Quantitative Biology 2023, 11(3): 343–358
https://doi.org/10.15302/J-QB-023-0333

RESEARCH ARTICLE

# Pattern discovery of long non-coding RNAs associated with the herbal treatments in breast and prostate cancers

**Elham Dalalbashi Esfahani[1,*], Esmaeil Ebrahimie[2,3,4], Ali Niazi[1], Manijeh Mohammadi Dehcheshmeh[2,3]**

[1] Institute of Biotechnology, Shiraz University, Shiraz 7196484334, Iran
[2] Genomics Research Platform, School of Agriculture, Biomedicine and Environment, La Trobe University, Melbourne, Victoria 3086, Australia
[3] School of Animal and Veterinary Sciences, The University of Adelaide, South Australia 5005, Australia
[4] School of BioSciences, The University of Melbourne, Victoria 3052, Australia
* Correspondence: elhamdallalbashi@gmail.com

Received July 30, 2022; Revised March 8, 2023; Accepted March 28, 2023

***Background*: Accumulating evidence shows that long non-coding RNAs (lncRNAs) play critical roles in cancer progression. The possible association between lncRNAs and herbal medicine is yet to be known. This study aims to identify medicinal herbs associated with lncRNAs by RNA-seq data for breast and prostate cancer.**
***Methods*: To develop the optimal approach for identifying cancer-related lncRNAs, we implemented two steps: (1) applying protein–protein interaction (PPI), Gene Ontology (GO), and pathway analyses, and (2) applying attribute weighting and finding the efficient classification model of the machine learning approach.**
***Results*: In the first step, GO terms and pathway analyses on differential co-expressed mRNAs revealed that lncRNAs were widely co-expressed with metabolic process genes. We identified two hub lncRNA-mRNA networks that implicate lncRNAs associated with breast and prostate cancer. In the second step, we implemented various machine learning-based prediction systems (Decision Tree, Random Forest, Deep Learning, and Gradient-Boosted Tree) on the non-transformed and Z-standardized differential co-expressed lncRNAs. Based on five-fold cross-validation, we obtained high accuracy (91.11%), high sensitivity (88.33%), and high specificity (93.33%) in Deep Learning which reinforces the biomarker power of identified lncRNAs in this study. As data originally came from different cell lines at different durations of herbal treatment intervention, we applied seven attribute weighting algorithms to check the effects of variables on identifying lncRNAs. Attribute weighting results showed that the cell line and time had little or no effect on the selected lncRNAs list. Besides, we identified one known lncRNAs, downregulated RNA in cancer (DRAIC), as an essential feature.**
***Conclusions*: This study will provide further insights to investigate the potential therapeutic and prognostic targets for prostate cancer (PC) and breast cancer (BC) in common.**

**Keywords:** RNA-Seq; lncRNA; cancer; co-expression; machine learning; attribute weighting

**Author summary:** Functionally characterized lncRNAs play critical roles in cancer progression but the potential relationship between lncRNAs and herbal medicine is yet to be known. To identify this association by RNA-seq data for breast and prostate cancer, a co-expression network in response to herbal medicines was performed. GO terms and pathway analyses on differential co-expressed mRNAs revealed that lncRNAs were widely co-expressed with metabolic process genes. On the other hand, various machine learning-based prediction systems on the differential co-expressed lncRNAs were implemented. Results show that the Deep Learning model could accurately forecast cancer-related lncRNAs.

## INTRODUCTION

Both breast cancer (BC) and prostate cancer (PC) indicate a substantial proportion of newly diagnosed cancers and cancer-related deaths that occur in men and women. While various methods are used for cancer therapy, complementary and alternative medicines are increasingly sought out by cancer patients worldwide. Herbal medicines have a substantial place among these complementary and alternative medicines. The side effects and high cost of most modern medicines, as well as the improvements in quality, efficacy, and safety of herbal medicines along with the development of science and technology are all reasons an increasing number of patients are turning to medicinal herbs as therapeutic targets [1,2]. Plants produce a variety of chemical compounds, the so-called secondary metabolites, with anti-cancer properties [2,3]. Well-known specific constituents of these compounds are alkaloids, terpenoids, flavonoids, pigments, and tannins [4]. The phytochemical and positive effects of medicinal herbs in BC and PC treatments have been extensively studied [5].

Worldwide, BC is the leading cause of cancer morbidity in women, while PC is the second most common cause of cancer morbidity in men [6]. These two cancers are genetic diseases involving malapropos gene expression due to gene network dysregulation in cancer cells [6,7]. Recent studies have revealed a high correlation between the prevalence of BC and PC, suggesting they are influenced by common aspects, including genetic, epidemiological, biochemical, and mechanical [6,8], confirming the existence of a common pathogenesis framework that hopes to lead to the same therapeutic targets, including diagnostic, monitoring, prevention, and treatment strategies.

Nowadays, it is important to note that promising biomarkers for cancer diagnosis, prognosis, and therapeutic response are long non-coding RNAs (lncRNAs) [9,10]. LncRNAs comprise of a significant class of non-coding RNAs (do not encode proteins) with transcripts longer than 200 nucleotides in length. Their widespread roles at different levels of gene expression, protein expression, and epigenetic regulation have been highlighted in various diseases, including diabetes, cancer, rheumatic, osteoporosis, cardiac dysfunction, and infectious diseases, all of which involve the aberrant expression of lncRNAs [11,12]. Previous studies revealed the critical role of lncRNAs in various pathological stages of BC and PC [13,14]. Despite intense research efforts, only a small number of lncRNAs have been clearly distinguished during the onset and metastasis of these two cancers.

Compared to coding genes, lncRNAs are expressed at a lower level, and in a more tissue- and cell-specific manner [15,16]. Although lncRNAs are emerging as cancer regulators, many of their biological processes and mechanisms of action remain unclear [15,17]. "Guilt-by-association" is an informatics term that Guttman *et al*. used as a classification method for the putative function of ncRNA [18,19]. According to this approach, tightly co-express lncRNAs and protein-coding genes are presumably co-regulated. Thus, we can predict diverse roles for lncRNAs according to protein-coding genes and pathways correlated with a given lncRNAs [20,21].

To identify the differentially expressed lncRNAs on a genomic scale and gain further insights into their potential biological function, a co-expression network of lncRNAs with well-annotated protein-coding genes was constructed. Next, PPI and enrichment analyses for differentially expressed lncRNAs were performed [7].

We hypothesized that these potential key lncRNAs might interact with their corresponding coding genes to regulate cell cycle progression. However, further research should be performed to verify the correlation between these lncRNAs and target mRNAs, and whether these lncRNA-mRNA axes play an essential role in the development of BC and PC.

Finding the relationship between lncRNA and disease can help us understand the disease's mechanism and accelerate biomarkers discovery. Since discovering the potential lncRNA-disease associations in practical ways is costly and time-consuming, many computational models and machine learning tools that utilize existing data have been offered to predict potential connection patterns [22,23]. Machine learning methods aim to discern meaningful relationships between regular and target features or variables to determine possible cryptic patterns among them. The purpose of this paper is to utilize various prediction models that use different data mining algorithms to compare their accuracies in order to detect worn parts.

In the current study, we analyzed the expression of mRNAs and lncRNAs of nine BC or PC studies treated by anti-cancer herbal medicines. We used RNA-sequencing data because it is a potent way to identify lncRNAs [24,25]. To reveal new information about the functional roles of BC- and PC-related lncRNAs, we conducted co-expression of differentially lncRNAs and mRNAs. Then, we carried out protein-protein interaction, Gene Ontology, and KEGG pathway analysis. On the other hand, different prediction systems, Deep Learning (DL), Decision Tree (DT), Random Forest (RF), and Gradient-Boosted Tree (GB) were applied to the differentially co-expressed lncRNAs in order to develop the best lncRNA prediction method based on herbal cancer treatment. Five-fold cross-validation was used to compute the accuracy of each prediction model.

Overall, we found that several lncRNAs in these two

cancers may be involved with the tumorigenesis. We identified that *RP4-536B24.2* and *FUT8-AS1* could be considered as potential candidate biomarkers for BC and PC diagnosis and prognosis. We identified novel BC- and PC-associated lncRNAs and predicted their potential biological roles by co-expressing lncRNA-mRNA and bioinformatics analysis. Seven attribute weighting algorithms were applied to check the effects of variables on identifying lncRNA. Little or no effect of cell line and time on the selected lncRNAs list was approved. A common cancer-related lncRNA in BC and PC, DRAIC (downregulated RNA in cancer), is recognized as an essential feature by attribute weighting algorithms. Our research with various machine learning-based prediction systems confirmed that DL was the most accurate model (accuracy of 91.11%) in predicting potential cancer-related lncRNAs associations with a high area under the receiver operating characteristic curve (AUC) of 0.956. This study may provide some significant evidence to guide subsequent experimental studies on the altered lncRNAs in BC and PC as biomarkers or as potential targets for treatment development.

## RESULTS

The flowchart of the analysis pipeline is outlined in Fig. 1. After conducting RNA-Seq analysis by CLC genomics, we have genome-wide expression of 28,032 lncRNA and 20,338 protein-coding mRNA. We performed differential expression analysis by DESeq2 package to detect quantitative changes in expression levels between two experimental conditions to identify significant lncRNAs and mRNAs. By the criteria of FDR adjusted *P*-value < 0.05, a total of 32 lncRNAs and 195 mRNAs were identified as significantly differentially expressed in cancer and treated cancer.

