



Integrated PPI- and WGCNA-Retrieval of Hub Gene Signatures Shared Between Barrett's Esophagus and Esophageal Adenocarcinoma

Asma Sindhoo Nangraj¹, Gurudeeban Selvaraj^{2*}, Satyavani Kaliamurthi², Aman Chandra Kaushik^{1,3}, William C. Cho⁴ and Dong Qing Wei^{1,2,5*}

¹ The State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, ² Center of Interdisciplinary Sciences-Computational Life Sciences, Henan University of Technology, Zhengzhou, China, ³ Wuxi School of Medicine, Jiangnan University, Wuxi, China, ⁴ Department of Clinical Oncology, Queen Elizabeth Hospital, Hong Kong, China, ⁵ Peng Cheng Laboratory, Shenzhen, China

OPEN ACCESS

Edited by:

Weiwei Xue,
Chongqing University, China

Reviewed by:

Esra Gov,
Adana Science and Technology
University, Turkey
Shuyan Li,
Lanzhou University, China

*Correspondence:

Gurudeeban Selvaraj
gurudeeb99@haut.edu.cn
Dong Qing Wei
dqwei@sjtu.edu.cn

Specialty section:

This article was submitted to
Pharmacology of Anti-Cancer Drugs,
a section of the journal
Frontiers in Pharmacology

Received: 05 January 2020

Accepted: 28 May 2020

Published: 31 July 2020

Citation:

Nangraj AS, Selvaraj G, Kaliamurthi S,
Kaushik AC, Cho WC and Wei DQ
(2020) Integrated PPI- and WGCNA-
Retrieval of Hub Gene Signatures
Shared Between Barrett's Esophagus
and Esophageal Adenocarcinoma.
Front. Pharmacol. 11:881.
doi: 10.3389/fphar.2020.00881

Esophageal adenocarcinoma (EAC) is a deadly cancer with high mortality rate, especially in economically advanced countries, while Barrett's esophagus (BE) is reported to be a precursor that strongly increases the risk of EAC. Due to the complexity of these diseases, their molecular mechanisms have not been revealed clearly. This study aims to explore the gene signatures shared between BE and EAC based on integrated network analysis. We obtained EAC- and BE-associated microarray datasets GSE26886, GSE1420, GSE37200, and GSE37203 from the Gene Expression Omnibus and ArrayExpress using systematic meta-analysis. These data were accompanied by clinical data and RNAseq data from The Cancer Genome Atlas (TCGA). Weighted gene co-expression network analysis (WGCNA) and differentially expressed gene (DEG) analysis were conducted to explore the relationship between gene sets and clinical traits as well as to discover the key relationships behind the co-expression modules. A differentially expressed gene-based protein-protein interaction (PPI) complex was used to extract hub genes through Cytoscape plugins. As a result, 403 DEGs were excavated, comprising 236 upregulated and 167 downregulated genes, which are involved in the cell cycle and replication pathways. Forty key genes were identified using modules of MCODE, CytoHubba, and CytoNCA with different algorithms. A dark-gray module with 207 genes was identified which having a high correlation with phenotype (gender) in the WGCNA. Furthermore, five shared hub gene signatures (SHGS), namely, pre-mRNA processing factor 4 (PRPF4), serine and arginine-rich splicing factor 1 (SRSF1), heterogeneous nuclear ribonucleoprotein M (HNRNPM), DExH-Box Helicase 9 (DHX9), and origin recognition complex subunit 2 (ORC2), were identified between BE and EAC. SHGS enrichment denotes that RNA metabolism and splicosomes play a key role in

esophageal cancer development and progress. We conclude that the PPI complex and WGCNA co-expression network highlight the importance of phenotypic identifying hub gene signatures for BE and EAC.

Keywords: bioinformatics analysis, Barrett's esophagus, hub gene signature, esophageal adenocarcinoma, weighted gene co-expression network analysis, protein-protein interaction

INTRODUCTION

Esophageal cancer is a deadly cancer considering its high mortality rate, with 572,034 newly diagnosed cases and 508,585 deaths in 2018 (Bray et al., 2018). Esophageal cancer is classified into two subcategories: esophageal adenocarcinoma (EAC; distal esophagus) and esophageal squamous cell carcinoma (ESCC; proximal esophagus). It starts from the esophageal epithelium, the innermost layer of the esophagus (Rustgi and El-Serag, 2014). Esophageal cancer is a very complex disease, as its various subtypes have different risk factors, time trends, and geographic patterns (Analysis et al., 2017) (Montgomery et al., 2014; Lordick et al., 2016). According to the geographic variation, EAC is more common in economically advanced regions than in low-income countries (Chai et al., 2019). The common risk factors of EAC are age, male sex, obese, gastroesophageal reflux disease (GERD), cigarette smoking, and diet (low in vegetables and fruits). Cook et al. (2014) report some common symptoms like vomiting/nausea and heartburn in EAC and GERD. Besides, Barrett's esophagus (BE) is considered as a precursor for EAC. BE is a metaplastic transformation from the normal squamous mucosa of the esophagus to a columnar lining; its presence conveys a 30–40-fold increased risk of EAC (Schneider and Corley, 2015). The tumor development is a step-by-step process that comprises constant changes from erosive esophagitis to non-dysplastic BE, low-grade dysplasia, high-grade dysplasia, adenocarcinoma *in situ*, and finally invasive adenocarcinoma (Anaparthi and Sharma, 2014). Due to poor prognosis, over 40% of patients are diagnosed with high-grade dysplasia. Additionally, the 5-year survival rate is less than 20% despite the advances in diagnosis and treatment (Tramontano et al., 2017). Certainly, surgical therapy has improved the patient's survival yet it is not suitable for advanced-stage cancer patients (Davies et al., 2014).

Thus, it is essential to discover biomarkers that can lead to the discovery of medication. Microarray analysis of gene expression profiles is a common practice for identifying key hub genes and key pathways (Wei et al., 2018; Sadhu and Bhattacharyya, 2019). In the current era of integrated bioinformatics, acquiring data is not an issue; rather, normalization seems to be a tough job (Campain and Yang, 2010). Considering all of these notions, we designed an integrated study to find key hub genes associated with BE and EAC. First, we extracted BE- and EAC-associated microarray datasets from the Gene Expression Omnibus (GEO) and ArrayExpress using systematic meta-analysis as well as RNA-seq data from TCGA. Preprocessing and normalization were conducted for further analysis. DEGs were identified using linear models for microarray data (LIMMA) algorithm. Meta-

analysis was performed using a network analysis tool. We analyzed functional and pathway enrichment of DEGs. Additionally, a protein-protein interaction (PPI) network was constructed to study the associations between the DEGs and to recognize target genes using different modules of Cytoscape software. Weighted gene co-expression network analysis (WGCNA) was conducted by the construction of the co-expression network to find a correlation between modules and clinical traits. Furthermore, clinically significant modules were identified. Finally, key hub genes were identified and validated using immunohistochemistry and survival analysis.

MATERIALS AND METHODS

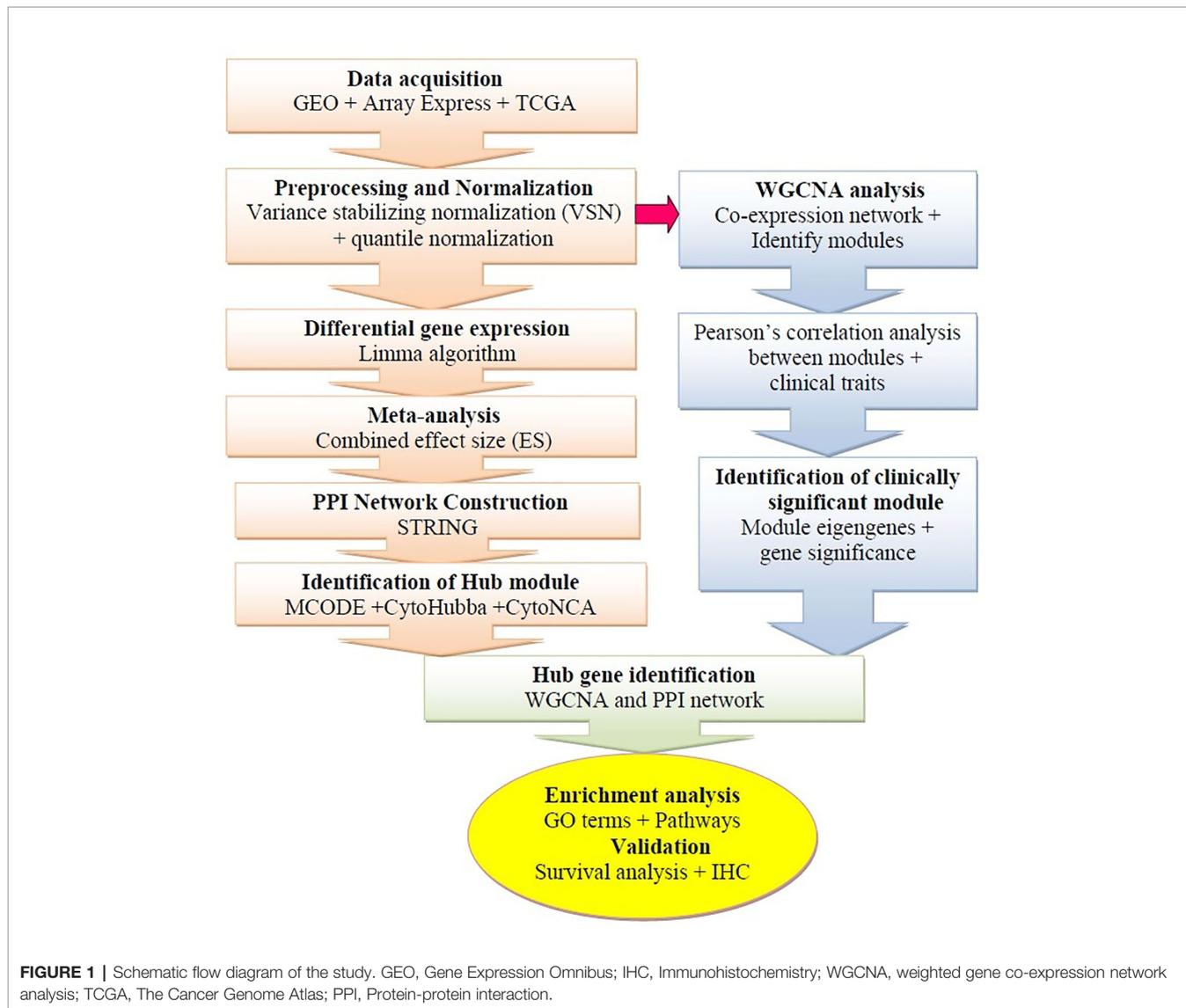
Data acquisition, Preprocessing, and Normalization

