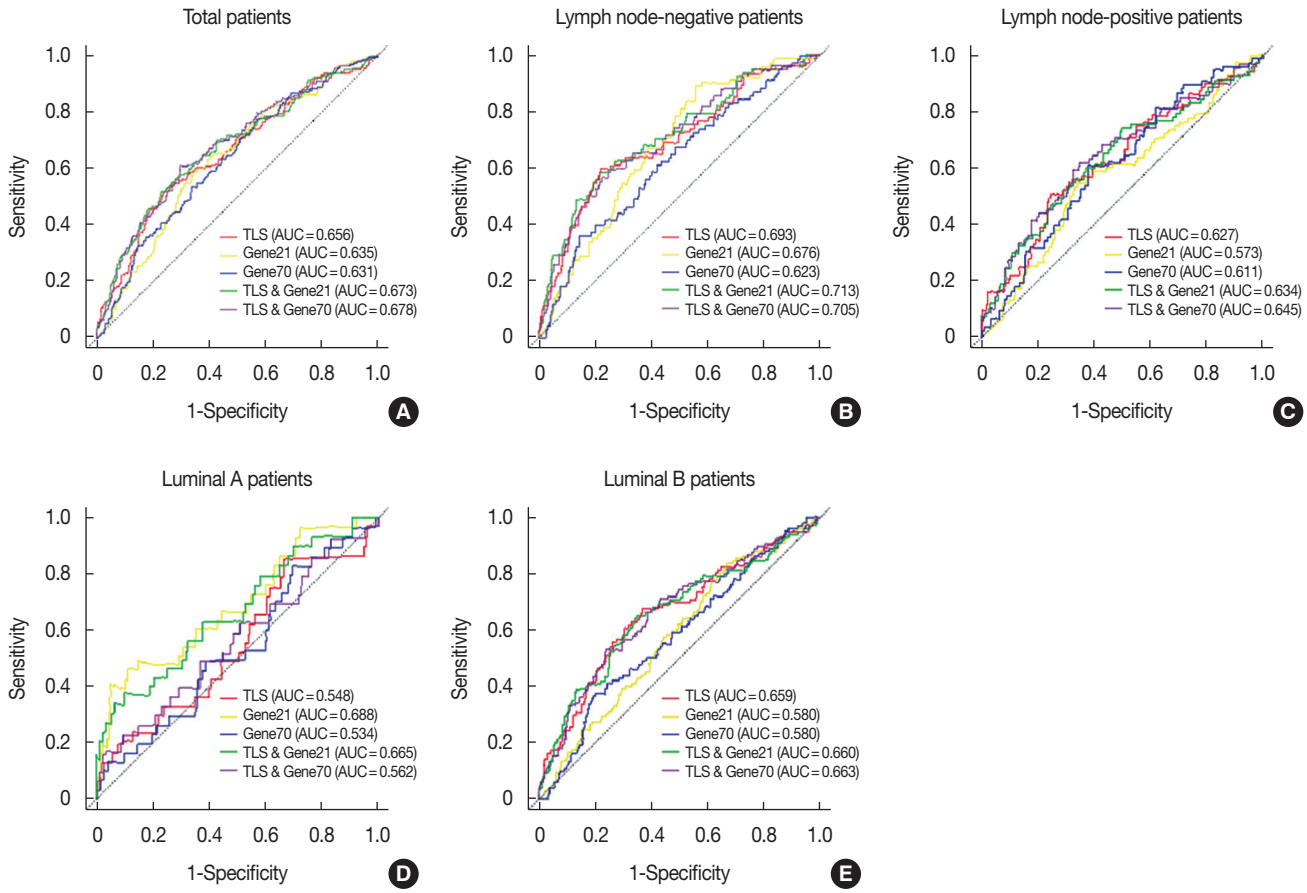


Figure 6. The comparison of the predictive power among 21-gene recurrence score (Gene21), 70-gene prognosis signature (Gene70), the tamoxifen efficacy-related long noncoding RNA signature (TLS), the integrated model of Gene21 with TLS and the integrated model of Gene70 with TLS in different subgroups of all patients. The receiver operating characteristic (ROC) of Gene21, Gene70, TLS, the integrated model of Gene21 with TLS and the integrated model of Gene70 with TLS were plotted and corresponding area under the ROC curve (AUC) was calculated in total tamoxifen treated breast cancer patients (A), all lymph node-negative patients (B), all lymph node-positive patients (C), all luminal A patients (D), and all luminal B patients (E).

horts. We also found that this TLS predicted disease outcome in different subgroups and the integration of TLS with other gene signatures improved outcome prediction.

