Evolution of the Mir-155 Family and Possible Targets in Cancers and the Immune System

Guang-Bing Xie¹, Wei-Jia Liu¹, Zhi-Jun Pan¹, Tian-Yin Cheng²*, Chao Luo¹*

Abstract

The mir-155 family is not only involved in a diversity of cancers, but also as a regulator of the immune system. However, the evolutionary history of this family is still unclear. The present study indicates that mir-155 evolved independently with lineage-specific gain of miRNAs. In addition, arm switching has occurred in the mir-155 family, and alternative splicing could produce two different lengths of ancestral sequences, implying the alternative splicing can also drive evolution for intragenic miRNAs. Here we screened validated target genes and immunity-related proteins, followed by analysis of the mir-155 family function by high-throughput methods like the gene ontology (GO) and Kyoto Enencyclopedin of Genes and Genemmes (KEGG) pathway enrichment analysis. The high-throughput analysis showed that the CCND1 and EGFR genes were outstanding in being significantly enriched, and the target genes cebpb and VCAM1 and the protein SMAD2 were also vital in mir-155-related immune reponse activities. Therefore, we conclude that the mir-155 family is highly conserved in evolution, and CCND1 and EGFR genes might be potential targets of mir-155 with regard to progress of cancers, while the cebpb and VCAM1 genes and the protein SMAD2 might be key factors in the mir-155 regulated immune activities.

Keywords: mir-155 family - evolution - function - interaction - cancer - immunity

Introduction

MicroRNAs (miRNAs) are approximately 22nt, endogenous and non-coding RNAs that play important roles in the post-transcriptional repression of organisms (Bartel., 2004). Generally, the miRNAs are conducted by the RNA polymerase II, forming the primary transcripts (pri-miRNAs). Then the pri-miRNAs are cleaved by RNase III enzyme Drosha with the help of DGCR8 (Landthaler et al., 2004; Price et al., 2011), followed by the release of a hairpin-shaped RNA molecule called miRNA precursor (pre-miRNA). The transport of pre-miRNAs into the cytoplasm by the Exportin 5 will then trigger the alternative splicing by Dicer enzyme and form an imperfect symmetry (miRNA: miRNA*duplex), consisted of one mature (3'arm) and one immature (5'arm) sequences (Yi et al., 2003; Zhang et al., 2004; MacRae et al., 2006). After that, the mature one enters the RNA-induced silencing complex (RISC) while the other will be degraded or retained in some species (Schwarz et al., 2003; Bartel, 2004). The sequence analysis of pre-miRNA and miRNA: miRNA*duplex in different species can help us to understand the distribution and evolution of a miRNA family. Besides, it has been hot topic to identify the regulation function of miRNAs in wide ranging signaling cascades and modulators of different cell-type and context dependent activities (Farooqi et al., 2014).

It is commonly believed that microRNAs are evolutionarily conserved among different organisms (Xu et al., 2012). However, a considerable amount of studies have stated that most microRNAs follows dynamic evolution, especially the novel microRNAs. For example, miR-941 was found to be a rapid regulatory evolution microRNA in the human lineage, which leads to a diversity of evolution patterns (Hu et al., 2012). Then many current studies either focus on the investigation of cluster evolution or the local duplication evolutionary patterns of a microRNA family, for example, the mir-2 family, the let-7 family, the mir-17 family, the mir-395 family, the mir-548 family (Guddeti et al., 2005; Concepcion et al., 2012; Hertel et al., 2012; Liang et al., 2012; Marco et al., 2012a). However, it has been seldom studied how the independent and single-member microRNA family evolves up to date. Moreover, the current studies have usually emphasized on the construction of protein-protein interaction (PPI) network, but pay rare attention on the miRNA regulatory network.

MiRNA-155 (mir-155) is an important regulatory factor participating in many physiological and pathological processes of organism (O’Connell et al., 2012; Elton et al., 2014),...
2013), definitely such as malignant growth (Babar et al., 2012; Mattiske et al., 2012), viral infections (Vargova et al., 2011; Wang et al., 2011) and cardiovascular diseases (Corsten et al., 2012). Further, the mir-155 family is an independent and single-member’s family with unclear evolutionary patterns. Therefore, we extracted all the messages about the mir-155 family in different species, and performed multiple sequence alignment, as well as comparative genomics in the current study. We expect the study can supply new insights in the evolutionary patterns of independent and single-member miRNA family and the miRNA regulatory network.