The microarray datasets were systematically extracted from the GEO¹ (Edgar et al., 2002) and the ArrayExpress² database (Brazma et al., 2003). The gene expression profiles based on RNA-sequencing were additionally obtained from The Cancer Genome Atlas (TCGA)³ (Zhu et al., 2014). The framework of this study is shown in **Figure 1**. For microarray profiles, we selected four datasets (GSE26886, GSE1420, GSE37200, and GSE37201) available by October 2019 (Kimchi et al., 2005; Silvers et al., 2010; Wang et al., 2013; Lin et al., 2015). The GEO accession number, sample size, description, platform, expression data, and references are extracted from each identified dataset (**Table 1**). The TCGA portal was accessed in October 2019, 184 esophageal cancer samples were retrieved. The tab-delimited text (.txt) files of microarray datasets were obtained from the GEO database. The Network Analyst (NA) web interface for integrative biological network analysis was employed for background correction preprocessing, normalization, probe identification, and meta-analysis of the datasets (Xia et al., 2015). The input files were prepared as per the description of the tool (first line #Name (sample ID); second line #class (sample type); genes in the rows and samples in the columns). We applied two different methods to normalize the datasets: first, variance stabilizing normalization (VSN), which improves DEG detection and reduces false-positive errors (Konishi, 1985), and second, quantile normalization, which can make two distributions equal in statistical methods (Hansen et al., 2012). The processed datasets were used for subsequent microarray meta-analysis.

¹ <http://www.ncbi.nlm.nih.gov/geo>

² <https://www.ebi.ac.uk/arrayexpress/>

³ <https://portal.gdc.cancer.gov/>



DEG Identification and Meta-Analysis

Differential gene expression analysis was performed with the R package LIMMA (linear models for microarray data), which is embedded in NA (Ritchie et al., 2015). Each gene expression was calculated based on the false discovery rate (FDR; $p < 0.05$) using the Benjamini–Hochberg method and t -test. In addition, the microarray

meta-analysis between EAC and BE samples was performed using combined effect size (ES). The combined ES is the difference between two group means divided by standard deviation, which is comparable across different studies. It can be calculated by two types of models, namely fixed-effect models (FEM) and random-effect models (REM). In FEM, the calculated effect size in each study

TABLE 1 | Relevant information about selected microarray datasets.

GSE Acc. No.	No. of Samples	Platform	Description	Country	PMID
GSE26886	21 vs 20	AHG-U133 Plus 2.0 Array	Gene expression profiling of Barrett's esophagus, adenocarcinoma, esophageal squamous epithelium, and squamous cell carcinoma	Germany	23514407
GSE1420	8 vs 8	AHG-U133A Array	Barrett's esophagus, Barrett's-associated adenocarcinomas and normal esophageal epithelium	USA	15833844
GSE37200	15 vs 31	AHG-U133A Array	Gene expression profiling of Barrett's esophageal tissues and esophageal adenocarcinoma	USA	26068949
GSE37201	22	AHG-U133A Array	Barrett's esophageal tissues and esophageal adenocarcinoma	USA	20332323

AHG, Affymetrix Human Genome.

is supposed to arise from an original true effect size plus measurement error. In REM, each study further contains a random effect that can incorporate unknown cross-study (different platforms) heterogeneities in the model. The FEM or REM can be chosen based on statistical heterogeneity estimated using Cochran's Q tests (Cochran, 1950). The method typically gives a lower number of DEGs but more confidence (Selvaraj et al., 2018).

Cochran's Q test equation:

$$\text{Cochran's Q test equation: } T = k(k-1) \frac{\sum_{j=1}^k (X_{\cdot j} - N/k)^2}{\sum_{i=1}^b X_i \cdot (k - X_i \cdot)}$$

where k is the number of samples; $X_{\cdot j}$ is the column total for the j^{th} sample; b is the number of genes; $X_i \cdot$ is the row total for the i^{th} gene; N is the grand total.

Gene Ontology and Pathway Enrichment Analysis

We used ClueGO v2.5.3, a Cytoscape⁴ plugin, for function and pathway enrichment analysis of DEGs (Bindea et al., 2009; Kohl et al., 2011). A list of DEGs or hub genes were provided as input into ClueGO with select specific parameters, for example, species, such as *Homo sapiens*, ID type—Entrez gene ID, different enrichment functions—biological process or cellular component or molecular function or KEGG pathways, for the analysis. Each enrichment was calculated based on the Bonferroni method (kappa score 0.96; cutoff value $p < 0.005$).

PPI Network Construction and Module Extraction

The search tool for retrieval of interacting genes/proteins (STRING)⁵ (Szklarczyk et al., 2017) is a database that is used to construct the PPI network. Currently, the database consists of 18,838 human proteins with 25,914,693 core network interactions. In this study, we constructed the PPI network from identified DEGs using the STRING interactome. The highest confidence interaction score was set to 0.9, which reduces the number of false-positive interactions (Bozhilova et al., 2019). Molecular complex detection (MCODE) is a Cytoscape plugin used to identify the finest clusters. MCODE calculates accurate correlation levels as well as identifying essential PPI network modules (Shannon et al., 2003). In addition, other add-ins of Cytoscape, namely, CytoHubba and CytoNCA, were employed to discover the highest linkage hub genes in the network (Chen et al., 2009; Tang et al., 2015).

WGCNA Analysis

The WGCNA package was employed to construct a gene co-expression network using a variant set of genes (12,701 genes). The analysis was performed based on the package instructions (Langfelder and Horvath, 2008). The connection strength between each pair of nodes was calculated using the adjacency matrix a_{ij} .

$$Z_{ij} = [\text{cor}(b_i, b_j)] a_{ij} = Z_{ij} \beta$$

While vectors (b_i and b_j) were expression values for genes, Pearson's correlation coefficient of gene i and j and a_{ij} were represented as the connection strength between genes. The soft-thresholding power of $\beta = 9$ was used to ensure scale-free topology. The hierarchical clustering of the weighting coefficient matrix was used to define the modules. The functional modules in the co-expression network with defined genes, the topological measure (TOM) indicating the concurrence in shared adjacent genes, was calculated as

$$\text{TOM}_{i,j} = \frac{\sum_{K=1}^N A_{i,j} \cdot A_{k,j} + A_{i,j}}{\min(K_i, K_j) + 1 - A_{i,j}}$$

where A is the weighted adjacency matrix described in the above formula. TOM-based dissimilarity measures with a minimum size of 100 for the gene dendrogram and average linkage hierarchical clustering were performed, and similar expression profiles were divided into the same gene modules using the dynamic tree cut package.

Identification of Clinically Significant Modules

Eigengene and gene significance methods were used to identify modules that were correlated with clinical traits of the GSE37200 microarray data set. The first principal component of each gene module and the expression of the module eigengene were defined as representative of the whole gene set and were described in the first eigengene module. The association between module eigengenes and clinical trait was used to calculate and identify the significant clinical module. Second, the gene significance was described as a mediated p -value of each gene in the linear regression between expression and clinical traits. Furthermore, the module significance was described as the average the gene significance of all genes associated with the module. The average absolute gene significance was defined as module significance. It was calculated to incorporate clinical traits into a co-expression network (Langfelder and Horvath, 2008).

Survival Analysis and Validation of SHGS

The SHGS were identified from the modules of WGCNA and the PPI network using an interactive Venn diagram. The R package *survival* was employed to calculate Kaplan–Meier (KM) survival plots with hazard ratio (HR) and log-rank tests of hubs, which was implemented in the *OSeac*⁶ (consensus survival analysis for EAC) web interface. *OSeac* retrieved the gene expression profiles and clinical data including TNM (Stage I, II, III, and IV), gender (male and female), race (White, Black, and African American), and grade (G1, G2, G3, and GX) of 198 patients from TCGA and GEO. We analyzed the overall survival rate of the shared gene signature as an input and obtained the plot from the tool (Wang et al., 2020). The Human Protein Atlas⁷ was used to validate the immunohistochemistry of SHGS (Uhlén et al., 2005; Uhlen et al., 2017).

⁴<https://cytoscape.org/>

⁵<https://string-db.org/>

⁶<http://bioinfo.henu.edu.cn/EAC/EACList.jsp>

⁷<https://www.proteinatlas.org/>

RESULTS

Physiognomies of Selected Studies

We collected a total of 682 studies from the GEO and ArrayExpress database up to October 2019. In all, 678 datasets/studies that did not satisfy the inclusion criteria were excluded. Finally, four potential studies were selected (**Supplementary Figure 1**). Among the four selected studies, three were conducted on the Affymetrix human genome U133A platform and one was performed on the Affymetrix human genome U133 plus 2.0 platform, which included 125 samples in total chosen in this study. In each study, EAC samples were compared with the adjacent BE samples. The dataset GSE37200 was used to construct a co-expression network with the relevant clinical trait information. After preprocessing and normalization, the GSE37200 dataset with 22,284 genes was further processed, and variant genes (12,701) were selected for WGCNA studies.

