Modular signature of long non-coding RNA association networks as a prognostic biomarker in lung cancer

Albert Li[1], Wen-Hsuan Yu[1], Chia-Lang Hsu[2], Hsuan-Cheng Huang[3*] and Hsueh-Fen Juan[1,4,5*]

Abstract

Background: Increasing amount of long non-coding RNAs (lncRNAs) have been found involving in many biological processes and played salient roles in cancers. However, up until recently, functions of most lncRNAs in lung cancer have not been fully discovered, particularly in the co-regulated lncRNAs. Thus, this study aims to investigate roles of lncRNA modules and uncover a module-based biomarker in lung adenocarcinoma (LUAD).

Results: We used gene expression profiles from The Cancer Genome Atlas (TCGA) to construct the lncRNA association networks, from which the highly-associated lncRNAs are connected as modules. It was found that the expression of some modules is significantly associated with patient's survival, including module N1 (HR = 0.62, 95% CI = 0.46–0.84, p = 0.00189); N2 (HR = 0.68, CI = 0.50–0.93, p = 0.00159); N4 (HR = 0.70, CI = 0.52–0.95, p = 0.0205) and P3 (HR = 0.68, CI = 0.50–0.92, p = 0.0123). The lncRNA signature consisting of these four prognosis-related modules, a 4-modular lncRNA signature, is associated with favourable prognosis in TCGA-LUAD (HR = 0.51, CI = 0.37–0.69, p value = 2.00e−05). Afterwards, to assess the performance of the generic modular signature as a prognostic biomarker, we computed the time-dependent area under the receiver operating characteristics (AUC) of this 4-modular lncRNA signature, which showed AUC equals 68.44% on 336th day. In terms of biological functions, these modules are correlated with several cancer hallmarks and pathways, including Myc targets, E2F targets, cell cycle, inflammation/immunity-related pathways, androgen/oestrogen response, KRAS signalling, DNA repair and epithelial-mesenchymal transition (EMT).

Conclusion: Taken together, we identified four novel LUAD prognosis-related lncRNA modules, and assessed the performance of the 4-modular lncRNA signature being a prognostic biomarker. Functionally speaking, these modules involve in oncogenic hallmarks as well as pathways. The results unveiled the co-regulated lncRNAs in LUAD and may provide a framework for further lncRNA studies in lung cancer.

Keywords: LncRNA association networks, Lung cancer, lncRNA modular prognostic biomarkers

Background

Lung cancer is one of the leading causes of cancer mortality around the globe [1]. Although a few oncogenic alterations can be targeted by certain medications, which showed greater therapeutic efficacy than chemotherapy [2, 3], the prognostic biomarkers assessing lung cancer

*Correspondence: hsuancheng@nycu.edu.tw; yukijuan@ntu.edu.tw
[1] Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei 10617, Taiwan
[3] Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei 11221, Taiwan
Full list of author information is available at the end of the article

Li *et al. BMC Medical Genomics*　　(2021) 14:290

Page 2 of 14

patients are rather limited currently [3]. Therefore, further exploration in biomarkers that assess the prognosis of lung cancer patients is warranted.

LncRNAs, transcripts with length greater than 200 nucleotides [4], regulate gene expression through various mechanisms, including cis- [5–7], trans-[8, 9], and post-transcriptional regulation. Aberrant lncRNA regulations were found in various cancers [10]. These perturbations in transcriptome can be revealed by lncRNA-mRNA co-expression networks, resulting in the discovery of several key lncRNAs and mRNAs that contributed to cancer progression [11, 12]. Subsequent in silico and clinical studies further suggested that multiple lncRNAs can be integrated into singular signatures to better predict cancer patients' survival [13–16], tumour relapse [17, 18] and response to immunotherapy [15]. Given that the underlying biological explanation of combining multiple lncRNAs into one signature may still be unclear, it was speculated that functionally similar lncRNA may co-appear within cells. Therefore, identifying highly associated lncRNAs and their roles in LUAD is crucial.

A lncRNA-mRNA co-expression network is a bi-partite network. The association between each lncRNA can be measured in a lncRNA-mRNA bi-partite network via computing the association index, a quantitative metric measuring the similarity between two lncRNAs according to the overlapping correlated mRNAs. By constructing the lncRNA association networks, similar lncRNAs would be clustered as modules [19]. Functions of the lncRNA modules can then be deduced by the module-correlated mRNAs under the norm of 'guilt by association' [20].

However, our understanding regarding the associated lncRNAs in LUAD has been primitive. The roles of LUAD-specific modules are to be confirmed. Furthermore, whether lncRNA modules signature can be a decent prognostic biomarker needs further investigation. Hence, we postulated that lncRNA modules are composed of co-regulated lncRNAs, and these modules play crucial roles in LUAD. To validate the hypothesis, we gathered gene expression profile from TCGA-LUAD project, constructed the lncRNA association networks, deduced impact of modules on patients' survival, and analysed biological roles of these lncRNA modules. In this study, we provided a module-based approach to uncover a novel biomarker in LUAD.

Results
Overview of the analytical pipeline
We proposed a computational pipeline (Fig. 1) to construct LUAD-specific lncRNA association networks and identify prognosis-related lncRNA modules. Firstly, the correlation between lncRNA and mRNA was quantified
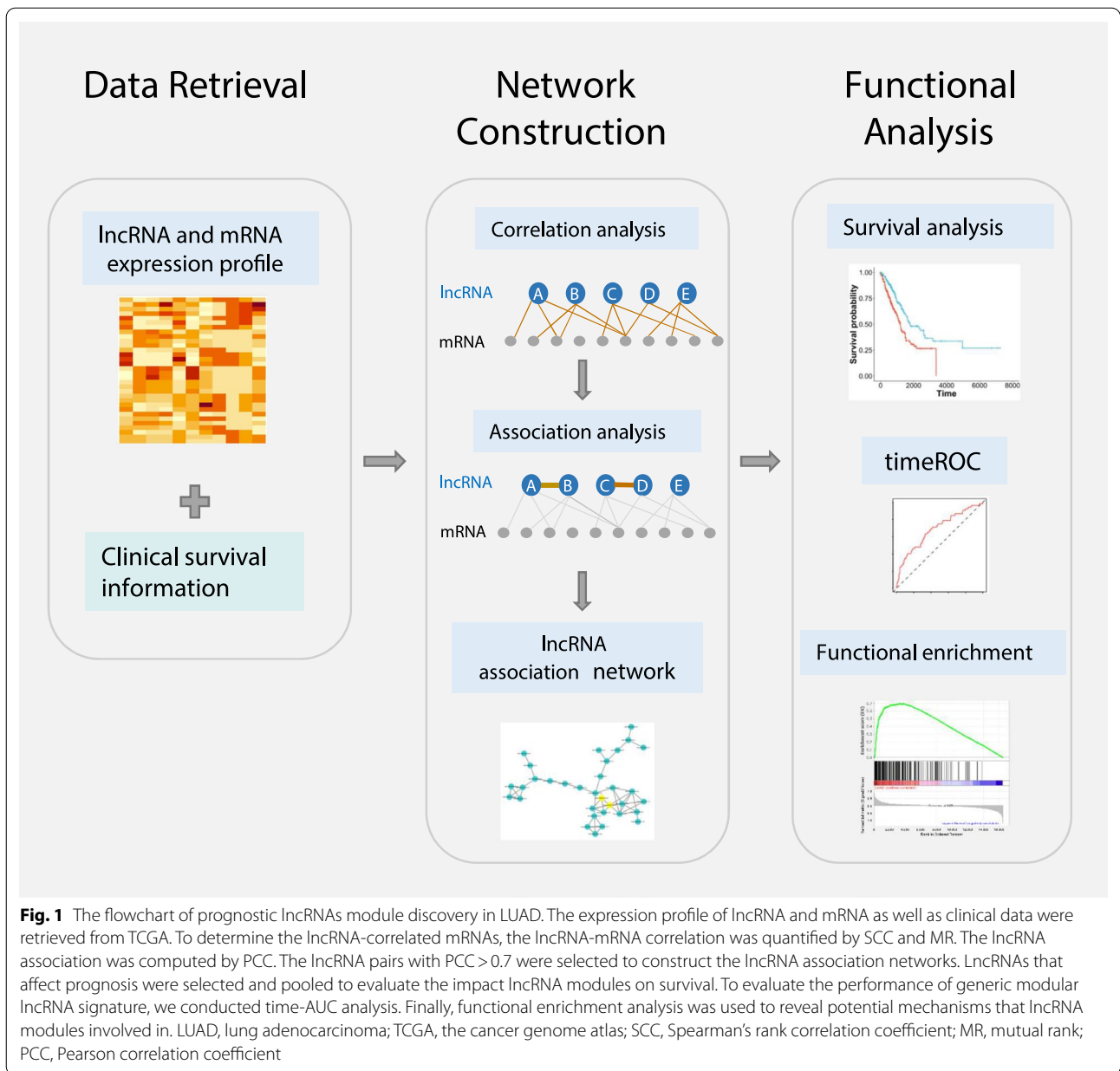
with Spearman correlation coefficient (SCC), which was then transformed to mutual rank (MR) index. The mRNAs within top- and bottom- MR index were selected as positive- and negative- correlated mRNAs respectively. Second, the association between each lncRNA was quantified by the association index. Next, we selected lncRNA pairs with top-scoring association index and constructed lncRNA association networks. Finally, we assessed roles of lncRNA modules in LUAD through survival analysis, time-dependent AUC, and functional enrichment analysis.

