


RESEARCH ARTICLE

Open Access



The cox-filter method identifies respective subtype-specific lncRNA prognostic signatures for two human cancers

Suyan Tian^{1*} , Chi Wang^{2,3}, Jing Zhang⁴ and Dan Yu^{5*}

Abstract

Background: The most common histological subtypes of esophageal cancer are squamous cell carcinoma (ESCC) and adenocarcinoma (EAC). It has been demonstrated that non-marginal differences in gene expression and somatic alteration exist between these two subtypes; consequently, biomarkers that have prognostic values for them are expected to be distinct. In contrast, laryngeal squamous cell cancer (LSCC) has a better prognosis than hypopharyngeal squamous cell carcinoma (HSCC). Likewise, subtype-specific prognostic signatures may exist for LSCC and HSCC. Long non-coding RNAs (lncRNAs) hold promise for identifying prognostic signatures for a variety of cancers including esophageal cancer and head and neck squamous cell carcinoma (HNSCC).

Methods: In this study, we applied a novel feature selection method capable of identifying specific prognostic signatures uniquely for each subtype – the Cox-filter method – to The Cancer Genome Atlas esophageal cancer and HNSCC RNA-Seq data, with the objectives of constructing subtype-specific prognostic lncRNA expression signatures for esophageal cancer and HNSCC.

Results: By incorporating biological relevancy information, the lncRNA lists identified by the Cox-filter method were further refined. The resulting signatures include genes that are highly related to cancer, such as H19 and NEAT1, which possess perfect prognostic values for esophageal cancer and HNSCC, respectively.

Conclusions: The Cox-filter method is indeed a handy tool to identify subtype-specific prognostic lncRNA signatures. We anticipate the method will gain wider applications.

Keywords: Long non-coding RNA (lncRNA), Prognostic signature, Head and neck squamous cell carcinoma (HNSCC), Esophageal cancer, Cox regression model

Background

Esophageal cancer is a cancer of the esophagus, the hollow tube that carries foods and liquids from throat to stomach. The causes of esophageal cancer are unclear, but it is commonly believed that both environmental and genetic factors play roles in its initiation and progression [1]. For instance, smoking, heavy alcohol consumption, obesity, and damage to the esophagus from acid reflux (Barrett esophagus) are thought to increase

the risk of developing esophageal cancer, while, the tendency of familial aggregation for esophageal cancer suggests that genetic components are of crucial importance. The most common histological subtypes of esophageal cancer are squamous cell carcinoma (ESCC) and adenocarcinoma (EAC). As far as prognosis is concerned, no evidence suggests any substantial difference between these two subtypes. Nevertheless, a study by The Cancer Genome Atlas research group [2] has demonstrated that non-marginal differences with regard to gene expression and somatic alteration exist between ESCC and EAC. Consequently, biomarkers that hold prognostic value for these two subtypes are expected to be distinct, at least to some extent.

* Correspondence: windytian@hotmail.com; yudan19792003@163.com

¹Division of Clinical Research, The First Hospital of Jilin University, 1Xinmin Street, Changchun, Jilin 130021, People's Republic of China

⁵Department of Otolaryngology Head and Neck Surgery, The Second Hospital of Jilin University, 218 Ziqiang Road, Changchun, Jilin 130041, People's Republic of China

Full list of author information is available at the end of the article



Head and neck squamous cell carcinoma (HNSCC) develops in mucous membranes of the mouth, nose and throat. Hypopharyngeal squamous cell carcinoma (HSCC), which originates in mucosa of the hypopharynx and accounts for approximately 3% of HNSCC cases, has one of the poorest prognoses among HNSCC patients [3]. Laryngeal squamous cell cancer (LSCC) accounts for relatively more HNSCC cases and has a better prognosis compared to HSCC even though the initial sites of these two diseases are anatomically very close. LSCC originates in the larynx, whereas HSCC originates in the lower part of the throat near the larynx (i.e., the hypopharynx). Therefore, finding molecular markers that can distinguish between the two subtypes is crucial for survival prediction.

Long non-coding RNAs (lncRNAs) are a class of RNA molecules that have a length of more than 200 nucleotides and are without protein-coding capacity [4]. Therefore, lncRNAs have previously been regarded as transcriptional “junk.” Nowadays, paramount investigations have demonstrated that lncRNAs can serve as novel biomarkers and therapeutic targets in complex diseases such as cancer. Identification of lncRNA signatures is in demand and usually requires the help of a feature selection method. The primary aims of feature selection are to reduce the number of features (e.g., genes or metabolites) under consideration to a manageable size, thus speeding up the learning process and facilitating biological interpretation and experimental validation [5].

Applying feature selection to lncRNA (vs mRNA) data might achieve better model parsimony because mRNA-based studies obtain signatures with a limited number of genes, and because the expression levels of lncRNAs are usually lower than those of mRNAs (thus less differentially expressed lncRNAs can be identified). Studies that aim to identify lncRNA signatures for esophageal cancers and HNSCC have increased dramatically. For example, studies by Cao et al. [6], Wang et al. [7] and Yao et al. [8] specifically aimed to identify lncRNA expression signatures with prognostic value for HNSCC patients, while several studies [9–12] identified relevant lncRNA signatures for esophageal cancer. Nevertheless, those studies usually considered HNSCC or esophageal cancer as a whole or only focused on one specific subtype.

In this study, we applied a novel feature selection method – the Cox-filter method [13] – to the cancer genome atlas (TCGA) esophageal cancer and HNSCC RNA-Seq data, with the objectives of constructing subtype-specific prognostic lncRNA expression signatures for EC and HNSCC. Precision medicine for those patients will only be possible once subtype-specific prognostic signatures become available.

Materials and methods

Experimental data

The lncRNA expression values, namely, FPKM (fragments per kilo-bases per million) for HNSCC were retrieved from the TANRIC (The Atlas of ncRNA in Cancer) database [14], version 1.0.6 (<https://www.tanric.org/>), which was last updated on 07/29/2015. Then the corresponding clinical information was retrieved from the the Genomic Data Commons (<https://gdc.cancer.gov/>) by matching the barcode IDs of samples in the TANRIC database [14] with those in the TCGA database. Patients without information on overall survival (OS), age, gender, pathological tumor stage and histological subtype were discarded. Only patients with LSCC and HSCC were retained for analysis. If the sum of FPKM values of lncRNA expression across all samples (LSCC and HSCC patients combined) was < 4, they were deleted. Finally, log 2 transformations on (FPKM counts + 1) were carried out, providing a better approximation to a normal distribution.

For the esophageal cancer study, both the expression profiles (RNA-Seq data) of TCGA ESCA cohort and clinical information such as overall survival time were downloaded from the Genomic Data Commons. Subsequently, the lncRNAs were collected by mapping the Ensemble IDs of RNA-Seq data to those in the TANRIC database [14] (given that the ESCA cohort is not included in the TANRIC database) so that expression profiles of lncRNAs were obtained.

The ratio of LSCC and HSCC is extremely high (89:6) while that for ESCC to EAC is very close to 1 (81:83), which represents the two extreme cases (huge imbalance of sample ratios versus perfect balance of sample ratios). Hence, using these two datasets, it is possible to examine the influence of subgroup size imbalance on the performance of a feature selection algorithm. The demographical characteristics of these two datasets are presented in Table 1.