Identification of DEGs and Enrichment Analysis

In total, 403 DEGs were obtained through microarray meta-analysis, which include 169 downregulated and 234 upregulated genes. A heatmap is a simple yet effective way to compare the content of multiple major gene lists. Major DEGs across all the datasets were represented in red, orange, and yellow in a heatmap. Gray indicates that the respective gene is not present in the gene list (**Supplementary Figure 2**). **Table 2** illustrates the top 10 upregulated and downregulated DEGs. Monocyte differentiation antigen CD14 (CD14), ribose 5-phosphate isomerase A (RPIA), tumor necrosis factor superfamily member 11 (TNFSF11), plexin D1 (PLXND1), major histocompatibility complex, class II DM beta (HLA-DMB), and

spliceosome-associated factor 3, and U4/U6 recycling protein (SART3) were highly expressed upregulated genes, whereas fucosyltransferase 2 (FUT2), SECIS binding protein 2 like (SECISBP2L), COP9 signalosome subunit 4 (COPS4), gelsolin (GSN), and glutathione peroxidase 3 (GPX3) were highly expressed downregulated genes. According to the gene ontology (GO) terms BP, MF, and CC, downregulated genes were significantly enriched in the mitotic cell cycle process, sister chromatid segregation, antigen processing, presentation of peptide antigen *via* MHC class I, chromosomal region, and MHC class I protein binding, whereas retinol dehydrogenase activity and fucosyltransferase activity were highly enriched in upregulated genes associated with EAC (**Figures 2A–C**). In KEGG, pathway enrichment demonstrated that the upregulated genes were enriched for viral myocarditis, cell cycle, DNA replication, and AGE-RAGE signaling pathways in diabetic complications. Downregulated genes were associated with pathways involved in fatty acid degradation, glycosphingolipid biosynthesis, and amino sugar and nucleotide sugar metabolism (**Figure 2D**).

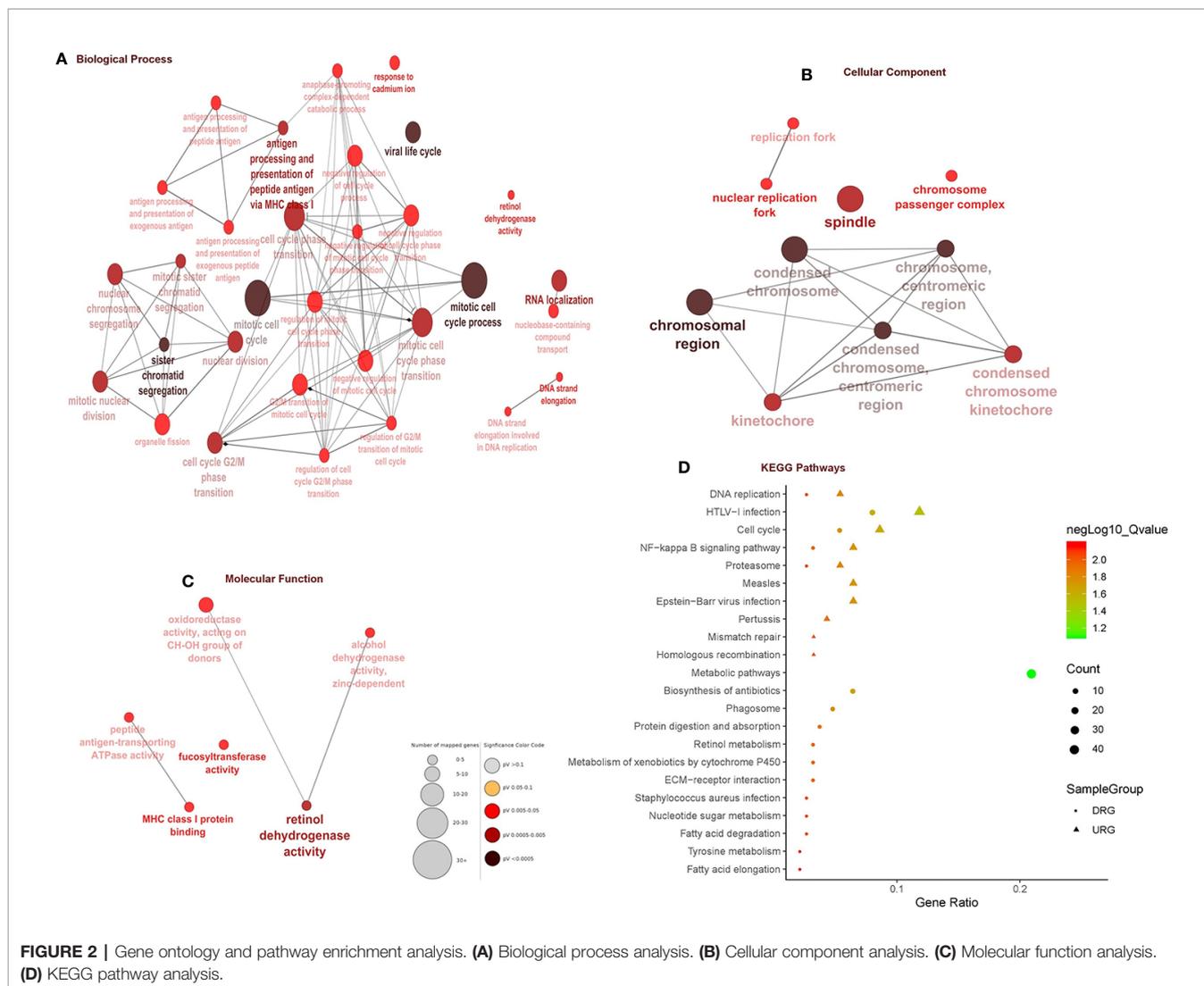
WGCNA and Clinically Significant Module Identification

A dendrogram of samples (GSE37200) with clinical trait was clustered using the average linkage method and Pearson's correlation method (**Figure 3A**). Co-expression analysis was carried out to construct the co-expression network. In this study, the power of $\beta = 9$ (scale-free $R^2 = 0.95$) was selected as the soft-thresholding parameter to ensure a scale-free network (**Figure 3B**). A dendrogram of all differentially expressed genes was clustered based on a dissimilarity measure (1-TOM) (**Supplementary Figure 3**). A total of 39 modules were identified through hierarchical

TABLE 2 | Top ten up- and downregulated genes.

S.No.	Gene	Gene name	Combined ES	P-value
Upregulated genes				
1	CD14	Monocyte differentiation antigen CD14	1.2948	4.98E-08
2	RPIA	Ribose 5-phosphate isomerase A	1.1376	4.63E-08
3	TNFSF11	Tumor necrosis factor super family member 11	1.1229	3.66E-08
4	PLXND1	Plexin D1	1.1223	3.60E-08
5	HLA-DMB	Major histo-compatibility complex, class II, DM beta	1.1168	3.97E-08
6	SART3	Spliceosome associated factor 3, U4/U6 recycling protein	1.1142	4.06E-08
7	OSBPL3	Oxysterol binding protein like 3	1.1119	4.79E-08
8	PRAF2	PRA1 domain family member 2	1.1117	4.61E-08
9	PILRB	Paired immunoglobulin like type 2 receptor beta	1.1078	4.67E-08
10	RGS16	Regulator of G protein signaling 16	1.1064	4.79E-08
Downregulated genes				
1	FUT2	Fucosyltransferase 2	-1.1095	4.67E-08
2	SECISBP2L	SECIS binding protein 2 like	-1.111	4.52E-08
3	COPS4	COP9 signalosome subunit 4	-1.1148	4.79E-08
4	GSN	Gelsolin	-1.117	4.04E-08
5	GPX3	Glutathione peroxidase 3	-1.1185	3.87E-08
6	ADH1A	Alcohol dehydrogenase 1A (class I), alpha polypeptide	-1.1271	4.39E-08
7	CORO2A	Coronin 2A	-1.2074	4.67E-08
8	ACADS	Acyl-CoA dehydrogenase short chain	-1.2157	3.57E-08
9	RCAN2	Regulator of calcineurin 2	-1.2172	5.07E-08
10	CLEC3B	C-type lectin domain family 3 member B	-1.2316	3.73E-08

*Combined ES, Cochran's combination test of random effect model (REM) or effect size (ES).



clustering. Light green (eigengene value = 0.41), dark gray (eigengene value = 0.62) and Sienna3 (eigengene value = 0.46) modules appeared to have the highest association with age, gender, and ethnicity. There was no module–trait relationship associated with tumor stage, denoted as NA in **Figure 3C**. Therefore, the dark gray module having the highest association with gender was selected as the clinically significant module for further analysis. There were 207 phenotypic genes identified in the dark-gray module (**Figure 3D**). In **Supplementary Figure 4**, the hierarchical clustering dendrogram of the eigengene network represents the relationships among the modules and the clinical trait weight.

Identification and Validation of Hub Genes

The PPI network was constructed with 403 DEGs using the STRING database. The interactive relationships between the key genes in the whole network were determined using the Cytoscape plugins (MCODE, Cytoscape, and CytoHubba). There are two clusters: 82 nodes and 938 edges in cluster 1, and 20 nodes and 168 edges in cluster 2, which were identified from MCODE based

on a scoring system (cutoff k-score = 12). In addition, the data were imported into another plugin, CytoHubba, which helped to identify 104 key genes through five different calculation methods, namely, EPC, MCC, DMNC, MNC, and Stress. Then, the two clusters were imported into the CytoNCA plugin, which helped to identify 40 key genes using five different algorithms, namely, betweenness, closeness, degree, eigenvector, and subgraph. We securely conceive that the key genes are the intersections between the PPI network and the dark-gray module with 207 genes (**Supplementary Table 1**) highly correlated with phenotype (gender) from the WGCNA analysis (**Figures 4A, B**). Finally, five SHGS, namely, pre-mRNA processing factor 4 (PRPF4), serine and arginine rich splicing factor 1 (SRSF1), heterogeneous nuclear ribonucleoprotein M (HNRNPM), DEXH-box helicase 9 (DHX9), and origin recognition complex subunit 2 (ORC2), are identified between BE and EAC. Pathway enrichment demonstrated that all the SHGS are involved in the metabolism of RNA, and its molecular functional terms include cell cycle, DNA binding, DNA topoisomerase binding, pre-mRNA splicing, and RNA helicase activity (**Figure 5**).