Construction of the lncRNA association networks
The gene expression profile of 553 LUAD samples were obtained from TCGA. After pre-processing, 4342 lncRNAs and 16,618 mRNAs remained. We quantified the correlation between lncRNAs and mRNAs with SCC (Fig. 2A). To identify the lncRNA-correlated mRNAs, SCCs were transformed to MR index [21]. The mRNAs with top and bottom 0.1% of MR index ($MR \leq \sqrt{4342 \times 16619} \times 0.001 \cong 8.49$) were selected as the lncRNA- positively or negatively correlated mRNAs. The number of co-expressed mRNA of each lncRNA was shown in Fig. 2B. The mRNAs that are highly correlated with lncRNAs were chosen to calculate the association index between lncRNAs (Fig. 2C). Several association indices are popular and commonly used, such as Jaccard index. We used Pearson Correlation Coefficient (PCC) in this study because PCC can group the associated genes (AUC = 0.903) and can best separate highly co-expressed gene pairs from the others (AUC = 0.750) [20]. We collected the lncRNA-lncRNA pairs having PCCs > 0.7 and constructed the lncRNA association networks (LAN). The LAN constructed from the positively or negatively correlated mRNAs were coined as positive lncRNA association network (PLAN) or negative lncRNA association network (NLAN), respectively (Fig. 3).

Characteristics of the lncRNA association networks
To verify the scale-free property of generic lncRNA association networks, we inspected the relation between the degree (k) of nodes and the degree density (pk) (Fig. 4A). The power law distribution was revealed when plotting k against pk. Further, when taking the logarithm on both axes, the regression line significantly fit the data points, with the slope coefficient = − 1.96, p value = 0.000292 on PLAN; the slope coefficient = − 2.24, p value = 6.51e-05 on NALN. (Fig. 4B). The above results suggested that the lncRNA association networks were non-random, and had the scale-free property. We selected modules with nodes no less than 12 for further analysis (Fig. 4C). Six modules were selected from the NLAN, and three modules were from PLAN (Additional file 1: Table S1). Characteristics

Li *et al. BMC Medical Genomics*   (2021) 14:290

Page 3 of 14

Fig. 1 The flowchart of prognostic lncRNAs module discovery in LUAD. The expression profile of lncRNA and mRNA as well as clinical data were retrieved from TCGA. To determine the lncRNA-correlated mRNAs, the lncRNA-mRNA correlation was quantified by SCC and MR. The lncRNA association was computed by PCC. The lncRNA pairs with PCC > 0.7 were selected to construct the lncRNA association networks. LncRNAs that affect prognosis were selected and pooled to evaluate the impact lncRNA modules on survival. To evaluate the performance of generic modular lncRNA signature, we conducted time-AUC analysis. Finally, functional enrichment analysis was used to reveal potential mechanisms that lncRNA modules involved in. LUAD, lung adenocarcinoma; TCGA, the cancer genome atlas; SCC, Spearman's rank correlation coefficient; MR, mutual rank; PCC, Pearson correlation coefficient

of each module were visualized with box plots, including the expression of lncRNAs (Additional file 1: Fig. S1A), node degree (Additional file 1: Fig. S1B), clustering coefficient (Additional file 1: Fig. S1C), and betweenness centrality (Additional file 1: Fig. S1D).
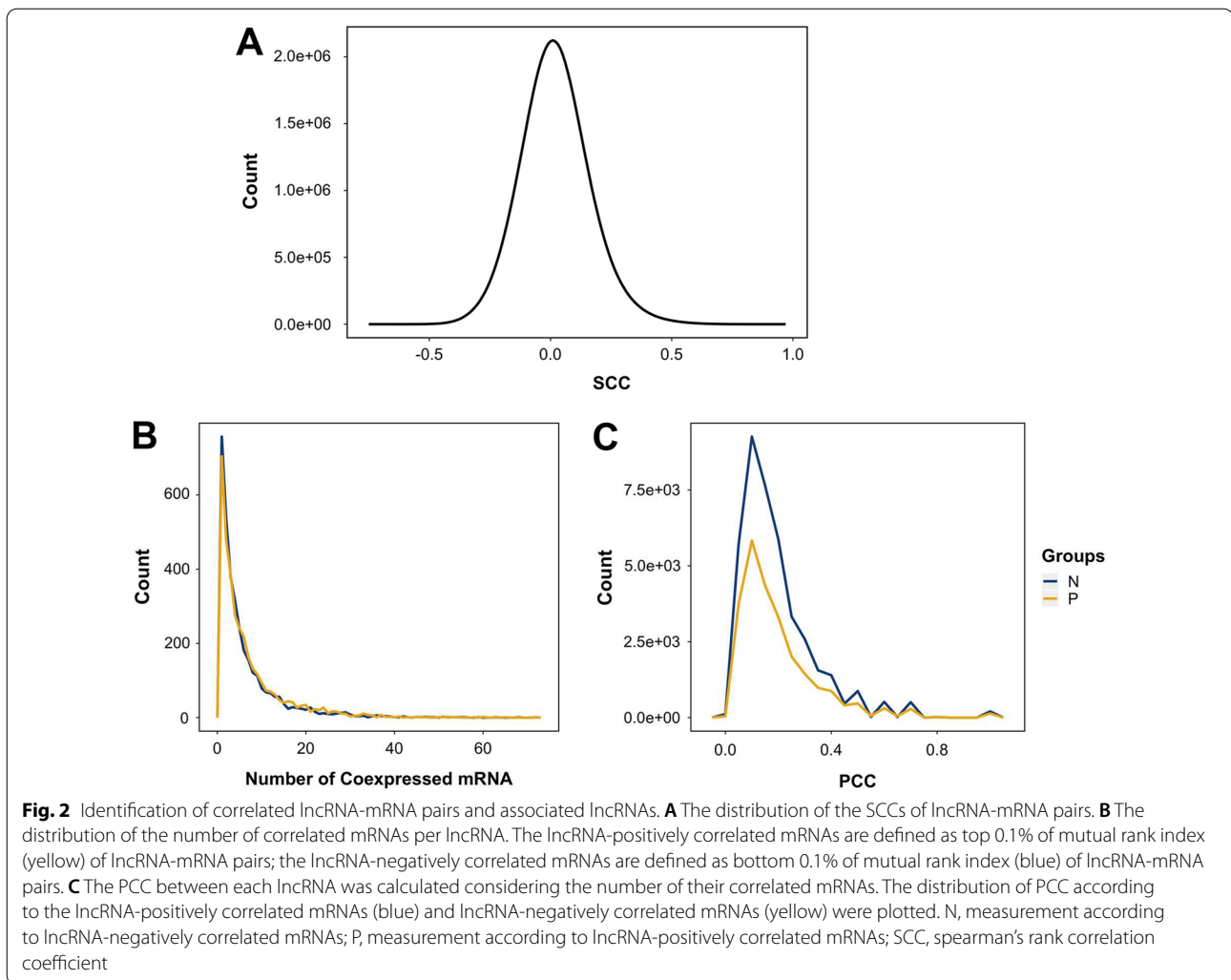
LncRNAs within the modules are highly associated
Since we postulated that modules are composed of highly associated lncRNAs, we compared the association index, PCC, between intra- and inter-module lncRNAs. It was found that intra-module PCCs are significantly higher than that in inter-module PCCs (Fig. 5A, B), implying

lncRNAs within the modules are more alike. The PCC of each lncRNA pair was then visualized with a heat map (Fig. 5C, D). It can be seen that lncRNAs within the same modules are highly associated. The above findings suggested that the association index is can group highly associated lncRNAs together as modules.

