

Research Article

Machine Learning Supports Long Noncoding RNAs as Expression Markers for Endometrial Carcinoma

Ana Carolina Mello,^{1,2} Martiela Freitas,^{1,2,3} Laura Coutinho,^{1,2,4} Tiago Falcon,^{1,2} and Ursula Matte ^{1,2,5}

¹Bioinformatics Core, Experimental Research Center, Hospital de Clínicas de Porto Alegre, Porto Alegre 90035-903, Brazil

²Gene Therapy Center, Experimental Research Center, Hospital de Clínicas de Porto Alegre, Porto Alegre 90035-903, Brazil

³Post-Graduation Program on Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul, Porto Alegre 91501-970, Brazil

⁴Undergraduation Program on Biotechnology/Bioinformatics, Universidade Federal do Rio Grande do Sul, Porto Alegre 91501-970, Brazil

⁵Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre 91501-970, Brazil

Correspondence should be addressed to Ursula Matte; umatte@hcpa.edu.br

Received 21 November 2019; Accepted 17 December 2019; Published 10 January 2020

Academic Editor: Rosaria Scudiero

Copyright © 2020 Ana Carolina Mello et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Uterine corpus endometrial carcinoma (UCEC) is the second most common type of gynecological tumor. Several research studies have recently shown the potential of different ncRNAs as biomarkers for prognostics and diagnosis in different types of cancers, including UCEC. Thus, we hypothesized that long noncoding RNAs (lncRNAs) could serve as efficient factors to discriminate solid primary (TP) and normal adjacent (NT) tissues in UCEC with high accuracy. We performed an *in silico* differential expression analysis comparing TP and NT from a set of samples downloaded from the Cancer Genome Atlas (TCGA) database, targeting highly differentially expressed lncRNAs that could potentially serve as gene expression markers. All analyses were performed in R software. The receiver operator characteristics (ROC) analyses and both supervised and unsupervised machine learning indicated a set of 14 lncRNAs that may serve as biomarkers for UCEC. Functions and putative pathways were assessed through a coexpression network and target enrichment analysis.

1. Introduction

Uterine corpus endometrial carcinoma (UCEC) is the second most common type of gynecological tumor, in either developed or underdeveloped countries [1]. According to the epidemiologic data from the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO), UCEC comprises 4.8% of cancer incidence worldwide and 2.1% of cancer-related mortality rate [2]. This cancer originates at glandular epithelial cells of the endometrium, which is the mucous layer of the inner uterus [3] and is specified according to clinical and endocrine characteristics by the International Federation of Gynecology and Obstetrics (FIGO): type I carcinomas are estrogen

dependent and associated with endometrial hyperplasia, whereas type II carcinomas are estrogen independent and associated with endometrial atrophy [3].

Obesity, aging, early menarche, late menopause, nulliparity, breast cancer, and diabetes mellitus history are some of the endogenous risk factors for developing the disease [4]. Other risk factors include dietetic factors, tamoxifen (<https://pubchem.ncbi.nlm.nih.gov/compound/2733526>) therapy [5], radiotherapy, and high levels of estrogen [6]. The standard treatment is surgery to remove fallopian tubes and the ovary, which is effective at most cases when treating stage I disease [7]. For advanced stages, surgery followed by treatments such as radiation therapy, chemotherapy, or a combination of both is the best treatment.

Several researches have recently shown the potential of different noncoding RNAs (ncRNAs) as biomarkers for prognostics and diagnosis in different types of cancers [3, 8–10], including UCEC [11, 12] although, currently, no biomarker is consistently used for those purposes in UCEC [13]. Long noncoding RNAs (lncRNAs) are a class of ncRNAs that are at least 200 base pairs long and have multiple functions, albeit these cannot be inferred by their sequences. Most of the lncRNAs are long intergenic (lincRNAs) [14], which are sequences that do not overlap messenger RNAs (mRNAs) [15]. Another class of lncRNAs are antisense RNAs (asRNAs), which are transcribed from the opposite strand of the sense transcripts of either protein-coding or non-protein-coding genes [16]. Some have been associated with chromatin-modifying complexes that confer either repressive or activating modifications [17, 18]; however, the specific functions of the majority of lncRNAs are still unknown.

In UCEC, the lncRNA *HOTAIR* was found to be overexpressed [19] and apparently contributes to the cisplatin-induced resistance by inhibiting autophagy [20]. The silencing of *in vivo* expression of *HOTAIR* suppressed significantly the endometrial tumorigenesis, leading to smaller tumors [21, 22]. Another lncRNA, *MALAT1*, is known to be overexpressed during endometrial hyperplasia and also during early carcinoma stages; however, its expression is significantly lower in advanced stages, as well as during metastasis [23].

Given their role in modulating UCEC progression, we hypothesized that lncRNAs could efficiently discriminate solid primary tumor (TP) and adjacent normal tissue (NT) with high level of confidence. To test our hypothesis, we performed an *in silico* differential expression analysis comparing TP and NT from a set of patients, targeting highly differentially expressed lncRNAs that could potentially serve as biomarkers for UCEC. These potential biomarkers were tested using both supervised and unsupervised machine learning. Also, their relationship with other genes was verified by gene coexpression networks and enrichment analysis. From our knowledge, previous works considered only the mRNAs [24, 25], focused in drivers of many cancers (not exclusive) [24], subtypes [26], or other types of uterine carcinoma (nonendometrial) [27]. Similar approaches, based on support vector machine (SVM), have been recently applied to detect biomarkers and key network elements in other tumors [28–31].

2. Results and Discussion

From the 46 selected samples (TP and NT from 23 patients) of the total RNA data, we obtained a total of 8,700 transcripts, after normalization and filtering, that went through the differential expression analysis. In order to assess the relationships with other regulatory RNAs and pathways, we also analyzed miRNA data, from which we selected 42 samples (TP and NT from 21 patients, all featured in the total RNA data as well), and 1,881 miRNAs were attained.

The total RNA normalization and filtering of the data guaranteed that the outliers had been removed, and the

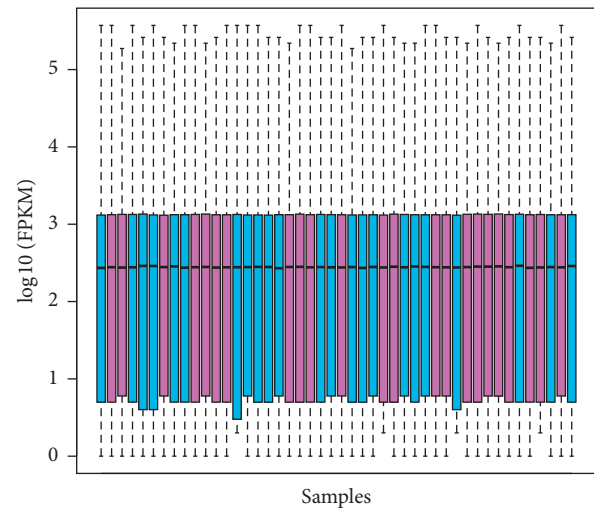


FIGURE 1: Boxplot after normalization of total RNA samples (training set data). Primary tumor samples (TP) (sky blue); adjacent tumor samples (TP) (purple). Black line inside the boxes indicates the median position. Dotted lines indicate the expression deviations.

samples were well normalized (Figure 1), meaning a reduction in the expression deviation and the obtention of an acceptable false positive rate [32, 33]. Both steps indicate that all the differences further found on the expression of the transcripts in TP when compared to NT were most likely caused by the different environments. Considering a $|\logFC| \geq 5$ and a $FDR < 0.01$ cutoff, a total of 191 transcripts were differentially expressed (Figure 2). Additionally, 67 of them were upregulated in tumor. We chose high values of fold change to ensure that we selected genes with high potential to discriminate the groups and serve as biomarkers.

Seventeen lncRNAs were significantly differentially expressed and were chosen as the main candidates to correctly distinguish TP from NT samples (Table 1). Sixteen of them were downregulated in tumor, while only one of them, *LINC01376*, was upregulated, indicating that its expression is essential for maintaining the tumor environment. In order to predict whether the candidates were good for classification models, we plotted AUC-ROC curves for each of them. Five candidates had an AUC greater than 0.7, thus demonstrating high confidence results: *LINC01376*, *BRWD1-AS1*, *LINC00244*, *LINC02475*, and *ZNF667-AS1* (Figure 3).