Adjuvant tamoxifen treatment can dramatically reduce the recurrence risk of ER-positive breast cancer [3]. However, various factors lead to primary or secondary tamoxifen resistance in ER-positive patients, including difficulties in ER binding, reactivation of ER-mediated downstream biological processes, stimulation from the tumor microenvironment and mutation of the ESR1 gene. Recent studies demonstrated that lncRNAs are related to reactivation of ER downstream pathways, providing insight into the mechanism of tamoxifen resistance [28]. In our study, the lncRNA-based model included six lncRNAs, which have not been thoroughly studied. Our GSEA analysis demonstrated that GSEA high-risk patients had different cell cycle (most gene sets were correlated with G0/G1 phase, some were related to S phase or G2/M phase) and nucleotide metabolism genes from those in the low-risk group (Supplementary Table 1, available online). Estrogen can accelerate the G1 to S phase transition to promote cell cycle progression. Tamoxifen inhibits breast cancer cell growth by arresting cells at G0/G1 phase, during which time nucleotides are prepared for the next step DNA synthesis [29,30]. Therefore, we predicted that patients with high TLS scores were resistant to tamoxifen therapy because these lncRNAs interfered with G0/G1 arrest induced by tamoxifen.

The PAM50 subtype, 21-gene recurrence score and 70-gene prognosis signature are the most widely used mRNA signatures to predict the prognosis of ER-positive breast cancer patients. These mRNA signatures are superior to traditional clinicopathological factors and can predict the efficacy of adjuvant chemotherapy and endocrine therapy in these patients. Patients with the luminal B subtype, high recurrence score, or poor 70-gene prognosis signature tumor have a higher recurrence risk, leading to adjuvant chemotherapy recommendation. In contrast, for patients with luminal A, low recurrence

score, or good prognosis signature tumor who have a low risk of relapse, adjuvant chemotherapy cannot provide further benefits in addition to endocrine therapy [6]. In the present study, we found TLS provided additional prognosis information compared with the PAM50 subtype, 21-gene recurrence score and 70-gene prognosis signature. Subgroup analysis and ROC analysis demonstrated that the TLS model could further classify patients into different relapse risk groups, particularly in lymph node-negative, luminal B, 21-gene recurrence score high-risk and 70-gene prognosis signature high-risk patients. Additionally, integration of TLS with other mRNA signatures also improved the prediction accuracy for ER-positive breast cancer patients treated with tamoxifen. As a result, chemotherapy can be administrated to patients in need more accurately.

Our study also had several limitations. First, the microarray platform and repurposing method limited the number of lncRNAs available for analysis, which may not have included potentially important lncRNAs. Second, clinicopathological factors including menopausal status, tumor grade, and pathological type of enrolled patients were not available. Thus, it remains uncertain whether TLS can predict the DMFS in certain subgroups. Moreover, it was unknown whether these clinicopathological factors can be integrated with TLS, which may further improve the prediction accuracy for ER-positive patients treated with tamoxifen.

In conclusion, we identified six lncRNAs useful for predicting DMFS in ER-positive breast cancer patients treated with tamoxifen. Our six lncRNA-based model further classified these patients into high TLS score and low TLS score groups, which was independently associated with DMFS and requires further clinical evaluation.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. Harbeck N, Gnant M. Breast cancer. Lancet 2017;389:1134-50.
2. Montemurro F, Aglietta M. Hormone receptor-positive early breast cancer: controversies in the use of adjuvant chemotherapy. Endocr Relat Cancer 2009;16:1091-102.
3. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005;365:1687-717.
4. Vendrell JA, Ghayad S, Ben-Larbi S, Dumontet C, Mechti N, Cohen PA. A20/TNFAIP3, a new estrogen-regulated gene that confers tamoxifen resistance in breast cancer cells. Oncogene 2007;26:4656-67.
5. Jirström K, Rydén L, Anagnostaki L, Nordenskjöld B, Stål O, Thorstenson S, et al. Pathology parameters and adjuvant tamoxifen response in a randomised premenopausal breast cancer trial. J Clin Pathol 2005;58:1135-42.
6. Kittaneh M, Montero AJ, Glück S. Molecular profiling for breast cancer: a comprehensive review. Biomark Cancer 2013;5:61-70.
7. Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet 2006;15 Spec No 1:R17-29.
8. Wang H, Guan Z, He K, Qian J, Cao J, Teng L. LncRNA UCA1 in anti-cancer drug resistance. Oncotarget 2017;8:64638-50.
9. Niknafs YS, Han S, Ma T, Speers C, Zhang C, Wilder-Romans K, et al. The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. Nat Commun 2016;7:12791.
10. Xue X, Yang YA, Zhang A, Fong KW, Kim J, Song B, et al. LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. Oncogene 2016;35:2746-55.
11. Wang G, Chen X, Liang Y, Wang W, Shen K. A long noncoding RNA signature that predicts pathological complete remission rate sensitively in neoadjuvant treatment of breast cancer. Transl Oncol 2017;10:988-97.
12. Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, et al. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. Nat Struct Mol Biol 2013;20:908-13.
13. Zhou M, Zhong L, Xu W, Sun Y, Zhang Z, Zhao H, et al. Discovery of potential prognostic long non-coding RNA biomarkers for predicting the risk of tumor recurrence of breast cancer patients. Sci Rep 2016;6:31038.
14. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy: analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004;20:307-15.
15. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8:118-27.
16. Reese SE, Archer KJ, Therneau TM, Atkinson EJ, Vachon CM, de Andrade M, et al. A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. Bioinformatics 2013;29:2877-83.
17. Jiang H, Wong WH. SeqMap: mapping massive amount of oligonucleotides to the genome. Bioinformatics 2008;24:2395-6.
18. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 2012;22:1760-74.
19. Budczies J, Klauschen F, Sinn BV, Győrffy B, Schmitt WD, Darb-Esfahani S, et al. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. PLoS One 2012;7:e51862.
20. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000;56:337-44.
21. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
22. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003;34:267-73.

23. Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, et al. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. Nucleic Acids Res 2011;39:3864-78.
24. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.
25. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1-13.
26. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 2010;5:e13984.
27. Gendoo DM, Ratanasirigulchai N, Schröder MS, Paré L, Parker JS, Prat A, et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. Bioinformatics 2016;32:1097-9.
28. Hayes EL, Lewis-Wambi JS. Mechanisms of endocrine resistance in breast cancer: an overview of the proposed roles of noncoding RNA. Breast Cancer Res 2015;17:40.
29. Osborne CK, Boldt DH, Clark GM, Trent JM. Effects of tamoxifen on human breast cancer cell cycle kinetics: accumulation of cells in early G1 phase. Cancer Res 1983;43:3583-5.
30. Lane AN, Fan TW. Regulation of mammalian nucleotide metabolism and biosynthesis. Nucleic Acids Res 2015;43:2466-85.

Supplementary Table 1. Detail of the Gene Set Enrichment Analysis for all six lncRNAs

Gene set	Gene	ES	NES	NOM *p*-val	FDR *q*-val	FWER *p*-val
REACTOME_RNA_POL_II_TRANSCRIPTION_PRE_INITIATION_AND_PROMOTER_OPENING	38	0.68	2.24	0.00E+00	9.59E-04	8.00E-03
REACTOME_MRNA_SPLICING_MINOR_PATHWAY	35	0.76	2.23	0.00E+00	9.60E-04	9.00E-03
REACTOME_MITOTIC_M_M_G1_PHASES	137	0.66	2.26	0.00E+00	1.04E-03	6.00E-03
REACTOME_G1_S_TRANSITION	93	0.70	2.24	0.00E+00	1.05E-03	8.00E-03
REACTOME_TRANSCRIPTION	165	0.58	2.22	0.00E+00	1.13E-03	1.00E-02
REACTOME_RNA_POL_II_PRE_TRANSCRIPTION_EVENTS	57	0.66	2.24	0.00E+00	1.15E-03	8.00E-03
REACTOME_HIV_LIFE_CYCLE	108	0.60	2.21	0.00E+00	1.18E-03	1.20E-02
REACTOME_DNA_REPLICATION	156	0.66	2.26	0.00E+00	1.19E-03	6.00E-03
REACTOME_LATE_PHASE_OF_HIV_LIFE_CYCLE	96	0.60	2.21	0.00E+00	1.26E-03	1.20E-02
REACTOME_CELL_CYCLE_CHECKPOINTS	94	0.70	2.24	0.00E+00	1.28E-03	8.00E-03
REACTOME_CLASS_I_MHC_MEDIATED_ANTIGEN_PROCESSING_PRESENTATION	186	0.57	2.28	0.00E+00	1.30E-03	4.00E-03
REACTOME_CELL_CYCLE_MITOTIC	253	0.61	2.26	0.00E+00	1.39E-03	6.00E-03
REACTOME_ANTIGEN_PROCESSING_UBIQUITINATION_PROTEASOME_DEGRADATION	153	0.59	2.29	0.00E+00	1.40E-03	3.00E-03
REACTOME_HIV_INFECTION	182	0.61	2.27	0.00E+00	1.67E-03	6.00E-03
REACTOME_CELL_CYCLE	322	0.60	2.31	0.00E+00	1.98E-03	2.00E-03
REACTOME_RNA_POL_II_TRANSCRIPTION	84	0.64	2.30	0.00E+00	2.10E-03	3.00E-03
REACTOME_M_G1_TRANSITION	67	0.71	2.17	0.00E+00	2.54E-03	2.30E-02
REACTOME_HOST_INTERACTIONS_OF_HIV_FACTORS	114	0.62	2.16	0.00E+00	2.57E-03	2.50E-02
REACTOME_MITOTIC_G1_G1_S_PHASES	112	0.64	2.15	0.00E+00	2.81E-03	3.20E-02
