Transcriptome Based System Biology Exploration Reveals Homogeneous Tumorigenicity of Alimentary Tract Malignancy

Yu-Chen Lu[1†], Jing-Qi Shi[1†], Zi-Xin Zhang[1†], Jia-Yi Zhou[1,2†], Hai-Kun Zhou[1], Yuan-Cai Feng[1], Zhen-Hua Lu[1], Shu-Ya Yang[1], Xi-Yang Zhang[1], Yang Liu[1], Zi-Chao Li[1], Yuan-Jie Sun[1], Lian-He Zheng[3*], Dong-Bo Jiang[1*] and Kun Yang[1*]

[1] Department of Immunology, School of Basic Medicine, The Fourth Military Medical University, Xi'an, China, [2] Aviation Psychology Research Office, Air Force Medical Center, Beijing, China, [3] Department of Orthopedics, The Tangdu Hospital, The Fourth Military Medical University, Xi'an, China

Malignancies of alimentary tract include esophageal carcinoma (ESCA), stomach adenocarcinoma (STAD), colon adenocarcinoma (COAD), and rectum adenocarcinoma (READ). Despite of their similarities in cancer development and progression, there are numerous researches concentrating on single tumor but relatively little on their common mechanisms. Our study explored the transcriptomic data of digestive tract cancers from The Cancer Genome Atlas database, yielding their common differentially expressed genes including 1,700 mRNAs, 29 miRNAs, and 362 long non-coding RNAs (lncRNAs). There were 12 mRNAs, 5 miRNAs, and 16 lncRNAs in the core competitive endogenous RNAs network by RNA-RNA interactions, highlighting the prognostic nodes of SERPINE1, hsa-mir-145, and SNHG1. In addition, the weighted gene co-expression network analysis (WGCNA) illustrated 20 gene modules associated with clinical traits. By taking intersections of modules related to the same trait, we got 67 common genes shared by ESCA and READ and screened 5 hub genes, including ADCY6, CXCL3, NPBWR1, TAS2R38, and PTGDR2. In conclusion, the present study found that SERPINE1/has-mir-145/SNHG1 axis acted as promising targets and the hub genes reasoned the similarity between ESCA and READ, which revealed the homogeneous tumorigenicity of digestive tract cancers at the transcriptome level and led to further comprehension and therapeutics for digestive tract cancers.

Keywords: alimentary tract malignancy, homogeneous tumorigenicity, transcriptome, competing endogenous RNA, weighted gene co-expression network analysis

INTRODUCTION

The incidence and mortality of alimentary tract malignancies have been high in the world, seriously endangering public health and human life (1). For instance, the principle malignant conditions of the digestive tract cause great damage, namely cancers of stomach (approximately 1.0 million new cases in 2018), esophagus (570,000 cases) and colorectum (1.8 million cases) (2). The alimentary

tract, from oropharynx to anal canal, are closely related in mainly organ functions and development. No wonder that they should share similar mechanisms in the progression of cancer. However, the mechanisms of digestive tract tumorigenicity in common remain to be explored.

An important reasons why digestive tract cancers contribute to numerous deaths is the lack of suitable and effective diagnosis (2). Besides, for the treatment, the novel strategy such as immunotherapy depends on predictive biomarkers and therapeutic targets pursuing (3). Recently, some molecules closely related to survival rates are considered as potential diagnostic and prognostic tools in digestive tract cancers (4). However, they are not suitable for diagnostic and prognostic biomarkers because of their low specificity and veracity. Thus, it is urged to investigate the molecular mechanisms, which are important for early diagnosis and treatment of digestive tract cancers.

With the development of high-throughput sequencing and omics profiling in life sciences, there are increasing reliable methods and constantly updated databases (5, 6). The Cancer Genome Atlas (TCGA), one of community resource databases, leads frontiers of information to identify novel biomarkers, which has been widely used in cancer research (7, 8). Despite the relative maturity of genomic analyses, researchers believe that progress will primarily come from transcriptome-based subtyping efforts (9). Given that integrated analyses usually emphasis cancer heterogenicity, this study tried another view of homogeneity of cancer development. By using bioinformatics analysis including differentially expressed genes calling, the competitive endogenous RNA (ceRNA) network constructing and the weighted gene co-expression network analysis (WGCNA) exploring, we identified novel biomarkers associated with clinical traits and common regulatory mechanisms of digestive tract cancers.

METHOD AND MATERIALS

Download and Pretreat the Data
The RNA-seq and miRNA-seq data were downloaded from the TCGA database using the GDC Data Portal (https://gdc-portal. nci.nih.gov/). The mRNA expression data of ESCA included a total of 171 tumor samples consisting of 160 tumor samples and 11 normal samples. The mRNA expression data of STAD included a total of 407 samples consisting of 375 tumor samples and 32 normal samples. The mRNA expression data of COAD included a total of 177 tumor samples consisting of 167 tumor samples and 10 normal samples. The mRNA expression data of READ included a total of 521 samples consisting of 480 tumor samples and 41 normal samples. The miRNA expression data of ESCA included a total of 198 samples consisting of 185 tumor samples and 13 normal esophageal samples. The miRNA expression data of STAD included a total of 491 samples consisting of 446 tumor samples and 45 normal samples. The miRNA expression data of COAD included a total of 458 tumor samples consisting of 450 tumor samples and 8 normal samples. The miRNA expression data of READ included a total of 84 samples consisting of 83 tumor samples and 1 normal sample.

The sequencing data were all publicly available and no ethical issues were involved. The Gencode database (https://www. gencodegenes.org/human/) was used to distinguish and annotate the mRNA and lncRNA from the RNA-seq. LncRNA and mRNA were extracted by Perl to make lncRNA and mRNA data matrixes. The edgeR package in Bioconductor was used to screen the differentially expressed miRNAs, differentially expressed lncRNAs and differentially expressed mRNAs (10). The miRNAs, lncRNA and mRNA were deemed to be different if |log2FoldChange| >0.69 and respectively, both with p-value <0.05. The intersection of four tumors was taken and Venn diagram was drawn to make the common differentially expressed mRNAs, differentially expressed lncRNAs and differentially expressed miRNAs among four tumors.

Gene Set Enrichment Analysis (GSEA) of Different Cancers
For the differentially expressed mRNAs of four digestive tract cancers, Gene set enrichment analysis (GSEA) was conducted to investigate the hallmarks of the four digestive tract cancers by online tool WebGestalt (http://www.webgestalt.org/). The hallmarks gene sets were downloaded from Molecular Signature database (https://www.gsea-msigdb.org/gsea/msigdb). And p-value <0.05 and FDR <0.05 were set as the cutoff criterion to screen the prominent hallmarks.

Functional Annotation
We performed GO and KEGG pathways enrichment analysis to detect the potential biological functions and pathways of the common differentially expressed mRNAs using DAVID (https://david.ncifcrf.gov/) online tool (11, 12). And p-value <0.0001 was set as the cutoff criterion. The cytoscape BiNGo plug-in was also used to analysis and draw the results of enrichment (13).

