Genomics 108 (2016) 93–101

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Co-expressed miRNAs in gastric adenocarcinoma

Sally Yepes [a,*], Rocío López [b,c], Rafael E. Andrade [b,c], Paula A. Rodriguez-Urrego [b,c], Liliana López-Kleine [d], Maria Mercedes Torres [a]

[a] Facultad de Ciencias, Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá D.C., Colombia
[b] Facultad de Medicina, Universidad de los Andes, Bogotá D.C., Colombia
[c] Departamento de Patología y Laboratorios, Hospital Universitario, Fundación Santa Fe de Bogotá, Bogotá D.C., Colombia
[d] Departamento de Estadística, Universidad Nacional de Colombia, Bogotá D.C., Colombia

ARTICLE INFO

Article history:
Received 20 February 2016
Received in revised form 7 July 2016
Accepted 12 July 2016
Available online 13 July 2016

Keywords:
miRNAs
Gastric adenocarcinoma
WGCNA
Node degree

ABSTRACT

Co-expression networks may provide insights into the patterns of molecular interactions that underlie cellular processes. To obtain a better understanding of miRNA expression patterns in gastric adenocarcinoma and to provide markers that can be associated with histopathological findings, we performed weighted gene correlation network analysis (WGCNA) and compare it with a supervised analysis. Integrative analysis of target predictions and miRNA expression profiles in gastric cancer samples was also performed. WGCNA identified a module of co-expressed miRNAs that were associated with histological traits and tumor condition. Hub genes were identified based on statistical analysis and network centrality. The miRNAs 100, let-7c, 125b and 99a stood out for their association with the diffuse histological subtype. The 181 miRNA family and miRNA 21 highlighted for their association with the tumoral phenotype. The integrated analysis of miRNA and gene expression profiles showed the let-7 miRNA family playing a central role in the regulatory relationships.

1. Introduction

Gastric cancer is the second leading cause of cancer-related deaths worldwide and the fourth most common malignant disease [1]. Despite these numbers, the molecular pathogenesis of gastric cancer remains to be fully explored. The role that miRNAs play in regulating gene expression in carcinogenesis suggests their promising potential as diagnostic markers of malignancy. A significant number of studies searching for differential expression profiles of miRNAs in gastric cancer, especially in normal vs. tumor samples, have been reported (for revision, see Wu WK et al.; Shrestha S et al. [2,3]). However, many studies have primarily used a limited statistical framework for analysis, such as pair-wise comparison based methods and relatively small numbers of patients. In addition, the characterization of miRNAs in histopathological subtypes is a limited field despite the widespread use of histological classification systems and the clear differences between groups.

The Laurén classification system [4] divides gastric cancers into two histological subtypes that show differences in their pathogenesis, outcomes and epidemiology. The diffuse subtype of gastric cancers is characterized by a lack of neoplastic precursor lesions. This is in contrast to the intestinal subtype, which progresses through a multistep process and has an important influence on the microenvironment that usually involves infection with *Helicobacter pylori*. Varieties of genetic and epigenetic abnormalities are defined in intestinal subtypes and precursor lesions, including point mutations, the loss of heterozygosity, microsatellite instability, and hypermethylation. On the other hand, E-cadherin deregulation via mutation or epigenetic silencing appears to be the key molecular event in diffuse cancer [5]. Some evidence has shown that patients with the intestinal subtype of gastric cancer have better outcomes than patients with the diffuse subtype [6,7].

Despite the relevance of the disease and the worldwide use of histopathological classification systems little work has been done regarding the molecular characterization of histological subtypes, particularly in regard to miRNAs expression. The main purpose of this paper was to characterize miRNA expression profiles and identifying key miRNAs related to gastric carcinogenesis and histopathological traits, using co-expression networks and supervised analysis.

The analyses were executed in two public and large gastric cancer cohorts that were previously profiled using microarrays and sequencing platforms to ensure a heterogeneous group of samples and numerous miRNAs. The integrated analysis of miRNA and gene expression was also performed to obtain a global picture of the regulatory relationships in the disease. Although data-driven conclusions provide valuable information, candidate biomarkers should demonstrate their biological plausibility or clinical utility. To contribute to achieving this requirement, markers associated with carcinogenesis were assayed in a series of 25 independent patients using real-time reverse transcription PCR (qRT-PCR).

* Corresponding author.
E-mail address: sl.yepes233@uniandes.edu.co (S. Yepes).

http://dx.doi.org/10.1016/j.ygeno.2016.07.002
0888-7543/

Weighted gene correlation network analysis (WGCNA) recognizes that gene expression data are more complex than a list of differentially expressed genes. Instead, they can be analyzed by considering correlations between genes across samples. Gene co-expression network reconstruction is a systems biology strategy that is used to explore the system-level functionality of a transcriptome. The method delineates modules of highly correlated genes, identifies the most central genes within each module and determines relationships with sample traits [8]. This analytical framework can assist in gene screening aimed at identifying candidate biomarkers or therapeutic targets [8]. Biological applications that exemplify the use of WGCNA have been reported in different settings [9–13].

2. Materials and methods

2.1. Samples and preprocessing

We applied the WGCNA approach to two gastric adenocarcinoma cohorts that were profiled for miRNA data. The first dataset corresponded to miRNASeq data from The Cancer Genome Atlas (TCGA) project (//tcga-data.nci.nih.gov/) and contained 1046 miRNAs in 295 gastric cancer samples. This cohort was used to search for relationships between miRNAs and histopathological traits. The second cohort corresponded to a microarray dataset that was retrieved from the Array Express database (www.ebi.ac.uk/arrayexpress) under the accession number E-TABM-341. This cohort profiled 315 miRNA genes in 353 gastric samples (184 cancers and 169 normal mucosae). Because this cohort contained information for tumor samples and non-tumor tissues, it was used to identify miRNA expression patterns that were associated with a cancerous condition. For the analyses, we departed from the quantile normalization data and read counts for the microarray and RNAseq, respectively. In the case of RNA-seq, the read counts were transformed using the variance-stabilizing transformation from the package DESeq2 [14]. Given that TCGA samples were processed in multiple batches, batch effect was corrected using The COMBAT method implemented in the SVA package [15]. Original expression values before preprocessing and samples traits are available as supplementary information, Tables S1 and S2.

2.2. Weighted gene co-expression analysis

Co-expression networks were built according to the protocols of WGCNA in an R environment, as previously described [8,16]. Briefly, we used a measure of similarity between the gene expression profiles that was based on a matrix of pairwise Pearson's correlation coefficients. This similarity is a measure of the level of concordance between gene expression profiles across the samples. Then, the similarity matrix was transformed into an adjacency matrix using a power adjacency function, which encodes the connection strengths between pairs of nodes [17]. The power $\beta = 3$ was chosen based on the scale-free topology criterion [17–19], resulting in a scale-free topology index (R2) of 0.925 for the TCGA cohort. For the E-TABM-341 cohort, we chose the power $\beta = 3$, which resulted in a scale-free topology index (R2) of 0.68.