LncRNA modules are associated with the prognosis of LUAD patients
To confirm the effects of the modules on patients' prognosis, we firstly screened the lncRNAs in the selected nine modules by univariate survival analysis (log-rank
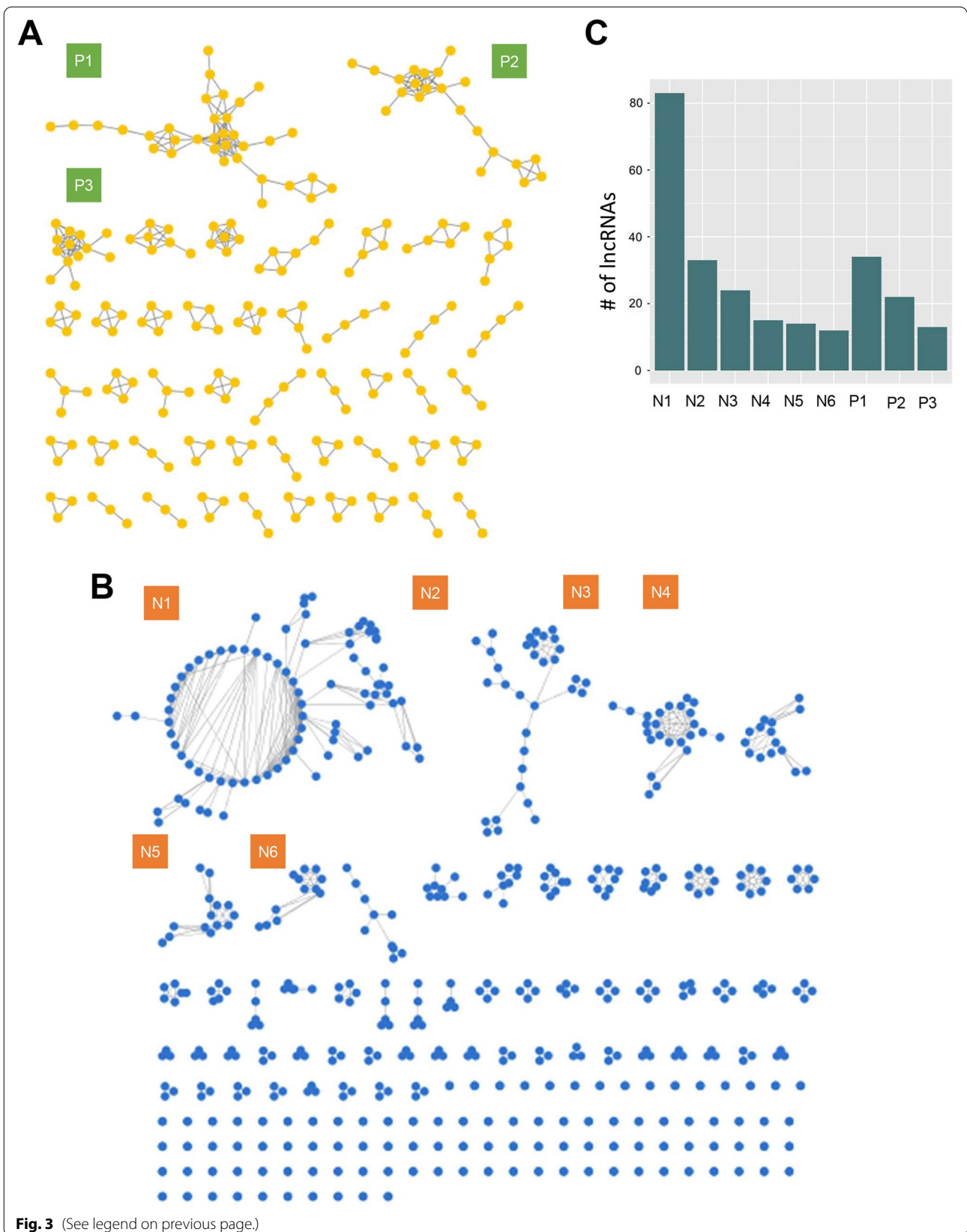
Li *et al. BMC Medical Genomics*   (2021) 14:290

Page 4 of 14

Fig. 2 Identification of correlated lncRNA-mRNA pairs and associated lncRNAs. **A** The distribution of the SCCs of lncRNA-mRNA pairs. **B** The distribution of the number of correlated mRNAs per lncRNA. The lncRNA-positively correlated mRNAs are defined as top 0.1% of mutual rank index (yellow) of lncRNA-mRNA pairs; the lncRNA-negatively correlated mRNAs are defined as bottom 0.1% of mutual rank index (blue) of lncRNA-mRNA pairs. **C** The PCC between each lncRNA was calculated considering the number of their correlated mRNAs. The distribution of PCC according to the lncRNA-positively correlated mRNAs (blue) and lncRNA-negatively correlated mRNAs (yellow) were plotted. N, measurement according to lncRNA-negatively correlated mRNAs; P, measurement according to lncRNA-positively correlated mRNAs; SCC, spearman's rank correlation coefficient

test). Because the results from univariate analysis may be confounded by factors that affect the outcome, we next utilised multivariate Cox proportional hazard model to adjust the cancer stage and the age of patients. Afterwards, we excluded the survival-related lncRNAs that were cancer-stage independent or not differentially expressed between cancer and normal tissue (see Additional file 1: Fig. S2). Finally, nine prognosis-related lncRNAs were selected (Table 1). Specifically, there are five prognostic lncRNAs (AC073655.2, AL031775.2, AL122010.1, AC079210.1, and AL391834.2) in module

N1, two lncRNAs (AL683813.2 and AC005759.1) in N2, one lncRNA (AC008937.3) in N4, and one lncRNA (AC099343.4) in P3. The modular expression was calculated by summing the coefficient-adjusted expression of the prognostic lncRNA(s). These four lncRNA modules are significantly associated with overall survival (OS) as shown in the Kaplan–Meier plots (Fig. 6A–D). We also inspected the impact of lncRNA modules on patients' progression free survival (PFS) (Fig. 6E–G). When adjusting the effects of cancer stage and patients' age on survival, the association between OS and lncRNA module is

(See figure on next page.)

Fig. 3 The lncRNA association networks. **A** The lncRNA-lncRNA pairs with PCC greater than 0.7 were selected to construct the positive lncRNA association network (PLAN). The calculation of PCC is based on the mRNAs that are positively correlated with lncRNAs. Modules P1-P3 (green square) are selected for following analyses. **B** As the above method, we measured the PCC based on the mRNAs that are negatively correlated with lncRNAs. The lncRNA pairs with PCC greater than 0.7 were collected to construct the negative lncRNA association network (NLAN). N1-N6 are modules selected for following analyses. **C** The number of nodes in each selected module
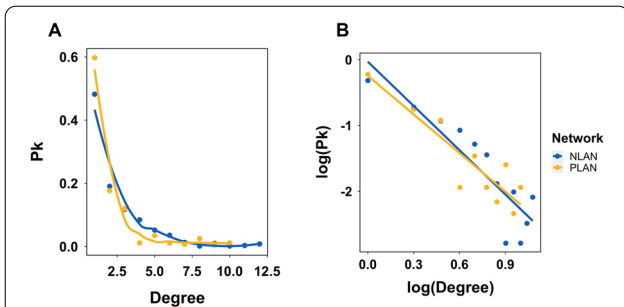
Li *et al. BMC Medical Genomics*   *(2021) 14:290*

Page 5 of 14

Fig. 3 (See legend on previous page.)