The Construction of Competing Endogenous RNA (ceRNA) Network
We used miRWalk 3.0 database (http://mirwalk.umm.uni-heidelberg.de/), including Targetscan database, miRDB database and miRTarBase database, aiming to predict target mRNA of different expressed miRNA and took the intersection between the predicted mRNA and common differentially expressed mRNA. We got the adjusted relationships between the miRNA and mRNA. We used the miRcode database (http://www.mircode.org/) to obtain the relationships of lncRNA-miRNA and took the intersection between the predicted lncRNA and common differentially expressed lncRNA. Thus, the visual ceRNA network was constructed by cytoscape software (14). According to ceRNA hypothesis, miRNA expression was negatively correlated with lncRNA or mRNA (15). Thus, for identification of potential ceRNA regulatory axes, the positively correlated lncRNA-miRNA pairs, and miRNA-mRNA pairs in the ceRNA network were discarded. The R package ggalluvial was used to demonstrate the ceRNA axes.

To discover the prognostic factors, we performed the overall survival of genes in the ceRNA network. After downloading the

clinical data from TCGA, the R packages survival was used for survival analysis. To avoid the effect of different cancers, we put the tumor samples which expressions were first quarter of each cancer together, treating as the high-expression group. Thus, the hsa-mir-145 high-expression group includes 106 COAD samples, 36 ESCA samples, 38 READ samples, and 100 STAD samples. Besides, the SERPINE1 and SNHG1 high-expression group includes 110 COAD samples, 36 ESCA samples, 39 READ samples, and 94 STAD samples. Just like the high-expression group, we set the low-expression group with the same numbers of samples. The Log-rank test was used to overall survival analysis and we selected the genes with P value <0.05.

For the lncRNA of highest connectivity, the GEPIA2 database (http://gepia2.cancer-pku.cn/#index) was used to show co-expression relationships between lncRNA and mRNA. Besides, the ImmLnc database (http://bio-bigdata.hrbmu.edu.cn/ImmLnc/) was used to show the Immune-related pathways.

The Weighted Gene Co-Expression Network Analysis (WGCNA)

The "WGCNA" package in R software was used to construct gene co-expression network and identify the co-expression genes modules (16). For four tumors, all genes were divided into different modules (colors) through cluster analysis and dynamic tree cut algorithm. We combined the gene matrixes with clinical traits to analyze the correlation between the different genes modules and clinical traits. We set the | correlation score| >0.3 and P <0.05 as threshold values. For the different gene modules of different cancers, which were related to same clinical trait, we took the intersection to get the common genes. For the common genes, STRING database (https://string-db.org/) was used to construct the PPI network and cytoscape MCODE plug-in was used to screen the hub genes (17). And Metascape database (https://metascape.org/) and David database (https://david.ncifcrf.gov/) were used to perform the GO and KEGG pathways enrichment analysis to detect the potential biological functions of common genes (11, 12). For the hub genes, the UALCAN database (http://ualcan.path.uab.edu/) was used to analyze the correlation between hub genes expression and clinical traits. For the methylation analysis, We explored the association between hub genes' expression levels and their methylation status by Disease Meth version 2.0 (http://bio-bigdata.hrbmu.edu.cn/diseasemeth/) and MethSurv (https://biit.cs.ut.ee/methsurv/). Besides, the online database TIMER2.0 (http://timer.cistrome.org/) was used to verify that the above key nodes in ceRNA network and hub genes were the common features in alimentary tract malignancies.

RESULT

Differentially Expressed Genes in ESCA, STAD, COAD, and READ Based on TCGA Data

In total, 3,999 mRNAs were identified as differentially expressed (DE) mRNAs in ESCA, including 1,776 up-regulated and 2,223 down-regulated mRNAs (**Supplementary Figure 2A** and **Supplementary Table 1**). For STAD, 6,520 mRNAs were identified as DE mRNAs, including 3,426 up-regulated and 3,094 down-regulated mRNAs (**Supplementary Figure 2B** and **Supplementary Table 1**). For COAD, 7,635 mRNAs were screened as DE mRNAs, including 4,203 up-regulated and 3,432 down-regulated mRNAs (**Supplementary Figure 2C** and **Supplementary Table 1**). And there were 7,184 DE mRNAs in READ, consisting of 3,780 up-regulated mRNAs and 3,404 down-regulated mRNAs (**Supplementary Figure 2D** and **Supplementary Table 1**). Finally, the number of common mRNAs was 1,700 among four tumors (**Figure 1B**).

Altogether, 187 miRNAs were considered as DE miRNAs in ESCA including 106 up-regulated and 81 down-regulated miRNAs (**Supplementary Figure 2E** and **Supplementary Table 2**). For STAD, 103 DE miRNAs were screened out, including 85 up-regulated and 18 down-regulated miRNAs (**Supplementary Figure 2F** and **Supplementary Table 2**). For COAD, 231 DE miRNAs were called out, including 133 up-regulated and 98 down-regulated miRNAs (**Supplement Figure 2G** and **Supplement Table 2**). Besides, there were 245 DE miRNAs in READ, including 174 up-regulated and 71 down-regulated miRNAs (**Supplementary Figure 2H** and **Supplementary Table 2**). In the end, there were 29 common DE miRNAs among the digestive tract cancers (**Figure 1C**).

We got 1,448 DE lncRNAs in ESCA, consisting of 744 up-regulated and 704 down-regulated lncRNAs (**Supplementary Figure 2I** and **Supplementary Table 3**). And for STAD, there were 4,422 DE lncRNAs, including 3,226 up-regulated and 1,196 down-regulated lncRNAs (**Supplementary Figure 2J** and **Supplementary Table 3**). And for COAD, there were 3,824 DE lncRNAs, including 2,584 up-regulated and 1,240 down-regulated lncRNAs (**Supplementary Figure 2K** and **Supplementary Table 3**). The number of DE lncRNAs in READ was 2,456, including 1,281 up-regulated and 1,175 down-regulated lncRNAs (**Supplementary Figure 2L** and **Supplementary Table 3**). Finally, the number of the common DE lncRNAs was 362 (**Figure 1D**).

Besides, as for the common up-regulated genes and down-regulated genes, the number of common up-regulated mRNAs was 769 and common down-regulated mRNAs was 808. About common up-regulated lncRNAs and down-regulated lncRNAs, the number was 216 and 135, respectively. Besides, there were 11 common up-regulated miRNAs and 12 common up-regulated miRNAs among the four cancers. And the common up-regulated genes and down-regulated genes were used to construct the ceRNA network.