### Co-expression network analysis

To construct the co-expression network, we used the "rcorr" function on the 227 differentially expressed RNAs (coding RNAs and lncRNAs). We performed the correlation analysis of the differentially expressed lncRNAs and protein-coding genes by calculating the Pearson correlation coefficient of all samples. lncRNA-mRNA pairs with |R| > 0.8 were selected for

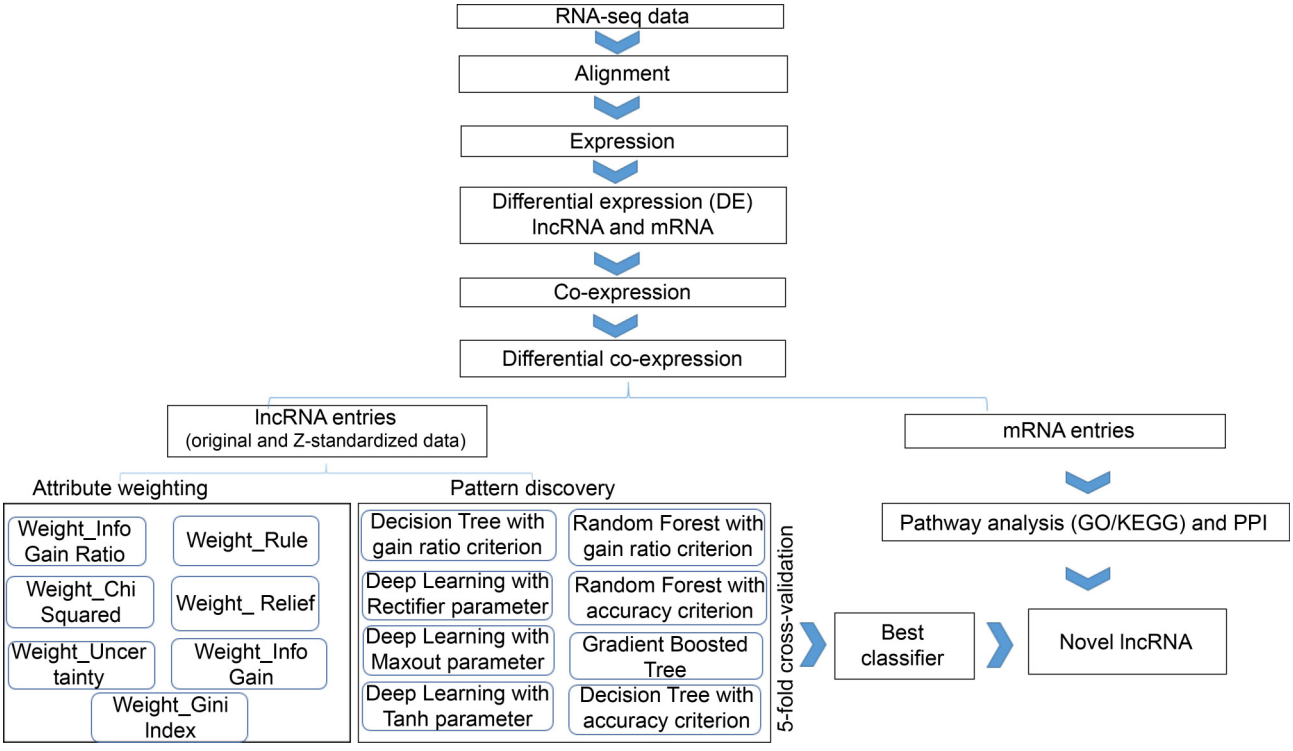


**Figure 1. The flowchart of the analysis pipeline.** After RNA-seq alignment and expression quantification of each sample, we carried out differential expression analysis to identify significant cancer-related lncRNAs and mRNAs. Then, we performed two co-expression analyses between differential lncRNAs and mRNAs to predict potential roles of lncRNAs. At this point, we have two paths to lncRNAs recognition, (1) applying attribute weighting and finding the best and most effective classification model of the machine learning approach (2) determining lncRNAs on a genomic scale by bioinformatics tools, including pathway analysis and protein–protein interaction (PPI).

co-expression network construction. Meanwhile, the common pairs in the tumor and treated groups were removed. As shown in Fig. 2A, 24 lncRNAs, 83 mRNAs, and 120 lncRNA-mRNA pairs were included in this network.

## Prediction the potential roles of lncRNAs

### PPI networks for differentially co-expressed mRNAs with lncRNAs

In this study, we found that lncRNAs are widely co-expressed with DEGs. In order to investigate the roles of these 83 differentially co-expressed genes, we constructed the PPI network analysis using the STRING database (Fig. 2B). Our network showed that several genes, including *GCLM*, *SLC7A11*, *CEBPB*, and *PGR*, play critical roles in BC and PC.

### Literature mining based PPI networks for differentially co-expressed mRNAs with lncRNAs

Text mining was used to identify cellular location of co-expressed mRNAs with lncRNAs. Also, direct interaction between proteins was extracted by text mining and extraction of sentences (Supplementary 1). As presented in Fig. 2C, *CEBPB* and *LIF* were the hub transcription factors and ligand in the network and key players.

### GO and KEGG analyses of differentially expressed lncRNAs

Among the set of 195 co-expressed protein-coding mRNAs the enriched GO terms for molecular functions, biological processes, and cellular components were the neutral amino acid transmembrane transporter activity (GO: 0015175, $P$ = 5.58E–6), the L-alpha-amino acid transmembrane transport (GO:1902475, $P$ = 7.5E–9), and the plasma membrane (GO: 0005886, $P$ = 4.43E–4) (Table 1), respectively.

GO analysis showed that the co-expressed mRNAs are biologically involved in L-alpha-amino acid transmembrane transport, proline metabolic process, proline catabolic process to glutamate, lung alveolus development, and 4-hydroxyproline catabolic process. Our GO analysis strongly supported these previous studies that lncRNAs play a vital role in tumorigenesis and regulate different aspects of cellular energy metabolism [26,27].

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was used to evaluate the biological significance of that mRNAs. The TNF signaling pathway that regulate immune cells and induce apoptosis, necrosis, angiogenesis, immune cell activation, differentiation, and cell migration [28] ($P$ = 7.9E–3), metabolic pathways that promote cancer cell survival and growth [29] ($P$ = 3.3E–2), the intriguing third pathway, prostate cancer ($P$ = 4.6E–2) and carbon metabolism ($P$ = 7.3E–2) are the most significant pathways in our KEGG analysis (Supplementary 2 and 3). Enrichment analysis based on the co-expression network indicated that lncRNAs might be associated with cancer-related pathways.

### Predicting the functions of lncRNA based on co-expression network

Although the majority of the lncRNAs may impact human cancers and diseases [11], their putative functions remain largely unknown. Nowadays, the prediction of the roles of lncRNAs may be inferred from the co-expression network [30]. In this study, we observed two hub lncRNAs, *RP4-536B24.2* (antisense to *SLC7A5*) and *FUT8-AS1* (*FUT8* antisense RNA 1), with the highest numbers of mRNAs associations and high co-expression with various cancer genes (Table 2).

## Data mining

### Attribute weighting algorithms selected three lncRNAs as the cancer-related lncRNAs

As data were normalized before running the attribute weighting models, all resulting weights were between 0 and 1. Features with weights closer to 1 reflect the importance of each variable regarding the target label. The results of attribute weighting algorithms application on eight approaches are presented in Supplementary 4−10. We calculated the average of attribute weighting algorithms for each gene and variable; a gene was assumed to be important if the average of the assigned weight was closer to 1 (Table 3). Attribute weighting results showed that the cell line and time had no or little effect on the selected gene list. Also, ENST00000498938.2, ENST00000502514.5, and ENST00000489011.1 were the key predictive parameters (Table 3). ENST00000498938.2 (DRAIC-known cancer-related lncRNA) was the most important lncRNA, confirmed by more than 50% of attribute weighting algorithms. Four algorithms (Info Gain Ratio, Gini Index, Uncertainty, and Info Gain) assigned the highest possible weight (1.0) to this lncRNA. ENST00000502514.5, ENST00000489011.1, ENST00000509144.2, and ENST00000558107.1 lncRNA placed second to fifth, respectively. Other variables did not attain notable weights. The lowest weights belonged to the cell line and time variable (Table 3).