Statistical methods

The Cox-filter method proposed by Tian et al. [13] screens genes one by one according to the significance level of the corresponding coefficients in a Cox model. Under the two-class cases (the model can easily be extended to multiple-class cases), the corresponding Cox model may be written as,

$$\lambda_{ijg}(t) = \lambda_{0g}(t) \exp(\beta_{1g} I_i(j = c_2) + \beta_{2g} X_{ijg} + (\beta_{1g} I_i(j = c_2) \times X_{ijg}))$$

Tian et al. [13] provided a detailed description of the definitions of parameters (i.e., β s and λ s) and a graphical illustration of all possible scenarios; those details are not presented here. For the current study, the features under

Table 1 Characteristics of head and neck squamous cell carcinoma and esophageal cancer data

	Patients (#)	Deaths (#)	Median survival time (days)	p-value (log-rank test)
Esophageal cancer				
Squamous Cell Carcinoma [ESCC]	81	29	763	
Adenocarcinoma [EAC]	83	38	801	0.721
Head and Neck Squamous Cell Carcinoma [HNSCC]				
Laryngeal [LSCC]	89	25	1838	
Hypopharyngeal [HSCC]	6	2	–	0.839

The log-rank tests indicated histological subtype but had no prognostic value for esophageal cancer or HNSCC. For esophageal cancer, this is consistent with previous results. For HNSCC the discrepancy may be attributable to the small sample size of HSCC subtype

consideration are lncRNAs, subtype-specific prognostic lncRNAs were those for which either β_{2g} or $(\beta_{2g} + \beta_{3g})$ is significantly different from zero. More specifically, $\beta_{2g} \neq 0$ implies that lncRNA g has a prognostic value for subgroup c_1 while $(\beta_{2g} + \beta_{3g}) \neq 0$ implies lncRNA g has a prognostic value for subgroup c_2 . Therefore, β_{2g} and β_{3g} are the parameters of interest and their significance levels determine if subtype-specific lncRNAs exist.

Statistical language and packages

All statistical analyses were carried out in the R language, version 3.5 (www.r-project.org).

Results

By applying the Cox-filter model to esophageal cancer data and setting the cutoff of adjusted p -values for these linear coefficients at 0.05, we identified 200 lncRNAs that have prognostic values for EAC and 96 for ESCC. Among them, there were 46 overlaps. We searched the GeneCards database for their biological relevance. For EAC, after removing 19 genes that are not be recognized by the GeneCards database (www.genecards.org), 58

lncRNAs were indicated to be directly related to cancers. For ESCC, 19 lncRNAs are unrecognizable as well. Among the remaining 77 lncRNAs, 27 of them were directly related to cancers. A Venn-diagram (Fig. 1) was made and the gene symbols were given, stratified by EAC-specific lncRNAs, ESCC-specific lncRNAs and overlapped lncRNAs between two subtypes. Among these unique 74 lncRNAs, 44 were regarded as being differentially expressed between cancer tissues and normal tissues.

For HNSCC, using a cutoff of 0.05 for adjusted p -values the Cox-filter method identified 126 LSCC lncRNAs (20 non-identifiable in the GeneCards database) and 89 HSCC lncRNAs (30 of which are non-identifiable in the GeneCards database). Fifty-six were directly related to cancers for LSCC and 16 for HSCC. Among these lncRNAs, 6 lncRNAs were shared by these two subtypes, and 44 lncRNAs were regarded as being differentially expressed between cancer tissues and normal tissues. Figure 2 presents gene symbols of those lncRNAs. From the gene symbols given in Figs. 1 and 2, we observed several microRNAs (e.g., MIR146A and

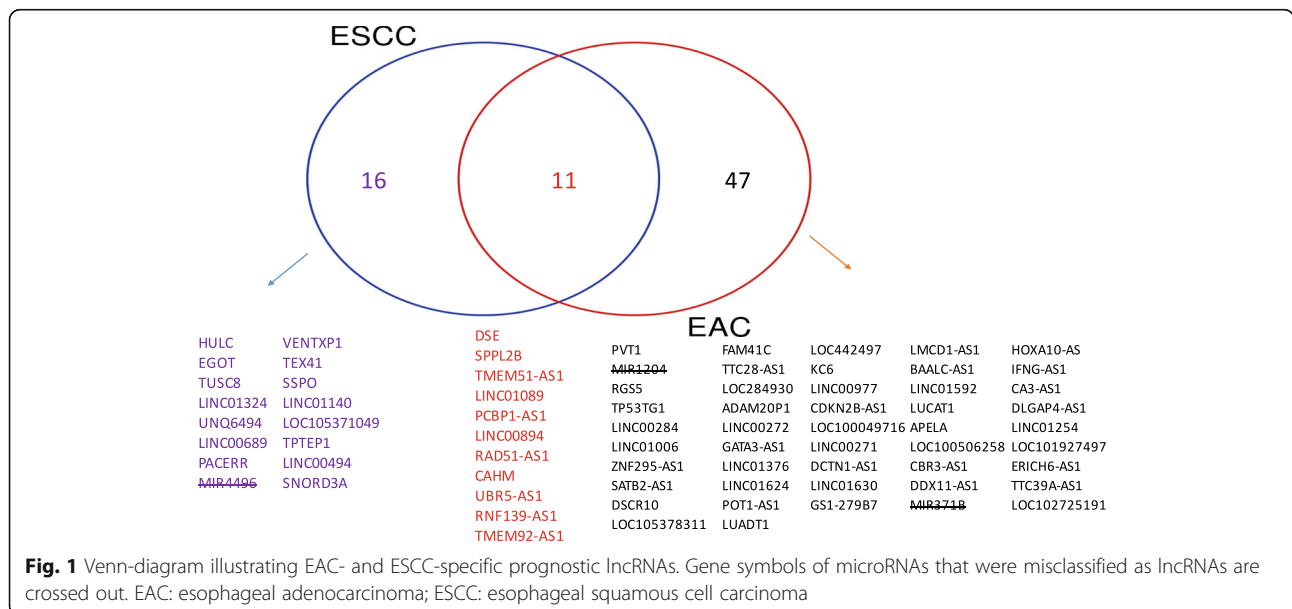
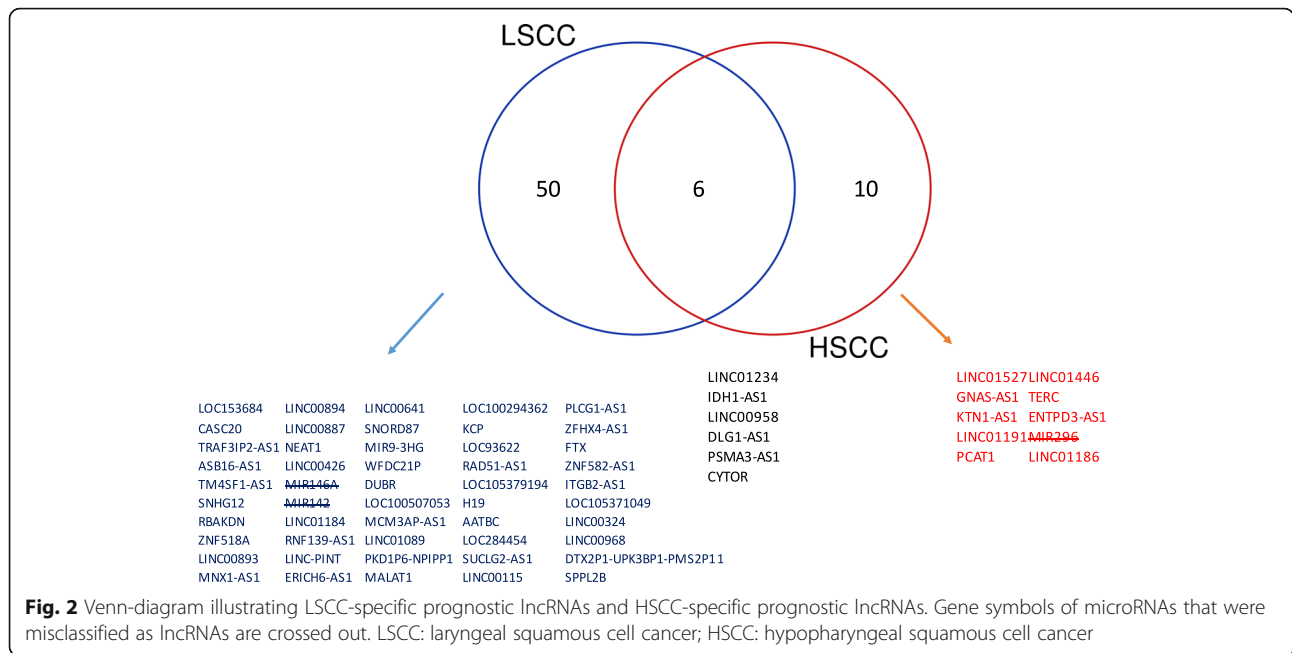