REACTOME_DNA_REPAIR	92	0.58	2.16	0.00E+00	2.96E-03	3.20E-02
REACTOME_S_PHASE	93	0.69	2.15	0.00E+00	3.15E-03	4.00E-02
REACTOME_MITOTIC_PROMETAPHASE	66	0.62	2.12	0.00E+00	3.68E-03	5.20E-02
REACTOME_MRNA_SPLICING	71	0.62	2.12	0.00E+00	3.72E-03	5.40E-02
KEGG_BASAL_TRANSCRIPTION_FACTORS	31	0.61	2.14	0.00E+00	3.74E-03	4.80E-02
REACTOME_SYNTHESIS_OF_DNA	78	0.70	2.12	0.00E+00	3.83E-03	5.20E-02
REACTOME_CHROMOSOME_MAINTENANCE	93	0.62	2.12	0.00E+00	3.86E-03	5.40E-02
REACTOME_SCFSKP2_MEDIATED_DEGRADATION_OF_P27_P21	51	0.74	2.11	0.00E+00	3.88E-03	5.90E-02
REACTOME_REGULATION_OF_MITOTIC_CELL_CYCLE	69	0.71	2.13	0.00E+00	3.90E-03	5.10E-02
KEGG_CELL_CYCLE	105	0.56	2.11	0.00E+00	3.91E-03	5.70E-02
REACTOME_REGULATION_OF_MRNA_STABILITY_BY_PROTEINS_THAT_BIND_AU_RICH_ELEMENTS	73	0.66	2.13	0.00E+00	3.97E-03	5.10E-02
REACTOME_APC_C_CDC20_MEDIATED_DEGRADATION_OF_MITOTIC_PROTEINS	58	0.72	2.07	0.00E+00	4.32E-03	8.80E-02
REACTOME_FORMATION_OF_RNA_POL_II_ELONGATION_COMPLEX_	41	0.65	2.08	0.00E+00	4.53E-03	8.80E-02
REACTOME_NUCLEOTIDE_EXCISION_REPAIR	47	0.60	2.08	0.00E+00	4.64E-03	8.80E-02
REACTOME_MRNA_PROCESSING	114	0.57	2.10	0.00E+00	4.64E-03	6.90E-02
KEGG_HUNTINGTONS_DISEASE	151	0.56	2.07	0.00E+00	4.73E-03	9.40E-02
REACTOME_CYCLIN_E_ASSOCIATED_EVENTS_DURING_G1_S_TRANSITION_	60	0.71	2.10	0.00E+00	4.75E-03	6.80E-02
REACTOME_ORC1_REMOVAL_FROM_CHROMATIN	56	0.71	2.08	0.00E+00	4.75E-03	8.80E-02
REACTOME_APC_C_CDH1_MEDIATED_DEGRADATION_OF_CDC20_AND_OTHER_APC_C_CDH1_TARGETED_PROTEINS_IN_LATE_MITOSIS_EARLY_G1	58	0.72	2.08	0.00E+00	4.85E-03	8.70E-02
KEGG_SPLICEOSOME	84	0.63	2.08	0.00E+00	4.95E-03	8.60E-02
REACTOME_AUTODEGRADATION_OF_CDH1_BY_CDH1_APC_C	51	0.73	2.06	0.00E+00	4.95E-03	9.80E-02
REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA	99	0.58	2.08	0.00E+00	5.05E-03	8.60E-02
REACTOME_P53_DEPENDENT_G1_DNA_DAMAGE_RESPONSE	50	0.71	2.08	0.00E+00	5.08E-03	8.40E-02
REACTOME_METABOLISM_OF_NON_CODING_RNA	39	0.65	2.05	0.00E+00	5.14E-03	1.06E-01
REACTOME_CLEAVAGE_OF_GROWING_TRANSCRIPT_IN_THE_TERMINATION_REGION_	27	0.68	2.08	0.00E+00	5.18E-03	8.40E-02
REACTOME_PROCESSING_OF_CAPPED_INTRONLESS_PRE_MRNA	17	0.75	2.05	0.00E+00	5.19E-03	1.04E-01
KEGG_PARKINSONS_DISEASE	96	0.64	2.05	0.00E+00	5.30E-03	1.09E-01
REACTOME_APOPTOSIS	128	0.52	2.08	0.00E+00	5.33E-03	8.40E-02
REACTOME_TRANSCRIPTION_COUPLED_NER_TC_NER	42	0.62	2.08	0.00E+00	5.34E-03	8.00E-02
REACTOME_VIF_MEDIATED_DEGRADATION_OF_APOBEC3G	47	0.75	2.04	0.00E+00	5.39E-03	1.12E-01
REACTOME_ASSEMBLY_OF_THE_PRE_REPLICATIVE_COMPLEX	54	0.71	2.04	0.00E+00	5.76E-03	1.27E-01
REACTOME_SIGNALING_BY_WNT	59	0.67	2.03	0.00E+00	5.95E-03	1.30E-01