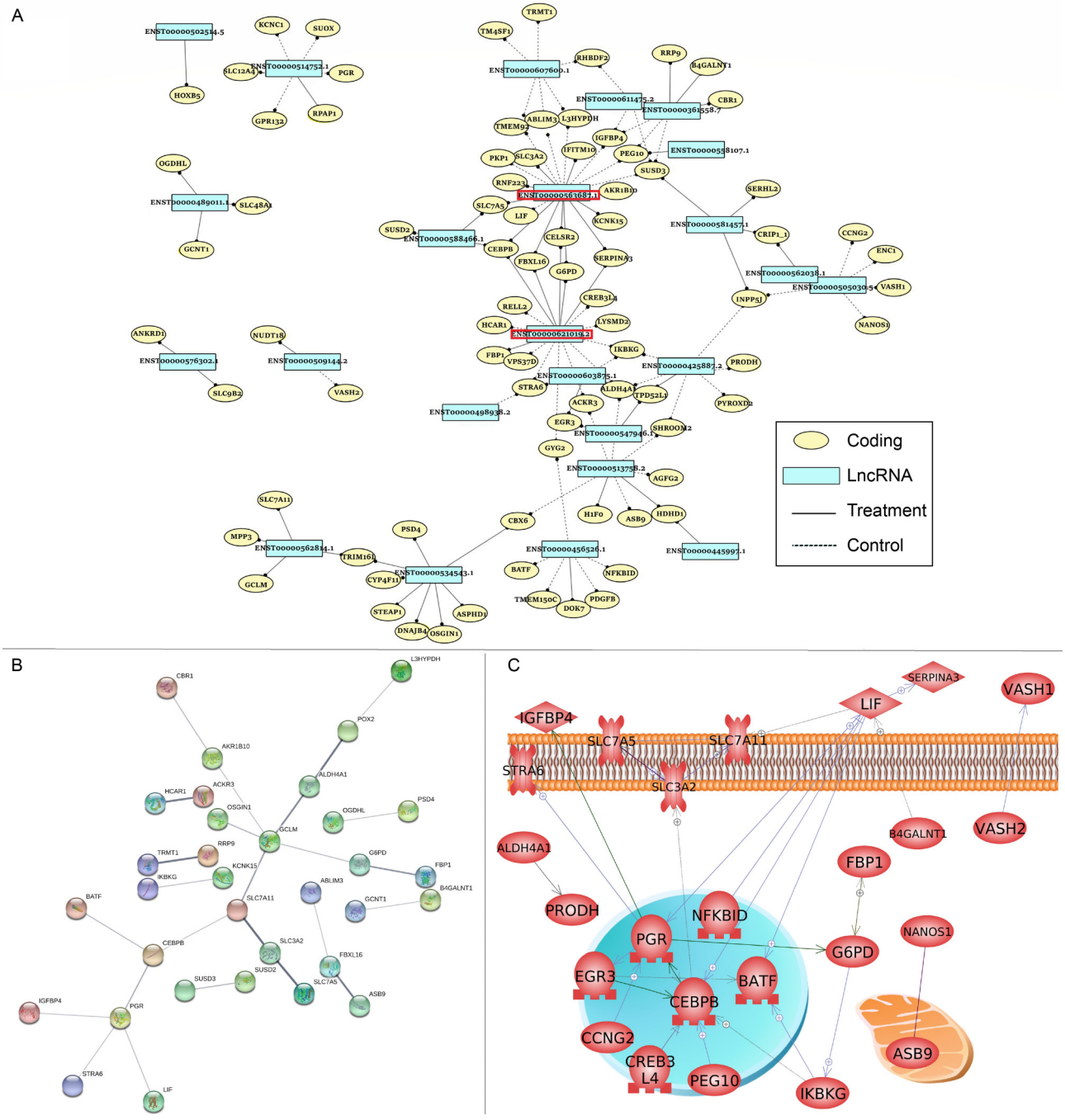


**Figure 2. Pathway analysis figures used to predict the potential roles of lncRNAs.** (A) Construction of lncRNA-mRNA co-expression networks of BC and PC. Nodes indicate the lncRNAs and mRNAs and edges represent the expression correlation between them. Blue rectangle-nodes represent lncRNAs, and yellow circle-nodes represent mRNAs. Node font size indicates the degree of a node. Edge shapes indicate the type of correlation between lncRNA and mRNAs: dashed edges for the control condition, solid edges for treatment condition, arrow edges for negative correlation, and pin edges for positive correlation. Two hub lncRNAs are highlighted with a red rectangle. (B) Construction of protein–protein interactions (PPI) networks for differentially expressed genes in BC and PC. STRING analysis of PPI of the 83 differentially co-expressed mRNAs, a total of 25 edges were found between 33 of the genes (disconnected nodes hid in the network). The thickness of edges indicates the confidence of interaction. (C) Construction of PPI networks for differentially expressed genes in BC and PC. Pathway studio (Elsevier) was used for analysis. BC, breast cancer; PC, prostate cancer.

**Table 1 Only the top 15 enriched Gene Ontology terms of differentially expressed genes in BC and PC herbal treatment are presented**

| GO ID | GO names | GO terms | COUNT | *P*-value |
|---|---|---|---|---|
| GO.0015175 | Neutral amino acid transmembrane transporter activity | MF | 2 | 5.58E–6 |
| GO.0015179 | L-amino acid transmembrane transporter activity | MF | 2 | 9.80E–6 |
| GO.0004181 | Metallocarboxypeptidase activity | MF | 2 | 3.26E–4 |
| GO.0001085 | RNA polymerase II transcription factor binding | MF | 2 | 8.02E–4 |
| GO.0001228 | DNA-binding transcription activator activity, RNA polymerase II-specific | MF | 5 | 1.97E–3 |
| GO.1902475 | L-alpha-amino acid transmembrane transport | BP | 2 | 7.50E–9 |
| GO.0006560 | Proline metabolic process | BP | 2 | 4.05E–8 |
| GO.0010133 | Proline catabolic process to glutamate | BP | 2 | 8.62E–8 |
| GO.0048286 | Lung alveolus development | BP | 4 | 8.86E–8 |
| GO.0019470 | 4-hydroxyproline catabolic process | BP | 2 | 2.90E–7 |
| GO.0005886 | Plasma membrane | CC | 27 | 4.43E–4 |
| GO.0009898 | Cytoplasmic side of plasma membrane | CC | 2 | 6.85E–4 |
| GO.0010008 | Endosome membrane | CC | 4 | 9.00E–4 |
| GO.0031093 | Platelet alpha granule lumen | CC | 2 | 1.77E–3 |
| GO.0016324 | Apical plasma membrane | CC | 4 | 4.26E–3 |

BP, biological process; CC, cellular component; MF, molecular function.

**Table 2 The five top lncRNAs with the highest numbers of mRNA associations in differential co-expression in BC and PC herbal treatment**

| lncRNA | Ensembl ID | LncRNA chromosome | Transcript type | mRNAs |
|---|---|---|---|---|
| *RP4-536B24.2* | ENST00000563687.1 | Chr16 | Antisense | 19 |
| *FUT8-AS1* | ENST00000621019.2 | Chr14 | Antisense | 16 |
| *CTD-2140G10.4* | ENST00000534543.1 | Chr11 | Antisense | 8 |
| *APOBEC3B-AS1* | ENST00000513758.2 | Chr22 | Antisense | 8 |
| *AC008268.1* | ENST00000425887.2 | Chr2 | LincRNA | 7 |

The most intriguing aspect of this part is ENST00000498938.2, known cancer-related lncRNA, DRAIC (downregulated RNA in androgen independent cells).

Figure 3 plots the expression response of DRAIC to herbal treatments in different BC and PC cell lines. In all of ECC-1, MCF-7, PC-3, and Pca, the expression of DRAIC significantly decreased in response to herbal treatment according to Bayesian Estimation Supersedes the t-test (BEST). This pattern was independent from the cell line type and time of herbal treatment as outlined by attribute weighting results. The developed pipeline in this study, integrating co-expression analysis with attribute weighting models, was successful in the selection of lncRNAs with a similar trend of response to herbal treatments across different cell lines.

Out of 6246 mined variables (6245 gene expression data and type of tissue), tissue type received the lowest weight, demonstrating the success of attribute weighting models in developing a tissue-independent signature of CR.

Figure 3 dot plots visualizing the expression of genes in the transcriptomic signature of CR (A) C1qa, (B) Plcg1, (C) Map4k2, and (D) Zbtb2, derived from attribute weighting models, in the hypothalamus, amygdala, pituitary, and adrenal glands of control and CR rats ($n$ = 5/group/region). Attribute weighting was successful in the selection of genes with a similar trend of response to long-term CR across different tissues. Values represent gene expression based on FPKM (fragments per kilobase of exon per million mapped fragments) values. *P*-value * ⩽ 0.05; ** ⩽ 0.01; *** ⩽ 0.001.

Bayesian Estimation Supersedes the t-test was used for further evaluation of key discovered herbal treatment responding lncRNAs, such as DRAIC. BEST applies Bayesian model for estimating the difference between the means of two groups and yields a probability distribution over the difference. Based on the distribution, the mean credible value can be considered as the best guess of the actual difference and the 95%

**Table 3 Attribute weighting of the Z-standardized lncRNAs in BC and PC herbal treatment, regardless of herbal medicine type and concentration variable, according to the 7 applied attribute weighting algorithms**

| Attribute | Weight_ Info Gain Ratio | Weight_ Rule | Weight_ Chi Squared | Weight_ Gini Index | Weight_ Uncertainty | Weight_ Relief | Weight_ Info Gain | Average |
|---|---|---|---|---|---|---|---|---|
| ENST00000498938.2 (DRAIC) | 1.0 | 0.7 | 0.8 | 1.0 | 1.0 | 0.8 | 1.0 | 0.9 |
| ENST00000502514.5 | 0.7 | 0.4 | 1.0 | 0.7 | 0.8 | 1.0 | 0.6 | 0.7 |
| ENST00000489011.1 | 0.7 | 1.0 | 0.6 | 0.4 | 0.6 | 0.9 | 0.5 | 0.7 |
| ENST00000509144.2 | 0.6 | 0.6 | 0.6 | 0.5 | 0.6 | 0.8 | 0.6 | 0.6 |
| ENST00000558107.1 | 0.7 | 0.9 | 0.7 | 0.6 | 0.6 | 0.4 | 0.5 | 0.6 |
| ENST00000588466.1 | 0.6 | 0.9 | 0.8 | 0.5 | 0.6 | 0.5 | 0.4 | 0.6 |
| ENST00000607600.1 | 0.4 | 0.4 | 0.6 | 0.5 | 0.5 | 0.9 | 0.4 | 0.5 |
| ENST00000361558.7 | 0.6 | 0.9 | 0.2 | 0.4 | 0.4 | 0.6 | 0.3 | 0.5 |
| ENST00000514752.1 | 0.6 | 0.3 | 0.5 | 0.4 | 0.5 | 0.6 | 0.4 | 0.5 |
| ENST00000445997.1 | 0.6 | 0.8 | 0.5 | 0.4 | 0.4 | 0.1 | 0.4 | 0.5 |
| ENST00000513758.2 | 0.7 | 0.5 | 0.4 | 0.5 | 0.3 | 0.0 | 0.6 | 0.4 |
| ENST00000576302.1 | 0.6 | 1.0 | 0.2 | 0.4 | 0.2 | 0.2 | 0.4 | 0.4 |
| ENST00000547946.1 | 0.5 | 0.3 | 0.3 | 0.3 | 0.4 | 0.6 | 0.3 | 0.4 |
| ENST00000611475.2 | 0.6 | 0.7 | 0.3 | 0.3 | 0.5 | 0.2 | 0.3 | 0.4 |
| ENST00000562814.1 | 0.5 | 0.5 | 0.4 | 0.3 | 0.5 | 0.2 | 0.3 | 0.4 |
| ENST00000505030.5 | 0.6 | 0.3 | 0.6 | 0.3 | 0.6 | 0.0 | 0.3 | 0.4 |
| ENST00000581457.1 | 0.4 | 0.8 | 0.5 | 0.1 | 0.6 | 0.0 | 0.1 | 0.4 |
| ENST00000603875.1 | 0.6 | 0.4 | 0.4 | 0.3 | 0.6 | 0.1 | 0.3 | 0.4 |
| ENST00000456526.1 | 0.5 | 0.3 | 0.6 | 0.2 | 0.6 | 0.1 | 0.2 | 0.4 |
| ENST00000562038.1 | 0.3 | 0.2 | 0.4 | 0.5 | 0.3 | 0.3 | 0.4 | 0.3 |
| ENST00000563687.1 | 0.5 | 0.4 | 0.2 | 0.3 | 0.3 | 0.2 | 0.3 | 0.3 |
| ENST00000534543.1 | 0.5 | 0.4 | 0.2 | 0.2 | 0.2 | 0.1 | 0.3 | 0.3 |
| ENST00000425887.2 | 0.4 | 0.4 | 0.3 | 0.1 | 0.6 | 0.0 | 0.1 | 0.3 |
| ENST00000621019.2 | 0.3 | 0.2 | 0.4 | 0.3 | 0.4 | 0.0 | 0.2 | 0.3 |
| Cell line | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.7 | 0.0 | 0.1 |
| Time | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.1 |