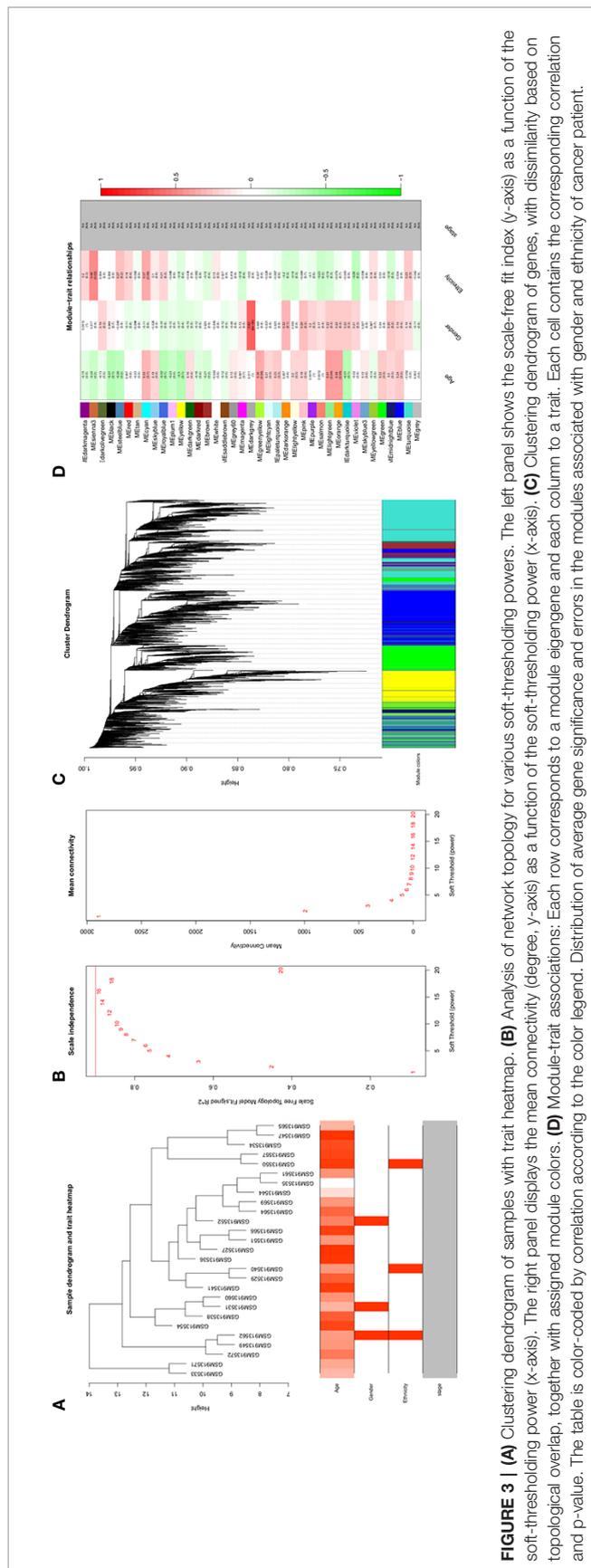


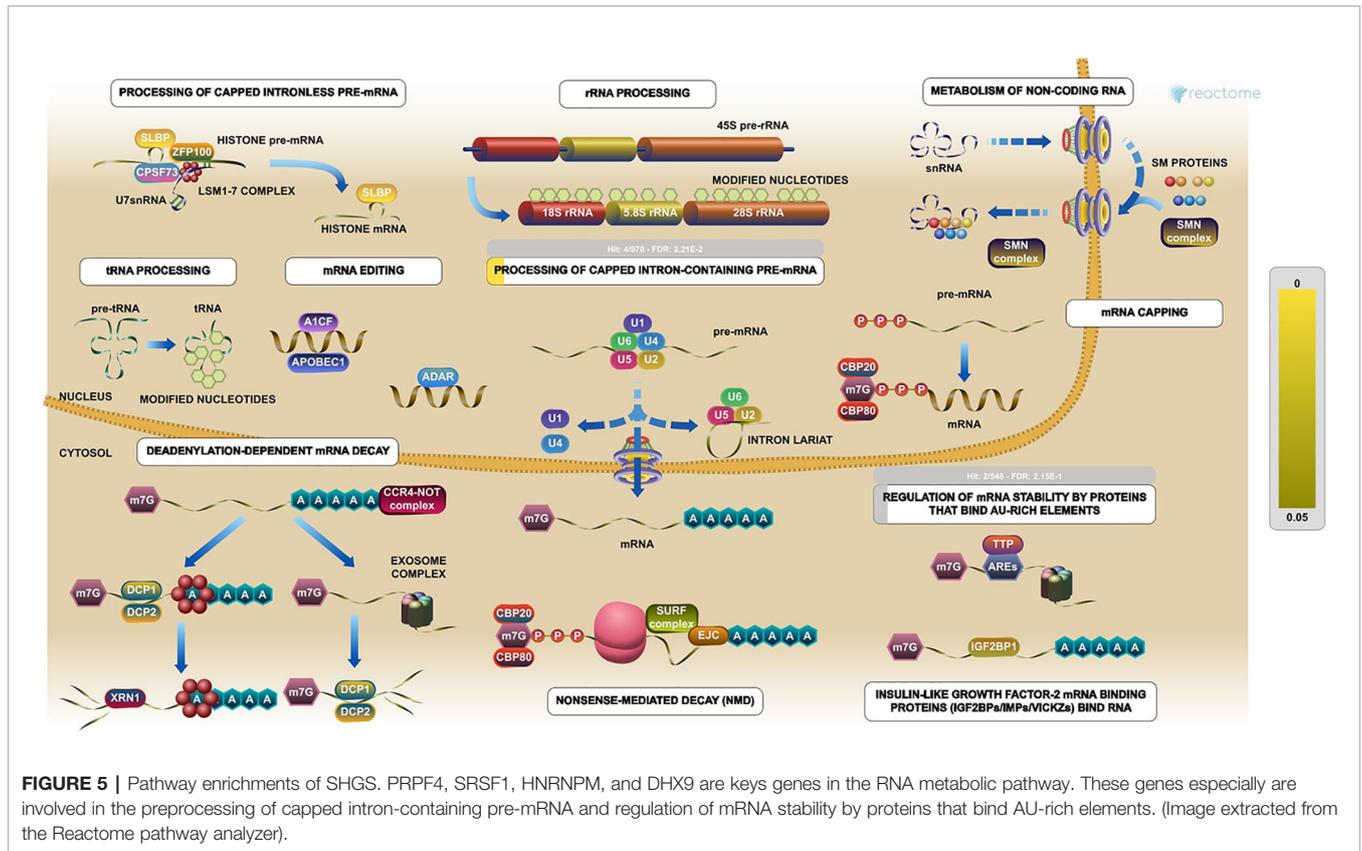
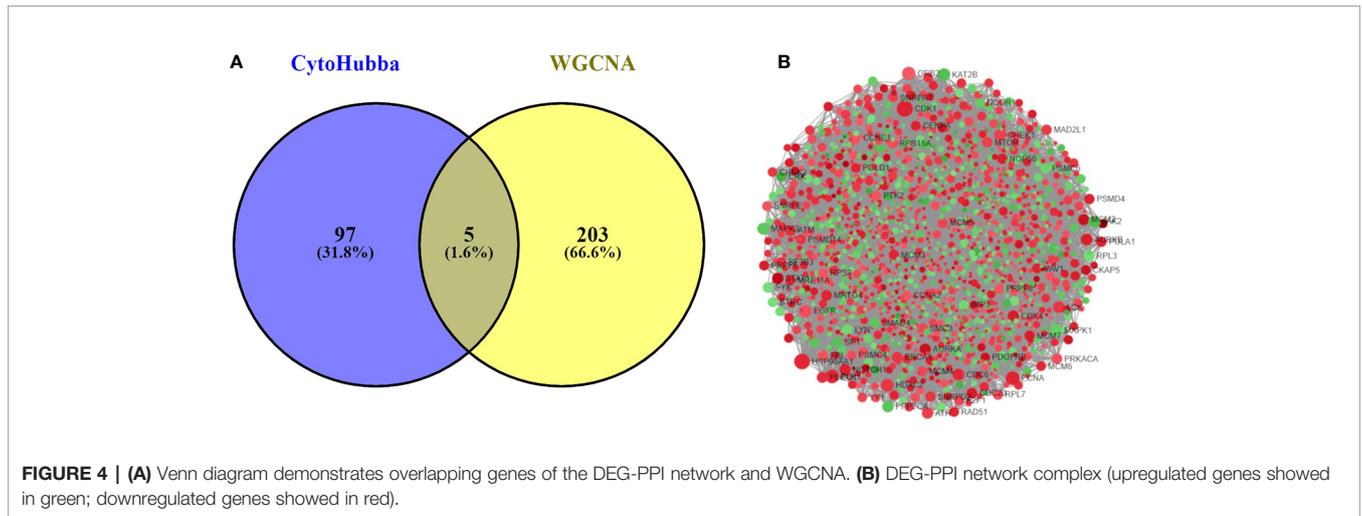
FIGURE 3 | (A) Clustering dendrogram of samples with trait heatmap. (B) Analysis of network topology for various soft-thresholding powers. The left panel shows the scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis). (C) Clustering dendrogram of genes, with dissimilarity based on topological overlap, together with assigned module colors. (D) Module-trait associations: Each row corresponds to a module eigengene and each column to a trait. Each cell contains the corresponding correlation and p-value. The table is color-coded by correlation according to the color legend. Distribution of average gene significance and errors in the modules associated with gender and ethnicity of cancer patient.