Three of the top five candidates are lincRNAs. The full range of biological function of lincRNAs remains to be deeply explored [34]. Dysregulation of the expression of lincRNAs may be pervasive in human cancers and drives cancer development and progression [35]. Previous researches demonstrated that there are SNPs on *LINC01376* associated with breast cancer [36, 37]. Notably, it was the only lncRNA upregulated in our analysis with a \logFC of 5.88, and it scored the best AUC (0.913) among all candidates.

lncRNAs may be a product of a partially or fully complementary region of a protein-coding gene and may act as a *cis* regulatory lncRNA, in accordance with the sequence

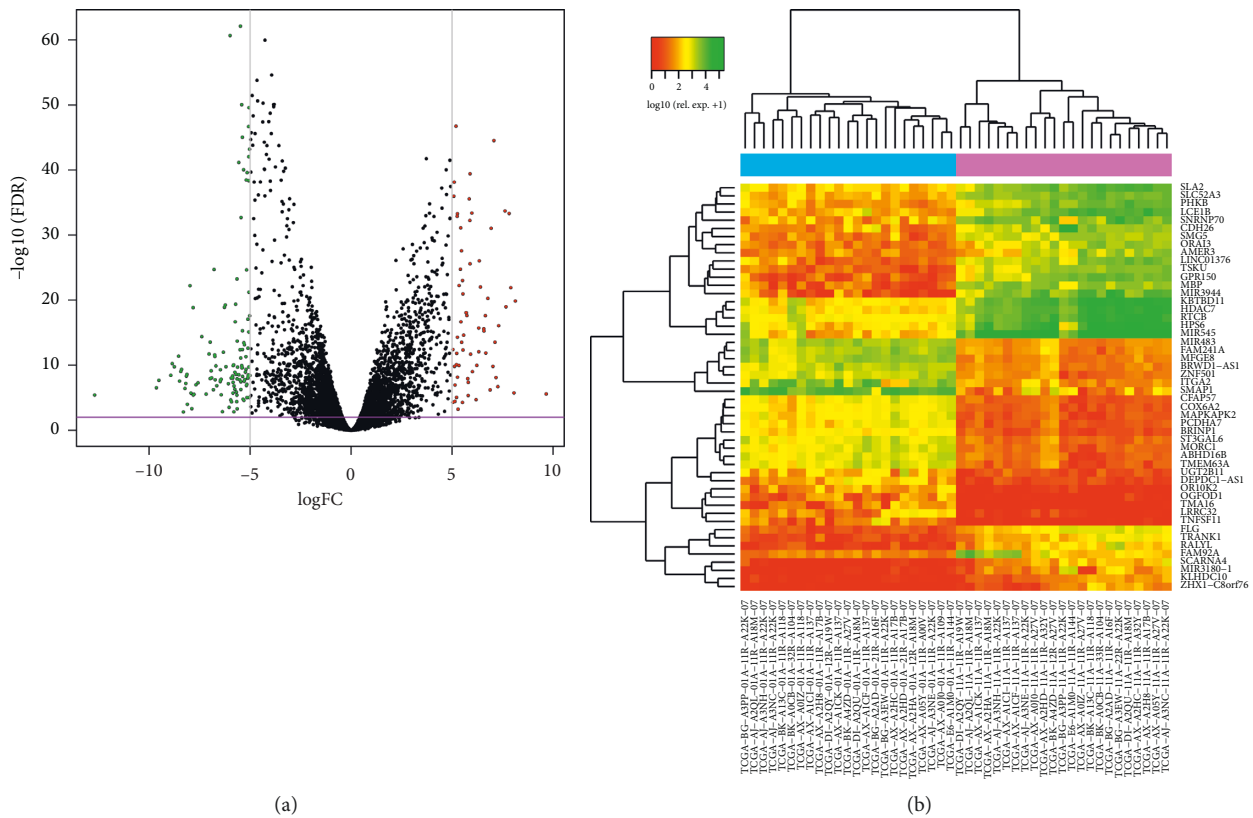


FIGURE 2: Differentially expressed genes highlighted. (a) Volcano plot of the differential expression analysis of total RNA in primary tumor (TP) compared with adjacent tissue (NT). TP upregulated transcripts (67 genes) with log fold change (\log_{FC}) > 5 (red). TP downregulated transcripts with $\log_{\text{FC}} < -5$ (124 genes) (green). Horizontal purple line indicates the $-\log_{10}(\text{FDR}) = 2$ ($\text{FDR} = 0.01$) cutoff. Grey vertical lines indicate $\log_{\text{FC}} = -5$ and $\log_{\text{FC}} = 5$ cutoffs. (b) Heat map of the expression of the 50 most differentially expressed transcripts. Dendrogram: the clustering of TP samples (sky blue); the clustering of NT samples (purple). Top left: colorkey for the heat expression quantification.

complementarity (antisense lncRNA), or may be a lincRNA, acting as a *trans* regulatory element. Also, lncRNAs can connect to DNA, regulating the transcription process, or bind to proteins affecting their stability ([38]). Antisense lncRNAs are known to regulate the transcription output RNA in mammals, affecting the mRNA expression [39, 40] and stabilizing the mRNA [40]. Some antisense lncRNAs may act on their neighboring coding genes, reducing the complementary mRNA expression [41, 42], while others clearly upregulate the expression of the corresponding mRNA and protein [43].

Members of the *BRWD1* gene family are involved in cellular processes such as cell cycle progression, signal transduction, gene regulation, and apoptosis [44]. *BRWD1-AS1* is downregulated in UCEC, with a \log_{FC} of approximately -5.05 , and scored the third highest AUC (0.820). It is possible that *BRWD1-AS1* could be essential for the normal regulation of the corresponding mRNA, thus its downregulation may lead to cell survival considering that *BRWD1* has an important role in apoptosis. It is important noticing, however, that *BRWD1* was not differentially expressed in our analysis. Another lncRNA-AS that stood out in our analysis was *ZNF667-AS1*, with a \log_{FC} of approximately -5.16 and an AUC of 0.717. Other studies have demonstrated that

ZNF667-AS1 is commonly downregulated in several cancer types, including UCEC [36, 37]. Vrba and colleagues [45] showed that it is expressed in all normal finite lifespan human cells examined to date and is downregulated or lost in immortalized human mammary epithelial cells. Additionally, they demonstrated that its downregulation is due to DNA hypermethylation [45]. In cervical cancer, *ZNF667-AS1* inhibited the proliferation of cancer cells and its downregulation was negatively correlated with the overall survival of patients, tumor size, and FIGO stage [46].

To evaluate if our combined set of differentially expressed lncRNAs is able to correctly separate TP from NT samples, we performed the SVM approach. From the seventeen lncRNAs, three of them (*PLCG1-AS1*, *LINC01411*, and *LINC02249*) presented elevated expression variation and low AUC and could not be considered good models. Thus, 14 remained during the SVM analysis. The results for all tested sets are shown in Table 2. The first set contained all 14 selected lncRNAs and obtained an accuracy of 0.9583, indicating that this set is an efficient classification model. The second set contained 5 lncRNAs that had an AUC greater than 0.7 and attained an accuracy of 0.9167. The third and final sets contained the top 2 lncRNAs that attained the best overall AUC, obtaining an accuracy of 0.9167 (Figure 4).

TABLE 1: Long noncoding RNAs that discriminate TP from NT samples.

lncRNA	logFC	FDR	AUC
<i>LINC01376</i>	5.88492063377296	2.37514982980919e ⁻²⁶	0.913
<i>BRWD1-AS1</i>	-5.05883480566089	2.3408685424323e ⁻⁵⁰	0.820
<i>LINC00244</i>	-5.11057542483241	2.25688066033936e ⁻¹³	0.892
<i>LINC02475</i>	-5.15921206745447	8.93670161058092e ⁻⁰⁶	0.744
<i>ZNF667-AS1</i>	-5.16386849714406	8.14134337771604e ⁻⁰⁸	0.717
<i>DEPDC1-AS1</i>	-5.17825782522368	2.06667461632931e ⁻²⁵	0.633
<i>LINC0001</i>	-5.20814071478136	4.04097876674846e ⁻¹⁰	0.601
<i>LINC00504</i>	-5.29670533561579	2.78034526756707e ⁻⁰⁸	0.535
<i>LINC00632</i>	-5.38175032295688	4.77496645881152e ⁻¹⁶	0.659
<i>MACC1-AS1</i>	-5.38578795348022	2.25613841195626e ⁻¹⁰	0.694
<i>TCF4-AS1</i>	-5.64269979170812	0.000474000179607125	0.516
<i>SHANK2-AS1</i>	-5.92292108357159	4.02699887083864e ⁻⁰⁵	0.624
<i>WDR86-AS1</i>	-6.32585914185104	9.39257125151877e ⁻¹³	0.643
<i>LINC01411</i>	-7.78029289441235	2.81489971171908e ⁻⁰⁶	0.471
<i>PLCG1-AS1</i>	-7.93247628000136	1.81453269242651e ⁻⁰⁶	0.674
<i>SLC8A1-AS1</i>	-8.15747923329883	4.51757852734348e ⁻⁰⁹	0.511
<i>LINC02249</i>	-9.64327159184569	2.76479236381279e ⁻⁰⁷	0.578

Thus, even a set composed only by the two top AUC representatives is good enough to discriminate the samples with an accuracy of 91.67%, identical to the accuracy of the top 5 set. The unsupervised hierarchical cluster (Figure 5) demonstrated that all NT samples clustered together with high group support. TP samples were more dispersed and some of them were closer to NTs than to other TPs. This might be explained due to the higher expression variability in TP samples usually observed in many cancer types, such as melanoma [47], breast cancer [48], lung cancer [49], and hepatocellular carcinoma [10].