To detect modules, we used average linkage hierarchical clustering with a dissimilarity measure (the Topological Overlap Measure-TOM). This measure represents the overlap observed between shared neighbors, which is a robust measure of network interconnectedness [19]. Modules were detected as branches of the dendrogram, which were cut using the Dynamic Tree-Cut algorithm [20].

To represent and summarize each module, the module eigengene was calculated. This measure is defined as the first principal component of a given module. It can be considered to be representative of gene expression profiles and to capture the maximal amount of variation in the module. To quantify module–trait associations, given that we had a summary profile (eigengene) for each module, we correlated eigengene with external traits and looked for the most significant associations. This calculation was referred to as the Module-Trait Relationships (MTRs) [8]. For intramodular analysis, we evaluate the Gene Significance (GS) and Module Membership (MM), the latter of which is also known as eigengene-based connectivity (kME). GS is the absolute value of the correlation between a specific gene and a trait. The MM is the correlation between the module eigengene and the gene expression profile. Using the GS and MM, we can identify genes that have a high significance for a clinical trait and important module membership [8]. A graphic depiction of the modules as networks was made using the program Gephi [21]. Node centrality has been shown to be useful to identify functionally critical genes; the node degree, number connections associated with a gene, was calculated and graphically visualized.

2.3. Supervised analysis

To evaluate the ability of miRNAs to classify gastric cancers and to thereby determine their potential usefulness as biomarkers in a large spectrum of patients, supervised analysis was used. The samples were classified according to their miRNA profile, their diffuse or intestinal subtype, and whether they were a non-tumor mucosa or a tumor sample using the TCGA and the E-TABM-341 datasets. Methods of supervised analysis were used: significance analysis of microarray (SAM) [22], Support Vector Machines (SVM) [23] and Random forest [24]. SAM was used as a feature extraction strategy to identify the most discriminatory genes for classification by SVM and Random forest. Before this analysis, samples were randomly divided into training and test groups. Prediction errors were calculated using a leave-one-out cross-validation method. All analyses were executed using R (R Core Team, 2014) [25].

2.4. Real-time reverse transcription PCR (qRT-PCR)

Formalin-fixed paraffin-embedded (FFPE) specimens of gastric tissues obtained during radical gastrectomies were retrieved from the samples archive at the Hospital Universitario-Fundación Santa Fe de Bogotá. All tissues were used with the approval of the ethics committee of this institution. A total of 25 paired tumor and no-tumoral tissues from the same patients were reviewed and manually microdissected by a pathologist. Expression of mature miRNAs, including 21, 199, 125b, 100, 106b, 146, 155, 315, 148a, and 181a–c, were evaluated by qRT-PCR analysis. To extract FFPE RNA, we used an Ambion's RecoverAll kit, and the selected miRNAs were quantified from total RNA using Taqman® miRNA qRT-PCR (Applied Biosystems, CA, USA) according to the manufacturer's instructions. RNU6B was used as an endogenous control. The 2(−Delta Delta C(T)) method [26] was used to analyze the relative changes in gene expression between non-tumoral mucosae and gastric adenocarcinomas.

2.5. Integrative analysis of target predictions and miRNAs expression profiles

To identify the most likely targets and obtain an overview of the regulatory relationships, we integrate the transcriptome and microRNAome obtained on the same patients (matched samples), using MAGIA web tool [27]. TCGA level 3 mRNAs information (RNAseq data), corresponding to normalized data, was retrieved from the TCGA web portal. MiRNAs and mRNA readouts were initially pre-processed (filtered, normalized, and corrected by batch effect) before integration analysis. We choose Pearson correlation as relatedness measure and the intersection of miRanda and PITA target prediction algorithms. MiRNA-mRNA anti-correlations were visualized by Cytoscape [28].

3. Results

3.1. Weighted gene co-expression analysis

Using a WGCNA approach, we analyzed gene co-expression patterns based on miRNA profiles in gastric adenocarcinoma to identify key miRNAs that regulate histopathological identity and gastric carcinogenesis in two datasets independently. In this type of analysis, correlated expression profiles imply that genes operate in collaboration or in related pathways and that they contribute together to the phenotype. Briefly, WGCNA calculates correlations among genes that are analyzed across samples, and the correlations are weighted using a power function to determine the connection strengths between genes. Co-regulated genes are grouped into modules base on similarities in their expression patterns. Finally, each module is summarized, and hub genes are identified based on intramodular statistical analysis and node centrality properties, linking network topological features to biological information.

In the TCGA cohort, after the preprocessing step, 853 miRNAs and 265 samples were used to construct the co-expression network. Clusters of highly co-expressed genes were identified and each was assigned a color. Fig. 1 demonstrates the 3 modules identified. The correlation of a module's eigengene to traits related to histopathology (Module-Trait Relationships-MTRs) show that some modules are more important than others in the disease. The module with the greatest association with clinical traits was the turquoise-colored module (Fig. 2). Therefore, this module should allow us to capture and summarize the most miRNA expression profiles. This module corresponded to 232 correlated miRNAs that were mainly associated with histological identity. Regarding the other modules that were identified, the blue and brown modules were not highly associated with the studied characteristics based on an intramodular analysis, and the gray module represents background genes.

The histological typing of gastric adenocarcinomas made by TCGA included the classification system of Laurén [3] (intestinal and diffuse carcinoma), the subclassification for intestinal types (tubular, papillary, and mucinous), and the classification system of the World Health Organization (WHO) (tubular, papillary, poorly cohesive and mucinous) [29]. The turquoise module showed a positive association with intestinal-type ($r = 0.52$; $p = 5e-20$) and a negative correlation with diffuse-type carcinoma ($r = -0.52$; $p = 5e-20$) according to the Laurén classification system. Using the WHO system, a positive correlation was observed for tubular ($r = 0.31$; $p = 2e-7$) and papillar subtypes ($r = 0.22$; $p = 4e-4$), and a negative association were observed for poorly cohesive carcinoma ($r = -0.52$, $p = 4e-20$). Other characteristics, such as tumor, nodal status and metastasis (TNM), and clinical staging, did not show a strong correlation with the modules. Despite the evident morphological differences that allow the classification of tissues into intestinal and diffuse subtypes, the co-expression patterns that differentiate these subtypes are subtle. This was much clearer when the correlations between genes and progression characteristics, such as TNM staging, were analyzed. This observation is demonstrated by the r-values of the correlations (Fig. 2).

The relationship between Module Membership and Gene Significance was highly correlated ($cor = 0.97$, $p = 2.9e-143$), which reflects the strong relationships identified between histological subtype and the turquoise module eigengene. The individual results for each miRNA and previous calculations can be found in the supplementary information. It might be concluded that, in taking into account this intramodular analysis, the turquoise module contains genes that have high significance for histological subtypes in addition to high module membership. MiRNA 100, 195, let-7c, 140, 99a, and 125b presented the highest correlations with the diffuse subtype and the miRNAs 210, 592, 130b, and 455, among others, were associated with the intestinal subtype (Table S3, supplementary data).