Survival Analysis and Immunohistochemistry

Kaplan–Meier plots demonstrated that the prognostic impact of the SHGS was identified from modules of the PPI network complex and WGCNA. The results revealed that high expression of HNRNPM and SRSF1 was associated with poor overall survival of BE and EAC patients ($p < 0.05$). Moreover, high expression of PRPF4, DHX9, and ORC2 was correlated with longer overall survival of BE and EAC patients (Figure 6). In addition, we plotted a gender-based survival curve to determine the correlation of WGCNA modules. The hazard ratio (HR) and 95% confidence interval were as follows in males: PRPF4 (HR =1.08; 95%CI - 0.46 ± 2.48; $p = 0.865$); SRSF1 (HR =3.08; 95%CI - 1.49 ± 6.37; $p = 0.002$); HNRNPM (HR =3.295; 95%CI - 1.54 ± 7.02; $p = 0.002$); DHX9 (HR =1.39; 95%CI - 0.64 ± 2.48; $p = 0.404$); ORC2 (HR =1.25; 95%CI - 0.58 ± 2.72; $p = 0.564$). Further, in female cases PRPF4 (HR =0.39; 95%CI - 0.03 ± 3.89; $p = 0.421$); SRSF1 (HR =1.49; 95%CI - 0.20 ± 10.79; $p = 0.689$); HNRNPM (HR =8.06; 95%CI - 0.82 ± 79.01; $p = 0.073$); DHX9 (HR =0.38; 95%CI - 0.04 ± 3.89; $p = 0.424$); ORC2 (HR =3.24; 95%CI - 0.20 ± 51.91; $p = 0.4061$). The results clearly demonstrated that the high expression of SHGS correlated to the poor prognosis of male compared to female. Furthermore, immunohistochemical slides of the Human Protein Atlas database indicated that the protein expressions of SHGS were drastically higher in cancerous tissues compared with in adjacent normal tissues, as shown in Figure 7. Therefore, these SHGS were all key genes that play an initiative role and might have a tendency to co-express.

DISCUSSION

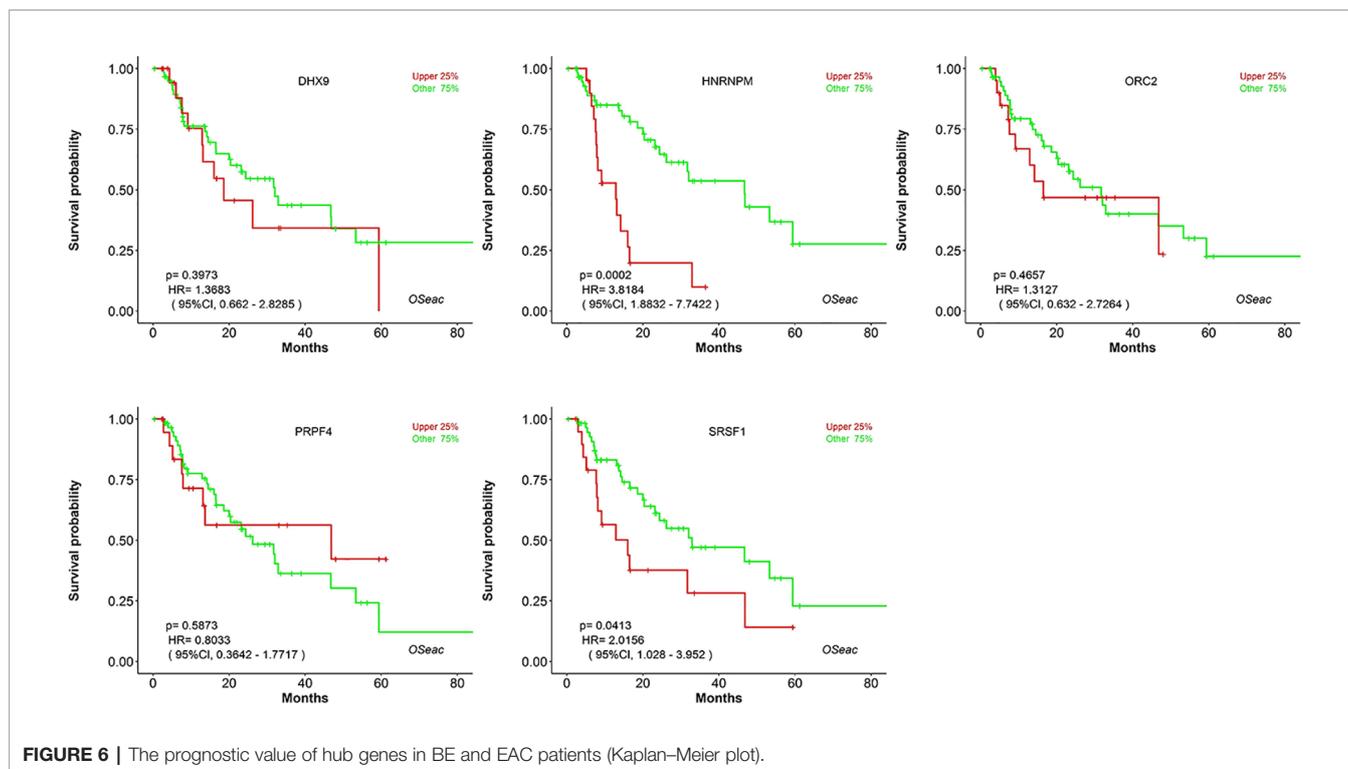
EAC is an obstinate type of cancer, which has a high mortality rate because of poor prognosis, metastatic rate, and treatment resistance (Tatarian and Palazzo, 2019). EAC usually arises from a premalignant variation in the lining of the esophagus known as BE (Thrift, 2016). Unfortunately, the treatment and diagnosis of EAC and BE are limited due to the lack of precise molecular targets. Therefore, we designed this study to explore SHGS between EAC and BE to improve the diagnosis and prognosis status of the patients. There are numerous advanced technologies that can quantify the enormous amount of transcripts in a parallel manner. Microarray and data mining are well-known approaches for cancer biomarker discovery (Selvaraj et al., 2019). Nevertheless, a single microarray dataset is not enough to deal with this obstinate disease. However, a comprehensive analysis of a number of microarray datasets with different platforms can assist with identifying the molecular mechanism of EAC and BE. Therefore, we selected four different microarray datasets to identify SHGS and the associated pathways between BE and EAC. Moreover, WGCNA is a powerful tool for searching effective biological mechanisms and key genes from gene expression microarrays. It provides module construction and correlation analysis within the gene expression data to determine the associations between genes. It also elucidates the biological significance of a gene module to provide insights into molecular and pathological characteristics in many diseases. All these characteristics make it a robust, reliable, and significant



method for analysis of large-scale data. There is no prior research employing WGCNA to do gene co-expression network analysis with BE and EAC. To explore SHGS, we decided to construct a gene co-expression network with relevant clinical trait information from the GSE37200 dataset.

Phenotype variants like age, gender, and ethnicity are factors that are intensively involved in the prognosis and diagnosis of BE and EAC (Ford et al., 2005; Runge et al., 2015). EAC usually appears at

the later stage of life, but it may start at a young age in the form of BE. Earlier studies have reported that male patients with BE are at low risk of malignant progression and predominantly die due to causes other than EAC (Sikkema et al., 2010). There are studies reported that there is a marked male prevalence of EAC with a male-to-female ratio of 9:1 due to sex hormone factors. Androgen exposure may increase the risk of EAC compared to estrogen (Xie and Lagergren, 2016; Kim et al., 2016). Furthermore, geographically



White people, especially White Americans, are at higher risk than other ethnicities (Schneider and Corley, 2015). A comprehensive study from January 2006 to December 2017 reported that high risk of male patients with esophageal diseases in the province of Madinah in Saudi Arabia is due to a variety of factors, including inflammatory disorders, infection, and neoplastic condition (Albasri et al., 2019). In addition, genomic analysis by restriction fragment length polymorphism indicated that the highest frequencies of Y-chromosomal haplogroups are associated with BE and EAC in White males (Westra et al., 2020). Recent case–control studies demonstrate that gastroesophageal reflux disease in male patients is highly associated with the development of BE in Germany (Schmidt et al., 2020). These reports supported the present results, indicating that predicted dark gray modules with the highest association with gender must have a clinically significant module.

Two types of biological materials, namely, GO and KEGG pathway data, are key to understanding the disease mechanism. CD14 acts as a co-receptor with toll-like receptors (TLRs) to identify evading pathogens and improve the immune system. It is reported that TLRs 1-10 are expressed in the normal esophagus and that there is a high association of TLRs 4, 5, and 9 with BE and EAC (Kaupilla and Selander, 2014). TNFSF11 is a key regulator of interactions between T cells and dendritic cells, which regulate the T-cell-dependent immune response and enhance bone-resorption in hypercalcemia of malignancy (Luan et al., 2012). Somja et al. (2013) observed that both metaplastic and malignant lesions of the esophagus are infiltrated by regulatory T cells. They concluded that soluble

factors secreted by epithelial cells during the EAC or BE influence tumor progression through tolerogenic dendritic cells, which can be a potential therapeutic tool. In addition, different cohort studies have reported that GSN is a serum glycoprotein biomarker used as a diagnostic tool for EAC and BE (Shah et al., 2015; Shah et al., 2018). Glycosphingolipid biosynthesis is an important pathway that can produce cell-surface glycans. These glycans are altered in the development from BE into EAC, with specific changes in lectin binding patterns. This binding is a key marker in endoscopic visualization of high-grade dysplastic lesions (Bird-Lieberman et al., 2012). These reports suggest that the predicted GO terms and pathways of DEGs are highly associated with EAC and BE.