In order to look for pathways and biological functions of the lncRNAs from our set, we performed a correlation analysis among the expression data of all differentially expressed transcripts (including miRNAs). A gene coexpression network was plotted from the result of the correlation analysis (Figure 6) and was assessed alongside enrichment analyses. ClueGO (v. 2.3.3) [50] indicated that CHD5 and WRB (downregulated) and EHMT1 (upregulated) were involved in the histone H3K27 methylation. Trimethylation of H3K27 has been associated with transcriptional inhibition of genes in endometrial cancer [51]. In addition, EZH2, which is a methyl-transferase for H3K27, is upregulated in many tumors [52, 53], consequently causing low expression levels of the genes regulated in this part of the genome. In this study, EZH2 is slightly upregulated in UCEC, although the difference was not statistically significant, considering a FDR < 0.01. Previous studies have demonstrated that downregulation of H3K27 methylation process in tumor may contribute to its progression by enhancing the expression of oncogenes [54, 55].

Many studies first identified CHD5 as a tumor suppressor gene in neuroblastomas [56–60]. Subsequent research showed increasing evidence that it functions as tumor suppressor in several other types of cancer, including ovarian [61], breast [62], lung [63], and colorectal cancer [64]. Thus, CHD5, commonly found downregulated in cancer due to deletion of region 1p36, where it is located, may act as a master regulator, controlling the key processes for the suppression of a variety of tumors [65]. Clonal alteration was found at the region 1p36 in endometrial cancer [66], which might explain why CHD5 is downregulated with a logFC of -5.40 in our analysis. Furthermore, in our coexpression network, CHD5 is positively and strongly linked to three lncRNAs of our set: *TCF4-AS1*, *LINC02249*, and *LINC02475*, which may indicate that they possibly play a role on the expression of CHD5. To our knowledge, all three lncRNAs have not been biologically associated with cancer to date, nor to any other disease. CHD5 was also positively correlated to hsa-mir-767, which is downregulated in UCEC. Downregulation of this miRNA has also been detected in lung adenocarcinoma cells [67], whereas upregulation was detected in human melanoma [68]. No miRNA prediction binding site tool (see Section 4) identified a statistically supported connection between hsa-mir-767 and CHD5.

Gene ontology enrichment analysis performed on pathfindR (v. 1.2.1) [69] showed that *GPR161* (upregulated) was associated with the negative regulation of the hedgehog signaling pathway. However, previous studies revealed that abnormal activation of this pathway is related to cell proliferation in endometrial cancer [70, 71]. According to the Human Protein Atlas [72], high expression of *GPR161* is associated with low survival probability in UCEC (66% probability of 5-year survival in *GPR161* high and 79% in *GPR161* low, $p = 0.0016$) but it is not prognostic for UCEC. Additionally, *GPR161* has been found to be upregulated in breast cancer as well, acting as a promoter of cell proliferation and invasion [73]. The coexpression network reveals that this gene is positively correlated with MIR545, a pre-mir, which has been linked to cell proliferation in colorectal cancer [74] and hepatocellular carcinoma [75]. We also found that MIR545 is upregulated in endometrial cancer with a 7.8 logFC. Those results indicate that *GPR161* and MIR545 are probably promoting cell proliferation on UCEC as well, but other elements may also be involved. Again, no miRNA prediction binding site tool (see Section 4) identified a statistically supported connection between MIR545 and *GPR161*.

KEGG analysis also performed on pathfindR (v. 1.2.1) [69] indicated that the upregulated genes *HDAC7*, *GTF2AIL*, and *MAPKAPK2* were involved in the viral carcinogenesis pathway. Previous studies have demonstrated that inhibition of histone deacetylases, such as *HDAC7*, induces apoptosis, cell cycle arrest, and growth inhibition in endometrial cancer cells [76–79]. Moreover, *LINC01376*, the only lncRNA that is upregulated in UCEC from our set, positively and strongly correlates with *HDAC7* ($r = 0.91$), which reinforces *LINC01376* as a marker. Again, based on the Human Protein Atlas [72], high expression of *HDAC7* is associated with low survival probability in UCEC

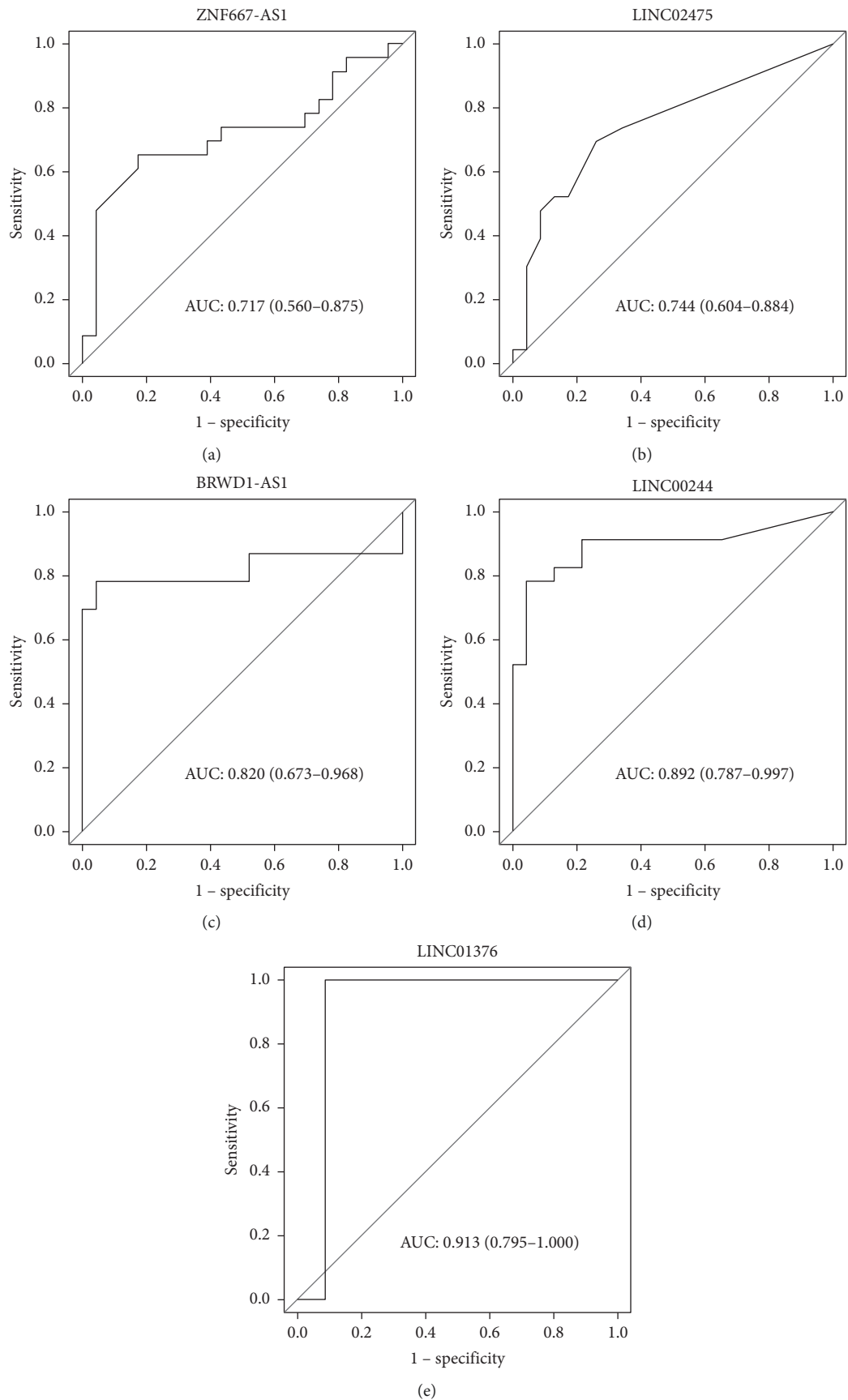


FIGURE 3: ROC curve of the lncRNAs that obtained AUC greater than 7; sensitivity (true positive rate) (y axis); 1-specificity (x axis) (false negative rate).