For the purposes of determining the relationships between genes and identifying the genes with central roles in the turquoise module, this module is represented as a network. As shown in Fig. 3, the miRNAs 100, let-7c, 125b, and 99a have high node degree; other genes with high degree score include miRNA 143, 145, 195, 1, 133, and 199. Degree defined as the number of edges incident upon a node. Therefore, these genes can be considered to be hubs in the turquoise module. The miR-99a/100, let-7 and miR-125b paralogs are encoded in two tri-cistrons on human chromosomes 11 and 21 [30]. It has been shown that miRNA families can be grouped in clusters and expressed from a same transcript [31].

We also built a gene co-expression network using the E-TABM-341 cohort. All samples were treated as one single group (both normal and tumor tissues). This network was intended to search for modules that distinguish between tumoral and normal conditions, which might therefore provide insight into the central differences between tissues with respect to the process of gastric carcinogenesis. WGCNA identified 3 co-expressed modules (Fig. 1B). As shown in the dendrogram and the MTR matrix (Fig. 4), the module represented by a brown color showed a positive association ($r = 0.5$; $p = 1e-22$) with a tumoral condition. Using the GS and MM measures, the brown module was shown to have high significance for the distinction between tumor and normal mucosae. In these calculations, the following miRNAs were found to be associated with cancer: 181b, 181d, 213, 21, 181a, 93, and 181c (Table S4, supplementary data). The network depiction highlighted

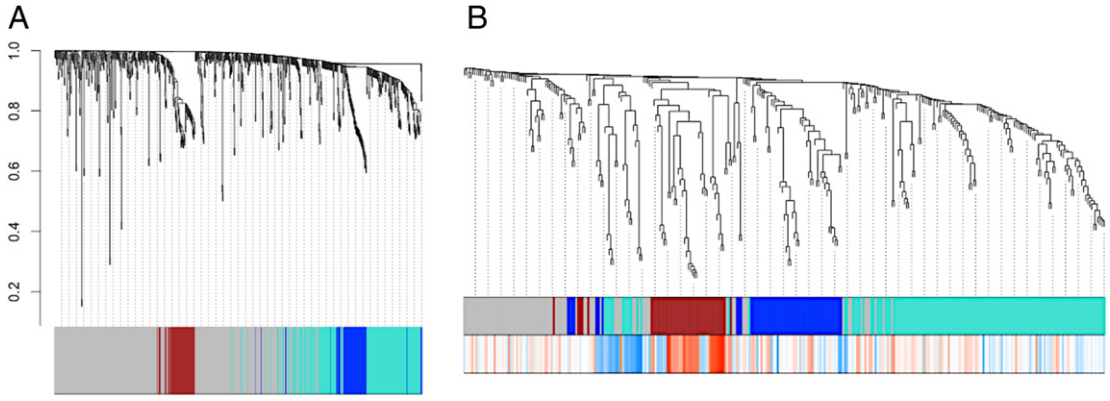

Fig. 1. Gene dendrogram showing the co-expression modules identified during the analysis of relationships with histological traits (A) and tumoral markers (B) (Average linkage, dissTOM). The color bands provide a visual comparison of the module assignments (branch cuttings) based on the dynamic hybrid cutting method. The first band shows the results of the WGCNA analysis, the second color band in panel B indicates the gene significance (GS) measure: red lines indicate a positive correlation with the tumoral condition.

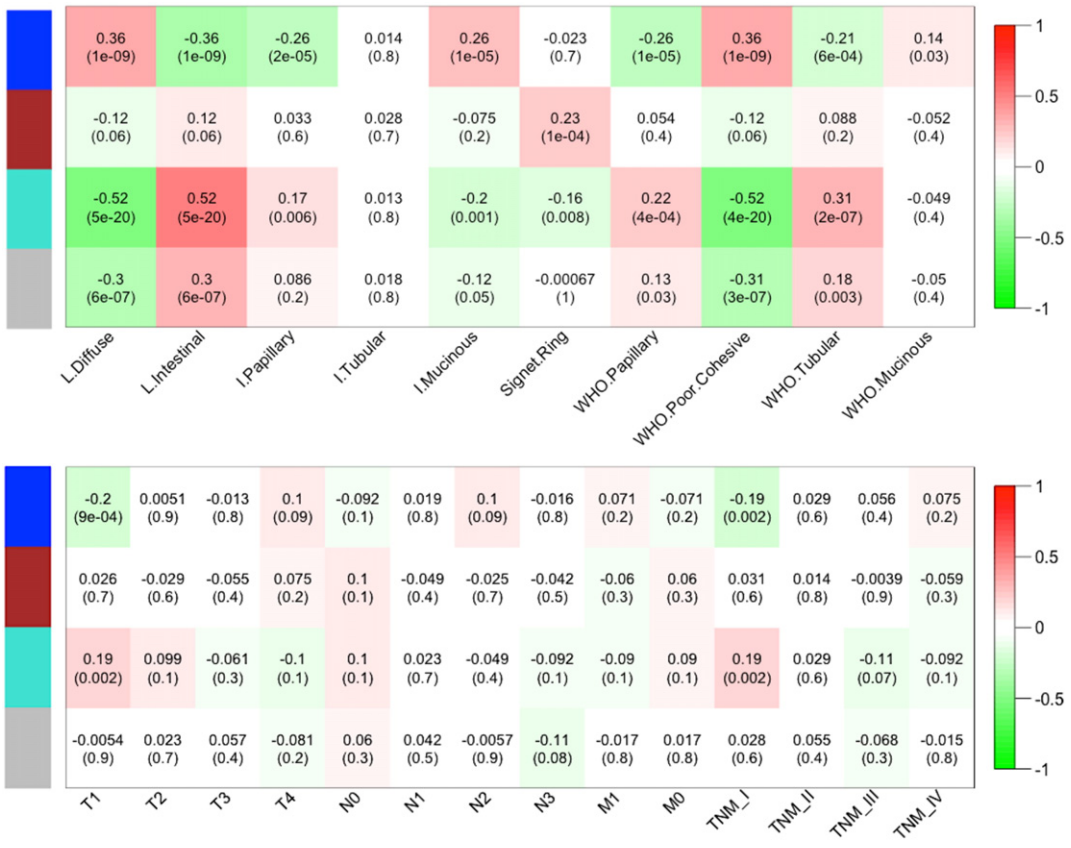

Fig. 2. Matrix of Module-Trait Relationships (MTRs) and p-values for selected traits. Each row corresponds to a module eigengene, and each column corresponds to a histopathological trait. Each cell contains a corresponding correlation and p-value. The table is color-coded by correlation according to the color legend. The histological evaluation was made using two systems, Laurén and WHO classification system. Abbreviations (L: Laurén class; I: intestinal subclass; WHO: World Health Organization class; T: extent of the tumor; N: extent of spread to the lymph nodes; M: presence of metastasis; TNM: staging system TNM).

the miRNAs with high node degree (miRNA 106a–b, 20, 93, 17, and 21, among others) (Fig. 5).