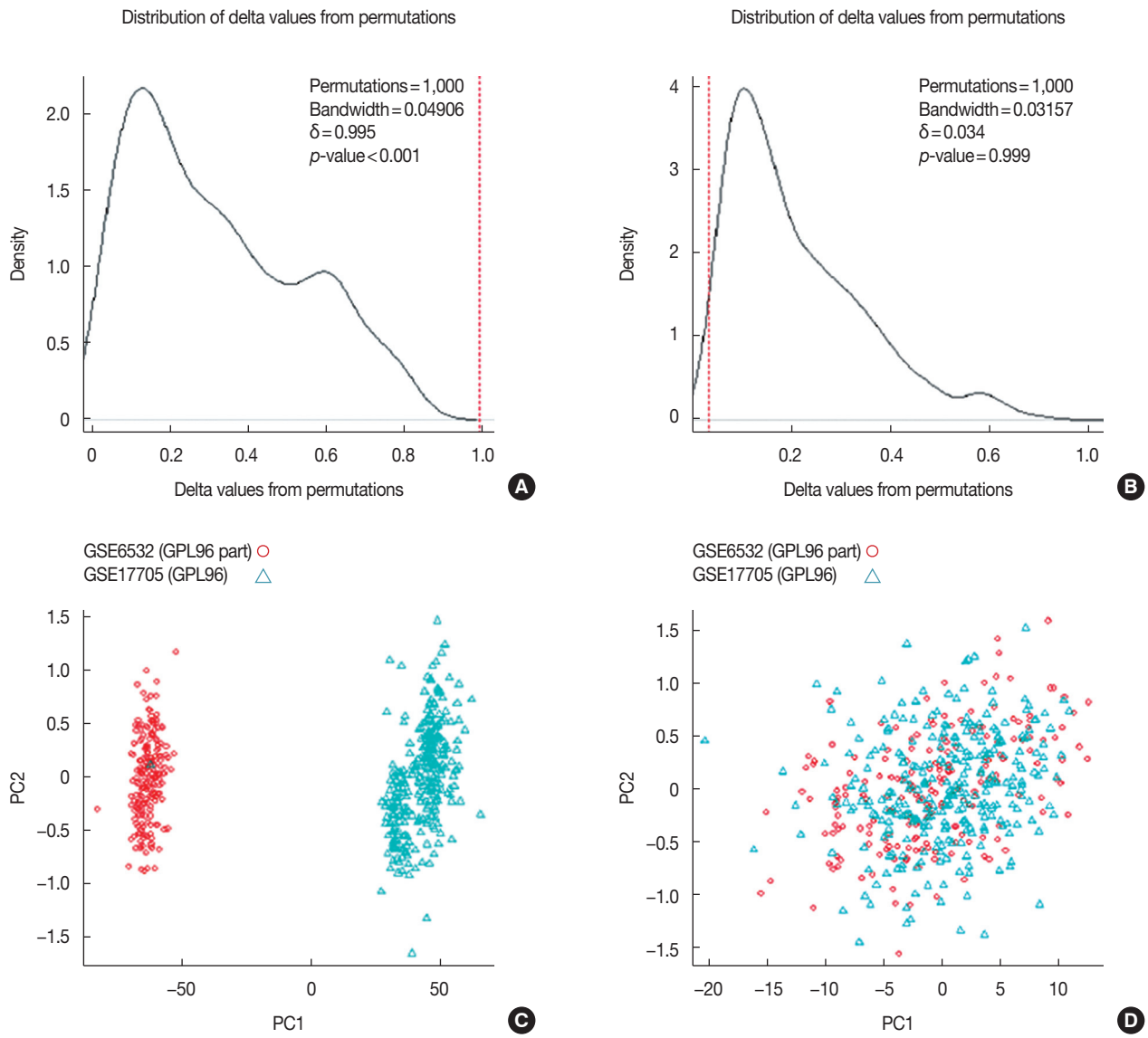
(*Continued to the next page*)