TABLE 2: SVM analysis for the three tested sets of lncRNAs.

Set*	Accuracy	CI (95%)	p value	Kappa	Sensitivity	Specificity
First set	0.9583	0.7888, 0.9989	$1.49e-06$	0.9167	0.9167	1.0000
Second set	0.9167	0.6262, 0.9526	$1.794e-05$	0.8333	0.9167	0.9167
Third set	0.9167	0.73, 0.9897	$1.794e-05$	0.8333	0.9167	0.9167

*The first set is comprised of 14 lncRNAs; the second set is composed of the five lncRNAs with AUC > 0.7; the third set is composed by the AUC's top two lncRNAs, as shown in Table 1.

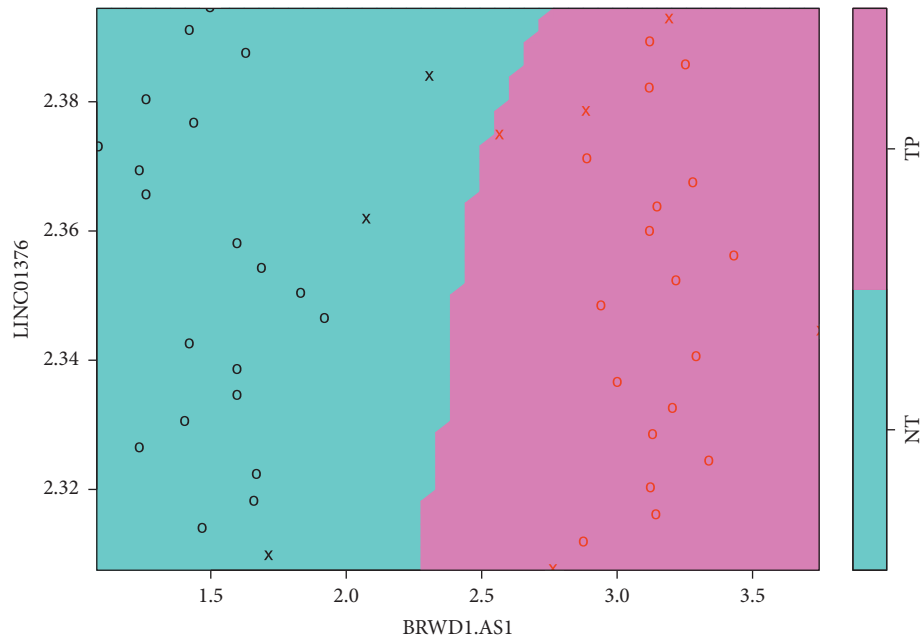


FIGURE 4: Scatter plot of the fitted support vector machine model based on the expression of the top 2 lncRNAs. Classes: primary tumor (TP) and adjacent tissue (NT). Support vectors instances: x. Expression values are in log 10 scale.

($p = 0.045$), but it is not considered a prognostic for UCEC. Based on KEGG (<https://www.genome.jp/kegg/>), *MAPKAPK2* also participates in cellular senescence.

In order to look for more putative pathways and biological functions of the lncRNAs from our set, we performed an *in silico* prediction of lncRNA function. As stated before, characterization of most lncRNAs remains to be done; therefore, we were able to predict the functions of seven lncRNAs from our set: *BRWD1-AS1*, *LINC00504*, *SHANK2-AS1*, *MACC1-AS1*, *WRD86-AS1*, *LINC00632*, and *ZNF667-AS1*. The first four were strongly associated with male reproductive system pathways, such as spermatogenesis, male gamete generation, and sperm motility. The alteration of this pathway is not a novelty in endometrial cancer [80, 81]. *LINC00632* and *ZNF667-AS1* were potentially related with neural functions such as neuron projection, synapse organization, and neurotransmitter secretion and transport, whereas *WRD86-AS1* was linked to the spliceosome complex and RNA splicing, as well as to breast, thyroid, and colorectal cancers.

Even though there have been lncRNAs reported as oncogenic drivers [82], survival analysis performed for each lncRNA from our set did not associate their expression profile to a specific prognostic; consequently, none of them can be reckoned as drivers. However, these lncRNAs can still

act as biomarkers, once we demonstrated that they effectively distinguish TP from NT.

3. Conclusions

In summary, our data suggest a set of 14 lncRNAs as highly effective biomarkers of UCEC, albeit none can be reckoned as drivers and coexpression associated genes are not prognostic. Moreover, this set can be reduced to the two top lncRNAs (*LINC01376* and *BRWD1-AS1*) with a minimal reduction in accuracy and specificity. This data must be validated in clinical samples, but it is predicted to contribute in the diagnosing UCEC. While there have been similar works on lncRNAs [83] that paved the way for our study, we focused on looking for prognostics and diagnostics biomarkers for UCEC, considering only highly expressed lncRNAs and honing on strong statistical analyses to support our findings. In addition, the elucidation of how these lncRNAs play a part in the establishment and progression of UCEC may contribute to new diagnostic options in the future.

4. Materials and Methods

All data analyzed in this study are available at the Cancer Genome Atlas (TCGA), and all analyses were performed