We have identified five different SHGS (PRPF4, SRSF1, HNRNPM, DHX9, and ORC2) between EAC and BE. PRPF4, SRSF1, and HNRNPM are U4/U6 small nuclear ribonucleoprotein Prp4, serine and arginine-rich splicing factor 1, and heterogeneous nuclear ribonucleoprotein M coding genes, respectively. These genes play an important role in pre-mRNA splicing and spliceosome assembly (Bertram et al., 2017). Pre-mRNA splicing is key to the pathology and has a substantial role in generating multiple oncogenic and tumor-suppressor proteins after the post-transcriptional process. Splicing is of different types such as amino acid addition, exon skipping, frame shift, intron retention, promoter usage, truncated C-terminus, and 5'-SS, which have various clinical applications including proliferation, metastasis, drug resistance, and radiotherapy (Guo et al., 2015; Di et al., 2019). In addition, there are studies reporting splicing signatures associated with the prognosis of

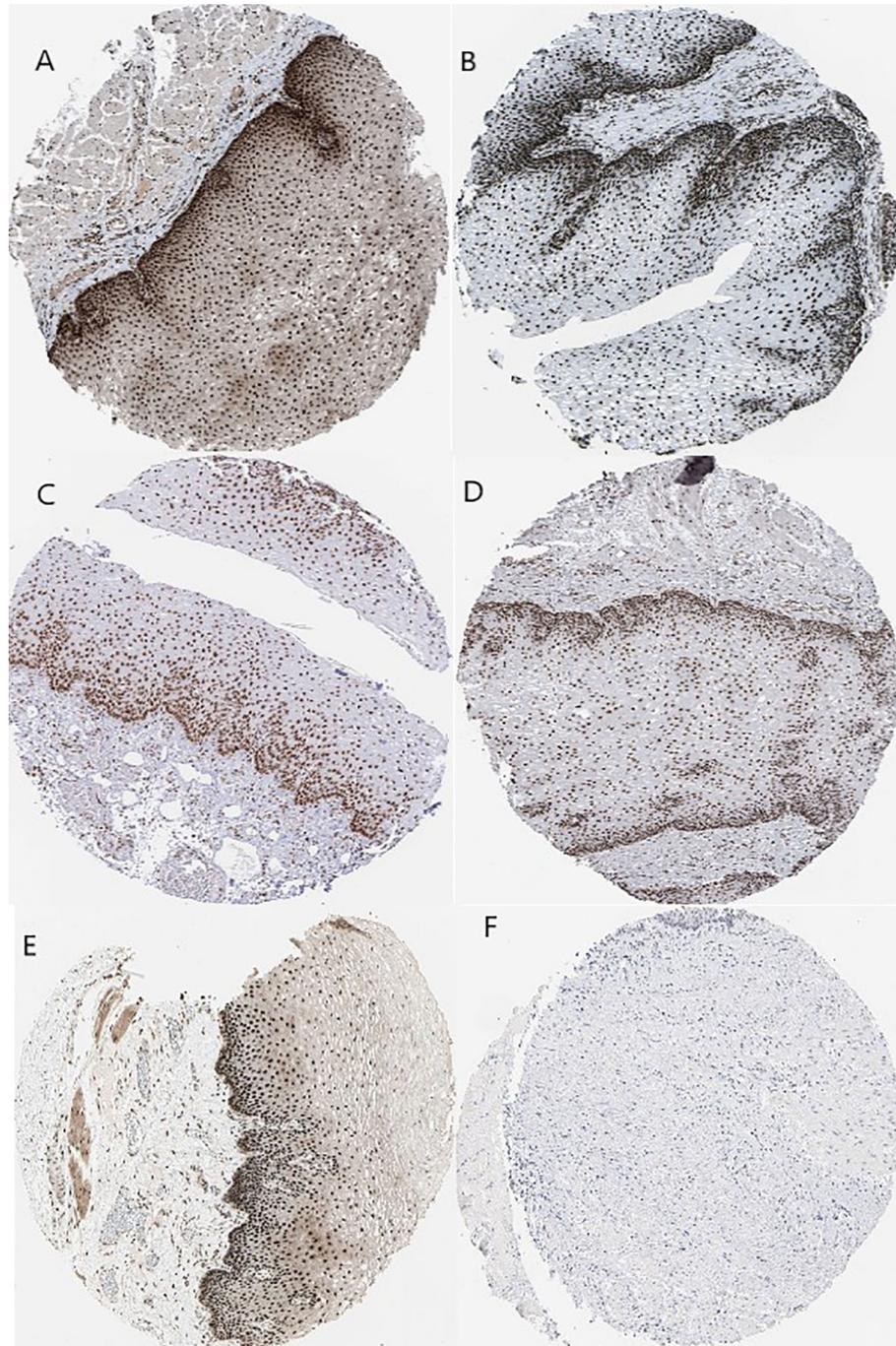


FIGURE 7 | Immunohistochemistry of the five hub genes based on the Human Protein Atlas. **(A)** Protein levels of PRPF4 in normal tissue (staining: high; intensity: strong; quantity: >75%). **(B)** Protein levels of SRSF1 in normal tissue (staining: high; intensity: strong; quantity: >75%). **(C)** Protein levels of SRSF1 in tumor tissue (staining: high; intensity: strong; quantity: >75%). **(D)** Protein levels of HNRNPM in normal tissue (staining: high; intensity: strong; quantity: >75%). **(E)** Protein levels of DHX9 in normal tissue (staining: high; intensity: strong; quantity: >75%). **(F)** Protein levels of ORC2 in normal tissue (staining: not detected; intensity: low; quantity: <25%).

esophageal cancer (Lin et al., 2018; Mao et al., 2019). Through the splicing mechanism, PRPF4, SRSF1, and HNRNPM regulate the cell proliferation, migration, and invasion in different cancers, including lung cancer (Choi, 2012; Chang and Lin, 2019), breast cancer

(Anczuków et al., 2015; Sun et al., 2017; Park et al., 2019), cutaneous squamous cell carcinoma (Zhang et al., 2018), hepatocellular carcinoma (Tu et al., 2019), esophagus dysplasia (Varghese et al., 2015; Fitzgerald et al., 2018), gastric cancer (Wu

et al., 2019), cervical cancer (Dong et al., 2019), and Ewing's sarcoma (Passacantilli et al., 2017).

DHX9 is an ATP-dependent RNA helicase A coding gene involved in DNA replication, transcriptional activation, post-transcriptional RNA regulation, mRNA translation, and RNA-mediated gene silencing (Capitanio et al., 2017). Knockdown of ATP-dependent RNA helicase inhibited the expression of β -catenin, c-Myc, and cyclin D1 in esophageal cancer cells through suppressing the Wnt/ β -catenin signaling pathway (Ma et al., 2017). In addition, ATP-dependent RNA helicase was reported to dysregulate distinct steps of mRNA and pre-ribosomal RNA metabolism in cancer cells (Awasthi et al., 2018). ORC2 is an origin recognition complex subunit 2 coding gene binding origins of replication (Shen et al., 2012). It can bind to different histone trimethylation proteins and stabilize leucine-rich repeat and WD repeat-containing protein 1 (LRWD1) through protecting it from ubiquitin-mediated proteasomal degradation (Chan and Zhang, 2012). Studies demonstrated that increased expressions of certain histone-mediated proteins correlate with advanced TNM stages, tumor grade, metastatic potential, and decreased overall and disease-free survival of patients with esophageal cancer (Schizas et al., 2018). This supportive information enhances the understanding of why the predicted DHX9, HNRNPM, ORC2, PRPF4, and SRSF1 genes are highly correlated to EAC and BE progression and act as potential biomarkers for diagnosis as well as prognosis.

CONCLUSION

This network pharmacology-based study provides new insights into BE and EAC patients for their diagnosis and prognosis. The results of microarray dataset-based PPI networks and WGCNA exhibited that the dark-gray module had the maximum association with EAC and BE, with the identification of five SHGS, namely PRPF4, SRSF1, HNRNPM, DHX9, and ORC2. The WGCNA-based gene co-expression network indicated that the relationships between co-expressed genes and clinical trait (gender of the patient) were associated with the progression of esophageal cancer. SHGS enrichment denotes that the RNA metabolic and spliceosome pathways play an essential role in the development and progress of esophageal cancer. Survival analysis demonstrates that the high expression of HNRNPM and SRSF1 in esophageal cancer might be a poor prognostic marker. The co-expression modules were established to preserve a reliable expression relationship independent of phenotype and may share similar biological functions. This approach shares the limitations of other data mining methods: the results of WGCNA can technically be biased due to tissue contamination or artifacts. To enhance the reliability of the WGCNA results, immunohistochemical data from the Human Protein Atlas were used for confirmation. However, we could not obtain all the related IHC data of tumor and adjacent normal samples for each gene due to the database constraint. These findings may support new therapeutic targets and potential useful for the advancement of prognostic biomarker evaluation.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Publicly available datasets were analyzed in this study. This data can be found at: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26886> <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1420> <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37200> <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37201> https://portal.gdc.cancer.gov/exploration?filters=%7B%22op%22%3A%22and%22%2C%22content%22%3A%5B%7B%22op%22%3A%22in%22%2C%22content%22%3A%7B%22field%22%3A%22cases.primary_site%22%2C%22value%22%3A%5B%22esophagus%22%5D%7D%7D%5D%7D <https://www.proteinatlas.org/ENSG00000136875-PRPF4/tissue/esophagus>.

<https://www.proteinatlas.org/ENSG00000136450-SRSF1/tissue/esophagus> <https://www.proteinatlas.org/ENSG00000099783-HNRNPM/tissue/esophagus> <https://www.proteinatlas.org/ENSG00000135829-DHX9/tissue/esophagus>

<https://www.proteinatlas.org/ENSG00000115942-ORC2/tissue/esophagus>.

AUTHOR CONTRIBUTIONS

DW, GS, and AN conceived and designed the experiment. AN, AK, and SK performed the meta-analysis and network analysis. AN and GS performed the WGCNA. AN, AK, and GS aided in preparation of the figures and supplemental materials. AN and SK performed Cytoscape analysis. AN, SK and GS compiled the manuscript. WC and DW provided expert advice in the study, performed editing and proofreading of the manuscript. All authors contributed to the study and approved the submitted version.