Fig. 1 Venn-diagram illustrating EAC- and ESCC-specific prognostic lncRNAs. Gene symbols of microRNAs that were misclassified as lncRNAs are crossed out. EAC: esophageal adenocarcinoma; ESCC: esophageal squamous cell carcinoma



MIR 296) that were mistakenly recognized as lncRNAs by the TANRIC database. Since TANRIC has not been updated since its initiation, it is natural to expect such errors. In the following results, those microRNAs were removed manually.

Discussion

In this study, Pvt1 oncogene (PVT1) with a confidence score of 25.4 is ranked on the second place for the EAC-specific prognostic lncRNAs. Based on the strategy of competitive endogenous RNA (ceRNA) networks [15], overexpression of PVT1 correlates with a poor prognosis [16] or a fast tumor progression [17] in esophageal cancer patients or in ESCC [18] In this study, PVT1 was indicated as an EAC-specific lncRNA since it does not belong to the intersection set between lncRNAs for these two subtypes.

CDKN2B antisense RNA 1 (CDKN2B-AS1), also known as ANRIL, was on the top of this list (i.e., cancer related EAC-specific prognostic lncRNAs), however, only three studies [19–21] have addressed its association with esophageal cancer. While the first two studies explored the association between CDKN2B-AS1 and esophageal cancer by way of genetic mutations, the third did so from the prospective of expression level. Other than esophageal cancer, CDKN2B-AS1 had been linked to a variety of cancer types such as acute lymphoblastic leukemia [22], gastric cancer [20, 23] and hepatocellular carcinoma (HCC) [24]. For other top-ranked lncRNAs, Yoon et al. [25] have demonstrated that LUCAT1 was over-expressed in tumor tissues compared to paired normal tissues and may promote carcinogenesis of ESCC.

Another recent study [26] has shown that up-regulation of CBR3-AS1 promoted cell proliferation and was positively correlated with pathologic stages of ESCC. Lastly, despite the absence of literature suggesting that TP53TG plays any role in the development and progression of esophageal cancer, this lncRNA can suppress tumor growth and is of importance for the correct response of P53 to DNA damage [27]. In addition, the association of TP53TG with other cancer types such as glioma and lung cancer has been reported in previous studies.

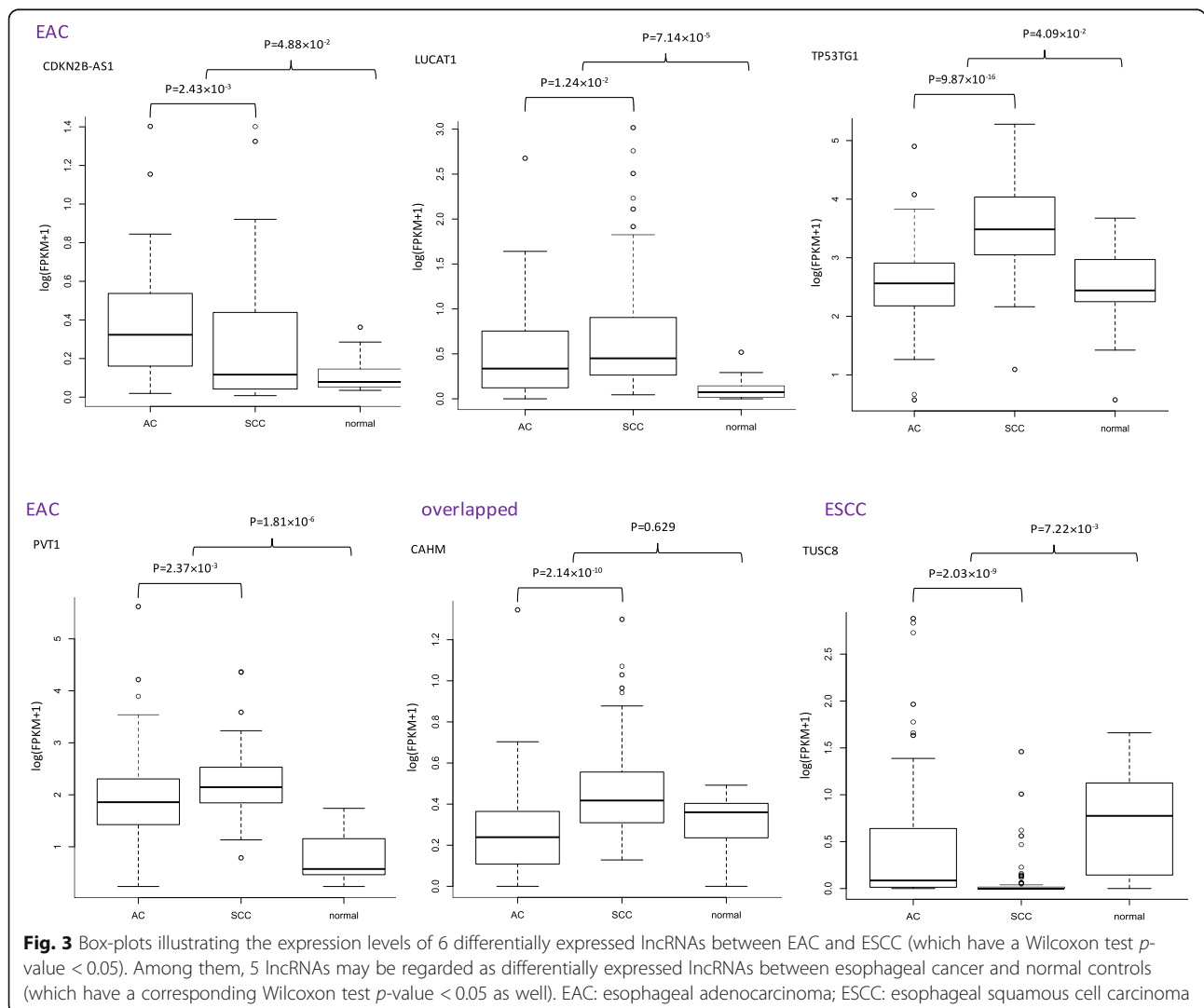
Besides the lower prevalence of lncRNA studies on EAC, another possible explanation for the links of top-5-ranked lncRNAs with ESCC instead of EAC is that racial disparities of ESCC between Asian and Caucasian populations existed at the molecular level [28]. Then, it is natural to observe a link between PVT1 and ESCC during the literature mining considering those studies were all carried out in East Asia. In contrast, our work is based on the TCGA RNA-Seq data in which most patients are Whites.