Highest Density Interval (HDI) as the range where the actual difference is with 95% credibility.

When comparing the results of Tables 2 and 3, certain lncRNAs not part of the list of the highest numbers of mRNA associations can be observed as very powerful and essential in attribute weighting results (Table 3), suggesting they, either alone or in association with other lncRNAs, are useful like ENST00000498938.2, but determining their specific function will require many clinical trials. In contrast, lncRNAs associated with more genes (Table 2) could be critical and act as master regulators, and their possible roles could be determined by the co-expression method.

## Cancer-related lncRNA prediction algorithms

To find the best statistical models for identifying and predicting the candidate lncRNAs responding to herbal treatment in BC and PC, we ran four popular algorithms (DL, DT, GBT, and RF) on 44 samples with 24 differential co-expressed lncRNAs and four categorical variables (cell line, extract, time, and concentration level).

The performance and repeatability of the models and selected lncRNAs were evaluated by five-fold cross-validation to examine the statistical performance of a model in analysis of unseen/future data and test how accurately a model may perform in practice. Cross validation has nested structure and generates training and testing subsets [31]. The training subset employs for the development of the model. Then, the training-originated model mines the testing subset. The testing step gives an indication of model performance, based on a range of performance indexes, including accuracy, AUC, sensitivity, and specificity. In five-fold cross-validation, the procedure was repeated by 5 times. Then, the average and standard deviation of performance indexes on testing subsets were recorded. All the models

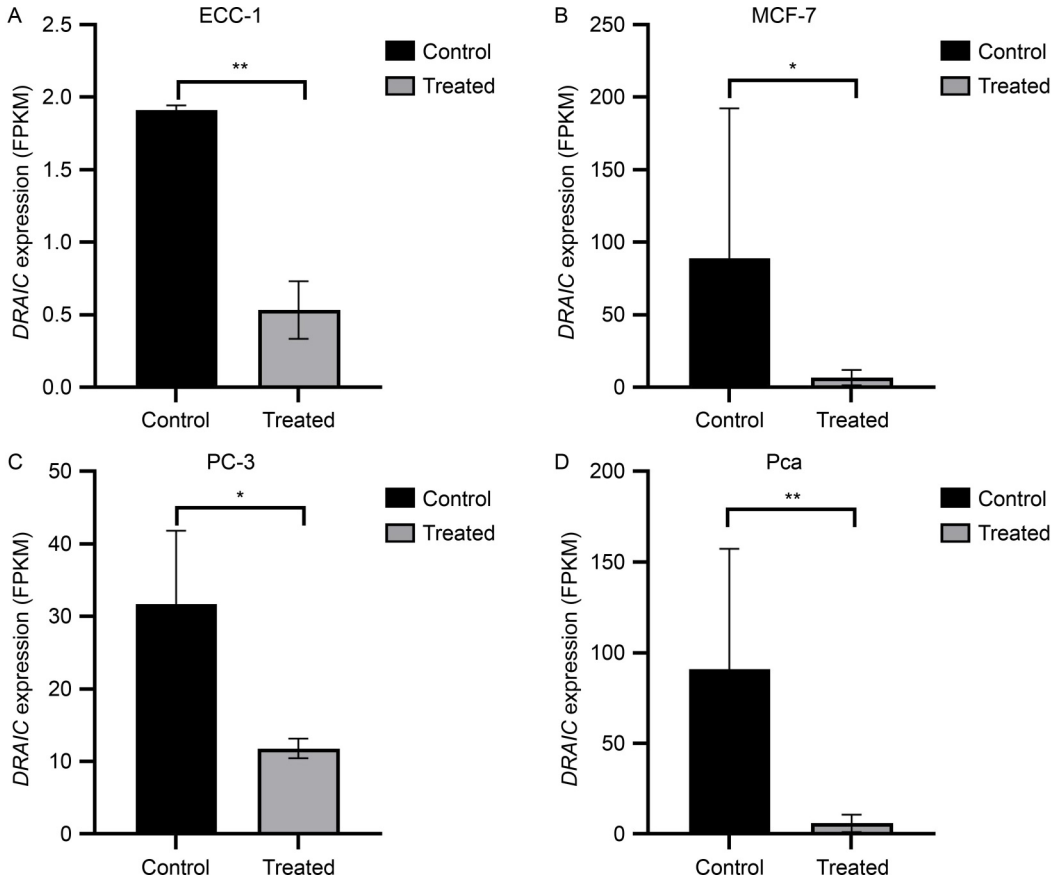


**Figure 3. DRAIC has a similar response pattern to herbal treatments in breast and prostate cancer cell lines.** *DRAIC* expression significantly decreased after herbal treatment in ECC-1 (A), MCF-7 (B), PC-3 (C), and Pca (D) cell lines of genes with a similar trend of response to long-term CR across different tissues, according to the Bayesian Estimation Supersedes the t-test (BEST). * demonstrates significant at $P = 0.05$ and ** demonstrates statistical significance at $P = 0.01$.

developed were trained and tested on both regular and Z-standardized datasets. To eliminate individual source dependency and batch effects, we focused on the Z-standardized dataset and the results of the approach in which the extract and concentration of medicinal herbs have was deleted. Validation of model performance on independent test sets using sensitivity, AUC, and specificity criteria offer a good estimation of model performance on future unseen data.

## The best presentation to show the association of cancer-related lncRNAs and parameters based on the high-performance Decision Tree algorithms

We investigated the Decision Tree for a straightforward interpretation, an inverted tree-like graph with a root at the top and grows downwards. The primary goal of the Decision Tree model is to create a classification model that predicts the value of label/target class based on several input features or variables (here 24 lncRNAs, extract, concentration, cell line, time). Interestingly, the same lncRNA selected as the essential feature by attribute weighting models was again selected by Decision Tree models to generate the trees, showing a high correspondence between attribute weighting models and Decision Tree models.

As shown in Fig. 4, the expression of ENST00000 498938.2 was set as the crucial feature in the tree's root. When its expression was equal to or more than −0.142, the cell line goes to cancer. However, cancer could be suppressed when its expression was equal or less than −0.142 and associated with other lncRNAs.

## Deep learning algorithms predicted cancer-related lncRNAs with up to 90% accuracy, sensitivity, specificity, F measure and high AUC

For assessing the quality of predictive models, accuracy is often the first notable criterion since it measures the ratio of correct predictions to the total number of cases evaluated. As presented in Supplementary 11, DL was the most accurate model for predicting cancer-related

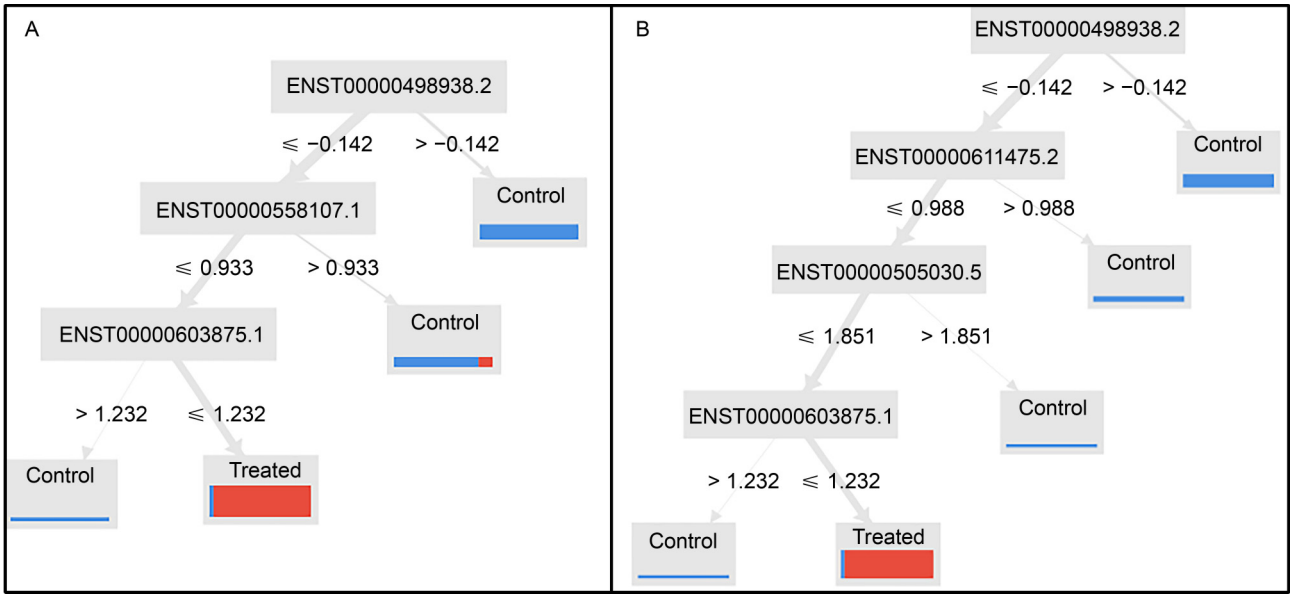


**Figure 4.** Decision Tree induced by (A) Decision Tree Accuracy algorithm, (B) Decision Tree Gain Ratio algorithm on Z-standardized in distinguishing cancer cell line in BC and PC from under treatment ones. The model shows the importance of lncRNAs on sub-clinical cancer classification (control=cancer cell line, treated=healthy cell line). A significant cancer occurrence pattern was discovered where cell lines with a high level of ENST00000498938.2 (>−0.142) could be seen. However, with the low expression of this lncRNA (<−0.142), cancer could be suppressed in almost all cell lines.