https://doi.org/10.4048/jbc.2018.21.e39

Supplementary Table 1. Continued

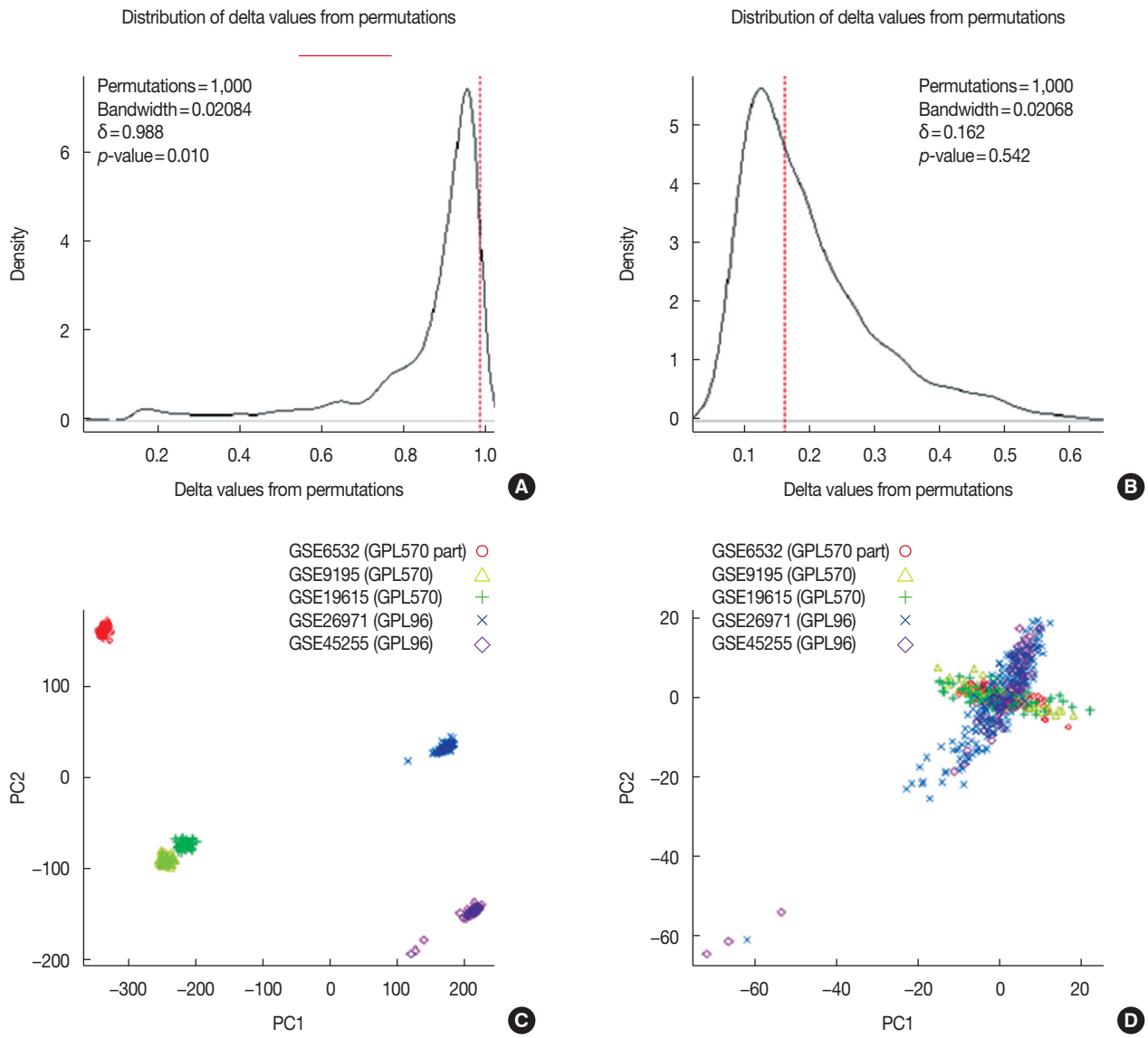
Gene set	Gene	ES	NES	NOM *p*-val	FDR *q*-val	FWER *p*-val
REACTOME_ACTIVATION_OF_NF_KAPPAB_IN_B_CELLS	56	0.69	2.03	0.00E+00	5.96E-03	1.33E-01
REACTOME_CDK_MEDIATED_PHOSPHORYLATION_AND_REMOVAL_OF_CDC6	44	0.74	2.03	0.00E+00	6.02E-03	1.37E-01
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	103	0.66	2.03	0.00E+00	6.02E-03	1.37E-01
REACTOME_SCF_BETA_TRCP_MEDIATED_DEGRADATION_OF_EMI1	46	0.74	2.02	0.00E+00	6.63E-03	1.52E-01
PID_BARD1_PATHWAY	28	0.62	2.02	0.00E+00	6.78E-03	1.61E-01
KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS	108	0.48	2.01	0.00E+00	6.88E-03	1.69E-01
REACTOME_TRNA_AMINOACYLATION	29	0.69	1.97	0.00E+00	8.05E-03	2.35E-01
REACTOME_REGULATION_OF_APOPTOSIS	51	0.67	1.98	0.00E+00	8.11E-03	2.24E-01
REACTOME_ACTIVATION_OF_GENES_BY_ATF4	21	0.62	1.97	0.00E+00	8.12E-03	2.35E-01
REACTOME_MRNA_CAPPING	27	0.65	1.98	0.00E+00	8.13E-03	2.24E-01
REACTOME_REGULATORY_RNA_PATHWAYS	21	0.64	1.97	0.00E+00	8.16E-03	2.33E-01
REACTOME_PURINE_METABOLISM	27	0.64	1.98	0.00E+00	8.17E-03	2.27E-01
REACTOME_CDT1_ASSOCIATION_WITH_THE_CDC6_ORC_ORIGIN_COMPLEX	45	0.73	1.98	0.00E+00	8.21E-03	2.18E-01
REACTOME_AUTODEGRADATION_OF_THE_E3_UBIQUITIN_LIGASE_COP1	44	0.73	1.98	0.00E+00	8.24E-03	2.21E-01
REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT	46	0.72	1.97	0.00E+00	8.24E-03	2.32E-01
REACTOME_RNA_POL_I_RNA_POL_III_AND_MITOCHONDRIAL_TRANSCRIPTION	95	0.57	1.99	0.00E+00	8.26E-03	2.04E-01
REACTOME_FORMATION_OF_TRANSCRIPTION_COUPLED_NER_TC_NER_REPAIR_COMPLEX	27	0.61	1.98	0.00E+00	8.32E-03	2.18E-01
REACTOME_FORMATION_OF_THE_HIV1_EARLY_ELONGATION_COMPLEX	31	0.65	1.96	0.00E+00	8.39E-03	2.45E-01