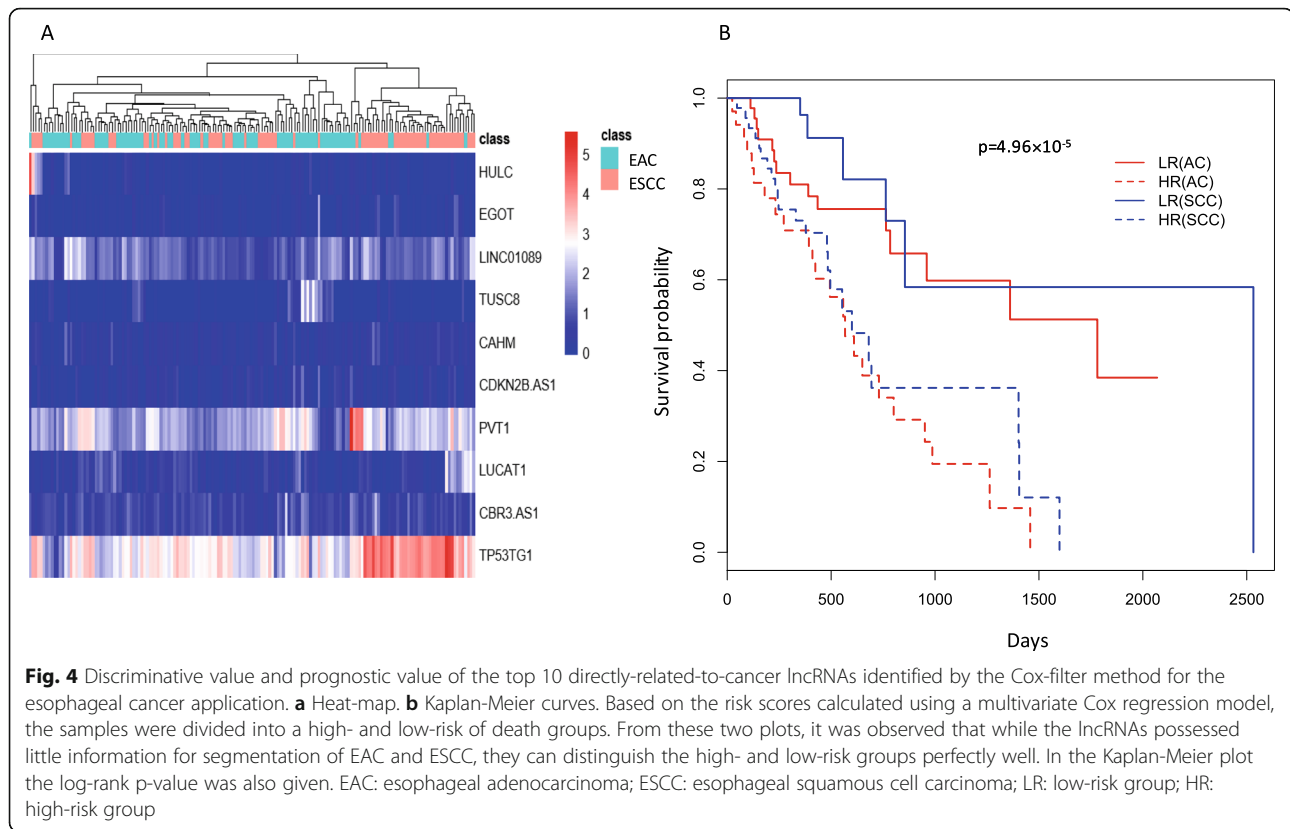
On the other hand, for the top 5 directly-related-to-cancer lncRNAs for the ESCC, only two studies provided experimental supports on the association of HULC [29] and EGOT [30] with esophageal cancer. For the remaining three lncRNAs – LINC01089, TUSC8 and CAHM — the lncRNADisease2 database [31] used computational methods and predicted they are associated with gastric cancer. Even though the identified lncRNAs are related to a variety of cancers, more focus on their correlations with ESCC and EAC are in demand. The expression levels of those 10 lncRNAs were compared between ESCC and EAC, between esophageal

cancer tissues and normal tissues using Wilcoxon tests. Among them, 6 (4 were specific for EAC, 1 for ESCC and 1 shared by both subtypes) had a corresponding p -value < 0.05 and may be considered as the differentially expressed lncRNAs between EAC and ESCC (Fig. 3). All these 6 lncRNAs except CAHM had corresponding Wilcoxon test p -values < 0.05 in the comparison of tumor tissues and normal tissues as well (Fig. 3). Nevertheless, as shown in Fig. 4a, these 10 lncRNAs hold very limited discriminative capacity to separate EAC from ESCC. In contrast, they can predict the prognosis status perfectly. In Fig. 4b, Kaplan-Meier curves were plotted for high-risk and low-risk groups (stratified according to the estimated risk scores of the multivariate Cox-regression model with these 10 lncRNAs as covariates), and then a log-rank test was performed to compare these survival curves. From Fig. 4b, we observed that within each subtype, the difference between the high-risk and low-risk groups was significant while within each risk group

(between subtype), the difference was less or not significant. This result is expectable given that the outcomes (i.e., dependent variables) considered in the segmentation of subtypes and prognosis prediction are distinct.

Among the overlapped 11 lncRNAs, in addition to that CAHM was experimentally validated to be associated with colorectal cancer by a qPCR study [32] and astrocytoma [33] by a microarray study, TMEM51-AS1 was with chromophobe renal cell carcinoma [34] and liver cancer [35] by qPCR studies, RAD51-AS1 was with only ovarian epithelial cancer [36], RNF139-AS1 was with only astrocytoma [37] and LINC01089 with breast cancer [38] by qPCR and astrocytoma [33] by a microarray study, all except DSE and SPPL2B (which is not recorded on LncRNADisease2 database) were predicted to be correlated with a variety of cancers such as gastric cancer by the LncRNADisease2 database. Further studies are warranted to investigate the roles that the identified lncRNAs (including overlapped ones and unique-to-





subtype ones by integrating the Cox-filter method and biological relevancy) may play during the development and progression of esophageal cancer.