3.2. Supervised analysis

To identify which miRNAs best discriminate between the intestinal and diffuse subtypes and to determine their potential classification power, supervised statistical methods were used. After performing SAM and choosing a false discovery rate (FDR) of zero, 47 miRNAs were identified (Table 1). Strong overlap was observed between the genes chosen using this feature selection method and those chosen in the co-expression network analysis. The genes that are underlined and bold in Table 1 were found to be highly correlated with both subtypes (diffuse and intestinal) in the turquoise module. The genes obtained from SAM allowed the classification of samples bases on histological subtypes through SVM with a prediction error of 17%. Despite the error rates in the supervised analysis and after taking into account the high molecular heterogeneity of gastric cancer, the gene profile used for classification still tended to achieve distinction: a 17% error rate implies that 83% of the cases were appropriately classified using these genes. When used to distinguish between tumor and non-tumor tissues, SAM detected 26 genes with an FDR of zero (Table 1), and the estimated prediction error resulting from SVM for these lists of genes was 6%. Several genes obtained through SAM analysis were also identified as highly correlated with a cancer condition in the brown module. These genes are shown as underlined and bold miRNAs in Table 1. Random forest algorithm was also applied achieving moderate error rates, 14,8% for histology classification and 10,7% for tumoral condition. Therefore discriminating miRNAs represent probable markers involved in a wide range of samples.

3.3. Real-time reverse transcription PCR (qRT-PCR)

The expression levels of selected miRNAs were evaluated in an independent group of 25 patients using qRT-PCR. The relative change in expression of miRNAs in gastric cancer compared with non-tumoral mucosae showed a consistent pattern, only for some miRNAs, in at least 80% of samples. MiRNA 21 increased an average of 3.76-fold, and miRNA 181 a, b and c were increased an average of 1.6, 1.8, and 1.7-fold, respectively. Some miRNA shown a decreased expression in tumor samples compared to non-tumor tissues. MiRNA 315 decreased an average of 0.25-fold and 148a decreased 0.3-fold. All miRNAs mentioned above showed differential expression in the supervised analysis. Other miRNAs did not show a consistent fold change across most samples.

3.4. MiRNA-mRNA integration analysis

The analysis combined target predictions with miRNAs and gene expression to identify, among predicted target genes for each considered miRNA, those regulatory relationships significantly supported by expression data [27]. From 13.120 mRNA and 853 miRNA transcripts resulting after preprocessing, the integrative analysis highlighted the significant role of let-7 family in the disease. As shown in the miRNA-mRNA network (Fig. 6), let-7 family members (let 7b, 7e, 7c, 7d, and 7i) had the higher degree values. A total number of 238 miRNA–gene anti-correlated interactions with a FDR value <0.01 was identified. Among the top miRNA-gene anti-correlated interactions we found TP53 highly associated to let-7e ($r = -0.36$; q value $= 9.58e-8$). The Table S5 in supplementary information shown the top anti-

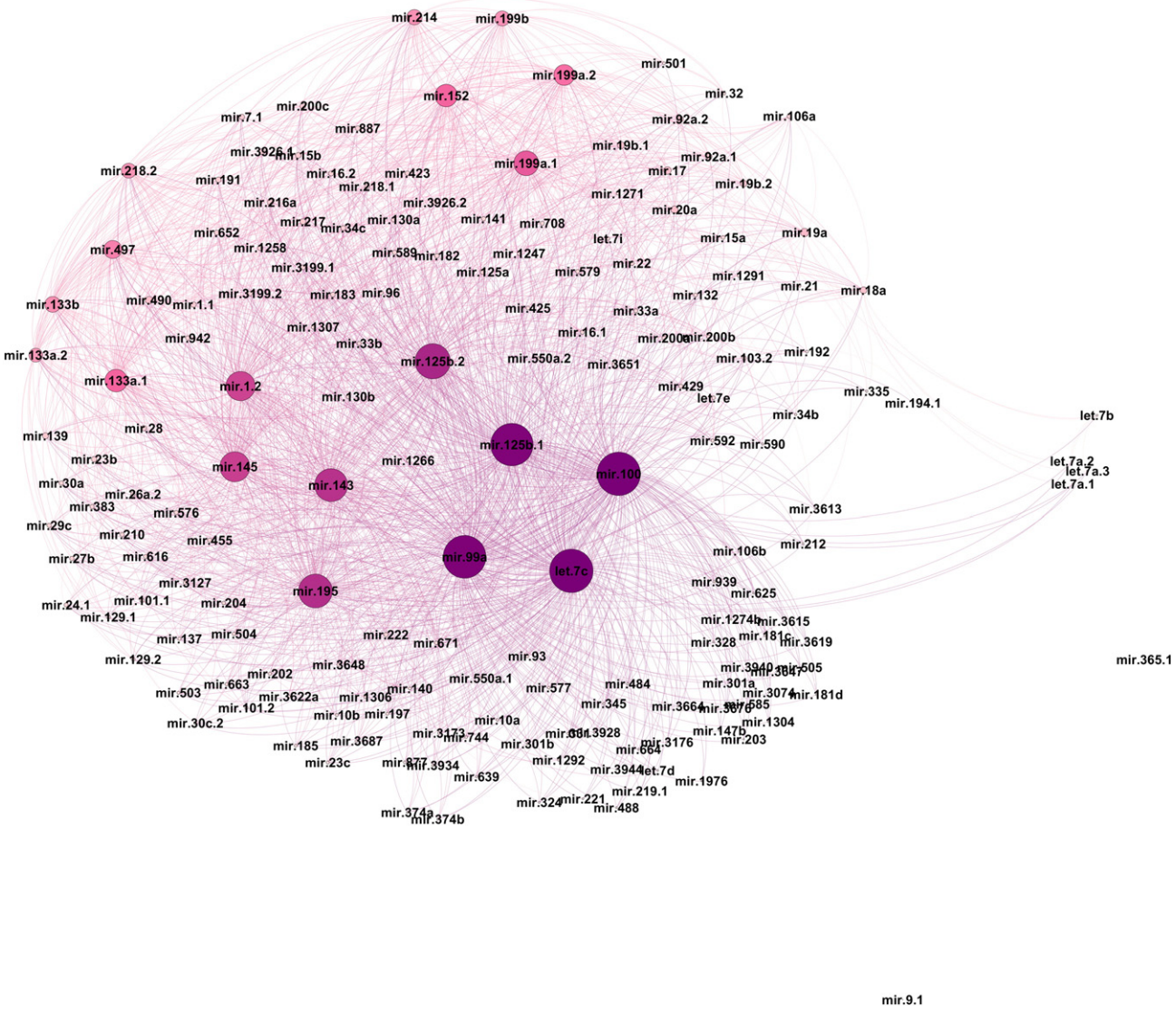

Fig. 3. Graphic depiction of the turquoise-color module. The hub miRNAs of the network are highlighted (miRNA 100, let-7c, 99a, and125b). Other genes with high degree score include miRNA 143, 145, 195, 1, 133, and 199. The bigger and darker the circle, the higher the degree score of the corresponding gene. A high node degree is an indicator that the gene plays a crucial biological role.