Materials and Methods

Data source

First, the Taxonomy Tools of the NCBI (http://www.ncbi.nlm.nih.gov/taxonomy/) was used to determine the distribution of the mir-155 family. Then all the members of mir-155 family, as well as their corresponding species, annotations, deep sequencing data, miRNA/miRNA* complex and pre-miRNA sequences from different animal species were extracted from the miRase database (Release 20 http://mirbase.org/index.shtml) (Kozomara and Griffiths-Jones, 2011).

The construction of Phylogenetic trees

To identify the ancestral sequence, Phylogenetic trees of mir-155 family were reconstructed with MEGA 4.0 (Tamura et al., 2007) by 1,000 bootstrap resampling using Neighbor-Joining (NJ) method. It has been reported that the mir-100 is the only microRNA which is shared by all bilateria (Grimson et al., 2008; Griffiths-Jones et al., 2011). Therefore, we also analyzed the relationship between the mir-100 and mir-155 using Neighbor-Joining (NJ) method. After that, we screened the species containing the ancestral sequences of mir-155 from the mir-100 family. Then we compared the sequences mir-155 and mir-100 of the same species to determine the origin of the mir-155 family.

Relative Arm Usage

To quantify the relative amount of mature miRNA products from each arm of the same miRNAhairpin precursor, we performed the “relative arm usage” measure as reported (Marco et al., 2010). The quantity of relative arm usage was calculated as follows: Relative arm usage=$\log^2 (N_1'/N_3')$; where $N_1'$ is the number of reads mapped to the 5’ arm of the hairpin precursor and $N_3'$ the number of reads from the 3’ arm. Positive values indicate a bias toward 5’ arm usage while the negative values a bias toward 3’ arm. Zero means that mature sequences are produced at equal levels from both arms. Besides, miRDB is an online database system used for microRNA target prediction and functional annotation (Wang, 2008). In the current study, we used the miRDB database to predict the target sites of mir-155.

Prediction of the target genes and the target mRNAs

BioMart is a freely available, open source, federated database system that provides a unified access to disparate, geographically distributed data sources (Zhang et al., 2011). The Ensembl and Ensembl Genomes BioMarts can supply us a point of access to the high-quality gene annotation, variation data, functional and regulatory annotation and evolutionary relationships from genomes spanning the taxonomic space (Kinsella et al., 2011). To predict the expressed target mRNAs of mir-155, we fetched the messages of 3’ UTR of the whole-genome from ENSEMBEL (Release 72) via Biomart. Here we only chose the canonical seed to predict the 3’UTRs (Bartel, 2009), and the target sites were performed with miRecords (Xiao et al., 2009) and miRTarbase (Hsu et al., 2011).

Functional and pathway enrichment analysis

Molecule Annotation System (MAS 3.0, http://bioinfo.capitalbio.com/mas3/) is convenient analysis management software for the analysis of significantly enriched Gene Ontology (GO), pathway terms with user-friendly web interface and stable and fast analysis procedure. The selected target genes were queried using MAS to obtain richness of the functions and Kyoto Encyclopedia OF Genes & Genomes (KEGG) pathway. Based on the hypergeometric distribution, the pathways or GO terms with $p<0.01$ and $q<0.01$ were chose for the next analysis.

Construction of an miRNA-target gene- cancer regulatory network

Cytoscape is a popular bioinformatics package for biological network visualization and data integration (Smoot et al., 2011). We selected the cancer-related pathways from the enriched KEGG pathways, and then constructed a regulatory network of mir-155-pathway-target gene with the help of Cytoscape 2.8.

Construction of an mir-155-target gene-protein regulatory network

Human Protein Reference Database (HPRD) an object-oriented database that integrates a great deal of information relevant to the function of human proteins in health and disease (Prasad et al., 2009). In the present study, we chose the genes which were regulated by mir-155 and enriched in the immune related functions after functional enrichment analysis. Then we analyzed and constructed the mir-155-target gene-protein regulatory network using Cytoscape 2.8. All the protein interaction data used here should be verified by experiment before.

Results

Mir-155 family all distributes in vertebrates

The annotation in the miRBase database (Release 20) showed that mir-155 family totally had 20 members. Based on the results of taxonomy browser of NCBI, we found that the members of mir-155 family all distributed in vertebrates (Figure 1).