For LSCC prognostic lncRNAs, H19, MALAT1, NEAT1, CYTOR and SNHG12 were ranked as the first five of this directly-related-to-cancer list. For HSCC, TERC, PCAT1, CYTOR, LINC01234 and LINC00958 made to the list. H19 is a well-known oncogene and acts as a driving force in a variety of cancers. For HNSCC specifically, a study by Guan et al. [39] demonstrated that overexpression of H19 is associated with tumor recurrence and poor prognosis by performing an experiment including 62 HNSCC patients (46 with LSCC and 14 with HSCC). A very recent study [40] also showed that the expression level of H19 was higher in patients with metastasized (vs non-metastasized) tongue squamous cell carcinoma, and was higher in tumor cells than normal squamous cells.

MALAT1 was found to be overexpressed in tumor tissues of oral squamous cell carcinoma (OSCC) patients by a real-time PCR experiment carried out by Zhou et al. [41]. Chang et al. [42] showed that inhibition of MALAT1 can prevent OSCC proliferation whereas its overexpression can promote OSCC. According to the ceRNA network, MALAT1 is a microRNA sponge of miR-125b of which STAT3 is predicted as a binding target. In addition, two studies [43, 44] provided

experimental supports for the association of MALAT1 and tongue squamous cell carcinoma. Using qRT-PCR, Wang et al. [45] examined and compared the expression level of NEAT1 in LSCC and adjacent non-neoplastic tissues and showed that NEAT1 was significantly overexpressed in LSCC. Hence, they concluded that “NEAT1 plays an oncogenic role in the tumorigenesis of LSCC.”

CYTOR, also known as LINC00152, was proved experimentally to be associated with progression and prognosis of tongue squamous cell carcinoma [46] and HNSCC [47]. Using TCGA RNA-Seq data and some bioinformatics tools, Guo et al. [48] identified CYTOR as an HNSCC-associated lncRNA and determined that its expression is positively correlated with lymph node metastasis and risk of death. Subsequently, its function was explored by cell-based experiments which suggested that CYTOR inhibited cell apoptosis after the treatment with chemotherapeutic drug diamminedichloroplatinum (DDP). Furthermore, acting as the microRNA sponge of miR-19-5p that combines with the 3'UTR region of WWP1, overexpression of SNHG12 may promote proliferation and invasion of LSCC [49]. In our analysis, CYTOR was shared by both LSCC and HSCC subtypes.

Even though no experimental evidence or computational prediction links TERC with HNSCC in the LncRNADisease2.0 database [31], literature mining in the PubMed database identified several studies to

support their association. For LSCC specifically, Liu et al. [50] detected TERC gene amplification in precancerous and cancerous tissues using fluorescent in situ hybridization. In a recent study [51], the expression values of PCAT1 in paired HNSCC tissues and adjacent non-tumor tissues were measured using qRT-PCR. The results showed that PCAT1 was over-expressed in the tumor tissues, which consisted with the results given by the online bioinformatics tool, GEPIA (<http://gepia.cancer-pku.cn>). In addition, that study also proved that after the knockdown of PCAT1, p38 MAPK and apoptosis signal-regulating kinase 1 which induced Caspase 9 and PARG mediated apoptosis were activated.

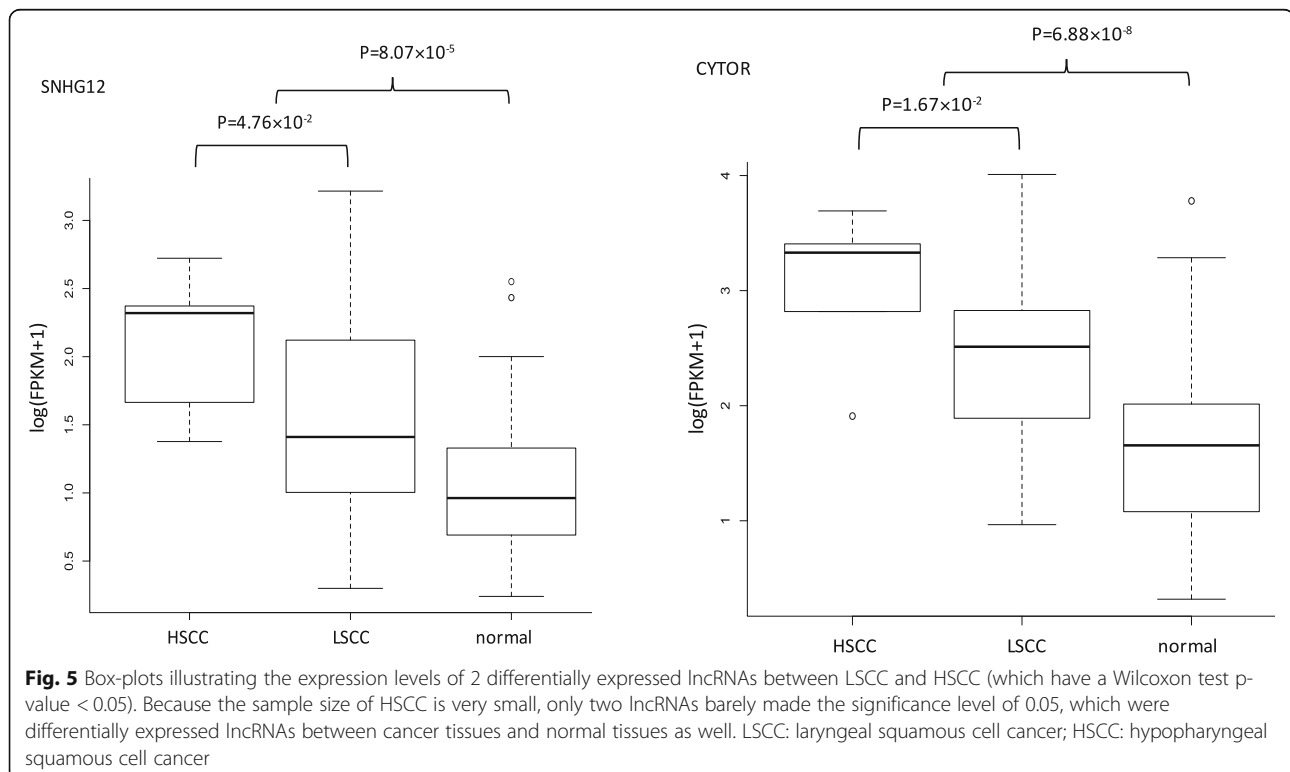
For the last two HSCC-specific lncRNAs, namely, LINC01234 and LINC00958, no evidence has been found to link them with HNSCC in either the LncRNA-Disease2 database (experimentally or computationally) or the PubMed literature search. Both of these genes overlapped the HSCC and LSCC subtypes. Likewise, for the final 3 overlapped lncRNAs, no support for a link with HNSCC can be found. Further studies are warranted.