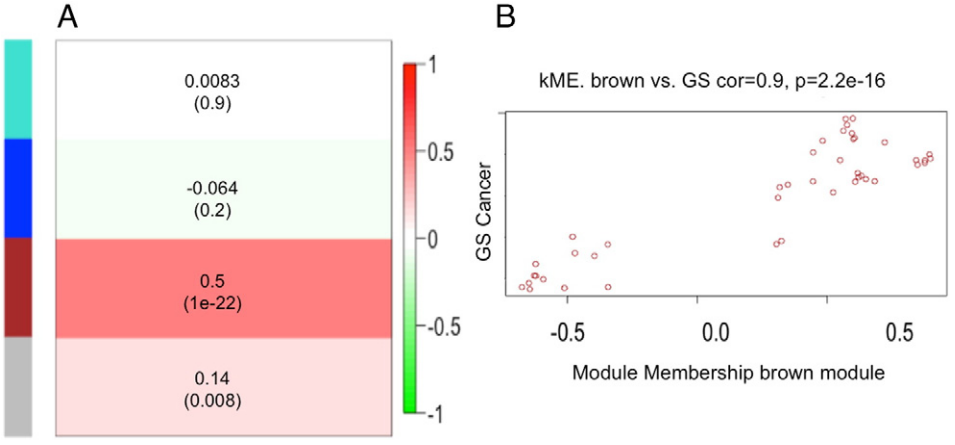

Fig. 4. Intramodular analysis of a cancer-related module (brown) A. Matrix of Module-Trait Relationships (MTRs) for the tumoral condition. Each cell states the correlation and p-value that resulted from correlating module eigengene with the tumoral phenotype. The table is color-coded by correlation according to the color legend. B. Scatterplot of Gene Significance (GS) for the cancer condition versus Module Membership (kME) for the tumor-related module. GS and kME are highly correlated, which reflects the strong correlations between the tumoral condition and the respective module eigengene.

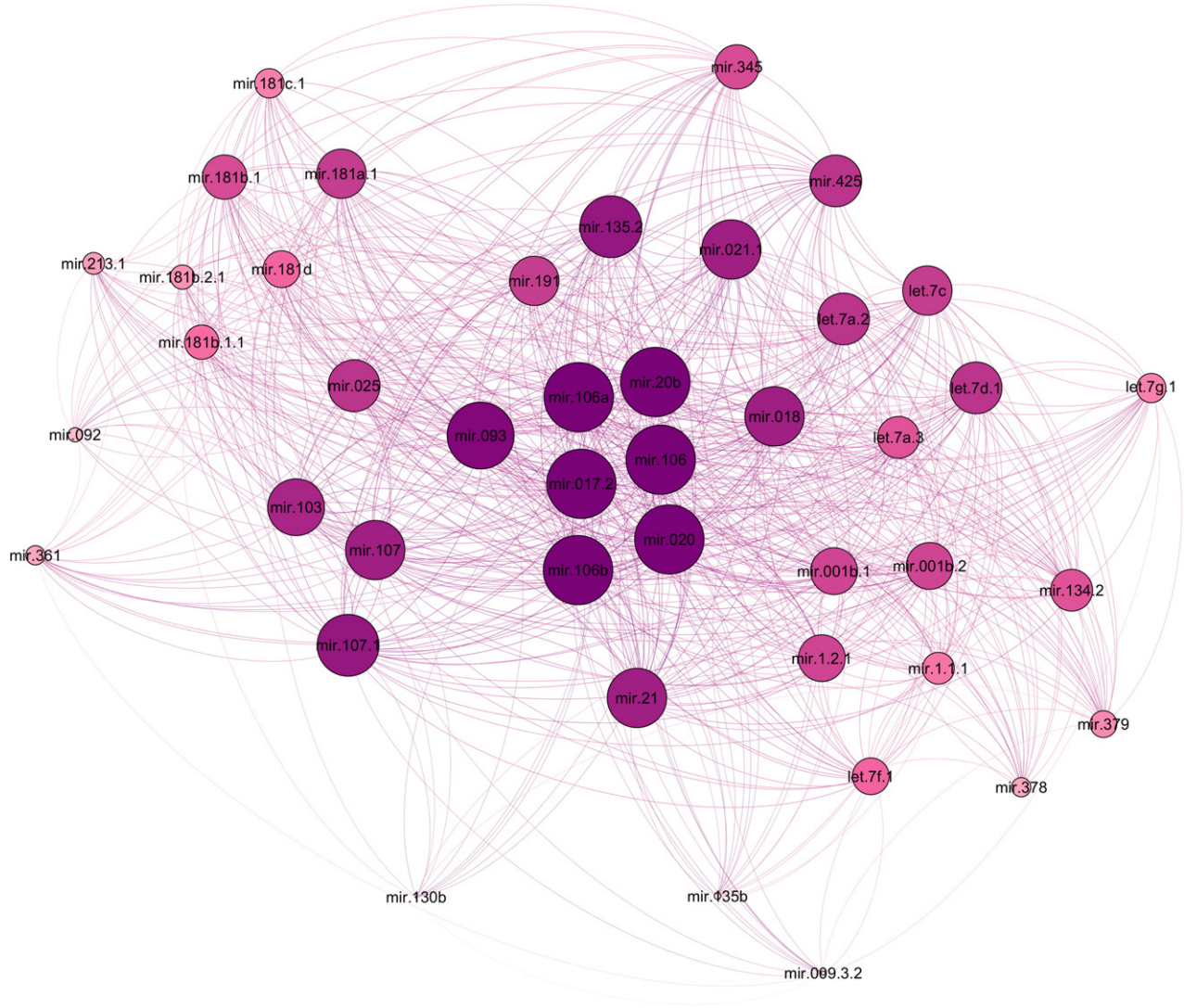

Fig. 5. Graphic depiction of the modules related to tumoral condition. The miRNA hubs of the module are highlighted (miRNA 106a, 106b, 20, 93, 17, and 21). The bigger and darker the circle, the higher their node degree score.

correlated miRNA-mRNA interactions, ranked by statistical significance. The same integrated analysis carried out independently for each histological subtype produces comparable results, let-7 family members (let 7b, 7e, 7c, 7d, and 7i) shown higher node degree values in both diffuse and intestinal miRNA-mRNA networks.

This result emphasizes the importance of integration of different genomic data (as mRNA-miRNA) in order to get more generalized results. Even though the individual analysis of one biological level (miRNA) gives information associated with the disease, the integration with other biological levels is necessary to have a more realistic view from a functional perspective. The integration results suggest and outstanding regulatory role of the let-7 family in the disease that deserves further study.

4. Discussion

Given differences in the pathogenesis, outcome and epidemiology between histological subtypes of gastric adenocarcinoma and the extensive use of histological classification systems, a molecular characterization of histological identity that includes different levels of genomic information is an objective to be pursued. We develop two co-expression network analyses; the first one used data from the TCGA project with the goal of analyzing histological aspects, and the second network used data from the EBI with the intention of study the oncogenesis process.

From 1046 miRNAs profiled in the TCGA cohort, we identified a group of 232 miRNAs that were co-expressed and showed that some of them are related to histological subtype. The centrality measure known as node degree highlighted the miRNAs 100, let-7c, 125b, 99a, and 143 as functionally critical components of the module and associated with diffuse subtype. Genes with strong network connections are likely to be co-regulated with their co-expressed genes and to lead phenotype-specific cellular processes [32]. These genes were also noted to have high correlations (GS), and they were also identified in the supervised analysis. Overall those analyses, emphasizing the biological implication of particular miRNAs, but clearly, they are not the only molecules associated with the phenotypes. A cancer phenotype is complex in nature, involving deregulation of different level of biological information and variety of molecules. We characterized here the miRNA layer and there would be numerous other variables that would be major contributors of the phenotypes.