lncRNA effect on herbal medicine, with high accuracy of about 91.11% in Z-standardized datasets, followed by RF and DT. The lowest accuracy was 76.67% seen for the GB model. The DL with Rectifier parameter performed the best with the highest sensitivity of 95%, and GB showed the lowest sensitivity (76.67%). Noticeable high specificity (>80%) in Table 1 was observed for DL, RF, and GB that document machine learning models received high power in identifying healthy samples. The DL with Maxout parameter performed the best with the highest specificity of 93.33%, and DT has shown poor performance at 63.62%. The F measure indices were 91.99% in DL with Rectifier parameter, and GB showed the lowest F measure (72.56%). Finally, the noticeable high AUC value (>0.900) was observed for all models, except for GB with the lowest value in the predicting model (0.812). According to Supplementary 12, the classification model's best performance was achieved when DL algorithms were applied to Z-standardized data, and poor performance was shown in GB algorithms, overall.

### ROC curve supporting DL as the most robust and efficient model in predicting cancer-related lncRNA association

The ROC curve shows the relationship between sensitivity (false positive rate) and specificity (true positive rate). It also is one of the best graphical ways to compare prediction models' performances. DL (DL-Maxout) showed the best area under the curve in predicting the true positive rate against the false positive rate (Fig. 5). The results of ROC curve on eight approaches are presented in Supplementary 13.

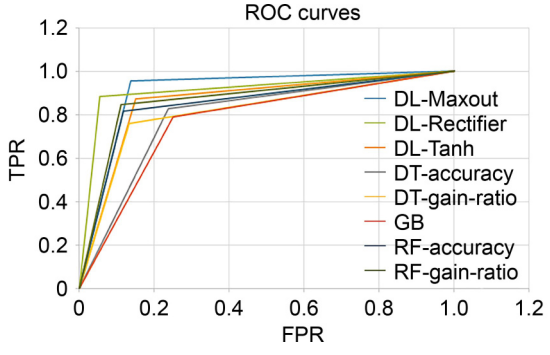


**Figure 5.** Comparing receiver operating characteristic (ROC) curves of machine learning models in BC and PC for predicting novel lncRNA-disease associations, run on Z-Standardized dataset.

## DISCUSSION

Co-expression analysis of herbal treatment in BC and PC in this study highlighted *FUT8-AS1* and *SLC7A5* as the top lncRNAs with the highest numbers of mRNA

associations. Antisense to *SLC7A5* (novel transcript), also known as *RP4-536B24.2*, was found differentially co-expressed with several protein-coding genes expressed in carcinomas. It was significantly positive co-expressed with its sense, *SLC7A5* (*LAT1*), the link of this gene with breast [32,33], and other cancers is well recognized recently. The amino acid transporter, *SLC7A5*, implicates in metabolic changes occurring in tumorigenesis, and its over-expression is reported in many human diseases [34]. It is worth noting that in most of the tumor types, the positive correlation between sense and antisense transcripts is present and implies that they are not regulated independently [35,36]. Also, *IGFBP4* [37], *PEG10* [38], *SUSD3* [39], *SLC3A2* [33,40], *CEBPB* [41], *AKR1B10* [42], *SLC7A5* [43], and *TMEM92* [44] are experimentally illustrated to be positive associated with multiple cancers, including BC and PC. Moreover, through the KEGG pathway of the co-expressed mRNAs of this lncRNA, we found one enrichment pathway "central carbon metabolism in cancer", and according to the GO analysis, they were enriched in the extracellular space and extracellular region part. In the aggregate, these results signify a oncogenic role for "antisense to *SLC7A5*" lncRNA through expression regulation of multiple cancer-related genes in BC and PC (Table 4).

*FUT8-AS1* significantly co-expressed with sixteen mRNAs in tumor and treatment breast and prostate tissue. For the genes highly co-expressed with *FUT8-AS1*, three of the sixteen genes (Table 4), *CEBPB* [41], *STRA6* [45], and *ACKR3* [46], are known cancer-related genes with documented positive association with various cancers. KEGG pathway analyses suggested these mRNAs are enriched in the TNF signaling pathway, pentose phosphate pathway, and prostate cancer. Furthermore, the top GO analysis indicated that the genes of lncRNA *FUT8-AS1* were mainly involved in the cellular carbohydrate biosynthetic process, leucine-zipper, carbohydrate biosynthetic process, protein dimerization activity, and glucose metabolic process. *FUT8-AS1* is associated with mRNAs involved in the most frequently reported metabolic alterations in cancer patients, including resting energy expenditure linked with increased metabolism of carbohydrate, lipid, and protein metabolism, which are all typical of cancer cells [47,48]. Association of lncRNA *FUT8-AS1* to these mRNAs suggests that this lncRNA can serve as a novel biomarker or therapeutic target candidate for patients with BC or PC.

The current study employed attribute weighting models for simultaneous analysis of numerical expression data of long non-coding RNAs with categorical (polynomial) data of cell line, time, and extract. Seven attribute weighting algorithms (Info Gain Ratio, Rule, Chi Squared, Gini Index, Uncertainty, Relief, and Info Gain) mined both numerical expression features, as well as categorical cell line and herbal extract data led to the discovery of transcriptomic signature of response to drug in the level of lncRNAs. Selection of biomarkers based on the commonality/intersection between attribute weighting models increased the confidence on selecting reliable key biomarkers, such as DRAIC. Recently, application of the above-mentioned attribute weighting algorithms has resulted in the development of a universal transcriptomic signature of long-term calorie restriction, independent from categorical variable of tissue (at the levels of hypothalamus, amygdala, pituitary, and adrenal glands). To this end, attribute weighting algorithms mined both numerical data of gene expression, as well as categorical data of tissue type [49]. At the next step, application of standard modeling (Decision Tree, Random Forest, Deep Learning, and Gradient-Boosted

**Table 4 Co-expressed cancer-related genes in BC and PC herbal treatment with two hub lncRNAs and their high correlation in our data**

| lncRNA | Cancer gene | Full name | Correlation coefficient | Correlation | Condition |
|---|---|---|---|---|---|
| *RP4-536B24.2* | *AKR1B10* | Aldo-Keto reductase family 1 member B10 | 0.880587 | Positive | Control |
| | *IGFBP4* | Insulin like growth factor binding protein 4 | 0.901975 | Positive | Control |
| | *PEG10* | Paternally expressed gene 10 protein | 0.857614 | Positive | Control |
| | *SUSD3* | Sushi domain containing protein 3 | 0.921415 | Positive | Control |
| | *TMEM92* | Transmembrane protein 92 | 0.862738 | Positive | Control |
| | *SLC3A2* | Solute carrier family 3 member 2 | 0.847517 | Positive | Treatment |
| | *SLC7A5* | Solute carrier family 7 member 5 | 0.863401 | Positive | Treatment |
| | | | 0.910333 | | |
| *FUT8-AS1* | *CEBPB* | CCAAT enhancer binding protein beta | 0.929094 | Positive | Treatment |
| | *ACKR3* | Atypical chemokine receptor 3 | 0.894196 | Positive | Control |
| | *STRA6* | Stimulated by retinoic acid 6 | 0.845782 | Positive | Control |

Tree) was successful in receiving high sensitivity and specificity.

DRAIC functions in tumorigenesis indicate this lncRNA's ability as a potential target or a prognostic marker in breast cancer treatment [50,51]. Also, RNA-seq analysis revealed up-regulation of DRAIC lncRNA in multiple tumors, including prostate cancer and lung cancer [50]. It is noteworthy that there is significant differential co-expression between DRAIC and *STRA6* (stimulated by retinoic Acid 6) in tumor conditions. According to previous reports, this gene markedly upregulated in human breast and colon tumors [45,52,53], confirming the biomarker potential of DRAIC.

There are some limitations in this study. Attribute weighting models select the most relevant features/attributes to the target labels. They do not indicate whether the attributes are independent of each other. In fact, as the lncRNAs originally come from co-expression analysis, it is possible that they are related to each other and have the same regulatory mechanisms that can lead to the discovery of key regulatory mechanisms in future studies. The employed attribute weighting and predictive models in this study can simultaneously mine both categorical variables (such as gender) along with numerical variables of expression. We used this capability and analyzed cell line type and extract in combination of expression data. Adding gender, age, and the other demographic data could improve the models. However, there are some limitations as the original raw sequencing (fastq) files have been downloaded from public repositories and the demographic information of some samples are not available in public repository.