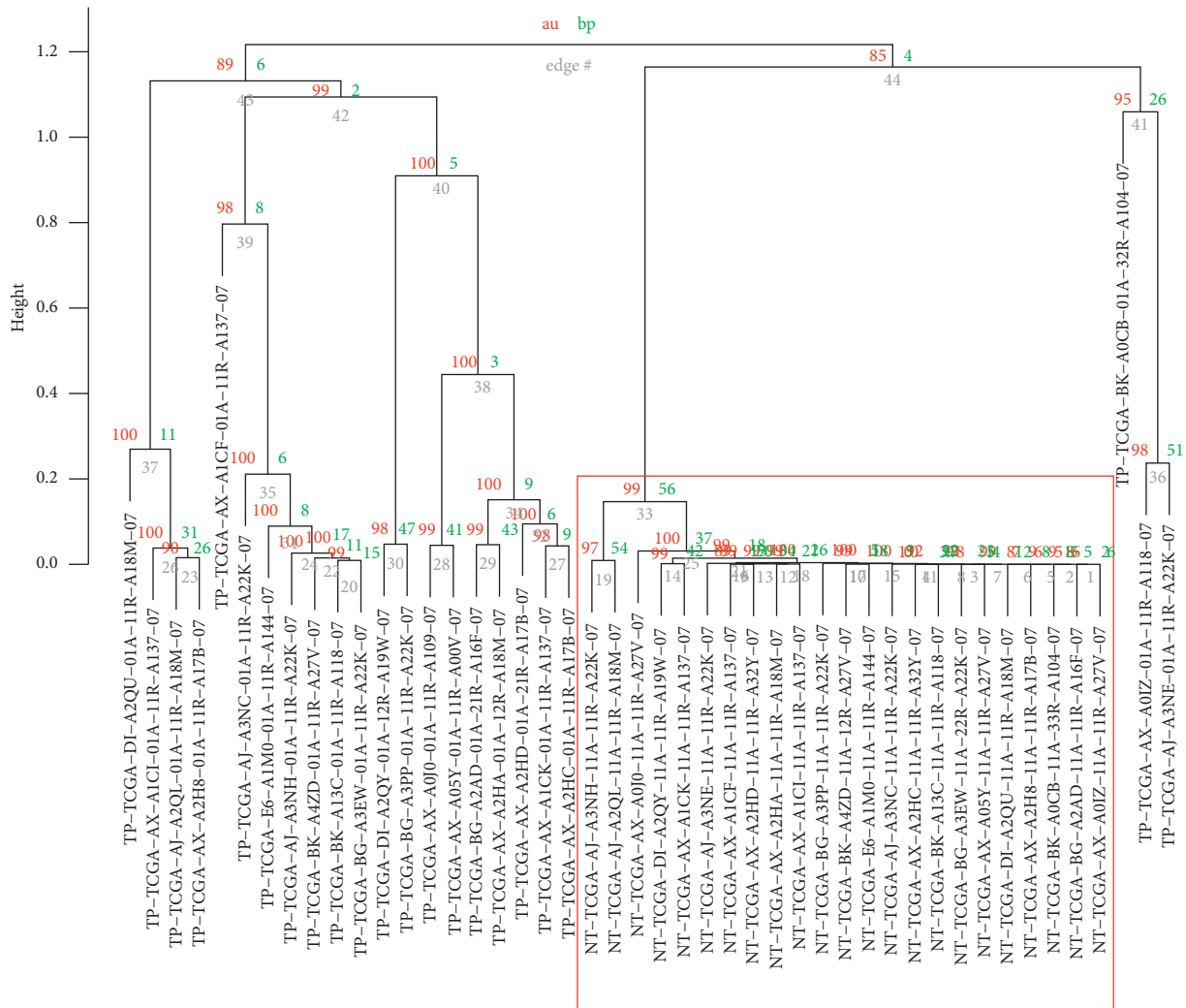


FIGURE 5: Hierarchical clustering analysis using the train set (23 patients with TP and NT samples for each) based on the expression of the 14 lncRNAs with $|\log \text{fold change}| > 5$. Before each sample ID, there is the indication if the samples belong to the TP or NT group. TP: primary solid tumor. NT: adjacent normal tissue. Red values (au) represent the group support. Green values (bp) represent the bootstrap support. Grey values (edge) represent the limit of the branches. Red square highlights the clustering of NT samples.

using the R software (v. 3.4.0) (<https://www.R-project.org/>). Data download, preprocessing, and differential expression analysis were performed using TCGABiolinks package (v 2.7.1) [84], available at Bioconductor digital repository (<https://www.bioconductor.org/>). Total RNA and miRNA data must be separately downloaded from TCGA. For the total RNA data, 587 samples were downloaded: 35 NTs and 551 TPs. We analyzed an initial set of 23 patients that had both NT and TP samples, thus a total of 46 samples. The results were then visualized on volcano plots, and the 50 most differentially expressed transcripts (top 50) were featured on a heatmap, plotted using the package gplots (v. 3.0.1). For the miRNA data, a total of 579 samples were downloaded, 33 NTs and 545 TPs. We analyzed a set of 21 patients (all included in the total RNA analysis) that had both TP and NT samples, thus a total of 42 samples.

For the putative lncRNA biomarkers, we first performed a supervised prediction model using the area under the curve

(AUC) of the receiver operating characteristic (ROC) with the pROC package (v. 1.13.0) [85], then a supervised learning model using support vector machines (SVM), with the caret package (v. 6.0–81). Next, we performed an unsupervised model using hierarchical cluster analysis with 1000 bootstrap replications. Clusters with unbiased grouping support p values (au) of approximately 95% were considered as statistically significant groups. Hierarchical clusters were plotted using the pvclust package (v. 2.0–0) [86]. We also performed a Kaplan–Meier survival analysis considering the expression of the lncRNAs only on TP samples for the purpose of appraising whether they could be considered drivers for UCEC. At last, in order to look for pathways and biological functions of the lncRNAs of our set, we set up a coexpression network that was assessed alongside enrichment analyses. We also used FuncPred [87], an online tool that performs in silico prediction of lncRNA function by using tissue specific and evolutionary conserved expression.

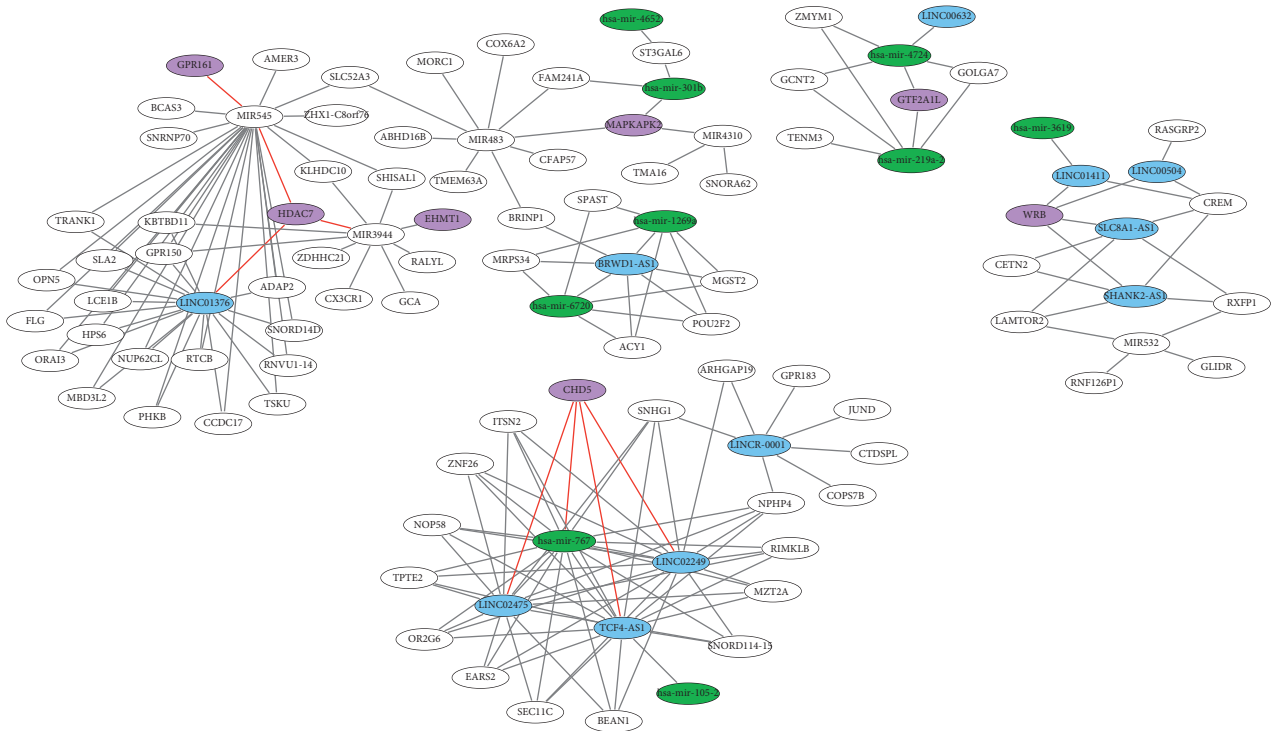


FIGURE 6: Representative gene coexpression networks. Blue circles are the lncRNAs; green circles are the miRNAs; purple circles are the highlighted genes; white circles are the genes. Each edge represents an $r > 0.7$ or $r < -0.7$ and p value < 0.05 . Red edges highlight the discussed pathways.

4.1. Download and Data Preprocessing. The UCEC harmonized total RNA expression data (hg38) and miRNA expression data downloaded from TCGA were obtained through RNA-seq using Illumina HiSeq platform. Total RNA and miRNA are separately download due to differences in the data composition. First, we set the query for the download of the total RNA data, using the function *GDCquery* with the following options: *project* = "TCGA-UCEC," *data.category* = "Transcriptome Profiling," *data.type* = "Gene Expression Quantification," *workflow.type* = "HTSeq—Counts" and *legacy* = FALSE. The query for the microRNA data uses the following options: *project* = "TCGA-UCEC," *data.category* = "Transcriptome Profiling," *data.type* = "miRNA Expression Quantification" and *legacy* = FALSE. Then, we downloaded both data using the function *GDCdownload* with the previously set queries and the option *method* = "api." The function *GDCprepare* was then used to transform the downloaded total RNA data into summarized experiment data, making it suitable for the analyses.

The data normalization was performed first by GC content, using the function *TCGAanalyze_Normalization* with the options *geneInfo* = *geneInfoHT* and *method* = "gcContent," and then by gene length, changing the *method* option to "geneLength." Then, a quantile filter was applied using the function *TCGAanalyze_Filtering* with the option *method* = "quantile" and *qnt.cut* = 0.25. Finally, the Spearman correlation among samples was checked using the R function *cor* and both NT and TP sample groups were separated with the function *TCGAquery_SampleTypes*.

4.2. Differential Expression and Survival Analysis. We considered as differentially expressed those genes with a false discovery rate (FDR) < 0.01 and log fold change ($\log FC$) $\geq |\pm 5|$, using the function *TCGAanalyze_DEA* with the options *fdr.cut* = 0.01, *logFC.cut* = 5, and *method* = "glmLRT." This high cutoff was chosen in sense to detect the genes with the highest potential for group differentiation.