The implication of the aforementioned miRNAs in gastric cancer has been noticed before, Ueda et al. 2010 [33], in a comprehensive microarray expression analysis, found miRNA125b, 199 and 100 as the most important miRNAs related to progression, they showed increasing expression levels according to clinical stage progression. Those miRNAs

Table 1
miRNAs obtained by SAM for histological and cancer classification.

Histological subtype Diffuse vs. intestinal			Type of gastric tissue Cancer vs. normal		
Gene	Direction	R.fold	Gene	Direction	R.fold
100	Up	3,25	148a	Down	0,31
let.7c	Up	3,34	148	Down	0,35
99a	Up	3,61	133a-2	Down	0,32
140	Up	1,51	148b	Down	0,43
125b.2	Up	3,16	375	Down	0,39
195	Up	2,28	133a-1	Down	0,39
125b.1	UP	2,54	204-1	Down	0,48
132	Up	1,56	**181d**	Up	1,78
130b	Down	0,54	**21**	Up	1,88
139	Up	2,14	**181b-1.1**	Up	1,71
210	Down	0,25	**181b-2.1**	Up	1,63
455	Down	0,44	**181b-1**	Up	1,72
3926.1	Up	2,2	**21**	Up	1,82
22	Up	1,41	451	Down	0,56
592	Down	0,34	**181a-1**	Up	1,59
497	Up	1,89	211-1	Down	0,6
33a	Down	0,47	**20b**	Up	1,58
1307	Down	0,58	**017-2**	Up	1,59
199a.1	Up	1,65	**20**	Up	1,62
141	Down	0,48	486	Down	0,57
216a	Up	2,45	**106**	Up	1,55
214	Up	1,67	**213-1**	Up	1,46
199a.2	Up	1,6	**181c-1**	Up	1,58
105.2	Down	0,21	**106a**	Up	1,55
152	Up	1,53	**93**	Up	1,5
217	Up	2,67	18	Up	1,7
577	Down	0,32			
218.2	Up	2,2			
33b	Down	0,49			
199b	Up	1,56			
345	Down	0,58			
127	Up	1,69			
19a	Down	0,6			
3199.2	Up	1,67			
767	Down	0,23			
887	Up	1,5			
379	Up	1,69			
105.1	Down	0,24			
130a	Up	1,52			
410	Up	1,78			
let-7i	Up	1,31			
212	Up	1,39			
655	Up	1,6			
143	Up	2,08			
380	Up	1,74			
381	UP	1,52			
218.1	Up	1,85			

Genes are listed in order of significance. Underlined and bold genes were found to be correlated with diffuse and intestinal subtypes in the turquoise module and with cancerous phenotypes in the brown module.

were identified in that paper through a class comparison statistical method. Interestingly, we confirm those miRNAs through a co-expression network analysis using 295 different patients and an alternative analytical platform (TCGA cohorts and RNAseq profiles). Therefore, those miRNas have the potential to be considered biomarkers and encourage further study.

The miRNA-125b-2 is located on chromosome 21 within a phylogenetically conserved cistron that also contains miR-99a and let-7c (99a/let-7c/125b-2). Other units with high homology reside on chromosome 11 (100/let-7a-2/125b-1), and another less-conserved cistron is located on chromosome 19 (99b/let-7e/125a) [30]. According to Emmrich et al. [30], miRNAs 99a/100–125b regulate hematopoietic stem cell and progenitor cell homeostasis through the TGFb and Wnt signaling pathways. Given the nature of WGCNA, this work supports the finding that some miRNAs are expressed in clusters. Increasing our understanding of how they regulate gene expression to induce phenotypes is an issue worth exploring in the future.

The first analysis of the TCGA cohort proposed a new genomic classification system of gastric cancer: EBV-infected tumours; MSI tumours; genomically stable tumours; and chromosomally unstable tumours [34]. This study used unsupervised clustering and various molecular levels. We confirm that the hubs mention earlier (miRNA 100, let-7c, 99a, and 125b) belong to the genomically stable group, other genes with high degree scores (miRNA 143, 145, 195, 1, 133, 199) were also found in that molecular subgroup. Our study highlights miRNAs that can be associated with the new molecular stable subtype proposed by the TCGA study. This subgroup is also enriched in mutations of RHOA or fusions involving RHO-family GTPase-activating proteins and alteration in cell adhesion pathways as reported by The Cancer Genome Atlas project [34].

Among the miRNAs with high values of gene significance associated with the intestinal subtype stands out the miRNA 210. The miRNA 210 is a candidate gene involved in epithelial-mesenchymal transition (EMT) and a potential target to inhibit invasion and metastasis of gastric cancer [35]. Recent evidence has shown that under hypoxic conditions, miRNA 210 becomes highly upregulated in response to hypoxia inducing factors, leading to deregulation of cell cycle, mitochondria function, apoptosis, angiogenesis and metastasis [36]. The roles of miRNA210 in different human tumor models suggest that it can be used as a therapeutic target [37].

The network construction that we used to identify miRNA expression patterns associated with the cancerogenesis, highlighted genes in the miRNA-106b-25 cluster, the 181 family and the miRNA 21, among others. The miRNA-106b-25 cluster encodes three miRNAs, miR-106b, miR-93, and miR-25, which are embedded within the MCM7 gene. This cluster is abnormally upregulated in gastric cancer compared to normal tissues, and these genes exert their oncogenic functions by suppressing Cip/Kip family members, which are Cdk inhibitors [38]. The cluster seems to modulate the cell cycle and impair TGFβ tumor suppressor pathways in gastric cancer [39–41].

Agreement between analyses at the level of individual genes was found, and the use of two different methodological frameworks reinforced the relevance of some miRNAs to observed phenotypes. The miRNA family 181 (181a-1, 181a-2, 181b-1, 181b-2, 181c and 181d) forms clusters of miRNAs that are located on chromosomes 1, 9 and 19. They have oncogenic functions that have been reported in gastric cancer [33]. They were identified in the supervised analysis and confirmed in the qRT-PCR analysis. Another gene that was identified in the network to have topological role and shown differential expression was miRNA 21, an important oncomir that is expressed in gastric cancer. MiRNA 21 promoted cell proliferation and inhibited apoptosis in gastric cancer cells, and it is upregulated not only in gastric cancer but also in *H. pylori* infections [42].