## CONCLUSION

We identified that 194 mRNAs and 32 lncRNAs were differentially expressed in BC and PC progression. Co-expression network analysis showed that lncRNAs were widely co-expressed with cellular energy metabolism-related genes in BC and PC, implicating the critical roles of these lncRNAs in cell growth and apoptosis regulation. Two novel dysregulated lncRNAs were identified in co-expression networks, and one known cancer-related lncRNA was selected as an essential feature by data mining. This study shows that these selected lncRNAs are independent of cell line and time. The DL model can accurately forecast cancer-related lncRNAs association and outperform other learning models on measurements of model sensitivity and specificity.

## METHODS

### Pipeline of lncRNAs biomarker selection in this study

To select the potential lncRNAs biomarkers of herbal treatment response in breast and prostate cancer, this study employs a set of conservative selection criteria and reliable statistical and bioinformatic pipelines. To announce a lncRNA as potential marker of herbal treatment response, that lncRNA needs to:

(1) Show association/co-expression analysis between that lncRNA and cancer-related mRNAs (genes).

(2) Receive overall high weights in 7 feature selection/attribute weighting algorithms with different statistical backgrounds (including Info Gain Ratio, Rule, Chi Squared, Gini Index, Uncertainty, Relief, and Info Gain). Examining the performance of each lncRNA in representing the herbal treatment by 7 different attribute weighting models increase the robustness and selection confidence as the points of 7 expert systems have been considered. Attribute weighting selected lncRNAs such as DRAIC in this study as key responding lncRNA to herbal treatment.

(3) The above methods have been complemented by application of bioinformatics methods, such as PPI, GO, and pathway analyses.

(4) Literature mining by MedScan, a Natural Language Processing (NLP), implemented in Pathway Studio webtool (Elsevier) have been applied to find the possible link between the lncRNA and their associated genes with cancer.

(5) The performance and repeatability of models and the selected lncRNAs were then evaluated by five-fold cross-validation to examine the statistical performance of a model in analysis of unseen/future data and test how accurately a model may perform in practice.

### Co-expression analysis

Data preparation and differential expression of lncRNAs and mRNAs

All sequence data in fastq format were obtained from GEO in NCBI website and ENA in EMBL-EBI website, originating from nine published studies. We downloaded and reanalyzed the raw RNA-seq data from [54] (prostate cancerous tissue treated by docetaxel chemotherapy) [55], (prostate cancerous cell line treated by Sulforaphane) [55], (prostate cancerous cell line treated by Sulforaphane), (Kyushu University, 2015) (breast cancerous cell line treated by phytoestrogen resveratrol) [56], (breast cancerous cell line treated by

genistein) [56], (breast cancerous cell line treated by liquiritigenin) [56], (breast cancerous cell line treated by S-equol) [57], (breast cancerous cell line treated by genistein) [57], (breast cancerous cell line treated by genistein). The information about accession numbers and sample types were summarized in Table 5.

Briefly, first, CLC genomics workbench software version 9.0.0 (CLC Bio, Aarhus, Denmark, QIAGEN Digital Insights website was used to obtain lncRNA and mRNA gene expression profiles. Second, we performed a differential expression analysis by the R package DEseq2 [58] to examine cancer-associated lncRNAs and mRNAs and to remove low expression values genes. Differentially expressed genes were considered with an adjusted *P*-value of 0.05.

### Normalization

While several methods for RNA normalization have been proposed, we selected the Reads Per Kilobase per Million of reads (RPKM) normalization methods. The RPKM value minimizes the effect of gene length bias when relating expression levels across genes, whereby longer genes will be sequenced deeper than shorter genes, so it was used for the gene co-expression network [59,60]. This step was performed by CLC Genomics.

### Co-expression analysis of cancer genes and lncRNAs

The co-expression network was constructed using the "rcorr" function of library Hmisc in the R environment (Harrell, 2006). The normalized expression data of significant coding-genes and lncRNAs were used as input. After calculating the Pearson correlation coefficient (PCC), the most popular co-expression measure, we used the *r* value to calculate the PCC correlation coefficient between the expression of the lncRNA and protein-coding mRNAs across all samples. We considered the correlation values are higher than 0.8 and smaller than −0.8 ($-0.8>r>0.8$), well below the significance threshold of 0.5 (*P*-value < 0.5).

## Prediction of the potential roles of lncRNAs

### Construction of PPI network

To investigate the molecular function of lncRNA, we analyzed the interaction co-expressed mRNAs by PPIs. STRING version 11.0 in STRING website, an online software, was used to search the interaction relationships of DEGs with medium confidence of 0.400 as the product criterion [61].

### Construction of literature mining based network

Differentially co-expressed mRNAs with lncRNAs were used as input to construct interaction networks based on text mining using NLP (natural language processing) by Pathway Studio Webtool (Elsevier), as previously described [62,63]. Gene Ontology database was used to identify the cellular location and class of protein, such as transcription factor, ligand, receptor.

### Enrichment analysis and KEGG pathway

To gain a good understanding of the functions of differentially expressed lncRNAs, we performed KEGG and Gene Ontology term enrichment by using the set of co-expressed mRNAs [64]. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) and the comparative GO web tool was used. Biological process (BP), cellular component (CC), and molecular function (MF) are three domains of GO. *P*-value $\leqslant$ 0.05 is recommended to denote the significance of the pathway correlations and GO term enrichment.

**Table 5 The selected original datasets of prostate and breast cancer herbal treatment**

| Simple source | Platforms | Tissue/cell | Herbal medicine | No. of samples | | Accession number |
|---|---|---|---|---|---|---|
| | | | | Treatment | Control | |
| *In vivo* | Illumina HiSeq 2000 (*Homo sapiens*) | Pca | Docetaxel | 6 | 6 | GSE51005 |
| *In vivo* | Illumina HiSeq 2000 (*Homo sapiens*) | LNCaP | Sulforaphane | 3 | 3 | GSE48812 |
| *In vivo* | Illumina HiSeq 2000 (*Homo sapiens*) | PC-3 | Sulforaphane | 3 | 3 | |
| *In vivo* | Illumina Genome Analyzer IIx (*Homo sapiens*) | MCF-7 | Phytoestrogen resveratrol | 2 | 2 | PRJDB1992 |
| *In vivo* | Illumina HiSeq 2000 (*Homo sapiens*) | MCF-7 | S-equol | 2 | 2 | GSE56066 |
| *In vivo* | Illumina HiSeq 2000 (*Homo sapiens*) | MCF-7 | Genistein | 2 | 2 | |
| *In vivo* | Illumina HiSeq 2000 (*Homo sapiens*) | MCF-7 | Liquiritigenin | 2 | 2 | |
| *In vivo* | Illumina HiSeq 2000 (*Homo sapiens*) | ECC-1 | Genistein | 2 | 2 | GSE38234 |
| *In vivo* | Illumina HiSeq 2000 (*Homo sapiens*) | T-47D | Genistein | 2 | 2 | |

### Network visualization

The Cytoscape software version 3.8.0 downloaded from cytoscape website (old versions) was used to illustrate network visualization. In this representation, each mRNA or lncRNA corresponded to a node, and when an edge connected mRNA and lncRNA, it indicated a strong correlation (*i.e.*, either positive or negative). The font size of the node was determined by connectivity.

## Data mining (machine learning and prediction models)

### Preparation of dataset

The original dataset had 44 recorded samples (or rows) with 24 differential co-expressed lncRNAs and four variables [Cell line, Time, Extract, and Concentration]. To investigate the best classification model and test whether the co-expressed lncRNAs are independent of cell line, time, extract, and concentration, we used four approaches. At the first approach, attribute weighting and classification models were run with all variables: Cell line, Time, Extract, Concentration. In the second approach, models were run with three variables: Cell line, Time, Concentration. In the third approach, models were run with three variables: Cell line, Time, Extract. In the end, they were run with two variables: Cell line, Time.

Also, as the dataset used for predicting cancer-related lncRNAs was collected from a limited number of humans, we transformed the original dataset to Z-standardization to create a reliable base for generalization of this study's findings (for each feature, subtracting the mean and dividing by the standard deviation).

Finally, all the datasets (four approaches for original datasets and four for transformed datasets) were imported into the Rapid Miner software (Rapid Miner 5.0.001, Rapid-I GmbH, Stochumer Str. 475, 44,227 Dortmund, Germany) to apply attribute weighting and prediction models (DL, GB, RF, and DT) to these data separately.

### Attribute weighting

Attribute weighting is a crucial aspect when modeling multi-attribute decision analysis problems. By reducing the size of attributes, attribute weighting models generate a more manageable set of attributes for modeling [65]. Seven attribute weighting algorithms, including Info Gain Ratio, Rule, Chi Squared, Gini Index, Uncertainty, Relief, and Info Gain that were able to mine both categorical and numerical data were employed, as previously described [49,66,67]. Weights for each model were normalized as values between 0 and 1; showing each attribute's importance to the target attribute [31]. This analysis can address whether the developed gene signature is cell line, time, extract, and concentration − independent or dependent. In other words, this analysis finds whether those categorical variables are critical feature in the treatment cancer process or not.

### Bayesian Estimation Supersedes the t-test

Bayesian Estimation Supersedes the t-test was used for further evaluation of key discovered herbal treatment responding lncRNAs, such as DRAIC. BEST applies Bayesian model for estimating the difference in means between two groups and yields a probability distribution over the difference [68,69]. Based on the distribution, the mean credible value can be considered as the best guess of the actual difference and the 95% Highest Density Interval (HDI) as the range where the actual difference has 95% credibility.

### Prediction models

Prediction models are an effective and reliable technique to determine the relationship between lncRNAs and control-treated samples, which can help us understand the mechanism underlying a disease mechanism and accelerate biomarkers discovery.

DL model uses multiple layers to progressively extract higher-level features from the raw input. The network can contain many hidden layers consisting of neurons with Tanh, Rectifier, and Maxout activation functions. This employed model can be applied to large-scale data and learn complex non-linear relationships through mini-batch stochastic gradient descent and non-linear activation function [23,70].