Mir-155 is a conserved microRNA family in vertebrates

According to the multiple alignment of all precursor sequences (Figure 2a), we found that the 5’ arm of the family precursor was highly conserved, and 3’ arm has some changes the nucleotides. The highly conserved
nucleotides are usually located in 3' arm in many microRNAs family, but in the mir-155 family, the highly conserved nucleotides were 5' arm instead of 3' arm (Figure 2a). Therefore, we speculated the arm switching had happened in mir-155 family.

The ancestral sequence and rare local duplication of the mir-155 family

To find out the ancestral sequences of mir-155 family, we constructed NJ phylogenetic tree against all the precursor sequences. According to the analysis of the Phylogenetic trees (Figure 3), the ancestral sequences of mir-155 family might have a mutation into two directions: a more inclined to fish, the other to amphibians, and then further to mammals, quadrumanas. The gene sequence of zebrafish or Ictalurus punctatus, etc which were the first branch separated from ancestral mir-155 gene families might have the more similar sequence with the ancestors. The ancestral sequences of mir-155 family were DRE-mir-155 and IPU-mir-155, and might come from two ways: one speculation is it occurs spontaneously; the other is it is a branch of the ancestral miRNA. Intriguingly, NJ tree showed that the two ancestral sequences of mir-155 shared the same nucleotide (Figure 4a, b). The only difference between the two ancestral sequences was that the sequence from D. rerio was longer than the sequence from I. punctatus (Figure 4c). Date have shown the mir-100 is the only microRNA shared by all bilateria (Grimson Srivastava et al., 2008; Griffiths-Jones Hui et al., 2011). Based on the comparison between the mir-100 and mir-155 family in the same species, we found that the original time of mir-155 family was just the time when the same sequences generated in mir-100 family. The original time of the mir-100 and mir-155 families was parallel, and mir-155 family came from the lineage-specific gain of miRNAs, but not from the original miRNAs (Figure 4a).

The mir-155 family is highly conserved in function

To verify whether the arm switching happened in mir-155 or not, we further calculated the “Relative Arm Usage” (Table 1). The 5' arm was highly expressed compared with the 3' arm, which suggested the dominantly mature microRNA was generated from the 5' arm. Based on the
results predicted by miRDB, we found the target sites of 5’ arm were more than those of the 3’ arm, which also supported the inference of arm switching in the mir-155 family. The seed sequence is 2-8 nucleotides of the mature microRNAs which play a crucial role in transcript targeting (Lewis et al., 2003; Nahvi et al., 2009). Based on the results of multiple sequence alignment against the mature sequences, the seed sequence of 5’ arm in the mir-155 family were 100% conserved while of 3’ arm, showed nucleotide diversity (Figure 4c). Therefore, the function of the mir-155 family was highly conserved.

The target gene-pathway-cancer regulatory network of mir-155

There were totally 738 target genes regulated by mir-155 were screened in the current study. All the targets were queried for the KEGG pathway and GO enrichments by using MAS. The KEGG pathway analysis showed mir-155 was involved in a diversity of pathways, especially the various cancers. There were totally 88 pathways, and of them, 11 pathways were presented relation with various cancers (Table 2), such as Colorectal cancer, Pancreatic cancer. Therefore, we constructed the target gene-pathway-cancer regulatory network of mir-155 (Figure 5) to uncover the relationship between mir-155 and cancers. The CCND1, EGFR genes were involved in most of the cancer-related pathways.

The construction of mir-155-target gene-protein regulatory network

Based on the GO analysis, the target genes were enriched in 366 terms of cellular molecular function, 533 of biological processes and 172 of cellular components. With the help of HPRD, there were totally 379 proteins with interaction data with mir-155 target genes were enriched in immune-related GO terms, such as CD31, CRKL and Fgr. The mir-155-target gene-protein regulatory network was shown in Figure 6. The target genes cebpb and VCAM1 and the protein SMAD2 might be the key factors in the mir-155 regulated immune activities.