REACTOME_ELONGATION_ARREST_AND_RECOVERY	31	0.64	1.98	0.00E+00	8.55E-03	2.18E-01
REACTOME_METABOLISM_OF_NUCLEOTIDES	56	0.57	1.98	0.00E+00	8.61E-03	2.16E-01
REACTOME_REGULATION_OF_ORNITHINE_DECARBOXYLASE_ODC	46	0.73	1.98	0.00E+00	8.71E-03	2.15E-01
KEGG_OXIDATIVE_PHOSPHORYLATION	98	0.63	1.98	0.00E+00	8.79E-03	2.14E-01
REACTOME_ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRESS	26	0.73	1.96	0.00E+00	9.10E-03	2.63E-01
REACTOME_DOWNSTREAM_SIGNALING_EVENTS_OF_B_CELL_RECEPTOR_BCR	78	0.58	1.95	0.00E+00	9.15E-03	2.76E-01
REACTOME_G2_M_CHECKPOINTS	31	0.71	1.95	0.00E+00	9.29E-03	2.75E-01
REACTOME_ANTIVIRAL_MECHANISM_BY_IFN_STIMULATED_GENES	60	0.56	2.05	1.96E-03	5.08E-03	1.04E-01
REACTOME_ER_PHAGOSOME_PATHWAY	56	0.69	1.99	1.96E-03	8.10E-03	2.01E-01
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS_	72	0.69	1.95	1.96E-03	9.44E-03	2.74E-01
REACTOME_DESTABILIZATION_OF_MRNA_BY_AUF1_HNRNP_D0	47	0.72	1.97	1.98E-03	8.09E-03	2.35E-01
KEGG_PROTEASOME	41	0.78	2.01	1.98E-03	6.86E-03	1.64E-01
REACTOME_ANTIGEN_PROCESSING_CROSS_PRESENTATION	68	0.63	2.00	1.99E-03	7.80E-03	1.92E-01
REACTOME_MITOCHONDRIAL_PROTEIN_IMPORT	40	0.63	1.97	2.05E-03	7.99E-03	2.35E-01
REACTOME_TELOMERE_MAINTENANCE	60	0.65	1.98	2.08E-03	8.44E-03	2.18E-01
KEGG_DNA_REPLICATION	32	0.73	2.08	2.09E-03	4.42E-03	8.80E-02
KEGG_PYRIMIDINE_METABOLISM	73	0.54	1.95	2.09E-03	9.24E-03	2.76E-01
REACTOME_CYTOSOLIC_TRNA_AMINOACYLATION	20	0.74	2.00	2.10E-03	7.91E-03	1.92E-01
REACTOME_DEPOSITION_OF_NEW_CENPA_CONTAINING_NUCLEOSOMES_AT_THE_CENTROMERE	45	0.65	1.95	2.13E-03	9.35E-03	2.74E-01
REACTOME_MICRORNA_MIRNA_BIOGENESIS	20	0.65	1.96	2.20E-03	8.46E-03	2.49E-01
PID_FANCONI_PATHWAY	37	0.63	2.06	2.21E-03	4.84E-03	9.80E-02
REACTOME_INTERACTIONS_OF_VPR_WITH_HOST_CELLULAR_PROTEINS	31	0.60	1.98	2.25E-03	8.15E-03	2.22E-01