The Gradient Boosted Tree (GBT) model is used in regression combination and classification tree models, such as Decision tree models. This model improves prediction power results through progressively improving estimations [31].

The RF classifier is an ensemble machine learning method based on the voting model of all possible tree induction. This operator generates a prediction result in the form of several random trees [31,71].

DT is a well-known and widely discussed technique for classification and prediction. This algorithm generates re iscursive partitioning (classification and regression trees) and repeatedly splits the attribute values used to extract the potential patterns between the

target and regular variables [31]. We performed DT and RF with two different criteria (Gain Ratio and Accuracy) and used 5-fold cross-validation to acquire the mean accuracy.

Five-fold cross-validation was used to compute each prediction model's accuracy. The dataset is randomly partitioned into five equal-sized subsamples to perform five-fold cross-validation. Prediction models were tested on four sub-sample sets, and the remaining subsample was used as evaluating data. The procedure was repeated five times and the average accuracy, AUC, receiver operating characteristic curve (ROC), F measure, sensitivity, and specificity of five runs was calculated by dividing the percentage of correct predictions over the total number of examples.

### SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at https://doi.org/10.15302/J-QB-023-0333.

### ACKNOWLEGEMENTS

This research has received no external funding.

### COMPLIANCE WITH ETHICS GUIDELINES

**Conflicts of interest** The authors Elham Dalalbashi Esfahani, Esmaeil Ebrahimie, Ali Niazi and Manijeh Mohammadi Dehcheshmeh declare that they have no competing interests.

The article does not contain any human or animal subjects performed by any of the authors.



## REFERENCES

1. Ekor, M. (2014) The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. Front. Pharmacol, 4, 177
2. Kuruppu, A. I., Paranagama, P. and Goonasekara, C. L. (2019) Medicinal plants commonly used against cancer in traditional medicine formulae in Sri Lanka. Saudi Pharm. J. 27, 565–573
3. Macharia, J.M., Mwangi, R.W., Rozmann, N., Zsolt, K., Varjas, T., Uchechukwu, P.O., Wagara, I.N. and Raposa, B.L. (2022) Medicinal plants with anti-colorectal cancer bioactive compounds: potential game-changers in colorectal cancer management. Biomed Pharmacother. 153, 113383
4. Kooti, W., Servatyari, K., Behzadifar, M., Asadi-Samani, M., Sadeghi, F., Nouri, B. and Zare Marzouni, H. (2017) Effective medicinal plant in cancer treatment, part 2: review study. J. Evid. Based Complementary Altern. Med., 22, 982−995
5. Iqbal, J., Abbasi, B.A., Mahmood, T., Kanwal, S., Ali, B., Shah, S.A. and Khalil, A.T. (2017) Plant-derived anticancer agents: a green anticancer approach. Asian Pac. J. Trop. Med., 7, 1129–1150
6. López-Otín, C. and Diamandis, E. P. (1998) Breast and prostate cancer: an analysis of common epidemiological, genetic, and biochemical features. Endocr. Rev., 19, 365−396
7. Wu, W., Wagner, E.K., Hao, Y., Rao, X., Dai, H., Han, J., Chen, J., Storniolo, A.M.V., Liu, Y. and He, C. (2016) Tissue-specific co-expression of long non-coding and coding RNAs associated with breast cancer. Sci. Rep., 6, 32731
8. Ren, Z.-J., Cao, D.-H., Zhang, Q., Ren, P.-W., Liu, L.-R., Wei, Q., Wei, W.-R. and Dong, Q. (2019) First-degree family history of breast cancer is associated with prostate cancer risk: a systematic review and meta-analysis. BMC Cancer, 19, 871
9. Zheng, Y., Xu, Q., Liu, M., Hu, H., Xie, Y., Zuo, Z. and Ren, J. (2019) LnCAR: a comprehensive resource for lncRNAs from cancer arrays. Cancer Res., 79, 2076−83
10. Beylerli, O., Gareev, I., Sufianov, A., Ilyasova, T. and Guang, Y. (2022) Long noncoding RNAs as promising biomarkers in cancer. Noncoding RNA Res., 7, 66−70
11. Wang, J., Shen, Y.-C., Chen, Z.-N., Yuan, Z.-C., Wang, H., Li, D.-J., Liu, K. and Wen, F.-Q. (2019) Microarray profiling of lung long non-coding RNAs and mRNAs in lipopolysaccharide-induced acute lung injury mouse model. Biosci. Rep., 39, BSR20181634
12. Silva, A. M., Moura, S. R., Teixeira, J. H., Barbosa, M. A., Santos, S. G. and Almeida, M. I. (2019) Long noncoding RNAs: a missing link in osteoporosis. Bone Res., 7, 10
13. Zhou, S., He, Y., Yang, S., Hu, J., Zhang, Q., Chen, W., Xu, H., Zhang, H., Zhong, S., Zhao, J., *et al*. (2018) The regulatory roles of lncRNAs in the process of breast cancer invasion and metastasis. Biosci. Rep., 38, BSR20180772
14. Cimadamore, A., Gasparrini, S., Mazzucchelli, R., Doria, A., Cheng, L., Lopez-Beltran, A., Santoni, M., Scarpelli, M. and Montironi, R. (2017) Long non-coding RNAs in prostate cancer with emphasis on second chromosome locus associated with prostate-1 expression. Front. Oncol., 7, 305
15. Yang, W., Li, Y., Song, X., Xu, J. and Xie, J. (2017) Genome-wide analysis of long noncoding RNA and mRNA co-expression profile in intrahepatic cholangiocarcinoma tissue by RNA sequencing. Oncotarget, 8, 26591−26599
16. Marttila, S., Chatsirisupachai, K., Palmer, D. and de Magalhaes, J.P. (2020) Ageing-associated changes in the expression of lncRNAs in human tissues reflect a transcriptional modulation in ageing pathways. Mech. Ageing. Dev., 185, 111177

 

17. Cogill, S.B. and Wang, L. (2014) Co-expression network analysis of human lncRNAs and cancer genes. Cancer Inform. 13, 49–59
18. Guttman, M. and Rinn, J. L. (2012) Modular regulatory principles of large non-coding RNAs. Nature, 482, 339−346
19. Sharma, A. and Capobianco, E. (2022) Non-coding RNAs are brokers in breast cancer interactome networks and add discrimination power between subtypes. J. Clin. Med., 11, 2103
20. Rinn, J. L. and Chang, H. Y. (2012) Genome regulation by long noncoding RNAs. Annu. Rev. Biochem., 81, 145−166
21. Ruan, X., Li,P., Chen, Y., Shi, Y., Pirooznia, M., Seifuddin, F., Suemizu, H., Ohnishi, Y., Yoneda, N., Nishiwaki, M., *et al*. (2020) *In vivo* functional analysis of nonconserved human lncRNAs associated with cardiometabolic traits. Nat. Commun., 11, 45
22. Han, S., Liang, Y., Li, Y., and Du, W. (2016) Long noncoding RNA identification: comparing machine learning based tools for long noncoding transcripts discrimination. BioMed Res. Int. 2016, 1–14
23. Ammunét, T., Wang, N., Khan, S. and Elo, L. L. (2022) Deep learning tools are top performers in long non-coding RNA prediction. Brief. Funct. Genomics, 21, 230−241
24. Zhu, B., Xu, M., Shi, H., Gao, X. and Liang, P. (2017) Genome-wide identification of lncRNAs associated with chlorantraniliprole resistance in diamondback moth *Plutella xylostella* (L. ). BMC Genomics, 18, 380
25. Zheng, H., Brennan, K., Hernaez, M. and Gevaert, O. (2019) Benchmark of long non-coding RNA quantification for RNA sequencing of cancer samples. Gigascience, 8, giz145
26. Sun, H., Huang, Z., Sheng, W. and Xu, M.-D. (2018) Emerging roles of long non-coding RNAs in tumor metabolism. J. Hematol. Oncol., 11, 106
27. Lin, W., Zhou, Q., Wang, C. Q., Zhu, L., Bi, C., Zhang, S., Wang, X. and Jin, H. (2020) LncRNAs regulate metabolism in cancer. Int. J. Biol. Sci., 16, 1194−1206
28. Wajant, H. (2009) The role of TNF in cancer. Results Probl. Cell Differ. 49,1–15
29. Boroughs, L. K. and DeBerardinis, R. J. (2015) Metabolic pathways promoting cancer cell survival and growth. Nat. Cell Biol., 17, 351−359
30. Hao, Y., Wu, W., Shi, F., Dalmolin, R. J., Yan, M., Tian, F., Chen, X., Chen, G. and Cao, W. (2015) Prediction of long noncoding RNA functions with co-expression network in esophageal squamous cell carcinoma. BMC Cancer, 15, 168
31. Ebrahimi, M., Mohammadi-Dehcheshmeh, M., Ebrahimi, E. and Petrovski, K. R. (2019) Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: deep learning and gradient-boosted trees outperform other models. Comp. Biol. Med., 114, 103456
32. El Ansari, R., Craze, M.L., Miligy, I., Diez-Rodriguez, M., Nolan, C.C., Ellis, I.O., Rakha, E.A. and Green, A.R. (2018) The amino acid transporter SLC7A5 confers a poor prognosis in the highly proliferative breast cancer subtypes and is a key therapeutic target in luminal B tumours. Breast Cancer Res., 20, 21
33. Alfarsi, L. H., El-Ansari, R., Craze, M. L., Masisi, B. K., Mohammed, O. J., Ellis, I. O., Rakha, E. A. and Green, A. R. (2020) Co-expression effect of SLC7A5/SLC3A2 to predict response to endocrine therapy in oestrogen-receptor-positive breast cancer. Int. J. Mol. Sci., 21, 1407
34. Scalise, M., Galluccio, M., Console, L., Pochini, L. and Indiveri, C. (2018) The human SLC7A5 (LAT1): the intriguing histidine/large neutral amino acid transporter and its relevance to human health. Front Chem., 6, 243−243
35. Boque-Sastre, R., Moura, M. C., Gomez, A., Guil, S. and Esteller, M. (2017) Abstract 3483: genome-wide analysis of the antisense transcriptome in cancer. Cancer Res., 77, 3483
36. Watanabe, Y., Numata, K., Murata, S., Osada, Y., Saito, R., Nakaoka, H., Yamamoto, N., Watanabe, K., Kato, H., Abe, K. *et al*. (2010) Genome-wide analysis of expression modes and DNA methylation status at sense-antisense transcript loci in mouse. Genomics, 96, 333–341
37. Durai, R., Davies, M., Yang, W., Yang, S.Y., Seifalian, A., Goldspink, G., Winslet, M. (2006) Biology of insulin-like growth factor binding protein-4 and its role in cancer (review). Int. J. Oncol., 28, 1317−1325
38. Li, X., Xiao, R., Tembo, K., Hao, L., Xiong, M., Pan, S., Yang, X., Yuan, W., Xiong, J. and Zhang, Q. (2016) PEG10 promotes human breast cancer cell proliferation, migration and invasion. Int. J. Oncol., 48, 1933−1942
39. Yu, Z., Jiang, E., Wang, X., Shi, Y., Shangguan, A. J., Zhang, L. and Li, J. (2015) Sushi domain-containing protein 3: a potential target for breast cancer. Cell Biochem. Biophys., 72, 321−324
40. Yang, Y., Toy, W., Choong, L. Y., Hou, P., Ashktorab, H., Smoot, D. T., Yeoh, K. G. and Lim, Y. P. (2012) Discovery of SLC3A2 cell membrane protein as a potential gastric cancer biomarker: implications in molecular imaging. J. Proteome Res., 11, 5736−5747
41. Sankpal, N.V., Moskaluk, C.A., Hampton, G.M. and Powell, S.M. (2006) Overexpression of CEBPβ correlates with decreased TFF1 in gastric cancer. Oncogene. 7, 643–649
42. Fukumoto, S., Yamauchi, N., Moriguchi, H., Hippo, Y., Watanabe, A., Shibahara, J., Taniguchi, H., Ishikawa, S., Ito, H., Yamamoto, S., *et al*. (2005) Overexpression of the aldo-keto reductase family protein AKR1B10 is highly correlated with smokers' non-small cell lung carcinomas. Clin. Cancer Res., 11, 1776−1785
43. Bhutia, Y. D., Babu, E., Ramachandran, S. and Ganapathy, V. (2015) Amino Acid transporters in cancer and their relevance to "glutamine addiction": novel targets for the design of a new class of anticancer drugs. Cancer Res., 75, 1782−1788
44. Morgan, R., Feng, G. and Pandha, H. S. (2013) Abstract A193: transmembrane protein TMEM92 as a novel target in prostate cancer. Mol. Cancer Ther., 12, A193
45. Szeto, W., Jiang, W., Tice, D. A., Rubinfeld, B., Hollingshead, P. G., Fong, S. E., Dugger, D. L., Pham, T., Yansura, D. G., Wong, T. A., *et al*. (2001) Overexpression of the retinoic acid-responsive gene *Stra6* in human cancers and its synergistic induction by Wnt-1 and retinoic acid. Cancer Res., 61, 4197−4205
46. Behnam Azad, B., Lisok, A., Chatterjee, S., Poirier, J. T., Pullambhatla, M., Luker, G. D., Pomper, M. G. and Nimmagadda, S. (2016) Targeted imaging of the atypical