The integrated analysis of miRNA and gene expression profiles highlighted the let-7 miRNA family as central in the regulatory relationships, the analysis inferred functional relationships and provides an integrated picture of the disease. Remarkably, Ueda et al. [33] analyzed correlation between miRNA expression and survival with Cox regression and Kaplan-Meier curves, their analyses highlighted let7 family members (as let-7e, let-7g, let-7i, and let-7b), among other genes, associated with patient's survival. It is clear in our work the relevant role of the let-7 family in the regulation of the disease, over 850 miRNAs only 5 were present in the top miRNA-mRNA network, all members of the Let-7 family. Let-7 miRNA family is downregulated in various solid tumours and has roles as tumor suppressor genes to target oncogenes such as high mobility group A2 (HMGA2) and RAS [43,44]. Loss of let-7 in cancer cells induces reverse embryogenesis and dedifferentiation [45]. The integrated approach manifest the importance of analyzing different levels of biological information in order to gain a better understanding of the cancer complexity.

The present study is a systems biology approach to the study of gastric adenocarcinoma that helped to reduce the complexity of multivariate datasets and allowed us to gain insights about the miRNAs

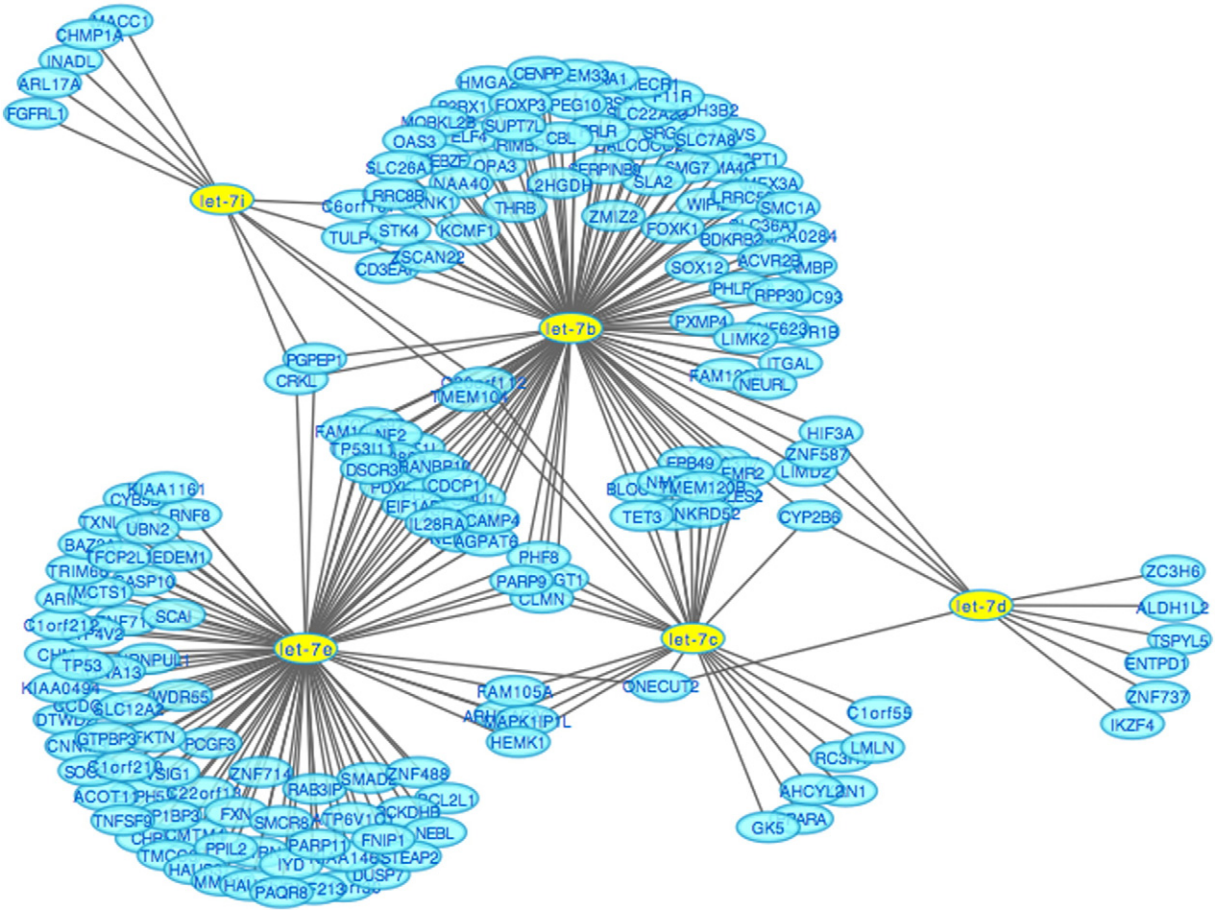

Fig. 6. Network for the top 250 most probable functional miRNA–mRNA interactions. Yellow ovals represent genes with high network centrality, node degree (miRNA let-7e, let-7b, let-7c, let-7d, and let-7i).

expression patterns. It is clear that further studies are needed to fully establish whether the miRNAs and their targets would be used in clinical settings.

5. Conclusion

The workflow followed here provides a systems biology perspective to characterizing miRNAs expression in gastric adenocarcinoma, a disease with a few molecular biomarkers. Contributing to the prioritization of miRNAs associated with histological subtyping and gastric carcinogenesis.

Author contributions

Conceived and designed the experiments: SY. Performed the experiments: SY, RL, REA, PARU. Analyzed the data: SY, RL, REA, PARU, LLK, MMT. Wrote the paper: SY.

Funding

Centro de Estudios e Investigación en Salud-Fundación Santa Fe de Bogotá, and Universidad de los Andes supported this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2016.07.002.

References

[1] J. Ferlay, H.R. Shin, F. Bray, D. Forman, C. Mathers, D.M. Parkin, Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008, Int. J. Cancer 127 (12) (Dec 15 2010) 2893–2917.
[2] W.K. Wu, C.W. Lee, C.H. Cho, D. Fan, K. Wu, et al., MicroRNA dysregulation in gastric cancer: a new player enters the game, Oncogene 29 (43) (Oct 28 2010) 5761–5771.
[3] S. Shrestha, S.D. Hsu, W.Y. Huang, H.Y. Huang, W. Chen, et al., A systematic review of microRNA expression profiling studies in human gastric cancer, Cancer Med. 3 (4) (Aug 2014) 878–888.
[4] P. Laurén, The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification, Acta Pathol. Microbiol. Scand. 64 (1965) 31–49.
[5] M. Vauhkonen, H. Vauhkonen, P. Sipponen, Pathology and molecular biology of gastric cancer, Best Pract. Res. Clin. Gastroenterol. 20 (4) (2006) 651–674.
[6] H. Zheng, H. Takahashi, Y. Murai, Z. Cui, K. Nomoto, S. Miwa, et al., Pathobiological characteristics of intestinal and diffuse-type gastric carcinoma in Japan: an immunostaining study on the tissue microarray, J. Clin. Pathol. 60 (3) (Mar 2007) 273–277.
[7] K. Yamashita, S. Sakuramoto, N. Katada, N. Futawatari, H. Moriya, K. Hirai, et al., Diffuse type advanced gastric cancer showing dismal prognosis is characterized by deeper invasion and emerging peritoneal cancer cell: the latest comparative study to intestinal advanced gastric cancer, Hepato-Gastroenterology 56 (89) (Jan-Feb 2009) 276–281.
[8] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (Dec 29 2008) 559.
[9] S. Bocklandt, W. Lin, M.E. Sehl, F.J. Sánchez, J.S. Sinsheimer, S. Horvath, et al., 2011 epigenetic predictor of age, PLoS One 6 (6) (2011), e14821.
[10] M.J. Hawrylycz, E.S. Lein, A.L. Guillozet-Bongaarts, E.H. Shen, L. Ng, J.A. Miller, et al., An anatomically comprehensive atlas of the adult human brain transcriptome, Nature 489 (7416) (Sep 20 2012) 391–399.
[11] F.E. Dewey, M.V. Perez, M.T. Wheeler, C. Watt, J. Spin, P. Langfelder, et al., Gene coexpression network topology of cardiac development, hypertrophy, and failure, Circ. Cardiovasc. Genet. 4 (1) (Feb 2011) 26–35.
[12] W.C. Chou, A.L. Cheng, M. Brotto, C.Y. Chuang, Visual gene-network analysis reveals the cancer gene co-expression in human endometrial cancer, BMC Genomics 15 (Apr 23 2014) 300.