Discussion

Mir-155 has been reported to play important roles in inhibition of malignant growth, viral infections and even in attenuating the progression of cardiovascular

Table 1. The Arm Switching of mir-155 Family

<table>
<thead>
<tr>
<th>MicroRNA</th>
<th>5’ reads</th>
<th>3’ reads</th>
<th>arm usage</th>
<th>5’ targets</th>
<th>3’ targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>has-mir-155</td>
<td>6099</td>
<td>163</td>
<td>2.094</td>
<td>512</td>
<td>259</td>
</tr>
<tr>
<td>mmu-mir-155</td>
<td>96729</td>
<td>2037</td>
<td>2.198</td>
<td>303</td>
<td>152</td>
</tr>
<tr>
<td>dre-mir-155</td>
<td>183</td>
<td>1</td>
<td>2.784</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>bta-mir-155</td>
<td>118</td>
<td>1</td>
<td>2.593</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The relative arm usage is defined: log2(5’/3’). Positive values indicate a bias toward 5’ arm usage and negative values a bias toward 3’ arm. Zero means that mature sequences are produced at equal levels from both arms (Marco et al., 2010). Targets were predicted by miRDB (Wienholds et al., 2005), and ‘-’ indicate there was no data available.

Table 2. The Enriched Cancer-related Pathways of Target Genes

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Count</th>
<th>p-Value</th>
<th>q-Value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal cancer</td>
<td>15</td>
<td>6.31E-14</td>
<td>1.44E-12</td>
<td>APC; MLH1; MSH2; MSH6; CTNNB1; KRAS; SMAD2; JUN; RAC1; SMAD4; EGFR; SMAD3; CCND1; MET; GSK3B</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>12</td>
<td>5.73E-11</td>
<td>4.06E-10</td>
<td>KRAS; SMAD2; RAC1; SMAD4; CDKN2A; EGFR; SMAD3; CCND1; STAT3; CDK4; NFkB1; E2F2</td>
</tr>
<tr>
<td>Endometrial cancer</td>
<td>10</td>
<td>4.59E-10</td>
<td>2.46E-09</td>
<td>APC; MLH1; FOXO3; CTNNB1; KRAS; CTNNA1; PDK1; EGFR; CCND1; GSK3B</td>
</tr>
<tr>
<td>Melanoma</td>
<td>11</td>
<td>7.05E-10</td>
<td>3.47E-09</td>
<td>FGF7; KRAS; MITF; FGF2; FGFR2; CDKN2A; EGFR; CCND1; MET; CDK4; E2F2</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td>9</td>
<td>1.24E-09</td>
<td>5.48E-09</td>
<td>KRAS; CDKN2A; EGFR; THBS1; CCND1; IL8; CDK4; MMP1; E2F2</td>
</tr>
<tr>
<td>Non-small cell lung cancer</td>
<td>8</td>
<td>2.20E-07</td>
<td>4.92E-07</td>
<td>FOXO3; KRAS; PDK1; CDKN2A; EGFR; CCND1; CDK4; E2F2</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>9</td>
<td>9.84E-07</td>
<td>1.69E-06</td>
<td>CTNNB1; KRAS; PDK1; EGFR; CCND1; CDK2; GSK3B; NFkB1; E2F2</td>
</tr>
<tr>
<td>Renal cell carcinoma</td>
<td>8</td>
<td>1.89E-06</td>
<td>3.03E-06</td>
<td>HIF1A; KRAS; ETS1; JUN; RAP1B; RAC1; PAK2; MET</td>
</tr>
<tr>
<td>Small cell lung cancer</td>
<td>8</td>
<td>8.07E-06</td>
<td>1.17E-05</td>
<td>NOS3; COL4A2; APAF1; CCND1; CDK2; CDK4; NFkB1; E2F2</td>
</tr>
<tr>
<td>Basal cell carcinoma</td>
<td>4</td>
<td>3.97E-03</td>
<td>2.23E-03</td>
<td>APC; CTNNB1; WNT5A; GSK3B</td>
</tr>
<tr>
<td>Thyroid cancer</td>
<td>3</td>
<td>4.67E-03</td>
<td>2.56E-03</td>
<td>CTNNB1; KRAS; CCND1</td>
</tr>
</tbody>
</table>

Evolution of the Mir-155 Family and Possible Targets in Cancers and the Immune System


Conflict of interest: None

Acknowledgements

This work is supported by College of Orient Science and Technology, College of Veterinary Medicine, College of Animal Science And Technology and Horticulture, and Landscape College, Hunan Agriculture University. Conflict of interest: None

References


Marco A, MacPherson JI, Ronshaugen M, Griffiths-Jones S (2012b). MicroRNAs from the same precursor have different targeting properties. Silence, 3, 8-.