 

chemokine receptor 3 (ACKR3/CXCR7) in human cancer xenografts. J. Nucl. Med., 57, 981−988
47. Fadaka, A., Ajiboye, B., Ojo, O., Adewale, O., Olayide, I., Emuowhochere, R.J. (2017) Biology of glucose metabolization in cancer cells. J. Oncol. Sci., 3, 45–51
48. Kareva, I. (2022) Understanding metabolic alterations in cancer cachexia through the lens of exercise physiology. Cells, 11, 2317
49. Govic, A., Nasser, H., Levay, E. A., Zelko, M., Ebrahimie, E., Mohammadi Dehcheshmeh, M., Kent, S., Penman, J. and Hazi, A. (2022) Long-term calorie restriction alters anxiety-like behaviour and the brain and adrenal gland transcriptomes of the ageing male rat. Nutrients, 14, 4670
50. Zhao, D. and Dong, J.-T. (2018) Upregulation of long non-coding RNA DRAIC correlates with adverse features of breast cancer. Noncoding RNA, 4, 39
51. Yao, Q., Zhang, X. and Chen, D. (2022) The emerging potentials of lncRNA DRAIC in human cancers. Front. Oncol., 12, 867670
52. Berry, D. C., Levi, L. and Noy, N. (2014) Holo-retinol-binding protein and its receptor STRA6 drive oncogenic transformation. Cancer Res., 74, 6341−6351
53. Muñiz-Hernández, S., Velázquez-Fernández, J. B., Díaz-Chávez, J., Mondragón-Fonseca, O., Mayén-Lobo, Y., Ortega, A., López-López, M. and Arrieta, O. (2020) STRA6 polymorphisms are associated with EGFR mutations in locally-advanced and metastatic non-small cell lung cancer patients. Front. Oncol., 10, 579561
54. Rajan, P., Stockley, J., Sudbery, I. M., Fleming, J. T., Hedley, A., Kalna, G., Sims, D., Ponting, C. P., Heger, A., Robson, C. N., *et al*. (2014) Identification of a candidate prognostic gene signature by transcriptome analysis of matched pre- and post-treatment prostatic biopsies from patients with advanced prostate cancer. BMC Cancer, 14, 977
55. Beaver, L. M., Buchanan, A., Sokolowski, E. I., Riscoe, A. N., Wong, C. P., Chang, J. H., Löhr, C. V., Williams, D. E., Dashwood, R. H. and Ho, E. (2014) Transcriptome analysis reveals a dynamic and differential transcriptional response to sulforaphane in normal and prostate cancer cells and suggests a role for Sp1 in chemoprevention. Mol. Nutr. Food Res., 58, 2001−2013
56. Gong, P., Madak-Erdogan, Z., Li, J., Cheng, J., Greenlief, C. M., Helferich, W., Katzenellenbogen, J. A. and Katzenellenbogen, B. S. (2014) Transcriptomic analysis identifies gene networks regulated by estrogen receptor α (ERα) and ERβ that control distinct effects of different botanical estrogens. Nucl. Recept. Signal., 12, e001
57. Gertz, J., Reddy, T. E., Varley, K. E., Garabedian, M. J. and Myers, R. M. (2012) Genistein and bisphenol A exposure cause estrogen receptor 1 to bind thousands of sites in a cell type-specific manner. Genome Res., 22, 2153−2162
58. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biol., 15, 550
59. Howard, J. T., Ashwell, M. S., Baynes, R. E., Brooks J. D., Yeatts J. L. and Maltecca, C. (2017) Gene co-expression network analysis identifies porcine genes associated with variation in metabolizing fenbendazole and flunixin meglumine in the liver. Sci. Rep. 7, 1357
60. Johnson, K. A. and Krishnan, A. (2022) Robust normalization and transformation techniques for constructing gene co-expression networks from RNA-seq data. Genome Biol. 23, 1
61. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., *et al*. (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res., 47, D607−D613
62. Alanazi, I. O., Al Shehri, Z. S., Ebrahimie, E., Giahi, H. and Mohammadi-Dehcheshmeh, M. (2019) Non-coding and coding genomic variants distinguish prostate cancer, castration-resistant prostate cancer, familial prostate cancer, and metastatic castration-resistant prostate cancer from each other. Mol. Carcinog., 58, 862−874
63. Alanazi, I. O., AlYahya, S. A., Ebrahimie, E. and Mohammadi-Dehcheshmeh, M. (2018) Computational systems biology analysis of biomarkers in lung cancer; unravelling genomic regions which frequently encode biomarkers, enriched pathways, and new candidates. Gene, 659, 29−36
64. Fruzangohar, M., Ebrahimie, E., Ogunniyi, A. D., Mahdi, L. K., Paton, J. C. and Adelson, D. L. (2013) Comparative GO: a web application for comparative Gene Ontology and Gene Ontology-based gene selection in bacteria. PLoS One, 8, e58759
65. Ebrahimi, M., Lakizadeh, A., Agha-Golzadeh, P., Ebrahimie, E. and Ebrahimi, M. (2011) Prediction of thermostability from amino acid attributes by combination of clustering with attribute weighting: a new vista in engineering enzymes. PLoS One, 6, e23146
66. Ghasemi, E., Ebrahimi, M. and Ebrahimie, E. (2022) Machine learning models effectively distinguish attention-deficit/hyper-activity disorder using event-related potentials. Cogn. Neurody-namics, 16, 1335−1349
67. Ebrahimie, E., Zamansani, F., Alanazi, I. O., Sabi, E. M., Khazandi, M., Ebrahimi, F., Mohammadi-Dehcheshmeh, M. and Ebrahimi, M. (2021) Advances in understanding the specificity function of transporters by machine learning. Comput. Biol. Med., 138, 104893
68. Meredith, M. and Kruschke, J.K. (2021) Bayesian Estimation Supersedes the t-test (computer software manual)
69. Kruschke, J. K. (2013) Bayesian estimation supersedes the t test. J. Exp. Psychol. Gen., 142, 573−603
70. Hu, J., Gao, Y., Li, J. and Shang, X. (2019) Deep learning enables accurate prediction of interplay between lncRNA and disease. Front. Genet., 10
71. Yao, D., Zhan, X., Zhan, X., Kwoh, C. K., Li, P., and Wang, J. (2020) A random forest based computational model for predicting novel lncRNA-disease associations. BMC Bioinf. 21, 126

 