[13] S. Horvath, B. Zhang, M. Carlson, K.V. Lu, S. Zhu, R.M. Felciano, et al., Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target, Proc. Natl. Acad. Sci. U. S. A. 103 (46) (Nov 14 2006) 17402–17407.
[14] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.
[15] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (6) (2012, Mar 15) 882–883.
[16] S. Horvath, Weighted Network Analysis. Applications in Genomics and Systems Biology, Springer Book, 2011 (ISBN: 978–1–4419-8818-8).
[17] B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis, Stat. Appl. Genet. Mol. Biol. 4 (2005), 17.
[18] H. Jeong, S.P. Mason, A.L. Barabási, Z.N. Oltvai, Lethality and centrality in protein networks, Nature 411 (6833) (May 3 2001) 41–42.
[19] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, A.L. Barabási, Hierarchical organization of modularity in metabolic networks, Science 297 (5586) (Aug 30 2002) 1551–1555.
[20] P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R, Bioinformatics 24 (5) (Mar 1 2008) 719–720.
[21] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks, Proceedings of the Third International ICWSM Conference, 2009.
[22] H. Schwender, siggenes: multiple testing using SAM and Efron's empirical Bayes approaches, R Package Version 1.42.0, 2012.
[23] Carey V, Gentleman R, Mar J, Vertrees, cfJ and Gatto L. MLInterfaces: uniform interfaces to R machine learning procedures for data in Bioconductor containers. R package version 1.48.0, 2015.
[24] L. Breiman, Random forests, Mach. Learn. 45 (1) (Oct 2001) 5–32.
[25] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014 (ISBN 3–900051–07-0, URL http://www.R-project.org/).
[26] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) method, Methods 25 (4) (Dec 2001) 402–408.
[27] G. Sales, A. Coppe, A. Bisognin, M. Biasiolo, S. Bortoluzzi, et al., MAGIA, a web-based tool for miRNA and Genes Integrated Analysis, Nucleic Acids Res. 38 (Web Server issue) (Jul 2010) W352–W359.
[28] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (Nov 2003) 2498–2504.
[29] World Health Organization classification of tumours, in: S.R. Hamilton, L.A. Aaltonen (Eds.), Pathology and Genetics of Tumours of the Digestive System, IARC Press, Lyon, 2000.
[30] S. Emmrich, M. Rasche, J. Schöning, C. Reimer, S. Keihani, A. Maroz, et al., miR-99a/100 ~ 125b tricistrons regulate hematopoietic stem and progenitor cell homeostasis by shifting the balance between TGFβ3 and Wnt signaling, Genes Dev. 28 (8) (Apr 15 2014) 858–874.
[31] L.S. Servín-González, A.J. Granados-López, J.A. López, Families of microRNAs expressed in clusters regulate cell signaling in cervical cancer, Int. J. Mol. Sci. 16 (6) (Jun 5 2015) 12773–12790.
[32] F.J. Azuaje, Selecting biologically informative genes in co-expression networks with a centrality score, Biol. Direct 9 (Jun 19 2014) 12.
[33] T. Ueda, S. Volinia, H. Okumura, M. Shimizu, C. Taccioli, et al., Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis, Lancet Oncol. 11 (2) (Feb 2010) 136–146.
[34] Cancer Genome Atlas Research Network, Comprehensive molecular characterization of gastric adenocarcinoma, Nature 513 (7517) (Sep 11 2014) 202–209.
[35] P. Yu, S. Fan, L. Huang, L. Yang, Y. Du, MIR210 as a potential molecular target to block invasion and metastasis of gastric cancer, Med. Hypotheses 84 (3) (Mar 2015) 209–212.
[36] K. Dang, K.A. Myers, The role of hypoxia-induced miR-210 in cancer progression, Int. J. Mol. Sci. 16 (3) (Mar 19 2015) 6353–6372.
[37] L. Hong, Y. Han, H. Zhang, Q. Zhao, Y. Qiao, miR-210: a therapeutic target in cancer, Expert Opin. Ther. Targets 17 (1) (Jan 2013) 21–28.
[38] Y.K. Kim, J. Yu, T.S. Han, S.Y. Park, B. Namkoong, D.H. Kim, et al., Functional links between clustered microRNAs: suppression of cell-cycle inhibitors by microRNA clusters in gastric cancer, Nucleic Acids Res. 37 (5) (Apr 2009) 1672–1681.
[39] T. Kan, F. Sato, T. Ito, N. Matsumura, S. David, Y. Cheng, et al., The miR-106b-25 polycistron, activated by genomic amplification, functions as an oncogene by suppressing p21 and Bim, Gastroenterology 136 (5) (May 2009) 1689–1700.
[40] F. Petrocca, R. Visone, M.R. Onelli, M.H. Shah, M.S. Nicoloso, I. de Martino, et al., E2F1-regulated microRNAs impair TGFb dependent cell-cycle arrest and apoptosis in gastric cancer, Cancer Cell 13 (3) (Mar 2008) 272–286.
[41] Y. Espinosa-Parrilla, X. Muñoz, C. Bonet, N. Garcia, A. Venceslá, N. Yiannakouris, et al., Genetic association of gastric cancer with miRNA clusters including the cancer-related genes MIR29, MIR25, MIR93 and MIR106: results from the EPIC-EURGAST study, Int. J. Cancer 135 (9) (Nov 1 2014) 2065–2076.
[42] Z. Zhang, Z. Li, C. Gao, P. Chen, J. Chen, W. Liu, et al., miR-21 plays a pivotal role in gastric cancer pathogenesis and progression, Lab. Investig. 88 (12) (Dec 2008) 1358–1366.
[43] S. Roush, F.J. Slack, The let-7 family of microRNAs, Trends Cell Biol. 18 (10) (Oct 2008) 505–516.
[44] B. Boyerinas, S.M. Park, A. Hau, A.E. Murmann, M.E. Peter, The role of let-7 in cell differentiation and cancer, Endocr. Relat. Cancer 17 (1) (Jan 29 2010) F19–F36.
[45] M.E. Peter, Let-7 and miR-200 microRNAs: guardians against pluripotency and cancer progression, Cell Cycle 8 (6) (Mar 15 2009) 843–852.