Considering the expression on TP samples of all differentially expressed lncRNAs, we performed a Kaplan–Meier survival analysis with the function *TCGA_SurvivalKM* and the following options: *Survresult* = TRUE, *p.cut* = 0.5, *ThreshTop* = 0.67, and *ThreshDown* = 0.33. Survival curves for each lncRNA were generated and analyzed separately.

4.3. ROC Analysis. For the ROC analysis, we focused in the differentially expressed lncRNAs. The individualized results were visualized on a AUC-ROC curve plot, which is a graph showing the performance of a classification model at all classification thresholds. The evaluation metric used was the AUC, which represents the degree or measure of separability, that is, how much the model is capable of distinguishing between two groups [88].

4.4. Support Vector Machine (SVM) Analysis. For the SVM, we first set a randomly chosen seed of 3033, making our work replicable, with the function *set.seed* (3033). Then, we used the function *trainControl*() that controls the

computational nuances of the *train()* method. At this point, we applied a 1000 bootstrap replication, with the option *method* = “boot” and *number* = 1000.

Next, we set our train and prediction data. The train data contained the selected differentially expressed lncRNAs from the 23 initial patients. Here, three lncRNAs were cut out of the analysis for presenting elevated expression variation. Our prediction model data contained the same lncRNAs, however, from 24 samples, different from the 46 samples we have been working on: 12 TPs and the remaining 12 NTs from TCGA database that did not have TP/NT correspondence in the same patient. The expression data from the 24 new patients were analyzed apart from the set of 23 patients. To train our data, we tested both land radial kernel methods, using the function *train()* with the previously set *trainControl* as a parameter, along with the options *preProcess* = *c* (“center,” “scale”), *tuneLength* = 10, and *method* = “svmLinear” or *method* = “svmRadial”. Next, we tested our classifier at specific cost (C) values for linear method, using the function *expand.grid* with cost values ranging from 0 to 5 and adding the *tuneGrid* parameter with the set grid to the *train()* function.

This process was repeated for three different sets of lncRNAs. The first set contained all selected lncRNAs, and the best performance was obtained using linear kernel method, with cost = 0.25. The second set had the ones that attained an AUC greater than 0.7. The best performance was obtained using the linear kernel method, with cost = 1. The third set had the two lncRNAs that attained the best AUC. The best performance was obtained using a radial basis function kernel method, with standard values of sigma = 0.07 and cost = 5.

4.5. Coexpression Network and Enrichment Analysis. A Spearman correlation analysis among the expression data of all differentially expressed transcripts and miRNAs was performed in order to plot a gene coexpression network. The cutoff for the statistically significant correlations was $r > 0.7$ and $r < -0.7$ and p value < 0.05 . The network was visualized using Cytoscape (v. 3.5.1) [89]. For better clearance of the network, only the following correlations of interest were plotted: lncRNA-miRNA, lncRNA-mRNA, and miRNA-mRNA. The coexpression network was analyzed alongside enrichment analyses performed on ClueGo (v. 2.3.3) [50] and pathfindR (v. 1.2.1) [69].

4.6. miRNA Interaction Prediction. The miRNAs and their putative targets were used to predict their interaction using the online software TargetScan (release 7.1) [90] and miRDB [91].

Data Availability

All data used in this work are publicly available at the Cancer Genome Atlas (TCGA) database (<https://cancergenome.nih.gov/>).

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors thank the fellowship from the Programa Nacional de Pós-Doutorado (PNPD) CAPES/HCPA to TF (Process no. 88887.160608/2017-00) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the fellowships provided to MF (Process no. 140002/2018-9) and UM (Process no. 307038/2015-7). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (CAPES)—Finance Code 001.

References

- [1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, “Global cancer statistics,” *CA: A Cancer Journal for Clinicians*, vol. 61, no. 2, pp. 69–90, 2011.
- [2] J. Ferlay, I. Soerjomataram, R. Dikshit et al., “Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012,” *International Journal of Cancer*, vol. 136, no. 5, pp. E359–E386, 2015.
- [3] M. Pichler and G. A. Calin, “MicroRNAs in cancer: from developmental genes in worms to their clinical application in patients,” *British Journal of Cancer*, vol. 113, no. 4, pp. 569–573, 2015.
- [4] S. Saso, J. Chatterjee, E. Georgiou, A. M. Ditri, J. R. Smith, and S. Ghaem-Maghami, “Endometrial cancer,” *BMJ*, vol. 343, no. jul06 2, p. d3954, 2011.
- [5] V. C. Jordan, “The role of tamoxifen in the treatment and prevention of breast cancer,” *Current Problems in Cancer*, vol. 16, no. 3, pp. 129–176, 1992.
- [6] R. Murali, R. A. Soslow, and B. Weigelt, “Classification of endometrial carcinoma: more than two types,” *The Lancet Oncology*, vol. 15, no. 7, pp. e268–e278, 2014.
- [7] P. Morice, A. Leary, C. Creutzberg, N. Abu-Rustum, and E. Darai, “Endometrial cancer,” *The Lancet*, vol. 387, no. 10023, pp. 1094–1108, 2016.
- [8] G. Diaz, M. Melis, A. Tice et al., “Identification of microRNAs specifically expressed in hepatitis C virus-associated hepatocellular carcinoma,” *International Journal of Cancer*, vol. 133, no. 4, pp. 816–824, 2013.
- [9] W. Ding, H. Yang, S. Gong et al., “Candidate miRNAs and pathogenesis investigation for hepatocellular carcinoma based on bioinformatics analysis,” *Oncology Letters*, vol. 13, no. 5, pp. 3409–3414, 2017.
- [10] T. Falcon, M. Freitas, A. C. Mello, L. Coutinho, M. R. Alvares-da-Silva, and U. Matte, “Analysis of the Cancer Genome Atlas (TCGA) data reveals novel putative ncRNAs targets in hepatocellular carcinoma,” *BioMed Research International*, vol. 2018, Article ID 2864120, 9 pages, 2018.
- [11] C. Vallone, G. Rigon, C. Gulia et al., “Non-coding RNAs and endometrial cancer,” *Genes*, vol. 9, no. 4, p. 187, 2018.
- [12] B. J. Chen, F. L. Byrne, K. Takenaka et al., “Transcriptome landscape of long intergenic non-coding RNAs in endometrial cancer,” *Gynecologic Oncology*, vol. 147, no. 3, pp. 654–662, 2017.
- [13] S. Hutt, A. Taylor, P. Ellis, A. Michael, S. Butler-Manuel, and J. Chatterjee, “The role of biomarkers in endometrial cancer