Li *et al. BMC Medical Genomics*   (2021) 14:290

Page 6 of 14

Fig. 4 Identification of the scale-free property of the lncRNA association networks. **A** The degree of each node in the networks were plotted against the degree density (pk). The fitted curve showed power law distribution. **B** The value in both axes underwent logarithm transformation. The fitted regression lines were significant, with $p = 6.51e−05$ in NALN and $p = 0.000292$ in PLAN respectively

still significant. (Fig. 7A). Next, since combining multiple lncRNAs into one signature can better predict the outcomes [13, 14, 16], four lncRNA modules (N1, N2, N4, and P3) were pooled into one lncRNA signature. The OS ($p < 0.0001$) and PFS ($p = 0.021$) are significant (Fig. 7B, C). The hazard ratio of the 4-module signature in the multivariable model is 0.51 (95% CI = 0.37, 0.69), and p value = $2.00e−05$ (Fig. 7A).

Assessment of the lncRNA modular signature as a novel prognostic biomarker in LUAD

To assess the lncRNA modules as a prognostic biomarker in LUAD, we interrogated patients' OS and compared the AUC at several time intervals (Fig. 7D). The results were consistent with the former survival analyses. The 4-module lncRNA signature is better than individual
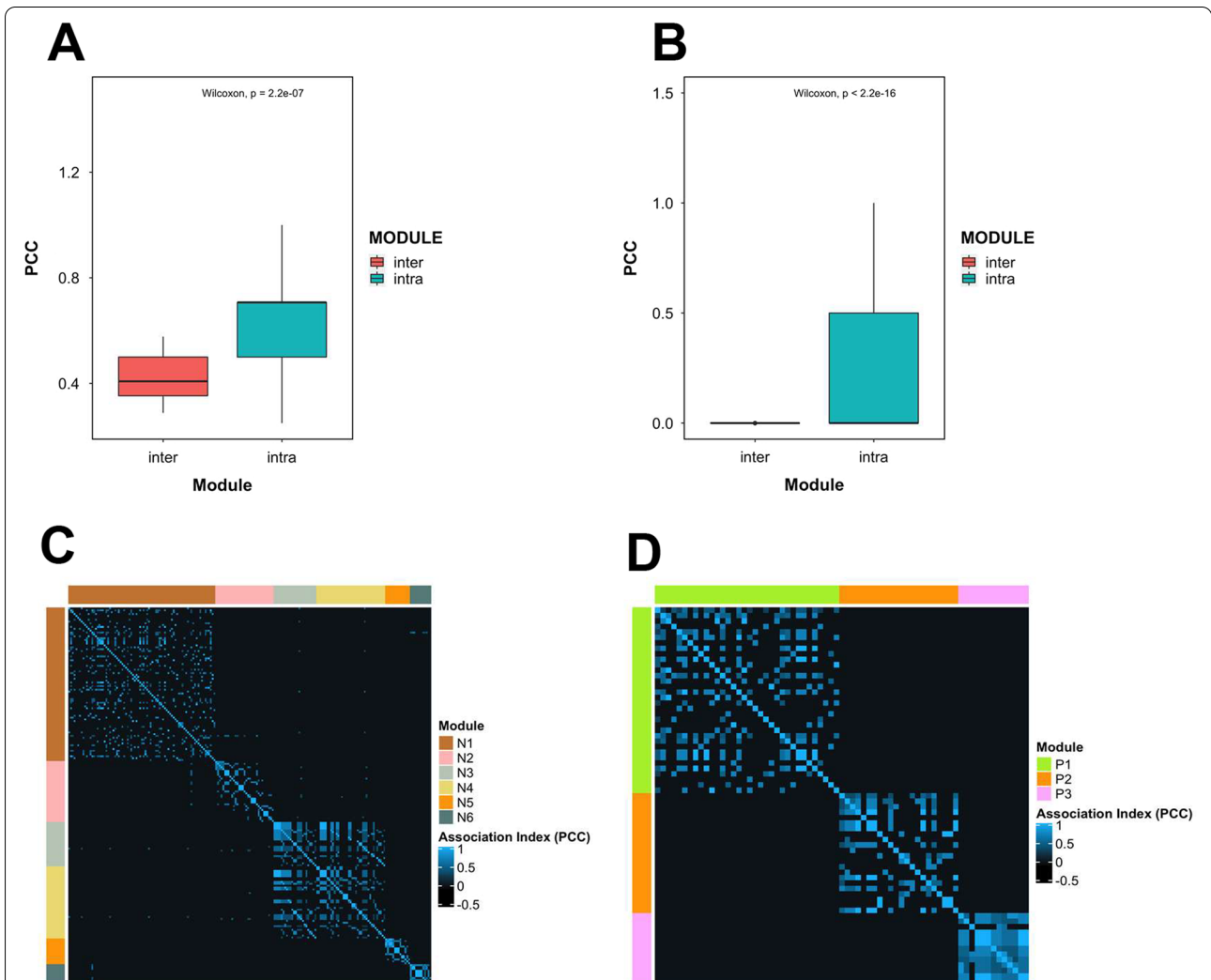
Fig. 5 Comparison of intra-module lncRNAs and inter-module lncRNAs. **A**, **B** The PCC between intra-modular lncRNAs and inter-modular lncRNAs were compared in NLAN and PLAN respectively. Wilcoxon test was used to compare two groups. **C**, **D** The PCCs of each lncRNA were visualized in heatmaps. The clusters in heatmpas were highly correspond to lncRNA modules identified in the association networks

Li *et al. BMC Medical Genomics*   (2021) 14:290

Page 7 of 14

Table 1 Survival-related lncRNAs in each prognostic lncRNA module and the overall survival

ID	Name	Module	Survival
ENSG00000258365	AC073655.2	N1	Favourable
ENSG00000272402	AL031775.2		Favourable
ENSG00000230163	AL122010.1		Favourable
ENSG00000267461	AC079210.1		Favourable
ENSG00000273226	AL391834.2		Favourable
ENSG00000232611	AL683813.2	N2	Favourable
ENSG00000268650	AC005759.1		Favourable
ENSG00000271828	AC008937.3	N4	Favourable
ENSG00000271646	AC099343.4	P3	Favourable

lncRNA modules at all time points. The highest AUC of the 4-module lncRNA signature appears on the 336th day (AUC = (0.68)).

The lncRNA modules are related to cancer hallmarks

Until recently, functions of many lncRNAs in cancer have still not been fully characterized [22]. We used pre-ranked gene set enrichment analysis to deduce the functions of the prognostic lncRNA modules [23]. It was found that Myc targets, E2F targets, and cell cycle were enriched in module N1, N2 and P3. Further, in module N4, inflammation/immunity-related terms, androgen/oestrogen response, and KRAS signalling were revealed. Of note, several pathways of other cancer types were enriched in the module P3, implying the involvement of other cancer types as well (Fig. 8). Some of our enrichment results were consistent with the findings validated in other published literatures, where the expression of the lncRNAs were linked to cancer-related processes such as E2F [24], c-Myc [25, 26], androgen receptor [27, 28] and oestrogen receptor [29]. All in all, the enrichment analysis revealed cancer hallmarks in which lncRNA modules involve. The enriched terms not only reproduce the characters of lncRNAs found previously, but as well reveal several novel mechanisms in which lncRNA modules may participate.

Discussion

This study utilized an association network approach to group similar lncRNAs, identify four prognosis-related lncRNA modules and assess the performance of the modular signature as a novel biomarker in LUAD. Biologically, these lncRNA modules involve in various oncogenic mechanisms. Altogether, this study not only revealed the roles of lncRNA modules but also came up with a new module-based biomarker in LAUD.

The genetic regulation of lncRNA is complex. Thus far, we merely made a correlation and association analyses to model the relation between lncRNAs and mRNAs. It may not be proper to infer the causal relationship between lncRNAs and mRNAs from this study. In fact, lncRNA can regulate mRNAs through cis- and trans- regulation, while mRNAs can also regulate the expression of lncRNAs through other mechanisms [30, 31]. The direction of regulation between lncRNAs and mRNAs still need further investigation in the future.