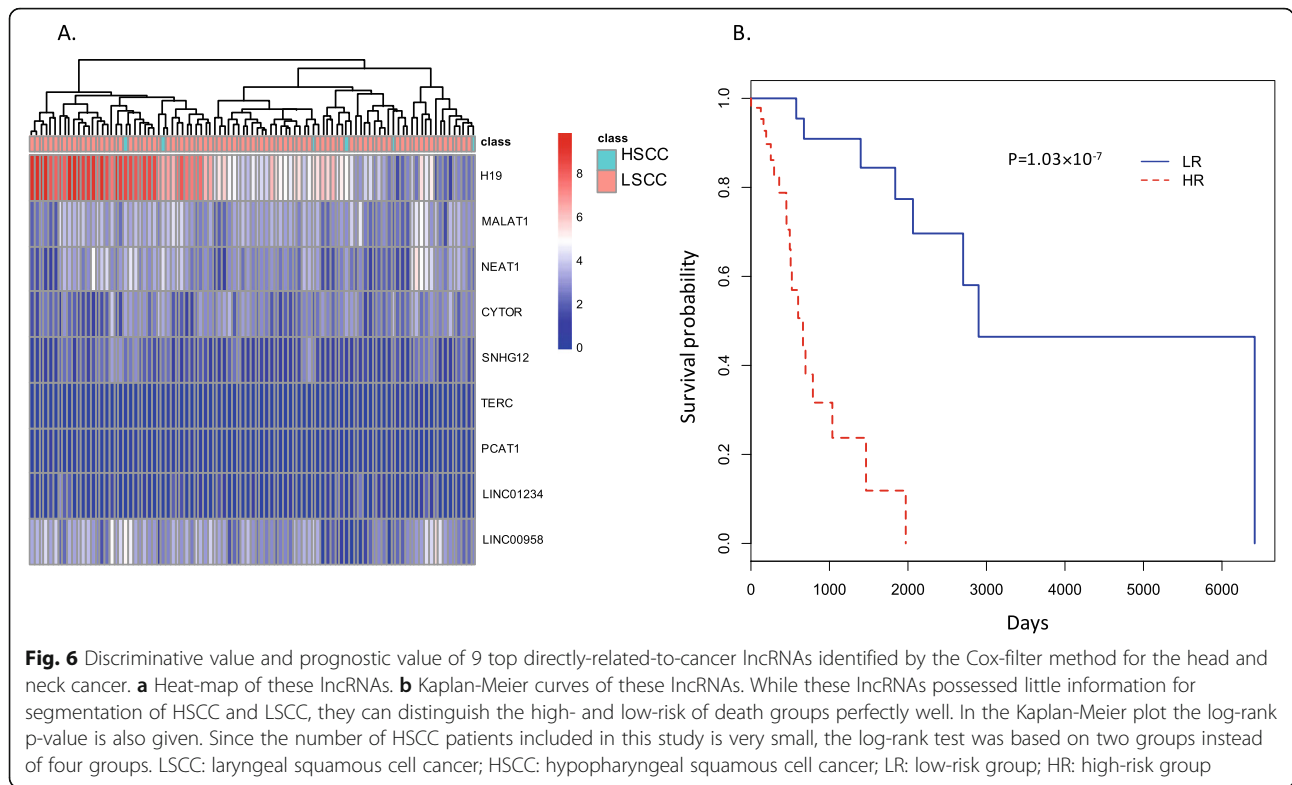
Among these 9 unique directly-related-to-cancer lncRNAs, only CYTOR and SNHG12 have Wilcoxon test p -values < 0.05 (Fig. 5) and may be loosely regarded as differentially expressed genes between LSCC and HSCC subtypes, and between cancer tissues and normal tissues. The small sample size of HSCC in this analysis

may explain the results to some degree. Similar to the results of esophageal cancer application, while these 9 lncRNAs cannot distinguish LSCC and HSCC, they do have prognostic value for predicting the risk of death for HNSCC patients (here, LSCC and HSCC were examined together given there were only 6 HSCC patients in this study). Corresponding heat-map and Kaplan-Meier curves are presented in Fig. 6. Lastly, the regulated mRNAs by the identified lncRNAs were retrieved from the lncRNADisease 2.0 database [31] and the pathway enrichment analysis was carried out using the String database [52]. The enriched GO terms and KEGG pathways for these four subtypes are presented in Table 2, from which we observe that no overlaps among these four subtypes occur.

Conclusions

The Cox-filter method is among the first efforts to develop feature selection algorithms capable of identifying prognostic genes specifically for different subtypes. When applied to gene expression profiles, it achieved satisfactory performance. In this study, we show that this method is applicable to lncRNA expression profiles, as illustrated by the two real-world applications in which the Cox-filter method identified many lncRNAs with meaningful implication with cancer. The ratio of the two distinct subtypes in these applications represent extreme cases: one with good balance case and one with bad





balance. The Cox-filter method can easily deal with the first case. In the second case, it can still estimate the significance level of lncRNAs in minor subtypes by borrowing some information from the dominant subtype. Therefore, the Cox-filter method is a handy tool to construct subtype-specific prognostic lncRNA signatures, indeed.

The big drawback of the Cox-filter method is inclusion of many false positives in the final models. To address this drawback, several extensions that incorporate biological information and prioritize genes with high connectivity levels have been proposed [53, 54]. When

applying to lncRNA profiles, the issue is still apparent and thus needs to be addressed as well. However, those extensions cannot be adopted to the lncRNA expression profiles directly because the biological pathway information was retrieved from a knowledgebase such as String [52] or HPRD [55], which focus on mRNAs (protein coding genes). A statistical model (e.g., the WGCNA method [56] with the capacity of constructing co-expression networks/modules is needed before implementing such Cox-filter extensions. Nevertheless, by combining biological relevancy information from the GeneCards database, we further refined the lncRNA lists

Table 2 Enriched pathway analysis for the mRNAs regulated by selected lncRNAs

	GO-BP	GO-CC	GO- MF	KEGG
LSCC	-	-	-	-
HSCC	peptide cross-linking, skin development, epidermis development, keratinization, epithelium development, cornification	cornified envelope		
ESCC	-	nuclear envelope, organelle envelope, organelle, membrane-bounded organelle, intracellular membrane-bounded organelle, organelle part, intracellular organelle part, organelle membrane, intracellular organelle, nuclear part	-	-
EAC	Nephron tubule development	-	-	Nitrogen metabolism

no enriched pathways. LSCC laryngeal squamous cell carcinoma, HSCC hypopharyngeal squamous cell carcinoma, ESCC esophageal squamous cell carcinoma, EAC esophageal adenocarcinoma, GO-BP gene ontology biological process category, GO-CC gene ontology cellular component category, GO-MF gene ontology molecular function category, KEGG Kyoto encyclopedia of genes and genomes pathways

identified by the Cox-filter method, and the resulting lncRNA signatures have been demonstrated to possess perfect prognostic value.

Abbreviations

ceRNA: Competitive endogenous RNA; EAC: Esophageal adenocarcinoma; ESCC: Esophageal squamous cell carcinoma; FPKM: Fragments per kilo-bases per million; HCC: Hepatocellular carcinoma; HNSCC: Head and neck squamous cell carcinoma; HSCC: Hypopharyngeal squamous cell carcinoma; lncRNA: Long non-coding RNAs; LSCC: Laryngeal squamous cell; TANRIC: The atlas of ncRNA in cancer; TCGA: The cancer genome atlas

Acknowledgements

The Markey Cancer Center's Research Communications Office assisted with preparation of the manuscript.

Authors' contributions

Conceived and designed the study: ST DY. Analyzed the data: ST CW JZ. Interpreted data analysis and results: ST DY. Wrote the paper: ST JZ CW DY. All authors reviewed and approved the final manuscript.

Funding

This study was supported by the Education Department of Jilin Province (grant No. JJKH20190032KJ) and the Natural Science Foundation of China (grant No. 31401123).