FUNDING

The authors are thankful for financial support from the Ministry of Science and Technology of China (Grant No.: 2016YFA0501703), the National Natural Science Foundation of China (Contract No. 61832019, 61503244), the Science and Technology Commission of Shanghai Municipality (Grant: 19430750600), Henan Natural Science (Grant No.:162300410060), as well as SJTU JiRLMDS Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14) to DQ.W; Doctoral Study Scholarship from Shanghai Jiao Tong University to ASN; China Postdoctoral Science Foundation (Grant No.: 2018M632766) to GS; Henan Province Postdoctoral Science Foundation (Grant No.: 001802029 & 001803035) to SK and GS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2020.00881/full#supplementary-material>

REFERENCES

- Albasri, A. M., Ansari, I. A., Anjum, S., Elsayaf, Z. M., and Alhujaily, A. S. (2019). Pattern of oesophageal diseases in Madinah region, Saudi Arabia: An 11 years experience. *J. Pak Med. Assoc.* 69 (9), 1365–1368.
- Analysis, Working Group, University Asan and Cancer Genome Atlas Research Network (2017). Integrated genomic characterization of oesophageal carcinoma. *Nature* 541 (7636), 169. doi: 10.1038/nature20805
- Anaparthi, R., and Sharma, P. (2014). Progression of Barrett oesophagus: role of endoscopic and histological predictors. *Nat. Rev. Gastroenterol. Hepatol.* 11 (9), 525. doi: 10.1038/nrgastro.2014.69
- Anczuków, O., Akerman, M., Cléry, A., Wu, J., Chen, S., Shirole, N. H., et al (2015). SRSF1-regulated alternative splicing in breast cancer. *Mol. Cell* 60 (1), 105–117. doi: 10.1016/j.molcel.2015.09.005
- Awasthi, S., Verma, M., Mahesh, A., Khan, M. I. K., Gayathri, G., Rajavelu, A., et al. (2018). DDX49 is an RNA helicase that affects translation by regulating mRNA export and the levels of pre-ribosomal RNA. *Nucleic Acids Res.* 46 (12), 6304–6317. doi: 10.1093/nar/gky231
- Bertram, K., Agafonov, D. E., Dybkov, O., Haselbach, D., Leelaram, M. N., Will, C. L., et al. (2017). Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. *Cell* 170 (4), 701–713. doi: 10.1016/j.cell.2017.07.011
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25 (8), 1091–1093. doi: 10.1093/bioinformatics/btp101
- Bird-Lieberman, E. L., Neves, AndréA., Lao-Sirieix, P., O'donovan, M., Novelli, M., Lovat, L. B., et al. (2012). Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus. *Nat. Med.* 182, 315. doi: 10.1038/nm.2616
- Bozhilova, L. V., Whitmore, A. V., Wray, J., Reinert, G., and Deane, C. M. (2019). Measuring rank robustness in scored protein interaction networks. *BMC Bioinf.* 20 (1), 446. doi: 10.1186/s12859-019-3036-6
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer J. Clin.* 68 (6), 394–424. doi: 10.3322/caac.21492
- Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., et al. (2003). ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 31 (1), 68–71. doi: 10.1093/nar/gkg091
- Campaign, A., and Yang, Y. H. (2010). Comparison study of microarray meta-analysis methods. *BMC Bioinf.* 111, 408. doi: 10.1186/1471-2105-11-408
- Capitanio, J. S., Montpetit, B., and Wozniak, R. W. (2017). Human Nup98 regulates the localization and activity of DEXH/D-box helicase DHX9. *Elife* 6, e18825. doi: 10.7554/eLife.18825
- Chai, T., Shen, Z., Zhang, P., Lin, Y., Chen, S., Zhang, Z., et al. (2019). Comparison of high risk factors (hot food, hot beverage, alcohol, tobacco, and diet) of esophageal cancer: A protocol for a systematic review and meta-analysis. *Medicine* 98 (17), e15176. doi: 10.1097/MD.00000000000015176
- Chan, K. M., and Zhang, Z. (2012). Leucine-rich repeat and WD repeat-containing protein 1 is recruited to pericentric heterochromatin by trimethylated lysine 9 of histone H3 and maintains heterochromatin silencing. *J. Biol. Chem.* 287 (18), 15024–15033. doi: 10.1074/jbc.M111.337980
- Chang, H.-L., and Lin, J.-C. (2019). SRSF1 and RBM4 differentially modulate the oncogenic effect of HIF-1 α in lung cancer cells through alternative splicing mechanism. *BiochimicaetBiophysicaActa (BBA)-Mol. Cell Res.* 1866 (12), 118550. doi: 10.1016/j.bbamcr.2019.118550
- Chen, S.-H., Chin, C.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., and Lin, C.-Y. (2009). “cyto-Hubba: A Cytoscape plug-in for hub object analysis in network biology”, in *20th International Conference on Genome Informatics*. (Pacifico Yokohama, Japan: Imperial College Press).
- Choi, S.-Y. (2012). Identification of PRPF4 as a novel cancer promoter through AKT signaling in lung cancer. *AACR; Cancer Res.* 72 (8 Suppl), 4174–4174. doi: 10.1158/1538-7445.AM2012-4174
- Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika* 37/3/4, 256–266. doi: 10.1093/biomet/37.3-4.256
- Cook, M. B., Corley, D. A., Murray, L. J., Liao, L. M., Weimin Ye, F., Gammon, M. D., et al. (2014). Gastroesophageal reflux in relation to adenocarcinomas of the esophagus: a pooled analysis from the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON). *PLoS One* 9 (7), e103508. doi: 10.1371/journal.pone.0103508
- Davies, A. R., Gossage, J. A., Zylstra, J., Mattsson, F., Lagergren, J., Maisey, N., et al. (2014). Tumor stage after neoadjuvant chemotherapy determines survival after surgery for adenocarcinoma of the esophagus and esophagogastric junction. *J. Clin. Oncol.* 32 (27), 2983–2990. doi: 10.1200/JCO.2014.55.9070
- Di, C., Zhang, Q., Chen, Y., Wang, Y., Zhang, X., Liu, Y., et al. (2019). Function, clinical application, and strategies of Pre-mRNA splicing in cancer. *Cell Death Differ.* 267, 1181–1194. doi: 10.1038/s41418-018-0231-3
- Dong, M., Dong, Z., Zhu, X., Zhang, Y., and Song, L. (2019). Long non-coding RNA MIR205HG regulates KRT17 and tumor processes in cervical cancer via interaction with SRSF1. *Exp. Mol. Pathol.* 111, 104322. doi: 10.1016/j.yexmp.2019.104322
- Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30 (1), 207–210. doi: 10.1093/nar/30.1.207
- Fitzgerald, R. C., Varghese, S., Newton, R., and Wernisch, L. “Methods and means for dysplasia analysis.” U.S. Patent Application 15/747,117, filed July 26, 2018.
- Ford, A. C., Forman, D., Reynolds, P.D., Cooper, B. T., and Moayyedi, P. (2005). Ethnicity, gender, and socioeconomic status as risk factors for esophagitis and Barrett's esophagus. *Am. J. Epidemiol.* 162 (5), 454–460. doi: 10.1093/aje/kwi218
- Guo, W., Wang, C., Guo, Y., Shen, S., Guo, X., Kuang, G., et al. (2015). RASSF5A, a candidate tumor suppressor, is epigenetically inactivated in esophageal squamous cell carcinoma. *Clin. Exp. Metastasis* 32 (1), 83–98. doi: 10.1007/s10585-015-9693-6
- Hansen, K. D., Irizarry, R. A., and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* 13 (2), 204–216. doi: 10.1093/biostatistics/kxr054
- Kaupilla, J. H., and Selander, K. S. (2014). Toll-like receptors in esophageal cancer. *Front. Immunol.* 5, 200. doi: 10.3389/fimmu.2014.00200
- Kim, Y. S., Kim, N., and Gwang Ha, K. (2016). Sex and gender differences in gastroesophageal reflux disease. *J. Neurogastroenterol. Motil.* 22 (4), 575. doi: 10.5056/jnm16138
- Kimchi, E. T., Posner, M. C., Park, J. O., Darga, T. E., Kocherginsky, M., Karrison, T., et al. (2005). Progression of Barrett's metaplasia to adenocarcinoma is associated with the suppression of the transcriptional programs of epidermal differentiation. *Cancer Res.* 65 (8), 3146–3154. doi: 10.1158/0008-5472.CAN-04-2490
- Kohl, M., Wiese, S., and Warscheid, B. (2011). “Cytoscape: software for visualization and analysis of biological networks.” in *Data mining in proteomics* (Humana Press), 291–303.
- Konishi, S. (1985). Normalizing and variance stabilizing transformations for intraclass correlations. *Ann. Inst. Stat. Math.* 37 (1), 87–94. doi: 10.1007/BF02481082
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9 (1), 559. doi: 10.1186/1471-2105-9-559
- Lin, J., Myers, A. L., Wang, Z., Nancarrow, D. J., Ferrer-Torres, D., Handlogten, A., et al. (2015). Osteopontin (OPN/SPP1) isoforms collectively enhance tumor cell invasion and dissemination in esophageal adenocarcinoma. *Oncotarget* 6 (26), 22239. doi: 10.18632/oncotarget.4161
- Lin, P., He, R. Q., Ma, F. C., Liang, L., He, Y., Yang, H., et al. (2018). Systematic analysis of survival-associated alternative splicing signatures in gastrointestinal pan-adenocarcinomas. *EBioMedicine* 34, 46–60. doi: 10.1016/j.ebiom.2018.07.040
- Lordick, F., Mariette, C., Haustermans, K., Obermannová, R., and Arnold, D. (2016). Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 27, suppl_5, v50–v57. doi: 10.1093/annonc/mdw329
- Luan, X., Lu, Q., Jiang, Y., Zhang, S., Wang, Q., Yuan, H., et al. (2012). Crystal structure of human RANKL complexed with its decoy receptor osteoprotegerin. *J. Immunol.* 189 (1), 245–252. doi: 10.4049/jimmunol.1103387
- Ma, Z., Feng, J., Guo, Y., Kong, R., Ma, Y., Sun, L., et al. (2017). Knockdown of DDX5 inhibits the proliferation and tumorigenesis in esophageal cancer. *Oncol. Res. Featuring Preclinical Clin. Cancer Ther.* 25 (6), 887–895. doi: 10.3727/096504016X14817158982636