Gene Set Enrichment Analysis (GSEA) of Different Cancers

Aiming to explore the heterogeneity of digestive tract cancers, Gene set enrichment analysis (GSEA) was used to show the hallmarks of different cancers (18). The results of GSEA showed that 11 genes sets were enriched in ESCA such as G2/M checkpoint, E2F targets and KRAS signaling set. For STAD, the number of enriched gene sets was 14, including G2/M
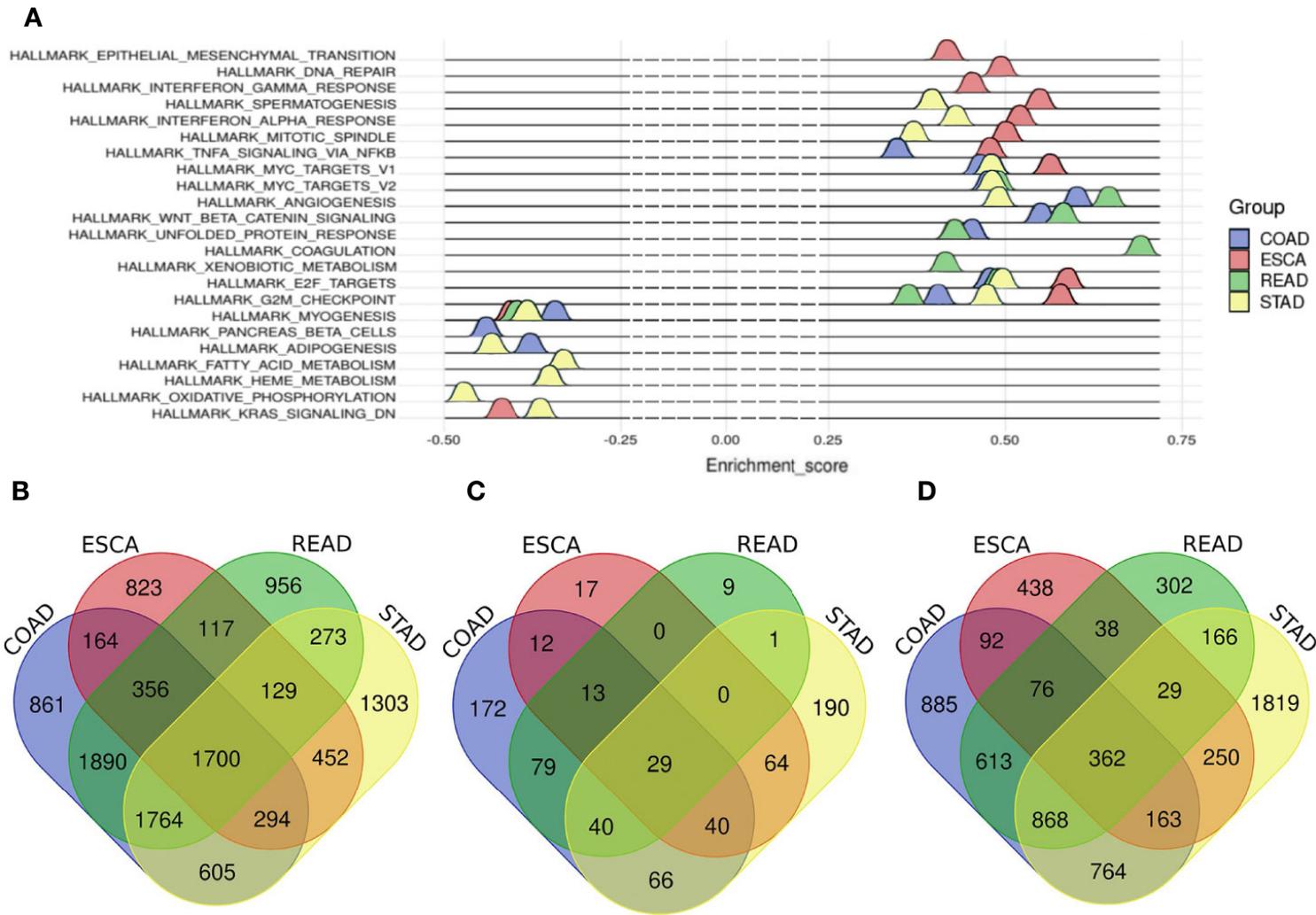
FIGURE 1 | The hallmarks of different tumors and Venn diagram of common m RNAs, common miRNAs and common lncRNAs among four tumors. **(A)** The hallmarks of different tumors. Blue represents COAD, red represents ESCA, green represents READ, and cyan represents STAD. **(B)** common differentially expressed mRNAs, **(C)** common differentially expressed miRNAs, **(D)** common differentially expressed lncRNAs. Blue represents COAD, red represents ESCA, green represents READ and yellow represents STAD.

checkpoint, E2F targets and adipogenesis. And there were 11 terms enriched in COAD and 10 terms in READ. For instance, the gene sets were enriched in G2/M checkpoint, MYC targets and E2F targets in COAD and coagulation and MYC targets in READ (**Figure 1A** and **Supplementary Table 4**).

GO and KEGG Pathways Enrichment Analysis of Common mRNAs

To gain insight into the biological functions of four tumors in common, GO and KEGG pathways analysis (11, 12) were performed on 1,700 shared DE mRNAs. GO terms were enriched in biological processes (BP) "DNA replication", "cell division", "G1/S transition of mitotic cell cycle", and "mitotic nuclear division" of biological processes (**Figure 2A** and **Supplementary Table 5**). The KEGG pathways analysis showed that the common DE mRNAs were significantly enriched in "Cell cycle", "DNA replication", "ECM-receptor interaction", and "Calcium signaling pathway" (**Figure 2B** and **Supplementary Table 6**). And Cytoscape BiNGo plug-in was utilized to visualize the major GO terms, which was shown in **Figure 2C** (13). Besides, to show the overall enrichment, we reset p-value <0.001 in **Supplementary Figure 3**.

The Construction of ceRNA Network

The Competing endogenous RNAs (ceRNAs) regulate RNA transcripts by competitively binding with shared miRNAs (15). According to recent investigations, ceRNAs play important parts in cancer initiation and progression and have cancer-specificity, supporting ceRNAs serve as diagnostic biomarkers or therapeutic targets (19). In the present study, the ceRNA network showed that there were 5 core miRNA nodes, 21 lncRNA nodes, 21 mRNAs nodes, and 66 edges (**Figure 3A**). We selected the negative interactions from the ceRNA network to construct regulatory axes, including 16 lncRNAs, 5 miRNAs, and 12 mRNAs (**Figure 3B**). For the lncRNA, the small nucleolar RNA host gene 14 (SNHG14), which had the highest connectivity and co-expression relationships with mRNAs of the ceRNA network, the ImmunLnc database showed that SNHG14 took part in immune pathways such as antigen processing and presentation in four tumors (**Supplementary Figures 4C–J**).

The correlations between ceRNAs and the survival outcomes of patients were analyzed to validate the prognostic value. The results showed serpin peptidase inhibitor (SERPINE1), hsa-mir-145 and small nucleolar RNA host gene 1 (SNHG1) were significantly correlated with overall survival (**Figures 3C–E**). Besides, they formed a potential regulatory axis SERPINE1/ hsa-mir-145/SNHG1 based on sponge effect in ceRNA network. And the correlations between the above three genes and overall survival were shown in **Supplementary Figure 5**.