Numerous studies also focused on the issue regarding the competing endogenous RNA (ceRNA) networks [32–34]. Via constructing a lncRNA–miRNA–mRNA ceRNA network, the molecular mechanisms of certain diseases can be further clarified. A LUAD-specific ceRNA network revealed that seven lncRNAs were associated with the prognosis through interacting with one miRNA (hsa-mir-31) and 16 mRNAs [35]. *Zhang* et al. identified a LUAD-specific ceRNA network through curating the experiment-supported databases. Based on this ceRNA network, they used random walking and restart algorithm to rank lncRNAs that are associated with LUAD. An in vitro validation was subsequently conducted to prove their prediction that MAPKAPK5-AS1, a novel LUAD-related lncRNA, plays crucial roles in tumour growth [36]. Compared with the construction of ceRNA network in LUAD, our approach mainly used the correlation between lncRNAs and mRNAs to infer the association network of lncRNAs, without considering the roles of miRNAs. However, the advantages of our study include using the real-world data to assess the impact of the lncRNA signature on patients' survival, and evaluating the lncRNA signature as a biomarker in LUAD.

Network biology has become popular in biomedical researches. Analysing networks can model the complex interactions of the biomolecules, providing more accurate predictions [19, 21, 37–41]. Furthermore, it was revealed that networks can delineate dissimilar molecules and gather similar nodes as modules [42, 43]. Therefore, in this study, we conjectured that similar lncRNAs would co-appear as modules and are highly associated. We further tested the biological functions of these generic modules, and discovered module-based biomarkers in LUAD. It is not denying that surveying the biomarkers within the lncRNA association networks may miss the prognosis-related lncRNAs that are not appear in the networks. However, the goal of this study was to construct a LUAD-specific lncRNA association network and assess whether these modules play important roles in LUAD, instead of revealing all lncRNAs related to the prognosis in LUAD.

One of the future applications of this study is that the distance of modules in the networks can be quantified by the shortest path length (proximity), which was proved to
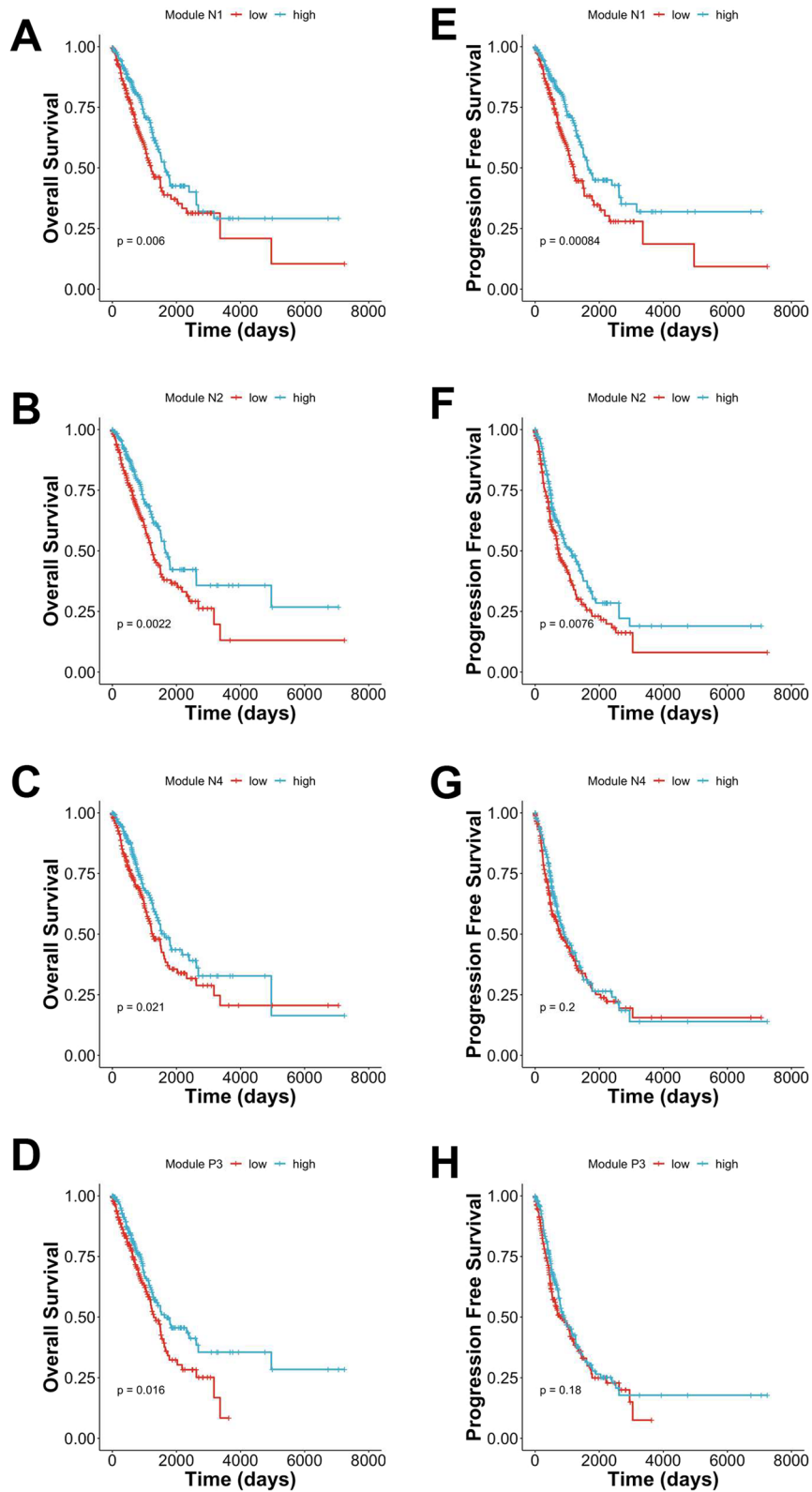
Li *et al. BMC Medical Genomics*   (2021) 14:290

Page 8 of 14

Fig. 6 Survival analysis of prognostic lncRNA modules. The median of modular gene expression was used to assign patients into high or low expression groups. Log-rank test was used to test statistical significance. **A**–**D**. Kaplan–Meier curve of the overall survival (OS) in module N1, N2, N4, and P3. **E**–**G**. Kaplan–Meier curve of the event-free survival (EFS) in module N1, N2, N4, and P3

Li *et al. BMC Medical Genomics*   (2021) 14:290

Page 9 of 14

be a successful method to judge the relatedness between clusters of nodes [38]. This fact can be further applied to the cancer drug repurposing. For instance, a recent study found that the proximity between disease-related proteins and drug targets in the protein interactome is an effective measure to predict the response to the drugs [39]. The accuracy of the proximity method even outweigh the traditional expression-based perturbation methods, such as The Library of Integrated Network-Based Cellular Signatures (LINCS) [39]. In reality, according to the latest LUAD guideline [3], targeting lncRNA modules is not feasible. Therefore, the drug repurposing method that interrogating the proximity between lncRNA module-related proteins and drug targets may be one of the solutions.

We referred our findings to The Cancer LncRNome Atlas [44], a database collecting lncRNA alterations across multiple human cancer types (http://fcgportal.org/TCLA/search.php). It can be seen that the lncRNAs revealed in this study were novel. Copy number variation (CNV) deletion were found in AC073655.2 and AC099343.4, but the alterations in expression level have been unclear. Four lncRNAs (AC008937.3, AL683813.2, AL031775.2, AL122010.1, and AC005759.1) were detectable in LUAD. The other two lncRNAs, AL391834.2, and AC079210.1, were not detectable in LUAD (Additional file 1: Table S2). It is believed that gene expression can be affected by CNV, where decreased gene expression may be attributed to deletions [45]. In our study, we found that decreased expression of AC073655.2 and AC099343.4 are associated with poorer prognosis in LUAD patients. Therefore, it may not be inappropriate to postulate that the underlying mechanisms might be associated with the deletions. However, further investigations will be needed to validate this hypothesis.