lncRNA = long noncoding RNA; ES = enrichment score; NES = normalized enrichment score; NOM *p*-val = nominal *p*-value; FDR *q*-val = false discovery rate *q*-value; FWER *p*-val = familywise-error rate *p*-value.

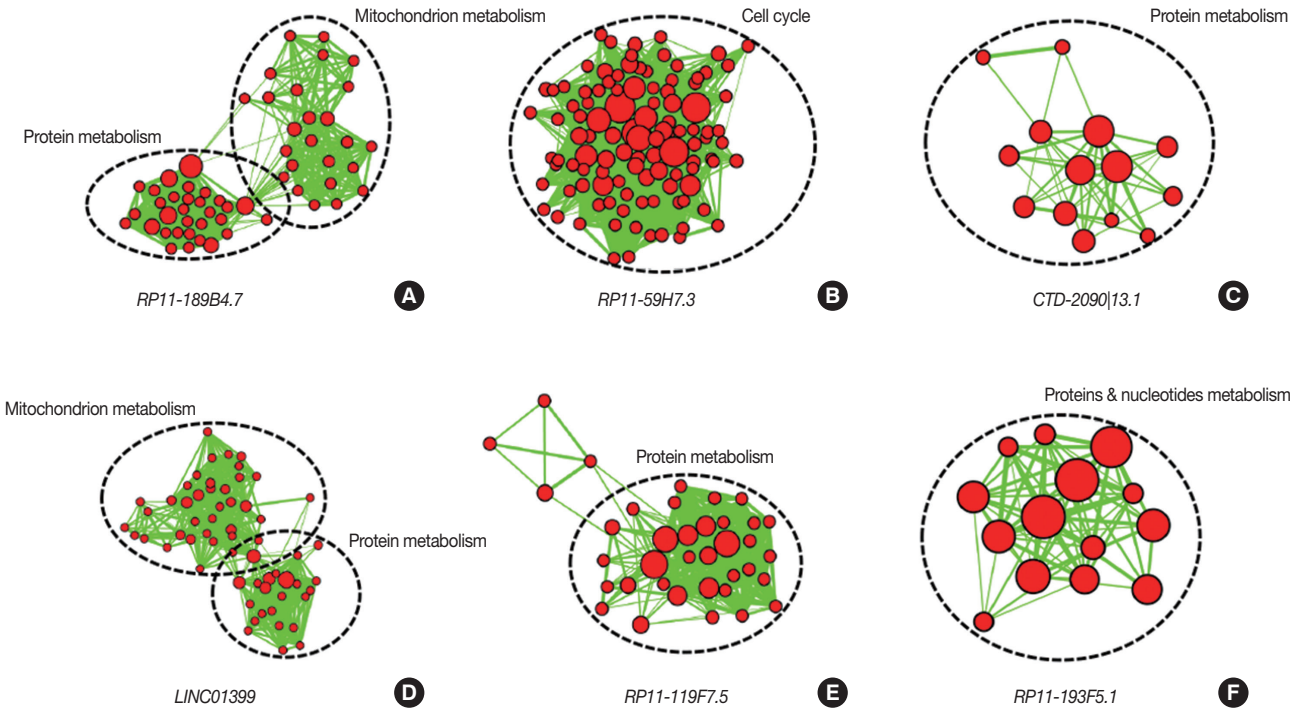


Supplementary Figure 1. The evaluation of batch effect before and after normalization in the discover dataset. The batch effect before and after Combat function normalization was assessed by Guided Principal Components Analysis package for discover dataset. The variance among batches was statistically assessed with *p*-value before (A) and after (B) Combat normalization for discover dataset. Meanwhile, the first two principle components (PC) of each batch were also compared before (C) and after (D) Combat normalization.

https://doi.org/10.4048/jbc.2018.21.e39



Supplementary Figure 2. The evaluation of batch effect before and after normalization in the validation dataset. The batch effect before and after Combat function normalization was assessed by Guided Principal Components Analysis package for validation dataset. The variance among batches was statistically assessed with *p*-value before (A) and after (B) Combat normalization for validation dataset. Meanwhile, the first two principle components (PC) of each batch were also compared before (C) and after (D) Combat normalization.



Supplementary Figure 3. Identification of the biological function associated with each individual long noncoding RNA (lncRNA) in the tamoxifen efficacy-related lncRNA signature (TLS). Messenger RNAs highly correlated with each lncRNA in TLS were input into the Database for Annotation, Visualization, and Integrated Discovery (DAVID) and results were visualized by Enrichment Map plugin in Cytoscape. (A) *RP11-189B4.7*, (B) *RP11-59H7.3*, (C) *CTD-2090|13.1*, (D) *LINC01399*, (E) *RP11-119F7.5*, and (F) *RP11-193F5.1*. Nodes represent DAVID annotation terms whose size is positively correlated with the number of genes in terms.

https://doi.org/10.4048/jbc.2018.21.e39