- Mao, S., Li, Y., Lu, Z., Che, Y., Sun, S., Huang, J., et al. (2019). Survival-associated alternative splicing signatures in esophageal carcinoma. *Carcinogenesis* 40 (1), 121–130. doi: 10.1093/carcin/bgy123
- Montgomery, E. A., Basman, F. T., Brennan, P., and Malekzadeh, R. (2014). Oesophageal cancer. *World Cancer Rep.* 15, 528–543.
- Park, S., Han, S.-H., Kim, H.-G., Jeong, J., Choi, M., Kim, H.-Y., et al. (2019). PRPF4 is a novel therapeutic target for the treatment of breast cancer by influencing growth, migration, invasion, and apoptosis of breast cancer cells via p38 MAPK signaling pathway. *Mol. Cell. Probes* 47, 101440. doi: 10.1016/j.mcp.2019.101440
- Passacantilli, I., Frisone, P., De Paola, E., Fidaleo, M., and Paronetto, M. P. (2017). hnRNPM guides an alternative splicing program in response to inhibition of the PI3K/AKT/mTOR pathway in Ewing sarcoma cells. *Nucleic Acids Res.* 45 (21), 12270–12284. doi: 10.1093/nar/gkx831
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47–e47. doi: 10.1093/nar/gkv007
- Runge, T. M., Abrams, J. A., and Shaheen, N. J. (2015). Epidemiology of Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterol. Clinics* 44 (2), 203–231. doi: 10.1016/j.gtc.2015.02.001
- Rustgi, A. K., and El-Serag, H. B. (2014). Esophageal carcinoma. *New Engl. J. Med.* 371 (26), 2499–2509. doi: 10.1056/NEJMra1314530
- Sadhu, A., and Bhattacharyya, B. (2019). Common subcluster Mining in Microarray Data for molecular biomarker discovery. *Interdiscip. Sci.: Comput. Life Sci.* 11 (3), 348–359. doi: 10.1007/s12539-017-0262-3
- Schizas, D., Mastoraki, A., Naar, L., Spartalis, E., Tsilimigras, D. I., Karachaliou, G.-S., et al. (2018). Concept of histone deacetylases in cancer: Reflections on esophageal carcinogenesis and treatment. *World J. Gastroenterol.* 2441, 4635. doi: 10.3748/wjg.v24.i41.4635
- Schmidt, M., Ankerst, D. P., Chen, Y., Wiethaler, M., Slotta-Huspenina, J., Becker, K.-F., et al. (2020). Epidemiologic Risk Factors in a Comparison of a Barrett Esophagus Registry (BarrettNET) and a Case–Control Population in Germany. *Cancer Prev. Res.* 13 (4), 377–384. doi: 10.1158/1940-6207.CAPR-19-0474
- Schneider, J. L., and Corley, D. A. (2015). A review of the epidemiology of Barrett's oesophagus and oesophageal adenocarcinoma. *Best Pract. Res. Clin. Gastroenterol.* 29 (1), 29–39. doi: 10.1016/j.bpg.2014.11.008
- Selvaraj, G., Kaliampurthi, S., Kaushik, A. C., Khan, A., Wei, Y.-K., Cho, W. C., et al. (2018). Identification of target gene and prognostic evaluation for lung adenocarcinoma using gene expression meta-analysis, network analysis and neural network algorithms. *J. Biomed. Inf.* 86, 120–134. doi: 10.1016/j.jbi.2018.09.004
- Selvaraj, G., Satyavani, K., Lin, S., Gu, K., and Wei, D.-Q. (2019). Prognostic impact of tissue inhibitor of metalloproteinase-1 in non-small cell lung cancer: Systematic review and meta-analysis. *Curr. Med. Chem.* 26, 7694. doi: 10.2174/0929867325666180904114455
- Shah, A. K., Lê Cao, K.-A., Choi, E., Chen, D., Gautier, Benoît, Nancarrow, D., et al. (2015). Serum glycoprotein biomarker discovery and qualification pipeline reveals novel diagnostic biomarker candidates for esophageal adenocarcinoma. *Mol. Cell. Proteomics* 14 (11), 3023–3039. doi: 10.1074/mcp.M115.050922
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome. Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303
- Shah, A. K., Gunter, H., Brown, I., Winterford, C., Na, R., Lê Cao, K.-A., et al. (2018). Evaluation of Serum Glycoprotein Biomarker Candidates for Detection of Esophageal Adenocarcinoma and Surveillance of Barrett's Esophagus. *Mol. Cell. Proteomics* 17 (12), 2324–2334. doi: 10.1074/mcp.RA118.000734
- Shen, Z., Chakraborty, A., Jain, A., Giri, S., Ha, T., Prasanth, K. V., et al. (2012). Dynamic association of ORCA with prereplicative complex components regulates DNA replication initiation. *Mol. Cell. Biol.* 32 (15), 3107–3120. doi: 10.1128/MCB.00362-12
- Sikkema, M., De Jonge, P. J. F., Steyerberg, E. W., and Kuipers, E. J. (2010). Risk of esophageal adenocarcinoma and mortality in patients with Barrett's esophagus: a systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* 8 (3), 235–244. doi: 10.1016/j.cgh.2009.10.010
- Silvers, A. L., Lin, L., Bass, A. J., Chen, G., Wang, Z., Thomas, D. G., et al. (2010). Decreased selenium-binding protein 1 in esophageal adenocarcinoma results from posttranscriptional and epigenetic regulation and affects chemosensitivity. *Clin. Cancer Res.* 16 (7), 2009–2021. doi: 10.1158/1078-0432.CCR-09-2801
- Somja, J., Demoulin, S., Roncarati, P., Herfs, Michaël, Bletard, N., Delvenne, P., et al. (2013). Dendritic cells in Barrett's esophagus carcinogenesis: an inadequate microenvironment for antitumor immunity? *Am. J. Pathol.* 182 (6), 2168–2179. doi: 10.1016/j.ajpath.2013.02.036
- Sun, H., Liu, T., Zhu, D., Dong, X., Liu, F., Liang, X., et al. (2017). HnRNPM and CD44s expression affects tumor aggressiveness and predicts poor prognosis in breast cancer with axillary lymph node metastases. *Genes Chromosomes Cancer* 56 (8), 598–607. doi: 10.1002/gcc.22463
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi: 10.1093/nar/gkw937
- Tang, Y., Li, M., Wang, J., Pan, Yi, and Wu, F.-X. (2015). CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems* 127, 67–72. doi: 10.1016/j.biosystems.2014.11.005
- Tatarian, T., and Palazzo, Francisco (2019). “Epidemiology, Risk Factors, and Clinical Manifestations of Esophageal Cancer,” in *Shackelford's Surgery of the Alimentary Tract, 2 Volume Set* (Philadelphia, United States: Elsevier), 362–367.
- Thrift, A. P. (2016). The epidemic of oesophageal carcinoma: where are we now? *Cancer Epidemiol.* 41, 88–95. doi: 10.1016/j.canep.2016.01.013
- Tramontano, A. C., Sheehan, D. F., Yeh, J. M., Kong, C. Y., Dowling, E. C., Rubenstein, J. H., et al. (2017). The impact of a prior diagnosis of Barrett's esophagus on esophageal adenocarcinoma survival. *Am. J. Gastroenterol.* 1128, 1256. doi: 10.1038/ajg.2017.82
- Tu, J., Chen, J., He, M., Tong, H., Liu, H., Zhou, B., et al. (2019). Bioinformatics analysis of molecular genetic targets and key pathways for hepatocellular carcinoma. *OncoTargets Ther.* 12, 5153. doi: 10.2147/OTT.S198802
- Uhlén, M., Björling, E., Agaton, C., Al-KhaliliSzigyarto, C., Amini, B., Andersen, E., et al. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* 4 (12), 1920–1932. doi: 10.1074/mcp.M500279-MCP200
- Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357 (6352), eaan2507. doi: 10.1126/science.aan2507
- Varghese, S., Newton, R., Ross-Innes, C. S., Lao-Sirieix, P., Krishnadath, K. K., O'Donovan, M., et al. (2015). Analysis of dysplasia in patients with Barrett's esophagus based on expression pattern of 90 genes. *Gastroenterology* 149 (6), 1511–1518. doi: 10.1053/j.gastro.2015.07.053
- Wang, Q., Ma, C., and Kemmer, W. (2013). Wdr66 is a novel marker for risk stratification and involved in epithelial-mesenchymal transition of esophageal squamous cell carcinoma. *BMC Cancer* 13 (1), 137. doi: 10.1186/1471-2407-13-137
- Wang, Q., Yan, Z., Ge, L., Li, N., Yang, M., Sun, X., et al. (2020). OSeac: An Online Survival Analysis Tool for Esophageal Adenocarcinoma. *Front. Oncol.* 10:3155. doi: 10.3389/fonc.2020.00315
- Wei, D.-Q., Selvaraj, G., and Kaushik, A. C. (2018). Computational Perspective on the Current State of the Methods and New Challenges in Cancer Drug Discovery. *Curr. Pharm. Des.* 24 (32), 3725. doi: 10.2174/138161282432190109105339
- Westra, W. M., Rygiel, A. M., Mostafavi, N., de Wit, G. M. J., Roes, A. L., Moons, L. M. G., et al. (2020). The Y-chromosome F haplogroup contributes to the development of Barrett's esophagus-associated esophageal adenocarcinoma in a white male population. *Dis. Esophagus.* doi: 10.1093/dote/daaa011
- Wu, Z.-H., Liu, C.-C., Zhou, Y.-Q., Hu, L.-N., and Guo, W.-J. (2019). OnclncRNA-626 promotes malignancy of gastric cancer via inactivated the p53 pathway through interacting with SRSF1. *Am. J. Cancer Res.* 9 (10), 2249.
- Xia, J., Gill, E. E., and Hancock, R. E. W. (2015). NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat. Protoc.* 10 (6), 823. doi: 10.1038/nprot.2015.052
- Xie, S.-H., and Lagergren, J. (2016). The male predominance in esophageal adenocarcinoma. *Clin. Gastroenterol. Hepatol.* 14 (3), 338–347. doi: 10.1016/j.cgh.2015.10.005
- Zhang, L., Qin, H., Wu, Z., Chen, W., and Zhang, G. (2018). Pathogenic genes related to the progression of actinic keratoses to cutaneous squamous cell carcinoma. *Int. J. Dermatol.* 57 (10), 1208–1217. doi: 10.1111/ijd.14131

Zhu, Y., Qiu, P., and Ji, Y. (2014). TCGA-assembler: open-source software for retrieving and processing TCGA data. *Nat. Methods* 11 (6), 599–600. doi: 10.1038/nmeth.2956

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Nangraj, Selvaraj, Kalamurthi, Kaushik, Cho and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.