Conclusions

In conclusion, by mining the expression profile from TCGA, we constructed lncRNA-association networks and revealed four crucial LUAD-specific lncRNA modules. These modules involve in several cancer-related biological mechanisms and affect patients' survival. The 4-module lncRNA signature can be used as a prognostic biomarker for LUAD. As far as we know, this is the first study using lncRNA association network to discover promising lncRNA modular biomarkers in LUAD. This method offered a novel insight into the co-regulated lncRNAs in

lung cancer and may become a framework for future studies, such as the lncRNA module-based target therapy.

Methods

Pre-processing of lncRNA and mRNA expression profiles

We retrieved the transcriptome data of lung adenocarcinoma (LUAD) from Genomic data commons (GDC) data portal. The lncRNA and mRNA expression profile were collected from the HTSeq-FPKM-UQ workflow of TCGA-LUAD project (n = 553) in February 2018. The pre-processing procedure of the gene expression profile was adjusted from our previous work [21]. In brief, genes with more than 20% of values to be NAs were removed. Next, the remaining NAs were filled with 0. We then adjusted the expression of genes with log2 transformation. To link the Ensembl gene ID with gene type, we used Ensembl 92 (GRCh38.p12) as the reference genome. Finally, The Ensembl gene ID was converted to gene symbol by referring to BioMart database [46]. The pre-processed gene expression file contained 4342 lncRNAs and 16618 mRNAs.

We also used identical pre-processing steps to gather the expression profile of adjacent normal tissue for subsequent analyses. These samples were used as control group and pooled from the adjacent normal tissue in both LUAD and lung squamous cell carcinoma (LUSC) project from TCGA, consisting of 108 samples in total.

Identification of LncRNA-correlated mRNAs

To model the relation between lncRNAs and mRNAs, we used Spearman's correlation coefficient (SCC) (1) to quantify the correlation of lncRNA and mRNA.

$$SCC = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{1}$$

where d_i = the difference between the two ranks of each investigated gene, and n = sample size.

The SCCs were then transformed to mutual rank (MR) index. The MR is the geometric average of the SCC rank [21]. That is, for the gene pair lncRNA X and mRNA Y, the mutual rank is defined as follow:

$$MR_{XY} = \sqrt{Rank_{X \to Y} \times Rank_{Y \to X}} \tag{2}$$

In order to select negatively and positively lncRNA-correlated mRNAs, we ranked the SCCs in ascending and

---

(See figure on next page.)

Fig. 7 Evaluation of multi-module lncRNA signature. **A** A forest plot showing the HR and p-value under multivariable Cox PH model to assess overall survival of LUAD patients with high *v.s.* low expression of the four lncRNA modules (N1, N2, N4, P3) and the 4-module signature. Age and TNM stage were adjusted. **B** Kaplan–Meier curve of the OS in the 4-module Signature. Log-rank test was used. **C** Kaplan–Meier curve of the EFS in the 4-module Signature. Log-rank test was used. **D** The comparison of AUC of module N1, N2, N4, P3 and the 4-module signature in five different time points
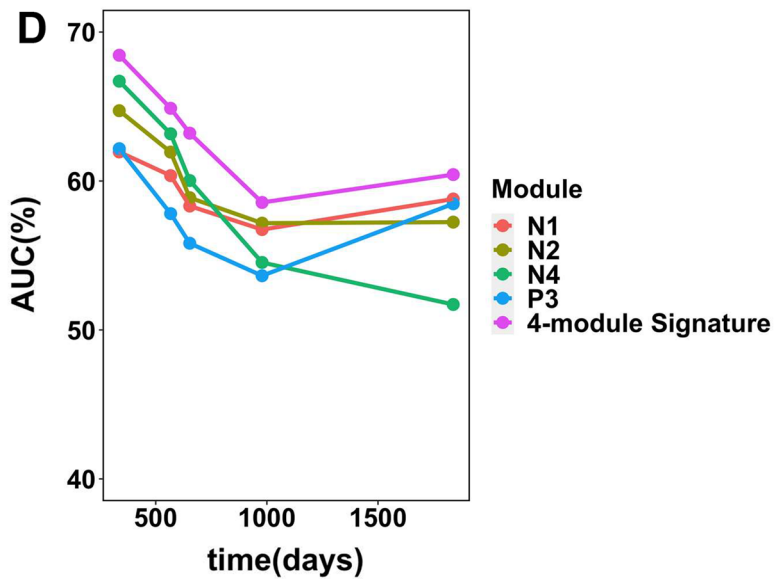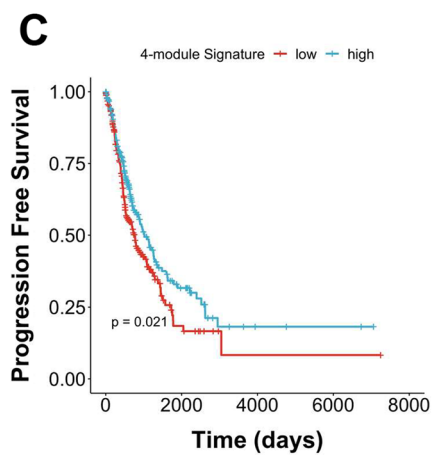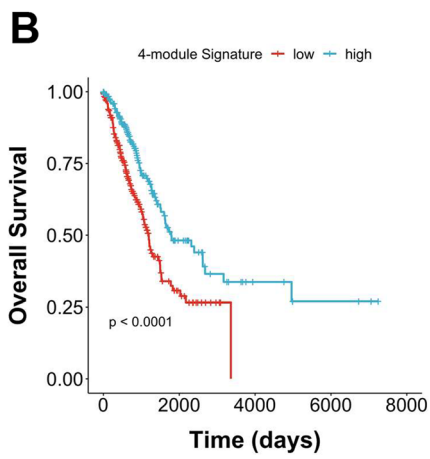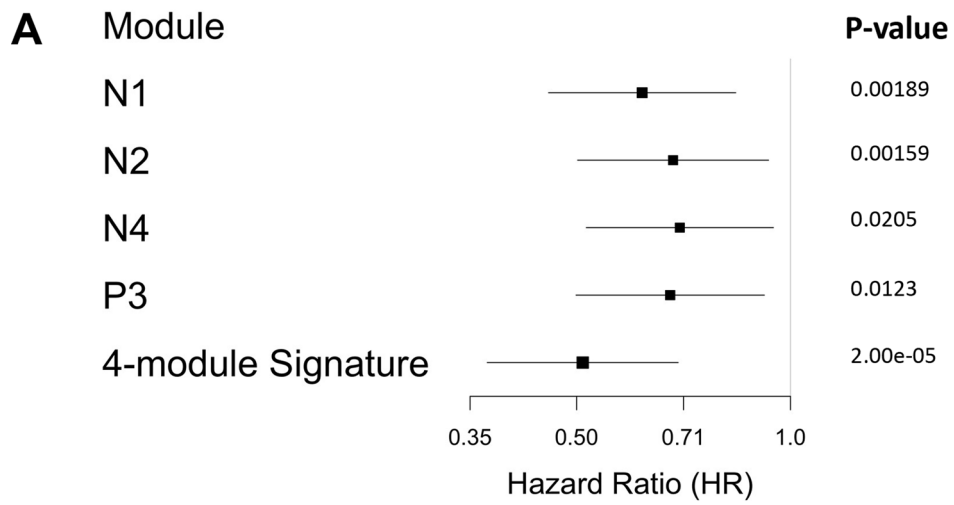
Fig. 7 (See legend on previous page.)