Availability of data and materials

Data for head and neck squamous cell carcinoma (the HNSC cohort) were downloaded from The Atlas of ncRNA in Cancer (TANRIC) database (<https://www.tanric.org/>), and data for esophageal cancer (the ESCA cohort) were downloaded from the Genomic Data Commons of The Cancer Genome Atlas (<https://gdc.cancer.gov>).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Division of Clinical Research, The First Hospital of Jilin University, 1Xinmin Street, Changchun, Jilin 130021, People's Republic of China. ²Department of Biostatistics, College of Public Health, University of Kentucky, 800 Rose St, Lexington, KY 40536, USA. ³Markey Cancer Center, University of Kentucky, 800 Rose St, Lexington, KY 40536, USA. ⁴School of Life Science, 2699 Qianjin Street, Changchun, Jilin 130012, People's Republic of China. ⁵Department of Otolaryngology Head and Neck Surgery, The Second Hospital of Jilin University, 218 Ziqiang Road, Changchun, Jilin 130041, People's Republic of China.

Received: 16 September 2019 Accepted: 29 January 2020

Published online: 05 February 2020

References

- Huang F-L, Yu S-J. Esophageal cancer: risk factors, genetic association, and treatment. *Asian J Surg.* 2018;41:210–5.
- Kim JY, Bowlby R, Mungall AJ, Robertson AG, Odze RD, Cherniack AD, et al. Integrated genomic characterization of oesophageal carcinoma. *Nature.* 2017;541:169–74.
- Garneau JC, Bakst RL, Miles B. Hypopharyngeal cancer: a state of the art review. *Oral Oncol.* 2018;86:244–50.
- St Laurent G, Wahlestedt C, Kapranov P. The landscape of long noncoding RNA classification. *Trends Genet.* 2015;31:239–51.
- Saeyns Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics.* 2007;23:2507–17.
- Cao W, Liu J, Liu Z, Wang X, Han Z-G, Ji T, et al. A three-lncRNA signature derived from the atlas of ncRNA in cancer (TANRIC) dA three-lncRNA expression signature predicts survival in head and neck squamous atabase predicts the survival of patients with head and neck squamous cell carcinoma. *Oral Oncol.* 2017;65:94–101. <https://doi.org/10.1016/j.oraloncology.2016.12.017>.
- Wang P, Jin M, Sun CH, Yang L, Shan LY, Wang X, et al. A three-lncRNA expression signature predicts survival in head and neck squamous cell carcinoma (HNSCC). *Biosci Rep.* 2018;38:BSR20181528.
- Yao Y, Chen X, Lu S, Zhou C, Xu G, Yan Z, et al. Circulating long noncoding RNAs as biomarkers for predicting head and neck squamous cell carcinoma. *Cell Physiol Biochem.* 2018;50:1429–40.
- Huang GW, Xue YJ, Wu ZY, Xu XE, Wu JY, Cao HH, et al. A three-lncRNA signature predicts overall survival and disease-free survival in patients with esophageal squamous cell carcinoma. *BMC Cancer.* 2018;18:147.
- Li CQ, Huang GW, Wu ZY, Xu YJ, Li XC, Xue YJ, et al. Integrative analyses of transcriptome sequencing identify novel functional lncRNAs in esophageal squamous cell carcinoma. *Oncogenesis.* 2017;6:e297.
- Li J, Chen Z, Tian L, Zhou C, He MY, Gao Y, et al. lncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut.* 2014;63:1700–10.
- Mao Y, Fu Z, Zhang Y, Dong L, Zhang Y, Zhang Q, et al. A seven-lncRNA signature predicts overall survival in esophageal squamous cell carcinoma. *Sci Rep.* 2018;8:1–10.
- Tian S, Wang C, An M-W. Test on existence of histology subtype-specific prognostic signatures among early stage lung adenocarcinoma and squamous cell carcinoma patients using a cox-model based filter. *Biol Direct.* 2015;10:15.
- Wu HUA, Yu DH, Wu MH, Huang T. Long non-coding RNA LOC541471: a novel prognostic biomarker for head and neck squamous cell carcinoma. *Oncol Lett.* 2019;17:2457–64.
- Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell.* 2011;146:353–8.
- Chen LP, Wang H, Zhang Y, Chen QX, Lin TS, Liu ZQ, et al. Robust analysis of novel mRNA-lncRNA cross talk based on ceRNA hypothesis uncovers carcinogenic mechanism and promotes diagnostic accuracy in esophageal cancer. *Cancer Manag Res.* 2019;11:347–58.
- Shen S, Li K, Liu Y, Yang C, He C, Wang H. Down-regulation of long noncoding RNA PVT1 inhibits esophageal carcinoma cell migration and invasion and promotes cell apoptosis via microRNA-145-mediated inhibition of FSCN1. *Mol Oncol.* 2019;13:2554–73.
- Yang S, Ning Q, Zhang G, Sun H, Wang Z, Li Y. Construction of differential mRNA-lncRNA crosstalk networks based on ceRNA hypothesis uncover key roles of lncRNAs implicated in esophageal squamous cell carcinoma. *Oncotarget.* 2016;7:85728–40.
- Lin X, Yan C, Gao Y, Du J, Zhu X, Yu F, et al. Genetic variants at 9p21.3 are associated with risk of esophageal squamous cell carcinoma in a Chinese population. *Cancer Sci.* 2017;108:250–5.
- Li W-Q, Pfeiffer RM, Hyland PL, Shi J, Gu F, Wang Z, et al. Genetic polymorphisms in the 9p21 region associated with risk of multiple cancers. *Carcinogenesis.* 2014;35:2698–705.
- Hu Z, Wu H, Li Y, Hou Q, Wang Y, Li S, et al. β -Elemene inhibits the proliferation of esophageal squamous cell carcinoma by regulating long noncoding RNA-mediated inhibition of hTERT expression. *Anticancer Drugs.* 2015;26:531–9.
- Iacobucci I, Sazzini M, Garagnani P, Ferrari A, Boattini A, Lonetti A, et al. A polymorphism in the chromosome 9p21 ANRIL locus is associated to Philadelphia positive acute lymphoblastic leukemia. *Leuk Res.* 2011;35:1052–9. <https://doi.org/10.1016/j.leukres.2011.02.020>.
- Zhang E, Kong R, Yin D, You L, Sun M, Han L, et al. Long noncoding RNA ANRIL indicates a poor prognosis of gastric cancer and promotes tumor growth by epigenetically silencing of miR-99a/miR-449a. *Oncotarget.* 2014;5:2276–92.
- Huang Y, Xiang B, Liu Y, Wang Y, Kan H. lncRNA CDKN2B-AS1 promotes tumor growth and metastasis of human hepatocellular carcinoma by targeting let-7c-5p/NAP1L1 axis. *Cancer Lett.* 2018;437:56–66.
- Yoon J-H, You B-H, Park CH, Kim YJ, Nam J-W, Lee SK. The long noncoding RNA LUCAT1 promotes tumorigenesis by controlling ubiquitination and stability of DNA methyltransferase 1 in esophageal squamous cell carcinoma. *Cancer Lett.* 2018;417:47–57.
- Wang C-M, Wu Q-Q, Li S-Q, Chen F-J, Tuo L, Xie H-W, et al. Upregulation of the long non-coding RNA PlncRNA-1 promotes esophageal squamous carcinoma cell proliferation and correlates with advanced clinical stage. *Dig Dis Sci.* 2014;59:591–7.