The Weighted Gene Co-Expression Network Analysis (WGCNA)

We used gene co-expression network analysis to screen gene modules which were closely related to clinical traits. For ESCA, transcriptome were divided into 18 modules and the blue, brown,
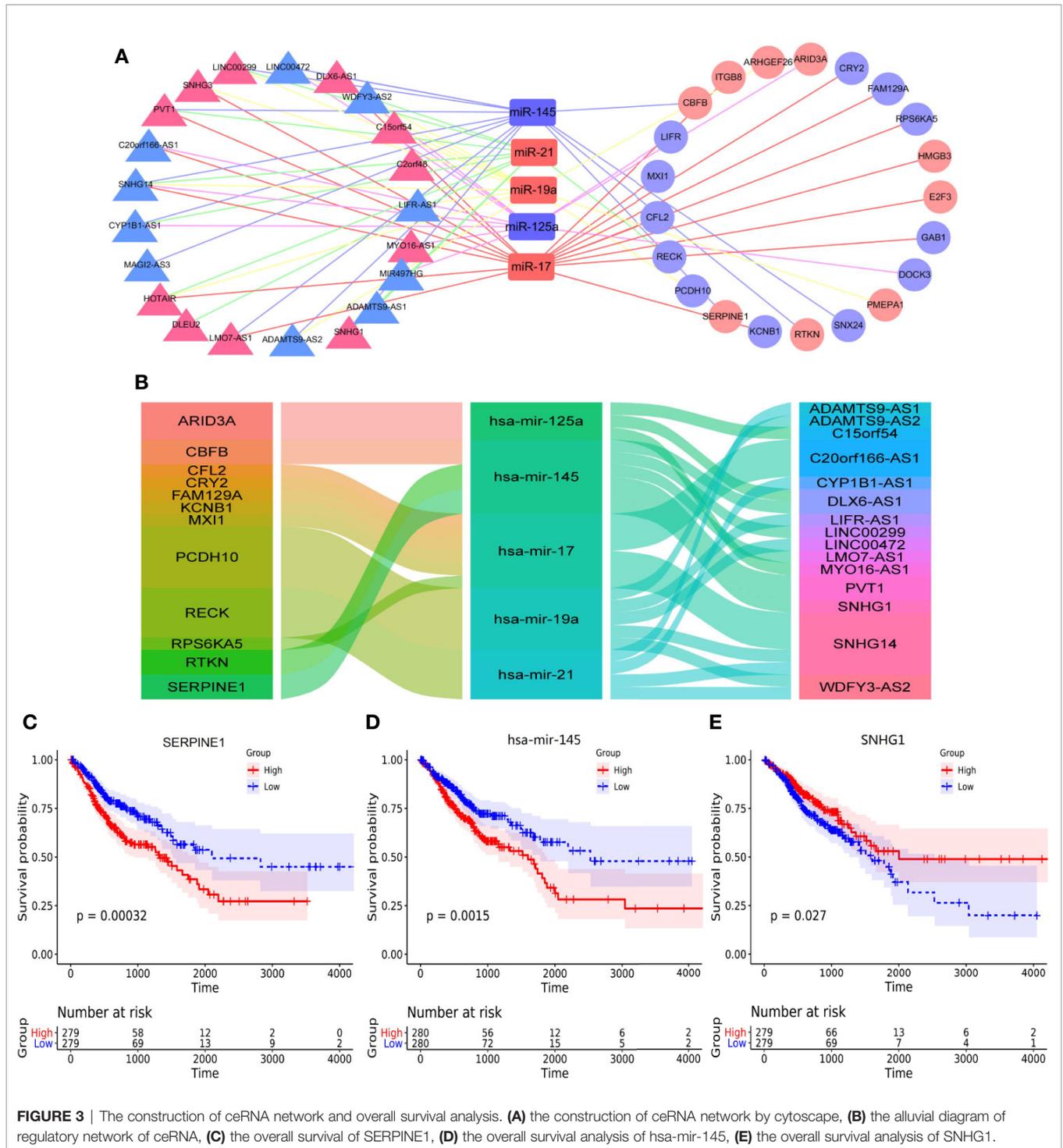
and salmon modules were related to clinical traits (**Figures 4A, E**). For STAD, genes were divided into 22 modules and there was no module screened associated to clinical traits (**Figure 4B**). For COAD, genes were divided into 34 modules and the cyan, pink, and purple modules were screened as the modules which were related to clinical traits (**Figures 4C, E**). For READ, the genes were divided into 33 modules and 14 of them were screened correlated to clinical traits (**Figures 4D, E**). Thus, there were 20 modules related to clinical traits in total. The relationships between other modules and clinical traits were shown in **Supplementary Figure 6**.

Besides, we took the intersection of modules related to the same clinical traits among different cancers. As a result, ESCA and READ had three common genes in neoplasm histologic grade and three common genes in the trait of age. And there were 30 common genes in the trait of overall survival between ESCA and READ. For the trait of race, we found 10 common genes between ESCA and COAD. Last but not the least, about the trait of histological types, there were 4 genes shared by ESCA, COAD and READ, 31 common genes between READ and ESCA, 32 common genes between COAD and ESCA and 26 common genes between COAD and READ. Thus, in total, the number of common genes between ESCA and READ was 67. Then between COAD and ESCA the number was 42 and between COAD and READ being 26. For the common genes between different tumors, GO and KEGG pathways enrichment analysis were performed to show the functions of these genes (**Supplementary Figure 7**). Considering that the number of common genes between READ and ESCA was the highest, these 67 common genes were performed further analysis. The results showed that they were significantly enriched in extracellular structure organization, chemotaxis, and positive regulation of cellular component movement (**Figures 5A–C**). And it screened out 5 hub genes of the 67 genes by protein-protein interaction (PPI) by cytoscape Mcode plug-in (**Figure 6A**) (17). Among them, the UALCAN database showed that CXCL3 was related to the trait of histological types in ESCA and READ, GPR44 (PTGDR2) was associated with tumor grade in ESCA and NPBWR1 was related to patients' age in ESCA and READ (**Figures 6B–F**). Furthermore, Disease Meth version 2.0 analysis showed that the mean methylation levels of NPBWR1 was higher in tumor tissue than normal tissue both in ESCA and READ (P < 0.01) (**Supplementary Figures 8A, B**).

In order to verify above genes representing the common features in alimentary tract malignancies, the exploration of database TIMER2.0 (http://timer.cistrome.org/) showed that SERPINE1, SNHG1, ADCY6, CXCL3, ADCY6, CXCL3, NPBWR1, TAS2R38, and PTGDR2 were mainly differentially expressed in alimentary tract malignancies (**Supplementary Figure 9**).

DISCUSSION

The morbidity of alimentary tract malignancies has declined though, it is still the leading cause of death in malignant tumors

FIGURE 2 | GO and KEGG enrichment analysis. **(A)** GO enrichment analysis including BP, CC, and MF, **(B)** KEGG pathway enrichment analysis, **(C)** the directed acycliic graph of GO terms. (P < 0.0001).

worldwide nowadays (2). Thus, it remains an essential task to explore the carcinogenesis and mechanisms in tumor growth. There are many investigations using RNA-seq to discover novel biomarkers and therapeutic targets for the digestive tract cancers, such as miR-378, miR-199a, and miR-92a which are closely related to survival rate of colon cancer (20). However, lots of investigations put emphasis on a certain type of cancer but relatively little concrete on their homogeneity. According to recent researches, upper gastrointestinal (GI) tumors with chromosomal instability are characterized in fragmented

FIGURE 3 | The construction of ceRNA network and overall survival analysis. **(A)** the construction of ceRNA network by cytoscape, **(B)** the alluvial diagram of regulatory network of ceRNA, **(C)** the overall survival of SERPINE1, **(D)** the overall survival analysis of hsa-mir-145, **(E)** the overall survival analysis of SNHG1.

genomes and lower GI tumors are enriched in mutations in SOX9 and PCBP1, which reveal the heterogeneity between upper GI tumors and lower GI tumors (21).