Li *et al. BMC Medical Genomics*   (2021) 14:290
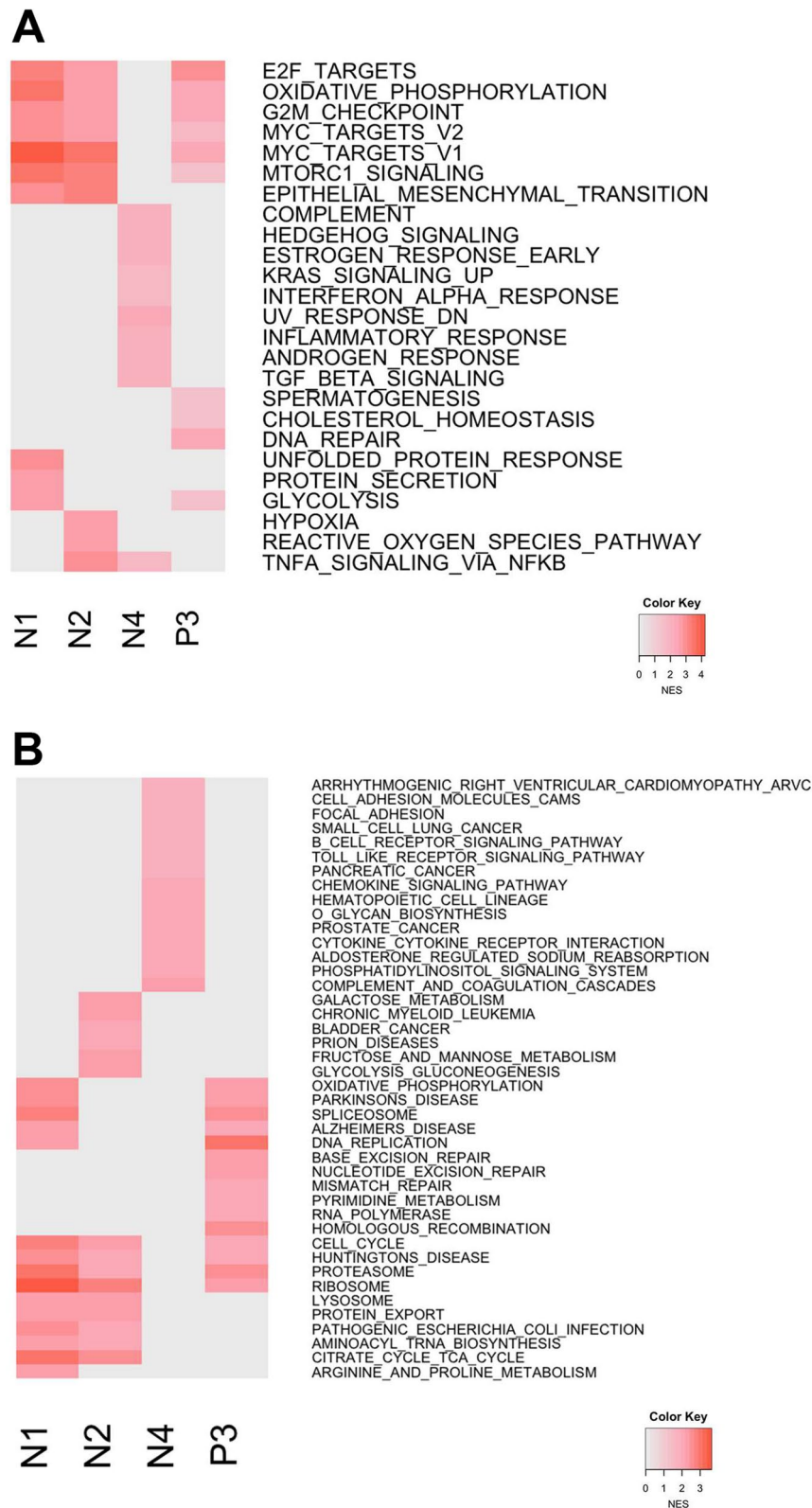
Page 11 of 14

Fig. 8 Comparing functions of prognostic lncRNA modules. **A** GSEA was used to conduct functional enrichment analysis where top 10 biological processes, based on normalized enrichment score (NES), in each module were shown (FDR q-value < 0.001). **B** Biological functions regarding KEGG pathways were investigated. Terms with top 15 NES were shown in each module (FDR q-value < 0.001)

Li *et al. BMC Medical Genomics*   (2021) 14:290

Page 12 of 14

descending order to represent negatively and positively correlated mRNAs respectively. We then defined lncRNA-mRNA pairs with top one percent of MR index as highly correlated lncRNA-mRNA pairs.

Construction of lncRNA association networks

We used the lncRNA mRNA correlation bi-partite networks to deduce the association between lncRNAs. Pearson correlation coefficient (PCC) (3), an association index, was computed to define the association between each lncRNA.

$$PCC_{XY} = \frac{|N(X) \cap n(Y)|n_y - |N(X)||N(Y)|}{\sqrt{|N(X)||N(Y)|\left(n_y - |N(X)|\right)\left(n_y - |N(Y)|\right)}}$$

(3)

where N(X) is the number of mRNAs correlated with lncRNA X, N(Y) is the number of genes correlated with lncRNA Y, n_y is the number of mRNAs correlated with both lncRNA X and lncRNA Y. We defined two lncR-NAs being associated if and only if they shared PCC > 0.7. We then collected lncRNA pairs with this criterion into a lncRNA-lncRNA edge list and visualized the lncRNA association network with Cytoscape [47]. The lncRNA association network constructed from negatively correlated mRNAs were coined as negative lncRNA association network (NALN). On the other hand, those considered positively correlated mRNAs were termed positive lncRNA association network (PLAN).

Survival analysis and biomarker assessment

To select lncRNA modules that determine prognosis of LUAD patients, we designed a three-step selection procedure, implemented with the R package: *survival* (https://CRAN.R-project.org/package=survival) and *survminer* (http://www.sthda.com/english/rpkgs/survminer/). Firstly, the survival time, age at diagnosis, TNM stage, event type (cancer progression or death), and censoring information of TCGA LUAD were collected from clinical data in GDC data portal. We divided patients into two subgroups based on the median of lncRNA expression. The univariate survival analysis was conducted by log-rank test. Secondly, to adjust known confounding factors, including patients' age and cancer stage, the genes that passed log-rank test were subsequently examined by multivariable Cox-proportional hazard model. Finally, the remaining lncRNAs without stage-dependency or differential expression were discarded. The stage-dependency analysis was conducted by comparing lncRNA expression across different TNM stage (from stage I to stage IV). Kruskal–Wallis test was used for hypotheses testing.

For the differential expression analysis, we compared the gene expression between cancer and adjacent normal tissue with Wilcoxon test.

For the biomarker assessment, time-dependent area under the receiver operating characteristic (time-AUC) was calculated by using the R package 'time ROC' [48].

Functional gene set enrichment analysis

To investigate functions of prognostic lncRNA modules, we interrogated lncRNA-correlated mRNAs to conduct pre-ranked GSEA [23]. The mRNAs were ranked based on the correlation (MR index) with the lncRNAs within each module. In PLAN, since the positively correlated mRNAs were more important, the SCCs were arranged in the ascending order, from small to large, which led to higher MR index in the positively correlated mRNAs. By contrast, SCCs were arranged in descending order, which resulted in higher MR index in negatively correlated mRNAs. Gene sets database, Hallmark and KEGG pathway, from MSigDB v7.1 [49, 50] were selected. The chip platform was from Human_ENSEMBL_Gene_MSigDB.v7.1.chip. The top 10 and 15 were selected from Hallmark and KEGG pathway respectively. All selected terms were significant, with FDR-q value < 0.001.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-021-01137-0.

Additional file 1. Detailed components in each lncRNA module.

Authors' contributions

HFJ and HCH conceived, designed, and supervised the study. AL, WHY and CLH developed the method. AL performed the analyses. AL, HFJ and HCH wrote the manuscript. All authors read and approved the final manuscript.

Li *et al. BMC Medical Genomics*   (2021) 14:290

Page 13 of 14

Availability of data and materials
The RNA-seq transcriptome profiling of lung adenocarcinoma from TCGA is freely available via GDC data portal (https://portal.gdc.cancer.gov/).

Declarations

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
[1]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei 10617, Taiwan. [2]Department of Medical Research, National Taiwan University Hospital, Taipei 10002, Taiwan. [3]Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei 11221, Taiwan. [4]Department of Life Science, National Taiwan University, Taipei 10617, Taiwan. [5]Center for Computational and Systems Biology, National Taiwan University, Taipei 10617, Taiwan.