27. Diaz-Lagares A, Crujeiras AB, Lopez-Serra P, Soler M, Setien F, Goyal A, et al. Epigenetic inactivation of the p53-induced long noncoding RNA TP53 target 1 in human cancer. *Proc Natl Acad Sci U S A*. 2016;113:E7535–44.
28. Deng J, Chen H, Zhou D, Zhang J, Chen Y, Liu Q, et al. Comparative genomic analysis of esophageal squamous cell carcinoma between Asian and Caucasian patient populations. *Nat Commun*. 2017;8:1533.
29. Kang M, Sang Y, Gu H, Zheng L, Wang L, Liu C, et al. Long noncoding RNAs POLR2E rs3787016 C/T and HULC rs7763881 a/C polymorphisms are associated with decreased risk of esophageal cancer. *Tumour Biol*. 2015;36:6401–8.
30. Xu S-P, Zhang J-F, Sui S-Y, Bai N-X, Gao S, Zhang G-W, et al. Downregulation of the long noncoding RNA EGOT correlates with malignant status and poor prognosis in breast cancer. *Tumour Biol*. 2015;36:9807–12.
31. Bao Z, Yang Z, Huang Z, Zhou Y, Cui Q, Dong D. LncRNADisease 2.0: An updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res*. 2019;47:D1034–7.
32. Pedersen SK, Mitchell SM, Graham LD, McEvoy A, Thomas ML, Baker RT, et al. CAHM, a long non-coding RNA gene hypermethylated in colorectal neoplasia. *Epigenetics*. 2014;9:1071–82.
33. Zhi F, Wang Q, Xue L, Shao N, Wang R, Deng D, et al. The use of three long non-coding RNAs as potential prognostic indicators of astrocytoma. *PLoS One*. 2015;10:e0135242.
34. He H-T, Xu M, Kuang Y, Han X-Y, Wang M-Q, Yang Q. Biomarker and competing endogenous RNA potential of tumor-specific long noncoding RNA in chromophobe renal cell carcinoma. *Onco Targets Ther*. 2016;9:6399–406.
35. Zhu J, Liu S, Ye F, Shen Y, Tie Y, Zhu J, et al. The long noncoding RNA expression profile of hepatocellular carcinoma identified by microarray analysis. *PLoS One*. 2014;9:e101707.
36. Zhang X, Liu G, Qiu J, Zhang N, Ding J, Hua K. E2F1-regulated long non-coding RNA RAD51-AS1 promotes cell cycle progression, inhibits apoptosis and predicts poor prognosis in epithelial ovarian cancer. *Sci Rep*. 2017;7:4469.
37. Mahmoudvand H, Kheirandish F, Ghasemi Kia M, Tavakoli Kareshk A, Yarahmadi M. Chemical composition, protoscolicidal effects and acute toxicity of *Pistacia atlantica* Desf. fruit extract. *Nat Prod Res*. 2016;30:1208–11.
38. Sas-Chen A, Aure MR, Leibovich L, Carvalho S, Eneka Y, Korner C, et al. LIMT is a novel metastasis inhibiting lncRNA suppressed by EGF and downregulated in aggressive breast cancer. *EMBO Mol Med*. 2016;8:1052–64.
39. Villalba M, Lopez L, Redrado M, Ruiz T, de Aberasturi AL, de la Roja N, et al. Development of biological tools to assess the role of TMRSS4 and identification of novel tumor types with high expression of this prometastatic protein. *Histol Histopathol*. 2017;32:929–40.
40. Kou N, Liu S, Li X, Li W, Zhong W, Gui L, et al. H19 facilitates tongue squamous cell carcinoma migration and invasion via sponging miR-let-7. *Oncol Res*. 2019;27:173–82.
41. Zhou X, Liu S, Cai G, Kong L, Zhang T, Ren Y, et al. Long non coding RNA MALAT1 promotes tumor growth and metastasis by inducing epithelial-Mesenchymal transition in Oral squamous cell carcinoma. *Sci Rep*. 2015;5:15972.
42. Chang S-M, Hu W-W. Long non-coding RNA MALAT1 promotes oral squamous cell carcinoma development via microRNA-125b/STAT3 axis. *J Cell Physiol*. 2018;233:3384–96.
43. Liang J, Liang L, Ouyang K, Li Z, Yi X. MALAT1 induces tongue cancer cells' EMT and inhibits apoptosis through Wnt/ β -catenin signaling pathway. *J Oral Pathol Med*. 2017;46:98–105.
44. Su Y, Xiong J, Hu J, Wei X, Zhang X, Rao L. MicroRNA-140-5p targets insulin like growth factor 2 mRNA binding protein 1 (IGF2BP1) to suppress cervical cancer growth and metastasis. *Oncotarget*. 2016;7:68397–411.
45. Wang P, Wu T, Zhou H, Jin Q, He G, Yu H, et al. Long noncoding RNA NEAT1 promotes laryngeal squamous cell cancer through regulating miR-107/CDK6 pathway. *J Exp Clin Cancer Res*. 2016;35:22.
46. Rehmani A, Judkins C, Whelan A, Nguyen M, Schultz C. Comparison of safety and efficacy of unfractionated heparin versus Bivalirudin in patients undergoing percutaneous coronary intervention. *Heart Lung Circ*. 2017;26:1277–81.
47. Haque S-U, Niu L, Kuhnell D, Hendershot J, Biesiada J, Niu W, et al. Differential expression and prognostic value of long non-coding RNA in HPV-negative head and neck squamous cell carcinoma. *Head Neck*. 2018;40:1555–64.
48. Guo Y-Z, Sun H-H, Wang X-T, Wang M-T. Transcriptomic analysis reveals key lncRNAs associated with ribosomal biogenesis and epidermis differentiation in head and neck squamous cell carcinoma. *J Zhejiang Univ Sci B*. 2018;19:674–88.
49. Li J, Sun S, Chen W, Yuan K. Small Nucleolar RNA host gene 12 (SNHG12) promotes proliferation and invasion of laryngeal Cancer cells via sponging miR-129-5p and potentiating WW domain-containing E3 ubiquitin protein ligase 1 (WWP1) expression. *Med Sci Monit*. 2019;25:5552–60.
50. Liu Y, Dong X, Tian C, Liu H. Human telomerase RNA component (hTERC) gene amplification detected by FISH in precancerous lesions and carcinoma of the larynx. *Diagn Pathol*. 2012;7:34.
51. Sur S, Nakanishi H, Steele R, Ray RB. Depletion of PCAT-1 in head and neck cancer cells inhibits tumor growth and induces apoptosis by modulating c-Myc-AKT1-p38 MAPK signalling pathways. *BMC Cancer*. 2019;19:354.
52. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43:D447–52.
53. Tian S. Identification of subtype-specific prognostic signatures using cox models with redundant gene elimination. *Oncol Lett*. 2018;15:8545–55.
54. Tian S, Wang C, Chang HH, Sun J. Identification of prognostic genes and gene sets for early-stage non-small cell lung cancer using bi-level selection methods. *Sci Rep*. 2017;7:46164. <https://doi.org/10.1038/srep46164>.
55. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human protein reference database--2009 update. *Nucleic Acids Res*. 2009;37(SUPPL.1):D767–72.
56. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