In this study, heterogeneity is analyzed among four tumors from the point of tumor hallmarks (22). And we mined 1,700 common DE mRNAs, 336 DE lncRNAs, and 29 DE miRNAs, mainly involved in functional terms where pathways and gene

clusters are widely reported in recent investigations and play important roles in the development of digestive tract cancers (23–27).

By the construction of ceRNA network, interactions among various types of RNA molecules reveal the frank regulation inside the whole transcriptome. Herein, TCGA malignancies of alimentary tract share the core ceRNA network with nodes of 5

FIGURE 4 | Identification of weighted gene co-expression network modules. Different colors represent different gene modules **(A)** ESCA, **(B)** STAD, **(C)** COAD, and **(D)** READ. **(E)** the heatmap of correlation between modules and clinical traits.
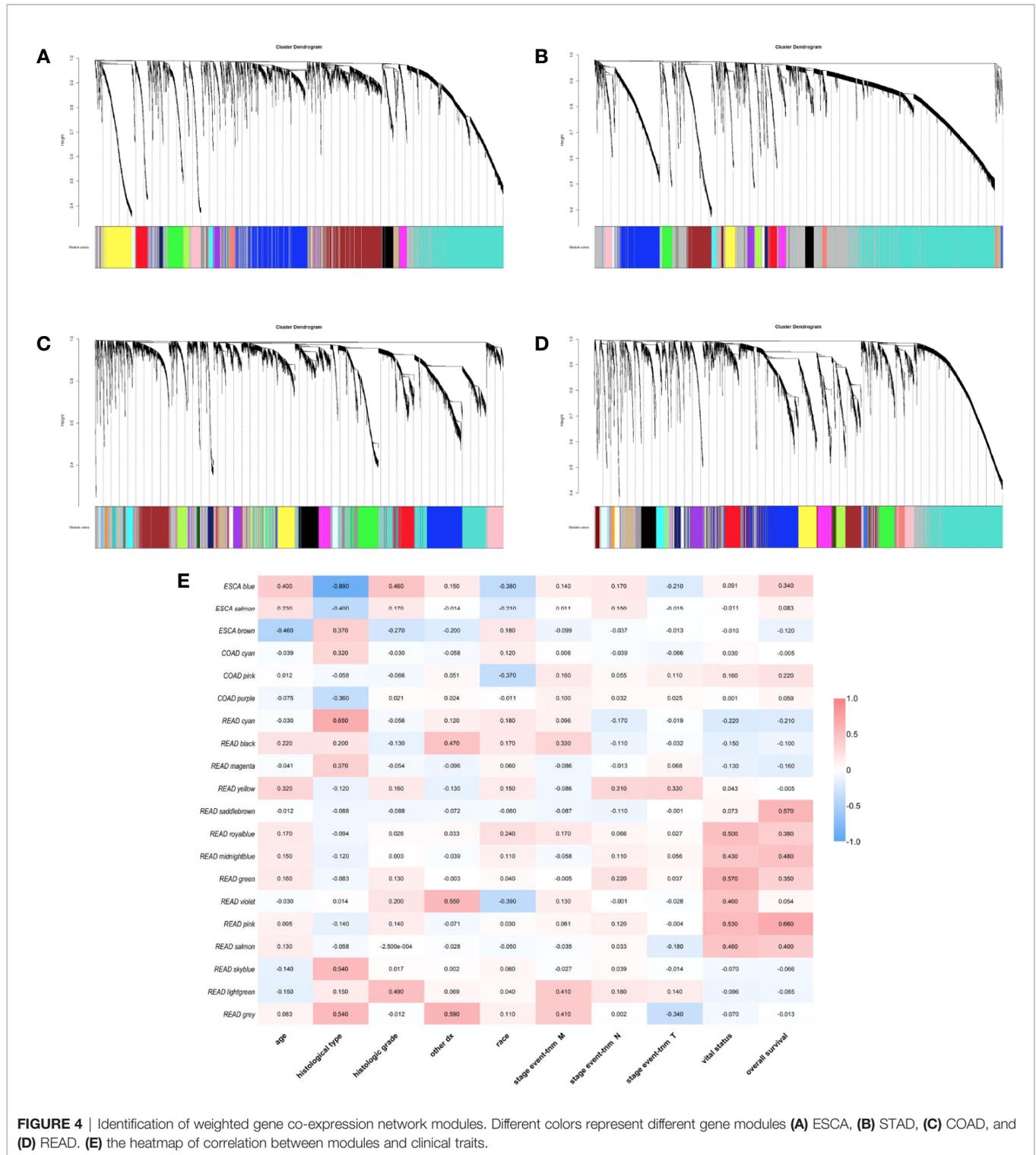
miRNAs, 21 lncRNAs, and 21 mRNAs. The SNHG14, one of lncRNAs, had the highest connectivity, with all 5 miRNAs in the network. Recently, it is reported that high expression of SNHG14 in tumor tissue promotes cancer cell invasion and metastasis by targeting hsa-mir-145 in gastric cancer and hsa-mir-944 in colorectal cancer (28, 29). On the contrary, there are some investigations showing that up-regulation of SNHG14 suppresses cell proliferation and metastasis of colorectal cancer by targeting hsa-mir-92b and hsa-mir-186 (30, 31). Intriguingly, the data of GEPIA2 showed that SNHG14 is a down–regulated lncRNA in gastrointestinal cancers in comparison to normal tissue and its expression is positively correlated with stage

FIGURE 5 | GO enrichment analysis. **(A)** the relation between common genes and GO terms of ESCA and READ, **(B)** Network of enriched terms: Colored by cluster ID, **(C)** the P-value of GO terms of ESCA and READ.

development (**Supplementary Figures 4A, B**). These contradictory results are probably caused by that they set different normal cell lines as control groups. For instance, the expression of SNHG14 in FHC cell lines is higher than other cancer cells but is lower in NCM460 cell lines (28, 30–33). Thus, the function of SNHG14 is complicated and might serve as both positive and negative roles in cancer initiation and progression corresponding to bio-context. Moreover, SNHG14 is associated with immunity and could be a promising target for the immunotherapy through PD-L1/PD-1 blockade in cancers (34). Our exploration based on ImmunLnc database supports this story that SNHG14 is associated with antigen processing and presentation (**Supplementary Figure 4J**).