References
1. Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, Abdel-Rahman O, Abdelalim A, Abdoli A, Abdollahpour I, Abdulle ASM, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: a systematic analysis for the global burden of disease study. JAMA Oncol. 2019;5(12):1749–68.
2. Reck M, Rabe KF. Precision diagnosis and treatment for advanced non-small-cell lung cancer. N Engl J Med. 2017;377(9):849–61.
3. Non-small Cell Lung Cancer (Version 6, 2020) https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf.
4. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. RNA Biol. 2013;10(6):925–33.
5. Martens JA, Laprade L, Winston F. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature. 2004;429(6991):571–4.
6. Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature. 2007;445(7128):666–70.
7. Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T, Ohta K. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature. 2008;456(7218):130–4.
8. Pandey GK, Mitra S, Subhash S, Hertwig F, Kanduri M, Mishra K, Fransson S, Ganeshram A, Mondal T, Bandaru S, et al. The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. Cancer Cell. 2014;26(5):722–37.
9. Marín-Béjar O, Marchese FP, Athie A, Sánchez Y, González J, Segura V, Huang L, Moreno I, Navarro A, Monzó M, et al. Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. Genome Biol. 2013;14(9):R104.
10. Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, Huarte M, Zhan M, Becker KG, Gorospe M. LincRNA-p21 suppresses target mRNA translation. Mol Cell. 2012;47(4):648–55.
11. Alaei S, Sadeghi B, Najafi A, Masoudi-Nejad A. LncRNA and mRNA integration network reconstruction reveals novel key regulators in esophageal squamous-cell carcinoma. Genomics. 2019;111(1):76–89.
12. Jiang L, Hong L, Yang W, Zhao Y, Tan A, Li Y. Co-expression network analysis of the lncRNAs and mRNAs associated with cervical cancer progression. Arch Med Sci. 2019;15(3):754–64.
13. Zhang H, Zhu M, Du Y, Zhang H, Zhang Q, Liu Q, Huang Z, Zhang L, Li H, Xu L, et al. A panel of 12-lncRNA signature predicts survival of pancreatic adenocarcinoma. J Cancer. 2019;10(6):1550–9.
14. Li J, Chen Z, Tian L, Zhou C, He MY, Gao Y, Wang S, Zhou F, Shi S, Feng X, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. Gut. 2014;63(11):1700–10.
15. Zhou M, Zhang Z, Bao S, Hou P, Yan C, Su J, Sun J. Computational recognition of lncRNA signature of tumor-infiltrating B lymphocytes with potential implications in prognosis and immunotherapy of bladder cancer. Brief Bioinform. 2020. https://doi.org/10.1093/bib/bbaa047.
16. Ma B, Li Y, Ren Y. Identification of a 6-lncRNA prognostic signature based on microarray re-annotation in gastric cancer. Cancer Med. 2020;9(1):335–49.
17. Tang J, Ren J, Cui Q, Zhang D, Kong D, Liao X, Lu M, Gong Y, Wu G. A prognostic 10-lncRNA expression signature for predicting the risk of tumour recurrence in breast cancer patients. J Cell Mol Med. 2019;23(10):6775–84.
18. Li J, Wang W, Xia P, Wan L, Zhang L, Yu L, Wang L, Chen X, Xiao Y, Xu C. Identification of a five-lncRNA signature for predicting the risk of tumor recurrence in patients with breast cancer. Int J Cancer. 2018;143(9):2150–60.
19. Fuxman Bass JI, Diallo A, Nelson J, Soto JM, Myers CL, Walhout AJM. Using networks to measure similarity between genes: association index selection. Nat Methods. 2013;10(12):1169–76.
20. Oliver S. Guilt-by-association goes global. Nature. 2000;403(6770):601–3.
21. Wu CH, Hsu CL, Lu PC, Lin WC, Juan HF, Huang HC. Identification of lncRNA functions in lung cancer based on associated protein–protein interaction modules. Sci Rep. 2016;6:35939.
22. Carlevaro-Fita J, Lanzós A, Feuerbach L, Hong C, Mas-Ponte D, Pedersen JS, Abascal F, Amin SB, Bader GD, Barenboim J, et al. Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. Commun Biol. 2020;3(1):56.
23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–50.
24. Feldstein O, Nizri T, Doniger T, Jacob J, Rechavi G, Ginsberg D. The long non-coding RNA ERIC is regulated by E2F and modulates the cellular response to DNA damage. Mol Cancer. 2013;12(1):131.
25. Hart JR, Roberts TC, Weinberg MS, Morris KV, Vogt PK. MYC regulates the non-coding transcriptome. Oncotarget. 2014;5(24):12543–54.
26. Kim T, Jeon YJ, Cui R, Lee JH, Peng Y, Kim SH, Tili E, Alder H, Croce CM. Role of MYC-regulated long noncoding RNAs in cell cycle regulation and tumorigenesis. J Natl Cancer Inst. 2015. https://doi.org/10.1093/jnci/dju505.
27. Wang LG, Johnson EM, Kinoshita Y, Babb JS, Buckley MT, Liebes LF, Melamed J, Liu XM, Kurek R, Ossowski L, et al. Androgen receptor overexpression in prostate cancer linked to Pur alpha loss from a novel repressor complex. Cancer Res. 2008;68(8):2678–88.
28. Takayama K, Horie-Inoue K, Katayama S, Suzuki T, Tsutsumi S, Ikeda K, Urano T, Fujimura T, Takagi K, Takahashi S, et al. Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer. EMBO J. 2013;32(12):1665–80.
29. Chakravarty D, Sboner A, Nair SS, Giannopoulou E, Li R, Hennig S, Mosquera JM, Pauwels J, Park K, Kossai M, et al. The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. Nat Commun. 2014;5:5383.
30. Sánchez Y, Segura V, Marín-Béjar O, Athie A, Marchese FP, González J, Bujanda L, Guo S, Matheu A, Huarte M. Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumour suppressor signature. Nat Commun. 2014;5:5812.

Li *et al. BMC Medical Genomics*   (2021) 14:290

Page 14 of 14

31. Huarte M. The emerging role of lncRNAs in cancer. Nat Med. 2015;21(11):1253–61.

32. Cong Z, Diao Y, Xu Y, Li X, Jiang Z, Shao C, Ji S, Shen Y, De W, Qiang Y. Long non-coding RNA linc00665 promotes lung adenocarinoma progression and functions as ceRNA to regulate AKR1B10-ERK signaling by sponging miR-98. Cell Death Dis. 2019;10(2):84.

33. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi Pier P. A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? Cell. 2011;146(3):353–8.

34. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. Nature. 2014;505(7483):344–52.

35. Wu X, Sui Z, Zhang H, Wang Y, Yu Z. Integrated Analysis of lncRNA-Mediated ceRNA Network in Lung Adenocarcinoma. Front Oncol. 2020;10:554759.

36. Zhang H, Wang Y, Lu J. Identification of lung-adenocarcinoma-related long non-coding RNAs by random walking on a competing endogenous RNA network. Ann Transl Med. 2019;7(14):339.

37. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004;5(2):101–13.

38. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási A-L. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science (New York, NY). 2015;347(6224):1257601.

39. Guney E, Menche J, Vidal M, Barábasi AL. Network-based in silico drug efficacy screening. Nat Commun. 2016;7:10331.

40. Cheng F, Kovács IA, Barabási AL. Network-based prediction of drug combinations. Nat Commun. 2019;10(1):1197.

41. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12(1):56–68.

42. van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform. 2018;19(4):575–92.

43. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:559.

44. Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, Zhang Y, Yang L, Shan W, He Q, et al. Comprehensive genomic characterization of long non-coding RNAs across human cancers. Cancer Cell. 2015;28(4):529–40.

45. Zhou C, Zhang W, Chen W, Yin Y, Atyah M, Liu S, Guo L, Shi Y, Ye Q, Dong Q, et al. Integrated analysis of copy number variations and gene expression profiling in hepatocellular carcinoma. Sci Rep. 2017;7(1):10570.

46. Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, Kasprzyk A. BioMart–biological queries made easy. BMC Genom. 2009;10:22–22.

47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.

48. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med. 2013;32(30):5381–97.

49. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739–40.

50. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417–25.

Publisher's Note