- and hyperplasia: a literature review," *Acta Oncologica*, vol. 58, no. 3, pp. 342–352, 2019.
- [14] K. Takenaka, B. J. Chen, S. C. Modesitt, F. L. Byrne, K. L. Hoehn, and M. Janitz, "The emerging role of long non-coding RNAs in endometrial cancer," *Cancer Genetics*, vol. 209, no. 10, pp. 445–455, 2016.
- [15] B. Hrdlickova, R. C. de Almeida, Z. Borek, and S. Withoff, "Genetic variation in the non-coding genome: involvement of micro-RNAs and long non-coding RNAs in disease," *Biochimica et Biophysica Acta (BBA)—Molecular Basis of Disease*, vol. 1842, no. 10, pp. 1910–1922, 2014.
- [16] V. Pelechano and L. M. Steinmetz, "Gene regulation by antisense transcription," *Nature Reviews Genetics*, vol. 14, no. 12, pp. 880–893, 2013.
- [17] A. M. Khalil, M. Guttman, M. Huarte et al., "Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression," *Proceedings of the National Academy of Sciences*, vol. 106, no. 28, pp. 11667–11672, 2009.
- [18] M. E. Dinger, P. P. Amaral, T. R. Mercer et al., "Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation," *Genome Research*, vol. 18, no. 9, pp. 1433–1445, 2008.
- [19] M. Smolle, M. Bullock, H. Ling, M. Pichler, and J. Haybaeck, "Long non-coding RNAs in endometrial carcinoma," *International Journal of Molecular Sciences*, vol. 16, no. 11, pp. 26463–26472, 2015.
- [20] M.-Y. Sun, J.-Y. Zhu, C.-Y. Zhang et al., "Autophagy regulated by lncRNA HOTAIR contributes to the cisplatin-induced resistance in endometrial cancer cells," *Biotechnology Letters*, vol. 39, no. 10, pp. 1477–1484, 2017.
- [21] L. Huang, L.-M. Liao, A.-W. Liu et al., "Overexpression of long noncoding RNA HOTAIR predicts a poor prognosis in patients with cervical cancer," *Archives of Gynecology and Obstetrics*, vol. 290, no. 4, pp. 717–723, 2014.
- [22] J. Huang, P. Ke, L. Guo et al., "Lentivirus-mediated RNA interference targeting the long noncoding RNA HOTAIR inhibits proliferation and invasion of endometrial carcinoma cells in vitro and in vivo," *International Journal of Gynecologic Cancer*, vol. 24, no. 4, pp. 635–642, 2014.
- [23] Y. Zhao, Y. Yang, J. Trovik et al., "A novel wnt regulatory axis in endometrioid endometrial cancer," *Cancer Research*, vol. 74, no. 18, pp. 5103–5117, 2014.
- [24] K. Banno, I. Kisu, M. Yanokura et al., "Biomarkers in endometrial cancer: possible clinical applications (Review)," *Oncology Letters*, vol. 3, no. 6, pp. 1175–1180, 2012.
- [25] E. Colas, C. Perez, S. Cabrera et al., "Molecular markers of endometrial carcinoma detected in uterine aspirates," *International Journal of Cancer*, vol. 129, no. 10, pp. 2435–2444, 2011.
- [26] T. A. O'Mara, M. Zhao, and A. B. Spurdle, "Meta-analysis of gene expression studies in endometrial cancer identifies gene expression profiles associated with aggressive disease and patient outcome," *Scientific Reports*, vol. 6, no. 1, Article ID 36677, 2016.
- [27] A. D. Santin, F. Zhan, S. Cane' et al., "Gene expression fingerprint of uterine serous papillary carcinoma: identification of novel molecular markers for uterine serous cancer diagnosis and therapy," *British Journal of Cancer*, vol. 92, no. 8, pp. 1561–1573, 2005.
- [28] W. Tang, S. Wan, Z. Yang, A. E. Teschendorff, and Q. Zou, "Tumor origin detection with tissue-specific miRNA and DNA methylation markers," *Bioinformatics*, vol. 34, no. 3, pp. 398–406, 2018.
- [29] X. Zeng, X. Zhang, and Q. Zou, "Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks," *Briefings in Bioinformatics*, vol. 17, no. 2, pp. 193–203, 2016.
- [30] Q. Zou, J. Li, L. Song, X. Zeng, and G. Wang, "Similarity computation strategies in the microRNA-disease network: a survey," *Briefings in Functional Genomics*, vol. 15, no. 1, pp. 55–64, 2016.
- [31] X.-R. Tang, Y.-Q. Li, S.-B. Liang et al., "Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study," *The Lancet Oncology*, vol. 19, no. 3, pp. 382–393, 2018.
- [32] J. H. Bullard, E. Purdom, K. D. Hansen, and S. Dudoit, "Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments," *BMC Bioinformatics*, vol. 11, no. 1, p. 94, 2010.
- [33] D. Risso, K. Schwartz, G. Sherlock, and S. Dudoit, "GC-content normalization for RNA-Seq data," *BMC Bioinformatics*, vol. 12, no. 1, p. 480, 2011.
- [34] J. D. Ransohoff, Y. Wei, and P. A. Khavari, "The functions and unique features of long intergenic non-coding RNA," *Nature Reviews Molecular Cell Biology*, vol. 19, no. 3, pp. 143–157, 2017.
- [35] R. A. Gupta, N. Shah, K. C. Wang et al., "Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis," *Nature*, vol. 464, no. 7291, pp. 1071–1076, 2010.
- [36] H. Behravan, J. M. Hartikainen, M. Tengström et al., "Machine learning identifies interacting genetic variants contributing to breast cancer risk: a case study in Finnish cases and controls," *Scientific Reports*, vol. 8, Article ID 13149, 2018.
- [37] M. Shi, K. M. O'Brien, D. P. Sandler, J. A. Taylor, D. V. Zaykin, and C. R. Weinberg, "Previous GWAS hits in relation to young-onset breast cancer," *Breast Cancer Research and Treatment*, vol. 161, no. 2, pp. 333–344, 2016.
- [38] V. Villegas and P. Zaphiropoulos, "Neighboring gene regulation by antisense long non-coding RNAs," *International Journal of Molecular Sciences*, vol. 16, no. 2, pp. 3251–3266, 2015.
- [39] S. Katayama, Y. Tomaru, T. Kasukawa et al., "Antisense transcription in the mammalian transcriptome," *Science*, vol. 309, no. 309, pp. 1564–1566, 2005.
- [40] P. Liu, L. Jin, L. Zhao et al., "Identification of a novel antisense long non-coding RNA PLA2G16-AS that regulates the expression of PLA2G16 in pigs," *Gene*, vol. 671, pp. 78–84, 2018.
- [41] M. Martinez-Moreno, T. M. O'Shea, J. P. Zepecki et al., "Regulation of peripheral myelination through transcriptional buffering of *Egr2* by an antisense long non-coding RNA," *Cell Reports*, vol. 20, no. 8, pp. 1950–1963, 2017.
- [42] P. Halley, B. M. Kadakkuzha, M. A. Faghihi et al., "Regulation of the apolipoprotein gene cluster by a long noncoding RNA," *Cell Reports*, vol. 6, no. 1, pp. 222–230, 2014.
- [43] Y. Xue, G. Ma, Z. Zhang et al., "A novel antisense long noncoding RNA regulates the expression of MDC1 in bladder cancer," *Oncotarget*, vol. 6, no. 1, pp. 484–493, 2014.
- [44] G. Stelzer, R. Rosen, I. Plaschkes et al., "The GeneCards suite: from gene data mining to disease genome sequence analyses," *Current Protocols in Bioinformatics*, vol. 54, no. 1, pp. 1.30.1–1.30.33, 2016.
- [45] L. Vrba, J. C. Garbe, M. R. Stampfer, and B. W. Futscher, "A lincRNA connected to cell mortality and epigenetically-silenced in most common human cancers," *Epigenetics*, vol. 10, no. 11, pp. 1074–1083, 2015.
- [46] L. P. 1 Zhao, R. H. Li, D. M. Han et al., "Independent prognostic Factor of low-expressed LncRNA ZNF667-AS1 for