As for prognostic values, our research illustrates that SERPINE1, hsa-mir-145, and SNHG1 are closely related to survival rate in the common ceRNA network of four digestive tract tumors. Hsa-mir-145 is an important microRNA widely regarded as a tumor suppressor, of which expression is down-regulated in various cancers (35). Our research shows that hsa-mir-145 is related to the overall survival of gastrointestinal cancers, consistent with latest finding in COAD, ESCA, and

STAD (36–38). Besides, it is reported that SERPINE1 is significantly associated with poor outcomes in gastric cancer and a potential biomarker for tumor aggressiveness in colorectal cancer (39, 40). Moreover, as a member of small nucleolar RNA host lncRNA family, SNHG1 acts as a useful tumor biomarker for cancer diagnosis, prognosis, and treatment (41). Three molecules above form a ceRNA regulatory axis, in which our research prompts that SNHG1 possibly performed as sponge to provoke the effect of hsa-mir-145 on the up-regulation of SERPINE1, enhancing tumor cell migration and invasion (42). Meanwhile, the SERPINE1/hsa-mir-145/SNHG1 axis is associated with overall survival and could serve as a critical target for therapy. Besides, to explore therapeutic value of the key nodes in other cancers, we found that SERPINE1/hsa-mir-145/SNHG1 axis mainly expressed in alimentary tract malignancies but not other cancers.

WGCNA is an R package for weighted correlation network analysis and is widely used in bioinformatics analysis. In order to explore the relationships between genes and clinical features, the R WGCNA package is used to construct Co-expression networks. WGCNAs yield 20 modules related to

FIGURE 6 | The identification of hub genes and the correlation between hub genes expression and clinical traits. **(A)** the interaction network of hub genes, **(B)** the correlation between CXCL3 expression and tumor histological types in ESCA, **(C)** the correlation between CXCL3 expression and tumor histological types in READ, **(D)** the correlation between GPR44 (PTGDR2) expression and tumor grades in ESCA, **(E)** the correlation between NPBWR1 expression and ages in READ, **(F)** the correlation between NPBWR1 expression and ages in ESCA. *p < 0.05, **p < 0.01, ***p < 0.001.

clinical traits and reveal common features shared between ESCA and READ. For instance, ESCA and READ have 67 common genes in the clinical traits of age, histological types, neoplasm histologic grade and overall survival. About the GO and KEGG pathways enrichment, the 67 common genes are related to chemotaxis, extracellular structure organization, acylglycerol homeostasis and regulation of lipid transport. Considering the similarities of ESCA and READ, esophagus and rectum are the beginning and end parts of the alimentary tract, both consisting of stratified squamous epithelium and simple columnar epithelium and functioning as the cargo of food or residue on the border between inner digestive system and outside. Considering their similar biofunction and bi-category tissue constitution, ESCA and READ should share the common epithelial carcinogenesis.

To explore the relationships of common genes between ESCA and READ, we screen the hub genes including, NPBWR1, TAS2R38, CXCL3, ADCY6, and PTGDR2, which are closely related to clinical traits and mainly express in alimentary tract malignancies. Neuropeptide B/W receptor-1 (NPBWR1/GPR7), a G protein-coupled receptor, implicates in the biological process of energy homeostasis, neuroendocrine function and modulating

inflammatory pain (43, 44). And methylation of NPBWR1 is significantly associated with prostate cancer prognosis (45). Similar with prostate, our research proves NPBWR1 expression in ESCA and READ is significantly associated with its methylation. From the data of MethSurv, we find that with age increasing, the methylation level is increasing, contrary to the expression of NPBWR1 (**Supplementary Figure 8D**). Thus, it can be deduced that age influences the expression of NPBWR1 by affecting the methylation level in ESCA. Similar with NPBWR1, PTGDR2, also named CRTH2 and GPR44, is another G protein-coupled receptor and the receptor of PGD2, preferentially expressing in CD4+ effector T helper 2 cells (46). Previous studies show that PTGDR2 inhibits tumor growth and tumorigenesis in gastric cancer and restricts angiogenesis in colon cancer (47, 48). Our research shows that PTGDR2 is associated with the tumors' grades in ESCA and its expression is higher with tumor grades increasing.

For the other hub genes, ADCY6 encodes a member of the adenylate cyclase protein family, and it is required for the synthesis of cAMP (49). In our research, the ADCY6 is associated with the clinical traits of overall survival in READ and ESCA. For the hub gene TAS2R38, encoding bitterness taste

receptor, influences the progression of adenomatous polyps by dietary vitamin C and folic acid (50). Other studies show that genetic variation in TAS2R38 is thought to be associated with gastrointestinal risks by modifying dietary intake (51, 52). In our research, this gene is associated with the trait of overall survival and the mean methylation levels of TAS2R38 was lower in READ (**Supplementary Figure 8C**).

CXCL3 is also known as CINC-2 alpha (Cytokine-induced neutrophil Chemoattractant) and is a number of signaling pathways such as ERK1/2 MAPK, by activating CXCR2 receptor (53–55). There are investigations showing that CXCL3 is associated with vascular invasion and tumor capsule formation in HCC and is targeted by IRF2 to suppress MDSC migration and infiltration in colorectal cancer (54, 56). In our research, we find that the CXCL3 is associated with tumor histology types and is in high expression in adenocarcinoma, especially in Mucinous adenocarcinoma. Besides, these five hub genes all take part in the G alpha (i) signaling pathway. NPBWR1 and PTGDR2 are both G protein-coupled receptors and TAS2R38 encodes a seven-transmembrane G protein-coupled receptor associated with taste glucosinolates (44, 57, 58). ADCY6 catalyzes the formation of the signaling molecule cAMP downstream of G protein-coupled receptors (49, 59). And CXCL3 encodes a secreted growth factor that signals through the G-protein coupled receptor, CXC receptor 2 (60, 61). Thus, these genes have co-expression relationships in G alpha (i) signaling pathway and reason the similarity in ESCA and READ.

In conclusion, by constructing ceRNA network and using WGCNA analysis, we find a potential regulatory axis and several significant gene modules in alimentary tract cancers. For the clinical treatment of these malignancies, the axis SERPINE1/hsa-mir-145/SNHG1 could serve as prognostic biomarkers and potential therapeutic targets. The hub gene ADCY6, CXCL3, NPBWR1, TAS2R38, and PTGDR2 highly influence the clinical traits of age, histology types, neoplasm histologic grade, and overall survival in ESCA and READ, leading to their similarity in these clinical traits. Further study is guaranteed for sure, aiming to identify the therapeutic value of SERPINE1/hsa-mir-145/

SNHG1 axis and fully reveal the hub genes contribution to the clinical outcomes.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

KY and D-BJ developed the original idea. KY, D-BJ, and L-HZ designed the whole study. Y-CL, J-QS, Z-XZ, and J-YZ ran the most of data processing. Y-CL, H-KZ, and Y-CF did data visualization. Z-HL and Z-CL helped to access the public data. S-YY, X-YZ, YL, and Y-JS reviewed the manuscript and raised suggestions. Y-CL and D-BJ drafted the manuscript. D-BJ and J-QS revised the writings. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.580276/full#supplementary-material

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492

2. Arnold M, Abnet CC, Neale RE, Vignat J, Giovannucci EL, McGlynn KA, et al. Global Burden of 5 Major Types Of Gastrointestinal Cancer. *Gastroenterology* (2020) 159(1):335–49. doi: 10.1053/j.gastro.2020.02.068