- cervical cancer and inhibitory function on the proliferation of cervical cancer," *European Review for Medical and Pharmaceutical Sciences*, vol. 21, no. 23, pp. 5353–5360, 2017.
- [47] S. M. Shaffer, M. C. Dunagin, S. R. Torborg et al., "Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance," *Nature*, vol. 546, no. 7658, pp. 431–435, 2017.
- [48] L. G. Martelotto, C. K. Ng, S. Piscuoglio, B. Weigelt, and J. S. Reis-Filho, "Breast cancer intra-tumor heterogeneity," *Breast Cancer Res*, vol. 16, no. 3, p. 210, 2014.
- [49] Y. Liu, J. Zhang, L. Li et al., "Genomic heterogeneity of multiple synchronous lung cancer," *Nature Communications*, vol. 7, no. 1, p. 13200, 2016.
- [50] G. Bindea, B. Mlecnik, H. Hackl et al., "ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks," *Bioinformatics*, vol. 25, no. 8, pp. 1091–1093, 2009.
- [51] Q. Li, N. Jia, X. Tao, K. Hua, and W. Feng, "The expression and significance of histone lysine methylation in endometrial cancer," *Oncology Letters*, vol. 14, no. 14, pp. 6210–6216, 2017.
- [52] Y. Gao, P. Hyttel, and V. J. Hall, "Regulation of H3K27me3 and H3K4me3 during early porcine embryonic development," *Molecular Reproduction and Development*, vol. 77, no. 6, pp. 540–549, 2010.
- [53] K. Li, M. K. Chen, J. Situ et al., "Role of co-expression of c-Myc, EZH2 and p27 in prognosis of prostate cancer patients after surgery," *Chinese Medical Journal*, vol. 126, no. 1, pp. 82–87, 2013.
- [54] Y. Wei, W. Xia, Z. Zhang et al., "Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers," *Molecular Carcinogenesis*, vol. 47, no. 9, pp. 701–706, 2008.
- [55] L. G. Pellakuru, T. Iwata, B. Gurel et al., "Global levels of H3K27me3 track with differentiation in vivo and are deregulated by MYC in prostate cancer," *The American Journal of Pathology*, vol. 181, no. 2, pp. 560–569, 2012.
- [56] P. M. Thompson, T. Gotoh, M. Kok, P. S. White, and G. M. Brodeur, "CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system," *Oncogene*, vol. 22, no. 7, pp. 1002–1011, 2003.
- [57] I. Garcia, G. Mayol, E. Rodríguez et al., "Expression of the neuron-specific protein CHD5 is an independent marker of outcome in neuroblastoma," *Molecular Cancer*, vol. 9, no. 1, p. 277, 2010.
- [58] T. Fujita, J. Igarashi, E. R. Okawa et al., "CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas," *JNCI: Journal of the National Cancer Institute*, vol. 100, no. 13, pp. 940–949, 2008.
- [59] I. Garcia, G. Mayol, J. Rios et al., "A three-gene expression signature model for risk stratification of patients with neuroblastoma," *Clinical Cancer Research*, vol. 18, no. 7, pp. 2012–2023, 2012.
- [60] H. Koyama, T. Zhuang, J. E. Light et al., "Mechanisms of CHD5 inactivation in neuroblastomas," *Clinical Cancer Research*, vol. 18, no. 6, pp. 1588–1597, 2012.
- [61] K. L. Gorringer, D. Y. H. Choong, L. H. Williams et al., "Mutation and methylation analysis of the chromodomain-helicase-DNA binding 5 gene in ovarian cancer," *Neoplasia*, vol. 10, no. 11, pp. 1253–IN32, 2008.
- [62] X. Wu, Z. Zhu, W. Li et al., "Chromodomain helicase DNA binding protein 5 plays a tumor suppressor role in human breast cancer," *Breast Cancer Research*, vol. 14, no. 3, p. R73, 2012.
- [63] R. Zhao, Q. Yan, J. Lv et al., "CHD5, a tumor suppressor that is epigenetically silenced in lung cancer," *Lung Cancer*, vol. 76, no. 3, pp. 324–331, 2012.
- [64] G. Ragnarsson, G. Eiriksdottir, J. T. Johannsdottir, J. G. Jonasson, V. Egilsson, and S. Ingvarsson, "Loss of heterozygosity at chromosome 1p in different solid human tumours: association with survival," *British Journal of Cancer*, vol. 79, no. 9–10, pp. 1468–1474, 1999.
- [65] V. Kolla, T. Zhuang, M. Higashi, K. Naraparaju, and G. M. Brodeur, "Role of CHD5 in human cancers: 10 years later," *Cancer Research*, vol. 74, no. 3, pp. 652–658, 2014.
- [66] A. Pazarbasi, M. Kasap, O. Demirhan et al., "Chromosomal abnormalities in endometrial and ovarian carcinomas," *Balkan Journal of Medical Genetics*, vol. 10, no. 2, pp. 61–69, 2007.
- [67] Y. L. Wan, H. J. Dai, W. Liu, and H. T. Ma, "miR-767-3p inhibits growth and migration of lung adenocarcinoma cells by regulating CLDN18," *Oncology Research Featuring Pre-clinical and Clinical Cancer Therapeutics*, vol. 26, no. 4, pp. 637–644, 2018.
- [68] K. Zhang and L. Guo, "MiR-767 promoted cell proliferation in human melanoma by suppressing CYLD expression," *Gene*, vol. 641, pp. 272–278, 2018.
- [69] E. Ulgen, O. Ozisik, and O. U. Sezerman, *pathfindR: An R Package for Pathway Enrichment Analysis Utilizing Active Subnetworks*, bioRxiv, Long Island, NY, USA, 2018.
- [70] Y.-Z. Feng, T. Shiozawa, T. Miyamoto et al., "Overexpression of hedgehog signaling molecules and its involvement in the proliferation of endometrial carcinoma cells," *Clinical Cancer Research*, vol. 13, no. 5, pp. 1389–1398, 2007.
- [71] X. Liao, M. K. Siu, C. W. Au et al., "Aberrant activation of hedgehog signaling pathway contributes to endometrial carcinogenesis through β -catenin," *Modern Pathology*, vol. 22, no. 6, pp. 839–847, 2009.
- [72] M. Uhlen, L. Fagerberg, B. M. Hallstrom et al., "Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, Article ID 1260419, 2015.
- [73] M. E. Feigin, B. Xue, M. C. Hammell, and S. K. Muthuswamy, "G-protein-coupled receptor GPR161 is overexpressed in breast cancer and is a promoter of cell proliferation and invasion," *Proceedings of the National Academy of Sciences*, vol. 111, no. 11, pp. 4191–4196, 2014.
- [74] X. Huang and S. Lu, "MicroR-545 mediates colorectal cancer cells proliferation through up-regulating epidermal growth factor receptor expression in HOTAIR long non-coding RNA dependent," *Molecular and Cellular Biochemistry*, vol. 431, no. 1–2, pp. 45–54, 2017.
- [75] Z. Liu, C. Dou, B. Yao et al., "Ftx non coding RNA-derived miR-545 promotes cell proliferation by targeting RIG-I in hepatocellular carcinoma," *Oncotarget*, vol. 7, no. 18, pp. 25350–25365, 2016.
- [76] S. Jiang, S. C. Dowdy, X. W. Meng et al., "Histone deacetylase inhibitors induce apoptosis in both Type I and Type II endometrial cancer cells," *Gynecologic Oncology*, vol. 105, no. 2, pp. 493–500, 2007.
- [77] S. Xu, J. Ren, H. Chen et al., "Cytostatic and apoptotic effects of DNMT and HDAC inhibitors in endometrial cancer cells," *Current Pharmaceutical Design*, vol. 20, no. 11, pp. 1881–1887, 2014.
- [78] N. Takai, J. C. Desmond, T. Kumagai et al., "Histone deacetylase inhibitors have a profound antigrowth activity in endometrial cancer cells," *Clinical Cancer Research*, vol. 10, no. 3, pp. 1141–1149, 2004.

- [79] J. Ren, J. Zhang, H. Cai et al., "HDAC as a therapeutic target for treatment of endometrial cancers," *Current Pharmaceutical Design*, vol. 20, no. 11, pp. 1847–1856, 2014.
- [80] P. Yu, L. Yan, H. Zhang, X. Lin, and X. Zhao, "Expression and clinical significance of Sperm-Associated Antigen 9 in patients with endometrial carcinoma," *International Journal of Gynecologic Cancer*, vol. 22, no. 1, pp. 87–93, 2012.
- [81] Q. Gao, S. Xiang, K. Wilson, M. Madondo, A. Stephens, and M. Plebanski, "Sperm protein 17 expression by murine epithelial ovarian cancer cells and its impact on tumor progression," *Cancers*, vol. 10, no. 8, p. 276, 2018.
- [82] M. M. Ali, V. S. Akhade, S. T. Kosalai et al., "PAN-cancer analysis of S-phase enriched lncRNAs identifies oncogenic drivers and biomarkers," *Nature Communications*, vol. 9, no. 1, p. 883, 2018.
- [83] R. Akrami, A. Jacobsen, J. Hoell, N. Schultz, C. Sander, and E. Larsson, "Comprehensive analysis of long non-coding RNAs in ovarian cancer reveals global patterns and targeted DNA amplification," *PLoS One*, vol. 8, no. 11, Article ID e80306, 2013.
- [84] A. Colaprico, T. C. Silva, C. Olsen et al., "TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data," *Nucleic Acids Research*, vol. 44, no. 8, p. e71, 2016.
- [85] X. Robin, N. Turck, A. Hainard et al., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, p. 77, 2011.
- [86] R. Suzuki and H. Shimodaira, "Pvclust: an R package for assessing the uncertainty in hierarchical clustering," *Bioinformatics*, vol. 22, no. 12, pp. 1540–1542, 2006.
- [87] U. Perron, P. Provero, and I. Molineris, "In silico prediction of lncRNA function using tissue specific and evolutionary conserved expression," *BMC Bioinformatics*, vol. 18, no. S5, p. 144, 2017.
- [88] R. Kumar and A. Indrayan, "Receiver operating characteristic (ROC) curve for medical researchers," *Indian Pediatrics*, vol. 48, no. 4, pp. 277–287, 2011.
- [89] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [90] V. Agarwal, G. W. Bell, J. W. Nam, and D. P. Bartel, "Predicting effective microRNA target sites in mammalian mRNAs," *eLife*, vol. 4, 2015.
- [91] N. Wong and X. Wang, "miRDB: an online resource for microRNA target prediction and functional annotations," *Nucleic Acids Research*, vol. 43, no. D1, pp. D146–D152, 2015.