3. Abolarinwa BA, Ibrahim RB, Huang Y. Conceptual Development of Immunotherapeutic Approaches to Gastrointestinal Cancer. *Int J Mol Sci* (2019) 20(18):4624. doi: 10.3390/ijms20184624

4. Zheng Q, Chen C, Guan H, Kang W, Yu C. Prognostic role of microRNAs in human gastrointestinal cancer: A systematic review and meta-analysis. *Oncotarget* (2017) 8(28):46611–23. doi: 10.18632/oncotarget.16679

5. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol* (2008) 5(10):588–99. doi: 10.1038/ncponc1187

6. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LJ, Kinzler KW. Cancer genome landscapes. *Science* (2013) 339(6127):1546–58. doi: 10.1126/science.1235122

7. Giordano TJ. The cancer genome atlas research network: a sight to behold. *Endocr Pathol* (2014) 25(4):362–5. doi: 10.1007/s12022-014-9345-4

8. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* (2013) 45(10):1113–20. doi: 10.1038/ng.2764

9. Bijlsma MF, Sadanandam A, Tan P, Vermeulen L. Molecular subtypes in cancers of the gastrointestinal tract. *Nat Rev Gastro Hepat* (2017) 14(6):333–42. doi: 10.1038/nrgastro.2017.33

10. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2010) 26(1):139–40. doi: 10.1093/bioinformatics/btp616

11. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* (2000) 25(1):25–9. doi: 10.1038/75556

12. Altermann E, Klaenhammer TR. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics* (2005) 6:60. doi: 10.1186/1471-2164-6-60

13. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* (2005) 21(16):3448–9. doi: 10.1093/bioinformatics/bti551

14. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* (2003) 13(11):2498–504. doi: 10.1101/gr.1239303

15. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* (2011) 146(3):353–8. doi: 10.1016/j.cell.2011.07.014

16. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* (2008) 9(1):559. doi: 10.1186/1471-2105-9-559

17. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf* (2003) 4(1):2. doi: 10.1186/1471-2105-4-2

18. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102

19. Qi X, Zhang D, Wu N, Xiao J, Wang X, Ma W. ceRNA in cancer: possible functions and clinical implications. *J Med Genet* (2015) 52(10):710–8. doi: 10.1136/jmedgenet-2015-103334

20. Weng W, Feng J, Qin H, Ma Y, Goel A. An update on miRNAs as biological and clinical determinants in colorectal cancer: a bench-to-bedside approach. *Future Oncol* (2015) 11(12):1791–808. doi: 10.2217/fon.15.83

21. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. *Cancer Cell* (2018) 33(4):721–35. doi: 10.1016/j.ccell.2018.03.010

22. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144(5):646–74. doi: 10.1016/j.cell.2011.02.013

23. Walker RC, Underwood TJ. Molecular pathways in the development and treatment of oesophageal cancer. *Best Pract Res Clin Gastroenterol* (2018) 36-37:9–15. doi: 10.1016/j.bpg.2018.11.013

24. Lin EW, Karakasheva TA, Hicks PD, Bass AJ, Rustgi AK. The tumor microenvironment in esophageal cancer. *Oncogene* (2016) 35(41):5337–49. doi: 10.1038/onc.2016.34

25. Berger H, Marques MS, Zietlow R, Meyer TF, Machado JC, Figueiredo C. Gastric cancer pathogenesis. *Helicobacter* (2016) 21:34–8. doi: 10.1111/hel.12338

26. Figueiredo C, Camargo MC, Leite M, Fuentes-Pananá EM, Rabkin CS, Machado JC. Pathogenesis of Gastric Cancer: Genetics and Molecular Classification. *Curr Top Microbiol Immunol* (2017) 400:277. doi: 10.1007/978-3-319-50520-6_12

27. Al-Sohaily S, Biankin A, Leong R, Kohonen-Corish M, Warusavitarne J. Molecular pathways in colorectal cancer. *J Gastroen Hepatol* (2012) 27(9):1423–31. doi: 10.1111/j.1440-1746.2012.07200.x

28. Pei Q, Liu GS, Li HP, Zhang Y, Xu XC, Gao H, et al. Long noncoding RNA SNHG14 accelerates cell proliferation, migration, invasion and suppresses apoptosis in colorectal cancer cells by targeting miR-944/KRAS axis through PI3K/AKT pathway. *Eur Rev Med Pharmacol Sci* (2019) 23(22):9871–81. doi: 10.26355/eurrev_201911_19551

29. Liu Z, Yan Y, Cao S, Chen Y. Long non-coding RNA SNHG14 contributes to gastric cancer development through targeting miR-145/SOX9 axis. *J Cell Biochem* (2018) 119(8):6905–13. doi: 10.1002/jcb.26889

30. Han Y, Zhou S, Wang X, Mao E, Huang L. SNHG14 stimulates cell autophagy to facilitate cisplatin resistance of colorectal cancer by regulating miR-186/ATG14 axis. *BioMed Pharmacother* (2020) 121:109580. doi: 10.1016/j.biopha.2019.109580

31. Zhang W, Duan W, Mo Z, Wang J, Yang W, Wu W, et al. Upregulation of SNHG14 suppresses cell proliferation and metastasis of colorectal cancer by targeting miR-92b-3p. *J Cell Biochem* (2019) 121(2):1998–2008. doi: 10.1002/jcb.29434

32. Matboli M, Shafei AE, Ali MA, El-Din Ahmed TS, Naser M, Abdel-Rahman T, et al. Role of extracellular LncRNA-SNHG14/miRNA-3940-5p/NAP12 mRNA in colorectal cancer. *Arch Physiol Biochem* (2019) 125:1–7. doi: 10.1080/13813455.2019.1650070

33. Di W, Weinan X, Xin L, Zhiwei Y, Xinyue G, Jinxue T, et al. Long noncoding RNA SNHG14 facilitates colorectal cancer metastasis through targeting EZH2-regulated EPHA7. *Cell Death Dis* (2019) 10(7):514. doi: 10.1038/s41419-019-1707-x

34. Zhao L, Liu Y, Zhang J, Liu Y, Qi Q. LncRNA SNHG14/miR-5590-3p/ZEB1 positive feedback loop promoted diffuse large B cell lymphoma progression and immune evasion through regulating PD-1/PD-L1 checkpoint. *Cell Death Dis* (2019) 10(10):731. doi: 10.1038/s41419-019-1886-5

35. Xu WX, Liu Z, Deng F, Wang DD, Li XW, Tian T, et al. MiR-145: a potential biomarker of cancer migration and invasion. *Am J Transl Res* (2019) 11(11):6739–53.

36. Tabrizi M, Khalili M, Vasei M, Nouraei N, Mansour SN, Khavanin A, et al. Evaluating the miR-302b and miR-145 expression in formalin-fixed paraffin-embedded samples of esophageal squamous cell carcinoma. *Arch Iran Med* (2015) 18(3):173–8. doi: 10.151803/AIM.007

37. Liang Y, Zhang C, Ma M, Dai D. Identification and prediction of novel non-coding and coding RNA-associated competing endogenous RNA networks in colorectal cancer. *World J Gastroentero* (2018) 24(46):5259–70. doi: 10.3748/wjg.v24.i46.5259

38. Liu J, Liu F, Shi Y, Tan H, Zhou L. Identification of key miRNAs and genes associated with stomach adenocarcinoma from The Cancer Genome Atlas database. *FEBS Open Bio* (2018) 8(2):279–94. doi: 10.1002/2211-5463.12365

39. Mazzoccoli G, Pazienza V, Panza A, Valvano MR, Benegiamo G, Vinciguerra M, et al. ARNTL2 and SERPINE1: potential biomarkers for tumor aggressiveness in colorectal cancer. *J Cancer Res Clin Oncol* (2012) 138(3):501–11. doi: 10.1007/s00432-011-1126-6

40. Zhou X, Huang X, Liang S, Tang S, Wu S, Huang T, et al. Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis. *Onco Targets Ther* (2018) 11:2815–30. doi: 10.2147/OTT.S163891

41. Thin KZ, Tu JC, Raveendran S. Long non-coding SNHG1 in cancer. *Clin Chim Acta* (2019) 494:38–47. doi: 10.1016/j.cca.2019.03.002

42. Simone TM, Longmate WM, Law BK, Higgins PJ. Targeted Inhibition of PAI-1 Activity Impairs Epithelial Migration and Wound Closure Following Cutaneous Injury. *Adv Wound Care* (2015) 4(6):321–8. doi: 10.1089/wound.2014.0611

43. Sakurai T. NPBWR1 and NPBWR2: Implications in Energy Homeostasis, Pain, and Emotion. *Front Endocrinol* (2013) 4:23. doi: 10.3389/fendo.2013.00023

44. O'Dowd BF, Scheideler MA, Nguyen T, Cheng R, Rasmussen JS, Marchese A, et al. The cloning and chromosomal mapping of two novel human opioid-somatostatin-like receptor genes, GPR7 and GPR8, expressed in discrete areas of the brain. *Genomics* (1995) 28(1):84–91. doi: 10.1006/geno.1995.1109

45. Cottrell S, Jung K, Kristiansen G, Eltze E, Semjonow A, Ittmann M, et al. Discovery and Validation of 3 Novel DNA Methylation Markers of Prostate Cancer Prognosis. *J Urol* (2007) 177(5):1753–8. doi: 10.1016/j.juro.2007.01.010

46. Hirai H, Tanaka K, Yoshie O, Ogawa K, Kenmotsu K, Takamori Y, et al. Prostaglandin D2 selectively induces chemotaxis in T helper type 2 cells, eosinophils, and basophils via seven-transmembrane receptor CRTH2. *J Exp Med* (2001) 193(2):255–61. doi: 10.1084/jem.193.2.255

47. Iwanaga K, Nakamura T, Maeda S, Aritake K, Hori M, Urade Y, et al. Mast cell-derived prostaglandin D2 inhibits colitis and colitis-associated colon cancer in mice. *Cancer Res* (2014) 11(74):3011–9. doi: 10.1158/0008-5472.CAN-13-2792

48. Zhang B, Bie Q, Wu P, Zhang J, You B, Shi H, et al. PGD2/PTGDR2 Signaling Restricts the Self-Renewal and Tumorigenesis of Gastric Cancer. *Stem Cells* (2018) 36(7):990–1003. doi: 10.1002/stem.2821

49. Gros R, Van Uum S, Hutchinson-Jaffe A, Ding Q, Pickering JG, Hegele RA, et al. Increased enzyme activity and beta-adrenergic mediated vasodilation in subjects expressing a single-nucleotide variant of human adenylyl cyclase 6. *Arterioscler Thromb Vasc Biol* (2007) 27(12):2657–63. doi: 10.1161/ATVBAHA.107.145557

50. Lucock M, Ng X, Boyd L, Skinner V, Wai R, Tang S, et al. TAS2R38 bitter taste genetics, dietary vitamin C, and both natural and synthetic dietary folic acid predict folate status, a key micronutrient in the pathoaetiology of adenomatous polyps. *Food Funct* (2011) 2(8):457–65. doi: 10.1039/c1fo10054h

51. Choi JH, Kim J. TAS2R38 Bitterness Receptor Genetic Variation and Risk of Gastrointestinal Neoplasm: A Meta-Analysis. *Nutr Cancer* (2019) 71(4):585–93. doi: 10.1080/01635581.2018.1559935

52. Lambert JD, VanDusen SR, Cockroft JE, Smith EC, Greenwood DC, Cade JE. Bitter taste sensitivity, food intake, and risk of malignant cancer in the UK Women's Cohort Study. *Eur J Nutr* (2019) 58(5):2111–21. doi: 10.1007/s00394-018-1772-4

53. Al-Alwan LA, Chang Y, Mogas A, Halayko AJ, Baglole CJ, Martin JG, et al. Differential roles of CXCL2 and CXCL3 and their receptors in regulating normal and asthmatic airway smooth muscle cell migration. *J Immunol* (2013) 191(5):2731–41. doi: 10.4049/jimmunol.1203421

54. Zhang L, Zhang L, Li H, Ge C, Zhao F, Tian H, et al. CXCL3 contributes to CD133(+) CSCs maintenance and forms a positive feedback regulation loop with CD133 in HCC via Erk1/2 phosphorylation. *Sci Rep* (2016) 6:27426. doi: 10.1038/srep27426

55. Doll D, Keller L, Maak M, Boulesteix AL, Siewert JR, Holzmann B, et al. Differential expression of the chemokines GRO-2, GRO-3, and interleukin-8 in colon cancer and their impact on metastatic disease and survival. *Int J Colorectal Dis* (2010) 25(5):573–81. doi: 10.1007/s00384-010-0901-1

56. Liao W, Overman MJ, Boutin AT, Shang X, Zhao D, Dey P, et al. KRAS-IRF2 Axis Drives Immune Suppression and Immune Therapy Resistance in Colorectal Cancer. *Cancer Cell* (2019) 35(4):559–72. doi: 10.1016/j.ccell.2019.02.008

57. Bufe B, Breslin PA, Kuhn C, Reed DR, Tharp CD, Slack JP, et al. The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. *Curr Biol* (2005) 15(4):322–7. doi: 10.1016/j.cub.2005.01.047

58. Marchese A, Sawzdargo M, Nguyen T, Cheng R, Heng HH, Nowak T, et al. Discovery of three novel orphan G-protein-coupled receptors. *Genomics* (1999) 56(1):12–21. doi: 10.1006/geno.1998.5655

59. Gao X, Sadana R, Dessauer CW, Patel TB. Conditional stimulation of type V and VI adenylyl cyclases by G protein betagamma subunits. *J Biol Chem* (2007) 282(1):294–302. doi: 10.1074/jbc.M607522200

60. Shibata F, Konishi K, Nakagawa H. Chemokine receptor CXCR2 activates distinct pathways for chemotaxis and calcium mobilization. *Biol Pharm Bull* (2002) 25(9):1217–9. doi: 10.1248/bpb.25.1217

61. Shibata F, Konishi K, Nakagawa H. Identification of a common receptor for three types of rat cytokine-induced neutrophil chemoattractants (CINCs). *Cytokine* (2000) 12(9):1368–73. doi: 10.1006/cyto.2000.0739